# Wirtschaftsstatistik

Dozent: Frau Dr. Merlins

**Konferenz am 13.10.2023**

Konferenzen via Adobe Connect

**Organisatorisches**

Fragen an

elena@merins.de

Termine für die Onlinekonferenzen

03.11.2023 19:00 - 20:30 Uhr

24.11.2023 19:00 - 20:30 Uhr

15.12.2023 19:00 - 20:30 Uhr

Präsenztermine:

20.10.2023 19:30 - 21:00 Uhr

17.11.2023 19:30 - 21:00 Uhr

08.12.2023 19:30 - 21:00 Uhr

12.01.2024 19:30 - 21:00 Uhr

Alle Übungsaufgaben lösen, Voraussetzung, um Klausur zu bestehen

Klausurtermine

Hilfsmittel

1. Formelsammlung
2. Taschenrechner
3. Schreibgerät
4. Lineal Grafen mit Linear zeichnen

Klausur ist schaffbar

* MC ist nicht sicher
* Und Rechenaufgaben

Vlt. 10 Fragen Single Choice (Theorie) 1 Antwort von 4 Fragen

* Zu Begriffen
* Zum Verständnis

**Wahrscheinlichkeitsrechnung ist sehr schwer**

Übern, üben, übern

Was ist Statistik?

Folie 4

Was ist Datengewinnung?

Folie 5

* Datenerhebung
* Primärerherbung nach Vorgaben
* Sekundärerhebung aus bereits vorhandenem Material
* Vollerhebung
* Teilerhebung (Stichprobe) n von N

Deskr. St. Vs. Geplante St.

Alles bis Modul 5 ist beschreibende Statistik

**Induktive Statistik**

Daten beurteilen basierend auf Stichprobe

**Schließende Statistik**

Der Schluss vom Teil auf das Ganze

Einfache Stichprobe

Jede mögliche Stichprobe besitzt dieselbe Chance ausgewählt zu werden

Geschichtete Stichprobe

Grundgesamtheit N wird in Schichten eingeteilt

Klumpenstichprobe

Folie 15

Auswahlen der Stichproben ist sehr wichtig

Willkürliche und bewusste Auswahlen (geschichtete Stichproben)

Phasen eines Statistikverfahren (17)

**5 Ds**

Definition

Was wollen wir

Design

Design-Entscheidung

Erhebungsart

Längsschnitt über größeren Zeitraum

Erhebungstechnik

Budget

Zeit

Tema

Datenerhebung

Vorbereitung der Datenauswertung

Datenauswertung und -analyse

Dateiaufbau

Datenbereinigung

…

Statistik richtig interpretieren, ist klausurrelevant

Dokumentation

Ist extrem wichtig

Folie 22 wichtig

„OPERATIONALISIERUNG“

Folie 25

Histogramm

Diagramm auf Seite 25 ist ein Histogramm

Graphische Darstellung S. 26 für Vorstandsabteilung

**Arten der Softwareprogramme sind nicht klausurrelevant**

Grundbegriffe der Statistik

klausurrelevant

Merkmalsträger

Statistische Masse

Merkmal

Merkmalsausprägung

MERKMALE

SKALENNIVEAU

Wichtig

Nominalskala • Ordinalskala

Es kann nicht mit den Variablen gerechnet werden

Stetige Klassierung

Von 0 Bis unter100 , damit nichts bis zu den Klassengrenzen verloren geht

Relative Häufigkeit nm besten nicht in % darstellen

Übung mit Häufigkeitstabelle (Folie 6)

Wichtig: in der Dokumentation hinweisen, dass die Statistik nur gültige Antworten berücksichtigt, wenn dem so ist

