Biometry practical 2

Illustrated (imperfect) practical guide

Preparatory work

1. Open in *MS Excel* the questionary data (file analysed already in previous practical),
2. rename 'Sheet3' to 'Praks2' (or 'Practical 2') and
3. make a copy of the data table (from worksheet 'Andmed') and paste it into the upper left corner of the worksheet 'Praks2'.

Exercise 1.

Construct the frequency table to variable 'HEAD' and illustrate it with histogram. Format the histogram.

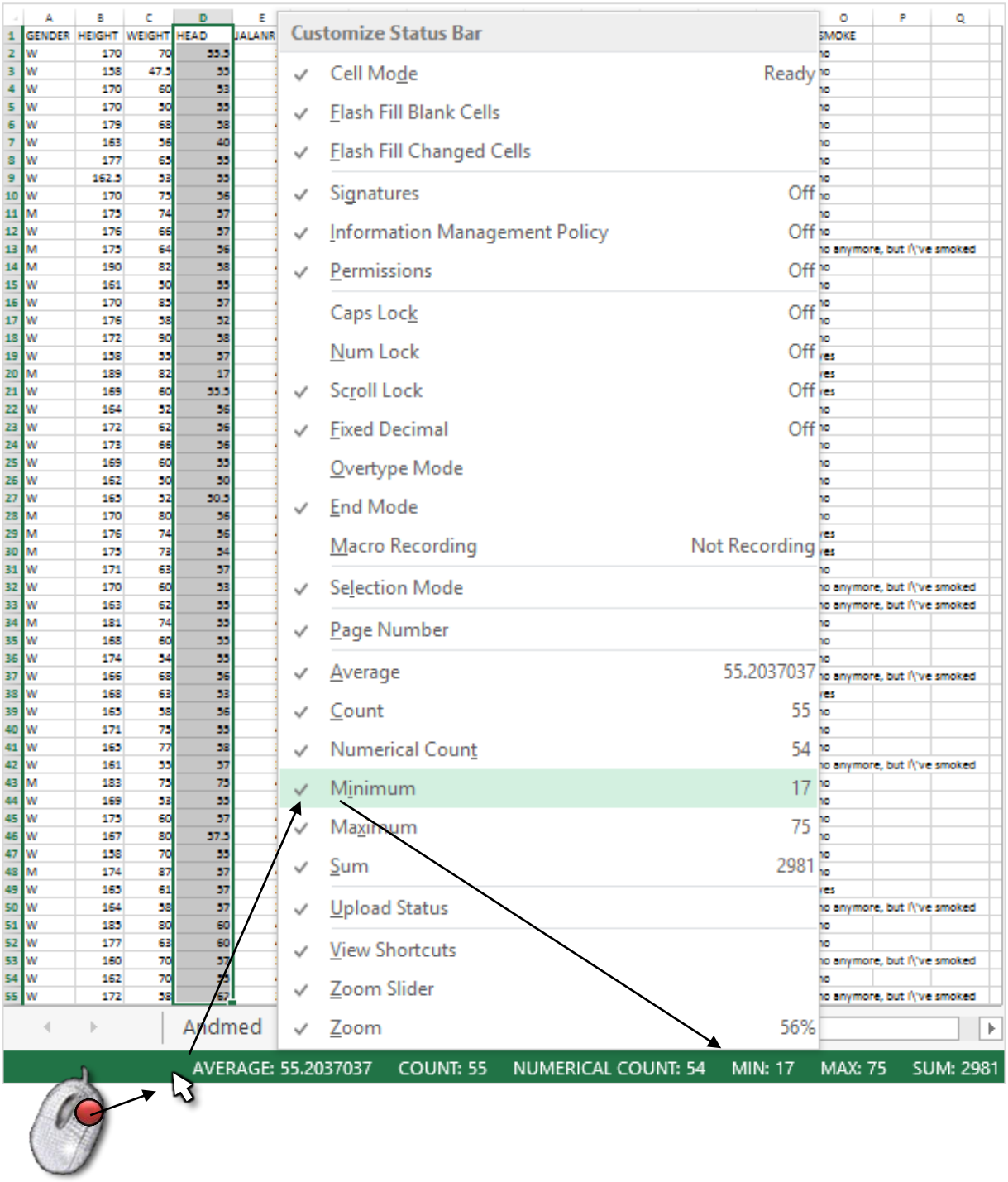
Guide

To construct the frequency table to continuous variable there are at least tree possibilities in Excel – statistical procedure *Histogram*, function *FREQUENCY* and *PivotTable*. The first two assume that there is also formed the table containing upper limits of continuous variable values' classes (intervals)

(actually can procedure *Histogram* form these intervals by itself, but mostly are the limits of these automatically generated intervals not rounded and therefore the description of values' distribution is more inconvenient).

Following it is introduced, how to construct the frequency table with procedure *Histogram*.

