Einführung in die Statistik

Prof. Dr. rer. soc. Berthold Löffler

Fakultät Soziale Arbeit,
Gesundheit und Pflege

Hochschule Ravensburg Weingarten

Inhaltsverzeichnis

Begriffsbestimmungen

- soz. Erwünschtheit ≙ Verzerrfaktor: was will Gesell. von mir hören? Niemand sagt, dass AFD wählen, weil von Gesell. abgelehnt
- Kumulieren: man addiert immer auf
- 1.246.871.951 Menschen in China → Zahl ist zu genau! | Statistik arbeitet mit Trick, dass bei Sachbezügen genaue / präzise Zahl raus kommt → geht eig. gar nicht
- kreative Statistik: stützt sich nicht auf Lügen
- Suggestionen: man kann Frage so formulieren, dass Antworten von vornherein klar | mit Fragestellung kann man Antwort beeinflussen (Suggestion: wie alt warst du, als du das 1. Mal geklaut hast? → neutral: ~~wie alt~~ Hast du schon mal geklaut?)
- Trendextrapolationen ≙ Form von Prognose: Bsp. Statistiken von früheren Verdiensten → Ausrechnen wie viel % Gehalt zugenommen → Annahme: Löhne sind von Jahr zu Jahr um 5% gestiegen ⇒ Ausrechnung der Spiegelung der Vergangenheit in Zukunft | alles was in Trendextrapolation ≠ sicher ⇒ man macht es um schon mal im Voraus zu überlegen was ich in Zukunft mache ≙ Planung der Zukunft ist damit möglich

Grundbegriffe der Statistik

1. Untersuchungszweck ≙ was möchte ich machen / wissen?
   ↳ klar definierte Überschrift
2. Grundgesamtheit ≙ kann ich Personen überhaupt befragen? → Studenten in ganz D ✓ aber in Wgt ☺
   ↳ welchen Aufwand will u. kann ich betreiben
   → Stichprobe aus Grundgesamtheit, wenn nicht alle mög. sind!
3. Stichprobe → muss Realität entsprechen! (Grundgesamtheit: 20% ♀, 80% ♂ → also auch so in Stichprobe)
   ↳ repräsentative Stichprobe mit von 2 Wegen: a) Zufallstichprobe, b) bewusste Stichprobe
   a) Bsp. ges. 5000 Studenten → Liste von allen → jeder S. wird genommen
   b) Bsp. Willkürstichprobe → stehe an Eingang → frage jeden, der rein kommt
   ↳ Quotenstichprobe ≙ verkleinertes Abbild der Realität

Grundregeln

1. logisch Wiederspruchfrei
2. kein offensichtlicher Unsinn — nicht argumentieren, was für andere nicht nachvollziehbar ist
3. intersubjektiv nachprüfbar sein — ich beschreibe was, der andere macht es → folgt meiner Anweisung ≙ überprüfbar

1 EINFÜHRUNG IN DIE STATISTIK

Statistik in Krieg ist immer Propaganda

Vertraue keiner Statistik, die Du nicht selber gefälscht hast.

(Statistiker – Kalauer)

man muss misstrauisch gegenüber Statistiken sein!

Definition:

Statistik ist das methodische Vorgehen bei der Beschaffung von Daten und deren Interpretation für Informations- und/oder Entscheidungszwecke. „Sammeln u. Erheben"

1.1 Aufgabe und Erkenntniswert der Statistik

- Statistik ist ein wissenschaftliches Werkzeug zur Beschaffung und Interpretation von Daten für Informations-, Entscheidungs- oder Erkenntniszwecke

- Die statistische Urteilsbildung ist das Ergebnis induktiver Vorgehensweise

- Statistische Aussagen informieren über typische, allgemeine, quantifizierbare Eigenschaften von Gesamtheiten, Mengen, Ereignissen usw.

- Statistische Urteile gelten für die Gesamtheit, nicht jedoch zwangsläufig für jedes Element dieser Gesamtheit

- Statistische Urteile enthalten Informationen über die Verteilung spezifischer Merkmalsausprägungen in einer Gesamtheit und/oder über die Beziehungen zwischen verschiedenen Variablen.

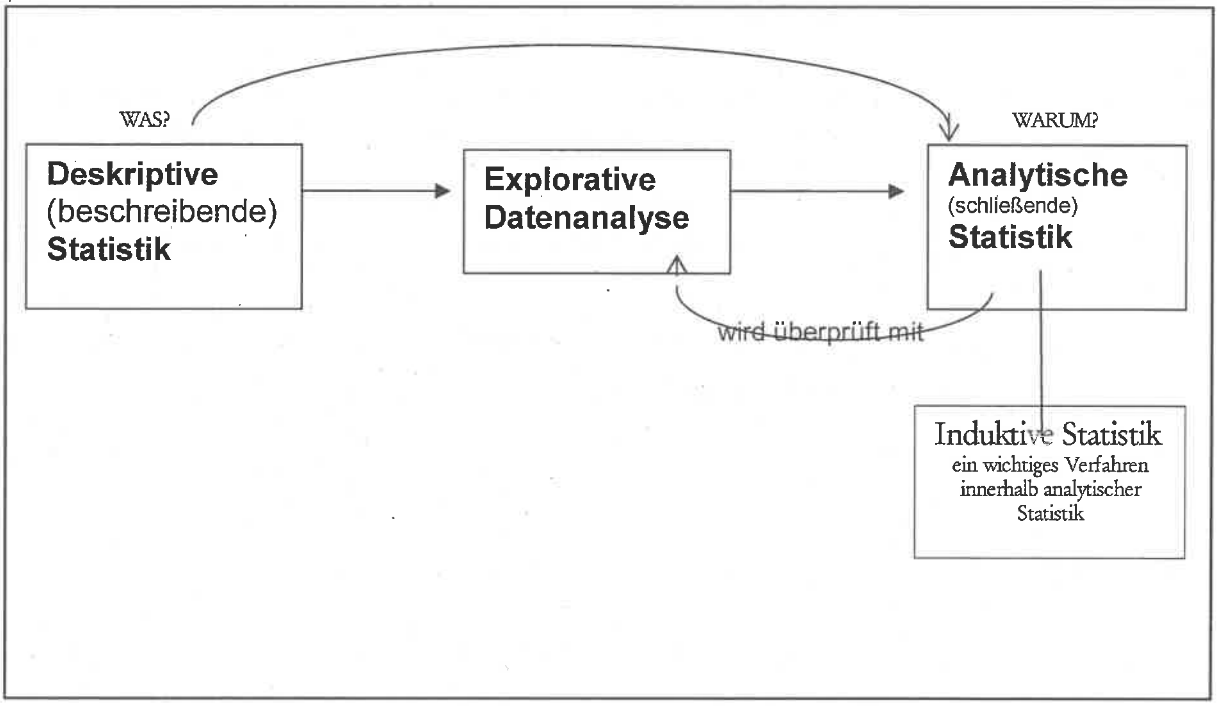
- Statistik kann keine Beweise in streng mathematischem Sinn führen, sondern nur eine rational begründbare Gewissheit (Evidenz) zugunsten bestimmter Hypothesen deutlich machen.

zunächst selbstverständliches erscheint durch EST in anderem Licht
↳ Verhalten sich Leute, die umweltbewusst sind als Leute, die es nicht sind? -> aus Stegreif JA, weil sie sich ökonomisch verhalten
=> Ergebnis: verhalten sich alle gleich!

· bei einer Statistik kann nicht pauschal das selbe rauskommen (abhängig von Land, Kultur etc.) -> DESHALB Wiederholungsstudie!
· Veränderungen (bspw. Wertewandel) kann nur bei Wiederholungsstudie herausgefunden werden

1.2 Teilbereiche der Statistik

1.2.1 Inhaltliche Unterteilung



Deskriptive Statistik:
Ausgangspunkt jeder Datenanalyse:
Beschreibung und Darstellung der Beobachtungsdaten anhand von Häufigkeitsverteilungen (z.B. Tabellen, Grafiken), statistischen Maßzahlen und Zusammenhangsmaßen.

Explorative Datenanalyse:
Suche nach Strukturen, möglichen Fragestellungen und Hypothesen. Die entstandenen Hypothesen werden im Anschluss mit Methoden der induktiven Statistik überprüft.

Induktive Statistik:
Handelt es sich beim Datensatz um eine repräsentative Stichprobe, so können mit den Methoden der induktiven Statistik Rückschlüsse auf die Grundgesamtheit getroffen werden. Diese Aussagen bergen zwar Unsicherheiten, lassen sich aber einigermaßen zuverlässig abschätzen.

(DULLER 2013:9)

Analytische Statistik:
Erklärung und Prognose möglicher Ursachen von Ereignissen durch Modelle (Hypothesenbildung) auf der Grundlage der Wahrscheinlichkeitsrechnung.

Beispiel für Aussagen der beschreibenden Statistik:

> Bei 100 Würfen mit einem Würfel fällt 14mal die 1, 17mal die 2, 17mal die 3, 18mal die 4, 19mal die 5 und 15mal die 6.
> Die mittlere Augenzahl beträgt 3,56. Der Median (die Zahl, die genau in der Mitte aller 100 Würfe liegt) ist die 4.
> Der Modus (die Zahl, die am häufigsten gewürfelt wurde) ist die 5.

Problemstellung der analytischen Statistik:

> Bei 100 Würfen trat die 1 nur 11mal, die 6 dagegen 25mal auf. Lässt sich daraus schließen, dass der Würfel gezinkt ist?

1.2.2 Unterscheidung nach Anzahl betrachteter Merkmale

Bei der Erhebung von Daten werden in der Regel mehrere Merkmale erhoben. Bei der Analyse kann jedoch jedes einzelne Merkmal für sich analysiert werden. Man spricht dann von der **univariaten Statistik**. Die Analyse von zwei Variablen wird **bivariate Statistik** genannt. **Multivariate Statistik** analysiert mehr als zwei Merkmale.

Fragebogen: womit lernt ihr? Lernhilfen?

(DULLER 2013:9)

☐ Lernvideo
☐ LernPDF
☐ Nachhilfe

MERKMAL
≙ Eigenschaft Bsp. Alter/Fam.stand
≙ angekreuzte Lernhilfe ☒ Lernvideo

MERKMALSAUSPRÄGUNG
≙ Menschen haben unendlich viele Merkmalseigenschaften, die in einem Oberbegriff zusammengefasst sind → Bsp. ledig, verheiratet
≙ Auswahlmöglichkeiten

MERKMALSTRÄGER
≙ etw. (Person/Institution), die durch unendliche Merkmalseigenschaften verfügen
≙ Zettel zum Ausfüllen

GRUNDGESAMTHEIT
≙ Menge aller Merkmalsträger / Menge der Zettel (Bsp. 100)

(n=100)
100 wurden befragt

Video	PDF	Nachhilfe	
70	20	10	Antworten in absoluten Zahlen
70/100	20/100	10/100	relative Häufigkeit
70%	20%	10%	

2 DESKRIPTIVE STATISTIK

2.1 Merkmale und statistische Meßskalen

Merkmalswerte *(x_i)* werden anhand von Beobachtung, Befragung oder Messung ermittelt. Die statistische Meßskala bildet hierfür das Instrument.

