

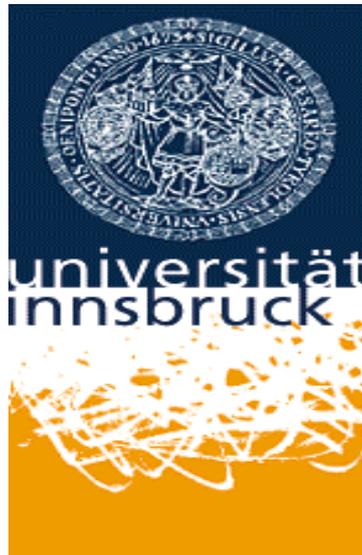
**Gerhard Havranek**

**Häufigkeit und Dauer von Aktivitäten  
Eine Analyse ihrer Wechselwirkungen**

**Diplomarbeit**

**eingereicht an der Leopold-Franzens-Universität Innsbruck  
Baufakultät (Architektur und Bauingenieurwesen)**

**Zur Erlangung des akademischen Grades DIPLOMINGENIEUR**



**Betreuer: o. Univ. Prof. Dr.-Ing. Kay W. Axhausen  
Institut für Strassenbau und Verkehrsplanung,  
Leopold-Franzenz-Universität Innsbruck  
bzw. Institut für Verkehrsplanung und Transporttechnik, Straßen- und  
Eisenbahnbau (IVT), Eidgenössische Technische Hochschule (ETH) Zürich**

**Innsbruck, im Oktober 2002**

# LEOPOLD-FRANZENS-UNIVERSITÄT INNSBRUCK

## VERPFLICHTUNGSERKLÄRUNG

**Titel der Diplomarbeit:**

**Häufigkeit und Dauer von Aktivitäten  
Eine Analyse ihrer Wechselwirkungen**

Verfasser: Gerhard Havranek  
Matrikel-Nr.: 9217000  
Studienrichtung: Bauingenieurwesen  
Studienzweig: Verkehrswesen und Infrastrukturplanung  
Betreuer: o.Univ.Prof. Dr.-Ing. Kay W. Axhausen

Ich erkläre, daß ich meine Diplomarbeit selbständig verfaßt und alle in ihr verwendeten Unterlagen, Hilfsmittel und die zugrundegelegte Literatur genannt habe.

Ich versichere, daß ich dieses Diplomarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin / einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

Ich nehme zur Kenntnis, daß auch bei auszugsweiser Veröffentlichung meiner Diplomarbeit das Institut, an dem die Diplomarbeit ausgearbeitet wurde, und der Betreuer zu nennen sind.

Innsbruck, am

---

Unterschrift

## DANKSAGUNGEN

Mein besonderer Dank gilt Herrn Professor Kay Axhausen einerseits für den Vorschlag des sehr reizvollen und interessanten Themas für diese Diplomarbeit und andererseits natürlich für die ausgezeichnete und geduldige Betreuung, die trotz der großen räumlichen und zeitlichen Distanz hervorragend geklappt hat. Bei den zahlreichen Treffen wurden meine offenen Fragen immer sehr detailliert und hilfreich beantwortet.

Weiters möchte ich mich speziell bei drei Universitätsassistenten an der Eidgenössischen Technischen Hochschule (ETH) in Zürich bedanken: DI Stefan Schönfelder für die prompte und sehr hilfreiche Beantwortung meiner Fragen in bezug auf SAS-Programmierung und Modellierung, DI Arnd König für die Hilfestellung bei Grafiken in SAS und DI Robert Schlich für die Findung interessanter Fragestellungen. Allen dreien und dem gesamten Institut für Verkehrsplanung und Transporttechnik, Straßen- und Eisenbahnbau natürlich einen herzlichen Dank für die gute Aufnahme bei meinen Zürichbesuchen.

Herzlich bedanken möchte ich mich bei meiner Familie, speziell meinen Eltern, für die Unterstützung und Geduld, die sie mir im Laufe dieser Diplomarbeit entgegengebracht haben. Ihnen möchte ich diese Arbeit widmen.

# INHALTSVERZEICHNIS

VERPFLICHTUNGSERKLÄRUNG.....	II
DANKSAGUNGEN.....	III
INHALTSVERZEICHNIS.....	IV
TABELLENVERZEICHNIS .....	VII
ABBILDUNGSVERZEICHNIS.....	VIII
KURZFASSUNG .....	1
ABSTRACT .....	2
<b>1 EINLEITUNG .....</b>	<b>3</b>
<b>2 DATENBESCHREIBUNG UND VERÄNDERUNGEN DER DATEN DER UPPSALA TRAVEL HOUSEHOLD SURVEY .....</b>	<b>5</b>
2.1 EINLEITUNG.....	5
2.2 UMFRAEGEBIET .....	6
2.3 HINTERGRUND.....	6
2.4 DEFINITION ALLER IN DER UMFRAGE VERWENDETEN AUSDRÜCKE .....	7
2.5 KORREKTUR BZW. VERVOLLSTÄNDIGUNG DES DATENSATZES .....	9
2.5.1 <i>Haushaltsnummern und andere logische Umcodierungen</i> .....	9
2.5.2 <i>Verkehrsmittelwahl bzw. Führerscheinbesitz</i> .....	10
2.5.3 <i>Dauer der Aktivitäten</i> .....	10
2.5.4 <i>Neuberechnung der tatsächlichen Arbeitszeiten</i> .....	11
2.5.5 <i>Ergänzung fehlender sozioökonomischer Daten</i> .....	12
2.5.6 <i>Aufteilung des Datensatzes</i> .....	13
2.5.7 <i>Erstellung der Aktivitätsdatei</i> .....	14
2.5.8 <i>Erstellung der Intervalldateien</i> .....	15
<b>3 BISHERIGE FORSCHUNGSARBEITEN MIT DEN DATEN DER UTHS .....</b>	<b>17</b>
3.1 EINLEITUNG.....	17
3.2 DER ARBEITSWEG ALS MEHRZWECKWEG IM STÄDTISCHEN VERKEHR .....	17
3.3 AKTIVITÄTSMUSTER DER GESCHLECHTER UND AUSWIRKUNGEN DER ERWERBSTÄTIGKEIT VON FRAUEN DARAUF .....	20
3.4 ZUGANGSMÖGLICHKEITEN UND INNERSTÄDTISCHER VERKEHR .....	23
3.5 WIEDERHOLUNG UND TAGTÄGLICHE VARIABILITÄT IN KOMPLEXEN REISEMUSTERN.....	24
<b>4 REISE- UND VERKEHRSVERHALTENSMUSTER, MODELLE UND DEREN AUSSEHEN.....</b>	<b>28</b>
4.1 EINLEITUNG UND GESCHICHTLICHER HINTERGRUND .....	28
4.2 WAS SIND AKTIVITÄTSMUSTER GENAU? .....	28
4.3 WARUM BRAUCHEN WIR AKTIVITÄTSMUSTER? .....	29
4.4 WELCHE FAKTOREN BEEINFLUSSEN DIESE AKTIVITÄTSMUSTER ? .....	30

4.4.1	<i>Haushaltseigenschaften</i> .....	31
4.4.2	<i>Lokale und geografische Gegebenheiten</i> .....	32
4.5	BEGRIFFSKLÄRUNG .....	34
4.5.1	<i>Gewohnheitsmässiges Verhalten bzw. Wiederholung</i> .....	34
4.5.2	<i>Variabilität</i> .....	34
4.5.3	<i>Rhythmische Muster und Periodizität</i> .....	35
<b>5</b>	<b>HAZARD MODELLE: EINLEITUNG</b> .....	<b>37</b>
5.1	HINTERGRUND .....	37
5.2	HAZARD MODELLE .....	39
5.3	ANSATZ, AUSWERTUNG UND DARSTELLUNG VON HAZARD MODELLEN .....	41
5.3.1	<i>Einleitung</i> .....	41
5.3.2	<i>Censoring problem</i> .....	43
5.3.3	<i>Time of Origin - Anfangszeitpunkt</i> .....	44
5.3.4	<i>Interpretation der Hazard Funktion</i> .....	45
<b>6</b>	<b>MODELLFORMEN</b> .....	<b>47</b>
6.1	EINLEITUNG .....	47
6.2	NONPARAMETRISCHE MODELLE MIT PROC LIFETEST .....	48
6.2.1	<i>Kaplan-Meier-Schätzwert</i> .....	48
6.2.2	<i>Lifetable- Methode</i> .....	49
6.3	SEMIPARAMETRISCHE MODELLE MIT PROC PHREG .....	51
6.3.1	<i>Proportional Hazard Models</i> .....	51
6.3.2	<i>(Maximum) Partial Likelihood</i> .....	53
6.3.3	<i>Besonderheiten</i> .....	54
6.4	VOLLPARAMETRISCHE MODELLE MIT PROC LIFEREG .....	56
6.4.1	<i>Accelerated Failure Time Model</i> .....	57
6.4.2	<i>MAXIMUM LIKELIHOOD Abschätzung (ML)</i> .....	61
6.4.3	<i>Besonderheiten</i> .....	64
<b>7</b>	<b>STUDIEN MIT HAZARDMODELLEN</b> .....	<b>66</b>
7.1	EINLEITUNG .....	66
7.2	WEITERE STUDIEN MIT DEN DATEN DER UTHS .....	66
7.3	STUDIEN MIT HAZARDMODELLEN ÜBER AKTIVITÄTSDAUERN UND INTERVALLDAUERN .....	69
7.3.1	<i>Arbeitspapier von Lyoong Oh (2000)</i> .....	69
7.3.2	<i>Arbeiten von Schönfelder und Axhausen (2000)</i> .....	72
7.4	WEITERE STUDIEN MIT HAZARDMODELLANALYSEN .....	75
<b>8</b>	<b>AKTIVITÄTSMUSTER</b> .....	<b>77</b>
8.1	EINLEITUNG .....	77
8.2	KLASSIFIZIERUNG DER PERSONEN UND ANDERE CHARAKTERISTIKA .....	77
8.2.1	<i>Beschreibung der Personen- bzw. Haushaltsgruppen</i> .....	77
8.2.2	<i>Arbeitsstatus der Personen</i> .....	80
8.2.3	<i>Arbeitstage und arbeitsfreie Tage</i> .....	81
8.2.4	<i>Aktivitätskategorien</i> .....	81
8.3	DURCHSCHNITTLLICHE AKTIVITÄTSDAUER PRO TAG BZW. AKTIVITÄT .....	82
8.4	AKTIVITÄTSDAUER UND -VERTEILUNG NACH WOCHENTAGEN .....	86
8.5	AKTIVITÄTSDAUERVERTEILUNG AUF ARBEITSTAGE UND FREIE TAGE .....	91
8.6	WEITERE ERGEBNISSE DER DESKRIPTIVEN ANALYSE .....	95
<b>9</b>	<b>HAZARDMODELL - ANALYSE</b> .....	<b>99</b>
9.1	EINLEITUNG .....	99
9.2	DAUER DER AKTIVITÄTEN .....	99
9.2.1	<i>Einleitung</i> .....	99

9.2.2	<i>Nonparametrische Modelle mit PROC LIFETEST</i> .....	100
9.2.3	<i>Semiparametrische Modelle mit PROC PHREG</i> .....	113
9.2.4	<i>Vollparametrische Modelle mit PROC LIFEREG</i> .....	122
9.3	DAUER DER INTERVALLE ZWISCHEN GLEICHARTIGEN AKTIVITÄTEN .....	134
9.3.1	<i>Einleitung</i> .....	134
9.3.2	<i>Nonparametrische Modelle mit PROC LIFETEST</i> .....	134
9.3.3	<i>Semiparametrische Modelle mit PROC PHREG</i> .....	146
9.3.4	<i>Vollparametrische Modelle mit PROC LIFEREG</i> .....	154
<b>10</b>	<b>SCHLUSSFOLGERUNGEN UND WEITERE AUSBLICKE</b> .....	<b>159</b>
10.1	ERGEBNISSE .....	159
10.2	SCHLUSSFOLGERUNGEN .....	161
10.3	WEITERE AUSBLICKE .....	163
	<b>LITERATURVERZEICHNIS</b> .....	<b>164</b>
	 <b>ANHÄNGE</b> .....	 <b>167</b>
<b>ANHANG I :</b>	<b>LISTE ALLER VARIABLEN</b> .....	<b>167</b>
<b>ANHANG II :</b>	<b>AKTIVITÄTENKATEGORISIERUNG</b> .....	<b>178</b>
<b>ANHANG III :</b>	<b>PARTIAL LIKELIHOOD</b> .....	<b>182</b>
<b>ANHANG IV :</b>	<b>LISTE DER MÖGLICHEN EINFLUSSVARIABLEN</b> .....	<b>184</b>
<b>ANHANG V :</b>	<b>HAZARD- UND SURVIVALKURVEN, AKTIVITÄTSDAUERN</b> .....	<b>185</b>
<b>ANHANG VI:</b>	<b>AUSWERTUNGEN VON <i>PROC LIFETEST</i>: WILCOXON UND LOG-RANK-TESTS DER 4 KATEGORIEN, AKTIVITÄTSDAUERN</b> .....	<b>203</b>
<b>ANHANG VII :</b>	<b>ERSTE GESAMTMODELLE DER SEMIPARAMETRISCHEN MODELLSCHÄTZUNGEN, AKTIVITÄTSDAUERN</b> .....	<b>222</b>
<b>ANHANG VIII :</b>	<b>ERSTE GESAMTMODELLE DER VOLLPARAMETRISCHEN MODELLSCHÄTZUNGEN, AKTIVITÄTSDAUERN</b> .....	<b>227</b>
<b>ANHANG IX :</b>	<b>ÜBERSICHTSTABELLEN SIGNIFIKANTE VARIABLEN, AKTIVITÄTSDAUERN</b> .....	<b>232</b>
<b>ANHANG X :</b>	<b>HAZARD- UND SURVIVALKURVEN, INTERVALLDAUERN</b> .....	<b>237</b>
<b>ANHANG XI:</b>	<b>AUSWERTUNGEN VON <i>PROC LIFETEST</i>: WILCOXON UND LOG-RANK-TESTS DER 4 KATEGORIEN, INTERVALLDAUERN</b> .....	<b>256</b>
<b>ANHANG XII :</b>	<b>ERSTE GESAMTMODELLE DER SEMIPARAMETRISCHEN MODELLSCHÄTZUNGEN, INTERVALLDAUERN</b> .....	<b>273</b>
<b>ANHANG XIII :</b>	<b>ERSTE GESAMTMODELLE DER VOLLPARAMETRISCHEN MODELLSCHÄTZUNGEN, INTERVALLDAUERN</b> .....	<b>278</b>
<b>ANHANG XIV :</b>	<b>ÜBERSICHTSTABELLEN SIGNIFIKANTE VARIABLEN, INTERVALLDAUERN</b> .....	<b>284</b>

# TABELLENVERZEICHNIS

Tabelle 8.1 : Verteilung der Geschlechter in den Personengruppen .....	79
Tabelle 8.2 : Verteilung der Haushaltsgröße in den Personengruppen .....	80
Tabelle 8.3 : Verteilung des Arbeitsstatus über die Personengruppen .....	80
Tabelle 8.4 : Verteilung der Personen auf die Arbeitsstatusgruppen .....	81
Tabelle 8.5 : Durchschnittliche Aktivitätsdauer außer Haus pro Person pro Tag in Stunden .....	83
Tabelle 8.6 : Durchschnittliche Aktivitätsdauer außer Haus pro Aktivität in Stunden .....	84
Tabelle 8.7 : Aktivitätshäufigkeit in allen Kategorien, nach Geschlecht gegliedert .....	85
Tabelle 8.8 : Häufigkeitsverteilung aller Aktivitäten nach Wochentagen .....	88
Tabelle 8.9 : Anzahl der Aktivitäten pro Person nach Wochentagen .....	88
Tabelle 8.10 : Häufigkeitsverteilung aller Aktivitäten außer Haus nach Wochentagen .....	89
Tabelle 8.11 : Anzahl der Aktivitäten außer Haus pro Person nach Wochentagen .....	89
Tabelle 8.12 : Durchschnittliche Aktivitätsdauer pro Aktivität nach Kategorien, Geschlecht und Arbeitstag (Samstag frei) .....	92
Tabelle 8.13 : Durchschnittliche Aktivitätsdauer pro Person pro Tag an Werktagen .....	94
Tabelle 8.14 : Durchschnittliche Aktivitätsdauer pro Person pro Tag an freien Tagen .....	94
Tabelle 9.1 : Übersicht der Signifikanz der möglichen Einflußvariablen .....	116
Tabelle 9.2 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Einkauf kurzfristig .....	119
Tabelle 9.3 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Freizeit allgemein .....	119
Tabelle 9.4 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Dienstleistung allgemein .....	120
Tabelle 9.5 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Treffen mit Freunden .....	120
Tabelle 9.6 : Log-Likelihoods der Modelle: .....	124
Tabelle 9.7 : Log-Likelihood-Ratio-Chi-Quadrat-Statistik: .....	124
Tabelle 9.8 : Modell der zweiten Stufe, Kategorie Einkauf kurzfristig .....	130
Tabelle 9.9 : Modell der zweiten Stufe, Kategorie Freizeit allgemein .....	130
Tabelle 9.10 : Modell der zweiten Stufe, Kategorie Dienstleistung allgemein .....	131
Tabelle 9.11 : Modell der zweiten Stufe, Kategorie Treffen mit Freunden .....	131
Tabelle 9.12 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Einkauf kurzfristig .....	151
Tabelle 9.13 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Freizeit allgemein .....	152
Tabelle 9.14 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Kategorie Dienstleistung allgemein .....	152
Tabelle 9.15 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Treffen mit Freunden .....	153
Tabelle 9.16 : Modell der zweiten Stufe, Kategorie Einkauf kurzfristig .....	155
Tabelle 9.17 : Modell der zweiten Stufe, Kategorie Freizeit allgemein .....	155
Tabelle 9.18 : Modell der zweiten Stufe, Kategorie Dienstleistung allgemein .....	156
Tabelle 9.19 : Modell der zweiten Stufe, Kategorie Treffen mit Freunden .....	156

# ABBILDUNGSVERZEICHNIS

Abbildung 2.1 : Musterfragebogen aus der Uppsala Travel Household Survey (UTHS).....	8
Abbildung 4.1 : Basiskonzept: Rhythmik des Verkehrsverhaltens .....	36
Abbildung 8.1 : Verteilung der Personen und Haushalte auf die Personengruppen in Prozent .....	79
Abbildung 8.2 : Verteilung der Aktivitätsdauern in der Kategorie „Arbeit“ .....	84
Abbildung 8.3 : Aktivitätshäufigkeit in allen Kategorien, nach Geschlecht gegliedert .....	85
Abbildung 8.4 : Durchschnittliche Aktivitätsdauern nach Wochentagen .....	86
Abbildung 8.5 : Durchschnittliche Aktivitätsdauern außer Haus nach Wochentagen .....	87
Abbildung 8.6 : Anzahl der Aktivitäten pro Person nach Wochentagen.....	89
Abbildung 8.7 : Häufigkeitsverteilung aller Aktivitäten außer Haus nach Wochentagen .....	90
Abbildung 8.8 : Dauer und Häufigkeit aller Aktivitäten außer Haus pro Person pro Tag nach Wochentagen und Geschlecht .....	91
Abbildung 8.9 : Durchschnittliche Dauer der Aktivität an Arbeitstagen und freien Tagen .....	92
Abbildung 8.10 : Aktivitätsdauer nach Kat. außer Haus, Geschlecht und Arbeitstag, Sa = frei.....	93
Abbildung 8.11 : Verkehrsmittelwahl nach Kategorien.....	95
Abbildung 8.12 : Verkehrsmittelwahl nach Geschlecht.....	96
Abbildung 8.13 : Durchschnittliche Anzahl von Wegen pro Tag und Person nach Verkehrsmitteln...	97
Abbildung 9.1 : Gesamte Aktivitäten.....	104
Abbildung 9.2 : Arbeit .....	104
Abbildung 9.3 : Einkauf langfristig .....	104
Abbildung 9.4 : Einkauf kurzfristig .....	104
Abbildung 9.5 : Freizeit allgemein.....	104
Abbildung 9.6 : Dienstleistung allgemein.....	104
Abbildung 9.7 : Erledigungen .....	105
Abbildung 9.8 : Besuch bei Verwandten.....	105
Abbildung 9.9 : Treffen mit Freunden .....	105
Abbildung 9.10 : Verein, Initiative, Kirche.....	105
Abbildung 9.11 : Sport (aktiv) .....	105
Abbildung 9.12 : Spaziergang / -fahrt .....	105
Abbildung 9.13 : Etwas Neues sehen .....	106
Abbildung 9.14 : Lokal, Kino, Theater .....	106
Abbildung 9.15 : Geglättete Hazardkurve der Altersschichtung, Kategorie Einkauf kurzfristig.....	107
Abbildung 9.16 : Geglättete Hazardkurve der Altersschichtung, Kategorie Freizeit allgemein .....	108
Abbildung 9.17 : Geglättete Hazardkurve der Altersschichtung, Kategorie Dienstleistung allg. ....	108
Abbildung 9.18 : Geglättete Hazardkurve der Altersschichtung, Kategorie Treffen mit Freunden....	109
Abbildung 9.19 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Einkauf kurzfristig .....	110
Abbildung 9.20 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Freizeit allgemein .....	110
Abbildung 9.21 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Dienstleistung allg. ....	111
Abbildung 9.22 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Treffen mit Freunden....	111
Abbildung 9.23 : Plot für die Evaluierung des Loglogistischen Modells, Einkauf kurzfristig .....	125
Abbildung 9.24 : Plot für die Evaluierung des Lognormalen Modells, Einkauf kurzfristig .....	125
Abbildung 9.25 : Cox-Snell-Residuenplot mit einer loglogistischen Verteilung, mit Kovariablen....	126
Abbildung 9.26 : Cox-Snell-Residuenplot mit einer lognormalen Verteilung, mit Kovariablen.....	126
Abbildung 9.27 : LIFEREG- Auswertung: Plot der Hazardkurve vs Zeit, Loglogistic, mit Kovariablen .....	127
Abbildung 9.28 : LIFEREG- Auswertung: Plot der Hazardkurve vs Zeit, Lognormal, mit Kovariablen .....	127
Abbildung 9.29 : Kategorie Einkauf kurzfristig .....	136
Abbildung 9.30 : Kategorie Freizeit allgemein .....	136
Abbildung 9.31 : Kategorie Dienstleistung allgemein .....	137

Abbildung 9.32 : Kategorie Treffen mit Freunden.....	137
Abbildung 9.33 : Arbeit .....	139
Abbildung 9.35 : Erledigung von Pflichten.....	139
Abbildung 9.37 : Verein, Initiative, Kirche.....	139
Abbildung 9.34 : Einkauf langfristig.....	139
Abbildung 9.36 : Besuch bei Verwandten.....	139
Abbildung 9.38 : Sport (aktiv) .....	139
Abbildung 9.39 :Spaziergang / -fahrt.....	140
Abbildung 9.40 : Lokal, Kino, Theater .....	140
Abbildung 9.41 : Aktivitäten allgemein .....	140

## KURZFASSUNG

Die Untersuchung des Verkehrsverhaltens verschiedener Personen und Haushalte mit Hilfe von Reise- und Aktivitätsmustern stellt seit jeher einen wichtigen Bestandteil der Verkehrsforschung dar. Bildeten früher ausschließlich Wege und Reisen die Grundlagen der Analyse, so wird seit Beginn der 90-er Jahre vermehrt mit aktivitäts-basierenden Ansätzen (*activity based approach*) gearbeitet. Dabei stützt man sich auf sogenannte Zeitdauerdaten (*duration data*), die sowohl die Abfolge, als auch die Dauer der Aktivitäten der befragten Personen während des Untersuchungszeitraums beinhalten und damit Aufschluß über die zeitlichen Aspekte des Verkehrsverhaltens geben. Von besonderem Interesse sind die Aktivitäten außer Haus, da sie eine Ortsveränderung notwendig machen und Verkehr induzieren.

Für diese Arbeit wurde die Hazardmodellschätzung aus einer Reihe von Ansatzmöglichkeiten gewählt, die auf der gleichnamigen Funktion und dadurch auf bedingten Wahrscheinlichkeiten aufbaut. Mit diesen Modellschätzungen wird versucht, die Rhythmen der Aktivitätsgestaltung verschiedener Personengruppen und Bevölkerungsschichten genauer zu beschreiben und das Auftreten von Ereignissen, hier Aktivitäten, in Abhängigkeit der bis zu diesem Moment vergangenen Zeitspanne, vorherzusagen. Grundlage der Analysen bildet eine Reisebefragung an 35 aufeinanderfolgenden Tagen im Jahr 1971 aus Uppsala in Schweden, bei der 144 Personen über alle unternommenen Wege Bericht erstatteten.

Nach einer ausführlichen deskriptiven Analyse werden die Hazardmodellschätzungen in drei Stufen dargestellt (nonparametrischer, semiparametrischer und vollparametrischer Ansatz), die mit Hilfe von Programmen des Statistikpaketes SAS durchgeführt wurden. Dabei sollten etwaige Einflußfaktoren - sowohl aus dem soziodemographischen, als auch dem aktivitätsbezogenen Bereich - auf die Aktivitätsdauer und die Dauer der Intervalle zwischen gleichartigen Aktivitäten außer Haus identifiziert werden. Zudem soll der tatsächliche Einfluß auf die zeitliche Abfolge und die Dauer bestimmter Aktivitäten geschätzt werden.

Im Vergleich zu früheren Studien zeigt diese Analyse einen weitaus weniger signifikanten Einfluß der soziodemographischen Faktoren auf die Länge der Aktivitätsdauern und Intervalldauern. Das Geschlecht und verschiedene Kenngrößen der Mobilität der Personen erweisen sich jedoch in vielen Fällen als signifikant. Bei den aktivitätsabhängigen Variablen spielen die Reisezeiten zu den Aktivitätsorten, die vorher unternommenen Aktivitäten, monetäre Ausgaben während der Aktivität und die Tageszeit in den vier näher untersuchten Aktivitätskategorien Einkauf kurzfristig, Freizeit allgemein, Dienstleistungen und Treffen mit Freunden eine signifikante Rolle.

## ABSTRACT

The analysis of travel behaviour of different persons and households based on travel diaries and activity patterns has always been a main field of investigation in transportation research. While earlier studies exclusively analysed trips and journeys, researchers in the 90's focused on an activity based approach using duration data. Such data include both sequence and duration of activities of all informants throughout the whole study period thus providing interesting insights into temporal aspects of individual travel behaviour. Out-of-home activities are of special interest because they necessitate spatial changes and induce traffic.

For the present study hazard model estimations which are based on the hazard function and conditional probabilities were chosen. Hazard models are used to describe rhythms of activity patterns of various groups sharing social or activity-related characteristics. In addition the occurrence of events, in this case beginning or end of activities, is predicted contingent on the activity duration or the duration of the interval since the last activity. The study utilizes a data set collected in Uppsala, Sweden, in 1971. It includes travel diaries of 144 persons over a period of 35 consecutive days.

Following a comprehensive descriptive analysis hazard model estimations are presented at three levels (nonparametric, semi-parametric and fully parametric approach); the calculations were carried out with the statistics package SAS. A detailed analysis was undertaken for the four activity categories „shortterm shopping“, „leisure“, „personal business“ and „meeting friends“. The purpose was to identify sociodemographic and activity-related factors influencing activity duration and duration of intervals between similar out-of-home activities. Moreover the magnitude of the influence of these factors on temporal sequence and duration of specific activities is estimated.

The influence of sociodemographic factors on duration of activities and intervals turns out to be far less significant in the present analysis than in earlier studies. However, sex and various parameters of people's mobility prove to be significant in many cases. Among the activity-related variables duration of traveltime to activity locations, previous activities, activity expenditures and time of day are statistically significant.

# 1 EINLEITUNG

Die Vorhersage von Mobilitätsbedürfnissen der Bevölkerung und den Abläufen ihrer tagtäglichen Verkehrswege ist ein Bestandteil der Analyse des Reiseverhaltens, der in den letzten Jahren immer wichtiger geworden ist. Früher waren sehr viele Studien nur auf die Wege bzw. Reisen verschiedener Personen fixiert und haben diese genauer analysiert. Wie aber Axhausen und Polak (1991) feststellten, sind es vor allem die Aktivitäten, die genauer beleuchtet und analysiert werden müßten:

*“...ABA - (activity based approach) argued that observed patterns of trip making should be seen as the consequence of individual's desire to participate in spatially and temporally distinct activities and the analysis of travel behaviour should therefore be based on an understanding of the linked sequence of activities in which people engage during a day.”*

Diese Theorie, daß Wege meist nur gemacht werden, um gewisse persönliche bzw. individuelle Bedürfnisse an verschiedenen Orten befriedigen zu können - sozusagen notwendiges Übel sind, da nicht alle Aktivitäten an einem Ort ausgeführt werden können und so das Verkehrsverhalten eines jeden Individuums zeitlichen und räumlichen Restriktionen bzw. Randbedingungen der dynamischen Umwelt unterliegt - wurde stark von der *time geography* von Hägerstrand (1970) und *use of urban land* von Chapin (1974) beeinflusst.

Die Intention dieser Arbeit ist es, die Rhythmen der Aktivitätsgestaltung verschiedener Personengruppen und Bevölkerungsschichten mit Hilfe von Zeitdauermodellen, die auf Hazardfunktionen basieren, darzustellen und geeignete Modelle zu schätzen. Dadurch sollten etwaige Einflußfaktoren gefunden werden und zudem deren tatsächlicher Einfluß auf die Dauer und die zeitliche Abfolge bestimmter Aktivitäten bzw. Aktivitätskategorien geschätzt werden. Die Grundlage der Modellschätzungen mit drei verschiedenen Stufen bildet eine Reisebefragung über 35 aufeinanderfolgende Tage aus Uppsala in Schweden aus dem Jahr 1971, bei der 144 Personen detailliert über alle unternommenen Wege Bericht erstatteten. Dieser großartige Datensatz der Uppsala-Umfrage (*Uppsala Travel Household Survey*, kurz UTHS) wurde schon in einer Reihe von Studien ausgewertet. Durch die Verwendung anderer Analysemethoden sollen keineswegs die Ergebnisse bisheriger Studien korrigiert bzw. in Frage gestellt werden. Vielmehr soll diese Arbeit einen weiteren Beitrag zum besseren Verständnis der komplexen Reise- und Aktivitätsverhaltensmuster der Umfrageteilnehmer aus Uppsala leisten.

Im Kapitel 2 werden das Umfragegebiet und die im Jahr 1971 erhobenen Daten aus Uppsala vorgestellt und zudem die für die Auswertungen nötigen Veränderungen detailliert beschrieben. Im Kapitel 3 werden einige Forschungsarbeiten, die mit den Daten der UTHS durchgeführt wurden, präsentiert. Von besonderer Bedeutung sind dabei die Arbeiten von S.E. Hanson, P.O. Hanson und Huff.

Im vierten Kapitel (4) werden grundlegende Konzepte der Reise- und Verkehrsverhaltensforschung besprochen. Dabei werden einige der wichtigsten Grundbegriffe wie gewohnheitsmäßiges Verhalten und Variabilität erklärt. Zudem werden jene Faktoren analysiert und klassifiziert, die die Verhaltensmuster verschiedener Personen beeinflussen können.

Das Kapitel 5 beschäftigt sich mit den Grundlagen der in der Arbeit verwendeten Hazardmodelle und der Überlebensanalyse (*survival analysis*). So werden neben den bisherigen Anwendungsgebieten von Hazardmodellen die verschiedenen Verteilungsfunktionen und Modellansätze vorgestellt. Im Kapitel 6 werden dann einige der möglichen Modelle besprochen und ihre Anwendung in der EDV bzw. insbesondere die praktische Ausführung mit Hilfe des Statistikprogrammes SAS detailliert beschrieben. Das Kapitel 7 bietet einen kurzen, wenn auch nicht umfassenden Überblick über Forschungsarbeiten, die in den letzten Jahren mit Hilfe solcher Hazardmodelle gemacht worden sind und geht vor allem auf Studien, deren Themenstellungen ähnlich dieser Arbeit sind, genauer ein.

Kapitel 8 und 9 stellen die eigenständige Analyse der UTHS-Daten dar. Im Kapitel 8 werden die Aktivitätsmuster und diverse andere Charakteristika der befragten Personen aus Uppsala nach verschiedenen Gesichtspunkten zusammengefaßt und diskutiert. Unterschiede zwischen Gesellschaftsgruppen, Altersgruppen und den beiden Geschlechtern werden dabei unter anderem auch mit Hilfe von Abbildungen und Tabellen dargestellt. Im Kapitel 9 werden sowohl für die Aktivitätsdauern, als auch für die Intervalldauern, also die Zeitspannen zwischen zwei gleichartigen Aktivitäten, die Auswertungen der Hazardmodelle in jeweils drei Stufen präsentiert. Für vier ausgewählte Aktivitätskategorien (Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden) werden zudem detailliert alle Ergebnisse der nonparametrischen, semiparametrischen und vollparametrischen Modellansätze dargestellt und jeweils im Anschluß diskutiert.

Der letzte Abschnitt, Kapitel 10, beinhaltet eine Zusammenfassung der erzielten Ergebnisse und etwaige Unterschiede zu anderen, ähnlich gelagerten Studien. Zudem wird auf die Schwierigkeiten eines direkten Vergleichs dieser Ergebnisse mit denen neuerer Datensätze hingewiesen. Den Abschluß bildet ein Ausblick auf mögliche weiterführende Forschungsarbeiten mit den hier verwendeten Modellansätzen.

## **2 DATENBESCHREIBUNG UND VERÄNDERUNGEN DER DATEN DER UPPSALA TRAVEL HOUSEHOLD SURVEY**

### **2.1 EINLEITUNG**

Die in dieser Diplomarbeit verwendeten Daten stammen aus der Uppsala Haushalts-Verkehrsbefragung (Uppsala Household Travel Survey, kurz UTHS), einer Umfrage über das Reiseverhalten von Personen bzw. Haushalten, die in Uppsala (Schweden) im Frühjahr 1971 durchgeführt wurde (Marble, P. Hanson und S. Hanson, 1972).

Über eine Beobachtungszeit von 5 Wochen bzw. 35 aufeinander folgenden Tagen führten 144 erwachsene Personen (über 16 Jahre) aus 92 Haushalten, die zufällig ausgewählt worden waren, detailliert Buch über alle Aktivitäten die sie außerhalb ihres Zuhauses unternahmen. Aufgrund einerseits sehr guter Betreuung durch die Umfragemitarbeiter (jeden zweiten Tag telefonischer Kontakt, jede Woche einmal persönlich) und andererseits Anreize in Form einer Tombola für alle zu Ende geführten Reisetagebücher am Schluß, ist in den Daten kaum eine Verschlechterung über die Zeit zu erkennen bzw. weisen nur etwa 5 Prozent der Personen in der Umfrage einen konstanten Abwärtstrend in der Anzahl der Wege pro Tag über die gesamte Umfragedauer gesehen auf (Hanson und Huff, 1982).

Die Uppsala Travel Household Survey stellt somit einen im Gegensatz zu anderen Umfragen unvergleichbaren Längsschnitt-Datensatz zur Verfügung, der erlaubt, speziell die räumlichen und zeitlichen Strukturen der tagtäglichen Bewegungen von Haushaltsmitgliedern in der Stadt zu beschreiben und zu analysieren. Sowohl durch die Dauer als auch die detaillierte Beschreibung aller zurückgelegten Wege (auch zu Fuß) und durchgeführten Aktivitäten unterscheidet er sich grundsätzlich von ähnlichen Umfragen (Huff und Hanson, 1986).

## 2.2 UMFRAEGEBIET

Uppsala ist die Hauptstadt der Uppsala Kommune und dessen administratives Zentrum. Sie liegt etwa über 70 km nordnordwestlich von Stockholm in einer Tiefebene und hatte zur Zeit der Umfrage etwa 120.000 Einwohner (2002 ca. 175.000), davon etwa 25.000 Studenten (Hanson und Hanson, 1980). Die Ausdehnung der Stadt erfolgt vor allem in drei Richtungen: in Richtung Süden zum See Ekoln, in nordöstlicher Richtung nach Vaksala und im Norden wächst sie mit Gamla Uppsala zusammen (heutzutage fast fließender Übergang).

Im geografischen, zugleich historischen und wirtschaftlichen Zentrum von Uppsala (CBD) herrschte Kfz-Fahrverbot und nur mit dem Fahrrad oder dem äußerst effizienten und gut ausgebauten Bussystem, der einzigen Form öffentlicher Verkehrsmittel in Uppsala, gelangt man von allen Richtungen in den wichtigsten Teil der Stadt.

## 2.3 HINTERGRUND

Der Grund, wieso gerade die schwedische Stadt Uppsala ausgewählt wurde, waren die guten Voraussetzungen, da von schwedischen Wissenschaftlern und Behörden schon früher Untersuchungen durchgeführt wurden und zudem ein geeignetes Bevölkerungsregister (Swedish Population Register) existierte. Zudem war die Tatsache wichtig, daß der Bezirk Uppsala (Uppsala Kommune) Ende der 60er Jahre Testgebiet für ein computergestütztes Landvermessungssystem war, das genaue Koordinateninformationen auf Basis einzelner Parzellen enthielt (Marble *et al.*, 1972).

Das Bevölkerungsregister beinhaltet für jeden Haushalt in allen Bezirken Schwedens Informationen über Name, Alter, Geschlecht und Adresse jedes einzelnen Haushaltsmitglieds. Dadurch konnten die Haushalte schon vor der Umfrage in sechs verschiedene Lebenszykluskategorien (life cycle groups) eingeordnet werden (beispielsweise beinhaltet Gruppe 2 ältere Personen ohne Kinder, Gruppe 3 junge Singles oder Gruppe 6 Haushalte mit mindestens je einem Erwachsenen zwischen 18 und 49 und einem Kind unter 7 Jahren, siehe auch Kapitel 8). Eine räumliche Strukturierung der Haushalte wurde hingegen nicht vorgenommen. Aus jeder Gruppe wurden 20 Haushalte zufällig gewählt und davon 13, die sich zu einer Teilnahme bereit erklärten, beobachtet (sollte ein Haushalt ausfallen, waren immer noch genügend zufällig gewählte als Reserve vorhanden). Jede dieser 6 Gruppen wurde dann erneut in 5 Ströme unterteilt, wobei jeder Strom an einem anderen Tag mit den Reisetagebüchern begann, so daß keine Abhängigkeit vom Wochentag gegeben war.

Die teilnehmenden Personen (gesamt 144) führten danach an 35 aufeinanderfolgenden Tagen (bei einigen Personen bis zu 37 Tagen) kontinuierlich Buch über jegliche Aktivitäten außer Haus. Die Osterferien bzw. Ostertage fielen genau in die Umfrageperiode, so daß einige Personen speziell in diesen Tagen freizeitdominierte Aktivitätsmuster zeigen, die stärker von ihrem täglichen arbeitsbezogenen Verhaltensmustern abweichen.

## 2.4 DEFINITION ALLER IN DER UMFRAGE VERWENDETEN AUSDRÜCKE

Jedes Verlassen der eigenen Wohnung initiiert eine Reise, die als Serie von Bewegungen zu verschiedenen Orten, an denen die Person etwas zu erledigen hatte oder diverse Aktivitäten ausführte, verstanden werden kann. Jede Reise hatte als Ausgangspunkt und Endpunkt die eigenen vier Wände. Die einzelnen Reisen unterscheiden sich natürlich in ihrer Komplexität durch die Anzahl der zurückgelegten Wege und der ausgeführten Aktivitäten.

Die Teilnehmer mußten nun für jeden Weg jeder Reise folgende Informationen in ihrem Reisetagebuch niederschreiben: Abfahrtszeit von bzw. Ankunftszeit an einem Ort, Art des besuchten Ortes (insgesamt wurden diese in 99 verschiedene Kategorien eingeteilt), die durchgeführte Aktivität oder Aktivitäten (wurden in 56 Kategorien gegliedert), den aufgewendeten Geldbetrag (falls Ausgaben getätigt wurden), das benutzte Verkehrsmittel, Angabe über andere eventuell mitfahrenden Haushaltsmitglieder und, ob der Weg schon zu Hause geplant worden war oder nicht. Unüblich (bzw. in vorherigen Umfragen nie berücksichtigt) war, dass dieser Datensatz alle Bewegungen, auch jene die zu Fuß oder mit dem Fahrrad zurückgelegt worden waren, beinhaltet. Alle Ziel- und Ausgangsorte wurden aufgrund der angegebenen Adressen in x-y Koordinaten geocodiert.

Einen Musterfragebogen zeigt Abbildung 2.1.

Zusätzlich zu den Reisedaten und den schon aus dem Bevölkerungsregister bekannten Daten wurden noch weitere soziodemografische Informationen zum Haushalt und zu den einzelnen Mitgliedern erhoben. Diese beinhalteten den Besitz eines oder mehrerer Kraftfahrzeuge im Haushalt, Besitz eines Führerscheines, Erwerbstätigkeit (plus Angabe der Arbeitsstunden außer Haus pro Woche), Grad der Ausbildung, Familienstand, Alter und Geschlecht jedes Haushaltsmitgliedes, Besitzverhältnisse, Haushaltsgröße und gesamtes Haushaltseinkommen.

Abbildung 2.1 : Musterfragebogen aus der Uppsala Travel Household Survey (UTHS)

Name Jan Svenson Date 16 / 4 1971When did you leave home? 7.50 hoursIs this a continuation from another sheet? Yes  No

Stop number <u>1</u>	Did you plan to make this stop when you left home?	<input checked="" type="checkbox"/> Yes	No
Means of Travel	1 foot 5 car (passenger)	2 bicycle 6 taxi	3 bus 7 moped
		<input checked="" type="checkbox"/> 4 car (driver)	8 other _____
Where you accompanied by someone from your household?	Yes <input checked="" type="checkbox"/> No	If yes, by how many? _____	
Where did you make this stop? (please give address)	<u>V. Ågatan</u>	New? <input type="checkbox"/>	
When did you arrive at this place?	<u>7.55</u> hours	When did you leave this place?	<u>12.05</u> hours
What did you do at this place?	1) <u>worked</u>		Expenditure _____
	2) <u>took a coffee break</u>		<u>1:00</u>
	3) _____		_____
	4) _____		_____

Stop number <u>2</u>	Did you plan to make this stop when you left home?	<input checked="" type="checkbox"/> Yes	No
Means of Travel	<input checked="" type="checkbox"/> 1 foot 5 car (passenger)	2 bicycle 6 taxi	3 bus 7 moped
		4 car (driver)	8 other _____
Where you accompanied by someone from your household?	Yes <input checked="" type="checkbox"/> No	If yes, by how many? _____	
Where did you make this stop? (please give address)	<u>Drottninggatan 12</u>	New? <input type="checkbox"/>	
When did you arrive at this place?	<u>12.10</u> hours	When did you leave this place?	<u>12.55</u> hours
What did you do at this place?	1) <u>ate lunch with an associate</u>		Expenditure _____
	2) _____		_____
	3) _____		_____
	4) _____		_____

Is this trip continued on the next sheet?  Yes No

If No, fill in the section below.

When did you return home?	_____ hours		
Means of Travel	1 foot 5 car (passenger)	2 bicycle 6 taxi	3 bus 7 moped
		4 car (driver)	8 other _____
Where you accompanied by someone from your household?	Yes No	If yes, by how many? _____	

## 2.5 KORREKTUR BZW. VERVOLLSTÄNDIGUNG DES DATENSATZES

Da der Datensatz (wie heute vorhanden) an einigen Stellen fehlerhaft war - manche Variablenwerte waren unlogisch bzw. fehlten gänzlich - und einige Strukturveränderungen für die spätere Auswertung notwendig waren, wurde der Datensatz vor der Analyse noch einmal mit Hilfe des Statistikprogrammes SAS bearbeitet.

Zudem wurde der gesamte Datensatz in vier einzelne Datensätze geteilt, die durchwegs verschiedenartige Variablen enthielten und auf anderen Gesichtspunkten aufgebaut waren. Für die Analyse mit Hazard Modellen oder linearer Regression wurde noch ein Aktivitätsdatensatz mit einigen beobachtungsübergreifenden Variablen, sowie Aktivitätsdauer und Intervalldauern generiert.

Im Folgenden werden die wichtigsten Veränderungen kurz beschrieben.

### 2.5.1 Haushaltsnummern und andere logische Umcodierungen

Bei insgesamt drei Haushalten, die der sozialen Gruppe 3 angehörten (Einzel-Haushalt) waren jeweils zwei Personen mit Wegen angegeben. Da in einem Einzelhaushalt jedoch nicht zwei Personen leben können, wurde die jeweils zweite Person dieser drei Haushalte als eigener Haushalt (Einzelhaushalt) definiert, sie blieb jedoch im gleichen Strom (gleicher Anfangstag der Umfrage). Änderungen betrafen die Personen mit der Personenummer 13012 (jetzt id 13071), 13032 (jetzt id 13081) und 33032 (jetzt id 33011). Deswegen ist auch die Anzahl der Haushalte im Gegensatz zu den von Hanson und Hanson behandelten Daten auf 95 angestiegen.

Bei ebenfalls drei Haushalten war die Anzahl der Haushaltsmitglieder mit 0 angegeben, die wahre Anzahl wurde genau ermittelt bzw. bei Familien mit Kindern, deren Wege in der Umfrage ja nicht gesondert angeführt sind, auf die Mindestanzahl gesetzt.

In der Codierung des jeweiligen Anfangstages der Umfrage wurde bei einigen Personen ein nicht möglicher Tag eingegeben (beispielsweise kann jemand aus dem Strom 6 nicht am 4. Tag der Umfrage beginnen.) Aus diesem Grund wurde bei all diesen Personen der Anfangstag dahingehend verändert, daß die erste Reise frühestens am ersten Umfragetag des Stromes stattfand.

Bei einigen Personen wurden zwei unterschiedliche Wege mit der selben Wegenummer bezeichnet oder teilweise vertauscht. Aus diesem Grund mußten bei insgesamt 18 Wegen Änderungen die Wege- oder Reisennummer betreffend vorgenommen werden.

Aufgrund der Reisegeschwindigkeitsberechnung bzw. der zurückgelegten Entfernungen oder aufgrund von Vergleichen innerhalb des Personenreisemusters wurde bei einigen Wegbeobachtungen das benutzte Verkehrsmittel (*mode*) geändert. Zudem wurden einige mit Tippfehlern eingegebene Koordinaten, die zu falschen bzw. unrealistischen Wegdistanzen führten, korrigiert. Ebenfalls durch Vergleiche mit ähnlichen Wegen der Person wurden weitere Koordinatenangaben geändert, um realistische Distanzen und Reisegeschwindigkeiten ermitteln zu können.

Bei der genaueren Analyse der Aktivitätsmuster konnten auch insgesamt acht Wegbeobachtungen ermittelt werden, deren Aktivitäts-Kategorisierung nicht stimmen konnte. Vier davon wurden von der Kategorie 0 (Nachhauseweg) auf die Kategorien 1 (fehlende Information) oder Kategorie 66 (Spaziergang) umkodiert. Die restlichen vier Beobachtungen wurden aufgrund der gleichzeitigen Aktivität des (Lebens-) Partners diesem angeglichen.

### **2.5.2 Verkehrsmittelwahl bzw. Führerscheinbesitz**

Einige der Befragten hatten bei manchen Wegen angegeben, diesen mit dem Auto als Fahrer zurückgelegt zu haben, obwohl die zusätzlich erhobenen sozioökonomischen Daten entweder gar keinen Führerscheinbesitz ausgaben oder in der entsprechenden Rubrik „nicht unterwegs“ eingetragen war. Bei diesen Personen wurde die Anzahl der Wege mit Verkehrsmittel 4 (Auto als Fahrer) ermittelt.

Falls weniger als 5 Wege auftraten, wurden die Angaben in den jeweiligen Kategorien auf „nicht gefahren“ und „kein Führerscheinbesitz“ korrigiert. Zudem wurden das benutzte Verkehrsmittel von 4 auf 5 geändert (Mitfahrer anstatt Autofahrer). Traten mehr als 5 Wege auf, wurde bei diesen Personen der Führerscheinbesitz als positiv eingetragen, die Frage, ob gefahren wurde, wenn nötig, auch mit ja beantwortet.

### **2.5.3 Dauer der Aktivitäten**

Bei jedem Halt wurde mit Hilfe der Ankunftszeit und der Abfahrtszeit an einem Ort die Dauer der Aktivität berechnet. In einigen Fällen endeten die Reisen oder Wege kurz nach Mitternacht oder auch erst im Laufe des nächsten Tages. Bei diesen Beobachtungen war das Datum der Ankunft jedoch nicht

genau angegeben, die Dauer der Reise mit höchstens einem Tag. Die ursprünglich negativen Ergebnisse der Aktivitätsdauerbetrachtung wurden erneut berechnet und zudem die Variable *jourdays* auf den Wert 1 gesetzt, falls die Aktivität nach 16.00 begonnen wurde.

Bei wenigen Beobachtungen waren unübersehbare Eingabefehler die Ankunfts- und Abfahrtszeit betreffend vorhanden und diese Verwechslungen wurden korrigiert, um positive Dauern der Aktivitäten zu erlangen. Bei jenen Wegen, die zu Hause endeten und erst an einem der darauffolgenden Kalendertage wieder fortgesetzt wurden, wurde der Wert der Aktivitätsdauer (falls notwendig) in der Wege-datei prinzipiell auf Null gesetzt, für die Aktivitätsdatei war jedoch eine exakte Angabe aller Aktivitätsdauern notwendig. Deshalb mußten auch einige dieser Aktivitäten neu berechnet werden. Bei den restlichen Beobachtungen mit ursprünglich negativen Werten der Aktivitätsdauer, die nicht auf Anheb logisch korrigiert werden konnten, wurde der Wert vorerst als fehlend (*missing value*) gesetzt.

Anschließend wurde für jede Person die Dauer der Aktivitäten in jeder einzelnen Kategorie aufsummiert. Die für Arbeit, Lebensmitteleinkäufe, Ausbildung und Einkäufe in Damenbekleidungsgeschäften aufgewendeten Zeiten wurden genauer betrachtet und als Grundlage für weitere Ergänzungen in die Datensätze genommen.

In der späteren Auswertung (SAS-Programme *recode\_I*, *split\_I* und *create\_act\_I*) mußten noch weitere Korrekturen bei sowohl negativen bzw. unlogischen Aktivitätsdauern als auch bei negativen Intervallauern zwischen zwei Aktivitäten gemacht werden (siehe Kapitel 2.5.7).

## 2.5.4 Neuberechnung der tatsächlichen Arbeitszeiten

Auf Grund der ermittelten gesamten Dauer der für die Aktivität „Arbeit“ aufgewendeten Stunden und der Anzahl der Tage, an denen gearbeitet wurde, wurde die Arbeitszeit pro Woche noch einmal berechnet. Einkalkuliert wurden auch die Ostertage bzw. jene Tage, die sich Arbeitnehmer frei genommen hatten. Dabei ergaben sich zum Teil gravierende Unterschiede zu den ursprünglich angegebenen Arbeitsstunden pro Woche. Aufgrund dieser neu ermittelten Arbeitsstunden wurden die Personen in vier verschiedene Beschäftigungsstufen eingeteilt:

Stufe 0 - nicht außer Haus arbeitend

Stufe 1 - Vollzeit beschäftigt, über 20 Stunden bzw. öfter als 15 Tage in den 5 Wochen

Stufe 2 - Teilzeit beschäftigt, zwischen 5 und 20 Stunden pro Woche

Stufe 3 - Marginal beschäftigt, zwischen 0.3 und 5 Stunden pro Woche

In jeder dieser Beschäftigungsstufen wurde mit Hilfe von SAS - Tabellen das Durchschnittsalter getrennt für Frauen und Männer berechnet.

### **2.5.5 Ergänzung fehlender sozioökonomischer Daten**

Bei den erhobenen sozioökonomischen Daten war das Fehlen einiger Variablenwerte festzustellen. Vor allem in den Kategorien Geburtsjahr (bzw. daraus folgend Alter), Geschlecht und Arbeitsstunden pro Woche war der Datensatz nicht vollständig. Die Leerstellen in der Kategorie Arbeitsstunden pro Woche konnten mit den vorher beschriebenen Berechnungen vervollständigt werden. Für die beiden anderen Kategorien wurden neben den ermittelten Durchschnittswerten auch verschiedene, teils konservativ anmutende Hierarchien verwendet.

Zuerst wurde - falls vorhanden - der Partner in diesem Haushalt genauer unter die Lupe genommen und die Variablenwerte der beiden Personen in den Kategorien Geschlecht, Geburtsjahr, Arbeitsstunden, Ausbildungsstunden und aufgewendete Zeiten für Einkäufe in Lebensmittelgeschäften und Damenbekleidungsgeschäften miteinander verglichen. Glücklicherweise waren, wenn mehrere Personen in einem Haushalt wohnten, bis auf wenige Ausnahmen die Daten aller übrigen Haushaltsmitglieder vollständig vorhanden, so daß meist nur für ein Haushaltsmitglied neue Variablenwerte generiert werden mußten. Zudem fehlten bei den meisten Personen nur einzelne Werte und nicht alle auf einmal, was das Generieren der neuen Werte wesentlich erleichterte und der Realität um ein großes Stück näherbrachte.

Falls bei beiden Personen das Geschlecht gefehlt hatte, wurden die Stunden für Einkäufe und Arbeit begutachtet und derjenige mit dem größeren Einkaufsvolumen und der geringeren Arbeitszeit als weiblich eingestuft. Waren Arbeitsstunden und Einkaufszeiten relativ ähnlich, so gab meist die Anzahl der Besuche in Damenbekleidungsgeschäften den Ausschlag. Diese eher konservativ klingende Annahme ist für die Umfragezeit Anfang der 70er Jahre durchaus realistisch und nachprüfbar und auch durch einige wissenschaftliche Studien über weibliche und männliche Reise-Aktivitäts-Muster gestützt (Hanson und Hanson, 1980, 1981).

Falls das Geschlecht eines anderen Haushaltsmitgliedes bekannt war, die Zeiten für Arbeit und Einkauf jedoch das gleiche Geschlecht ergeben hätten, konnte die Person nicht der Ehe- oder Lebenspartner sein. Deshalb wurde zusätzlich, falls vorhanden, das Geburtsjahr und die für Ausbildung aufgewendete Zeit betrachtet und entschieden, in welchem Verhältnis die Person zu den übrigen Haushaltsmitgliedern steht (Kind, Jugendlicher, älterer Mensch oder Elternteil, Geschwister des Haushalts- bzw. Familienoberhaupts, etc.). Fehlte hingegen auch das Alter bzw. Geburtsjahr, so spielte natürlich

die von vornherein bekannte Gruppe im Lebenszyklus (*life cycle group*), wie anfangs dieses Kapitels beschrieben, einen entscheidenden Faktor, beispielsweise konnte bei jungen Paaren mit zumindest einem Kind im Schulalter die dritte Person mit unbekanntem Daten neben den Eltern kein Erwachsener sein.

Fehlte das Alter der jeweilig beobachteten Person wurde diese Leerstelle mit Hilfe einer Arithmetik aus dem Alters des Ehepartners und des Durchschnittsalters aus den SAS-Tabellen für die jeweilige Arbeitsklasse berechnet. War die jeweils andere Person männlich, so wurde das Alter der Frau meist mit zwei Jahren weniger festgesetzt, was in der damaligen Zeit der durchschnittliche Altersunterschied zwischen den beiden Ehepartnern war. Waren die Geburtsjahre beider Ehepartner nicht bekannt oder gab es nur eine Person in diesem Haushalt, wurde nur der aus den SAS-Tabellen für die einzelnen Arbeitsklassen errechnete Alterswert eingesetzt (für Männer und Frauen unterschiedlich) und eventuell leicht korrigiert, wenn der Altersunterschied der beiden Partner zu gravierend war.

Durch diese korrigierenden und ergänzenden Annahmen konnte in dem meisten Fällen eine sehr gute Annäherung an die Realität erzielt werden.

### 2.5.6 Aufteilung des Datensatzes

Da für die weitere Auswertung detailliertere Informationen benötigt wurden, wurde der gesamte Datensatz mit Hilfe des Statistikprogrammes SAS (Programm 3 *split\_1*) in vier verschiedene Dateien unterteilt: Personen-, Haushalts-, Reise- und Wege- Datei. In jeder einzelnen Datei wurden Identifikationsnummern eingeführt, zusätzliche Variablen generiert und aus den vorliegenden die für die jeweilige Kategorie benötigten herausgefiltert bzw. zusammengefaßt.

In der Personendatei wurden so beispielsweise die maximale und die durchschnittliche Anzahl von Reisen pro Tag oder Wegen pro Reise Tag berechnet, Variablenwerte über die Anzahl der Arbeitswege, der Arbeitstage und der maximalen/durchschnittlichen Arbeitswege pro Tag, etc. angegeben.

Die Haushaltsdatei beinhaltet neben den gegebenen sozioökonomischen Daten des Haushalts auch Haushaltsgröße, Anzahl der Kinder im Vorschulalter / Schulalter und diverse Durchschnittswerte in bezug auf Arbeitszeit, Arbeitswege und Einkäufe des Haushalts.

In der Reisedatei und der Wegedatei wurden zusätzlich zu den vorhandenen Variablen vor allem Hauptaktivitäten, Hauptverkehrsmittel und andere typische Eigenschaften der Reise oder des Weges (Gesamt-reisezeit, -reiseausgaben, Anteil der geplanten Wege, Anteil der Wege in Begleitung eines

weiteren Haushaltmitglieds, etc.) ermittelt. Die Reihung dieser Variablen folgte einer vorher festgelegten bzw. bestimmten Hierarchie (Axhausen, 1999).

Eine ausführliche Auflistung aller Variablen sehen Sie in den Tabellen im Anhang I, Seite 167.

### 2.5.7 Erstellung der Aktivitätsdatei

Nach diversen Korrekturen und der Aufteilung in vier verschiedene Datensätze wurde mit Hilfe des SAS-Programms 4 (*create\_act\_I*) ein zur Analyse der Aktivitätsmuster benötigter Datensatz erstellt. Dieser 17291 Beobachtungen umfassende Aktivitätsdatensatz beinhaltet neben den Variablen der jeweiligen Aktivität auch beobachtungsübergreifende Informationen. So wurden neben wegspezifischen Variablen wie Reisegeschwindigkeit, Reisedistanz von und zur Aktivität, benütztes Verkehrsmittel, Benützung eines unabhängigen bzw. eines unabhängigen motorisierten Verkehrsmittels, Zeitspanne von der letzten bzw. zur nächsten Aktivität, etc. auch Informationen über die vergangene und die nachfolgende Aktivität (Aktivitätsdauer, -kategorie, etc.) berechnet und angegeben. Zusätzlich wurden auch noch einige informative soziodemografische Variablen erzeugt, die vor allem die Organisation des Haushalts bzw. die Familienstruktur der jeweiligen Person (z.B. Haushaltsvorstand, Anzahl der Kraftfahrzeuge im Haushalt, Durchschnittsalter des Haushalts, durchschnittliche Anzahl der Wege, und Reisen pro Haushalt, etc.) betreffen.

Ein Großteil der Korrekturen wurde in den Programmen 2 (*recode\_I*) und 3 (*split\_I*) gemacht und schon oben beschrieben (siehe Kapitel 2.5.1 bis 2.5.6). Die genauen Korrekturen und Generierung von zusätzlichen Variablen sind auch in allen SAS-Programmen ausreichend kommentiert.

Teilweise mußten Reisezeiten geringfügig verändert werden, da einige Beobachtungen gleiche Ankunfts- und Abfahrtszeiten beinhalteten. Dabei wurden unter Berücksichtigung der vorherigen und nachfolgenden Aktivitäten und Wegen (Dauer, Distanz, Kategorie, Verkehrsmittel und Vergleiche mit ähnlichen Wegen der Person) Ankunfts- oder Abfahrtszeiten dahingehend leicht verändert, daß positive und realistische Werte von 1 bis 5 Minuten für Reise- und Aktivitätszeiten entstanden.

Bei den Aktivitäten Nummer 38 (Spaziergang mit dem Hund), 66 (Spaziergang) und 67 (motorisierte Spazierfahrt) wurden zudem die Aktivitätsdauer und die Reisedauer des Weges auf den gleichen Wert gesetzt und mit der darauffolgenden Wegebeobachtung (meist der Weg nach Hause, der zuvor mit einer Reisedauer und Aktivitätsdauer von 0 bis 5 Minuten angegeben war) in Einklang gebracht. Für die gesamte Reisedauer der Person wurden diese beiden Wegebeobachtungen, die mit der Zusatzva-

riable *round\_tr=1* (*Roundtrip* - Rundweg/Rundfahrt) gekennzeichnet wurden, jedoch nur einmal berücksichtigt.

Da noch weitere Aktivitätsdauern und Intervalldauern negativ waren oder laut Eingabe 0 Minuten dauerten, wurden einige Korrekturen vorgenommen, so daß nur noch positive reelle Werte vorhanden waren. Mit Hilfe der vorhergegangenen Aktivitäten und Wege (betreffend Aktivitätskategorie, -dauer, Reisezeit, Verkehrsmittel, Wegdistanz, etc.) oder der nachfolgenden wurden die Aktivitätsdauern mit Werten von 1, 2 oder 5 Minuten als geringfügig positiv festgelegt. Die Intervalldauern waren dadurch bis auf zwei Ausnahmen, die separat korrigiert werden mußten, ebenfalls alle positiv und realistisch.

Für die Auswertungen in den einzelnen Kategorien wurde der gesamte Datensatz in mehrere kleinere gegliedert, die nur noch Aktivitäten der entsprechenden Kategorie enthalten. Dabei wurden unterschiedliche Grenzen der Aktivitätsdauer festgelegt und der Datensatz jeweils verkleinert. In der Kategorie Einkauf kurzfristig wurde beispielsweise eine maximale Aktivitätsdauer von 5 Stunden festgelegt, wodurch sich die Anzahl der Aktivitäten von 3231 auf 3221 Beobachtungen reduzierte. Auch in den restlichen Kategorien wurde eine Obergrenze von 1, 5, 10 oder 15 Stunden festgelegt und die Datensätze dementsprechend verkleinert. Dadurch konnten unrealistische Dauern von Aktivitäten eliminiert und die Datensätze einheitlicher strukturiert werden. Vor allem für die vier in den Hazardmodellen detaillierter betrachteten Kategorien Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden wurden weitere beschreibende Größen erzeugt (z.B. Tageszeit des Mittelpunktes der Aktivität, Möglichkeit zur Benützung eines Kraftfahrzeugs, Anzahl der Aktivitäten bzw. Wege pro Tag, monetäre Ausgaben), sodaß sich die Anzahl der Variablen pro Beobachtung auf 173 erhöhte.

### 2.5.8 Erstellung der Intervalldateien

Nachdem neben einer Auswertung der Aktivitätsdauern in einem zweiten Schritt bei den Hazardmodellen auch die Dauern der Intervalle zwischen gleichartigen Aktivitäten genauer betrachtet worden sind, wurden mit dem Programm 6 (*Hazard\_LIFETEST\_I*) weitere Datensätze generiert, die auf den einzelnen Intervallen basieren. Wie bei den Aktivitätsdauern wurde zuerst ein gesamter Datensatz (mit 17147 Beobachtungen) erstellt, der dann in die einzelnen Kategorien unterteilt wurde. Dabei wurden jene Intervalle, die länger als 14 Tage andauern sowohl im gesamten Datensatz, als auch in den Kategoriedatensätzen sofort eliminiert. Um keine zensurierten Daten vorliegen zu haben, wurde auch die jeweils erste und letzte Beobachtung bzw. das Intervall vom Umfragebeginn bis zur ersten durchgeführten Aktivität und das von der letzten Aktivität bis zum Ende der Umfrage aus dem Datensatz entfernt. Es wären dies jeweils nicht vollständige Intervalle, da man den genauen Zeitpunkt der letzten

Aktivität vor der ersten in der Umfragephase berichteten und den Zeitpunkt der ersten Aktivität nach der Umfrage nicht genau festlegen kann.

Um für die späteren Auswertungen eine zufriedenstellende Datengrundlage und vor allem eine ausreichende Anzahl an möglichen Einflußvariablen garantieren zu können, wurden weitere intervallbezogene und intervallübergreifende Variablen erzeugt. Die Dauer der Zeitspanne zwischen zwei beliebigen Aktivitäten (*intervall*) und zwischen zwei gleichartigen Aktivitäten (*inter*), die Art des Verkehrsmittels vor dem Intervall (*a\_p\_mm\_c*) die Dauer der vorherigen Aktivität in dieser Kategorie (*a\_p\_durst\_c*) und der vorherigen beliebigen Aktivität (*a\_p\_durst\_a*) bzw. die Frage, ob es ein unabhängiges Verkehrsmittel war (*mm\_c\_d*) sind darin enthalten.

## **3 BISHERIGE FORSCHUNGSARBEITEN MIT DEN DATEN DER UTHS**

### **3.1 EINLEITUNG**

Die Tatsache, daß die Uppsala Travel Household Survey (UTHS) eine vollständige und detaillierte Auskunft über das Reiseverhalten von 144 Personen über 35 aufeinanderfolgende Tage geben kann, machten sich natürlich in den vergangenen Jahren einige Wissenschaftler und Verkehrsplaner zunutze. In diversen Studien wurden verschiedenste Themenbereiche bearbeitet - wie zum Beispiel die Wiederholung und Variabilität in den Verhaltensmustern der Personen oder die Verteilung der Geschlechterrolle oder der Arbeitsweg als Mehrzweckweg - und Probleme diskutiert („Wie systematisch sind Wiederholungen im Verhaltensmuster?“ oder „Ist das Reismuster prinzipiell gewohnheitsmäßig?“).

In diesem Kapitel sollen einige Studien im Überblick behandelt werden.

### **3.2 DER ARBEITSWEG ALS MEHRZWECKWEG IM STÄDTISCHEN VERKEHR**

Die Studie von Susan Hanson aus dem Jahr 1980 zeigt die Wichtigkeit des Weges zur Arbeit als Mehrzweckweg einerseits und als Teil des gesamten Reismusters andererseits. Ein großer Teil aller vom Haushalt unternommenen Wege und Reisen wird in Verbindung mit dem Weg von und zur Arbeit zurückgelegt (44%). Trotzdem wurden vorher in diversen Studien für den Weg von und zur Arbeit nur die Weglänge, Reisezeiten, Verkehrsmittel- oder Routenwahl genauer untersucht bzw. der Arbeitsweg als reine Bewegung Heimstätte - Arbeitsstätte betrachtet. Doch gerade diese Komplexität, daß sehr viele Wege, die an sich nichts mit dem Weg von und zur Arbeit zu tun haben und als freiwillig bzw. frei wählbar bezeichnet werden, regelmäßig an den Arbeitsweg angehängt werden, spielt eine sehr große organisatorische Rolle im Reiseverhalten des Haushaltes.

Vorher scheiterte eine genauere Bearbeitung dieses Themas auch am Fehlen geeigneter Datensätze, denn in den meisten Umfragen waren Wege, die zu Fuß oder mit dem Fahrrad zurückgelegt worden waren, nicht beachtet worden. Zudem waren Ziele nur in Zonen, nicht aber mit punktgenauen Koordinaten geocodiert gewesen. All diese Voraussetzungen sind mit dem Datensatz der Uppsala Umfrage jedoch vorhanden.

Chapin (1974) bezeichnet den Weg zur Arbeit als nicht frei wählbar (sowohl zeitlich als auch räumlich) und unterscheidet ihn dadurch von anderen Wegen des innerstädtischen Verkehrs. Weniger künstlich und ein besserer Lösungsansatz wäre es jedoch, den Weg zur Arbeit als dominanten Zweck anzusehen, dem dann andere untergeordnete, freiwillige bzw. frei wählbare Aktivitäten zugeordnet und angehängt werden können (Westelius, 1973).

Doch wie aus den Auswertungen zu erkennen ist, sind viele Aktivitäten (wie Einkaufen, soziale Erledigungen, Freizeit, etc.), die vorher als „nicht pflichtmäßig“ eingestuft waren, so oft mit dem „pflichtmäßigen bzw. notwendigen“ Weg zur Arbeit verbunden, daß dieses „frei wählbar - nicht frei wählbar“ - Konzept überdacht werden muss.

Hanson fand für die Uppsala-Daten heraus, daß 53 % aller Arbeitswege Mehrzweckwege waren (bei den Reisen, die nicht mit der Arbeit verbunden sind, waren es nur 38 %) und 57 % aller Haushalte mehr Halte auf Arbeitswegen als auf anderen Wegen machten. In Bezug auf Flächennutzung und Aktivitäten, die am häufigsten auf dem Arbeitsweg erledigt wurden, zeigte sich, daß nicht nur einige wenige, sondern eine sehr große Anzahl verschiedenster Aktivitäten öfters frequentiert wurde. Auch die Wahrscheinlichkeit, daß der Halt nicht zu Hause geplant wurde, war bei vielen Aktivitäten am Arbeitsweg wesentlich größer als auf anderen Wegen. Betrachtet man alle Wege, so waren 95 % aller Halte geplant, Schaufensterbummel beispielsweise waren auf Arbeitswegen zu 52%, auf anderen Wegen nur zu 25% nicht zu Hause geplant worden. Das heißt, daß zum Beispiel am Heimweg noch schnell eine Erledigung gemacht wird, weil man gerade noch Zeit hat oder zufällig auf dem Weg an diesem Geschäft vorbeikommt. Dies hat natürlich Einfluß auf diverse Geschäfte, die von solchen „zufälligen“ Kunden bzw. Personen, die kurzfristig und ungeplant ins Geschäftslokal kommen, abhängig sind. Räumlich gesehen sind dies vor allem Geschäfte im CBD (Central Business District), also im gesellschaftlichen und wirtschaftlichen Zentrum von Uppsala, da ca. 1/3 aller Reiseaktivitäten im zentralen Bezirk stattfinden (mit den eigentlichen Wegen zur Arbeit mehr als die Hälfte).

Diejenigen Flächennutzungen und Aktivitäten, die mehr als „erwartet“ mit dem Arbeitsweg in Verbindung gebracht wurden (also mit mehr als 44 % der Halte), waren zum Großteil kommerzielle Einrichtungen bzw. Essen gehen oder kurze Erledigungen.

Diese Zahlen sagen jedoch überhaupt nichts über die Zeit, zu der die Aktivitäten an den Arbeitsweg angehängt wurden, ob vor der Arbeit, in der Mittagspause oder auf dem Weg nach Hause. Zur Arbeit wurden vorrangig Behörden, Versicherungen und andere öffentliche Büros aufgesucht, zu Mittag Restaurants oder Fotogeschäfte und auf dem Heimweg am ehesten Kiosk, Bank, Autowerkstätten und Spirituosen-geschäfte. Der Großteil der Wege wurde jedoch nicht in der Früh auf dem Weg zur Arbeit erledigt, da die zeitlichen Rahmenbedingungen ein pünktliches Erscheinen am Arbeitsplatz vorschreiben (70 % der Wege zu Arbeit wurden von zu Hause ohne Zwischenhalt getätigt), sondern auf dem Heimweg, wo die Zeit naturgemäß nicht so drängt.

Solche Resultate haben natürlich auch Auswirkungen auf Verkehrsmodelle wie das von Selinger (1977) über fließende Arbeitszeiten, da nicht nur differierende Zeiten für den Weg von und zur Arbeit auftreten würden (und so Staus vermieden werden sollten), sondern sich zum Beispiel Geschäfte mit ihren Öffnungszeiten anpassen müßten und sich dann auch das Einkaufsverhalten vieler Arbeitnehmer wieder ändern würde, da sie eventuell vor ihren Erledigungen zuerst nach Hause fahren und erst dann von zu Hause aus einkaufen gehen.

Weiters wurde mit Hilfe von Ausgangspunkt / Endpunkt - Matrizen festgestellt, daß viele Flächennutzungen bzw. Aktivitäten entweder mit dem Arbeitsplatz oder mit der Wohnung in Verbindung zu bringen sind, nur sehr wenige mit beiden oder anderen Orten.

Bei der Frage nach den auf den mit der Arbeit verbundenen Wegen benutzten Verkehrsmitteln, stellte die Autorin eine geringe Anzahl an Reisen mit inflexiblen Mitteln wie Bus (trotz eines sehr guten Bus-systems in Uppsala) und als Autobefahrer fest, unabhängige Verkehrsmittel (freie Wahlmöglichkeit der Haltepunkte und Reiserouten) wie der eigene Pkw, das Fahrrad oder der Weg zu Fuß wurden jedoch relativ oft gewählt.

Nachdem Verkehrsplaner den Arbeitsweg hauptsächlich aufgrund der Verkehrserzeugung und der Spitzenstunden untersucht haben, schlägt diese Arbeit vor, den Arbeitsweg auf die Weise genauer zu betrachten, daß er für viele nicht arbeitsverbundene Reiseverknüpfungen verantwortlich ist. Genau aus diesem Grund stellt er einen nicht zu unterschätzenden Faktor in den Reismustern jedes Einzelnen dar.

### 3.3 AKTIVITÄTSMUSTER DER GESCHLECHTER UND AUSWIRKUNGEN DER ERWERBSTÄTIGKEIT VON FRAUEN DARAUFG

In früheren Studien wurden zwar einerseits städtische Verhaltensmuster verschiedener Personen untersucht und andererseits Aktivitätsmuster von außer Haus erwerbstätigen Männern und Frauen in Bezug auf eine eventuelle Änderung der Rollenverteilung im Haushalt (wenn die Frau einem Beruf außer Haus nachgeht) verglichen, „jedoch geschah dies immer mit sogenannten Zeit-Budget-Analysen, die größtenteils der Frage nachgehen, wie einzelne Personen ihre 24 Stunden verbringen“ (Szalai, 1972, Chapin 1974, Thrift, 1977, Anderson, 1971), räumliche Komponenten fanden dabei keine Beachtung oder wurden nur rudimentär als „innerhalb“ oder „außerhalb des Hauses“ klassifiziert.

In der ersten Studie (Susan Hanson und Perry Hanson, 1980) sollten mit Einbezug aller räumlichen Komponenten (genaue geocodierte Ortsangaben, Distanzen, Verkehrsmittel, etc.) geeignete und verständliche Außer-Haus-Aktivitätsmuster der erwerbstätigen Männer und Frauen gefunden werden, um dabei auch prinzipielle Unterschiede im Verkehrsverhalten aufzeigen zu können (arbeitende Frauen haben einen wesentlich engeren Zeitrahmen als ihre männlichen Kollegen und auch als nicht oder teilweise erwerbstätige Frauen, da die häusliche Verantwortung nicht signifikant reduziert ist).

Die Resultate dieser Studie zeigen deutliche Unterschiede im täglichen Verkehrsverhalten, die weitgehend der traditionell geprägten Meinung über die Rollenverteilung zwischen Mann und Frau entsprechen. Arbeitende Frauen machten weniger Wege, legten kürzere Distanzen zurück, erledigten mehr Einkäufe und wendeten weniger Zeit für Freizeitaktivitäten auf als ihre männlichen Gegenüber, aber auch als nicht arbeitende Frauen, während die Männer häufiger Freizeitstätten und Arbeitsplatz frequentierten. Auch in der Verkehrsmittelverteilung hinkte die arbeitende Frau stark ihrem männlichen Counterpart hinterher, da sie wesentlich weniger Zugangsmöglichkeiten zu unabhängigen Verkehrsmitteln hatte, Ausnahme dabei war nur die sehr häufige Benutzung des Fahrrades. Dafür verantwortlich sind vor allem die in Uppsala von jeher große Akzeptanz des Fahrrades als Verkehrsmittel zur Arbeit (Hanson und Hanson, 1976, 1977) und die Tatsache, daß viele nicht Erwerbstätige Mütter von kleinen Kindern sind, die am Fahrrad nur schwierig transportiert werden können.

Die zweite Studie (Susan Hanson und Perry Hanson, 1981) beschäftigt sich dann eingehend mit der Frage, wie die Erwerbstätigkeit der Frau ihr eigenes Verhaltensmuster beziehungsweise auch das ihres Ehemannes beeinflusst. Es werden zuerst die soziodemografischen Daten von 6 verschiedenen Gruppen (Vollzeit- Teilzeit- und nicht außer Haus arbeitende Frauen und jeweils deren Ehemänner) miteinander verglichen, wobei auffällt, daß bei Frauen mit kleinen Kindern 60 % zu Hause und nur 10 %

vollzeitbeschäftigt sind. Bei jungen Frauen ohne Kinder verhält sich dies genau umgekehrt (60% arbeiten Vollzeit und nur 16% gar nicht außer Haus) und bei Müttern mit Kindern im Schulalter ist mit 55 % die Teilzeitarbeit am gefragtesten.

Statistisch signifikante Unterschiede gibt es sowohl bei Männern als auch bei Frauen aber nur in der Haushaltsgröße (2,6 Personen pro Haushalt bei vollzeitbeschäftigten Frauen, 2,9 bei Teilzeit und 3,1 bei nicht außer Haus beschäftigten Frauen), alle anderen soziodemografischen Eigenschaften (wie Ausbildungsgrad, Haushaltseinkommen, Autobesitz im Haushalt, Führerscheinbesitz und auch die räumlichen Voraussetzungen (Wegstrecken von Zuhause zur Arbeit oder ins Zentrum)) weisen keine gravierenden Unterschiede auf.

Mit Hilfe von 4 Meßwerten als Indizes der Variabilität (Anzahl der verschiedenen besuchten Orte, Anzahl der verschiedenen Flächennutzungen, Anzahl der verschiedenen durchgeführten Aktivitäten und Anteil der im Zentrum gemachten Wege) werden die Veränderungen in den Verkehrsverhaltensmustern einerseits der Frauen selber, andererseits ihrer Ehemänner, wenn die Frau außer Haus beschäftigt ist, beleuchtet. Signifikante Unterschiede treten bei den Frauengruppen natürlich bei den arbeitsbezogenen Wegen auf, alle anderen Vergleiche zeigen nur sehr kleine Unterschiede. So neigen Vollzeit beschäftigte Frauen beispielsweise mehr zu Einzel-Zweck-Arbeitswegen (teilzeitbeschäftigte Frauen finden eher die Zeit, am Arbeitsweg noch eine Erledigung zu machen), machen etwas mehr Wege am Wochenende, legen für manche Aktivitäten (Einkaufen, Erholung) größere Distanzen pro Weg zurück und wenden nur geringfügig weniger Zeit für gewisse Aktivitäten abseits der Arbeit auf. Insgesamt machen vollzeitbeschäftigte Frauen aber eher zeitliche denn räumliche Einschränkungen ihrer Außer-Haus-Aktivitäten. Bei Teilzeitbeschäftigung der Frau besteht nicht unmittelbar die Notwendigkeit, das Verkehrsmuster gravierend zu verändern (bis auf die zusätzlichen Arbeitswege), was durch die Auswertungen bestätigt wird.

Die Verhaltensmuster der Ehemänner-Gruppen zeigen noch weniger Unterschiede, statistisch signifikant sind sie überhaupt nirgends. Ein interessantes Muster tritt jedoch insofern auf, daß je länger die Arbeitszeit der Frau ist, desto häufiger sind die Wege der Ehemänner für soziale Zwecke, Einkaufen und Freizeitaktivitäten, weniger oft hingegen für persönliche Erledigungen (Post, Bank, etc.).

Insgesamt bleiben also, obwohl eine Frau vollzeitbeschäftigt ist, weiter große Differenzen zwischen ihrem Verkehrsmuster und dem ihres Mannes bestehen. Sie legt mehr Wege für persönliche Erledigungen und Einkaufen, jedoch weniger für Arbeit, Sozial- und Freizeitaktivitäten zurück und ist zudem in der Verkehrsmittelwahl stark benachteiligt (eine nur geringe Anzahl der Wege wird mit dem Auto zurückgelegt, der Mann hingegen (in allen drei Gruppen) legt mehr Wege mit dem Auto als mit irgendeinem anderen Verkehrsmittel zurück).

Die Studie kommt zum Schluß, daß die Arbeit der Frau außer Haus keine große Rollenverschiebung im Haushalt mit sich bringt, sie behält die Hauptverantwortung für den Haushalt und die Familie und es wird hauptsächlich ihr Verkehrs-Aktivitäts-Muster an die geänderte Arbeitssituation angepaßt, nicht das ihres Mannes.

Im Anschluß an die Analyse werden noch einige Auswirkungen dieser empirischen Studien für politische Entscheidungsträger und Verkehrsexperten diskutiert, wobei es zwei Alternativen zu geben scheint: Entweder die traditionelle Rollenverteilung zu akzeptieren und Maßnahmen zur Erleichterung der doppelten Belastung (Arbeit und Haushalt) für vollzeitbeschäftigte Frauen einzuleiten oder die Unterschiede zwischen Männer- und Frauenrolle durch verschiedene Maßnahmen auszugleichen. Zur ersten Alternative gehören Dinge wie flexiblere Arbeitszeiten, größere Anreize (auch finanzieller Natur) für Teilzeitarbeit, sodaß Frauen nicht mehr Vollzeit arbeiten müssen, ein Ausbau der öffentlichen Verkehrsmittel (da dies das von Frauen meistgenutzte Verkehrsmittel ist), vermehrte Ansiedlung von Geschäften des täglichen Bedarfs in Wohngebieten oder ein Anlegen von sicheren Radwegnetzen (auch für Kinder alleine befahrbar), etc. Diese Maßnahmen würden zwar wahrscheinlich keine großen Veränderungen der Rollenverteilung bringen, dennoch eine gewisse Erleichterung in den täglichen Verkehrsmustern der erwerbstätigen Frau bewirken. Zur zweiten Alternative gehören Maßnahmen, wie sie unter anderem auch durch schwedische Regierungsprogramme eingeleitet wurden. So wurde vom schwedischen Parlament 1975 und 1976 das Schulalter von 7 auf 6 Jahre herabgesetzt bzw. die Einführung von Vorschulklassen, die Einrichtung von 100.000 Ganztages- und 50.000 Nachmittags-Kinderbetreuungsplätzen beschlossen, um Müttern den Einstieg in die Erwerbstätigkeit zu erleichtern. Eine weitere Maßnahme (1974) war die Entscheidung, welcher Ehepartner die 9 Monate Karenzzeit (bei 90%iger Bezahlung) in Anspruch nimmt, den Eltern selbst zu überlassen. Zusätzlich kann einer der beiden Elternteile pro Jahr 18 Tage zur Kinderbetreuung zu Hause bleiben, was ebenfalls die Rolle des Mannes im Haushalt verfestigen sollte. Schlußendlich wurde im Jahr 1979 den schwedischen Arbeitnehmern bewilligt, als Eltern von Kindern unter 8 Jahren, ihre tägliche Arbeitsleistung auf 6 Stunden zu reduzieren. Die schwedische Gewerkschaft drängte daraufhin auf eine viertägige Arbeitswoche als Alternativlösung.

Es sind noch weitere Maßnahmen gefordert bzw. teilweise beschlossen worden: gleiche Bezahlung für gleiche Arbeit, Steuererleichterungen für Betriebe, die vermehrt Frauen anstellen, unterschiedliche Besteuerung für Alleinverdiener und für Mehrpersonenhaushalte oder Familien, Um- und Einschulung von Frauen in Männerberufe und umgekehrt. Diese Maßnahmen haben jedoch alle keinen direkten Einfluß aufs Verkehrsverhalten. „Wie sich diese Maßnahmen auswirken, die die schwedische Regierung ergriffen hat, um selbst aktiv zu gewissen sozialen Änderungen beizutragen und die Ungleich-

heiten in unserer Gesellschaft zwischen Männern und Frauen etwas zu mildern, wird man aber erst in einigen Jahren sehen können.“ (Hanson und Hanson, 1981, Seite 181).

### 3.4 ZUGANGSMÖGLICHKEITEN UND INNERSTÄDTISCHER VERKEHR

Diese Studie von Hanson und Schwab (1985) versucht die grundlegende Annahme, daß das innerstädtische Verkehrsverhalten des Einzelnen sehr stark von seinen Zugangsmöglichkeiten abhängt, genauer zu erforschen bzw. zu widerlegen. Diverse Modelle bzw. Untersuchungen - beispielsweise die gleichmäßige und gerechte Verteilung von privaten und öffentlichen Einrichtungen über das Stadtgebiet (Knox, 1978) oder die Abschätzung der Inanspruchnahme von Alternativrouten im Straßennetz (Koenig, 1980) – hatten diese bisher kaum untersuchte und weitverbreitete Annahme zur Grundlage genommen. Im interurbanen Bereich hat die Annahme, daß räumliche Strukturvariablen (beispielsweise die Dichte von Aktivitätsorten) Reismuster beeinflussen können durchaus gewisse Richtigkeit (Daniels und Warnes, 1980), sie kann aber nicht ohne weiteres auf den innerstädtischen (intraurbanen) Bereich übertragen werden.

Hanson und Schwab versuchten nun zu allererst geeignete Meßmethoden zu finden, um die Zugangsmöglichkeiten bestimmter Bevölkerungsgruppen adäquat genau bestimmen zu können. Dabei spielten einerseits die Dichte von potentiellen Aktivitätsorten im Umkreis des Arbeitsplatzes und des Zuhauses der jeweiligen Person, andererseits auch die räumliche Trennung der Arbeit und Aktivitäten eine Rolle, die in komplizierten Modellen miteinander verbunden wurden.

In einem weiteren Schritt untersuchten sie die Beziehungen zwischen Zugänglichkeitsindizes und den verschiedenen Meßwerten des Verkehrsverhaltens (Wegehäufigkeit, Verkehrsmittelwahl, Wegdistanzen, Komplexität der Wege, Reisen in Verbindung mit dem Arbeitsweg und Größe des Aktivitätsraumes). Dabei kamen zum Schluß, daß bei weitem nicht so große Wechselwirkungen zwischen Zugänglichkeit und Verkehrsverhalten bestehen, wie in der Literatur immer angenommen wurde. Trotz dieser schwachen Zusammenhänge gab es einige faszinierende Muster: hohe (Zugänglichkeits-)Dichten sind sehr oft mit nichtmotorisiertem Verkehr, weniger Autoverkehr, reduzierten Weglängen für gewisse frei wählbare Zwecke und räumlich strikter eingeschränkten Aktivitätsrahmen verbunden. Daraus könnte die Schlußfolgerung gezogen werden, daß höhere Dichten in der städtischen Bauweise Energieeinsparungen mit sich bringen.

Durch die Verwendung von Zugänglichkeitsmeßwerten, die einerseits auf der Arbeitsstätte basierten, andererseits auf der Wohnung, konnten die verschiedenen Auswirkungen auf die Verhaltensmuster

besser abgeschätzt werden. So beeinflussen wohnungsbezogene Meßwerte die allgemeinen Reismuster mehr (z.B. bei arbeitenden Männern die für Einkaufen und persönliche Erledigungen zurückgelegten Distanzen), arbeitsplatzbezogene Meßwerte eher die Verkehrsmittelwahl und Weglängen, keineswegs jedoch die Wegehäufigkeit.

Für zukünftige Studien schlagen sie vor, die Meßwerte für Zugangsmöglichkeiten noch genauer zu unterscheiden (beispielsweise neben wohnungs- und arbeitsplatzbasierenden noch weitere Meßwerte einzuführen). Zudem sollten Zugangsmöglichkeiten für Modelle der Wegerzeugung und der Wegehäufigkeit nicht als entscheidende Faktoren eingebracht werden, für Untersuchungen der Verkehrsmittelwahl und Wegverteilung jedoch müßten sie auf jeden Fall betrachtet werden.

### **3.5 WIEDERHOLUNG UND TAGTÄGLICHE VARIABILITÄT IN KOMPLEXEN REISEMUSTERN**

Die folgenden drei Studien von Susan Hanson und James O. Huff bzw. Perry Hanson beschäftigen sich jeweils mit der Variabilität und den beobachteten Wiederholungen in den Verhaltensmustern von Einzelpersonen.

Die erste Arbeit von Susan Hanson und James O. Huff aus dem Jahr 1982 stellt geeignete Meßmethoden für individuelle Verhaltensmuster vor, um darin tagtägliche Variabilität entdecken und abschätzen zu können. Aufgrund einer oder mehrerer charakteristischer Eigenschaften wurden die Wege einer Person in sogenannte Äquivalenzklassen eingeteilt und diese miteinander verglichen. Dadurch lassen sich Tage mit ähnlichen Reismustern identifizieren. Ebenso können durch Aufsummierung der Zeitabstände zwischen Wegen der gleichen Äquivalenzklasse Einblicke in eventuelle zyklische Verhaltensmuster gewonnen werden oder mit Hilfe der häufigsten Wegkombinationen atypische Verhaltensmuster entdeckt werden. Die präsentierten Meßmethoden wurden an den Daten aus Uppsala getestet und man bestätigte frühere Ergebnisse von Zeit-Budget-Studien, daß Eintagesbefragungen als Grundlage für Variabilitäts- und Wiederholungsuntersuchungen absolut nicht ausreichend sind bzw. zu falschen Ergebnissen führen können. Zudem bestätigten die Autoren, daß die Qualität von longitudinalen Datensätzen über mehrere Wochen Beobachtungszeit durchaus konstant bleiben kann. Wie schon im Kapitel 2.1 erwähnt, weisen nur etwa 5 Prozent der Umfrage einen Abwärtstrend in der Anzahl der Wege oder Aktivitäten pro Tag über die gesamte Umfragedauer gesehen auf.

Bei der Analyse fanden die Autoren ansatzweise noch einige interessante Dinge über Wiederholung und Variabilität heraus, die jedoch erst in der nächsten Studie genauer untersucht und dargestellt worden sind.

Die zweite Studie von Huff und Hanson (1986) stellte fest, daß das Reise-Aktivitäts-Muster des Einzelnen natürlich sowohl Wiederholungen als auch Abwechslung oder Variabilität beinhaltet. Die beobachtete Wiederholung ist verständlicherweise wesentlich um so geringer, je komplexer die Muster sind, die miteinander verglichen werden. Werden nur gleiche Aktivitäten beobachtet, die in ein oder zwei Eigenschaften übereinstimmen müssen, so werden sich diese im Gegensatz zu größeren Zusammenhängen auch öfter wiederholen. Eine wichtige Erkenntnis aus dieser Analyse ist, daß jede Person mehr als ein charakteristisches oder atypisches tägliches Verhaltensmuster an den Tag legt. Diese beobachteten Muster unterscheiden sich meist grundlegend voneinander. Für manche Personen stellen auch Tage ohne einen einzigen Weg ein signifikantes Reismuster dar. Gewisse grundlegende Segmente des Reiseverhaltens wiederholen sich, jedoch nicht so sehr, daß man mit einem einzigen Tag eine adäquate Beschreibung der typischen Routine- Reise einer Person geben kann. Dafür sind die repräsentativen Tagesmuster einfach zu verschieden.

In der dritten Studie von Susan und Perry Hanson aus dem Jahre 1988 wurden nun diese vermuteten systematischen Variabilitäten im sich wiederholenden Verkehr genauer untersucht. Zuerst wurden die 5 repräsentativsten Tage jedes Einzelnen eruiert (in der Studie als R-days bezeichnet) und die restlichen 30 Tage diesen 5 zugeordnet. Dabei waren im Schnitt 12,7 Tage dem repräsentativsten Tag am ähnlichsten, 6,8 ähnelten dem zweiten R-day am meisten, usw.. Dabei waren die Reismuster der verschiedenen repräsentativen Tage teilweise stark unterschiedlich. Das heißt, daß jeder Mensch nicht ein einziges typisches Verhaltensmuster oder tägliches Routinemuster hat (auch nicht ein werktagspezifisches), sondern mehrere verschiedene, die sich eventuell in gewissen Zyklen wiederholen. Nachdem schon vorher der einzelne Tag als Zyklus ausgefallen war, wurde als nächstmögliche Wiederholungsperiode gleichartiger Verhaltensmuster die wöchentliche Periode betrachtet. Sie spiegelt zwar bei weitem nicht alle spezifischen, unterschiedlichen Verhaltensweisen wider, jedoch hat sie große Vorteile gegenüber einer Ein-Tages-Beobachtung, denn es besteht eine sehr große Wahrscheinlichkeit, daß in dieser zufällig gewählten Woche einige der typischen täglichen Verhaltensmuster des Einzelnen auftauchen. In der Tat tauchten bei über 90% der Personen zwischen 3 und 5 verschiedene repräsentative Muster bzw. R-days in dieser zufällig gewählten Woche auf. Daraus ersieht man einerseits, daß eine einwöchige Befragung doch gewisse Informationen über die wiederkehrenden täglichen individuellen Muster beinhaltet und andererseits, dass Tage, die mit dem repräsentativsten Tag am stärksten korrelieren, die größte Wahrscheinlichkeit haben, in jeder beliebigen Woche vorzukommen.

Ein Versuch, gewisse Zyklen zu erkennen, wurde mit Hilfe der sogenannten Kernwege bzw. deren Äquivalenzklassen unternommen. Kernwege sind oft wiederkehrende Wege (zumindest 4 mal in der gesamten Umfragezeit, sodaß eine wöchentliche Wiederholung möglich wäre), die sich in 4 Eigenschaften gleichen müssen (Aktivität, Ort, Zeit und benutztes Verkehrsmittel). Ein Beispiel ist hier der Weg zur Schule, der um 7.30 beginnt und mit dem Bus unternommen wird, es müssen aber immer die selben Eigenschaften wiederkehren. Ändert sich eines der Attribute (z.B. statt des Schulgebäudes wird die 1 km entfernte Sporthalle als erstes besucht oder das Verkehrsmittel ist ausnahmsweise das Auto des Vaters statt des Buses), so wurde dieser Weg von Hanson und Hanson nicht als Kernweg erkannt. Wie erwartet, tauchte bei 69% aller Befragten die Arbeit als Kernweg auf. Einen noch größeren Anteil wies jedoch die Aktivität Einkaufen auf, denn bei 83 % der Befragten war dies ein Kernweg. Auffallend war, daß nur bei 47% der außer Haus beschäftigten Personen Arbeit die häufigste Kernaktivität war, was jedoch in anderen Studien vorher schon erwartet worden war.

Gesamt gesehen ist die tagtägliche Variabilität sehr hoch, abhängig von unterschiedlichen Umwelteinflüssen und Rahmenbedingungen. Teile dieser Variabilität sind jedoch erklärbar und vorhersehbar. Der Schlüssel zu diesem Argument liegt in der Tatsache, daß Kernwege sowohl die Struktur der individuellen Reismuster bestimmen, als auch systematisch in den Beobachtungen immer wiederkehren. Nach Untersuchungen in Bezug auf 3 verschiedene Ursachen (Unterschied zwischen Wochentag – Wochenende, Regelmäßigkeiten in der Anzahl der Wege pro Woche für eine gegebene Kernwegeklasse und Systematik in den Wartezeiten zwischen Wegen einer bestimmten Klasse) stellten die Autoren fest, daß eine haufenförmige Verteilung der Kernwege wesentlich öfter vorkam als eine gleichförmige Verteilung. Trotzdem relativ wenige Kernwege in einem gleichförmigen, regelmäßigen Muster auftreten, sind sie dennoch sehr wichtig für das gesamte Reise-Aktivitäts-Muster des Einzelnen. Sie stellen nicht nur den größten Teil aller Außer-Haus-Wege, sondern verankern auch den Großteil der restlichen Wege.

Die Ergebnisse, daß in den individuellen Verhaltensmustern nur sehr geringfügig Wiederholungen gemacht werden und die Abwechslung (Variabilität) extrem hoch ist, hängen natürlich neben dem verwendeten Datensatz auch von den Meßmethoden ab. Fordert man rigoros (wie hier gemacht), daß die Verhaltensmuster in komplexeren Zügen übereinstimmen müssen und beispielsweise nicht nur in einer einzigen Eigenschaft oder nur gleiche Aktivitäten miteinander verglichen werden, dann zeigen sich natürlich auch wesentlich weniger Wiederholungen. Nachdem die Varianz in der Anzahl der Kernwege pro Woche relativ hoch ist, kann eine einwöchige Befragung keine Grundlage für eine annehmbar genau Abschätzung der Häufigkeit des Auftretens von Kernwegen eines bestimmten Typs darstellen. Obwohl eine einwöchige Befragung lang genug ist, um verschiedene typische tägliche Verhaltensmuster ziemlich ausreichend zu beinhalten, so ist sie bei weitem zu kurz, die Abstände bzw. Häufigkeiten, in der eine Person Kernwege verschiedener Art macht, ausreichend zu dokumentieren.

Genau diese Häufigkeiten (bzw. die Messung derselben) sind es aber, die notwendig sind, um glaubwürdige und aussagekräftige Modelle für Reiseverhalten liefern zu können oder sinnvolle Maßnahmen ergreifen zu können. Denn vom Standpunkt eines Entscheidungsträgers ist es durchaus wichtig, ob beispielsweise 10 % der Bevölkerung jeden Tag den Bus benutzen oder jeder in der Bevölkerung 10 % der Zeit den Bus benützt. Gerade solche Häufigkeiten - der Unterschied zwischen inter- und intra-individueller Variabilität - können mit einwöchigen Umfragen nicht beantwortet werden (mit Ein-Tages-Umfragen, wie vorher gezeigt, sowieso nicht).

## **4 REISE- UND VERKEHRSVERHALTENSMUSTER, MODELLE UND DEREN AUSSEHEN**

### **4.1 EINLEITUNG UND GESCHICHTLICHER HINTERGRUND**

Lange bevor die heutige Garde der Verkehrsplaner und Soziologen sich mit der Frage der täglichen Aktivitäten der Menschen beschäftigten begann, führte Martha Ballard (1735 - 1812) 27 Jahre lang detailliert Buch über jede Sequenz ihres Tagesablaufes in Augusta, US-Bundesstaat Maine. Der Historiker L. T. Ulrich bereitete 1990 die Tagebücher von Ballard über ihr Leben als Hebamme, Mutter von neun Kindern und Großmutter dutzender Enkelkinder, als Frau eines Bürgermeisters, Gärtners und Händlers und als Nachbarin in einer Studie über das Leben im späten 18. und frühen 19. Jahrhundert in Neuengland auf und dokumentierte damit einen Vorläufer der modernen Verkehrsverhaltensmuster. Obwohl Ballard keinen schlaun Soziologen oder Sozialforscher hinter sich hatte, berichtete sie vor ca. 200 Jahren genau über jene Aspekte des täglichen Lebens, die normalerweise in Studien des späten 20. Jahrhunderts über städtische Aktivitätsmuster enthalten sind.

### **4.2 WAS SIND AKTIVITÄTSMUSTER GENAU?**

Einer der Pioniere der heutigen Studien über Reise- und Aktivitätsmuster war Verkehrsplaner Chapin. Er bezeichnete sie 1974 als den „Versuch, ein Fenster zu den Menschen zu öffnen, um Einblicke in deren täglichen Lebensablauf zu gewinnen und vor allem zu hinterfragen, wie sie Gebrauch von der Stadt und ihren Einrichtungen machen“. Dadurch erhoffen sich die Forscher, mehr über die Lebensqualität der verschiedenen Gruppen von Stadtbewohnern herauszufinden und so Verkehrsplanern bei der Gestaltung einer lebenswerteren Stadt helfen zu können.

Ursprünglich entstanden Studien über Aktivitätsmuster in den 60er und 70er Jahren vor allem wegen der Unzufriedenheit mit den vorher entwickelten „*Transportation Studies*“, die nicht genügend Einblicke in das tägliche Leben der Menschen oder ihre tagtäglichen Entscheidungen über Wegewahl oder Verkehrsmittelwahl, etc. ermöglichten. Die neuen Studien über Aktivitätsmuster konzentrieren sich vermehrt auf die durchgeführten Aktivitäten und weniger auf die Reisen und streichen dabei die Rolle der Rahmenbedingungen auf das Verhalten der Personen und die komplexen Verknüpfungen zwischen den verschiedenen Aktivitäten heraus. Meist bestehen sie aus einer sehr detaillierten und sorgfältig strukturierten Umfrage für eine große Anzahl an befragten Personen. Sie versuchen mehrere Fragen wie „Was machen die Personen?“, „Wo machen sie es?“ und „Wann machen sie es?“ zu beantworten, oft auch die Frage „Wieso machen sie das?“. Im Gegensatz zu den Tagebüchern von Martha Ballard, die sehr detailliert über alle Aktivitäten berichtete (auch häusliche Arbeiten wie kochen, Brot backen und nähen etc.), beschränken sich die Umfragen und Studien unserer Zeit auf die Aktivitäten außer Haus. Dabei gibt es drei grundlegende Prinzipien: (1) jedem, egal ob alt oder jung, reich oder arm, Mann oder Frau, stehen 24 Stunden pro Tag zur Verfügung; (2) jeder kann nur an einem einzigen Ort gleichzeitig sein; und (3) niemand kann sich ununterbrochen bzw. ohne Zeitverschiebung von einem Platz zum anderen fortbewegen. Menschen sind an Zeit und Raum gebunden. (Hanson und Hanson, 1993).

Einige Wissenschaftler (Hägerstrand, 1970; Carlstein, Parkes und Thrift, 1978; Lenntorp, 1976) versuchten mit Zeit-Raum-Prismen die Auswirkungen von Rahmenbedingungen auf das tägliche Leben der Menschen grafisch darzustellen, wobei die Höhe immer 24 Stunden entsprach, die Breite von der räumlichen Entfernung der Aktivitäten abhängig war und mit den benützten Verkehrsmitteln bzw. den ausgegebenen Geldbeträgen variierte (vergleiche z.B. Hägerstrand, 1970).

### **4.3 WARUM BRAUCHEN WIR AKTIVITÄTSMUSTER?**

Auf diese Frage gibt es nicht nur eine einzige Antwort.

Ein Hauptgrund für die Studien über Aktivitätsmuster ist sicherlich die Begierde der Verkehrsplaner, mehr über den Tagesablauf der Menschen zu erfahren und Hintergründe über die Art, die räumliche Ausdehnung und die Verschiedenheit der menschlichen Aktivitätsmuster zu erforschen. Vor allem interessiert die Forscher dabei die Frage, wie unterschiedlich die Aktivitätsmuster verschiedener Bevölkerungsgruppen aussehen, was die Gründe dafür und die resultierenden Auswirkungen sind. Diese Ergebnisse könnten den Verkehrsplanern helfen, Auswirkungen von beispielsweise Änderungen im öffentlichen Verkehrsangebot auf das tägliche Leben verschiedener Gesellschaftsgruppen voraussagen

zu können (Hanson, 1977). Ebenso ergeben sich daraus aber zusätzliche Fragen, ob gewisse von Verkehrsplanern durchgeführte Projekte an den Bedürfnissen der Menschen vorbeigehen bzw. welche Projekte für bestimmte Bevölkerungsgruppen notwendig sind.

Eine zweite Motivation für Studien über Aktivitätsmuster ist der Wunsch, mehr Einsicht in die Zusammenhänge zwischen menschlichem Verhalten und der städtischen Umgebung zu erlangen, speziell die Frage, wie sich Verhalten durch geänderte Bewegungsvoraussetzungen bzw. ein anderes Umfeld in der Stadt entwickelt. Deswegen spezialisieren sich Studien dieser Art vor allem auf die Wechselwirkungen zwischen Verhalten und lokalen Begebenheiten bzw. örtlichem Umfeld. Zusätzlich gibt es auch Studien, wie zum Beispiel die „HATS“- Studie (*Household Activity Travel Simulator*), die von Verkehrsforschern der Universität Oxford Mitte der 70er Jahre durchgeführt wurde, die explizit anhand einer kleineren Gruppe von Personen deren geändertes Verkehrsverhalten aufgrund von Veränderungen des Umfeldes untersuchen. Änderung bei den Schulstunden oder beim Arbeitsbeginn, Fahrplanänderungen oder Tarifumgestaltungen beim Bussystem und ähnliche Dinge haben ja durchaus nicht nur Auswirkungen auf entweder Schüler oder deren Eltern, sondern meist auf die Aktivitätsmuster der gesamten Familie. Mit Hilfe dieser Studien konnten die Forscher die gravierenden Veränderungen auf alle Haushaltsmitglieder näher untersuchen.

Ein weiterer Grund, wieso Studien über Aktivitätsmuster auch in Zukunft sehr gefragt sein werden, ist die zunehmende Vernetzung von Verkehr und Telekommunikation. Die rasante Entwicklung in der Computer- und Telekommunikationsbranche hat auch enorme Auswirkung auf die Verkehrsentwicklung; viele vorher benötigte Fahrten werden überflüssig, andere, eventuell von der Arbeit unabhängige, Wege kommen aufgrund der zusätzlich zur Verfügung stehenden Zeit wieder hinzu. In der Tat hat vor allem in den letzten Jahren der Freizeitverkehr bei allen Bevölkerungsgruppen und Gesellschaftsschichten deutlich zugenommen, genauere Untersuchungen mit Aktivitäts- und Verhaltensmustern sind aber mit Sicherheit noch von Nöten.

Politische und verkehrsplanerische Maßnahmen beeinflussen mit Sicherheit das Verkehrsverhalten der Menschen und deswegen sollte man vorher über diese Verhaltensmuster genau Bescheid wissen.

#### **4.4 WELCHE FAKTOREN BEEINFLUSSEN DIESE AKTIVITÄTSMUSTER ?**

Hierbei gibt es mehrere Erklärungsansätze: Chapin (1974) behauptet, daß Verkehrsverhalten und Beteiligung an Aktivitäten das Ergebnis von Entscheidungen innerhalb von Rahmenbedingungen bzw.

Zwängen seien, Geografen stufen jedoch meist die Rahmenbedingungen als die wichtigeren Faktoren ein, Entscheidungen werden als weniger ausschlaggebend bewertet.

Vor allem Hägerstrand (1970) sah die Zeit-Raum-Bewegungen der Individuen über den gesamten Tag gesehen in Abhängigkeit von drei Grund-Rahmenbedingungen: (1) *coupling constraints*, (2) *capability constraints* und (3) *authority constraints*.

*Coupling constraints* treten auf, weil die Rolle einer Person ihr vorschreibt, zu bestimmten Zeiten an gewissen Orten zu sein. Dazu gehören die Anwesenheit bei der Arbeit, Treffen anderer Personen, etc.. Zweitens die *capability constraints*, die mit den Mitteln zusammenhängen, die eine Person zur Verfügung hat, um am Verkehr teilzunehmen. Diese Rahmenbedingungen können der Besitz bzw. Zugang zu einem Kraftfahrzeug oder die Zugangsmöglichkeiten zum öffentlichen Verkehr sein. Die dritte Kategorie von Rahmenbedingungen sind *authority constraints*, die erklären, daß gewisse Aktivitätsorte nur zu gewissen Tageszeiten zugänglich sind (z.B. haben Banken oder Kinos nicht 24 Stunden pro Tag geöffnet, der Mensch muß sich anpassen). (Hägerstrand, 1970). Schon wegen dieser grundlegenden Rahmenbedingungen wird erwartet, daß jene Personen, bei denen *coupling* oder *capability constraints* enger sind, einen wesentlich niedrigeren Grad an Variabilität in ihrem Reiseverhalten aufweisen als Personen, die frei von solchen Rahmenbedingungen agieren können (Pas und Koppelmans, 1987).

In den modernen Studien über städtische Aktivitätsmuster beschränkt man sich generell auf zwei Arten erklärender Faktoren, die beide sowohl Entscheidungen als auch Rahmenbedingungen beeinflussen können. Ein Faktor sind Haushalts- und Familienvoraussetzungen bzw. -eigenschaften, der andere bezieht sich vornehmlich auf geografisch lokale Gegebenheiten, die auch die Charakteristiken des städtischen öffentlichen Verkehrsmittelnetzes beinhalten. Insgesamt sind jedoch die haushaltsabhängigen Einflußfaktoren laut einigen Studien deutlich wichtiger als geografische Strukturen.

#### 4.4.1 Haushaltseigenschaften

Zu den Haushaltseigenschaften gehören unter anderem Haushaltsgröße, Anzahl der beschäftigten Personen im Haushalt, Haushaltseinkommen, Autobesitz, aber natürlich auch Alter, Geschlecht und Ausbildungsstandard der einzelnen Familienmitglieder. Ein wichtiger Einflußfaktor ist neben der Rolle der einzelnen Person im Haushalt auch die Haushaltsstruktur. Mehrere Studien haben gezeigt, daß die Zugehörigkeit zu einer bestimmten Gesellschaftsschicht und die Zusammensetzung des Haushalts in bezug auf Altersstrukturen die Aktivitätsmuster des gesamten Haushalts stark beeinflussen. So haben Kinder eine Auswirkung auf die Muster der Erwachsenen im Haushalt, sind nur arbeitende Erwachsene

ne im Haushalt, so sehen die Aktivitätsmuster generell verschieden aus (S. Hanson und P. Hanson, 1981a; Jones, Dix, Clarke und Heggie, 1983; Pas, 1984; Townsend, 1991). Normalerweise wird diese Stufe im Lebenszyklus in einer nominal skalierten Variable bemessen, die Haushalte aufgrund der Anzahl, des Alters und dem Familienstand der Erwachsenen und dem Alter der Kinder in Kategorien einteilt.

Die Rolle des Individuum innerhalb des Haushaltes wird meist mit dem Erwerbstätigkeitsstatus und dem Geschlecht beschrieben. Der Grad der Erwerbstätigkeit beeinflusst die Häufigkeit der Teilnahme an diversen Aktivitäten (S. Hanson und P. Hanson, 1981a), das Geschlecht hat jedoch wahrscheinlich die größten Auswirkungen auf die Aktivitätsmuster. So sind die Wegdistanzen bei Frauen kürzer, ihre Muster bei den Variablen Verkehrsmittelwahl und Aktivitätsteilnahme unterscheiden sich signifikant (S. Hanson und P. Hanson, 1981b; Pas, 1984) und die Aktivitätsmuster variieren von Tag zu Tag bei Frauen wesentlich stärker als bei Männern (Jones und Clarke, 1988; Pas, 1988).

Goulias und Kitamura konnten 1992 in einer Simulationsstudie, die auf 6 verschiedenen Umfrage- wellen aus den Niederlanden zwischen 1984 und 1989 (Dutch National Mobility Panel) basierte, zeigen, wie sich Veränderungen im Familien- und Haushaltskontext (Struktur des Haushalts, Arbeitssituation, Einkommen, Pkw-Besitz, etc.) auf die Mobilität auswirken.

#### **4.4.2 Lokale und geografische Gegebenheiten**

Eine Schlüsselrolle bei diesem zweiten wichtigen Faktor, dem lokalen Kontext, spielen unter anderem Charakteristiken des Transportsystems, die Verteilung von möglichen Aktivitätsorten, das Wetter und Gefährdung der persönlichen Sicherheit.

Die ausschlaggebenden Komponenten beim Transportsystem sind sicherlich Kosten und Verfügbarkeit des Netzes im Vergleich zum eigenen Pkw bzw. die Strukturen der Tarifgestaltung: Ist der Fahrpreis distanzabhängig oder gilt er für die gesamte Region? Zahlt der Benutzer selbst oder kommt, wie in Japan, sein Arbeitgeber für die Fahrtkosten auf? Ideale Voraussetzungen für den Fahrradverkehr (z.B. Niederlande, China, Kalifornien) bzw. die Qualität des Radwegnetzes beeinflussen die Aktivitätsmuster, die Menschen machen in diesen Städten einfach mehr Wege mit dem Rad. Auch wenn die geografischen Voraussetzungen nicht perfekt sind, das Angebot an Radwegen, verkehrsberuhigten Zonen, genügend Abstellmöglichkeiten aber ausreichend ist, wird das Fahrrad häufig als Verkehrsmittel gewählt (Beispiel Uppsala, Schweden).

Die Verteilung der Aktivitätsorte fand bisher in den Studien vor allem in bezug auf die Dichte dieser Orte um die Heim- oder Arbeitsstätte Beachtung. Hanson und Schwab (1987) untersuchten die Zusammenhänge zwischen den Zugangsmöglichkeiten der Individuen zu verschiedenen Aktivitätsorten und ihren Aktivitätsmustern. Dabei zeigte sich, daß Verkehrsmittelwahl und Wegdistanzen mehr von den Zugangsmöglichkeiten abhängig waren als beispielsweise die Wegehäufigkeit.

