

# Korrekturmöglichkeiten

---

## **Bildung von Gewichtungs- bzw. Adjustierungsklassen $C_j$**

1. Unterteilung der Stichprobe nach den Werten bzw. Wertekombinationen von Hintergrundvariablen, um für jede der resultierenden Adjustierungsklassen
2. die Inverse der Antwortrate als Nonresponsegewicht zu verwenden

[entspricht in der Wirkung der Imputation des mittleren (konditionalen) beobachteten Wertes für die jeweilige Klasse]

## A, Adjustierungsklasse C (Region)

C=1

C=2

C=3

Antwortrate	80/100	70/100	50/100
Mittleres Eink.	9.800	11.600	13.600
Totales Eink.	780.000	815.000	680.000

$$\bar{y}_R = \frac{780.000 + 815.000 + 680.000}{80 + 70 + 50} = 11.375$$

$$\bar{y}_A = \frac{780.000(100/80) + 815.000(100/70) + 680.000(100/50)}{100 + 100 + 100} = 11.664$$

## Korrekturmöglichkeiten

---

### **Gewichtung nach Antwortneigung** (propensity weighting)

1. Regression des  $[0,1]$  Indikators für unit response  $r_i$  auf ein geeignetes, zugleich für Respondenten und Nichtrespondenten verfügbares Set von Hintergrundmerkmalen (logistische Regression);
2. Berechnung *erwarteter* Antwortwahrscheinlichkeiten auf der Basis dieser Regression, und
3. Gruppierung dieser Wahrscheinlichkeiten zu fünf oder sechs Werten, um
4. über den so gruppierten *propensity score* jeweils eine Adjustierungs- bzw. Gewichtungsklasse  $C_j$  zu bilden.  
(response propensity stratification)

## Alternative

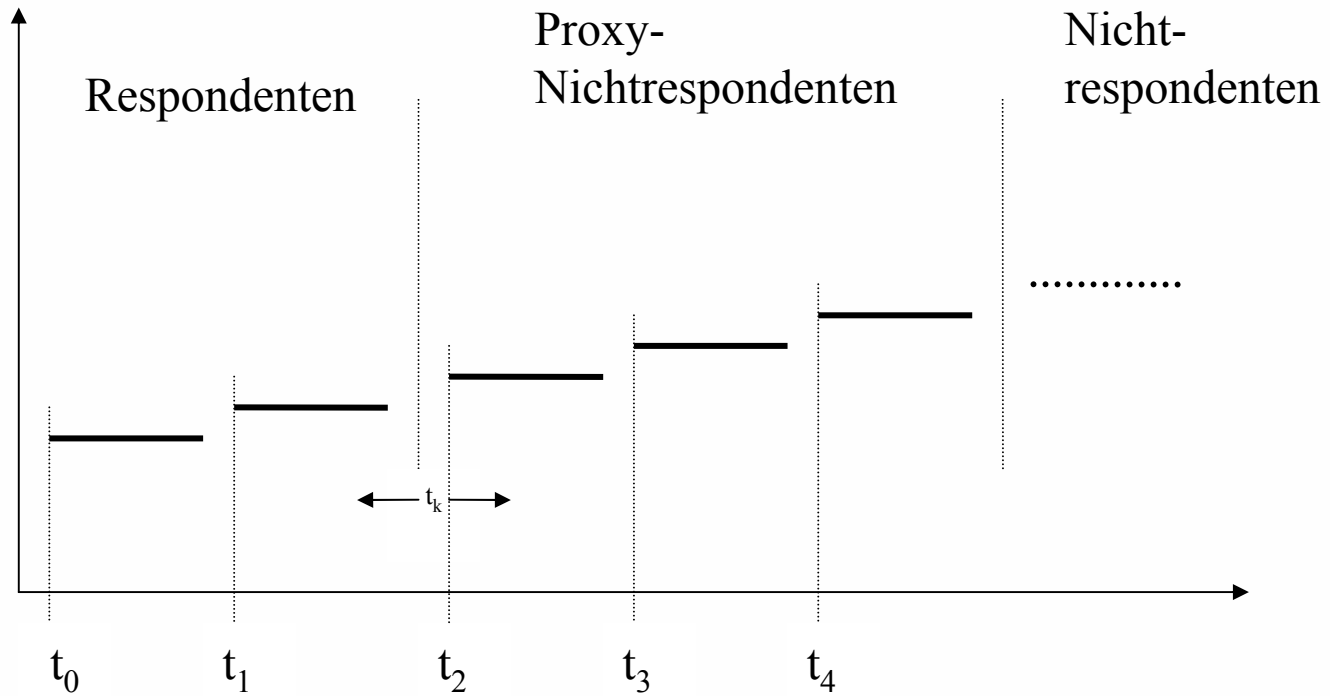
Verzicht auf Schritt 4 (Gruppierung) und gleich

