Übung 2 - Lösung:

Aufgabe 2.1 – Lösung

Um die Graphik zu erstellen müssen die geordneten Daten verwendet werden. Basierend auf der von Benjamin & Cornell vorgeschlagenen Faustregel (Vorlesungsskript Faber, Gleichung 3.1) kann die Anzahl der Klassen auf 6 festgelegt werden. Tabelle 3.1.2 zeigt die Übersicht über die erhobenen Daten.

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

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

Die Klassen werden eingeteilt in:

$$\begin{aligned}
&24500-26500, \quad \text{Klassenschwerpunkt} = 25500\\
&26500-28500, \quad \text{Klassenschwerpunkt} = 27500\\
&28500-30500, \quad \text{Klassenschwerpunkt} = 29500\\
&30500-32500, \quad \text{Klassenschwerpunkt} = 31500\\
&32500-34500, \quad \text{Klassenschwerpunkt} = 33500\\
&34500-36500, \quad \text{Klassenschwerpunkt} = 35500
\end{aligned}$$

	Intervall (Anzahl der Autos $*10^3$)	Intervall Mittelpunkt (Anzahl der Autos $*10^3$)	Abs. Häufigkeit in der Klasse	Rel. Häufigkeit [%]	kumulative Häufigkeit
Richtung 1	24.5-26.5	25.5	3	10.000	0.100
	26.5-28.5	27.5	1	3.333	0.133
	28.5-30.5	29.5	3	10.000	0.233
	30.5-32.5	31.5	3	10.000	0.333
	32.5-34.5	33.5	16	53.333	0.867
	34.5-36.5	35.5	4	13.333	1.000
	Intervall (Anzahl der Autos $*10^3$)	**Intervall Mittelpunkt (Anzahl der Autos $*10^3$)**	**Abs. Häufigkeit in der Klasse**	**Rel. Häufigkeit [%]**	**kumulative Häufigkeit**
Richtung 2	17.5-20.0	18.75	3	10.000	0.100
	20.0-22.5	21.25	2	6.667	0.167
	22.5-25.0	23.75	4	13.333	0.300
	25.0-27.5	26.25	2	6.667	0.367
	27.5-30.0	28.75	8	26.667	0.633
	30.0-32.5	31.25	11	36.667	1.000

Tabelle 2.1.2 Übersicht über die beobachteten Verkehrsflüsse

Abbildung 2.1.1 and Abbildung 2.1.2 zeigen die relative Häufigkeitsverteilung und die relative kumulierte Häufigkeitsverteilung der Verkehrsflüsse. Obwohl man die Werte der Summenhäufigkeit von Tabelle 2.1.2 verwenden könnte, um die Summenhäufigkeit zu bilden, werden die Quantile der

Beobachtungen (Tabelle 2.1.3) stattessen verwendet. Das ist, wie im Skript erwähnt (Abschnitt C.3), da die Beobachtungen bekannt sind. Die Summenhäufigkeiten in Tabelle 2.1.2 würden verwendet werden wenn nur die Intervalle der Beobachtungen bekannt wären. Versuche jedoch, die Summenhäufigkeiten zur Übung selbst aufzutragen unter Verwendung des Intervalls und der Summenhäufigkeiten von Tabelle 2.1.2.

Nummer (*i*)	**Richtung 1**		**Richtung 2**	
	geordnet	**Quantile =** $\frac{i}{n+1}$	**geordnet**	**Quantile =** $\frac{i}{n+1}$
1	24846	0.0323	17805	0.0323
2	24862	0.0645	18123	0.0645
3	25365	0.0968	19735	0.0968
4	28252	0.1290	20903	0.1290
5	29224	0.1613	21145	0.1613
6	29976	0.1935	22762	0.1935
7	30035	0.2258	22828	0.2258
8	30613	0.2581	23141	0.2581
9	32158	0.2903	24609	0.2903
10	32472	0.3226	26525	0.3226
11	32618	0.3548	26846	0.3548
12	32962	0.3871	27746	0.3871
13	33091	0.4194	28117	0.4194
14	33197	0.4516	28858	0.4516
15	33198	0.4839	28877	0.4839
16	33245	0.5161	29080	0.5161
17	33380	0.5484	29586	0.5484
18	33406	0.5806	29965	0.5806
19	33788	0.6129	29994	0.6129
20	33888	0.6452	30263	0.6452
21	33937	0.6774	30313	0.6774
22	34007	0.7097	30366	0.7097
23	34013	0.7419	30629	0.7419
24	34076	0.7742	30680	0.7742
25	34425	0.8065	30788	0.8065
26	34455	0.8387	30958	0.8387
27	34576	0.8710	31074	0.8710
28	35237	0.9032	31405	0.9032
29	35843	0.9355	31994	0.9355
30	35852	0.9677	32384	0.9677

