Ch 8 Definitions

- Statistical hypothesis testing - a procedure that allows us to evaluate hypotheses about population parameters based on sample statistics
- Null hypothesis (h0) - opposite of the research hypothesis to assess how likely the research hypothesis is correct. Usually states that there isn't a difference between the population mean & the specified value
- one-tailed test - is a type of hypothesis test that involves a directional hypothesis, specifies that the values of 1 group are either larger (right-tailed test) or smaller (left-tailed test) than some specified population value
- two-tailed test - is a type of hypothesis test that involves a non-direction research hypothesis. We are equally interested in whether the values are less than or greater than one another. The sample outcome may be located at both the low & high ends of the sampling distribution
- P value - the probability associated w/the obtained value of Z
- Alpha (a) - the level of probability at which the null hypothesis is rejected (usually set at .05 or .01, or .001 level)
- Type 1 error (false rejection error) - probability associated with rejecting a true null hypothesis
- Type 2 error (false acceptance error) - probability associated with failing to reject a false null hypothesis
- T statistic - used to test null hypothesis instead of z-statistic when population standard deviation isn't known & is instead estimated using the sample standard deviation
- degrees of freedom (df) - represent the number of scores that are free to vary in calculating a statistic
- t distribution - a family of curves, each determined by its degrees of freedom (df). Used when the population standard deviation is unknown & the standard error is estimated from the sample standard deviation. Appendix C summarizes the t distribution and shows df, probabilities or alpha, significance levels for both one-tailed & two-tailed tests.

Ch 9 Definitions

- Bivariate analysis - a statistical method to detect & describe relationship between 2 vars
- Cross-tabulation - a technique for analyzing the relationship between two vars that have been organized in a table
- Positive relationship - the two vars vary in the same direction. When the independent var increases, so does the dependent var

- Negative relationship - the two vars vary in opposite directions. When one var increases, the other decreases.
- Elaboration - processes designed to further explore a bivariate relationship. We look at other vars called control vars to see if they explain the bivariate relationship under consideration
- Direct causal relationship - a bivariate relationship that can't be accounted for by other theoretically relevant vars
- Spruious relationship - a relationship in which both the independent var & dependent var are influenced by a causally prior control var & there is no casual link between them. The relationship between the IV & DV is said to be "explained away" by the control var.
- Intervening var - a control var that follows an independent var but comes before the dependent var in a causal sequence
- Intervening relationship - a relationship in which the control var intervenes between the independent & dependent vars

Chapter 10 definitions

- Chi-square test - inferential statistic technique to test for significant relationships between 2 variables organized in a bivariate table. Can be applied to nominally or ordinally measured vars
- Statistical Independence - the absence of association between 2 cross-tabulated vars. The percentage distributions of the dependent var within each category of the independent var are identical.
- Research hypothesis (h1) - proposes that the 2 vars are related in the population
- Null hypothesis (h0) - states that no association exists between the 2 cross-tabulated vars in the population, & therefore the vars are statistically independent
- Expected frequencies (fe) - the cell frequencies that would be expected in a bivariate table if the 2 tables were statistically independent
- Observed frequencies (fo) - the cell frequencies actually observed in a bivariate table
- Measure of association - a single summarizing number that reflects the strength of the relationship, indicates usefulness of predicting the dependent var from the independent var, & often shows the direction of the relationship
- Proportional Reduction of Error (PRE) - concept that underlies the definition & interpretation of several measures of association. Measures are derived by comparing the errors made in predicting the dependent var while ignoring the independent var, with errors made when making predictions that use info about the indpendent var
- Lambda - asymmetrical measure of association, suitable for use w/nominal vars. Provides us w/an indication of the strength of an association between the independent & dependent vars..

- Cramer's V - Based on the value of chi-square. Since it cannot take negative values, it's considered a nondirectional measure

- Symmetrical Measure of Association - a measure whose value will be the same when either the var is considered the independent var or the dependent var

- Gamma - Symmetrical measure of association used to determine if there is an association between the independent & dependent vars. Gamma helps us determine if there is an association & if the association is positive or negative. Suitable for use with ordinal vars or dichotomous nominal vars (vars that can only have 2 possible values, EX: gender)

- Magnitude - closer to the absolute value of 1, the stronger the association. If the measure equals 0, then there's no relationship between the 2 vars.

- Direction - the sign on the measure indicates if the relationship is positive or negative. In a positive relationship, when one var is high, so is the other. In a negative relationship, when one var is high, the other is low.

