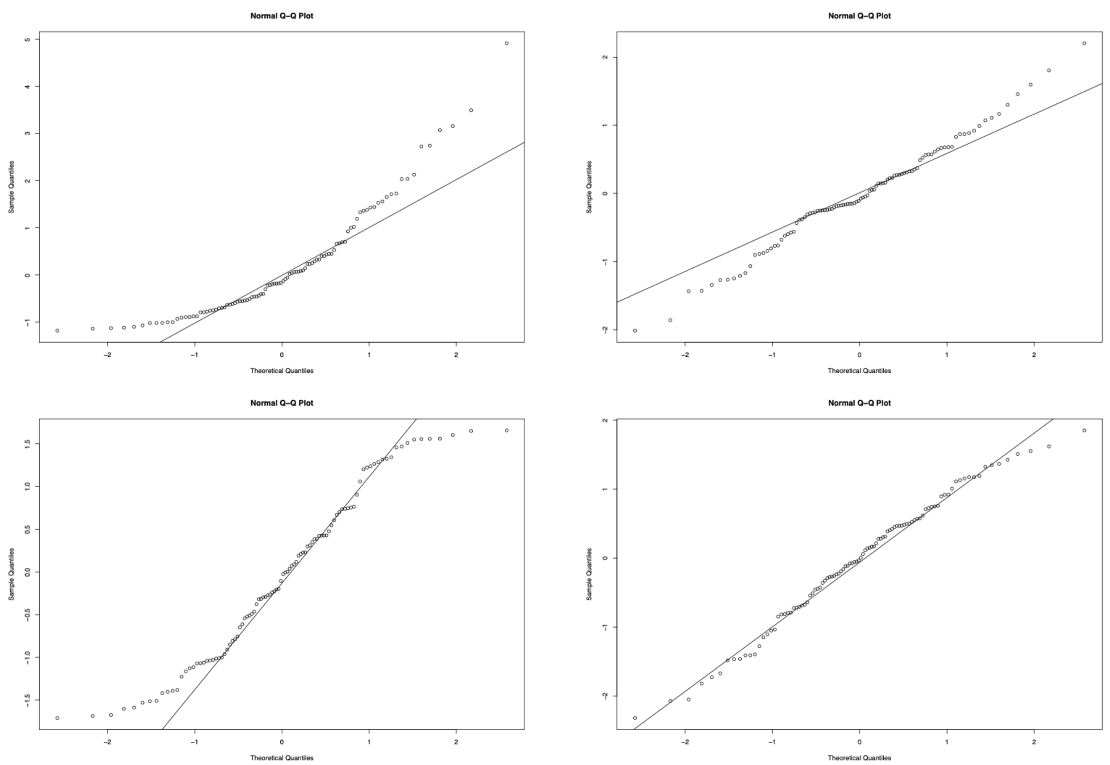
Problem 1
Submitted as a Markdown file.

Problem 2
Q-Q plots. For each of the following Q-Q plots, describe the distribution of the data (whether it is Normal or heavy tailed, etc.).

Figure 1: Q-Q plots



Solution:

(a) This relationship is right-skewed.

(b) This relationship is heavy tailed.

(c) This relationship is heavy tailed.

(d) This looks very normal, with some possible outliers at the percentile ends.

Problem 3
Coefficient of determination. Show that

$$R^2 = r^2, r = \text{sign}\{\beta_1\}\sqrt{R^2}$$

where R^2 is the coefficient of determination when regressing Y onto X and r is the sample correlation coefficient between X and Y.

Solution: Observe,

$$R^2 = \frac{SSR}{SSTO} = \frac{\beta_1^2 \left(\sum_{i=1}^n (X_i - \overline{X})^2\right)^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2} = \frac{\left[\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})\right]^2}{\left(\sum_{i=1}^n (X_i - \overline{X})\right)\left(\sum_{i=1}^n (Y_i - \overline{Y})^2\right)} = r^2$$

and since

$$sign(r) = sign(\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})) = sign(\beta_1)$$

we can conclude that

$$r = \text{sign}\{\beta_1\}\sqrt{R^2}$$

Problem 4
Confirm the formula for inverting a 2×2 matrix.

Solution:

$$\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \left(\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \right) &= \frac{1}{ad-bc} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\
&= \frac{1}{ad-bc} \begin{pmatrix} ad-bc & -ab+ab \\ cd-cd & -bc+ad \end{pmatrix} \\
&= \frac{1}{ad-bc} \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

Problem 5
Projection matrices. Show the following are projection matrices, i.e., being symmetric and idempotent. What are the ranks of these matrices? Here $\mathbf{H}$ is the hat matrix from a simple linear regression model with n cases (where the X values are not all equal) , $\mathbf{I_n}$ is the $n \times n$ identify matrix, and $\mathbf{J_n}$ is the $n \times n$ matrix with all ones.

