## **Types of Data Science Questions**

Question 1

Which of the following describes a predictive analysis?

Finding if one variable is related to another one

Showing the effect on a variable of changing the values of another variable

**Using data collected in the past to predict values in the future**

Question 2

We collect data on all the songs in the Spotify catalogue and want to summarize how many are country western, hip-hop, classic rock, or other. What type of analysis is this?

Exploratory

**Descriptive**

Predictive

Question 3

We collect data on a small sample of songs from the Spotify catalogue and want to figure out the relationship between the use of the word "truck" and whether a song is country western. What type of analysis is this?

Descriptive

**Inferential**

Exploratory

## **Experimental Design**

Question 1

In a study measuring the effect of diet on BMI, cholesterol, lipid levels, triglyceride levels, and glycemic index, which is an independent variable?

**Diet**

BMI

Lipid levels

Question 2

Which of the following is NOT a method to control your experiments?

Control group

**Placebo effect**

Blinding

Question 3

What might a confounder be in an experiment looking at the relationship between the prevalence of white hair in a population and wrinkles?

Socioeconomic status

Smoking status

**Age**

Question 4

According to Leek group recommendations, what data do you need to share with a collaborating statistician?

The raw data

A tidy data set

A code book describing each variable and its values in the tidy data set

An explicit and exact recipe of how you went from the raw data to the tidy data and the code book

**All of the above**

Question 5

If you set your signifance level at p-value ≤ 0.01, how many significant tests would you expect to see by chance if you carry out 1000 tests?

**10**

50

100

Question 6

What is an experimental design tool that can be used to address variables that may be confounders at the design phase of an experiment?

Using regression models

**Randomization**

Using data from a database

Question 7

Which of the following describes a descriptive analysis?

Use your sample data distribution to make predictions for the future

Draw conclusions from your sample data distribution and infer for the larger population

**Generate a table summarizing the number of observations in your dataset as well as the central tendencies and variances of each variable**

# **Big Data**

Question 1

Which is NOT one of the three V's of Big Data?

Volume

**Valuable**

Variety

Question 2

Which one of the following is an example of structured data?

**A table of names and student grades**

Lung x-ray images

The text from a series of books

Question 3

What is the reason behind the explosion of interest in big data?

**The price and difficulty of collecting and storing data has dramatically dropped**

There have been massive improvements in machine learning algorithms

There have been massive improvements in statistical analysis techniques

## **Module Four Summative Quiz**

Question 1

What is the format for including a link that appears as blue text in your markdown document?

**[text that is shown](link.com)**

(link.com)[text that is shown]

(text that is shown)[link.com]

Question 2

Which of the following describes a predictive analysis?

**Using data collected in the past to predict values in the future**

Finding if one variable is related to another one

Showing the effect on a variable of changing the values of another variable

Question 3

We collect data on all the songs in the Spotify catalogue and want to summarize how many are country western, hip-hop, classic rock, or other. What type of analysis is this?

Exploratory

**Descriptive**

Predictive

Question 4

What might a confounder be in an experiment looking at the relationship between the prevalence of white hair in a population and wrinkles?

**Age**

Socioeconomic status

Sex

Question 5

Which one of the following is an example of structured data?

The text from a series of books

Lung x-ray images

**A table of names and student grades**