# Attribution Statement:

1. Homework 7 by Parin Patel: I did this homework by myself, with help from the book and the professor. In addition, I used the following websites to help with the r code for the Chi-squared distribution help:
   1. <http://onlinestatbook.com/2/chi_square/distribution.html#:~:targetText=The%20degrees%20of%20freedom%20of,a%20single%20normal%20deviate%20squared.>

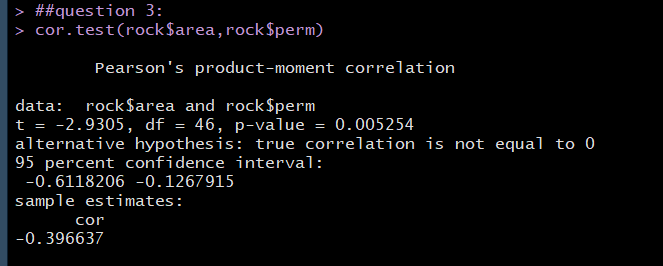
# Exercises:

1. **Run cor.test() on the correlation between “area” and “perm” in the rock data set and**

**interpret the results. Note that you will have to use the “$” accessor to get at each of**

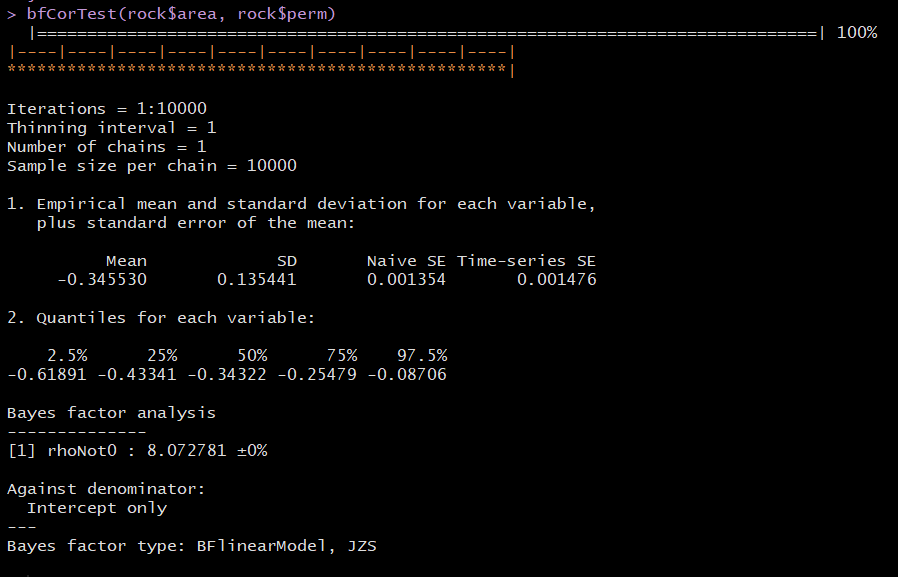
**the two variables (like this: rock$area). Make sure that you interpret both the confidence interval and the p-value that is generated by cor.test().**

The confidence interval results show that a 95% confidence range of the difference between the area and perm of the rock is [-0.612, -0.127]. The results of the confidence interval does not include 0, and therefore shows support for a significant difference between the two groups. Also, it shows of a potential negative correlation. This is further emphasized by our p-value (0.0053), which is less than our alpha of 0.05. Therefore, based on our output, we can reject the null hypothesis and support the alternative hypothesis that the true correlation is not equal to 0. In fact, sample correlation estimate, shown in Figure 1 below, shows that the rock area and permeability have slightly inverse correlation of -0.40.



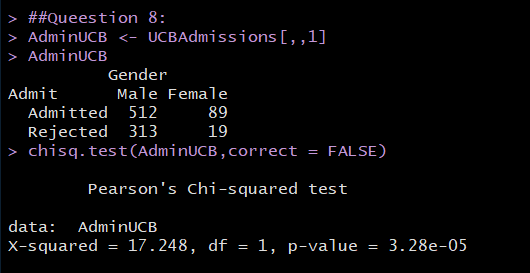
1. **Create a copy of the bfCorTest() custom function presented in this chapter. Don’t for‑ get to “source” it (meaning that you have to run the code that defines the function one time to make R aware of it). Conduct a Bayesian analysis of the correlation between “area” and “perm” in the rock data set.**

Based on the Bayesian method conducted, the output shows a mean correlation of -0.35 in the posterior distribution of the rho. This Bayesian outputs are slightly larger than the result of the correlation test. In addition, the high density interval (HDI) for the Bayesian method is [-0.62,-0.-8]. This is also slightly wider than the range for the correlation test. This range allows for more certainty of the true correlation, which is likely to be 0 or lower, and therefore supports that there is likely a true difference between the two groups. In addition, the Bayes factor supports the alternative hypothesis because 7 according to the “rule of thumb by Kass and Raftery (1995), any odds ratio between 3:1 and 20:1 are positive evidence for the favored hypothesis.”



