**Angewandte empirische Sozialforschung**

Jemand möchte etwas über mich wissen, als Teil eines Ganzen. Datenbeschaffung zum Zwecke höherer Erkenntnisse.

* Bsp.: Postleitzahl abfrage bei Einkäufen
* Straßenbefragungen über allgemeine Themen
* Straßenbefragungen über gezielte Themen.

Marktforschung ist einer der häufigsten Gründe für empirische Sozialforschung. Demoskopie (Meinungsforschung) ist eine ebenfalls weit verbreitete Methode für empirische Sozialforschung.

**Quantitativ**:

* Beginnt anders als bei Qualitativ. Wichtig ist, beide Elemente getrennt zu betrachten.
* Hypothese
  + Vermutung welche einem wissenschaftlichen Standard entspricht
  + Realitätsbezug
  + Informationsgehalt
  + „Wenn“ „dann“ – Sätze

**Fragebogen**:

* Fragt ganz stur.
* Häufig durch geschlossene (Standardisiert) Fragen geprägt. Häufig fehlen hier wichtige (für einen selbst) Begriffe. Dadurch wird immer das Feld „Sonstige“ beigefügt. Geschlossene Fragen beeinflussen das Urteilsvermögen. Man denkt nicht mehr selbst nach, sondern sieht die vorgegebenen.

25.10.2019

**Trendextrapolationen**:

* Ergebnisse werden auf die Zukunft ausgelegt und dienen als *Prognose*.
* Je weiter dies in der Zukunft liegt, desto unwahrscheinlicher die Vorhersage.
* Wissen aus der Vergangenheit (Daten) werden bis heute genommen und Mithilfe von Formeln werden die prozentuale Zunahme (oder Abnahme) von Gehältern errechnen.
  + Nominal und Reallohn.
    - Nominal ist das was ich erhalte (sagt nichts aus) abhängig von Inflation.
* Bsp.: Zeitraum von 1970 bis heute wird die Steigerung der Löhne ermittelt (Ergebnis in %). Diese Prozentzahl nehme ich und errechne bis in das Jahr 2030 die Lohnentwicklung.
  + Unbekannte Variablen welche bis 2030 eintreten können, welche nicht vorhersehbar sind, werden vernachlässigt.

**Kumulieren**:

* Aufaddieren von Werten.

**Scheinpräzision**:

* Eine Zahl welche Eindeutigkeit signalisieren soll und dadurch die Glaubwürdigkeit unterstreichen soll.
* Die Statistik zeigt eigentlich etwas was nicht präzise ausgesagt werden kann und bildet dies visuell aus. Eine Präzise Zahl kann hier nie/kaum ausgegeben werden. Statistik bildet nie die Realität ab (Anhaltspunkte)
* Genaue Zahlen sagen nichts aus und sind stets falsch. Solche Statistiken sind generell als unwahr zu beurteilen.
* Statistik im Krieg ist stehts Kriegspropaganda und soll das eigene Land „gut“ darstellen.

Statistik muss nicht zwingend falsch sein. (Krebsstation Bsp.). Wenn die Wirklichkeit offensichtlich verändert wird, kann dadurch eine Statistik positiv/negativ beeinflusst werden.

Mit der Fragestellung kann man sehr viel beeinflussen (Bsp. IG-Metall). Man kann eine Frage so formulieren, dass die Antwort quasi direkt mitgeliefert wird „Suggestion“

* Bsp.: Wie alt waren Sie, als Sie das erste Mal geklaut haben 🡪 Jedem wird unterstellt, einmal geklaut zu haben.

**Begriffe**:

* Definition von Begriffen welchen man sucht.
  + Bsp.: Blick aus der Juristischen Seite: Was ist ein Jugendlicher 🡪 Jugendlich ist wer das 14 Lebensjahr…
  + Bsp.: Blick aus der Soz.Päd.: Was ist ein Jugendliche 🡪 Jugendlich ist wer…

**Grundbegriffe der Statistik**

* Untersuchungszweck: Was möchte ich genau wissen?, häufig ist dies oft nicht klar und dadurch werden Statistiken nicht eindeutig.
* Bsp.: Lohnt studieren? 🡪 Frage dahinter: Empfinden Absolventen ihr Studium als ein lohnendes Studium.
  + Untersuchungszweck muss genau definiert werden um daraus abzuleiten was meine Untersuchungs(Grund)gesamtheit (Population) ist.
  + Wen Frage ich ? für diese Fragestellung sollten keine 1. Semester Studenten befragt werden.
    - Alle Studenten, welche sich im letzten Semester ihres Studiums befinden.
  + Kann ich überhaut die Grundgesamtheit befragen ?
    - Örtlich begrenzte Befragungen bspw. Studenten nur in Weingarten, aber es gibt auch noch Studenten anderer Hochschulen. (Häufig mit Kosten verbunden)
  + Stichproben aus der Befragung ziehen! Diese Stichprobe muss repräsentativ sein!
    - Soz. Merkmale oder was zu erfassen gilt, muss prozentual in einer Stichprobe enthalten sein.
    - Bsp.: 80% sind Frauen, 20% sind Männer (Grundgesamtheit), daher muss die Stichprobe aus 80% Frauen bestehen. 🡪 Die Stichprobe muss der Realität entsprechen.
    - Bsp.: Was wäre wen am Sonntag Wahl wäre, was würden Sie wählen: Grundgesamtheit wäre: Alle wahlberechtigten (deutsch, volljährig).
      * Stichprobe 66% wahlberechtigte wäre nicht ausreichend, daher eine repräsentative Befragung.
    - Repräsentative Stichproben können auf zwei Wegen gewonnen werden:
      * Zufallsstichprobe:
        + 500 SA-Studierende, will eine Stichprobe von 100, Liste mit Namen und jeder 5te wird herausgezogen.
        + Montagmorgen am Eingang von Gebäude B, jeder der hereinläuft frage ich (Kein Zufall, da die Leute, welche dieses Gebäude betreten müssen befragt werden, andere fallen heraus). Jedes Element einer Zufallsstichprobe muss dieselbe Chance haben gefragt zu werden.
        + Das meiste was als repräsentativ dargestellt wird, ist es häufig gar nicht.
        + Willkürstichprobe / Bewusste Stichprobe

Quotenstichprobe ist die bekannteste (Marktforscher)

Anhand von Sozialstatistischen Merkmale wird geschaut (häufig 5 Merkmale). Anschließend konstruiert man Personen, welche diese Merkmale haben. Diese spiegeln dann die Realität. (Wenn ein Modell, die Realität so perfekt wie möglich darstellt. Je detaillierter, desto besser.

* + - **Grundregel**: Ich darf alles machen, muss allerdings dies gut beschreiben, *sie müssen logisch widerspruchsfrei sein*, *kein offensichtlicher Unsinn* (nicht mit etwas argumentieren, was für andere nicht nachvollziehbar ist) und *was gemacht wird muss intersubjektiv* (zwischen dir und mir bsp.: Ein anderer muss dies anhand einer Handlungsstrategie nachvollziehen/nachmachen können) *nachprüfbar sein*.
* Merkmal
  + Alter
* Merkmalsausprägung
  + Menschen haben unendlich viele Merkmalseigenschaften welche in einem Oberbegriff eingeordnet werden 🡪 einem Merkmal
    - Alter
    - Geschlecht
* Merkmalträger
  + Befragte welche die Informationen haben.
  + Institutionen (Bsp.: Beratungsstellen)
  + Menschen haben unendliche viele Merkmalseigenschaften.
* Bsp.: Anja (Vorname und Geschlecht bekannt)
  + Anja studiert Soziale Arbeit im 8. Semester
  + Sammelt Briefmarken
  + Anja trägt eine Brille
  + Konservativ eingestellt
  + 27 Jahre alt.
  + Schuhgrüße 38
  + Größe 1,62m
  + Einkommen 280 €
  + Staatsangehörigkeit: Dänisch
  + Blond
  + Hilfsbereit, kommunikativ, trinkfreudig, nah am Wasser gebaut
  + Lebend in einer Partnerschaft.
* 1. Kategorie: Metrische Merkmale
  + Körpergröße
  + Schuhgröße
  + Einkommen
  + Alter
* 2. Kategorie: Komparativ oder ordinale Merkmale (Rangmerkmale/ mehr oder weniger ausgeprägt)
  + Konservativ
  + Bsp.: Sehr gesprächig 🡨🡪 ganz ruhige = Rang
* 3. Kategorie: Klassifikatorisch oder nominale Merkmale (Unterschiedsmerkmale)
  + Name
  + Haarfarbe
  + Nationalität / Staatsangehörigkeit
  + Ausbildung
  + Geschlecht
  + Hobbie
  + Partnerschaft
  + Brille
* Innerer Zusammenhang von den drei Kategorien
  + Nicht alle Merkmale sind wirklich eindeutig.
    - Bsp.: Anja hat einen IQ von 140 (metrisch weil messbar, komparativ weil mehr oder weniger) Aber die, die mit IQ arbeiten, sagen, dass IQ messbar ist.
    - Blond ist ein nominales Merkmal. Wenn man sich auf Blondinen Witze bezieht, deutet man Blonde als blöd und somit ist man in einer ordinalen Rangfolge.
  + Aus statistischer Sicht ist das metrische am besten und das nominale am schlechtesten. Oberste Ebene hat am meisten Informationen, die unterste Ebene am wenigsten. Bsp.: Anja ist 1,70 und Tanja 1,80 (Metrisch)

Bsp.: Anja ist kleiner als Tanja, Tanja ist größer als Anja (Komparativ). Wesentliche Informationen verschwinden. Ist Tanja nur 1 cm größer oder 10 cm größer als Anja.

