

Statistical and Predictive Modeling I (DATA 1204)
Assignment #5 – Linear Regression (15% of Total Grade)
Professor: Fatma Tetikoglu

The following questions will use the Expenses.csv dataset.

The Ask: Mr. John Hughes wants us to conduct a Simple Linear Regression Analysis using R to understand the effect of sex on expenses. He wants a proper Simple Linear Regression Analysis with a Hypothesis Testing with the assumption that the predictor (or independent) variable for our linear regression will be “sex” and the dependent variable (the one we’re trying to predict) will be “expenses”.

Questions: Create a PowerPoint (PPT) presentation that includes the following:

1. Cover Page (1%)
2. Description of the analysis: Please describe the context of Simple Linear Regression and the specific analysis you will be conducting. (2%)
3. Conduct the analysis in the ask. What insights have you gained from the Simple Linear Regression results? Please explain your answer, elaborate on your steps for the hypothesis testing and summarize your findings. (10%)
4. What other two (2) variables would you include to help increase the accuracy of your model? Please explain the reasons why you would choose them and how this analysis would help us. An analysis on R for these 2 variables is NOT needed. (2%)

Hint: Leverage the Week #12 – Exercise

Please post your PowerPoint (.ppt) including R Code via assignments under Assignment#5 by 11:59 pm on Friday, August 2, 2024.

Rubric

| Question | Exemplary (14-15) | Proficient (10-13) | Incomplete (7-9) | Needs Improvement (0-6) |
|----------|--|--|--|--|
| 1 | Cover page presents the student's name, course name, date, etc. | Cover page is incomplete. | Cover page is incorrect. | Cover page is absent. |
| 2 | Description of the analysis is explained properly. The context of Simple Linear Regression Analysis is summarized correctly. | Description of the analysis is incomplete. There is no explanation about Simple Linear Regression Analysis. | Description of the analysis is incorrect. | Description of the analysis is absent. |
| 3 | A summary for the Simple Linear Regression analysis is given properly. Step-by-step outline on how to prove (or dis-prove) the hypothesis is properly explained in 5 steps. Insight from the analysis conducted is described properly and explained clearly. Findings and conclusion are given correctly. Input and output from R are presented and explained. | A summary for the Simple Linear Regression analysis is incomplete. Steps of the hypothesis testing are incomplete or missing some information and explanation. Insight from the analysis conducted is described incomplete and not explained clearly. Findings and conclusion are not given properly and correctly. Input and output from R are presented but not explained. | Insight from the analysis conducted is described incorrectly and in a generic form. Findings and conclusion are not correct. Input and output from R are absent. | A summary for the Simple Linear Regression analysis is absent. Findings, conclusion, and insight are absent. |
| 4 | Two variables for further analysis and improve accuracy are given correctly, properly and explained in detail. | Two variables for further analysis and improve accuracy are given without any explanation. | Two variables for further analysis and improve accuracy are given incorrectly. | Two variables for further analysis and improve accuracy are not presented. |