Gewichtung der Respondenten  $i$  mit der Inversen  
des geschätzten propensity scores,

$$[\hat{p}_i]^{-1}$$

## Informationen über Nichtrespondenten

- über separate follow-up Surveys
- über Registerinformationen
  - Einwohnermeldeamt
  - Access Panels (wenn Sampling Frame selbst unverzerrt ist)
- über Informationen aus anderen Wellen einer Panelstudie
- über „Proxy – Nichtrespondenten“



## Schätzungen der Prävalenz affektiver Störungen\*

$$\Delta\% = (p_{NR} * 100) = 8\%$$

Antwortrate	74% 82%				
Schwierigkeitsgrad, die Personen zu erreichen	gering	mittel	hoch	gering +mittel	gering + mittel + hoch
Prävalenz	0,218	0,285	0,312	0,223	0,239

$$Bias = p_{NR} (\bar{x}_R - \bar{x}_{NR}) \quad Bias = 0,08(0,218 - 0,285) = -0,0054$$

## Schätzungen der Prävalenz affektiver Störungen\*

Schwierigkeitsgrad, die Personen zu erreichen	gering	mittel	hoch	gering +mittel	gering + mittel + hoch
Prävalenz	0,218	0,285	0,312	0,223	0,239
Standardfehler	0,0059				0,0075

$$MSE = 0,0059^2 + 0,021^2 = 0,00048$$

$$\text{Bias} = 0,239 - 0,218 = 0,021$$

$$MSE = 0,0075^2 + 0^2 = 0,00006$$

$$\text{Bias} = 0 \text{ (unterstellt)}$$



## **Flexible** Survey Designs bieten ..

- zugeschnittenen („tailored“) Interviewer-vermittelten Kontakt (im Sinne der „leverage-saliency“ Theorie)
- in Abhängigkeit von der Stärke der Survey Kooperation:*
- Einsatz einer langen, intermediären, kurzen Fragebogenversion
  - Flexibles Zeitmanagement von Interviews
  - „Mixed-Mode“ Datenerhebung
    - f2f, tel., web, mail, je nach Wahl der angefragten Person

## Flexibles/Mixed Mode Survey Design

### Schritte in einer Studie zur Verkehrsmittelwahl\*

1. Vorausbrief (falls möglich)
2. Initialer interviewervermittelter Kontakt per Telefon, um Termin für ein **volles** face-to-face zu vereinbaren; falls ungenügende Teilnahmemotivation:
3. *hilfsweise*: Angebot der Möglichkeit eines **gekürzten** Interviews per Telefon (von in etwa halbierten Länge); falls noch unzureichende Teilnahmemotivation:
4. Erbitten von Minimum Information per Telefon (~ 1 Fragebogenseite)

\* siehe Engel/Hansen (2001),  
<http://www.sozialforschung.uni-bremen.de/methdfig.pdf>  
Erhebung in 1999 in Berlin und Brandenburg

autofahr Überwiegende Nutzung \* intview (0) 'tel' Kreuztabelle

			intview (0) 'tel'		Gesamt
			,00	1,00	
autofahr Überwiegende Nutzung	,00 nein	Anzahl	20	141	161
		% von intview (0) 'tel'	22,5%	43,8%	39,2%
	1,00 ja,häufig-immer	Anzahl	69	181	250
		% von intview (0) 'tel'	77,5%	56,2%	60,8%
Gesamt	Anzahl		89	322	411
	% von intview (0) 'tel'		100,0%	100,0%	100,0%

↑  
Kurzfassung

↑  
Langfassung

autofahr Überwiegende Nutzung \* intview (0) 'tel' \* fr\_31 Herkunft alte/neue Bundesländer? Kreuztabelle

fr_31 Herkunft alte/neue Bundesländer?					intview (0) 'tel'		Gesamt
					,00	1,00	
1 alte Bundesländer	autofahr Überwiegende Nutzung	,00 nein	Anzahl	7	67	74	
			% von intview (0) 'tel'	23,3%	58,3%	51,0%	
		1,00 ja,häufig-immer	Anzahl	23	48	71	
			% von intview (0) 'tel'	76,7%	41,7%	49,0%	
	Gesamt		Anzahl	30	115	145	
			% von intview (0) 'tel'	100,0%	100,0%	100,0%	
2 neue Bundesländer	autofahr Überwiegende Nutzung	,00 nein	Anzahl	12	70	82	
			% von intview (0) 'tel'	20,7%	34,7%	31,5%	
		1,00 ja,häufig-immer	Anzahl	46	132	178	
			% von intview (0) 'tel'	79,3%	65,3%	68,5%	
	Gesamt		Anzahl	58	202	260	
			% von intview (0) 'tel'	100,0%	100,0%	100,0%	

