

Biomeetria praks 7

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excel*'is 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 sugu ja auto omamine on seotud, st kas autot omavate tudengite arv on mees- ja naistudengite hulgas erinev?

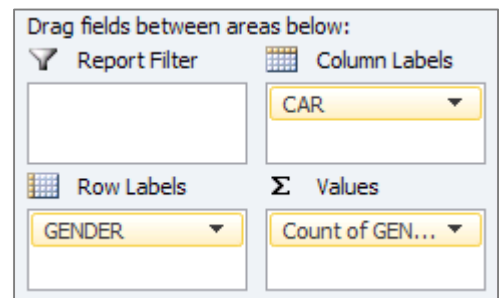
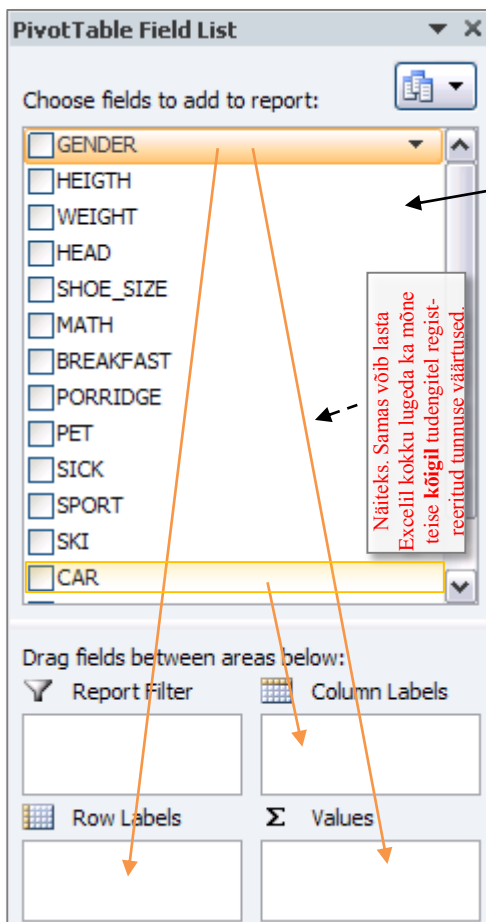
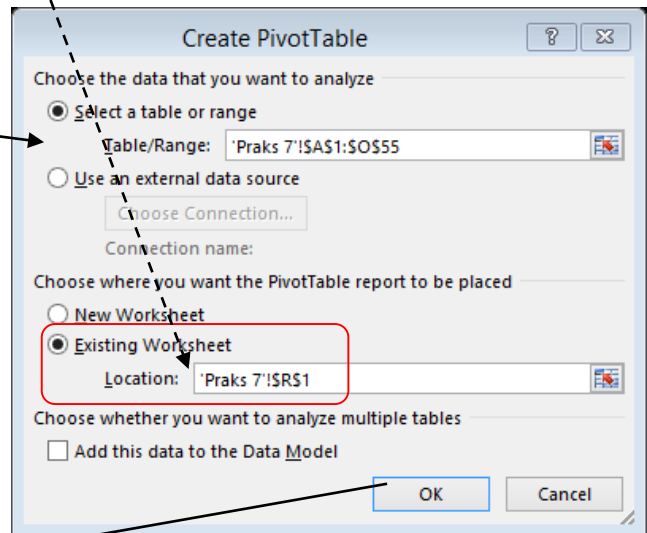
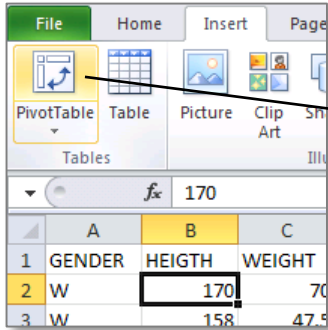
1. Konstrueerige kahemõõtmeline sagedustabel tunnustele 'GENDER' ja 'CAR'. Tudeng, kes ei tea, kas tal on auto või mitte, jätke analüüsist välja.
 2. Leidke sellesse tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii veeru- kui ka reaprotsendid).
 3. Kommenteerige tabelit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

