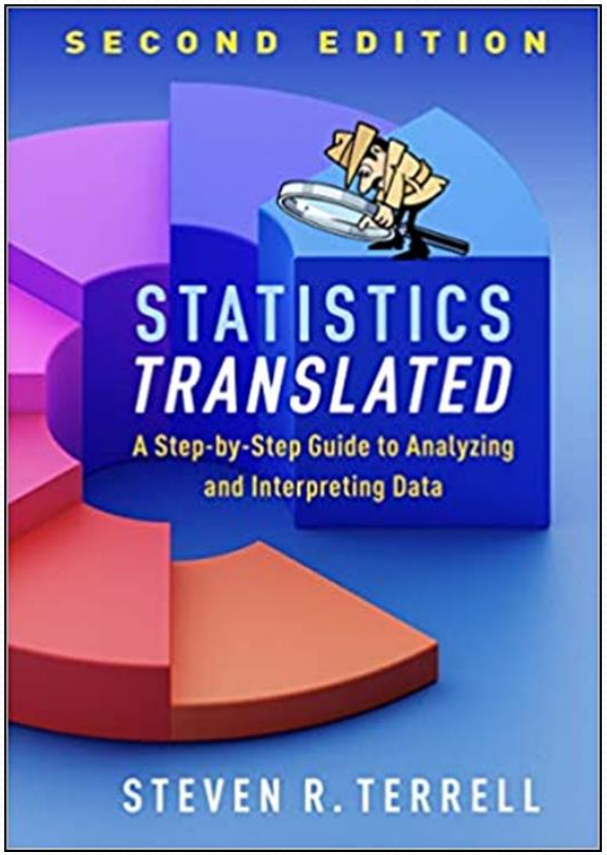
Data Types and Quantitative Analysis

Steven R. Terrell, Ph.D.
www.finishyourphd.com
info@finishyourphd.com



Click on Picture for Book Details.

Quantitative Research

- Deductive by nature (tests a hypothesis based on a theory)
- Hypotheses and research procedures stated prior to starting the study
- Numeric data is collected and analyzed
- Cause and effect is examined
- Descriptive and inferential statistical procedures are use to test the hypotheses.
- The appropriate inferential statistical test can generally be determined by examining the component parts of the hypothesis.

Independent and Dependent Variables

Independent variable – anything that we believe makes a difference in a behavior (i.e., the cause).

Levels of the Independent Variable – Categories within the Independent variable (e.g., gender has two levels – males and females).

Dependent variable – anything that is being measured (i.e., the effect).

For example, if we look at the difference in income by gender, gender is the Independent variable (the cause), the Levels are male and female, and income is the Dependent variable.

Hypothesis:

Children who are assigned to technology-based classrooms will have higher achievement than children assigned to lecture-based classrooms.

In this hypothesis, the independent variable is the type of instruction received. It has two levels: technology-based and lecture-based. An independent variable can, theoretically, have any number of levels although more than four or five levels is unusual. This is called a manipulated independent variable in that we are actively assigning members to each group.

The dependent variable in this hypothesis is achievement.

Hypothesis:

High school seniors will have a significantly higher number of absences than children in the other grades.

In this hypothesis, the independent variable is the high school year in which the student falls. There are four levels: freshman, sophomore, junior and senior. This is a latent independent variable in that we are not putting children into a specific class, we are measuring them in the class into which they naturally fall.

The dependent variable in this hypothesis is the number of absences.

Type of Hypothesis

- Statistical (null) hypothesis:
  - There will be no significant difference (i.e., different due to reasons other than chance) in SAT scores between children who are home-schooled and those who attend traditional schools.
  - The average SAT of home-schooled students minus the average SAT of traditional school students equals zero (i.e., null).
- Research (alternate) hypothesis:
  - Students who are home-schooled will score higher on the SAT than students who attend traditional schools.
  - Students who are home-schooled will score lower on the SAT than students who attend traditional schools.

Directional Research Hypotheses

- A directional hypothesis contains a "greater than" or "less than" comparison:
  - Students sleeping less than eight hours per night will have lower levels of achievement than students getting eight or more hours of sleep
  - Students receiving graphical report cards will have higher levels of intrinsic motivation than students receiving traditional report cards

Non-directional Research Hypotheses

- A non-directional hypothesis has an inferred "not equal" condition:
  - There will be a significant difference in motivational levels of students receiving weekly report cards and those receiving monthly report cards.

Testing Your Hypotheses

The purpose of your study will be to test your hypothesis by either "rejecting" or "failing to reject" your null hypothesis.

Rejecting your null hypothesis means that the null hypothesis is not true – in other words, there is a significant difference (i.e., not due to chance) between the groups you have measured. This is the same as saying that the research hypothesis is true.

