

Andenårsvirkning af fosfortilførsel på forsøgsmarker

Vi piner endnu engang datasættet `fosfor.txt` fra ugeopgave 1.3 og 1.9 (se link fra kursusoversigten) for at se om det vil afsløre noget mere om andenårsvirkningen af fosforgødskning. Lad os lige repetere forsøgsdesignet: marken var inddelt i 3 blokke med 9 parceller i hver, og der var 9 gødskningskombinationer (behandlinger) som blev afprøvet en gang i hver blok. De 9 kombinationer er sammensat af de tilførte fosformængder i 1981 og 1982, der er angivet som `p81` og `p82` i datafilen. Som responsvariabel benyttes udbyttet (`udbytte`) i 1982 for de 27 parceller i forsøget. Datasættet er organiseret således, at hver datalinje svarer til en parcel.

Baggrunden for forsøget var, at man ved et-årige gødningsforsøg på forsøgsmarker undertiden brugte marker, hvor der også året forinden havde været gødningsforsøg. Derfor kunne resultaterne blive upålidelige, hvis der var en effekt af den gødning, der havde været brugt året før på de enkelte parceller. Mistanken var især rettet mod fosfor. Samtidigt havde forsøget det mål at kvantificere langtidsvirkningen af fosfor; her dog kun andenårsvirkningen i forhold til førstearsvirkningen (altså virkningen på samme års udbytte). Forsøget er udført af Karsten Attermann Nielsen (KVL-speciale).

Formålet med ugens case er dels at få repeteret den additive model for blokforsøg, og dels at opnå fortrolighed med hvad denne og beslægtede modeller (herunder regression) udtrykker. Ved at vende og dreje modellerne håber vi desuden på, at få kvantificeret langtidseffekten af gødskning med fosfor på en letfortolkelig måde. Selv om der kun er 27 observationer her, kræver det faktisk en ret dyb analyse at komme helt til bunds i problemstillingen.

Link til datasættet `fosfor.txt` kan findes på kursusoversigten.

Første del: Den additive model for tofaktor forsøg

- 1a) Gem filen med datasættet på din PC og indlæs dernæst datasættet i R som en dataframe med navnet `kart`. Forsøg at danne dig et overblik over datasættet: hvilke vektorer har du til rådighed, og hvad er længden af dem?
- 1b) Overvej følgende: Hvilke faktorer er der i forsøget og hvilke niveauer har de? Find ud af om R opfatter de forskellige variable som faktorer eller numeriske vektorer? Hvilke faktorer er balancerede? Afprøv eventuelt kommandoerne `levels(...)` og `table(...)` på forskellige variable i datasættet.

Forsøgets struktur er i sin helhed for omfattende til, at du har lært rutine-variansanalysen af forsøget, men du kan alligevel komme et langt stykke ad vejen ved hjælp af (den additive model for) tosidet variansanalyse og regressionsanalyser. Dette skal du prøve i det følgende.

- 1c) Hvis du har mod på det, kan du først på egen hånd forsøge at undersøge, hvordan udbyttet afhænger af gødskningsfaktoren og blokfaktoren. Fit nogle modeller i R og tænk over, hvilke modeller det giver mening at teste imod hinanden. Husk at gødskningsfaktoren er sammensat af de tilførte fosformængder i 1981 og 1982 og brug dette i analysen. Hvis du går i stå, kan det være en god ide, hvis du i stedet lader dig styre af de følgende spørgsmål d)-h).
- 1d) Forsøgsdesignet beskrevet i Opgave 1.3 (-se opgavearket fra tirsdag) er et eksempel på et fuldstændigt blokforsøg med to faktorer: hver række (nummereret 1-3) kan man tænke på som en mark (=blok), der hver især er opdelt i 9 parceller. Inden for hver blok har man sørget for at afprøve hver af de 9 fosforbehandlinger på netop en af parcellerne. Prøv at fitte en model (vekselvirkningsmodel eller additiv model) her, og forsøg at hive konklusioner ud vedrørende behandlingseffekten (af variabelen fosfor) ved alene at se på parameterestimaterne. Du skal her køre en analyse i R af formen `model1 <- lm(...)`, men du skal også opskrive modellen på papir!
- 1e) Test hypotesen om, at der ikke er nogen effekt af behandlingen. Hvilke spørgsmål får du derved svar på? Hvilke spørgsmål fra den oprindelige problemstilling giver denne analyse ikke svar på?
- 1f) Hvis der ikke er nogen andenårseffekt af fosforgødningen, kan behandlingsfaktoren (=fosfor) reduceres til nogle færre niveauer (Hvor mange og hvilke?) Denne "nye" behandlingsfaktor skal vi bruge i en reduceret model, som du nu skal fitte i R med `model2 = lm(...)`. Skriv på papir, hvilken hypotese `model2` svarer til i `model1`.
- 1g) Test ovennævnte hypotese med `anova(model2, model1)`. Hvad konkluderer du heraf?
- 1h) For at benytte de metoder, du kender, ville det være rart at kunne se bort fra blokeffekterne. Er det rimelig at se bort fra blokeffekten her?

