

Biomeetria praks 4

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

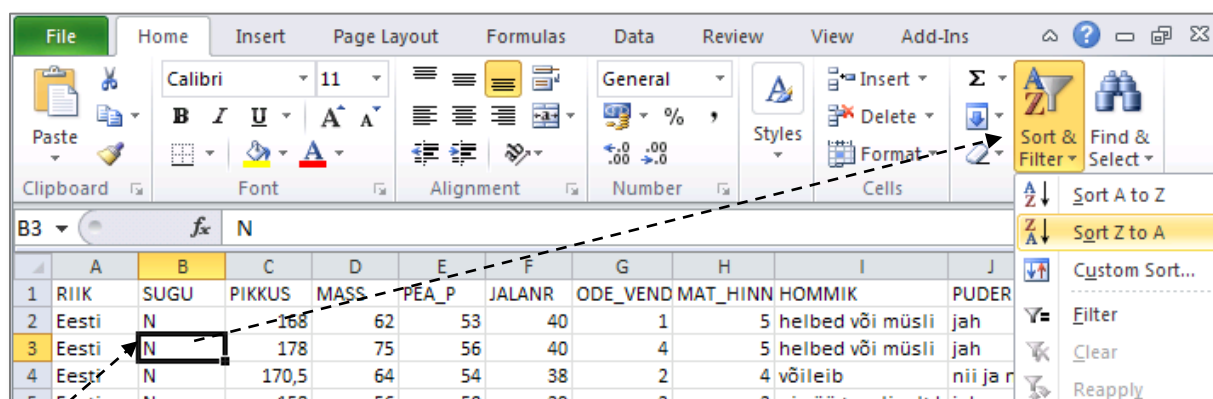
1. Avage MS Excel'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks4' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks4' ülemisse vasakusse nurka.

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 168 cm)?

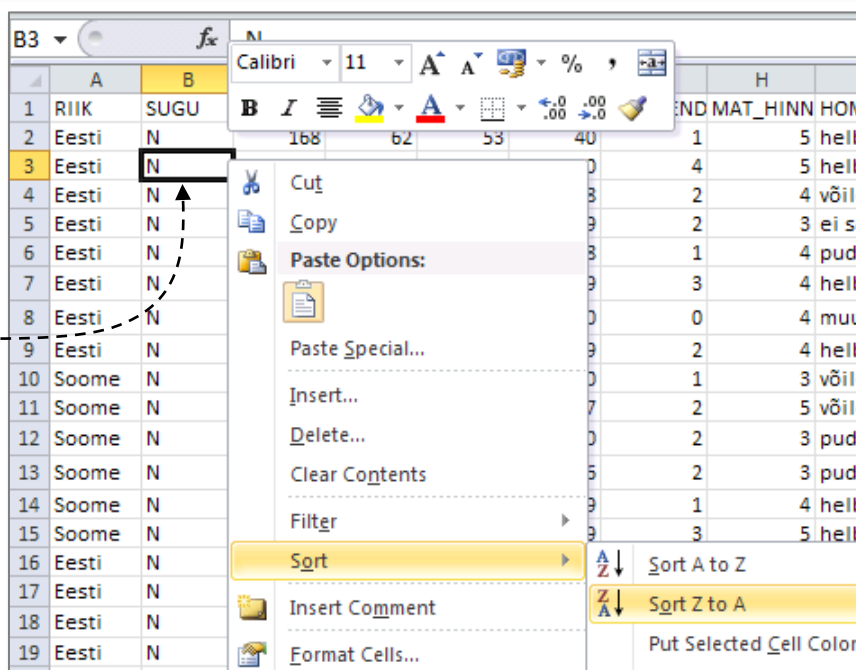
Tööjuhend

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.

... ▾ X ✓ f_x =STDEV.S(C2:C53)

