Stat 6021: Homework Set 9

1. In an experiment testing the effect of a toxic substance, 1,500 insects were divided at random into six groups of 250 each. The insects in each group were exposed to a fixed dose of the toxic substance. A day later, the insects were observed. Death from exposure was recorded as 1, and survival 0. The file "insects.txt" contains the data from the experiment. The first column denotes the dose level on a logscale, x; the 2nd column denotes the sample size for each group, n_i; and the 3rd column denotes the number of insects that died in the group, y_i. You will probably want to read the data in and give appropriate names for the columns.

   (a) Plot the sample log odds against the log dose, x. Does this plot suggest fitting a logistic regression model to be appropriate?

   (b) Use R to fit the logistic regression model, and write the estimated logistic regression equation.

   (c) Interpret the estimated coefficient $\hat{\beta}_1$ in context.

   (d) What are the estimated odds of death at log dose level $x = 2$?

   (e) Predict the probability of death from exposure at log dose level $x = 2$.

   (f) Carry out both the Pearson's χ^2 and deviance goodness of fit tests to check the fit of this logistic regression model. Clearly state the null and alternative hypotheses, test statistic, and conclusion.

2. (No R needed) A health clinic sent fliers to its clients to encourage everyone to get a flu shot. In a follow-up study, 159 elderly clients were randomly selected and asked if they received a flu shot. A client who received a flu shot was coded $y = 1$, and a client who did not receive a flu shot was coded $y = 0$. Data were also collected on their age, x_1, health awareness rating on a 0-100 scale (higher values indicate greater awareness), x_2, and gender, x_3, where males were coded $x_3 = 1$ and females were coded $x_3 = 0$. A first order logistic regression model was fitted and the output is displayed below.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.17716    2.98242  -0.395  0.69307
age          0.07279    0.03038   2.396  0.01658 *
aware       -0.09899    0.03348  -2.957  0.00311 **
gender       0.43397    0.52179   _____  _______

    Null deviance: 134.94  on 158  degrees of freedom
Residual deviance: 105.09  on 155  degrees of freedom
```

(a) Interpret the estimated coefficient for x_3, gender, in context.

(b) Conduct the Wald test for β_3. State the null and alternative hypotheses, calculate the test statistic, and make a conclusion in context.

(c) Calculate a 95% confidence interval for β_3, and interpret the interval in context.

(d) Comment on whether your conclusions from parts 2b and 2c are consistent.

(e) Suppose you want to drop the coefficients for age and gender, β_1 and β_3. A logistic regression model for just awareness was fitted, and the output is shown below.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.91133    1.62651    3.02  0.00253 **
aware       -0.11931    0.03013   -3.96  7.5e-05 ***

    Null deviance: 134.94  on 158  degrees of freedom
Residual deviance: 113.20  on 157  degrees of freedom
```

Carry out the appropriate hypothesis test to see if the coefficients for age and gender can be dropped.

(f) Based on your conclusion in question 2e, what are the estimated odds of a client receiving the flu shot if the client is 70 years old, has a health awareness rating of 65, and is male? What is the estimated probability of this client receiving the flu shot?

(g) Based on your conclusion in question 2e, interpret the estimated coefficient for age, in context. What does this interpretation tell us about the estimated probability of a client receiving the flu shot if the client is 71 years old, has a health awareness rating of 65, and is male, compared to the estimated probability found for a similar client who is a year younger?