

Wiederholung



Aufgabe Creme: Sie haben eine Feuchtigkeitscreme gekauft. Auf der Verpackung lesen Sie „Sie hilft zu 78 %“ (getestet an 16 Konsumenten).

- nur 12 von 16“ 78 → nicht 100 % aber schon viel (100 % vermutlich unerreichbar, gut wären 95 %)
- Formulierung richtiger „sie hilft bei 78 %“
- Stichprobengröße zu klein → ungenaues Ereignis

Aufgabe Einkommen der Nachbarn: Drei Ihrer Nachbarn verdienen monatlich 1000 Euro und einen der 10000 Euro verdient. Würden Sie nun sagen, dass ihre Nachbarn im Durchschnitt 3250 Euro verdienen?

- Kann man sagen, Mittelwert ist richtig errechnet.
- Bei 4 Werten könnte man auch jeden einzelnen nennen.

ABER:

- Mittelwert suggeriert Daten zwischen 3000 und 4000 Euro (also leichte Abweichungen)
- Die Spannweite / Variation ist groß → genau hinschauen → keine Verallgemeinerung
- Ausreißer „verfälschen“ aber sollten doch bedacht werden

Aufgabe Film: Sie möchten einen Film anschauen. Zuvor betrachten sie die Bewertung des Films. Die mögliche Bewertung geht von 1 (sehr schlecht) bis 10 (sehr gut). Unter Männern vergeben 6,9 % 10 Punkte und 16,7 % nur einen Punkt. 19 % der Frauen geben 10 Punkte und 9 % einen Punkt. Wer ist die Zielgruppe des Films?

- Frauen (oder Pärchen)
 - Wir sehen hier nicht, wie viel Leute eine Bewertung abgegeben haben → das ist aber ok
 - Höhere Prozentzahl an Frauen als Männer geben positive Bewertung und
 - Höhere Prozentzahl (fast 1/5) der Frauen gibt höhere als schlechtere Punkte.
- Versuchen zu veranschaulichen
- Es fehlen Angaben für Bewertungen von 2 - 9 → macht Aussage schwierig
 - Möglich: dass Männer zu 80 % 9 gewählt haben.

Warum Statistik?

1. Abhängigkeit von Noten und besuchten Übungsstunden. Es ist ein Scatterplott/Streudiagramm

Weitere Abhängige Variablen z.B. Fähigkeiten.

3 Gruppen → siehe rechts

Allgemeine Vermutungen:

zu viel Übung → Unsicherheit

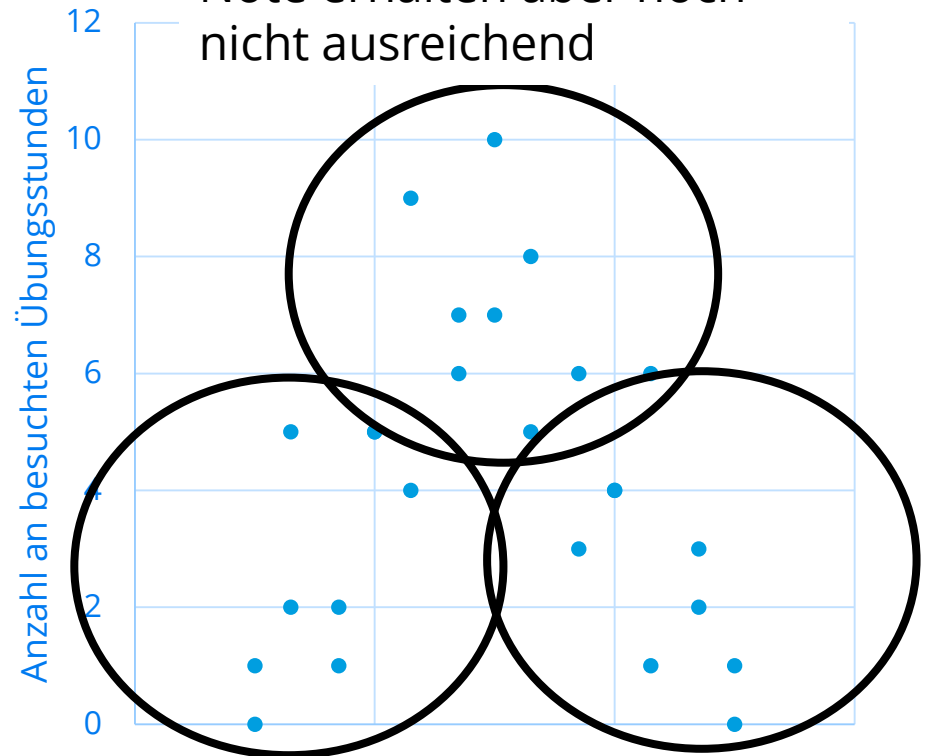
zu viel Übung → Überlastung

Übungsstunden bringen nix

Es ist mehr als nur x führt zu mehr y

U-förmige Zusammenhang ist definitiv interessant!