Summenzeile muss zu JEDER Häufigkeitstabelle angegeben werden

Für Summenhäufigkeit gibt es keine Summenzeile

Die Summe in Spalte absolute. Summenhäufigkeit = die Summe aller Häufigkeiten

Häufigkeitsverteilung

Kontrolle Klassenmitte muss tatsächlich in den Klassengrenzen stehen

z. B. 0 bis 2 🡆 Klassenmitte = 10

16

Empirische Verteilungsfunktion

Bei klassierten Daten wir der Punkt zur unteren klassengrenze eingezeichnet

**wichtig bei Diagrammen**

* Achsenbeschriftung
* Skala
* Quellenanagabe
* Diagramm muss
  + anschauliches Bild der Daten liefern
  + das Wesentliche der Verteilung aufzeigen
  + anschaulich und korrekt präsentieren
* Skalierung darf die Präsentation der Daten nicht verfälschen (Manipulation durch optische Täuschung)

**wichtig bei Tabellen**

* Quellenanagabe

**Histogramm muss immer Skala mit geordneten Messwerten (Ordinalwerte) haben!**

Die Breite einer Säule = Klassenbreite

Höhe = klassenhäufigkeit

Häufigkeitstabelle

Summe in der letzten Spalte ist wichtig!!!

Die Summe alle Häufigkeitsverteilungen muss 1 oder 100%sein

Merkmalsausprägung kann mit verschiedenen Variablen i oder j oder m oder … bezeichnnet werden

Folie 9

Summe zu allen absoluten Häufigkeit = der Summe zur letzten absoluten Häufigkeit

**Folie 24**

**Histogramm**

**Wichtig, klausurrelevant**

Graphen ab Seite 26 sind nicht mehr klausurrelevant

**03.11.2023**

Bei Modus muss man nichts rechnen, ist der Beobachtungswert mit der größten Häufigkeit.

XD ist Modus

Achtung: Gibt es kein Beobachtungswert mit mehreren Häufigkeiten (also ist bei jedem Beobachtungswert die Häufigkeit == 1, gibt es auch keinen Modus!)

Median ist Zentralwert daher Xz

Median ist Mitte aus der geordneten Reihe

Ist robustes Maß, weniger anfällig gegen Ausreißer

Bei Median muss die Reihe der Ausprägungen sortiert sein!

Mittelwert ist nicht dasselbe wie Median

Für Ordinalwerte (Rangmerkmale)

Unterscheidung zwischen Anzahl der Merkmalsausprägungen (gerade und ungerade)

Beim Median immer genau 50% links und rechts

Bei ungeraden Werten die überschaubare sind, muss nichts gerechnet werden

Bei gerader Anzahl der Merkmalsausprägungen wird Median aus den beiden Werten in der Mitte berechnet

Bei Bestimmung des Median mit geraden Zahlen muss der Median Sinn ergeben (Folie 11)

z. B. Zeugnisnote

bei metrischen Daten kein exakte M

Folie 12

Folie 14

0,5 sind 50% in der y-Achse

Folie 15

Immer mit Dezimalwerten nicht mit Prozentwerten rechnen (auch wenn Tabelle Prozentwerte hat)

Arithmetisches Mittel, Mittelwert

Mittelwert ist kein stabiler Wert in der Mitte, kann auch etwas neben der Mitte liegen (Folie 16)

Bei Mittelwertberechnung zu absoluten Häufigkeiten muss die Anzahl n berücksichtigt werden

Arithmetisches Mittel Nicht verwechseln mit median!

Folie 24

Bei Klassierten Daten auch nur Näherungswert

Statt der Ausprägung wird die Klassenmitte verwendet

Modus ist einfach, ist die am häufigsten Vorkommen der absoluten Häufigkeiten Folie 333

Median immer auf einer nach der Größe der Beobachtungsdaten (geordnete) sortierte Reihenfolge verwenden (xi muss sortiert sein!)

Median ist ein Q2 Quantile Folie 35 (Quartile)

Q1 und Q3 Quantile werden nach derselben Methodik (Algorithmus) wie Median berechnet, weil Median ein Q2 Quantil ist

Q2 Quartil bei unklassierten Daten

**Bei gerader Summe der absoluten Häufigkeiten** (gerade Anzahl der Ausprägungen) bzw. gerader Anzahl der Häufigkeiten nach Berechnung des Q2 Quartils

Q2 = (xn/2 + xn/2+ 1 ) / 2

Q1 = (xxn/22 + xxn/22+1) / 2

Q3 = (xxn/22 \* 3 + xxn/22 \* 3 +1) / 2

**Bei ungerader Summe der absoluten Häufigkeiten** (ungerade Anzahl der Ausprägungen) bzw. ungerader Anzahl der Häufigkeiten nach Berechnung des Q2 Quartils

Q2 = x(n+1) / 2

Q1 = x(n+1) / 2

Q3 = xn/2 + x(n+1) / 2

Bsp.

