

806109 TILASTOTIETEEN PERUSMENETELMÄT I
Muut kuin taloustieteiden tiedekunnan opiskelijat
Harjoitus 6, viikko 42, syksy 2008

HUOM! Tällä viikolla suoritusmerkintään riittää 2 tehtävää.

1. Varuskunnassa järjestettiin ampumahiihtokilpailut. Kilpailun tulokset olivat seuraavat:

kilpailija	sotilasarvo	loppuaika
T. Tarkka	luutnantti	1.23.30
K. Kiri	kersantti	1.24.25
R. Riuska	alikersantti	1.24.25
L. Latu	alikersantti	1.26.20
O. Osuma	vääpeli	1.27.15
V. Vauhti	vänrikki	1.27.20
L. Luoti	luutnantti	1.28.05
H. Huti	sotilasmestari	1.29.15
K. Kunto	kapteeni	1.29.30

Sotilasarvojen järjestys (alhaisimmasta korkeimpaan) on: alikersantti, kersantti, vääpeli, sotilasmestari, vänrikki, luutnantti ja kapteeni. Tutki sotilasarvon ja kilpailussa menestymisen välistä riippuvuutta sopivan tunnusluvun avulla. Tulkitse tulos ja perustele käyttämäsi tunnusluvun valinta lyhyesti.

2. Lukion poikien liikuntaryhmässä järjestettiin Cooperin testi. Aluksi jokainen poika arvioi, kuinka pitkän matkan ehtii 12 minuutin aikana juosta (=arvio). Lisäksi kullekin pojalle laskettiin painoindeksi (=painoindeksi). Tämän jälkeen Cooperin testi juostiin ja jokaisen juoksijan testitulos mitattiin (=tulos). Liitteessä 1 on esitetty aineistoon liittyvää R-ohjelman tulostusta. Käytä näitä tuloksia hyväksesi seuraavissa kohdissa.

- Laske ennakoarvion ja testituloksen välinen korrelaatiokerroin r ja tulkitse tulos lyhyesti.
 - Sovita aineistoon regressioyhtälö $\hat{y} = a + bx$, jossa ennakoarvion avulla selitetään testitulosta. Tulkitse regressioyhtälön kertoimet selväkielisesti.
 - Määrä b)-kohdan regressioyhtälön determinaatikerroin (eli selityssaste) ja tulkitse tulos.
 - Ennen testiä Akseli arvioi juoksevansa 2600 metriä. Testitulos oli 2860 metriä. Määrä Akselille b)-kohdan regressioyhtälön mukainen ennustearvo ja residuaali.
3. 10 lapsidiabeetikolta mitattiin diabeteksen kesto ($= x$, vuosia) ja aamulla käytettävän pitkävaikutteisen insuliinin määrä ($= y$, insuliiniyksikköä) ja saatiin seuraavat tulokset:

x	5.4	1.2	7.1	0.7	6.0	2.1	4.8	13.6	12.0	1.3
y	16	14	12	4	16	8	18	18	20	4

- a) Tutki muuttujien välistä riippuvuutta
- a1) graafisesti,
 - a2) (Pearsonin tulomomentti)korrelaatiokertoimen avulla.
- b) Sovita aineistoon regressiosuora $y = a + bx$, missä y =aamulla käytettävän pitkävaikutteisen insuliinin määrä ja x =diabeteksen kesto. Tulkitse kertoimet. Määrää myös regressioyhtälön determinaatikerroin eli selitysaste ja tulkitse se.
- c) Lisää b)-kohdan regressiosuora a1)-kohdan kuvioon. Paljonko regressioyhtälö ennustaa insuliinin määrän olevan, jos diabeteksen kesto on 10.0 vuotta?
- d) Määrää havaintoaineiston viimeisen tilastoyksikön (jolla muuttujien havaitut arvot ovat $x=1.3$ ja $y=4$) ennustearvo \hat{y}_i ja residuaali $e_i = y_i - \hat{y}_i$.
- e) Liitteessä 2 on esitetty R-ohjelman tulostus b)-kohtaan liittyen. Vertaa R:n tulostuksen antamaa informaatiota b)-kohdassa saamiisi tuloksiin.
4. Pienen ala-asteen kakkosluokan liikuntaryhmän pojat ($n=11$) ottelivat neliottelun, jossa lajeina olivat 60 metrin juoksu, 100 metrin juoksu, pallonheitto ja pituushyppy.
- a) Pallonheitossa lopputulokset (metreinä) olivat
18.7 23.4 15.9 19.6 21.0 19.7 18.5 19.7 21.8 18.4 20.1
Laske pallonheiton lopputulosten aritmeettinen keskiarvo ja keskihajonta.
 - b) Liitteessä 3 on esitetty tarkasteltavasta aineistosta saatu R:n tulostus. Kommentoi 60 metrin juoksun ja 100 metrin juoksun välistä riippuvuutta korrelaatiokertoimen r perusteella. Missä korrelaatiodiagrammin kuvassa/kuvissa (1-12) kuvataan 60 metrin juoksun ja 100 metrin juoksun välistä riippuvuutta?
 - c) Määrää regressioyhtälö $y = a + bx$ ja tulkitse kertoimet, kun vastemuuttujana on 100 metrin juoksu ja selittävänä muuttujana 60 metrin juoksu. Määrää lisäksi regressioyhtälön determinaatikerroin (eli selitysaste) ja tulkitse se.
- Matin tulokset neliottelussa olivat: 60 metrin juoksu = 12.5 sekuntia, 100 metrin juoksu = 23.2 sekuntia, pallonheitto = 19.6 metriä ja pituushyppy 2.20 metriä. Ennusta Matin vastemuuttujan arvo regressioyhtälön avulla.

HUOM!

1. välikoe perjantaina 17.10. klo 12.00-16.00 salissa L6.

Tästä johtuen torstaina 16.10. ei ole luentoa. Harjoituksen 6 (vko 42) laskupäivä:

torstaina 16.10. klo 12-14 salissa SÄ118 ja

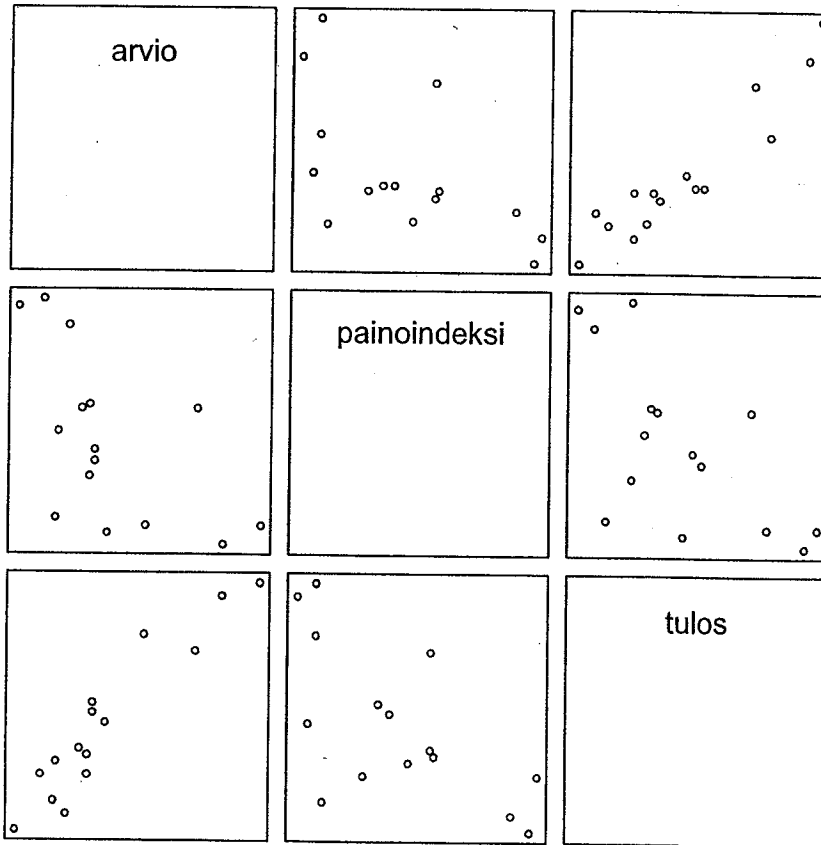
perjantaina 17.10. klo 8-12 salissa PR101

Voit osallistua jompaan kumpaan tai molempiin.

