



Synthetic Population of the Canton Zürich for the Year 2000

M Frick

Synthetic Population of the Canton Zürich for the Year 2000

M Frick
IVT
ETH Zürich
Zürich

Phone: +41 44 633 37 13
Fax: +41 44 633 10 57

frick@ivt.baug.ethz.ch

Abstract

The generation of synthetic populations represents a substantial contribution to the acquisition of useful data for large scale agent based microsimulations in the field of transport planning. Basically, the observed data are available from various sources, i.e. censuses (microcensus) in which the data is available in terms of simple summary tables of demographics, such as the number of persons per household for census-block-group-sized areas. Nevertheless, there is a need of more disaggregated personal data, and thus another type of data source is considered. The Public Use Sample (PUS), often used in transportation studies, is a 5% representative sample of complete census records, including bad records, for each individual, excluding addresses and unique identifiers. The problem is, to generate a large number of individual agents (~1Mio.) with appropriate characteristic values of the demographic variables for each agent, interacting in the microsimulation. The main techniques used to generate the agents are IPF and simple MC. In this paper we present further results of our effort to disaggregate the available census data. First, we present agents with age and sex as the sociodemographic characteristics for all municipalities in Switzerland using data from census 2000 and microcensus 2000. Second, we added more sociodemographic variables like driver licence ownership, car availability, employment, accessibility of halbtax and GA periodic tickets to obtain more realistic agents for all of Switzerland. Third we made some effort to disaggregate the data to a hectare based level for employment, age, and sex of the agents. The current state of this work will be presented.

Keywords

Synthetic Population , Canton Zürich , IPF , Entropy

Preferred citation style

Frick, M. (2005) Synthetic Population of the Canton Zürich for the Year 2000 , *Travel Survey Metadata Series*, **18**, Institute for Transport Planning and Systems (IVT); ETH Zürich, Zürich.

1.0 Document Description

Citation

Title:	Synthetic Population of the Canton Zürich for the Year 2000
Subtitle:	None
Alternative Title:	None
Parallel Title:	None
Identification Number:	SYNPOP2000CantonZH
Authoring Entity:	Martin Frick (IVT)
Other identifications and acknowledgements:	None
Producer:	Martin Frick
Copyright:	(c) IVT, ETH Zürich
Date of Production:	2005-12-09
Software used in Production:	Nesstar Publisher

Distributor: Institute for Transpor Planning and Systems

Depositor: Institute for Transpor Planning and Systems

2.0 Study Description

Citation

Title:	Synthetic Population of the Canton Zürich for the Year 2000
Subtitle:	None
Alternative Title:	None
Parallel Title:	None
Identification Number:	SYNPOP2000CantonZH
Authoring Entity:	Martin Frick (Institute for Transpor Planning and Systems)
Producer:	Martin Frick
Date of Production:	2005-12-09
Software used in Production:	Nesstar Publisher
Funding Agency/Sponsor:	ETH Zürich
Grant Number:	None

Distributor: Martin Frick

Depositor: Institute for Transport Planning and Systems

Version: Initial Version 0.0.-1 (alpha)

Study Scope

Topic Classification: Synthetic Population

Abstract:

The generation of synthetic populations represents a substantial contribution to the acquisition of useful data for large scale agent based microsimulations in the field of transport planning. Basically, the observed data are available from various sources, i.e. censuses (microcensus) in which the data is available in terms of simple summary tables of demographics, such as the number of persons per household for census-block-group-sized areas. Nevertheless, there is a need of more disaggregated personal data, and thus another type of data source is considered. The Public Use Sample (PUS), often used in transportation studies, is a 5% representative sample of complete census records, including bad records, for each individual, excluding addresses and unique identifiers. The problem is, to generate a large number of individual agents (~1Mio.) with appropriate characteristic values of the demographic variables for each agent, interacting in the microsimulation. The main techniques used to generate the agents are IPF and simple MC. In this paper we present further results of our effort to disaggregate the available census data. First, we present agents with age and sex as the sociodemographic characteristics for all municipalities in Switzerland using data from census 2000 and microcensus 2000. Second, we added more sociodemographic variables like driver licence ownership, car availability, employment, accessibility of halbtax and GA periodic tickets to obtain more realistic agents for all of Switzerland. Third we made some effort to disaggregate the data to a hectare based level for employment, age, and sex of the agents. The current state of this work will be pre-sented.