Anzahl der Häufigkeiten = 20

Q2 = (x20/2 + x20/2+ 1 ) / 2 = (x10 + x11 ) / 2

Q1 = (x20/4 + x20/4+ 1 ) / 2 = (x5 + x6 ) / 2

Q3 = (x20/4 \* 3 + x20/4 \* 3 + 1 ) / 2 = (x15 + x16 ) / 2

Anzahl der Häufigkeiten = 22

Q2 = (x22/2 + x22/2+ 1 ) / 2 = (x11 + x12 ) / 2

Q1 = x(11+ 1) / 2 = x6

Q3 = x22/2 + (11+ 1) / 2 = x11 + 6 = x17

Bei klassierten Daten

Q2 = xk-1 + (xk - xk-1) \* (0,50 - Fk-1) / fk

Q1 = xk-1 + (xk - xk-1) \* (0,25 - Fk-1) / fk

Q3 = xk-1 + (xk - xk-1) \* (0,75 - Fk-1) / fk

MODUL 5: STREUUNGSPARAMETER

Folie 1

Lageparameter können nicht alles korrekt beschreiben

Folie 4

Noten der Mädchen streuen weniger als Noten der Jungs

Für die Beschreibung der Streuung wird Bezugspunkt benötigt wird

Folie 15

Varianz ist der Ausgangswert

s² ist die Varianz

**Standardabweichung 🡆 wie hoch ist die Varianz zum Mittelwert**

Standardabweichung ist Wurzel aus der Varianz s =

Variationskoeffizient ist Quotient aus Standardabweichung / Mittelwert

v = s / x̅

**Präsenz am 17.11.2023**

**Übung zur Bestimmung der Lager- und Streuungsparameter**

Nicht alle Wert zur xi sind gegeben!

**Geg:**

x x ist das Merkmal (in der Übung die **Anzahl der Handys**)

xi xi ist die Merkmalsausprägung (in der Übung die konkrete **Anzahl der Handys**    
 = **Beobachtungswert**)

n Summe der absoluten Häufigkeiten (Anzahl der Ausprägungen)

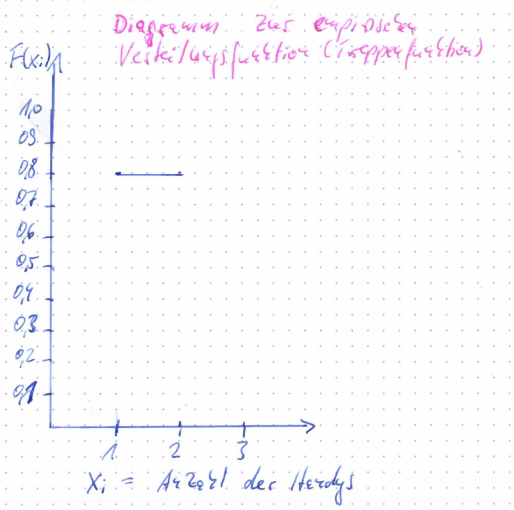
hi ist die absolute Häufigkeit

Benutzer sind die Merkmalsträger

**x1 = 1** ist die **Merkmalsausprägung** ((in der Übung **1 Handy**) = **Beobachtungswert**)

**h1 = 16**  die Häufigkeit zur ersten Merkmalsausprägung

**x̅ = 1,25**



F(x1) = 0,8 relative Summenhäufigkeit zur ersten Merkmalsausprägung aus dem Diagramm zur   
 empirischen Verteilungsfunktion (Treppenfunktion)

insgesamt gibt es 3 Merkmalsausprägungen 🡆 x1 = 1, x2 = 2, x3 =3

* Bestimmen Sie die absoluten Häufigkeiten, die relativen Häufigkeiten, die absoluten Summenhäufigkeiten und die relativen Summenhäufigkeiten zu allen Merkmalsausprägungen.
* Bestimmen Sie die Quartile, die Spannweite, die Varianz, die Standardabweichung und den Varianzkoeffizient.
* Vervollständigen Sie das Diagramm zur Treppenfunktion

