

## Homework 6

Kwasi Mensah

2022-11-20

Author: Kwasi Mensah Homework\_Number: 6 Output: pdf\_document Attribution statement:

I did the homework by myself, with help from the book and the professor #R Markdown  
#Run these three functions to get a clean test of homework code

```
dev.off() #Clear the graph window  
  
## null device  
##          1  
  
cat('\014') #Clear the console  
  
rm(list = ls()) #Clear user objects from the environment
```

#R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#####  
#####
```

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(rjags)  
  
## Warning: package 'rjags' was built under R version 4.2.2  
  
## Loading required package: coda  
  
## Linked to JAGS 4.3.1  
  
## Loaded modules: basemod,bugs
```

```

data(InsectSprays)
?InsectSprays

## starting httpd help server ...

## done

summary(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

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 ...

table(InsectSprays)

##      spray
## count A B C D E F
##    0  0 0 2 0 0 0
##    1  0 0 4 0 2 0
##    2  0 0 2 1 1 0
##    3  0 0 2 2 4 0
##    4  0 0 1 2 1 0
##    5  0 0 0 5 2 0
##    6  0 0 0 1 2 0
##    7  1 1 1 0 0 0
##    9  0 0 0 0 0 1
##   10  2 0 0 0 0 1
##   11  0 2 0 0 0 1
##   12  1 0 0 1 0 0
##   13  1 1 0 0 0 2
##   14  3 1 0 0 0 0
##   15  0 0 0 0 0 2
##   16  0 1 0 0 0 1
##   17  1 3 0 0 0 0
##   19  0 1 0 0 0 0
##   20  2 0 0 0 0 0
##   21  0 2 0 0 0 0
##   22  0 0 0 0 0 1
##   23  1 0 0 0 0 0
##   24  0 0 0 0 0 1
##   26  0 0 0 0 0 2

```

*# The dependent variable in the data set is "count" (Insect count) and the independent variable "spray" type of spray. There are a total of 72 observations and 2 variables.*

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.

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

```
##              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 first line or "spray" group is the variance "between groups", which is Mean Sq of 533.8. This is calculated based on the spread of means. The second line or "Residuals" group is the variance "Within groups", which is Mean Sq of 15.4. This is calculated based on the variance in the raw data. We will reject the null hypothesis due to the probability of a larger F-ratio, 2e-16, being less than .05 and statistically significant.*

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 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?

[533.8/15.4](#)

```
## [1] 34.66234
```

*# The F-ratio is calculated by dividing the mean squares from the two groups in our ANOVA results. we would divide 533.8/15.4 = 34.66234. Our F-ratio is 34.66234. The F-ratio should be close to or greater than 1. The f-ratio is greater than 1 so it is statistically significant. We will reject the null hypothesis. The Pr value is less than .05 which also helps us to conclude that it is significant.*

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.

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

```
##              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 degrees of freedom for the between groups is 5 and the degrees of freedom for the within groups is 66. The degree of freedom indicates how many elements of the data set are free to vary once an initial set of statistics have been calculated. When we calculate the degree of freedom from a data set of 72 observations we lose one degree of freedom for calculating the grand mean, thus having a total degree of freedom of 71. If we have k groups the degree of freedom will be (k-1).*

