

Biomeetria praks 4

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

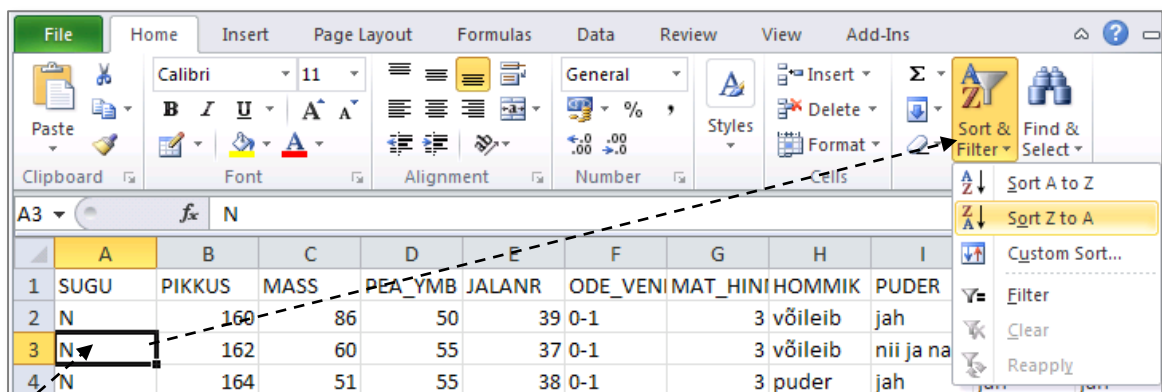
1. Avage MS Excel'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks4' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks4' ülemisse vasakusse nurka.

Ülesanne 1.

Kas teie kursuse neidude pikkused erinevad Eesti standardist (Eesti naiste keskmine pikkus on 168 cm)?

Tööjuhend

1. Sorteeri kogu andmetabel veeru 'SUGU' järgi.



1) Näiteks pannes kursori veergu 'SUGU' ja valides Home-sakilt käsud

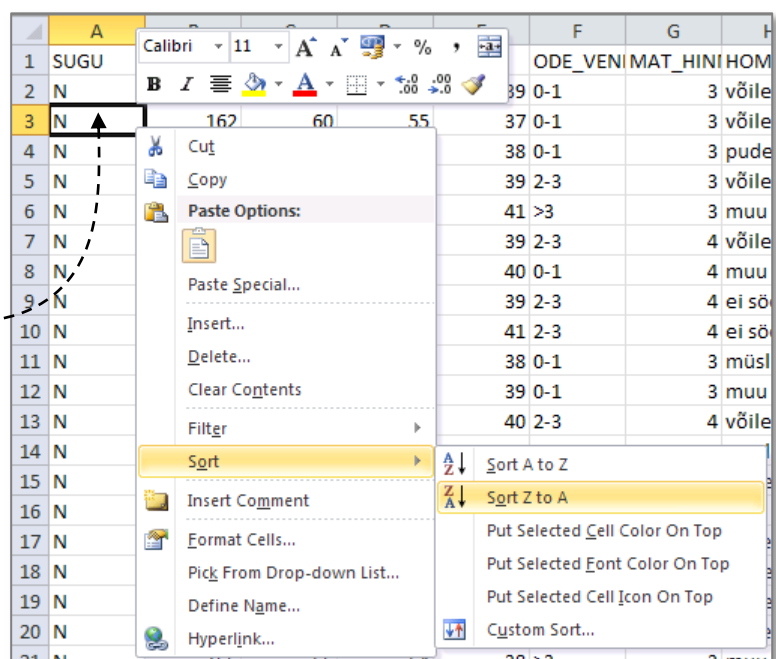
Sort & Filter ->

Sort Z to A

(sorteerimaks nende 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.

Seega on 42 teie kursuse neiu keskmine pikkus 167,8 cm standardhälbega 5,0 cm

- st, et keskmiselt erineb neidude tegelik pikkus 167,8 sentimeetrist 5,0 cm võrra;
- ehk, eeldades, et pikkus jaotub normaaljaotuse järgi, jääb ligikaudu 68,3% neidude pikkus vahemikku $167,8 \pm 5,0$ cm ($\bar{x} \pm s$) ja ligikaudu 95,5% neidude pikkus vahemikku $167,8 \pm 10,0$ cm ($\bar{x} \pm 2s$).