Gruppe 1: Viele Übungsstunden besucht und dadurch (?) relativ gute Note erhalten aber noch nicht ausreichend



Gruppe 3: Sehr gute Studenten, die die Übungsstunden nicht brauchen (?)

Statist

Gruppe 2: Keine Übungsstunden besucht (faul) und deshalb (?) schlechte Note.

Warum Statistik? Lügen aufdecken

Aufgabe Mitarbeiter prüfen:

Ein Mitarbeiter behauptet, dass er täglich 100 Stück eines Produktes herstellen kann. Nach dem sie 10 Tage lang seine Produktion überprüft haben, stellen Sie fest, dass der Mitarbeiter im Durchschnitt nur 99 Stück herstellt. Hat der Mitarbeiter gelogen?

Dauer: 1 Minuten **Abgabeformat:** KEINE

NEIN: Der Mitarbeiter hatte vielleicht einen schlechten Tag
Natürliche Schwankung
99 liegt im Toleranzbereich

→ Was ist der Toleranzbereich: Hier muss man eine Grenze setzen
Toleranz ist abhängig von Art des Untersuchungsobjekts

Warum Statistik? Definition / Messung von Konstrukten

Die Zahl der Armen fällt!

- Definition hat sich verändert → Die Einkommensgrenze wird verändert
- Falls basierend auf Umfragedaten, wird es noch schwieriger

Die Zahl an Krebsfällen steigt!

- Bessere Diagnostik/mehr Tests → mehr Fälle werden erkannt
- Art der Stichprobe (pensionierte Bergarbeiter oder junge Sportstudenten)

Aufgabe Arme und Krebsfälle:

In einer Zeitung stehen die Überschriften „Die Zahl der Armen fällt!“ und „Die Zahl an Krebsfällen steigt!“

Wie kann es zu beiden Ergebnissen kommen, ohne das in Wirklichkeit sich etwas geändert hat?

(Die Überschrift dieser Folie gibt Ihnen einen Hinweis)

Dauer: 5 Minuten

Abgabeformat: Textdatei

Warum Statistik? Formulierung von Fragen

Statistik braucht Daten. Daten müssen einwandfrei sein. Dazu gehören auch gute Fragen in einem Fragebogen.

Sind Sie nicht auch der Meinung, dass man mehr für den Umweltschutz tun sollte?

Ja

Nein

Fragen sollten objektiv sein, diese Frage ist

- suggestiv und
- nicht wertfrei
- enthält eine doppelte Verneinung
- Unspezifisch/schlecht operationalisiert:
Was ist Umweltschutz
- (umgangssprachlich)

Variablen und Fälle

	A	B	C
1	Name	Geschlecht	Note
2	Mara	1	1
3	Silke	1	2
4	Gerd	2	3
5	Klaus	2	4
6	Peter	2	4
7	Ilse	1	3
o			



Zusätzliche Informationen: Ordinale Daten können manchmal als metrisch interpretiert werden

Ordinal oder metrisch.

Skala von 1 bis 5



Aufgabe Fälle/Variablen:

1. Was kann von dem Folgendem alles ein **Fall** sein?: Menschen, Personen, Gruppen, Staaten, Organisationen, Planeten

→ ALLES

2. Was von dem Folgendem kann eine **Variable** sein? Alter, Geschlecht, IQ oder Größe, Jahresgewinn eines Unternehmens

→ ALLES



Übung Dateneingabe 1: Lösungen

Name	Alter	Geschlecht	Note durchschnitt	Wohnort (Stadtteil)	Abitur 1 = ja 2 = nein	Wie oft Fitnessstraining	Motivation für Schule (1-5)	Berufswunsch
Mia	9	W	3	Neustadt	1	Eher selten	1	Eventmanager
Tom	10	M	4	Altstadt	2	Sehr selten	3	Bäcker
Max	11	M	2	Umland	1	Sehr häufig	5	999
Pia	9	W	1	Neustadt	1	Eher häufig	2	Richter
	<i>Metrisch</i>	<i>Nominal</i>	<i>Metrisch</i>	<i>Nominal</i>	<i>(nominal / ordinal)</i>	<i>Ordinal / Metrisch</i>	<i>Ordinal / Metrisch</i>	<i>Nominal</i>