Failing to reject your null hypothesis means there is no significant difference (i.e., not due to chance) between the groups being measured; in other words, the null hypothesis is true. This is the same as saying that the research hypothesis is not true.

Some "hypothetical" closing thoughts…

- The hypothesis must be a logical extension of the review of literature
- Don't state your hypothesis and then look for a problem to match it!
- If you conduct a good study, it's not important whether or not you reject your null hypothesis. What is important is that your results are valid.

Choosing the Right Statistical Test

- State the hypothesis that you are investigating
- Identify the independent variable and it's levels (there may be more than one independent variable)
- Identify and statistically "describe" the dependent variables (there may be more than one dependent variable)
- Select and run the appropriate statistical test
- Interpret the hypothesis in light of the results

Before we describe our dependent variable, we must understand data types

- Discrete data
  - Nominal (categorical) level data
  - Ordinal (rank) level data
- Continuous (quantitative) data
  - Interval level data
  - Ratio level data

Discrete Data

- Nominal (categorical) – frequencies of occurrence are counted.
  - Gender
  - Class or group

- Ordinal (rank) – Numbers represent importance or significance of an observation or experience.
  - Ranking of service department

  - .

Continuous (Quantitative) Data

- Interval data – numbers are assigned on a scale where intervals are constant throughout the scale. There is no absolute zero.
  - Intelligence
  - Achievement
  - Aptitude

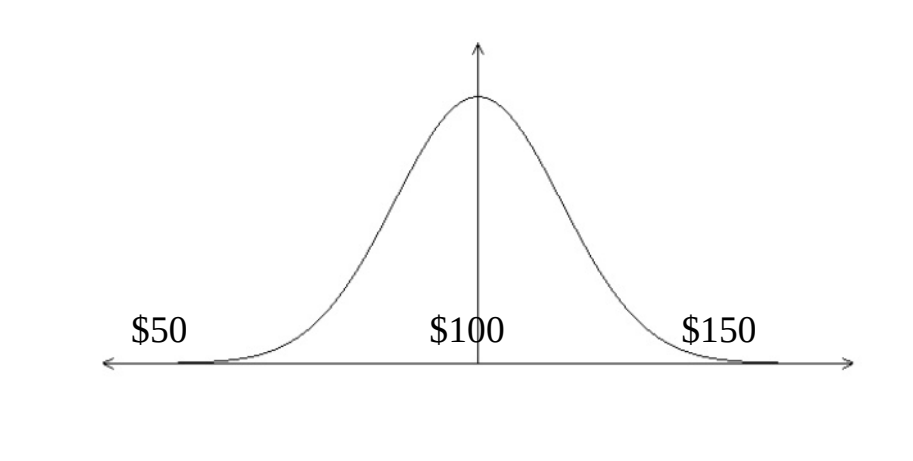
- Ratio data – numbers are assigned so that statements of ratio can be made. Numbers represent units from absolute zero.
  - Weight
  - Elapsed Time
  - Distance

Descriptive Statistics Help Us Decide Which Statistical Test to Use

- Measures of Central Tendency
  - Mean, median, mode
- Measures of Variation
  - Range, standard Deviation, variance
- Measures of Relative Standing
  - z scores, T-scores, percentiles
- Graphical presentations
  - Histograms, pie charts, box-and-whisker plots

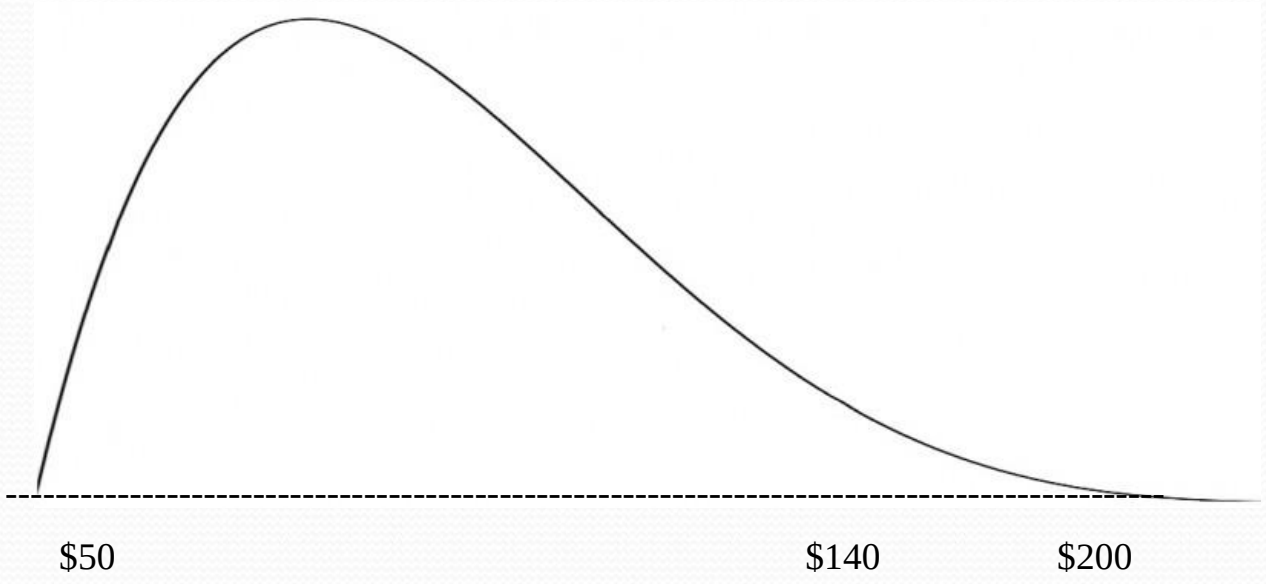
What is a Normal Distribution?

