

806109P TILASTOTIETEEN PERUSMENETELMÄT I

1. välikoe 9.3.2012 (Jari Pääkkilä)

VALITSE VIIDESTÄ TEHTÄVÄSTÄ NELJÄ JA VASTAA VAIN NIIHIN!

1. Valitse kohdissa A-F oikea (vain yksi) vaihtoehto. Oikeasta vastauksesta saat +1 pistettä, väärästä et menetä pisteitä. Perusteluita ei vaadita.

A) Muuttujasta x on saatu havaintoaineistoon viisi havaintoa: 2, 2, 0, 4, ja -1. Summalausekkeen $\sum_{i=1}^5 (2x_i - 1)$ arvo on

a1) -1, a2) 2, a3) 4, a4) 8, a5) 9, a6) 13.

B) Konsistenssilla tarkoitetaan mittauksen

b1) harhattomuutta, b2) helppoutta, b3) herkkyyttä,

b4) mitta-asteikkoa, b5) standardointia, b6) tarkkuutta.

C) Otantatutkimuksen kohdepopulaatio koostuu tilastoyksiköistä $\{a, b, c, d, e, f, g, h, i, j\}$ ja kehikkopopulaatio koostuu puolestaan tilastoyksiköistä $\{a, b, c, e, f, g, i, j, k, l\}$. Mikä seuraavista väitteistä on **tos**i?

c1) Jos otanta suoritetaan yksinkertaisella satunnaisotannalla palauttamatta, voi otannan lopputulos olla: b, f, b, j .

c2) Jos otanta suoritetaan yksinkertaisella satunnaisotannalla palauttaen, voi otannan lopputulos olla: b, d, b, j .

c3) Otantatutkimusta ei voi tehdä, koska kohde- ja kehikkopopulaatio eivät vastaa täysin toisiaan.

c4) Otannan ainoa mahdollinen lopputulos on a, b, c, e, f, g, i, j .

c5) Asetelmassa on ns. ylipeittoa, johon kuuluvat tilastoyksiköt k ja l .

c6) Koska kohde- ja kehikkopopulaatio eivät vastaa täysin toisiaan, täytyy kummastakin populaatiosta poimia oma otos.

D) Mikä seuraavista keskihajontaa koskevista väitteistä on **epätosi**?

d1) Keskihajonnan arvo ei muutu, jos kaikista havaintoarvoista vähennetään sama vakio.

d2) Keskihajonta on mittayksiköstä riippumaton tunnusluku.

d3) Keskihajonta ei voi olla koskaan negatiivinen.

d4) Keskihajonta voi olla $= 0$.

d5) Standardoidun muuttujan keskihajonta on aina yksi.

d6) Keskihajonnan laskeminen on sallittua vain välimatka- ja suhdeasteikon muuttujalle.

E) Linnanmaan alueella asui vuoden 2004 alussa 16–18 -vuotiaita 63, 19–24 -vuotiaita 1058, 25–64 -vuotiaita 1329 ja 65-vuotiaita tai sitä vanhempia 26. Tästä populaatiosta poimitaan 200 suuruinen otos ositetulla otannalla (ositteina edellä mainitut ikäryhmät) suhteellista kiintiöintiä käyttäen. Otokseen tulee tällöin 19–24 -vuotiaita

e1) 50, e2) 85, e3) 25, e4) 43, e5) 107, e6) 75.

F) R-ohjelmalla muodostettu runko-lehti –esitys erääseen havaintoaineistoon kuuluvan muuttujan x jakaumasta on seuraava:

```
1 | 2: represents 1.2
leaf unit: 0.1
```

```
-0 | 347899
0 | 1259
1 | 0
2 | 6
3 | 9
```

Esityksen perusteella voidaan päätellä, että muuttujan x

- f1) mitta-asteikko on järjestysasteikko,
- f2) minimi on -0.3,
- f3) maksimi on 39,
- f4) arvo on mitattu kaikkiaan viideltä tilastoyksiköltä,
- f5) jakauma on vino vasemmalle,
- f6) vaihteluvälin pituus on 4.8.

