

MTTTP1, luento 4.4.2019

7.7.3 χ^2 -riippumattomuustesti

Kahden muuttujan välinen riippuvuustarkastelu ristiintaulukon perusteella, kaava (12), hypoteesit

H_0 : ei riippuvuutta

H_1 : on riippuvuutta

Testisuure χ^2 -riippumattomuustestisuure, joka noudattaa nk. χ^2 -jakaumaa vapausastein $(I-1)(J-1)$. I ja J ristiintaulukon muuttujien luokkien lukumäärät. Ks. χ^2 -jakauman tiheysfunktio

<https://fi.wikipedia.org/wiki/%CE%A7%C2%B2-jakauma>

Esim. 7.7.8

Taustamusiikki

Ostettu viini	Ei	Ranskalainen	Italialainen	Yhteensä
Ranskalainen	30 (36%)	39 (52%)	30 (36%)	99
Italialainen	11 (13%)	1 (1%)	19 (23%)	31
Muu	43 (51%)	35 (47%)	35 (42%)	113
Yhteensä	84	75	84	243

Kun ristiintaulukosta lasketaan

χ^2 -riippumattomuustestisuure, sen arvoksi saadaan 18,279 ja p -arvoksi 0,001. Tehdään päättely 1 %:n riskitasolla. Koska p -arvo on pienempi kuin valittu riskitaso, niin päätellään taustamusiikin ja viinin valinnan välillä olevan riippuvuutta.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4,286 ^a	4	,369
Likelihood Ratio	4,220	4	,377
Linear-by-Linear Association	,013	1	,910
N of Valid Cases	174		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 9,31.

Hyväksytään H_0 : ei riippuvuutta, koska p -arvo (Pearson Chi-Square) on $0,369 > 0,05$.

Testiä voidaan käyttää

a) $(I-1)(J-1) = 1$

- $n > 40$
- $20 \leq n \leq 40$, kaikkien teoreettisten frekvenssien (expected count) oltava ≥ 5

b) $(I-1)(J-1) > 1$

- kaikkien teoreettisten frekvenssien oltava > 1 ja enintään 20 % saa olla alle 5

Esim. 7.7.6

Opintojakson työläys * opsuunta Crosstabulation

			opsuunta		Total
			hallinto	taloust	
Opintojakson työläys	työläs	Count	13	16	29
		Expected Count	8,5	20,5	29,0
		% within opsuunta	68,4%	34,8%	44,6%
	sopiva	Count	5	15	20
		Expected Count	5,8	14,2	20,0
		% within opsuunta	26,3%	32,6%	30,8%
	vähätöinen	Count	1	15	16
		Expected Count	4,7	11,3	16,0
		% within opsuunta	5,3%	32,6%	24,6%
Total	Count	19	46	65	
	Expected Count	19,0	46,0	65,0	
	% within opsuunta	100,0%	100,0%	100,0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7,668 ^a	2	,022
Likelihood Ratio	8,680	2	,013
Linear-by-Linear Association	7,548	1	,006
N of Valid Cases	65		

a. 1 cells (16,7%) have expected count less than 5. The minimum expected count is 4,68.

Testin käyttöön liittyvät oletukset tällä luokituksella kunnossa, vain 16,7 % odotetusta frekvensseistä alle 5 ja kaikki > 1 .

Pienin riskitaso, jolla H_0 voidaan hylätä, on 0,022.

Tätä suuremmilla riskeillä H_0 hylätään, pienemmällä hyväksytään.

7.7.4 Odotusarvojen yhtäsuuruuden testaaminen t-testillä

Tutkitaan kahden populaation odotusarvojen yhtäsuuruutta riippumattomien otosten t-testillä, kaava (13), hypoteesit

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2 \quad (\text{tai yksisuuntaisena})$$

Oletukset t-testin käyttöön liittyen: riippumattomat otoksen normaalijakaumista, joiden varianssit tuntemattomia, mutta yhtä suuret (yhtäsuuruutta voidaan testata).