Jede der folgenden Skalen ist verbunden mit:

- einem gewissen Informationsniveau.
- einer Reihe von statistischen Verfahren, die eingesetzt werden dürfen.

(DULLER 2013:13)

Messqualität

2.2 Das Informationsniveau von Meßskalen

2.2.1 Klassifikatorische/Nominale Merkmale (Unterschiedsmerkmale)

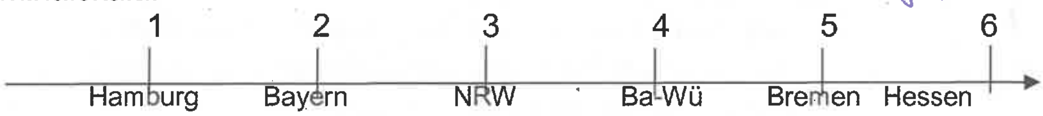
Merkmale sind nominal, wenn:

- ihre Ausprägungen nicht in eindeutiger Weise geordnet werden können.
- eine sinnvolle Interpretation von Abständen nicht möglich ist.
- sie nur aufgrund ihrer Bezeichnungen unterschieden werden können.

(DULLER 2013:13; Sibbertsen/Lehne 2012:4)

- ohne Wertung
- sind Sachen, die einfach so sind, bleibt auch so (Bsp. Anja heißt Anja)

Nominalskala:



=> Geschlecht ♂ od. ♀

=> Vorname

=> Herkunftsland

=> Haarfarbe

=> Farbe allg.

...

→ Charakter-eigenschaften

2.2.2 Komparative/Ordinale Merkmale (Rangmerkmale)

Messqualität

Merkmale sind ordinal, wenn:

- ihre Ausprägungen nur in einer relativ unbestimmten Rangbeziehung zueinander stehen

• 1.|2.|3. Platz

Bsp.: besser – schlechter, größer – kleiner

Ordinalskala:

Hotel 1 Stern	Hotel 2 Sterne	Hotel 3 Sterne	Hotel 4 Sterne

Likertskala:

stark unzufrieden	unzufrieden	weiß nicht	zufrieden	sehr zufrieden

→ man kann damit alles rechnen

2.2.3 Metrische Merkmale

Messqualität

Merkmale sind metrisch, wenn:

- ihre Ausprägungen Vielfache einer Einheit sind.
- die Ausprägungen sich voneinander unterscheiden.
- sie eine eindeutige Anordnung haben.
- sie einen eindeutig definierten Abstand haben.

• Zahlen

(BOURIER 2013:14)

metrische Skala:

0,0	0,5	1,0	1,5	2,0

innerer Zsmhang von 3 Kategorien/Merkmalen:

(1) nicht alle Merkmale sind eindeutig → Anja hat IQ von 140 – metrisch, weil messbar? / ordinal, weil mehr od. weniger? → od. beides?

(2) blond ≙ nominal $\xrightarrow{\text{kann zu ordinal werden}}$ wenn Merkmal eine Eigenschaft zugeteilt bekommt ⇒ blond ≙ blöd → je heller, desto blöder)

(3) metrisch (beste) → ordinal (Mitte) → nominal (schlechteste) ⇒ Verlieren an Infos! metrisch → alles, ordinal → vgl., nominal → Unterschiede

Bsp. Anja 1,60 m, Tanja 1,80

Anja 1,60 u. Tanja 1,80 → metrisch

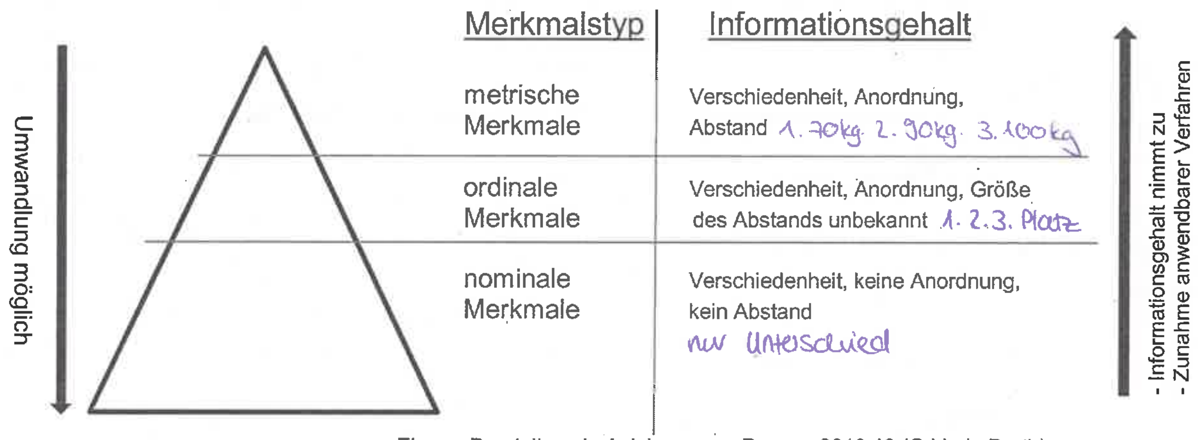
Anja kleiner als Tanja → ordinal

beide haben Körpergröße → nominal

2.2.4 Übersicht und Konsequenzen

Merkmalstyp	Charakterisierung	Messmethode	Messergebnis
klassifikatorisch	Art/Klasse	nominal	Klassen
komparativ	Intensität	ordinal	Rangordnung
metrisch	Zahlen	kardinal	Größenmäßig festgelegte Werte

Aus den verschiedenen Informationsniveaus von Merkmalen resultiert folgender hierarchischer Aufbau:



Eigene Darstellung in Anlehnung an DULLER 2013:13 (© Linda Barth)

Aus dieser Hierarchie der Merkmale ergeben sich zwei Konsequenzen:

- Jedes Merkmal aus einer höheren Hierarchiestufe kann durch Zusammenfassen und Umbenennen von Merkmalsausprägungen in ein Merkmal der niedrigeren Stufe umgewandelt werden, allerdings entsteht dadurch ein Informationsverlust.

  Beispiel: Das Merkmal Körpergröße kann in cm gemessen werden, es sind jedoch auch die Ausprägungen klein – mittel – groß möglich.

- Alle Verfahren, die für ein Merkmal aus einer bestimmten Stufe zulässig sind, sind auch zulässig für Merkmale aus darüber liegenden Stufen.

(DULLER 2013:13)

Merkmal ≙ Fam.stand (Überbegriff)
Merkmalsausprägung ≙ ledig, verheiratet, ...

2.3 Häufigkeitsverteilungen

Häufigkeitsverteilungen können in zwei Gruppen aufgeteilt werden:

Bedingungen, dass Tabelle richtig:
• Überschrift (wann, was, wo)
• Quelle

- **Eindimensionale Häufigkeitsverteilung**:
  Die statistische Untersuchung beschränkt sich auf ein Merkmal
- Zweidimensionale Häufigkeitsverteilung: 2 Merkmale
- **Mehrdimensionale Häufigkeitsverteilung**:
  Die statistische Untersuchung erstreckt sich auf mehrere Merkmale

(BOURIER 2013:38)

Verteilung des Fam.standes auf ges. Bev.

Beispiele für eindimensionale Häufigkeitsverteilungen: ZEITVERGLEICH (alle 10 Jahre)
Wohnbevölkerung der BRD nach Familienstand

WO WANN WAS => muss drin stehen!

NOMINAL
KLASSIFIKATORISCH

Wohnbevölkerung der BRD am 31.12.1986 nach Familienstand

Familienstand (x)	Häufigkeit in 1000 (absolute Häufigkeit)	Häufigkeit in Prozent (relative Häufigkeit)
ledig	24.172	39,5
verheiratet	29.401	48,1
verwitwet	5.366	8,8
geschieden	2.198	3,6
Wohnbevölkerung gesamt	61.137	100,00

Quelle: Statistisches Jahrbuch 1988 für die BRD (SJ 1988)

=> MOMENTAUFNAHMEN

39,5 % von 100% bzw. von ges. Bev. ≙ ledig -> Dreisatz!

je mehr Daten, desto besser können Unterschiede erkannt werden
ABER: Zahlen können Wirklichkeit verschleiern

Wohnbevölkerung der BRD am 31.12.1996 nach Familienstand

Familienstand (x)	Häufigkeit in 1 000	Häufigkeit in Prozent
ledig	33.429	40,7 ~
verheiratet	38.103	46,5 ↓
verwitwet	6.463	8,3
geschieden	4.018	4,9 ↑
Wohnbevölkerung gesamt	82.012	100,0

Quelle: Statistisches Jahrbuch 1998 für die BRD (SJ 1998)

d.h. -> viel mehr Frauen -> Witwen (wg. ~~Kr~~ Krieg)

Wohnbevölkerung der BRD im Jahre 2014 nach Familienstand

Familienstand (x)	Häufigkeit in 1000 (absolute Häufigkeit)	Häufigkeit in Prozent (relative Häufigkeit)
Ledig	32.926	40,8 ~
verheiratet	36.793	45,5 ↓
Verwitwet/geschieden	11.083	13,7 ↑
Wohnbevölkerung gesamt	80.802	100,0

Quelle:
https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/HaushalteMikrozensus/HaushalteFamilien2010300147004.pdf?__blob=publicationFile

2 Merkmalsausprägungen wurden zsmgefasst !! => das darf eig. nicht passieren

=> nur Momentaufnahme: vllt. ~~morgen~~ 2. Frage schon geschieden!
↳ 1. Umfrage: verheiratet, 2. Umfrage: geschieden liegen ja 10 Jahre auseinander

=> eig. Neustrukturierungen von Fam., Singlehaushalte, Alleinerziehend, Patch-work
↳ wenig Bewegung in Tabelle!

=> Kinder werden mit einberechnet -> demogr. Wandel: Kinder ↓
-> ledige ↑ (mit Kinder u. Erwachsene!)

absolute Häufigkeit: => kann sein dass sich absolute Zahlen nicht ändern
↳ Vgl. eher schwer!, weil kein exaktes ~~Zahlenwert~~ Zahlenvgl. mögl.
↳ vllt früher 2 Kandidaten, heute mehr Wähler
=> Umweltfaktoren berücksichtigen

relative Häufigkeit:
↳ Veränderungen über Jahre hinweg vgl.

