

gesis

Leibniz-Institut
für Sozialwissenschaften



Gewichtung von Erhebungsdaten:
Kalibrierung, Anpassungsgewichtung und
Imputation

Meet the Experts

Best-Practice-Methoden in der Umfrageforschung

Dr. Christian Bruch und Dr. Matthias Sand, 26.11.2020

Referenten

Dr. Matthias Sand

- Senior Researcher bei GESIS, Team Survey Statistics
- Tätigkeitsschwerpunkte: Komplexe Stichprobenziehung, Gewichtung von Erhebungsdaten, sowie Imputation fehlender Werte
- Dissertation zum Thema Gewichtung von Erhebungsdaten aus mehreren Auswahlgrundlagen

Dr. Christian Bruch

- Senior Researcher bei GESIS, Team Survey Statistics
- Tätigkeitsschwerpunkte: Gewichtungsverfahren, Varianzschätzung und Imputation fehlender Werte
- Dissertation zum Thema Varianzschätzung unter Imputation und bei komplexen Stichprobendesigns

Eckpunkte zum Vortrag

- Der Vortrag wird aufgezeichnet; die anschließende Diskussion wird nicht aufgezeichnet
- Teilnehmende sind stummgeschaltet
- Fragen bitte ausschließlich per Chat privat an „meetexperts“ stellen
- Wenn der Chat „an alle“ geht, dann sind die Nachrichten für alle sichtbar (inkl. Name)
- Fragen werden nach dem Vortrag beantwortet

- I Mechanismen fehlender Werte
- II Item-Nonresponse und Imputation
- III Unit-Nonresponse und
Kalibrierung/Anpassungsgewichtung

MISSING COMPLETELY AT RANDOM (MCAR)

- Die Antwortwahrscheinlichkeit der Untersuchungseinheiten hängt weder von den beobachteten Hilfsvariablen \mathbf{x}_i , der interessierenden Survey Variable y_i , noch von anderen Parameter der Erhebung ab.
- Wenn der Nonresponse Mechanismus MCAR ist, sind die Schätzungen approximativ unverzerrt (unter SRS).
- Wird angenommen wenn der Nonresponse Mechanismus ignoriert wird.

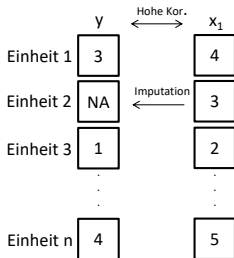
MISSING AT RANDOM (MAR)

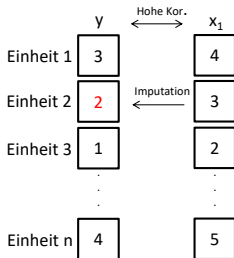
- Die Antwortwahrscheinlichkeit hängt von \mathbf{x}_i ab, aber nicht von y_i .
- **Ignorierbare Nonresponse Mechanismus:** Das Modell kann den Nonresponse Mechanismus erklären und dieser kann ignoriert werden, nachdem dieser durch das Modell berücksichtigt wurde.
- Falsche Annahme: Die fehlenden Werte können vollständig ignoriert werden und die Daten können behandelt werden wie ein vollständiger Datensatz.

NOT MISSING AT RANDOM (NMAR)

- Die Antwortwahrscheinlichkeit hängt (teilweise) von y_i ab und kann nicht alleine durch \mathbf{x}_i erklärt werden
- Fehlende Werte können nicht alleine durch die beobachteten Daten erklärt werden
- **Nichtignorierbarer Nonresponse Mechanismus**

- **Unit-Nonresponse:** Eine Befragungseinheit verweigert die vollständige Teilnahme an der Erhebung
→ eher Gewichtung
- **Item-Nonresponse:** Eine Person nimmt grundsätzlich an der Erhebung teil, gibt jedoch bei bestimmten Fragen keine Antwort
→ eher Imputation





Die Imputation ist der Prozess der Zuweisung von (konstruierten oder tatsächlich existierenden) Werten zu fehlenden Werten einer Variable unter einem möglichen Rückgriff auf Hilfsvariablen (Prädiktoren), um einen vollständigen Datensatz zu erhalten.