Esim. 7.7.9

$$H_0 : \mu_{\text{poijat}} = \mu_{\text{tytöt}}$$

$$H_1 : \mu_{\text{poijat}} \neq \mu_{\text{tytöt}}$$

Group Statistics

sex		N	Mean	Std. Deviation	Std. Error Mean
paino	poika	65	3640,46	438,244	54,357
	tyttö	55	3451,27	523,280	70,559

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
paino	Equal variances assumed	,293	,589	2,156	118	,033	189,189	87,765	15,390	362,987
	Equal variances not assumed			2,124	105,703	,036	189,189	89,069	12,595	365,783

H_0 : Populaatioiden varianssit samoja

Testisuure (Levene's Test for Equality of Variances)

Koska p -arvo on $0,589 > 0,05$, H_0 hyväksytään, t-testin tulokset 1. riviltä.

$$H_0 : \mu_{\text{pojat}} = \mu_{\text{tytöt}}$$

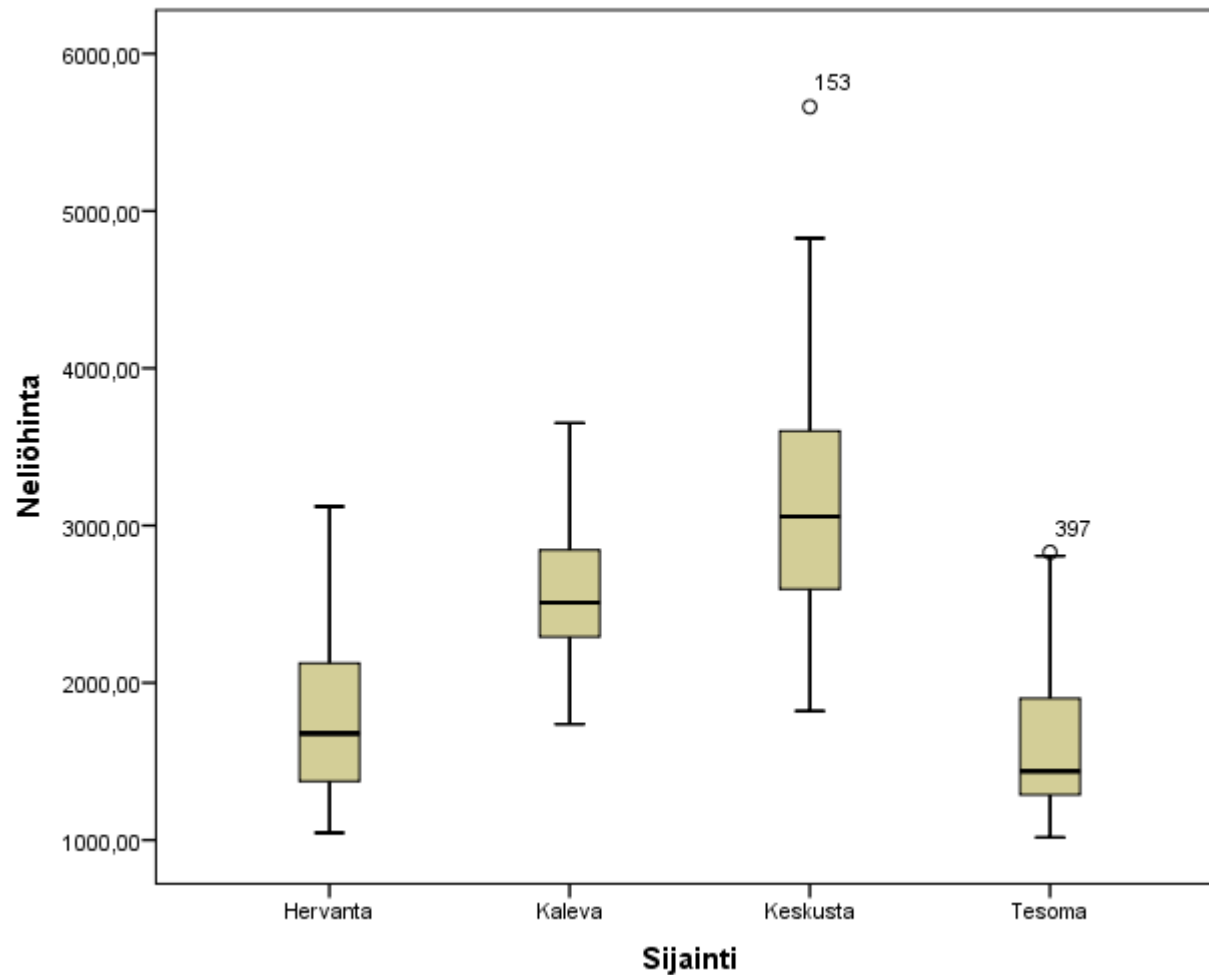
Testisuureen arvo on 2,156, kaksisuuntaisen testin p -arvo 0,033.

