STAT 659 Spring 2016
Homework 11 Solution

8.2

(a) $H_0 : \pi_{1+} = \pi_{+1}$ $v.s.$ $H_a : \pi_{1+} \neq \pi_{+1}$

```
            McNemar's Test
--------------------------------------
Statistic (S)                 119.1260
DF                                   1
Asymptotic Pr > S               <.0001
Exact Pr >= S                9.556E-35
```

Since P-value < 0.0001, we will reject H_0 and conclude that the population proportions answering yes were not identical for heaven and hell.

(b) 90% CI is (0.094, 0.125). Since 0 is not in the interval of (0.094, 0.125), we may conclude that there is significant evidence of difference between the population proportions.

8.3

(a)

```
           Odds Ration Estimates
-------------------------------------------
                 Point              95%Wald
Effect         Estimate         Confidence Limits
x               2.018            1.630   2.500
```

The population odds of belief in heaven estimated to be 2.02 times population odds of belief in hell.

(b) A subjects estimated odds of a yes response of believing in heaven are $\frac{125}{2} = 62.5$ times that of believing in hell.

8.5

(a) This is probability, under H_0, of observed or more extreme result, with more extreme defined in direction specified by H_a.

(b) Mid P-value includes only half observed probability, added to probability of more exteme results.

(c) When binomial parameter =0.50, binomial is symmetric, so two-sided P-value=2(one-sided P-value) in (a) and (b)

8.6

(a) $H_0 : \pi^F_{s+} = \pi^F_{+s}$ $v.s.$ $H_a : \pi^F_{s+} \neq \pi^F_{+s}$

```
              McNemar's Test
--------------------------------------
Statistic (S)                 150.5208
DF                                   1
Asymptotic Pr > S               <.0001
Exact Pr >= S                8.291E-41
```

Since P-value < 0.0001 ,we will reject H_0 and conclude that there is strong evidence that for females the proportions of supporting government action for these two items are not identical.

(b) 90% CI is (0.4314,0.5290). Since 0 is not in the interval of (0.4314,0.5290), we may conclude that there is significant evidence of difference between the population proportions.

(c) i. Marginal Model
The odds ratio estimate for a logistic model for the probability of a "support" response for the marginal model is $\frac{295 \cdot 229}{59 \cdot 125} = 9.16$. Thus, the population odds of supporting the information option is estimated to be 9.16 times the population odds of supporting the health option.
ii. The Conditional Model
The odds ratio estimate for a logistic model for the probability of a "support" response for the conditional model is $\frac{n_{12}}{n_{21}} = \frac{181}{11} = 16.45$. The odds of a person supporting the information option is estimated to be 16.45 times the odds of a person supporting the health option.

(d) SE for the difference for the males is $\sqrt{(160+6)-(160-6)^2/267}/267 = 0.03290258$. SE for the difference for the females $\sqrt{(181+11)-(181-11)^2/354}/354 = 0.02967602$. Thus the SE for the difference between males and females in their differences of proportion of support is $\sqrt{0.03290258^2 + 0.02967602^2} = 0.04430853$. So the 90% CI is

$$\left(\frac{236}{267} - \frac{82}{267}\right) - \left(\frac{295}{354} - \frac{125}{354}\right) \pm 1.645 \times 0.0443$$

, which is $(0.0237, 0.1694)$.

8.7

We know that $p_1 = 0.314$ and $p_2 = 0.292$. Then $p_1 - p_2 = 0.022$, $SE = \sqrt{\frac{0.314(1-0.314)}{1144} + \frac{0.292(1-0.292)}{1144}} = 0.0192$. Then 95% CI for the true difference of proportion is $0.022 \pm 2 \times 0.019$, that is, (-0.016, 0.060), which is wider than dependent samples.
The SE in 8.1.2 is 0.0135 and the resulting CI is (-0.04, 0.048) . The paired approach with dependent samples yields a smaller SE and narrower CI.

