

Maria Neubauer
Berufsakademie Riesa

Statistik und Quantitative Methoden

2021

Gliederung der Veranstaltung

1. Vorstellung
2. Struktur und Ziele der Veranstaltung
3. Organisatorische Fragen /Lernmaterialien

4. Warum Statistik

5. Datenmanagement
6. Häufigkeiten

Vorstellung von mir

Dr. Maria Neubauer

Technische Universität Dresden, Fakultät
Wirtschaftswissenschaften, Lehrstuhl für
Wirtschaftspädagogik und
Wirtschaftsinformatik



Kontaktieren Sie mich: maria.neubauer@ba-riesa.de

Ihr Vorwissen in Statistik und Ihre Erwartungen

Was wissen Sie über das Thema Statistik?

Aufgabe Statistik-Begriffe:

Welche 10 Begriffe aus der Statistik fallen Ihnen spontan ein?

Bei dieser Aufgabe geht es vor allem darum, dass ich mir ein Bild über Ihr Vorwissen machen möchte.

Dauer: 5 Minuten

Abgabeformat: Chat im virtuellen Klassenzimmern

Was erwarten Sie sich von einer Lehrveranstaltung zur Statistik?

Kennenlernstatistik

Bitte füllen Sie diese Umfrage aus:

	Sehr gut	Eher gut	Mittel	Eher schlecht	Sehr schlecht
Wie gut schätzen Sie Ihre Vorkenntnisse in Mathematik ein?					
Wie gut schätzen Sie Ihre Vorkenntnisse in Statistik ein?					

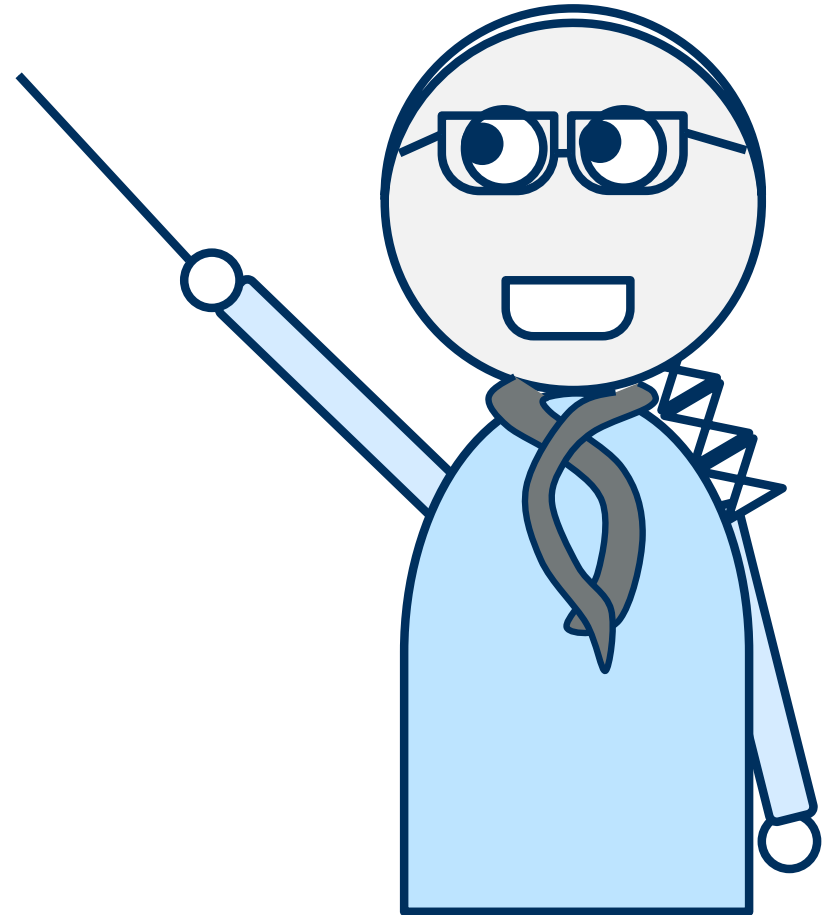
	Ja	Nein
Hatten Sie bereits ein Studium mit einem mathematischen Anteil absolviert?		
Hatten Sie Matheleistungskurs im Abitur?		
Haben Sie bereits im beruflichen oder privaten (aber nicht (hoch-)schulischen) Kontext eine statistische Auswertung durchgeführt		

Didaktischer Hinweis

Lehre = Zeit zum **Fehler** machen

Lehre = Zeit um etwas selbst tun

Lehre = Zeit für Ihre Fragen, Kritik, Anmerkungen, Vorschläge



Welche Vorteile bringt aktive Teilnahme?

- Selbst was tun: Eigene Fragen, Interaktive Elemente
- Lücken im Skript füllen
- „Geheimes“ Wissen rund um die Statistik, das nicht in klassische Lehrbüchern auftaucht
- Einfluss auf Klausurgestaltung (Sie reflektieren die Prüfungsfragen)

Lerninhalte aus der Modulbeschreibung

Wahrscheinlichkeitsrechnung

- Grundbegriffe (Ereignis, klassische Definition Wahrscheinlichkeitsrechnung, Unabhängigkeit), Zufallsgröße, Zufallsvektor, Erwartungswert, Varianz, Kovarianz, Korrelationskoeffizient, spezielle Verteilungen (Binominal, Exponential, Gleich- und Normalverteilung)
- Zentraler Grenzwertsatz
- Testverteilungen (Chi², Student-Verteilung)

Beschreibende Statistik

- Häufigkeitsverteilungen (ein- und zweidimensional)
- Statistische Maßzahlen (empirischer Mittelwert, empirische Streuung, Median, Modus)
- Regressions- Korrelationsrechnung (Regressions-Gerade und -parabel, empirischer Korrelationskoeffizient)
- Schiefe und Wölbung

