Datenerhebung

– Beispiel: Mediennutzung von Jugendlichen

Kennnummer / Fall	Alter	TV-Nutzungsdauer in h pro Woche	TV informativ
1	16	23	1
2	15	23	1
3	14	28	1
4	13	8	1
5	15	18	0
⋮	⋮	⋮	⋮

Datenerhebung

– Beispiel: Bücher lesen – Häufigkeitstabelle für die 2.945 untersuchten Personen

Merkmalsausprägung	Codierung	Absolute Häufigkeiten	Relative Häufigkeiten	Kumulierte absolute Häufigkeiten	Kumulierte relative Häufigkeiten
Nie	$x_1 = 0$	$n_1 = 550$	$r_1 = \frac{550}{2945} = 0{,}1868$	$\sum_{i=1}^{1} n_i = 550$	$\sum_{i=1}^{1} r_i = \frac{550}{2945} = 0{,}1868$
Weniger als einmal im Monat	$x_2 = 1$	$n_2 = 892$	$r_2 = \frac{892}{2945} = 0{,}3029$	$\sum_{i=1}^{2} n_i = 1442$	$\sum_{i=1}^{2} r_i = \frac{1442}{2945} = 0{,}4896$
Mind. einmal im Monat, aber nicht mehr als einmal die Woche	$x_3 = 2$	$n_3 = 375$	$r_3 = \frac{375}{2945} = 0{,}1273$	$\sum_{i=1}^{3} n_i = 1817$	$\sum_{i=1}^{3} r_i = \frac{1817}{2945} = 0{,}6170$
Mind. einmal pro Woche, aber nicht täglich	$x_4 = 3$	$n_4 = 589$	$r_4 = \frac{589}{2945} = 0{,}2000$	$\sum_{i=1}^{4} n_i = 2406$	$\sum_{i=1}^{4} r_i = \frac{2406}{2945} = 0{,}8170$
Täglich	$x_5 = 4$	$n_5 = 539$	$r_5 = \frac{539}{2945} = 0{,}1830$	$\sum_{i=1}^{5} n_i = 2945$	$\sum_{i=1}^{5} r_i = \frac{2945}{2945} = 1{,}0000$

Statistische Kennwerte

14 Personen verbringen einen Abend zusammen in der Bar, die Variable B gibt an, wie viele Biere sie am Ende insgesamt getrunken haben.

Person (n = 14)	Roman	Andi	Tom	Anika	Tanja	Svenja	Lukas	Maxi	Simone	Franzi	Tim	Finnja	Dennis	Alicia
B	8	0	5	3	3	10	4	8	6	4	13	4	5	4

Statistische Kennwerte

Minimum & Maximum

- Minimum: Kleinster Wert, den die Variable annimmt
- Maximum: Größter Wert, den die Variable annimmt

Person (n = 14)	Roman	Andi	Tom	Anika	Tanja	Svenja	Lukas	Maxi	Simone	Franzi	Tim	Finnja	Dennis	Alicia
B	8	0	5	3	3	10	4	8	6	4	13	4	5	4

- Minimum: 0
- Maximum: 13

Statistische Kennwerte

- Absolute & Relative Häufigkeiten

B	Absolute Häufigkeiten	Relative Häufigkeiten
0	1	1/14
3	2	2/14 = 1/7
4	4	4/14 = 2/7
5	2	2/14 = 1/7
6	1	1/14
8	2	2/14 = 1/7
10	1	1/14
13	1	1/14

Statistische Kennwerte

Modus

- Ausprägung auf die mehr Einheiten entfallen, als auf alle anderen Ausprägungen

B	Absolute Häufigkeiten	Relative Häufigkeiten
0	1	1/14
3	2	2/14 = 1/7
4	4	4/14 = 2/7
5	2	2/14 = 1/7
6	1	1/14
8	2	2/14 = 1/7
10	1	1/14
13	1	1/14

- Modus: 4

Statistische Kennwerte

Arithmetisches Mittel

- Landläufig bekannt als „der Durchschnitt“
- Die Variablenwerte aller Einheiten werden addiert und abschließend durch die Stichprobengröße geteilt
- $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
- $\bar{x} = \frac{1}{n} (\sum_{j=1}^{k} n_j \cdot x_j)$

Statistische Kennwerte

- Median

  - Der Wert, der in der Mitte einer Datenverteilung liegt
  - Die eine Hälfte der Variablenwerte ist größer, die andere ist kleiner