A	B	C	N	O	P	Q	R	S	T	U
1	SUGU	PIKKUS	MASS	OLU	SUIITS	KINDO	TEATER		Neidude pikkus	
2	N	160	86	ei joo	jah	viimase	irohkem kui aasta tagasi		Vaatluste arv	42
3	N	162	60	ei joo	ei	rohkem	irohkem kui aasta tagasi		Keskmine	167,762
4	N	164	51	ei joo	ei	viimase	irohkem kui aasta tagasi		Standardhälve	=STDEV.S(B2:B43)
5	N	165	68	ei joo	ei	viimase	irohkem kui aasta tagasi			
6	N	168	58	ei joo	enam ei	viimase	irohkem kui aasta tagasi			
7	N	168	68	üle 0,5 l	ei	viimase	irohkem kui aasta tagasi			
8	N	169	60	kuni 0,5 l	ei	rohkem	irohkem kui aasta tagasi			
9	N	169	65	ei joo	jah	viimase	irohkem kui aasta tagasi			
10	N	169	79	ei joo	jah	viimase	irohkem kui aasta tagasi			
11	N	170	62	üle 0,5 l	ei	viimase	irohkem kui aasta tagasi			
12	N	170	65	ei joo	ei	viimase	irohkem kui aasta tagasi			
13	N	170	70	üle 0,5 l	jah	rohkem	irohkem kui aasta tagasi			
14	N	172	65	ei joo	ei	rohkem	irohkem kui aasta tagasi			
15	N	153	47	ei joo	ei	viimase	viimase aasta jooksul			
16	N	159	57	üle 0,5 l	ei	viimase	viimase aasta jooksul			
17	N	160	52	ei joo	ei	viimase	viimase aasta jooksul			
18	N	162	49	ei joo	enam ei	viimase	viimase aasta jooksul			
19	N	163	55	ei joo	ei	viimase	viimase aasta jooksul			
20	N	165	70	ei joo	ei	viimase	viimase aasta jooksul			
21	N	166	56	kuni 0,5 l	ei	viimase	viimase aasta jooksul			
22	N	166	60	kuni 0,5 l	ei	viimase	viimase aasta jooksul			
23	N	169	64	ei joo	ei	viimase	viimase aasta jooksul			
24	N	169	65	kuni 0,5 l	ei	viimase	viimase aasta jooksul			
25	N	169	70	ei joo	jah	viimase	viimase aasta jooksul			
26	N	169	73	ei joo	ei	viimase	viimase aasta jooksul			
27	N	170	55	ei joo	ei	viimase	viimase aasta jooksul			
28	N	170	56	üle 0,5 l	jah	viimase	viimase aasta jooksul			
29	N	170	64	ei joo	ei	viimase	viimase aasta jooksul			
30	N	170	78	ei joo	ei	viimase	viimase aasta jooksul			
31	N	171	70	ei joo	ei	viimase	viimase aasta jooksul			
32	N	172	75	kuni 0,5 l	ei	viimase	viimase aasta jooksul			
33	N	174	64	ei joo	ei	viimase	viimase aasta jooksul			
34	N	175	80	kuni 0,5 l	enam ei	viimase	viimase aasta jooksul			
35	N	176	60	ei joo	ei	viimase	viimase aasta jooksul			
36	N	179	72	kuni 0,5 l	ei	viimase	viimase aasta jooksul			
37	N	160	56	ei joo	ei	viimase	viimase aasta jooksul			
38	N	171	57	ei joo	ei	viimase	viimase aasta jooksul			
39	N	165	81	ei joo	ei	viimase	viimase aasta jooksul			
40	N	167	60	üle 0,5 l	jah	viimase	viimase aasta jooksul			
41	N	169	50	üle 0,5 l	ei	viimase	viimase aasta jooksul			
42	N	169	63	kuni 0,5 l	jah	viimase	viimase aasta jooksul			
43	N	172	58	ei joo	ei	viimase	viimase aasta jooksul			
44	M	173	78	üle 0,5 l	enam ei	viimase	irohkem kui aasta tagasi			

Neidude pikkus	
Vaatluste arv	42
Keskmine	167,7619
Standardhälve	4,981965

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

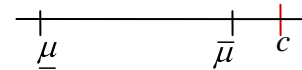
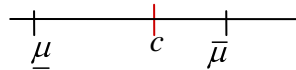
Neidude pikkus	
Vaatluste arv	42
Keskmine	167,7619
Standardhälve	4,981965
H ₀ : kursuse neidude pikkused vastavad Eesti standardile (168 cm)	
H ₁ : kursuse neidude pikkused ei vasta Eesti standardile (168 cm)	
või	
H ₀ : kursuse neidude keskmine pikkus ei erine 168 sentimeetrist	
H ₁ : kursuse neidude keskmine pikkus erineb 168 sentimeetrist	
või	
H ₀ : $\mu_T = 168$	μ_T - 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!

