Exercise for Lecture 14

Rei Monden

Exercise 14-1

```r
birthw <- read.csv("Datasets/birthweight.csv", header =TRUE)

# 2x2 contingency table
birth.tab <- xtabs(~ Normal + Low, data = birthw)
birth.tab
```

```
##        Low
## Normal   0   1
##      0 159  22
##      1   8  14
```

```r
mcnemar.test (birth.tab,
              correct = FALSE)
```

```
##
##  McNemar's Chi-squared test
##
## data:  birth.tab
## McNemar's chi-squared = 6.5333, df = 1, p-value = 0.01059
```

$\chi^2(1) = 6.53$, $p = 0.01$. We reject the marginal homogeneity assumption at $\alpha = 5\%$. In other words, the sample marginal proportions of smoking for both variables,

$$\widehat{P}(\text{Normal} = 1) = \frac{8 + 14}{203} = 0.11 \quad \text{and} \quad \widehat{P}(\text{Low} = 1) = \frac{22 + 14}{203} = 0.18$$

are deemed statistically different from each other.

Exercise 14-2

a.

```r
heaven <- read.csv("Datasets/Heaven.csv", header =TRUE)

# 2x2 contingency table
heaven.tab <- xtabs(~ Heaven + Hell, data = heaven)
heaven.tab
```

```
##         Hell
## Heaven   0   1
##      0 160   2
##      1 125 833
```

```r
mcnemar.test (heaven.tab,
              correct = FALSE)
```

1

```
##
##  McNemar's Chi-squared test
##
## data:  heaven.tab
## McNemar's chi-squared = 119.13, df = 1, p-value < 0.00000000000000022
```

The test result is $\chi^2(1) = 119.13$, $p < 0.001$. We reject the marginal homogeneity assumption at $\alpha = 5\%$. That is, the sample marginal proportions of belief in either heaven or hell,

$$\widehat{P}(\text{Heaven} = 1) = \frac{125 + 833}{1120} = 0.86 \quad \text{and} \quad \widehat{P}(\text{Hell} = 1) = \frac{2 + 833}{1120} = 0.75$$

are statistically different from each other.

b.

```
library(PropCIs)

# Wald 90% CI:
diffpropci.Wald.mp(heaven.tab[2,1],
                   heaven.tab[1,2],
                   sum(heaven.tab),
                   0.90)
```

```
##
##
##
## data:
##
## 90 percent confidence interval:
##  -0.12546701 -0.09417584
## sample estimates:
## [1] -0.1098214
```

```
# Score 90% CI:
scoreci.mp(heaven.tab[2,1],
           heaven.tab[1,2],
           sum(heaven.tab),
           0.90)
```

```
##
##
##
## data:
##
## 90 percent confidence interval:
##  -0.12637784 -0.09504691
```

If we repeat the experiment many times, approximately 90% of the confidence intervals calculated in this way would include the true difference between the population proportions.

Exercise 14-3

a.

```
neuros <- read.table("Datasets/Neurologists.dat", header = TRUE)

neuros.fit <- glm(count ~ factor(A) + factor(B),
                  family = poisson(link = "log"),
```

```
                  data    = neuros)
summary(neuros.fit)
```

```
##
## Call:
## glm(formula = count ~ factor(A) + factor(B), family = poisson(link = "log"),
##      data = neuros)
##
## Coefficients:
##               Estimate Std. Error z value            Pr(>|z|)
## (Intercept)    3.21106    0.16709  19.217 < 0.0000000000000002 ***
## factor(A)2     0.06596    0.20977   0.314              0.7532
## factor(A)3    -0.22884    0.22649  -1.010              0.3123
## factor(A)4    -0.64870    0.25730  -2.521              0.0117 *
## factor(B)2    -0.81990    0.19731  -4.155      0.00003247831 ***
## factor(B)3    -2.03292    0.32065  -6.340      0.00000000023 ***
## factor(B)4    -1.59760    0.26595  -6.007      0.00000000189 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 161.740  on 15  degrees of freedom
## Residual deviance:  69.163  on  9  degrees of freedom
## AIC: 136.44
##
## Number of Fisher Scoring iterations: 6
```