  - $x_{med} := \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{für} \quad \text{n ungerade} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & \text{für} \quad \text{n gerade} \end{cases}$

Statistische Kennwerte

– Geordnete Urliste:

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14
B	0	3	3	4	4	4	4	5	5	6	8	8	10	13

$$\frac{1}{2}\left(x\left(\frac{14}{2}\right) + x\left(\frac{14}{2} + 1\right)\right) =$$

$$\frac{1}{2}\left(x(7) + x(8)\right) =$$

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

4,5

Statistische Kennwerte

Quantile

- Der Median ist das 0,5-Quantil (p = 0,5)
- Quantile lassen sich jedoch auch für jedes andere p bilden
- Das 0,25-Quantil und das 0,75-Quantil nennt man Quartile
- $\tilde{x}_p := \begin{cases} x^{(i)}\ , falls\ np\ nicht\ ganzzahlig\ mit\ der\ kleinsten\ ganzen\ Zahl\ i > np \\ \frac{1}{2}\left(x^{(np)} + x^{(np+1)}\right), falls\ np\ ganzzahlig \end{cases}$

n × p = 14 × 0,25 = 3,5 (nächstgrößerer Wert ist die 4)

$X^{0,25}$ = **4**

n × p = 14 × 0,75 = 10,5 (nächstgrößerer Wert ist die 11)

