MODUL C - LÖSUNGEN

Aufgabe C.1 – Lösung

Um die Grafik zu erstellen müssen die geordneten Daten verwendet werden. Basierend auf der von Benjamin & Cornell vorgeschlagenen Faustregel (Vorlesungsskript Faber, Gleichung C.8) kann die Anzahl der Intervalle auf 6 festgelegt werden. Tabelle C.1.2 zeigt die Übersicht über die erhobenen Daten.

<u>Histogramm ausschliesslich für Richtung 1:</u>

Die maximale Beobachtung in Richtung 2 ist 35852 und die minimale Beobachtung ist 24846. Die Länge der Intervalle kann daher wie folgt gewählt werden:

$$\frac{35852-24846}{6}=1834\approx 2000$$

Die 6 Intervalle werden eingeteilt in:

$$\begin{aligned}
&24000-26000, && \text{Intervallschwerpunkt} = 25000\\
&26000-28000, && \text{Intervallschwerpunkt} = 27000\\
&28000-30000, && \text{Intervallschwerpunkt} = 29000\\
&30000-32000, && \text{Intervallschwerpunkt} = 31000\\
&32000-34000, && \text{Intervallschwerpunkt} = 33000\\
&34000-36000, && \text{Intervallschwerpunkt} = 35000
\end{aligned}$$

Tabelle C.1.2: *Übersicht über die beobachteten Verkehrsflüsse für Richtung 1*

	Intervall [Anzahl der Autos *10^3]	Intervall-Mittelpunkt [Anzahl der Autos *10^3]	absolute Häufigkeit [-]	relative Häufigkeit [-]	kumulative rel. Häufigkeit [-]
Richtung 1	24.0-26.0	25.0	3	0.10	0.10
	26.0-28.0	27.0	0	0.00	0.10
	28.0-30.0	29.0	3	0.10	0.20
	30.0-32.0	31.0	2	0.07	0.27
	32.0-34.0	33.0	13	0.43	0.70
	34.0-36.0	35.0	9	0.30	1.00

<u>Histogramm zum Vergleich der Verkehrsdaten beider Richtungen:</u>

Die maximale Beobachtung beider Richtungen ist 35852 und die minimale Beobachtung ist 17805. Um beide Histogramme miteinander vergleichen zu können, muss die Intervalleinteilung auf der x-Achse identisch sein. Hierfür eignet sich folgende Einteilung in 10 Intervalle:

$$16000-18000, \quad \text{Intervallschwerpunkt} = 17000$$
$$18000-20000, \quad \text{Intervallschwerpunkt} = 19000$$
$$20000-22000, \quad \text{Intervallschwerpunkt} = 21000$$
$$22000-24000, \quad \text{Intervallschwerpunkt} = 23000$$
$$24000-26000, \quad \text{Intervallschwerpunkt} = 25000$$
$$26000-28000, \quad \text{Intervallschwerpunkt} = 27000$$
$$28000-30000, \quad \text{Intervallschwerpunkt} = 29000$$
$$30000-32000, \quad \text{Intervallschwerpunkt} = 31000$$
$$32000-34000, \quad \text{Intervallschwerpunkt} = 33000$$
$$34000-36000, \quad \text{Intervallschwerpunkt} = 35000$$

Tabelle C.1.2: *Übersicht über die beobachteten Verkehrsflüsse für beide Richtungen*

	Intervall [Anzahl der Autos *10^3]	Intervall-Mittelpunkt [Anzahl der Autos *10^3]	absolute Häufigkeit [-]	relative Häufigkeit [-]	kumulative rel. Häufigkeit [-]
Richtung 1	16.0-18.0	17.0	0	0.00	0.00
	18.0-20.0	19.0	0	0.00	0.00
	20.0-22.0	21.0	0	0.00	0.00
	22.0-24.0	23.0	0	0.00	0.00
	24.0-26.0	25.0	3	0.10	0.10
	26.0-28.0	27.0	0	0.00	0.10
	28.0-30.0	29.0	3	0.10	0.20
	30.0-32.0	31.0	2	0.07	0.27
	32.0-34.0	33.0	13	0.43	0.70
	34.0-36.0	35.0	9	0.30	1.00
	Intervall [Anzahl der Autos *10^3]	Intervall-Mittelpunkt [Anzahl der Autos *10^3]	absolute Häufigkeit [-]	relative Häufigkeit [-]	kumulative rel. Häufigkeit [-]
Richtung 2	16.0-18.0	17.0	1	0.03	0.03
	18.0-20.0	19.0	2	0.07	0.10
	20.0-22.0	21.0	2	0.07	0.17
	22.0-24.0	23.0	3	0.10	0.27
	24.0-26.0	25.0	1	0.03	0.30
	26.0-28.0	27.0	3	0.10	0.40
	28.0-30.0	29.0	7	0.23	0.63
	30.0-32.0	31.0	10	0.33	0.97
	32.0-34.0	33.0	1	0.03	1.00
	34.0-36.0	35.0	0	0	1.00