Privathaushalte in der BRD am 30.04.1986 nach Personenzahl

Merkmal u. -ausprägung | absolute Häufigkeit | relative Häufigkeit

Anzahl der Personen im Haushalt	Häufigkeit in 1000 (absolute Häufigkeit)	Häufigkeit in Prozent (relative Häufigkeit)
1	9.177	34,3
2	7.886	29,5
3	4.564	17,1
4	3.516	13,1
5	1.596	6,0
insgesamt	26.739	100,0

Zeitvgl.: alle 10 Jahre

Raumvgl.: geht nur wenn rel. Häufigkeit da

Quelle: Statistisches Jahrbuch 1988

METRISCH

Privathaushalte in der BRD im April 1997 nach Personenzahl

Anzahl der Personen im Haushalt	Häufigkeit in 1 000 (absolute Häufigkeit)	Häufigkeit in Prozent (relative Häufigkeit)
1	13 259	35,4 ↑
2	12.221	32,6
3	5.725	15,3
4	4.537	12,1
5 und mehr	1.715	4,6
insgesamt	37.457	100,0

Quelle: Statistisches Jahrbuch 1998

Prozent <–> Prozentpunkte: 20 Prozentpunkte sind 40 Prozent

Privathaushalte in der BRD im Jahre 2014 nach Personenzahl

Anzahl der Personen im Haushalt	Häufigkeit in 1000 (absolute Häufigkeit)	Häufigkeit in Prozent (relative Häufigkeit)
1	16.411	40,8 ↑
2	13.837	34,4
3	4.988	12,4
4	3.660	9,1
5 und mehr	1.327	3,3
insgesamt	40.223	100,0

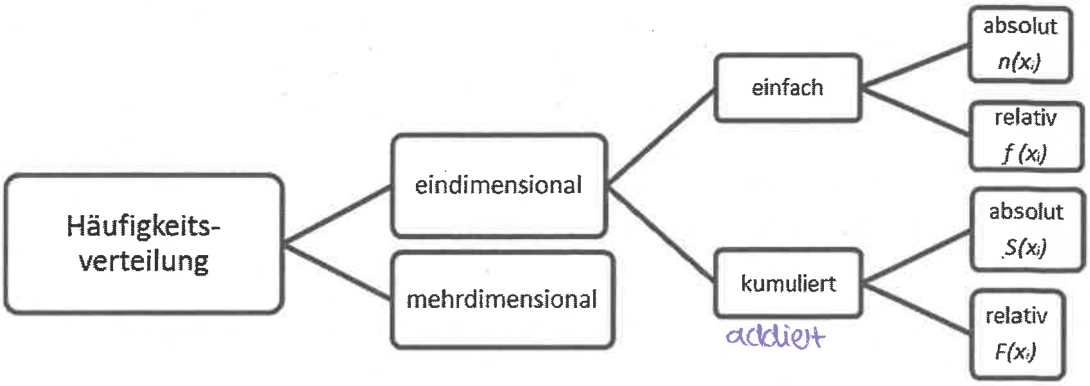
Quelle:
https://www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/Bevoelkerung/lrbev05.html

quantitative Statistik: 1. Auswertung, 2. Interpretation der Werte
↳ Zahl allein sagt nichts aus!

Fragenkatalog bei Tabellenvgl.:
• Worüber reden wir? Was verändert sich mit der Zeit?
• Warum ist das so? Ursachen?

relative Häufigkeit – Probleme:
• Artikel: tödl. Haiangriffe steigen um 100 %
↳ von 1 Person auf 2 gestiegen
• von 4 Spuren auf 6 Spuren -> steigert Spurenkapazität um 50 %
DANN zu viele Unfälle -> wieder von 6 auf 4 Spuren
=> Spurenkapazität um 33,3 gesunken => im Vgl. zu 50 %: um 17 % ist Spurenkapazität gestiegen!
=> WICHTIG: immer noch absolute Zahlen ansehen! -> führen sonst in Irre

2.4 Übersicht über Häufigkeitssummenfunktion und empirische Verteilungsfunktion



Eigene Darstellung in Anlehnung an BOURIER 2013:38 ff. (© Linda Barth)

Häufigkeitsverteilung von Familien nach Zahl der Kinder (Kinder unter 18 Jahren) in der BRD im Jahre 2015

METRISCH

Kinderzahl x_i	Familien-anzahl (in 1000) $n(x_i)$	relative Häufigkeit $f(x_i)$	Häufigkeits-summenfunktion (in 1000) $S(x_i)$	empirische Verteilungsfunktion $F(x_i)$
0	32.732	0,803	32.732	0,803
1	4.248	0,104	36.980	0,907
2	2.924	0,072	39.904	0,979
3	701	0,017	40.605	0,996
4	127	0,003	40.732	0,999
5 und mehr	43	0,001	40.775	1,000
	40.775	1,000		

Quelle:
https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/HaushalteMikrozensus/HaushalteFamilien2010300157004.pdf?__blob=publicationFile

Häufigkeitssummenfunktion

36.980 ≙ 32.732 + 4248
↳ wer hat mind. 2 Kinder
=> Zahlen werden summiert

empirische Verteilungsfunktion
≙ kumulierte
↳ 80% → kein Kind
90% → keins od. ein Kind

3 Qualitätskriterien in ESF:
(1) Objektivität -> WER untersucht? => egal, wer es macht
(2) Reliabilität -> das, was misst kann man vgl. => zum selben Sachverhalt (≙ Zuverlässigkeit)
(3) Validität -> Wert / Gültigkeit => messe ich das, was ich überhaupt wissen will?

Entstehung einer Häufigkeitsverteilung

Bsp. Raumtemperatur:
(1) hat jeder Student Thermometer?
(2) Thermometer misst immer 20 °C
(3) Thermometer hat Gültigkeit -> Temp messe
↳ mit Uhr nicht messen ≙ Gültigkeit nein

Beispiel: metrisch -> Zeit wird angeschaut

Die Unternehmensberatung “Hire & Fire” ist auf Beschluss des Kreistages nun auch im Ravensburger Landratsamt – Kreissozialamt zugange. Offiziell verkündetes Ziel ist es, die Belegschaft zu dezimieren, was heutzutage auch „Lean Clean Team Management“ genannt wird. Natürlich wissen die Kreisräte, dass im öffentlichen Dienst faktisch niemand entlassen werden kann. Aber in der Öffentlichkeit suggeriert der Einsatz von Unternehmensberatern die Möglichkeit einer Effizienzsteigerung, die es in Wirklichkeit nicht einmal in der Privatwirtschaft gibt. Die Unternehmensberater messen, wie lange die Sachbearbeiter für einen durchschnittlichen Antrag auf ALG II brauchen. Von 80 Bearbeitungsvorgängen wird die Zeit genommen (in Minuten):

sind unrealistisch (eher 800 u. mehr)

52 45 59 32 46 48 30 53 44 44 58 46 40 37 54 43 39 35 55 44
47 50 46 40 29 48 37 42 38 53 40 43 52 58 38 45 42 41 57 55
53 39 47 56 45 42 30 47 48 61 50 47 44 33 43 49 49 33 42 51
54 40 35 44 54 35 41 46 51 37 38 48 45 57 46 56 49 50 43 41

Urliste

man kann auch nur etw. messen, wenn man es vergleichen kann! (ungenaue Methode)

Bearbeitungsdauer x_i	Häufigkeit $n(x_i)$	**Bearbeitungsdauer** x_i	Häufigkeit $n(x_i)$
29	1	47	4
30	2	48	4
32	1	49	3
33	2	50	3
35	3	51	2
37	3	52	2
38	3	53	3
39	2	54	3
40	4	55	2
41	3	56	2
42	4	57	2
43	4	58	2
44	5	59	1
45	4	61	1
46	5		n = 80

-> Sortierung der Urliste: alle Messzahlen werden der Größe nach aufgelistet und dann gezählt, wie viele in welcher Gruppe

-> noch keine Häufigkeitsverteilung (zu detailliert)
↳ Leistungsklassen bilden, damit STRUKTUR

Aufbereitung von Messwerten immer so, dass Struktur vorhanden!

Klassenintervall von....bis unter Klassierte Werte	Klassenmitte x	absolute Häufigkeit $n(x_i)$	relative Häufigkeit $f(x_i)$
27,5 – 32,5	30	4	0,05
32,5 – 37,5	35	8	0,10
37,5 – 42,5	40	16	0,20
42,5 – 47,5	45	22	0,275
47,5 – 52,5	50	14	0,175
52,5 – 57,5	55	12	0,15
57,5 – 62,5	60	4	0,05
$\sum =$		80	1,00

wo gehört die 5 hin?

=> Struktur entsteht!

Zeitspannen

Klassenbreite von 5 -> sind noch 7 Werte übrig

Vorteil: STRUKTUR entsteht

wo sieht man die Struktur in den Spalten? - absolute Häufigkeit (in welchem Zeitintervall am meisten)

=> Gaußsche Normalverteilung

Körpergröße, Intelligenz ist normal verteilt (gr. Masse befindet sich in der Mitte)

Welche Struktur?

Zahlen so detailliert wie mögl. darstellen u. Struktur entstehen lassen

Bsp.: „Ich sehe den Wald vor lauter Bäumen nicht mehr."
(da stehe ich im Wald)
↳ muss einen Platz finden, wo Wald von oben PLUS einzelne Bäume sehen

Klassenbreite mache ich 10: nur 3 Abschnitte -> viel zu weit weg vom Wald -> Struktur ist weg.

Intervalle müssen immer gleich breit sein, damit Vgl. möglich!
(deshalb auch vorher und nachher weniger od. mehr)

wichtig: der 29-Wert muss drin sein, Intervallbreite vorne u. hinten gleich viel

27,5 - 32,4$\bar{9}$ -> 32,5 - 37,5
Schreibweise !

von 27,5 bis unter 32,5 (also 32,4$\bar{9}$)

3 STATISTISCHE MAẞZAHLEN

Definition:

Statistische Maßzahlen haben die Aufgabe, relativ aufwendig darstellbare Häufigkeitsverteilungen in wenigen Werten zu beschreiben.

3.1 Übersicht: Eindimensionale statistische Maßzahlen

Diese Maßzahlen lassen sich in 2 Gruppen gliedern:

Mittelwerte Mittelwerte kennzeichnen die zentrale Lage oder die zentrale Tendenz einer Verteilung. (CLAUß/FINZE 2011:27 ff.)		**Streuungswerte** Streuungswerte sind die Maßzahlen zur Bewertung der Variabilität der Messewerte, also der Breite einer Verteilung. (CLAUß/FINZE 2011:27 ff.)	
Durchschnittswerte, Zentralwerte	**Abkürzung**	**Streuungsmaße, Variabilitätsmaße**	**Abkürzung**
rechnerischer Wert Ø 1) arithmetisches Mittel (Durchschnittswert, Mittelwert) nur bei metrisch mögl.	$\bar{x}$	1) Spannweite (Variationsbreite, Variationsweite)	Sw
1 2 3 4 5 2) Medianwert (Zentralwert, Stellungsmittel, mittelster Wert) für ordinal u. metrisch	Z	2) durchschnittliche Abweichung (mittlere Abweichung)	e
3) Modalwert/Modus (Dichtemittel, häufigster Wert) für nominal	D	3) Varianz	s^2
4) Geometrisches Mittel	G	4) Standardabweichung (mittlere quadratische Abweichung) Ø Abweichung vom Ø	s
5) Harmonisches Mittel	H	5) Quartilsabstand (Hälftespielraum)	QA
Die drei gebräuchlichsten Mittelwerte sind der *Modus*, der *Median* und das *arithmetische Mittel*.		Die drei gebräuchlichsten Streuungswerte sind die *Spannweite*, die *Varianz* und die *Standardabweichung*.	

3.1.1 Der Modus

Definition:

Der Modus kennzeichnet innerhalb einer Häufigkeitsverteilung diejenige Merkmalsausprägung einer Variablen, die die größte Häufigkeit aufweist.
Haben wir eine bimodale Verteilung, d.h. weisen zwei Werte die größte Häufigkeit auf, dann besitzt die Verteilung zwei Modalwerte.

