AFRICAN INSTITUTE FOR MATHEMATICAL SCIENCES

(AIMS RWANDA, KIGALI)

Name: Yuusf Brima — Assignment Number: 1

Course: Statistical Machine Learning for Data Science — Date: January 10, 2021

Exercise 1

Let $D_n = (\mathrm{x}_i, yi) \overset{\text{iid}}{\sim} p_{\mathrm{x},y}(\mathrm{x}, y), \mathrm{x}_i \in \mathbb{R}, y_i \in \mathbb{R}, i = 1, ..., n$. Consider using the data $\mathscr{D}_n$, to build mappings $f : \mathrm{x} \rightarrow Y$, such that $f \in \mathscr{H}$, where

$$\mathscr{H} := \left\{ \mathrm{x} \rightarrow f(\mathrm{x}) = \theta \mathrm{x}, \theta \in \mathbb{R}^* \right\} \tag{1}$$

We suppose that $\forall i \in [n]$, we have

$$\mathrm{p}(y_i|\mathrm{x}_i, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(y_i - f(\mathrm{x}_i))^2} \tag{2}$$

where $f \in \mathscr{H}$. We let $\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, ..., \mathrm{x_n})^{\mathrm{T}}$ and $\mathbf{y} = (y_1, y_2, y_3, ..., y_n)^{\mathrm{T}}$ and define

$$\mathbf{SSE}_n(a) = \sum_{i=1}^{n} (y_i - a\mathrm{x}_i)^2 = (\mathbf{y} - a\mathbf{x})(\mathbf{y} - a\mathbf{x})^{\mathrm{T}} = ||\mathbf{y} - a\mathbf{x}||_2^2. \tag{3}$$

(1) The input space in this problem, and its dimensionality is thus: $\mathscr{X} \in \mathbb{R}$ and $\dim(\mathscr{X}) = 1$

(2) The output space in this problem, and its dimensionality is thus: $\mathscr{Y} \in \mathbb{R}$ and $\dim(\mathscr{Y}) = 1$

(3) The dimensionality of $\mathbf{x}$ is $\mathbb{R}^{n\times 1}$

(4) The dimensionality of $\mathbf{y}$ is $\mathbb{R}^{n\times 1}$

(5) The assumed conditional distribution of $\mathbf{y}_i$ given x_i and deduce the distribution $\mathbf{y}$ given $\mathbf{x}$ is thus:

$$\mathrm{p}(y_i|\mathrm{x}_i) = \phi(y_i, f(\mathrm{x}_i), \sigma^2)$$

where

$$\begin{aligned} y_i = f(\mathrm{x}_i + \epsilon_i) \\ &= \mathrm{x}_i\theta + \epsilon_i \\ &= \mathrm{x}_i\theta \end{aligned}$$

$$
\begin{aligned}
\mathrm{p}(y_i|\mathrm{x}_i) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(y_i - f(\mathrm{x}_i))^2} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(y_i - \mathrm{x}_i\theta)^2} \\
&= \eta(f(\mathrm{x}_i), \sigma^2) \\
&= \eta(\mathrm{x}_i\theta, \sigma^2)
\end{aligned}
$$

(6) Rewriting the model defined by Equation (2) in its additive form featuring the deterministic component (signal) and the stochastic component (noise or error term). Be sure to reflect the fact that $f \in \mathscr{H}$ as defined in Equation (1)

$$
y_i = f(\mathrm{x}_i) + \epsilon_i \quad \text{where } \epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)
$$
$$
\underbrace{y_i}_{\text{observation}} = \underbrace{\mathrm{x}_i\theta}_{\text{deterministic signal}} + \underbrace{\epsilon_i}_{stochasticnoise} \quad \text{and } f \in \mathscr{H}
$$

(7) Type Statistical Machine Learning task is being solved is Simple Linear Regression (SLR) and the justification is $\mathrm{x}_i \in \mathbb{R}$ which is a single predictor variable and $y_i \in \mathbb{R}$ which is continuous.

(8) Using the appropriate tools to find $\frac{\partial \mathbf{SSE}_n(a)}{\partial_a}$

$$
\begin{aligned}
\frac{\partial \mathbf{SSE}_n(a)}{\partial_a} &= \sum_{i=1}^{n} (y_i - \mathrm{x}_i a)^2 = (y_i - \mathrm{x}_i a)(y_i - \mathrm{x}_i a)^{\mathrm{T}} = ||y_i - \mathrm{x}_i a||_2^2. \\
&= \frac{\partial}{\partial a}(\sum_{i=1}^{n} (y_i - \mathrm{x}_{ia})^2) \\
&= -\sum_{i=1}^{n} 2\mathrm{x}_i(y_i - \mathrm{x}_{ia}) \\
&= -2(\sum_{i=1}^{n} \mathrm{x}_i y_i - a\sum_{i=1}^{n} \mathrm{x}_i^2) \\
&= -2(\mathrm{x}^{\mathrm{T}} y - a\mathrm{x}^{\mathrm{T}}\mathrm{x})
\end{aligned}
$$

(9) Show that

$$
\hat{\theta} = \underset{\theta \in \mathbb{R}^*}{\mathbf{argmin}}\{\mathbf{SSE}_n(\theta)\} = \frac{\mathrm{x}^{\mathrm{T}}\mathrm{y}}{\mathrm{x}^{\mathrm{T}}\mathrm{x}}
$$

$$
\begin{aligned}
\frac{\partial \mathbf{SSE}_n(a)}{\partial_a} = 0 &\iff 2(\mathrm{x}^{\mathrm{T}} - a\mathrm{x}^{\mathrm{T}}\mathrm{x}) = 0 \\
a\mathrm{x}^{\mathrm{T}}\mathrm{x} &= \mathrm{x}^{\mathrm{T}}\mathrm{y} \\
a &= \frac{\mathrm{x}^{\mathrm{T}}\mathrm{y}}{\mathrm{x}^{\mathrm{T}}\mathrm{x}}
\end{aligned}
$$

$$
\underset{\theta \in \mathbb{R}^*}{\mathbf{argmin}}\{\mathbf{SSE}_n(\theta) = \frac{\mathrm{x}^{\mathrm{T}}\mathrm{y}}{\mathrm{x}^{\mathrm{T}}\mathrm{x}} = (\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\mathrm{y}
$$

