

## Biomeetria praks 4

### Illustreeritud (mittetäielik) tööjuhend

#### Eeltöö

1. Avage MS Excel'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht (Insert / Lisa -> Worksheet / 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 vastavad Eesti standardile (Eesti naiste keskmine pikkus on 165 cm)?

#### Tööjuhend

1. Sorteerige andmetabel veeru 'SUGU' järgi (Data / Andmed -> Sort / Sordi).

**NB! Sorteerida tuleb alati KOGU andmetabel!**

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

Seega on 44 teie kursuse neiu keskmine pikkus 168,5 cm standardhällbega 6,4 cm

- st, et keskmiselt erineb neidude tegelik pikkus 168,5 sentimeetrist 6,4 cm võrra;
- ehk, eeldades, et pikkus jaotub normaaljaotuse järgi, jääb ligikaudu 68,3% neidude pikkus vahemikku  $168,5 \pm 6,4$  cm ( $\bar{x} \pm s$ ) ja ligikaudu 95,5% neidude pikkus vahemikku  $168,5 \pm 12,7$  cm ( $\bar{x} \pm 2s$ ).

	A	B	C	P	Q	R	S	T	U	V	W
1	RIIK	SUGU	PIKKUS	OLU	SUITS	TEATER	KINO		Neidude pikkus		
2	Eesti	M	175	2	ei	viimase a	viimase kuu jooksul		Vaatluste arv	44	
3	Eesti	M	185	0,25	ei	viimase a	viimase aasta jooksul		Keskmine	168,489	
4	Eesti	N	167	0,5	ei	viimase a	viimase kuu jooksul		Standardhälve	=STDEV(C4:C47)	
5	Eesti	N	161,5	0	ei	viimase a	viimase kuu jooksul				
6	Soome	N	178	0	ei	viimase a	viimase 10 päeva jooksul				
7	Eesti	N	175	1	ei	rohkem ku	viimase aasta jooksul				
8	Eesti	N	170	0	ei	viimase ku	rohkem kui aasta tagasi				
9	Eesti	N	172	0	ei	viimase ku	viimase kuu jooksul				
10	Eesti	N	168	0	ei	viimase 1	viimase 10 päeva jooksul				
11	Eesti	N	175	0,5	enam ei	rohkem ku	viimase kuu jooksul				
12	Eesti	N	167	0	ei	rohkem ku	viimase kuu jooksul				
13	Eesti	N	165	0	ei	rohkem ku	viimase aasta jooksul				
14	Eesti	N	158	6	enam ei	viimase ku	viimase aasta jooksul				
15	Eesti	N	176	0	ei	rohkem ku	rohkem kui aasta tagasi				
16	Soome	N	164	0	ei	rohkem ku	viimase aasta jooksul				
17	Eesti	N	181	0	ei	viimase a	viimase kuu jooksul				
18	Eesti	N	177	0	ei	viimase ku	viimase kuu jooksul				
19	Eesti	N	175	0	ei	viimase a	viimase 10 päeva jooksul				
20	Eesti	N	170	1,5	jah	viimase a	viimase 10 päeva jooksul				
21	Eesti	N	163	0	ei	viimase a	viimase kuu jooksul				
22	Eesti	N	171	1,5	jah	viimase a	viimase kuu jooksul				
23	Eesti	N	160	5	jah	viimase a	viimase kuu jooksul				
24	Eesti	N	160	1	jah	viimase ku	viimase kuu jooksul				
25	Eesti	N	175	0	ei	rohkem ku	viimase aasta jooksul				
26	Eesti	N	165	0	jah	viimase a	viimase kuu jooksul				
27	Soome	N	173	0	ei	viimase a	viimase aasta jooksul				
28	Eesti	N	180	0	ei	rohkem ku	viimase kuu jooksul				
29	Eesti	N	161	0	enam ei	rohkem ku	viimase aasta jooksul				
30	Eesti	N	170	0	enam ei	viimase a	viimase 10 päeva jooksul				
31	Soome	N	170	4	jah	viimase a	viimase aasta jooksul				
32	Eesti	N	173	0	ei	viimase a	viimase aasta jooksul				
33	Eesti	N	161	1	ei	viimase a	viimase 10 päeva jooksul				
34	Eesti	N	164	0	ei	viimase a	viimase kuu jooksul				
35	Eesti	N	163	0,3	ei	rohkem ku	rohkem kui aasta tagasi				
36	Eesti	N	170	0	enam ei	viimase a	viimase 10 päeva jooksul				
37	Eesti	N	167	0,01	ei	viimase ku	viimase kuu jooksul				
38	Eesti	N	160	0	ei	viimase a	viimase 10 päeva jooksul				
39	Soome	N	172	0	ei	viimase 1	viimase kuu jooksul				
40	Eesti	N	174	3	ei	rohkem ku	viimase aa				
41	Soome	N	170	0	ei	viimase a	viimase aa				
42	Eesti	N	160	0	ei	rohkem ku	rohkem kui				
43	Eesti	N	174	1	ei	viimase ku	viimase 10				
44	Eesti	N	164	0,5	ei	viimase a	rohkem kui				
45	Eesti	N	159	0	ei	viimase a	viimase ku				
46	Eesti	N	175	0,5	ei	viimase a	viimase ku				
47	Eesti	N	160	0	ei	rohkem ku	viimase aa				
48			163	0	ei	rohkem ku	viimase aasta jooksul				