8.11

$\hat{\beta} = \log\left(\frac{132}{107}\right) = 0.20997$,
$\hat{var}\left(\beta\right) = \frac{1}{n_{12}} + \frac{1}{n_{21}} = \frac{1}{132} + \frac{1}{107} = 0.01692$
95 % CI for β is $\log\left(\frac{132}{107}\right) \pm 1.96\sqrt{\frac{1}{132} + \frac{1}{107}} = (-0.04499, 0.46494)$. The corresponding CI for odds ratio is $(e^{-0.04498933}, e^{0.4649355}) = (0.9560, 1.5919)$.
The odds of a yes response was between 0.96 and 1.60 higher for paying higher taxes than for accepting a cut in living standard.

8.13

(a) Equation 8.9, $\gamma_{ij} = \frac{n_{ij}-n_{ji}}{\sqrt{n_{ij}-n_{ji}}}$. More people tend to move from category 2 to 1, from category 1 to 4, and from category 2 to 4 under the assumption if the symmetry model holds.

(b) Quasi-symmetry model has deviance $G^2 = 2.3$ with $df = 3$. P-value=0.5125209
There is no lack of fit in terms of the quasi-symmetry model.

(c) H_0 : Symmetry vs. H_a : Quasi symmetry, $G^2\left(S|QS\right) = 148.2676$ with $df = 6 - 3 = 3$ and P-value < 0.0001.

(d) Bhapka's test statistic is 144.25 wih $df = 3$,p-value< 0.0001. Also, the LR staistic for baseline category logit model is 155.34 with $df = 3$, p-value< 0.0001 Hence, We can reject the H_0. This is strong indication of the inadequacy of the symmetry model. There is marginally heterogeneous.

8.14

(a) Symmetry:
$G^2\left(S\right) = 134.4519$, $df = 6$, P-value < 0.0001
The Symmetry model provides a very poor fit.
Quasi-Symmetry :
$G^2\left(QS\right) = 3.9324$, $df = 3$, P-value $= 0.2688568$
The Quasi-Symmetry model provides an adequate fit.

(b) H_0 : Symmetry $v.s.$ H_a : Quasi symmetry
$G^2\left(S|QS\right) = 134.4519 - 3.39324 = 130.52$ with $df = 6 - 3 = 3$, P-value < 0.0001

We reject H_0. This is strong indication of the inadequacy of the symmetry model. There is marginally heterogeneous.

8.16

For the symmetry model, $G^2 = 445.2268$ with $df = 3$ and $p-value < 0.001$, indicating that the symmetry model does not fit the data well.
For the quasi-symmetry model, $G^2 = 1.2266$ with $df = 1$ and $p-value = 0.268$, indicating that the quasi-symmetry model fits the data well.
$G^2(S) - G^2(QS) = 444.0002$,$df = 2$,$p-value < 0.0001$. Hence we reject the marginal homogeneity. For the ordinal quasi-symmetry model, $G^2 = 2.4688$ with $df = 2$ and $p-value = 0.2910$, indicating that the ordinal quasi-symmetry model fits the data well.

8.17

Test fit of symmetry model :
$G^2_S = 40.5765$, $df = 6$, P-value < 0.0001
The Symmetry model provides a very poor fit

Test fit of quasi-symmetry model:
$G^2_{QS} = 2.2273$, $df = 3$, P-value $= 0.5265912$
The quasi-symmetry model provides a very good fit

$G^2_S - G^2_{QS} = 40.5765 - 2.2273 = 38.3492$, $df = 6 - 3 = 3$, P-value < 0.0001
reject marginal homogeneity.

