

806109P TILASTOTIETEEN PERUSMENETELMÄT I

1. välikoe 21.10.2011 (Marjatta Mankinen)

VALITSE VIIDESTÄ TEHTÄVÄSTÄ NELJÄ JA VASTAA VAIN NIIHIN!

1. Valitse kohdissa A–F oikea vaihtoehto (vain yksi oikea!). Perusteluja ei vaadita, mutta ne on kuitenkin hyvä laittaa näkyviin. Jokaisessa kohdassa saat oikeasta vastauksesta pisteen, väärästä vastauksesta et menetä pisteitä.

A) Lottoarvonta voidaan tulkita

- a1) ositetuksi otannaksi,
- a2) kiintiöpöiminnäksi,
- a3) systemaattiseksi otannaksi,
- a4) ryväotannaksi,
- a5) yksinkertaiseksi satunnaisotannaksi palauttaen,
- a6) yksinkertaiseksi satunnaisotannaksi palauttamatta.

B) Mikä seuraavista termeistä **ei liity** kokeelliseen tutkimukseen?

- b1) käsittely,
- b2) satunnaistus,
- b3) tasainen kiintiöinti,
- b4) kaltaistetut parit,
- b5) yhdistely- eli tekijäkoe,
- b6) lohkojen muodostaminen.

C) Kuntosalin naisasiakkaiden lukumäärä on muuttuja, joka on

- c1) diskreetti ja välimatka-asteikkoa,
- c2) jatkuva ja välimatka-asteikkoa,
- c3) diskreetti ja suhdeasteikkoa,
- c4) jatkuva ja suhdeasteikkoa,
- c5) diskreetti ja järjestysasteikkoa,
- c6) diskreetti ja luokitteluasteikkoa.

- D) Erään yrityksen työntekijöiden (60) päivittäisiä työaikoja on tarkasteltu yhden viikon ajalta. Seuraavassa taulukossa on esitetty työntekijöiden ylityöpäivien (= normaalin työajan ylittävien päivien) lukumäärän frekvenssijakauma ko. viikolla.

Ylityöpäivien lkm	frekvenssi
0	3
1	11
2	31
3	13
4	2
5	0
Yhteensä	60

Ylityöpäivien lukumäärän keskihajonta (kahden desimaalin tarkkuudella) on

- d1) 1.87 d2) 0.86 d3) 11.52
d4) 0.41 d5) 0.75 d6) 2.41

- E) Mikä seuraavista väitteistä on **epätosi**?

- e1) Keskihajonta ei voi koskaan olla negatiivinen.
e2) Keskihajonta voi olla = 0.
e3) Keskihajonnan laskeminen on sallittu vain välimatka- ja suldeasteikon muuttujille.
e4) Keskihajonnan arvo ei muutu, jos kaikista havaintoarvoista vähennetään sama vakio.
e5) Keskihajonta on mittayksiköstä riippumaton.
e6) Standardoidun muuttujan keskihajonta on aina yksi.

- F) Suhdeasteikkoa olevan muuttujan x arvoista (kaikki eri suuria kokonaislukuja) on saatu R:llä seuraava tulostus:

```
> numSummary(data[, "x"], statistics=c("mean", "sd", "quantiles"),  
+ quantiles=c(0, .25, .5, .75, 1))  
mean      sd 0%   25%  50%  75% 100%  n NA  
44.7 18.97117 10 26.75 44.5 61.25 78 20 1
```

Tulostuksesta voidaan päätellä, että

- f1) x:n jakauma on selvästi oikealle vino,
f2) x:n arvoista yksi on outlier (=poikkeava havainto),
f3) x:n variaatiokerroin on 0.772,
f4) viisi x:n arvoista sijoittuu välille (26.75, 44.5),
f5) x:n arvoista välille (26.75, 61.25) sijoittuu noin 68.4%,
f6) x:n vaihteluväli on 68.

Vain yksi eo. väitteistä on **tosi**, mikä?

2. Liitteessä 1 on annettu havaintomatriisi, jossa havaintoyksikköinä ovat Euroopan unionin maat ja muuttujina euroalueeseen kuuluminen: 1 = kuuluu, 2 = ei kuulu (=euro), työttömyysaste kesäkuussa 2011 (=työttaste), bruttokansantuotteen volyymin kasvuprosentti vuodesta 2009 vuoteen 2010 (= gdp10) ja kansainvälisen Standard & Poor's -luottoluokituslaitoksen elokuussa 2011 antama luottoluokka (= luottoluok). Standard & Poor's luokittelee maat asteikolla AAA (korkein mahdollinen), AA+, AA, AA-, ..., CC, SD, D (alin mahdollinen).

