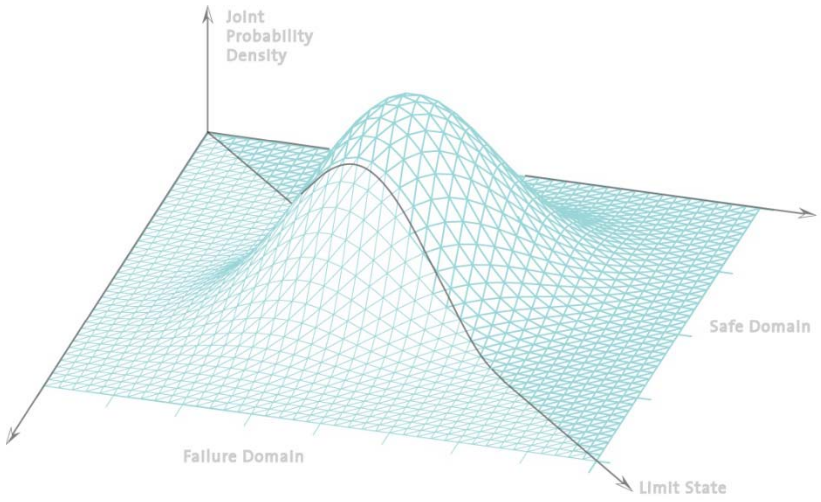


Statistik und Wahrscheinlichkeitsrechnung

Übung 2

Inhalt der heutigen Übung

- Beschreibende Statistik
- Gemeinsames Lösen der Übungsaufgaben
  - 2.1: Häufigkeitsverteilung
  - 2.2: Tukey-Boxplot
  - 2.5: Korrelation
- Vorstellen der Gruppenaufgabe 2.4
- Berechnungsbeispiele mit EXCEL

Numerische Zusammenfassungen

Mittelwerte:

Arithmetisches Mittel:	Schwerpunkt der Stichprobe
Median:	Mittlerer Wert einer Stichprobe
Modalwert:	Am häufigsten vorkommender Wert

Streuungsmasse:

Varianz / Standardabweichung:	Verteilung um den Mittelwert
Variationskoeffizient :	Variabilität relativ zum Mittelwert

Andere Masse:

- Schiefekoeffizient:	Schiefe relativ zum Mittelwert
- Kurtosis:	Wölbung um den Mittelwert

Masse für Korrelation:

- Kovarianz:	Tendenz für paarweise beobachtete Eigenschaften
- Korrelationskoeffizient :	Normalisierter Koeffizient zwischen -1 und +1

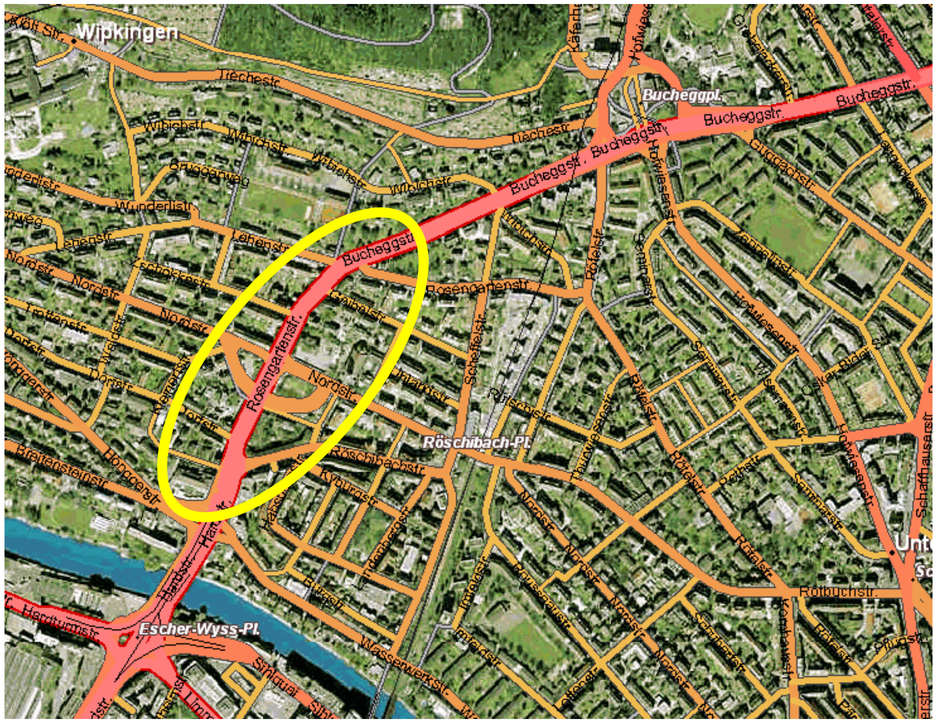
Zusammenfassung Graphische Darstellung

Ein-dimensionales Streudiagramm	Veranschaulicht den Bereich und die Verteilung von Datenreihen entlang einer Achse, und zeigt Symmetrie.
Zwei-dimensionales Streudiagramm	Veranschaulicht den paarweisen Zusammenhang von Daten.
Histogramm	Stellt die Verteilung von Daten über einem Bereich von Datenreihen dar, zeigt Modalwert und Symmetrie.
Quantile Plot	Stellt Median, Verteilung und Symmetrie dar.
Tukey – Boxplot	Stellt Median, obere/untere Quartile, Symmetrie und Verteilung dar.
Q-Q Plot	Vergleicht zwei Datenreihen, relatives Bild.
Mittel-über-Differenz Plot	Vergleicht zwei Datenreihen, relatives Bild.

Beschreibende Statistik

Uns wurden die unten dargestellten Verkehrsdaten der Rosengartenstrasse in Zürich aus dem Monat April 2001 zur Auswertung übergeben. Richtung 1 gibt die Verkehrsbelastung zum Bucheggplatz, Richtung 2 die Belastung zum Escher-Wyss-Platz an.