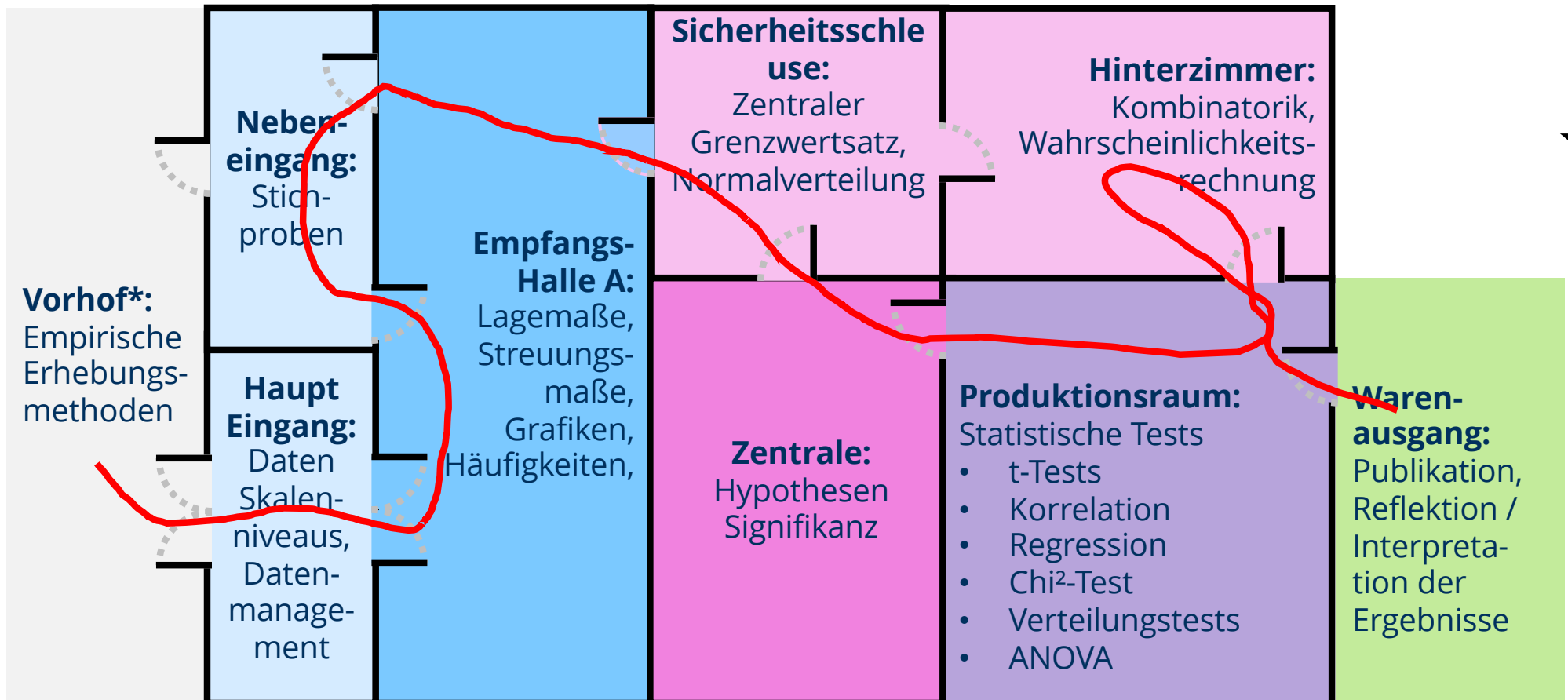
Induktive Statistik

- Einführung in die Stichprobentheorie
- Schätzverfahren (Maximum-Likelihood-Methode)
- Testverfahren (
 - Chi²-Anpassungstest,
 - Chi²-Unabhängigkeitstest und
 - Streuungstest,
 - Gaußtest,
 - t-Test)

Quantitative Verfahren 2

- Uni-, bi- und multivariate Analyseverfahren
- Parametrische und nicht-parametrische Testverfahren
- Entscheidungsunterstützung bei der Auswahl des geeigneten Testverfahrens
- Anwendung einer statistischen Standardsoftware (z.B. SPSS)

(Produktions-)Haus der Statistik und Übersicht über die Veranstaltung Statistik



*nicht Bestandteil dieser Veranstaltung

10 wichtige Lehrziele: Studierende können ...

1. aus wagen Konstrukten konkrete Variablen formen und einem konkreten Skalenniveau zuordnen.
2. darlegen, wie im Forschungsprozess alles zusammen passen muss von der Hypothese über Datenerhebung, Dateneingabe und Auswertung und Interpretation.
3. grundlegende Diagramme, Kennwerte und Streuungsmaße erstellen/errechnen und interpretieren.
4. begründen, warum Stichproben zu zufälligen Ergebnissen führen und welche Rolle der Zufall für die Rückschlüsse in der Statistik haben und können daraus ableitend erklären, warum statistische Test notwendig sind.
5. für eine konkrete Fragestellung den richtigen statistischen Test wählen und rechnen auch die Voraussetzungen des Testes beachten.
6. wenden den Begriff der Signifikanz für statistische Ergebnisse richtig an.
7. beim Lesen von statistische Ergebnissen, typische Fehler erkennen und ein schätzen, wie die Rahmenbedingungen der Studie das Ergebnis beeinflusst.
8. die Ergebnisse statistischer Tests sachlich richtig interpretieren.
9. aus vorhandenen statistischen Ergebnissen verantwortungsvolle Handlungsoptionen ableiten.
10. Fragestellungen der Kombinatorik und Wahrscheinlichkeitsrechnung mit der richtigen Methode beantworten.

