Statistik

Grundlagen

- Welche Daten habe ich zur Verfügung?
- Welche stat. Maße stehen mir zur Verfügung?

Grundgesamtheit (GG)

Die Grundgesamtheit ist die Menge alle möglichen Daten.
Dies kann endlich sein oder theoretisch unendlich.

Stichprobe (SP)

Dabei handelt es sich um eine Teilmenge der Grundgesamtheit.

Beispiel: Zoohandlung
Männerunterhose (10cm) BLMUH

Wie erzeuge ich eine Stichprobe?

Zufällig: Die Elemente der Stichprobe werden ohne Wissen einer Eigenschaft ausgewählt.
→ Computer

Bestandsmassen

→ Massen die sich auf einen best. Zeitpunkt beziehen

Bewegungsmassen

→ Massen die sich auf einen best. Zeitraum beziehen.

Abgrenzung der Grundgesamtheit

Abgrenzungsmerkmale

→ sachliche Abgrenzung
→ örtliche Abgrenzung
→ zeitliche Abgrenzung

epräsentativ:

Es wird versucht ein Spiegelbild der Grundgesamtheit in der Stichprobe zu erzeugen.

Beispiel: $\boxed{100 Stk}$

M_1	M_2	M_3	M_4
40%	20%	25%	15%
40Stk	20Stk	25Stk.	15Stk

An hand dieser Stichprobe schließe ich wieder auf die Grundgesamtheit.

Merkmal:
Die Eigenschaft, die untersucht werden soll.

Merkmalsausprägungen:
Die Werte die eine best. Merkmal annehmen kann.

Merkmalsträger:
Die stat. Einheit welche die Merkmale besitzt.

Die Schreibweise für eine Grundgesamtheit / Stichprobe:

$$S = \{x_1, x_2, \dots, x_n\}$$

n: Anzahl der Elemente
x_n: Alle Elemente; Werte können auch mehrfach vorkommen.

Schreibweise der Merkmalsausprägungen:

$$T = \{a_1, a_2, \dots, a_m\}$$

m: Anzahl der verschiedenen Ausprägungen
a_m: Jede Ausprägung kommt nur einmal vor.

Skalenniveaus

Kriterien:

→ Wert (braun; 17; ....)
→ Reihenfolge (Note, ...)
→ Verhältnisse/ Differenzen (Anzahl der Geschw., ..)

Beispiel: 60 Personen wurden nach ihrem Familienstand befragt.

ledig – 0
verh. – 1
gesch. – 2
verw. – ~~3~~ 20

Codierung

Fassen Sie die Erhebung zusammen?

⇒ Der durchschnittliche Familienstand beträgt ~~2,14~~. 17,1

⇒ Die Daten die zugrunde liegen, lassen diese Berechnung nicht zu.

Beispiele:

Beispiel:
Haarfarbe

Wert ✓
Reihenfolge –
Verh. / Diff –

"Nominalskala"

Beispiel
Mathe / Stat-Note

Wert ✓
Reihenfolge ✓
Verh. / Diff –

"Ordinalskala"

Beispiel
Anzahl der Kinder

Wert ✓
Reihenfolge ✓
Verh. / Diff. ✓

"metrische Skala"

Beispiel: Jemand sagt er habe 1,5 Kinder.

→ Für diese Frage gibt es als Antwort nur ganze Zahlen.

Diskret:
In einem best. Intervall gibt es keine weiteren Werte mehr.

Stetig:
In einem best. Intervall gibt es bel. (theoretisch) viele weitere Werte.

Beispiel: Länge eine Person

180 - 181 cm
Durch Steigerung der Messgenauigkeit können bel. viele Zwischenwerte erzeugt werden.

Beispiel: Geldbetrag

Diskret / stetig ?

Zur Übung rechnen Sie
bitte folgende Aufgaben
aus de Aufgabensammlung

13/16/21

Feststellung des Skalenniveaus

- Wert
- Reihenfolge
- Verhältnisse / Differenzen

a) Jahresumsatz eines Unternehmens
metrischen Skala

b) Körperlänge von männlichen Stud.
metrische Skala

c) Nationalität von Sportlern
Nominalskala

d) Geschlecht der Stud. einer Hochschule
Nominalskala

e) Haushaltsgröße (Personenanzahl)
metrische Skala

f) Schulnoten von 1 bis 6
Ordinalskala

diskret - stetig

a) Einkommen - diskret /stetig

b) Haarfabe - diskret

c) Körperlänge - stetige

d) Anzahl de Personen in einem ICE
diskrete

Welche Merkmale sind stetig

A: Erlernter Beruf - diskret

B: Gründe für die Wahl einer Partei
diskret

C: Einwohnerzahl einer Stadt
diskret

D: Stromverbrauch in kWh
stetig / diskret

E: Körpergröße
stetig / diskret

D + E sind richtig

Tabellarische Aufbereitung von Stichprobenweten

Urliste:
Die Urliste ist im Bereich de Statistik das direkte Ergebnis eine Datenerhebung, also die ursprüngliche Aufzeichnung der Beobachtungs- oder Messwerte.

Stichprobenwerte:
Die einzelnen Beobachtungswerte de Urliste heißen Stichprobenwerte.
Werte können auch mehrfach auftreten.
Diese Mehrfachwerte dürfen auf keinen Fall gelöscht werden.

$$\boxed{\begin{array}{l} A = \{x_1, x_2, \ldots x_n\} \\ n\text{: Anzahl de Stichprobenwerte} \end{array}}$$

Merkmalsausprägungen:
Die Beobachts- oder Messwerte die in einer Stichprobe vorkommen.
Jede Merkmalsausprägung kommt nur einmal vor.

$$B = \{a_1, a_2, \ldots, a_m\}$$

m: Anzahl der verschiedenen Ausprägungen in der Stichprobe

In der Praxis kommt eine Urliste zwar immer vor, praktisch wird aber diese Urliste sofort in einer Strichliste dargestellt.