Tööjuhend

1. Konstrueerige kahemõõtmeline sagedustabel tunnustele 'GENDER' ja 'CAR'.

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

Loodav tabel paigutage samale 'Praks7'-lehele:



Tulemus:

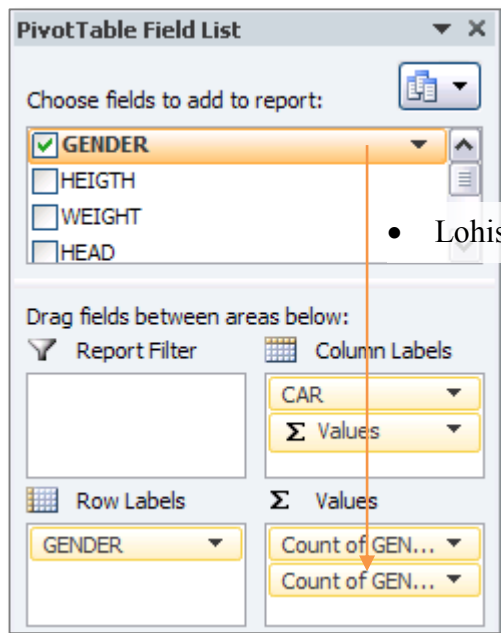
Count of GENDER	Column Labels	no	yes	(blank)	Grand Total
M		2	8		10
W		20	23	1	44
Grand Total		22	31	1	54

- Tudeng, kes ei tea, kas tal on auto või mitte, jätke analüüsist välja.

Tulemus:

Count of GENDER	Column Labels	no	yes	Grand Total
M		2	8	10
W		20	23	43
Grand Total		22	31	53

2. Leidke tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii reakui ka veeruprotsendid).

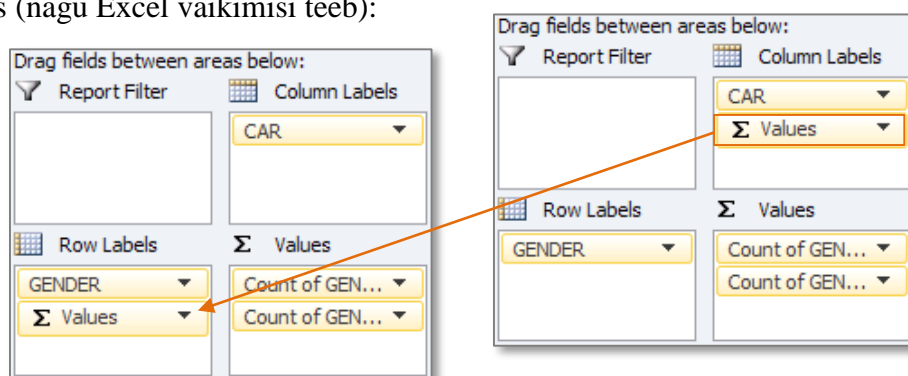


- Lohistage tunnus 'GENDER' ka teine kord väärtuste lahtrisse.

Tulemus:

Row Labels	Column Labels		no		yes		Total Count of GENDER	Total Count of GENDER2
	Count of GENDER	Count of GENDER2	Count of GENDER	Count of GENDER2				
M	2	2	8	8	10	10		
W	20	20	23	23	43	43		
Grand Total	22	22	31	31	53	53		