Beispiele für Prüfungsfragen

Gegeben: Z.B. Frage einer Studentin, die jeden Tag 100 mal die Temperatur ihrer Katze misst und so jeden Tag einen Mittelwert notiert. „Was kann ich erwarten wenn ich an 1000 Tagen hintereinander so einen Mittelwert errechne?“ *Gesucht:* **Antwort** auf die Frage mit einer Skizze die den tagesmittelwert 1000 Tage hin darstellt + Fachwort für diese Verteilung

Gegeben: Datenreihe einer metrischen Variable: 23, 23, 25, 21, 20, 25, 26, 27
Gesucht: **Berechnung** Häufigkeit, zentrale Tendenz, Streuungen, Box-Plot. Histogramm, Konfidenzintervall.

Gegeben: Ergebnisse eines statistischen Testes. *Gesucht:* **Interpretation** der Ergebnisse in ein bis zwei Sätzen. Welcher statistische Test wurde gewählt. Was bedeutet der Test (**Theorie:** Nullhypothese, Teststatistik) und welche Handlungsempfehlungen lassen sich für die **Praxis** daraus ableiten.

Organisatorische Fragen

Haben Sie auf Ihrem Rechner Excel (oder ähnliche Programme)? Sie brauchen es für diese Veranstaltung.

Welches Vorwissen in Excel / OpenOffice Calc haben Sie?

Welche Statistiksoftware nutzen wir, warum?

	Tabellenkalkulation, Excel, OpenOffice Calc, Google Tabellen	PSP	SPSS / STATA	Software R
Verfügbarkeit z.B. auf Rechner eines Unternehmens	+++	++	--	++
Datenmanagement	+	++	+++	---?
Anzahl an statistischen Prozeduren	--	+	++	+++
Anzahl an statistischen Prozeduren				
Grafiken	++	0	+	-/+ (Programmierkenntnisse müssen erarbeitet werden)
Kosten	-/+	+++	--- / -- (SPSS: nur 1-Jahres-Lizenzen, STATA: teuer)	+++
Skriptsprache / Klickmenü	Hauptsächlich Klickmenüs	Beides	Beides	Nur Skriptsprache
Bonus	Schöne Grafiken, einfacher Einstieg, Werkzeug für viele andere Anwendung		Standard in Sozialwissenschaft	Werkzeug für viele andere Anwendung

Tipp für Datenanalyse in Excel. Add-In Datenanalyse

Verbesserte Datenanalysen in Excel lassen sich mit dem Add-In „Datenanalyse“ erstellen. Dafür müssen Sie nichts installieren sondern **nur aktivieren**.

Das Add-In „Datenanalyse“ wird so aktiviert:

- Gehen Sie auf „Datei“
- Dann „Optionen“
- Dort auf der linken Seite auf „Add-Ins“
- Unten unter „Verwalten“ „Excel-Add-Ins“ einstellen und auf → Los klicken
- Haken setzen bei Analyse-Funktionen → OK

Nun ist das Add-In unter → Daten → Analyse → Datenanalyse verfügbar.

Wir werden in dieser Veranstaltung immer wieder auf das Add-In zurückgreifen.

Empfohlene Lernmaterialien: Lehrbücher



Umfassend lexikonartig: Bortz, J.; Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler. Berlin, Heidelberg: Springer.

Grundlagen übersichtlich beschrieben: Fahrmeir, L. u.a. (2010): Statistik – Der Weg zur Datenanalyse: 7. Aufl., Berlin: Springer

Lösungen zu Fahrmeier u.a.: Caputo, A. u. a. (2009): Arbeitsbuch Statistik, 5. Aufl., Berlin: Springer

Vertiefende Behandlung stat. Tests: Backhaus, K; Erichson, B.; Plinke, W.; Weiber, R. (2016) Multivariate Analysemethoden: Eine anwendungsorientierte Einführung, 14. Aufl., Wiesbaden: Springer Gabler.

Weniger Statistik, dafür Datenerhebung, -bearbeitung, ebenfalls umfassend lexikonartig: Döring, N.; Bortz, J. (2016): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, 5. Aufl., Berlin: Springer

Vertiefende Behandlung stat. Tests: Hair, J. u.a. (2014): Multivariate Data Analysis. 7 Aufl. Harlow.

Empfohlene Lernmaterialien: Bücher

Videos: Youtube (z.B. „Mathe by Daniel Jung“, „Sebastian Stoll“)

Tafelwerke: Ein Tafelwerk das Kombinatorik, Wahrscheinlichkeitsrechnung und Statistik enthält (Achtung überall unterschiedliche Notation)

Erkläreseiten:

- Für Lehrer*innen: https://lehrerfortbildung-bw.de/u_matnatech/mathematik/gym/bp2004/fb2/modul4/
- <https://matheguru.com>; www.statistik-nachhilfe.de;
<https://statpages.info/#WhichAnalysis>

Sonstige: Gezielte Internetrecherche nach *Stichworten* aus der Veranstaltung

In Präsenzveranstaltung/Chat: Fragen stellen

Folien dieser Veranstaltung: ... *erhalten sie alle*

IHR TIPP:

Warum Statistik?

