

**806109 TILASTOTIETEEN PERUSMENETELMÄT I**  
**Harjoitus 5, viikko 7, kevät 2012**  
**(Muut kuin taloustieteiden tiedekunnan opiskelijat)**

1. Jatkoa harjoituksen 4 tehtäviin 4 ja 5:

- a) Yhdeksältä miesopiskelijalta mitattiin leposyke (lyönteinä minuutissa) ja havaitut arvot olivat

72 70 58 77 62 63 66 72 81

Laske leposykkeen

a1) keskihajonta, a2) varianssi, a3) variaatiokerroin.

- b) Irtokarkkien yhteispainoarvioiden yksiulotteinen frekvenssijakauma eli suora jakauma on seuraava:

Arvioitu yhteispaino (g)	Frekvenssi	%-osuus
500 – 690	14	24
700 – 890	21	36
900 – 1090	9	15
1100 – 1290	13	22
1300 – 1490	2	3
Yhteensä	59	100

Määrää tarkasteltavan muuttujan

b1) keskihajonta, b2) varianssi.

2. Tero osallistui kilpailuun, joka koostui kahdesta osasta: kahdentoista minuutin aikana juostusta matkasta (=Cooperin testin tuloksesta) ja tikanheitosta (5 heiton yhteistulos). Alla olevassa R-ohjelman tulostuksessa on esitetty 15 kilpailuun osallistuneen juoksutuloksista (juostu matka metreinä) laskettuja tunnuslukuja:

```
> summary(cooper)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
2143   2466   2576   2548   2634   2920
> sd(cooper) # keskihajonta
[1] 186.9356
```

Tikkakisan lopputulokset on esitetty alla olevassa runko-lehti -esityksessä:

```
The decimal point is 1 digit(s) to the right of the |

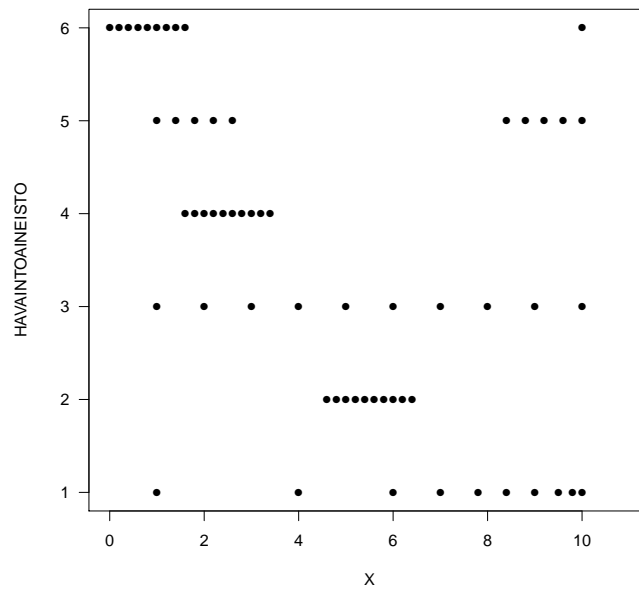
0 | 8
1 | 78
2 | 02235578
3 | 113
4 | 1
```

Laskujen lyhentämiseksi todettakoon, että tikkakisan lopputulosten summa on 371 ja varianssi on 60.9.

- a) Kummassa kisan osassa (juoksussa vai tikanheitossa) tulosten vaihtelu oli suhteellisesti suurempaa?
- b) Teron tulos juoksussa oli 2350 metriä ja tikanheitossa 22. Kummassa kisan osassa Tero menestyi suhteellisesti paremmin?

3. Tunnusluvut A–F liittyvät alla olevan pistekuvion havaintoaineistoihin 1-6, joissa jokaisessa on kymmenen havaintoa. Yhdistä tunnusluvut A–F oikeisiin aineistoihinsa.

Aineisto	$\bar{x}$	Md	$s_x$	$g_1$
A	5.5	5.5	3.9	0.0
B	5.5	5.5	0.6	0.0
C	2.5	2.5	0.6	0.0
D	7.3	8.1	2.9	-0.9
E	1.7	0.9	3.0	2.1
F	5.5	5.5	3.0	0.0



4. Tarkastellaan kahta osa-aineistoa (A ja B), joista suhdeasteikkoa olevan muuttujan  $x$  osalta tiedetään, että

Tunnusluku	Aineisto	
	A	B
$n$	100	100
$x_{(1)}$	3	53
$x_{(50)}$	23	74
$x_{(51)}$	23	74
kvartiiliväli	(17,32)	(65,84)
vaihteluvälin pituus	46	39
keskihajonta	10.5	10.9

- a) Osa-aineistot A ja B yhdistetään yhdeksi 200 tilastoyksikköä sisältäväksi aineistoksi. Muodosta muuttujan  $x$  laatikko-jana -kuvio yhdistetyssä aineistossa.
- b) Osa-aineistossa A kaikille muuttujan  $x$  arvoille tehdään muunnos  $y = -x$ . Määrää näin saadun muuttujan  $y$
- b1) mediaani,    b2) kvartiilivälin pituus,    b3) maksimi,    b4) keskihajonta.