- Data are quantitative.
- Data are approximately evenly distributed around the mean in the shape of a bell.
- Below, the average income is $100; 50% of people make between $50 and $100; 50% make between $100 and $150.



What is a Non-normal Distribution?

- Data are quantitative.
- Data are skewed to the right or left.
- If we added larger values to the prior figure, the mean would be greater and the distribution would be skewed to the right.



Inferential (Decision Making) Statistics

- Discrete (i.e., nominal and ordinal) data use non-parametric statistics.
- Continuous (i.e., interval and ratio) data generally use parametric statistics if the data are normally distributed (i.e., bell-shaped).
- Continuous data may use non-parametric statistics if distribution is not normal (i.e., not bell-shaped).

Common Parametric Statistics

- One sample *t*-test
  - A sample size of less than 30.
  - Data representing a known value from each member of the sample.
  - One fixed value to compare against the sample data.
  - There will be a significant difference in IQ between our students and the national average (i.e., 100).
- One sample *z*-test
  - A sample size of 30 or greater.
  - Data representing a known value from each member of the sample.
  - One fixed value to compare against the sample data.
  - There will be a significant difference in average income between citizens in the southern United States and the national average.

Common Parametric Statistics

- Independent samples *t*-test
  - One independent variable with two levels that are independent of one another (e.g., gender).
  - One dependent variable with continuous data.
  - There will be a significant difference in GPA between males and females.

- Dependent samples *t*-test
  - One independent variable with two levels that are related to one another (e.g., pre-test / post-test scores).
  - One dependent variable with continuous data.
  - There will be a significant difference in test scores before the new teaching strategy is used, and after the new teaching strategy is used.

Common Parametric Statistics

- One-way analysis of variance (ANOVA)
  - One independent variable with three or more levels
  - One dependent variable with continuous data
  - There will be a significant difference in absences between undergraduate, masters and doctoral students.

- N-way analysis of variance (factorial ANOVA)
  - More than one independent variable with two or more levels.
  - One dependent variable with continuous data.
  - There will be a significant difference in absences between male and female students in undergraduate, masters and doctoral programs.

Common Non-Parametric Statistics

- Chi-square
  - Used with nominal data to determine if cell sizes are statistically equivalent.
  - There is no significant difference in the percentage of males and the percentage of females in a gifted program.
- Mann-Whitney U Test
  - Equivalent to *t*-test when data are not normally distributed or data are ordinal in nature.
- Kruskal-Wallis H Test
  - Equivalent to ANOVA when data are not normally distributed or data are ordinal in nature.