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

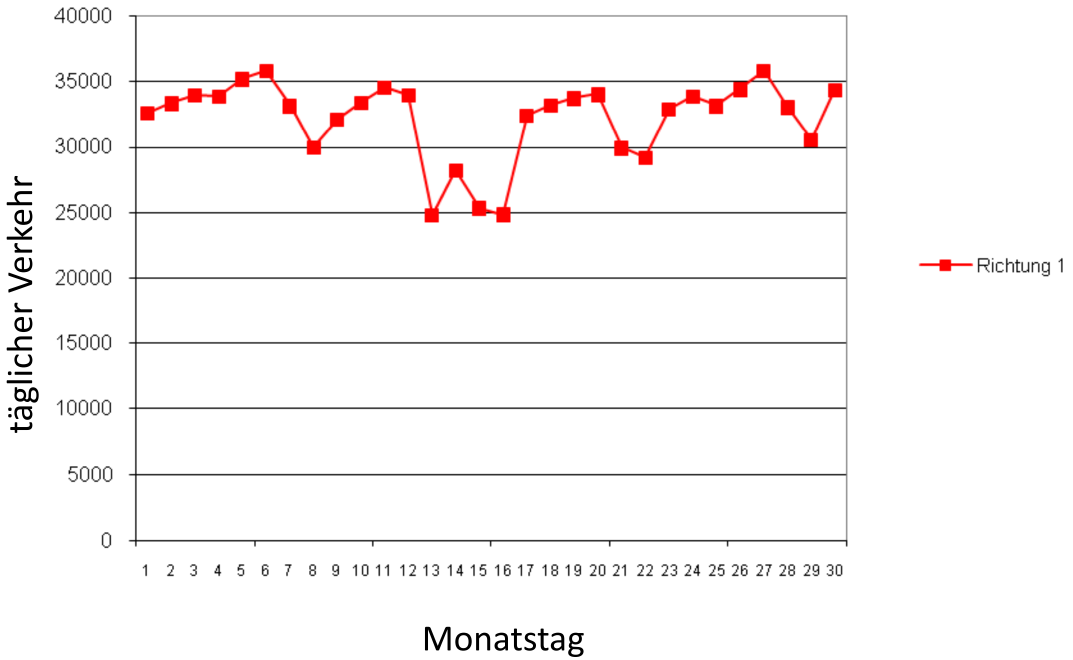


Beschreibende Statistik

Was wollen wir wissen?

Wie können wir das erfahren? - Grafik, Histogramm
- numerische Zusammenfassung etc.

Beispiel:
Man will etwas über die Änderung des Verkehrs in Richtung 1 im Monat April erfahren.



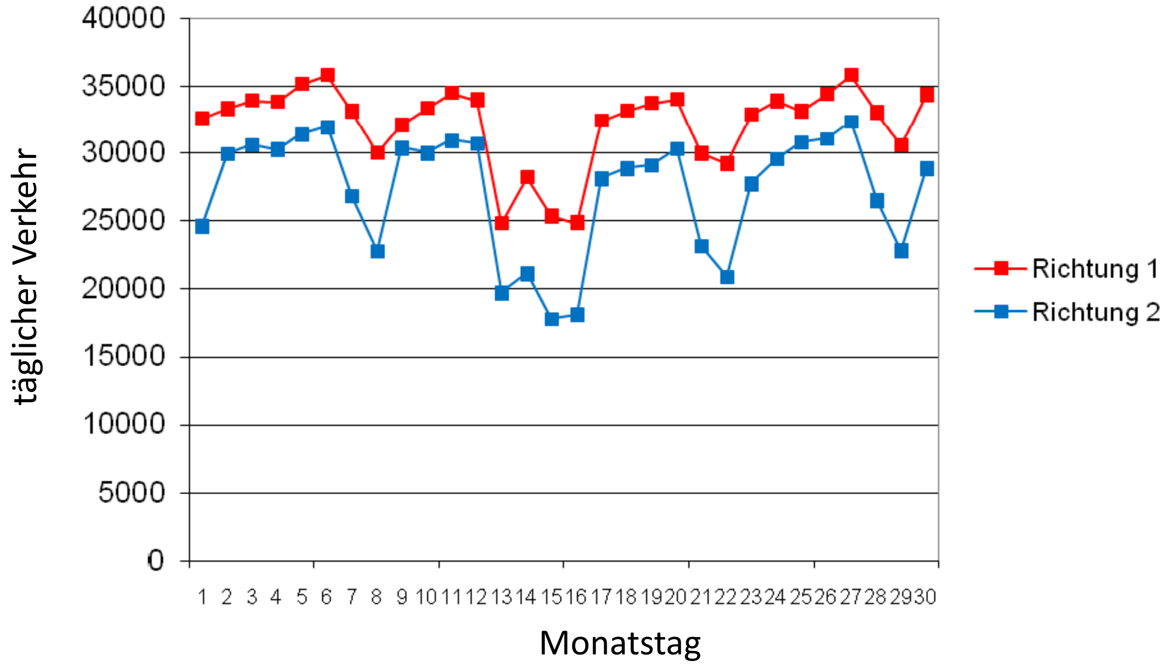
Beschreibende Statistik

Was wollen wir wissen?

Wie können wir das erfahren?
- Grafik, Histogramm
- numerische Zusammenfassung etc.

Beispiel:

Man will zusätzlich etwas über das Verhältnis der beiden Verkehrsströme erfahren.



Beschreibende Statistik

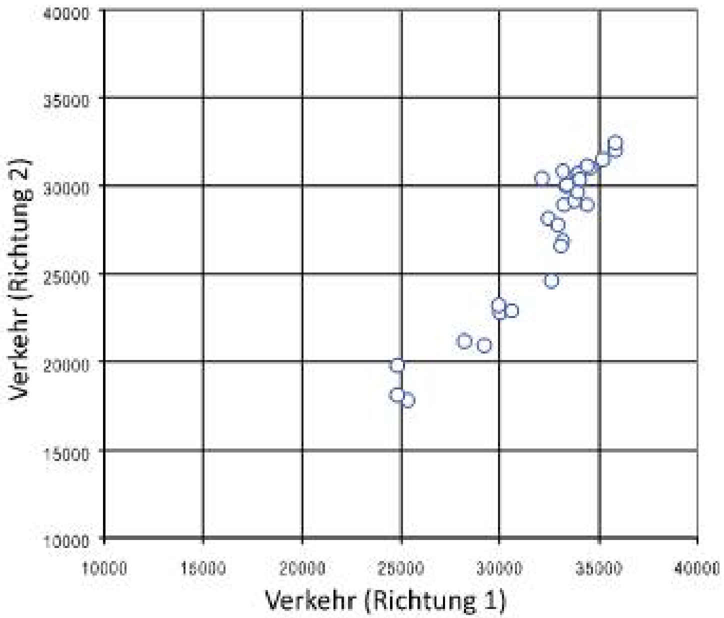
Was wollen wir wissen?

Wie können wir das erfahren?

- Grafik, Histogramm
- numerische Zusammenfassung etc.

Beispiel:

Man will zusätzlich über das Verhältnis der beiden Verkehrsströme erfahren, man benötigt aber keine Informationen der Zeit.



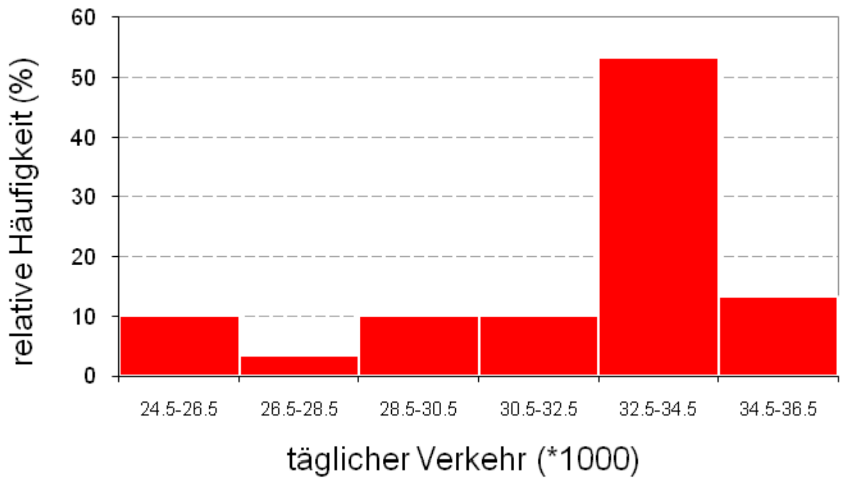
Beschreibende Statistik

Was wollen wir wissen?

