

## Biometry practical 4

### Illustrated (imperfect) practical guide

#### Preparatory work

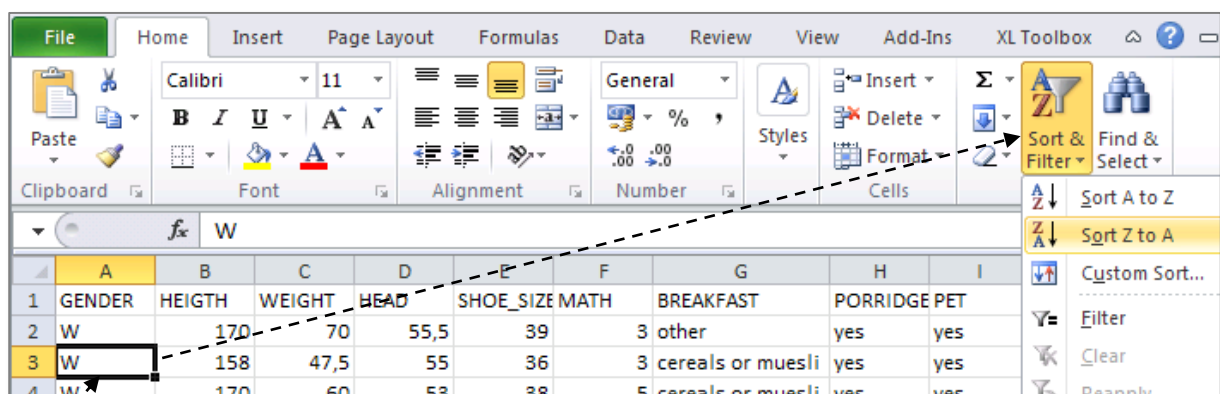
1. Open in *MS Excel* the questionnaire data (file analysed already in previous practical),
2. insert new worksheet, rename new worksheet to 'Praks4' (or 'Practical4') and
3. make a copy of the data table (from worksheet 'Andmed') and paste it into the upper left corner of the new worksheet.

#### Exercise 1.

Suppose that the first year students of Institute of Veterinary Medicine and Animal Sciences (students in our dataset) are just a random sample from all first year students of Estonian University of Life Sciences. Knowing, that the average height of Estonian women is 169 cm, test the hypothesis: is the average height of first year female students of Estonian University of Life Sciences different from Estonian average 169 cm?

#### Guide

1. Sort the datatable by column 'GENDER'.



- 1) For example, by putting the cursor into column 'GENDER' and selecting from *Home*-tab commands