Excel spreadsheet showing the use of the CONFIDENCE.NORM function. The spreadsheet contains data for a population mean and standard deviation. The function arguments dialog box is open, showing the following values:

Argument	Value	Result
Alpha	0,05	= 0,05
Standard_dev	T4	= 4,981965152
Size	T2	= 42

The dialog box also displays the formula result: 1,506690782. A red dashed box highlights the text: "Kuna tahame leida 95% usaldusintervalli, siis olulisuse nivoo $\alpha = 0,05$." Below the dialog box, the formula bar shows the function call: `=CONFIDENCE.NORM(0,05;T4;T2)`. A result box shows the output: **Funktsioon CONFIDENCE.NORM** with the value **1,506691**.

b) funktsiooni CONFIDENCE.T abil

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

Tulemus:

Funktsioon CONFIDENCE.NORM	
	1,506691
Funktsioon CONFIDENCE.T	
	1,55249

5. Arvutage alumine ja ülemine usalduspiir mõlema tulemuse alusel.

	R	S	T	U	V	W	X	Y
1		Neidude pikkus						
2		Vaatluste arv	42					
3		Keskmine	167,7619					
4		Standardhälve	4,981965					
5								
6								
7		H ₀ : kursuse neidude pikkused vastavad Eesti standardile (168 cm)						
8		H ₁ : kursuse neidude pikkused ei vasta Eesti standardile (168 cm)						
9								
10		või						
11								
12		H ₀ : kursuse neidude keskmine pikkus ei erine 168 sentimeetrist						
13		H ₁ : kursuse neidude keskmine pikkus erineb 168 sentimeetrist						
14								
15		või						
16								
17		H ₀ : μ _T = 168	μ _T - kursuse neidude keskmine pikkus					
18		H ₁ : μ _T ≠ 168						
19								
20								
21		Funktsioon CONFIDENCE.NORM						
22			1,506691		Alumine usalduspiir	166,2552	=T3-T22	
23					Ülemine usalduspiir	169,2686		
24		Funktsioon CONFIDENCE.T						
25			1,55249		Alumine usalduspiir	166,2094	=T3-T25	
26					Ülemine usalduspiir	169,3144		
27								

Kumb 95%-usaldusintervallidest on laiem? Miks?

Vastus. Funktsiooniga CONFIDENCE.T arvatatu 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 $t_{1-\alpha/2, n-1} = t_{0,975; 41} = 2,02$ 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;41) ja teine funktsiooniga =NORM.S.INV(0,975).

Kui aru ei saanud, tutvuge 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 teie kursuse tütarlaste pikkused vastavad Eesti standardile (168 cm)? **Sõnastage lõppjärelus (koos põhjendusega) ja pange see kirja.**

Spikker. Et Eesti naiste keskmine pikkus 168 cm jääb teie kursuse neidude keskmise pikkuse 95% usaldusintervalli sisse, $168 \in (166,2; 169,3)$, siis ei ole alust lugeda tõestatuks alternatiivset hüpoteesi keskmise pikkuse erinevusest 168 sentimeetrist ja tuleb jääda nullhüpoteesi H_0 juurde: kursuse neidude keskmine pikkus ei erine 168 sentimeetrist.

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

Ülesanne 2.

Kas meeste ja naiste keskmised peaümberrõõdud on erinevad?