1. To decide, how many and how big intervals to form, there must be knowledge about the count and range of values of studied variable.
   - The count $n = 54$ (Did you get the same number? How?) gives the hint, that the values of students head circuits should be divided into 7 or 8 intervals (because $\sqrt{54} \approx 7{,}3$; at the same time this is not absolute truth, if it is more suitable, there can be formed also 6 or 9 intervals).
   - The range of values is given by minimum and maximum value.
     To find these characteristics there are several ways in Excel. The most quick variant is (look at the scheme in next page also):
     a. select the cells containing values of study variable,
     b. click with mouse right key on Status Bar under the worksheet and select wanted functions (for example *Min*),
     c. the values of selected functions are displayed on the Status Bar.



- Maximal head girth is 75 and minimal 17 cm. Are these values real? If it is hard to imagine the girth, then maybe it is easier to imagine the diameter. But how to calculate the diameter? How big is the head diameter, if head girth is 17 cm? And is this number really possible?

  Hint. The girth is calculated as $P = 2\pi r$, from which the diameter is $d = 2r = P/\pi$ (to calculate the diameter corresponding to girth 17 cm in Excel you can use the formula '=17/PI()'). The result is 5.4 cm. To compare – the diameter of CD is 12 cm and the diameter of usual coffee cup is about 7 cm. So, the head girth 17 cm and corresponding head diameter 5.4 cm are clearly too small. Additionally there is one head girth value 40 cm in data table, which is also too small (corresponding head diameter is 12.7 cm).
  And the maximal value 75 cm is too big for head girth.

  Obviously are values 17, 40 and 75 cm measuring or typing errors.

As **traditional statistical analyses methods are not derived to analyse unreal or very exceptional data**, it is usual to omit these very different values. At the present situation it is the most reasonable to just delete the head girth vales 17, 40 and 75 cm.

And to avoid future confusion these values should be deleted in both worksheets 'Praks2' ('Practical2') and 'Andmed' ('Data').

- Minimal head girth after deleting unreal values is 50 and maximal 62 cm.
- To find the approximate width of intervals the difference between maximal and minimal value must be divided by the desired number of intervals: (62-50)/7= 1.7… cm.

  To get round interval limits it is necessary to slightly increase or decrease the calculated value, also the number of intervals can be changed. At the present situation we can fix the width of intervals to **2 cm**.
- To fix the intervals' limits it is natural to start with the first interval and fix it in way that it contains the minimal value. At the present situation the first interval can be for example 48-50 cm.
- Altogether we can form **seven 2-centimeter** intervals of students head girth:
  48-50; 50-52; 52-54; 54-56; 56-58; 58-60; 60-62.

  As Excel puts the values equal to interval limit into the lower interval, it is more correct to present the intervals as half-closed intervals (because how to known otherwise into witch interval belongs the value 41 cm, for example):

  (48,50], (50,52], (52,54], (54,56], (56,58], (58,60] and (60,62].

NB. But instead these intervals somebody can form also **nine 1.5-centimeter** intervals of the form

(49;50.5], (50.5;52], (52;53.5], (53.5;55], (55;56.5], (56.5;58], (58;59.5], (59.5;61], (61;62.5];

or **seven 2-centimeter** intervals of the form

(49,51], (51,53], (53,55], (55,57], (57,59], (59,61], (61,63].

- To give these intervals to Excel there must be constructed a special table containing the upper limits of intervals:

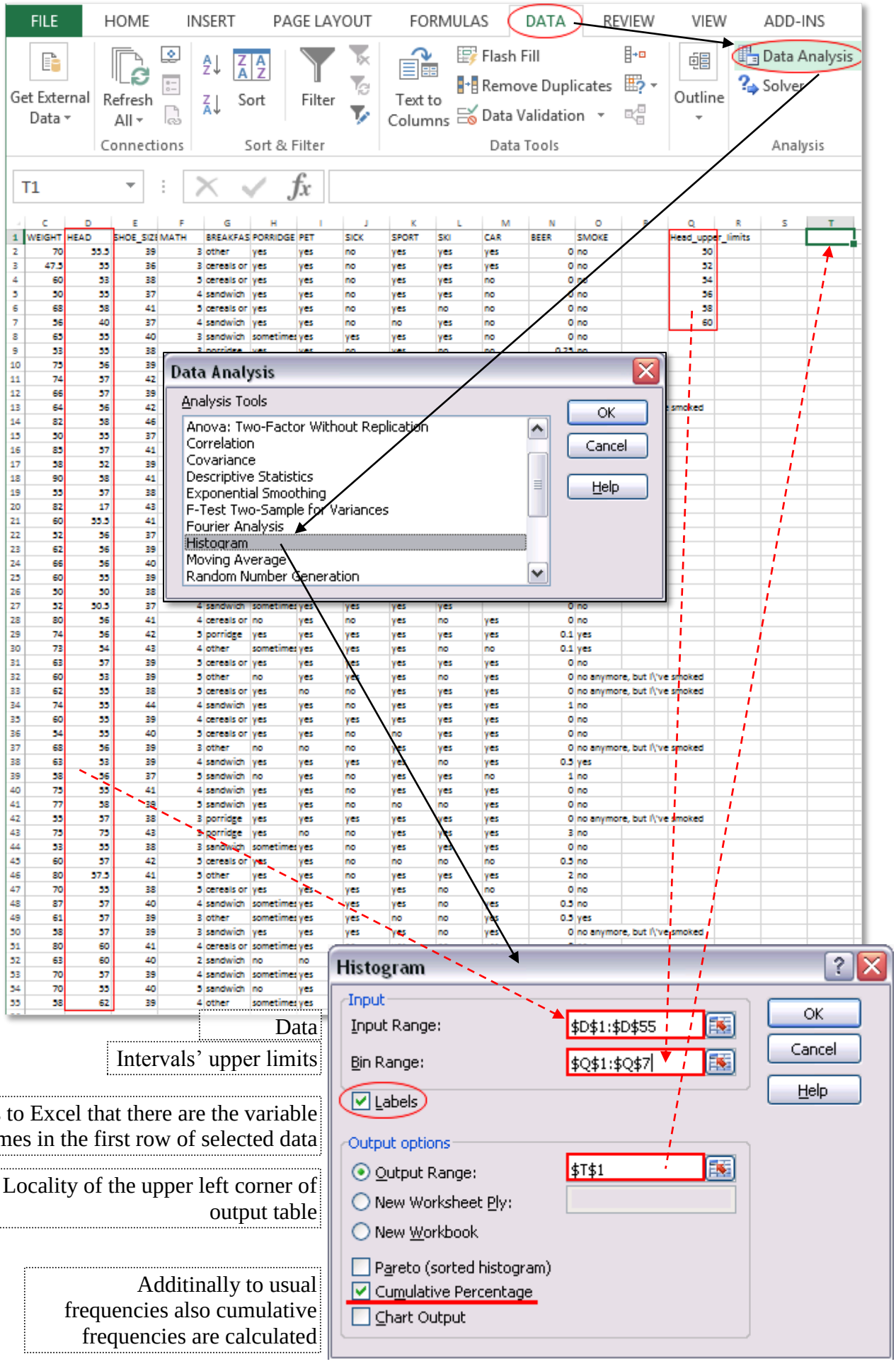
O	P	Q
SMOKE		Head_upper_limits
no		50
no		52
no		54
no		56
no		58
no		60
no		

- Remarks:
  – between initial data table and any additional tables should be at least one empty column (or row), why?;

  – Excel understood the given values as follows:
  '51' means '≤51';
  '53' means '51<x≤53'; …

  – there is **no need to specify the last interval '62'=(60,62]**, because constructing the frequency table Excel automatically makes one class for all values not belonging into any given intervals. As all head girth values less or equal to 60 belong into specified intervals, the additionally made class will contain only values bigger than 60.

2. Frequency table with procedure *Histogram*: *Data*-tab → *Data Analysis…*→ *Histogram*



Data

Intervals' upper limits

Says to Excel that there are the variable names in the first row of selected data

Locality of the upper left corner of output table

Additionally to usual frequencies also cumulative frequencies are calculated

Expected result of procedure *Histogram*:

(**cumulative frequency** shows the percentage,
of values less or equal to the upper limit of interval)

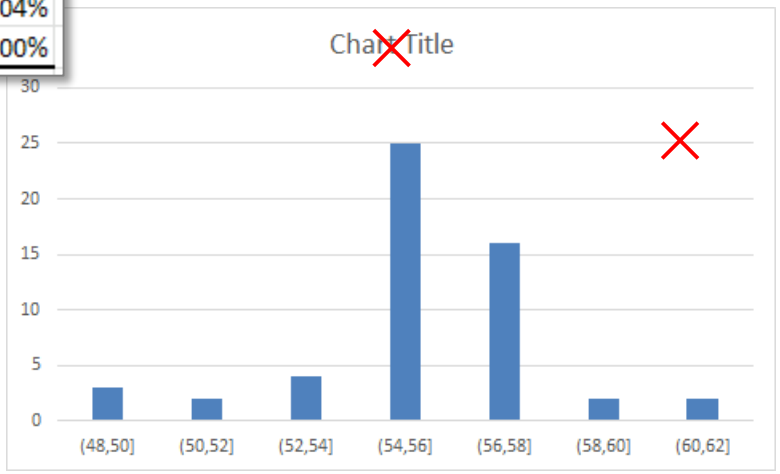
	T	U	V
	Head_upper_lim	*Frequency*	*Cumulative %*
	50	1	1.96%
	52	2	5.88%
	54	4	13.73%
	56	25	62.75%
	58	16	94.12%
	60	2	98.04%
	More	1	100.00%

