Homework 5

Due Friday, 2/26/20 on Canvas.

Treat the data analysis problem as a report to a client, and do NOT turn in any R, SAS or python code. For example, when reporting coefficient estimates, your answer should just be $\hat{\beta} = 2.47$ and NOT

```
> fit$coefficients[2]
[1] 2.469829453
```

It is up to you as to how many decimal places to report, but please be reasonable and consistent. Two or three decimal places is sufficient for most applications. You should only report what is necessary, i.e. estimates, confidence intervals, P values, plots, etc., in a clear and concise manner.

1. You observe an outcome Y_i that can take on values 1,2, and 3 and a continuous covariate x_i. Consider fitting the proportional odds model where:

$$P(Y_i = j) = \pi_{ij} > 0, \quad j = 1, 2, 3, \quad i = 1, \ldots, n.$$

$$\log\left(\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3}}\right) = \alpha_1 - \beta x_i,$$

$$\log\left(\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3}}\right) = \alpha_2 - \beta x_i.$$

   (a) What is π_{i2} in terms of $\alpha_1, \alpha_2, \beta$ and x_i?

   (b) Show that $\alpha_1 \leq \alpha_2$.

   (c) The file Q1c.txt on Canvas contains a sample data set with $n = 200$. Using the data, fit two separate logistic regression models:

   (Mi) $logit(\pi_{i1}) = \beta_{01} - \beta_{11} x_i$,

   (Mii) $logit(\pi_{i1} + \pi_{i2}) = \beta_{02} - \beta_{12} x_i$.

   Using R, make two plots:

      (i) On the same axes, plot the estimated probabilities from (Mi) and the estimated probabilities from (Mii) as a function of x_i.

      (ii) On the same axes, plot the logit of the estimated probabilities from (Mi) and the logit of the estimated probabilities from (Mii) as a function of x_i.

   (d) The file Q1d.txt on the Courseweb also contains a sample data set with $n = 200$. Repeat part (c) using this data set.

   (e) From the plots in parts (c) and (d), are you more conformable fitting the proportional odds model to Q1c.txt or Q1d.txt and why?