Quelle: Bruch (2016), Haziza (2009), Kim und Rao (2009), Chauvet et al. (2011)

- Geeignet und überwiegend benutzt für Item-Nonresponse
- Manche Methoden können theoretisch auch für Unit-Nonresponse verwendet werden, in diesem Fall werden jedoch eher Gewichtungsverfahren angewandt
- Die meisten Methoden können im Fall von Missing at Random oder Missing Completely at Random angewendet werden
- NMAR: Berücksichtigung so vieler Hilfsvariablen (Prädiktoren) wie möglich, um MAR anzunehmen

Quelle: Van Buuren (2012)

- Möglichkeit mit vollständigen Daten zu arbeiten
- Reduzierung des Nonresponse Bias
- Reproduktion der Kovarianz (Korrelations) Struktur und der Randverteilungen der Variablen
- Verwendung von zusätzlichen Informationen in der Schätzung (im Vergleich zur Complete Case Analysis)
- Der imputierte Wert soll plausibel sein, d.h. Kombinationen wie die *3-jährige Großmutter* sollen vermieden werden.
- Imputationsmethoden sollten die verschiedenen Skalenniveaus berücksichtigen
- In der Sozialwissenschaft häufig im Zusammenhang mit der Gewichtung und Split Questionnaire Designs verwendet (siehe Raghunathan und Grizzle, 1995).

Quelle: siehe z.B. Van Buuren (2012)

Ausgangspunkt:

- Es sollen so viele Information wie möglich berücksichtigt werden, damit möglichst viele Prädiktoren verwendet werden (auch zur Sicherstellung der MAR-Annahme)
- Bei zu vielen Variablen kann dies zu Problemen aufgrund von Multikollinearitäten, Freiheitsgraden und Laufzeit des Algorithmus führen
- Van Buuren (2012) empfiehlt nicht mehr als 15 bis 25 Prädiktoren zu verwenden

Welche Variablen sollen als Prädiktoren verwendet werden?

- Analysevariable → Ansonsten Zerstörung der Korrelation zwischen Variablen
- Variablen, die hoch korreliert sind zum Nonresponse Mechanismus und/oder zu der zu imputierenden Variable
- Prädiktor und zu imputierende Variable sollten ausreichend gemeinsame Beobachtungen haben, um Model zuverlässig zu schätzen

Quelle: siehe z.B. Van Buuren (2012), Schafer (1997), Meng (1994), Collins et al. (2001), Nicoletti und Peracchi (2006)

Einfache Imputation	Multiple Imputation
Für jeden fehlenden Wert wird nur ein imputierter Wert verwendet	Für jeden fehlenden Wert werden mehr als nur ein Imputationswert gezogen
Ein Datensatz liegt am Ende vor	Mehr als ein Datensatz wird gebildet
Keine (direkte) Berücksichtigung der Unsicherheit der Imputation	Berücksichtigung der Unsicherheit der Imputation

Quelle: Little und Rubin (2002)

- Multiple Imputation wurde von Donald B. Rubin in den 70er Jahre entwickelt
- Für jeden fehlenden Wert werden M Werte imputiert. Daraus resultieren M vollständige Datensätze
- Für jeden vollständigen Datensatz m mit $m = 1, \dots, M$ wird die interessierende Statistik θ geschätzt (angegeben durch $\hat{\theta}_m$) sowie der dazugehörigen Varianzschätzer \hat{V}_m

Quelle: Rubin (1987), Van Buuren (2012)

	Unimp.	m=1	m=2	m=3	m=4	m=5
Einheit 1	3	3	3	3	3	3
Einheit 2	NA	2	4	4	2	5
Einheit 3	NA	1	4	2	5	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Einheit n	4	4	4	4	4	4

	↓	↓	↓	↓	↓
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$

Quelle: Rubin (1987)

Punktschätzer	$\bar{\theta} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_m$
Within Imputationsvarianz	$\bar{W} = \frac{1}{M} \sum_{m=1}^M \hat{V}_m$
Between Imputationsvarianz	$B = \frac{1}{M-1} \sum_{m=1}^M (\hat{\theta}_m - \bar{\theta})^2$
Gesamtvarianz	$\bar{T} = \bar{W} + \frac{M+1}{M} \cdot B$

Quelle: Rubin (1987), Bjørnstadt (2007), Van Buuren (2012)