Neidude pikkus	
Vaatluste arv	44
Keskmine	168,4886
Standardhälve	6,360752

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

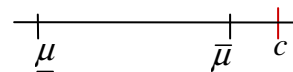
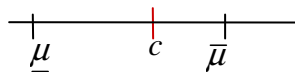
Neidude pikkus					
Vaatluste arv	44				
Keskmine	168,4886				
Standardhälve	6,360752				
H <sub>0</sub> : kursuse neidude pikkused vastavad Eesti standardile (165 cm)					
H <sub>1</sub> : kursuse neidude pikkused ei vasta Eesti standardile (165 cm)					
või					
H <sub>0</sub> : kursuse neidude keskmine pikkus ei erine 165 sentimeetrist					
H <sub>1</sub> : kursuse neidude keskmine pikkus erineb 165 sentimeetrist					
või					
H <sub>0</sub> : $\mu_T = 165$	$\mu_T$ - kursuse neidude keskmine pikkus				
H <sub>1</sub> : $\mu_T \neq 165$					

### 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 abil (sellele funktsioonile tuleb ette anda 3 argumenti: olulisuse nivoo  $\alpha$ , neidude pikkuste standardhälve ja neidude arv (kelle pikkus oli teada);

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!

**Function Arguments**

CONFIDENCE

Alpha 0,05 = 0,05

Standard\_dev V4 = 6,360752329

Size V2 = 44

Returns the confidence interval for a population mean.

Size is the sample size.

Formula result = 1,879447672

Help on this function

OK Cancel

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

	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1		<b>Neidude pikkus</b>									
2		Vaatluste arv	44								
3		Keskmine	168,4886								
4		Standardhälve	6,360752								
5											
6	ul										
7		H <sub>0</sub> : kursuse neidude pikkused vasta									
8	si	H <sub>1</sub> : kursuse neidude pikkused ei va									
9											
10	ul	või									
11											
12		H <sub>0</sub> : kursuse neidude keskmine pikk									
13		H <sub>1</sub> : kursuse neidude keskmine pikk									
14											
15	si	või									
16											
17		H <sub>0</sub> : $\mu_T = 165$	$\mu_T$ - kursuse ne								
18		H <sub>1</sub> : $\mu_T \neq 165$									
19	ul										
20	ul										
21		Funktsioon CONFIDENCE									
22		=CONFIDENCE(0,05;V4;V2)									
23											

