Problem: Using the model for *galton* data (Lecture-4 group work), check the LINE conditions and report.
Solution: To test the LINE assumptions for the model:

$$\hat{Y} = 23.94 + 0.65\,X,$$

where Y is the children height in inches and X is the mid-parent height in inches. Remember, that the model is coming from group work-1.

I.) Linear between children and mid-parent height assumption

To test whether there is a linear relation between the children and mid-parent height:
Plot a) in Figure 1 below shows a positive, linear relationship between the children and mid-parent height. There is no obvious outlier or non-linearity.
Plot b) in Figure 1 below shows that the estimated value of children height (using model 1) agrees with the observed (sample) children heights.
Plot c) in Figure 1 below is the residual plot which doesn't show any clear pattern.
Plot d) in Figure 1 below is Tukey's curve which shows barely any deviation from the flat line, indicating absence of non-linearity.
All plots show that linearity assumption is not failing.
Next, we do the Tukey's curve test:

Step 1: Null hypothesis (Ho): Linearity assumption holds
Alternative Hypothesis (Ha): Linearity assumption fails
Step 2: Set Level of significance α as 0.05 (default value)
Step 3: Test statistic= 2.154
Step 4: P-value=0.031
Step 5: If P-value $\leq \alpha$, so we can reject Ho

Conclusion: At 5% level of significance, linearity assumption failed. It is important to note that while we rejected the linearity assumption, the P-value was not very small and of one used $\alpha = 0.01$, we would not reject the linearity assumption.

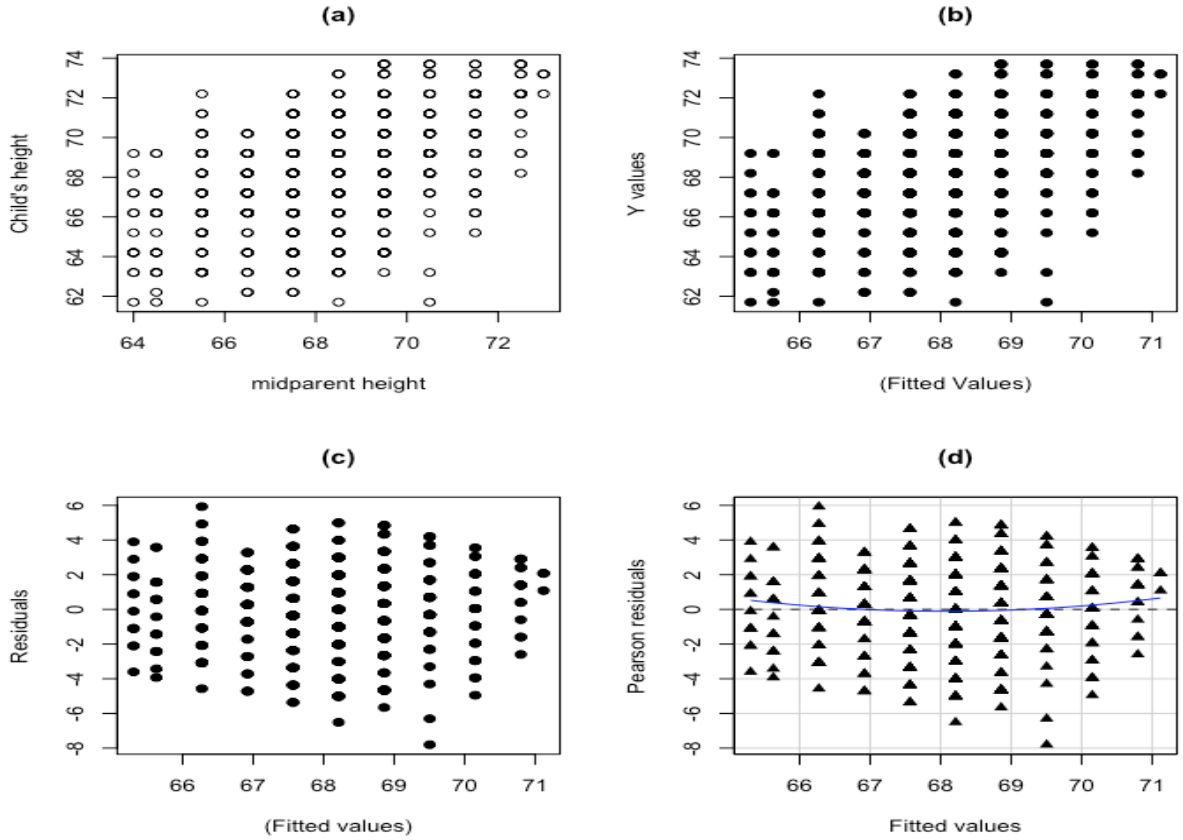

Figure 1: (a) scatterplot of children and mid-parent height, (b): scatterplot of children height and its estimate, (c) Residual plot and (d) Tukey's curve test, basically residual plot with a curve fitted.