Bsp.: Anja und Tanja haben eine Körpergröße (Nominal)

08.11.2019

**Eindimensionale Häufigkeitsverteilung** (Skript Seite 9)

Beispiele:

* Wohnbevölkerung der BRD nach Familienstand
  + Aufgegliedert in Merkmalsausprägungen (Bsp. Alter 28)
  + Familienstand = Klassifikatorisch
  + Tabelle ist Eindimensional, weil nur ein Merkmal untersucht wird.
  + Überschrift aus welcher hervorgeht um was es geht „Was Sache ist“ muss vorhanden sein. Sache, wo und ein zeitliches wann 🡪 *Wohnbevölkerung der BRD am 31.12.1986*
  + Wichtig ist die Quelle, aus welcher die Daten stammen. (Muss bei jeder Tabelle vorhanden sein, ansonsten ist diese Tabelle unseriös).
  + (Links) Merkmalsausprägung (hier: ledig, verheiratet etc.)
  + (Mitte) Absolute Häufigkeit
  + (Rechts) Relative Häufigkeit 🡪 39,5 % von 100 % sind ledig [24172 \* 100 / 61137]
    - „Vergleich“ ist entscheidend
    - Die Frage „was bedeutet die Zahl für mich“ und wie ist deren Interpretation.
  + Absolute Zahlen sind für die Feinanalyse nicht die geeigneten Daten. Es wäre sehr mühsam dort Informationen herauszulesen. Erst die relativen Daten ergeben genauere Kenntnisse über die Daten.
  + 2014 werden zwei Merkmalsausprägungen zusammengefasst. Es können hier keine genauen Daten ausgegeben werden, wie viele Menschen sich scheiden lassen.
  + Phase 1: Was verändert sich (in einem Zeitvergleich)
  + Phase 2: Die Frage, warum ist das so. Was ist die Ursache.
  + Häufig sin dies Momentaufnahmen. Das bedeutet, aktuell stimmen die Zahlen so, aber was ist, wenn ein paar Wochen später alle heiraten würden ?
  + In 30 Jahren Entwicklung, ist die Zahl der ledigen kaum verändert. 1,3% sind kaum erwähnenswert.
    - In diesen 30 Jahren Entwicklung, ist ganz wenig Bewegung vorhanden. Scheinbar gibt es kaum einen Wandel in Deutschland. Diese Zahlen würden in einem extremen Gegensatz zu Aussagen vom Gesellschaftlichen Wandel stehen.
  + Entscheidend: Die Geburtenzahl ging offensichtlich zurück, daher gibt es mehr ledige Personen (auch Kinder werden hier als ledig dazugezählt). 🡪 Demographischer Wandel. Ein Kind kann nur den Familienstand „ledig“ haben, da es vom Gesetz her nicht heiraten darf.
  + Je mehr Daten, desto mehr unterschiede können auffallen.
  + Die Daten in diesem Beispiel wird suggeriert, dass es keinen Wandel in unserer Gesellschaft gibt. Es werden wesentliche Inhalte verschleiert.
* Erhebung von Privathaushalte der BRD am 30.04.1986 nach Personenzahl
  + Was hat dies für ein Merkmal: Metrisch

Häufigkeitssummenfunktion und empirische Verteilungsfunktion

* Kinderzahl = Metrisches Merkmal
* 4 Spalte, Häufigkeitssummenfunktion = Zahlen werden addiert (32732 + 4248 = 36980)
* 5 Spalte, empirische Verteilungsfunktion

**Problem Relativer Häufigkeit**

* Jeder kann mit diesen Daten etwas anfangen.
* Leute werden an der „Nase herumgeführt“
* Relative Häufigkeit ist ohne die Absolute Häufigkeit nichts aussagend!
* Beispiel- Wursthysterie
* Beispiel- Denkzettel für Amtsinhaber Dieter Salomon
  + Jahr 2010: 50,5%
  + Jahr 2018: 31,3% 🡪 verliert augenscheinlich mehr als 19% seiner Stimmen. (Falsch)
  + Genau gesehen hat Herr Salomon 40% der Stimmen verloren.
  + Entscheidend ist hier der Unterschied zwischen Prozent und Prozentpunkte.
* Relative Häufigkeiten sind oft fehlinterpretiert. Es ist wichtig hier, immer die Absoluten Häufigkeiten dazugeben zu lassen.

15.11.2019

Studie der Barmer:

* Keine Absolute Zahlen
* Die Zahl 190.000 ist nicht ein zu ordnen, da keine Vergleichszahl vorhanden ist.
* Frage, sind dies nur Barmer versicherte ?
* Die Aussage „Zahl hat sich verdoppelt“ woher sind diese Infos? Zahlenstamm?
* Altersspanne uneindeutig.
* Die Diagnose Depression wurde 2005 anders ausgelegt als heute. Diagnoseverfahren hat sich verändert.
* Denkbar wäre, dass nicht die Depressionen zunehmen sondern der Stress in der Schule und Familie (Druck Leistung zu erbringen)
* Hypothese, Eltern schicken ihre Kinder schneller zum Arzt um eine Diagnose zu erhalten. Pädagogisch/Präventiv wird nicht interveniert.
  + Bsp. ADS/ADHS Diagnosen, werden schneller gestellt und dementsprechende Medikamente verteilt

Seite 12 (Skript)

* Metrisch – Zeit wird angeschaut.
* Ungenaue Methode, da hier Stress durch die Mitarbeiter zu verfälschten Ergebnissen führen kann.
* Man kann auch nur etwas messen, wenn man es vergleichen kann.
* Daten müssen aufbereitet (sortiert) werden.
* Sortiert, von der kleinsten bis zur größten Zeit
* Sortiert, wie häufig eine Zeit erreicht wurde.
* Mehr Struktur muss in die Aufbereitung eingebaut werden.
* Klassierte Werte, mehrere Merkmale werden untergebracht.
  + Im Beispiel eine Klassenbreite von 5
* Bei der absoluten Häufigkeit entsteht eine Struktur.
  + In welchem Zeitintervall die meisten Personen liegen.
  + Gaußsche Normalverteilung. (Beispiel: Intelligenz)
* Zahlen so detailliert wie möglich darstellen und zeitgleich so viel Struktur wie möglich.
* Intervalle müssen immer gleich breit sein [Intervall-Breite]
* Am Besten bei einem Punkt beginnen, welcher nicht existiert. Im Beispiel hier 27,5
* Es darf keine Überlappung der Klassen geben. Im Beispiel gibt es zwei gleiche Werte (32.5)
* Nomenklatur „von…bis unter“ = von **27,5 bis unter 32,5**

**Drei Qualitätskriterien**

* Objektivität
  + Der Untersuchende spielt keine Rolle (kann jederzeit ausgetauscht werden)
  + Wer untersucht?
* Reliabilität (Zuverlässigkeit)
  + Das was gemessen wir, muss vergleichbar sein zu anderen Untersuchung (unter gleichen Bedingungen)
  + Bsp. Fragebogen, jeder erhält dieselbe Frage.
  + Bei einer gleichen Frage, kann man die Antworten vergleichen (Dies ist bei unterschiedlichen Fragen nicht möglich)
  + Zum selben Sachverhalt kann Stellung genommen werden.
* Validität
  + Wert oder Gültigkeit
  + Messe ich mit dem was ich Frage überhaupt das, was ich messen möchte.

