

Wie funktioniert Statistik?



- Wer sind die eigentlich?
 - Der Trick mit der Stichprobe!
- Was ist eigentlich wieviel von was?
 - Jonglieren mit Prozenten!
- Und wer fehlt hier?
 - Missingwerte und ihr Tücken
- Was hat sich wie verändert?
 - Das Kreuz mit den Kreuztabellen

Stichprobe

- „Mehr als die Hälfte (63,5 Prozent) der HSV-Fans hat einen Hochschulabschluss“
- Quelle: [SHZ](#)
- Stichprobe: 45.000 Fußballfans
- und zwar in Profilen von XING-Mitgliedern

Teilkategorie Produktqualität			
Kundenbefragung Antifalten-Creme 2014			
Unternehmen	Punkte*	Rang	Qualitätsurteil
Sebamed	85,0	1	sehr gut
Garnier	83,5	2	
Helena Rubinstein	81,2	3	
Christian Dior	81,2	4	
Estée Lauder	80,9	5	
Dr. Hauschka	80,4	6	
L'Oréal	80,3	7	
Olaz	79,9	8	gut
Nivea	79,5	9	
Biotherm	79,3	10	
Chanel	79,1	11	
Weleda	78,1	12	
Vichy	77,6	13	
Avène	77,0	14	
Shiseido	76,8	15	
Aok	76,5	16	
Balea	76,1	17	
Louis Widmer	75,3	18	
Florena	75,3	19	
Eucerin	74,8	20	
Clinique	74,7	21	
Lancôme	72,4	22	
Lancaster	71,3	23	
Diadermine	70,4	24	
Alterra	68,7	25	befriedigend
Rival de Loop	67,3	26	
Sonstige**	78,5		

„Berücksichtigt wurden alle Marken, zu denen sich jeweils mindestens 80 Kunden geäußert hatten.“

Mittelwerte

- A: Meine Klienten sind im Durchschnitt 40 Jahre alt
- B: Auch meine Klienten haben ein Durchschnittsalter von 40 Jahren
- A: Prima, dann haben wir ja ein ganz ähnliches Klientel
- C: Komisch, meine Klienten sind eigentlich alle Anfang 20, aber mein Mittelwert sagt 40 Jahre?

- A:
- Mittelwert 40,0

- B:
- Mittelwert 40,0

- C:
- Mittelwert 40,0

- A hat 6 Klienten:
- 18, 20, 22, 58, 60 und 62 Jahre

- B hat ebenfalls 6 Klienten
- 38, 39, 40, 40, 41, 42 Jahre

- C hat ebenfalls 6 Klienten
- 20, 23, 24, 25, 25 Jahre

- A:
- Mittelwert 40,0
- Standardabweichung 21,98
- B:
- Mittelwert 40,0
- Standardabweichung 1,41
- C:
- Mittelwert 40,0
- Standardabweichung 39,21 (??)

- Schauen wir die eingegebenen Werte zu C noch einmal genau an:
- 120, 23, 24, 25, 25 Jahre