Es sind also anhand

* der Häufigkeit zur ersten Merkmalsauprägung (hi = 16)
* dem arithmetischen Mittel x̅ ( x̅ = 1,25) und
* der relativen Summenhäufigkeit (Fi =0,8 aus der Grafik zur Treppenfunktion

die gefragten Werte zu bestimmen

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Anzahl der Handys  xi | hi | xi \* hi | fi | Hi | Fi | x̅ | (xi – x̅ )² | (xi – x̅ )² \* hi |
| **1** | **16** | 16 | 0,80 | 16 | **0,8** | **1,25** | 0,0625 | 1,00 |
| **2** | 3 | 6 | 0,15 | 19 | 0,95 | **1,25** | 0,5625 | 1,69 |
| **3** | 1 | 3 | 0,05 | 20 | 1,00 | **1,25** | 3,0625 | 3,06 |
| ∑ | 20 |  | 1,00 | - | - | - |  | 5,75 |

Werte in grüner Schrift sind in der Aufgabenstellung gegeben

**x1 = 1**

**x2 = 2**

**x3 = 3**

**h1 = 16**

**x̅ = 1,25**

**F1 = 0,8 (**aus dem Diagramm zur empirischen Verteilungsfunktion (Treppenfunktion))

**Berechnung der Summe zu den absoluten Häufigkeiten 🡆 n**

fi = xi / n

f1 = 0,8

x1 = 16

**Werte für f1 und x1 einsetzen**

0,8 = 16 / n | :1 (nach n umstellen)

1 / 0,8 = n / 16 | \* 16

n = 16 / 0,8

**n = 20**

**NR für h2**

n = h1 + h2 + h3

h1 + h2 + h3 = n

n = 20, h1 = 16

16 + h2 + h3 = 20 | -16

h2 + h3 = 4 | -4

h2 = **4 – h3**

**h2 und h3 berechnen**

x̅ = 1/n ∑xi \* hi

x̅ = 1/n \* (x1 \* h1 + x2 \* h2 + x3 \* h3) | Werte für x̅, n, x1, h1, x2, x3 einsetzen