1. Lieblingsfach	2. Lieblingsfach	3. Lieblingsfach
Sport	Mathe	Deutsch
Musik	Mathe	Englisch
Deutsch	Musik	Sport
Mathe	Physik	Deutsch

oder

Sport	Mathe	Deutsch	Englisch	Physik
1	1	1		
1	1		1	
1		1		1
1	1			1

Aufgabe Dateneingabe 2:

Fügen Sie folgende zwei Fragebogenfragen auch in die Tabelle ein:

Welche Sportarten übst Du aus (Joggen, Fahrradfahren, Fußballspielen, Sonstiges und zwar _____):

- Joggen
- Radfahren
- Fußballspielen
- sonstiges und zwar: _____

4 Variablen (metrisch) und eine 5. für „Sonstiges“ angekreuzt (numerisch) und eine 6. für den Text (String); alle nominal

Sortieren Sie folgende Dinge nach Beliebtheit?

Äpfel, Birnen, Pflaumen, Bananen?

4 Variablen (metrisch) für jede Obstsorte eine Variable. Alternativ für jeden Rang eine Variable und Nummern für die Obstsorten.

Dauer: 10 Minuten **Abgabeformat:** (Excel-)Tabelle



Aufgabe Dateneingabe 3:

Fügen Sie folgende zwei Fragebogenfragen auch in die Tabelle ein:

Wissensfrage 1: Wann ist Weihnachten? Im November, im Dezember, Im Januar.

- Im November *Falsche Antwort*
- Im Dezember *Richtige Antwort*
- Im Januar *Falsche Antwort*

[1 Variable; Numerisch; Wertelabels: 0 = falsche Antwort, 1 = richtige Antwort; metrisch]

Wissensfrage 2: Wann ist Ostern?

- Im Frühling *Richtige Antwort*
- Im Herbst *Falsche Antwort*
- Im Advent *Falsche Antwort*
- Zwischen Fasching und Pfingsten *Richtige Antwort*

[Hier müssen Sie entscheiden ob 4 Variablen, oder eine zusammengefasste]

Dauer: 10 Minuten **Abgabeformat:** (Excel-)Tabelle



Aufgabe Neuberechnung Formel:

A. Erstellen Sie eine Kundendatei: Erfinden Sie vier Variablen mit der Summe in Euro an bestellten Waren für 1. bis 4. Quartal

Name	1. Quartal	2. Quartal	3. Quartal	4. Quartal	Durchschnittlich	Summe	Differenz	...
Müller	1	2	3	4	2,5	10	3	
Meier	2	3	4	5	3,5	14	3	
Schneider	1	1	2	2	1,5	6	1	
Schmidt	1	1	1	1	1	4	0	

8. Variable:

- Prozent von Quartalssumme
- Mittelwert aller Quartale
- Totale Summe plus 19 % Umsatzsteuer
- Abweichung von Quartal 4 zum Durchschnitt

Aufgabe Umcodieren:

Sie haben eine Online-Befragung mit 15 Teilnehmern durchgeführt. Sie haben Sie eine vollständig ausgefüllte Excel-Datei zum Auswertungszeitpunkt erhalten. Die Befragung wurde auf einer Skala von 1-4 (sehr motiviert, eher motiviert, eher unmotiviert, unmotiviert) durchgeführt. Leider hat die Onlinesoftware die Zahlen umcodiert. Sie möchten die ursprünglichen Zahlen wiederherstellen. Sie wollen überall dort wo originale eine 1 ist eine 4 usw. Was wäre der schnellste Schritt in Excel für die Umcodierung?

- 1 → 5
- 2 → 4
- 3 → 3
- 4 → 2
- 5 → 1

Einfache Lösung =6-X

Komplizierte Lösung mit „=Wenn“-Funktion

Alt	NEU
2	4
3	3
4	2
2	3
1	5
2	4
4	2
5	1
1	5
4	2
3	3
2	4

Fehlende Werte

Was tun mit fehlenden Werten?

- Fehlende Werte zunächst gesondert codieren z.B. mit -999 oder 9999, k.A. usw.
- Einfachste Lösung: Den entsprechenden Fall löschen. (Achtung Excel würde mit -999 rechnen)
- Fortgeschrittene Lösung: ein statistisches Verfahren das fehlende Werte anhand der restlichen Daten einfach ersetzt.. Mittelwerte
- Schwierig: Nachbefragung, eine weitere Person befragen
- Schwierig: Vergleichbares Profil nutzen
- Fragliche Lösung: Den entsprechenden Fall nur dann nicht beachten, wenn die entsprechende Variable für die Analyse genutzt wird.



Häufigkeiten

Verkaufte Produkte pro Monat

	Absolute Häufigkeit	Relative Häufigkeit (Prozente)	Kumulierte Häufigkeit
1. Quartal	23	0,23	23
2. Quartal	34	0,33	57
3. Quartal	32	0,32	89
4. Quartal	12	0,11	101
	101	1	

Leichtigkeit der Interpretation (ist auch subjektiv!)

