

HW6_Videtti

#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?

#The dependent variable is the number of insects killed by the spray.
#The independent variable is the type of insecticide.
#Given there were 12 observations for each of the 6 insecticides, there should be $12 \times 6 = 72$ observations. We can confirm this by looking at the number of rows in the *InsectSprays* data set.

```
nrow(InsectSprays)
```

```
## [1] 72
```

#There are indeed 72 observations!

#2. 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.

```
summary(aov(count~spray,data = InsectSprays))
```

```
##              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 "Mean Sq" for spray corresponds to the between-groups variance, while the "Mean Sq" for Residuals corresponds to the within-groups variance. This is because the variance between the sample means is the "Mean Sq" for count and the variance among all the samples is the "Mean Sq" for Residuals. For further proof, note that the F ratio/F value is 34.7. We know that the F

ratio is the between-groups variance divided by the within-groups variance. We also know that $533.8 / 15.4$ is approximately 34.7 and $15.4 / 533.8$ is approximately 0.0288.

#3. 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 learned 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?

#As we discussed in question 2, the F-ratio is 34.7. For samples that come from the same population, we expect F-ratios to be close to 1 and mostly somewhere between 0 and 1, so an F-ratio of 34.7 sticks out like a sore thumb. For this reason, if we had all the information needed for a NHST, I would likely reject the null hypothesis, barring the p-value a.k.a. $\Pr(>F)$.

#4. 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 between-groups variance is calculated using 6 groups. This means that there are 5 degrees of freedom between groups ($6 - 1 = 5$).

#The within-groups variance is calculated using 72 observations. We lose one degree of freedom from calculating the mean between all observations in order to calculate the variance, and we lose another 5 degrees of freedom due to calculating the between-groups variance as discussed above. This means that there are 66 degrees of freedom within groups ($72 - 1 - 5 = 66$).

#The sum of the degrees of freedom between groups and the degrees of freedom within groups is 71 ($66 + 5 = 71$). We see that this is one less than the

total number of observations (72). This is due to the fact that we calculated the grand mean in order to calculate the within-groups variance, meaning that we lost a degree of freedom since only 5 of the 6 sample means are free to vary after we calculate the grand mean.

#5. 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
```

#For between groups, we have 5 degrees of freedom, a sum of squares of 2669, and a variance of 533.8. For within groups, we have 66 degrees of freedom, a sum of squares of 1015, and a variance of 15.4. This gives us a F-ratio of 34.7. These are all things we have discovered, discussed, and interpreted in previous questions.

#Something new we haven't yet discussed is the p-value or $Pr(>F)$. We have an extremely small p-value meaning that we have significant evidence to reject the null hypothesis.

#The null hypothesis is that the average number of insects killed was the same for every different type of insecticide.

#The alternative hypothesis is that the average number of insects killed was different for at least one type of insecticide.

#Thus, we have significant evidence to reject the null hypothesis and conclude that there is a difference in the average number of insects killed for at least one of the six different types of insecticide.

#6. 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?

```
library(BayesFactor)

## Loading required package: coda

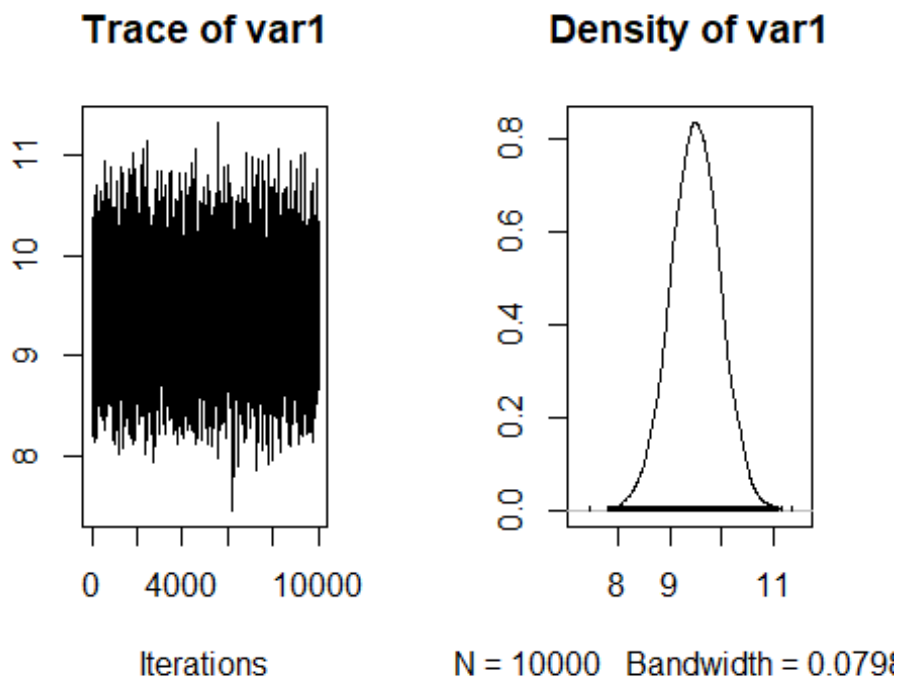
## Loading required package: Matrix

## *****
## Welcome to BayesFactor 0.9.12-4.3. If you have questions, please contact
## Richard Morey (richarddmorey@gmail.com).
##
## Type BFManual() to open the manual.
## *****

Exercise6 <- anovaBF(count~spray,data = InsectSprays)
Exercise6

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

plot(posterior(Exercise6,iterations = 10000)[,"mu"])
```



