STA841 HW5

Dipesh Gautam

December 3, 2015

1 Problem 1, Agresti 11.12

In this problem we look at matched pairs of reviews by two reviewers and the reviews are categorized con, mixed or pro.

1.1 1a

We fit the symmetry, quasi-independence and quasi-symmetry model summary of each is given below.

```
Call:
glm(formula = count ~ sym, family = poisson(log), data = movies)

Deviance Residuals:
      1        2        3        4        5        6        7        8
 0.0000   0.0000   0.4332   0.0000   0.0000   0.3112  -0.4525  -0.3217
      9
 0.0000

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.1781     0.2041  15.569  < 2e-16 ***
sym1,2       -1.0986     0.3227  -3.404 0.000664 ***
sym1,3       -0.7357     0.2918  -2.521 0.011692 *
sym2,2       -0.6131     0.3444  -1.780 0.075015 .
sym2,3       -0.8755     0.3028  -2.892 0.003833 **
sym3,3        0.9808     0.2394   4.098 4.17e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 102.21480  on 8  degrees of freedom
Residual deviance:   0.59276  on 3  degrees of freedom
AIC: 52.784

Number of Fisher Scoring iterations: 4

Call:
glm(formula = count ~ sym + ebert, family = poisson(log), data = movies)

Deviance Residuals:
       1          2          3          4          5          6          7          8
 0.00000   -0.03454    0.02727    0.03482    0.00000   -0.02949   -0.03092    0.03281
```

```
       9
 0.00000

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.17805    0.20412  15.569  < 2e-16 ***
sym1,2      -1.11095    0.37783  -2.940  0.00328 **
sym1,3      -0.86571    0.35370  -2.448  0.01438 *
sym2,2      -0.63763    0.51888  -1.229  0.21913
sym2,3      -1.01631    0.44190  -2.300  0.02146 *
sym3,3       0.73580    0.42928   1.714  0.08652 .
ebert2       0.02452    0.38813   0.063  0.94962
ebert3       0.24503    0.35636   0.688  0.49171
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1.0221e+02  on 8  degrees of freedom
Residual deviance: 6.0515e-03  on 1  degrees of freedom
AIC: 56.198

Number of Fisher Scoring iterations: 3

Call:
glm(formula = count ~ siskel + ebert + D1 + D2 + D3, family = poisson(log),
    data = movies)

Deviance Residuals:
       1         2         3         4         5         6         7         8
 0.00000  -0.03454   0.02727   0.03482   0.00000  -0.02949  -0.03092   0.03281
       9
 0.00000

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.21771    0.39572   5.604 2.09e-08 ***
siskel2     -0.15060    0.35064  -0.430  0.66755
siskel3      0.09464    0.38394   0.246  0.80530
ebert2      -0.12608    0.37462  -0.337  0.73645
ebert3       0.33967    0.37687   0.901  0.36744
D1           0.96035    0.44527   2.157  0.03102 *
D2           0.62392    0.48302   1.292  0.19645
D3           1.50687    0.41578   3.624  0.00029 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1.0221e+02  on 8  degrees of freedom
Residual deviance: 6.0515e-03  on 1  degrees of freedom
AIC: 56.198

Number of Fisher Scoring iterations: 3
```

We see that symmetry model has residual deviance of 0.593 on 3 degrees of freedom whereas quasi-symmetry and quasi-independence both have residual deviance of 0.006 on 1 degree of freedom. So, we can conclude that both the quasi-symmetry and quasi-indepence model fit much better than the symmetry model.

1.2 1b

```
> a = anova(fit.sym, fit.qsym)
> dev = round(a$Deviance[2],3)
> df = a$Df[2]
> dev

[1] 0.587

> df

[1] 2
```

We see that the deviance between symmetry and quasi-symmetry or quasi-independence model is 0.587 on 2 degrees of freedom. So, we can see the evidence of marginal homogeneity.