Ein weiterer, in letzter Zeit immer wichtiger werdender Faktor ist die Komponente der persönlichen Sicherheit, speziell am Abend bzw. in der Nacht. Die Aktivitätsmuster von alten Menschen, Kindern und vor allem Frauen sind stark abhängig von der Sicherheit des Weges und des benutzten Verkehrsmittels. Um deren Mobilität zu erhöhen, sind viele Städte bestrebt, Plätze und Straßen zu beleuchten, Frauen-Nachttaxis zu finanzieren und allgemein eine sichere Stadt zu schaffen. Eine Studie in Chicago (Rollinson, 1991) ermittelte, daß „Sicherheitsgedanken die Aktivitätsmuster aller Menschen sehr stark beeinflussen“.

Abhängig vom Ziel der Studie wurden in den letzten Jahren auch noch verschiedenste andere mögliche Einflußfaktoren des lokalen Kontext untersucht. Lenntorp (1976) inkludierte die Öffnungszeiten der Geschäfte und anderer Einrichtungen (vergleichbar mit den authority constraints von Hägerstrand), S. Hanson und P. Hanson (1977) nahmen die täglichen Wettermeßwerte in ihre Untersuchungen und die Baltimore Studie berücksichtigte Parkkosten zu Hause und am Arbeitsplatz als mögliche Faktoren.

Weitere Einflußfaktoren sind sicherlich noch vorhanden, sowohl geografische als auch haushaltsabhängige, jedoch sind sie sehr schwer zu erfassen und einzuordnen (z.B. Psychologische Faktoren).

Leider sind Umfragen über längere Zeit, die jedoch für die genaue Analyse von Verhaltensmustern nötig sind, extrem teuer und werden daher nicht so oft durchgeführt. Einzig eine Befragung aus den Niederlanden (Dutch National Mobility Panel) erlaubt Einsichten in die Änderungen in solchen Mustern, da sie von 1984 bis 1989 erhoben wurde (Kitamura, 1991). Notwendig für aussagekräftige Verhaltensmuster ist natürlich auch eine längere Laufzeit einer Umfrage, wie beispielsweise in Uppsala über 5 Wochen. Ist die Laufzeit zu kurz (in Österreich werden z.B. die meisten Umfragen (vor allem aufgrund der Kosten) nur auf einen Tag begrenzt), so können eventuelle Zyklen, Wiederholungen oder Gewohnheitsmuster gar nicht erfaßt werden. Brög und Erl (1980) beispielsweise stellten Gewohnheiten im Verhaltensmuster fest, nachdem sie 1-Tages-Beobachtungen durchführten und die Teilnehmer fragten, was sie am Tag vorher machten bzw. wie ihr typischer Tagesablauf aussieht. Wie in einigen Studien gezeigt wurde, kommen bei Umfragen über den Zeitraum von einer Woche (z.B. die 1973 erhobene „Reading Activity Diary Survey (Pas und Koppelman, 1986)) zwar bis zu drei repräsentative Tage im Verhaltensmuster dieser Person vor, jedoch kann man aufgrund der begrenzten Umfragedauer keinerlei Zyklen erkennen (S. Hanson und P. Hanson, 1988a; siehe auch Kapitel 3.5).

## 4.5 BEGRIFFSKLÄRUNG

In der Fachliteratur kommen immer wieder Begriffe vor, deren genaue Bedeutung hier kurz geklärt werden soll.

### 4.5.1 Gewohnheitsmässiges Verhalten bzw. Wiederholung

Vorrangig wird gewohnheitsmässiges Verhalten erwartet, weil sich die moderne Gesellschaft extrem nach der Uhrzeit und der Notwendigkeit, die eigenen Verhaltensweisen mit denen anderer zu koordinieren und zu synchronisieren, richtet. Geschäfte, Schulen, Arbeitszeiten, Fahrpläne der Verkehrsmittel und sogar der Verkehrsfluß sind von der Uhrzeit abhängig und in den meisten Verhaltensmustern der Menschen wird dies reflektiert. Wechselwirkungen mit anderen Menschen, vor allem zu Hause oder am Arbeitsplatz, benötigen eine gewissen zeitliche und räumliche Festlegung bzw. Wiederholung, um zu guter letzt auch etwas vorhersehbar zu sein.

Nach Cullen (1978) muss gewohnheitsmässiges tägliches Verhalten erwartet werden, weil es für den Einzelnen ein Weg ist, mit der Komplexität und Vielfalt der städtischen Umgebung fertig zu werden, denn ein Routineverhalten zu verfolgen, erspart ihm, täglich mühevoll Entscheidungen treffen zu müssen.

### 4.5.2 Variabilität

Variabilität ist kurz erklärt die Abwechslung in den Verhaltensmustern der Personen.

Grundsätzlich kann man die Variabilität in zwei verschiedene Kategorien einteilen: einerseits die interpersonelle Variabilität, die Unterschiede im Verkehrsverhalten verschiedener Personen bzw. Haushalte aufgrund der Charakteristiken dieser Personen und ihrer Umgebung beschreibt, andererseits die intrapersonelle Variabilität, die tägliche Unterschiede in den Verhaltensmustern Variabilität einer einzigen Person erklärt. In den meisten Modellen und Analysen wurde bisher nur die interpersonelle Variabilität untersucht, vor allem aufgrund des Fehlens geeigneter Datensätze. Nur eine längere Befragung (zumindest über eine Woche) kann eine geeignete Grundlage für die Untersuchung von intrapersoneller Variabilität darstellen. Die fünfwöchige Befragung aus Uppsala stellt so einen geeigneten Datensatz dar und wurde deshalb wiederholt für derartige Untersuchungen verwendet (S. Hanson und P. Hanson, 1981, 1986 und 1988; Huff und Hanson, 1988,; etc.).

Ein Grund für Variabilität ist sicherlich die Langeweile der kompletten Wiederholung und der Wunsch, immer den Streß der Entscheidungsfällung und des andauernden Suchens nach neuen Aktivitäten, Verkehrsmitteln oder Zielen zu minimieren und insofern auszugleichen, daß die räumlichen und zeitlichen Eigenschaften vor allem von nicht obligatorischem Verkehr (also Freizeitverkehr, gewisse Erledigungen, etc., eben nicht tägliche Wege, die gemacht werden müssen) abgewechselt und ausgetauscht werden.

Einige Erklärungsversuche für andauernde Variabilität in der Zielwahl nach dem Ansatz für gewohnheitsmäßiges Verhalten wurden von Smith (1978) („der grundlegende Mechanismus des Verhaltens ist der Wunsch der Menschen, Risiken zu verteilen“), Hay und Johnston (1979) („die Variabilität in der Wahl der Ziele reflektiert die Notwendigkeit von kontinuierlichem Suchen angesichts immer neuer Angebote“) und auch von Hanson (1980) („nachdem die Kombination von Waren und Service zu verschiedenen Zeiten unterschiedlich sind, variieren auch die Destinationen (einer funktionalen Klasse) in einem Mehrzweckweg dementsprechend.) gemacht. Ähnliche Ansichten vertreten Brög (1980) bzw. Brög und Erl (1983): „Verhaltensweisen variieren in kurzen Abständen, da sich die Umwelt, die einen entscheiden läßt und in der Verkehr stattfindet, ebenfalls schnell ändern und auch die Rahmenbedingungen, die einen Großteil des Verhaltens beeinflussen, von Tag zu Tag variieren“.

Grundsätzlich ist es jedoch notwendig, gewohnheitsmäßigen Verhaltensweisen noch genauer zu untersuchen, denn auf diese Annahme sind im Prinzip bis jetzt alle räumlichen Wahl-Modelle aufgebaut. Denn oft wurden nur einzelne Aktivitäten beobachtet und falls diese wieder auftauchten, wurde das Aktivitätsmuster als sich grundsätzlich wiederholend angesehen, Variabilität (in Verkehrsmittel oder Ziel) wurde zwar eingeräumt, jedoch eher im Zusammenhang mit der Theorie, dass der Mensch andauernd neues dazulernt.

Je komplexer die Verhaltensmuster sind, die untersucht werden, desto weniger Gewohnheit steckt in ihnen und desto größer ist die Variabilität.

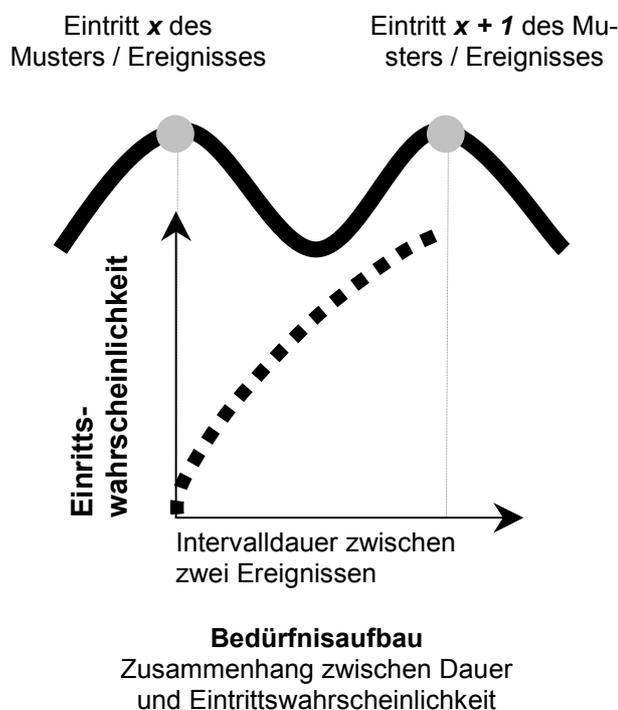
### **4.5.3 Rhythmische Muster und Periodizität**

Diese beiden Begriffe hängen eng mit den in den vorigen Abschnitten definierten Begriffen zusammen, wurden jedoch erst in den 90-er Jahren vermehrt verwendet. Jede Person mit einem gewohnheitsmäßigen Verhalten hat bestimmte rhythmische Muster in ihrem Verkehrsverhalten, die sich wiederholen. Diese Wiederholung bzw. Periodizität kann in täglichen, wöchentlichen oder durchaus auch anderen Abständen vorkommen. Jede Person hat aber nicht nur ein einziges rhythmische Muster, son-

dem weist für verschiedene Aktivitäten auch unterschiedliche Muster auf. Um eine erneute Aktivität durchzuführen, muß jedoch erst ein Bedürfnisaufbau stattfinden, der durch eine Wahrscheinlichkeitsfunktion ausgedrückt werden kann. Am Ende bzw. Höhepunkt dieses Bedürfnisaufbaus findet dann ein erneutes Ereignis statt bzw. wird in diesem Fall eine erneute Aktivität durchgeführt. Beispielsweise wird eine Person erst dann wieder für den täglichen Bedarf einkaufen, wenn neue Lebensmittel von Nöten sind, nicht aber sofort nach dem soeben getätigten Einkauf. Genauso wenig wird ein Kinobesuch sofort an den nächsten anschließen, sondern erst dann wieder stattfinden, wenn das Verlangen danach groß genug ist.

Die Periodizität des Verkehrsverhaltens und die Rhythmik der Verhaltensmuster stellen die Grundlagen des aktivitäts-basierten Ansatzes (*activity based approach*) dar, der vor allem von Axhausen und Polak (1991) geprägt wurde. Dabei werden die rhythmischen Muster als Ereignisse angesehen, deren Eintrittswahrscheinlichkeit in Abhängigkeit der schon vergangenen Zeitspanne seit dem letzten gleichartigen Muster bzw. Ereignis geschätzt werden können (Schönfelder und Axhausen, 2000b). In der folgenden Abbildung 4.1 wird das Basiskonzept des rhythmischen Verkehrsverhaltens mit dem oben angeführten Bedürfnisaufbau und dem daraus resultierenden Ereigniseintritt, der periodisch vor sich geht, gezeigt (siehe Kapitel 7.3.2).

Abbildung 4.1 : Basiskonzept: Rhythmik des Verkehrsverhaltens



Quelle: Schönfelder und Axhausen, 2000b

## 5 HAZARD MODELLE: EINLEITUNG

### 5.1 HINTERGRUND

Daten, die zeitliche Angaben beinhalten und bei denen zeitliche Abfolgen bis zum Auftreten eines bestimmten Ereignisses von Bedeutung sind, werden in der Literatur üblicherweise als *Duration*<sup>1</sup> Daten (Zeitdauer- oder Überlebensdaten) bezeichnet (Hensher und Mannering, 1993). Eine sehr gängige Möglichkeit, solche Daten zu analysieren, stellen Hazard Modelle bzw. *Duration* Modelle (Risiko- bzw. Zeitdauermodelle) dar, die in den Kontext der Überlebensanalyse (*Survival Analysis*)<sup>2</sup> fallen.

In der Regel verwendet man zur Untersuchung von Verteilungen und des Einflusses bzw. der Korrelation von Determinanten eines bestimmten Phänomens lineare Regressionsmodelle oder ähnliche stochastische Analysemethoden. Bei Zeitdauerdaten können jedoch klassische Regressionsmodelle, die mit kleinster-Quadrate-Schätzung arbeiten, nicht verwendet werden, da empirische Zeitdauern nur positive Werte annehmen können. Eine weitere Eigenheit von Zeitdauerdaten, das sogenannte *censoring problem* (siehe Kapitel 5.3.2), macht zudem herkömmliche statistische und stochastische Methoden (beispielsweise die Mittelwertberechnung) unmöglich (Schönfelder und Axhausen, 2000). Diese Restriktionen haben zur Entwicklung der sogenannten *survival analysis* geführt. *Survival Analysis* ist eine Gruppe statistischer Methoden, deren Ziel es ist, Auftreten und zeitliche Abfolgen von Ereignissen zu studieren und die beobachtete Verteilung mit Hilfe von Hazard Funktionen abzubilden.

Ursprünglich wurden Hazard Modelle in der Medizin, der Biochemie und der Biologie entwickelt, um zu sehen, wann ein bestimmtes Ereignis (das Ableben) unter der Voraussetzung eines exakt definierten Anfangspunktes und einer Zeitskala auftritt und welche Einflußfaktoren dabei eine Rolle spielen.

---

<sup>1</sup> Im Folgenden wird für Überlebenszeit bzw. Zeitdauer der englische Ausdruck *duration* verwendet,

<sup>2</sup> Weiters wird für die Überlebensanalyse der englische Ausdruck *Survival Analysis* benützt.

In der Medizin wird die Zeit zwischen der Diagnose eines bestimmten Zustandes und dem Ende der Behandlung ( positiv - Entlassung oder negativ - Ableben) oder der Operation und dem Tod des Patienten als *Duration Time* (Überlebenszeit) definiert, das Ereignis ist jedoch meist negativ (Tod). Aus diesem Grund werden *Duration* Daten in der Medizin als Überlebensdaten und die zugehörigen statistischen Analysen als Überlebensanalysen bezeichnet.

In der Biologie werden Hazard Modelle beispielsweise für kleinere Biosysteme verwendet, wo nach einem Zusammenbruch (Kollaps) des Systems oder dem Aussterben bestimmter Populationen die Entwicklung unter gewissen Voraussetzungen (Temperaturen, Umwelteinflüsse, menschliche Regulierung, etc.) auf einer Zeitachse bis zum nächsten Zusammenbruch beobachtet wird.

Auch in der Wirtschaft und in sozialen Belangen werden Hazard Modelle verwendet. Zum Beispiel wird die Zeit, die ein vormals Arbeitsloser braucht, um wieder Arbeit zu finden bzw. an einem Umschulungsprogramm teilnehmen kann ( Pedessen,1991) oder die Zeit bis zu einer erneuten Straftat eines ehemaligen Häftlings (Kleinbaum,1996) oder die Rückfallzeit eines Drogensüchtigen nach einer Entziehungskur in solchen *Duration* Datensätzen angegeben.

In der Technikbranche kann die Zeit bis zum Versagen eines Bauteils bei elektronischen oder mechanischen Geräten mit derartigen Daten am besten angezeigt und mit Hazardmodellen analysiert werden (Filiou,1994; Wilson, 1994).

Ein weiteres Anwendungsgebiet für Hazard Modelle stellt die Erdbebenforschung dar. Hierbei wird vor allem die Wahrscheinlichkeit des nächsten Ausbruchzeitpunktes an einem bestimmten Ort berechnet und vorausgesagt, bezugnehmend auf den letzten Ausbruch und tektonische bzw. geologische Voraussetzungen.

Erst seit den 80er Jahren werden Hazard Modelle auch in einigen Bereichen der Verkehrsverhaltensforschung angewendet. So wird zum Beispiel die Zeit, die vergeht, bis eine Person ihre täglichen Routen oder benützten Verkehrsmittel ändert, in solchen Daten erfaßt; Oder die Zeit zwischen Verkehrsunfällen (beispielsweise von Taxifahrern (Hamed und Hensher, 1995), die Zeit, bis ein Verkehrsunfall oder -hindernis beseitigt ist und die Straße wieder frei befahrbar ist, Umsteigzeiten bei öffentlichen Verkehrsmitteln oder Wartezeiten von Autoschlangen vor Ampeln, Mautstellen und internationalen Grenzübergängen (Hensher und Mannering, 1993). Auch bei den Verhaltensmustern von einzelnen Personen, beispielsweise wieviel Zeit zwischen den einzelnen Entscheidungen, Wege zu machen bzw. zwischen der Entscheidung und der tatsächlichen Abfahrt, vergeht, spielen diese Daten und Modelle eine große Rolle. Einige Beispiele sind im Kapitel 7 genauer beschrieben.

## 5.2 HAZARD MODELLE

Alle standardmäßigen Annäherungen an die *Survival Analysis* sind probabilistisch oder stochastisch. Das bedeutet, daß Zeitpunkte, an denen Ereignisse stattfinden, stark mit Zufallsprozeß und Zufallsvariablen in Zusammenhang stehen. So werden dafür der Großbuchstabe  $T$ , der die Zufallsvariable für die Überlebenszeit einer Person oder das Auftreten eines bestimmten Ereignisses darstellt und alle nichtnegativen Werte annehmen kann und das kleine  $t$  als jeder beliebige Wert, der in bezug auf  $T$  von Interesse ist, eingeführt. Wollen wir beispielsweise wissen, ob eine Person nach einer Operation 3 Jahre überlebt, so nimmt  $t$  den Wert 3 an und wir schauen dann, ob  $T$  diesen Wert 3 erreicht. Falls ja, wird eine zusätzliche binomiale Variable  $\delta$  (Delta), die die Werte 0 und 1 annehmen kann, auf 1 gesetzt. Bei Nichtauftreten des Ereignisses wird die Überlebenszeit durch das Ende der Beobachtungszeit zensuriert und  $\delta = 0$  gesetzt.

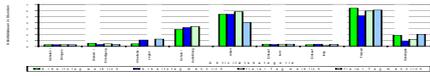
Generell gibt es drei verschiedene Arten, Wahrscheinlichkeitsverteilungen zu beschreiben:

- (a) Kumulative Verteilungsfunktion (KVF)
- (b) Dichte-Wahrscheinlichkeitsverteilung (DWV)
- (c) Hazard Funktionen

Die kumulative Verteilungsfunktion (KVF) kann bei allen Zufallsvariablen angewendet werden. Sie gibt uns die Wahrscheinlichkeit  $f(t)$  an, daß die Variable weniger oder gleich einem von uns gewählten Wert  $t$  ist, das heißt  $F(t) = P(T \leq t)$ . Wenn wir den Wert von  $F$  für jeden Wert  $t$  wissen, können wir die Verteilung  $T$  komplett beschreiben. In der *Survival Analysis* wird jedoch die genau entgegengesetzte Funktion  $S(t)$ , die sogenannte *Survival*-Funktion, verwendet.  $S(t)$  gibt die Wahrscheinlichkeit an, daß eine Person länger als eine beliebig festgelegte Zeit  $t$  überlebt bzw. daß die Zufallsvariable  $T$  die spezifizierte Zeit  $t$  erreicht, d.h.  $(S(t) = P(T > t) = 1 - F(t))$ . Die Funktion kann natürlich grafisch dargestellt werden, wobei  $S(t=0)$  immer den Wert 1 annimmt und  $S(t = \infty) = 0$  ist. Die Kurve ist zudem stetig fallend. In der Praxis wird die Kurve meist als Stufenfunktion (siehe Graf) angegeben und nimmt zudem am Ende der Studie (bzw. des Beobachtungszeitraums) noch nicht exakt den Wert 0 an, da ja nicht jeder das Ereignis erfahren hat (siehe Kapitel 5.3.2).

Wenn die Zufallsvariablen kontinuierlich sind (wie beispielsweise die Zeit), ist die Dichte-Wahrscheinlichkeits-Verteilung (kurz DWV) eine weitere gängige Art der Beschreibung. Sie ist definiert als  $f(t) = dF(t) / dt = -dS(t) / dt$ , das heißt, sie ist die erste Ableitung der KVF nach der Zeit. Obwohl diese Definition auf den ersten Blick weniger einleuchtend erscheint, hat diese Form weit mehr Bezug zu unserer Vorstellung von Wahrscheinlichkeitsverteilungen bzw. der bildlichen Form von Verteilungen.

Für kontinuierliche Überlebensdaten (*duration data*) wird jedoch am häufigsten die Hazard Funktion verwendet. Sie ist definiert als



$$(1)$$

wobei manche Autoren  $h(t)$  auch oft mit  $\lambda(t)$  oder  $r(t)$  bezeichnen, hier wird jedoch grundsätzlich  $h(t)$  (außer  $\lambda_0(t)$  für die Basishazardfunktion (*baseline hazard*)) verwendet.

Das Hauptziel der Hazardfunktion  $h(t)$  ist - im Gegensatz zur Überlebensfunktion - , die unstetige (*instantaneous*) Wahrscheinlichkeit, daß ein Ereignis in der Zeit  $t$  stattfindet, zu beschreiben bzw. zu quantifizieren. Sie wird auch oft als negative oder umgekehrte Überlebensfunktion bezeichnet. Je größer  $S(t)$  für ein gegebenes  $t$  ist, desto kleiner ist  $h(t)$ . Unstetig ist das Potential aufgrund der Tatsache, daß es nur für einen bestimmten Moment angegeben werden kann und sich im Laufe der Zeit durchaus verändern kann (Beispiel: 60 km pro Stunde (60 km/h) bedeutet, daß ich vom Anfangszeitpunkt in einer Stunde exakt 60 km weit fahren kann, wenn ich die Geschwindigkeit über die gesamte Zeitdauer konstant beibehalte.).

Nachdem die Zeitkurve kontinuierlich verläuft, ist die Wahrscheinlichkeit, daß ein Ereignis exakt zum Zeitpunkt  $t$  stattfindet, gleich Null. Wir können jedoch die Wahrscheinlichkeit betrachten, daß ein Ereignis im kleinen Zeitintervall  $t + \Delta t$  stattfindet. Zusätzlich wollen wir, daß dies eine bedingte Wahrscheinlichkeit ist, abhängig davon, ob die Person die Beobachtungszeit bis zum Zeitpunkt  $t$  überlebt hat. Denn bei einer bereits verstorbenen Person (bzw. nach dem Ereignis selbst) ist kein Risiko zu sterben (bzw. das Ereignis zu erfahren) mehr vorhanden. So ist die Hazardfunktion als unstetiges Potential pro Zeiteinheit für das Auftreten des Ereignisses unter der Voraussetzung, daß die Person bis zum Zeitpunkt  $t$  überlebt hat bzw. das Ereignis noch nicht stattgefunden hat, zu sehen. Sie ist demnach eine bedingte Wahrscheinlichkeit (*conditional probability*).

Wie man sieht, sind die drei Beschreibungsformen für kontinuierliche Wahrscheinlichkeitsverteilungen (*survivor function, probability density function, hazard function*) sehr ähnlich und man kann ohne größere Probleme die Zusammenhänge mathematisch darstellen:

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \quad (2)$$

$f(t)$  corresponding density function

$F(t)$  cumulative distribution function

## 5.3 ANSATZ, AUSWERTUNG UND DARSTELLUNG VON HAZARD MODELLEN

### 5.3.1 Einleitung

Allgemein kann die Hazard Funktion auch als  $h(t | X)$  angegeben werden, wobei  $X$  ein Vektor erklärender Variablen ist. Um die *Duration* Modelle, denen Hazardfunktionen zugrunde liegen, besser abschätzen zu können, werden neben der Zeitabhängigkeit auch die Zusammenhänge und Wechselwirkungen zwischen Zeit und beschreibenden Variablen, die sich klarerweise während der Beobachtungsdauer verändern können, berücksichtigt und in die Modelle eingebaut. Grundsätzlich gibt es in der *Survival Analysis* eine Reihe verschiedener Rechenmodelle: *Lifetables*, *Kaplan-Meier-Estimators*, *exponential regression*, *log-normal regression*, *proportional hazard regression*, *competing risk models* und *discrete-time methodes*, um nur einige zu nennen. Meist ist die verwendete Art des Modells abhängig vom Datensatz und der Art der Auswertung bzw. der Aussage. Oft sind durchaus zwei verschiedene Modellarten anwendbar und nur durch die Einschränkung, ob beispielsweise die Zeit als diskrete oder kontinuierliche Variable angenommen wird, erhält man mit einem Modell aussagekräftigere Ergebnisse. Bei Hazardmodellen wurden prinzipiell zwei verschiedene Modellansätze entwickelt [Jones, Janssen & Mannering, 1991] :

- *proportional hazard models* (mehr dazu siehe Kapitel 6.3.1)
- *accelerated lifetime models* (mehr dazu siehe Kapitel 6.4.1)

*Proportional Hazard Models* beruhen auf der Annahme, daß Kovariablen (z.B. Faktoren, die die Dauer beeinflussen) multiplikativ auf eine bestimmte grundlegende Hazard Funktion (*baseline hazard*) wirken. Die Proportionalität besteht darin, daß sich die Hazard Rate in einen zeitabhängigen und in einen Kovariablen-abhängigen Term gliedern läßt (Hensher und Mannering, 1993). Die Wirkung von Zeitregressoren (erklärenden Variablen) besteht in der Multiplikation der *baseline hazard* mit einem Skalierungsfaktor, was mathematisch folgendermaßen beschrieben werden kann [Jones et al., 1991]:

$$h(t, \beta, X, h_0) = h_0 * t * y(\beta, X) \quad (3)$$

$t$  ..... Zeit

$X$ ..... Vektor von erklärenden/beschreibenden Variablen

$\beta$  ..... Vektor von Abschätzparametern

$h_0(t)$ ... Basis-Hazard-Funktion

$y(\beta)$  .. Skalierungsfaktor

Die *accelerated Lifetime Modelle* dagegen verwenden die beschreibenden Variablen, um die Zeit (z.B. *accelerate time*) direkt in einer grundlegenden Überlebensfunktion abzubilden. Diese Funktion tritt genau dann auf, wenn alle Kovariablen gleich Null sind. Der Hazard sieht dann folgendermaßen aus:

$$h(t, \beta, X) = h_0[t * y(\beta, X)] * y(\beta, X). \quad (4)$$

Für beide Modelle, also sowohl *Proportional Hazard Models*, als auch *Accelerated Lifetime Models* hat der Skalierungsfaktor diese Struktur:

$$y(\beta, X) = \exp(\beta, X) \quad (5)$$

wobei der Parameter  $\beta$  mit Hilfe von *maximum likelihood*-Techniken abschätzbar ist [Jones et al., 1991].

Eine grundlegende Aufgabe in der *Survival Analysis* ist es mit Sicherheit, die maßgebenden Variablen herauszufinden.

Jede Hazard Funktion hat eine korrespondierende Wahrscheinlichkeitsverteilung, die, je komplizierter die Hazard Funktion ist, schwieriger anzugeben und vor allem zu interpretieren ist. Deshalb werden hier einige häufig verwendete, simple Hazardmodelle diskutiert:

Die einfachste Funktion beschreibt einen über die Zeit konstanten Hazard, das heißt, daß  $h(t)=\lambda$  oder äquivalent  $\log h(t)=\mu$  ist und so die zugehörige Überlebensfunktion  $S(t)=e^{-\lambda t}$  ist. Dieses Überlebensmodell wird als exponentiell bezeichnet.

Der nächste Schritt in der Komplexität der Funktionen ist, den natürlichen Logarithmus des Hazard eine lineare Funktion der Zeit sein zu lassen:  $\log h(t)=\mu+\alpha t$  bzw.  $h(t)=\lambda\gamma^t$ , wobei  $\lambda=e^\mu$  und  $\gamma=e^\alpha$  sind. Diese Art der Darstellung nennt man *lognormal survival* und sie folgt der Gompertz Verteilung (z.B. Tuberkulose Patienten).

Alternativ dazu kann  $\log h(t)=\lambda t^\alpha$  mit  $\lambda=e^\mu$  als Hazard Funktion angenommen werden. Sie wird Weibull-Verteilung bezeichnet, wobei zwischen einer wachsenden Weibullverteilung (z.B. für Leukämie-Patienten, wo das Risiko zu sterben immer größer wird, je weiter die Beobachtungszeit fortgeschritten ist) und einer fallenden Weibullverteilung (z.B. Erholung nach einer Entziehungskur) unterschieden wird. Die drei verschiedenen Modelle gehören alle in die Gruppe der proportionalen Hazard Modelle.

### 5.3.2 *Censoring problem*

Die Datensätze sind jedoch nicht immer vollständig. Für die meisten Überlebensanalysen muß ein Schlüsselproblem, das in der Literatur unter *censoring problem* (Zensurierung) bekannt ist, betrachtet werden. Dies taucht auf, wenn wir zwar einige Informationen über die individuelle Überlebenszeit haben, sie jedoch nicht genau kennen (Kleinbaum 1996). Einzelne Beobachtungen müssen auch aufgrund von Zeitüberschreitungen oder Unvollständigkeit zensuriert werden.

Meistens wird bei der Zensurierung das gesamte, aber nicht exakt bekannte Überlebenszeitintervall *rechtsseitig* abgeschnitten. Dies tritt ein, wenn der Beobachtungszeitraum endet, bevor die betreffende Person das Ereignis erfahren hat (z.B. bei medizinischen Daten, wenn der Patient über den Beobachtungszeitraum hinaus überlebt) oder die Person nicht mehr beobachtet werden kann (beispielsweise bei schlechter Betreuung oder einem Umzug der Person) oder die Person sich von der Studie zurückzieht bzw. keine Beobachtungen mehr zuläßt.

Neben dieser rechtsseitigen Zensurierung gibt es noch eine weit weniger häufig auftretende *linksseitige Zensurierung*, bei der man den genauen Anfang- oder Auslösezeitpunkt, der zeitlich vor dem eigentlichen Beobachtungsbeginn liegt, nicht kennt. Ein Beispiel dafür ist die Infizierung mit dem HIV-Virus, bei der der genaue Zeitpunkt meist unbekannt ist und der eigentliche Beobachtungszeitraum erst mit der Diagnose beginnt.

Auch manche Beobachtungen für Verkehrsverhalten sind linkszensuriert, weil man nicht weiß, wann beispielsweise die letzte Aktivität „Einkaufen“ oder der letzte Unfall, etc. vor dem Beobachtungsbeginn gewesen ist. Diese Zensurierung wird in dieser Arbeit jedoch weitgehend vernachlässigt.

Der dadurch entstehende Unterschied zwischen zensurierten Ereigniszeiten und tatsächlich beobachteten Überlebenszeiten wird oft in einer eigenen zusätzlichen Variable in der Analyse berücksichtigt (Hensher und Mannering, 1993, Hamed, 1993 ).

Manche Computerprogramme für *Survival Analysis* erlauben keine linkszensurierten Datensätze, die hier verwendeten jedoch schon. So arbeitet das *LIFEREG*-modell beispielsweise sogar mit Intervallzensurierung, also kombinierten rechts- und linkszensurierten Daten.

Es gibt zudem noch verschiedene Arten des Zensurierens:

*Typ I, singly censored*, wobei Typ I heißt, daß eine fixierte Zensurzeit vorhanden ist und *singly*, daß alle Beobachtungen dieselbe Beobachtungsdauer haben.

*Typ II* liegt dann vor, wenn die Beobachtung nach einer bestimmten Anzahl von eingetretenen Ereignissen abgebrochen wird.

**Zufällige Zensierung** (*random censoring*), die dritte Möglichkeit, tritt auf, wenn Beobachtungen aus Gründen, die nicht im Einflußbereich des Umfrageleiters stehen, beendet werden (beispielsweise Umzug, Rückzug von der Studie, Tod [wenn ein anderes Ereignis maßgebend ist], etc.) oder auch, wenn die einzelnen Personen zufällig verteilte Beobachtungseintritte, jedoch ein fixes Beobachtungsende haben (z.B. Untersuchung der Überlebensdauer bis zu einem fixen Zeitpunkt nach einer Herzattacke).

Die Mehrzahl der Computerprogramme behandelt jedoch alle verschiedenen Arten als rechtszensuriert, bei zufälliger Zensierung muß zusätzlich noch der Informationsgehalt ausreichend sein.

### 5.3.3 *Time of Origin* - Anfangszeitpunkt

Alle Modelle für Überlebensdaten sind grundlegend mit dem zeitlichen Eintreten von Ereignissen verknüpft. Dabei gibt es einerseits eine zeitliche Skala, die in Minuten, Stunden, Wochen, Monaten oder Jahren problemlos unterteilt und bemessen werden kann (linear oder linear im Logarithmus des Hazard oder der Ereigniszeit), andererseits den Anfangszeitpunkt, der nicht immer leicht festzustellen ist bzw. manchmal auch gar nicht bekannt ist.

Es gibt mehrere Möglichkeiten, den geeignetsten Anfangszeitpunkt anzugeben, was sehr stark von der Art der Studie abhängt. Für einfache Studien über die Lebensdauer einer bestimmten Bevölkerungsschicht wird das Alter gebraucht und dementsprechend die Geburt als Anfangszeitpunkt genommen. Will man jedoch die Lebensdauer beispielsweise nach einem Schlaganfall oder bei Krebserkrankung untersuchen, so ist der geeignetere Anfangszeitpunkt das Erkrankungsdatum. Sollen in einer Studie verschiedene Behandlungsarten miteinander verglichen werden, so ist nicht der Ausbruch der Krankheit, sondern der Behandlungsbeginn der entscheidende Anfangszeitpunkt. Die Zeit zwischen Krankheitsausbruch und Behandlungsbeginn kann mit Hilfe einer Kovariable in die Analyse miteinfließen.

Es gibt noch einige weitere Anfangszeitpunkte, die je nach Ziel der Studie gewählt werden können:

- **Kalendarisches Datum:** Werden alle Studienteilnehmer ab dem selben Tag beobachtet und liegen keine näheren Informationen über die einzelnen Personen vor, muß dieses Datum als Anfangszeitpunkt genommen werden.

- Zeitpunkt eines anderen Ereignisses, das mit der Studie in Zusammenhang steht: Bei Studien über den Scheidungszeitpunkt beispielsweise wird das Hochzeitsdatum als maßgeblichen Anfangszeitpunkt herangezogen.
- Zeitpunkt des letzten gleichartigen Ereignisses: Ist ein Ereignis wiederholbar, ist sicherlich das letztmalige Auftreten der geeignetste Anfangszeitpunkt.

Für die hauptsächlich verwendeten Zeit-kontinuierlichen Methoden (*continuous-time-methods*) ist ein einziger, eindeutiger Anfangszeitpunkt notwendig, bei den diskreten Methoden (*discrete time methods*) können auch mehrere Zeitpunkte flexibler gewählt werden.

Für die Auswertung ist es wichtig, den Anfangszeitpunkt mit der stärksten bzw. effektivsten Wirkung auf den Hazard zu nehmen, da sonst der Hazard unkontrollierbar wird bzw. andere Kovariablen (vor allem zeitabhängige) grundlegend beeinträchtigt.

### 5.3.4 Interpretation der Hazard Funktion

Obwohl es hilfreich sein kann, den Hazard als unstetige Wahrscheinlichkeit eines Ereignisses am Zeitpunkt  $t$  zu sehen, ist er nicht wirklich eine Wahrscheinlichkeit, da er infolge der Definition (Division durch  $\Delta t$ ) auch größer als 1,0 sein kann. Trotz der fehlenden Obergrenze ist er jedoch immer größer als 0! Hazardraten sind demnach latente Variablen, die als Intensität des Übergangs von einem Zustand in einen weiteren aufgefaßt werden können (Schneider, 1991).

Nachdem der Hazard als Wahrscheinlichkeit angegeben wird (und nie direkt gemessen werden kann), ist er eine unbeobachtete Meßzahl. Wir können ihn mit Daten abschätzen, was aber immer noch eine Abschätzung bleibt.

Der Hazard bleibt eine Charakterisierung des Individuums und nicht der Masse bzw. Gesamtbevölkerung (obwohl jeder Einzelne in dieser Masse gleich ist). Jeder Einzelne mag eine von Grund auf verschiedene Hazard Funktion haben (siehe dazu Kapitel 6.3.3: Problem der *heterogeneity effects*).

Da der Hazard eine Abschätzung der Form „Anzahl von Ereignissen pro Zeitintervall“ ist, wird er manchmal als *rate* bezeichnet. Um den Wert des Hazard interpretieren zu können, muß die Maßeinheit der Zeit bekannt sein (beispielsweise Monate, Wochen). Ist z.B. mein Hazard, an Grippe zu erkranken zu einem bestimmten Zeitpunkt 0,015 (1,3) mit der Zeiteinheit Monate (Jahre), so kann ich erwarten,

im Zeitabschnitt Monat 0,15 mal (bzw. 1,3 mal in diesem Jahr) an Grippe zu erkranken, unter der Voraussetzung, daß mein Hazard über diesen Monat (Jahr) konstant bleibt.

Der Hazard ändert sich jedoch ständig mit den Randbedingungen. Sitze ich beispielsweise im Büro, ist mein Hazard, einen schweren Unfall zu haben, der einen Krankenhausaufenthalt erfordert, natürlich geringer, als jener Hazard, wenn ich die Stiege hinuntergehe, mich im Verkehr fortbewege oder eine gefährliche Sportart ausübe. Ebenso ändert sich mein beispielsweise der Hazard bzw. das Bedürfnis, einen bestimmten Ort zu besuchen oder eine bestimmte Aktivität durchzuführen, je nachdem wann ich das letzte mal dort war oder die Aktivität vorher durchgeführt habe (vgl. Bedürfnisaufbau, Schönfelder und Axhausen, 2000). Die Hazard Funktion ist offensichtlich speziell individuell und zeitlich sehr stark von den Randbedingungen und Umwelteinflüssen abhängig. Mit Hilfe von zeitabhängigen Kovariablen, die die veränderlichen Randbedingungen beinhalten, können wir in einem Regressionsmodell (Cox Regression Modell - *PROC PHREG*, siehe Kapitel 6.3) auch die Auswirkungen auf den Hazard einer ganzen Gruppe von Individuen abschätzen.

## 6 MODELLFORMEN

### 6.1 EINLEITUNG

In den letzten Jahren wurden viele Modelle für die EDV-mäßige Verarbeitung von *Duration* Daten implementiert. Im Kapitel 6 werden einige für SAS relevante und für unsere Daten interessante Hazard-Modelle kurz vorgestellt. Grundsätzlich kann zwischen drei Stufen der Modellentwicklung unterschieden werden: Nonparametrische, Semiparametrische und Parametrische Modelle.

Bei den Nonparametrischen Modellen wird auf die Untersuchung von Determinanten verzichtet und nur eine deskriptive bzw. univariate Analyse durchgeführt. Ziel dabei ist es, basierend auf den empirischen Beobachtungen der Aktivitätsausübung eine Schätzung der Survival Funktion und der korrespondierenden Hazard Funktion zu erhalten, um Verteilungen möglichst wahrheitsgetreu abbilden zu können.

Bei den semiparametrischen Modellen wirkt ein Vektor ( $X$ ), der alle zusätzlichen erklärenden Variablen beinhaltet, multiplikatorisch auf eine grundlegende Hazard-Funktion (*baseline hazard*). Es muß dabei aber keine exakte Verteilung angenommen werden, womit gewisse Fehlerquellen ausgeschlossen werden und das Modell (im Gegensatz zu vollparametrischen Modellen) robuster wirkt. Bei diesen proportionalen Hazardmodellen oder Cox-Modellen ändern sich somit die Eigenschaften der Hazardfunktionen proportional zur Ausprägung der Kovariablen, wenn der *hazard ratio* konstant bleibt. Ein weiterer Vorteil der semiparametrischen Modelle ist die Möglichkeit, zeitabhängige Kovariablen in die Berechnung miteinfließen zu lassen.

Bei vollparametrischen bzw. parametrischen Modellen wird eine exakte Verteilungsannahme der zeitlichen Abläufe des *baseline hazard* getroffen. Hier wird mit sogenannten *Accelerated Failure Time Models* (AFT-Modelle) gearbeitet, die dann mit Hilfe der Maximum Likelihood Methode geschätzt werden.

## 6.2 NONPARAMETRISCHE MODELLE MIT *PROC LIFETEST*

Vor 1970 war die Abschätzung von Überlebensfunktionen (*survivor functions*) mit Hilfe von Überlebenskurven die wichtigste und vorherrschende Methode der *Survival Analysis*, heutzutage liegt der Schwerpunkt bei der Cox Regression (siehe *PROC PHREG*) (Allison, 1995). Trotzdem werden auch heute noch sehr oft Überlebenskurven für die Vorausscheidung von Daten, die Berechnung aus Regressionsmodellen und die Überprüfung der Paßformen von Regressionsmodellen verwendet. *PROC LIFETEST* wendet zwei Methoden für die Abschätzung von Überlebensfunktionen an:

- Kaplan-Meier-Methode, die vor allem für kleinere Datensätze mit exakt bemessenen Ereigniszeiten geeignet ist.
- *Life-table-* oder *actuarial-*Methode, die bei größeren Datenmengen oder ungenaueren Messungen der Ereigniszeiten zur Anwendung kommt.

Zusätzlich zur Berechnung und Darstellung von geschätzten Überlebensfunktionen sieht *PROC LIFETEST* drei alternative Statistiken vor, die Nullhypothese zu überprüfen, daß die Funktionen von zwei oder mehr verschiedenen Gruppen beispielsweise bei unterschiedlichen experimentellen Behandlungsmethoden gleich sind ( $S_1(t) = S_2(t)$  für alle  $t$ ). Es sind dies der Log-rank-Test (auch als Mantel-Haenszel-Test bekannt), der Wilcoxon-Test und die Likelihood-ratio-statistic, die unter der zusätzlichen Annahme von exponential verteilten Ereigniszeiten berechnet werden kann.

### 6.2.1 Kaplan-Meier-Schätzwert

Der Kaplan-Meier-Schätzwert (KM) - auch bekannt unter *product-limit-estimator* - ist die am häufigsten verwendete Methode, um Überlebensfunktionen abzuschätzen. Obwohl sie schon lange vorher verwendet worden war, zeigten Kaplan und Meier erst 1958, daß es sich tatsächlich um eine *maximum likelihood* Abschätzung handelt.

Der KM-Schätzwert  $\hat{S}(t)$  ist ohne Zensurierung oder bei nur rechtsseitiger Zensurierung allein der Beobachtungsanteil mit einer Ereigniszeit größer als  $t$ . Sind beispielsweise bei 40% der Beobachtungen des Ereignisses größer als der Zeitpunkt  $t=10$  (Wochen), so ist  $\hat{S}(10) = 0,40$  (bei rechtsseitiger Zensurierung unter der Voraussetzung, daß  $t \leq c$  (des Zensurzeitpunktes) ist; für  $t > c$  ist  $\hat{S}(t)$  nicht definiert). Allgemein ist der KM-Schätzwert definiert als:

$$S(t) = \prod_{j:t_j \leq t} \left[ 1 - \frac{d_j}{n_j} \right] \quad \text{für } t_1 \leq t \leq t_k \quad (6)$$

wobei zu jeder Zeit  $j$   $n_j$  Personen mit Hazardrisiko (d.h. sie haben das Ereignis noch nicht vor dem Zeitpunkt  $t_j$  erlebt) vorhanden sind und  $d_j$  Personen, die bis zum Zeitpunkt  $t_j$  verstorben sind.

Neben dem KM-Schätzwert (*survival*) erhält man noch die Variable Failure (=1 - *survival*) und den *survival standard error*, der durch die Greenwood-Formel berechnet wird und mit dessen Hilfe man jeweils zu einem bestimmten Zeitpunkt zugehörige Konfidenz-Intervalle abschätzen kann. Für Konfidenzregionen haben *Harris und Albert (1991)* eigene Methoden (und SAS Macros) entwickelt. Mit Hilfe der Funktion Graphics kann die Überlebenskurve stufenförmig ausgedruckt werden, wobei bei jedem Todesfall die Kurve abfällt.

Mit dem Befehl STRATA können nicht nur zwei Gruppen miteinander verglichen werden, sondern auch Gruppeneinteilungen mit mehreren alphanumerischen oder numerischen Variablen getroffen werden, um für diese Gruppen alle nötigen Statistiken zu berechnen (Varianz, Kovarianz, Chi-Quadrat, p-Werte, etc.).

### 6.2.2 Lifetable- Methode

Wenn eine größere Anzahl an Beobachtungen vorliegt und die Ereigniszeiten exakt gemessen wurden, werden durch die KM-Methode aufgrund vieler gleicher Ereigniszeiten sehr lange Tabellen erzeugt, die sowohl für die Präsentation als auch für die Interpretation schlecht geeignet sind. Die Lifetable-Methode (auch unter *actuarial method* oder Sterbetafel-Methode bekannt) löst diese Probleme einerseits durch die Gruppierung der Ereigniszeiten in Intervalle, deren Größe der Bearbeiter selbst bestimmen kann (beispielsweise 24 Stunden- bzw. 1 Tages-Intervalle), und andererseits durch zusätzliche Abschätzungen und Grafiken der Hazard Funktion, die mit der KM-Methode nicht erzeugt werden können. Natürlich kann es auch ein Nachteil sein, wenn die Intervalle willkürlich gewählt werden, aber dieser geringe Verlust an Information ist in der Regel akzeptabel. Die Intervalllängen müssen nicht exakt dieselbe Ausdehnung über die gesamte Beobachtungsdauer haben, vor allem gegen Ende der Beobachtung können die Intervalle länger werden, um anschauliche Ergebnisse zu erzielen.

Die bedingte Wahrscheinlichkeit des Ereigniseintritts (*conditional probability of failure*) ist ein Schätzwert der Wahrscheinlichkeit einer Person, das Ereignis im Intervall zu erfahren, unter der Voraussetzung, daß sie es bis zum Start des Intervalls geschafft hat. Berechnet wird sie mit:

Anzahl der Ereigniseintritte (*number failed*) / effektive Stichprobengröße im Intervall.

Der Lifetable-Schätzwert (LT-Schätzwert) ist die Wahrscheinlichkeit, daß ein Ereignis zu einem späteren Zeitpunkt als dem Intervallbeginn eintritt. Die Berechnung sieht folgendermaßen aus:

$$\hat{S}(t_i) = \prod_{j=1}^{i-1} (1 - q_j) \quad (7)$$

wobei  $t_i$  der Startzeitpunkt des Intervalls  $i$  und  $q_i$  die bedingte Wahrscheinlichkeit des Ereigniseintritts sind. Für  $i=1$  und  $t_i=0$  wird die Überlebenswahrscheinlichkeit (LT-Schätzwert) auf 1 gesetzt.

Der sogenannte PDF-wert ist eine Abschätzung der Wahrscheinlichkeitsdichtefunktion am Intervallmittelpunkt. Die Abschätzung der korrespondierenden Hazard Funktion am Intervallmittelpunkt, die von größtem Interesse ist, wird mit der Formel:

$$h(t_{im}) = \frac{d_i}{b_i(n_i - \frac{w_i}{2} - \frac{d_i}{2})} \quad (8)$$

ermittelt, wobei für das Intervall  $i$ ,  $t_{im}$  der Mittelpunkt,  $d_i$  die Anzahl der Ereignisse,  $b_i$  die Intervallbreite,  $n_i$  die Anzahl der am Intervallbeginn mit Risikofaktor behafteten Beobachtungen (*risk set*) und  $w_i$  die Anzahl der zensierten Beobachtungen sind. Eine bessere Abschätzung des Hazards kann mit  $d_i/T_i$  erreicht werden, wobei  $T_i$  die gesamte *exposure time* im Intervall  $i$  ist. Für jedes Individuum ist die *exposure time* das Ausmaß der tatsächlich beobachteten Zeit im Intervall. Da die genaue *exposure time* selten zu ermitteln ist, stellt der Nenner in der Formel oben eine Annäherung an  $T_i$  dar, der annimmt, daß alle Ereignisse und Zensierungen zum Intervallmittelpunkt stattfinden.

Nebst dem Vergleich verschiedener Gruppen kann man mit *PROC LIFETEST* quantitative Kovariablen überprüfen, ob sie mit der Überlebenszeit korrelieren oder nicht. Dabei wird eine Teststatistik jeder einzelnen Variable, die restlichen ignorierend, erstellt. Diese Statistik ist eine Verallgemeinerung des Log-rank-tests und des Wilcoxon-tests und kann auch als nonparametrischer Koeffiziententest des *accelerated failure time* Modells (siehe *PROC LIFEREG*) interpretiert werden. Der große Vorteil dieser Methode ist sicherlich, daß eine große Menge an Kovariablen überschlagsmäßig begutachtet werden kann, bevor man Regressionsmodelle erstellt. Iterative Kalkulationen werden dafür nicht benötigt, was die Rechenzeit stark herabsetzt. Die bei semi- oder vollparametrischen Methoden inkludierten Tests sind jedoch genauer.

Neben den normalen Survivalkurven können mit *PROC LIFETEST* auch noch zwei weitere Plots erzeugt werden, die hilfreiche Informationen über den Verlauf der Hazardfunktion beinhalten: Log Survival (LS) ( $-\log \hat{S}(t)$  zu  $t$ ) und Log-log-survival (LLS) (also  $\ln(-\ln \hat{S}(t))$  versus  $\log t$ ). Sie werden bei weiteren Modellschätzungen (beispielsweise *Proportional Hazard Models*, siehe Kapitel 6.3) benötigt.

Eine Abschätzung der Log-survivalkurve gibt uns Informationen, ob die Hazardfunktion mit der Zeit steigt, fällt oder konstant bleibt (im Prinzip die erste Ableitung). Um die Hazardkurven „benutzerfreundlicher“ zu machen, gibt es statt der mit der *Lifetable*-Methode in wahllose Intervalle eingeteilten Daten, die Möglichkeit, mit Hilfe von *Smooth Macros*, die ungruppierten Daten zu glätten.

### 6.3 SEMIPARAMETRISCHE MODELLE MIT *PROC PHREG*

Neben der aktuellen Dauer eines Zustandes werden bei Hazardmodellen als wesentliche Determinanten für die Wahrscheinlichkeit eines Ereigniseintritts auch weitere Bestimmungsgrößen berücksichtigt. Dazu gehören neben soziodemografischen Charakteristika, sowohl der Individuen als auch der Haushalte, Öffnungszeiten von Einrichtungen (*authority constraints*), Arbeitszeitstrukturen oder persönliche Prämissen der Verkehrsteilnehmer. Überlebenszeitmodelle, die den Einfluß solcher Determinanten schätzen können, fallen in die Klasse der *Proportional* oder *Parametric Hazard Models*.

#### 6.3.1 Proportional Hazard Models

Bei den *Proportional Hazard Models* (PHM) wird eine grundlegende Hazard-Funktion (*baseline hazard*) angenommen, auf die multiplikatorisch ein Vektor ( $X$ ) wirkt, der alle zusätzlichen erklärenden Variablen beinhaltet. Ein großer Vorteil dieses semiparametrischen Modells ist, daß keine exakte Wahrscheinlichkeitsverteilung angenommen werden muß, womit gewisse Fehlerquellen ausgeschlossen werden und das Modell (im Gegensatz zu vollparametrischen Modellen) robuster ist. Bei den PHM ändern sich somit die Eigenschaften der Hazardfunktionen proportional zur Ausprägung der Kovariablen, unter der Voraussetzung, daß die Hazardrate (*hazard ratio*) konstant bleibt. Bei dem im Kapitel 5.2 beschriebenen Basismodell wird hingegen angenommen, daß alle erklärenden Variablen und Kovariablen den Wert 0 haben und damit die Intervalldauern in keiner Art beeinflussen.

Ein weiterer Vorteil von semiparametrischen Modellen ist die Möglichkeit, zeitabhängige Variablen und Kovariablen, in die Berechnung miteinfließen zu lassen. Als zeitabhängige Variablen werden jene bezeichnet, die im Laufe der Beobachtung (je nach betrachtetem Zeitpunkt) verschiedene Werte annehmen können (z.B. Wettereinflüsse oder Erfahrungszuwächse aufgrund kognitiver Lernprozesse).

Obwohl *PROC PHREG* die jüngste SAS- Prozedur der Survival Analysis ist, wird sie am häufigsten verwendet. Sie basiert auf der Regressionsmethode von Sir David COX (1972), dem *COX-*

*Proportional Hazard Model*. Neben den oben genannten Vorteilen gibt es bei der sehr populären COX Regression noch weitere, nicht so bekannte Vorteile:

- Sie erlaubt eine Art geglätteter Analyse, die effektive Kontrollen für Störvariablen ermöglicht.
- Es können Zeitperioden, in denen kein Risiko eines Ereigniseintritts besteht, festgestellt werden.
- Diskrete und kontinuierliche Meßwerte von Ereigniszeiten können in Einklang gebracht werden.

Trotz all dieser Vorteile ist es manchmal besser, *PROC LIFETEST* oder *PROC LIFEREG* zu verwenden. Die Gründe dafür sind einerseits, daß bei *PROC PHREG* relativ wenige Grafikmöglichkeiten im Programm selbst eingebaut sind und andererseits, daß sogenannte *heterogenety effects* nicht berücksichtigt werden können. In Cox-Modellen wird implizit von der These ausgegangen, daß aufgrund der Proportionalität des Modells Individuen oder Prozesse mit identischen Kovariablen auch identische Hazardraten aufweisen müssen. Werden keine Kovariablen berücksichtigt, existiert für die gesamte Stichprobe eine einzige Hazardrate. Dies ist jedoch offensichtlich eine unangemessene Voraussetzung, da sich Individuen und ihre Umgebung in vielerlei Hinsicht unterscheiden und die gemessenen bzw. gewählten Kovariablen kaum das gesamte Spektrum der Heterogenität der Stichprobe abdecken können (Schönfelder, 2000, vgl. Allison, 1995, 233ff.) Berücksichtigt man diese Heterogenitätseffekte jedoch nicht, können die Schätzergebnisse von Modellen verzerrt werden. Es wird jedoch in dieser Ausarbeitung, wie auch in vielen anderen ähnlichen Studien, auf eine Integration der Heterogenität dezidiert verzichtet, da man über die Unterschiede zwischen den Individuen bzw. deren Rahmenbedingungen erneut Verteilungsannahmen treffen müßte, die wiederum gewissen Unsicherheiten unterliegen. Im Cox-Ansatz selbst bleibt genügend Spielraum für Zufallseffekte in den Zusammenhängen zwischen den Schätzwerten und den empirischen Dauerdaten bestehen (vgl. Schönfelder, 2000).

Normalerweise wird das PHM grundlegend, also ohne Einbezug von zeitabhängigen Kovariablen oder nonproportionalen Hazards, angegeben mit:

$$h_i(t) = \lambda_0(t) \exp\{\beta_1 x_{i1} + \dots + \beta_k x_{ik}\} \quad (9)$$

Der erste Teil der Gleichung ist die grundlegende Hazard Funktion, deren einzige Einschränkung darin besteht, daß sie nicht negativ sein darf. Der zweite Teil ist eine lineare Funktion eines Satzes von  $k$  fixierten Kovariablen, der im Exponenten steht. Nimmt man von beiden Seiten den Logarithmus, so ergibt sich die Formel:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad , \quad \text{wobei } \alpha(t) = \log \lambda_0(t) \text{ ist.} \quad (10)$$

Einzelne Modelle unterscheiden sich nur durch ihren *baseline hazard* (pdf und  $\alpha(t)$ ). Bei

$\alpha(t) = \alpha$  bekommen wir das Exponential Modell (konstante Hazardrate) oder bei

$\alpha(t) = \alpha t$  das Gompertz Modell (log-logistische Verteilung) oder bei

$\alpha(t) = \alpha \log t$  das Weibull Modell (monoton steigend oder fallend).

Wie wir aber in weiterer Folge sehen werden, sind solche Spezifikationen bei der Coxregression gar nicht nötig, es wäre für  $\alpha(t)$  sogar eine Stufenfunktion denkbar.

Der Name *Proportional Hazard Models* (PHM) kommt von der Tatsache, daß der Hazard für ein Individuum oder einen Prozeß ein fixer Anteil (Proportion) des Hazards für jede andere Person bzw. die gesamte Stichprobe ist. Um dies anschaulich zu machen, berechnet man den ratio der Hazards für zwei Personen  $i$  und  $j$ :  $h_i(t) / h_j(t) = \exp\{\beta_1 (x_{i1} - x_{j1}) + \dots + \beta_k (x_{ik} - x_{jk})\}$ , in der sich  $\lambda_0(t)$  kürzt und der Ratio zeitunabhängig ist bzw. sogar über die Zeit konstant bleibt.

### 6.3.2 (Maximum) Partial Likelihood

Die Popularität des Cox-Modelles läßt sich damit erklären, daß mit Hilfe des Prinzips der Maximierung der *partial likelihood* die  $\beta$ -Koeffizienten des PHM abgeschätzt werden können, ohne daß vorher explizite Verteilungsannahmen zur *baseline hazard* getroffen werden müssen. Diese Prinzip der *Maximum Partial Likelihood* soll hier nur kurz angerissen werden.

Die Likelihood Funktion für das PHM der obigen Gleichung kann in zwei Teile gegliedert werden: Ein Teil, der sowohl von  $\lambda_0(t)$ , als auch von  $\beta$  (dem Koeffizientenvektor) abhängig ist und ein zweiter Teil, der nur von  $\beta$  abhängt. *Partial Likelihood* läßt nun den ersten Teil außer Acht und behandelt den zweiten als normale Likelihood-Funktion (siehe Kapitel 6.4). Es werden so Schätzwerte für  $\beta$  gefunden, die die *partial likelihood* Funktion maximieren. Natürlich sind diese Schätzwerte nicht vollständig genau, da im ersten außer Acht gelassenen Teil Anteile von  $\beta$  enthalten sind, jedoch sind auch bei geringfügig höheren Standardfehlern der *partial likelihood* die Unterschiede zur Likelihood Funktion nicht übermäßig. Mehr noch, die Schätzwerte sind robuster, da sie von der aktuellen Kurve bzw. Form der Hazard Funktion unabhängig sind und immer noch konsistent und asymptotisch normal sind. Ein weiteres interessantes Detail der *partial likelihood* Abschätzung ist die Unabhängigkeit von Zahlenwerten. Es zählt nur die Reihung der Ereigniszeiten, nicht aber der genaue Wert. Eine monotone Transformation der Ereigniszeitpunkte (Addition einer Konstante, Multiplikation mit einer Konstante, Logarithmus, etc.) würde keinerlei Änderung der Koeffizienten bewirken.

Die allgemeine Form der *partial likelihood* für Daten mit fixen Kovariablen eines PHM sieht dann folgendermaßen aus:

$$PL(\beta) = \prod_{i=1}^n \left[ \frac{e^{\beta x_i}}{\sum_{j=1}^n y_{ij} e^{\beta x_j}} \right]^{\delta_i} \quad (11)$$

wobei  $y_{ij}$  die Werte 0 oder 1 annehmen kann, je nachdem ob derjenige das Ereignis schon erfahren hat.

$PL(\beta)$  = *partial likelihood* von  $\beta$

$\beta$  = Vektor der Schätzparameter

$n$  = Anzahl aller Dauern

$X$  = Vektor der Kovariablen

Nach der Logarithmierung auf beiden Seiten kann man wieder, wie bei der maximum Likelihood Abschätzung, mit Hilfe des Newton-Raphson-Algorithmus diejenigen Koeffizienten  $\beta$  schätzen, die die Funktion maximieren.

### 6.3.3 Besonderheiten

#### 6.3.3.1 **TIED DATA - Ereignisse zum selben Zeitpunkt**

Die oben erwähnte allgemeine Formel für die partial likelihood gilt uneingeschränkt jedoch nur dann, wenn keine zwei Ereignisse zum selben Zeitpunkt stattfinden. Daher braucht man alternative Lösungsformen, wenn dies doch geschieht. Die Breslow-Annäherung (bei nur wenigen gleichen Ereigniszeiten) oder die Annäherung von Efron (1977) sind solche Möglichkeiten.

In *PROC PHREG* werden zwei verschiedene Lösungsmöglichkeiten angeboten, mit gleichen Ergebnissen umzugehen: Die *EXACT*-Methode nimmt dabei eine wahre, aber nicht bekannte Reihung von gleichen Ereigniszeitpunkten vor (Zeit ist kontinuierlich), die *DISCRETE*-Methode nimmt alle Ereigniszeitpunkte wirklich so an, als wären sie exakt zum selben Zeitpunkt passiert. Die Terminologie der beiden Methoden ist jedoch etwas irreführend, da beide, wenn auch auf unterschiedlichen Wegen, exakte Wahrscheinlichkeiten (Likelihoods) berechnen.

### 6.3.3.2 **TIME DEPENDENT COVARIATES - Zeitabhängige Kovariablen**

Zeitabhängige Kovariablen sind jene, die ihren Wert im Verlauf der Beobachtung verändern können. Um das COX-Modell auch für zeitabhängige Kovariablen einsatzfähig zu machen, wird nur ein zusätzlicher Term mit einer zeitabhängigen Variable  $x_{i2}(t)$  eingeführt. Die computermäßige Auswertung ist jedoch um einiges schwieriger und fehleranfälliger bzw. können Fehler nicht so ohne weiteres sofort entdeckt werden. Für ein Modell mit einer fixen und einer zeitabhängigen Kovariablen gilt:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2}(t) \quad (12)$$

Das heißt: Der Hazard zum Zeitpunkt  $t$  hängt sowohl von der Variable  $x_1$  ab, als auch von der Variable  $x_2$  zum Zeitpunkt  $t$ , wobei  $x_2(t)$  auch Informationen vor dem Zeitpunkt  $t$  beinhalten kann.

Andere Zeitursprungspunkte (*Time Origins*), können auch als Kovariablen, manchmal auch als zeitabhängige Kovariablen angegeben werden. Der Gebrauch von zeitverschobenen Kovariablen ist öfters essentiell, um Verteilungen von kausalen Ordnungen oder Systemen zu erhalten.

Die Formel lautet dann:

$$\log h(t) = \alpha(t) + \beta_1 x_1 + \beta_2 x_2(t) + \beta_3 x_3(t), \quad (13)$$

wobei  $x_3$  die zeitabhängige Variable ist.  $x_3(t) = x_3(0) + t$  und  $x_3(0) =$  Variable am Zeitursprung ergibt:

$$\log h(t) = \alpha^*(t) + \beta_1 x_1 + \beta_2 x_2(t) + \beta_3 x_3(0) \quad \text{mit } \alpha^*(t) = \alpha(t) + \beta_3 t. \quad (14)$$

Mit Hilfe des Tricks, ein Modell mit zeitabhängiger Altersvariable in eines mit fixen bzw. zeitkonstanten Altersvariablen zu konvertieren, fällt diese zufällige Funktion der Zeit aus der Abschätzung heraus und man erhält wieder exakt dieselben Ergebnisse. Dieser Trick funktioniert jedoch nur, wenn der Logarithmus des Hazard eine lineare Funktion des alternativen Zeitursprungs ist.

Werden gewisse Informationen nur in regelmäßigen Abständen (beispielsweise nur einmal monatlich gesammelt), das Ereignis aber zu einem genau definierten Zeitpunkt stattfindet, gibt es diverse Ad-hoc-Methoden, die Kovariablenwerte einem bestimmten Zeitpunkt zuweisen können. Kovariablen, die in einigen Intervallen nicht definiert sind, können auch in die Studien mit einbezogen werden, indem sie als zeitabhängige Variablen mit dem Wert 0 vor einem bestimmten Ereignis behandelt werden.

### 6.3.3.3 COX Modelle mit *Nonproportional Hazards*

Die eben behandelten Modelle mit zeitabhängigen Variablen sind streng genommen keine PHM mehr, da diese zeitabhängigen Kovariablen zu unterschiedlichen Zeitpunkten andere Ergebnisse für die Individuen ergeben und so ihre Hazardsratios nicht konstant bleiben. Um herauszufinden, ob gewisse Variablen des PHM - obwohl nicht konkret als solche definiert - zeitabhängig sind, kann man eine zusätzliche zeitabhängige Variable ins Modell nehmen, die die Wechselwirkungen der zu überprüfenden Variable und der Zeit repräsentiert. Hat diese Wechselwirkungsvariable keinerlei statistische Signifikanz, so weiß man, daß auch die PH-Annahme für diese Variable nicht gestört ist. Bzw. andererseits ist bei einer signifikanten Wechselwirkungsvariable die Nonproportionalität bewiesen. Eine weitere Methode, Nonproportionalität festzustellen, ist die *Stratification*, die besonders bei Kovariablen, die sowohl kategorisch, als auch von keinem direkten Interesse sind, hilfreich ist.

Wenn wir keine definierten zeitabhängigen Kovariablen haben, wissen wir jedoch immer noch nicht, ob das PHM so stimmt. Es sind bei allen Schätzungen (auch bei Regressionsmodellen, etc.) kleinere Ungenauigkeiten enthalten. Besser sollte man deswegen der Frage nachgehen: Sind alle relevanten Kovariablen im Modell inkludiert und berücksichtigt?

## 6.4 VOLLPARAMETRISCHE MODELLE MIT *PROC LIFEREG*

Eine weitere Spezifizierung der Charaktere von Zeitdauerdaten, sowohl in theoretischer als auch in konzeptioneller Hinsicht, die von den bisher dargestellten minimalen Annahmen abrücken, lassen die Erstellung von vollparametrischen Modellen sinnvoll erscheinen. Die oben beschriebenen semiparametrischen Modelle bieten zwar eine robustere Modellschätzung und gewisse andere Vorteile (beispielsweise die Möglichkeit, zeitabhängige Variablen zu integrieren,  $\beta$ -Koeffizienten abzuschätzen, ohne vorher genaue Annahmen über die Survival-Verteilung treffen zu müssen, etc.), der Einsatz von vollparametrischen Modellen ist dennoch öfters von Vorteil.

Die im Statistikprogramm SAS dafür verwendete Methode ist *PROC LIFEREG*, die Abschätzungen von parametrischen Regressionsmodellen mit zensierten Überlebensdaten (*survival data*) unter Verwendung der *maximum likelihood* Methode liefert. Die Vorteile gegenüber den semiparametrischen Modellen sind folgende:

- Es können neben rechtszensierten auch links- und intervall-zensierte Daten verwendet werden.
- Es können einzelne Hypothesen über den Verlauf der Hazardfunktion überprüft werden, *PROC PHREG* liefert dagegen nur nonparametrische Abschätzungen der Überlebensfunktion, die oft schwierig zu interpretieren sind. Dezidierte Hypothesentests sind bei *PROC PHREG* nicht oder nur sehr schwer möglich.
- Kennt man den Verlauf der Survival-Verteilung (beispielsweise aus einer Abschätzung mit *PROC LIFETEST*), liefert *PROC LIFEREG* genauere Ergebnisse mit kleineren Standardfehlern bzw. -abweichungen.
- *PROC LIFEREG* liefert automatisch einen Satz von Dummy- oder Indikatorvariablen, um kategoriale Variablen mit multiplen Werten darzustellen.

Die größte Einschränkung bei vollparametrischen Modellen ist jedoch sicherlich die Tatsache, daß keine zeitabhängigen Kovariablen eingebaut werden können.

Ein großer Vorteil von *PROC LIFEREG* gegenüber vollparametrischen Modellen in anderen Statistikprogrammen ist die große Bandbreite an Survival-Verteilungen. Die Verwendung einer verallgemeinerten Gamma-Verteilung macht Likelihood-Ratio-Parform-Tests für viele andere Wahrscheinlichkeitsverteilungen möglich.

#### 6.4.1 Accelerated Failure Time Model

Die Art von Regressionsmodellen, die mit *PROC LIFEREG* abgeschätzt wird, ist unter *Accelerated Failure Time Model* (AFT-Modell) bekannt. In der allgemeinen Form wird dabei ein Vergleich zwischen den Survivor Funktion zweier Individuen beschrieben:

$$S_i(t) = S_j(\Phi_{ij}t) \quad (15)$$

für alle  $t$ , wobei  $S_i(t)$  die Survivor Funktion für Individuum  $i$ ,  $S_j(t)$  die für das andere Individuum  $j$  und  $\Phi_{ij}$  eine für das Paar  $(i,j)$  spezielle Konstante sind.

*PROC LIFEREG* schätzt einen speziellen Fall dieses Modells, das der Form des linearen Regressionsmodells sehr ähnlich ist:

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \sigma \varepsilon_i, \quad (16)$$

wobei  $T_i$  die Zufallsvariable bzw. der Ereigniszeitpunkt,  $\varepsilon_i$  ein zufälliger Störfaktor und  $\beta_0, \dots, \beta_k$  und  $\sigma$  die abzuschätzenden Parameter sind.

Die einzigen Unterschiede dieser Gleichung zu der normalen linearen Regressionsformel sind das  $\sigma$ , das als Konstante auch weggelassen werden kann bzw. zur Unterscheidung verschiedener Individuen dient und die Logarithmus-Form der abhängigen Variable  $T_i$ , damit sie unabhängig von den Werten für  $x$  und  $\beta$  auf jeden Fall positiv bleibt.

Sind keine zensierten Daten vorhanden, kann das Modell sofort mit Hilfe der kleinsten Quadrate-Methode (*ordinary least square*, OLS) abgeschätzt werden. Dabei wird der Logarithmus der Ereigniszeit als abhängige Variable in der REG-Prozedur verwendet. So erhält man die besten linearen und unvoreingenommenen Abschätzungen der  $\beta$ -Koeffizienten, unabhängig vom Verlauf der  $\varepsilon$ -Verteilung. Ist diese normalverteilt, ergeben sich die OLS-Abschätzungen zusätzlich als *maximum likelihood*-Abschätzungen und haben die minimalsten Varianzen unter allen Abschätzern, egal ob linear oder nichtlinear.

Prinzipiell muß man bei vollständig parametrischen Modellen, die mit *PROC LIFEREG* erzeugt werden, die Verteilung der  $\varepsilon$  wissen, um die Verteilung von  $T$  abschätzen zu können. *PROC LIFEREG* erlaubt neben der Normalverteilung noch weitere Verteilungsmöglichkeiten für  $\varepsilon$ :

- a) Extremwerte-verteilung mit 2 Parametern,
- b) Extremwerte-verteilung mit 1 Parameter,
- c) Log-gamma-Verteilung und
- d) logistische Verteilung,

die jeweils eine korrespondierende Verteilung für  $T$  haben: Weibull (a), Exponential (b), Gamma (c) bzw. Log-logistic (d).

Der Grund, neben der Normalverteilung noch andere Verteilungen zuzulassen bzw. zu untersuchen, sind die verschiedenen Auswirkungen auf die Hazardfunktionen, was wiederum zu grundlegend unterschiedlichen Interpretationen führen kann.

Es werden folgend kurz die einzelnen Verteilungen bzw. AFT-Modelle beschrieben:

#### 6.4.1.1 Exponential-Modell

Das einfachste Modell, das durch *PROC LIFEREG* abgeschätzt werden kann, zeichnet sich durch eine standardisierte Extremwertverteilung von  $\varepsilon$ , die auch unter Gumble- oder doppelte Exponential-

verteilung bekannt ist, und durch  $\sigma = 1$  aus.  $T$  selbst ist dann ebenfalls exponentiell verteilt bzw. die dazugehörige Hazardfunktion konstant. Die Formel des exponentiellen Regressionsmodells:

$$\log h(t) = \beta_0^\circ + \beta_1^\circ x_1 + \dots + \beta_k^\circ x_k \quad (17)$$

unterscheidet sich von der Formel (16) ( $\log T_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \sigma \varepsilon_i$ ) nur durch die Notation von  $\beta_j = -\beta_j^\circ$ . Negatives Vorzeichen deswegen, weil ein hoher Hazard ein schnelles Ereignis und kurze Überlebenszeiten voraussagt.

#### 6.4.1.2 Weibull-Modell

Bei diesem Modell bleibt die Annahme einer Standard-Extremwertverteilung von  $\varepsilon$  aufrecht, jedoch muß  $\sigma$  nicht gleich 1 sein. Durch diesen kleinen Unterschied ergeben sich gewaltige Differenzen zum Exponential-Modell.  $T$  hat dadurch abhängig von den Kovariablen eine Weibullverteilung.

Diese Weibullverteilung was aus zwei Gründen lange Zeit das beliebteste parametrische Modell.

Einerseits hat die Überlebensfunktion eine relativ leicht benutzbare mathematische Form:

$$S_i(t) = \exp \left\{ - \left[ t_i e^{-\beta x_i} \right]^{1/\sigma} \right\} \quad (18)$$

und zweitens ist das Weibull-Modell neben dem AFT auch ein *Proportional Hazard Model*. Dies bedeutet, daß seine Koeffizienten bei richtiger Transformation als relative Hazardraten interpretiert werden können. Die Umwandlung der log-hazard-Form des Weibull-Modells  $\log h(t) = \alpha \log t + \beta_0^\circ + \beta_1^\circ x_1 + \dots + \beta_k^\circ x_k$  und dem log-survival-Zeitmodell  $\log T = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \sigma \varepsilon$  kann mit  $\beta_j^\circ = -\beta_j / \sigma$  und  $\alpha = 1 / \sigma$  leicht durchgeführt werden. Die Umwandlung von Standardfehler und Konfidenzintervallen ist jedoch nur mit komplizierten Formeln möglich (Collett, 1994, p.282).

#### 6.4.1.3 Log-Normal-Modell

Im Gegensatz zum Weibull-Modell hat das Log-Normal-Modell keine monoton verlaufende Hazardfunktion. Bei  $t = 0$  ist der Hazard 0, hat dann irgendwo eine Spitze und nähert sich bei  $t \rightarrow \infty$  wieder 0. Dieses Modell ist kein *Proportional Hazard Model* und sein Hazard kann auch nicht genauer dargestellt werden. Es kann jedoch als Regressionsmodell mit dem Hazard als abhängige Variable angegeben werden:  $\log h(t) = \log h_0(t e^{-\beta x}) - \beta x$ , wobei  $h_0$  die Hazard Funktion für ein Individuum bei  $x = 0$  angibt. Diese Formel gilt in weiterer Folge auch für die Modelle 6.4.1.4 und 6.4.1.5 (Log-Logistic und Gamma Modelle), wobei  $h_0$  jeweils andere Formen annimmt.

Für wiederholbare Ereignisse ist die umgekehrte U-Form des Log-normal Hazard oftmals gut geeignet. Wenn ein Ereignis, wie z.B. Kirchenbesuch oder sportliche Aktivität etc. gerade eingetreten ist, ist es relativ unwahrscheinlich, daß es sofort wieder auftritt, es muß erst ein Bedarf aufgebaut werden.

#### 6.4.1.4 Log-Logistic Modell

Hierbei hat eine logistische (logit) Verteilung der Form

$$f(\varepsilon) = \frac{\varepsilon^\varepsilon}{(1 + \varepsilon^\varepsilon)^2} \quad (19)$$

, die mit einem Mittel von 0 von der Form her der Normalverteilung ähnelt.

Diese logistische Verteilung des logistischen Regressionsmodells kann (a) durch eine Abschätzung eines linearen Modells mit Hilfe eines logistisch verteilten Fehlerterms oder (b) durch Dichotomisierung der abhängigen Variable erreicht werden. Die Verteilung des Logarithmus von T ist dann ebenfalls logistisch verteilt, jedoch ohne einem Mittelwert von 0. T ist also log-logistisch verteilt mit einer Hazardfunktion der Form:

$$h(t) = \frac{\lambda \gamma (\lambda t)^{\gamma-1}}{1 + (\lambda t)^\gamma} \quad (20)$$

wobei  $\gamma = 1 / \sigma$  und  $\lambda = \exp\{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)\}$ .

Die Überlebensfunktion hat hier das relativ einfache Aussehen:  $S(t) = 1 / (1 + (\lambda t)^\gamma)$ .

Mit dem Einsetzen der Terme für  $\gamma$  und  $\lambda$  wird mit  $\beta_i^\circ = \beta_i / \sigma$  diese Form erreicht:

$$\log \left[ \frac{S(t)}{1 - S(t)} \right] = \beta_0^\circ + \beta_1^\circ x_1 + \dots + \beta_k^\circ x_k - \gamma \log t \quad (21)$$

Aus dem Grund, daß S(t) die Wahrscheinlichkeit des Überlebens bis zum Zeitpunkt t ist und S(t) / 1 - S(t) das Gegenteil ist, spricht man bei Log-logistic Modellen auch von einem Vertreter der proportional odds Modelle, wobei das Log-logistic Modell das einzige sowohl *proportional odds* Modell als auch AFT Modell ist.

#### 6.4.1.5 Gamma Modell

In der Survival Analysis Literatur unterscheidet man zwischen standardisierten 2-Parameter Modellen und dem in *PROC LIFEREG* verwendeten generalisierten 3-Parameter Modell. Durch diesen zusätzli-

chen Parameter (im Gegensatz zu den Modellen 6.4.1.1 bis 6.4.1.4) kann die Hazard Funktion eine ganze Reihe von Formen annehmen. Die vorher behandelten Modelle (mit Ausnahme des Log-logistic Modells) sind jeweils Spezialfälle des Gamma Modells. Aufgrund von zwei Tatsachen ist es jedoch nicht die beste Möglichkeit, immer das generalisierte Gamma Modell zu verwenden bzw. auszuwählen: Erstens ist die Formel für die Hazard Funktion, die die Gammafunktion und die unvollständige Gammafunktion beinhaltet, wesentlich komplizierter als die der Submodelle. Zweitens ist die Rechenzeit für das generalisierte Gammamodell vor allem für größere Datensätze um ein vielfaches länger als die der vorher beschriebenen Modelle. Das standardisierte Gamma Modell kann mit *PROC LIFEREG* nicht direkt abgeschätzt werden. Am einfachsten geht es mit der Gleichsetzung von *SCALE* = *SHAPE* aus dem generalisierten Gamma Modell, wobei der Wert hierfür abgeschätzt werden muß. Um das standardisierte Gamma Modell interpretieren zu können, gelten die Formeln:

$$f(t) = \frac{\lambda (\lambda t)^{K-1} e^{-\lambda t}}{\Gamma(K)} \quad (22)$$

mit  $k = 1 / \delta^2$  (, dem *Shape* Parameter in *PROC LIFEREG*),

$\lambda = \exp\{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)\}$  und

$\Gamma(K) = (K - 1)!$  , wenn  $K$  ein Integer ist.

Die Survivor und die Hazard Funktion der standardisierten Gamma Verteilung sind jedoch sehr unhandlich, da sie die unvollständige Gamma Funktion beinhaltet, die als Integral ausgedrückt wird.

## 6.4.2 MAXIMUM LIKELIHOOD Abschätzung (ML)

Alle Modelle in *PROC LIFEREG* werden mit der Methode von ML abgeschätzt. Aus zwei Gründen hat die ML als genereller Versuch der Abschätzung in viele Anwendungsgebieten Einzug gefunden:

Erstens produziert ML für große Datensätze Schätzwerte mit guten Eigenschaften wie Konsistenz, asymptotische Effizienz und asymptotische Normalität. Konsistenz heißt, daß sich die Abschätzungen, je größer die Stichprobe wird, weiter an Wahrscheinlichkeiten der wahren Werte annähern, vorausgesetzt sie sind weitgehend unabhängig. Asymptotische Effizienz bedeutet, daß bei großen Datensätzen nur Standardfehler entstehen, die so klein wie bei jeder anderen Schätzmethode sind. Asymptotische Normalität heißt, daß die Verteilungen der Schätzung normal verteilt sind bzw. Normal- und Chi-Quadrat-Verteilung angenommen werden kann, um Konfidenzintervalle und P-Werte zu berechnen. Für kleinere Datensätze gibt es keine genauer definierten Eigenschaften. Deshalb kann aber noch

keineswegs behauptet werden, daß sich die ML-Methode nicht dafür eignen sollte. Vielmehr wird sie sowohl bei großen als auch bei kleinen Datensätzen verwendet.

Der zweite Grund für die Popularität von ML ist das Fehlen geeigneter Alternativen, vor allem bei zensierten Datensätzen. Beispielsweise ergibt die Kleinste-Quadrate-Methode mit gewissen Voraussetzungen oft relativ hohe Standardfehler und hat zudem zu wenig theoretischen Hintergrund, Hypothesen und Konfidenzintervalle fundiert zu konstruieren.

Das **Grundprinzip der ML-Methode** besteht darin, jene Werte als Schätzwerte herauszufinden, die die Wahrscheinlichkeit maximieren, das zu sehen, was tatsächlich beobachtet wurde. Dazu gibt es 2 Schritte: (1) die Angabe eines Ausdrucks für die Wahrscheinlichkeit als Funktion der unbekannt Parameter und (2) jene Werte für die unbekannt Parameter herauszufinden, die die Werte dieses Wahrscheinlichkeitsausdrucks möglichst groß machen.

Der erste Schritt, die Konstruktion der Likelihood Funktion, bedarf der Aufstellung eines Modells, das darauf hinausläuft, eine Wahrscheinlichkeitsverteilung für die abhängige Variable zu erstellen und eine Form zu suchen, die Verteilungsparameter mit den Werten der Kovariablen in Einklang bringt. Die Maximierung, der zweite Schritt, wird mit Hilfe von iterativen numerischen Methoden, die meist computerdominiert sind, erreicht.

Grundlegende mathematische Züge der MAXIMUM LIKELIHOOD Abschätzung:

Für  $n$  unabhängige Personen haben wir jeweils  $t_i$ ,  $\delta_i$  und  $x_i$  vorliegen, also Zeitpunkt  $t_i$  des Ereignisses oder der Zensur,  $\delta_i$  als binomiale Variable, ob zensiert wurde oder nicht und  $x_i$  als Vektor von Kovariablen-werten.

Die Wahrscheinlichkeit (*likelihood*) des gesamten Datensatzes entspricht dem Produkt aller Einzelwahrscheinlichkeiten, die mit Hilfe der pdf (Wahrscheinlichkeitsdichtefunktion) als

$$L = \prod_{i=1}^n f_i(t_i) \quad \text{für alle nichtzensierten bzw. als} \quad (23)$$

$$L = \prod_{i=1}^r f_i(t_i) \prod_{i=r+1}^n S_i(t_i) \quad \text{unter Einbezug der zensierten Fälle} \quad (24)$$

angegeben werden kann.

Nachdem  $\delta_i$  nur 0 oder 1 annehmen kann, kann die Wahrscheinlichkeit mit

$$L = \prod_{i=1}^r [f_i(t_i)]^{\delta_i} [S_i(t_i)]^{1-\delta_i} \quad \text{angegeben werden.} \quad (25)$$

Hat man ein bestimmtes AFT Modell ausgewählt, so kann man die Wahrscheinlichkeitsdichtefunktion und die Überlebensfunktion entsprechend einsetzen.

Beim einfachsten Modell, dem Exponentialmodell, ergibt sich dabei mit  $f_i(t_i) = \lambda_i e^{-\lambda_i t_i}$  und  $S_i(t_i) = e^{-\lambda_i t_i}$  nach Umformung der Ausdruck:

$$L = \prod_{i=1}^n \lambda_i^{\delta_i} e^{-\lambda_i t_i} \quad (26)$$

bzw. durch Logarithmieren auf beiden Seiten, das gewisse Vorteile bringt:

$$\log L = -\beta \sum_{i=1}^n \delta_i x_i - \sum_{i=1}^n t_i e^{-\beta x_i} \quad \text{mit } \beta \text{ als Koeffizientenvektor.} \quad (27)$$

Im zweiten Schritt sollen Werte für  $\beta$  gefunden werden, die diesen vorher aufgestellten Wahrscheinlichkeitsausdruck maximal machen. Hierbei wird der bekannte Ansatz verwendet, bei der die erste Ableitung gleich Null gesetzt wird. Das Problem der nichtlinearen Terme von  $\beta$  kann nur iterativ gelöst werden, wobei bei *PROC LIFEREG* der Newton-Raphson-Algorithmus (Gradient und Hessian) verwendet wird. Mit Hilfe eines Startwertes  $\beta_0$  werden alle  $\beta$  auf 1 % genau berechnet und zudem eine Kovarianzenmatrix angegeben.

Fehlschlagen kann dieser Algorithmus eigentlich nur in den seltensten Fällen, wenn extrem kleine Stichproben vorliegen, fast alle Ereignisse zensiert oder zu viele Parameter geschätzt werden müssen. Gänzlich schlägt die Methode fehl, wenn alle Ereignisse zum selben Zeitpunkt zensiert werden bzw. liegt dann Nonkonvergenz vor.

### 6.4.3 Besonderheiten

#### 6.4.3.1 Hypothesen-Tests

Im Gegensatz zum semiparametrischen Modell können hier Nullhypotesentests durchgeführt werden. Normalerweise werden Chi-Quadrat-tests angewendet, um die Hypothese, daß jeder Koeffizient 0 ist, zu überprüfen. Meist sind dies Wald-Tests oder Lagrange-Chi-Quadrat-Statistiken (bei Exponentialmodellen) oder Likelihood-Ratio-Tests (die ähnlich dem F-Test bei linearen Regressionsmodellen sind). Meist sind nur geringe Unterschiede in den Ergebnissen zu erkennen, obwohl die Methoden jeweils anders sind.

#### 6.4.3.2 Tests der Paßform (*Goodness-of-fit-Tests*)

Meist ergeben sich bei verschiedenen AFT-Modellen unterschiedliche Verteilungen für T, aber ähnliche Koeffizientenabschätzungen und p-Werte. Deswegen braucht man gewisse Kriterien, zwischen den Modellen mit ihren unterschiedlichen Hazardfunktionskurvenformen entscheiden zu können.

Likelihood-Ratio-Statistiken helfen uns bei der Unterscheidung zwischen verschiedenen miteinander vernetzten Modellen. Für jedes Modell werden die *log likelihoods* berechnet und mit dem des generalisierten Gamma Modells verglichen. Unter der Voraussetzung, daß das generalisierte Gamma Modell die Charakteristik der Daten einigermaßen genau trifft, was aber aufgrund der großen Bandbreite an möglichen Verteilungen meist gut paßt, wird dann die Güte der Paßform jedes einzelnen Modells festgestellt.

Auch grafische Methoden zur Überprüfung der Paßform gibt es: Mit Hilfe von *PROC LIFETEST* (PLOTS = (LS,LLS)) kann man feststellen, ob die Verteilung der Ereigniszeiten einer exponentiellen Verteilung oder einer Weibullverteilung nahe kommt. Genauso kann man mit *PROC LIFETEST* die log-logistische und die log-normale Verteilungen evaluieren. Wobei hierbei die Schwierigkeit besteht, daß all diese Grafiken auf der Annahme basieren, daß die Stichprobe aus einer homogenen Bevölkerung gezogen wurden und keine Kovariablen mit der Ereigniszeit verknüpft sind. Das heißt, die Modelle sehen oft gut aus, solange keine Kovariablen mit einbezogen werden. Eine Möglichkeit, dies zu korrigieren ist, Grafiken zu erstellen, die auf den *Residuals* der Regressionsmodelle beruhen, wie beispielsweise die *Cox-Snell-Residuals*.

### 6.4.3.3 **Integration von links- und intervallzensierten Daten (*Left and Intervall Censoring*)**

Wie anfangs schon erwähnt, hat man bei vollparametrischen Modellen den großen Vorteil, problemlos zensierte Daten behandeln zu können, da für Intervalle von  $a$  nach  $b$  gilt:  $S_i(a) - S_i(b)$ , wenn linkszensiert  $a = 0$ , rechtszensiert  $b = \infty$  und bei Intervallzensierung  $a =$  untere Grenze und  $b =$  obere Grenze.

### 6.4.3.4 **Das schrittweise Exponential Modell (*piecewise exponential model*)**

Alle bisher behandelten AFT- Modelle setzen einen ruhigen Verlauf und einen zeitlich einfache Funktion des Hazards voraus. Das COX-Modell (mit *PROC PHREG* generiert) ist hierbei weit weniger restriktiv, aber es entbehrt der Möglichkeit, Hypothesen über den Verlauf der Hazard Funktion zu testen. Mit Hilfe des *piecewise exponential model* kann die Flexibilität des COX Modells ohne Verlust der Hypothesentestmöglichkeit und dem Einbezug von zeitabhängigen Kovariablen erhalten bleiben.

### 6.4.3.5 **Generating Predictions und Hazard Funktionen**

Nachdem man ein Modell gefunden und angepaßt hat, ist es öfters angenehm, Überlebenszeiten bzw. Ereigniszeitpunkte voraussagen, Wahrscheinlichkeiten für gewisse Zeitpunkte anzugeben oder die Hazard Funktion in Relation zur Zeit darzustellen. Dies geschieht mit gewissen SAS-Befehlen mit von Allison geschriebenen Macros (*PREDICT* bzw. *LIFEHAZ*).

## 7 STUDIEN MIT HAZARDMODELLEN

### 7.1 EINLEITUNG

*Duration* bzw. Zeitdauerdaten, die zeitliche Angaben beinhalten oder bei denen zeitliche Abfolgen bis zum Auftreten eines bestimmten Ereignisses von Bedeutung sind, werden seit Beginn der 80er Jahre auch in einigen Bereichen der Verkehrsverhaltensforschung vermehrt mit Hilfe von Hazardmodellen analysiert. Diese Modelle aus dem Bereich der *Survival Analysis* bieten im Vergleich zu klassischen linearen Regressionsmodellen oder ähnlichen stochastischen Analysemethoden, mit denen vorher Verteilungen oder der Einfluß von Determinanten eines bestimmten Phänomens untersucht worden sind, ein sehr breit gefächertes Feld von Überprüfungsmethoden, um als Instrument bei der Auswertung von Datensätzen mit Zeitdauern geeignet eingesetzt werden zu können. Eine detaillierte Betrachtung aller Vorteile von Hazardmodellen ist im Kapitel 5 enthalten.

In der Fachliteratur finden sich vor allem aus dem letzten Jahrzehnt einige Studien, die mit Hilfe von Hazardmodellen Zeitdauerdaten analysiert haben. Es ist jedoch eine Auswahl getroffen worden und es werden im Folgenden nur ein paar vorgestellt und vor allem jene, die ein ähnliches Ziel wie diese Arbeit verfolgen, näher behandelt und kurz diskutiert.

### 7.2 WEITERE STUDIEN MIT DEN DATEN DER UTHS

Mit den Daten der *Uppsala Travel Household Survey* (kurz UTHS), die auch in dieser Arbeit als Datengrundlage genommen worden sind, wurden schon einige Studien gemacht. Einige Arbeiten, die hauptsächlich von Susan und Perry Hanson verfaßt wurden, sind schon im Kapitel 3 behandelt. In diesen Studien wurden aber vielfach herkömmliche statistische und stochastische Methoden verwendet. Analysen mit Hazardmodellen wurden mit den Daten der UTHS erst von Wilson (1994) und Filiou (1994) jeweils im Rahmen einer Abschlußarbeit an der Universität von London gemacht. Sie untersuchten dabei speziell das Verkehrsverhalten von unbeschäftigten bzw. beschäftigten Umfrageteilnehmern.

Filiou stellt zu allererst die Wichtigkeit der Tag zu Tag Variabilität (*day-to-day variability*) im Verkehrsverhalten des Einzelnen in den Vordergrund. Vorher basierten die meisten relevanten Arbeiten und Analysen, um longitudinale Veränderungen in der Mobilität (bzw. deren unterschiedliche Stufen) bemessen zu können, auf kurzzeitigen Datensätzen, die entweder aus Ein-Tages-Befragungen, höchstens Umfragen über eine Woche oder aus Tagebüchern erstellt worden waren (Pas und Koppelman, 1986; Kitamura, 1987). Auch Filiou streicht die Vorteile einer 5-wöchigen Befragung, wie sie mit den Daten der UTHS vorliegen, heraus. Mit Hilfe eines longitudinalen Datensatzes ist es möglich, die relative Häufigkeit, wie oft verschiedene Aktivitäten unternommen worden sind, anzugeben bzw. zu bemessen, die relative Bedeutung verschiedener Aktivitäten für den Einzelnen herauszufinden oder eine Abschätzung der Varianz der von der Person unternommenen Aktivitäten und besuchten Orte machen zu können. „Wenn Aktivitäten über eine längere Zeitspanne beobachtet werden, gibt es eine Möglichkeit, die Aspekte des individuellen Reiseverhaltens herauszufinden“ (Hanson und Hanson, 1988).

Filiou versucht, ähnlich wie in der hier vorliegenden Arbeit, mit Hilfe von Hazardfunktionen, die Effekte der Charakteristika der Individuen auf die Dauer von Aktivitäten und die Intervalle zwischen Aktivitäten, statistisch abzuschätzen. Er beschränkt sich dabei jedoch auf eine kurze beschreibende (deskriptive) Analyse und einen Vergleich der Auswertungen des Proportional Hazard Modells (PHM) mit dem *Accelerated Lifetime* Modell oder *Accelerated Failuretime* Modell (AFT-Modell). Eine weitere Einschränkung macht Filiou bei der Auswahl der Umfrageteilnehmer, indem er nur die in einem regelmäßigen Arbeitsverhältnis stehenden Personen in die Studie mit einbezieht. Er beschäftigt sich demnach mit den Aktivitätsmustern von 67 der insgesamt 144 Personen.

In der deskriptiven Analyse der Aktivitätsmuster erläutert Filiou zuerst die Klassifizierungen von Personen, Zeitpunkten, Aktivitäten und speziellen Funktionen (Haupteinkäufer, Fahrer, etc.) und beschreibt danach die durchschnittliche zeitliche Verteilung der Aktivitäten auf die einzelnen Kategorien, sowie die Verkehrsmittelwahl in den unterschiedlichen Gruppen.

Die Hazardmodell Analyse beginnt wiederum rein beschreibend mit einer Liste von Diagrammen, die jeweils die Verteilung der Aktivitätsdauern bzw. der Intervalldauern in den insgesamt sechs einzelnen Kategorien (Kommunikation, Einkauf, persönliche Erledigungen, soziale Belange, Arbeit und Freizeit) darstellen. Bei den Intervalldauern unterscheidet Filiou zusätzlich zwischen einer Verteilung aller Intervalle und der Verteilung der zensurierten Intervalle. Um einen Indikator für die Güte der Modelle zu erhalten, zeigt Filiou die Devianz-Residualverteilungen für die proportionalen Hazardmodelle wieder in Form von Diagrammen. Da die *martingale residuals* Eigenschaften ähnlich denen der linearen Regressionsmodelle zeigen, aber nicht symmetrisch um den Nullpunkt verteilt sind, werden sie zu *deviance residuals* transformiert, die dann genau dort liegen, wenn die Güte des Modells annehmbar

hoch ist. Dieselbe Möglichkeit der Überprüfung besteht bei den AFT-Modellen direkt mit Hilfe des SAS-Programmes LIFEREG jedoch nicht. Es können dabei nur die Cox-Snell-Residuenplots ausgegeben und damit die *martingale* und *deviance residuals* berechnet werden, was in der Arbeit jedoch nur erwähnt wird, nicht aber durch Ergebnisse präzisiert wird.

Den Schluß der Hazardmodell-Analyse stellt der tabellarische Vergleich der PH-Modelle und der AFT-Modelle für jede der 6 Aktivitätskategorien dar. Dabei werden die Parameterschätzwerte von sieben ausgewählten möglichen Einflußgrößen und die Signifikanz der Einflüsse präsentiert. Diese Einflußgrößen sind die im Vorfeld erläuterten, die Personen bzw. Gesellschaftsschichten oder den Zeitpunkt klassifizierenden Charakteristika (Alleine lebend, Jung, Kinder im Haushalt, Fahrer, Haupteinkäufer, Wochentag, Osterfeiertage), deren Größe und Richtung der Beeinflussung auf die Aktivitätsdauer bzw. Intervalldauer abgeschätzt werden soll.

Die Schlußfolgerungen aus den gewonnenen Ergebnissen versucht Filiou im abschließenden Kapitel für jede Aktivitätskategorie separat zu ziehen. Dabei ist jedoch kritisch anzumerken, daß nur ein Bruchteil der möglichen Einflußvariablen in die Untersuchungen mit einbezogen wurde und aus diesem Grund die Schlußfolgerungen in bezug auf Einflüsse gewisser Variablen nicht umfassend aussagekräftig sein können. Er verwendet alle sieben Einflußgrößen, ohne irgendwelche Korrelationen untereinander zu berücksichtigen oder weitere mögliche Variablen zu begutachten (vgl. Oh, 2000, der die Bedeutung möglichst vieler Einflußgrößen hervorhebt bzw. die Auswertungen von Kansky, 1967, der zwar 27 beschreibenden Variablen und 22 Meßwerte des Verkehrsverhaltensmusters einfließen läßt, diese jedoch zu einer einzigen resultierenden Variable zusammenfaßt und deswegen keine sehr guten Ergebnisse erhält).

Wilson (1994) arbeitet in sehr ähnlicher Art und Weise wie Filiou mit demselben Datensatz, den er jedoch nur in bezug auf die nicht in einem regelmäßigen Arbeitsverhältnis bzw. nicht arbeitenden insgesamt 72 Personen begutachtet. Nach einer kurzen deskriptiven Analyse der zeitlichen Verteilung der Aktivitätsdauern und Intervalldauern in Form von Diagrammen zeigt auch Wilson die Residuenplots für die PH-Modelle für die einzelnen Kategorien. Unterschiede zu Filiou's Auswertungen bestehen nebst der vorher erwähnten geänderten Auswahl von nicht arbeitenden Personen, nur in der Angabe der *martingalen* Residuenplots (statt *deviance* Residuenplots) und der Tatsache, daß die Aktivitätskategorie Kommunikation nicht mehr aufgelistet ist und demnach nur noch 5 verschiedene Kategorien vorhanden sind. Auch er bringt einen Vergleich des PH-Modells mit dem AFT-Modell und zeigt die Parameterschätzwerte bzw. die Signifikanzniveaus von insgesamt jeweils 7 verschiedenen ausgewählten möglichen Einflußgrößen. Diese Schlußfolgerungen aus diesen Ergebnissen bringt er am Ende noch stichwortartig zu Papier und führt zudem fünf Gründe an, wieso die Auswertungen mit Fehlern behaftet sein könnten.

## 7.3 STUDIEN MIT HAZARDMODELLEN ÜBER AKTIVITÄTSDAUERN UND INTERVALLDAUERN

### 7.3.1 Arbeitspapier von Lyoong Oh (2000)

Das Arbeitspapier von Lyoong Oh (2000) bzw. die Überarbeitung seines Arbeitspapiers aus dem Jahre 1999 zeigen Auswertungen und Diskussionen von Modellabschätzungen für die Dauer von Aktivitäten für verschiedene Aktivitätstypen. Grundlage der Auswertungen stellt ein Datensatz mit Reisetagebüchern von 300 Befragten über einen Zeitraum von 7 Tagen dar, der einerseits über 8000 Beobachtungen und andererseits auch wichtige reisebezogene und soziodemografische Attribute enthält.

Oh beruft sich in seiner Arbeit wie einige andere Wissenschaftler auch auf die von Axhausen und Polak (1991) formulierten Vorteile eines auf Aktivitäten basierenden Datensatzes im Gegensatz zu einem Datensatz, der vornehmlich mit Hilfe von Wegen und Reisen aufgebaut ist. Aktivitätsbezogene Annäherungen (ABA) beruhen auf einem viel kompletteren bzw. der Wahrheit näher liegenden Gerüst, um Verkehrsverhalten charakterisieren zu können, als es rein wegebezogene Annäherungen (WBA) können. Menschen machen nämlich prinzipiell Reisen bzw. Wege nicht um die Reise selbst zu machen, sondern hauptsächlich, um gewisse Bedürfnisse an verschiedenen Orten befriedigen zu können. Stark beeinflusst wird diese Theorie bzw. Formulierung der ABA natürlich von Hägerstrand (1970) (*Die Ausübung von verschiedenen Aktivitäten an verschiedenen Orten unterliegt immer zeitlichen und räumlichen Umwelteinflüssen bzw. Randbedingungen.*) und Chapin (1974).

Für die Modellschätzung greift auch Oh auf Hazard *Duration* Modelle zurück, da sie ob ihrer enormen Flexibilität in der Anwendung und ihres breit gefächerten Einsatzbereiches sehr gute Annäherungen bzw. Ergebnisse liefern und zudem signifikante kausale Faktoren, die den zeitlichen Mechanismen unterliegen, identifiziert und bemessen werden können. Es sind sowohl nonparametrische Hazardmodelle, die vor allem die deskriptive Analyse der Hazard und Survivalkurven beinhalten, als auch semiparametrische und vollparametrische Modelle angeführt. Zusätzlich wird versucht, bei den vollparametrischen Modellen noch etwaige Heterogenitätsfragen zu klären. Es wird in der gesamten Arbeit grundlegend nur zwischen zwei verschiedenen Aktivitätstypen, einerseits den obligatorischen bzw. pflichtmäßigen (z.B. Arbeit) und andererseits den freiwilligen bzw. ungezwungenen Aktivitäten (z.B. Freizeit oder Sozialbelange) unterschieden. Eine detailliertere Aktivitätenkategorisierung und weitere Modellschätzungen stellt Oh für die nächste Überarbeitung des Papiers in Aussicht.

Die nonparametrischen Hazard- bzw. Zeitdauermodelle werden in der Studie nur relativ kurz abgehandelt. Es wird jeweils die Kurve der Survival- und der Hazardfunktion zuerst für alle Aktivitäten, danach für pflichtmäßige und freiwillige Aktivitäten grafisch und tabellarisch dargestellt. Hierbei sind eine relativ steil abfallende Survivalkurve und höhere Hazardfunktionswerte bei geringen Aktivitätsdauern für freiwillige Aktivitäten (da diese meist kürzer andauern) und eine für die Arbeitskategorie typische Hazardkurve mit einem Höhepunkt bei ca. 8 Stunden zu sehen. Hierbei ist jedoch ein etwaiger Weg in der Mittagspause nicht berücksichtigt. Die Survivalkurve aller Aktivitäten hat - wie vermutet - einen exponentiellen Verlauf und die Hazardkurve verläuft relativ ruhig bzw. konstant mit einem Höhepunkt gegen Ende.

Auch die Auswertung und Ergebnisse der semiparametrischen Modelle werden von Oh nur kurz diskutiert. Er versucht dabei, die Einflussgrößen bzw. statistischen Signifikanzen von insgesamt 14 Variablen, die aus dem individuellen, sozioökonomischen, reisebezogenen und verantwortungsabhängigen Bereich gewählt worden sind, zu bemessen. Dabei sind bei den pflichtmäßigen Aktivitäten insgesamt 8 von 14 Kovariablen als beeinflussende Faktoren (im 90% Konfidenzintervall) erhoben worden, bei den freiwilligen jedoch nur 3 als statistisch signifikant erkannt worden. Das Vorzeichen der Koeffizienten gibt jeweils den Einfluss auf Verminderung (positives Vorzeichen) oder Erhöhung der durchschnittlichen Aktivitätsdauer an. Laut Oh liegt die Verbesserung des Loglikelihoods bei Null auf den Loglikelihood bei Konvergenz jedoch unter 1%, was den Schluß zulässt, dass entweder andere, in der Auswertung nicht berücksichtigte Kovariablen, Einfluss auf die Länge der Aktivitätsdauern haben könnten oder, dass die Faktoren von unbeobachteten Effekten in dieser Modellschätzung signifikant sind. Da die semiparametrischen Coxmodelle jedoch nicht genügend Flexibilität in bezug auf die Veränderung des Modells für Verbesserungen bieten, müssen etwaige Heterogenitätseffekte bei den vollparametrischen Modellen genauer betrachtet werden. Die Survivalkurven der semiparametrischen Modelle haben in ihrem Verlauf sehr große Ähnlichkeit mit den Ergebnissen der nonparametrischen Modellschätzungen.

Da bei den vollparametrischen Hazardmodellen eine Annahme für die Verteilung des Hazards bzw. für die Form der Funktion, wie sich Kovariablen im Modell gegenseitig beeinflussen, getroffen werden muss, stellt Oh zuerst ein vollparametrisches Modell ohne jegliche Kovariableneffekte auf. Die erhaltenen Ergebnisse decken sich ziemlich gut mit denen anderer Verkehrsmodelle aus der Literatur (Kiefer, 1998): eine konstante Hazardkurve für die Exponentialverteilung; einen exponentiellen Verlauf für die Weibullverteilung und einen exponentiellen Verlauf nach einem Höhepunkt zu einem sehr frühen Zeitpunkt der Zeitskala für die loglogistische Verteilungsannahme. Sowohl die Statistik des Loglikelihoods bei Konvergenz als auch der grafische Vergleich der negativen logarithmischen Survivalraten (Integrierte Hazardraten für theoretische und empirische Daten) schlagen eine Weibullverteilung der grundlegenden Hazardfunktion vor (*Weibull baseline hazard distributional assumption*).

Durch eine Analyse der Maximal-, Durchschnitts- und Standardabweichung zwischen den einzelnen Verteilungskurven der theoretischen und empirischen Daten wird dieser Vorschlag zusätzlich unterstützt. Die letztendlich getroffene Wahl einer Weibullverteilung als die beste Art zur Beschreibung der zeitlichen Verhaltensmuster bei Aktivitätsdauern geht auch mit vielen in der Literatur erwähnten Beispielen konform (Hamed & Mannering, 1993 und Bhat, 1996). Oh verwendet sie konsequent in allen Analysen mit vollparametrischen Hazardmodellen. Wie in den Auswertungen dieser Diplomarbeit zu sehen ist, gilt diese Aussage, dass sich die Weibullverteilung bei Verkehrsdaten immer am besten eignet, jedoch nicht immer. Die von Oh abgeschätzten Kovariableneffekte der vollparametrischen Modelle weichen geringfügig von den Ergebnissen der semiparametrischen Modelle ab. Es sind wiederum dieselben 14 maßgebenden Variablen in die Abschätzung miteinbezogen. Auffällig ist hier, dass die Dauern der pflichtmäßigen Aktivitäten von Variablen aller vier Kategorien beeinflusst werden, die der freiwilligen Aktivitäten jedoch vor allem von Variablen aus den Bereichen reiseabhängig und verantwortungsabhängig. Im Gegensatz zu den semiparametrischen Modellen (unter 1% Verbesserung) ist eine deutliche Steigerung des Wertes des Loglikelihood bei Null auf den Wert bei Konvergenz zu sehen (19.4 bzw. 4.6% für pflichtmäßige bzw. freiwillige Aktivitäten). Die Survivalraten hier zeigen deutliche längere Dauern für pflichtmäßige Aktivitäten im Vergleich zu freiwilligen Aktivitäten und logischerweise sind die Hazardraten für die Beendigung einer pflichtmäßigen Aktivität weitaus geringer.

Oh behandelt auch das Problem der Heterogenität, das eventuell zu falschen Schlüssen beim Verlauf der Hazardkurve und bei Abschätzungen der Kovariablenkoeffizienten führen kann. Bei proportionalen Hazardmodellen wird üblicherweise angenommen, dass sich die Survivalfunktion homogen über die Stichprobe verteilt und so jegliche Kovariableneffekte im Vektor  $Z$  enthalten sind. Unbeobachtete Heterogenität tritt jedoch dann auf, wenn nicht beobachtete Variablen die Aktivitätsdauern beeinflussen. Eine gute Möglichkeit, ein Modell zu erstellen, das Heterogenitätseffekte beinhaltet, ist, mit der bedingten Dauerdichtefunktion zu arbeiten. Ähnlich wie bei vollparametrischen Modellen ohne Kovariableneffekte wird auch hier versucht, mit Hilfe der Maximierung der Loglikelihoodfunktion eine geeignete Verteilung für die Heterogenitätsmodelle zu finden. Alle vier von Oh in Betracht gezogenen Modelle (Exponential, Exponential & Gammaheterogenität, Weibull und Weibull & Gammaheterogenität) bestätigen die in den vorhergehenden Modellschätzungen gewonnenen Erkenntnisse. Es sind nur geringfügige Heterogenitätseffekte erkennbar. Zudem gehen die von Oh gefundenen Ergebnisse mit jenen von Heckman und Singer (1984) konform, dass beispielsweise leicht höhere Parameterabschätzungen erzielt werden, wenn die Gammaheterogenität für unbeobachtete Effekte in den Modellen spezifiziert wird. Aufgrund der nur sehr geringen Heterogenitätseffekte ändern sich auch die Survival- und Hazardkurve nur unwesentlich.

Oh führt gegen Ende seiner Studie noch die Modelloption der Statusabhängigkeit (*state dependency*) an, wodurch man den Einfluss von früheren Erfahrungen auf die jetzt gemachten Beobachtungen untersuchen kann. Solche Abhängigkeiten können wichtige gewohnheitsmäßige Verhaltenseffekte mit einbeziehen, die große Auswirkungen auf die Aktivitätsdauer haben können. Heckman und Borjas (1979) haben die Abhängigkeiten in verschiedene Kategorien gegliedert (Markov (Status einer vorherigen Aktivität), Häufigkeit (Status mehrerer vorheriger Aktivitäten), Dauer (Status der Dauer bis zu einem beobachteten Punkt dieser Aktivität) und verzögerte Dauer (Status der Dauer der vorherigen Aktivität)). Oh führt diese Modelloption der Statusabhängigkeit und die von Heckman und Borjas erarbeiteten Formeln aber nur kurz an und nicht näher aus.

Als eine der möglichen Zukunftsperspektiven bzw. zukünftigen Forschungsrichtungen schlägt Oh den mehrstufigen Modellansatz vor. Dies bedeutet, dass neben den schon gemachten Spezifikationen (non-, semi- und vollparametrische Modelle, Verteilungsannahmen, unbeobachtete Effekte, etc.) mehr Wert auf soziologische Hierarchiestrukturen der Individuen in ihrem täglichen Leben gelegt werden sollte. Außerdem sollten Modelle gefunden werden, die unterschiedliche planmäßige Entscheidungen voraussagen können.

### 7.3.2 Arbeiten von Schönfelder und Axhausen (2000)

Schönfelder und Axhausen versuchen in ihren Arbeiten (Periodizität im Verkehrsverhalten: Erste Ergebnisse mit Überlebenszeitmodellen, 2000 und *Analysing the rhythms of travel using survival analysis*, 2000) ebenfalls, der Problematik der Periodizität des Verkehrsverhaltens mit Hilfe von stochastischen Modellen aus der Überlebenszeitanalyse bzw. Hazardmodellen Herr zu werden und die rhythmischen Muster des Verkehrsverhaltens näher zu beschreiben. Sie beziehen sich in beiden Studien auf die sehr umfangreichen über einen Zeitraum von 6 Wochen erhobenen Datensätze aus dem Forschungsprojekt *Mobidrive*, das im Herbst 1999 in den beiden deutschen Städten Halle/Saale und Karlsruhe durchgeführt wurde. Dabei wurden neben der aktivitätsbezogenen Befragung wesentlich mehr Informationen über die Sozio-Demografie der Reisenden (Beruf, Einkommensverhältnisse, Fahrzeugverfügbarkeit, Dauerkarten für den öffentlichen Verkehr, Telekommunikationsstand) und insbesondere deren Rolle im Haushalt (Haushaltszusammensetzung, Haushaltsvorstand, Anzahl und Alter der Kinder, Aufgabenverteilung im Haushalt, etc.) als beispielsweise in der relativ alten UTHS erhoben. Zusätzlich wurde parallel eine weitere Befragung durchgeführt, die Informationen bezüglich Gruppen- und Vereinszugehörigkeit, der Einteilung des disponiblen Einkommens oder der Einstellungen zu verschiedenen Verkehrsmitteln beinhaltet.