Funktsioon CONFIDENCE

1,879448

**b) statistikaprotseduuri *Descriptive Statistics* abil**  
 (märkides üksnes lisavaliku *Confidence Level for Mean*).

*Tools / Tööriistad -> Data Analysis... -> Descriptive Statistics*

A	B	C	I	J	K	L	M	N	O	P	Q	R	S	T	U		
1	RIIK	SUGU	PIKKUS	HOMMIK	PUDER	LEMMIK	HAIGE	SPORT	SUUSK	AUTO	OLU	SUITS	TEATER	KINO	Neidude pikkus		
2	Eesti	M	175	võileib	jah	jah	ei	ei	ei	ei	2	ei	viimase a	viimase kuu	jooksul	Vaatluste arv	
3	Eesti	M	185	muu	jah	jah	ei	jah	jah	jah	0,25	ei	viimase a	viimase aasta	jooksul	Keskmine	
4	Eesti	N	167	ei söö	tav	ei	ei	jah	jah	ei	0,5	ei	viimase a	viimase kuu	jooksul	Standardhälve	
5	Eesti	N	161,5	puder	jah	jah	ei	jah	ei	ei	0	ei	viimase a	viimase kuu	jooksul		
6	Soome	N	178	puder	jah	jah	ei	jah	jah	ei	0	ei	viimase a	viimase 10 päeva	jooksul		
7	Eesti	N	175	helbed	võ	nii ja na	jah	ei	jah	jah	ei	1	ei	rohkem ku	viimase aasta	jooksul	H0: kursuse neidude p
8	Eesti	N	170	võileib	nii ja na	jah	ei	ei	ei	ei	0	ei	viimase k	rohkem kui	aasta tagasi	H1: kursuse neidude p	
9	Eesti	N	172	ei söö	tav	nii ja na	jah	ei	ei	ei	0	ei	viimase k	viimase kuu	jooksul		
10	Eesti	N	168	võileib	ei	jah	ei	jah	ei	ei	0	ei	viimase 1	viimase 10 päeva	jooksul	või	
11	Eesti	N	175	võileib	jah	jah	ei	ei	ei	ei	0,5	enam ei	rohkem ku	viimase kuu	jooksul		
12	Eesti	N	167	puder	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase kuu	jooksul	H0: kursuse neidude k	
13	Eesti	N	165	ei söö	tav	ei	jah	ei	jah	ei	0	ei	rohkem ku	viimase aasta	jooksul		
14	Eesti	N	158	ei söö	tav	nii ja na	jah	ei	jah	jah	ei	6	enam ei	viimase k	viimase aasta	jooksul	
15	Eesti	N	176	helbed	võ	jah	jah	ei	jah	ei	0	ei	rohkem ku	rohkem kui	aasta	jooksul	
16	Soome	N	164	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
17	Eesti	N	181	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
18	Eesti	N	177	võileib	nii ja na	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul	H1: $\mu_T \neq 165$	
19	Eesti	N	175	ei söö	tav	nii ja na	jah	ei	jah	ei	0	ei	rohkem ku	viimase aasta	jooksul		
20	Eesti	N	170	võileib	nii ja na	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
21	Eesti	N	163	helbed	võ	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul	Funktsioon CONFIDEI	
22	Eesti	N	171	puder	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul	1	
23	Eesti	N	160	võileib	nii ja na	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
24	Eesti	N	160	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul	Descriptive Statistics	
25	Eesti	N	175	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
26	Eesti	N	165	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
27	Soome	N	173	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
28	Eesti	N	180	puder	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
29	Eesti	N	161	puder	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
30	Eesti	N	170	muu	nii ja na	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
31	Soome	N	170	puder	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
32	Eesti	N	173	võileib	nii ja na	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
33	Eesti	N	161	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
34	Eesti	N	164	võileib	ei	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
35	Eesti	N	163	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
36	Eesti	N	170	helbed	võ	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
37	Eesti	N	167	puder	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
38	Eesti	N	160	muu	nii ja na	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
39	Soome	N	172	võileib	ei	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
40	Eesti	N	174	puder	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
41	Soome	N	170	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
42	Eesti	N	160	võileib	jah	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		
43	Eesti	N	174	muu	jah	jah	jah	jah	jah	jah	1	ei	viimase k	viimase 10 päeva	jooksul		
44	Eesti	N	164	puder	jah	jah	jah	jah	jah	jah	0,5	ei	viimase a	rohkem kui	aasta tagasi		
45	Eesti	N	159	võileib	ei	jah	jah	ei	jah	jah	0	ei	viimase a	viimase kuu	jooksul		
46	Eesti	N	175	ei söö	tav	ei	jah	jah	jah	jah	0,5	ei	viimase a	viimase kuu	jooksul		
47	Eesti	N	160	võileib	jah	jah	jah	jah	ei	jah	0	ei	rohkem ku	viimase aasta	jooksul		
48	Eesti	N	163	helbed	võ	ei	jah	jah	ei	ei	0	ei	rohkem ku	viimase aasta	jooksul		

**Descriptive Statistics**

Input  
 Input Range:

Grouped By:  
 Columns  
 Rows

Labels in first row

Output options  
 Output Range:    
 New Worksheet Ply:  
 New Workbook

Summary statistics  
 Confidence Level for Mean:  %

Kth Largest:   
 Kth Smallest:

OK Cancel Help

Ka selle vahepealkirja võiks ise trükkida, eristamaks erinevate meetoditega arvutatud usalduspiire.

Descriptive Statistics	
Column1	
Confidence Level(95,0%)	1,933845

## 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	168,4886					
4		Standardhälve	6,360752					
5								
6								
7		HO: kursuse neitude pikkused vastavad Eesti standardile (165 cm)						
8		H1: kursuse neitude pikkused ei vasta Eesti standardile (165 cm)						
9								
10		või						
11								
12		HO: kursuse neitude keskmine pikkus ei erine 165 sentimeetrist						
13		H1: kursuse neitude keskmine pikkus erineb 165 sentimeetrist						
14								
15		või						
16								
17		H <sub>0</sub> : μ <sub>T</sub> = 165	μ <sub>T</sub> - kursuse neitude keskmine pikkus					
18		H <sub>1</sub> : μ <sub>T</sub> ≠ 165						
19								
20								
21		Funktsioon CONFIDENCE						
22			1,879448		Alumine usalduspiir	166,6092 =V3-V22		
23					Ülemine usalduspiir	170,3681		
24								
25		Descriptive Statistics						
26		Column1			Alumine usalduspiir	166,5548 =V3-V28		
27					Ülemine usalduspiir	170,4225		
28		Confidence Level(95,0%)	1,933845					

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

Vastus. Protseduuriga *Descriptive Statistics* arvatatu on pisut laiem.

