Tilastotieteen perusmenetelmät I
1. välikoe 21.10.2011

Tehtävien ratkaisut

(1.) A a6

B b3

C c3

D d2

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

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{r} f_i x_i = \frac{1}{60} (3 \cdot 0 + 11 \cdot 1 + 31 \cdot 2 + 13 \cdot 3 + 2 \cdot 4) = 2$$

$$s = \sqrt{\frac{1}{60-1} \left[3(0-2)^2 + 11(1-2)^2 + 31(2-2)^2 + 13(3-2)^2 + 2(4-2)^2 \right]}$$

$$= \sqrt{\frac{1}{59} \cdot 44} = \underline{0.86}$$

E e5

F f4

2. a) euro: luokitteluasteikko, diskreetti
työttaste: suhdeasteikko, jatkuva
gdpi0: välimatka-asteikko, jatkuva
luottoluok: järjestysasteikko, diskreetti 2p

b) Euromaiden työttömyysasteet suuruusjärjestyksessä

3.9, 4.1, 4.6, 6.1, 6.7, 7.0, 7.0, 7.8, 7.9, 8.0, 9.8, 12.5, 12.8, 13.3, 14.4, 16.7, 21.0

min: $\underline{3.9}$, max = $\underline{21.0}$, Md: $\frac{n}{2} = \frac{17}{2} = 8.5 \Rightarrow$ Md = 9. arvo = $\underline{7.9}$

Q_1: $\frac{n}{4} = \frac{17}{4} = 4.25 \Rightarrow Q_1 = 5.$ arvo = $\underline{6.7}$

Q_3: $\frac{3 \cdot n}{4} = \frac{3 \cdot 17}{4} = 12.75 \Rightarrow Q_3 = 13.$ arvo = $\underline{12.8}$

Ei-euromaiden työttömyysasteet suuruusjärjestyksessä

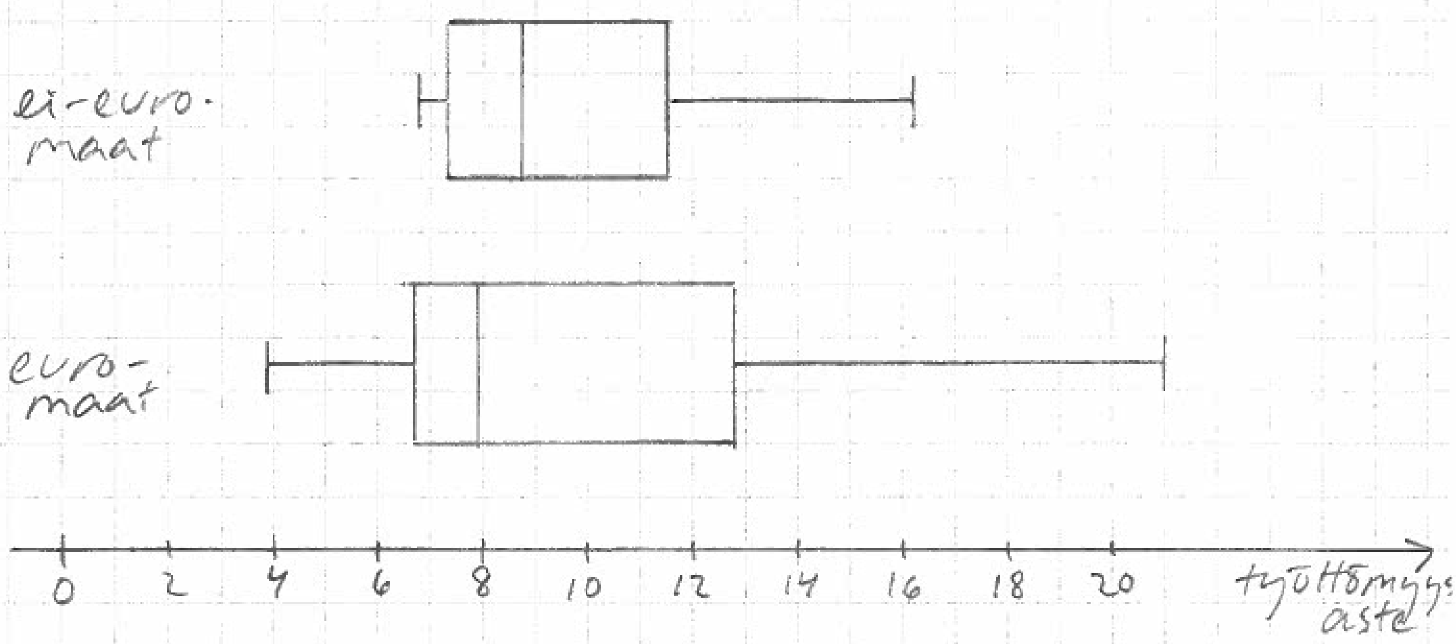
6.8, 7.3, 7.3, 7.4, 8.0, 9.5, 10.9, 11.5, 15.6, 16.2

min = $\underline{6.8}$, max = $\underline{16.2}$

Md: $\frac{10}{2} = 5 \Rightarrow$ Md = $\frac{5.\text{arvo} + 6.\text{arvo}}{2} = \frac{8.0 + 9.5}{2} = \underline{8.75}$