5. Seuraavassa runko-lehti -kuviossa on kuvattu 282 vastasyntyneen napaverestä mitatut seerumin triglyseridipitoisuudet (mmol/l) siten, että esityksen rungolla olevat luvut viittaavat käytetyn mittayksikön kymmenesosiin ja lehdellä olevat luvut sadasosiin, ts. havaintoarvot ovat 0.15, 0.16, ..., 1.64 ja 1.66 mmol/l.  
(Aineiston lähde: <http://www-users.york.ac.uk/~mb55/intro/refint.htm>.)

The decimal point is 1 digit(s) to the left of the |

```

1 | 56
2 | 0000124566677788888888999999
3 | 00000000011222222233333334444444555555666666677778888899999
4 | 00000000000000001111111222222444444445555556666666777778888888888
5 | 00002222223444444555555566666667788889999
6 | 0000000122344444566666778
7 | 00002245566888888
8 | 00222234446788
9 | 5669
10 | 12248
11 | 1
12 | 08
13 |
14 |
15 |
16 | 46

```

- a) Kommentoi tarkasteltavan muuttujan jakauman muotoa yllä esitetyn runko-lehti -kuvion perusteella lyhyesti. Onko jakauma symmetrinen, vino oikealle vai vino vasemmalle?
- b) Kommentoi tarkasteltavan muuttujan jakauman muotoa lyhyesti, kun tiedetään, että muuttujan havaituista arvoista laskettu vinoustunnusluku  $g_1 = 1.72$  ja huipukkuustunnusluku  $g_2 = 4.95$ .
- c) Laske triglyseridipitoisuuden ( $= x$ ) aritmeettinen keskiarvo ja keskihajonta. Hyödynnä laskujen lyhentämiseksi seuraavia apusummia:  $\sum_{i=1}^{282} x_i = 142.66$  ja  $\sum_{i=1}^{282} (x_i - \bar{x})^2 = 13.494$ .  
Ovatko keskiarvo ja keskihajonta tässä tilanteessa hyviä tunnuslukuja kuvaamaan tarkasteltavan jakauman sijaintia ja hajontaa?

6. Jatkoa edelliseen tehtävään: Tässä tehtävässä pyritään määrittelemään vastasyntyneen napaverestä mitatun seerumin triglyseridipitoisuudelle 95% viiteväliä. Viitevälillä tarkoitetaan tässä yhteydessä sellaista triglyseridipitoisuusarvojen väliä, jolle sijoittuu 95% kaikista havaintoarvoista.

- a) Luennoilla esiteltiin muuttujan keskihajonnan ”tulkitsemiseksi” mm. seuraava muistisääntö: Mikäli muuttujan jakauma on normaalijakauman kaltainen, välille  $[\bar{x} - 2s, \bar{x} + 2s]$  sijoittuu noin 95% kaikista havaintoarvoista. Määrää triglyseridipitoisuuden 95% viitearvo edellä esitetyn muistisäännön perusteella. Hyödynnä laskelmasasi edellisen tehtävän c)-kohdan tuloksia.
- b) Määrää triglyseridipitoisuuden 95% viiteväli laskemalla ko. muuttujan jakauman 2.5% ja 97.5%-fraktiilit. Esim. 2.5%-fraktiili on sellainen muuttujan arvo, että sitä pienempiä havaintoarvoja on korkeintaan 2.5 prosenttia ja sitä suurempia on korkeintaan 97.5 prosenttia (vrt. esim. alakvartiili  $Q_1$ , joka on 25%-fraktiili).
- c) Kumpi edellä käytetyistä menetelmistä antaa realistisemmän viitevälin? Miksi?

7. Selvitettäessä tentissä viipymisen (aika minuuteissa) ja tenttituloksen välistä riippuvuutta saatiin seuraavat mittaustulokset:

Opiskelija	1	2	2	4	5	6	7	8	9	10
viipymisaika	200	150	100	180	230	160	170	175	210	190
pistemäärä	24	15	5	22	23	18	19	19	20	21

Tutki muuttujien välistä riippuvuutta graafisesti.

8. Eräessä tutkimuksessa selvitettiin lasten korvatulehduksiin (otititeihin) liittyviä riskitekijöitä. Tutkimuksen mittaustuloksista saatiin muodostettua mm. seuraava ristiintaulukko:

Allergia	korvatulehdusten lkm kahden ensimmäisen ikävuoden aikana			Yhteensä
	0	1-2	vähintään 3	
ei	199	301	220	720
kyllä	22	52	56	130
Yhteensä	221	353	276	850

Tutki lapsen allergisuuden ja lapsen kahden ensimmäisen ikävuoden aikana sairastettujen korvatulehdusten lukumäärän välistä riippuvuutta ehdollisten prosenttijakaumien avulla.

Vastauksia tehtäviin:

- a1) 7.4 a2) 55.3 a3) 0.11 b1) 235.8 b2) 55616.6
- b1) -23 b2) 15 b3) -3 b4) 10.5
- c) 0.506 ja 0.219
- a) [0.07, 0.94] b) [0.22, 1.02]