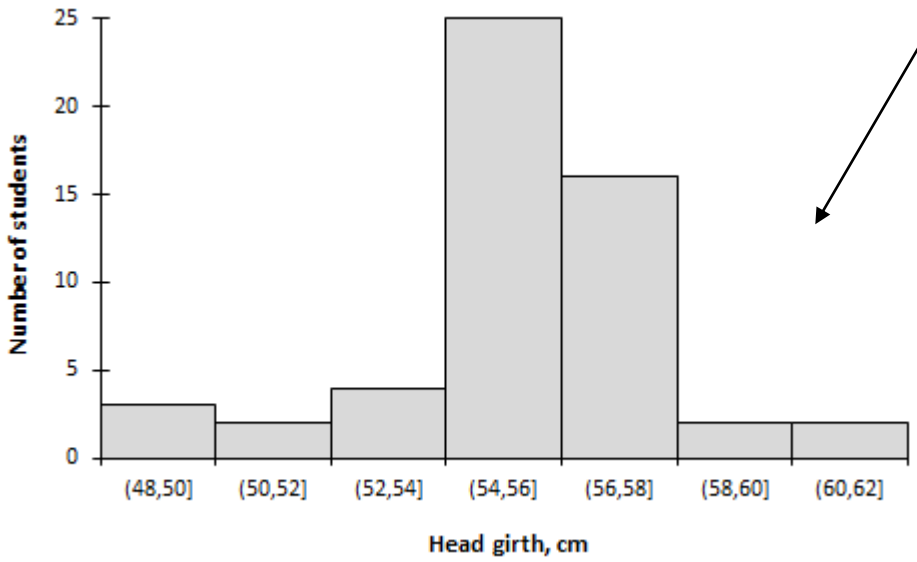
3. Add in front of the table real interval limits in the form of half-closed intervals.

	Head_upper_lim	*Frequency*	*Cumulative %*
(48,50]	50	1	1.96%
(50,52]	52	2	5.88%
(52,54]	54	4	13.73%
(54,56]	56	25	62.75%
(56,58]	58	16	94.12%
(58,60]	60	2	98.04%
(60,62]	More	1	100.00%

NB! To select separate cells hold down the 'Ctrl'-key …

- Delete title and gridlines
- Remove the border around the chart area
- Add axes and fix the range of y-axis from 0 to 25 with step 5
- Add titles to x- and y-axis
- Take gap between bars = 0
- Format axis units and names with font size 10 and black color
- Fill the bars with favorite color and add darker countour to bars





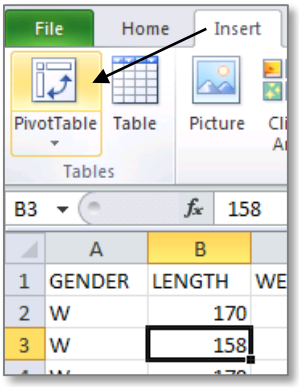
4. **Write down at least one sentence using absolute frequencies and one sentence using cumulative frequencies**.

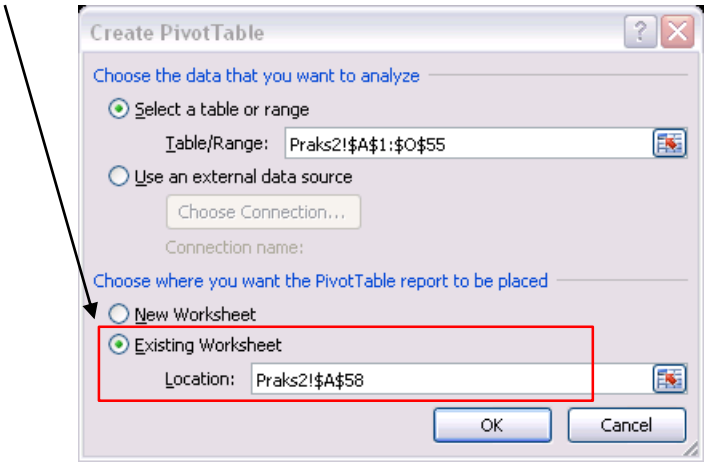
Exercise 2.

Use the *PivotTable* to describe the students' weight depending on the porridge eating.