Jos riskitasoksi valitaan 5 %, niin nollahypoteesi hylätään (koska p -arvo $< 0,05$) ja tehdään päätelmä, että tytöt ja pojat ovat syntyessään keskimäärin eri painoisia.

Jos otettaisiin riski, joka olisi pienempi kuin 3,3 % (vaikkapa 1 %) niin tehtäisiin päinvastainen päätelmä.

Yksisuuntaiseen testiin liittyvä p -arvo on $0,033/2 = 0,0165$.

Esim. 7.7.11 Tutkitaan, poikkeavatko Hervannan ja Tesoman keskimääräiset neliöhinnat toisistaan, aineisto Tre_myydyt_asunnot_2012



Group Statistics

Sijainti	N	Mean	Std. Deviation	Std. Error Mean
Neliöhinta Hervanta	127	1752,6063	456,78817	40,53340
Tesoma	54	1593,3333	484,13026	65,88178

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Neliöhinta	Equal variances assumed	,009	,926	2,108	179	,036	159,27297	75,55123	10,18732	308,35861
	Equal variances not assumed			2,059	94,993	,042	159,27297	77,35222	5,70924	312,83669

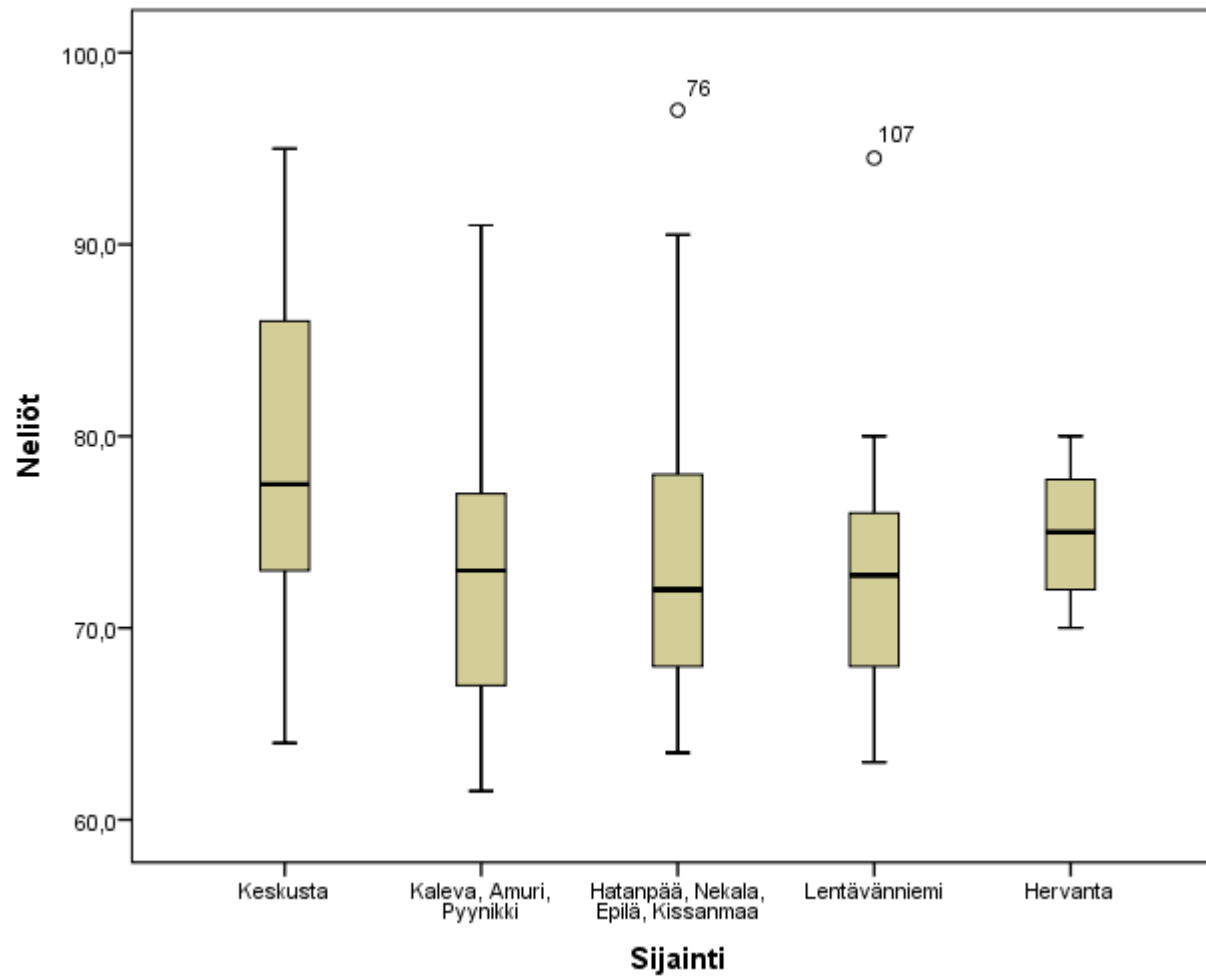
Otoskeskiarvojen ero 159 euroa, t-testin arvo 2,108, johon liittyvä p -arvo on 0,036 (kaksisuuntainen testi). 5 %:n riskitasolla päätellään, että keskihinnat eroavat, 1 %:n riskitasolla päätellään, että ei eroja.