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
```

*# The data set is about the counts of insects in agricultural experimental units treated with different insecticides. There are 6 different types of sprays which is the independent variable. The data set has 12 groups of insects with each group having a different count of insects, this is the dependent variable.*

*# We assume that there are no difference between the sprays for our initial hypothesis. For our alternative hypothesis we can assume that there is in fact difference between the insect sprays. When we run the ANOVA test we can see differences between the mean square. The total F-value is 71 by adding the df of the between group and within group, which is higher than 1, thus we can conclude our analysis is significant and we could reject the null hypothesis. The  $Pr(>F)$  value being  $2e-16$ , which is less than  $.05$ , we will reject the null hypothesis. This will support our alternative hypothesis.*

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)

## Warning: package 'BayesFactor' was built under R version 4.2.2

## Loading required package: Matrix

## *****
## Welcome to BayesFactor 0.9.12-4.4. If you have questions, please contact
## Richard Morey (richarddmorey@gmail.com).
##
## Type BFManual() to open the manual.
## *****

library(BEST)

## Warning: package 'BEST' was built under R version 4.2.2

## Loading required package: HDInterval

## Warning: package 'HDInterval' was built under R version 4.2.2

# Bayes Factor for ANOVA
spraysBayesOut <- anovaBF(count ~ spray, data=InsectSprays)
spraysBayesOut

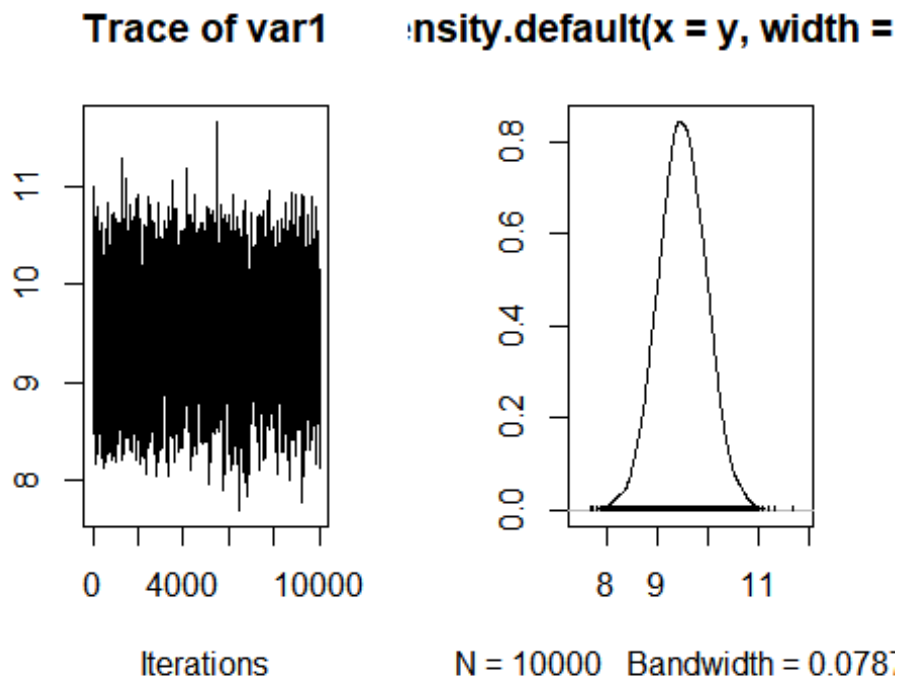
## Bayes factor analysis
## -----
## [1] spray : 1.506706e+14 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS

# Produce posterior distribution
spraysMCMCOut <- posterior(spraysBayesOut, iterations=10000)
summary(spraysMCMCOut)

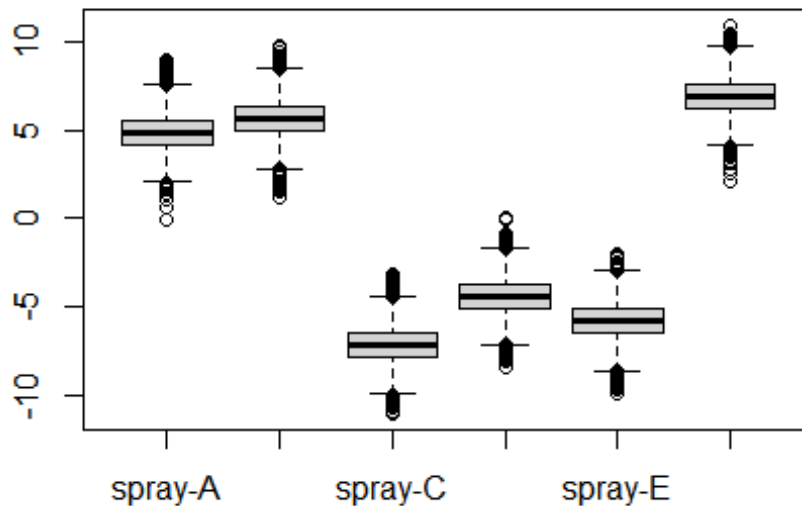
##
## Iterations = 1:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##           Mean      SD Naive SE Time-series SE
## mu          9.504 0.4691 0.004691      0.004691
## spray-A     4.822 1.0403 0.010403      0.010562
## spray-B     5.610 1.0477 0.010477      0.010401
## spray-C    -7.133 1.0449 0.010449      0.010854

