

Homework 6

Daniel Caley

8/14/2021

Contents

Question 1	2
Understanding InsectSprays	2
What is The Outcome Variable	2
What is the total number of observations?	2
Question 2	3
Question 3	3
Calculate the F-Ratio	3
F-Ratios	3
Calculate the Degree of Freedom	3
Explain the Results	4
AOV Command	4
Question 6	5
BayesFactor Package	5
Produce Posterior Distributions	5
Interpret the results briefly in your own words	6
Question 7	7
Post-Hoc Testing	7
Interpret the results of this t-test.	8

The homework for week 6 is exercises 1-7 on pages 117 and 118.

Question 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.

Understanding InsectSprays

```
My_InsectSprays <- InsectSprays  
str(My_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 ...
```

```
summary(My_InsectSprays)
```

```
##      count      spray  
## Min.   : 0.00   A:12  
## 1st Qu.: 3.00   B:12  
## Median : 7.00   C:12  
## Mean    : 9.50   D:12  
## 3rd Qu.:14.25   E:12  
## Max.    :26.00   F:12
```

```
head(My_InsectSprays, 5)
```

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

What is The Outcome Variable

In this experiment, what is the dependent variable (outcome) and what is the independent variable?

- The dependent variable is the count column or the count of insects killed by the spray.
- The independent variable is the spray.

What is the total number of observations?

```
length(My_InsectSprays$count)
```

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

Question 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.

- Between group variance is the Mean Sq for spray of 533.8. The between groups shows how much of the variation in the means between the groups being tested.
- Within group variance is the Mean Sq for residuals is 15.4. The within group variance shows how much of the variation is due to mixing up all of the data.

Question 3

Calculate the F-Ratio

Based on the information in question 2 and your response to that question, calculate an F-ratio by hand or using a calculator.

```
MeanSq_Spray <- 533.8
MeanSq_Residual <- 15.4

FRation <- MeanSq_Residual / MeanSq_Spray
FRation
```

```
## [1] 0.02884976
```

F-Ratios

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?

This is different from what we learned previously where we would state an alpha and then if our P-value was within that percentage we would either reject or fail to reject the Null Hypothesis. In this case when the F is statistically significant the results say nothing about which means are different. We would reject the Null Hypothesis, but no nothing about which means are different among the insect sprays.

Calculate the Degree of Freedom

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.

```
TotalObs <- length(My_InsectSprays$spray)
TotalSpray <- length(unique(My_InsectSprays$spray))

WthinDF <- TotalObs - TotalSpray
WthinDF
```

```
## [1] 66
```

```
BtwnDF <- TotalObs - WthinDF - 1
```

```
BtwnDF
```

```
## [1] 5
```

- Within Groups DF is 66
- Between Groups DF is 5

Explain the Results

Explain why the sum of these two values adds up to one less than the total number of observations in the data set.

- When looking at the data set of 72 observation the degree of freedom is always 1 minus the total when calculating for the grand mean.
- Only 5 out of the 6 groups of means can vary freely which leaves 66 within groups.

AOV Command

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=My_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
```

- The Null Hypothesis is that the sprays are not different
- Alternative Hypothesis is that the sprays are different from each other
- Results of the null hypothesis significance test leads me to conclude that due to the P value being less than 0.05. These results do not tell me which groups are different from each other.

Question 6

BayesFactor Package

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.

```
insectResultsBF <- anovaBF(count ~ spray, data=My_InsectSprays)
insectResultsBF

