Midterm 2 Model Solutions

Q1. a. Since we wish to examine risks over one month, the denominator for each child is the number of months of age past their 2nd birthday.

b. With the offset above, the model predictions provide estimates of expected number of infections in one month. For a child living in a house with 2 smokers and attending daycare, the relevant prediction on the log scale is $\hat{\lambda} = -3.31869 + 0.49933 + 0.39936 = -2.42$, which yields $\hat{\mu} = e^{-2.42} = 0.08892 \approx .09$. The estimated expected number of infections per month is .09.

c. The log-linear regression coefficient for day care is 0.39936 so that the estimated relative risk is $e^{0.39936} \approx 1.49$. A confidence interval on the log scale is $0.39936 \pm 1.96 \times .1101$ which after exponentiation yields an interval of 1.2 to 1.85

d. To test the significance of using number of smokers in relation to the dichotomous variable, we can apply the Likelihood Ratio Test based on the deviances for the model fitted in b) and the deviance for Model 1. The observed statistic is $LRT = 320.19 - 315.31 \approx 4.9$. The diffence in degrees of freedom between the two models is 2, so we refer to χ^2 on 2 degrees of freedom. From tables we have that p-value follows between .05 and .10, so there is no statistically significant evidence for improved predictions using number of smokers.

e. The first residual plot is useful in looking for signs of non-linearity. There is a slight curvature in the scatterplot smooth in the plot. The quantile-quantile plot indicates non-normality of the residuals, but this is not a deviation from assumptions for the Poisson model. The plot of $\sqrt{\text{standardized residuals}}$ shows a slight decreasing trend. The plot of standardized residuals versus the leverage values does not reveal any overly influential observations, since the contours of the Cook's distance don't appear on the plot, indicating that they lie outside the boundaries of the plotting region.

Q2. a. There are various ways to test the significance of the association. Perhaps the simplest is to use Pearson's X^2. Here is the table of observed counts:

```
     asthma
pets  Yes No
  Yes  38 12
  No   12 38
```

The table of expected counts is

```
     asthma
pets  Yes No
  Yes  25 25
  No   25 25
```

The absolute value of observed - expected is 13 in each cell, so that Thus we have $X^2 = \frac{4\times 13^2}{25} \approx 27$ Referring to $\chi^2_{(1)}$ tables yields $P < .001$, indicating that there is strong evidence for association.

b. The estimated odds ratio from the above table is $\hat{OR} = \frac{38\times 38}{12\times 12} = 10.02$ The standard error for the logarith of this estimate is $\sqrt{1/38 + 1/12 + 1/12 + 1/38} = .4683$ yielding 95% confidence interval of $10.02 \times e^{\pm 1.95\times .04683}$ or 4.0 to 25.1.

c. The odds ratio allows us to relate the odds of disease between the groups. If the probability of asthma amongst those without pets is .2, the odds is then .2/.8 or .25. The odds for asthma amongst those whose families had pets is $10.02 \times .25 \approx 2.5$. Solving $\frac{p}{1-p} = 2.5$ yields $p = .71$, so that would be the estimated risk of disease in that group.