

## Biomeetria praks 4

### Illustreeritud (mittetäielik) tööjuhend

#### Eeltöö

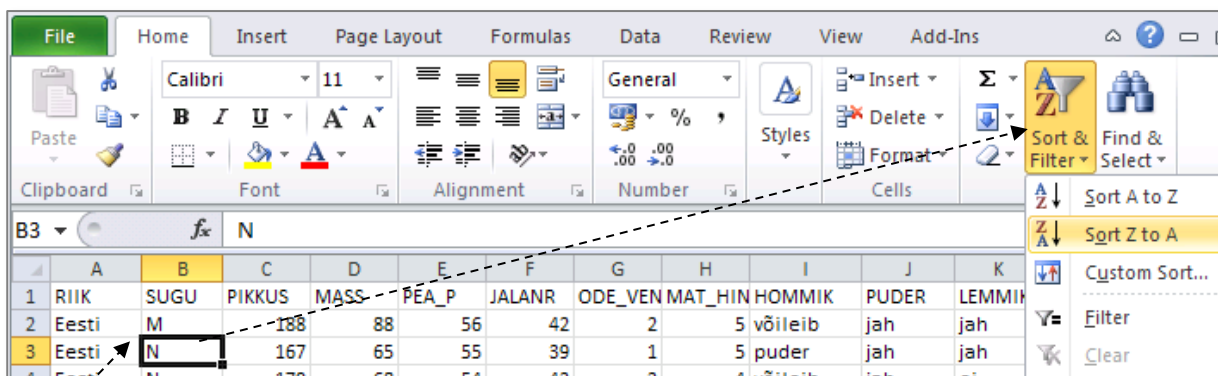
1. Avage MS Excel'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks4' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks4' ülemisse vasakusse nurka.

#### Ülesanne 1.

Tõlgendades teie kursuse neide kui valimit kogu Maaülikooli esimese kursuse neidudest, uurige, kas Maaülikooli esimese kursuse neidude keskmine pikkus erineb Eesti standardist (Eesti naiste keskmine pikkus on 169 cm)?

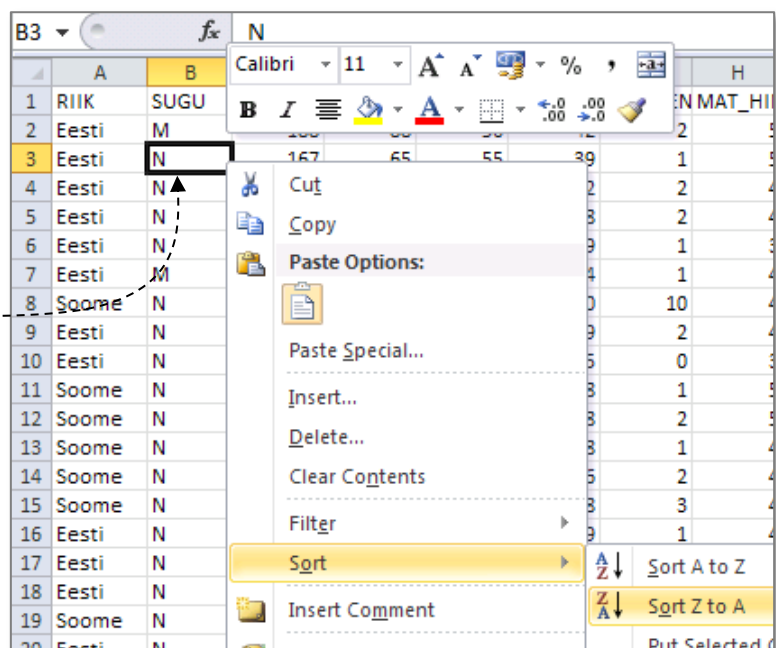
#### Tööjuhend

1. Sorteerige kogu andmetabel veeru 'SUGU' järgi.



- 1) Näiteks pannes kursori veergu 'SUGU' ja valides *Home*-sakilt käsud *Sort & Filter* -> *Sort Z to A* (sorteerimaks neide noormeestest ettepoole).

- 2) Või klikkides suvalisel lahtril veerus 'SUGU' hiire parempoolse klahviga ja valides avanenud rippmenüüst käsud *Sort* -> *Sort Z to A*.



