STAT331 Applied Linear Models
Practice Problems for Chap7-Chap9

1. Consider modelling a two-sample problem with n observations in two ways:

   (1) $y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$,

   (2) $y_i = \gamma_1 z_{i,1} + \gamma_2 z_{i,2} + \epsilon_i$,

   where

   $$x_{i1} = \begin{cases} 1, & \text{if} \quad i = 1, \ldots, m, \\ 0, & \text{if} \quad i = m+1, \ldots, n, \end{cases} \quad x_{i2} = \begin{cases} 0, & \text{if} \quad i = 1, \ldots, m, \\ 1, & \text{if} \quad i = m+1, \ldots, n, \end{cases}$$

   and

   $$z_{i1} = 1, z_{i2} = \begin{cases} 0, & \text{if} \quad i = 1, \ldots, m, \\ 1, & \text{if} \quad i = m+1, \ldots, n. \end{cases}$$

   Furthermore, assume that $\gamma_1 = \beta_1$, and $\gamma_2 = \beta_2 - \beta_1$. Therefore, these two models are essentially equivalent.

   (a) Write out the design matrices for model (1) and (2).

   (b) Interpret β_1, β_2, γ_1, and γ_2.

   (c) Show that $E(\hat{\beta}_1) = E(\hat{\gamma}_1)$ and $E(\hat{\beta}_2) = E(\hat{\gamma}_1) + E(\hat{\gamma}_2)$.

   (d) Show that $Var(\hat{\beta}_1) = Var(\hat{\gamma}_1)$ and $Var(\hat{\beta}_2) = Var(\hat{\gamma}_1 + \hat{\gamma}_2)$.

2. To predict performance on the y =GRE (graduate entrance exam) score, 36 undergraduate college students were asked to follow 1 of 3 study programs. x = High school GPA was also recorded. The population model considered was

$$E(y) = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \beta_3 x + \beta_4 P_1 x + \beta_5 P_2 x$$

where the indicator $P_1 = 1$ for program 1 and 0 otherwise, and $P_2 = 1$ for program 2 and 0 otherwise.

(a) Based on the model, write three separate models for the relationship between y and x for students who followed program 1, 2 and 3.

(b) Give an interpretation for each of the following parameters: β_3, β_4, and β_5.

(c) In terms of the model parameters, write the null hypothesis and the alternative hypothesis for testing whether interaction is present between GPA and study program. Also, write the regression model reduced by this null hypothesis.

(d) Using your answer from part c), briefly explain how you would carry out the test for interaction (write the test statistic and how do you draw the conclusion).

3. In weighted least square estimation, define H^{WLS} to be $X(X^TV^{-1}X)^{-1}X^TV^{-1}$. One can verify that $\hat{Y}^{WLS} = H^{WLS}Y$.

   (a) Is H^{WLS} symmetric? Is H^{WLS} idempotent?

   (b) Is H^{WLS} a projection matrix? In other words, is it guaranteed that $Y - H^{WLS}Y$ is orthogonal to $H^{WLS}Y$?

   (c*) Define the space spanned by $\{H^{WLS}Z : Z \in \mathbb{R}^n\}$ to be W. Prove that W is equivalent to the column space of X.

   (d) Consider the ordinary least square estimator $\hat{\beta}^{OLS}$. Prove that the covariance between $\hat{\beta}^{OLS}_j$ and $\hat{\beta}^{WLS}_j$ is always non-negative for any j.

4. Recall that when $y_i \in \{0, 1\}$, the likelihood function for Bernoulli distribution is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}.$$

Now we are going to consider another labeling method. In particular, let $y_i \in \{-1, 1\}$ instead of $y_i \in \{0, 1\}$. In other words, we denote the first class by 1, and the second class by -1.

(a) Prove that the likelihood function when $y_i \in \{-1, 1\}$ is

$$L_{new}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_i^{(1+y_i)/2} (1 - \pi_i)^{(1-y_i)/2}.$$

(b) Prove that to maximize the likelihood function L_{new} is the same as to

$$\min \ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log\{1 + \exp(-y_i \mathbf{x}_i^T \boldsymbol{\beta})\}.$$

Therefore, the benefit of using the labels $y_i \in \{-1, 1\}$ is that we can treat the optimization in a simpler way. This function $\log\{1 + \exp(-y\mathbf{x}^T \boldsymbol{\beta})\}$ is called the logistic deviance loss function.