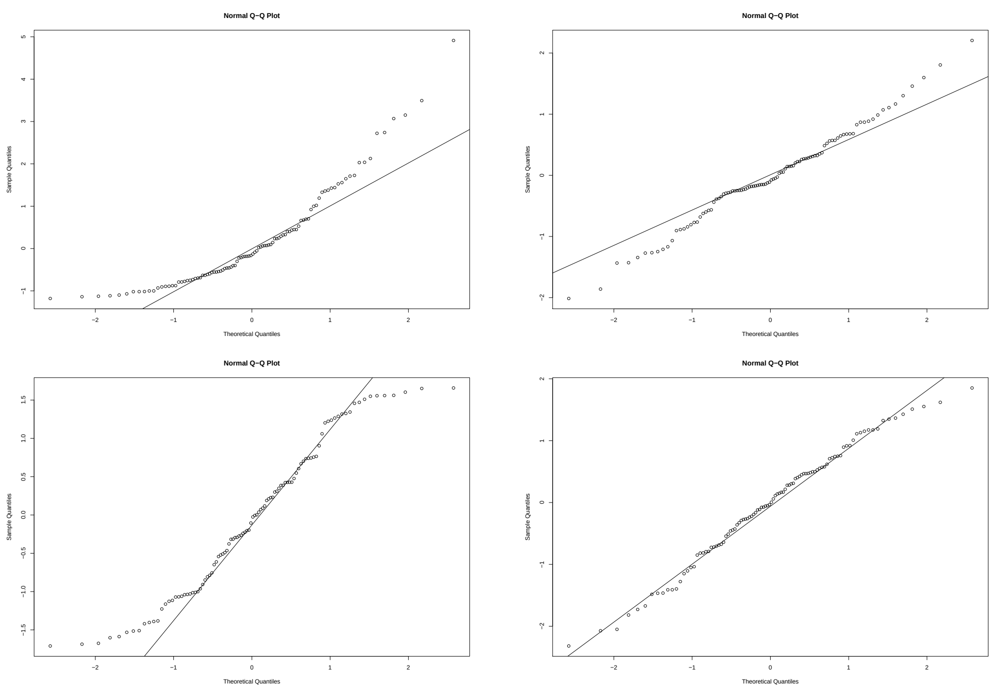
Statistics 206

Homework 3 (Solution)

1. **Q-Q plots.** For each of the Q-Q plot in Figure 1, describe the distribution of the data (whether it is Normal or heavy tailed, etc.).

Figure 1: Q-Q plots



Looking at it in anticlockwise fashion,

- Top left: right skewed
- Bottom left: light tailed
- Bottom right: approximately normal
- Top right : heavy tailed

2. **Coefficient of determination.** Show that

$$R^2 = r^2, \quad r = \text{sign}\{\hat{\beta}_1\}\sqrt{R^2},$$

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

Proof.

$$R^2 = SSR/SSTO = \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 / \sum (y_i - \bar{y})^2$$
$$= (\sum (x_i - \bar{x})(y_i - \bar{y}))^2 / \sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 = r^2$$

□

3. Confirm the formula for inverting a 2×2 matrix.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Check if the following equality holds.

$$\begin{aligned} & \tfrac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ = & \tfrac{1}{ad-bc} \begin{bmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{bmatrix} \\ = & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

4. **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).

(a) $\mathbf{I}_n - \mathbf{H}$

$$(\mathbf{I}_n - \mathbf{H})' = \mathbf{I}_n' - \mathbf{H}' = \mathbf{I}_n - \mathbf{H}$$
$$(\mathbf{I}_n - \mathbf{H})^2 = \mathbf{I}_n^2 - \mathbf{I}_n\mathbf{H} - \mathbf{H}\mathbf{I}_n + \mathbf{H}^2 = \mathbf{I}_n - \mathbf{H}$$

It projects a vector onto the linear subspace of $\mathbf{R}^n$ that is orthogonal to the column space of X. Its rank is $n - p = n - 2$.