	A	B	C	D	P	Q	R	S	T	U	V	W
1	RIIK	SUGU	PIKKUS	MASS	OLU	SUITS	TEATER	KINO			Neidude pikkus	
2	Eesti	N	168	62	0	ei	rohkem kui aa	viimase kuu jooksul			Vaatluste arv	52
3	Eesti	N	178	75	0	jah	viimase kuu jo	viimase kuu jooksul			Keskmine	168,33
4	Eesti	N	170,5	64	0	ei	viimase aasta j	viimase aasta jooksul			Standardhälve	=STDEV.S(C2:C53)
5	Eesti	N	158	56	0	ei	rohkem kui aa	viimase 10 päeva jooksul				
6	Eesti	N	163	60	0	ei	viimase aasta j	viimase aasta jooksul				
7	Eesti	N	177	70	0	ei	viimase aasta j	viimase kuu jooksul				
8	Eesti	N	170	51	0	ei	viimase aasta j	rohkem kui aasta tagasi				
9	Eesti	N	170	56	0	ei	rohkem kui aa	viimase aasta jooksul				
10	Soome	N	167	60	0	ei	viimase kuu jo	viimase aasta jooksul				
11	Soome	N	157	46	0	ei	rohkem kui aa	viimase aasta jooksul				
12	Soome	N	175	65	0	ei	viimase aasta j	viimase aasta jooksul				
13	Soome	N	157	70	1	enam ei, ag	viimase aasta j	viimase 10 päeva jooksul				
14	Soome	N	173	62	0	ei	rohkem kui aa	viimase aasta jooksul				
15	Soome	N	161	65	2	jah	rohkem kui aa	viimase aasta jooksul				
16	Eesti	N	169	68	0	ei	viimase kuu jo	viimase aasta jooksul				
17	Eesti	N	178	56	0	ei	viimase aasta j	viimase aasta jooksul				
18	Eesti	N	164	62	0	ei	viimase aasta j	viimase aasta jooksul				
19	Eesti	N	159	55	0	ei	viimase aasta j	viimase kuu jooksul				
20	Eesti	N	164	55	0	ei	viimase aasta j	viimase kuu jooksul				
21	Eesti	N	174	74	0,5	ei	viimase kuu jo	viimase 10 päeva jooksul				
22	Eesti	N	171	67	0	ei	viimase aasta j	viimase 10 päeva jooksul				
23	Soome	N	160	56	3	ei	rohkem kui aa	viimase aasta jooksul				
24	Soome	N	170	76	0	jah	rohkem kui aa	viimase kuu jooksul				
25	Soome	N	165	55	0	ei	rohkem kui aa	viimase aasta jooksul				
26	Eesti	N	168	55	0	ei	viimase aasta j	viimase aasta jooksul				
27	Eesti	N	159	48	0	ei	viimase aasta j	viimase aasta jooksul				
28	Soome	N	166	58	0	jah	viimase aasta j	viimase aasta jooksul				
29	Soome	N	169	59	0	ei	viimase kuu jo	viimase 10 päeva jooksul				
30	Soome	N	160	57	0	ei	rohkem kui aa	viimase aasta jooksul				
31	Soome	N	178	70	0	ei	viimase aasta j	viimase kuu jooksul				
32	Soome	N	169	68	0	ei	viimase aasta j	viimase aasta jooksul				
33	Eesti	N	174	62	0	enam ei, ag	viimase aasta j	viimase kuu jooksul				
34	Eesti	N	165	59	1	jah	viimase aasta j	viimase kuu jooksul				
35	Eesti	N	177	76	0	ei	viimase aasta j	viimase aasta jooksul				
36	Soome	N	170	80	0	ei	rohkem kui aa	n				
37	Eesti	N	173	75	0	ei	rohkem kui aa	v				
38	Eesti	N	175,5	66	0	ei	viimase aasta j	v				
39	Eesti	N	172	66	0	jah	viimase aasta j	v				
40	Eesti	N	165	60	3	ei	rohkem kui aa	v				
41	Eesti	N	170	60	7	ei	rohkem kui aa	v				
42	Eesti	N	168	55	0	ei	rohkem kui aa	v				
43	Soome	N	158	75	0	ei	viimase aasta j	v				
44	Eesti	N	170	54	0	ei	viimase aasta j	viimase aasta jooksul				
45	Soome	N	181	72	0,5	ei	rohkem kui aa	viimase kuu jooksul				
46	Soome	N	167	70	0	ei	viimase aasta j	viimase 10 päeva jooksul				
47	Soome	N	166	49	0	ei	viimase aasta j	viimase 10 päeva jooksul				
48	Soome	N	174	51	0	ei	ei ole kunagi ki	viimase 10 päeva jooksul				
49	Soome	N	160	60	0	ei	rohkem kui aa	viimase aasta jooksul				
50	Soome	N	170	65	0	ei	viimase aasta j	viimase aasta jooksul				
51	Soome	N	170	66	0	ei	viimase aasta j	viimase kuu jooksul				
52	Eesti	N	168	65	0	ei	viimase kuu jo	viimase 10 päeva jooksul				
53	Eesti	N	172	62	0	ei	rohkem kui aa	rohkem kui aasta tagasi				
54	Soome	M	183	90	4	jah	viimase kuu jo	viimase aasta jooksul				
55	Soome	M	187	75	3	ei	rohkem kui aa	viimase aasta jooksul				
56	Eesti	M	171	75	0	enam ei, ag	rohkem kui aa	viimase kuu jooksul				
57	Eesti	M										