Bisher haben Untersuchungen zu rhythmischen Mustern der Mobilität bei den Forschungen zu zeitlichen Aspekten des Verkehrsverhaltens auf der Ebene von Personen und Haushalten eine untergeordnete Rolle gespielt, Verhaltensroutinen, Verhaltensvariabilität und Verhaltensdynamik wurden indes eingehend untersucht. Mit Hilfe des grundlegenden Datensatzes mit langfristigem Raum-Zeit-Verhalten der Verkehrsteilnehmer wird es jedoch möglich, die Periodizität auf der Wochen und Monatebene abzubilden. Somit kann eine weitere zeitliche Perspektive der Analysen des aktivitätenbasierten Ansatzes erreicht werden. Für verkehrsrelevante Planungen und Massnahmen öffentlicher sowie privater Akteure stellt nämlich das Verständnis individueller Mobilitätsgewohnheiten wie rhythmischer Muster eine essentielle Entscheidungsgrundlage dar.

Rhythmische Muster werden als wiedererkennbare Elemente verstanden, die in regelmässigen Abständen im Verkehrsverhalten von Personen und Haushalten beobachtet werden können. In diesen Arbeiten sind insbesondere Rhythmen gemeint, die über längere Zeiträume andauern (Wochen- oder Monatebene) und damit weit über den üblichen Beobachtungszeitraum eines Stichtages (vgl. KONTIV) hinausgehen. Zu den rhythmischen Mustern würden komplette Tagesprogramme mit übereinstimmenden Attributen oder Teilsequenzen bzw. einzelne Hauptaktivitäten oder –wege gehören, was in diesen ersten Arbeiten nur angerissen wird. Rhythmen können generell durch ihre Perioden und Amplituden typisiert werden. Schönfelder beschränkt sich bei seinen Arbeiten jedoch auf die Analyse der Periodizität des Verkehrsverhaltens, die durch die Anordnung und Länge der Zeitintervalle zwischen gleichartigen Aktivitäten (Mustern) beschrieben werden kann (vgl. Intervalldauern in dieser Diplomarbeit). Die Amplituden der rhythmischen Muster können beliebig komplex gewählt werden, von der eher eindimensionalen Betrachtung der Aktivitäten mit lediglich einem Attribut (z.B. Zweck) bis zu einer komplexen multiplen Charakterisierung von Mustern nach verschiedenen Attributen. Auch er nimmt die gewählten Kategorisierungen der Aktivitäten, Wege, etc. aus früheren vergleichbaren Studien (z.B. Pas, 1983; Hanson und Huff, 1986; Wilson, 1998).

Die konzeptionelle Grundlage des Ansatzes von Schönfelder bei der Verwendung der Hazardmodelle ist die Vorstellung, dass rhythmische Muster als Ereignisse angesehen werden können, deren Eintrittswahrscheinlichkeiten über einen Zeitraum annähernd genau bestimmt werden können. Im Zeitintervall zwischen zwei Ereignissen (Aktivitäten) findet ein Bedürfnisaufbau statt, der durch eine Wahrscheinlichkeitsfunktion ausgedrückt werden kann und an dessen Ende (relatives Maximum) der Ereigniseintritt steht. (Beginn einer neuen Aktivität). Dieser Ansatz wird auch in anderen Studien als durchaus legitimer verwendet (vgl. Abbildung 4.1, Seite 36).

Zu Beginn steht eine kurze deskriptive Analyse (ebenfalls mit Hilfe des SAS-Programms PROC LIFETEST), in der vor allem die Hazard- und Survivalkurven für verschiedene ausgewählte Aktivitätskategorien gezeigt werden. Nicht näher wird auf die soziodemografischen Merkmale der Bevölke-

rungsstichprobe in bezug auf Verkehrsmittelwahl, Altersstrukturierung, Haushaltsstrukturen, etc. eingegangen.

In den nachfolgenden ersten Modellschätzungen mit dem semiparametrischen Cox-Modell werden exemplarisch die Ergebnisse der beiden Kategorien Kurzfristiger Einkauf und Sport, die beide bei den empirischen Untersuchungen der Intervalle eine deutliche Tendenz zur Periodizität gezeigt haben, separat für die beiden Befragungsstädte präsentiert und kurz diskutiert. Zur Überprüfung der grundlegenden Annahme, daß das geschätzte Hazardmodell tatsächlich proportional zur *baseline hazard* steht, sind Proportionalitätstests durchgeführt worden. Schönfelder wählt weiter 10 beschreibende Variablen aus den Bereichen Personenbezogen bzw. Haushaltsbezogen und zudem den Beschäftigungsstatus als mögliche Einflußgrößen aus, deren Parameterschätzwert und Risk ratio dann in den Ergebnissen der Hazardmodellschätzungen angegeben werden. Bei beiden Kategorien haben die ad hoc gewählten Kovariablen laut Autor nur einen bedingten Einfluss auf die Intervalldauern, da das berechnete  $R^2$  nur sehr klein ist. Er geht kurz auf Unterschiede zwischen den beiden Fallstudien ein. Dabei fallen neben etwa 20% mehr berichteten Wegen insgesamt in Karlsruhe wesentlich kürzere Intervalle zwischen Einkaufsaktivitäten der vermeintlich (auto-)mobileren Verkehrsteilnehmern (Führerschein / Autobesitz des Haushaltes) aus Karlsruhe und die wesentlich kürzeren Intervalle bei Männern auf. Ähnliche Tendenzen sind bei der Signifikanz der Vollzeitberufstätigkeit (deutlich längere Intervalle zwischen den Einkäufen) und beim Einfluss des höheren Haushaltseinkommens (höhere Frequenz der Einkaufsaktivitäten) zu sehen. Beim Modell für die Aktivität Sport sind die Variablen Geschlecht, Führerscheinbesitz und Vereinsmitgliedschaft signifikant für beide Städte. Männer und – wie zu erwarten war – Befragte, die in Vereinen organisiert sind, weisen kürzere Intervalle zwischen den Sportaktivitäten auf. Ebenso interessant sind die wesentlich höhere Aktivitätsteilnahme (kürzere Intervalle) von Führerscheininhabern und Personen, die nicht Vollzeitbeschäftigten (mehr als 30 Stunden Wochenarbeitszeit).

In der ersten Arbeit ist in einem methodischen und inhaltlichen Ausblick unter anderem noch die Entwicklung eines vollparametrischen Modells und die Einrechnung von unbeobachteten bzw. bisher nicht unberücksichtigte Heterogenitäten innerhalb der Stichprobe in Aussicht gestellt.

Im zweiten bis dato vorliegenden Arbeitspapier sind auch erste Ergebnisse der vollparametrischen Modellschätzung mit der Annahme einer Weibullverteilung der Basisfunktion enthalten. Die Weibullverteilung wählt Schönfelder vor allem aufgrund der häufigen Verwendung in vergleichbarer Verkehrsliteratur (Hamed und Mannering, 1993; Hensher und Mannering, 1994; Bhat, 1996a,b), die jeweils diese Verteilungsart als wichtigen Faktor für den Zusammenhang zwischen Verhalten und Zeitdauer (bzw. vor allem Intervalldauer) ansehen. Als Überprüfung der Paßform führt er noch den besse-

ren Loglikelihoodwert des Weibullmodells gegenüber weiteren möglichen Modellen (Exponentialmodell oder Log-logistisches Modell) an, führt aber keine anderen Überprüfungen aus.

Die Ergebnisse der vollparametrischen Modelle sind jenen der semiparametrischen sehr ähnlich. Auffällig und unerwartet ist jedoch der Effekt der Weibullverteilung auf den *baseline hazard*. Für die Kategorie Einkauf kurzfristig bzw. täglicher Einkauf wird eine fallende Hazardkurve erzeugt, die jedoch nach ca. 1 Tag so gut wie konstant zu verlaufen scheint. Bei einer genaueren Betrachtung der anderen Kategorien, die in der Arbeit nicht weiter ausgeführt sind, fällt diese Tendenz ebenfalls auf. Für häufige d.h. regelmäßige Aktivitäten mit relativ kurzen Intervallen verursacht das Weibullmodell im Gegensatz dazu eine abnehmende bzw. fallende Hazardkurve.

In der abschließenden Diskussion bzw. dem methodischen Ausblick unterstreicht Schönfelder, wie wichtig eine gediegene Auswahl von Kovariablen der verwendeten parametrischen Modelle durch geeignete und detaillierte Signifikanztests ist. Weiters sollte bei den vollparametrischen Modellen in Zukunft doch eine weitere Diskussion ob der passenden Verteilungsannahme der *baseline hazard* geführt werden und zensurierte (Intervall-)zeiten inkludiert werden. Die Berücksichtigung von Heterogenitätseffekten sollte ebenfalls in weiteren Studien berücksichtigt werden.

## 7.4 WEITERE STUDIEN MIT HAZARDMODELLANALYSEN

In der Fachliteratur sind noch einige weitere, interessante Studien mit Hazardmodellanalysen aus der Verkehrsforschung zu finden. Diese Arbeiten alle vorzustellen und zu diskutieren, würde jedoch hier den Rahmen sprengen. Deshalb sollen nur ein paar noch kurz erwähnt werden.

Chandra Bhat (1996) beschäftigt sich beispielsweise in seiner Studie mit den Faktoren, die auf die Dauer von Einkaufsaktivitäten speziell auf dem Weg von der Arbeit nach Hause Einfluß nehmen und versucht, ein einheitliches methodisches Grundgerüst zu entwickeln, um ein stochastisches hazard-basierendes Zeitdauermodell aus gruppierten, in Intervalle abgestuften Ereignisdatensätzen abschätzen zu können.

Das Ziel der Studie von Hamed und Mannering (1993) ist es, eine geeignete Methode zu finden, um die freizeitmässigen Aktivitätsmuster bzw. jene, die nach der Arbeit auftreten, abschätzen zu können. Es wird einerseits eine Reihe von unstetigen und stetigen Modellen erstellt, die dazu beitragen sollten, die Aktivitätswahl (unstetig), die Reisezeit zur Aktivität und wieder nach Hause (stetig) und die Aktivitätsdauer (stetig) des Reisenden beschreiben und voraussagen zu können. Andererseits erstellen die

Autoren noch mit Hilfe der Hazardmodellanalyse *Duration*-Modelle, um die Einflußfaktoren abschätzen zu können, die Personen motiviert, ihr Haus zu verlassen und gewisse Aktivitäten außer Haus durchzuführen. Zusätzlich wird noch ein einfaches Simulationsexperiment der erarbeiteten Modelle gezeigt.

Hensher und Mannering (1994) behandeln in ihrer Arbeit vor allem die Problematik der Hazardmodelle an sich. Sie bemängeln, daß solche Modelle mit Zeitdauerdaten zwar überall in Verkehrsbereich vorkommen, jedoch viel zu wenige Studien bisher sich damit beschäftigt haben und diese Phänomene statistisch analysiert worden sind. Sie geben in ihrer Arbeit einen Überblick, in welchen Bereichen der Verkehrsproblematik Hazardmodelle geeignet eingesetzt werden könnten, stellen die verschiedenen Arten von Hazardmodellen auch in bezug auf Datenstrukturierung, lückenhafte Datensätze, mögliche bzw. geeignete Verteilungsannahmen und die damit verbundenen Auswirkungen vor und behandeln weitere Probleme, die infolge von Heterogenität und State Dependence auftreten können. Am Ende der Arbeit werden noch ein paar damals veröffentlichte Studien mit Hazardmodellen (vor allem von Jovanis, Hamed, Hensher und Mannering) behandelt und weitere Zukunftsaussichten vorgestellt.

## **8 AKTIVITÄTSMUSTER**

### **8.1 EINLEITUNG**

Wie schon im Kapitel 4 beschrieben, kann man mit Hilfe verschiedenster Beobachtungen, die vor allem die Reisen und Aktivitäten der befragten Personen beschreiben, und daraus erarbeiteten Statistiken das Reiseverhalten analysieren und Aktivitätsmuster der betreffenden Einzelpersonen, unterschiedlicher Personenschichten oder spezifischer Gruppen erstellen. Die vorhandenen 144 Einzelpersonen lassen sich in verschiedene Gruppen mit speziellen Merkmalen einteilen, wobei sowohl soziodemografischen als auch aktivitätenspezifische Faktoren eine Rolle spielen. Prinzipiell wurden in den Auswertungen Einteilungen bezüglich der Personenschicht, des Geschlechts, der Altersschicht, der Haushaltsgröße, des Arbeitsstatus (Vollzeit, Teilzeit, marginal oder nicht beschäftigt), des Führerscheinbesitzes, etc. gemacht.

In den folgenden Unterkapiteln werden zunächst die Gruppenklassifizierung und die durchschnittliche Aktivitätsdauer pro Tag bzw. pro Aktivität behandelt, darauf folgen die Aktivitätsdauer und Aktivitätshäufigkeit und deren Verteilung auf die Wochentage, auf Arbeitstage und arbeitsfreie Tage, sowie auf die einzelnen Kategorien, immer mit der zusätzlichen Unterscheidung der beiden Geschlechter. Abschließend werden noch einige primäre Ergebnisse der deskriptiven Analyse der Verkehrsmittelverteilung bzw. der anschließenden Aktivitäten beleuchtet. Auch die Verteilung der Aktivitätsdauern und die Intervalldauern zwischen gleichartigen Aktivitäten werden dargestellt.

### **8.2 KLASSIFIZIERUNG DER PERSONEN UND ANDERE CHARAKTERISTIKA**

#### **8.2.1 Beschreibung der Personen- bzw. Haushaltsgruppen**

Das vorhandene und gut ausgearbeitete schwedische Bevölkerungsregister (*Swedish Person Register*) bildete die Grundlage für die Auswahl jener Personen, die an der Umfrage teilnahmen. Dieses Register stellt vor allem für die damalige Zeit eine ideale Grundlage für die Befragung dar, da nebst Namen,

Adresse und Haushaltskoordinaten auch soziodemografische Daten der Personen wie beispielsweise Geburtstag, Geschlecht, Wochenarbeitsstunden, Ausbildungsjahre, Führerscheinbesitz, Berufsbezeichnung und Haushaltsgröße angeführt sind. Aufgrund der Einteilung in 6 verschiedene vordefinierte Personengruppen wurden später die Stichproben ausgewählt. Die Kriterien für die weitere Strukturierung der Befragung wurden im Kapitel 2.3 näher beschrieben. Nachdem in den folgenden Auswertungen noch einige Male diese Gruppeneinteilung vorkommt, ist es nötig, die 6 Gruppen (*life cycle groups*) zu beschreiben (nach Marble *et. al.*, 1972):

Gruppe 1, als "*elderly households*" definiert, bezeichnet Haushalte, deren Vorstand im November 1970 67 Jahre alt oder älter war. Haushalte in dieser Gruppe bestanden aus ein oder zwei Personen, der Großteil (13 von 19) Haushalten bestand aus nur einer Person.

Gruppe 2, "*older households without children*", bezeichnet Haushalte, deren Vorstand zwischen 50 und 66 Jahren alt war und ohne Kinder im selben Haushalt. Daher scheinen in Gruppe 2 nur Ein- oder Zwei-Personen-Haushalte auf (Unverheiratete, getrennte, geschiedene oder verwitwete Personen oder Ehepaare ohne Kinder oder ohne noch im selben Haushalt wohnende Kinder).

Gruppe 3, "*younger, single person households*", bezeichnet Haushalte, deren Vorstand zwischen 18 und 49 Jahren alt war und prinzipiell aus nur einer Person bestand.

Gruppe 4, "*younger, married couples with no children*", bezeichnet Haushalte, deren Vorstand zwischen 18 und 49 Jahren war und die prinzipiell aus zwei miteinander verheirateten Personen bestanden, Kinder waren jedoch keine im Haushalt. Ein einziger Haushalt (hh\_id = 1409) mit 3 Personen ist enthalten, wobei die dritte jedoch ebenfalls ein Erwachsener ist.

Gruppe 5, "*younger couples with no pre-school children*", bezeichnet Haushalte, deren Vorstand zwischen 18 und 49 Jahren war und die prinzipiell aus zwei miteinander verheirateten Personen und einem oder mehreren Kindern, die alle über 7 Jahre alt waren, bestanden. Einzelfälle mit nur einem Erwachsenen und einem Kind über 7 Jahren sind jedoch enthalten.

Gruppe 6, "*younger couples with pre-school children*", bezeichnet Haushalte, deren Vorstand zwischen 18 und 49 Jahren war und die prinzipiell aus zwei miteinander verheirateten Personen und zumindest einem unter 7 Jahre alten Kind bestanden. Einzelfälle mit nur einem Erwachsenen und einem Kind unter 7 Jahren sind enthalten.

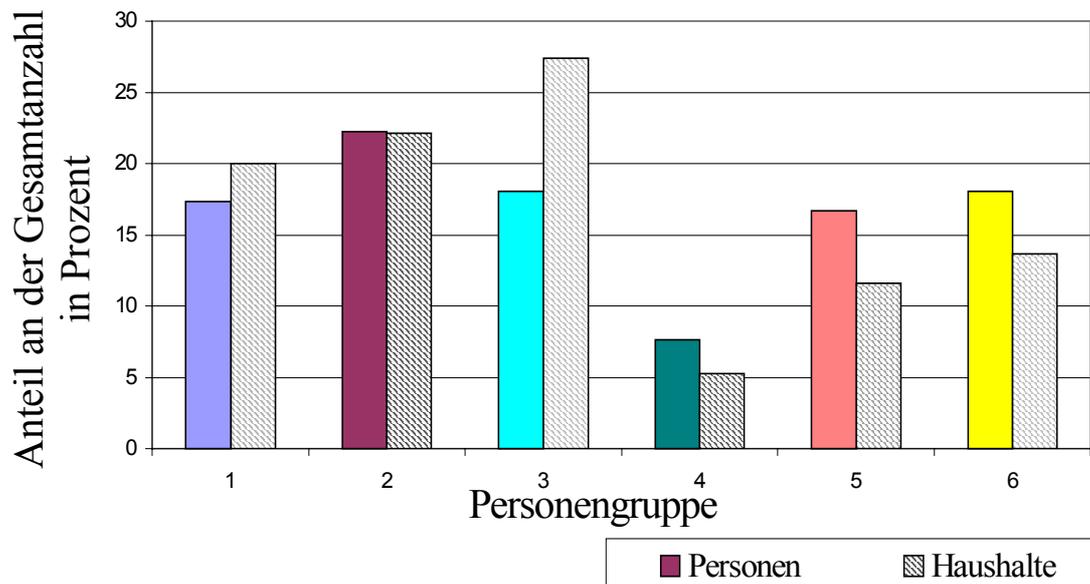
Die besonderen Merkmale der einzelnen Personengruppen und grundlegende Unterschiede sind in den folgenden Grafiken und Tabellen dargestellt.

Tabelle 8.1 : Verteilung der Geschlechter in den Personengruppen

Personengruppe	1	2	3	4	5	6	Gesamt
Weiblich	19	19	14	5	11	13	81
Männlich	6	13	12	6	13	13	63
Gesamt	25	32	26	11	24	26	144

Die Gesamtanzahl der 144 befragten Personen teilt sich in 81 weibliche und 63 männliche Studienteilnehmer auf, die genaue Verteilung auf die einzelnen Personengruppen sind in der Tabelle 8.1 dargestellt. Vor allem in den ersten beiden Gruppen, in denen hauptsächlich ältere Ehepaare oder alleinstehende Personen leben, überwiegen die weiblichen Personen. In der Gruppe 1 sind sogar in keinem der 13 Einpersonenhaushalte männliche Personen vorhanden. In der Single-gruppe 3 und den Gruppen 4, 5 und 6 sind weibliche und männliche Bewohner ziemlich ausgeglichen verteilt.

Abbildung 8.1 : Verteilung der Personen und Haushalte auf die Personengruppen in Prozent



Die prozentuelle Verteilung der Personen an der Gesamtanzahl einerseits und der Haushalte andererseits ist in Abbildung 8.1 dargestellt. Dabei fällt auf, daß vor allem in der Gruppe 4 sehr wenige Personen (11 von 144 Personen = 7,64 %) die Umfrage beendet haben, die restlichen Gruppen mit Ausnahme der Gruppe 2 (22,22 %) alle bei ca. 17 bis 18 % (24-26 von 144 Personen) gleichverteilt liegen.

Bei den Haushalten liegt eindeutig die Gruppe 3 (Single-Haushalt) mit über 27 % vorne, die Ein- oder Zweipersonenhaushalte aus den Gruppen 1 und 2 liegen mit knapp über 20 % vor den größeren Haushalten der Gruppen 5 und 6. Auch hier stellt die Gruppe 4 mit Abstand den kleinsten Anteil (5,26 %).

Tabelle 8.2 : Verteilung der Haushaltsgröße in den Personengruppen

Gruppe	1	2	3	4	5	6	Gesamt
1 Person	13	10	26	0	0	0	49
2 Personen	6	11	0	4	1	0	22
3 Personen	0	0	0	1	6	5	12
4+ Personen	0	0	0	0	4	8	12
Gesamt	19	21	26	5	11	13	95

Die Haushaltsgrößen und deren Verteilung auf die einzelnen Personengruppen sind in Tabelle 8.2 dargestellt und zeigen ein deutliches Übergewicht der Drei- und Vier-oder-mehr- Personenhaushalte in den Gruppen 5 und 6, sowie ein großes Vorkommen von Einzelpersonenhaushalten in den Gruppen 1 und 2, sowie auch Gruppe 3, was dort aber die einzig möglichen Variante ist.

### 8.2.2 Arbeitsstatus der Personen

Aufgrund der tatsächlichen Gesamtarbeitszeiten, den Tagen, an denen gearbeitet wurde und die Häufigkeit der Aktivität Arbeit über die Umfragedauer von 35 Tagen wurde bei jeder Person erneut die tatsächliche Arbeitszeit in den 5 Wochen der Befragung berechnet und mit der angegebenen Wochenarbeitszeit verglichen. Dabei ergaben sich teilweise gravierende Unterschiede, die eine andere Arbeitsstatus-klassifizierung zur Folge hatten. Zudem wurde die Klasse „marginal arbeitend“ eingeführt, die eine Arbeitszeit zwischen 0.3 und 5 Stunden pro Monat einschließt. Die Verteilung auf die verschiedenen Personengruppen bzw. Geschlechter wird in den Tabellen 8.3 und 8.4 gezeigt.

Tabelle 8.3 : Verteilung des Arbeitsstatus über die Personengruppen

Personengruppe	1	2	3	4	5	6
Vollzeit	3	17	18	9	18	16
Teilzeit	3	5	5	1	1	1
Marginal	0	0	1	0	1	1
Nicht beschäftigt	19	10	2	1	4	8

In der ersten Gruppe sind relativ viele Personen schon im Ruhestand und daher nicht erwerbstätig (19 von 25 Personen), in den restlichen 5 Gruppen überwiegen die Vollzeitbeschäftigten, nur in den Gruppen 2 und 6 sind ebenfalls viele nicht Erwerbstätige (17 bzw. 31 %, hauptsächlich (Ehe-) Frauen) enthalten. Teilzeitarbeit ist auch eher bei den älteren Bevölkerungsgruppen bzw. Gruppen ohne Kinder im Haushalt denn in den Gruppen 4, 5 und 6 (jeweils nur eine Person) beliebt.

Tabelle 8.4 : Verteilung der Personen auf die Arbeitsstatusgruppen

Beschäftigung	Vollzeit	Teilzeit	Marginal	Nicht	Gesamt
Weiblich	35	10	1	35	81
Männlich	46	6	2	9	63
Gesamt	81	16	3	44	144

Während bei Nichterwerbstätigen der Großteil Frauen sind (knapp 80 %), überwiegt bei den Vollzeitbeschäftigten mit 57 % die Gruppe der Männer, bei den restlichen beiden Arbeitsstatusgruppen (teilweise bzw. marginal beschäftigt) sind die allgemein geringe Anzahl (gesamt 16 bzw. 3 von insgesamt 144) und der ziemlich ausgeglichene Anteil von weiblichem und männlichem Geschlecht auffällig.

Insgesamt sind also die Aktivitäten von 81 Vollzeit beschäftigten Personen und von 63 anders oder nicht arbeitenden Personen in der Studie berichtet.

### 8.2.3 Arbeitstage und arbeitsfreie Tage

Von den möglichen 35 (bzw. bei manchen Personen bis zu 37) Befragungstagen sind 5 Samstage und 5 Sonntage, sowie Freitag, der 12. 4. 1971 (Karfreitag) und Montag, der 15.4.1971 (Ostermontag) als insgesamt 12 arbeitsfreie Tage definiert, womit noch 23 potentielle Arbeitstage übrigbleiben. Es sind dies zwar nicht bei allen auch wirklich Arbeitstage gewesen, jedoch als solche deklariert.

### 8.2.4 Aktivitätskategorien

Da die ursprünglich 56 Aktivitätskategorien aus der Uppsala-Umfrage für eine übersichtliche Betrachtung ungeeignet sind, wurde in der Datenaufarbeitung sowohl die allgemeine KONTIV-Kategorisierung mit 10 verschiedenen Klassen, als auch die in der City-mobil-Umfrage 23-teilige Kategorisierung (vgl. Götz, Jahn und Schultz, (1997), Schönfelder und Axhausen, (2000)) eingeführt. Um eine geeignete und übersichtliche Darstellung der Ergebnisse zu erlangen, werden die Auswertun-

gen immer mit den 10 Kategorien der Standardeinteilung (KONTIV) oder mit einer neunteiligen Kategorisierung für alle Aktivitäten außer Haus gezeigt. Die exakten Kategorienbezeichnungen sowie die Unterschiede zwischen den einzelnen Kategorisierungen sind im Anhang II, Seite 178 zu finden.

### 8.3 DURCHSCHNITTLICHE AKTIVITÄTSDAUER PRO TAG BZW. AKTIVITÄT

Grundsätzlich können die durchschnittlichen Aktivitätsdauern zumindest für zwei verschiedene Bezugsebenen berechnet werden. Einerseits ist die durchschnittliche Aktivitätsdauer pro Person und Tag ein aussagekräftiges Argument in der Verkehrsmusteranalyse, andererseits stellt auch die durchschnittliche Aktivitätsdauer pro Aktivität eine geeignete Grundlage dar.

In der folgenden Tabelle (Tabelle 8.5) ist die durchschnittliche Aktivitätsdauer außer Haus pro Person und pro Tag dargestellt, wobei in jeder Aktivitätskategorie alle Zeiten aufsummiert und durch die Gesamtanzahl aller Umfragebeteiligten und durch die gesamte Umfragedauer von 35 Tagen dividiert worden sind. Dadurch ergeben sich vor allem in jenen Kategorien, in denen einige Personen gar nie Aktivitäten durchführten oder mit nur wenigen Aktivitäten, kurze Aktivitätsdauern. So liegt beispielsweise die durchschnittliche Aktivitätsdauer in der Kategorie Arbeit für die Gesamtbevölkerung bei etwas über zweieinhalb Stunden pro Tag. Würde man nur die arbeitenden Personen in diese Statistik einbeziehen, ergebe sich eine durchschnittliche Dauer von knapp vier Stunden (3,92) pro Tag bzw. von knapp sechs Stunden (5,97) pro Arbeitstag (bezogen auf 23 Arbeitstage in der Umfragezeit).

Durch die geringe Anzahl von Aktivitäten in den Kategorien „Jemanden Bringen/Abholen“, „Dienstleistung / Erledigungen“, „Arbeitsbezogen“ und „langfristiger Einkauf“ (siehe auch Abbildung 8.9) sind auch vor allem im Vergleich mit den Aktivitätsdauern in der nachfolgenden Tabelle 8.6 die durchschnittlichen Aktivitätsdauern hier sehr gering. Sie liegen alle bei unter 10 Minuten pro Tag pro Person, bei der Kategorie „Schule / Ausbildung“ bei 11 Minuten pro Tag. Diese Zahlen beziehen sich jedoch immer auf alle 144 Personen der Umfrage, wobei nicht jeder Umfrageteilnehmer auch Aktivitäten in dieser Kategorie unternommen haben muß. Vor allem in der Kategorie Schule/Ausbildung sind nur sehr wenige Personen involviert, was man am deutlichen Unterschied in der Aktivitätsdauer bezogen auf alle 144 Personen (11 Minuten pro Tag) bzw. auf jene 31, die zumindest eine solche Aktivität durchgeführt haben (ca. eine halbe Stunde pro Tag bzw. 0.78 Stunden pro Arbeitstag) sieht.

Tabelle 8.5 : Durchschnittliche Aktivitätsdauer außer Haus pro Person pro Tag in Stunden

Aktivitätskategorie	Weiblich	Männlich	Gesamt
Freizeit	3,17	3,93	3,50
Arbeit	2,00	3,65	2,72
Einkauf kurz	0,26	0,16	0,22
Schule / Ausbildung	0,07	0,16	0,11
Sonstiges	0,09	0,10	0,10
Dienstleistung / Erledigung	0,09	0,05	0,07
Einkauf lang	0,02	0,02	0,02
Bringen / Abholen	0,02	0,02	0,02
Arbeitsbezogen	0,00	0,02	0,01

Die größte durchschnittliche Aktivitätsdauer pro Tag pro Person ist in der Kategorie Freizeit mit dreieinhalb Stunden pro Tag gegeben, wobei die Männer sogar knapp 4 Stunden täglich freizeitorientierte Aktivitäten aufweisen. Hierbei sind aber neben den tatsächlichen Freizeitaktivitäten wie Erholung in der Natur, der Besuch kultureller Veranstaltungen, Sport, etc. auch Aktivitäten wie Verwandten- oder Freundesbesuche, Vereinsaktivitäten, Kommunikationsbelange (Telefon, etc.) oder Essen mit Freunden und Arbeitskollegen inbegriffen.

Die durchschnittlichen Aktivitätsdauern pro Aktivität sind in Tabelle 8.6 dargestellt. Auf den ersten Blick sind in allen Kategorien die durchschnittlichen Dauern länger als in der Tabelle 8.5, was aber aufgrund der Tatsache, daß Aktivitäten nicht in jeder Kategorie jeden Tag durchgeführt werden können, durchaus plausibel ist. Auch hier täuscht die berechnete durchschnittliche Dauer der Aktivität „Arbeit“ etwas, da sehr viele Personen ihre tägliche Arbeit in zwei Teilen angegeben haben, weil sie durch einen Weg zum Mittagessen oder diversen Erledigungen und einen Weg zurück zur Arbeitsstätte die Arbeit unterbrochen hatten. Diese Zweiteilung wird durch Abbildung 8.2 bestätigt, wo die Verteilung der Aktivitätsdauern dargestellt ist. Man sieht einen unproportional hohen Anteil von vierstündigen Arbeitsdauern, was durch die Mittagspause erklärt werden kann (Filiou, 1994).

Auch in dieser Auswertung dauern die Aktivitäten der Kategorie „Freizeit“ mit durchschnittlich knapp 6 Stunden am längsten, knapp gefolgt von Arbeitsaktivitäten mit ca. 5 Stunden und 20 Minuten. Arbeitsbezogene Aktivitäten dauerten im Schnitt ca. eine Stunde, kurze Einkäufe rund 20 Minuten. Für Schule oder Ausbildung wurden im Schnitt 3 Stunden aufgewendet, (im Vergleich zur durchschnittlichen täglichen Aktivitätsdauer aller Personen (11 Minuten) natürlich ungleich mehr). Aktivitäten der Kategorie „Jemanden Bringen / Abholen“ dauerten im Schnitt 15 Minuten, Dienstleistungen und Erledigungen rund 8 Minuten länger.

Abbildung 8.2 : Verteilung der Aktivitätsdauern in der Kategorie „Arbeit“

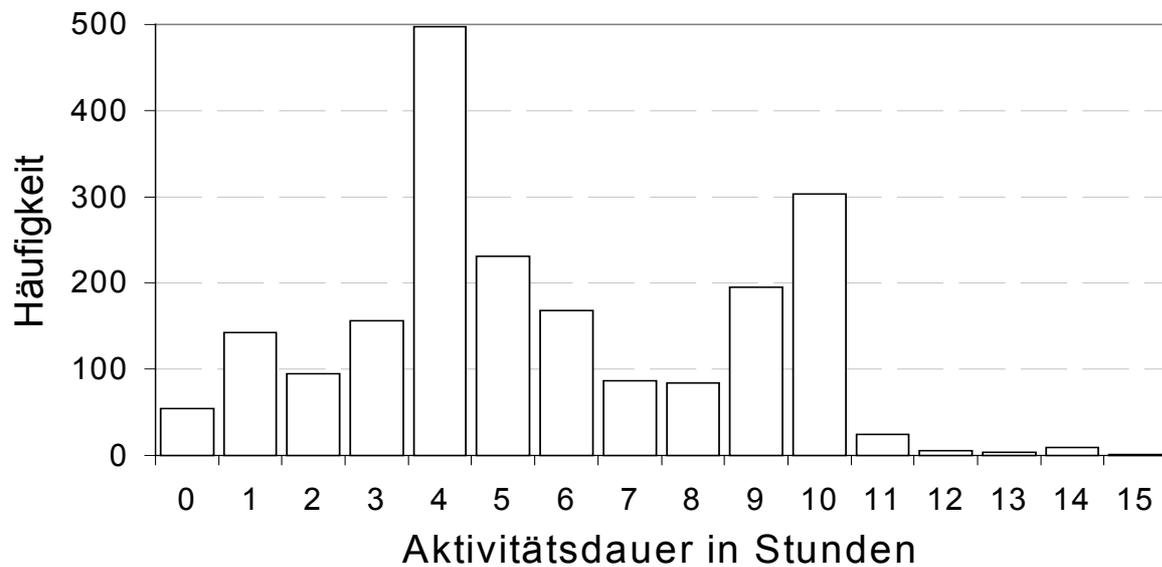


Tabelle 8.6 : Durchschnittliche Aktivitätsdauer außer Haus pro Aktivität in Stunden

Aktivitätskategorie	Weiblich	Männlich	Gesamt
Freizeit	6,18	5,50	5,83
Arbeit	5,41	5,25	5,31
Schule / Ausbildung	2,81	3,12	3,00
Sonstiges	1,72	1,24	1,46
Arbeitsbezogen	0,41	1,08	0,97
Dienstleistungen / Erledigung	0,48	0,27	0,39
Einkauf kurz	0,35	0,31	0,34
Einkauf lang	0,23	0,35	0,28
Bringen / Abholen	0,23	0,25	0,24

Die berichteten Aktivitäten mit einer Gesamtanzahl von 17003 teilen sich wie in Abbildung 8.3 und Tabelle 8.7 ersichtlich zu einem knappen Drittel (5863) auf die Aktivität „Zu Hause“ und zu jeweils ca. einem Fünftel auf die Aktivitäten „Arbeit“, „Freizeit“ und „kurzfristiger Einkauf“ (zwischen 2581 und 3231) bzw. zu einem geringen Prozentsatz auf die restlichen sechs Kategorien auf. In den meisten Kategorien ist eine relativ ausgeglichene Häufigkeitsverteilung zwischen den Geschlechtern gegeben, ausgenommen davon sind Kurzfristiger Einkauf, den die weiblichen Umfrageteilnehmer fast doppelt so oft tätigten und die Aktivität „Arbeit“, die Männer eineinhalb Mal so oft durchführten. Auch in den

Kategorien „Arbeitsbezogen“ und „Schule / Ausbildung“ werden von den männlichen Umfrageteilnehmern wesentlich öfter Aktivitäten durchgeführt, jedoch sind die Aktivitätshäufigkeiten zu gering.

Abbildung 8.3 : Aktivitätshäufigkeit in allen Kategorien, nach Geschlecht gegliedert

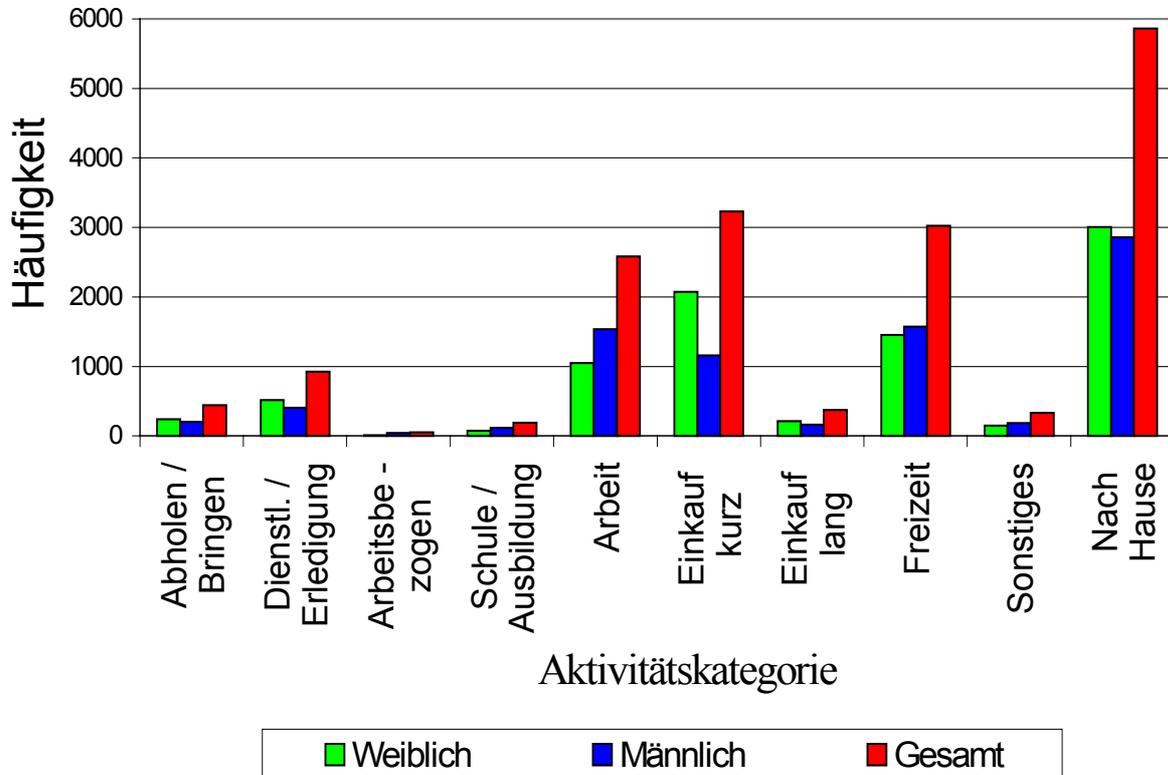


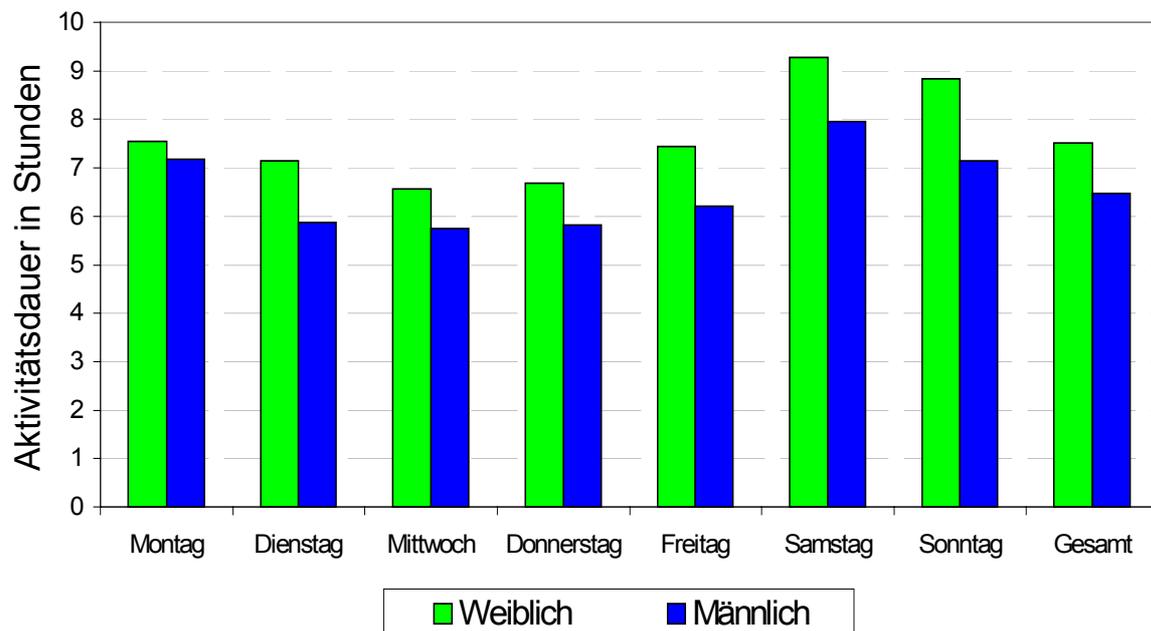
Tabelle 8.7 : Aktivitätshäufigkeit in allen Kategorien, nach Geschlecht gegliedert

Aktivitätskategorie	Weiblich	Männlich	Gesamt
Arbeitsbezogen	8	43	51
Schule / Ausbildung	73	112	185
Sonstiges	149	183	332
Einkauf lang	214	158	372
Bringen / Abholen	238	203	441
Dienstleist./ Erledigung	515	407	922
Arbeit	1047	1534	2581
Freizeit	1452	1573	3025
Einkauf kurz	2072	1159	3231
Nach Hause	3007	2856	5863

## 8.4 AKTIVITÄTSDAUER UND -VERTEILUNG NACH WOCHENTAGEN

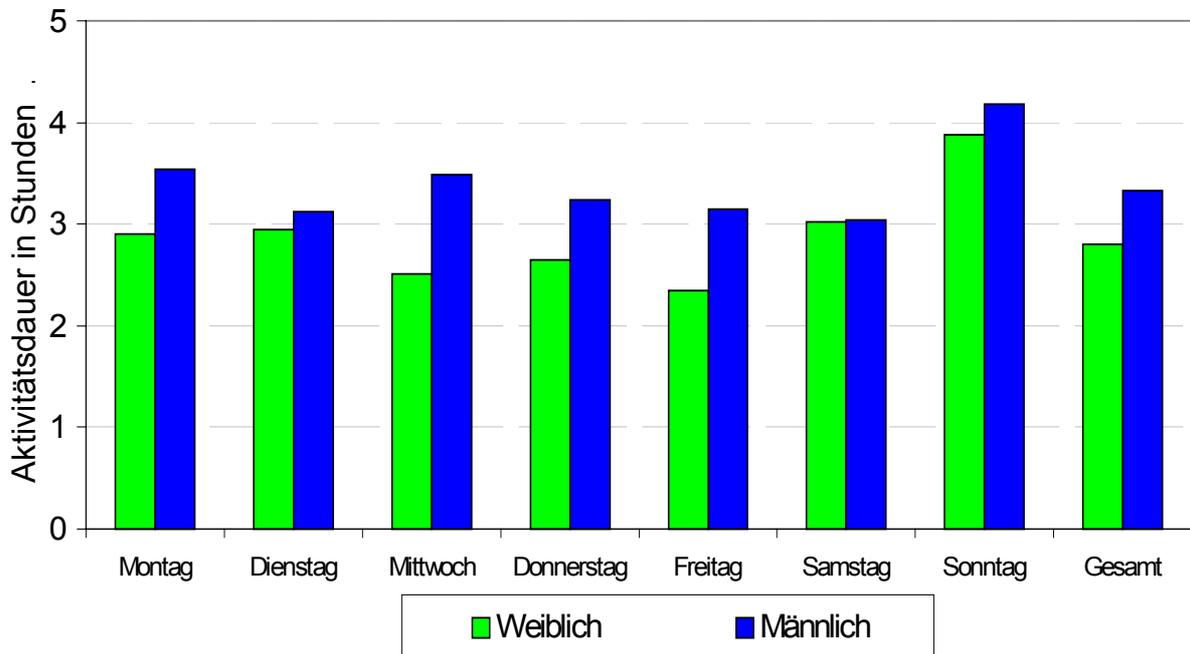
In den folgenden Abbildungen 8.4 bis 8.8 ist die Verteilung aller Aktivitäten bzw. der Aktivitäten außer Haus auf die einzelnen Wochentage dargestellt. Sowohl die Verteilung der Aktivitätsdauern als auch der Aktivitätshäufigkeiten können eine sehr gute Informationsbasis für die Erstellung von Verkehrsverhaltensmustern der befragten Personen bieten. In den folgenden Abbildungen und Tabellen sind jeweils die durchschnittlichen täglichen Aktivitätsdauern angegeben.

Abbildung 8.4 : Durchschnittliche Aktivitätsdauern nach Wochentagen



Die in der Abbildung 8.4 dargestellte Verteilung der durchschnittlichen täglichen Aktivitätsdauern aller Aktivitäten zeigt einige interessante Details. Die Kurve der Aktivitätsdauer zeigt einen Tiefpunkt in der Wochenmitte und danach einen kontinuierlichen Anstieg bis zum Wochenende. An Samstagen liegt mit durchschnittlich über 8,5 Stunden Gesamtaktivitätsdauer der höchste Wert vor, gefolgt vom Sonntag mit 8 Stunden, am Mittwoch hingegen haben die Befragungsteilnehmer im Schnitt mit knapp über 6 Stunden am wenigsten Zeit für diverse Aktivitäten. Verblüffend dabei ist jedoch die Tatsache, daß gerade an Mittwochen praktisch gleich lange Aktivitäten der Freizeitkategorie gefördert wird, wie an Samstagen oder Sonntagen, hingegen am Freitag extrem wenig Zeit für Freizeitaktivitäten aufgewendet wird. (Die Häufigkeitsverteilung sieht noch einmal anders aus.) Den Ausschlag für die wesentlich größeren Aktivitätsdauern an Wochenenden geben jedoch die gravierend höheren Aktivitätsdauern in der Aktivität „Zu / Nach Hause“ an Wochenenden (ca. 18 Stunden) im Gegensatz zu den anderen Tagen (zwischen 12,76 und 15 Stunden).

Abbildung 8.5 : Durchschnittliche Aktivitätsdauern außer Haus nach Wochentagen



Sieht man sich nur die Aktivitäten außer Haus im Schnitt an, so ergibt sich ein etwas anderes Bild. Die Spitzen am Samstag sind lange nicht mehr so groß bzw. liegen unter dem Wochenschnitt. Die längsten durchschnittlichen Aktivitätsdauern liegen nun mit großem Abstand mit ca. 4 Stunden am Sonntag vor, danach fällt die Dauerkurve stetig im Laufe der Woche, bis am Freitag mit knapp  $2\frac{3}{4}$  Stunden die niedrigste Aktivitätsdauer vorliegt. Die Verteilung der Dauer der Aktivität „Arbeit“ hat hierbei gar keinen Einfluß, da sie ziemlich gleich über die Wochentage verteilt ist, an Werktagen im Schnitt ca. 5 Stunden 20 Minuten, an Wochenenden knapp unter 5 Stunden. Bei der Aktivität „Kurzfristiger Einkauf“ ist generell kein eindeutiger Trend feststellbar. Die durchschnittlichen Aktivitätsdauern bewegen sich an den Wochentagen zwischen 17 und 21 Minuten, an Samstagen im Schnitt bei 25 Minuten und an Sonntagen bei lediglich 10 Minuten.

Ein interessantes Detail stellt auch die unterschiedliche Verteilung der Aktivitätsdauern für männliche und weibliche Umfrageteilnehmer dar, wenn man einerseits alle Aktivitäten und andererseits nur die Aktivitäten außer Haus betrachtet. Sind die durchschnittlichen Aktivitätsdauern bei allen Aktivitäten, die auch die Aktivität „zu Hause sein“ beinhalten, bei den Männern noch um ca. eine Stunde kürzer (zwischen 23 Minuten am Montag und 100 Minuten am Sonntag), so sieht man den gegenteiligen Effekt, wenn man nur die Dauer der Aktivitäten außer Haus betrachtet. Hier dauern die Aktivitäten der Männer um rund eine halbe Stunde länger an (zwischen einer Stunde am Mittwoch und 2 Minuten am Samstag).

Tabelle 8.8 : Häufigkeitsverteilung aller Aktivitäten nach Wochentagen

	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag	Sonntag
Weiblich	1310	1357	1398	1401	1427	1147	735
Männlich	1177	1297	1187	1380	1359	1111	717
Gesamt	2487	2654	2585	2781	2786	2258	1452

Tabelle 8.9 : Anzahl der Aktivitäten pro Person nach Wochentagen

	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag	Sonntag
Weiblich	3,23	3,35	3,45	3,46	3,52	2,83	1,81
Männlich	3,74	4,12	3,77	4,38	4,31	3,53	2,28
Gesamt	3,45	3,69	3,59	3,86	3,87	3,14	2,02

Sieht man sich die in Tabelle 8.8 dargestellte Häufigkeitsverteilung aller Aktivitäten der 144 Personen genauer an, fällt auf, daß am Sonntag extrem wenige Aktivitäten (1452) durchgeführt wurden und auch der Samstag mit 2258 Aktivitäten unter dem Durchschnitt von ca. 2450 liegt. Ähnlich hohe Aktivitätshäufigkeiten (beide knapp unter 2800) weisen Donnerstag und Freitag auf, die Werte von Montag bis Mittwoch liegen knapp über dem Durchschnitt. Ähnlich liegt auch bei der Häufigkeitsverteilung aller Aktivitäten außer Haus (siehe Tabelle 8.10) der Freitag vorne (1850 Aktivitäten), eng gefolgt von den restlichen Wochentagen und mit gewissem Abstand folgen Samstag (1436) und abgeschlagen der Sonntag mit nur 894 Aktivitäten außer Haus.

Die Abbildungen 8.6 und 8.7 bzw. Tabellen 8.9 und 8.11 zeigen die durchschnittliche Anzahl der Aktivitäten pro Wochentag, einerseits für alle Aktivitäten, andererseits nur für die Aktivitäten außer Haus. Hier ist jedoch im Gegensatz zu den gesamten Aktivitäten eine wesentlich höhere Anzahl von Aktivitäten bei den männlichen Umfrageteilnehmern zu sehen. Von der Verteilung auf die Wochentage ändert sich natürlich nichts zu den Tabellen 8.8. und 8.10, zwischen den Geschlechtern treten jedoch teils gravierende Unterschiede auf. Im Durchschnitt machten die männlichen Umfrageteilnehmer ca. 3,7 Aktivitäten (ca. 2,4 außer Haus) pro Tag, die Frauen mit 3,1 Aktivitäten (ca. 2 außer Haus) pro Tag durchschnittlich rund eine halbe Aktivität weniger. Die größten Unterschiede zwischen den beiden Geschlechtern ergeben sich an Donnerstagen und Freitagen, an denen Männer durchschnittlich 4,4 bzw. 4,3 Aktivitäten durchführen, Frauen jedoch nur jeweils 3,5, die geringsten Unterschiede am Mittwoch mit durchschnittlich 3,8 Aktivitäten außer Haus bei Männern und 3,5 bei Frauen.

Abbildung 8.6 : Anzahl der Aktivitäten pro Person nach Wochentagen

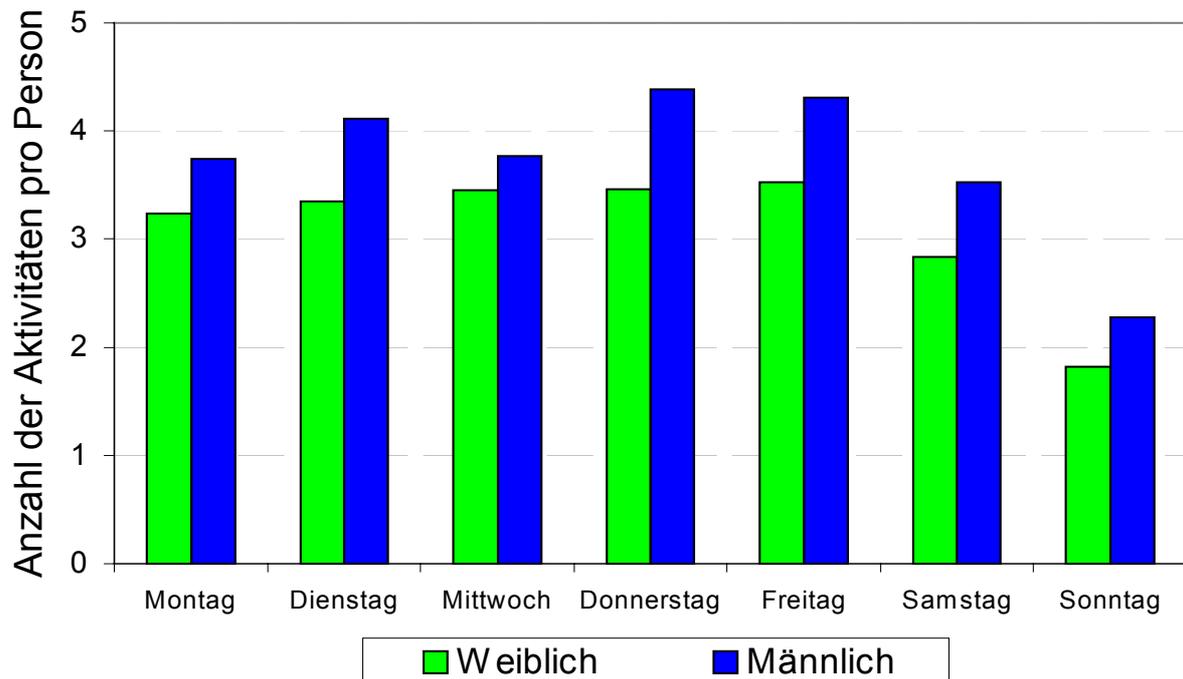


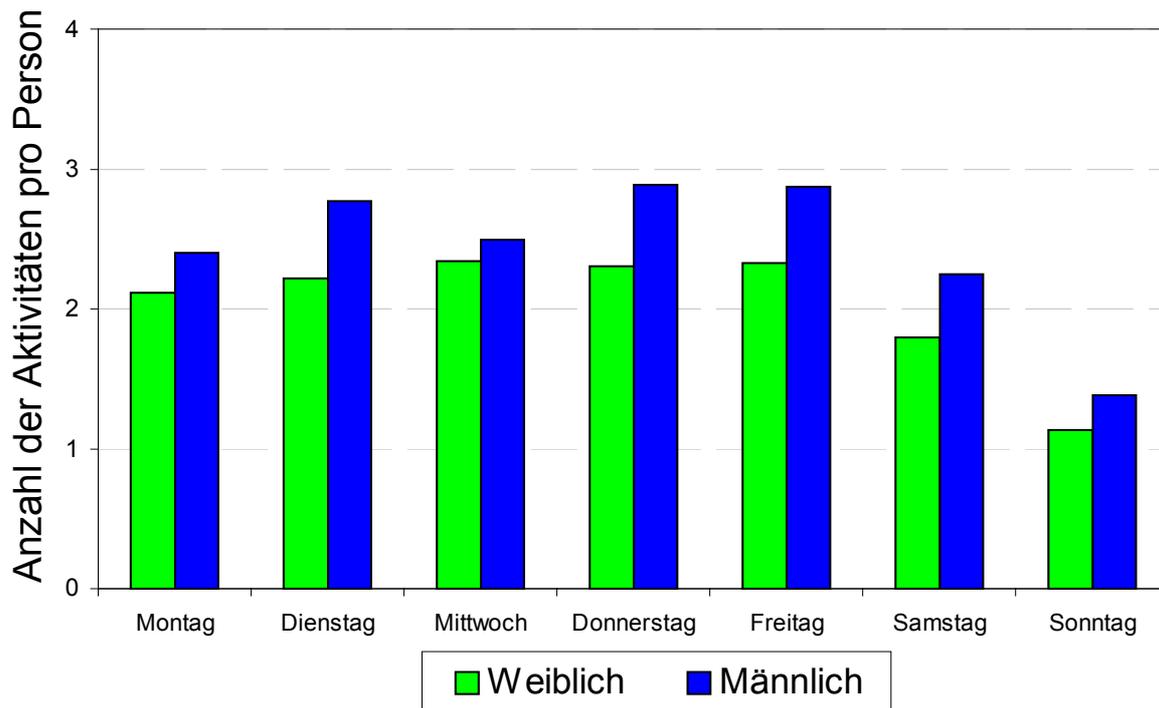
Tabelle 8.10 : Häufigkeitsverteilung aller Aktivitäten außer Haus nach Wochentagen

	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag	Sonntag
Weiblich	857	899	948	933	944	728	459
Männlich	757	872	785	909	906	708	435
Gesamt	1614	1771	1733	1842	1850	1436	894

Tabelle 8.11 : Anzahl der Aktivitäten außer Haus pro Person nach Wochentagen

	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag	Sonntag
Weiblich	2,12	2,22	2,34	2,30	2,33	1,80	1,13
Männlich	2,40	2,77	2,49	2,89	2,88	2,25	1,38
Gesamt	2,24	2,46	2,41	2,56	2,57	1,99	1,24

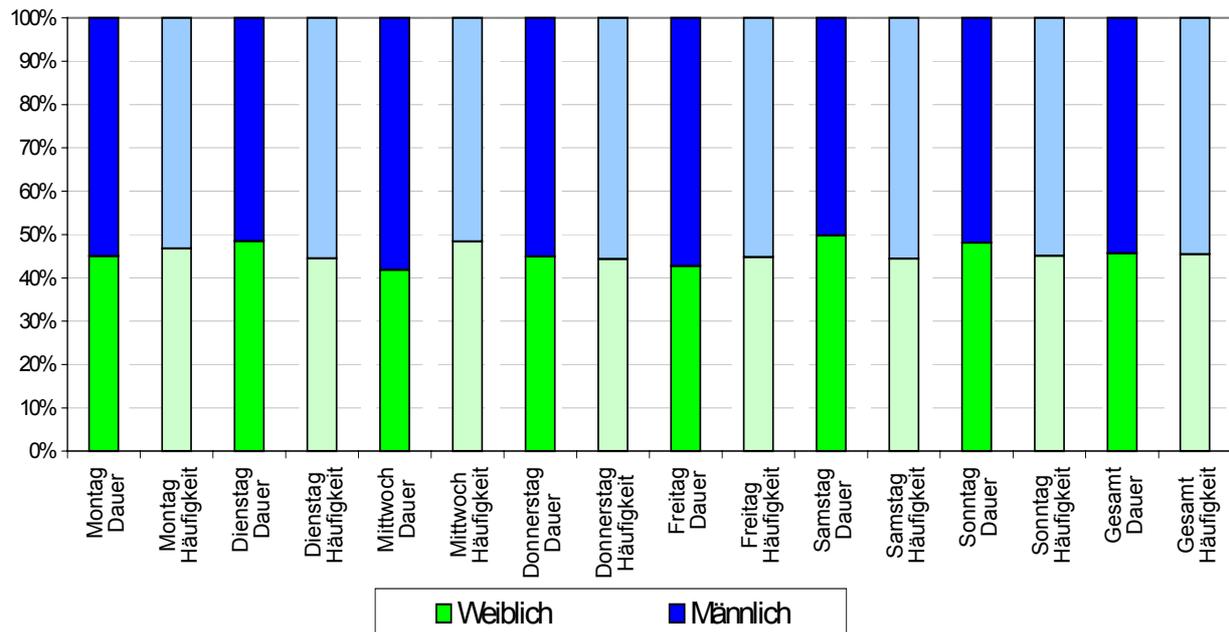
Abbildung 8.7 : Häufigkeitsverteilung aller Aktivitäten außer Haus nach Wochentagen



Die Häufigkeit der Freizeitaktivitäten (siehe Tabellen und Diagramme im Anhang XVI) nimmt jedoch ebenso wie die Kurve in der Abbildung 8.4 vom geringsten Wert am Mittwoch (322 Freizeitaktivitäten) bis zum Samstag (550) zu und dann wieder von Sonntag (541) bis Mittwoch stetig ab. In der Kategorie „Arbeit“ liegen die Häufigkeiten am Dienstag, Mittwoch und Donnerstag bei über 500, am Montag und Freitag um die 400 und am Wochenende mit nur 132 bzw. 73 durchgeführten Arbeitsaktivitäten sehr abgeschlagen am Ende. Die Häufigkeiten der Kategorie „kurzfristiger Einkauf“ steigen von Montag (463) stetig bis Freitag (571), sind am Samstag mit einer Anzahl von 521 noch recht hoch und sacken am Sonntag auf einen Wert von 168 ab.

In der Abbildung 8.8 sind noch einmal die Unterschiede zwischen den Geschlechtern für alle Aktivitäten außer Haus prozentuell grafisch gegenüber gestellt und mit Zahlenangaben in der Datentabelle ergänzt. Die Angaben der Aktivitätsdauer sind hier ausnahmsweise in Minuten. Dabei sind die insgesamt eine halbe Stunde kürzere Dauer der Aktivitäten außer Haus bei Frauen und die geringfügig höhere Anzahl von Aktivitäten außer Haus (52 zu 48%) ersichtlich.

Abbildung 8.8 : Dauer und Häufigkeit aller Aktivitäten außer Haus pro Person pro Tag nach Wochentagen und Geschlecht



## 8.5 AKTIVITÄTSDAUERVERTEILUNG AUF ARBEITSTAGE UND FREIE TAGE

Im Kapitel 8.2.3 wurden gesamt 23 Arbeitstage und 12 freie Tage (Samstage und Sonntage, sowie Karfreitag und Ostermontag) definiert. Die Abbildungen 8.9 bis 8.12 zeigen nun die wochentag- und wochenendspezifischen Merkmale der Aktivitätsverteilung in bezug auf Dauer und Häufigkeit.

Die durchschnittliche Aktivitätsdauer ist, wie aus Abbildung 8.9 und Tabelle 8.12 ersichtlich, an freien Tagen in den Kategorien „Arbeitsbezogen“, „Schule / Ausbildung“, „Freizeit“ (jeweils ca. 15 Minuten) geringfügig und in der Aktivitätskategorie „Zu Hause“ (ca. 4,5 Stunden) um einiges höher als an Arbeitstagen. Allein Aktivitäten der Kategorie „Arbeit“ dauerten im Schnitt an Arbeitstagen um eine halbe Stunde länger als an freien Tagen. Die restlichen Aktivitätskategorien weisen eine ziemlich ausgeglichene Dauer pro Aktivität an Arbeitstagen bzw. freien Tagen auf.

Abbildung 8.9 : Durchschnittliche Dauer der Aktivität an Arbeitstagen und freien Tagen

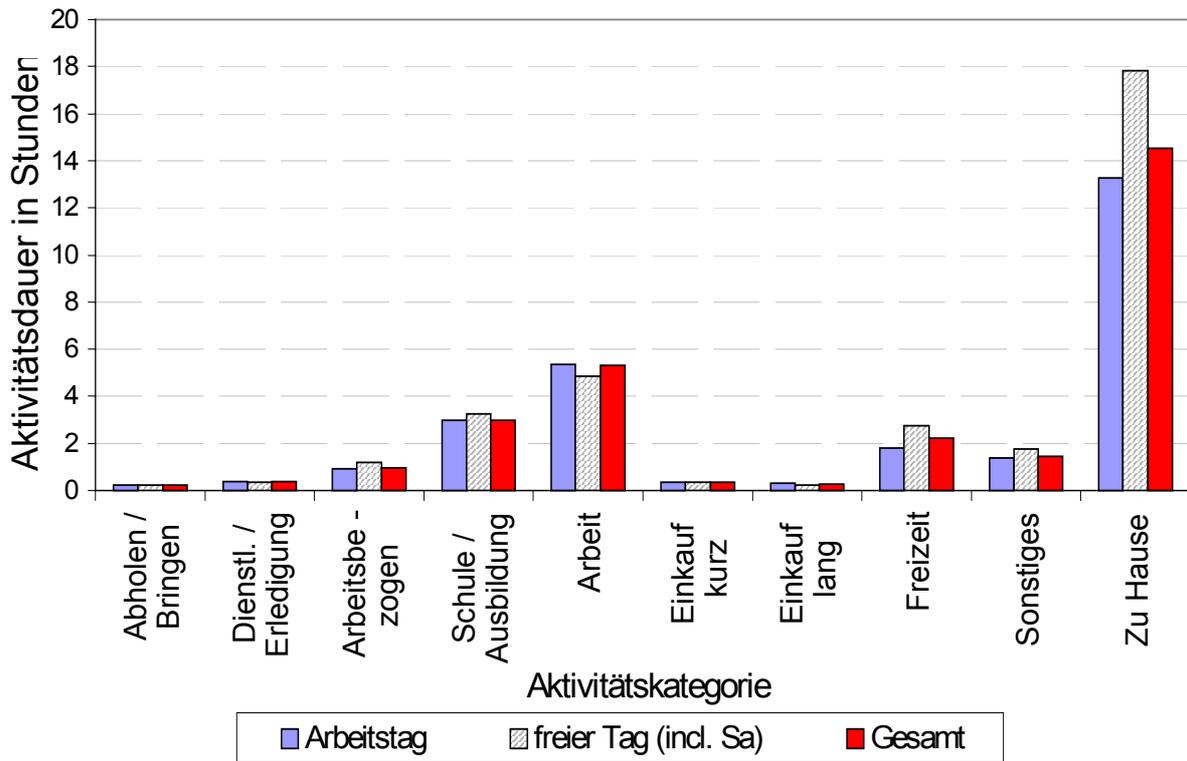
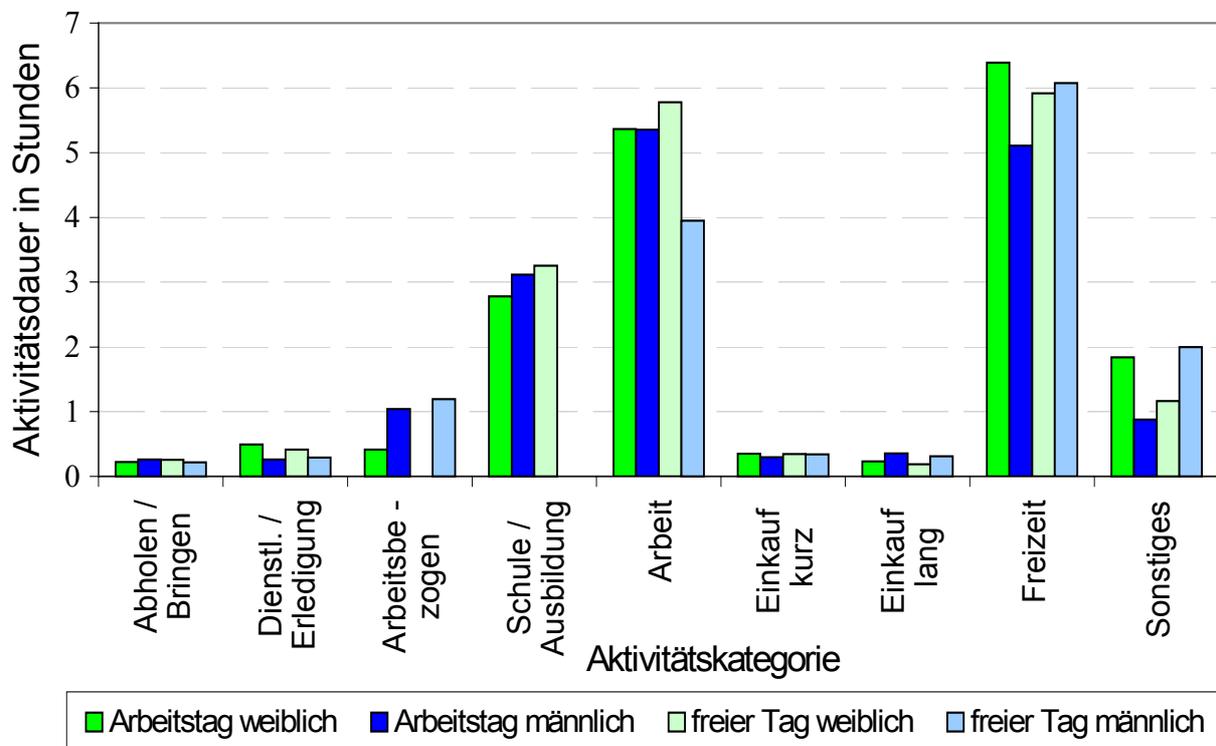


Tabelle 8.12 : Durchschnittliche Aktivitätsdauer pro Aktivität nach Kategorien, Geschlecht und Arbeitstag (Samstag frei)

	Arbeitstag			Freier Tag			Gesamt
	Weiblich	Männlich	Gesamt	Weiblich	Männlich	Gesamt	
Zu Hause	15,30	11,13	13,27	19,85	15,73	17,81	14,52
Freizeit	6,39	5,11	5,71	5,92	6,07	5,99	5,83
Arbeit	5,36	5,35	5,36	5,78	3,95	4,86	5,31
Schule/Ausbildung	2,78	3,12	2,99	3,25	0,00	3,25	3,00
Sonstiges	1,84	0,88	1,36	1,16	1,99	1,75	1,46
Arbeitsbezogen	0,41	1,05	0,93	0,00	1,19	1,19	0,97
Dienstleist./Erledigung	0,49	0,27	0,39	0,42	0,30	0,36	0,39
Einkauf kurz	0,35	0,30	0,33	0,35	0,35	0,35	0,34
Einkauf lang	0,24	0,36	0,29	0,19	0,31	0,25	0,28
Abholen/Bringen	0,22	0,26	0,24	0,26	0,22	0,23	0,24

Sieht man sich aber die geschlechtsspezifischen Unterschiede in Abbildung 8.10 an, so sieht man, daß zwar in der Kategorie „Arbeit“ gesamt an Arbeitstagen um eine halbe Stunde länger pro Aktivität gearbeitet wurde, Frauen aber an Wochenenden pro Arbeitsaktivität noch länger als an Werktagen (20 Minuten) und vor allem ca. 2 Stunden länger als Männer arbeiten. Jedoch liegen die Häufigkeiten der Arbeitsaktivität am Wochenende insgesamt bei knapp über 100, an Werktagen hingegen für Frauen bei 934 bzw. bei 1419 für Männer. Daher sollte nicht der Schluß gezogen werden, daß an freien Tagen überdurchschnittlich viel und lange gearbeitet wird, es liegen nur die Aktivitätsdauern pro wirklich durchgeführter Aktivität höher. Die tatsächlichen durchschnittlichen Aktivitätsdauern pro Person und Tag sind in den Tabellen 8.13 und 8.14 zu sehen.

Abbildung 8.10 : Aktivitätsdauer nach Kategorien außer Haus, Geschlecht und Arbeitstag, Sa = frei



Ebenfalls gravierende Unterschiede zwischen den Geschlechtern weisen die Aktivitätsdauern in den Kategorien „Arbeitsbezogen“, „Schule / Ausbildung“ und „Freizeit“ auf. Frauen haben am Wochenende keinerlei arbeitsbezogene Aktivitäten unternommen, daher liegen auch keine Aktivitätsdauern vor. Keiner der befragten Männer hat im Gegensatz dazu am Wochenende eine schulische Ausbildung in Anspruch genommen. Die durchschnittliche Dauer der Freizeitaktivitäten ist bei Frauen an Werktagen höher als an Wochenenden, bei Männern und der Gesamtstudie jedoch umgekehrt.

Tabelle 8.13 : Durchschnittliche Aktivitätsdauer pro Person pro Tag an Werktagen

Aktivitätskategorie	Weiblich	Männlich	Gesamt
Zu Hause	18,01	15,84	17,06
Arbeit	2,69	5,24	3,81
Freizeit	2,79	3,27	3,00
Einkauf kurz	0,31	0,17	0,25
Schule / Ausbildung	0,10	0,24	0,16
Sonstiges	0,12	0,08	0,10
Dienstleistungen / Erledigung	0,12	0,06	0,09
Abholen / Bringen	0,03	0,03	0,03
Einkauf lang	0,02	0,03	0,03
Arbeitsbezogen	0,00	0,03	0,01

Tabelle 8.14 : Durchschnittliche Aktivitätsdauer pro Person pro Tag an freien Tagen

Aktivitätskategorie	Weiblich	Männlich	Gesamt
Zu Hause	16,62	19,61	17,93
Arbeit	0,67	0,60	0,64
Freizeit	3,88	5,18	4,45
Einkauf kurz	0,15	0,14	0,15
Schule / Ausbildung	0,02	0,00	0,01
Sonstiges	0,03	0,16	0,08
Dienstleistungen / Erledigung	0,03	0,03	0,03
Abholen / Bringen	0,01	0,02	0,01
Einkauf lang	0,01	0,01	0,01
Arbeitsbezogen	0,00	0,01	0,01

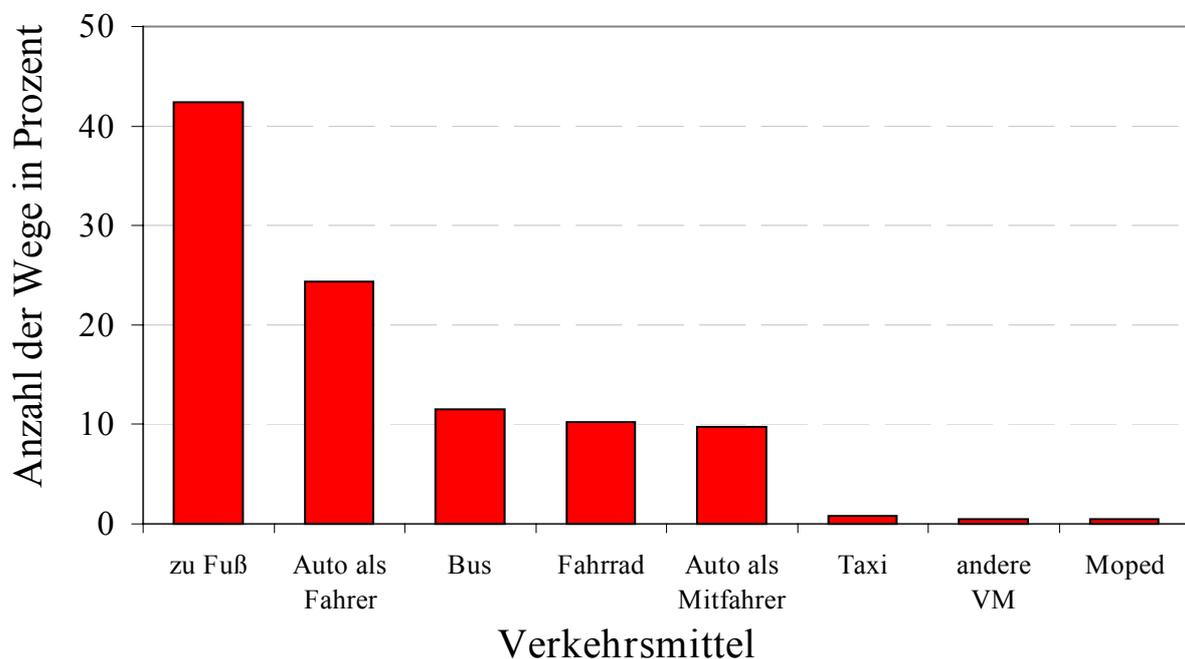
In den Tabellen 8.13 und 8.14 wird der Unterschied bei der durchschnittlichen Aktivitätsdauer pro Tag deutlich gemacht. Die auffälligsten Differenzen liegen sicherlich bei der Dauer von Arbeits- und Ausbildungsaktivitäten vor, die pro Werktag um ein vielfaches über dem Wert an freien Tagen liegen (Gesamt sind dies 3,81 Stunden an Werktagen und nur 0,64 Stunden an freien Tagen bei der Arbeit bzw. 10 Minuten durchschnittliche Ausbildungsdauer pro Person pro Werktagen und nicht einmal eine Minute im Mittel pro Person an freien Tagen). Die durchschnittliche Dauer ist einzig in der Aktivität Freizeit an freien Tagen länger, dort erhöht sich der Wert von drei Stunden auf knapp viereinhalb Stunden pro Tag pro Person.

## 8.6 WEITERE ERGEBNISSE DER DESKRIPTIVEN ANALYSE

Zusätzlich zu den oben beschriebenen Verteilungen von Aktivitäten und Haushaltsmerkmalen wurden noch einige interessante Charakteristika für die später folgenden Hazardmodelle erarbeitet. So ist beispielsweise die Verkehrsmittelverteilung oder die Aktivitätsverteilung nach einer bestimmten Verkehrsmittelbenützung ein aussagekräftiges Detail für Verkehrsverhaltensmuster.

In den Abbildungen 8.11 und 8.12 sieht man die Verteilung der Wege auf die einzelnen Verkehrsmittelkategorien, deren Dauer bzw. geschlechtsspezifische Unterschiede. Fast die Hälfte aller Wege wurden von den befragten Personen zu Fuß zurückgelegt, rund ein Viertel als Autofahrer und jeweils ca. 10 % als Automitfahrer, Busbenützer bzw. mit dem Fahrrad. Eine sehr untergeordnete Rolle bei der Verkehrsmittelwahl spielen Taxi, Moped und auch andere Verkehrsmittel wie Eisenbahn, Überlandbusse und Flugzeug in der täglichen Wahl.

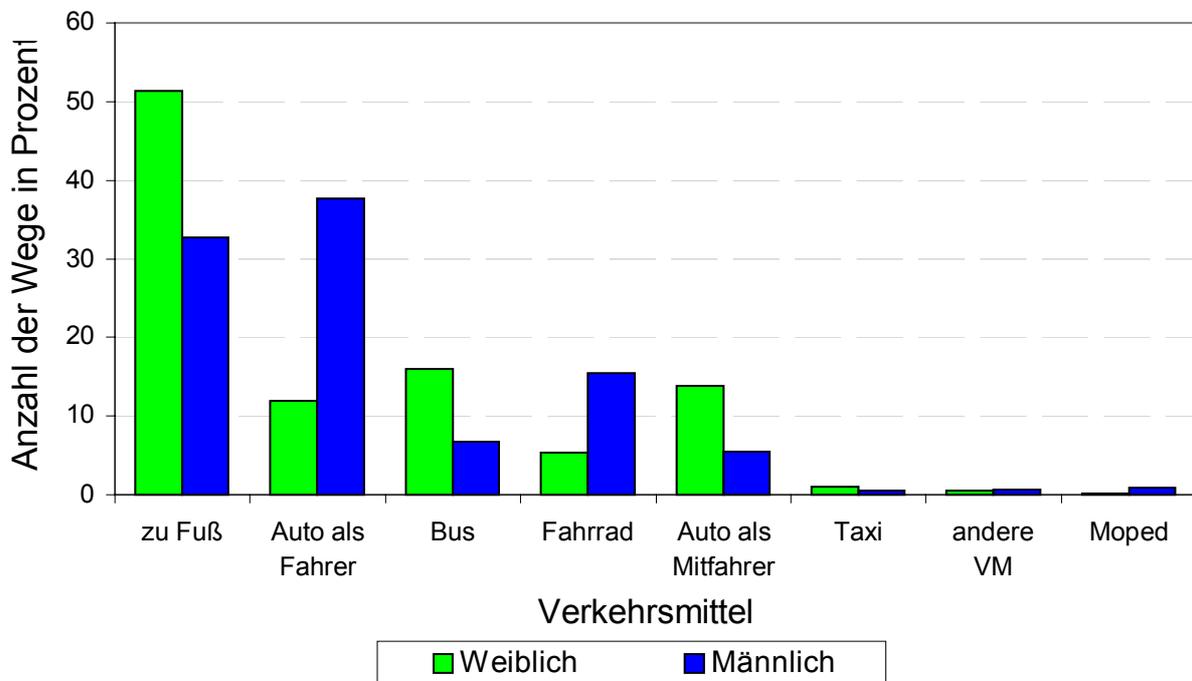
Abbildung 8.11 : Verkehrsmittelwahl nach Kategorien



Bei den geschlechtsspezifischen Unterschieden sind vor allem der überdurchschnittlich hohe Anteil von männlichen Autofahrern (75% aller Wege als Autolenker wurden von Männern zurückgelegt), Mopedbenützern (88%) und Fahrradfahrern (72%), sowie der hohe Anteil von weiblichen Fußgängern

(63%), PKW-Mitfahrern (73%) und Busbenutzern (72%) auffällig. Nur in den Kategorien Taxi und andere Verkehrsmittel liegt eine ausgeglichene Bilanz vor.

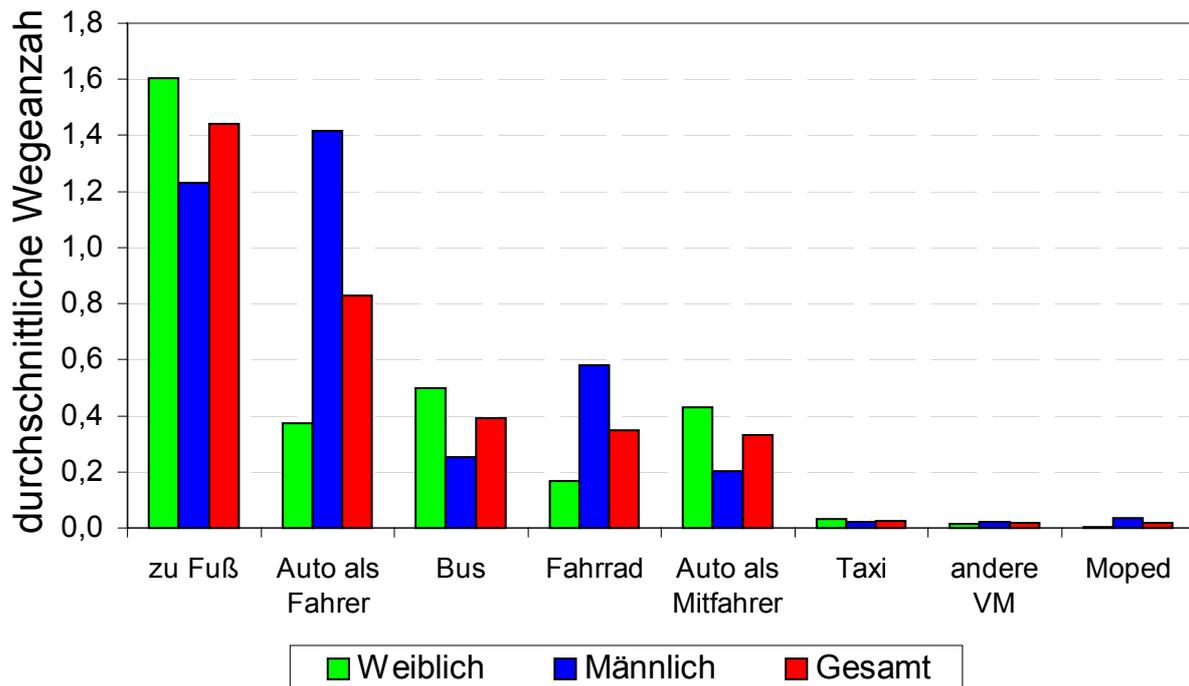
Abbildung 8.12 : Verkehrsmittelwahl nach Geschlecht



In Abbildung 8.12 sieht man, daß Frauen mehr als die Hälfte all ihrer Wege zu Fuß zurücklegen und auch die Häufigkeiten von Busbenützung (16%) und Mitfahrer (14%) häufiger als die der eigenen Autobenützung (ca. 12%) sind. Bei Männern hingegen überwiegen mit knapp 38% die Wege als Autofahrer, gefolgt von knapp 32% Wegen zu Fuß und ca. 15% per Fahrrad. Bus und Mitfahrgelegenheiten werden jedoch nur jeweils zu etwa 6% in Anspruch genommen.

Abbildung 8.13 bestätigt die äußerst geringe Anzahl von Wegen als Autofahrer bei Frauen, durchschnittlich werden nur knapp 0.4 Wege pro Tag bzw. nicht einmal 3 pro Woche zurückgelegt. Männer machen hingegen pro Tag im Durchschnitt einen Weg mehr als Frauen (1.4) und kommen so auf etwa 10 Wege als Autofahrer pro Woche. Sie legen damit als Autofahrer pro Woche rund einen Weg mehr als zu Fuß (8.6 Wege) zurück.

Abbildung 8.13 : Durchschnittliche Anzahl von Wegen pro Tag und Person nach Verkehrsmitteln



Ein weiterer interessanter Aspekt für die Analyse der Verhaltensmuster ist die Verteilung der Aktivitäten nach der Benützung eines bestimmten Verkehrsmittels und die Dauer dieser Aktivitäten. Da dies jedoch eine sehr detaillierte Auswertung beinhaltet und viele Tabellen und Diagramme gezeigt werden müssten, sind diese nur im Anhang angeführt. Einige interessante oder unerwartete Aspekte werden hier kurz beschrieben.

Die prinzipiell längsten Aktivitätsdauern folgen auf die Benützung der Verkehrsmittel Moped, Bus und „zu Fuß“, nicht nur in den Kategorien Freizeit und Einkauf, sondern beispielsweise auch in der Kategorie Arbeit (mit Ausnahme des Verkehrsmittels „zu Fuß“, denn nach dessen Benützung ist die Arbeitsdauer weit unter dem Durchschnitt, da der Weg von und zur Mittagspause meist ein Fußweg ist und dadurch die Arbeitszeit sich nur mehr über den Nachmittag erstreckt). Zu erklären sind diese großen Aktivitätsdauern meist mit einer größeren Zeitreserve der betreffenden Personen bzw. der größeren Uneingeschränktheit in bezug auf den nach Hägarstrand definierten Zeitfaktor bei der Verkehrsmittelwahl und Verkehrswegwahl (siehe Kapitel 4.4). Die längsten Aktivitätsdauern der Kategorien Einkauf kurz und Einkauf lang treten nach Benützung des Autos sowohl als Fahrer als auch als Mitfahrer auf. Nach der Benützung von Mopeds scheinen lange Aktivitätsdauern in den Kategorien Arbeit, Ausbildung / Schule und Dienstleistungen / Erledigungen bei Männern und Sonstiges bei Frauen auf, jedoch spielt die Häufigkeit der Mopedbenützung eine sehr untergeordnete Rolle. Nach der Be-

nützung anderer Verkehrsmittel, die Zug, Überlandbusse und Flugzeuge inkludieren, treten in zwei Kategorien extrem lange Aktivitätsdauern auf. Diese überdurchschnittlichen Dauern bei Arbeit und Freizeit sind meist der großen Entfernung zum Heimatort und dem damit verbundenen seltenen Besuch dieser Aktivitätsorte zuzuschreiben.

Die häufigste Aktivität nach dem Verkehrsmittel „zu Fuß“ ist der kurze Einkauf mit knapp 2000 Aktivitäten, vor allem bei Frauen (1464 Wege), gefolgt von Freizeit (gesamt 1301) und Arbeit (gesamt 739). Nach der Benützung von Fahrrad, Moped oder anderen Verkehrsmitteln ist die häufigste Aktivität Arbeit, meist knapp gefolgt von Freizeit. Nach der Benützung von Bus oder Auto als Fahrer folgen in ziemlich ähnlicher Anzahl Aktivitäten der Kategorien Arbeit, Einkauf kurz oder Freizeit, jedoch mit dem starken geschlechtsspezifischen Unterschied, daß der Bus in all diesen Aktivitätskategorien in weit größerem Maße von Frauen, das Auto als Fahrer stark von Männern in Anspruch genommen wird. Extrem häufig ist vor allem bei Frauen auch die Benützung des Autos als Mitfahrer vor Freizeitaktivitäten, das Verkehrsmittel Taxi wird am häufigsten, wenngleich auch nicht sehr oft, für Freizeit oder Erledigungen genutzt. Die Aktivität Jemanden Bringen / Abholen wurde verständlicherweise am öftesten mit den Verkehrsmitteln Auto als Fahrer (333) bzw. Mitfahrer (62) durchgeführt.

Alle Tabellen und Diagramme dazu sind in den Anhängen XV bis XVIII: Deskriptive Analysen der Aktivitätsdauern zu sehen.

## **9 HAZARDMODELL - ANALYSE**

### **9.1 EINLEITUNG**

In diesem Kapitel wird die Modellschätzung der Uppsala Daten mit Hilfe von Hazard Modellen beschrieben. Die dazu nötigen Umstrukturierungen und Ergänzungen des Datensatzes wurden schon im Kapitel 2.5 eingehend behandelt. Nach der Definition von abhängigen und unabhängigen Variablen werden Modellschätzungen mit Hilfe des Statistikprogrammes SAS (siehe Kapitel 6) gemacht, um die Verteilung der Aktivitätsdauern einerseits und der Dauer von Intervallen zwischen gleichartigen Aktivitätskategorien andererseits zu charakterisieren. Diese Modelle werden dann mit Hilfe von statistischen Auswertungen auf ihre Paßform überprüft und anschließend die Resultate der einzelnen Modelle kurz diskutiert.

### **9.2 DAUER DER AKTIVITÄTEN**

#### **9.2.1 Einleitung**

Mit Hilfe der Informationen aus den Rohdaten bzw. den nach gewissen Hierarchien und Korrekturkriterien veränderten Rohdaten der UTHS (siehe Kapitel 2) werden die Dauern der einzelnen Aktivitäten berechnet und ebenso die Dauern der Intervalle zwischen gleichartigen Aktivitäten festgestellt (dazu mehr im Kapitel 9.3). In diesem Kapitel wird die Aktivitätsdauer (in Stunden) als jeweilige abhängige Variable definiert und eine Reihe erklärender Variablen, sowohl soziodemografischer als auch aktivitätsbezogener Natur ermittelt, die einen Einfluß auf die abhängige Variable haben könnten. Natürlich ist es eher sinnlos, alle vorhandenen Variablen in die Berechnungen miteinfließen zu lassen. Daher wird mit Hilfe gewisser Hierarchien und Entscheidungskriterien (beispielsweise einer Korrelationsmatrix) eine Auswahl getroffen, die nur noch interessante mögliche Einflußvariablen beinhaltet.

Es werden nun folgend alle drei verschiedenen Schritte der Modellschätzung (nonparametrisch, semiparametrisch und vollparametrisch), die mit Hilfe der SAS-Programme durchgeführt worden sind, vorgestellt und diskutiert.

## 9.2.2 Nonparametrische Modelle mit *PROC LIFETEST*

Wie schon im Kapitel 6.2 beschrieben, dienen nonparametrische Modelle vorrangig dazu, die Form der Survivalkurven und Hazardkurven möglichst wahrheitsgetreu abzubilden und zusätzlich eventuelle Unterschiede zwischen einzelnen Gruppen mit ähnlichen oder gleichen beschreibenden Variablen herauszufinden. Die Auswertungen mit *PROC LIFETEST* sind aus diesem Grunde auch eher kurz gehalten und beschränken sich auf Abbildung und Diskussion der Hazard- und Survivalkurven, eine kurze Abhandlung der Gruppenunterschiede und einen sogenannten *Log-Rank Test with Forward Stepwise Sequence of Chi-Squares* von möglichen Einflußvariablen. Detailliertere Variablenüberprüfungen werden bei den semiparametrischen bzw. vollparametrischen Modellen durchgeführt.

Zu Beginn werden für die Gesamtdaten bzw. alle Aktivitätskategorien Survival- und Hazardkurven erstellt und Unterschiede zwischen verschiedenen Alters- und Geschlechtsgruppen aufgezeigt. Danach werden für all jene Aktivitätskategorien, die für spätere Auswertungen ausgewählt werden, einzeln die entsprechenden Kurven erstellt und verglichen. Es sind dies die Kategorien Arbeit, Einkauf langfristig, Einkauf kurzfristig, Freizeit, Dienstleistung allgemein, persönliche Erledigungen wie Post, Bank, etc., Treffen mit Freunden, Besuch bei Verwandten, Verein / Bürgerinitiative / Kirche, Ausübung von Sport, Spaziergang / Spazierfahrt, Etwas Neues sehen, Lokal / Kino / Theater. Für jede Aktivitätskategorie sehen Sie die entsprechenden Survival- bzw. Hazardkurven in den Abbildungen 9.2 bis 9.14 .

### 9.2.2.1 Gesamte Aktivitäten

Die Hazard- und Survivalkurve für den gesamten Datensatz sehen Sie in Abbildung 9.1 bzw. detaillierter im Anhang V.

Die Survivalkurve fällt leicht exponentiell ab, die Hazardkurve verläuft nach einem anfänglichen Startwert von ca. eins bis zu einer Aktivitätsdauer von knapp über 20 Stunden relativ konstant und steigt erst bei jenen Dauern, die nicht mehr sonderlich oft vorkommen, stärker an. Für den gesamten Datensatz ist also auch eine exponentielle Verteilung der Wahrscheinlichkeitsdichtefunktion zu erwarten.

Für die Gesamtaktivitäten werden verschiedene Gruppen mit ähnlichen Charakteristiken miteinander verglichen. Dazu gehören einerseits die verschiedenen Altersschichten und andererseits beispielsweise die beiden Geschlechtsgruppen (Abbildungen im Anhang V). Man könnte noch einige Gruppen mit ähnlichen Charakteristika zusammenfassen und untersuchen, jedoch sind solche Untersuchungen für die einzelnen Aktivitätskategorien wesentlich interessanter und aussagekräftiger.

Bei einer Aktivitätsdauer von ca. 10 Stunden haben alle Altersgruppen einen ähnlichen Survivalwert, das heißt, die Wahrscheinlichkeit, daß eine Aktivität ca. 10 Stunden andauert, ist bei allen Gruppen relativ ähnlich hoch. Bei Aktivitäten, die bis zu 3 Stunden andauern, haben die Personen bis 25 Jahre die höchste Wahrscheinlichkeit, bei längeren Aktivitätsdauern (über 10 Stunden) liegt der Survivalwert der über 70- bzw. 75-jährigen am höchsten. Die relativ aussagekräftigen 50%-Quartilen liegen bei den 60-jährigen bei 2.75, für 30 und 50-jährige jedoch bei nur 1.5 bzw. 1.33 Stunden. Dies bedeutet, daß die Hälfte der Aktivitäten dieser Personengruppen bei einer Dauer von 2.75 (1.5 bzw. 1.33) Stunden abgeschlossen sind.

Ebenso kann man Unterschiede bei den beiden Geschlechtern erkennen. Bei einer Aktivitätsdauer von ca. 0,5 bis 5 Stunden ist der Survivalwert der Männer leicht höher, zwischen 5 und 10 Stunden nahezu identisch und über 10 Stunden ist der Wert für Frauen wesentlich höher. Die beiden Variablen sind mit allen drei Tests (Log-Rank, Wilcoxon und doppelter negativer Loglineartest) statistisch signifikant.

### 9.2.2.2 **Verschiedene Aktivitätskategorien**

Für insgesamt 13 Aktivitätskategorien werden Survival- und Hazardkurven angegeben und verglichen. Für die bei den semiparametrischen Modellen genauer betrachteten vier Kategorien „Einkauf kurz“, „Freizeit allgemein“, „Dienstleistung allgemein“ und „Treffen mit Freunden“ werden danach zusätzlich noch stufenförmige Log-rank-Tests (siehe Kapitel 0) durchgeführt.

Die Survivalkurve in der Kategorie Arbeit (Abbildung 9.2) verläuft im Vergleich zu den meisten anderen Kurven nicht exponentiell. Einem mäßigen Abfall folgt bei einer Aktivitätsdauer von ca. 3 bis 4 Stunden ein steiler Abfall (von 0.8 auf 0.6), dann wieder eine Verflachung, bevor bei ca. 10 Stunden ein erneut steil abfallender Kurvenverlauf sichtbar wird (von 0.2 auf 0.05). Das bedeutet, daß Aktivitäten eine sehr hohe Wahrscheinlichkeit haben, nach ca. 3 bis 4 Stunden bzw. nach 10 Stunden zu enden, sehr kurze Aktivitätszeiten sind hingegen relativ unwahrscheinlich (nur 20% aller Aktivitäten haben eine Dauer unter 3 Stunden). Diese Ergebnisse sind bei der Kategorie Arbeit jedoch nicht weiter verwunderlich. Die hohe Wahrscheinlichkeit einer Arbeitsaktivität von ca. 3 bis 4 Stunden ist nicht vorrangig durch Teilzeitarbeit, sondern durch das Einlegen einer Mittagspause zu erklären. Die Survivalkurve der täglichen Gesamtarbeitszeit würde anders aussehen, jedoch haben die meisten arbeiten-

den Personen in ihrer Mittagspause andere Aktivitäten erledigt (Einkauf, Treffen mit Freunden, etc.) oder haben ihre Mittagspause an einem anderen Ort verbracht (Restaurant, Lokal, etc.) und so ihre tägliche Arbeitszeit unterteilt. Dennoch haben einige Personen entweder ohne Pause gearbeitet oder die Arbeitspausen im selben Gebäude (evtl. Kantine, etc.) verbracht und haben so eine durchgehende, für heutige Verhältnisse oft übermäßig lange Arbeitszeit. Arbeitsaktivitäten über 11 Stunden kommen vor, sind jedoch schon sehr unwahrscheinlich, der Survivalwert liegt unter 0.01. Die Hazardkurve zeigt einen relativ ruhigen und niedrigen Verlauf, relative Spitzen sind bei ca. 4 Stunden und 6,5 Stunden und deutlich ausgeprägte Spitzen bei 10 Stunden und 14 Stunden zu sehen. Wesentlich aussagekräftiger sind jedoch dabei die Spitzen im linken Teil der Zeitachse, da zu diesen Zeitpunkten noch eine große Anzahl an Aktivitäten vorhanden ist. Im rechten Teil - wie aus der Survivalkurve ersichtlich - ist nur noch eine geringe Aktivitätsdichte vorhanden, wodurch die Bedeutung dieser Werte der Hazardkurve stark herabgesetzt wird.

Vergleicht man diese Ergebnisse mit anderen Studien (z.B. Oh, 2000), so sieht man, daß die Hazardkurve hier nicht nur eine deutliche Spitze bei 8 Stunden Arbeitsaktivität, sondern auch eine bei einer Dauer von ca. 4 Stunden aufweist. Der Grund dafür sind eindeutig Pausen, die nicht am Arbeitsplatz selbst gemacht wurden bzw. für andere Erledigungen genutzt wurden und dementsprechend die Dauer der Arbeitszeit unterbrechen. Der Grund, daß die zweite relative Spitze erst bei ca. 10 Stunden erkennbar ist, sind die im Vergleich zur heutigen Zeit wesentlich längeren Arbeitszeiten.

Bei den Kategorien Einkauf langfristig und Einkauf kurzfristig (Abbildung 9.3 und Abbildung 9.4) zeigen die Survivalkurven einen ähnlichen, exponentiellen Verlauf. Der einzig gravierende Unterschied ist die Skalierung der Zeitachse, da die wenigen langfristigen Einkäufe (nur 367 insgesamt) meist nur sehr kurz andauern, die kurzfristigen Einkäufe, also jene des täglichen Bedarfs (gesamt 3221), etwas länger. Nach ca. 30 Minuten sind rund 90 % der langfristigen Einkäufe abgeschlossen, nach derselben Zeitspanne nur knapp 80 % der kurzfristigen. Die Hazardkurve der langfristigen Einkäufe verläuft sehr zackig mit hohen Werten, was hauptsächlich auch auf die geringe Anzahl von Aktivitäten zurückzuführen ist. Die Hazardkurve der kurzfristigen Einkäufe verläuft ebenfalls sehr zackig mit relativ hohen Werten zu Beginn. Die geglätteten Kurvenverläufe sind im Anhang V zu sehen.

Sehr konstante Hazardkurvenverläufe mit relativ niedrigen Werten weisen die Kategorien Freizeit allgemein, Besuch bei Verwandten, Treffen mit Freunden und Lokal, Kino, Theater auf (Abbildung 9.5, 9.8, 9.9 und 9.14), was auf eine exponentielle Verteilung der Wahrscheinlichkeitsdichtefunktion schließen läßt. Der Verlauf der Survivalkurve zeigt bei allen vier Kategorien eine exponentielle, wenn auch leicht differenzierte Form. Bei den Freizeitaktivitäten liegt die 50%-Quartile bei ca. 1,5 Stunden, bei Verwandtenbesuchen bei 3 Stunden, bei Treffen mit Freunden bei 2,8 Stunden und bei Lokal- und Kinobesuchen bei etwa 1 Stunde. Deutliche Spitzen der Hazardkurven sind jeweils nur in jenen Bereichen festzustellen, in denen die Anzahl der Aktivitäten nicht mehr von relevanter Bedeutung ist.

Die Hazardkurve der Kategorie Dienstleistung allgemein (Abbildung 9.6) beginnt auf einem sehr hohen Niveau und fällt bis zu einer Aktivitätsdauer von ca. 3,5 Stunden mit einem relativen Höhepunkt nach 1,5 Stunden auf Null ab. Die verbleibende Spitze bei 4,2 Stunden Aktivitätsdauer hat keinerlei Einfluß mehr. Eine sehr zackige Hazardkurve mit zum Teil extrem hohen Werten zeigt auch die Kategorie Erledigungen (Abbildung 9.7), was auf sehr viele Ereignisse, die jeweils gleich lange andauern (10 Minuten, 15 Minuten, 25 Minuten und 30 Minuten) schließen läßt. Dies kann teilweise auch auf die Gewohnheit der Befragten zurückzuführen sein, normalerweise gerundete Schätzwerte der Aktivitätsdauern anzugeben. Vor allem bei zeitlich eher kurzen Aktivitäten (wie in dieser Kategorie vornehmlich vorhanden), die erst am Abend ins Befragungsformular eingetragen werden, ist eine solche Vorgangsweise durchaus vorstellbar.

Eine genauere Betrachtung der Hazard- und Survivalkurve der Kategorie Spaziergang / Spazierfahrt (Abbildung 9.12) läßt ebenfalls den Schluß zu, daß gerundete Zeitangaben gemacht worden sein könnten, da relativ hohe Spitzenwerte bei genau einer halben, einer, eineinhalb und zwei Stunden vorliegen. Längere Spaziergänge kommen eher selten vor, nach 40 Minuten ist rund die Hälfte schon beendet, nach 1 Stunde etwa 70 % und nur rund 5 % der Spaziergänge und -fahrten dauern länger als 2 Stunden an.

Die Hazardwerte der Kategorien Verein, Initiative, Kirche und Sport aktiv (Abbildung 9.10 und Abbildung 9.11) steigen bis zu einem Spitzenwert bei knapp 3 Stunden Aktivitätsdauer fast kontinuierlich an, die danach folgenden Spitzen der Hazardkurve sind aufgrund der geringen Aktivitätsanzahl wiederum nahezu bedeutungslos. Die Survivalkurve verläuft jedoch bei beiden Kategorien relativ flach bzw. keineswegs exponentiell fallend, das heißt, daß die Aktivitäten relativ konstant über die Zeit verteilt sind. So endet die Hälfte aller Aktivitäten der beiden Kategorien erst nach einer Dauer von knapp 2 Stunden, mehr als 4 Stunden dauern jedoch auch nahezu keine Aktivitäten an.

Die Auswertung der Kategorie Etwas Neues sehen (Abbildung 9.13) gestaltet sich ob der sehr geringen Anzahl an Aktivitäten schwierig bzw. können durch die stufenförmige Survivalkurve und die extrem zackige Hazardkurve keine allgemein gültigen Schlußfolgerungen gezogen werden.

Abbildung 9.1 : Gesamte Aktivitäten

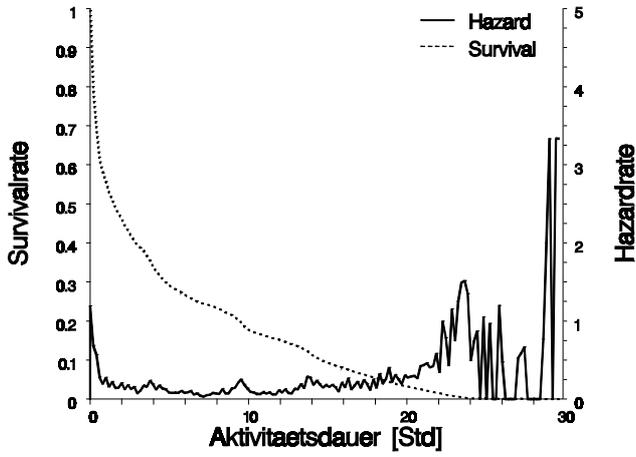


Abbildung 9.2 : Arbeit

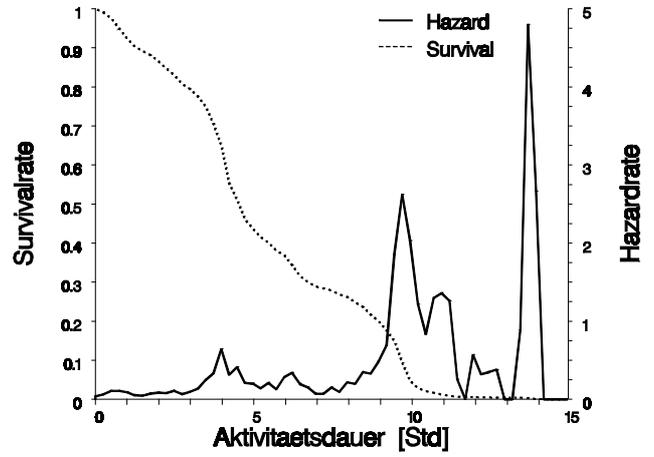


Abbildung 9.3 : Einkauf langfristig

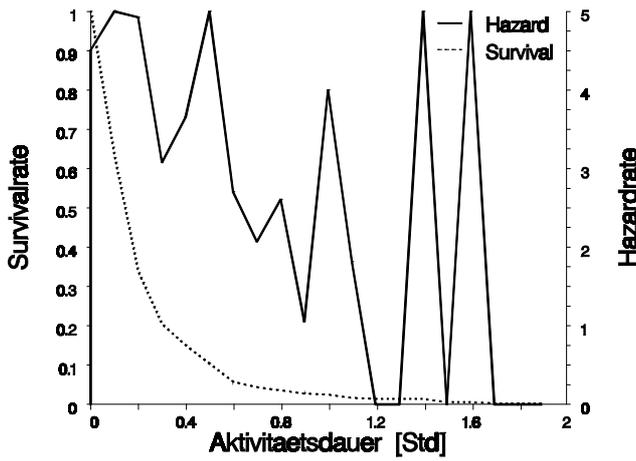


Abbildung 9.4 : Einkauf kurzfristig

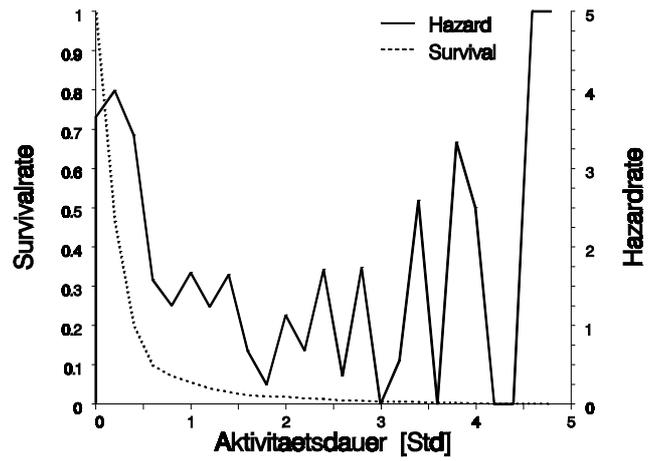


Abbildung 9.5 : Freizeit allgemein

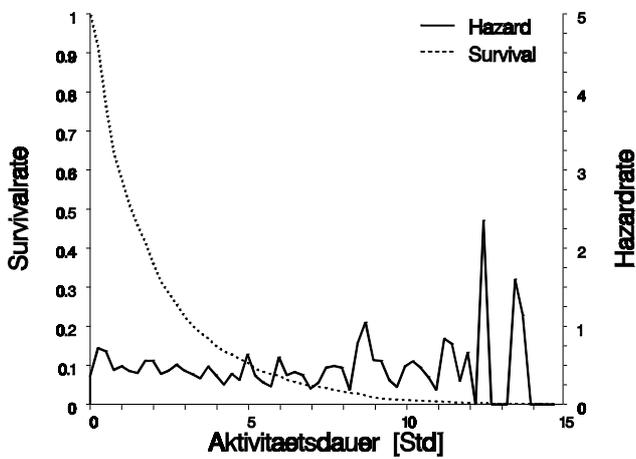


Abbildung 9.6 : Dienstleistung allgemein

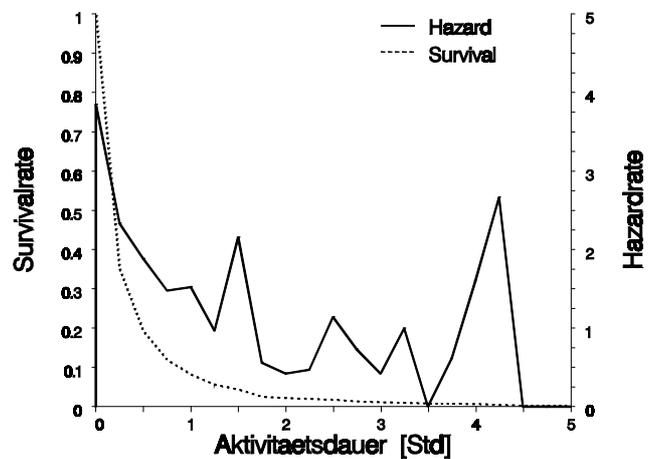


Abbildung 9.7 : Erledigungen

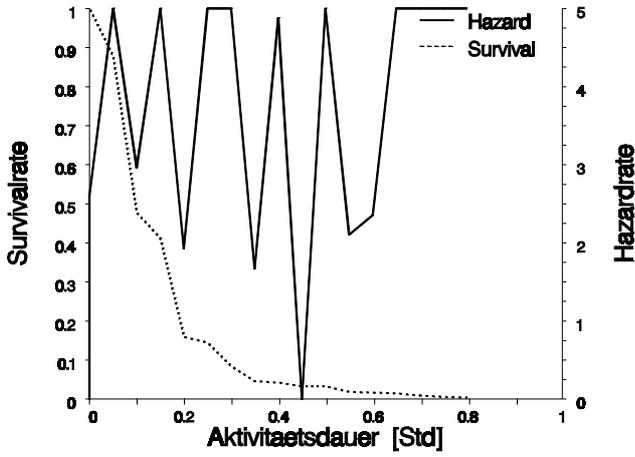


Abbildung 9.8 : Besuch bei Verwandten

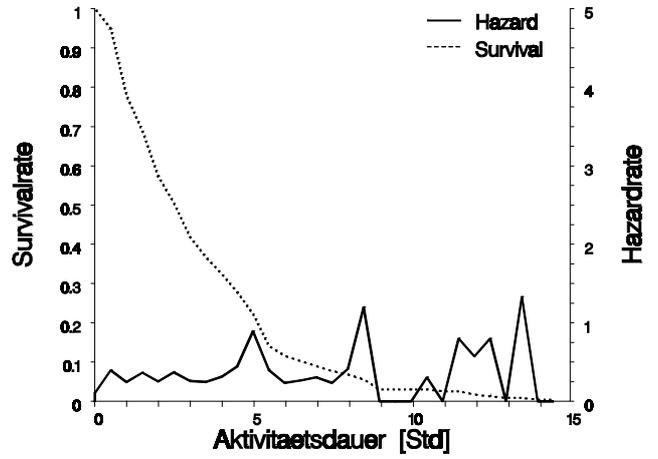


Abbildung 9.9 : Treffen mit Freunden

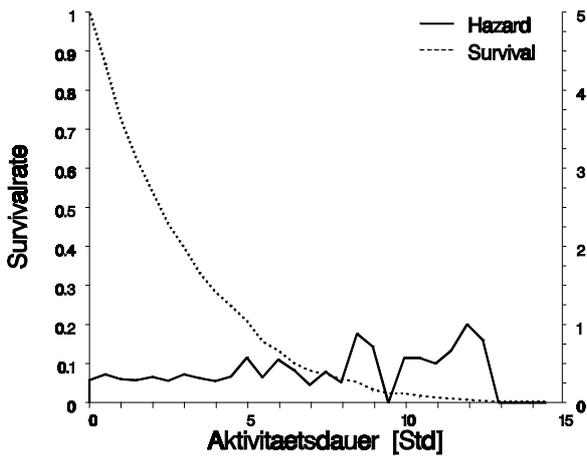


Abbildung 9.10 : Verein, Initiative, Kirche

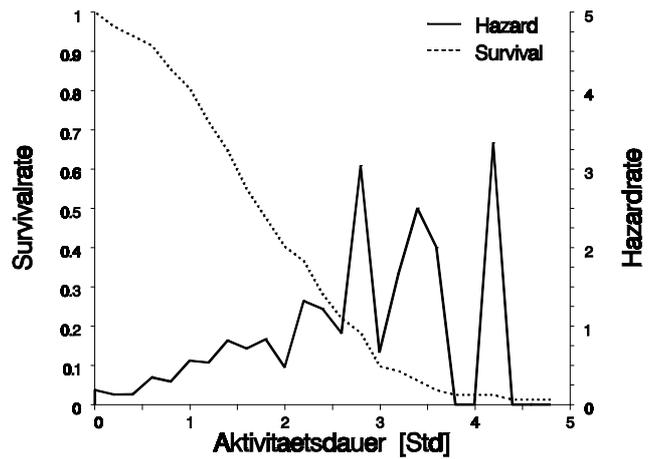


Abbildung 9.11 : Sport (aktiv)

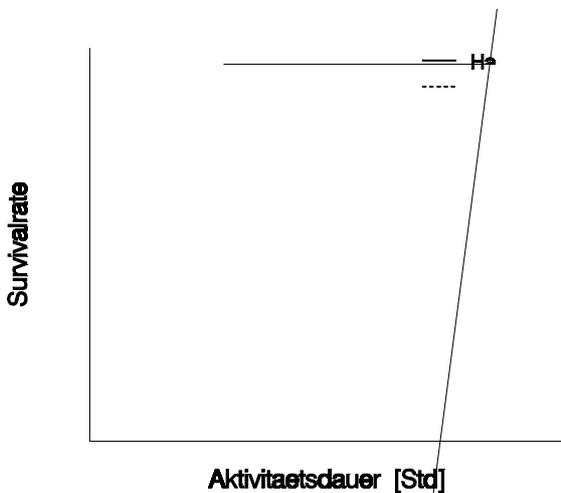


Abbildung 9.12 : Spaziergang / -fahrt

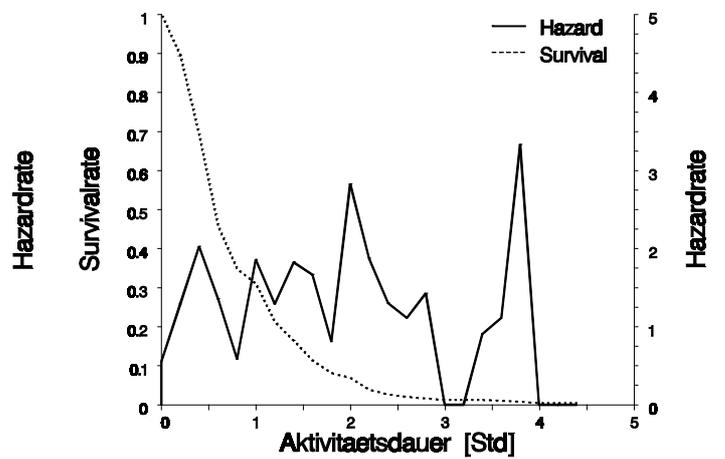


Abbildung 9.13 : Etwas Neues sehen

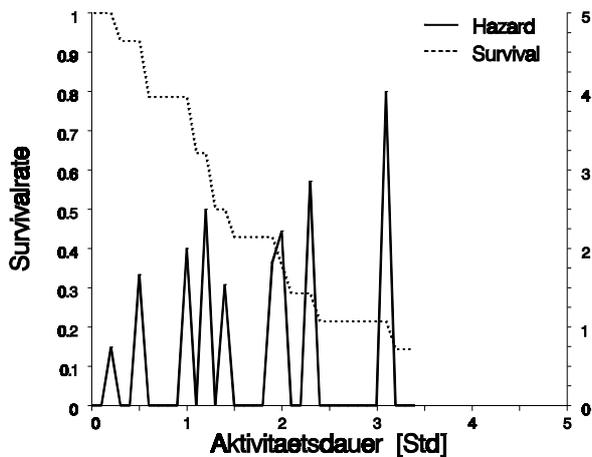
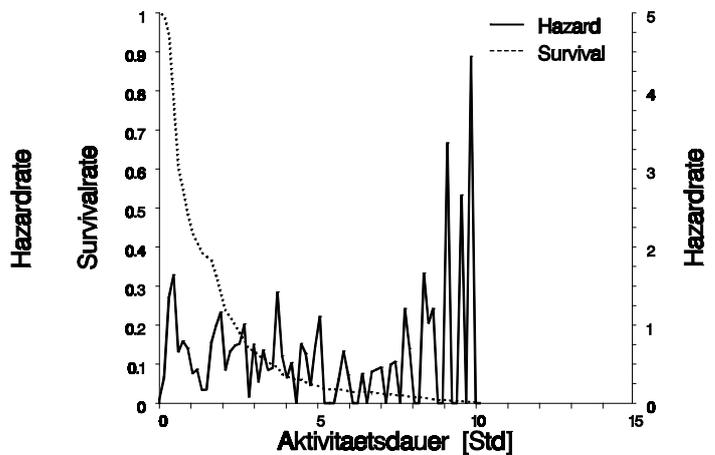


Abbildung 9.14 : Lokal, Kino, Theater



### 9.2.2.3 Genauere Betrachtung ausgewählter Aktivitätskategorien

Wie zuvor erwähnt werden bei den semiparametrischen und vollparametrischen Modellschätzungen (Kapitel 9.2.3 und 9.2.4) vier Aktivitätskategorien näher untersucht. Diese sind „Einkauf kurzfristig“, „Freizeit allgemein“, „Dienstleistung allgemein“ und „Treffen mit Freunden“. Es wird nun für diese Gruppen einerseits versucht, etwaige Unterschiede zwischen den einzelnen Altersschichten und zwischen den beiden Geschlechtern aufzuzeigen und andererseits werden sogenannte stufenweise Log-rank-tests durchgeführt, um aus den möglichen Einflußvariablen die signifikanten herauszufinden.

