STAT310 - HWK Solution 12

1. Null hypothesis is there is no relationship between taking vitamin C and the incidence of colds.

   (a) Fisher's Exact Test

```
> o = matrix(c(31, 17, 109, 122), 2, 2)
> fisher.test(o, alternative="greater")

        Fisher's Exact Test for Count Data

data:  o   p-value = 0.02052
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 1.134558      Inf
sample estimates: odds ratio
  2.035861
```

   P-value = 0.02052, thus we reject the null hypothesis. There is relationship between taking vitamin C and the incidence of colds.

   or we may use dhyper() function

```
> sum(dhyper(31:48, 140, 139, 48))
[1] 0.02052272
```

   (b) Likelihood Ratio Test
   Null model: Under the assumption of independence of $\theta_{ij} = \frac{f_{i\bullet}}{n} * \frac{f_{\bullet j}}{n}$, where $f_{i\bullet}$ is the total counts of ith row, $f_{\bullet j}$ is the total counts of jth column, and n is the total counts of all cells.

   Full model: $\theta_{ij} = \frac{f_{ij}}{n}$, where f_{ij} is the counts of $cell_{ij}$

   The likelihood function is $L(\theta|s) = \Pi_{i,j}\theta_{ij}^{f_{ij}}$, then the log-likelihood function is $l(\theta|s) = log(L(\theta|s)) = \Sigma_{ij} f_{ij} log(\theta_{ij})$

   The $X^2 = -2log\Lambda = -2log\frac{L(\theta|s)_0}{L(\theta|s)_f} = -2(l(\theta|s)_0 - l(\theta|s)_f) = 4.8177$
   Then $p - value = P(X^2 > 4.8177) = 0.0273$. Thus we reject the null hypothesis.

   (c) Chi-squared Test

```
> chisq.test(o, correct = F)

        Pearson's Chi-squared test

data:  o X-squared = 4.8114, df = 1, p-value = 0.02827
```

   Thus we reject the null hypothesis.

2. The null hypothesis is that there is no relationship between hair color and gender.

```
> # Put the observed data in a matrix named 'o'
> o = matrix(c(592, 544, 119, 97, 849, 677, 504, 451, 36, 14), 2, 5)
>
> # n is the sample size
> n = sum(o)
>
> # x is the row sums
> x = apply(o,1,sum)
>
> # y is the column sums
> y = apply(o, 2, sum)
>
> # the expected counts under the null model
> # the use of the outer product operator x %o% y which produces a matrix with ij element x
> e = (x %o% y)/n
>
> # The chi-square test statistic is sum of (observed-expected)^2/expected
> w = sum( (o-e)^2/e )
>
> # The number of degrees of freedom is (a-1)*(b-1).
> df.1 = (length(x)-1)*(length(y)-1)
>
> # p-value from a chi-square distribution.
> p.1 = 1 - pchisq(w, df.1)
>
> # Also examine the standardized residuals
> theta.null = e/n
> r = (o-e)/sqrt(e*(1-theta.null))
>
> # Print some results
>
> #Observed Data
> print(o)
     [,1] [,2] [,3] [,4] [,5]
[1,]  592  119  849  504   36 [2,]  544   97  677  451   14
>
> #Expected Counts (rounded)
> print(round(e,1))
      [,1]  [,2]  [,3]  [,4] [,5]
[1,] 614.4 116.8 825.3 516.5   27 [2,] 521.6  99.2 700.7 438.5   23
>
> #Standardized residuals (rounded)
> print(round(r,2))
      [,1]  [,2]  [,3]  [,4]  [,5]
[1,] -0.98  0.21  0.93 -0.59  1.73 [2,]  1.05 -0.22 -0.99  0.63
-1.88
>
> #Chi-square test statistic
> print(w)
```

```
[1] 10.46745
>
> #Degrees of freedom
> print(df.1)
[1] 4
>
> #P-value
> print(signif(p.1,2))
[1] 0.033
```

P-value given above is 0.033, thus we reject the null hypothesis. There is relationship between hair color and gender.

Or you may use chisq.test() function

```
> chisq.test(o, correct = F)

        Pearson's Chi-squared test

data:  o X-squared = 10.4674, df = 4, p-value = 0.03325
```

3. Step 1: Prove that if $\frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}} = 1, then \theta_{ij} = \theta_{i\bullet}\theta_{\bullet j}$ for all i, j
   $\theta_{1\bullet}\theta_{\bullet 1} = (\theta_{11} + \theta_{12})(\theta_{11} + \theta_{21}) = \theta_{11}^2 + \theta_{11}\theta_{12} + \theta_{11}\theta_{21} + \theta_{12}\theta_{21}$
   From $\frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}} = 1$, we have $\theta_{11}\theta_{22} = \theta_{12}\theta_{21}$
   Then $\theta_{1\bullet}\theta_{\bullet 1} = \theta_{11}^2 + \theta_{11}\theta_{12} + \theta_{11}\theta_{21} + \theta_{11}\theta_{22} = \theta_{11}(\theta_{11} + \theta_{12} + \theta_{21} + \theta_{22}) = \theta_{11}$
   Thus we have proved that $\theta_{11} = \theta_{1\bullet}\theta_{\bullet 1}$. Using similar method, we can prove $\theta_{12} = \theta_{1\bullet}\theta_{\bullet 2}$, $\theta_{21} = \theta_{2\bullet}\theta_{\bullet 1}$, and $\theta_{22} = \theta_{2\bullet}\theta_{\bullet 2}$.

   Step 2: Prove that if $\theta_{ij} = \theta_{i\bullet}\theta_{\bullet j}$ for all i,j, then $\frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}} = 1$
   $\frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}} = \frac{(\theta_{1\bullet}\theta_{\bullet 1})(\theta_{2\bullet}\theta_{\bullet 2})}{(\theta_{1\bullet}\theta_{\bullet 2})(\theta_{2\bullet}\theta_{\bullet 1})} = 1$

   Therefore, the row and column variables are independent if and only if $\frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}} = 1$.