Chapter 8 Workshop

Table of contents

Dataset `alfalfa`

This dataset is from the R package `faraway`. The alfalfa dataset frame has 25 rows and 4 columns. Data comes from an experiment to test the effects of seed inoculum, irrigation and shade on alfalfa yield.

This data frame contains the following columns:

shade - Distance of location from tree line divided into 5 shade areas

irrigation - Irrigation effect divided into 5 levels

inoculum - Four types of seed incolum, A-D with E as control

yield - Dry matter yield of alfalfa

```
library(tidyverse)

data(alfalfa, package="faraway")

head(alfalfa, 25)
```

Exercise 8.1

Obtain the main effects and interaction plots.

Old style Main effects plots:

```
mod1 <- aov(yield ~ shade + irrigation + inoculum,
            data=alfalfa)

library(effects)

plot(allEffects(mod1))
```

Old style Interaction effects plots:

```
mod2 <- aov(
  yield ~ shade*irrigation*inoculum - shade:irrigation:inoculum,
  data = alfalfa
  )


library(effects)

mod1 <- lm(yield ~ shade * irrigation, data = alfalfa)

effect('shade:irrigation',
       mod = mod1) |>
  plot(multiline = TRUE)
```

```
mod2 <- lm(yield ~ shade * inoculum, data = alfalfa)

effect('shade:inoculum',
       mod=mod2) |>
  plot(multiline=TRUE)
```

```
mod3 <- lm(yield ~ irrigation * inoculum, data = alfalfa)

effect('irrigation:inoculum',
       mod=mod3) |>
  plot(multiline=TRUE)
```

ggplot2 can produce good main effects and interaction plots but the R codes for this task are not short. For main effects plot-

```
library(ggplot2)

plot1 <- ggplot(alfalfa) +
  aes(x = shade, y = yield) +
  stat_summary(fun = mean, geom = "point", aes(group = 1)) +
  stat_summary(fun = mean, geom = "line", aes(group = 1)) +
  geom_hline(aes(yintercept = mean(yield)), alpha = .7) +
  ggtitle("Main effect of shade")

plot2 <- ggplot(alfalfa) +
  aes(x = irrigation, y = yield) +
  stat_summary(fun = mean, geom = "point", aes(group = 1)) +
  stat_summary(fun = mean, geom = "line", aes(group = 1)) +
  geom_hline(aes(yintercept = mean(yield)), alpha = .7) +
  ggtitle("Main effect of irrigation")

plot3 <- ggplot(alfalfa) +
  aes(x = inoculum, y = yield) +
  stat_summary(fun = mean, geom = "point", aes(group = 1)) +
  stat_summary(fun = mean, geom = "line", aes(group = 1)) +
  geom_hline(aes(yintercept = mean(yield)), alpha = .7) +
  ggtitle("Main effect of inoculum")

library(patchwork)
plot1+plot2+plot3
```

For interaction plot-

```
#Interactions Plot
plot4 <- ggplot(alfalfa) +
  aes(x = shade, y = yield,
      group = irrigation, colour = irrigation) +
```

```
  stat_summary(fun=mean, geom="point")+
  stat_summary(fun=mean, geom="line")+
  geom_hline(aes(yintercept = mean(yield)), alpha = .7) +
  ggtitle("shade*irrigation interaction")

plot5 <- ggplot(alfalfa) +
  aes(x = inoculum, y = yield,
      group = irrigation, colour = irrigation) +
  stat_summary(fun=mean, geom="point")+
  stat_summary(fun=mean, geom="line")+
  geom_hline(aes(yintercept = mean(yield)), alpha = .7) +
  ggtitle("inoculum*irrigation interaction")

plot6 <- ggplot(alfalfa) +
  aes(x = shade, y = yield,
      group = inoculum, colour = inoculum) +
  stat_summary(fun=mean, geom="point")+
  stat_summary(fun=mean, geom="line")+
  geom_hline(aes(yintercept = mean(yield)), alpha = .7) +
  ggtitle("shade*inoculum interaction")

plot4 / plot5 / plot6
```

Exercise 8.2

Fit one-way ANOVA models to this dataset.

```
anova1 <- aov(yield ~ shade, data = alfalfa)
summary(anova1)

anova2 <- aov(yield ~ irrigation, data = alfalfa)
summary(anova2)

anova3 <- aov(yield ~ inoculum, data = alfalfa)
summary(anova3)

plot(TukeyHSD(anova1))
```

```
plot(TukeyHSD(anova2))
```

```
plot(TukeyHSD(anova3))
```

Exercise 8.3

Fit a three-factor (additive) ANOVA model without interactions.

```
anova1 <- aov(yield ~ shade + irrigation + inoculum,
              data = alfalfa)

summary(anova1)

library(ggfortify)
autoplot(anova1, 1)
```

For the base R style four diagnostic plots, use `plot(anova1)` and set the `par`.

```
par(mfrow=c(2,2))
plot(anova1)
```

Exercise 8.4

Fit the indicator variable regression model of `yield ~ inoculum`

```
alfalfa <- alfalfa |>
  mutate(
    I.A = as.numeric(inoculum=="A"),
    I.B = as.numeric(inoculum=="B"),
    I.C = as.numeric(inoculum=="C"),
    I.D = as.numeric(inoculum=="D")
  )

indicator.reg <- lm(yield ~ I.A + I.B + I.C + I.D,
                    data = alfalfa)

summary(indicator.reg)
```

Note that this regression allows the treatments to be compared with the control.

Exercise 8.5

`shade` is a categorical variable of factor codes but let us (incorrectly) treat it as numerical (and if the actual distances are given, then the `distance` variable becomes a covariate). Fit ANCOVA of yield on the `inoculum` factor and `shade` covariate.

R:

```
ancova.model <- lm(yield ~ as.numeric(shade) * inoculum,
                   data = alfalfa)
summary(ancova.model)
```

Exercise 8.6

In a two-factor experiment, one of the factors was assigned to main plot (main-plot factor), the second factor, called the subplot factor, was assigned into subplots. The dataset `https://www.massey.ac.nz/~kgovinda/data/plots.RData` gives the experimental set up. Perform the ANOVA for this basic split-plot experiment.

R:

```
url1 <- "https://www.massey.ac.nz/~anhsmith/data/plots.RData"
download.file(url = url1, destfile = "plots.RData")
load("plots.RData")

plots
```

```
sp.model <- aov(yield ~ block + A*B + Error(block/A),
                data=plots)
summary(sp.model)
```

```
#Incorrect model
summary(aov(yield ~ block + A*B ,
            data=plots))
```

- More R code examples are here