Tööjuhend

1. Moodustage samale töölehele abitabel veergudest 'PEAYMB' ja 'SUGU' ning sorteerige abitabel soo järgi.

	A	B	C	D	E	F	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	SUGU	PIKKUS	MASS	PEA_YMB	JALANR	KINO	TEATER		Neitude pikkus								SUGU	PEA_YMB
2	N	160	66	50	39	viimase aa rohkem kui aasta tagasi			Vaetluste arv	42							N	50
3	N	162	60	55	37	rohkem kui rohkem kui aasta tagasi			Keskmine	167,7619							N	55
4	N	164	51	55	38	viimase aa rohkem kui aasta tagasi			Standardhälve	4,981965							N	55
5	N	165	68	55	39	viimase 10 rohkem kui aasta tagasi											N	55
6	N	168	58	56	41	viimase aa rohkem kui aasta tagasi											N	56
7	N	168	68	52	39	viimase ku rohkem kui aasta tagasi			H ₀ : kursuse neitude pikkused vastavad Eesti standardile (168 cm)								N	52
8	N	169	60	50	40	rohkem kui rohkem kui aasta tagasi			H ₁ : kursuse neitude pikkused ei vasta Eesti standardile (168 cm)								N	50
9	N	169	65	55	39	viimase 10 rohkem kui aasta tagasi											N	55
10	N	169	79	55	41	viimase ku rohkem kui aasta tagasi			või								N	55
11	N	170	62	58	38	viimase ku rohkem kui aasta tagasi											N	58
12	N	170	65	53	39	viimase ku rohkem kui aasta tagasi			H ₀ : kursuse neitude keskmine pikkus ei erine 168 sentimeetrist								N	53
13	N	170	70	55	40	rohkem kui rohkem kui aasta tagasi			H ₁ : kursuse neitude keskmine pikkus erineb 168 sentimeetrist								N	55
14	N	172	65	53	40	rohkem kui rohkem kui aasta tagasi											N	53
15	N	173	47	53	37	viimase aa viimase aasta jooksul			või								N	53
16	N	179	57	54	39	viimase aa viimase aasta jooksul											N	54
17	N	180	52	55,5	38	viimase ku viimase aasta jooksul			H ₀ : µ = 168 µ _T - kursuse neitude keskmine pikkus								N	55,5
18	N	162	49	55	36	viimase 10 viimase aasta jooksul			H ₁ : µ ≠ 168								N	55
19	N	163	55	55	37	viimase aa viimase aasta jooksul											N	55
20	N	165	70	52	39	viimase aa viimase aasta jooksul											N	52
21	N	166	56	54	38	viimase 10 viimase aasta jooksul			Funktsioon CONFIDENCE.NORM								N	54
22	N	166	60	54	38	viimase 10 viimase aasta jooksul			1,306691		Alumine usalduspiir	166,2552 =T3-T22					N	54
23	N	169	64	56	40	viimase ku viimase aasta jooksul					Ülemine usalduspiir	169,2686					N	56
24	N	169	65	58,5	38	viimase ku viimase aasta jooksul			Funktsioon CONFIDENCE.T								N	58,5
25	N	169	70	55	40	viimase ku viimase aasta jooksul			1,53249		Alumine usalduspiir	166,2094 =T3-T25					N	55
26	N	169	73	56	38	viimase ku viimase aasta jooksul					Ülemine usalduspiir	169,3144					N	56
27	N	170	55	58	38	viimase 10 viimase aasta jooksul											N	58
28	N	170	56	55	39	viimase 10 viimase aasta jooksul											N	55
29	N	170	64	56	39	viimase aa viimase aasta jooksul											N	56
30	N	170	78	57	38	viimase aa viimase aasta jooksul											N	57
31	N	171	70	54	40	viimase aa viimase aasta jooksul											N	54
32	N	172	75	55	39	viimase 10 viimase aasta jooksul											N	55
33	N	174	64	56	39	viimase ku viimase aasta jooksul											N	56
34	N	175	80	52	41	viimase aa viimase aasta jooksul											N	52
35	N	176	60	55	39	viimase ku viimase aasta jooksul											N	55
36	N	179	72	57,5	41	viimase ku viimase aasta jooksul											N	57,5
37	N	180	56	53	38	viimase ku viimase kuu jooksul											N	53
38	N	171	57	54	37	viimase ku viimase kuu jooksul											N	54
39	N	165	81	55	41	viimase aa viimase kuu jooksul											N	55
40	N	167	60	52,5	39	viimase ku viimase kuu jooksul											N	52,5
41	N	169	50	54	39	viimase ku viimase kuu jooksul											N	54
42	N	169	63	53	39	viimase ku viimase kuu jooksul											N	53
43	N	172	58	57	38	viimase ku viimase kuu jooksul											N	57
44	M	173	78	49	44	viimase 10 rohkem kui aasta tagasi											M	49
45	M	175	62	56	43	viimase aa rohkem kui aasta tagasi											M	56
46	M	175	74	58	43	ei ole kunagi rohkem kui aasta tagasi											M	58
47	M	175	77	56	43	viimase ku rohkem kui aasta tagasi											M	56
48	M	178	110	58	42	rohkem kui rohkem kui aasta tagasi											M	58
49	M	180	80	56	43	viimase aa rohkem kui aasta tagasi											M	56
50	M	181	88	56	43	rohkem kui rohkem kui aasta tagasi											M	56
51	M	182	74	54	43	viimase aa rohkem kui aasta tagasi											M	54
52	M	184	63	57	44	viimase ku rohkem kui aasta tagasi											M	57
53	M	192	89	55,5	46	viimase ku rohkem kui aasta tagasi											M	55,5
54	M	193	92	52	45	viimase 10 rohkem kui aasta tagasi											M	52
55	M	195	83	56	46	viimase ku rohkem kui aasta tagasi											M	56
56	M	176	61	54	43	viimase aa viimase aasta jooksul											M	54
57	M	178	59	58	42	viimase aa viimase aasta jooksul											M	58
58	M	180	80	57	43	viimase ku viimase aasta jooksul											M	57
59	M	183	150	63	46	viimase ku viimase aasta jooksul											M	63
60	M	185	72	57	43	viimase 10 viimase aasta jooksul											M	57
61	M	185	77	55	44	viimase aa viimase aasta jooksul											M	55
62	M	187	98	57	43	viimase aa viimase aasta jooksul											M	57
63	M	190	85	58	46	viimase 10 viimase aasta jooksul											M	58
64	M	193	100	57,5	46	viimase aa viimase aasta jooksul											M	57,5
65	M	194	88	54	47	viimase 10 viimase aasta jooksul											M	54
66		165	54	54	37	viimase ku rohkem kui aasta tagasi												
67		173	80	57	44	rohkem kui rohkem kui aasta tagasi												
68		172	55	54	38	viimase ku viimase aasta jooksul												
69		168	54	56	37	viimase ku viimase kuu jooksul												