2. Leidke neidude arv, nende keskmine pikkus ja pikkuse standardhälve kasutades funktsioone COUNT, AVERAGE ja STDEV.S.

A	B	C	D	N	O	P	Q	R	S	T	U	V	W
1	RIIK	SUGU	PIKKUS	MASS	SUUSK	AUTO	OLU	SUITS	TEATER	KINO		Neidude pikkus	
2	Eesti	N	167	65	jah	jah	0	enam ei	ei	viimase aas	viimase kuu jooksul	Vaatluste arv	44
3	Eesti	N	179	68	ei	jah	0	ei	rohkem kui	viimase aasta jooksul		Keskmine	168
4	Eesti	N	178	60	jah	ei	2	ei	viimase aas	viimase kuu jooksul		Standardhälve	=STDEV.S(C2:C45)

Neidude pikkus	
Vaatluste arv	44
Keskmine	168
Standardhälve	6,36122

Seega on 44 teie kursuse neiu keskmine pikkus 168,0 cm standardhälbega 6,4 cm

- st, et keskmiselt erineb neidude tegelik pikkus 168,0 sentimeetrist 6,4 cm võrra (vastavalt standardhälbe olemusele);
- ehk, eeldades, et pikkus jaotub normaaljaotuse järgi, jääb vastavalt normaaljaotuse omadustele ligikaudu 68,3% EMÜ esimese kursuse neidude pikkus vahemikku  $168,0 \pm 6,4$  cm ( $\bar{x} \pm s$ ) ja ligikaudu 95,5% neidude pikkus vahemikku  $168,0 \pm 12,8$  cm ( $\bar{x} \pm 2s$ ).

3. Sõnastage kontrollitav hüpoteeside paar ja pange see leitud arv-karakteristikute alla ka kirja.

Näiteks:

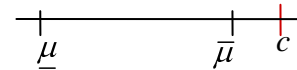
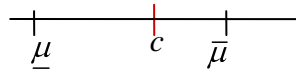
Neidude pikkus	
Vaatluste arv	44
Keskmine	168,0
Standardhälve	6,36122
H <sub>0</sub> : EMÜ 1. kursuse neidude pikkused vastavad Eesti standardile (169 cm)	
H <sub>1</sub> : EMÜ 1. kursuse neidude pikkused ei vasta Eesti standardile (169 cm)	
või	
H <sub>0</sub> : EMÜ 1. kursuse neidude keskmine pikkus ei erine 169 sentimeetrist	
H <sub>1</sub> : EMÜ 1. kursuse neidude keskmine pikkus erineb 169 sentimeetrist	
või	
H <sub>0</sub> : $\mu_T = 169$	$\mu_T$ - EMÜ 1. kursuse neidude keskmine pikkus
H <sub>1</sub> : $\mu_T \neq 169$	

### Meeldetuletuseks teooriast - seos hüpoteeside kontrolli ja usalduspiiride vahel

- Juhul, kui kontrollitavaks hüpoteesiks on mingi andmete alusel hinnatud suuruse võrdlemine konstandiga (mingi arvuga), tehakse otsus sageli baseeruvana uuritava suuruse usaldusintervallil:
  - kui arv, millega andmeist arvatud suurust võrreldakse, jääb usalduspiiride vahele, siis **ei ole alust väita**, et **arvatud suurus erineb** ette antud **konstandist**;
  - kui aga ette antud arv jääb usaldusintervallist väljapoole, **on arvatud suurus konstandist erinev**.
- Näiteks kui soovitakse võrrelda andmeist arvatud keskmist mingi konstandiga (et kas andmed vastavad teatud standardile), on kontrollitav hüpoteeside paar kujul:

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

Kui nüüd  $c \in [\underline{\mu}, \bar{\mu}]$ , siis kehtib  $H_0: \mu = c$ ;      kui aga  $c \notin [\underline{\mu}, \bar{\mu}]$ , siis kehtib  $H_1: \mu \neq c$ .



4. Arvutage liidetav neidude keskmise pikkuse 95% usaldusintervalli leidmiseks (so pool usaldusintervalli laiust) ja tehke seda **kahel viisil**:

**a) funktsiooni CONFIDENCE.NORM abil**

(funktsioonile tuleb ette anda 3 argumenti:

olulisuse nivoo  $\alpha$ , neidude pikkuste standardhälve ja neidude arv);

The image shows a screenshot of the Microsoft Excel 'Insert Function' dialog box. The 'Or select a category' dropdown is set to 'Statistical'. In the 'Select a function' list, 'CONFIDENCE.NORM' is selected and highlighted. Below the list, the function signature is shown: **CONFIDENCE.NORM(alpha;standard\_dev;size)** with a description: 'Returns the confidence interval for a population mean, using a normal distribution.' There are 'OK' and 'Cancel' buttons at the bottom right.

On the left side of the image, there is a smaller screenshot of the Excel ribbon showing the 'Funktioon CONFIDENCE.NORM' button. A dashed box around it contains the following text:

Selle vahepealkirja võiks ise trükkida, et oleks selgem, mis funktsiooni on rakendatud. Ja kursor pange enne funktsiooni tellimist ikka sellesse lahtrisse, kuhu soovite tulemust saada!

The screenshot shows an Excel spreadsheet with the following data in rows 2-4:

1	Neidude pikkus								
2	Vaatluste arv	44							
3	Keskmine	168,0							
4	Standardhälve	6,36122							

The 'Function Arguments' dialog box for CONFIDENCE.NORM is shown with the following arguments:

- Alpha: 0,05
- Standard\_dev: V4
- Size: V2

The dialog box also displays the formula result as 1,879585773 and includes a note: "Kuna tahame leida 95% usaldusintervalli, siis olulisuse nivoo alpha = 0,05." An arrow points from the dialog box to a small table at the bottom right:

Funktsioon CONFIDENCE.NORM	
	1,879586

**b) funktsiooni CONFIDENCE.T abil**

(funktsiooni süntaks on analoogne funktsiooni CONFIDENCE.NORM süntaksiga).

