Statistical Methods - Homework #1

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

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

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

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

<u>Solution:</u> Differentiate and set to zero:

$$\frac{dS}{d\beta_1} = -2 \sum_{i=1}^{n} x_i \left(y_i - \hat{\beta}_1 x_i \right) = 0 \tag{1}$$

$$\Rightarrow \boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}.$$

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

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

<u>Solution:</u> Note

$$\begin{aligned}
\sum_{i=1}^{n} y_i^2 &= \sum_{i=1}^{n} (y_i - \hat{y}_i + \hat{y}_i)^2 \\
&= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} \hat{y}_i^2 + 2 \sum_{i=1}^{n} (y_i - \hat{y}_i) \hat{y}_i \\
&= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} \hat{y}_i^2 + 2\hat{\beta}_1 \underbrace{\sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i) x_i}_{=0},
\end{aligned}$$

because of (1). Hence,

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

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

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

<u>Solution:</u> Starting from the estimator in (A),

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i (\beta_1 x_i + \varepsilon_i)}{\sum_{i=1}^{n} x_i^2} = \beta_1 + \frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}.$$

By linearity of expectation and $\mathrm{E}(\varepsilon_i) = 0$,

$$\mathrm{E}(\hat{\beta}_1) = \beta_1 + \frac{\sum_{i=1}^n x_i \,\mathrm{E}(\varepsilon_i)}{\sum_{i=1}^n x_i^2} = \boxed{\beta_1}. \tag{2}$$

For the variance, use independence of the errors and $\mathrm{Var}(\varepsilon_i) = \sigma^2$:

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}\right) = \frac{1}{\left(\sum_{i=1}^n x_i^2\right)^2} \sum_{i=1}^n x_i^2 \,\mathrm{Var}(\varepsilon_i) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} = \boxed{\frac{\sigma^2}{\sum_{i=1}^n x_i^2}}. \tag{3}$$

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

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

Solution: Since $\hat{y}_i = \hat{\beta}_1 x_i$,

$$\mathrm{E}(\hat{y}_i) = \mathrm{E}(\hat{\beta}_1)\, x_i = \boxed{\beta_1 x_i} \quad \text{by (2)}.$$

For the variance, pull out the constant x_i and use (3):

$$\mathrm{Var}(\hat{y}_i) = \mathrm{Var}(\hat{\beta}_1)\, x_i^2 = \frac{\sigma^2}{\sum_{j=1}^n x_j^2} \cdot x_i^2 = \boxed{\frac{\sigma^2 x_i^2}{\sum_{j=1}^n x_j^2}}.$$

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

Solution: Since $y_i = \beta_1 x_i + \varepsilon_i$ and $\hat{y}_i = \hat{\beta}_1 x_i$,

$$\mathrm{E}(e_i) = \mathrm{E}(y_i - \hat{y}_i) = \beta_1 x_i - \mathrm{E}(\hat{\beta}_1)\, x_i = \boxed{0}.$$

In linear regression, fitted values and residuals are uncorrelated, so

$$\mathrm{Var}(y_i) = \mathrm{Var}(y_i - \hat{y}_i) + \mathrm{Var}(\hat{y}_i).$$

With $\mathrm{Var}(y_i) = \sigma^2$ and from (D) $\mathrm{Var}(\hat{y}_i) = \frac{\sigma^2 x_i^2}{\sum_{j=1}^n x_j^2}$, we get

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

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. ($\text{rank}(X) = p$) Letting $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, (4) can equivalently be written as

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

(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

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

<u>Solution:</u> Differentiate and set to zero:

$$\begin{aligned} \frac{\partial S}{\partial \boldsymbol{\beta}} &= -2 \sum_{i=1}^{n} \boldsymbol{x}_i \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} \right) = 0 \\ &\Rightarrow \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \right) \hat{\boldsymbol{\beta}} = \sum_{i=1}^{n} \boldsymbol{x}_i y_i. \end{aligned} \tag{5}$$

Since $\text{rank}(X) = \text{rank}(X^\top X) = p$, $X^\top X$ is invertible. Therefore,

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

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

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

(Hint: $\text{Var}(A\boldsymbol{Z}) = A \, \text{Var}(\boldsymbol{Z}) \, A^\top$ for a constant matrix A)
<u>Solution:</u> By linearity of expectation,

$$\text{E}(\hat{\boldsymbol{\beta}}) = (X^\top X)^{-1} X^\top \text{E}(\boldsymbol{y}) = (X^\top X)^{-1} X^\top X \boldsymbol{\beta} = \boxed{\boldsymbol{\beta}}.$$

For the variance,

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}) &= \text{Var}\big((X^\top X)^{-1} X^\top \boldsymbol{\varepsilon} \big) \\ &= (X^\top X)^{-1} X^\top \text{Var}(\boldsymbol{\varepsilon}) \left((X^\top X)^{-1} X^\top \right)^\top \\ &= (X^\top X)^{-1} X^\top (\sigma^2 I_n) \, X (X^\top X)^{-1} \\ &= \sigma^2 (X^\top X)^{-1} X^\top X (X^\top X)^{-1} \\ &= \boxed{\sigma^2 (X^\top X)^{-1}}. \end{aligned}$$

(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 $\boldsymbol{y} \in \mathbb{R}^n$, $E(\boldsymbol{y}^\top A\boldsymbol{y}) = \mathrm{trace}(A\mathrm{Var}(\boldsymbol{y})) + \mathrm{E}(\boldsymbol{y})^\top A\mathrm{E}(\boldsymbol{y})$.)

