

Assessment: Correlation is Not Causation

Question 1

1/1 point (graded)

In the videos, we ran one million tests of correlation for two random variables, X and Y.

How many of these correlations would you expect to have a significant p-value ($p \leq 0.05$), just by chance?

☐ 5,000

☒ 50,000

☐ 100,000

☐ It's impossible to know



Answer

Correct:

Correct. In this example, the chance of finding a correlation when none exists is $0.05 \times 1,000,000$ chances.

Explanation

The p-value is defined as the probability of finding the observed result when the null hypothesis (no correlation) is true. When we have a p-value of 0.05, this means the chance of finding a correlation when none exists is 5% - e.g., $0.05 \times 1,000,000$ chances, which is 50,000.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 2

1/1 point (graded)

Which of the following are examples of p-hacking?

Select ALL that apply.

- ☒ Looking for associations between an outcome and several exposures and only reporting the one that is significant.
- ☒ Trying several different models and selecting the one that yields the smallest p-value.
- ☒ Repeating an experiment multiple times and only reporting the one with the smallest p-value.
- ☐ Using a Monte Carlo simulations in an analysis.



Answer

Correct:

Correct, this is one of the three examples of multiple testing included in the video.

Correct, this is one of the three examples of multiple testing included in the video.

Correct, this is one of the three examples of multiple testing included in the video.

Explanation

Repeating an experiment multiple times and only reporting the one with the smallest p-value, looking for associations between an outcome and several exposures and only reporting the one that is significant, and trying several different models and selecting the one that yields the smallest p-value are all examples of p-hacking.

Monte Carlo simulations do not necessarily lead to multiple testing problems such as p-hacking in and of themselves.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 3

1/1 point (graded)

The Spearman correlation coefficient is robust to outliers because:

- ☐ It drops outliers before calculating correlation.
- ☐ It is the correlation of standardized values.
- ☒ It calculates correlation between ranks, not values.



Explanation

Because the Spearman correlation coefficient uses ranks instead of values to calculate correlation, it is more robust to outliers.

Submit

You have used 1 of 1 attempt

i Answers are displayed within the problem

Question 4

1/1 point (graded)

Which of the following may be examples of reversed cause and effect?

Select ALL that apply.

☒ Past smokers who have quit smoking may be more likely to die from lung cancer.

☐ Tall fathers are more likely to have tall sons.

☒ People with high blood pressure tend to have a healthier diet.

☒ Individuals in a low social status have a higher risk of schizophrenia.



Answer

Correct:

Correct - it is possible that a smoker would quit after being diagnosed, but is at a higher risk of dying. It does not mean that quitting smoking increases your risk of dying from lung cancer.

Correct. This could be reverse causation. Individuals with high blood pressure may have a healthier diet because they are trying to improve their blood pressure, not because a healthy diet causes high blood pressure.

Correct. This could be reverse causation. It is possible that individuals with schizophrenia are more likely to have a lower social status due to the challenges of managing their disease.

Explanation

"Past smokers who have quit smoking may be more likely to die from lung cancer" is an example because it is possible that a smoker would quit after being diagnosed, but is at a higher risk of dying. However, it does not mean that quitting smoking increases your risk of dying from lung cancer. For the statement "tall fathers are more likely to have tall sons," we know the order of causation because the fathers came before the sons, so this is not a case of reversed cause and effect.

"People with high blood pressure tend to have a healthier diet" could be reverse causation. Individuals with high blood pressure may have a healthier diet because they are trying to improve their blood pressure, not because a healthy diet causes high blood pressure.

"Individuals in a low social status have a higher risk of schizophrenia" could be reverse causation. Individuals with schizophrenia may be more likely to have a lower social status due to the challenges of managing their disease.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 5

1/1 point (graded)

What can you do to determine if you are misinterpreting results because of a confounder?

- ☐ Nothing. If the p-value says the result is significant, then it is.
- ☒ More closely examine the results by stratifying and plotting the data.

☐ Always assume that you are misinterpreting the results.

☐ Use linear models to tease out a confounder.



Answer

Correct:

Correct. Although you can sometimes use linear models, you can't always and exploratory data analysis (stratifying and plotting data) will help determine if there is a confounder.

Explanation

Exploratory data analysis (stratifying and plotting data) can help determine if there is a confounder. Linear models cannot be used in all situations.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 6

1/1 point (graded)

Look again at the admissions data presented in the confounders video using `?admissions`.

What important characteristic of the table variables do you need to know to understand the calculations used in this video?

- ☐ The data are from 1973.

- ☐ The columns `major` and `gender` are of class character, while `admitted` and `applicants` are numeric.
- ☐ The data are from the `ds1labs` package.
- ☒ The column `admitted` is the percent of students admitted, while the column `applicants` is the total number of applicants.



Answer

Correct:

Correct. In all data science projects, it is important to understand the data that you are working with.

Explanation

Several of these statements are true but not relevant to understanding the calculations in the video. The only statement that is critical for the analysis is that "The column `admitted` is the percent of students admitted, while the column `applicants` is the total number of applicants." In all data science projects, it is important to understand the data that you are working with.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 7

1/1 point (graded)

In the example in the confounders video, major selectivity confounds the relationship between UC Berkley admission rates and gender because:

- ☐ It was harder for women to be admitted to UC Berkeley.
- ☒ Major selectivity is associated with both admission rates and with gender, as women tended to apply to more selective majors.
- ☐ Some majors are more selective than others.
- ☐ Major selectivity is not a confounder.



Explanation

Major selectivity is a confounder because it is associated with both admission rate and with gender.

Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Question 8

1/1 point (graded)

Admission rates at UC Berkeley are an example of Simpson's Paradox because:

- ☒ It appears that men have higher a higher admission rate than women, however, after we stratify by major, we see that on average women have a higher admission rate than men.
- ☐ It was a paradox that women were being admitted at a lower rate than men.
- ☐ The relationship between admissions and gender is confounded by major selectivity.



Answer

Correct:

Correct, this is a good explanation of why this example is considered an example of Simpson's Paradox

Explanation

Simpson's Paradox refers specifically to cases where the sign of the correlation flips when comparing the entire dataset vs. specific strata, so only the first statement is correct.

Submit

You have used 1 of 1 attempt

i Answers are displayed within the problem