(10) Proving that mean$[\hat{\theta}] = \mathbb{E}[\hat{\theta}]$.

$$\begin{aligned}
\mathbb{E}[\hat{\theta}] &= \mathbb{E}[(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\mathrm{y}] \\
&= (\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\mathrm{y}
\end{aligned}$$

But

$$\mathrm{y} = \theta\mathrm{x} + \epsilon$$

Therefore

$$\begin{aligned}
\mathbb{E}[\hat{\theta}] &= \theta \underbrace{(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}((\mathrm{x}^{\mathrm{T}}\mathrm{x}))}_{\text{identity matrix}} + \underbrace{(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\mathbb{E}[\epsilon]}_{0} \\
&= \theta \quad \text{so, } \hat{\theta} \text{ is an unbiased estimator}
\end{aligned}$$

(11) Proving that variance$[\hat{\theta}] = \mathbb{V}[\hat{\theta}]$

$$\begin{aligned}
\mathbb{V}[\hat{\theta}] &= \mathbb{V}[\theta(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\mathrm{x} + (\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\epsilon] \\
&= \mathbb{V}[\theta + (\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\epsilon] \\
&= \mathbb{V}[(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}\epsilon] \\
&= ((\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}})^{\mathrm{T}}\epsilon(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}} \\
&= (\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\sigma^2\mathcal{I}_n(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}} \\
&= \sigma^2(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}
\end{aligned}$$

(12) Finding the mean$(y_i|\mathrm{x}_i) = \mathbb{E}[\mathrm{x}_i] \quad \forall i \in [n]$ and deduce mean$(y|\mathrm{x}) = \mathbb{E}|x]$

$$\begin{aligned}
\mathbb{E}[y_i|\mathrm{x}_i] &= \mathbb{E}[\theta\mathrm{x}_i + \epsilon_i] \\
&= f(\mathrm{x}_i) \\
&= \theta\mathbb{E}[\mathrm{x}_i], \quad \mathrm{x}_i \in \mathbb{R} \text{ and} \\
&= \theta\mathrm{x}_i, \quad \forall i \in [n] \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[y|\mathrm{x}] &= f(\mathrm{x}) \\
&= \theta\mathrm{x}, \quad \mathrm{x} \in \mathbb{R}^{n\times 1}, \theta \in \mathbb{R}
\end{aligned}$$

(13) To write down y as a function of $\mathrm{x}\theta$ and all other necessary parts of the assumed model in keeping with Equation (2) is as follows:

$$\begin{aligned}
y &= f(\mathrm{x}) + \epsilon \quad \text{where } \epsilon \sim \mathcal{N}(0, \sigma^2\mathcal{I}_n) \\
&= \theta\mathrm{x} + \epsilon \quad \mathrm{x} \in \mathbb{R}^{n\times 1} \\
\text{and } \mathrm{p}(y, \mathrm{x}, \sigma^2) &= \underset{=}{1} e^{\frac{-((y-\theta\mathrm{x})^2}{2\sigma^2}}
\end{aligned}$$

(14) To write $\hat{y}$ as function of x and y is as follows:

$$\begin{aligned}
\hat{y} &= \mathrm{x}\hat{\theta} \\
&= \mathrm{x}\{(\mathrm{x}^{\mathrm{T}}\mathrm{x})^{-1}\mathrm{x}^{\mathrm{T}}y\}
\end{aligned}$$

(15) To find variance $[y_i|\mathrm{x}_i] = \mathbb{V}[y_i|\mathrm{x}_i] \forall i \in [n]$ and deduce variance$(y|\mathrm{x})$

$$\begin{aligned}\mathbb{V}|\mathrm{x}_i] &= \mathbb{V}[f(\mathrm{x}_i) + \epsilon_i] \\ &= \ \mathbb{V}[f(\mathrm{x}_i)] + \mathbb{V}[\epsilon_i] \\ &= 0 + \sigma^2\end{aligned}$$

$$\begin{aligned}\mathbb{E}(\hat{y}_i|\mathrm{x}_i) =&\mathbb{E}[(X^T X)^{-1} X^{\mathrm{T}} y \mathrm{x}_i] \\ =&\theta \mathrm{x}_i (\mathrm{x}_i^{\mathrm{T}} \mathrm{x})^{-1} \mathrm{x}^{\mathrm{T}} \mathrm{x} \\ =&\theta \mathrm{x}_i\end{aligned}$$

Deducing the mean of $y|x$, we have $\mathbb{E}(\hat{Y}|x) = \theta \mathrm{x}$.

(16)

$$\begin{aligned}\mathbb{V}(\hat{Y}_i|\mathrm{x}_i) =&\mathbb{V}(\hat{\theta}\mathrm{x}_i|\mathrm{x}_i) \\ =&\mathbb{V}(X_i(X^T X)^{-1} X^T(\theta X + \epsilon)) \\ =&\mathbb{V}(\theta x_i (X^T X)^{-1} X^T X + x_i (X^T X)^{-1} X^T \epsilon) \\ =&\mathbb{V}(\mathrm{x}_i (X^T X)^{-1} X^T \epsilon) \\ =&\mathbb{V}(x_i (X^T X)^{-1} X^T \epsilon) \\ =&\mathrm{x}_i^2((X^T X)^{-1} X^T \mathbb{V}(\epsilon)(X^T X)^{-1} X^T) \\ =&\mathrm{x}_i^2 X (X^T X)^{-1} \sigma^2 \mathbb{I}_n (X^T X)^{-1} X^T \\ =&\mathrm{x}_i^2 \sigma^2 (X^T X)^{-1}\end{aligned}$$

