rechtssteile Verteilung (linksschief): xD (Modus, Modalwert) > xZ(Median o. Zentralwert) > x̅ (Mittelwert)

linkssteile Verteilung (rechtsschief): xD < xZ > x̅, metrische Verteilung xZ = x̅

2-dimensionale Häufigkeitsverteilung 🡆 Kreuztabelle

M = Merkmal (A oder B), G = Geschlecht

Randverteilung für eindim.. Häufigkeitsverteilung von G

absolute Häufigkeit

folgende Werte sind in d. Tabelle einzutragen

*relat. Spaltenhäufigkeit (z. B 400 von 1000)*

*relat. Zeilenhäufigkeit (z. B 400 von 1200)*

*% unter den Summen sind relative Werte der Zeilen- / Spaltensumme zu den Randverteilungen* (Gesamtsumme)

|  |  |  |  |
| --- | --- | --- | --- |
| M/G | m | m | ∑ (**Randverteilung**) |
| A | **400**  *40%*  *33,33%* | **800**  *80%*  *66,66%* | *1200*  60% v **2000** |
| B | **600**  *60%*  *75%* | **200**  *20%*  *25%* | *800*  40% v **2000** |
| ∑ | *1000*  *50%* v. **2000** | *1000*  *50%* v. 2000 | ***2000***  100% |

fi =

**hi = fi \* n**

n = ∑hi = h1 + h2 +h3

20 = 16 + h2 + h3  | | -16

4 = h2 + h3 | - h3

h2 = 4 – h3

x̅ = ∑(xi \* hi)

1,25 =  **\*** 16 + 2h2 +3 h3 |

25 = 16 + 2 \* (4-h3) + h3 | - 16

9 = 8 **– 2h3 + h3** | - 8

1 = **– 2h3 + h3  🡆 h3 = 1**

h2 = 4 – h3  = 4 – 1 = 3

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Merkmals-  ausprä-  gung  Anzahl der Handys | absolute Häufigk.  Anz. Handy  -Nutzer |  | relative Häufigk.  **hi / n** | abs. Summen-  häufigk.  **hi+hi+1** | rel. Summen-  häufigk.  **fi + fi+1** | arithm.  Mittel  **∑(xi \* hi) / n**  **25 / 20** | Zwischen-  rechnung für Varianz |
| xi | hi | xi \* hi | fi | Hi | Fi | x̅ | (xi – x̅)2 \* hi |
| **1** | **16** | 16 | **0,80** | 16 | **0,8** | **1,25** | 1 |
| **2** | 3 | 6 | 0,15 | 19 | 0,95 | **1,25** | 1,687 |
| **3** | 1 | 3 | 0,05 | **20** | **1,00** | **1,25** | 3,063 |
| ∑ | **n = 20** | **25** | **1,00** | - | - | - | **5,75** |

**im Casio**

**Variable speichern: Wert eingeben 🡆 SHIFT 🡆 RCL 🡆 (–) 🡆 speichert Wert in Variable „A“**

**Variable verwenden: aus Variable „A“: ALPHA 🡆 (–)**

**SHIFT** 🡆 Taste **„1“** 🡆 **„1: Type“ (oder MODE** 🡆 **„2: STAT“ 🡆 „1: 1-VAR“)**

**SHIFT** 🡆 Taste **„1“** 🡆 **„2: Data“ 🡆 Dateneingabe in Tabelle (Spalte x, Spalte „FREQ“ = absolute Häufigkeiten)**  🡆 mit Taste **„AC“** speichern und verlassen (über **SHIFT** 🡆 Taste **„1“** 🡆 **„2: Data“** Daten prüfen

**n SHIFT** 🡆 Taste **„1“** 🡆 **„1: n“**

**Summe xi \* hi SHIFT** 🡆 Taste **„1“** 🡆 **„3: Sum“ 🡆 „2: ∑ x“**

**Mittelwert x̅ SHIFT** 🡆 Taste **„1“** 🡆 **„4: Var“** 🡆 **„2: x̅“ 🡆 x̅2**

**Mittelwert bei klassierten Daten** wie bei unklassierten Daten nur mit Klassenmitte mi, statt mit xiarithmetische Mittel aus unterer und oberer Klassengrenzem**i (**Klassenmitte) **= ,** hi = absol. Häufigkeit der Klasse

bei klassierten Daten können x̅ , Varianz s², Std.-Abw. s, Varianzko-effizient v wie bei unklassierten Daten im Casio berechnet werden.

