

806109 TILASTOTIETEEN PERUSMENETELMÄT I
Harjoitus 5, viikko 7, kevät 2013
(Muut kuin taloustieteiden tiedekunnan opiskelijat)

1. Jatkoa harjoituksen 4 tehtäviin 4 ja 2:

- 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) Neljän irtokarkin yhteispainon yksiulotteinen frekvenssijakauma eli suora jakauma on:

Yhteispaino (g)	Frekvenssi	%-osuus
10 – 19	4	5
20 – 29	27	37
30 – 39	34	47
40 – 49	7	10
50 – 59	1	1
Yhteensä	73	100

Määrää tarkasteltavan muuttujan

- b1) keskihajonta, b2) varianssi.

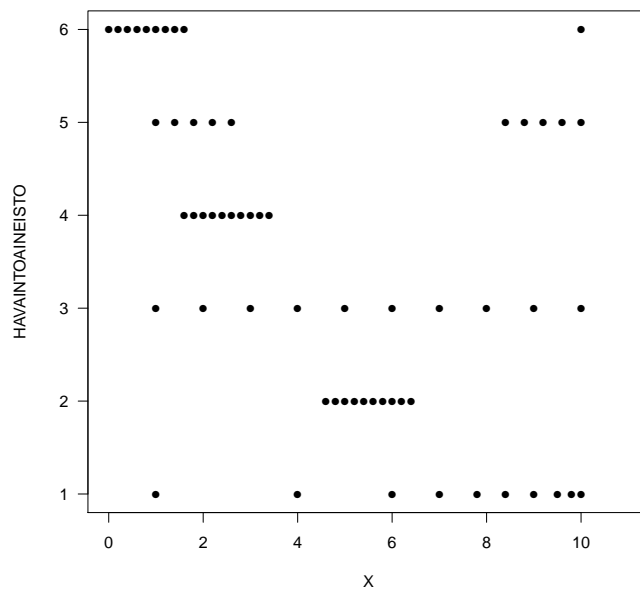
2. Jatkoa tehtävään 1 a): Yhdeksän miesopiskelijan lisäksi leposyke (lyönteinä minuutissa) mitattiin myös kuudeltatoista naisopiskelijalta. Alla on esitetty naisten leposykkeen jakaumaan liittyvien tunnuslukujen arvoja (R-ohjelman tulostusta):

```
> summary(naiset) # 1st. Qu.=alakvartiili, Median=mediaani
                  # Mean=aritm. keskiarvo, 3st. Qu.=yläkvartiili
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
60.00  67.00   73.00   73.19   78.25   90.00
> sd(naiset)     #keskihajonta
8.78
```

- a) Kummalla sukupuolella leposykkeen vaihtelu oli suhteellisesti suurempaa?
- b) Oletetaan, että sukupuoli vaikuttaa ihmisen leposykkeeseen. Olli ja Elli kuuluvat käsiteltävään aineistoon. Sekä Ollilla että Ellillä leposykkeen arvo oli 63 lyöntiä minuutissa. Kummalla heistä leposyke oli suhteellisesti alhaisempi?

3. Tunnusluvut A–F liittyvät seuraavan sivun 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

- Vertaile muuttujan x jakaumia ryhmissä A ja B laatikko-jana -kuvion avulla.
- Osa-aineistot A ja B yhdistetään yhdeksi 200 tilastoyksikköä sisältäväksi aineistoksi. Muodosta muuttujan x laatikko-jana -kuviota yhdistetyssä aineistossa.
- Osa-aineistossa A kaikille muuttujan x arvoille tehdään muunnos $y = -x$. Määrää näin saadun muuttujan y
 - mediaani,
 - kvartiilivälin pituus,
 - maksimi,
 - keskihajonta.

5. Muuttuja x kuvaa lämpötilaa, joka on mitattu celsiusasteina ($^{\circ}C$). Havaintoaineistoon x :stä on saatu viisi havaintoarvoa: -3, 0, 3, -3 ja 3.

- Laske muuttujan x aritmeettinen keskiarvo ja keskihajonta.
- Muunnettaessa celsiusasteet fahrenheitasteiksi ($^{\circ}F$) voidaan käyttää kaavaa $y = 1.8x + 32$, missä x kuvaa lämpötilaa celsiusasteina ja y fahrenheitasteina. Muunna tämän tehtävän viisi lämpötila-arvoa fahrenheitasteiksi.
- Laske b)-kohdassa muodostetun lämpötilamuuttujan aritmeettinen keskiarvo ja keskihajonta. Hyödynnä laskuissasi a)-kohdan tuloksia ja luentomonisteen luvun 4.3.2 huomautusta 3.

6. 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 | 00000000000000001111112222224444444444555555666666677778888888888
5 | 0000222222234444445555555666666667788889999
6 | 00000000122344444566666778
7 | 00002245566888888
8 | 00222234446788
9 | 5669
10 | 12248
11 | 1
12 | 08
13 |
14 |
15 |
16 | 46
```

- Kommentoi tarkasteltavan muuttujan jakauman muotoa yllä esitetyn runko-lehti -kuvion perusteella lyhyesti. Onko jakauma symmetrinen, vino oikealle vai vino vasemmalle?
- Kommentoi tarkasteltavan muuttujan jakauman muotoa lyhyesti, kun tiedetään, että muuttujan havaituista arvoista laskettu vinoustunnusluku $g_1 = 1.72$ ja huipukkuustunnusluku $g_2 = 4.95$.
- 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?

7. Jatkoa edelliseen tehtävään: Tässä tehtävässä pyritään määrittelemään vastasyntyneen naperesta 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?

8. Sähkölämmitteisen loma-asunnon sähkönkulutusta ja ulkoilman lämpötilaa seurattiin seitsemänä vuorokautena. Seurannassa saatiin seuraavat havainnot:

vuorokausi:	1	2	3	4	5	6	7
Sähkön kulutus (kWh):	32	28	23	21	30	28	22
Ulkoilman lämpötila (C):	5	8	12	10	-1	3	7

Tutki muuttujien välistä riippuvuutta graafisesti.

Vastauksia tehtäviin:

1. a1) 7.4 a2) 55.3 a3) 0.11 b1) 7.9 b2) 62.1
4. c1) -23 c2) 15 c3) -3 c4) 10.5
5. a) 0 ja 3 c) 32 ja 5.4
6. c) 0.506 ja 0.219
7. a) [0.07, 0.94] b) [0.22, 1.02]