Q_1: $\frac{10}{4} = 2.5 \Rightarrow Q_1 = 3.$ arvo = $\underline{7.3}$

Q_3: $\frac{3 \cdot 10}{4} = 7.5 \Rightarrow Q_3 = 8.$ arvo = $\underline{11.5}$



Euroalueeseen kuuluvien ja euroalueeseen kuulumattomien EU-maiden työttömyysasteen (kesäkuussa 2011) laatikko-jana-kuviot $2\frac{1}{2}$p

c) c1) sijainti: Ei-euromaiden työttömyysaste on keskimäärin korkeampi kuin euromaiden ($Md_{ei\text{-}euro} = 8.75$ ja $Md_{euro} = 7.9$)

c2) hajonta: Työttömyysasteen hajonta on sekä vaihteluvälin että kvartiilivälin pituudella mitattuna suurempaa euromaissa.

$w_{euro} = 21.0 - 3.9 = 17.1$, $w_{ei\text{-}euro} = 16.2 - 6.8 = 9.4$

$Q^*_{euro} = 12.8 - 6.7 = 6.1$, $Q^*_{ei\text{-}euro} = 11.5 - 7.3 = 4.2$

c3) vinous: Sekä euromaiden että ei-euromaiden työttömyysasteen jakaumat ovat oikealle vinoja.