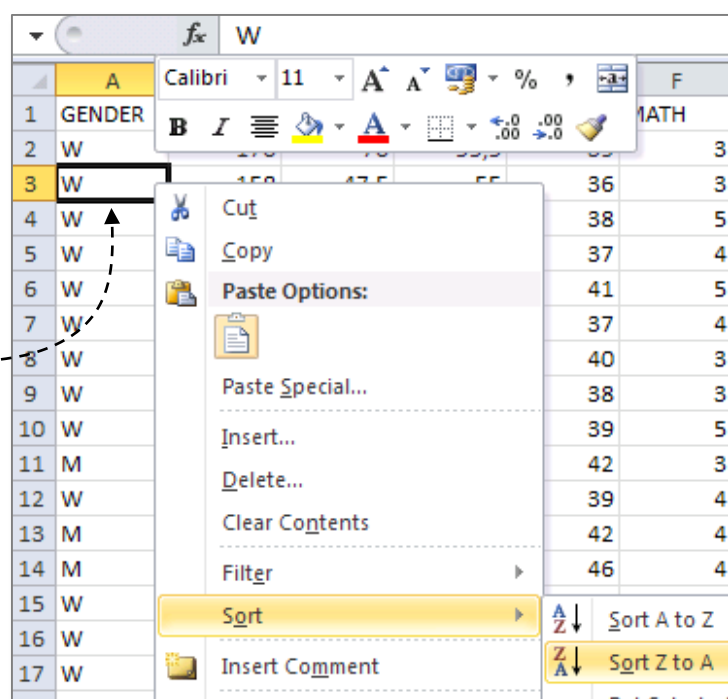
*Sort & Filter* ->

*Sort Z to A*

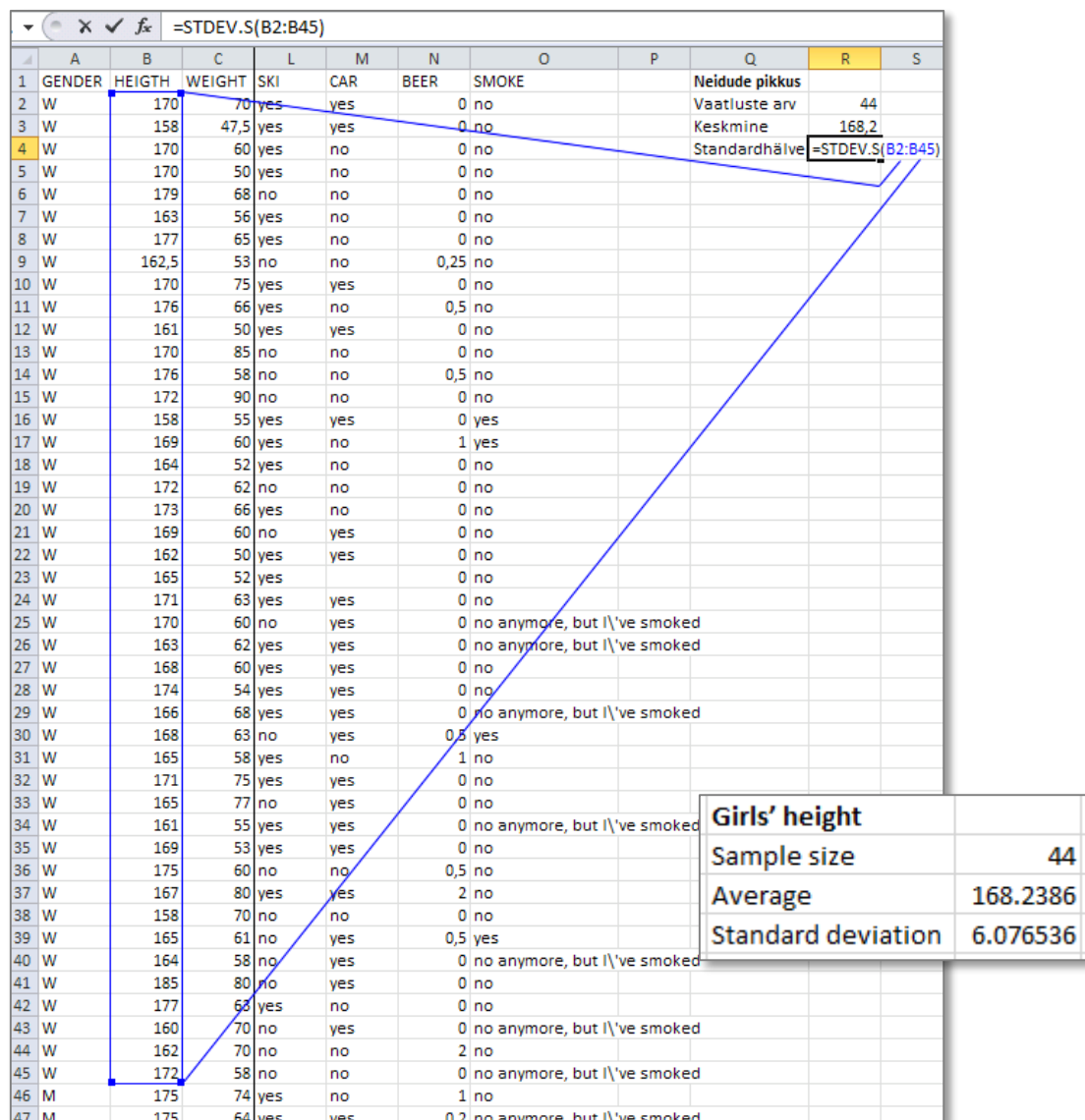
(to get women before men).

- 2) Or clicking with right mouse button on an arbitrary cell in column 'GENDER' and selecting from the drop down menu commands

*Sort* -> *Sort Z to A*.



2. Find the number of girls and their average height and standard deviation of height using functions COUNT, AVERAGE and STDEV.S.



So, there are 44 girls with average height 168.2 cm and standard deviation 6.1 cm;

- this says, that the average difference of girls' height from 168.2 cm is 6.1 cm; or, expecting that the height follows normal distribution, then according to the properties of normal distribution

\*) approximately 68.3% of the heights of first year female students is in interval  $168.2 \pm 6.1$  cm ( $\bar{x} \pm s$ ) and

\*) approximately 95.5% of the heights of first year female students is in interval  $168.2 \pm 12.2$  cm ( $\bar{x} \pm 2s$ ).

3. Formulate the hypothesis pair and write it down.

For example:

Girls' height	
Sample size	44
Average	168.2386
Standard deviation	6.076536
$H_0$ : the heights of first year female students correspond to Estonian standard (169 cm) $H_1$ : the heights of first year female students does not correspond to Estonian standard (169 cm)	
$H_0$ : the average height of first year female students does not differ from 169 cm $H_1$ : the average height of first year female students differs from 169 cm	
$H_0$ : $\mu_F = 169$ $H_1$ : $\mu_F \neq 170$	$\mu_F$ - average height of first year female students

**Remainder from theory – relationships between hypothesis testing and confidence intervals**

- If the task of hypothesis testing is to compare some estimated parameter with constant value, the decision is often made based on the confidence interval of the studied parameter:
  - if the **constant is between** confidence limits, **then there is no reason to reject null hypothesis** and it can be concluded **that the studied parameter does not differ from given constant**;
  - if constant is outside confidence interval, then the **studied parameter is differ from given constant**.
- For example, if you want to compare the average value with given constant (does the data correspond to standard), the hypothesis pair is:

$$H_0: \mu = c \text{ and } H_1: \mu \neq c.$$

if  $c \in [\underline{\mu}, \bar{\mu}]$ , then  $H_0: \mu = c$  is true;      if  $c \notin [\underline{\mu}, \bar{\mu}]$ , then  $H_1: \mu \neq c$  is true.