(a) $\mathbf{I_n} - \mathbf{H}$

(b) $\mathbf{I_n} - \frac{1}{n}\mathbf{J_n}$

(c) $\mathbf{H} - \frac{1}{n}\mathbf{J_n}$

Solution:

(a)
- *Idempotent*:

$$\begin{aligned}
(I_n - H)^2 &= I_n^2 - X(X^TX)^{-1}X^T - X(X^TX)^{-1}X^T + X(X^TX)^{-1}X^T \cdot X(X^TX)^{-1}X^T \\
&= I_n - 2H + X\left[(X^TX)^{-1}X^TX\right](X^TX)^{-1}X^T \\
&= I_n - 2H + X(X^TX)^{-1}X^T \\
&= I_n - 2H + H \\
&= I_n - H
\end{aligned}$$

- *Symmetric*: $(I_n - H)' = I_n' - H' = I_n - H$
- *Rank*: $n - 2$

(b)
- *Idempotent*: $(I_n - \frac{1}{n}J_n)^2 = I_n^2 - 2\frac{1}{n}J_n + \frac{1}{n^2} \cdot \mathbf{1} \cdot \mathbf{1^T} \cdot \mathbf{1} \cdot \mathbf{1^T} = \mathbf{I_n} - \mathbf{2}\frac{\mathbf{1}}{\mathbf{n}}\mathbf{J_n} + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J_n} = \mathbf{I_n} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J_n}$
- *Symmetric*: $(I_n - \frac{1}{n}J_n)' = I_n' - \frac{1}{n}J_n' = I_n - \frac{1}{n}(\mathbf{1} \cdot \mathbf{1^T})^\mathbf{T} = \mathbf{I_n} - \frac{\mathbf{1}}{\mathbf{n}}((\mathbf{1^T})^\mathbf{T} \cdot \mathbf{1^T} = \mathbf{I_n} - \frac{\mathbf{1}}{\mathbf{n}}(\mathbf{1} \cdot \mathbf{1^T} = \mathbf{I_n} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J_n}$

- *Rank*: $n-1$

(c)
- *Idempotent*:

$$\begin{aligned}(H - \frac{1}{n}J_n)^2 &= H^2 - \frac{1}{n}HJ_n - \frac{1}{n}J_nH + \frac{1}{n}J_n^2 \\ &= H - \frac{1}{n}HJ_n - \frac{1}{n}J_nH + \frac{1}{n}J_n \\ &= H - \frac{1}{n}X(X^TX)^{-1}X^T\mathbf{11^T} - \mathbf{\frac{1}{n}11^TX(X^TX)^{-1}X^T} + \mathbf{\frac{1}{n}J_n}\end{aligned}$$

- *Symmetric:* $(H - \frac{1}{n}J_n)' = H' - \frac{1}{n}J_n' = H - \frac{1}{n}J_n$
- *Rank:* 1

Problem 6
Under the simple linear regression model, using matrix algebra, show that:

(a) The residual vector $\mathbf{e}$ is uncorrelated with the fitted values vector $\hat{\mathbf{Y}}$ and the LS estimator $\hat{\beta}$. *Hint: If* $\mathbf{Z}$ *is an* $r \times 1$ *random vector,* $\mathbf{A}$ *is an* $s \times r$ *non-random matrix, and* $\mathbf{B}$ *is a* $t \times r$ *non-random matrix, then* $Cov(\mathbf{AZ}, \mathbf{BZ}) = \mathbf{A}\sigma^{\mathbf{2}}\{\mathbf{Z}\}\mathbf{B}'$*.)*

(b) With Normality assumption on the error terms, SSE is independent with the LS estimator $\hat{\beta}$ and SSR.

Solution:

(a)
- Observe,

$$\begin{aligned}Cov(e, \hat{Y}) &= Cov((1-H)Y, HY) \\ &= (1-H)\sigma^2(Y)H^T \\ &= (1-H)\sigma^2(Y)H \\ &= \sigma^2(Y)(1-H)H && \text{since } \sigma^2(Y) = \sigma^2 I \\ &= \sigma^2(Y)(H - H^2) \\ &= \sigma^2(Y)(H-H) \\ &= 0\end{aligned}$$

-

$$\begin{aligned}Cov(e, \hat{\beta}) &= Cov((I-H)Y, (X^TX)^{-1}X^TY) \\ &= (I-H)\sigma^2(Y)X(X^TX)^{-T} && \text{since } \sigma^2(Y) = \sigma^2 I \\ &= \sigma^2(Y)(I - X(X^TX)^{-1}X^T)X(X^TX)^{-T} \\ &= \sigma^2(Y)(X(X^TX)^{-T} - X(X^TX)^{-T}) \\ &= 0\end{aligned}$$

(b) Define the function:

$$SSE = f(e) = e^Te = \|(I-H)Y\|_2$$

Therefore, if we let $g = id : \mathbb{R}^n \to \mathbb{R}$, then we see that by the fact that if two sets of random variables, say $(Z_1, \ldots, Z_s)$ and $(W_1, \ldots, W_t)$, are independent with each other, then their functions, say $f(Z_1, \ldots, Z_s)$ and $g(W_1, \ldots, W_t)$, are independent, then it must follow that

$$e \perp\!\!\!\perp \hat{\beta} \implies f(e) \perp\!\!\!\perp g(\hat{\beta}) \implies SSE \perp\!\!\!\perp \hat{\beta}$$

Similarly, consider the function:

$$SSR = g(\hat{\beta}, \hat{Y}) = \left\| (H - \frac{1}{n} J_n) Y \right\|_2 = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$$

Then it must follow that

$$e \perp\!\!\!\perp \hat{\beta} \text{ and } \hat{Y} \implies f(e) \perp\!\!\!\perp g(\hat{\beta}, \hat{Y}) \implies SSE \perp\!\!\!\perp SSR$$