

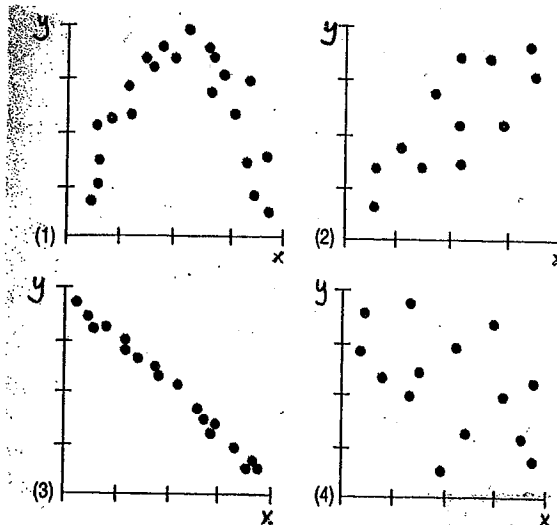
806109 TILASTOTIETEEN PERUSMENETELMÄT I
Taloustieteiden tiedekunnan opiskelijat
Harjoitus 6, viikko 42, syksy 2008

1. Aineistona YRITYS2007
(jatkoa harjoituksen 5 tehtävään 5)

Laske metallialan yrityksillä

- vuosien 2006 ja 2007 TE-arvosanojen välinen Pearsonin tulomomenttikorrelaatiokerroin r ,
- gearingin ja sijoitetun pääoman tuoton välinen Spearmannin järjestyskorrelaatiokerroin ρ . Tulkitse tulokset.
- Jos a)-kohdassa jätetään Luvata ja Componenta muista yrityksistä selvästi poikkeavina pois, miten arvioit käyvän Pearsonin tulomomenttikorrelaatiokertoimelle?

2. Alla on neljä erilaista hajontakuviota.



Missä kuvioista muuttujien x ja y välillä

- ei ole lainkaan tai vain heikkoa riippuvuutta,
- negatiivista riippuvuutta,
- lineaarista riippuvuutta,
- kohtalaista tai melko voimakasta riippuvuutta,
- voimakasta riippuvuutta?

Hajontakuvioita vastaavista aineistoista lasketut Pearsonin tulomomenttikorrelaatiokertoimet ovat -0.923 , -0.487 , 0.006 ja 0.777 . Yhdistä kertoimet oikeisiin kuvioihin.

3. Eräessä suuressa yrityksessä haluttiin selvittää, onko työntekijän sukupuolella yhteyttä siihen, kannattaako yrityksen johdon ehdottama uudistusta vai ei. Yrityksen koko henkilökunnan osalta tulokset olivat seuraavat:

	Sukupuoli	Yht.
	Mies	Nainen
Suhtautuminen kannattaa	60	65
uudistukseen ei kannata	170	150
Yht.	230	215
		445

Koska yrityksen henkilökunta oli jaetavissa kahteen eri organisaatioon tuloksia tarkastettiin myös erikseen näillä organisaatioosilla (A ja B) ja saatiin seuraavat tulokset:

Organisaatio A

	Sukupuoli	Yht.
	Mies	Nainen
Suhtautuminen kannattaa	10	45
uudistukseen ei kannata	20	90
Yht.	30	135
		165

Organisaatio B

	Sukupuoli	Yht.
	Mies	Nainen
Suhtautuminen kannattaa	50	20
uudistukseen ei kannata	150	60
Yht.	200	80
		280

Tutki ristitulosuhteen ja ehdollisten prosenttijakamien avulla sukupuolen ja uudistukseen suhtautumisen välistä riippuvuutta.

- a) koko aineistossa,
b) erikseen eri organisaatioosilla.
Mistä päätelmiä voit näistä tuloksista tehdä?
4. (jatkoa tehtävään 1a))
a) Määrittää metallialaan yrityksille regressioyhtälö $y = a + bx$, missä y =TE-arvosana v. 2007 ja x =TE-arvosana v. 2006. Tulkitse saadun yhtälön kertoimet.
b) Piirrä regressioyhtälön kuvaaja harjoituksen 5 tehtävässä 5a) piirrettyyn hajontakuviioon.