Abbildung C.1.1 und Abbildung C.1.2 zeigen die relative Häufigkeitsverteilung der Verkehrsflüsse.

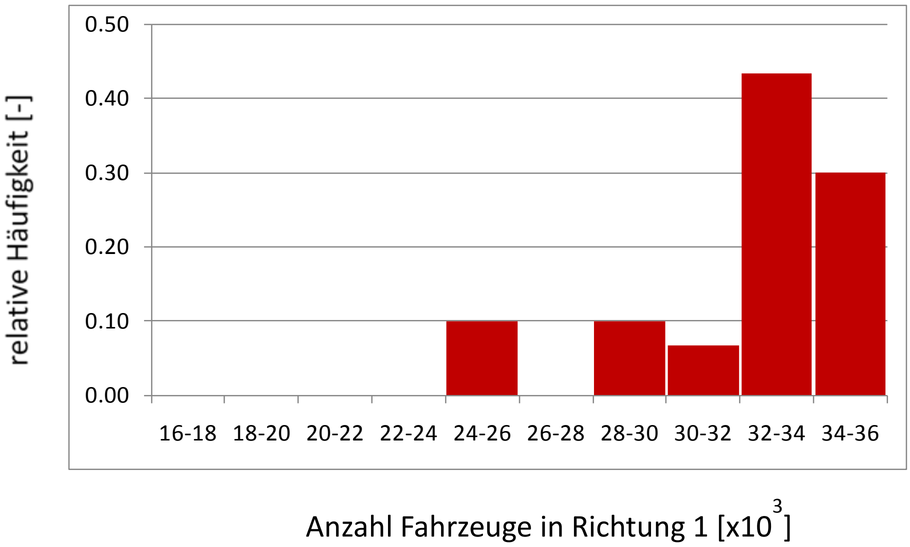


Abbildung C.1.1: *Histogramm der relativen Häufigkeitsverteilung der Anzahl der Fahrzeuge in Richtung 1.*

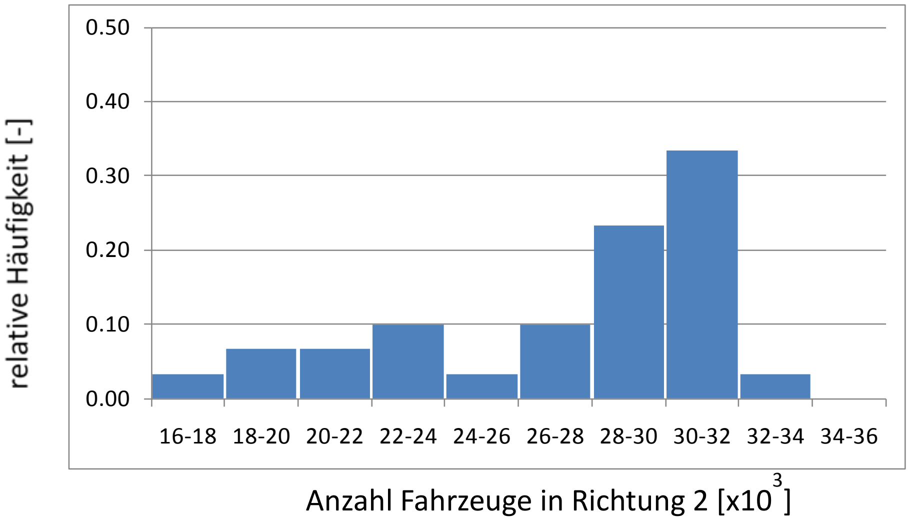


Abbildung C.1.2: *Histogramm der relativen Häufigkeitsverteilung der Anzahl der Fahrzeuge in Richtung 2.*

Abbildung C.1.3 veranschaulicht die kumulierten relativen Häufigkeitsdiagramme der Verkehrsdaten in beide Richtungen, basierend auf den Intervallen aus Tabelle C.1.2.