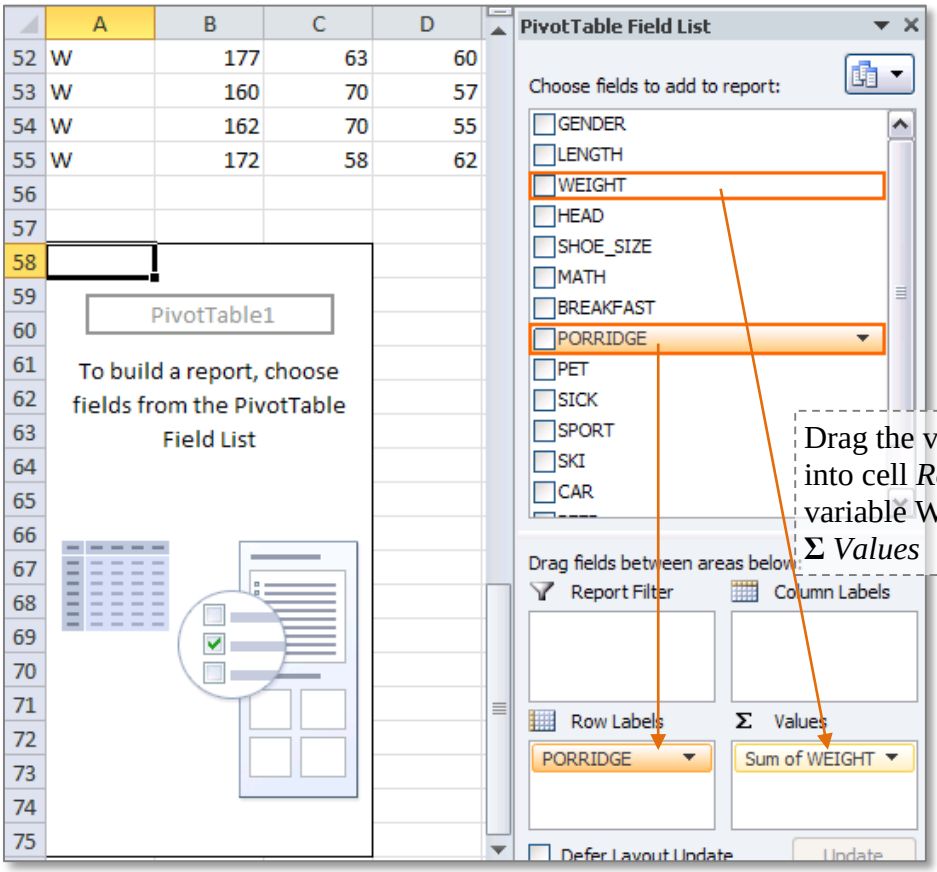
Guide

1. Put the cursor into data table (in worksheet 'Praks2').
2. *Insert*-tab → *PivotTable*
3. Place the output table under the data table:





4.

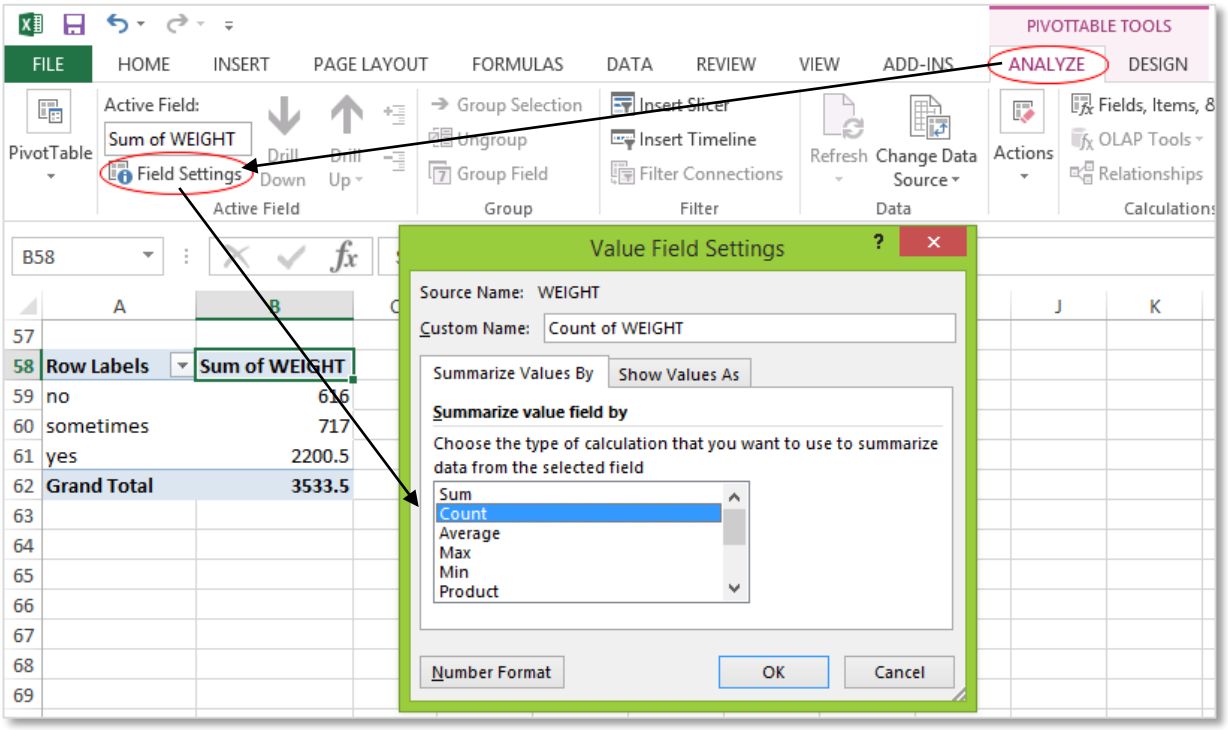


Drag the variable PORRIDGE into cell *Row Labels* and variable WEIGHT into cell **Σ** *Values*