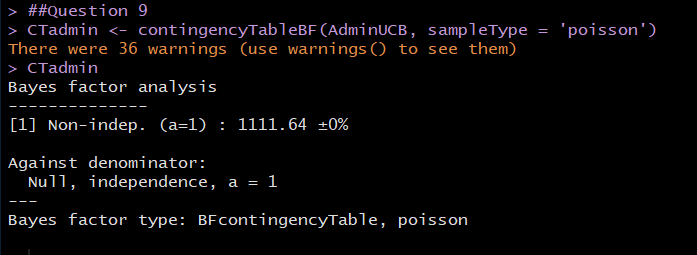
**8. Not unexpectedly, there is a data set in R that contains these data. The data set is called UCBAdmissions and you can access the department mentioned above like this: UCBAdmissions[ , ,1]. Make sure you put two commas before the 1: this is a three-dimensional contingency table that we are subsetting down to two dimensions. Run chisq.test() on this subset of the data set and make sense of the results.**

The chi-squared tests assessed the admittance of students by gender at the University of California, Berkeley. The results reported a chi-squared value of 17.25 with 1 degree of freedom (df), meaning a chi square test is the distribution of a single normal deviate squared. Therefore, only 1 variable is able to be distributed. The p-value of the is 3.28e-5, which is very low, therefore we would reject the null hypothesis that there is an independent relationship between the applicant’s gender and their admittance decision. Therefore, based on this test, we conclude that these two factors are not independent and by inspecting the 2x2 contingency table, we can see that the proportion of males admitted is much lower than the proportion of females admitted.



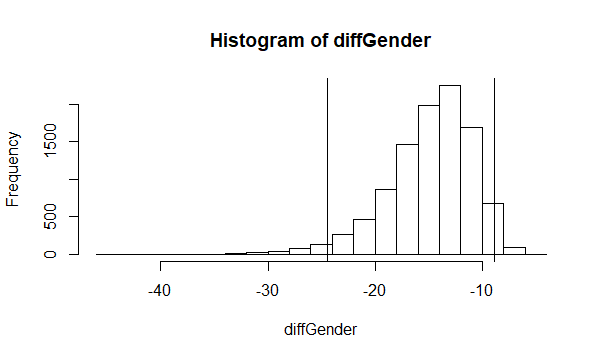
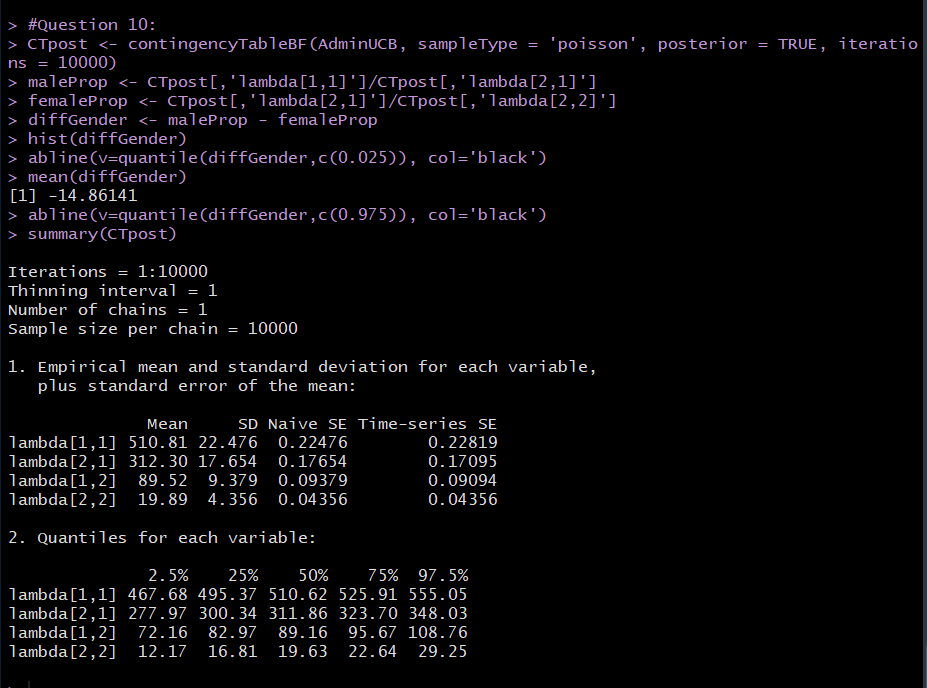
**9. Use contingencyTableBF() to conduct a Bayes factor analysis on the UCB admissions data. Report and interpret the Bayes factor.**

The Bayes Factor of 1112:1 strongly supports the alternative hypothesis that that the two factors are not independent of each other (two factors are associated).



**10. Using the UCBA data, run contingencyTableBF() with posterior sampling. Use the results to calculate a 95% HDI of the difference in proportions between the columns.**

The HDI, which represents the range of each variable, was calculated for 95%. Each lambda value shown in the figure below represents a factor from the UCBA dataset. The first lambda represents male accepted, followed by male rejected, then by female accepted, and finally the last lambda is for female rejected. For male accepted, the HDI varies from 468 to 555. The male rejection’s HDI varied from 278 to 348. Likewise, the female accepted HDI varied from 72 to 109. Finally, the female rejected HDI varied from 12 to 29. All these values, again at for at 95% HDI. The mean different in proportion of the column is determined to be -14.86.



