

Tilastomenetelmien peruskurssi(TILP150)

Syksy 2006

3. harjoitukset/viikko 39.

Näissä harjoituksissa käsitellään Lukua 6 (lukuun ottamatta Luku 6.2.2). 1.-5. tehtävät liittyvät Lukuun 6.1, ja 6.-8.tehtävät Lukuun 6.2.

1. Syksyn 2006 TILP150-kurssilaisiin liittyvän aineiston perusteella tutkitaan, onko muuttujien MATPIT = lyhyt/pitkä matematiikka (1 = lyhyt, 2 = pitkä) ja PAKKO = Tilastomenetelmien peruskurssi on minulle ”pakkopullaa” (1 = täysin samaa mieltä, 2 = jonkin verran samaa mieltä, 3 = jonkin verran eri mieltä, 4 = täysin eri mieltä) välillä riippuvuutta ja millaista se on?

MATPIT		lyhyt	pitkä	yht.
PAKKO	täysin samaa mieltä	13	8	21
	jonkin verran samaa mieltä	30	33	63
	jonkin verran eri mieltä	23	50	73
	täysin eri mieltä	2	10	12
yht.		68	101	169

- a) Kirjaa muuttujien mitta-asteikot.
 - b) Poimi muuttujien frekvenssijakaumat ristiintaulukosta.
 - c) Laske sarakeprosentit (ts. muuttujan PAKKO pros. frekvenssijakaumat ehdolla muuttujan MATPIT arvot) sekä muuttujan PAKKO pros. frekvenssijakauma.
 - d) Vertaile sanallisesti c) kohdassa laskemiasi prosentteja.
2. Jatkoa 1. tehtävään.
 - a) Laske odotetut frekvenssit, standardoidut jäännökset ja χ^2 -luvun arvo (vastaus $\chi^2 = 10.6$).
 - b) Tulkitse a) kohdan tulokset sanallisesti.
 3. *Miten paljon pienempi riski poliolääkettä syöneillä on sairastua polioon verrattuna niihin, jotka söivät lumelääkettä?* Tarkastellaan aineistoa, joka on esitetty luentomonisteen Luvussa 3.3.

		Poliolääke		
		kyllä(= 1)	ei (= 2)	yht.
Sairastunut	kyllä (= 1)	56	142	198
	ei (= 2)	199 944	199 858	399 802
yht.		200 000	200 000	400 000

Laske riskisuhde RR ja tulkitse se sanallisesti.

4. a) *Eroavatko naisten ja miesten käyttäytymiset tupakoinnin suhteen ja jos eroavat, niin miten?* Tutkitaan aineistoa:

		SP		
		mies	nainen	yht.
TUP	kyllä	5	15	20
	ei	15	5	20
	yht.	20	20	40

Aineistosta laskettu $\chi^2 = 10$. Laske sarakeprosentit. Kerro tulokset sanallisesti.

- b) *Eroavatko naisten ja miesten käyttäytymiset tupakoinnin suhteen ja jos eroavat, niin miten?* Tutkitaan aineistoa:

		SP		
		mies	nainen	yht.
TUP	kyllä	15	5	20
	ei	5	15	20
	yht.	20	20	40

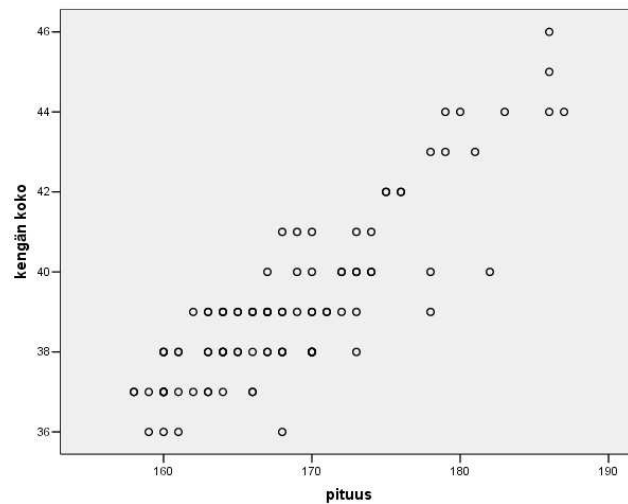
Aineistosta laskettu $\chi^2 = 10$. Laske sarakeprosentit. Kerro tulokset sanallisesti.

5. Täytä seuraava 2×2 -ristiintaulukko haluamallasi tavalla. (Voit keksiä myös muuttujat.)

	1	2	yht.
1			20
2			20
yht.	20	20	40

Onko taulukossasi muuttujien välillä riippuvuutta ja millaista se on?

6. Analysoidaan kevään 2005 TILA03-kurssilaisiin liittyvän aineiston perusteella, riippuvatko pituus ja kengänkoko toisistaan ja miten.
- a) Sirontakuviot on esitetty kuvassa 1. Tarkastele sirontakuviota ja kerro omin sanoin, millaisesta riippuvuudesta on kyse. Arvaa Pearsonin korrelaatiokertoimen arvo.



