2104529 Computational Methods in Industrial Engineering

Workshop 6: Time Series and Regression

Question 1

Consider '`AirPassengers`' dataset and explore the following data mining procedure for a time series data

(a) Plot the decomposition of the data and separate data into training set (1948-1958) and testing set (1959)

(b) Develop the following smoothing forecasting models and compute forecasting accuracy

(c) Develop DOS/Java application for exponential series forecasting that takes any csv file as an input (**hint:** this require separate R script and basic on Batch file)

Question 2

Consider self-report weight/height and actual values of such measures of 200 participants available at https://tinyurl.com/2104529PublicFolder)

(a) import/check for NA value separate into groups based on availability of data

(b) For participants who report both weight and height, separate data into testing data (15%) and training data (85%)

(c) For training set, check difference between self-report and actual values

(d) For training set, build a simple linear regression model to predict actual weight with actual height

Question 3

Consider '`height.csv`' dataset available at https://tinyurl.com/2104529PublicFolder), construct a logistic regression model into to predict sex of participant and compare with classification method.

Question 4

Consider the '`award`' (available at www.ats.ucla.edu/stat/data/poisson_sim.csv) dataset of difference student groups with different Mathematic score and follow the data mining steps

(a) Import using its URL like and analysis the data

(b) Construct Poisson Regression to predict number of awards

(c) Compare the poisson regression with simple linear model and calculate AIC of each model

Question 5

Consider the '`Professor Salaries`' (available at https://tinyurl.com/2104529PublicFolder) dataset and follow these instruction

(a) Import data from MS Excel file with '`xlsx`'

(b) Observe the pattern of '`salary`' in the dataset and explain the relationship of each column

(c) Visualize histogram of '`salary`' base on main factors

(d) Construct a linear regression model to predict '`salary`'

last updated: October 31, 2022