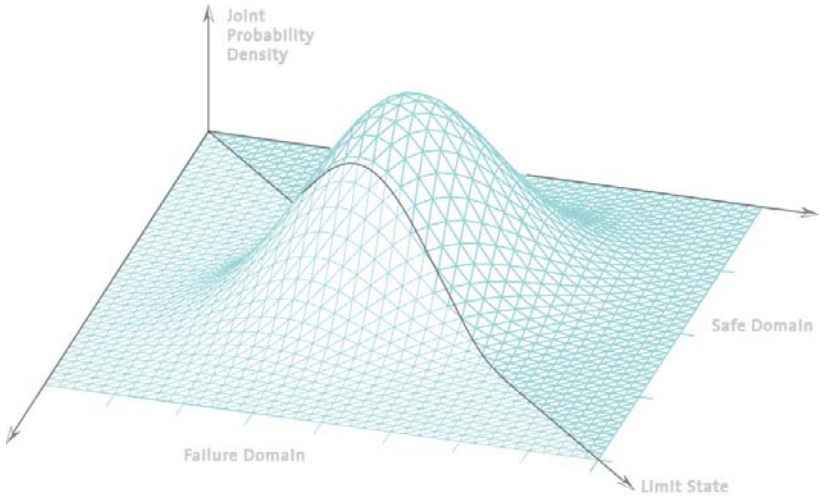


Exercises Tutorial 3

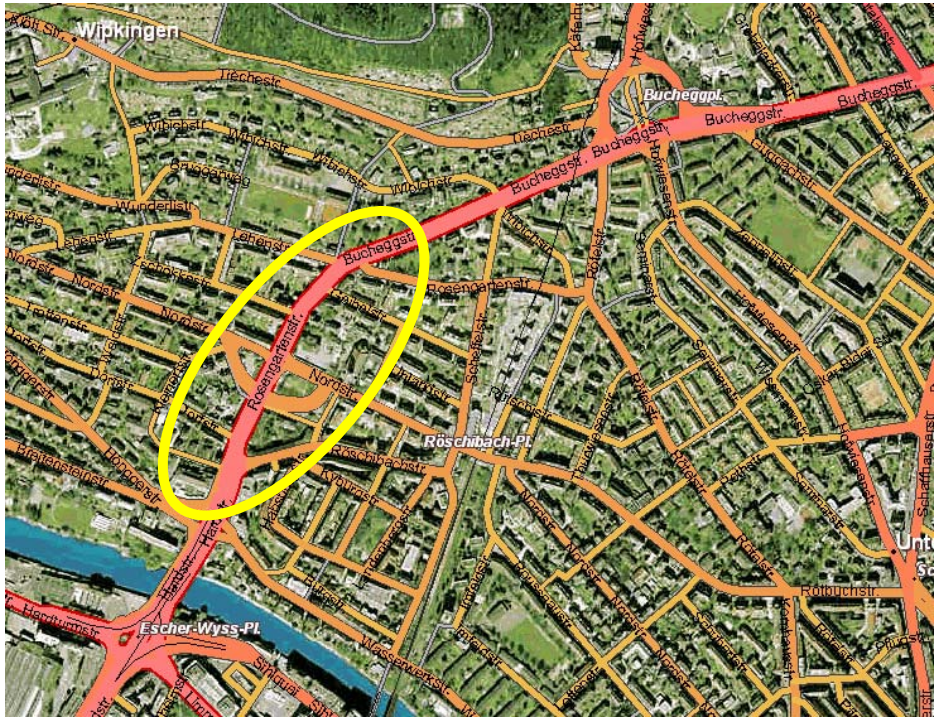
Statistics and Probability Theory

Prof. Dr. Michael Havbro Faber
Swiss Federal Institute of Technology Zurich
ETHZ

Exercise 3.1 (Descriptive Statistics)

Two sets of data are provided, each of which represents the daily traffic flow in Rosengartenstrasse in Zurich during the month of April 2001
Direction 1 corresponds to driving towards Bucheggplatz, while direction 2 corresponds to driving towards Escher Wyss Platz.

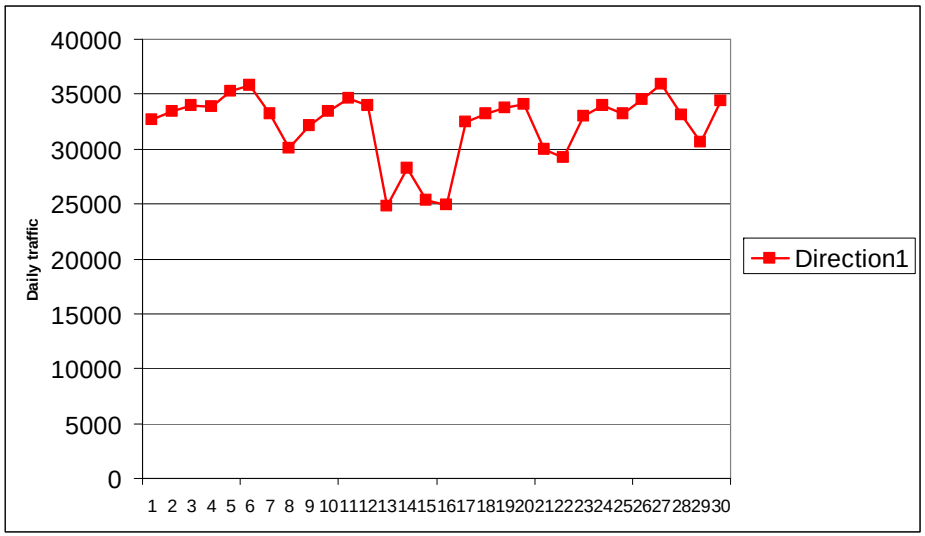
Date	Direction 1	Direction 2
01.04.2001	32618	24609
02.04.2001	33380	29965
03.04.2001	34007	30629
04.04.2001	33888	30263
05.04.2001	35237	31405
06.04.2001	35843	31994
07.04.2001	33197	26846
08.04.2001	30035	22762
09.04.2001	32158	30366
10.04.2001	33406	29994
11.04.2001	34576	30958
12.04.2001	34013	30680
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



What do we want to know?