Põhjuseks see, et protseduur *Descriptive Statistics* arvutab usalduspiirid *t*-jaotuse baasil valemist

$$\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}},$$

funktsioon CONFIDENCE 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  $t_{1-\alpha/2, n-1} = t_{0,975,43} = 2,02$  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).

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

6. Otsus püstitatud hüpoteesi osas – kas teie kursuse tütarlaste pikkused vastavad Eesti standardile (165 cm)? **Sõnastage lõppjärelus (koos põhjendusega) ja pange see kirja.**

Spikker. Et Eesti naiste keskmine pikkus 165 cm ei jää teie kursuse neitude keskmise pikkuse 95% usaldusintervalli sisse, 165∉(166,6; 170,4), siis võib olulisuse nivool 0,05 lugeda tõestatuks alternatiivse hüpoteesi H<sub>1</sub>: kursuse neitude keskmine pikkus erineb 165 sentimeetrist.

**Mida tähendab väljend „olulisuse nivool 0,05“?**

Aga seda, et kui väita, et kursuse neitude keskmine pikkus ei vasta Eesti standardile (165 cm), on eksimise tõenäosus väiksem kui 5%.

## Ülesanne 2.

Kas viimase poolaasta jooksul viirushaiguseid põdenud tudengite keskmised pikkused on erinevad?

