

Stk 4070 våren 2005: Prosjektoppgave.

Frist: Torsdag 19.5 kl. 1400. (Post, mail eller levering i B821.)

Hvis flere samarbeider om å løse oppgaven, må likevel hver student formulere sin egen besvarelse med sine egne ord. Det må gå fram av besvarelsen hvem du har samarbeidet med. Skriv fødselsdatoen din (ddmmåå) øverst på arket sammen med navnet ditt.

Oppgave 1.

Denne opppgaven tar for seg et såkalt split-plot forsøk. Et slikt forsøk er randomisert i to trinn: I første trinn randomiseres relativt store enheter, deretter randomiseres små enheter innenfor disse.

Her er et eksempel: Kull med rotter føres likt innen kull, men forskjellige kull randomiseres til forskjellig fôr. Innenfor hvert kull gis hver rotte en injeksjon av et kjemisk stoff, forskjellige stoffer for hver rotte innenfor et kull. En respons y måles, for eksempel levetid.

Anta at a fôrtyper prøves, og at b forskjellige injeksjoner prøves på b rotter innen hvert kull. Eventuelle overskytende rotter i et kull tas ut av forsøket. La c være antall kull pr. fôrtype som undersøkes. En modell for forsøket kan da være:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ik} + f_{ijk}.$$

Her er α_i effekten av fôr, β_j er effekten av injeksjon, γ_{ij} er samspillet mellom fôr og injeksjon, e_{ik} er effekten av kull, og f_{ijk} er effekten av rotte. De siste to effektene er tilfeldige; resten er faste.

a) Sett opp variasjonsområdet for indeksene i , j og k , og hvilke restriksjoner effektene α_i , β_{ij} og γ_{ij} må tilfredsstille.

b) Formuler naturlige (enklest mulige) forutsetninger for de to tilfeldige effektene e_{ik} og f_{ijk} . Anta $\text{Var}(e_{ik}) = \tau^2$ og $\text{Var}(f_{ijk}) = \sigma^2$.

c) Vis at middelet pr. kull tilfredsstillt kravet til en vanlig enveis variansanalysemodell. Sett opp de kvadratsommene som trengs, finn forventningene til disse, og beskriv ut fra dette hvordan en variansanalysetest for før-effekt kan gjennomføres.

d) Finn estimatorer for β_j og for γ_{ij} , og finn variansene til disse ut fra modellen.

e) Finn forventningen til restkvadratsummen

$$SSR = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{i.k} - \bar{y}_{ij.} + \bar{y}_{i..})^2.$$

(Hint: Det uttrykket du skal regne på, vil for fast i være som et samspill.)

f) Hvor mange frihetsgrader har SSR ?

g) Hvilken varians kan estimeres ved middelkvadratsummen som svarer til SSR ?

h) Beskriv hvordan du kan finne et 95% konfidensintervall for β_j .

i) Anta at de to første injeksjonstypene er beslektet, mens den tredje er av en annen type. Sett opp to relevante kontraster, og vis hvordan du for hver av disse tester på 1% nivå en hypotese om at kontrasten er 0.

j) Finn en kvadratsum for injeksjon og forventningen til denne.

k) Finn en kvadratsum for samspill og forventningen til denne. (Hint: Sammenlign med kvadratsummen for samspill i toveis variansanalyse.)

l) Sett opp en fullstendig variansanalysetabell for split-plot forsøket over, og indiker hvordan de forskjellige effektene skal testes.

m) Hvilke effekter blir mest nøyaktig testet ved et split-plot forsøk?

Oppgave 2.

Se på situasjonen i Oppgave 8.2, side 235 i Fitzmaurice et al., og hent ned datasettet `cd4.dat` knyttet til denne oppgaven. For at alle skal få litt forskjellige datasett arbeide med, gjør følgende: Regn ut tverrsummen n i fødselsnummeret ditt `ddmmåå`, og stryk linje nummer n i datasettet.

Ta med i besvarelsen nødvendige programmeringssteg og utskrifter.

a) Konstruer to nye tidsvariable, $w1$, som er *Week* inntil uke 16 og 0 deretter, samt $w2$, som er 0 inntil uke 16 og deretter lik *Week*. Vis at $Week = w1 + w2$, og at $w1$ og $w2$ er ortogonale vektorer.

b) Vi vil gjerne beskrive virkningen av behandlingsgruppene ved 3 ortogonale kontraster. Vis at vi kan få til dette ved å definere $x1$ lik 1 for

gruppene 1 og 2, -1 ellers, x_2 lik 1 for gruppene 1 og 3, -1 ellers, og x_3 lik 1 for gruppene 1 og 4, -1 ellers. Konstruer variablene x_1 , x_2 og x_3 .

c) Diskuter hvordan kovariablene for behandling skal innføres i modell-ligningen. Det er rimelig at forventet virkning er 0 ved tid 0. Videre vil vi gjerne ha en mulighet for at responsen er forskjellig for tidsrommet inntil uke 16 og tidsrommet etter.

d) Sett opp og kjør en modell-ligning med de kovariablene som du finner rimelig. Som tilfeldige effekter, ta w_1 og w_2 innen pasient.

e) Diskuter utskriften fra denne kjøringen. Hva kan du lære fra denne utskriften om forskjellen mellom behandlingene?

f) Diskuter om det er rimelig å prøve en ny modell etter det du har lært. Gjør i tilfelle nødvendige kjøringer og diskuter resultatet.

g) Gjenta også kjøringen fra d), men nå med bare tilfeldig intercept, ikke w_1 og w_2 . Er en slik modell rimelig? Gjør en modifisert likelihood-test av denne modellen mot modellen i d).

h) Vis følgende: Dersom vi har p variable t_i som alle er t -fordelte med samme nevner og uavhengige tellere, er

$$\frac{1}{p} \sum_{i=1}^p t_i^2$$

F -fordelt med p frihetsgrader i telleren og det antall frihetsgrader du venter deg i nevneren.

i) Bruk e) og h) til å finne og utføre en total test av hypotesen om at det ikke er noen forskjell i virkningen av de forskjellige behandlingene på cd4-responsen.