Solution: Since $\hat{\boldsymbol{y}} = H\boldsymbol{y}$, we have $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (I - H)\boldsymbol{y}$. Hence

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

because H (thus $I - H$) is symmetric idempotent.

With $A = I - H$, $\mathrm{E}(\boldsymbol{y}) = X\boldsymbol{\beta}$, $\mathrm{Var}(\boldsymbol{y}) = \sigma^2 I_n$, and $(I - H)X = \mathbf{0}$, the hint gives

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

since $\mathrm{tr}(H) = \mathrm{tr}\big((X^\top X)^{-1}X^\top X\big) = \mathrm{tr}(I_p) = p$.

(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 $BX = \mathbf{0}$.)

Solution: Using the hint, write $C = (X^\top X)^{-1}X^\top + B$; then $\mathrm{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta} \iff BX = \mathbf{0}$.

Since $\mathrm{Var}(\boldsymbol{y}) = \sigma^2 I_n$,

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2 CC^\top, \qquad \mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (X^\top X)^{-1}.$$

Expand $CC^\top$ and use $BX = \mathbf{0}$ to kill cross terms:

$$CC^\top = (X^\top X)^{-1} + BB^\top.$$

Therefore

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 BB^\top \succeq 0.$$

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

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

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

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

Solution: For fixed σ^2, maximizing $\ell(\boldsymbol{\beta}, \sigma^2)$ over $\boldsymbol{\beta}$ is equivalent to minimizing $(\boldsymbol{y} - X\boldsymbol{\beta})^\top(\boldsymbol{y} - X\boldsymbol{\beta})$ over $\boldsymbol{\beta}$. Differentiating with respect to $\boldsymbol{\beta}$ and setting to zero gives the normal equations $X^\top(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$.
Since $X^\top X$ is invertible, the MLE of $\boldsymbol{\beta}$ is

$$\boxed{\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y}}.$$

Plugging $\hat{\boldsymbol{\beta}}$ into the log-likelihood, the σ-dependent part becomes $-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\text{RSS}$, where
$\text{RSS} = (\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}) = \boldsymbol{y}^\top(I - H)\boldsymbol{y}$.
Differentiating with respect to σ^2 and setting to zero yields $\hat{\sigma}^2_{\text{MLE}} = \text{RSS}/n$.
Therefore,

$$\boxed{\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}) = \frac{1}{n}\boldsymbol{y}^\top(I - H)\boldsymbol{y}}.$$

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

Solution: The statistic $\left(\hat{\boldsymbol{\beta}}, \text{RSS} = \boldsymbol{y}^\top(I - H)\boldsymbol{y}\right)$ is sufficient for $(\boldsymbol{\beta}, \sigma^2)$, and it is complete for this full-rank normal linear model.

By the Lehmann–Scheffé theorem, any unbiased estimator that is a function of $\left(\hat{\boldsymbol{\beta}}, \text{RSS}\right)$ is the UMVUE.
From Question 2.(B), $\text{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, and $\hat{\boldsymbol{\beta}}$ is already a function of the complete sufficient statistic, so

$$\boxed{\text{UMVUE of } \boldsymbol{\beta} \text{ is } \hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y}}.$$

Also, from Question 2.(C), $\text{E}(\text{RSS}) = (n - p)\sigma^2$, so

$$\boxed{\text{UMVUE of } \sigma^2 \text{ is } \frac{\text{RSS}}{(n-p)} = \frac{1}{n-p}\boldsymbol{y}^\top(I - H)\boldsymbol{y}}.$$

(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 $\frac{1}{\sigma^2}\boldsymbol{y}^\top A \boldsymbol{y} \sim \chi^2(k)$.)

Solution: Under the model, $\boldsymbol{y} \sim MVN(\boldsymbol{\mu}, \sigma^2 I_n)$ with $\boldsymbol{\mu} = X\boldsymbol{\beta}$.
Let $A = I - H$; then A is symmetric idempotent and
$\text{rank}(A) = \text{trace}(I - H) = n - \text{trace}(X(X^\top X)^{-1}X^\top) = n - \text{trace}((X^\top X)^{-1}X^\top X) = n - p.$

Since $HX = X$, we have $AX = (I - H)X = \mathbf{0}$, so $\boldsymbol{\mu}^\top A \boldsymbol{\mu} = (X\boldsymbol{\beta})^\top A (X\boldsymbol{\beta}) = 0$.