(b) $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$

$$(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)' = \mathbf{I}_n' - \frac{1}{n}\mathbf{J}_n' = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$$
$$(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n)^2 = \mathbf{I}_n^2 - \mathbf{I}_n\frac{1}{n}\mathbf{J}_n - \frac{1}{n}\mathbf{J}_n\mathbf{I}_n + \frac{1}{n^2}\mathbf{J}_n^2 = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$$

It projects a vector onto the linear subspace of $\mathbf{R}^n$ that is orthogonal to the subspace spanned by $\mathbf{1}_n$. Its rank is $n - 1$.

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

$$(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)' = \mathbf{H}' - \frac{1}{n}\mathbf{J}_n' = \mathbf{H} - \frac{1}{n}\mathbf{J}_n$$

For the rest, notice that $\mathbf{H}\mathbf{J}_n = \mathbf{J}_n$ because $\mathbf{H}$ is the projection matrix onto the column space of X and every column of $\mathbf{J}_n$, namely $\mathbf{1}_n$, is in the column space of X. So,

$$(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)^2 = \mathbf{H} - \frac{1}{n}\mathbf{J}_n\mathbf{H} - \mathbf{H}\frac{1}{n}\mathbf{J}_n + \frac{1}{n^2}\mathbf{J}_n^2 = \mathbf{H} - \frac{1}{n}\mathbf{J}_n\mathbf{H} - \mathbf{H}\frac{1}{n}\mathbf{J}_n + \frac{1}{n}\mathbf{J}_n = \mathbf{H} - \frac{1}{n}\mathbf{J}_n,$$

where $\mathbf{J}_n = \mathbf{J}_n\mathbf{H}$ follows from

$$\mathbf{J}_n\mathbf{H} = \mathbf{J}_n^t\mathbf{H}^t = (\mathbf{H}\mathbf{J}_n)^t = \mathbf{J}_n^t = \mathbf{J}_n.$$

It projects a vector onto the linear subspace of column space of X that is orthogonal to the subspace spanned by $\mathbf{1}_n$. Its rank is $p - 1 = 1$.

5. Under the simple linear regression model, using matrix algebra, show that:

   The residuals vector $\mathbf{e}$ is uncorrelated with the fitted values vector $\hat{\mathbf{Y}}$ and the LS estimator $\hat{\boldsymbol{\beta}}$.

   Proof.

$$e = (I - \mathbf{H})Y, \quad \hat{\beta} = (X'X)^{-1}X'Y$$

$$\mathrm{Cov}(e, \hat{\beta}) = (I - \mathbf{H})\mathrm{Cov}(Y)((X'X)^{-1}X')' = \sigma^2(I - \mathbf{H})X(X'X)^{-1} = 0,$$

   since $(I - \mathbf{H})X = X - X = 0$. Therefore $\hat{\beta}$ and the residuals e are uncorrelated.

   - $\hat{Y} = X\hat{\beta}$. Hence, $\mathrm{Cov}(\hat{Y}, e) = \mathrm{Cov}(X\hat{\beta}, e) = X\mathrm{Cov}(\hat{\beta}, e) = 0$. Therefore $\hat{Y}$ and the residuals e are uncorrelated.
   - (Alternative)

$$\hat{Y} = \mathbf{H}Y, \quad e = (I - \mathbf{H})Y$$

$$\mathrm{Cov}(\hat{Y}, e) = \mathrm{Cov}(\mathbf{H}Y, (I - \mathbf{H})Y) = H\mathrm{Cov}(Y)(I - \mathbf{H})^t = \sigma^2\mathbf{H}(I - \mathbf{H}) = 0$$

     since $\mathbf{H}(I - \mathbf{H}) = \mathbf{H} - \mathbf{H} = 0$. Therefore $\hat{Y}$ and the residuals e are uncorrelated.

$\square$