MATH 4044 – Statistics for Data Science

Practical Week 7

Question 1

The data for this practical is stored in a SAS data file called `store.sas7bdat` located in `mydata` library on the SAS OnDemand server. Variables in that file are as follows:

Variable name	Description
Region	Region of the country (North, East, South, West)
Advertising	Advertising (Yes or No)
Gender	Gender of shopper (M or F)
Book_Sales	Amount spent on books
Music_Sales	Amount spent on music
Electronics_Sales	Amount spent ion electronics
Total_Sales	Total sales

(a) Check the necessary assumptions and perform an ANOVA test to determine whether there is statistically significant difference in average music sales by region. Interpret the results.

(b) Suppose we want to test the hypothesis that music sales in the East region are different than in the rest of the country. Obtain relevant SAS output and interpret your results.

(c) Perform appropriate post-hoc tests and interpret the results.

(d) Define and estimate a multiple regression model for music sales with Region as the explanatory variable. Define appropriate dummy variables using *Region* = *East* as the baseline. Interpret the results.

Note: You can write your own code (see topic notes) or use Tasks. If you choose to use the Tasks menu, make sure that PROC GLM is used. In Enterprise Guide go to ***Tasks > ANOVA > Linear Models…*** and in SAS Web Editor to ***Tasks and Snippets > Introductory Statistics > One-Way ANOVA.***

SAS actually has a procedure called ANOVA, however this procedure requires equal sample sizes per group and it is not as useful PROC GLM.