Deducing the variance of $\hat{Y}|x$, we have $\mathbb{V}(\hat{Y}|x) = \sigma^2 \mathcal{I}_n$. Where,

$$\begin{aligned}\mathbb{V}(\hat{Y}|\mathrm{x}) =&\mathbb{V}(X(X^T X)^{-1} X^T X + X(X^T X)^{-1} X^T \epsilon) \\ =&\mathbb{V}(\theta X + (X^T X)^{-1} X^T \epsilon) \\ =&X(X^T X)^{-1} \sigma^2 \mathbb{I}_n X (X^T X)^{-1} X^T \\ =&\sigma X (X^T X)^{-1} X^T \\ =&\sigma^2 \mathcal{I}_n\end{aligned}$$

(17)

$$\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2 (\mathrm{X}^{\mathrm{T}} X)^{-1}\right)$$

(18) $\hat{Y}_i|\mathrm{x}_i \sim \mathbb{N}\left(\theta \mathrm{x}_i, \mathrm{x}_i^2 \sigma^2 (X^T X)^{-1}\right)$
$\hat{Y}|\mathrm{x} \sim \mathcal{N}(\theta X, \sigma^2 \mathbb{I}_n)$.
Where $\mathcal{I}_n$ is the identity matrix with n-dimension.

(19)

$$\begin{aligned}\hat{\sigma}^2 =&\frac{\sum_{i=1}^{2}(Y_i - \hat{Y}_i)^2}{n-p} \\ =&\frac{\sum_{i=1}^{2}(Y_i - \hat{Y}_i)^2}{n-1}\end{aligned}$$

Where p is the number of regression parameters and in our case $p = 1$.

(20)

$$
\begin{aligned}
\mathbb{V}(Y_i|\mathrm{x}_i) =& \mathbb{V}(f(\mathrm{x}_i) + \epsilon_i) \\
=& \mathbb{V}(f(\mathrm{x}_i)) + \mathbb{V}(\epsilon_i) \\
=& 0 + \sigma^2 \qquad \text{Since we have} \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

Deducing the variance of $Y|\mathrm{x}$, we have $\mathbb{V}(Y|\mathrm{x}) = \sigma^2\mathcal{I}_n$.

Exercise 2

(1) Find by all means possible the history and description of this data and comment on it.

(2) Plot the distribution of the response for the dataset and comment.

