Code

One-Way ANOVA on Numerical Tables

Objective of the lesson

In class we handed out ``Examples of ANOVA". Below we redo the example using R.

1. **Set up an analysis of variance test for the following data showing yields from three fertilizers- F1, F2, and F3.**

F1: 18.2, 20.1, 17.6, 16.8, 18.8, 19.7, 19.1

F2: 17.4, 18.7, 19.1, 16.4, 15.9, 18.4, 17.7

F3: 15.2, 18.8, 17.7, 16.5, 15.9, 17.1, 16.7

There are three groups with seven observations per group. We denote group i values by yi:

Hide

```
y1 = c(18.2, 20.1, 17.6, 16.8, 18.8, 19.7, 19.1)
y2 = c(17.4, 18.7, 19.1, 16.4, 15.9, 18.4, 17.7)
y3 = c(15.2, 18.8, 17.7, 16.5, 15.9, 17.1, 16.7)
```

Now we combine them into one long vector, with a second vector, group, identifying group membership:

Hide

```
y = c(y1, y2, y3)
n = rep(7, 3)
Fertilizer = rep(1:3, n)
```

Now we show summary statistics by group and overall. We locally define a temporary function, `tmpfn` , to make this easier.

Hide

```
tmpfn = function(x) c(sum = sum(x), mean = mean(x), var = var(x), n = length(x))
```

Hide

```
tapply(y, Fertilizer, tmpfn)
```

```
$`1`
      sum       mean        var          n
130.300000  18.614286   1.358095   7.000000

$`2`
      sum       mean        var          n
123.600000  17.657143   1.409524   7.000000

$`3`
      sum       mean        var          n
117.900000  16.842857   1.392857   7.000000
```

While we could show you how to use R to mimic the computation of SS by hand, it is more natural to go directly to the ANOVA table.

Hide

```
data = data.frame(y_values = y, Fertilizer = factor(Fertilizer))
```

Hide

```
knitr::kable(data, format="markdown")
```

y_values	Fertilizer
18.2	1
20.1	1
17.6	1
16.8	1
18.8	1
19.7	1
19.1	1
17.4	2
18.7	2
19.1	2
16.4	2
15.9	2
18.4	2
17.7	2
15.2	3
18.8	3
17.7	3
16.5	3
15.9	3
17.1	3
16.7	3

Displaying the dataframe in a better format

Hide

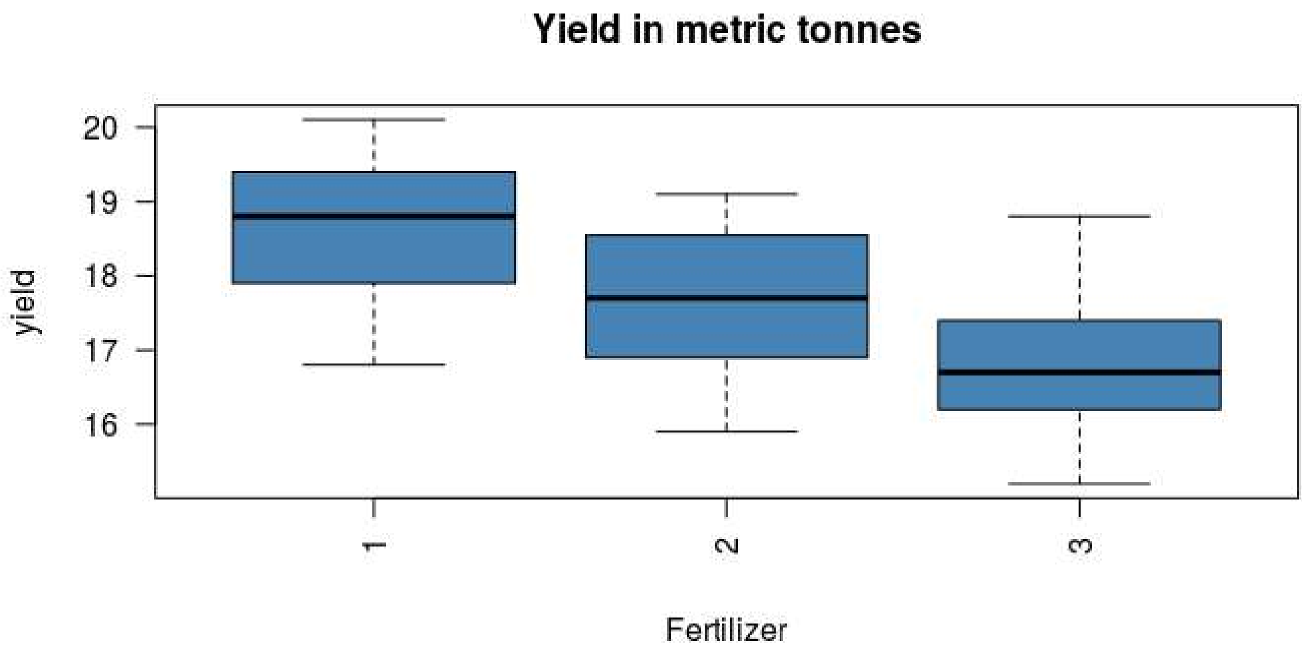
```
library(stargazer)
stargazer(data,                 # Export txt
          summary = FALSE,
          type = "text",
          out = "data_stargazer_txt.txt")
```

```
======================
   y_values Fertilizer
----------------------
1   18.200      1
2   20.100      1
3   17.600      1
4   16.800      1
5   18.800      1
6   19.700      1
7   19.100      1
8   17.400      2
9   18.700      2
10  19.100      2
11  16.400      2
12  15.900      2
13  18.400      2
14  17.700      2
15  15.200      3
16  18.800      3
17  17.700      3
18  16.500      3
19  15.900      3
20  17.100      3
21  16.700      3
----------------------
```

