Derin Gezgin, Johnny Andreasen, Sababa Ahmed
STA 207: Advanced Regression Techniques | Group Work #2

Test LINE assumptions for *galton* data.

Scatterplot of Y versus X | Plot (A)

Fitted Values v. Real Values | Plot (B)

Residual Plot | Plot (C)

Tukey's Test Plot | Plot (D)
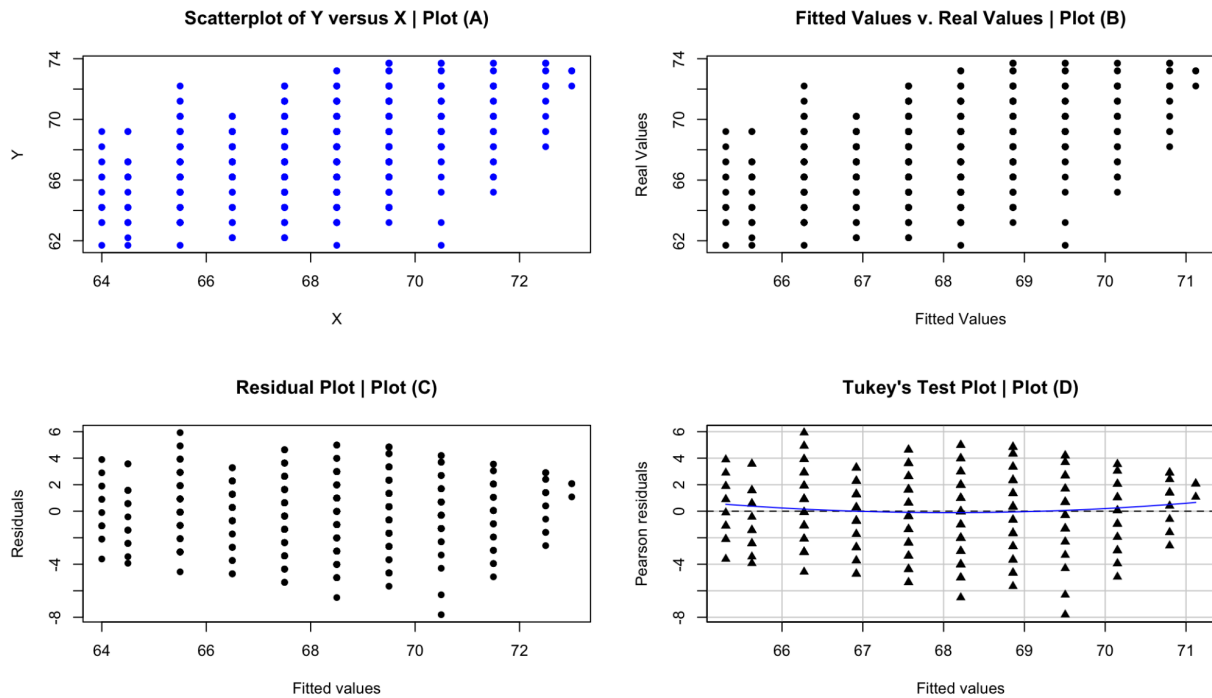
I.   Linearity Between "X" and "Y"

a.   Plot (A) shows …

A relationship in the positive direction. But it's not strongly positive, considering how spread the points are.

b.   Plot (B) shows …

We can see that the fitted values v. Reals Values plot shows a correlation between the predictions.

c.   Plot (C) shows …

There is a pattern in the residuals points to nonlinearity.

d.   Plot (D) shows…

A blue line close to the flat 0 line means that the residuals do not change much, and the linearity assumption holds false, but Tukey's test plot (D) shows that the blue line is close to the 0 line, which shows that the residuals do not change much. This would make the linearity assumption false.
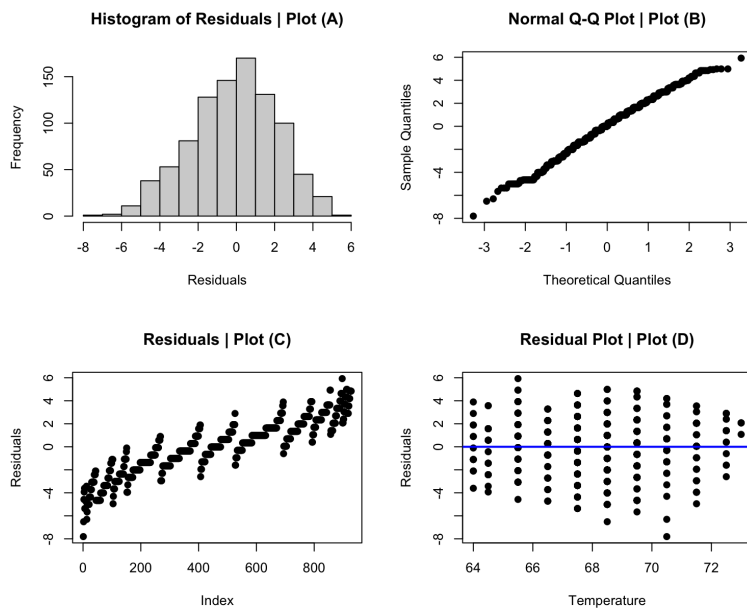
Tukey's Test

-   H_0 = Linearity assumption holds
-   H_a = Linearity assumption does not hold
-   **Significance level =** 0.05 or 5%
-   **Test Statistic = 2.15**
-   **P value = 0.031**
-   **Check: P-Value < 0.05**

- **P-Value is less than the significance level.**

At a 5% level of significance, we have evidence that there is no linear relationship between mid parents' height and the child's height

Independence of Errors

Plot (C) shows errors have an upward, cyclic, repetitive, and linear pattern. This shows a pattern in the residuals, which **fails the independence assumption.**

Normality of Errors

Plot (A) shows there's a normal distribution, with a slight left skew, in the residuals.

Plot (B) shows there's a 45 degree linear relation in the Q-Q plot. This means that the sample quantiles and the theoretical quantiles agree.

Thus, from plots (A) and (B), we can say that the Normality Assumption holds.

Errors are Homoscedastic
There's no significant spread out in the plot (D) so we can say that there is no heteroskedasticity.

BP test
- **H_0** = Error variance is constant, homoscedasticity
- **H_A** = Error variance is not constant, heteroscedasticity
- **Confidence level** = 0.05 or 5%
- **Test statistic = 0.34**

Derin Gezgin, Johnny Andreasen, Sababa Ahmed
STA 207: Advanced Regression Techniques | Group Work #2

- **P value = 0.558**
- **Check P- Value vs confidence level**
    - **P value > confidence level**

At a 5% level of significance we have evidence that the residuals are homoscedastic