*Beispiel für die drei Begriffe:*

*Raumtemperatur wird gemessen:*

* *Objektiv ist die Raumtemperatur angenommen 20°. Objektiv gesehen bedeutet, wenn jeder Student im Raum ein Thermometer nimmt und dies überprüft. Jeder muss das selbe heraus bekommen.*
* *Reliabilität bedeutet, dass das Thermometer die 20° auch misst. Es dürfen keine unterschiedliche Daten herauskommen. Sollte dies nicht der Fall sein, ist es unzuverlässig.*
* *Validität wäre hier, dass das Thermometer eine Gültigkeit hat. Wenn jemand mit seiner Uhr die Temperatur misst, wäre das Ergebnis nicht gültig.*

29.11.2019

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **X**  **Von bis unter** | **n(xi)** | **f(xi)** | **S(xi)** | **F(xi)** | **KM** |
| 155 – 160 | 6 | 0,17 | 6 | 0,17 | 157,5 |
| 160 – 165 | 6 | 0,17 | 12 | 0,34 | 162,5 |
| 165 – 170 | 9 | 0,25 | 21 | 0,59 | 167,5 |
| 170 – 175 | 11 | 0,31 | 32 | 0,90 | 172,5 |
| 175 – 180 | 3 | 0,08 | 35 | 0,98 | 177,5 |
| 180 – 185 | 1 | 0,02 | 36 | 1,0 | 182,5 |
|  |  |  |  |  |  |
|  | n = 36 |  |  |  |  |

Median = n/2 = 36/2 = 18

Z = 165 + (0,5 – 0,34)\*(5/0,25)

ZKörper = 168,2 cm

Median, an welcher Stelle er liegt und welchen Wert dieser hat.

Skript Seite 21, in dieser Tabelle (Intelligenz) liegt der Median bei n/2, also bei 14,5 (29 /2 ). Welcher IQ wäre dies ? = 100

14,5 liegt zwischen 13 und 20 (bei der kumulierten Häufigkeit). Das bedeutet, jeder der bis zur 13ten Stelle ist hat den IQ 90, alles was über 13 ist hat den IQ 100

Skript Seite 23: Intervall 10.

Durchschnittsintelligenz berechnen: 109,8

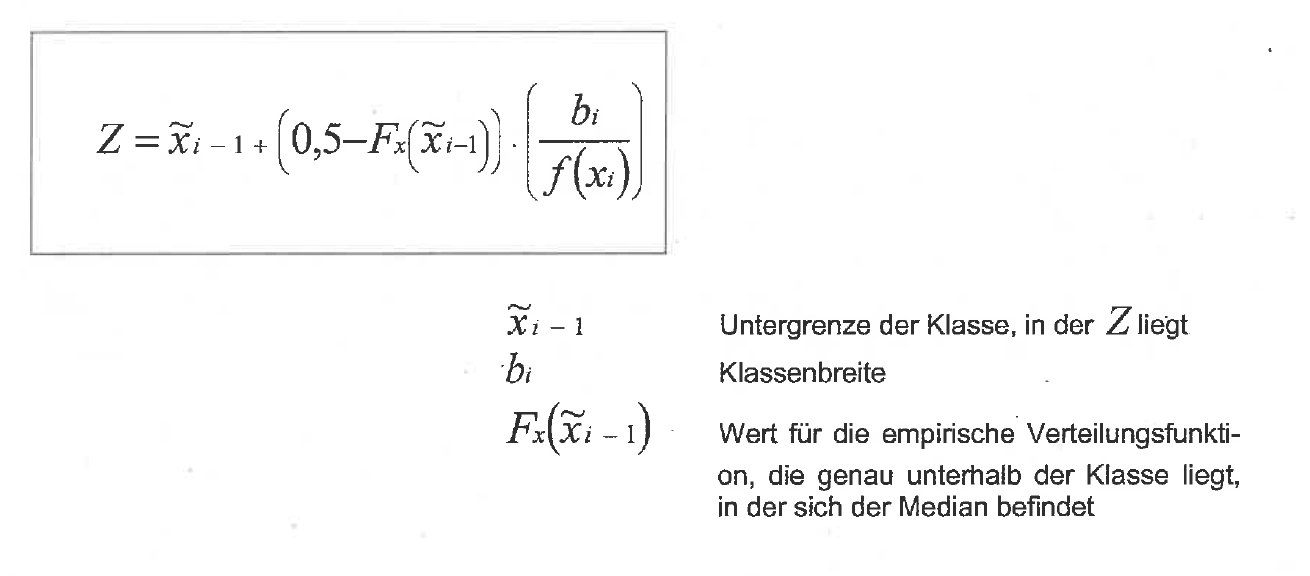
Medianstelle: n/2 = 225/2 = 112,5 🡪 IQ 100 – 110 ist der Median (113)

Median Formel auf der Seite 23, funktioniert nur wenn wir den Median kennen.

F(x) Wert, der eine Stufe unter dem Median liegt

Klassenbreite = 10

f(xi) ist die Zeile wo der Median steht



Z = 100 + (0,5 – 0,280) \* (10 / 0,223)

ZIQ = 109,9 Das bedeutet der IQ ist normal verteilt!

**Die Spannweite**

Ist die Differenz zwischen dem höchsten und niedrigsten Wert. Bsp. Alter, der jüngste ist 20 und der älteste 36 = 36 – 20 = 16 –> Streuungsmaß

Menschen welche nun knapp über 20 sind, können nicht genau beschrieben werden, da die Spannweite zu groß ist bis zum ältesten.

Totale Homogene Verteilung, wären alle gleich alt!

34 sind 25, 1 ist 20, 1 ist 40 = Dies wäre Homogen

80% Grüne, 20% verteilt sich = relativ homogen

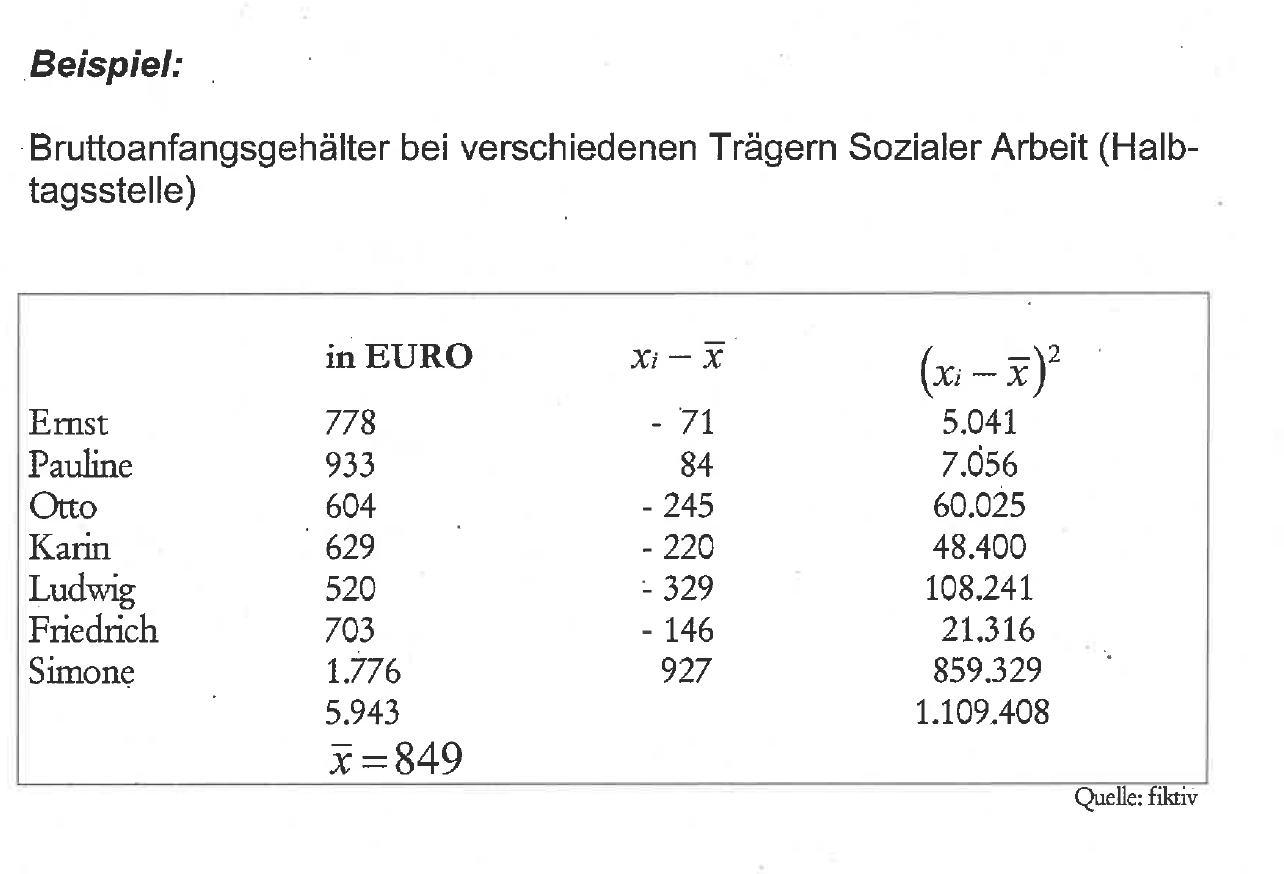
Sehr viele sind um 20 Jahre, gleich viele um 40 rum = nicht homogen

Je mehr Menschen sich um einen Durchschnittswert herum eingeordnet sind (oder gar durchschnitt) sind, desto homogener. Die Streuung nahezu 0. Wenn alle 25 sind, ist die Streuung 0.

