

Homework 7

Daniel Caley

8/18/2021

Contents

Question 3	2
Interpret the Results	2
Question 4	3
Interpret the Results	3
Question 8	5
Interpret the Results	5
Question 9	6
Interpret the Results	6
Question 10	7
Interpret the Results	8

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

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`).

```
set.seed(3)

MyRock <- rock

RockCor <- cor.test(MyRock$area, MyRock$perm)
RockCor

##
## Pearson's product-moment correlation
##
## data: MyRock$area and MyRock$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
```

Interpret the Results

Make sure that you interpret both the confidence interval and the `p-value` that is generated by `cor.test()`.

- The confidence interval does not pass 0
- The confidence interval shows that there is a 95% confidence that the correlation are somewhere between -0.6118 and -0.1267.
- Due to the P-Value being at 0.052, provides further confidence that the correlation could be near -0.396637.

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.

```
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(MyRock$area,MyRock$perm)
```

```
##
## 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
##      -0.347974      0.137807      0.001378      0.001552
##
## 2. Quantiles for each variable:
##
##      2.5%      25%      50%      75%      97.5%
## -0.62128 -0.43917 -0.34554 -0.25479 -0.08356
##
## Bayes factor analysis
## -----
## [1] rhoNot0 : 8.072781 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

Interpret the Results

- The point estimate is close to the `cor.test` at -0.347974.
- The 95% HDI is somewhere between -0.6118 and -0.1267 and does not cross 0.
- Due to the `rhoNot0` at 8.07 provides further evidence that the odds are in favor of the alternative hypothesis.

- In order to make 8.07 more intuitive, inverting the value like the following $1/8.0727$ comes out to be 0.12:1 which is not great odds in favor of the null hypothesis.
- Any odds ratio weaker than 3:1 is not worth mentioning.

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.

```
MyUCB <- UCBAdmissions[ , ,1]
```

```
MyUCB
```

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

```
chisq.test(MyUCB)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  MyUCB
## X-squared = 16.372, df = 1, p-value = 5.205e-05
```

Interpret the Results

- In this case the chi-squared test is looking at a 2x2 table in relation to admittance of students at UC Berkeley by gender.
- The Chi-Squared is 16.372 with 1 degree of freedom.
- Degrees of freedom are the values that have the freedom to vary in the admittance data.
- Due to having a 2 X 2 table on only 1 variable can vary or could be manipulated.
- The p-value of 5.205e-05 is extremely low therefore Male and Females are not independent and specifically UC Berkley would reject a male applicants over woman.

Question 9

Use `contingencyTableBF()` to conduct a Bayes factor analysis on the UCB admissions data.

```
BayesUCB <- contingencyTableBF(MyUCB, sampleType = 'poisson')
BayesUCB
```

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

Interpret the Results

Report and interpret the Bayes factor.

- The Non-independent number of 1111.64 + or - 0% is the Bayes factor and is help to measure if the odds are favor of the alternative Hypothesis.
- Due to having 1111.64 the odds are in the favor and further help prove that the Chi-Square is true.
- Typically anything over 150:1 is illustrates very strong evidence in favor of the hypothesis.

Question 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.

```
BayesUCB_Post <- contingencyTableBF(MyUCB, sampleType = 'poisson', posterior = TRUE, iterations = 100000)
summary(BayesUCB_Post)
```

