

8.2

- a) We reject the null hypothesis and conclude that the marginal probabilities of answering Yes are different. The 95% confidence interval for the difference in the two proportions is:

$$N = 833 + 125 + 2 + 160$$

$$p_{.1p} = (833 + 125) / N; \quad p_{.p1} = (833 + 2) / N$$

$$p_{.1p} - p_{.p1} + c(-1, 1) * 1.96 * \sqrt{((125 + 2) - ((125 - 2)^2 / N)) / N^2}$$

$$[1] \quad 0.09117822 \quad 0.12846463$$

- b) 90% confidence interval for the difference in proportions

$$p_{.1p} - p_{.p1} + c(-1, 1) * 1.645 * \sqrt{((125 + 2) - ((125 - 2)^2 / N)) / N^2}$$

$$[1] \quad 0.09417445 \quad 0.12546841$$

8.3

$$\text{logit}[P(Y_2 = 1)] = \log\left(\frac{(835/1120)}{1 - (835/1120)}\right) = 1.075$$

$$\beta = \log\left(\frac{(958/162)}{(835/285)}\right) = .702$$

$$\text{Marginal Model} = 1.075 + .702x$$

- a) The odds that someone believes in both heaven and hell is $\exp(.702) = 2.02$ times greater than someone who does not believe in heaven and hell
- b) The odds that someone believes in hell conditional that they believe in heaven is $\exp(1.075 + .702) = 5.91$ times the odds that someone believes in heaven, conditional that they believe in hell

8.5

- a) If the probabilities of “No” for both questions is the same then probability of a response being a No in one question, but not in the other is .5.
- b) The mid pvalue is 1/2 the probability of one more “No” response plus the probability of “No” for the same question.
- c) The binomial distribution is discrete and so it cannot calculate exact probabilities. The Mid pvalue is used as a correction to try to get closer to the exact pvalue.

8.6

- a) We reject the null hypothesis and conclude that there is a significant difference in proportions

```
mcnemar.test(x = matrix(c(114, 11, 181, 48), nrow = 2))
```

McNemar's Chi-squared test with continuity correction

```
data: matrix(c(114, 11, 181, 48), nrow = 2)
McNemar's chi-squared = 148.76, df = 1, p-value < 2.2e-16
```

- b) The 90% confidence interval for difference in proportions of a female opposing info and opposing health opinions is between .43 and .53

```
N = 114 + 181 + 11 + 48
```

```
p.1p = (114 + 181) / N; p.p1 = (114 + 11) / N
```

```
p.1p - p.p1 + c(-1, 1) * 1.645 * sqrt( ((181 + 11) - ((181 - 11)^2 / N)) / N^2)
```

```
[1] 0.4314089 0.5290430
```

- c) The odds ratios for Supporting Info Opinion is $\exp(2.214) = 9.152$ times higher than Supporting Health Opinion. The conditional odds ratio of Supporting Info Opinion is $\exp(-.605 + 2.214) = 4.99$ times higher than Supporting Health Opinion.

$$\text{logit}[P(Y_2 = 1)] = \log\left(\frac{125/354}{1 - (125/354)}\right) = -.605$$

$$\beta = \log\left(\frac{295/59}{125/229}\right) = 2.214$$

$$\text{Marginal Model} = -.605 + 2.214x$$

- d) You could create a new 2x2 table by taking the difference in the counts between Males and Females for each question.

8.7

```
(.314 - .292) + c(-1, 1) * 1.96 * sqrt( ((359 + 334) - ((359 - 334)^2 / 1144)) / 1144^2 )
= (-.023, .067)
```

The interval is wider than the interval in the book. The estimate could gain higher precision by adding more observations.

