Biometry practical 3

Illustrated (imperfect) practical guide

Preparatory work

1. Open in *MS Excel* the questionary data (file analysed already in previous practical),
2. insert new worksheet,
3. rename new worksheet to 'Praks3' (or 'Practical3') and

4. make a copy of the data table (from worksheet 'Andmed') and paste it into the upper left corner of the worksheet 'Praks3'.

Exercise 1.

- Leave at least one empty row below data table and calculate for all numerical variables number of observations (*n*), average ($\bar{x}$), median, standard deviation (*s*), standard error (*se*), minimum and maximum using *Excel* functions.
- Add into data table new variable BMI (body mass index) and calculate it's values for all students applying the formula

$$\text{BMI} = \text{Weight, kg} / (\text{Height, m})^2.$$

Calculate all mentioned characteristics also for new variable.

Guide

1. Leave at least one empty row below data table

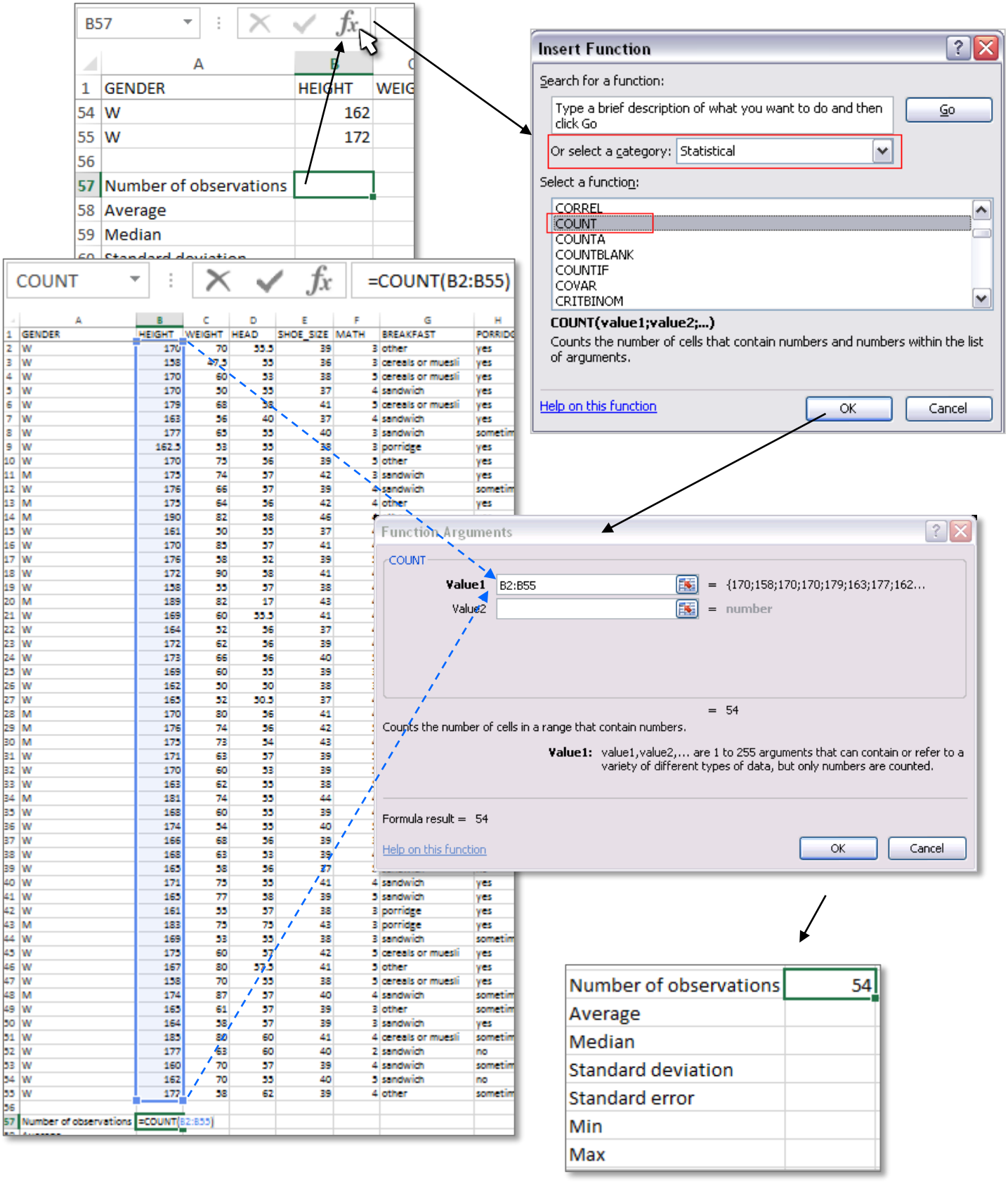
   (it is necessary to separate summary statistics from initial data to avoid considering these calculated values as a part of database – for example in case of sorting the data or applying the PivotTable)

   and write into the first column the names of calculable characteristics (then it is later more easier to understand, what is done).