Beispiel:
In zwei Parallelklassen wurde das Alter der Schüler ermittelt. Die Ergebnisse sehen Sie im Anschluss

A-klasse:

a_i	16	17	18	19
n_i	4	16	3	2
h_i	4/25	16/25	3/25	2/25
h_i	0,16	0,64	0,12	0,08
h_i	16%	64%	12%	8%

Σ25

B-klasse:

a_i	16	17	18	19
n_i	4	13	2	1
h_i	4/20	13/20	2/20	1/20
h_i	0,20	0,65	0,10	0,05
h_i	20%	65%	10%	5%

Σ20

Frage:
In welcher Klasse ist der Anteil der 17-jährigen größer?

⇒ In der B-klasse ist der Anteil der 17-jährigen um 1% höher als in der A-klasse.

→ Absolute Häufigkeiten können oftmals nicht verwendet werden, wenn eine unterschiedliche Anzahl von Stichprobenweiten vorliegt.

Relative Häufigkeit

Tritt die Merkmalsausprägung a_i in einer Urliste mit n Stichprobenweiten n_i mal auf, so nennt man

$$\boxed{h_i = \frac{n_i}{n} = \frac{\text{Anzahl günstige Elemente}}{\text{Anzahl aller Elemente}}}$$

die relative Häufigkeit von a_i in der Urliste.

Häufigkeitstabelle

→ absolute Häufigkeit n_i
→ relative Häufigkeit h_i
→ absolute Summenhäufigkeit N_i
absolute kumulierte Häufigkeit
→ relative Summenhäufigkeit H_i
relative kumulierte Häufigkeit

kumulierte Häufigkeit:
Die kumulierte Häufigkeit umfasst die bis zur betreffenden Ausprägung aufsummierten absoluten bzw. relativen Häufigkeiten.
Sehr oft bezeichnet man dieses auch als Summenhäufigkeit.

Beispiel:

In einem Betrieb mit 60 Beschäftigten sind

6 Mitarbeiter bis 20 Jahre alt
18 Mitarbeiter über 20-30 Jahre alt
9 Mitarbeiter über 30-40 Jahre alt
12 Mitarbeiter über 40-50 Jahre alt
15 Mitarbeiter über 50 Jahre alt

Erstellen Sie eine Häufigkeitstabelle.
Fragen zu der Tabelle im Anschluss.

a_i	n_i	h_i	N_i	H_i
bis 20	6	10%	6	10%
ü20-30	18	30%	24	40%
ü30-40	9	15%	33	55%
ü40-50	12	20%	45	75%
ü50	15	25%	60	100%
	Σ60	Σ100%		

Fragen:

Wieviele Mitarbeiter (Anzahl / Prozent) sind über 30 bis 40 Jahre?

→ 9 Mitarbeiter / 15%

Wieviele Mitarbeiter (Anzahl / Prozent) sind 40 Jahre oder jünger?

→ 33 Mitarbeiter / 55%

In der folgenden Tabelle sind 60 Preise für den Kraftstoff Diesel. Diese Werte wurden an 60 verschiedenen Tankstellen zur gleichen Zeit an verschiedenen Orten ermittelt.

Tankstellen nummer	Preis in €		Tankstellen nummer	Preis in €		Tankstellen nummer	Preis in €		Tankstellen nummer	Preis in €	
1	1,16		16	1,17		31	1,12	10	46	1,09	15
2	1,15	1	17	1,18		32	1,13	11	47	1,11	16
3	1,14	2	18	1,19		33	1,14	12	48	1,11	17
4	1,09	3	19	1,22		34	1,15	13	49	1,12	18
5	1,15	4	20	1,21		35	1,16		50	1,13	19
6	1,08	5	21	1,11	7	36	1,17		51	1,14	20
7	1,21		22	1,23		37	1,18		52	1,15	21
8	1,22		23	1,24		38	1,19		53	1,16	
9	1,23		24	1,09	8	39	1,09	14	54	1,21	
10	1,33		25	1,26		40	1,21		55	1,18	
11	1,18		26	1,04	9	41	1,22		56	1,19	
12	1,17		27	1,28		42	1,23		57	1,16	
13	1,16		28	1,29		43	1,24		58	1,21	
14	1,17		29	1,29		44	1,25		59	1,22	
15	1,12	6	30	1,31		45	1,26		60	1,23	

Berechnen Sie die relative kumulierte Häufigkeit für folgende Aussage:

Der Spritpreis betrug höchstens 1,15 Euro.

$$\frac{21}{60} = \underline{35\%}$$

Häufigkeitstabelle:

Fehltage	n_i	h_i	N_i	H_i
0	5	10%	5	10%
3	9	18%	14	28%
5	13	26%	27	54%
9	9	18%	36	72%
12	8	16%	44	88%
18	4	8%	48	96%
21	2	4%	50	100%
	$\Sigma 50$	$\Sigma 100\%$		

$\Rightarrow$ 54%

In der folgenden Tabelle sind die Fehlzeiten von den 50 Mitarbeitern der Backplasa AG des letzten Jahres aufgelistet.

Fehlzeit in Tage	0	3	5	9	12	18	21
Anzahl der MA	5	9	13	9	8	4	2

Welcher Anteil der gesamten Fehlzeit entfällt auf die oberen (kränksten) acht Mitarbeiter?

Gesamtanzahl de Fehltage:

Gesamt $= 0 \cdot 5 + 3 \cdot 9 + 5 \cdot 13 + 9 \cdot 9 + 12 \cdot 8 + 18 \cdot 4 + 21 \cdot 2$

Gesamt $=$ 383 Tage

oberen $= 2 \cdot 21 + 4 \cdot 18 + 2 \cdot 12 =$ 138 Tage

$$\frac{138}{383} = 0{,}3603 \mathrel{\hat{=}} 36{,}03\%$$

x_i: Stichprobenwerte

Beispiel:

1 2 3 4 5

Durchschnittliche Lebenserwartung?

$$\bar{x} = \frac{1}{5} \cdot (1+2+3+4+5)$$

$$\bar{x} = 3 \text{ Jahre}$$

Eigenschaften:

- Empfindlich bzgl. Ausreißern
- metrische Skala
- Das Ergebnis der Berechnung ist in den Stichprobenwerten nicht enthalten
- Überlegen wie man rundet

Median

→ Der Median ist die Merkmalsausprägung des genau in der Mitte liegenden Einzelwertes.
→ Er teilt die der Größe nach geordneten Messergebnisse in zwei Hälften
→ Er wird auch als Zentralwert bezeichnet.
→ In jeder dieser Hälften befinden sich 50% der Werte.

