

806109 TILASTOTIETEEN PERUSMENETELMÄT I
Harjoitus 5, syksy 2010

1. Kohdissa a)-f) on väittämiä, jotka ovat alla annettujen tietojen perusteella tosia ($=T$) tai epätosia ($=E$).

Erääseen kokeeseen osallistui 40 naista ja 60 miestä. Koe tarkastetaan ja saadaan taulukon 1 mukaiset naisten ja miesten pistejakaumat.

Taulukko 1. Naisten ja miesten pistejakaumat

Koepisteet	Naiset	Miehet
0	5%	10 %
1	20%	15 %
2	30%	20 %
3	35%	40 %
4	10%	15 %

Merkitään Md =koepisteiden mediaani, Md_N =naisten koepisteiden mediaani, Md_M =miesten koepisteiden mediaani.

- a) $Md_M > Md_N$
- b) $Md = 2$
- c) 7.5% osallistujista sai nolla pistettä.
- d) Miesten koepisteiden jakauma on vasemmalle vino.
- e) Koepisteiden keskiarvo on 2.31.
- f) Koepisteiden moodi = koepisteiden yläkvartiili.
2. Erään tentin pistemäärien keskiarvo ja keskihajonta olivat pojilla (yht. 30) ja tytöillä (yht. 20) seuraavat:

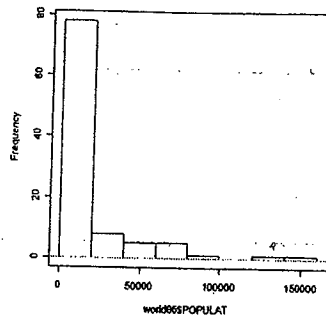
	Pojat	Tytöt
Keskiarvo	14	17
Keskihajonta	2	3

- a) Laske koko aineiston (yht. 50) keskiarvo.
- b) Jos jokaisen pojan pistemäärä kerrotaan (-5):llä ja näin saatuun tulokseen lisätään 10, miten poikien keskiarvo ja keskihajonta muuttuvat?

3. Viidentoista miesopiskelijan lepopulssi (lyöntiä minuutissa) mitattiin ja tulosten keskiarvoksi saatiin 67.6 ja keskihajonnaksi 7.9. Samoilta miehiltä mitattiin sydämenlyöntien lukumäärä myös 400 metrin juoksun jälkeen (= rasiuspulssi). Keskiarvoksi saatiin nyt 165.4 ja keskihajonnaksi 13.9.

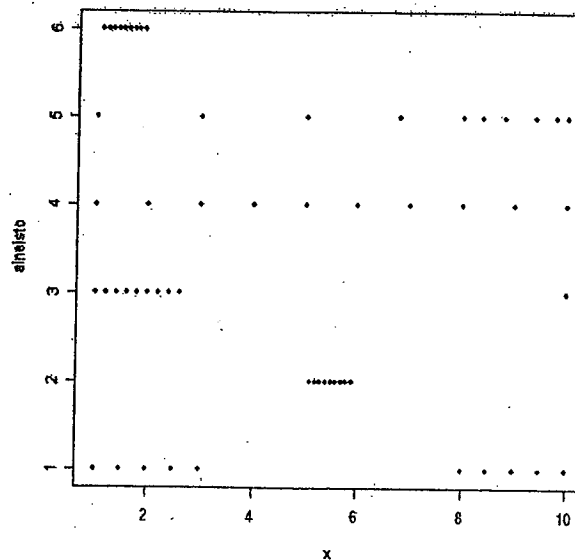
- Kumman muuttujan, lepopulssin vai rasiuspulssin, arvot vaihtelivat enemmän?
- Eeron lepopulssi oli 63 ja rasiuspulssi 159. Kumpi pulsseista oli Eerolla suhteellisesti alhaisempi?

4. Alla on esitetty WORLD2006-aineiston asukasluvun (POPULAT, tuhansia) histogrammi. Kommentoi ko. muuttujan jakauman muotoa (symmetrinen vai vino, jos vino, mihin suuntaan vino)



- histogrammin perusteella,
- vinousmitan g_1 arvon 2.75 ja huipukkuusmitan g_2 arvon 3.64 perusteella.

5. Tunnuksluvut A-F liittyvät alla olevan pistekuvion havaintoaineistoihin 1-6. Yhdistä tunnuksluvut A-F oikeisiin aineistoihinsa.



Aineisto	Tunnusluvut
A	$\bar{x}=5.5$, $Md=5.5$, $s_x=3.8$, $g_1=0.0$
B	$\bar{x}=2.6$, $Md=1.9$, $s_x=2.6$, $g_1=2.1$
C	$\bar{x}=7.0$, $Md=8.2$, $s_x=3.1$, $g_1=-0.8$
D	$\bar{x}=1.5$, $Md=1.5$, $s_x=0.3$, $g_1=0.0$
E	$\bar{x}=5.5$, $Md=5.5$, $s_x=3.0$, $g_1=0.0$
F	$\bar{x}=5.5$, $Md=5.6$, $s_x=0.3$, $g_1=-0.1$

6. Eräessä tutkimuksessa tuhannelta opiskelijalta kysyttiin mm. mieliväriä. Mieliväri oli valittava kolmesta väristä: punainen, keltainen ja sininen. Miehiä tutkimuksessa oli mukana 450 ja heistä punaisen valitsi mieliväriksi 90, keltaisen 135 ja sinisen 225. Mikä tulisi olla mielivärien jakauma tutkimukseen osallistuneilla naisilla, jotta sukupuoli ja mieliväri olisivat toisistaan täysin tilastollisesti riippumattomia? Mikä olisi tässä tilanteessa suhteellisen kontingenssikertoimen arvo?
7. a) Täydennä ristiintaulukko solufrekvensseillä siten, että muuttujat x ja y ovat täysin riippumattomia toisistaan.

		x			yht.
		A	B	C	
y	0				20
	1				30
yht.		25	10	15	50

- b) Täydennä ristiintaulukko solufrekvensseillä siten, että muuttujien x ja y välinen riippuvuus on mahdollisimman voimakasta.

		x		yht.
		A	B	
y	0			42
	1			8
yht.		10	40	50

8. Tutkitaan odottavan äidin tupakoinnin vaikutusta lapsen syntymäpainoon prof. Paula Rantakallion havaintoaineiston avulla:

Lapsen synt.paino	Äidin tupakointi			
	yli 10 sav/p	1-10 sav/p	Ei ollenkaan	
< 2500	23	95	345	463
≥ 2500	253	1331	9360	10944
	276	1426	9705	11407

Tutki riippuvuutta

- a) ehdollisten prosenttijakaumien avulla,
 b) sopivan riippuvuusluvun avulla.