Tulemus:

Funktsioon CONFIDENCE.NORM	
	1,879586
Funktsioon CONFIDENCE.T	
	1,933988

## 5. Arvutage alumine ja ülemine usalduspiir mõlema tulemuse alusel.

	T	U	V	W	X	Y	Z	AA
1	<b>Neitude pikkus</b>							
2		Vaatluste arv	44					
3		Keskmine	168,0					
4		Standardhälve	6,36122					
5								
6								
7		H <sub>0</sub> : EMÜ 1. kursuse neitude pikkused vastavad Eesti standardile (169 cm)						
8		H <sub>1</sub> : EMÜ 1. kursuse neitude pikkused ei vasta Eesti standardile (169 cm)						
9								
10		või						
11								
12		H <sub>0</sub> : EMÜ 1. kursuse neitude keskmine pikkus ei erine 169 sentimeetrist						
13		H <sub>1</sub> : EMÜ 1. kursuse neitude keskmine pikkus erineb 169 sentimeetrist						
14								
15		või						
16								
17		H <sub>0</sub> : μ <sub>T</sub> = 169	μ <sub>T</sub> - EMÜ 1. kursuse neitude keskmine pikkus					
18		H <sub>1</sub> : μ <sub>T</sub> ≠ 169						
19								
20								
21		Funktsioon CONFIDENCE.NORM						
22			1,879586		Alumine usalduspiir	166,1204	=V3-V22	
23					Ülemine usalduspiir	169,8796		
24		Funktsioon CONFIDENCE.T						
25			1,933988		Alumine usalduspiir	166,066	=V3-V25	
26					Ülemine usalduspiir	169,934		
27								

## Kumb 95%-usaldusintervallidest on laiem? Miks?

Vastus.

Funktsiooniga CONFIDENCE.T arvatu on pisut laiem.

Põhjuseks see, et funktsioon CONFIDENCE.T arvutab usalduspiirid  $t$ -jaotuse baasil valemist  $\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}$ ,

funktsioon CONFIDENCE.NORM arvutab aga **asümptootilised** (ligikaudsed, kusjuures täpsus on seda suurem, mida rohkem on andmeid) ja väikeste valimite korral pisut liiga kitsad usalduspiirid standardse normaaljaotuse baasil valemist  $\bar{x} \pm z_{1-\alpha/2} \frac{s}{\sqrt{n}}$ .

Suurused  $t_{1-\alpha/2, n-1} = t_{0,975;43} = 2,017$  ja  $z_{1-\alpha/2} = z_{0,975} = 1,96$  näitavad, milline on see väärtus, millest vastava  $t$ -jaotuse või siis standardse normaaljaotuse korral on väiksemad 97,5% väärtustest (ehk, millest suuremaid väärtusi võib esineda vaid tõenäosusega 0,025), neist suurustest esimene on *Excel* 2010-s leitav näiteks funktsiooniga =T.INV(0,975;43) ja teine funktsiooniga =NORM.S.INV(0,975).

## Kui aru ei saanud, tutvu 3. loengu materjalidega.

**NB!** Vanemates *Exceli* versioonides ei leidu kumbagi funktsioonidest CONFIDENCE.T ja CONFIDENCE.NORM. Neist teisega, mis leiab usalduspiirid normaaljaotuse baasil, on analoogne funktsioon CONFIDENCE,  $t$ -jaotusel baseeruvate usalduspiiride leidmiseks aga vanemais *Exceli* versioonides funktsioon puudub ja kasutada tuleb protseduuri *Descriptive Statistics* valikut *Confidence Level for Mean*, mis arvutab usalduspiiride leidmiseks vajaliku liidetava  $t$ -jaotuse baasil (vt eelmise praktikumi viimast ülesannet).

**6. Otsus püstitatud hüpoteesi osas – kas EMÜ esimese kursuse neidude keskmine pikkus erineb Eesti standardist (169 cm)?**

**Sõnastage lõppjäreltus (koos põhjendusega) ja pange see kirja.**

Spikker. Et Eesti naiste keskmine pikkus 169 cm jääb EMÜ esimese kursuse neidude keskmise pikkuse 95% usaldusintervalli sisse,  $169 \in (166,1; 169,9)$ , siis ei ole alust lugeda tõestatuks alternatiivset hüpoteesi keskmise pikkuse erinevusest 169 sentimeetrist ja tuleb jääda nullhüpoteesi  $H_0$  juurde: EMÜ esimese kursuse neidude keskmine pikkus ei erine 169 sentimeetrist.

**7. Lisa. Naiste keskmine pikkus maailmas on 154 cm. Kas on alust väita, et EMÜ esimese kursuse tütarlaste pikkused erinevad maailma keskmisest?**

## Ülesanne 2.

Kas autot omavate ja mitte omavate tudengite keskmised kehamassid on erinevad?