Beispiel: 1 2 3 4 5

$\bar{x} = 3$
$x_{Med} = 3$ ⇒ keine Ausreißer

1 2 3 4 10.000

$\bar{x} = 2.002$
$x_{Med} = 3$ ⇒ Es sind Ausreißer vorhanden.

Eigenschaften:
→ Die Stichprobenwerte müssen aufsteigend sortiert werden.
→ mind. Ordinalskala
→ Unempfindlich bzgl. Ausreißern.

Formel:

Die Anzahl der Stichprobenwerte ist ungerade:

$$\boxed{x_{Med} = x_{\left(\frac{n+1}{2}\right)}}$$

→ Das Ergebnis ergibt die Nummer des gesuchten Wertes in der aufsteigend sortierten Liste der Stichprobenwerte.

Beispiel: 1 2 3 4 5

x_1 x_2 x_3 x_4 x_5

$$x_{Med} = x_{\left(\frac{5+1}{2}\right)} = x_3 \Rightarrow x_{Med} = 3$$

Die Anzahl der Stichprobenwerte ist gerade:

$$\boxed{x_{Med} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)}$$

Beispiel: 1 6 18 36 37 41

x_1 x_2 x_3 x_4 x_5 x_6

$$x_{Med} = \frac{1}{2}\left(x\left(\frac{6}{2}\right) + x\left(\frac{6}{2}+1\right)\right) = \frac{1}{2}(x_3 + x_4)$$
$$= \frac{1}{2}(18+36) = 27$$

Modus

Der Modus (Modalwert) ist der Merkmalswert, welcher am häufigsten vorkommt.

Eigenschaften:
- mind. Nominalskala
- keinen Modus –
  eindeutigen Modus ✓
  mehrere Modi –

Zur Übung rechnen Sie bitte folgende Aufgabe aus der Aufgabensammlung

28

Beantworten Sie die Frage jeweils nur mit ja oder nein.

Zu den Daten 18, 13, 16, 13, 19, 12 ist der Median kleiner als der arithmetische Mittelwert.

Zu den Daten 18, 13, 16, 13, 19, 12 ist der Modalwert kleiner als der arithmetische Mittelwert.

Zu den Daten 18, 13, 16, 12, 19, 19 ist der arithmetische Mittelwert kleiner als der Modalwert.

Zu den Daten 19, 18, 19, 12, 12 ist der arithmetische Mittelwert kleiner als der Median.

Zu den Daten 18, 13, 16, 12, 19, 19, 19 ist der Median kleiner als der Modalwert.

Zu den Daten 18, 13, 16, 13, 19, 12, 22 ist der Modalwert kleiner als der Median.

Zu den Daten 188, 130, 160, 121, 190, 190 ist der arithmetische Mittelwert kleiner als der Modalwert.

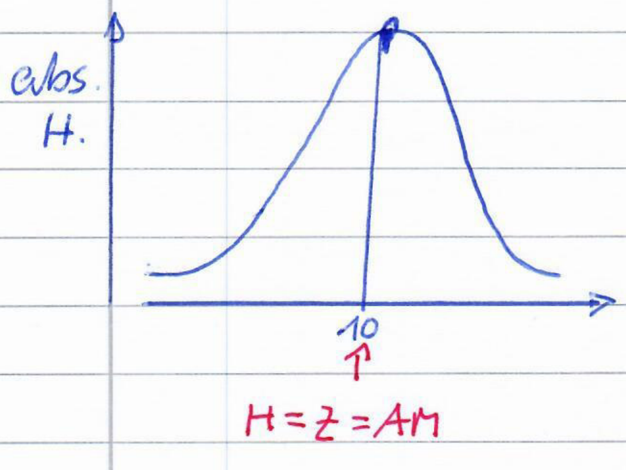
Antwort: Ja

→ Median: aufsteigend zu sortieren.

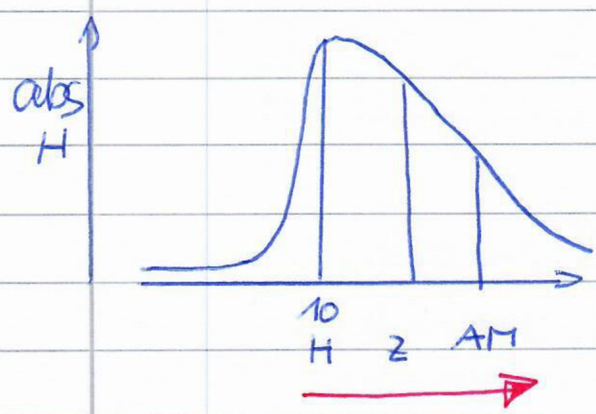
Schiefe

Die Schiefe ist ein Maß für die Abweichung eine Stichprobe vom Mittelwert eine symmetrischen Verteilung.