Tabelle 2.1.3 Quantile der Verkehrsflussdaten

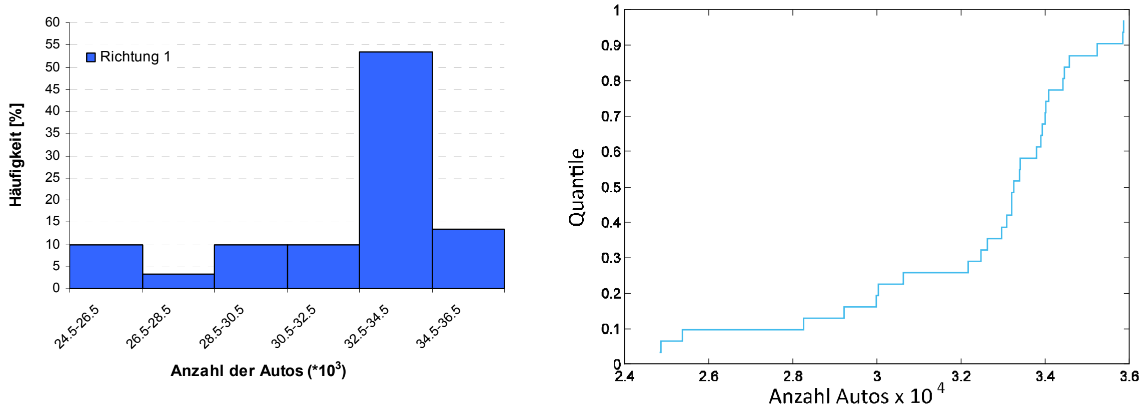


Abbildung 2.1.1 Häufigkeitsverteilung und kumulierte Häufigkeitsverteilung

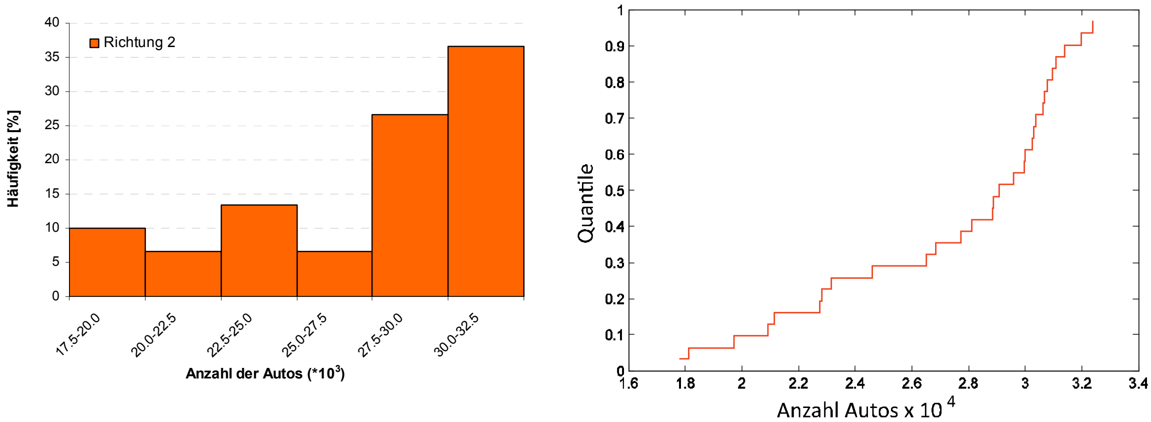


Abbildung 2.1.2 Häufigkeitsverteilung und kumulierte Häufigkeitsverteilung.