Seega on 52 teie kursuse
neiu keskmine pikkus
168,3 cm standardhõlbega
6,1 cm

- st, et keskmiselt erineb
neidude tegelik pikkus
168,3 sentimeetrist
6,1 cm võrra (vastavalt
standardhõlbe olemusele);
- ehk, eeldades, et pikkus
jaotub normaaljaotuse
järgi, jääb vastavalt
normaaljaotuse
omadustele
ligikaudu 68,3% EMÜ
esimese kursuse neidude
pikkus vahemikku
 $168,3 \pm 6,1$ cm ($\bar{x} \pm s$) ja
ligikaudu 95,5% neidude
pikkus vahemikku
 $168,3 \pm 12,2$ cm ($\bar{x} \pm 2s$).

3. Sõnastage kontrollitav
hüpoteeside paar
ja pange see leitud
arvarakteristikute alla
ka kirja.

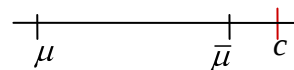
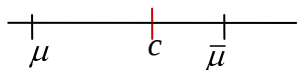
Neidude pikkus	
Vaatluste arv	52
Keskmine	168,3269
Standardhälve	6,119628
H_0 : EMÜ 1. kursuse neidude pikkused vastavad Eesti standardile (168 cm)	
H_1 : EMÜ 1. kursuse neidude pikkused ei vasta Eesti standardile (168 cm)	
või	
H_0 : EMÜ 1. kursuse neidude keskmine pikkus ei erine 168 sentimeetrist	
H_1 : EMÜ 1. kursuse neidude keskmine pikkus erineb 168 sentimeetrist	
või	
H_0 : $\mu_T = 168$	μ_T - EMÜ 1. kursuse neidude keskmine pikkus
H_1 : $\mu_T \neq 168$	

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 α , neidude pikkuste standardhälve ja neidude arv);

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!

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.

Help on this function

OK Cancel

Excel spreadsheet showing data and the CONFIDENCE.NORM function dialog box.

	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1		Neidude pikkus									
2		Vaatluste arv	52								
3		Keskmine	168,3269								
4		Standardhälve	6,119628								
5											
6											
7		H_0 : EMÜ 1. kursuse nei									
8		H_1 : EMÜ 1. kursuse nei									
9											
10		või									
11											
12		H_0 : EMÜ 1. kursuse nei									
13		H_1 : EMÜ 1. kursuse nei									
14											
15		või									
16											
17		$H_0: \mu_T = 168$	$\mu_T - E$								
18		$H_1: \mu_T \neq 168$									
19											
20											
21		Funktsioon CONFIDENCE.NORM									
22			=CONFIDENCE.NORM(0,05;V4;V2)								

Function Arguments

CONFIDENCE.NORM

Alpha: 0,05 = 0,05

Standard_dev: V4 = 6,119628189

Size: V2 = 52

Formula result = 1,663303325

Returns the confidence interval for a population mean, using a normal distribution.

Alpha is the significance level used to compute the confidence level, a greater than 0 and less than 1.

Kuna tahame leida 95% usaldusintervalli, siis olulisuse nivoo $\alpha = 0,05$.

OK

Funktsioon CONFIDENCE.NORM

1,663303

b) funktsiooni CONFIDENCE.T abil

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

Tulemus:

Funktsioon CONFIDENCE.NORM	
1,663303	
Funktsioon CONFIDENCE.T	
1,703715	

5. Arvutage alumine ja ülemine usalduspiir mõlema tulemuse alusel.

Kumb 95%-usaldusintervallidest on laiem? Miks?

