FAIRLEIGH DICKINSON UNIVERSITY

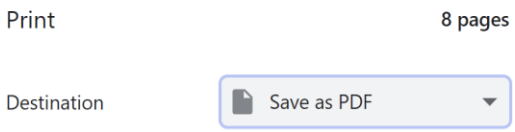
DSCI 6780 Analyze Homework 4 Analyze Part II

Due: 6:00pm Friday Nov 17, 2023. Please submit on WebCampus before the weekend starts.

Submission Instructions:

All three documents should be submitted to WebCampus.

- Main Document: This word document including all your detailed answers for each question. **Excel and Python outputs, graphs, and plots should be included in your answers.** Readers of your main file should **NOT** need to look into your supporting Python or Excel file to find answers.
- Supporting Document 1: One **SINGLE** Excel file consists of **4 tabs for Q3-Q6.**
- Supporting Document 2: One Single Python **HTML file printed as PDF format** (After you download the Jupyter Notebook file as html, open the html, print, choose "Save as PDF").



Question 1 Supplier Selection (10%):

Data: *HW4Q1.xlsx*

Task: Using Python.

You need to select from three suppliers who supply surfactant to your company. The requirement is that the mean readings of surfactant should be **> 0.2**. Do a proper statistical analysis to make a decision at 1% significance level. Please show your work.

Question 2 Supplier Selection (10%):

Data: *HW4Q2.xlsx*

Task: Using Python.

Three suppliers supply surfactant to your company. You want to select the supplier that has the largest mean readings of surfactant. Do a proper statistical analysis to make a decision at 1% significance level. Please show your work.

Question 3 Simple Regression (10%):

Data: *HW4Q3.xlsx*

Task 1: Using Excel.

1. Produce a **scatter** plot, showing the relationship between the inventory level and the customer service level. Please add Trendline, Equation, and R square to the scatter plot.
2. Investigate and describe the level of **correlation** between the inventory level and the customer service level. What can you say about the result?
3. Explore the relationship between inventory level (y) and customer service levels (X) using **regression** analysis.
   Report and briefly discuss the model output.

Task 2: Using Python.

1. Build a regression model to answer the question: how do customer service levels (X) predict inventory level (y)?
2. Produce a **scatter** plot, showing the relationship between the inventory level and the customer service level in Python. Add a trendline to your plot.
3. Did you see any difference in Python output compared to Excel regression results?

Question 4 Simple Regression (20%):

Data: *HW4Q4.xlsx*

Task 1: Using Excel.

1. Investigate the relationship between the number of employees and the number of customer complaints per month through **correlation** analysis.
   What can you say about the result?
2. Investigate the relationship between the number of employees and the number of customer complaints per month through **regression** analysis. Using common business sense, what is your y variable and what is your X variable?
   What can you say about the result?

Task 2: Using Python.

1. Build a regression model to investigate the relationship between the number of employees and the number of customer complaints per month.
2. Did you see any difference in Python output compared to Excel regression results?

Question 5 Multiple Regression (25%):

Data: *HW4Q5.xlsx*

Task 1: Using Excel.

1. These data present the costs of moving varying weights of goods over different distances. Investigate the relationship between costs and the other variables provided. Derive a regression model to predict freight costs.
2. Please briefly explain 1) R square; 2) F test; 3) write down the equation you can use to predict freight cost.
3. What is the predicted freight cost to move a shipment of 2000 lbs over 1500 miles?

Task 2: Using Python.

1. Build a regression model to investigate the cost incurred of moving varying weights of goods over different distances.
2. Use Python to produce a 3D plot showing the relationship between the different variables. Make necessary changes to X, Y labels in your code. Attach your plot in below.
3. Did you see any difference in Python output compared to Excel regression results?

Question 6 Multiple Regression (25%):

Data: *HW4Q6.xlsx*

Task 1: Using Excel.

1. These data show the fuel consumed according to different variables. Undertake an analysis of the relationship of these variables with fuel consumption and derive an appropriate regression model to predict fuel consumption.
2. Please briefly explain 1) R square; 2) F test; 3) write down the equation you can use to predict fuel consumption.

Task 2: Using Python.

1. Build a regression model to investigate the relationship between fuel cost and other variables.
2. Did you see any difference in Python output compared to Excel regression results?