Homework 4

IEOR: 4574

Due date: March 07

Syed Haider, Ph.D.

Feb 21, 2023

- **Question 1:** Run the OLS related code shared in the "GLM.py" file for the 'us_macro_quarterly.xlsx' data: [**Points 30**]
  - Explain the model summary when all the data is fit. Discuss R-squared, Df Residuals, Df Model, Method, F-statistic, Prob (F-statistic), Log-Likelihood, AIC, BIC, Coef, Std err, t, and $P > |t|$.
  - Run with the train-test split based on the ordering and compare the results of the test with the above model fitted on the entire data.
  - Run all three variations of ANOVA and discuss results on SSR, df_diff, ss_dif, df_resid, F and $Pr(> F)$.
  - Submit appropriately labeled visuals.
- **Question 2:** Run the code shared in the "GLM.py" file for the "chip_dataset.csv" data: [**Points 50**]
  - Explain the model summary when the entire data is fitted. Discuss R-squared, Df Residuals, Df Model, Method, F-statistic, Prob (F-statistic), Log-Likelihood, AIC, BIC, Coef, Std err, t, and $P > |t|$.
  - Run with the train-test split when the order is not important and compare the results of the test with the above model fitted on the entire data.
  - Run all three variations of ANOVA and discuss results on SSR, df_diff, ss_dif, df_resid, F and $Pr(> F)$.
  - Submit appropriately labeled visuals.
- **Question 3:** Run the code: GLM using Poisson Regression on the 'Smokers_Age.xlsx' data. And discuss the summary results [**Points 20**].