Statistik I - Exercise session 2
12.5.2014 & 19.5.2014

Info

- Classroom: SPA1 220
- Time: Mondays, 16:15 - 17:45
- in English
- Assignments on webpage (lvb>staff>PB)

Contact: Petra Burdejova
petra.burdejova@hu-berlin.de
Office: SPA1 R400 (upon agreement)

Schedule:

Date	Week	Exercises
28.04.14	E1	1-2, 1-3 (even), 1-10
05.05.14	E1	1-2, 1-3 (even), 1-10
12.05.14	E2	1-20, 1-22, 1-32
19.05.14	E2	1-20, 1-22, 1-32
26.05.14	E3	1-80, 1-83, (1-98)
02.06.14	E3	1-80, 1-83, (1-98)
09.06.14	–	–
16.06.14	E4	TBA
23.06.14	E5	TBA
30.06.14	E5	TBA
07.07.14	E6	TBA
14.07.14	E6	TBA

Review

- week 3 (holidays) & week 4
- Slides: Descriptive Statistics (cca 30-60)

Mean

average value, calculated as a sum of sampled values divided by the number of observations (items in the sample)

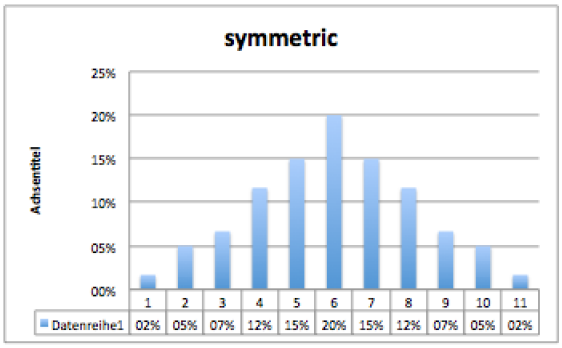
Mode

value having the highest frequency (i.e. appears most often in sample)

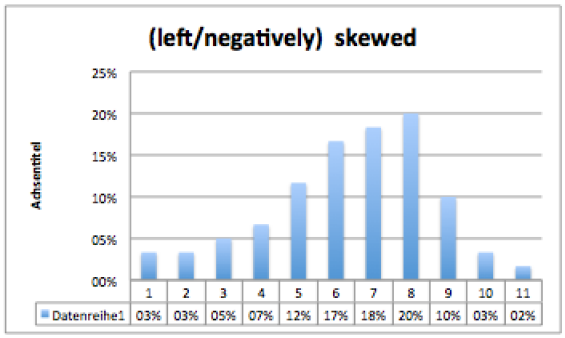
Median

middle value, separating lower half of ordered sample from upper half

value	frequency	rel.frequency		cumulative
1	1	0,017	1,7%	1,7%
2	3	0,050	5,0%	6,7%
3	4	0,067	6,7%	13,4%
4	7	0,117	11,7%	25,0%
5	9	0,150	15,0%	40,0%
6	12	0,200	20,0%	60,0%
7	9	0,150	15,0%	75,0%
8	7	0,117	11,7%	86,7%
9	4	0,067	6,7%	93,4%
10	3	0,050	5,0%	98,4%
11	1	0,017	1,7%	100,0%
TOTAL	60	1,000	100%	
MEAN	6			
MEDIAN	6			
MODE	6			



value	frequency	rel.frequency		cumulative
1	2	0,033	3,3%	3,3%
2	2	0,033	3,3%	6,6%
3	3	0,050	5,0%	11,6%
4	4	0,067	6,7%	18,3%
5	7	0,117	11,7%	30,0%
6	10	0,167	16,7%	46,6%
7	11	0,183	18,3%	65,0%
8	12	0,200	20,0%	85,0%
9	6	7,000	10,0%	95,0%
10	2	0,033	3,3%	98,3%
11	1	0,017	1,7%	100,0%
TOTAL	60	1,000	100%	
MEAN	6,4			
MEDIAN	7			
MODE	8			



