

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 kehikopopulaatio koostuu puolestaan tilastoyksiköistä $\{a, b, c, e, f, g, i, j, k, l\}$. Mikä seuraavista väitteistä on tosi?

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

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

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

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

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

c6) Koska kohde- ja kehikopopulaatio 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 mitayksiköstä 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) Linnamaan 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 suuruisen otos ositetulla otannalla (ositteina edellä mainitut ikäryhmät) suhteellista kiintiöintä 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 eräeseen 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) korkeakoulututkimnon suorittaneiden (= koulutusaste) keskuudessa työllisten prosenttiosuus vuoden 2009 lopussa (= työlliset09) ja työllisten prosenttiosuuden muutos vuoteen 2007 verrattuna (= muutos09_07). Vastaa liitteen 1 tietoja apuna käyttäen seuraaviin kysymyksiin.

a) Mikä on tarkasteitavan aineiston havaintoyksikkö? Ilmoita lisäksi aineiston muuttujista (asuinmaakunta, koulutusaste, työlliset09 ja muutos09_07) mitta-asteikko ja onko muuttuja jatkuva vai diskreetti. Perusteluja ei vaadita, mutta voit halutessasi sisällyttää ne vastaukseen. (2.5 p)

b) Piirrä laatikko-jana-kuviot (samaa kuvaan!) vuoden 2009 työllisten prosenttiosuuksista erikseen alennan ja ylempään korkeakoulututkimnon suorittaneille. (2.5 p)

c) Vertaile ja tulkitse lyhyesti b)-kohdassa piirtämiä kuvioiden perusteella alennan ja ylempään korkeakoulututkimnon suorittaneiden vuoden 2009 työllisten prosenttiosuusten jakaumien sijaintia ja hajontaa. (1 p)

2. a) Havaintoyksikkö: maakunta

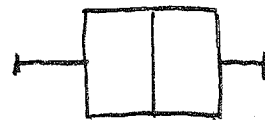
muuttuja mitta-asteikko

asuinmaakunta	luokittelu (diskreetti)
koulutusaste	järjestys (diskreetti)
työlliset 09	suhde (jatkuva)
muutos 09-07	välimatka (jatkuva)

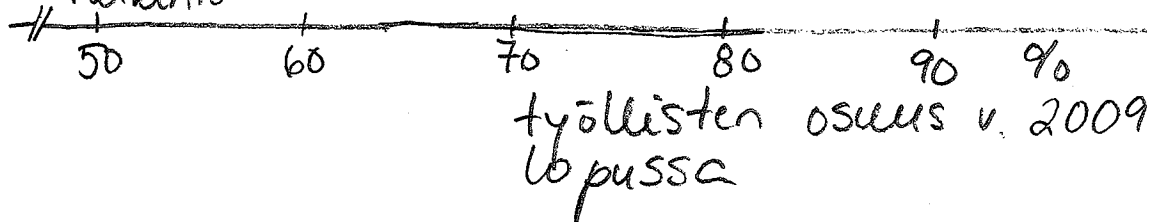
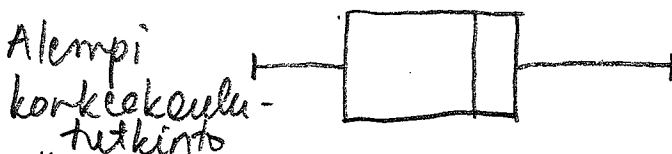
b) Esimerkiksi seuraavasta vastauksesta sai täydet pisteet:

laskimesta: ylempi korkeakoulu tutkinto
min = 80,1 ; $Q_1 = 82,7$; med = 86,6 ;
 $Q_3 = 89$; max = 91,6

Ylempi korkeakoulututkinto



Alempi korkeakoulututkinto



c) Esimerkiksi:

"Ylemmän korkeakoulututkinnon suorittaneiden työllisyysprosentti on yleisesti ottaen korkeampi kuin alemmän korkeakoulututkinnon suorittaneilla."

"Ylemmän korkeakoulututkinnon suorittaneiden työllisyysaste on vähemmän hajontaa."