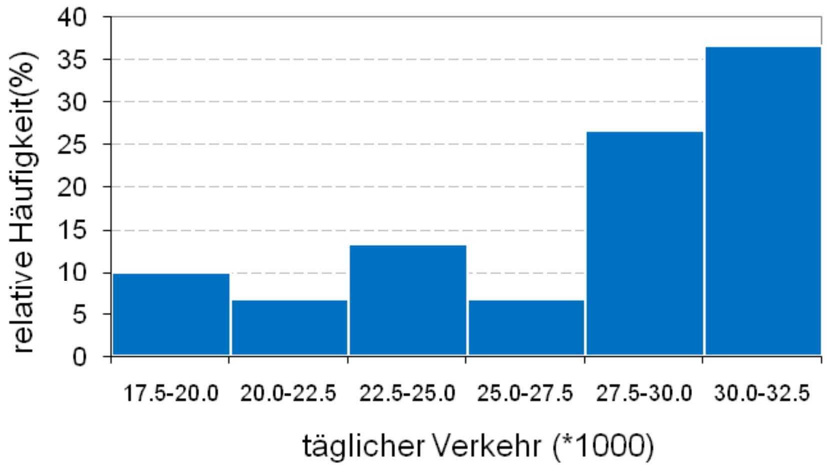
Wie können wir das erfahren?
- Grafik, Histogramm
- numerische Zusammenfassung etc.

Beispiel:

Man interessiert sich für die täglichen Verkehrsströme, in jede Richtung



Richtung 1



Richtung 2

…was ist bei diesem Vergleich nicht gut gelöst?

Beschreibende Statistik

Wir üben heute…

wie man die Eigenschaften von gegebenen Daten darstellen kann, und zwar:

- Grafisch
    - Häufigkeitsverteilung (Histogramm)
    - kumulative Häufigkeitsverteilung
- Numerisch
    - Mittelwert
    - Standardabweichung
- Zusammenfassungen
    - Tukey-Box-Plot
- Korrelation von Datenreihen

Anmerkung: Du kannst Excel, Matlab und/oder andere Statistikprogramme verwenden.

ABER !!!!!

Stelle IMMER sicher, Funktionen selbst einzusetzen oder zu prüfen, ob die Funktionen, die du vom verwendeten Programm bereitgestellt bekommst, mit denen des Skripts übereinstimmen!

Aufgabe 2.1

Erstelle von den erhobenen Verkehrsdaten nach deren Einteilung in Klassen eine Häufigkeitsverteilung sowie eine kumulierte Häufigkeitsverteilung und stelle deren Verläufe in den geeigneten Graphen dar.

Vergleiche die Verkehrsflüsse beider Richtungen.

Aufgabe 2.1

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

Schritte:

1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

Aufgabe 2.1

Schritt 1 (Sortieren der Daten)

Schritte:

1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

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

sortieren →

Richtung 1	Richtung 2
24846	17805
24862	18123
25365	19735
28252	20903
29224	21145
29976	22762
30035	22828
30613	23141
32158	24609
32472	26525
32618	26846
32962	27746
33091	28117
33197	28858
33198	28877
33245	29080
33380	29586
33406	29965
33788	29994
33888	30263
33937	30313
34007	30366
34013	30629
34076	30680
34425	30788
34455	30958
34576	31074
35237	31405
35843	31994
35852	32384

Aufgabe 2.1

Schritte:
1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

Schritt 2 (Wahl der Anzahl der Klassen)

Es gibt keine allgemeingültige Regel, aber eine Faustregel:

$$k = 1 + 3.3 \log_{10} n$$

dabei ist *k* die Anzahl an Klassen und *n* die Anzahl an Daten.

In unserem Fall, n = 30

$$k = 1 + 3.3 \log_{10} 30 = 5.87 \approx 6$$

Richtung 1,

Minimum = 24846

Maximum= 35852

Wir könnten folgende Klassem wählen:

(24.5,26.5] ,(26.5,28.5], (28.5,30.5],(30.5,32.5],(32.5,34.5],(34.5,36.5]

Aufgabe 2.1

Schritte:

1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

Schritt 3 (Zähle die Anzahl in jeder Klasse)

Richtung 1

24846
24862
25365
28252
29224
29976
30035
30613
32158
32472
32618
32962
33091
33197
33198
33245
33380
33406
33788
33888
33937
34007
34013
34076
34425
34455
34576
35237
35843
35852

Zählen →

	Intervall (Anzahl der Autos $*10^3$)	Intervall Mittelpunkt (Anzahl der Autos $*10^3$)	Abs. Häufigkeit in der Klasse
Richtung 1	24.5-26.5	25.5	3
	26.5-28.5	27.5	1
	28.5-30.5	29.5	3
	30.5-32.5	31.5	3
	32.5-34.5	33.5	16
	34.5-36.5	35.5	4

Aufgabe 2.1

Schritte:
1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

Schritt 4 (Zeichne die Häufigkeitsverteilung)

Zuerst einige Berechnungen ....

	Intervall (Anzahl der Autos *10³)	Intervall Mittelpunkt (Anzahl der Autos *10³)	Abs. Häufigkeit in der Klasse	Rel. Häufigkeit [%]
Richtung 1	24.5-26.5	25.5	3	10.000
	26.5-28.5	27.5	1	3.333
	28.5-30.5	29.5	3	10.000
	30.5-32.5	31.5	3	10.000
	32.5-34.5	33.5	16	53.333
	34.5-36.5	35.5	4	13.333

$$rel.Häufigkeit\% = \frac{n_k}{n_{ges}}100$$

$$= \frac{3}{30}100 = 10$$

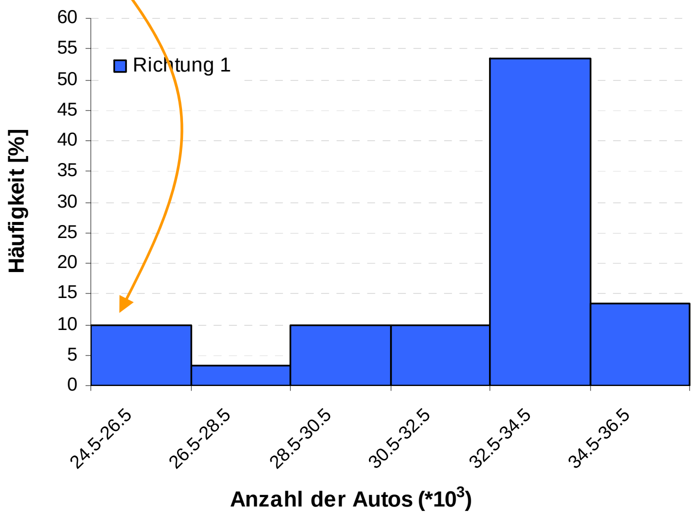
Aufgabe 2.1

Schritt 4 (Zeichne die Häufigkeitsverteilung)