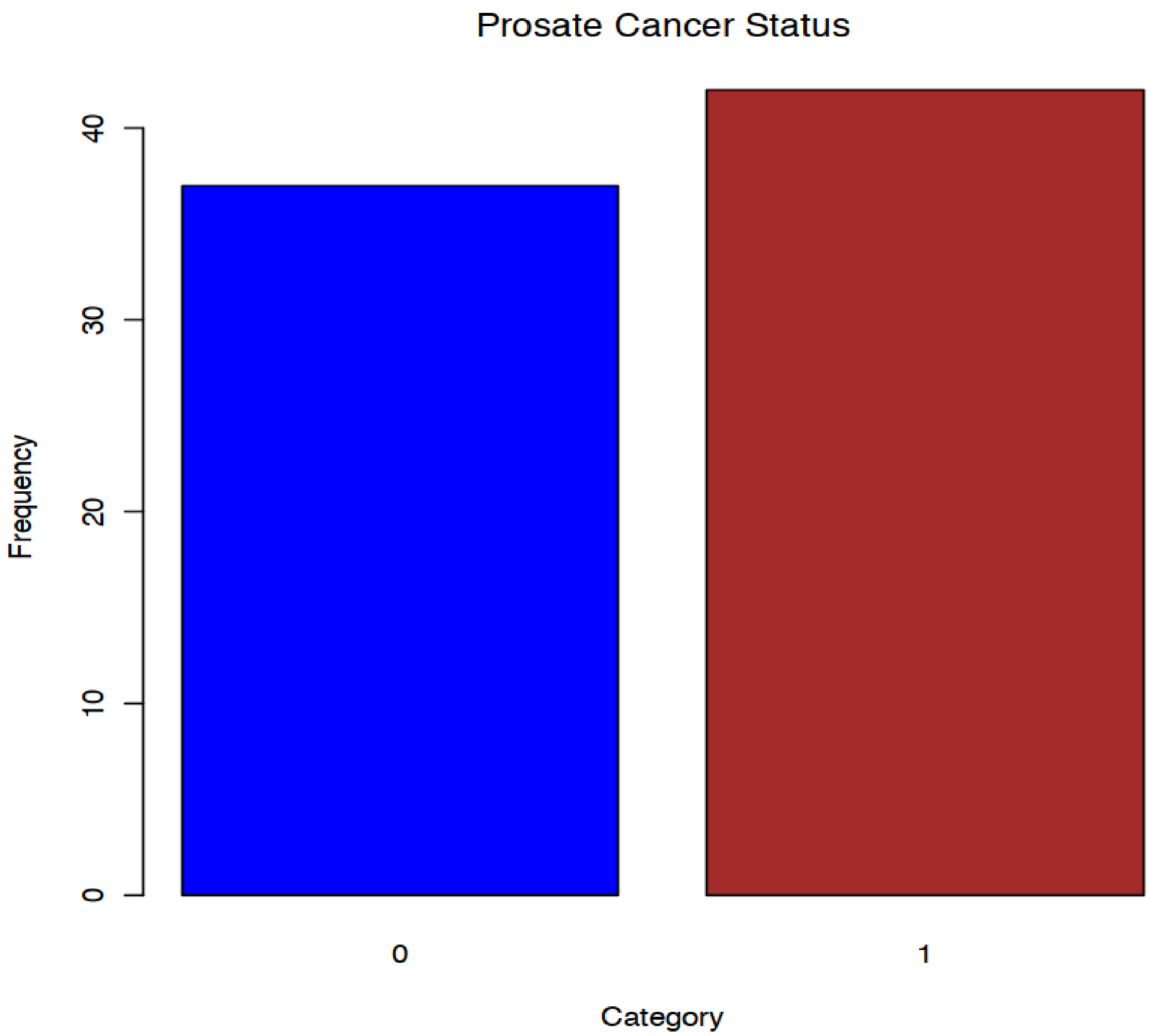

Figure 1: Barchart distribution of response variable

Prostate Cancer Microarray Gene Expression dataset has a size of 79 observations and 500 variables in which 0 indicates people not suffering from prostate cancer while 1 represent people with prostate cancer . And from Figure 1, it means that 37 out of 79 are not suffering while 42 out of 79 are suffering from prostate cancer. In percentage terms, 46.84 % are not suffering while 53.16 % are suffering prostate cancer.

(3) Comment on the shape of this dataset in terms of the sample size and the dimensionality of the input space.
From the observation of the dataset $n = 79$ and $p = 500$, which implies $p >> n$ therefore the dataset is ultra high dimensional.

(4) Comment succinctly from the statistical perspective on the type of data in the input space. It is absolutely crucial here to provide as many details as possible including distributional aspects via boxplots and others on randomly selected subsets of variables.

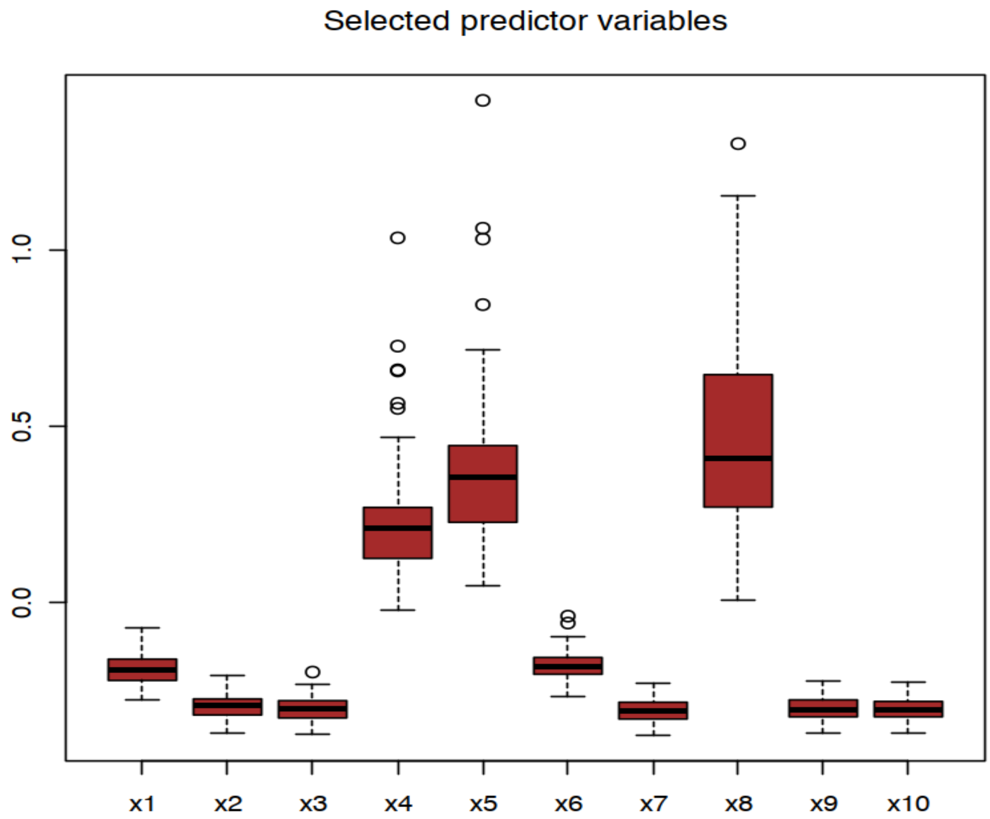

Figure 2: A distribution bloxplot of selected independent variables

- From the figure 3, we can clearly observe that the median line from selected variables x_2, x_3, x_7, x_9, x_10 are almost equal, hence being symmetrically distributed (Normally distributed).
- Expect variable x_3, x_4, x_5, x_6, x_8, we can see that there is no outliers in other selected variables.
- Apart from x_4, x_5, x_8, all other remaining variables are in the same range.

Exercise 3

Given

$$X = \begin{bmatrix} 1 & -2 \\ -2 & 1 \\ 4 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} -5 \\ 4 \\ -3 \end{bmatrix}$$

(1) Compute $X^{\mathrm{T}}X$

$$\begin{aligned} X^{\mathrm{T}}X &= \begin{bmatrix} 1 & -2 & 4 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 1 \\ 4 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 21 & 0 \\ 0 & 6 \end{bmatrix} \end{aligned}$$

(2) Comment on the shape of $X^{\mathrm{T}}X$

- $X^{\mathrm{T}}X$ is a diagonal matrix.