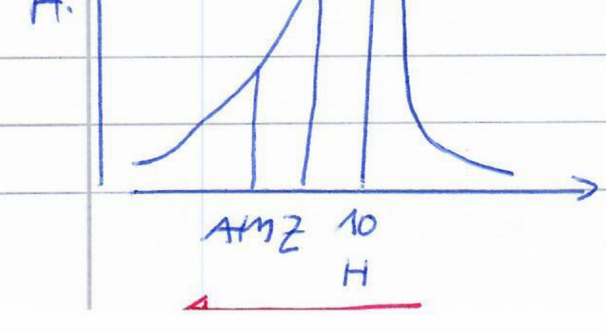
AM - aritm. Mittel
Z - Median
H - Modus / Modalwert



"Symmetrische Verteilung"



"linkssteile / rechtsschiefe Verteilung"



"rechtssteile / linksschiefe Verteilung

Formel:

$$\boxed{V = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3}$$

n: Gesamtanzahl der Stichprobenwerte
x_i: Stichprobenwerte
$\bar{x}$: arithm. Mittel
s: Standardabweichung

Eigenschaften:

- metrische Skala
- empfindlich bzgl. Ausreißern

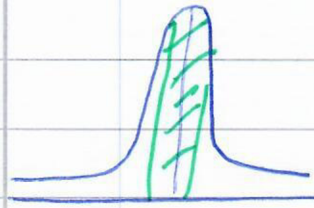
Interpretation:
→ $V > 0$ → rechtsschief
→ $V < 0$ → linksschief
→ $V = 0$ → symmetrisch

Wölbung

→ Gibt die Krümmung der Verteilung an.


flachgewölbt / platykurtisch


mittelgewölbt / mesokurtisch


hochgewölbt / leptokurtisch

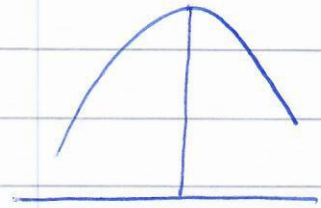
Formel:

$$\boxed{W = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^4}$$

Eigenschaften:
- Wie bei Schiefe

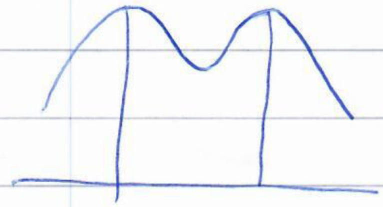
Intepretation:
- Je größe W desto größer ist auch die Wölbung.

Modalität



"unimodal"

Eindeutigen Modus



"bimodal"

mehrere (zwei) Modi

Exzess

Hier wird ein Vergleich der Wölbung zu einer Normalverteilung hergestellt.

Normalverteilung Wölbung von 3

$$\boxed{\text{Exzess} = \text{Wölbung} - 3}$$

Exzess < 0 flachgipflig

Exzess $= 0$ normalgipflig

Exzess > 0 steilgipflige

Gewogenes (gewichtetes) arithm. Mittel

Wird dann verwendet wenn die Einzelwerte gehäuft vorkommen.

Beispiel:
100 Frauen sind im Durchschnitt 168 cm groß.
50 Männer sind im Durchschnitt 180 cm groß.
Wie groß ist die Gesamtgruppe im Durchschnitt?

$$\bar{x} = \frac{1}{150} \cdot (\underbrace{168 + \ldots + 168}_{100 \text{ mal}} + \underbrace{180 + \ldots + 180}_{50 \text{ mal}})$$

$$\bar{x} = \frac{1}{150} \cdot (168 \cdot 100 + 180 \cdot 50) = 172 \text{ cm}$$

→ Gleiche Werte werden zusammengefasst.

$$\boxed{\bar{x}_g = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i \cdot n_i}$$

k: Anzahl unterschiedlich Stichprobenwerte
n_i: abs. Häufigkeit
x_i: unterschiedliche Stichprobenwerte

$$\bar{x}_g = \frac{1}{150} \cdot \sum_{i=1}^{2} (168 \cdot 100 + 180 \cdot 50) = \underline{\underline{172 \text{ cm}}}$$

Geometrisches Mittel

Beispiel:
Wir beobachten eine Bakterienkultur und notieren das Wachstum.

Bestand	1000	1800	2520	3276	4586
Veränderungsfaktor		1,8	1,4	1,3	1,4
Zuwachs		800	720	756	1310

Durchschnittliche Veränderungsfaktor?

$$\bar{x} = \frac{1}{4}(1{,}8 + 1{,}4 + 1{,}3 + 1{,}4) = 1{,}4750 \ (+47{,}50\%)$$

Probe: $1000 \cdot 1{,}475^4 = 4733$

=> Mit dem arithm. Mittel kann hier nicht gerechnet werden.

Definition:
Das geometrische Mittel wird immer dann verwendet, wenn multiplikativ verknüpfte Werte vorliegen.

$$\bar{x}_{geo} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Beispiel: $\bar{x}_{geo} = \sqrt[4]{1{,}8 \cdot 1{,}4 \cdot 1{,}3 \cdot 1{,}4} = 1{,}4634$

Probe: $1000 \cdot 1{,}4634 = 4586$

Harmonisches Mittel

Beispiel:
Ein Autofahrer fährt staubedingt 50km mit 20km/h und danach 50km mit 125km/h.
Wie groß war die Durchschnittsgeschw.?

$$\bar{x} = \frac{1}{2}\left(20\,\text{km/h} + 125\,\text{km/h}\right) = \cancel{72{,}5}\ \text{km/h}$$

falsch

Definition:
Mit dem harmonischen Mittel lassen sich Mittelwerte von Verhältniszahlen berechnen.
Es ist ein Lageparameter und kommt bei Verhältniszahlen zur Anwendung.
Man berechnet mit ihm den Mittelwert dieser Zahlen. Als Verhältniszahlen werden Brüche bezeichnet, die eine Beziehung wiederspiegeln.

Beispiele:
Studenten pro Einwohner
Preis pro m^2

$$\boxed{\bar{x}_n = \frac{a_1 + a_2 + \ldots + a_n}{\frac{a_1}{x_1} + \frac{a_2}{x_2} + \ldots + \frac{a_n}{x_n}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}}$$

a_n: Anteilswete

x_n: Zugehorigen Wete

→ Sollte $a_1 - a_n$ gleich groß sein, so kann auch die zweite Formel verwendet werden.

Beispiel

$$\bar{x}_n = \frac{50 + 50}{\frac{50}{20} + \frac{50}{125}} = \frac{2}{\frac{1}{20} + \frac{1}{125}} = 34{,}48 \frac{km}{h}$$

Beispiel:

Erwin kauft Apfelsinen jeweils für 12€.