1,25 = 1/20 \* (16 + 2h2 + 3 h3) | : 1/20

25 = (16 + 2h2 + 3h3) | - 16

9 = (2h2 + 3h3) | h2 aus NR einsetzen

9 = 2 \* (**4-h3**) + h3 | ausmultiplizieren

9 = 8 – 2h3 + h3

9 = 8 + h3 | -8

1 = h3

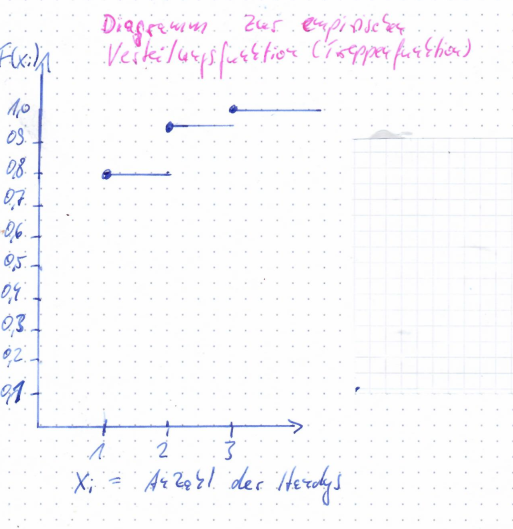
**h3 = 1**

h2 = 4- 1

**h2 = 3**

**h2 und h3 in Tabelle einsetzen und Werte für fi, xi \* hi, Hi, Fi, (xi \* hi)² berechnen**

**Diagramm zur empirischen Verteilungsfunktion (Treppenfunktion) vervollständigen**



**empirische Verteilungsfunktion F(x)**

* ist **relative Summenfunktion**
* gibt für jede reelle Zahl x den Anteil der Merkmalsträger an, für die das Merkmal X einen Wert xi annimmt, der kleiner oder gleich x ist
* **Wertebereich: 0 ≤ F(x) ≤ 1**
* ist monoton nichtfallend (**steigt oder ist konstant**)
* ist eine **Treppenfunktion mit Sprungstellen bei x1, x2, ..., xi**
* Die **Größe der Sprünge beträgt fi = F(xi) - F(xi-1))**

**Varianz s² berechnen**

s² = 1/n ( (x1 – x̅)² \* h1 + (x2 – x̅)² \* h2 +(x3 – x̅)² \* h3)

s² = 1/20 ( (1 – 1,25)² \*16 + (2 – 1,25)² \* 3+(3 – 1,25)² \* 1)

s² = 1/20 \* (1,00 + 1,69 + 3,06)

s² = 1/20 \* 5,75

s² = 0,2875

**Standardabweichung s berechnen**

**Varianzkoeffizient v berechnen**

v = s / x̅

**v = 0,536 / 1,25 = 0,4288**

**Varianzkoeffizient**

* Relatives Streuungsmaß (Streuungsparameter)
* dimensionslose Größe.
* prozentuale Verhältnis der Standardabweichung zum arithmetischen Mittel
* dient zum Vergleich der Streuung zwischen verschiedenen Erhebungen

**Spannweite w berechnen**

w = xmax - xmin

w = 3 – 1 = 2

**Übung zu klassierten Daten**

Zu folgender Tabelle sind für jede Gruppe die Quartile zu berechnen

Wieviel % der Befragten haben in jeder Gruppe ein alter zwischen 15 und 75 Jahren?

Boxplot zu den Streuungsmaßen Spannweite und Quartilsabstand und Lagemaß Median zeichnen

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Merkmal: Alter | | | | |
|  | relative Häufigkeit  fi | | relative Summenhäufigkeit  Fi | |
| Merkmalsausprägungen  xi | fi mit  Migrations-hintergrund in % | fi ohne  Migrations-hintergrund in % | Fi mit  Migrations-hintergrund in % | Fi ohne  Migrations-hintergrund in % |
| b. u. 15 | 22% | 12% | 22% | 12% |
| 15 b. u. 35 | 31% | 23% | Einfallsklasse für Q1  Einfallsklasse für Q2  53% | Einfallsklasse für Q1  35% |
| 35 b. u. 55 | 29% | 33% | Einfallsklasse für Q3  82% | Einfallsklasse für Q2  68% |
| 55 b. u. 75 | 15% | 25% | 97% | Einfallsklasse für Q3  93% |
| 75 und älter | 3 | 7% | 100% | 100% |
| ∑ | 100% | 100% | - | - |

**mit Migrationshintergrund**

**Median (Quartil Q2) berechnen**

**WICHTIG. WERTE ZU xi MÜSSEN NACH GRÖSSE SORTIERT SEIN**

**Einfallsklasse bestimmen**

Einfallsklasse für Q2 (Median) ist jene, wo relative Summenhäufigkeit Fi ≥ 50%

Einfallsklasse k ist die 2. Klasse (xi = 15 b. u. 35, Fi = 53%)

Q2 = xk-1 + (xk – xk-1) \* (0,5 – Fk-1) / fk

Q2 = 15 + (35 – 15) \* (0,5 – 0,22) / 0,31

Q2 = 15 + 20 \* (0,50 – 0,22) / 0,31 Punktrechnung vor Strichrechnung

Q2 = 15 + 20 \* 0,28 / 0,31

Q2 = 15 + 20 \* 0,90

Q2 = 15 + 18

**Q2 = 33,03**

**Quartil Q1 berechnen**

**Einfallsklasse bestimmen**

Einfallsklasse für Q1 ist jene, wo relative Summenhäufigkeit Fi ≥ 25%

Einfallsklasse k ist die 2. Klasse (xi = 15 b. u. 35, Fi = 53%)

