

## HW7\_Videtti

*##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()`.*

```
?rock
```

```
## starting httpd help server ... done
```

*#area = area in pores space, in pixels out of 256 by 256*

*#perm = permeability in milli-Darcies*

```
head(rock)
```

```
##   area    peri    shape perm
## 1 4990 2791.90 0.0903296  6.3
## 2 7002 3892.60 0.1486220  6.3
## 3 7558 3930.66 0.1833120  6.3
## 4 7352 3869.32 0.1170630  6.3
## 5 7943 3948.54 0.1224170 17.1
## 6 7979 4010.15 0.1670450 17.1
```

```
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
```

*#First, we see that our alternative hypothesis is that the true correlation between area and perm is not equal to 0, meaning that our null hypothesis is that the true correlation between area and perm is equal to 0.*

*#We see that the 95% confidence interval ranges from -0.6118206 to -0.1267915. Note that this interval is all negative values, more importantly that it does not contain 0. We also see that the p-value is 0.005254, which is less than 0.05. For these reasons, we have significant evidence to reject the null hypothesis.*

*##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.*

```
library(BayesFactor)
```

```
## Loading required package: coda
## Loading required package: Matrix
## *****
```

```
## Welcome to BayesFactor 0.9.12-4.3. If you have questions, please contact
Richard Morey (richarddmorey@gmail.com).
```

```
##
```

```
## 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(rock$area,rock$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
##    -0.344841    0.137182    0.001372    0.001532
```

```
##
```

```
## 2. Quantiles for each variable:
```

```
##
```

```
##      2.5%      25%      50%      75%      97.5%
## -0.61931 -0.43342 -0.34225 -0.25145 -0.07916
```

```
## Bayes factor analysis
## -----
## [1] rhoNot0 : 8.072781 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS
```

*#Point estimate for actual correlation (rho) is -0.344841.*  
*#95% HDI ranges from -0.61931 to -0.07916. Note that this interval does not contain 0.*  
*#Bayes factor is 8.072781, meaning that there are approximately 8:1 odds in favor of the alternative hypothesis. Using a rule of thumb odds cutoff value of 3:1, this is significant odds in favor of the alternative hypothesis.*

*#All these things considered (more specifically the 95% HDI and Bayes factor), we have significant evidence to reject the null hypothesis that the correlation between rock\$area and rock\$perm is 0. That is, we can say that there is a nonzero correlation between rock\$area and rock\$perm.*

**##8. Not unexpectedly, there is a data set in R that contains these data. The data set is called UCBAAdmissions and you can access the department mentioned above like this: UCBAAdmissions[, ,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.**

```
UCBAAdmissions[, ,1]

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

chisq.test(UCBAAdmissions[, ,1])

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

*#Chi-squared value is 16.372  
#Since we have 2 rows and 2 columns, we have  $(2-1)*(2-1) = 1$  degree of freedom.  
#We have a very low p-value of 0.00005205, meaning we would reject the null hypothesis that admissions are independent of gender. That is, we have significant evidence to say that admissions and gender are not independent.*

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

```
contingencyTableBF(UCBAdmissions[, , 1], 'poisson')  
## 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 is 1111.64, meaning that there are 1111.64:1 odds in favor of the alternative hypothesis that admissions and gender are not independent. Since these odds are more favorable than 3:1, we can choose to reject the null hypothesis, meaning that we can say that admissions and gender are not independent.*

**##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.**

```
Exercise10 <- contingencyTableBF(UCBAdmissions[, , 1], 'poisson', posterior =  
TRUE, iterations = 10000)  
summary(Exercise10)
```

```
##
## 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] 510.84 22.751 0.22751      0.22751
## lambda[2,1] 312.73 17.585 0.17585      0.17663
## lambda[1,2]  89.47  9.420 0.09420      0.09700
## lambda[2,2]  19.89  4.427 0.04427      0.04207
##
## 2. Quantiles for each variable:
##
##           2.5%    25%    50%    75%   97.5%
## lambda[1,1] 467.43 495.24 510.27 525.47 556.48
## lambda[2,1] 278.75 300.74 312.37 324.32 347.86
## lambda[1,2]  71.94  82.93  89.03  95.69 108.64
## lambda[2,2]  12.27  16.73  19.54  22.71  29.37

MaleProp <- Exercise10[, "lambda[1,1]"]/Exercise10[, "lambda[2,1]"]
FemaleProp <- Exercise10[, "lambda[1,2]"]/Exercise10[, "lambda[2,2]"]
GenderProp <- MaleProp - FemaleProp
quantile(GenderProp, 0.025)

##           2.5%
## -5.959818

quantile(GenderProp, 0.975)

##           97.5%
## -1.195032
```

*#Thus, our 95% HDI of the difference in proportions between the Male and Female columns ranges from -5.959818 and -1.195032.*