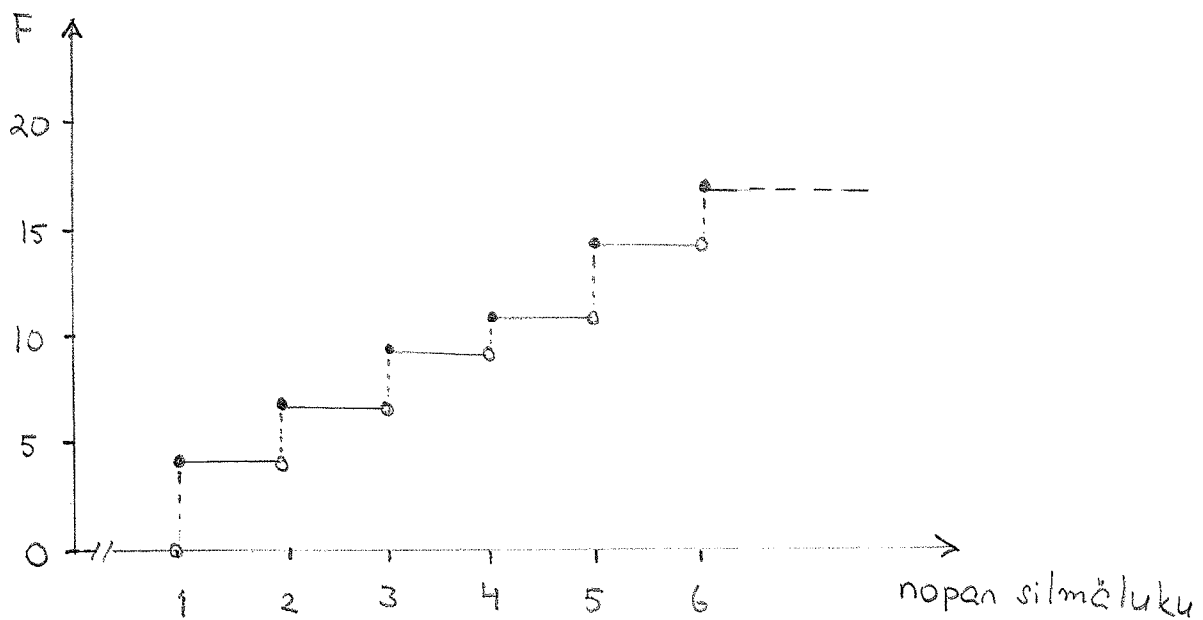
3.) $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \approx \underline{\underline{3,35}}$ } laskimen tilastotaiminnoilla

a) $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \approx 1,9020$

\Rightarrow varianssi: $s^2 \approx 1,9020^2 \approx \underline{\underline{3,62}}$

b)

nopan silmäluku	f	%f	F	%F
1	4	23	4	23
2	3	18	7	41
3	2	12	9	53
4	2	12	11	65
5	3	18	14	83
6	3	18	17	100
Yhteensä	17	100		



c) standardointi: $z_i = \frac{x_i - \bar{x}}{s_x}$

c1) 10-sivuinen noppa: $z_{\text{kimmo}} = 0,979$

6-sivuinen noppa: $z_{\text{kimmo}} = \frac{x_{\text{kimmo}} - \bar{x}}{s_x} = \frac{5 - 3,54}{1,811} \approx 0,806$

\Rightarrow 10-sivuisen nopan standardoitu arvo on suurempi eli kimmo oli suhteellisesti onnekkaampi heittäessään 10-sivuisia noppaa.

c2) Kimmon 10-sivuisen nopan heiton standardoitu arvo oli 0,979,

$$\Rightarrow \frac{X_{\text{kimmo}} - \bar{X}}{s_x} = z_{\text{kimmo}}$$

$$\Rightarrow \frac{8 - \bar{X}}{2,768} = 0,979$$

$$\Rightarrow \bar{X} = 8 - 2,768 \cdot 0,979$$
$$\approx \underline{\underline{5,29}}$$

Tehtävä 4

a1) Tarkastellaan alkoholin käyttöä sukupuolittain.