12€ pro Stück 0,50€	24 Stk.
12€ pro Stück 0,40€	30 Stk.
12€ pro Stück 0,30€	40 Stk
	94 Stk

Stückpreis:

$$\frac{36}{94}\ €/\text{Stk} = 0{,}38\ €/\text{Stk}.$$

$$\bar{x}_h = \frac{12+12+12}{\frac{12}{0{,}5}+\frac{12}{0{,}4}+\frac{12}{0{,}3}} = \frac{3}{\frac{1}{0{,}5}+\frac{1}{0{,}4}+\frac{1}{0{,}3}} = 0{,}38\ €/\text{Stk}$$

Zur Übung rechnen Sie
bitte folgende Aufgaben
aus de Aufgaben-
sammlung

36/39/45/46/51

$$\bar{x} = \frac{1}{13}(2+3+6+5+2+8+7+2+4+3+1+3+0)$$

$$\underline{\underline{\bar{x} = 3,54}}$$

Verhältniszahlen:

$$\bar{X}_h = \frac{15+25+20}{\frac{15}{45}+\frac{25}{100}+\frac{20}{40}} = 55{,}38 \ km/h$$

Arithm. Mittel:

$$\bar{x} = \frac{1}{10} \cdot (32+27+29+25+34+28+36+30+32+39)$$

$$\bar{x} = 31{,}2$$

Median:
- Die Stichprobenwerte müssen der Größe nach aufsteigend sortiert werden.

~~25 28 29 30~~

25 27 28 29 30 32 32 34 36 39

$x_5 \quad x_6$

$$x_{Med} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) = \frac{1}{2}(x_5 + x_6)$$

$$x_{Med} = \frac{1}{2}(30+32) = \underline{\underline{31}}$$

Multiplikativ verknüpfte Messwerte

$$\bar{X}_{geo} = \sqrt[4]{1{,}08 \cdot 1{,}15 \cdot 0{,}96 \cdot 1{,}12}$$

$$\bar{X}_{geo} = 1{,}075 \quad \Rightarrow 7{,}5\%$$

Arithm. Mittel:

$\bar{x} = \frac{1}{50} \cdot (0 \cdot 5 + 3 \cdot 9 + 5 \cdot 13 + 9 \cdot 9 + 12 \cdot 8 + 18 \cdot 4 + 21 \cdot 2$

$\bar{x} = 7{,}66 \Rightarrow \bar{x} = 8$

Median:

$x_{Med} = \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right) = \frac{1}{2}(x_{25} + x_{26})$

$x_{Med} = \frac{1}{2}(5+5) = 5$

Modus / Modalwert: $x_{Mod} = 5$

Schiefe: Ausreißer nach oben.

Streumaße / Dispersionsmaße

Spannweite

$$\boxed{R = X_{Max} - X_{Min}}$$

Eigenschaften:

→ mind. Ordinalskala
→ Die Stichprobenwerte müssen der Größe nach sortiert werden.

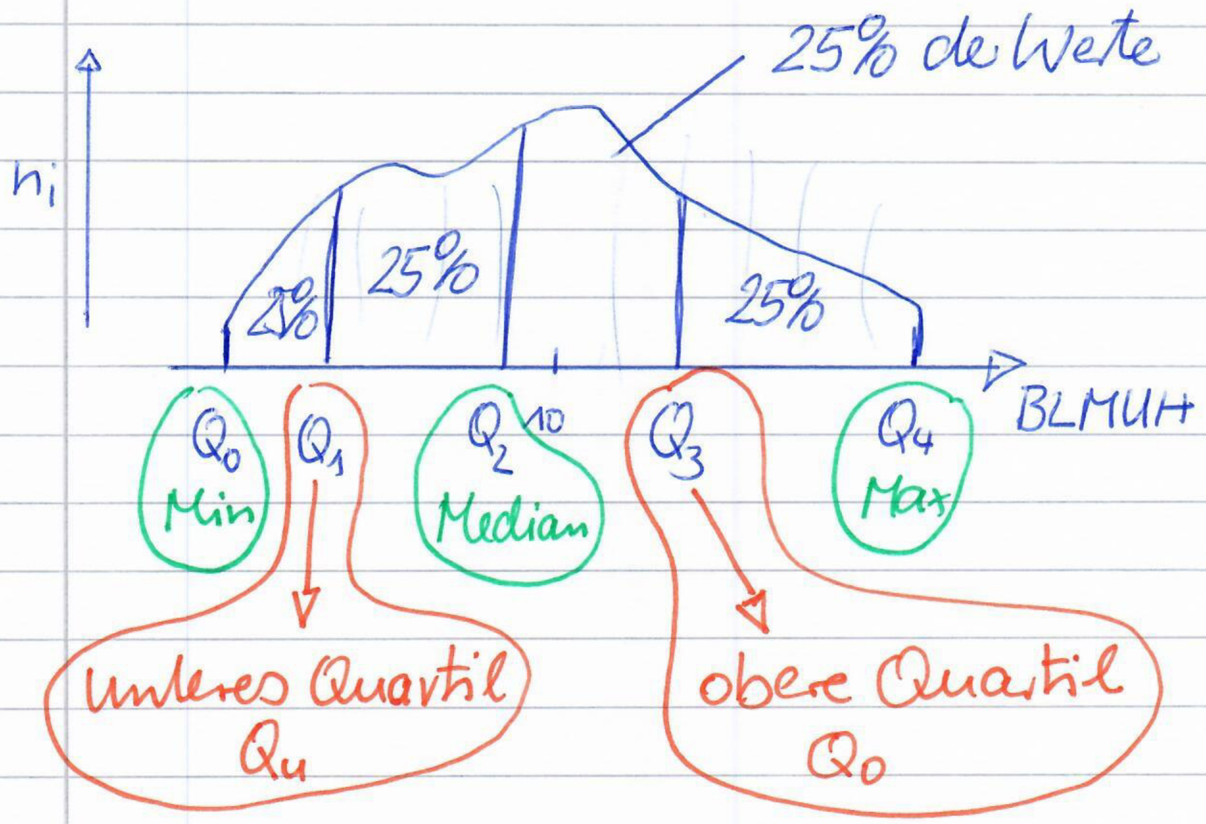
Quantil

Die gesamte Verteilung (Stichprobenwerte) werden in n gleiche Teile aufgeteilt. In jedem diese Teile befinden sich die gleiche Anzahl von Stichprobenwerten. Es soll berechnet werden wo die Grenzen der einzelnen Teile sich befinden.
Aus den entstehenden Teilen (Intervalle) können best. Schlüsse gezogen werden.