Formel:

$$D = n(x_i)\max$$
$$i = 1,2,\ldots,k$$

D	Modalwert (Dichtemittel, häufigster Wert)
$n(x_i)$	absolute Häufigkeit
k	Anzahl der verschiedenen Ausprägungen
i	Zahl der Teilnehmer

Beispiel:

Musikinteressen von Jugendlichen in der Stadt x im Jahre y

Musikart	Häufigkeit
Klassik	47
Volkstümliche Musik	302 Modus
Rockmusik	259
Country	44
Hip Hop / Rap	123
Jazz	34
Techno	111
Popmusik	80
	1.000

Quelle : fiktiv

volkstümliche Musik ≙ am häufigsten gewählt
(wenn 2 mal ~~die~~ 302, dann hat man halt 2 Modi)

Bsp. Median – Ø

Median

wie viele Kinder pro Fam?

(1) 0 0 0 0 1 1 1 2 3 => $\tilde{z}_1$ ≙ 1 Kind

(2) 0 0 0 0 1 1 1 2 11 => $\tilde{z}_2$ ≙ 1 Kind obwohl Fam mit 11 Kinder

(1) 0+0+0+0+1+1+1+2+3 = 8 : 9 = 0,9 Kinder ≙ Ø = $\bar{x}_1$

(2) 0+0+0+0+1+1+1+2+11 = 16 : 9 = 1,8 Kinder ≙ Ø = $\bar{x}_2$

$\tilde{z} = 1 \Leftrightarrow \bar{x} = 1{,}8$
ist Ausreißer
Vermutung: 1 Fam

***Anmerkung*:**

Für alle Merkmalstypen zulässig.

Vor- und Nachteile:

+	-
lässt sich ohne Rechenaufwand aus der Häufigkeitsverteilung ablesen	relative Unzuverlässigkeit
gegen Ausreißer unempfindlich	

(CLAUß/FINZE 2011:29)

3.1.2 Das arithmetische Mittel

Definition:

Das arithmetische Mittel wird im Alltag auch als Mittelwert bezeichnet. Die meisten Durchschnittswerte sind arithmetische Mittel.

(DULLER 2013:90)

3.1.2.1 Berechnung des arithmetischen Mittels aus Einzelwerten

Dieses Maß wird aus der Summe aller Merkmalsausprägungen (Messwerte) einer Variablen, geteilt durch ihre Anzahl, berechnet.
Bei ungeordneten Daten ist das arithmetische Mittel über folgende ***Formel*** definiert:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_i + \ldots + x_n}{N}$$

oder in abgekürzter Schreibweise:

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{k}(x_i)}$$

$\overline{x}$	arithmetisches Mittel (Durchschnittswert)
n	Gesamtheit der Merkmalsträger
k	Anzahl der verschiedenen Ausprägungen
x_i	Merkmalswert

3.1.2.2 Bestimmung des arithmetischen Mittels aus einer Häufigkeitsverteilung

Formel:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} x_i \cdot n(x_i)$$

Anmerkung:
Ausschließlich für metrische Merkmale zulässig.

> „(...) Wenn man einer Londoner Staatsanwältin glauben darf, ist dieses Verschwinden der Großfamilie nur zu begrüßen. Denn Großfamilien machen kriminell. Nur in wenigen Fällen von Jugendkriminalität, mit denen die Staatsanwältin befasst war, kamen die Übeltäter aus Ein-Kind-Familien. Je größer die Familien, desto krimineller. Auch hier der gleiche Fehler beim Umrechnen von Haushalten auf Personen: Kleine Haushalte machen zwar einen großen Prozentsatz der Haushalte, aber einen weit kleineren Prozentsatz der Personen aus.“

(KRÄMER 2012:61f.)

Beispiel 1:
Anzahl der zu leistenden Sozialstunden im Rahmen von Urteilen nach dem Jugendgerichtsgesetz von 40 Jugendlichen (nicht klassierte Werte)

Anzahl der Sozialstunden	Anzahl der hierzu verurteilten Jugendlichen	
15	4	60
25	12	300
35	10	350
45	6	270
55	7	385
150	1	150
	N=40	Σ = 1515

Quelle fiktiv

Beispiel 2:
Anzahl der zu leistenden Sozialstunden im Rahmen von Urteilen nach dem Jugendgerichtsgesetz von 40 Jugendlichen (klassierte Werte)

Anzahl der Sozialstunden von...bis unter	Anzahl der hierzu verurteilten Jugendlichen	
0-20	16	160
20-40	7	210
40-60	14	700
60-80	3	210
	N=40	Σ = 1280

Quelle fiktiv

Vor- und Nachteile:

+	-
	reagiert empfindlich auf extreme Messwerte → nur unter Berücksichtigung der Verteilung interpretierbar

3.1.3 Der Median

Definition:

Er ist der Wert, der eine nach der Größe geordnete Reihe von Messwerten halbiert, d.h. der Median ist der Wert, unter dem 50% und über dem 50% aller Messwerte der Verteilung liegen.

Um den Median zu ermitteln, wird die Urliste der Größe nach geordnet. Eine ungerade Anzahl n im Datensatz hat genau eine Ausprägung in der Mitte. Sie damit den Median. Handelt es sich beim Datensatz um eine gerade Anzahl n, stehen zwei Ausprägungen in der Mitte. Der Median errechnet sich dann aus dem arithmetischen Mittel dieser beiden Ausprägungen, falls es sich um ein metrisches Merkmal handelt.

(DULLER 2013:92)

3.1.3.1 Berechnung des Median aus Einzelwerten

(Grundformel)

$$\frac{n}{2}$$

n = Gesamtheit der Merkmalsträger

Anmerkung:

- Medianstelle $\neq Z$, denn Z = der Wert x_i, der der Medianstelle entspricht
- Median kann nur bei ordinal- und intervallskalierten Daten ermittelt werden

Beispiel:

Das 1. Semester besteht aus 27 Studierenden im Alter von 20 bis 36 Jahren

20 21 21 21 22 22 22 23 23 23 23 23 23 23 23 24 24 25 26 26 27 28 29 29 30 34 36

50% aller Fälle | 14. Wert | 50% aller Fälle

3.1.3.2 Medianberechnung aus Häufigkeitsverteilungen

Beispiel:

Intelligenz von Jugendlichen, gemessen in IQ (I)

was ist Median?
$\frac{n}{2} = \frac{30}{2} = 15$
wann wird 15 überschritten?
↳ 15
=> also IQ 90

IQ x_i	**Anzahl der Jugendlichen** $n(x_i)$	**kumulierte Häufigkeit** $S(x_i)$
80	9	9
90	6	15
100	6	21
110	6	27
120	3	30
	N = 30	30

Intelligenz von Jugendlichen, gemessen in IQ (II)

$\frac{n}{2} = 14,5$
wann wird 14,5 überschritten?
↳ 20 schon überschritten

IQ x_i	**Anzahl der Jugendlichen** $n(x_i)$	**kumulierte Häufigkeit** $S(x_i)$
80	9	9
90	4	13
100	7	20
110	6	26
120	3	29
	N = 29	29

14,5 → kumulierte Häufigkeit: 13 ist zu niedrig, 20 ist zu hoch
=> liegt bei 20 J. => also IQ 100

Beispiel:
Altersverteilung eines Semesters

Alter	Häufigkeit $n(x_i)$	relative Häufigkeit $f(x_i)$	Häufigkeits-summenf. $S(x_i)$	empirische Verteilungsf. $F(x_i)$
20	1	0,037	1	0,037
21	3	0,111	4	0.148
22	3	0,111	7	0,259
23	8	0,296	15	0,555
24	2	0,074	17	0,629
25	1	0,037	18	0,666
26	2	0,074	20	0,740
27	1	0,037	21	0,777
28	1	0,037	22	0,814
29	2	0,074	24	0,888
30	1	0,037	25	0,925
34	1	0,037	26	0,925
36	1	0,037	27	0,962
$\sum$	27	≈1,000	27	≈1,000

$z = \frac{n}{2} = \frac{27}{2} = 13{,}5$ $\tilde{x}_i - 1 =$

Modus = $n(x_i)$ (max)

3.1.3.3 Medianberechnung bei klassierten Werten

Liegen Werte nur in Form von Merkmalsklassen vor, muss man bei der Ermittlung des Median anders vorgehen. Hierfür ist zunächst die richtige Medianklasse zu ermitteln, anschließend wird dann der Wert des Median mit Hilfe einer linearen Interpolation geschätzt. Die nach Merkmalsklassen geordnete Tabelle unserer Variable „Alter“ hat folgende Form:

Tabelle:

Variable Alter in Altersgruppen

Altersgruppe von ... bis unter	Häufigkeit n(x_i)	relative Häufigkeit f(x_i)	Prozent-werte	empirische Verteilungsf. F(x_i)
20-22	7	0,259	25,9	0,259
23-25	11	0,407	40,7	0.666
26-28	4	0,148	14,8	0,814
29-31	3	0,111	11,1	0,925
32-34	1	0,037	3,7	0,962
35-37	1	0,037	3,7	0,999
$\sum$ =	27	≈1,000	≈100,0	≈1,000

Die Medianklasse lässt sich anhand der empirischen Verteilungsfunktion ermitteln. Der Median befindet sich in der Klasse, bei der die empirische Verteilungsfunktion einen kumulierten Wert aufweist, der zum ersten Mal größer als 0,5 ist. In unserem Beispiel ist das der Wert 0.666 bzw. die Klasse 23-bis 25 Jahre.

Formel:

wo liegt Median genau?

$$Z = \tilde{x}_{i-1} + \left(0{,}5 - F_x(\tilde{x}_{i-1})\right) \cdot \left(\frac{b_i}{f(x_i)}\right)$$

100 – 110 ≙ 100 ≙ Untergrenze

$\tilde{x}_{i-1}$	Untergrenze der Klasse, in der Z liegt
b_i	Klassenbreite ≙ Abstand
$F_x(\tilde{x}_{i-1})$	Wert für die empirische Verteilungsfunktion, die genau unterhalb der Klasse liegt, in der sich der Median befindet

$f(x_i)$ ≙ imme das $f(x_i)$, ~~das~~ in dem der Median liegt

Beispiel von S. 24:

$\frac{n}{2} \hat{=} \frac{225}{2} = 112{,}5$ → unterste Grenze $S(x_i) = 63$

$\rightarrow \tilde{x}_i - 1 \hat{=} 63 \rightarrow \tilde{x}_i - 1 = 100$

$b_i = 10$

$f(x_i) = 0{,}223$

$F_x(\tilde{x}_i - 1) = 0{,}280$

$\Rightarrow z = \tilde{x}_i - 1 + (0{,}5 - F_x(\tilde{x}_i - 1)) \cdot \left(\frac{b_i}{f(x_i)}\right)$

$= 100 + (0{,}5 - 0{,}28) \cdot \left(\frac{10}{0{,}223}\right)$

$z_{iQ} = 109{,}9$

$x_{iQ} = 110 = \varnothing$

> Normalverteilung ↳ kommt zu keinen Ausreißen

Beispiel:
Intelligenzquotientenmessung bei Studierenden Technikmanagement

	IQ von... bis unter	$n(x_i)$	Normalverteilung $f(x_i)$	$S(x_i)$	$F(x_i)$
75	70-80	3	0,013	3	0,013
85	80-90	15	0,067	18	0,080
95	90-100	45	0,200	63	0,280
105	100-110	50	0,223	113	0,503
115	110-120	57	0,253	170	0,756
125	120-130	36	0,160	206	0,916
135	130-140	12	0,053	218	0,969
145	140-150	7	0,031	225	1,000
		$\sum =$ 225 TN	1,000		