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

Produce Posterior Distributions

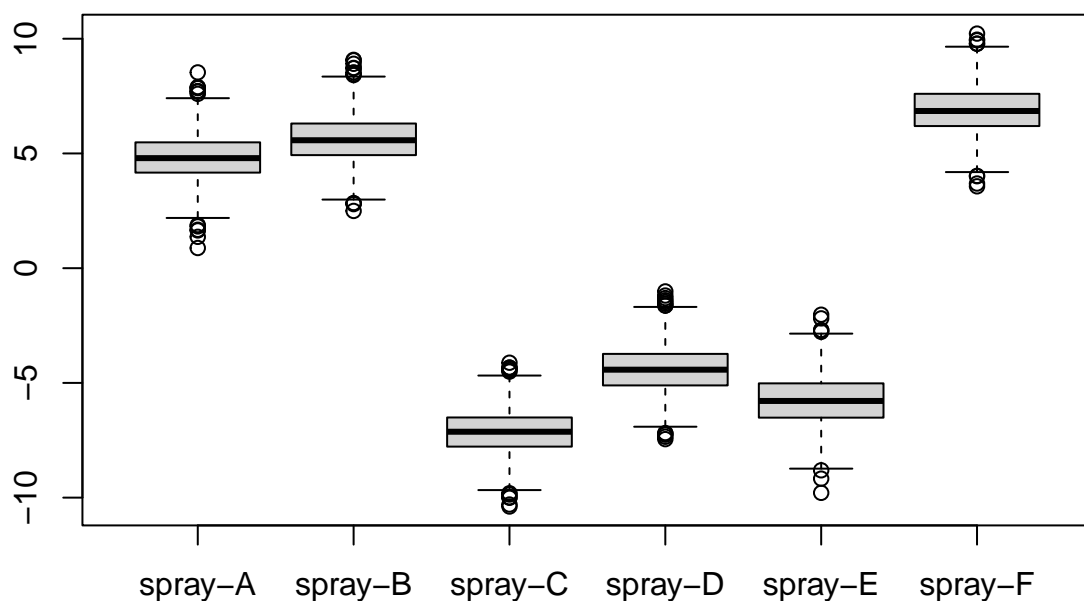
Produce posterior distributions with the `posterior()` command and display the resulting HDIs.

```
insectResultsBF <- posterior(insectResultsBF, iterations = 1000)
summary(insectResultsBF)

##
## Iterations = 1:1000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##           Mean      SD Naive SE Time-series SE
## mu          9.474 0.4746 0.01501      0.01501
## spray-A     4.827 1.0261 0.03245      0.03245
## spray-B     5.630 1.0212 0.03229      0.03502
## spray-C    -7.140 1.0196 0.03224      0.03224
## spray-D    -4.421 1.0413 0.03293      0.03293
## spray-E    -5.782 1.0872 0.03438      0.02981
## spray-F     6.885 1.0680 0.03377      0.03377
## sig2       16.003 2.8897 0.09138      0.10258
## g_spray     3.569 4.3911 0.13886      0.13886
##
## 2. Quantiles for each variable:
##
##           2.5%    25%    50%    75%   97.5%
## mu          8.5563  9.172  9.484  9.783 10.374
## spray-A     2.8728  4.166  4.796  5.484  6.938
```

```
## spray-B 3.7175 4.929 5.576 6.304 7.682
## spray-C -9.1825 -7.776 -7.129 -6.507 -5.122
## spray-D -6.4379 -5.109 -4.424 -3.736 -2.284
## spray-E -7.9903 -6.514 -5.787 -5.021 -3.632
## spray-F 4.8047 6.195 6.850 7.599 9.003
## sig2 11.6110 13.920 15.612 17.595 22.714
## g_spray 0.8386 1.640 2.478 4.208 11.023
```

```
boxplot(as.matrix(insectResultsBF[,2:7]))
```



Interpret the results briefly in your own words

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?

- Null Hypothesis is that all the groups are the same.
- Alternative Hypothesis is that the groups are different.
- The strength of the results is that 15067060000000000000:1 in favor of an effect for insect spray is pretty strong and therefore in favor of the alternative hypothesis. Kass and Raftery give a rule of thumb of an odds ratios of more than 150:1 are very strong evidence for the favored hypothesis. When looking at the boxplot Spray A, B, and F are all very similar to each other with C, D and E also being very similar to each other. With that being said A, B, and F are superior sprays with an estimate of 4.8, 5.8, and 6.8 respective increase over the overall mean of 9.56. That is with no overlap and therefore superior to all of the other sprays.