$X^{0,75}$ = **8**

Statistische Kennwerte

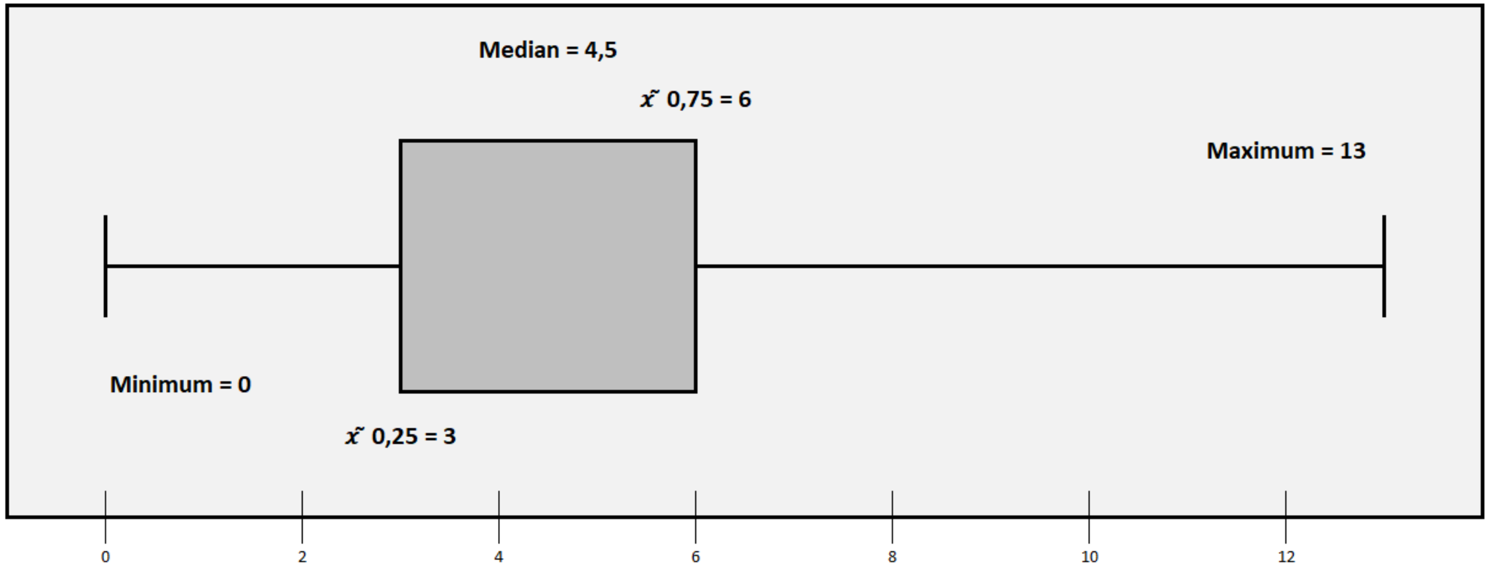
Quartilsabstand

- d_Q
- $\tilde{x}_{0,75} - \tilde{x}_{0,25}$

- $d_Q = 6 - 3 = 3$

Statistische Kennwerte

Boxplot



Statistische Kennwerte

Zäune zur Ermittlung des modifzierten Boxplots

- $Z_U = \tilde{x}_{0,25} - 1{,}5\ d_Q$
- $Z_O = \tilde{x}_{0,75} + 1{,}5\ d_Q$

$Z_U = \tilde{x}_{0,25} - 1{,}5\ d_Q = 3 - 1{,}5 \times 3 = 0$

$Z_O = \tilde{x}_{0,75} + 1{,}5\ d_Q = 6 + 4{,}5 = 10{,}5$

Statistische Kennwerte

Spannweite

- Abstand zwischen kleinstem und größtem Wert
- $R = x_{max} - x_{min} = x_n - x_{(1)}$

Spannweite: 13-0 =**13**

Statistische Kennwerte

Varianz und Standardabweichung

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

$$S = \sqrt{S^2}$$

$S^2 = 10{,}88$

$S = 3{,}3$

Uebungs aufgaben

Aufgabe 1

Elf Journalisten wurden zu ihrem Jahreseinkommen (in Tsd. Euro) befragt und haben folgende Zahlen angegeben:

Journalist j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Jahreseinkommen X	70	110	75	55	55	90	90	140	70	65	90

– Bestimme alle bekannten Lage- sowie Streuungsmaße und stelle beide Arten des Boxplots dar.

Aufgabe 1 - Lösung

geordnete Urliste bilden!

1	2	3	4	5	6	7	8	9	10	11
55	55	65	70	70	75	90	90	90	110	140

Lagemaße + Streuungsmaße

Lagemaße:
- min
- $\tilde{x}_{0,25}$
- $\tilde{x}_{med} = \tilde{x}_{0,5}$
- $\tilde{x}_{0,75}$
- max
- Modus
- dQ
- zu
- zo
- R

Streuungsmaße:
- arith. Mittel $= \bar{x}$
- Standardabweichung $= s$
- Varianz $= s^2$

Streuungsmaße:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i =$$

$$= \frac{1}{11} \cdot (55+55+65+70+70+75+90+90+90+110+140) =$$

$$= \underline{\underline{82,73}}$$

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

$$= \frac{1}{11-1} \left[(55-82,73)^2 + \ldots + (140-82,73)^2 \right] =$$

$$= \underline{\underline{641,82}}$$

$$s = \sqrt{s^2} = \sqrt{641,82} = \underline{\underline{25,33}}$$

Lagemaße:

$min = 55$

$max = 140$

aufrunden!

$n \cdot p = 11 \cdot 0,25 = 2,75$ $\quad \tilde{x}_{0,25} = x_{(3)} = 65$

$\tilde{x}_{med} = x_{\left(\frac{11+1}{2}\right)} = x_{(6)} = 75$

aufrunden

$n \cdot p = 11 \cdot 0,75 = 8,25$ $\quad \tilde{x}_{0,75} = x_{(9)} = 90$

Modus: 90 (kommt 3x vor)

$R = x_{max} - x_{min} = x_{11} - x_1 =$
$= 140 - 55 = 85$

$dQ = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 90 - 65 =$
$= 25$

$z_u = \tilde{x}_{0,25} - 1,5\, dQ =$
$= 65 - 1,5 \cdot 25 = 27,5$

$z_o = \tilde{x}_{0,75} + 1,5\, dQ =$
$= 127,50$

Aufgabe 2.1 der AS

Elf Franchise-Filialen einer Fast-Food-Kette erzielen 2020 folgende Umsätze (in Tsd. Euro):

Filiale j	1	2	3	4	5	6	7	8	9	10	11
Umsatz X	140	90	90	55	90	75	70	55	70	10	70

a) Geben Sie das arithmetische Mittel, die (empirische) Standardabweichung und den Variationskoeffizienten an.
b) Bestimmen Sie rechnerisch das untere und obere Quartil sowie den Median. Zeichnen Sie den zugehörigen (einfachen) Boxplot.