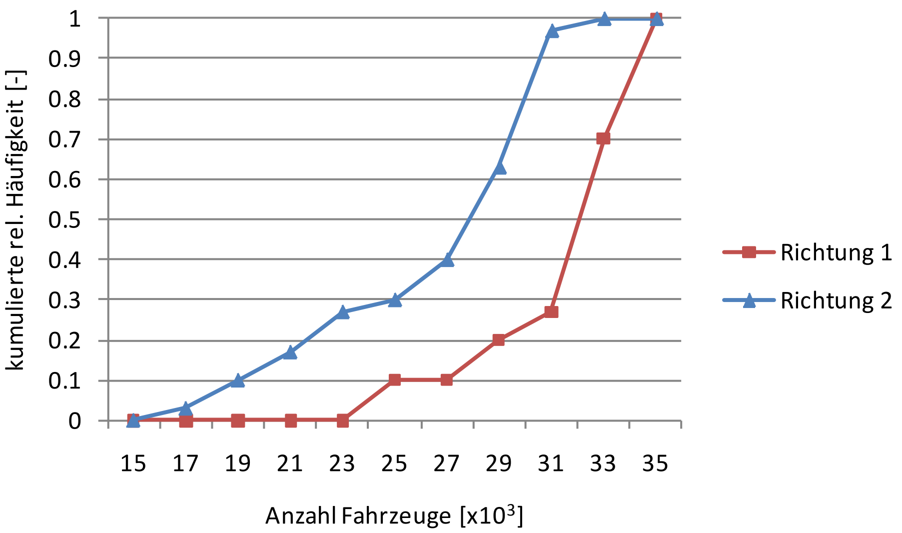


Abbildung C.1.3: *Kumulierte relative Häufigkeitsdiagramme beider Verkehrsrichtungen auf Basis von 7 Intervallen.*

Die einzelnen kumulierten relativen Häufigkeiten (y-Achse) wurden über den Intervallmittelpunkten (x-Achse) aufgetragen. Es ist zu beachten, dass mit der Wahl der Klassenanzahl (hier 10) eine erhebliche Datenreduktion stattfinden kann und somit auch Information verloren gehen könnte.

Aus dieser graphischen Präsentation der Daten ist ersichtlich, dass der Verkehrsfluss in Richtung 2 geringer als der in Richtung 1 ist. Dies ist dadurch zu erkennen, dass die kumulative relative Häufigkeitsverteilung der Datenreihe Richtung 1 nach rechts verschoben ist.

Aufgabe C.2 - Lösung

Um den Tukey Box Plot zu zeichnen, werden 5 Kennwerte benötigt, die die zentrale Tendenz der Datenreihen beschreiben (Vorlesungsskript Faber, Tabelle C.8):

- unteres Quartil
- unterer Nachbarschaftswert
- Median
- oberer Nachbarschaftswert
- oberes Quartil

Betrachten wir den Verkehrsfluss in Richtung 1. Basierend auf Gleichung C.10 des Vorlesungsskriptes wird der Wert wie folgt berechnet:

$$i = \upsilon(n+1) = n\upsilon + \upsilon$$

Für das untere Quartil (d.h. 0.25-Quantil) ergibt sich:

$$i = 30 \cdot 0.25 + 0.25 = 7.75$$

Der resultierende Wert teilt sich in einen ganzzahligen-Teil $k = 7$ und in eine reelle Zahle zwischen 0 und 1: $p = 0.75$. Daher ist der Interpolationswert x_i^o :

$$x_i^o = (1-p)x_7^o + px_{7+1}^o = (1-0.75)\cdot 30035 + 0.75 \cdot 30613 = 30468.5 \approx 30469 \text{ Autos}$$