The test result is $\chi^2(9) = 69.16$, $p < .001$ (the critical value is 16.92). We conclude that the independence model fits significantly worse than the saturated model. But, as explained in the lecture, this is not a major problem, as what we are mostly interested in is in the residuals from this model.

```
neuros.stdres <- rstandard(neuros.fit, type = "pearson")
matrix(neuros.stdres, ncol = 4, byrow = TRUE)
```

```
##            [,1]       [,2]       [,3]       [,4]
## [1,]  4.778145 -2.4633322 -2.2309078 -2.270845
## [2,]  2.311913 -0.2738636 -0.3167425 -2.973579
## [3,] -3.792088  2.3745061  1.7855553  1.219738
## [4,] -4.556948  0.6762936  1.1290412  5.260539
```

We can see that, in the main diagonal, the 1st and 4th standardized residuals are indeed large and positive, indicating that agreement for these categories is greater than expected by chance (i.e., under independence). But for ratings 2 and 3 the standardized residuals are not so clearly large in the positive direction. Altogether, we may say that there is at least some evidence indicating that the independence model has missed some existing agreement between the raters.

    b.

```
neuros.qind <- glm(count ~ factor(A) + factor(B) + factor(diag),
                   family = poisson(link = "log"),
                   data    = neuros)
summary(neuros.qind)
```

```
##
## Call:
## glm(formula = count ~ factor(A) + factor(B) + factor(diag), family = poisson(link = "log"),
```

```
##     data = neuros)
##
## Coefficients:
##               Estimate Std. Error z value   Pr(>|z|)
## (Intercept)     1.4145     0.4832   2.927   0.003418 **
## factor(A)2      1.8919     0.4984   3.796   0.000147 ***
## factor(A)3      1.1554     0.4603   2.510   0.012069 *
## factor(A)4      0.3172     0.5053   0.628   0.530176
## factor(B)2      0.1301     0.2993   0.435   0.663800
## factor(B)3     -1.8205     0.4406  -4.132 0.00003591 ***
## factor(B)4     -1.8488     0.4090  -4.520 0.00000617 ***
## factor(diag)1   2.2231     0.5097   4.362 0.00001291 ***
## factor(diag)2  -1.0385     0.4516  -2.300   0.021458 *
## factor(diag)3   0.8601     0.6556   1.312   0.189540
## factor(diag)4   2.4197     0.5846   4.139 0.00003492 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 161.740  on 15  degrees of freedom
## Residual deviance:  22.045  on  5  degrees of freedom
## AIC: 97.32
##
## Number of Fisher Scoring iterations: 6
```

The quasi-independence model does seem to fit better than the independence model: $\chi^2(5) = 22.05$, $p = 0.0005134$. Furthermore, in accordance with the findings from the previous question, we now see that the δ_i parameters for ratings 1 and 4 are indeed large and positive, thus showing strong agreement for these ratings. This is not the case for ratings 2 and 3.

c.

```r
library(psych)
neuros.mat <- matrix(neuros$count,
                     ncol  = 4,
                     byrow = TRUE)
neuros.mat
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   38    5    0    1
## [2,]   33   11    3    0
## [3,]   10   14    5    6
## [4,]    3    7    3   10
```

```r
cohen.kappa(neuros.mat)
```

```
## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels,
##     w.exp = w.exp)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
##                   lower estimate upper
## unweighted kappa   0.11     0.21  0.31
## weighted kappa     0.41     0.52  0.64
##
##  Number of subjects = 149
```

The difference between the observed agreement to the expected agreement if the ratings were independent is 21% of the maximum possible difference (as expressed by kappa's denominator).