

Homework #7

3. Run correlation test function on correlation between "area" and "perm" in rock data set

```
> 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
```

The null hypothesis for this test is ρ equal to 0 with the alternative hypothesis of ρ not equal to 0 (there is no difference between the rock area and its permeability). The t-test shows a likelihood of significance at ~ -2.93 . The degrees of freedom for this correlation is 46 as one degree of freedom was lost for the calculation of the mean in each of the two samples from the 48×2 matrix. Using the conventional $p < .05$ threshold for alpha, we have rejected the null hypothesis. The procedure provided a 95% confidence interval ranging ~ -0.6118 to ~ -0.1268 , where, if we repeated this sampling process many times and each time constructed a confidence interval around the calculated value, about 95% of those constructed intervals would contain the true population value.

4. Create copy of bfCorTest() function, source it, and run Bayesian analysis on the correlation between "area" and "perm"

```
> library("BayesFactor")
> 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"])) # Get 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
-0.345230	0.137588	0.001376	0.001538

2. Quantiles for each variable:

2.5%	25%	50%	75%	97.5%
-0.62087	-0.43622	-0.34325	-0.25260	-0.07856

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

```
Against denominator:
Intercept only
---
```

```
Bayes factor type: BFlinearModel, JZS
```

The point estimate for rho is ~ -0.3436 with a 95% HDI range from -0.6178 to -0.0688 but does not straddle 0. Finally, the Bayes factor of ~ 8.0728 shows positive evidence toward the alternative hypothesis.

8. Access admission data

```
> UCBA admissions[, , 1]
      Gender
Admit   Male Female
Admitted  512     89
Rejected  313     19
```

Store 2x2 table of admission and gender

```
> UCBAAdmit <- UCBA admissions[, , 1]
```

Run chi-square test on subset of data set to make sense of results

```
> chisq.test(UCBAdmit, correct = FALSE)
```

Pearson's Chi-squared test

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

The output shows 1 degree of freedom, which is typical for a 2x2 table. The observed chi-square test is 17.248 with an associated p-value of 3.28e-05, below the alpha threshold of $p < .05$. Thus, we reject the null hypothesis.

9. Run Bayes factor contingency table on UCB admissions data
Applied Poisson sample type as observations assumed fixed

```
> AdminMCMCOut <- contingencyTableBF(UCBAdmit, sampleType =
"poisson", posterior = TRUE, iterations = 10000)
> summary(AdminMCMCOut)
```

```
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]	510.48	22.563	0.22563	0.22712
lambda[2,1]	312.53	17.650	0.17650	0.17650
lambda[1,2]	89.73	9.497	0.09497	0.09497
lambda[2,2]	19.96	4.467	0.04467	0.04340

2. Quantiles for each variable:

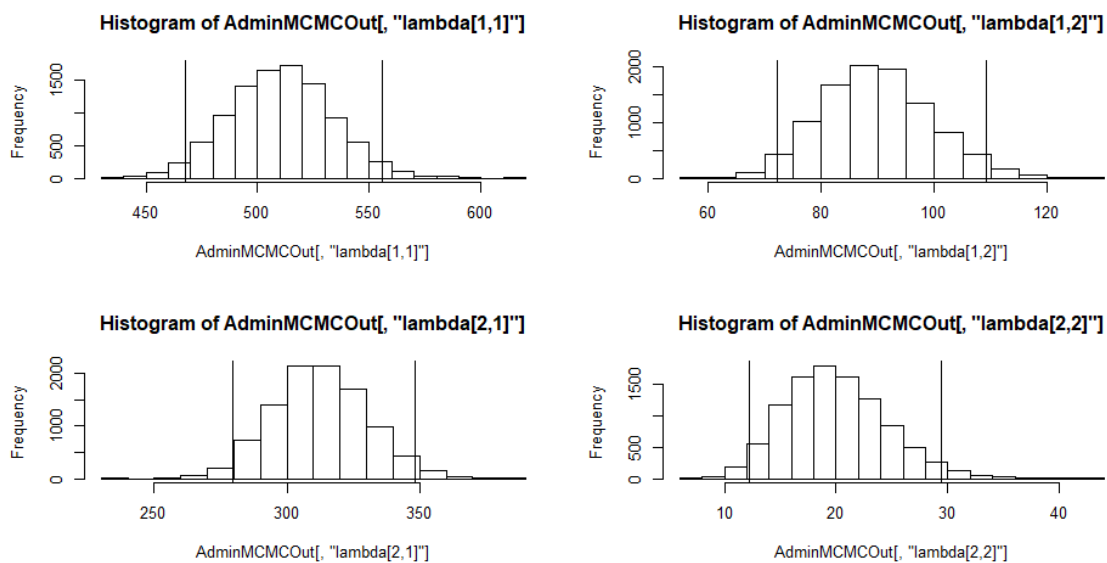
	2.5%	25%	50%	75%	97.5%
lambda[1,1]	467.63	495.10	510.34	525.46	555.53
lambda[2,1]	279.55	300.58	312.02	324.06	348.24
lambda[1,2]	72.19	82.96	89.43	95.90	109.19
lambda[2,2]	12.19	16.75	19.62	22.83	29.44