Aufgabe 2.1 der AS - Lösung

x_j	1	2	3	4	5	6	7	8	9	10	11
Umsatz X	10	55	55	70	70	70	75	90	90	90	140

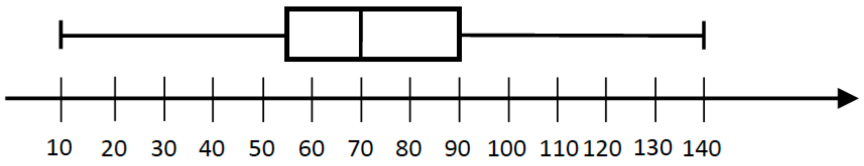
- $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{11} \cdot (10 + 55 + 55 + 70 + 70 + 70 + 75 + 90 + 90 + 90 + 140) = \frac{815}{11} \approx 74{,}09$
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{11-1} \cdot (10 - \frac{815}{11})^2 + (55 - \frac{815}{11})^2 + (55 - \frac{815}{11})^2 + \dots + (140 - \frac{815}{11})^2 \approx 999{,}09$

  $S = \sqrt{S^2} \approx 31{,}61$
- $v = \frac{s}{\bar{x}} = \frac{31{,}61}{74{,}09} \approx 0{,}43$

b)

- $\tilde{x}_{0,25} = x_3 = 55$
- $\tilde{x}_{0,75} = x_9 = 90$
- $x_{med} = x_{(\frac{11+1}{2})} = x_6 = 70$

c)



Tutorium – Statistik für JoKo

Aufgabe 2.2 der AS

Bei einer Klausur sind zwölf Aufgaben zu bearbeiten, wobei pro Aufgabe ein Punkt erzielt werden kann. Als *nicht bestanden* gilt eine Klausur, wenn ein Kandidat weniger als 5 Punkte erreicht. Die Korrektur einer Klausur ergibt folgende Häufigkeitsverteilung der erreichten Punktzahlen a_j.

a_j	0	1	2	3	4	5	6	7	8	9	10	11	12
$h(a_j)$	6	2	5	0	9	6	16	24	13	8	7	4	0

a) Berechnen Sie das arithmetische Mittel, den Modus und den Median.
b) Berechnen Sie außerdem die Spannweite, den Interquartilsabstand, die Varianz, die Standardabweichung und den Variationskoeffizienten.
c) Zeichnen Sie einen Boxplot.

Aufgabe 2.2 der AS - Lösung

geordnete Urliste! ← beginnt immer bei 1!

1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	2	4	5	6	6	7	8	9	13	16	24

a) $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{13} \cdot (0 + 0 + 2 + \ldots + 24) = 7{,}69$

$x_{mod} =$ 0 ∧ 6, da sie jeweils 2mal vorkommen!

welcher Wert kommt am häufigsten vor?

$x_{med} = x^{(7)} = 6$

$n \cdot p = 13 \cdot 0{,}5 = 6{,}5 = 7$

b) Spannweite $= x_{max} - x_{min} = 24 - 0 = 24$

$\tilde{x}_{0,25} = x^{(4)} = 4$

$n \cdot p = 13 \cdot 0{,}25 = 3{,}25$

$\tilde{x}_{0,75} = x^{(10)} = 9$

$n \cdot p = 13 \cdot 0{,}75 = 9{,}75$

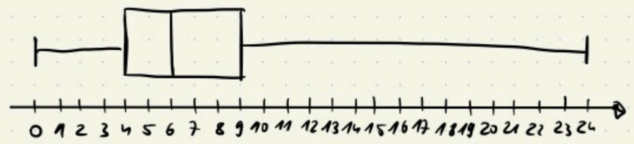
$d_Q = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 9 - 4 = 5$

$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{12} \cdot [(0 - 7{,}69)^2 + \ldots (24 - 7{,}69)^2] = 45{,}23$

$s = \sqrt{s^2} = \sqrt{45{,}23} = 6{,}73$

$v = \frac{s}{\bar{x}} = \frac{6{,}73}{7{,}69} = 0{,}875$

c)



Aufgabe 2.3 der AS

Für die Bevölkerung der Regionen der Erde werden folgende Flächenproportionen - d.h. Bewohner je Quadratkilometer – angegeben: 94 96 95 96 98 97 102 97

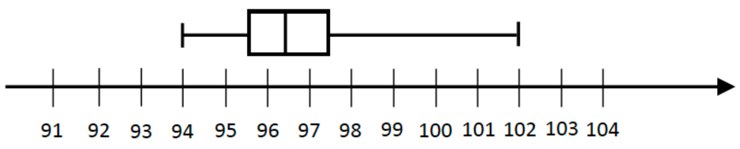
Erstellen Sie einen einfachen und einen modifizierten Boxplot.

