SCMA Assignments

Overall instructions:

1. *Download required packages onto your local machine:*
    a. *Download R studio https://www.rstudio.com/products/rstudio/download/#download*
    b. *Download Python Anaconda packages (refer to https://docs.anaconda.com/anaconda/user-guide/tasks/install-packages/ for instructions)*
2. *Listed below are the SCMA632 coding assignments. There will be separate assignments for R and Python. These are weekly assignments, and their deadlines invariably falls on Friday midnight. The SCMA team will start grading the assignments on Saturday. The team can call some or all of you and ask you to run the codes and explain the results on a video call.*
3. *Coding Assignments will be of two types - R coding assignments and Python coding assignments. Each assignment will have a starter code and a reference to the data set. You must modify starter codes to suit your data sets. This may involve pointing to a different dataset or indexing data as per your data set or any other data set specific actions needed to get the desired results.*
4. *After completing the assignments, you should Submit codes as well as your results and its interpretations. Submission must be made on GitHub Classroom.*
    a. *Use link https://classroom.github.com/a/p7gImS4t to access your GitHub R assignment repository. Once you access, copy paste your codes in the file "myRcodes." Copy paste your results to the file "myRresults"*
    b. *Use link https://classroom.github.com/a/XMn45fVn to access your GitHub Python assignment repository. Once you access, copy paste your codes in the file "myPythoncodes"*
    c. *Copy-paste codes to the file "myPythonresults"*
5. *Those who want to learn more about GitHub can watch tutorial https://www.youtube.com/watch?v=PQsJR8ci3J0*

A1a: "Data Cleaning" (Total points 60; Grading weight 8%; Due on Fri, July 2nd, 11.59 PM)

(1) Consider the NSSO data. This is raw data. We want to clean the data so that we can perform various statical analysis using it

(2) Get data for the state/UT assigned to you

*(2) Use recommended practices to handle missing values. The practices may include getting rid of rows or columns with missing data or filling in missing values with mean, median, or mode**. (10 points)***

(3) Find the administrative divisions listed for your state/UT data set in the URL; Assign numerical codes 1,2,3…to each of the administrative divisions in the state/UT assigned to you. This will be your administrative division index. Add a new column "Index" to your database and fill Index values for each row (10 Points)https://en.wikipedia.org/wiki/List_of_divisions_in_India).

*(4) Modify starter codes to compute administrative-division-wise descriptive statistics as well as descriptive statistics for your assigned state/UT. You must do it for each variable in your data set**(10 Points)***

*(5) Modify starter codes to find and treat outlier data points using a 3-standard deviation rule for each division and the entire state/UT **(10 Points)***

*(6) Modify starter code to Visualize your division wise as well as state/UT data; **(10 Points)***

*(7) Write interpretations of your results: Use tables and graphs wherever necessary **(10 points).***

(8) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)

A1b & A1c: ***Distribution theory (Total Points: 30; Grading weight 2%; due on Fri, July 2nd , 11.59 PM)***

(1) Consider the data set "IPL Ball-by-Ball 2008-2020". It has the ball-by-ball data on the runs scored by various cricket batsmen or wickets taken by the bowlers. We want to know which distribution fits for the runs and wickets data

(2) The ball by ball may follow different distributions. Distributions are of two types viz., discrete, and continuous. Distributions like Poison, Binomial, Negative Binomial, Hypergeometric distribution are discrete distributions whereas distributions like Normal, beta, gamma, etc are continuous distributions.

*(3) Find the most appropriate distribution for the runs scored by the batsman assigned to you **(10 Points)***

*(4) Find out the most appropriate distribution for the wickets taken by the bowler assigned to you **(10 Points**)*

*(5)) Write interpretations of your results: Use tables and graphs wherever necessary **(10 Points**).*

(6) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)

Assignment A2: "Estimating and Testing" (Total Points: 60; Grading weight 10%; due on Fri, July 9th, 11.59 PM)
(1) Consider the NSSO data for the state/UT assigned to you. We want to find out the levels of consumptions in each division and the differences in the levels of consumptions across divisions.
(2) State hypotheses to estimate distribution parameters for consumption variable for each division ***(10 points)***
(3) Use an appropriate distribution to estimate distribution parameters for each consumption variables ***(10 Points)***
(4) Construct confidence intervals to convey the reliability of your estimates ***(10 points)***
(4) State hypothesis to test whether there exist differences between the divisions on the consumption pattern ***(10 Points)***
(5) Use statistical tests to say if the differences between the divisions are statistically significant or not ***(10 points)***
(6) Write interpretations of your results: Use tables and graphs wherever necessary ***(10 Points****).*
(7) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)