In den folgenden Abbildungen (9.15 bis 9.18 bzw. 9.19 bis 9.22) werden die geglätteten Hazardkurven der sieben Altersgruppen bzw. die ungeglätteten der beiden Geschlechter der vier Kategorien grafisch dargestellt. Für die Kurvenglättung wurde ein spezielles SAS-Makro (Allison, 1995) angewandt. Bei den Geschlechtsunterschieden sind jedoch die ungeglätteten Hazardkurven, bei denen auch die erste Stunde angezeigt werden kann und nicht in die Kurvenverläufe aggregiert wird, wesentlich anschaulicher und werden daher gezeigt. Die sieben unterschiedlichen Kurven in den ersten vier Abbildungen (9.15 bis 9.18) stellen jeweils eine bestimmte Altersgruppe dar, die im Vorfeld definiert worden sind. Die erste Gruppe sind unter 25-jährige Personen, die danach folgenden Gruppen sind um jeweils ein 10-Jahresintervall älter, die siebente Gruppe sind Personen, die älter als 75 Jahre sind. Bei den beiden Geschlechtern sind die weiblichen Personen wie üblich in grün, die männlichen Umfrageteilnehmer in blau dargestellt.

Die Hazardkurven der einzelnen Altersgruppen scheinen meist recht unterschiedlich. Nicht nur die Werte der Hazardfunktion, sondern auch die Bandbreite der Kurve auf der Skala der Aktivitätsdauer lassen zum Teil gravierende Unterschiede zwischen den einzelnen Altersgruppen erkennen.

In der Kategorie Dienstleistung allgemein (Abbildung 9.17) sind extreme Unterschiede zwischen den einzelnen Altersgruppen erkennbar. Zudem fällt auf, daß ältere Personen ab 65, aber auch Personen unter 25 ihre Erledigungen alle in relativ kurzer Zeit hinter sich bringen, nach rund 2 Stunden sind alle beendet. Nur die Gruppe der 50-jährigen macht bei den Personen mittleren Alters eine Ausnahme, da bis auf eine einzige längere Aktivität alle anderen spätestens nach 70 Minuten abgeschlossen sind.

Sieht man sich die ungeglätteten Survival- und Hazardkurven der Kategorien Einkauf kurzfristig und Freizeit allgemein an (siehe Anhang V), so fällt auf, daß dort die Unterschiede gar nicht so extrem sind. Ältere Personen haben jedoch im vorderen Bereich der Zeitachse höhere Survivalwerte, das heißt im Schnitt etwas längere Zeitspannen für ihre Freizeit- und kurzfristigen Einkaufsaktivitäten, 30- und 50-jährige haben im Gegensatz dazu relativ viele kurze Aktivitäten unternommen. Die Hazardkurve der Kategorie Treffen mit Freunden (Abbildung 9.18) zeigt einen ähnlichen Verlauf wie die allgemeine Freizeitkategorie (Abbildung 9.16), jedoch differieren bei den einzelnen Altersgruppen die Werte der Survivalkurve etwas mehr. Auffällig ist die weitaus geringere Häufigkeit von Aktivitäten bei älteren Menschen in dieser Kategorie, auch erkennbar am stufenförmigen Verlauf der Survivalkurve.

Abbildung 9.15 : Geglättete Hazardkurve der Altersschichtung der Kategorie Einkauf kurzfristig

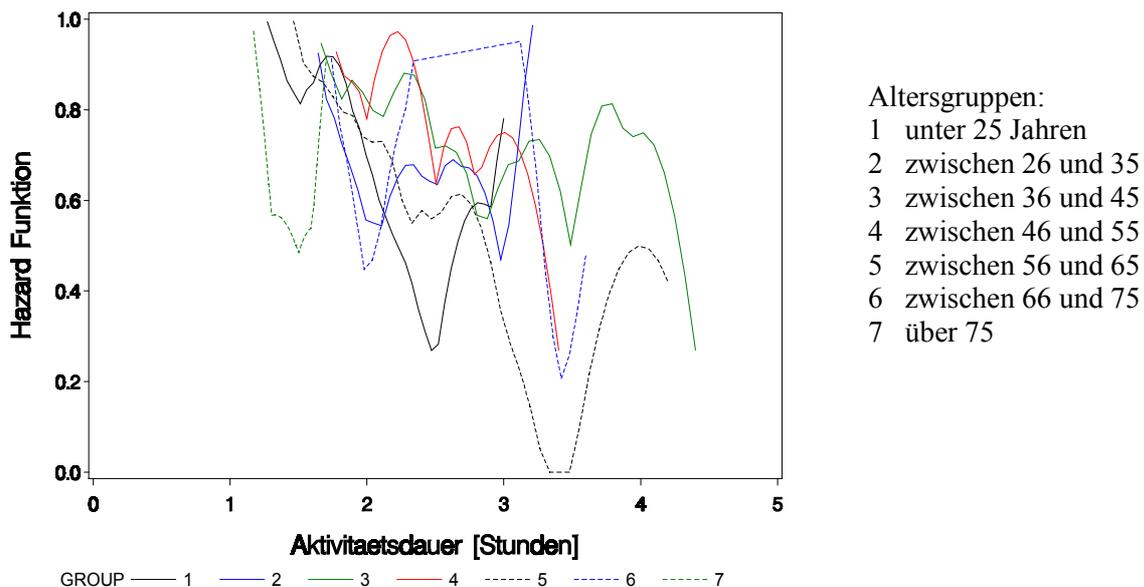


Abbildung 9.16 : Geglättete Hazardkurve der Altersschichtung der Kategorie Freizeit allgemein

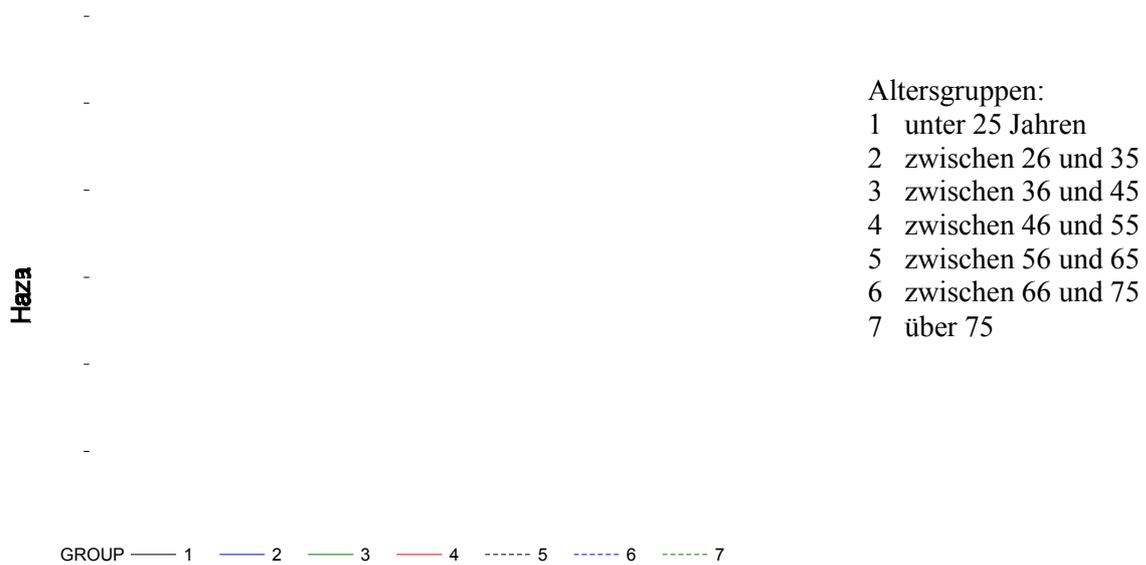


Abbildung 9.17 : Geglättete Hazardkurve der Altersschichtung der Kategorie Dienstleistung allgemein

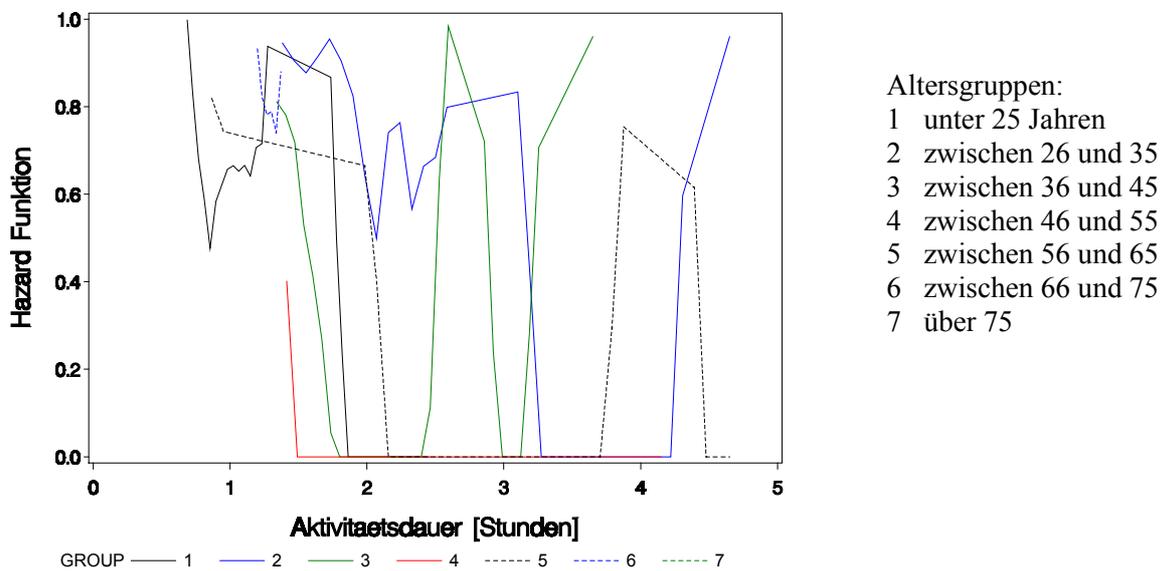
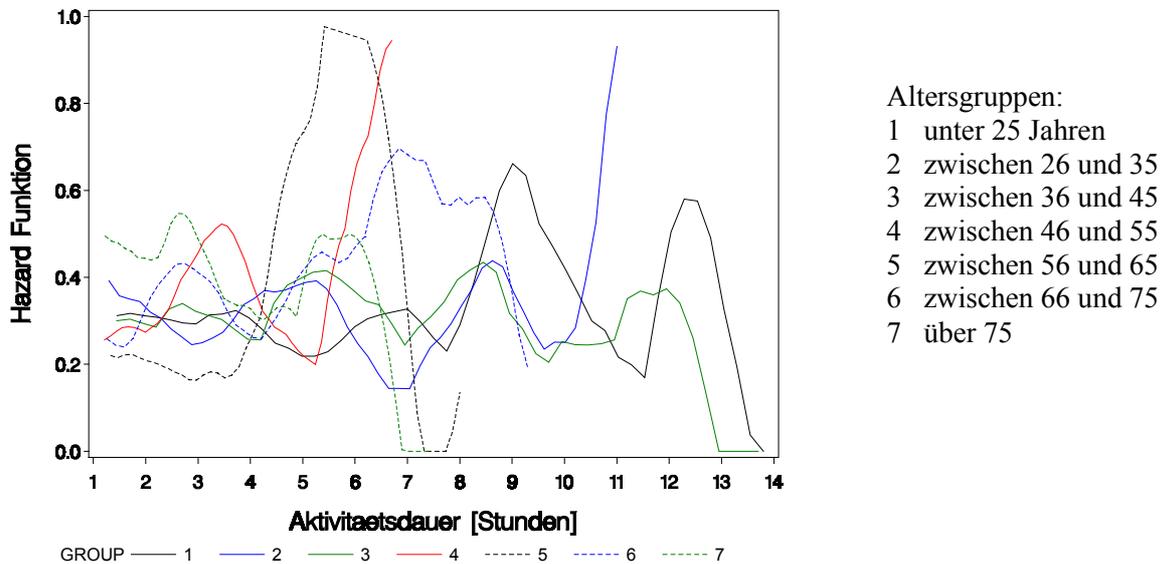


Abbildung 9.18 : Geplättete Hazardkurve der Altersschichtung der Kategorie Treffen mit Freunden



Die Unterschiede zwischen den beiden Geschlechtern sind, wie in den folgenden Abbildungen deutlich sichtbar, weitaus geringer. Größere Differenzen im Verlauf der Hazardkurven gibt es wiederum nur in der Kategorie Dienstleistung allgemein, hierbei aber auch erst im hinteren Aktivitätsdauerbereich, also nicht im Bereich bis zu einer Stunde, der vor allem in dieser Kategorie von Bedeutung ist. Der Hauptgrund für die Darstellung von ungeplätteten Kurven hier ist die Tatsache, daß Werte bis zu einer Aktivitätsdauer von einer Stunde bei der geplätteten Kurve nicht mehr berücksichtigt werden können. Nachdem aber bis zu diesem Zeitpunkt der Großteil der Aktivitäten schon beendet sind (in der Kategorie Dienstleistung beispielsweise 95% bei Männern, 90% bei Frauen, in der Kategorie Einkauf bei beiden Geschlechtern knapp über 90%), ist es sinnvoll, diesem Bereich entsprechende Bedeutung zuzumessen. Auch für die Kategorien Freizeit allgemein und Treffen mit Freunden wird die ungeplättete Kurve gezeigt, da sie wesentlich mehr Aussagekraft bezüglich der Geschlechtsunterschiede hat. Die Spitzenwerte gegen Ende der Zeitachse haben aufgrund der äußerst geringen Aktivitätenhäufigkeit keinerlei Bedeutung. Man kann nur die jeweils letzte Aktivität gut erkennen.

Abbildung 9.19 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Einkauf kurzfristig

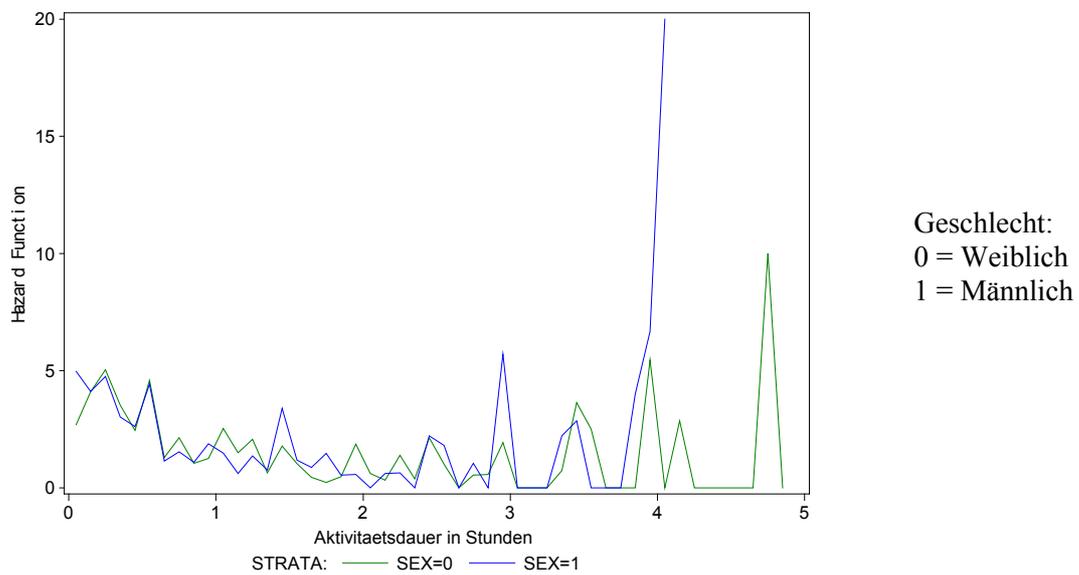


Abbildung 9.20 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Freizeit allgemein

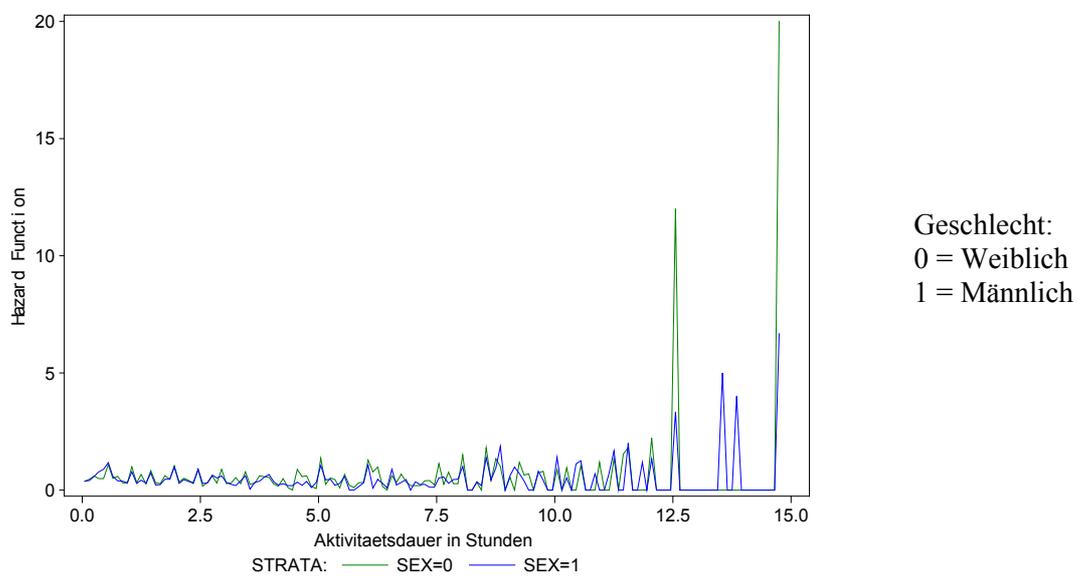


Abbildung 9.21 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Dienstleistung allgemein

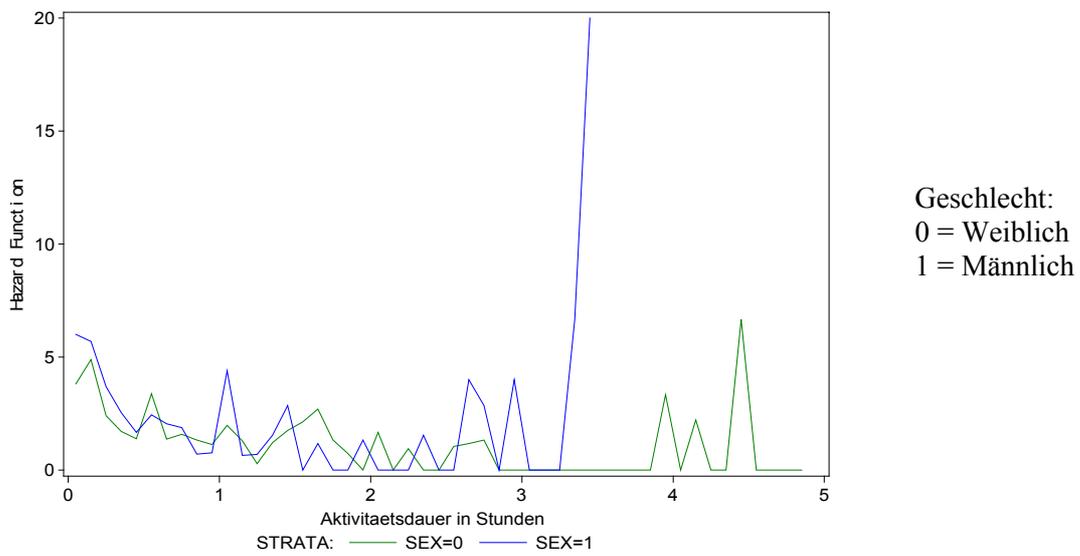
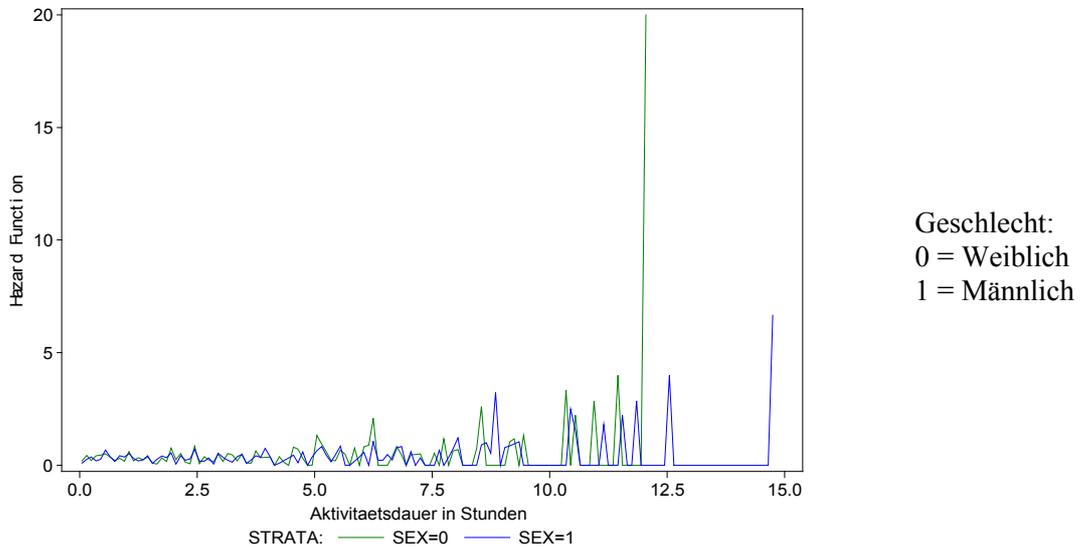


Abbildung 9.22 : Ungeglättete Hazardkurve der Geschlechtsunterschiede, Treffen mit Freunden



#### 9.2.2.4 Vergleich mit den Ergebnissen der semiparametrischen Auswertung

Für die ausgewählten vier Kategorien sind zudem noch als Überprüfung der möglichen Einflußvariablen der sogenannte Log-Rank-Test und der Wilcoxon-Test gemacht worden. Dabei sind jeweils zwei Stufen der Auswertung möglich. Einerseits können durch sogenannte *Univariate Chi-Squares* die einzelnen Variablen separat getestet werden, ohne, daß sie sich gegenseitig in irgendeiner Weise beeinflussen bzw. korrelieren, andererseits können auch *Forward Stepwise Sequences of Chi-Squares* erstellt werden, also stufenförmige Auswertungen von Signifikanzen, die in Abhängigkeit derjenigen Variable, die bei singulärer Betrachtung den höchsten Chi-Quadratwert ergibt, angegeben werden. Diese stufenförmigen Log-Rank-Tests können auch als nonparametrische Koeffiziententests des *accelerated failure time model (AFT-Modell, siehe Kapitel 6.4.1)* interpretiert werden. Sie sind auch den Hypothesentests bei semiparametrischen Auswertungen mit *PROC PHREG*, daß alle Koeffizienten Null sind bzw. alle Variablen keinen Einfluß haben, sehr ähnlich (Allison, S.52). Da die Tests jedoch keine Koeffizientenabschätzungen bzw. Angaben über die Größe des Einflusses der Kovariablen geben, sind es meist nur überblicksmäßige bzw. vergleichende Auswertungen. Sie können als Grundlage für Regressionsmodelle bei den semiparametrischen Modellen verwendet werden.

Die Ergebnisse der Überprüfung von jeder Variable separat (*univariate Chi-Square-Tests*) sind im Anhang VI dargestellt. Aufgrund der Tatsache, daß die Wilcoxon-Tests wesentlich mehr Gewichtung auf frühe Ereigniszeiten (bzw. kurze Aktivitätsdauern hier) legen (Allison, S. 39), sind in den Kategorien Einkauf kurzfristig und Dienstleistung allgemein nur die Ergebnisse des Wilcoxon-Tests präsentiert, da hier 90 bis 95 % der Aktivitätsdauern schon nach rund einer Stunde beendet sind, die Gewichtung also begründeterweise auf kurze Aktivitätsdauern gelegt wird. In den Kategorien Freizeit allgemein und Treffen mit Freunden werden hingegen die Ergebnisse des Log-Rank-Tests dargestellt.

Die Auswertungen der stufenförmigen Wilcoxon- und Log-Rank-Tests sind ebenfalls im Anhang VI dargestellt, zeigen jedoch meist keine gravierenden Unterschiede zu den in den semiparametrischen Modellen gewonnenen Erkenntnissen. Geringfügig andere Ergebnisse in der Signifikanz der überprüften Variablen sind manchmal festzustellen, was jedoch meist auf die Vorgangsweise bei diesen Tests zurückzuführen ist. Da die stufenförmigen Tests immer nur jene Variablen, die vorher überprüft worden sind, in die Berechnung mit einbeziehen, nicht aber jene, die erst danach untersucht werden, kann es zu unrichtigen Schlußfolgerungen kommen. Zudem sind in den semiparametrischen Modellen mit Hilfe von Korrelationsmatrizen Variablen mit zu hohen Korrelationen vorab aussortiert worden, die bei den stufenförmigen Tests noch inkludiert sind. Detaillierte Unterschiede zwischen den nonparametrischen und semiparametrischen Auswertungen sind ebenfalls im Anhang VI erläutert. Aus all den oben genannten Gründen ist den Ergebnissen, die im folgenden Kapitel (9.2.3) präsentiert werden, wesentlich mehr Bedeutung beizumessen.

### 9.2.3 Semiparametrische Modelle mit *PROC PHREG*

#### 9.2.3.1 Einleitung

In diesem Kapitel werden nur noch die Auswertungen der vier ausgewählten Aktivitätskategorien Einkauf kurzfristig (mit 3231 Aktivitäten insgesamt), Freizeit allgemein (3025), Dienstleistung allgemein (922) und Treffen mit Freunden (687) dargestellt und diskutiert. Nachdem relativ wenige Aktivitäten sehr lange andauern, werden in allen vier Kategorien jene Aktivitäten, die nach einer bestimmten Zeit noch nicht abgeschlossen sind, aus den Modellschätzungen ausgeschlossen, da sie ansonsten nur das Bild verzerren würden. Diese Grenzen werden bei kurzfristigen Einkäufen nach einer Aktivitätsdauer von 5 Stunden gesetzt, bei Freizeitaktivitäten bei 15 Stunden, bei Dienstleistungen nach 2 Stunden und bei Treffen mit Freunden ebenfalls bei 15 Stunden. Durch diese Limits der Aktivitätsdauer reduzieren sich die in die Modellschätzung einbezogenen Fälle auf 3221 Aktivitäten für Einkäufe bzw. 2987 (Freizeit), 896 (Dienstleistung) und 669 (Treffen mit Freunden).

Andere als die ausgewählten Aktivitätskategorien wären für eine detailliertere Betrachtung nur mit Vorbehalten in Frage gekommen. So wäre beispielsweise die Kategorie Sport aktiv zwar recht interessant, jedoch werden viel zu wenige Aktivitäten (insgesamt nur 119 von 23 Personen) in den Umfragen berichtet, sodaß eine aussagekräftige Schlußfolgerung der Ergebnisse nicht gewährleistet wäre. Die Aktivitätskategorie Bar / Diskothek / Lokal mit 622 Aktivitäten wäre eventuell auch noch interessant, als Unterkategorie der Freizeitaktivitäten wird jedoch Treffen mit Freunden ausgewählt. In der Kategorie Spaziergang / Spazierfahrt mit 497 berichteten Aktivitäten mußten zu viele Korrekturen gemacht werden. Alle anderen Aktivitätskategorien sind von keinem größeren Interesse, da nicht genügend Aktivitäten berichtet worden sind.

Einzig die Aktivitätskategorie Arbeit mit 2589 berichteten Aktivitäten wäre an und für sich noch interessant, da sie einen Vergleich zwischen erzwungenen und frei wählbaren Aktivitäten offenbaren würde (vgl. dazu *Authority constraints* von Hägarstrand). Diese Kategorie wird jedoch in vielen bisherigen Studien eingehend behandelt (vergleiche Oh, 1999 (*comparison of the duration dynamics of the obligatory (eg., work) and discretionary (eg., recreational or social) activities*); Filiou, 1994; Wilson, 1994 etc.) und Auswertungen mit dieser Kategorie werden deshalb hier nicht weiter verfolgt.

Die grundlegenden Ziele der semiparametrischen Auswertung in dieser Arbeit bestehen darin, aus einer Reihe möglicher Einflußvariablen jene herauszufinden, bei denen tatsächlich ein Einfluß auf die Dauer der Aktivität festzustellen ist und in einem zweiten Schritt die Größe und Richtung dieser Beeinflussung mit geeigneten Zahlenwerten anzugeben. Alle Modellschätzungen sind mit semiparametrischen Regressionsmethoden mit *PROC PHREG* in mehreren Arbeitsschritten durchgeführt worden.

Dazu sind zuerst für jede Kategorie mit Hilfe einer singulären Betrachtung alle möglichen Variablen (siehe dazu Liste der möglichen Einflußvariablen im Anhang IV) auf ihre Einzelsignifikanz überprüft worden. Mit Hilfe von Korrelationsmatrizen sind in einem zweiten Schritt aus jenen Variablen, deren Korrelation zu hoch war (über einem Wert von 0.6), die mit den niedrigsten Chi-Quadratwerten eliminiert worden. Für die verbliebenen Variablen ist anschließend jeweils die Proportionalität mit Hilfe einer üblichen grafischen Methode des Vergleichs der negativen Loglog-Survivalkurven überprüft worden. Diese Kurven sind simple Transformationen der geschätzten Survivalkurven (mathematisch: Multiplikation des logarithmierten Ergebnisses mit dem Wert 2  $[-\ln(\ln \hat{S})]$ ), die möglichst parallel verlaufen sollten (Details dazu im Programm Hazard\_PHREG\_I). In einem weiteren Schritt ist ein erstes Gesamtmodell erstellt worden, mit Hilfe dessen Ergebnisse danach weitere nicht signifikante Variablen eliminiert worden sind. Die zweite Gesamtmodellschätzung beinhaltet demnach nur mehr stark signifikante Variablen. Durch Probieren verschiedenster Variablensätze kann zwar noch eine gewisse Steigerung der Paßform bzw. Güte des Modells (*goodness of model fit*, Allison, S.138 ff ) erreicht werden, die Unterschiede sind jedoch meist nicht mehr gravierend.

Bei der Schätzung eines zweiten Gesamtmodells, in dem nur noch signifikante Variablen aus dem ersten Gesamtmodell weiter verwendet worden sind, kann es vorkommen, daß die Güte des Modells etwas unter der des ersten Gesamtmodells liegt. Hier ist jedoch eine geringfügige Qualitätsverschlechterung des Modells in bezug auf die Güte zu Gunsten eines Modells, in dem nur noch signifikante Variablen aus dem ersten Gesamtmodell und keine die Aktivitätsdauer überhaupt nicht beeinflussenden Größen vorhanden sind, in Kauf genommen worden.

Die Ergebnisse dieses zweiten Gesamtmodells werden nun auf den folgenden Seiten detailliert präsentiert und diskutiert. Zudem wird noch ein weiterer Vergleich für zwei Gruppen angestellt, der die Häufigkeit an Aktivitäten in der jeweiligen Kategorie betrifft. Einen zusätzlichen interessanten Aspekt hält die Aufgliederung der Variablen in eine soziodemografische und eine aktivitätenspezifische Sparte bereit, die im Zusatzanhang XXI (PHREG AD Semiparametrische Modelle) genauer zu sehen ist.

Bei der Berechnung gibt es zwei unterschiedliche Methoden, wie man gleichzeitig stattfindende Ereignisse bzw. zum selben Zeitpunkt endende Aktivitätsdauern berücksichtigen kann. In den Programmen werden beide (Exact- und Efron-methode) angewendet, die Ergebnisse liegen jedoch knapp beieinander. Große Differenzen gibt es hauptsächlich bei den unterschiedlich langen Rechenzeiten.

### 9.2.3.2 Einzelne Kategorien

Vergangene Studien haben meist nur eine sehr geringe Anzahl an möglichen Einflußvariablen untersucht bzw. die Umfrageteilnehmer schon vor den Modellschätzungen aufgrund spezieller individueller Charakteristika in verschiedene Gruppen eingeteilt. Kansky (1967) ließ das erste Mal eine größere Menge an Variablen in die Auswertungen einfließen, faßte diese 27 beschreibenden Variablen und 22 Meßwerte des Verkehrsverhaltensmusters jedoch zu einer einzigen resultierenden Variable zusammen. Hanson und Hanson (1981b) haben erstmals nicht die Umfrageteilnehmer aufgrund ihrer beschreibenden Charakteristika eingeteilt und danach untersucht, sondern zuerst mit einem größeren Satz von Variablen das Verkehrsverhalten untersucht und die Umfrageteilnehmer erst aufgrund ihres Verkehrsverhaltens in verschiedene Gruppen eingeteilt.

Hier wird bewußt auf eine ähnliche Vorgangsweise verzichtet, da vorrangig untersucht werden soll, welche möglichen Einflußgrößen tatsächlich Auswirkungen auf die Aktivitätsdauer aufweisen, eine Einteilung in verschiedene Gruppen ist nicht vorgesehen. Aus diesem Grund ist eine relativ große Anzahl an möglichen Einflußgrößen aus den Umfragen erarbeitet worden.

Bei insgesamt 50 Variablen sind Korrelationen untereinander jedoch unvermeidbar und so sind schon vor jedem ersten Gesamtmodell einige Variablen eliminiert worden. Verhältnismäßig hohe Korrelationen liegen verständlicherweise zwischen den Variablen Führerscheinbesitz, Autos im Haushalt und Autoverfügbarkeit vor. Ebenso sind die Variablen Gesellschaftsgruppe, Alter, Kinder im Haushalt meist recht eng miteinander verknüpft. Zwischen dem benützten Verkehrsmittel vor der Aktivität und dem Verkehrsmittel, das zur nächsten Aktivität benützt wird, besteht auch ein sehr enger Zusammenhang. Jemand, der mit dem eigenen Auto einkaufen fährt, wird mit ziemlich großer Sicherheit auch wieder den Heimweg oder einen möglichen weiteren Weg mit seinem Fahrzeug antreten. Die Fragen, ob am Aktivitätsort Geld ausgegeben worden ist bzw. wie hoch diese monetären Ausgaben gewesen sind, sind natürlich ebenfalls stark miteinander verbunden. Sind die Korrelationen zwischen zwei oder mehreren Variablen zu groß gewesen, so ist die Variable mit dem höchsten Qui-Quadratwert der singulären Variablenschätzung für ein erstes Gesamtmodell weiter verwendet worden.

Eine Übersicht aller möglichen Einflußvariablen für die vier Kategorien sehen Sie in der Tabelle 9.1, in der die stark signifikanten Variablen nach einer Modellschätzung der zweiten Stufe mit einem Stern (\*) markiert sind. Stark signifikant bedeutet in diesem Fall, daß sie einen Signifikanzwert unter 0.05 aufweisen bzw. im 95% Konfidenzintervall liegen. Die marginal im 90% Konfidenzintervall signifikanten Variablen sind mit zwei Sternen (\*\*) gekennzeichnet.

Tabelle 9.1 : Übersicht der Signifikanz der möglichen Einflußvariablen

Variable	Variablen- abkürzung	Einkauf kurzfristig	Freizeit allgemein	Dienstl. allgemein	Treffen mit Freunden	Dummy
Geschlecht	SEX	*		*		X
Altersgruppe	AGE_**					X
Alter	AGE			*		
Unter 50 Jahren j/n	YOUNG					X
Führerscheinbesitz j/n	LICENSE	*				X
Haushaltsvorstand j/n	HEAD_O_HH					X
arbeitend / nicht arbeitend	WORKED			**		X
Kinder im Haushalt j/n	CHILDREN				*	X
Single bzw. alleine lebend j/n	SINGLE	**				X
Autoverfügbarkeit j/n	CAR_POSS					X
Ausbildung während der Umfrage	EDUC					X
Gesellschaftsgruppe Nr. * j/n	G*_D					X
Anzahl der Ausbildungsjahre	EDUC_YR					
Nummer der Gesellschaftsgruppe	G	*				
Anzahl der Autos im Haushalt	CAR	*				
Anzahl der Führerscheine im HH	HH_NR_L				*	
Anzahl der Haushaltsmitglieder	HH_MEMB		*			
Begleitung j/n	ACC_D					X
Aktivität geplant j/n	PLANNED	*	*		**	X
Ausgaben am Aktivitätsort j/n	A_EXP_D		**	*	*	X
Wochenende j/n	WEEKEND		*			X
Tageszeit (vor 13 Uhr j/n)	A_DAYT_D		*		*	X
Unabhängiges VM zur Aktivität	A_P_MM_D		*			X
Unabhängiges VM nach der A	A_N_MM_D	*				X
Unabh. motorisiertes VM zur A	A_P_MMMD					X
Unabh. motor. VM von der A	A_N_MMMD	*				X
Anzahl der Begleitpersonen	ACC	*		**	*	
Ausgabenhöhe am Aktivitätsort	A_EXP	*				
Tageszeit (vo/na/ab)	A_DAYT	*				
Distanz zum Aktivitätsort	A_P_DIST		*			
Reisezeit zum Aktivitätsort	A_P_TT	*		*	*	
Distanz zum nächsten Aktivitätsort	A_N_DIST					
Reisezeit zum nächsten Aktivitätsort	A_N_TT	*	**			
Verkehrsmittel zum Aktivitätsort	A_P_MM			*		
Verkehrsmittel vom Aktivitätsort	A_N_MM				*	
Anzahl der Wege gesamt	NR_T_SUM					
Anzahl der Reisen gesamt	NR_J_SUM		*			
Vorangegangene Aktivitätskategorie	A_P_PUR	*	*			
Nachfolgende Aktivitätskategorie	A_N_PUR		*	*	*	
Anzahl der A in dieser Kategorie	NR_A_SUMC	*			*	
Gesamtakt. in dieser Kategorie	N	3221	2987	896	669	

\* stark signifikant  $\alpha < 0.05$  , \*\* marginal signifikant  $\alpha < 0.10$

Im Folgenden werden einige auffällige Erkenntnisse aus den Auswertungen näher betrachtet.

In allen vier Kategorien wird die Aktivitätsdauer von der Frage beeinflusst, ob bzw. wieviel Ausgaben gemacht worden sind. Bei kurzfristigen Einkäufen und bei Dienstleistungen verlängern Ausgaben die Dauer der Aktivität, Freizeitaktivitäten werden durch Ausgaben verkürzt, Treffen mit Freunden sogar sehr deutlich.

Die Dauer der Reisezeit beeinflusst ebenfalls in allen vier Kategorien die Aktivitätsdauer. Nachdem die Reisezeit vor bzw. nach der Aktivität meist recht eng miteinander verknüpft sind, ist eine der beiden Variablen aufgrund der Korrelation schon vor dem ersten Gesamtmodell eliminiert worden und so scheint bis auf die Kategorie Einkauf nur eine der beiden in den Auswertungen auf. Trotzdem hat diese Variable stets signifikanten Einfluß auf die Dauer der Aktivität. Je länger die Reisezeit, desto länger ist auch in allen Kategorien die Aktivitätsdauer.

Auch die Art der Aktivität davor oder danach hat einen signifikanten Einfluß auf die Dauer dieser Aktivität, die Richtung des Einflusses ist immer negativ. Das heißt, je höher die Einstufung der Aktivitätskategorie ist, desto länger ist auch die Aktivitätsdauer.

Die Tageszeit hat nur in der Kategorie Dienstleistung keinen signifikanten Einfluß. Beim Einkaufen sind die Aktivitäten umso kürzer, je später die Aktivität stattfindet. Bei den beiden Freizeitkategorien verhält es sich umgekehrt. Findet die Aktivität am Vormittag statt, so hat sie eine wesentlich größere Wahrscheinlichkeit, früher beendet zu werden als am Abend bzw. Nachmittag.

Die Frage des Verkehrsmittels vor bzw. nach der Aktivität spielt eine weitaus geringere Rolle als angenommen bzw. erwartet. Bei Freizeitaktivitäten hat beispielsweise nur die Frage, ob vor der Aktivität ein von Fahrplänen etc. unabhängiges Verkehrsmittel benützt worden ist, oder nicht, einen Einfluß, beim Einkaufen die eines unabhängigen Verkehrsmittels danach. Bei Dienstleistungen spielt nur die Art des benützten Verkehrsmittel zum Aktivitätsort eine Rolle, bei Treffen mit Freunden das Verkehrsmittel danach.

Auffällig ist zudem, daß weitaus weniger soziodemografische Charakteristika einen Einfluß auf die Aktivitätsdauer haben, als ursprünglich erwartet worden ist. Dies steht auch in einem Gegensatz zu vielen anderen früheren Studien, die gerade die individuellen beschreibenden Variablen als die maßgebenden Einflußgrößen annehmen. So hat das Geschlecht nur in den Kategorien Einkauf und Dienstleistungen einen signifikanten Einfluß dahingehend, daß Männer eine niedrigere Wahrscheinlichkeit haben, längere Aktivitäten zu unternehmen. Die in manchen Studien beobachtete größere Flexibilität aufgrund der Möglichkeit, ein eigenes unabhängiges Kraftfahrzeug benützen zu können,

scheint hier mit Ausnahme der Einkaufsaktivitäten keinen allzu großen Einfluß auf die Aktivitätsdauer zu haben. Eventuell spielen auch zeitliche Rahmenbedingungen der Umfrageteilnehmer eine gewichtige Rolle, die Dauer der Aktivität nicht gravierend verlängern zu können. Der Ausbildungsgrad der befragten Personen spielt scheinbar ebenso wie das Alter oder die Frage, ob die Person in der Umfragezeit gearbeitet hat oder nicht, nur eine untergeordnete Rolle.

Wie schon vorher kurz erwähnt, ist ein weiteres Detail von grundlegendem Interesse. Die Frage, ob Personen mit sporadischen Aktivitäten in der betreffenden Kategorie das Gesamtergebnis beeinflussen, wird hier mittels eines Vergleichs des jeweiligen Gesamtmodells der zweiten Stufe mit einer Modellschätzung der zweiten Stufe für Personen mit regelmäßiger Aktivitätsteilnahme beantwortet. Gleichzeitig ist in den folgenden Tabellen (Tabellen 9.2 bis 9.5) auch jeweils eine Angabe der Parameterschätzwerte  $\beta$  und der Hazardrate (*Risk Ratio*) für jede Variable enthalten. Mit deren Hilfe kann eine Angabe über die Größe und die Richtung des Einflusses der entsprechenden Variable gemacht werden. Die Hazardrate gibt die Größe der Änderung in bezug auf die vorherige Aktivitätsdauer an. Ist beim Parameterschätzwert  $\beta$  ein negatives Vorzeichen vorhanden, so verkürzt diese Einflußgröße die Aktivitätsdauer, bei positivem Parameterschätzwert wird die Dauer der Aktivität dementsprechend verlängert. Sind Variablen in einer Modellschätzung der zweiten Stufe signifikant im 95%-Konfidenzintervall, so werden sie fett gedruckt und mit einem Stern (\*) in der Spalte SN (Signifikanzniveau) gekennzeichnet, bei marginal signifikanten Variablen wird eine kursive Schrift und zwei Sterne (\*\*) verwendet. Nicht signifikante Variablen der zweiten Modellstufe sind durch normale Schrift ohne Stern erkennbar.

Tabelle 9.2 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Einkauf kurzfristig

Gesamtmodell 2. Stufe				Modell für regelmäßige Einkäufer			
Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN	Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN
<b>SEX</b>	<b>0.19179</b>	1.211	*	<b>SEX</b>	<b>0.18581</b>	1.204	*
<b>LICENSE</b>	<b>0.33082</b>	1.392	*	<b>LICENSE</b>	<b>0.32659</b>	1.386	*
SINGLE				SINGLE			
<b>G</b>	<b>0.03290</b>	1.033	*	<b>G</b>	<b>0.03112</b>	1.032	*
<b>CAR</b>	<b>-0.12983</b>	0.878	*	<b>CAR</b>	<b>-0.12476</b>	0.883	*
<b>PLANNED</b>	<b>-0.35602</b>	0.700	*	<b>PLANNED</b>	<b>-0.35494</b>	0.701	*
<b>A_N_MM_D</b>	<b>0.20487</b>	1.227	*	<b>A_N_MM_D</b>	<b>0.20030</b>	1.222	*
<b>A_N_MMMD</b>	<b>-0.37187</b>	0.689	*	<b>A_N_MMMD</b>	<b>-0.36965</b>	0.691	*
<b>ACC</b>	<b>-0.08232</b>	0.921	*	<b>ACC</b>	<b>-0.08206</b>	0.921	*
<b>A_EXP</b>	<b>-0.00691</b>	0.993	*	<b>A_EXP</b>	<b>-0.00691</b>	0.993	*
<b>A_DAYT</b>	<b>0.08895</b>	1.093	*	<b>A_DAYT</b>	<b>0.09316</b>	1.098	*
<b>A_P_TT</b>	<b>-0.01926</b>	0.981	*	<b>A_P_TT</b>	<b>-0.01928</b>	0.981	*
<b>A_N_TT</b>	<b>-0.00679</b>	0.993	*	<b>A_N_TT</b>	<b>-0.00692</b>	0.993	*
<b>A_P_PUR</b>	<b>-0.03478</b>	0.966	*	<b>A_P_PUR</b>	<b>-0.03337</b>	0.967	*
<b>NR_A_SUMC</b>	<b>0.00894</b>	1.009	*	<b>NR_A_SUMC</b>	<b>0.00888</b>	1.009	*

n = 3221; Likelihood Ratio Chi Quadrat = 748.91    n=3199; Likelihood Ratio Chi Quadrat = 756.03

Tabelle 9.3 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Freizeit allgemein

Gesamtmodell 2. Stufe				Modell bei regelmäßige Aktivitäten			
Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN	Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN
<b>HH_MEMB</b>	<b>0.07595</b>	1.079	*	<b>HH_MEMB</b>	<b>0.07663</b>	1.080	*
<b>PLANNED</b>	<b>-0.54313</b>	0.581	*	<b>PLANNED</b>	<b>-0.54663</b>	0.579	*
<i>A_EXP_D</i>	<i>0.10351</i>	1.109		<i>A_EXP_D</i>	<i>0.10186</i>	1.107	
<b>WEEKEND</b>	<b>-0.24196</b>	0.785	*	<b>WEEKEND</b>	<b>-0.24354</b>	0.784	*
<b>A_DAYT_D</b>	<b>-0.56016</b>	0.571	*	<b>A_DAYT_D</b>	<b>-0.55914</b>	0.572	*
<b>A_P_MM_D</b>	<b>0.15873</b>	1.172	*	<b>A_P_MM_D</b>	<b>0.16111</b>	1.175	*
<b>A_P_DIST</b>	<b>-3.80E-6</b>	1.000	*	<b>A_P_DIST</b>	<b>-3.75E-6</b>	1.000	*
<i>A_N_TT</i>	<i>-0.00098</i>	0.999		<i>A_N_TT</i>	<i>-0.00094</i>	0.999	
<b>A_N_MM</b>	<b>-0.16604</b>	0.847	*	<b>A_N_MM</b>	<b>-0.16592</b>	0.847	*
<b>NR_J_SUM</b>	<b>0.00862</b>	1.009	*	<b>NR_J_SUM</b>	<b>0.00874</b>	1.009	*
<b>A_P_PUR</b>	<b>-0.03343</b>	0.967	*	<b>A_P_PUR</b>	<b>-0.03365</b>	0.967	*
<b>A_N_PUR</b>	<b>-0.08263</b>	0.921	*	<b>A_N_PUR</b>	<b>-0.08344</b>	0.920	*

n = 2985; Likelihood Ratio Chi Quadrat = 980.78    n = 2963; Likelihood Ratio Chi Quadrat = 983.11

Tabelle 9.4 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Dienstleistung allgemein

Gesamtmodell 2. Stufe				Modell bei regelmäßige Aktivitäten			
Variable	Parameter-schätzung (β)	Risk Ratio	SN	Variable	Parameter-schätzung (β)	Risk Ratio	SN
<b>SEX</b>	<b>0.21291</b>	1.237	*	<b>SEX</b>	<b>0.18257</b>	1.200	*
<i>WORKED</i>	<i>0.14855</i>	1.160		<b>WORKED</b>	<b>0.18644</b>	1.205	*
<b>AGE</b>	<b>-0.00456</b>	0.995	*	<i>AGE</i>	<i>-0.00391</i>	0.996	
<b>A_EXP_D</b>	<b>-0.27321</b>	0.761	*	<b>A_EXP_D</b>	<b>-0.24368</b>	0.784	*
<i>ACC</i>	<i>-0.13119</i>	0.877		<b>ACC</b>	<b>-0.15636</b>	0.855	*
<b>A_P_TT</b>	<b>-0.01892</b>	0.981	*	<b>A_P_TT</b>	<b>-0.02093</b>	0.979	*
<b>A_P_MM</b>	<b>-0.07578</b>	0.927	*	<b>A_P_MM</b>	<b>-0.06438</b>	0.938	*
<b>A_N_PUR</b>	<b>-0.03878</b>	0.962	*	<b>A_N_PUR</b>	<b>-0.03501</b>	0.966	*

n = 896; Likelihood Ratio Chi Quadrat = 112.20    n = 778; Likelihood Ratio Chi Quadrat = 94.33

Tabelle 9.5 : Vergleich des Gesamtmodells mit einem Modell für Personen mit regelmäßigen Aktivitäten, Kategorie Treffen mit Freunden

Gesamtmodell 2. Stufe				Modell bei regelmäßige Aktivitäten			
Variable	Parameter-schätzung (β)	Risk Ratio	SN	Variable	Parameter-schätzung (β)	Risk Ratio	SN
<b>CHILDREN</b>	<b>0.20606</b>	1.229	*	<b>CHILDREN</b>	<b>0.26743</b>	1.307	*
<b>HH_NR_L</b>	<b>0.16126</b>	1.175	*	<b>HH_NR_L</b>	<b>0.14068</b>	1.151	*
<i>PLANNED</i>	<i>-0.28199</i>	0.754		<i>PLANNED</i>	<i>-0.26096</i>	0.770	
<b>A_EXP_D</b>	<b>0.54492</b>	1.724	*	<i>A_EXP_D</i>			
<i>WEEKEND</i>				<b>WEEKEND</b>	<b>-0.21544</b>	0.806	*
<b>A_DAYT_D</b>	<b>-0.63482</b>	0.530	*	<b>A_DAYT_D</b>	<b>-0.88365</b>	0.413	*
<b>ACC</b>	<b>-0.14743</b>	0.863	*	<b>ACC</b>	<b>-0.13476</b>	0.874	*
<b>A_P_TT</b>	<b>-0.00769</b>	0.992	*	<b>A_P_TT</b>	<b>-0.00932</b>	0.991	*
<b>A_N_MM</b>	<b>-0.07597</b>	0.927	*	<b>A_N_MM</b>	<b>-0.06962</b>	0.933	*
<b>A_N_PUR</b>	<b>-0.06291</b>	0.939	*	<b>A_N_PUR</b>	<b>-0.06287</b>	0.939	*
<b>NR_A_SUMC</b>	<b>-0.01233</b>	0.988	*	<b>NR_A_SUMC</b>	<b>-0.01348</b>	0.987	*

n = 669; Likelihood Ratio Chi Quadrat = 120.37    n = 519; Likelihood Ratio Chi Quadrat = 114.74

Die Unterschiede zwischen den Aktivitäten aller Personen und jenen mit regelmäßigen Aktivitäten sind jedoch nur sehr geringfügig. So ist in den Kategorien Einkauf kurzfristig und Freizeit (Tabelle 9.2 und Tabelle 9.3) in der Signifikanz der Variablen keinerlei Unterschied festzustellen und nur die Parameterschätzwerte und Hazardraten geringfügig verschieden.

In der Kategorie Dienstleistung allgemein sind im Gesamtmodell die Variablen „gearbeitet ja/nein“ und „Begleitung ja/nein“ nur im 90% Konfidenzintervall, bei Personen mit regelmäßigen Dienstleistungsaktivitäten sind diese Variablen stark signifikant und die Variable „Alter“ hingegen marginal signifikant (Tabelle 9.4).

In der Kategorie Freizeit hat bei jenen Personen, die sich regelmäßig mit ihren Freunden treffen, die Frage, ob Ausgaben getätigt worden sind, oder nicht, keinen Einfluß auf die Dauer der Aktivität (Tabelle 9.5). Ein möglicher Grund für dieses Ergebnis ist die Tatsache, daß diese Treffen sehr regelmäßig stattfinden und dabei der Ort bzw. die Art des Treffpunkts ziemlich konstant bleibt. Trifft man sich zweimal wöchentlich beim Wirt zum Kartenspielen mit seinen Freunden, so gibt man auf jeden Fall Geld aus, meist sogar in ähnlicher Menge. Die Dauer der Aktivität ist bei solch regelmäßigen Treffen auch ziemlich konstant. Anders verhält es sich bei Personen mit sporadischen Treffen mit Freunden, die auch einen Einfluß auf das Gesamtmodell haben. Bei ihnen verkürzt die Ausgabe von Geld während der Aktivität die Dauer derselben enorm, so daß im Gesamtmodell eine fast doppelt so lange (Faktor 1,7) Aktivitätszeit für jene Aktivitäten, bei denen Ausgaben gemacht worden sind, erwartet wird. Genau umgekehrt verhält es sich bei der Frage, ob die Aktivität am Wochenende gemacht wurde oder nicht. Bei Personen mit regelmäßigen Treffen wird am Wochenende eine geringfügig längere Aktivitätsdauer erwartet, beim Gesamtmodell hat diese Variable jedoch keinen signifikanten Einfluß.

Bei der Modellschätzung für sporadische Aktivitätsteilnehmer (näheres im Zusatzanhang XXI) sind meist keine oder nur sehr wenige Variablen signifikant die Aktivitätsdauer beeinflussend. Die Güte dieser Modelle ist aber aufgrund ihrer sehr geringen Anzahl an Ereignissen eher schlecht und die Ergebnisse daher zu vernachlässigen.

Bei der Aktivitätskategorie Einkauf fällt auf, daß relativ viele der untersuchten soziodemografischen Variablen einen Einfluß auf die Aktivitätsdauer nehmen, bei den anderen Kategorien sind es nur noch zwei (Freizeit allgemein) oder drei (Dienstleistung allgemein und Treffen mit Freunden).

## 9.2.4 Vollparametrische Modelle mit *PROC LIFEREG*

Die maßgebende Änderung gegenüber den semiparametrischen Modellschätzungen ist bei den vollparametrischen Modellen die Annahme einer geeigneten grundlegenden Verteilungsfunktion (*baseline hazard distribution*), die multiplikativ auf die Kovariablen wirkt (vgl. Formeln im Kapitel 6.4). Es wird angenommen, daß diese Verteilungsannahme teilweise von Verhaltensmustern beeinflusst wird (Mannering *et al.*, 1994). Die verschiedenen Möglichkeiten der Verteilungsannahme sind im Kapitel 6.4.1 genauer beschrieben. Grundsätzlich können auch hier alle 5 verschiedenen Modelle in Betracht gezogen werden bzw. mit dem generalisierten Gammamodell verglichen werden. Aufgrund mehrerer Kriterien und Entscheidungsgrundlagen soll für jede Kategorie jene Verteilungsannahme getroffen werden, die am besten den wahren Verlauf der Hazard- und der Survivalkurve widerspiegelt.

In den meisten bisherigen Studien wird grundsätzlich die Annahme einer Weibullverteilung getroffen. Es wird gesagt, sie komme an die Anforderungen der verkehrsspezifischen Belange zumeist am nächsten heran bzw. könne man damit reisebezogene Entscheidungen laut Verkehrsverhaltensforschern am besten beschreiben. Oh (1999) entschied sich beispielsweise aufgrund des grafischen Vergleichs der theoretischen Verteilungskurven der einzelnen von ihm näher in Betracht gezogenen Modellansätze (Weibullverteilung, Exponentialverteilung und Log-Logistische Verteilung) mit den tatsächlichen durch nonparametrische Auswertungen gefundenen Kurven der Hazard- und Survivalfunktion für die Weibullverteilung. Zusätzlich verdeutlicht er die grafische Entscheidung durch eine Angabe der maximalen, minimalen und Standard-Abweichung der jeweiligen Verteilungskurve von der empirischen Kurve. Die Wahl dieses Modells geht auch mit früheren aktivitätsbasierenden Verkehrsverhaltensstudien (Hamed & Mannering, 1993 und Bhat, 1996) konform. Filiou (1994) und Wilson (1994) bevorzugen ebenfalls die Weibullverteilung als grundlegende Verteilungsannahme einerseits aufgrund der Einfachheit der Verteilungsfunktion und andererseits aufgrund der engen Zusammenhänge von Extremwertverteilungen mit Bedarfsaufbau und -abdeckung, was mit Hilfe der Weibullverteilung gut dargestellt werden kann (Mannering, 1993). Schönfelder und Axhausen (2000) beziehen sich bei ihrer Wahl des Weibullmodells als grundlegende Verteilungsfunktion hauptsächlich auf bisherige Studien (Hamed und Mannering, 1993; Hensher und Mannering, 1994; Bhat, 1996).

Geht man jedoch strikt nach den Entscheidungskriterien, die für die Paßform der verschiedenen Annahmen der einzelnen Modelle ausschlaggebend sind, so treten vereinzelt Unterschiede auf bzw. findet man Modellansätze, die eventuell die tatsächlichen Verläufe der Verkehrsverhaltensentscheidungen besser beschreiben können. Deshalb werden zur Evaluierung der hier vorgestellten Modelle einige Überprüfungsmethoden angewandt.

Zuerst werden mit Hilfe von Likelihood-Ratio-(Chi-Quadrat)-Statistiken (Allison, S. 88 ff.) miteinander vernetzte Modelle verglichen. Dabei werden die Log-Likelihood-werte berechnet, die signifikante Unterschiede zwischen dem jeweiligen Submodell und dem übergeordneten generalisierten Gammamodell, das aufgrund seiner großen Bandbreite an möglichen Verteilungsformen die Charakteristik der Daten meist recht gut beschreibt, aufzeigen. Einzig die Log-logistische Verteilung kann damit nicht überprüft werden, da sie mit den restlichen Verteilungsmodellen nicht verknüpft ist. Aufgrund dieser Statistiken kann in allen vier betrachteten Kategorien das Modell der Exponentialverteilung jeweils ausgegliedert werden.

Danach werden mit Hilfe grafischer Methoden (Allison, S. 91ff.) die für den jeweiligen Modellansatz ausschlaggebenden Überprüfungen durchgeführt: Der Verlauf der durch nonparametrische Auswertungen gefundenen Log-Survival-Kurve zeigt, ob das exponentielle Modell passen würde, eine möglichst lineare Log-Logsurvival-Kurve bezeichnet eine gute Paßform für einen Modellansatz der Weibullverteilung; Die Kurven für das Log-logistische Modell und das Log-normale Modell, die mathematisch jeweils etwas komplizierter definiert sind, sollten ebenfalls jeweils möglichst linear verlaufen, um die Aussage treffen zu können, daß das Modell gut paßt.

Eine Einschränkung liegt mit der Annahme vor, daß die Stichprobe aus einer homogenen Bevölkerungsverteilung gezogen worden ist und keine Kovariablen mit der Überlebenszeit verknüpft sind. Es kann dadurch passieren, daß Modelle, deren Kurvenverläufe ohne Berücksichtigung von Kovariablen gut aussehen, in der Praxis das wirkliche Verkehrsverhalten mit einflußnehmenden Kovariablen dennoch nicht gut genug abbilden. Aus diesem Grunde können Grafiken der Residuen (*residuals*) aus den Regressionsmodellen erstellt werden, die unabhängig vom verwendeten Modelltyp und der Berücksichtigung von Kovariablen verwendet und interpretiert werden können. Mehrere Arten von Residuenplots stehen zur Auswahl, am geeignetsten für unsere Bedürfnisse erscheinen die sogenannten Cox-Snell Residuen (*residuals*) (Allison, S.94). Auf eine genauere Definition wird hier verzichtet. Ähnlich wie bei der Überprüfung der verschiedenen Grafikkurven sollten auch hier die Kurven der Residuenplots möglichst linear verlaufen, sodaß das gewählte Modell eine gute Paßform hat. Wie Allison in seinen Studien anführt, scheint diese Überprüfungsmethode in der Theorie recht gut zu sein, in der Praxis jedoch nur geringe Unterschiede zu zeigen.

Exemplarisch werden nun die einzelnen Schritte der Überprüfung, wie oben genannt, für die Kategorie Einkauf kurzfristig durchgemacht und wie folgt dargestellt:

Nachdem die Parameterschätzwerte von *scale* und *shape* des generalisierten Gammamodells viel zu weit auseinander liegen, wird das standardisierte Gammamodell von vornherein nicht in die Überprüfungen mit einbezogen. Der *shape*-Parameterwert liegt zudem unter bzw. knapp bei Null, was auf eine

gute Näherung durch das lognormale Modell bzw. eine schlechtere Annäherung durch das Weibullmodell (Wert sollte dann knapp bei 1 liegen) hindeuten würde. Es werden also Weibullmodell, Exponentialmodell, Loglogistisches und Lognormales Modell getestet. Zuerst werden die Log-Likelihoods für die einzelnen Modelle und die Log-Likelihood-Ratio-Chi-Quadrat-Statistik erstellt, wodurch man diese Werte erhält:

Tabelle 9.6 : Log-Likelihoods der Modelle:

Art der Verteilung	LL ohne Kovariablen	LL mit Kovariablen
WEIBULL	-4877.09	-4450.42
EXPONENT	-4881.48	-4477.69
LOGLOGISTC	-4469.60	-3976.89
LOGNORMAL	-4485.48	-4033.47
G. GAMMA	-4478.18	-4024.20

Tabelle 9.7 : Log-Likelihood-Ratio-Chi-Quadrat-Statistik:

LL-Ratio ohne Kovariablen	LL-Ratio mit Kovariablen	Verteilung	vs.	Verteilung
8.78	54.54	Exponential	vs.	Weibull
806.60	906.98	Exponential	vs.	G. Gamma
797.82	852.44	Weibull	vs.	G. Gamma
14.60	18.54	Lognormal	vs.	G. Gamma

Daraus ergibt sich die Schlußfolgerung, daß die Annahme einer Lognormalen Verteilung die günstigste Lösung wäre, die oft verwendete Weibullverteilung scheidet hier eindeutig aus. Wie vorher erwähnt, kann das Loglogistische Verteilungsmodell mit dieser Methode jedoch nicht überprüft werden. Der Loglikelihoodwert (-3976.89) liegt zudem noch unter dem des generalisierten Gamma-Modells (-4024.20) und auch aufgrund anderer Kriterien würde sich das Modell an sich recht gut eignen.

Aus den grafischen Überprüfungen der Kurven für das Lognormale bzw. das Loglogistische Modell (Abbildung 9.23 und 9.24) kann nicht mit eindeutiger Überzeugung das bessere gefunden werden. Beide Kurven haben einen annähernd linearen Verlauf, nur im obersten Teil weicht die loglogistische Kurve mehr von der gewünschten Linearität nach oben ab.

Abbildung 9.23 : Plot für die Evaluierung des Loglogistischen Modells, Kategorie Einkauf kurzfristig

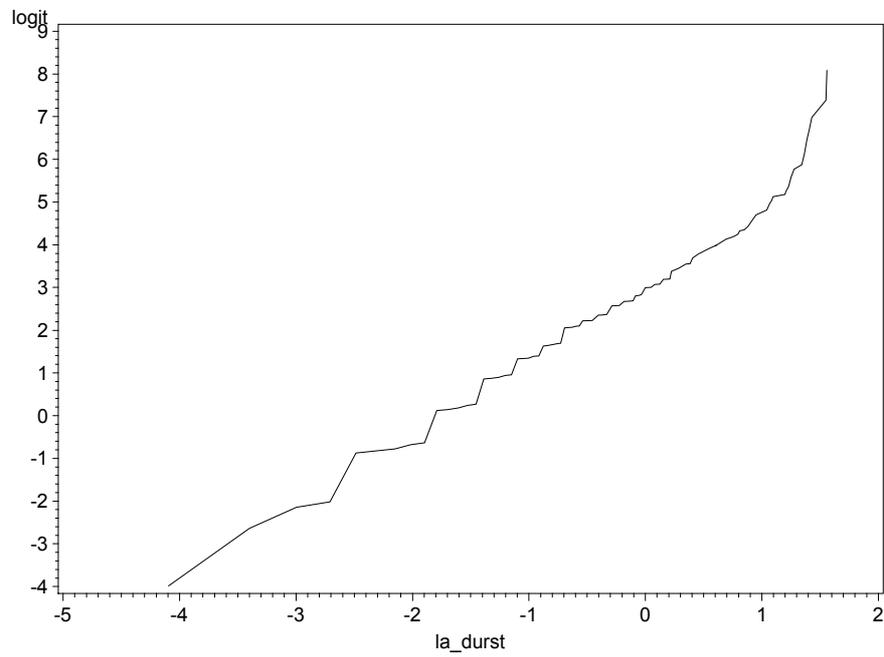
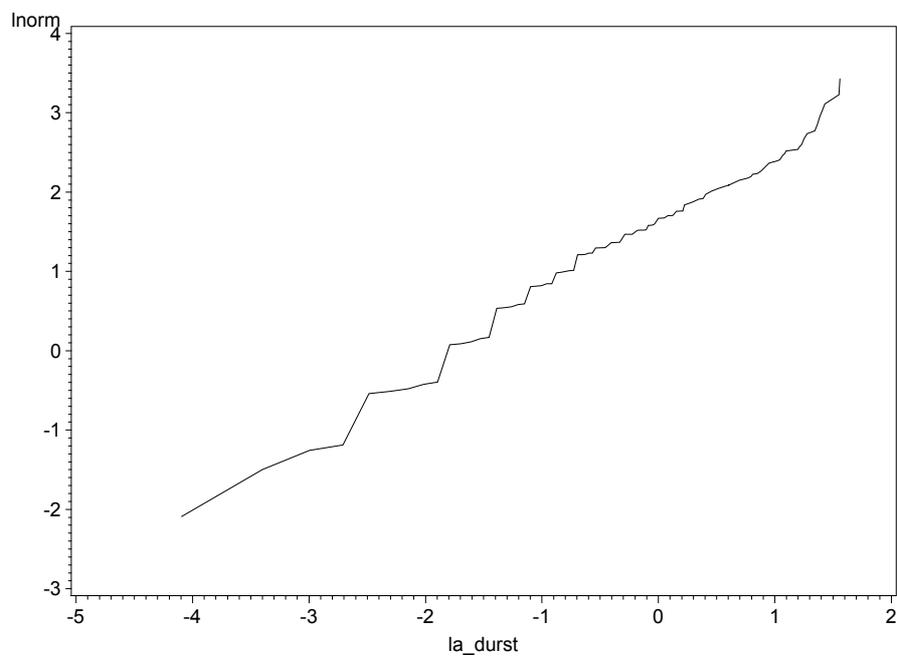


Abbildung 9.24 : Plot für die Evaluierung des Lognormalen Modells, Kategorie Einkauf kurzfristig



Auch der Vergleich der Cox-Snell-Residuen (Abbildungen 9.25 und 9.26 ) ergibt keine eindeutige Entscheidungsgrundlage, beide Kurven verlaufen ähnlich linear. Etwas besser wird die geforderte Linearität über den gesamten Kurvenverlauf betrachtet von der loglogistischen Kurve getroffen.

Abbildung 9.25 : Cox-Snell-Residuenplot mit einer loglogistischen Verteilung, mit Kovariablen

---

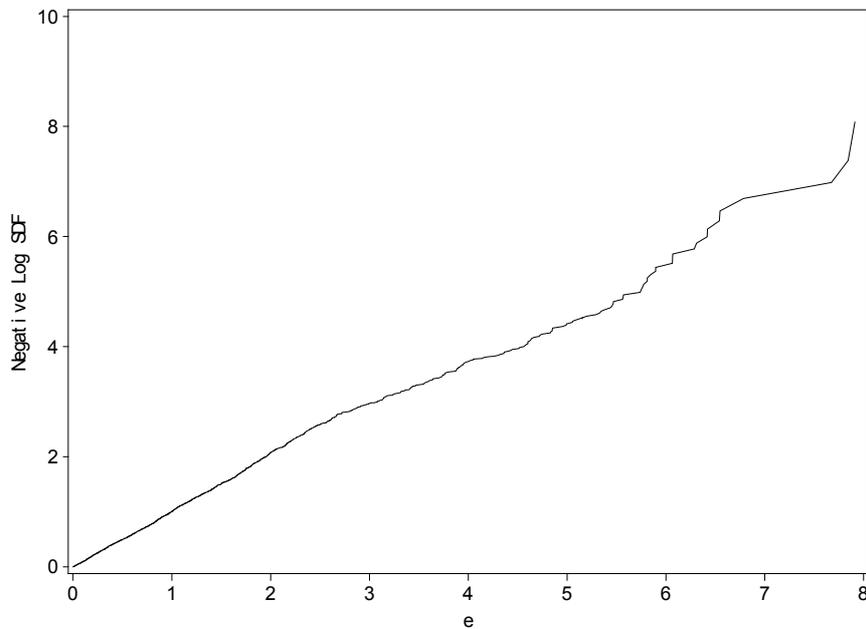
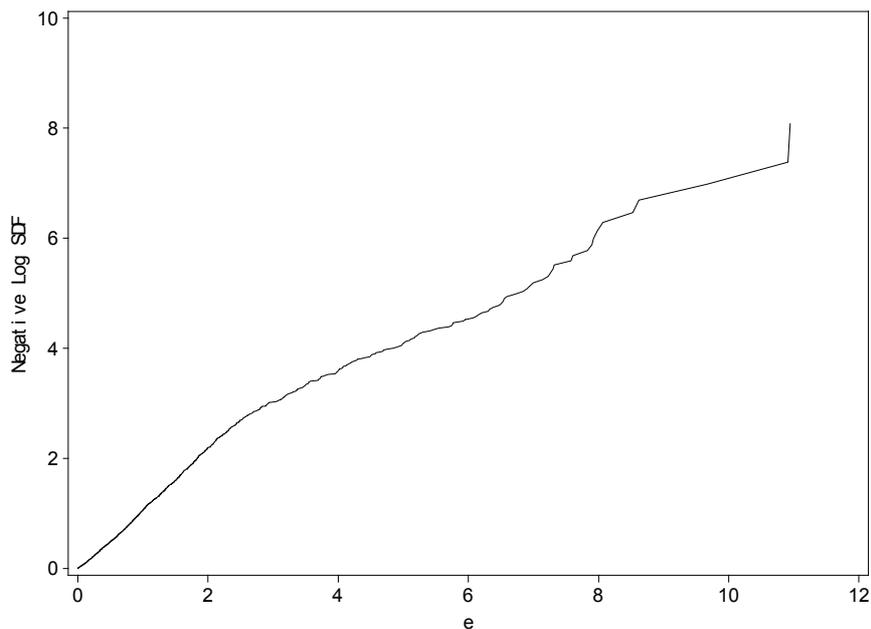


Abbildung 9.26 Cox-Snell-Residuenplot mit einer lognormalen Verteilung, mit Kovariablen

---



Zu guter Letzt wird noch die mit Hilfe des *Lifehaz*-Macros (nach Allison) erzeugte Hazardkurve der Modelle mit loglogistischer bzw. lognormaler Verteilung (siehe Abbildung 9.27 und 9.28) mit der in den nonparametrischen Auswertung gefundenen Hazardkurve verglichen. Dabei ergeben sich ebenfalls nur sehr geringfügige Unterschiede zwischen den beiden Modellen, sodaß die Entscheidung zu Gunsten eines loglogistischen Verteilungsmodelles hauptsächlich aufgrund des wesentlich besseren Loglikelihoodwertes fällt.

Abbildung 9.27 : LIFEREG- Auswertung: Plot der Hazardkurve vs Zeit, Loglogistic, mit Kovariablen

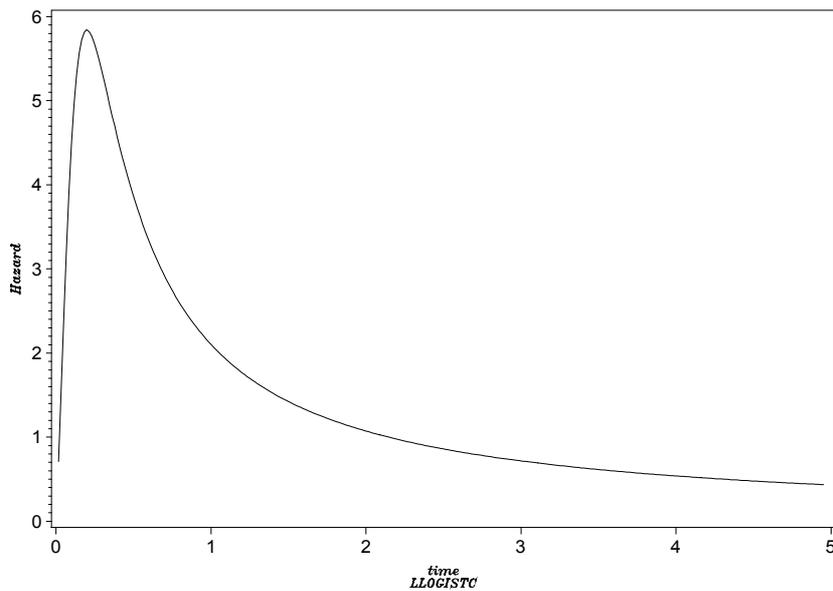
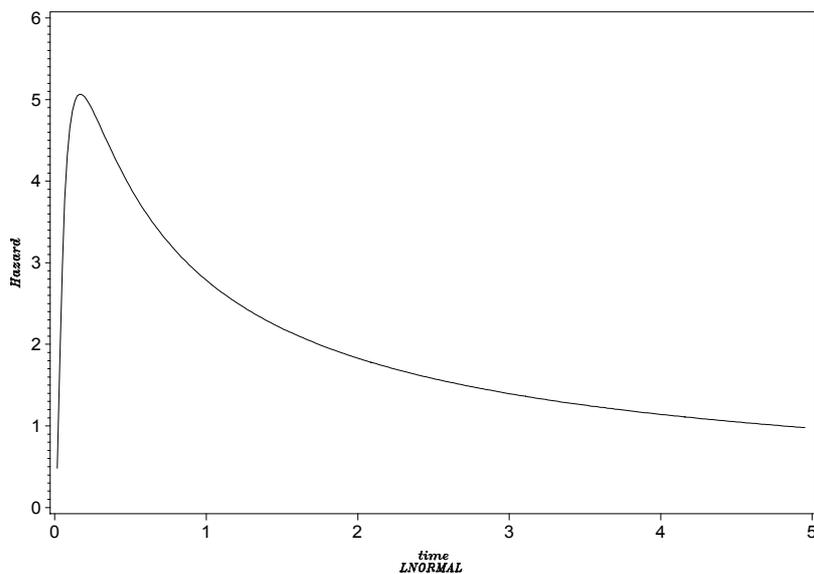


Abbildung 9.28 : LIFEREG- Auswertung: Plot der Hazardkurve vs Zeit, Lognormal, mit Kovariablen



Mit dieser nun aufgrund aller gemachten Überprüfungen gewählten Loglogistischen Annahmeverteilung werden alle weiteren Auswertungen für die Kategorie Einkauf kurzfristig gemacht.

In den restlichen drei Kategorien wird analog vorgegangen, die detaillierten Ergebnisse der einzelnen Überprüfungen finden sie im Zusatzanhang XXII (LIFEREG AD Vollparametrische Modelle).

In der Kategorie Freizeit allgemein liegen bei der Überprüfung eines generalisierten Gammamodells die Parameterschätzwerte für shape und scale nicht extrem weit auseinander, d. h., daß man ein standardisiertes Gammamodell erstellen kann. Dessen Parameterwerte für shape und scale werden mit 0.86 festgelegt, da dann der maximale Wert für den Loglikelihood erzielt wird. Die Log-Likelihood-Ratio-Chi-Quadrat-Statistik läßt die Annahme eines Weibullmodells und eines standardisierten Gammamodells zu. Das Exponentialmodell, das aufgrund der Parameterwerte des generalisierten Gammamodells ebenfalls zur Auswahl steht, muß nach der Log-Likelihood-Ratio-Chi-Quadrat-Statistik mit Kovariablen signifikant zurückgewiesen werden. Der Loglikelihoodwert für ein loglogistisches Modell ist zu klein, um das Modell ebenfalls in die Überlegungen mit einzubeziehen. Bei der grafischen Überprüfung der ausschlaggebenden Log-logsurvivalkurve für die Weibullverteilung zeigt sich eine annehmbare gute Annäherung an eine lineare Form. Die zweite mögliche Verteilung, die eines standardisierten Gammamodells, kann mit Hilfe dieser Kurven nicht überprüft werden, ebensowenig wie mit Hilfe der Cox-Snell-Residuenplots. Die Cox-Snell-Residuenkurve für die Weibullverteilung mit Kovariablen nähert sich zwar im vorderen Teil sehr gut einer linearen Kurve an, im letzten Abschnitt weicht sie jedoch sehr stark nach oben ab. Aus diesem Grund ist, wie auch schon aufgrund des besseren Loglikelihoodwertes des standardisierten Gammamodells, das letztgenannte Annahmeverteilungsmodell vorzuziehen. Die beiden Werte für die Parameter shape und scale werden mit jeweils 0.86 festgelegt. Die weiteren Auswertungen in der Kategorie Freizeit allgemein werden deshalb mit der Annahme einer standardisierten Gammaverteilung gemacht.

In der Kategorie Dienstleistung allgemein liegen ähnlich wie in der Kategorie Einkauf kurzfristig die Werte von shape und scale beim generalisierten Gammamodell sowohl mit, als auch ohne Kovariablen, relativ weit auseinander (-0.16 und 1.02 ohne Kovariablen bzw. -0.20 und 0.94 mit Kovariablen), wodurch die Erstellung eines standardisierten Gammamodells wenig Sinn hat. Der Shape-Parameter liegt relativ knapp bei 0, was auf ein Naheverhältnis zum lognormalen Modell schließen läßt. Nach der Log-Likelihood-Ratio-Chi-Quadrat-Statistik scheiden, wie schon nach der Überprüfung der shape- und scale-Parameterwerte des generalisierten Gammamodells vermutet (Weibull und Exponentialverteilung wären nur bei Shape-Parameterwerten von ungefähr 1 gute Annäherungen), diese beiden Verteilungsmöglichkeiten aus. Die in dieser Statistik nicht betrachtete loglogistische Verteilungsannahme hat hingegen, ebenso wie die lognormale Verteilung sehr gute Loglikelihoodwerte. Aus den beiden Kurven für die Evaluierung des loglogistischen Modells bzw. des lognormalen Modells läßt sich keine

endgültige Entscheidung treffen, etwas linearer über den ganzen Bereich verläuft jedoch die Lognormalverteilungskurve. Speziell im letzten Abschnitt weicht die Loglogistische Kurve etwas mehr nach oben ab. Auch die Cox-Snell-Residuenplots der beiden möglichen Verteilungen ergeben, daß die Kurve des loglogistischen Modells vor allem im letzten Teil etwas weiter nach oben von der linearen Form abweicht. Die Wahl fällt hier auch aufgrund des besseren Loglikelihoodwert auf die eines Lognormalen Annahmeverteilungsmodells, mit dem alle weiteren Auswertungen gemacht werden.

Beim generalisierten Gammamodell in der Kategorie Treffen mit Freunden liegt der Shape-Parameter relativ nahe bei 1 und zudem nahe am Scale-Parameter (0.87 bzw. 0.83). Aufgrund dieser Werte ist sowohl die Exponentialverteilung, als auch die Weibullverteilung, als auch die standardisierte Gammaverteilung möglich, Loglogistisches und Lognormales Modell scheiden hingegen aus. Aufgrund der Likelihood-Ratio-Chi-Quadrat-Statistik steht aber auch das Exponentialmodell nicht mehr zur Diskussion, es muß signifikant abgelehnt werden. Sowohl das standardisierte Gammamodell, als auch das Weibullmodell weisen hingegen eine sehr gute Statistik auf. Die Überprüfung der Kurven (Log-LogSurvivalkurve bzw. Cox-Snell-Residuenplot), die nur für das Weibullmodell gemacht werden können, ergeben jeweils einen relativ geradlinigen, linearen Verlauf. Die Annahme eines solchen Verteilungsmodells wäre also durchaus gerechtfertigt. Nachdem auch der Loglikelihoodwert des Weibullmodells sehr gut aussieht, kann das Modell (anstatt eines standardisierten Gammamodells) für die weiteren Auswertungen herangezogen werden.

Man sieht also, daß im Gegensatz zu vielen bisherigen Studien in nur einem Fall die Annahme einer Weibullverteilung die geeignetste Methode darstellt. In den Kategorien Einkauf kurzfristig und Dienstleistung allgemein verfehlt diese Möglichkeit vollständig ihr Ziel, in der Kategorie Freizeit wäre sie eventuell denkbar, im Spezialfall Treffen mit Freunden sinnvoll.

Ähnlich wie bei den semiparametrischen Auswertungen werden auch hier in jeder Kategorie Korrelationsmatrizen erstellt und eine Einzelanalyse der Variablen durchgeführt und aufgrund dieser Ergebnisse entschieden, welche möglichen Einflußfaktoren als Kovariablen in einem ersten Gesamtmodell verwendet werden. Die Ergebnisse der jeweiligen ersten Gesamtmodelle sind im Anhang VIII zu sehen. Um eine gute Gegenüberstellung zu gewährleisten, sind im Anhang VII zusätzlich die ersten Gesamtmodelle der semiparametrischen Auswertungen für die vier Kategorien dargestellt.

In einem weiteren Schritt wird wiederum in jeder Kategorie ein Modell der zweiten Stufe erstellt, das nur noch die im ersten Gesamtmodell statistisch signifikanten Kovariablen enthält. Diese Ergebnisse werden in den Tabellen 9.8 bis 9.11 präsentiert und kurz diskutiert.

Tabelle 9.8 : Modell der zweiten Stufe, Kategorie Einkauf kurzfristig

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	-1.49682	0.10363	208.6242	<.0001
<b>SEX</b>	-0.24875	0.03430	52.6046	<.0001
<b>LICENSE</b>	-0.32622	0.04569	50.9863	<.0001
<b>G</b>	-0.03262	0.00921	12.5374	<b>0.0004</b>
<b>CAR</b>	0.17029	0.04572	13.8697	<b>0.0002</b>
<b>PLANNED</b>	0.19893	0.04498	19.5622	<.0001
<b>A_P_MM_D</b>	-0.14651	0.05523	7.0377	<b>0.0080</b>
<b>A_N_MM_D</b>	-0.20377	0.05231	15.1748	<.0001
<b>A_P_MMMD</b>	0.18343	0.06021	9.2797	<b>0.0023</b>
<b>A_N_MMMD</b>	0.12180	0.05867	4.3097	<b>0.0379</b>
<b>ACC</b>	0.10800	0.02732	15.6330	<.0001
<b>A_EXP</b>	0.00696	0.00051	188.4759	<.0001
<b>A_DAYT</b>	-0.11476	0.02688	18.2319	<.0001
<b>A_P_DIST</b>	-0.00001	5.33E-6	5.4540	<b>0.0195</b>
<b>A_P_TT</b>	0.01626	0.00177	84.2240	<.0001
<b>A_N_TT</b>	0.00435	0.00127	11.6811	<b>0.0006</b>
<b>A_P_PUR</b>	0.01933	0.00543	12.6800	<b>0.0004</b>
<b>NR_A_SUMC</b>	-0.00924	0.00094	96.1044	<.0001
<b>SCALE</b>	0.46432	0.00686		

n = 3221; Log-Likelihood = - 3982.59 , Verteilungsannahme: Loglogistisch

Tabelle 9.9 : Modell der zweiten Stufe, Kategorie Freizeit allgemein

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	-1.12021	0.10660	110.4340	<.0001
<b>HH_MEMB</b>	-0.07242	0.01467	24.3788	<.0001
<b>PLANNED</b>	0.49290	0.05837	71.2975	<.0001
<i>A_EXP_D</i>	-0.09239	0.04812	3.6867	<i>0.0548</i>
<b>WEEKEND</b>	0.22523	0.03401	43.8631	<.0001
<b>A_DAYT_D</b>	0.51494	0.04033	162.9980	<.0001
<b>A_P_MM_D</b>	-0.13311	0.05337	6.2201	<b>0.0126</b>
<b>A_P_DIST</b>	3.21E-6	8.48E-7	14.3190	<b>0.0002</b>
<i>A_N_TT</i>	0.00087	0.00046	3.5469	<i>0.0597</i>
<b>A_N_MM</b>	0.14907	0.01023	212.1954	<.0001
<b>NR_J_SUM</b>	-0.00748	0.00068	121.5626	<.0001
<b>A_P_PUR</b>	0.02832	0.00709	15.9599	<.0001
<b>A_N_PUR</b>	0.07722	0.00792	94.9691	<.0001
<b>SCALE</b>	0.86000	0		
<b>SHAPE</b>	0.86000	0		

n = 2985; Log-Likelihood = - 4137.07 , Verteilungsannahme: Standardisierte Gammverteilung

Tabelle 9.10 : Modell der zweiten Stufe, Kategorie Dienstleistung allgemein

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	-2.12102	0.17883	140.6784	<b>&lt;.0001</b>
<b>SEX</b>	-0.20891	0.07208	8.4002	<b>0.0038</b>
<b>A_EXP_D</b>	0.29675	0.07451	15.8602	<b>&lt;.0001</b>
<b>A_P_TT</b>	0.02086	0.00339	37.9485	<b>&lt;.0001</b>
<b>A_P_MM</b>	0.07177	0.02332	9.4735	<b>0.0021</b>
<b>NR_T_SUM</b>	-0.00156	0.00064	6.0588	<b>0.0138</b>
<i>WORKED</i>	-0.13016	0.07790	2.7917	<i>0.0948</i>
CAR_POSS	0.08669	0.08321	1.0855	0.2975
AGE	0.00313	0.00211	2.2002	0.1380
<b>A_N_TT</b>	0.00527	0.00260	4.1157	<b>0.0425</b>
SCALE	0.94771	0.02239		

n = 896; Log-Likelihood = - 1223.25 , Verteilungsannahme: Lognormal

Tabelle 9.11 : Modell der zweiten Stufe, Kategorie Treffen mit Freunden

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	-0.40465	0.19644	4.2431	<b>0.0394</b>
<b>CHILDREN</b>	-0.16269	0.08140	3.9950	<b>0.0456</b>
<b>HH_NR_L</b>	-0.12884	0.04603	7.8351	<b>0.0051</b>
<b>PLANNED</b>	0.28808	0.12029	5.7354	<b>0.0166</b>
<b>A_EXP_D</b>	-0.48593	0.17121	8.0556	<b>0.0045</b>
<b>A_DAYT_D</b>	0.54355	0.08718	38.8757	<b>&lt;.0001</b>
<b>ACC</b>	0.12201	0.04549	7.1945	<b>0.0073</b>
<b>A_P_TT</b>	0.00637	0.00183	12.1776	<b>0.0005</b>
<b>A_N_MM</b>	0.05185	0.02062	6.3242	<b>0.0119</b>
<b>A_N_PUR</b>	0.04858	0.01513	10.3041	<b>0.0013</b>
<b>NR_A_SUMC</b>	0.00899	0.00250	12.9697	<b>0.0003</b>
SCALE	0.81123	0.02452		

n = 669; Log-Likelihood = - 910.53 , Verteilungsannahme: Weibull

Grundsätzlich weisen die hier erzielten Ergebnisse keine gravierenden Unterschiede zu denen der semiparametrischen Auswertungen auf, einzig das Vorzeichen ist bei allen Werten umgedreht. Dies ist jedoch mit der Art der Angabe bei den beiden Modellschätzungsmethoden zu erklären. Während bei den semiparametrischen Modellen die Parameterschätzwerte im Log-Survival-Zeitformat angegeben werden, wird bei der vollparametrischen Methode das Log-Hazard-Format angewandt (Allison, S.117). Ein negatives Vorzeichen vor dem Parameterschätzwert bedeutet hier also eine negative Beeinflussung der Aktivitätsdauer bzw. eine Verkürzung derselben.

Wie im Kapitel 9.2.3 hat auch bei den vollparametrischen Modellschätzungen die Frage, ob bzw. wieviel Geld auf der Aktivität ausgegeben worden ist, einen signifikanten Einfluß auf die Aktivitätsdauer. In den Kategorien Einkauf kurzfristig und Dienstleistungen allgemein verlängern Ausgaben die Dauer der Aktivität, Freizeitaktivitäten dauern bei finanziellen Aufwendungen kürzer an, Treffen mit Freunden sogar deutlich kürzer. Die Dauer der Reisezeit entweder vor oder nach der Aktivität, die meist recht eng miteinander verknüpft sind und deshalb nicht immer beide ins Modell genommen werden können, beeinflußt die Aktivitätsdauer ebenfalls in allen vier Kategorien signifikant. Je länger die Reisezeit, desto länger ist auch in allen Kategorien die Aktivitätsdauer.

Auch bei den vollparametrischen Auswertungen hat die Tageszeit der Aktivität nur in der Kategorie Dienstleistung keinen signifikanten Einfluß. Kurzfristige Einkäufe dauern umso kürzer an, je später am Tag sie stattfinden, allgemeine Freizeitaktivitäten und auch spezielle wie Treffen mit Freunden sind jedoch meist nachmittag bzw. abends signifikant um ca. 50% länger.

Wie bei den semiparametrischen Modellschätzungen läßt sich auch hier nur ein relativ geringer Einflußfaktor von soziodemografischen bzw. beschreibenden Charakteristika feststellen. Das Geschlecht hat bei Einkaufs- und Dienstleistungsaktivitäten einen Einfluß, bei Freizeitaktivitäten jeglicher Art aber nicht. Weitere beschreibende bzw. individuelle Faktoren wie Führerschein- oder Autobesitz, die eine größere Flexibilität auch in bezug auf die Aktivitätsdauer erwarten lassen, der Ausbildungsgrad, das Anstellungsverhältnis oder das Alter der betreffenden Person nehmen nur vereinzelt Einfluß auf die Dauer der Aktivität.

Bei den kurzfristigen Einkäufen haben bei einem vollparametrischen Modell mit loglogistischer Verteilung noch mehr mögliche Kovariablen (17 anstatt 14) einen signifikanten Einfluß als bei den semiparametrischen Auswertungen, speziell die vor der Aktivität zurückgelegte Distanz und das benutzte Verkehrsmittel. Bei den allgemeinen Freizeitaktivitäten sind nicht nur dieselben möglichen Einflußgrößen signifikant, sondern die Werte nahezu identisch. Ebenso verhält es sich in der speziellen Freizeitkategorie „Treffen mit Freunden“. In der Kategorie der allgemeinen Dienstleistungen sind bei beiden Hazardmodellarten nur relativ wenige signifikante Einflußgrößen zu erkennen. Die Variablen

Alter und Art der nächstfolgenden Aktivität gehören bei den vollparametrischen Modellen nicht dazu, die Variablen Anzahl aller Wege und Reisezeit nach der Aktivität hingegen schon.

*Heterogenitätseffekte und State dependence:*

Eine Vereinfachung, die hier getroffen worden ist, unbeobachtete bzw. nicht berücksichtigte Heterogenitäten nicht in die Modellschätzungen mit ein zu beziehen, kann grundsätzlich rein theoretisch zu irreführenden Schlußfolgerungen in bezug auf die Form der Hazardkurve und die Koeffizientenabschätzungen der Kovariablen führen. Die Annahme, die implizit bei proportionalen Hazardmodellen getroffen wird, ist jene, daß die Überlebensfunktion  $[S(t)=Pr\{T \geq t\}]$  homogen über den gesamten beobachteten Bevölkerungsquerschnitt verteilt ist (Hensher und Mannering, 1994). So wird auch weiters angenommen, daß die gesamte Variation bei den Aktivitätsdauern mit dem Kovariablenvektor  $\beta$  beschrieben wird. Unbeobachtete oder bisher nicht berücksichtigte Heterogenitäten treten aber dann auf, wenn andere, nicht im Vektor  $\beta$  enthaltene Faktoren, die Aktivitätsdauer beeinflussen. Dies können sowohl allgemeine Heterogenitäten zwischen den einzelnen Befragten, als auch nicht berücksichtigte Effekte vorhergehender Aktivitäts- oder Intervalldauern (sogenanntes *state dependence*) sein.

Nachdem jedoch einerseits die Stichproben bei der diesen Auswertungen zugrunde liegenden Umfrage mit Hilfe des schwedischen Bevölkerungsregisters nach einer vorherigen Kategorisierung der Befragungsteilnehmer in sechs verschiedene Bevölkerungsgruppen gezogen wurden, und andererseits bei einigen dieser Arbeit vorangegangenen Studien (vgl. Oh, 1999) keinerlei gravierende Veränderungen der Ergebnisse durch die Berücksichtigung von Heterogenitätseffekten festgestellt werden konnten, wird hier auf eine genauere Betrachtung solch möglicher Effekte verzichtet. Wie man vor allem bei den Studien von Oh sieht, sind zwar sowohl bei Exponential, als auch bei Weibull-verteilten Modellen geringe Heterogenitätseffekte sichtbar, es verbessern sich dadurch die Loglikelihoodwerte bei Konvergenz etwas, die Hazard und Survivalkurven verändern sich jedoch nur unwesentlich.

## 9.3 DAUER DER INTERVALLE ZWISCHEN GLEICHARTIGEN AKTIVITÄTEN

### 9.3.1 Einleitung

Die aus den Rohdaten der UTHS (siehe Kapitel 2) erzeugte Aktivitätsdatei, in der die Dauern der einzelnen Aktivitäten vorhanden sind, ist nochmals verändert worden und es sind zusätzliche, speziell intervallbezogene Variablen generiert worden. So scheinen in den Intervalldateien, die einmal gesamt und zudem für jede einzelne der ausgewählten Kategorien vorliegen, die Dauer der Zeitspanne zwischen zwei beliebigen aufeinanderfolgenden Aktivitäten (*intervall*) und zwischen zwei gleichartigen Aktivitäten (*inter*) auf. Zusätzlich sind eine Reihe erklärender Variablen, sowohl soziodemografischer als auch aktivitätsbezogener Natur vorhanden, die einen Einfluß auf die abhängige Variable (Intervalldauer zwischen zwei Aktivitäten derselben Kategorie in Tagen, *inter\_t*) haben könnten.

Wie auch schon im Kapitel 9.2 werden nach mehreren Auswahlverfahren mit Hilfe gewisser Hierarchien und Entscheidungskriterien (beispielsweise Einzelvariablenanalyse, Korrelationsmatrix, Proportionalitätsüberprüfung, etc.) alle drei verschiedenen Schritte der Modellschätzung (nonparametrisch, semiparametrisch und vollparametrisch) mit Hilfe der SAS-Programme durchgeführt und anschließend präsentiert.

Die Auswertungsschritte bei der Überprüfung möglicher Einflußvariablen auf die Dauer der Intervalle zwischen gleichartigen Aktivitäten sind stark an jene Schritte der im Kapitel 9.2 beschriebenen Modelle für die Aktivitätsdauern angelehnt und werden deshalb nicht noch einmal separat erläutert.

### 9.3.2 Nonparametrische Modelle mit *PROC LIFETEST*

Nonparametrische Modelle dienen vorrangig zur möglichst wahrheitsgetreuen Abbildung der Form der Survivalkurven und Hazardkurven und zusätzlich, eventuelle Unterschiede zwischen einzelnen Gruppen mit ähnlichen oder gleichen beschreibenden Variablen herauszufinden. Die Auswertungen mit *PROC LIFETEST* sind aus diesem Grunde auch eher kurz gehalten und beschränken sich auf Abbildung und Diskussion der Hazard- und Survivalkurven, eine kurze Abhandlung der Gruppenunterschiede und einen sogenannten *Log-Rank Test* bzw. *Wilcoxon Test with Forward Stepwise Sequence of Chi-Squares* von möglichen Einflußvariablen. Detailliertere Variablenüberprüfungen werden bei den semiparametrischen Modellen durchgeführt.

Zu Beginn werden für die vier ausgewählten und bei den semiparametrischen und vollparametrischen Auswertungen näher betrachteten Aktivitätskategorien Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden jeweils die Hazard- und Survivalkurve in einer gemeinsamen Grafik dargestellt (Abbildung 9.29 bis 9.32). Die Kurven sind mit Hilfe einer nonparametrischen Auswertung mit PROC LIFETEST aus den Datensätzen uppsdur.Kategorie\_int erzeugt, mit Lifehaz-Macros und Grafikprogrammen verändert und in je eine Grafik importiert worden.

Überblicksmäßig werden noch die Survival- und Hazardkurven für alle im Kapitel 9.2 ausgewählten Aktivitätskategorien (Arbeit, Einkauf langfristig, persönliche Erledigungen wie Post, Bank, etc., Besuch bei Verwandten, Verein / Bürgerinitiative / Kirche, Ausübung von Sport, Spaziergang / Spazierfahrt und Lokal / Kino / Theater) in den Abbildungen 9.33 bis 9.40 dargestellt. Einzig die Kategorie „Etwas Neues sehen/ Sightseeing“ wird nicht angegeben, da sie nur zwei Beobachtungen beinhaltet und daher absolut nicht aussagekräftig ist. Bei allen Kategorien ist als abhängige Variable bzw. Größe auf der Zeitachse die Dauer der Intervalle zwischen zwei gleichartigen Aktivitäten in Tagen (*inter\_t*) gewählt. Die letzte Grafik (Abbildung 9.41) stellt die Dauer des Intervalls zwischen zwei beliebigen Aktivitäten aus dem gesamten Datensatz dar. Die Einheit der Zeit ist hier in Stunden gewählt (*intervall\_st*). Die Zeitdauer von der Beendigung der letzten bis zum Beginn der nächsten Aktivität stellt jedoch genau die Reisezeit dar und hat aus diesem Grund nicht extrem viel Sinn. Die geglätteten Hazardkurven sind im Anhang X (Abbildung X.1 bis X.13) ersichtlich.

Wie im Kapitel 9.2.2.3 (Genauere Betrachtung ausgewählter Aktivitätskategorien) werden auch hier die Unterschiede zwischen den Geschlechtern und den verschiedenen Altersgruppen in den vier ausgewählten Kategorien etwas näher beleuchtet.

### 9.3.2.1 **Verschiedene Kategorien**

In der Kategorie Einkauf kurzfristig (Abbildung 9.29) verläuft die Hazardkurve relativ ruhig mit jeweils ausgeprägten Höhepunkten im Tagesabstand. Größere Spitzen sind nur am Ende der Intervalldauer-skala bei 11 bzw. 13 Tagen zu sehen, wo jedoch die Anzahl der Beobachtungen nur mehr so gering ist, daß sie keinerlei Einfluß mehr haben. Der Verlauf der Hazardkurve gibt eine große Wahrscheinlichkeit an, daß die Intervalle jeweils nach vollen Tagen beendet werden. Dies bedeutet, daß sehr viele Personen bei ihren Einkäufen einen deutlichen Rhythmus im Tagesabstand erkennen lassen. Sie kaufen selten einmal nachmittags und das nächste Mal vormittags ein, sondern meist um die selbe Tageszeit, so daß größtenteils ganztägige Abstände zustande kommen.

Abbildung 9.29 : Kategorie Einkauf kurzfristig

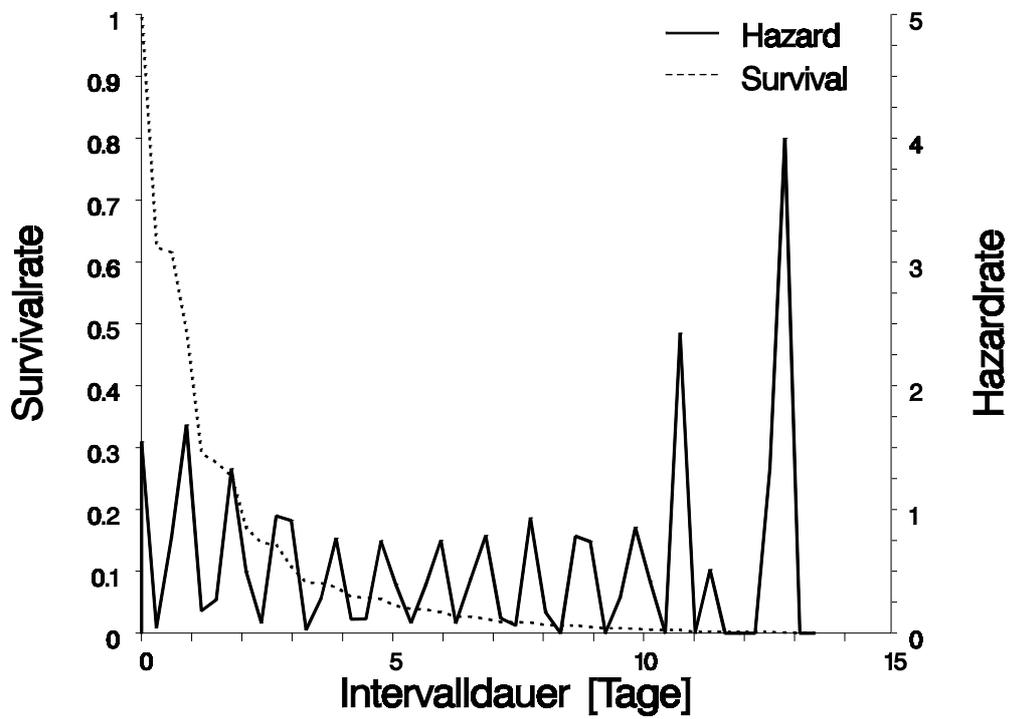


Abbildung 9.30 : Kategorie Freizeit allgemein

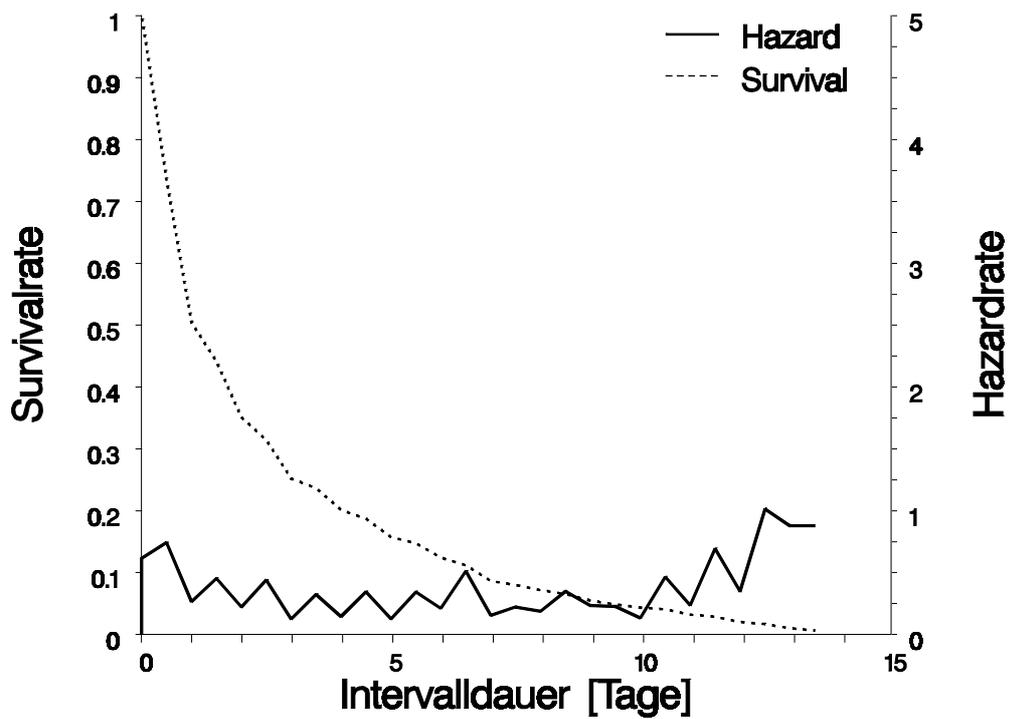


Abbildung 9.31 : Kategorie Dienstleistung allgemein

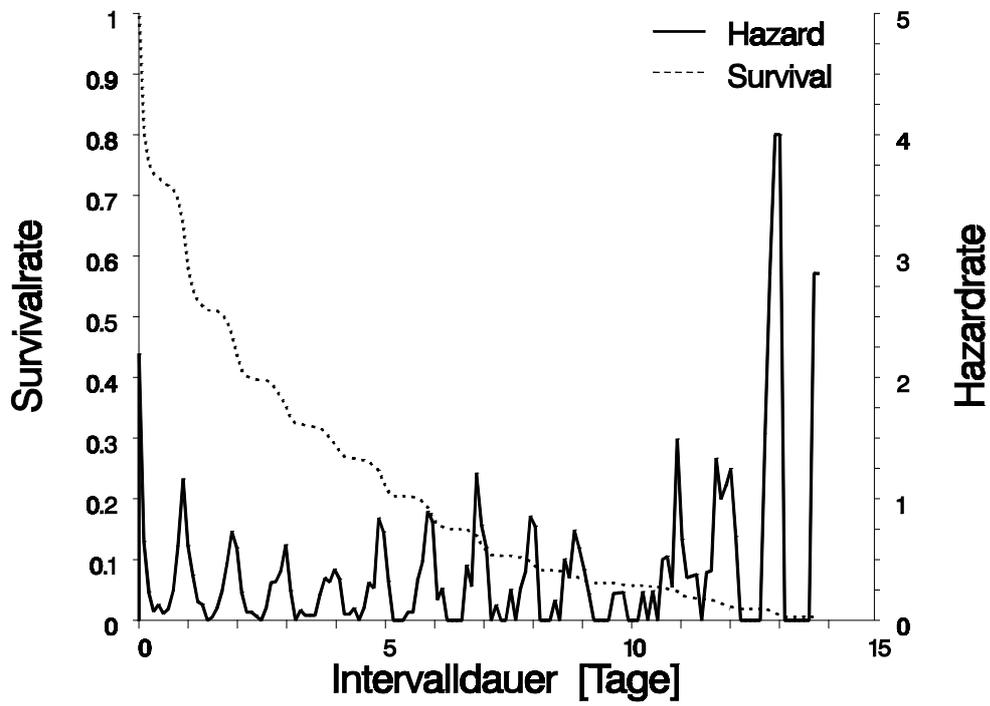
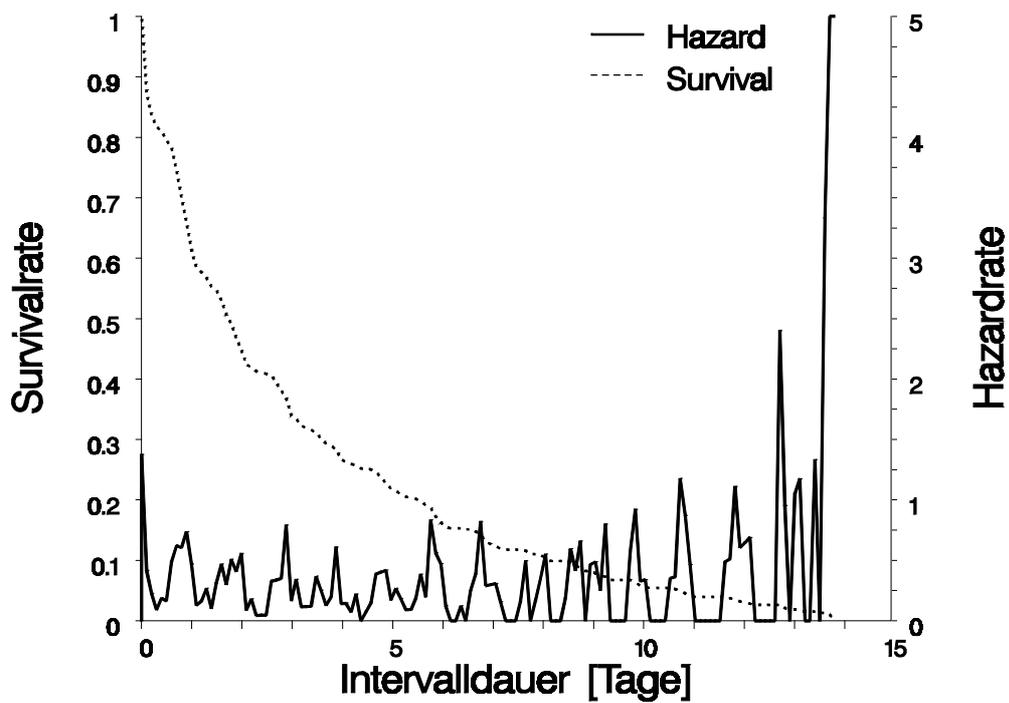


Abbildung 9.32 : Kategorie Treffen mit Freunden



Die Survivalkurve in der Kategorie Einkauf kurzfristig (Abbildung 9.29) hat einen wesentlich steileren Abfall als alle anderen näher betrachteten Kategorien, was auf eine durchschnittlich viel kürzere Zeitspanne zwischen zwei aufeinander folgenden Einkaufsaktivitäten schließen läßt. 51 % aller Intervalle zwischen zwei Einkäufen dauern höchstens einen Tag an, knapp  $\frac{3}{4}$  aller Einkäufe werden spätestens im Abstand von 2 Tagen erledigt. Trotzdem dauern ca. 10 Prozent aller Intervalle zwischen zwei Einkaufsaktivitäten zwischen 3 und 10 Tagen an.

