

Homework 6

Kyle Walter

5/20/2021

Student Name: Kyle Walter

Homework Number: 6

Due Date: 05/23/2021

Attribution statement: I did this homework by myself, with help from the book and the professor.

1.

```
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 ...
```

```
dim(InsectSprays)
```

```
## [1] 72  2
```

```
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
```

As we can see from the summary there are 6 separate sprays which represent the independent variables and counts of 12 tests each with the number representing the number of dead insect observed, the dependent variable, collected.

Problem 2. Perform ANOVA on insect sprays

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

##              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
```

Next I ran an anova test to look at the variance among all the groups. 533.8 represents the mean variance between groups, while 15.4 represent the mean variance observed within the groups.

3

Also seen in the output from the anova in problem 2 is the F statistic. It is calculated by taking the mean square between the group and dividing it by the mean square within the groups. $533.8/15.4 = 34.6623377$. The F Value is quite large, usually values drop off substantially from 1 and by about 4 become very rare. Additionally next to the F value we have the Probability of F which is $2e-16$ or virtually zero. Given an alpha of .05 with this value being less, we would reject the null hypothesis, and say there is statistical evidence point to the means of bugs killed by due to spray are not the same.

4

Calculating the degrees of Freedom works as follows. Starting with 72 for the number of observations in the spray data. We remove 1 from this, as a single observation is cannot be compared with anything. Then to figure the degrees of freedom within groups, we look at the independent variables, of which we have 6 different sprays. Similar to what was done with observations we remove 1 to get the degrees of freedom between groups as 5. Lastly to get the degrees of Freedom within groups we remove 5 from total degrees of Freedom 71 to get 66

Problem 5

```
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
```

Similar to experiment results in problem 2, I ran the anova on the insect sprays data by the differnt sprays. The Null hypothesis would be that we expect to observer no difference in the average number of dead insects among the different sparys.

The Alternative hypothesis would be that we would see support to show that the spray makes a difference in how many insects are killed.

The anova results show an F significance of 34.7 which is well above an F value of 1. This is further supported by the fact the P value is less than 2e-16 and smaller than the .05 that is used as a test threshold. This allows us to conclude based on the evidence that there is statistical support for the alternative hypothesis. We thus reject the null hypthosis.

6. Perform bayes factor analysis on the insect sprays dataset

```
library("BayesFactor")
```

```
## Warning: package 'BayesFactor' was built under R version 4.0.5
```

```
## Loading required package: coda
```

```
## Loading required package: Matrix
```

```
## *****
```

```
## Welcome to BayesFactor 0.9.12-4.2. 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 4.0.5
## Loading required package: HDInterval
## Warning: package 'HDInterval' was built under R version 4.0.5

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
```

The Bays Factor Anova shows us a large Bayes Factor, with a number above 500 showing significant support of the Alternative Hypothesis, echoing what has been seen previously in the frequentist anova tests.

Generate posterior probabilities

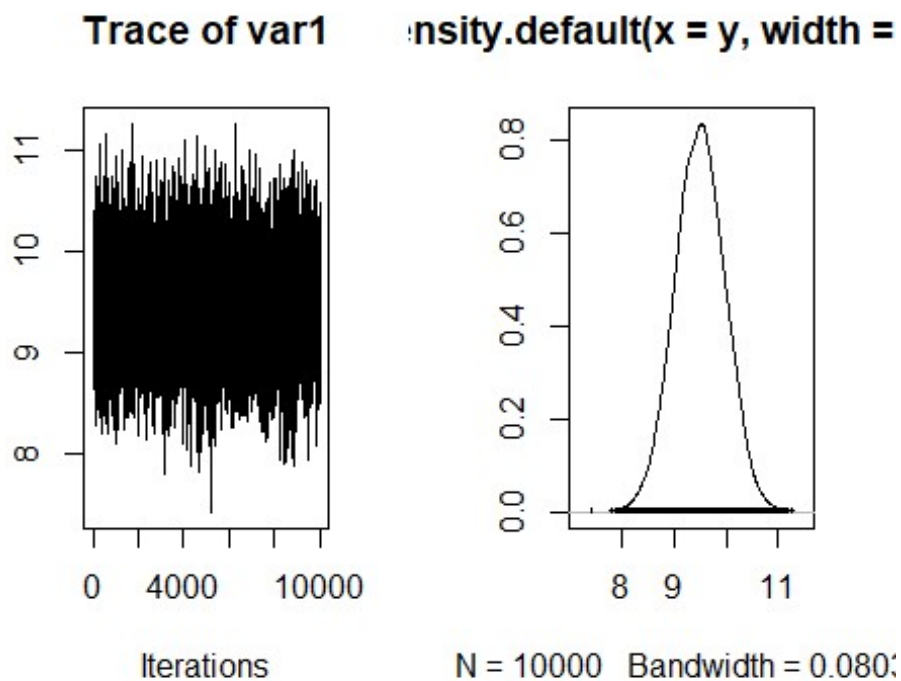
```
spraysMCMCOut <- posterior(spraysBayesOut, iterations=10000)
summary(spraysMCMCOut)