Mean

ungrouped variables $\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

classified/grouped and discrete variable $\overline{x} = \frac{1}{n} \cdot \sum_{j=1}^{k} x_j \cdot h(x_j) = \sum_{j=1}^{k} x_j \cdot f(x_j)$

weighted $\overline{x} = \frac{\sum_{i=1}^{n} x_i \cdot g_i}{\sum_{i=1}^{n} g_i}$

pooled $\overline{x} = \sum_{\ell=1}^{r} \frac{n_\ell}{n} \overline{x}_\ell$

n_ℓ observations and $\overline{x}_\ell$ middle value in group ℓ

Mode/Modus

unclassified data $x_D = \left\{ x_j \mid h_j = \max_{x_k} h_k \text{ or } f_j = \max_{x_k} f(x_k) \right\}$

classified data $x_D = x_j^u + \frac{\hat{f}(x_j) - \hat{f}(x_{j-1})}{2 \cdot \hat{f}(x_j) - \hat{f}(x_{j-1}) - \hat{f}(x_{j+1})} \cdot (x_j^{up} - x_j^{down})$

Median

unclassified data $x_{0,5} = x_{\left(\frac{n+1}{2}\right)}$ if n odd

$x_{0,5} = \frac{1}{2} \cdot \left\{ x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right\}$ if n even

classified data $x_{0,5} = x_j^{down} + \frac{0,5 - F(x_j^{down})}{f(x_j)} \cdot (x_j^{up} - x_j^{down})$

p - Quantile

classified data $x_p = x_{(k)}$ if $n \cdot p$ not integer and k is next integer after $n \cdot p$

$x_p = \frac{1}{2} \cdot \left\{ x_{(k)} + x_{(k+1)} \right\}$ if $n \cdot p$ is integer and $k = n \cdot p$

classified data $x_p = x_j^u + \frac{p - F(x_j^{down})}{f(x_j)} \cdot (x_j^{up} - x_j^{down})$ for $0 < p \leq 1$

Quartiles

0,25-quantile, 0,5-quantile(median), 0.75-quantile

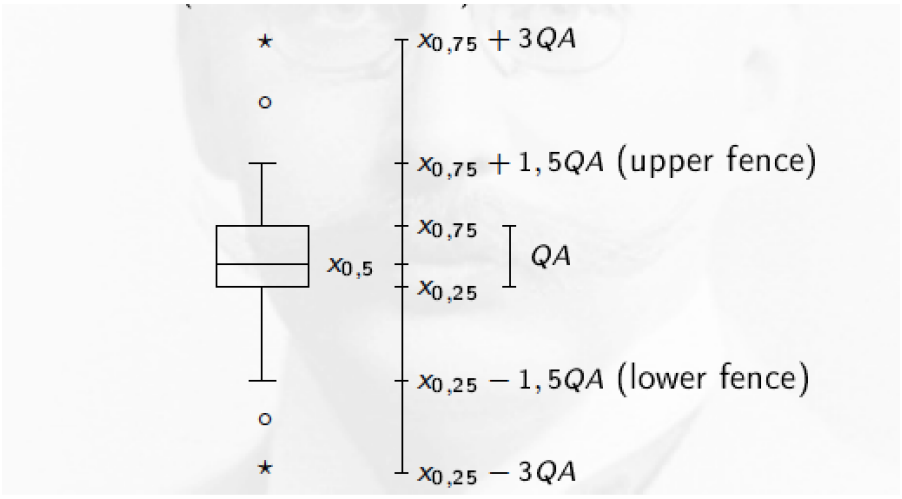
Harmonic mean

simple $\overline{x}_H = \dfrac{n}{\sum\limits_{i=1}^{n} \frac{1}{x_i}}$

weighted $\overline{x}_H = \dfrac{\sum\limits_{j=1}^{k} g_j}{\sum\limits_{j=1}^{k} \frac{g_j}{x_j}}$ with $x_j = \dfrac{g_j}{h_j}, \quad j = 1, \ldots, k$

	Qualitative		Quantitative		
	Nominal	Ordinal	Interval	Ratio	Absolute
Mean	✗	✗	✓	✓	✓
Median	✗	✓	✓	✓	✓
Mode	✓	✓	✓	✓	✓