2. Liitteen 1 havaintoaineistoon on laskettu kussakin maakunnassa (= **asuinmaakunta**) korkeakoulututkinnon suorittaneiden (= **koulutusaste**) keskuudessa työllisten prosenttiosuus vuoden 2009 lopussa (= **tyolliset09**) ja työllisten prosenttiosuuden muutos vuoteen 2007 verrattuna (= **muutos09_07**). Vastaa liitteen 1 tietoja apuna käyttäen seuraaviin kysymyksiin.

- a) Mikä on tarkasteltavan aineiston havaintoyksikkö? Ilmoita lisäksi aineiston muuttujista (**asuinmaakunta**, **koulutusaste**, **tyolliset09** ja **muutos09_07**) mitta-asteikko ja onko muuttuja jatkuva vai diskreetti. Perusteluja ei vaadita, mutta voit halutessasi sisällyttää ne vastaukseesi. (2.5 p)
- b) Piirrä laatikko-jana -kuviot (samaa kuvaan!) vuoden 2009 työllisten prosenttiosuuksista erikseen alemman ja ylemmän korkeakoulututkinnon suorittaneille. (2.5 p)
- c) Vertaile ja tulkitse lyhyesti b)-kohdassa piirtämiesi kuvioiden perusteella alemman ja ylemmän korkeakoulututkinnon suorittaneiden vuoden 2009 työllisten prosenttiosuuksien jakaumien sijaintia ja hajontaa. (1 p)

3. Lautapelikerho Lie-or-Die on hankkinut uusia arpakuutiota (tavanomaisia 6-sivuisia noppia). Arpakuutioiden toimivuuden testaamiseksi yhtä uutta arpakuutiota heitettiin yhteensä sata kertaa ja heittojen tulokset (silmäluvut) talletettiin havaintomatriisiin. Näistä heittotuloksista on poimittu seuraava satunnaisotos ($n = 17$):

3, 6, 1, 2, 5, 1, 3, 2, 1, 5, 1, 4, 2, 6, 5, 4, 6

- a) Laske otoksen heittotulosten aritmeettinen keskiarvo ja varianssi. (2 p)
- b) Muodosta otoksen heittotulosten summajakauma ja esitä se graafisesti. (2 p)
- c) Kimmo päätti kokeilla onneaan uuden nopan kanssa ja sai heitollaan silmäluvun 5. Hän halusi verrata tulostaan toiseen noppaan. Kerhotilan lattialta löytyi 10-sivuinen noppa, jota Kimmo heitti saaden silmäluvun 8. Koska kerholla on ollut tapana testata aina uudet noppansa, niin 10-sivuiselle nopalle on tehty myös sadan heiton testaus (kuten uudelle 6-sivuiselle nopalle). 6- ja 10-sivuisien noppien 100 heiton testausarjojen tuloksista on saatu R-ohjelmalla seuraava tulostus:

```
> numSummary(noppakoe, statistics=c("mean", "sd", "quantiles"),
+ quantiles=c(0,.25,.5,.75,1))
              mean      sd  0% 25% 50% 75% 100%    n
6.sivuinen   3.54 1.811   1   2   4   5   6  100
10.sivuinen  x.xx 2.768   1   3   5   7  10  100
```

Kimmon 10-sivuisen nopan heiton standardoitu arvo on 0.979.

- c1) Kumpaa noppaa (6- vai 10-sivuisista) heittäessään Kimmo oli suhteellisesti onnekkaampi (eli sai suhteellisesti suuremman silmäluvun) edellä annetun testi-informaation perusteella? (1 p)
- c2) Edellä R-ohjelman tulostuksesta on peitetty 10-sivuisen nopan sadan heiton testausarjan keskiarvo. Laske peitetty keskiarvo. (1 p)

4. a) Eräessä perusjoukossa on todettu ihmisten nauttivan silloin tällöin alkoholipitoisia juomia. Kyseistä perusjoukkoa koskevassa tutkimuksessa haluttiin selvittää miesten ja naisten tottumuksia keskimääräisessä viikottaisessa alkoholin käytössä. Kerätty aineisto on esitetty alla olevassa ristiintaulukossa.