Aufgabe 2.3 der AS - Lösung

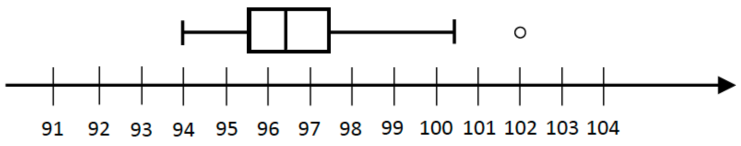
x_j	1	2	3	4	5	6	7	8
	94	95	96	96	97	97	98	102

- $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{8}(94 + 95 + 96 + \ldots + 102) = \frac{775}{8} \approx 96{,}875$
- $x_{med} = \frac{1}{2}\left(x_{(\frac{8}{2})} + x_{(\frac{8}{2}+1)}\right) = \frac{1}{2}(x_4 + x_5) = \frac{1}{2} \cdot (96 + 97) = 96{,}5$
- $\tilde{x}_{0,25} = \frac{1}{2}(x_{8\cdot 0,25} + x_{8\cdot 0,25+1}) = \frac{1}{2}(x_2 + x_3) = \frac{1}{2}(95 + 96) = 95{,}5$

  $\tilde{x}_{0,75} = \frac{1}{2}(x_{8\cdot 0,75} + x_{8\cdot 0,75+1}) = \frac{1}{2}(x_6 + x_7) = \frac{1}{2}(97 + 98) = 97{,}5$



- $d_Q = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 97{,}5 - 95{,}5 = 2$
- $z_u = 95{,}5 - 1{,}5 \cdot 2 = 92{,}5 \rightarrow z_u = 94$
- $z_o = 97{,}5 + 1{,}5 \cdot 2 = 100{,}5$



Aufgabe 2.4 der AS

Die Evaluation des Partyverhaltens des Statistikkurses ergibt, dass 32 Personen abends immer zu Hause bleiben. 78 Befragte gehen einmal pro Woche, 112 zweimal pro Woche, 142 dreimal pro Woche und 212 mindestens viermal aus.

a) Ermitteln Sie den Modus, den Median und das arithmetische Mittel des Merkmals „Anzahl der Tage des abendlichen Ausgehens pro Woche“ der befragten Studierenden.
b) Ermitteln Sie die mittlere quadratische Abweichung der „Anzahl der Tage des abendlichen Ausgehens pro Woche“.

Aufgabe 2.5 der AS

Die 30 Arbeitnehmer einer Gewerkschaft erhalten folgende Bruttomonatslöhne X:

X	1000	1400	2000	2500	4100
h(X)	4	8	10	5	3

a) Geben Sie die drei Quartile und den Interquartilsabstand der Löhne an.
b) Zeichnen Sie die modifizierte Form des Boxplots, die üblicherweise in der Praxis verwendet wird.

Aufgabe 2.5 der AS - Lösung

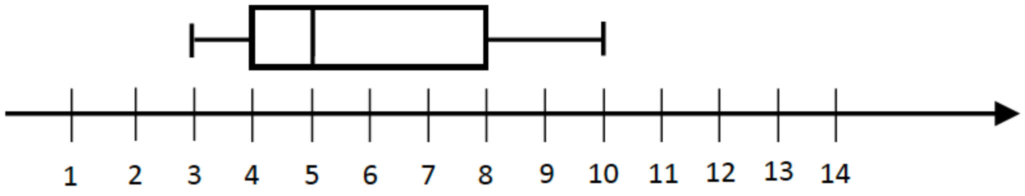
x_j	1	2	3	4	5
	3	4	5	8	10

a)

- $\tilde{x}_{0,25} = x_2 = 4$
- $\tilde{x}_{0,50} = x_3 = 5 = x_{med}$
- $\tilde{x}_{0,75} = x_4 = 8$

b)

- $d_Q = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 8 - 4 = 4$
- $z_u = 4 - 1{,}5 \cdot 4 = -2 \rightarrow z_u = 3$
- $z_o = 8 + 1{,}5 \cdot 4 = 14 \rightarrow z_o = 10$



it`s time for a quiz!