Lesson X Practical Exercise

STUDENT NAME

School of Public Affairs: Pennsylvania State University

PADM 504, Section 001: Data Analysis for Policy and Administration

Professor Lauren Azevedo

DATE

**Purpose**

Orient the reader to the purpose of your analysis. Provide the research question and how you are answering it.

**Method**

To evaluate heteroscedasticity & outliers of the regression model the following methods are used:

**Heteroscedasticity**

If a residuals vs fitted values plot of a regression model shows any pattern, then there exists heteroscedasticity in the regression model. Breusch-Pagan test is also use to detect the presence of heteroscedasticity.

**Outliers**

If a boxplot shows values outside the lower hings & upper hings, then those values are treated as outliers. Outliers can make model inaccurate. Outliers should be excluded from the data before modeling.

**Results**

Table-01: Estimate of Multiple Regression Model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t statistic | p-value | R square | Adjusted R square |
| Intercept | 52.5009 | 1.5401 | 34.089 | 0.0000 \*\*\* | 0.6159 | 0.6056 |
| Urban | 0.1874 | 0.0336 | 5.572 | 0.0000 \*\*\* |  |  |
| GDP | 0.0004 | 0.0002 | 2.883 | 0.0047 \*\* |  |  |
| Hospbed | 0.0880 | 0.0269 | 3.267 | 0.0014 \*\* |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

F-statistic: 59.86 on 3 and 112 DF, p-value: < 0.0000

Since the p-value of the F statistic is less than 0.05, the overall model is okay.

The intercept term is 52.5009. If the contribution of all the regressors is zero, then the female life expectancy will be 52.5009 years, on average. The coefficient of urban is 0.1874 which implies that for every one unit increase in percentage of urbanization, the female life expectancy will be increased by 0.1874 years, on average. The coefficient of GDP is 0.0004 which implies that for every one unit increase in GDP, the female life expectancy will be increased by 0.0004 years, on average. The coefficient of Hospbed is 0.0880 which implies that for every one unit increase in hospital bed per 1000 people, the female life expectancy will be increased by 0.0880 years, on average. All of the regression coefficients in the above regression model is statistically significant as the p-values are less than 0.05. The value of R square is 0.6159 implies that the 61.59% variation in female life expectancy can be explained by urban, GDP & hospbed.

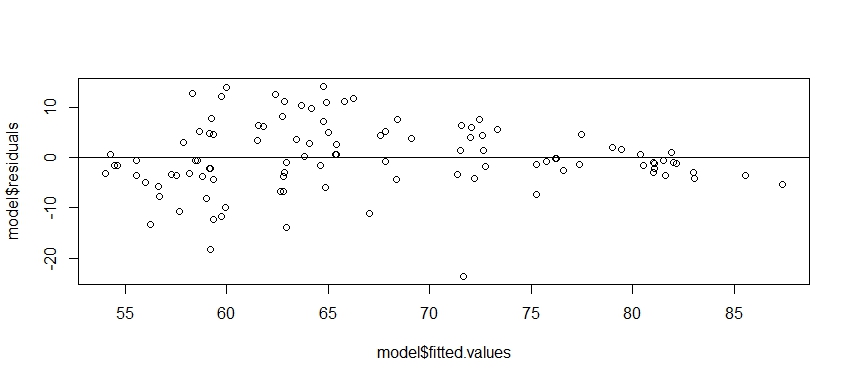


Figure-01: Residual vs Fitted value plot

From the residual vs fitted value plot, we observe a funnel pattern which implies that there exists heteroscedasticity in the above regression model. Now, we need to test the following hypothesis,

: There is no heteroscedasticity vs : There is some heteroscedasticity

Table-02: Breusch-Pagan test for detecting heteroscedasticity

|  |  |  |
| --- | --- | --- |
| BP statistic | df | p-value |
| 8.7161 | 3 | 0.03331 |

Since the p-value of the corresponding BP statistic is less than 0.05, we can reject null hypothesis at 5% level of significance. That means, there exists heteroscedasticity. In the presence of heterosedasticity, weight least square estimate provides better result.