(3) Find $(X^{\mathrm{T}}X)^{-1}$ in the most straightforward way

$$(X^{\mathrm{T}}X)^{-1} = \mathrm{Diag}(u) \quad \text{where } u = \left(\tfrac{1}{21}, \tfrac{1}{6}\right)^{\mathrm{T}}$$

(4) Compute $\hat{\theta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y$

$$X^{\mathrm{T}}Y = \begin{bmatrix} 1 & -2 & 4 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} -5 \\ 4 \\ -3 \end{bmatrix} = \begin{bmatrix} -25 \\ 11 \end{bmatrix}$$

$$(X^{\mathrm{T}}X)^{-1} = \begin{bmatrix} \frac{1}{21} & 0 \\ 0 & \frac{1}{6} \end{bmatrix}$$

$$\begin{aligned} (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y &= \begin{bmatrix} \frac{1}{21} & 0 \\ 0 & \frac{1}{6} \end{bmatrix} \begin{bmatrix} -25 \\ 11 \end{bmatrix} \\ &= \begin{bmatrix} \frac{-25}{21} \\ \frac{11}{6} \end{bmatrix} \end{aligned}$$

(5) Compute the vector $\hat{Y}$ of estimated responses

$$\begin{aligned} \hat{Y} &= X\hat{\Theta} \\ &= \begin{bmatrix} 1 & -2 \\ -2 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} \frac{-25}{21} \\ \frac{11}{6} \end{bmatrix} \\ &= \begin{bmatrix} \frac{-34}{7} \\ \frac{59}{14} \\ \frac{-41}{14} \end{bmatrix} \end{aligned}$$

(6) Compute the vector e of the residual values

$$e == Y - \hat{Y}$$
$$= \begin{bmatrix} -5 \\ 4 \\ -3 \end{bmatrix} - \begin{bmatrix} \frac{-34}{7} \\ \frac{59}{14} \\ \frac{-41}{14} \end{bmatrix} \qquad = \begin{bmatrix} \frac{-1}{7} \\ \frac{-3}{14} \\ \frac{-1}{14} \end{bmatrix}$$

(7) Find the value of $\mathrm{SSE}(\hat{\theta})$

$$\begin{aligned} \mathrm{SSE}(\hat{\theta}) &= \sum (Y - \hat{Y})^2 \\ &= \frac{1}{49} + \frac{9}{196} + \frac{1}{196} \\ &= \frac{1}{14} \end{aligned}$$

(8) Find the estimated $\hat{\sigma^2}$

$$\hat{\sigma^2} = \frac{\mathrm{SSE}(\hat{\theta})}{n-2}$$

Since $n = 3$

$$\frac{\mathrm{SSE}(\hat{\theta})}{n-2} = \frac{1}{14}$$

(9) Find and write down Variance($\hat{\sigma^2}$)

$$\begin{aligned} \mathrm{Variance}(\hat{\sigma^2}) &= \hat{\sigma^2}(X^{\mathrm{T}}X)^{-1} \\ &= \frac{1}{14}\begin{bmatrix} \frac{1}{21} & 0 \\ 0 & \frac{1}{6} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{294} & 0 \\ 0 & \frac{1}{84} \end{bmatrix} \end{aligned}$$

(10) Write the matrix $(X^{\mathrm{T}}X)$ in a function of the identity matrix when the data matrix is given by:

$$\begin{aligned} X &= \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \\ X^{\mathrm{T}}X &= \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} \end{aligned}$$

In the form of an identity matrix

$$X^{\mathrm{T}}X = 5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Exercise 4

(1) Generate an upper triangular pairwise scatterplot for this data, and comment based onthe scatterplots regarding which of the explanatory variables aremore strongly related tothe response. Can you tell from the plot the strongest of all the predictor variables?