Für das obere Quantil (d.h. 0.75-Quantil) ergibt sich:

$$i = 30 \cdot 0.75 + 0.75 = 23.25$$

Mit Hilfe von Tabelle 2.1.1 bekomme

n wir:

$$x_i^o = (1-p)x_{23}^o + px_{23+1}^o = (1-0.25)\cdot 34013 + 0.25 \cdot 34076 = 34028.75 \approx 34029 \text{ Autos}$$

Tabelle C.2.1: *Daten der Verkehrsbeobachtung auf der Rosengartenstrasse.*

Datum	Richtung 1	Richtung 2	Richtung 1 sortiert	Richtung 2 sortiert
01.04.2001	32618	24609	24846	17805
02.04.2001	33380	29965	24862	18123
03.04.2001	34007	30629	25365	19735
04.04.2001	33888	30263	28252	20903
05.04.2001	35237	31405	29224	21145
06.04.2001	35843	31994	29976	22762
07.04.2001	33197	26846	30035	22828
08.04.2001	30035	22762	30613	23141
09.04.2001	32158	30366	32158	24609
10.04.2001	33406	29994	32472	26525
11.04.2001	34576	30958	32618	26846
12.04.2001	34013	30680	32962	27746
13.04.2001	24846	19735	33091	28117
14.04.2001	28252	21145	33197	28858
15.04.2001	25365	17805	33198	28877
16.04.2001	24862	18123	33245	29080
17.04.2001	32472	28117	33380	29586
18.04.2001	33245	28858	33406	29965
19.04.2001	33788	29080	33788	29994
20.04.2001	34076	30313	33888	30263
21.04.2001	29976	23141	33937	30313
22.04.2001	29224	20903	34007	30366
23.04.2001	32962	27746	34013	30629
24.04.2001	33937	29586	34076	30680
25.04.2001	33198	30788	34425	30788
26.04.2001	34455	31074	34455	30958
27.04.2001	35852	32384	34576	31074
28.04.2001	33091	26525	35237	31405
29.04.2001	30613	22828	35843	31994
30.04.2001	34425	28877	35852	32384

Der Median ergibt sich aus folgender Beziehung:

$$i = 30 \cdot 0.5 + 0.5 = 15.5$$

$$x_i^o = (1-p)x_{15} + px_{15+1} = (1-0.5)*33198 + 0.5*33245 = 33221.5 \approx 33222 \text{ Autos}$$

Um die Nachbarschaftswerte zu berechnen wird zuerst die interquartile Differenz ermittelt:

$$r = Q_{0.75} - Q_{0.25} = 34029 - 30469 = 3560$$

Der untere Nachbarschaftswert ist die kleinste Beobachtung grösser/gleich dem unteren Grenzwert. Dieser berechnet sich wie folgt:

$$Q_{0.25} - 1.5r = 30469 - 1.5 \cdot 3560 = 25129$$

Aus Tabelle C.2.1 kann somit der untere Nachbarschaftswert mit **25365** abgegriffen werden.

Auf gleiche Weise wird der obere Grenzwert berechnet:

$$Q_{0.75} + 1.5r = 34029 + 1.5 \cdot 3560 = 39369$$

Von Table C.2.1 kann somit der obere Nachbarschaftswert abgelesen werden. Dies ist der grösste Wert der geordneten Datenreihe kleiner/gleich 39369. Das ist 35852. Dieser Wert entspricht in diesem Fall dem Maximum der Datenreihe.
Tabelle C.2.2 zeigt eine Übersicht für beide Datenreihen. Die Daten für den Verkehrsfluss in Richtung 2 weisen hierbei keine Ausreisser auf.

Tabelle C.2.2: *Statistische Werte für den Tukey-Box-Plot*

Kennwerte	Richtung 1	Richtung 2
unteres Quartil	30469	23063
unterer Nachbarschaftswert	25365	17805
Median	33222	28979
oberer Nachbarschaftswert	35852	32384
oberes Quartil	34029	30642
Ausreisser	24846 24862	

Abbildung C.2.1 zeigt die Tukey Box Plots für beide Richtungen. Es ist ersichtlich, dass alle Masszahlen der Datenreihe 1 tendenziell grösser sind als die entsprechenden Werte der Datenreihe 2. Ebenso ist ersichtlich, dass die Daten nicht symmetrisch verteilt sind, und die oberen Balken kürzer als die unteren sind.