```
##
## 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] 511.11 22.553 0.22553      0.22961
## lambda[2,1] 312.55 17.814 0.17814      0.17814
## lambda[1,2]  89.67  9.384 0.09384      0.09384
## lambda[2,2]  19.86  4.475 0.04475      0.04475
##
## 2. Quantiles for each variable:
##
##           2.5%   25%   50%   75%  97.5%
## lambda[1,1] 467.72 495.75 510.72 525.81 556.74
## lambda[2,1] 278.87 300.51 312.10 324.52 348.25
## lambda[1,2]  72.14  83.13  89.36  95.93 109.07
## lambda[2,2]  12.07  16.76  19.51  22.61  29.61
```

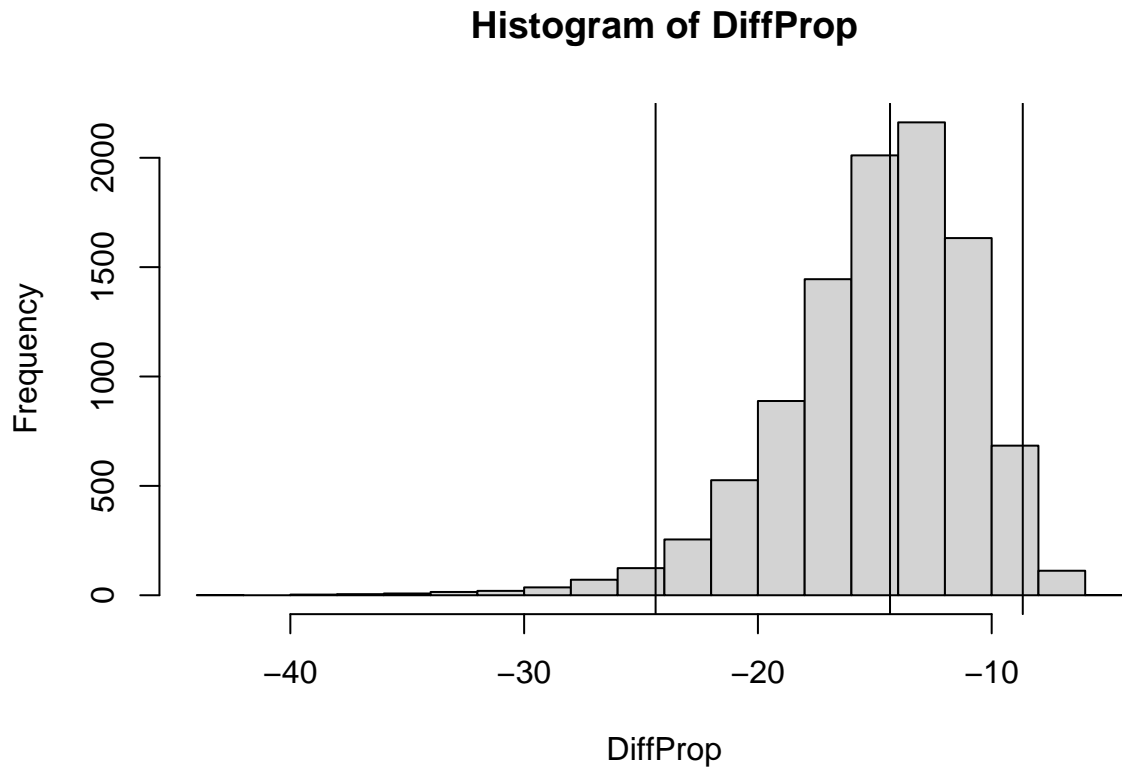
```
MaleProp <- BayesUCB_Post[, 'lambda[1,1]']/BayesUCB_Post[, 'lambda[2,1]']
FemaleProp <- BayesUCB_Post[, 'lambda[2,1]']/BayesUCB_Post[, 'lambda[2,2]']

DiffProp <- MaleProp - FemaleProp

mean(DiffProp)
```

```
## [1] -14.93356
```

```
hist(DiffProp)
abline(v=quantile(DiffProp,c(0.025)), col='black')
abline(v=quantile(DiffProp,c(0.975)), col='black')
abline(v=quantile(DiffProp,c(0.500)), col='black')
```



Interpret the Results

- When looking at the HDI, each represents males admitted and rejected; and females admitted and rejected.
- At 95% confidence males accepted range is from 468 to 557.
- Alternatively at 95% confidence the Males rejected was 279 to 348.
- The female accepted between a range of 72 and 109 at 95% confidence.
- At 95% confidence the female rejections numbers were between 12 to 30.