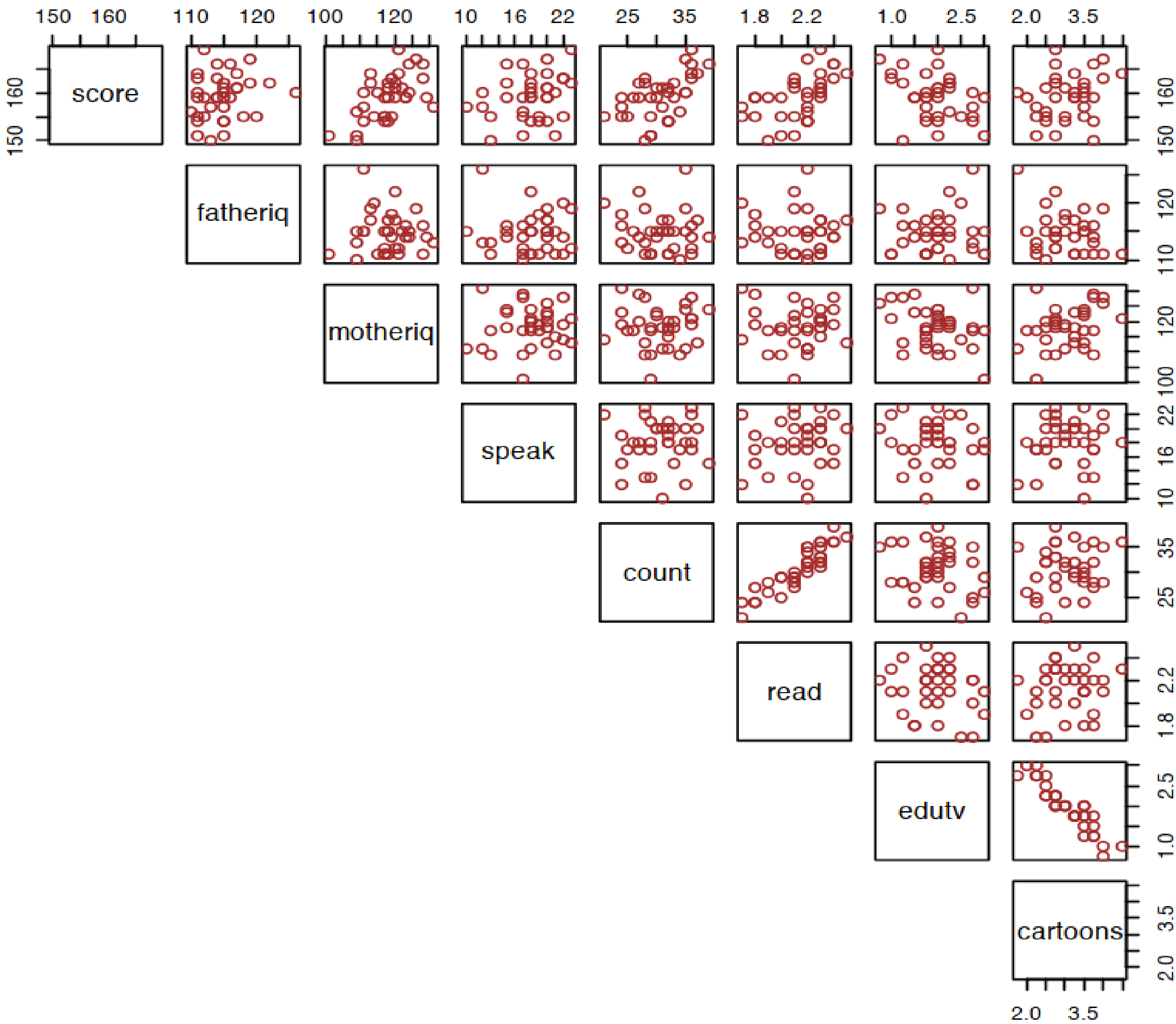

Figure 3: Correlation plot of independent variables to score

From the scatterplots in 3, 3 variables are linearly correlated to score which are **motheriq, count** and **read** amongst which motheriq has the strongest linear correlation to the response score.

(2) Generate the correlation matrix for this data (please do not include the p-values in the-matrix for now). Which variable does the correlation matrix appear to indicate as the strongest?

	score	fatheriq	motheriq	speak	count	read	edutv	cartoons
score	1.00	0.19	0.57	0.27	0.54	0.53	-0.37	0.25
fatheriq	0.19	1.00	-0.02	-0.03	-0.08	-0.07	0.12	-0.25
motheriq	0.57	-0.02	1.00	0.07	0.02	-0.04	-0.33	0.34
speak	0.27	-0.03	0.07	1.00	0.06	0.19	-0.15	0.11
count	0.54	-0.08	0.02	0.06	1.00	0.91	-0.22	0.15
read	0.53	-0.07	-0.04	0.19	0.91	1.00	-0.17	0.13
edutv	-0.37	0.12	-0.33	-0.15	-0.22	-0.17	1.00	-0.92
cartoons	0.25	-0.25	0.34	0.11	0.15	0.13	-0.92	1.0

Table 1: Correlation matrix

From table 1 above, the correlation matrix clearly indicates the predictor variable **motheriq** has the strongest linear correlation with score (the response variable).

(3) Plot a histogram of the response variable and also perform a test of normality on it.

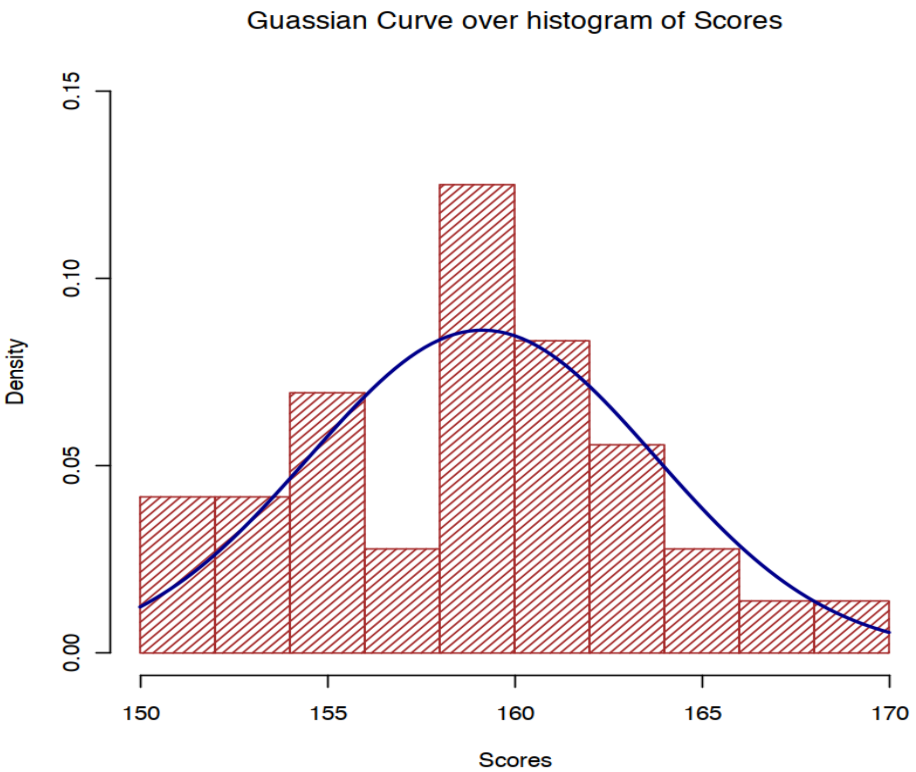

Figure 4: histogram of response variable score

Formal Test of Normality
$\mathbb{H}_0$ The response variable **score** is normally distributed while $\mathbb{H}_1$ **score** is not normally distributed.

```
Shapiro-Wilk normality test
```

```
data:  gifted$score
W = 0.98051, p-value = 0.7625
```

From Figure 4 above and the Shapiro-Wilk Normality test where the p-value = 0.7625 which is greater than the Significance Level $\alpha = 0.05$, we fail to reject $\mathbb{H}_0$ and conclude that the response variable is normally distributed.

