Data Wrangling.

By
Herminnia Jessica Flores

Scenario #1

A store is investigating the influence of gender upon whether customers sign up for a discount club card. Options for gender are male and female, and options for signing up for the club card are signed up and not signed up.

- 1) The IV is the Gender. It Is a Binary Categorical Variable.
- 2) There are 2 binary categorical levels of the IV, Male & Female.
- 3) DV is signing up for the club card .
- 4) There are 2 Binary Categorical levels of the DV, signed up & not signed up.
- 5) an Independent Chi Square would be the appropriate analysis to run as we are NOT comparing a sample to a population and both the IV & DV are categorical variables with only 2 levels

Scenario #2

This same store has just expanded their club card system. They now have three different tiers - silver, gold, and platinum. They would like to know whether the type of club card the customer has dictates how much money the customer spends.

- 1)The IV is The Club Card System. It is Categorical.
- 2)There are Levels of the single IV. Silver, Gold, & Platinum.
- 3)The DV is The Amount of Money The Customer Spends. It is Continuous.
- 4)There Are No Levels to the DV
- 5) The Appropriate Analysis Would Be ANOVA Because We DO NOT Want To Control Other Factors.

Scenario #3

Now, the store manager would like to know: Do people spend more money before or after they get a club card?

- 1)The IV is The Club Card. It is Categorical.
- 2)There Are 2 Levels of the IV. Before Having One & After.
- 3)The DV is The Amount of Money Spent. It Is Continuous.
- 4)There are no Levels of the DV.
- 5)Simple Linear Regression Would Be The Best Analysis because there is only one dependent variable.

Scenario #4

Lastly, the store manager would like to know if the age of a customer predicts whether that customer will sign up for a club card or not.

- 1)The IV Is The Age. It Is Continuous.
- 2)It Is A Multilevel IV.
- 3)The DV Is Club Card and It Is Categorical.
- 4)There Are 2 Levels To The DV. Signed Up or Not.
- 5)The Mcnemar Chi Square Would Be Appropriate For The Analysis because there are only 2 levels to the Dependent Variable.