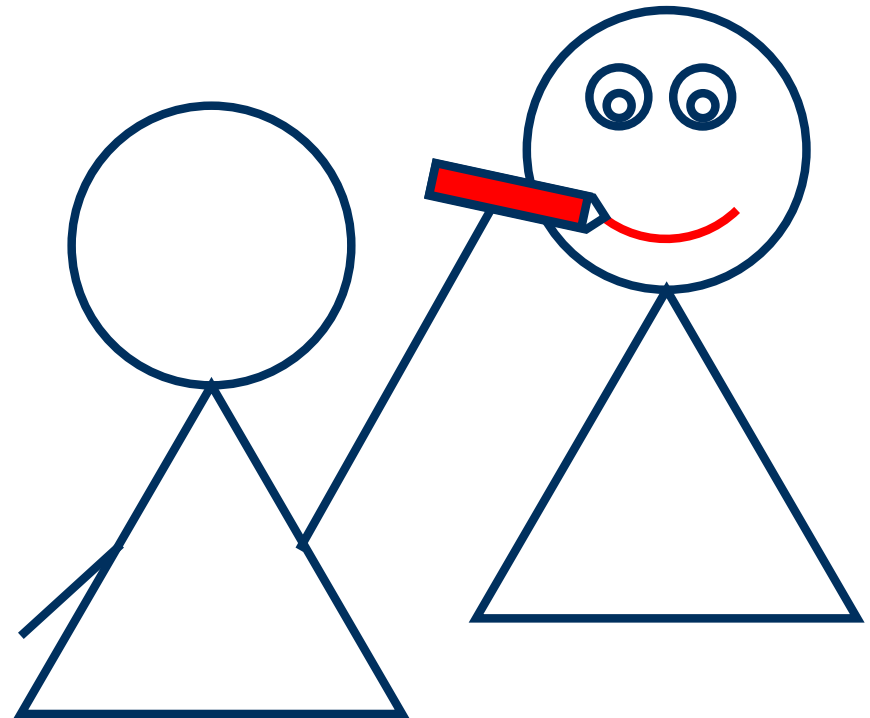
Wann und Wo Statistik?

Aufgabe Warum Statistik:

Stellen Sie sich für wenige Sekunden vor, in welchen Situationen Sie Statistikwissen in Zukunft im beruflichen und privaten Leben brauchen werden und es schon verwendet haben. Hinweis: Das Lernen wird Ihnen leichter fallen, wenn Sie wissen, wozu Sie etwas lernen. Am besten Sie haben eigene Ideen.

Dauer: 1 Minute

Abgabeformat: Chat



Erwartungen und Ängste für Statistik

Scrabble: Als Vorstellungsrunde oder Erwartungen

Mit dem Begriff:

Statistik Veranstaltung

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

Aufgabe Frageformulierung:

- Was ist an dieser Frage falsch formuliert?
- Das ist keine rein statistische Frage.

Dauer: 2 Minuten **Abgabeformat:** Chat

Prozente und Mittelwerte verstehen

Aufgabe Creme: Sie haben eine Feuchtigkeitscreme gekauft. Auf der Verpackung lesen Sie „Sie hilft zu 75 %“ (getestet an 16 Konsumenten). Wie interpretieren Sie dies?

Dauer: 2 Minuten **Abgabeformat:** Chat

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

Dauer: 2 Minuten **Abgabeformat:** Chat

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?

Dauer: 2 Minuten **Abgabeformat:** Chat

Warum Statistik? Streuung verheimlichen

- Aussage: Im Durchschnitt machen die Deutschen 32 Überstunden im Jahr.
- Das Problem hier: das trifft nicht für jeden zu, besser wäre die Angabe eines Intervalls.
- Es gibt ein Intervall: Das *Konfidenzintervall*, es umfasst den „wahren Wert“ zu 95 Prozent %.

Bedeutung von Datenmustern erkennen

Beispiel: Auswertung einer Statistikklausur

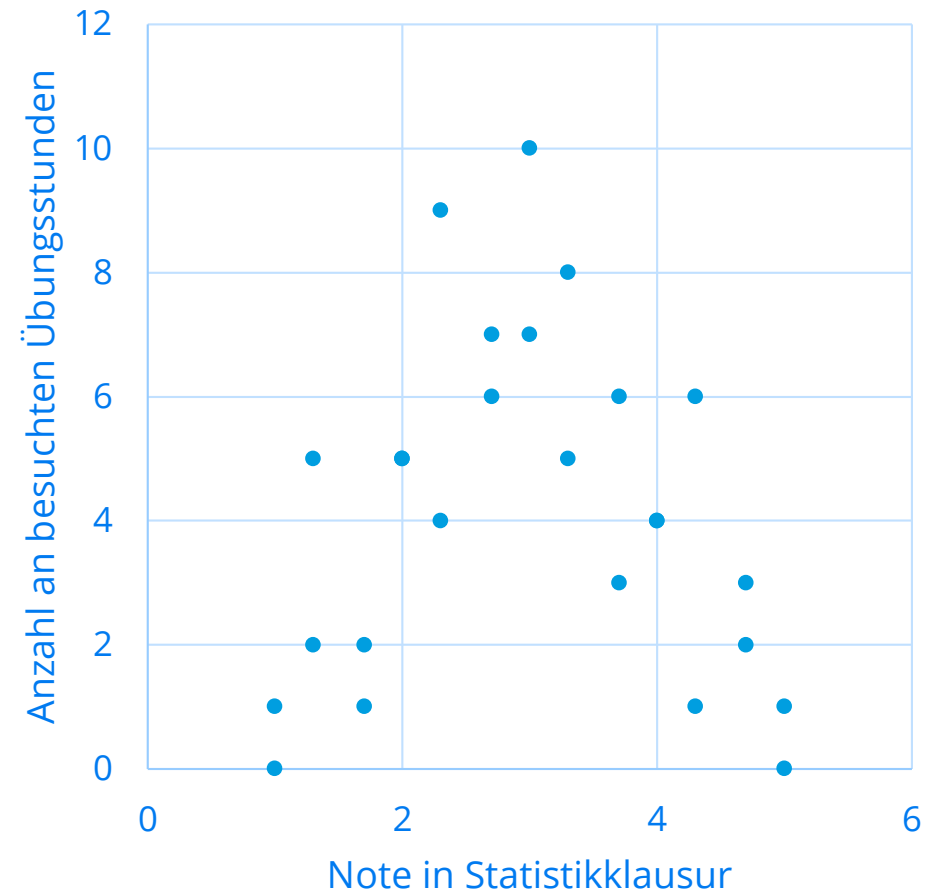
Aufgabe Übungsstunden und

Note: 1. Wie interpretieren Sie die nebenstehende Grafik?