Observing the means in the first section shows a close resemblance to the original table contents. The following charts show the 95% HDI mirrored in the second section of the summary where the range for admitted males ranges from ~ 468 to ~ 556 and females ~ 72 to ~ 109 compared to the not admitted ranges ~ 279 to ~ 348 and ~ 12 to ~ 30 for males and females respectively.

```

> par(mfrow = c(2, 2))
> hist(AdminMCMCOut[, "lambda[1,1]"])
> abline(v = quantile(AdminMCMCOut[, "lambda[1,1]"], 0.025))
> abline(v = quantile(AdminMCMCOut[, "lambda[1,1]"], 0.975))
> hist(AdminMCMCOut[, "lambda[1,2]"])
> abline(v = quantile(AdminMCMCOut[, "lambda[1,2]"], 0.025))
> abline(v = quantile(AdminMCMCOut[, "lambda[1,2]"], 0.975))
> hist(AdminMCMCOut[, "lambda[2,1]"])
> abline(v = quantile(AdminMCMCOut[, "lambda[2,1]"], 0.025))
> abline(v = quantile(AdminMCMCOut[, "lambda[2,1]"], 0.975))
> hist(AdminMCMCOut[, "lambda[2,2]"])
> abline(v = quantile(AdminMCMCOut[, "lambda[2,2]"], 0.025))
> abline(v = quantile(AdminMCMCOut[, "lambda[2,2]"], 0.975))
> par(mfrow = c(1, 1))

```



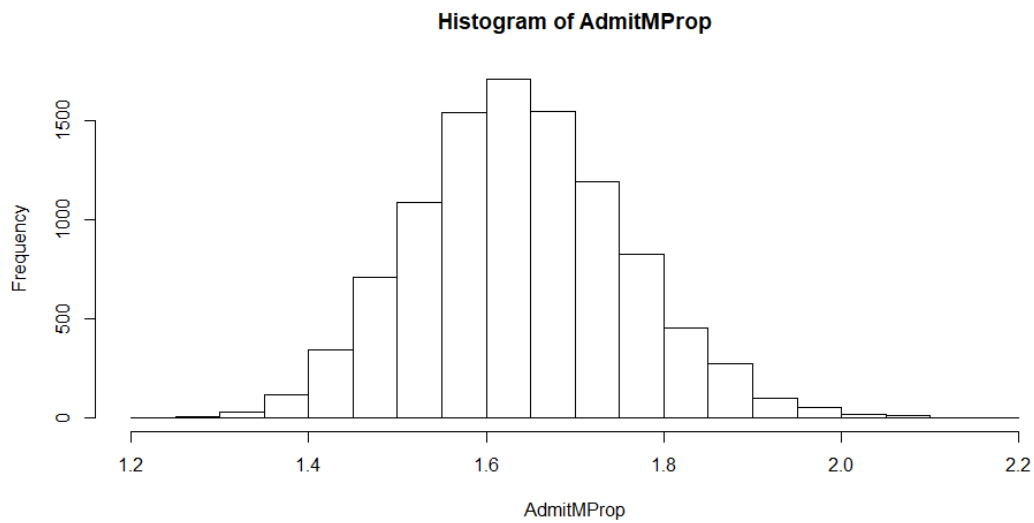
10. Use results from Bayes factor contingency table calculate 95% HDI of difference in proportions between columns