Anzahl der Intervalle:

$n=4 \rightarrow 25\%$ der Werte – $\boxed{\text{Quartil}}$
$n=5 \rightarrow 20\%$ der Werte – Quintil
$n=10 \rightarrow 10\%$ der Werte – Dezil
$n=100 \rightarrow 1\%$ der Werte – Perzentil

Quartil



Beispiel:
2, 4, 7, -20, ... (20 Stichprobe)

⇒ Die Werte müssen aufsteigend sortiert werden.
⇒ mind. Ordinalskala

-20; -4; -1; -1; 0; 1; 2; 3; 4; 4; 7; 7; 8; 11; 11; 12; 12; 15; 18; 22

$Q_0 = -20$ 0 $Q_1 = 0{,}25$ $Q_2 = 5{,}5$ $Q_3 = 11{,}75$ $Q_4 = 22$

20,25 5,25 6,25 10,25

⇒ Je breiter ein Intervall ist, desto größer ist die Streuung.

Q_0 (Min.) $= -20$

Q_4 (Max.) $= 22$

Q_2 (Median) $= \frac{1}{2}\left(X\left(\frac{n}{2}\right) + X\left(\frac{n}{2}+1\right)\right) = \frac{1}{2}\left(X_{10} + X_{11}\right)$

$= \frac{1}{2} \cdot (4+7) = 5{,}5$

Q_1 (untere Quartil) $=$ Round$(0{,}25 \cdot (n+1))$

Round$(0{,}25 \cdot (20+1))$

Round$(5{,}25) \rightarrow X_{5,25}$-Wert

$X_{5,25}$-Wert

gerundet

X_5-Wert

$\Rightarrow Q_1 = 0$

genau

$X_{5,25}$-Wert

X_5-Wert $= 0$

X_6-Wert $= 1$] Diff $= 1$

$X_{5,25} = X_5 + 0{,}25 \cdot 1 = 0{,}25$

$Q_1 = 0{,}25$

Q_3 (obere Quartil) = Round (0,75 · (n+1))

Round (0,75 · (20+1))

Round (15,75)

$x_{15,75}$-Wert

gerundet

x_{16}-Wert

$Q_3 = 12$

genau

$x_{15,75}$-Wert

$x_{15} = 11$] Diff = 1

$x_{16} = 12$

$x_{15,75} = x_{15} + 0,75 \cdot 1$

$x_{15,75} = 11,75$

$Q_3 = 11,75$

Berechnung eine bel. Größe von Intervallen.

Beispiele:

Unteres 2%iges Intervall

$$\boxed{x = \text{Round}(0{,}02 \cdot (n+1))}$$

oberes 3%-iges Intervall

$$\boxed{x = \text{Round}(0{,}97 \cdot (n+1))}$$

$\Rightarrow$ Quantil

Quantilsabstand

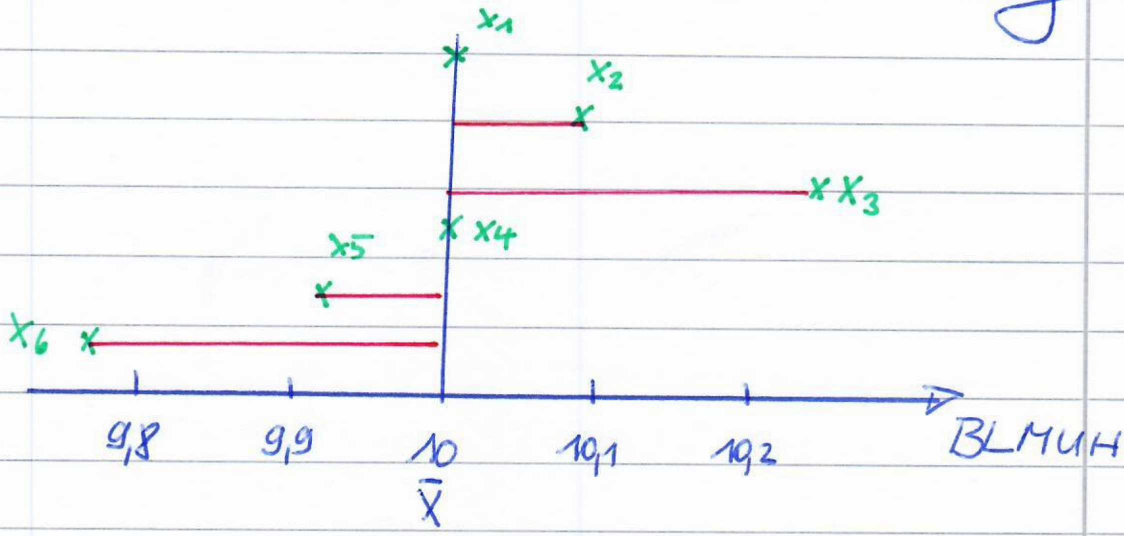
$$\boxed{QA = Q_3 - Q_1}$$

→ In diesem Intervall befinden sich 50% der Stichprobenwerte

Dezilabstand

$$\boxed{DZ = D_9 - D_1}$$

Durchschnittliche Abweichung



→ Der Abstand sollte minimal sein.

