Josh Rhoades
Biostatistics HW 5

**1.**

```
combined6.14<-array(c(152,335,183,797),,dim=c(2,2,1))
 chisq.test(combined6.14)

        Chi-squared test for given probabilities

data:  combined6.14
X-squared = 725.3027, df = 3, p-value < 2.2e-16
```

the p value is highly significant, there is something going on with less than 5 cups a day group.

```
strat6.14 <-
array(c(7,55,31,269,7,20,18,112,7,33,24,114,40,88,45,172,34,
50,24,55,27,55,24,58,30,34,17,17),dim=c(2,2,7))

 mantelhaen.test(strat6.14 )
        Mantel-Haenszel chi-squared test with continuity
correction

data:  strat6.14
Mantel-Haenszel X-squared = 5.0709, df = 1, p-value =
0.02433
alternative hypothesis: true common odds ratio is not equal to
1
95 percent confidence interval:
 1.050526 1.800678
sample estimates:
common odds ratio
         1.375376
```

The odds ratio of 1 is outside of the CI,, smoking matters.

```
 chisq.test(strat6.14)
        Chi-squared test for given probabilities

data:  strat6.14
X-squared = 1643.367, df = 27, p-value < 2.2e-16
```

Smoking is a confounding factor, people that smoke are also more likely to drink more coffee, so the carrelation between coffee and MI is confounded by smoking.

**2.**

```
Admit <- c(rep(1,512), rep(0,313), rep(1,89), rep(0,19),
rep(1,353), rep(0,207), rep(1,17), rep(0,8),
rep(1,120),rep(0,205), rep(1,202), rep(0,391),
rep(1,138),rep(0,279), rep(1,131), rep(0,244), rep(1,53),
rep(0,138), rep(1,94), rep(0,299), rep(1,22), rep(0,351),
rep(1,24) , rep(0,317))

Gender <- c(rep('Male',512), rep('Male', 313),
rep('Female',89), rep('Female',19), rep('Male', 353),
rep('Male', 207), rep('Female',17), rep('Female',8), rep('Male',
120), rep('Male',205), rep('Female',202), rep('Female',391),
rep('Male', 138), rep('Male',279), rep('Female',131),
rep('Female',244), rep('Male',53), rep('Male', 138),
rep('Female',94), rep('Female',299), rep('Male',22), rep('Male',
351), rep('Female',24), rep('Female',317))

Dept <- c(rep('A',512), rep('A', 313), rep('A',89), rep('A',19),
rep('B',353),rep('B',207),rep('B',17),rep('B',8),
rep('C',120),rep('C',205),rep('C',202),rep('C',391),rep('D',138),
rep('D',279),rep('D',131),rep('D',244),rep('E',53),rep('E',138),r
ep('E',94),rep('E',299),rep('F',22),rep('F',351),rep('F',24),
rep('F', 317))

> glm(Admit~Gender*Dept, family=binomial)
Call:  glm(formula = Admit ~ Gender * Dept, family =
binomial)
Coefficients:
    (Intercept)       GenderMale            DeptB            DeptC
DeptD            DeptE
         1.5442          -1.0521          -0.7904          -2.2046
-2.1662          -2.7013
          DeptF  GenderMale:DeptB  GenderMale:DeptC
GenderMale:DeptD  GenderMale:DeptE  GenderMale:DeptF
         -4.1250           0.8321           1.1770           0.9701
1.2523           0.8632

Degrees of Freedom: 4525 Total (i.e. Null);  4514 Residual
Null Deviance:      6044
Residual Deviance: 5167         AIC: 5191
```

Departments A and B accept many more than it rejects, this is quite different from all other departments. Departments C D and E all reject fewer than they accept, they do so at similar rates.  Department F rejects far more than they accept, they are more stringent than all other departments. Department A has a higher rate of female acceptance, the gender of applicants does confound the results, it is likely that due to over application of males, a higher rate of rejection exists when compared to under represented females.

**3.**

**A**

```
> chisq.test(c(78,71,87,86), p=c(1/4,1/4,1/4,1/4))
        Chi-squared test for given probabilities

data:  c(78, 71, 87, 86)
X-squared = 2.0994, df = 3, p-value = 0.552
```

We only use the row that applies to SIDS.

```
> chisq.test(c(78,71,87,86), p=c(1/4,1/4,1/4,1/4))$expected
[1] 80.5 80.5 80.5 80.5
```

**B**

```
chisq.test(c(40,19,40,43), p=c(1/4,1/4,1/4,1/4))
        Chi-squared test for given probabilities

data:  c(40, 19, 40, 43)
X-squared = 10.3944, df = 3, p-value = 0.01549
```

I could not figure out how to do the 10% significance level.

**C**

```
 chisq.test(c(50,48,46,34), p=c(1/4,1/4,1/4,1/4))
        Chi-squared test for given probabilities

data:  c(50, 48, 46, 34)
X-squared = 3.4831, df = 3, p-value = 0.323

chisq.test(c(30,40,36,35), p=c(1/4,1/4,1/4,1/4))
        Chi-squared test for given probabilities

data:  c(30, 40, 36, 35)
X-squared = 1.4397, df = 3, p-value = 0.6963
```

The p values are large, so that means we cannot reject the null hypothesis. Therefore it is likely that the deaths for both asphyxiation and immaturity are evenly spread across seasons.

**D** Ho= pie = ¼, it does not matter

**4**

```
ownership<-
array(c(583,139,524,145,24,59,86,24,74,182,31,145),dim=c(3
,4))
 chisq.test(ownership)
        Pearson's Chi-squared test

data:  ownership
X-squared = 26.722, df = 6, p-value = 0.0001632

 residuals<-chisq.test(ownership)$residuals
residuals
          [,1]      [,2]       [,3]      [,4]
[1,] -1.3132641  3.048726 -0.51442802  0.3858087
[2,]  0.3673363 -0.131866  0.91987349 -1.2395370
[3,]  1.2719903 -3.328760  0.09369146  0.2163034

> residualpvalue<-pnorm(abs(residuals),lower.tail=F)*2
> residualpvalue
          [,1]        [,2]      [,3]      [,4]
[1,] 0.1890940 0.0022981429 0.6069528 0.6996383
[2,] 0.7133682 0.8950902423 0.3576389 0.2151467
[3,] 0.2033766 0.0008723341 0.9253543 0.8287513
```

**5**

```
aov(Ozone~Month,data = airquality)
Call:
   aov(formula = Ozone ~ Month, data = airquality)

Terms:
                  Month Residuals
Sum of Squares    3387.2  121755.9
Deg. of Freedom        1       114

Residual standard error: 32.68079
Estimated effects may be unbalanced
37 observations deleted due to missingness

summary(output)
            Df Sum Sq Mean Sq F value Pr(>F)
Month        1   3387    3387   3.171 0.0776 .
Residuals   114 121756    1068
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
37 observations deleted due to missingness
```

Months p value not significant.

I could not get the TukeyHSD to work, I kept getting errors:

```
summary(output<- aov(Ozone~Month,data = airquality))
            Df Sum Sq Mean Sq F value Pr(>F)
Month        1   3387    3387   3.171 0.0776 .
Residuals   114 121756    1068
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
37 observations deleted due to missingness

> TukeyHSD(output, "Month", ordered = FALSE)
Error in TukeyHSD.aov(output, "Month", ordered = FALSE) :
  no factors in the fitted model
In addition: Warning message:
In replications(paste("~", xx), data = mf) : non-factors
ignored: Month

boxplot(Ozone ~ Month, data= airquality)
```