### Tööjuhend

1. Moodustage samale töölehele abitabel veergudest 'MASS' ja 'AUTO' ning sorteerige abitabel veeru 'AUTO' järgi.

	A	B	C	D	E	O	P	Q	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	AUTO	OLU	SUITS		Neidude pikkus								MASS	AUTO
2	Eesti	N	167	65	55	jah		0 enam ei, aeg		Vaatluste arv	44							65	jah
3	Eesti	N	179	68	54	jah		0 ei		Keskmine	168,0							68	jah
4	Eesti	N	178	60	55	ei		2 ei		Standardhälve	6,36122							60	ei
5	Eesti	N	161	70	56,5	ei		0,1 ei										70	ei
6	Soome	N	174	58	58	ei		0 ei										58	ei
7	Eesti	N	171	55		jah		0 ei		H <sub>0</sub> : EMÜ 1. kursuse neidude pikkused vastavad Eesti standardile (169 cm)								55	jah
8	Eesti	N	165	55		ei		0 ei		H <sub>1</sub> : EMÜ 1. kursuse neidude pikkused ei vasta Eesti standardile (169 cm)								55	ei
9	Soome	N	167	65	58,5	jah		0 ei										65	jah
10	Soome	N	163	69	58	jah		0 ei		või								69	jah
11	Soome	N	165	55	58	jah		0 ei										55	jah
12	Soome	N	169	61	58	jah		0 ei		H <sub>0</sub> : EMÜ 1. kursuse neidude keskmine pikkus ei erine 169 sentimeetrist								61	jah
13	Soome	N	170	68	58	ei		0 ei		H <sub>1</sub> : EMÜ 1. kursuse neidude keskmine pikkus erineb 169 sentimeetrist								68	ei
14	Eesti	N	165	51	56	jah		0 ei										51	jah
15	Eesti	N	170	57	56	jah		0 ei		või								57	jah
16	Eesti	N	170	63	56	ei		0 ei										63	ei
17	Soome	N	162	65	58	jah		0,5 ei		H <sub>1</sub> : $\mu_T = 169$	$\mu_T$ - EMÜ 1. kursuse neidude keskmine pikkus							65	jah
18	Eesti	N	168	80	56	ei		0 jah		H <sub>1</sub> : $\mu_T \neq 169$								80	ei
19	Eesti	N	177	60		ei		0 ei										60	ei
20	Eesti	N	168	73	53	jah		0 ei										73	jah
21	Eesti	N	165	50,1	52	jah		0,33 ei		Funktsioon CONFIDENCE.NORM								50,1	jah
22	Eesti	N	162	63	56	ei		0,5 ei			1,879586		Alumine usalduspiir	166,1204	=V3-V22			63	ei
23	Soome	N	162	52	55	ei		0,5 ei					Ülemine usalduspiir	169,8796				52	ei
24	Soome	N	161	47	62	jah		0 ei		Funktsioon CONFIDENCE.T								47	jah
25	Soome	N	179	80	62	ei		0 ei			1,933988		Alumine usalduspiir	166,066	=V3-V25			80	ei
26	Soome	N	179	68	57	jah		0 ei					Ülemine usalduspiir	169,934				68	jah
27	Soome	N	176	80	55	jah		0 ei										80	jah
28	Soome	N	155	52	53	ei		0,5 jah										52	ei
29	Eesti	N	162	57	53	ei		1 ei										57	ei
30	Eesti	N	167	66	54	ei		0 ei										66	ei
31	Eesti	N	180	71	56	jah		0 jah										71	jah
32	Eesti	N	164	52	56	jah		0 ei										52	jah
33	Eesti	N	179	75	56	jah		0 ei										75	jah
34	Venemaa	N	167	46,5	55	ei		0 ei										46,5	ei
35	Eesti	N	172	64	55	ei		0 ei										64	ei
36	Soome	N	167	55	50	jah		1,5 ei										55	jah
37	Soome	N	171	75	56	jah		0 ei										75	jah
38	Soome	N	160	61	54	jah		0 ei										61	jah
39	Soome	N	168	75	58	ei		0 ei										75	ei
40	Soome	N	163	50	54	jah		0 ei										50	jah
41		N	169	65	57	jah		0,5 ei										65	jah
42	Soome	N	172	54	55	ei		0 enam ei, aeg										54	ei
43	Eesti	N	163	52	48	ei		0 ei										52	ei
44	Eesti	N	164	55	55,5	ei		0 ei										55	ei
45	Soome	N	156	48	50	ei		0 ei										48	ei
46	Eesti	M	188	88	56	jah		3 ei										88	jah
47	Eesti	M	185	69	56	ei		0 ei										69	ei
48	Eesti	M	188	80	58	jah		0 ei										80	jah
49	Eesti	M	183	73	54	jah		2,5 jah										73	jah
50	Eesti	M	190	85	58	jah		4 ei										85	jah
51	Eesti	M	184	78	55	jah		2 ei										78	jah
52	Eesti	M	175	62	56,5	jah		0 ei										62	jah
53	Eesti	M	183	80	56	jah		4 enam ei, aeg										80	jah
54			163	53	58	ei		0 jah										53	ei

Copy -> Paste  
+  
Sort ...

