

Tilastotieteen peruskurssi 5 op

Tentti 27.03.2020

1. Jokien kiintoainekulkeumaa (yksikkö = miljoonaa tonnia / vuosi, muuttuja kulkeuma) mallinnettiin lineaarisella regressiomallilla, jossa selittävinä tekijöinä olivat joen valuma-alueen pinta-ala (muuttuja alue) ja virtaama (kuvaava joen virtausnopeutta). Muuttujan alue yksikkö on 1000 km^2 eli valuma-alue ilmaistaan tuhansissa neliökilometreissä. Muuttujan virtaama yksikkö on m^3/s eli kuutiometriä sekunnissa.

Alla on tuloste analyysistä:

```
> m <- lm(kulkeuma ~ alue + virtaama, data=ain)
> summary(m)
```

```
Call:
lm(formula = kulkeuma ~ alue + virtaama, data = ain)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-3.972 -2.041 -0.491  1.293  6.970
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.06070     2.15356  -0.493   0.637
    alue      0.09736     0.03557   2.737   0.029 *
    virtaama  0.01264     0.00509   2.483   0.0379 *
```

```
Residual standard error: 3.631 on 7 degrees of freedom
Multiple R-squared:  0.5772,    Adjusted R-squared:  0.4564
F-statistic: 4.779 on 2 and 7 DF,  p-value: 0.04913
```

- (a) Tulkitse muuttujan alue kertoimen estimaatti.
- (b) Ennusta kiintoainekulkeuma joelle, jossa valuma-alueen pinta-ala on 50 tuhatta neliökilometriä ja virtausnopeus on 20 kuutiometriä sekunnissa.
- (c) Kirjoita näkyviin mallin systemaattinen osa tulosteen estimaattien avulla.
- (d) Tulkitse muuttujaan virtaama liittyvä testi. Hypoteesit näkyviin, $\alpha = 0.05$.
- (e) Tulkitse mallin selitysaste.
- (f) Oletetaan, että valuma-alue kasvaa 100 km^2 ja virtaama kasvaa $1 \text{ m}^3/\text{s}$. Miten keskimääräinen muuttuu keskimääräinen kiintoainekulkeuma?

Käännä

2. Eläintieteilijä selvitti vesikauhun esiintymistä haisunäädissä kahdella eri maantieteellisellä alueella (alueet itäinen ja läntinen). Yhteenvedo aineistoon tulleet haisunäädistä on alla:

vesikauhu	Maantieteellinen alue	
	Itäinen	Läntinen
Kyllä	72	79
Ei	130	89

Analyysiä varten muuttujat koodattiin seuraavasti: vesikauhu = 1, kun haisunäädellä on vesikauhu ja 0, kun haisunäädellä ei ole vesikauhua. alue = 0 = Läntinen, kun haisunäätä on läntiseltä alueelta ja alue = 1 = itäinen, kun haisunäätä on itäiseltä alueelta. Analyysi tehtiin logistisella regressiolla, tuloste on alla:

```
> m <- glm(vesikauhu ~ alue, data = logreg.dat)
> summary(m)

Call:
glm(formula = vesikauhu ~ alue, data = logreg.dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4702  -0.3564  -0.3564   0.5298   0.6436

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.47024    0.03777  12.451  <2e-16 ***
alueItäinen   -0.11380    0.05112  -2.226   0.0266 *
---
(Dispersion parameter for gaussian family taken to be 0.2396408)

Null deviance: 89.376  on 369  degrees of freedom
Residual deviance: 88.188  on 368  degrees of freedom
AIC: 525.42

Number of Fisher Scoring iterations: 2
```

- tulkitse parametri β_1 (muuttujan alue kerroin) vetosuhteen OR avulla.
- tulkitse selittävään muuttujaan alue liittyvä testi.
- laske, tulosteen ei aineiston avulla, vedonlyöntisuhde (vesikauhun puolesta) ryhmässä Läntinen alue
- laske, tulosteen ei aineiston avulla, vedonlyöntisuhde (vesikauhun puolesta) ryhmässä Itäinen alue

Käännä

3. Kokeessa tarkasteltiin kahden pesuainemerkin PesuPaavo ja PyykkiPekka tahranpoistokykyä veden eri lämpötiloissa (kylmä, lämmin, kuuma). Koe toteutettiin seuraavasti: 24 kangaspalaan tehtiin 15 likatahraa jokaiseen. Neljä kangaspalaa arvottiin täysin satunnaisesti kuhunkin käsittelyyn. Käsittelyt:

(PesuPaavo, kylmä vesi) (PesuPaavo, lämmin vesi) (PesuPaavo, kuuma vesi)
(PyykkiPekka, kylmä vesi) (PyykkiPekka, lämmin vesi) (PyykkiPekka, kuuma vesi)

Pesun jälkeen jokaisesta kangaspalasta laskettiin kuinka monta likatahraa pesu on poistanut. Vastemuuttuja pesutulos on siis pesussa poistuneiden likatahrojen lukumäärä. Tästä seuraa se, että suurempi poistuneiden likatahrojen lukumäärä ilmaisee parempaa pesutulosta.

Alla olevien tulosteiden perusteella päättelä, onko selittäjillä (pesuaine ja veden lämpötila) vaikutusta vastemuuttujaan pesutulos. Jos vaikutusta on, niin vastaa seuraaviin kysymyksiin. Mikä käsittely antaa keskimäärin huonoimman pesutuloksen ja mikä on kyseisessä käsittelyssä poistunut keskimääräinen tahrojen lukumäärä? Mikä käsittely antaa keskimäärin parhaimman pesutuloksen ja mikä on kyseisessä käsittelyssä poistunut keskimääräinen tahrojen lukumäärä?

Analysis of Variance Table

Response: lukumaara

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pesuaine	1	32.667	32.667	14.700	0.001216 **
veden.lampotila	2	160.333	80.167	36.075	5.044e-07 ***
pesuaine:veden.lampotila	2	44.333	22.167	9.975	0.001215 **
Residuals	18	40.000	2.222		

>

>

>

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.5000	0.7454	7.379	7.59e-07 ***
pesuainePyykkiPekka	-0.5000	1.0541	-0.474	0.640960
veden.lampotilaLammin	1.5000	1.0541	1.423	0.171830
veden.lampotilaKuuma	5.0000	1.0541	4.743	0.000162 ***
pesuainePyykkiPekka:veden.lampotilaLammin	6.5000	1.4907	4.360	0.000377 ***
pesuainePyykkiPekka:veden.lampotilaKuuma	2.0000	1.4907	1.342	0.196394

Käännä

4. Aikakausilehden kuukausiraportti vertaili einesvalmistajan naudanlihahampurilaisten ja kanahampurilaisten energiasisältöjä (yksikkö kcal/100g). Aineiston hampurilaiset ostettiin kaupoista täysin satunnaisesti. Hampurilaisten energiasisällöt:

	Energiasisältö					
Naudanliha	266	261	256	229	264	270
Kana	230	233	203	207	195	204

Testaa käyttäen kurssilla esitettyä menetelmää, poikkeavatko naudanliha- ja kanahampurilaisten ko. valmistajan keskimääräiset energiasisällöt toisistaan.
 $\alpha = 0.05$