



Medien Institut



Statistik II: Datenbereinigung

Dr. Andreas Vlašić

Medien Institut

☎ (0621) 52 67 44

✉ vlastic@medien-institut.de



Praxisbeispiel: Evaluation einer Imagekampagne - Fragestellung

- Ziele der Kampagne:
 - ➔ Steigerung der Bekanntheit des Unternehmens
 - ➔ Markenprofilierung und Imageaufbau
 - ➔ Gewinnung von Sympathieträgern in der Zielgruppe Meinungsführer
- Maßnahme: Anzeigenmotive in ausgewählten Printmedien (*Focus*, *SPIEGEL*, *Wirtschaftswoche* und *Capital*)
- Grundgesamtheit: Wirtschafts- oder Politikexperten mit Ratgeberfunktion, die zumindest gelegentlich *Focus*, *SPIEGEL*, *Wirtschaftswoche* oder *Capital* lesen
- Methode: Telefonische Befragung (CATI) von jeweils 300 Personen der Zielgruppe (gemäß Selbsteinschätzung nach AWA) in zwei Wellen
- Stichprobe: Gestufte Zufallsauswahl
 - ➔ zufällige Auswahl von Wohnorten und Telefonnummern
 - ➔ Zielperson: Last Birthday
 - ➔ Experten-Screening

Aufbau des Fragebogens

1.) Screening von VPN durch Filterfragen

- Frage 1: Interesse am Thema Politik
- Frage 2: Meinungsführerschaft beim Thema Politik
- Frage 3: Interesse am Thema Wirtschaft
- Frage 4: Meinungsführerschaft beim Thema Wirtschaft
- Frage 5: Lesen der Zeitschriften

2.) Kern

- Frage 6: Spontane Assoziationen zu Unternehmen X
- Frage 7: Eigenschaften des Unternehmens X
- Frage 8: Öffentliche Wahrnehmung des Unternehmens X
- Frage 9: Aufgaben des Unternehmens X (offen)
- Frage 10: Aufgaben des Unternehmens X
- Frage 11: Wichtigkeit des Unternehmens X

3.) Soziodemographie

- Frage 12: Alter
- Frage 13: Bildung
- Frage 14: Beruf
- Frage 15: Einkommen

Vom Fragebogen zur Datenmatrix

Frage 1: Sie haben also schon einmal vom Unternehmen X gehört. Wenn Sie sich nun bitte einmal das Unternehmen X vor ihrem geistigen Auge vorstellen, welche Eigenschaften bringen Sie mit ihr in Verbindung? Ich nenne Ihnen gleich ein paar Eigenschaften und Sie sagen mir bitte jeweils, wie stark diese Ihrer Meinung nach auf das Unternehmen X zutreffen:

	trifft voll zu	trifft zu	trifft weniger zu	trifft nicht zu	weiß nicht
Das Unternehmen X ist ...					
▪ sympathisch					
▪ vertrauenswürdig					
▪ kompetent					
▪ fair					
▪ ehrlich					
▪ modern					
▪ freundlich					
▪ bürokratisch					
▪ verlässlich					
▪ flexibel					
▪ dynamisch					
▪ innovativ					
▪ entgegenkommend					
▪ engagiert					

Vom Fragebogen zur Datenmatrix /2

- Zuordnen von Codierungen für Antworten -> Codeplan

Frage 1

Eigenschaft	Variablenname	Ausprägungen				
		trifft voll zu	trifft zu	trifft weniger zu	trifft nicht zu	weiß nicht
sympathisch	e_symp	1	2	3	4	5
vertrauenswürdig	e_vert	1	2	3	4	5
kompetent	e_komp	1	2	3	4	5
fair	e_fair	1	2	3	4	5
ehrlich	e_ehrl	1	2	3	4	5
modern	e_mode	1	2	3	4	5
freundlich	e_freu	1	2	3	4	5
bürokratisch	e_buer	1	2	3	4	5
verlässlich	e_verl	1	2	3	4	5
flexibel	e_flex	1	2	3	4	5
dynamisch	e_dyna	1	2	3	4	5
innovativ	e_inno	1	2	3	4	5
entgegenkommend	e_entg	1	2	3	4	5
engagiert	e_engag	1	2	3	4	5

Mehrfachantworten

- Wenn zu einer Frage mehrere Antworten gleichzeitig gegeben werden können, müssen diese in der Datenmatrix auf mehrere Variablen verteilt werden („Normalisierung“)
- Analyse in SPSS unter „Mehrfachantworten“

Frage: Welche Faktoren sind für Sie beim Kauf eines Computers wichtig?
(*Mehrfachantworten möglich!*)

- 1 = neueste Technik
- 2 = gute Hotline
- 3 = schickes Aussehen



	Faktoren		
VPN 1	1, 3		
VPN 2	2		
VPN 3	2, 3		
...			

	Faktor_1	Faktor_2	Faktor_3
VPN 1	1	0	1
VPN 2	0	1	0
VPN 3	0	1	1
...			

Erste Schritte der Datenbereinigung

- Überblick über Daten verschaffen
- Fehleingaben identifizieren
 - ➔ fehlerhafte Codierungen
 - ➔ unplausible Codierungen
- Verarbeitung von fehlenden Werten
 - ➔ „weiß nicht“ vs. „keine Antwort“
 - ➔ Systemdefinierte fehlende Werte
 - ➔ Ergänzung von fehlenden Werten
- Zusammenfassen von Variablen (Bsp. Altersklassen)
- Fehlerhafte Daten zeigen sich oftmals erst bei der Auswertung!
 - ➔ Zwischenstände
 - ➔ Syntaxdatei

Key Facts

- Praxisbeispiel:
 - ➔ Forschungsdesign
 - ➔ Definition und Rekrutierung der Stichprobe
 - ➔ Umsetzung des Fragebogens
- Was versteht man unter einem Codeplan?
- Welche ersten Schritte gibt es bei der Datenbereinigung?
- Was sind „fehlende Werte“, und wie werden sie behandelt?
- Wie importiert man Daten aus Excel?
- Wie kann das Labeln von Variablen und Werten automatisiert werden?