(4) Perform a simple linear regression (SLR) model fitting featuring the response and the variable you singled out as the most important.

```
SLR <-  lm(score  motheriq, data = gifted)
```

Simple Linear Regression Model:

$$\hat{score} = 111.0930 + 0.4066 \times \text{motheriq}$$

(5) Generate the 4 residual analysis plots and comment on the suitability of your built model.Are there observations that might have badly influenced the estimation of your parameters? We can infer the following properties of the model from the residual plot

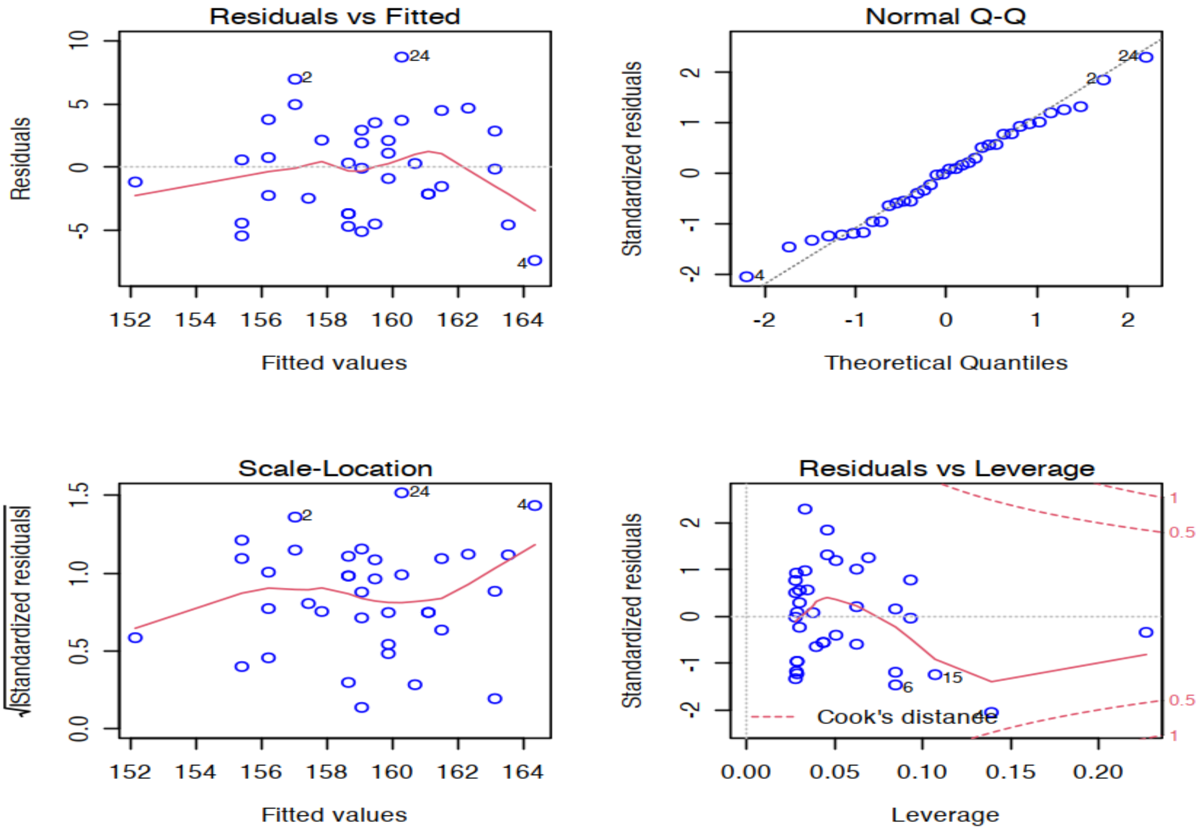

Figure 5: SLR Residual Plot

in Figure 5. The plot of Residuals vs Fitted indicates the linearity assumption is met. Normal Q-Q plot shows the we have normality. From Scale-Location plot the model has homoskedacity and Residuals vs Leverage shows that we have outliers.