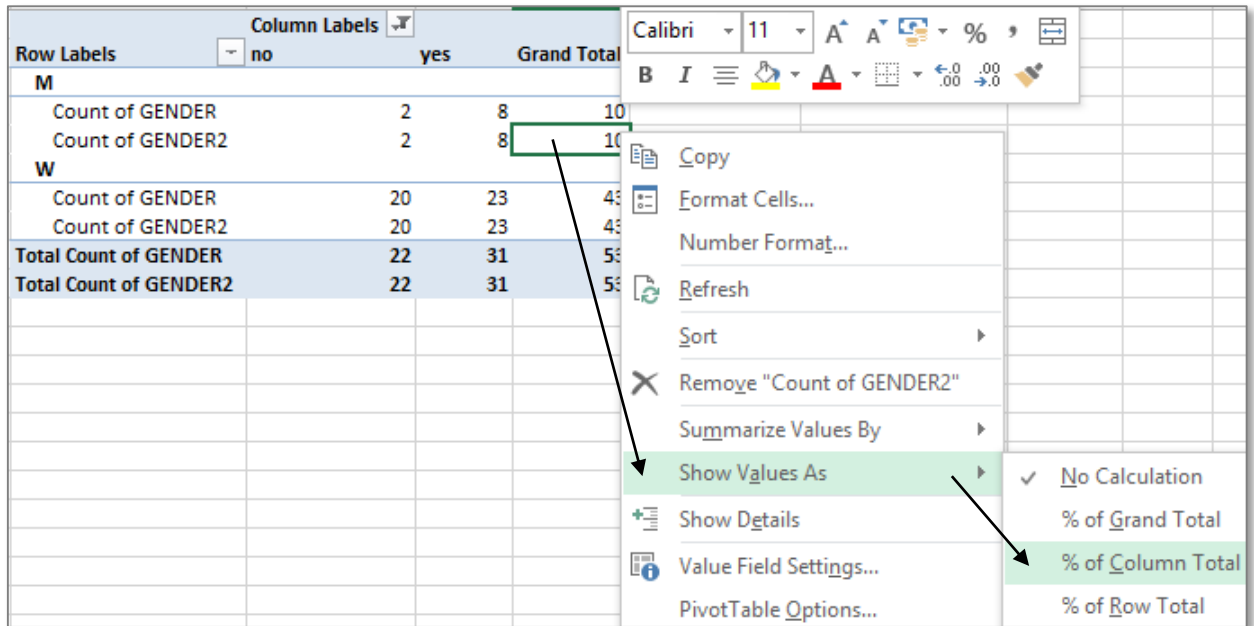
- Tabeli selgema esituse huvides jagage tabel arvutatavate väärtuste järgi ridadeks, mitte veergudeks (nagu Excel vaikimisi teeb):



Tulemus:

Row Labels	Column		Grand Total
	no	yes	
M			
Count of GENDER	2	8	10
Count of GENDER2	2	8	10
W			
Count of GENDER	20	23	43
Count of GENDER2	20	23	43
Total Count of GENDER	22	31	53
Total Count of GENDER2	22	31	53

- Esitamaks Exceli poolt uuesti kokku loetud tudengite arvu veeruprotsendina:



Tulemus:

Row Labels	no	yes	Grand Total
M			
Count of GENDER	2	8	10
Count of GENDER2	9.09%	25.81%	18.87%
W			
Count of GENDER	20	23	43
Count of GENDER2	90.91%	74.19%	81.13%
Total Count of GENDER	22	31	53
Total Count of GENDER2	100.00%	100.00%	100.00%

- Analoogselt arvutage ka reaprotsendid ... (**% of Row Total**).

(alustage sarnaselt punktis 2 kirjeldatule)

Row Labels	no	yes	Grand Total
M			
Count of GENDER	2	8	10
Count of GENDER2	9.09%	25.81%	18.87%
Count of GENDER3	20.00%	80.00%	100.00%
W			
Count of GENDER	20	23	43
Count of GENDER2	90.91%	74.19%	81.13%
Count of GENDER3	46.51%	53.49%	100.00%
Total Count of GENDER	22	31	53
Total Count of GENDER2	100.00%	100.00%	100.00%
Total Count of GENDER3	41.51%	58.49%	100.00%

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 kahemõõtmeline sagedustabel tunnustele 'GENDER' ja 'CAR', mis seekord sisaldab üksnes absoluutseid sagedusi.