### Tööjuhend

1. Moodustage samale töölehele abitabel veergudest 'PIKKUS' ja 'HAIGE' ning sorteerige abitabel haigestumise järgi.

U	PIKKUS	MASS	LEMMIK	HAIGE	SPORT	SUUSK	AUTO	DLU	SUITS	TEATER	KINO	Neidude pikkus	PIKKUS	HAIGE	
170	64	jah	jah	jah	jah	jah	ei	0	enam	ei	viimase a	viimase 10 päeva jooksul	175	ei	
170	80	jah	ei	jah	jah	jah	1,5	jah	viimase a	viimase 10 päeva jooksul	Vaatuste arv	44	185	ei	
163	64	jah	ei	ei	jah	jah	0	ei	viimase a	viimase kuu jooksul	Keskmine	168,693	167	ei	
161,5	52	jah	ei	jah	ei	ei	0	ei	viimase a	viimase kuu jooksul	Standardhälve	6,90148	161,5	ei	
171	75	jah	ei	jah	jah	jah	1,5	jah	viimase a	viimase kuu jooksul			178	ei	
178	70	jah	ei	jah	jah	ei	0	ei	viimase a	viimase 10 päeva jooksul	H0: kursuse neidude pikkused vastavad Eesti standardile (165 cm)		175	ei	
175	67	jah	ei	jah	jah	ei	1	ei	rohkem ku	viimase aasta jooksul	H1: kursuse neidude pikkused ei vasta Eesti standardile (165 cm)		170	ei	
160	52	jah	ei	jah	ei	jah	5	jah	viimase a	viimase kuu jooksul			172	ei	
160	55	jah	ei	jah	ei	jah	1	jah	viimase ku	viimase kuu jooksul	või		168	ei	
170	69	jah	ei	ei	ei	ei	0	ei	viimase ku	rohkem kui aasta tagasi			175	ei	
167	57	ei	ei	jah	jah	ei	0,5	ei	viimase a	viimase kuu jooksul	H0: kursuse neidude keskmine pikkus ei erine 165 sentimeetrist		167	ei	
172	59	jah	ei	ei	ei	ei	0	ei	viimase ku	viimase kuu jooksul	H1: kursuse neidude keskmine pikkus erineb 165 sentimeetrist		165	ei	
167	80	jah	jah	jah	jah	ei	0,01	ei	viimase ku	viimase kuu jooksul			158	ei	
175	63	jah	ei	jah	ei	jah	0	ei	rohkem ku	viimase aasta jooksul	või		176	ei	
173	73	ei	jah	ei	jah	ei	0	ei	viimase a	viimase aasta jooksul			164	ei	
168	65	jah	ei	jah	ei	ei	0	ei	viimase 1	viimase 10 päeva jooksul	H0: $\mu_T = 165$	$\mu_T$ - kursuse neidude keskmine pikkus	181	ei	
170	80	jah	jah	jah	jah	jah	0	ei	viimase a	viimase aasta jooksul	H1: $\mu_T \neq 165$		177	ei	
165	70	jah	ei	jah	jah	jah	0	jah	viimase a	viimase kuu jooksul			175	ei	
170	60	jah	ei	jah	ei	ei	4	jah	viimase a	viimase aasta jooksul			170	ei	
175	60	jah	ei	ei	ei	ei	0,5	enam	ei	rohkem ku	viimase kuu jooksul	Funktsioon CONFIDENCE	163	ei	
160	50	jah	jah	jah	jah	ei	0	ei	viimase a	viimase 10 päeva jooksul	2,03922	Alumine usalduspiir	166,654	171	ei
160	51	jah	jah	ei	ei	jah	0	ei	rohkem ku	rohkem kui aasta tagasi		Ülemine usalduspiir	170,732	160	ei
167	55	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase kuu jooksul			160	ei	
161	54	ei	jah	jah	jah	ei	1	ei	viimase a	viimase 10 päeva jooksul	Descriptive Statistics		175	ei	
165	54,3	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta jooksul	Column 1	Alumine usalduspiir	166,759	165	ei
158	55	jah	ei	jah	jah	ei	6	enam	ei	viimase ku	viimase aasta jooksul	Ülemine usalduspiir	170,627	173	ei
173	62	jah	ei	jah	ei	jah	0	ei	viimase a	viimase aasta jooksul	Confidence Level(95,0%)	1,93385	180	ei	
174	58	jah	jah	jah	jah	jah	1	ei	viimase ku	viimase 10 päeva jooksul			161	ei	
175	76	jah	ei	ei	ei	ei	2	ei	viimase a	viimase kuu jooksul			170	ei	
180	65	jah	ei	ei	ei	jah	0	ei	rohkem ku	viimase kuu jooksul			170	ei	
176	64	jah	ei	jah	ei	ei	0	ei	rohkem ku	rohkem kui aasta tagasi			173	jah	
164	63	jah	ei	jah	ei	ei	0	ei	rohkem ku	viimase aasta jooksul			161	jah	
172	70	jah	jah	jah	ei	ei	0	ei	viimase 1	viimase kuu jooksul			164	jah	
161	56	jah	ei	jah	ei	jah	0	enam	ei	rohkem ku	viimase aasta jooksul			163	jah
181	60	jah	ei	jah	ei	ei	0	ei	viimase a	viimase kuu jooksul			170	jah	
164	55	jah	jah	jah	jah	jah	0,5	ei	viimase a	rohkem kui aasta tagasi			167	jah	
177	71	ei	jah	jah	ei	ei	0	ei	viimase ku	viimase kuu jooksul			160	jah	
175	70	jah	ei	jah	ei	ei	0	ei	viimase a	viimase 10 päeva jooksul			172	jah	
170	110	jah	ei	jah	jah	jah	0	enam	ei	viimase a	viimase 10 päeva jooksul			174	jah
164	58	ei	jah	jah	jah	ei	0	ei	viimase a	viimase kuu jooksul			170	jah	
163	50,5	ei	jah	jah	ei	jah	0,3	ei	rohkem ku	rohkem kui aasta tagasi			160	jah	
159	57	jah	jah	ei	jah	jah	0	ei	viimase a	viimase kuu jooksul			174	jah	
163	54	jah	jah	ei	ei	ei	0	ei	rohkem ku	viimase aasta jooksul			164	jah	
175	55	jah	jah	jah	jah	jah	0,5	ei	viimase a	viimase kuu jooksul			159	jah	
160	70	jah	jah	jah	ei	jah	0	ei	rohkem ku	viimase aasta jooksul			175	jah	
185	82	jah	ei	jah	jah	jah	0,25	ei	viimase a	viimase aasta jooksul			160	jah	
174	60	jah	jah	jah	ei	ei	3	ei	rohkem ku	viimase aasta jooksul			163	jah	

Kopeeri / Copy -> Kleebi / Paste  
+  
Andmed / Data -> Sordi ... / Sort ...

2. Leidke nii viirushaiguseid põdenud kui ka mitte põdenud tudengite arvud, keskmised pikkused ja pikkuste standardhälbed.

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

PIKKUS	HAIGE	Haigus		
		Ei	Jah	
175	ei			
185	ei	Tudengite arv	30	17
167	ei	Keskmine pikkus	170,3	166,4
161,5	ei	Standardhälve	6,9	5,8
178	ei			
175	ei			
170	ei			
172	ei			
168	ei			
175	ei			
167	ei			
165	ei			
158	ei			
176	ei			
164	ei			
181	ei			
177	ei			
175	ei			
170	ei			
163	ei			
171	ei			
160	ei			
160	ei			
175	ei			
165	ei			
173	ei			
180	ei			
161	ei			
170	ei			
170	ei			
173	jah			
161	jah			
164	jah			
163	jah			
170	jah			
167	jah			
160	jah			
172	jah			
174	jah			
170	jah			
160	jah			
174	jah			
164	jah			
159	jah			
175	jah			
160	jah			
163	jah			