4. Calculate the half of the 95% confidence interval using functions CONFIDENCE.NORM and CONFIDENCE.T:

**a) function CONFIDENCE.NORM**

(this function has 3 arguments: significance level  $\alpha$ , standard deviation and number of female students in dataset);

You can just type this text (to be clear later, which function was used).  
And after that put the cursor into result cell!

**Insert Function**  
 Search for a function:  
 Type a brief description of what you want to do and then click Go  
 Or select a category: Statistical  
 Select a function:  
 CHISQ.INV.RT  
 CHISQ.TEST  
CONFIDENCE.NORM  
 CONFIDENCE.T  
 CORREL  
 COUNT  
 COUNTA  
**CONFIDENCE.NORM(alpha;standard\_dev;size)**  
 Returns the confidence interval for a population mean, using a normal distribution.

The screenshot shows an Excel spreadsheet with the following data:

	P	Q	R	S	T	U	V	W	X	Y	Z
1		Girls' height									
2		Sample size	44								
3		Average	168.2386								
4		Standard deviation	6.076536								
7		H <sub>0</sub> : the heights of first year fem									
8		H <sub>1</sub> : the heights of first year fem									
10		H <sub>0</sub> : the average height of first y									
11		H <sub>1</sub> : the average height of first y									
13		H <sub>0</sub> : μ <sub>f</sub> = 169	μ <sub>f</sub> - average								
14		H <sub>1</sub> : μ <sub>f</sub> ≠ 170									
17		Function CONFIDENCE.NORM									
18			=CONFIDENCE.NORM(0.05;R4;R2)								

The dialog box 'Function Arguments' for CONFIDENCE.NORM shows:

- Alpha: 0.05
- Standard\_dev: R4
- Size: R2

Formula result = 1.795468576

Help on this function

OK

If we want to calculate 95% confidence interval, then the significance level  $\alpha = 0.05$ .

Function CONFIDENCE.NORM

1.795469

**b) function CONFIDENCE.T**

(arguments of this function are similar to the function CONFIDENCE.NORM).

Result:

Function CONFIDENCE.NORM	
	1.795469
Function CONFIDENCE.T	
	1.847436

5. Calculate the lower and upper confidence limits based on both functions.

	P	Q	R	S	T	U	V	W
1		Girls' height						
2		Sample size	44					
3		Average	168.2386					
4		Standard deviation	6.076536					
5								
6								
7		H <sub>0</sub> : the heights of first year female students correspond to Estonian standard (169 cm)						
8		H <sub>1</sub> : the heights of first year female students does not correspond to Estonian standard						
9								
10		H <sub>0</sub> : the average height of first year female students does not differ from 169 cm						
11		H <sub>1</sub> : the average height of first year female students differs from 169 cm						
12								
13		H <sub>0</sub> : μ <sub>F</sub> = 169	μ <sub>F</sub> - average height of first year female students					
14		H <sub>1</sub> : μ <sub>F</sub> ≠ 170						
15								
16								
17		Function CONFIDENCE.NORM			Lower confidence limit	166.4432 = R3-R18		
18			1.795469		Upper confidence limit	170.0341		
19								
20		Function CONFIDENCE.T			Lower confidence limit	166.3912 = R3-R21		
21			1.847436		Upper confidence limit	170.0861		

Which of these 95% confidence intervals is wider? Why?

Answer.

The confidence interval got with function CONFIDENCE.T is slightly wider.

The reason is, that function CONFIDENCE.T calculates the confidence limits based on t-distribution following the formula  $\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}$ ,

but function CONFIDENCE.NORM calculates **asymptotic** (approximate) confidence interval (whereby the accuracy is increasing when the sample size is increasing) based on normal distribution:  $\bar{x} \pm z_{1-\alpha/2} \frac{s}{\sqrt{n}}$ , got range is little under estimated in case of small data set.

Parameters  $t_{1-\alpha/2, n-1} = t_{0,975;43} = 2,017$  and  $z_{1-\alpha/2} = z_{0,975} = 1,96$  are 97.5%-points (values, from which the bigger values can occur only with probability 0.025) of t-distribution and standard normal distribution, respectively, the first of these quantities is calculable in Excel 2010 with function =T.INV(0,975;43) and the second with function =NORM.S.INV(0,975).

**NB!** In the older *Excel* versions there is no functions CONFIDENCE.T and CONFIDENCE.NORM. There is only function CONFIDENCE, which is equivalent to functions CONFIDENCE.NORM, to calculate the confidence interval based on t-distribution the corresponding option *Confidence Level for Mean* of procedure *Descriptive Statistics* can be used.

- 6. Make a final conclusion and write it down – does the average height of first year female students differ from Estonian average (169 cm)? The answer must contain also the argumentation, why you made this decision.**

Example. As the average height of Estonian women (169 cm) is between 95% confidence limits of the first year female students average height:  $169 \in (166,4; 170,1)$ , then there is no reason to reject nullhypothesis  $H_0$ : the average heights of first year female students does not differ from 169 cm.

