Homework Chapter 6 - Mark Cappiello

1. The data sets package (installed in R by default) contains a data set called InsectSprays that shows the results of an experiment with six different kinds of insecticide. For each kind of insecticide, n = 12 observations were conducted. Each observation represented the count of insects killed by the spray. In this experiment, what is the dependent variable (outcome) and what is the independent variable? What is the total number of observations?

library(datasets)  
data("InsectSprays")  
head(InsectSprays)

## count spray  
## 1 10 A  
## 2 7 A  
## 3 20 A  
## 4 14 A  
## 5 14 A  
## 6 12 A

str(InsectSprays)

## 'data.frame': 72 obs. of 2 variables:  
## $ count: num 10 7 20 14 14 12 10 23 17 20 ...  
## $ spray: Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...

In the InsectSprays data set the dependent variable is “count” and the independent variable is “spray”.

the total number of observation in the InsectSprays data set is 72.

1. After running the aov() procedure on the InsectSprays data set, the “Mean Sq” for spray is 533.8 and the “Mean Sq” for Residuals is 15.4. Which one of these is the between-groups variance and which one is the within-groups variance? Explain your answers briefly in your own words.

model <- aov(count ~ spray, data = InsectSprays)  
summary(model)

## Df Sum Sq Mean Sq F value Pr(>F)   
## spray 5 2669 533.8 34.7 <2e-16 \*\*\*  
## Residuals 66 1015 15.4   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The between-groups variance is the “Mean Sq” for spray (533.8)

The between-groups variance is used to compare the variability of data when you are comparing data across groups. This shows variability that the variability here is due to the different sprays.

The within-groups variance is the “Mean Sq” for Residuals (15.4)

The within-groups variance is used to compare the variability of data when you are comparing data inside the same group. This shows that the variability here is most likely due to random error.

1. Based on the information in question 2 and your response to that question, calculate an F-ratio by hand or using a calculator. Given everything you have earned about F-ratios, what do you think of this one? Hint: If you had all the information you needed for a Null Hypothesis Significance Test, would you reject the null? Why or why not?

Calculating:

F-ratio = Mean Sq between-groups / Mean Sq within-groups

F-ratio = 533.8 / 15.4

F-ratio = 34.66

This F-ratio seems to be pretty large so it means that the variability between-groups is much larger than the variability within-groups.

A high F-ratio says that the differences between the group means are statistically significant. So if you were conducting a Null Hypothesis Significance Test you would reject the null hypothesis. Because a high F-ratio shows that the group means are not all equal, and the type of spray or the independent variable most likely has a significant effect.

It would be a reasonable conclusion to say that the type of insect spray has a significant effect on the outcome.

1. Continuing with the InsectSprays example, there are six groups where each one has n = 12 observations. Calculate the degrees of freedom between groups and the degrees of freedom within groups. Explain why the sum of these two values adds up to one less than the total number of observations in the data set.

the degrees of freedom between groups

df between-groups = number groups - 1

df between-groups = 6 - 1

df between-groups = 5

the degrees of freedom within groups

df within-groups = total observations - number of groups

df within-groups = 72 - 6

df within-groups = 66

df within + df between = observations - 1

66 + 5 = 71

Estimating the mean of all observations uses up one degree of freedom. This leaves n−1 degrees of freedom that is split between the variability between the groups and the variability within the groups.

1. Use R or R-Studio to run the aov() command on the InsectSprays data set. You will have to specify the model correctly using the “~” character to separate the dependent variable from the independent variable. Place the results of the aov() command into a new object called insectResults. Run the summary() command on insectResults and interpret the results briefly in your own words. As a matter of good practice, you should state the null hypothesis, the alternative hypothesis, and what the results of the null hypothesis significance test lead you to conclude.

insectResults <- aov(count ~ spray, data = InsectSprays)  
summary(insectResults)

## Df Sum Sq Mean Sq F value Pr(>F)   
## spray 5 2669 533.8 34.7 <2e-16 \*\*\*  
## Residuals 66 1015 15.4   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Null Hypothesis: The type of spray has no effect on the number of insects counted.

Alternative Hypothesis: The type of spray will affect the number of insects counted.

The output from summary(insectResults) returns an ANOVA table with the following columns:

Df: Degrees of freedom

Sum Sq: Sum of squares

Mean Sq: Mean square or the variance

F value: F-statistic - very high. we will reject the null hypothesis

Pr(>F): same as p-value. very small, much less than 0.05, we will reject the null hypothesis.

1. Load the BayesFactor package and run the anovaBF() command on the InsectSprays data set. You will have to specify the model correctly using the “~” character to separate the dependent variable from the independent variable. Produce posterior distributions with the posterior() command and display the resulting HDIs. Interpret the results briefly in your own words, including an interpretation of the BayesFactor produced by the grouping variable. As a matter of good practice, you should state the two hypotheses that are being compared. Using the rules of thumb offered by Kass and Raftery (1995), what is the strength of this result?

# install and library the BayesFactor package  
#install.packages("BayesFactor")  
library(BayesFactor)

## Loading required package: coda

## Loading required package: Matrix

## \*\*\*\*\*\*\*\*\*\*\*\*  
## Welcome to BayesFactor 0.9.12-4.7. If you have questions, please contact Richard Morey (richarddmorey@gmail.com).  
##   
## Type BFManual() to open the manual.  
## \*\*\*\*\*\*\*\*\*\*\*\*

# run the anovaBF() command on the InsectSprays data set  
bf <- anovaBF(count ~ spray, data = InsectSprays)  
bf

## Bayes factor analysis  
## --------------  
## [1] spray : 1.506706e+14 ±0%  
##   
## Against denominator:  
## Intercept only   
## ---  
## Bayes factor type: BFlinearModel, JZS

# Produce posterior distributions with the posterior() command  
posterior\_samples <- posterior(bf, iterations = 10000)  
  
# display the resulting HDIs  
library(coda)  
hdi\_values <- HPDinterval(as.mcmc(posterior\_samples))  
hdi\_values

## lower upper  
## mu 8.5593199 10.438916  
## spray-A 2.7664750 6.846414  
## spray-B 3.6580519 7.742681  
## spray-C -9.2554415 -5.129272  
## spray-D -6.4342115 -2.367867  
## spray-E -7.8969917 -3.778694  
## spray-F 4.7985299 8.948955  
## sig2 10.9509702 21.934738  
## g\_spray 0.4888246 9.107478  
## attr(,"Probability")  
## [1] 0.95

Null Hypothesis: The type of spray has no effect on the number of insects counted.

Alternative Hypothesis: The type of spray will affect the number of insects counted.

Bayes factor analysis of “spray : 1.506706e+14 ±0%” is very large, so we will reject the null hypothesis

The denominator refers to the “Intercept only” model, the model with only the intercept, or mean, and no group effects, or no spray effect.

The Bayes Factor is comparing the likelihood of the data under the model that includes the spray variable against the likelihood of the data under the null model (intercept only).

Strength of Evidence:

the rules of thumb by Kass and Raftery (1995) say that a Bayes Factor > 100 is decisive evidence against the null hypothesis.

Here, the Bayes Factor much greater than 100, this means that the evidence against the null hypothesis is decisive.

In summary, the Bayes Factor of 1.5×10 14th indicates that the type of spray has a significant effect on insect counts, with extremely strong evidence supporting this conclusion.