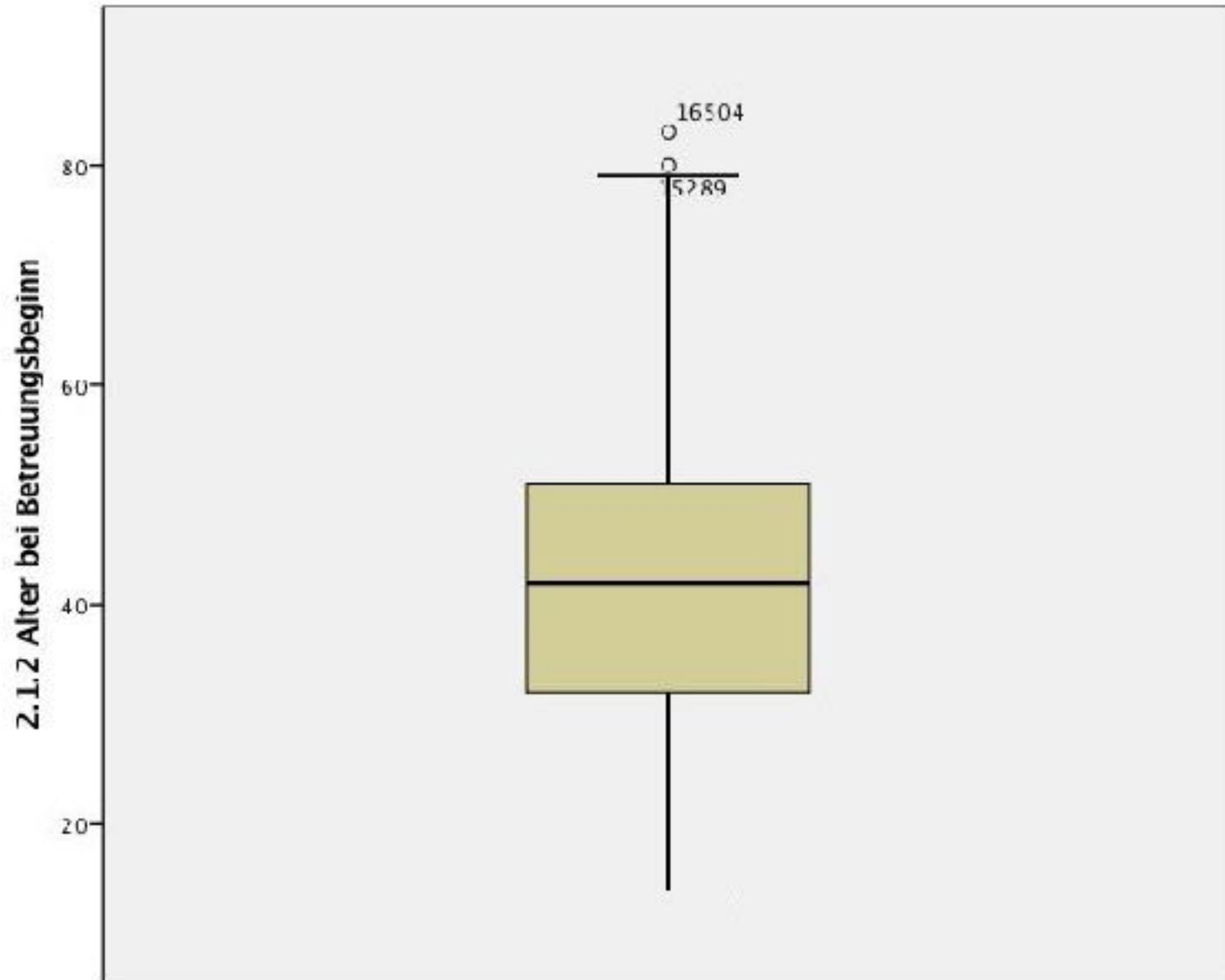
Statistiken

2.1.2 Alter bei Betreuungsbeginn

N Gültig 19097

Fehlend 0

Mittelwert	41,50
Median	42,00
Modus	48
Std.abweichung	12,176
Minimum	14
Maximum	83



		A	B	C
N	Gültig	6	6	6
Mittelwert		40,0	40,0	40,0
Median		40,0	40,0	24,5
Modus		18*	40	23*
Standardabweichung		21,98	1,41	39,20

*Mehrere Modi, der niedrigste Wert wurde verwendet

Der Median ist deutlich „robuster“ gegen Ausreißer.

Prozente

- Der Anteil der weiblichen Angestellten konnte im vergangenen Jahr um 100% gesteigert werden.
- Frau Weber hat eine Kollegin bekommen

- Eine dänische Untersuchung besagt:
 - 25 % der Jungen fühlen sich in der KiTa unwohl
 - 10 % der Mädchen fühlen sich in der KiTa unwohl
- Damit fühlen sich mehr als $\frac{1}{3}$ der Kinder in der KiTa unwohl !!!

- In der Kita sind 80 Jungen und 100 Mädchen
- 20 Jungen (25%) fühlen sich unwohl
- 10 Mädchen (10%) fühlen sich unwohl
- damit fühlen sich 30 Kinder unwohl
- 30 von 180 sind 16,7 %

- Nach der Armutsstudie des Paritätischen liegt die Armutsquote bei 15,5%
- – definiert als die Menschen, die pro Monat weniger als 60 Prozent des Durchschnittseinkommens zur Verfügung haben
- Dieser Prozentsatz bleibt gleich, auch wenn sich das reale Einkommen aller Bundesbürger verdoppelt

- Wien im Januar 2015:
- die Durchschnittstemperatur betrug 3,5 ° Celsius
- das langjährige Mittel beträgt 0,8° Celsius
- eine Steigerung um 2,7° Celsius = **337%** (2,7/0,8*100)
- Dramatisch !! ??
- Rechnen wir das Ganze mal um in ° Fahrenheit
- 3,5° C = 38,5° F / 0,7° C = 33,4° F
- Steigerung um 5,1° F = **15,3 %** (5,1/33,4*100)

Missingwerte

2.3.11 Problematische Schulden

	Häufigkeit	Prozent	Häufigkeit	Gültige Prozente
Keine Angabe	1262	15,3		
Keine Schulden	4638	56,1	4638	66,1
Bis 10.000 Euro	1232	14,9	1232	17,6
Bis 25.000 Euro	541	6,5	541	7,7
Bis 50.000 Euro	334	4,0	334	4,8
Über 50.000 Euro	267	3,2	267	3,8
Gesamt	8274	100,0	7012	100,0

- 56,1% haben keine Schulden
 - 28,6% haben Schulden
 - 15,3% machen keinen Angabe
-
- Sind alle ohne Angabe schuldenfrei, so wären 71,4% ohne Schulden
 - Haben alle ohne Angaben Schulden, so wären 43,9% mit Schulden