2. Wie kann es zu diesem Zusammenhang gekommen sein?

Dauer: 5 Minuten

Abgabeformat: Chat



Schwellenwerte festlegen

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:** Chat

Warum Statistik? Den Begriff der Signifikanz richtig verstehen

Typische Formulierungen in wissenschaftlichen Texten verstehen:

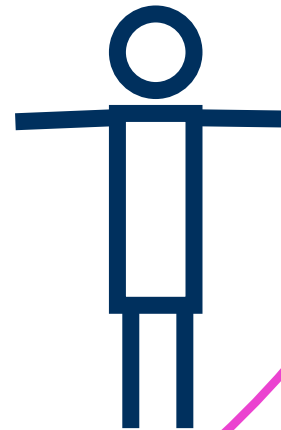
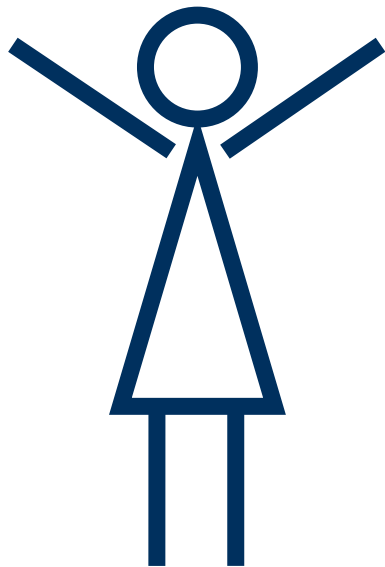
Das A Medikament ist signifikant wirksamer als B.

→ Einfach: A ist besser als B.

Von 100 getesteten Mitteln hatten fünf *signifikante* Wirkung.

→ Achtung! Am Ende der Statistik Veranstaltung werden Sie verstehen, dass dies kein gutes Zeichen ist.

Datengrundlagen berücksichtigen: Amazons Algorithmus zur Einstellung von Bewerbern bevorzugte Männer



Aufgabe Amazon:

Lesen Sie den Bericht über die Amazons (verworfenen) Einstellungspraxis bei Bewerbern: <https://www.golem.de/news/machine-learning-amazon-verwirft-sexistisches-ki-tool-fuer-bewerber-1810-137060.html>

Dauer: 3 Minuten **Abgabeformat:** keine

Definition / Messung von Konstrukten

Die Zahl der Armen fällt!

Die Zahl an Krebsfällen steigt!

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?

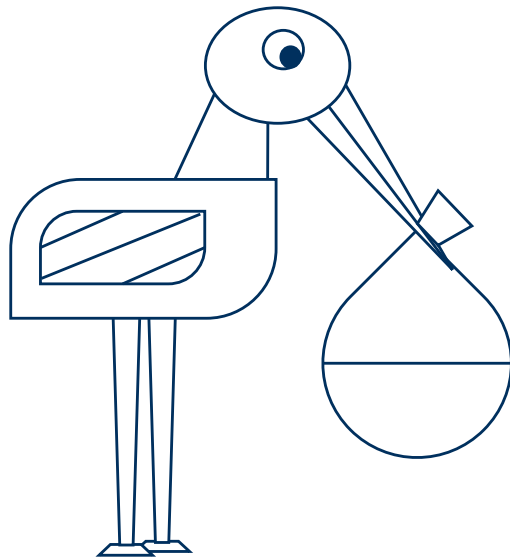
(Die Überschrift dieser Folie gibt Ihnen einen Hinweis)

Dauer: 2 Minuten

Abgabeformat: Chat

Scheinkorrelation entdecken

Scheinkorrelation
(Scheinkausalität) kennen Sie auch
von dem Zusammenhang von
Störchen und Babys



- | **Aufgabe Störche:**
- | Welche Scheinerklärung kennen Sie,
| die den Zusammenhang zwischen
| Babys und Störchen erklärt? Geht es
| um Raum oder Zeit?
- | Lesen Sie auf Wikipedia die Erklärung:
| <https://de.wikipedia.org/wiki/Scheinkorrelation>
- | **Dauer:** 5 Minuten **Abgabeformat:**
| keine Abgabe

Scheinkorrelation entdecken

Ich habe immer gedacht, dass Korrelation Kausalität bedeutet. Aber dann habe ich einen Statistikkurs gemacht. Nun denke ich das nicht mehr.

Hört sich so an, als ob der Kurs geholfen hat.

Ja vielleicht

Weitere verrückte Scheinkorrelationen finden Sie hier:
<https://scheinkorrelation.jimdofree.com/>

Scheingenauigkeit mit Schlussfolgerungen in der Sozialwissenschaft, Physik, Mathematik

Ein Sozialwissenschaftler, ein Physiker und ein Mathematiker fahren mit einem Zug durch die Schweiz. Als sie aus dem Fenster schauen, entdecken sie auf einem Acker ein schwarzes Schaf. Der Sozialwissenschaftler, der noch nie in der Schweiz war und hier das erste und bisher einzige Schaf dieses Landes kennen lernt, folgert messerscharf: "Aha – in der Schweiz sind alle Schafe schwarz!" Der Physiker denkt, er wäre schlauer und macht sich sogleich über den Sozialwissenschaftler lustig: "Das ist eine völlig unerlaubte Verallgemeinerung – das einzige, was du sagen kannst, ist: Es gibt in der Schweiz ein schwarzes Schaf." Der Mathematiker, der sich bisher nicht an der Diskussion beteiligt hatte, kann daraufhin nur müde lächeln und meint: "Auch das ist völliger Unsinn. Du kannst nur behaupten: Es gibt in der Schweiz ein Schaf, das von einer Seite schwarz ist!"

