MSIS 638

Case 5.3a

Jia Liang Ma

1. Remember that the first and second assumptions of the regression analysis implies that residuals (or errors) must have constant variance for different values of the explanatory variable. The below figure shows three behaviors of residuals. Notice that although residuals have a mean of zero in all three cases, only the third one shows a constant variance for residuals. The first and second cases show a varying constant for residuals. This situation is referred to as *heteroscedasticity*.

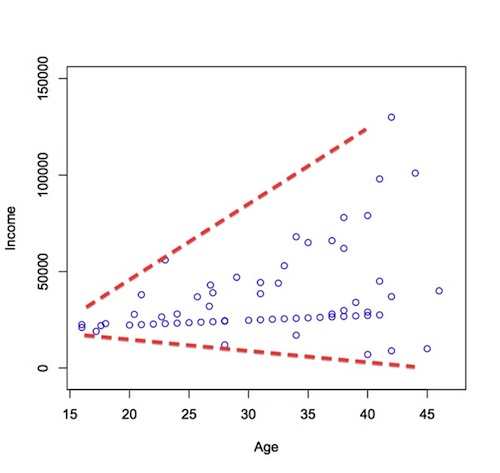
Chart, scatter chart

Description automatically generated

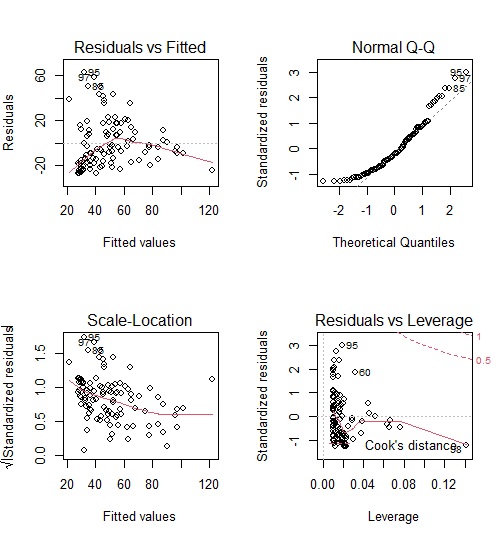
1. Search the internet to find a *real-world* dataset that shows *heteroscedasticity*. Show the corresponding scatter diagram to visualize the varying variance for the residuals. Your scatter diagram showing the relationship between the dependent and explanatory variable should look like the following (in case of an increasing trend for variance of residuals). Note how observations of the variable get wider and wider around the regression line for larger values of .

A computer screen capture

Description automatically generated with low confidence

Example: 

*Image Reference:* [*http://www.statsmakemecry.com/smmctheblog/confusing-stats-terms-explained-heteroscedasticity-heteroske.html*](http://www.statsmakemecry.com/smmctheblog/confusing-stats-terms-explained-heteroscedasticity-heteroske.html)



*Demo Dataset Reference:* [*https://help.xlstat.com/s/article/breusch-pagan--white-heteroscedasticity-tests-in-excel?language=en\_US*](https://help.xlstat.com/s/article/breusch-pagan--white-heteroscedasticity-tests-in-excel?language=en_US)

*Method: R, Graphic test.*

1. Why do you think the dataset that you found shows a heteroscedasticity behavior? You need to justify this behavior based on your understanding of the data.

We can see the Heteroscedasticity through the Residuals vs Leverage plot. (Lowe right)

We also use The Breusch-Pagan Test to double check the residuals.

1. Search the internet to find out what issues’ heteroscedasticity causes in a regression analysis. *Make sure you use your own words in your response.*

The direct cause for heteroscedasticity is that the test results cannot be trust, same as the violation of other assumptions. Therefore, heteroscedasticity is a major concern in regression analysis and ANOVA test.

*Reference:* [*https://statisticsbyjim.com/regression/heteroscedasticity-regression/*](https://statisticsbyjim.com/regression/heteroscedasticity-regression/)

[*https://en.wikipedia.org/wiki/Heteroscedasticity*](https://en.wikipedia.org/wiki/Heteroscedasticity)

1. Propose at least three ways to deal with heteroscedasticity and resolve it. *Make sure you use your own words in your response.*

a. Perform a sensitivity analysis, use a different specification for the model (different X variables or non-linear one).

b. We can transform the data into logs that has the effect of reducing the effect of large errors relative to small ones and reducing the extreme value to make the fitted line on the right place.

c. Using a weighted least squares estimation method, so that OLS can observed the difference in transformation of weighted values of X and Y. The weights are changed by the observation. There is a positive relationship between dependent variable and OLS regression.