

## Harjoitus 5, viikko 39: mikroluokkatehtävät

1. Jatketaan kotitehtävissä käsitellyn fungisidi-aineiston analysointia R-ohjelmistolla.

- (a) Sijoita vehnäsadon (tn/ha) havaitut arvot kahteen vektoriin: vertailukäsittelyn jälkeiset satotulokset vektoriin `sato1` ja koekäsittelyn jälkeiset tulokset vektoriin `sato2` siten, että kummassakin maatilojen järjetyks on sama. Yhdistä em. vektorit yhdeksi vektoriksi `sato`.

```
> sato1 <- c(5.0, 4.3, 5.9)
> sato2 <- c(6.1, 5.7, 7.0)
> sato <- c(sato1, sato2)
```

- (b) Määrittää koeyksiköille 1, ..., 6 tekijän `fung` ja `tila` tasot funktiolla `gl()` (= “generate levels”). Tämä funktio soveltuu käytettäväksi luokitteluasteikon tekijöiden tasojen generointiin, kun aineisto on säännönmukaisesti lajiteltu määrättyyn järjestykseen kuten tässä ensin käsittelyn ja sitten tilan suhteen. Funktion kutsussa 1. argumentti kertoo tasojen lukumäärän, 2. argumentti sen, kuinka monen havainnon jälkeen taso vaihtuu seuraavaksi, ja 3. argumenttina on havaintojen kokonaismäärä. Listaa lopuksi luomasi aineisto:

```
> fung <- gl(2, 3, 6, labels = c(" ei", " kyllä") ) ; fung
> tila <- gl(3, 1, 6) ; tila
> data.frame(sato, fung, tila)
```

Moodle 1

- (c) Piirrä havainnot rinnakkaisiin pistekuvioihin, joissa saman tilan havainnot on yhdistetty janalla.

```
> interaction.plot(fung, tila, sato, type = 'b', lty=1, pch = 0:2 )
```

Moodle 2

- (d) Analysoidaan aineisto pareittaisiin erotuksiin perustuvalla menetelmällä. Laske vasteen havaittujen arvojen erotukset käsittelyjen välillä jokaisen maatiloiden kohdalta erikseen ja talleta ne vektoriin `ero`. Lasketaan erotusmuuttujan keskiarvo ja keskihajonta. Sen jälkeen kutsu funktiota `t.test()` tällä pääargumentilla ja muiden argumenttien osalta oletusarvoillaan:

```
> ero <- sato2 - sato1 ; ero
> mean(ero); sd(ero)
> t.test(ero)
```

Saitko tulostukseen saman luottamusvälin kuin kotitehtävässä 3 (c)? Moodle 3

- (e) Todettakoon, että sama analyysi voidaan toteuttaa myös seuraavasti:

```
> t.test(sato2, sato1, paired =T)
```

Moodle 4

- (f) Jos sen sijaan jätät asettamatta loogisen argumentin `paired` arvoksi `T` (=TRUE), niin funktio `t.test()` analysoi aineiston sillä oletusarvolla, että kyseessä olisivat riippumattomat rinnakkaiset ryhmät (ks. luentomonisteen luku 4: 1-suuntainen varianssianalyysi).

```
> t.test(sato2, sato1, var.equal=T)
```

Vertaa saamaasi tulostusta kotitehtävän 1 tuloksiin. Ovatko tulokset samat? Huomaa, kuinka paljon epätarkempi ja näytöltään heikompi tulos tällä analyysillä saadaan verrattuna aiempaan analyysiin, jossa tilojen välinen vaihtelu eliminoidaan käsittelyjen vertailusta. Moodle 5

2. Jatkoa edelliseen tehtävään sekä kotitehtäviin 2. ja 3. Analysoidaan aineisto 2-suuntaisella varianssianalyysillä.

- (a) Sovita aineistoon 2-suuntainen malli, jossa vasteena on vehnäsadon määrä, koetekijänä fungisidikäsitteily ja lohkoetekijänä maatala. Tulosta myös mallin parametrien estimaatit, keskivirheet, jne.

```
> va2 <- lm(sato ~ fung + tila)
> round( cbind( summary(va2)$coef , confint(va2) ), 4)
```

Moodle 6

Vertaa fungisidikäsittelystä saatuja tuloksia edellisessä tehtävässä saatuuihin.