		HAIGE			
Data		ei	jah		Grand Total
Count of PIKKUS		30	17		47
Average of PIKKUS2		170,25	166,4117647		168,8617021
StdDev of PIKKUS3		6,914190805	5,766918843		6,721721868

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

<b>t-test</b>	
$H_0$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on võrdsed.	
$H_1$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on erinevad.	
või	
$H_0: \mu_{ei} = \mu_{jah}$	$\mu_{ei}$ - viirushaiguseid mitte põdenud tudengite keskmine pikkus
$H_1: \mu_{ei} \neq \mu_{jah}$	$\mu_{jah}$ - viirushaiguseid põdenud tudengite keskmine pikkus

### 4. Millist t-testi nende keskmiste võrdlemisel kasutada?

**NB! t-testi on 3 tüüpi, vt biomeetria praktikumi nr 4 juhendit WebCT-is (ül. 2, punkt 4).**

- 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 erinevatesse gruppidesse kuuluvate tudengite pikkuste varieeruvuse võrdumise või mittevõrdumise üle (funktsioon FTEST).

NB! Protseduur *F-test (Tools / Tööriistad -> 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>	
$H_0$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on võrdsed.	
$H_1$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on erinevad.	
või	
$H_0: \mu_{ei} = \mu_{jah}$	$\mu_{ei}$ - viirushaiguseid mitte põdenud tudengite keskmine pikkus
$H_1: \mu_{ei} \neq \mu_{jah}$	$\mu_{jah}$ - viirushaiguseid põdenud tudengite keskmine pikkus
Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsustamaks, millist t-testi kasutada.	
<b><u>F-test (võrdleme dispersioone)</u></b>	
$H_0: \sigma_{ei}^2 = \sigma_{jah}^2$	(pikkuste varieeruvus viirushaiguseid põdenud ja mitte põdenud tudengite gruppides ei ole erinev)
$H_1: \sigma_{ei}^2 \neq \sigma_{jah}^2$	(pikkuste varieeruvus viirushaiguseid põdenud ja mitte põdenud tudengite gruppides on erinev)



PIKKUS	HAIGE	Haigus	
		Ei	Jah
175	ei		
185	ei		
167	ei		
161,5	ei		
178	ei		
175	ei		
170	ei		
172	ei		
168	ei		
175	ei		
167	ei		
165	ei		
158	ei		
176	ei		
164	ei		
181	ei		
177	ei		
175	ei		
170	ei		
163	ei		
171	ei		
160	ei		
160	ei		
175	ei		
165	ei		
173	ei		
180	ei		
161	ei		
170	ei		
170	ei		
173	jah		
161	jah		
164	jah		
163	jah		
170	jah		
167	jah		
160	jah		
172	jah		
174	jah		
170	jah		
160	jah		
174	jah		
164	jah		
159	jah		
175	jah		
160	jah		
163	jah		

	HAIGE			Grand Total
Data	ei	jah		
Count of PIKKUS	30	17		47
Average of PIKKUS2	170,25	166,4117647		168,8617021
StdDev of PIKKUS3	6,914190805	5,766918843		6,721721868

t-test

$H_0$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on võrdsed.

$H_1$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on erinevad.

või

$H_0: \mu_{ei} = \mu_{jah}$        $\mu_{ei}$  - viirushaiguseid mitte põdenud tudengite keskmine pikkus

$H_1: \mu_{ei} \neq \mu_{jah}$        $\mu_{jah}$  - viirushaiguseid põdenud tudengite keskmine pikkus

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

F-test (võrdleme dispersioone)

$H_0: \sigma_{ei}^2 = \sigma_{jah}^2$       (pikkuste varieeruvus viirushaiguseid põdenud ja mitte

$H_1: \sigma_{ei}^2 \neq \sigma_{jah}^2$       (pikkuste varieeruvus viirushaiguseid põdenud ja mitte

funktsioon FTEST =FTEST(AB2:AB31;AB32:AB48)

**Function Arguments**

FTEST

**Array1** AB2:AB31 = {175;185;167;161,5

**Array2** AB32:AB48 = {173;161;164;163;1

= 0,449650686

Returns the result of an F-test, the one-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 = 0,449650686

[Help on this function](#)

OK Cancel

6. Sõnastage järeldus F-testist, põhjendage.

See ongi põhjendus. Aru saite?