Schritte:
1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

	Intervall (Anzahl der Autos $*10^3$)	Intervall Mittelpunkt (Anzahl der Autos $*10^3$)	Abs. Häufigkeit in der Klasse	Rel. Häufigkeit [%]
Richtung 1	24.5-26.5	25.5	3	10.000
	26.5-28.5	27.5	1	3.333
	28.5-30.5	29.5	3	10.000
	30.5-32.5	31.5	3	10.000
	32.5-34.5	33.5	16	53.333
	34.5-36.5	35.5	4	13.333

Zeichnen



Aufgabe 2.1

Schritt 5 (Zeichne die kumulative Häufigkeits-verteilung)

Schritte:

1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

kumulieren

/100

	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

Aufgabe 2.1

Schritt 5 (Zeichne die kumulative Häufigkeits-verteilung)

Schritte:
1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

kumulieren

	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	.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

+ /100

Aufgabe 2.1

Schritt 5 (Zeichne die kumulative Häufigkeits-verteilung)

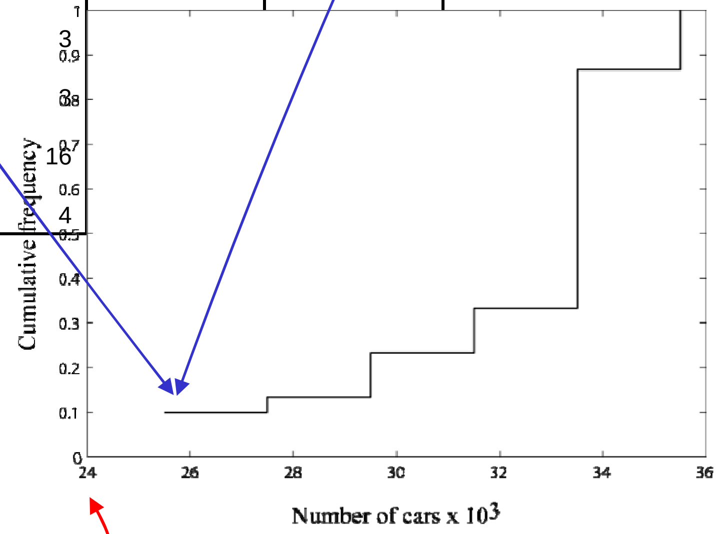
Schritte:
1. Daten sortieren
2. Geeignete Anzahl von Klassen wählen
3. Die Anzahl der Daten für jede Klasse zählen
4. Häufigkeitsverteilung zeichnen
5. Kumulative Häufigkeitsverteilung zeichnen

kumulieren

	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		
	30.5-32.5	31.5	3		
	32.5-34.5	33.5	16		
	34.5-36.5	35.5	4		

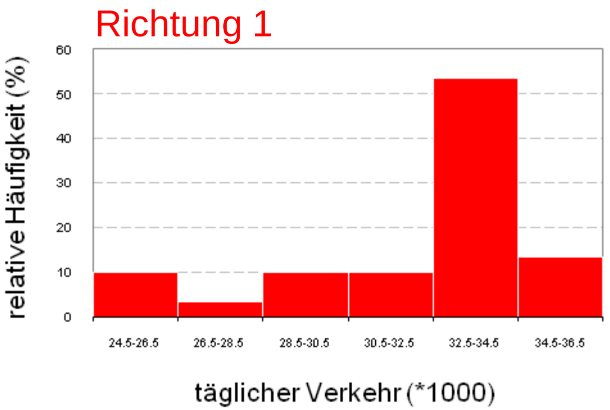
zeichnen

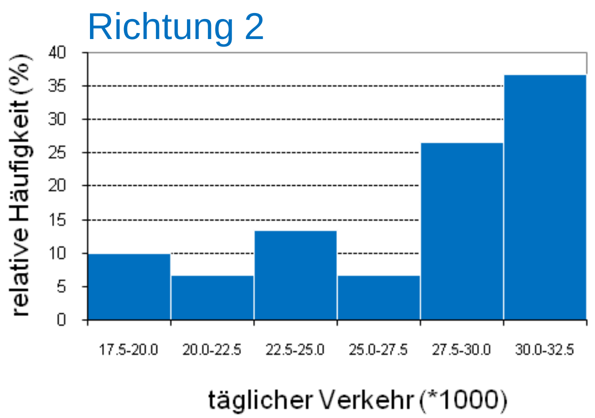
y=0

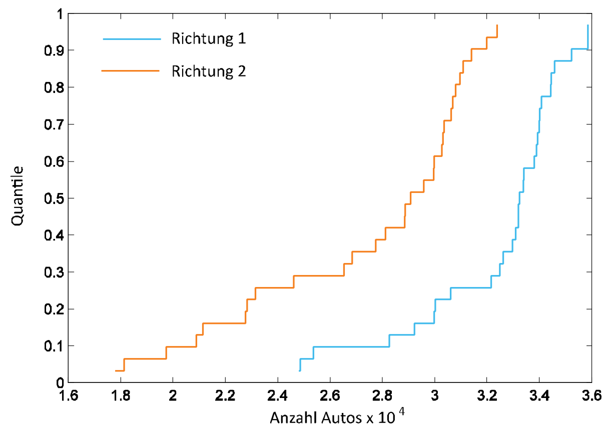


Aufgabe 2.1

Dasselbe kann nun für Richtung 2 durchgeführt werden.
Was kann aus diesen Diagrammen erkannt werden?







- Der Verkehrsstrom in Richtung 1 ist grösser als in Richtung 2.
- Es sind grosse Variationen in beiden Richtungen zu erkennen.
- etc.…

Diese Diagramme geben einen guten ersten Überblick über die Daten!

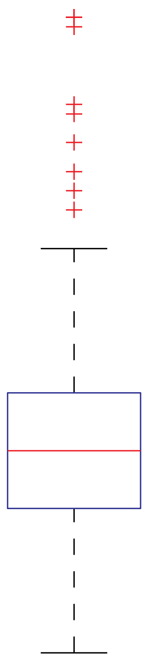
Quantile

- Definition :
  - Die Q-Quantile korrespondiert mit dem Wert der Stichprobe, welcher mit dem Wert 100% - Q x 100% überschritten wird.
  - D.h. zum Beispiel: das 0.75-Quantil wird von 100% - 0.75 x 100% = 25% der Daten überschritten.
  - Die Quantile werden von der geordneten Stichprobe berechnet: $x_1^o \leq x_2^o \leq \ldots \leq x_n^o$

$$Q_i = \frac{i}{n+1}, \; n: \text{ Gesamt Anzahl der Beobachtungen, } i=1,2\ldots,n$$

Aufgabe 2.2

Verwende für beide Datenreihen der Verkehrsdaten den Tukey-Boxplot, um eine zusammenfassende Übersicht über die Eigenschaften der jeweiligen Verteilung zu bekommen. Trage beide Darstellungen in die gleiche Graphik auf, um die Datenreihen der Verkehrsmessung anschaulich vergleichen zu können und beurteile diese hinsichtlich ihrer Symmetrie.



