

806109 TILASTOTIETEEN PERUSMENETELMÄT I
Harjoitus 6, viikko 42, syksy 2010

HUOM!

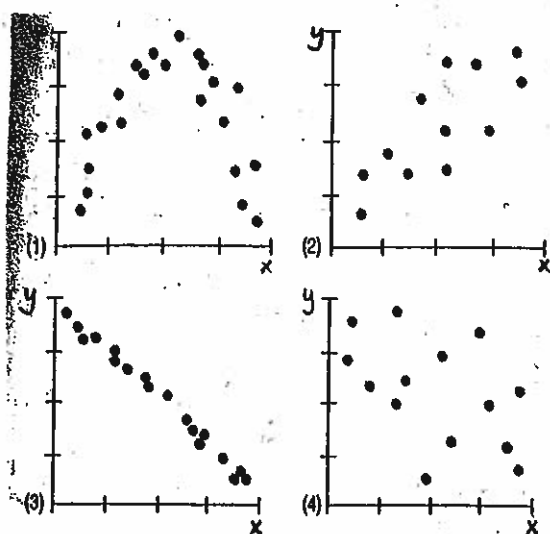
1. välikoe pe 22.10. klo 12.00-16.00 saleissa L3 ja L4.

Saliin L3 menevät ne, joiden sukunimi alkaa A-L.

Saliin L4 menevät ne, joiden sukunimi alkaa M-Ö.

Välikokeeseen ei tarvitse ilmoittautua!

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 tulomomenttikorrelaatioker-
toimet 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
VantaanEnergia	13	20	70	8.8	8.8	12.8
TurkuEnergia	12	31	61	8.5	9.1	7.5
EVPEnergia	5	90	43	6.7	6.7	2.4

- a) Laske ja tulkitse sijoitetun pääoman tuotto-% ja omavaraisuusasteen välinen Pearsonin tulomomenttikorrelaatiokerroin r . ($r = +0.282$)
- b) Mikä r :n arvoksi saadaan, jos a)-kohdassa RAONordic (outlier) jätetään laskuista pois? Tulkitse saatu arvo. ($r = +0.748$)
- c) Laske ja tulkitse gearingin ja omavaraisuusasteen välinen Spearmanin järjestyskorrelaatiokerroin. ($r = -0.951$)
- d) (vapaaehtoinen!) Laske ja tulkitse vuosien 2007 ja 2008 TE-arvosanojen Pearsonin tulomomenttikorrelaatiokerroin r . ($r = +0.575$)

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. Sähkölämmitteisen loma-asunnon sähkön kulutusta ja ulkoilman lämpötilaa seurattiin viikon ajan. Tällöin saatiin seuraavat havainnot:

Vuorokausi	1	2	3	4	5	6	7
Sähkön kulutus (kWh):	32	28	23	21	30	28	22
Ulkoilman keskilämpötila (C):	5	8	12	10	-1	3	7