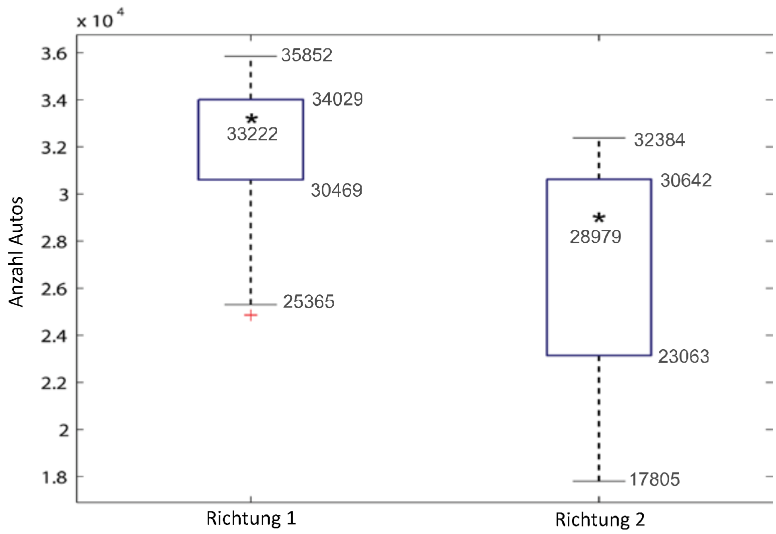


Abbildung C.2.1: *Tukey Box Plot der Verkehrsflüsse in der Rosengartenstrasse.*

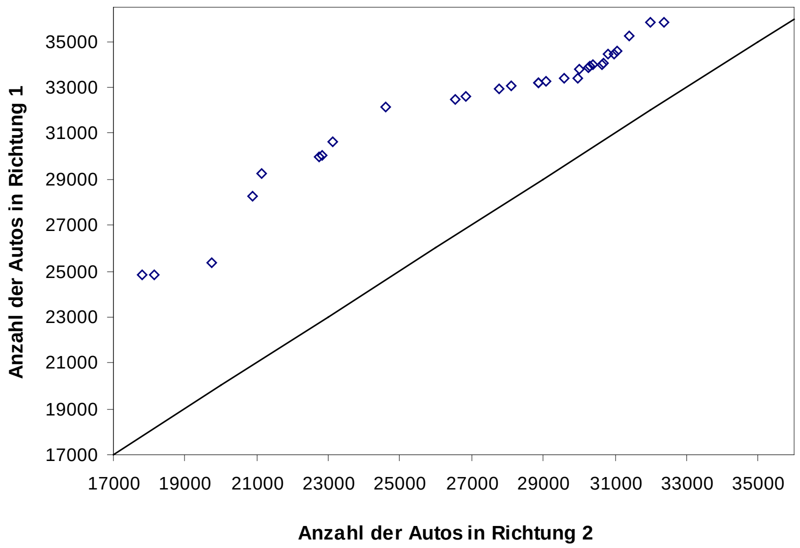
Aufgabe C.3 - Lösung

Bei der Erstellung des Q-Q-Plots muss zuerst die Anzahl jeder Datenreihe beachtet werden. Für beide Datenreihen beträgt sie in diesem Fall jeweils 30 Beobachtungstage. Damit ist der Q-Q-Plot ein einfaches Auftragen der beiden beobachteten Datenreihen gegeneinander (Abbildung C.3.1). Für den Mittelwert-Differenz-Plot ist zunächst die Differenz und der Mittelwert der Daten in beiden Verkehrsrichtungen zu erstellen (Tabelle C.3.1).

Tabelle C.3.1: *Daten für den Mittelwert-Differenz-Plot.*

Richtung 2	Richtung 1	y_i-x_i	$(y_i+x_i)/2$
17805	24846	7041	21325.5
18123	24862	6739	21492.5
19735	25365	5630	22550.0
20903	28252	7349	24577.5
21145	29224	8079	25184.5
22762	29976	7214	26369.0
22828	30035	7207	26431.5
23141	30613	7472	26877.0
24609	32158	7549	28383.5
26525	32472	5947	29498.5
26846	32618	5772	29732.0
27746	32962	5216	30354.0
28117	33091	4974	30604.0
28858	33197	4339	31027.5
28877	33198	4321	31037.5
29080	33245	4165	31162.5
29586	33380	3794	31483.0
29965	33406	3441	31685.5
29994	33788	3794	31891.0
30263	33888	3625	32075.5
30313	33937	3624	32125.0
30366	34007	3641	32186.5
30629	34013	3384	32321.0
30680	34076	3396	32378.0
30788	34425	3637	32606.5
30958	34455	3497	32706.5
31074	34576	3502	32825.0
31405	35237	3832	33321.0
31994	35843	3849	33918.5
32384	35852	3468	34118.0

a)



b)

