# Math-661: Assignment 3

1. **Exercise 1 (Agresti 5.20)**

   Let $y_i, i = 1, \ldots, N$, denote $N$ independent binary random variables.

   (a) Derive the log-likelihood for the probit model $\Phi^{-1}[\pi(\boldsymbol{x}_i)] = \sum_j \beta_j x_{ij}$.

   (b) Show that the log-likelihood equations for the logistic and probit regression models are

   $$\sum_{i=1}^{N} (y_i - \hat{\pi}_i) z_i x_{ij} = 0, \qquad j = 1, \ldots, p,$$

   where $z_i = 1$ for the logistic case and $z_i = \phi(\sum_j \hat{\beta}_j x_{ij})/[\hat{\pi}_i(1 - \hat{\pi}_i)]$ for the probit case.

2. **Exercise 2 (based on Agresti 5.32)**

   For the horseshoe crab dataset (`Crabs.txt`), let $y = 1$ if a female crab has at least one satellite, and let $y = 0$ if a female crab does not have any satellite.

   (a) Fit a main-effects logistic model using color and weight as explanatory variables.

      i. Interpret the regression coefficients.

      ii. Show how to conduct inference about the color and weight effects (i.e., evaluate statistical significance).

   (b) Allow interaction between color and weight in their effects on $y$.

      i. Interpret the regression coefficients.

      ii. Test whether this model provides a significantly better fit compared to the main-effects model.

3. **Exercise 3: Survival of the Donner Party**

   In 1846, a group of 87 people (called the Donner Party) were headed west from Springfield, Illinois, to California. The leaders attempted a new route through the Sierra Nevada and were stranded throughout the winter. The harsh weather conditions and lack of food resulted in the death of many people within the group. Social scientists have used the data to study the theory that females are better able than males to survive harsh conditions. The data are saved under `Donner.txt`.

   (a) Create a logistic regression model using gender and age as predictors and provide the equation of the estimated model.

   (b) Interpret the regression coefficients.

   (c) Estimate the survival probability of a 20-year old female (show your calculation).

   (d) Explain why the deviance or Pearson goodness-of-fit tests are not appropriate.

   (e) Assess the model goodness-of-fit.