Scheingenauigkeit im Umgang mit Zahlen in den Sozialwissenschaften

1. 55,2304358% aller Statistiken täuschen eine viel zu hohe Genauigkeit vor und sind damit immun vor Kritik.
2. Und 115 % aller Statistiken beruhen auf übertriebenen Zahlen.
3. Es werden **immerhin** in 3 von 18 Fällen Grafiken richtig dargestellt, während **nur** in 20 Prozent Tabellen richtig sind.
4. Die Krebsrate der Bewohner um Atomkraftwerke ist 500 % größer als normalerweise, sie beträgt 0,0005 %.
5. Achtung 99,999 % alle Statistiken auf dieser Folie sind ad hoc entwickelt.
6. Die Statistik beweist, dass 97% aller Statistiken uns nichts sagen. (Erhard Blanck)

Ablauf einer statistischen Analyse

Der Prozess der Datenerhebung und Auswertung in der Statistik

In Statistiklehrbüchern

Planung

Datenerhebung

Datenaufbereitung und -darstellung

Datenanalyse und -interpretation.

Beim Statistischen Bundesamt

Bedarfsermittlung und Gesetzgebung

Erhebung von Daten

Prüfung und Auswertung

Datenschutz und statistische
Geheimhaltung

Veröffentlichung



Zoom auf die Planung der Datenerhebung

Fragestellung aufstellen

Definition der Grundgesamtheit

Variablen definieren, die erhoben werden

Variablen operationalisieren

Erhebungseinheiten definieren

Größe und Art der Stichprobe bestimmen



Dies ist nicht Teil der
Statistikveranstaltung

Datenmanagement: Bedeutung

- Wir beginnen nun mit dem Datenmanagement. Denn Statistik beruht auf Daten.
- Benötigte Zeit für das Datenmanagement im Vergleich zu anderen Schritten:

Vorbereitung der Daten (Eingabe, Fehlersuche, Deskriptive Statistiken; Berechnung neuer Variablen)



Fragestellung bilden
Erstellung Erhebungsinstruments

Erhebung der Daten:
Befragte kontaktieren;
Warten auf Rücklauf

Berechnung
Ergebnisse
Verschriftlichung der Ergebnisse

Datenmanagement

Datentypen

Dateneingabe

Datenvorbereitung

Klassische Situation für die Nutzung von Statistik

Sie haben in Ihrem Unternehmen z.B. eine neue Kommunikationsplattform probiert, die z.B. die Mitarbeiterzufriedenheit heben soll. Nach und vor der Einführung haben Sie eine **Mitarbeiterbefragung** gemacht mit Hilfe eines Fragebogens aus Papier.

Sie wollen wissen,

- gibt es eine Verbesserung der Zufriedenheit.
- unterscheiden sich die Zufriedenheit in Abteilung A und B?

Folgende drei Schritte bieten sich nun an:

- 1. Schritt: Bringen Sie die Daten in Tabellenform.**
2. Schritt: Analysieren Sie die Daten deskriptiv (z.B. Häufigkeiten, Mittelwert)
3. Schritt: Analysieren Sie die Daten mittels schließender Statistik

Dateneingabe: Fälle und Variablen

Bei der Übertragung von Daten in (Excel-)Tabellen sollten Sie so vorgehen, wie es auf dieser Folie gezeigt wird. Dies ist der Standard.

Zeilen = **Fälle** (auch Beobachtung genannt)
Personen, Staaten, Schulklassen, allgemein auch als Untersuchungsobjekt bezeichnet

Spalten = **Variablen**, Merkmale/Eigenschaften von Personen/Untersuchungsobjekten (Größe, Name) Frage im Fragebögen

	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			

Aufgabe Fälle/Variablen:

1. Was kann von dem Folgendem alles ein **Fall** sein?: Menschen, Personen, Gruppen, Staaten, Organisationen, Planeten
2. Was von dem Folgendem kann eine **Variable** sein? Alter, Geschlecht, IQ oder Größe, Jahresgewinn eines Unternehmens
3. Fällt Ihnen etwas ein, was niemals ein Fall sein kann?

Dauer: 4 Minuten **Abgabeformat:** Chat

Skalenniveaus

Variablen können verschiedene Arten/Typen oder besser „Skalen“ haben.

Symbol bei SPSS	Skala	Merkmale
	Nominal-Skala	Klassifizierung qualitativer Merkmale
	Ordinal-Skala	Rangwerte mit Ordinalzahlen <i>Diskrete Daten</i>
	Metrische-Skala	Gleiche Abstände <i>(Stetig Daten)</i>

Zur Vollständigkeit: die metrische wird noch in Verhältnisskala und Intervallskala unterschieden (Die Verhältnisskala hat einen natürlichen Nullpunkt. Die Intervallskala hat einen willkürlichen Nullpunkt (z.B. Celsius) In der sozialwissenschaftlichen statistischen Praxis spielt das aber selten eine Rolle was der Nullpunkt ist).



Zusätzliche Informationen: Ordinale Daten können manchmal als metrisch interpretiert werden

Die Information auf dieser Seite können Sie sich mal überfliegen. Sie kann für Ihre praktische Tätigkeit relevant werden. Tatsächlich gibt es in der Wissenschaft heftige Diskussionen über dieses Thema.

- Es gibt statistische Tests für ordinale und statistisches Tests für metrische Daten. Theoretisch und praktisch sind die Ergebnisse meist sehr ähnlich.
- Es gibt nun empirische Belege, dass einige ordinale Skalen auch als metrische Skalen interpretiert werden können: z.B. wenn es sich z.B. um Einschätzungsfragen auf eine Skala von 1 bis 10 handelt.
- Bei ordinalen Daten wie der Schultyp (Hauptschule, Realschule, Gymnasium) – wo die Abstände zwischen den einzelnen Kategorien nicht gleich definiert sind, muss die Skala ordinal sein.
- Im Notfall können Sie die ordinalen Skalen auch als nominale interpretieren und z.B. als Gruppen behandeln. Dann können Sie je nach Art der Daten einen t-Test / Anova oder Chi-Quadrat-Test durchführen.

