

Biomeetria praks 7

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage MS Excel'is oma kursuse ankeedivastuseid sisaldav (**eelmistes praktikumides puhastatud**) andmestik,
 2. lisage uus tööleht (*Insert / Lisa -> Worksheet / Arvutustabel*), nimetage see ümber leheküljeks 'Praks7' ja
 3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks7' ülemisse vasakusse nurka.
-

Ülesanne.

Kas ebameeldivaima aine grupp ja õnnelikkus on seotud, st kas õnnelike tudengite esinemissagedus on erinev, sõltuvalt ebameeldivaima aine grupist?

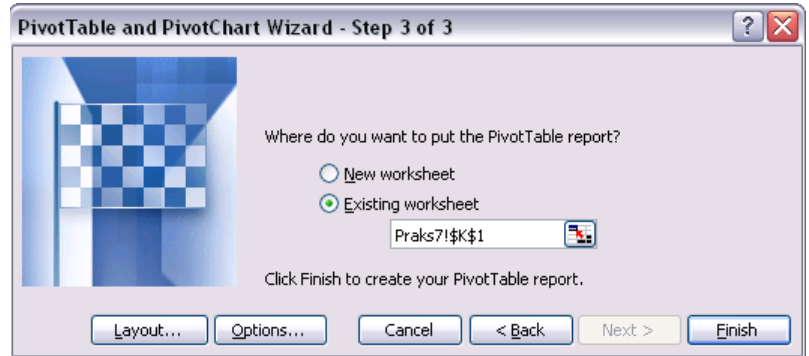
1. Konstrueerige kahemõõtmeline sagedustabel tunnustele 'AINEKOOD' ja 'ÕNNELIK'.
 2. Leidke sellesse tabelisse lisaks absoluutsetele sagedustele ka kahed suhtelised sagedused (nii veeru- kui ka reaprotsendid).
 3. Kommenteerige tabelit (sõnastades laused vähemalt kahe arvutatud protsendi kohta)!
 4. Testimaks tuvastatud seose/erinevuse statistilist olulisust, pange kirja vastav hüpoteeside paar.
 5. Tehke uus 2-mõõtmeline sagedustabel tunnustele 'AINEKOOD' ja 'ÕNNELIK', 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 lõppjärelendus (kas seos on statistiliselt oluline, mille alusel te seda otsustasite ja milles see seos ikkagi seisneb?).
-

Tööjuhend

1. Konstrueerige kahemõõtmeline sagedustabel tunnustele 'AINEKOOD' ja 'ÕNNELIK'.

- Paigutage kursor andmetabeli suvalisse lahtrisse (leheküljel 'Praks 7').
- Data / Andmed* → *PivotTable and PivotChart Report... / PivotTable ja PivotChart Aruanne...*

Loodav tabel paigutage samale 'Praks7'-lehele, näiteks:



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

- Jätke tabelist välja märkimata ebameeldivaima ainega tudengid (blank) ja ka ainegrupp 'Muu', sest selle esinemissagedus on korrektse analüüsi tarvis liiga väike.

ÕNNELIK	Hum	Reaal	Grand Total
ei	3	1	4
jah	16	27	43
nii ja naa	11	5	16
Grand Total	30	33	63

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

The screenshot illustrates the process of creating a PivotTable in Excel. It shows a PivotTable with the following data:

ÕNNELIK	Hum	Reaal	Grand Total
ei	3	1	4
jah	16	27	43
nii ja naa	11	5	16
Grand Total	30	33	63

The PivotTable Field List shows the source data and the field 'ÕNNELIK' is added to the Row Area. The context menu for the PivotTable includes 'Field Settings...'. The PivotTable Field dialog box shows the source field 'ÕNNELIK' and the summary function 'Count'. The 'Show data as:' dropdown menu is set to 'Normal'.