By the hint, $\frac{1}{\sigma^2}\boldsymbol{y}^\top (I - H)\boldsymbol{y} = \frac{1}{\sigma^2}\boldsymbol{y}^\top A \boldsymbol{y} \sim \chi^2(n-p)$.

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

(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.)

Solution: Write $\boldsymbol{y} = H\boldsymbol{y} + (I - H)\boldsymbol{y}$, and note that $(H\boldsymbol{y}, (I - H)\boldsymbol{y})$ is jointly multivariate normal since both are linear transforms of $\boldsymbol{y}$.

$\text{Cov}(H\boldsymbol{y}, (I - H)\boldsymbol{y}) = H(\sigma^2 I_n)(I - H) = \sigma^2(H - H^2) = \mathbf{0}$, so $H\boldsymbol{y}$ and $(I - H)\boldsymbol{y}$ are independent.

Because $X^\top (I - H) = \mathbf{0}$, $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y} = (X^\top X)^{-1} X^\top (H\boldsymbol{y})$, so $\hat{\boldsymbol{\beta}}$ depends only on $H\boldsymbol{y}$.

Also, $\hat{\sigma}^2 = \frac{1}{n-p}\boldsymbol{y}^\top (I - H)\boldsymbol{y}$ depends only on $(I - H)\boldsymbol{y}$, so $\boxed{\hat{\boldsymbol{\beta}} \text{ and } \hat{\sigma}^2 \text{ are independent.}}$

(E) Consider the hypothesis test:

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

Under H_0, show that

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

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

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

Solution: Under $H_0 : \boldsymbol{\beta} = \mathbf{0}$, we have $\boldsymbol{y} \sim MVN(\mathbf{0}, \sigma^2 I_n)$.

The matrix H is symmetric idempotent with $\text{rank}(H) = \text{trace}(H) = p$ and $\mathbf{0}^\top H \mathbf{0} = 0$, so by the hint $\boxed{\frac{1}{\sigma^2}\boldsymbol{y}^\top H \boldsymbol{y} \sim \chi^2(p)}$.

By (C), $\frac{1}{\sigma^2}\boldsymbol{y}^\top (I - H)\boldsymbol{y} \sim \chi^2(n-p)$, and by (D) these two quadratic forms are independent.

Hence $\frac{(\boldsymbol{y}^\top H \boldsymbol{y})/p}{(\boldsymbol{y}^\top (I - H)\boldsymbol{y})/(n-p)} \sim F(p, n-p)$, i.e., $\boxed{\frac{\text{ESS}/p}{\text{RSS}/(n-p)} \sim F(p, n-p)}$.

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

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

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

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

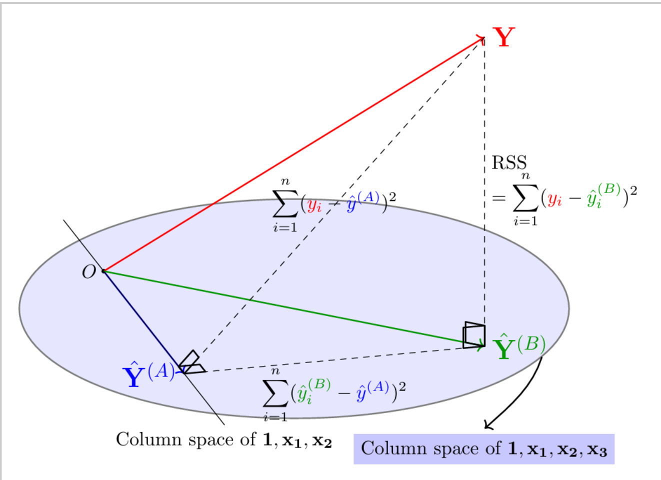
(a) $\sum_{i=1}^n (y_i - \hat{y}_i^{(\text{A})})^2$ and $\sum_{i=1}^n (y_i - \hat{y}_i^{(\text{B})})^2$

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

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

Solution:

Model B's regressor space contains that of Model A (since $\{1, x_1, x_2\} \subseteq \{1, x_1, x_2, x_3\}$). The projection can be graphically illustrated as follows:



That is, Model B projects $\mathbf{Y}$ onto a larger subspace, so the projection is at least as close to $\mathbf{Y}$ as in Model A. Hence the residual sum of squares satisfies

$$\sum_{i=1}^n (y_i - \hat{y}_i^{(\text{A})})^2 \;\geq\; \sum_{i=1}^n (y_i - \hat{y}_i^{(\text{B})})^2.$$

From the projection property we also have

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

Therefore, the explained sum of squares (ESS) satisfies

$$\sum_{i=1}^n (\hat{y}_i^{(\text{A})} - \bar{y})^2 \;\leq\; \sum_{i=1}^n (\hat{y}_i^{(\text{B})} - \bar{y})^2.$$

Finally, note that the coefficient of determination is $R^2 = \text{ESS}/\text{TSS}$, where TSS is the same for both models. Thus

$$R^2_{\text{A}} \;\leq\; R^2_{\text{B}}.$$