**SPLH 861 HW8: Models for Binary Outcomes (7 points total)  
Due Friday 11/21/2014 by 11:59 PM via Blackboard  
Revision Due by Friday 12/19/14 by 11:59 PM via Blackboard  
  
Please submit all relevant files (word document, syntax, and output)  
using this naming convention: 861\_Firstname\_Lastname\_HW8**

**General Instructions:**

This homework features data from yet another study conducted at Midwestern Red State University, this time examining the extent to which graduate school activities predict subsequent employment opportunities. A total of 500 graduate students reported their number of publications (in print or in press) and their number of conference posters, as evaluated in December of the first year they applied for jobs. Students also reported their amount of advanced quantitative training, which was then categorized into Low (met minimum program requirements), Medium (took 1–2 additional classes), or High (took every quantitative class they possibly could). Whether the student was offered a job in their first year on the market served as the dependent variable. To maintain equivalency across fields, graduate students who did not apply for jobs until after completing a postdoctoral fellowship were not included in the analysis.

Your task is to use maximum likelihood (i.e., using a logit link and Bernoulli response distribution) to estimate two generalized linear models, each predicting a positive job offer binary outcome: an empty model (i.e., with no predictors) and a conditional model 1 that examines the main effects of number of publications, number of conference posters, and level of quantitative training. Because their 0 is already meaningful, do not center number of publications or number of conference posters, but do use two contrasts to represent the group differences in quantitative training. All values should be reported to the nearest .01 to be correct. Use an alpha level of *p* < .05 to denote significant effects.

After answering each set of questions, you will then complete a results section about the model by inserting the answers you found where needed and selecting the correct answer out of the list of possible choices below **[leave the bold text and brackets in please]**:

**[PROGRAM]** = program you used to complete this assignment **[VALUE]** = corresponding value calculated separately  
**[EST]** = corresponding fixed effect estimate (coefficient)  
**[A]** = simple main effect or main effect  
**[B]** = significantly or nonsignificantly   
**[C]** = higher or lower  
**[D]** = larger or smaller  
**[E]** = more publications, more conference posters, either more publications or more conference posters

**Section 1**

Estimate an empty model (for which you should get −2LL = 662.07) to answer these questions.

**Questions:**

1. What is the mean logit of getting a job across the sample?
2. What is the expected probability of getting a job across the sample?

**Results:**

The extent to which graduate school activities could predict employment success was examined in a sample of 500 graduate students from at a Midwestern Red State University. Predictors included number of publications, number of conference posters, and level of advanced quantitative training, including Low (met minimum program requirements; the reference group), Medium (took 1-2 additional classes), or High (took every quantitative class they possibly could). Whether the student was offered a job in their first year on the market served as the dependent variable. To maintain equivalency across fields, graduate students who did not apply for jobs until after completing a postdoctoral fellowship were not included in the analysis.

Two generalized linear models were estimated using maximum likelihood within **[PROGRAM]**, in which the binary outcome variable of getting a job offer was predicted using a logit link and a Bernoulli distribution. As such, the log of the odds of the probability of an offer was predicted directly by the model, and predicted probabilities were found through the inverse logit link of exp(logit) / [1+exp(logit)]. The significance of fixed effects was evaluated using Wald tests (i.e., the *p*-value for the ratio of each estimate to its standard error). First, an empty model was estimated to provide a baseline for assessing improvement in model fit. The estimated fixed intercept was **[EST]**, which is the expected logit of getting a job offer, which translates into an overall expected probability of **[VALUE]** of getting a job offer within this sample.

**Section 2**

Continue by adding fixed main effects of the number of publications, the number of conference posters, and level of quantitative training (low, medium, or high) to the model predicting getting a job offer (new −2LL = 604.98) to answer the questions below.

**Questions:**

1. What is the expected logit of getting a job for someone with no publications, no posters, and low quantitative training? Report the estimate and SE:
2. What is the expected probability of getting a job for someone with no publications, no posters, and low quantitative training? Report the value and SE:
3. What is the expected logit of getting a job for someone with no publications, no posters, and medium quantitative training? Report the estimate and SE:
4. What is the expected probability of getting a job for someone with no publications, no posters, and medium quantitative training? Report the value and SE:
5. What is the expected logit of getting a job for someone with no publications, no posters, and high quantitative training? Report the estimate and SE:
6. What is the expected probability of getting a job for someone with no publications, no posters, and high quantitative training? Report the value and SE:
7. What is the difference in the logit of getting a job between low and medium quantitative training? Report the fixed effect, SE, and *p*-value:
8. What is the difference in the logit of getting a job between medium and high quantitative training? Report the fixed effect, SE, and *p*-value:
9. What is the change in the logit of getting a job for every additional publication?  
   Report the fixed effect, SE, and *p*-value:
10. What is the change in the logit of getting a job for every additional conference poster?  
    Report the fixed effect, SE, and *p*-value:
11. What is the difference in the logit for one additional publication versus one additional conference poster? Report the fixed effect, SE, and *p*-value:

**Create Figure 1 A and B:** **(A)** create model-predicted outcomes on the logit scale for a graduate student with low, medium or high quantitative training who had 0, 2, 4, 6, or 8 publications, holding conference posters constant at 5. Plot the predicted outcomes on the y-axis (scaled from −2 to 4 by 0.5) against number of publications on the x-axis (scaled from 0 to 8) with separate lines for each of levels of quantitative training. **(B)** Show the same model predictions in probability instead (in which the y-axis is scaled from 0 to 1 by .10).

**Create Figure 2 A and B:** **(A)** create model-predicted outcomes on the logit scale for a graduate student with low, medium or high quantitative training who had 0, 2, 4, 6, or 8 conference posters, holding publications constant at 3. Plot the predicted outcomes on the y-axis (scaled from −2 to 4 by 0.5) against number of conference posters on the x-axis (scaled from 0 to 8) with separate lines for each of levels of quantitative training. **(B)** Show the same model predictions in probability instead (in which the y-axis is scaled from 0 to 1 by .10).

**[Insert Figure 1A and 1B, Figure 2A and 2B]**

**Results:**

Second, a conditional model was estimated to examine the effect of number of publications, number of conference posters, and level of quantitative training. Results in terms of model-predicted log-odds are shown in Figure 1A and Figure 2A; results in terms of model-predicted probabilities are shown in Figure 1B and Figure 2B.

With respect to research productivity, the **[A]** of number of publications indicated that each additional publication was related to a **[EST]** difference of in the logit of receiving a job offer, such that students with more publications had a **[B C]** probability of receiving a job offer. In addition, the **[A]** of number of conference posters indicated that each additional poster was related to a **[EST]** difference of in the logit of receiving a job offer, such that students with more conference posters had a **[B C]** probability of receiving a job offer. Interestingly, the effect of an additional publication was **[B D]** than the effect of an additional conference poster, suggesting that students searching for jobs would be better advised to put their time into **[E]**. Finally, we examined the effects of quantitative training. Relative to students with a low level of quantitative training, achieving a medium amount of training resulted in a **[B C]** logit of receiving a job offer, whereas achieving a high amount of quantitative training instead of a medium amount resulted in a **[B C]** logit of receiving a job offer. For instance, given 0 publications and 0 conference posters, the probability of receiving a job offer for persons with low, medium, or high quantitative training was **[VALUE]**, **[VALUE]**, and **[VALUE]**, respectively.