MA478 HW 1

Clark

Due 30 January

1. Suppose $\boldsymbol{X}$ and $\boldsymbol{W}$ are two design matrices with n rows and $C(\boldsymbol{X}) = C(\boldsymbol{W})$. Show that $\boldsymbol{P_X} = \boldsymbol{P_W}$

   Hint: Show that $\boldsymbol{P_X P_w} = \boldsymbol{P_X}$ and $\boldsymbol{P_W P_X} = \boldsymbol{P_W}$.

   What does this mean about $(\boldsymbol{P_X} - \boldsymbol{P_w})^T(\boldsymbol{P_X} - \boldsymbol{P_w})$?

2. Up to now we have been considering models of the form $E[Y_i] = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$. What would happen if, instead, we fit the model: $E[Y_i] = \beta_0 + \beta_1(x_{1,i} - \bar{x_1}) + \beta_2(x_{2,i} - \bar{x_2})$ where $\bar{x_j}$ denotes the column averages? Why does this make sense in light of what you showed above?

3. Next you will explore, analyze, and model a dataset for IOCT times. Your objective is to build a multiple linear regression model that best determines the answer to the question, does height matter for the IOCT? To do this properly, you should ensure you explore relationships between your covariates and you should ensure (or address the impacts of) any violations of the linear regression assumptions. Provide a professional write up for your client, a professor within DPE who hasn't taken stats since college. Ensure you address:

   - Data Exploration - Consider plots that show distributions of variables of interest. Issues of collinearity. Modifications of covariates if necessary. Too much detail will definitely cause DPE to lose interest, too little detail will result in questions for your final models.
   - Data preparation - Consider outliers. What will you do with high points that have high leverage? Will you transform any of your variables?
   - Build models - Consider at least 3 different models, using different variables (or different transformations). If your models are nested conduct formal tests to decide on which model is the most appropriate. If the models are not nested consider using AIC or BIC to decide.

- Final model - Present your best model. Why did you select it? What is your answer to the research question? What are some limitations for your study? Based off of your model what is your predicted IOCT time?

  This should be 2-3 pages. Include figures as an appendix. Also ensure you include your R code.