$$\boxed{x_D = \frac{1}{n} \sum_{i=1}^{n} |(x_i - \bar{x})|}$$

Beispiel: 1; 3; 6; 7; 8

$$\bar{x} = \frac{1}{5}(1+3+6+7+8) = 5$$

$$x_D = \frac{1}{5}(|1-5| + |3-5| + |6-5| + |7-5| + |8-5|)$$

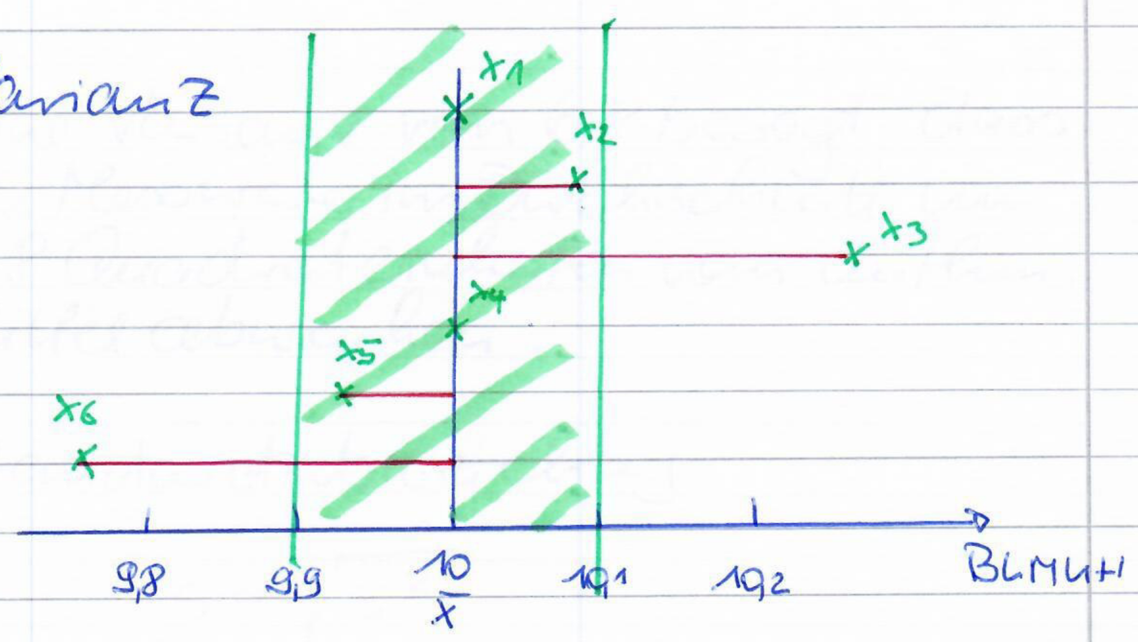
$$x_D = \frac{1}{5}(4 + 2 + 1 + 2 + 3)$$

$$x_D = 2{,}4$$

Eine durchschnittliche Abweichung von 2,4 besagt, dass die Messwerte im Durchschnitt um 2,4 Einheiten vom arithm. Mittel abweichen.

$5 \pm 2{,}4$

Varianz



→ Alle Werte die nicht in dem geforderten Intervall liegen müssen in die Berechnung stärker einfließen.
→ Diese Werte erhalten eine "Gewichtung"

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

Beispiel: 1; 3; 6; 7; 8 $\bar{x} = 5$

$s^2 = \frac{1}{5} \cdot \left((1-5)^2 + (3-5)^2 + (6-5)^2 + (7-5)^2 + (8-5)^2\right)$

$s^2 = \frac{1}{5} \cdot (16 + 4 + 1 + 4 + 9)$

4 2 1 2 3

↳ höhere Gewichtung ↳ bleibt es gleich

$s^2 = 6{,}8$

Eine Varianz von 6,8 besagt, dass die Messwerte im Durchschnitt um 6,8 Quadrateinheiten vom arithm. Mittel abweichen.

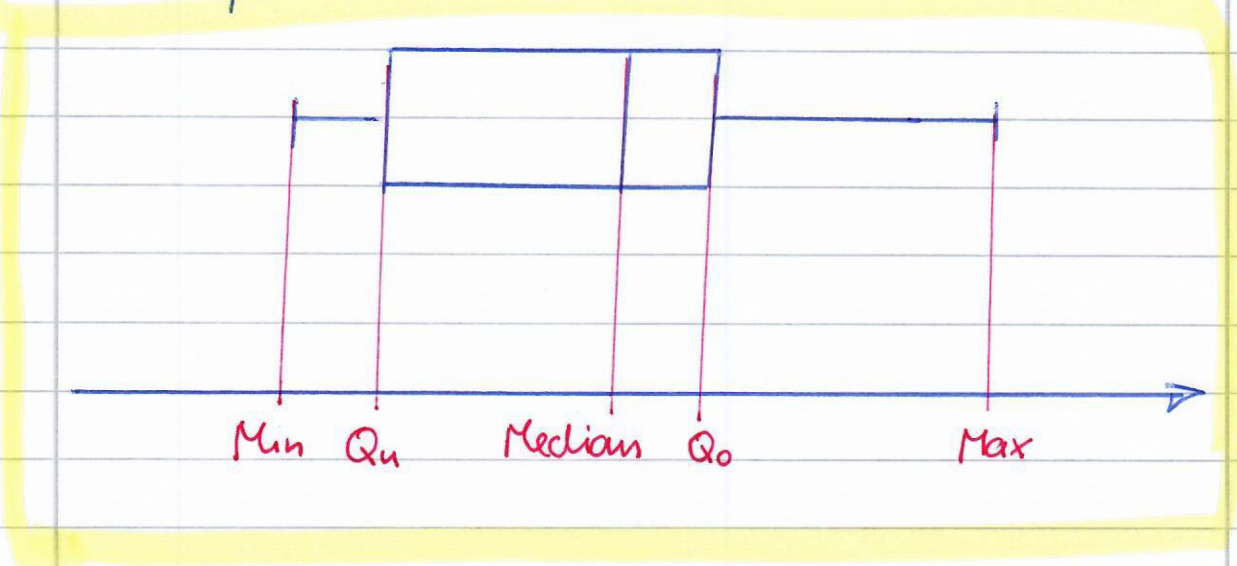
Standardabweichung

$$\boxed{s = \sqrt{s^2}}$$

Beispiel: $s = \sqrt{6{,}8} = 2{,}61$

Eine Standardabweichung von 2,61 besagt, dass die Messwerte im Durchschnitt um 2,61 Einheiten vom arithm. Mittel abweichen. Dabei sind ~~d~~ größere Abweichungen stärker gewichtet.