$1\frac{1}{2}$p

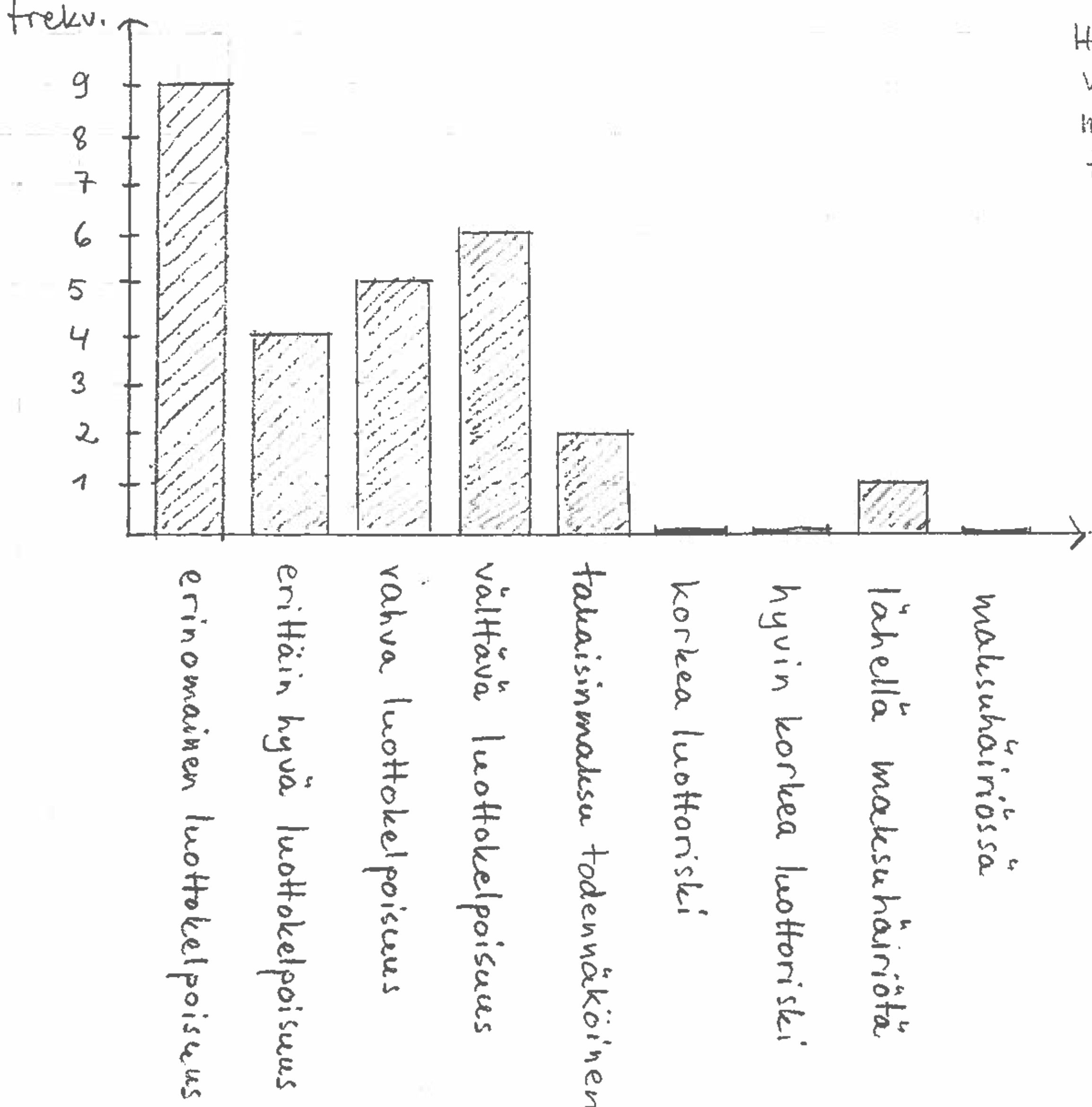
Taulukko 1

③ a) Euroopan unionin maiden luottoluokan frekvenssijakauma elokuussa 2011

Luottoluokka	Frekvenssi
1. erinomainen luottokelpoisuus	9
2. erittäin hyvä luottokelpoisuus	4
3. vahva luottokelpoisuus	5
4. välttävä luottokelpoisuus	6
5. takaisinmaksu todennäköinen	2
6. korkea luottoriski	–
7. hyvin korkea luottoriski	–
8. lähellä maksuhäiriötä	1
9. maksuhäiriössä	–
Yhteensä	27

1p

b) Luottoluokka järjestysasteikkoa ⇒ pylväskuvio



Huom. luokat voi laittaa myös päinvastaiseen järjestykseen

1p

Kuvio 1. Euroopan unionin maiden luottoluokan frekvenssijakauma elokuussa 2011

c) Sopivat keskiluvut moodi $=Mo$ ja mediaani $=Md$
Sopivat hajontaluvut vaihteluväli W ja kvartiiliväli Q

Mo = erinomainen luottokelpoisuus

$\frac{n}{2} = \frac{27}{2} = 13.5 \Rightarrow$ Md = 14. arvo = vahva luottokelpoisuus

W = [erinomainen luottokelpoisuus, lähellä maksuh.]

tai

W = [lähellä maksuh., erinomainen luottokelpoisuus]

$Q = [Q_1, Q_3]$

$\frac{n}{4} = \frac{27}{4} = 6.75 \Rightarrow Q_1 = 7.$ arvo = erinomainen luottokelp.

$\frac{3n}{4} = \frac{3 \cdot 27}{4} = 20.25 \Rightarrow Q_3 = 21.$ arvo = välttävä luottokelp.

Q = [erinomainen luottokelp., välttävä luottokelp.]

tai jos lähtee lopusta

Q = [välttävä luottokelp., erinomainen luottokelp.]

2p

d)

Luotto-luokka	Euroalueeseen kuuluminen kuuluu	ei kuulu	Yhteensä
erinomainen	6	3	9
jokin muu	11	7	18
Yhteensä	17	10	27

1p

Ristitulosuhde $OR = \frac{f_{11} \cdot f_{22}}{f_{12} \cdot f_{21}} = \frac{6 \cdot 7}{3 \cdot 11} = \frac{42}{33} = \underline{1.27}$ $\frac{1}{2}$p

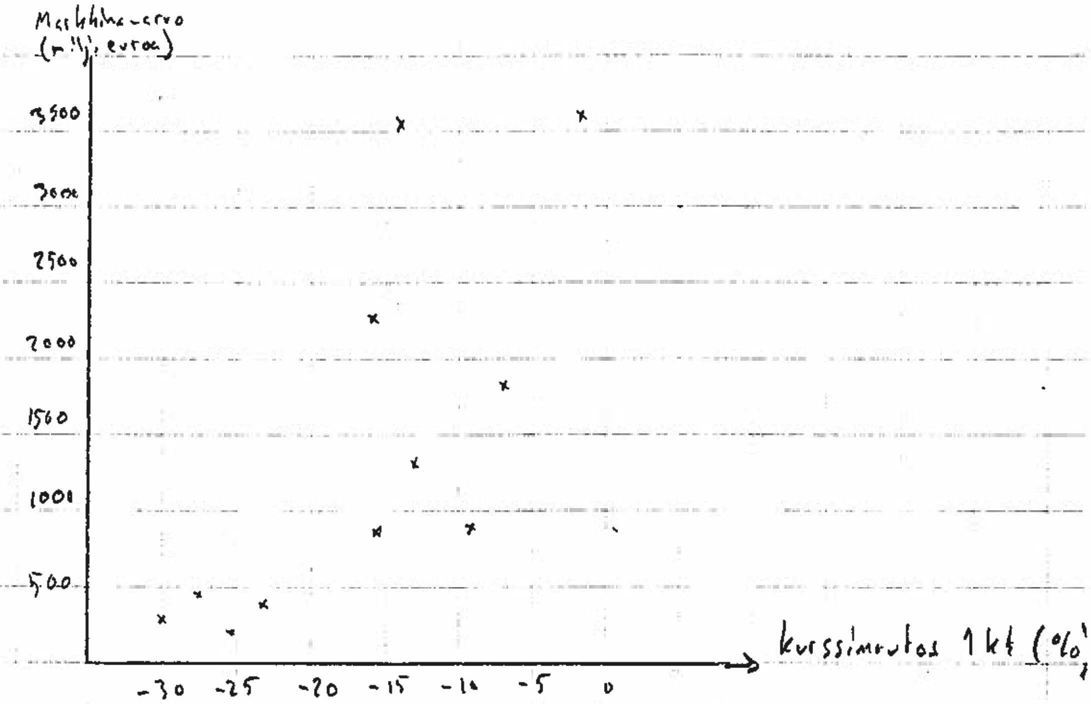
Jos OR = 1, muuttujien välillä ei ole riippuvuutta