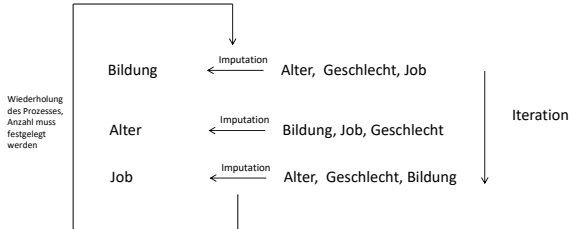
- Häufig Anwendung von Standard-Varianzschätzverfahren auf Basis imputierter Werte, die behandelt werden, als wären sie beobachtet worden
- ⇒ Vorgehen kann zu einer starken Unterschätzung der Varianz führen
- Bei Verwendung der einfachen Imputation werden spezielle Anpassungen bei der Varianzschätzung benötigt, um den Imputationsprozess zu berücksichtigen
- Beispiel Resampling Verfahren: Reimputation oder die Adjustierung der imputierten Werte in Shao (2002) oder die Methoden und Korrekturfaktoren in Mashreghi et al. (2014)
- Notwendigkeit der Berücksichtigung der Annahmen, auf denen die Verfahren basieren und insb. der Varianzkomponenten die sie genau schätzen

Quelle: siehe z.B. auch Bruch (2019) oder Bruch (2016)

- Mittelwertimputation
- Deterministische Regressionsimputation
- Stochastische Regressionsimputation
- Hot Deck Random Imputation
- Nearest Neighbour Imputation
- Predictive Mean Matching
- Classification and regression trees (CART)

Quellen: Van Buuren (2012), Little und Rubin (2002), Chen und Shao (2001), Rubin (1986), Little (1988), Koller-Meinfelder (2009)

Bildung, Alter und Job besitzen fehlende Werte
 Geschlecht ist vollständig beobachtet



Quelle: Van Buuren et al. (2006), Van Buuren und Groothuis-Oudshoorn (2011), Van Buuren (2012)

Gewichtung

Grundlage der Kalibrierung ist der designgewichtete Schätzer einer Erhebung.

$$\hat{\tau}_y = \sum_{i \in \mathcal{S}} \frac{1}{\pi_i} y_i = \sum_{i \in \mathcal{S}} d_i y_i$$

Basierend auf der Stichprobe \mathcal{S} bestimmt sich ein Schätzer der Populationsgröße (\hat{N}) anhand von:

$$\hat{N} = \sum_{i \in \mathcal{S}} d_i$$

- Im Falle des Horvitz-Thompson Schätzers wird das Designgewicht d_i für ein Element i berechnet über die Inverse seiner Inklusionswahrscheinlichkeit
- Unter Abwesenheit von "nicht-ignorierbarem" Nonresponse gilt der HT-Schätzer als unverzerrt
- Fehlende Werte sind im seltensten Fall MCAR

Grundlage der Kalibrierung ist der designgewichtete Schätzer einer Erhebung.

$$\hat{\tau}_y = \sum_{i \in S} \frac{1}{\pi_i} y_i = \sum_{i \in S} d_i y_i$$

Basierend auf der Stichprobe S bestimmt sich ein Schätzer der Populationsgröße (\hat{N}) anhand von:

$$\hat{N} = \sum_{i \in S} d_i$$

- Im Falle des Horvitz-Thompson Schätzers wird das Designgewicht d_i für ein Element i berechnet über die Inverse seiner Inklusionswahrscheinlichkeit
 - Unter Abwesenheit von "nicht-ignorierbarem" Nonresponse gilt der HT-Schätzer als unverzerrt
 - Fehlende Werte sind im seltensten Fall MCAR
- ⇒ Die meisten Kalibrierungsverfahren nehmen MAR an

Warum Kalibrierung?

- Zur Reduktion der Verzerrung des Schätzer aufgrund von Nonresponse (wenn MAR-Annahme zutrifft)
- Zur Reduktion der Varianz des Schätzers

Zur Bestimmung von Kalibrierungsgewichten werden immer Hilfsvariablen, für die Informationen in der Population und der Erhebung vorliegen, benötigt.