Copy -> Paste
+
Sort ...

2. Leidke nii mees- kui ka naistudengite arvud, keskmised peaümbermõõdud ja peaümbermõõdu standardhälbed.

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

SUGU	PEA_YMB		Sugu	
			N	M
N	50			
N	55	Tudengite arv	42	22
N	55	Keskmine	54,5	56,1
N	55	Standardhälve	1,95	2,67
N	56			
N	52			
N	50			
N	55			
N	53			
N	58			
N	53			
N	55			
N	53			
N	55			
N	53			
N	53			
N	54			
N	53,5			
N	55			
N	55			
N	52			
N	54			
N	54			
N	56			
N	58,5			
N	55			
N	56			
N	58			
N	55			
N	56			
N	57			
N	54			
N	55			
N	56			
N	52			
N	55			
N	57,5			
N	53			
N	54			
N	55			
N	52,5			
N	54			
N	53			
N	57			
M	49			
M	56			
M	58			
M	56			
M	58			
M	56			
M	56			
M	54			
M	57			
M	55,5			
M	52			
M	56			
M	54			
M	58			
M	57			
M	63			
M	57			
M	55			
M	57			
M	58			
M	57,5			
M	54			

	Column Labels			
Values	N	M	Grand Total	
Count of PEA_YMB	42	22	64	
Average of PEA_YMB2	54,5	56,1	55,1	
StdDev of PEA_YMB3	1,95	2,67	2,32	

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

t-test	
H_0 : Mees- ja naistudengite keskmised peaümberrõõdud on võrdsed.	
H_1 : Mees- ja naistudengite keskmised peaümberrõõdud on erinevad.	
või	
$H_0: \mu_M = \mu_N$	μ_M - meestudengite keskmine peaümberrõõd
$H_1: \mu_M \neq \mu_N$	μ_N - naistudengite keskmine peaümberrõõd

