

Matthew L. Pergolski
IST 772
Dr. Block
11/25/2021

Homework 7

Beginning Statement

“I produced the material below with no assistance [direct quote from IST 772 class syllabus].” Note: Homework questions from the book may have been copied/pasted into the document for both the student and viewer’s convenience.

The homework for week seven is exercises 3,4, 8, 9, and 10 on pages 155 and 156.

Homework Question 3

Question:

3. 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()`.

Answer/Student Response:

The following was developed and observed:

```
# data
rock
data(rock)
rock$area
rock$perm

# correlation
cor.test(rock$area, rock$perm)
```

```
> 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
```

Homework Question 4

The null hypothesis assumes that ρ , the population correlation coefficient, is equal to 0. We can interpret that there is evidence of a negative correlation due to a minus sign in front of the t-test value (i.e., $t = -2.9305$; this t-value is observed on 46 degrees of freedom). We also observe that the p-value of 0.005254 is less than our standard alpha value of 0.05, so we can reject the null hypothesis. Another important aspect to note is the confidence interval, which does not straddle 0, so we have confidence that the correlation is indeed negative (the confidence interval ranges from -0.6118206 to -0.1267915).

Question:

4. Create a copy of the `bfCorTest()` custom function presented in this chapter. Don't forget 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.

Answer/Student Response:

The following was ran and observed:

```
# function copied from the class textbook page 136
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(rock$area, rock$perm)
```

```

> bfCorTest(rock$area, rock$perm)
|=====|
0
|----|----|----|----|----|----|----|----|----|
*****|
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
-0.345277  0.138783  0.001388    0.001635

2. Quantiles for each variable:

 2.5%   25%   50%   75%   97.5%
-0.62324 -0.43819 -0.34345 -0.25289 -0.07491

Bayes factor analysis
-----
[1] rhoNot0 : 8.072781 ±0%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS

> |

```

We see that evidence for the alternative hypothesis is about 8:1, which is not necessarily considered ‘strong’ when comparing our rule of thumb of 150+:1.

Homework Question 8

Question:

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.

Answer/Student Response:

The following was generated and observed:

```

UCBAdmissions
UCBAdmissions[, ,1]
chisq.test(UCBAdmissions[, ,1], correct = FALSE)

```

```
> chisq.test(UCBAdmissions[, ,1], correct = FALSE)

Pearson's Chi-squared test

data:  UCBAdmissions[, , 1]
X-squared = 17.248, df = 1, p-value = 3.28e-05
```

The observed chi-square calculated is 17.248 with an associated p-value of 3.28e-05 on one degree of freedom, which indicates we can reject the null hypothesis since this p-value value is below our standard of 0.05.

Homework Question 9

Question:

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

Answer/Student Response:

The following was generated and observed:

```
> contingencyTableBF(UCBAdmissions[, ,1], 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
```

From the Bayes Factor results, we see there is a 1111.64:1 ratio exists in favor of the alternative hypothesis, indicating that there is a strong association between the two variables. If we remember to our rule of thumb, any ratio greater than 150:1 is considered strong evidence, so we can reject the null hypothesis.

Homework Question 10

Question:

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

Answer/Student Response:

The following was generated and observed:

```

# Monte Carlo method
MCMC <- contingencyTableBF(UCBAAdmissions[, ,1], sampleType = "poisson", posterior = TRUE, iterations = 10000)
summary(MCMC)

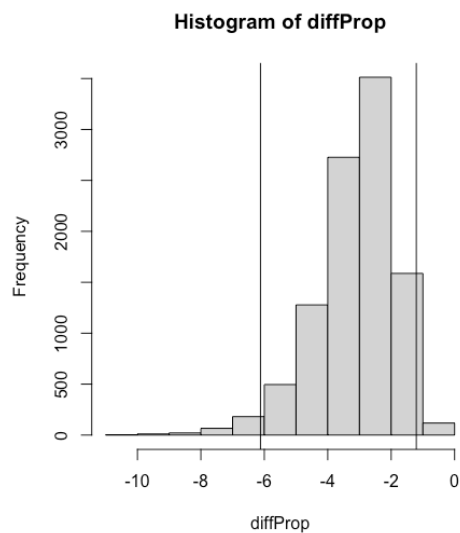
# group 1
mProp <- MCMC[, "lambda[1,1]" ] / MCMC[, "lambda[2,1]" ]
mProp
hist(mProp)

# group 2
fProp <- MCMC[, "lambda[1,2]" ] / MCMC[, "lambda[2,2]" ]
fProp
hist(fProp)

# difference between groups
diffProp <- mProp - fProp
diffProp
hist(diffProp)

abline(v=quantile(diffProp, c(.025)))
abline(v=quantile(diffProp, c(.975)))

```



We now see a histogram of the posterior distribution of differences in proportions between the two columns. The two black lines indicate the 95% HDI. The interval does not straddle 0, we have further evidence to reject the null hypothesis.