Stat100, vår 2018

**Oppgaver uke 16**

**Merk at hvis du vil** **gjøre variansanalys og kontrastanalyse i R bruk følgende oppskrift:**

Hvis nødvendig: Stack data (data < active dataset <………)

Gjør variansanalyse og aktiver en model **Statistics < Means < One-way ANOVA <** ……..

Ved kontrastanalyse bruk den aktiverte modellen **Models < Test contrasts in ANOVA.**

Ved modellkritikk først bør du Lagre residualer og tilpassede verdier. **Model > add observation statistics to data.** Da har du residualene og gjennonsnittsverdiene.

**Oppgave 1**

Avl på melkemengde går hovedsakelig gjennom okse. 5 okser (med navn Åge, Hynar, Trygve, Bull og Solve) ble valgt ut for å testes hvor gode de er med hensyn på melkemengde hos døtre. Alle disse oksene har mange døtre, vi plukker tilfeldig ut 8 av disse fra hver okse som vi nummerer fra 1 til 8. For disse døtrene har vi observert årlig melkemengde (kilo melk) i ett år.

Vi lar Yij være melkemengde fra datter nummer j av okse nr. i.

Data finner du i Tabell 2 og litt av resultat av analysen i Tabell 1. Figuren nederst på neste side er residualer plottet mot tilpassede verdier.

Df Sum Sq Mean Sq F value

okse 1251009

Residuals 3901166

mean

Åge 6689

Hynar 6282

TRygve 6325

Bull 6671

Solve 6622

*Tabell 1: Resultat fra kjøring i R-commander.*

1. Skriv ned modellen som blir brukt.

Hvorfor er dette en meningsløs modell: Yij = j + ij.

Gi en grundig forklaring på *alle* størrelser som inngår i modellen.

Hva blir estimatene til alle parametere i modellen?

Finn R2 og gi en konkret forklaring på hva denne måler.

b) Utfør en test på om okse har betydning for melkeproduksjon

Skriv ned nullhypotese og alternativ hypotese.

Trekk en konklusjon på bakgrunn av testen, bruk 5 % signifikansnivå.

Åge og Solve er såkalt NRF okser (vanlig norsk rase), mens de 3 andre er såkalt Holstein Frieser okser.

c) Kan du med sikkerhet si at det er slik at generelt gir døtre av disse to NRF oksene mer melk enn døtre av disse tre Holstein Frieser oksene?

d) Hva blir residualene for datter nr. 1 og datter nr. 8 til Åge?

Hva betyr det at et residual er negativt eventuelt positivt?

Hva blir summen av samtlige residualer?

Hvilket dyr har den minste residualen og hva betyr dette?

Åge Hynar Lars Bull Gorm

6363 6337 6330 6250 6613

6697 6179 6571 6393 6317

6634 6478 6676 6992 6905

7033 6171 5970 7181 6609

6362 6465 6480 6147 6484

6038 6184 6022 6678 6777

6825 6073 6011 6434 6834

7562 6369 6546 7294 6442

*Tabell 2: Data til oppgave I*

**

**Oppgave 2 (Bruk R-commander, data ligger som R-data kalt fluer 3)** Se også uke 15.

a-e gjort i uke 15

Bananfluer ble selektert både for motstand (M) mot og ømfintlighet (O) for miljøgifter. I tillegg hadde vi en kontrollgruppe (K) som ikke ble utsatt for seleksjon. For alle tre gruppene talte vi opp antall egg lagt av hunnfluer i løpet av 14 dager tidlig i deres livssyklus

Vi ønsker å undersøker effekten av selektering med hensyn på fruktbarhet. Vi antar at seleksjoner ikke påvirker variasjoner i fruktbarhet.

a-e se også uke 15

*Finn gjennomsnitt og standardavvik i hver gruppe.*

*R\_commander: STATISTICS < SUMMARIES < NUMERICAL SUMMARIES*

*Ser det ut til at antagelsen om likt (populasjons) standardavvik er akseptabel?*

*Finn en passende grafisk framstilling av data, det det fokuseres på eventuelle forskjeller mellom grupper. Prøv diverse i GRAPHS*

*Sett opp modellen for datanalysen du vil utføre og estimer alle parametere i modellen.*

*Kan du påvise signifikant forskjell mellom gruppene med hensyn på fruktbarhet?*

*Skriv ned nullhypotese og alternativ hypotese.*

Etter at variansanalysen er kjørt svar på følgende ved hjelp av kontraster

Har seleksjon signifikant betydning?

Er det signifikant forskjell mellom det at selekteres på motstand mot eller ømfintlighet for miljøgifter?

Sjekk modellen.

Har du problemer med modellantagelsene?