Aus dieser graphischen Präsentation der Daten ist ersichtlich, dass der Verkehrsfluss in Richtung 2 geringer als derjenige in der Gegenrichtung. Richtung 1 zeigt die grösste Häufigkeit in der Klasse mit den Grenzen 32500 und 34500 Autos pro Tag, wohingegen das grösste Aufkommen in Richtung 2 im Intervall 30000 – 32500 Autos pro Tag liegt. Zusätzlich kann man sehen, dass beide Verteilungen linksschief sind.

Trägt man die Häufigkeitsdiagramme im gleichen Massstab auf (Abbildung 2.1.3), so erhält man einen direkten visuellen Vergleich der beiden Datenreihen. Es ist leicht zu erkennen, dass die kumulative Häufigkeitsverteilung der Datenreihe Richtung 1 signifikant nach rechts verschoben ist. Dies zeigt wiederum den höheren Verkehrsfluss in Richtung 1.

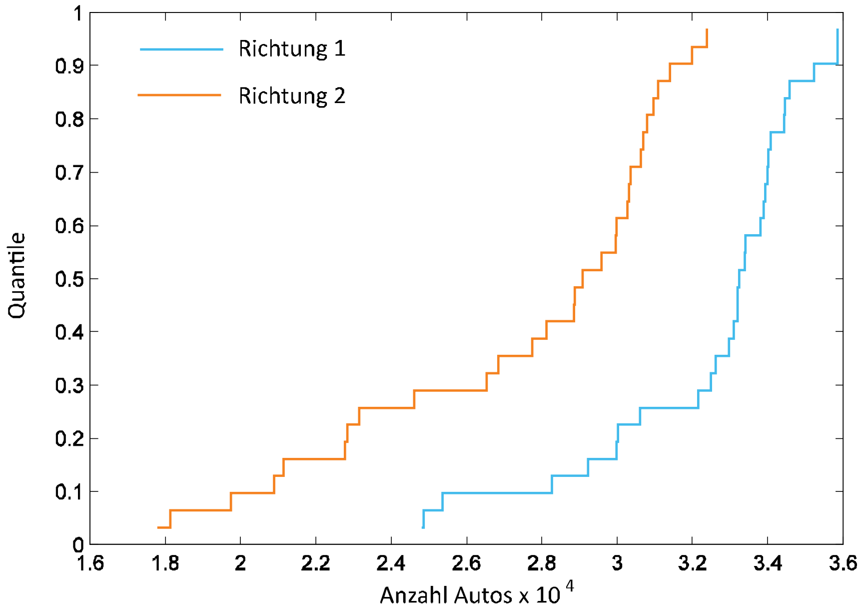
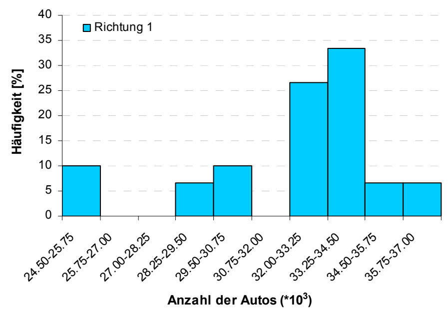


Abbildung 2.1.3 **Kumulative Häufigkeitsverteilung des Verkehrsflusses in der Rosengartenstrasse (Richtung 1 und Richtung 2).**

Mit der gewählten Klassenzahl (6) gehen aus den dargestellten Diagrammen viele Informationen verloren. Folgende Diagramme stellen die Lösungen für eine grössere Klassenzahl (10) dar.

