Review of the Course

Good luck!

Your Final

Your final exam will be an open-book open-notes exam.

Calculators are allowed (but I don't think you need a calculator).

Other electronic devices (smartphones, tablets, and laptops etc.) are not allowed.

You can bring any print materials (e.g., slides, your own notes, any books).

Your Final

All questions in your final come from the slides.

No questions on data scraping.

No questions on using Tableau; however, there may be questions asking you the meaning of Tableau plots.

Slides marked with “optional” will not be tested. Maximum likelihood would not be tested.

Emphasizes on data analysis methodologies.

Your Final

All questions in your final come from the slides.

50 questions, 120 minutes – you should have sufficient time to answer the questions.

Single choice; four options (ABCD) for each question – choose the most appropriate answer only.

Detailed arrangement on final will be sent to you later.

Sample Questions for your final

Which of the following statements is FALSE regarding algorithms?

A. An algorithm only works for a specific programming language
B. An algorithm can be used to solve a specific problem.
C. Algorithms can be used for purposes other than data analysis.
D. Algorithms are instructions that can be implemented by computer programs.

Sample Questions for your final

Suppose that you have an R data frame named "Individuals", and the data frame has one variable called "Name". Which of the followings help you select that variable?

A. Individual#Name
B. Individual@Name
C. Individual$Name
D. Individual%Name

Sample Questions for your final

Suppose that you want to run some codes when variable A is equal to B, you should write

A. If (A==B) {your codes}
B. If (A==B) then {your codes}
C. If (A=B) {your codes}
D. If (A=B) then {your codes}

Sample Questions for your final

Suppose that you want to run some codes when variable A is equal to B, you should write

A. If (A==B) {your codes}
B. If (A==B) then {your codes}
C. If (A=B) {your codes}
D. If (A=B) then {your codes}

Sample Questions for your final

When you write runit(2), you are generating ___.

A. A random number between 0 and 1.
B. Two random numbers between 0 and 1.
C. A random number between 0 and 2.
D. Two random numbers between 0 and 2.

Sample Questions for your final

Consider the following R codes:

```
x <- '111'
y <- '222'
Z <- x+y
```

What is the value of Z?

A. 333
B. 111 222
C. 111222
D. You will get an error message

Sample Questions for your final

Suppose that your dependent variable is the outcome of the final exam, which takes two values (pass or fail). Your independent variables are hours of study and class fixed effects. Which of the following methods is most appropriate for your analysis?

A. Linear regression
B. Linear-Log regression
C. Probit regression
D. Multinomial logit model

Sample Questions for your final

Fixed effects can be used in

A. Linear regression
B. Multinomial logit regression
C. Logistic regression
D. All of the above

Sample Questions for your final

Suppose that your dependent variable is the program choice of a master student (e.g., BSc Marketing, BSc BA, BSc Finance). Which of the following methods is most appropriate for your analysis?

A. Linear regression
B. Linear-Log regression
C. Probit regression
D. Multinomial logit model

Sample Questions for your final

You should consider taking the log transformation of a variable when

A. The variable will be used as a fixed effect
B. The variable is a binary variable, i.e., it only takes values 0 and 1
C. The variable has a left skewed distribution
D. The variable has a right skewed distribution

Sample Questions for your final

Which of the followings is FALSE regarding experiments?

A. Participants are randomly assigned into different groups.
B. Experiments can be used to prove causal relationship.
C. It is often impossible to run an experiment.
D. In an experiment, each group must have the same number of participants.

Sample Questions for your final

Suppose that you are regressing Y on X, and you use Z as an instrumental variable. Which of the followings is FALSE?

A. Z should be correlated with X
B. Your regression equation is $Y = a + b1\ X + b2\ Z$
C. Z should not affect Y directly
D. There may be multiple variables that can be used as your instrumental variable.

Sample Questions for your final

Which of the followings is TRUE regarding sentiment analysis?

A. Sentiment analysis only works for English
B. The lexicon-based approach for sentiment analysis deals with sarcasm perfectly
C. Sentiment analysis allows us to understand the polarization (positive vs negative) of a text
D. The lexicon-based approach is more accurate than the machine-learning based approach for sentiment analysis.

Sample Questions for your final

Suppose that you are analysing the sentiment of the following sentence: “HKU has a lot of problems, but it is still my favorite school.” Which of the following statements is TRUE?

A. The overall sentiment of the sentence is neutral.
B. We should put more weight on the second part of the sentence.
C. We should put more weight on the first part of the sentence.
D. The sentiment cannot be analyzed.

Sample Questions for your final

Through data plotting, you find out that your dependent variable and independent variable has an inverted U shape relationship. Which of the following regression models will you consider?

A. $Y = a + b\,X$
B. $Y = a + b\,X^2$
C. $Y = a + b_1X + b_2X^2$
D. $Y = a + b_1X + b_2X^2 + b_3X^3$

Sample Questions for your final

Suppose that you are regressing students' final grades on their IQ level and hours study. However, you believe that students with higher IQ studies more efficiently and can understand concepts easier and faster. Then, you should use

A. IQ fixed effects
B. Interaction between IQ and hours of study
C. Neural network models
D. None of the above