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 erinevat sugu tudengite peaümberrõõdude 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 : Mees- ja naistudengite keskmised peaümberrõõdud on võrdsed.	
H_1 : Mees- ja naistudengite keskmised peaümberrõõdud on erinevad.	
või	
$H_0: \mu_M = \mu_N$	μ_M - meestudengite keskmine peaümberrõõd
$H_1: \mu_M \neq \mu_N$	μ_N - naistudengite keskmine peaümberrõõd
<hr/>	
Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsustamaks, millist t-testi kasutada.	
F-test (võrdleme dispersioone)	
$H_0: \sigma_M^2 = \sigma_N^2$	(peaümberrõõdude varieeruvus mees- ja naistudengite hulgas ei ole erinev)
$H_1: \sigma_M^2 \neq \sigma_N^2$	(peaümberrõõdude varieeruvus mees- ja naistudengite hulgas on erinev)

Excel spreadsheet showing data for 'SUGU' and 'PEA_YMB' columns. The formula bar shows `=F.TEST(AB2:AB43;AB44:AB65)`. A 'Function Arguments' dialog box is open, showing the following details:

Column Labels	N	M	Grand Total
Count of PEA_YMB	42	22	64
Average of PEA_YMB2	54,5	56,1	55,1
StdDev of PEA_YMB3	1,95	2,67	2,32

t-test
 H_0 : Mees- ja naistudente keskmised peaümberrõõdud on võrdsed.
 H_1 : Mees- ja naistudente keskmised peaümberrõõdud on erinevad.
 või
 $H_0: \mu_M = \mu_N$ μ_M - meestudente keskmine peaümberrõõd
 $H_1: \mu_M \neq \mu_N$ μ_N - naistudente keskmine peaümberrõõd

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

F-test (võrdleme dispersioone)
 $H_0: \sigma_M^2 = \sigma_N^2$ (peaümberrõõdude varieeruvus mees- ja naistudente hulgas ei ole erinev)
 $H_1: \sigma_M^2 \neq \sigma_N^2$ (peaümberrõõdude varieeruvus mees- ja naistudente hulgas on erinev)

funktsioon F.TEST =F.TEST(AB2;AB43;AB44;AB65)

Function Arguments

F.TEST

Array1 AB2:AB43 = {50;55;55;55;56;52;50;55;53;58;53...}

Array2 AB44:AB65 = {49;56;58;56;58;56;56;54;57;55,5;...}

= 0,084097921

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, arrays, or references that contain numbers (blanks are ignored).

Formula result = 0,084097921

[Help on this function](#)

OK Cancel

6. Sõnastage järeldus F-testist, põhjendage.

See ongi põhjendus. Aru saite?

F-test (võrdleme dispersioone)		
$H_0: \sigma_M^2 = \sigma_N^2$	(peaümberrõõdude varieeruvus mees- ja naistudente hulgas ei ole erinev)	
$H_1: \sigma_M^2 \neq \sigma_N^2$	(peaümberrõõdude varieeruvus mees- ja naistudente hulgas on erinev)	
funktsioon F.TEST	0,084097921	= $p > 0,05 \Rightarrow H_0$: peaümberrõõdude varieeruvus võrreldavais gruppides ei ole erinev
Kasutada võib t-testi, mis eeldab peaümberrõõdude võrdset varieeruvust võrreldavates gruppides.		

Otsustusreegel ja otsus

Järeldus

7. Viige läbi t-test võrdlemaks keskmisi peaübermõõte.

Tehke seda kahel viisil:

a) kasutades funktsiooni T.TEST:

Excel spreadsheet showing data for a t-test. The formula bar shows `=T.TEST(AB2:AB43;AB44:AB65;2;2)`. The spreadsheet contains columns for gender (SUGU), PEA_YMB, and calculated statistics (N, M, Grand Total, etc.).