II.) Independence of errors assumption

Plot c) in Figure 2 below is the index plot of errors. There is a clear repetitive (cyclic) and positive relation among the errors. Therefore, from this we can see that the independence of errors assumption fails.

III.) Normality of errors assumption

Plot a) in Figure 2 shows histogram of errors which is mostly normal curve with a slight skew to the left.
Plot b) in Figure 2 is the QQ plot and it shows a clear agreement between the sample and normal quantiles as most values are aligned in a strong, positive linear pattern. Therefore, from plots (a) and (b) in Figure 2, we can say that the normality of errors assumption holds true.

IV.) Errors are homoscedastic assumption

Plot d) in Figure 2 shows the residual plot with a mean of errors line. We do not see the variance of errors changing as mid-parent height (our X variable) increases. Thus, there is no evidence of homoscedasticity failing.
Using the Breusch-Pagan test:
Step 1: H0: Errors are homoscedasticity)
Ha: Errors are heteroscedasticity
Step 2: $\alpha = 0.05$ (default value)
Step 3: Test statistic = 0.34256
Step 4: P-value = 0.5584
Step 5: Since P-value > α, so we do not reject the H0.
Conclusion: At 5% level of significance, we have enough evidence that errors are homoscedastic.

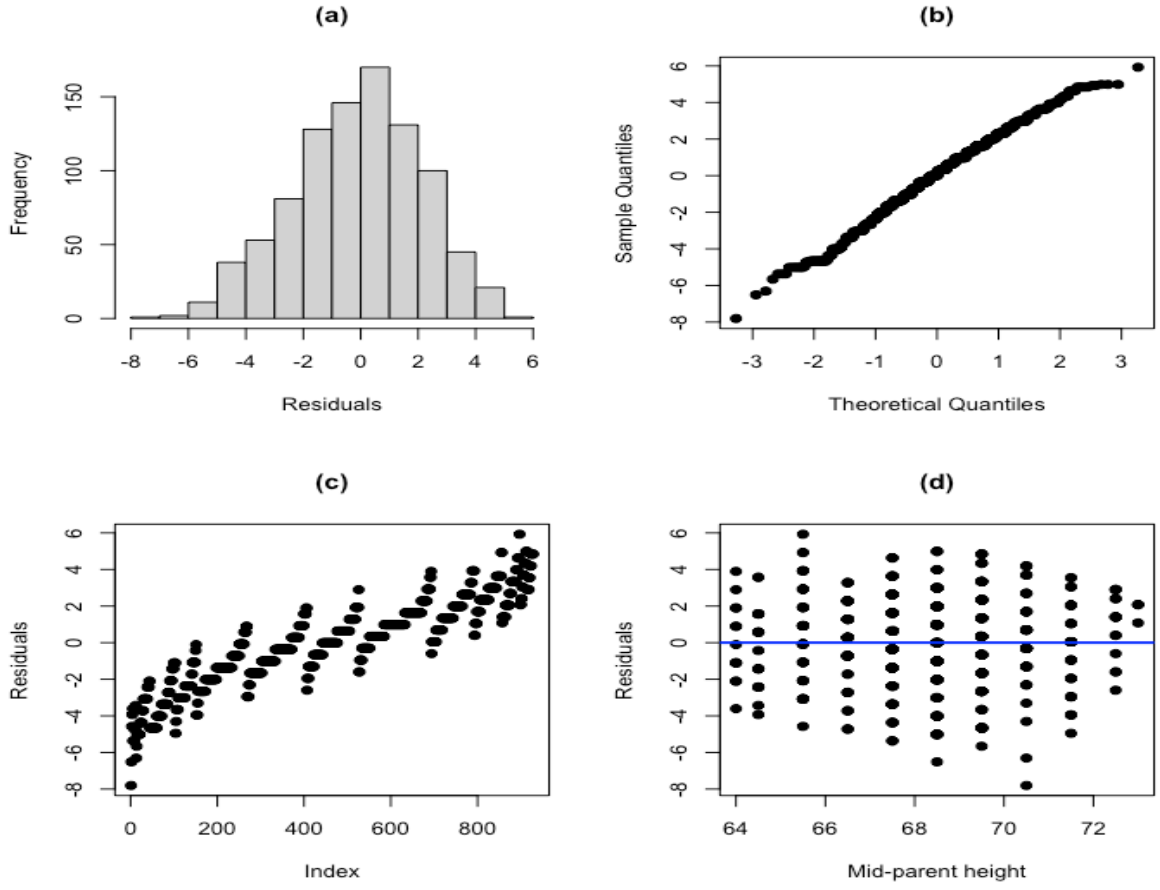

Figure 2: (a) Histogram of errors, (b) QQ plot of errors, (c) Index plot of residuals, and (d) Residual plot.