APPLIED BUSINESS ANALYTICS



2024-25, Spring Semester

Exercíses to Practice - Week 3

- 1. (Conceptual Question) Suppose that you collect a random sample of 250 salaries for the salespersons employed by a large PC manufacturer. Furthermore, assume that you find that two of these salaries are considerably higher than the others in the sample. Before analysing this data set, should you delete the unusual observations? Explain why or why not.
- 2. (Conceptual Question) Suppose you have a crosstabs of two "Yes/No" categorical variables, with the counts shown as percentages of row totals. What will these percentages look like if there is absolutely no relationship between the variables?
- 3. (Conceptual Question) Suppose there are two commodities A and B with strongly negatively correlated daily returns, such as a stock and gold. Is it possible to find another commodity with daily returns that are strongly negatively correlated with both A and B?
- 4. (Conceptual Question) Explain the relationship between a categorical variable and the series of binary dummy variables derived from it. How many dummy binary variables are required to capture the information in a categorical variable with N categories?
- 5. The file LaptopSalesJanuary2008.csv contains data for all sales of laptops at a computer chain in London in January 2008. This is a subset of the full dataset that includes data for the entire year. The data dictionary is in the end of this document.
 - a. Create a **bar chart**, showing the average retail price by store. Which store has the highest average? Which has the lowest?
 - b. To better compare retail prices across stores, create side-by-side boxplots of retail price by store. Now compare the prices in the two stores from (a). Does there seem to be a difference between their price distributions?
- 6. The file **ToyotaCorolla.csv** contains data on used cars (Toyota Corollas) on sale during late summer of 2004 in the Netherlands. It has 1436 records containing details on 38 attributes, including Price, Age, Kilometers, HP, and other specifications. The goal will be to predict the price of a used Toyota Corolla based on its specifications.
 - a. Identify the categorical variables.
 - b. Use Python to convert the categorical variables in this dataset into dummy variables, and explain in words, for one record, the values in the derived binary dummies.
 - c. Use Python to produce a correlation matrix and matrix plot. Comment on the relationships among variables.