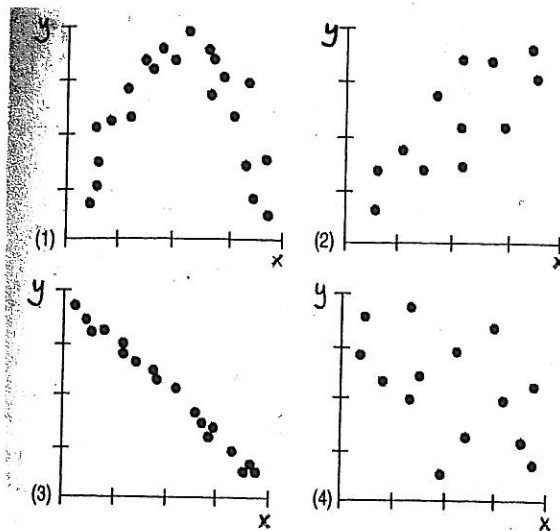


806109 TILASTOTIETEEN PERUSMENETELMÄT I
Harjoitus 6, viikko 43, syksy 2009

1. 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.

2. (jatkoa harjoituksen 5 tehtävään 6)

Ko. havaintomatriisi alla.

yritys	sipo	gearing	omavaraste	te08	te07	nettotulospros
NesteOil	6	46	46	7.3	9.3	0.7
Fortum	15	73	41	7.9	8.4	27.1
Gasum	9	5	55	8.7	10.0	2.2
PohjolanVoima	0	254	24	4.8	4.9	-6.6
Vattenfall	8	116	37	6.9	5.7	6.2
KymppiVoimaHankinta	3	599	6	4.6	8.5	0.3
FingridYhtiöt	6	174	27	6.1	6.4	9.2
SavonVoima	13	8	76	9.3	9.7	8.7
RAONordic	60	45	49	9.1	6.9	11.5
VantaanEnergiä	13	20	70	8.8	8.8	12.8
TurkuEnergiä	12	31	61	8.5	9.1	7.5
EVPEnergiä	5	90	43	6.7	6.7	2.4

- a) Laske ja tulkitse sijoitetun pääoman tuottoprosentin ja omavaraisuusasteen välinen Pearsonin tulomomenttikorrelaatiokerroin r .
Mikä r :n arvoksi saadaan, jos RAONordic (outlier) jätetään pois? Tulkitse saatu arvo.
- b) Laske ja tulkitse gearingin ja omavaraisuusasteen välinen Spearmanin järjestyskorrelaatiokerroin.
3. Eräessä suuressa yrityksessä haluttiin selvittää, onko työntekijän sukupuolella yhteyttä siihen, kannattaako yrityksen johdon ehdottamaa uudistusta vai ei. Yrityksen koko henkilökunnan osalta tulokset olivat seuraavat:

	Sukupuoli		Yht.
	Mies	Nainen	
Suhtautuminen kannattaa	60	65	125
uudistukseen ei kannata	170	150	320
Yht.	230	215	445

Koska yrityksen henkilökunta oli jaettavissa kahteen eri organisaatioon tuloksia tarkasteltiin myös erikseen näillä organisaatiotasolla (A ja B) ja saatiin seuraavat taulukot:

Organisaatiotaso A

	Sukupuoli		Yht.
	Mies	Nainen	
Suhtautuminen kannattaa	10	45	55
uudistukseen ei kannata	20	90	110
Yht.	30	135	165

Organisaatiotaso B

	Sukupuoli		Yht.
	Mies	Nainen	
Suhtautuminen kannattaa	50	20	70
uudistukseen ei kannata	150	60	210
Yht.	200	80	280

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

- a) koko aineistossa,
b) erikseen eri organisaatiotasolla.
Mitä päätelmiä voit näistä tuloksista tehdä?