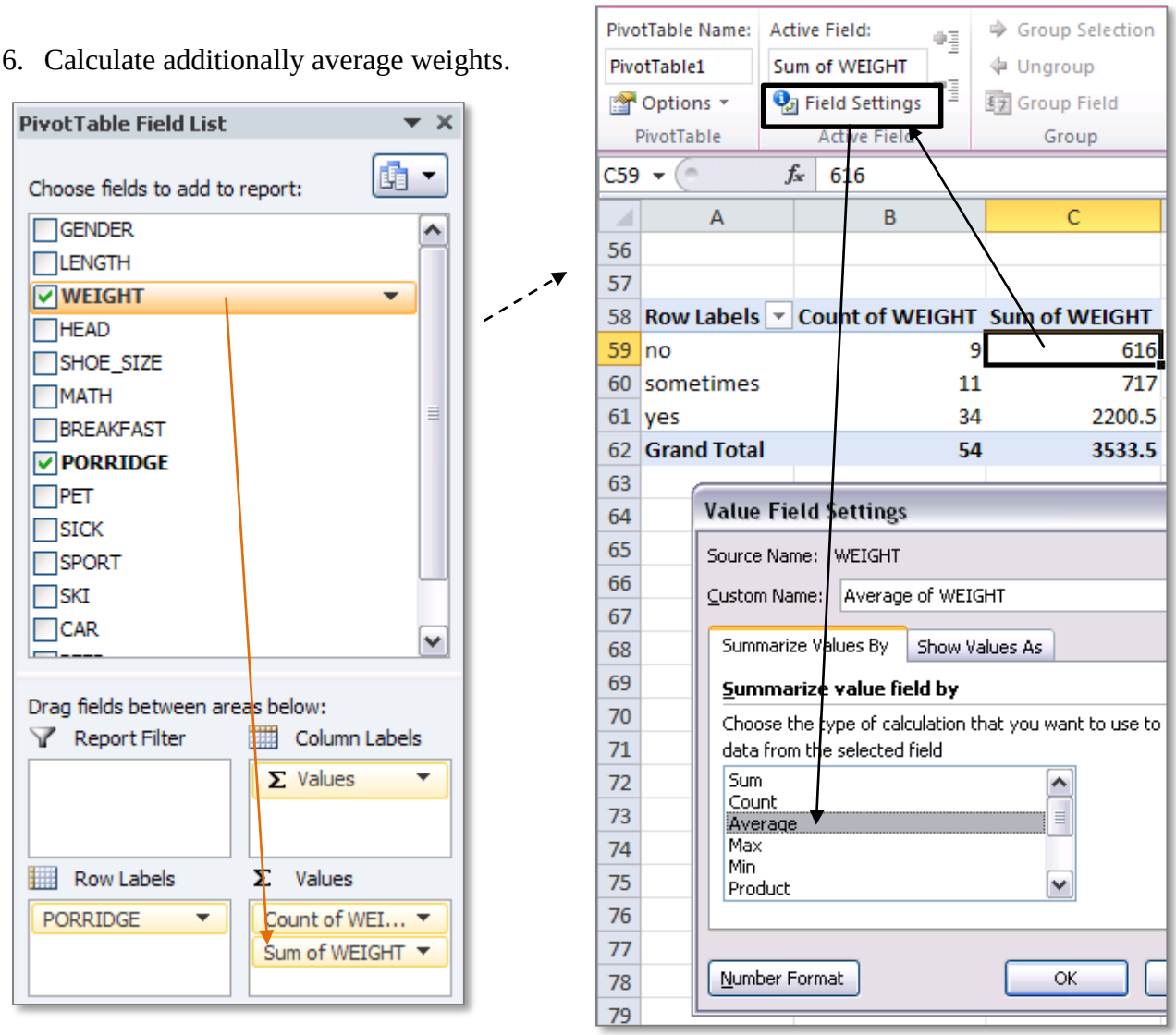
Result:

Row Labels	Sum of WEIGHT
no	616
sometimes	717
yes	2200.5
Grand Total	**3533.5**

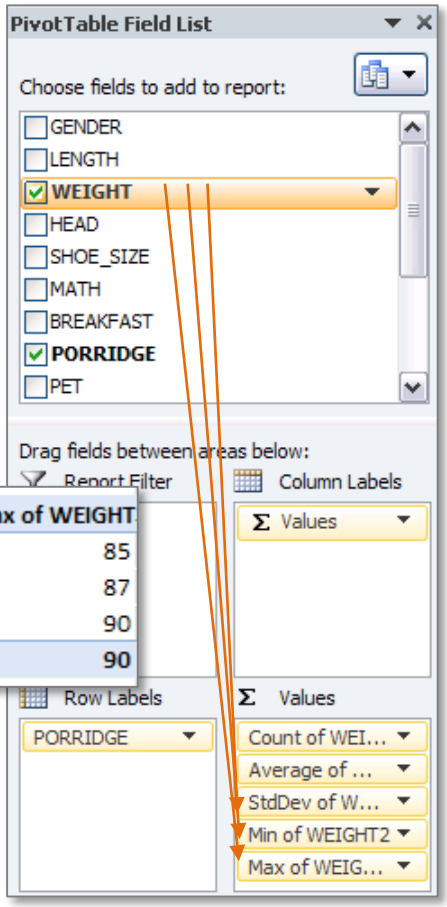
5. Instead of sum of weights (calculated by Excel by default), calculate the **count of students** in different groups.