Koska OR = 1.27 ≠ 1, luottoluokan ja euroalueeseen kuulumisen välillä on riippuvuutta. $\frac{1}{2}$p

4. a) $r = \frac{s_{xy}}{s_x s_y} = \frac{66.0}{\sqrt{66.3} \cdot \sqrt{367.9}} = 0,422650 \approx 0,42$

Lyhytaikaisen kurssimuutoksen ja pitkäaikaisen kurssimuutoksen välillä on kohtalaista positiivista lineaarista riippuvuutta ko. aineistossa

b) 1)



kurssimuutoksen ja markkina-arvon välillä on positiivista monotonista riippuvuutta ko. aineistossa

4, b2)

Pörssiyhtiö	Kurssimuutos 1kk	$R(x_i)$	Markkina-arvo (milj. euroa)	$R(y_i)$	$d_i = R(x_i) - R(y_i)$
Basware	-26	9	230	11	-2
Cramo	-30	11	272	10	1
Kemira	-13	4	1320	5	-1
Kesko	-16	6,5	2277	3	3,5
Metso	-14	5	3514	2	3
Neste Oil	-7	2	1820	4	-2
Pöyry	-23	8	345	9	-1
Ramirent	-27	10	473	8	2
Sponda	-9	3	821	6	-3
Talvivaara	-16	6,5	806	7	-0,5
Wärtsilä	-2	1	3598	1	0

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6\left((-2)^2 + 1^2 + (-1)^2 + 3{,}5^2 + 3^2 + (-2)^2 + (-1)^2 + 2^2 + (-3)^2 + (-0{,}5)^2 + 0^2\right)}{11^3 - 11}$$

$$= 1 - \frac{6 \cdot (4 + 1 + 1 + 12{,}25 + 9 + 4 + 1 + 4 + 9 + 0{,}25 + 0)}{1331 - 11} = 1 - \frac{6 \cdot 45{,}5}{1320} = 0{,}793181\ldots$$

$$\approx 0{,}79$$

kurssimuutoksen ja markkina-arvon välillä on melko vahvaa positiivista riippuvuutta ko. aineistossa

5. a) Tarkasteltavat muuttujat ovat suhdeasteikkoa ja niiden välillä on lineaarista riippuvuutta sirontakuvion perusteella, joten Pearsonin tulomomenttikorrelaatiokertoimen käyttö on mielekästä.
Neliöhinta on selitettävä $=y$ ja asunnon ikä selittävä muuttuja $=x$

b) $a = 2172,35$ (R:n tulostuksesta)

$$b = r_{xy}\frac{s_y}{s_x} = -0.949\frac{\sqrt{26931}}{\sqrt{0.9444}}$$

$$\approx -160.26$$

$$\hat{y} = 2172 - 160.3\,x$$

Tulkinnat:

a uuden asunnon (ikä 0 vuotta) neliöhinta on keskimäärin 2172.35 €.

b kun asunto vanhenee vuodella (kun asunnon ikä kasvaa yhdellä vuodella), vähenee neliöhinta keskimäärin 160,26 €.

c) $\hat{R}^2 = r_{xy}^2 = (-0,949)^2 = 0,9006$

Asunnon ikä selittää noin 90 % asunnon neliöhinnon kokonaisvaihtelusta.

d) $\hat{y} = 2172,35 - 2,5 \cdot 160,26$
$\approx 1771,70$ €

(Jos a)-kohdassa ajateltu, että neliöhinta on selittävä muuttuja, tässä tulee olla 2100 sijoitettuna.)