- c) Määrittää regressioyhtälön determinaatikerroin (= selitysaste) ja tulkitse se.
d) Laske regressioyhtälön antama ennustearvo d1) Outokummun, d2) Luvtan vuoden 2007 TE-arvosanaksi. Mitkä ovat vastaavat residuaalit?

5. Eräessä kunnassa selvitetiin 12 kunnallisuusvaalehdokkaan ikä (ikä, vuosina), vaalimainontaan sijoittama rahamäärä (mainoskust, 1000 euroina) ja vaaleissa saatu äänimäärä (äänet). Tästä aineistosta R:llä laskettuja tuloksia on esitetty liitteessä 1.

- a) Laske korrelaatiomatriisin pelitetty arvo.

- b) Tulkitse ehdokkaan

- b1) iän ja mainontaan sijoittaman rahamäärän,

- b2) mainontaan sijoittaman rahamäärän ja äänimäärän välisiä riippuvuuksia liitteestä löytyvien kuvioiden ja riippuvuuslukujen avulla.

- c) Tulkitse liitteessä 1 esitetyn regressioanalyysin tulokset (regressioyhtälö ja sen kertoimien tulkinat, determinaatikerroin eli selitysaste ja sen tulkinat).

6. Liitteessä 2 on R:llä saatuja tuloksia aineistosta, joka sisältää seuraavat tiedot 50:stä Oulussa keväällä 2002 myytävänä olleesta rivitaloasunnosta: hintayhtiö (1000 euroina), neljännäärä, ikä (vuosia, väh. yksi) ja etäisyys keskustasta (km).

- a) Mikä mukana olevista muuttujista on paras selittävä muuttuja hintayhtiölle? Perustele vastauksesi. Määrittää kyseinen regressioyhtälö ja tulkitse yhtälön kertoimet. Määrittää myös regressioyhtälön determinaatikerroin.

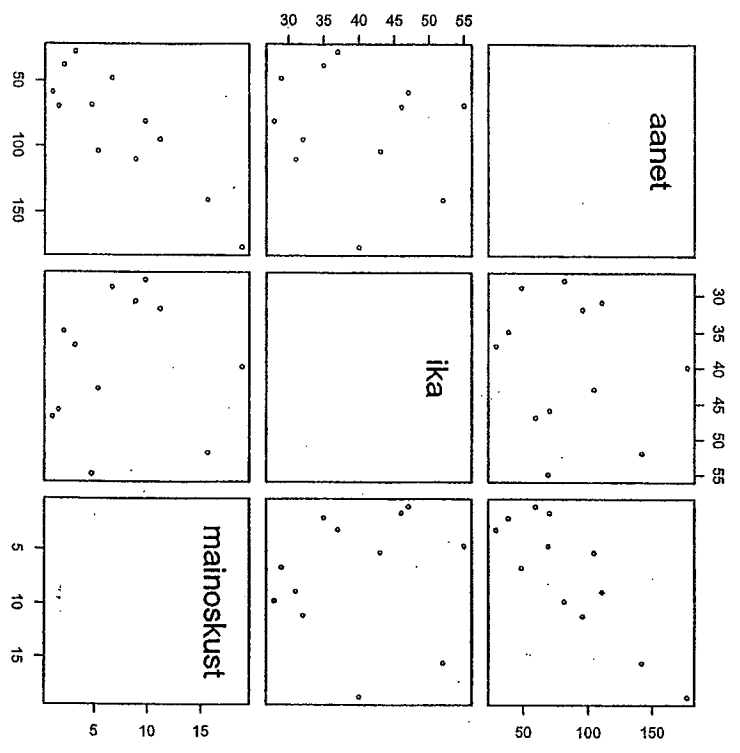
- b) Tulkitse regressioanalyysin tulokset liitteen kohdasta b) (regressioyhtälö, kertoimien tulkinat, determinaatikerroin).