Vastus.
Funktsiooniga CONFIDENCE.T arvutatu 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 $t_{1-\alpha/2, n-1} = t_{0,975;51} = 2,01$ 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 *Excelis* 2010 leitav näiteks funktsiooniga =T.INV(0,975;51) 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 (168 cm)? **Sõnastage lõppjäreldus (koos põhjendusega) ja pange see kirja.**

Spikker. Et Eesti naiste keskmine pikkus 168 cm jääb EMÜ esimese kursuse neidude keskmise pikkuse 95% usaldusintervalli sisse, $168 \in (166,6; 170,0)$, siis ei ole alust lugeda tõestatuks alternatiivset hüpoteesi keskmise pikkuse erinevusest 168 sentimeetrist ja tuleb jääda nullhüpoteesi H_0 juurde: EMÜ esimese kursuse neidude keskmine pikkus ei erine 168 sentimeetrist.

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

...		fx		=V3+V25				
	T	U	V	W	X	Y	Z	AA
1		Neidude pikkus						
2		Vaatluste arv	52					
3		Keskmine	168,3269					
4		Standardhälve	6,119628					
5								
6								
7		H ₀ : EMÜ 1. kursuse neidude pikkused vastavad Eesti standardile (168 cm)						
8		H ₁ : EMÜ 1. kursuse neidude pikkused ei vasta Eesti standardile (168 cm)						
9								
10		või						
11								
12		H ₀ : EMÜ 1. kursuse neidude keskmine pikkus ei erine 168 sentimeetrist						
13		H ₁ : EMÜ 1. kursuse neidude keskmine pikkus erineb 168 sentimeetrist						
14								
15		või						
16								
17		H ₀ : $\mu_T = 168$	μ_T - EMÜ 1. kursuse neidude keskmine pikkus					
18		H ₁ : $\mu_T \neq 168$						
19								
20								
21		Funktsioon CONFIDENCE.NORM						
22			1,663303		Alumine usalduspiir	166,6636	=V3-V22	
23					Ülemine usalduspiir	169,9902		
24		Funktsioon CONFIDENCE.T						
25			1,703715		Alumine usalduspiir	166,6232	=V3-V25	
26					Ülemine usalduspiir	170,0306		
27								