Die Kategorie Freizeit allgemein (Abbildung 9.30) zeigt einen extrem ruhigen Verlauf ihrer Hazardkurve, die erst im letzten Abschnitt über die 0.5 Marke hinausgeht. Zuvor bewegt sich der Graf stets um den Wert von 0.3, was auf eine sehr gleichmäßige Verteilung der Intervalldauern schließen läßt. In der Tat nähert sich die Survivalkurve der exponentiellen Verteilungskurve stark an. Das bedeutet, daß der zeitliche Abstand zwischen zwei Freizeitaktivitäten meist keinem bestimmten Muster folgt und Freizeitaktivitäten auch nicht vorwiegend zur selben bestimmten Tageszeit durchgeführt werden, so daß nicht nur ganztägige Abstände entstehen. Ein kleiner Höhepunkt ist dennoch bei 6 bis 7 Tagen zu sehen, was auf vermehrte Freizeitaktivitäten an Wochenenden bzw. im Wochenabstand schließen läßt.

Die Hazardkurve in der Kategorie Dienstleistung allgemein (Abbildung 9.31) hat wiederum sehr ausgeprägte Spitzen an jeweils vollen Tagen, also eine deutliche Steigerung der Wahrscheinlichkeit, daß ein Intervall nach einer Dauer von jeweils vollen Tagen beendet wird bzw. eine erneute Aktivität in dieser Kategorie durchgeführt wird. Extreme Höhepunkte sind auch wieder erst am Ende der Zeitachse zu sehen, wo nur noch sehr grosse Intervalldauern und wenige Beobachtungen vorhanden sind und daher die Bedeutung nicht mehr allzu groß ist. Die Survivalkurve hat einen stufenförmigen Verlauf, der mit der Zeit flacher wird. Rund ein Viertel alle Intervalle dauern weniger als 5 Stunden, das heißt, daß oft mehrere Dienstleistungsaktivitäten hintereinander an einem Tag innerhalb kürzerer Zeit gemacht werden. Weitere 20 Prozent aller Intervalle dauern genau einen Tag (24 Stunden) an, danach geht es stufenweise in 5 bis 10 % - Schritten pro zusätzlichem Tag weiter.

Die spezielle Freizeitaktivitätskategorie Treffen mit Freunden (Abbildung 9.32) zeigt einen sehr ähnlichen Verlauf wie die allgemeine Freizeitkategorie, sowohl in bezug auf die Hazardkurve, als auch auf die Survivalkurve. Man kann keinen eindeutig definierbaren Verlauf der Hazardkurve erkennen. Sie verläuft relativ ruhig ohne ausgeprägtes Muster und steigt auch erst im letzten Teil stark an, was aber vorrangig auf die geringe Anzahl an Beobachtungen zurückzuführen ist. Es sind mit Ausnahme der Intervalldauern von 6 bzw. 7 Tagen keine starken Spitzen im Ganztagesabstand zu sehen. Diese Höhepunkte könnten durchaus auch wiederum mit vermehrten Freundesbesuchen jeweils an Wochenenden oder bestimmten Wochentagen im Wochenabstand (beispielsweise Kegelabend jeden Donnerstag) zusammenhängen, wie schon in der allgemeinen Freizeitkategorie.

Abbildung 9.33 : Arbeit

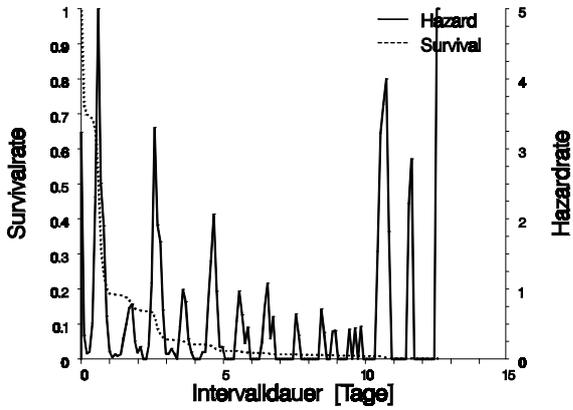


Abbildung 9.34 : Einkauf langfristig

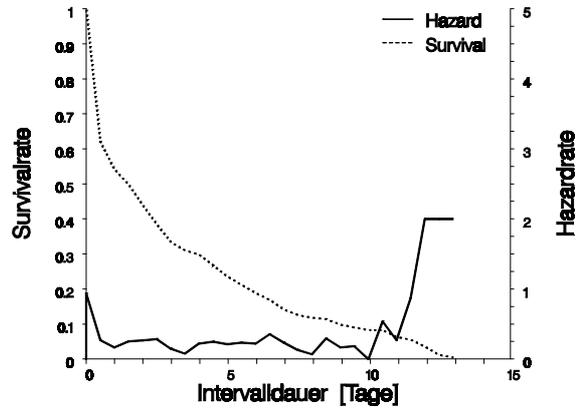


Abbildung 9.35 : Erledigung von Pflichten

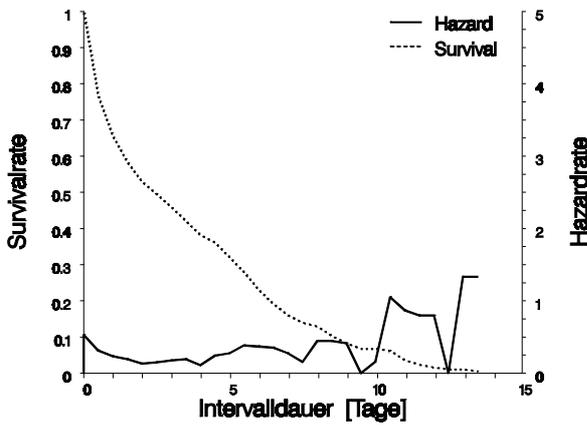


Abbildung 9.36 : Besuch bei Verwandten

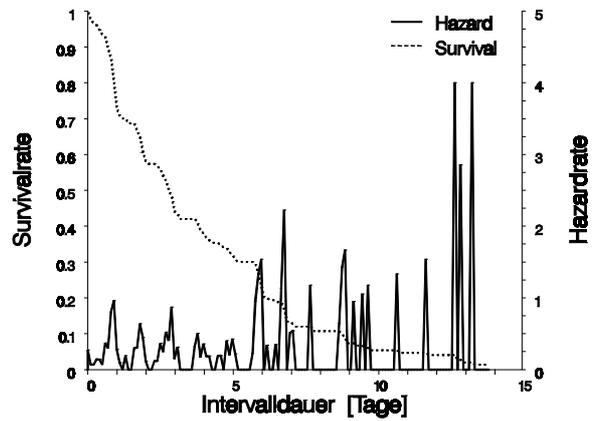


Abbildung 9.37 : Verein, Initiative, Kirche

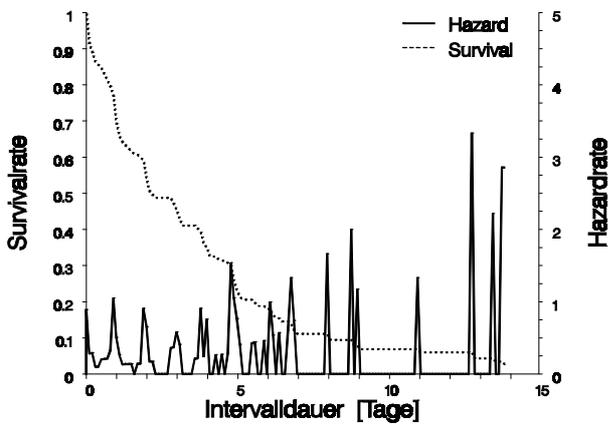


Abbildung 9.38 : Sport (aktiv)

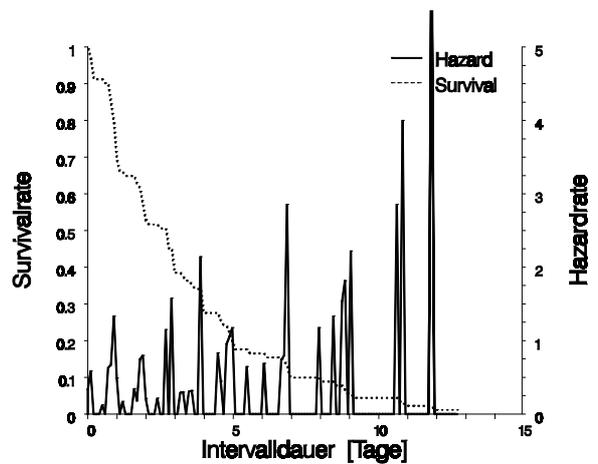


Abbildung 9.39 :Spaziergang / -fahrt

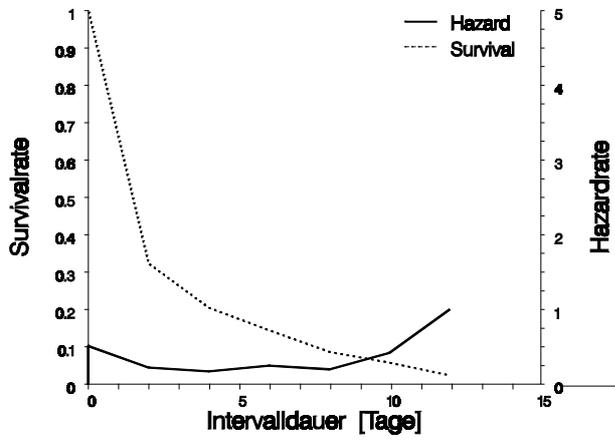


Abbildung 9.40 : Lokal, Kino, Theater

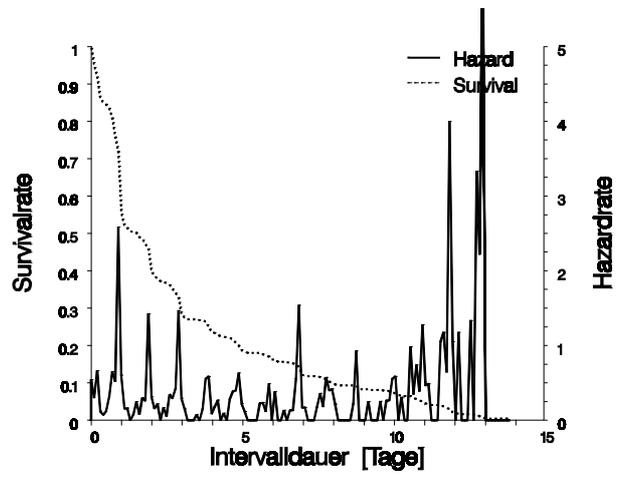
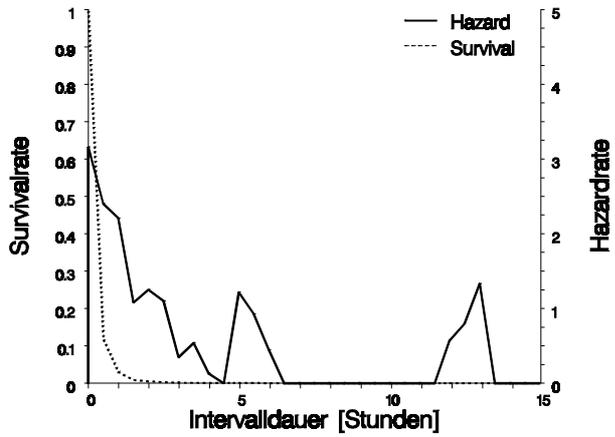


Abbildung 9.41 : Aktivitäten allgemein



Die weiteren Kategorien haben teilweise sehr unterschiedliche Hazard- und Survivalkurven. Die Survivalkurve der Kategorie Arbeit (Abbildung 9.33) beispielsweise beginnt mit einem sehr steilen Abfall, was auf eine große Anzahl an sehr geringen Intervalldauern zwischen zwei Arbeitsaktivitäten hinweist. Das ist jedoch hauptsächlich mit dem Einlegen einer Mittagspause zu erklären. Die Hazardkurve ist relativ zackig mit ausgeprägten Extrema, wobei sehr große Hazardraten bei einer Intervalldauer von knapp einem Tag bzw. knapp drei Tagen zu sehen sind. Die Intervalldauern sind hier natürlich nicht genau im Tagesabstand, da die Aktivitätsdauer in der Kategorie Arbeit prinzipiell um einiges höher liegt als beispielsweise bei Erledigungen oder Einkäufen. Um einen geregelten täglichen Arbeitsbeginn in der Früh zu gewährleisten, kann das Intervall vom Ende der letzten bis zum Anfang der darauffolgenden Arbeitsaktivität natürlich nicht 24 Stunden sein, sondern dauert nur vom Arbeitsende des einen Tages bis zum Arbeitsbeginn am nächsten Morgen.

Die Survivalkurven der Intervalle zwischen zwei Besuchen bei Verwandten und zwischen sozialen Aktivitäten der Kategorie Verein, Initiative, Kirche (Abbildungen 9.36 und 9.37) zeigen einen sehr ähnlichen Verlauf wie die der Kategorie Freizeit allgemein. Relativ ähnlich verlaufen auch die Survivalkurven der Kategorien Lokal, Kino, Theater und Sport aktiv (Abbildungen 9.40 und 9.38). Der einzig gravierende Unterschied besteht im Verlauf der Hazardkurven. Während dieser sich bei der allgemeinen Freizeitkategorie als sehr ruhig darstellt, gibt es in den Unterkategorien Verein, Initiative, Kirche und Verwandtenbesuche zu Beginn einen relativ ruhigen Verlauf und gegen Ende große Spitzenwerte, die jedoch aufgrund der geringen Anzahl an Beobachtungen nicht mehr stark ins Gewicht fallen. In der Kategorie Sport aktiv stellt sich der gesamte Verlauf wesentlich zackiger dar, regelmäßige Muster, die auf einen eventuellen Tages- oder Wochenrhythmus schließen lassen könnten, sind jedoch auch nicht vorhanden. In der Kategorie Lokal/Kino/ Theater ist überraschenderweise nach einem, zwei und drei Tagen Intervalldauer zwischen zwei Besuchen von Lokalitäten dieser Art jeweils ein deutlicher Höhepunkt der Hazardkurve zu sehen. Vor allem die Regelmäßigkeit, daß Aktivitäten dieser Kategorie im Tagesabstand (ca. 45 Prozent aller Intervalle dauern nicht länger als einen Tag an) gemacht werden, verblüfft etwas.

Die Hazardkurven der Kategorien langfristiger Einkauf und Erledigung von Pflichten (Abbildungen 9.34 und 9.35) zeigen einen sehr ruhigen Verlauf und die dazugehörige Survivalkurve einen stetigen, aber nicht extrem steilen Abfall. Das heißt, daß die Aktivitäten gleichverteilt sind und nach keinem regelmäßigen Muster gemacht werden, sondern genau dann, wenn die Erledigung ansteht oder Zeit dafür vorhanden ist. Die Hazardkurve der Kategorie Spaziergang / Spazierfahrt (Abbildung 9.39) zeigt einen extrem ruhigen Verlauf, der nach einem leichten Hoch zu Beginn (Hazardwert von 0.5) langsam fällt und ab einer Intervalldauer von zwei Tagen bis zum Ende hin leicht ansteigt (Wert bei 1). Die Survivalkurve fällt bis zu einer Intervalldauer von zwei Tagen relativ steil ab (65 % aller Intervalle sind nicht länger als 2 Tage) und zeigt danach einen stetigen, jedoch leicht abfallenden Verlauf.

### 9.3.2.2 Genauere Betrachtung der vier ausgewählten Kategorien

Wie schon bei den Aktivitätsdauern werden auch bei den Intervalldauern die vier Aktivitätskategorien „Einkauf kurzfristig“, „Freizeit allgemein“, „Dienstleistung allgemein“ und „Treffen mit Freunden“ bei den semiparametrischen und vollparametrischen Modellschätzungen (Kapitel 9.3.3 und 9.3.4) näher untersucht. Daher werden nonparametrische Auswertungen der Unterschiede zwischen einzelnen Altersgruppen und den beiden Geschlechtern durchgeführt und anhand von Hazard- und Survivalkurven grafisch dargestellt. Die entsprechenden Grafiken sehen Sie gesamt im Anhang X (LIFETEST I Hazard- und Survivalkurven) in den Abbildungen X.14 bis X.37. In den Abbildungen der Altersgruppen sind jeweils sieben Kurven zu sehen, wobei die erste Kurve die Gruppe der unter 25-jährige Personen darstellt, die danach folgenden Gruppen um jeweils ein 10-Jahresintervall älter sind und die letzte Gruppe Personen über 75 Jahre sind. In den Abbildungen für die geschlechtlichen Unterschiede sind die weiblichen Personen wie üblich in grün, die männlichen Umfrageteilnehmer in blau dargestellt. Folgend werden einige interessante Grafiken zur Diskussion herausgenommen.

Die geglätteten Hazardkurven der einzelnen Altersgruppen bzw. den Geschlechtern scheinen meist recht unterschiedlich. Nicht nur die Werte der Hazardfunktion, sondern auch die Bandbreite der Kurve auf der Skala der Intervalldauer lassen zum Teil gravierende Unterschiede zwischen den einzelnen Altersgruppen und den beiden Geschlechtern vermuten. Sieht man sich jedoch die ungeglätteten Kurven näher an, sind die Unterschiede meist gar nicht so groß.

In der Kategorie Einkauf kurzfristig sind vor allem die wesentlich größeren Schwankungen der Hazardkurve (Abbildungen X.14 und X.15) bzw. der stufenförmige Verlauf der Survivalkurve (Abbildung X.30) und der frühere Abbruch der Kurve bei älteren Menschen auffallend. Dies bedeutet, daß ältere Menschen ein wesentlich ausgeprägteres von der Tageszeit abhängiges Muster bei Einkaufsaktivitäten haben als jüngere Umfrageteilnehmer, da die Wahrscheinlichkeit, daß der nächstfolgende Einkauf genau einen oder genau zwei, drei, vier etc. Tage später folgt, sehr viel höher liegt. Sehr junge Menschen (die Gruppe der unter 25-jährigen und der 30-jährigen Umfrageteilnehmer) haben hingegen eher niedrige und ausgeglichene Hazardwerte, was auf eine sehr hohe Bandbreite der zeitlichen Verteilung der Einkäufe über den gesamten Tag gesehen schließen läßt. Sie haben meist keinen fixen Zeitpunkt am Tag, zu dem sie ihren täglichen Einkaufsbedarf decken. Zudem kommt es bei jungen Altersgruppen auch zu oft sehr langen Intervalldauern zwischen zwei Einkäufen.

Unterschiede zwischen den Geschlechtern sind nur sehr geringe zu erkennen (Abbildungen X.16 und X.17 bzw. X.31). Weibliche Umfrageteilnehmer haben etwas höhere Spitzen der Hazardwerte bei kürzeren Intervalldauern, männliche hingegen bei Intervalldauern ab etwa 7 Tagen. Dies bedeutet, daß

Frauen mit einer größeren Wahrscheinlichkeit ihre Einkaufsaktivitäten in kürzeren Abständen durchführen als ihre männlichen Umfragekollegen.

Sieht man sich die ungeglätteten Hazard- und Survivalkurven der Kategorie Freizeit allgemein an (Abbildungen X.18 und X.32), so fällt auf, daß die Unterschiede zwischen den Altersgruppen gar nicht so extrem sind. Ältere Personen haben jedoch im vorderen Bereich der Zeitachse höhere Survivalwerte, das heißt im Schnitt etwas längere Zeitspannen zwischen ihren Freizeitaktivitäten, 30-, 40- und 50-jährige haben im Gegensatz dazu relativ viele Aktivitäten in kürzeren Zeitabständen unternommen. Bei der Hazardkurve stechen immer wieder höhere Spitzenwerte der Gruppe der über 75-jährigen und der 50-jährigen hervor, was auf eine relativ regelmäßige Inanspruchnahme von Freizeitmöglichkeiten im Abstand von jeweils ganzen Tagen hindeutet (beispielsweise Besuch von Freunden alle zwei Tage, Kartenspiel- oder Bridgerunde jede Woche einmal, etc.). Die jüngeren Umfrageteilnehmer weisen hohe Hazardwerte nur bei sehr kurzen Intervallauern von ein bis zwei Tagen auf.

Die Unterschiede der beiden Geschlechter sind vorrangig in einem Intervallbereich von 5 bis 9 Tagen bzw. in der Verteilung zu sehen (Abbildungen X.21 und X.33). Aufgrund der wesentlich höheren Anzahl an Freizeitaktivitäten bei männlichen Umfrageteilnehmern fällt auch die Survivalkurve wesentlich rascher ab, sie machen also mehr Freizeitaktivitäten in kürzeren Abständen. Frauen haben nach 5 bzw. 6 Tagen eine wesentlich höhere Wahrscheinlichkeit, eine erneute Freizeitaktivität durchzuführen, Männer eine sehr hohe, ihren Aktivitäten im Wochenrhythmus zu frönen.

Die Kurven der Kategorie Treffen mit Freunden (Abbildung X.26 bis X.29) zeigen einen ähnlichen Verlauf wie die der allgemeinen Freizeitkategorie, jedoch differieren bei den einzelnen Altersgruppen die Werte der Survival- und der Hazardkurve etwas mehr. So haben beispielsweise die 50-jährigen eine sehr hohe Wahrscheinlichkeit, ihre Freundesbesuche im 5, 6, 7 oder 8-Tagesrhythmus durchzuführen, unter 25-jährige im 1- bzw. 2-Tagesrhythmus, ältere Personen nach 1, 2, 3 oder 5 Tagen und 40-jährige nach genau einer Woche. Bei den übrigen Altersgruppen sind im aussagekräftigen Intervallbereich (jener Bereich, wo noch genügend Beobachtungen vorhanden sind) keine extremen Hazardwerte erkennbar, sondern eher ein ruhiger Verlauf, der auf eine relativ ausgeglichene Verteilung der Treffen mit Freunden über den Tag bzw. die Umfragezeit schließen läßt.

Die geschlechtlichen Unterschiede sind zwar vorhanden, jedoch nicht einfach zu beschreiben. Grundsätzlich haben Frauen ebenfalls wesentlich längere Intervalle zwischen den einzelnen Treffen mit Freunden bzw. treffen sich Männer um ein Vielfaches häufiger mit Freunden und dadurch auch in kürzeren Abständen. Hohe Wahrscheinlichkeiten, daß Männer ein erneutes Treffen mit Freunden haben, sind nach 1 bzw. 3, 6 oder 9 Tagen, bei Frauen nach 2 Tagen, 7 Tagen oder noch längeren Intervallauern.

In der Kategorie Dienstleistung allgemein (Abbildung X.22 und X.23) sind keine extremen Unterschiede zwischen den einzelnen Altersgruppen erkennbar. Es fällt nur auf, daß ältere Personen ab 75, aber auch Personen unter 25 ihre Erledigungen gerne im Abstand von 2 oder noch lieber 5 Tagen machen. Nach 9 Tagen sind zusätzlich zu diesen beiden Gruppen die Werte der 50-jährigen sehr hoch.

Die Unterschiede zwischen Männern und Frauen (Abbildung X.24 und X.25) sind ebenfalls meist nur geringfügig, nach 2 oder 7 Tagen haben Männer eine sehr hohe Wahrscheinlichkeit, eine erneute Dienstleistungsaktivitäten durchzuführen, Frauen hingegen nach 1 oder 3 Tagen.

### 9.3.2.3 Vergleich mit den Ergebnissen der semiparametrischen Auswertung

Wie bei den Auswertungen für Aktivitätsdauern sind auch hier als Überprüfung der möglichen Einflußvariablen in den vier Kategorien der Log-Rank-Tests und der Wilcoxon-Test mit jeweils zwei Stufen gemacht worden. Mit Hilfe der sogenannten *Univariate Chi-Squares*, bei denen die einzelnen Variablen separat ohne gegenseitige Beeinflussung bzw. Korrelation getestet werden, und die *Forward Stepwise Sequences of Chi-Squares*, also stufenförmige Auswertungen von Signifikanzen, die in Abhängigkeit derjenigen Variable erstellt werden, die bei singulärer Betrachtung den höchsten Chi-Quadratwert ergibt, können nonparametrische Analysen der möglichen Einflußgrößen durchgeführt werden. Die Ergebnisse der Überprüfungen sind im Anhang XI, Seite 256 übersichtlich dargestellt.

Nachdem ein Großteil der Intervalle in allen vier Kategorien nur sehr kurz andauert (rund  $\frac{3}{4}$  der Intervalle werden innerhalb des ersten Viertels der Zeitachse beendet), werden in allen Kategorien, wenn alle möglichen Einflußgrößen in der Modellschätzung enthalten sind, nur die Ergebnisse des Wilcoxon-Tests, sowohl bei den univariaten, als auch bei den stufenförmigen Tests präsentiert, weil dabei wesentlich mehr Gewichtung auf frühe Ereigniszeiten (bzw. kurze Intervalldauern hier) gelegt wird. Bei den Ergebnissen der Modelle mit genau jenen Variablen, die auch im ersten Gesamtmodell der semiparametrischen Auswertung verwendet werden, sind zur besseren Übersicht auch die Ergebnisse der Log-Rank-Tests angegeben. Sie sind jedoch meistens etwas weiter von den Ergebnissen der semiparametrischen Tests (Kapitel 9.3.3) entfernt.

Die stufenförmigen Wilcoxon- und Log-Rank-Tests, die auch als nonparametrische Koeffiziententests des *accelerated failure time model (AFT-Modell)*, siehe Kapitel 6.4.1) interpretiert werden können, zeigen in den seltensten Fällen gravierende Unterschiede zu den in den semiparametrischen Modellen gewonnenen Erkenntnissen. Nur wenige Einflußgrößen weisen unterschiedliche Signifikanz auf, was jedoch meist auf die Vorgangsweise bei diesen Tests zurückzuführen ist. Da die stufenförmigen Tests

immer nur jene Variablen, die vorher überprüft worden sind, in die Berechnung mit einbeziehen, nicht aber jene, die erst danach untersucht werden, kann es zu unrichtigen Schlußfolgerungen kommen.

In der Kategorie Einkauf kurzfristig (Tabelle XI.4) steht beispielsweise die Variable Führerscheinbesitz ganz am Schluß mit absolut keinem signifikanten Einfluß, bei den semiparametrischen Auswertungen ist sie jedoch signifikant. Umgekehrt ist die Art der beliebigen Aktivität vor dem Intervall und die Anzahl aller Reisen der Person hier stark signifikant, bei den semiparametrischen Tests gar nicht.

Bei den Intervallen zwischen Freizeitaktivitäten (Tabelle XI.10) scheinen exakt dieselben Variablen wie in den semiparametrischen Modellen auf, es gibt nur bei der Größe der Beeinflussung Unterschiede.

Ähnlich verhält es sich bei den Ergebnissen der Kategorie Dienstleistung allgemein (Tabelle XI.16). Es kommen nur die beiden Variablen Ausbildung während der Umfrage und Anzahl der Führerscheine im Haushalt mit marginaler Signifikanz hinzu, die Dauer zwischen der das Intervall abschließenden Aktivität und der beliebigen Aktivität vorher beeinflusst die Intervalldauer wie beim semiparametrischen Modell für Personen mit regelmäßigen Dienstleistungsaktivitäten stark signifikant.

Bei Treffen mit Freunden (Tabelle XI.22) fällt nur die Anzahl aller gemachten Wege einer Person im Gegensatz zu den semiparametrischen Auswertungen aus den signifikanten Variablen heraus. Sonst ist ebenfalls nur die Größe der Beeinflussung unterschiedlich.

Aufgrund der eventuell irreführenden Abhängigkeit von der Variable mit dem jeweilig höchsten Chi-Quadrat-wert und den geringfügigen Unterschieden zu den semiparametrischen Modellschätzungen ist den im Kapitel 9.3.3 präsentierten Ergebnissen wesentlich mehr Bedeutung beizumessen.

### 9.3.3 Semiparametrische Modelle mit *PROC PHREG*

#### 9.3.3.1 Einleitung

In diesem Kapitel werden nur noch die Auswertungen der vier ausgewählten Aktivitätskategorien Einkauf kurzfristig (mit 3231 Aktivitäten bzw. 3087 Intervallen insgesamt), Freizeit allgemein (3025 bzw. 2576), Dienstleistung allgemein (922 bzw. 738) und Treffen mit Freunden (687 bzw. 564) dargestellt und diskutiert. Um eine einheitliche Auswertung zu garantieren, werden in allen vier Kategorien die jeweils ersten Intervalle bzw. letzten Intervalle von vornherein nicht mit berücksichtigt, sie wären links- bzw. rechtszensuriert. Zudem werden all jene Intervalle, die größer als 14 Tage sind, ebenfalls aus den Datensätzen entfernt, um das Bild nicht allzu sehr zu verzerren. Durch diese Limits und die schon im Kapitel 9.2.3 angewendeten Grenzen bei den Aktivitätsdauern in jeder Kategorie reduzieren sich die in die Modellschätzung einbezogenen Fälle auf 3069 Intervalle zwischen zwei Einkäufen bzw. 2487 (Freizeit), 692 (Dienstleistung) und 536 (Treffen mit Freunden).

Die grundlegenden Ziele der semiparametrischen Auswertung sind im Kapitel 9.2.3 schon eingehend erklärt, für die Auswertung der Intervalle werden grundsätzlich dieselben Maßstäbe angewendet. Es gilt, aus einer Reihe möglicher Einflußvariablen jene herauszufinden, bei denen tatsächlich ein Einfluß auf die Dauer des Intervalls zwischen zwei Aktivitäten festzustellen ist und in einem zweiten Schritt die Größe und Richtung dieser Beeinflussung mit geeigneten Zahlenwerten anzugeben. Alle Modellschätzungen sind nach einer Reihe von Auswahlkriterien für die möglichen Einflußvariablen (Einzelanalyse aller Variablen, Korrelationsmatrizen, Proportionalitätstests, etc.) mit semiparametrischen Regressionsmethoden mit *PROC PHREG* in mehreren Arbeitsschritten durchgeführt worden. Dazu gehören die Erstellung eines ersten Gesamtmodells, ein Modell der zweiten Stufe und zusätzliche Analysen der Unterschiede zwischen Personen mit sporadischen bzw. regelmäßigen Aktivitätsteilnahmen. Es werden wiederum beide Berechnungsarten (Exact- und Efron-methode) angewendet, um gleichzeitig stattfindende Ereignisse bzw. zum selben Zeitpunkt endende Intervalldauern berücksichtigen zu können. Auch hier liegen die Ergebnisse knapp beieinander.

Die Ergebnisse der jeweils zweiten Gesamtmodelle werden nun auf den folgenden Seiten detailliert präsentiert und diskutiert. Die gesamten Ergebnisse inklusive des Vergleichs zwischen Personen mit regelmäßigen und sporadischen Teilnahmen an Aktivitäten und die Aufgliederung der Variablen in eine soziodemografische und eine aktivitätenspezifische Sparte sind im Zusatzanhang XXIV ersichtlich.

### 9.3.3.2 Einzelne Kategorien

Zusätzlich zu den bei der Auswertung der möglichen Einflußgrößen auf die Aktivitätsdauern verwendeten Variablen sind bei den Intervallen noch ein paar aktivitätsvernetzende Variablen beigefügt worden. So sind in den Intervalldateien die Dauer der Zeitspanne zwischen zwei beliebigen Aktivitäten in Tagen und Stunden (*intervall\_t*, *intervall\_st*) und zwischen zwei gleichartigen Aktivitäten (*inter\_t*, *inter\_st*) vorhanden. Weiters werden die Dauer der vorigen Aktivität allgemein und in dieser Kategorie, die Art des Verkehrsmittels zur Aktivität des Intervalls und die Dauer des vorigen Intervalls zwischen zwei Aktivitäten dieser Kategorie mitberücksichtigt. Aus der Liste der möglichen Einflußvariablen wird die Anzahl der Aktivitäten bzw. Intervalle in dieser Kategorie (*nr\_a\_sum\_c* bzw. *nr\_i\_sum\_c*) genommen, da diese Variablen selbstverständlich einen Einfluß zeigen, der jedoch direkt abhängig ist. Werden in einer festgesetzten Zeitspanne von 5 Wochen mehr Aktivitäten durchgeführt, so muß klarerweise auch die Intervalldauer zwischen zwei Aktivitäten sinken.

Aus den insgesamt 57 möglichen Variablen werden mit Hilfe von Einzelanalysen nur jene in einen nächsten Schritt mitgenommen, die statistisch signifikant sind. Korrelationen untereinander sind ebenfalls unvermeidbar und so sind schon vor dem ersten Gesamtmodell einige Variablen eliminiert worden. Hohe Korrelationen treten wie bei den Auswertungen der Aktivitätsdauern zwischen den Variablen Führerscheinbesitz, Autos im Haushalt und Autoverfügbarkeit auf, ebenso zwischen den Variablen Gesellschaftsgruppe, Alter und Kinder im Haushalt. Die Korrelationen zwischen dem Verkehrsmittel vor und nach den jeweiligen Aktivitäten des Intervalls sind bei weitem nicht so groß wie bei den Auswertungen der Aktivitäten. Auch die Fragen, ob am Aktivitätort Geld ausgegeben worden ist bzw. wie hoch diese monetären Ausgaben gewesen sind, sind miteinander verknüpft, spielen aber lang keine so große Rolle. Sind die Korrelationen zwischen zwei oder mehreren Variablen zu groß, so bleibt die Variable mit dem höchsten Chi-Quadratwert der singulären Variablenschätzung für ein erstes Gesamtmodell in der Auswertung. Weiters sind Proportionalitätsüberprüfungen bei allen nicht numerischen Variablen (Dummyvariablen) durchgeführt worden, numerische Variablen können damit zwar getestet werden, die Ergebnisse sind aber nicht aussagekräftig. Alle getesteten Variablen können positiv begutachtet werden. Die ersten Gesamtmodelle sind mit 21 (Einkauf kurzfristig), 23 (Freizeit allgemein), 17 (Dienstleistung allgemein) und 11 Variablen (Treffen mit Freunden) erstellt worden und die Ergebnisse sind zum besseren Vergleich im Anhang XII zu sehen.

Eine Übersicht aller möglichen Einflußvariablen für die vier Kategorien sehen Sie im Anhang XIV, in der die stark signifikanten Variablen nach einer Modellschätzung der zweiten Stufe mit einem Stern (\*) markiert sind. Stark signifikant bedeutet in diesem Fall, daß sie einen Signifikanzwert unter 0.05 aufweisen bzw. im 95% Konfidenzintervall liegen. Die marginal im 90% Konfidenzintervall signifikanten Variablen sind mit zwei Sternen (\*\*) gekennzeichnet.

Im Folgenden werden einige auffällige Erkenntnisse aus den Auswertungen näher betrachtet und diskutiert bzw. in den Tabellen 9.12 bis 9.15 dargestellt.

In allen vier Kategorien wird die Intervalldauer von der Länge der letzten durchgeführten beliebigen Aktivität insofern beeinflusst, daß eine längere Dauer der Aktivität auch eine längere Dauer des Intervalls zur Folge hat. Die Dauer der letzten gleichartigen Aktivität vor dem Intervall hat nur in den beiden Kategorien Einkauf (Tabelle 9.12) und Freizeit (Tabelle 9.13) einen signifikanten Einfluß. Auch hier dauert das Intervall länger, je länger die Dauer der Aktivität vor dem Intervall ist. Macht also beispielsweise jemand einen längeren Einkauf, so ist es eher wahrscheinlich, daß der nächste Einkauf nicht gleich am nächsten Tag, sondern erst zwei Tage später folgt.

Das Geschlecht hat nur bei den Dienstleistungsaktivitäten (Tabelle 9.14) keinen signifikanten Einfluß, in allen anderen Kategorien ist ein deutlicher Einfluß zu sehen, wenn auch nicht in dieselbe Richtung. Männer haben demnach eine wesentlich höhere Wahrscheinlichkeit, größere Intervalle zwischen Einkaufsaktivitäten zu legen. Frauen hingegen führen ihre allgemeinen Freizeitaktivitäten mit hoher Wahrscheinlichkeit in wesentlich größeren Abständen durch und treffen sich auch nicht so oft oder in kurzen Zeitintervallen mit Freunden. Diese Ergebnisse sind natürlich direkt verknüpft mit der Anzahl der insgesamt vom jeweiligen Geschlecht durchgeführten Aktivitäten in der Kategorie. So weisen Frauen eine durchschnittliche Anzahl von 26 Einkäufen in den fünf beobachteten Wochen auf, Männer hingegen nur ca. 18. Im Gegensatz dazu liegen Männer bei Freizeitaktivitäten mit durchschnittlich 25 gegenüber 18 Aktivitäten im Beobachtungszeitraum bei Frauen voran.

Die Tageszeit der am Intervallende durchgeführten Aktivität hat bei Einkaufs- und Freizeitaktivitäten einen Einfluß, jedoch auch wieder mit unterschiedlichen Vorzeichen. Je früher am Tag der Einkauf durchgeführt wird, desto höher ist die Wahrscheinlichkeit, daß die Zeitspanne zum nächsten Einkauf klein bleibt. Bei Freizeitaktivitäten finden hingegen Aktivitäten, die zu einer späteren Tageszeit angesiedelt sind, mit einer signifikant höheren Wahrscheinlichkeit in kürzeren Zeitintervallen statt.

Ob ein motorisiertes Verkehrsmittel zu der das Intervall abschließenden Aktivität benützt wird, hat außer bei Treffen mit Freunden überall einen signifikanten Einfluß. Hat ein Umfrageteilnehmer die Möglichkeit, selbst zu fahren oder zumindest ein motorisiertes Fahrzeug mit zu benützen, so ist die Wahrscheinlichkeit, daß seit der letzten gleichartigen Aktivität nicht so viel Zeit verstrichen ist, wesentlich größer. Es sind die Intervalle bei mobileren Personen demnach wesentlich kürzer.

Die Anzahl aller gemachten Wege der jeweiligen Person hat in den beiden Freizeitkategorien einen signifikanten, jeweils positiven Einfluß auf die Intervalldauer. Das heißt, je mehr Wege bzw. Aktivi-

täten die Person insgesamt gemacht hat, desto höher ist die Wahrscheinlichkeit, daß der Abstand zwischen zwei Freizeitaktivitäten klein ist. Dies bedeutet jedoch noch nicht unbedingt, daß man den Schluß ziehen kann, daß Personen mit vielen Wegen auch signifikant mehr Freizeitaktivitäten machen. Bei den anderen beiden Kategorien hat die Anzahl aller Wege jedoch keinen signifikanten Einfluß. Die Dauer des Intervalls zwischen der Aktivität der jeweiligen Kategorie zu Beginn des betrachteten Intervalls und der beliebigen Aktivität vorher, was prinzipiell der Reisezeit zur Aktivität entspricht, nimmt hingegen in den Kategorien Einkauf und Dienstleistung signifikant Einfluß. Je länger dieses Intervall bzw. die Reisezeit ist, desto größer ist das abhängige Intervall. Die Reisezeit zur Aktivität am Ende des Intervalls hat hingegen bei Treffen mit Freunden einen marginalen und bei kurzfristigen Einkäufen einen signifikanten Einfluß in dieselbe Richtung, wenngleich weitaus geringer als bei der Reisezeit vor dem Intervall.

Etwas überraschend ist der Einfluß der Dauer des Intervalls zwischen zwei gleichartigen Aktivitäten vorher. Betrachtet man alle 144 Umfrageteilnehmer bzw. jene, die in den einzelnen Kategorien Aktivitäten unternommen haben, so hat die Dauer des Intervalls vorher mit Ausnahme der Kategorie Treffen mit Freunden einen signifikanten Einfluß auf die Dauer dieses Intervalls. Je länger das vorherige Intervall andauert, desto größer ist auch die folgende Intervalldauer. Bei Personen mit regelmäßigen Aktivitäten in der jeweils betrachteten Kategorie ist dieser Einfluß aber mit Ausnahme der allgemeinen Freizeitkategorie nicht signifikant. Das wiederum bedeutet, daß Personen mit mehr als einer berichteten Aktivität pro Woche bzw. insgesamt mehr als 5 Aktivitäten im Beobachtungszeitraum, die in den Tabellen als Personen mit regelmäßigen Aktivitätsteilnahmen bezeichnet werden, eventuell gar keine so große Regelmäßigkeit in ihrer Aktivitätsverteilung haben. Werden zwei Aktivitäten in einem längeren Zeitabstand durchgeführt, so bedeutet dies nicht unbedingt, daß die Dauer des nächsten Intervalls ebenfalls wieder größer ist. Es kann jedoch durchaus sein, daß beispielsweise ein wöchentlich ähnlicher Rhythmus bzw. ein regelmäßiges Verhaltensmuster vorliegt, was durch die Betrachtung von jeweils zwei aufeinanderfolgenden Intervallen nicht herausgefunden werden kann.

Auffällig sind auch die wesentlich höhere Anzahl an Einfluß nehmenden soziodemografischen Variablen wie Führerscheinbesitz bzw. Zugangsmöglichkeit zu einem Fahrzeug, Alter, Geschlecht, Beschäftigungsgrad, Haushaltsstand, etc. in den Kategorien Einkauf und Freizeit allgemein im Gegensatz zu den anderen beiden Kategorien, die zudem wesentlich weniger Variablen mit signifikanten Einflüssen aufweisen. Trotzdem dominieren in allen vier Kategorien die aktivitätsabhängigen Variablen.

Diverse andere Variablen haben in den einzelnen Kategorien noch einen signifikanten Einfluß, jedoch sind diese unterschiedlich. Bei Freizeit- und Einkaufsaktivitäten haben beispielsweise die Fragen, ob die Aktivität am Ende des Intervalls geplant worden ist oder ob sie an einem oder über ein Wochenende

de stattfindet, einen signifikanten Einfluß. Wenn sie geplant worden ist, ist das Intervall mit großer Wahrscheinlichkeit länger, wenn sie an einem Wochenende stattfindet, dementsprechend kürzer.

Der Ausbildungsgrad der befragten Personen, die Anzahl der Haushaltsmitglieder oder das Alter spielen ebenso wie die Fragen, ob Kinder im Haushalt sind, ob die Person Haushaltsvorstand ist, ob bzw. wie viele Ausgaben getätigt worden sind, ob die Person in der Umfragezeit gearbeitet hat oder nicht, gar keine oder nur eine untergeordnete Rolle.

Wie vorher schon erwähnt, ist ein weiteres Detail von grundlegendem Interesse. Die Frage, ob Personen mit sporadischen Aktivitäten in der betreffenden Kategorie, also weniger als 5 Aktivitäten insgesamt, das Gesamtergebnis beeinflussen, wird hier wie schon bei den Auswertungen der Aktivitäten mittels eines Vergleichs des jeweiligen Gesamtmodells der zweiten Stufe mit einer Modellschätzung der zweiten Stufe für Personen mit regelmäßiger Aktivitätsteilnahme beantwortet. In den folgenden Tabellen (9.12 bis 9.15) sind jeweils die Parameterschätzwerte  $\beta$  und die Hazardraten für jede Variable angegeben, mit deren Hilfe die Größe und die Richtung des Einflusses der entsprechenden Variable abgeschätzt werden kann. Ist beim Parameterschätzwert  $\beta$  ein negatives Vorzeichen vorhanden, so verkürzt diese Einflußgröße die Intervalldauer. Sind Variablen in einer Modellschätzung der zweiten Stufe signifikant im 95%-Konfidenzintervall, so werden sie fett gedruckt und mit einem Stern (\*) in der Spalte SN (Signifikanzniveau) gekennzeichnet, bei marginal signifikanten Variablen wird eine kursive Schrift und zwei Sterne (\*\*) verwendet. Nicht signifikante Variablen der zweiten Modellstufe sind durch normale Schrift ohne Stern erkennbar.

Tabelle 9.12 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Einkauf kurzfristig

Gesamtmodell 2. Stufe				Modell für regelmäßige Einkäufer			
Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN	Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN
<b>SEX</b>	-0.30394	0.738	*	<b>SEX</b>	-0.30033	0.741	*
<b>LICENSE</b>	0.17594	1.192	*	<i>LICENSE</i>	0.13215	1.141	**
<b>AGE</b>	0.00252	1.003	*	<i>AGE</i>	0.00210	1.002	**
<b>HH_NR_L</b>	-0.05414	0.947		<b>CAR</b>	0.15440	1.167	*
<b>PLANNED</b>	-0.18464	0.831	*	<b>HH_NR_L</b>	-0.09976	0.905	*
<b>WEEKEND</b>	0.19733	1.218	*	<b>PLANNED</b>	-0.14544	0.865	*
<b>A_DAYT</b>	-0.10015	0.905	*	<b>WEEKEND</b>	0.17669	1.193	*
<b>A_P_TT</b>	-0.01032	0.990	*	<b>A_DAYT</b>	-0.10419	0.901	*
<b>INTER_T_P</b>	-0.01789	0.982	*	<b>A_P_TT</b>	-0.01020	0.990	*
<b>A_P_DURST_A</b>	-0.01944	0.981	*	<b>A_P_DURST_A</b>	-0.01882	0.981	*
<b>A_P_DURST_C</b>	-0.06503	0.937	*	<i>A_P_DURST_C</i>	-0.05879	0.943	**
<b>MM_C_D</b>	0.14556	1.157	*	<b>MM_C_D</b>	0.14336	1.154	*
<b>INTERVALL_ST_P</b>	-0.49697	0.608	*	<b>INTERVALL_ST_P</b>	-0.47897	0.619	*

n = 2920; Likelihood Ratio Chi Quadrat = 407.04      n=3034; Likelihood Ratio Chi Quadrat = 397.17

Tabelle 9.13 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Freizeit allgemein

Gesamtmodell 2. Stufe				Modell für regelmäßige Aktivitäten			
Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN	Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN
<b>SEX</b>	0.13649	1.146	*	<b>SEX</b>	0.13716	1.147	*
<b>LICENSE</b>	-0.37431	0.688	*	<b>LICENSE</b>	-0.37437	0.688	*
<b>WORKED</b>	-0.24400	0.783	*	<b>WORKED</b>	-0.24349	0.784	*
<b>SINGLE</b>	0.31796	1.374	*	<b>SINGLE</b>	0.31668	1.373	*
<b>CAR</b>	0.44833	1.566	*	<b>CAR</b>	0.44949	1.568	*
<b>PLANNED</b>	-0.42155	0.656	*	<b>PLANNED</b>	-0.42270	0.655	*
<b>WEEKEND</b>	0.29377	1.341	*	<b>WEEKEND</b>	0.29336	1.341	*
<b>A_DAYT</b>	0.17023	1.186	*	<b>A_DAYT</b>	0.16927	1.184	*
<b>NR_T_SUM</b>	0.00511	1.005	*	<b>NR_T_SUM</b>	0.00512	1.005	*
<b>A_P_DURST_A</b>	-0.01347	0.987	*	<b>A_P_DURST_A</b>	-0.01345	0.987	*
<b>A_P_DURST_C</b>	-0.04002	0.961	*	<b>A_P_DURST_C</b>	-0.03976	0.961	*
<b>INTER_T_P</b>	-0.02429	0.976	*	<b>INTER_T_P</b>	-0.02402	0.976	*
<b>MM_C_D</b>	0.23553	1.266	*	<b>MM_C_D</b>	0.23400	1.264	*

n = 2704; Likelihood Ratio Chi Quadrat = 663.89      n=2697; Likelihood Ratio Chi Quadrat = 661.83

Tabelle 9.14 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Kategorie Dienstleistung allgemein

Gesamtmodell 2. Stufe				Modell für regelmäßige Aktivitäten			
Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN	Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN
<b>HH_NR_L</b>	0.06925	1.072		<b>HH_NR_L</b>	0.07236	1.075	**
<b>A_P_PUR</b>	-0.13727	0.872	*	<b>A_P_PUR</b>	-0.13035	0.878	*
<b>A_P_DURST_A</b>	-0.00499	0.995	*	<b>A_P_DURST_A</b>	-0.00455	0.995	
<b>INTER_T_P</b>	-0.02160	0.979	*	<b>MM_C_D</b>	0.36401	1.439	*
<b>MM_C_D</b>	0.33266	1.395	*	<b>INTERVALL_ST_P</b>	-0.30949	0.734	*
<b>INTERVALL_ST_P</b>	-0.26342	0.768					

n = 652; Likelihood Ratio Chi Quadrat = 120.79      n=703; Likelihood Ratio Chi Quadrat = 114.88

Tabelle 9.15 : Vergleich des semiparametrischen Gesamtmodells der zweiten Stufe mit einem Modell der zweiten Stufe für Personen mit regelmäßigen Aktivitäten, Treffen mit Freunden

Gesamtmodell 2. Stufe				Modell für regelmäßige Aktivitäten			
Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN	Variable	Parameter- schätzung ( $\beta$ )	Risk Ratio	SN
<b>SEX</b>	0.32839	1.389	*	<b>SEX</b>	0.26614	1.305	*
<b>SINGLE</b>	0.27145	1.312	*	<b>SINGLE</b>	0.26268	1.300	*
<b>ACC_D</b>	-0.30288	0.739	*	<b>ACC_D</b>	-0.35154	0.704	*
<i>A_P_TT</i>	-0.00363	0.996	**	<b>NR_T_SUM</b>	0.00197	1.002	*
<b>NR_T_SUM</b>	0.00251	1.003	*	<b>A_P_DURST_A</b>	-0.01027	0.990	*
<b>A_P_DURST_A</b>	-0.00783	0.992	*				

n = 534; Likelihood Ratio Chi Quadrat = 75.53      n=466; Likelihood Ratio Chi Quadrat = 54.28

Grundsätzlich sind nur sehr geringfügige Unterschiede zwischen den Intervallen aller Personen und jenen mit regelmäßigen Aktivitäten zu sehen. Bei allgemeinen Freizeitaktivitäten (Tabelle 9.13) sind überhaupt dieselben Einflußgrößen signifikant und nur die Parameterschätzwerte und Hazardraten geringfügig verschieden. Bei der Kategorie Einkauf kurzfristig (Tabelle 9.12) hat nur die Anzahl der Kraftfahrzeuge im Haushalt bei Personen mit regelmäßigen Einkäufen zusätzlich einen signifikanten Einfluß, ansonsten sind ebenfalls dieselben Einflußgrößen ausschlaggebend. Je mehr Autos vorhanden sind, desto höher ist die Wahrscheinlichkeit von kürzeren Intervallen zwischen zwei Einkäufen eines regelmäßigen Einkäufers. Wie vorher schon erwähnt, hat hingegen die Dauer des Intervalls zwischen zwei vorher durchgeführten Einkäufen bei regelmäßigen Einkäufern, wie auch in der Kategorie Dienstleistung allgemein (Tabelle 9.14), keinen signifikanten Einfluß. In den Kategorien Treffen mit Freunden und Dienstleistung allgemein (Tabelle 9.15 und 9.14) fällt auf, daß bei Personen mit regelmäßigen Aktivitäten wesentlich geringere Einflüsse zu erkennen sind bzw. auch, daß sich durch den Einbezug von weniger möglichen Variablen bessere Modelle ergeben.

### 9.3.4 Vollparametrische Modelle mit *PROC LIFEREG*

Wie auch bei den vollparametrischen Auswertungen der Aktivitätsdauern wird bei den semiparametrischen Modellschätzungen der Intervalle diejenige Verteilungsfunktion (*baseline hazard distribution*) gesucht, die am geeignetsten multiplikativ auf die Kovariablen wirkt (vgl. Formeln im Kapitel 6.4). Es werden wiederum alle 5 verschiedenen Möglichkeiten der Verteilungsannahme (vgl. Kapitel 6.4.1: Weibullverteilung, Exponentialverteilung, Log-Logistische Verteilung, Lognormale Verteilung und Standardisierte Gammaverteilung) in Betracht gezogen und aufgrund mehrerer Kriterien und Entscheidungsgrundlagen soll für jede Kategorie jene Verteilungsannahme getroffen werden, die am besten den wahren Verlauf der Hazard- und der Survivalkurve widerspiegelt.

Die gesamten Überprüfungen (Likelihood-Ratio-(Chi-Quadrat)-Statistiken, grafische Methoden, Cox-Snell Residuenplots) und ihre Ergebnisse und Schlußfolgerungen sind im Zusatzanhang XXV (LIFEREG I Vollparametrische Modelle) zu sehen. Sie werden jedoch grundsätzlich äquivalent zu jenen im Kapitel 9.2.4 für die Kategorie Einkauf kurzfristig gezeigten Überprüfungen durchgeführt.

Zum Unterschied zu den Aktivitätsdauern ist hier tatsächlich in jeder untersuchten Kategorie die Annahme einer Weibullverteilung, wie es in vielen anderen Studien gemacht worden ist, eine sehr gute, wenngleich in der Kategorie Dienstleistung allgemein doch die Lognormale Verteilungsannahme geringfügig besser zu sein scheint. In der Kategorie Treffen mit Freunden sind die beiden Annahmen einer Weibullverteilung bzw. einer standardisierten Gammaverteilung nach diversen Überprüfungen absolut gleichwertig. Aufgrund des geringfügig besseren Log-Likelihood-Wertes und einer nachvollziehbaren Nähe zu den Parameterschätzwerten des semiparametrischen Modells wird hier die standardisierte Gammaverteilung gewählt.

Nach der Wahl der geeigneten Verteilungsannahme werden auch hier wie bei den semiparametrischen Auswertungen in jeder Kategorie eine Einzelanalyse der Variablen durchgeführt, Korrelationsmatrizen erstellt und aufgrund dieser Ergebnisse entschieden, welche möglichen Einflußfaktoren als Kovariablen in ein erstes Gesamtmodell einfließen. Die Ergebnisse der ersten Gesamtmodelle sind im Anhang XIII (LIFEREG I Erste Gesamtmodelle der vollparametrischen Auswertung) übersichtlich dargestellt und können so auch problemlos mit den ersten Gesamtmodellen der semiparametrischen Auswertung verglichen werden (siehe Anhang XII).

In einem weiteren Schritt wird wiederum in jeder Kategorie ein Modell der zweiten Stufe erstellt, das nur noch die im ersten Gesamtmodell statistisch signifikanten Kovariablen enthält. Diese Ergebnisse werden in den Tabellen 9.16 bis 9.19 präsentiert und anschließend kurz diskutiert.

Tabelle 9.16 : Modell der zweiten Stufe, Kategorie Einkauf kurzfristig

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	-0.89518	0.15645	32.7381	<.0001
<b>SEX</b>	0.45948	0.07350	39.0854	<b>&lt;.0001</b>
<b>AGE</b>	-0.00423	0.00181	5.4701	<b>0.0193</b>
CAR	-0.12440	0.07619	2.6657	0.1025
<b>PLANNED</b>	0.33304	0.10526	10.0115	<b>0.0016</b>
<b>WEEKEND</b>	-0.32627	0.08106	16.1994	<b>&lt;.0001</b>
A_P_TT	0.01448	0.00393	13.5486	<b>0.0002</b>
INTER_T_P	0.02890	0.01582	3.3385	0.0677
A_P_DURST_A	0.03717	0.00335	123.3442	<b>&lt;.0001</b>
A_P_DURST_C	0.09669	0.05381	3.2285	0.0724
<b>MM_C_D</b>	-0.26858	0.07157	14.0828	<b>0.0002</b>
<b>INTERVALL_ST_P</b>	0.83330	0.16988	24.0625	<b>&lt;.0001</b>
SCALE	1.78958	0.02736		

n = 2917; Log-Likelihood = - 6385.46, Verteilungsannahme: Weibull

Tabelle 9.17 : Modell der zweiten Stufe, Kategorie Freizeit allgemein

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	0.85588	0.13786	38.5417	<.0001
<b>SEX</b>	-0.32141	0.07421	18.7572	<b>&lt;.0001</b>
<b>LICENSE</b>	0.41461	0.09169	20.4455	<b>&lt;.0001</b>
<b>HEAD_OHH</b>	0.20541	0.08963	5.2523	<b>0.0219</b>
<b>WORKED</b>	0.27970	0.07056	15.7144	<b>&lt;.0001</b>
<b>SINGLE</b>	-0.49550	0.06550	57.2298	<b>&lt;.0001</b>
<b>CAR</b>	-0.46129	0.10256	20.2316	<b>&lt;.0001</b>
<b>PLANNED</b>	0.50056	0.09348	28.6749	<b>&lt;.0001</b>
<b>WEEKEND</b>	-0.32274	0.05408	35.6124	<b>&lt;.0001</b>
<b>A_DAYT</b>	-0.28152	0.03789	55.2080	<b>&lt;.0001</b>
<b>NR_T_SUM</b>	-0.00659	0.00043	234.1669	<b>&lt;.0001</b>
A_P_DURST_A	0.01529	0.00216	50.0338	<b>&lt;.0001</b>
A_P_DURST_C	0.06765	0.01142	35.0837	<b>&lt;.0001</b>
<b>MM_C_D</b>	-0.28577	0.05148	30.8126	<b>&lt;.0001</b>
SCALE	1.32302	0.01976		

n = 2790; Log-Likelihood = - 5191.82, Verteilungsannahme: Weibull

Tabelle 9.18 : Modell der zweiten Stufe, Kategorie Dienstleistung allgemein

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	-4.41501	0.38179	133.7283	<.0001
<b>HH_NR_L</b>	-0.24069	0.08796	7.4884	<b>0.0062</b>
<i>NR_T_SUM</i>	0.00282	0.00149	3.5975	<i>0.0579</i>
<b>A_P_PUR</b>	0.48071	0.02774	300.3853	<b>&lt;.0001</b>
<b>A_P_DURST_A</b>	0.01225	0.00431	8.0906	<b>0.0044</b>
<i>INTER_T_P</i>	0.03279	0.01879	3.0448	<i>0.0810</i>
<i>A_P_MM_C</i>	0.10411	0.05520	3.5579	<i>0.0593</i>
<b>MM_C_D</b>	-0.55759	0.16201	11.8453	<b>0.0006</b>
<b>INTERVALL_ST_P</b>	0.77414	0.28810	7.2201	<b>0.0072</b>
SCALE	1.90722	0.05282		

n = 652; Log-Likelihood = - 1346.11 , Verteilungsannahme: Lognormal

Tabelle 9.19 : Modell der zweiten Stufe, Kategorie Treffen mit Freunden

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	1.40772	0.19540	51.9046	<.0001
<b>SEX</b>	-0.40384	0.11939	11.4413	<b>0.0007</b>
<b>SINGLE</b>	-0.34757	0.12030	8.3474	<b>0.0039</b>
<b>ACC_D</b>	0.36026	0.13687	6.9281	<b>0.0085</b>
<i>A_P_TT</i>	0.00502	0.00270	3.4413	<i>0.0636</i>
<b>NR_T_SUM</b>	-0.00268	0.00095	7.8864	<b>0.0050</b>
<i>A_P_DURST_A</i>	0.00852	0.00437	3.8025	<i>0.0512</i>
SCALE	1.30244	0.04590		

n = 533; Log-Likelihood = - 1000.03 , Verteilungsannahme: Weibull

Grundsätzlich weisen die hier erzielten Ergebnisse wiederum keine gravierenden Unterschiede zu denen der semiparametrischen Auswertungen auf. Einzig das Vorzeichen der Parameterschätzwerte ist bei allen Werten umgedreht, weil die vollparametrischen Werte wieder im Log-Survival-Zeitformat angegeben werden. Ein negatives Vorzeichen vor dem Parameterschätzwert bedeutet hier also eine Verkürzung der Intervalldauer bzw. eine negative Beeinflussung derselben. Es ist jedoch keine einzige Auswirkung einer Variable in den vollparametrischen Modellen anders als im semiparametrischen Modell. Verlängert eine Einflußgröße dort die Intervalldauer, so hat sie hier dieselbe Richtung.

Das Geschlecht hat auch hier in allen anderen Kategorien, ausgenommen den Dienstleistungsaktivitäten, einen deutlichen Einfluß. Männer haben wesentlich größere Intervalle zwischen Einkaufsaktivitäten,

ten, Frauen hingegen führen ihre allgemeinen Freizeitaktivitäten und Freundestreffen mit hoher Wahrscheinlichkeit in wesentlich größeren Abständen durch.

In allen vier Kategorien wird wie im semiparametrischen Modell die Intervalldauer von der Dauer der letzten durchgeführten beliebigen Aktivität verlängert. Die Signifikanz der Einflußgrößen ist jedoch etwas geringer. Die Dauer der letzten gleichartigen Aktivität vor dem Intervall hat wieder nur in den beiden Kategorien Einkauf und Freizeit einen signifikanten Einfluß. Auch hier dauert das Intervall länger, je länger die Dauer der Aktivität vor dem Intervall ist.

Die Dauer des Intervalls zwischen zwei gleichartigen Aktivitäten vor dem betrachteten Intervall hat weitaus geringeren Einfluß als bei den semiparametrischen Auswertungen. In der Freizeitkategorie fällt sie gänzlich heraus, in den Kategorien Einkauf und Dienstleistungen verlängert ein größeres Intervall die begutachtete Intervalldauer mit marginaler Signifikanz. Diesen Effekt sieht man auch bei den semiparametrischen Modellen für Personen mit regelmäßigen Aktivitäten in der jeweiligen Kategorie. Die Dauer des Intervalls zwischen der Aktivität dieser Kategorie zu Beginn des betrachteten Intervalls und der beliebigen Aktivität vorher ist wieder in den Kategorien Einkauf und Dienstleistung stark signifikant. Je länger dieses Intervall ist, desto größer ist auch das abhängige Intervall.

Die Tageszeit der das Intervall abschließenden Aktivität hat bei Einkaufsaktivitäten im Gegensatz zu den semiparametrischen Modellen keinen Einfluß. Die Signifikanz bei Freizeitaktivitäten bleibt bestehen. Je später am Tag sie durchgeführt werden, desto höher ist die Wahrscheinlichkeit, daß sie in kürzeren Zeitintervallen stattfinden.

Ob ein motorisiertes Verkehrsmittel zu der das Intervall abschließenden Aktivität benützt wird, hat außer bei Treffen mit Freunden wieder überall einen signifikanten Einfluß. Mit der Möglichkeit, ein motorisiertes Fahrzeug mit bzw. selbst zu benützen, wird die Wahrscheinlichkeit einer kürzeren Intervalldauer zwischen zwei gleichartigen Aktivitäten wesentlich größer.

Eine höhere Anzahl aller gemachten Wege der jeweiligen Person hat in den beiden Freizeitkategorien wiederum eine signifikant kürzere Intervalldauer zwischen zwei Freizeitaktivitäten zur Folge. Bei Dienstleistungsaktivitäten ist hier ein marginal signifikanter Einfluß zu sehen. Bei mehr Wegen pro Person insgesamt, steigt jedoch die Intervalldauer zwischen zwei Dienstleistungsaktivitäten.

Wenige weitere Einflußgrößen sind zu den semiparametrischen Modellschätzungen verschieden. So haben in der Kategorie Einkauf kurzfristig (Tabelle 9.16) die Anzahl der Führerscheine im Haushalt, der Führerscheinbesitz, sowie die Tageszeit der das Intervall abschließenden Aktivität hier keinerlei signifikanten Einfluß. Sie sind (mit Ausnahme der Tageszeit) jedoch auch in den semiparametrischen

Modellen teilweise nur marginal bzw. gar nicht signifikant. Bei Freizeitaktivitäten (Tabelle 9.17) hat der Haushaltsvorstand die signifikant größeren Intervalldauern als die übrigen Familienmitglieder, die vorherige Intervalldauer hat hingegen keinen Einfluß. In der Kategorie Dienstleistung allgemein (Tabelle 9.18) verlängern bei den vollparametrischen Modellen höhere Werte der Einflußgrößen Anzahl der Wege gesamt und Art des Verkehrsmittels zur Aktivität vor dem Intervall die Intervalldauer signifikant. In der Kategorie Treffen mit Freunden (Tabelle 9.19) sind exakt dieselben Einflußgrößen wie bei den semiparametrischen Auswertungen ausschlaggebend. Wie auch bei den semiparametrischen Modellen sind sehr wenige soziodemografische Variablen signifikant. Es dominieren wieder in allen vier Kategorien die aktivitätsabhängigen bzw. intervallbezogenen Einflußgrößen.

Auffällig ist jedoch die Tatsache, daß mit Ausnahme der Treffen mit Freunden in allen Kategorien die Parameterschätzwerte deutlich höher als bei den semiparametrischen Modellschätzungen liegen. Bei Intervallen zwischen Dienstleistungsaktivitäten liegen die Werte sogar bei den meisten Variablen beim dreifachen Wert, bei den Intervallen zwischen Einkaufsaktivitäten größtenteils beim zumindest doppelten Parameterschätzwert.

## 10 SCHLUSSFOLGERUNGEN UND WEITERE AUSBLICKE

### 10.1 ERGEBNISSE

In früheren Studien wurden sehr oft die soziodemografischen Charakteristika von Personen und Haushalten als jene Faktoren eingestuft, die das Reiseverhalten und die Aktivitätsmuster maßgebend beeinflussen. Wie aus den Auswertungen der deskriptiven Analyse im Kapitel 8 und denen der Hazardmodelle im Kapitel 9 hervorgeht, sind die Einflüsse der soziodemografischen Merkmale hier jedoch weit aus geringer. Nur wenige personenspezifische Variablen standen nach diversen Überprüfungen mit Korrelationsmatrizen, Proportionalitätstests und jeweils einem ersten Gesamtmodell überhaupt noch für ein zweites Gesamtmodell zur Verfügung bzw. zeigten in diesem dann signifikante Einflüsse. In den aussagekräftigen Modellen überwiegen die aktivitäts- und reisespezifischen Variablen. Die wichtigsten Ergebnisse der Modellauswertungen werden hier sowohl für die Aktivitätsdauern, als auch für die Intervalldauern noch einmal zusammengefaßt, um geeignete Schlußfolgerungen ziehen zu können.

Die soziodemografischen Faktoren haben generell nur vereinzelt signifikanten Einfluß auf die Dauer der Aktivitäten. Einzige Ausnahme bildet dabei das Geschlecht, das in allen Kategorien Auswirkungen hat. Weibliche Umfrageteilnehmer haben bei Einkäufen und Dienstleistungsaktivitäten eine signifikant längere Aktivitätsdauer aufzuweisen, bei Freizeitaktivitäten und Freundestreffen sind keinerlei gravierende Unterschiede zwischen den Geschlechtern festzustellen. Andere Faktoren, die eigentlich Indikatoren einer größeren Flexibilität in bezug auf die Aktivitätsgestaltung und insofern auch auf die Dauer von Aktivitäten sind, wie Führerscheinbesitz oder Zugangsmöglichkeiten zu einem privaten, unabhängigen PKW, etc. scheinen mit Ausnahme der Einkaufsaktivitäten keinen allzu großen Einfluß auf die Aktivitätsdauer zu haben. Zeitliche Rahmenbedingungen, die in dieser Studie vor allem aufgrund der nicht vorhandenen Informationen nicht in die Auswertungen mit eingehen, scheinen eine noch gewichtigere Rolle zu spielen. Der Ausbildungsgrad der befragten Personen, das Alter oder die Frage, ob die Person in der Umfragezeit gearbeitet hat oder nicht, nehmen keinen signifikanten Einfluß auf die Aktivitätsdauer.

Bei der Intervalldauer haben jedoch sowohl das Geschlecht, als auch die Flexibilitätsfaktoren bei allen ausgewählten Kategorien einen signifikanten Einfluß. So haben Männer eine weitaus höhere Wahrscheinlichkeit, lange Intervalle zwischen zwei Einkaufsaktivitäten und signifikant kürzere Abstände

zwischen Freizeitaktivitäten und Treffen mit Freunden zu haben, als die weiblichen Umfrageteilnehmer. Die Möglichkeit, ein motorisiertes, unabhängiges Verkehrsmittel benutzen zu können, erhöht die Wahrscheinlichkeit in allen Kategorien mit Ausnahme der Treffen mit Freunden, die Aktivitäten in kürzeren Abständen durchzuführen. Dies bedeutet, daß mobilere Personen zwar nicht unbedingt längere Aktivitäten durchführen, aber wesentlich häufiger. Ausbildungsgrad, Anzahl der Haushaltsmitglieder, Kinder im Haushalt, Alter, etc. spielen beim Einfluß auf die Intervalldauer meist eine untergeordnete oder gar keine Rolle.

Das tatsächlich benützte Verkehrsmittel vor bzw. nach der Aktivität spielt in bezug auf die Aktivitätsdauer eine weitaus geringere Rolle als angenommen bzw. erwartet, auf die Intervalldauer hat es überhaupt keinen signifikanten Einfluß. Bei Freizeitaktivitäten hat nur die Frage, ob vor der Aktivität ein von Fahrplänen etc. unabhängiges Verkehrsmittel benützt worden ist, einen Einfluß, beim Einkaufen ein unabhängiges Verkehrsmittels danach. Bei Dienstleistungen spielt nur die Art des benützten Verkehrsmittel zum Aktivitätsort eine Rolle, bei Treffen mit Freunden das Verkehrsmittel danach.

Die Reisezeit hat sowohl auf die Aktivitätsdauer, als auch auf die Intervalldauer zumeist einen signifikanten Einfluß. Je länger die Reisezeit vor oder nach der Aktivität ist - wobei aufgrund der zu großen Korrelation nur eine der beiden ins Modell eingegangen ist - desto länger ist auch in allen Kategorien die Dauer der Aktivität an sich. Bei den Intervalldauern hat eine längere Reisezeit nur in der Kategorie Freizeit keine signifikante Verlängerung zur Folge. Vor allem bei Einkäufen und Dienstleistungsaktivitäten macht es einen Unterschied, wie lange der Weg dorthin ist. Je länger die Anreisezeit zum Geschäft ist, desto weniger oft wird die Tätigkeit durchgeführt.

Monetäre Ausgaben während der Aktivität verlängern kurzfristige Einkäufe und Dienstleistungen, sie verkürzen hingegen Freizeitaktivitäten. Treffen mit Freunden sind sogar deutlich kürzer, wenn Geld ausgegeben wird. Bei den Intervalldauern spielen die Ausgaben keinerlei Rolle.

Die Tageszeit der Aktivität hat meist nur bei der Dauer einen signifikanten Einfluß. Je später am Tag eine Einkaufsaktivität stattfindet, desto kürzer ist sie, bei Freizeitaktivitäten und Treffen mit Freunden verhält es sich genau umgekehrt. Sie dauern am Nachmittag rund 50% länger. Für die Intervalldauern spielt die Tageszeit nur bei Freizeitaktivitäten eine Rolle. Je später am Tag die Aktivität stattfindet, desto höher ist die Wahrscheinlichkeit, daß bis zur nächsten gleichartigen weniger Zeit vergeht. Hierbei wäre eine Variable die Vereinszugehörigkeit betreffend sehr interessant, die in den Umfragen jedoch nicht erhoben wurde.

Weitere signifikant beeinflussende Variablen sind die Art der Aktivität vorher auf die Aktivitätsdauer bzw. die Länge des vorherigen Intervalls auf die Dauer des nächstfolgenden. Je länger das Intervall

zwischen zwei Aktivitäten derselben Kategorie war, desto höher ist die Wahrscheinlichkeit, daß wiederum eine lange Zeitspanne zwischen den nächsten beiden Aktivitäten vergeht. Dies kann jedoch auch auf ein gewisses regelmäßiges bzw. gewohnheitsmäßiges Verhalten hindeuten. Es sagt im Prinzip nur aus, daß die meisten Umfrageteilnehmer ihre Aktivitäten nicht abwechselnd in kurzen und langen Zeitabständen, sondern konstant verteilt durchführten.

Zwischen den Auswertungen der semiparametrischen und der vollparametrischen Modellstufe sind bei den Aktivitätsdauern meist nur sehr geringfügige Unterschiede zu erkennen, in den Freizeitkategorien sogar fast identische Ergebnisse vorhanden. Bei den Intervalldauern sind zwar nahezu dieselben Variablen signifikant, die Parameterschätzwerte der vollparametrischen Modelle nehmen jedoch meist den dreifachen Wert an, d.h. die Größe des Einflusses wird unterschiedlich bewertet. Der Grund hierfür kann nicht genau angegeben werden. Es kann einerseits eine bessere Paßform der von vornherein robusten semiparametrischen Modelle oder andererseits auch eine gute Wahl der Verteilungsannahmen für die vollparametrischen Modelle sein.

## 10.2 SCHLUSSFOLGERUNGEN

Nachdem die Umfrage der UTHS (Uppsala Household Travel Survey) in Uppsala (Schweden) vor mehr als 30 Jahren im Frühjahr 1971 durchgeführt worden ist (Marble *et al.*, 1972), sind die Ergebnisse dieser Arbeit nicht unbedingt sofort auf die heutige Zeit übertragbar. Zu viele Faktoren haben in den letzten Jahren das Gesamtbild des Verkehrs geprägt und verändert. So ist beispielsweise das Gesamtverkehrsaufkommen auf der ganzen Welt sehr stark angestiegen, aber auch die Anzahl der motorisierten Wege und Reisen jeder einzelnen Person hat meist zugenommen. Die Zugangsmöglichkeiten zu motorisierten unabhängigen Fahrzeugen und zu Fahrberechtigungen sind hingegen für alle Bevölkerungsschichten größtenteils unabhängig von ihrem finanziellen Niveau weitaus erleichtert worden. Die regionalen Netze des öffentlichen Verkehrs wurden und werden stark ausgebaut, erweitert und verbessert und so sind die Möglichkeiten, mit dem öffentlichen Verkehr viele Aktivitätsorte zu erreichen, stark angestiegen. Dem gegenüber steht jedoch auch die gegenüber den 70er Jahren ungemein höhere Nachfrage nach Erholungsaufenthalten in der Natur und bei sportlichen Aktivitäten und ein dadurch gesteigertes Reiseverhalten. Dieser Freizeitverkehr hat vor allem in den Industrieländern der ersten Welt rapide zugenommen.

Aus all diesen genannten Gründen ist ein direkter Vergleich mit heutigen Umfragen relativ schwierig. Die Auswertungen und Ergebnisse dieser Studie lassen jedoch einen kleinen Einblick in die Aktivitätsmuster der befragten Personen in dieser Zeit zu. Es werden in der Arbeit nebst der überblicksmäßi-

gen Betrachtung aller relevanten Aktivitätskategorien vier Kategorien näher analysiert. Bei den vier detailliert angeführten Kategorien (Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden) wird vor allem versucht, mit Hilfe der Hazardmodell-Analyse die Unterschiede zwischen den Geschlechtern und verschiedenen Altersgruppen heraus zu arbeiten und die einzelnen möglichen Einflußgrößen auf ihre Wirkung die Dauer der Aktivitäten einerseits bzw. andererseits die Dauer der Intervalle zwischen den Aktivitätsteilnahmen, also die Periodizität der Ereignisse, betreffend, zu ermitteln und auch zahlenmäßig festzulegen. Am Schluß wird jeweils noch ein Vergleich zwischen Personen mit regelmäßigen Teilnahmen an Aktivitäten der entsprechenden Kategorie und jenen Personen, die nur sporadisch solche Aktivitäten durchführen, angestellt. Die wichtigsten Ergebnisse sind bereits in der Einleitung dieses Kapitels zusammengefaßt.

Sicherlich wäre eine Analyse aller durchgeführten Aktivitäten jeder Person sehr interessant, um die Vollständigkeit der Aktivitätsmuster zu gewährleisten und die Komplexität des Reise- und Aktivitätsverhaltens aufzuzeigen. Es ist jedoch einerseits die Periodizität der Aktivitätsteilnahme und auch die Anzahl der Aktivitäten in den einzelnen Kategorien ziemlich unterschiedlich und andererseits liegen in mehreren Aktivitätskategorien einfach viel zu wenige Ereignisse vor, um diese als eine repräsentative Stichprobengröße ansehen zu können. Die Ergebnisse der semiparametrischen und vollparametrischen Untersuchungen hätten mit einer zu geringen Anzahl an Aktivitäten bzw. Intervallen zu wenig bis gar keine Aussagekraft.

Der direkte Vergleich der Ergebnisse des Datensatzes aus Uppsala mit neueren Umfragen wird zudem durch eine andere und meist ausführlichere Art der Fragestellung bei Befragungen in der heutigen Zeit bzw. die aus unserer Sicht partielle Unvollständigkeit der vorliegenden Daten aus der Uppsala-Umfrage erschwert. Um ein umfassendes Bild der Aktivitätsmuster zu erhalten und die Gründe für unterschiedliche Verhaltensweisen eruieren zu können, ist eine sehr detaillierte Befragung immer von Vorteil. Es konnten zwar damals erstmals mit Hilfe des schwedischen Bevölkerungsregisters einige soziodemografische Kenngrößen der befragten Personen sofort ermittelt werden und auch weitere interessante Charakteristika in den Datensatz integriert werden, jedoch ist der Umfang der nebst den eigentlichen Aktivitätsteilnahmen erhobenen Faktoren und Daten in keinster Weise mit heutigen Umfragen vergleichbar. So wurden beispielsweise bei der *Mobidrive*- Befragung in Halle/Saale und Karlsruhe im Herbst 1999 (vgl. Axhausen und Schönfelder, 2000) weitaus mehr Informationen über die Sozio-Demografie der Reisenden (Beruf, Einkommensverhältnisse, Fahrzeugverfügbarkeit, Dauerkarten für den öffentlichen Verkehr, Telekommunikationsstand) und insbesondere deren Rolle im Haushalt (Haushaltszusammensetzung, Haushaltsvorstand, Anzahl und Alter der Kinder, Aufgabenverteilung im Haushalt, etc.) erhoben und in den Datensatz integriert. Auch eine zusätzliche Befragung außerhalb der eigentlichen Mobilitätsbefragung, die unter anderem die Gruppen- und Vereinszugehörigkeit, die Einteilung des disponiblen Einkommens oder die Einstellung zu verschiedenen Verkehrs-

mitteln zum Inhalt hatte, wurde durchgeführt, um Aufschlüsse über die Werthaltungen und Einstellungen der Befragten zu erhalten. Mit Hilfe gewisser Hierarchien und Berechnungen konnten für diese Arbeit aus dem Uppsala-Datensatz zwar einige wenige Kenngrößen ermittelt werden (Zugangsmöglichkeiten zu motorisierten Fahrzeugen, Haushaltsstruktur, Hauptverkehrsmittel der befragten Personen, etc.), ein großes Informationsdefizit zu heutigen Umfragen bleibt jedoch bestehen. Dieses Defizit an geeigneten zusätzlichen Charakteristika und Verhaltensweisen der Umfrageteilnehmer bemängelten auch schon Filiou (1994) und Wilson (1994), die ebenfalls mit den Daten aus Uppsala arbeiteten und damit Studien erstellten (siehe auch Kapitel 7.2).

### **10.3 WEITERE AUSBLICKE**

Die hier entwickelten und angewendeten Methoden zur Bestimmung der Einflußfaktoren von Aktivitätsdauer und Dauer von Intervallen zwischen gleichartigen Aktivitäten in ihrer Art und Größe können jedoch durchaus auf die Analysen neuerer Daten übertragen werden. Trotz des weitaus geringeren Informationsgehalts würde auch einer eventuellen Betrachtung der Unterschiede der Verhaltensmuster der damaligen Zeit mit denen neuerer Umfragen nichts im Wege stehen. Ein solcher Vergleich wäre sinnvoll, um Veränderungen des Verkehrsverhaltens und der Aktivitätsteilnahme bzw. der Verteilung derselben im Laufe der Jahre aufzeigen zu können. Eine Variante wäre dabei ein Vergleich mit den ebenfalls über einen längeren Zeitraum (von 6 Wochen) erhobenen Daten aus Halle/Saale und Karlsruhe (Axhausen und Schönfelder, 2000), wobei natürlich die unterschiedlichen sozialen, wirtschaftlichen und verkehrstechnischen Veränderungen im Umfeld im Laufe der Zeit mit einem gewissen Faktor eingerechnet werden müssten.

Noch interessanter wäre natürlich der Vergleich der Umfragedaten aus den 70er Jahren mit einer neuerlichen Befragung in Uppsala, da dabei die Veränderungen im selben Gebiet beobachtet werden könnten. Lediglich geänderte Einflußfaktoren und Umweltbedingungen in bezug auf die wirtschaftlichen und arbeitstechnischen Randbedingungen, die geänderte Sozialstruktur der Gesellschaft und der Ausbau des öffentlichen Personennahverkehrs, sowie des Straßen- und Radwegnetzes würden dann das Verkehrs- und Reiseverhalten der neuen Umfrageteilnehmer direkt beeinflussen. Da solche Umfragen jedoch sehr aufwendig und kostenintensiv sind, ist eine neuerliche Befragungsstudie in Uppsala eher unwahrscheinlich.

# LITERATURVERZEICHNIS

## REFERENCES

- Allison, P. D. (1995) *Survival Analysis Using the SAS<sup>®</sup> System: A Practical Guide*, SAS Institute Inc., Cary.
- Armoogum, J. und J.-L. Madre (1997) Item (Sampling, Weighting, Nonresponse), for Transport Surveys; *Raising the Standard, International Conference of Transport Survey Quality and Innovation*, Grainau, May 1997.
- Axhausen, K.W. und J.W. Polak (1991) in G. Papageorgiou (Hrsg.) *Concise Encyclopaedia of Traffic and Transportation Systems*, Oxford, Pergamon, 564-569.
- Axhausen, K.W., A. Zimmermann, S. Schönfelder, G. Rindsfuser und T. Haupt (2002) Observing the rhythms of daily life: A six-week travel diary, *Transportation*, **29** (2) 95-124.
- Bhat, C.R. (1995) A model of post home-arrival activity participation behavior, University of Massachusetts, Amherst.
- Bhat, C.R. (1996) A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity, *Transportation Research B*, **30** (3) 189 – 207.
- Bhat, C.R. (1996) A generalized multiple durations proportional hazard model with an application to activity behaviour during the evening work-to-home commute, *Transportation Research B*, **30** (6) 465 - 480.
- Bhat, C.R. (1998) A comprehensive and operational analysis framework for generating the daily activity-travel pattern of workers, Technical Paper, University of Texas, Austin.
- Chapin, F.S. (1974) *Human activity patterns in the city: Things people do in time and space*. John Wiley and Sons, London.
- Cox, D.R. (1972) Regression models and life tables, *Journal of the Royal Statistical Society*, **26B**, 186-220.
- Dörner, D. (1989) *Die Logik des Mißlingens, Strategisches Denken in komplexen Situationen*, Rowohlt Verlag GmbH, Reinbek.
- Filiou, A. (1994) *Hazard models: A new approach to understanding the activity patterns of employed people in Uppsala, Sweden*, MS Thesis, University of London Centre for Transport Studies, Imperial College, London.
- Götz, K., T. Jahn, und I. Schultz (1997) Mobilitätsstile: Ein sozial-ökologischer Untersuchungsansatz, *Forschungsbericht Stadtverträgliche Mobilität*, **7**, Forschungsverbund City:mobil, Frankfurt am Main.

- Hägerstrand, T. (1970) What about people in regional science?, *Papers of the Regional Science Association*, **24** (1) 7-21.
- Hamed, M.M., D.A. Hensher und H.R. Al-Masaeid (1995) The duration between traffic accidents in the taxi sector: An empirical inquiry, *Working Paper*, **IST-WP-95-2**, University of Sydney, Sydney.
- Hamed, M.M. und F.L. Mannering (1990) Modeling travelers' post-work activity involvement: Toward a new methodology, *Working Paper*, Department of Civil Engineering, University of Washington, Seattle.
- Hanson S.E. (1980) The Importance of the multi-purpose journey to work in urban travel behaviour, *Transportation*, **9** (3) 229-248.
- Hanson S.E. (1982) Relative Location and Sociodemographic Factors, in R.G. Golledge und R.J. Stimmien *The determinants of daily travel patterns in urban geography*, **3** (3) 179 – 201.
- Hanson S.E. und P.O. Hanson (1980) Gender and urban activity patterns in Uppsala, Sweden, *The Geographical Review*, **70** (2) 291-299.
- Hanson S.E. und P.O. Hanson (1981a) The impact of married women's employment on household travel patterns: A Swedish example, *Transportation*, **10** (2) 165-183.
- Hanson S.E. und P.O. Hanson (1981b) The travel-activity pattern of urban residents: Dimensions and relationships to sociodemographic characteristics, *Economic Geography*, **57W** (4) 332-347.
- Hanson S.E. und P.O. Hanson (1993) The geography of everyday life, in Gärling, T. & Golledge, R. G. *Behaviour and Environment: Psychological and Geographical Approaches*, 249 – 269, Elsevier Science Publishers, Oxford.
- Hanson, S.E. und J.O. Huff (1982) Assessing day-to-day variability in complex travel patterns, *Transportation Research Record*, **891**, 18 - 24.
- Hanson, S.E. und J.O. Huff (1988) Repetition and day-to-day variability in individual travel patterns: implications for classification, in R.G. Golledge und H. Timmermanns (Hrsg.) *Behavioural Modelling in Geography and Planning*, 368 – 399, Croom Helm, London.
- Hanson, S.E. und J.O. Huff (1988) Systematic variability in repetitious travel, *Transportation*, **15** (1-2) 111–135.
- Hanson S.E. und I. Johnston (1985) Gender differences in work-trip length: Explanations and implications, *Urban Geography*, **6** (3) 193 – 219.
- Hanson, S.E. und M. Schwab (1987) Accessibility and intraurban travel, *Environment and Planning A*, **19** (7) 735 – 748.
- Hensher, D.A. und F.L. Mannering (1993) Hazard-based duration models and their application to transport analysis, *Working Paper*, **IST-WP-93-1**, Institute of Transport Studies, University of Sydney, Sydney.
- Huff, J.O. und S.E. Hanson (1986) Repetition and variability in urban travel, *Geographical Analysis*, **18** (2) 97 – 114.

- Huff, J.O. und S.E. Hanson (1990) Measurement of habitual behaviour: Examining systematic variability in repetitive travel, in P.M. Jones (Hrsg.) *New Approaches in Dynamic and Activity Based Approaches to Travel Analysis*, 229 – 249, Avebury, Aldershot.
- Jones, P.M. (1981) Activity approaches to understanding travel behaviour, in P.R. Stopher, A.H. Meyerburg und W. Brög *New Horizons in Travel-Behaviour Research*, 253-266, Lexington, Massachusetts / Toronto.
- Kalbfleisch, J.D. und R.L. Prentice (1980) *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Kiefer, N.M. (1988) Economic duration data and hazard functions. *Journal of Economic Literature*, **26** (2) 646-679.
- Kitamura, R. (1988) An evaluation of activity-based travel analysis, *Transportation*, **15** (1) 9-34.
- Kleinbaum, D. G. (1996) *Survival Analysis, A Self-Learning Text*, Springer Verlag, New York.
- Lancaster, T. (1990) *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.
- Marble, D.F., P.O. Hanson and S.E. Hanson (1972) Household Travel Behavior Study, *Report No. 1*, Northwestern University.
- Oh, L. (2000) Identification of the causal factors of activity duration dynamics using hazard based duration models, Working paper, Imperial College, London.
- Pas, E.I. und A.S. Harvey (1997) Time use research and travel demand analysis modelling, in P.R. Stopher und M.E.H Lee-Gosselin (Hrsg.) *Understanding Travel Behaviour in an Era of Change*, 315-338, Pergamon, Oxford.
- Schönfelder, S. und K.W. Axhausen (2000a) *Analysing the rhythms of travel using survival analysis, Working Paper*, Institut für Verkehrsplanung und Transporttechnik, ETH, Zürich.
- Schönfelder, S. und K.W. Axhausen (2000b) Periodizität im Verkehrsverhalten: Erste Ergebnisse mit Überlebenszeitmodellen, *Tagungsband AMUS 2000 - Stadt Region Land*, **69**, 131 - 144, Institut für Stadtbauwesen, RWTH Aachen.
- Wilson, P. (1994) *The Use of Hazard Models to Analyse Travel Behaviour of the Unemployed*, MSC Thesis, University of London Centre for Transport Studies, Imperial College, London.
- Wilson, W.C. (1998) Activity pattern analysis by means of sequence-alignment methods, *Environment and Planning A*, **30** (6) 1017-1038.
- Zmud, J. und C.H. Arce (1997) Item Nonresponse in Travel Surveys: Causes and Solutions for Transport Surveys: *Raising the Standard: International Conference of Transport Survey Quality and Innovation*, Grainau, May 1997.

# Anhang I : Liste aller Variablen

Im folgenden Anhang I sind die in den einzelnen Datensätzen verwendeten Variablen detailliert erklärt. In der ersten Spalte sind die Variablenabkürzungen bzw. Bezeichnungen, die in den insgesamt fünf verschiedenen Datensätzen verwendet worden sind, angeführt, in der zweiten Spalte dann die jeweilige Erklärung in deutscher Sprache. Im Zusatzanhang XXX sind diese Erläuterungen auch in einer englischen Version zu sehen.

Die Grunddaten sind durch die SAS-Programme split und create\_act in fünf prinzipiell unterschiedliche Datensätze aufgeteilt worden. Dadurch liegen nun ein Personendatensatz (Tabelle I.1), ein Haushaltsdatensatz (Tabelle I.2), ein Reisedatensatz (Tabelle I.3), ein Wegedatensatz (Tabelle I.4) und ein Aktivitätsdatensatz (Tabelle I.5) vor, wobei der letztgenannte für die unterschiedlichen Berechnungen und Modellschätzungen noch einmal aufgeteilt bzw. verändert worden ist. Die in Tabelle I.5 angeführten Variablenbezeichnungen sind jedoch für alle vorhandenen und verwendeten Aktivitätsdatensätze gültig. Im grundlegenden Aktivitätsdatensatz (upps.act) sind nur die am Ende der Tabelle I.5. angeführten Variablen nicht vorhanden. In den für die Auswertungen verwendeten Aktivitätsdatensätzen für Zeitdauern (upps.act\_dur) und Intervalldauern (upps.act\_int) sind alle 193 beschriebenen Variablen angeführt, es ist jedoch die Anzahl der Beobachtungen (17291 bzw. 17147) unterschiedlich. Vor allem aus den Intervalldatensätzen der einzelnen für die Modellierung relevanten und interessanten Kategorien sind all jene Beobachtungen, deren Intervalldauer mehr als 2 Wochen beträgt, entfernt worden, wodurch sich nur die Anzahl der Beobachtungen, nicht aber die der Variablen verändert hat.

Tabelle I.1 : Personendaten.- person data, deutsche Version

---

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
id	Personen-identifikations-nummer
y_o_b	Geburtsjahr
sex	Geschlecht
shop_nr	Anzahl der Einkäufe
work_new	Arbeitsklasse (abhängig von den Arbeitsstunden)
hours_wk	Angegebene Anzahl der Wochenarbeitsstunden
hourn_wk	Berechnete Anzahl der Wochenarbeitsstunden
wife_emp	Beschäftigung der Ehefrau (ja/nein)
license	Führerscheinbesitz (ja/nein)
driven	Autobenützung in der Umfrageperiode (ja/nein)
work_sum	Berechnete Gesamtarbeitszeit
n_t_work	Anzahl der Arbeitswege

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
n_d_work	Anzahl der Arbeitstage
n_tdwork	Anzahl der Arbeitswege pro Tag , Maximum
shgr_sum	Summe der Einkaufszeiten (Lebensmittel)
n_t_shgr	Anzahl der Einkäufe (Lebensmittel)
age	Alter
n_t_shwc	Anzahl der Einkäufe (Frauenbekleidungsgeschäft)
hh_id	Haushaltsidentifikationsnummer
nr_t_m	Maximale Anzahl der Wege der Person pro Tag
t_nr_s	Gesamte Anzahl der Wege der Person im Umfragezeitraum
s	Strom
g	Gruppe
h	Haushalt
i	Individuum
nr_j_sum	Gesamtanzahl der Reisen der Person im Umfragezeitraum
city	Stadt der Umfrage (Uppsala)
citycode	Codierung der Stadt der Umfrage (3)
hh_nr	Haushaltsnummer
p_nr	Personennummer im Haushalt
study	Ort der Befragung
studycode	Codierung des Befragungsortes
nr_t_sum	Gesamtanzahl der Wege der Person im Umfragezeitraum (korr.)
p_tt	Fahrzeit der Person
p_actt	Aktivitätszeit der Person
p_j_time	Reisezeit der Person
p_acc_s	Anzahl aller begleiteten Wege
p_acc	Anteil der Reisen in Begleitung
nr_j_dav	durchschnittliche Anzahl der Reisen pro Tag
nr_t_jav	durchschnittliche Anzahl der Wege pro Reise
wave	Welle
children	Kinder im Haushalt
single	alleinlebend (ja/nein)
young	unter / über 50
educ_yr	Anzahl der Ausbildungsjahre
profcode	Codierung des Berufs
worked	Beschäftigung im Beobachtungszeitraum (ja/nein)
educ	Status des Ausbildungsgrades
job	Status des Beschäftigungsgrades
school	Schulbildung im Umfragezeitraum (ja/nein)
nr_j_max	maximale Anzahl an Reisen der Person
mp_act_r	Rang der Hauptaktivität der Person
mp_exp	höchste Ausgabe der Person
mp_expac	höchste Ausgabe der Person auf der Hauptaktivität
mp_explu	höchste Ausgabe der Person auf dem Hauptaktivitätsort
mp_mode	Hauptverkehrsmittel der Person

Tabelle I.2 : Haushaltsdaten - Household data, deutsche Version

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
hh_id	Haushalts-identifikations-nummer
car	Autobesitz (ja / nein)
hh_nr_l	Anzahl der Führerscheine im Haushalt
hh_nr_d	Anzahl der Personen im Haushalt, die unterwegs waren
hhshopnr	Summe der Einkaufszeiten pro Haushalt
hhhoursw	Summe der angegebenen Wochenarbeitszeit pro Haushalt
hhhournw	Summe der berechneten Wochenarbeitszeit pro Haushalt
hhnr_j_s	Anzahl aller Reisen der Haushaltsmitglieder
hhwork_s	Summe der gesamten Arbeitszeit pro Haushalt (in 5 Wochen)
hhntwork	Anzahl der Arbeitswege (wie oft gearbeitet wurde) pro Haushalt
hhndwork	Anzahl der Arbeitstage pro Haushalt
hhshgr_s	Summe der Einkaufszeiten (Lebensmittel) pro Haushalt
hhntshgr	Anzahl der Einkäufe (Lebensmittel) pro Haushalt
hh_age	Gesamalter des Haushalts
hhntshwc	Anzahl der Einkäufe (Frauenbekleidung) pro Haushalt
hhnr_t_s	Anzahl der Wege aller Haushaltsmitglieder
s	Strom nach UTHS
g	Gruppe nach UTHS
h	Haushaltsnummer nach UTHS
home_add	Individuum im Haushalt nach UTHS
hh_nr	Adresse (code)
marst	Haushaltsnummer
hh_memb	Marst-angabe (UTHS)
hhnr_j_ma	Anzahl der Haushaltsmitglieder
hhnr_t_ma	Maximale Anzahl der Reisen pro Haushalt
shop_nrm	Maximale Anzahl der Wege pro Haushalt
hours_wm	Durchschnittliche Einkaufszeit (Haushalt)
hourn_wm	Durchschnittliche angegebene Wochenarbeitszeit (Haushalt)
nr_j_sm	Durchschnittliche berechnete Wochenarbeitszeit (Haushalt)
nr_t_sm	Durchschnittliche Anzahl der Reisen (Haushalt)
work_sm	Durchschnittliche Arbeitszeit des Haushalt (über 5 Wochen)
n_t_wm	Durchschnittliche Anzahl der Arbeitswege (Haushalt)
n_d_wm	Durchschnittliche Anzahl der Arbeitstage (Haushalt)
n_tdw	Durchschnittliche Anzahl der Arbeitswege pro Arbeitstag (Haushalt)
shgr_sm	Durchschnittliche Einkaufszeit (Lebensmittel) (Haushalt)
ntshgrm	Durchschnittliche Anzahl der Einkäufe (Lebensmittel) (Haushalt)
agem	Durchschnittsalter des Haushalts
ntshwcm	Durchschnittliche Einkaufszeit (Frauenbekleidung) (Haushalt)
hhnrjdav	Durchschnittliche Anzahl der Reisen pro Tag (Haushalt)
hhnrtdav	Durchschnittliche Anzahl der Wege pro Reise (Haushalt)
hh_ageav	Durchschnittsalter des Haushalt

Tabelle I.3 : Reisedaten - journey data, deutsche Version

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
j_id	Reise Identifikationsnummer
id	Identifikationsnummer
nr_trips	Anzahl der Wege auf dieser Reise
dayofst	Tag der Befragung
t_l_h	Abfahrtszeit von zu Hause
j_time_s	Dauer der gesamten Reise
nr_act	Anzahl der Aktivitäten während dieser Reise
s	Strom
g	Gruppe
h	Haushalt
i	Individuum
jtravt_s	Gesamtreisezeit im Verkehrsmittel (nur Bewegung)
jactt_s	Gesamtzeit der Aktivitäten der Reise
mj_act_r	Hauptaktivität der Reise (Rang)
mj_exp	Hauptausgabe der Reise
mj_moder	Hauptverkehrsmittel der Reise (Rang)
j_nr	Gesamtausgaben der Reise
j_day	Nummer der Reise an diesem Tag
t_nr_s	Anzahl der Wege auf dieser Reise (gesamt)
exp_j_s	Summe der Ausgaben auf dieser Reise
aplanned	Anteil der geplanten Wege auf dieser Reise
acc_p_av	Anteil der Wege in Begleitung von anderen
a_acc	Anteil der begleiteten Wegen auf dieser Reise
mj_pur	Hauptreisezweck (Uppsala)
j_pur	Reisezweck (Kontiv)
city	Ort der Umfrage (Uppsala)
citycode	Codierung des Ortes der Umfrage (3)
d_o_rp	Tag der Beobachtungszeit der Person
d_o_s	Tag der Umfrage
d_o_w	Wochentag (Ziffer)
d_o_y	fortlaufende Tageszahl (beginnend mit 1.1.1970)
hh_nr	Haushaltsnummer
study	Name der Studie (Uppsala)
studycode	Codierung der Studie (Uppsala =3)
p_nr	Nummer der Person
f_dayo_j	Erster Tag der Reise
l_dayo_j	Letzter Tag der Reise
tr_nrmax	Größte Anzahl von Wegen auf dieser Reise
j_travt	Gesamte Fahrzeit auf dieser Reise
j_actt	Gesamte Aktivitätszeit auf dieser Reise
j_travpt	Maximale Fahrzeit auf dieser Reise
j_tt	korrigierte Gesamtfahrzeit auf dieser Reise
j_arr	Uhrzeit der Ankunft im eigenen Haushalt
j_time_s	Gesamte Reisezeit (Fahrzeit und Aktivitätszeit)
wave	Welle der Befragung (1)

Tabelle I.4 : Wegedaten - Trip data, deutsche Version

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
id	Personenidentifikationsnummer
s	Strom
g	Gruppe
h	Haushalt
i	Individuum
dayofst	Beobachtungstag
t_l_h	Abfahrtszeit von daheim
nr_act	Anzahl der Aktivitäten
trip_nr	Wegenummer
planned	zu Hause geplant (ja/nein)
mode	Verkehrsmittel
acc	begleitet (ja/ nein)
str_code	Geocode des Ortes
str_nr	Straßennummer
str_lett	Straßenbuchstabe?
arr_time	Ankunftszeit
dep_time	Abfahrtszeit
landuse1	Art des Ortes der Aktivität Nummer 1
act1	Aktivität Nummer 1
exp1	Ausgabe bei Aktivität Nummer 1
landuse2	Art des Ortes der Aktivität Nummer 2
act2	Aktivität Nummer 2
exp2	Ausgabe bei Aktivität Nummer 2
landuse3	Art des Ortes der Aktivität Nummer 3
act3	Aktivität Nummer 3
exp3	Ausgabe bei Aktivität Nummer 3
landuse4	Art des Ortes der Aktivität Nummer 4
act4	Aktivität Nummer 4
exp4	Ausgabe bei Aktivität Nummer 4
newplace	neuer Ort (ja/nein)
jourdays	Anzahl der Tage der Reise
xcoorfr	x-Koordinate des Abfahrtsortes
ycoorfr	y-Koordinate des Abfahrtsortes
xcoorto	x-Koordinate des Ankunftsortes
ycoorto	y-Koordinate des Ankunftsortes
travtime	Reisezeit (angegeben)
j_id	Reise-identifikationsnummer
trip_id	Wege-identifikationsnummer
act_time	Dauer der Aktivität
act_r1	Rang der Aktivität 1
act_r2	Rang der Aktivität 2
act_r3	Rang der Aktivität 3
act_r4	Rang der Aktivität 4
mt_act_r	Haupt Aktivität auf diesem Weg
mt_pur	Hauptzweck des Weges
mode_r	Haupt Verkehrsmittel (Rang)
mt_l_use	Haupt Aktivitätsort auf diesem Weg
mt_actex	Haupt Ausgabe falls Haupt Aktivität
exp_t_s	Gesamtausgaben auf dem Weg
al	Hilfsvariable

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
a2	Hilfsvariable
mt_exp	Hauptausgabe (muss nicht Hauptaktivität sein)
mt_expac	Haupt Aktivität falls Hauptausgabe
mt_explu	Haupt Aktivitätsort falls Hauptausgabe
mt_mode	Hauptverkehrsmittel
j_day	Tag der Reise
id_aux	Hilfsvariable Identitätsnummer
j_nr	Nummer der Reise
t_pur1	Zweck 1 des Weges (Uppsala-codierung)
t_pur2	Zweck 2 des Weges (Uppsala-codierung)
t_pur3	Zweck 3 des Weges (Uppsala-codierung)
t_pur4	Zweck 4 des Weges (Uppsala-codierung)
t_pur	Hauptzweck des Weges (Uppsala-codierung)
cm_pur	Hauptzweck des Weges (Citymobil-codierung)
kont_pur	Hauptzweck des Weges (KONTIV-codierung)
pur_lei	Freizeit Zweck (ja/nein)
pur_oth	anderer Zweck als Freizeit Zweck (ja/nein)
city	Ort der Umfrage (Uppsala)
citycode	Codierung des Ortes der Umfrage (3)
d_o_rp	Tag der Beobachtungszeit der Person
d_o_s	Tag der Umfrage
d_o_w	Wochentag (Ziffer)
d_o_y	fortlaufende Tageszahl
hh_nr	Haushaltsnummer
p_nr	Nummer der Person im Haushalt
t_nr	Nummer des Weges
study	Name der Studie (Uppsala)
studycode	Codierung der Studie (Uppsala =3)
t_dest_x	x-Koordinaten des Ankunftsortes
t_dest_y	y-Koordinaten des Ankunftsortes
t_dur	zeitliche Dauer des Weges
t_mm_b_s	Hauptverkehrsmittel nach Typ
t_mm_b_t	Hauptverkehrsmittel nach Zeitdauer (nicht codiert)
t_orig_x	x-Koordinaten des Abfahrtsortes
t_orig_y	y-Koordinaten des Abfahrtsortes
wave	Welle der Befragung (1)
t_dep	Abfahrtszeit des Weges (Zeit8-codierung)
t_arr	Ankunftszeit des Weges (Zeit8-codierung)
t_depnext	Abfahrtszeit des nächsten Weges
t_nr_ges	Anzahl der Wege der Person (fortlaufend)
t_arrnext	Ankunftszeit des nächsten Weges
t_deppre	Abfahrtszeit des vorigen Weges
t_arrpre	Ankunftszeit des vorigen Weges
t_p_pur	Hauptzweck des vorherigen Weges (KONTIV)
t_pdayof	Tag des vorherigen Weges
t_pjdays	Dauer des vorherigen Weges in Tagen
t_n_mpur	Hauptzweck des nächsten Weges (KONTIV)
t_n_cm	Hauptzweck des nächsten Weges (Citymobil)
t_n_pur	Hauptzweck des nächsten Weges (Uppsala)
t_ndayof	Tag des nächsten Weges
t_n_jdays	Dauer des nächsten Weges in Tagen
act_plus	Hilfsvariable Aktivitätszeit
actrtime	Hilfsvariable Aktivitätszeit
act_xxx	Hilfsvariable Aktivitätszeit

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
t_p_mpur	Hauptzweck des vorherigen Weges (Uppsala)
t_p_cm	Hauptzweck des vorherigen Weges (Citymobil)
round_tr	Rundweg (gleicher Abfahrts- und Ankunftsort) (ja/nein)
w_o_rp	Woche der Beobachtungszeit der Person
w_o_s	Woche der Befragung
w_o_y	Woche des Jahres
distx	Distanz des Weges in X-Richtung
disty	Distanz des Weges in Y-Richtung
t_distair	Distanz des Weges, Luftentfernung
t_dist	Distanz des Weges, Bodenentfernung
speed	Geschwindigkeit des Weges
ff_pur	nicht kodiert
dog	nicht kodiert
ldj	nicht kodiert
s_o_p	nicht kodiert
s_o_p_hh	nicht kodiert
s_o_p_o	nicht kodiert
t_at	nicht kodiert
t_ett	nicht kodiert
t_dt	nicht kodiert
t_et	nicht kodiert
t_exp	nicht kodiert
t_ivt	nicht kodiert
t_mc	nicht kodiert
t_pt	nicht kodiert
t_tt	nicht kodiert
t_wait	nicht kodiert
t_wt	nicht kodiert
seqt_nr	nicht kodiert
travplus	Hilfsvariable Aktivitätszeit

Tabelle I.5 : Aktivitätsdaten - activity data, deutsche Version

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
a_st_d	Anfangstag der Aktivität
a_start	Anfangszeit der Aktivität
a_end_d	Schlußtag der Aktivität
a_end	Zeit des Aktivitätsendes
id	Personen-identifikations-nummer
hh_id	Haushalts-identifikations-nummer
seqt_nr	fortlaufende Wege- /Aktivitätsnummer
a_n_tt	Reisezeit zur nächsten Aktivität
a_n_mm	Hauptverkehrsmittel zur nächsten Aktivität
a_n_desx	X-Koordinaten des nächsten Aktivitätsortes
a_n_desy	Y-Koordinaten des nächsten Aktivitätsortes
a_n_dist	Distanz zur nächsten Aktivität
a_n_sp	Geschwindigkeit zur nächsten Aktivität
a_n_pur	Zweck der nächsten Aktivität (Uppsala)
a_n_cm	Zweck der nächsten Aktivität (KONTIV)
a_n_mpur	Zweck der nächsten Aktivität (Citymobil)
a_n_trid	Identifikations-nummer der nächsten Aktivität
planned	Aktivität bzw. Weg zu Hause geplant (ja/nein)
acc	Anzahl der Begleitpersonen aus dem Haushalt
act1	Aktivität 1
act2	Aktivität 2
act3	Aktivität 3
act_time	berichtete Aktivitätszeit
j_day	Nummer der Reise an diesem Tag
j_nr	Nummer der Reise
city	Ort der Umfrage (Uppsala)
citycode	Codierung des Ortes der Umfrage (3)
d_o_rp	Tag der Beobachtungszeit der Person
d_o_s	Tag der Umfrage
d_o_w	Wochentag (Ziffer)
d_o_y	fortlaufende Tageszahl
t_nr	Nummer des Weges
study	Name der Studie (Uppsala)
studycode	Codierung der Studie (Uppsala =3)
wave	Welle der Befragung (1)
w_o_rp	Woche der Beobachtungszeit der Person
w_o_s	Woche der Befragung
w_o_y	Woche des Jahres
ldj	Reise mit großer Entfernung (ja/nein)
a_dur	Dauer der Aktivität
a_p_trid	Identifikations nummer der vorigen Aktivität
a_nr	Nummer der Aktivität
a_p_mm	Hauptverkehrsmittel von der vorigen Aktivität
a_p_tt	Reisezeit von der vorigen Aktivität
a_p_orix	x-Koordinaten des vorigen Abfahrtsortes
a_p_ory	y-Koordinaten des vorigen Abfahrtsortes
a_p_dist	Distanz von der vorigen Aktivität
a_p_sp	Geschwindigkeit von der vorigen Aktivität
a_p_pur	Zweck der vorigen Aktivität (Uppsala)
a_p_cm	Zweck der vorigen Aktivität (KONTIV)
a_p_mpur	Zweck der vorigen Aktivität (Citymobil)

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
a_locx	x-Koordinaten des Aktivitätsortes
a_locy	y-Koordinaten des Aktivitätsortes
a_exp	Ausgaben bei dieser Aktivität
a_pur	Hauptzweck der Aktivität (Uppsala)
a_mpur	Hauptzweck der Aktivität (KONTIV)
a_cm	Hauptzweck der Aktivität (Citymobil)
a_act1	Aktivität 1 an diesem Ort (Uppsala)
a_act2	Aktivität 2 an diesem Ort (Uppsala)
a_act3	Aktivität 3 an diesem Ort (Uppsala)
a_act4	Aktivität 4 an diesem Ort (Uppsala)
at_home	Aktivität zu Hause (ja/nein)
h_a_dur	Dauer der Aktivitäten zu Hause
oh_a_dur	Dauer der Aktivitäten außer Haus
ho_overn	Übernachtung zu Hause (ja/nein)
more24ah	mehr als 24 Stunden zu Hause (ja/nein)
a_pc	Parkkosten bei der Aktivität
a_s_o_p	Anzahl der Personen bei der Aktivität
act_id	Aktivitäts-Identifikations-nummer
sex	Geschlecht
shop_nr	Anzahl der Einkäufe
work_new	Arbeitsklasse (abhängig von den Arbeitsstunden)
hourn_wk	Angegebene Anzahl der Wochenarbeitsstunden
wife_emp	Beschäftigung der Ehefrau (ja/nein)
license	Führerscheinbesitz (ja/nein)
work_sum	Berechnete Gesamtarbeitszeit
n_t_work	Anzahl der Arbeitswege
n_d_work	Anzahl der Arbeitstage
shgr_sum	Summe der Einkaufszeiten (Lebensmittel)
n_t_shgr	Anzahl der Einkäufe (Lebensmittel)
age	Alter
n_t_shwc	Anzahl der Einkäufe (Frauenbekleidungsgeschäft)
i	Individuum im Haushalt nach UTHS
nr_j_sum	Maximale Anzahl der Wege der Person pro Tag
nr_t_sum	Gesamte Anzahl der Wege der Person im Umfragezeitraum
p_j_time	Reisezeit der Person
children	Kinder im Haushalt (ja/nein)
single	alleinlebend (ja/nein)
young	unter / über 50
educ_yr	Anzahl der Ausbildungsjahre
profcode	Codierung des Berufs
worked	Beschäftigung im Beobachtungszeitraum (ja/nein)
educ	Status der Ausbildung im Umfragezeitraum
job	Status des Beschäftigungsgrades
school	Schulbildung im Umfragezeitraum (ja/nein)
age25	Alter unter 25 Jahren (ja/nein)
age35	Alter zwischen 26 und 35 Jahren (ja/nein)
age45	Alter zwischen 36 und 45 Jahren (ja/nein)
age55	Alter zwischen 46 und 55 Jahren (ja/nein)
age65	Alter zwischen 56 und 65 Jahren (ja/nein)
age75	Alter über 66 Jahre (ja/nein)
car	Autobesitz (ja / nein)
hh_nr_1	Anzahl der Führerscheine im Haushalt
hh_nr_d	Anzahl der Personen im Haushalt, die unterwegs waren
hhshopnr	Summe der Einkaufszeiten pro Haushalt

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
hhhoursw	Summe der angegebenen Wochenarbeitszeit pro Haushalt
hhhournw	Summe der berechneten Wochenarbeitszeit pro Haushalt
hhnr_j_s	Anzahl aller Reisen der Haushaltsmitglieder
hhwork_s	Summe der gesamten Arbeitszeit pro Haushalt (in 5 Wochen)
hhntwork	Anzahl der Arbeitswege (wie oft gearbeitet wurde) pro Haushalt
hhndwork	Anzahl der Arbeitstage pro Haushalt
hhshgr_s	Summe der Einkaufszeiten (Lebensmittel) pro Haushalt
hhntshgr	Anzahl der Einkäufe (Lebensmittel) pro Haushalt
hh_age	Gesamalter des Haushalts
hhntshwc	Anzahl der Einkäufe (Frauenbekleidung) pro Haushalt
hhnr_t_s	Anzahl der Wege aller Haushaltsmitglieder
s	Strom nach UTHS
g	Gruppe nach UTHS
h	Haushaltsnummer nach UTHS
home_add	Adresse (code)
hh_nr	Haushaltsnummer
marst	Marst-angabe (UTHS)
hh_memb	Anzahl der Haushaltsmitglieder
hhnr_jma	Maximale Anzahl der Reisen pro Haushalt
hhnr_tma	Maximale Anzahl der Wege pro Haushalt
shop_nrm	Durchschnittliche Einkaufszeit (Haushalt)
hours_wm	Durchschnittliche angegebene Wochenarbeitszeit (Haushalt)
hourn_wm	Durchschnittliche berechnete Wochenarbeitszeit (Haushalt)
nr_j_sm	Durchschnittliche Anzahl der Reisen (Haushalt)
nr_t_sm	Durchschnittliche Anzahl der Wege (Haushalt)
work_sm	Durchschnittliche Arbeitszeit des Haushalt (über 5 Wochen)
n_t_wm	Durchschnittliche Anzahl der Arbeitswege (Haushalt)
n_d_wm	Durchschnittliche Anzahl der Arbeitstage (Haushalt)
n_tdw	Durchschnittliche Anzahl der Arbeitswege pro Arbeitstag (Hh)
shgr_sm	Durchschnittliche Einkaufszeit (Lebensmittel) (Haushalt)
ntshgrm	Durchschnittliche Anzahl der Einkäufe (Lebensmittel) (Haushalt)
agem	Durchschnittsalter des Haushalts
ntshwcm	Durchschnittliche Einkaufszeit (Frauenbekleidung) (Haushalt)
hhnrjdv	Durchschnittliche Anzahl der Reisen pro Tag (Haushalt)
hhnrtdv	Durchschnittliche Anzahl der Wege pro Tag (Haushalt)
hhnrtdjv	Durchschnittliche Anzahl der Wege pro Reise (Haushalt)
hh_ageav	Durchschnittsalter des Haushalts
status	Status der Zensurierung
acc_d	Begleitung (ja/nein)
a_p_mm_d	Unabhängiges Verkehrsmittel (VM) zur Aktivität' (ja/nein)
a_n_mm_d	Unabhängiges VM nach der Aktivität' (ja/nein)
a_p_mmm_d	Unabhängiges motorisiertes VM zur Aktivität' (ja/nein)
a_n_mmm_d	Unabhängiges motorisiertes VM nach der Aktivität' (ja/nein)
act_mp	Uhrzeit des Aktivitätsmittelpunktes
a_dayt	Zeitpunkt des Aktivitätsmittelpunktes (Vormittag/Mittag/Abend)
a_dayt_d	Zeitpunkt des Aktivitätsmittelpunktes (vor/nach 12.30)
head_ohh	Haushaltsvorstand (ja/nein)
a_exp_d	Ausgaben auf der Aktivität (ja/nein)
weekend	Aktivität am oder über ein Wochenende (ja/nein)
car_poss	Möglichkeit einer KFZ-(Mit-)Benutzung
g1_d	Zugehörigkeit zu Gesellschaftsgruppe 1 (ja/nein)
g2_d	Zugehörigkeit zu Gesellschaftsgruppe 2 (ja/nein)
g3_d	Zugehörigkeit zu Gesellschaftsgruppe 3 (ja/nein)
g4_d	Zugehörigkeit zu Gesellschaftsgruppe 4 (ja/nein)

<i>Variablenabkürzung</i>	<i>Variablenname deutsch</i>
g5_d	Zugehörigkeit zu Gesellschaftsgruppe 5 (ja/nein)
g6_d	Zugehörigkeit zu Gesellschaftsgruppe 6 (ja/nein)
sonntag	Aktivität an oder über einen Sonntag (ja/nein)
a_durst	Aktivitätsdauer in Minuten
study	Name der Studie (Uppsala)
city	Ort der Umfrage (Uppsala)
xid	vorige Identifikationsnummer
inter	Intervall zwischen zwei Aktivitäten dieser Kategorie
inter_st	Intervall zwischen zwei Aktivitäten dieser Kategorie in Stunden
inter_t	Intervall zwischen zwei Aktivitäten dieser Kategorie in Tagen
intervall	Intervall zwischen zwei Aktivitäten beliebiger Kategorien
xhh_ida	vorige Haushalts-Identifikationsnummer
xida	vorige Personen-Identifikationsnummer
xa_enda	Zeitpunkt des Endes der vorigen Aktivität
xa_end_da	Tag des Endes der vorigen Aktivität
xa_st_da	Tag des Starts der vorigen Aktivität
xd_o_wa	Wochentag des Starts der vorigen Aktivität
intervall_st	Intervall zw. zwei Aktivitäten beliebiger Kategorien in Stunden
intervall_t	Intervall zwischen zwei Aktivitäten beliebiger Kategorien in Tagen
a_p_durst_a	Aktivitätsdauer der letzten Aktivität beliebiger Kategorie
intervall_st_p	Intervall zw. zwei Aktivitäten beliebiger Kat. vorher in Stunden.
xa_end_d	Tag des Endes der vorigen gleichen Aktivität (korrigiert)
a_p_durst_c	Aktivitätsdauer der letzten Aktivität in dieser Kategorie
inter_st_p	Dauer des vorigen Intervalls zw Aktivitäten dieser Kat. in Stunden
inter_t_p	Dauer des vorigen Intervalls zw Aktivitäten dieser Kat. in Tagen
a_p_mm_c	Hauptverkehrsmittel zur vorigen gleichen Aktivität
mm_c_d	selbes Haupt-VM wie zur vorigen gleichen Aktivität (ja/nein)
xhh_id	vorige Haushalts-Identifikationsnummer (selbe Akt.kategorie)
xa_end	Zeitpunkt des Endes der vorigen gleichen Aktivität (korrigiert)
xd_o_w	Wochentag des Starts der vorigen gleichen Aktivität (korrigiert)
xa_st_d	Tag des Starts der vorigen gleichen Aktivität (korrigiert)

## Anhang II : Aktivitätenkategorisierung

Im folgenden Anhang II werden die in der UTHS Umfrage angegebenen Aktivitäten nach verschiedenen Gesichtspunkten kategorisiert. Die erste Möglichkeit stellt eine allgemeine und sehr oft verwendete Standardkategorisierung mit Hilfe von 10 verschiedenen Aktivitätssparten dar, wie sie beispielsweise auch bei der Kategorisierung nach KONTIV gemacht wird. Die zweite in den Auswertungen verwendete Möglichkeit ist die auch von Axhausen und Schönfelder (2000) verwendete Citymobilkategorisierung, die auf *Götz, Jahn und Schultz, (1997)* zurückgeht und die Aktivitäten in 23 verschiedene Gruppen gliedert. Die dritte Aktivitätenkategorisierung ist jene, die vor allem auf Hanson und Hanson zurückgeht und schon im Rahmen der Uppsala-studie bzw. der Erstellung des UTHS-Datensatzes eingeführt worden ist. Sie unterscheidet die einzelnen Aktivitäten sehr detailliert und teilt sie in insgesamt 56 verschiedene Kategorien ein.