- (b) Talleta mallin sovitustuloksista sovitteet ja jäännöstermit omiin vektoreihinsa ja listaa ne yhdessä havaintojen kanssa. Tulosta mallitukseen liittyvä ANOVA-taulu.

```
> sovite <- fitted(va2)
> jaannos <- resid(va2)
> data.frame(fung, tila, sato, sovite, jaannos)
> anova(va2)
```

Moodle 7

Ovatko luvut samoja kuin kotitehtävässä 3?

Moodle 8

Moodle 9

**3.** Valkosipulia kasvatetaan penkeissä, joiden leveys on 1 m ja penkkien välit ovat 0.5 m. Haluttiin mm. selvittää, kuinka moneen riviin penkin pituussuunnassa taimia kannattaa istuttaa. Tähän tarkoitukseen suunniteltiin koe, jossa koetekijänä oli rivien määrä per penkki tasoinaan 3, 4 ja 5 riviä.

Koepopulaatioon valittiin 12 samankokoista penkkiä, jotka jaettiin 4 lohkoksi, joten yhteen lohkoon kuuluu 3 penkkiä. Kussakin lohkossa vertailtavat 3 käsittelyä arvottiin penkkien kesken. Edelleen, jokaisessa rivissä taimien etäisyys toisistaan oli 10 cm. Satotulosta mitattiin grammoina per neliömetri (muuttuja `sipuli.g`). Tulokset on annettu seuraavassa taulukossa:

Käsittely	Lohko			
	1	2	3	4
1: 3 riviä	371	378	464	463
2: 4 riviä	410	373	498	654
3: 5 riviä	642	536	507	588

- (a) Talletetaan vastemuuttujan havainnot vektoriin `sipuli.g`

```
> sipuli.g <- c(371, 378, 464, 463, # käsittely 1: 3 riviä
+              410, 373, 498, 654, # käsittely 2: 4 riviä
+              642, 536, 507, 588) # käsittely 3: 5 riviä
```

- (b) Muunna vasteen arvot niin, että mittayksiköksi tulee alkuperäisen  $\text{g/m}^2$  sijasta  $\text{kg/m}^2$ . Määrittele käsittelytekijä `rivit` ja generoi sen 3 tasoa funktiolla `gl()`. Generoi vastaavasti tasot ( $b = 4$ ) lohkoketekijälle `lohko`. Listaa näin luotujen vektorien muodostama datakehikko.

```
> library(gmodels)
> Y <- sipuli.g/1000 # vasteen mittayksiköksi kg/m^2
> rivit <- gl(3,4,12)
> levels(rivit) <- c(" 3r", " 4r", " 5r")
> lohko <- gl(4,1,12)
> data.frame(rivit, lohko, Y)
```

Moodle 10

- (c) Laske ja tulosta vasteen keskiarvot ja keskihajonnat sekä käsittelyryhmittäin että lohkoittain käyttäen funktiota `tapply()`. Laske myös kokonaiskeskiarvo.

```
> Ym.r <- tapply(Y, rivit, mean) # vasteen keskiarvot käsittelyryhmissä
> Ysd.r <- tapply(Y, rivit, sd)  # keskihajonnat ryhmittäin
> round(cbind(keskiarvo=Ym.r, hajonta=Ysd.r), 3) # siisti tulostus
> round(tapply(Y, lohko, mean), 3) # lohkokeskiarvot
> round(tapply(Y, lohko, sd), 3)   # lohkohajonnat
> round(mean(Y), 3)               # kokonaiskeskiarvo
```

Miten rivilukumäärä näyttäisi vaikuttavan vasteeseen?

