MATH-650 Assignment 12

Saket Choudhary (USCID: 2170058637) (skchoudh@usc.edu)

12/4/2015

Chapter 19: Problem 10

```
coffee.data <- read.csv('data_10.csv')
n11 <- coffee.data[1,2]
n12 <- coffee.data[1,3]
n21 <- coffee.data[2,2]
n22 <- coffee.data[2,3]
R1 <- n11+n12
R2 <- n21+n22
C1 <- n11+n21
C2 <- n12+n22
T1 <- R1+R2
```

```
e11 <- R1*C1/T1
e12 <- R1*C2/T1
e21 <- R2*C1/T1
e22 <- R2*C2/T1
chi2 <-  T1*(abs(n11*n22-n12*n21)-T1/2)^2/(R1*C1*R2*C2)
chi2
```

```
## [1] 7.22013
```

```
pval <- pchisq(chi2, 1, lower.tail = F)
pval
```

```
## [1] 0.00720905
```

Check with `chisq.test`

```
chisq.test(coffee.data[,-1])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  coffee.data[, -1]
## X-squared = 7.2201, df = 1, p-value = 0.007209
```

Conclusion: Thus with a p-value of `0.007209`, there is a convincing evidence that drinking alcohol and being sexually active are not independent.

Chapter 19: Problem 13

```
smoker.data <- read.csv('data_13.csv')
smoker.data
```

```
##      Smoking Cancer Control
## 1    Smokers     83      72
## 2 NonSmokers      3      14
```

```
n11 <- smoker.data$Cancer[1]
n12 <- smoker.data$Control[1]
n21 <- smoker.data$Cancer[2]
n22 <- smoker.data$Control[2]
R1 <- n11+n12
R2 <- n21+n22
C1 <- n11+n21
C2 <- n12+n22
T1 <- R1+R2

ex <- R1*C1/T1
excess <- n11-ex
variance <- R1*R2*C1*C2/(T1*T1*(T1-1))
Z <- excess/variance
pval.excess <- pnorm(Z, lower.tail = F)
if (n11>ex){
  possiblek <- seq(n11,min(R1,C1),1);
}else{
possiblek <- seq(0,n11,1)
}

pval.fisher <- sum(dhyper(possiblek, R1, R2, C1))
```

```
pval.excess
```

```
## [1] 0.07668854
```

```
pval.fisher
```

```
## [1] 0.004411402
```

Fisher's Exact test p-value: 0.0044114

Excess test p-value: 0.0766885

Thus on a threshold of significance = 0.05, we can reject using Fisher's exact test but not using excess test where the null hypothesis is that the observed count of smoking persons is not a random allocation based on cancer or control.