

Lesson 9 Hands-On

March 2023

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 and data type
Gender, which is categorical
2. Levels of IV
2 levels: male and female
3. DV and data type
Categorical: Yes or no to signing up for a discount club card
4. Levels of DV
2: yes or no
5. Most appropriate analysis

Because both the IV and DV are categorical, we're in "Chi-Squared Land". Then, we are comparing what we may suspect to be unrelated categories (male and female), we would run an **Independent Chi-Square**.

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 and data type
The type of club card
2. Levels of IV
3 tiers- Silver, gold and platinum.
3. DV and data type
Continuous: Money the customer spends
4. Levels of DV
n/a
5. Most appropriate analysis

We have a continuous DV, and more than 2 levels of our single independent variable. We would run an ANOVA. We are not controlling for anything else, so it's a straight **ANOVA**.

Scenario 3

Now, the store manager would like to know: Do people spend more money before or after they get a club card?
Rephrased: Does getting a club card influence how much money a customer spends?

1. The IV and data type
Categorical: has a club card or doesn't have a club card
2. Levels of IV
two
3. DV and data type
Continuous: the amount of money spent
4. Levels of DV
na
5. Most appropriate analysis
The discerning factor is we are examining changes over time (before and after a customer has a club card). Therefore, we should run a **Dependent t-test**.

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 and data type
Continuous: customer age
2. Levels of IV
na
3. DV and data type
Categorical: yes or no to club card
4. Levels of DV
two
5. Most appropriate analysis
 - a. We have a continuous IV and a categorical DV, so we are focused on logistical regression. There are two levels of our DV, so we will do a form of binary logistic regression. We do not want to know how much influence each age has (or a specific age), so we are going to do **binary logistic regression**. (not totally sure if it should be stepwise binary logistic regression because I am unclear about the "how much influence something has" as I would assume they would want to know if the age could predict it, but there are not trying to tell if something is more important, so I'm sticking with this answer)