Statistical Methods - Homework #1

1. Consider the simple linear regression model without an intercept:

$$y_i = \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim (0, \sigma^2) \ \text{ independently for } i = 1, \ldots, n.$$

(A) Find the least squares estimate of β_1, denoted $\hat{\beta}_1$, that minimizes

$$S(\beta_1) = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2.$$

(B) Let the fitted values be $\hat{y}_i = \hat{\beta}_1 x_i$. Using the result in (A), show that

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} \hat{y}_i^2.$$

(C) Find $\mathrm{E}(\hat{\beta}_1)$. Also, show that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

(D) Find $\mathrm{E}(\hat{y}_i)$. Also, show that

$$\mathrm{Var}(\hat{y}_i) = \frac{\sigma^2 x_i^2}{\sum_{j=1}^{n} x_j^2}.$$

(E) Let $e_i = y_i - \hat{y}_i$. Find $\mathrm{E}(e_i)$. Also, show that

$$\mathrm{Var}(e_i) = \sigma^2 \left(1 - \frac{x_i^2}{\sum_{j=1}^{n} x_j^2}\right).$$

(Hint: use $\mathrm{Var}(y_i) = \mathrm{Var}(y_i - \hat{y}_i) + \mathrm{Var}(\hat{y}_i)$)

2. Consider the multiple linear regression model:

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim (0, \sigma^2) \text{ independently for } i = 1, \ldots, n, \tag{4}$$

where $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, \boldsymbol{\beta} \in \mathbb{R}^p$ and $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n \times p}$ is full rank. ($\operatorname{rank}(X) = p$) Letting $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, (4) can equivalently be written as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \text{ and } \operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n.$$

(A) Show that the least squares estimate of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}$, that minimizes

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2$$

is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^\top \right)^{-1} \sum_{i=1}^n \boldsymbol{x}_i y_i = \left(X^\top X\right)^{-1} X^\top \boldsymbol{y}.$$

(B) Find $\mathrm{E}(\hat{\boldsymbol{\beta}})$. Also, show that

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 \left(X^\top X\right)^{-1}.$$

(Hint: $\operatorname{Var}(A\boldsymbol{Z}) = A \operatorname{Var}(\boldsymbol{Z})\, A^\top$ for a constant matrix A)

(C) Let the fitted values be $\hat{y}_i = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$ and the hat matrix be $H = X(X^\top X)^{-1} X^\top$. Show that

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \boldsymbol{y}^\top (I - H)\boldsymbol{y}.$$

Also, show that

$$\mathrm{E}(\boldsymbol{y}^\top (I - H)\boldsymbol{y}) = (n - p)\sigma^2.$$

(Hint: For a constant matrix $A \in \mathbb{R}^{n\times n}$ and a random vector $\boldsymbol{y} \in \mathbb{R}^n$, $E(\boldsymbol{y}^\top A\boldsymbol{y}) = \mathrm{trace}(A\mathrm{Var}(\boldsymbol{y})) + \mathrm{E}(\boldsymbol{y})^\top A\mathrm{E}(\boldsymbol{y})$.)

(D) Show the Gauss-Markov Theorem: For all estimators of the form $\tilde{\boldsymbol{\beta}} = C\boldsymbol{y}$ with an $p \times n$ matrix C such that $\mathrm{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$, we have

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}) \text{ is nonnegative definite.}$$

(Hint: let $C = (X^\top X)^{-1}X^\top + B$ for a $p \times n$ matrix B, and note that the unbiasedness is equivalent to $BX = \mathbf{0}$.)

3. Consider the multiple linear regression model:

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \ \text{ independently for } i = 1, \ldots, n, \tag{5}$$

where $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, \boldsymbol{\beta} \in \mathbb{R}^p$ and $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n\times p}$ is full rank. ($\text{rank}(X) = p$) Let $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$. Letting $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, (5) can equivalently be written as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \sigma^2 I_n).$$

(A) Find the Maximum Likelihood Estimator(MLE) of $\boldsymbol{\beta}$ and σ^2. Hint: the log-likelihood is

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})^\top(\boldsymbol{y} - X\boldsymbol{\beta}).$$

(B) Find the Uniformly Minimum Variance Unbiased Estimator(UMVUE) of $\boldsymbol{\beta}$ and σ^2.
(Hint : $(\hat{\boldsymbol{\beta}}, \text{RSS})$ is a complete and sufficient statistic for $(\boldsymbol{\beta}, \sigma^2)$.)

(C) Let the hat matrix be $H = X(X^\top X)^{-1}X^\top$. Show that

$$\frac{1}{\sigma^2}\boldsymbol{y}^\top(I - H)\boldsymbol{y} \sim \chi^2(n-p)$$

(Hint : If $\boldsymbol{y} \sim MVN(\boldsymbol{\mu}, \sigma^2 I_n)$, A is a symmetric idempotent matrix with $\text{rank}(A) = \text{trace}(A) = k$, and $\boldsymbol{\mu}^\top A\boldsymbol{\mu} = 0$, then $\frac{1}{\sigma^2}\boldsymbol{y}^\top A\boldsymbol{y} \sim \chi^2(k)$.)

(D) Show that the estimators

$$\hat{\boldsymbol{\beta}} = \left(X^\top X\right)^{-1} X^\top \boldsymbol{y}, \text{ and } \hat{\sigma}^2 = \frac{1}{n-p}\boldsymbol{y}^\top (I - H)\boldsymbol{y}$$

are independent. (Hint: for jointly multivariate normal vectors, zero covariance implies independence.)

(E) Consider the hypothesis test:

$$H_0 : \boldsymbol{\beta} = \mathbf{0}, \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}.$$

Under H_0, show that

$$\frac{1}{\sigma^2}\boldsymbol{y}^\top H \boldsymbol{y} \sim \chi^2(p), \text{ and}$$

$$\frac{\mathrm{ESS}/p}{\mathrm{RSS}/(n-p)} = \frac{\boldsymbol{y}^\top H \boldsymbol{y}/p}{\boldsymbol{y}^\top (I-H)\boldsymbol{y}/(n-p)} \sim F(p, n-p)$$

(Hint : If $\boldsymbol{y} \sim MVN(\boldsymbol{\mu}, \sigma^2 I_n)$, A is a symmetric idempotent matrix with $\mathrm{rank}(A) = \mathrm{trace}(A) = k$, and $\boldsymbol{\mu}^\top A \boldsymbol{\mu} = 0$, then $\dfrac{1}{\sigma^2}\boldsymbol{y}^\top A \boldsymbol{y} \sim \chi^2(k)$.)

4. Suppose that we have the two multiple linear regression models:

$$\begin{aligned} &\text{Model A: } y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \\ &\text{Model B: } y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \end{aligned}$$

for $i = 1, \ldots, n$, where $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.

Let $\hat{y}_i^{(\mathrm{A})}$ and $\hat{y}_i^{(\mathrm{B})}$ be the fitted values for Model A and Model B, respectively. Compare each of the following quantities. *(Hint: use projection.)*

(a) $\sum_{i=1}^n (y_i - \hat{y}_i^{(\mathrm{A})})^2$ and $\sum_{i=1}^n (y_i - \hat{y}_i^{(\mathrm{B})})^2$

(b) $\sum_{i=1}^n (\hat{y}_i^{(\mathrm{A})} - \bar{y})^2$ and $\sum_{i=1}^n (\hat{y}_i^{(\mathrm{B})} - \bar{y})^2$

(c) R^2 for Model A and R^2 for Model B