2. Leidke nii autot omavate ja mitte omavate tudengite arvud, keskmised kehamassid ja kehamassi standardhälbed

Kasutada võite nii vastavaid funktsioone kui ka Pivot Table'i abi. Kui soovite, arvutage mõlemal viisil.

MASS	AUTO		Auto		
46,5	ei		Ei	Jah	
48	ei	Tudengite arv	23	30	
52	ei	Keskmine mass	60,9	66,0	
52	ei	Standardhälve	9,51	11,27	
52	ei				
53	ei				
54	ei				
55	ei				
55	ei				
57	ei				
58	ei				
60	ei				
60	ei				
63	ei				
63	ei				
64	ei				
66	ei				
68	ei				
69	ei				
70	ei				
75	ei				
80	ei				
80	ei				
47	jah				
50	jah				
50,1	jah				
51	jah				
52	jah				
55	jah				
55	jah				
55	jah				
57	jah				
61	jah				
61	jah				
62	jah				
65	jah				
65	jah				
65	jah				
65	jah				
68	jah				
68	jah				
69	jah				
71	jah				
73	jah				
73	jah				
75	jah				
75	jah				
78	jah				
80	jah				
80	jah				
80	jah				
85	jah				
88	jah				

Values	Column Labels		Grand Total
	ei	jah	
Count of MASS	23	30	53
Average of MASS2	60,89130435	65,97	63,76603774
StdDev of MASS3	9,505927005	11,27062554	10,74848725



3. Sõnastage kontrollitav hüpoteeside paar ja pange see ka kirja.

<b>t-test</b>	
$H_0$ : Autot omavate ja mitte omavate tudengite keskmised massid on võrdsed	
$H_1$ : Autot omavate ja mitte omavate tudengite keskmised massid on erinevad	
või	
$H_0: \mu_{EI} = \mu_{Jah}$	$\mu_{EI}$ - autot mitte omavate tudengite keskmine kehamass
$H_1: \mu_{EI} \neq \mu_{Jah}$	$\mu_{Jah}$ - autot omavate tudengite keskmine kehamass

4. Millist t-testi nende keskmiste võrdlemisel kasutada?

**NB! t-testi on 3 tüüpi, vt lk 12 (punkt 7 b).**

- Et tegu on sõltumatute vaatlustega (võrreldavad grupid koosnevad erinevatest tudengitest), tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsustamaks, millist t-testi kasutada (kas seda, mis eeldab keskmiste võrdlemisel võrdset varieeruvust, või seda, mis arvutab mõlema grupi tarvis eraldi dispersioonid).
- Varieeruvuse (dispersioonide) võrdlemiseks kasutatakse **F-testi**.

5. Pange kirja kontrollitav hüpoteeside paar ja viige läbi **F-test**, otsustamaks autot omavate ja mitte omavate tudengite kehamasside varieeruvuse võrdumise või mittevõrdumise üle (**funktsioon F.TEST**).

NB! Protseduur *F-test (Data-sakk -> Data Analysis... -> F-Test Two-Sample for Variances)* testib vaid ühepoolset hüpoteesi ega ole seetõttu otseselt rakendatav, otsustamaks dispersioonide võrdumise või mittevõrdumise üle.

<b>t-test</b>	
$H_0$ : Autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad	
$H_1$ : Autot omavate ja mitte omavate tudengite keskmised massid on erinevad	
või	
$H_0: \mu_{EI} = \mu_{Jah}$	$\mu_{EI}$ - autot mitte omavate tudengite keskmine kehamass
$H_1: \mu_{EI} \neq \mu_{Jah}$	$\mu_{Jah}$ - autot omavate tudengite keskmine kehamass
Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsustamaks, millist t-testi kasutada.	
<b>F-test (võrdleme dispersioone)</b>	
$H_0: \sigma_{EI}^2 = \sigma_{Jah}^2$	(autot omavate ja mitte omavate tudengite kehamasside varieeruvus ei ole erinev)
$H_1: \sigma_{EI}^2 \neq \sigma_{Jah}^2$	(autot omavate ja mitte omavate tudengite kehamasside varieeruvus on erinev)

Excel spreadsheet showing an F-test analysis. The formula bar displays `=F.TEST(AC2:AC24;AC25:AC54)`. The spreadsheet contains data for 'ei' and 'jah' groups, including counts, averages, and standard deviations. A summary table shows the results of the F-test, with a p-value of 0,414168071. A 'Function Arguments' dialog box is open, showing the arrays AC2:AC24 and AC25:AC54, and the resulting p-value.

