

## Biomeetria praks 7

### Illustreeritud (mittetäielik) tööjuhend

#### Eeltöö

1. Avage *MS Excel*'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks7' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks7' ülemisse vasakusse nurka.

---

#### Ülesanne.

##### Kas õdede-vendade arv ja auto omamine on seotud?

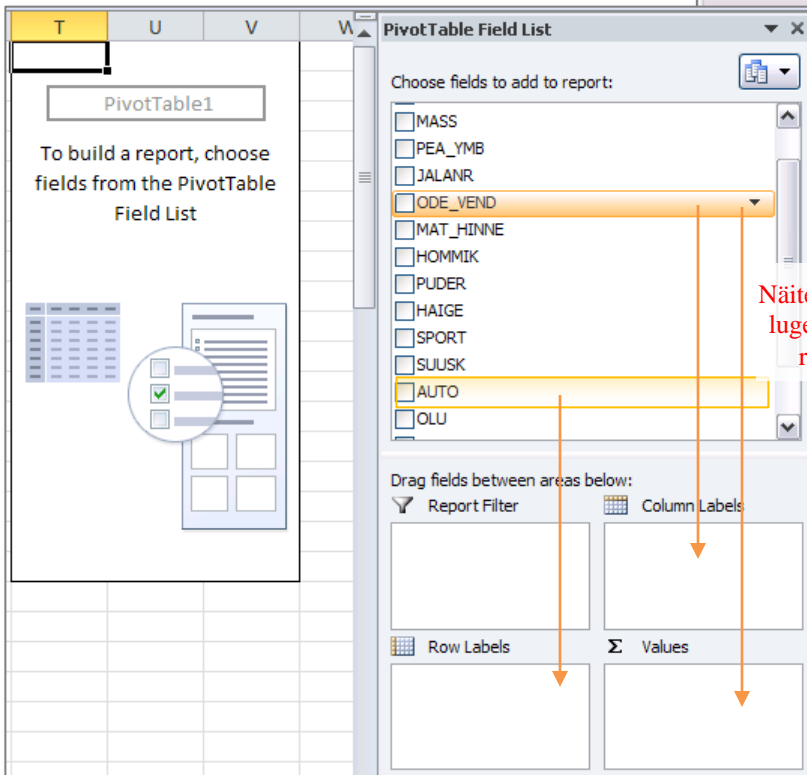
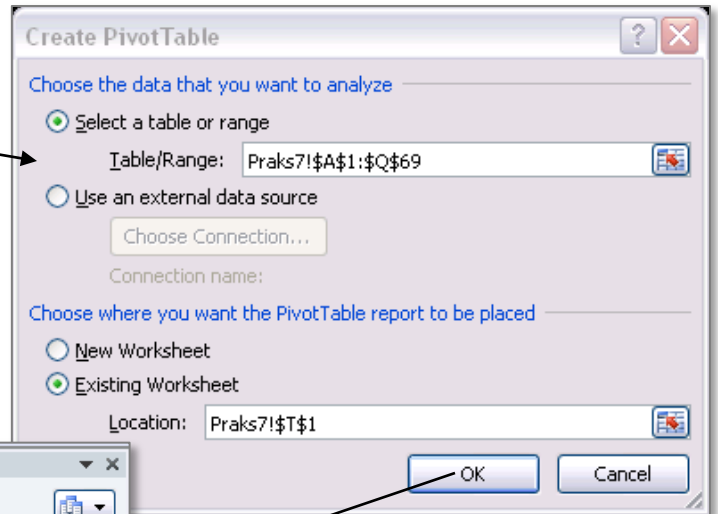
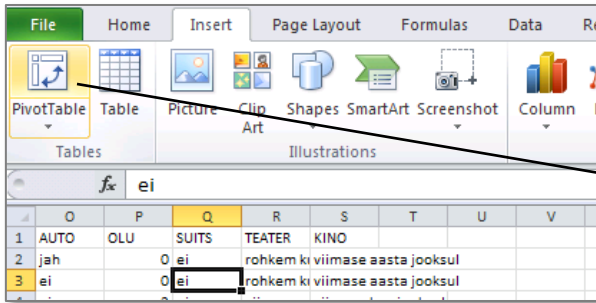
1. Konstrueerige kahemõõtmeline sagedustabel tunnustele 'ODE\_VEND' ja 'AUTO'.
2. Leidke sellesse tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii veeru- kui ka reaprotsendid).
3. Kommenteerige tabelit (sõnastades laused vähemalt kahe arvutatud protsendi kohta)!
4. Testimaks tuvastatud seose/erinevuse statistilist olulisust, pange kirja vastav hüpoteeside paar.
5. Tehke uus 2-mõõtmeline sagedustabel tunnustele 'ODE\_VEND' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi;
6. arvutage viimase alusel nullhüpoteesile (seost ei ole) vastavad sagedused ja
7. viige läbi hii-ruut-test (leidke  $p$ -väärtus).
8. Sõnastage lõppjärelendus (kas seos on statistiliselt oluline, mille alusel te seda otsustasite ja milles see seos ikkagi seisneb?).