Vor- und Nachteile: Median 112,5 , Ø = ~~108~~ 109,8
sehr nah an 113

+	-
Unempfindlich gegenüber Ausreißern	kommt u.U. als Merkmalswert selbst nicht vor
einfach zu ermitteln	

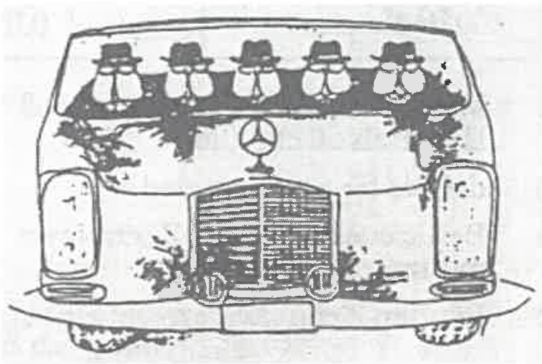
welches Intervall passt bei 112,5:
↳Klasse 100-110 passt

„Sollen wir das arithmetische Mittel als durchschnittliche Körpergröße nehmen und den Gegner erschrecken, oder wollen wir ihn einlullen und nehmen den Median?“

(KRÄMER 2012:71)

Rechenbeispiel:

Im fünfköpfigen Vorstand der X - AG sitzen Mänädscher im Alter von 48, 53, 53, 55 und 62 Jahren. Man plant eine Geschäftsreise nach Bangkok. Das älteste Vorstandsmitglied kann jedoch nicht mitreisen, weil ihm sein Arzt wegen hohen Blutdrucks eindringlich von der möglicherweise sehr anstrengenden Reise abgeraten hat. An seiner Stelle kann nun ein junger dynamischer Prokurist im Alter von 35 Jahren mitreisen. Wie ändert sich der Zentralwert und das arithmetische Mittel der Altersverteilung der Geschäftsleute?



http://www.von-der-lippe.org/dokumente/Des-auf.pdf S.13

3.1.4 Die Spannweite

Definition:

Bsp.:
jüngstes Alter: 20
höchstes Alter: 36
Spannweite:
36 - 20 = 16

Die Spannweite gibt die Länge des Bereiches an, über den sich die Merkmalswerte verteilen. Sie ergibt sich aus der Differenz des größten und des kleinsten beobachteten Merkmalswertes.

(BOURIER 2013:89)

Anmerkung:
Die Spannweite findet in der Regel nur bei metrischen Daten Anwendung. Liegen klassierte oder komparative Merkmale vor, kann die Spannweite nur näherungsweise bestimmt werden. Man erreicht dies, indem man die kleinste Klassengrenze von der größten Klassengrenze abzieht.

Formel:

$$Sw = x_n - x_1$$

Sw Spannweite
x_n größter beobachteter Merkmalswert
x_1 kleinster beobachteter Merkmalswert

Beispiel 1:

Schulgrößen in einer Großstadt (Zahl der Schüler)

Klassen von…bis unter	**Anzahl**	
200 - 400	4	300 x 4 = 1.200
400 - 600	9	500 x 9 = 4.500
600 - 800	9	700 x 9 = 6.300
800 - 1000	12	900 x 12 = 10.800
1000 - 1200	7	1100 x 7 = 7.700
1200 - 1400	2	1300 x 2 = 2.600
	$n = 43$	$\bar{x} = 770$

Quelle: fiktiv

Beispiel 2:

Im 1. Semester sind 30 Studierende im Alter zwischen 19 und 42 Jahren. Die altersmäßige Spannweite berechnet sich wie folgt:

42 – 19 Jahre = 23 Jahre

Sw = 23 Jahre

Vor- und Nachteile:

+	-
einfach zu berechnen	Ausreißer haben großen Einfluss
schneller erster Eindruck über Streuung	keine Aussage über die Streuung zwischen den beiden Extremwerten

- Bsp.: jüngster ≙ 20, ältester = 40 -> man weiß nur Extremwerte, Rest nicht
- wie sieht Verteilung aus? einheitlich ≙ homogen, untersch. ≙ heterogen
  min Streuung · max. Streuung

3.1.5 Standardabweichung und Varianz

Definition Varianz:

Die Varianz ergibt sich aus der Summe der quadrierten Abweichungen der Merkmalswerte vom arithmetischen Mittel. Diese Summe wird dann durch die Anzahl der Merkmalsträger dividiert.

Definition Standardabweichung:

Die Standardabweichung berechnet sich aus der Quadratwurzel der Varianz.

(BOURIER 2013:97)

3.1.5.1 Die Standardabweichung

a) Berechnung aus Einzelwerten

Folgende sechs Schritte führen zur Standardabweichung:

1. Berechnung von $\bar{x}$ ⌀ 1a) Spannweite
2. Berechnung der Differenzen zwischen den Merkmalswerten und $\bar{x}$ Bsp. 1.776 − (⌀ ≙) 849 = 927
3. Quadrieren der Differenzen von 2
4. Addieren der quadrierten Differenzen
5. Teilen der Summe der quadrierten Differenzen durch die Anzahl der Werte / durch n / Teilnehmerzahl ≙ Varianz
6. Wurzel aus dem unter 5. berechneten Durchschnitt ≙ Standardabweichung $s = \sqrt{\text{Varianz}}$
7. Variationskoeffizient $v = \frac{s}{\bar{x}}$

Bsp. S. 28:

1. $\bar{x} = 849$
2. $778 - 849 = -71$ (als Bsp)
3. $(-71)^2 = 5041$ (als Bsp.)
4. Addieren von allen quadrierten Sachen ≙ 1.109.408
5. 1.109.408 : n = 1.109.408 : 7 Pers. = 158.486,9 (≙ Varianz)
6. $s = \sqrt{158.486{,}9} = 398{,}1$

s ist noch nicht mal Hälfte von $\bar{x}$ (398,1 > 424,5)

=> alle s, die unterhalb ½ ⌀ ≙ Homogenität
alle s, die oberhalb ½ ⌀ ≙ Heterogenität

=> Variationskoeffizient $v = \frac{\text{Standardabweichung } s}{\text{arithm. Mittel } \bar{x}} = \frac{398{,}1}{849} = 0{,}47$

wie ist 0,5 Trennlinie zw. Homogenität u. Heterogenität!
$v < 0{,}5$ = Homogenität
$v > 0{,}5$ = Heterogenität

Formel:

$$s = \sqrt{\left(\frac{1}{n} \cdot \sum_{i=1}^{k} (x_i - \overline{x})^2\right)}$$

s Standardabweichung
n Gesamtheit der Merkmalsträger
x_i Merkmalswert
k Anzahl der verschiedenen Ausprägungen
$\overline{x}$ arithmetisches Mittel

Anmerkung:
Nur für metrische Merkmale.

Beispiel:

wenn jeder gleich viel verdienen, dann Abweichung = 0

Bruttoanfangsgehälter bei verschiedenen Trägern Sozialer Arbeit (Halbtagsstelle)

Streuung wird kleiner wenn Ludwig u. Simone raus! → geringeres Streuungsmaß

	in EURO x_i	hohe Streuung $x_i - \overline{x}$	$(x_i - \overline{x})^2$ → durch quadrieren gl. Zahlen
Ernst	778	- 71	5.041
Pauline	933	84	7.056
Otto	604	- 245	60.025
Karin	629	- 220	48.400
Ludwig	520	- 329	108.241
Friedrich	703	- 146	21.316
Simone	1.776	$\xrightarrow{-849}$ 927	859.329
7 Personen	5.943		1.109.408 → Zwischenergebnis
	Ø $\overline{x} = 849$ = th. Gemeinsamkeit / Idealzustand: alle verdienen gleich		

Quelle: fiktiv

① Spannweite = 1.776 − 520
= 1.256
→ hohe Spannweite ☹

② arithmetisches Mittel ≙ 849

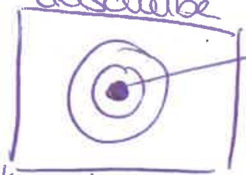
⇒ Feststellung: wie groß ist Streuung?
Bezugspunkt ≙ arithmetisches Mittel
↳ man muss wissen, wovon etw. streut!
↳ um Streuung zu ermitteln

⇒ wie sehr weichen die einzelnen u. alle vom arithmetischen Mittel ab? (Differenz Ø u. Einkommen)

⇒ arithmetisches Mittel: orientiert sich an Realität (nicht sagen „was sollte man in diesem Job Ø verdienen?")

Summe der quadr. Abweichungen vom Ø durch 7 teilen, DANN bekommt man Varianz
1.109.408 : 7 = 158.486,9 €
↳ Maß der Streuung
⇒ Varianz = unpraktisch weil so weit draußen, schlecht bewertet!
⇒ DESHALB $\sqrt{158.486,9}$ = 398,10 €
durchschn. Abweichung vom Ø ≙ S

Zielscheibe



hier hinzielen, weiß man!
→ alle 5 Pfeile ins Mitte
≙ Bezugspunkt
↳ keine Streuung
= totale Homogenität

S_2

S_1

→ je größer Strecke, desto größer Streuung

b) Berechnung aus einer Häufigkeitsverteilung

Formel:

$$s = \sqrt{\left(\frac{1}{n} \cdot \sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n(x_i)\right)}$$

Beispiel:

Tägliche Kosten von Fremdunterbringungen bei verschiedenen Trägern im Rahmen der HzE

Kosten von...bis unter...EURO	Anzahl Inanspruchnahmen HzE $n(x_i)$	Klassenmitte	$x_i \cdot n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
100-200	3	150	450	-130	16 900	50 700
200-300	9	250	2 250	-30	900	8 100
300-400	7	350	2 450	70	4 900	34 300
400-500	1	450	450	170	28 900	28 900
N =	20		Σ= 5 600			
			$\bar{x}$ = 280			Σ = 122 000

Aussagewert der Standardabweichung:

- Die Standardabweichung s ist ein Maß der Streuung von Abweichungen um $\bar{x}$
- s = 0 heißt, alle Merkmale haben dieselbe Merkmalsausprägung und liegen damit auf der Geraden von $\bar{x}$.
- Je größer s, desto größer die Streuung um $\bar{x}$; je geringer s, desto geringer die Streuung um $\bar{x}$.
- Das Verhältnis des Wertes von s zu $\bar{x}$ gibt einen Anhaltspunkt für die Streuung .Diesen Anhaltspunkt liefert der sog. Variationskoeffizient.

  Der Variationskoeffizient wird nach folgender Formel berechnet:

Formel:

$$\boxed{v = \frac{s}{\bar{x}}}$$

v Variationskoeffizient
s Standardabweichung
$\bar{x}$ arithmetisches Mittel

Nimmt v Werte über 0,5 an, so kann man sagen, dass die untersuchte statistische Masse inhomogen ist (Faustregel).

3.1.5.2 Die Varianz s^2

Definition:

Die Varianz s^2 ist ein weiteres und ebenfalls sehr häufig gebrauchtes, aber für den Ungeübten recht unhandliches Streuungsmaß. Die Varianz erhält man dadurch, dass man vorgeht wie bei s, am Ende wird jedoch darauf verzichtet, die Wurzel zu ziehen.