Schritte

1. Berechne den Median
2. Berechne die 75%- and 25%- Quantile.
3. Berechne die verbundenen Werte.
4. Ausreisser
5. Zeichne den Tukey-Box-Plot

Aufgabe 2.2

Schritte
1. Berechne den Median
2. Berechne die 75%- und 25%- Quantile.
3. Berechne die verbundenen Werte.
4. Ausreisser
5. Zeichne den Tukey-Box-Plot

Schritt 1 (Berechne den Median)

Es ist der Zentralwert (50%-Quantil).

Aber …. wenn die Anzahl der Daten gerade ist, ist das nicht möglich!

In diesem Fall müssen wir linear interpolieren.

Richtung 1
24846
24862
25365
28252
29224
29976
30035
30613
32158
32472
32618
32962
33091
33197
33198
33245
33380
33406
33788
33888
33937
34007
34013
34076
34425
34455
34576
35237
35843
35852

Der Median ist $$\frac{33198 + 33245}{2} = 33221.5$$

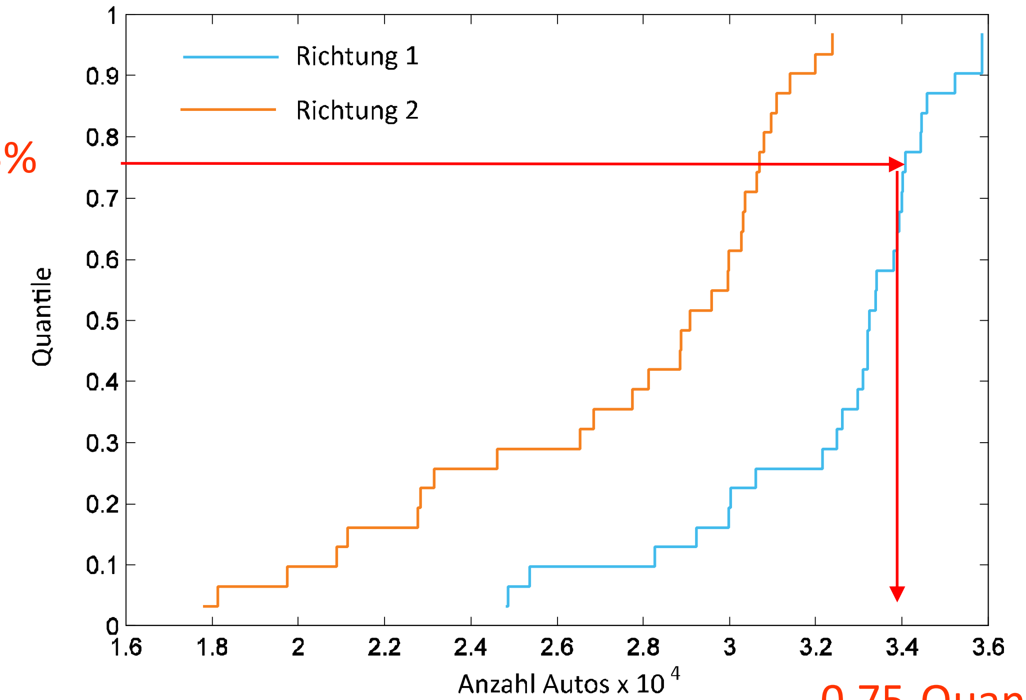
Aufgabe 2.2

Schritte
1. Berechne den Median
2. Berechne die 75%- und 25%- Quantile.
3. Berechne die verbundenen Werte.
4. Ausreisser
5. Zeichne den Tukey-Box-Plot

Schritt 2 (Berechnung der Quartile)

Anschaulich bedeutet das:

0.75 = 75%



0.75-Quantil

Aufgabe 2.2

Schritt 2 (Berechnung der Quartile)

Genaue Berechnung:

$$Q_i = \frac{i}{n+1},$$

n : Gesamt Anzahl der Beobachtungen

Richtung 1	i	i/31
24846	1	0.03
24862	2	0.06
25365	3	0.10
28252	4	0.13
29224	5	0.16
29976	6	0.19
30035	7	0.23
30613	8	0.26
32158	9	0.29
32472	10	0.32
32618	11	0.35
32962	12	0.39
33091	13	0.42
33197	14	0.45
33198	15	0.48
33245	16	0.52
33380	17	0.55
33406	18	0.58
33788	19	0.61
33888	20	0.65
33937	21	0.68
34007	22	0.71
34013	23	0.74 ← 75%
34076	24	0.77
34425	25	0.81
34455	26	0.84
34576	27	0.87
35237	28	0.90
35843	29	0.94
35852	30	0.97

Schritte
1. Berechne den Median
2. Berechne die 75%- und 25%- Quantile.
3. Berechne die Nachbarschaftswerte.
4. Ausreisser
5. Zeichne den Tukey-Box-Plot

Aufgabe 2.2

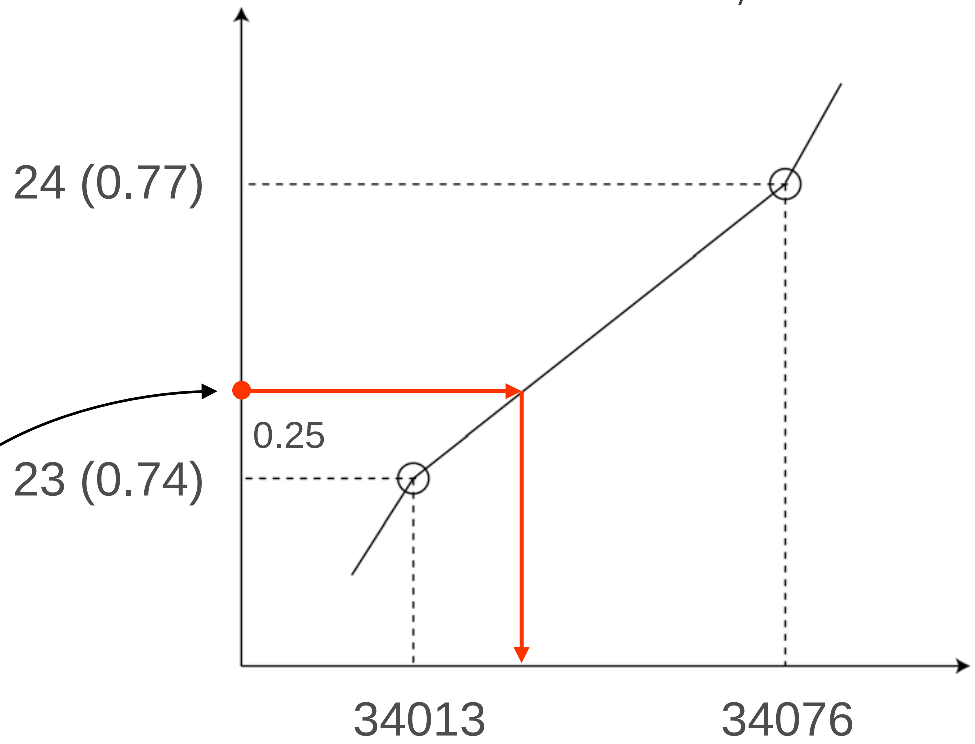
Schritt 2 (Berechnung der Quartile)