Boxplot:
plot of 5-values-summary



Exercises

Exercise 1-20

Based on Exercise 1-10 (see table below)

No.	Name	Studies	No. of Siblings	Income
1	Martin A.	Economics	0	924
2	Ute A.	Social S.	1	789
3	Wilhelm A.	Business	0	1365
4	Kurt B.	Business	1	683
5	Sylvia B.	Polit. S.	1	744
6	Elke D.	Polit. S.	2	640
7	Klaus D.	Social S.	2	631
8	Theo E.	Economics	1	814
9	Jean F.	Polit. S.	1	778
10	Elvira G.	Business	0	1062
11	Karl H.	Business	0	1230
12	Andreas K.	Economics	1	700
13	Thomas K.	Business	0	850
14	Chris L.	Social S.	3	641
15	Uwe L.	Polit. S.	2	640
16	Axel M.	Business	0	850
17	Maria M.	Business	1	683
18	Ruth M.	Social S.	0	616
19	Bärbel N.	Business	1	683
20	Armin R.	Business	2	683
21	Christa R.	Economics	1	660
22	Bernd S.	Business	1	1440
23	Claudia S.	Social S.	3	794
24	Udo T.	Economics	0	660
25	Clausia W.	Polit. S.	1	640

Field of studies	h	f
Econom.	5	0.2
Business	10	0.4
Polit. S	5	0.2
Social S.	5	0.2

No of sibl.	h	f
0	8	0.32
1	11	0.44
2	4	0.16
3	2	0.08

Income	h	f	F
$[600; 650)$	6	0.24	0.24
$[650; 700)$	6	0.24	0.48
$[700; 900)$	8	0.32	0.8
$[900; 1200)$	2	0.08	0.88
$[1200; 1450)$	3	0.12	1.0

a) Determine an appropriate ”average/middle value” for the survey characteristic X: ”Studies”.

b) Determine three appropriate average/middle values for the survey characteristic Y: ”Number of siblings”.

c) Based on the original list calculate the average income of students.

d) Based on the frequency distribution of classified/grouped data for variable Z: ”Income” calculate 90% quantile and quartiles

e) State the 5-values-summary for variable ”Income” and plot the box-plot.

Exercise 1-22

Mr. Meier owns a garden business with three branches: Berlin, New York and Flensburg. At the end of fiscal year he wants to have an overview of business situation. Therefore he asks all of three branches for information about the orders received last year.

A) Berlin sent the following information:

Value of orders	Number of orders
0 - 20 000	15
20 000 - 50 000	30
50 000 - 150 000	45
150 000 - 300 000	10

a) What is the examined variable? How is it scaled?

b) For which statistical unit was this variable observed?

c) Calculate the average value per order.

B) New york replied to Mr. Meier briefly:
Y: ”the value (in \$) per order” ; 95 orders ; $\bar{y} = 60000$ \$.
For a better comparison, Mr. Meier wants to see this statement in €.
Calculate the average value per order when exchange rate 1 \$ = 1.5 €.

C) From Flensburg Mr. Meier received the following data:
2 000 €, 12 000 €, 17 000 €, 12 000 €, 200 000 €.

a) What average value(s) can you state meaningfully?
Justify their advantages/disadvantages in this situation.

b) Fax corrects the value of the last order to 2 000 €.
Calculate the average value per order from corrected data.

D) Calculate the average value per order for the entire company.

Exercise 1-32

In a rolling department 4 workers (A,B,C,D) use different modern machines. They need the following average times for rolling a piece of metal sheet:

A	20 sec/piece
B	30 sec/piece
C	60 sec/piece
D	60 sec/piece

I. Consider a random variable X = Processing Time (sec/piece).

a) Suppose the workers work the same long time. What is the average time per piece needed in this department?

b) Suppose the workers have the following quotas:

A	1000 pieces
B	500 pieces
C	300 pieces
D	200 pieces

What is the average time per piece needed in the department now?

II. Consider a random variable Y = pieces produced per hour

a) Suppose the workers work the same long time. How many pieces per hour are produced on average in this department?

b) Suppose the workers have the the same quotas as in I-b).
How many pieces per hour will now be produced on average?