

Case 7: Konfunderinger i ufuldstændigt blokforsøg

Ugens case handler om konfundering, partiel konfundering og ufuldstændige blokforsøg. En speciel pointe med casen er at forsøge at give lidt mere forståelse af forskellen på at modellere blokeffekten som systematisk eller tilfældig; denne forskel er snævert knyttet til begrebet “recovery of inter-block information”. Casen er bygget over datamateriale og forsøgsdesign fra kompendiets (BMS) opgave 9.4.

Link til datasættet `bms_exercise9-4.txt` findes på kursusplanen for uge 7.

Læs først teksten til kompendiets opgave 9.4. Brug helst max. en halv time på at besvare de første to spørgsmål (svarende til opgave 9.4 i kompendiet) og undgå såvidt muligt at bruge hjælp hertil eller at læse videre i opgaveteksten.

- Hvilken slags forsøg er der tale om og hvordan udføres randomisering i et sådant forsøg. Tænk herunder over hvilke faktorer der indgår i forsøget, hvilke niveauer de har og om de skal modelleres som tilfældige eller systematiske.
- Opskriv en statistisk model på sædvanlig (matematisk) form og vha. R-kode, og analysér data og drag konklusioner.
- Der er tale om konfundering i dette forsøg; hvilke(n) effekt(er) er konfunderet?

Formålet med delspørgsmål d)-i) er at undersøge, hvordan forsøgsdesignet har indflydelse på muligheden for at estimere / kvantificere effekten af forskellige vekselvirkninger på baggrund af det indsamlede datamateriale. Opdelingen i de 6 delspørgsmål er vejledende og kan virke lidt forvirrende indtil den overordnede pointe bliver helt klar. Prøv derfor i første omgang blot at gøre nogle ting som virker relevante, også selvom du ikke lige følger den foreslåede rækkefølge.

- Kør modellen med alle hoved- og vekselvirkninger mellem de tre faktorer af interesse, både med og uden blok som additiv (altså uden vekselvirkninger) systematisk effekt. Hvordan kommer svaret fra c) til udtryk i R-udskrifterne?
- Effekten af vekselvirkningen mellem N og P kan repræsenteres ved dobbeltdifferencen

$$(\gamma(N_2, P_2) - \gamma(N_2, P_1)) - (\gamma(N_1, P_2) - \gamma(N_1, P_1)),$$

i modellen hvor $\gamma(\dots)$ er leddet som beskriver vekselvirkningen. Det er denne størrelse som automatisk anføres, når R fitter modellen. Beregn et estimat og et konfidensinterval for effekten af vekselvirkningen mellem N og P.

- f) Forsøg at gøre det samme for tre-faktorvekselvirkningen mellem N, P og K, som tilsvarende kan specificeres (og i R udskrives) som en tre-dobbelt differens

$$\begin{aligned} & \left((\eta(N_2, P_2, K_2) - \eta(N_2, P_1, K_2)) - (\eta(N_1, P_2, K_2) - \eta(N_1, P_1, K_2)) \right) \\ - & \left((\eta(N_2, P_2, K_1) - \eta(N_2, P_1, K_1)) - (\eta(N_1, P_2, K_1) - \eta(N_1, P_1, K_1)) \right) \end{aligned}$$

ud fra tre-faktorvekselvirkningen $\eta(\dots)$ i modellen. Hvad sker der? Forsøg at forklare hvorfor?

- g) Besvar spørgsmål (e) igen, men nu med blokke som tilfældig (random) effekt i modellen. Sammenlign resultaterne og kommentér.
- h) Besvar spørgsmål (f) igen, men nu med blokke som tilfældig (random) effekt i modellen. Sammenlign resultaterne og kommentér.
- i) Opsummer resultaterne fra (e)–(h) i ord, og forsøg derved at forklare virkningen af at modellere blokeffekten som tilfældig. Kan du se en analogi til split-plot forsøg?
- j) Kan du foreslå en forsøgsplan som i stedet for at konfundere tre-faktorvekselvirkningen helt ville have gjort det muligt i højere grad at estimere og teste alle hoved- og vekselvirkninger, uanset om blokke indgår med systematisk eller tilfældig effekt?
- k) Hvis du har tid: Hvor mange blokke ville du gerne (mindst) have til rådighed for at kunne estimere (og teste) alle syv hoved- og vekselvirkninger lige godt? Hvordan kan man i så fald også karakterisere det forsøg der fremkommer?
- l) Hvis du stadig har tid: Hvor stor en LSD-værdi ville du få til sammenligning af to behandlinger i forsøget fra spg. (k)? (her skal bruges estimator for varianskomponenterne fra analysen i (b)).

Hjælp til nogle af spørgsmålene

- a) Læs afsnit 9.3 hvis du er i tvivl her.
- b) Her er vi interesseret i effekterne af N, P og K. Lad blokkene indgå som en faktor med tilfældig virkning, selv om det ikke har den store betydning lige her.
- c) Se på blok 1 og 2 alene. Hvilken effekt er konfunderet der? Har du problemer her, så læs mere i afsnit 9.3. Samme spørgsmål for blok 3 og 4. Og igen blok 5 og 6. Hvilke effekter er slet ikke konfunderet i nogle af parrene af blokke?
- d) Prøv ligesom i (b) at gå i gang med testene og se hvad der sker. Hvis der er en effekt du ikke kan teste så drop den fra modellen og fortsæt for at se om der skulle være flere af den slags.
- e) Kør `summary()` på den relevante model som du skal bruge til at estimere vekselvirkningen mellem N og P. Så kan du se hvilke parametre der indgår og i hvilken rækkefølge. Du skal bruge `estimable` på det koefficientsæt, lavet med `c(...)`, som svarer til den funktion vi vil have, og du skal kun bruge estimerer der vedrører vekselvirkningen, resten af koefficienterne skal være 0.

Hvis du strandeder i dette spørgsmål, vil du heller ikke kunne løse (f)-(i), og i så fald skal du springe ned til (j) og senere sætte dig ind i hvordan `estimable()` benyttes.
- f) Samme metode som i (e), blot nu vedrørende trefaktor vekselvirkningen.
- g) Som i (e).
- h) Som i (f).
- j) Det virker lidt extravagant at konfundere $N \times P \times K$ tre gange: i blok 1 og 2 delforsøget, i blok 3 og 4, og igen i blok 5 og 6. Mon ikke du kan finde på noget bedre?
- k) Hvor mange blokke skal du bruge til en komplet gentagelse af alle behandlinger med præcis en effekt konfunderet? Hvad så hvis du vil konfundere alle effekter lige ofte? Det vil bevirke at alle 2^3 behandlinger sammenlignes med samme præcision, selv om det fortsat er et ufuldstændigt blokforsøg. Siger det dig noget?
- l) Med den rette forsøgsplan fra spørgsmål (k) kan du bruge formel (9.11) på side 169 i BMS. hertil skal bruges variansestimaterne som du får fra R med `VarCorr(...)`.