	A	B	C	
1	GENDER	HEIGHT	WEIGHT	H
54	W	162	70	
55	W	172	58	
56				
57	Number of observations			
58	Average			
59	Median			
60	Standard deviation			
61	Standard error			
62	Min			
63	Max			

2. Use *Excel* functions to calculate desired characteristics for students' height (variable HEIGHT).

a. For this you can select appropriate function from the list of *Excel* functions (the names of functions are listed at the beginning of next page):



b. Knowing the function name, you can just type the desired formula into the corresponding cell.
(NB! Don't forget to start the formula with sign =)

Number of observations	=COUNT(B2:B55)
Average	=AVERAGE(B2:B55)
Median	=MEDIAN(B2:B55)
Standard deviation	=STDEV.S(B2:B55)
Standard error	
Min	=MIN(B2:B55)
Max	=MAX(B2:B55)

All these functions can be applied also in previously described way – select yourself, which is more simple for you (try both variants).

c. As in *Excel* there is no function for **standard error**, the calculations must be performed following the standard error formula

$$se = s/\sqrt{n}$$

(this means, that you must input the formula just typing necessary commands):

	A	B
1	GENDER	HEIGHT
56		
57	Number of observations	54
58	Average	170.1944444
59	Median	170
60	Standard deviation	7.397358444
61	Standard error	=B60/SQRT(B57)
62	Min	158
63	Max	190

3. Apply the same functions to calculate the desired characteristics for all numerical variables.

	A	B	C	D	E	F
1	GENDER	HEIGHT	WEIGHT	HEAD	SHOE_SIZE	MATH
54	W	162	70	55	40	5
55	W	172	58	62	39	4
56						
57	Number of observations	54				
58	Average	170.19444				
59	Median	170				
60	Standard deviation	7.3973584				
61	Standard error	1.006653				
62	Min	158				
63	Max	190				

4. Round the average, standard deviation and standard error values to one decimal place.

Result:

1	GENDER	HEIGHT	WEIGHT	HEAD	SHOE_SIZE	MATH
56						
57	Number of observations	54	54	54	54	54
58	Average	170.2	65.4	55.2	39.6	4.0
59	Median	170	63	56	39	4
60	Standard deviation	7.4	10.7	6.6	2.0	0.8
61	Standard error	1.0	1.5	0.9	0.3	0.1
62	Min	158	47.5	17	36	2
63	Max	190	90	75	46	5

5. **Write some sentences about the location and variability on studied variables.**

6. Add into data table new column (after column WEIGHT), name the new variable as 'BMI' (body mass index) and calculate it's values for all students applying the formula

$$\text{BMI} = \text{Weight, kg} / (\text{Height, m})^2.$$

	A	B	C	D
1	GENDER	HEIGHT	WEIGHT	BMI
2	W	170	70	=C2/((B2/100)^2)
3	W	158	47.5	

NB! Follow the number and position of bracket!

Do you understand this *Excel* formula?

Remarks.

- **Usually you can get the power sign ^ by key combination 'AltGr' + 'Ä'.**
- Alternative to find the square is just to multiply the value with itself: (B2/100)*(B2/100);
- another alternative is to use the power function: POWER(B2/100;2) – here the first argument is the base of a power and the second argument is exponent.

	A	B	C	D	H
1	GENDER	HEIGHT	WEIGHT	BMI	
2	W	170	70	24.221	
3	W	158	47.5		
4	W	170	60		
5	W	170	50		
6	W	179	68		
7	W	163	56		
8	W	177	65		
9	W	162.5	53		
10	W	170	75		
11	M	175	74		
12	W	176	66		
13	M	175	64		
14	M	190	82		
15	W	161	50		
16	W	170	85		
17	W	176	58		
18	W	172	90		
19	W	158	55		
20	M	189	82		
21	W	169	60		
22	W	164	52		
23	W	172	62		
24	W	173	66		
25	W	169	60		
26	W	162	50		
27	W	165	52		
28	M	170	80		
29	M	176	74		
30	M	175	73		
31	W	171	63		
32	W	170	60		
33	W	163	62		
34	M	181	74		
35	W	168	60		
36	W	174	54		
37	W	166	68		
38	W	168	63		
39	W	165	58		
40	W	171	75		
41	W	165	77		
42	W	161	55		
43	M	183	75		
44	W	169	53		
45	W	175	60		
46	W	167	80		
47	W	158	70		
48	M	174	87		
49	W	165	61		
50	W	164	58		
51	W	185	80		
52	W	177	63		
53	W	160	70		
54	W	162	70		
55	W	172	58		

