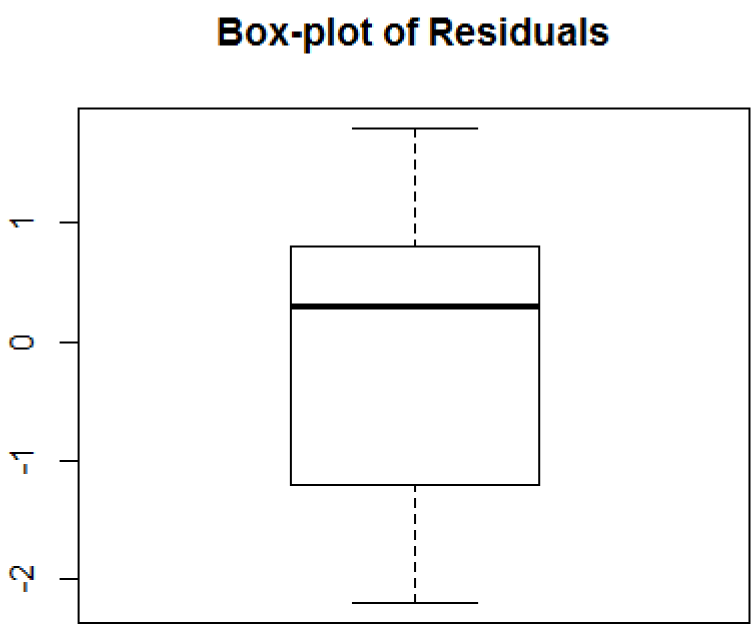
MA 542: Homework 4 Solutions

1. (a) Full model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$. Reduced model: $Y_i = \beta_0 + \epsilon_i$.

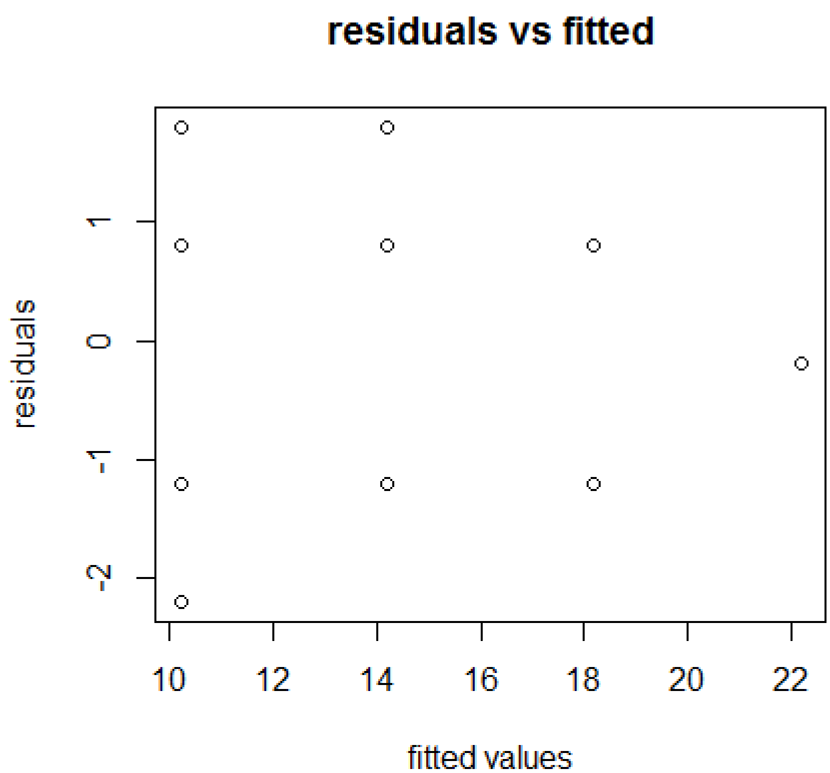
   (b) $SSE(F) = 455,273,165$. $SSE(R) = 548,736,108$. $df_F = 82$. $df_R = 83$.
   $F^* = \frac{SSE(R)-SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} = 16.82$.
   Decision rule: If $F^* < F$, conclude the reduced model. If $F^* > F$, conclude the full model.