### ***Standardkategorisierung (vgl. KONTIV o.ä.)***

- 1 Jemanden bringen / holen
- 2 Dienstleitungen / Erledigungen
- 3 Arbeitsbezogen
- 4 Schule / Ausbildung
- 5 Arbeit
- 6 Kurzfristiger Einkauf
- 7 Langfristiger Einkauf
- 8 Freizeit
- 9 Sonstiges
- 10 Nach Hause

***City-mobil-Kategorisierung (vgl. Götz, Jahn und Schultz, (1997),  
Schönfelder und Axhausen, (2000))***

- 1 Arbeit
- 2 Beruflicher Zusammenhang
- 3 Ausbildungsplatz
- 4 Fort-/Weiterbildung
- 5 Versorgung von Kindern oder anderen Familienangehörigen
- 6 Einkauf des täglichen Bedarfs
- 7 Einkauf großer Vorräte und besondere Anschaffungen
- 8 Einkaufsbummel
- 9 Erledigung von Pflichten (Bank, Post, ...)
- 10 Besuch bei Verwandten
- 11 Treffen mit Freunden
- 12 Verein, Bürgerinitiative, Kirche, etc.
- 13 Arzt, Zahnarzt, Friseur
- 14 Autopflege, Tanken, etc.
- 15 Ausübung von Sport
- 16 Größerer Ausflug in die Natur
- 17 Spaziergang, Spazierfahrt
- 18 Kurzurlaub, Wochenendausflug
- 19 Garten, Gartenhaus
- 20 Etwas Neues sehen (Stadt, Sightseeing, etc.)
- 21 Disko, Lokal, Kneipe, Restaurant
- 22 Nach Hause
- 23 Sonstiges

## ***Uppsala-Kategorisierung (vgl. Hanson und Hanson, (1971))***

- 1 Fehlende Information
- 2 Jemanden bringen / holen
- 3 Verkehrsmittelwechsel
- 4 Benützung öffentlicher Anlagen
- 5 Fehlgeschlagener Versuch, eine Tätigkeit auszuführen
- 10 Kommunikation generell, keine bestimmte Kategorie
- 11 Kommunikation mit einem Familienmitglied
- 12 Kommunikation mit Freunden oder Nachbarn
- 13 Kommunikation mit einem Geschäftsmann oder einer Behörde
- 14 Kommunikation mit einem Verkäufer / einer Verkäuferin in einem Geschäft
- 15 Kommunikation mit Geschäftspartnern oder Kollegen
- 16 Telefonanruf tätigen
- 17 Telegramm / Fax senden
- 18 Brief oder Paket aufgeben / verschicken
- 20 Einkaufen generell, keine andere Kategorie
- 21 Waren erstehen oder andere monetäre Services in Anspruch nehmen
- 22 Schaufensterbummel
- 23 Waren / Kleidung anprobieren
- 24 Waren tauschen
- 25 Warenbestellung ohne Geldausgabe
- 26 Warenbestellung mit Geldausgabe
- 27 Warenabholung ohne Geldausgabe
- 28 Waren heimtransportieren
- 29 Ausgeliehene Ware retournieren
- 30 Persönliche Erledigung (personal business), keine andere Kategorie
- 31 Bankerledigung (Geldabhebung oder Geldgeschäft)
- 32 Andere Bankerledigungen
- 33 Rechnung einzahlen
- 34 Treffen ausmachen / Termin arrangieren
- 35 Konsultation von Arzt, Rechtsanwalt, etc.
- 36 Mittagessen alleine
- 37 Kirchenbesuch
- 38 Spaziergang mit dem Hund

- 39 Wäsche
- 40 Soziale Belange, keine andere Kategorie
- 41 Verwandtenbesuch
- 42 Besuch bei Freunden
- 43 Vereinstreffen / Klubtreffen
- 44 Mittagessen mit anderen
- 45 Bar- /Lokalbesuch mit anderen
- 46 Begleitung eines anderen Einkaufenden ohne selbst Geld auszugeben
- 50 Arbeit
- 51 Arbeitsbezogene Aktivität
- 52 Anstellung suchen / Arbeitsuche
- 53 Unentgeltliche Arbeit (beispielsweise Hilfe bei Hausarbeit, Autoreparatur bei Freunden)
- 60 Freizeit / Erholung generell, keine andere Kategorie
- 61 Konzert, Kino, Theater, etc.
- 62 Besuch sportlicher Wettkämpfe in der Halle als Zuschauer (passiv)
- 63 Sportausübung in der Halle aktiv
- 64 Besuch sportlicher Wettkämpfe im Freien als Zuschauer (passiv)
- 65 Sportausübung im Freien aktiv
- 66 Spaziergang
- 67 Spazierfahrt
- 68 Auf Kinder aufpassen im Freien (beispielsweise am Spielplatz)
- 69 Sitzen im Park etc.
- 70 Belegung eines Ausbildungskurses

# Anhang III : Partial Likelihood

## *Partial Likelihood - Mathematische und Computermäßige Details*

Im folgenden Anhang werden die in der Arbeit nicht ausführlich betrachteten mathematischen Grundlagen des für die Hazardmodelle sehr wichtigen Ansatzes der *Partial Likelihood* näher betrachtet und erklärt. Zusätzlich werden auch einige die Auswertung mit dem Computer bzw. den SAS-Programmen betreffenden Details erläutert.

Die normale Likelihood Funktion wird als Produkt der Wahrscheinlichkeiten (likelihoods) aller Individuen (im Datensatz) angeschrieben. Die *partial likelihood* wird hingegen nur als Produkt der Likelihoods aller beobachteten Ereignisse J angegeben:

$$PL = \prod_{j=1}^J L_j \quad , \text{ wobei } L_j \text{ die Likelihood des } j\text{-ten Ereignisses ist.}$$

In einem Beispiel von Brustkrebspatientinnen kommt dieser Unterschied recht gut zum Ausdruck. Von insgesamt 45 beobachteten Frauen sterben schon 26 von im Verlauf der Beobachtungszeit, demnach gehen 26 Fälle in die *partial likelihood* ein. Die Wahrscheinlichkeit, daß ein Patient genau zu einem Zeitpunkt stirbt und nicht ein anderer, ist sein Hazard dividiert durch die Summe der Hazards all jener Personen, die zu diesem Zeitpunkt noch nicht gestorben sind, sich also noch im *Riskset* befinden, also:

$$L_1 = \frac{h_1(5)}{h_1(5) + h_2(5) + \dots + h_j(5)} = \frac{\lambda_0(5)e^{\beta x_1}}{\lambda_0(5)e^{\beta x_1} + \lambda_0(5)e^{\beta x_2} + \dots + \lambda_0(5)e^{\beta x_j}}$$

Hier kürzen sich alle  $\lambda_0(5)$  heraus, so wie auch bei allen anderen  $L_j$ . Dadurch ist das Produkt aller  $L_j$  von  $j=1$  bis  $j=J$ , die *partial Likelihood*, natürlich auch unabhängig von allen  $\lambda$  bzw. muß man keine grundlegende Hazard Funktion bestimmen, um trotzdem Koeffizienten  $\beta$  abschätzen zu können. Diese einfache Vorgangsweise mit wenigen Voraussetzungen ist auch, wie im Kapitel 6.3.2 beschrieben wird, ein großer Vorteil der *partial Likelihood*.

Hier sieht man auch, daß die partial Likelihood wirklich nur von der Abfolge der Ereignisse, nicht aber von ihrem Wert abhängt. Ereignis 1 kann beispielsweise von Monat 0 bis 8 (exklusive) passieren, das Ergebnis würde sich nicht ändern.

Die allgemeine Form der partial Likelihood für Daten mit fixen Kovariablen eines PHM sieht dann folgendermaßen aus:

$$PL(\beta) = \prod_{i=1}^n \left[ \frac{e^{\beta x_i}}{\sum_{j=1}^n y_{ij} e^{\beta x_j}} \right]^{\delta_i}$$

wobei  $y_{ij}$  die Werte 0 oder 1 annehmen kann, je nachdem ob die beobachtete Person  $j$  das Ereignis schon erfahren hat bzw. das Ereignis allgemein eingetreten ist.

$PL(\beta)$  = partial likelihood von  $\beta$

$\beta$  = Vektor der Schätzparameter

$n$  = Anzahl aller Dauern

$X$  = Vektor der Kovariablen

Nach der Logarithmierung auf beiden Seiten kann man wieder, wie bei der maximum Likelihood Abschätzung, mit Hilfe des Newton-Raphson-Algorithmus diejenigen Koeffizienten  $\beta$  schätzen, die die Funktion maximieren.

Bei binomialen (dichotomischen) Variablen ist der äquivalente Test zum Log-Rank-Test von PROC LIFEREG der Partial Likelihood Score-Test.

## Anhang IV : Liste der möglichen Einflußvariablen

Variablenbezeichnung	Variablenname	dummy	Einsatzbereich
SEX	Geschlecht	X	Aktivitäten, Intervalle
AGE_**	Altersgruppe	X	Aktivitäten, Intervalle
AGE	Alter		Aktivitäten, Intervalle
YOUNG	Unter 50 Jahren j/n	X	Aktivitäten, Intervalle
LICENSE	Führerscheinbesitz j/n	X	Aktivitäten, Intervalle
HEAD_O_HH	Haushaltsvorstand j/n	X	Aktivitäten, Intervalle
WORKED	arbeitend / nicht arbeitend	X	Aktivitäten, Intervalle
CHILDREN	Kinder im Haushalt j/n	X	Aktivitäten, Intervalle
SINGLE	Single bzw. alleine lebend j/n	X	Aktivitäten, Intervalle
CAR_POSS	Autoverfügbarkeit j/n	X	Aktivitäten, Intervalle
EDUC	Ausbildung während der Umfrage j/n	X	Aktivitäten, Intervalle
G*_D	Gesellschaftsgruppe Nr. * j/n	X	Aktivitäten, Intervalle
EDUC_YR	Anzahl der Ausbildungsjahre		Aktivitäten, Intervalle
G	Nummer der Gesellschaftsgruppe		Aktivitäten, Intervalle
CAR	Anzahl der Autos im Haushalt		Aktivitäten, Intervalle
HH_NR_L	Anzahl der Führerscheine im Haushalt		Aktivitäten, Intervalle
HH_MEMB	Anzahl der Haushaltsmitglieder		Aktivitäten, Intervalle
ACC_D	Begleitung j/n	X	Aktivitäten, Intervalle
PLANNED	Aktivität geplant j/n	X	Aktivitäten, Intervalle
A_EXP_D	Ausgaben am Aktivitätsort j/n	X	Aktivitäten, Intervalle
WEEKEND	Wochenende j/n	X	Aktivitäten, Intervalle
A_DAYT_D	Tageszeit (vor 13 Uhr j/n)	X	Aktivitäten, Intervalle
A_P_MM_D	Unabhängiges Verkehrsmittel (VM) zur Aktivität	X	Aktivitäten, Intervalle
A_N_MM_D	Unabhängiges VM nach der Aktivität	X	Aktivitäten, Intervalle
A_P_MMMD	Unabhängiges motorisiertes VM zur Aktivität	X	Aktivitäten, Intervalle
A_N_MMMD	Unabhängiges motorisiertes VM von der Aktivität	X	Aktivitäten, Intervalle
ACC	Anzahl der Begleitpersonen		Aktivitäten, Intervalle
A_EXP	Höhe der Ausgaben am Aktivitätsort		Aktivitäten, Intervalle
A_DAYT	Tageszeit (vormittag/nachmittag/abends)		Aktivitäten, Intervalle
A_P_DIST	Distanz zum Aktivitätsort		Aktivitäten, Intervalle
A_P_TT	Reisezeit zum Aktivitätsort		Aktivitäten, Intervalle
A_N_DIST	Distanz zum nächsten Aktivitätsort		Aktivitäten, Intervalle
A_N_TT	Reisezeit zum nächsten Aktivitätsort		Aktivitäten, Intervalle
A_P_MM	Verkehrsmittel zum Aktivitätsort		Aktivitäten, Intervalle
A_N_MM	Verkehrsmittel vom Aktivitätsort		Aktivitäten, Intervalle
NR_T_SUM	Anzahl der Wege gesamt		Aktivitäten, Intervalle
NR_J_SUM	Anzahl der Reisen gesamt		Aktivitäten, Intervalle
A_P_PUR	Vorangegangene Aktivitätskategorie		Aktivitäten, Intervalle
A_N_PUR	Nachfolgende Aktivitätskategorie		Aktivitäten, Intervalle
NR_A_SUMC	Anzahl der Aktivitäten in dieser Kategorie		Aktivitäten
INTER_T_P	Dauer des Intervalls vorher in Tagen		Intervalle
INTERVALL_ST	Intervalldauer zur vorigen beliebigen Aktivität		Intervalle
A_P_DURST_A	Dauer der beliebigen Aktivität vorher		Intervalle
A_P_DURST_C	Dauer der gleichen Aktivität vor dem Intervall		Intervalle
A_P_MM_C	Verkehrsmittel zur Aktivität vor Intervall		Intervalle
MM_C_D	Unabhängiges VM zur Aktivität vor Intervall	X	Intervalle
INTERVALL_ST_P	Intervalldauer zu beliebiger A vor Intervall		Intervalle

Die Variablen sind alle jeweils im Datensatz upps.act\_dur oder upps.act\_int enthalten. und im Anhang I detailliert beschrieben.

## Anhang V : Hazard- und Survivalkurven, Aktivitätsdauern

In den Abbildungen V.1 bis V.14 werden für alle Aktivitätskategorien die geglätteten Hazardkurven dargestellt. Im Kapitel 9.2.2.2 sind die Hazard- und Survivalkurven in jeweils einer einzigen Grafik zu sehen. In den Abbildungen V.15 bis V.36 sind die geglätteten und ungeglätteten Hazard- und Survivalkurven der Altersunterschiede und Geschlechtsunterschiede für die vier ausgewählten Kategorien Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden abgebildet, die teilweise im Kapitel 9.2.2.3 zu sehen sind.

Abbildung V.1 : Geglättete Hazardkurve aller Aktivitäten

---

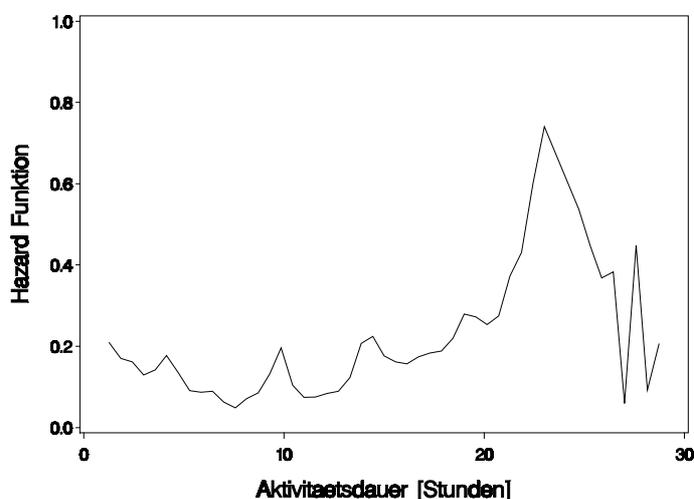


Abbildung V.2 : Geglättete Hazardkurve der Kategorie Arbeit

---

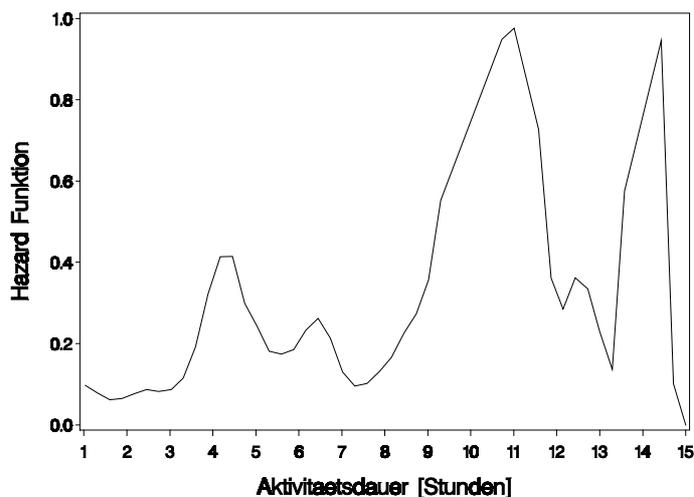


Abbildung V.3a : Geglättete Hazardkurve, Einkauf langfristig, Aktivitätsdauer bis 2 Stunden

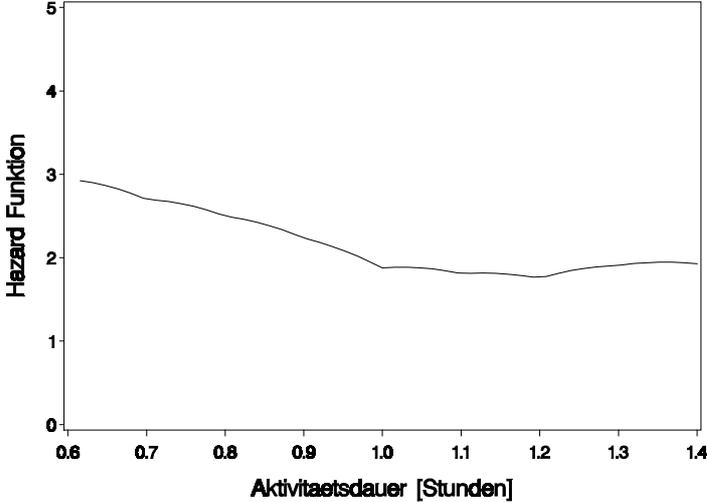


Abbildung V.3b : Geglättete Hazardkurve, Einkauf langfristig, ursprünglicher Datensatz

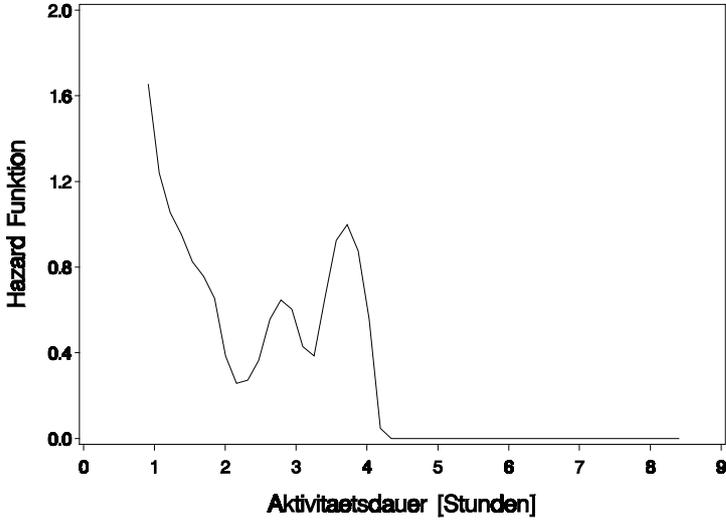


Abbildung V.4a : Geglättete Hazardkurve, Einkauf kurzfristig, Aktivitätsdauer bis 5 Stunden

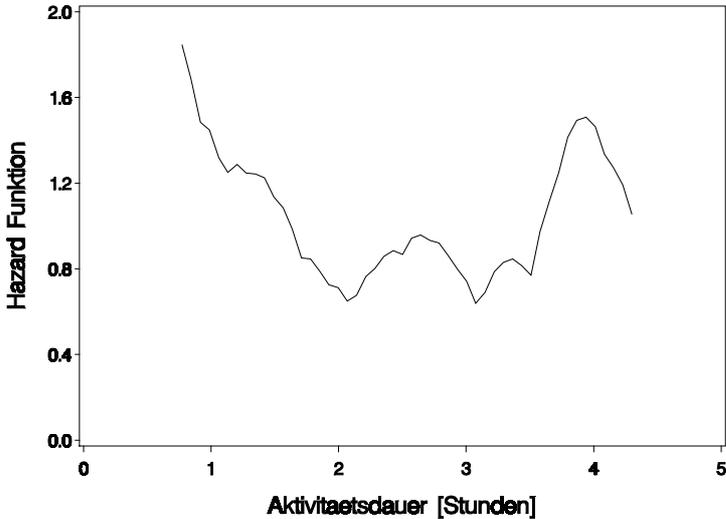


Abbildung V.4b : Geglättete Hazardkurve, Einkauf kurzfristig, ursprünglicher Datensatz

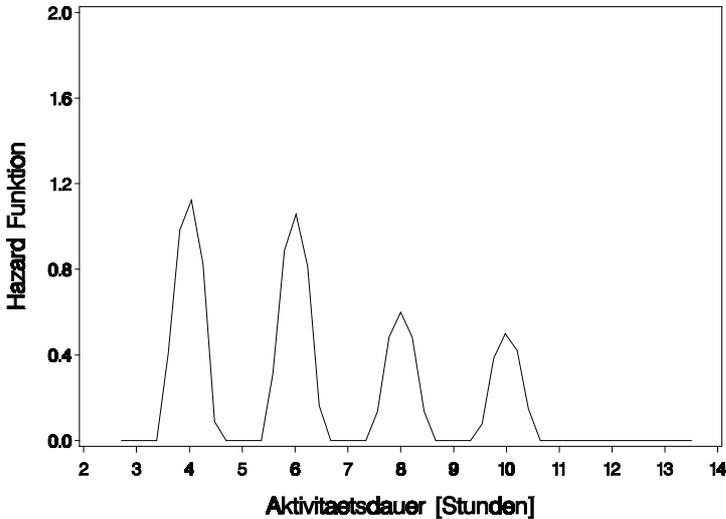


Abbildung V.5 : Geglättete Hazardkurve, Freizeit allgemein

---

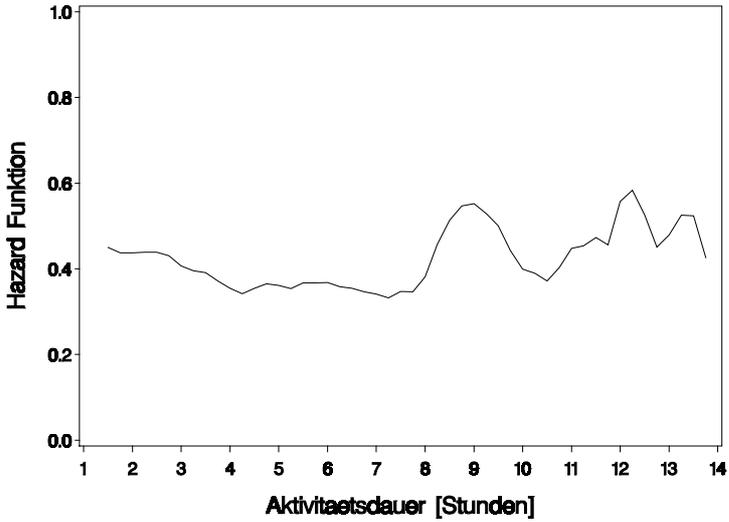


Abbildung V.6 : Geglättete Hazardkurve, Dienstleistung allgemein

---

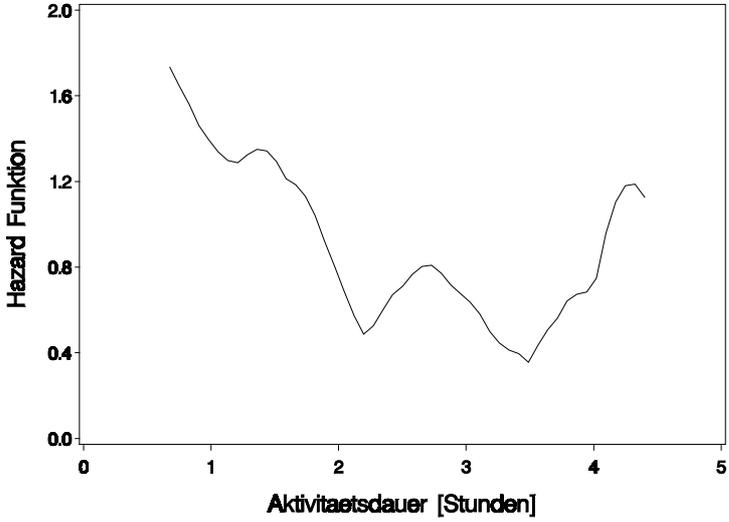


Abbildung V.7 : Geglättete Hazardkurve, Erledigungen

---

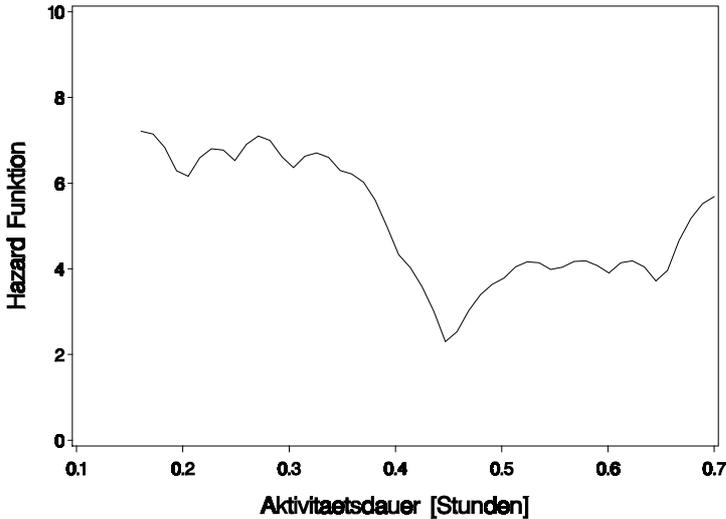


Abbildung V.8 : Geglättete Hazardkurve, Besuch bei Verwandten

---

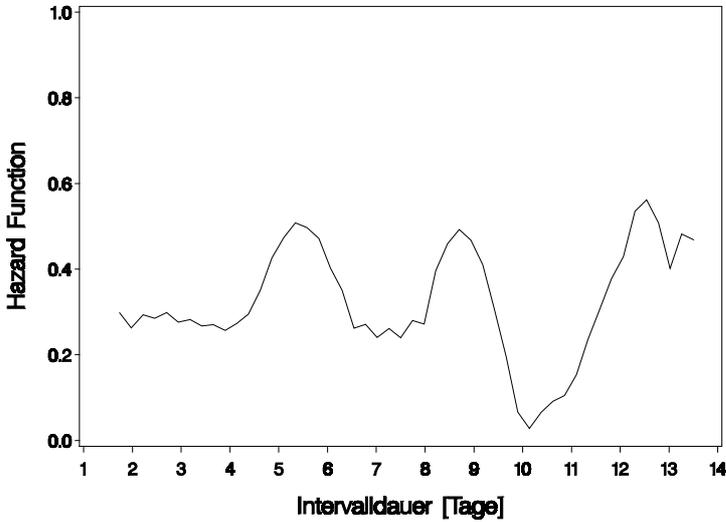


Abbildung V.9 : Geglättete Hazardkurve, Treffen mit Freunden

---

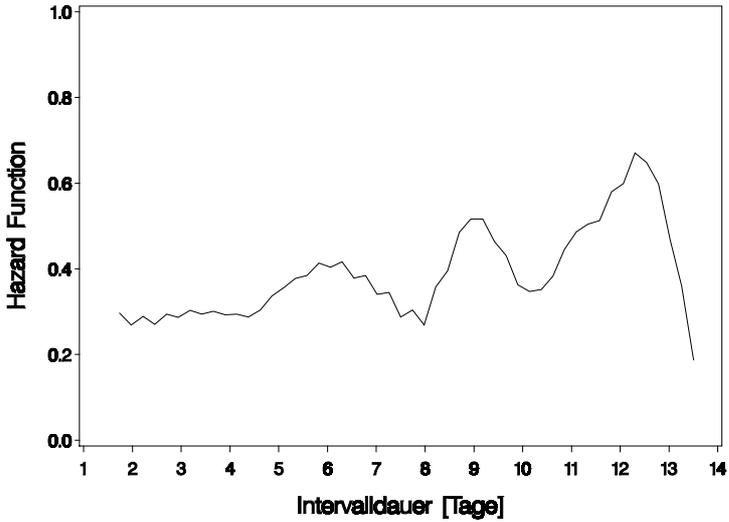


Abbildung V.10 : Geglättete Hazardkurve, Verein, Initiative, Kirche

---

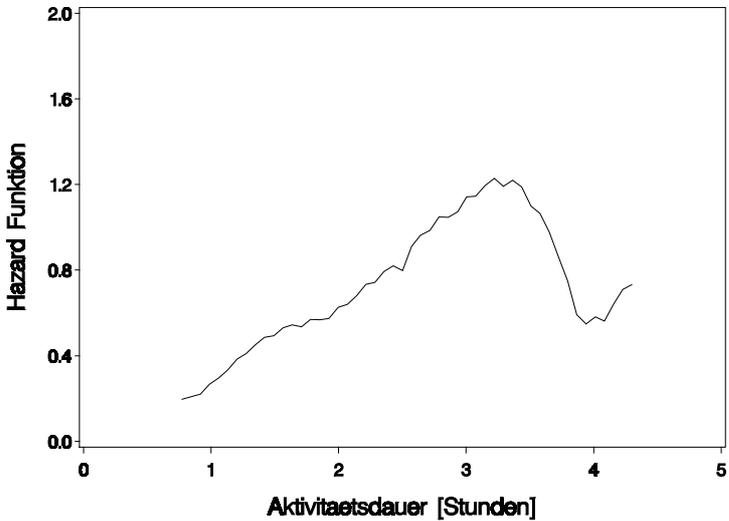


Abbildung V.11 : Geglättete Hazardkurve, Sport (aktiv)

---

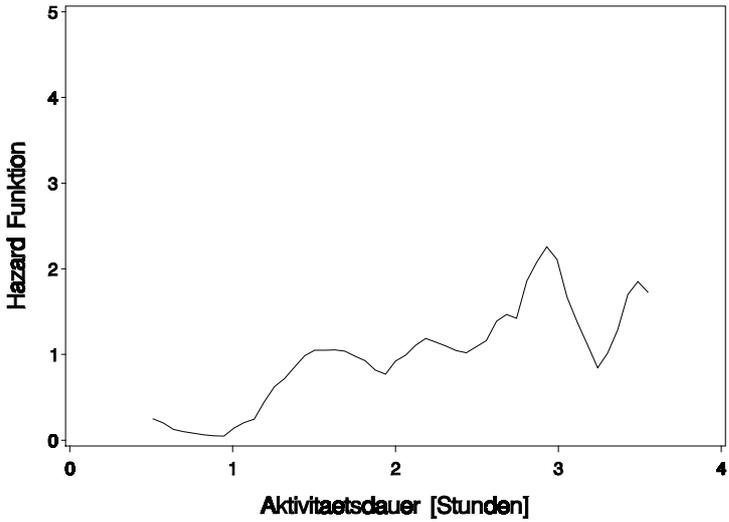


Abbildung V.12 : Geglättete Hazardkurve, Spaziergang / -fahrt

---

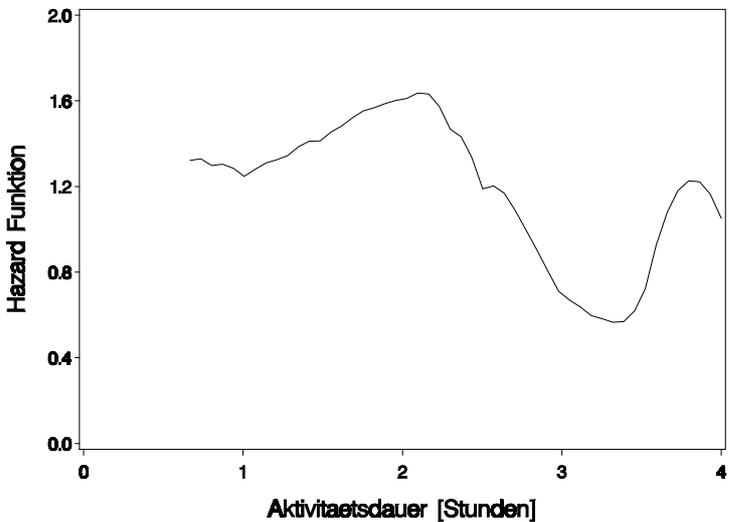


Abbildung V.13 : Geglättete Hazardkurve, Etwas Neues sehen/Sightseeing

---

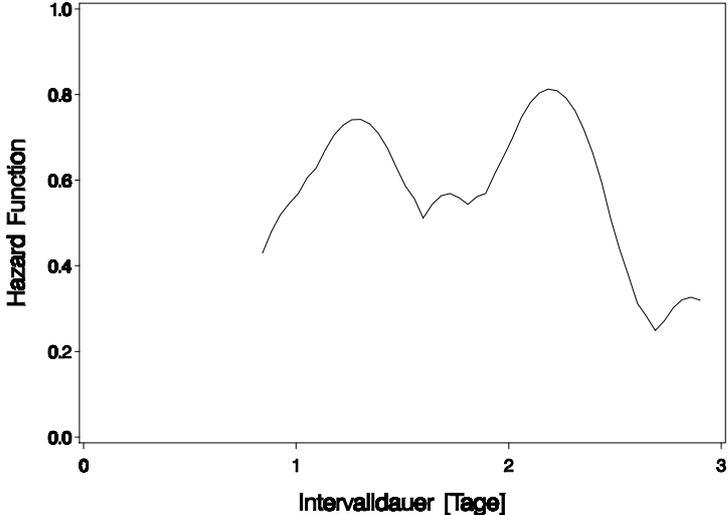


Abbildung V.14 : Geglättete Hazardkurve, Lokal, Kino, Theater

---

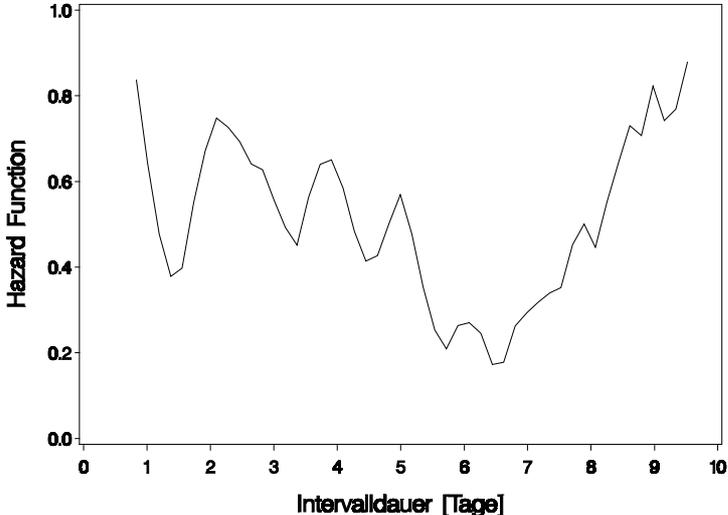


Abbildung V.15 : Survivalkurve Kategorie Einkauf kurzfristig, Altersschichtung

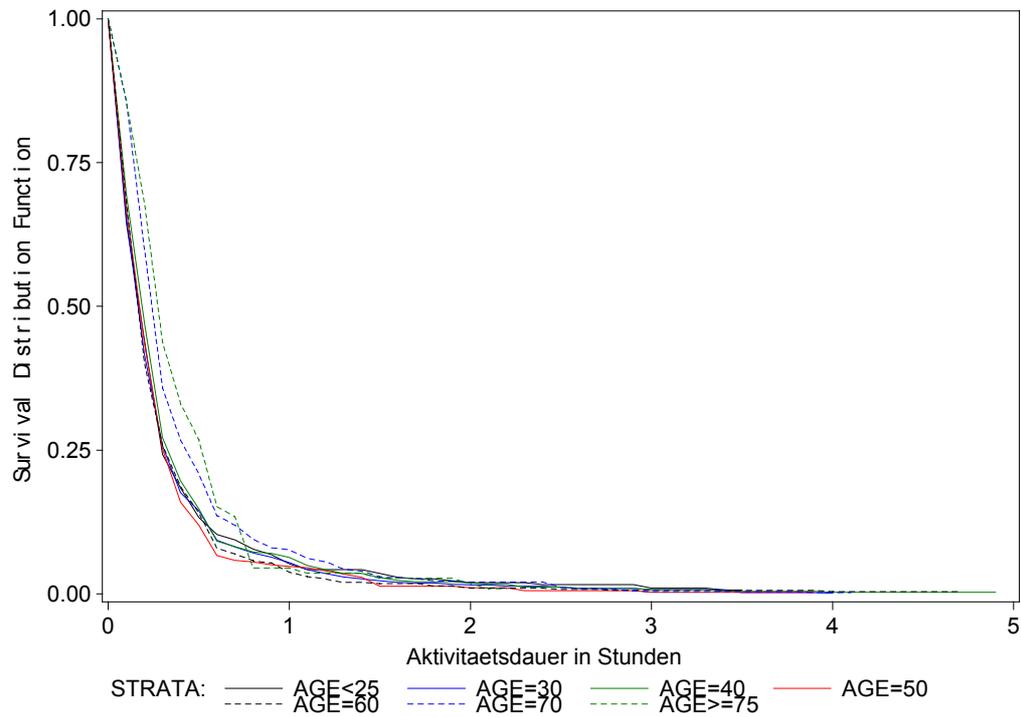


Abbildung V.16 : Hazardkurve Kategorie Einkauf kurzfristig, Altersschichtung

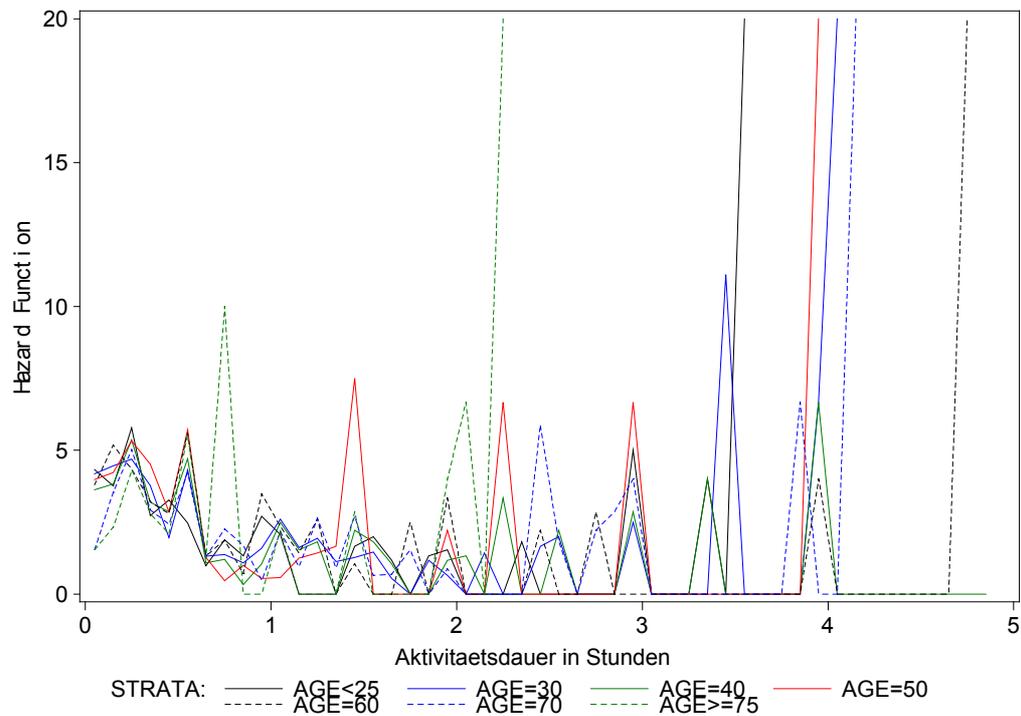


Abbildung V.17 : Survivalkurve Kategorie Einkauf kurzfristig, Schichtung der Geschlechter

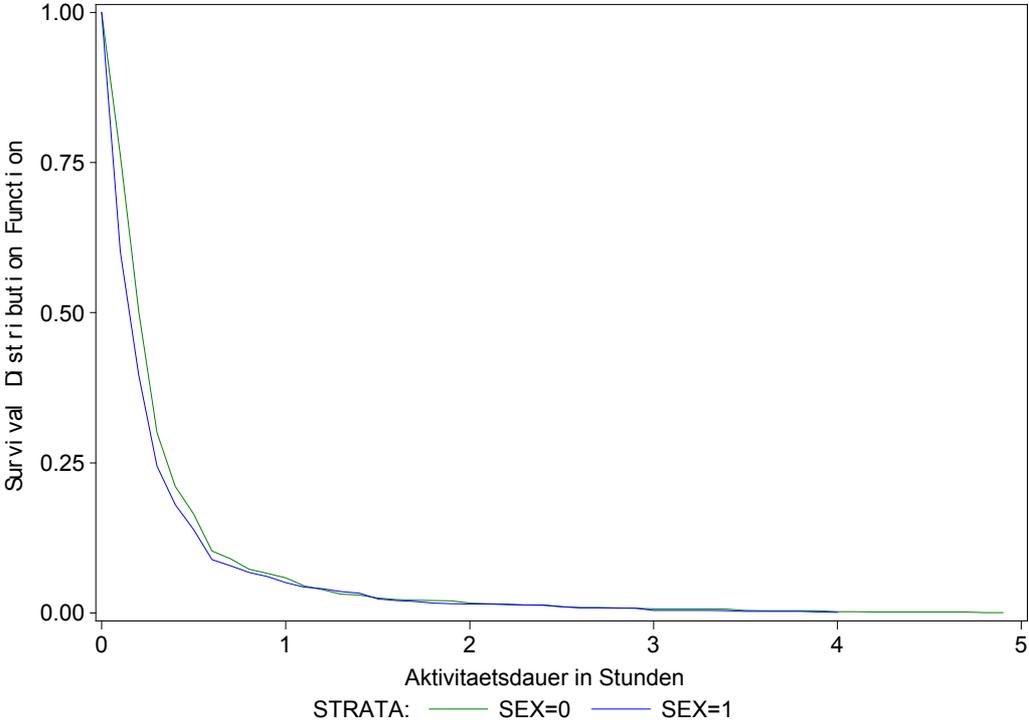


Abbildung V.18 : Hazardkurve Kategorie Einkauf kurzfristig, Schichtung der Geschlechter

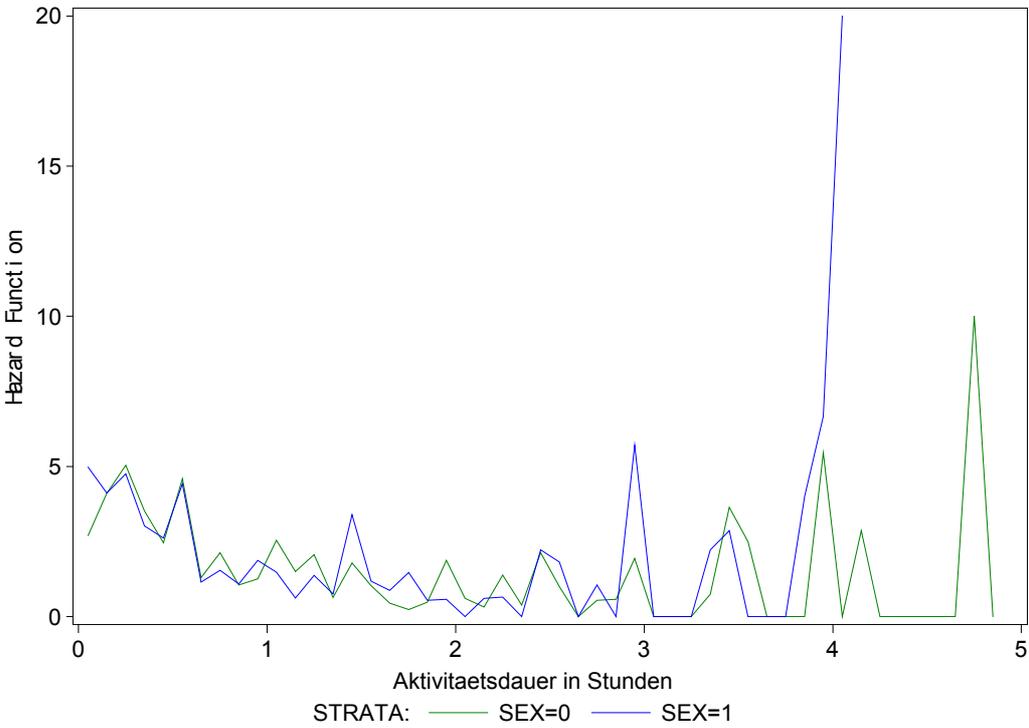


Abbildung V.19 : Survivalkurve Kategorie Freizeit allgemein, Altersschichtung

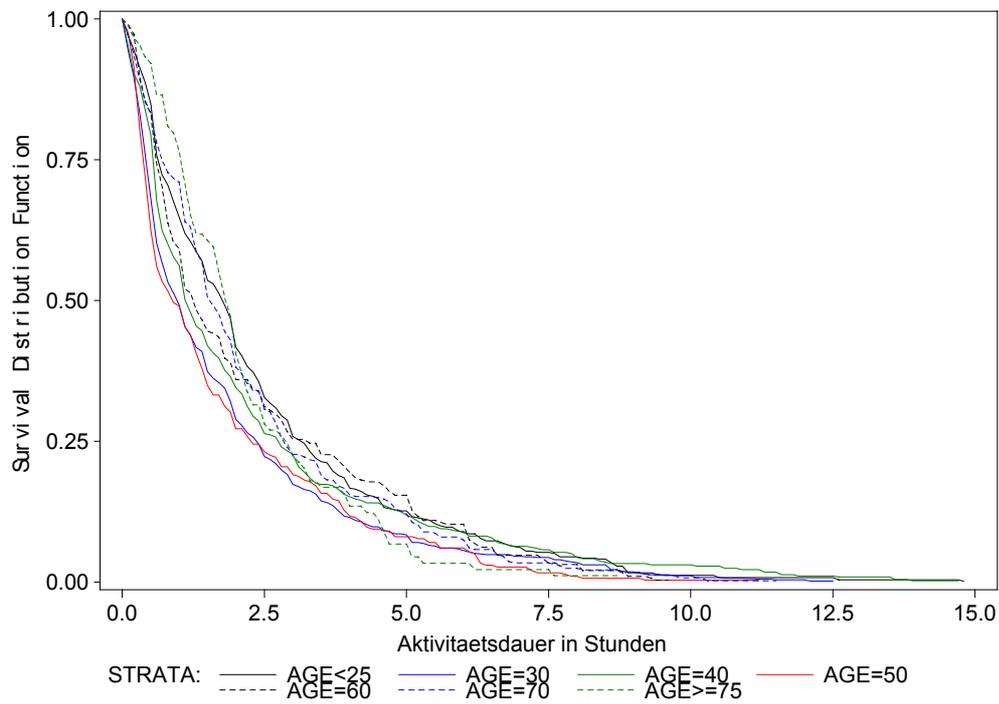


Abbildung V.20 : Hazardkurve Kategorie Freizeit allgemein, Altersschichtung

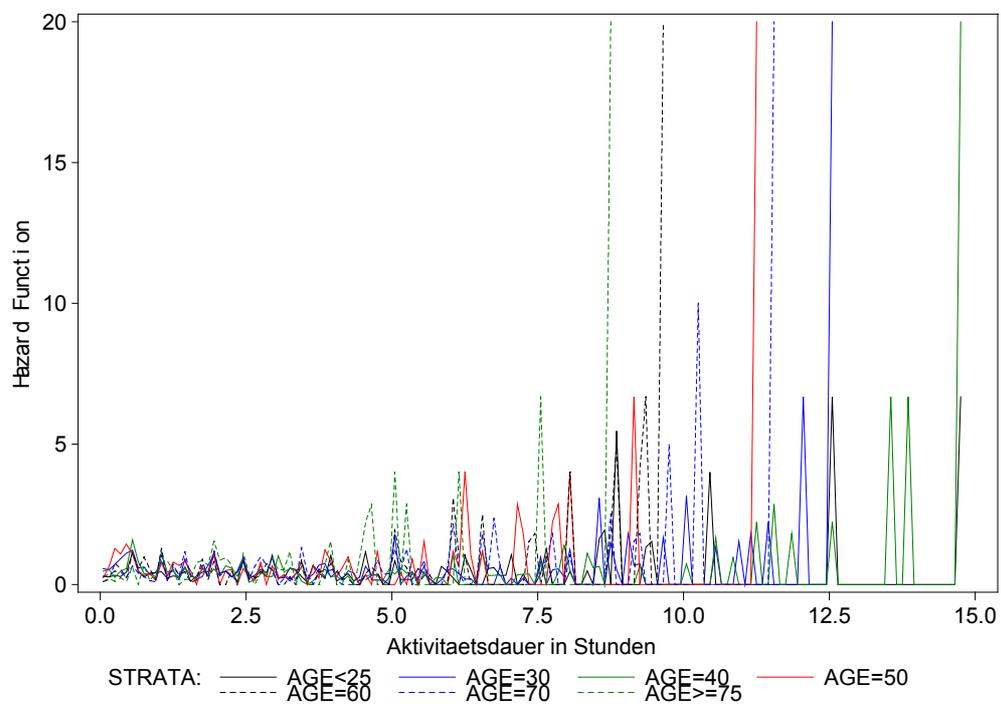


Abbildung V.21 : Survivalkurve Kategorie Freizeit allgemein, Schichtung der Geschlechter

---

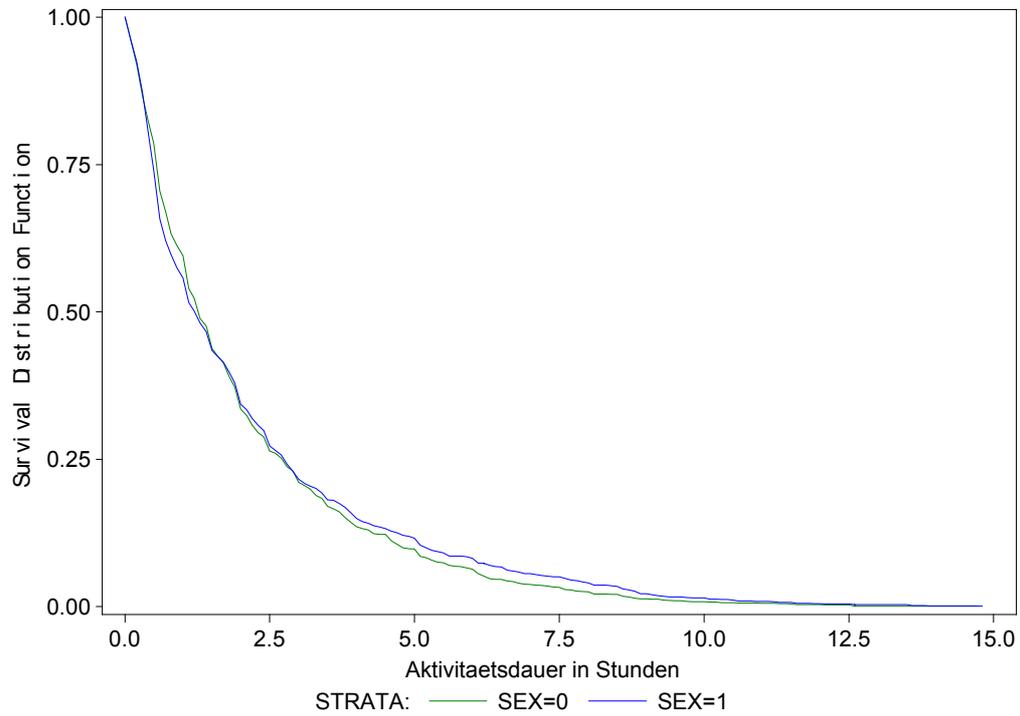


Abbildung V.22 : Hazardkurve Kategorie Freizeit allgemein, Schichtung der Geschlechter

---

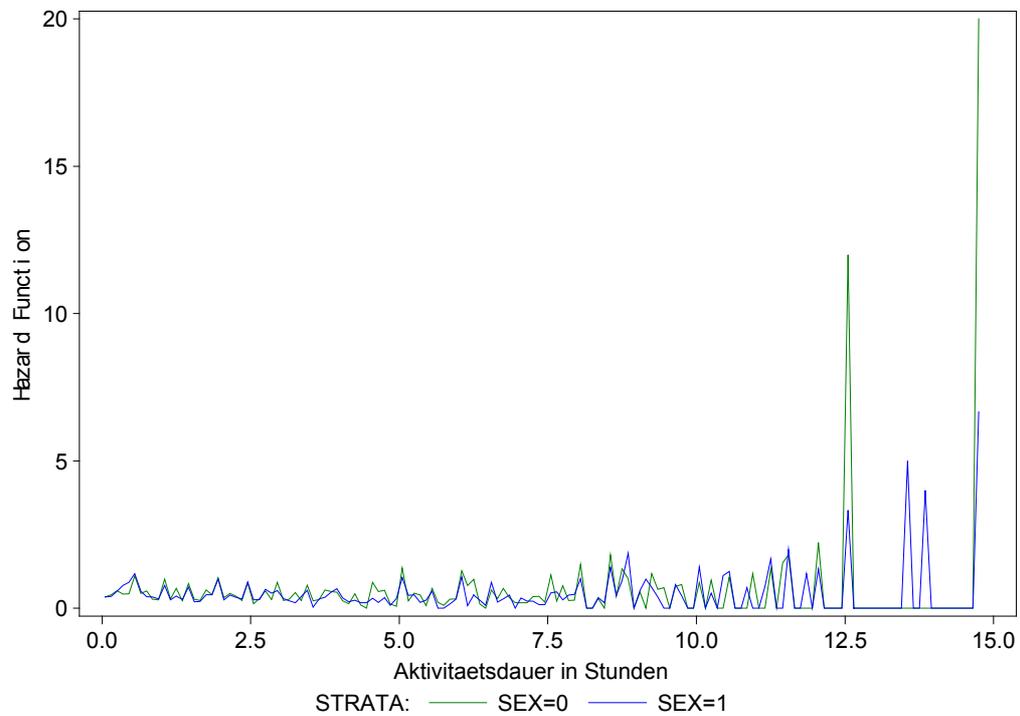


Abbildung V.23 : Survivalkurve Kategorie Dienstleistung allgemein, Altersschichtung

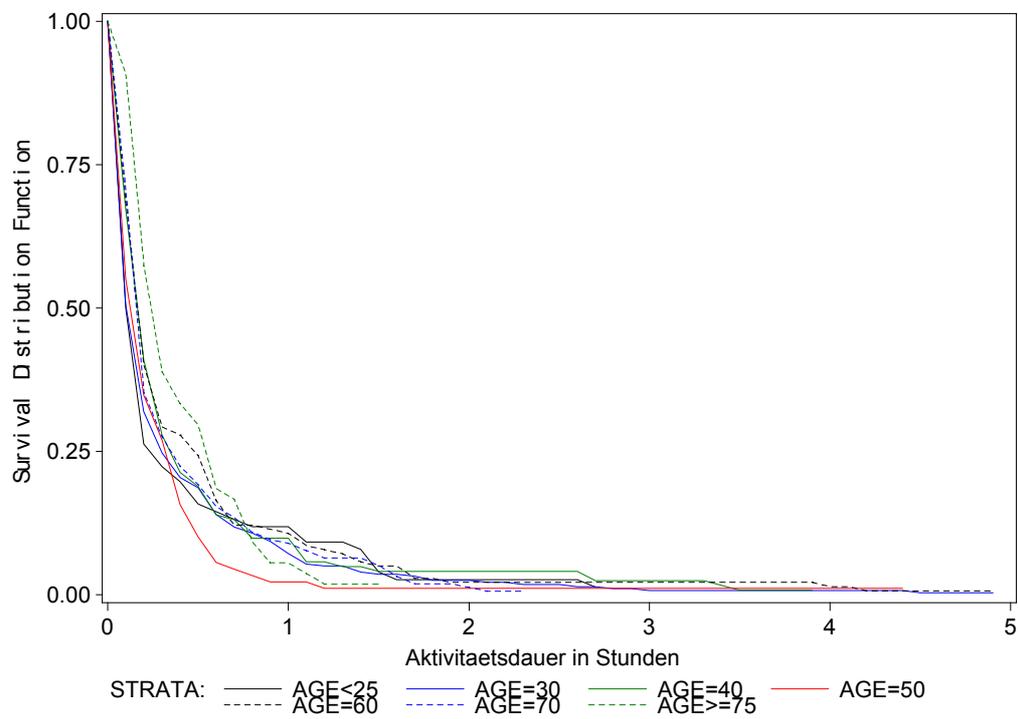


Abbildung V.24 : Hazardkurve Kategorie Dienstleistung allgemein, Altersschichtung

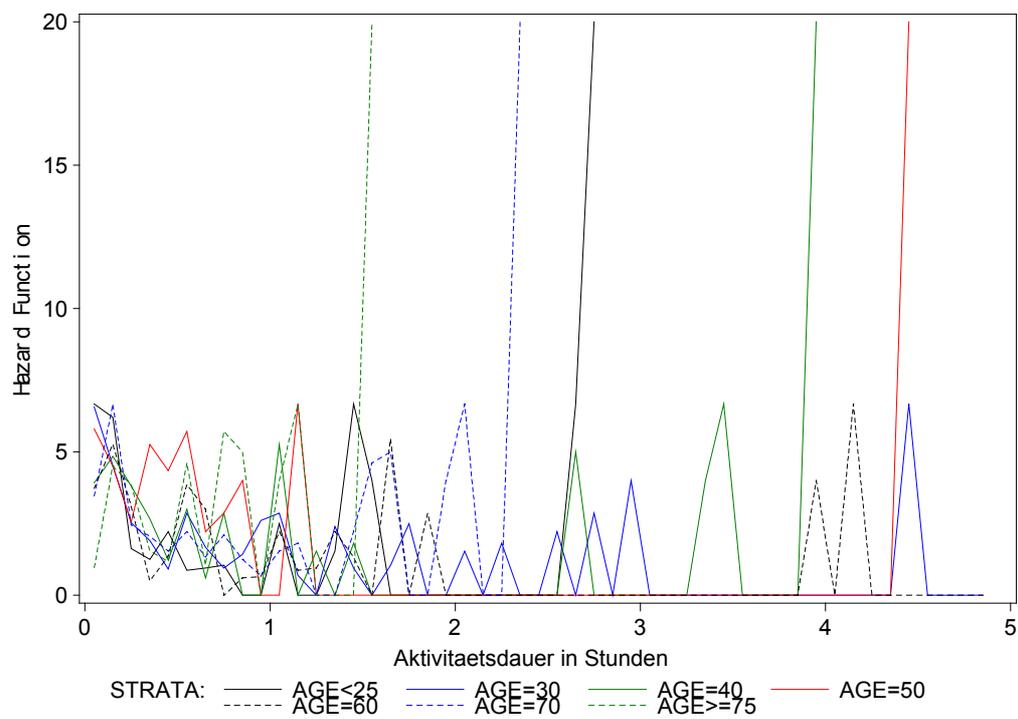


Abbildung V.25 : Survivalkurve Kategorie Dienstleistung allgemein, Schichtung der Geschlechter

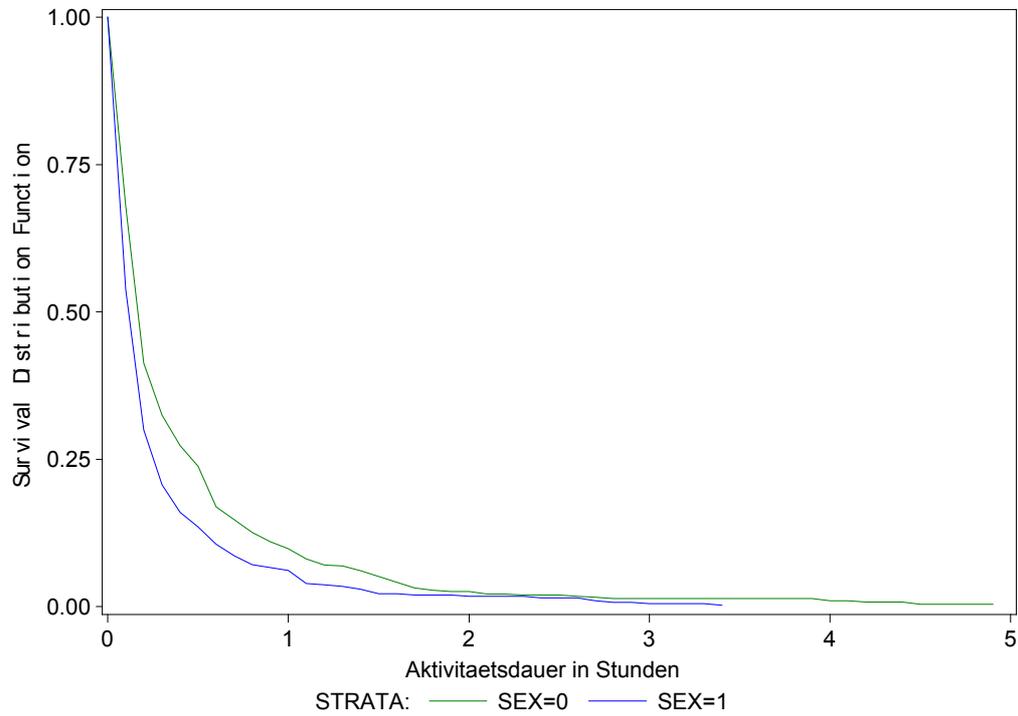


Abbildung V.26 : Hazardkurve Kategorie Dienstleistung allgemein, Schichtung der Geschlechter

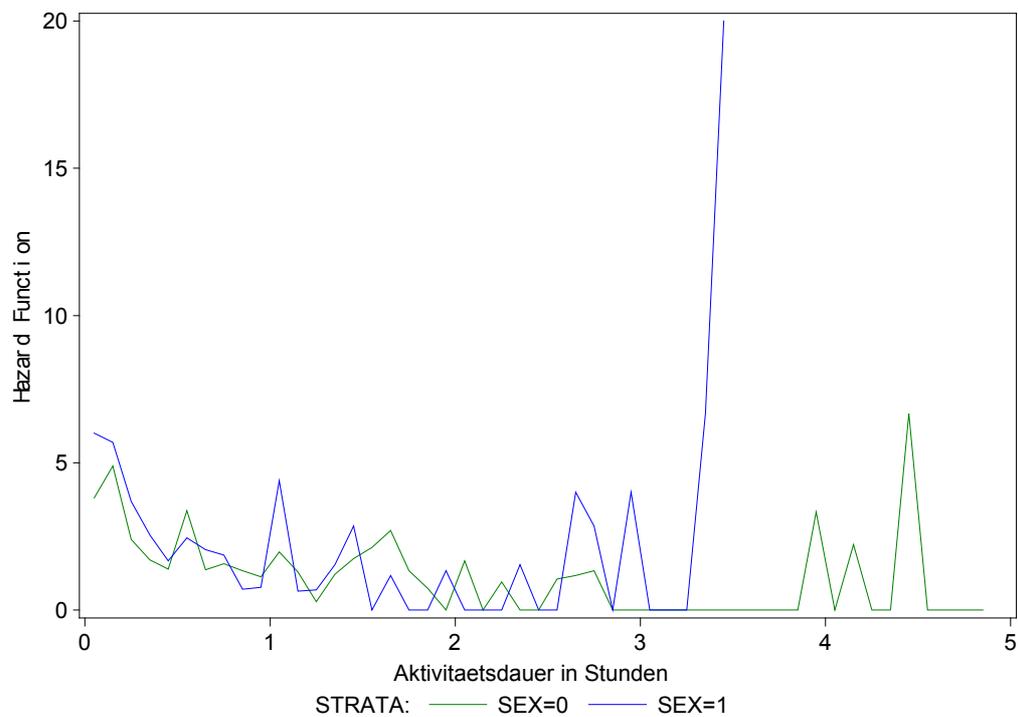


Abbildung V.27 : Survivalkurve Kategorie Treffen mit Freunden, Altersschichtung

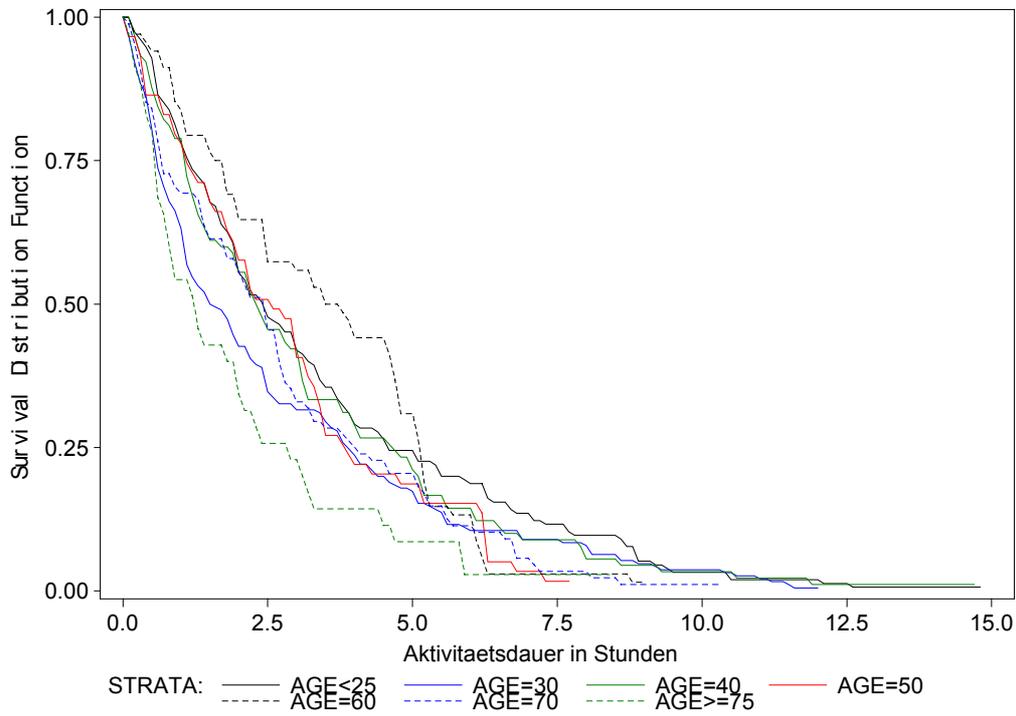


Abbildung V.28 : Hazardkurve Kategorie Treffen mit Freunden, Altersschichtung

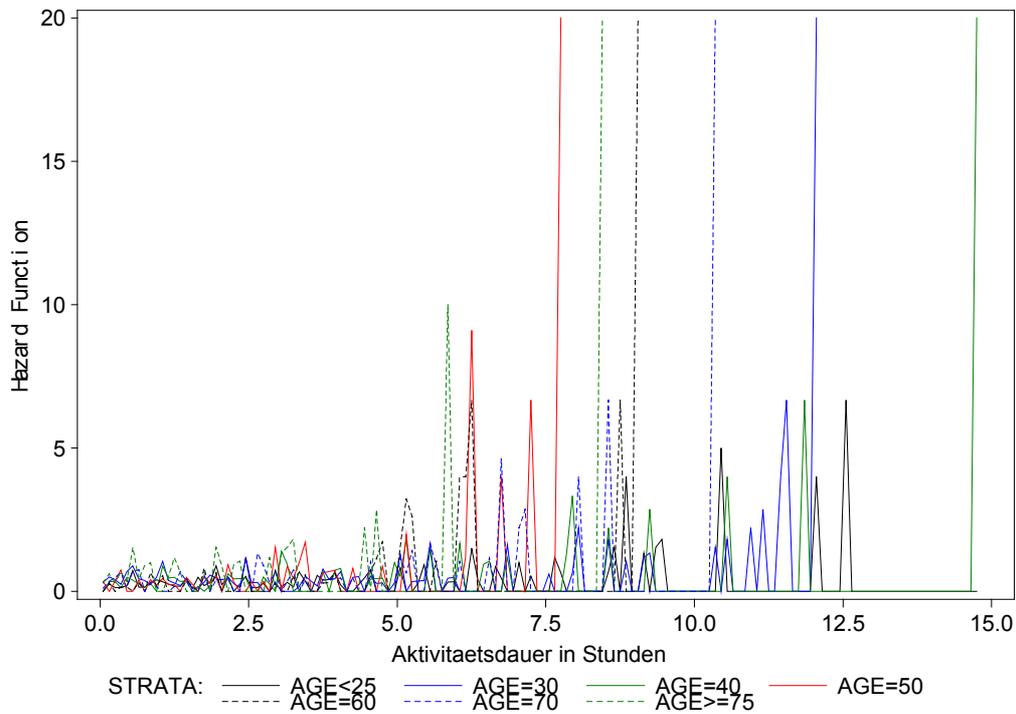


Abbildung V.29 : Survivalkurve Kategorie Treffen mit Freunden, Schichtung der Geschlechter

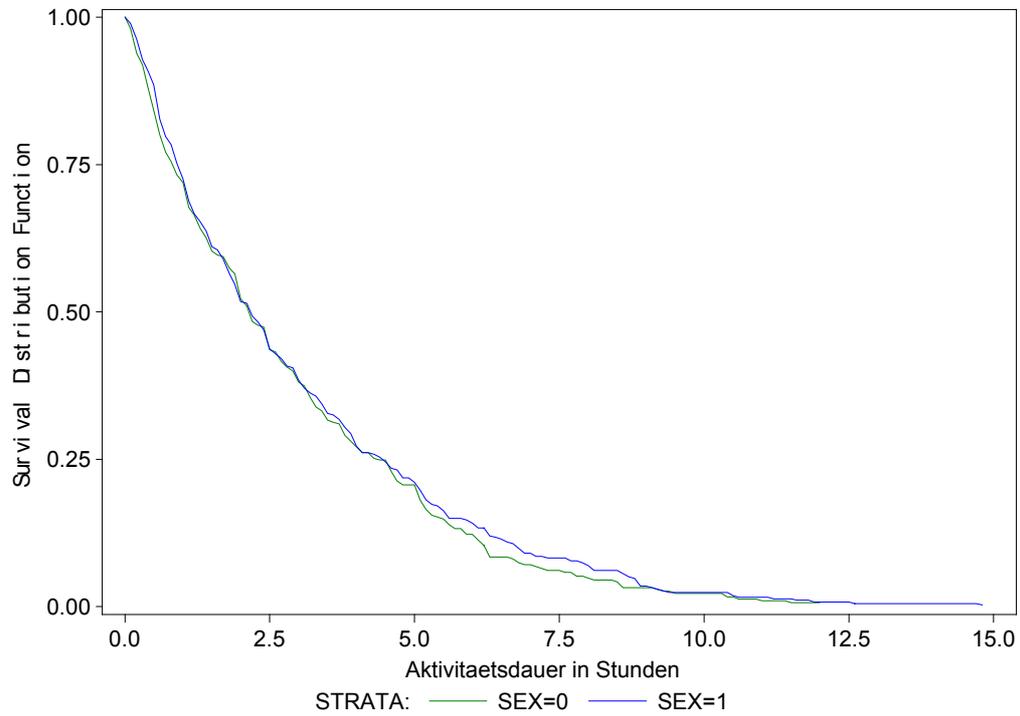
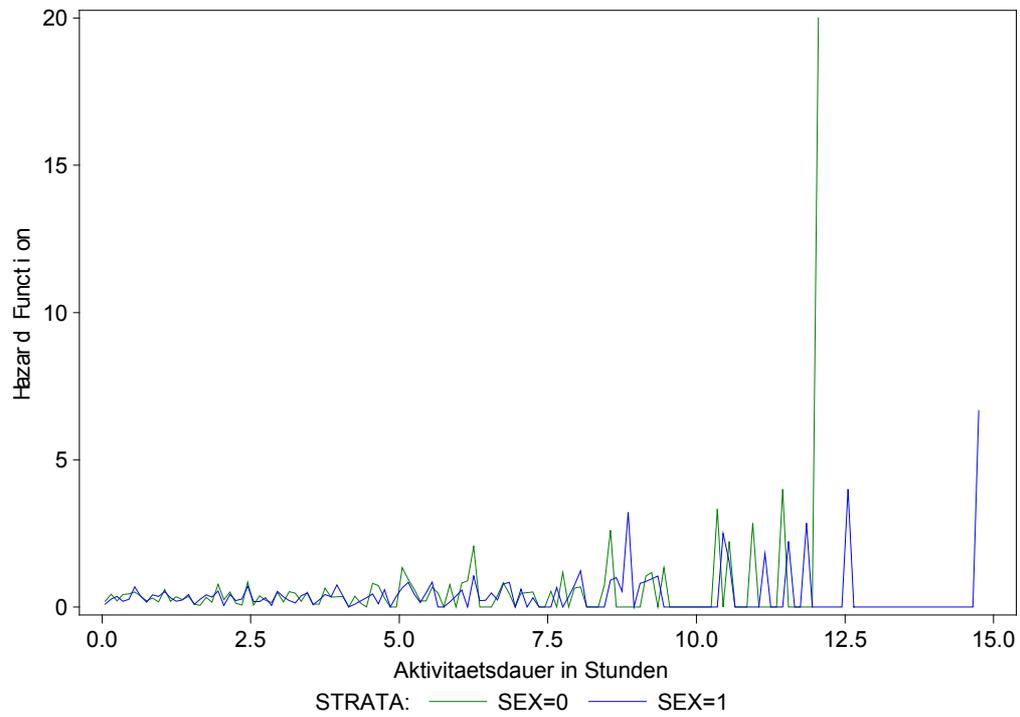


Abbildung V.30 : Hazardkurve Kategorie Treffen mit Freunden, Schichtung der Geschlechter



Zum Vergleich:

Abbildung V.31 : Geglättete Hazardkurve der Geschlechtsschichtung, Einkauf kurzfristig

---

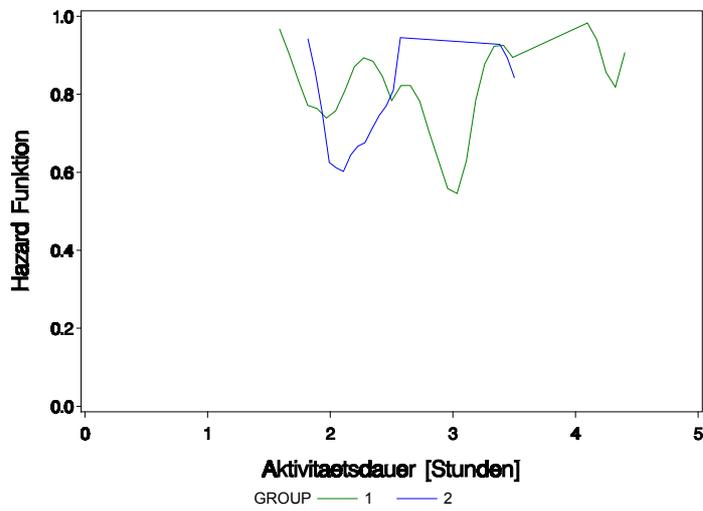


Abbildung V.32 : Geglättete Hazardkurve der Geschlechtsschichtung, Freizeit allgemein

---

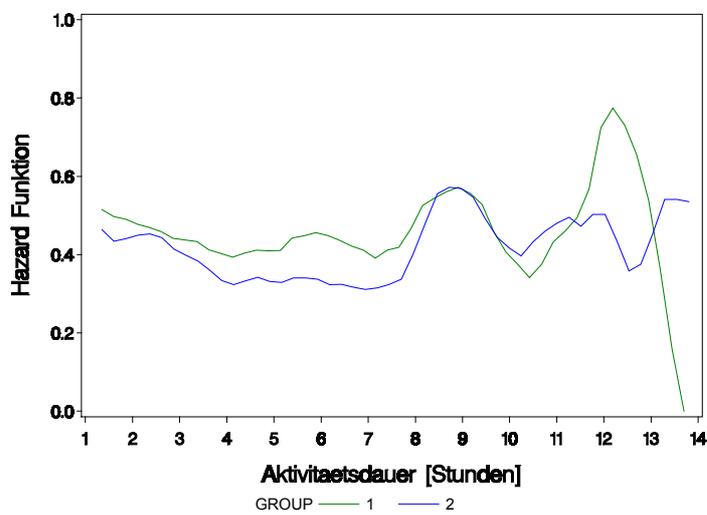


Abbildung V.33 : Geglättete Hazardkurve der Geschlechtsschichtung, Dienstleistung allgemein

---

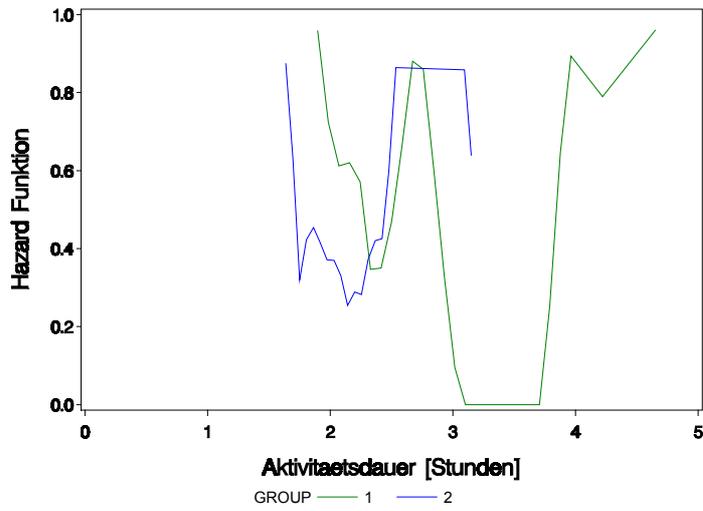
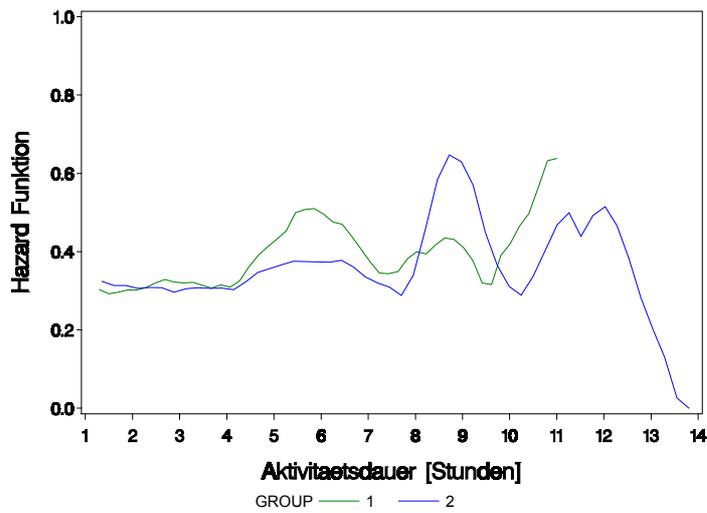


Abbildung V.34 : Geglättete Hazardkurve der Geschlechtsschichtung, Treffen mit Freunden

---



## **Anhang VI: Auswertungen von *PROC LIFETEST*: Wilcoxon und Log-Rank-Tests der 4 Kategorien, Aktivitätsdauern**

**Einerseits mit allen möglichen Variablen, andererseits mit all jenen, die in den semiparametrischen Auswertungen im 1. Gesamtmodell verwendet wurden.**

**Übersicht:**

**Tabellen VI.1 bis VI.5 Kategorie Einkauf kurzfristig**

**Tabellen VI.6 bis VI.10 Kategorie Freizeit allgemein**

**Tabellen VI.11 bis VI.15 Kategorie Dienstleistung allgemein**

**Tabellen VI.16 bis VI.20 Kategorie Treffen mit Freunden**

In allen Auswertungen ist die abhängige Variable „a\_durst“, also Aktivitätsdauer in Stunden.

Für jede Kategorie wurden zuerst ein Wilcoxon Test mit 3 verschiedenen Tests (Univariater Chi-Quadrat Test, Kovarianz Matrix für die Wilcoxon Statistiken und Stufenförmiger Chi-Quadrat-Test) und dann der Log-Rank-Test (ebenfalls mit den 3 verschiedenen Tests) gemacht. Danach wurden zusätzlich für jede der vier Kategorien noch diese beiden Tests (mit jeweils 3 Auswertungsteilen) für alle Variablen, die bei den semiparametrischen Regressionsmodellen (mit PROC PHREG) im 1. Gesamtmodell verwendet worden sind, durchgeführt. Die wichtigsten Ergebnisse werden hier präsentiert, die gesamte Auswertung ist im Zusatzanhang XIX (LIFETEST AD Logrank und Wilcoxontest) zu sehen.

Die Erläuterung zu den wichtigsten Ergebnissen finden Sie zudem im Kapitel 9.2.2.4.

Prinzipiell wird beim Wilcoxontest den frühen Ereigniszeiten, in diesem Fall also eher kurzen Aktivitätsdauern, mehr Bedeutung zugemessen, beim Log-Rank-Test gehen längere Aktivitätsdauern stärker gewichtet in die Berechnungen mit ein. Aus diesem Grund können auch schon bei den univariaten Tests unterschiedliche Ergebnisse entstehen (wie im Zusatzanhang XIX zu sehen.) In diesem Anhang werden bei den Kategorien Freizeit allgemein und Treffen mit Freunden die Ergebnisse des Log-Rank-Tests präsentiert, bei den Kategorien Einkauf kurzfristig und Dienstleistung allgemein die Ergebnisse des Wilcoxon-Tests, da es sich hier um meist kurze Aktivitätsdauern handelt bzw. 90 bis 95% der Aktivitäten schon im ersten Fünftel der Zeitachse beendet werden. Die univariaten und stufenförmigen Tests jener Variablen, die in den semiparametrischen Auswertungen im ersten Gesamtmodell verwendet werden, werden ebenso gezeigt, vor allem um Vergleiche zu den semiparametrischen Modellergebnissen anstellen zu können. Es sind dabei jeweils die stufenförmigen Log-Rank- und Wilcoxontests angegeben.

## Liste der möglichen Einflußvariablen und verwendete Variablendefinition:

Variablenbezeichnung	Variablenname	Dummyvariable
SEX	Geschlecht	X
YOUNG	Unter 50 Jahren j/n	X
LICENSE	Führerscheinbesitz j/n	X
HEAD_OHH	Dummyvariable Haushaltsvorstand j/n	X
WORKED	Dummyvariable gearbeitet j/n	X
CHILDREN	Kinder im Haushalt j/n	X
SINGLE	Allein lebend j/n	X
CAR_POSS	Fahrzeugzugangsmöglichkeit j/n	X
EDUC	Ausbildung während der Umfrage j/n	X
AGE	Alter	
EDUC_YR	Anzahl der Ausbildungsjahre	
G	Gesellschaftsgruppe	
CAR	Fahrzeug im Haushalt	
HH_NR_L	Anzahl der Führerscheine im HH	
HH_MEMB	Anzahl der Haushaltsmitglieder	
ACC_D	Begleitung j/n	X
PLANNED	Weg geplant j/n	X
A_EXP_D	Geld ausgegeben oder nicht	X
WEEKEND	Wochenende j/n	X
A_DAYT_D	Tageszeit des Akt.mittelp. (vo/na)	X
A_P_MM_D	Unabhängiges VM zur Aktivität j/n	X
A_N_MM_D	Unabhängiges VM nach der Aktivität j/n	X
A_P_MMMD	Unabhängiges motorisiertes VM zur Aktivität j/n	X
A_N_MMMD	Unabhängiges motorisiertes VM nach der Akt. j/n	X
ACC	Anzahl der Begleitpersonen aus dem HH	
A_EXP	Höhe der Ausgaben	
A_DAYT	Tageszeit des Aktivitätsmittelpunktes (vo/na/ab)	
A_P_DIST	Distanz von der vorigen Aktivität	
A_P_TT	Reisezeit von der vorigen Aktivität	
A_N_DIST	Distanz zur nächsten Aktivität	
A_N_TT	Reisezeit zur nächsten Aktivität	
A_P_MM	Hauptverkehrsmittel von voriger Aktivität	
A_N_MM	Hauptverkehrsmittel zur nächsten Aktivität	
NR_T_SUM	Anzahl der Wege gesamt	
NR_J_SUM	Anzahl der Reisen gesamt	
A_P_PUR	Voriger Aktivitätszweck (kontiv)	
A_N_PUR	Nächster Aktivitätszweck (kontiv)	
NR_A_SUMC	Anzahl der Aktivitäten in dieser Kategorie	

# Kategorie Einkauf kurzfristig, alle Variablen

Tabelle VI.1 : Univariate Chi-Squares for the Wilcoxon Test

---

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square Label
SEX	-118.6	15.1180	61.5506	<.0001
YOUNG	-103.3	15.9869	41.7449	<.0001
LICENSE	-158.0	15.8954	98.8434	<.0001
HEAD_OHH	-77.3241	15.7649	24.0573	<.0001
WORKED	-126.3	14.8965	71.8592	<.0001
CHILDREN	-7.4940	14.8714	0.2539	0.6143
SINGLE	-39.1664	14.4355	7.3614	0.0067
CAR_POSS	-127.1	16.0168	62.9426	<.0001
EDUC	-66.6791	13.7821	23.4072	<.0001
AGE	4856.7	568.9	72.8794	<.0001
EDUC_YR	-1225.4	140.5	76.0976	<.0001
G	-241.9	56.9252	18.0573	<.0001
CAR	-81.2945	15.6087	27.1262	<.0001
HH_NR_L	-184.2	28.2174	42.6092	<.0001
HH_MEMB	-3.2481	34.7521	0.00874	0.9255
ACC_D	104.0	13.1062	62.9891	<.0001
PLANNED	89.9253	10.1702	78.1813	<.0001
A_EXP_D	-0.9083	11.4302	0.00631	0.9367
WEEKEND	-14.3408	13.0148	1.2142	0.2705
A_DAYT_D	-63.0206	15.3997	16.7471	<.0001
A_P_MM_D	-72.4786	9.0398	64.2845	<.0001
A_N_MM_D	-74.8722	9.6107	60.6916	<.0001
A_P_MMMD	64.0280	12.9520	24.4380	<.0001
A_N_MMMD	65.9321	13.1034	25.3178	<.0001
ACC	125.7	19.4524	41.7434	<.0001
A_EXP	19335.8	1120.4	297.8	<.0001
A_DAYT	-89.3035	17.4899	26.0713	<.0001
A_P_DIST	429791	90487.1	22.5601	<.0001
A_P_TT	4229.2	300.7	197.8	<.0001
A_N_DIST	561052	350403	2.5637	0.1093
A_N_TT	4697.0	498.3	88.8675	<.0001
A_P_MM	316.3	42.8334	54.5214	<.0001
A_N_MM	331.1	43.1279	58.9331	<.0001
NR_T_SUM	-26690.7	1875.1	202.6	<.0001
NR_J_SUM	-4784.4	600.5	63.4753	<.0001
A_P_PUR	668.8	85.9828	60.5016	<.0001
A_N_PUR	55.7045	74.5531	0.5583	0.4550
NR_A_SUMC	-6548.2	534.6	150.0	<.0001

---

**Tabelle VI.2 :Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_EXP	1	297.8	<.0001	297.8	<.0001
NR_T_SUM	2	515.8	<.0001	217.9	<.0001
A_P_TT	3	661.2	<.0001	145.5	<.0001
A_N_MM	4	710.6	<.0001	49.4222	<.0001
SEX	5	780.8	<.0001	70.1352	<.0001
PLANNED	6	821.1	<.0001	40.3256	<.0001
NR_A_SUMC	7	860.0	<.0001	38.9343	<.0001
LICENSE	8	891.6	<.0001	31.5617	<.0001
CAR_POSS	9	922.0	<.0001	30.4346	<.0001
A_DAYT	10	948.5	<.0001	26.5028	<.0001
A_P_PUR	11	968.1	<.0001	19.5618	<.0001
A_EXP_D	12	986.1	<.0001	17.9563	<.0001
EDUC_YR	13	1001.9	<.0001	15.8612	<.0001
ACC_D	14	1015.8	<.0001	13.8628	0.0002
EDUC	15	1026.7	<.0001	10.9271	0.0009
A_N_TT	16	1037.3	<.0001	10.6384	0.0011
WEEKEND	17	1048.3	<.0001	10.9860	0.0009
HH_NR_L	18	1052.2	<.0001	3.9036	0.0482
SINGLE	19	1057.9	<.0001	5.6850	0.0171
CAR	20	1063.5	<.0001	5.6160	0.0178
AGE	21	1067.5	<.0001	3.9403	0.0471
A_P_MMMD	22	1070.8	<.0001	3.3007	0.0693
A_N_MM_D	23	1075.2	<.0001	4.4340	0.0352
A_N_MMMD	24	1078.4	<.0001	3.1707	0.0750
A_DAYT_D	25	1080.8	<.0001	2.4581	0.1169
HEAD_OHH	26	1083.0	<.0001	2.2071	0.1374
A_P_DIST	27	1085.1	<.0001	2.0872	0.1485
A_P_MM_D	28	1087.0	<.0001	1.8248	0.1767
A_P_MM	29	1088.8	<.0001	1.8934	0.1688
ACC	30	1090.3	<.0001	1.4314	0.2315
YOUNG	31	1091.8	<.0001	1.4748	0.2246
G	32	1092.6	<.0001	0.8393	0.3596
CHILDREN	33	1093.6	<.0001	1.0006	0.3172
WORKED	34	1094.3	<.0001	0.7188	0.3965
A_N_DIST	35	1094.6	<.0001	0.3279	0.5669
A_N_PUR	36	1094.8	<.0001	0.1974	0.6568
HH_MEMB	37	1094.9	<.0001	0.0805	0.7767
NR_J_SUM	38	1095.0	<.0001	0.1099	0.7403

Die Variablen A\_EXP, NR\_T\_SUM, A\_P\_TT, A\_N\_MM, SEX, PLANNED, NR\_A\_SUMC, LICENSE, CAR\_POSS, A\_DAYT, A\_P\_PUR, A\_EXP\_D, EDUC\_YR, ACC\_D, EDUC, A\_N\_TT, WEEKEND, HH\_NR\_L, SINGLE, CAR, AGE und A\_N\_MM\_D sind stark signifikant (im 95% Konfidenzintervall), A\_P\_MMMD und A\_N\_MMMD marginal signifikant (im 90% Konfidenzintervall). Im Vergleich zu den semiparametrischen Auswertungen sind hier jedoch keinerlei Korrelationen mit einberechnet. Auch findet für eine Variable auch nur jeweils eine Prüfung der darüber angeführten möglichen Einflußgrößen statt, nicht aber aller möglichen Variablen wie bei den semiparametrischen Modellen mit PROC PHREG. Die Ergebnisse können also irreführend sein bzw. gelten nur für die Nullhypothese, daß bei der PROC PHREG - Analyse alle Koeffizienten Null sind.

# Kategorie Einkauf kurzfristig, Variablen des 1. Gesamtmodells

Tabelle VI.3 : Univariate Chi-Squares for the Wilcoxon Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square	Label
SEX	-114.9	15.4812	55.0426	<.0001	
LICENSE	-168.0	16.1717	107.9	<.0001	
HEAD_OHH	-74.0341	16.0721	21.2186	<.0001	
WORKED	-125.8	15.0172	70.1829	<.0001	
SINGLE	-42.1114	14.6634	8.2476	0.0041	
EDUC	-67.1831	13.8582	23.5022	<.0001	
AGE	4934.2	571.5	74.5369	<.0001	
G	-238.8	57.8659	17.0355	<.0001	
CAR	-82.0043	15.8227	26.8602	<.0001	
PLANNED	91.1065	10.5351	74.7860	<.0001	
A_P_MM_D	-77.8929	9.3401	69.5491	<.0001	
A_N_MM_D	-81.6322	10.0520	65.9510	<.0001	
A_P_MMMD	58.5583	13.2431	19.5524	<.0001	
A_N_MMMD	56.8966	13.3991	18.0311	<.0001	
ACC	135.8	19.8273	46.9361	<.0001	
A_EXP	20762.6	1155.3	323.0	<.0001	
A_DAYT	-90.6254	17.7801	25.9795	<.0001	
A_P_DIST	444518	92358.6	23.1645	<.0001	
A_P_TT	4474.1	304.2	216.3	<.0001	
A_N_DIST	563134	350433	2.5823	0.1081	
A_N_TT	4910.2	503.5	95.1068	<.0001	
NR_T_SUM	-26743.1	1889.0	200.4	<.0001	
A_P_PUR	662.3	87.5548	57.2179	<.0001	
NR_A_SUMC	-6520.7	539.2	146.3	<.0001	

Bis auf die Variable A\_N\_DIST, die in den semiparametrischen Auswertungen im 1. Gesamtmodell aufgrund einer singulären Analyse der jeweiligen Variable mit PROC PHREG berücksichtigt wurde, sind auch hier alle möglichen Einflußvariablen singulär (univariat) betrachtet, stark signifikant im 95% Konfidenzintervall.

Im Vergleich dazu sind beim univariaten Log-Rank-Test alle begutachteten möglichen Einflußvariablen stark signifikant.

**Tabelle VI.4 : Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_EXP	1	323.0	<.0001	323.0	<.0001
A_P_TT	2	536.6	<.0001	213.6	<.0001
NR_T_SUM	3	697.8	<.0001	161.2	<.0001
LICENSE	4	745.4	<.0001	47.5817	<.0001
A_P_MMD	5	796.0	<.0001	50.6516	<.0001
A_N_TT	6	831.9	<.0001	35.8521	<.0001
PLANNED	7	864.6	<.0001	32.6559	<.0001
SEX	8	888.5	<.0001	23.9197	<.0001
NR_A_SUMC	9	920.7	<.0001	32.2348	<.0001
A_DAYT	10	945.4	<.0001	24.6826	<.0001
CAR	11	962.2	<.0001	16.8368	<.0001
A_N_MM_D	12	980.9	<.0001	18.6558	<.0001
A_P_PUR	13	993.4	<.0001	12.4889	0.0004
A_N_MMD	14	1003.4	<.0001	10.0250	0.0015
ACC	15	1009.7	<.0001	6.2585	0.0124
G	16	1019.7	<.0001	10.0437	0.0015
A_P_DIST	17	1024.7	<.0001	5.0157	0.0251
A_P_MM_D	18	1029.7	<.0001	4.9796	0.0256
SINGLE	19	1034.1	<.0001	4.4188	0.0355
EDUC	20	1036.6	<.0001	2.4456	0.1179
WORKED	21	1037.7	<.0001	1.1031	0.2936
HEAD_OHH	22	1038.3	<.0001	0.6141	0.4333
A_N_DIST	23	1038.6	<.0001	0.3076	0.5792
AGE	24	1038.7	<.0001	0.1268	0.7218

**Tabelle VI.5 : Forward Stepwise Sequence of Chi-Squares for Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_TT	1	141.8	<.0001	141.8	<.0001
NR_A_SUMC	2	234.5	<.0001	92.6664	<.0001
A_EXP	3	305.0	<.0001	70.4865	<.0001
PLANNED	4	361.2	<.0001	56.1872	<.0001
SEX	5	403.2	<.0001	42.0230	<.0001
A_N_MMMD	6	439.7	<.0001	36.5400	<.0001
LICENSE	7	483.5	<.0001	43.7886	<.0001
A_P_PUR	8	510.9	<.0001	27.4399	<.0001
A_N_MM_D	9	533.7	<.0001	22.7066	<.0001
A_N_TT	10	544.3	<.0001	10.6415	0.0011
SINGLE	11	553.9	<.0001	9.5987	0.0019
A_DAYT	12	561.1	<.0001	7.2349	0.0071
ACC	13	566.1	<.0001	4.9385	0.0263
G	14	571.3	<.0001	5.2281	0.0222
A_P_MM_D	15	573.9	<.0001	2.6390	0.1043
A_P_MMMD	16	576.7	<.0001	2.7714	0.0960
CAR	17	578.1	<.0001	1.3933	0.2379
HEAD_OHH	18	580.3	<.0001	2.2154	0.1366
A_N_DIST	19	581.5	<.0001	1.2145	0.2704
AGE	20	582.6	<.0001	1.0235	0.3117
EDUC	21	583.4	<.0001	0.8624	0.3531
NR_T_SUM	22	583.8	<.0001	0.4049	0.5246
A_P_DIST	23	584.0	<.0001	0.1677	0.6822
WORKED	24	584.0	<.0001	0.00184	0.9658

Die Variablen A\_P\_TT, NR\_A\_SUMC, A\_EXP, PLANNED, SEX, A\_N\_MMMD, LICENSE, A\_P\_PUR, A\_N\_MM\_D, A\_N\_TT, SINGLE, A\_DAYT, ACC und G sind stark signifikant (im 95% Konfidenzintervall), A\_P\_MMMD marginal signifikant (im 90% Konfidenzintervall). Im Vergleich zu den semiparametrischen Auswertungen sind hier jedoch keinerlei Korrelationen miteinberechnet. Auch findet für eine Variable auch nur jeweils eine Prüfung der darüber angeführten möglichen Einflußgrößen statt, nicht aber aller möglichen Variablen wie bei den semiparametrischen Modellen mit PROC PHREG. Die Ergebnisse können also irreführend sein bzw. gelten nur für die Nullhypothese, daß alle Koeffizienten Null sind.

# Kategorie Freizeit allgemein, alle Variablen

Tabelle VI.6 : Univariate Chi-Squares for the Log-Rank Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square Label
SEX	18.8531	26.6913	0.4989	0.4800
<b>YOUNG</b>	<b>-116.1</b>	<b>25.2738</b>	<b>21.1068</b>	<b>&lt;.0001</b>
<i>LICENSE</i>	<i>-48.7540</i>	<i>26.0696</i>	<i>3.4975</i>	<i>0.0615</i>
HEAD_OHH	33.8594	24.6211	1.8912	0.1691
WORKED	-20.5529	23.8290	0.7439	0.3884
<b>CHILDREN</b>	<b>-130.3</b>	<b>25.0282</b>	<b>27.0844</b>	<b>&lt;.0001</b>
SINGLE	19.9798	25.4999	0.6139	0.4333
<b>CAR_POSS</b>	<b>-55.9386</b>	<b>26.3743</b>	<b>4.4984</b>	<b>0.0339</b>
EDUC	-30.2648	23.8917	1.6047	0.2052
<b>AGE</b>	<b>2644.3</b>	<b>986.4</b>	<b>7.1856</b>	<b>0.0073</b>
<b>EDUC_YR</b>	<b>-1718.5</b>	<b>210.9</b>	<b>66.3765</b>	<b>&lt;.0001</b>
<b>G</b>	<b>-509.3</b>	<b>90.9017</b>	<b>31.3904</b>	<b>&lt;.0001</b>
<b>CAR</b>	<b>-104.7</b>	<b>24.9586</b>	<b>17.5947</b>	<b>&lt;.0001</b>
<b>HH_NR_L</b>	<b>-171.4</b>	<b>44.8111</b>	<b>14.6319</b>	<b>0.0001</b>
<b>HH_MEMB</b>	<b>-273.1</b>	<b>58.6650</b>	<b>21.6780</b>	<b>&lt;.0001</b>
<b>ACC_D</b>	<b>168.1</b>	<b>26.2440</b>	<b>41.0378</b>	<b>&lt;.0001</b>
<b>PLANNED</b>	<b>95.6038</b>	<b>11.1984</b>	<b>72.8844</b>	<b>&lt;.0001</b>
<b>A_EXP_D</b>	<b>-83.1734</b>	<b>15.7953</b>	<b>27.7278</b>	<b>&lt;.0001</b>
<b>WEEKEND</b>	<b>273.6</b>	<b>26.4039</b>	<b>107.4</b>	<b>&lt;.0001</b>
<b>A_DAYT_D</b>	<b>257.2</b>	<b>17.6853</b>	<b>211.5</b>	<b>&lt;.0001</b>
<b>A_P_MM_D</b>	<b>-78.7664</b>	<b>17.6826</b>	<b>19.8422</b>	<b>&lt;.0001</b>
<b>A_N_MM_D</b>	<b>-56.1601</b>	<b>16.5373</b>	<b>11.5326</b>	<b>0.0007</b>
<b>A_P_MMMD</b>	<b>347.7</b>	<b>26.2397</b>	<b>175.6</b>	<b>&lt;.0001</b>
<b>A_N_MMMD</b>	<b>384.7</b>	<b>25.9433</b>	<b>219.9</b>	<b>&lt;.0001</b>
<b>ACC</b>	<b>279.8</b>	<b>47.0448</b>	<b>35.3800</b>	<b>&lt;.0001</b>
<b>A_EXP</b>	<b>1059.6</b>	<b>1543.0</b>	<b>0.4715</b>	<b>0.4923</b>
<b>A_DAYT</b>	<b>436.1</b>	<b>34.4321</b>	<b>160.4</b>	<b>&lt;.0001</b>
<b>A_P_DIST</b>	<b>17729289</b>	<b>2023209</b>	<b>76.7894</b>	<b>&lt;.0001</b>
<b>A_P_TT</b>	<b>19444.2</b>	<b>3789.4</b>	<b>26.3288</b>	<b>&lt;.0001</b>
<b>A_N_DIST</b>	<b>19437264</b>	<b>2283412</b>	<b>72.4606</b>	<b>&lt;.0001</b>
<b>A_N_TT</b>	<b>20880.2</b>	<b>3202.0</b>	<b>42.5235</b>	<b>&lt;.0001</b>
<b>A_P_MM</b>	<b>1529.4</b>	<b>88.9608</b>	<b>295.6</b>	<b>&lt;.0001</b>
<b>A_N_MM</b>	<b>1537.6</b>	<b>86.2662</b>	<b>317.7</b>	<b>&lt;.0001</b>
<b>NR_T_SUM</b>	<b>-38748.6</b>	<b>3146.6</b>	<b>151.6</b>	<b>&lt;.0001</b>
<b>NR_J_SUM</b>	<b>-13628.1</b>	<b>1087.8</b>	<b>157.0</b>	<b>&lt;.0001</b>
<b>A_P_PUR</b>	<b>967.2</b>	<b>113.9</b>	<b>72.1067</b>	<b>&lt;.0001</b>
<b>A_N_PUR</b>	<b>1502.3</b>	<b>90.7288</b>	<b>274.2</b>	<b>&lt;.0001</b>
<b>NR_A_SUMC</b>	<b>-9680.6</b>	<b>1206.0</b>	<b>64.4325</b>	<b>&lt;.0001</b>

**Tabelle VI.7 : Forward Stepwise Sequence of Chi-Squares for Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_N_MM	1	317.7	<.0001	317.7	<.0001
A_N_PUR	2	537.9	<.0001	220.2	<.0001
NR_J_SUM	3	682.6	<.0001	144.8	<.0001
A_DAYT_D	4	811.2	<.0001	128.5	<.0001
PLANNED	5	882.2	<.0001	71.0591	<.0001
WEEKEND	6	933.0	<.0001	50.7486	<.0001
A_DAYT	7	976.3	<.0001	43.2652	<.0001
A_N_DIST	8	1010.2	<.0001	33.9210	<.0001
A_P_MM	9	1032.9	<.0001	22.7568	<.0001
HEAD_OHH	10	1053.0	<.0001	20.0521	<.0001
EDUC_YR	11	1069.5	<.0001	16.5489	<.0001
A_P_PUR	12	1080.6	<.0001	11.0179	0.0009
NR_A_SUMC	13	1093.1	<.0001	12.5939	0.0004
A_EXP_D	14	1100.1	<.0001	6.9603	0.0083
A_N_TT	15	1105.6	<.0001	5.4724	0.0193
A_P_MM_D	16	1110.6	<.0001	5.0047	0.0253
SEX	17	1116.7	<.0001	6.1532	0.0131
CAR	18	1123.4	<.0001	6.6663	0.0098
A_P_MMMD	19	1132.1	<.0001	8.6600	0.0033
HH_MEMB	20	1135.6	<.0001	3.5029	0.0613
ACC	21	1142.9	<.0001	7.3134	0.0068
ACC_D	22	1149.0	<.0001	6.0926	0.0136
EDUC	23	1153.6	<.0001	4.6471	0.0311
SINGLE	24	1157.4	<.0001	3.7540	0.0527
NR_T_SUM	25	1159.2	<.0001	1.8449	0.1744
A_P_TT	26	1161.0	<.0001	1.7745	0.1828
CHILDREN	27	1162.2	<.0001	1.1706	0.2793
A_N_MMMD	28	1163.4	<.0001	1.2100	0.2713
A_N_MM_D	29	1164.7	<.0001	1.2956	0.2550
G	30	1165.6	<.0001	0.9772	0.3229
AGE	31	1170.1	<.0001	4.4740	0.0344
CAR_POSS	32	1170.5	<.0001	0.3945	0.5300
HH_NR_L	33	1170.9	<.0001	0.4063	0.5239
A_P_DIST	34	1171.0	<.0001	0.1114	0.7386
LICENSE	35	1171.1	<.0001	0.0434	0.8349
YOUNG	36	1171.1	<.0001	0.0568	0.8117
A_EXP	37	1171.2	<.0001	0.0259	0.8722
WORKED	38	1171.2	<.0001	0.00850	0.9265

Die Variablen A\_N\_MM, A\_N\_PUR, NR\_J\_SUM, A\_DAYT\_D, PLANNED, WEEKEND, A\_DAYT, A\_N\_DIST, A\_P\_MM, HEAD\_OHH, EDUC\_YR, A\_P\_PUR, NR\_A\_SUMC, A\_EXP\_D, A\_N\_TT, A\_P\_MM\_D, SEX, CAR, A\_P\_MMMD, ACC, ACC\_D, EDUC und AGE sind stark signifikant (im 95% Konfidenzintervall), HH\_MEMB und SINGLE marginal signifikant (im 90% Konfidenzintervall). Im Vergleich zu den semiparametrischen Auswertungen sind hier jedoch keinerlei Korrelationen miteinberechnet. Auch findet für eine Variable auch nur jeweils eine Prüfung der darüber angeführten möglichen Einflußgrößen statt, nicht aber aller möglichen Variablen wie bei den semiparametrischen Modellen mit PROC PHREG. Die Ergebnisse können also irreführend sein bzw. gelten nur für die Nullhypothese, daß alle Koeffizienten Null sind.