Values	ei	jah	Grand Total
Count of MASS	23	30	53
Average of MASS2	60,89130435	65,97	63,76603774
StdDev of MASS3	9,505927005	11,27062554	10,74848725

**F-test (võrdleme dispersioone)**  
 $H_0: \sigma_{ei}^2 = \sigma_{jah}^2$  (autot omavate ja mitte omavate tudengite kehamasside varieeruvus ei ole erinev)  
 $H_1: \sigma_{ei}^2 \neq \sigma_{jah}^2$  (autot omavate ja mitte omavate tudengite kehamasside varieeruvus on erinev)

Formula result = 0,414168071

6. Sõnastage järeldus F-testist, põhjendage.

Summary of the F-test results and conclusion:

<b>F-test (võrdleme dispersioone)</b>	
$H_0: \sigma_{ei}^2 = \sigma_{jah}^2$ (autot omavate tudengite kehamasside varieeruvus ei ole erinev)	
$H_1: \sigma_{ei}^2 \neq \sigma_{jah}^2$ (autot omavate tudengite kehamasside varieeruvus on erinev)	
funktsioon F.TEST	0,414168071
	$= p > 0,05 \Rightarrow H_0$ : kehamasside varieeruvus võrreldavais gruppides ei ole erinev

Kasutada võib t-testi, mis eeldab kehamasside võrdset varieeruvust võrreldavates gruppides.

See ongi põhjendus. Aru saite?

7. Viige läbi t-test võrdlemaks keskmisi kehamasse.

Tehke seda kahel viisil:

a) kasutades funktsiooni T.TEST:

Excel spreadsheet showing a t-test analysis. The formula bar shows `=T.TEST(AC2:AC24;AC25:AC54;2;2)`. The spreadsheet data includes columns for 'MASS' and 'AUTO' (ei/jah). A summary table shows 'Count of MASS' (23 ei, 30 jah) and 'Average of MASS2' (60,89130435 ei, 65,97 jah). The t-test results show a p-value of 0,414168071, which is greater than 0,05, leading to the conclusion that the variance is not significantly different.

**t-test**  
 $H_0$ : Autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad  
 $H_1$ : Autot omavate ja mitte omavate tudengite keskmised massid on erinevad

**F-test (võrdleme dispersioone)**  
 $H_0$ :  $\sigma^2_{ei} = \sigma^2_{jah}$  (autot omavate ja mitte omavate tudengite kehamasside varieeruvus ei ole erinev)  
 $H_1$ :  $\sigma^2_{ei} \neq \sigma^2_{jah}$  (autot omavate ja mitte omavate tudengite kehamasside varieeruvus on erinev)  
 funktsioon F.TEST 0,414168071 = p > 0,05 =>  $H_0$ : kehamasside varieeruvus võrreldavais gruppides ei ole erinev  
 Kasutada võib t-testi, mis eeldab kehamasside võrdset varieeruvust võrreldavates gruppides.

**Võrdleme keskmisi**  
 funktsioon T.TEST =T.TEST(AC2:AC24;AC25:AC54;2;2)

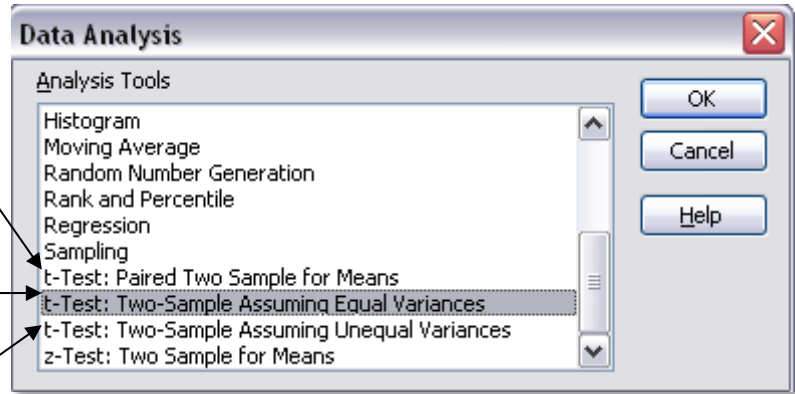
**Function Arguments dialog box:**  
 T.TEST  
 Array1 AC2:AC24 = {46,5;48;52;52;52;53;54;55;55;57;58;...}  
 Array2 AC25:AC54 = {47;50;50,1;51;52;55;55;55;57;61;...}  
 Tails 2 (Testime kahepoolset hüpoteesi)  
 Type 2 (Võrdse varieeruvuse eeldusel (miks võib seda eeldada?) sobiv t-testi tüüp)  
 Returns 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 = 0,088306401

**b) vastava statistikaprotseduuri abil (Data-sakk -> Data Analysis... -> t-Test: ...):**

Sõltuvate gruppide (paariviisiline) võrdlus; funktsioonis T.TEST tüüp nr 1

Sõltumatute gruppide võrdlus võrdsete dispersioonide eeldusel; funktsioonis T.TEST tüüp nr 2

Sõltumatute gruppide võrdlus mittevõrdsete dispersioonide eeldusel; funktsioonis T.TEST tüüp nr 3