Aufgabe: Dateneingabe 1:

Sie haben einen Fragebogen (für Jugendliche) erstellt der folgende Fragen enthält. Erstellen Sie ein Datenblatt dass alle Variablen enthält. Achtung einige Variablen sind einfach andere schwerer umzusetzen.

- Name (offene Frage)
- Alter (in Jahren)
- Geschlecht (weiblich, divers, männlich)
- Besitz eines Fahrrades (ja/nein)
- Wie oft machen Sie Fitnessstraining (sehr selten, eher selten, eher häufig, sehr häufig)
- Wie oft fahren Sie mit dem Rad (1 Mal im Jahr, 1 Mal im Monat, 1 Mal in der Woche, täglich)
- Wie hoch ist Ihre Motivation zur Schule zu gehen? (Skala von 1 (keine Motivation) bis 5 (hoch motiviert))
- Ihr Berufswunsch? (offene Frage)
- Was sind Ihre **drei** Lieblingsfächer? (Liste der Fächer: Sport, Deutsch, Mathe, Physik, Biologie, Sozialwissenschaft, Musik)

Erfinden Sie vier Fälle und erfinden Sie dafür die Daten.

Welches Skalenniveau liegt jeweils vor?

Dauer: 20 Minuten **Abgabeformat:** Tabelle

Aufgabe Dateneingabe 2:

Fügen Sie folgende zwei Fragebogenfragen auch in die Tabelle ein:

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

Sortieren Sie folgende Dinge nach Beliebtheit?

Äpfel, Birnen, Pflaumen, Bananen?

Dauer: 5 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*

- Wissensfrage 2: Wann ist Ostern?
 - Im Frühling *Richtige Antwort*
 - Im Herbst *Falsche Antwort*
 - Im Advent *Falsche Antwort*
 - Zwischen Fasching und Pfingsten *Richtige Antwort*

Dauer: 5 Minuten Abgabeformat: Tabelle

Tipps: Datenmanagement in Excel

Sortieren Sie Ihre Daten so: Erste Zeile = Überschriften der einzelnen Fragen.

Nummerieren Sie **jeden** Papierfragebogen auch auf dem Papier und nutzen Sie die Nummer im Datenblatt

A	B	C	D	E	F	
Nummer	Bef	Geschlecht	Alter	Frage 1	Frage 2	Frage 3
1		1	34	2	das hatte ich	3
2		2	23	4	keine Ahnung	4
3		2	25	9999	das mussten	2
4		1	30	1	weil wird immer	1
6		1	29	2	9999	4

Im Tabellenblatt Legende beschreiben Sie für jede Variable die Codierung u.Ä. z.B. was bedeutet eine 9999.

Im Tabellenblatt Daten/Rohdaten etc. ist die Originalversion ihrer Daten.

In weiteren Tabellenblättern können Sie Ihre Analysen eingeben. Dort können Sie ggf. Kopien der Variablen einfügen, für eine leichtere Analyse.

Umcodieren Excel: Die Funktionen

- Für Anfänger etwas schwierig ist das „Ausdenken und Neuformulieren“ von Variablen und das Anpassen von Variablen an gewünschte statistische Tests daher üben wir das auf den nächsten Folien.
- Wenn Sie erprobt im Umgang mit Excel sind, dann ist das eine Wiederholung für Sie, wie in Excel Funktionen ausgedrückt werden. Wir benötigen einfache Funktionen um neue Variablen zu berechnen. Hier stellt sich eine Formel vor:

=Summe(A1:A4)

Hallo! Ich bin eine Funktion, das erkennt man an dem „Gleichheitszeichen“

„Folgendes addieren

... die Werte in den Zellen A1, A2, A3, A4“

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
- B. Tragen Sie die Antworten für fünf fiktive Kunden ein.
- C. Erstellen und errechnen Sie vier neue Variablen für alle Kunden mit
 - 5. Variable: Durchschnittlichen Summe in Euro an bestellten Waren
 - 6. Variable: Totale Summe in Euro an bestellten Waren
 - 7. Variable: Differenz zwischen maximaler und minimaler Summe in Euro an bestellten Waren
 - 8. Variable: Erfinden Sie eine weitere Variable, die auf (1-7) vorangegangenen Variablen basiert

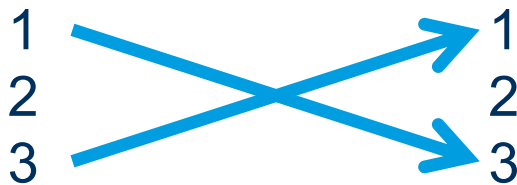
Dauer: 20 Minuten **Abgabeformat:** Tabelle

Umcodieren

- Wenn Sie einen Fragebogen haben, bei dem die Daten eine andere Form haben, als Sie brauchen müssen Sie die Daten oft umcodieren.
- **Simple Möglichkeiten:**
 - Sortieren und manuell mit Werten ersetzen
 - Suchen und Ersetzen (evtl. in eigenem Tabellenblatt)
 - Einfache Berechnungen (siehe nächste Folie)
- **Fortgeschrittene Möglichkeiten:**
 - In Excel die Funktion „=Wenn()“ nutzen.
 - Tipp: Wenn Sie Papierfragebögen haben, dann sollten Sie die Daten gleich in das ideale Datenformat überbringen.