- a) Ilmoita aineiston muuttujista (euro, työttaste, gdp10 ja luottoluok) mitta-asteikko ja onko muuttuja jatkuva vai diskreetti. (2 p)
- b) Piirrä (samaa kuvioon!) euromaiden (=euroalueeseen kuuluvien) ja ei-euromaiden (= euroalueeseen kuulumattomien) työttömyysasteen laatikko-jana -kuviot. (2.5 p)
- c) Vertaile ja tulkitse lyhyesti piirtämiesi kuvioiden perusteella euromaiden ja ei-euromaiden työttömyysasteen jakaumien c1) sijaintia, c2) hajontaa ja c3) vinoutta. (1.5 p)

3. (tehtävän 2 aineisto)

- a) Luokittele luottoluokka seuraavasti: (esitetty Kalevassa 3.8.2011)
 AAA= erinomainen luottokelpoisuus,
 AA+, AA, AA- = erittäin hyvä luottokelpoisuus,
 A+, A, A- = vahva luottokelpoisuus,
 BBB+, BBB, BBB- = välttävä luottokelpoisuus,
 BB+, BB, BB- = takaisinmaksu todennäköinen.
 B+, B, B- = korkea luottoriski,
 CCC+, CCC, CCC- = hyvin korkea luottoriski,
 CC = lähellä maksuhäiriötä,
 SD, D = maksuhäiriössä
 ja muodosta tähän luokitteluun perustuva EU-maiden luottoluokan frekvenssijakauma elokuussa 2011. (1 p)
- b) Esitä a)-kohdan frekvenssijakauma graafisesti. (1 p)
- c) Määrä (ja laske) sopivat keski- ja hajontaluvut a)-kohdan jakaumalle. (2 p)
- d) Muodosta ristiintaulukko, jossa muuttujina ovat euroalueeseen kuuluminen ja luottoluokka luokiteltuna kahteen luokkaan (erinomainen, jokin muu) ja tutki näiden muuttujien välistä riippuvuutta ristitulosuhteen avulla. (2 p)

4. Tarkastellaan 59 Helsingin pörssissä noteeratun pörssiyhtiön markkina-arvoa (milj. euroa) sekä viimeisen kuukauden ja vuoden kurssimuutosta (%) 13.9.2011. (Lähde: Talouselämä nro 31, 2011)

- a) Näihin muuttujiin liittyvä kuvio on liitteessä 2 ja alla on R-ohjelman tulostusta.

```
> round(cov(osakkeet[,c("markkinaarvo", "kurssimuutos1kk",
+ "kurssimuutos1v")]),1)
              markkinaarvo kurssimuutos1kk kurssimuutos1v
markkinaarvo    1 051 024.9          -115.8          5 130.9
kurssimuutos1kk    -115.8           66.3           66.0
kurssimuutos1v     5 130.9           66.0          367.8
```

Laske ja tulkitse lyhyt- ja pitkäaikaisen kurssimuutoksen välinen Pearsonin tulomomenttikorrelaatiokerroin. (1.5 p)

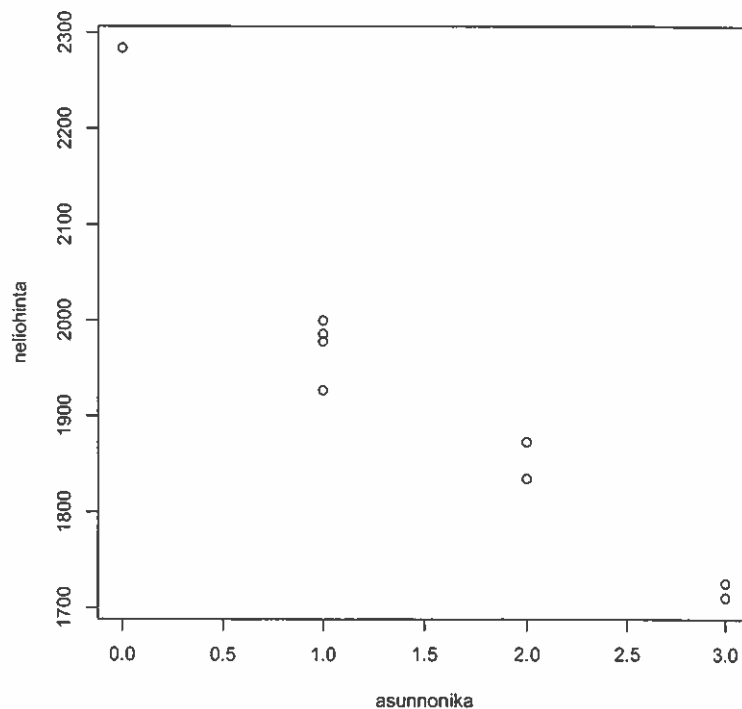
- b) Pörssiyrityksiä poimittiin satunnaisotannalla 11 yhtiötä. Alla olevassa havaintomatriisissa on näiden yhtiöiden edeltävän kuukauden kurssimuutokset (%) ja markkina-arvot (milj. euroa). Tutki ja kommentoi ko. muuttujien välistä riippuvuutta tässä otoksessa

