Solution for Stat 475 Hw5

March 20, 2018

Chap 4, Problem 4

(a) The average number of cars in a unit time is different at different periods. For example, the intensity (number of cars in a unit time) is higher at rush hours. Thus to overcome this problem, we can collect the data at multiple different time periods

(b) Maybe not, because they are consecutive. Any periods of heavier or lighter traffic (for example coming from lights at neighboring intersections) might affect the number of cars at the intersection for more than one light cycle. We must assume that these things are not happening. Similarly, events that temporarily affect traffic (church/school ending for example) must be assumed not to occur during any time during observation.

(c) $P(Y \geq 9) = 0.0179$. It seems fairly improbable, but will likely happen quite often during a day, given the number of cycles that the light will go through in a day. We also dont know how much the traffic at this light increases

(d) Binomial with 60 trials and probability $p = 0.0179$ yields $P(W \geq 1) = 1 - P(W = 0) = 0.339$

Chap 4, Problem 12

(a) Make the variables to be factors, then fit a model (fit.nominal); make the variables to be numerical (scores) then fit 3 models. From the summary (model) to find the deviances for the 4 models to be

$$3.245, 8.399, 12.95, 63.6$$

Thus the nominal model is the best and the next is score 1:2:3:4:5

```
data <- read.csv("http://www.chrisbilder.com/categorical/Chapter3/pol_ideol_data.csv")
s1 <- (rep(1:5, times = 4))
s2 <- (rep(c(0,2,3,4,6), times = 4))
s3 <- (rep(c(2,1,0,1,2), times = 4))
dataall <- data.frame(data, s1 = s1, s2 = s2, s3 = s3)
fit.nominal <- glm(count ~ (gender + party + ideol)^2,
                   family = poisson(link = "log"), data = dataall)
fit.s1 <- glm(count ~ gender*party + gender*ideol + party*s1,
                   family = poisson(link = "log"), data = dataall)
fit.s2 <- glm(count ~ gender*party + gender*ideol + party*s2,
                   family = poisson(link = "log"), data = dataall)
fit.s3 <- glm(count ~ gender*party + gender*ideol + party*s3,
                   family = poisson(link = "log"), data = dataall)
```

(b)

(Null model) $$\log(\mu_{ijk}) = \beta_0 + \beta_i^G + \beta_j^P + \beta_k^I + \beta_i^G\beta_j^P + \beta_i^G \times \beta_k^I + \beta_j^P s_k^I$$

(model s1) $$\log(\mu_{ijk}) = \beta_0 + \beta_i^G + \beta_j^P + \beta_k^I + \beta_i^G\beta_j^P + \beta_i^G \times s_k^I + \beta_j^P s_k^I$$

```
fit.null <- glm(count ~ (gender + party)^2 + ideol + gender*ideol +
                    party*s1, family = poisson(link = "log"), data = dataall)
fit.s1   <- glm(count ~ (gender + party)^2 + ideol + gender*s1    +
                    party*s1, family = poisson(link = "log"), data = dataall)
anova(fit.s1, fit.null,test = "LR")

Goodness-of-fit tests for the three score sets:
Scores    - 2 log(Lambda)   p-value
1-2-3-4-5    8.44      0.0377
0-2-3-4-6    8.40      0.0336
2-1-0-1-2    3.23      0.3576
```

(c i) GI linear association model

(Model) $$\log(\mu_{ijk}) = \beta_0 + \beta_i^G + \beta_j^P + \beta_k^I + \beta_i^G\beta_j^P + \beta_i^G s_k^I + \beta_j^P s_k^I$$

$$\log(OR_{MF,kl}^{GI}) = (\beta_M^{GI} - \beta_F^{GI})(s_k^I - s_l^I)$$

From the model, $\beta_M^{GI} = 0.2884$ and $\beta_F^{GI} = 0$, $\log(OR_{MF,kl}^{GI}) = \exp(0.2084) = 1.23$ for $k-l = 1$. The estimated odds that a person with a given extremeness score is male are increased by 23% as the extremeness score is raised by 1 unit. The standard error is 0.08454 leading to 95% Wald confidence interval of 1.04 to 1.45.

```
fit.s1   <- glm(count ~ (gender + party)^2 + ideol + gender*s1    +
                    party*s1, family = poisson(link = "log"), data = dataall)
fit.s1$coefficient
 (Intercept)  genderM  partyR  ideolSC   ideolSL   ideolVC    ideolVL
   4.6329   -0.6304    -1.5133   -1.1888    -0.7167    -1.2784   -0.7005

genderM:partyR     genderM:s1      partyR:s1
 0.2884          0.0551          0.4310
```

(c ii) Compare this odds ratio with the odds ratio from each comparison of political ideologies that results in a 1-unit change in score. These are: VC:SC, VC:SL, VL:SC, VL:SL, SC:M, and SL:M. Some of these ORs can be obtained directly from the PolIdeolOrd.R program, but others need to be inverted because they represent the OR for a 1 unit change in score (e.g., M:SL). Odds ratios and confidence intervals are below:
Odds Ratio Estimate Lower CI Upper CI

```
GI VC:SC | M:F 0.88 0.55 1.42
GI VC:SL | M:F 1.22 0.75 1.98
GI VL:SL | M:F 1.15 0.68 1.92
GI VL:SC | M:F 0.83 0.49 1.41
GI SC:M | M:F 1.75 1.15 2.66
GI SL:M | M:F 1.27 0.83 1.92
```

(d) The association between Gender and Political Ideology seems to be reasonably well represented as an ordinal association. This association is statically significant and indicates that males tend to hold more extreme ideological views than females, with a 23% increase in the odds of a male for each 1-unit increase in extremeness of ideology

Chap 4, Problem 30

Fit a loglinear model with variables guild and degradation. We can add association between the variables as ordinal in degradation or as nominal. Trying both and testing the difference in fit of the models (H0: Association is ordinal, Ha: association is not ordinal) gives $2log(\Lambda) = 23.44, p - value = 0.0006$, so the ordinal association model is not a good fit. We will use the nominal association model to summarize the relationship between the variables. There are many ways to do this; here, we present the ratio of mean counts in high vs. low degradation. Fruit guild had capture rates that may be the same in low and high degradation areas. All other guilds are lower than 1:1 odds, indicating fewer captures in high degradation. Assuming equal opportunity for captures in all locations (time, number of nets, etc.), this suggests that all guilds except fruit feeders are affected by habitat degradation. Other summaries involving moderate degradation could be considered, as well as pairwise comparisons of guilds using odds ratios.


```
data <- read.csv("BirdCounts.csv")
model <- glm(count ~ guild * degrad, family = poission(link = "log" ), data = data)
c <- coefficient(summary(model))
exp(c)
```