8.11

The odds of someone answering yes to cuts in paying higher taxes is .95 to 1.59 times that of those who answer yes to cut in living standard. Because the 95% confidence interval crosses 1, the odds ratio of the two questions are not significantly different.

$$\log(132/107) \pm 1.96 * \sqrt{(1/132 + 1/107)} \rightarrow \exp(-.0449, .4649) = [.956, 1.59]$$

8.13

- a) $G^2 = 150.6 > 79 = qchisq(.95, 6)$ Reject the null hypothesis that there is symmetry in the responses
- b) $G^2 = 2.3 < 7.8 = qchisq(.95, 3)$ Do not reject the null hypothesis there is quasi symmetry
- c) $G^2 = 150.6 - 2.3 = 148.3 > 7.8 = qchisq(.95, 3)$ Reject the null hypothesis that there is marginal homogeneity
- d)

8.14

a)

Call:

```
glm(formula = count ~ symm1 + symm2 + symm3 + symm4 + symm5 +
     symm6 + symm7, family = poisson(link = log))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-13.4108	-7.2282	-0.4995	1.3303	24.3118

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.70410	0.04255	110.558	< 2e-16 ***
symm1	1.34799	0.06452	20.891	< 2e-16 ***
symm2	0.80744	0.03005	26.869	< 2e-16 ***
symm3	0.64786	0.01858	34.866	< 2e-16 ***
symm4	-0.52535	0.04927	-10.662	< 2e-16 ***
symm5	-0.14302	0.02104	-6.797	1.07e-11 ***
symm6	-0.15522	0.01922	-8.074	6.79e-16 ***
symm7	-0.19120	0.02040	-9.374	< 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: 4481.8 on 15 degrees of freedom
 Residual deviance: 1114.2 on 8 degrees of freedom
 AIC: 1222.7

Number of Fisher Scoring iterations: 6

Call:

```
glm(formula = count ~ r.16 + r.2004 + symm1 + symm2 + symm3 +
     symm4 + symm5 + symm6 + symm7, family = poisson(link = log))
```

Deviance Residuals:

1	2	3	4	5	6	7
0.00000	-0.70846	0.40207	-0.07500	1.10133	0.00000	-0.35440
8	9	10	11	12	13	14
0.00000	-1.16275	0.54702	0.00000	0.07902	0.20851	0.00000

	15	16
-0.08348	0.00000	

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.18811	5.65209	-0.387	0.6987
r.16ne	1.12626	2.92019	0.386	0.6997
r.16s	1.01299	2.87508	0.352	0.7246
r.16w	3.64395	2.92959	1.244	0.2136
r.2004ne	0.05887	2.84044	0.021	0.9835
r.2004s	1.92691	2.79125	0.690	0.4900
r.2004w	4.65784	2.72725	1.708	0.0877 .
symm1	7.05506	0.39501	17.860	<2e-16 ***
symm2	4.25354	2.82613	1.505	0.1323
symm3	1.96530	0.08702	22.583	<2e-16 ***
symm4	1.01553	0.69309	1.465	0.1429
symm5	0.92124	0.56922	1.618	0.1056
symm6	0.57861	0.04195	13.792	<2e-16 ***
symm7	0.19720	0.42515	0.464	0.6428

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4481.7873 on 15 degrees of freedom
 Residual deviance: 3.7156 on 2 degrees of freedom
 AIC: 124.22

Number of Fisher Scoring iterations: 4

b) $G^2 = 1114 - 125 = 989 > 12.59 = qchisq(.95, 6)$ Reject Null and conclude that there marginal homogeneity

8.16

There is evidence of symmetry $G^2 = 364 > 11.07 = qchisq(.95, 5)$. There is also strong evidence of quasi symmetry with $G^2 = 807 > 14 = qchisq(.95, 7)$. Marginal homogeneity is also present with $G^2 = 445 - 1.22 = 443.98 > 6 = qchisq(.95, 2)$.

Call:

```
glm(formula = count ~ symm1 + symm2 + symm3 + symm4 + symm5,
     family = poisson(link = log))
```

Deviance Residuals:

1	2	3	4	5	6	7	8
0.000	7.396	7.416	-9.608	0.000	6.689	-11.295	-8.419
9							
0.000							

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.33729	0.08084	53.651	< 2e-16 ***
symm1	-0.14764	0.14727	-1.003	0.31609
symm2	0.77302	0.04827	16.013	< 2e-16 ***
symm3	0.11495	0.04189	2.744	0.00607 **
symm4	0.13826	0.02537	5.451	5.02e-08 ***
symm5	0.10910	0.02032	5.369	7.92e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 809.19 on 8 degrees of freedom
 Residual deviance: 445.23 on 3 degrees of freedom
 AIC: 513.04

Number of Fisher Scoring iterations: 5

Call:

```
glm(formula = count ~ Chem + Recycle + symm1 + symm2 + symm3 +
     symm4 + symm5, family = poisson(link = log))
```

Deviance Residuals:

1	2	3	4	5	6	7	8
0.0000	-0.1346	0.1669	0.3316	0.0000	-0.1380	-0.9812	0.2979
9							
0.0000							

Coefficients: (2 not defined because of singularities)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.22231	0.18109	17.794	< 2e-16 ***
Chem2	0.38767	0.10419	3.721	0.000199 ***
Chem3	-1.60598	0.15339	-10.470	< 2e-16 ***
Recycle2	2.21156	0.18925	11.686	< 2e-16 ***
Recycle3	1.77467	0.16684	10.637	< 2e-16 ***

symm1	0.96735	0.21897	4.418	9.97e-06	***
symm2	0.03089	0.06625	0.466	0.641076	
symm3	0.43038	0.06501	6.621	3.58e-11	***
symm4	NA	NA	NA	NA	
symm5	NA	NA	NA	NA	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 809.1941 on 8 degrees of freedom
 Residual deviance: 1.2266 on 1 degrees of freedom
 AIC: 73.037

Number of Fisher Scoring iterations: 4

Call:

```
glm(formula = count ~ Chem * Recycle + symm1 + symm2 + symm3 +
     symm4 + symm5, family = poisson(link = log))
```

Deviance Residuals:

[1] 0 0 0 0 0 0 0 0 0 0

Coefficients: (5 not defined because of singularities)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.29892	0.19219	17.165	< 2e-16	***
Chem2	0.36464	0.10628	3.431	0.000602	***
Chem3	-2.20031	0.60850	-3.616	0.000299	***
Recycle2	2.12603	0.20333	10.456	< 2e-16	***
Recycle3	1.71172	0.17398	9.838	< 2e-16	***
symm1	0.89074	0.22823	3.903	9.51e-05	***
symm2	0.04686	0.06798	0.689	0.490602	
symm3	0.62393	0.20354	3.065	0.002174	**
symm4	NA	NA	NA	NA	
symm5	NA	NA	NA	NA	
Chem2:Recycle2	NA	NA	NA	NA	
Chem3:Recycle2	0.64656	0.62890	1.028	0.303910	
Chem2:Recycle3	NA	NA	NA	NA	
Chem3:Recycle3	NA	NA	NA	NA	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 8.0919e+02 on 8 degrees of freedom
 Residual deviance: 6.5059e-14 on 0 degrees of freedom
 AIC: 73.81

Number of Fisher Scoring iterations: 3

8.17

Deviance for the symmetry model is 40.5 on 6 degrees of freedom. Deviance from the quasi-symmetry model is 27.4 on 5 degrees of freedom $G^2 = 40.5 - 27.4 = 13.1 > 3.84 = qchisq(.95, 1)$ Reject the null and conclude that the symmetry model is not an adequate fit.

Output from the symmetry model

Parameter	DF	Estimate	SE	Wald 95%	CL	Chi-Square	Pr > ChiSq
Intercept	1	3.2581	0.1961	2.8737	3.6425	276	<.0001
symm1	1	1.2958	0.2213	0.862	1.7296	34.27	<.0001
symm2	1	0.976	0.2138	0.557	1.395	20.84	<.0001
symm3	1	0.1919	0.2331	-0.265	0.6488	0.68	0.4104
symm4	1	-1.3863	0.3397	-2.0521	-0.7205	16.66	<.0001
symm5	1	1.6017	0.215	1.1804	2.0231	55.51	<.0001
symm6	1	1.4287	0.2075	1.0219	1.8354	47.39	<.0001
symm7	1	-1.118	0.3119	-1.7294	-0.5067	12.85	0.0003
symm8	1	2.1929	0.2068	1.7877	2.5982	112.48	<.0001
symm9	1	0.7122	0.2189	0.2833	1.1411	10.59	0.0011
symm10	0	0	0	0	0	.	.

8.19

Deviance for the independence model 4188 and deviance for the quasi-independence model is 3.7. There is a large improvement in the model fit because the quasi-independence model uses an identity matrix to weight the observations in the trace.