Assignment A3: "Regression Models" (Total Points 50; Grading Weight 10%; due on Fri, July 16th, 11.59 PM)
(1) Consider the NSSO data for the state/UT assigned to you. We want to identify factors influencing the consumption.
(2) Perform Regression analysis to identify factors influencing the consumption using demographic variables at state as well as at district level ***(20 points)***
(3) Carry out diagnostic checking of the regression like outlier detection in Y and X's ***(10 Points)***
(4) Compute the price and income elasticity of demand of cereals, Milk, and Pulses ***(10 points)***
(5) Write interpretations of your results: Use tables and graphs wherever necessary ***(10 Points).***
(6) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)

Assignment A4: "Panel Data": (Total Points 40; Grading Weight 10%; Fri, July 23rd, 11.59 PM)
(1) Consider 'Bank 2019' and 'Bank 2020' data. We want to find out if the mergers have improved the bank efficiency or not?
(2) Fit a regression model using panel data ***(20 Points)***
(3) Draw inferences on the impact of merger and Write interpretations of your results: Use tables and graphs wherever necessary ***(20 Points)***
(4) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)

Assignment A5: "Logistic Regression" (Total Points 50; Grading Weight 10%; Due on Fri, July 30th, 11.59 PM)
(1) Consider the NSSO data for the state/UT assigned to you. We want to identify factors influencing consumption using a binary independent variable viz., non-vegetarians and vegetarians.
*(2) Use Logistic regression to find factors that influence consumption patterns among non-vegetarians and vegetarians****. (30 points)***
(3) Use Probit/Tobit regression and inverse Mills ratio to handle zero values in the data ***(10 Points)***
(4) Write interpretations of your results: Use tables and graphs wherever necessary ***(10 Points)***
(5) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)

Assignment A6: "Conjoint Analysis" (Toal Points 100; Grading Weight 10%; Due on Fri, Aug 6th, 11.59 PM)
(1) Consider the product or a service category assigned to you.
(2) Identify critical attributes and their levels for the product/service assigned to you ***(20 points)***
(3) Develop concept cards to measure critical attribute levels ***(20 Points)***
(4) Create a google form for the cards, circulate among your classmates, and get primary data. You must have at least 60 respondents ***(20 Points)***
(5) Perform the Conjoint Analysis ***(20 points)***
(6) Write interpretations of your results: Use tables and graphs wherever necessary ***(20 points)***
(7) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)

Assignment A7a: "Time Series Analysis I": (Total Points:100; Weight 5%; Due on Fri, Aug 13th, 11.59 PM)
(1) Go to the website Investing.com, download the time series data for the commodity or currency assigned to you
*(2) Check for outliers, identify and treat them if necessary (****10 Points****)*
*(3) Convert the data to weekly and monthly frequencies (****10 Points****)*
*(4) Estimate the trend and seasonality in the prices for the monthly data (****10 Points****)*
*(5) Fit an ARIMA process using ARIMA models for daily data (1****0 Points****)*
*(6) Estimate using Artificial neural networks (ANN), Facebook profit (****20 Points – 10 points each****)*

(7) *Forecast the respective model for 50 days and present the forecast of the three models (3**0 Points – 10 points each**)*
(8) *Write interpretations of your results: Use tables and graphs wherever necessary (**10 points**)*
(9) *Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)*

Assignment A7b: "Time Series Analysis II": (Total Points:100; Weight 5%; Due on Fri, Aug 13th, 11.59 PM)

(1) Go to website Investing.com, download the time series data on oil prices, INR/USD exchange rate, and Dow Jones index
*(2) Estimate the inter-relationship between oil, exchange rate, and Dow Jones index using transfer function **(10 Points)***
*(3) Estimate Value at Risk(VaR) using the ARIMA, ANN, and Facebook Profit (**30 Points**)*
*(3) Forecast the risk using the ARCH/GARCH model (**20 Points**)*
*(4) Using the same data, fit vector autoregressive model and vector error correction model. Study Granger causality (**20 Points**)*
*(5) estimate the impulse response function and interpret the results (**10 Points**)*
(6) Write interpretations of your results: Use tables and graphs wherever necessary (10 points)
(7) Copy-paste your codes and interpretations on to your respective files in your GitHub repository (Repo)