Row Labels	no	yes	Grand Total
M			
Count of GEI	2	8	10
Count of GEI	9.09%	25.81%	18.87%
Count of GEI	20.00%	80.00%	100.00%
W			
Count of GEI	20	23	43
Count of GEI	90.91%	74.19%	81.13%
Count of GEI	46.51%	53.49%	100.00%
Total Count of G	22	31	53
Total Count of G	100.00%	100.00%	100.00%
Total Count of G	41.51%	58.49%	100.00%
Kommentaariid.			
H ₀ : Sugu ja auto omamine ei ole seotud.			
H ₁ : Sugu ja auto omamine on seotud.			
Count of GENDER			
Row Labels	no	yes	Grand Total
M	2	8	10
W	20	23	43
Grand Total	22	31	53

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

Kuidas? Vt järgnevaid alapunkte!

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

Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M	2	8	10
W	20	23	43
Grand Total	22	31	53

Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M	2	8	10
W	20	23	43
Grand Total	22	31	53

- seejärel kustutage ära kopeeritud tabeli **sisu**

Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M	2	8	10
W	20	23	43
Grand Total	22	31	53

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

	R	S	T	U
29	Count of GENDE	Column Labels		
30	Row Labels	no	yes	Grand Total
31	M	=U31*S33/U33		10
32	W			43
33	Grand Total	22	31	53

Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M	4.1509434	25.1509434	10
W	17.849057	=U32*T33/U33	43
Grand Total	22	31	53

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

Et oleks lihtsam aru saada, milliste arvude võrdlemisel χ^2 -test baseerub (ehk siis millised tabelite osad tuleb Exceli funktsioonile CHISQ.TEST ette anda), võib vastavad lahtrid selguse mõttes ära värvida ja ka tabelitele peale kirjutada, mida neis olevad arvud näitavad.

	R	S	T	U
21	Andmete alusel arvutatud (empiirilised) sagedused			
22	Count of GENDER Column Labels			
23	Row Labels	no	yes	Grand Total
24	M	2	8	10
25	W	20	23	43
26	Grand Total	22	31	53
27	Teoreetilised nullhüpooteesile vastavad sagedused			
29	Count of GENDER Column Labels			
30	Row Labels	no	yes	Grand Total
31	M	4.150943	5.849057	10
32	W	17.84906	25.15094	43
33	Grand Total	22	31	53
34				
35				
36	Hii-ruut-test			

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

Function Arguments

CHISQ.TEST

Actual_range S24:T25 = {2,8;20,23}

Expected_range S31:T32 = {4.15094339622642,5.84905660377...}

= 0.125385274

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 = 0.125385274

[Help on this function](#)

OK Cancel

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

al_range;expected_range)

independence: the value from the chi-squared distribution for appropriate degrees of freedom.

OK Cancel

NB! Vanemais Exceli versioonides testib sama hüpooteesi funktsioon CHITEST.

Tulemus:

8. Sõnastage lõppjärelus

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

Kui noormeeste hulgas on autoomanikke 80,0%, siis neidude hulgas vaid 53,5%. Kõigist autoomanikest on 25,8% mehed ja 74,2% naised (naisterahvaid on lihtsalt rohkem). Hii-ruut testi kohaselt ei ole kirjeldatud erinevus autoomanike proportsioonis meeste ja naiste vahel siiski statistiliselt oluline ($p = 0,125 > 0,05$), st et väites, nagu kehtiks antud andmeist ilmnenu erinevus mees- ja naistudengite vahel üldisemalt, on eksimise tõenäosus 12,5%, mis on üldistuse tegemiseks pisut liiga suur viga (seega – seos ei ole statistiliselt oluline).

H ₀ : Sugu ja auto omamine ei ole seotud.			
H ₁ : Sugu ja auto omamine on seotud.			
Andmete alusel arvutatud (empiirilised) sagedused			
Count of GEI Column Labels			
Row Labels	no	yes	Grand Total
M	2	8	10
W	20	23	43
Grand Total	22	31	53
Teoreetilised nullhüpooteesile vastavad sagedused			
Count of GEN Column Labels			
Row Labels	no	yes	Grand Total
M	4.150943	5.849057	10
W	17.84906	25.15094	43
Grand Total	22	31	53
Hii-ruut-test	0.125385	= p > 0.05 ->	H ₀ : Sugu ja auto omamine ei ole seotud.