Ch 8 Deeper

Assumptions of Statistical Hypothesis Testing:

- the sample is a random sample

- the level of measurement is interval-ratio

- the population is normally-distributed or that the sample size is larger than 50

The Research Hypothesis:

- Statement reflected the substantive hypothesis. It is always expressed in terms of population parameters since we're interested in making statements about population parameters based on our sample statistics

- Symbolized as H1

The research hypothesis specifies that the **population** is one of the following:

- Not equal to some specified value

- Greater than some specified value

- Less than some specified value

Null Hypothesis:

- Rather than directly testing the research hypothesis, we test the null hypothesis

- We hope to reject the null hypothesis in order to indirectly provide support for the research hypothesis

- The null hypothesis is a statement of no difference which contradicts the research hypothesis & is always expressed in terms of population parameters

Research & Null Hypotheses:

One-tail test - specifies the hypothesized direction

- Research Hypothesis:

- Null Hypothesis:

Two-tail test - direction isn't specified (more common)

- Research Hypothesis:

- Null Hypothesis:

Two-Tailed Test:

- Sometimes, we believe there's a difference between groups but we can't anticipate the direction of that difference, when we don't have an idea of the direction in the research hypothesis, we conduct a two-tailed test.

- A two-tailed test is the type of hypothesis test that involes a non-direction research hypothesis. We are equally interested in whether the values are less than or greater than one another. The sample outcome may be located at both the low and high ends of the sampling distribution

Probability Values & Alpha:

- Probability or the P value is the probability associated with the obtained value of Z

- Alpha (a) is the level of probability at which the null hypothesis is rejected (usually set alpha at the .05, .01, or .001 level). Alpha is considered a specific cut off point below which P must fall to reject the null hypothesis.

  - 0.05 is a 5 in 100 chance and it's most likely to make a Type 1 error with but it's the easiest to reject null with
  - 0.001 is a .1 in 100 chance and it's the most likely to make a Type 2 error with and it's the hardest to reject null with
  - 0.01 is a 1 in 100 chance
- If the P value is equal to or less than our alpha level, our obtained Z statistic is considered statistically significant. In other words, it's very unlikely to have occurred by random chance or sampling error.

- If P value is equal to or below the alpha level (a), we reject null hypothesis

- If P value is above the alpha level (a), we fail to reject the null hypothesis

Z for a Two-Tailed Test:

- To find P for a two-tailed test, look up the area in column C of appendix B that corresponds to your obtained Z, then multiply it by 2 to obtain the two-tailed probability

- NOTE: Rejecting null hypothesis for a two-tailed test is more difficult than it is for 1-tailed test

The 5 Steps to Hypothesis Testing:

1. Making assumptions
2. Stating research & null hypotheses & selecting alpha
3. Selecting the sampling distribution & specifying the test statistic
4. Computing test statistic
5. Making a decision & interpreting the results

The T-Statistic:

- In most situations, standard deviation isn't known as it was for the Z statistic. In these cases, we use the T statistic instead of the Z statistic to test the null hypothesis

- Formula:

    - Replace denominator with standard error
    - After calculation, look into C Column which tells us how improbable or probable (p value) it is to pull the sample mean from the population mean
    - Converting the sample mean to a Z-score equivalent is called computing the test statistic, but here, it'd be converting the sample mean to a T-score equivalent.
    - A negative Z-statistic would mean the difference would have to be evaluated at the left tail of the distribution
    - A positive Z-statistic would mean the difference would have to be evaluated at the right tail of the distribution

T Distributions & Degrees of Freedom:

- T distribution is a family of curves, each determined by its degrees of freedom

- Degrees of freedom (df) represent the number of scores that are free to vary in calculating each statistic

- When the degrees of freedom is small, the t distribution is much flatter than the normal curve. But as the df increases, the shape of the distribution gets closer to the normal distribution, until the two are almost identical when the df is greater than 130

Calculating the Degrees of Freedom:

- When calculating degrees of freedom, you **must** known the sample size & whether there are any restrictions in calculating that statistic

- Then the number of restrictions is subtracted from the sample size
- Amount of restrictions is defined as the number of samples

T Distribution:

- Appendix C summarizes the t distribution
- The t table differs from the z table by showing the df, probabilities or alpha, significance levels are also shown for both one-tailed tests and two-tailed tests

T-Tests for Difference Between Means:

- The sampling distribution allows us to compare our sample results with all possible sample outcomes & estimate the likelihood of their occurence
- To get the t-test score, we need the difference between the 2 means, the standard error, & the degrees of freedom, farther down are the equations required to find these.
- Formula:

    - Find mean
    - Find N (number of samples)
    - Find variance (to get variance from standard deviation, square the standard deviation)

