

806109 TILASTOTIETEEN PERUSMENETELMÄT I  
 Taloustieteiden tiedekunnan opiskelijat  
 Harjoitus 5, viikko 41, syksy 2008

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

Eräaseen kokeeseen osallistui 40 naista ja 60 miestä. Koe tarkastettiin ja saatiin tulokson 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%

Merkittään  $Md$ =koepisteiden mediaani,  $Md_N$ =naisten koepisteiden mediaani,  $Md_M$ =miesten koepisteiden mediaani.

- $Md_M > Md_N$
  - $Md = 2$
  - 75% osallistujista sai nolla pistettä.
  - Miesten koepisteiden jakauma on vasemmalle vino.
  - Koepisteiden keskiarvo on 2.31.
  - 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

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

3. Viidentoista miesopiskelijan lepopulssi (työntä minuttissa) mitattiin ja tulosten keskiarvoksi saatiin 67.6 ja keskihajonnaksi 7.9. Samoilta miehiltä mitattiin sydämenlyöntien lukumäärä myös 400 metrin juoksen 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 elintarvikkealan yrityksiä R:llä saatuja tuloksia henkilöstön määrää kuvaavasta muuttujasta (=henkilö). Aineistona *Yleises 2007*.

```
> stem.leaf(elint2007$henkil, style="bare", unit=100, trim.outliers=FALSE,
depths=FALSE, reverse.negative.leave=FALSE, na.rm=TRUE)
[1] "warning: NA elements have been removed!"
1 | 2: represents 1200
Leaf unit: 100
n: 23
```

0	00011444445579
1	0112
2	2
3	3
4	2
5	9
6	7
7	8
8	9
9	10
10	11
11	12
12	13
13	4

Huom!  
 N/A tarkoittaa  
 puuttuvia arvoja.

- Tulkitse runko-lehti-kuvion antamat arvot.
- Mitäk sijaintia ja hajontaa kuvaavat tunnusluvut selviävät b)-kohdan tulostuksesta.

```
> numSummary(elint2007[, "henkil"], statistics=c("mean", "sd", "quantiles"),
quantiles=c(0, .25, .5, .75, 1))
      mean      sd 0%   25% 50%   75% 100%  n NA
2027.696 3203.805 23 419.5 725 1720 13470 23 2
```

```
> skewness(elint2007$henkil, na.rm=T)
[1] 2.25141 = 9_1
```

- Kommentoi jakauman muotoa.
- Mikä keskiarvo ja mikä hajontaluku on mielestäsi paras kuvaamaan henkilöstön määrää kuvaavan muuttujan jakaumaa? Perustele vastauksesi.

5. Aineistona YRITYSS2007  
 Tutki graafisesti metallialan yrityksillä

- a) vuosien 2006 ja 2007 TE-arvosanojen,  
 b) gearingin ja sijoitetun pääoman tuoton välistä riippuvuutta.
6. a) Täydennä ristiintaulukko solufrekvensseillä siten, että muuttujat  $x$  ja  $y$  ovat täysin riippumattomia toisistaan.

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

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

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

7. 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 prosenttijakamien,  
 b) sopivan riippuvuusluvun avulla.