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	N	O	P	Q	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	SUUSK	AUTO	OLU	SUITS		Neidude pikkus								MASS	AUTO
2	Eesti	N	168	62	53	jah	ei	0	ei		Vaatluste arv	52							62	ei
3	Eesti	N	178	75	56	ei	ei	0	jah		Keskmine	168,33							75	ei
4	Eesti	N	170,5	64	54	ei	ei	0	ei		Standardhälve	6,1196							64	ei
5	Eesti	N	158	56	50	jah	ei	0	ei										56	ei
6	Eesti	N	163	60	54	jah	ei	0	ei										60	ei
7	Eesti	N	177	70	53	jah	jah	0	ei		H ₀ : kursuse neidude pikkused vastavad Eesti standardile (168 cm)								70	jah
8	Eesti	N	170	51	52	jah	jah	0	ei		H ₁ : kursuse neidude pikkused ei vasta Eesti standardile (168 cm)								51	jah
9	Eesti	N	170	56		jah	jah	0	ei										56	jah
10	Soome	N	167	60	54	ei	jah	0	ei		või								60	jah
11	Soome	N	157	46	54	ei	ei	0	ei										46	ei
12	Soome	N	175	65	56	jah	ei	0	ei		H ₀ : kursuse neidude keskmine pikkus ei erine 168 sentimeetrist								65	ei
13	Soome	N	157	70	56	jah	ei	1	enam ei, agi		H ₁ : kursuse neidude keskmine pikkus erineb 168 sentimeetrist								70	ei
14	Soome	N	173	62	56	ei	ei	0	ei										62	ei
15	Soome	N	161	65	58	ei	jah	2	jah		või								65	jah
16	Eesti	N	169	68	55,5	jah	ei	0	ei										68	ei
17	Eesti	N	178	56	56	jah	jah	0	ei		H ₀ : $\mu_T = 168$	μ_T - kursuse neidude keskmine pikkus							56	jah
18	Eesti	N	164	62	55	ei	ei	0	ei		H ₁ : $\mu_T \neq 168$								62	ei
19	Eesti	N	159	55	54	jah	jah	0	ei										55	jah
20	Eesti	N	164	55	54,5	ei	ei	0	ei										55	ei
21	Eesti	N	174	74	58	jah	jah	0,5	ei		Funktsioon CONFIDENCE.NORM								74	jah
22	Eesti	N	171	67	66	ei	jah	0	ei		1,6633			Alumine usalduspiir	166,66	=V3-V22			67	jah
23	Soome	N	160	56	57	ei	ei	3	ei					Ülemine usalduspiir	169,99				56	ei
24	Soome	N	170	76	56	jah	jah	0	jah		Funktsioon CONFIDENCE.T								76	jah
25	Soome	N	165	55	56	jah	ei	0	ei		1,7037			Alumine usalduspiir	166,62	=V3-V25			55	ei
26	Eesti	N	168	55	57	jah	ei	0	ei					Ülemine usalduspiir	170,03				55	
27	Eesti	N	159	48	52	ei	ei	0	ei										48	ei
28	Soome	N	166	58	57	jah	jah	0	jah		2,0076								58	jah
29	Soome	N	169	59	58	ei	jah	0	ei										59	jah
30	Soome	N	160	57	58	jah	jah	0	ei										57	jah
31	Soome	N	178	70	53	ei	ei	0	ei										70	ei
32	Soome	N	169	68	58	jah	ei	0	ei										68	ei
33	Eesti	N	174	62	58	jah	ei	0	enam ei, agi										62	ei
34	Eesti	N	165	59	59	jah	jah	1	jah										59	jah
35	Eesti	N	177	76	55	jah	jah	0	ei										76	jah
36	Soome	N	170	80	59	ei	jah	0	ei										80	jah
37	Eesti	N	173	75	55	jah	ei	0	ei										75	ei
38	Eesti	N	175,5	66	55	jah	ei	0	ei										66	ei
39	Soome	N	172	66	56	jah	ei	0	jah										66	ei
40	Eesti	N	165	60	57	jah	ei	3	ei										60	ei
41	Eesti	N	170	60	58	jah	ei	7	ei										60	ei
42	Eesti	N	168	55	58	jah	ei	0	ei										55	ei
43	Soome	N	158	75	58	jah	ei	0	ei										75	ei
44	Eesti	N	170	54	53	jah	ei	0	ei										54	ei
45	Soome	N	181	72	57	jah	jah	0,5	ei										72	jah
46	Soome	N	167	70	51,5	jah	ei	0	ei										70	ei
47	Soome	N	166	49	55	jah	ei	0	ei										49	ei
48	Soome	N	174	51	55	jah	ei	0	ei										51	ei
49	Soome	N	160	60	57	jah	jah	0	ei										60	jah
50	Soome	N	170	65	59	jah	jah	0	ei										65	jah
51	Soome	N	170	66	58	jah	ei	0	ei										66	ei
52	Eesti	N	168	65	55	jah	jah	0	ei										65	jah
53	Eesti	N	172	62	53	ei	ei	0	ei										62	ei
54	Soome	M	183	90	60	jah	jah	4	jah										90	jah
55	Soome	M	187	75	55	jah	jah	3	ei										75	jah
56	Eesti	M	171	75	56	ei	ei	0	enam ei, agi										75	ei
57	Eesti	M	177	69	55	ei	ei	1	enam ei, agi										69	ei

Copy -> Paste
+
Sort ...

2. Leidke nii autot omavate ja mitte omavate tudengite arvud, keskmised kehamassid ja kehamassi standardhälbed (selle tudengi andmed, kes ei tea, kas tal on auto või ei ole, jätke analüüsist välja).

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

MASS	AUTO		Auto	
62	ei		Ei	Jah
75	ei	Tudengite arv	33	22
64	ei	Keskmine mass	62,3	65,7
56	ei	Standardhälve	8,07	9,84
60	ei			
46	ei			
65	ei			
70	ei			
62	ei			
68	ei			
62	ei			
55	ei			
56	ei			
55	ei			
48	ei			
70	ei			
68	ei			
62	ei			
75	ei			
66	ei			
66	ei			
60	ei			
60	ei			
55	ei			
75	ei			
54	ei			
70	ei			
49	ei			
51	ei			
66	ei			
62	ei			
75	ei			
69	ei			
70	jah			
51	jah			
56	jah			
60	jah			
65	jah			
56	jah			
55	jah			
74	jah			
67	jah			
76	jah			
58	jah			
59	jah			
57	jah			
59	jah			
76	jah			
80	jah			
72	jah			
60	jah			
65	jah			
65	jah			
90	jah			
75	jah			
55				