Interpolation

33888	20	0.65
33937	21	0.68
34007	22	0.71
34013	23	0.74
34076	24	0.77
34425	25	0.81

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

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



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

Schritte
1. Berechne den Median
2. Berechne die 75%- und 25%- Quantile.
3. Berechne die verbundenen Werte.
4. Ausreisser
5. Zeichne den Tukey-Box-Plot

Aufgabe 2.2

Schritt 3 (Berechnung der verbundenen Werte)

interquartiler Bereich

$$\left.\begin{aligned} Q_{0.75} &= 34029 \\ Q_{0.25} &= 30469 \end{aligned}\right\} r \equiv Q_{0.75} - Q_{0.25} = 34029 - 30469 = 3560$$

Grösster verbundener Wert: $\text{grösster Wert} \leq (75\%\ \text{Quantil}) + 1.5 \cdot r$

In diesem Fall, grösster Wert kleiner als $34029 + 1.5 \times 3560 = 39363$

33198
33245
33380
33406
33788
33888
33937
34007
34013
34076
34425
34455
34576
35237
35843
35852

Der grösste Wert der Datenreihe kleiner als der berechnete, ist der *grösste verbundene Wert*.

Grösster verbundener Wert = 35852

Aufgabe 2.2

Schritte
1. Berechne den Median
2. Berechne die 75%- und 25%- Quantile.
3. Berechne die verbundenen Werte.
4. Ausreisser
5. Zeichne den Tukey-Box-Plot

Schritt 3 (Berechnung der Nachbarschaftswerte)

$$\left.\begin{aligned} Q_{0.75} &= 34029 \\ Q_{0.25} &= 30469 \end{aligned}\right\} \; r \equiv Q_{0.75} - Q_{0.25} = 34029 - 30469 = 3560$$

Kleinster verbundener Wert: $\textit{kleinster Wert} \geq (25\%\ \textit{quantile}) - 1.5 \cdot r$

In diesem Fall, kleinster Wert grösser als $30469 - 1.5 \times 3560 = 25129$

Direction 1
24846
24862
25365
28252
29224
29976
30035
30613
32158
32472
32618
32962
33091
33197
33198

25129

Kleinster verbundener Wert = 25365

Aufgabe 2.2

Schritt 4 (Ausreisser)

Schritte
1. Berechne den Median
2. Berechne die 75%- und 25%- Quantile.
3. Berechne verbundenen Werte.
4. Ausreisser
5. Zeichne den Tukey-Box-Plot

Richtung 1	i	i/31
24846	1	0.03
24862	2	0.06
25365	3	0.10
28252	4	0.13
29224	5	0.16
29976	6	0.19
30035	7	0.23
30613	8	0.26
32158	9	0.29
32472	10	0.32
32618	11	0.35
32962	12	0.39
33091	13	0.42
33197	14	0.45
33198	15	0.48
33245	16	0.52
33380	17	0.55
33406	18	0.58
33788	19	0.61
33888	20	0.65
33937	21	0.68
34007	22	0.71
34013	23	0.74
34076	24	0.77
34425	25	0.81
34455	26	0.84
34576	27	0.87
35237	28	0.90
35843	29	0.94
35852	30	0.97

Ausreisser:

Ausserhalb der oberen und unteren Nachbarschaftswerte

24846

24862

Zusammenfassung

Grösster verbundener Wert:	35852
75% Quantil:	34029
Median :	33222
25% Quantil :	30469
Kleinster verbundener Wert:	25365

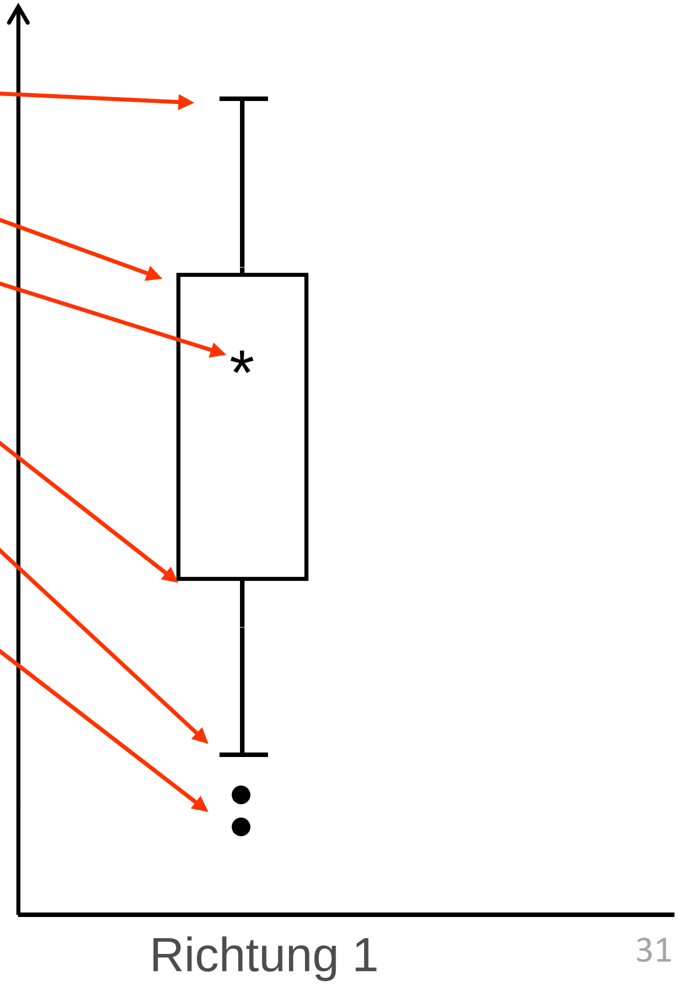
Aufgabe 2.2

Schritt 5 (Zeichne Tukey-Boxplot)

