

806109 TILASTOTIETEEN PERUSMENETELMÄT I
Harjoitus 5, viikko 7, kevät 2011
(Muut kuin taloustieteiden tiedekunnan opiskelijat)

1. Jatkoa harjoituksen 4 tehtäviin 4 ja 5:

- a) Yhdeksältä mies- ja 16 naisopiskelijalta mitattujen leposykkeiden (lyöntiä minuutissa) havaitut arvot on esitetty alla olevissa runko-lehti -kuvioissa (R-ohjelman tulostusta).

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

naiset:	miehet:
6 00348	5 8
7 1224667	6 236
8 224	7 0227
9 0	8 1

Laske **miesten** leposykkeen

- a1) keskihajonta, a2) varianssi, a3) variaatiokerroin.
- b) Erään yrityksen työntekijöiden kuukausipalkan jakauma on

palkka (euroa)	1000–1490	1500–1990	2000–2490	2500–3490	3500–5490	Yhteensä
frekvenssi	3	15	23	8	4	53

Laske työntekijöiden kuukausipalkan

- b1) keskihajonta, b2) varianssi.

2. Jatkoa tehtävään 1 a): 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 on suhteellisesti suurempaa?
- b) Oletetaan, että sukupuoli vaikuttaa ihmisen leposykkeeseen. Olli ja Elli kuuluvat käsiteltävään aineistoon. Sekä Ollilla että Ellillä leposyke on 63 lyöntiä minuutissa. Kummalla heistä leposyke on suhteellisesti alhaisempi?

3. Tarkastellaan kahta osa-aineistoa (A ja B), joista suhteasteikkoa 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.

4. 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 | 00000000011222222233333333444444445555556666666677778888899999
4 | 00000000000000000111111122222244444444445555556666666677778888888888
5 | 000022222223444444555555666666667788889999
6 | 00000000122344444566666778
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 ko. tunnusluvut tässä tilanteessa hyviä kuvaamaan tarkasteltavan jakauman sijaintia ja hajontaa?

5. 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 (ja edellisen tehtävän c)-kohdassa laskettujen ko. tunnuslukujen arvojen) perusteella.
- b) Määrää triglyseridipitoisuuden 95% viiteväli laskemalla ko. muuttujan jakauman 2.5% ja 97.5%-fraktilit. 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?

6. Sähkölämmitteisen loma-asunnon sähkön kulutusta ja ulkoilman lämpötilaa seurattiin seitsemän vuorokautta. Tällöin 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.

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

Allergia	korvatulehdusten lkm			Yhteensä
	kahden ensimmäisen ikävuoden aikana			
	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:

1. a1) 7.4 a2) 55.3 a3) 0.11 b1) 773.6 b2) 598421.6

3. b1) -23 b2) 15 b3) -3 b4) 10.5

4. c) 0.506 ja 0.219

5. a) [0.07, 0.94] b) [0.22, 1.02]