**Problem 1 – Corruption Perception Index**

The dataset *country.csv contains* Corruption Perception Index and Gini Index of 20 countries. Corruption Perception Index close to 100 indicates low corruption and close to 0 indicates high corruption. Gini Index is a measure of income distribution among citizens of a country (high Gini indicates high inequality). Corruption Index is taken from Transparency International, while Gini Index is sourced from Wikipedia.

1. Develop a simple linear regression model (Y = b0 + b1X) between corruption perception index (Y) and Gini index (X).
2. What is the change in the corruption perception index for every one unit increase in Gini index?
3. What proportion of the variation in corruption perception index is explained by Gini index?
4. Is there a statistically significant relationship between corruption perception index and Gini index at alpha value 0.1?
5. Highlight the 95% confidence interval for the regression coefficient b1.

**Problem 2 – Bollywood Movies**

The box-office collection of a Bollywood movie across different regions and the corresponding social media engagement (like + dislikes) is provided in attached file.

1. Develop an SLR Model for the data shown in the bollywood\_collections excel file.
2. Is there any evidence that the box-office collection (Y) of the movie has a statistically significant relationship with its social media engagement.
3. Should Bollywood movie producers invest more to promote their movies through social media?

**Problem 3: Diamonds**

A regression model (Model 1) based on data of 6,000 diamonds is developed using price as the dependent variable and carat as the independent variable

Model 1: Y = B0 + B1 X Carat

The model parameters are as follows

|  |  |  |  |
| --- | --- | --- | --- |
|  | B | t-value | p-value |
| Constant | -12,738.581 | -63.439 | .008 |
| Carat | 18,381.261 | -8.489 | .003 |

Question

1. Comment on the validity of the model
2. For every one-carat increase in the diamond weight, the price of the diamond increases by what amount.

Model 2: ln(Y) = B0 + B1 x Carat

|  |  |  |  |
| --- | --- | --- | --- |
|  | B | t-value | p-value |
| Constant | 7.265 | 682.302 | .000 |
| Carat | 1.375 | 183.004 | 0.000 |

Question

1. Calculate the maximum possible price of a specific diamond that weighs 0.4 grams.

**Problem 4: Hospital**

Use the data provided in the Hospital excel file on body weight of patients and their treatment costs to construct a SLR model.

Part 1

1. Comment on the validity of the model
2. What will be the average treatment cost of a patient weighing 50 kgs?
3. Calculate the R-square.
4. Does the model satisfy homoscedasticity assumption?
5. Are the residuals normally distributed?

Part 2

1. Do a transformation of variables and attempt to construct a better SLR model than above.
2. Does the new model satisfy homoscedasticity assumption?
3. Are the residuals normally distributed?

Instruction: FileName: SLR\_assignment\_firstname

Email: [ashok.21aug@gmail.com](mailto:ashok.21aug@gmail.com)