LIITE 1

```
# Tunnuslukuja
> numSummary(Cooper[,c("arvio", "painoindeksi", "tulos")],
             statistics=c("mean", "quantiles"))
             mean    0%    25%    50%    75%   100%  n
arvio      2454.66667 2100.0 2280.0 2380.0 2525.00 3050.0 15
painoindeksi 22.33333 19.6 20.2 22.1 23.25 26.1 15
tulos      2585.33333 2260.0 2430.0 2510.0 2730.00 3020.0 15

# Kovarianssimatriisi
> cov(Cooper[,c("arvio", "painoindeksi", "tulos")], use="complete.obs")
             arvio painoindeksi    tulos
arvio      72540.9524  -353.30952 60094.7619
painoindeksi -353.3095    4.82381 -334.0476
tulos      60094.7619  -334.04762 56383.8095
```



```

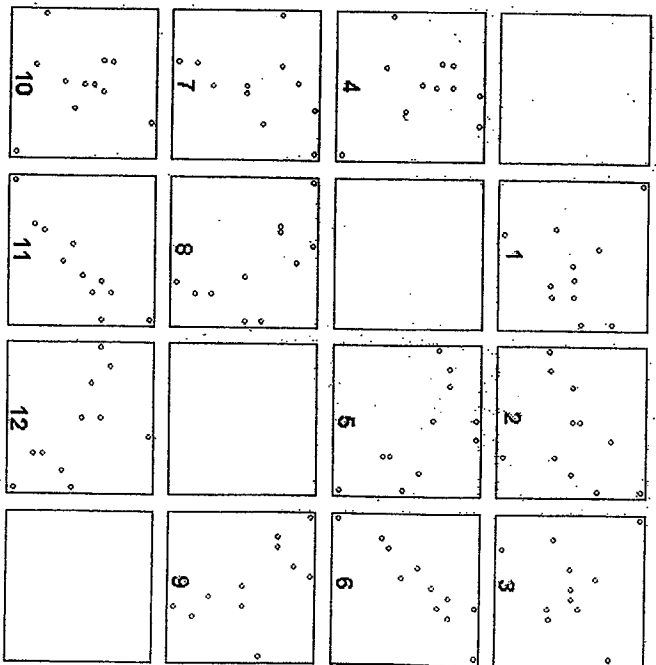
> RegModel.1 <- lm(insuliinin.maara ~ diabeteksen.kesto, data=diabetes)
> summary(RegModel.1)

Call:
lm(formula = insuliinin.maara ~ diabeteksen.kesto, data = diabetes)

Residuals:
    Min       1Q   Median       3Q      Max
-5.0168 -2.8374 -0.5759  2.8743  5.5994

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.7599      2.1089   3.680 0.00622 **
diabeteksen.kesto  0.9668      0.3058   3.162 0.01336 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.124 on 8 degrees of freedom
Multiple R-Squared:  0.5554, Adjusted R-Squared:  0.4999
F-statistic: 9.995 on 1 and 8 DF, p-value: 0.01336
    
```



```

#keskiarvot:
> mean(juoksu.100m)
[1] 23.3182
> mean(juoksu.60m)
[1] 12.6636
> mean(pituushyppy)
[1] 2.1318

#varianssit:
> var(juoksu.100m)
[1] 1.5796
> var(juoksu.60m)
[1] 0.1605
> var(pituushyppy)
[1] 0.0186
    
```

```

#korrelaatiomatriisi
> cor(neliottelu, use="complete.obs")

          juoksu.100m  juoksu.60m  pallonheitto  pituushyppy
juoksu.100m      1.0000      0.9328      -0.1067     -0.6302
juoksu.60m       0.9328      1.0000      0.0322     -0.5983
pallonheitto    -0.1067      0.0322      1.0000      0.3907
pituushyppy     -0.6302     -0.5983      0.3907      1.0000
    
```