(6) Let's assume for a little while that you are to use the above SLR model. Then give an interpretation of your estimated slope in layman's term.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 111.0930    11.8567   9.370 6.02e-11 ***
motheriq      0.4066     0.1002   4.058 0.000274 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.856 on 34 degrees of freedom
Multiple R-squared:  0.3263,Adjusted R-squared:  0.3065
F-statistic: 16.47 on 1 and 34 DF,  p-value: 0.000274
```

From the above summary, a unit increase in **motheriq** will increase **score** by 0.4066.

(7) Generate both confidence bands and the prediction bands for this model and provide intelligent comments on what you get.

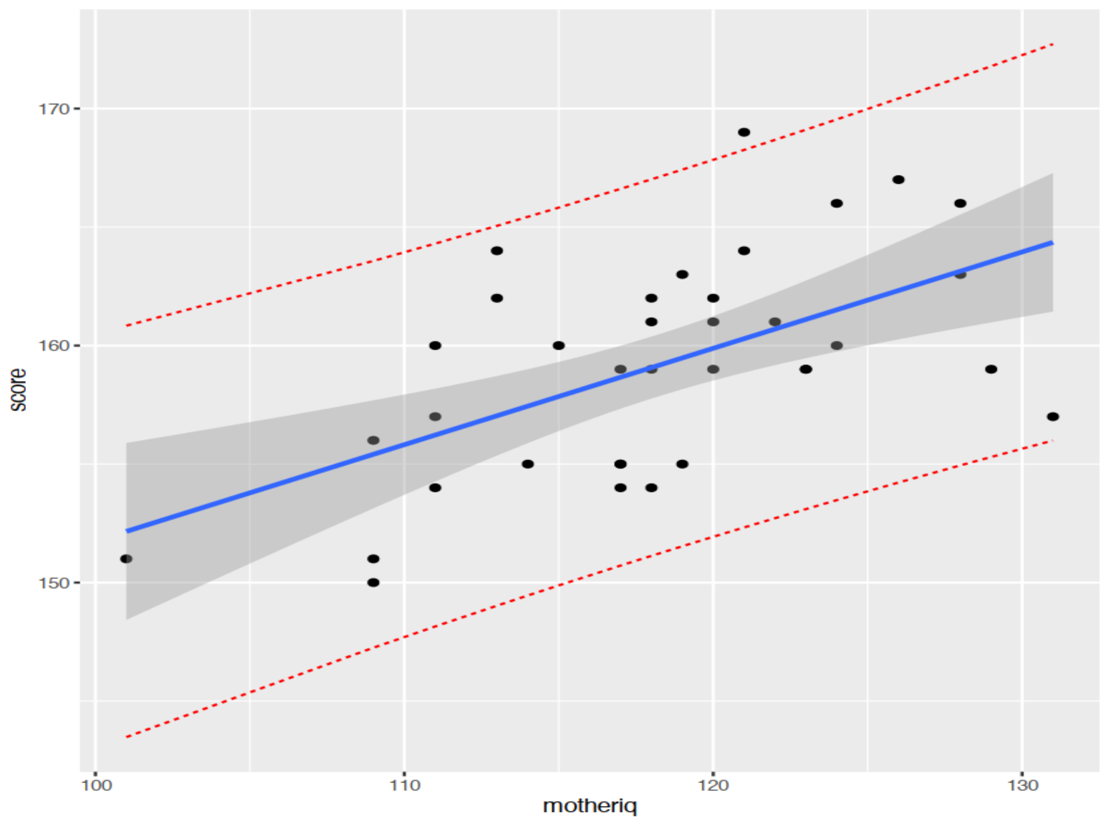

Figure 6: SLR Prediction Band

The Prediction interval (PI) in Figure 6 is an estimate of an interval in which a future observation will fall, with a certain confidence level, given the observations that were already observed. From Figure 7, the confidence interval (lower line and upper red lines)

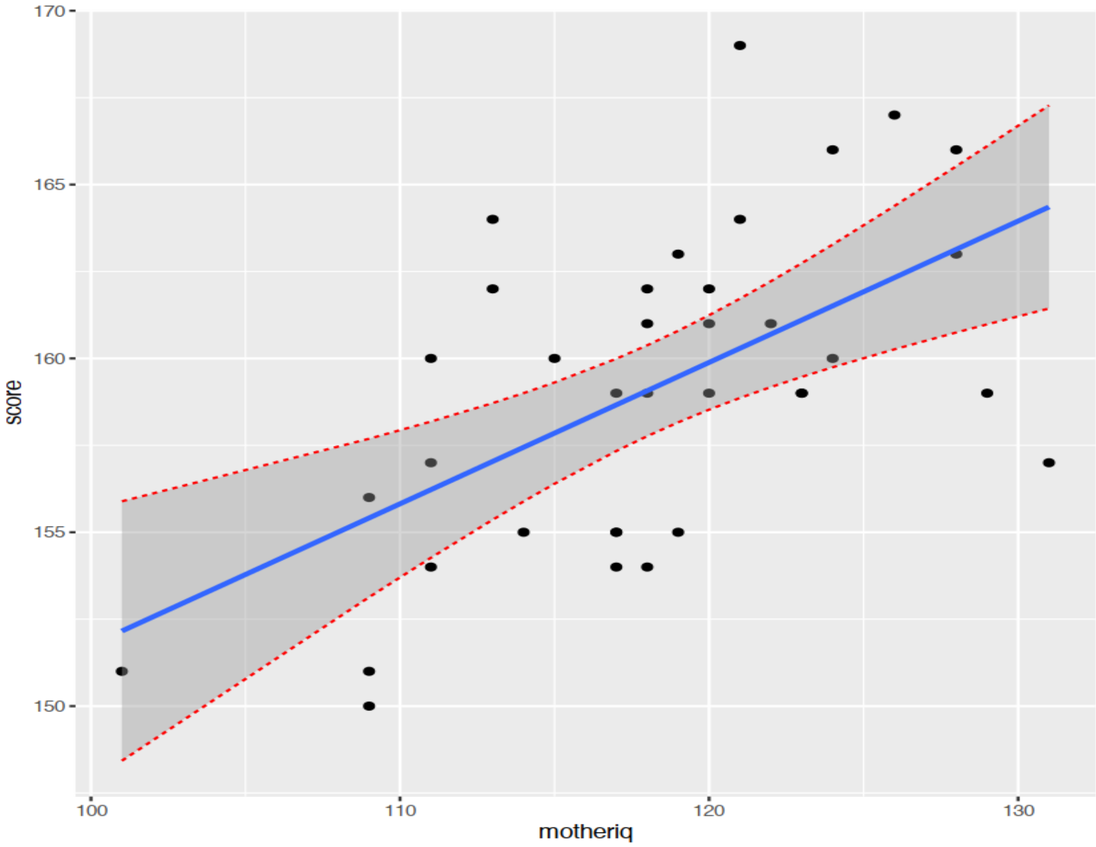

Figure 7: SLR Confidence Band

signifies the range in which the true population parameter lies at a 95% level of confidence. This functionally means is that we're 95% confident that the true regression line lies somewhere in that gray zone.

(8) Build a multiple linear regression (MLR) model for this data comprising of all the provided explanatory variables.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 75.50849   24.02618   3.143  0.00393 **
motheriq     0.40007    0.07291   5.488 7.33e-06 ***
fatheriq     0.25249    0.13756   1.835  0.07707 .
speak        0.18764    0.14767   1.271  0.21429
count        0.20649    0.26631   0.775  0.44462
read         7.54405    5.58640   1.350  0.18769
edutv       -4.20244    2.24503  -1.872  0.07170 .
cartoons    -3.33899    2.01808  -1.655  0.10919
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.591 on 28 degrees of freedom
Multiple R-squared:  0.7496,Adjusted R-squared:  0.687
F-statistic: 11.97 on 7 and 28 DF,  p-value: 5.803e-07
```

From observing the $\Pr(>|t|)$ values of the predictor variables, we can draw a conclusion that **motheriq** has the strongest factor of variation (explanatory power) in the response variable as compared to the others because is has a p-value significantly lower than the Significance Level $\alpha = 0.05$.