# Appendix A: Final Script:

##question 3:

cor.test(rock$area,rock$perm)

##Question 4:

install.packages("BayesFactor")

library("BayesFactor")

#bfCorTest() , source it, area and perm

##Use pae 136 -

bfCorTest <- function (x,y)

{

zx <- scale(x) # scale X

zy <- scale(y) # scale Y

zData <- data.frame(x=zx,rhoNot0=zy) #create df

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(rock$area, rock$perm)

##Queestion 8:

AdminUCB <- UCBAdmissions[,,1]

AdminUCB

chisq.test(AdminUCB,correct = FALSE)

##Question 9

CTadmin <- contingencyTableBF(AdminUCB, sampleType = 'poisson')

CTadmin

#Question 10:

CTpost <- contingencyTableBF(AdminUCB, sampleType = 'poisson', posterior = TRUE, iterations = 10000)

maleProp <- CTpost[,'lambda[1,1]']/CTpost[,'lambda[2,1]']

femaleProp <- CTpost[,'lambda[2,1]']/CTpost[,'lambda[2,2]']

diffGender <- maleProp - femaleProp

hist(diffGender)

abline(v=quantile(diffGender,c(0.025)), col='black')

mean(diffGender)

abline(v=quantile(diffGender,c(0.975)), col='black')

summary(CTpost)

# Appendix B: Final Output:

> ##question 3:

> cor.test(rock$area,rock$perm)

Pearson's product-moment correlation

data: rock$area and rock$perm

t = -2.9305, df = 46, p-value = 0.005254

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.6118206 -0.1267915

sample estimates:

cor

-0.396637

> ##Question 4:

> install.packages("BayesFactor")

Error in install.packages : Updating loaded packages

> install.packages("BayesFactor")

Installing package into ‘C:/Users/parin/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

Warning in install.packages :

package ‘BayesFactor’ is in use and will not be installed

> library("BayesFactor")

> #bfCorTest() , source it, area and perm

> ##Use pae 136 -

> bfCorTest <- function (x,y)

+ {

+ zx <- scale(x) # scale X

+ zy <- scale(y) # scale Y

+ zData <- data.frame(x=zx,rhoNot0=zy) #create df

+ 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(rock$area, rock$perm)

|==============================================================================| 100%

|----|----|----|----|----|----|----|----|----|----|

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*|

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

-0.347404 0.137282 0.001373 0.001512

2. Quantiles for each variable:

2.5% 25% 50% 75% 97.5%

-0.61763 -0.43963 -0.34659 -0.25621 -0.07651

Bayes factor analysis

--------------

[1] rhoNot0 : 8.072781 ±0%

Against denominator:

Intercept only

---

Bayes factor type: BFlinearModel, JZS

> ##Queestion 8:

> AdminUCB <- UCBAdmissions[,,1]

> AdminUCB

Gender

Admit Male Female

Admitted 512 89

Rejected 313 19

> chisq.test(AdminUCB,correct = FALSE)

Pearson's Chi-squared test

data: AdminUCB

X-squared = 17.248, df = 1, p-value = 3.28e-05

> ##Question 9

> CTadmin <- contingencyTableBF(AdminUCB, sampleType = 'poisson')

> CTadmin

Bayes factor analysis

--------------

[1] Non-indep. (a=1) : 1111.64 ±0%

Against denominator:

Null, independence, a = 1

---

Bayes factor type: BFcontingencyTable, poisson

> #Question 10:

> CTpost <- contingencyTableBF(AdminUCB, sampleType = 'poisson', posterior = TRUE, iterations = 10000)

> maleProp <- CTpost[,'lambda[1,1]']/CTpost[,'lambda[2,1]']

> femaleProp <- CTpost[,'lambda[2,1]']/CTpost[,'lambda[2,2]']

> diffGender <- maleProp - femaleProp

> hist(diffGender)

> abline(v=quantile(diffGender,c(0.025)), col='black')

> mean(diffGender)

[1] -14.88043

> abline(v=quantile(diffGender,c(0.975)), col='black')

> summary(CTpost)

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] 510.67 22.798 0.22798 0.22798

lambda[2,1] 312.83 17.701 0.17701 0.17389

lambda[1,2] 89.46 9.525 0.09525 0.09525

lambda[2,2] 19.93 4.462 0.04462 0.04462

2. Quantiles for each variable:

2.5% 25% 50% 75% 97.5%

lambda[1,1] 467.71 494.97 510.27 525.85 556.7

lambda[2,1] 278.85 300.61 312.56 324.54 348.5

lambda[1,2] 71.94 82.80 89.21 95.64 109.0

lambda[2,2] 12.16 16.73 19.55 22.76 29.7

>