Time Period:

-

Country:

Switzerland

Geographic Coverage:

Canton Zürich

Geographic Unit(s): Hectare

Unit of Analysis: Individuals

Universe: about 1.2 Mio Agents representing permanently residing persons
in the Canton Zürich

Methodology and Processing

Time Method: None

Sampling Procedure: None

Sources Statement

Weighting: Several

3.0 File Description

File: Kopie von population.windows.NSDstat

- Number of cases: 9
- No. of variables per record: 15
- Type of File: NSDstat 200203

4.0 Variable Description

Variable Groups

- [Socio-demographics](#)
- [Mobility Tools](#)
- [Spatial Info](#)
- [Home Locations](#)
- [Work Locations](#)

Socio-demographics

Variables within *Socio-demographics*

- [Municipality Number 2000](#)
- [the age of the agent](#)
- [Sex](#)
- [The driver licence ownership for the agent](#)
- [Employment status of the agent](#)
- [Household Monthly Income](#)

Mobility Tools

Variables within *Mobility Tools*

- [Municipality Number 2000](#)
- [The car availability for the agent](#)
- [Half Fare Ticket Ownership](#)
- [General Abonnement Ownership](#)
- [Commuter Trip Mode](#)

Spatial Info

Variable Groups within *Spatial Info*

- [Home Locations](#)

- [Work Locations](#)

Home Locations

Variables within *Home Locations*

- [Municipality Number 2000](#)
- [Home Location x 100](#)
- [Home Location y 100](#)

Work Locations

Variables within *Work Locations*

- [Municipality Number 2000](#)
- [Work Location Municipality Number 2000](#)
- [Work Location x 100](#)
- [Work Location y 100](#)

Variables

Variable: Municipality Number 2000

Value	Label	Frequency
1 .		9

Range of Valid Data Values: 1 to 1

Summary Statistics:

Minimum : 1

Maximum : 1

Mean : 1

Standard deviation : 0

Variable Format: numeric

Variable: the age of the agent

Value	Label	Frequency
1.	0	0
2.	1-4	0
3.	5-9	1
4.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	1
5.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	1
6.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0
7.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0

8.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	1
9.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	1
10.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	1
11.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	1
12.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	2
13.	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0
14.	0 1-4 5-9 10-14 15-19 20-24	0

	25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	
15 .	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0
16 .	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0
17 .	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0
18 .	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0
19 .	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85-89 90-94 95-99 100-104 gt105	0
20 .	0 1-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64	0

65-69 | 70-74 | 75-79 | 80-84 |
 85-89 | 90-94 | 95-99 | 100-104 |
 gt105

0 | 1-4 | 5-9 | 10-14 | 15-19 | 20-24
 | 25-29 | 30-34 | 35-39 | 40-44 |
 21 . 45-49 | 50-54 | 55-59 | 60-64 | 0
 65-69 | 70-74 | 75-79 | 80-84 |
 85-89 | 90-94 | 95-99 | 100-104 |
 gt105

Range of Valid Data Values: 3 to 12

Summary Statistics:

Minimum : 3

Maximum : 12

Mean : 8.222

Standard deviation : 3.456

Variable Format: numeric

Variable: Sex

Value	Label	Frequency
1 .		5
2 .		4

Range of Valid Data Values: 1 to 2

Summary Statistics:

Variable Format: numeric

Variable: Home Location x 100

Range of Valid Data Values: 679000 to 679700

Summary Statistics:

Minimum : 679000

Maximum : 679700

Variable Format: numeric

Variable: Home Location y 100

Range of Valid Data Values: 235500 to 237800

Summary Statistics:

Minimum : 235500

Maximum : 237800

Variable Format: numeric

Variable: The driver licence ownership for the agent

Value	Label	Frequency
1 .		5
2 .		2
3 .		2

Range of Valid Data Values: 1 to 3

Summary Statistics:

Variable Format: numeric

Variable: The car availability for the agent

Value	Label	Frequency
1 .		5
4 .		4

Range of Valid Data Values: 1 to 4

Summary Statistics:

Variable Format: numeric

Variable: Employment status of the agent

Value	Label	Frequency
1 .		7
2 .		2

Range of Valid Data Values: 1 to 2

Summary Statistics:

Variable Format: numeric

Variable: Half Fare Ticket Ownership

Value	Label	Frequency
2 .		2
4 .		5
5 .		2

Range of Valid Data Values: 2 to 5

Summary Statistics:

Variable Format: numeric

Variable: General Abonnement Ownership

Value	Label	Frequency
2 .		9

Range of Valid Data Values: 2 to 2

Summary Statistics:

Variable Format: numeric

Variable: Household Monthly Income

Value	Label	Frequency
1 .		3
2 .		4
5 .		1
6 .		1

Range of Valid Data Values: 1 to 6

Summary Statistics:

Variable Format: numeric

Variable: Work Location Municipality Number 2000

Range of Valid Data Values: -2 to 261

Summary Statistics:

Minimum : -2

Maximum : 261

Mean : 109.444

Standard deviation : 130.11

Variable Format: numeric

Variable: Commuter Trip Mode

Range of Valid Data Values: -2 to 32

Summary Statistics:

Minimum : -2

Maximum : 32

Mean : 16.778

Standard deviation : 11.498

Variable Format: numeric

Variable: Work Location x 100

Range of Valid Data Values: -2 to 691500

Summary Statistics:

Minimum : -2

Maximum : 691500

Mean : 530299.556

Standard deviation : 300680.342

Variable Format: numeric

Variable: Work Location y 100

Range of Valid Data Values: -2 to 250200

Summary Statistics:

Minimum : -2

Maximum : 250200

Mean : 189088.444

Standard deviation : 107421.053

Variable Format: numeric