

STAT 442 Midterm
November 13, 2008

Time: 75 minutes

Aids allowed: One page (2-sided) 8.5"x11" formula sheet, Calculator.

Grading: Question 1 is worth 20 marks, question 2 is worth 10 marks.

Instructions: Please answer all questions on the provided booklet(s). Be sure to record your name on all exam booklets.

1. Researchers conducted a study to investigate potential risk factors for ear infections in children between the ages of 2 and 5. 257 children between the ages of 3 and 5 were included in the study. The researchers visited the families of the children to determine the number of ear infections experienced since they turned two years of age, whether or not the child was cared for at home or in a daycare setting, and the number of smokers in the household. The collected data set was entered in a file with one row for each child with the following variable headings and definitions.

age: Age of child in years at time of interview

earinf: Number of infections experienced by child since turning 2 years of age.

anySmoke: No = no smokers in household, Yes = one or more smokers

smokers: Number of smokers in four categories: 0, 1, 2, 3+ (i.e. 3 or more)

daycare: No = child cared for at home, Yes = child attended daycare

The data was read into a statistical package for analysis by Poisson regression.

a. The researchers wished to fit a model that would provide estimated risks of infection within a one month period. How should the *denominator* or *offset* variable be defined.

b. An initial model including the categorical variables **smokers** and **daycare** was fitted by Poisson regression, with the following results:

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.31869    0.10748 -30.878 < 2e-16 ***
smokers1       0.35107    0.14071   2.495 0.012599 *
smokers2       0.49933    0.13932   3.584 0.000338 ***
smokers3+     0.03768    0.21403   0.176 0.860269
daycareDaycare 0.39936    0.11091   3.601 0.000317 ***
```

```
Null deviance: 343.96 on 256 degrees of freedom
Residual deviance: 315.31 on 252 degrees of freedom
```

Based on the above output, what is the expected number of infections over one month for a child who attends daycare and lives in a house with 2 smokers?

c. Based on the above model, what is the estimated relative risk of ear infection for a child attending day care compared to a child who doesn't. Provide a 95% confidence interval for this estimate.

d. The researcher wished to assess whether using number of smokers provided more accurate risk estimates, compared with just using the dichotomous variable. Conduct a relevant statistical test. Relevant information may be found in part b and/or in the following output.

```
Model 1: glm(formula = earInf ~ offset(denom) + anySmoke + daycare,
             family = poisson)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.3190	0.1075	-30.884	< 2e-16 ***
anySmokeYes	0.3654	0.1162	3.145	0.001662 **
daycareDaycare	0.4000	0.1108	3.609	0.000307 ***

Null deviance: 343.96 on 256 degrees of freedom

Residual deviance: 320.19 on 254 degrees of freedom

```
Model 2: glm(formula = earInf ~ offset(denom) + anySmoke + smokers + daycare,
             family = poisson)
```

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.31869	0.10748	-30.878	< 2e-16 ***
anySmokeYes	0.03768	0.21403	0.176	0.860269
smokers1	0.31339	0.21924	1.429	0.152878
smokers2	0.46165	0.21835	2.114	0.034492 *
smokers3+	NA	NA	NA	NA
daycareDaycare	0.39936	0.11091	3.601	0.000317 ***

Null deviance: 343.96 on 256 degrees of freedom

Residual deviance: 315.31 on 252 degrees of freedom

```
Model 3: glm(formula = earInf ~ offset(denom) + daycare, family = poisson)
```

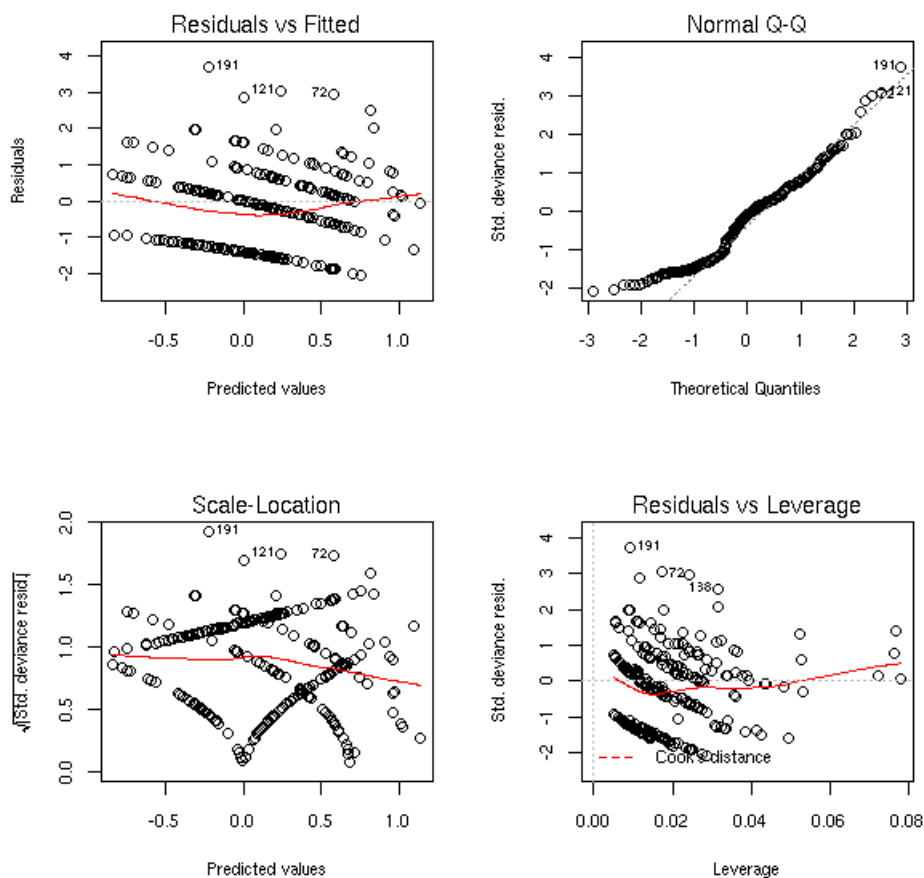
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.10186	0.07785	-39.844	< 2e-16 ***
daycareDaycare	0.40975	0.11078	3.699	0.000217 ***

Null deviance: 343.96 on 256 degrees of freedom

Residual deviance: 330.40 on 255 degrees of freedom

e. The following diagnostic plots were obtained from one of the fitted models. What would you conclude from these plots?



2. A case control study was conducted to examine the risk association between exposure to pets and asthma. 50 asthma sufferers and 50 controls were asked whether their family owned a pet dog or cat while they were growing up. Of the 50 asthma sufferers, 38 had grown up with a dog or cat in the household. For the controls the corresponding number was 12.

a. Test the hypothesis that an association between pet exposure and asthma exists. What do you conclude?

b. Provide an estimate and associated 95% confidence interval for the odds ratio measuring increased risk in those who grew up with pets compared to those who grew up without pets.

c. Suppose probability of having asthma in the population without pets was known to be 20%. Provide an estimate of the probability in those with pets.