- 7. Supplementary task.**

**The average height of women over the world is 154 cm. Can you conclude that the average height of first year female students in Estonian University of Life Sciences differs from the world average?**

NB! You don't need to make any additional calculations. The decision can be made just based on already calculated confidence interval.

## Exercise 2.

Are weights of students owning and not owning a car different?

### Guide

1. Make an additional table containing only columns 'WEIGHT' and 'CAR' and sort it by column 'CAR'.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	HEIGHT	WEIGHT	HEAD	SHOE_SIZ	MATH	BREAKFA	PORRIDGE	PET	SICK	SPORT	SKI	CAR	BEER	SMOKE	Girls' height									WEIGHT	CAR
2	170	70	35.5	39	3	other	yes	yes	no	yes	yes	yes	0	no	Sample size	44								70	yes
3	158	47.5	35	36	3	cereals	or	yes	yes	no	yes	yes	0	no	Average	168.239								47.5	yes
4	170	60	33	38	3	cereals	or	yes	yes	no	yes	yes	0	no	Standard deviation	6.07634								60	no
5	170	50	33	37	4	sandwich	yes	yes	no	yes	yes	no	0	no										50	no
6	179	68	38	41	3	cereals	or	yes	yes	no	yes	no	0	no										68	no
7	169	56		37	4	sandwich	yes	yes	no	no	yes	no	0	no										56	no
8	177	65	35	40	3	sandwich	sometime	yes	yes	yes	yes	no	0	no										65	no
9	162.5	33	35	38	3	porridge	yes	yes	no	yes	no	no	0.25	no										33	no
10	170	75	36	39	3	other	yes	yes	no	no	yes	yes	0	no										75	yes
11	176	66	37	39	4	sandwich	sometime	no	no	yes	yes	no	0.3	no										66	no
12	161	30	35	37	4	nothing	no	yes	yes	yes	yes	yes	0	no										30	yes
13	170	85	37	41	4	cereals	or	no	yes	no	no	no	0	no										85	no
14	176	38	32	39	3	cereals	or	yes	yes	no	no	no	0.5	no										38	no
15	172	90	38	41	4	porridge	yes	yes	no	no	no	no	0	no										90	no
16	158	55	37	38	4	cereals	or	yes	yes	yes	no	yes	0	yes										55	yes
17	169	60	35.5	41	4	cereals	or	yes	yes	no	yes	no	1	yes										60	no
18	164	51	36	37	4	other	sometime	no	no	yes	yes	no	0	no										51	no
19	172	62	36	39	4	sandwich	yes	yes	no	no	no	no	0.5	no										62	no
20	173	66	36	40	3	cereals	or	yes	yes	no	yes	yes	0	no										66	no
21	169	60	35	39	3	other	yes	yes	no	yes	no	no	0	no										60	yes
22	162	30	30	38	3	porridge	yes	yes	yes	no	yes	yes	0	no										30	yes
23	165	32	30.5	37	4	sandwich	sometime	yes	yes	yes	yes	yes	0	no										32	no
24	171	63	37	39	3	cereals	or	yes	yes	yes	yes	yes	0	no										63	yes
25	170	60	33	39	3	other	no	yes	yes	yes	no	yes	0	no										60	yes
26	163	62	35	38	3	cereals	or	yes	no	yes	yes	yes	0	no										62	yes
27	168	60	35	39	4	cereals	or	yes	yes	yes	yes	yes	0	no										60	yes
28	174	34	35	40	3	cereals	or	yes	no	no	yes	yes	0	no										34	yes
29	166	68	36	39	3	other	no	no	no	yes	yes	yes	0	no										68	yes
30	168	63	33	39	4	sandwich	yes	yes	yes	yes	no	yes	0.5	yes										63	yes
31	165	58	36	37	3	sandwich	no	yes	no	yes	yes	no	1	no										58	no
32	171	75	33	41	4	sandwich	yes	yes	no	yes	yes	yes	0	no										75	yes
33	165	77	38	39	3	sandwich	yes	yes	no	no	no	no	0	no										77	yes
34	161	55	37	38	3	porridge	yes	yes	yes	yes	yes	yes	0	no										55	yes
35	169	55	35	38	3	sandwich	sometime	yes	no	yes	yes	yes	0	no										55	yes
36	173	60	37	42	3	cereals	or	yes	yes	no	no	no	0.5	no										60	yes
37	167	80	37.5	41	3	other	yes	yes	no	yes	yes	yes	2	no										80	yes
38	158	70	35	38	3	cereals	or	yes	yes	yes	no	no	0	no										70	no
39	165	61	37	39	3	other	sometime	yes	yes	no	no	no	0.5	yes										61	yes
40	164	38	37	39	3	sandwich	yes	yes	yes	yes	no	yes	0	no										38	yes
41	185	80	60	41	4	cereals	or	sometime	yes	no	yes	no	0	no										80	yes
42	177	63	60	40	2	sandwich	no	no	no	yes	yes	no	0	no										63	no
43	160	70	37	39	4	sandwich	sometime	yes	yes	yes	no	yes	0	no										70	yes
44	162	70	35	40	3	sandwich	no	yes	no	no	no	no	2	no										70	no
45	172	38	62	39	4	other	sometime	yes	no	yes	no	no	0	no										38	no
46	173	74	37	42	3	sandwich	yes	yes	no	yes	yes	no	1	no										74	no
47	173	64	36	42	4	other	yes	yes	no	yes	yes	yes	0.2	no										64	yes
48	180	82	38	46	4	other	yes	yes	yes	yes	yes	yes	3	no										82	yes
49	189	82	43	43	4	cereals	or	no	yes	no	yes	yes	2.5	yes										82	yes
50	170	80	36	41	4	cereals	or	no	yes	no	yes	no	0	no										80	yes
51	176	74	36	42	3	porridge	yes	yes	yes	yes	yes	yes	0.1	yes										74	yes
52	173	73	34	43	4	other	sometime	yes	yes	yes	no	no	0.1	yes										73	no
53	181	74	35	44	4	sandwich	yes	yes	no	yes	yes	yes	1	no										74	yes
54	183	75	43	43	3	porridge	yes	no	no	yes	yes	yes	3	no										75	yes
55	174	87	37	40	4	sandwich	sometime	yes	yes	yes	no	no	0.5	no										87	yes