```

```
## spray-D -4.424 1.0412 0.010412      0.010590
## spray-E -5.780 1.0577 0.010577      0.010751
## spray-F  6.905 1.0381 0.010381      0.010763
## sig2     16.153 2.9287 0.029287      0.034482
## g_spray  3.424 3.2554 0.032554      0.034772
##
## 2. Quantiles for each variable:
##
##           2.5%   25%   50%   75%  97.5%
## mu          8.5871  9.192  9.500  9.820 10.419
## spray-A     2.7969  4.122  4.825  5.510  6.847
## spray-B     3.5374  4.904  5.618  6.317  7.641
## spray-C    -9.1680 -7.846 -7.132 -6.453 -5.040
## spray-D    -6.4883 -5.107 -4.434 -3.727 -2.382
## spray-E    -7.8744 -6.489 -5.780 -5.058 -3.717
## spray-F     4.8534  6.197  6.908  7.606  8.919
## sig2       11.4225 14.044 15.798 17.841 22.845
## g_spray     0.8165  1.671  2.528  4.037 11.429
plot(spraysMCMCOut[, "mu"], main=NULL)
```



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



# The Lower bound is 8.5602 and the upper bound is 10.436. The output yields a Bayes Factor of  $1.506706e+14 \pm 0:1$  which is well over 150:1. According to the rules by Kass and Raftery (1995), any odds ratio in excess of 150:1 is considered very strong evidence. This confirms support for an alternative hypothesis of credible differences among these group means, as a result we will reject the null hypothesis. The box plot shows posterior distributions for deviations of each group from the grand mean. The grand mean is represented on this figure as 0 on the y-axis. There is separation above and below zero on the graph and none of the box plots overlap which tells us there are differences among each spray. There are no extreme outliers in the trace plot and density histogram, and the two show the the mean is between 9 & 10.

# Given a null hypothesis of there being no difference between the insect sprays and an alternative hypothesis of there being a difference, the result suggest we will reject the null hypothesis as there is a difference among the insect sprays. This separation is shown in the empirical means and quantiles between spray A,B,F ,being positive, vs. C,D,E, being negative.

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.

```
observationC_F <- InsectSprays[InsectSprays$spray=="C" |
InsectSprays$spray=="F",]
Cspray <- InsectSprays[InsectSprays$spray=="C",1]
```

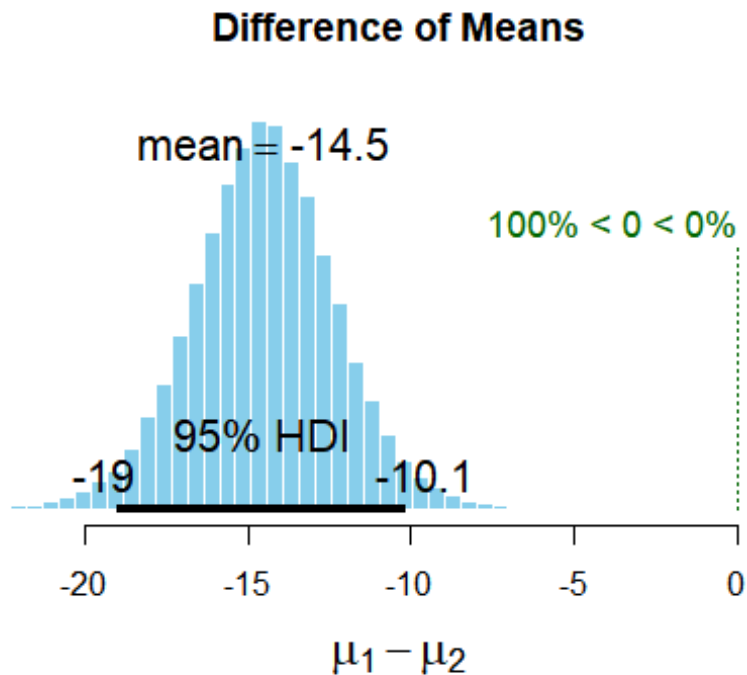
```

Fspray <- InsectSprays[InsectSprays$spray=="F",1]
observations <- data.frame("C" = Cspray, ".F" = Fspray)

# T-test
plot(BESTmcmc(InsectSprays[InsectSprays$spray=="C",1],
                  InsectSprays[InsectSprays$spray=="F",1]))

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 24
##   Unobserved stochastic nodes: 5
##   Total graph size: 67
##
## Initializing model
##
##
## Sampling from the posterior distributions:

```



```

sprayB_F <- anovaBF(count ~ spray, data=observationC_F)
summary(sprayB_F)

## Bayes factor analysis
## -----
## [1] spray : 90005.78 ±0%
##

```



```
## Against denominator:
```

```
##   Intercept only
```

```
## ---
```

```
## Bayes factor type: BFlinearModel, JZS
```

*# The entire distribution is below 0. There is a 95% chance of the mean difference lies between -19 and -10.1, and since these intervals do not pass 0 our analysis is statistically significant. The population mean difference of is most likely -14.5. The F insect spray is more effective than the C insect spray.*