Vor- und Nachteile:

+	-
	Rechenvorgang inhaltlich nicht nachvollziehbar, nicht interpretierbar (Bourier 2013: 99)
	Informationsgehalt gering
Größere Abweichungen werden durch die Quadrierung stärker berücksichtigt als Kleinere. Ob das ein Vorteil oder Nachteil ist, hängt vom jeweiligen Untersuchungsgegenstand ab.	

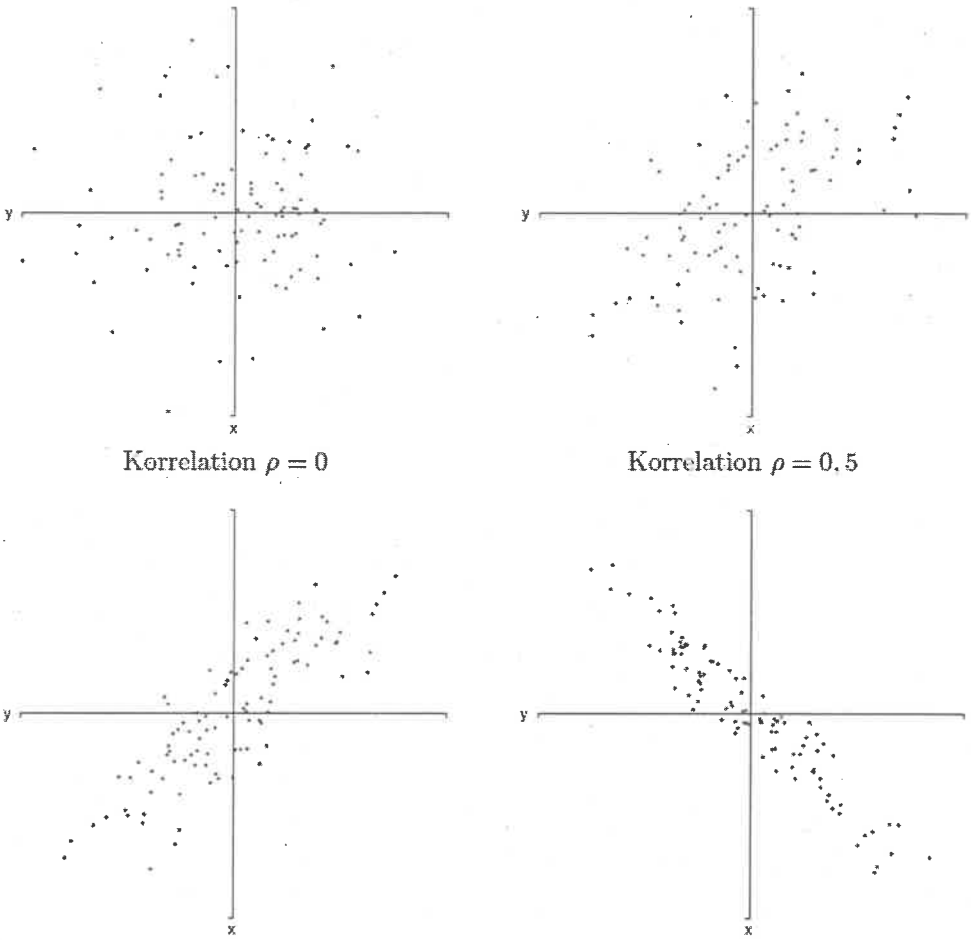
Berechnungsoptionen Merkmalstyp/Maßzahl:

	Modus	Median	arithm. Mittel
metrisch	✓	✓	✓
komparativ	✓	✓	✗
klassifikatorisch	✓	✗	✗

3.2 Zweidimensionale statistische Maßzahlen

Bei bivariablen (zweidimensionalen) Häufigkeitsverteilungen sind sogenannte Beobachtungspaare Gegenstand der statistischen Untersuchung. Wie das Wort bereits sagt, werden beide Variablen gemeinsam erhoben und betrachtet. Es geht letztlich darum, den Zusammenhang zwischen beiden herauszufinden.

(CLAUß/FINZE/PARTSCH 2013:54)



Korrelation $\rho = 0$ Korrelation $\rho = 0,5$

Korrelation $\rho = 0,85$ Korrelation $\rho = -0,95$

(DULLER 2013:131)

3.2.2 Zweidimensionale Häufigkeitsverteilungen (bivariate Statistik): Kontingenz-/ Kreuztabelle

Beispiel:
Für 50 Frauen und Männer werden die Schuhgröße und das Monatseinkommen aufgelistet. Jedem Merkmalsträger werden zwei Merkmale zugeordnet.

Person (Geschlecht)	Schuhgröße (x)	Einkommen in € (y)	Person (Geschlecht)	Schuhgröße (x)	Einkommen in € (y)
1 (m)	46	10.000	26 (m)	41	2.000
2 (m)	46	10.000	27 (w)	41	1.000
3 (m)	46	3.000	28 (m)	40	5.000
4 (m)	45	10.000	29 (w)	40	3.000
5 (m)	45	10.000	30 (w)	40	3.000
6 (m)	45	5.000	31 (w)	40	2.000
7 (m)	44	10.000	32 (w)	40	2.000
8 (m)	44	5.000	33 (w)	40	1.000
9 (m)	44	5.000	34 (w)	39	5.000
10 (w)	44	3.000	35 (m)	39	3.000
11 (m)	43	10.000	36 (w)	39	2.000
12 (m)	43	5.000	37 (w)	39	1.000
13 (m)	43	5.000	38 (w)	39	1.000
14 (m)	43	3.000	39 (w)	39	1.000
15 (m)	43	2.000	40 (w)	39	1.000
16 (m)	42	5.000	41 (w)	39	1.000
17 (m)	42	5.000	42 (w)	38	3.000
18 (w)	42	3.000	43 (w)	38	2.000
19 (m)	42	3.000	44 (w)	38	1.000
20 (m)	42	1.000	45 (w)	38	1.000
21 (m)	41	5.000	46 (w)	37	5.000
22 (m)	41	3.000	47 (w)	37	2.000
23 (m)	41	3.000	48 (w)	37	2.000
24 (w)	41	2.000	49 (w)	37	1.000
25 (w)	41	2.000	50 (w)	37	1.000

3.2.3 Absolute und relative Häufigkeit (integrierte Kreuztabelle)

Die gebräuchlichste Darstellungsform von bivariaten Datenberechnungen ist die Kontingenz-/ oder Kreuztabelle. Sie lässt bereits erste Rückschlüsse über den Zusammenhang von zwei Variablen zu.

Einkommen (y) / Schuhgröße (x)	mtl. Einkommen in €					
	1.000	**2.000**	**3.000**	**5.000**	**10.000**	
46	0 (0,00)	0 (0,00)	1 (0,02)	0 (0,00)	2 (0,04)	3
45	0 (0,00)	0 (0,00)	0 (0,00)	1 (0,02)	2 (0,04)	3
44	0 (0,00)	0 (0,00)	1 (0,02)	2 (0,04)	1 (0,02)	4
43	0 (0,00)	1 (0,02)	1 (0,02)	2 (0,04)	1 (0,02)	5
42	1 (0,02)	0 (0,00)	2 (0,04)	2 (0,04)	0 (0,00)	5
41	1 (0,02)	3 (0,06)	2 (0,04)	1 (0,02)	0 (0,00)	7
40	1 (0,02)	2 (0,04)	2 (0,04)	1 (0,02)	0 (0,00)	6
39	5 (0,10)	1 (0,02)	1 (0,02)	1 (0,02)	0 (0,00)	8
38	2 (0,04)	1 (0,02)	1 (0,02)	0 (0,00)	0 (0,00)	4
37	2 (0,04)	2 (0,04)	0 (0,00)	1 (0,02)	0 (0,00)	5
	12	10	11	11	6	50

3.2.4 Übersicht statistischer Maßzahlen für zweidimensionale Häufigkeitsverteilungen

Klassifikatorische Merkmale	Assoziationsmaße	z.B. die statistische Unabhängigkeit SU bzw. der Kontingenzkoeffizient C
Komparative Merkmale	Kontingenzmaße	z.B. der Rangkorrelationskoeffizient R von Krueger-Spearman
Metrische Merkmale	Korrelationsmaße	z.B. der Maßkorrelationskoeffizient r von Bravais-Pearson

3.2.5 Mögliche Kombinationen: Merkmal/Maßzahl

Merkmal y / **Merkmal x**	**Nominalskala**	**Ordinalskala**	**Metrische Skala**
Nominalskala	SU/C	SU/C	SU/C
Ordinalskala	SU/C	R	R
Metrische Skala	SU/C	R	r

Treffen unterschiedliche Messniveaus aufeinander, wird die Maßzahl für das jeweils niedrigere Merkmal herangezogen.

3.2.6 Interpretation der Maßzahlen

Mit der Berechnung von Korrelationskoeffizienten kann zunächst nur ein rein mathematischer Zusammenhang aufgezeigt werden. Ob tatsächlich ein kausaler Zusammenhang besteht, ist damit nicht unbedingt gesagt. Häufig handelt es sich um Scheinkorrelationen.

Oft liegt eine indirekte Abhängigkeit vor, weil zwei Merkmale kausal mit einem dritten Merkmal (intervenierende Variable) zusammenhängen.

Beispiel:

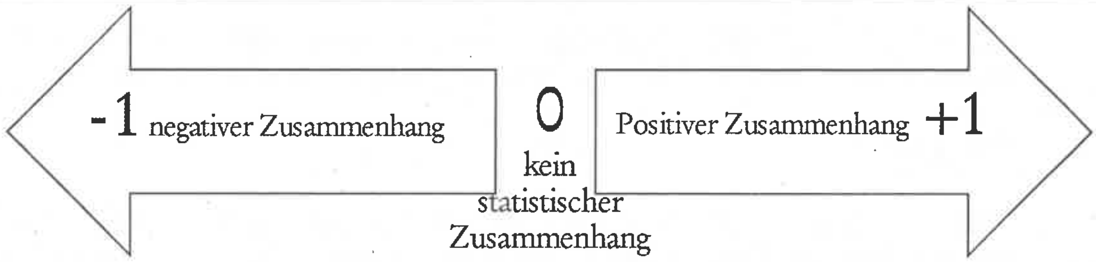
Wenn festgestellt wird, dass Männer im Straßenverkehr mehr Unfälle verursachen als Frauen, dann muss das nicht darauf zurückzuführen sein, dass Männer schlechter Auto fahren. Es kann einfach daran liegen, dass die Männer im Untersuchungszeitraum mehr Kilometer gefahren sind als die Frauen und daher ein erhöhtes Unfallrisiko hatten. Das Merkmal „Geschlecht" ist dann nur indirekt über das Merkmal „Kilometerleistung" mit dem Merkmal „Unfallhäufigkeit verbunden.