Kuva 1: Kevään 2005 Tilastomenetelmien peruskurssilaisten kengännumeron ja pituuden sirontakuvio.

- b) Aineistosta laskettu Pearsonin korrelaatiokerroin on 0.86. Mitä Pearsonin korrelaatiokerroin arvon suuruus ja etumerkki kertovat riippuvuudesta?
7. Kahvi on useiden kehitysmaiden johtava vientituote. Kun kahvin hinta on korkea, viljelijät kaatavat usein metsää istuttaakseen lisää kahvipuita. Alla on aineisto hinnoista, joita on maksettu kahvin kasvattajille Indonesiassa sekä metsän uudistusosuuksista viiden vuoden aikana (Lähde: Moore & McCabe, 2006).

<i>Hinta</i>	<i>Osuus (%)</i>
29	0.49
40	1.59
54	1.69
55	1.82
72	3.10

- a) Piirrä sirontakuvio. Kumpi muuttujista olisi selittäjä? Tarkastele, missä neljänniksissä havainnot sijaitsevat (ks. alla aputuloksia). Onko muuttujien välillä riippuvuutta sirontakuvion perusteella ja jos on, niin millaista riippuvuus on?
- b) Laske (Pearsonin) korrelaatiokerroin r_{xy} (käytä alla olevia aputuloksia). Tulkitse sanallisesti saamasi tulos.

Aputuloksia: x -havainnoille $\bar{x} = 50$ ja $s_x = 16.32$ sekä y -havainnoille $\bar{y} = 1.74$ ja $s_y = 0.93$.

8. Syksyn 2006 TILP150-kurssilaisiin liittyvän aineistosta tarkastellaan muuttujia

KA = ylioppilastodistuksen keskiarvo

AIKARV = äidinkielen numero

MATARVPAIN = painotettu matematiikan numero

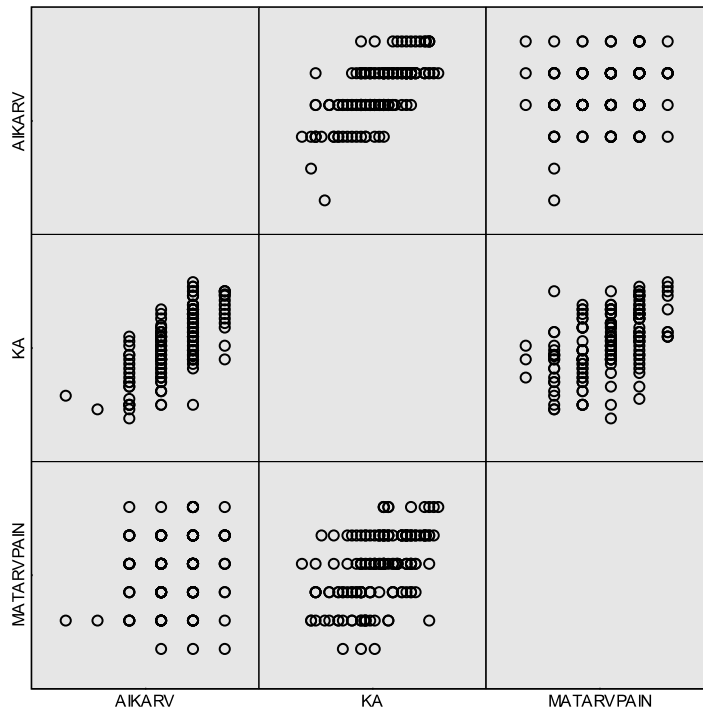
Edellä on muodostettu uusi muuttuja MATARVPAIN = matematiikan numero (asteikko 4-12.5). Se on saatu siten, että pitkän matematiikan numero on kerrottu 1.25:lla ja lyhyen matematiikan numero on käytetty sellaisenaan (painokerroin on esitetty artikkelissa Väisänen & Ylönen (2004): Matemaattiset taidot ja matemaattinen minäkäsitys tilastollisten menetelmien oppimisessa, *Kasvatus*, nro 4).

Tutkitaan muuttujien välisiä riippuvuuksia. Muuttujien kovarianssimatriisi on seuraava:

	KA	AIKARV	MATARVPAIN
KA	0.45	0.40	0.57
AIKARV	0.40	0.77	0.35
MATARVPAIN	0.57	0.35	2.83

- Poimi kovarianssimatriisista kaikkien muuttujien varianssit ja laske vastaavat keskihajonnat.
- Muodosta muuttujien korrelaatiomatriisi.
- Tulkitse sanallisesti korrelaatiomatriisissa esiintyviä korrelaatioita ja kommentoi niiden suuruutta. Tarkastele samalla Kuvassa 2 olevia sirontakuvioita. Siinä sirontakuviot on esitetty ”matriisissa”, jossa kahden muuttujan sirontakuvio sijaitsee ikkunassa, jota sarake ja rivi leikkaa.

Demohyvyitys: 7 tehtävää = 1/2 pistettä, 4 tehtävää = 1/4 pistettä



Kuva 2: Muuttujien KA, AIKARV ja MATARVPAIN sirontakuviot.