Statt xi wird mi (Klassenmitte) in Tabelle eingetragen

🡆 Bsp. Klasse „20 b. u. 35 “ 🡆 **m1** = ½ \* (20 + 35) = **27,5** 🡆 **h1 = 15**, 🡆 **m1 \* h1 = 27,5 \* 15 = 412,50**  
 Klasse „35 b. u. 55“ 🡆 **m2** = ½ \* (35 + 55) = **45,0** 🡆 **h2 = 20** 🡆 **m2 \* h2 = 45,0 \* 20 = 900,00**  
 n = h1 + h2 = **15 + 20 = 35**  **∑ 1312,50**

**x̅ = 1312,50 : 35 = 37,50**

Klassenbreite Bsp. Klasse „20 b. u. 35 “: Klassenbreite bi = 35 – 20 = 15

Variablendefinition in Klassen: k: Klassenobergrenze (im Bsp. 35), k-1: Klassenuntergrenze (im Bsp, 20)

**Varianz s² SHIFT** 🡆 Taste **„1“** 🡆 **„4: Var“** 🡆 **„3: Sigma x“ 🡆 „x²“**

**arithmetische Mittel der Abweichungsquadrate**

wichtiger Streuungsparameter, für metrische Merkmale, Ausgangswert für Standardabweichung und Variationskoeffizient

**s² = ( 1/n \* ( (x1² \* h(x1)  +  (x2² \* h(x2)  +  (x3² \* h(x3) ) – x̅2)**

s² = (1/20 \* 37) - 1,252 = 1,85 – 1,5625 = 0,2875

**Varianz s² bei klassierten Daten**

**s² =** wie bei unklassierten Daten nur mit Klassenmitte mi statt mit xi

**Standardabweichung (durchschn. Abweichung) s SHIFT 🡆 Taste „1“ 🡆 „4: Var“ 🡆 „3: Sigma x“**

**Varianzkoeffizient v SHIFT 🡆 Taste „1“ 🡆 „4: Var“ 🡆 „3: Sigma x“ : x̅**

**v = s / x̅** Standardabweichung s / Mittelwert x̅

Relatives Streuungsmaß (Streuungsparameter), dimensionslose Größe, prozentuales Verhältnis der Standardabweichung zum arithmetischen Mittel, zum Vergleich der Streuung zwischen verschiedenen Erhebungen

**Modus x̅D** 🡆 Beobachtungswert mit der größten Häufigkeit 🡆 mehrere Merkmalsausprägungen mit der gleichen maximalen Häufigkeit = mehrere Modalwerte = Beobachtungswert mit Häufigkeit == 1 = kein Modus!) 🡆 Modus xD aus o. g. Tabelle ist 1 🡆 x1 = 1 hat die max. hi = 16

bei klassierten Daten ist Modalwert die Mitte der Klasse mit den größten Häufigkeiten (Modalklasse)

im o. g. Bsp. ist Klasse k2 mit h2 = 20 die Modalklasse und der Modalwert = mi = ½ \* (35+55) = 45

**Spannweite w**

w = xmax - xmin

Spannweite bei klassierten Daten w = mimax - mimin

Bsp. Klassen „0 b. u. 25“, „25 b. u. 40“, „40 b. u. 55“, „55 b. u. 70“, „70 b. u. 90“

mi 12,5 7,5 7,5 7,5 10

w = mimax - mimin = 12,5 – 7,5 = 5

**Quartile** 🡆 Q1, Q2, Q3 🡆 Q2-Quartil = **Median / Zentralwert x̅Z**

**Beobachtungs-Werte müssen geordnet sein!**

wenn xn eine gerade Zahl **x̅z =**bei 14 Beobachtungswerten x̅z = (x14:2 + x14:2+1) \* ½ = (x7 + x8) \* ½ 🡆 Werte von x7 und x8 werden addiert und mit 0,5 multipliziert

wenn xn eine ungerade Zahl **x̅z =**

bei 13 Beobachtungswerten x̅z = x13+1 = x7 🡆 Wert von x7 wird ausgelesen

Q1-Quartil und Q3-Quartil nach demselben Algorithmus (bei xn = gerade = xn/2 + xn/2+1) ausgehend vom Q2-Quartil