Arten der Kalibrierung:

- Poststratifizierung
- Raking-Verfahren
- Generalisierter Regressionsschätzer (GREG)

Häufige Fehler vermeiden:

- Kalibrierung kann keine mangelhafte Erhebungspraxis ausgleichen
- Populationsvariablen müssen sich alle auf die gleiche Population beziehen
- Kalibrierung sollte immer für einen designgewichteten Schätzer verwendet werden

$$\hat{Y} = \sum_{i \in \mathcal{R}} w_i y_i = \sum_{i \in \mathcal{R}} g_i d_i y_i$$

w_i entspricht dem Gesamtgewicht der Einheit i ; g_i dem Kalibrierungsgewicht

Wenn die **gemeinsame Verteilung** der Hilfsvariablen sowohl für die Erhebungs- als auch für die Populationsdaten bekannt sind, lassen sich die Poststratifizierungsgewichte (g_i^{Post}) anhand des Verhältnisses der Zellen (Poststrata, $h = 1, \dots, H$) der Zellpopulation (N_h) und deren Schätzwert (\hat{N}_h) bestimmen.

Die Zellpopulation berechnet sich nach:

$$N_h = \sum_{i \in \mathcal{R}_h} g_i^{Post} * d_i$$

Da $\hat{N}_h = \sum_{i \in \mathcal{R}_h} d_i$, gilt

$$g_i^{Post} = \frac{N_h}{\hat{N}_h} \quad \forall i \in \mathcal{R}_h$$

- Dieser Vorgang wird für jede Zelle durchgeführt
- Jedes Element der Zelle bekommt das gleiche Kalibrierungsgewicht

Gemeinsame Verteilung Population

	keine	niedrig	mittel	hoch
20 - 29	240	862	748	641
30 - 39	289	1086	895	763
40 - 49	209	729	590	447
50 - 59	164	523	452	379
60+	114	364	269	236

Designgewichtete gemeinsame Verteilung

	keine	niedrig	mittel	hoch
20 - 29	142,9	857,1	714,3	757,1
30 - 39	214,3	785,7	871,4	1000
40 - 49	114,3	600	700	528,6
50 - 59	157,1	457,1	542,9	557,1
60+	128,6	285,7	328,6	257,1

Poststratifizierungsgewichte

	keine	niedrig	mittel	hoch
20 - 29	1,68	1,01	1,05	0,85
30 - 39	1,35	1,38	1,03	0,76
40 - 49	1,83	1,22	0,84	0,85
50 - 59	1,04	1,14	0,83	0,68
60+	0,89	1,27	0,82	0,92

$$240/142,9 = 1,68$$

Voraussetzungen und Probleme:

- Gemeinsame Verteilung der Anpassungsvariablen ist in Erhebung und Population bekannt
 - ⇒ Oft für Populationswerte problematisch
- Steigende Anzahl von Hilfsvariablen kann zu Problemen führen
 - Verfügbarkeit
 - Zu viele Merkmalskombinationen führen (oft) zu (annähernd) unbesetzten Zellen
- Gemeinsame Verteilung in der Erhebung führt zu sog. "Strukturnullen"
- ⇒ Zellen sind in der Population besetzt, in der Erhebung jedoch nicht
- Eine Vielzahl von schwach besetzten Zellen kann zu einer großen Spannweite der Gewichte führen

- Raking beschreibt eine Art der Kalibrierung, bei der iterativ lediglich an die Randverteilung der Hilfsvariablen der Population angepasst wird
- Iterative Proportional Fitting (IPF) - Algorithmus von Deming and Stephan (1940)
- Wird verwendet, wenn:
 - Nur die Randverteilungen der Population bekannt sind
 - Die gemeinsame Verteilung der Hilfsvariablen strukturelle Nullen beinhaltet
 - Die gemeinsame Verteilung mit anderen Problemen behaftet ist (bspw. wenige Beobachtungen in einzelnen Zellen)
- Die Randverteilungen der Hilfsvariablen stimmen in SP und ZP (annähernd) überein
- Die gemeinsame Verteilung der Hilfsvariablen jedoch nicht

Vorgehen bei 2 Hilfsvariablen:

- 1 Die Randverteilung der ersten Variable in der Population wird durch deren Schätzwert geteilt (N_k/\hat{N}_k). Die daraus resultierenden Gewichte werden zur Schätzung der zweiten Randverteilung verwendet.

Vorgehen bei 2 Hilfsvariablen:

- 2 Die Randverteilung der zweiten Variablen wird durch deren (neuen) Schätzwert geteilt. Die daraus resultierenden Gewichte werden zur Schätzung der ersten Randverteilung verwendet.