Je mehr verteilt desto heterogen ist die Verteilung.

Wann geht die Homogenität in die Heterogenität über ?

Bsp.



Spannweite wäre 1776 – 520 = 1256 €

Arithmetisches Mittel = 849

Gemeinsamer Bezugspunkt ist 849 €

In Spalte 3 sind die Differenzen… Ernst hat 71 € weniger… usw.

Im Beispiel eine sehr große Streuung.

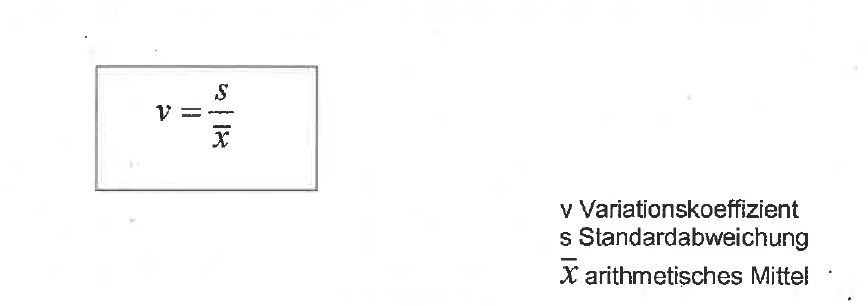
Nimmt man zwei heraus, wird das Arithmetische Mittel kleiner und die Streuung kleiner (kleineres Streuungsmaß)

Summe der quadrierte Abweichungen vom Arithmetischen Mittel durch sieben (Anzahl der TN) teilen erhält man die **Varianz** = 158.468,9 €. Varianz ist unpraktikabel, da diese sehr schlecht zu bewerten ist.

Um dies besser darzustellen zieht man die Wurzel aus der Varianz = 398,10 €. Was bedeutet diese Zahl = Durchschnittliche Abweichung vom Arithmetische Mittel. Alle Sieben weichen im Durchschnitt von 398,10 vom Arithmetischen Mittel ab.

Ist eine gemeinsame Durchschnittliche Abweichung von 398 € vom Durchschnittsgehalt von 849 € viel ? Streuung ist in diesem Fall sehr hoch, aber es ist nicht mal die Hälfte vom Durchschnittsgehalt. Festlegung von viel oder wenig in dem Fall wichtig.

Alle Durchschnittliche Abweichungen, welche nicht höher sind als die Hälfte zeigen noch eine Art von Homogenität an und alles was drüber liegt wäre heterogen.



V = 398,10 / 849  
V = 0,47

Standardabweichung = Wurzel aus der Varianz

Bsp.: Auf die Größe

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **X**  **Von bis unter** | **n(xi)** | **f(xi)** | **S(xi)** | **F(xi)** | **KM** | **xi-x** | **(xi – X)2** | **(xi – X)2 \* n(xi)** |
| 155 – 160 | 6 | 0,17 | 6 | 0,17 | 157,5 | -10,3 | 106,09 | 631,5 |
| 160 – 165 | 6 | 0,17 | 12 | 0,34 | 162,5 | -5,3 | 28,09 | 168,5 |
| 165 – 170 | 9 | 0,25 | 21 | 0,59 | 167,5 | -0,3 | 0,09 | 0,9 |
| 170 – 175 | 11 | 0,31 | 32 | 0,90 | 172,5 | 4,7 | 22,09 | 242,9 |
| 175 – 180 | 3 | 0,08 | 35 | 0,98 | 177,5 | 9,7 | 94,09 | 282,3 |
| 180 – 185 | 1 | 0,02 | 36 | 1,0 | 182,5 | 14,7 | 216,09 | 216,1 |
|  |  |  |  |  |  |  |  |  |
|  | n = 36 |  |  |  |  |  | 466,54 | 1547,6 |

(x)Arithmetisches Mittel 167,8

157, 5 – 167,8 =

Varianz = 466,54 / 36 = 12,96

Standardabweichung = Wurzen Varianz = 3,6

V = 3,6 / 167,8

V = 0,021 🡪 homogen, da unter 0,5 (V = 0,04 Richtige Lösung)

06.12.2019

Standardabweichung gehört zu den Streuungswerte und ist die durchschnittliche Abweichung vom Durchschnitt (Mittelwert).  
Varianz entspricht dem Vorergebnis der Standardabweichung, meist sehr große Zahlen (wegen dem quadrieren).

* Wenn alle Personen gleich viel verdienen 🡪 jeder ist durchschnittlich, Abweichung ist dann 0
* Wenn alle gleich viel verdienen etc. entspricht dies total homogen 🡺 Standardabweichung bei 0. Solange durchschnittliche Abweichung vom durchschnitt nicht größer als 0,5, dann noch homogen.
  + Kleiner, gleich 0,5 homogen (oder ½)
  + Größer, gleich 0,5 heterogen (oder ½)
* Wenn alle das selbe machen = Abweichung vorhanden!
  + Wenn zwei Personen selbe Daten:
    - Keine Objektive Interpretation! (keine falschen und richtigen)

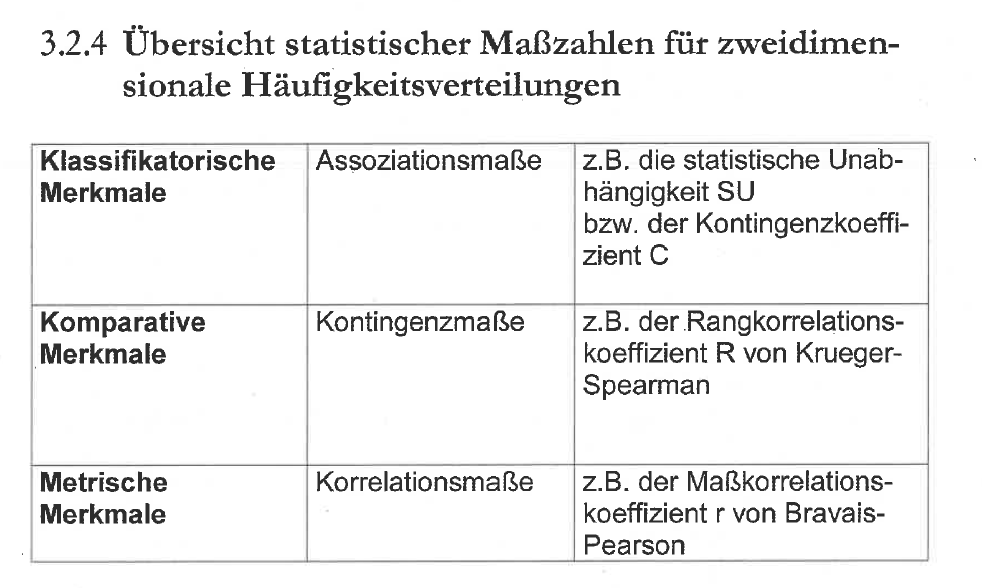
Skript: Lebenswelten junger Muslime in Deutschland