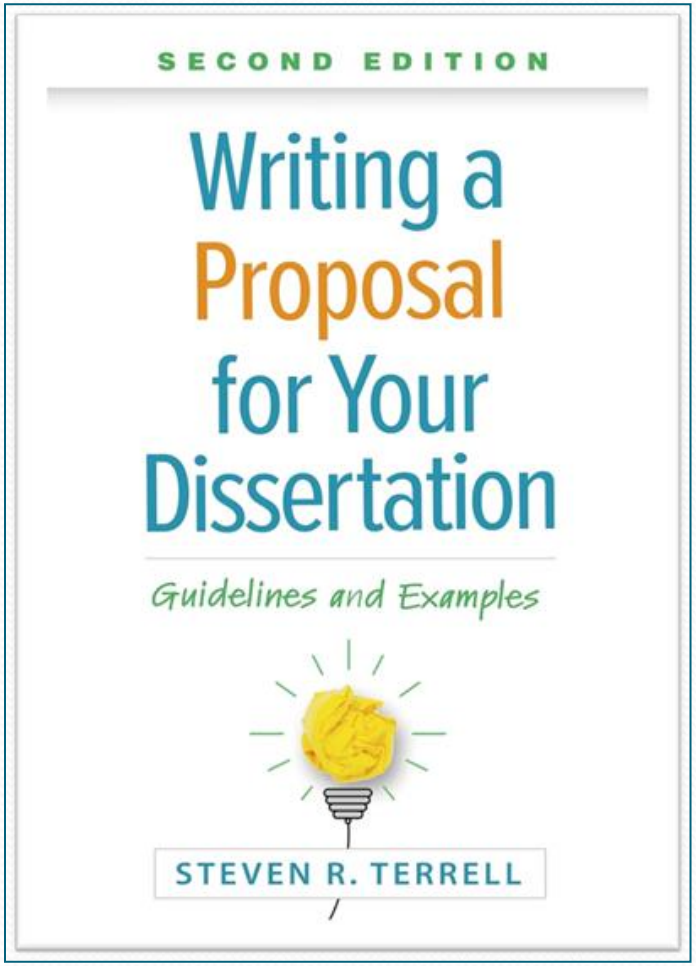
Number of Independent Variables	Number of Levels in the Independent Variable	Number of Dependent Variables	Type of Data the Dependent Variable Represents	Statistical Test to Use	Alternate Statistical Test**
N/A*	N/A*	1	Quantitative	One sample *z*-test or One sample *t*-test	N/A
1	2	1	Quantitative	Dependent sample *t*-test or Independent sample *t*-test	Wilcoxon *t*-test or Mann-Whitney *U* test
1	3 or more	1	Quantitative	Analysis of Variance (ANOVA)	Kruskal-Wallis *H* Test
2	2 or more	1	Quantitative	Factorial ANOVA	N/A
0***	0	0***	Quantitative	Pearson Correlation	N/A
0***	0	0***	Ordinal	Spearman Correlation	N/A
1 or more	2 or more	1 or more	Nominal	Chi-square	N/A

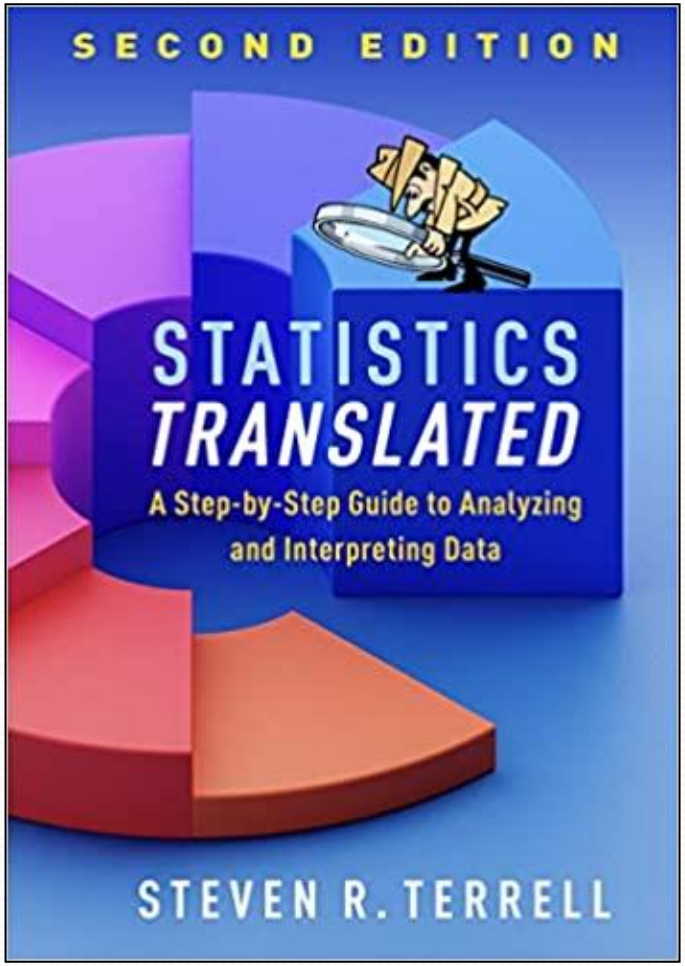
* Not all parametric tests have a non-parametric equivalent or the test mentioned is already a non-parametric test.

** When you're working with one-sample *z* tests or one-sample *t*-tests, it is really not logical to point out the independent variable or its levels.

*** In correlational procedures, you simply want to look at the relationship between two variables; these are call predictor and criterion variables.

References





Click on Picture for Book Details.