Which is the best way to know it? – plot, histogram, statistics etc.

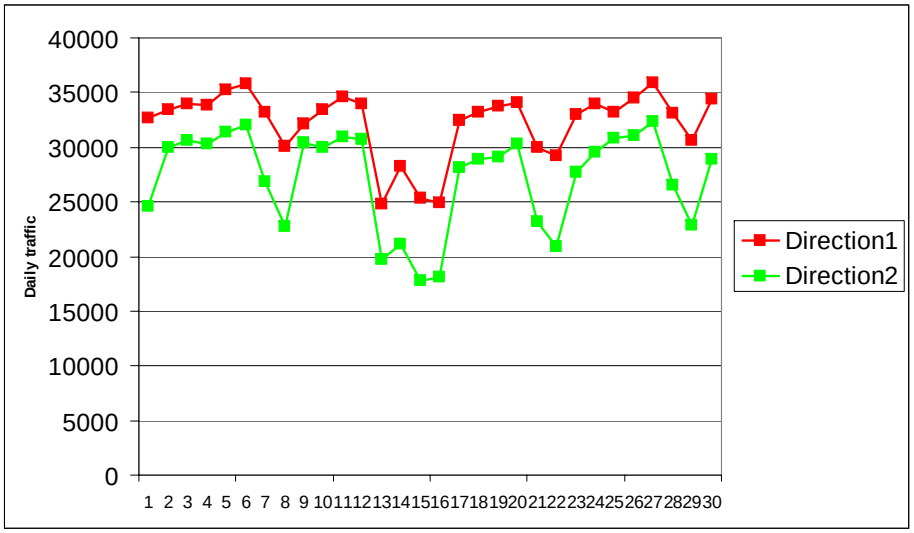
For example,
if you are interested in:
the change in the traffic of direction 1 during the month



What do we want to know?

Which is the best way to know it? – plot, histogram, statistics etc.

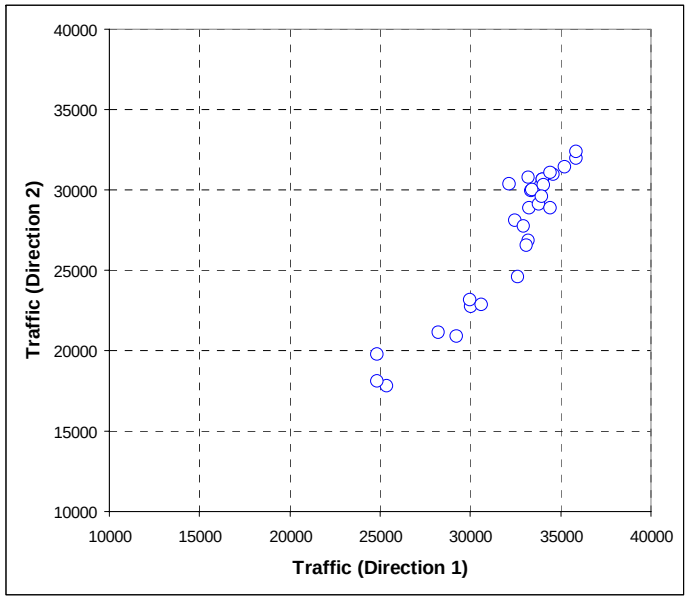
For example,
if you are interested in:
the relation between the traffic of direction 1 and that of direction 2,



What do we want to know?

Which is the best way to know it? – plot, histogram, statistics etc.

For example,
if you are interested in:
the relation between the traffic of direction 1 and that of direction 2,
but you are not interested in the time element

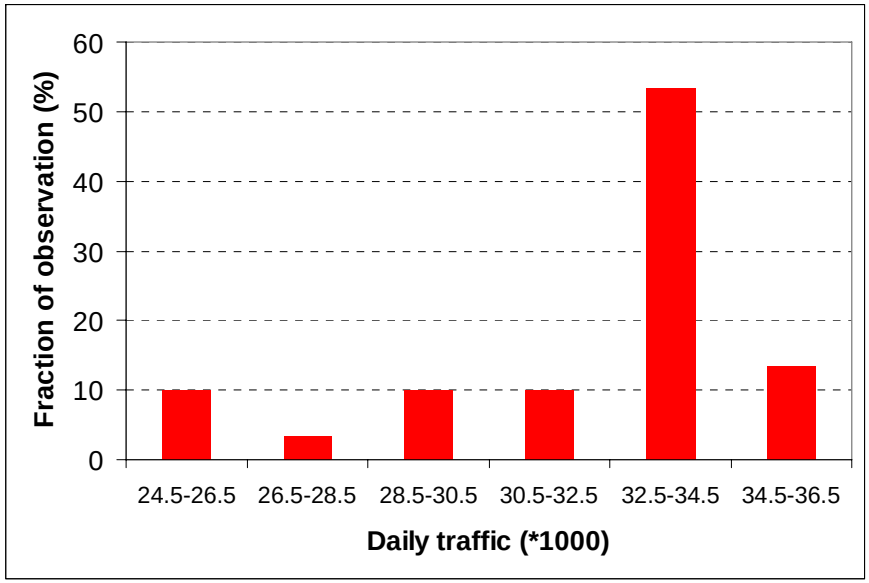


Correlated!