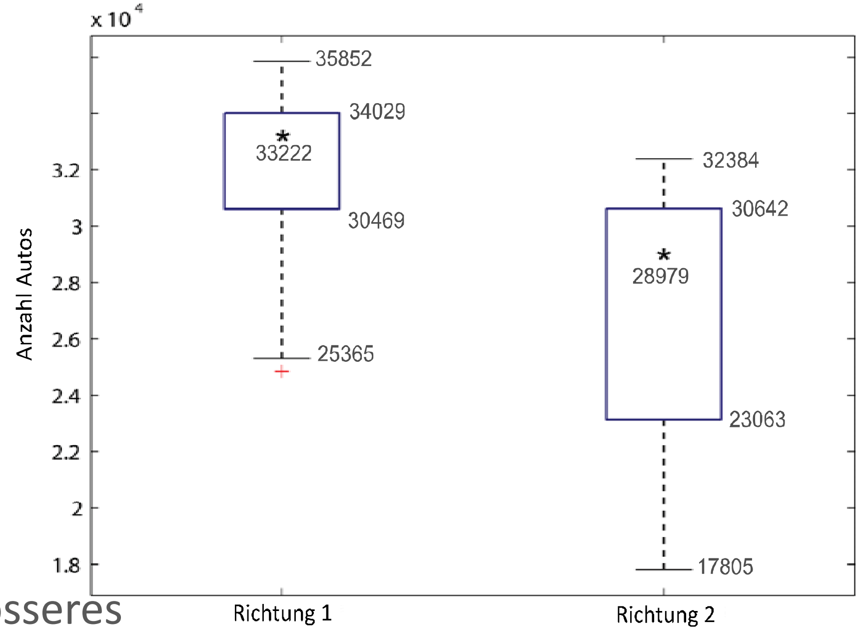
Schritte
1. Berechne den Median
2. Berechne die 75%- und 25%- Quantile.
3. Berechne die verbundenen Werte.
4. Ausreisser
5. Zeichne den Tukey-Boxplot

Zusammenfassung

grösster verbundener Wert:	35852
75% Quantil:	34029
Median :	33222
25% Quantil :	30469
kleinster verbundener Wert:	25365
Ausreisser:	24846
	24862



Aufgabe 2.2 - Lösung



Eigenschaften der jeweiligen Verteilung:

- Median
- verbundenen Werte
- oberes und unteres Quartil
(0,75 und 0.25%-Quantil)
- Ausreisser

Vergleich der Datenreihen:

- Alle Eigenschaften sind in Richtung1 grösser. Grösseres Verkehrsvolumen in Richtung 1
- Grössere Interquartiler Bereich in Richtung 2: Beobachtungen sind weiter gestreut um den Medium

Symmetrie der Datenreihen:

- keine Symmetrie beobachtet.
- Der Median ist bei beiden Datenreihen näher am oberen
- Nachbarschaftswert
- linksschief.

Aufgabe 2.5

Die Tabelle zeigt die Anzahl an Studienanfängern X und die Gesamtanzahl an Studierenden Y an einer Universität.
Die Korrelation dieser Zahlen ist mit Hilfe des Berechnungsblattes zu bestimmen.

	Uni A	Uni B	Uni C	Uni D	Uni E	Uni F
Studienanfänger	3970	732	499	1300	3463	2643
Studentenzahl	24273	5883	2847	5358	23442	17076

Tabelle 2.5.1 **Anzahl Studienanfänger und Studentenzahl (gesamt)**

Aufgabe 2.5

Die Korrelation dieser Beobachtungen ist mit Hilfe des Berechnungsblattes zu bestimmen.

	Uni A	Uni B	Uni C	Uni D	Uni E	Uni F
Studienanfänger	3970	732	499	1300	3463	2643
Studentenzahl	24273	5883	2847	5358	23442	17076

Tabelle 2.5.1 **Anzahl Studienanfänger und Studentenzahl (gesamt)**

Was ist bekannt?

Studienanfänger: X
Studentenzahl: Y
Studienanfänger: x_i, $i=1,\ldots,6$
Studentenzahlen: y_i, $i=1,\ldots,6$
Beobachtungen/Universität: $n=6$

Aufgabe 2.5

Die Korrelation dieser Beobachtungen ist mit Hilfe des Berechnungsblattes zu bestimmen.

	Uni A	Uni B	Uni C	Uni D	Uni E	Uni F
Studienanfänger	3970	732	499	1300	3463	2643
Studentenzahl	24273	5883	2847	5358	23442	17076

Tabelle 2.5.1 **Anzahl Studienanfänger und Studentenzahl (gesamt)**

Was ist bekannt?

Studienanfänger: X
Studentenzahl: Y
Studienanfänger: x_i, $i=1,\ldots,6$
Studentenzahlen: y_i, $i=1,\ldots,6$
Beobachtungen/Universität: $n=6$

Was wird gesucht?

Korrelation: $r_{XY} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})(y_i-\bar{y})}{s_X s_Y}$

Mittelwert: $\bar{x}$ $\bar{y}$

Standardabweichung: s_X s_Y

Aufgabe 2.5

Bestimmen Sie die Korrelation dieser Zahlen mit Hilfe des Berechnungsblattes.

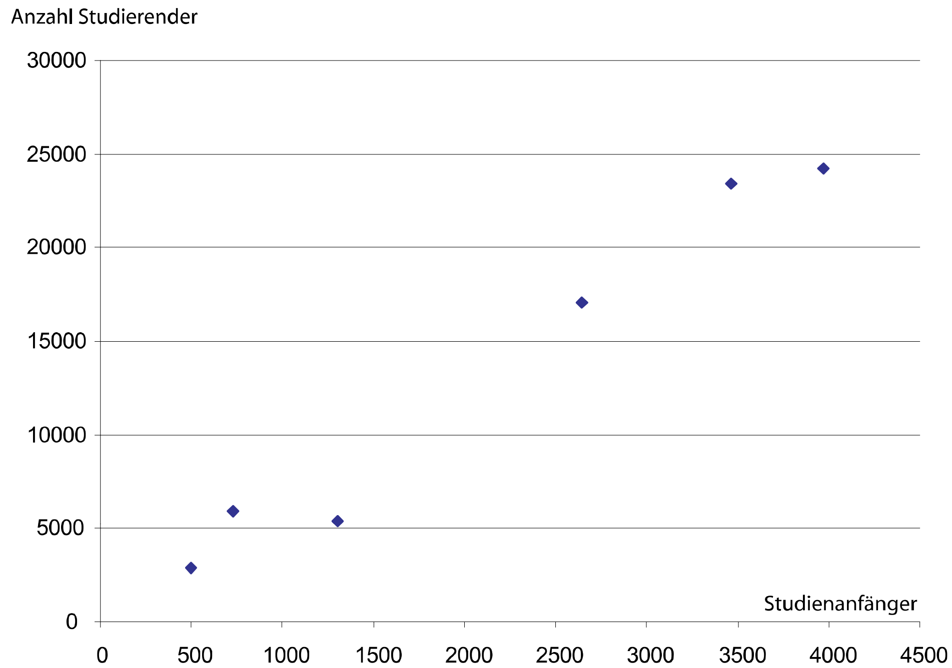
	Uni A	Uni B	Uni C	Uni D	Uni E	Uni F
Studienanfänger	3970	732	499	1300	3463	2643
Studentenzahl	24273	5883	2847	5358	23442	17076

Tabelle 2.5.1 **Anzahl Studienanfänger und Studentenzahl (gesamt)**

Auf den ersten Blick sind sie korreliert, oder??????

Gebe eine grobe Schätzung des Korrelationskoeffizienten !!!!

$$-1 \leq r_{XY} \leq 1$$



Aufgabe 2.5

Was ist bekannt?