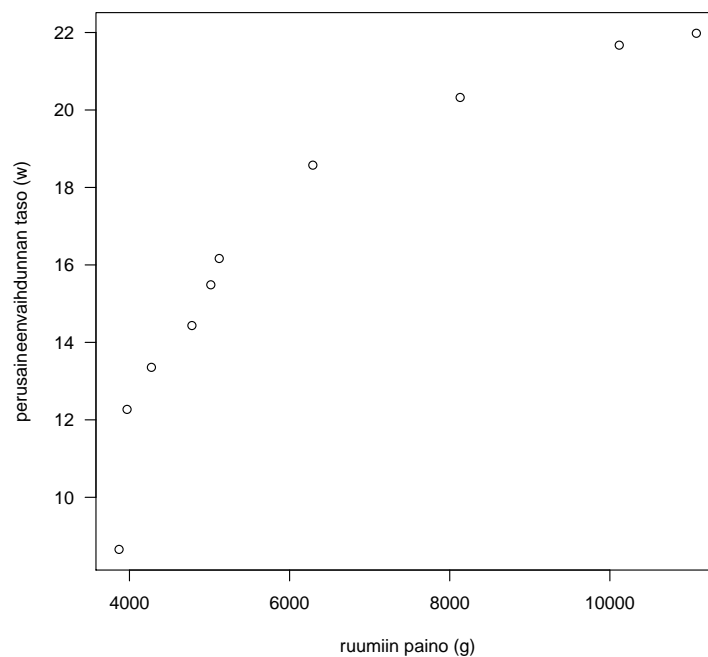
Sukupuoli	Alkoholin käyttökertoja viikossa			Yhteensä
	1	2-3	yli 3	
Miehet	2	9	14	25
Naiset	6	28	25	59
Yhteensä	8	37	39	84

- a1) Tutki (ja kommentoi lyhyesti) tarkasteltavien muuttujien välistä riippuvuutta ehdollisten prosenttijakaumien avulla. (2 p)
- a2) Kuinka monta yli kolme kertaa viikossa alkoholia nauttivaa miestä aineistossa tulisi keskimäärin olla, jos tarkasteltavien muuttujien välillä vallitsisi täydellinen riippumattomuus? (0.5 p)

- b) Linnuilla perusaineenvaihdunnan tason on havaittu riippuvan ruumiin koosta. Todetakaan, että perusaineenvaihdunta on se energia, joka tarvitaan lepotilassa paaston aikana elimistön toimintoihin, sydämen toimintaan, verenkiertoon, hengitykseen, jne. Kymmeneltä lintulajilta mitatut ruumiin painot (=paino, grammoina) ja perusaineenvaihdunnan tasot (=PAV, watteina) ovat seuraavat:

Lintulaji	1	2	3	4	5	6	7	8	9	10
Paino (g)	3870	3970	4274	4780	5016	5121	6290	8130	10117	11080
PAV (w)	8.66	12.27	13.36	14.44	15.49	16.17	18.58	20.32	21.67	21.98