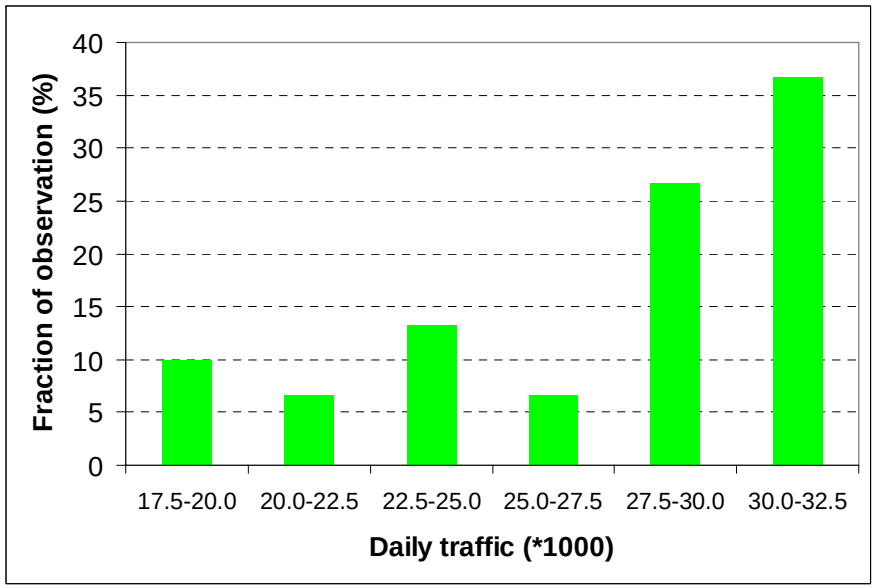
What do we want to know?

Which is the best way to know it? – plot, histogram, statistics etc.

For example,
if you are interested in:
traffic volume of each direction



Direction 1



Direction 2

We will look today into…

how to represent and compare the properties of sets of data which you have

- graphically
  - frequency distribution (histogram)
  - cumulative frequency distribution
- numerically
  - median
  - quantile
- a summary plot
  - Tukey box plot
- Correlation between data sets

You can use excel, matlab and/or other programming/statistics software….BUT

Make sure ALWAYS to insert functions by yourself or check that the functions provided by the used program agree with those of the used script!

Exercise 3.1

Provide frequency distributions and cumulative frequency distributions of the observed data. What is your first impression of the data? Try to make comparison between the two directions.

Date	Direction 1	Direction 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

Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Step 1 (sort the data)

Date	Direction 1	Direction 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

sort/order → in ascending order

Direction 1	Direction 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

Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Step 2 (select the number of intervals)

No general rule but suggestion - (script Equation (C.8))

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

k is the number of the intervals, n is the number of the data.

In this case, n = 30

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

For direction 1,

minimum = 24846
max = 35852

we may select the intervals as follows:

[24.5 26.5 28.5 30.5 32.5 34.5 36.5] (*1000)

Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Step 3 (count the data in each interval)

Direction 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

Count →

Direction 1	Interval (Number of cars *10^3)	Interval Midpoint (Number of cars *10^3)	Number of observations
	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

Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Step 4 (draw the frequency distribution)

But first some calculations….

	Interval (Number of cars *10^3)	Interval Midpoint (Number of cars *10^3)	Number of observations	Frequency %
Direction 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

$$\text{Frequency\%} = \frac{n_o}{n} 100$$

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

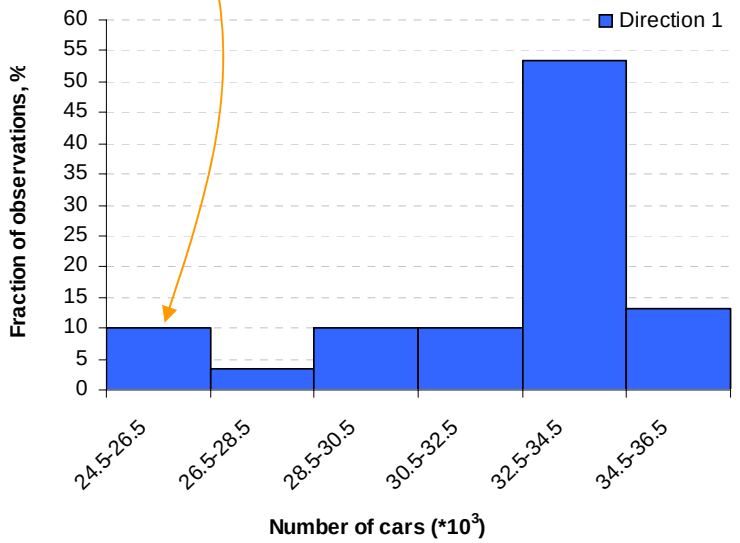
Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Step 4 (draw the frequency distribution)

	Interval (Number of cars *10^3)	Interval Midpoint (Number of cars *10^3)	Number of observations	Frequency %
Direction 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

Draw



Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Step 5 (draw the cumulative frequency distribution)

Cumulate →

	Interval (Number of cars *10^3)	Interval Midpoint (Number of cars *10^3)	Number of observations	Frequency %	Cumulative frequency
Direction 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

/100

Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

Step 5 (draw the cumulative frequency distribution)

Cumulate →

	Interval (Number of cars *10^3)	Interval Midpoint (Number of cars *10^3)	Number of observations	Frequency %	Cumulative frequency
Direction 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

/100 +

Steps

1. sort the data
2. select the number of intervals
3. count the data in each interval
4. draw the frequency distribution
5. draw the cumulative frequency distribution

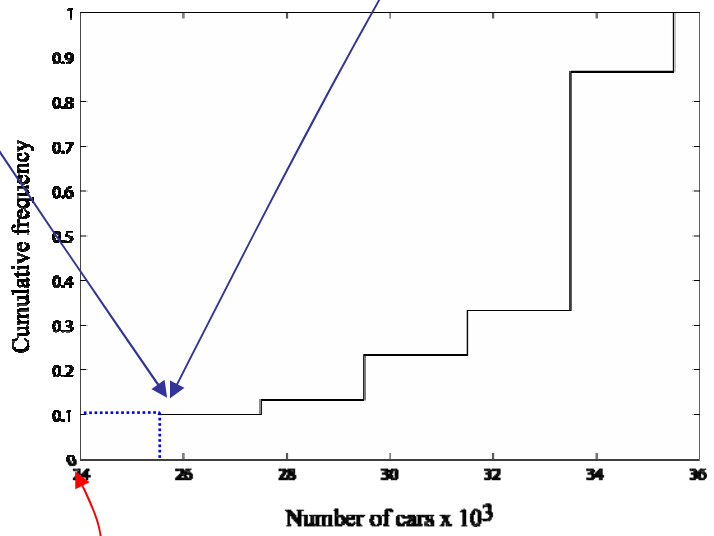
Step 5 (draw the cumulative frequency distribution)