	Intervall (Anzahl Autos *10^3)	Intervallmitte (Anzahl Autos *10^3)	Abs. Häufigkeit	Rel. Häufigkeit [%]	kumulative Häufigkeit
Richtung 1	24.50-25.75	25.125	3	10.000	0.100
	25.75-27.00	26.375	0	0.000	0.100
	27.00-28.25	27.625	0	0.000	0.100
	28.25-29.50	28.875	2	6.667	0.167
	29.50-30.75	30.125	3	10.000	0.267
	30.75-32.00	31.375	0	0.000	0.267
	32.00-33.25	32.625	8	26.667	0.533
	33.25-34.50	33.875	10	33.333	0.867
	34.50-35.75	35.125	2	6.667	0.933
	35.75-37.00	36.25	2	6.667	1.000
	Intervall (Anzahl Autos *10^3)	**Intervallmitte (Anzahl Autos *10^3)**	**Abs. Häufigkeit**	**Rel. Häufigkeit [%]**	**kumulative Häufigkeit**
Richtung 2	17.5-19.0	18.25	2	6.667	0.067
	19.0-20.5	19.75	1	3.333	0.100
	20.5-22.0	21.25	2	6.667	0.167
	22.0-23.5	22.75	3	10.000	0.267
	23.5-25.0	24.25	1	3.333	0.300
	25.0-26.5	25.75	0	0.000	0.300
	26.5-28.0	27.25	3	10.000	0.400
	28.0-29.5	28.75	4	13.333	0.533
	29.5-31.0	30.25	10	33.333	0.867
	31.0-32.5	31.75	4	13.333	1.000

Tabelle 2.1.3 **Übersicht über den beobachteten Verkehrsfluss**



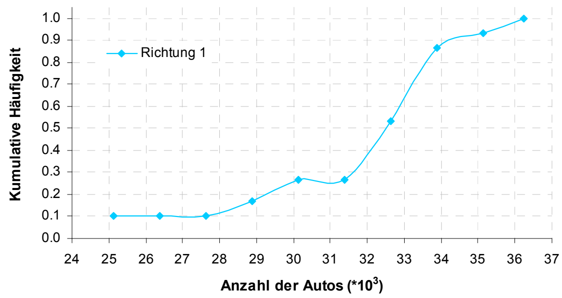


Abbildung 2.1.4 Häufigkeitsverteilung und kumulierte Häufigkeitsverteilung des Verkehrsflusses in der Rosengartenstrasse (Richtung 1).

Aus Abbildung 2.1.4 kann man sehen, dass eine grössere Klassenzahl eine bessere Visualisierung der Verteilungen ermöglicht.

Aufgabe 2.2 - Lösung

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

- Unteres Quantil
- Unterer Nachbarschaftswert
- Median
- Oberer Nachbarschaftswert
- Oberes Quantil

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

$$\nu = nQ_\nu + Q_\nu$$

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

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

ν hat keinen ganzzahligen Wert. Der Wert teilt sich in einen ganzzahligen-Teil $k = 7$ und einen fraktionalen Teil $p = 0.75$. Daher ist x_ν^o :

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

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

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

Mit Hilfe von Tabelle 2.1.1 bekommen wir:

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

Der Median ergibt sich aus folgender Beziehung:

$$v = 30*0.5 + 0.5 = 15.5$$

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

Um die Nachbarschaftswerte zu berechnen wird die interquartile Differenz benötigt:

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

Der untere Nachbarschaftswert ist die kleinste Beobachtung, die grösser oder gleich des unteren Quantils minus 1.5r ist:

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

Von Tabelle 3.1.1 kann somit der untere Nachbarschaftswert mit 25365 abgelesen werden.

Auf gleiche Weise wird der obere Nachbarschaftswert berechnet:

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

Von Table 2.1.1 kann somit der obere Nachbarschaftswert abgelesen werden. Dies ist ein Wert der kleiner oder gleich zu 39369 ist, das ist 35852 welcher in diesem Fall mit dem höchsten Wert der Datenreihe übereinstimmt.
Tabelle 2.2.1 zeigt eine Übersicht für beide Datenreihen. Die Daten für den Verkehrsfluss in Richtung 2 weisen hierbei keine Ausreisser auf.

Statistik	**Richtung 1**	**Richtung 2**
unteres Quartil	30613	23141.0
unterer Nachbarschaftswert	28252	17805.0
Median	33221.5	28978.5
oberer Nachbarschaftswert	35852	32384.0
oberes Quartil	34013	30629.0
Ausreisser	24846 24862 25365	

Tabelle 2.2.1 Statistische Werte für den Tukey-Box-Plot

Abbildung 3.2.1 zeigt Tukey-Box-Plots für beide Richtungen. Es ist ersichtlich, dass alle Masszahlen der zentralen Tendenz für die Datenreihe 1 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.

