

## BIOS 6612 Homework 4: Poisson regression

A study has been carried out to determine the relationship between incidence of non-melanoma skin cancer (outcome) and age and city of residence (predictors) among women. The file `skincancer.csv` contains the data from this study. There are four variables:

- `city`: a factor with two levels for city of residence, either Minneapolis or Dallas.
- `age.group`: a factor with age in years, given in ranges.
- `cases`: number of incident cases of non-melanoma skin cancer ( $Y_i$ ).
- `py1000`: total person-time at risk, in 1000s of person-years ( $T_i$ ).

You will need to fit several Poisson regression models to answer the following questions. Provide code and output for your analyses with your answers to this assignment.

1. **(15 points)** For descriptive purposes, the simple regression model

$$\log \mathbb{E}(Y_i) = \log T_i + \beta_0 + \beta_1 \text{Dallas}_i$$

is of interest. Let  $\text{Dallas}_i$  be an indicator variable for city of residence, with  $\text{Dallas}_i = 1$  if the city of residence is Dallas and 0 if the city of residence is Minneapolis.

- (a) Fit this model; provide estimates and standard errors for each regression coefficients. **(5 points)**
  - (b) Interpret this model. Include estimates and confidence intervals for the rate of non-melanoma skin cancer in each city as part of your response. Interpret the intercept or explain why it is not interpretable. **(10 points)**
2. **(30 points)** It may be important to adjust for age when modeling the incidence of non-melanoma skin cancer.
    - (a) Fit a Poisson regression model for the effect of city on rate of non-melanoma skin cancer adjusting for age. That is, include terms for both age group and city in your model. Report your estimated model coefficients and standard errors. **(5 points)**

- (b) A previous study found that age-adjusted skin cancer rates in Dallas were double those in Minneapolis. Carry out a hypothesis test of whether results from this study are consistent with those from the previous study. Make sure to explicitly state your null and alternative hypothesis, the test statistic and its null distribution, and the conclusion of your hypothesis test. **(10 points)**
  - (c) Carry out a test of the hypothesis that rates of skin cancer in the 15–24 and 25–34 age groups are equal. **(10 points)**
  - (d) Based on this model, what is the estimated rate of skin cancer among women in Minneapolis aged 45–54? **(5 points)**
3. **(15 points)** The saturated model for this data includes (in addition to main effects) the interaction between age and city of residence.
- (a) Fit this model and carry out a likelihood ratio test of the null hypothesis that the rate ratio for skin cancer between Dallas and Minneapolis does not depend on age. **(10 points)**
  - (b) Explain another way to perform this test *without* actually fitting the interaction model. **(5 points)**