Cumulate

Direction 1	Interval (Number of cars *10^3)	Interval Midpoint (Number of cars *10^3)	Number of observations	Frequency %	Cumulative frequency
	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

Draw

y=0

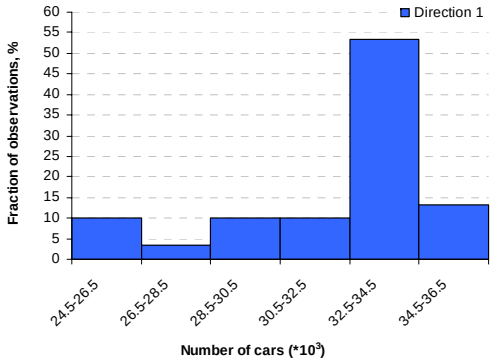


Answer 3.1

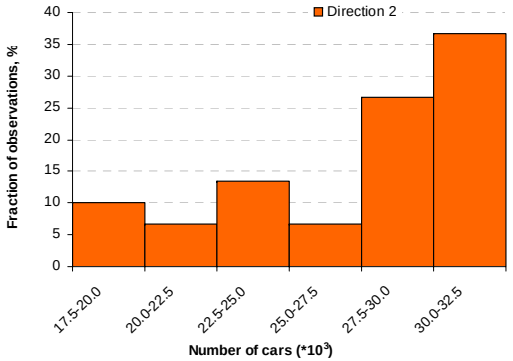
Do the same for direction 2.

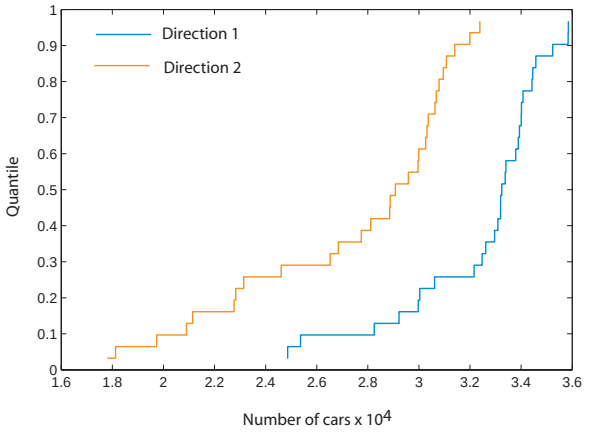
What can we know from these plots?

Direction 1



Direction 2





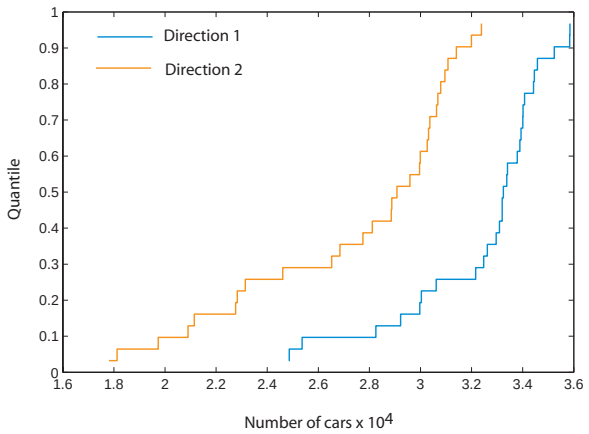
These figures give nice overviews of the data!

Answer 3.1

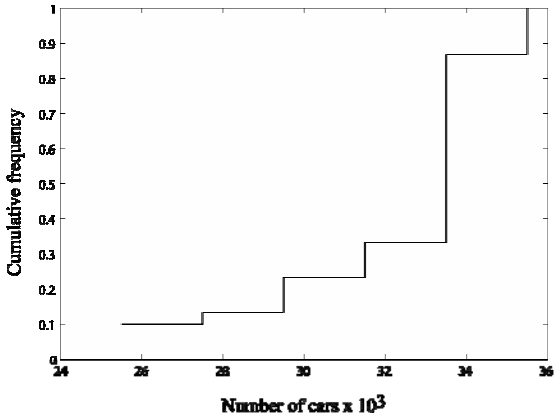
a. When we have in hand all observations:
prefer to plot the cumulative distribution plot using the quantiles of the data!

b. If we have in hand only the intervals observed and the frequency of observations within each interval
a. is not possible so…plot the cumulative frequency!

a.



b.



Quantiles

A quantile is related to a given percentage α, for which α% of all observations in the data set have smaller values.

e.g. the 0.65 quantile of a given data set of observations corresponds to the observation for which 65% of all observations in the data set have smaller values

Quantiles

A quantile is related to a given percentage α, for which α% of all observations in the data set have smaller values.

e.g. the 0.65 quantile of a given data set of observations corresponds to the observation for which 65% of all observations in the data set have smaller values

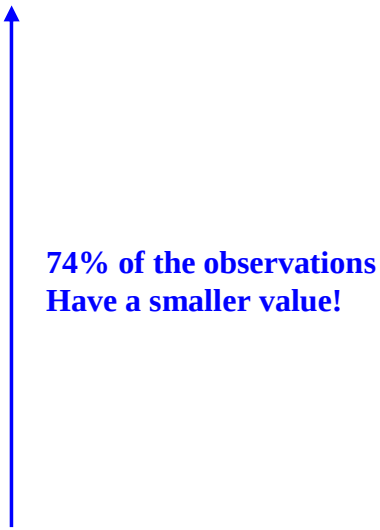
Date	Direction 1	Direction 2	
01.04.2001	32618	24609	
02.04.2001	33380	29965	
03.04.2001	34007	30629	
04.04.2001	33888	30263	
05.04.2001	35237	31405	
06.04.2001	35843	31994	
07.04.2001	33197	26846	
08.04.2001	30035	22762	
09.04.2001	32158	30366	
10.04.2001	33406	29994	
11.04.2001	34576	30958	
12.04.2001	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	**Q=0.74**
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	

Quantiles

A quantile is related to a given percentage α, for which α% of all observations in the data set have smaller values.

e.g. the 0.65 quantile of a given data set of observations corresponds to the observation for which 65% of all observations in the data set have smaller values

Date	Direction 1	Direction 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



74% of the observations Have a smaller value!