Statistical test results shown in the spreadsheet:

- t-test**
 - $H_0: \mu_1 = \mu_2$ (mees- ja naistudentide keskmised peaübermõõdud on võrdsed.)
 - $H_1: \mu_1 \neq \mu_2$ (mees- ja naistudentide keskmised peaübermõõdud on erinevad.)
- F-test (võrdleme dispersiooni)**
 - $H_0: \sigma_1^2 = \sigma_2^2$ (peaübermõõdude varieeruvus mees- ja naistudentide hulgas ei ole erinev)
 - $H_1: \sigma_1^2 \neq \sigma_2^2$ (peaübermõõdude varieeruvus mees- ja naistudentide hulgas on erinev)

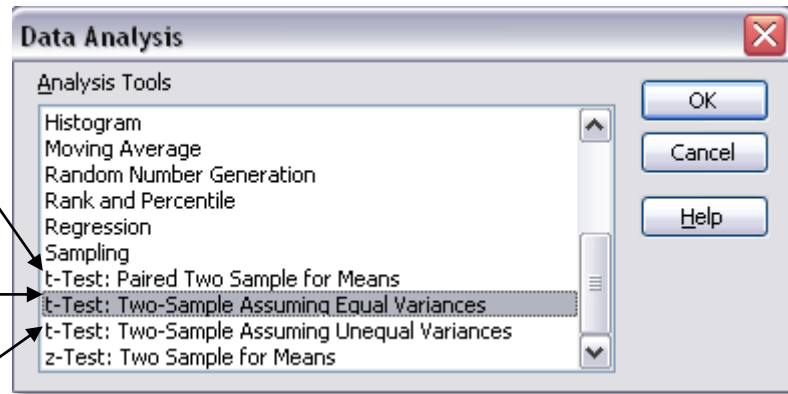
Function Arguments dialog box for T.TEST:

- Array1:** AB2:AB43 = {50;55;55;55;56;52;50;55;53;58;53;55}
- Array2:** AB44:AB65 = {49;56;58;56;58;56;56;54;57;55,5;...}
- Tails:** 2 (Testime kahepoolset hüpoteesi)
- Type:** 2 (Võrdse varieeruvuse eeldusel (miks võib seda eeldada?) sobiv t-testi tüüp)

Formula result = 0,009278651

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



	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1		SUGU	PEA_YMB			Sugu				
2	N		50			N	M			
3	N		55		Tudengite arv	42	22			
4	N		55		Keskmine	54,5	56,1			
5	N		55		Standardhälve	1,95	2,67			
6	N		56							
7	N		52							
8	N		50							
9	N		55							
10	N		53							
11	N		58		Values					
12	N		53		Count of PEA_YMB2	42	22	64		
13	N		55		Average of PEA_YMB2	54,5	56,1	55,1		
14	N		53		StdDev of PEA_YMB3	1,95	2,67	2,32		
15	N		53							
16	N		54							
17	N		53,5		t-test					
18	N		55		H ₀ : Mees- ja naistudengite keskmised peaübermõõdud on võrdsed.					
19	N		55		H ₁ : Mees- ja naistudengite keskmised peaübermõõdud on erinevad.					
20	N		52		või					
21	N		54		H ₀ : $\mu_M = \mu_N$	μ_M - meesstudengite keskmine peaübermõõt				
22	N		54		H ₁ : $\mu_M \neq \mu_N$	μ_N - naistudengite keskmine peaübermõõt				
23	N		56		Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsustamaks, millist t-testi kasutada.					
24	N		58,5							
25	N		55		F-test (võrdleme dispersioone)					
26	N		56		H ₀ : $\sigma_M^2 = \sigma_N^2$ (peaübermõõtude varieeruvus mees- ja naistudengite hulgas ei ole erinev)					
27	N		58		H ₁ : $\sigma_M^2 \neq \sigma_N^2$ (peaübermõõtude varieeruvus mees- ja naistudengite hulgas on erinev)					
28	N		55		funktsioon F.TEST	0,08409792	= p > 0,05 => H ₀ : peaübermõõtude varieeruvus võrreldavates gruppides.			
29	N		56							
30	N		57							
31	N		54							
32	N		55							
33	N		56		Võrdleme keskmisi					
34	N		52							
35	N		55		funktsioon T.TEST	0,009278651				
36	N		57,5		protseduur t-test					
37	N		53							
38	N		54							
39	N		55							
40	N		52,5							
41	N		54							
42	N		53							
43	N		57							
44	M		49							
45	M		56							
46	M		58							
47	M		56							
48	M		58							
49	M		56							
50	M		56							
51	M		54							
52	M		57							
53	M		55,5							
54	M		52							
55	M		56							
56	M		54							
57	M		58							
58	M		57							
59	M		63							
60	M		57							
61	M		55							
62	M		57							
63	M		58							
64	M		57,5							
65	M		54							

t-Test: Two-Sample Assuming Equal Variances

Input

Variable 1 Range: \$AB\$2:\$AB\$43

Variable 2 Range: \$AB\$44:\$AB\$65

Hypothesized Mean Difference: []