b1) sopivan graafisen esityksen avulla. (1.5 p)

b2) laskemalla ja tulkitsemalla Spearmannin järjestyskorrelaatiokerroin. (3 p)

Pörssiyritys	Kurssimuutos 1 kk (%)	Markkina-arvo (milj. euroa)
Basware	-26	230
Cramo	-30	272
Kemira	-13	1 320
Kesko	-16	2 277
Metso	-14	3 514
Neste Oil	-7	1 820
Pöyry	-23	345
Ramirent	-27	473
Sponda	-9	821
Talvivaara	-16	806
Wärtsilä	-2	3 598

5. Kymmenen lokakuussa 2010 Oulussa myynnissä olevan kerrostaloasunnon aineiston analysointi R-ohjelmalla tuotti seuraavan tulostuksen (osa tuloksista peitetty merkinnällä xx.xx).



```
> summary(Model)
```

```
Call:
```

```
lm(formula = neliohinta ~ asunnonika)
```

```
Residuals:
```

```
    Min      1Q  Median      3Q     Max
-85.12 -23.81 -12.12  20.68 111.65
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2172.35      33.17   65.490 3.29e-12
asunnonika      xx.xx      18.84   -8.505 2.80e-05
---
```

```
Residual standard error: 54.93 on 8 degrees of freedom
```

```
Multiple R-squared: xx.xx, Adjusted R-squared: xx.xx
```

```
F-statistic: 72.34 on 1 and 8 DF, p-value: 2.803e-05
```

Aineiston muuttujat ovat neliöhinta (euroa/neliö)=(neliohinta) ja asunnon ikä vuosina=(asunnonika). Aineistosta laskettu Pearsonin tulomomenttikorrelaatiokerroin on -0.949. Neliöhinnan varianssi on 26931 ja asunnon iän varianssi on 0.9444.

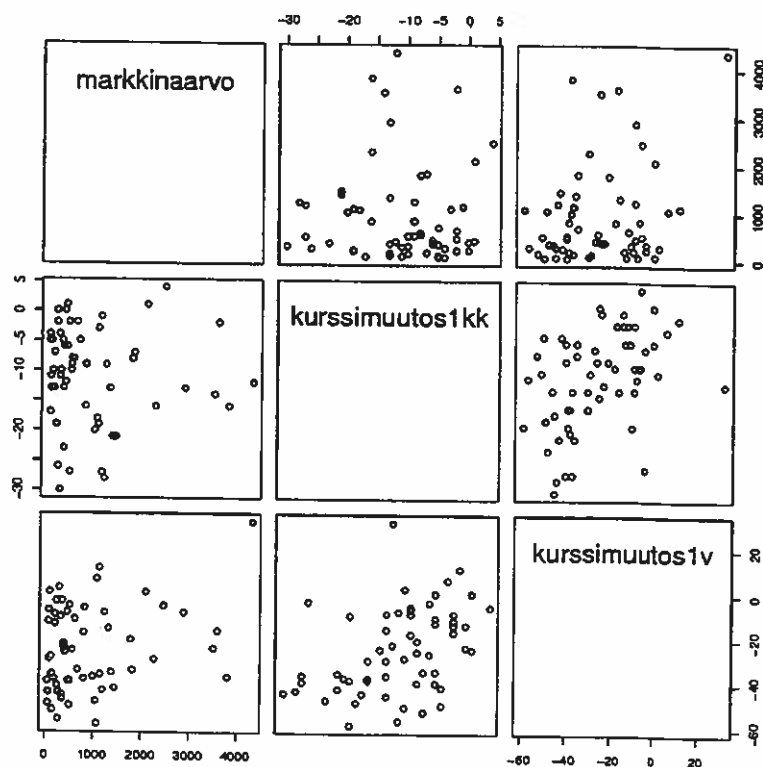
- Onko Pearsonin tulomomenttikorrelaatiokerroin käyttäminen mielekästä/luvallista tälle aineistolle? Perustele lyhyesti vastauksesi. Kumpi aineiston muuttujista on tulostuksen regressiomallissa selitettävä ja kumpi selittävä muuttuja? (1.5 p)
- Määrää regressioyhtälö ja tulkitse sen kertoimet. (2.5 p)
- Määrää ja tulkitse myös regressioyhtälön determinaatikerroin (selitysaste). (1 p)
- Laske mallin mukainen ennustearvo, kun asunnon neliöhinta on 2100 euroa ja ikä on 2.5 vuotta. (1 p)

