Project Details

ECO 204 - Statistics for Business and Economics - II, Summer 2025
Faculty: Shaikh Tanvir Hossain, **TA:** Habiba Afroz
Project Submission Due Date: 3rd September, 2025

§

Pick a dataset from the list provided on the course webpage (examples available in the project folder include: `approval`, `athlet2`, `earns`, `fertil3`, `intdef`, `inven`, `okun`, `prminwge`, `rdchem`, `traffic1`). For your chosen dataset, complete the tasks below and submit an **Excel or R Markdown** file (with HTML output) that reproduces every result. (**Note:** If you do the complete task in R and LaTeXand submit you will get 3 additional points in the final exam.)

1. (a) For your chosen dataset, **pick a dependent variable and some (at least three) independent variables**. In this case you have to come up with a story around the data and the variables you choose. Qualitatively explain why the chosen independent variables may affect the dependent variable.

   (b) **Perform the regression** using Excel or R. Write a formatted regression table in the report.

   (c) **Interpret** the estimated coefficients, goodness of fit measures (e.g., R-squared, adjusted R-squared), and any relevant model diagnostics.

   (d) Do **model diagnostic checking**. At minimum include: residuals vs fitted plot, residuals vs each regressor, a QQ-plot of residuals. Briefly describe what each plot reveals.

   (e) Conduct **individual significance testing (t-tests)** for the coefficients and report the results.

   (f) Conduct a **joint significance test (F-test)** for all variables. Report the F-statistic, degrees of freedom, and p-value, and interpret the result.

   (g) Perform at least **one restricted vs unrestricted test** other than the joint test. State the restricted model, estimate both models, compute the test statistic (F or LR), and conclude which model is preferred at conventional significance levels.

   (h) Add **interaction variable(s)** (at least one interaction between two regressors) and re-run the regression. Interpret the interaction term(s) and report whether they change the main conclusions.

   (i) You can think some **Non-linear transformations** of the regressors (e.g., polynomial terms, logarithmic transformations) that might improve model fit. Implement these transformations, re-run the regression, and compare the results to the original model.

   (j) Do a **logistic regression** where you have to pick a dependent variable that is binary in nature (e.g., success/failure, yes/no) and at least one independent variable. Report the estimated coefficients, odds ratios, and any relevant model diagnostics.