Case Study 5

Multivariable Thinking

Multivariable thinking involves investigating how multiple variables interact with a response variable and with each other. Looking at trends in data with multiple variables can help improve statistical thinking, and enhance your ability to ask important questions about documented research. Many research questions are so complex that they demand multivariable statistics, and now can be explored with the advancement of technology.(Wuensch, 2019). Further, individuals who possess a single-variable mental model of causality risk not thinking deeply enough to accurately detect and understand how the world works (Ramsey, 2014).

In this activity, you will explore multivariable thinking by exploring a paradox in statistics called Simpson's paradox, and practice identifying confounding variables through a variety of contexts.

Simpson's Paradox

Simpson's paradox is a phenomenon in probability and statistics where a trend appears in several different groups of data but disappears or reverses when these groups are combined.

Learning Objectives

- Understanding characteristics of graphs
- Explain Simpson's paradox in detail
- Discover the value in multivariable thinking, including identifying and describing confounding variables

Please read the following article and answer the questions below:

https://towardsdatascience.com/simpsons-paradox-how-to-prove-two-opposite-arguments-using-one-dataset-1c9c917f5ff9

- Describe the characteristics of the scatter plot: "Probability vs Exercise over 50." Comment on strength, direction, outliers, and form.

- Describe the same characteristics in the plot: "Probability vs Exercise Combined." What do you notice that is different from the previous plot?

Please watch the video here and answer the following questions below: https://www.youtube.com/watch?v=ebEkn-BiW5k

- In your own words, describe why you think Simpson's paradox could be dangerous for scientific publications.

Part 2: Confounding Variables

- Corneisha is doing an experiment to determine whether the amount of time a person spends reading affects their SAT score. After a few of the subjects come to the lab and read for a certain amount of time, one of Corneisha's lamps burns out and therefore makes the lab much darker for the rest of the subjects. Identify the explanatory variable, response variable, and the potential confounding variable in the context above.

- It's your turn! Come up with a research question that explores the relationship between two variables. With your knowledge of confounding variables, identify a third variable you would also wish to consider. Justify WHY this is a confounding variable.

- Provide a recommendation on how researchers could help alleviate Simpson's paradox in their publications. Be creative. Post your your response to the discussion post "Simpson's Paradox" on D2L.

- Reflect on the learning objectives above. If you have any questions or comments, please post on the "Simpson's Paradox Questions" discussion post.

Extra Information on Confounding Variables

https://www.scribbr.com/methodology/confounding-variables/

(Optional Reading): Simpson's Paradox in the world today

Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects

https://arxiv.org/pdf/2005.07180.pdf