Relative Häufigkeiten sind präziser zu interpretieren, weil sie einzelne Ergebnisse im Vergleich zur Gesamtzahl setzen.

ABER: Die meisten Menschen (ohne statistische Ausbildung) können sich absoluten Häufigkeiten besser vorstellen, als relative Häufigkeiten oder kumulierten Häufigkeiten.

Quelle: Sedlmeier, Peter, Gigerenzer, Gerd Teaching (2001): Bayesian Reasoning in Less Than Two Hours. Journal of Experimental Psychology: General Vol. 130, No. 3. 380-400

Übungen

Aufgabe Häufigkeiten Studierende:

Wir haben einen Datensatz der folgende Variablen für 15 Fälle beinhaltet:

ID Nummer des Befragten

A) Studiendauer in Semester

B) Engagement im Studium: Skala 1 (sehr engagiert) bis 5 (gar nicht engagiert)

C) Note der Abschlussarbeit

ID	B	C	E
1	5	3	3
2	5	3	2
3	7	4	1,2
4	7	5	3
5	9	2	2,5
6	5	4	1,7
7	6	1	2,1
8	7	2	1
9	8	3	3,2
10	6	1	4
11	7	4	0
12	6	3	2,4
13	8	5	1,5
14	7	5	2,7
15	6	3	3

Aufgabe Studierende

I. Erstellen Sie eine Häufigkeitstabelle mit der Variable Note (absolute Häufigkeit, relative Häufigkeit, kumulierte Häufigkeit)

- Manuell Bereiche zuordnen

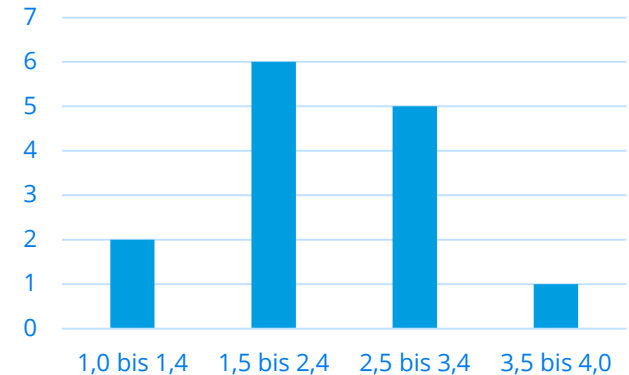
- Achtung Unterschied zwischen einfacher „kumulierter“ Häufigkeit und „kumulierte relative“ Häufigkeit

II. Erstellen Sie ein Säulendiagramm der Variable Note. → Ist eine metrische Variable → besser Histogramm

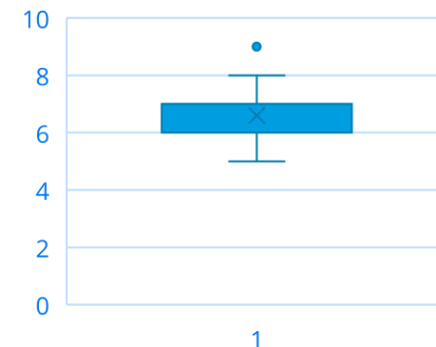
III. Erstellen Sie ein Box-Plot für die Studiendauer

IV. Erstellen Sie ein Histogramm für die Variable Engagement im Studium.

→ Klassen aufstellen, und zu jeder Klasse die Häufigkeit aufbauen.



Kastengrafik



Ziel von Statistik

„Statistiken lügen zwar nicht,
vertuschen aber sehr gern.“

„Mit Statistiken kann man alles
beweisen, nur nicht die Wahrheit“



Ihre Fragen:

Sind Statistiken () nicht dafür geeignet, die Wahrheit zu zeigen?

Wann kann man einer Statistik / einer Statistikerin trauen?

Wie vertuschen / lügen Statistiken?

Wie kann ich vertuschende Statistiken enttarnen?

Beleuchten Statistiken alle Seiten eines Problems?

Kann man alles ausschließlich mit Statistiken beweisen?

Was ist die Wahrheit?

Dient eine Statistik dem Beweis von Wahrheit?

Warum kann man die Wahrheit nicht beweisen?

Haben Statistiken einen Stellenwert für die Menschheit?

Ist lügen und vertuschen dasselbe?

Was ist „alles“?

Wie kann man kritisch mit Statistiken umgehen?

Hinweis für mögliche Prüfungsfragen

- Interpretieren Sie Ergebnisse mit mehr als nur ja / nein
- Beschreiben Sie Grafiken / Daten auch einfach, was Sie dort sehen