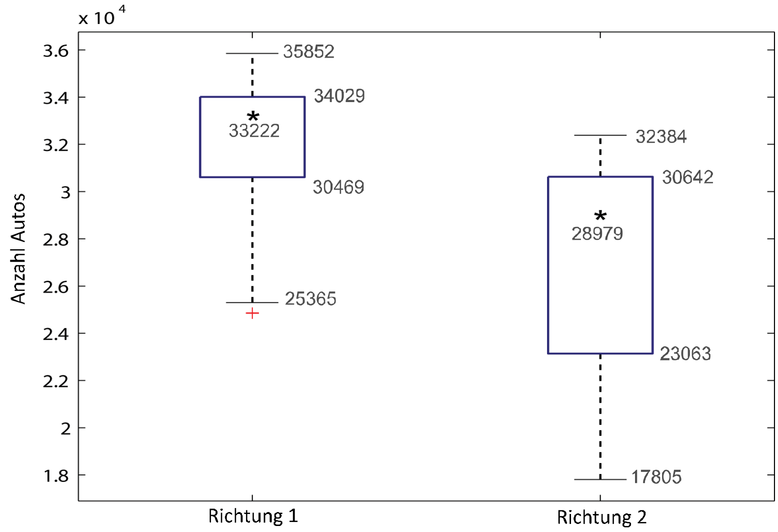


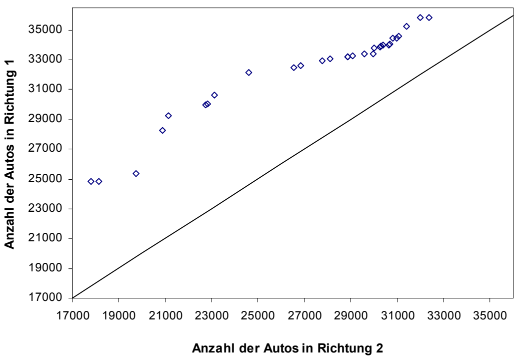
Abbildung 2.2.1: Tukey-Box-Plot der Verkehrsflüsse in der Rosengartenstrasse.

Aufgabe 2.3 - 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 2.3.1). Für den Tukey mean-difference-Plot ist zunächst die Differenz und der Mittelwert der Daten in beiden Verkehrsrichtungen zu erstellen (Tabelle 2.3.1).

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

Tabelle 2.3.1: Daten für den Tukey mean-difference-Plot.



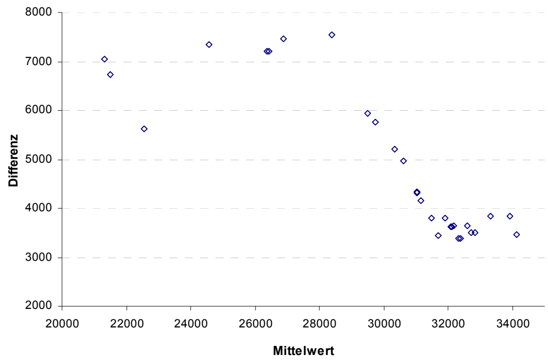


Abbildung 2.3.1 Q-Q-Plot der Verkehrsflussdaten und Tukey Mean-Difference-Plot.

Abbildung 2.3.1 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 Tukey-Mean-Difference 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 2.5- Lösung

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

Mittelwert aus den Zahlen der Studentenanzahl (gesamt): $\overline{y}$ =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):
$$r_{XY} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\overline{x})(y_i-\overline{y})}{s_X s_Y} = \frac{11604968}{1337 \cdot 8801} = 0.99\,.$$

	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 2.6- 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 2.6.1 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>

