

## OPRE 6304 Fall 2024 Homework 1 (80 points)

(Homework 1 will be due on Sept 12<sup>th</sup>, Thursday, 11:59pm (CT). Please submit a single PDF write-up including answers to all questions on eLearning, and attach the Excel/R script files to your submission)

### 1. Time-Series Forecasting (Adapted from Problem 2.2 in Snyder & Shen) (20 points)

The demand for a new brand of dog food has been steadily rising at the local PetMart store. The previous 26 weeks' worth of demand (number of bags) are given in the file 'dog\_food.xlsx'

- a. Use the following methods to forecast next week's demand:
  - i. (2 points) Naïve forecast
  - ii. (2 points) Average forecast
  - iii. (3 points) Moving average method with time window  $n = 3$
  - iv. (3 points) Weighted average method with weights (0.5, 0.3, 0.2)
  - v. (5 points) Exponential Smoothing with  $\alpha = 0.4$
- b. (5 points) Report the MSE, MAD and MAPE for the predictions in (a) made for week 4 – 26.

### 2. Linear Trend Model for US Gasoline Expenditure (20 points)

The US gasoline expenditure from 1954 – 2004 has been recorded in 'gas.csv'. You can find the description of each data column in the following table. Using a simple linear regression (assuming  $D = \beta_0 + \beta_1 \text{YEAR} + \epsilon$ ), forecast the gasoline expenditure. Use data from years 1954 – 1993 as training data and years after 1993 as test data.

- a. (5 points) Report your forecast for the gasoline expenditure in 2005.
- b. (5 points) Report the estimated values for  $\beta_0$  and  $\beta_1$ . How do you interpret these two values?
- c. (5 points) Report the R<sup>2</sup> and OSR<sup>2</sup>. How do you interpret their meaning?
- d. (5 points) Report the p-value associated with  $\beta_1$ . How do you interpret this p-value?

Variables

Year = Year, 1954-2004,

GasExp = Total U.S. gasoline expenditure

Pop = U.S. total population in thousands in the previous year

GasP = Price index for gasoline in the previous year

Income = Per capita disposable income in the previous year

Pnc = Price index for new cars in the previous year

Puc = Price index for used cars in the previous year

### 3. Multiple Linear Regression for US Gasoline Expenditure (40 points)

The US gasoline expenditure from 1954 – 2004 has been recorded in ‘gas.csv’. You can find the description of each data column in the table above. Use a multiple linear regression, assuming

$$D = \beta_0 + \beta_1 YEAR + \beta_2 POP + \beta_3 GASP + \beta_4 INCOME + \beta_5 PNC + \beta_6 PUC + \epsilon,$$

forecast the gasoline expenditure. Use data from years 1954 – 1993 as training data and years after 1993 as test data.

- a. (5 points) Report the R2 and OSR2. What can you imply from the R2 and OSR2?
- b. (5 points) Report the estimated coefficient associated with POP. Does the sign of the coefficient make sense to you? What can you imply from it?
- c. (5 points) Report the VIFs for all the variables. What can you imply from the VIFs?
- d. (20 points) Build a better model using feature selection. Show, step by step, which variable you removed and why you chose to remove it.
- e. (5 points) Report your final model and the final R2 and OSR2.