6. Calculate additionally average weights.



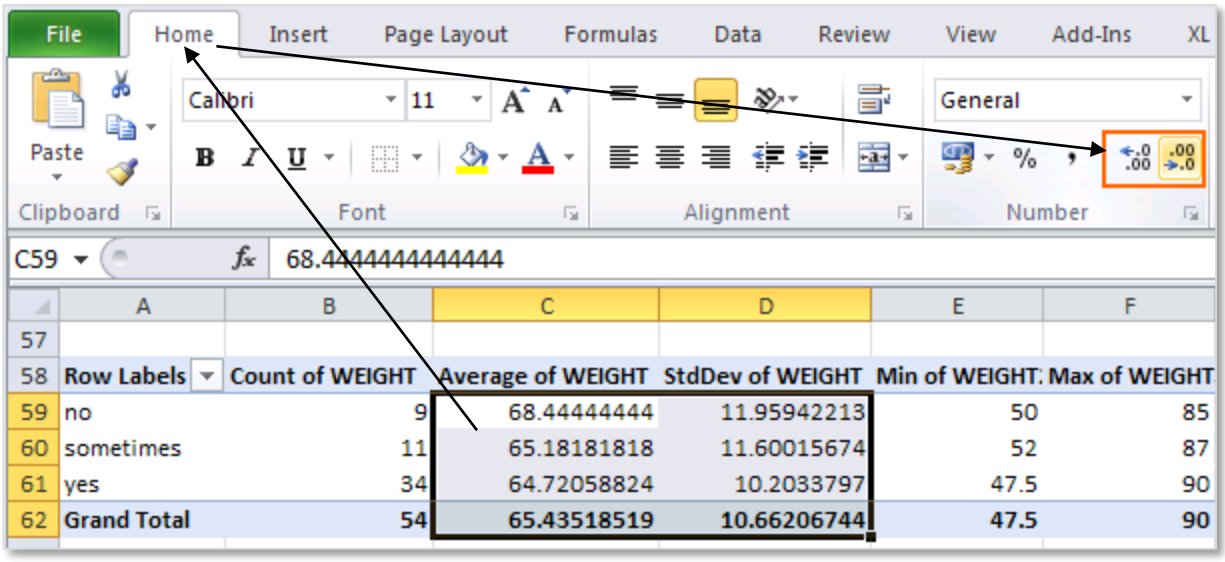
7. Calculate also standard deviations, minimal and maximal values of weights depending on porridge eating.



Result:

Row Labels	Count of WEIGHT	Average of WEIGHT	StdDev of WEIGHT	Min of WEIGHT	Max of WEIGHT
no	9	68.44444444	11.95942213	50	85
sometimes	11	65.18181818	11.60015674	52	87
yes	34	64.72058824	10.2033797	47.5	90
Grand Total	**54**	**65.43518519**	**10.66206744**	**47.5**	**90**

8. Round the averages and standard deviations to one decimal place (using corresponding command at *Home*-tab, for example).



	A	B	C	D	E	F
57						
58	**Row Labels**	**Count of WEIGHT**	**Average of WEIGHT**	**StdDev of WEIGHT**	**Min of WEIGHT**	**Max of WEIGHT**
59	no	9	68.44444444	11.95942213	50	85
60	sometimes	11	65.18181818	11.60015674	52	87
61	yes	34	64.72058824	10.2033797	47.5	90
62	**Grand Total**	**54**	**65.43518519**	**10.66206744**	**47.5**	**90**

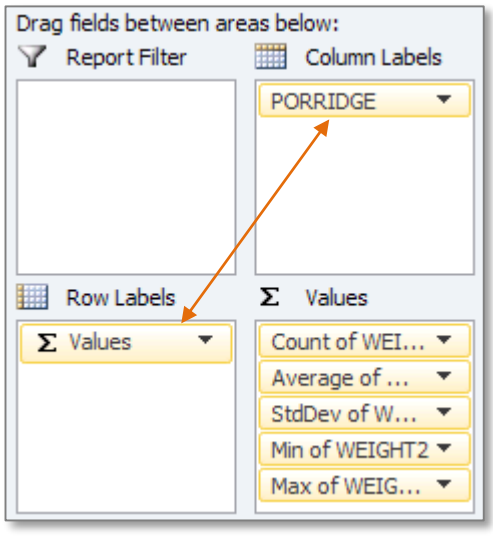
Result:

Row Labels	Count of WEIGHT	Average of WEIGHT	StdDev of WEIGHT	Min of WEIGHT	Max of WEIGHT
no	9	68.4	12.0	50	85
sometimes	11	65.2	11.6	52	87
yes	34	64.7	10.2	47.5	90
Grand Total	**54**	**65.4**	**10.7**	**47.5**	**90**

9. Present the table in form where different characteristics are in different rows and for each porridge-eating group corresponds own column:

	Column L			
Values	**no**	**sometimes**	**yes**	**Grand Total**
Count of WEIGHT	9	11	34	54
Average of WEIGHT	68.4	65.2	64.7	65.4
StdDev of WEIGHT	12.0	11.6	10.2	10.7
Min of WEIGHT2	50	52	47.5	47.5
Max of WEIGHT3	85	87	90	90

Hint:

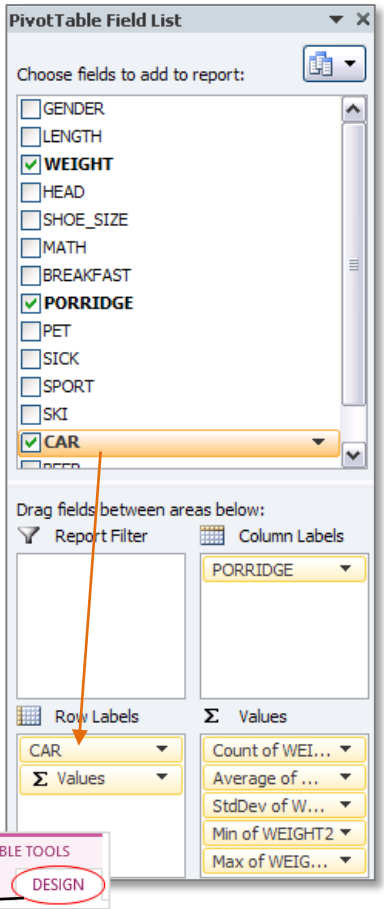


10. And now divide the table into rows according to the variable 'CAR' values and omit one student who did not know does she or he have a car or not.

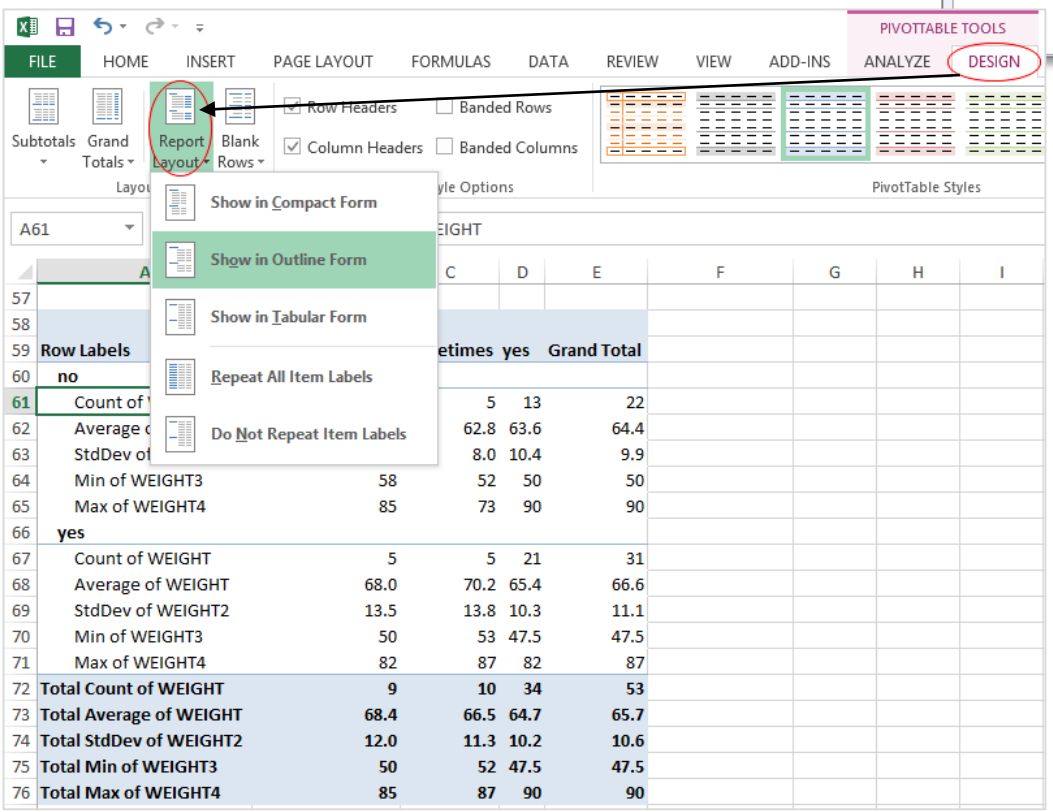
Expected result:

Row Labels	Column: no	sometimes	yes	Grand Total
no				
Count of WEIGHT	4	5	13	22
Average of WEIGHT	69.0	62.8	63.6	64.4
StdDev of WEIGHT	11.7	8.0	10.4	9.9
Min of WEIGHT2	58	52	50	50
Max of WEIGHT3	85	73	90	90
yes				
Count of WEIGHT	5	5	21	31
Average of WEIGHT	68.0	70.2	65.4	66.6
StdDev of WEIGHT	13.5	13.8	10.3	11.1
Min of WEIGHT2	50	53	47.5	47.5
Max of WEIGHT3	82	87	82	87
Total Count of WEIGHT	**9**	**10**	**34**	**53**
Total Average of WEIGHT	**68.4**	**66.5**	**64.7**	**65.7**
Total StdDev of WEIGHT	**12.0**	**11.3**	**10.2**	**10.6**
Total Min of WEIGHT2	**50**	**52**	**47.5**	**47.5**
Total Max of WEIGHT3	**85**	**87**	**90**	**90**

Hint:



11. Try different *PivotTable* layouts (sometimes helps some layout better understand the structure of the table):



12. Can you say something about each number in the last table?

Write down some sentences about similarities or differences of average body weights of students

- **depending on car owning,**
- **depending on porridge eating.**