Tulemus:

ÕNNELIK	Data	Hum	Reaal	Grand Total
ei	Count of ÕNNELIK	3	1	4
	Count of ÕNNELIK2	75,00%	25,00%	100,00%
jah	Count of ÕNNELIK	16	27	43
	Count of ÕNNELIK2	37,21%	62,79%	100,00%
nii ja naa	Count of ÕNNELIK	11	5	16
	Count of ÕNNELIK2	68,75%	31,25%	100,00%
Total Count of ÕNNELIK		30	33	63
Total Count of ÕNNELIK2		47,62%	52,38%	100,00%

Analoogselt arvutage ka veeruprotsendid ...

		AINEKOOD ▾		
ÕNNELIK ▾	Data ▾	Hum	Reaal	Grand Total
ei	Count of ÕNNELIK	3	1	4
	Count of ÕNNELIK2	75,00%	25,00%	100,00%
	Count of ÕNNELIK3	10,00%	3,03%	6,35%
jah	Count of ÕNNELIK	16	27	43
	Count of ÕNNELIK2	37,21%	62,79%	100,00%
	Count of ÕNNELIK3	53,33%	81,82%	68,25%
nii ja naa	Count of ÕNNELIK	11	5	16
	Count of ÕNNELIK2	68,75%	31,25%	100,00%
	Count of ÕNNELIK3	36,67%	15,15%	25,40%
Total Count of ÕNNELIK		30	33	63
Total Count of ÕNNELIK2		47,62%	52,38%	100,00%
Total Count of ÕNNELIK3		100,00%	100,00%	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 2-mõõtmeline sagedustabel tunnustele 'AINEKOOD' ja 'ÕNNELIK', mis seekord sisaldab üksnes absoluutseid sagedusi.

Kommentaariid.			
H ₀ : Õnnelikkus ja ebameeldivaima aine grupp ei ole seotud.			
H ₁ : Õnnelikkus ja ebameeldivaima aine grupp on seotud.			
Count of ÕNNELIK		AINEKOOD ▾	
ÕNNELIK ▾	Hum	Reaal	Grand Total
ei	3	1	4
jah	16	27	43
nii ja naa	11	5	16
Grand Total	30	33	63

Et χ^2 -testi tulemused ei pruugi väga väikeste (< 5) sageduste korral olla õiged, on seose statistilise olulisuse testimisel mõttekas eemaldada tabelist väheesinevad väärtused.

Antud juhul on mitteõnnelike tudengite arv (4) nende analüüsi kaasamiseks liiga väike, mistõttu tuleks χ^2 -testi tarvis tunnuse 'ÕNNELIK' väärtus 'ei' tabelist eemaldada.

Count of ÕNNELIK		AINEKOOD ▾		
ÕNNELIK ▾	Hum	Reaal	Grand Total	
jah	16	27	43	
nii ja naa	11	5	16	
Grand Total	27	32	59	

Alternatiivina võib mõnikord sarnased väärtused grupeerida (siis ei vähene andmestiku suurus, samas muutuvad järeldused üldisemaks).

6. Eelmises punktis moodustatud sagedustabeli lõppvariandile arvutage nullhüpoteesile vastavad sagedused (**tehke uus tabel – vt järgnevaid alapunkte selle leheküljel!**).

Lihtsaim variant on ehk

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

The image shows a PivotTable in Excel with the following data:

Count of ÖNNELIK	AINEKOOD			
ÖNNELIK	Hum	Reaal	Grand Total	
jah		16	27	43
nii ja naa		11	5	16
Grand Total		27	32	59

The 'Paste Special' dialog box is open with the following settings:

- Paste: Values
- Operation: None
- Skip blanks
- Transpose

- seejärel kustutada ära kopeeritud tabeli sisu

Count of ÖNNELIK	AINEKOOD			
ÖNNELIK	Hum	Reaal	Grand Total	
jah		16	27	43
nii ja naa		11	5	16
Grand Total		27	32	59

- ning arvutada sinna asemele sõltumatuse juhule (nullhüpoteesile) vastavad sagedused

(valemist $n_{ij} = n_i \cdot n_j / n$).

