GB704 - Homework 1
Due : Monday, September 17th

Problem 1 (25pts)

A company claims that a plastic injection press is properly calibrated, which should mean that the rejection rate is 1% or less. You plan to use a preliminary run to test that claim.

1. Your manager wants you to select the run size so that the standard error of the mean rejection rate is 0.2% – under the hypothesis that the default rate is 1%. How large must the run be?
2. You run a test of the corresponding size and find that the proportion of rejects is 1.35%. Based on a one-sided test can you reject the hypothesis that the machine's rejection rate is 1% or less with 95% confidence?

Problem 2 (25 pts)

Find a publicly traded stock for which at least 5 years of historical data exist. Throughout this problem, use data at a **monthly frequency**, which will give you 12 observations per year and will make computations easier.

1. Regress your company's monthly return on the S&P500's return
2. If I tell you that the S&P500 is going to return 5% this coming month and based on the model you just estimated, what would be your forecast for your stock's return over that same period?

Problem 3 (50pts)

Download dataset data1D2D.xlsx from my webpage.

1. Regress ln(spending) on gender. Does gender have a significant effect on spending according to that regression?
2. Now regress ln(spending) on income and gender. What happens to the significance of gender? Explain in a sentence or two what caused this change (if any).

3. Now we want to forecast what a new male customer with income $\$225,000$ and age 40 is going to spend using a regression model with ln(spending) on the left-hand side. Possible explanatory variables are $age, age^2, income, income^2$ and *gender*. Choose the model that you feel is best for that purpose (explain why you chose it) and use that model to forecast the spending of our new customer.

4. (Nearest neighbors forecasting) Perform the same forecasting exercise as in part 3 by averaging the spending of the 10 closest consumers to our target in our existing data. Use the same notion of distance as in class.

5. (Cross-validation) Estimate your preferred model from part 3 after leaving the first 500 observations out. Use the resulting model to forecast the 500 observations you left out. Plot observed outcomes vs predicted outcomes. Do this for logs first, then for dollar spending. Fit a line through those dots and interpret the resulting R^2 for each of the two charts in **one** sentence.