1.3 1c

```
> data.tab = xtabs(count~siskel+ebert, data = movies)
> kappa = Kappa(data.tab, weights = "Equal-Spacing")
> kappa.unweighted = kappa$Unweighted[1]
> kappa.unweighted.se = kappa$Unweighted[2]
> kappa.weighted = kappa$Weighted[1]
> kappa.weighted.se = kappa$Weighted[2]
> kappa.unweighted

    value
0.3888385

> kappa.unweighted.se

       ASE
0.05979313

> kappa.weighted

   value
0.426874

> kappa.weighted.se

       ASE
0.06349523
```

2 Problem 2, Agresti 11.17

We fit the Bradley-Terry model on the frequency of journal citations among four statistics journals in order to infer the prestige rankings among the journals. the summary of the fit is given below.

```
Call:
glm(formula = y ~ sym + cited, family = poisson(log), data = journal)

Deviance Residuals:
       1         2         3         4         5         6         7         8
 0.00000  -0.82716   0.40752  -0.13329   0.18483   0.00000   0.02593  -0.34205
       9        10        11        12        13        14        15        16
-0.32264  -0.08931   0.00000   0.44598   0.15202   1.58360  -0.65702   0.00000

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      6.57088    0.03742 175.579  < 2e-16 ***
sym1,2           0.01531    0.05232   0.293 0.769762
sym1,3          -0.34586    0.05621  -6.153 7.61e-10 ***
sym1,4          -1.18296    0.07066 -16.742  < 2e-16 ***
sym2,2           2.43028    0.11945  20.345  < 2e-16 ***
sym2,3           0.60851    0.07776   7.825 5.08e-15 ***
sym2,4          -1.19892    0.09785 -12.252  < 2e-16 ***
sym3,3           0.88597    0.07749  11.434  < 2e-16 ***
sym3,4          -1.08085    0.08347 -12.950  < 2e-16 ***
sym4,4          -1.60340    0.10834 -14.800  < 2e-16 ***
citedcommunstat -2.94907    0.10255 -28.759  < 2e-16 ***
citedjasa       -0.47957    0.06059  -7.915 2.47e-15 ***
citedjrss-b      0.26895    0.07083   3.797 0.000146 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 3748.0412  on 15  degrees of freedom
Residual deviance:    4.2934  on  3  degrees of freedom
AIC: 147.79

Number of Fisher Scoring iterations: 4
```

We get residual devianc of 4.293 on 3 degrees of freedom. So, we are fairly satisfied with the model fit. We have the Biometrika parameter as 0, Communications in statistics as -2.949, JASA as -0.48 and JRSS-B as 0.269. So the prestige ranking is:

1. JRSS-B
2. Biometrika
3. JASA
4. Communications in Statistics

For citations involving Communications in Statistics and JRSS-B, the probability that Commun. Stat cites JRSS-B article is 0.962