F-test (võrdleme dispersioone)	
$H_0: \sigma_{ei}^2 = \sigma_{jah}^2$	(pikkuste varieeruvus viirushaiguseid põdenud ja mitte põdenud tudengite gruppides ei ole erinev)
$H_1: \sigma_{ei}^2 \neq \sigma_{jah}^2$	(pikkuste varieeruvus viirushaiguseid põdenud ja mitte põdenud tudengite gruppides on erinev)
funktsioon FTEST	0,449650686 = $p > 0,05 \Rightarrow H_0$ : pikkuste varieeruvus võrreldavais gruppides ei ole erinev
Kasutada võib t-testi, mis eeldab pikkuste võrdset varieeruvust võrreldavates gruppides.	

7. Viige läbi t-test võrdlemaks keskmisi pikkusi.

Tehke seda kahel viisil:

a) kasutades funktsiooni TTEST:

**=TTEST(AB2:AB31;AB32:AB48;2;2)**

AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
PIKKUS	HAIGE			Haigus					
175	ei			Ei	Jah				
185	ei		Tudengite arv	30	17				
167	ei		Keskmine pikkus	170,3	166,4				
161,5	ei		Standardhälve	6,9	5,8				
178	ei								
175	ei								
170	ei								
172	ei								
168	ei								
175	ei		Data	ei	jah	Grand Total			
167	ei		Count of PIKKUS	30	17	47			
165	ei		Average of PIKKUS2	170,25	166,4117647	168,8617021			
158	ei		StdDev of PIKKUS3	6,914190805	5,766918843	6,721721868			
176	ei								
164	ei		t-test						
181	ei		H <sub>0</sub> : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on võrdsed.						
177	ei		H <sub>1</sub> : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on erinevad.						
175	ei		või						
170	ei		H <sub>0</sub> : $\mu_{ei} = \mu_{jah}$	$\mu_{ei}$ - viirushaiguseid mitte põdenud tudengite keskmine pikkus					
163	ei		H <sub>1</sub> : $\mu_{ei} \neq \mu_{jah}$	$\mu_{jah}$ - viirushaiguseid põdenud tudengite keskmine pikkus					
171	ei								
160	ei		Tegu on sõltumatute vaatlustega. Seega tuleb enne keskmiste võrdlemist võrrelda dispersioone, otsust.						
160	ei								
175	ei		F-test (võrdleme dispersioone)						
165	ei		H <sub>0</sub> : $\sigma_{ei}^2 = \sigma_{jah}^2$	(pikkuste varieeruvus viirushaiguseid põdenud ja mitte põden					
173	ei		H <sub>1</sub> : $\sigma_{ei}^2 \neq \sigma_{jah}^2$	(pikkuste varieeruvus viirushaiguseid põdenud ja mitte põden					
180	ei								
161	ei		funktsioon FTEST	0,449650686	= p > 0,05 => H <sub>0</sub> : pikkuste varieeruvus võrrel				
170	ei								
170	ei		Kasutada võib t-testi, mis eeldab pikkuste võrdset varieeruvust võrreldavates grup						
173	jah								
161	jah		Võrdleme keskmisi						
164	jah								
163	jah		funktsioon TTEST	=TTEST(AB2:AB31;AB32:AB48;2;2)					
170	jah								
167	jah								
160	jah								
172	jah								
174	jah								
170	jah								
160	jah								
167	jah								
160	jah								
172	jah								
174	jah								
164	jah								
159	jah								
175	jah								
160	jah								
163	jah								

**Function Arguments**

TTEST

**Array1** AB2:AB31 = {175;185;167;161,5}

**Array2** AB32:AB48 = {173;161;164;163;1}

**Tails** 2 [Testime kahepoolset hüpoteesi]

**Type** 2 [Võrdse varieeruvuse eeldusel (miks võib seda eeldada?) sobiv t-testi tüüp.]

Returns the probability associated with the t-test.

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

Formula result = 0,059114685

[Help on this function](#) [OK] [Cancel]

**b) vastava statistikaprotseduuri abil (Tools / Tööriistad -> Data Analysis... -> t-Test: ...):**

Sõltuvate gruppide (paariviisiline) võrdlus; funktsioonis TTEST tüüp nr 1

Sõltumatute gruppide võrdlus võrdsete dispersioonide eeldusel; funktsioonis TTEST tüüp nr 2

Sõltumatute gruppide võrdlus mittevõrdsete dispersioonide eeldusel; funktsioonis TTEST tüüp nr 3

PIKKUS	HAIGE		
175	ei		
185	ei		
167	ei		
161,5	ei		
178	ei		
175	ei		
170	ei		
172	ei		
168	ei		
175	ei		
167	ei		
165	ei		
158	ei		
178	ei		
164	ei		
181	ei		
177	ei		
175	ei		
170	ei		
163	ei		
171	ei		
160	ei		
160	ei		
175	ei		
165	ei		
173	ei		
180	ei		
161	ei		
170	ei		
170	ei		
173	jah		
161	jah		
164	jah		
163	jah		
170	jah		
167	jah		
160	jah		
172	jah		
174	jah		
170	jah		
160	jah		
174	jah		
164	jah		
159	jah		
175	jah		
160	jah		
163	jah		

**t-test**

$H_0$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on võrdsed.