7. Calculate all descriptive characteristics also for new variable.

Copy → Paste

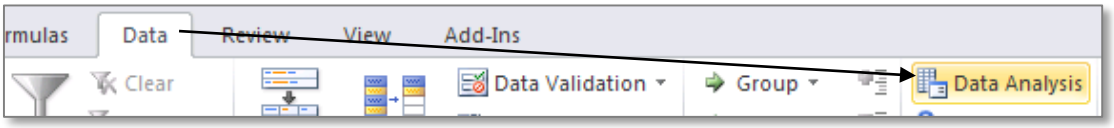
	A	B	C	D	E
1	GENDER	HEIGHT	WEIGHT	BMI	HEAD
56					
57	Number of observations	54	54	54	54
58	Average	170.2	65.4	22.6	55.2
59	Median	170	63	21.55482	56
60	Standard deviation	7.4	10.7	3.2	6.6
61	Standard error	1.0	1.5	0.4	0.9
62	Min	158	47.5	17.30104	17
63	Max	190	90	30.42185	75

Exercise 2.

- Apply the procedure *Descriptive Statistics* (*Data*-tab → *Data Analysis…*) to calculate descriptive statistics for numerical variables HEIGHT, WEIGHT, BMI, HEAD and SHOE_SIZE.
- Calculate also 90%, 95% or 99% confidence interval of the mean. What you can conclude based on confidence interval?

Guide

1. To calculate descriptive statistics: *Data*-tab → *Data Analysis…* → *Descriptive Statistics*



If study variables are located in columns side by side, all of them can be analysed together.

Selection 'Labels in first row' is necessary, if data are specified with variable names in the first row.

About additional options 'Summary statistics' etc look at the next page.

Location of the left upper corner of output table.

Explanation of additional options of procedure *Descriptive Statistics*:

- option '*Summary statistics*' asks *Excel* to calculate values of 12 basic characteristics;
- option '*Confidence Level for Mean: 95%*' asks *Excel* to calculate one half of the **confidence interval of mean** (this value must be added to and subtracted from the average value to get the confidence limits); instead of default **confidence level** 95% some other value can be typed (for example 90 or 99);
- options '*Kth Largest* = 1' and '*Kth Smallest* = 1' ask *Excel* to output the maximum and minimum value; as these values include already in the summary statistics table calculated according to the first option '*Summary statistics*', it is more meaningful to output the second largest and the second smallest values by specifying '*Kth Largest* = 2' and '*Kth Smallest* = 2'.

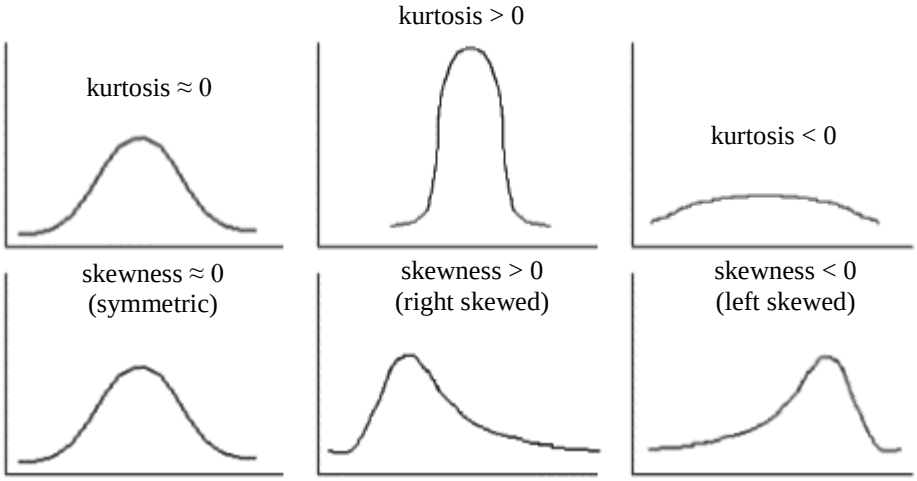
- Result:

Result of option *Summary statistics*'