```
Call: glm(formula = count ~ r.16 + r.2004, family = poisson(link = log))
```

Coefficients:

(Intercept)	r.16ne	r.16s	r.16w	r.2004ne	r.2004s	r.2004w
5.1016	-0.2065	0.2085	-0.3163	-0.3272	0.4309	-0.0876

Degrees of Freedom: 15 Total (i.e. Null); 9 Residual

Null Deviance: 4482

Residual Deviance: 4168 AIC: 4274

```
Call: glm(formula = count ~ r.16 + r.2004 + mob.symint(r.16, r.2004) +
  mob.qi(r.16, r.2004), family = poisson(link = log))
```

Coefficients:

(Intercept)	r.16ne	r.16s
6.319	-3.329	-2.917
r.16w	r.2004ne	r.2004s
-3.604	-4.355	-1.963
r.2004w	mob.symint(r.16, r.2004)2:2	mob.symint(r.16, r.2004)2:3
-2.493	3.709	3.310
mob.symint(r.16, r.2004)2:4	mob.symint(r.16, r.2004)3:3	mob.symint(r.16, r.2004)3:4
3.105	2.604	2.600
mob.symint(r.16, r.2004)4:4	mob.qi(r.16, r.2004)2	mob.qi(r.16, r.2004)3
2.946	NA	NA

```

mob.qi(r.16, r.2004)4      mob.qi(r.16, r.2004)5
                        NA                        NA

```

```

Degrees of Freedom: 15 Total (i.e. Null); 3 Residual
Null Deviance:      4482
Residual Deviance: 3.932    AIC: 122.4

```

8.20

The quasi symmetry model fits much better than the independence model given that the residual deviance is much lower. The kappa coefficient indicates that the difference between the agreements is .21 that of the maximum possible difference under the independence model. We conclude that there is significant differences in the agreements of the two neurologists.

Call:

```

glm(formula = C ~ A + B, family = poisson, data = dta)

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-3.3392  -2.2512   0.1611   1.2448   3.4648

```

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.21106     0.16709  19.217 < 2e-16 ***
A2           -0.81990     0.19731  -4.155 3.25e-05 ***
A3           -2.03292     0.32065  -6.340 2.30e-10 ***
A4           -1.59760     0.26595  -6.007 1.89e-09 ***
B2             0.06596     0.20977   0.314  0.7532
B3           -0.22884     0.22649  -1.010  0.3123
B4           -0.64870     0.25730  -2.521  0.0117 *
---

```

```

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

Call:

```

glm(formula = C ~ A + B + mob.symint(A, B), family = poisson,
    data = dta)

```

Deviance Residuals:

```

      1      2      3      4      5      6      7      8      9     10
0.00000 0.10284 0.08135 -0.44703 -0.25299 0.00000 -0.19933 0.54573 -0.71421 0.48189
     11     12     13     14     15     16
0.00000 -0.27616 1.41749 -1.64094 0.21185 0.00000