Q=0.74

Quantiles

A quantile is related to a given percentage α, for which α% of all observations in the data set have smaller values.

e.g. the 0.65 quantile of a given data set of observations corresponds to the observation for which 65% of all observations in the data set have smaller values

How to calculate it????

$$Q_i = \frac{i}{n+1}, \; n: \text{total number of obervations}$$

Exercise 3.2

Use the Tukey box plot to provide a summary of the main features of the distribution of each data set. Plot the Tukey box plots on the same graph so that you are able to compare these features. Do you observe any symmetry in the data sets?

Steps

1. calculate the median
2. calculate the 75%- and 25%- quantile.
3. calculate the adjacent values.
4. check for outside values
5. draw the Tukey box plot

Steps

1. calculate the median
2. calculate the 75%- and 25%- quantile.
3. calculate the adjacent values.
4. check for outside values
5. draw the Tukey box plot

Step 1 (calculate the median)

Just take the central value (50%-quantile).

but…..

if the number of data is even, this is not possible!

In that case, take the two values around the center, then take the average.

Direction 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

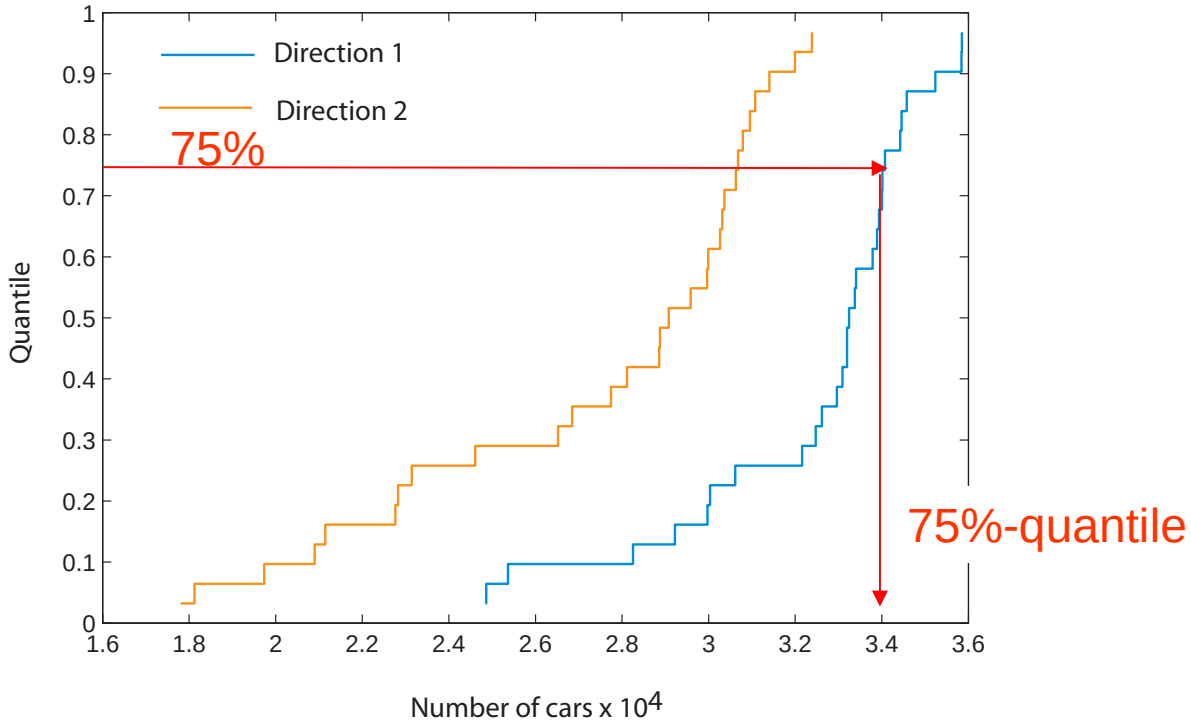
Median is $\frac{33198 + 33245}{2} = 33221.5$

Steps

1. calculate the median
2. calculate the 75%- and 25%- quantile.
3. calculate the adjacent values.
4. check for outside values
5. draw the Tukey box plot

Step 2 (calculate the quantiles)

Roughly speaking,



Step 2 (calculate the quantiles)

More strictly speaking,

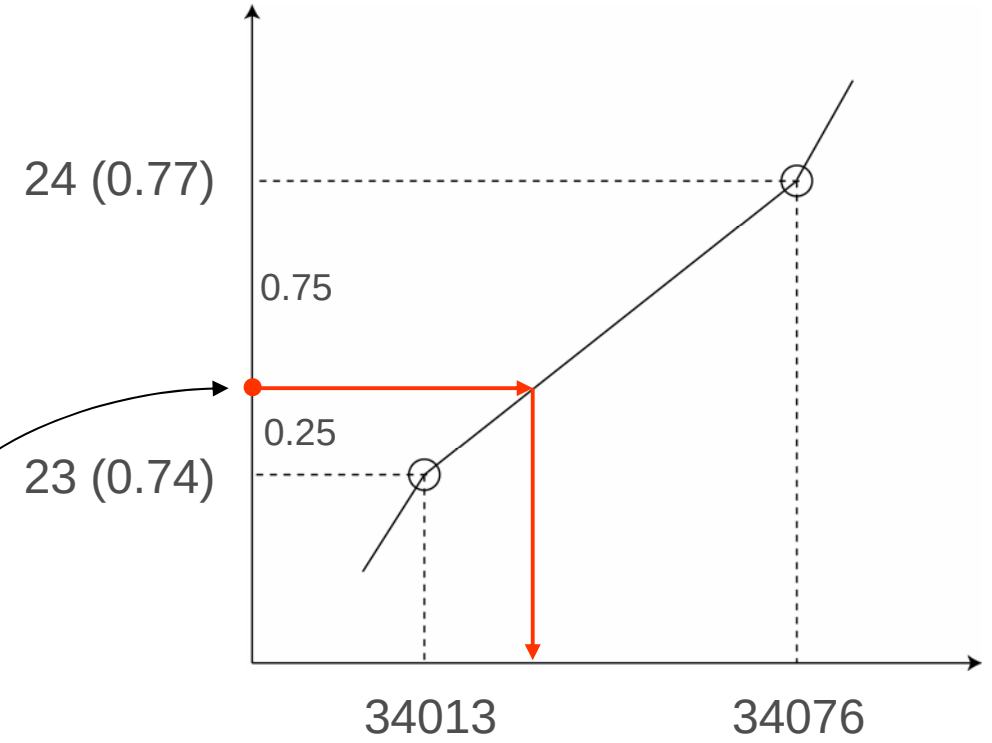
$$Q_i = \frac{i}{n+1},\ n: \text{total number of obervations}$$