Anden del: estimation af effekterne - regressionsanalyser

Du burde nu gennem første del af casen have dokumenteret, at der er en anden-års effekt af fosfor (og tilsvarende ville du kunne eftervise en første-års effekt). Vi ønsker nu at kvantificere effekten og har til det formål knapt så meget brug for gentagelserne over blokkene. Derfor regner vi i anden del alene på de 9 gennemsnit der fremkommer ved for hver behandling at tage gennemsnit over de tre blokke.

Link til datasættet med gennemsnittene `fosfor-means.txt` kan findes på kursusoversigten under uge 1.

- 2a) Hvor mange faktorer (og tilhørende niveauer) er der nu?
- 2b) Indlæs data. Find en god måde at tegne disse data i en figur. (Tænk på strukturen af data).
- 2c) Opskriv (på papir) og tilpas (i R) en additiv variansanalysemodel.
- 2d) Hvad bliver estimatorne, og hvad betyder det for andenårseffekten?
- 2e) Estimatorne antyder, at der godt kunne være en næsten retlinet sammenhæng mellem p82 og udbyttet. Hvorfor? Gælder det tilsvarende for p81?
- 2f) Tilpas (i R) en ret linie til de tre middel-udbytter med `p81=0` som funktion af fosfortilførslen i 1982. Gør det samme for de to andre værdier af p81. Er der væsentlig forskel på de 3 hældninger?
- 2g) Prøv at opstille en model for de 9 middel-udbytter, hvori der er retlinet effekt af såvel p81 som p82. Kan du give koefficienterne en fortolkning, som giver et kvantitativt mål for, hvor stor andenårsvirkningen er i forhold til førsteårsvirkningen?
- 2h) Det vil være flot, hvis du også kan tilpasse denne sidstnævnte model i R, selv om det ikke direkte er en model, du sikkert ikke direkte har set i SD1 (eller på tidligere kurser).

Mon ikke tiden er ved at være brugt op for de fleste af jer her? Det kan dog bemærkes at en sidste nuance knytter sig til de tre hældningskoefficienter fra (2f). Disse er nemlig lidt forskellige, omend vi ikke her formelt har undersøgt, om hældningerne er signifikant forskellige. Dog går forskellene i en naturlig retning som antyder en “fosfor-mætning” eller i hvert fald afdæmpning af effekten af fosfortilførslen i 1982 med stigende mængde tilsat fosfor i 1981.

Hjælp til nogle af spørgsmålene

- 1a) Højreklik på linket og vælg “gem destinationen som”.
Brug dernæst `read.table(file.choose(), header=T)`. Udskriv hele datasættet på skærmen i R.
- 1b) En af faktorerne er behandlingen (`fosfor`). Den hænger sammen på en naturlig måde med to af de andre.
- 1d) Du skal bruge den additive model for blokforsøg. Husk brugen af `factor(...)` i R.
- 1e) Får du at vide, om der er forskel på (effekten af) de 9 gødningsbehandlinger? Effekt af fosfortilførslen i 1981? i 1982?
- 1f) Den nye faktor står allerede i datasættet. Hvilke grupper inddeler den observationerne i, og hvilke grupper fra før bliver derved slået sammen?
- 1g) Du skulle gerne nå frem til at udtale dig om effekten af enten p81 eller p82. Hvilken handler hypotesen om?
- 1h) Det kan desuden være værd at overveje, hvad konsekvensen er af at ignorere blokinddelingen, hvis den faktisk har betydning.
- 2b) Hvis der var to faktorer: behandling og blok, ville der være en standard-figur. Her er det to andre faktorer, men samme tegning kan bruges.
- 2c) Additiv model for to-faktor forsøg. Med de rigtige to faktorer.
- 2e) Prøv at indtegne de tre estimater svarende til p82 i en figur, hvis du er i tvivl. Tilsvarende for p81.
- 2g) Hædningskoefficienten er stigningen i udbytte pr. tilført enhed fosfor.