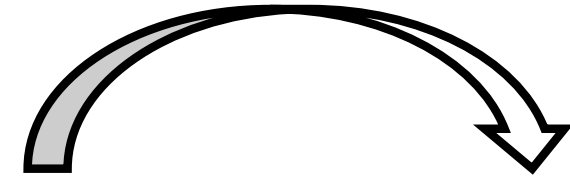
Umcodieren: Beispiele

- Beispiel 1: Aus dem Wert 1 soll der Wert 3 gemacht werden



- In dem Fall ist das relativ einfach mit einer Funktion: =4-Wert

Beispiel 2: Aus den Originaldaten („Motivation“) wurde eine neue Variable generiert („MotiviertZahlen“).



ID	Motivation	MotiviertZahlen
1	Sehr motiviert	1
2	Eher motiviert	2
3	Sehr unmotiviert	4
4	Sehr motiviert	1
5	Sehr unmotiviert	4
6	Sehr motiviert	1
7	Eher motiviert	2
8	Eher unmotiviert	3
9	Eher unmotiviert	4
10	Sehr motiviert	2

Aufgabe Umcodieren:

Sie haben eine Online-Befragung mit 15 Teilnehmern durchgeführt. Nun haben Sie eine vollständig ausgefüllte Tabelle zum Thema Motivation auf einer Skala von 1-4 (sehr motiviert, eher motiviert, eher unmotiviert, sehr unmotiviert). Leider hat die Onlinesoftware die Zahlen anders codiert, als Sie es möchten. Sie wollen überall dort wo originale eine 1 ist eine 5, wo eine 2 ist eine 4 usw. Was wäre der schnellste Schritt in Excel für diese Umcodierung?

1 → 5

2 → 4

3 → 3

4 → 2

5 → 1

Dauer: 3 Minuten **Abgabeformat:** Tabelle und Chat

Fehlende Werte

Wenn ein Teilnehmer auf eine Frage keine Antwort gegeben hat, nennt man das fehlende Werte.

Nr.	Motivati on	Arbeitsl eistung
1	2	3
2	3	4
3	4	2
4	5	
5	2	2

Aufgabe fehlende Werte:

Was machen Sie mit fehlenden Daten?
Nennen Sie mindestens zwei Lösung.

Dauer: 4 Minuten **Abgabeformat:** Chat

Weitere Tipps für das Datenmanagement

- Grundlage der für die Analyse sind beispielsweise Fragebögen. Diese sollten einwandfrei sein (Pretest, mehrmals Korrektur lesen lassen, Literatur zur Fragebogenkonstruktion).
- Für eine bessere Übersicht:
 - neu berechnete Variablen immer wieder löschen, wenn sie nicht gebraucht werden.
 - Aber: Niemals Stammvariablen löschen
 - (In SPSS/Stata Syntax verwenden, damit jederzeit die Variablen schnell neu berechnet werden können.)
- Das Ausdenken und Neuformulieren von neuen Variablen und das Anpassen von Variablen an gewünschte statistische Tests erfordert Erfahrung und auch Kreativität.

Häufigkeiten

Häufigkeiten

Verkaufte Produkte pro Quartal

	Absolute Häufigkeit	Kumulierte Häufigkeit	Relative Häufigkeit
1. Quartal	23		
2. Quartal	34		
3. Quartal	32		
4. Quartal	12		
Summe			

Aufgabe Häufigkeiten:

- a) Bilden Sie die relativen und kumulierten Häufigkeiten.
- b) Bilden Sie die Summe, wenn es sinnvoll ist.
- c) Welche der drei Häufigkeitsauswertungen sind leichter zu interpretieren? Begründen Sie Ihre Antwort.

Dauer: 10 Minuten **Abgabeformat:** Tabelle a und b; Chat c.

Übungen

Aufgabe Häufigkeiten Studierende:

Wir haben einen Rohdatensatz der folgende Variable für 10 Fälle beinhaltet:

ID) Nummer des Befragten

A) Studiendauer in Semester

Erstellen Sie eine Häufigkeitstabellen für diese Variable mit absoluter Häufigkeit, relative Häufigkeit und kumulierte Häufigkeit.

ID	A
1	5
2	5
3	7
4	7
5	9
6	5
7	6
8	7
9	8
10	6

Dauer: 10 Minuten **Abgabeformat:** Tabelle

Zweidimensionale Häufigkeitsverteilung: Kontingenztafel

Zweidimensionale Häufigkeitsdarstellungen sind Sie sicher schon begegnet. Wir werden Sie später noch mal brauchen als Grundlage für einen statistischen Test.

	Kartenzahler	Barzahler	Gesamt
Frauen	2	4	6
Männer	3	1	4
Gesamt	5	5	10

Übung Zweidimensionale Häufigkeitsverteilung

Aufgabe Häufigkeiten Studierende:

A. Bilden Sie einen eigenen Rohdatensatz aus dem eine zweidimensionale Häufigkeitsverteilung gebildet werden kann. Mit 10 Fällen.

Hilfe:

- Sie brauchen zwei Variablen.
- Jede Variable sollte min. 2 maximal 2-4 Ausprägungen haben.

B. Bilden Sie die Zweidimensionale Häufigkeitsverteilung.

Dauer: 15 Minuten **Abgabeformat:** Tabelle

Aufgabe: Mit Daten arbeiten

Gehen Sie auf folgende Seite und wählen Sie eine beliebige Tabelle.

GENSIS Datenbank vom Statistischem Bundesamt: <https://www-genesis.destatis.de/genesis/online>

Teilen Sie die Tabelle mit Ihrer Gruppe und beschreiben Sie die Tabelle.

- a) Auf welche Fälle/Erhebungseinheit beruht diese Tabelle.
- b) Erklären Sie mindestens zwei Variablen.
- c) Erklären Sie (oder stellen Sie eine Vermutung an) wie diese Daten erfasst wurden und von wem.
- d) Erfinden Sie eine (interessante) Fragestellung, die Sie mit den Daten eventl. beantworten könnten.
- e) Welche Variablen benötigen Sie für Fragestellung aus d)

Dauer: 15 Minuten **Abgabe:** Gruppendiskussion