4. (jatkoa harjoituksen 5 tehtävään 6)
- Määrittää regressioyhtälö $\hat{y} = a + bx$, missä y =TE-arvosana v. 2008 ja x =TE-arvosana v. 2007.
Tulkitse saadun yhtälön kertoimet a ja b .
 - Piirrä regressioyhtälön kuvaaja harjoituksen 5 tehtävässä 6c) muodostettuun hajontakuvioon.
 - Määrittää regressioyhtälön determinatiokerroin (= selitysaste) ja tulkitse se.
 - Oulun Energian TE-arvosana v. 2007 oli 9.6. Laske regressioyhtälön antama ennustearvo Oulun Energian vuoden 2008 TE-arvosanaksi. Oulun Energian todellinen TE-arvosana v. 2008 oli 9.1. Mikä on todellisen ja ennustetun arvon erotus eli residuaali?
5. Liitteessä 1 on R:llä saatuja tuloksia (osa peitetty!) aineistosta, joka sisältää 39 valtiota koskevia tietoja vuodelta 2005. Aineiston muuttujia ovat mm.
Fert = syntyvyysaste = synnytysten lkm/nainen ja FemEc = naisten taloudellinen aktiivisuus = naisten lukumäärä työvoimassa sataa työvoimassa olevaa miestä kohti.
- Kommentoi syntyvyysasteen ja naisten taloudellisen aktiivisuuden välistä riippuvuutta
 - liitteessä annetun kuvion perusteella,
 - laskemalla sopiva riippuvuusluku ja tulkitsemalla se.
 - Tulkitse regressioanalyysin tulokset (regressioyhtälö ja sen kertoimien tulkinta).
 - Laske ja tulkitse regressioyhtälön determinatiokerroin (= selitysaste).
6. Liitteessä 2 on R:llä saatuja tuloksia aineistosta, joka sisältää seuraavat tiedot 50:stä Oulussa keväällä 2002 myytävänä olleesta rivitaloasunnosta: hintapyyntö (1000 euroina), neliömäärä, ikä (vuosia, väh. yksi) ja etäisyys keskustasta (km).
- Mikä mukana olevista muuttujista on paras selittävä muuttuja hintapyyntölle? Perustele vastauksesi. Määrittää kyseinen regressioyhtälö ja tulkitse yhtälön kertoimet. Määrittää myös regressioyhtälön determinatiokerroin.
 - Tulkitse regressioanalyysin tulokset liitteen kohdasta b) (regressioyhtälö, kertoimien tulkinta, determinatiokerroin).
 - 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.

HUOM!

- välikoe pe 23.10. 14.00-18.00 saleissa L3 ja L4.
Saliin L3 menevät ne, joiden sukunimi alkaa A-L.
Saliin L4 menevät ne, joiden sukunimi alkaa M-Ö.

Välikokeeseen ei tarvitse ilmoittautua!

```

> numSummary(nations[,c("FemEc", "Fert")], statistics=c("mean", "sd",
"quantiles"), quantiles=c( 0.25, 5.75, 1 ))
mean sd 0% 25% 50% 75% 100% n NA
FemEc 65.7 16.34 29.0 52.0 67.0 81.0 91.0 37 2
Fert 2.1 0.76 1.3 1.7 1.9 2.5 4.3 37 2

> cov(nations[,c("FemEc", "Fert")], use="complete.obs")
FemEc Fert
FemEc 266.9 -7.29
Fert -7.3 0.58

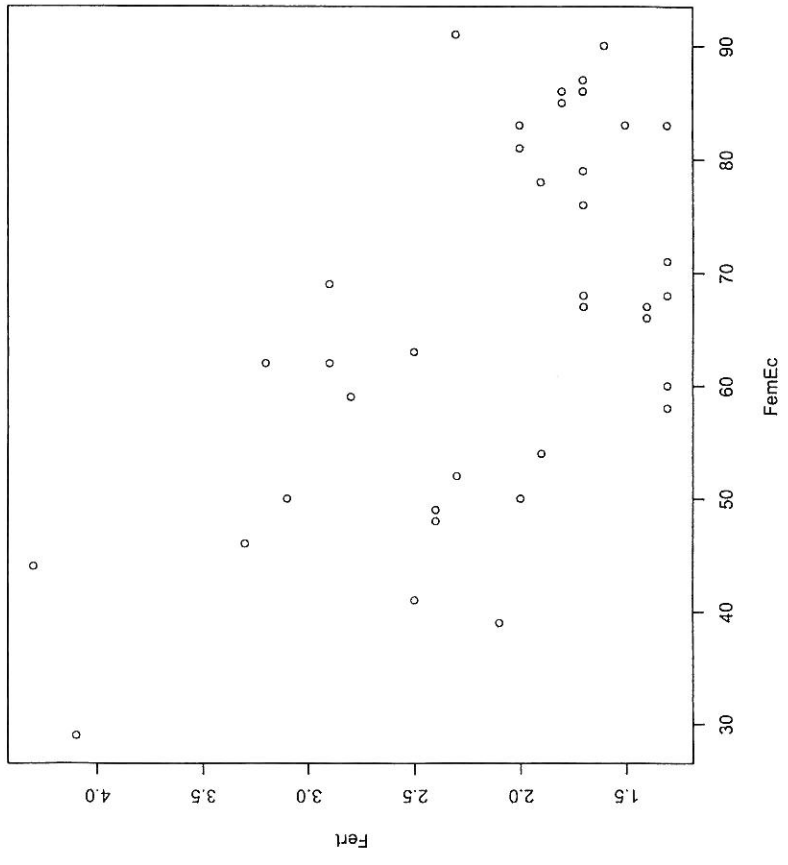
> malli <- lm(Fert~FemEc, data=nations)
> summary(malli)

Call:
lm(formula = Fert ~ FemEc, data = nations)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0475 -0.4017 -0.0738  0.3355  1.5700

Coefficients:
(Intercept)  3.93208  0.43086  9.13  8.8e-11
FemEc      -0.02732  0.00637  -4.29  0.00013
---
Residual standard error: 0.62 on 35 degrees of freedom
Multiple R-squared: 0.00637
F-statistic: 18.4 on 1 and 35 DF, p-value: 0.000135

```



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

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