- a) Piirrä muuttujien välinen hajontakuvio ja kommentoi kuvion perusteella muuttujien välistä riippuvuutta.
- b) Sovita aineistoon regressiosuora $y = a + bx$, missä y = sähkön kulutus ja x = ulkoilman keskilämpötila vuorokaudessa. Tulkitse saadut kertoimet a ja b . Määrää myös regressioyhtälön determinaatikerroin (=selitysaste) ja tulkitse se.
- c) Piirrä regressioyhtälön kuvaaja a)-kohdassa muodostettuun hajontakuvioon.
- d) Paljonko regressioyhtälö ennustaa loma-asunnon lämmönkulutuksen olevan (=ennustearvo), jos ulkoilman keskilämpötila vuorokaudessa on 9 astetta?
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.
- a) Kommentoi syntyvyysasteen ja naisten taloudellisen aktiivisuuden välistä riippuvuutta
 a1) liitteessä annetun kuvion perusteella,
 a2) laskemalla sopiva riippuvuusluku ja tulkitsemalla se.
- b) Tulkitse regressioanalyysin tulokset (regressioyhtälö ja sen kertoimien tulkinta).
- c) Laske ja tulkitse regressioyhtälön determinaatikerroin (= selitysaste).
6. Liitteessä 2 on R:llä saatuja tuloksia Oulun läänin kuntia (47) koskevasta aineistosta. Muuttujat liitteessä ovat:
 Koulutus = korkeakoulututkinnon suorittaneiden prosenttiosuus kunnan 15 vuotta täyttäneestä väestöstä v. 2007
 Tulotaso = verotettava tulo euroa/asukas v. 2007
 Seniorit = 65 vuotta täyttäneiden prosenttiosuus kunnan väestöstä v. 2007
 Veroaste = veroprosentti v. 2007
- a) Laske korrelaatiomatriisin peitetyt arvot.
- b) Tulkitse muuttujien tulotaso ja seniorit välistä riippuvuutta liitteessä olevan hajontakuvion ja riippuvuusluvun avulla.
- c) Määrää regressioyhtälö $y = a + bx$, jossa koulutuksella selitetään tulotasoa. Tulkitse regressioyhtälön kertoimet.
- d) Määrää c)-kohdan regressioyhtälön determinaatikerroin (= selitysaste) ja tulkitse se.

```

> numSummary(nations[,c("FemEc", "Fert")], statistics=c("mean", "sd",
"quantiles"), quantiles=c( 0.25, 0.5, 0.75, 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

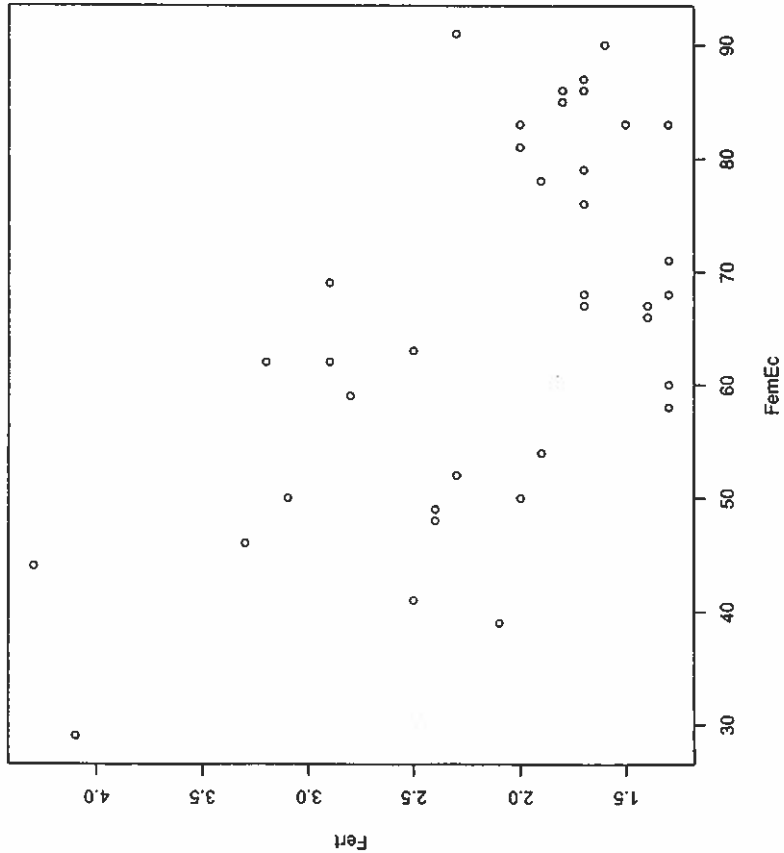
> malli <- lm(Fert~FemEc, data=nations)
> summary(malli)

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: 0.43086 Adjusted R-squared: 0.40013
F-statistic: 18.4 on 1 and 35 DF, p-value: 0.00013

```



```
> round(cor(oulu[,c("Koulutus", "Seniorit", "Tulotaso", "Veroaste")],
+ use="complete.obs"), 3)
```