Question 7

Post-Hoc Testing

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).

```
SprayC <- InsectSprays$count[InsectSprays$spray == "C"]
SprayF <- InsectSprays$count[InsectSprays$spray == "F"]
```

```
set.seed(3)
```

```
BaysianSpray <- BESTmcmc(SprayC, SprayF)
```

```
## Waiting for parallel processing to complete...done.
```

```
BaysianSpray
```

```
## MCMC fit results for BEST analysis:
```

```
## 100002 simulations saved.
```

```
##           mean      sd median  HDIlo  HDIup  Rhat n.eff
## mu1       1.972  0.6427  1.962  0.7083  3.242 1.000 52898
## mu2      16.495  2.1355 16.486 12.2473 20.730 1.000 56626
## nu       32.341 29.2438 23.484  1.0411 89.655 1.002 18746
## sigma1    2.051  0.5794  1.964  1.0855  3.249 1.000 27037
## sigma2    6.877  1.8414  6.565  3.9342 10.661 1.000 30781
```

```
##
```

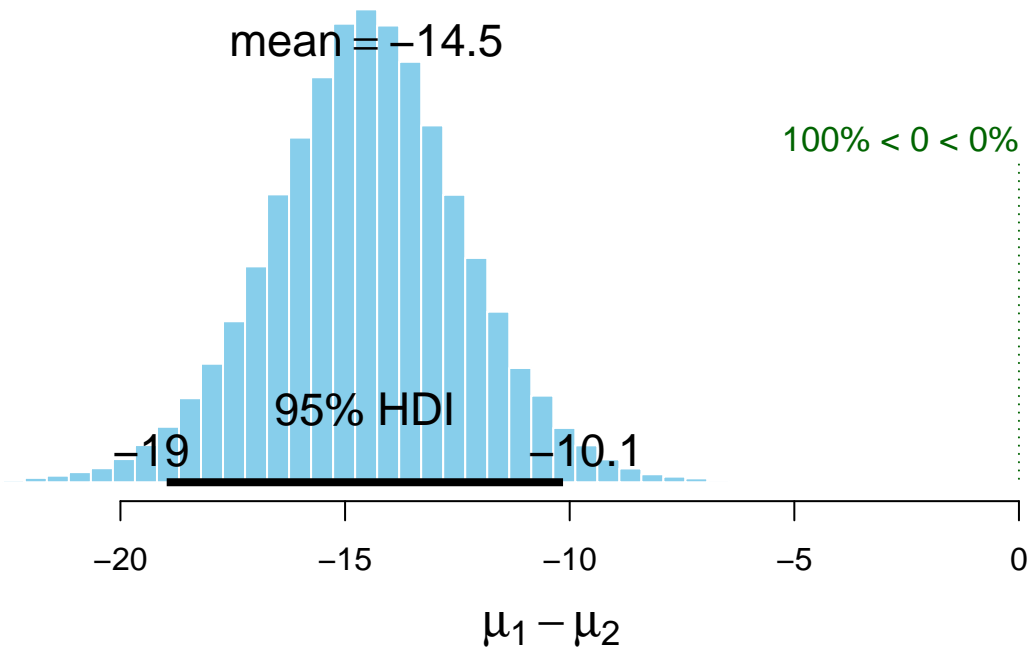
```
## 'HDIlo' and 'HDIup' are the limits of a 95% HDI credible interval.
```

```
## 'Rhat' is the potential scale reduction factor (at convergence, Rhat=1).
```

```
## 'n.eff' is a crude measure of effective sample size.
```

```
plot(BaysianSpray)
```

Difference of Means



Interpret the results of this t-test.

- The 95% HDI goes from -19 to -10.2 without crossing 0. Meaning that the mean difference between group C and F have a 95% change of falling between the ranges.
- The results show that most likely the mean difference falls at -14.5.
- Due to the results falling below 0 there is a clear difference between the 2 groups with F having a higher count of insect than C. The objective of the insect spray products was to kill more insects, spray F is therefore the clear winner due to having more dead insects.