Achtung: **Es werden die Werte unter bzw. über dem Median berücksichtig!**

Bsp. bei 14 Beobachtungswerten

Für Q1 werden die Werte x1 bis x6 (kleinster xi-Wert für Q2 – 1 🡆 hier x7 – 1 = x6) berücksichtigt. x6 🡆 6 ist gerade Anzahl, daher **Q2  =**

Für Q3 werden die Werte x9 bis x14 berücksichtigt

Für Q3 werden vom letzten xn-Wert für Q2 die für Q1 ermittelten xi-Werte (hier 3 und 4) addiert, um die xi-Werte zu erhalten **Q3 =**

**Median bei klassierten Daten**

k = Einfallsklasse 🡆 Klasse mit rel. Summenhäufigkeit Fi ≥ 50%,

**x**k-1 = Untergrenze d. Einfallsklasse, **x**k = Obergrenze d. Einfallsklasse,

Fk-1 = relative Summenhäufigkeit der Klasse unterhalb (vor) der Einfallsklasse

fk = relative Häufigkeit der Einfallsklasse

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| xi | hi | fi | Hi | Fi |
| 10 b. u. 30 | 60 | 30 | 60 | **30** |
| **30** b. u. 40 | 80 | 40 | 140 | 70 |
| 40 b. u. 70 | 60 | 30 | 200 | 100 |
| ∑ | 200 | 100 | - | - |

**Einfallsklasse ist k2 „30 b. u. 40“, da Fi ≥ 50** (in Klasse „10 b. u. 30“ ist Fi < 50

**Hinweis. %-Werte immer als Dezimalwerte schreiben**

xk-1 = K-Untergrenze, xk = K-Obergrenze

bei Q1, Q3 wird genauso verfahren.

Die Einfallsklasse muss rel. Summenhäufigkeit

Fi ≥ 25% (bei Q1) bzw. Fi ≥ 75% (bei Q3) haben.

Für den %-Wert im Zähler des Bruchs wird statt

0,5 🡆 0,25 (für Q1) bzw. 0,75 (für Q3) genommen

**Interquartilsabstand IQR = Q3 – Q1**

Streuungsmaß

Differenz zwischen oberem und unterem Quartil

enthält 50% der Verteilung.

**Boxplot**:  
enthält Minimum, Q1, Median (Lokalisation),

Q3, Maximum

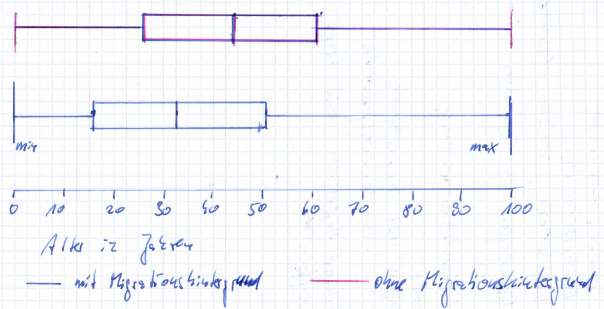
Streuungsmaße Spannweite (xmax – xmin), IQR

Informationen über die Schiefe und Ausreiße

Achsenbeschriftung und Legend („mit Migrationshintergr.“) nicht vergessen, beides unterhalb das Boxplots“

Achse enthält die Skala zu den x-Werten.

Boxplot enthält die zum Median und zu den Quartilen berechneten x-Werte und Xmax- und Xmin sowie den IQR



**Korrelationsanalyse 🡆 Zusammenhangsanalyse (Interdependenzanalyse) zwischen 2 metrischen Merkmalen X und Y**K-Analyse prüft, **ob zwei Variablen X und Y linear zusammenhängen** und **prüft die Stärke des Zusammenhangs**

**analysiert Wechselwirkung der Variablen X, Y untereinander** (wie beeinflussen sich die Variablen gegenseitig?)  
**Korrelationskoeffizient rx,y: Zusammenhangsmaß (Assoziationsmaß) -1 ≤ rxy ≤ +1**

**Korrelation 🡆 zahlenmäßiger statistischer Zusammenhang zwischen 2 Merkmalen X und Y**

positive K. (**Korrelationskoeffizient r > 0)**: **gleichförmige Entwicklung von X und Y (hoher Wert von X 🡆 auch hoher Wert von Y**)

negative K. (**r < 0)**: **gegenläufige Entwicklung von X und Y (hoher Wert von X 🡆 jedoch niedriger Wert von Y)**

r = -1: extrem starker negativer linearer Zusammenhang vor🡆 Punktewolke auf einer Geraden mit negativer Steigung

r = +1: extrem starker positiver linearer Zusammenhang 🡆 Punktewolke liegt auf einer Geraden mit positiver Steigung.

