



# EINSETZUNGEN IN DER STRUKTURERHEBUNG

Informationsveranstaltung VZ  
10. Oktober 2013

Christian Panchard, Daniel Kilchmann

BFS / METH



# Inhalt

**Vorgehen der Nächste-Nachbarn-Einsetzung**

**Umsetzung der Nächster-Nachbarn-Einsetzung**

**Resultate**

**Weiteres Vorgehen**



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Eidgenössisches Departement des Innern EDI  
**Bundesamt für Statistik BFS**

# Vorgehen der Nächste-Nachbarn-Einsetzung

Statistik Schweiz



## Nächster-Nachbarn-Einsetzung

### ► Einsetzung :

- ersetzen von fehlenden Werten mit vorhandenen Werten,
- ändern inkonsistenter Werte in Konsistente,
- ändern einer minimalen Anzahl inkonsistenter Werte.

→ Kontrollregeln um fehlende und inkonsistente Werte zu entdecken.

### ► Nächste Nachbarn :

- Auswahl einer Beobachtung, die Werte "spendet", damit eine Beobachtung, die mindestens eine Kontrollregel verletzt, vollständig und konsistent wird ("Spender" → "Empfänger").
- Auswahl von Spendern, die dem Empfänger ähnlich (nahe) sind und keine Fehler aufweisen.  
→ Definition einer Distanz.



# Umsetzung der Nächster-Nachbarn-Einsetzung



## Strukturerhebung 2010

- ▶ 317'221 Beobachtungen des Personenfragebogens.
- ▶ 689'914 Beobachtungen für Haushalts- und Wohnungsteil.

- 317'221 Beobachtungen im Einsetzungsverfahren.  
(Für Bearbeitung Haushaltsmerkmale noch zu definieren.)
- Total 197 Variablen (nicht alle werden eingesetzt).



# Auswahl der Einsetzungsvariablen

4 Variablenpakete, die separat eingesetzt werden :

## 1. Ausbildung und Beruf :

- ▶ Höchste abgeschlossene Ausbildung.
- ▶ Gegenwärtige Ausbildung.
- ▶ Situation auf dem Arbeitsmarkt.
- ▶ Stellung im Beruf.

## 2. Sprachen, Migration und Wohnsituation.

## 3. Mobilität.

## 4. Haushalt.



# Kreieren von Hilfsvariablen für die Einsetzung

Ziel : beibehalten der Konsistenz zwischen den Einsetzungspaketen.

- ▶ Anzahl Antworten zu Fragen bezüglich Arbeit (übliche Sprache, Adresse, Arbeitsweg).
- ▶ Anzahl Antworten zu Fragen bezüglich Ausbildung (übliche Sprache, Adresse, Schulweg).





## Kontrollregeln

- ▶ Auswahl der Kontrollregeln, welche nur die Einsetzungsvariablen des jeweiligen Einsetzungspakets betreffen.
- ▶ Hinzufügen von Regeln für die Hilfsvariablen.

→ Kontrollregeln

- ▶ für fehlende Werte,
  - ▶ höchste abgeschlossene Ausbildung = .,
- ▶ bezüglich Inkonsistenzen,
  - ▶  $\text{Alter} < 60$  und Arbeitsmarktstatus = "Rentner/in",
- ▶ bezüglich Hilfsvariablen (weiche Regeln),
  - ▶ Anzahl Antworten zu Fragen bezüglich Arbeit  $> 3$  und Arbeitsmarktstatus  $\neq$  "Erwerbstätig".



## Paket 1 : Anzahl Beobachtungen mit Fehlern

- ▶ 11'793 Beobachtungen mit mindestens einem fehlenden Wert (3.7%).
- ▶ 11'234 Beobachtungen mit mindestens einer Inkonsistenz (3.5%).

→ Einsetzungen bei 22'653 Beobachtungen (7.1%).



# Abgrenzung der Nachbarschaft

## 1. Distanzdefinition :

- ▶ Basiert auf Einsetzungs-, Hilfs- und Registervariablen.
- ▶ Vergleicht alle Variablen des Empfängers mit jenen des Senders.
  - ▶ Diskrete Distanz : z.B.  $\{0, 1\}$  für das Geschlecht.
  - ▶ Stetige Distanz : z.B.  $[0, 1]$  für das Alter.

→ Gewichtete Kombination der Distanzen pro Variable.



# Abgrenzung der Nachbarschaft

## 2. Festlegen der Einsetzungsklassen :

- ▶ Kreuzung der Kantone und der Haushaltsgrössen (1, ..., 5, 6+).
- ▶ Wenn innerhalb der Einsetzungsklasse kein Spender gefunden wird → Suche innerhalb aller Einsetzungsklassen.



## Auswahl des Spenders

- ▶ Aufteilen der Daten in potentiellen Spender und Empfänger.
- ▶ Berechnen einer Distanz zwischen allen potentiellen Spendern und allen Empfängern.
- ▶ Zufällige Auswahl eines Spenders mit der Wahrscheinlichkeit proportional zum Kehrwert der Distanz.



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Eidgenössisches Departement des Innern EDI  
**Bundesamt für Statistik BFS**

# Resultate



## Vergleich vor vs nach Einsetzungen

- ▶ Alle Empfänger wurden konsistent und vollständig eingesetzt.
- ▶ Höchste abgeschlossene Ausbildung :

		nach Einsetzung						
		keine	obligat.	berufliche	allgemeinbildende	höhere	universitäre	Total
vor Einsetzung	.	1 016	988	1 413	240	284	390	4 331
	keine	27 485						27 485
	obligat.		54 397					54 397
	berufliche			121 926				121 926
	allgemeinbildende				24 879			24 879
	höhere	7	3	1		36 402		36 413
	universitäre	15	7	2			47 766	47 790
	Total	28 523	55 395	123 342	25 119	36 686	48 156	317 221



# Vergleich Randverteilung

Höchste abgeschlossene Ausbildung :

	<i><b>vor Einsetzung</b></i>		<i><b>nach Einsetzung</b></i>		<i><b>Differenz</b></i>	
	<i>Anzahl</i>	<i>%</i>	<i>Anzahl</i>	<i>%</i>	<i>Anzahl</i>	<i>%</i>
.	88 663	-	-	-	-	-
<i>keine</i>	575 057	8.9%	596 385	9.1%	21 328	0.2%
<i>obligat.</i>	1 161 038	18.1%	1 181 962	18.1%	20 924	0.1%
<i>berufliche</i>	2 578 565	40.1%	2 607 458	40.0%	28 893	-0.1%
<i>allgemeinbildende</i>	502 858	7.8%	507 838	7.8%	4 980	0.0%
<i>höhere</i>	741 753	11.5%	747 249	11.5%	5 496	-0.1%
<i>universitäre</i>	871 319	13.5%	878 361	13.5%	7 042	-0.1%
<i>Total</i>	6 430 590	100.0%	6 519 253	100.0%		





## Qualitätsmasse

- Anzahl potentielle Spender pro Empfänger :

Anzahl pot. Spender	Anzahl Empfänger
1 - 10	2'941
11 - 50	6'785
51 - 100	4'368
101 -	8'559

- Wiederverwendung desselben Spenders :

Anzahl Verwendungen	Anzahl Empfänger
1	20'268
2	1'109
3	53
4	2



## Weiteres Vorgehen



- ▶ Einsetzungen der Pakete 2 bis 4.
- ▶ Einsetzungen für die Strukturerhebungen 2011 und folgende.