Direction 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	

Step 2 (calculate the quantiles)

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



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

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

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

Steps
1. calculate the median
2. calculate the 75% and 25% quantile.
3. calculate the adjacent values.
4. check for outside values
5. draw the Tukey box plot

Step 3 (calculate the adjacent values)

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

Interquartile range

Upper adjacent value: largest observation $\leq$ (75% *quantile*) + 1.5*r*

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

Upper adjacent value =

Steps
1. calculate the median
2. calculate the 75% and 25% quantile.
3. calculate the adjacent values.
4. check for outside values
5. draw the Tukey box plot

Step 3 (calculate the adjacent values)

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

Lower adjacent value: smallest observation $\geq$ (25% *quantile*) - 1.5r

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

lower adjacent value =

Steps

1. calculate the median
2. calculate the 75% and 25% quantile.
3. calculate the adjacent values.
4. check for outside values
5. draw the Tukey box plot

Direction 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

Step 4
(check for outside values)

Outside values:
Outside the upper and lower adjacent values

24846
24862

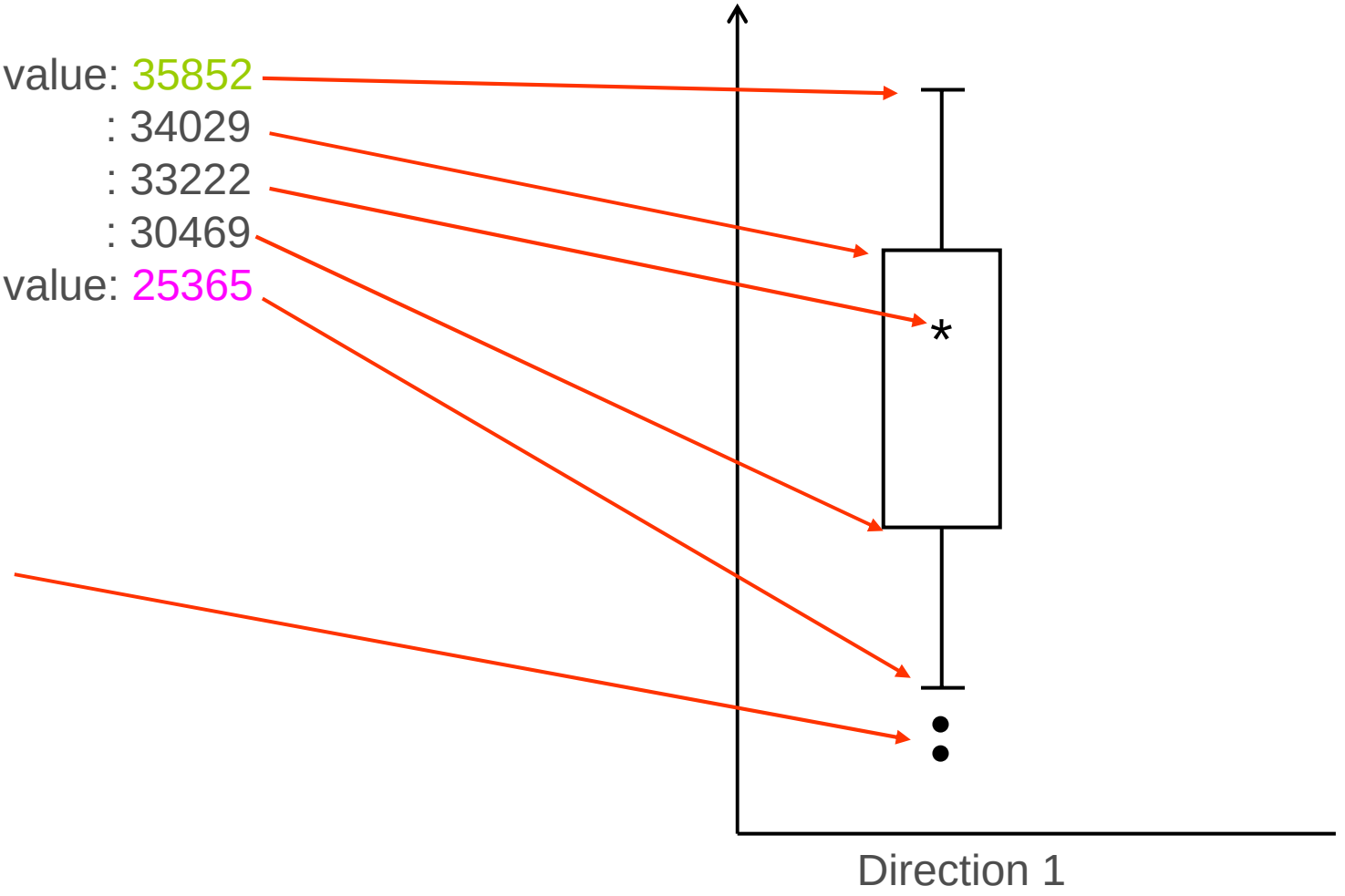
summary
Upper adjacent value: 35852
75% quantile : 34029
Median : 33222
25% quantile : 30469
Lower adjacent value: 25365

Steps
1. calculate the median
2. calculate the 75% and 25% quantile.
3. calculate the adjacent values.
4. check for outside values
5. draw the Tukey box plot