Q1 = xk-1 + (xk – xk-1) \* (0,25 – Fk-1) / fk

Q1 = 15 + (35 – 15) \* (0,25 – 0,22) / 0,31

**Q1 = 16,93**

**Quartil Q3 berechnen**

**Einfallsklasse bestimmen**

Einfallsklasse für Q3 ist jene, wo relative Summenhäufigkeit Fi ≥ 75%

Einfallsklasse k ist die 3. Klasse (xi = 35 b. u. 55, Fi = 82%)

Q3 = xk-1 + (xk – xk-1) \* (0,75 – Fk-1) / fk

Q3 = 35 + (55 – 35) \* (0,75 – 0,53) / 0,29

**Q3 = 50,17**

**ohne Migrationshintergrund**

**Median (Quartil Q2) berechnen**

**Einfallsklasse bestimmen**

Einfallsklasse für Q2 (Median) ist jene, wo relative Summenhäufigkeit Fi ≥ 50%

Einfallsklasse k ist die 3. Klasse (xi = 35 b. u. 55, Fi = 68%)

Q2 = xk-1 + (xk – xk-1) \* (0,5 – Fk-1) / fk

Q2 = 35 + (55 – 35) \* (0,5 – 0,35) / 0,33

**Q2 = 44,09**

**Quartil Q1 berechnen**

**Einfallsklasse bestimmen**

Einfallsklasse für Q1 ist jene, wo relative Summenhäufigkeit Fi ≥ 25%

Einfallsklasse k ist die 2. Klasse (xi = 15 b. u. 35, Fi = 35%)

Q1 = xk-1 + (xk – xk-1) \* (0,25 – Fk-1) / fk

Q1 = 15 + (35 – 15) \* (0,25 – 0,12) / 0,23

**Q1 = 26,30**

**Quartil Q3 berechnen**

**Einfallsklasse bestimmen**

Einfallsklasse für Q3 ist jene, wo relative Summenhäufigkeit Fi ≥ 75%

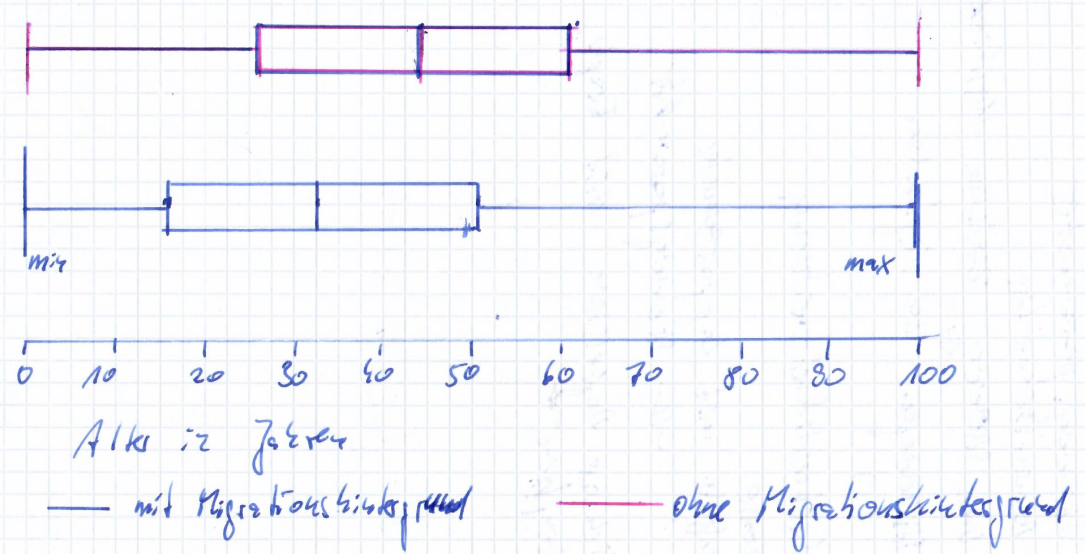
Einfallsklasse k ist die 3. Klasse (xi = 55 b. u. 75, Fi = 93%)