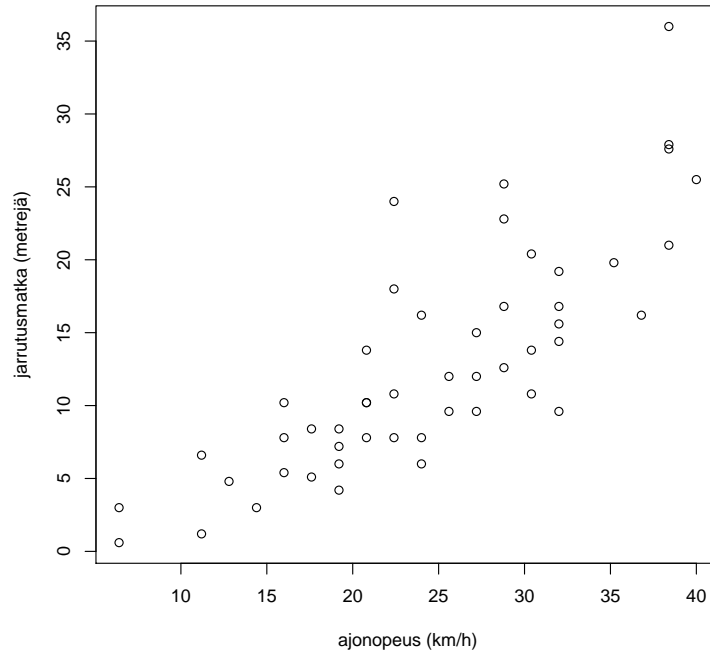
Havaintoaineiston perusteella muodostettu hajontakuviot on puolestaan seuraava:



Tutki muuttujien välistä riippuvuutta tilanteeseen sopivan tunnusluvun avulla. Tulkitse tulos lyhyesti. (3 p)

- c) Oletetaan, että lintujen painot olisikin mitattu kilogrammoina. Tällöin siis esimerkiksi ensimmäisen lintulajin kohdalla painon havaintoarvo olisi alkuperäisen havaintoarvon 3870 g sijasta 3.870 kg. Muuttuisiko b)-kohdassa käyttämäsi tunnusluvun arvo painon mittayksikön muutoksen seurauksena ja jos muuttuisi, niin miten? Perustele vastauksesi lyhyesti. (0.5 p)

5. Museoautojen harrastajat tutkivat kesätapaamisessaan ajonopeuden ja jarrutusmatkan välistä riippuvuutta. Tutkimukseen osallistui 50 museoautoa. Jokaisella autolla ajettiin tutkaan, joka mittasi auton nopeuden (km/h). Välittömästi nopeusmittauspisteen jälkeen auto pysäytettiin jarruttamalla mahdollisimman nopeasti ja jarrutuksen aloittamisesta auton pysähtymiseen kulunut matka mitattiin metreissä. Seuraavassa on esitetty R-ohjelman tulostusta, joka on saatu analysoitaessa tutkimuksessa kerättyä aineistoa.



