IST772 Problem Set 7

Abhijith Anil Vamadev

The homework for week 8 is based on exercises 3, 4, 8, 9, 10 on pages 155-156 but with changes as noted in this notebook (i.e., follow the problems as given in this document and not the textbook).

Attribution statement: (choose only one) 1. I did this homework by myself, with help from the book and the professor

Chapter 7, Exercise 3

Run cor.test() on the correlation between "mpg" and "hp" in the mtcars data set (type "? cars" to see the documentation) and interpret the results. (1 pt) Make sure that you interpret both the confidence interval and the p-value that is generated by cor.test(). (1 pt)

```
cars <- mtcars #assigning variable
cor.test(mtcars$mpg, mtcars$hp) #correlation test
```

```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$hp
## t = -6.7424, df = 30, p-value = 1.788e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.8852686 -0.5860994
## sample estimates:
##        cor
## -0.7761684
```

- Pearson's product moment correlation, yielded a t(30) = -6.7, with a p-value of 1.78e-07. The p-value is less than the alpha, so we reject the null hypothesis, that rho is 0, showing that the test-is statistically signficant.
- The 95% confidence interval around the point estimate of r = -0.77. If we repeated this sampling process many times and each time constructed a confidence interval around the calculated value of r, about 95% of those constructed intervals would contain the true population value, rho.
- The 95% confidence interval for rho ranged from -0.885 and -0.58. Since the confidence interval, does not contain 0, so we have a sense of certainty that the correlation is negative.

Chapter 7, Exercise 4

Below is a copy of the bfCorTest() custom function presented in this chapter; you can instead use the correlationBF function from the BayesFactor library. Conduct a Bayesian analysis of the correlation between "mpg" and "hp" in the mtcars data set. (1 pt) Report the results. (1 pt)

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

```
## 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
##
## Type BFManual() to open the manual.
## ************
```

```
bfCorTest <- function (x,y) # Get r from BayesFactor
{
  zx <- scale(x)    # standardize X
  zy <- scale(y)    # standardize Y
  zData <- data.frame(x=zx,rhoNot0=zy) # put in a data frame
  bfOut <- generalTestBF(x ~ rhoNot0, data=zData) # linear coefficient
  mcmcOut <- posterior(bfOut,iterations=10000) # posterior samples
  print(summary(mcmcOut[,"rhoNot0"])) # Show the HDI for r
  return(bfOut) # Return Bayes factor object
}
bfCorTest(cars$mpg, cars$hp)
```

```
##
## 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
##      -0.738780      0.124788       0.001248       0.001381
##
## 2. Quantiles for each variable:
##
##    2.5%     25%     50%     75%   97.5%
## -0.9794 -0.8218 -0.7397 -0.6574 -0.4898
```

```
## Bayes factor analysis
## --------------
## [1] rhoNot0 : 56963.84 ±0.01%
##
## Against denominator:
##    Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

- The point estimate for rho is -0.739. The 95% HDI ranges from -0.98 to -0.48, which does not include 0 we are very confident that rho could not be equal to 0.

- The bayes factor shows that the odds are in favor of the alternate hypothesis by a huge margin. showing that rho could not be equal to 0, for the population relation of mpg and hp. Taking all of this evidence together, we can say with some credibility that the population correlation is a negative value lying somewhere in the range of –0.98 up to –0.489, and probably close to a central value of –0.74.

Chapter 7, Exercise 8

The data set called UCBAdmissions (see "? UCBAdmissions" for documentation) contains data on applicants to graduate school at Berkeley for the six largest departments in 1973 classified by admission and sex. You can access the data for the first department like this: UCBAdmissions[, , 1]. Make sure you put two commas before the 1: this is a three dimensional contingency table that we are subsetting down to two dimensions. Run chisq.test() on the subset of the data set for department 1 (1 pt) and make sense of the results. (1 pt)

```
library(stats)
uca <- UCBAdmissions[, , 1] #choosing the 1st department
uca
```

```
##          Gender
## Admit      Male Female
##   Admitted  512     89
##   Rejected  313     19
```

```
chisq.test(uca,correct = FALSE) #doing the chisq test
```

```
##
##  Pearson's Chi-squared test
##
## data:  uca
## X-squared = 17.248, df = 1, p-value = 3.28e-05
```

- The observed chi-swuared value is 17.248 with df of 1 with a p-value of 3.28e-05. As this is a 2 by 2 contigency table the df is 1. Because the p-value is much smaller than the alpha threshold of $p < 0.05$, we reject the null hypothesis, of independence between Gender and Admited students.
- These two factors are not independent, as we can see from the contingency table where the ratio of males admitted is much higher than the females who were admitted. # Chapter 7, Exercise 9

Use contingencyTableBF() to conduct a Bayes factor analysis on the UCB admissions data for department 1. (1 pt) Report and interpret the Bayes factor. (1 pt)

```
contingencyTableBF(uca, sampleType = "poisson", posterior = FALSE)
```

```
## Bayes factor analysis
## --------------
## [1] Non-indep. (a=1) : 1111.64 ±0%
##
## Against denominator:
##   Null, independence, a = 1
## ---
## Bayes factor type: BFcontingencyTable, poisson
```

- The Bayes factor of 1111.64:1 is in favor of the alternative hypothesis that the two factors are not independent from one another (in other words, that the two factors are associated). Because the reported Bayes factor is in excess of 150:1, we can treat it as very strong positive evidence in favor of non-independence. Therefore, in this research situation, the Bayes factor and the null hypothes concurs with each other. # Chapter 7, Exercise 10

Using the UCBAdmissions data for department 1, run contingencyTableBF() with posterior sampling. (1 pt) Use the results to calculate a 95% HDI of the difference in proportions between the columns. (1 pt for extracting proportions, 1 pt for HDI, 1 pt for interpretation)