Odotusarvojen erotuksen luottamusvälin perusteella odotusarvojen erotuksen arvellaan olevan välillä 10 € – 308 €. Jos päättely tehdään luottamusvälin eikä testisuureen perusteella, niin päätellään odotusarvojen poikkeavan toisistaan, koska nolla ei kuulu luottamusvälille.

Varianssit voitiin olettaa yhtä suuriksi, koska p-arvo on 0,926 ($> 0,05$).

Esim. Tre_myydyt_kolmiot_2010,
http://www.sis.uta.fi/tilasto/tiltp_aineistoja/Tre_myydyt_kolmiot_2010.sav , keskimääräiset koot alueittain, eroja joidenkin alueiden välillä.

Tehdään päättelyt 1 %:n riskitasolla kaksisuuntaisessa testissä.



Group Statistics

Sijainti		N	Mean	Std. Deviation	Std. Error Mean
Neliöt	Keskusta	30	78,850	8,3213	1,5193
	Kaleva, Amuri, Pyyrikki	37	73,273	8,3594	1,3743

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Neliöt	Equal variances assumed	,014	,906	2,721	65	,008	5,5770	2,0496
	Equal variances not assumed			2,722	62,283	,008	5,5770	2,0486

On eroja, p-arvo $0,008 < 0,01$

Group Statistics				
Sijainti	N	Mean	Std. Deviation	Std. Error Mean
Neliöt Kaleva, Amuri, Pyynikki	37	73,273	8,3594	1,3743
Hatanpää, Nekala, Epilä, Kissanmaa	33	73,682	7,7740	1,3533

Independent Samples Test								
		Levene's Test for Equality of Variances			t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Neliöt	Equal variances assumed	,240	,626	-,211	68	,833	-,4088	1,9369
	Equal variances not assumed			-,212	67,872	,833	-,4088	1,9287

Ei eroja, p-arvo 0,833 > 0,01

Group Statistics

Sijainti		N	Mean	Std. Deviation	Std. Error Mean
Neliöt	Lentävänniemi	38	72,303	5,9121	,9591
	Hervanta	36	75,000	3,2049	,5342

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Neliöt	Equal variances assumed	4,103	,047	-2,421	72	,018	-2,6974	1,1143
	Equal variances not assumed			-2,457	57,650	,017	-2,6974	1,0978

Ei eroja, p-arvo $0,018 > 0,01$

Kaikissa tarkasteluissa populaatioiden varianssit voidaan olettaa samoiksi, Levenen testin p-arvot 0,906, 0,626, 0,047.

7.7.5 Lineaarinen riippuvuus

Populaation korrelaatiokertoimen testaus

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

http://www.sis.uta.fi/tilasto/mtt1/syksy2018/luento_runko.pdf#page=93

KERTAUSTA

Kaavat esimerkkeineen

http://www.sis.uta.fi/tilasto/mtt1/syksy2018/esimerkit_kaavoihin.pdf