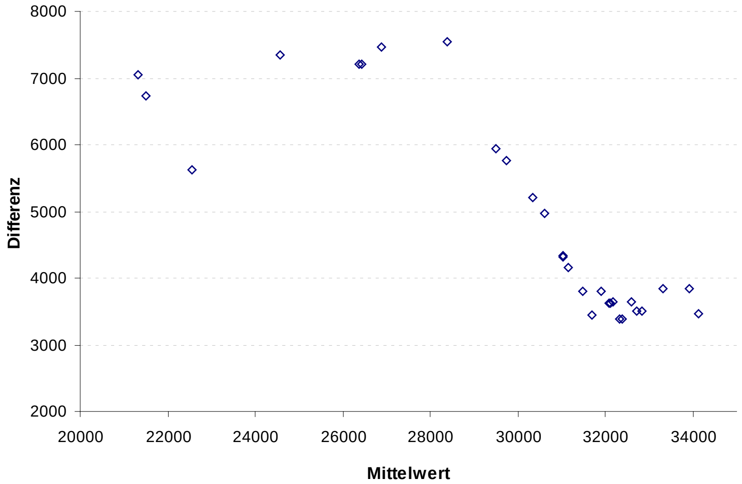


Abbildung C.3.1: ***a)*** *Q-Q-Plot der Verkehrsflussdaten und* ***b)*** *Mittelwert-Differenz-Plot.*

Abbildung C.3.1 zeigt das Verhältnis der beiden Datenreihen zueinander. Die Punkte liegen entfernt von der Mittelhalbierenden, verschoben in Richtung des Verkehrsflusses in Richtung 1. Die Verkehrsbelastung in Richtung 1 ist folglich grösser als in Richtung 2.

Aus dem Mittelwert-Differenz Plot ist ebenfalls zu sehen, dass eine Konzentration der Werte zum Verkehrsfluss Richtung 1 hin vorliegt. Für einen Grossteil der Daten sind die beobachteten Werte des Verkehrsflusses in Richtung 1 um etwa 3500 Autos pro Tag höher als in Richtung 2.

Aufgabe C.5- Lösung

Mittelwert aus den Zahlen der Studienanfänger : $\bar{x}=2101$

Mittelwert aus den Zahlen der Studentenanzahl (gesamt): $\bar{y}=13147$

Standardabweichung aus den Studienanfängerzahlen: $s_X=1337$

Standardabweichung aus der Studentenanzahl: $s_Y=8801$

Totale Anzahl der Beobachtungen: $n=6$.

Korrelationskoeffizient aus der Studienanfängerzahlen und der Studentenanzahl (gesamt):

$$r_{XY}=\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})(y_i-\bar{y})}{s_X s_Y}=\frac{11604968}{1337\cdot 8801}=0.99\,.$$

	x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(y-\bar{y})^2$	$(x_i-\bar{x})(y-\bar{y})$
A	3970	24273	1868	11126	3493161	123787876	20793574
B	732	5883	-1369	-7264	1874161	52765696	9944942
C	499	2847	-1602	-10300	2566404	106090000	16501516
D	1300	5358	-801	-7789	641601	60668521	6239887
E	3463	23442	1362	10295	1855044	105987025	14020755
F	2643	17076	542	3929	293764	15437041	2129134
Σ	12607	78879	-	-	10724135	464736159	69629807
Σ/n	2101	13147	-	-	1787356	77456026.5	11604968
$\sqrt{\Sigma/n}$	-	-	-	-	1337	8801	-

Aufgabe C.6- Lösung

Die Beziehung zwischen der Höhe der Messstation und der maximalen Temperatur, und der Höhe der Messstation und der minimalen Temperatur im Mai sind in Abbildung C.6.1 ersichtlich.