Moodle 11

Onko lohkojen kesken näh-

tävissä merkittävää vaihtelua?

Moodle 12

- (d) Piirrä funktiolla `interaction.plot( )` kuvio, jossa vasteen arvot ovat  $y$ -akselilla ja käsittelyryhmä  $x$ -akselilla siten, että saman lohkon havainnot on yhdistetty murtoviivalla. Piirrä myös käsittelykohtaiset keskiarvot yhdistäen ne paksulla sinisellä murtoviivalla.

```
> interaction.plot(rivit, lohko, Y, type = 'o', pch = 1,
+   xlab= "Rivien lkm", ylab="Satotulos (kg/m^2)" )
> lines( 1:3, Ym.r, type="o", pch=4, cex = 1.5, lwd=2, col='blue')
```

Mitä tämä kuvio kertoo yhtäältä käsittelyn vaikutuksen muodosta ja toisaalta lohkojen välisistä tasoeroista?

4. Sovita aineistoon 1-suuntaisen varianssianalyysin malli, jossa ainoana selittäjänä on käsittelytekijä eli rivien lukumäärä penkissä (käytetään parametrintapaa (b)). Tulosta parametrien estimaatit, keskivirheet,  $t$ -suureet,  $P$ -arvot ja 95 % luottamusvälit sekä ANOVA-taulu.

```
> m1 <- lm(Y ~ rivit)
> summary(m1)      Moodle 13
> confint(m1)      Moodle 14
> anova(m1)
```

Mitä päätelmiä teet käsittelytekijän vaikutuksesta yhtäältä kertoimien estimaateista,  $P$ -arvoista ja luottamusväleistä ja toisaalta Anova-taulun  $F$ -suureesta? Ovatko nämä keskenään sopusoinnussa vai havaitsetko jotain ristiriitaa? Moodle 15 Voidaanko esim. päätellä, että odotetun vasteen erosta rivi-lkm:ien 5 ja 3 välillä on näyttöä mutta näyttö ei ole riittävä koskien rivi-lkm:ien 4 ja 3 välistä eroa?

5. Sovita aineistoon 2-suuntaisen varianssianalyysin malli, jossa käsittelytekijän lisäksi mukana on lohkotekijä. Tulosta parametrien estimaatit, luottamusvälit ja ANOVA-taulu.

```
> m2 <- lm(Y ~ rivit + lohko)
> summary(m2)      Moodle 16
> confint(m2)      Moodle 17
> anova(m2)
```

Vertaa parametriestimaatteja, keskivirheitä ja jäännöskeskineliösummaa 1-suuntaisen mallin vastaviin lukuihin. Moodle 18 Mitä havaintoja teet? Onko lohkotus tehostanut analyysiä? Muuttuvatko päätelmäsi käsittelytekijän vaikutuksesta ja millä tavalla?

## 6. Kontrastit.

- (a) Estimoi ns. lineaarinen kontrasti kerroinvektorinaan  $(-1, 0, +1)$  rivilukumäärän eri tasojen kesken ensin 1-suuntaisesta mallista `m1` (tehtävä 4), jossa lohkotekijä ei ollut mukana.

```
> library(gmodels)
> fit.contrast(m1, rivit, c(-1, 0, 1), conf.int = 0.95)
```

Miten tulkitset tätä tulosta verrattuna tehtävässä 4 saamiisi tuloksiin? Moodle 19

- (b) Estimoi vielä lineaarinen kontrasti rivilukumäärän eri tasojen kesken 2-suuntaisesta mallista `m2` (tehtävä 5), jossa lohkotekijä oli mukana.

```
> fit.contrast(m2, rivit, c(-1, 0, 1), conf.int = 0.95)
```

Muuttuvatko päätelmäsi verrattuna kohtaan (a) ja jos, niin millä tavalla? Moodle 20