	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1		MASS	AUTO				Auto		
2		46,5	ei			Ei	Jah		
3		48	ei		Tudengite arv	23	30		
4		52	ei		Keskmine mass	60,9	66,0		
5		52	ei		Standardhälve	9,51	11,27		
6		52	ei						
7		53	ei						
8		54	ei						
9		55	ei						
10		55	ei		Values	Column Labels		Grand Total	
11		57	ei		Count of MASS	ei	jah		
12		58	ei		Average of MASS2	60,89130435	65,97	63,76603774	
13		60	ei		StdDev of MASS3	9,505927005	11,27062554	10,74848725	
14		60	ei						
15		63	ei						
16		63	ei		<b>t-test</b>				
17		64	ei		H <sub>0</sub> : Autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad				
18		66	ei		H <sub>1</sub> : Autot omavate ja mitte omavate tudengite keskmised massid on erinevad				
19		68	ei		või				
20		69	ei		H <sub>0</sub> : $\mu_{ei} = \mu_{jah}$	$\mu_{ei}$ - autot mitte omavate tudengite keskmine kehamass			
21		70	ei		H <sub>1</sub> : $\mu_{ei} \neq \mu_{jah}$	$\mu_{jah}$ - autot omavate tudengite keskmine kehamass			
22		75	ei						
23		80	ei						
24		80	ei						
25		47	jah						
26		50	jah						
27		50,1	jah						
28		51	jah						
29		52	jah						
30		55	jah						
31		55	jah						
32		55	jah						
33		57	jah						
34		61	jah						
35		61	jah						
36		62	jah						
37		65	jah						
38		65	jah						
39		65	jah						
40		65	jah						
41		68	jah						
42		68	jah						
43		69	jah						
44		71	jah						
45		73	jah						
46		73	jah						
47		75	jah						
48		75	jah						
49		78	jah						
50		80	jah						
51		80	jah						
52		80	jah						
53		85	jah						
54		88	jah						

Values	ei	jah	Grand Total
Count of MASS	23	30	53
Average of MASS2	60,89130435	65,97	63,76603774
StdDev of MASS3	9,505927005	11,27062554	10,74848725

**t-test**

H<sub>0</sub>: Autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad

H<sub>1</sub>: Autot omavate ja mitte omavate tudengite keskmised massid on erinevad

või

H<sub>0</sub>:  $\mu_{ei} = \mu_{jah}$       $\mu_{ei}$  - autot mitte omavate tudengite keskmine kehamass

H<sub>1</sub>:  $\mu_{ei} \neq \mu_{jah}$       $\mu_{jah}$  - autot omavate tudengite keskmine kehamass

Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone

**F-test (võrdleme dispersioone)**

H<sub>0</sub>:  $\sigma^2_{ei} = \sigma^2_{jah}$  (autot omavate ja mitte omavate tudengite kehamass)

H<sub>1</sub>:  $\sigma^2_{ei} \neq \sigma^2_{jah}$  (autot omavate ja mitte omavate tudengite kehamass)

funktsioon F.TEST     0,414168071 = p > 0,05 => H<sub>1</sub>: kehamasside varieer

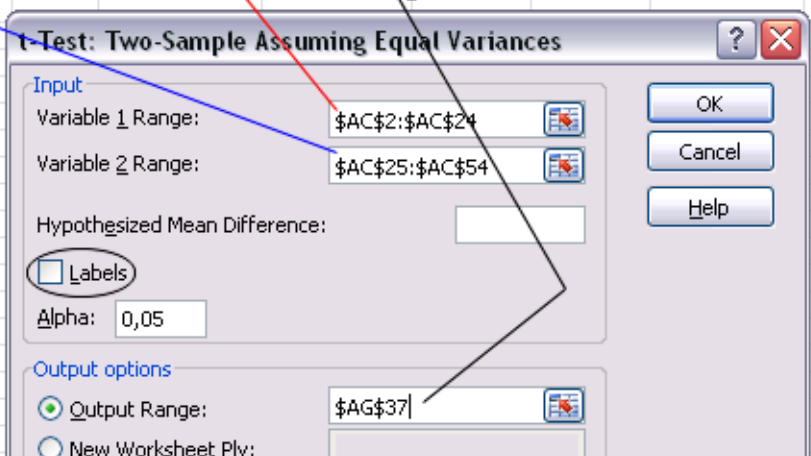
Kasutada võib t-testi, mis eeldab kehamasside võrdset varieeruvust võrre

**Võrdleme keskmisi**

funktsioon T.TEST     0,088306401

protseduur t-test



## 8. Sõnastage lõppjärelus, põhjendage.

funktsioon T.TEST	0,088306401	<u>= p &gt; 0,05 =&gt; H<sub>0</sub>: autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad</u>	
protseduur t-test	t-Test: Two-Sample Assuming Equal Variances		
		<i>Variable 1</i>	<i>Variable 2</i>
Mean	60,89130435	65,97	
Variance	90,36264822	127,027	
Observations	23	30	
Pooled Variance	111,2110051		
Hypothesized Me	0		
df	51		
t Stat	-1,737660446		
P(T<=t) one-tail	0,0441532		
t Critical one-tail	1,67528495		
P(T<=t) two-tail	0,088306401	<u>= p &gt; 0,05 =&gt; H<sub>0</sub>: autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad</u>	
t Critical two-tail	2,00758377		