## Tööjuhend

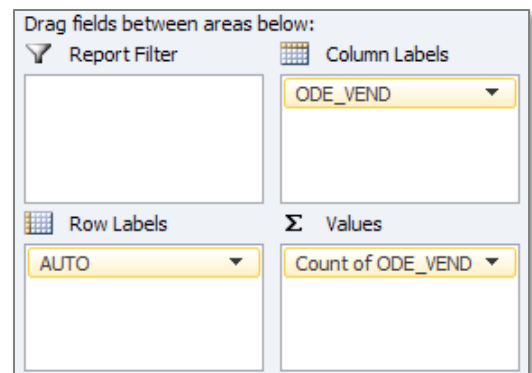
1. Konstrueerige kahemõõtmeline sagedustabel tunnustele ' ODE\_VEND' ja 'AUTO'.

- Paigutage kursor andmetabeli suvalisse lahtrisse (leheküljel 'Praks 7').
- *Insert*-sakk → *PivotTable*

Loodav tabel paigutage samale 'Praks7'-lehele:



Näiteks. Samas võib lasta *Excelil* kokku lugeda ka mõne teise **kõigil** tudengitel registreeritud tunnuse väärtused.



Count of ODE_VEND	Column Labels	>3	0-1	2-3	(blank)	Grand Total
Row Labels						
ei		2	18	14		34
jah		8	13	9		30
(blank)			1	2		3
<b>Grand Total</b>		<b>10</b>	<b>32</b>	<b>25</b>		<b>67</b>

- Kaotage konstrueeritud tabelist puuduvatele väärtustele (blank) vastavad read ja veerud.

	Count of ODE_VEND	Column Labels				
Row Labels	>3	0-1	2-3	(blank)	Grand Total	
ei	2	18	14		34	
jah	8	13	9		30	
Grand Total	10	32	25		67	

Analoogselt ka õdede-vendade arvuga

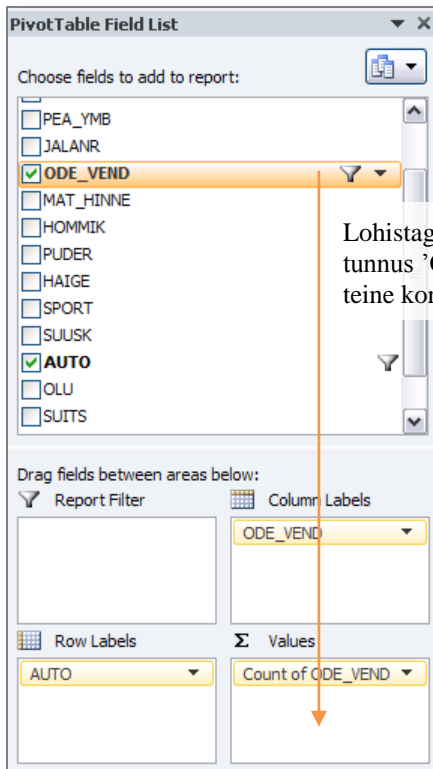
- Viimaks tõstke õdede-vendade arv '>3' viimasele kohale – siis on tabeli veerud loogilises järjestuses.