# Kategorie Freizeit allgemein, Variablen aus dem 1. Gesamtmodell

**Tabelle VI.8 : Univariate Chi-Squares for the Log-Rank Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square	Label
CAR	-94.4749	25.2307	14.0209	0.0002	
HH_MEMB	-267.2	60.9428	19.2272	<.0001	
ACC_D	173.1	26.7727	41.7875	<.0001	
PLANNED	98.9710	11.9763	68.2918	<.0001	
A_EXP_D	-83.7579	16.4106	26.0497	<.0001	
WEEKEND	281.0	26.9193	109.0	<.0001	
A_DAYT_D	265.4	17.7542	223.5	<.0001	
A_P_MM_D	-76.3870	17.9636	18.0822	<.0001	
A_N_MM_D	-50.9661	16.7631	9.2439	0.0024	
A_P_DIST	17811570	2026210	77.2744	<.0001	
A_N_TT	20820.4	3208.9	42.0989	<.0001	
A_N_MM	1596.0	88.6798	323.9	<.0001	
NR_J_SUM	-13895.5	1088.4	163.0	<.0001	
A_P_PUR	964.4	118.1	66.6919	<.0001	
A_N_PUR	1531.6	94.0031	265.5	<.0001	

Conclusio: Alle bei den semiparametrischen Auswertungen verwendeten Variablen sind auch hier statistisch signifikant im 95%-Konfidenzintervall. Beim univariaten Wilcoxonstest sind ebenfalls alle Variablen statistisch signifikant.

**Tabelle VI.9 : Forward Stepwise Sequence of Chi-Squares for Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_N_MM	1	323.9	<.0001	323.9	<.0001
A_N_PUR	2	540.6	<.0001	216.7	<.0001
NR_J_SUM	3	691.7	<.0001	151.1	<.0001
A_DAYT_D	4	831.5	<.0001	139.7	<.0001
PLANNED	5	897.3	<.0001	65.8072	<.0001
WEEKEND	6	944.2	<.0001	46.9483	<.0001
HH_MEMB	7	967.2	<.0001	22.9520	<.0001
A_P_DIST	8	990.8	<.0001	23.6472	<.0001
A_P_PUR	9	1007.8	<.0001	17.0088	<.0001
A_EXP_D	10	1013.2	<.0001	5.3853	0.0203
A_N_TT	11	1016.3	<.0001	3.1290	0.0769
A_P_MM_D	12	1018.7	<.0001	2.3237	0.1274
ACC_D	13	1019.3	<.0001	0.6484	0.4207
CAR	14	1019.8	<.0001	0.4632	0.4961
A_N_MM_D	15	1019.8	<.0001	0.0305	0.8613

Die Variablen A\_N\_MM, A\_N\_PUR, NR\_J\_SUM, A\_DAYT\_D, PLANNED, WEEKEND, HH\_MEMB, A\_P\_DIST, A\_P\_PUR und A\_EXP\_D sind stark, A\_N\_TT im Vergleich zum Wilcoxonstest nur noch marginal signifikant. Die beiden im Wilcoxonstest noch (marginal) signifikanten Variablen A\_P\_MM\_D und A\_N\_MM\_D sind nicht mehr signifikant! Im Vergleich zu den semiparametrischen Auswertungen sind hier jedoch keinerlei Korrelationen miteinander berechnet.

**Tabelle VI.10 : Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_N_MM	1	345.5	<.0001	345.5	<.0001
A_N_PUR	2	589.9	<.0001	244.4	<.0001
NR_J_SUM	3	784.0	<.0001	194.1	<.0001
A_DAYT_D	4	878.0	<.0001	94.0409	<.0001
PLANNED	5	958.1	<.0001	80.1007	<.0001
A_N_TT	6	1026.9	<.0001	68.7700	<.0001
WEEKEND	7	1074.6	<.0001	47.6552	<.0001
HH_MEMB	8	1095.4	<.0001	20.8010	<.0001
A_P_DIST	9	1111.9	<.0001	16.5280	<.0001
A_P_MM_D	10	1123.3	<.0001	11.4438	0.0007
A_P_PUR	11	1131.7	<.0001	8.3533	0.0038
A_EXP_D	12	1136.7	<.0001	4.9782	0.0257
A_N_MM_D	13	1139.7	<.0001	3.0570	0.0804
CAR	14	1141.1	<.0001	1.3761	0.2408
ACC_D	15	1141.9	<.0001	0.8513	0.3562

Die Variablen A\_N\_MM, A\_N\_PUR, NR\_J\_SUM, A\_DAYT\_D, PLANNED, A\_N\_TT, WEEKEND, HH\_MEMB, A\_P\_DIST, A\_P\_MM\_D, A\_P\_PUR und A\_EXP\_D sind stark signifikant (im 95% Konfidenzintervall), A\_N\_MMMD marginal signifikant (im 90% Konfidenzintervall).

# Kategorie Dienstleistung allg., alle Variablen

Tabelle VI.11 : Univariate Chi-Squares for the Wilcoxon Test

---

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square	Label
<b>SEX</b>	<b>-40.2353</b>	<b>8.3170</b>	<b>23.4032</b>	<b>&lt;.0001</b>	
<b>YOUNG</b>	<b>-39.8188</b>	<b>8.3256</b>	<b>22.8743</b>	<b>&lt;.0001</b>	
<b>LICENSE</b>	<b>-32.2259</b>	<b>8.1881</b>	<b>15.4898</b>	<b>&lt;.0001</b>	
HEAD_OHH	-12.2576	7.6049	2.5980	0.1070	
<b>WORKED</b>	<b>-39.1635</b>	<b>7.7726</b>	<b>25.3883</b>	<b>&lt;.0001</b>	
<b>CHILDREN</b>	<b>-28.9824</b>	<b>7.8618</b>	<b>13.5902</b>	<b>0.0002</b>	
SINGLE	5.6141	7.5461	0.5535	0.4569	
<b>CAR_POSS</b>	<b>-30.7612</b>	<b>8.2473</b>	<b>13.9117</b>	<b>0.0002</b>	
EDUC	-4.4976	7.5918	0.3510	0.5536	
<b>AGE</b>	<b>1664.3</b>	<b>309.0</b>	<b>29.0052</b>	<b>&lt;.0001</b>	
<b>EDUC_YR</b>	<b>-473.2</b>	<b>81.9833</b>	<b>33.3121</b>	<b>&lt;.0001</b>	
<b>G</b>	<b>-142.4</b>	<b>30.7699</b>	<b>21.4203</b>	<b>&lt;.0001</b>	
<b>CAR</b>	<b>-23.3694</b>	<b>8.0634</b>	<b>8.3996</b>	<b>0.0038</b>	
<b>HH_NR_L</b>	<b>-45.3706</b>	<b>16.2984</b>	<b>7.7493</b>	<b>0.0054</b>	
<b>HH_MEMB</b>	<b>-53.3247</b>	<b>18.4875</b>	<b>8.3195</b>	<b>0.0039</b>	
ACC_D	10.1376	6.2356	2.6431	0.1040	
PLANNED	-1.6871	3.6402	0.2148	0.6430	
<b>A_EXP_D</b>	<b>30.8894</b>	<b>7.1439</b>	<b>18.6958</b>	<b>&lt;.0001</b>	
WEEKEND	-4.3035	5.5488	0.6015	0.4380	
A_DAYT_D	-9.7494	8.2538	1.3952	0.2375	
<b>A_P_MM_D</b>	<b>-22.0153</b>	<b>4.8838</b>	<b>20.3204</b>	<b>&lt;.0001</b>	
<b>A_N_MM_D</b>	<b>-14.3400</b>	<b>4.0285</b>	<b>12.6710</b>	<b>0.0004</b>	
A_P_MMMD	8.8365	7.0686	1.5627	0.2113	
A_N_MMMD	5.3400	6.9082	0.5975	0.4395	
ACC	12.6235	7.7453	2.6563	0.1031	
A_EXP	-676.2	2523.7	0.0718	0.7887	
A_DAYT	-11.5424	9.5649	1.4562	0.2275	
A_P_DIST	145288	77778.0	3.4893	0.0618	
<b>A_P_TT</b>	<b>1396.5</b>	<b>167.8</b>	<b>69.3014</b>	<b>&lt;.0001</b>	
A_N_DIST	131390	88534.5	2.2024	0.1378	
<b>A_N_TT</b>	<b>1283.4</b>	<b>213.1</b>	<b>36.2797</b>	<b>&lt;.0001</b>	
<b>A_P_MM</b>	<b>88.3482</b>	<b>23.8976</b>	<b>13.6674</b>	<b>0.0002</b>	
<b>A_N_MM</b>	<b>57.9494</b>	<b>23.7499</b>	<b>5.9535</b>	<b>0.0147</b>	
<b>NR_T_SUM</b>	<b>-7872.2</b>	<b>1030.1</b>	<b>58.4031</b>	<b>&lt;.0001</b>	
<b>NR_J_SUM</b>	<b>-1824.0</b>	<b>361.7</b>	<b>25.4247</b>	<b>&lt;.0001</b>	
A_P_PUR	119.0	47.2790	6.3340	0.0118	
A_N_PUR	126.5	42.0009	9.0678	0.0026	
<b>NR_A_SUMC</b>	<b>-430.4</b>	<b>161.9</b>	<b>7.0660</b>	<b>0.0079</b>	

---

**Tabelle VI.12 : Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_TT	1	<b>69.3014</b>	<b>&lt;.0001</b>	<b>69.3014</b>	<b>&lt;.0001</b>
NR_T_SUM	2	<b>105.7</b>	<b>&lt;.0001</b>	<b>36.3635</b>	<b>&lt;.0001</b>
A_EXP_D	3	<b>124.3</b>	<b>&lt;.0001</b>	<b>18.6057</b>	<b>&lt;.0001</b>
SEX	4	<b>134.3</b>	<b>&lt;.0001</b>	<b>9.9818</b>	<b>0.0016</b>
A_P_MM	5	<b>145.0</b>	<b>&lt;.0001</b>	<b>10.7751</b>	<b>0.0010</b>
A_P_DIST	6	<b>153.1</b>	<b>&lt;.0001</b>	<b>8.0537</b>	<b>0.0045</b>
A_N_TT	7	<b>159.8</b>	<b>&lt;.0001</b>	<b>6.7401</b>	<b>0.0094</b>
A_EXP	8	<b>165.3</b>	<b>&lt;.0001</b>	<b>5.4994</b>	<b>0.0190</b>
NR_J_SUM	9	168.7	<.0001	3.3837	0.0658
NR_A_SUMC	10	<b>173.1</b>	<b>&lt;.0001</b>	<b>4.4079</b>	<b>0.0358</b>
PLANNED	11	176.4	<.0001	3.2960	0.0695
SINGLE	12	178.4	<.0001	1.9596	0.1616
EDUC_YR	13	179.8	<.0001	1.4311	0.2316
A_P_MMMD	14	181.0	<.0001	1.2461	0.2643
A_P_MM_D	15	182.5	<.0001	1.4345	0.2310
AGE	16	183.9	<.0001	1.4227	0.2330
CAR	17	186.2	<.0001	2.3024	0.1292
YOUNG	18	188.3	<.0001	2.1192	0.1455
HH_MEMB	19	<b>196.0</b>	<b>&lt;.0001</b>	<b>7.6641</b>	<b>0.0056</b>
A_N_DIST	20	196.9	<.0001	0.9103	0.3400
A_N_MM_D	21	197.8	<.0001	0.8949	0.3442
HEAD_OHH	22	198.5	<.0001	0.6648	0.4149
CAR_POSS	23	199.3	<.0001	0.8827	0.3475
WORKED	24	200.0	<.0001	0.6158	0.4326
HH_NR_L	25	200.4	<.0001	0.4304	0.5118
ACC	26	200.8	<.0001	0.3730	0.5414
ACC_D	27	201.6	<.0001	0.8211	0.3649
WEEKEND	28	202.0	<.0001	0.4016	0.5263
CHILDREN	29	202.2	<.0001	0.2129	0.6445
G	30	202.3	<.0001	0.1081	0.7423
A_P_PUR	31	202.4	<.0001	0.0830	0.7733
A_N_PUR	32	202.5	<.0001	0.0682	0.7940
A_N_MMMD	33	202.5	<.0001	0.0326	0.8568
A_N_MM	34	202.6	<.0001	0.0913	0.7625
A_DAYT	35	202.6	<.0001	0.0257	0.8725
A_DAYT_D	36	202.8	<.0001	0.2273	0.6336
EDUC	37	202.8	<.0001	0.0131	0.9088
LICENSE	38	202.8	<.0001	0.000212	0.9884

Die Variablen A\_P\_TT, NR\_T\_SUM, A\_EXP\_D, SEX, A\_P\_MM, A\_P\_DIST, A\_N\_TT, A\_EXP, NR\_A\_SUMC und HH\_MEMB sind stark signifikant (im 95% Konfidenzintervall), NR\_J\_SUM und PLANNED marginal signifikant (im 90% Konfidenzintervall). Im Vergleich zu den semiparametrischen Auswertungen sind hier jedoch keinerlei Korrelationen miteinberechnet. Auch findet für eine Variable auch nur jeweils eine Prüfung der darüber angeführten möglichen Einflußgrößen statt, nicht aber aller möglichen Variablen wie bei den semiparametrischen Modellen mit PROC PHREG. Die Ergebnisse können also irreführend sein bzw. gelten nur für die Nullhypothese, daß alle Koeffizienten Null sind.

## Kategorie Dienstleistung allgemein, Variablen aus dem 1. Gesamtmodell

**Tabelle VI.13 : Univariate Chi-Squares for the Wilcoxon Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square	Label
SEX	-39.1449	8.5776	20.8268	<.0001	
HEAD_OHH	-11.2575	7.8057	2.0800	0.1492	
WORKED	-39.1650	7.8635	24.8067	<.0001	
CAR_POSS	-29.2118	8.4853	11.8519	0.0006	
AGE	1681.0	310.9	29.2363	<.0001	
HH_MEMB	-60.1962	19.3277	9.7001	0.0018	
A_EXP_D	29.0089	7.3062	15.7645	<.0001	
A_P_MM_D	-24.4125	5.1285	22.6591	<.0001	
A_N_MM_D	-14.8194	4.2270	12.2911	0.0005	
ACC	12.3489	7.8252	2.4904	0.1145	
A_P_TT	1469.5	171.7	73.2384	<.0001	
A_N_TT	1278.5	214.7	35.4496	<.0001	
A_P_MM	96.2698	24.3522	15.6280	<.0001	
NR_T_SUM	-7830.9	1037.6	56.9585	<.0001	
A_P_PUR	120.3	49.5476	5.8910	0.0152	
A_N_PUR	133.5	43.7289	9.3183	0.0023	
NR_A_SUMC	-436.3	162.6	7.2043	0.0073	

Nur die beiden Variablen HEAD\_OHH und ACC sind in der semiparametrischen Auswertung verwendet worden, da sie in der singulären PROC PHREG Variablenbetrachtung statistisch signifikant waren. Auch beim Log-Rank-Test sind nur diese beiden Variablen statistisch nicht signifikant.

**Tabelle VI.14 : Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_TT	1	73.2384	<.0001	73.2384	<.0001
NR_T_SUM	2	107.8	<.0001	34.5881	<.0001
A_EXP_D	3	123.3	<.0001	15.4781	<.0001
A_N_TT	4	132.1	<.0001	8.7627	0.0031
A_P_MM	5	140.5	<.0001	8.4315	0.0037
SEX	6	150.4	<.0001	9.9054	0.0016
WORKED	7	155.2	<.0001	4.8034	0.0284
AGE	8	157.1	<.0001	1.9115	0.1668
NR_A_SUMC	9	157.9	<.0001	0.8014	0.3707
A_N_MM_D	10	158.7	<.0001	0.7850	0.3756
CAR_POSS	11	159.5	<.0001	0.8168	0.3661
A_N_PUR	12	159.6	<.0001	0.1061	0.7446
HEAD_OHH	13	159.7	<.0001	0.0417	0.8382
ACC	14	159.7	<.0001	0.0176	0.8944
A_P_PUR	15	159.7	<.0001	0.00969	0.9216
HH_MEMB	16	159.7	<.0001	0.00890	0.9248
A_P_MM_D	17	159.7	<.0001	0.000268	0.9869

**Tabelle VI.15 : Forward Stepwise Sequence of Chi-Squares for Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_TT	1	<b>32.6834</b>	<.0001	<b>32.6834</b>	<.0001
NR_T_SUM	2	<b>52.0977</b>	<.0001	<b>19.4142</b>	<.0001
A_EXP_D	3	<b>62.2034</b>	<.0001	<b>10.1058</b>	<b>0.0015</b>
A_N_PUR	4	<b>70.0850</b>	<.0001	<b>7.8815</b>	<b>0.0050</b>
SEX	5	<b>75.7538</b>	<.0001	<b>5.6688</b>	<b>0.0173</b>
A_P_MM	6	<b>80.5830</b>	<.0001	<b>4.8293</b>	<b>0.0280</b>
ACC	7	<i>83.2955</i>	<i>&lt;.0001</i>	<i>2.7125</i>	<i>0.0996</i>
AGE	8	84.7282	<.0001	1.4327	0.2313
CAR_POSS	9	85.9058	<.0001	1.1776	0.2778
A_N_TT	10	87.0060	<.0001	1.1002	0.2942
NR_A_SUMC	11	87.8471	<.0001	0.8411	0.3591
A_P_PUR	12	88.6957	<.0001	0.8486	0.3570
A_P_MM_D	13	89.2742	<.0001	0.5785	0.4469
HEAD_OHH	14	89.7334	<.0001	0.4592	0.4980
HH_MEMB	15	90.6595	<.0001	0.9261	0.3359
A_N_MM_D	16	90.9955	<.0001	0.3361	0.5621
WORKED	17	91.0641	<.0001	0.0686	0.7935

zum Vergleich!

# Kategorie Treffen mit Freunden, alle Variablen

Tabelle VI.16 : Univariate Chi-Squares for the Log-Rank Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square	Label
SEX	10.4840	12.3664	0.7187	0.3966	
YOUNG	-11.8319	11.7852	1.0079	0.3154	
LICENSE	-11.0474	12.2795	0.8094	0.3683	
HEAD_OHH	9.7542	11.2978	0.7454	0.3879	
WORKED	11.8959	11.2216	1.1238	0.2891	
<b>CHILDREN</b>	<b>-25.7975</b>	<b>10.7442</b>	<b>5.7651</b>	<b>0.0163</b>	
SINGLE	2.8801	12.2284	0.0555	0.8138	
CAR_POSS	-11.9757	12.3234	0.9444	0.3312	
EDUC	-12.4837	10.5370	1.4037	0.2361	
AGE	-167.9	467.4	0.1291	0.7194	
EDUC_YR	-131.3	96.3978	1.8548	0.1732	
<i>G</i>	<i>-63.1049</i>	<i>38.2701</i>	<i>2.7190</i>	<i>0.0992</i>	
CAR	-11.3826	11.6876	0.9485	0.3301	
<b>HH_NR_L</b>	<b>-38.0916</b>	<b>19.2472</b>	<b>3.9167</b>	<b>0.0478</b>	
HH_MEMB	-39.2646	27.3891	2.0552	0.1517	
<i>ACC_D</i>	<i>20.3670</i>	<i>11.7566</i>	<i>3.0012</i>	<i>0.0832</i>	
<b>PLANNED</b>	<b>14.3041</b>	<b>5.3540</b>	<b>7.1378</b>	<b>0.0075</b>	
<b>A_EXP_D</b>	<b>-11.8296</b>	<b>3.4435</b>	<b>11.8015</b>	<b>0.0006</b>	
<b>WEEKEND</b>	<b>40.9837</b>	<b>12.3453</b>	<b>11.0209</b>	<b>0.0009</b>	
<b>A_DAYT_D</b>	<b>46.0781</b>	<b>7.1458</b>	<b>41.5808</b>	<b>&lt;.0001</b>	
A_P_MM_D	-3.4080	8.2918	0.1689	0.6811	
A_N_MM_D	-12.9837	8.6737	2.2407	0.1344	
<b>A_P_MMMD</b>	<b>48.0981</b>	<b>12.1073</b>	<b>15.7818</b>	<b>&lt;.0001</b>	
<b>A_N_MMMD</b>	<b>41.9580</b>	<b>12.1073</b>	<b>12.0099</b>	<b>0.0005</b>	
<b>ACC</b>	<b>46.3209</b>	<b>21.8830</b>	<b>4.4806</b>	<b>0.0343</b>	
A_EXP	-166.6	91.2626	3.3321	0.0679	
<b>A_DAYT</b>	<b>85.0432</b>	<b>15.8964</b>	<b>28.6210</b>	<b>&lt;.0001</b>	
<b>A_P_DIST</b>	<b>1304771</b>	<b>419349</b>	<b>9.6809</b>	<b>0.0019</b>	
<b>A_P_TT</b>	<b>2055.3</b>	<b>538.0</b>	<b>14.5940</b>	<b>0.0001</b>	
<b>A_N_DIST</b>	<b>1582605</b>	<b>433700</b>	<b>13.3158</b>	<b>0.0003</b>	
<b>A_N_TT</b>	<b>1608.9</b>	<b>517.1</b>	<b>9.6798</b>	<b>0.0019</b>	
<b>A_P_MM</b>	<b>188.7</b>	<b>38.5989</b>	<b>23.9064</b>	<b>&lt;.0001</b>	
<b>A_N_MM</b>	<b>185.5</b>	<b>38.5271</b>	<b>23.1766</b>	<b>&lt;.0001</b>	
NR_T_SUM	-2095.0	1500.5	1.9494	0.1626	
NR_J_SUM	6.7586	551.6	0.000150	0.9902	
A_P_PUR	56.1586	48.9301	1.3173	0.2511	
<b>A_N_PUR</b>	<b>203.2</b>	<b>45.4686</b>	<b>19.9671</b>	<b>&lt;.0001</b>	
<b>NR_A_SUMC</b>	<b>809.2</b>	<b>386.1</b>	<b>4.3927</b>	<b>0.0361</b>	

**Tabelle VI.17 : Forward Stepwise Sequence of Chi-Squares for Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_DAYT_D	1	41.5808	<.0001	41.5808	<.0001
A_P_MM	2	64.9086	<.0001	23.3279	<.0001
A_N_PUR	3	80.3008	<.0001	15.3922	<.0001
NR_A_SUMC	4	90.4505	<.0001	10.1497	0.0014
A_N_DIST	5	106.4	<.0001	15.9981	<.0001
HH_NR_L	6	117.8	<.0001	11.3071	0.0008
A_EXP_D	7	125.7	<.0001	7.9856	0.0047
PLANNED	8	133.6	<.0001	7.8200	0.0052
ACC	9	138.7	<.0001	5.1302	0.0235
G	10	144.0	<.0001	5.3408	0.0208
A_DAYT	11	149.0	<.0001	4.9829	0.0256
A_EXP	12	153.4	<.0001	4.3662	0.0367
A_N_MM_D	13	157.6	<.0001	4.1830	0.0408
ACC_D	14	161.7	<.0001	4.1690	0.0412
A_P_PUR	15	166.7	<.0001	4.9198	0.0266
A_N_MM	16	169.7	<.0001	3.0740	0.0796
SEX	17	172.5	<.0001	2.8185	0.0932
CAR_POSS	18	176.0	<.0001	3.4777	0.0622
CHILDREN	19	179.6	<.0001	3.5690	0.0589
WEEKEND	20	181.9	<.0001	2.3065	0.1288
HH_MEMB	21	184.1	<.0001	2.1683	0.1409
EDUC	22	185.5	<.0001	1.3976	0.2371
NR_T_SUM	23	186.7	<.0001	1.2824	0.2575
NR_J_SUM	24	188.6	<.0001	1.8470	0.1741
WORKED	25	189.8	<.0001	1.1712	0.2791
A_P_DIST	26	190.6	<.0001	0.8569	0.3546
A_P_TT	27	191.5	<.0001	0.8412	0.3591
A_P_MM_D	28	192.5	<.0001	1.0546	0.3045
EDUC_YR	29	192.9	<.0001	0.3786	0.5383
HEAD_OHH	30	193.3	<.0001	0.4428	0.5058
YOUNG	31	193.8	<.0001	0.4121	0.5209
A_N_MMMD	32	194.1	<.0001	0.3755	0.5400
AGE	33	194.3	<.0001	0.1975	0.6568
LICENSE	34	194.4	<.0001	0.1150	0.7346
A_P_MMMD	35	194.5	<.0001	0.0256	0.8728
CAR	36	194.5	<.0001	0.00250	0.9601
A_N_TT	37	194.5	<.0001	0.000234	0.9878
SINGLE	38	194.5	<.0001	0.000046	0.9946

## Kategorie Treffen mit Freunden, Variablen aus dem 1. Gesamtmodell

**Tabelle VI.18 : Univariate Chi-Squares for the Log-Rank Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square Label
CHILDREN	-27.5236	11.2679	5.9665	0.0146
HH_NR_L	-41.2991	19.5765	4.4505	0.0349
PLANNED	12.5436	5.9821	4.3968	0.0360
A_EXP_D	-11.6968	3.4646	11.3983	0.0007
WEEKEND	43.2317	12.7317	11.5301	0.0007
A_DAYT_D	48.1266	7.3929	42.3785	<.0001
A_P_MMD	45.5764	12.5799	13.1259	0.0003
A_N_MMD	38.6140	12.5757	9.4280	0.0021
ACC	42.4708	22.6615	3.5124	0.0609
A_EXP	-164.3	91.6562	3.2117	0.0731
A_DAYT	90.3702	16.4715	30.1014	<.0001
A_P_DIST	1346274	431788	9.7213	0.0018
A_P_TT	2174.1	560.4	15.0523	0.0001
A_N_DIST	1600707	445955	12.8837	0.0003
A_N_TT	1616.9	543.4	8.8529	0.0029
A_P_MM	186.0	40.2675	21.3457	<.0001
A_N_MM	180.2	40.5095	19.7942	<.0001
A_N_PUR	205.1	48.6943	17.7353	<.0001
NR_A_SUMC	833.0	390.7	4.5461	0.0330

Hier ergeben sich erstmals größere Unterschiede zwischen Log-Rank und Wilcoxon-Test: Sind beim Log-Rank-Test die Variablen ACC und A\_EXP nur marginal signifikant, so sind beim Wilcoxon-Test die Variablen A\_EXP (0.1417) und NR\_A\_SUMC (0.2141) nicht mehr signifikant im 90%-Konfidenzintervall.

**Tabelle VI.19 : Forward Stepwise Sequence of Chi-Squares for Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_DAYT_D	1	42.3785	<.0001	42.3785	<.0001
A_P_MM	2	62.2676	<.0001	19.8890	<.0001
A_N_PUR	3	75.1307	<.0001	12.8632	0.0003
A_N_DIST	4	85.6839	<.0001	10.5531	0.0012
NR_A_SUMC	5	100.6	<.0001	14.9044	0.0001
HH_NR_L	6	112.4	<.0001	11.7684	0.0006
A_EXP_D	7	120.3	<.0001	7.9926	0.0047
ACC	8	124.7	<.0001	4.3540	0.0369
CHILDREN	9	129.1	<.0001	4.3597	0.0368
A_DAYT	10	133.0	<.0001	3.9499	0.0469
A_EXP	11	136.9	<.0001	3.8920	0.0485
PLANNED	12	139.9	<.0001	2.9617	0.0853
WEEKEND	13	141.9	<.0001	2.0410	0.1531
A_P_TT	14	144.0	<.0001	2.1297	0.1445
A_P_DIST	15	147.2	<.0001	3.1887	0.0741
A_N_MMMD	16	148.0	<.0001	0.7457	0.3878
A_N_MM	17	152.3	<.0001	4.3175	0.0377
A_P_MMMD	18	153.4	<.0001	1.1358	0.2866
A_N_TT	19	153.5	<.0001	0.0507	0.8218

**Tabelle VI.20 : Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_DAYT_D	1	62.3244	<.0001	62.3244	<.0001
A_N_DIST	2	88.4146	<.0001	26.0903	<.0001
A_N_PUR	3	107.4	<.0001	19.0262	<.0001
HH_NR_L	4	121.8	<.0001	14.3861	0.0001
A_P_MM	5	135.7	<.0001	13.9121	0.0002
PLANNED	6	145.0	<.0001	9.2689	0.0023
A_DAYT	7	152.8	<.0001	7.8231	0.0052
WEEKEND	8	159.6	<.0001	6.8133	0.0090
A_EXP_D	9	164.2	<.0001	4.5868	0.0322
A_EXP	10	169.0	<.0001	4.7622	0.0291
ACC	11	172.8	<.0001	3.8523	0.0497
CHILDREN	12	177.1	<.0001	4.2625	0.0390
A_N_MMMD	13	180.0	<.0001	2.9080	0.0881
A_N_MM	14	183.6	<.0001	3.6210	0.0571
NR_A_SUMC	15	186.2	<.0001	2.5642	0.1093
A_P_MMMD	16	187.1	<.0001	0.9356	0.3334
A_P_TT	17	188.1	<.0001	0.9276	0.3355
A_P_DIST	18	188.9	<.0001	0.7904	0.3740
A_N_TT	19	189.1	<.0001	0.2304	0.6312

zum Vergleich!

## **Anhang VII : Erste Gesamtmodelle der semiparametrischen Modellschätzungen, Aktivitätsdauern**

Auf den folgenden Seiten wird in jeder der vier ausgewählten Kategorien ein erstes Gesamtmodell präsentiert. Aus der Liste der insgesamt 50 möglichen Einflußvariablen sind jeweils durch singuläre Variablenanalyse, Korrelationsmatrizen und Proportionalitätstests diejenigen bestimmt worden, die in einem ersten Gesamtmodell als Kovariablen verwendet werden können. Die Modellinformationen werden hier bewußt nicht näher gezeigt, sie sind ausführlich im Zusatzanhang XXI (PHREG AD Semiparametrische Modelle) zu sehen.

Um eine rasche Gegenüberstellung mit den Auswertungen der vollparametrischen Modellschätzungen der Aktivitätsdauern zu gewährleisten, sind diese Ergebnisse im Anhang VIII (LIFEREG AD Vollparametrische Modelle) zu sehen.

In einem weiteren Schritt wird ein Modell der zweiten Stufe berechnet, in dem nur noch die im ersten Gesamtmodell statistisch signifikanten Variablen enthalten sind. Die Ergebnisse des zweiten Modells sind im Zusatzanhang XXI (PHREG AD Semiparametrische Modelle) angeführt bzw. direkt im Kapitel 9.2.3 diskutiert.

## Kategorie Einkauf kurzfristig

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
<b>SEX</b>	0.25028	0.05641	19.6823	<b>&lt;.0001</b>	1.284
<b>LICENSE</b>	0.38008	0.06719	31.9949	<b>&lt;.0001</b>	1.462
HEAD_OHH	-0.08437	0.06378	1.7501	0.1859	0.919
WORKED	0.01367	0.05189	0.0694	0.7922	1.014
<b>SINGLE</b>	0.13184	0.05794	5.1769	<b>0.0229</b>	1.141
EDUC	0.03798	0.04440	0.7315	0.3924	1.039
AGE	0.00370	0.00206	3.2398	0.0719	1.004
<b>G</b>	0.05753	0.02014	8.1631	<b>0.0043</b>	1.059
<b>CAR</b>	-0.16529	0.06895	5.7470	<b>0.0165</b>	0.848
<b>PLANNED</b>	-0.36536	0.05852	38.9781	<b>&lt;.0001</b>	0.694
A_P_MM_D	0.13282	0.06945	3.6577	0.0558	1.142
A_N_MM_D	0.17801	0.06663	7.1377	<b>0.0075</b>	1.195
A_P_MMMD	-0.08866	0.06869	1.6659	0.1968	0.915
A_N_MMMD	-0.31480	0.06630	22.5460	<b>&lt;.0001</b>	0.730
<b>ACC</b>	-0.07968	0.03358	5.6284	<b>0.0177</b>	0.923
<b>A_EXP</b>	-0.00685	0.00068	102.7746	<b>&lt;.0001</b>	0.993
<b>A_DAYT</b>	0.09736	0.03491	7.7756	<b>0.0053</b>	1.102
A_P_DIST	-6.45E-6	5.78E-6	1.2435	0.2648	1.000
<b>A_P_TT</b>	-0.01664	0.00237	49.4191	<b>&lt;.0001</b>	0.983
A_N_DIST	2.49E-6	1.88E-6	1.7502	0.1858	1.000
<b>A_N_TT</b>	-0.00771	0.00182	18.0352	<b>&lt;.0001</b>	0.992
NR_T_SUM	-0.00031	0.00051	0.3763	0.5396	1.000
<b>A_P_PUR</b>	-0.03379	0.00675	25.0668	<b>&lt;.0001</b>	0.967
<b>NR_A_SUMC</b>	0.00968	0.00167	33.6948	<b>&lt;.0001</b>	1.010

---

## Kategorie Dienstleistung allgemein

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
CAR	0.02894	0.04740	0.3727	0.5415	1.029
<b>HH_MEMB</b>	0.06869	0.01885	13.2783	<b>0.0003</b>	1.071
ACC_D	0.03159	0.04431	0.5082	0.4759	1.032
<b>PLANNED</b>	-0.54299	0.06839	63.0403	<b>&lt;.0001</b>	0.581
A_EXP_D	0.10373	0.05661	3.3575	0.0669	1.109
<b>WEEKEND</b>	-0.24558	0.04033	37.0824	<b>&lt;.0001</b>	0.782
<b>A_DAYT_D</b>	-0.55941	0.04779	137.0067	<b>&lt;.0001</b>	0.572
<b>A_P_MM_D</b>	0.17695	0.07405	5.7094	<b>0.0169</b>	1.194
A_N_MM_D	-0.05642	0.07823	0.5201	0.4708	0.945
<b>A_P_DIST</b>	-3.75E-6	1.04E-6	12.9485	<b>0.0003</b>	1.000
A_N_TT	-0.00100	0.00054	3.3789	0.0660	0.999
<b>A_N_MM</b>	-0.16881	0.01244	184.0678	<b>&lt;.0001</b>	0.845
<b>NR_J_SUM</b>	0.00866	0.00085	105.0295	<b>&lt;.0001</b>	1.009
<b>A_P_PUR</b>	-0.03349	0.00838	15.9761	<b>&lt;.0001</b>	0.967
<b>A_N_PUR</b>	-0.08334	0.00929	80.5235	<b>&lt;.0001</b>	0.920

---

## Kategorie Freizeit allgemein

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
SEX	0.11084	0.11378	0.9490	0.3300	1.117
HEAD_OHH	0.14136	0.12712	1.2365	0.2661	1.152
WORKED	0.07169	0.09120	0.6180	0.4318	1.074
CAR_POSS	-0.09645	0.09851	0.9586	0.3275	0.908
AGE	-0.00373	0.00254	2.1559	0.1420	0.996
HH_MEMB	0.04417	0.04254	1.0783	0.2991	1.045
<b>A_EXP_D</b>	<b>-0.28271</b>	<b>0.08124</b>	<b>12.1106</b>	<b>0.0005</b>	<b>0.754</b>
A_P_MM_D	0.04935	0.13032	0.1434	0.7049	1.051
A_N_MM_D	0.08797	0.15458	0.3239	0.5693	1.092
ACC	-0.11340	0.07410	2.3419	0.1259	0.893
<b>A_P_TT</b>	<b>-0.01592</b>	<b>0.00453</b>	<b>12.3362</b>	<b>0.0004</b>	<b>0.984</b>
A_N_TT	-0.00322	0.00328	0.9644	0.3261	0.997
<b>A_P_MM</b>	<b>-0.06393</b>	<b>0.02635</b>	<b>5.8866</b>	<b>0.0153</b>	<b>0.938</b>
NR_T_SUM	0.00158	0.00087	3.3082	0.0689	1.002
A_P_PUR	-0.00529	0.01292	0.1677	0.6822	0.995
<b>A_N_PUR</b>	<b>-0.03786</b>	<b>0.01347</b>	<b>7.9020</b>	<b>0.0049</b>	<b>0.963</b>
NR_A_SUMC	-0.00488	0.00463	1.1097	0.2922	0.995

---

## Kategorie Treffen mit Freunden

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
<b>CHILDREN</b>	0.20891	0.10090	4.2869	<b>0.0384</b>	1.232
<b>HH_NR_L</b>	0.16725	0.05835	8.2165	<b>0.0042</b>	1.182
PLANNED	-0.27721	0.15071	3.3832	0.0659	0.758
<b>A_EXP_D</b>	0.50087	0.21351	5.5031	<b>0.0190</b>	1.650
WEEKEND	-0.11061	0.08728	1.6060	0.2051	0.895
<b>A_DAYT_D</b>	-0.66169	0.10915	36.7481	<b>&lt;.0001</b>	0.516
<b>ACC</b>	-0.13450	0.05894	5.2078	<b>0.0225</b>	0.874
A_P_TT	-0.00469	0.00285	2.7005	0.1003	0.995
A_N_TT	-0.00433	0.00279	2.4042	0.1210	0.996
<b>A_N_MM</b>	-0.06798	0.02562	7.0371	<b>0.0080</b>	0.934
<b>A_N_PUR</b>	-0.06210	0.01872	11.0063	<b>0.0009</b>	0.940
<b>NR_A_SUMC</b>	-0.01218	0.00317	14.7609	<b>0.0001</b>	0.988

---

## **Anhang VIII : Erste Gesamtmodelle der vollparametrischen Modellschätzungen, Aktivitätsdauern**

Auf den folgenden Seiten werden in jeder der vier ausgewählten Kategorien die Modellinformationen und ein erstes Gesamtmodell präsentiert. Aus der Liste der möglichen Einflußvariablen sind nach einer Einzelanalyse (siehe semiparametrische Modellschätzungen) jeweils mit Hilfe von Korrelationsmatrizen diejenigen bestimmt worden, die in einem ersten Gesamtmodell als Kovariablen verwendet werden können. Durch Proportionalitätstest sind sie ein weiteres Mal untersucht und dann in ein erstes Gesamtmodell integriert worden.

Alle notwendigen Überprüfungen in bezug auf die gewählte Verteilungsannahme sehen Sie im Zusatzanhang XXII (LIFEREG AD Vollparametrische Modelle).

Um einen guten Vergleich mit den Auswertungen der semiparametrischen Modelle zu erhalten, sind diese Ergebnisse der ersten Gesamtmodelle für die entsprechenden vier Kategorien im Anhang VII (PHREG AD Erste Gesamtmodelle der semiparametrischen Auswertung) zu sehen.

In einem weiteren Schritt wird ein Modell der zweiten Stufe berechnet, in dem nur noch die statistisch signifikanten Variablen des ersten Gesamtmodells enthalten sind. Die Ergebnisse des Modells zweiter Stufe sind im Zusatzanhang XXII (LIFEREG AD Vollparametrische Modelle) bzw. direkt im Kapitel 9.2.4 zu sehen.

## Kategorie Einkauf kurzfristig : Loglogistisches Modell mit Kovariablen

---

### Modell Information

Data Set	WORK.DUMMY	
Dependent Variable	Log(a_durst)	Aktivitätsdauer in Stunden
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	3221	
Noncensored Values	3221	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Name of Distribution	LLOGISTC	
Log Likelihood	-3976.89356	

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	<b>-1.38525</b>	0.17411	63.3028	<b>&lt;.0001</b>
<b>SEX</b>	<b>-0.23171</b>	0.04602	25.3472	<b>&lt;.0001</b>
<b>LICENSE</b>	<b>-0.27255</b>	0.04992	29.8081	<b>&lt;.0001</b>
HEAD_OHH	-0.03521	0.05161	0.4656	0.4950
WORKED	-0.04474	0.04081	1.2015	0.2730
<i>SINGLE</i>	<i>-0.08013</i>	0.04735	2.8635	<i>0.0906</i>
EDUC	-0.04812	0.03549	1.8385	0.1751
AGE	0.00013	0.00160	0.0070	0.9332
<b>G</b>	<b>-0.03643</b>	0.01560	5.4561	<b>0.0195</b>
<b>CAR</b>	<b>0.11011</b>	0.05201	4.4823	<b>0.0342</b>
<b>PLANNED</b>	<b>0.17826</b>	0.04551	15.3449	<b>&lt;.0001</b>
<b>A_P_MM_D</b>	<b>-0.13789</b>	0.05526	6.2257	<b>0.0126</b>
<b>A_N_MM_D</b>	<b>-0.19967</b>	0.05223	14.6129	<b>0.0001</b>
<b>A_P_MMMD</b>	<b>0.18337</b>	0.06026	9.2599	<b>0.0023</b>
<b>A_N_MMMD</b>	<b>0.11604</b>	0.05875	3.9008	<b>0.0483</b>
<b>ACC</b>	<b>0.09928</b>	0.02783	12.7208	<b>0.0004</b>
<b>A_EXP</b>	<b>0.00690</b>	0.00051	184.2858	<b>&lt;.0001</b>
<b>A_DAYT</b>	<b>-0.10719</b>	0.02716	15.5744	<b>&lt;.0001</b>
<b>A_P_DIST</b>	<b>-0.00001</b>	5.25E-6	5.4066	<b>0.0201</b>
<b>A_P_TT</b>	<b>0.01611</b>	0.00177	82.5101	<b>&lt;.0001</b>
<b>A_N_DIST</b>	<b>-5.96E-7</b>	1.45E-6	0.1689	0.6811
<b>A_N_TT</b>	<b>0.00439</b>	0.00129	11.5839	<b>0.0007</b>
NR_T_SUM	0.00004	0.00040	0.0084	0.9270
<b>A_P_PUR</b>	<b>0.01876</b>	0.00545	11.8502	<b>0.0006</b>
<b>NR_A_SUMC</b>	<b>-0.00960</b>	0.00137	49.2858	<b>&lt;.0001</b>
SCALE	0.46347	0.00685		

---

n = 3221; Log-Likelihood = - 3976.89

## Kategorie Freizeit allgemein: Standardisierte Gammaverteilung mit Kovariablen

---

### Model Information

Data Set	WORK.DUMMY8	
Dependent Variable	Log(a_durst)	Aktivitätsdauer in Stunden
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	2985	
Noncensored Values	2985	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Missing Values	2	
Name of Distribution	GAMMA	
Log Likelihood	-4136.298782	

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	-1.15798	0.11296	105.0892	<.0001
CAR	-0.01773	0.04078	0.1890	0.6637
<b>HH_MEMB</b>	-0.06601	0.01616	16.6922	<.0001
ACC_D	-0.03259	0.03795	0.7374	0.3905
<b>PLANNED</b>	0.49399	0.05852	71.2478	<.0001
<b>A_EXP_D</b>	-0.09486	0.04828	3.8598	<b>0.0495</b>
<b>WEEKEND</b>	0.22978	0.03459	44.1405	<.0001
<b>A_DAYT_D</b>	0.51677	0.04039	163.6835	<.0001
<b>A_P_MM_D</b>	-0.15283	0.06348	5.7952	<b>0.0161</b>
A_N_MM_D	0.05836	0.06707	0.7571	0.3842
<b>A_P_DIST</b>	3.19E-6	8.57E-7	13.9023	<b>0.0002</b>
A_N_TT	0.00092	0.00047	3.8905	<b>0.0486</b>
<b>A_N_MM</b>	0.15136	0.01048	208.7042	<.0001
<b>NR_J_SUM</b>	-0.00759	0.00071	113.6275	<.0001
<b>A_P_PUR</b>	0.02843	0.00718	15.6844	<.0001
<b>A_N_PUR</b>	0.07790	0.00800	94.8268	<.0001
SCALE	0.86000	0		
SHAPE	0.86000	0		

---

n = 2985; Log-Likelihood = - 4136.30

### Lagrange Multiplier Statistics

Variable	Chi-Square	Pr > ChiSq
Scale	1.4802	0.2237
Shape1	23.5051	<.0001

---

## Kategorie Dienstleistung allgemein: Lognormalverteilung mit Kovariablen

---

### Modell Information

Data Set	WORK.DUMMY2	
Dependent Variable	Log(a_durst)	Aktivitätsdauer in Stunden
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	896	
Noncensored Values	896	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Name of Distribution	LNORMAL	
Log Likelihood	-1222.275174	

---



---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	-2.07778	0.28773	52.1462	<b>&lt;.0001</b>
<b>SEX</b>	-0.23320	0.10274	5.1524	<b>0.0232</b>
HEAD_OHH	0.05722	0.11495	0.2478	0.6186
WORKED	-0.12779	0.08679	2.1683	0.1409
CAR_POSS	0.11308	0.08925	1.6053	0.2051
AGE	0.00306	0.00229	1.7782	0.1824
HH_MEMB	-0.00097	0.03882	0.0006	0.9801
<b>A_EXP_D</b>	0.29967	0.07538	15.8032	<b>&lt;.0001</b>
A_P_MM_D	-0.04992	0.12272	0.1655	0.6842
A_N_MM_D	-0.09681	0.14299	0.4584	0.4984
ACC	0.04221	0.07470	0.3193	0.5720
<b>A_P_TT</b>	0.01999	0.00355	31.6557	<b>&lt;.0001</b>
A_N_TT	0.00441	0.00270	2.6794	0.1017
<b>A_P_MM</b>	0.06562	0.02453	7.1568	<b>0.0075</b>
<b>NR_T_SUM</b>	-0.00157	0.00072	4.7216	<b>0.0298</b>
A_P_PUR	-0.00003	0.01187	0.0000	0.9982
A_N_PUR	0.01016	0.01274	0.6363	0.4250
NR_A_SUMC	0.00048	0.00425	0.0125	0.9109
SCALE	0.94668	0.02236		

---

n = 896; Log-Likelihood = - 1222.28

## Kategorie Treffen mit Freunden: Weibullverteilungsmodell mit Kovariablen

---

### Modell Information

Data Set	WORK.DUMMY11	
Dependent Variable	Log(a_durst)	Aktivitätsdauer in Stunden
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	669	
Noncensored Values	669	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Missing Values	7	
Name of Distribution	WEIBULL	
Log Likelihood	-908.36281	

---



---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
<b>INTERCEPT</b>	-0.43166	0.19567	4.8665	<b>0.0274</b>
<b>CHILDREN</b>	-0.16202	0.08132	3.9693	<b>0.0463</b>
<b>HH_NR_L</b>	-0.13083	0.04678	7.8226	<b>0.0052</b>
<b>PLANNED</b>	0.28345	0.12006	5.5743	<b>0.0182</b>
<b>A_EXP_D</b>	-0.45193	0.17129	6.9612	<b>0.0083</b>
WEEKEND	0.09128	0.07034	1.6842	0.1944
<b>A_DAYT_D</b>	0.55969	0.08740	41.0032	<b>&lt;.0001</b>
<b>ACC</b>	0.10893	0.04734	5.2956	<b>0.0214</b>
<i>A_P_TT</i>	0.00428	0.00228	3.5011	<i>0.0613</i>
<i>A_N_TT</i>	0.00298	0.00225	1.7532	0.1855
<b>A_N_MM</b>	0.04581	0.02058	4.9528	<b>0.0260</b>
<b>A_N_PUR</b>	0.04747	0.01516	9.8019	<b>0.0017</b>
<b>NR_A_SUMC</b>	0.00867	0.00253	11.7739	<b>0.0006</b>
SCALE	0.80815	0.02445		

---

n = 669; Log-Likelihood = - 908.36

## Anhang IX : Übersichtstabellen signifikante Variablen, Aktivitätsdauern

In diesem Anhang werden zum Vergleich die in den nonparametrischen stufenweisen Chi-Quadrat-Tests bzw. die in den semiparametrischen und vollparametrischen Modellschätzungen verwendeten möglichen Einflußvariablen für alle vier ausgewählten Aktivitätskategorien Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden tabellarisch dargestellt.

In der ersten Tabelle sind die Ergebnisse der stufenweisen Chi-Quadrat-Tests (*Forward Stepwise Sequence of Chi-Squares*) der Variablen des ersten Gesamtmodells für den jeweilig günstigeren Test (entweder Log-Rank oder Wilcoxon Test) für alle vier Kategorien dargestellt. Für die Kategorien Einkauf kurzfristig und Dienstleistung allgemein werden daher die signifikanten Variablen des Wilcoxon Tests präsentiert, für die beiden Freizeitkategorien (Freizeit allgemein und Treffen mit Freunden) die des Log-Rank-Tests. In der zweiten und dritten Tabelle sind übersichtlich die jeweils in der zweiten Modellstufe signifikanten Variablen, sowohl marginal signifikant im 90%-Konfidenzintervall (Signifikanzwert unter  $\alpha = 0.10$ ) mit \*\*, als auch stark signifikant im 95%-Konfidenzintervall (Signifikanzwert unter  $\alpha = 0.05$ ) mit \* gekennzeichnet. Die exakten Werte und andere Ergebnisse sieht man im Zusatzanhang XIX (LIFETEST AD Logrank und Wilcoxontest), im Zusatzanhang XXI (PHREG AD Semiparametrische Modelle) sowie im Zusatzanhang XXII (LIFEREG AD Vollparametrische Modelle).

Die hier dargestellten Ergebnisse gelten nur für die Einflüsse der Variablen auf die Aktivitätsdauern, nicht aber auf die Intervalldauern zwischen zwei Aktivitäten der gleichen Kategorie. Diese Ergebnisse sind im Anhang XIV (I Übersichtstabelle signifikante Variablen) dargestellt.

## Übersicht der Signifikanz der möglichen Einflußvariablen, nonparametrische Modellschätzung

Variable	Einkauf kurzfristig	Freizeit allgemein	Dienstl. allgemein	Treffen mit Freunden	Dummy	Variablenname
SEX	*		*		X	Geschlecht
AGE_**					X	Altersgruppe
AGE						Alter
YOUNG					X	Unter 50 Jahren j/n
LICENSE	*				X	Führerscheinbesitz j/n
HEAD_O_HH					X	Haushaltsvorstand j/n
WORKED			*		X	arbeitend / nicht arbeitend
CHILDREN				*	X	Kinder im Haushalt j/n
SINGLE	*				X	Single bzw. alleine lebend j/n
CAR_POSS					X	Autoverfügbarkeit j/n
EDUC					X	Ausbildung während der Umfrage j/n
G*_D					X	Gesellschaftsgruppe Nr. * j/n
EDUC_YR						Anzahl der Ausbildungsjahre
G	*					Nummer der Gesellschaftsgruppe
CAR	*					Anzahl der Autos im Haushalt
HH_NR_L				*		Anzahl der Führerscheine im HH
HH_MEMB		*				Anzahl der Haushaltsmitglieder
ACC_D					X	Begleitung j/n
PLANNED	*	*		**	X	Aktivität geplant j/n
A_EXP_D		*	*	*	X	Ausgaben am Aktivitätssort j/n
WEEKEND		*			X	Wochenende j/n
A_DAYT_D		*		*	X	Tageszeit (vor 13 Uhr j/n)
A_P_MM_D	*				X	Unabhängiges VM zur Aktivität
A_N_MM_D	*				X	Unabhängiges VM nach der A
A_P_MMMD	*				X	Unabhängiges motorisiertes VM zur A
A_N_MMMD	*				X	Unabhängiges motorisiertes VM von der A
ACC	*			*		Anzahl der Begleitpersonen
A_EXP	*			*		Höhe der Ausgaben am Aktivitätssort
A_DAYT	*			*		Tageszeit (vorm/nachm/abends)
A_P_DIST	*	*		**		Distanz zum Aktivitätssort
A_P_TT	*		*			Reisezeit zum Aktivitätssort
A_N_DIST				*		Distanz zum nächsten Aktivitätssort
A_N_TT	*	**	*			Reisezeit zum nächsten Aktivitätssort
A_P_MM			*	*		Verkehrsmittel zum Aktivitätssort
A_N_MM		*		*		Verkehrsmittel vom Aktivitätssort
NR_T_SUM	*		*			Anzahl der Wege gesamt
NR_J_SUM		*				Anzahl der Reisen gesamt
A_P_PUR	*	*				Vorangegangene Aktivitätskategorie
A_N_PUR		*		*		Nachfolgende Aktivitätskategorie
NR_A_SUMC	*			*		Anzahl der A in dieser Kategorie

Ergebnisse aus den stufenförmigen Chi-Quadrat-tests der Variablen aus dem ersten Gesamtmodell mit *PROC LIFETEST*, siehe auch Anhang VI und Zusatzanhang XIX (LIFETEST AD Logrank und Wilcoxonstest).

Markierung mit \* : Signifikant im 95 % - Konfidenzintervall bzw. Signifikanzwert unter  $\alpha = 0.05$

Markierung mit \*\* : Signifikant im 90 % - Konfidenzintervall bzw. Signifikanzwert unter  $\alpha = 0.10$

## Übersicht der Signifikanz der möglichen Einflußvariablen, semiparametrische Modellschätzung

Variable	Einkauf kurzfristig	Freizeit allgemein	Dienstl. allgemein	Treffen mit Freunden	Dummy	Variablenname
SEX	*		*		X	Geschlecht
AGE_**					X	Altersgruppe
AGE			*			Alter
YOUNG					X	Unter 50 Jahren j/n
LICENSE	*				X	Führerscheinbesitz j/n
HEAD_O_HH					X	Haushaltsvorstand j/n
WORKED			**		X	arbeitend / nicht arbeitend
CHILDREN				*	X	Kinder im Haushalt j/n
SINGLE	**				X	Single bzw. alleine lebend j/n
CAR_POSS					X	Autoverfügbarkeit j/n
EDUC					X	Ausbildung während der Umfrage j/n
G*_D					X	Gesellschaftsgruppe Nr. * j/n
EDUC_YR						Anzahl der Ausbildungsjahre
G	*					Nummer der Gesellschaftsgruppe
CAR	*					Anzahl der Autos im Haushalt
HH_NR_L				*		Anzahl der Führerscheine im HH
HH_MEMB		*				Anzahl der Haushaltsmitglieder
ACC_D					X	Begleitung j/n
PLANNED	*	*		**	X	Aktivität geplant j/n
A_EXP_D		**	*	*	X	Ausgaben am Aktivitätssort j/n
WEEKEND		*			X	Wochenende j/n
A_DAYT_D		*		*	X	Tageszeit (vor 13 Uhr j/n)
A_P_MM_D		*			X	Unabhängiges VM zur Aktivität
A_N_MM_D	*				X	Unabhängiges VM nach der A
A_P_MMMD					X	Unabhängiges motorisiertes VM zur A
A_N_MMMD	*				X	Unabhängiges motorisiertes VM von der A
ACC	*		**	*		Anzahl der Begleitpersonen
A_EXP	*					Höhe der Ausgaben am Aktivitätssort
A_DAYT	*					Tageszeit (vorm/nachm/abends)
A_P_DIST		*				Distanz zum Aktivitätssort
A_P_TT	*		*	*		Reisezeit zum Aktivitätssort
A_N_DIST						Distanz zum nächsten Aktivitätssort
A_N_TT	*	**				Reisezeit zum nächsten Aktivitätssort
A_P_MM			*			Verkehrsmittel zum Aktivitätssort
A_N_MM				*		Verkehrsmittel vom Aktivitätssort
NR_T_SUM						Anzahl der Wege gesamt
NR_J_SUM		*				Anzahl der Reisen gesamt
A_P_PUR	*	*				Vorangegangene Aktivitätskategorie
A_N_PUR		*	*	*		Nachfolgende Aktivitätskategorie
NR_A_SUMC	*			*		Anzahl der A in dieser Kategorie

Ergebnisse aus den Gesamtmodellen der zweiten Stufe der semiparametrischen Auswertungen mit *PROC PHREG*, siehe auch Zusatzanhang XXI (PHREG AD Semiparametrische Modelle).

Markierung mit \* : Signifikant im 95 % - Konfidenzintervall bzw. Signifikanzwert unter  $\alpha = 0.05$

Markierung mit \*\* : Signifikant im 90 % - Konfidenzintervall bzw. Signifikanzwert unter  $\alpha = 0.10$

## Übersicht der Signifikanz der möglichen Einflußvariablen, vollparametrische Modellschätzung

Variable	Einkauf kurzfristig	Freizeit allgemein	Dienstl. allgemein	Treffen mit Freunden	Dummy	Variablenname
SEX	*		*		X	Geschlecht
AGE_**					X	Altersgruppe
AGE						Alter
YOUNG					X	Unter 50 Jahren j/n
LICENSE	*				X	Führerscheinbesitz j/n
HEAD_O_HH					X	Haushaltsvorstand j/n
WORKED			**		X	arbeitend / nicht arbeitend
CHILDREN				*	X	Kinder im Haushalt j/n
SINGLE					X	Single bzw. alleine lebend j/n
CAR_POSS					X	Autoverfügbarkeit j/n
EDUC					X	Ausbildung während der Umfrage j/n
G*_D					X	Gesellschaftsgruppe Nr. * j/n
EDUC_YR						Anzahl der Ausbildungsjahre
G	*					Nummer der Gesellschaftsgruppe
CAR	*					Anzahl der Autos im Haushalt
HH_NR_L				*		Anzahl der Führerscheine im HH
HH_MEMB		*				Anzahl der Haushaltsmitglieder
ACC_D					X	Begleitung j/n
PLANNED	*	*		*	X	Aktivität geplant j/n
A_EXP_D		**	*	*	X	Ausgaben am Aktivitätsort j/n
WEEKEND		*			X	Wochenende j/n
A_DAYT_D		*		*	X	Tageszeit (vor 13 Uhr j/n)
A_P_MM_D	*	*			X	Unabhängiges VM zur Aktivität
A_N_MM_D	*				X	Unabhängiges VM nach der A
A_P_MMMD	*				X	Unabhängiges motorisiertes VM zur A
A_N_MMMD	*				X	Unabhängiges motorisiertes VM von der A
ACC	*			*		Anzahl der Begleitpersonen
A_EXP	*					Höhe der Ausgaben am Aktivitätsort
A_DAYT	*					Tageszeit (vorm/nachm/abends)
A_P_DIST	*	*				Distanz zum Aktivitätsort
A_P_TT	*		*	*		Reisezeit zum Aktivitätsort
A_N_DIST						Distanz zum nächsten Aktivitätsort
A_N_TT	*	**	*			Reisezeit zum nächsten Aktivitätsort
A_P_MM			*			Verkehrsmittel zum Aktivitätsort
A_N_MM		*		*		Verkehrsmittel vom Aktivitätsort
NR_T_SUM			*			Anzahl der Wege gesamt
NR_J_SUM		*				Anzahl der Reisen gesamt
A_P_PUR	*	*				Vorangegangene Aktivitätskategorie
A_N_PUR		*		*		Nachfolgende Aktivitätskategorie
NR_A_SUMC	*			*		Anzahl der A in dieser Kategorie

Ergebnisse aus den Gesamtmodellen der zweiten Stufe der vollparametrischen Auswertungen mit *PROC LIFEREG*, siehe auch Zusatzanhang XXII (LIFEREG AD Vollparametrische Modelle).

Markierung mit \* : Signifikant im 95 % - Konfidenzintervall bzw. Signifikanzwert unter  $\alpha = 0.05$

Markierung mit \*\* : Signifikant im 90 % - Konfidenzintervall bzw. Signifikanzwert unter  $\alpha = 0.10$

## Vergleich der einzelnen parametrischen Modellschätzungen

Variable	Einkauf kurzfristig			Freizeit allgemein			Dienstl. allgemein			Treffen mit Freunden		
	non	semi	voll	non	semi	voll	non	semi	voll	non	semi	voll
SEX	*	*	*				*	*	*			
AGE_**												
AGE								*				
YOUNG												
LICENSE	*	*	*									
HEAD_O_HH												
WORKED							*	**	**			
CHILDREN										*	*	*
SINGLE	*	**										
CAR_POSS												
EDUC												
G*_D												
EDUC_YR												
G	*	*	*									
CAR	*	*	*									
HH_NR_L										*	*	*
HH_MEMB				*	*	*						
ACC_D												
PLANNED	*	*	*	*	*	*				**	**	*
A_EXP_D				*	**	**	*	*	*	*	*	*
WEEKEND				*	*	*						
A_DAYT_D				*	*	*				*	*	*
A_P_MM_D	*		*		*	*						
A_N_MM_D	*	*	*									
A_P_MMMD	*		*									
A_N_MMMD	*	*	*									
ACC	*	*	*					**		*	*	*
A_EXP	*	*	*							*		
A_DAYT	*	*	*							*		
A_P_DIST	*		*	*	*	*				**		
A_P_TT	*	*	*				*	*	*		*	*
A_N_DIST										*		
A_N_TT	*	*	*	**	**	**	*	*	*			
A_P_MM							*	*	*	*	*	*
A_N_MM				*		*				*	*	*
NR_T_SUM	*						*		*			
NR_J_SUM				*	*	*						
A_P_PUR	*	*	*	*	*	*						
A_N_PUR				*	*	*		*		*	*	*
NR_A_SUMC	*	*	*					*		*	*	*

## Anhang X : Hazard- und Survivalkurven, Intervall dauern

In den Abbildungen X.1 bis X.13 werden für alle Aktivitätskategorien die geglätteten Hazardkurven dargestellt. Im Kapitel 9.3.2.1 sind die Hazard- und Survivalkurven für die Intervall dauern in jeweils einer einzigen Grafik zu sehen. In den Abbildungen X.14 bis X.29 sind die geglätteten und uneglätteten Hazardkurven der Altersunterschiede und Geschlechtsunterschiede für die vier ausgewählten Kategorien Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden abgebildet, die teilweise im Kapitel 9.3.2.2 zu sehen sind. Anschließend in den Abbildungen X.30 bis X.37 sind noch die Survivalkurven der vier Kategorien dargestellt.

Abbildung X.1 : Geglättete Hazardkurve aller Aktivitäten

---

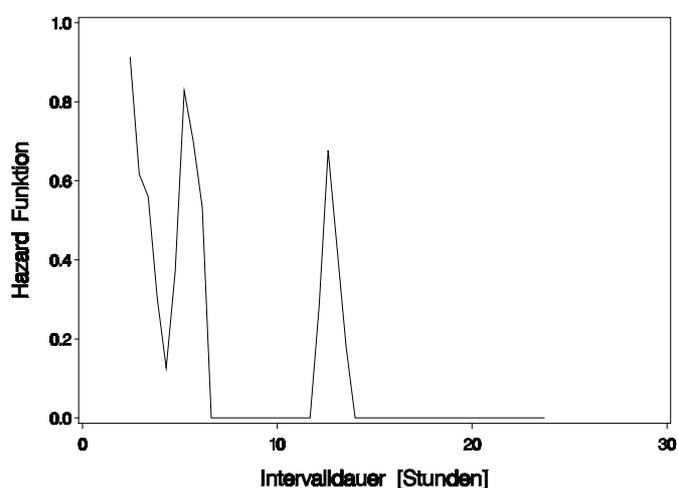


Abbildung X.2 : Geglättete Hazardkurve, Arbeit

---

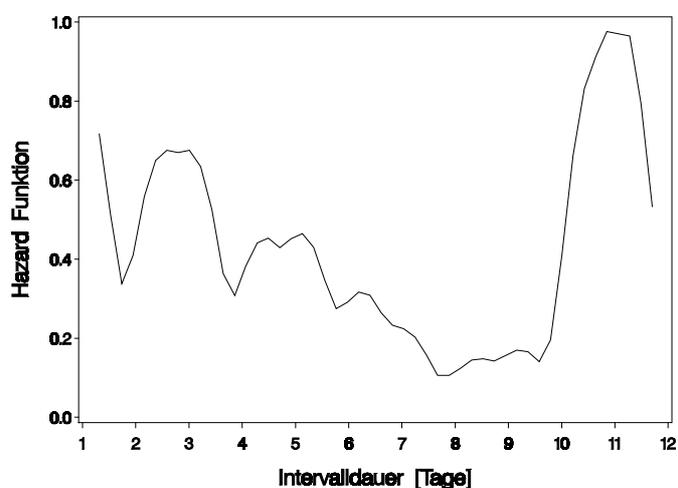


Abbildung X.3 : Geglättete Hazardkurve, Einkauf langfristig, Intervalldauer bis 14 Tage

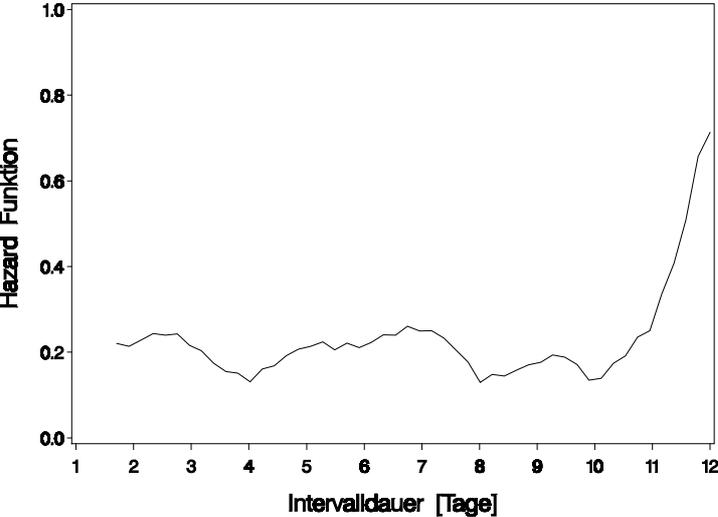


Abbildung X.4a : Geglättete Hazardkurve, Einkauf kurzfristig, Intervalldauer bis 14 Tage

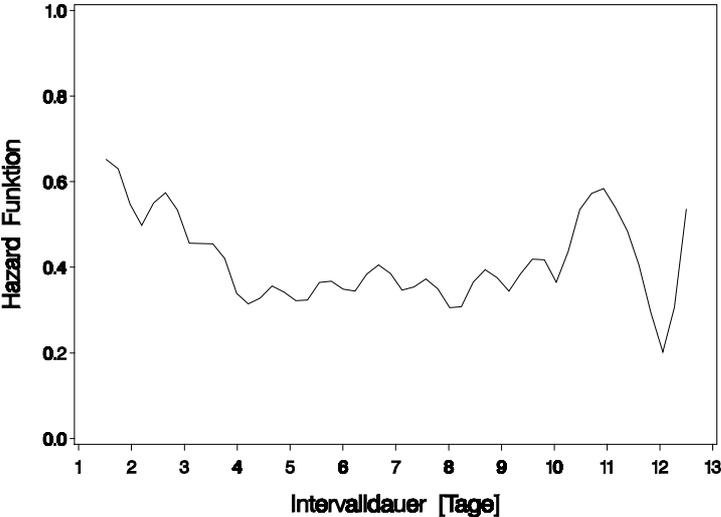


Abbildung X.4b : Geglättete Hazardkurve, Einkauf kurzfristig, ursprünglicher Datensatz

---

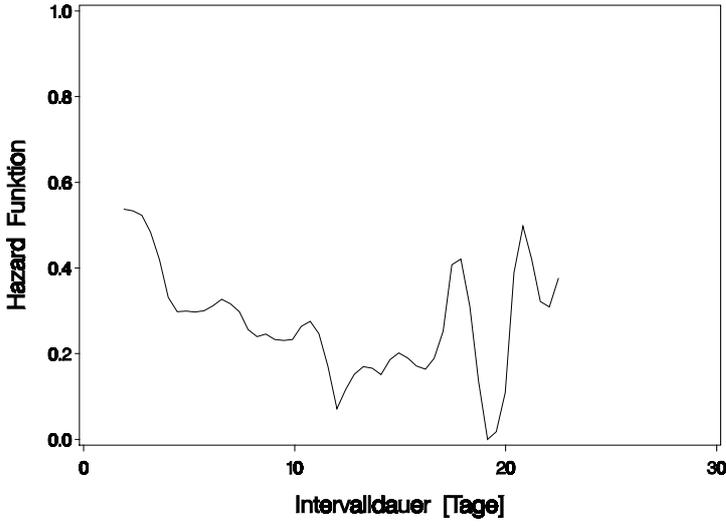


Abbildung X.5 : Geglättete Hazardkurve, Freizeit allgemein

---

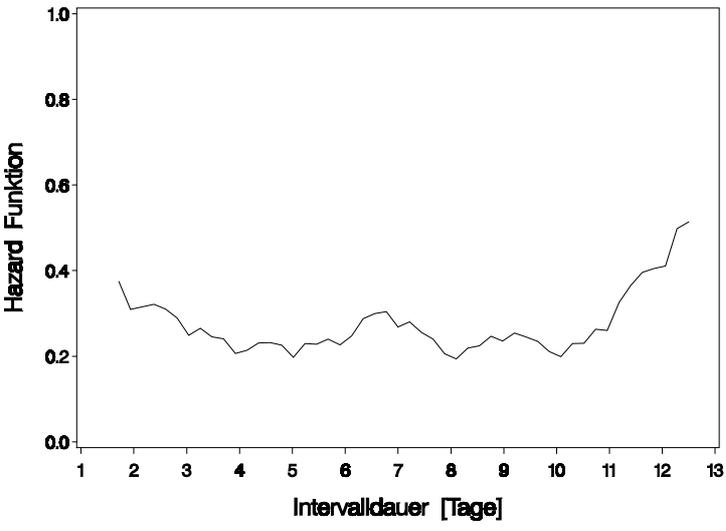


Abbildung X.6 : Geglättete Hazardkurve, Dienstleistung allgemein

---

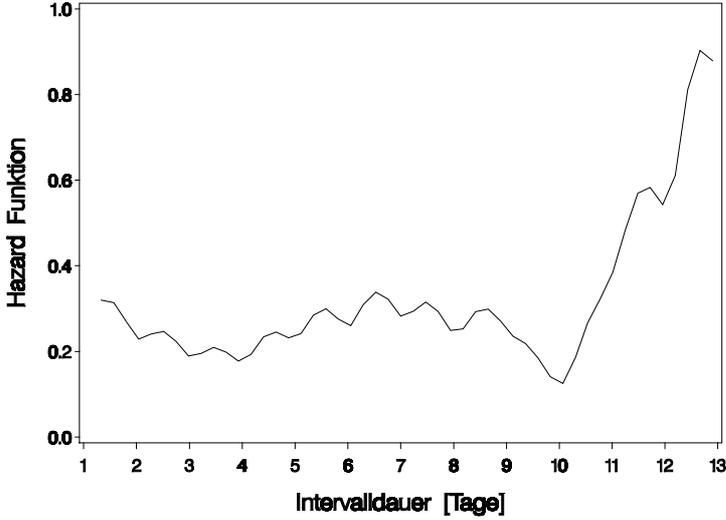


Abbildung X.7 : Geglättete Hazardkurve, Erledigungen

---

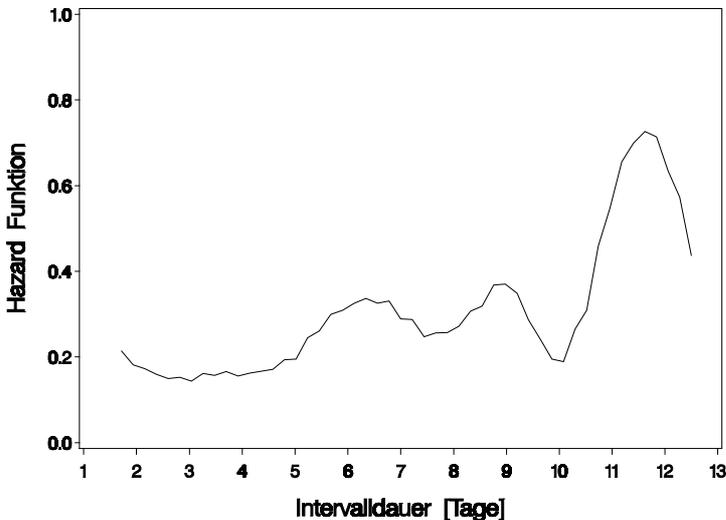


Abbildung X.8 : Geglättete Hazardkurve, Besuch bei Verwandten

---

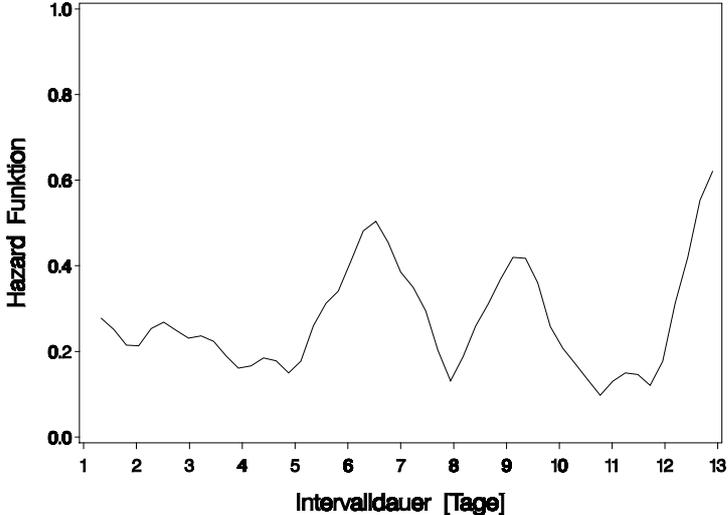


Abbildung X.9 : Geglättete Hazardkurve, Treffen mit Freunden

---

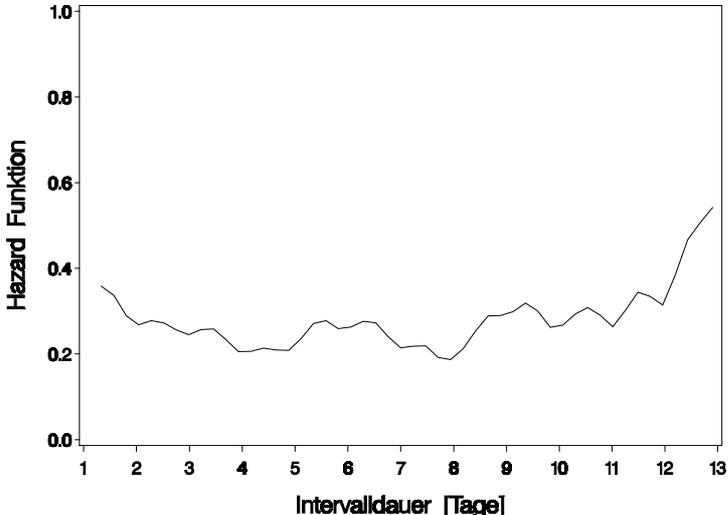


Abbildung X.10 : Geglättete Hazardkurve, Verein, Initiative, Kirche

---

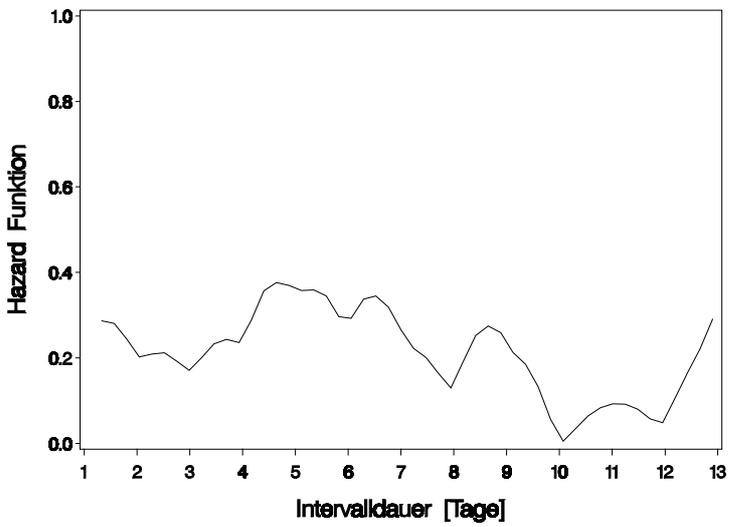


Abbildung X.11 : Geglättete Hazardkurve, Sport (aktiv)

---

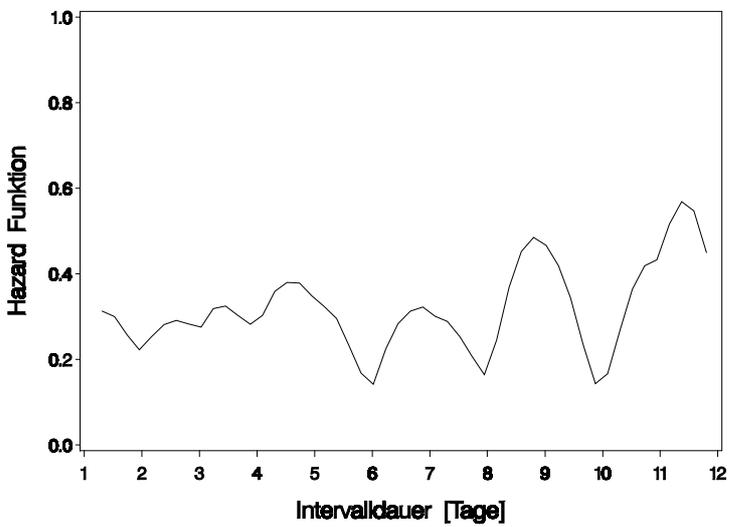


Abbildung X.12 : Geglättete Hazardkurve, Spaziergang / -fahrt

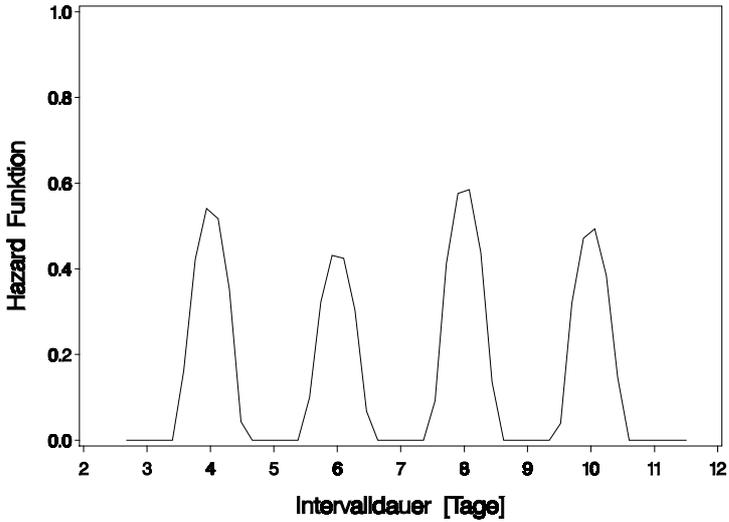


Abbildung X.13 : Geglättete Hazardkurve, Lokal, Kino, Theater

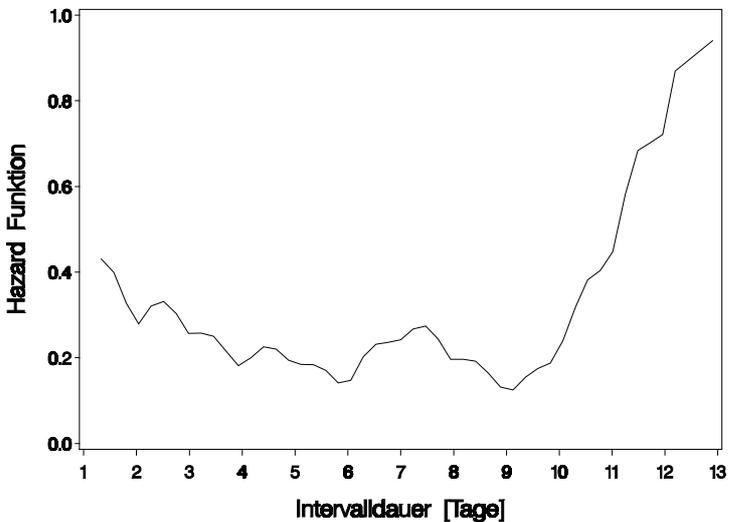


Abbildung X.14 : Ungeglättete Hazardkurve, Einkauf kurzfristig, Altersschichtung

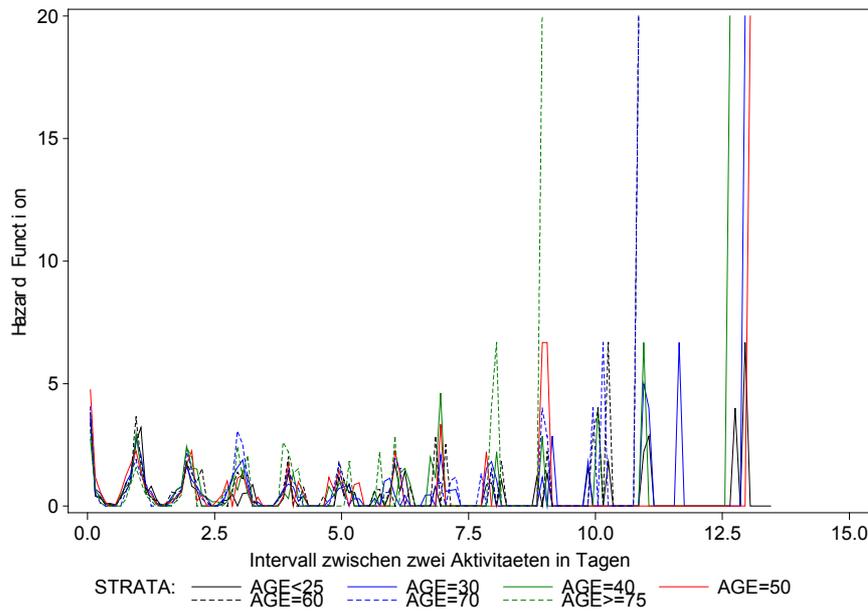
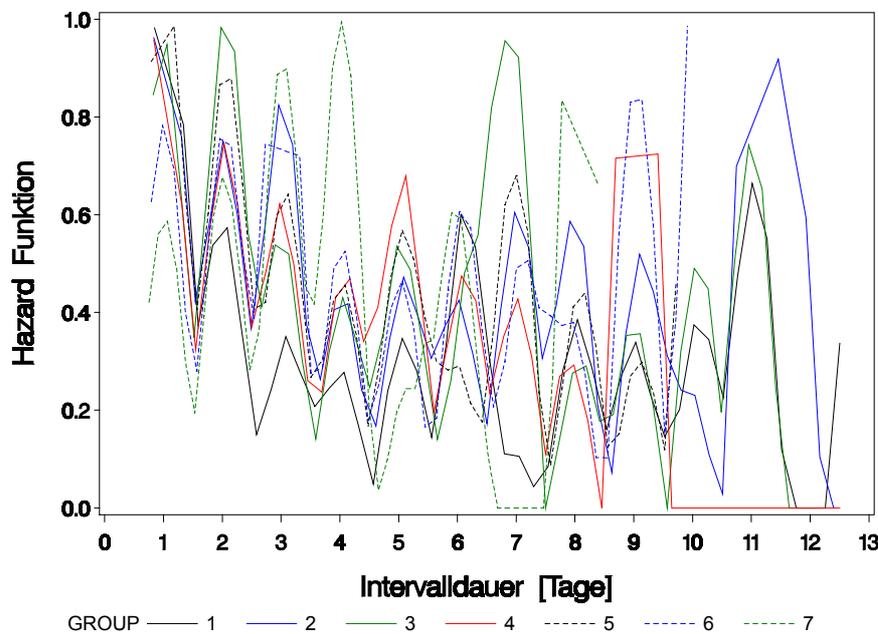


Abbildung X.15 : Geglättete Hazardkurve, Einkauf kurzfristig, Altersschichtung



In den folgenden Abbildungen der Geschlechtsunterschiede sind jeweils die grünen Hazard- bzw. Survivalkurve die der weiblichen Umfrageteilnehmer, die blauen die der männlichen Befragten.

Abbildung X.16 : Ungeglättete Hazardkurve, Einkauf kurzfristig, Geschlechtsunterschiede

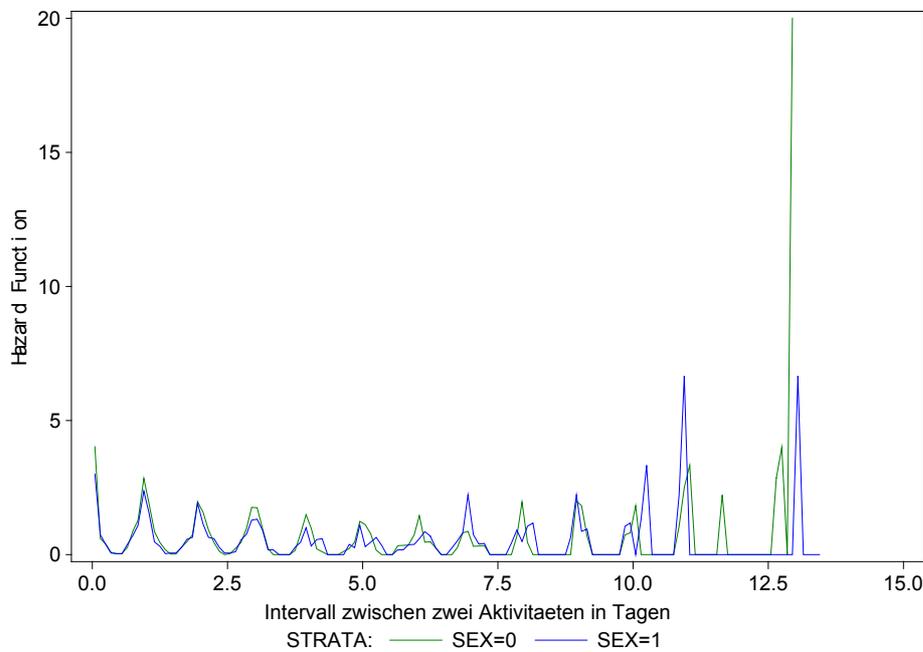


Abbildung X.17 : Geglättete Hazardkurve, Einkauf kurzfristig, Geschlechtsunterschiede

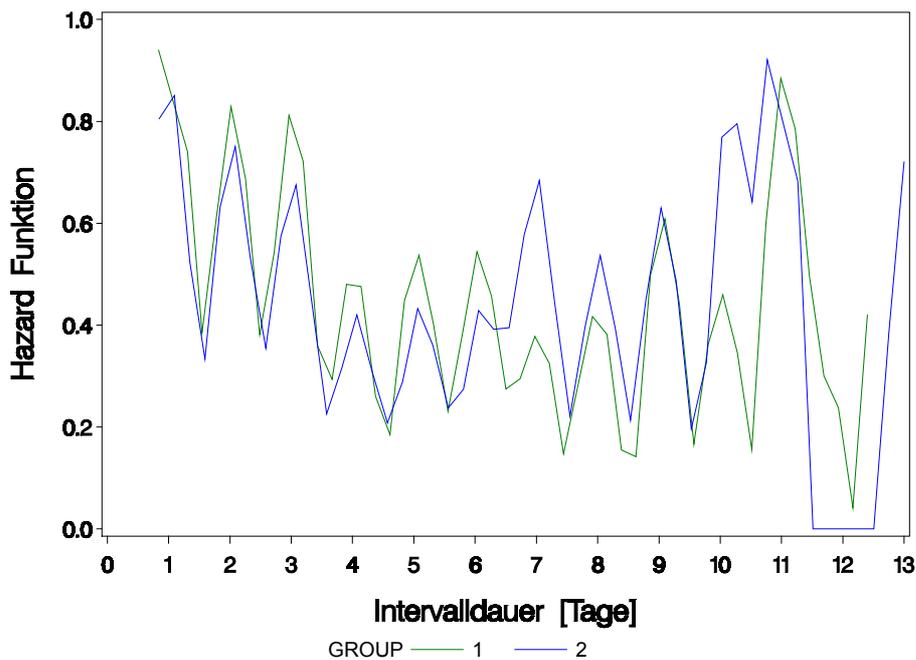


Abbildung X.18 : Ungeglättete Hazardkurve, Freizeit allgemein, Altersschichtung

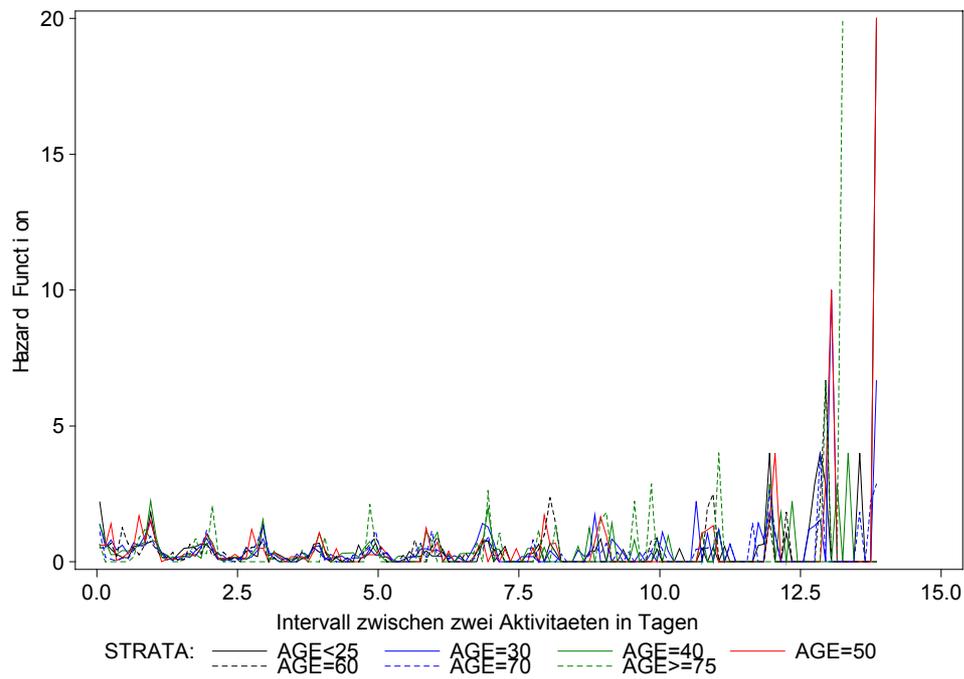


Abbildung X.19 : Geglättete Hazardkurve, Freizeit allgemein, Altersschichtung

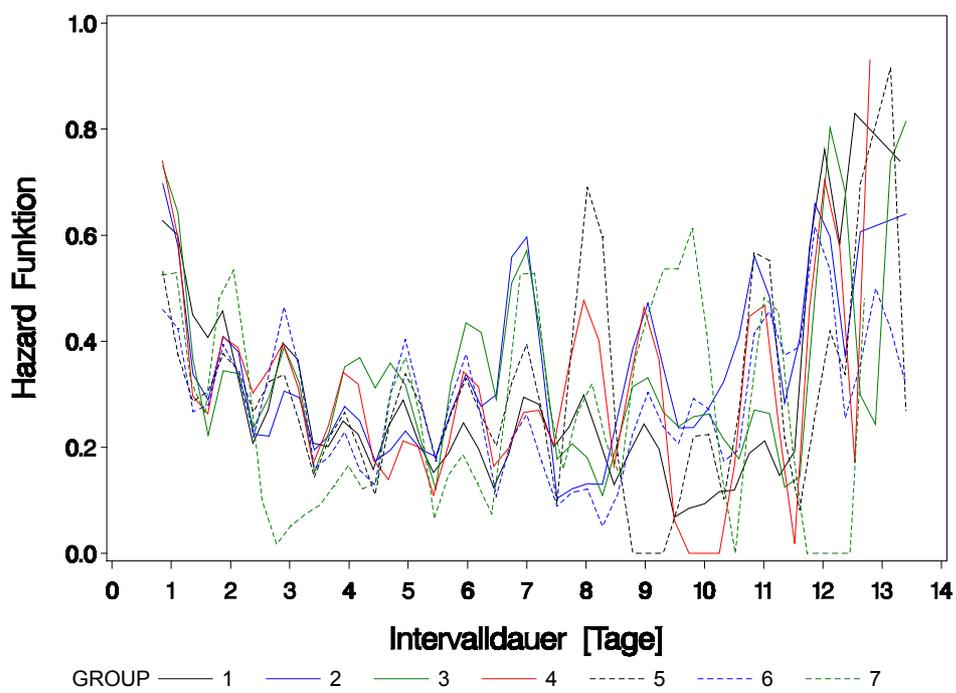


Abbildung X.20 : Ungeglättete Hazardkurve, Freizeit allgemein, Geschlechtsunterschiede

---

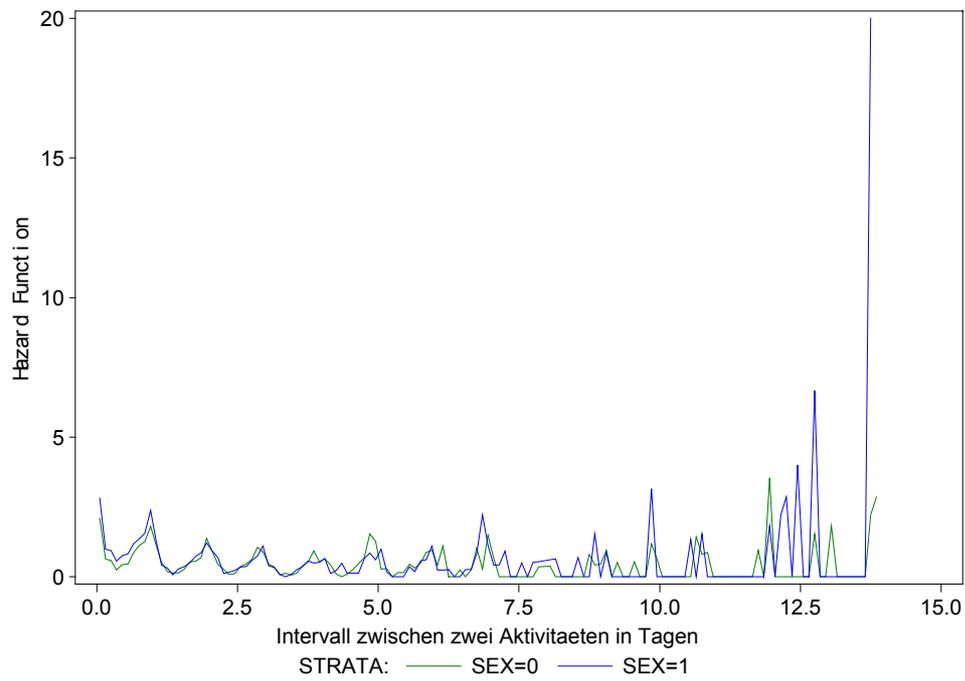


Abbildung X.21 : Geglättete Hazardkurve, Freizeit allgemein, Geschlechtsunterschiede

---

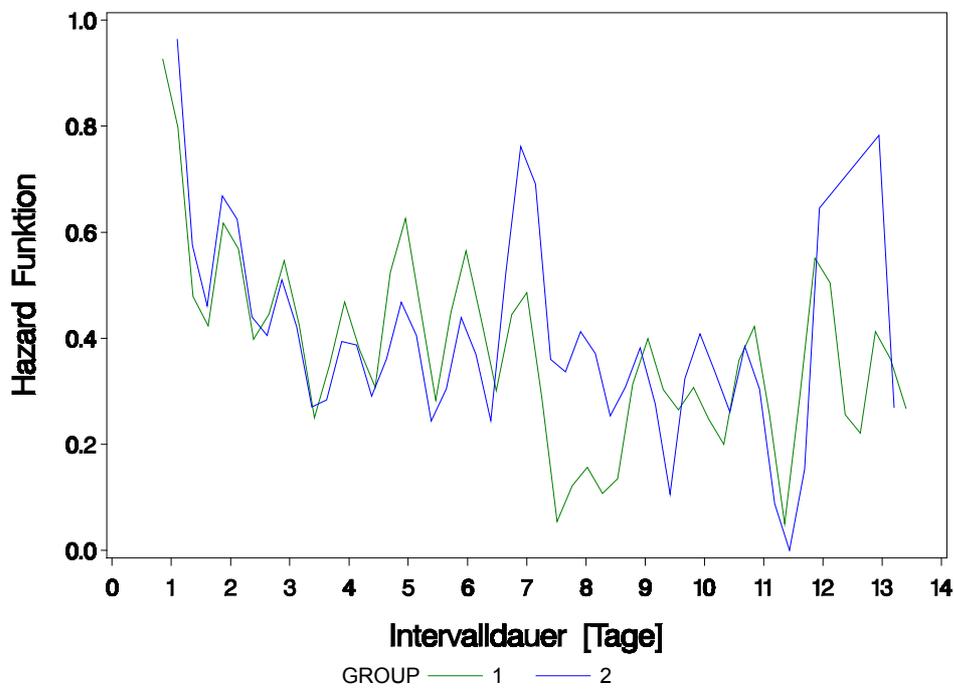


Abbildung X.22 : Ungeglättete Hazardkurve, Dienstleistung allgemein, Altersschichtung

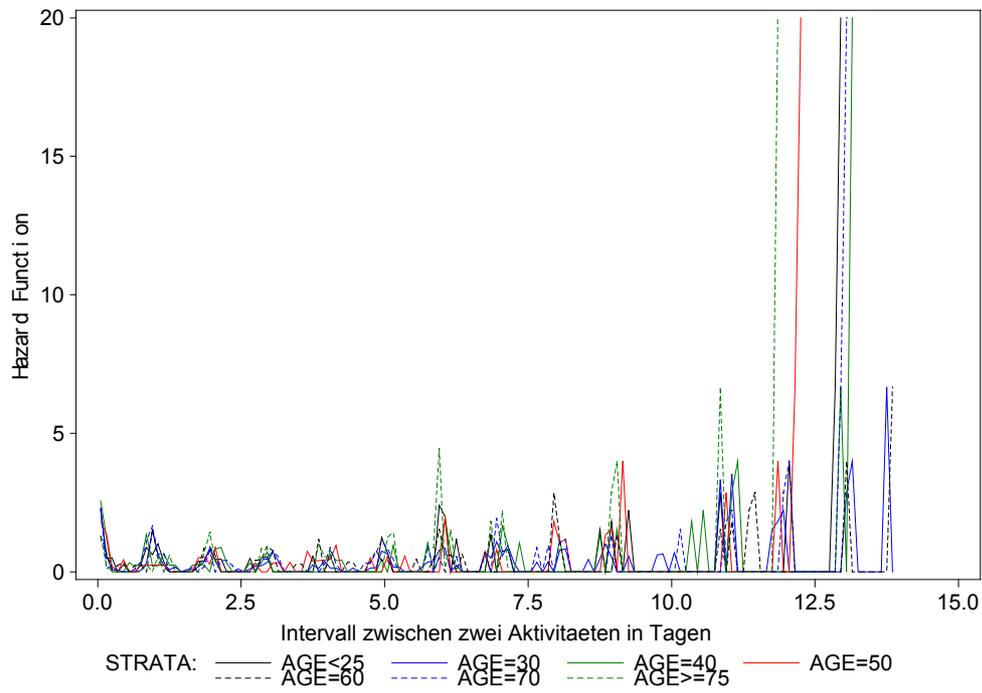


Abbildung X.23 : Geglättete Hazardkurve, Dienstleistung allgemein, Altersschichtung

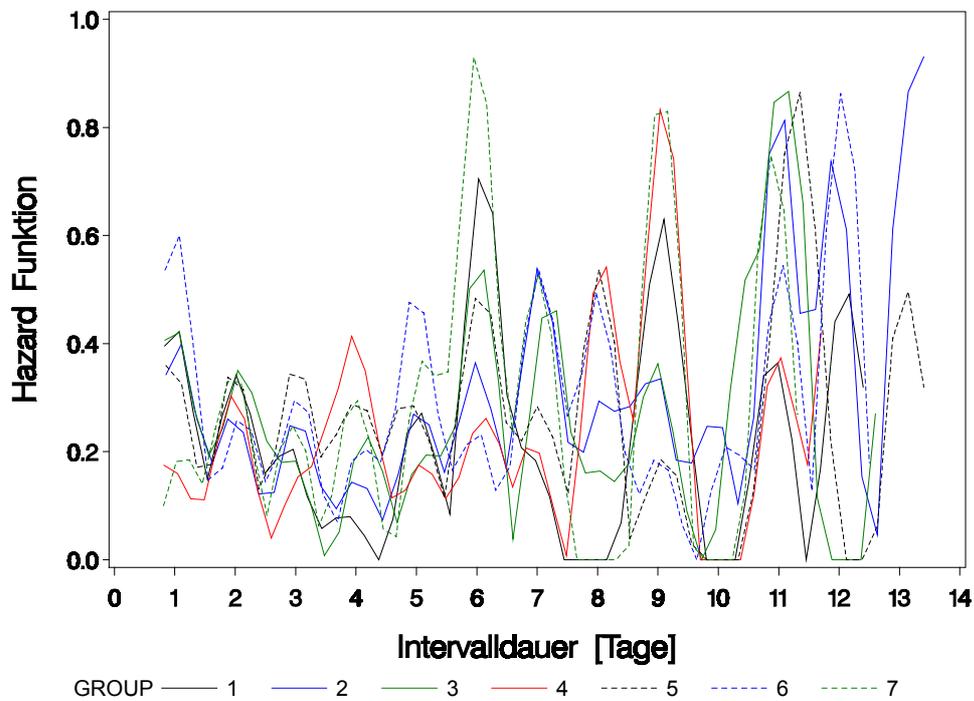


Abbildung X.24 : Ungeglättete Hazardkurve, Dienstleistung allgemein, Geschlechtsunterschiede

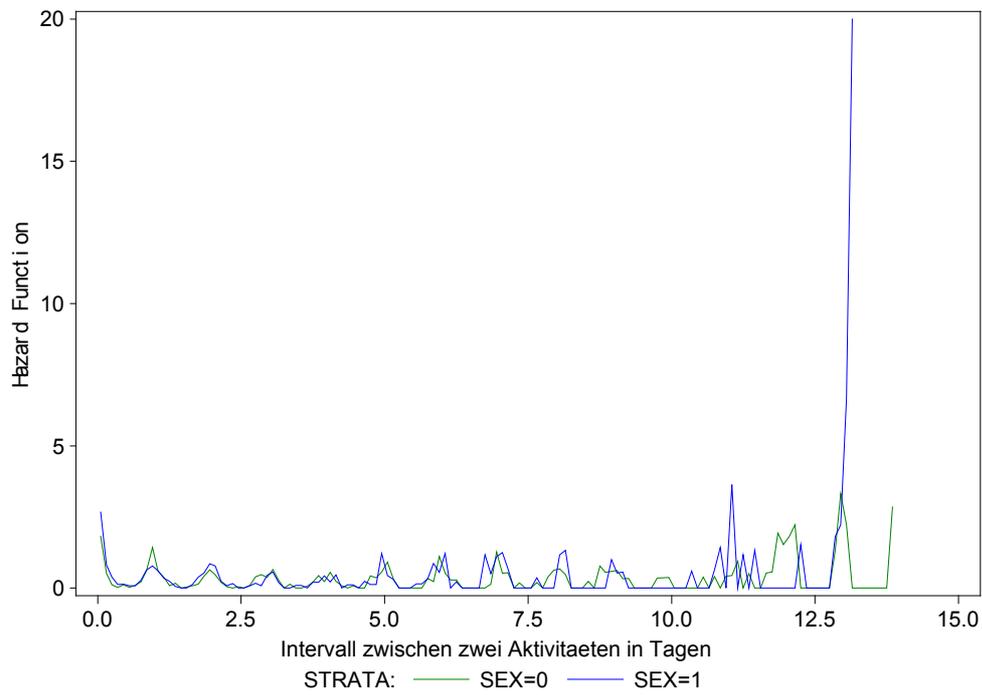


Abbildung X.25 : Geglättete Hazardkurve, Dienstleistung allgemein, Geschlechtsunterschiede

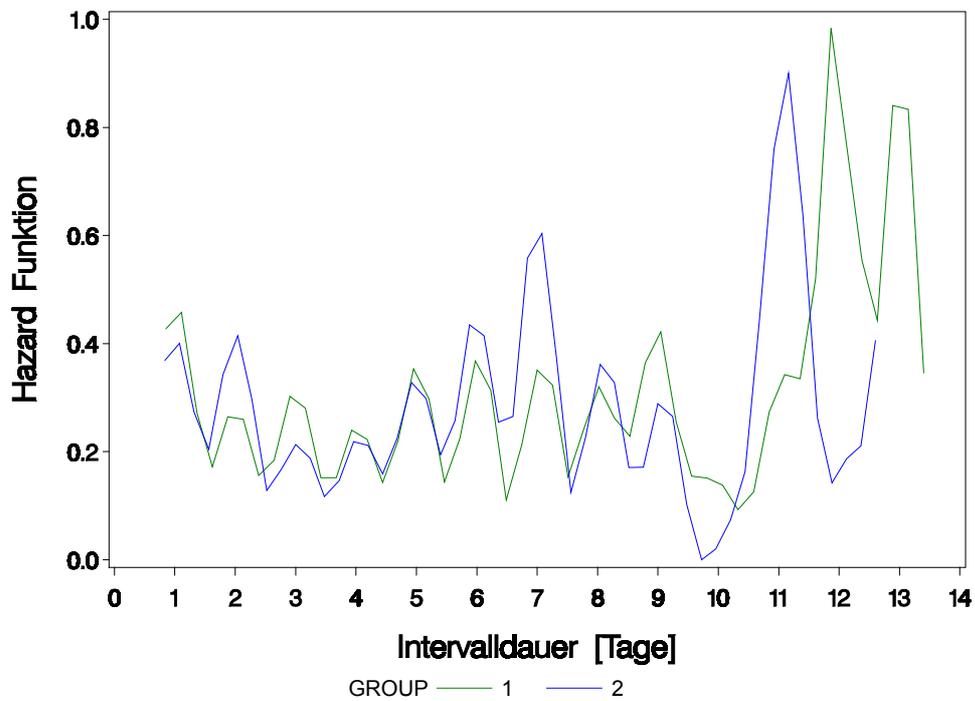


Abbildung X.26 : Ungeglättete Hazardkurve, Treffen mit Freunden, Altersschichtung

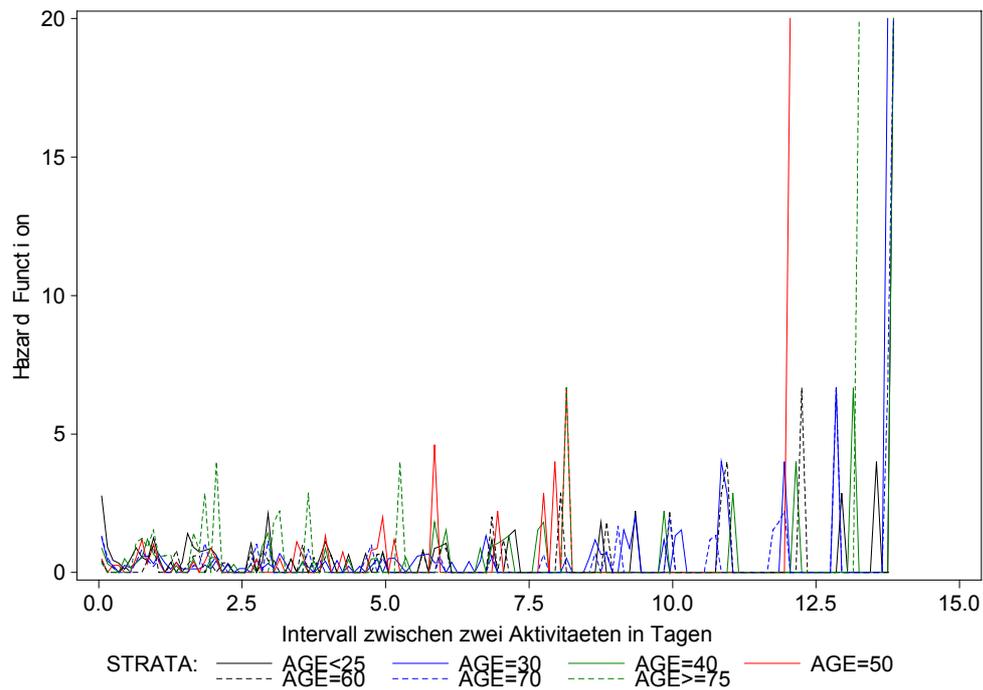


Abbildung X.27 : Geglättete Hazardkurve, Treffen mit Freunden, Altersschichtung

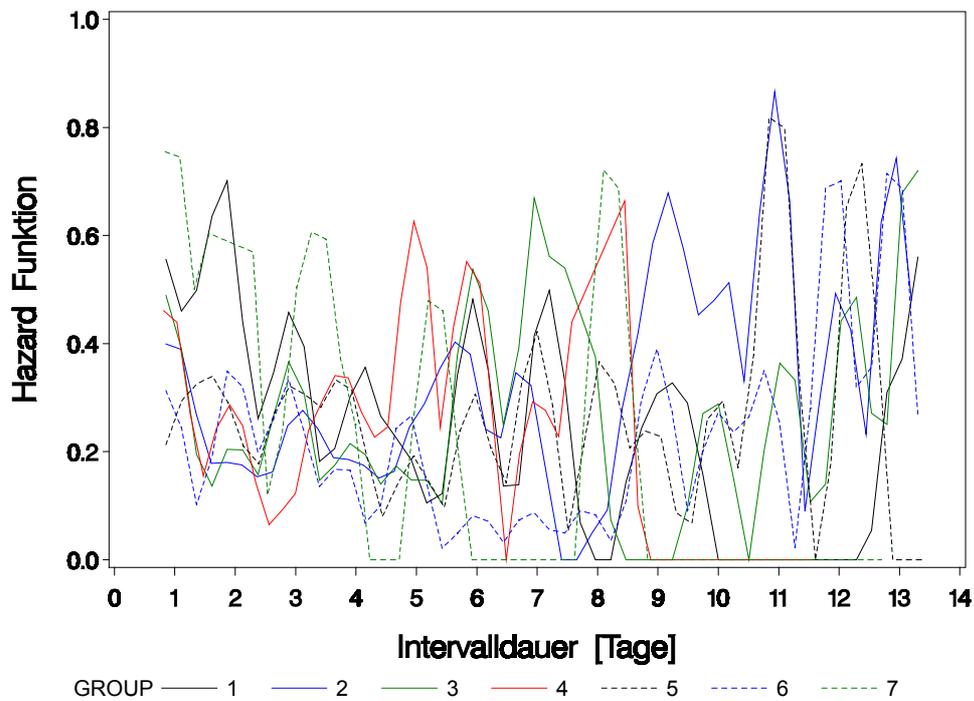


Abbildung X.28 : Ungeglättete Hazardkurve, Treffen mit Freunden, Geschlechtsunterschiede

