Paula Joy B. Martinez
IV – BS Economics
IE 198 Case 4

3.1. Import the Data in R and Fit a Linear Regression Model on the Full Dataset.

3.1.1. What is the p-Value of the Model?: 7.183e-16

3.1.2. What are the coefficients of the following variables:

3.1.2.1. **NetworkCBS**: **-5.402e+04**
3.1.2.2. **DayTH**: **5.620e+04**
3.1.2.3. **D1849Rating**: **-4.205e+04**
3.1.2.4. **Twitter**: **3.443e-02**
3.1.2.5. **TypeC**: **0**

3.2. Outlier Analysis and Model Validation.

3.2.1. Identify 4 outliers in the dataset. Which shows are outliers?

- Americal Idol
- Gossip Girl
- Bones
- NCIS

3.2.2. Does the full model satisfy the non-constant variance assumption of regression? Draw the residuals versus plots graph here.

- No, the full model doesn't satisfy the non-constant variance assumption of regression. From Figure 1 below, the residuals seem to have a double-bow pattern.

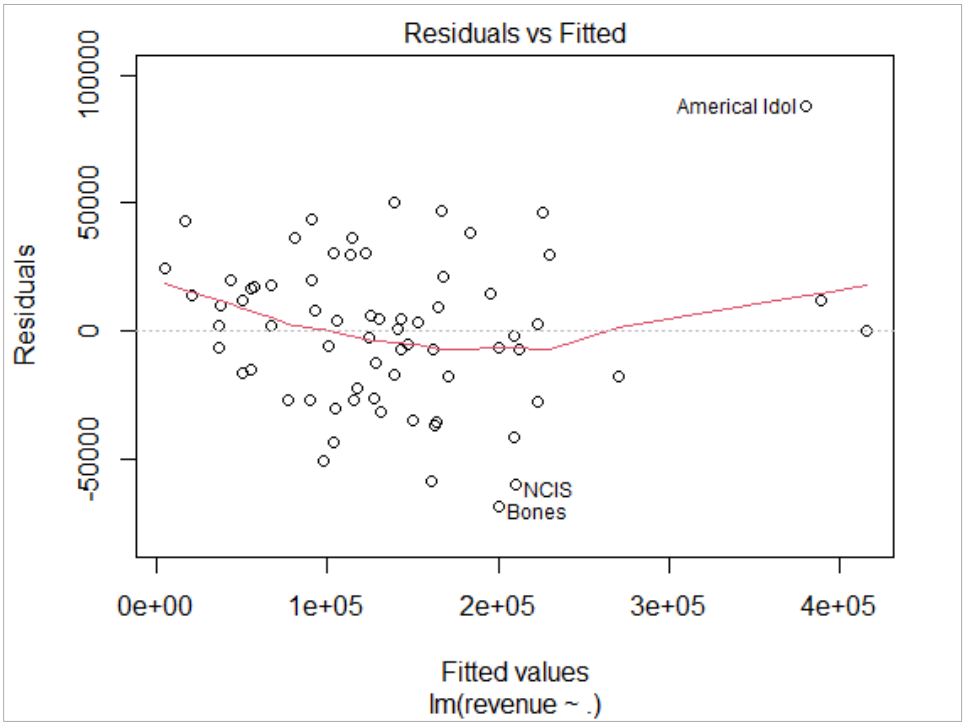

Figure 1 Residuals vs Fitted Plot

Paula Joy B. Martinez
IV – BS Economics
IE 198 Case 4

3.3. Reduce the original dataset by removing the identified outliers. Do Stepwise Regression on the reduced dataset.

3.3.1. Which of the variables are significant in the final iteration?

- network, day, viewers, d1849rating, facebooklikes, facebooktalkingabout, twitter, type

3.3.2. Write down the final reduced regression model with only the significant variables present.

$$revenue = \beta_0 + \beta_1\, network + \beta_2\, day + \beta_3\, viewers + \beta_4\, d1849rating + \beta_5\, facebooklikes + \beta_6\ facebooktalkingabout + \beta_7\ twitter + \beta_8\ type + \epsilon$$

3.4. Standardize the reduced dataset with significant variables.

3.4.1. Which variable is the most influential?

- viewers

3.4.2. Which Day has the biggest cost in advertising?

- Friday

3.4.3. Which Network has the biggest cost in advertising?

- CBS

3.4.4. Which is more influential? Twitter or Facebook? Why?

- Comparing the P value of the variable *twitter* (0.011880) with the variables *facebooklikes* (0.023940) and *facebooktalkingabout* (0.169893), we can say that Twitter is more influential than Facebook because its P value is less than the P value of the Facebook variables. This implies that when *twitter* is the last variable to be added in the model, it will have more value-adding information than in the similar situation with *facebooklikes* and *facebooktalkingabout*.