

806109 TILASTOTIETEEN PERUSMENETELMÄT I  
Harjoitus 5, viikko 41, syksy 2010

1. Liitteessä on mikroluokkaharjoituksessa 2 tehtävässä 4b) saatu kuvio yritys2008-aineistosta. Tulkitse ja vertaile kuvion antamia tietoja eri toimialoilta.
2. 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 tarkastettiin ja saatiin 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.
3. 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) Miten tyttöjen keskiarvo ja keskihajonta muuttuisivat, jos jokaisen tytön alkuperäinen pistemäärä kerrottaisiin kymmenellä ja näin saadusta luvusta vähennettäisiin kolme?

4. 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 (= rasisuspulssi). Keskiarvoksi saatiin nyt 165.4 ja keskihajonnaksi 13.9.

a) Kumman muuttujan, lepopulssin vai rasisuspulssin, arvot vaihtelivat suhteellisesti enemmän?

b) Eeron lepopulssi oli 63 ja rasisuspulssi 159. Kumpi pulsseista oli Eerolla suhteellisesti alhaisempi?

5. Yritys73-aineiston muuttujalle sipo (= sijoitetun pääoman tuotto prosentti), jota on käsitelty aiemmin mm. harjoituksen 3 tehtävässä 5, on laskettu R:llä tunnuslukuja ja saatu seuraava tulostus:

```
> numSummary(yritys73[, "sipo"], statistics=c("mean", "sd", "quantiles"),
+ quantiles=c(0, .25, .5, .75, 1))
  mean      sd  0% 25% 50% 75% 100%  n
15.67123 21.60379 -30   5  10  17   87 73
```

```
> skewness(yritys73$sipo)
[1] 1.414039
attr(,"method")
[1] "moment"
```

```
> kurtosis(yritys73$sipo)
[1] 2.537945
attr(,"method")
[1] "excess"
```

a) Mitkä sijaintia ja hajontaa kuvaavat tunnusluvut yo. tulostuksesta selviävät?

b) Tulkitse vinous- ja huipukkuusmitan arvot.

c) Jos pitäisi valita vain yksi (=paras) keskiluku ja vain yksi (=paras) hajontaluku kuvaamaan sijoitetun pääoman tuotto prosenttien jakaumaa, mitkä valitsisit? Perustele vastauksesi.

6. Alla olevassa havaintomatriisissa on kahdentoista liikevaihdoltaan Suomen suurimman energia-alan yrityksen kannattavuutta ja rahoitusrakennetta kuvaavia tunnuslukuja tilikaudelta 2008. Ko. tunnuslukujen tarkat laskukaavat löytyvät mikroluokkaharjoituksen 2 liitteestä.

	yritys	sipo	gearing	omavaraste	te08	te07	nettotulospros
	NesteOil	6	46	46	7.3	9.3	0.7
	Fortum	15	73	41	7.9	8.4	27.1
	Gasum	9	5	55	8.7	10.0	2.2
	PohjolanVoima	0	254	24	4.8	4.9	-6.6
	Vattenfall	8	116	37	6.9	5.7	6.2
	KymppivoimaHankinta	3	599	6	4.6	8.5	0.3
	FingridYhtiöt	6	174	27	6.1	6.4	9.2
	SavonVoima	13	8	76	9.3	9.7	8.7
	RAONordic	60	45	49	9.1	6.9	11.5
	VantaanEnergiä	13	20	70	8.8	8.8	12.8
	TurkuEnergiä	12	31	61	8.5	9.1	7.5
	EVPEnergiä	5	90	43	6.7	6.7	2.4

Tutki graafisesti

- sijoitetun pääoman tuottoprosentin ja omavaraisuusasteen,
- gearingin ja omavaraisuusasteen,
- vuosien 2007 ja 2008 TE-arvosanojen välistä riippuvuutta.

7. a) Täydennä ristiintaulukko solufrekvensseillä siten, että muuttujat  $x$  ja  $y$  ovat täysin riippumattomia toisistaan.

		$x$			yht.
		A	B	C	
$y$	0				300
	1				120
	2				180
yht.		60	420	120	600

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

		$x$		yht.
		A	B	
$y$	0			35
	1			15
yht.		20	30	50

8. 180 osakesijoittajasta selvitettiin tulotaso (= vuositulot luokiteltuna neljään luokkaan) ja osakesalkun riskitaso (alhainen, keskitasoa vai korkea) ja saatiin seuraavat tulokset:

Tulotaso	Riskitaso			Yht.
	alhainen	keskitasoa	korkea	
melko alhainen	11	7	1	19
keskitasoa	22	30	11	63
melko korkea	11	20	20	51
korkea	11	11	25	47
Yht.	55	68	57	180

Tutki sijoittajan tulotason ja osakesalkun riskitason välistä riippuvuutta

- ehdollisten prosenttijakaumien,
- sopivan riippuvuusluvun avulla.

