PBHS 32410 / STAT 22401 **Problem Set 5** **Winter 2025**

Possible points 100

1. (24 pts) **Confounding**. forced expiratory volume (FEV) is a widely used measure of lung capacity and function. In this study, factors related to FEV in children were being examined.

   The data are on the website in the file

   `fev_by_age.dta`

   (a) Examine the variable *age* in relation to FEV (via a plot and regression mode). What is the relationship?

   (b) Examine the indicator variable *nsmoke* in relation to FEV. Assuming smoking is bad for lung function, does this relationship make sense?

   (c) Use both variables in a regression model of FEV. Interpret the results. What changes here and why?

   (d) Do the following additional analyses: 1) Run the regression separately by smoking status and describe the effect of age on FEV in each group. 2) Create the age by smoking interaction term and test it. How is what is going on in (c) different from an interaction effect?

2. (20 pts) **Infection Data**. In a study of Zika virus infection over a fixed seasonal period, the potential preventative effect of night-time mosquito net protection is investigated. Data on infection occurrences are collected on 50 villages over a fixed period. Net use is coded as 0 (no) or 1 (yes).

   The data are on the website in file

   `Zika_12.dta`

   (a) Examine summary statistics on infections overall. Does the Poisson model seem reasonable? Repeat this evaluation for the data stratified by net use (no/yes).

   (b) Fit the Poisson regression model with net use as the predictor variable. Write out the estimated model on the log(counts) scale and the counts scale.

(c) Check that the model reproduces the mean counts by netting strata. Then fit the model with IRR and check that the relative infection reduction is produced (i.e., ratio of infection rate for net vs. not). Summarize the result (what is the net effect, haha)?

(d) Predict the counts and list out in comparison to the actual counts. The results may seem odd in that many individual predictions are 'off'. However, this simple model can only produce limited predictions (based on values of the netting variable). Nonetheless, what is notable about the netting vs. non-netting group that the model captures (think about the mean and also range of counts between groups).

3. (56 pts) **Factors related to kidney stone episodes.** The data here are taken from long-term followup of 1423 patients being treated at a kidney-stone clinic. The data set contains the number of kidney stones that each patient has formed from the beginning of treatment at the clinic, together with the number of years of follow-up. Patient age and sex are recorded. A small number of patients have only one functional kidney, either because the kidney was surgically removed or never functioned; this is also recorded. Only patients with at least one year of follow-up are included.

   The data can be found on Canvas in

   `stones_12.dta`

   (a) Calculate the incidence rate (IR) overall (events/person-years). Hint: to compute, consider that the (mean kidney stones) times N is the sum, and likewise for person time. Since both have the same denominator, the ratio of means is equal to the ratio needed for the IR.

   (b) Calculate the incidence rate ratio for males vs. females.

   (c) Using sex as the only predictor, fit the Poisson model. Be sure to include the follow-up time as the 'exposure' variable. What is the interpretation of the parameter for sex? Compare the estimate obtained by the model to the estimate in (b). Given the estimate and the corresponding 95% CI for the sex parameter, do men and women form stones at significantly different rates?

(d) Why it is important to include the follow-up time in the model?

(e) Do individuals with only one functional kidney have different kidney-stone rates, ignoring other effects (i.e., with one kidney yes/no as only predictor)? If so, how does the rate of kidney-stone formation differ from those with two kidneys?

(f) What is the effect of age on rate of kidney-stone formation, ignoring other possible factors?

(g) Fit a model with all three effects considered above. Provide a prediction of the rate of kidney-stone formation for a 45 years old female patient having only one functional kidney.