		Column Labels: T			
Values		ei	jah	Grand Total	
Count of MASS		33	22	55	
Average of MASS2		62,33333333	65,72727273	63,69090909	
StdDev of MASS3		8,068715305	9,83764305	8,890012555	

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

t-test	
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	
$H_0: \mu_{EI} = \mu_{Jah}$	μ_{EI} - autot mitte omavate tudengite keskmine kehamass
$H_1: \mu_{EI} \neq \mu_{Jah}$	μ_{Jah} - autot omavate tudengite keskmine kehamass

4. Millist t-testi nende keskmiste võrdlemisel kasutada?

NB! t-testi on 3 tüüpi, vt lk 11 (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.

t-test	
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	
$H_0: \mu_{EI} = \mu_{Jah}$	μ_{EI} - autot mitte omavate tudengite keskmine kehamass
$H_1: \mu_{EI} \neq \mu_{Jah}$	μ_{Jah} - autot omavate tudengite keskmine kehamass
Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsustamaks, millist t-testi kasutada.	
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)

The screenshot displays an Excel spreadsheet with data for an F-test. The data is organized into columns labeled AB through AL. The first column (AB) contains row numbers 1 through 34. The second column (AC) contains the word 'MASS'. The third column (AD) contains the word 'AUTO'. The fourth column (AE) contains the word 'jah'. The fifth column (AF) contains the word 'Ei'. The sixth column (AG) contains the word 'Jah'. The seventh column (AH) contains the word 'Grand Total'. The eighth column (AI) contains the word 'Count of MASS'. The ninth column (AJ) contains the word 'Average of MASS2'. The tenth column (AK) contains the word 'StdDev of MASS3'. The eleventh column (AL) contains the word 't-test'.

The 'Function Arguments' dialog box is open, showing the formula `=F.TEST(AC2:AC34;AC35:AC56)`. The dialog box includes the following information:

- Function:** F.TEST
- Array1:** AC2:AC34
- Array2:** AC35:AC56
- Result:** = 0,304537147
- Description:** Returns the result of an F-test, the two-tailed probability that the variances in Array1 and Array2 are not significantly different.
- Array1:** is the first array or range of data and can be numbers or names, arrays, or references that contain numbers (blanks are ignored).
- Formula result:** = 0,304537147
- Buttons:** OK, Cancel

6. Sõnastage järeldus F-testist, põhjendage.

See ongi põhjendus.
Aru saite?

F-test (võrdleme dispersioone)		Aru saite?	
$H_0: \sigma^2_{E1} = \sigma^2_{Jah}$	(autot oma)	vate tudengite kehamasside varieeruvus ei ole erinev)	
$H_1: \sigma^2_{E1} \neq \sigma^2_{Jah}$	(autot oma)	Otsustusreegel ja otsus	Järeldus s on erinev)
funktsioon F.TEST	0,304537147	= p > 0,05 => H ₀ : kehamasside varieeruvus võrreldavais gruppides ei ole erinev	
Kasutada võib t-testi, mis eeldab kehamasside võrdset varieeruvust võrreldavates gruppides.			

7. Viige läbi t-test võrdlemaks keskmisi kehamasse.

Tehke seda kahel viisil:

a) kasutades funktsiooni T.TEST:

Excel spreadsheet showing data for car mass (MASS) and type (Auto) and the results of a t-test.

Formula bar: `=T.TEST(AC2:AC34;AC35:AC56;2;2)`

AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
	MASS	AUTO				Auto				
1	62	ei				Ei	Jah			
2	75	ei								
3	64	ei		arv	33	22				
4	56	ei		Keskmine mass	62,3	65,7				
5	60	ei		Standardhälve	8,07	9,84				
6	46	ei								
7	65	ei								
8	70	ei								
9	62	ei								
10	68	ei								
11	62	ei								
12	55	ei								
13	56	ei								
14	55	ei								
15	48	ei								
16	70	ei								
17	68	ei								
18	62	ei								
19	75	ei								
20	66	ei								
21	66	ei								
22	60	ei								
23	60	ei								
24	55	ei								
25	75	ei								
26	54	ei								
27	70	ei								
28	49	ei								
29	51	ei								
30	66	ei								
31	62	ei								
32	75	ei								
33	69	ei								
34	70	jah								
35	51	jah								
36	56	jah								
37	60	jah								
38	65	jah								
39	56	jah								
40	55	jah								
41	74	jah								
42	67	jah								
43	76	jah								
44	58	jah								
45	59	jah								
46	57	jah								
47	59	jah								
48	76	jah								
49	80	jah								
50	72	jah								
51	60	jah								
52	65	jah								
53	65	jah								
54	90	jah								
55	75	jah								