Sukupuoli	Alk. käyttökertoja / viikko			
	1	2-3	> 3	
Miehet	2	9	14	25
Naiset	6	22	25	53
	8	31	39	78

prosentit riveittäin

$$2/25 = 8\%$$

$$6/53 \approx 11\%$$

$$9/25 = 36\%$$

$$22/53 = 41\%$$

$$14/25 = 56\%$$

$$25/53 = 47\%$$

Ehdolliset prosenttijakaumat:

Sukupuoli	1	2-3	> 3	
Miehet	8	36	56	100
Naiset	11	41	47	100
	10	44	46	100

Ehdolliset prosenttijakaumat näyttävät poikkeavan jonkin verran toisistaan, joten muuttujien välillä on hieman riippuvuutta. Naisten verrattuna suhteellisesti suurempi osa miehistä näyttää kuluttavan yli kolme kertaa viikossa alkoholia, 2-3 kertaa viikossa alkoholia käyttävistä henkilöistä suhteellisesti suurempi osa on naisia.

a2) Täydellisen riippumattomuuden tilanteessa $f_{ij} = e_{ij}$

$$e_{13} = \frac{25 \cdot 39}{78} = 11,607$$
$$\approx \underline{\underline{12}}$$

b) muuttujien välinen riippuvuus on käyrämuotoista, joten käytetään Spearmanin järjestyskorrelaatiokerrointa S

$$S = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad d_i = R(x_i) - R(y_i)$$

Lintulaji (i)	A	B	C	D	E	F	G	H	I	J
$R(x_i)$	1	2	3	4	5	6	7	8	9	10
$R(y_i)$	1	2	3	4	5	6	7	8	9	10
d_i	0	0	0	0	0	0	0	0	0	0
d_i^2	0	0	0	0	0	0	0	0	0	0

$$\Rightarrow \sum_{i=1}^{10} d_i^2 = 0$$

$$S = 1 - \frac{6 \cdot 0}{n^3 - n} = 1 - 0 = \underline{\underline{1}}$$

Tulkinta: muuttujien välillä vallitsee täydellinen positiivinen monotoninen riippuvuus. Mitä painavampi lintu sitä korkeampi perusaineenvaihdunnan taso sillä on.

c) Ei muuttuisi, koska lintujen painon suuruusjärjestys pysyisi samana.

5. Havaintoaineistoon on sovitetty regressiosuora

a) $y = a + bx$, missä y = jarrutusmatka (m) ja
 x = ajonopeus jarrutushetkellä (km/h)

Tulostuksesta: $\hat{y} = -5.27373 + 0.73733x$

tulkinnat: b: Ajonopeuden kasvaessa yhdellä kilometrillä/tunti, kasvaa jarrutusmatka keskimäärin noin 0.74 metriä.

a: kun ajonopeus on jarrutushetkellä nolla km/h, on jarrutusmatka keskimäärin noin -5.27 metriä (järjetön tulkinta käytännön kannalta!)

Determinaatiokerroin $\hat{R}^2 = 0.6511$ eli regressiomallilla (jarrutus-
hetken ajonopeudella) voidaan selittää noin
65.1% jarrutusmatkan kokonaisvaihtelusta.

b) ennuste: $\hat{y} = -5.27373 + 0.73733 \cdot 20.8 \approx \underline{10.1}$ metriä

residuaali: $e = y - \hat{y} \approx 10.2 - 10.1 = \underline{0.1}$ metriä

c) a)-kohdan mallille pätee, että $\hat{R}^2 = r_{xy}^2$ ja $b = s_{xy}/s_x^2 = r_{xy} \cdot s_y/s_x$

Nyt $\hat{R}^2 = r_{xy}^2 = 0.6511 \Rightarrow r_{xy} = \sqrt{0.6511} \approx 0.8069$ (+ etumerkki plus,
koska kovarianssi
on positiivinen)

$b = s_{xy}/s_x^2 \Rightarrow s_x^2 = s_{xy}/b = 52.775/0.73733 \approx 71.576$
 $\Rightarrow s_x \approx 8.46$

$b = r_{xy} \cdot s_y/s_x \Rightarrow s_y = bs_x/r_{xy} \approx \frac{0.73733 \cdot 8.46}{0.8069} \approx 7.73$

Lisäksi tiedetään, että $\bar{y} = 12.89$ ja $\bar{x} = 24.64$.

Nyt y ja x vaihtavat "rooleja" a)-kohtaan verrattuna, joten

$\bar{y} = 24.64$, $\bar{x} = 12.89$, $s_y \approx 8.46$, $s_x \approx 7.73$, $s_{xy} = s_{yx} = 52.775$ ja $r_{xy} = r_{yx} \approx 0.8069$

$\Rightarrow b = s_{xy}/s_x^2 \approx 52.775/7.73^2 \approx \underline{0.88}$

$a = \bar{y} - b\bar{x} \approx 24.64 - 0.88 \cdot 12.89 \approx \underline{13.30}$