

806109 TILASTOTIETEEN PERUSMENETELMÄT I  
Harjoitus 5, viikko 41, syksy 2011

HUOM. 1. välikoe on pe 21.10. klo 14-18 salissa L1. Välikokeeseen on ilmoit-  
tauduttava weboodissa 19.10. klo 12 mennessä. Ilmoittautumiskohdassa lukee  
OJ Tilastotieteen perusmenetelmät I s11

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 tau-  
lukon 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 poikien keskiarvo ja keskihajonta muuttuisivat, jos jokaisen pojan alku-  
peräinen pistemäärä kerrottaisiin kuudella 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.
- Kumman muuttujan, lepopulssin vai rasisuspulssin, arvot vaihtelivat suhteellisesti enemmän?
  - Eeron lepopulssi oli 63 ja rasisuspulssi 159. Kumpi pulsseista oli Eerolla suhteellisesti alhaisempi?
5. Yritys2010-aineiston muuttujalle tearv (= TE-arvosana v. 2010), jota on käsitelty aiemmin harjoituksen 3 tehtävässä 5 ja harjoituksen 4 tehtävässä 1, on laskettu R:llä tunnuslukuja ja saatu seuraava tulostus:

```
> numSummary(yritys2010[, "tearv"], statistics=c("mean", "sd", "quantiles"),
+ quantiles=c(0, .25, .5, .75, 1))
  mean   sd  0% 25% 50%  75% 100%  n
7.666 1.457 4.2 6.7 7.8 8.725  10 80

> skewness(yritys2010$tearv)
[1] -0.2221
attr(,"method")
[1] "moment"

> kurtosis(yritys2010$tearv)
[1] -0.7065
attr(,"method")
[1] "excess"
```

- Mitkä sijaintia ja hajontaa kuvaavat tunnusluvut yo. tulostuksesta selviävät?
  - Tulkitse vinous- ja huipukkuusmitan arvot.
  - Jos pitäisi valita vain yksi (=paras) keskiluku ja vain yksi (=paras) hajontaluku kuvaamaan TE-arvosanan 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.

```
> energia
      yritys sipo gearing omavaraste
1      NesteOil      6      46      46
2      Fortum      15      73      41
3      Gasum       9       5      55
4      PohjolanVoima 0     254     24
5      Vattenfall   8     116     37
6      KymppiVoimaHankinta 3   599     6
7      FingridYhtiöt 6     174     27
8      SavonVoima  13       8     76
9      RAONordic   60      45     49
10     VantaanEnergia 13     20     70
11     TurkuEnergia 12     31     61
12     EVPENERGIA   5     90     43
```

Tutki graafisesti

- a) sijoitetun pääoman tuottoprosentin ja omavaraisuusasteen,  
 b) gearingin ja omavaraisuusasteen,  
 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

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

LIITE