3 Problem 3, Remaining from HW4

We fit the various model to thest the given hypotheses. To test hypothesis 1, we fit the log-linear intercept model. Similarly fo the column totals and row totals we, added an extra covariate, column(or row) factor. Finally fo the one involving the difference between numbers < 45 or ≥ 45, we added an indicator covariate. The summaries of all these fits are below:

```
Call:
glm(formula = count ~ 1, family = poisson(link = log), data = pick6,
    subset = (tot > 0), offset = log(tot))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4592  -0.7834  -0.3313   0.7778   1.9289

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.98898    0.05661  -70.46   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.787  on 53  degrees of freedom
Residual deviance: 61.787  on 53  degrees of freedom
AIC: 253.49

Number of Fisher Scoring iterations: 4

Call:
glm(formula = count ~ 1 + column, family = poisson(link = log),
    data = pick6, subset = (tot > 0), offset = log(tot))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8788  -0.8663  -0.4205   0.8264   2.0294

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.951e+00  1.826e-01 -21.642   <2e-16 ***
column1      5.407e-02  2.442e-01   0.221    0.825
column2     -1.823e-01  2.582e-01  -0.706    0.480
column3      5.407e-02  2.442e-01   0.221    0.825
column4      1.777e-01  2.379e-01   0.747    0.455
column5      1.257e-15  2.582e-01   0.000    1.000
column6      6.852e-16  2.582e-01   0.000    1.000
column7     -1.054e-01  2.653e-01  -0.397    0.691
column8     -3.567e-01  2.845e-01  -1.254    0.210
column9     -1.823e-01  2.708e-01  -0.673    0.501
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.787  on 53  degrees of freedom
Residual deviance: 55.463  on 44  degrees of freedom
AIC: 265.16

Number of Fisher Scoring iterations: 5

Call:
glm(formula = count ~ 1 + row, family = poisson(link = log),
```

```
    data = pick6, subset = (tot > 0), offset = log(tot))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.81726  -0.56806   0.03798   0.68657   2.01668

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.82935    0.12804 -29.908   <2e-16 ***
row10       -0.45503    0.19912  -2.285   0.0223 *
row20        0.01787    0.17574   0.102   0.9190
row30        0.01787    0.17574   0.102   0.9190
row40       -0.32441    0.19184  -1.691   0.0908 .
row50       -0.47856    0.25301  -1.892   0.0586 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.787  on 53  degrees of freedom
Residual deviance: 48.270  on 48  degrees of freedom
AIC: 249.97

Number of Fisher Scoring iterations: 4

Call:
glm(formula = count ~ 1 + geq45, family = poisson(link = log),
    data = pick6, subset = (tot > 0), offset = log(tot))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.59107  -0.49025   0.02068   0.67539   1.74590

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.4139     0.1715 -25.737  < 2e-16 ***
geq45         0.4918     0.1817   2.707  0.00679 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.787  on 53  degrees of freedom
Residual deviance: 53.461  on 52  degrees of freedom
AIC: 247.16

Number of Fisher Scoring iterations: 4
```

We can conclude that variation of frequencies, row totals are not consistent with hypothesis of uniformity. Especially for the row totals, 10+ and 40+ and 50+ have lower chance of getting picked. Also that frequency of occurence of numbers 45-54 is not the same as the remaining numbers. Variation of column totals seem to be consistent with the hypothesis of uniformity.
With the additional information that in the first 24 weeks only numbers 1-44 were drawn and in the last 52 weeks nubmers were drawn from 1-54, we adjusted the total available for each number from the same 312 for all to 312 for numbers 1-44 and 168 for numbers 45-54. As the numbers 45-54 had only 168 chances to

be picked instead of 312 for the other numbers.
Summaries for fits to test hypotheses 1 and 2 again under this information is given below:

```
Call:
glm(formula = count ~ 1, family = poisson(link = log), data = pick6,
    subset = (tot > 0), offset = log(tot1))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.63605  -0.54429   0.06963   0.64246   1.68380

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.89964    0.05661  -68.88   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 51.845  on 53  degrees of freedom
Residual deviance: 51.845  on 53  degrees of freedom
AIC: 243.55

Number of Fisher Scoring iterations: 4

Call:
glm(formula = count ~ 1 + column, family = poisson(link = log),
    data = pick6, subset = (tot > 0), offset = log(tot1))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0470  -0.7469   0.1816   0.6355   1.7971

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.854e+00  1.826e-01 -21.111   <2e-16 ***
column1      3.726e-02  2.442e-01   0.153    0.879
column2     -1.991e-01  2.582e-01  -0.771    0.441
column3      3.726e-02  2.442e-01   0.153    0.879
column4      1.609e-01  2.379e-01   0.676    0.499
column5      8.381e-16  2.582e-01   0.000    1.000
column6      7.313e-16  2.582e-01   0.000    1.000
column7     -1.054e-01  2.653e-01  -0.397    0.691
column8     -3.567e-01  2.845e-01  -1.254    0.210
column9     -1.823e-01  2.708e-01  -0.673    0.501
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 51.845  on 53  degrees of freedom
Residual deviance: 45.847  on 44  degrees of freedom
AIC: 255.55

Number of Fisher Scoring iterations: 5
```

Still, the variation of frequency is not consistent with the hypothesis of uniformity. Variation of column totals still is.