2. (a) Box-plot of the residuals:



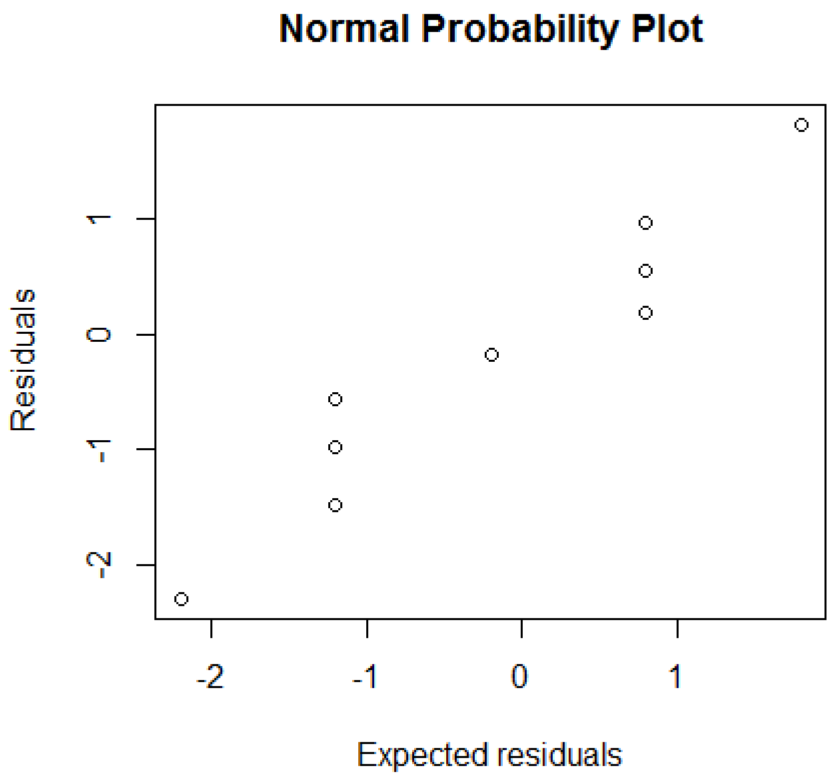
   The box plot shows that the residuals are a bit asymmetric and a range roughly from -2.2 to 1.8.

(b) Plot residuals vs fitted values:



The variance of the residuals appears to decrease as the fitted values increase. The non-constancy of error variance is highly suspected.

(c) Normal probability plot of the residuals:



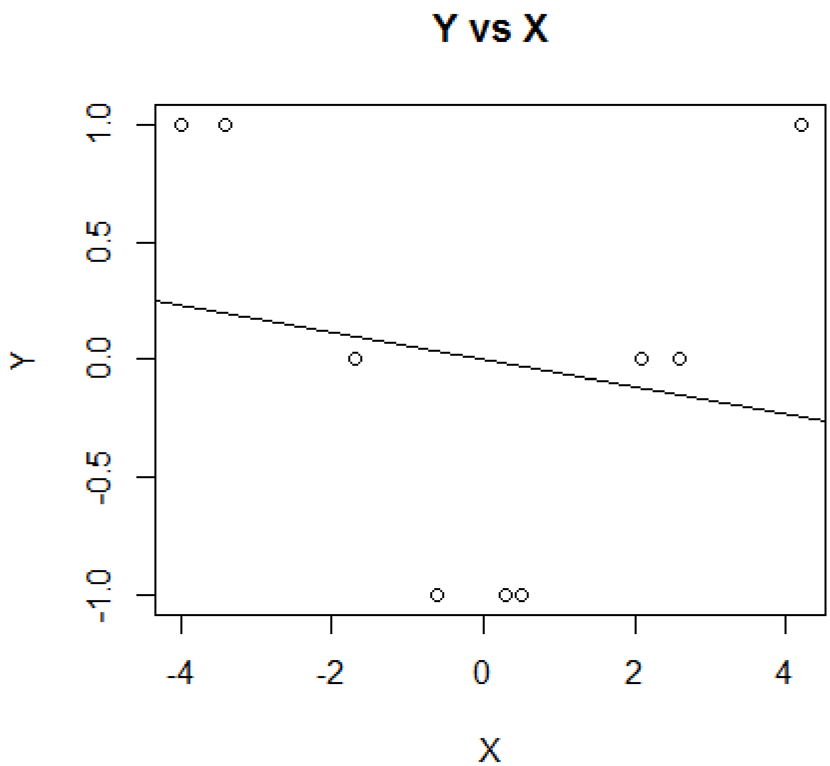
The normal probability plot appears to have a good linear relationship. We conclude that there is no significant departure of normality can be found by normal probability plot.