Vorgehen bei 2 Hilfsvariablen:

- ... Schritt 1 und 2 werden so lange wiederholt, bis sich keine relevante Veränderung mehr zwischen den Iterationen einstellt und die Randverteilungen beider Variablen in SP und ZP übereinstimmen

Randverteilung Population

Alter

20 - 29	30 - 39	40 - 49	50 - 59	60+
2491	3033	1975	1518	983

Bildung

keine	niedrig	mittel	hoch
1016	3564	2954	2466

Geschätzte Verteilung

Start

	keine	niedrig	mittel	hoch	
20 - 29	142,9	857,1	714,3	757,1	2471,4
30 - 39	214,3	785,7	871,4	1000	2871,4
40 - 49	114,3	600	700	528,6	1942,9
50 - 59	157,1	457,1	542,9	557,1	1714,3
60+	128,6	285,7	328,6	257,1	1000
	757,1	2985,7	3157,1	3100	

Randverteilung Population

Alter

20 - 29	30 - 39	40 - 49	50 - 59	60+
2491	3033	1975	1518	983

Bildung

keine	niedrig	mittel	hoch
1016	3564	2954	2466

Geschätzte Verteilung

Schritt 1: Alter

	keine	niedrig	mittel	hoch	
20 - 29	144	863,9	719,9	763,1	2491
30 - 39	226,3	829,9	920,5	1056,3	3033
40 - 49	116,2	609,9	711,6	537,3	1975
50 - 59	139,1	404,8	480,7	493,4	1518
60+	126,4	280,9	323,0	252,8	983
→	752	2989,4	3155,7	3102,8	

Gewichte: 1,01, 1,06, 1,02, 0,89, 0,93

Randverteilung Population

Alter

20 - 29	30 - 39	40 - 49	50 - 59	60+
2491	3033	1975	1518	983

Bildung

keine	niedrig	mittel	hoch
1016	3564	2954	2466

Geschätzte Verteilung

Schritt 2: Bildung

	keine	niedrig	mittel	hoch	↓
20 - 29	194,5	1030,0	673,9	606,5	2504,9
30 - 39	305,8	989,4	861,6	839,5	2996,3
40 - 49	157,0	727,2	666,1	427,0	1977,2
50 - 59	188,0	482,6	450,0	392,1	1512,7
60+	170,7	334,8	302,3	200,9	1008,8
	1016	3564	2954	2466	

Gewichte: 1,35, 1,19, 0,94, 0,79

Ende der ersten Iteration

Randverteilung Population

Alter

20 - 29	30 - 39	40 - 49	50 - 59	60+
2491	3033	1975	1518	983

Bildung

keine	niedrig	mittel	hoch
1016	3564	2954	2466

Geschätzte Verteilung

Nach 3 Iterationen

	keine	niedrig	mittel	hoch	
20 - 29	193,7	1024,6	670,2	602,5	2491
30 - 39	309,9	1002,0	872,3	848,9	3033
40 - 49	157,0	726,6	665,4	426,1	1975
50 - 59	188,9	484,5	451,6	393,1	1518
60+	166,6	326,3	294,6	195,5	983
	1016	3564	2954	2466	

Kalibrierungsgewichte: 1,4, 1,2, 0,9, 0,8, 1,4, 1,3, 1,0, 0,8, 1,4, 1,2, 1,0, 0,8, 1,2, 1,1, 0,8, 0,7, 1,3, 1,1, 0,9, 0,8

Für komplexe Erhebungen kann der Hovitz-Thompson Schätzer so verändert werden, dass dieser einem Verhältniss- oder Regressionsschätzer entspricht und den Vektor der Hilfsvariablen \mathbf{x} beinhaltet.

⇒ Der GREG-Schätzer kombiniert Designgewichtung und Kalibrierung direkt

$$\hat{Y}^{GREG} = \hat{Y}^{HT} + \left(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in S} \mathbf{x}_i * d_i \right)' * \hat{\mathbf{B}}$$

Die Anpassung des Designgewichtes g_i^{GREG} berechnet sich nach:

$$g_i^{GREG} = \frac{1}{q_i} \left(1 + c_i \left(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in R} \frac{1}{\pi_i q_i} \mathbf{x}_i \right)' \left(\sum_{i \in R} \frac{c_i}{\pi_i q_i} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_i \right)$$

Eigenschaften:

- Berücksichtigt Auswahlwahrscheinlichkeiten (und Sampling Design)
 - Kann zur Varianzreduktion des Punktschätzers beitragen
 - ⇒ Effekt direkt durch Zusammenhang zwischen x und y bedingt
 - Flexibel anwendbar aufgrund verschiedener Möglichkeiten, weitere Gewichte zu integrieren
 - Approx. unverzerrt
- ⇒ Auch ohne Nonresponse ratsam
- ⇒ Da es grundsätzlich ein Regressionsschätzer ist, können auch negative Gewichte entstehen

- In einem Unternehmen arbeiten **1.000 Frauen** und **300 Männer**; eine SP von **100** Personen wird zufällig gezogen.
- **30 Männer** und **50 Frauen** antworten bezüglich ihrer Arbeitszufriedenheit
- **20 Männer** und **10 Frauen** zufrieden
- Annahme: 0,3 aller Männer und 0,5 aller Frauen antworten grundsätzlich auf Befragungen

$$20 \frac{1}{\frac{100}{300} 0,3} + 10 \frac{1}{\frac{100}{1000} 0,5} = 400$$

⇒ Ca. 31 % sind zufrieden

- In einem Unternehmen arbeiten **1.000 Frauen** und **300 Männer**; eine SP von **100** Personen wird zufällig gezogen.
- **30 Männer** und **50 Frauen** antworten bezüglich ihrer Arbeitszufriedenheit
- **20 Männer** und **10 Frauen** zufrieden
- Annahme: 0,24 aller Männer und 0,5 aller Frauen antworten grundsätzlich auf Befragungen

$$20 \frac{1}{\frac{100}{300} 0,24} + 10 \frac{1}{\frac{100}{1000} 0,5} = 450$$

- Unzufriedene Mitarbeiter:

$$10 \frac{1}{\frac{100}{300} 0,24} + 40 \frac{1}{\frac{100}{1000} 0,5} = 925$$

⇒ Geschätzte AZ Mitarbeiter: **1.375**

- In einem Unternehmen arbeiten **1.000 Frauen** und **300 Männer**; eine SP von **100** Personen wird zufällig gezogen.
- **30 Männer** und **50 Frauen** antworten bezüglich ihrer Arbeitszufriedenheit
- **20 Männer** und **10 Frauen** zufrieden
- Kalibrierung durch $g_i^{GREG^*}$, mit $x_i = 1$, $c_i = 1$ und $\pi^M = 1/3$, $\pi^F = 1/10$

$$g_i^{GREG^*} = \frac{1}{\pi_i q_i} * \frac{N}{\sum_{i \in \mathcal{R}} \frac{1}{\pi_i q_i}} = 130/11$$

- für Männer und 208/11 für Frauen

$$20 \frac{130}{11} + 10 \frac{208}{11} = 425,45; 10 \frac{130}{11} + 40 \frac{208}{11} = 874,55$$

⇒ Ca. 32,72 % zufrieden

Imputation:

- Zentral ist die Bildung von Imputationsmodellen
 - Bestimmung guter Prädiktoren
 - Insb. auch deren Anzahl
- Geeignete Imputationsmethode auswählen, die sich am konkreten Anwendungsfall orientiert
- Bei der Varianzschätzung ist es wichtig, den Imputationsprozess zu berücksichtigen

Kalibrierung:

- Kann Reduktion von Verzerrung bewirken
 - Abhängig von Zusammenhang der Hilfsvariablen mit dem Ausfallprozess
 - Keine vollständige Behebung von Verzerrung
- Auch ohne Nonresponse sinnvoll
- Alle Subgruppen, für die Analysen gemacht werden, sollten enthalten sein