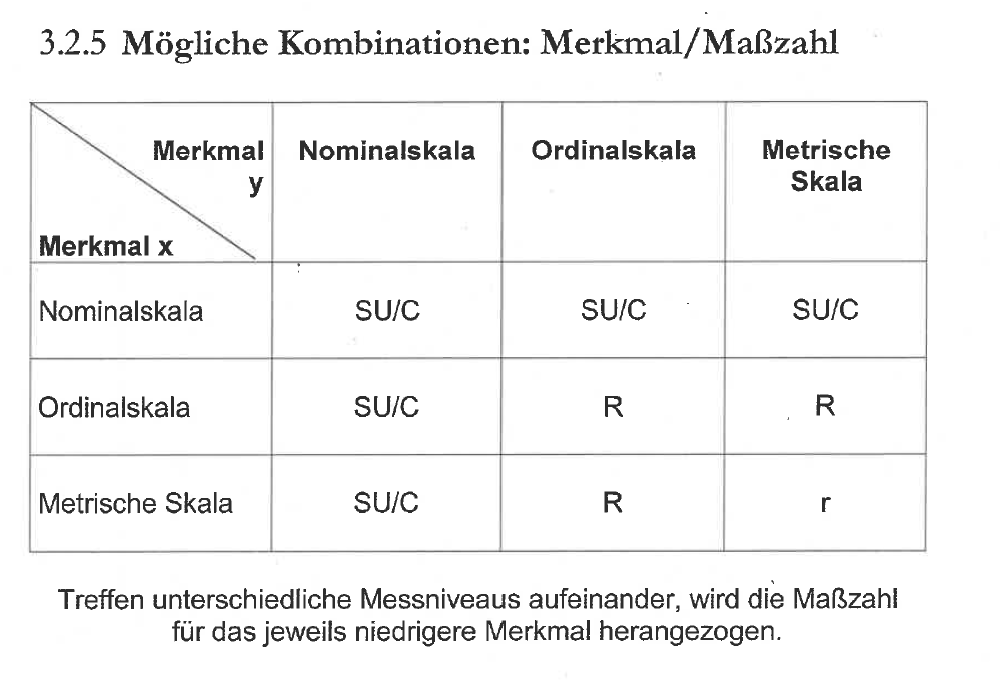
* Wäre bei dem Punkt „Macht“ in unterschiedliche Bereiche aufgeteilt worden, wären die Ergebnisse vermutlich deutlicher auseinander gegangen. Da ggf. manchen Geld wichtiger ist als Soz. Status etc. oder muslimischen Männern Macht wichtiger wäre als deutschen Männern etc…
* Variationskoeffizient = Standardabweichung/arithmetisches Mittel
  + 0,97 / 2,86 = 0,34
  + 1,21 / 3,14 = 0,38 Streuung bei „Macht“ am größten
  + 1,18 / 3,26 = 0,36
  + 0,72 / 4,58 = 0,16
  + 0,52 / 4,84 = 0,11
  + 0,31 / 4,92 = 0,06
  + 0,78 / 4,15 = 0,19
  + 0,88 / 4,46 = 0,2
  + 0,63 / 4,67 = 0,13
  + 1,12 / 3,11 = 0,36
  + 0,99 / 3,94 = 0,25 0,36 am höchsten (größte Abweichung), 0,23 geringste
  + 0,97 / 4,24 = 0,23 Abweichung 🡺 Muslime reagieren bei Tradition ziemlich homogen.

13.12.2019

Zweidimensionale Häufigkeiten (Tabelle auf Seite 33)

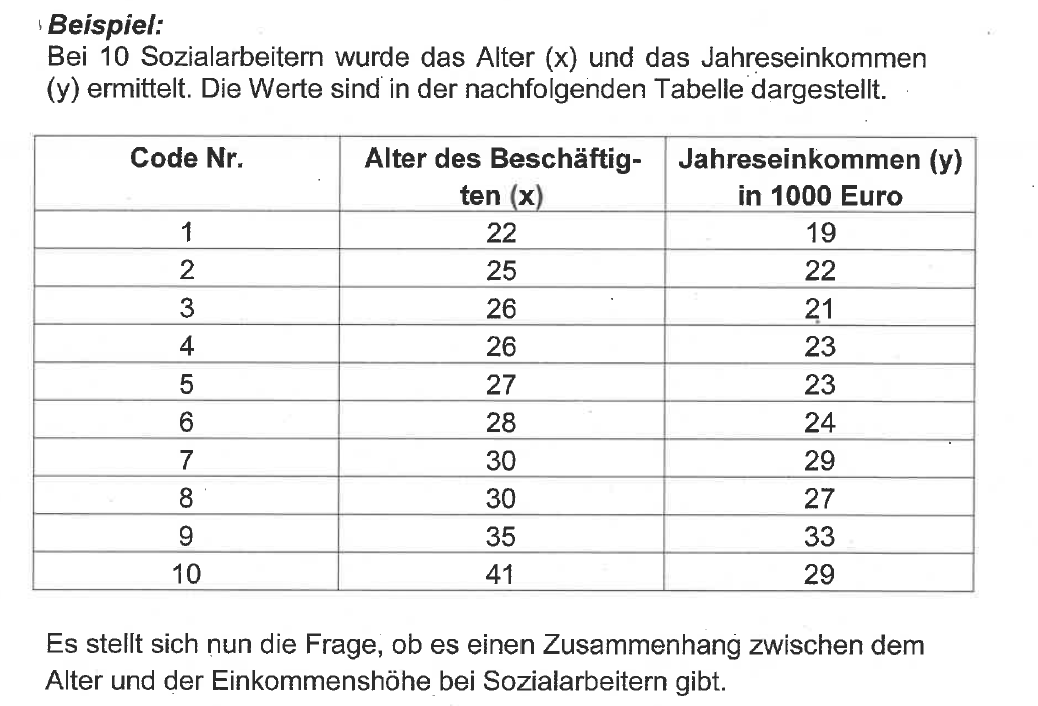
* Männer haben häufig eine größere Schuhgröße und häufig ein größeres Einkommen! (Vermutung und Aussage)
* Kreuztabelle oder Kontingenz
* Einzelne Merkmale sind miteinander verbunden (Schuhgröße und Einkommen). Dies nennt man gemeinsame Häufigkeit
* Vertikale und horizontale Schreibweise als Merkmalsausprägung.
* Gemeinsame Häufigkeit im inneren (schwarz umrandet) (Seite 34)
  + Es gibt niemanden mit der Schuhgröße 46 und dem Einkommen von 1000€
  + Es gibt zwei Personen mit der Schuhgröße 46 und dem Einkommen von 10.000 €
* Auch möglich die Anzahl der Schuhgrößen abzulesen ohne das Einkommen
* Es gibt einen signifikanten Zusammenhang zwischen Schuhgröße und Einkommen
* Gibt es einen Kausalen Zusammenhang zwischen Schuhgröße und Einkommen
* Gesunder Menschenverstand sagt: NEIN! Diesen Zusammenhang gibt es nicht.
* **Scheinkorrelation!**
* Ist die Scheinkorrelation Unsinn oder gibt es indirekte Verbindungen dieser Merkmale?
* Scheinkorrelationen sind häufig hinweise auf eine dritte Variable (intervenierende Variable)
* Bei der Schuhgröße ist die intervenierende Variable das **GESCHLECHT**
* Mit dieser dritten Variablen, macht das Beispiel, Schuhgröße und Einkommen wieder Sinn!
  + Es gibt einen rechnerischen Zusammenhang, allerdings ist dieser oft unsinnig (kausal)
  + Es gibt viele Fälle, in welchen die Menschen dies geglaubt haben (hier Schuhgröße ist kein Kausal Zusammenhang, dies ist nur rechnerisch interessant)
  + In der Medizin z.B. Kaffeekonsum in Verbindung mit Krebserkrankungen (Scheinkorrelation).
    - Weitere Variable angeschaut, Kaffeetrinker rauchen häufig in ihren Pausen.
  + Zusammenhang von zwei Merkmalen muss gefunden werden. Diese Maßzahl kann rechnerisch bestimmt werden und zeigt dann ob es einen Zusammenhang gibt oder nicht. Korrelation Koeffizient!
  + Bsp.: In DE gibt es einen Zusammenhang zwischen Alter und Einkommen bei abhängig Beschäftigten (Arbeitnehmerverhältnis). Dieser ist sehr hoch!
    - Öffentlicher Dienst, alle Jahre erhöht sich das Einkommen gebunden an das Alter. (Zusammenhang = 1)
    - Tarifverträge (sind oft angelegt an das Alter)
      * Betriebszugehörigkeit
      * Einzahlungsdauer in das Steuersystem
    - Beides steigt, Einkommen und Alter (positiver Zusammenhang)
    - Negativer Zusammenhang (-1), ein Merkmal steigt, das andere fällt. (Je älter, desto weniger Einkommen) Bsp. in Amerika! bei harter körperlicher Arbeit
    - *„Provokantes Beispiel: Je mehr Kinder, desto geringer der Bildungsstand oder andersherum, je mehr Bildung desto weniger Kinder“*

wichtigste!