Summary statistics:

Values	ei	jah	Grand Total
Count of MASS	33	22	55
Average of MASS2	62,33333333	65,72727273	63,69090909
StdDev of MASS3	8,068715305	9,83764305	8,890012555

t-test

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

$H_0: \mu_{ei} = \mu_{jah}$ μ_{ei} - autot mitte omavate tudengite keskmine kehamass
 $H_1: \mu_{ei} \neq \mu_{jah}$ μ_{jah} - autot omavate tudengite keskmine kehamass

Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsustamaks, millist t-testi kasutada.

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 0,304537147 = $p > 0,05 \Rightarrow H_0$: 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:AC34;AC35:AC56;2;2)

Function Arguments

T.TEST

Array1 AC2:AC34 = {62;75;64;56;60;46;65;70;62;68;62;55}

Array2 AC35:AC56 = {70;51;56;60;65;56;55;74;67;76;58...}

Tails 2

Type 2

Testime kahepoolset hüpoteesi

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 = 0,167552083

Help on this function

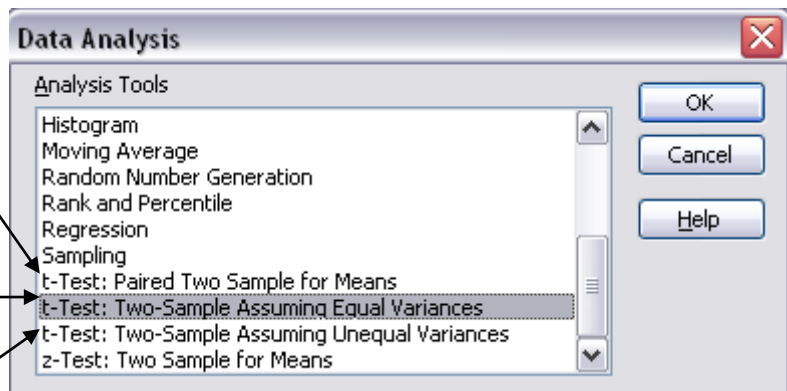
OK Cancel

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

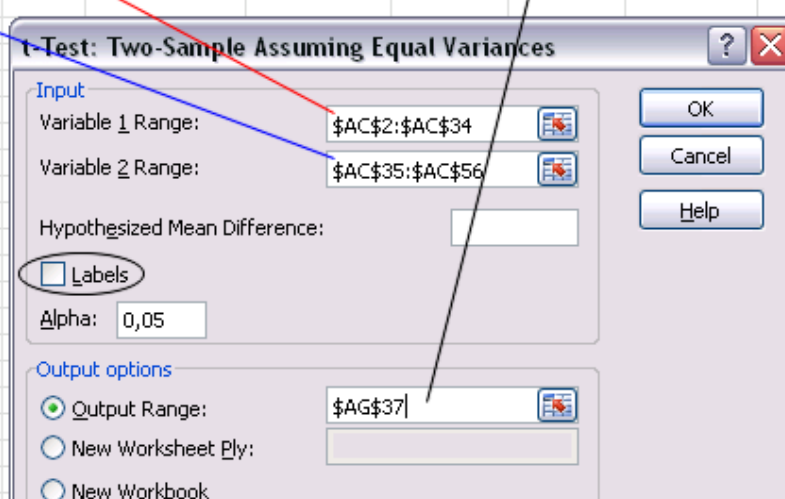
Sõltuvate gruppide (paariviisiline) võrdlus;
funktsioonis T.TEST tüüp nr 1