(d) $r = .9952 >$table value$= .91$ (n=10, $\alpha = .05$), so the normal assumption is not violated significantly.

3. (a) Linear regression function: $\hat{Y}_i = 2.575 - .324X_i$.

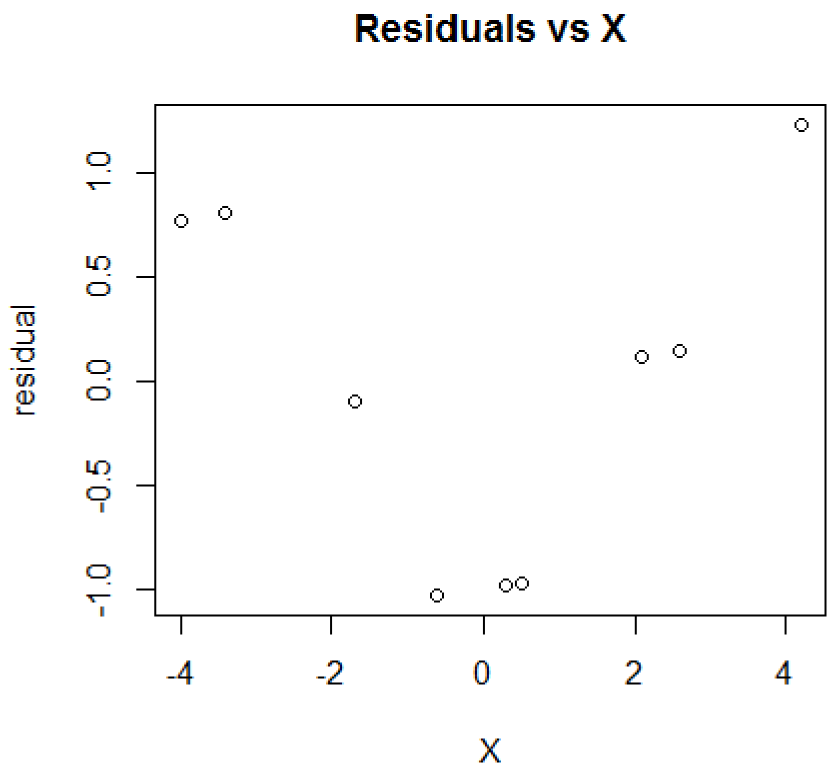
   (b) We test $H_0 : Y_{ij} = \beta_0 + \beta_1 X_i + \epsilon_{ij}$ (reduced) versus H_a: $Y_{ij} = \mu_j + \epsilon_{ij}$ (full). We reject H_0 in favor of H_a if $F^* > F$. Since $F^* = MSLF/MSPE = 58.603 > F(.025, 3, 10)$, we reject H_0 and conclude H_a.

4. (a) Non-linearity of regression function:



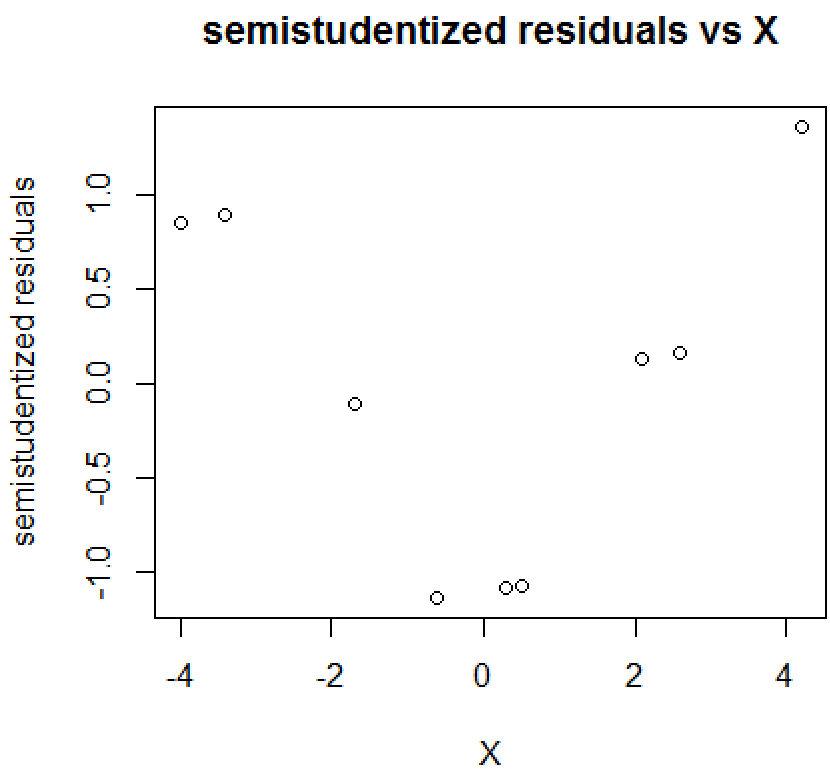
The scatter plot shows certain non-linearity of the regression function.