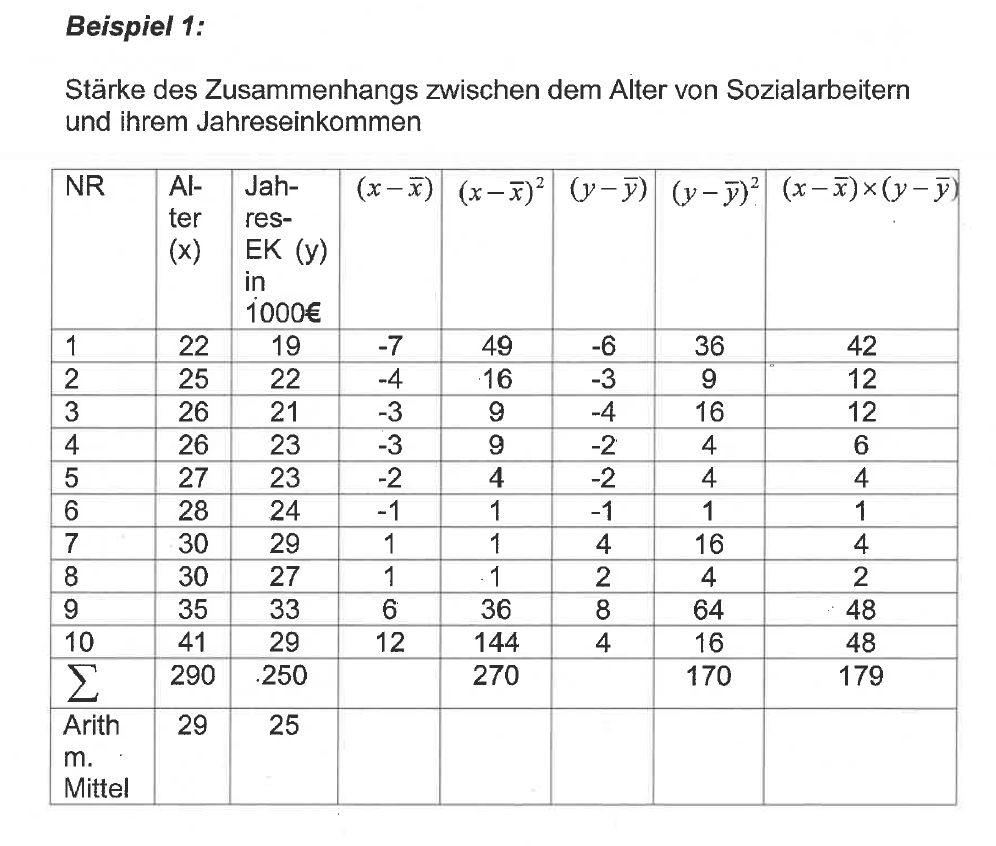
Bsp. Trifft ein metrisches auf ein Rang, muss der Rangkorrelationskoeffizient genommen werden!

**Rechnung**:

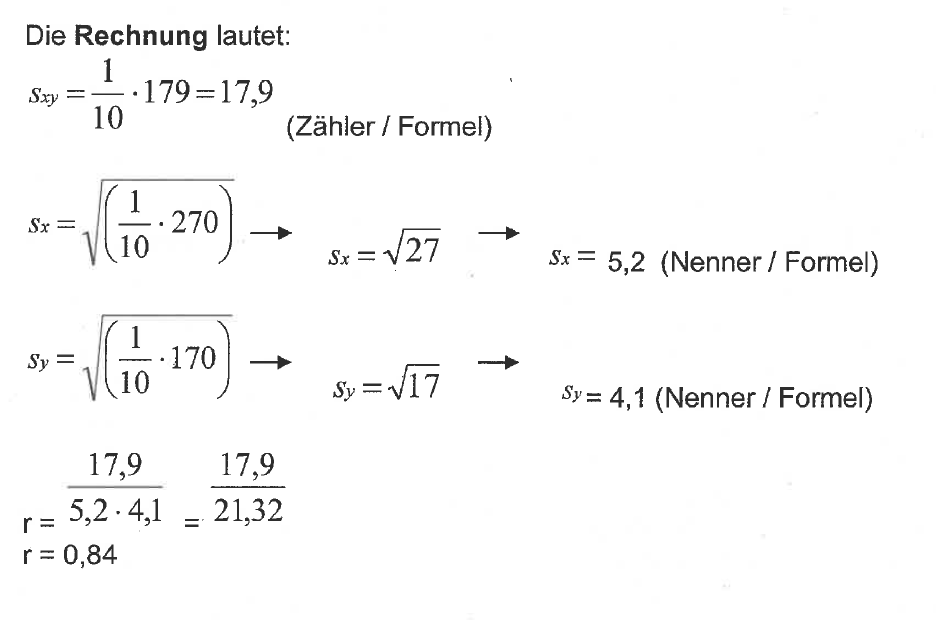


(Tendenz je älter desto mehr Einkommen)

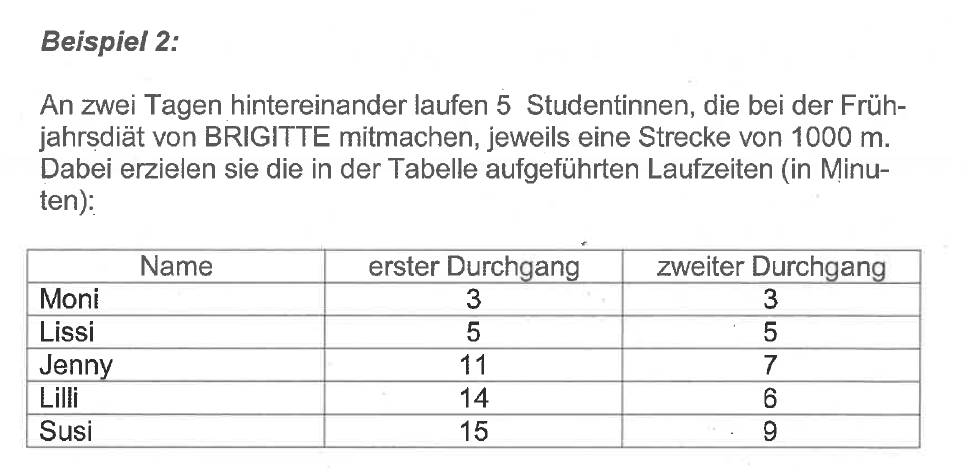
* „Korr-Koeff.“ bewegen sich zwischen 0 und 1 und sie können positiv oder negativ sein
* 1. Schritt Arithmetisches Mittel Berechnen
* 2. Schritt Standardabweichungen Berechnen (Für beide Merkmale)
* 3. Schritt Ko-Varianz. Differenzen für X und Y werden miteinander Multipliziert.



* Was müsste man mit den 270 machen um zur Standardabweichung zu kommen ?
  + Durch 10 teilen