LIITE 1

> EUmaat

	maa	euro	tyottaste	gdp10	luottoluok
1	Itävalta	1	3.9	2.3	AAA
2	Belgia	1	7.0	2.3	AA+
3	Kypros	1	7.0	1.1	BBB+
4	Viro	1	12.8	2.3	AA-
5	Suomi	1	7.8	3.6	AAA
6	Ranska	1	9.8	1.5	AAA
7	Saksa	1	6.1	3.7	AAA
8	Kreikka	1	16.7	-3.5	CC
9	Irlanti	1	14.4	-0.4	BBB+
10	Italia	1	8.0	1.3	A+
11	Luxemburg	1	4.6	2.7	AAA
12	Malta	1	6.7	2.7	A
13	Hollanti	1	4.1	1.7	AAA
14	Portugal	1	12.5	1.4	BBB-
15	Slovakia	1	13.3	4.0	A+
16	Slovenia	1	7.9	1.4	AA
17	Espanja	1	21.0	-0.1	AA
18	Bulgaria	2	11.5	0.2	BBB
19	Tsekki	2	6.8	2.7	A
20	Tanska	2	7.3	1.7	AAA
21	Unkari	2	10.9	1.3	BBB-
22	Latvia	2	16.2	-0.3	BB+
23	Liettua	2	15.6	1.4	BBB
24	Puola	2	9.5	3.9	A-
25	Romania	2	7.3	-1.9	BB+
26	Ruotsi	2	7.4	5.6	AAA
27	Britannia	2	8.0	1.8	AAA

LIITE 2



KAAVAKOKOELMA

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^r f_i X_i \quad (1)$$

$$G = \sqrt[n]{x_1 x_2 \dots x_n} \quad (2)$$

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (3)$$

$$K = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|, \quad K = \frac{1}{n} \sum_{i=1}^r f_i |X_i - \bar{x}| \quad (4)$$

$$s = s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (5)$$

$$s = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right)}, \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^r f_i (X_i - \bar{x})^2}$$

$$V = s/\bar{x} \quad (6)$$

$$z_i = \frac{x_i - \bar{x}}{s} \quad (7)$$

$$H/H_{max} = \frac{-\sum_{i=1}^r p_i \log p_i}{\log r} \quad (8)$$

$$g_1 = \frac{m_3}{s^3}, \quad m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (9)$$

$$g_2 = \frac{m_4}{s^4} - 3, \quad m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \quad (10)$$

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^r \frac{(f_{ij} - e_{ij})^2}{e_{ij}}, \quad e_{ij} = \frac{f_{i.} f_{.j}}{n} \quad (11)$$

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \quad (12)$$

$$C_{\max} = \sqrt{\frac{q-1}{q}}, \quad q = \min(m, r) \quad (13)$$

$$v = \sqrt{\frac{\chi^2}{nu}}, \quad u = \min(m-1, r-1) \quad (14)$$

$$\text{OR} = \frac{f_{11}f_{22}}{f_{12}f_{21}} \quad (15)$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right) \quad (16)$$

$$r = r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (17)$$

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (18)$$

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (19)$$

$$\hat{y} = a + bx, \quad b = \frac{s_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x}, \quad a = \bar{y} - b\bar{x} \quad (20)$$