##
## 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          9.502 0.4785 0.004785      0.004785
## spray-A     4.817 1.0515 0.010515      0.011113
## spray-B     5.625 1.0451 0.010451      0.010451
## spray-C    -7.152 1.0627 0.010627      0.011059
## spray-D    -4.416 1.0356 0.010356      0.010468
## spray-E    -5.772 1.0537 0.010537      0.010537
## spray-F     6.898 1.0517 0.010517      0.011134
## sig2       16.134 2.9206 0.029206      0.034524
## g_spray     3.486 3.4884 0.034884      0.037689
##
## 2. Quantiles for each variable:
##
##           2.5%    25%    50%    75%   97.5%
## mu          8.5696  9.181  9.504  9.823 10.432
## spray-A     2.7264  4.116  4.807  5.528  6.848
```

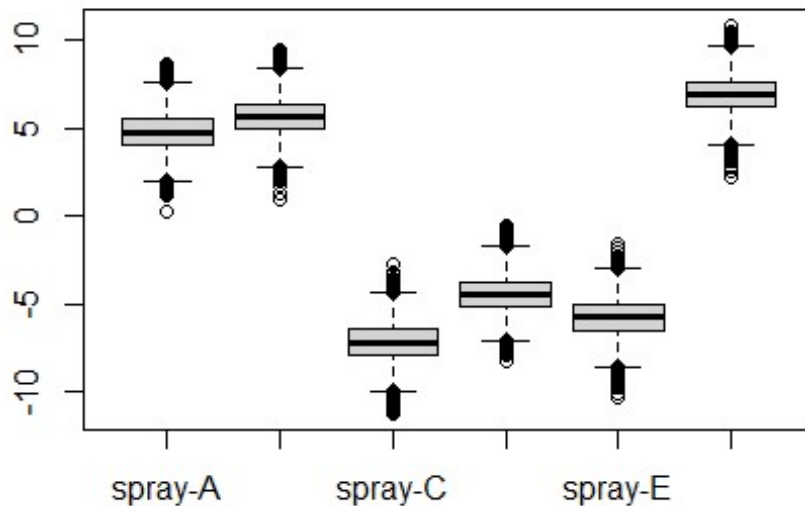
```
## spray-B 3.5883 4.935 5.626 6.328 7.701
## spray-C -9.1951 -7.860 -7.176 -6.448 -5.071
## spray-D -6.4515 -5.105 -4.418 -3.746 -2.328
## spray-E -7.8403 -6.471 -5.771 -5.070 -3.696
## spray-F 4.8043 6.202 6.902 7.601 8.941
## sig2 11.3615 14.041 15.813 17.857 22.723
## g_spray 0.8083 1.679 2.577 4.077 11.656
```

In addition the monte carolo mean and standard deviation shows us the means of each spay and it's standard deviation. We can see that A, B, and F have significantly larger impact while C,D, and E all have smaller impact. None of them however pass through the zero. Continuing to support the alternative hypothesis.

```
plot(spraysMCMCOut[, "mu"], main=NULL)
```



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

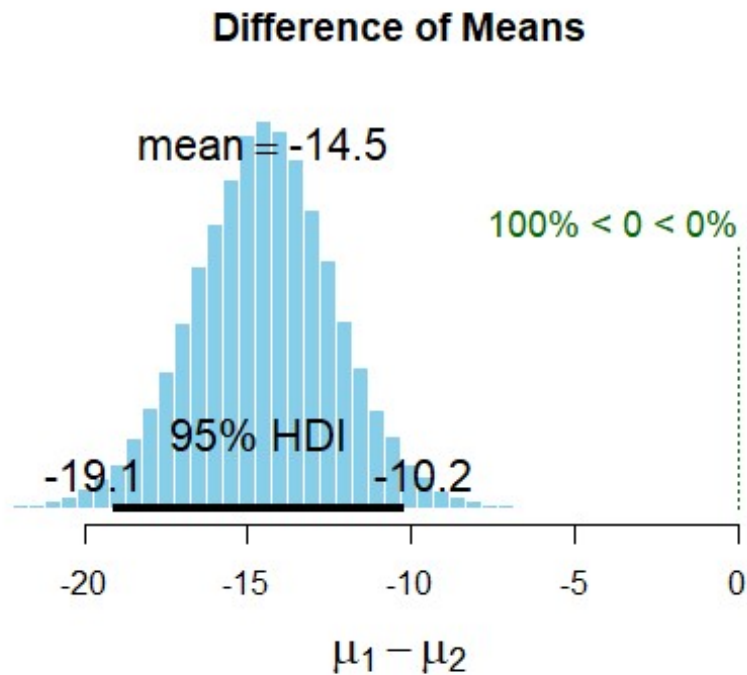


Lastly, in a more visual view, we can see the HDI shows a range for the entire group between 8 and 10.5 not passing through zero. And the box plots show all six sprays mean and deviation with none pass through zero all supporting the rejection of the null hypothesis.

Problem 7. Perform a Bayesian t-test between spray-C and spray-F

```
CandF_obs <- InsectSprays[InsectSprays$spray=="C" | InsectSprays$spray=="F",]
sprayC_obs <- InsectSprays[InsectSprays$spray=="C",1]
sprayF_obs <- InsectSprays[InsectSprays$spray=="F",1]
obs <- data.frame("C" = sprayC_obs, ".F" = sprayF_obs)
# Perform a Bayes t-test against just the C and F sprays.
plot(BESTmcmc(InsectSprays[InsectSprays$spray=="C",1],
               InsectSprays[InsectSprays$spray=="F",1]))

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



```
spray_BFOut <- anovaBF(count ~ spray, data=CandF_obs)
summary(spray_BFOut)

## Bayes factor analysis
## -----
## [1] spray : 90005.78 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

Comparing the mean of spray C to spray F using the MCMC HDI we see that the model predicts that this a mean variance. The HDI ranges from -18.9 to -10.1 with the model suggesting a population mean difference around -14.5. None of the data passes through the null hypothesis where the mean variance would be 0, so based on the observation there is statistical support for rejecting the null hypothesis between Sprays C and F