QMB 3200: Advanced Quantitative Methods Dr. Sravani Vadlamani
Fall 2020 Florida Polytechnic University

Final Project Guidelines

Assignments are due on the due date, unless otherwise specified. Late submission will not be accepted. Please ensure the guidelines for submissions are followed to receive full credit.

The final project involves application of regression concepts on real time datasets. You are assigned one (1) of the options listed below as indicated in Table 1.

Table 1: Project Assignments by Option

Name	Option	Name	Option
Stacie Akinyi	2	Timothy Glisson	1
River Anderson	1	Jake Greenberg	2
Alexis Baugher	3	Marie Hasegawa	3
Lauren Bentzel	1	Maverick Hope	2
Jonathan Boada	2	Ariel Kranz	1
Ryan Bogdanowicz	1	David Lausberg	1
Chevaughn Brown	2	Gus Lipkin	3
Marckenrold Cadet	3	Logan Miller	2
Elton Campbell	2	Joshua Morales	2
Michael Cassidy	1	Connor Oberhofer	3
Miguel Cecchini Do Amaral	3	Michael Ortiz	3
Dylan Christensen	1	Rene Perez	2
Jacinto Diego	1	Alex Pierstorff	3
Dawid Dobryniewski	2	Malcolm Pinkston-Jones	1
Donavan Dodson	3	Jacob Rogers	2
Bryce Elfers	1	Joyelle Saun	2
Nicole Ely	3	Hailey Skoglund	3
Brandon Ervin	3	Samuel Storrs	1
Nathaniel Fuller	3	Kristian Trevino	2
Calvin Genzlinger	2	Edward Von Leue	3
Bryce Gerhart	3	Cole Young	1
Graham Gilbert	3	Zoe Zumbro	2

The goal is to estimate linear regression models using the best predictors (can also include interaction terms). Your work will comprise of (1) calculating descriptive statistics (2) estimating linear regression models and choosing the best model (3) report writing. There are weekly deliverables to ensure you are on track and do not have last minute surprises. The deliverables for the various tasks and their due dates are given in Table 2

Table 2: Deliverables and Due Dates

Assignment	Due Date
Summary Statistics	November 15
Models	November 21
Final Report	December 4

Option 1

You are provided with data on medical costs (Insurance.xlsx) billed by insurance companies and you are tasked to estimate at least 4 statistical models to predict the medical costs based on age, gender, BMI, children, smoking status and region. You may have to recode some categorical variables prior to estimating your models.

Option 2

You are provided with data extracted from the 1998 national household travel survey conducted in The Netherlands (Data_Option2.xlsx). This sample provides the number of weekly trips undertaken by 1631 households (the corresponding variable in the dataset is called ntrips). The variables included in the dataset are shown in the following table. Please estimate at least 4 appropriate statistical models to predict the number of weekly trips and determine the best fit model.

Variable name	Description
hhsize	Number of persons living in the household
nchlt12	Number of children < 12 years of age in the household
nchgt12	Number of children >=12 years of age in the household
nworker	Number of workers in the household
nstudent	Number of students in the household
ncar	Number of cars owned by the household
income	Household income
resloc	Residential location (1=household resides in city, 2=household resides in suburb, and 3=household resides in rural area)

Option 3

This option involves the use of the NHTS Phoenix-Mesa sub-sample to develop a cross-classification matrix of trip generation and linear regression models of person and household trips. You are provided the corresponding household (Hhldtrips.xlsx) and person trip (persontrips.xlsx) files.

1. Develop cross-classification matrix of household trip rates by household six (1, 2, 3 or more), number of vehicles (0, 1, 2, 3 or more), and number of workers (0, 1, 2 or more). Identify two variables that you think are important factors affecting trip generation but missing from the cross-classification matrix developed here.
2. Estimate at least two multiple linear regression models of total person trips. Try several different model specifications and representations for explanatory variables including, for example, the representation of age and income as a series of dummy variables.
3. Estimate at least two multiple linear regression models of total household trips. Try several different model specifications and representations for explanatory variables including, for example, the representation of age and income as a series of dummy variables.
4. Estimate separate Person Trip Linear Regression Models for ADULT Males and Females

Data dictionary for person trip file

Variable	Description and categories
driver	`-1 appropriate skip, 1 yes a driver, 2 not a driver
worker	1 = yes , 2 = no
educ	`-1 appropriate skip, -7 refused, 1 less than high school, 2 and greater - greater than HS
hhincttl	see household file
numadlt	see household file
drvrcnt	see household file
r_age	age of person
r_sex	1= male, 2 = female
hhsize	see household file
homeown	1 = own, 2 = rent
pertrips	number of person trips

Data dictionary for household trip file

Variable	Description and Categories
homeown	1 = own, 2 = rent
hhvehcnt	number of vehicles in household
hhsize	Number of people in household
drvrcnt	number of drivers in household
wrkcount	number of workers in household
numadlt	number of adults in household
trpmiles	total number of miles traveled in household

hhincttl	-7 = Refused -8 = Don't know -9 = Not ascertained 01 = < $5,000 02 = $5,000 - $9,999 03 = $10,000 - $14,999 04 = $15,000 - $19,999 05 = $20,000 - $24,999 06 = $25,000 - $29,999 07 = $30,000 - $34,999 08 = $35,000 - $39,999 09 = $40,000 - $44,999 10 = $45,000 - $49,999 11 = $50,000 - $54,999 12 = $55,000 - $59,999 13 = $60,000 - $64,999 14 = $65,000 - $69,999 15 = $70,000 - $74,999 16 = $75,000 - $79,999 17 = $80,000 - $99,999 18 = > = $100,000

Model Requirements

For any models estimated, perform a detailed analysis to ensure statistical significance of various explanatory variables, goodness of fit, heteroscedasticity and normality assumptions.

Final Report

Please submit a professional report (PDF version only) on CANVAS. Please ensure to follow these guidelines:

1. Your report should look professional and include a title page, introduction, methodology, results and conclusion sections. Your Stata code should be included in the appendix.
2. The outputs (tables and graphs) from utilizing the statistical software should be cleaned up for improved legibility and comprehension. Figures and tables using snipping tool are NOT acceptable.
3. Your write up should be at most 15 pages (including figures, tables and Stata code). This is not very long and you should be concise and to the point.
4. Please include your Stata Code in the appendix.
5. When submitting any work, your objective is to communicate information to the reader (in this case, your instructor) in a Clear, Concise, Complete, Careful, and Courteous manner (5 C’s of good writing). If your work does not possess the “5C” qualities and/or does not adhere to the guidelines specified below, you will definitely lose A LOT OF credit even though you may have the correct answer(s).

Submission

1. Please follow the deliverables and due dates in Table 2 of this document.