Veel võimalusi lõppjäreluse sõnastamiseks:

- „**autot omavate ja autot mitteomavate tudengite keskmised kehamassid ei ole statistiliselt oluliselt erinevad ( $p > 0,05$ )**“,  
so selline pisut korrektsemalt ja teaduslikumalt sõnastatud järelus;
- „kehamass ei sõltu sellest, kas tudengil on auto või mitte“ (vähe teise nurga alt sõnastatud järelus, aga ka õige).

Märkus. Olulisuse tõenäosuse väärtus  $p = 0,088$  näitab, et

- väites, et kõigi autoomanikest tudengite (ka nende, kelle kohta meil andmeid ei ole) keskmine kehamass erineb kõigi autot mitteomavate tudengite keskmisest kehamassist, eksiksime 8,8%-lise tõenäosusega;
- eeldades, et autot omavate ja mitteomavate tudengite keskmised kehamassid (arvutatuna üle kõigi tudengite, st üldkogumis) on tegelikult võrdsed, siis tõenäosus saada nii suur kehamasside erinevus, nagu ilmnes meie andmetes lihtsalt tänu juhusele, on 0,088. Et see tõenäosus ei ole väga väike (formaalselt on piiriks 0,05), siis ei ole alust lükata ümber nullhüpoteesi kehamasside võrdsusest.

9. Aga mida nendest protseduuri *t-Test: Two-Sample Assuming Equal Variances* tulemustest veel välja saab lugeda?

- Näiteks keskmise kehamassi, kehamasside dispersiooni ja vaatluste arvu võrreldavates gruppides (seejuures on mõistlik peale analüüsi teostamist kirjutada veergudele peale, mis grupi kohta seal olevad arvud käivad):

	Auto='Ei'	Auto='Jah'
	Variable 1	Variable 2
Mean	60,89130435	65,97
Variance	90,36264822	127,027
Observations	23	30

- Ülejäänud osa väljatrükist on seotud juba kahe grupi keskmiste võrdlemisega:

Pooled Variance	111,2110051	Võrdsete dispersioonide eeldusel arvutatud ühine dispersioon
Hypothesized Me	0	
df	51	<i>t</i> -statistiku empiiriline (andmete alusel arvutatud) väärtus
t Stat	-1,737660446	1-poolsele hüpoteesile vastav <i>p</i> -väärtus
P(T<=t) one-tail	0,0441532	1-poolsele hüpoteesile vastav <i>t</i> -statistiku kriitiline väärtus
t Critical one-tail	1,67528495	2-poolsele hüpoteesile vastav <i>p</i> -väärtus
P(T<=t) two-tail	0,088306401	2-poolsele hüpoteesile vastav <i>t</i> -statistiku kriitiline väärtus
t Critical two-tail	2,00758377	

**Kahepoolne hüpotees** tähendab statistikas võrdumise ja mittevõrdumise testimist:

$$H_0: \mu_{Ei} = \mu_{Jah},$$

$$H_1: \mu_{Ei} \neq \mu_{Jah}.$$

**Ühepoolne hüpotees** testib vaid ühepoolset erinevust – kas näiteks üks keskmine on teisest väiksem või mitte. Seejuures konstrueerib *Excel* hüpoteeside paari vastavalt keskmiste väärtustele, testides alati seda, kas suurem keskmine on ikka statistiliselt oluliselt suurem või mitte. Seega on antud juhul testitav ühepoolsete hüpoteeside paar kujul

$$H_0: \mu_{Ei} \geq \mu_{Jah},$$

$$H_1: \mu_{Ei} < \mu_{Jah},$$

sest autoomanikest tudengite keskmine kehamass on suurem.

Olulisuse tõenäosuse *p* asemel võib otsuse vastu võtmisel lähtuda ka teststatistiku empiirilise (andmete alusel arvutatud) väärtuse absoluutväärtuse  $|t|$  ja teststatistiku kriitilise väärtuse  $t_{critical}$  võrdlusest (vt hüpoteeside kontroll, 3. loeng).

Nimelt, kuna  $|t| = 1,74 < 2,01 = t_{critical}$ , siis ei ole põhjust nullhüpoteesi ümber lükata (*t*-statistiku väärtus jääb sellesse piirkonda, kuhu ta nullhüpoteesi kehtides 95%-tõenäosusega peakski jääma).

10. Kas testides ühepoolset hüpoteesi võinuks lugeda alternatiivse hüpoteesi  $H_1$  tõestatuks?

Kuidas kõlab lõppjärelus testitud ühepoolse hüpoteesi kohta?