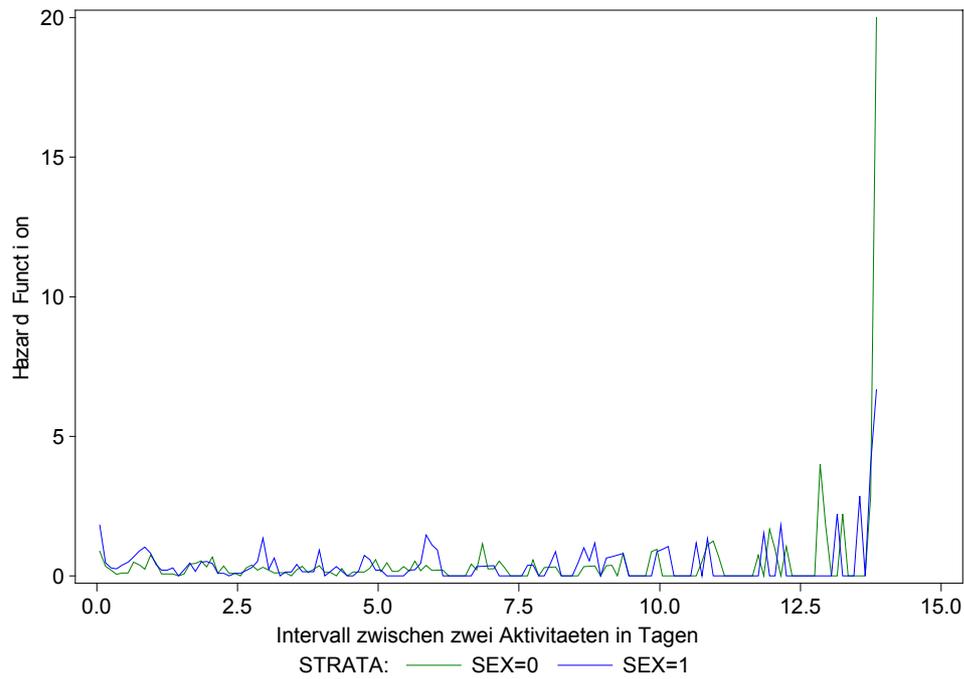


Abbildung X.29 : Geglättete Hazardkurve, Treffen mit Freunden, Geschlechtsunterschiede

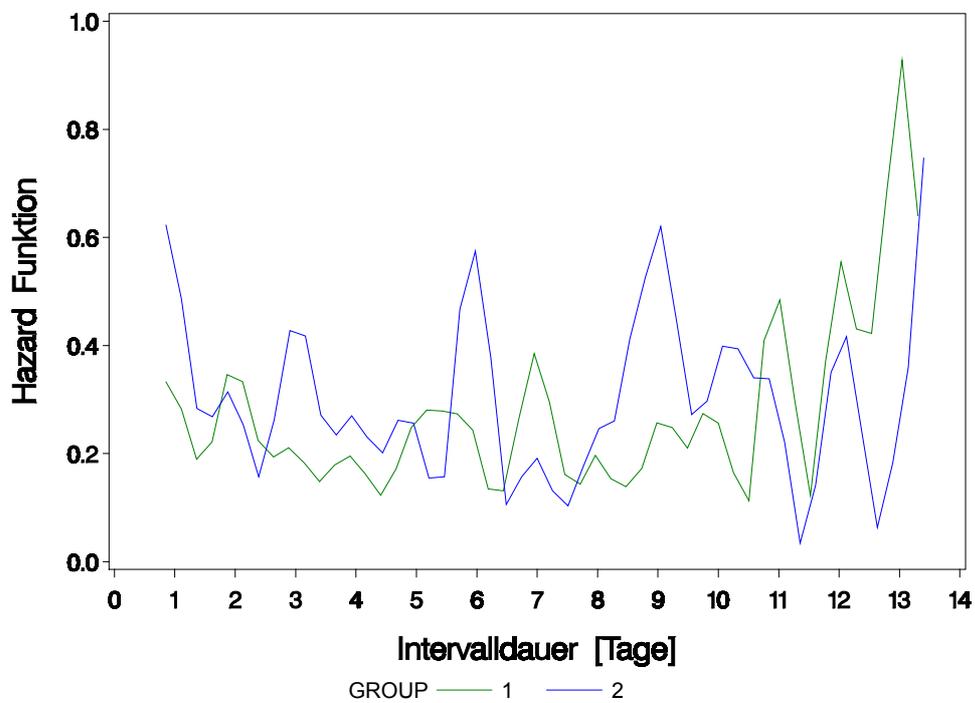


Abbildung X.30 : Survivalkurve Kategorie Einkauf kurzfristig, Altersschichtung

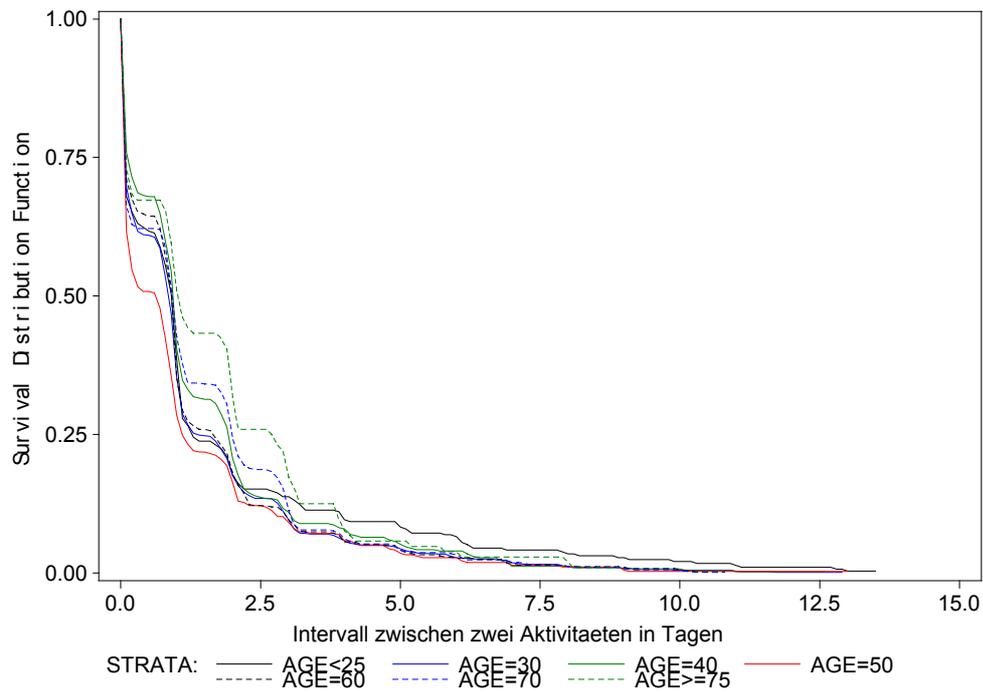


Abbildung X.31 : Survivalkurve Kategorie Einkauf kurzfristig, Geschlechtsunterschiede

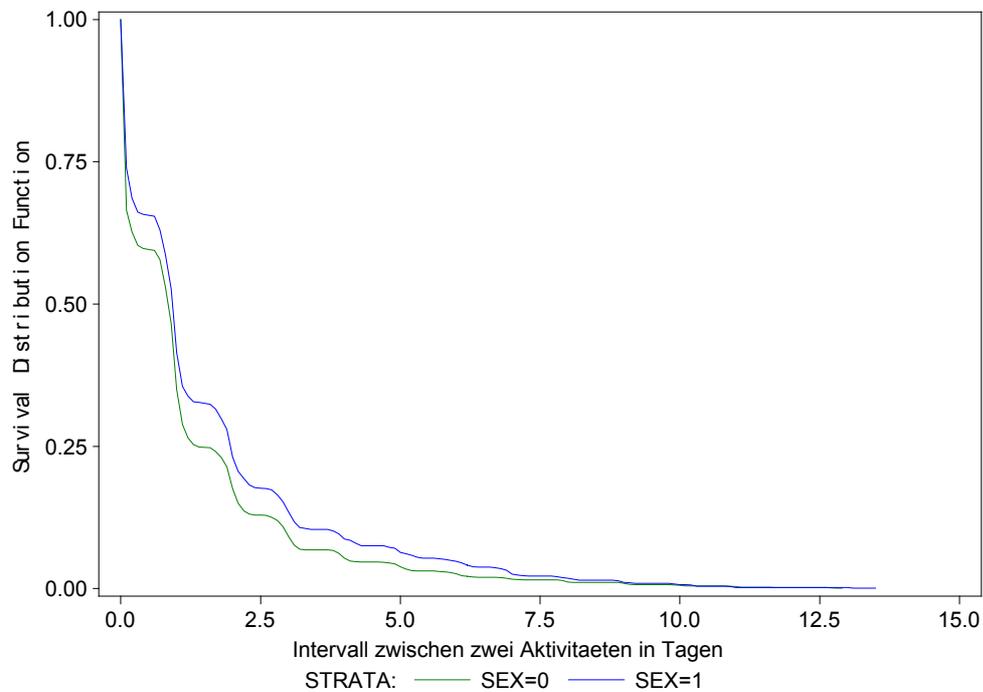


Abbildung X.32 : Survivalkurve Kategorie Freizeit allgemein, Altersschichtung

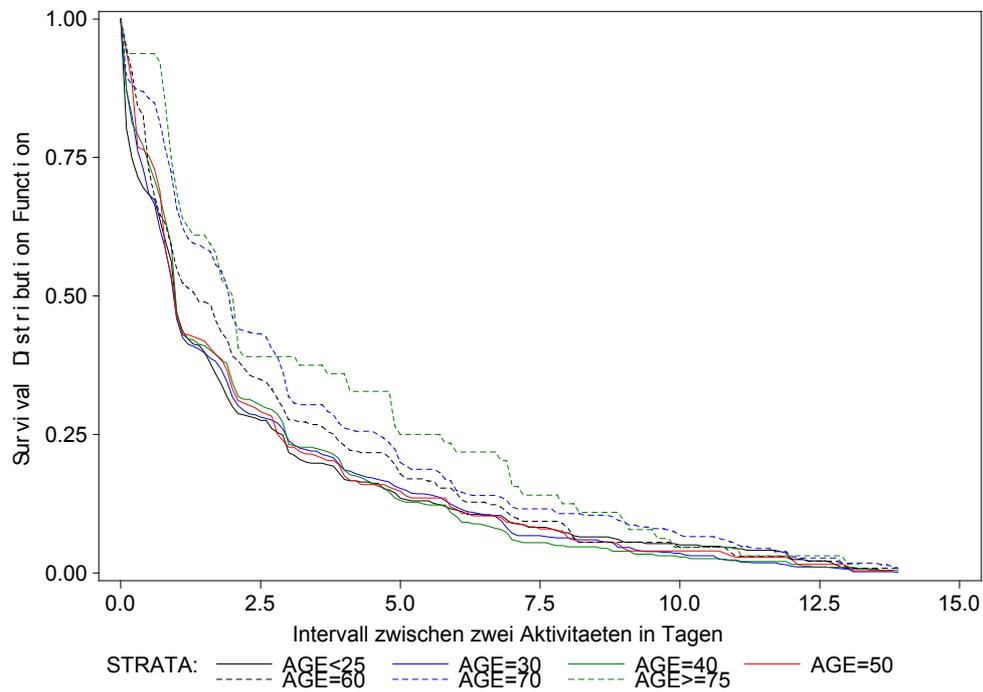


Abbildung X.33 : Survivalkurve Kategorie Freizeit allgemein, Geschlechtsunterschiede

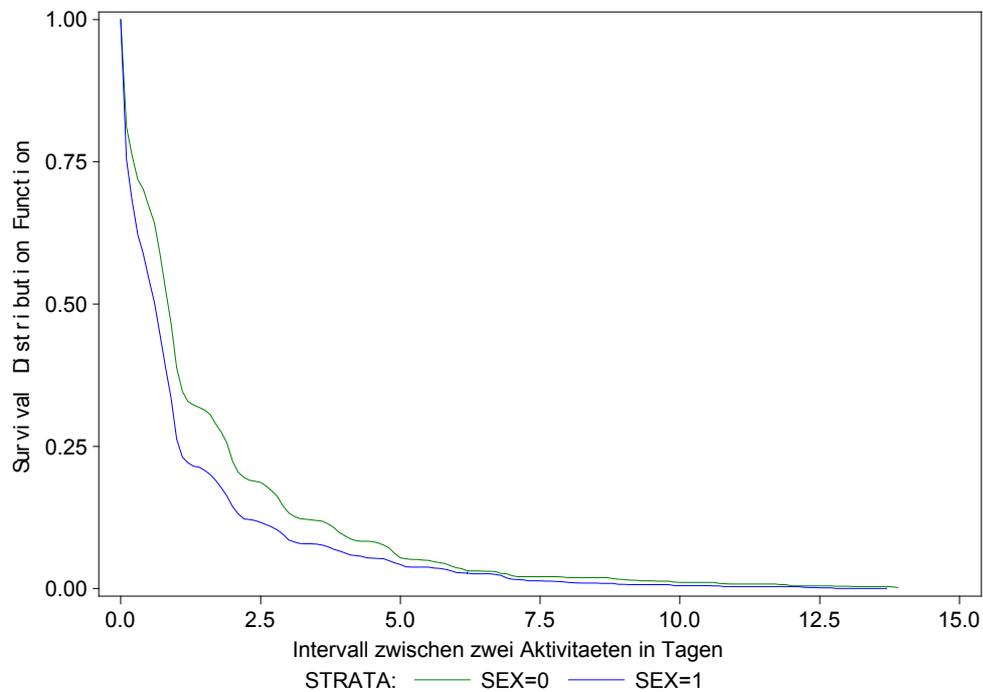


Abbildung X.34 : Survivalkurve Kategorie Dienstleistung allgemein, Altersschichtung

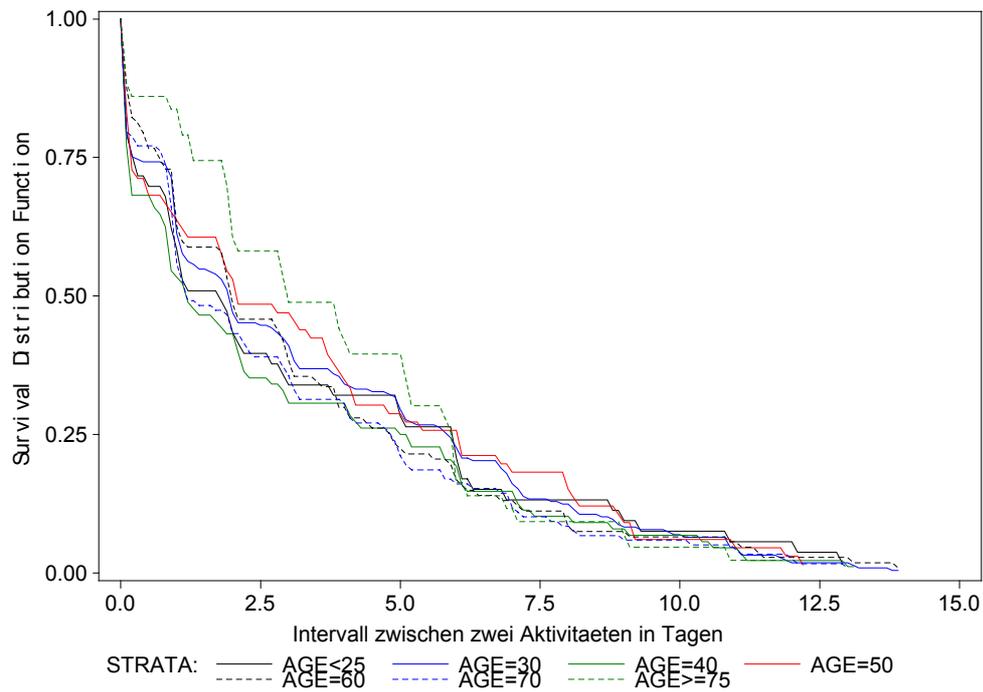


Abbildung X.35 : Survivalkurve Kategorie Dienstleistung allgemein, Geschlechtsunterschiede

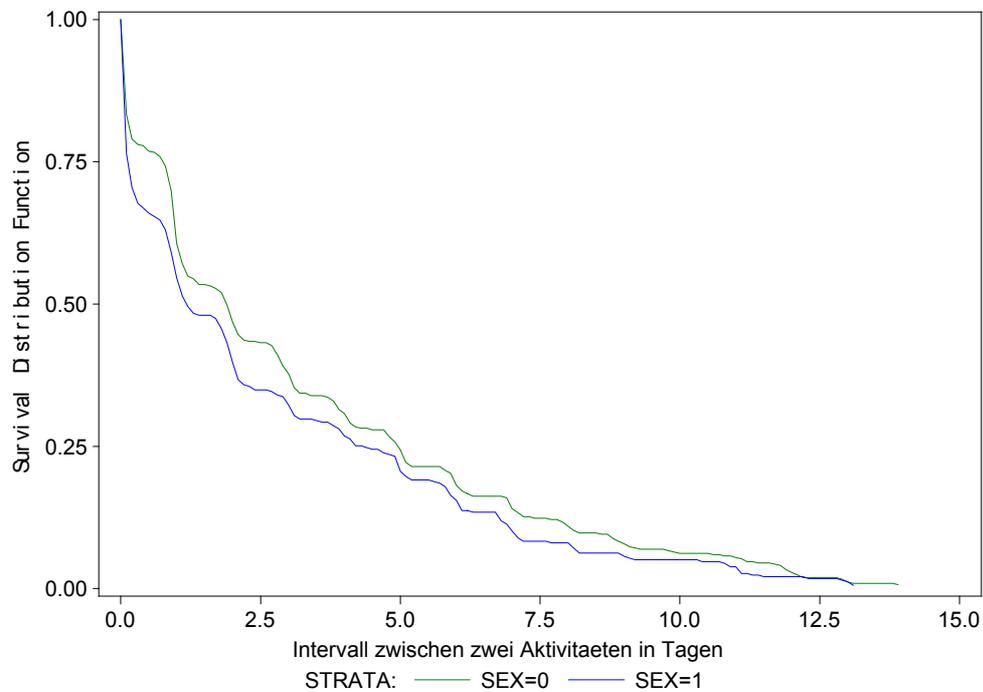


Abbildung X.36 : Survivalkurve Kategorie Treffen mit Freunden, Altersschichtung

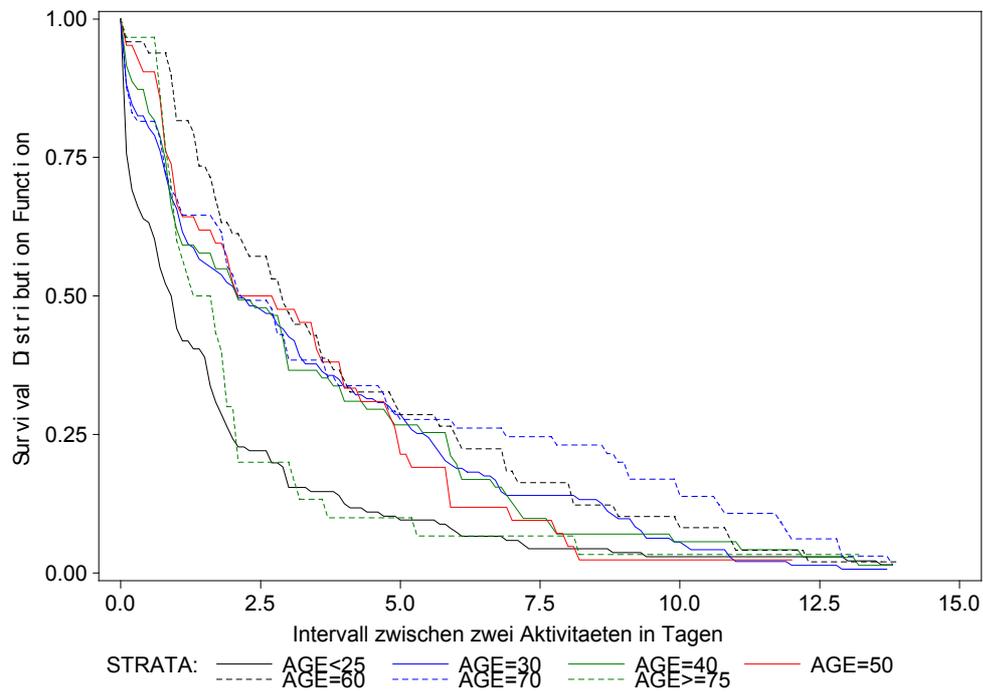
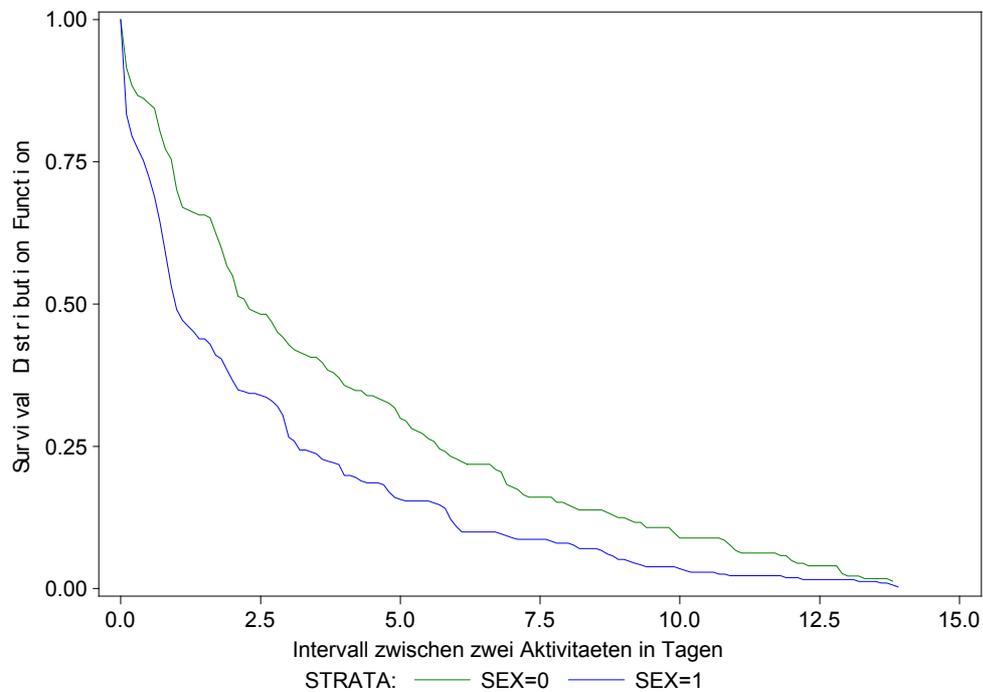


Abbildung X.37 : Survivalkurve Kategorie Treffen mit Freunden, Geschlechtsunterschiede



## **Anhang XI: Auswertungen von *PROC LIFETEST*: Wilcoxon und Log-Rank-Tests der 4 Kategorien, Intervalldauern**

**Einerseits mit allen möglichen Variablen, andererseits mit all jenen, die in den semiparametrischen Auswertungen im 1. Gesamtmodell verwendet wurden.**

**Übersicht:**

**Tabellen XI.1 bis XI.6 Kategorie Einkauf kurzfristig**

**Tabellen XI.7 bis XI.12 Kategorie Freizeit allgemein**

**Tabellen XI.13 bis XI.18 Kategorie Dienstleistung allgemein**

**Tabellen XI.19 bis XI.24 Kategorie Treffen mit Freunden**

In allen Auswertungen ist die abhängige Variable *inter\_t*, also Intervalldauer zwischen zwei gleichartigen Aktivitäten in Tagen.

Für jede Kategorie sind ein Wilcoxon Test und der Log-Rank-Test mit jeweils 3 verschiedenen Testmethoden (Univariater Chi-Quadrat Test, Kovarianz Matrix für die Wilcoxon Statistiken und Stufenförmiger Chi-Quadrat-Test) gemacht worden. Die Auswertungen sind zuerst mit allen möglichen Einflußvariablen durchgeführt worden und in einem zweiten Schritt für jede der vier Kategorien noch zusätzlich mit all jenen Variablen, die bei den semiparametrischen Regressionsmodellen (mit *PROC PHREG*) im ersten Gesamtmodell verwendet worden sind.

Die wichtigsten Ergebnisse werden hier präsentiert, die Erläuterung finden Sie im Kapitel 9.3.2.3. Die gesamte Auswertung ist im Zusatzanhang XXIII (LIFETEST I Logrank und Wilcoxontest) zu sehen.

Prinzipiell wird beim Wilcoxontest den frühen Ereigniszeiten, in diesem Fall also eher kurzen Intervalldauern, mehr Bedeutung zugemessen, beim Log-Rank-Test gehen längere Intervalldauern stärker gewichtet in die Berechnungen mit ein. Aus diesem Grund können auch schon bei den univariaten Tests, bei denen die Variablen einzeln untersucht werden, unterschiedliche Ergebnisse entstehen (wie im Zusatzanhang XXIII zu sehen). In diesem Anhang werden in der Kategorie Einkauf kurzfristig die Ergebnisse des Wilcoxon-Tests präsentiert, da es sich hierbei meist um kurze Intervalldauern handelt bzw. 91% der Intervalle schon nach 3 Tagen beendet werden. Bei den Kategorien Freizeit allgemein, Treffen mit Freunden und Dienstleistung allgemein werden ebenfalls die Ergebnisse des Wilcoxon Tests verwendet, da mehr als die Hälfte aller Intervalle schon im ersten Drittel der Zeitachse beendet werden. Die univariaten und stufenförmigen Tests jener Variablen, die in den semiparametrischen Auswertungen im ersten Gesamtmodell verwendet werden, werden ebenso gezeigt, vor allem um Vergleiche zu den semiparametrischen Modellergebnissen anstellen zu können. Es sind dabei jeweils die Ergebnisse sowohl des stufenförmigen Log-Rank- als auch des Wilcoxontests angegeben.

# Kategorie Einkauf kurzfristig, alle Variablen

Tabelle XI.1 : Univariate Chi-Squares for the Wilcoxon Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
<b>SEX</b>	<b>66.4104</b>	<b>14.6008</b>	<b>20.6879</b>	<b>&lt;.0001</b>
YOUNG	0.5702	15.2201	0.00140	0.9701
LICENSE	-29.0555	15.0963	3.7044	0.0543
<b>HEAD_OHH</b>	<b>79.0682</b>	<b>14.9290</b>	<b>28.0506</b>	<b>&lt;.0001</b>
WORKED	-9.9936	13.9015	0.5168	0.4722
CHILDREN	-2.5550	14.1662	0.0325	0.8569
<b>SINGLE</b>	<b>27.4825</b>	<b>13.8648</b>	<b>3.9290</b>	<b>0.0475</b>
<b>CAR_POSS</b>	<b>-37.7890</b>	<b>15.2548</b>	<b>6.1365</b>	<b>0.0132</b>
EDUC	3.2527	13.0983	0.0617	0.8039
AGE	864.4	532.3	2.6371	0.1044
<b>EDUC_YR</b>	<b>-581.3</b>	<b>131.3</b>	<b>19.5877</b>	<b>&lt;.0001</b>
G	-52.6168	53.3288	0.9735	0.3238
<b>CAR</b>	<b>-35.1318</b>	<b>14.8158</b>	<b>5.6228</b>	<b>0.0177</b>
<b>HH_NR_L</b>	<b>-55.8717</b>	<b>26.5054</b>	<b>4.4434</b>	<b>0.0350</b>
HH_MEMB	-36.7557	33.0731	1.2351	0.2664
ACC_D	-10.8183	12.2510	0.7798	0.3772
<b>PLANNED</b>	<b>41.6532</b>	<b>9.6614</b>	<b>18.5875</b>	<b>&lt;.0001</b>
A_EXP_D	5.9328	10.6989	0.3075	0.5792
<b>WEEKEND</b>	<b>-34.4387</b>	<b>12.8295</b>	<b>7.2057</b>	<b>0.0073</b>
<b>A_DAYT_D</b>	<b>34.0145</b>	<b>14.5359</b>	<b>5.4757</b>	<b>0.0193</b>
<b>A_P_MM_D</b>	<b>-55.4394</b>	<b>8.9418</b>	<b>38.4407</b>	<b>&lt;.0001</b>
A_N_MM_D	13.1078	9.1747	2.0412	0.1531
<b>A_P_MMM_D</b>	<b>50.0601</b>	<b>12.6484</b>	<b>15.6642</b>	<b>&lt;.0001</b>
A_N_MMM_D	-6.2534	12.6715	0.2435	0.6217
ACC	-11.3139	18.4521	0.3760	0.5398
A_EXP	1635.9	1182.0	1.9157	0.1663
<b>A_DAYT</b>	<b>47.0612</b>	<b>16.7431</b>	<b>7.9005</b>	<b>0.0049</b>
<b>A_P_DIST</b>	<b>616596</b>	<b>104368</b>	<b>34.9034</b>	<b>&lt;.0001</b>
<b>A_P_TT</b>	<b>3025.1</b>	<b>295.9</b>	<b>104.5</b>	<b>&lt;.0001</b>
A_N_DIST	-59221.4	330752	0.0321	0.8579
A_N_TT	215.4	550.5	0.1531	0.6956
<b>A_P_MM</b>	<b>312.6</b>	<b>41.6726</b>	<b>56.2723</b>	<b>&lt;.0001</b>
A_N_MM	-9.8286	41.6767	0.0556	0.8136
<b>NR_T_SUM</b>	<b>-13044.9</b>	<b>1823.0</b>	<b>51.2053</b>	<b>&lt;.0001</b>
<b>NR_J_SUM</b>	<b>-1166.5</b>	<b>550.1</b>	<b>4.4973</b>	<b>0.0339</b>
<b>A_P_PUR</b>	<b>660.1</b>	<b>84.8376</b>	<b>60.5387</b>	<b>&lt;.0001</b>
A_N_PUR	43.7130	69.5493	0.3950	0.5297
<b>INTER_T_P</b>	<b>167.6</b>	<b>62.4008</b>	<b>7.2168</b>	<b>0.0072</b>
<b>INTERVALL_ST</b>	<b>59.4345</b>	<b>6.7901</b>	<b>76.6168</b>	<b>&lt;.0001</b>
<b>A_P_DURST_A</b>	<b>8589.1</b>	<b>297.7</b>	<b>832.3</b>	<b>&lt;.0001</b>
<b>A_P_DURST_C</b>	<b>63.9348</b>	<b>21.1803</b>	<b>9.1120</b>	<b>0.0025</b>
<b>A_P_MM_C</b>	<b>137.7</b>	<b>40.4721</b>	<b>11.5709</b>	<b>0.0007</b>
<b>MM_C_D</b>	<b>-89.4171</b>	<b>14.5824</b>	<b>37.5995</b>	<b>&lt;.0001</b>
<b>INTERVALL_ST_P</b>	<b>77.5520</b>	<b>6.8132</b>	<b>129.6</b>	<b>&lt;.0001</b>

**Tabelle XI.2 : Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_DURST_A	1	832.3	<.0001	832.3	<.0001
A_DAYT	2	921.8	<.0001	89.5162	<.0001
INTERVALL_ST_P	3	986.9	<.0001	65.1006	<.0001
SEX	4	1033.9	<.0001	47.0114	<.0001
A_P_PUR	5	1077.6	<.0001	43.6586	<.0001
A_P_TT	6	1115.9	<.0001	38.3089	<.0001
NR_J_SUM	7	1142.5	<.0001	26.5898	<.0001
MM_C_D	8	1165.9	<.0001	23.4243	<.0001
WEEKEND	9	1181.3	<.0001	15.4165	<.0001
A_N_MM_D	10	1195.1	<.0001	13.7710	0.0002
NR_T_SUM	11	1210.3	<.0001	15.2556	<.0001
WORKED	12	1227.3	<.0001	17.0049	<.0001
SINGLE	13	1243.3	<.0001	15.9514	<.0001
A_P_MMMD	14	1254.9	<.0001	11.6436	0.0006
A_N_MMMD	15	1283.4	<.0001	28.4363	<.0001
EDUC	16	1294.6	<.0001	11.2439	0.0008
PLANNED	17	1304.0	<.0001	9.4167	0.0022
A_P_DIST	18	1307.9	<.0001	3.8874	0.0486
HH_NR_L	19	1311.7	<.0001	3.7825	0.0518
LICENSE	20	1321.4	<.0001	9.7187	0.0018
ACC_D	21	1326.7	<.0001	5.2873	0.0215
INTERVALL_ST	22	1330.5	<.0001	3.7618	0.0524
HEAD_OHH	23	1333.8	<.0001	3.3779	0.0661
A_N_DIST	24	1336.3	<.0001	2.4837	0.1150
A_EXP_D	25	1338.8	<.0001	2.5163	0.1127
CHILDREN	26	1341.7	<.0001	2.8315	0.0924
HH_MEMB	27	1352.7	<.0001	11.0181	0.0009
INTER_T_P	28	1353.8	<.0001	1.1109	0.2919
ACC	29	1354.9	<.0001	1.0532	0.3048
CAR	30	1355.8	<.0001	0.9755	0.3233
A_P_DURST_C	31	1356.6	<.0001	0.8087	0.3685
A_N_TT	32	1357.0	<.0001	0.3583	0.5494
A_DAYT_D	33	1357.3	<.0001	0.3416	0.5589
A_P_MM	34	1357.6	<.0001	0.2363	0.6269
A_P_MM_C	35	1357.8	<.0001	0.2175	0.6409
G	36	1358.0	<.0001	0.1743	0.6763
A_N_MM	37	1358.1	<.0001	0.1376	0.7107
A_P_MM_D	38	1358.4	<.0001	0.3038	0.5815
EDUC_YR	39	1358.5	<.0001	0.0896	0.7646
A_N_PUR	40	1358.5	<.0001	0.0389	0.8436
CAR_POSS	41	1358.6	<.0001	0.0277	0.8677
A_EXP	42	1358.6	<.0001	0.0155	0.9009
AGE	43	1358.6	<.0001	0.0136	0.9070
YOUNG	44	1358.6	<.0001	0.000419	0.9837

# Kategorie Einkauf kurzfristig, Variablen des 1. Gesamtmodells

**Tabelle XI.3 : Univariate Chi-Squares for the Wilcoxon Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
SEX	72.9127	14.9083	23.9195	<.0001
LICENSE	-30.0096	15.3570	3.8186	0.0507
AGE	826.6	534.9	2.3880	0.1223
CAR	-35.0635	15.0080	5.4584	0.0195
HH_NR_L	-61.2453	26.7510	5.2416	0.0221
PLANNED	41.5292	9.9967	17.2582	<.0001
WEEKEND	-35.9823	13.0142	7.6444	0.0057
A_EXP	1897.1	1330.9	2.0316	0.1541
A_DAYT	47.1952	17.0200	7.6891	0.0056
A_P_DIST	661476	104912	39.7540	<.0001
A_P_TT	3191.3	300.0	113.1	<.0001
A_P_MM	313.2	42.6133	54.0220	<.0001
NR_J_SUM	-1209.8	554.7	4.7562	0.0292
A_P_PUR	669.3	86.3870	60.0309	<.0001
INTER_T_P	186.2	64.7231	8.2728	0.0040
INTERVALL_ST	62.2321	6.8752	81.9321	<.0001
A_P_DURST_A	8695.6	301.4	832.5	<.0001
A_P_DURST_C	63.0385	21.2725	8.7816	0.0030
A_P_MM_C	142.2	41.3145	11.8522	0.0006
MM_C_D	-89.6960	14.8787	36.3424	<.0001
INTERVALL_ST_P	79.7044	6.8786	134.3	<.0001

**Tabelle XI.4 : Forward Stepwise Sequence of Chi-Squares for Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_DURST_A	1	832.5	<.0001	832.5	<.0001
A_DAYT	2	921.8	<.0001	89.3119	<.0001
INTERVALL_ST_P	3	990.7	<.0001	68.8325	<.0001
SEX	4	1043.5	<.0001	52.8478	<.0001
A_P_PUR	5	1086.4	<.0001	42.8349	<.0001
A_P_TT	6	1128.7	<.0001	42.3771	<.0001
NR_J_SUM	7	1154.7	<.0001	25.9521	<.0001
MM_C_D	8	1176.8	<.0001	22.1016	<.0001
WEEKEND	9	1192.9	<.0001	16.1516	<.0001
PLANNED	10	1203.1	<.0001	10.1905	0.0014
INTER_T_P	11	1210.9	<.0001	7.7717	0.0053
AGE	12	1215.5	<.0001	4.6328	0.0314
INTERVALL_ST	13	1219.6	<.0001	4.0430	0.0444
A_P_MM	14	1223.5	<.0001	3.9493	0.0469
A_P_DIST	15	1225.5	<.0001	1.9301	0.1648
A_P_DURST_C	16	1227.2	<.0001	1.6905	0.1935
CAR	17	1228.3	<.0001	1.1620	0.2810
HH_NR_L	18	1229.0	<.0001	0.7310	0.3926
A_EXP	19	1229.2	<.0001	0.2032	0.6522
A_P_MM_C	20	1229.4	<.0001	0.1451	0.7033
LICENSE	21	1229.5	<.0001	0.0896	0.7647

**Tabelle XI.5 : Univariate Chi-Squares for the Log-Rank Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
SEX	115.6	26.3290	19.2683	<.0001
LICENSE	-53.7779	26.7994	4.0268	0.0448
AGE	1063.9	966.9	1.2109	0.2712
CAR	-49.2640	26.2925	3.5107	0.0610
HH_NR_L	-70.3696	46.3692	2.3031	0.1291
PLANNED	55.5138	16.1145	11.8677	0.0006
WEEKEND	-59.2773	21.5783	7.5465	0.0060
A_EXP	3129.9	2679.4	1.3646	0.2427
A_DAYT	50.8723	28.9565	3.0865	0.0789
A_P_DIST	948361	226369	17.5515	<.0001
A_P_TT	4120.1	551.5	55.8170	<.0001
A_P_MM	368.8	74.9235	24.2260	<.0001
NR_J_SUM	-3752.3	1000.1	14.0781	0.0002
A_P_PUR	721.7	155.9	21.4208	<.0001
INTER_T_P	382.5	124.9	9.3835	0.0022
INTERVALL_ST	87.5504	22.1057	15.6859	<.0001
A_P_DURST_A	15572.5	1377.4	127.8	<.0001
A_P_DURST_C	102.1	38.0427	7.2088	0.0073
A_P_MM_C	312.7	75.4308	17.1803	<.0001
MM_C_D	-127.5	26.2989	23.4987	<.0001
INTERVALL_ST_P	119.1	14.4703	67.8002	<.0001

**Tabelle XI.6 : Forward Stepwise Sequence of Chi-Squares, Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_DURST_A	1	127.8	<.0001	127.8	<.0001
A_P_TT	2	176.3	<.0001	48.4489	<.0001
SEX	3	207.0	<.0001	30.7855	<.0001
INTERVALL_ST_P	4	228.8	<.0001	21.7560	<.0001
WEEKEND	5	246.7	<.0001	17.8760	<.0001
MM_C_D	6	260.2	<.0001	13.5211	0.0002
PLANNED	7	270.9	<.0001	10.7319	0.0011
LICENSE	8	278.0	<.0001	7.0201	0.0081
A_P_MM	9	286.6	<.0001	8.6142	0.0033
A_P_DURST_C	10	291.1	<.0001	4.5467	0.0330
AGE	11	295.2	<.0001	4.1265	0.0422
INTER_T_P	12	298.8	<.0001	3.5991	0.0578
A_DAYT	13	301.0	<.0001	2.1595	0.1417
A_P_MM_C	14	303.1	<.0001	2.1413	0.1434
INTERVALL_ST	15	304.2	<.0001	1.0855	0.2975
CAR	16	304.9	<.0001	0.6789	0.4099
HH_NR_L	17	306.4	<.0001	1.4833	0.2233
A_P_PUR	18	307.1	<.0001	0.7372	0.3905
NR_J_SUM	19	307.9	<.0001	0.7366	0.3908
A_EXP	20	308.1	<.0001	0.2743	0.6005
A_P_DIST	21	308.1	<.0001	0.0101	0.9199

# Kategorie Freizeit allgemein, alle Variablen

Tabelle XI.7 : Univariate Chi-Squares for the Wilcoxon Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
SEX	-99.5437	14.7180	45.7434	<.0001
YOUNG	-126.3	13.1765	91.9233	<.0001
LICENSE	-69.2532	14.1570	23.9298	<.0001
HEAD_OHH	-47.8281	13.8063	12.0009	0.0005
WORKED	-40.8057	12.9130	9.9859	0.0016
CHILDREN	-23.1730	14.4366	2.5765	0.1085
SINGLE	-91.2673	13.9436	42.8428	<.0001
CAR_POSS	-66.8810	14.3401	21.7522	<.0001
EDUC	-35.4167	13.4488	6.9351	0.0085
AGE	5058.1	512.0	97.6102	<.0001
EDUC_YR	-651.1	120.2	29.3501	<.0001
G	-236.0	50.6397	21.7105	<.0001
CAR	-93.8437	12.9228	52.7345	<.0001
HH_NR_L	-95.8426	23.9149	16.0613	<.0001
HH_MEMB	18.3722	34.5504	0.2828	0.5949
ACC_D	53.2749	13.9987	14.4833	0.0001
PLANNED	69.4095	6.9180	100.7	<.0001
A_EXP_D	23.0369	9.7710	5.5587	0.0184
WEEKEND	-75.9852	14.0969	29.0541	<.0001
A_DAYT_D	-69.8589	12.8301	29.6474	<.0001
A_P_MM_D	-52.5323	8.7579	35.9797	<.0001
A_N_MM_D	-21.5764	7.9775	7.3151	0.0068
A_P_MMMD	-45.1943	14.5292	9.6757	0.0019
A_N_MMMD	-41.0342	14.5811	7.9198	0.0049
ACC	49.3532	24.3482	4.1086	0.0427
A_EXP	631.1	623.1	1.0260	0.3111
A_DAYT	-171.7	21.2505	65.2976	<.0001
A_P_DIST	1621727	634634	6.5299	0.0106
A_P_TT	5650.0	1174.5	23.1399	<.0001
A_N_DIST	1847551	617937	8.9393	0.0028
A_N_TT	3601.4	1066.4	11.4056	0.0007
A_P_MM	-3.4114	49.4054	0.00477	0.9450
A_N_MM	-54.8137	49.4067	1.2309	0.2672
NR_T_SUM	-29244.0	2004.6	212.8	<.0001
NR_J_SUM	-9101.0	833.7	119.2	<.0001
A_P_PUR	0.3209	70.8554	0.000021	0.9964
A_N_PUR	-64.3821	62.3066	1.0677	0.3015
INTER_T_P	485.4	59.7391	66.0342	<.0001
INTERVALL_ST	71.1550	16.9854	17.5492	<.0001
A_P_DURST_A	6672.0	270.2	609.8	<.0001
A_P_DURST_C	603.5	67.4448	80.0655	<.0001
A_P_MM_C	82.2228	48.6565	2.8556	0.0911
MM_C_D	-121.1	14.5728	69.0085	<.0001
INTERVALL_ST_P	-35.0308	8.1591	18.4340	<.0001

**Tabelle XI.8 : Forward Stepwise Sequence of Chi-Squares, Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_DURST_A	1	609.8	<.0001	609.8	<.0001
WEEKEND	2	714.7	<.0001	104.9	<.0001
NR_T_SUM	3	813.6	<.0001	98.9148	<.0001
INTERVALL_ST_P	4	890.9	<.0001	77.2755	<.0001
A_P_PUR	5	966.6	<.0001	75.7292	<.0001
PLANNED	6	1065.0	<.0001	98.3227	<.0001
MM_C_D	7	1134.4	<.0001	69.4102	<.0001
A_P_DURST_C	8	1185.0	<.0001	50.6622	<.0001
SINGLE	9	1231.6	<.0001	46.5738	<.0001
CAR	10	1250.4	<.0001	18.7716	<.0001
CAR_POSS	11	1283.6	<.0001	33.1749	<.0001
WORKED	12	1294.0	<.0001	10.4316	0.0012
A_P_MMMD	13	1306.4	<.0001	12.3785	0.0004
SEX	14	1316.4	<.0001	9.9859	0.0016
INTER_T_P	15	1324.6	<.0001	8.2965	0.0040
EDUC	16	1330.5	<.0001	5.8380	0.0157
A_P_TT	17	1335.8	<.0001	5.2824	0.0215
HH_NR_L	18	1340.7	<.0001	4.9631	0.0259
NR_J_SUM	19	1344.3	<.0001	3.6151	0.0573
LICENSE	20	1348.1	<.0001	3.7993	0.0513
A_EXP_D	21	1351.6	<.0001	3.4145	0.0646
A_EXP	22	1355.6	<.0001	4.0725	0.0436
YOUNG	23	1357.9	<.0001	2.2484	0.1338
A_N_TT	24	1359.9	<.0001	2.0181	0.1554
A_N_DIST	25	1362.1	<.0001	2.2198	0.1363
A_DAYT	26	1362.7	<.0001	0.5925	0.4415
A_N_MM_D	27	1363.2	<.0001	0.4527	0.5010
A_P_MM_D	28	1364.3	<.0001	1.1383	0.2860
A_P_DIST	29	1364.5	<.0001	0.2337	0.6288
CHILDREN	30	1364.7	<.0001	0.1682	0.6818
HH_MEMB	31	1364.9	<.0001	0.1575	0.6914
A_P_MM_C	32	1365.0	<.0001	0.0931	0.7602
A_N_MMMD	33	1365.1	<.0001	0.0991	0.7529
AGE	34	1365.1	<.0001	0.0756	0.7834
A_N_PUR	35	1365.2	<.0001	0.0552	0.8142
G	36	1365.2	<.0001	0.0456	0.8309
A_DAYT_D	37	1365.3	<.0001	0.0377	0.8461
INTERVALL_ST	38	1365.3	<.0001	0.0220	0.8820
EDUC_YR	39	1365.3	<.0001	0.0167	0.8972
HEAD_OHH	40	1365.3	<.0001	0.0145	0.9043
A_P_MM	41	1365.3	<.0001	0.00218	0.9628
ACC_D	42	1365.3	<.0001	0.00139	0.9702
ACC	43	1365.3	<.0001	0.00234	0.9614
A_N_MM	44	1365.3	<.0001	0.000411	0.9838

# Kategorie Freizeit allgemein, Variablen aus dem 1. Gesamtmodell

**Tabelle XI.9 : Univariate Chi-Squares for the Wilcoxon Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
SEX	-106.9	14.9992	50.7895	<.0001
LICENSE	-76.8509	14.3702	28.6003	<.0001
HEAD_OHH	-54.6239	14.0917	15.0258	0.0001
WORKED	-41.9572	13.0038	10.4105	0.0013
SINGLE	-106.1	14.2561	55.3459	<.0001
EDUC	-33.8830	13.5619	6.2420	0.0125
AGE	5049.7	514.1	96.4749	<.0001
G	-206.6	51.3497	16.1888	<.0001
CAR	-98.5906	13.0992	56.6480	<.0001
ACC_D	62.7760	14.2844	19.3135	<.0001
PLANNED	79.0099	7.2710	118.1	<.0001
WEEKEND	-74.1042	14.3738	26.5793	<.0001
A_P_MM_D	-51.8476	8.9584	33.4960	<.0001
A_N_MM_D	-24.1817	8.1932	8.7110	0.0032
A_DAYT	-178.7	21.5751	68.6273	<.0001
A_P_DIST	1664814	639092	6.7859	0.0092
A_P_TT	5760.5	1181.3	23.7779	<.0001
NR_T_SUM	-30004.0	2022.0	220.2	<.0001
A_P_DURST_A	6835.2	275.6	615.0	<.0001
A_P_DURST_C	596.3	70.4979	71.5520	<.0001
INTER_T_P	530.9	62.2433	72.7439	<.0001
A_P_MM_C	118.9	49.7623	5.7136	0.0168
MM_C_D	-126.6	14.8745	72.4450	<.0001

**Tabelle XI.10 : Forward Stepwise Sequence of Chi-Squares, Wilcoxon Test**

Variable	DF	Chi-Square	Pr >Chi-Square	Chi-Square Increment	Pr >Increment
A_P_DURST_A	1	615.0	<.0001	615.0	<.0001
WEEKEND	2	715.1	<.0001	100.2	<.0001
NR_T_SUM	3	820.5	<.0001	105.4	<.0001
PLANNED	4	902.3	<.0001	81.8317	<.0001
MM_C_D	5	973.7	<.0001	71.3086	<.0001
A_P_DURST_C	6	1005.3	<.0001	31.6918	<.0001
SINGLE	7	1038.4	<.0001	33.0798	<.0001
WORKED	8	1065.3	<.0001	26.9178	<.0001
CAR	9	1093.2	<.0001	27.8922	<.0001
A_DAYT	10	1116.1	<.0001	22.8381	<.0001
LICENSE	11	1132.3	<.0001	16.2666	<.0001
INTER_T_P	12	1144.3	<.0001	11.9284	0.0006
SEX	13	1149.8	<.0001	5.5331	0.0187
ACC_D	14	1152.4	<.0001	2.6383	0.1043
A_P_TT	15	1153.6	<.0001	1.1555	0.2824
A_P_MM_C	16	1154.5	<.0001	0.9519	0.3292
A_P_MM_D	17	1155.3	<.0001	0.7593	0.3836
HEAD_OHH	18	1156.0	<.0001	0.6812	0.4092
EDUC	19	1156.4	<.0001	0.4345	0.5098
G	20	1156.9	<.0001	0.5194	0.4711
AGE	21	1157.7	<.0001	0.7673	0.3811
A_P_DIST	22	1157.7	<.0001	0.00747	0.9311
A_N_MM_D	23	1157.7	<.0001	0.00733	0.9318

**Tabelle XI.11 : Univariate Chi-Squares for the Log-Rank Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr >Chi-Square
SEX	-161.6	26.0123	38.5945	<.0001
LICENSE	-122.8	25.5293	23.1298	<.0001
HEAD_OHH	-66.8948	24.7335	7.3150	0.0068
WORKED	-51.5984	23.2141	4.9405	0.0262
SINGLE	-140.8	23.9194	34.6393	<.0001
EDUC	-69.9173	22.6654	9.5157	0.0020
AGE	6659.1	938.2	50.3800	<.0001
G	-381.4	91.3238	17.4447	<.0001
CAR	-167.8	24.4644	47.0512	<.0001
ACC_D	138.5	25.5209	29.4583	<.0001
PLANNED	79.0531	12.1873	42.0750	<.0001
WEEKEND	-54.1855	24.7166	4.8060	0.0284
A_P_MM_D	-76.8040	17.3839	19.5197	<.0001
A_N_MM_D	-46.1341	15.9492	8.3669	0.0038
A_DAYT	-158.0	35.5554	19.7373	<.0001
A_P_DIST	3292631	1400854	5.5246	0.0188
A_P_TT	8525.6	2244.6	14.4273	0.0001
NR_T_SUM	-54033.4	2985.2	327.6	<.0001
A_P_DURST_A	11073.2	1085.8	104.0	<.0001
A_P_DURST_C	985.8	138.6	50.5629	<.0001
INTER_T_P	858.6	121.3	50.0818	<.0001
A_P_MM_C	337.3	91.0604	13.7179	0.0002
MM_C_D	-175.7	26.0367	45.5147	<.0001

**Tabelle XI.12 : Forward Stepwise Sequence of Chi-Squares, Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Chi-Square Increment
NR_T_SUM	1	327.6	<.0001	327.6	<.0001
A_P_DURST_A	2	376.4	<.0001	48.7482	<.0001
MM_C_D	3	417.8	<.0001	41.3893	<.0001
PLANNED	4	453.9	<.0001	36.1702	<.0001
WORKED	5	485.7	<.0001	31.8123	<.0001
WEEKEND	6	521.0	<.0001	35.2253	<.0001
A_DAYT	7	557.7	<.0001	36.7359	<.0001
SINGLE	8	577.9	<.0001	20.2330	<.0001
CAR	9	594.8	<.0001	16.8513	<.0001
LICENSE	10	624.6	<.0001	29.7867	<.0001
A_P_DURST_C	11	637.9	<.0001	13.3282	0.0003
SEX	12	651.6	<.0001	13.6986	0.0002
INTER_T_P	13	656.4	<.0001	4.8513	0.0276
HEAD_OHH	14	660.3	<.0001	3.8734	0.0491
A_P_MM_C	15	662.3	<.0001	1.9670	0.1608
A_P_DIST	16	663.6	<.0001	1.3322	0.2484
A_N_MM_D	17	664.2	<.0001	0.6135	0.4335
A_P_MM_D	18	665.4	<.0001	1.1889	0.2756
G	19	665.6	<.0001	0.1828	0.6690
EDUC	20	665.8	<.0001	0.1695	0.6805
A_P_TT	21	665.8	<.0001	0.0592	0.8078
ACC_D	22	665.8	<.0001	0.00730	0.9319
AGE	23	665.8	<.0001	0.00157	0.9684

# Kategorie Dienstleistung allgemein, alle Variablen

Tabelle XI.13 : Univariate Chi-Squares for the Wilcoxon Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
SEX	-13.2573	7.0932	3.4932	0.0616
YOUNG	-14.4256	7.1463	4.0747	0.0435
LICENSE	-14.3269	7.0244	4.1599	0.0414
HEAD_OHH	-6.5971	6.5440	1.0163	0.3134
WORKED	1.3058	6.5836	0.0393	0.8428
CHILDREN	-17.3835	6.7311	6.6696	0.0098
SINGLE	7.3883	6.4267	1.3216	0.2503
CAR_POSS	-16.5324	7.0608	5.4823	0.0192
EDUC	-13.1667	6.6131	3.9641	0.0465
AGE	313.6	265.6	1.3942	0.2377
EDUC_YR	-129.1	69.6682	3.4316	0.0640
G	-81.0987	26.8910	9.0952	0.0026
CAR	-13.4612	6.9317	3.7713	0.0521
HH_NR_L	-41.8544	14.3269	8.5345	0.0035
HH_MEMB	-40.0696	15.8634	6.3803	0.0115
ACC_D	-0.1537	5.4323	0.000801	0.9774
PLANNED	4.9790	3.0507	2.6636	0.1027
A_EXP_D	4.9061	5.9139	0.6882	0.4068
WEEKEND	4.5113	5.0160	0.8089	0.3684
A_DAYT_D	-9.5129	7.0910	1.7997	0.1797
A_P_MM_D	1.0939	4.1819	0.0684	0.7937
A_N_MM_D	-1.9790	3.2179	0.3782	0.5386
A_P_MMMD	2.8544	5.9618	0.2292	0.6321
A_N_MMMD	-2.4094	5.8251	0.1711	0.6792
ACC	1.2265	7.2634	0.0285	0.8659
A_EXP	-1274.0	1731.5	0.5413	0.4619
A_DAYT	-11.0809	8.7402	1.6073	0.2049
A_P_DIST	-59815.9	71903.9	0.6920	0.4055
A_P_TT	333.6	124.3	7.2047	0.0073
A_N_DIST	-32025.8	90767.8	0.1245	0.7242
A_N_TT	315.3	193.8	2.6481	0.1037
A_P_MM	14.8722	20.6200	0.5202	0.4708
A_N_MM	4.4838	20.2008	0.0493	0.8243
NR_T_SUM	-3220.8	875.5	13.5332	0.0002
NR_J_SUM	-538.1	303.4	3.1469	0.0761
A_P_PUR	454.1	34.0532	177.8	<.0001
A_N_PUR	-5.9854	35.8021	0.0279	0.8672
INTER_T_P	153.8	55.1065	7.7911	0.0053
INTERVALL_ST	5.7064	2.0811	7.5183	0.0061
A_P_DURST_A	2301.7	224.7	104.9	<.0001
A_P_DURST_C	18.9605	10.6453	3.1724	0.0749
A_P_MM_C	28.6715	20.7672	1.9061	0.1674
MM_C_D	-22.7443	6.9695	10.6498	0.0011
INTERVALL_ST_P	19.4317	4.0056	23.5331	<.0001

**Tabelle XI.14 : Forward Stepwise Sequence of Chi-Squares, Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_PUR	1	177.8	<.0001	177.8	<.0001
A_P_DURST_A	2	213.3	<.0001	35.4321	<.0001
WORKED	3	230.6	<.0001	17.2914	<.0001
INTER_T_P	4	241.3	<.0001	10.7216	0.0011
MM_C_D	5	249.8	<.0001	8.5459	0.0035
A_DAYT_D	6	255.5	<.0001	5.6429	0.0175
INTERVALL_ST_P	7	260.1	<.0001	4.6154	0.0317
A_DAYT	8	264.1	<.0001	3.9911	0.0457
AGE	9	267.2	<.0001	3.1355	0.0766
HH_NR_L	10	274.5	<.0001	7.2388	0.0071
A_P_MM_D	11	277.7	<.0001	3.2869	0.0698
A_N_MM_D	12	280.9	<.0001	3.1252	0.0771
EDUC_YR	13	282.8	<.0001	1.9320	0.1645
A_N_DIST	14	284.6	<.0001	1.7658	0.1839
INTERVALL_ST	15	286.5	<.0001	1.9136	0.1666
PLANNED	16	288.3	<.0001	1.8095	0.1786
EDUC	17	289.9	<.0001	1.5792	0.2089
A_P_MM_C	18	291.3	<.0001	1.4209	0.2333
HEAD_OHH	19	292.2	<.0001	0.9000	0.3428
SEX	20	293.8	<.0001	1.5663	0.2107
ACC_D	21	295.0	<.0001	1.2765	0.2586
A_EXP_D	22	295.7	<.0001	0.7166	0.3973
HH_MEMB	23	296.4	<.0001	0.6621	0.4158
CHILDREN	24	299.1	<.0001	2.6513	0.1035
G	25	299.9	<.0001	0.8385	0.3598
CAR	26	300.8	<.0001	0.8544	0.3553
A_P_DIST	27	301.3	<.0001	0.5579	0.4551
NR_T_SUM	28	301.7	<.0001	0.4028	0.5257
A_N_TT	29	302.1	<.0001	0.3517	0.5532
WEEKEND	30	302.5	<.0001	0.3996	0.5273
A_P_MM	31	302.9	<.0001	0.4732	0.4915
A_N_MMMD	32	303.6	<.0001	0.7034	0.4016
YOUNG	33	303.9	<.0001	0.2875	0.5918
SINGLE	34	305.2	<.0001	1.2413	0.2652
A_P_DURST_C	35	305.4	<.0001	0.2623	0.6086
A_N_PUR	36	305.6	<.0001	0.1702	0.6799
A_P_MMMD	37	305.8	<.0001	0.1538	0.6949
A_EXP	38	305.9	<.0001	0.1366	0.7117
A_P_TT	39	306.0	<.0001	0.1214	0.7275
LICENSE	40	306.1	<.0001	0.1220	0.7269
CAR_POSS	41	306.5	<.0001	0.3869	0.5339
NR_J_SUM	42	306.7	<.0001	0.1338	0.7145
ACC	43	306.7	<.0001	0.0498	0.8234
A_N_MM	44	306.7	<.0001	0.0188	0.8909

## Kategorie Dienstleistung allgemein, Variabel aus dem 1. Gesamtmodell

**Tabelle XI.15 : Univariate Chi-Squares for the Wilcoxon Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
<b>SEX</b>	<b>-17.6830</b>	<b>7.3076</b>	<b>5.8555</b>	<b>0.0155</b>
LICENSE	-13.7963	7.2052	3.6663	0.0555
EDUC	-11.2113	6.7111	2.7908	0.0948
<b>G</b>	<b>-81.3905</b>	<b>27.5829</b>	<b>8.7070</b>	<b>0.0032</b>
<b>HH_NR_L</b>	<b>-36.8622</b>	<b>14.5802</b>	<b>6.3920</b>	<b>0.0115</b>
A_EXP_D	5.6723	6.0999	0.8647	0.3524
A_N_TT	286.8	195.2	2.1596	0.1417
A_P_MM	21.0000	21.1958	0.9816	0.3218
<b>NR_T_SUM</b>	<b>-3272.5</b>	<b>884.6</b>	<b>13.6845</b>	<b>0.0002</b>
<b>A_P_PUR</b>	<b>502.8</b>	<b>35.8206</b>	<b>197.0</b>	<b>&lt;.0001</b>
<b>INTERVALL_ST</b>	<b>6.6884</b>	<b>2.1279</b>	<b>9.8794</b>	<b>0.0017</b>
<b>A_P_DURST_A</b>	<b>2388.2</b>	<b>226.6</b>	<b>111.1</b>	<b>&lt;.0001</b>
A_P_DURST_C	19.6796	11.1179	3.1332	0.0767
<b>INTER_T_P</b>	<b>129.3</b>	<b>57.0229</b>	<b>5.1435</b>	<b>0.0233</b>
A_P_MM_C	38.4885	21.4477	3.2203	0.0727
<b>MM_C_D</b>	<b>-26.2649</b>	<b>7.1686</b>	<b>13.4241</b>	<b>0.0002</b>
<b>INTERVALL_ST_P</b>	<b>20.0170</b>	<b>4.0210</b>	<b>24.7811</b>	<b>&lt;.0001</b>

**Tabelle XI.16 : Forward Stepwise Sequence of Chi-Squares, Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
<b>A_P_PUR</b>	<b>1</b>	<b>197.0</b>	<b>&lt;.0001</b>	<b>197.0</b>	<b>&lt;.0001</b>
<b>A_P_DURST_A</b>	<b>2</b>	<b>232.5</b>	<b>&lt;.0001</b>	<b>35.4803</b>	<b>&lt;.0001</b>
<b>MM_C_D</b>	<b>3</b>	<b>243.7</b>	<b>&lt;.0001</b>	<b>11.2199</b>	<b>0.0008</b>
<b>INTER_T_P</b>	<b>4</b>	<b>250.3</b>	<b>&lt;.0001</b>	<b>6.5741</b>	<b>0.0103</b>
<b>INTERVALL_ST_P</b>	<b>5</b>	<b>256.2</b>	<b>&lt;.0001</b>	<b>5.9460</b>	<b>0.0148</b>
EDUC	6	259.0	<.0001	2.8197	0.0931
A_P_MM_C	7	261.2	<.0001	2.1160	0.1458
NR_T_SUM	8	263.1	<.0001	1.9602	0.1615
HH_NR_L	9	266.9	<.0001	3.7825	0.0518
LICENSE	10	268.3	<.0001	1.4023	0.2363
G	11	270.3	<.0001	2.0168	0.1556
INTERVALL_ST	12	271.5	<.0001	1.1909	0.2751
SEX	13	271.9	<.0001	0.3467	0.5560
A_P_DURST_C	14	272.1	<.0001	0.2210	0.6383
A_N_TT	15	272.3	<.0001	0.1748	0.6759
A_EXP_D	16	272.3	<.0001	0.0796	0.7778
A_P_MM	17	272.4	<.0001	0.0268	0.8699

**Tabelle XI.17 : Univariate Chi-Squares for the Log-Rank Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
<b>SEX</b>	<b>-24.8828</b>	<b>12.5322</b>	<b>3.9423</b>	<b>0.0471</b>
LICENSE	-20.0704	12.5717	2.5488	0.1104
EDUC	-14.3128	11.4641	1.5587	0.2119
G	-74.6678	46.3617	2.5939	0.1073
<b>HH_NR_L</b>	<b>-49.4914</b>	<b>23.9655</b>	<b>4.2647</b>	<b>0.0389</b>
A_EXP_D	17.1441	11.0027	2.4279	0.1192
A_N_TT	383.7	302.0	1.6147	0.2038
A_P_MM	45.2705	37.5998	1.4496	0.2286
<b>NR_T_SUM</b>	<b>-4902.8</b>	<b>1590.8</b>	<b>9.4982</b>	<b>0.0021</b>
<b>A_P_PUR</b>	<b>532.6</b>	<b>55.6944</b>	<b>91.4363</b>	<b>&lt;.0001</b>
<b>INTERVALL_ST</b>	<b>10.6349</b>	<b>3.5621</b>	<b>8.9136</b>	<b>0.0028</b>
<b>A_P_DURST_A</b>	<b>3732.1</b>	<b>747.8</b>	<b>24.9087</b>	<b>&lt;.0001</b>
A_P_DURST_C	45.1749	28.2406	2.5589	0.1097
<b>INTER_T_P</b>	<b>269.1</b>	<b>113.1</b>	<b>5.6577</b>	<b>0.0174</b>
<i>A_P_MM_C</i>	<i>64.7856</i>	<i>38.3493</i>	<i>2.8539</i>	<i>0.0912</i>
<b>MM_C_D</b>	<b>-38.5539</b>	<b>12.5769</b>	<b>9.3969</b>	<b>0.0022</b>
<b>INTERVALL_ST_P</b>	<b>24.2623</b>	<b>8.0594</b>	<b>9.0628</b>	<b>0.0026</b>

**Tabelle XI.18 : Forward Stepwise Sequence of Chi-Squares, Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
<b>A_P_PUR</b>	<b>1</b>	<b>91.4</b>	<b>&lt;.0001</b>	<b>91.4363</b>	<b>&lt;.0001</b>
<b>MM_C_D</b>	<b>2</b>	<b>104.7</b>	<b>&lt;.0001</b>	<b>13.2769</b>	<b>0.0003</b>
<b>INTER_T_P</b>	<b>3</b>	<b>110.5</b>	<b>&lt;.0001</b>	<b>5.8232</b>	<b>0.0158</b>
<b>INTERVALL_ST_P</b>	<b>4</b>	<b>115.3</b>	<b>&lt;.0001</b>	<b>4.8127</b>	<b>0.0283</b>
<i>HH_NR_L</i>	<i>5</i>	<i>118.5</i>	<i>&lt;.0001</i>	<i>3.1159</i>	<i>0.0775</i>
<i>A_P_DURST_A</i>	<i>6</i>	<i>121.2</i>	<i>&lt;.0001</i>	<i>2.7622</i>	<i>0.0965</i>
A_P_MM_C	7	123.3	<.0001	2.1148	0.1459
LICENSE	8	124.8	<.0001	1.4616	0.2267
NR_T_SUM	9	125.5	<.0001	0.6839	0.4083
A_P_MM	10	126.3	<.0001	0.8114	0.3677
EDUC	11	126.7	<.0001	0.3699	0.5430
G	12	127.0	<.0001	0.3718	0.5420
A_EXP_D	13	127.3	<.0001	0.2817	0.5956
A_N_TT	14	127.5	<.0001	0.2156	0.6424
INTERVALL_ST	15	127.9	<.0001	0.3258	0.5681
SEX	16	127.9	<.0001	0.0362	0.8492
A_P_DURST_C	17	127.9	<.0001	0.0001	0.9919

# Kategorie Treffen mit Freunden, alle Variablen

Tabelle XI.19 : Univariate Chi-Squares for the Wilcoxon Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
<b>SEX</b>	<b>-21.7017</b>	<b>5.7210</b>	<b>14.3892</b>	<b>0.0001</b>
<b>YOUNG</b>	<b>-16.9804</b>	<b>5.4102</b>	<b>9.8510</b>	<b>0.0017</b>
LICENSE	-3.6895	5.7712	0.4087	0.5226
HEAD_OHH	-4.9804	5.3716	0.8597	0.3538
WORKED	-0.3888	5.4496	0.00509	0.9431
CHILDREN	3.6553	5.2038	0.4934	0.4824
<b>SINGLE</b>	<b>-21.7359</b>	<b>5.7958</b>	<b>14.0648</b>	<b>0.0002</b>
CAR_POSS	-4.9120	5.7784	0.7226	0.3953
EDUC	-3.8582	5.1165	0.5686	0.4508
<b>AGE</b>	<b>951.2</b>	<b>234.1</b>	<b>16.5162</b>	<b>&lt;.0001</b>
EDUC_YR	48.9413	47.3582	1.0680	0.3014
G	-4.1418	18.0109	0.0529	0.8181
<b>CAR</b>	<b>-11.9584</b>	<b>5.3866</b>	<b>4.9285</b>	<b>0.0264</b>
HH_NR_L	-4.3594	9.1262	0.2282	0.6329
<b>HH_MEMB</b>	<b>29.9976</b>	<b>13.2176</b>	<b>5.1507</b>	<b>0.0232</b>
<b>ACC_D</b>	<b>22.0293</b>	<b>4.9008</b>	<b>20.2052</b>	<b>&lt;.0001</b>
PLANNED	1.5403	3.2672	0.2223	0.6373
A_EXP_D	-1.2689	1.9658	0.4167	0.5186
WEEKEND	-7.2738	5.6650	1.6487	0.1991
A_DAYT_D	3.1002	4.4309	0.4896	0.4841
<b>A_P_MM_D</b>	<b>-7.5575</b>	<b>3.8434</b>	<b>3.8664</b>	<b>0.0493</b>
A_N_MM_D	1.4743	3.9061	0.1425	0.7058
<b>A_P_MMM_D</b>	<b>-11.8851</b>	<b>5.8111</b>	<b>4.1829</b>	<b>0.0408</b>
A_N_MMM_D	-5.6186	5.8141	0.9339	0.3339
<b>ACC</b>	<b>35.5232</b>	<b>8.5241</b>	<b>17.3671</b>	<b>&lt;.0001</b>
A_EXP	6.4792	41.2896	0.0246	0.8753
A_DAYT	-6.9878	8.4387	0.6857	0.4076
A_P_DIST	5014.9	181008	0.000768	0.9779
<b>A_P_TT</b>	<b>630.8</b>	<b>195.3</b>	<b>10.4349</b>	<b>0.0012</b>
A_N_DIST	90491.2	178111	0.2581	0.6114
A_N_TT	127.7	262.9	0.2362	0.6270
A_P_MM	-11.9364	18.3697	0.4222	0.5158
A_N_MM	-8.3081	18.3325	0.2054	0.6504
<b>NR_T_SUM</b>	<b>-2239.8</b>	<b>710.2</b>	<b>9.9475</b>	<b>0.0016</b>
<i>NR_J_SUM</i>	<i>-444.3</i>	<i>266.5</i>	<i>2.7805</i>	<i>0.0954</i>
A_P_PUR	3.6015	25.9410	0.0193	0.8896
<b>A_N_PUR</b>	<b>60.6064</b>	<b>20.9122</b>	<b>8.3991</b>	<b>0.0038</b>
<b>INTER_T_P</b>	<b>84.7281</b>	<b>43.2051</b>	<b>3.8458</b>	<b>0.0499</b>
INTERVALL_ST	4.2619	8.3725	0.2591	0.6107
<b>A_P_DURST_A</b>	<b>506.5</b>	<b>100.6</b>	<b>25.3331</b>	<b>&lt;.0001</b>
A_P_DURST_C	12.3820	36.6163	0.1143	0.7352
A_P_MM_C	-1.2396	18.6579	0.00441	0.9470
<b>MM_C_D</b>	<b>-17.7971</b>	<b>5.7011</b>	<b>9.7448</b>	<b>0.0018</b>
INTERVALL_ST_P	-1.0577	4.9761	0.0452	0.8317

**Tabelle XI.20 : Forward Stepwise Sequence of Chi-Squares, Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_DURST_A	1	25.3331	<.0001	25.3331	<.0001
ACC_D	2	49.3278	<.0001	23.9947	<.0001
SEX	3	57.1175	<.0001	7.7897	0.0053
SINGLE	4	65.0582	<.0001	7.9407	0.0048
A_N_PUR	5	71.1028	<.0001	6.0446	0.0139
MM_C_D	6	76.4051	<.0001	5.3023	0.0213
A_P_PUR	7	81.8400	<.0001	5.4349	0.0197
WORKED	8	86.3828	<.0001	4.5428	0.0331
AGE	9	92.6650	<.0001	6.2822	0.0122
YOUNG	10	97.1632	<.0001	4.4982	0.0339
A_P_MM_D	11	101.1	<.0001	3.9537	0.0468
A_DAYT_D	12	104.1	<.0001	3.0071	0.0829
EDUC_YR	13	106.1	<.0001	1.9322	0.1645
NR_T_SUM	14	108.0	<.0001	1.9216	0.1657
HH_NR_L	15	110.8	<.0001	2.8166	0.0933
A_P_MM_C	16	113.0	<.0001	2.2398	0.1345
A_P_MMMD	17	115.0	<.0001	1.9572	0.1618
A_P_MM	18	118.0	<.0001	2.9976	0.0834
A_P_TT	19	119.4	<.0001	1.4461	0.2292
A_P_DIST	20	122.3	<.0001	2.8913	0.0891
CAR	21	123.4	<.0001	1.0616	0.3028
LICENSE	22	126.0	<.0001	2.6128	0.1060
WEEKEND	23	127.1	<.0001	1.0716	0.3006
HEAD_OHH	24	127.8	<.0001	0.7118	0.3988
EDUC	25	128.6	<.0001	0.7971	0.3720
ACC	26	129.2	<.0001	0.6147	0.4330
A_N_MM_D	27	129.8	<.0001	0.6161	0.4325
A_N_MM	28	130.2	<.0001	0.4296	0.5122
CAR_POSS	29	130.6	<.0001	0.3636	0.5465
CHILDREN	30	131.0	<.0001	0.3491	0.5546
G	31	133.6	<.0001	2.6386	0.1043
A_DAYT	32	133.9	<.0001	0.2665	0.6057
INTER_T_P	33	134.1	<.0001	0.2427	0.6222
PLANNED	34	134.3	<.0001	0.1681	0.6818
A_N_MMMD	35	134.4	<.0001	0.1641	0.6854
A_P_DURST_C	36	134.6	<.0001	0.1206	0.7284
A_EXP	37	134.6	<.0001	0.0913	0.7625
A_EXP_D	38	134.8	<.0001	0.1346	0.7137
NR_J_SUM	39	134.9	<.0001	0.0993	0.7527
INTERVALL_ST	40	135.0	<.0001	0.0880	0.7667
HH_MEMB	41	135.0	<.0001	0.0601	0.8063
A_N_DIST	42	135.1	<.0001	0.0517	0.8202
INTERVALL_ST_P	43	135.1	<.0001	0.0126	0.9107
A_N_TT	44	135.1	<.0001	0.00801	0.9287

## Kategorie Treffen mit Freunden, Variablen aus dem 1. Gesamtmodell

**Tabelle XI.21 : Univariate Chi-Squares for the Wilcoxon Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
SEX	-25.9350	5.8889	19.3955	<.0001
SINGLE	-26.2892	6.0544	18.8545	<.0001
AGE	925.6	238.9	15.0083	0.0001
CAR	-12.8812	5.6021	5.2871	0.0215
ACC_D	24.7780	5.0045	24.5141	<.0001
PLANNED	3.0471	3.5424	0.7399	0.3897
A_P_TT	588.5	197.9	8.8387	0.0029
NR_T_SUM	-2500.4	716.1	12.1931	0.0005
INTERVALL_ST	3.6875	8.4670	0.1897	0.6632
A_P_DURST_A	648.2	110.9	34.1419	<.0001
INTER_T_P	106.4	44.9903	5.5963	0.0180

**Tabelle XI.22 : Forward Stepwise Sequence of Chi-Squares, Wilcoxon Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
A_P_DURST_A	1	34.1419	<.0001	34.1419	<.0001
ACC_D	2	63.7411	<.0001	29.5992	<.0001
SEX	3	73.8928	<.0001	10.1517	0.0014
SINGLE	4	82.9066	<.0001	9.0138	0.0027
A_P_TT	5	86.6164	<.0001	3.7098	0.0541
NR_T_SUM	6	89.2027	<.0001	2.5863	0.1078
AGE	7	89.7421	<.0001	0.5394	0.4627
CAR	8	90.0434	<.0001	0.3013	0.5831
INTERVALL_ST	9	90.1114	<.0001	0.0680	0.7943
PLANNED	10	90.1549	<.0001	0.0435	0.8347
INTER_T_P	11	90.1684	<.0001	0.0135	0.9076

**Tabelle XI.23 : Univariate Chi-Squares for the Log-Rank Test**

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
<b>SEX</b>	<b>-42.9502</b>	<b>10.4200</b>	<b>16.9902</b>	<b>&lt;.0001</b>
<b>SINGLE</b>	<b>-30.4182</b>	<b>10.3491</b>	<b>8.6390</b>	<b>0.0033</b>
<b>AGE</b>	<b>1035.6</b>	<b>400.5</b>	<b>6.6874</b>	<b>0.0097</b>
<i>CAR</i>	<i>-17.3443</i>	<i>9.7533</i>	<i>3.1624</i>	<i>0.0754</i>
<b>ACC_D</b>	<b>46.0804</b>	<b>9.7827</b>	<b>22.1879</b>	<b>&lt;.0001</b>
PLANNED	6.4320	5.7004	1.2732	0.2592
<b>A_P_TT</b>	<b>863.9</b>	<b>424.4</b>	<b>4.1430</b>	<b>0.0418</b>
<b>NR_T_SUM</b>	<b>-4597.1</b>	<b>1204.8</b>	<b>14.5599</b>	<b>0.0001</b>
INTERVALL_ST	6.4816	9.8378	0.4341	0.5100
<b>A_P_DURST_A</b>	<b>1513.3</b>	<b>467.9</b>	<b>10.4621</b>	<b>0.0012</b>
<b>INTER_T_P</b>	<b>207.4</b>	<b>84.2550</b>	<b>6.0587</b>	<b>0.0138</b>

**Tabelle XI.24 : Forward Stepwise Sequence of Chi-Squares, Log-Rank Test**

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
<b>ACC_D</b>	<b>1</b>	<b>22.1879</b>	<b>&lt;.0001</b>	<b>22.1879</b>	<b>&lt;.0001</b>
<b>SEX</b>	<b>2</b>	<b>38.8342</b>	<b>&lt;.0001</b>	<b>16.6463</b>	<b>&lt;.0001</b>
<b>NR_T_SUM</b>	<b>3</b>	<b>46.2337</b>	<b>&lt;.0001</b>	<b>7.3995</b>	<b>0.0065</b>
<b>SINGLE</b>	<b>4</b>	<b>50.4962</b>	<b>&lt;.0001</b>	<b>4.2626</b>	<b>0.0390</b>
A_P_DURST_A	5	52.9803	<.0001	2.4840	0.1150
A_P_TT	6	55.4586	<.0001	2.4784	0.1154
INTER_T_P	7	56.1903	<.0001	0.7317	0.3923
INTERVALL_ST	8	56.7857	<.0001	0.5954	0.4404
AGE	9	56.8958	<.0001	0.1101	0.7400
PLANNED	10	56.9614	<.0001	0.0656	0.7978
CAR	11	56.9614	<.0001	0.000017	0.9967

## **Anhang XII : Erste Gesamtmodelle der semiparametrischen Modellschätzungen, Intervall dauern**

Auf den folgenden Seiten wird in jeder der vier ausgewählten Kategorien ein erstes Gesamtmodell präsentiert. Aus der Liste der insgesamt 57 möglichen Einflußvariablen sind jeweils durch singuläre Variablenanalyse, Korrelationsmatrizen und Proportionalitätstests diejenigen bestimmt worden, die in einem ersten Gesamtmodell als Kovariablen verwendet werden können. Die Modellinformationen werden hier bewußt nicht näher gezeigt, sie sind ausführlich im Zusatzanhang XXIV (PHREG I Semiparametrische Modelle) zu sehen.

Um eine rasche Gegenüberstellung mit den Auswertungen der vollparametrischen Modellschätzungen für Intervall dauern zu gewährleisten, sind diese Ergebnisse im Anhang XIII (LIFEREG I Vollparametrische Modelle) zu sehen.

In einem weiteren Schritt wird ein Modell der zweiten Stufe berechnet, in dem nur noch die im ersten Gesamtmodell statistisch signifikanten Variablen enthalten sind. Die Ergebnisse des zweiten Modells sind im Zusatzanhang XXIV (LIFEREG I Semiparametrische Modelle) angeführt bzw. direkt im Kapitel 9.3.3 diskutiert.

## Kategorie Einkauf kurzfristig

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
<b>SEX</b>	-0.29932	0.04512	44.0142	<b>&lt;.0001</b>	0.741
<b>LICENSE</b>	0.16695	0.07038	5.6263	<b>0.0177</b>	1.182
<b>AGE</b>	0.00243	0.00122	3.9709	<b>0.0463</b>	1.002
CAR	0.11556	0.07260	2.5331	0.1115	1.122
<i>HH_NR_L</i>	-0.08137	0.04728	2.9620	<i>0.0852</i>	0.922
<b>PLANNED</b>	-0.18763	0.05977	9.8559	<b>0.0017</b>	0.829
<b>WEEKEND</b>	0.18730	0.04775	15.3843	<b>&lt;.0001</b>	1.206
A_EXP	-0.00011	0.00042	0.0734	0.7864	1.000
<b>A_DAYT</b>	-0.09497	0.03666	6.7109	<b>0.0096</b>	0.909
A_P_DIST	-5.16E-6	7.86E-6	0.4309	0.5116	1.000
<b>A_P_TT</b>	-0.00795	0.00271	8.5823	<b>0.0034</b>	0.992
A_P_MM	-0.02085	0.01673	1.5533	0.2126	0.979
NR_J_SUM	-0.00031	0.00108	0.0809	0.7760	1.000
A_P_PUR	0.01150	0.00735	2.4491	0.1176	1.012
<i>INTER_T_P</i>	-0.01581	0.00902	3.0755	<i>0.0795</i>	0.984
INTERVALL_ST	-0.07630	0.08040	0.9006	0.3426	0.927
<b>A_P_DURST_A</b>	-0.02052	0.00198	107.9229	<b>&lt;.0001</b>	0.980
<i>A_P_DURST_C</i>	-0.05977	0.03217	3.4518	<i>0.0632</i>	0.942
A_P_MM_C	-0.01837	0.01583	1.3469	0.2458	0.982
<b>MM_C_D</b>	0.11643	0.04291	7.3615	<b>0.0067</b>	1.123
<b>INTERVALL_ST_P</b>	-0.49170	0.09628	26.0836	<b>&lt;.0001</b>	0.612

---

## Kategorie Dienstleistung allgemein

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
<b>SEX</b>	0.17837	0.05834	9.3483	<b>0.0022</b>	1.195
<b>LICENSE</b>	-0.35792	0.07221	24.5711	<b>&lt;.0001</b>	0.699
HEAD_OHH	-0.08294	0.07122	1.3561	0.2442	0.920
<b>WORKED</b>	-0.22919	0.06012	14.5345	<b>0.0001</b>	0.795
<b>SINGLE</b>	0.37307	0.06872	29.4750	<b>&lt;.0001</b>	1.452
EDUC	0.05883	0.04535	1.6828	0.1946	1.061
AGE	0.00123	0.00201	0.3744	0.5406	1.001
G	0.00765	0.02077	0.1358	0.7125	1.008
<b>CAR</b>	0.41454	0.08142	25.9195	<b>&lt;.0001</b>	1.514
ACC_D	0.00176	0.04696	0.0014	0.9701	1.002
<b>PLANNED</b>	-0.41735	0.07053	35.0153	<b>&lt;.0001</b>	0.659
<b>WEEKEND</b>	0.30322	0.04307	49.5572	<b>&lt;.0001</b>	1.354
A_P_MM_D	0.08521	0.08102	1.1060	0.2930	1.089
A_N_MM_D	-0.01055	0.08408	0.0157	0.9002	0.990
<b>A_DAYT</b>	0.16890	0.02952	32.7268	<b>&lt;.0001</b>	1.184
A_P_DIST	-6.65E-7	9.20E-7	0.5223	0.4698	1.000
A_P_TT	-0.00021	0.00055	0.1520	0.6966	1.000
<b>NR_T_SUM</b>	0.00511	0.00038	179.8198	<b>&lt;.0001</b>	1.005
A_P_DURST_A	-0.01311	0.00180	52.7910	<b>&lt;.0001</b>	0.987
A_P_DURST_C	-0.04339	0.00933	21.6178	<b>&lt;.0001</b>	0.958
<b>INTER_T_P</b>	-0.02239	0.01034	4.6859	<b>0.0304</b>	0.978
A_P_MM_C	0.01794	0.01246	2.0726	0.1500	1.018
<b>MM_C_D</b>	0.23881	0.04060	34.6037	<b>&lt;.0001</b>	1.270

---

## Kategorie Freizeit allgemein

---

1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
SEX	-0.00597	0.09273	0.0041	0.9487	0.994
LICENSE	-0.16697	0.17420	0.9187	0.3378	0.846
EDUC	0.05704	0.09417	0.3668	0.5447	1.059
G	-0.02234	0.02729	0.6702	0.4130	0.978
<b>HH_NR_L</b>	0.19618	0.09475	4.2871	<b>0.0384</b>	1.217
A_EXP_D	-0.03434	0.09725	0.1247	0.7240	0.966
A_N_TT	0.00253	0.00376	0.4526	0.5011	1.003
A_P_MM	-0.03082	0.03193	0.9317	0.3344	0.970
NR_T_SUM	-0.00072	0.00079	0.8238	0.3641	0.999
<b>A_P_PUR</b>	-0.14254	0.01813	61.7925	<b>&lt;.0001</b>	0.867
INTERVALL_ST	-0.23966	0.34493	0.4828	0.4872	0.787
<b>A_P_DURST_A</b>	-0.00526	0.00246	4.5500	<b>0.0329</b>	0.995
A_P_DURST_C	0.00104	0.04483	0.0005	0.9814	1.001
<b>INTER_T_P</b>	-0.02308	0.00995	5.3766	<b>0.0204</b>	0.977
A_P_MM_C	-0.04291	0.03162	1.8420	0.1747	0.958
<b>MM_C_D</b>	0.27701	0.08871	9.7518	<b>0.0018</b>	1.319
INTERVALL_ST_P	-0.31361	0.17449	3.2303	0.0723	0.731

---

## Kategorie Treffen mit Freunden

---

1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
<b>SEX</b>	0.32096	0.11737	7.4776	<b>0.0062</b>	1.378
SINGLE	0.17449	0.10831	2.5953	0.1072	1.191
AGE	-0.00029	0.00303	0.0092	0.9234	1.000
CAR	-0.02599	0.13936	0.0348	0.8521	0.974
<b>ACC_D</b>	-0.41378	0.12096	11.7027	<b>0.0006</b>	0.661
PLANNED	-0.04592	0.16630	0.0762	0.7824	0.955
<i>A_P_TT</i>	-0.00509	0.00307	2.7449	<i>0.0976</i>	0.995
<b>NR_T_SUM</b>	0.00180	0.00085	4.4527	<b>0.0348</b>	1.002
INTERVALL_ST	0.05595	0.08522	0.4310	0.5115	1.058
<b>A_P_DURST_A</b>	-0.00816	0.00389	4.4071	<b>0.0358</b>	0.992
INTER_T_P	-0.01097	0.01349	0.6616	0.4160	0.989

---

## **Anhang XIII : Erste Gesamtmodelle der vollparametrischen Modellschätzungen, Intervall dauern**

Auf den folgenden Seiten werden in jeder der vier ausgewählten Kategorien die Modellinformationen und ein erstes Gesamtmodell präsentiert. Aus der Liste der möglichen Einflußvariablen sind nach einer Einzelanalyse (siehe semiparametrische Modellschätzungen) jeweils mit Hilfe von Korrelationsmatrizen diejenigen bestimmt worden, die in einem ersten Gesamtmodell als Kovariablen verwendet werden können. Durch Proportionalitätstest sind sie ein weiteres Mal untersucht und dann in ein erstes Gesamtmodell integriert worden.

Alle notwendigen Überprüfungen in bezug auf die gewählte Verteilungsannahme sehen Sie im Zusatzanhang XXV (LIFEREG I Vollparametrische Modelle).

Um einen guten Vergleich mit den Auswertungen der semiparametrischen Modelle zu erhalten, sind diese Ergebnisse der ersten Gesamtmodelle für die entsprechenden vier Kategorien im Anhang XII (PHREG I Erste Gesamtmodelle der semiparametrischen Auswertung) zu sehen.

In einem weiteren Schritt wird ein Modell der zweiten Stufe berechnet, in dem nur noch die statistisch signifikanten Variablen des ersten Gesamtmodells enthalten sind. Die Ergebnisse des Modells zweiter Stufe sind im Zusatzanhang XXV (LIFEREG I Vollparametrische Modelle) bzw. zum besseren Vergleich im Zusatzanhang XXIX und direkt im Kapitel 9.2.4 zu sehen.

## Kategorie Einkauf kurzfristig : Weibullverteilungsmodell mit Kovariablen

---

### Modell Information

Data Set	WORK.DUMMY6I	
Dependent Variable	Log (inter_t)	Intervall zwischen zwei Aktivitäten in Tagen
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	2917	
Noncensored Values	2917	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Missing Values	139	
Zero or Negative Response	3	
Name of Distribution	WEIBULL	
Log Likelihood	-6377.590077	

---



---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	-1.03363	0.22711	20.7127	<.0001
<b>SEX</b>	0.46610	0.08025	33.7310	<b>&lt;.0001</b>
LICENSE	-0.19049	0.12530	2.3113	0.1284
<b>AGE</b>	-0.00521	0.00218	5.7134	<b>0.0168</b>
CAR	-0.23387	0.12909	3.2825	0.0700
HH_NR_L	0.12165	0.08313	2.1413	0.1434
<b>PLANNED</b>	0.34624	0.10660	10.5492	<b>0.0012</b>
<b>WEEKEND</b>	-0.30504	0.08481	12.9374	<b>0.0003</b>
A_EXP	-0.00001	0.00074	0.0003	0.9859
A_DAYT	0.03891	0.06508	0.3575	0.5499
A_P_DIST	3.23E-6	0.00001	0.0540	0.8163
<b>A_P_TT</b>	0.01194	0.00479	6.1999	<b>0.0128</b>
A_P_MM	0.04726	0.02981	2.5137	0.1129
NR_J_SUM	0.00324	0.00198	2.6749	0.1019
A_P_PUR	-0.01943	0.01309	2.2008	0.1379
INTER_T_P	0.02640	0.01593	2.7447	0.0976
INTERVALL_ST	0.11002	0.13986	0.6188	0.4315
<b>A_P_DURST_A</b>	0.04046	0.00385	110.4144	<b>&lt;.0001</b>
A_P_DURST_C	0.09408	0.05462	2.9669	0.0850
A_P_MM_C	0.01134	0.02811	0.1627	0.6867
<b>MM_C_D</b>	-0.21789	0.07649	8.1147	<b>0.0044</b>
<b>INTERVALL_ST_P</b>	0.88378	0.17318	26.0426	<b>&lt;.0001</b>
SCALE	1.78662	0.02733		

---

n = 2917; Log-Likelihood = - 6377.59

## Kategorie Freizeit allgemein: Weibullverteilungsmodell mit Kovariablen

---

### Modell Information

Data Set	WORK.DUMMY8I	
Dependent Variable	Log (inter_t)	Intervall zw. zwei Aktivitäten in Tagen
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	2654	
Noncensored Values	2654	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Missing Values	136	
Zero or Negative Response	49	
Name of Distribution	WEIBULL	
Log Likelihood	-4935.695937	

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	0.96429	0.29005	11.0525	0.0009
<b>SEX</b>	-0.25560	0.07786	10.7763	<b>0.0010</b>
<b>LICENSE</b>	0.41738	0.09575	19.0007	<b>&lt;.0001</b>
HEAD_OHH	0.15320	0.09489	2.6065	0.1064
<b>WORKED</b>	0.27841	0.07998	12.1168	<b>0.0005</b>
<b>SINGLE</b>	-0.47556	0.09117	27.2111	<b>&lt;.0001</b>
EDUC	-0.09350	0.06036	2.3995	0.1214
AGE	-0.00180	0.00268	0.4507	0.5020
G	-0.00919	0.02775	0.1098	0.7404
<b>CAR</b>	-0.50296	0.10788	21.7353	<b>&lt;.0001</b>
ACC_D	0.00973	0.06282	0.0240	0.8769
<b>PLANNED</b>	0.51095	0.09660	27.9767	<b>&lt;.0001</b>
<b>WEEKEND</b>	-0.38183	0.05776	43.7019	<b>&lt;.0001</b>
A_P_MM_D	-0.07535	0.10738	0.4924	0.4829
A_N_MM_D	-0.00134	0.11230	0.0001	0.9905
<b>A_DAYT</b>	-0.27574	0.03896	50.0818	<b>&lt;.0001</b>
A_P_DIST	1.23E-6	1.28E-6	0.9179	0.3380
A_P_TT	0.00097	0.00076	1.6202	0.2031
<b>NR_T_SUM</b>	-0.00618	0.00049	159.0835	<b>&lt;.0001</b>
<b>A_P_DURST_A</b>	0.01713	0.00244	49.3610	<b>&lt;.0001</b>
<b>A_P_DURST_C</b>	0.06135	0.01247	24.2050	<b>&lt;.0001</b>
INTER_T_P	0.02020	0.01357	2.2160	0.1366
A_P_MM_C	-0.01506	0.01668	0.8153	0.3665
<b>MM_C_D</b>	-0.28862	0.05408	28.4793	<b>&lt;.0001</b>
SCALE	1.32344	0.02027		

---

n = 2654; Log-Likelihood = - 4935.70

## Kategorie Dienstleistung allgemein: Lognormalverteilung mit Kovariablen

---

### Modell Information

Data Set	WORK.DUMMY2I	
Dependent Variable	Log (inter_t)	Intervall zw. zwei Aktivitäten in Tagen
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	652	
Noncensored Values	652	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Missing Values	102	
Name of Distribution	LNORMAL	
Log Likelihood	-1344.449912	

---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	-4.49105	0.44909	100.0080	<.0001
SEX	-0.11378	0.18279	0.3874	0.5337
LICENSE	0.33310	0.32906	1.0247	0.3114
EDUC	-0.15664	0.18259	0.7360	0.3910
G	0.02823	0.05375	0.2759	0.5994
<b>HH_NR_L</b>	-0.34573	0.17542	3.8842	<b>0.0487</b>
A_EXP_D	-0.09969	0.18361	0.2948	0.5872
A_N_TT	-0.00328	0.00647	0.2570	0.6122
A_P_MM	-0.01538	0.05925	0.0674	0.7952
NR_T_SUM	0.00261	0.00157	2.7500	0.0973
<b>A_P_PUR</b>	0.48209	0.02809	294.4596	<b>&lt;.0001</b>
INTERVALL_ST	0.62324	0.58949	1.1178	0.2904
<b>A_P_DURST_A</b>	0.01262	0.00440	8.2233	<b>0.0041</b>
A_P_DURST_C	-0.04694	0.09179	0.2615	0.6091
INTER_T_P	0.03169	0.01889	2.8137	0.0935
A_P_MM_C	0.10629	0.05868	3.2814	0.0701
<b>MM_C_D</b>	-0.51841	0.16700	9.6367	<b>0.0019</b>
<b>INTERVALL_ST_P</b>	0.72064	0.29157	6.1088	<b>0.0135</b>
SCALE	1.90237	0.05268		

---

n = 652; Log-Likelihood = - 1344.45

Conclusio: Das Modell mit lognormaler Verteilung kommt den in den semiparametrischen Auswertungen gemachten Modellanalysen wesentlich näher als das Modell mit Weibullverteilung. Zudem sind wesentlich mehr Variablen statistisch signifikant und können somit in ein Modell der zweiten Stufe übertragen werden. Es wird demnach ein Lognormales Verteilungsmodell verwendet.

## Kategorie Treffen mit Freunden: Zwei Möglichkeiten

### Weibullverteilungsmodell mit Kovariablen

---

#### Modell Information

Data Set	WORK.DUMMY2I	
Dependent Variable	Log (inter_t)	Intervall zw. zwei Aktivitäten in Tagen
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	443	
Noncensored Values	443	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Missing Values	90	
Zero or Negative Response	1	
Name of Distribution	WEIBULL	
Log Likelihood	-834.5870682	

---



---

#### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	1.00586	0.38225	6.9243	0.0085
<b>SEX</b>	-0.37505	0.15153	6.1262	<b>0.0133</b>
<i>SINGLE</i>	-0.26135	0.14055	3.4580	<i>0.0629</i>
AGE	0.00128	0.00396	0.1045	0.7465
CAR	0.02661	0.17957	0.0220	0.8822
<b>ACC_D</b>	0.45343	0.15672	8.3712	<b>0.0038</b>
PLANNED	0.01144	0.21693	0.0028	0.9580
A_P_TT	0.00617	0.00396	2.4213	0.1197
NR_T_SUM	-0.00179	0.00112	2.5847	0.1079
INTERVALL_ST	-0.05198	0.10853	0.2294	0.6320
<i>A_P_DURST_A</i>	0.00855	0.00493	3.0095	<i>0.0828</i>
INTER_T_P	0.01230	0.01749	0.4950	0.4817
SCALE	1.30836	0.05082		

---

n = 443; Log-Likelihood = - 834.59

## Standardisiertes Gamma-Verteilungsmodell mit Kovariablen

---

### Modell Information

Data Set	WORK.DUMMY2I	
Dependent Variable	Log (inter_t)	Intervall zw. zwei Aktivitäten in Tagen
Censoring Variable	STATUS	
Censoring Value(s)	0	
Number of Observations	443	
Noncensored Values	443	
Right Censored Values	0	
Left Censored Values	0	
Interval Censored Values	0	
Missing Values	90	
Zero or Negative Response	1	
Name of Distribution	ST. GAMMA	
Log Likelihood	-828.3766287	

---



---

### 1. Gesamtmodell: Analyse der Parameter-Schätzwerte

Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
INTERCEPT	1.15311	0.35986	10.2677	0.0014
<b>SEX</b>	-0.36791	0.14449	6.4830	<b>0.0109</b>
SINGLE	-0.21598	0.13394	2.6004	0.1068
AGE	0.00010	0.00375	0.0007	0.9790
CAR	0.04658	0.17028	0.0748	0.7844
<b>ACC_D</b>	0.43419	0.14809	8.5967	<b>0.0034</b>
PLANNED	0.02293	0.20573	0.0124	0.9112
A_P_TT	0.00577	0.00377	2.3417	0.1259
NR_T_SUM	-0.00176	0.00104	2.8480	<i>0.0915</i>
INTERVALL_ST	-0.06669	0.10521	0.4018	0.5262
A_P_DURST_A	0.00754	0.00465	2.6267	0.1051
INTER_T_P	0.01405	0.01663	0.7141	0.3981
SCALE	1.24000	0		
SHAPE	1.24000	0		

---

n = 443; Log-Likelihood = - 828.38

---

### Lagrange Multiplier Statistics

Variable	Chi-Square	Pr > ChiSq
Scale	0.3861	0.5344
Shape1	2.0792	0.1493

---

## Anhang XIV : Übersichtstabellen signifikante Variablen, Intervall dauern

In diesem Anhang werden zum Vergleich die in den nonparametrischen stufenweisen Chi-Quadrat-Tests bzw. die in den semiparametrischen und vollparametrischen Modellschätzungen verwendeten möglichen Einflußvariablen für alle vier ausgewählten Aktivitätskategorien Einkauf kurzfristig, Freizeit allgemein, Dienstleistung allgemein und Treffen mit Freunden tabellarisch dargestellt.

In der ersten Tabelle sind die Ergebnisse der stufenweisen Chi-Quadrat-Tests (*Forward Stepwise Sequence of Chi-Squares*) der Variablen des ersten Gesamtmodells für den jeweilig günstigeren Test (entweder Log-Rank oder Wilcoxon Test) für alle vier Kategorien dargestellt. Nachdem ein Großteil der Intervalle in allen vier Kategorien nur sehr kurz andauert (rund  $\frac{3}{4}$  der Intervalle werden innerhalb des ersten Viertels der Zeitachse beendet), werden in allen Kategorien die Ergebnisse des Wilcoxon-Tests präsentiert, weil dabei wesentlich mehr Gewichtung auf frühe Ereigniszeiten bzw. kurze Intervall dauern gelegt wird. In der zweiten und dritten Tabelle sind übersichtlich die jeweils in der zweiten Modellstufe signifikanten Variablen, sowohl marginal signifikant im 90%-Konfidenzintervall (Signifikanzwert unter  $\alpha = 0.10$ ) mit \*\*, als auch stark signifikant im 95%-Konfidenzintervall (Signifikanzwert unter  $\alpha = 0.05$ ) mit \* gekennzeichnet.

Die exakten Werte und andere Ergebnisse sieht man im Zusatzanhang XXIII (LIFETEST I Logrank und Wilcoxon-Test), im Zusatzanhang XXIV (PHREG I Semiparametrische Modelle) sowie im Zusatzanhang XXV (LIFEREG I Vollparametrische Modelle).

Die hier dargestellten Ergebnisse gelten nur für die Einflüsse der Variablen auf die Intervall dauern zwischen zwei Aktivitäten der gleichen Kategorie, nicht aber auf die Aktivitätsdauern. Diese Ergebnisse sind im Anhang IX (LIFETEST AD Übersichtstabellen signifikante Variablen) dargestellt.

## Übersicht der Signifikanz der möglichen Einflußvariablen, nonparametrische Modellschätzung

Variable	Einkauf kurzfristig	Freizeit allgemein	Dienstl. allgemein	Treffen mit Freunden	Dummy	Variablenname
SEX	*	*		*	X	Geschlecht
AGE_**					X	Altersgruppe
AGE	*					Alter
YOUNG					X	Unter 50 Jahren j/n
LICENSE		*			X	Führerscheinbesitz j/n
HEAD_O_HH					X	Haushaltsvorstand j/n
WORKED		*			X	arbeitend / nicht arbeitend
CHILDREN					X	Kinder im Haushalt j/n
SINGLE		*		*	X	Single bzw. alleine lebend j/n
CAR_POSS					X	Autoverfügbarkeit j/n
EDUC			**		X	Ausbildung während der Umfrage j/n
G*_D					X	Gesellschaftsgruppe Nr. * j/n
EDUC_YR						Anzahl der Ausbildungsjahre
G						Nummer der Gesellschaftsgruppe
CAR		*				Anzahl der Autos im Haushalt
HH_NR_L			**			Anzahl der Führerscheine im HH
HH_MEMB						Anzahl der Haushaltsmitglieder
ACC_D				*	X	Begleitung j/n
PLANNED	*	*			X	Aktivität geplant j/n
A_EXP_D					X	Ausgaben am Aktivitätsort j/n
WEEKEND	*	*			X	Wochenende j/n
A_DAYT_D					X	Tageszeit (vor 13 Uhr j/n)
A_P_MM_D					X	Unabhängiges VM zur Aktivität
A_N_MM_D					X	Unabhängiges VM nach der A
A_P_MMMD					X	Unabhängiges motorisiertes VM zur A
A_N_MMMD					X	Unabhängiges motorisiertes VM von der A
ACC						Anzahl der Begleitpersonen
A_EXP						Höhe der Ausgaben am Aktivitätsort
A_DAYT	*	*				Tageszeit (vormittag/nachmittag/abends)
A_P_DIST						Distanz zum Aktivitätsort
A_P_TT	*			**		Reisezeit zum Aktivitätsort
A_N_DIST						Distanz zum nächsten Aktivitätsort
A_N_TT						Reisezeit zum nächsten Aktivitätsort
A_P_MM	*					Verkehrsmittel zum Aktivitätsort
A_N_MM						Verkehrsmittel vom Aktivitätsort
NR_T_SUM		*				Anzahl der Wege gesamt
NR_J_SUM	*					Anzahl der Reisen gesamt
A_P_PUR	*		*			Vorangegangene Aktivitätskategorie
A_N_PUR						Nachfolgende Aktivitätskategorie
INTER_T_P	*	*	*			Dauer des Intervalls vorher in Tagen
INTERVALL_ST	*					Intervalldauer zur vorigen beliebigen A
A_P_DURST_A	*	*	*	*		Dauer der beliebigen Aktivität vorher
A_P_DURST_C		*				Dauer der gleichen A vor dem Intervall
A_P_MM_C						Verkehrsmittel zur Aktivität vor Intervall
MM_C_D	*	*	*		X	Unabhäng. VM zur Aktivität vor Intervall
INTERVALL_ST_P	*		*			Intervalldauer zu beliebiger A vor Intervall

## Übersicht der Signifikanz der möglichen Einflußvariablen, semiparametrische Modellschätzung

Variable	Einkauf kurzfristig	Freizeit allgemein	Dienstl. allgemein	Treffen mit Freunden	Dummy	Variablenname
SEX	*	*		*	X	Geschlecht
AGE_**					X	Altersgruppe
AGE	*					Alter
YOUNG					X	Unter 50 Jahren j/n
LICENSE	*	*			X	Führerscheinbesitz j/n
HEAD_O_HH					X	Haushaltsvorstand j/n
WORKED		*			X	arbeitend / nicht arbeitend
CHILDREN					X	Kinder im Haushalt j/n
SINGLE		*		*	X	Single bzw. alleine lebend j/n
CAR_POSS					X	Autoverfügbarkeit j/n
EDUC					X	Ausbildung während der Umfrage j/n
G*_D					X	Gesellschaftsgruppe Nr. * j/n
EDUC_YR						Anzahl der Ausbildungsjahre
G						Nummer der Gesellschaftsgruppe
CAR		*				Anzahl der Autos im Haushalt
HH_NR_L						Anzahl der Führerscheine im HH
HH_MEMB						Anzahl der Haushaltsmitglieder
ACC_D				*	X	Begleitung j/n
PLANNED	*	*			X	Aktivität geplant j/n
A_EXP_D					X	Ausgaben am Aktivitätsort j/n
WEEKEND	*	*			X	Wochenende j/n
A_DAYT_D					X	Tageszeit (vor 13 Uhr j/n)
A_P_MM_D					X	Unabhängiges VM zur Aktivität
A_N_MM_D					X	Unabhängiges VM nach der A
A_P_MMMD					X	Unabhängiges motorisiertes VM zur A
A_N_MMMD					X	Unabhängiges motorisiertes VM von der A
ACC						Anzahl der Begleitpersonen
A_EXP						Höhe der Ausgaben am Aktivitätsort
A_DAYT	*	*				Tageszeit (vorm/nachm/abends)
A_P_DIST						Distanz zum Aktivitätsort
A_P_TT	*			**		Reisezeit zum Aktivitätsort
A_N_DIST						Distanz zum nächsten Aktivitätsort
A_N_TT						Reisezeit zum nächsten Aktivitätsort
A_P_MM						Verkehrsmittel zum Aktivitätsort
A_N_MM						Verkehrsmittel vom Aktivitätsort
NR_T_SUM		*		*		Anzahl der Wege gesamt
NR_J_SUM						Anzahl der Reisen gesamt
A_P_PUR			*			Vorangegangene Aktivitätskategorie
A_N_PUR						Nachfolgende Aktivitätskategorie
INTER_T_P	*	*	*			Dauer des Intervalls vorher in Tagen
INTERVALL_ST						Intervalldauer zur vorigen beliebigen A
A_P_DURST_A	*	*	*	*		Dauer der beliebigen Aktivität vorher
A_P_DURST_C	*	*				Dauer der gleichen A vor dem Intervall
A_P_MM_C						Verkehrsmittel zur Aktivität vor Intervall
MM_C_D	*	*	*		X	Unabhäng. VM zur Aktivität vor Intervall
INTERVALL_ST_P	*					Intervalldauer zu beliebiger A vor Intervall

## Übersicht der Signifikanz der möglichen Einflußvariablen, vollparametrische Modellschätzung

Variable	Einkauf kurzfristig	Freizeit allgemein	Dienstl. allgemein	Treffen mit Freunden	Dummy	Variablenname
SEX	*	*		*	X	Geschlecht
AGE_**					X	Altersgruppe
AGE	*					Alter
YOUNG					X	Unter 50 Jahren j/n
LICENSE		*			X	Führerscheinbesitz j/n
HEAD_O_HH		*			X	Haushaltsvorstand j/n
WORKED		*			X	arbeitend / nicht arbeitend
CHILDREN					X	Kinder im Haushalt j/n
SINGLE		*		*	X	Single bzw. alleine lebend j/n
CAR_POSS					X	Autoverfügbarkeit j/n
EDUC					X	Ausbildung während der Umfrage j/n
G*_D					X	Gesellschaftsgruppe Nr. * j/n
EDUC_YR						Anzahl der Ausbildungsjahre
G						Nummer der Gesellschaftsgruppe
CAR		*				Anzahl der Autos im Haushalt
HH_NR_L			*			Anzahl der Führerscheine im HH
HH_MEMB						Anzahl der Haushaltsmitglieder
ACC_D				*	X	Begleitung j/n
PLANNED	*	*			X	Aktivität geplant j/n
A_EXP_D					X	Ausgaben am Aktivitätsort j/n
WEEKEND	*	*			X	Wochenende j/n
A_DAYT_D					X	Tageszeit (vor 13 Uhr j/n)
A_P_MM_D					X	Unabhängiges VM zur Aktivität
A_N_MM_D					X	Unabhängiges VM nach der A
A_P_MMMD					X	Unabhängiges motorisiertes VM zur A
A_N_MMMD					X	Unabhängiges motorisiertes VM von der A
ACC						Anzahl der Begleitpersonen
A_EXP						Höhe der Ausgaben am Aktivitätsort
A_DAYT		*				Tageszeit (vormittag/nachmittag/abends)
A_P_DIST						Distanz zum Aktivitätsort
A_P_TT	*			**		Reisezeit zum Aktivitätsort
A_N_DIST						Distanz zum nächsten Aktivitätsort
A_N_TT						Reisezeit zum nächsten Aktivitätsort
A_P_MM						Verkehrsmittel zum Aktivitätsort
A_N_MM						Verkehrsmittel vom Aktivitätsort
NR_T_SUM		*	**	*		Anzahl der Wege gesamt
NR_J_SUM						Anzahl der Reisen gesamt
A_P_PUR			*			Vorangegangene Aktivitätskategorie
A_N_PUR						Nachfolgende Aktivitätskategorie
INTER_T_P	**		**			Dauer des Intervalls vorher in Tagen
INTERVALL_ST						Intervalldauer zur vorigen beliebigen A
A_P_DURST_A	*	*	*	**		Dauer der beliebigen Aktivität vorher
A_P_DURST_C	**	*				Dauer der gleichen A vor dem Intervall
A_P_MM_C			**			Verkehrsmittel zur Aktivität vor Intervall
MM_C_D	*	*	*		X	Unabh. VM zur Aktivität vor Intervall
INTERVALL_ST_P	*		*			Intervalldauer zu beliebiger A vor Intervall

## Vergleich der einzelnen parametrischen Modellschätzungen

Variable	Einkauf kurzfristig			Freizeit allgemein			Dienstl. allgemein			Treffen mit Freunden		
	non	semi	voll	non	semi	voll	non	semi	voll	non	semi	voll
SEX	*	*	*	*	*	*				*	*	*
AGE_**												
AGE	*	*	*									
YOUNG												
LICENSE		*		*	*	*						
HEAD_O_HH						*						*
WORKED				*	*	*						*
CHILDREN												
SINGLE				*	*	*				*	*	*
CAR_POSS												
EDUC							**					
G*_D												
EDUC_YR												
G												
CAR				*	*	*						
HH_NR_L							**		*			
HH_MEMB												
ACC_D										*	*	*
PLANNED	*	*	*	*	*	*						
A_EXP_D												
WEEKEND	*	*	*	*	*	*						
A_DAYT_D												
A_P_MM_D												
A_N_MM_D												
A_P_MMMD												
A_N_MMMD												
ACC												
A_EXP												
A_DAYT	*	*		*	*	*						
A_P_DIST												
A_P_TT	*	*	*							**	**	**
A_N_DIST												
A_N_TT												
A_P_MM	*											
A_N_MM												
NR_T_SUM				*	*	*			**		*	*
NR_J_SUM	*											
A_P_PUR	*						*	*	*			
A_N_PUR												
INTER_T_P	*	*	**	*	*		*	*	**			
INTERVALL_ST	*											
A_P_DURST_A	*	*	*	*	*	*	*	*	*	*	*	**
A_P_DURST_C		*	**	*	*	*						
A_P_MM_C									**			
MM_C_D	*	*	*	*	*	*	*	*	*			
INTERVALL_ST_P	*	*	*				*		*			