Admitted males to not admitted males

```

> AdmitMProp <- AdminMCMCOut[, "lambda[1,1]"]/AdminMCMCOut[,
"lambda[2,1]"]
> hist(AdmitMProp)

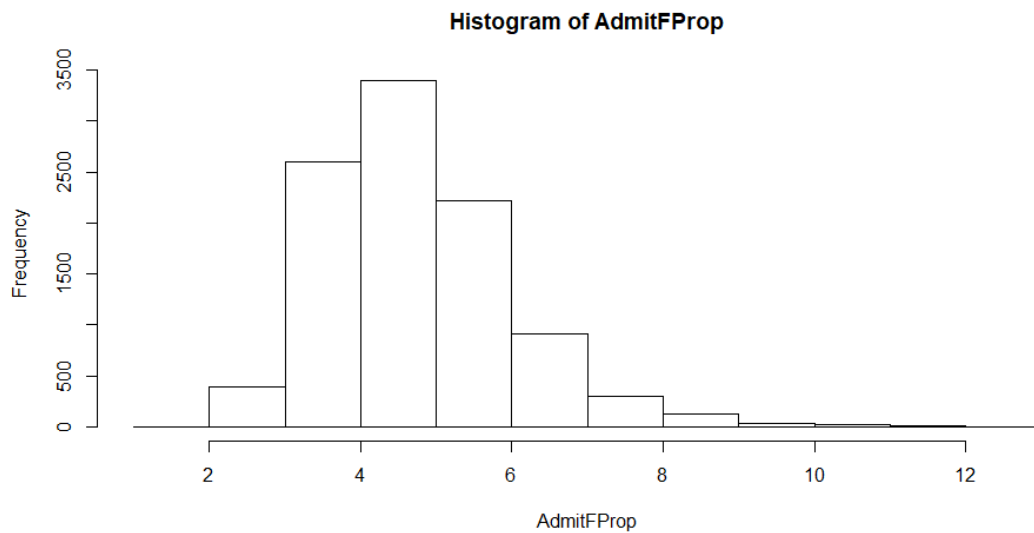
```



```
> mean(AdmitMProp)
[1] 1.638672
```

Admitted females to not admitted females

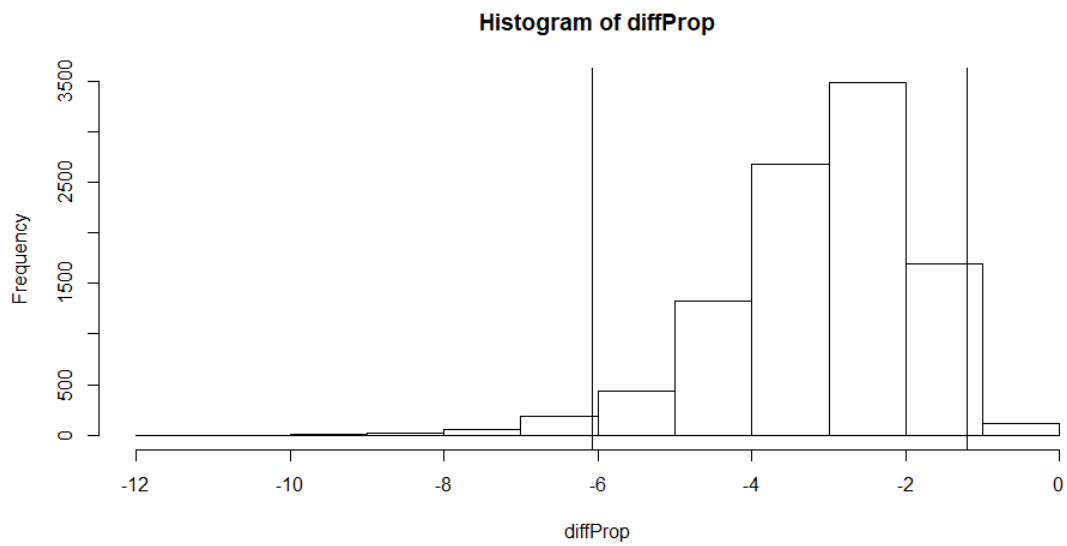
```
> AdmitFProp <- AdminMCMCOut[, "lambda[1,2]"]/AdminMCMCOut[,
"lambda[2,2]"]
> hist(AdmitFProp)
```



```
> mean(AdmitFProp)
[1] 4.732119
```

Compare proportion distribution for admitted males to admitted females

```
> diffProp <- AdmitMProp - AdmitFProp
> hist(diffProp)
```



```
> abline(v = quantile(diffProp, c(0.025)))  
> abline(v = quantile(diffProp, c(0.975)))  
> mean(diffProp)  
[1] -3.093447
```

This is how much admitted:not admitted ratio decreases as we switch from male to female with a 95% HDI range of ~ -6 to ~ -1.6 and the average at ~ -3 . This means that 3 times as many males are admitted as females.