$H_1$ : Viirushaiguseid põdenud ja mitte põdenud tudengite keskmised pikkused on erinevad.

või

$H_0: \mu_{ei} = \mu_{jah}$       $\mu_{ei}$

$H_1: \mu_{ei} \neq \mu_{jah}$       $\mu_{jah}$

Tegu on sõltumatute vaatlus

F-t

$H_0$ :

$H_1$ :

fun

Ka

**Võrdleme keskmisi**

funktsioon TTEST     0,059114685

protseduur t-test

## 8. Sõnastage lõppjäreltus, põhjendage.

funktsioon TTEST	0,059114685 = $p > 0,05 \Rightarrow H_0$ : viirushaiguseid põdenud ja mitte põdenud tudengite		
	keskmised pikkused ei ole erinevad		
protseduur t-test	t-Test: Two-Sample Assuming Equal Variances		
		<i>Variable 1</i>	<i>Variable 2</i>
	Mean	170,25	166,4117647
	Variance	47,80603448	33,25735294
	Observations	30	17
	Pooled Variance	42,63316993	
	Hypothesized Mean	0	
	df	45	
	t Stat	1,936394949	
	P(T<=t) one-tail	0,029557343	
	t Critical one-tail	1,679427393	
	P(T<=t) two-tail	0,059114685 = $p > 0,05 \Rightarrow H_0$ : viirushaiguseid põdenud ja mitte põdenud	
	t Critical two-tail	2,014103359	tudengite keskmised pikkused ei ole erinevad

Veel võimalusi lõppjäreltuse sõnastamiseks:

- „viirushaiguseid põdenud ja mitte põdenud tudengite keskmiste pikkuste vaheline erinevus ei ole statistiliselt oluline ( $p > 0,05$ )“, so sihuke teaduslikumalt sõnastatud järeltus;
- „viirustega nakatumine ei sõltu tudengi pikkusest“ (vähe teise nurga alt sõnastatud järeltus, aga kah õige).

Märkus. Kui nüüd tulemusi veelkord vaadata, siis ilmneb, et viirushaiguseid mitte põdenud tudengid on peaaegu neli sentimeetrit pikemad võrreldes viirushaiguseid põdenud tudengitega. Tõsi ta on. Ainult et tulemuse üldistamiseks (et näiteks soovitada teie kursuse neidudel viirushaigustest hoidumiseks kõrgema kotsaga kingi kanda) on andmeid natuke vähe (või on varieeruvus liiga suur, et erinevust selgelt tuvastada; või on kolme-nelja-sentimeetrine erinevus lihtsalt liiga väike).

Seda, et tuvastatud erinevus eksisteerib vaid analüüsitud valimis ega pruugi esineda üldkogumis, näitabki olulisuse tõenäosuse  $p$  väärtus 0,06. Seega väites, et kõik viirushaiguseid mitte põdenud tudengid (ka need, kelle kohta meil andmeid ei ole) on keskmiselt pikemad, kui viirushaiguseid põdenud tudengid, eksiksime me 6%-lise tõenäosusega.

Märkus2. Omaette küsimus on muidugi püstitatud hüpoteesi sisuline mõttekus reaalse uuringu jaoks ...

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

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

	<b>Ei</b>	<b>Jah</b>
	<i>Variable 1</i>	<i>Variable 2</i>
Mean	170,25	166,4117647
Variance	47,80603448	33,25735294
Observations	30	17

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

Pooled Variance	42,63316993	← Võrdsete dispersioonide eeldusel arvatud ühine dispersioon
Hypothesized Mean	0	
df	45	← <i>t</i> -statistiku empiiriline väärtus
t Stat	1,936394949	
P(T<=t) one-tail	0,029557343	← 1-poolsele hüpoteesile vastav <i>p</i> -väärtus
t Critical one-tail	1,679427393	← 1-poolsele hüpoteesile vastav <i>t</i> -statistiku kriitiline väärtus
P(T<=t) two-tail	0,059114685	← 2-poolsele hüpoteesile vastav <i>p</i> -väärtus
t Critical two-tail	2,014103359	← 2-poolsele hüpoteesile vastav <i>t</i> -statistiku kriitiline väärtus

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.

Nimelt, kuna  $|t| = 1,94 < 2,01 = t_{critical}$ , siis ei ole põhjust nullhüpoteesi ümber lükata

(*t*-statistiku väärtus jääb sellesse piirkonda, kuhu ta nullhüpoteesi kehtides 95%-tõenäosusega peabki jääma).

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

Mida see ühepoolne hüpotees õieti väidab (kuidas on sõnastatud)?