↑  
Kurzfassung

↑  
Langfassung

fr\_15\_3 Vorschlag: Technik neuester Stand \* intview (0) 'tel' Kreuztabelle

			intview (0) 'tel'		Gesamt
			,00	1,00	
fr_15_3 Vorschlag: Technik neuester Stand	0 nein	Anzahl	13	90	103
		% von intview (0) 'tel'	18,1%	36,9%	32,6%
	1 ja	Anzahl	59	154	213
		% von intview (0) 'tel'	81,9%	63,1%	67,4%
Gesamt		Anzahl	72	244	316
		% von intview (0) 'tel'	100,0%	100,0%	100,0%

fr\_15\_3 Vorschlag: Technik neuester Stand \* intview (0) 'tel' \* fr\_31 Herkunft alte/neue Bundesländer? Kreuztabelle

fr_31 Herkunft alte/neue Bundesländer?				intview (0) 'tel'		Gesamt
				,00	1,00	
1 alte Bundesländer	fr_15_3 Vorschlag: Technik neuester Stand	0 nein	Anzahl	3	33	36
			% von intview (0) 'tel'	13,0%	44,0%	36,7%
		1 ja	Anzahl	20	42	62
			% von intview (0) 'tel'	87,0%	56,0%	63,3%
	Gesamt		Anzahl	23	75	98
			% von intview (0) 'tel'	100,0%	100,0%	100,0%
2 neue Bundesländer	fr_15_3 Vorschlag: Technik neuester Stand	0 nein	Anzahl	10	56	66
			% von intview (0) 'tel'	20,4%	33,5%	30,6%
		1 ja	Anzahl	39	111	150
			% von intview (0) 'tel'	79,6%	66,5%	69,4%
	Gesamt		Anzahl	49	167	216
			% von intview (0) 'tel'	100,0%	100,0%	100,0%

Logit  
(Log-Odd)

$$\ln\left(\frac{\pi_1}{1 - \pi_1}\right) = \beta_0 + \beta_1 x$$

Odd

$$\frac{\pi_1}{1 - \pi_1} = e^{\beta_0 + \beta_1 x} = e^{\beta_0} \times e^{\beta_1 x}$$

Wahrschein-  
lichkeit

$$\pi_1 = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Survey Partizipation Prob. (volle Länge)	b	e <sup>b</sup>
Intercept	.29	1.33
Im Besitz einer „BahnCard“	1.37*	3.94
Region: Ost (vs. West)	.37	1.45
Geschlecht (weiblich)	-.31	.74
Bildung	.34*	1.41
Häufige/Regelmäßige Nutzung des Autos	-.86*	.42

\*  $p < .05$  $R^2$  (Nagelkerke): .128

n=414\*\*

\*\* 322 face-to-face-Interviews  
+ 92 Telefoninterviews

## Modellierung der Antwortneigung (response propensity)

$$\Pr\{R_t\} = \Pr\{I_t|C_t, L_t\} \times \Pr\{C_t|L_t\} \times \Pr\{L_t\}$$

- Wahrscheinlichkeit, den Zielhaushalt/die Zielperson zu **lokalisieren** [ $\Pr\{L_t\}$ ]
- **Kontakt**wahrscheinlichkeit, gegeben die Lokalisierung [ $\Pr\{C_t|L_t\}$ ]
- **Kooperations**wahrscheinlichkeit, gegeben Kontakt und Lokalisierung [ $\Pr\{I_t|C_t, L_t\}$ ]

$$\begin{aligned} &\Pr\{R_t|x_L, x_C, x_I\} \\ &= \Pr\{I_t|C_t, L_t, x_I\} \times \Pr\{C_t|L_t, x_C\} \times \Pr\{L_t, x_L\} \end{aligned}$$



# Kontaktwahrscheinlichkeit

- Haushaltsgröße
- Zahl der erwachsenen Haushaltsmitglieder
- Vorhandensein von kleinen Kindern und Senioren
- Anrufzeiten
- Zahl der Kontaktversuche  
(designseitig begrenzt?; Länge der Feldphase)

- Physikalische Zugangshindernisse
  - z.B. verschlossene Gebäude
  - technische Vorkehrungen, dass nur erwünschte Anrufe ankommen
- Handicaps
- „Urbanizität“
  - ländl. Raum > städtischen und großstädt. Raum
  - Unterschiede in Haushaltszusammensetzung, Zugangshindernissen, Bevölkerungsdichte, Kriminalitätsrate, Anteil von Miets- und Mehrfamilienhäusern

# Konzeptueller Rahmen für die Survey **Kooperation**\*