- Schuldenfrei sind
56,1% oder 66,1% oder 71,4%
- Mit Schulden sind
28,4% oder 33,9% oder 43,9%

Kreuztabellen

Behandlungsbeginn	Erwerbssituation Katamnesezeitpunkt				
	Keine Angabe	Erwerbstätig	Arbeitslos	Nicht erwerbstätig	Gesamt
Keine Angabe	0	32	23	18	73
Erwerbstätig	30	1064	196	196	1480
Arbeitslos	13	367	542	163	1085
Nicht erwerbstätig	1	74	39	382	496
Gesamt	44	1537	794	759	3134

Sind es 50,0% oder 68,3% oder 17,3%

50,0 % der zu Beginn Arbeitslosen sind auch zur Katamnese arbeitslos
(Basis 1085)

68,3 % der zur Katamnese Arbeitslosen waren auch zu Beginn arbeitslos
(Basis 794)

17,3 % aller Klienten waren zu Beginn und zur Katamnese arbeitslos
(Basis 3134)

Behandlungsbeginn	Erwerbssituation Katamnesezeitpunkt			
	Erwerbstätig	Arbeitslos	Nicht erwerbstätig	Gesamt
Erwerbstätig	1064	100	196	1450
Arbeitslos	367	542	163	1072
Nicht erwerbstätig	74	59	382	495
Gesamt	1505	771	741	3017

Sind es 50,6% oder 70,3% oder 18,0%

50,6 % der zu Beginn definitiv Arbeitslosen sind auch zur Katamnese definitiv arbeitslos (Basis 1072)

70,3 % der zur Katamnese definitiv Arbeitslosen waren auch zu Beginn definitiv arbeitslos (Basis 771)

18,0 % aller Klienten mit Angaben waren zu Beginn und zur Katamnese arbeitslos (Basis 3017)

Zeilenprozent	Erwerbssituation Katamnese			
	Erwerbstätig	Arbeitslos	Nicht erwerbstätig	Gesamt
Behandlungsbeginn				
Erwerbstätig	73,4 %	13,1 %	13,5 %	100,0 %
Arbeitslos	34,2 %	50,6 %	15,2 %	100,0 %
Nicht erwerbstätig	14,9 %	7,9 %	77,2 %	100,0 %
Gesamt	49,9 %	25,6 %	24,6 %	100,0 %

Spaltenprozent	Erwerbssituation Katamnese			
	Erwerbstätig	Arbeitslos	Nicht erwerbstätig	Gesamt
Behandlungsbeginn				
Erwerbstätig	70,7 %	24,6 %	26,5 %	48,1 %
Arbeitslos	24,4 %	70,3 %	22,0 %	35,5 %
Nicht erwerbstätig	4,9 %	5,1 %	51,6 %	16,4 %
Gesamt	100,0 %	100,0 %	100,0 %	100,0 %

Tabellenprozent	Erwerbssituation Katamnese			
	Erwerbstätig	Arbeitslos	Nicht erwerbstätig	Gesamt
Behandlungsbeginn				
Erwerbstätig	35,3 %	6,3 %	6,5 %	48,1 %
Arbeitslos	12,2 %	18,0 %	5,4 %	35,5 %
Nicht erwerbstätig	2,5 %	1,3 %	12,7 %	16,4 %
Gesamt	49,9 %	25,6 %	24,6 %	100,0 %

- Stichprobe genau beschreiben
 - Auswahl und Entstehung
 - z.B alle entlassenen Klienten im Jahr
- Stichprobe nicht wechseln
 - und wenn, dann wieder beschreiben warum gewechselt wurde
- Stichprobe mit Bedacht auswählen
 - Begonnene Behandlungen und Auswertung auf Entlassungsform?

- Mittelwerte nie allein angeben
 - mindestens mit Standardabweichung
 - besser noch mit Median und Modalwert
- Für Mittelwertvergleiche gibt es spezielle statistische Verfahren
- Bei der Berechnung auf „Künstliche“ Messings achten
 - 99=unbekannt

- Bei Prozentwerten immer die Basis angeben
- Prozente addieren oder subtrahieren nur bei gleicher Basis

- In Kreuztabellen zunächst mit absoluten Zahlen arbeiten
- Genau überlegen welche Prozente angegeben werden
- Bei Verläufen vorzugsweise Tabellenprozente
- Missingwerte genau beachten und ggf. herausrechnen

- Korrelation ist nicht immer gleich ein kausaler Zusammenhang

In Dörfern mit vielen Storchennestern werden mehr Kinder geboren, als in Dörfern mit wenig Storchennestern.

Also bringen die Störche Kindersegen!

Mögliche tatsächliche Erklärung:

Dörfer mit vielen Nestern haben schlichtweg mehr Häuser in den Menschen leben.