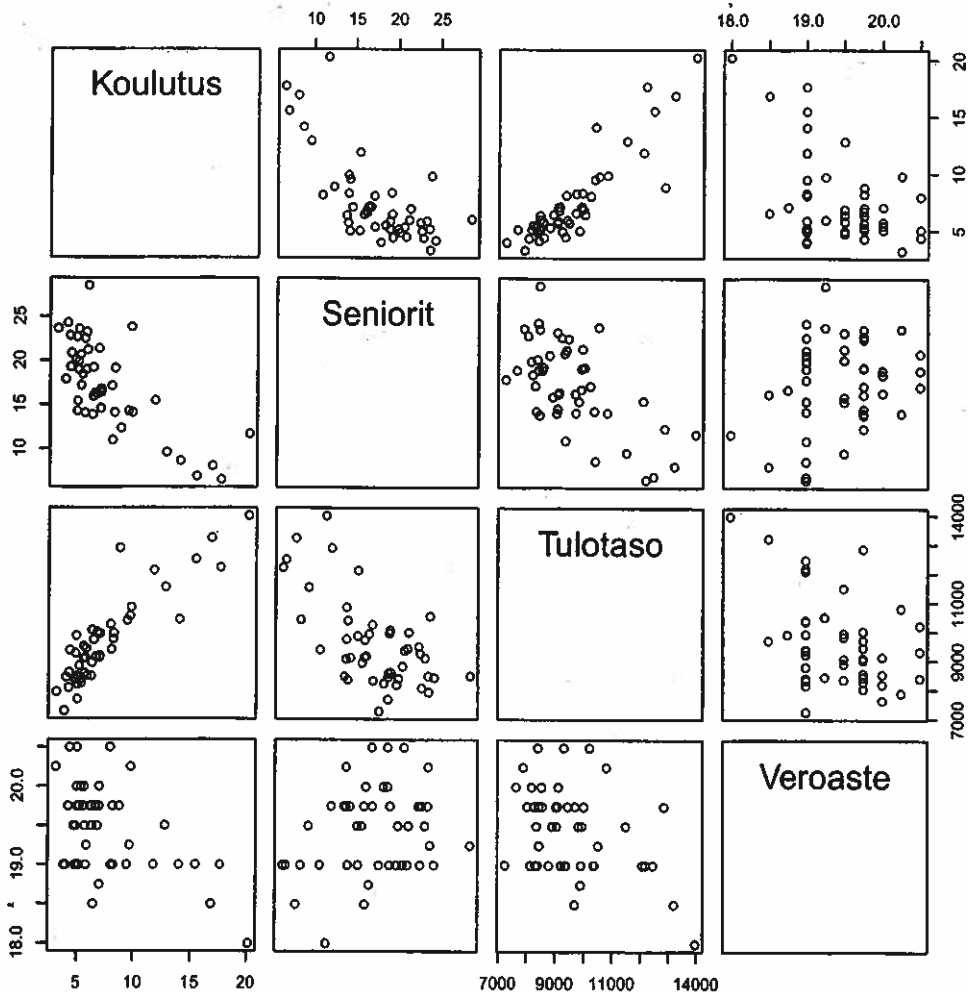
	Koulutus	Seniorit	Tulotaso	Veroaste
Koulutus	1.000		0.890	-0.486
Seniorit		1.000	-0.629	0.268
Tulotaso	0.890	-0.629	1.000	-0.411
Veroaste	-0.486	0.268	-0.411	1.000

```
> round(cov(oulu[,c("Koulutus", "Seniorit", "Tulotaso", "Veroaste")],
+ use="complete.obs"), 3)
```

	Koulutus	Seniorit	Tulotaso	Veroaste
Koulutus	14.937	-13.889	5253.633	-1.044
Seniorit	-13.889	24.602	-4767.826	0.739
Tulotaso	5253.633	-4767.826	2332582.874	-349.216
Veroaste	-1.044	0.739	-349.216	0.309

```
> numSummary(oulu[,c("Koulutus", "Seniorit", "Tulotaso", "Veroaste")],
+ statistics=c("mean", "sd", "quantiles"), quantiles=c(0, .25, .5, .75, 1))
```

	mean	sd	0%	25%	50%	75%	100%	n
Koulutus	7.4928	3.8649	3.190	5.025	6.250	8.205	20.17	47
Seniorit	17.2762	4.9600	6.540	14.215	17.200	20.770	28.54	47
Tulotaso	9577.8956	1527.2796	7230.762	8427.555	9205.402	10103.258	13957.26	47
Veroaste	19.4521	0.5557	18.000	19.000	19.500	19.750	20.50	47



Pearsonin tulomomenttikorrelaatiokertoimen r arvojen tulkinnassa voi käyttää (tässä vaiheessa kurssia) seuraavia adjektiiveja:

- $-1 \leq r < -0.8$: vahva (voimakas) negatiivinen lineaarinen riippuvuus
- $-0.8 \leq r < -0.6$: melko vahva (melko voimakas) negatiivinen lineaarinen riippuvuus
- $-0.6 \leq r < -0.4$: kohtalainen (melkoinen) negatiivinen lineaarinen riippuvuus
- $-0.4 \leq r < -0.2$: melko heikko negatiivinen lineaarinen riippuvuus
- $-0.2 \leq r < -0.1$: heikko negatiivinen lineaarinen riippuvuus
- $-0.1 \leq r \leq 0.1$: ei lineaarista riippuvuutta
- $0.1 < r \leq 0.2$: heikko positiivinen lineaarinen riippuvuus
- $0.2 < r \leq 0.4$: melko heikko positiivinen lineaarinen riippuvuus
- $0.4 < r \leq 0.6$: kohtalainen (melkoinen) positiivinen lineaarinen riippuvuus
- $0.6 < r \leq 0.8$: melko vahva (melko voimakas) positiivinen lineaarinen riippuvuus
- $0.8 < r \leq 1$: vahva (voimakas) positiivinen lineaarinen riippuvuus

Huom. Samat adjektiivit sopivat myös Spearmanin ρ :ta ja suhteellista kontingenssikerrointa C/C_{\max} tulkittaessa:

Esim.,

$\rho = 0.45$: muuttujien välillä vallitsee kohtalainen positiivinen riippuvuus

$C/C_{\max} = 0.45$: muuttujien välillä vallitsee kohtalainen riippuvuus