**VIELEN DANK FÜR IHRE
AUFMERKSAMKEIT**

- Bjørnstad, J. F. (2005). Non-Bayesian Multiple Imputation. Discussion Papers 421, Statistics Norway, Research Department.
- Bondarenko, I. and Raghunathan, T. (2016). Graphical and numerical diagnostic tools to assess suitability of multiple imputations and imputation models. *Statistics in medicine*, 35:3007–3020.
- Bruch, C. (2016). *Varianzschätzung unter Imputation und bei komplexen Stichprobendesigns*. doctoralthesis, Universität Trier.
- Bruch, C. (2019). Applying the rescaling bootstrap under imputation: a simulation study. *Journal of Statistical Computation and Simulation*, 89(4):641–659.
- Buuren, S. V., Brand, J. P., Groothuis-Oudshoorn, C. G., and Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12):1049–1064.
- CHAUVET, G., DEVILLE, J.-C., and HAZIZA, D. (2011). On balanced random imputation in surveys. *Biometrika*, 98(2):459–471.
- Chen, J. and Shao, J. (2001). Jackknife variance estimation for nearest-neighbor imputation. *Journal of the American Statistical Association*, 96(453):260–269.
- Collins, L. M., Schafer, J. L., and Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4):330–351.
- Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.*, 11(4):427–444.
- Gabler, S., Kolb, J.-P., Sand, M., and Zins, S. (2015). *Gewichtung*.
- Haziza, D. (2009). Imputation and inference in the presence of missing data. In Pfeffernann, D. and Rao, C., editor, *Handbook of Statistics 29A: Sample Surveys: Design, Methods and Applications*, chapter 10, pages 215–246. Elsevier BV, Amsterdam.
- Kim, J. K. and Rao, J. N. K. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika*, 96(4):917–932.
- Koller-Meinleider, F. (2010). *Analysis of Incomplete Survey Data - Multiple Imputation via Bayesian Bootstrap Predictive Mean Matching*. doctoralthesis, Otto-Friedrich-Universität Bamberg.
- Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, 6(3):287.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley and Sons.
- Lohr, S. L. (2009). *Sampling - Design and Analysis*. CENGAGE Learning.
- Lumley, T. (2010). *Complex Surveys*. John Wiley & Sons, Inc.
- Mashreghi, Z., Léger, C., and Haziza, D. (2014). Bootstrap methods for imputed data from regression, ratio and hot-deck imputation. *Canadian Journal of Statistics*, 42(1):142–167.
- Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statist. Sci.*, 9(4):538–558.
- Nicoletti, C. and Peracchi, F. (2006). The effects of income imputation on microanalyses: evidence from the european community household panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3):625–646.
- Raghunathan, T. E. and Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, 90(429):54–63.
- Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*, 4(1):87.
- Rubin, D. B., editor (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc.
- Sand, M. and Gabler, S. (2018). Gewichtung von (dual-frame -) telefonstichproben. In *Arbeitsbuch Telefonstichproben*.
- Särndal, C.-E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*.
- Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.
- Shao, J. (2002). Replication methods for variance estimation in complex surveys with imputed data. In Groves, R. M., Dilman, D. A., Eltinge, J. L., and A. R. J., editors, *Survey Nonresponse*, pages 303–314. Wiley, New-York.
- Valliant, R., Dever, J. A., and Kreuter, F. (2018). *Practical Tools for Designing and Weighting Survey Samples*. Springer International Publishing.
- van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Chapman and Hall/CRC.
- van Buuren, S. and Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–65.

Beratung von wissenschaftlichen Forschungsprojekten

GESIS bietet individuelle Beratung, um geeignete Lösungen zur Umsetzung Ihres wissenschaftlichen Umfrageprojekts zu finden.

Wer wird beraten?

- Kostenfrei beraten werden Wissenschaftlerinnen und Wissenschaftler, wenn sie
 - ▶ wissenschaftliche institutionelle oder Drittmittelprojekte an Hochschulen und öffentlich finanzierten Forschungsinstituten durchführen
 - ▶ wissenschaftliche Projekte an Einrichtungen von Bund und Ländern oder sonstigen öffentlichen Einrichtungen durchführen.
- Für weitere Projekte bieten wir bei vorhandener Kapazität kostenpflichtige Beratung an.



Kontakt Referenten: christian.bruch@gesis.org
matthias.sand@gesis.org

Allgemeine Projektberatung: [hotline projektberatung@gesis.org](mailto:hotline_projektberatung@gesis.org)

Website: <https://www.gesis.org/angebot/studien-planen-und-daten-erheben/projektplanung>

Weitere Angebote

- In den *GESIS Survey Guidelines* finden Sie kurze, praxisorientierte Texte zu häufig wiederkehrenden Beratungsthemen
<https://www.gesis.org/gesis-survey-guidelines/home>
- Bleiben Sie auf dem Laufenden über Neuigkeiten aus dem Institut
<https://www.gesis.org/institut/presse-und-medien/gesis-report>
- Nehmen Sie an wissenschaftlichen Weiterbildungsveranstaltungen teil
<https://www.gesis.org/angebot/wissen-vermitteln/gesis-training>
- Besuchen Sie unseren GESIS-Blog *Growing Knowledge in the Social Sciences*
<https://blog.gesis.org/>