Labels

Alpha: 0,05

Output options

Output Range: \$AE\$37

New Worksheet Ply: []

New Workbook

Buttons: OK, Cancel, Help

8. Sõnastage lõppjärelus, põhjendage.

funktsioon T.TEST	0,009278651 = $p < 0,05 \Rightarrow H_1$: mees- ja naistudengite keskmised peaümberrõõdud on erinevad		
protseduur t-test	t-Test: Two-Sample Assuming Equal Variances		
		<i>Variable 1</i>	<i>Variable 2</i>
Mean	54,52380952	56,09090909	
Variance	3,792102207	7,11038961	
Observations	42	22	
Pooled Variance	4,916038263		
Hypothesized Mea	0		
df	62		
t Stat	-2,685560633		
P(T<=t) one-tail	0,004639326		
t Critical one-tail	1,669804163		
P(T<=t) two-tail	0,009278651 = $p < 0,05 \Rightarrow H_1$: mees- ja naistudengite keskmised peaümberrõõdud on erinevad		
t Critical two-tail	1,998971517		

Veel võimalusi lõppjäreluse sõnastamiseks:

- „**mees- ja naistudengite keskmised peaümberrõõdud on statistiliselt oluliselt erinevad ($p < 0,05$)**“,
so selline pisut korrektsemalt ja teaduslikumalt sõnastatud järelus;
- „peaümberrõõd sõltub tudengi soost“ (vähe teise nurga alt sõnastatud järelus, aga ka õige).

Märkus. Olulisuse tõenäosuse väärtus $p = 0,009$ näitab, et

- väites, et kõigi naistudengite (ka nende, kelle kohta meil andmeid ei ole) keskmine peaümberrõõd erineb kõigi meestudengite keskmisest peaümberrõõdust, eksiksime 0,9%-lise tõenäosusega;
- eeldades, et nais- ja meestudengite keskmised peaümberrõõdud (arvutatuna üle kõigi tudengite, st üldkogumis) on tegelikult võrdsed, siis tõenäosus saada nii suur peaümberrõõdude erinevus, nagu ilmnes meie andmetes lihtsalt tänu juhusele, on 0,009. Et see tõenäosus on väga väike (formaalselt $p < 0,05$), siis ei ole eeldus keskmiste peaümberrõõdude võrdumisest üldkogumis ilmselt korrektne.

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

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

	Naised	Mehed
	Variable 1	Variable 2
Mean	54,52380952	56,09090909
Variance	3,792102207	7,11038961
Observations	42	22

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

Pooled Variance	4,916038263	Võrdsete dispersioonide eeldusel arvatud ühine dispersioon
Hypothesized Mean Difference	0	
df	62	<i>t</i> -statistiku empiiriline (andmete alusel arvatud) väärtus
t Stat	-2,685560633	1-poolsele hüpoteesile vastav <i>p</i> -väärtus
P(T<=t) one-tail	0,004639326	1-poolsele hüpoteesile vastav <i>t</i> -statistiku kriitiline väärtus
t Critical one-tail	1,669804163	2-poolsele hüpoteesile vastav <i>p</i> -väärtus
P(T<=t) two-tail	0,009278651	2-poolsele hüpoteesile vastav <i>t</i> -statistiku kriitiline väärtus
t Critical two-tail	1,998971517	

Kahepoolne hüpotees tähendab statistikas võrdumise ja mittevõrdumise testimist:

$$H_0: \mu_M = \mu_N,$$

$$H_1: \mu_M \neq \mu_N.$$

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

$$H_0: \mu_M \leq \mu_N,$$

$$H_1: \mu_M > \mu_N,$$

sest meesstudengite keskmine peaümbermõõt on suurem.

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

Nimelt, kuna $|t| = 2,69 > 1,67 = t_{critical}$, siis võib lugeda tõestatuks alternatiivse hüpoteesi (*t*-statistiku väärtus ei jää sellesse piirkonda, kuhu ta nullhüpoteesi kehtides 95%-tõenäosusega peaks jääma).

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

Kuidas kõlab lõppjärelus testitud ühepoolse hüpoteesi kohta?