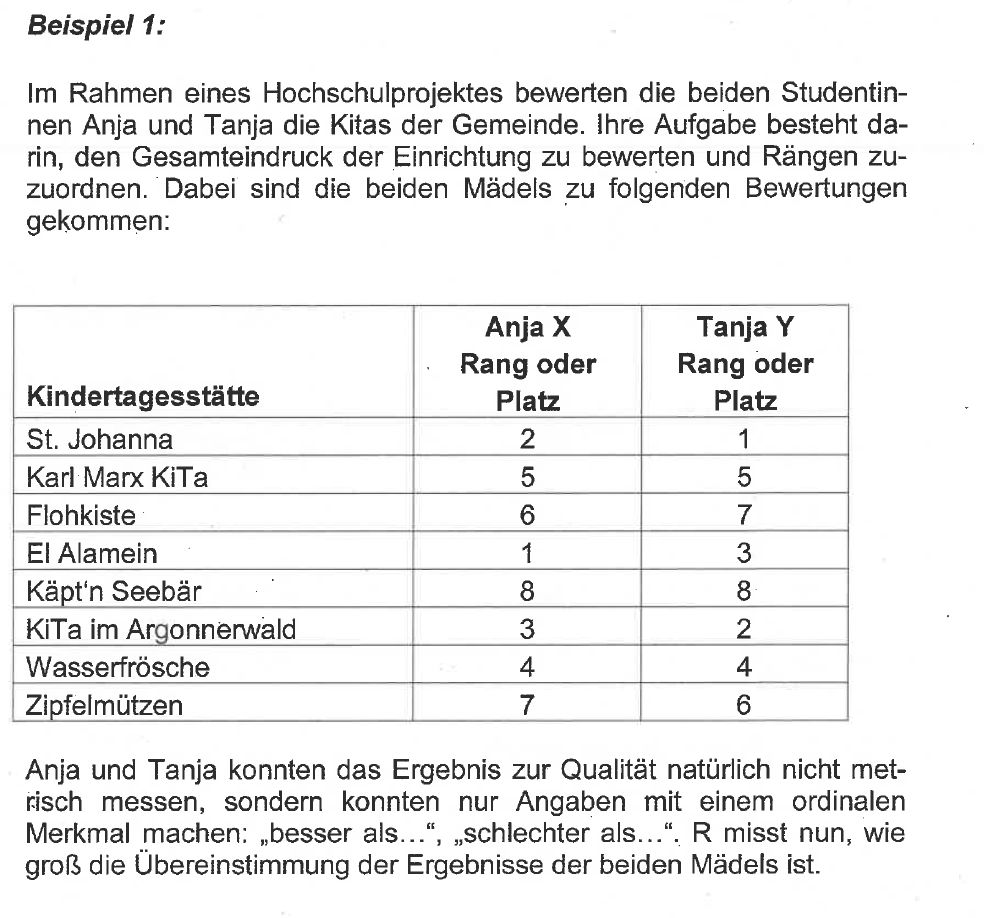


* r = 0,84 bedeutet das die Prognose der Wahrheit entspricht = Alter und Einkommen!
* Mit steigendem Alter steigt das Einkommen, da die Zahl knapp bei 1 ist läuft diese nicht parallel (also Alter und Einkommen steigen gleich) sondern hat einige kleinere Abweichungen. 0,84 ist ein hochpositiver Zusammenhang, je älter man ist, desto mehr verdient man.
* Wenn zwei Merkmale mit einander korrelieren, dann geht daraus nicht hervor wie das Kausalverhältnis zwischen den beiden Merkmalen ist.
* Bei zwei Variablen ist immer eine die UV und eine die AV.
  + Alter = UV (Alter wirkt sich auf das Einkommen aus)
  + Einkommen = AV (Einkommen wirkt nicht auf das Alter aus)
  + Das Alter kann auf das Einkommen wirken, aber das Einkommen wirkt sich nicht auf das Alter aus!
* UV und AV müssen berücksichtigt werden, damit man bei der Interpretation keinen Fehler macht.

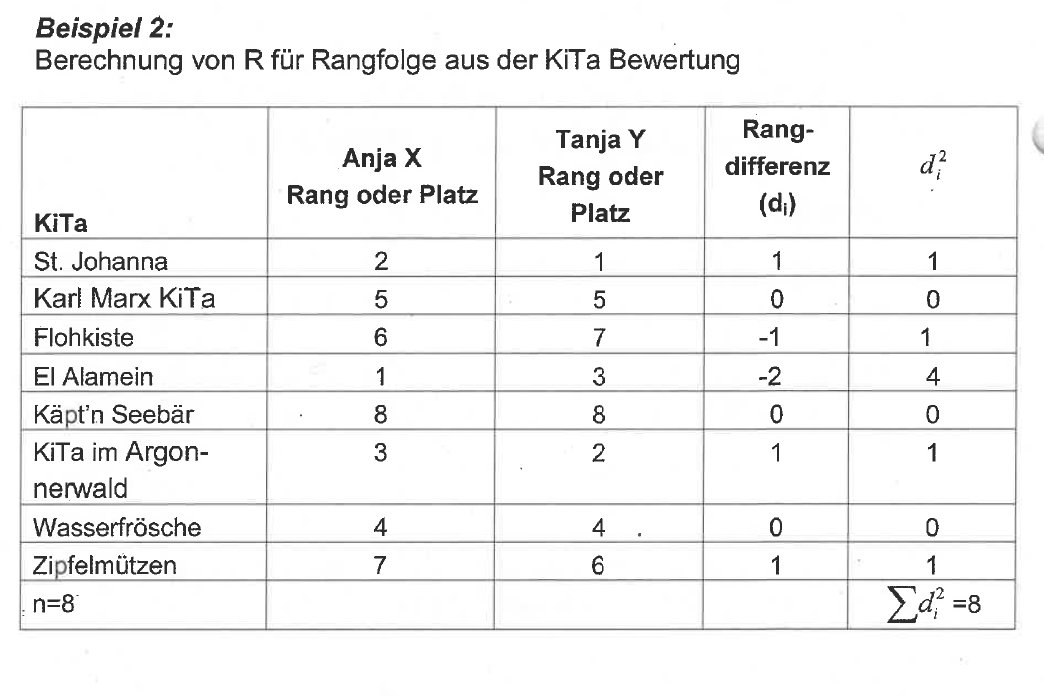


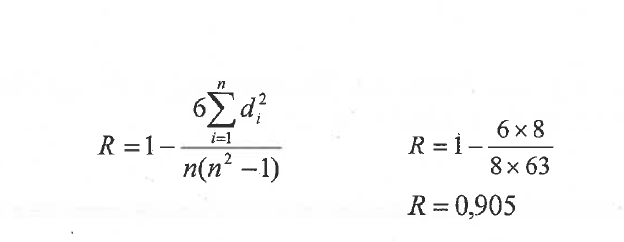
* r = 0,875 (Lösung) (Metrische Merkmale)
* Hoher Positiver Zusammenhang, die Laufzeiten verbessern sich
* Verbesserung vom 1. auf den 2. Durchgang

**Rangkorrelationskoeffizient (R)**:



* Sinn dieser Untersuchungen:
  + Wie Subjektiv/Objektiv die Meinung der beiden Personen ist.
* Rangdifferenz wird berechnet.
  + 2-1
  + 5-5
  + …
* Differenzen werden quadriert d2i



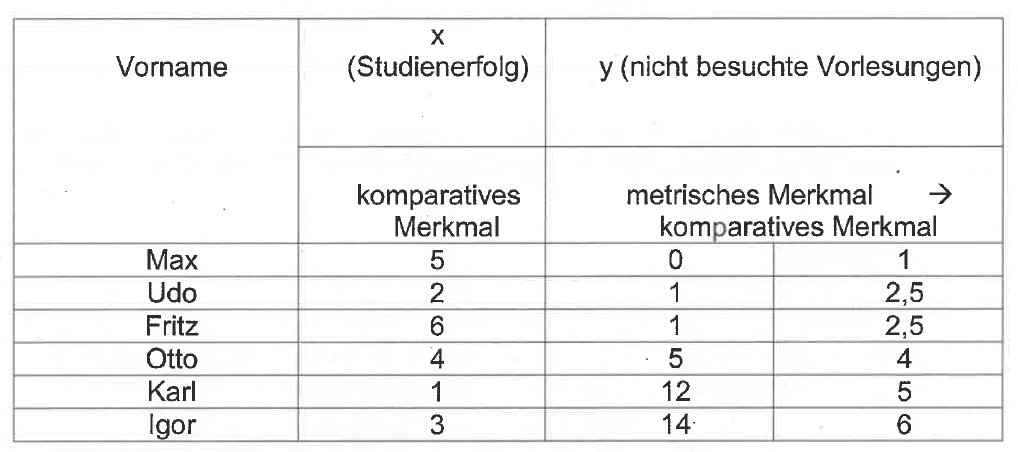


* Hoher Zusammenhang zwischen den beiden Beurteilern.

