IST772 Problem Set 6

Shivani Sanjay Mahaddalkar

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

Chapter 6, Exercise 1

The datasets package (installed in R by default) contains a dataset called warpbreaks (see "? warpbreaks" for documentation) that shows the number of warp thread breaks per loom for different tensions (again, we will not consider the type of wool). For each tension, n = 18 observations were conducted. In this experiment, what is the dependent variable (outcome) and what is the independent variable? What is the total number of observations? (1 pt) For the experiment, the dependent variable is the number of warp thread breaks per loom and the independent variable is the type of tension in these threads(namely L, M or H). The total number of observations are 54 which are divided into three groups of 18 each.

Chapter 6, Exercise 2

After running the aov() procedure on the warpbreaks data set, the "Mean Sq" for tension is 1017.1 and the "Mean Sq" for Residuals is 141.1 (1 pt). 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. (1 pt)

```
aov(breaks~tension, data = warpbreaks)

## Call:
##    aov(formula = breaks ~ tension, data = warpbreaks)
##
## Terms:
##                  tension Residuals
## Sum of Squares  2034.259  7198.556
## Deg. of Freedom        2        51
##
## Residual standard error: 11.88058
## Estimated effects may be unbalanced
```

The sum Sq for tension is 1017.1 which is the variance between groups. It is the sum of variance between each group mean and the overall mean. The Sum Sq for residuals is 141.1 which is the variance within-groups. It is the sum of variance of the breaks for L, M and H tensions.

Chapter 6, Exercise 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? (1 pt) The F-ratio would be (1017.1/141.1) which is 7.213. Since the F-ratio is much greater than 1, it can be said that there is sufficient proof to reject the Nul Hypothesis given the F-ratio is significant.

Chapter 6, Exercise 4

Continuing with the warpbreaks example, there are three groups where each one has n = 18 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. (1 pt) Starting with 54 observations, we lose one degree of freedom to calculating the overall mean of the data set. When calculating the between-groups variance, we borrow 2 degrees of freedom for three different groups which leaves us with 51 degrees of freedom as shown in the analysis.

Chapter 6, Exercise 5

Use R or R-Studio to run the aov() command on the warpbreaks 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 warpbreakResults. Run the summary() command on warpbreakResults 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. (1 pt) (Here's a summary of the APA recommended format for reporting statistical tests
http://www2.psych.utoronto.ca/users/reingold/courses/resources/handouts_apa/Stats1.pdf.
)

```
warpbreakResults <- aov(breaks~tension, data = warpbreaks)
summary(warpbreakResults)

##              Df Sum Sq Mean Sq F value  Pr(>F)
## tension       2   2034  1017.1   7.206 0.00175 **
## Residuals    51   7199   141.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reporting a significant F test for ANOVA: Null Hypothesis: All three groups were sampled from the same population i.e. the difference between the population means of the groups is zero. Alternate Hypothesis: Atleast one group has different population mean than the others. For a significant F test(Pr = 0.00175 < 0.05), with F value of 7.206(>1), we have

strong evidence to reject the null hypothesis. This indicates that the three groups have different population means and were not sampled from the same population.

Chapter 6, Exercise 6

Load the BayesFactor package and run the anovaBF() command on the warpbreaks 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. (1 pt) Interpret the results briefly in your own words, including an interpretation of the BayesFactor produced by the grouping variable. (1 pt) 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? (1 pt)

```r
library(BayesFactor)

## Warning: package 'BayesFactor' was built under R version 4.0.4

## 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.
## ************

warpbreakResultsBayes <- anovaBF(breaks~tension, data = warpbreaks)
mcmcOut <- posterior(warpbreakResultsBayes, iterations = 10000)
boxplot(as.matrix(mcmcOut[,2:4])) # Boxplot the posteriors for the groups
```
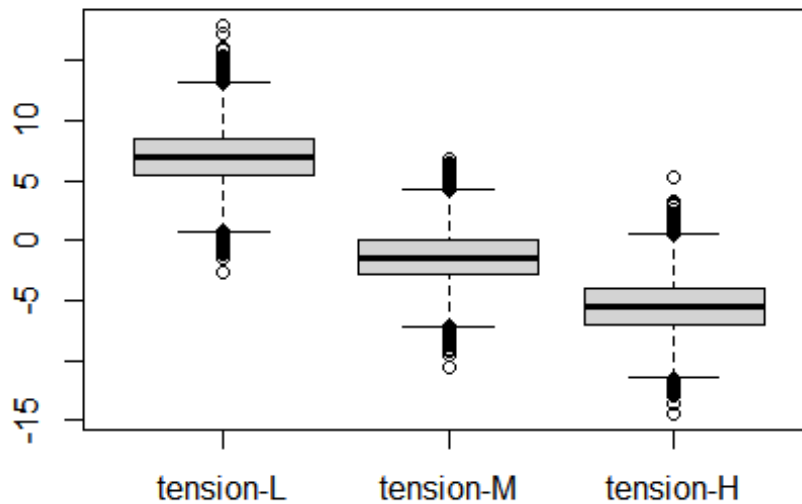
```
summary(mcmcOut)