Boxplot



⇒ Grafische Darstellung von Lage- und Streumaßen.

Zur Übung rechnen Sie

bitte folgende Aufgaben

aus der

Aufgabensammlung

52/53/56/64/65/66/67

$5 = \sqrt{5^2} \rightarrow$ wahr / ja

Erfahrener:

$$\bar{x} = \frac{1}{10}(4\cdot 1 + 2\cdot 2 + 3 + 2\cdot 4 + 5) = 2{,}4$$

$$s^2 = \frac{1}{10}\cdot\left(4\cdot(1-2{,}4)^2 + 2(2-2{,}4)^2 + (3-2{,}4)^2 + 2(4-2{,}4)^2 + (5-2{,}4)^2\right) = 2{,}04$$

$$s = \sqrt{2{,}04} = 1{,}43$$

Unerfahrener

$$\bar{x} = \frac{1}{11}(1 + 4\cdot 2 + 6\cdot 3) = 2{,}45$$

$$s^2 = \frac{1}{11}\left((1-2{,}45)^2 + 4(2-2{,}45)^2 + 6\cdot(3-2{,}45)^2\right) = 0{,}43$$

$$s = \sqrt{0{,}43} = 0{,}66$$

→ Beim arithm. Mittel unterscheiden sich die beiden Prüfer kaum.
→ Bei der Streuung unterscheiden sich die beiden Prüfer sehr.
Erfahrene → hohe Streuung
Unerfahrene → niedrige Streuung

A: Min $= 9$; Max $= 19$; $R = 19 - 9 = 10$

Modus $= 9$;

9 9 10 11 12 14 19 Median $= 11$

$\bar{x} = \frac{1}{7} \cdot (9+9+10+11+12+14+19) = 12$

B: Min $= 11$; Max $= 14$; $R = 14 - 11 = 3$

Modus $= 11, 12$;

11 11 11 12 12 12 13 14 Median $= 12$

$\bar{x} = \frac{1}{8} (3 \cdot 11 + 3 \cdot 12 + 13 + 14) = 12$

A: rechtsschief

B: symmetrische

$$s^2 = \frac{1}{7} \cdot (2 \cdot (9-12)^2 + (10-12)^2 + (11-12)^2 + (12-12)^2 + (14-12)^2 + (19-12)^2$$

$s^2 = 10{,}86$

$s = \sqrt{10{,}86} = 3{,}30$

$$s^2 = \frac{1}{8} \cdot (3 \cdot (11-12)^2 + 3 \cdot (12-12)^2 + (13-12)^2 + (14-12)^2$$

$s^2 = 1$; $s = 1$

$$v = \frac{1}{8} \cdot \left(3 \cdot \left(\frac{11-12}{1}\right)^3 + 3 \cdot \left(\frac{12-12}{1}\right)^3 + \left(\frac{13-12}{1}\right)^3 + \left(\frac{14-12}{1}\right)^3 \right)$$

$$v = \frac{1}{8} \cdot (-3 + 0 + 1 + 8) = \frac{6}{8} = \frac{3}{4}$$

$\Rightarrow v > 0$ (rechtsschief)

Obere Quartil

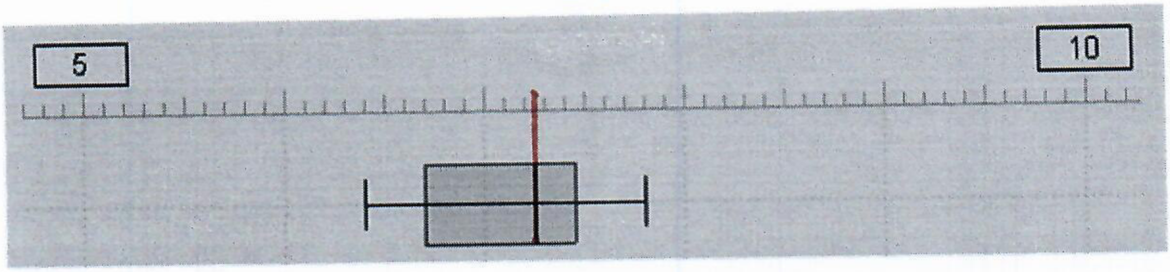
25 50 51 52 54 55 56 56 56 56 58

$Q_o = Round(0{,}75 \cdot (11+1) = 9$

x_9-Wert $= 56$

$Q_o = 56$

$\Rightarrow$ Aussage stimmt nicht



Median = 7,25

unteres Quartil:

$$x = \text{Round}(0{,}25 \cdot (50+1))$$

$$x = \text{Round}(12{,}75$$

$x_{12,75}$-Wert

gerundet

x_{13}-Wert:

$Q_u = 3$

genau

$x_{12,75}$-Wert

$x_{12} = 3$

$x_{13} = 3$

$Q_u = 3$

obere Quartil:

$$x = \text{Round}(0{,}75 \cdot (50+1)$$

$$x = \text{Round}(38{,}25)$$

$x_{38,25}$-Wert

gerundet

x_{39}-Wert $= 12$

$Q_o = 12$

genau

$x_{38,25}$-Wert

$x_{38} = 12$

$x_{39} = 12$

$Q_o = 12$

Variationskoeffiziant

Der Variationskoeffizient misst die Variation im Vergleich zum Mittelwert.

$$\boxed{V = \frac{s}{\bar{x}} \cdot 100\%}$$

→ Zum vergleichen versch. Messunge mit unterschiedlichen ~~Eigen~~ Einheiten.