HEIGHT		*WEIGHT*		*BMI*		*HEAD*		*SHOE_SIZE*	
Mean	170.194	Mean	65.4352	Mean	22.5569	Mean	55.2037	Mean	39.6481
Standard Error	1.00665	Standard Error	1.45092	Standard Error	0.43855	Standard Error	0.90204	Standard Error	0.27365
Median	170	Median	63	Median	21.5548	Median	56	Median	39
Mode	170	Mode	60	Mode	20.7612	Mode	55	Mode	39
Standard Deviation	7.39736	Standard Devia	10.6621	Standard Devi	3.2227	Standard Devia	6.62864	Standard Devi	2.01089
Sample Variance	54.7209	Sample Varian	113.68	Sample Variar	10.3858	Sample Varian	43.9389	Sample Variar	4.04368
Kurtosis	0.34724	Kurtosis	-0.67517	Kurtosis	-0.16936	Kurtosis	22.4574	Kurtosis	0.76174
Skewness	0.58075	Skewness	0.39892	Skewness	0.78119	Skewness	-3.43933	Skewness	0.77969
Range	32	Range	42.5	Range	13.1208	Range	58	Range	10
Minimum	158	Minimum	47.5	Minimum	17.301	Minimum	17	Minimum	36
Maximum	190	Maximum	90	Maximum	30.4218	Maximum	75	Maximum	46
Sum	9190.5	Sum	3533.5	Sum	1218.07	Sum	2981	Sum	2141
Count	54	Count	54	Count	54	Count	54	Count	54
Largest(2)	189	Largest(2)	87	Largest(2)	29.4118	Largest(2)	62	Largest(2)	44
Smallest(2)	158	Smallest(2)	50	Smallest(2)	17.8359	Smallest(2)	40	Smallest(2)	37
Confidence Level(95.0%)	2.01909	Confidence Lev	2.91018	Confidence Le	0.87963	Confidence Lev	1.80927	Confidence Lev	0.54887

- **Additional reading – description of the shape of the distribution**

  Most of the characteristics calculated by procedure *Descriptive Statistics* are introduced already earlier (and in the lecture).
  However, there are still two previously not described characteristics, which can be used to describe the shape of the distribution – *kurtosis* and *skewness*. The nature of these characteristics is illustrated with the next figure.



  It is worth to talk about the remarkable difference from the normal distribution only if whichever of these characteristics has value over 1 or under -1.

  However, these characteristics are not used too often.

  At the present moment the kurtosis of head girth (1.94) is slightly higher than values of the same characteristic for other studied body measurements – this is implying that most of the head circumference values are located in quite narrow range but at the same time there exist few much smaller and/or much bigger values.

- **To decide about the symmetry of distribution, often the comparison of mean and median values is used (instead of calculation of skewness)**.
  Namely, as the mean (average) is sensitive to the unusual values (outliers), then

  $\bar{x} > med$ refer to the positively (right) skewed distribution (there exist few much bigger values, and so the skewness value > 0),

  $\bar{x} < med$ refer to the negatively (left) skewed distribution (there exist few much smaller values, and so the skewness value < 0).

- **Look at the calculated values of mean, median and skewness – does these described relations apply also for students body measurements?**

2. Calculate 90%, 95% **or** 99% confidence limits for mean. What is the meaning of these values?

As *Excel* does not calculate the confidence limits, it must be done by user based for example on the output of procedure *Descriptive Statistics*. Just add to the output table of procedure *Descriptive Statistics* two rows – one for the **lower 95% confidence limit** and second for the **upper 95% confidence limit**.

	R	S
1	*HEIGHT*	
2		
3	Mean	170.19
4	Standard Error	1.0067
5	Median	170
6	Mode	170
7	Standard Deviation	7.3974
8	Sample Variance	54.721
9	Kurtosis	0.3472
10	Skewness	0.5808
11	Range	32
12	Minimum	158
13	Maximum	190
14	Sum	9190.5
15	Count	54
16	Largest(2)	189
17	Smallest(2)	158
18	Confidence Level(95.0%)	2.0191
19		
20	Alumine 95% usalduspiir	=S3-S18
21	Ülemine 95% usalduspiir	=S3+S18

The confidence limits of the mean are calculated by the formula

$$\bar{x} \pm t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

As *Excel* outputs both parts of this equation, it is simple to calculate desired confidence limits.

Lower 95% confidence limit	168.18
Upper 95% confidence limit	172.21

So, considering the analysed dataset as a sample from the first year students' population, it can be concluded that the **average height** of students is with 95% probability in interval from 168.2 cm to 172.2 cm. This means, that measuring the height of all first year students (the whole population) and calculating the actual population mean, this value should stay between calculated limits with 95% probability.

- If somebody calculated 90% or 99% confidence limits, then these should be (168.5, 171.9) and (167.5, 172.9). **Why is the 90% confidence interval narrower?**

- **Calculate the confidence intervals also for other variables and try to formulate the conclusion for at least one of them!!**