Lesson 9 - Strawberries

LTC James K. Starling

Last compiled on 31 January, 2022

A study (Smith & Skog, "Postharvest Carbon Dioxide Treatment Enhances Firmness of Several Cultivars of Strawberry," American Society for Horticulture Science, 27(5), 1992) was conducted in order to investigate the effects of storage method on the firmness of strawberries. As strawberries are harvested, they are placed into storage containers called "clamshells." Because clamshells have holes in them, air can circulate from outside the clamshell into the clamshell and vice versa while the container still protects the fruit. To test the potential effect of changing the composition of the air in which strawberries are stored, researchers obtained 15 clamshells filled with strawberries of several different varieties. Five clamshells were randomly assigned to be stored in regular air (approximately 21% oxygen and 0.04% carbon dioxide) and five clamshells were randomly assigned to be stored in a modified air method (18% oxygen and 15% carbon dioxide). After two days of refrigeration at 0.5°C in one of the two different atmospheres, the strawberries were then left at room temperature in regular air for two more days. The firmness of the strawberries was then measured by the force (measured in newtons, N) needed for a probe to pierce the exterior of the strawberry. The remaining five clamshells were randomly assigned to a "control" condition which measured the strawberries immediately (not after four days like the other two conditions).

```
### Loading packages
library(tidyverse)
library(ggResidpanel)
## Load data

## three ways to read in the data:
# berries <- read.table('http://www.isi-stats.com/isi2/data/StrawberryStorageRCBD.txt',
#                       header = TRUE, stringsAsFactors = TRUE)
# berries <- read.table('https://raw.githubusercontent.com/jkstarling/MA376/main/StrawberryStorageRCBD.
#                       header = TRUE, stringsAsFactors = TRUE)

setwd('C:/Users/james.starling/OneDrive - West Point/Teaching/MA376/JimsLessons/Block II/LSN9/')
berries <- read.table('StrawberryStorageRCBD.txt', header = TRUE, stringsAsFactors = TRUE)
```

Background

(1) Identify the observational units, the explanatory variable, and the response variable. Identify hypothesized Sources of explained variation and unexplained variation.

Observational Units:

Response Variable:

Explanatory Variable:

Hypothesized Explained Variation:

Hypothesized Unexplained Variation:

(2) Calculate the effects for each storage method. (Hint: use `summarise()` and `group_by()` as necessary)

```
#Calculate the overall average
# Mean <-
#calculate the group effects
# berries %>%
```

(3) Does the storage method appear to explain some of the variability in the firmness levels? How are you deciding?

(4) Calculate the ANOVA table for the (one variable) separate means model. (Use `aov()` or a linear model with `anova` to display the output.)

What is the proportion of variation explained by storage method?

```
# anova.simple <-
# summary(anova.simple)
# ##or
# model.simple <-
# anova(model.simple)
#R-squared
```

Model Update Rather than limiting themselves to one type of strawberry (i.e., using inclusion criteria of one type of strawberry) which would most likely have given them a smaller SSError, the researchers decided to include five different varieties of strawberries in the study allowing them to generalize to a more diverse population of strawberries. Suspecting that different varieties of strawberries have different firmness ratings, that is, believing variety is associated with firmness (the response variable), the researchers opted to **block** on strawberry variety. In this way the variation explained by variety will be accounted for in the analysis by removing it from the unexplained variation.

(5) Create a plot of Firmness vs Variety? (Use `geom_point()` with `ggplot()`. Set `x=Firmness`, `y=Variety`, and `color=Variety`) Do the data confirm or refute the researchers' prediction that different strawberry varieties tend to have different firmness ratings? Is there a strawberry variety that tends to have very low firmness? Very high?

```
# berries %>%
```

(6) Color your plot from the previous question by `Storage`. For the Vesper variety, which storage method had the largest firmness? Is this the case for all of the varieties? Looking at this graph, do you think the storage method effect will be statistically significant after adjusting for variety? Explain your reasoning.

```
# berries %>%
```

A block study design creates blocks of experimental units that are similar to each other, randomly assigns the treatments within each block, and then analyzes the data in a way which accounts for block-to-block variations. When there are only two groups being compared a block study design is called a matched pairs design. The term block comes from the first block designs which were agricultural experiments in large fields where separate parts of the field were called "blocks."

(7) How would we implement randomization when we block on strawberry type?

(8) Consider the study we did last lesson with baking chips (chocolate and butterscotch). What were the 'blocks' that we considered? Explain how exploring the melting time of a white-chocolate chip would make the study a block study design.

Modifying the Analysis to Consider Blocking. Now that we know more about the actual way this study was designed, we know that an analysis that ignores this "restricted randomization" is incorrect. Instead, the analysis should reflect the study design that was used (and account for the blocking).

(9) Conduct an analysis of a Two-variable ANOVA accounting for treatment and blocks. (Use `aov()` or a linear model with `anova()` (after using effect encoding)). Calculate the R-squared value. Compare the results with the single-variable ANOVA calculated above.

```
## Using AOV
# blocked.aov <-
# summary(blocked.aov)

## Calculating an anova table from a linear regression model using Effect Encoding to account for varie
```

```
# berries2 <- berries
# contrasts(.....
# contrasts(.....

# blocked.model <- ....
# summary(blocked.model)
# anova(blocked.model)

# R-squared
```

(10) Do we meet the validity conditions for a two variable ANOVA? (see p. 163)

```
# resid_panel(...
#
# berries %>% ....
```

References

Nathan Tintle et al.(2019). Intermediate Statistical Investigations.