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

<u>Kovarianzen:</u>

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

<u>Korrelationskoeffizienten:</u>

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

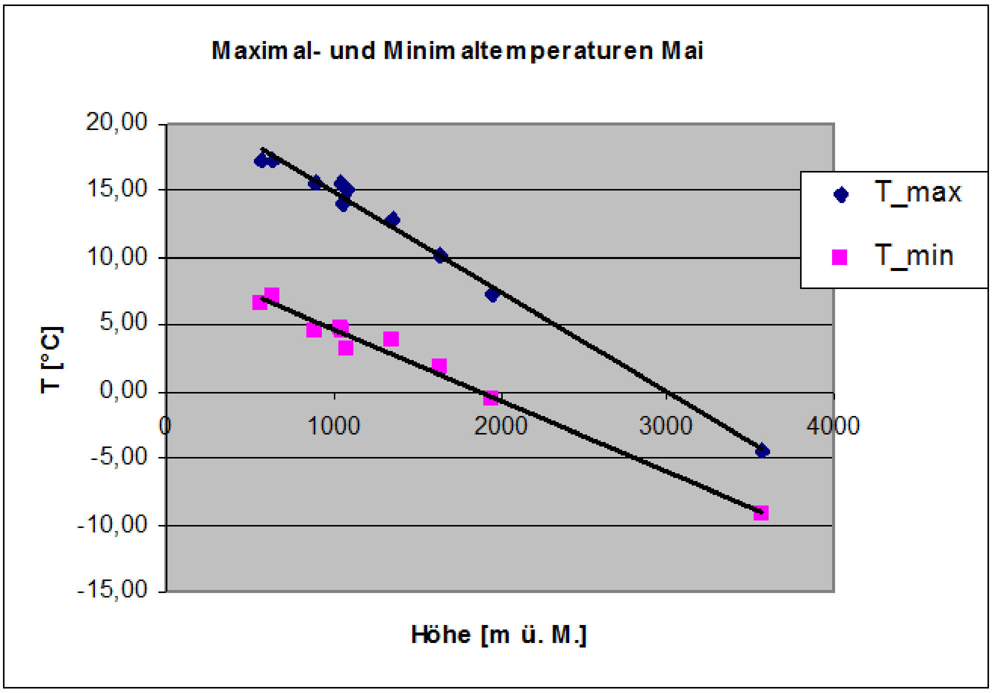


Abbildung 2.6.1: Beziehung zwischen Höhe der Messtation und maximalen/minimalen Temperaturen

x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
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 2.6.2: 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 3.6.3: Berechnungsblatt für Höhe – T_{min} Beziehung

Aufgabe 2.7 - Lösung

Die relative und kumulierte Häufigkeit entnimmt man aus Tabelle 3.7.2. Abbildung 3.7.1 zeigt das Histogramm.

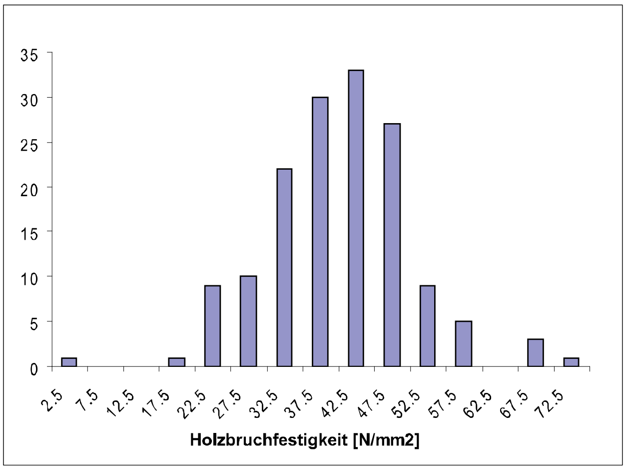


Abbildung 2.7.1: Histogram.

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

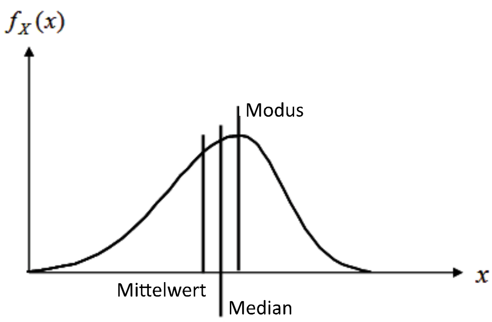
b. $$P[B]=\frac{\sum_{i=1}^{5} n_i}{n}=\frac{(1+0+0+1+9)}{151}=\frac{11}{151}=0.073$$

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 2.7.2: Klassifizierte Wertetabelle der Versuchsreihe zur Bruchfestigkeit von Holz.

Aufgabe 2.8:

<u>linksschiefe</u> Verteilung



<u>symmetrische</u> Verteilung