Copy -> Paste  
+  
Sort ...

2. Calculate the average and standard deviation of weights depending on the owning of car (NB! Omit the student, who does not know has she or he a car or not).

You can use corresponding functions or PivotTable. If you wish, you can try both variants.

WEIGHT	CAR		CAR		
			No	Yes	
60	no				
50	no	No of students	22	31	
68	no	Average	64.40909091	66.59677419	
56	no	Standard deviation	9.945849055	11.10886985	
65	no				
53	no				
66	no				
85	no	Values	Column Labels	-II	
			no	yes	Grand Total
58	no	Count of WEIGHT	22	31	53
90	no	Average of WEIGHT2	64.40909091	66.59677419	65.68867925
60	no	StdDev of WEIGHT3	9.945849055	11.10886985	10.59854236
52	no				
62	no				
66	no				
58	no				
60	no				
70	no				
63	no				
70	no				
58	no				
74	no				
73	no				
70	yes				
47.5	yes				
75	yes				
50	yes				
55	yes				
60	yes				
50	yes				
63	yes				
60	yes				
62	yes				
60	yes				
54	yes				
68	yes				
63	yes				
75	yes				
77	yes				
55	yes				
53	yes				
80	yes				
61	yes				
58	yes				
80	yes				
70	yes				
64	yes				
82	yes				
82	yes				
80	yes				



3. Formulate the hypothesis pair and write it down.

<b>t-test</b>	
$H_0$ : The <b>average weights</b> of students owning and not owning a car <b>are not different</b>	
$H_1$ : The <b>average weights</b> of students owning and not owning a car <b>are different</b>	
or	
$H_0: \mu_{No} = \mu_{Yes}$	$\mu_{No}$ - average weights of students not owning a car
$H_1: \mu_{No} \neq \mu_{Yes}$	$\mu_{Yes}$ - average weights of students owning a car

4. Which t-test to use?

**NB! There are three types of t-tests, look at page 12 (step 7b).**

- As compared **groups are independent** (there are different students in groups), before the comparison of means the **variances must be compared** to decide, which t-test to use (this, which assumes equal variances, or this, which assumes unequal variances).
- To compare variances the F-test can be used.**

5. To decide is the weights' variability of students with and without car equal or not, formulate the **hypothesis pair for variances comparison** and **perform F-test (function F.TEST)**.

NB! There is also statistical procedure *F-test (Data-tab -> Data Analysis... -> F-Test Two-Sample for Variances)*, but this tests only one side hypothesis and can't be directly applied to decide about equality of variances.

<b>t-test</b>	
$H_0$ : The <b>average weights</b> of students owning and not owning a car <b>are not different</b>	
$H_1$ : The <b>average weights</b> of students owning and not owning a car <b>are different</b>	
or	
$H_0: \mu_{No} = \mu_{Yes}$	$\mu_{No}$ - average weights of students not owning a car
$H_1: \mu_{No} \neq \mu_{Yes}$	$\mu_{Yes}$ - average weights of students owning a car
<u>The groups are independent. Before comparison of means the variances must be compared to decide, which t-test to use.</u>	
<b>F-test (comparison of variances)</b>	
$H_0: \sigma^2_{No} = \sigma^2_{Yes}$	(the weights' <b>variability</b> of students owning and not owning a car <b>is not different</b> )
$H_1: \sigma^2_{No} \neq \sigma^2_{Yes}$	(the weights' <b>variability</b> of students owning and not owning a car <b>is different</b> )

AD28    =F.TEST(Y2:Y23;Y24:Y54)

Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	WEIGHT	CAR			CAR				
2	60	no		No	Yes				
3	30	no	No of students	22	31				
4	68	no	Average	64.40809091	66.59677419				
5	36	no	Standard deviation	9.945849055	11.10886983				
6	65	no							
7	53	no							
8	66	no							
9	85	no	Column Labels	no	yes	Grand Total			
10	58	no	Values	Count of WEIGHT	22	31	53		
11	90	no	Average of WEIGHT2	64.40809091	66.59677419	65.68867925			
12	60	no	StdDev of WEIGHT3	9.945849055	11.10886983	10.58854236			
13	32	no							
14	62	no							
15	68	no	t-test						
16	58	no	H <sub>0</sub> : The average weights of students owning and not owning a car are not different						
17	60	no	H <sub>1</sub> : The average weights of students owning and not owning a car are different						
18	70	no	or						
19	63	no	H <sub>0</sub> : $\mu_{no} = \mu_{yes}$	$\mu_{no}$ - average weights of students not owning a car					
20	70	no	H <sub>1</sub> : $\mu_{no} \neq \mu_{yes}$	$\mu_{yes}$ - average weights of students owning a car					
21	58	no							
22	74	no	The groups are independent. Before comparison of means the variances must be compared to decide, which t-test to use.						
23	73	no							
24	70	yes	F-test (comparison of variances)						
25	47.5	yes	H <sub>0</sub> : $\sigma^2_{no} = \sigma^2_{yes}$	(the weights' variability of students owning and not owning a car is not different)					
26	75	yes	H <sub>1</sub> : $\sigma^2_{no} \neq \sigma^2_{yes}$	(the weights' variability of students owning and not owning a car is different)					
27	50	yes							
28	55	yes	function F.TEST	=F.TEST(Y2:Y23;Y24:Y54)					
29	60	yes							
30	50	yes							
31	63	yes							
32	60	yes							
33	62	yes							
34	60	yes							
35	34	yes							
36	68	yes							
37	63	yes							
38	75	yes							
39	77	yes							
40	55	yes							
41	33	yes							
42	80	yes							
43	61	yes							
44	58	yes							
45	80	yes							
46	70	yes							
47	64	yes							
48	82	yes							
49	82	yes							
50	80	yes							
51	74	yes							
52	74	yes							
53	75	yes							

Function Arguments

**F.TEST**

Array1: Y2:Y23 = {60;50;68;56;65;53;66;85;62;60;34;68;63;75;77;55;33;80;61;58;80;70;64;82;82;80;74;74;75}

Array2: Y24:Y54 = {70;47.5;75;50;55;60;50;60;62;60;34;68;63;75;77;55;33;80;61;58;80;70;64;82;82;80;74;74;75}

= 0.605560196

Returns the result of an F-test, the two-tailed probability that the variances in Array1 and Array2 are significantly different.

**Array1** is the first array or range of data and can be numbers, text, or arrays, or references that contain numbers (blanks and text are ignored).

Formula result = 0.60556

[Help on this function](#)    OK

6. Write down **justified** conclusion based on F-test.

<b>F-test (comparison of variances)</b>		
H <sub>0</sub> : $\sigma^2_{Ei} = \sigma^2_{Jah}$	(the weights' variability of students with and without car is not different)	
H <sub>1</sub> : $\sigma^2_{Ei} \neq \sigma^2_{Jah}$	(the weights' variability of students with and without car is different)	
function F.TEST	0.60556	= p > 0,05 => H <sub>0</sub> : the weights' variability in compared groups is not different
<b>The t-test assuming equal variances must be used.</b>		

This is the justification.  
Do you understand?

7. Perform the t-test to compare average weights.

Using both

a) function T.TEST:

The screenshot displays an Excel spreadsheet with the following data and analysis:

WEIGHT	CAR
60	no
50	no
68	no
56	no
65	no
53	no
66	no
85	no
58	no
90	no
60	no
52	no
62	no
66	no
58	no
60	no
70	no
63	no
70	no
58	no
74	no
73	no
70	yes
47,5	yes
75	yes
50	yes
55	yes
60	yes
50	yes
63	yes
60	yes
62	yes
60	yes
54	yes
68	yes
63	yes
75	yes
77	yes
55	yes
53	yes
80	yes
61	yes
58	yes
80	yes
70	yes
64	yes
82	yes
82	yes
80	yes
74	yes
74	yes
75	yes
87	yes
52	

**Summary Statistics:**

	no	yes	Grand Total
Count of WEIGHT	22	31	53
Average of WEIGHT	64.40909091	66.5967742	65.68867925
StdDev of WEIGHT	9.945849055	11.1088698	10.59854236

**t-test**  
 $H_0$ : The average weights of students owning and not owning a car are not different  
 $H_1$ : The average weights of students owning and not owning a car are different  
 or  
 $H_0$ :  $\mu_{No} = \mu_{Yes}$       $\mu_{No}$  - average weights of students not owning a car  
 $H_1$ :  $\mu_{No} \neq \mu_{Yes}$       $\mu_{Yes}$  - average weights of students owning a car

The groups are independent. Before comparison of means the variances must be compared to decide, which t-test to use.