Count of ODE_VEND	Column Labels				
Row Labels	>3	0-1	2-3	Grand Total	
ei	2	18	14	34	
jah	8	13	9	30	
Grand Total	10	31	23	64	

Tulemus:

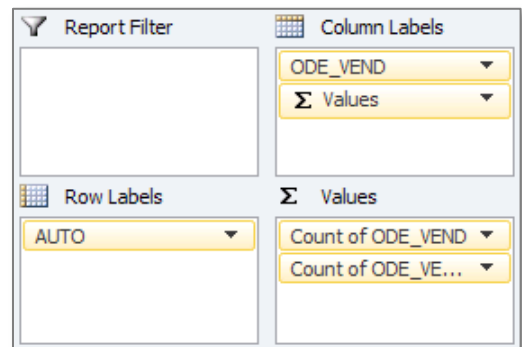
Count of ODE_VEND	Column Labels				
Row Labels	0-1	2-3	>3	Grand Total	
ei	18	14	2	34	
jah	13	9	8	30	
Grand Total	31	23	10	64	

- 2. Leidke tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii rea- kui ka veeruprotsendid).



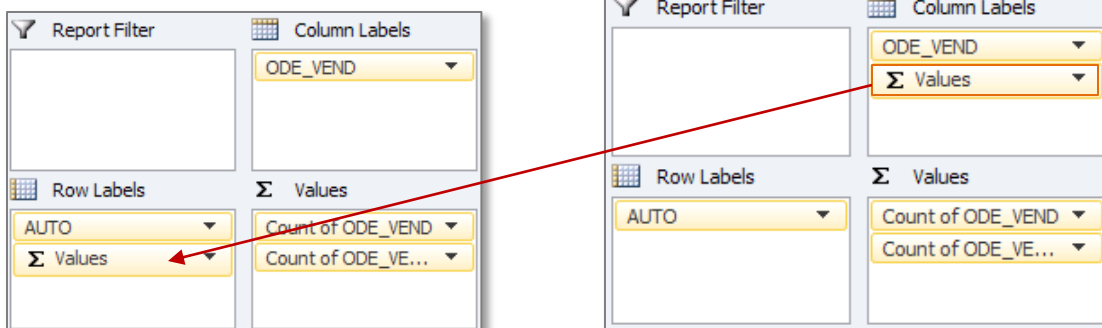
Lohistage tunnus 'ODE\_VEND' ka teine kord väärtuste lahtrisse.

Tulemus:



Row Labels	Count of ODE_VEND	Count of ODE_VEND2	Count of ODE_VEND	Count of ODE_VEND2	Count of ODE_VEND	Count of ODE_VEND2	Total Count of ODE_VEND	Total Count of ODE_VEND2
ei	18	18	14	14	2	2	34	34
jah	13	13	9	9	8	8	30	30
<b>Grand Total</b>	<b>31</b>	<b>31</b>	<b>23</b>	<b>23</b>	<b>10</b>	<b>10</b>	<b>64</b>	<b>64</b>

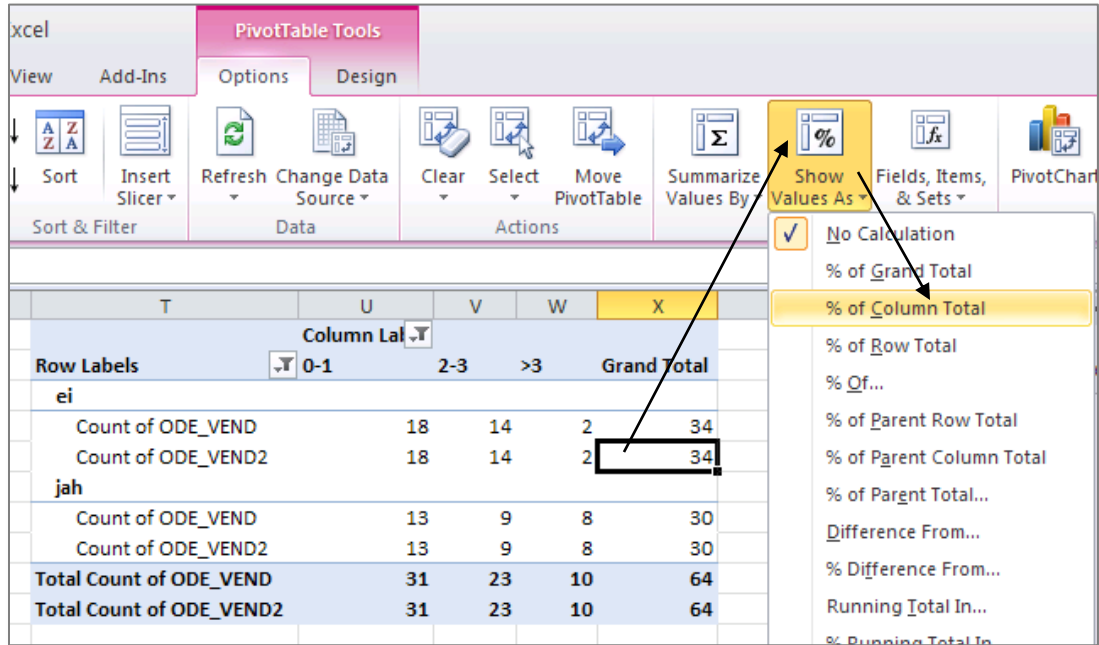
- Tabeli selgema esituse huvides jagage tabel arvutatavate väärtuste järgi ridadeks, mitte veergudeks (nagu Excel vaikimisi teeb):