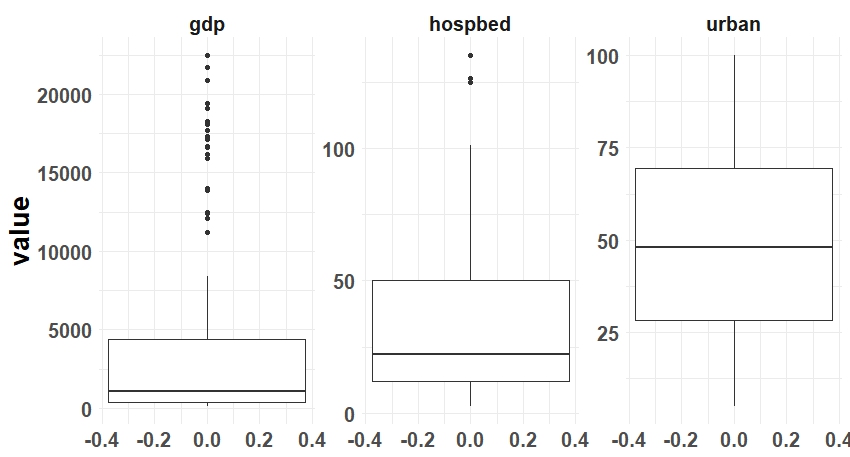


Figure-02: Boxplot of the regressors for detecting outliers

From the figure-02, it can be observed that GDP & hospbed has some outlier values. This outlier can make the result inaccurate. For the purpose of better result, outlier values are excluded then a corrected regression model is fitted.

Table-03: Weighted Least Square Estimate of Corrected Multiple Regression Model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Coefficients | Estimate | Std. Error | t statistic | p-value | R square | Adjusted R square |
| Intercept | 54.1669 | 1.7499 | 30.955 | 0.0000 \*\*\* | 0.6882 | 0.6790 |
| Urban | 0.1727 | 0.0305 | 5.661 | 0.0000 \*\*\* |  |  |
| GDP | 0.0004 | 0.0001 | 4.049 | 0.0001 \*\*\* |  |  |
| Hospbed | 0.0677 | 0.0199 | 3.405 | 0.0009 \*\*\* |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

F-statistic: 74.31 on 3 and 101 DF, p-value: < 0.0000

Since the p-value of the F statistic is less than 0.05, the overall model is okay.

The intercept term is 54.1669. If the contribution of all the regressors is zero, then the female life expectancy will be 54.1669 years, on average. The coefficient of urban is 0.1727 which implies that for every one unit increase in percentage of urbanization, the female life expectancy will be increased by 0.1727 years, on average. The coefficient of GDP is 0.0004 which implies that for every one unit increase in GDP, the female life expectancy will be increased by 0.0004 years, on average. The coefficient of Hospbed is 0.0677 which implies that for every one unit increase in hospital bed per 1000 people, the female life expectancy will be increased by 0.0677 years, on average. All of the regression coefficients in the above regression model is statistically significant at 5% level of significance as the p-values are less than 0.05. The value of R square is 0.6882 implies that the 68.82% variation in female life expectancy can be explained by urban, GDP & hospbed.

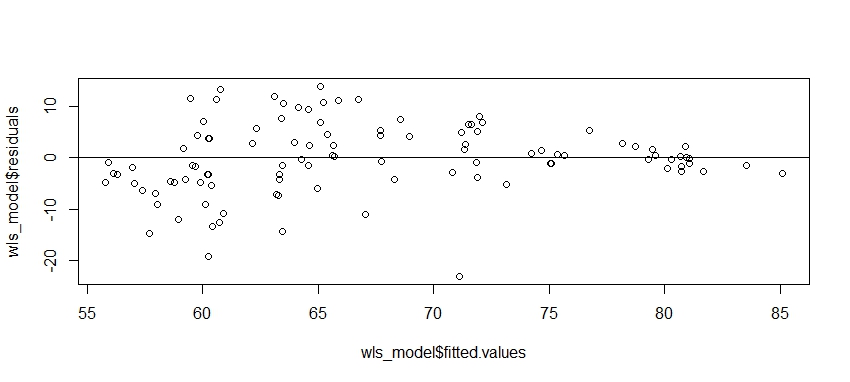


Figure-03: Residual vs Fitted value plot of the WLS model

From the residual vs fitted value plot, we observe a funnel pattern which implies that there exists heteroscedasticity in the above regression model. Now, we need to test the following hypothesis,

: There is no heteroscedasticity vs : There is some heteroscedasticity

Table-02: Breusch-Pagan test for detecting heteroscedasticity

|  |  |  |
| --- | --- | --- |
| BP statistic | df | p-value |
| 1.05 | 3 | 0.7892 |

Since the p-value of the corresponding BP statistic is 0.7892 which is greater than 0.05, we can not reject null hypothesis at 5% level of significance. That means, the heteroscedasticity is removed from the regression model.

**Discussion and Policy Implications/Recommendations**

Most of the practical exercises ask you to offer policy implications or recommendations based on your findings. First discuss the overall meaning of your analysis and then present your policy implications and recommendations to the reader.

**References** (if used)