Studienanfänger: *X*
Studentenzahl: *Y*
Anzahl Studienanfänger: x_i, $i=1,\dots,6$
Anzahl von Studentenzahlen: y_i, $i=1,\dots,6$
Anzahl Beobachtungen/Universität: $n=6$

	Uni A	Uni B	Uni C	Uni D	Uni E	Uni F
Studienanfänger	3970	732	499	1300	3463	2643
Studentenzahl	24273	5883	2847	5358	23442	17076

Tabelle 2.5.1 **Anzahl Studienanfänger und Studentenzahl (gesamt)**

Was ist gesucht?

Korrelation:

$$r_{XY} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \overline{x})(y_i - \overline{y})}{s_X\, s_Y}$$

Mittelwert:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Standardabweichung:

$$s_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad s_Y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Aufgabe 2.5 - Lösung

	Uni A	Uni B	Uni C	Uni D	Uni E	Uni F
Studienanfänger	3970	732	499	1300	3463	2643
Studentenzahl	24273	5883	2847	5358	23442	17076

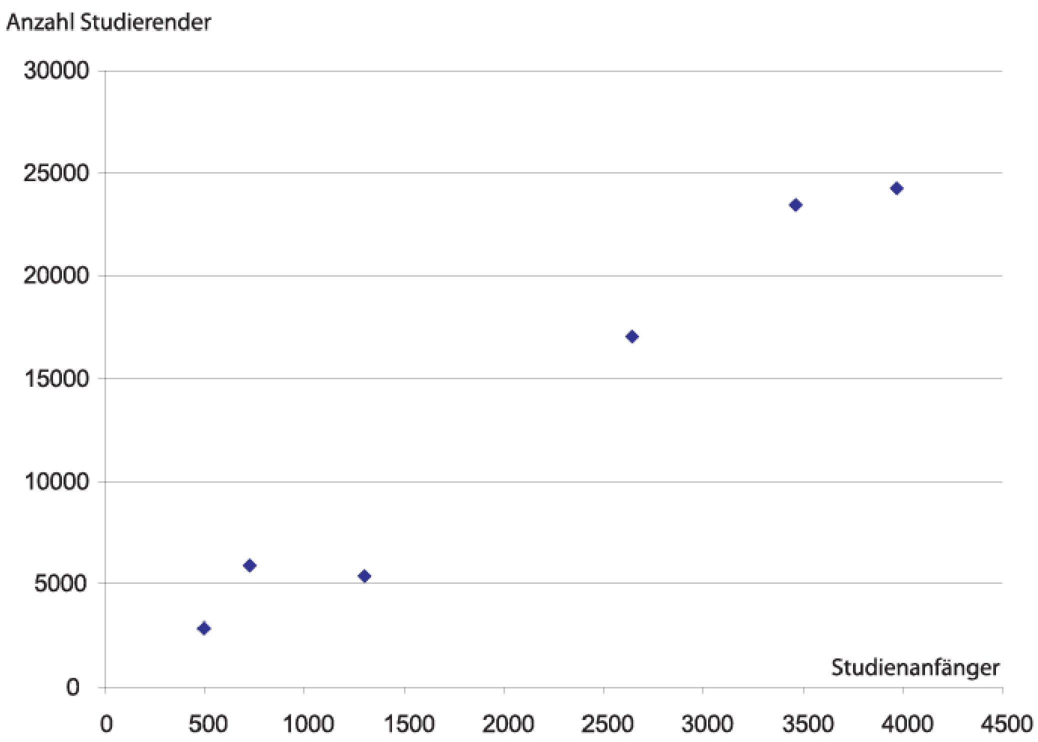
Tabelle 2.5.1 **Anzahl Studienanfänger und Studentenzahl (gesamt)**

	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 2.5 - Lösung

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

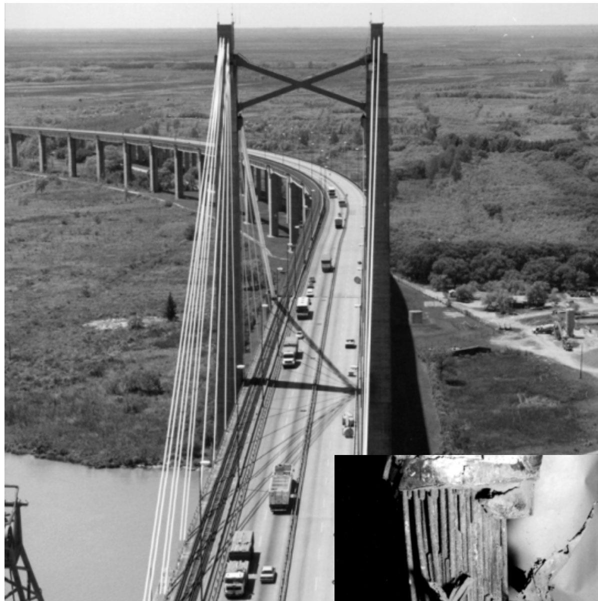
Wie erwartet, ist der Korrelationskoeffizient positiv.



Aufgabe 2.4 (Gruppenaufgabe)

Potentialfeldmessungen helfen dabei, die mögliche Korrosion in Brückentragwerken vorherzusagen. Während einer routinemässigen Untersuchung an einer Brücke wurden die Daten in folgender Tabelle durch Potentialfeldmessungen entlang der beiden Fahrspuren (Richtung 1 und 2) erhoben:

Messung Nr. (*i*)	Richtung 1 Widerstand (kOhm)	Richtung 2 Widerstand (kOhm)
1	20.2	3.8
2	20.4	5.6
3	22.1	6.5
4	23.8	7.1
5	24.3	7.9
6	24.7	8.2
7	25.3	9.1
8	25.6	9.3
9	25.7	9.6
10	25.9	9.8
11	26.2	10.3
12	26.7	10.9
13	26.9	11.1
14	27.3	11.7
15	27.6	12.2
16	27.6	12.6
17	27.8	12.9
18	27.9	13.8
19	28.3	13.9
20	28.7	14.5
21	28.9	15
22	28.9	15.4
23	29.3	17.1
24	29.4	17.8
25	29.9	23.4



Aufgabe 2.4 (Gruppenaufgabe)

a) Nutze die beiden Datenreihen aus der Tabelle und fertige zwei Tukey-Boxplots an (Richtung 1 und 2). Zeige die Hauptmerkmale der Tukey-Boxplots und schreibe deren Werte neben die korrespondierenden Punkte auf das Diagramm. Zeichne auch vorhandene Werte die ausserhalb liegen ein.

b) Der Tukey-Boxplot ist ein hilfreiches Werkzeug zur Bewertung der Symmetrie von Datenreihen. Diskutiere Symmetrie/Schiefe der Potentialfeldmessdaten der beiden Fahrspuren.

c) Wähle eine geeignete Anzahl von Klassen und zeichne ein Histogramm für die Potentialfeldmessdaten von Richtung 1.

d) Viel Erfolg ;-)