Tulemus:

Row Labels	Count of ODE_VEND	Count of ODE_VEND2	Total Count of ODE_VEND	Total Count of ODE_VEND2
ei	18	14	2	34
jah	13	9	8	30
<b>Total Count of ODE_VEND</b>	<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>
<b>Total Count of ODE_VEND2</b>	<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>

- Esitamaks *Exceli* poolt uuesti kokku loetud tudengite arvu veeruprotsendina:



Tulemus:

Row Labels	0-1	2-3	>3	Grand Total	
<b>ei</b>					
Count of ODE_VEND		18	14	2	34
Count of ODE_VEND2		18	14	2	34
<b>jah</b>					
Count of ODE_VEND		13	9	8	30
Count of ODE_VEND2		13	9	8	30
<b>Total Count of ODE_VEND</b>		<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>
<b>Total Count of ODE_VEND2</b>		<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>

- Arvutage analoogselt ka reaprotsendid ... ( % of Row Total ).

(alustage sarnaselt punktis 2 kirjeldatule)

Row Labels	0-1	2-3	>3	Grand Total	
<b>ei</b>					
Count of ODE_VEND		18	14	2	34
Count of ODE_VEND2		58,06%	60,87%	20,00%	53,13%
Count of ODE_VEND3		52,94%	41,18%	5,88%	100,00%
<b>jah</b>					
Count of ODE_VEND		13	9	8	30
Count of ODE_VEND2		41,94%	39,13%	80,00%	46,88%
Count of ODE_VEND3		43,33%	30,00%	26,67%	100,00%
<b>Total Count of ODE_VEND</b>		<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>
<b>Total Count of ODE_VEND2</b>		<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>
<b>Total Count of ODE_VEND3</b>		<b>48,44%</b>	<b>35,94%</b>	<b>15,63%</b>	<b>100,00%</b>

### 3. Kommenteerige tabelit

(sõnastades laused vähemalt ühe rea- ja ühe veeruprotsendi kohta)!

4. Seose statistilise olulisuse testimiseks **pange** esmalt **kirja hüpoteeside paar** (et oleks ikka selge, mida testite).

5. Tehke uus 2-mõõtmeline sagedustabel tunnustele 'ODE\_VEND' ja 'AUTO', mis seekord sisaldab üksnes absoluutseid sagedusi.

Row Labels	0-1	2-3	>3	Grand Total
<b>ei</b>				
Count of ODE_VE	18	14	2	34
Count of ODE_VE	58,06%	60,87%	20,00%	53,13%
Count of ODE_VE	52,94%	41,18%	5,88%	100,00%
<b>jah</b>				
Count of ODE_VE	13	9	8	30
Count of ODE_VE	41,94%	39,13%	80,00%	46,88%
Count of ODE_VE	43,33%	30,00%	26,67%	100,00%
<b>Total Count of ODE_V</b>	<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>
<b>Total Count of ODE_V</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>
<b>Total Count of ODE_V</b>	<b>48,44%</b>	<b>35,94%</b>	<b>15,63%</b>	<b>100,00%</b>
Kommentaariid. ....				
H <sub>0</sub> : Ödede-vendade arv ja auto omamine ei ole seotud.				
H <sub>1</sub> : Ödede-vendade arv ja auto omamine on seotud.				
<b>Count of ODE_VEND</b>				
Row Labels	0-1	2-3	>3	Grand Total
ei	18	14	2	34
jah	13	9	8	30
<b>Grand Total</b>	<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>

6. Tehke uus, viimati loodud tabeliga analoogne sagedustabel, kuhu arvutage nullhüpoteesile vastavad sagedused.

Kuidas? Vt järgnevaid alapunkte!