- c) Laske sekä a)- että b)-kohdan yhtälöä käyttäen ennustearvo sellaisen rivitaloasunnon hinnaksi v. 2002, joka on ollut silloin 10 vuotta vanha, kooltaan 100 neliötä ja sijainnut 5 km:n päässä keskustasta.

HUOMI

1. vaihtoe pe 17.10. 12.00-16.00 sajeissa L4 ja PR104.
Salin L4 menevät ne, joiden sukunimi alkaa A-Pa.
Salin PR104 menevät ne, joiden sukunimi alkaa Pe-O.

Välikohteseen ei tarvitse ilmoittautua!



```
#Keskiarvot
> mean(ika)
[1] 39.58333
> mean(mainoskust)
[1] 7.408333
> mean(aanet)
[1] 85
```

```
#Korrelaatiomatriisi
aanet      ika      mainoskust
aanet      1.0000000  0.8688182
ika        1.0000000 -0.0954054
mainoskust 0.8688182 -0.0954054  1.0000000
```

```
#Kovarianssimatriisi
aanet      ika      mainoskust
aanet      1880.54545 70.272727 213.181818
ika        70.27273 82.265152 -4.896212
mainoskust 213.18182 -4.896212 32.015379
```

```
> RegModel.1 <- lm(aanet~mainoskust, data=vaa1itc)
> summary(RegModel.1)
```

```
Call:
lm(formula = aanet ~ mainoskust, data = vaa1itc)

Residuals:
    Min       1Q   Median       3Q      Max
-32.283 -16.417  1.744  16.277  33.039

Coefficients:
(Intercept)      35.670
mainoskust       1.200

Estimate Std. Error t value Pr(>|t|)
(Intercept)  35.670    11.013    3.239 0.008888
mainoskust   1.200     1.200    1.000 0.000244
```

```
Residual standard error: 22.52 on 10 degrees of freedom
Multiple R-Squared: 0.7548, Adjusted R-squared: 0.7303
F-statistic: 30.79 on 1 and 10 DF, p-value: 0.0002445
```

```
a) > round(corr(rivitalot[,c("etäisyys", "hinta", "ikä", "neliöt")],
use="complete.obs"), 3)
```

	etäisyys	hinta	ikä	neliöt
etäisyys	1.000	-0.285	-0.473	-0.218
hinta	-0.285	1.000	-0.218	0.850
ikä	-0.473	-0.218	1.000	0.139
neliöt	-0.218	0.850	0.139	1.000

```
> numsummary(rivitalot[,c("etäisyys", "hinta", "ikä", "neliöt")],
statistics=c("mean", "sd", "quantiles"), quantiles=c(0, .25, .5, .75, 1))
```

	mean	sd	0%	25%	50%	75%	100%	n
etäisyys	5.216	1.819549	0.8	4.100	5.50	6.300	8.6	50
hinta	100.606	33.225116	42.0	74.125	99.55	119.925	185.0	50
ikä	16.120	12.874021	1.0	3.750	14.00	23.750	46.0	50
neliöt	76.756	24.822004	33.0	56.475	77.00	96.750	124.0	50

```
b) > malli <- lm(hinta~etäisyys+ikä+neliöt, data=rivitalot)
> summary(malli)
```

Call:
lm(formula = hinta ~ etäisyys + ikä + neliöt, data = rivitalot)

Residuals:
 Min 1Q Median 3Q Max
-17.886 -5.770 -2.727 5.180 23.068

Coefficients:
(Intercept) 65.94189 7.73437 8.526 5.02e-11
etäisyys -6.08546 0.87227 -6.977 9.88e-09
ikä -1.27212 0.12151 -10.469 9.28e-14
neliöt 1.13232 0.05691 19.898 < 2e-16

Residual standard error: 9.642 on 46 degrees of freedom
Multiple R-squared: 0.9209, Adjusted R-squared: 0.9158
F-statistic: 178.6 on 3 and 46 Df, p-value: < 2.2e-16