> „Solche übersehenen Hintergrundvariablen produzieren Nonsenskorrelationen zuhauf. Angefangen bei den Klapperstörchen, deren Zahl hoch positiv mit den bundesdeutschen Geburten korreliert, über die Zahl der unverheirateten Tanten eines Menschen und den Kalziumgehalt seines Skeletts (negative Korrelation), Heuschnupfen und Weizenpreis (negative Korrelation), Schuhgröße und Lesbarkeit der Handschrift (positive Korrelation), Schulbildung und Einkommen (positive Korrelation) bis zu Ausländeranteil und Kriminalität (positive Korrelation) spannt sich ein weiter Bogen eines falsch verstandenen bzw. absichtlich missbrauchten Korrelationsbegriffs."
>
> (KRÄMER 2012:172)

Definition:

Korrelationskoeffizienten messen die Stärke des Zusammenhangs zweier Merkmale. Er liegt im Wertebereich ($-1 \leq 0 \leq +1$),

d.h.
bei-1: hoch negativer Zusammenhang, hohes x gepaart mit niedrigem y
bei +1: hoch positiver Zusammenhang, hohes x gepaart mit hohem y
bei 0: beide Variablen in keinem statistischen Zusammenhang



eigene Darstellung (© Linda Barth)

Relevant für die Interpretation der Korrelationskoeffizienten sind das Vorzeichen und der Betrag. Aus dem Vorzeichen geht die Richtung des Zusammenhangs hervor. Der Betrag ermöglicht eine Aussage bezüglich der Stärke des Zusammenhangs.

(vgl. DULLER 2013:124)

3.2.7 Statistische Unabhängigkeit (SU für klassifikatorische Merkmale)

Frage:

Besteht ein Zusammenhang zwischen gemeinsam untersuchten Merkmalen?

Unabhängigkeit besteht immer dann, wenn die Verteilung auf die einzelnen Merkmalsausprägungen der Zeilen bzw. Spalten den jeweiligen Randverteilungen entspricht.

Beispiel:

Zusammenhang der Merkmale „Schultyp des Kindes“ und „soziale Stellung der Eltern“

Soziale S. der Eltern y / Schultyp des Kindes x	Arbeiter	Angestellter	Beamter	Selbständiger	
Hauptschule	12	4	2	2	20
Realschule	8	4	4	4	20
Gymnasium	0	2	4	4	10
	20	10	10	10	50

Beispiel:

Zusammenhang der Merkmale „Schultyp des Kindes“ und „soziale Stellung der Eltern“

Soziale S. der Eltern y / Schultyp des Kindes x	Arbeiter	Angestellter	Beamter	Selbständiger	
Hauptschule	$\frac{20 \cdot 20}{50} = 8$	$\frac{20 \cdot 10}{50} = 4$	$\frac{20 \cdot 10}{50} = 4$	$\frac{20 \cdot 10}{50} = 4$	20
Realschule	$\frac{20 \cdot 20}{50} = 8$	$\frac{20 \cdot 10}{50} = 4$	$\frac{20 \cdot 10}{50} = 4$	$\frac{20 \cdot 10}{50} = 4$	20
Gymnasium	$\frac{20 \cdot 10}{50} = 4$	$\frac{10 \cdot 10}{50} = 2$	$\frac{10 \cdot 10}{50} = 2$	$\frac{10 \cdot 10}{50} = 2$	10
	20	10	10	10	50

Die SU ist ein unpräzises Instrument zur Feststellung eines Zusammenhangs. Eine brauchbare Messung liefert der sog. Kontingenzkoeffizient C.

3.2.8 Der Kontingenzkoeffizient

Definition:

Kontingenzkoeffizienten beschreiben die Stärke des Zusammenhangs zwischen zwei Merkmalen, von denen mindestens eines nominalskaliert ist.

(BOURIER 2013:223)

Besteht keine Abhängigkeit, nimmt der Kontingenzkoeffizient C den Wert 0 an. Mit zunehmender Abhängigkeit wird der Kontingenzkoeffizient C größer. Bei vollständiger Abhängigkeit erreicht C den maximal möglichen Wert C_{max}.

(BOURIER 2013:226)

Formel:

$$C = \sqrt{\left(\frac{\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{n_{ij}^2}{n_i \cdot n_j} - 1}{\min.\{(k-1),(m-1)\}} \right)}$$

k ≙ Merkmal y -> Bsp.: 4 -> Arbeiter, Angest., Beamter, Selbständig

m ≙ Merkmal x -> Bsp.: 3 -> HS, RS, Gym

Formalisierte Tabelle:

Soziale S. der Eltern y / Schultyp des Kindes x	Arbeiter y_1	Angestellter y_2	Beamter y_3	Selbständiger y_4	Zeilen-summe
HS x_1	12 n_{11}	4 n_{12}	2 n_{13}	2 n_{14}	20 n_1
RS x_2	8 n_{21}	4 n_{22}	4 n_{23}	4 n_{24}	20 n_2
Gym x_3	0 n_{31}	2 n_{32}	4 n_{33}	4 n_{34}	10 n_3
Spaltensumme	n_1 20	n_2 10	n_3 10	n_4 10	n 50

Zu berechnen ist also:

$$\frac{n_{11}^2}{n_1 \cdot n_1}+\frac{n_{12}^2}{n_2 \cdot n_1}+\frac{n_{13}^2}{n_3 \cdot n_1}+\frac{n_{14}^2}{n_4 \cdot n_1}+\frac{n_{21}^2}{n_1 \cdot n_2}+\frac{n_{22}^2}{n_2 \cdot n_2}+\frac{n_{23}^2}{n_3 \cdot n_2}+\frac{n_{24}^2}{n_4 \cdot n_2}+$$

$$\frac{n_{31}^2}{n_1 \cdot n_3}+\frac{n_{32}^2}{n_2 \cdot n_3}+\frac{n_{33}^2}{n_3 \cdot n_3}+\frac{n_{34}^2}{n_3 \cdot n_4}-1$$

Setzt man die entsprechenden Zahlen aus der Tabelle ein, so ergibt sich: $\frac{(12)^2 = (\text{Anzahl Arbeite in HS})^2}{\text{Zeilen} \cdot \text{Spaltensumme}}$

$$\frac{12^2}{20 \cdot 20}+\frac{4^2}{10 \cdot 20}+\frac{2^2}{10 \cdot 20}+\frac{2^2}{10 \cdot 20}+\frac{8^2}{20 \cdot 20}+\frac{4^2}{10 \cdot 20}+\frac{4^2}{10 \cdot 20}+\frac{4^2}{10 \cdot 20}+$$

$$\frac{0^2}{20 \cdot 10}+\frac{2^2}{10 \cdot 10}+\frac{4^2}{10 \cdot 10}+\frac{4^2}{10 \cdot 10}-1$$

=> mit allen Feldern machen und dann $\boxed{-1}$ => −1 nicht vergessen!

= 0,36 + 0,08 + 0,02 + 0,02 + 0,16 + 0,08 + 0,08 + 0,08 + 0 + 0,04 + 0,16 + 0,16-1
= 1,24 - 1
= 0,24

Bei Kontingenzkoeffizient:
↳ Formel Bsp. S. 40
$\sqrt{\frac{0,24}{\min.((4-1),(3-1))}}$
=> wieso dann $\frac{0,24}{2}$?
weil 2<3?

Der Wert von C beträgt folglich:

$$C=\sqrt{\frac{0,24}{\min.\{(4-1),(3-1)\}}}=\sqrt{\frac{0,24}{2}}=0,35$$

C = 0,35

Interpretation: Zsmhang nicht groß (nicht erkennbar!) ≙ schwach mittlerer Zsmhang

3.2.9 Rangkorrelationskoeffizient R von Krueger – Spearman (für komparative Merkmale)

Formel:

für Rangmerkmal
PLUS metrisch dann aber umcodieren! Bsp. s. S. 43

$$R = 1-\frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2-1\right)}$$

d_i Differenz des Rangplatzpaares $(x_i - y_i)$
n Anzahl der Rangplätze

Zur Berechnung:
Gegeben sind zwei Rangfolgen X und Y. Gefragt ist nach dem Grad des Zusammenhangs zwischen ihnen.

Beispiel 1:

Im Rahmen eines Hochschulprojektes bewerten die beiden Studentinnen Anja und Tanja die Kitas der Gemeinde. Ihre Aufgabe besteht darin, den Gesamteindruck der Einrichtung zu bewerten und Rängen zuzuordnen. Dabei sind die beiden Mädels zu folgenden Bewertungen gekommen:

Kindertagesstätte	**Anja X Rang oder Platz**	**Tanja Y Rang oder Platz**
St. Johanna	2	1
Karl Marx KiTa	5	5
Flohkiste	6	7
El Alamein	1	3
Käpt'n Seebär	8	8
KiTa im Argonnerwald	3	2
Wasserfrösche	4	4
Zipfelmützen	7	6

Anja und Tanja konnten das Ergebnis zur Qualität natürlich nicht metrisch messen, sondern konnten nur Angaben mit einem ordinalen Merkmal machen: „besser als…“, „schlechter als…“. R misst nun, wie groß die Übereinstimmung der Ergebnisse der beiden Mädels ist.

Beispiel 2:
Berechnung von R für Rangfolge aus der KiTa Bewertung

KiTa	**Anja X Rang oder Platz**	**Tanja Y Rang oder Platz**	**Rang-differenz (d_i)** $(X-Y)$	d_i^2 $(X-Y)^2$
St. Johanna	2	1	1	1
Karl Marx KiTa	5	5	0	0
Flohkiste	6	7	-1	1
El Alamein	1	3	-2	4
Käpt'n Seebär	8	8	0	0
KiTa im Argonnerwald	3	2	1	1
Wasserfrösche	4	4	0	0
Zipfelmützen	7	6	1	1
n=8				$\sum d_i^2 = 8$

$$R = 1 - \frac{6 \cdot (\text{Summe } d_i^2)}{n \cdot (n^2-1)} = 1 - \frac{6 \cdot 8}{8 \cdot (8^2-1)} = 1 - \frac{48}{8 \cdot 63} = 1 - \frac{48}{504}$$

$$= 1 - \frac{48}{504}$$

$= 0{,}905 \Rightarrow$ hoher Zsmhang zw. beiden $\nearrow_0$

$$R = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$R = 1 - \frac{6 \times 8}{8 \times 63}$$

$$R = 0{,}905$$

R= 0,905: Großer Zusammenhang zwischen der Einschätzung von Anja und Tanja besteht. Was bedeutet das Ergebnis? Welche Hypothesen?

Beispiel 3:
Berechnung von R für den Zusammenhang zwischen Leistungseinschätzungen in einer berufsbildenden Werkstätte

Ausbilder / **Auszubildende**	**Meister/ Rang oder Platz**	**Sozialarbeiter/ Rang oder Platz**	**Rangdifferenz (d_i)**	d_i^2
Hans	1	3	-2	4
Georg	2	4	-2	4
Susi	3	1	2	4
Maike	4	2	2	4
Jenny	5	9	-4	16
Maik	6	5	1	1
Peter	7	7	0	0
Lisa	8	6	2	4
Harald	9	8	1	1
n=9				$\sum d_i^2$ =38

$$R = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$R = 1 - \frac{6 \times 38}{9 \times 80}$$

$$R = 0{,}683$$

R= 0,683 drückt aus, dass es einen mittleren Zusammenhang gibt zwischen der Rangfolge der Einschätzung des Meisters und des Sozialarbeiters. Was bedeutet das Ergebnis? Welche Hypothesen sind möglich?

nahe 1 → hohe pos. Zsmhang
nahe Mitte → mittlere pos. Zsmhang
nahe 0 → niedrige pos. Zsmhang

nahe -1 → hohe neg. Zsmhang
nahe Mitte → mittlere neg. Zsmhang
nahe 0 → niedrige neg. Zsmhang

Die Verwandlung von metrischen Messwerten in Rangplätze
Beispiel: Vorlesungsbesuch und Studienerfolg

0x gefehlt: 1. Platz
Udo u. Fritz haben beide 1x gefehlt
↳ eig. 2. u. 3. Platz
⇒ gemeinsam auf 5 ⇒ 5 : 2 = 2,5
Otto: Platz 3
usw.