Hide

```
par(mar=c(8, 4.1, 4.1, 2.1))

#create boxplots
boxplot(y_values ~ Fertilizer,
data = data,
main = "Yield in metric tonnes",
xlab = "Fertilizer",
ylab = "yield",
col = "steelblue",
border = "black",
las = 2 #make x-axis labels perpendicular
)
```



Hide

```
my_anova=aov(y ~ Fertilizer, data)
summary(my_anova)
```

Perform Post-Hoc Tests (If Necessary)

If the p-value in the ANOVA output is less than .05, we reject the null hypothesis. This tells us that the mean value between each group is not equal. However, it doesn't tell us which groups differ from each other.

In order to find this out, we must perform a post hoc test. In R, we can use the `TukeyHSD()` function to do so:

Hide

```
TukeyHSD(my_anova)
```

So it is reasonable to think that the difference between Fertilizer 1 and 3 is significant at 5% level.

2. **Set up an analysis of variance table for the following per acre production data for three varities of wheat, each grown on 4 plots and state if the variety differences are significant.**

V1: 6 7 3 8

V2: 5 5 3 7

V3: 5 4 3 4

Two-way anova using R

1. **Set up analysis of variance table for the following two-way design results.**

Hide

```
dw <- read.table(header=T, text='
 Fertilizer A  B C
   W   6    5     5
   X   7    5     4
   Y   3    3     3
   Z   8    7     4
 ')
dw
```

Hide

```
library(tidyr)
library(dplyr)
```

Hide

```
data=dw %>%
  gather(v, yield, A:C) %>%
  separate(v, c("seed")) %>%
  arrange(Fertilizer)
```

Hide

```
data
```

Hide

```
table(data$Fertilizer, data$seed)
```

Exploring the Data

Before we even fit the two-way ANOVA model, we can gain a better understanding of the data by finding the mean and standard deviation of weight loss for each of the six treatment groups using the dplyr package:

Hide

```
data %>%
  group_by(Fertilizer) %>%
  summarise(mean = mean(yield),
            sd = sd(yield))
```

Visualization

We can also create a boxplot for each of the six treatment groups to visualize the distribution of weight loss for each group:

Hide

```
par(mar=c(8, 4.1, 4.1, 2.1))

#create boxplots
boxplot(yield ~ Fertilizer,
data = data,
main = "Yield in metric tonnes",
xlab = "Fertilizer",
ylab = "Seed",
col = "steelblue",
border = "black",
las = 2 #make x-axis labels perpendicular
)
```

Anova

Hide

```
model = aov(yield~Fertilizer*seed, data = data)
#view the model output
summary(model)
```