T-Tests for Difference Between Means if Population Variances appear Equal:

T-Tests for Difference Between Means if Population Variances appear Unequal:

Ch 9

Bivariate Analysis:

- a bivariate analysis is a statistical method designed to detect & describe the relationship between 2 variables
- EX of a relationship: when the temperature rises, there is more criminal activity

Cross-Tabulation:

- Cross-tabulation (cross-tabs) is the most common technique used in the analysis of relationships between 2 variables

- The variables are organized on a table to demonstrate the strength of the association & when appropriate its direction

Independent & Dependent Variables:

- For some variables, whether it is the independent or dependent variable depends on the research question. EX: People who live in poverty have a longer problem with drug use, here poverty would be the independent variable, influencing the drug use. VS. People who use drugs tend to live in poverty, here drug use is the independent variable influencing poverty

How to Construct a Bivariate Table:

- A bivariate table displays the distribution of one variable across the categories of another variable

- Think of it as a series of frequency distributions joined to make one table

Bivariate Tables Should Have the Following Features:

- a title that is descriptive

- it should have 2 dimensions, one for each variablce (usually the independent variable is the column variable and the dependent variable is the row variable)

- The intersection of a row & a column is called a cell

- The column & row totals are the frequency distribution for each variable. The column total is the frequency distribution for the independent variable & the row total is the frequency distribution for the dependent variable

- We refer to tables as RxC, row by column. Think of rows as side-to-side (row-boat) and columns as Greek columns (up & down)

- The source of the data should always be clearly noted so that we're giving credit to the person who came up with the info

- You should know what a bivariate table is, a column variable, a row variable, and marginals

How to Compute Percentages in a Bivariate Table:

- When we use percentages it's easier for the reader to see the results

- To calculate percentages you must calculate the percentages within each category of the independent variable

- Interpret the table by comparing the percentage point difference for different categories of the independent variable

Calculating Percentages within Each Category:

- The first rule means that we have to calculate percentages within each category of the variable that the investigator defines as the independent variable

Comparing the Percentages:

- The second rule says you must compare the percentages across different categories of the independent variable

Properties of a Bivariate Table:

3 questions we can ask when examining a bivaraite relationship:

1. Does there appear to be a relationship?
2. How strong is the relationship?
3. What is the direction of the relationship

The Existence of the Relationship:

- When we calculate the percentage for the categories, we can then compare the percentages to see if there is a relationship between the variables

The Strength of the Relationship:

- Once you have established that there is a relationship, the next step is to see the strength of the relationship

- The easiest way to examine the percentage difference across the different categories of the independent variable. The larger the difference across the categories, the stronger the association.

The Direction of the Relationship:

- When both the independent & dependent variable in the bivariate table are measured at the ordinal level, we can look at the direction of the relationship.

- There are 2 types of relationships: Positive relationships & negative relationships

Positive Relationship:

- In a positive bivariate relationship, the two variables vary in the same direction

- When the independent variable increases, so does the dependent variable

Negative Relationship:

- In a negative bivariate relationship, the two variables vary in opposite directions

- When one variable increases, the other variable decreases

Elaboration:

- Elaboration is the process designed to further explore a bivariate relationship

- We look at the other variables called control variables to see if they explain the bivariate relationship under consideration

3 Goals of Elaboration:

- Elaboration allows us to test for nonspuriousness

- Elaboration clarifies the causal sequence of bivariate relationships by introducing variables hypothesized to intervene between the independent variable & dependent variable

- Elaboration specifies the different conditions under which the original bivariate relationship might hold

Why Would we Introduce a Control Variable:

1) Elaboration allows us to test for nonspuriousness:

- This shows us if there is a cause & effect relationship between the independent & dependent variables in the bivariate relationship
- It also establishes the time order between them
- Also, we want to know if there is another variable that better fits the relationship

2) Elaboration clarifies the causal sequence:

- It does this by introducing variables hypothesized to to intervene between the independent & dependent variables
- EX: What else might cause the dependent variable besides the independent variable?

3) Elaboration specifies conditional relationships:

- A conditional relationship is a relationship between the independent & dependent variables that differs for different conditions of the control variable
- What other variable may explain the relationship between father's education & son's income

Intervening Variable:

- Is there a direct causal link between the 2 variables. Sometimes the variables are linked but, there is an intervening variable that explains the relationship even more:

Limitations of Elaboration:

- Elaboration really is just a stab in the dark, you're trying to see if other variables (control variables) will explain the relationship better
- The problem arises when you miss an important variable that may help explain the relationship.