**F-test (comparison of variances)**  
 $H_0$ :  $\sigma^2_{No} = \sigma^2_{Yes}$  (the weights' variability of students owning and not owning a car is not different)  
 $H_1$ :  $\sigma^2_{No} \neq \sigma^2_{Yes}$  (the weights' variability of students owning and not owning a car is different)

function F.TEST     0.60556 = p > 0,05 =>  $H_0$ : the weights' variability in compared groups is not different

The t-test assuming equal variances must be used.

**Comparison of means**  
 function T.TEST     =T.TEST(Y2:Y23;Y24:Y54;2;2)

**Function Arguments Dialog:**

- T.TEST
- Array1: Y2:Y23 = {60;50;68;56;65;53;66;85;58;90;60;52}
- Array2: Y24:Y54 = {70;47,5;75;50;55;60;50;63;60;62;...}
- Tails: 2 (To test two-side hypothesis)
- Type: 2 (Type of t-test assuming equal variability in compared groups)

Returns the probability associated with a Student's t-Test.

**Type** is the kind of t-test: paired = 1, two-sample equal variance (homoscedastic) = 2, two-sample unequal variance = 3.

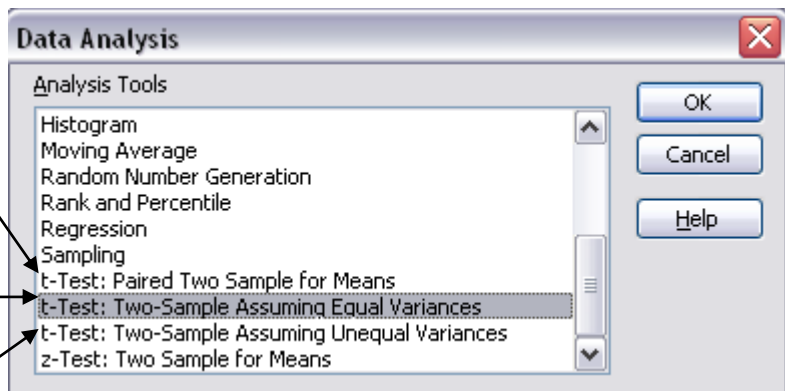
Formula result = 0,464386809

**b) corresponding statistical procedure (Data-tab -> Data Analysis... -> t-Test: ...):**

Comparison of **dependent groups** (pair-wise comparison);  
type 1 in function T.TEST

Comparison of **independent groups** assuming **equal variances**;  
type 2 in function T.TEST

Comparison of **independent groups** assuming **unequal variances**;  
type 3 in function T.TEST



	X	Y	Z	AA	AB	AC	AD	AE	AF
1		WEIGHT	CAR						
2		60	no			CAR			
3		50	no			No	Yes		
4		68	no		No of students	22	31		
5		56	no		Average	64.40909091	66.5967742		
6		65	no		Standard deviation	9.945849055	11.1088698		
7		53	no						
8		66	no						
9		85	no						
10		58	no						
11		90	no						
12		60	no						
13		52	no						
14		62	no						
15		66	no						
16		58	no						
17		60	no						
18		70	no						
19		63	no						
20		70	no						
21		58	no						
22		74	no						
23		73	no						
24		70	yes						
25		47,5	yes						
26		75	yes						
27		50	yes						
28		55	yes						
29		60	yes						
30		50	yes						
31		63	yes						
32		60	yes						
33		62	yes						
34		60	yes						
35		54	yes						
36		68	yes						
37		63	yes						
38		75	yes						
39		77	yes						
40		55	yes						
41		53	yes						
42		80	yes						
43		61	yes						
44		58	yes						
45		80	yes						
46		70	yes						
47		64	yes						
48		82	yes						
49		82	yes						
50		80	yes						
51		74	yes						
52		74	yes						
53		75	yes						
54		87	yes						

Values	no	yes	Grand Total
Count of WEIGHT	22	31	53
Average of WEIGHT2	64.40909091	66.5967742	65.68867925
StdDev of WEIGHT3	9.945849055	11.1088698	10.59854236

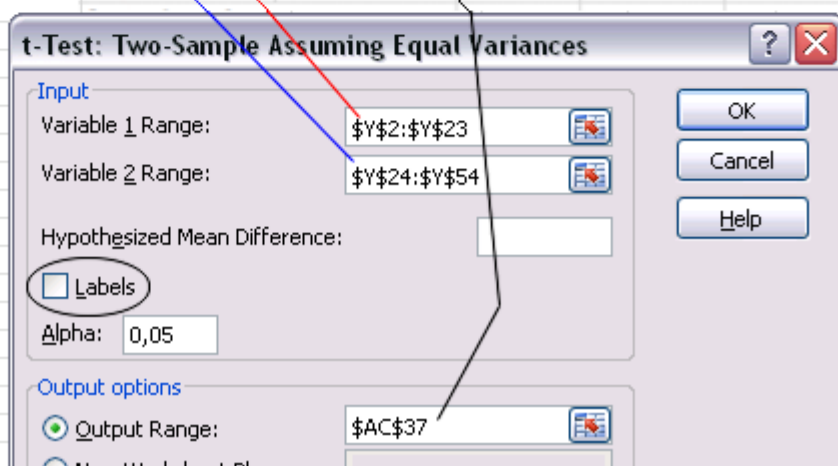
**t-test**  
 $H_0$ : The average weights of students owning and not owning a car are not different  
 $H_1$ : The average weights of students owning and not owning a car are different  
 or  
 $H_0: \mu_{No} = \mu_{Yes}$      $\mu_{No}$  - average weights of students not owning a car  
 $H_1: \mu_{No} \neq \mu_{Yes}$      $\mu_{No}$  - average weights of students owning a car

The groups are independent. Before comparison of means the variances must be compared.