Vorname	x (Studienerfolg)	y (nicht besuchte Vorlesungen)	
	komparatives Merkmal	metrisches Merkmal → komparatives Merkmal	
Max	5	0	1
Udo	2	1	2,5
Fritz	6	1	2,5
Otto	4	5	4
Karl	1	12	5
Igor	3	14	6

Umformung

Der Rangplatzunterschied zwischen Otto und Karl erscheint ebenso groß wie derjenige zwischen Karl und Igor, obwohl die Zahlen x_i ausweisen, dass sich die Häufigkeit des Fehlens bei Otto (5mal) und Karl (12mal) viel stärker unterscheidet als zwischen Karl (12mal) und Igor (14mal). Errechnen Sie R. Was bedeutet das Ergebnis?

R = −0,53 ⇒ mittlerer neg. Zsmhang
Interpretation: umso öfter fehlen, umso besser

3.2.10 Maßkorrelationskoeffizient r von Bravais – Pearson (metrische Merkmale)

nur für metrisch

Definition:

Der Maßkorrelationskoeffizient nach Bravais-Pearson wird mit r abgekürzt. Meist ist nur vom Korrelationskoeffizienten nach Pearson die Rede. Wie der Name es ausdrückt, kennzeichnet diese Maßzahl die Stärke des linearen (statistischen) Zusammenhangs zwischen den einzelnen Werten von zwei Variablen. Man spricht auch vom Zusammenhang von zwei Stichproben metrischer oder intervallskalierter Werte.

Im Unterschied zum Korrelationskoeffizienten von Bravais-Pearson misst der Rangkorrelationskoeffizient von Krueger-Spearman den Zusammenhang zwischen den Merkmalen X und Y indirekt, da der Zusammenhang zwischen den Rangziffern gemessen wird. Der Rangkorrelationskoeffizient ermittelt, wie stark die Tendenz ausgeprägt ist, dass mit einem höheren Rangplatz für Merkmal X ein höherer (oder niedrigerer) Rangplatz für Merkmal Y verbunden ist.

(BOURIER 2013:221)

Beispiel:
Bei 10 Sozialarbeitern wurde das Alter (x) und das Jahreseinkommen (y) ermittelt. Die Werte sind in der nachfolgenden Tabelle dargestellt.

Code Nr.	Alter des Beschäftigten (x)	Jahreseinkommen (y) in 1000 Euro
1	22	19
2	25	22
3	26	21
4	26	23
5	27	23
6	28	24
7	30	29
8	30	27
9	35	33
10	41	29

Es stellt sich nun die Frage, ob es einen Zusammenhang zwischen dem Alter und der Einkommenshöhe bei Sozialarbeitern gibt.

Code Nr	Alter x	Einkommen y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})\cdot(y-\bar{y})$

immer Summe berechnen

(1) arithm. Mittel von $x = \bar{x}$ u. $y = \bar{y}$ berechnen u. Tabelle füllen!

Formel:

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

$$= \frac{17{,}9}{\sqrt{27}\cdot\sqrt{17}} = 0{,}84$$

hochpositiver Zsmhang

Anmerkung: s_{xy} wird als Kovarianz bezeichnet. Sie errechnet sich so:

$\boxed{\text{Zähler / Nenner}}$

Summe Spalte $(x-\bar{x})\cdot(y-\bar{y})$

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l}(x_i-\bar{x})(y_j-\bar{y}) = \frac{1}{10}\cdot 179 = 17{,}9$$

ZÄHLER der Formel

s_x und s_y ist die jeweilige Standardabweichung:

$$s_x = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{k}(x_i-\bar{x})^2\right)}\;;\quad s_y = \sqrt{\left(\frac{1}{n}\sum_{j=1}^{l}(y_j-\bar{y})^2\right)}$$

$$= \sqrt{\frac{1}{10}\cdot 270} \qquad \sqrt{\frac{1}{10}\cdot 17}$$

$= \sqrt{27}$ NENNER der Formel

$= \sqrt{17}$ NENNER der Formel

Beispiel 1:

Stärke des Zusammenhangs zwischen dem Alter von Sozialarbeitern und ihrem Jahreseinkommen

NR	Alter (x)	Jahres-EK (y) in 1000€	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})\times(y-\bar{y})$
1	22	19	-7	49	-6	36	42
2	25	22	-4	16	-3	9	12
3	26	21	-3	9	-4	16	12
4	26	23	-3	9	-2	4	6
5	27	23	-2	4	-2	4	4
6	28	24	-1	1	-1	1	1
7	30	29	1	1	4	16	4
8	30	27	1	1	2	4	2
9	35	33	6	36	8	64	48
10	41	29	12	144	4	16	48
$\sum$	290	250		270		170	179
Arithm. Mittel	29	25					

Die **Rechnung** lautet:

$$s_{xy} = \frac{1}{10} \cdot 179 = 17{,}9$$

(Zähler / Formel)

$$s_x = \sqrt{\left(\frac{1}{10} \cdot 270\right)} \rightarrow s_x = \sqrt{27} \rightarrow s_x = 5{,}2$$ (Nenner / Formel)

$$s_y = \sqrt{\left(\frac{1}{10} \cdot 170\right)} \rightarrow s_y = \sqrt{17} \rightarrow s_y = 4{,}1$$ (Nenner / Formel)

$$r = \frac{17{,}9}{5{,}2 \cdot 4{,}1} = \frac{17{,}9}{21{,}32}$$

r = 0,84

Beispiel 2:

An zwei Tagen hintereinander laufen 5 Studentinnen, die bei der Frühjahrsdiät von BRIGITTE mitmachen, jeweils eine Strecke von 1000 m. Dabei erzielen sie die in der Tabelle aufgeführten Laufzeiten (in Minuten):

Name	erster Durchgang	zweiter Durchgang
Moni	3	3
Lissi	5	5
Jenny	11	7
Lilli	14	6
Susi	15	9

Aufgaben:

- Berechnen Sie bitte, ob es einen Zusammenhang gibt zwischen den Werten der beiden Durchgänge.
- Kommentieren Sie das Ergebnis und interpretieren Sie es hinsichtlich seiner Bedeutung für „Lernen durch Üben“.

4 BEGRIFFSERKLÄRUNGEN

Eindimensionale Häufigkeitsverteilung: Wenn Merkmalsträger hinsichtlich eines einzigen Merkmals (Dimension) untersucht werden. Sie beschreibt, wie sich die Merkmalsträger auf die Merkmalswerte des einen Merkmals verteilen (häufen). (BOURIER 2013:38)

Erhebungseinheit: Ein einzelnes Element der Grundgesamtheit. Die Anzahl der Erhebungseinheiten bildet den Umfang der Grundgesamtheit (=N). (DULLER 2013:8)

Grundgesamtheit: Die Menge aller Objekte, über die man Informationen gewinnen will. Eine exakte räumliche, zeitliche und sachliche Abgrenzung ist notwendig. (DULLER 2013:8)

Kontingenztabelle: Darstellungsmöglichkeit für zweidimensionale Häufigkeitsverteilungen

Kumulierte Häufigkeit: Die kumulierte Häufigkeit (Summenhäufigkeit) gibt die Anzahl bzw. den Anteil der Merkmalsträger an, die einen bestimmten Merkmalswert nicht überschreiten. (BOURIER 2013:40)

Merkmal: Die interessierende Eigenschaft der Erhebungseinheiten. Jedes Merkmal besitzt verschiedene Ausprägungen. (DULLER 2013:8) Die statistische Größe nennt man Merkmal. (SIBBERTSEN/LEHNE 2012:3)

Merkmalsausprägung: Den Wert, den ein Merkmal bei einem Merkmalsträger annimmt, nennt man Merkmalsausprägung. (SIBBERTSEN/LEHNE 2012:3)

Merkmalsträger: Objekte, beispielsweise befragte Personen, an denen statistische Größen gemessen werden, nennt man Merkmalsträger. (SIBBERTSEN/LEHNE 2012:3)

Repräsentative Stichprobe: Die Stichprobe zeichnet ein möglichst genaues Abbild der Grundgesamtheit. (DULLER 2013:8)

Stichprobe: Eine Teilmenge der Grundgesamtheit. (DULLER 2013:8)

Urliste: Nach einer Erhebung liegen die Daten bzw. Merkmalswerte (Urwerte, Urdaten) zunächst in Form einer sogenannten Urliste (statistische Reihe) vor. (BOURIER 2013:34)

5 SYMBOL- UND ABKÜRZUNGSVERZEICHNIS

C	Kontingenzkoeffizient
D	Modalwert (Dichtemittel, häufigster Wert)
e	durchschnittliche Abweichung (mittlere Abweichung)
$f(x_i)$	relative Häufigkeit
$F(x_i)$	kumulierte relative Häufigkeit (empirische Verteilungsfunktion)
G	geometrisches Mittel
H	harmonisches Mittel
i	Zahl der Teilnehmer
k	Anzahl der verschiedenen Ausprägungen
n	Gesamtheit der Merkmalsträger
$n(x_i)$	absolute Häufigkeit
QA	Quartilsabstand
r	Maßkorrelationskoeffizient von Bravais-Pearson
R	Rangkorrelationskoeffizient von Krueger Spearman
s	Standardabweichung (mittlere quadratische Abweichung
s^2	Varianz
SU	statistische Unabhängigkeit
Sw	Spannweite (Variationsbreite, Variationsweite)
$S(x_i)$	kumulierte Häufigkeit
x_i	Merkmalswert
$\overline{x}$	arithmetisches Mittel (Durchschnittswert)
Z	Medianwert (Zentralwert, Stellungsmittel, mittlerer Wert)

6 QUELLENVERZEICHNIS

BOURIER, GÜNTHER: Beschreibende Statistik – Praxisorientierte Einführung Mit Aufgaben und Lösungen; 11. Auflage; Springer Fachmedien Wiesbaden 2013

CLAUß, GÜNTER/FINZE, FALK-RÜDIGER/PARTZSCH, LOTHAR: Grundlagen der Statistik. Für Soziologen, Pädagogen, Psychologen und Mediziner; 6. Korrigierte Auflage; Frankfurt am Main 2011

DULLER, CHRISTINE: Einführung in die Statistik mit EXCEL und SPSS – Ein anwendungsorientiertes Lehr- und Arbeitsbuch; 3. Auflage; Springer Berlin Heidelberg 2013

Krämer, Walter: So lügt man mit Statistik; Überarbeitete Neuausgabe; Campus Verlag GmbH, München 2012

SIBBERTSEN, PHILIPP/LEHNE, HELMUT: Statistik – Einführung für Wirtschafts- und Sozialwissenschaftler; Springer-Verlag Berlin Heidelberg 2012