#H0: The average number of insects killed is the same for every different type of spray.

#H1: The average number of insects killed is not the same for every different type of spray.

#Here, we see that we have a Bayes Factor of approximately $1.5e+14$. This means that the odds favoring the alternative hypothesis are $1.5e+14$ to 1, meaning there are extremely high odds favoring the alternative hypothesis. Thus, we would reject the null hypothesis.

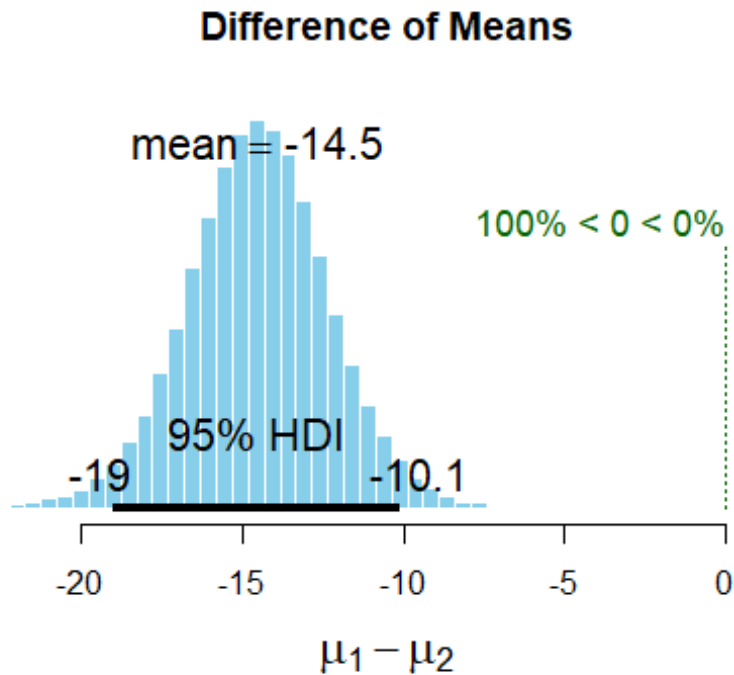
#That is, we have significant evidence to say that at least one of the sprays is more/less effective than the rest at killing insects.

#7. In situations where the alternative hypothesis for an ANOVA is supported and there are more than two groups, it is possible to do post-hoc testing to uncover which pairs of groups are substantially different from one another. Using the InsectSprays data, conduct a t-test to compare groups C and F (preferably a Bayesian t-test). Interpret the results of this t-test.

```
library(BEST)
```

```
## Loading required package: HDInterval
```

```
plot(BESTmcmc(InsectSprays$count[InsectSprays$spray ==  
'C'],InsectSprays$count[InsectSprays$spray == 'F']))  
## Waiting for parallel processing to complete...  
## done.
```



#We see here that our 95% HDI is -19.1 to -10.2. More importantly, this means that there is a 95% chance that the difference between the population means of insects killed from spray C and insects killed from spray F is negative. To elaborate, we can say we are 95% sure that spray F is more effective than spray C at killing insects.