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		62	ei			Ei	Jah		
3		75	ei		Tudengite arv	33	22		
4		64	ei		Keskmine mass	62,3	65,7		
5		56	ei		Standardhälve	8,07	9,84		
6		60	ei						
7		46	ei						
8		65	ei						
9		70	ei						
10		62	ei		Values	Column Labels			
11		68	ei		ei	jah	Grand Total		
12		62	ei		Count of MASS	33	22	55	
13		55	ei		Average of MASS2	62,33333333	65,72727273	63,69090909	
14		56	ei		StdDev of MASS3	8,068715305	9,83764305	8,890012555	
15		55	ei						
16		48	ei		t-test				
17		70	ei		H ₀ : Autot omavate ja mitte omavate tudengite keskmised massid on võrdsed				
18		68	ei		H ₁ : Autot omavate ja mitte omavate tudengite keskmised massid on erinevad				
19		62	ei		või				
20		75	ei		H ₀ : $\mu_{ei} = \mu_{jah}$	μ_{ei} - autot mitte omavate tudengite keskmine kehamass			
21		66	ei		H ₁ : $\mu_{ei} \neq \mu_{jah}$	μ_{jah} - autot omavate tudengite keskmine kehamass			
22		66	ei						
23		60	ei		Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone				
24		60	ei						
25		55	ei		F-test (võrdleme dispersioone)				
26		75	ei		H ₀ : $\sigma^2_{ei} = \sigma^2_{jah}$	(autot omavate ja mitte omavate tudengite kehamass)			
27		54	ei		H ₁ : $\sigma^2_{ei} \neq \sigma^2_{jah}$	(autot omavate ja mitte omavate tudengite kehamass)			
28		70	ei						
29		49	ei		funktsioon F.TEST	0,304537147	= p > 0,05 => H ₀ : kehamasside varieer		
30		51	ei						
31		66	ei		Kasutada võib t-testi, mis eeldab kehamasside võrdset varieeruvust võrre				
32		62	ei						
33		75	ei		Võrdleme keskmisi				
34		69	ei						
35		70	jah		funktsioon T.TEST	0,167552083			
36		51	jah						
37		56	jah		protseduur t-test				
38		60	jah						
39		65	jah						
40		56	jah						
41		55	jah						
42		74	jah						
43		67	jah						
44		76	jah						
45		58	jah						
46		59	jah						
47		57	jah						
48		59	jah						
49		76	jah						
50		80	jah						
51		72	jah						
52		60	jah						
53		65	jah						
54		65	jah						
55		90	jah						
56		75	jah						



8. Sõnastage lõppjärelus, põhjendage.

funktsioon T.TEST	0,167552083 = $p > 0,05 \Rightarrow H_0$: autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad		
protseduur t-test	t-Test: Two-Sample Assuming Equal Variances		
		Variable 1	Variable 2
Mean	62,33333333	65,72727273	
Variance	65,10416667	96,77922078	
Observations	33	22	
Pooled Variance	77,65465981		
Hypothesized Mean Difference	0		
df	53		
t Stat	-1,399288573		
P(T<=t) one-tail	0,083776041		
t Critical one-tail	1,674116237		
P(T<=t) two-tail	0,167552083	= $p > 0,05 \Rightarrow H_0$: autot omavate ja mitte omavate tudengite keskmised massid ei ole erinevad	
t Critical two-tail	2,005745995		

Veel võimalusi lõppjärelduse sõnastamiseks:

- „autot omavate ja autot mitte omavate 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 $p = 0,168$ näitab, et

- väites, et kõigi autoomanikest tudengite (ka nende, kelle kohta meil andmeid ei ole) keskmine kehamass erineb kõigi autot mitte omavate tudengite keskmisest kehamassist, eksiksime 16,8%-lise tõenäosusega;
- eeldades, et autot omavate ja mitte omavate 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 0,168. 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	62,33333333	65,72727273
Variance	65,10416667	96,77922078
Observations	33	22

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

Pooled Variance	77,65465981	Võrdsete dispersioonide eeldusel arvutatud ühine dispersioon
Hypothesized Mean Difference	0	
df	53	t-statistiku empiiriline (andmete alusel arvutatud) väärtus
t Stat	-1,399288573	
P(T<=t) one-tail	0,083776041	1-poolsele hüpoteesile vastav p-väärtus
t Critical one-tail	1,674116237	1-poolsele hüpoteesile vastav t-statistiku kriitiline väärtus
P(T<=t) two-tail	0,167552083	2-poolsele hüpoteesile vastav p-väärtus
t Critical two-tail	2,005745995	2-poolsele hüpoteesile vastav t-statistiku kriitiline väärtus

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 keskmistele väärtustele, testides alati seda, kas suurem keskmine on ikka statistiliselt oluliselt suurem või mitte. Seega on antud juhul testitav ühepoolsete hüpoteeside paar kujul

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

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

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 $|t| = 1,40 < 1,67 = t_{critical}$, 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 lõppjäreldus testitud ühepoolse hüpoteesi kohta?