Q3 = xk-1 + (xk – xk-1) \* (0,75 – Fk-1) / fk

Q3 = 55 + (75 – 55) \* (0,75 – 0,68) / 0,25

**Q3 = 60,60**

**Boxplot**



**Wichtig**

* Legende („mit Migrationshintergrund“, „ohne Migrationshintergrund“) nicht vergessen
* Legend darf NICHT in das Diagramm, sondern muss unterhalb des Diagramms!
* Achsenbeschriftung („Alter in Jahren“) nicht vergessen

**Boxplot enthält**

* Lokalisation (Lage des Median)
* Streuungsmaße:
  + Spannweite = Ausdehnung eines Boxplots (Differenz w = xmax – xmin)
  + Quartilsabstand = Ausdehnung der Box (Differenz IQR = Q3 – Q1)

eines Datensatzes

Aus Boxplot lassen sich neben Median, Q1 und Q3 Parameter ( = Quartilsabstand) Informationen über die Schiefe (Vergleich der beiden Hälften der Box oder der Längen der Whisker) und Ausreißer entnehmen

**Anteil der Befragten im alter von 15 bis 75 Jahren**

mit Migrationshintergrund

31 + 29 + 15 = 75%

ohne Migrationshintergrund

23 + 33 + 25 = 81%

**Klausurinhalte**

* Theoretische Fragen
* Aufgabe zu Lageparameter (evtl. mit Graph)
* Aufgabe zu Streuungsparameter (evtl. mit Graph)
* Aufgabe zu linearen Degression (evtl. mit Graph)
* Aufgabe zu Wahrscheinlichkeitsrechnung

**Konferenz am 24.11.2023**

**Reression**

Analyse von Zusammenhängen

gemeinsame Auswertung mehrerer Merkmale

im Kurs nur 2 Merkmale

univariate Analyse 🡆 Auswertung einzelner Merkmale

multivariate Analyse 🡆 Auswertung mehrerer Merkmale

Rangkorrelationskoeffizient und Kontingenz (S. 6)

Werden wir nicht berechnen, aber Theorie muss bekannt sein

Abhängigkeit (s. 7)

Wir unterstellen vorab eine Richtung der Abhängigkeit

Erst Streudiagramm erstellen, danach die Regressionsgerade berechnet und eingezeichnet

S. 8

Abhängigkeit

Gerichteter Zusammenhang

S. 9

Streudiagramm

<0 und > 0-Werte in den Quadranten

Sagt die Stärke des Zusammenhanges auf

S. 10

Korrelation sagt noch nicht über kausalen Zusammenhang (Korrelation ist keine Kausalität)

Ist für Zusammenhang notwendig aber nicht

S. 11

Scheinkorrelation

Korrelation ist statistischer Zusammenhang, was berechnet wird

S. 14

Scheinkorrelation nur Annahme / gesunder Menschenverstand

K-koeffizient 🡆 dimensionslos 🡆 lässt sich vergleichen mit anderen K-Koeffizienten

S. 24

Multiple kein Thema im Kurs, Wissen muss aber bekannt sein!

S. 25

wichtig

Für die Regressionsfunktion bei multipl. 🡆 Zuerst prüfen, ob es einen Zusammenhang zwischen der Variablen gibt

S. 26

Schwarze Quadrate sind die Wertepaare (Punkte im Streudiagramm)

S. 28

A und b der Nenner ist derselbe und muss nur 1x berechnet werden

(Minute 35)

S. 30

Bsp. zur Interpretation (Schätzung)

S. 31

Bsp. für Schätzung (Prognosemodell)

S. 32

Regressionsrechnung ist einfaches Modell der Realität

Woher weiß man, ob Modell gut ist

Bestimmtheitsmaß als Gütemaß wird benötigt

Y Dach sind geschätzte Werte

S. 40einfache lineare Regression

(nur 1 unabhängig Variable= 🡆 Korrelationskoeffizient

Wichtig ist Bedeutung

50% lassen sich die Umsätze durch Verkaufsfläche erklären

50% nur durch andere Faktoren