Call:

```
lm(formula = jarrutusmatka ~ ajonopeus)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.7207	-2.8576	-0.6816	2.7644	12.9604

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.27373	2.02753	-2.601	0.0123
nopeus	0.73733	0.07791	9.464	1.49e-12

Residual standard error: 4.614 on 48 degrees of freedom

Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438

F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

- Tulkitse tulostuksen regressioanalyysin tulokset (määrää regressioyhtälö, kertoimien selväkielinen tulkinta, determinaatikertoimen arvo ja sen tulkinta). (3 p)
- Yhdellä aineistoon kuuluvalla museoautolla ajonopeus oli 20.8 km/h ja jarrutusmatka oli 10.2 metriä. Laske kyseiselle havaintoyksikölle a)-kohdan regressiomallin mukainen vasteen ennuste ja residuaali. (1 p)
- Kyseessä olevassa aineistossa ajonopeuden ja jarrutusmatkan aritmeettiset keskiarvot olivat 24.64 km/h ja 12.89 metriä. Lisäksi tiedetään, että ajonopeuden ja jarrutusmatkan välinen kovarianssi oli 52.775. Sovita aineistoon regressiosuora $y = a + bx$, missä jarrutusmatkalla selitetään ajonopeutta. (2 p)

Liite 1

> tyolliset

	asuinmaakunta	koulutusaste	tyolliset09	muutos09_07
1	Pohjois-Karjala	Ylempi korkeakoulututkinto	80.1	1.5
2	Keski-Suomi	Ylempi korkeakoulututkinto	80.4	0.3
3	Pohjois-Pohjanmaa	Ylempi korkeakoulututkinto	81.0	-2.7
4	Lappi	Ylempi korkeakoulututkinto	81.4	-5.8
5	Varsinais-Suomi	Ylempi korkeakoulututkinto	82.7	-1.4
6	Pirkanmaa	Ylempi korkeakoulututkinto	83.1	-3.9
7	Satakunta	Ylempi korkeakoulututkinto	84.3	-3.1
8	Etelä-Karjala	Ylempi korkeakoulututkinto	84.7	-5.3
9	Kainuu	Ylempi korkeakoulututkinto	86.0	-2.9
10	Keski-Pohjanmaa	Ylempi korkeakoulututkinto	86.6	-0.2
11	Pohjois-Savo	Ylempi korkeakoulututkinto	86.7	-0.4
12	Etelä-Savo	Ylempi korkeakoulututkinto	86.9	4.1
13	Päijät-Häme	Ylempi korkeakoulututkinto	87.4	-1.1
14	Pohjanmaa	Ylempi korkeakoulututkinto	88.3	-1.7
15	Kanta-Häme	Ylempi korkeakoulututkinto	89.0	1.6
16	Etelä-Pohjanmaa	Ylempi korkeakoulututkinto	89.0	-0.1
17	Kymenlaakso	Ylempi korkeakoulututkinto	89.1	-1.9
18	Uusimaa	Ylempi korkeakoulututkinto	89.4	-1.1
19	Itä-Uusimaa	Ylempi korkeakoulututkinto	91.6	4.8
20	Keski-Suomi	Alempi korkeakoulututkinto	57.5	-6.7
21	Etelä-Pohjanmaa	Alempi korkeakoulututkinto	58.7	-5.9
22	Kainuu	Alempi korkeakoulututkinto	59.2	-31.7
23	Pohjois-Karjala	Alempi korkeakoulututkinto	60.1	-3.3
24	Päijät-Häme	Alempi korkeakoulututkinto	60.8	-7.6
25	Pohjois-Pohjanmaa	Alempi korkeakoulututkinto	62.4	-6.4
26	Varsinais-Suomi	Alempi korkeakoulututkinto	64.1	-3.2
27	Pirkanmaa	Alempi korkeakoulututkinto	64.7	-4.1
28	Pohjois-Savo	Alempi korkeakoulututkinto	65.1	-14.7
29	Lappi	Alempi korkeakoulututkinto	67.8	-8.5
30	Etelä-Savo	Alempi korkeakoulututkinto	67.9	-11.1
31	Keski-Pohjanmaa	Alempi korkeakoulututkinto	68.6	-24.3
32	Kymenlaakso	Alempi korkeakoulututkinto	69.0	-10.6
33	Satakunta	Alempi korkeakoulututkinto	69.8	-8.0
34	Pohjanmaa	Alempi korkeakoulututkinto	70.2	0.4
35	Etelä-Karjala	Alempi korkeakoulututkinto	71.6	-8.4
36	Uusimaa	Alempi korkeakoulututkinto	73.5	-4.5
37	Kanta-Häme	Alempi korkeakoulututkinto	74.5	-8.4
38	Itä-Uusimaa	Alempi korkeakoulututkinto	77.4	0.5

```
> numSummary(tyolliset[, "tyolliset09"], groups= tyolliset$koulutusaste,
+ statistics=c("mean", "sd", "quantiles"), quantiles=c(0, .25, .5, .75, 1))
              mean      sd  0%  25%  50%  75% 100%  n
Alempi korkeakoulututkinto 66.46842 5.739924 57.5 61.6 67.8 70.00 77.4 19
Ylempi korkeakoulututkinto xx.xxxxxx x.xxxxxxx xx.x xx.x xx.x xx.xx xx.x 19
```