	K	L	M	N
30	Count of ÖNNELIK	AINEKOOD		
31	ÖNNELIK	Hum	Reaal	Grand Total
32	jah	=N32*L34/N34		43 n_1
33	nii ja naa			16
34	Grand Total	27	32	59 n

.....

	K	L	M	N
30	Count of ÖNNELIK	AINEKOOD		
31	ÖNNELIK	Hum	Reaal	Grand Total
32	jah	19,6779661	23,3220339	43
33	nii ja naa	7,322033898	=N33*M34/N34	16 n_2
34	Grand Total	27	32	59 n

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

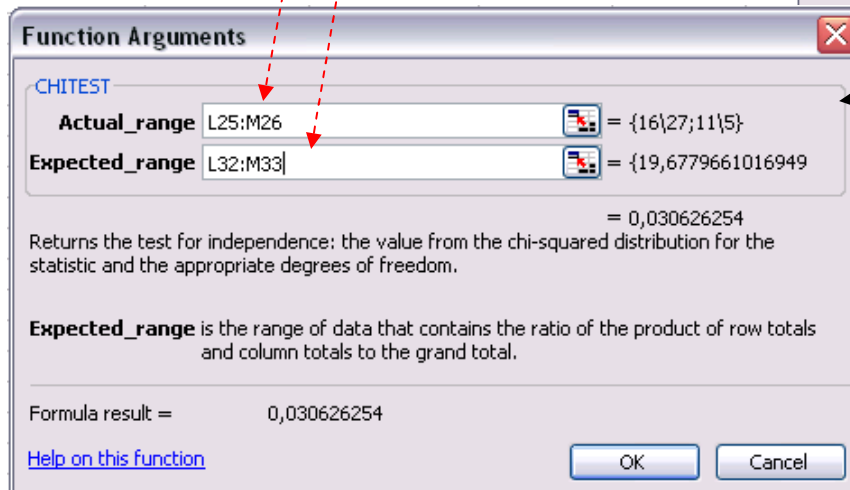
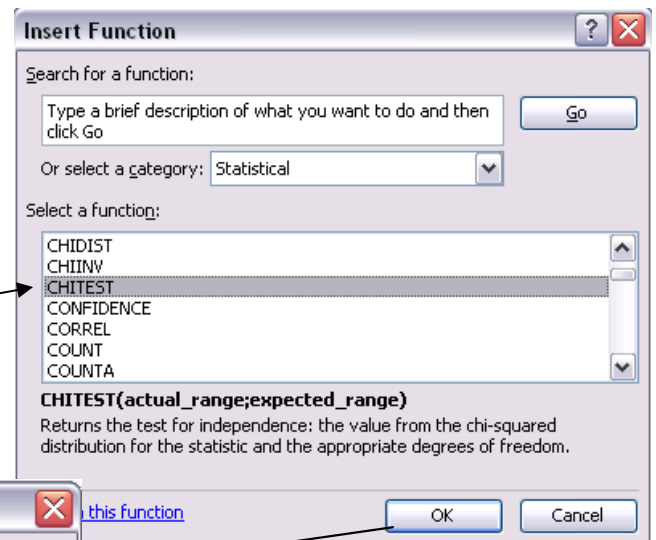
Et oleks lihtsam aru saada, milliste arvude võrdlemisel χ^2 -test baseerub (ehk siis millised tabelite osad tuleb *Excel*'i funktsioonile CHITEST ette anda), võib vastavad lahtrid selguse mõttes näiteks ära värvida.

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

Count of ÖNNELIK	AINEKOOD		
ÖNNELIK	Hum	Reaal	Grand Total
jah	16	27	43
nii ja naa	11	5	16
Grand Total	27	32	59

Count of ÖNNELIK	AINEKOOD		
ÖNNELIK	Hum	Reaal	Grand Total
jah	19,6779661	23,3220339	43
nii ja naa	7,322033898	8,677966102	16
Grand Total	27	32	59

Hii-ruut test



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