**F-test (comparison of variances)**  
 $H_0: \sigma^2_{No} = \sigma^2_{Yes}$     (the weights' variability of students owning and not owning a car are the same)  
 $H_1: \sigma^2_{No} \neq \sigma^2_{Yes}$     (the weights' variability of students owning and not owning a car are different)  
 function F.TEST    0.60556 = p > 0,05 =>  $H_0$ : the weights' variability are the same

The t-test assuming equal variances must be used.

**Comparison of means**  
 function T.TEST    0.4643868  
 procedure t-test   



8. Write down the final conclusion, justify it(!).

Comparison of means			
function T.TEST	0,4643868	$= p > 0,05 \Rightarrow H_0$ : the average weights of students with and without car are not different	
procedure t-test	t-Test: Two-Sample Assuming Equal Variances		
		<i>Variable 1</i>	<i>Variable 2</i>
Mean	64,40909	66,596774	
Variance	98,91991	123,40699	
Observations	22	31	
Pooled Variance	113,3241		
Hypothesized Mean Difference	0		
df	51		
t Stat	-0,73719		
P(T<=t) one-tail	0,232193		
t Critical one-tail	1,675285		
<b>P(T&lt;=t) two-tail</b>	<b>0,464387</b>	$= p > 0,05 \Rightarrow H_0$ : the average weights of students with and without car are not different	
t Critical two-tail	2,007584		

Other possibilities for the final conclusion:

- „the average weights of students owning and not owning a car are not **statistically significantly different** ( $p > 0,05$ )“, this is a little more scientific and accurate conclusion;
- „the students’ weight does not depend on owning of a car“ (this is little differently phrased conclusion but also correct).

Remark. Significance probability (p-value)  $p = 0.464$  is showing that

- the probability to make a mistake concluding that the average weights are different is 46.4%;
- assuming, that in population there is no difference between compared groups, the probability to get the sample with observed difference just by chance is 0.464.

As from one hand the probability to make a wrong conclusion that the groups are different is too big (formally the traditional limit is 0.05) and from other hand the probability to get observed difference just by chance is also big (once again, the formal limit is 0.05), there is no reason to reject null hypothesis about equality of average weights.

9. But what are showing the other quantities calculated by procedure *t-Test: Two-Sample Assuming Equal Variances*?

- The first part of output contains basic descriptive statistics of compared groups (mean, variance and number of observations; NB! Be careful and don't mix up the compared groups; you may want to write instead abstract names 'Variable 1' and 'Variable 2' the real group names):

	Car='No'	Car='Yes'
	Variable 1	Variable 2
Mean	64,40909091	66,59677419
Variance	98,91991342	123,4069892
Observations	22	31

- The other output parts are already related with hypothesis testing:

Pooled Variance	113,3240757	Common variance calculated assuming the equal variances in compared groups
Hypothesized Mean C	0	
df	51	Empirical (sample based) value of t-statistic
t Stat	-0,737186899	p-value corresponding to one-side hypothesis
P(T<=t) one-tail	0,232193405	Critical value of t-statistic corresponding to one-side hypothesis
t Critical one-tail	1,67528495	p-value corresponding to two-side hypothesis
P(T<=t) two-tail	0,464386809	Critical value of t-statistic corresponding to two-side hypothesis
t Critical two-tail	2,00758377	

**Two-side (or two-tailed) hypothesis** means testing “non-equal *against* equal”:

$$H_0: \mu_{No} = \mu_{Yes},$$

$$H_1: \mu_{No} \neq \mu_{Yes}.$$

**One-side (or one-tailed) hypothesis** means testing “less than *against* more or equal than” (or “more than *against* less or equal than”; in which direction to test depends on data, for example Excel performs one-side t-test testing always is bigger average also statistically significantly bigger):

$$H_0: \mu_{No} \geq \mu_{Yes},$$

$$H_1: \mu_{No} < \mu_{Yes}.$$

Instead of p-value the final conclusion can be made also based on **comparing the empirical value of teststatistic  $|t|$  with its critical value  $t_{critical}$** . If the empirical value is smaller than critical value,  $|t| = 1,68 < 2,01 = t_{critical}$ , then there is no reason to reject the null hypothesis (the empirical value of teststatistic is in interval, where it should be with 95% probability if the null hypothesis is true – so, our data does not allow to reject null hypothesis).

10. Can you reject the null hypothesis of one-side t-test?

How about the conclusion of this one-side test (how to phrase it)?