Sei x_i, y_i und z_i ($i = 1, 2, ..., 10$) die Höhe, die maximale Temperatur und die minimale Temperatur der *i-ten* Station sein. Wir erhalten:

<u>Mittelwerte:</u>

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1379, \qquad \bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = 13.7, \qquad \bar{z} = \frac{1}{10}\sum_{i=1}^{10} z_i = 4.36$$

<u>Standardabweichungen:</u>

$$s_X = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2} = 834, \; s_Y = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(y_i - \bar{y})^2} = 1.99, \; s_Z = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(z_i - \bar{z})^2} = 3.69$$

<u>Kovarianzen:</u>

$$s_{XY} = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})\cdot(y_i - \bar{y}) = -1513, \qquad s_{XZ} = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})\cdot(z_i - \bar{z}) = -2887$$

<u>Korrelationskoeffizienten:</u>

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y} = -0.91, \qquad r_{XZ} = \frac{s_{XZ}}{s_X \cdot s_Z} = -0.94$$

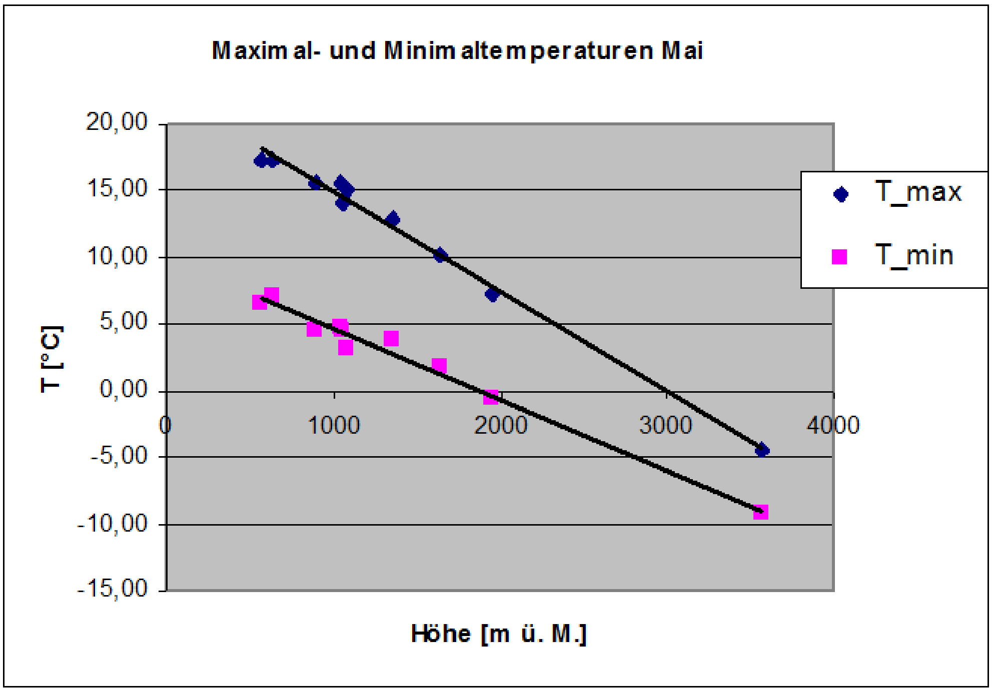


Abbildung C.6.1: *Beziehung zwischen Höhe der Messtation und maximalen/minimalen Temperaturen.*

x_i	y_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$y_i - \overline{y}$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
Höhe [m]	T_{max} [°C]					
1355	12.2	-24.1	580.81	-1.5	2.25	36.15
890	14.6	-489.1	239218.81	0.9	0.81	-440.19
1950	13.4	570.9	325926.81	-0.3	0.09	-171.27
1040	14	-339.1	114988.81	0.3	0.09	-101.73
1085	14.6	-294.1	86494.81	0.9	0.81	-264.69
1055	13.4	-324.1	105040.81	-0.3	0.09	97.23
574	16.4	-805.1	648186.01	2.7	7.29	-2173.77
3572	9.2	2192.9	4808810.4	-4.5	20.25	-9868.05
632	16.4	-747.1	558158.41	2.7	7.29	-2017.17
1638	12.8	258.9	67029.21	-0.9	0.81	-233.01

