

[MTTTA1] TILASTOMENETELMIEN PERUSTEET, KEVÄT 2019

<https://coursepages.uta.fi/mttta1/kevat-2019/>

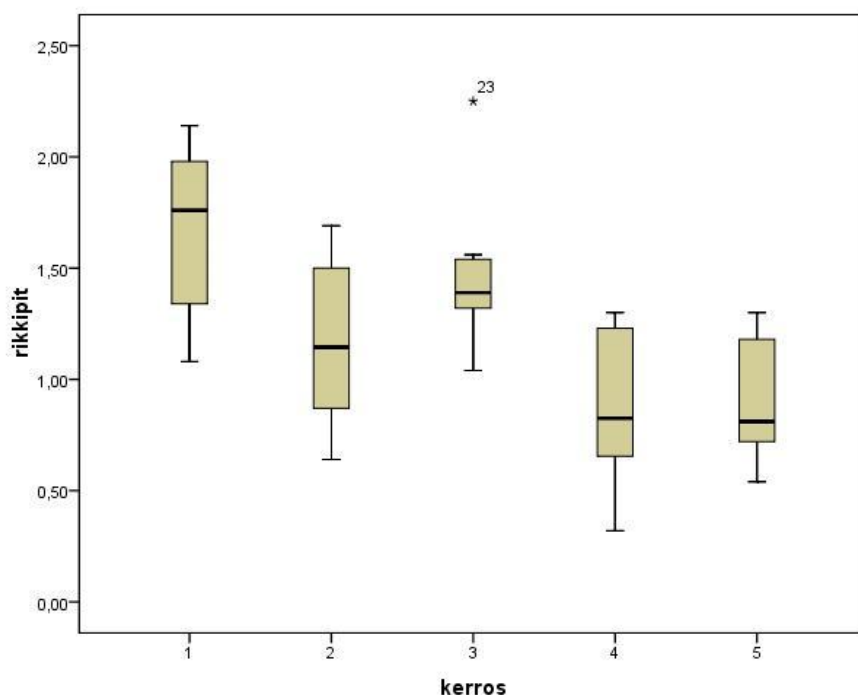
HARJOITUS 3 viikko 6

RYHMÄT:

ke	12.15–13.45	Is. C6	Leppälä
to	08.30–10.00	Is. C6	Korhonen
to	12.15–13.45	Is. C6	Korhonen
to	14.15–15.45	Is. C8	Leppälä

 Aiheet: Varianssianalyysi, χ^2 -riippumattomuustesti, regressioanalyysi

1. Ohessa SPSS -harjoitusten 1 tehtävän 2 analyysin tulokset kivihiilikerrosten vaikutuksesta rikkipitoisuuksiin. Tulkitse tulokset.



Descriptives

rikkipit

	N	Mean	Std. Deviation
1,00	7	1,6600	,41861
2,00	8	1,1700	,37887
3,00	9	1,4600	,33853
4,00	8	,8800	,35006
5,00	10	,8800	,27207
Total	42	1,1895	,45351

Test of Homogeneity of Variances

rikkipit

Levene Statistic	df1	df2	Sig.
,838	4	37	,510

ANOVA

rikkipit

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3,935	4	,984	8,095	,000
Within Groups	4,497	37	,122		
Total	8,432	41			

2. Ohessa on syksyn 2011 tilastotieteen johdantokurssin opiskelijapalautteen perusteella tehty ristiintaulukko. Täydennä puuttuvat kohdat A) - H). Aseta tilanteeseen liittyvät hypoteesit ja suorita testaus. Miten testisuure on laskettu?