Step 4
(draw the Tukey box plot)

summary
Upper adjacent value: 35852
75% quantile : 34029
Median : 33222
25% quantile : 30469
Lower adjacent value: 25365

Outside values:
24846
24862



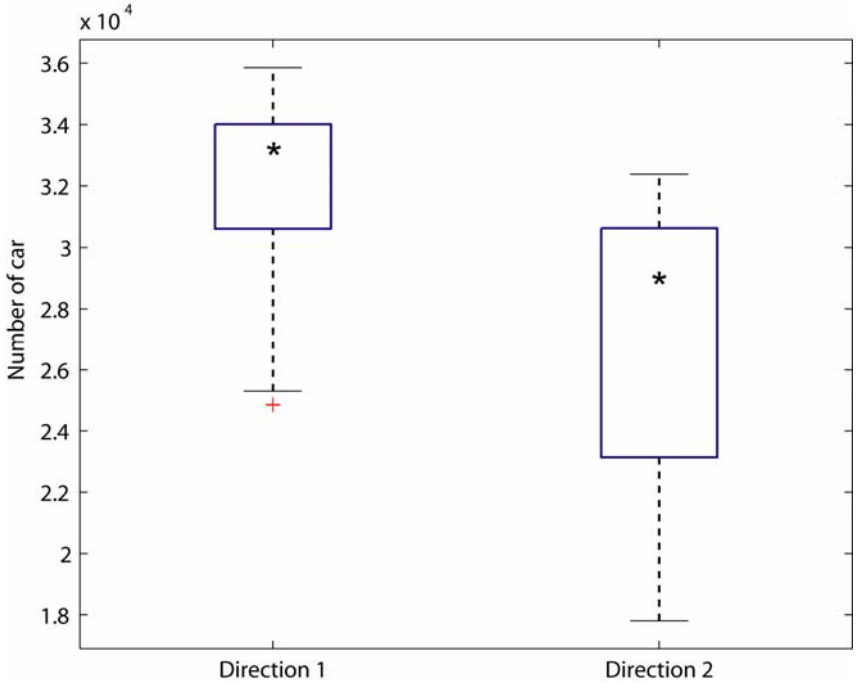
Answer 3.2

Use the Tukey box plot to provide a summary of the main features of the distribution of each data set.

- median
- Adjacent values
- Upper and lower quartiles
- Outside values

Plot the Tukey box plots on the same graph so that you are able to compare these features.
???

Do you observe any symmetry in the data sets?
???



Exercise 3.5

The data sets in Table 3.5.1 show the number of newcomers to the university and the number of total students at the university.
Estimate the correlation of these numbers using the following calculation sheet.

	Univ. A	Univ. B	Univ. C	Univ. D	Univ. E	Univ. F
Newcomer	3970	732	499	1300	3463	2643
Total students	24273	5883	2847	5358	23442	17076

Table 3.5.1 Number of newcomers to the university and the number of total students at the university.

Exercise 3.5

Estimate the correlation of these numbers using the following calculation sheet.

	Univ. A	Univ. B	Univ. C	Univ. D	Univ. E	Univ. F
Newcomer	3970	732	499	1300	3463	2643
Total students	24273	5883	2847	5358	23442	17076

Table 3.5.1 Number of newcomers to the university and the number of total students at the university.

What is known?

Newcomers: X
total students: Y
Number of newcomers: x_i, $i=1,\ldots,6$
Number of total students: y_i, $i=1,\ldots,6$
Number of observations/university: $n=6$

Exercise 3.5

Estimate the correlation of these numbers using the following calculation sheet.

	Univ. A	Univ. B	Univ. C	Univ. D	Univ. E	Univ. F
Newcomer	3970	732	499	1300	3463	2643
Total students	24273	5883	2847	5358	23442	17076

Table 3.5.1 Number of newcomers to the university and the number of total students at the university.

What is known?

Newcomers: X
total students: Y
Number of newcomers: x_i, $i=1,\ldots,6$
Number of total students: y_i, $i=1,\ldots,6$
Number of observations/university: $n=6$

What is required?

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

Need to:
Calculate the sample mean values: $\bar{x}$ $\bar{y}$

Calculate sample standard deviations: s_X s_Y

Exercise 3.5

Estimate the correlation of these numbers using the following calculation sheet.

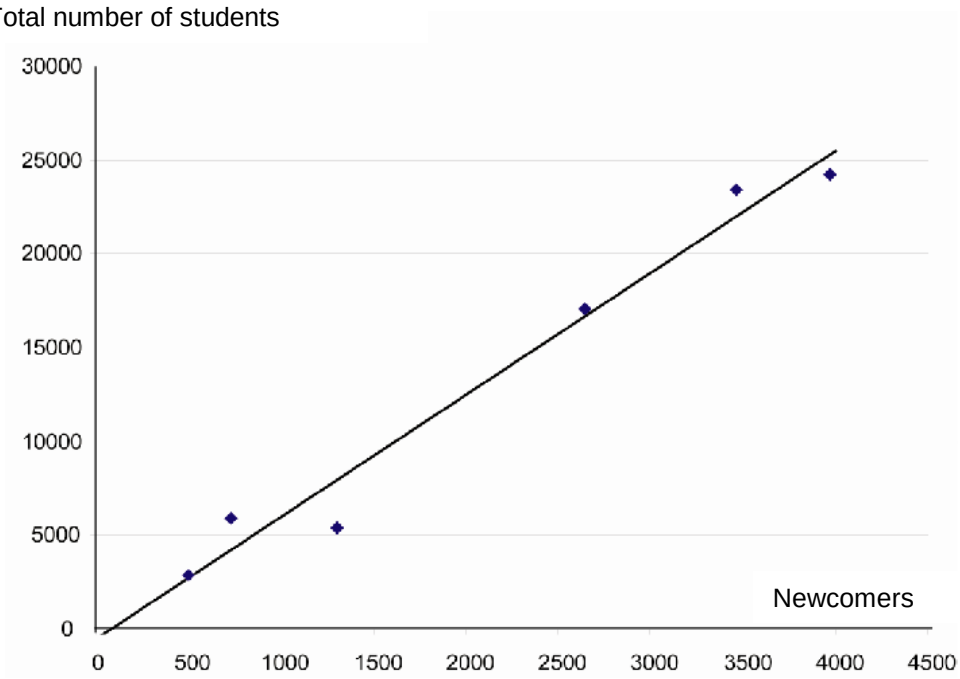
	Univ. A	Univ. B	Univ. C	Univ. D	Univ. E	Univ. F
Newcomer	3970	732	499	1300	3463	2643
Total students	24273	5883	2847	5358	23442	17076

Table 3.5.1 Number of newcomers to the university and the number of total students at the university.