(b) Non-constancy of error variance:



The residual vs x plot shows that residuals have some patterns, indicating the possibility of non-constancy of error variance.

(c) Presence of outliers:

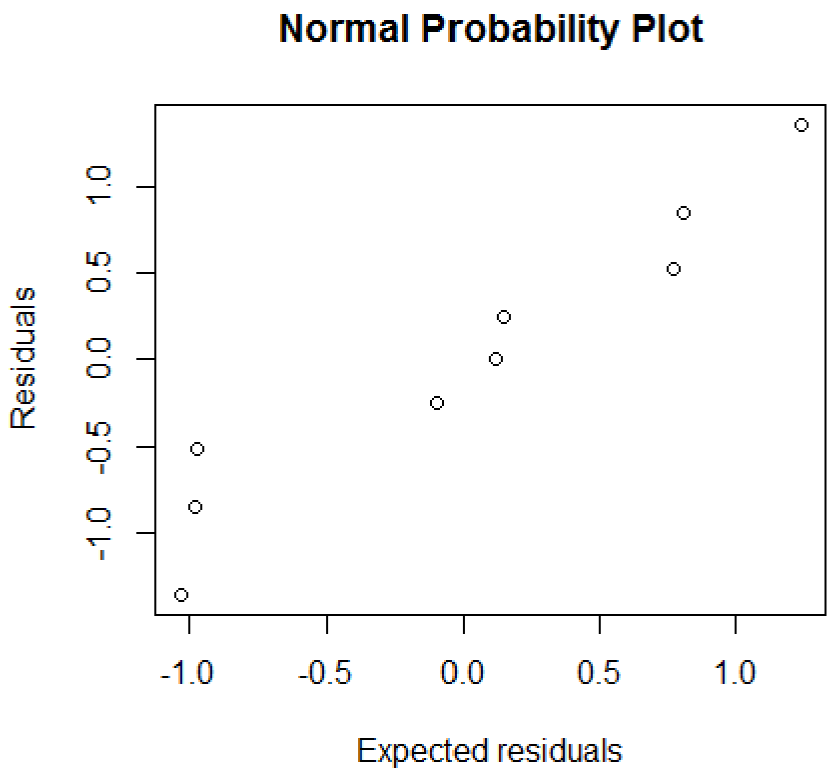


The plot shows that the semi-studentized residuals are all within the range (-4,4), indicating no evidence for presence of outliers.

(d) Non-independence of error terms:

The plot in part (b) appears to follow some patterns, indicating the non-independence of error terms.

(e) Non-normality of error terms:



The normal probability plot shows no evidence for the non-normality of error terms.

Since the residuals failed in most of the departure checks, the SLR model seems to be not reasonable here.

5. If $|t^*| < t_{1-\alpha/2;n-2}$, conclude error variance is constant. If $|t^*| > t_{1-\alpha/2;n-2}$, conclude error variance is not constant. $t^* = .8558 < t = 2.145$, so we conclude error variance is constant.

6. $SSE = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \hat{Y}_{ij})^2 = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j + \bar{Y}_j - \hat{Y}_{ij})^2$
$= \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)^2 + \sum_{j=1}^{c}\sum_{i=1}^{n_j}(\bar{Y}_j - \hat{Y}_{ij})^2 + \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)(\bar{Y}_j - \hat{Y}_{ij})$
$= SSPE + SSLF + \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)(\bar{Y}_j - \hat{Y}_{ij})$.
Next, we show the last term is 0. Since $\bar{Y}_j$ and $\hat{Y}_{ij}$ are both invariant if j is fixed
$\sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)(\bar{Y}_j - \hat{Y}_{ij}) = \sum_{j=1}^{c}(\bar{Y}_j - \hat{Y}_{ij})\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j) = \sum_{j=1}^{c}(\bar{Y}_j - \hat{Y}_{ij}) \cdot 0 = 0.$