Tabelle C.6.2: Berechnungsblatt für Höhe – T_{max} Beziehung

x_i	z_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$z_i - \overline{z}$	$(z_i - \overline{z})^2$	$(x_i - \overline{x})(z_i - \overline{z})$
Höhe [m]	T_{min} [°C]					
1355	2.3	-24.1	580.81	-2.06	4.2436	49.646
890	6.3	-489.1	239218.81	1.94	3.7636	-948.854
1950	4.7	570.9	325926.81	0.34	0.1156	194.106
1040	4.3	-339.1	114988.81	-0.06	0.0036	20.346
1085	6.3	-294.1	86494.81	1.94	3.7636	-570.554
1055	5.1	-324.1	105040.81	0.74	0.5476	-239.834
574	8.3	-805.1	648186.01	3.94	15.5236	-3172.094
3572	-5.3	2192.9	4808810.4	-9.66	93.3156	-21183.414
632	8.1	-747.1	558158.41	3.74	13.9876	-2794.154
1638	3.5	258.9	67029.21	-0.86	0.7396	-222.654

Tabelle C.6.3: Berechnungsblatt für Höhe – T_{min} Beziehung

Aufgabe C.7 - Lösung

Die relative und kumulierte Häufigkeit entnimmt man aus Tabelle C.7.2. Abbildung C.7.1 zeigt das Histogramm.

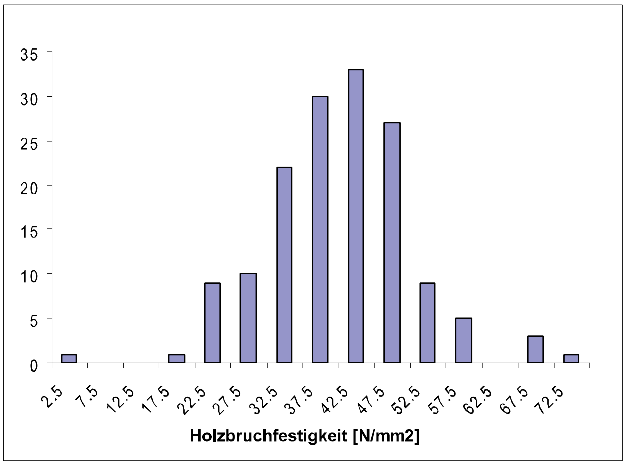


Abbildung C.7.1: Histogram.

a. die Wahrscheinlichkeit dass die gemessene Belastungswerte im Bereich 20 – 25 N/mm^2 liegen. $P[A] = \frac{n_k}{n} = \frac{9}{151} = 0.06$

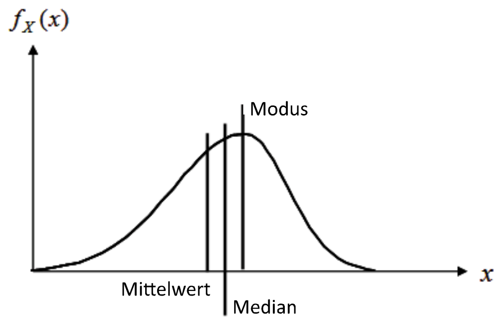
b. $$P[B] = \frac{\sum_{i=1}^{5} n_i}{n} = \frac{(1+0+0+1+9)}{151} = \frac{11}{151} = 0.073$$

Obere Klassengrenze [N/mm2]	Klassen-schwerpunkt [N/mm2]	abs. Häufigkeit	rel. Häufigkeit	kum rel. Häufigkeit
5	2.5	1	0.007	0.007
10	7.5	0	0.000	0.007
15	12.5	0	0.000	0.007
20	17.5	1	0.007	0.013
25	22.5	9	0.060	0.073
30	27.5	10	0.066	0.139
35	32.5	22	0.146	0.285
40	37.5	30	0.199	0.483
45	42.5	33	0.219	0.702
50	47.5	27	0.179	0.881
55	52.5	9	0.060	0.940
60	57.5	5	0.033	0.974
65	62.5	0	0.000	0.974
70	67.5	3	0.020	0.993
75	72.5	1	0.007	1.000

Table C.7.2: Klassifizierte Wertetabelle der Versuchsreihe zur Bruchfestigkeit von Holz.

Aufgabe C.8:

linksschiefe Verteilung



symmetrische Verteilung