r = 0: kein linearer Zusammenhang

kausaler Zusammenhang: Ursache-Wirkungs-Beziehung zwischen X und Y 🡆 Veränderung des abhängigen Merkmals Y basiert auf Veränderung von X

Korrelation: keine Aussage zum kausalen Zusammenhang und zur Kausalitätsrichtung

**Korrelationskoeffizient rxy:**

**dimensionsloses Maß für den Grad des linearen Zusammenhangs** (**Stärke und Richtung des linearen Zusammenhangs**)

für 2 mind. intervallskalierte Merkmale X und Y mit positiver Kovarianz cov(x,y) und Standardabweichung s

**Regressionsanalyse 🡆 Abhängigkeitsanalyse (Dependenzanalyse) für gerichteten Zusammenhang**

Abhängigkeitsanalyse zwischen 2 Merkmalen X, Y durch Regressionsanalyse

Unterscheidung zwischen unabhängigen und abhängigen Merkmalen

R-Analyse basiert auf Regressionsfunktion ŷ = a + b\*x 🡆 R-Funktion ermittelt Abhängigkeitsmaß

Streudiagramm und Regressionsgerade

**Diagramme**  
WICHTIG: sprechende Achsenbeschriftungen, sinnvolle Skaleneinteilung, ggf. Legende

**Säulendiagramm**   
- höhenproportionale Häufigkeitsverteilung (senkrechte nicht aneinandergrenzende Säulen); Säulen können beliebig breit sein; y-Achse zur abs. Häufigkeit hi, x-Achse zu x-Werten, für wenige Ausprägungen

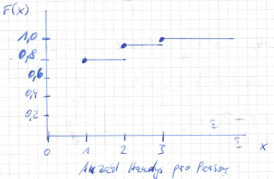
**Stabdiagramm** (Liniendiagramm) 🡆 wie Säulendiagramm nur mit schmalen Säulen

**Balkendiagramm** 🡆 am häufigsten verwendet, wie Säulendiagramm nur mit horizontalen Balken ( y-Achse zu x-Werten; x-Achse zur abs. Häufigkeit hi)

**Kreisdiagramm** 🡆 nur für eine Datenreihe, keine negativen Werte, keine 0-Werte, Kategorien repräsentieren Teile des gesamten Kreisdiagramms, max. 7 Teilwerte

**Histogramm**   
- **flächenproportionale Darstellung der absoluten oder relativen** (beide sind möglich) **Häufigkeiten** von ausschließlich klassierten Daten   
- **x-Achse**: Werte müssen auf Skala geordnet sein und gleiche Abstände haben,   
 x-Achse enthält **aneinandergrenzende Rechtecke zur Klassenbereite**, **keine Abstände zwischen den Flächen**

- **y-Achse**: **Höhe = Häufigkeitskoeffizient** berechnet aus (**hi : bi** (Klassenbreite)) oder (**fi : bi** ),   
 Skaleneinteilung entsprechend den berechneten Werten, **in den Rechtecken den Wert zur relativen Häufigkeit fi** eintragen

**empirische Verteilungsfunktion** (Treppenfunktion) F(x) ist (**relative) Summenhäufigkeitskurve**

- **x-Achse enthält x-Werte oder Klassengrenzen**,

- bei klassierten Daten werden die Punkte / Striche im Diagramm bei K-Obergrenze gezeichnet

- y-Achse enthält Skala zur relativen Summen-Häufigkeit Fi:;

- **Punkt** oder Strich (oder Punkt mit Strich) wird **zur Summenhäufigkeit auf y-Achse** **und**

**zum x-Wert bzw. zur K-Obergrenze auf x-Achse** gezeichnet

**Streudiagramm**

graphische Darstellung zur Abhängigkeit (Regression) und zum Zusammenhang (Korrelation) von **beobachteten Wertepaaren zweier Merkmale X, Y**  
Wertepaare werden in kartesisches Koordinatensystem eingetragen 🡆 ergibt Punktwolke  
Stärke und Richtung des Zusammenhangs anhand der Lage und Form der Punktwolke;   
erste Hinweise über mögliche Abhängigkeit zwischen Merkmalen

**Boxplot**

**W-Graph** (Baumdiagramm)