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6376	0.1622	22.424	< 2e-16 ***
A2	-1.9171	0.4413	-4.344	1.40e-05 ***
A3	-5.0039	0.7683	-6.513	7.35e-11 ***
A4	-5.4845	0.9021	-6.079	1.21e-09 ***
B2	-0.1590	0.2384	-0.667	0.504792
B3	-1.3608	0.3559	-3.824	0.000131 ***
B4	-2.2915	0.5265	-4.352	1.35e-05 ***
mob.symint(A, B)2:2	0.4182	0.2828	1.479	0.139136
mob.symint(A, B)2:3	2.3322	0.5874	3.970	7.18e-05 ***
mob.symint(A, B)2:4	2.3033	0.7610	3.027	0.002473 **
mob.symint(A, B)3:3	2.1683	0.5163	4.200	2.67e-05 ***
mob.symint(A, B)3:4	4.9118	0.9455	5.195	2.05e-07 ***
mob.symint(A, B)4:4	3.2205	0.6365	5.060	4.20e-07 ***

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: 6.184 on 3 degrees of freedom
AIC: 85.459

Number of Fisher Scoring iterations: 6

Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha)

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

8.22

```
library(BradleyTerry2)

dta = data.frame(
  win.1 = c(20, 19, 31),
  win.2 = c(29, 28, 19),
  player.1 = c("Pepsi", "Classic Coke", "Coke"),
  player.2 = c("Coke", "Pepsi", "Classic Coke")
)

mdl = BTm(cbind(dta$win.1, dta$win.2), player.1, player.2, id = "drink", data = dta)
summary(mdl)
```

Call:

```
BTm(outcome = cbind(dta$win.1, dta$win.2), player1 = player.1,
    player2 = player.2, id = "drink", data = dta)
```

Deviance Residuals:

```
      1      2      3
-0.3032 -0.3099 -0.3078
```

Coefficients:

```
      Estimate Std. Error z value Pr(>|z|)
drinkCoke    0.5796     0.2399   2.416  0.0157 *
drinkPepsi    0.2959     0.2399   1.233  0.2175
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 6.30488  on 3  degrees of freedom
Residual deviance: 0.28271  on 1  degrees of freedom
AIC: 17.193
```

Number of Fisher Scoring iterations: 3

```
fitted(mdl)
```

```
      1      2      3
0.4295572 0.4265597 0.6409661
```

a) Rankings: Coke, Pepsi, Classic Coke

b)

Comparison	Predicted	Sample
Coke vs Pepsi	.57	.592

8.23

	citing			
cited	Biometrika	Comm Statist	JASA	JRSS-B
Biometrika	714	730	498	221
Comm Statist	33	425	68	17
JASA	320	813	1072	142
JRSS-B	284	276	325	188

Call:

```
BTm(outcome = cbind(win1, win2), player1 = journal1, player2 = journal2,
    data = citations.sf)
```

Deviance Residuals:

1	2	3	4	5	6
0.8476	-0.5198	0.2022	-0.0930	1.6201	-0.7941

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
..Comm Statist	-2.94907	0.10255	-28.759	< 2e-16 ***
..JASA	-0.47957	0.06059	-7.915	2.47e-15 ***
..JRSS-B	0.26895	0.07083	3.797	0.000146 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1925.2329 on 6 degrees of freedom
 Residual deviance: 4.2934 on 3 degrees of freedom
 AIC: 46.394

Number of Fisher Scoring iterations: 4

1	2	3	4	5	6
0.95021963	0.61764628	0.43316389	0.07802400	0.03849296	0.32114304

- a) Prestige Ranking: JRSS-B, Biometrika, JASA, Comm Statist
- b) The Probability that the Comm Stat journal cites the JRSS-B journal is .038

8.27

- a) $\log(\mu_{ij}/\mu_{ji}) = \log(\mu_{ij}) - \log(\mu_{ji}) \rightarrow \exp[\log(\mu_{ij}) - \log(\mu_{ji})] = \mu_i - \mu_j = \beta_i - \beta_j$
- b) $\lambda_i^X = \lambda_j^Y \rightarrow \lambda_i^X - \lambda_j^Y = 0 \rightarrow \log(\mu_{ij}) = \lambda + \lambda_i + \lambda_j + \lambda_{ij}$
- c) $I(i \neq j) = 0 = \lambda_{ij} \rightarrow \log(\mu_{ij}) = \lambda + \lambda_i + \lambda_j + \lambda_{ij}$