```
summary(contingencyTableBF(uca, sampleType = "poisson", posterior = TRUE,
                           iterations=10000)) #summary of table
```

```
##
## 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
## lambda[1,1] 511.08 22.292  0.22292        0.22292
## lambda[2,1] 312.76 17.823  0.17823        0.17823
## lambda[1,2]  89.70  9.451  0.09451        0.09451
## lambda[2,2]  19.94  4.489  0.04489        0.04299
##
## 2. Quantiles for each variable:
##
##               2.5%    25%    50%    75%  97.5%
## lambda[1,1] 468.09 495.74 511.31 526.00 555.20
## lambda[2,1] 278.67 300.45 312.37 324.62 348.85
## lambda[1,2]  72.11  83.30  89.33  95.81 109.39
## lambda[2,2]  12.08  16.82  19.57  22.74  29.68
```

```
ctMCMCout <- contingencyTableBF(uca, sampleType = "poisson", posterior = TRUE,
                                iterations=10000) #assigning variable
male_prop <- ctMCMCout[,"lambda[1,1]"]/ctMCMCout[,"lambda[2,1]"]
#choosing male prop
female_prop <- ctMCMCout[,"lambda[1,2]"]/ctMCMCout[,"lambda[2,2]"]
#female propoation
diff_prop <- male_prop - female_prop #differnece in proportion
hist(diff_prop) #histogram
abline(v = quantile(diff_prop, c(0.025)), col = 'black') #the 2.5% percentile
quantile(diff_prop, c(0.025))
```

```
##      2.5%
## -6.007925
```

```
quantile(diff_prop, c(0.975))
```

```
##      97.5%
## -1.205579
```

```
abline(v = quantile(diff_prop, c(0.975)), col = 'black')#the 97.5% percentile
```

Histogram of diff_prop
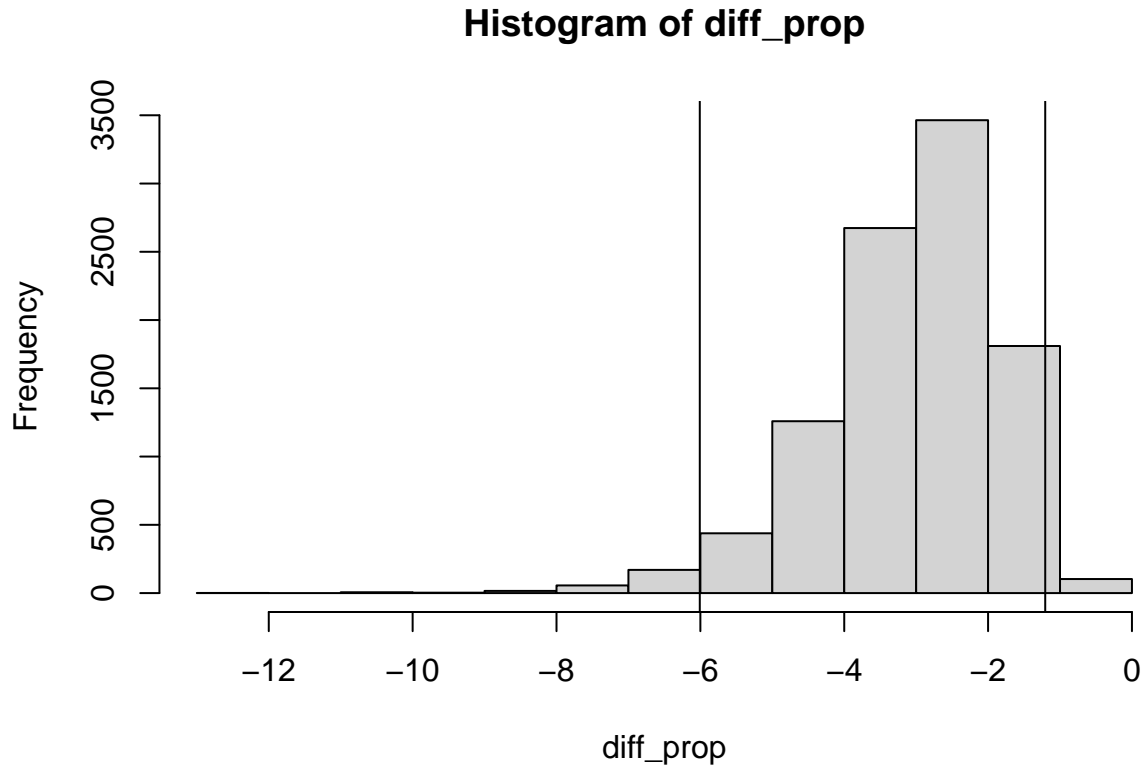
* The lower bound of the HDI is -5.93, and the upper bound of the HDI is -1.22. * Looking at the HDI, we can see that does not overlap with 0. This evidence accords with both the Bayes factor and the null hypothesis test on chi-square: there is credible evidence that in the population there is an association between admission into graduate school and gender of students.