Harjoitustyön vaikeusaste * Onko ensimmäisen vuoden opiskelija? Crosstabulation

			Onko ensimmäisen vuoden opiskelija?		
			Kyllä	Ei	Total
Harjoitustyön vaikeusaste	Helppo tai melko helppo	Count	A)	B)	9
		Expected Count	C)	D)	9,0
		% within Onko ensimmäisen vuoden opiskelija?	15,0%	23,1%	19,6%
	Sopiva	Count	12	10	22
		Expected Count	9,6	12,4	22,0
		% within Onko ensimmäisen vuoden opiskelija?	60,0%	38,5%	47,8%
	Vaikea tai melko vaikea	Count	5	10	15
		Expected Count	6,5	8,5	15,0
		% within Onko ensimmäisen vuoden opiskelija?	25,0%	38,5%	32,6%
	Total	Count	20	26	46
		Expected Count	20,0	26,0	46,0
		% within Onko ensimmäisen vuoden opiskelija?	100,0%	100,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2,102 ^a	E)	F)
Likelihood Ratio	2,116	2	,347
Linear-by-Linear Association	,064	1	,801
N of Valid Cases	46		

a. 1 cells (G %) have expected count less than 5. The minimum expected count is H) .

3. Kevytlevitteen valmistaja tutkii tunnistavatko miehet ja naiset voin ja kevytlevitteen eron samalla tavalla. Tutkimuksessa oli mukana 200 miestä ja 200 naista. Miehistä eron tunnisti 108 ja naisista 120. Aseta tilanteeseen liittyvät hypoteesit ja suorita testaus 5 %-n riskitasolla.

4. Markkinointitutkija halusi selvittää, vaikuttaako taustamusiikki asiakkaan tekemään viinin valintaan. Hän teki tutkimuksen eräässä kaupassa Pohjois-Irlannissa. Kerätyn aineiston perusteella hän sai ristiintaulukon

Ostettu viini	Taustamusiikki			Yhteensä
	Ei musiikkia	Ranskalainen	Italialainen	
Ranskalainen	30	39	30	99
Italialainen	11	1	19	31
Muu	43	35	35	113
Yhteensä	84	75	84	243

Aseta tilanteeseen sopivat hypoteesit ja suorita testaus 1 %:n riskitasolla.

Voit halutessasi käyttää testisuureen laskussa esimerkiksi sivua

http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html tai <http://vassarstats.net/newcs.html>

Jos laskit testisuureen käsin, niin arvioi taulukon avulla pienintä riskitasoa, jolla H_0 voidaan hylätä, tai määritä tarkempi arvo käyttäen sopivaa ohjelmaa (esim. Excel (funktio CHIDIST), SPSS (funktio CDF.CHISO), <http://vassarstats.net/vsdist.html>).

5. Tarkastellaan erään sairauden A esiintymistä tytöillä ja pojilla. Saadaan tutkimustulokset:

	Ikä 0 - 6		Ikä 7 - 12	
	Tytöt	Pojat	Tytöt	Pojat
On sairaus	10	25	25	10
Ei ole sairautta	100	80	100	100
Yhteensä	110	105	125	110

Onko sairauden esiintymisessä eroja tyttöjen ja poikien välillä, jos otetaan lapsen ikä huomioon? Aseta tilanteeseen sopivat hypoteesit, suorita testaus ja tee johtopäätelmät. Onko iän huomioiminen tarpeellista? Miksi?

Voit halutessasi käyttää laskuissa esimerkiksi sivua

http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html tai <http://vassarstats.net/newcs.html>

6. Tutkittiin miehillä iän (x , vuosina) vaikutusta reagointiaikoihin (y , millisekunteina) ja saatiin aineisto

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	37	190	1369	7030	36100
2	35	197	1225	6895	38809
3	41	205	1681	8405	42025
4	43	210	1849	9030	44100
5	42	218	1764	9156	47524
6	50	226	2500	11300	51076
7	49	228	2401	11172	51984
8	54	230	2916	12420	52900
9	60	234	3600	14040	54756
10	65	240	4225	15600	57600
	476	2178	23530	105048	476874

Piirrä riippuvuutta kuvaava pisteparvi ja estimoi yhden selittäjän regressiomalli. Kuinka paljon vuoden lisäys iässä muuttaa keskimääräistä reagointiaikaa?

7. (jatkoa tehtävään 6) Piirrä estimoimasi regressiosuora pisteparveen. Laske mallin selitysprosentti $100(r_{xy})^2$, missä r_{xy} on muuttujien välinen korrelaatiokerroin. Estimoi 45-vuotiaan miehen reagointiaika.