- Tehke koopia *Pivot Table*'i abil konstrueeritud tabeli struktuurist ja väärtustest:

- seejärel kustutage ära kopeeritud tabeli **sisu**

Row Labels	0-1	2-3	>3	Grand Total
ei	<del>18</del>	<del>14</del>	<del>2</del>	34
jah	<del>13</del>	<del>9</del>	<del>8</del>	30
<b>Grand Total</b>	<b>31</b>	<b>23</b>	<b>10</b>	<b>64</b>

- ning arvutage sinna asemele sõltumatus juhule (nullhüpoteesile) vastavad sagedused (valemist  $n_{ij} = n_i \cdot n_j / n$ ).

	T	U	V	W	X
29	Count of ODE_VEND	Column Labels			
30	Row Labels	0-1	2-3	>3	Grand Total
31	ei	=X31*U33/X33			34
32	jah				30
33	Grand Total	31	23	10	64

$n_{1.1}$

.....

	T	U	V	W	X
29	Count of ODE_VEND	Column Labels			
30	Row Labels	0-1	2-3	>3	Grand Total
31	ei	16,46875	12,21875	5,3125	34
32	jah	14,53125	10,78125	=X32*W33/X33	30
33	Grand Total	31	23	10	64

$n_{2.2}$

7. Viige läbi hii-ruut test (leidke  $p$ -väärtus) – funktsioon CHISQ.TEST.

Et oleks lihtsam aru saada, milliste arvude võrdlemisel  $\chi^2$ -test baseerub (ehk siis millised tabelite osad tuleb Excel'i funktsioonile CHISQ.TEST ette anda), võib vastavad lahtrid selguse mõttes näiteks ära värvida ja/või kirjutada tabelitele peale, milliste sagedustega tegu on.

	T	U	V	W	X
20					
21	Tegelikud andmete alusel arvatud (empiirilised) sagedused				
22	Count of ODE_VEND	Column Labels			
23	Row Labels	0-1	2-3	>3	Grand Total
24	ei	18	14	2	34
25	jah	13	9	8	30
26	Grand Total	31	23	10	64
27					
28	Teoreetilised nullhüpoteesile vastavad sagedused				
29	Count of ODE_VEND	Column Labels			
30	Row Labels	0-1	2-3	>3	Grand Total
31	ei	16,46875	12,21875	5,3125	34
32	jah	14,53125	10,78125	4,6875	30
33	Grand Total	31	23	10	64
34					
35					
36	Hii-ruut-test				

Järgnevalt, nagu funktsioonide puhul ikka, tuleb kursor panna lahtrisse, kuhu soovetakse tulemust saada (ja juurde võiks enne ka kirjutada, mida arvutama hakatakse).



**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
- CHISQ.INV.RT
- CHISQ.TEST**
- CONFIDENCE.NORM
- CONFIDENCE.T
- CORREL
- COUNT

**CHISQ.TEST(actual\_range;expected\_range)**  
Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

OK Cancel

**Function Arguments**

CHISQ.TEST

**Actual\_range** U24:W25 = {18\14\2;13\9\8}

**Expected\_range** U31:W32 = {16,46875\12,21875\5,3125;14,53125\10,78125\4,6875}

= 0,07193551

Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

**Actual\_range** is the range of data that contains observations to test against expected values.

Formula result = 0,07193551

OK Cancel

NB! Vanemais Exceli versioonides testib sama hüpoteesi funktsioon CHITEST.

Tulemus:

H <sub>0</sub> : Õdede-vendade arv ja auto omamine ei ole seotud.				
H <sub>1</sub> : Õdede-vendade arv ja auto omamine on seotud.				
Tegelikud andmete alusel arvatud (empiirilised) sagedused				
Count of ODE_VEND Column L				
Row Labels	0-1	2-3	>3	Grand Total
ei	18	14	2	34
jah	13	9	8	30
Grand Total	31	23	10	64
Teoreetilised nullhüpoteesile vastavad sagedused				
Count of ODE_VEND Column Labels				
Row Labels	0-1	2-3	>3	Grand Total
ei	16,46875	12,21875	5,3125	34
jah	14,53125	10,78125	4,6875	30
Grand Total	31	23	10	64
Hii-ruut-test	0,07193551			

## 8. Sõnastage lõppjärelus

(Kas seos on statistiliselt oluline? **Miks te nii otsustasite?** Milles see seos ikka seisneb?).