From a first view are they correlated ???????

Give a rough estimation for the correlation coefficient!!!!

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



Solution 3.5

	Univ. A	Univ. B	Univ. C	Univ. D	Univ. E	Univ. F
Newcomer	3970	732	499	1300	3463	2643
Total students	24273	5883	2847	5358	23442	17076

Table 3.5.1 Number of newcomers to the university and the number of total students at the university.

What is known?

Newcomers: X
total students: Y
Number of newcomers: x_i, $i=1,\ldots,6$
Number of total students: y_i, $i=1,\ldots,6$
Number of observations/university: $n=6$

What is required?

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

Need to:
Calculate the sample mean values: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Calculate sample standard deviations: $s_X = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \qquad s_Y = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})$

Solution 3.5

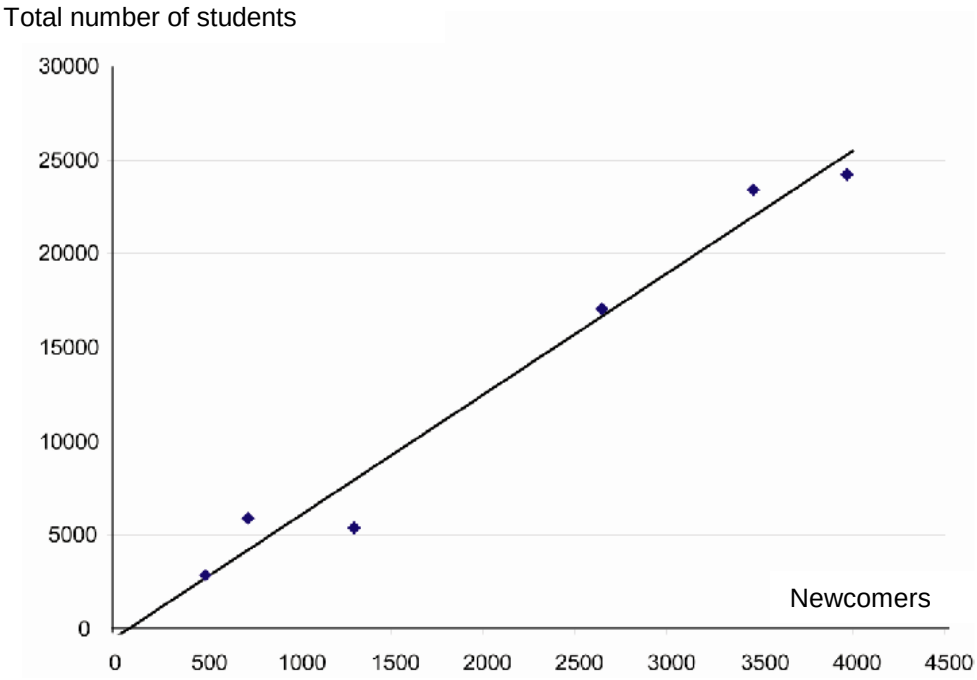
	Univ. A	Univ. B	Univ. C	Univ. D	Univ. E	Univ. F
Newcomer	3970	732	499	1300	3463	2643
Total students	24273	5883	2847	5358	23442	17076

Table 3.5.1 Number of newcomers to the university and the number of total students at the university.

	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
A	3970	24273	1868.83	11126.5	3492538	123799002	20793574
B	732	5883	-1369.17	-7263.5	1874617	52758432	9944942
C	499	2847	-1602.17	-10299.5	2566938	106079700	16501516
D	1300	5358	-801.17	-7788.5	641868	60660732	6239887
E	3463	23442	1361.83	10295.5	1854590	105997320	14020755
F	2643	17076	541.83	3929.5	293583	15440970	2129134
Σ	12607	78879	-	-	10724135	464736158	69629808
Σ / n	2161.17	13146.5	-	-	1787356	77456026	11604968
$\sqrt{\Sigma / n}$	-	-	-	-	1336.92	8800.91	-

Solution 3.5

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



Exercise 3.4 (Group exercise- to be presented on 12.04.07)

3.4.1

Resistivity measurements help to predict the possible corrosion of bridge structures. During a general bridge inspection the data shown in Table 3.2 were obtained from resistivity measurements along the two bridge lanes (direction 1 and 2):

a. Draw two box plots for the data provided in Table 3.4.1 (direction 1 and direction 2). Show the main features of the box plots and write their values next to the corresponding points on the diagrams. Plot also the outside values, if any.

b. Tukey box plot is a helpful tool for assessing the symmetry of data sets. Discuss the symmetry/skewness of the resistivity data for both lanes.

c. Choose a suitable number of intervals and plot the histogram for the resistivity data of direction 1.

Exercise 3.4 (Group exercise- to be presented on 12.04.07)

a. Draw two box plots for the data provided in Table 3.4.1 (direction 1 and direction 2). Show the main features of the box plots and write their values next to the corresponding points on the diagrams. Plot also the outside values, if any.

b. Tukey box plot is a helpful tool for assessing the symmetry of data sets. Discuss the symmetry/skewness of the resistivity data for both lanes.

According to exercise 3.2!!!

c. Choose a suitable number of intervals and plot the histogram for the resistivity data of direction 1.

According to exercise 3.1!!!

What should be in the presentation of the solution?

a, b and c!!

An example of calculation where applicable e.g. features of the Tukey box plot etc….

Try to work with a simple calculator, diagrams can be on a transparency made by hand☺

You can try for yourself to solve in e.g. excel or matlab or other.