BT1101 Final Take-Home Assessment

BT1101 Student. REPLACE WITH YOUR NAME

27 April 2020, 5:00 - 7:00 PM

Instructions

- **Rename your R Markdown file** `FA_[MatricNumber].rmd`, and the output will automatically be `FA_[MatricNumber].html`.
- Select `output: html_document`.
- Include all code chunks, so include `echo=TRUE` in all chunks.
- Replace the placeholder text, "Type your answer here.", with your own.
- Preinstall and include any `library('package_name')` statements before the assessment starts. Remember that there is no Internet connection when the assessment is in progress.
- Please copy and paste the question code from Examplify to this Rmarkdown file in the order of exam questions.
- You could code and type your answer in this Rmarkdown file during exam but **keep in mind that you need to copy and paste all your answer (r-chunk and text) into Essay Answer section for each question in Examplify.**
- Submit your both R Markdown file (.rmd) and HTML (.html) to the folder "Final Take-Home Assessment Submission" in Luminus after Examplify submission and Internet reconnection.
- This Rmarkdown file serves as a reference. **Only answers submitted in Examplify will be graded.** Zero point will be given for blank submission in Examplify even if you have submitted a complete Rmarkdown and HTML files.

Preparation

```
# general use
library(readxl)
library(dplyr)
library(tidyr)
library(ggplot2)
library(wooldridge)

# descriptive analytics
library(psych)
library(Rmisc)
library(rcompanion)
library (rpivotTable)
library(EnvStats)
library(car)

# predictive/prescriptive analytics
library(TTR)
library(forecast)
library(olsrr)
library(factoextra)
library(caret)
```

```
library(tseries)
library(lpSolve)
```

Question BT1101-CEOSalary Descriptive Analytics

Context: Use data('ceosal1'). This is a data set from a 1991 issue of Business Week of CEO's salaries in 1990.

The data.frame has 209 observations on 12 variables: - `salary`: 1990 salary, thousands $ - `pcsalary`: percent change salary, 1989-1990 - `sales`: 1990 firm sales, millions $ - `roe`: return on equity, 1988-1990 avg - `pcroe`: percent change roe, 1988-1990 - `ros`: return on firm's stock, 1988-1990 - `indus`: =1 if industrial firm - `finance`: =1 if financial firm - `consprod`: =1 if consumer product firm - `utility`: =1 if transport or utilities - `lsalary`: natural log of salary - `lsales`: natural log of sales

```
data('ceosal1')
CS<-ceosal1
```

(a) We will start by understanding a bit more about the CEO salaries.

-i) Plot a histogram for `salary`. What can you deduce about the distribution of CEO salaries from this histogram? (1 mark)

ii) Compute the prediction intervals for the `salary` of a CEO using alpha=0.05. What does this prediction interval tell us (2 marks)?

(b) What proportion of CEO's had `salary` greater than or equal to $1 million? (1 mark)

In the past, the proportion of CEO with salary greater than or equal to $1 million has averaged around 0.50 (i.e. population proportion for `salary` >= 1million is 0.50). Based on the sample in this dataset, is there sufficient evidence to conclude that the population proportion for salary being greater than or equal to $1million is equal to 0.5? State your hypotheses and conclusions in your answers. (3 marks)

(c) Finally, we would like to explore if CEO salaries vary across firm types.

- (i) First, create a variable `firm` and label it as "1" if it is an industrial firm, "2" if it is a financial firm, "3" if it is a consumer product firm, and "4" if it is a transport or utilities firm. (1 mark) - (ii) Now, conduct the appropriate test(s) to assess if the mean CEO salary is significantly different across firm type. Present your hypotheses, the results and conclusions. (6 marks)

Question BT1101-FamilySavings Predictive Analytics (Linear Regression)

Preparation

```
# general use
library(dplyr)
library(tidyr)
library(ggplot2)
library(wooldridge)

# descriptive analytics
```

```
library(psych)
library(rcompanion)
library (rpivotTable)
library(EnvStats)
library(car)

# predictive/prescriptive analytics
library(tseries)
library(forecast)
library(lpSolve)

# for any randomness
set.seed(1)

# for Monte Carlo simulation, sample size set to be 100, e.g.
n_samples = 100
```

Question (total 15 points)

- Data set: `saving` in `wooldridge`.

```
# load the data set, make sure you already load `wooldridge` package
data(saving)
```

This data set contains 100 observations with following variables:

- `sav`: annual savings, $
- `inc`: annual income, $

- `size`: family size
- `educ`: years education, household head
- `age`: age of household head
- `black`: =1 if household head is black
- `cons`: annual consumption, $

(a) It is interesting to explain the individual saving `sav` from various explanatory variables.

- **Run a linear regression model of `sav ~ inc + cons`, report the regression output and write out the fitted line.** (2 points)
- **Interpret the intercept as well as the coefficients before `inc` and `cons`.** (3 points)
- **Can we use `sav`, `cons` and `inc` altogether as explantory variables in linear regression? Why or why not?** (2 points)

Type your answer here

(b) Some labor economists believe that individual consumption is affected not only by income but by his or her education level nonlinearly as well. Let's first create the following variable `educsq` which is equal to square of `educ`.

```
saving$educsq = saving$educ^2
```

- **Investigate the relationship between `cons` and `inc` visually. Is there a linear relationship between the two?** (1 point)
- **Run and report a linear regression of `inc`, `educ` and `educsq` to explain `cons`. What's the marginal effect of education on consumption for those individuals' education level being 6 year? Interpret this marginal effect.** (3 points)

Type your answer here

(c) From the regression output in part (b), - **Is there a nonlinear relationship between education and consumption? Why or why not?** (1 points) - **Check the assumptions and discuss what you find. Which assumption(s) are you worry about?** (3 points)

Type your answer here

Question BT1101-AssemblingPC Prescriptive Analytics (Linear Optimization) Question (total 15 points)

Consider that you are a gamer who tries to build your own gaming PC. Suppose a working PC requires at least a CPU, a GPU, a RAM and a motherboard. Luckily your best friend gave you a motherboard as your birthday present. The motherboard supports exactly one CPU and comes with two GPU and four RAM slots. You visited your favorite local PC gearshop where you collected information about the other parts they offered:

Parts	In Stock	Category	Delivered Performance (FPS)	Price ($)
Intel i9	1	CPU	20	499
AMD Ryzen	1	CPU	18	450
Nvidia RTX 2080	1	GPU	90	1099
AMD Radeon 56	3	GPU	75	627
8G RAM card	8	RAM	10	100

Assume that the FPS performance of the gaming PC is simply sum of delivered FPS of each part. As a first-person shooting gamer, you only care about the performance of frame-per-second (FPS) and requires at least 144 FPS for your new PC. As a student, you obviously would like to spend the least amount of money to assemble a working PC meeting your standard.

(a) Write down your decision variables, the objective function, and ALL constraints that apply for this optimization problem in a table. You do not need to solve the problem for now. (5 points)

Type your answer here

(b) Your trustworthy shop specialist strongly recommended if you choose to get the Nvidia RTX 2080 GPU, your PC needs at least 16GB of RAM.

- **Write down the constraint in its math expression.** (2 points)

- **As you decide to follow the specialist's suggestion, solve this problem in R. Which part(s) should you purchase to assemble your gaming PC? How much do you spend on it?** (4 points)

Type your answer here

(c) Now suppose that the price for Nvidia RTX 2080 GPU is uncertain, fluctuating $\pm 55\%$ uniformly based on the current price. You have a budget and thus can pay at most $1500 for your new PC. What's the prbability that you are not able to get the PC you want at the moment? (4 points)

Type your answer here