20.12.2019

**Rangkorrelationskoeffizient**

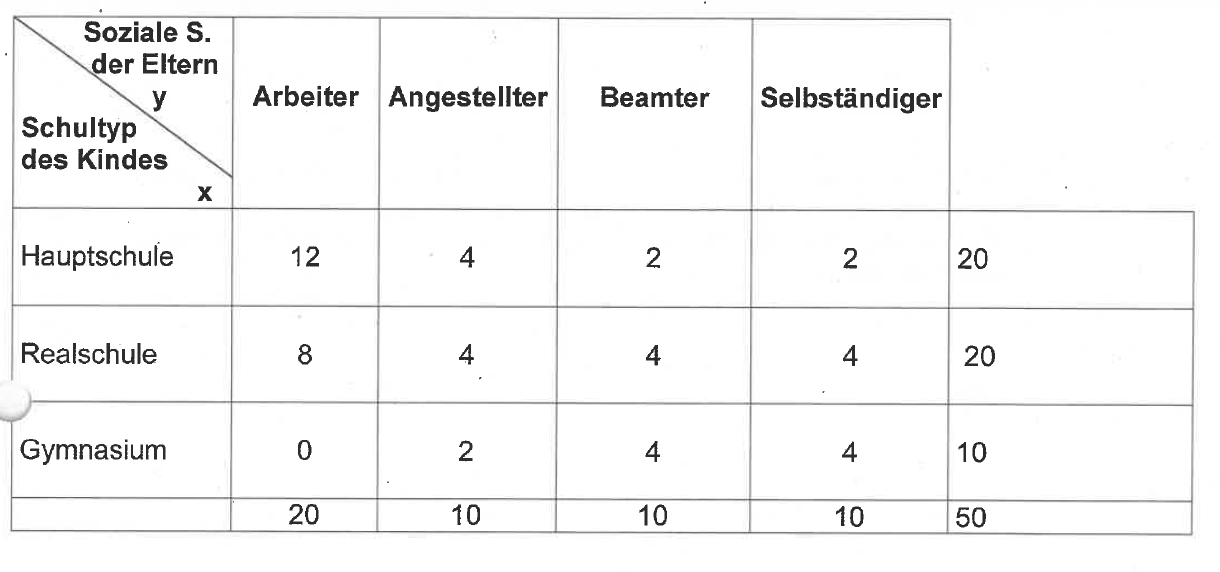
* Rangmerkmal und ein Metrisches Merkmal zustande kommt.
* Metrischen Werte müssen in Rangmerkmale umgewandelt werden



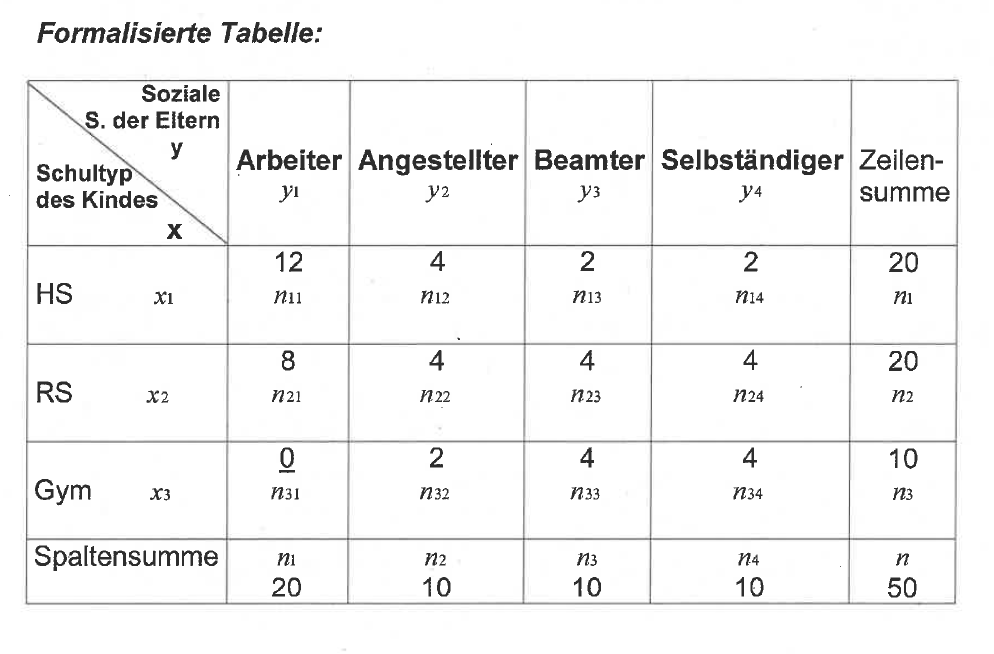
* Umformulierung der metrischen Merkmale in Ränge.
  + Max hat 0 mal gefehlt, dadurch erhält er Rang 1
  + Igor hat 14 mal gefehlt, dadurch erhält Rang 6
  + Platz, 2 ist doppelt vergeben, daher wird Platz 2 und 3 zusammengezählt, dies ergibt
  + 5 und das arith. Mittel ist dann 2,5. Wichtig ist hier, dass dann mit dem Rang 4 weiter gemacht wird.
* R = -0,53 🡪 um so öfters die Studenten fehlen, desto besser sind sie. 🡪 Mittlerer negativer Zusammenhang!

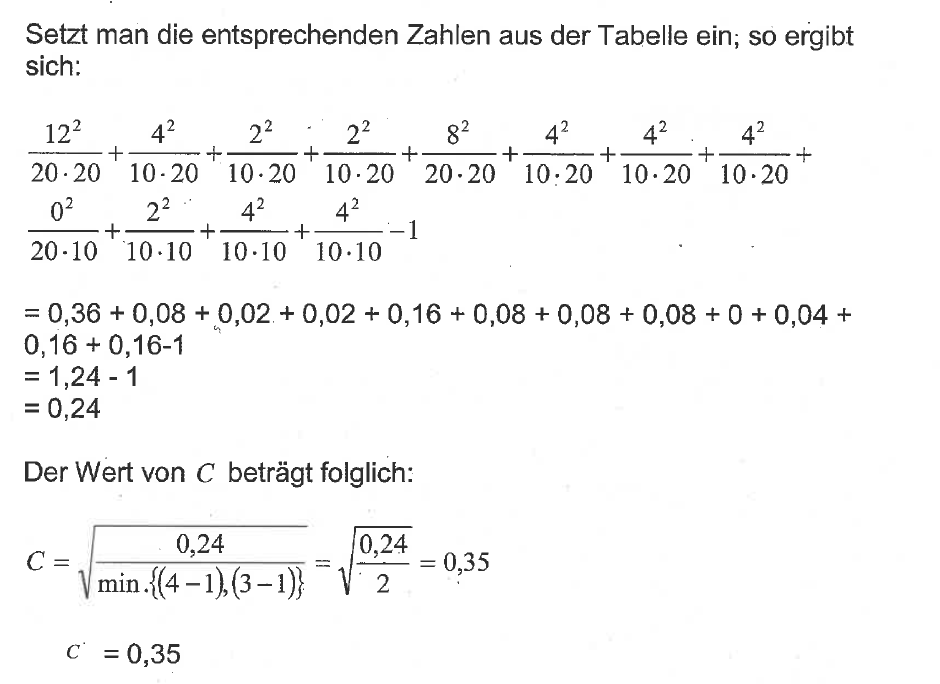


**Kontingenzkoeffizient**



* n=50, also 50 Fälle
* 20 Kinder kommen aus Arbeiterhaushalte… usw.
* 20 Kinder gehen auf die Hauptschule … usw.
* Je höher der berufliche Status der Eltern, desto höher die schulische Bildung.
* Aus der Arbeitergruppe ist kein Kind auf dem Gymnasium
* Wie müsste die Tabelle aussehen, damit man nicht sagen kann, dass die Soziale Herkunft und die Schulbildung aussagekräftig sind.
* Von 50 Kinder, gehen 10 auf das Gymnasium.
* Ordinales Merkmal bei der Einteilung der Schulen.





* 0,24 🡪 ist die Summe der Brüche Minus
* Woher kommt die 4 und woher die 3
  + 4 Kommt von den 4 Merkmalsausprägungen (Arbeiter, Selbständiger etc.)
  + 3 Kommt von den drei Schularten.
* C = 0,35 🡪 Zusammenhang ist nicht so groß, eher schwach. Soziale Schicht hat einen schwachen Zusammenhang zur Schulart.
  + Bildungspolitik ist super, es gibt keine Ungleichbehandlung. Dies sagt dieser Koeffizient aus. (Tabelle wurde nicht durch Erhebung erstellt! Nicht realitätsnah!)
* Korrelationskoeffizient kann in die Irre führen, hier in diesem Beispiel verdeckt dieser, dass die Arbeiterpartei ein Bildungsproblem hat. (Keine Arbeiterkinder sind auf dem Gymnasium). C betrachtet stets die ganze Tabelle, dadurch können schwächere Parteien herausfallen.
* Bei C stets aufpassen was dahinter stecken könnte. Immer die Ursprungszahlen anschauen und mitdenken.

Beispiel Franzosen arbeiten erheblich weniger als Deutsche

* Es wird suggeriert, dass Frankreich fauler ist, als andere Nationen.
* Wir wissen nicht, wie viel die Rumänischen Beschäftigten während der Arbeitszeit nicht arbeiten.
* Tabelle hat eine verzerrte Sichtweise, Skalierung ist sehr einseitig.

Arbeitslosen von Mariental

* Fragen werden dazu gestellt.
* Das Buch lesen genügt.
* Begriff: Müden Gemeinschaft 🡪 Ausdruck das die Wirkung der Arbeitslosigkeit, Auswirkungen auf das Gesellschaftliche Leben hat.