##
## 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          28.1346   1.650  0.01650        0.01594
## tension-L    6.9273   2.332  0.02332        0.02918
## tension-M   -1.4652   2.136  0.02136        0.02243
## tension-H   -5.4622   2.252  0.02252        0.02552
## sig2       147.3120 30.463  0.30463        0.31997
## g_tension    0.8627   4.396  0.04396        0.04396
##
## 2. Quantiles for each variable:
##
##                 2.5%      25%      50%        75%    97.5%
## mu          24.84621  27.0492  28.1316  29.231092   31.371
## tension-L    2.33454   5.3827   6.9444   8.471881   11.465
## tension-M   -5.70002  -2.8803  -1.4633  -0.006711    2.720
## tension-H   -9.89193  -6.9530  -5.4585  -3.952673   -1.107
```

```
## sig2      99.71148 125.8212 143.1756 164.646029 218.395
## g_tension  0.06913   0.1947   0.3717   0.744425   4.205
```

```
warpbreakResultsBayes
```

```
## Bayes factor analysis
## --------------
## [1] tension : 21.451 ±0.01%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

Null Hypothesis: All three groups were sampled from the same population i.e. the difference between the population means of the groups is zero. Alternate Hypothesis: Atleast one group has different population mean than the others. This analysis shows odds of 21.45:1 in favour of the altenative hypothesis. According to the rule of thumb provided by Kass and Raftery, odds exceeding 150:1 is considered strong evidence. The Bayesian analysis gives us weak evidence in support of the alternate hypothesis which contradicts the anova t-test.

Chapter 6, Exercise 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 warpbreaks data, conduct a t-test to compare groups "M" and "L" (preferably a Bayesian t-test). Interpret the results of this t-test. (1 pt) In addition to the Bayesian t-test, feel free to run Tukey's HSD or another post-hoc procedure.

```r
library(BEST)
```

```
## Warning: package 'BEST' was built under R version 4.0.4
```

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

```
## Warning: package 'HDInterval' was built under R version 4.0.4
```

```r
BESTmcmc(warpbreaks$breaks[warpbreaks$tension=='M'],
warpbreaks$breaks[warpbreaks$tension=='L'])
```
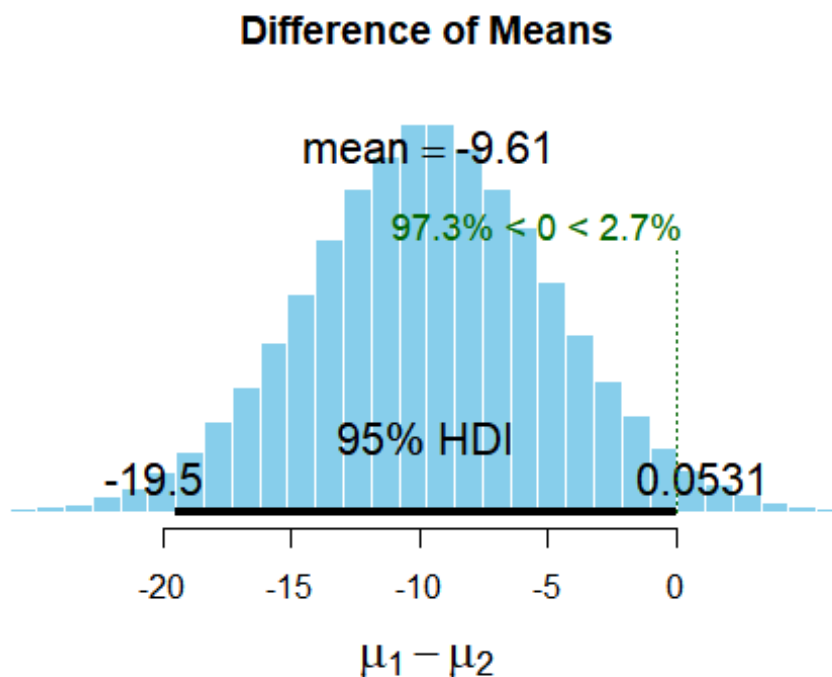
```
## Waiting for parallel processing to complete...
```

```
## done.
```

```
## MCMC fit results for BEST analysis:
## 100002 simulations saved.
##          mean    sd median  HDIlo  HDIup  Rhat  n.eff
## mu1     26.312  2.416 26.320 21.639  31.18 1.000 60194
```

```
## mu2     35.904  4.306 35.902 27.191  44.27 1.000 59966
## nu      38.981 31.061 30.437  2.397 100.41 1.001 20916
## sigma1  9.704  1.940  9.431  6.363  13.56 1.001 38923
## sigma2 17.264  3.507 16.785 11.185  24.36 1.000 37797
##
## '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(BESTmcmc(warpbreaks$breaks[warpbreaks$tension=='M'],
warpbreaks$breaks[warpbreaks$tension=='L']))
```

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


Difference of Means

mean = -9.61

97.3% < 0 < 2.7%

95% HDI

-19.5   0.0531

-20   -15   -10   -5   0

$\mu_1 - \mu_2$

26.305 is the point estimate for the number of breaks for the group M level of tension, and 95% of estimates of the population mean for the M level of tension fell between 21.58 and 31.03. 35.931 is the point estimate for the number of breaks for the group L level of tension, and 95% of estimates of the population mean for the M level of tension fell between 27.31 and 44.40. With the analysis, we might say that the population difference between the two groups is somewhere near-9.58 breaks, with the 95% highest density interval ranging from -19.6 to -0.12. Since the HDI does not involve 0, we have strong evidence that the groups have different population means. This confirms the ANOVA t-test we conducted in the first instance that atleast one group has a different population mean than the others.