Test fit of ordinal quasi-symmetry model $G^2_{OQS} = 27.4894$, $df = 5$, P-value < 0.0001
The ordinal quasi-symmetry model provides a very poor fit
The quasi-symmetry model seems to be the optimal model among the tree models.
The population odds of belief that car pollution is extremely dangerous to the environment are estimated to be $e^{1.2070} = 3.34$ times the population odds of belief that greenhouse gases are extremely dangerous to the environment. And thus 95 % CI for odds ratio is $(e^{0.7301}, e^{1.6838}) = (2.075, 5.386)$

8.19

Independence model:
for independent model $G^2 = 4167.6348$, $df = 9$ The model does not fit well.

Quasi-independence (QI) model:
for quasi-independence model $G^2 = 9,7032$, $df = 5$. This model fits better.
There is an additional term, $\delta_i I\,(i = j)$, in the quasi-independence model.
It treats the main diagonal differently from the rest of the table.

8.20

(a) Independence model:
for independent model $G^2 = 69.1626$ with $df = 9$ and P-value < 0.0001
The model does not fit well.
The standard residuals are about 4 in two of the main diagonal cells,while they are small in the quasi-independence model. Thus, the quasi-independence model fits cells on main diagonal perfectly.

(b) Fit with Quasi-Symmetric (QS) model:
$G^2 = 6.1840$ with $df = 3$ and P-value $= 0.1029934$
The model fits adequately.
The standard residuals are smaller than those in the independent model. Thus, the Quasi-Symmetric model fits better.

(c) The weighted kappa is an appropriate measure since the categories are ordinal. This value of weighted kappa is 0.3797 indicates that the two neurologists agree more frequently than expected under independence, but the agreement is not particularly strong.

8.22

(a) Bradley-Terry model :
$logit\,(\pi_{ij}) = \beta_i - \beta_j$ where $\beta_{coke} = 0.2837$, $\beta_{classic} = -0.2959$, $\beta_{Pepsi} = 0$
Ranking : 1-Coke, 2- Pepsi, 3-Classic Coke

(b) $\pi_{Coke,Pepsi} = \frac{\exp(0.2837)}{\exp(0.2837)+\exp(0)} = 0.5704531$
The estimated probability is 0.5704531
Sample proportion is $\frac{29}{49} = 0.5918367$. They are very close.

8.23

(a) $G^2 = 4.2934$ with $df = 3$ and P-value $= 0.2313886$
The model fits well
Prestige ranking: 1-JSSR-B, 2-Biometrika, 3-JASA, 4-Commun Statist

(b) $\beta_{CO} = -3.2180$, $\beta_{JR} = 0$
$\pi_{JE,CO} = \frac{\exp(0)}{\exp(0)+\exp(-3.2181)} = 0.9615098$
The estimated probability is 0.9615

Only for students having taken STAT 414,610 or 630

8.27

(a)

$$\log\left(\frac{\pi_{ij}}{\pi_{ji}}\right) = \log\left(\frac{\mu_{ij}}{\mu_{ji}}\right) = \log(\mu_{ij}) - \log(\mu_{ji})$$
$$= \left(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}\right) - \left(\lambda + \lambda_j^X + \lambda_i^Y + \lambda_{ji}\right) = \left(\lambda_i^X - \lambda_i^Y\right) - \left(\lambda_j^X - \lambda_j^Y\right)$$

Take $\beta_i = \left(\lambda_i^X - \lambda_i^Y\right)$ then $\log\left(\frac{\pi_{ij}}{\pi_{ji}}\right) = \beta_i - \beta j$ for al i, j

(b) Quasi-symmetric: $\lambda_i^X = \lambda_i^Y = \lambda_i$ for all i , then $\log\left(\mu_{ij}\right) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij} = \lambda + \lambda_i + \lambda_j + \lambda_{ij} = \log\left(\mu_{ji}\right)$
that is $\mu_{ij} = \mu_{ji}$, the symmetric model.

(c) If $\lambda_{ij} = 0$ for $i \neq j$, then model adds to independence model a term for each cell on main diagonal. That is, $\log\left(\mu_{ij}\right) = \lambda + \lambda_i^X + \lambda_j^Y + \delta I\left(i = j\right)$, a Quasi-independence model.