Lab 5: Discrete outcome regression

Objective: To gain some practice dealing with discrete-outcome regression.

Exercises: Below is a set of exercises that we will go through individually, in small groups, and/or together as appropriate and as time permits. They all pertain to the following scenario:

A study was conducted to evaluate the association between kidney stone history (X) and coronary artery calcification (Y). A total of N = 270 study participants were sampled and surveyed regarding their kidney stone history, and their degree of coronary artery calcification was measured by the Agatston score and classified as a three-level variable.

The data from this study were then tabulated; the results are shown in the table below.

	Y = 2 (severe)	Y = 1 (mild)	Y = 0 (none)
X = 2 (2+ prior)	72	8	10
X = 1 (1 prior)	60	12	18
X = 0 (0 prior)	45	15	30

Consider the following three regression models that could be used to evaluate the association between kidney stone history and coronary artery calcification:

logit
$$P(Y \ge 1 \mid X = x) = \alpha_0 + \alpha_1 1(x \ge 1)$$
 (MODEL 1)

logit
$$[P(Y = k | X = x)/P(Y=0 | X = x)] = \beta_{0k} + \beta_{1k}1(x=1) + \beta_{2k}1(x=2); k = 1, 2$$
 (MODEL 2)

logit
$$P(Y \le k | X = x) = y_0 - y_1 1(x=1) - y_2 1(x=2); k = 0, 1$$
 (MODEL 3)

Exercise 1: Which model(s) is/are saturated?

Exercise 2: Determine estimates of the values of α_0 , β_{21} , and γ_2 .

Exercise 3: Briefly discuss the relative advantages and disadvantages of each of the three models.

Exercise 4: Suppose the investigators had settled on fitting a model in the spirit of **MODEL 1**, but also adjusting for age. Describe two settings that would lead you to support the decision to adjust for age.

Exercise 5: Suppose the investigators had settled on fitting a model in the spirit of **MODEL 3**, but also adjusting for coronary heart disease. Describe a setting that would lead you to be nervous about the decision to adjust for coronary heart disease.