M K O

189 334 180

238 181 149

126 215 247

237 268 93

195 268 262

158 191 278

109 190 147

65 327 244

295 363 169

131 332 245

231 301 197

243 249 104

310 295 150

152 233 151

213 133 168

169 195 161

220 252 234

218 208 109

248 313 157

257 168 163

298 311 232

207 129 179

297 322 298

220 361 183

111 188 371

*Tabell 1: Den første kolonnen representerer antall egg for fluer selektert på motstandsdyktighet, den andre antall egg for uselekterte og den tredje antall egg for fluer selektert på mottagelighet.*

**Oppgave 3** (bruk R-commander, **data ligger som R-data kalt bread**).

Vi har 6 brødoppskrifter vi vil teste ut. For hver oppskrift ble det bakt 8 brød. Deretter ble 48 personer tilfeldig trukket ut (anta fra populasjonen alle Nordmenn som spiser brød) og gitt et brød hver. De ble bedt om å gi dette et poeng fra 0 til 10 Vi fikk følgende resultater, hver kolonne representerer en oppskrift.

R1 R2 R3 R4 R5 R6

4 7 6 4 3 7

4 6 6 3 5 8

6 5 8 5 5 6

5 3 10 4 6 4

10 5 7 6 6 10

8 6 10 7 9 6

9 9 8 3 2 4

7 4 8 5 2 4

Gjort til uke 15: Kan du bevise at det er forskjell mellom brødoppskrifter med hensyn på hva befolkningen mener. Kan du tenke deg et bedre forsøksoppsett? (STAT210 oppgave)

**Uke 16**

De tre første oppskriftene er bakt på grov hvete, de tre siste på fin.

Oppskrift 1, 3 og 5 er svenske oppskrifter de andre er norske.

Kan du bevise (ved analyse ***av kontraster***) at

Folk flest foretrekker grovbrød?

- Det er forskjell på folks oppfatning av svensk og norsk oppskrifter?

- Hvis du skal bruke Norske oppskriftene, er grovbrød bedre likt av befolkningen enn finbrød?

**Oppgave 4** (bruk R-commander på flere av delspørsmålene. Data finner du i en Rdata fil kalt fisk)

I et forsøk på å øke effektiviteten i utnytting av naturressurser, ble det på en forsøksstasjon på Filippinene prøvd å integrere husdyrhold og akvakultur. En gravde ut 24 forskjellige fiskedammer med standard dybde og med overflate på 1200 kvadratmeter, og nummererte alle dammene. I hver dam ble det satt ut 2000 eksemplarer av arten Nile Tilapia (en liten fisk, svært mye brukt i akvakultur i u-land). Over 6 dammer ble det bygd geitehus, over 6 andre kyllinghus, i alle disse 12 dammene hang husa delvis utover overflaten, slik at all husdyrgjødsla gikk i vannet. Husa ble deretter fylt med det antall dyr som ga omtrent 3 tonn gjødsel i måneden. I 6 andre dammer ble kunstgjødsel tilsvarende omtrent samme mengde nitrogen tilført en gang i måneden. For de siste 6 dammene ble ingen næring tilsatt i det hele tatt. Etter 4 måneder ble dammene tømt, og total vekt av fisk beregnet. Dette ga følgende resultat:

geit ingen kunstgjodsel kylling

161 120 161 176

157 117 155 167

150 126 153 160

150 134 141 160

166 140 138 170

164 125 158 181

*Tabell 1. Vekt av fisk med forskjellig type gjødsel.*

1. Hvilken modell vil du bruke for å se om det var effekt av gjødslingstyper?

Forklar de ansatte på forskningsstasjonen hvorfor du velger denne modellen, og hva alle parameterene i denne betyr.

1. Utfør en test for å sjekke om det er effekt av gjødselslag.

Skriv ned nullhypotese og alternativ hypotese

Trekk konklusjon på basis av p-verdi.

Forklar de ansatte på forskningsstasjonen hva denne p-verdien uttrykker.

Siden du fikk effekt av gjødseltyper, gå dypere inn i materialet og prøv å bevise følgende:

1. Har gjødsling i forhold til å ikke gjødsle signifikant positiv effekt?
2. Er det signifikant forskjell mellom husdyrgjødsel og kunstgjødsel?
3. Er det signifikant forskjell mellom geitegjødsel og kyllinggjødsel?
4. Er det noen problemer med modellantagelsene her?

**Oppgave 5**

Merk deg hvordan eksamensoppgaver *høst 2016 Del B oppgave 1* og *vår 2017 oppgave 1* er formulert. Prøv deg på disse to. *Det finnes ikke løsningsforslag på disse to oppgavene*, og det vil heller ikke bli laget. Spør gruppelærer om hjelp.