

RESEARCH METHODS

For Business and Marketing

GEORGE SELF

January, 2019 – Edition 1

George Self: *Research Methods*, For Business and Marketing, January, 2019

This work is licensed under a **Creative Commons** "Attribution-NonCommercial-ShareAlike 4.0 International" license.



FOREWORD

I have taught BASV 316, *Introductory Methods of Analysis*, on-line for the University of Arizona in Sierra Vista since 2010 and enjoy working with students on research methodology. I wanted a textbook that presented research in a practical way so students could use the lessons learned in their own research projects. I found an excellent book but over the years the cost of that book increased to the point that I felt like it was an unfair burden on students.

I began by looking for an acceptable “open source” book since authors make those available to students free of charge and I could modify the book to meet my own objectives. I could not find any that were focused on business research though I tried for several years—and keep looking to this day. I did, though, find a few open source books about research in the social and psychological sciences that were reasonably close to what I needed. So, I modified those books to emphasize business research and then provided my work to students free of charge.

Bhattacharjee[1], Blackstone[2], and Price[7] all released books about research that formed the major sources for this book. Those books are all open source and published under a Creative Commons license that permitted me to copy and modify them.

Three goals shaped the choices made about the topics covered by the text and how those topics are presented.

- The topics must have relevance for business students.
- Both qualitative and quantitative research methods are given roughly equal attention since both types of research are used in business.
- The text is engaging and readable.

While the book is useful in its current form, I will continually update it based on emerging trends in research.

This book is published under a Creative Commons **Attribution-NonCommercial-ShareAlike** license, just like the books that provided its foundation. The source is available at my GitHub account: <http://bit.ly/2xIjzXL>. It is my hope that students can use this book to learn about business research and other instructors can modify and use it for their own classes.

— George Self

BRIEF CONTENTS

I	BACKGROUND	1
II	QUANTITATIVE METHODS	3
1	DATA	5
III	QUALITATIVE METHODS	27
IV	MIXED METHODS	29
V	REPORTING	31
VI	APPENDIX	33
2	APPENDIX	35
	GLOSSARY	37
	BIBLIOGRAPHY	45

CONTENTS

I	BACKGROUND	1
II	QUANTITATIVE METHODS	3
1	DATA	5
1.0.1	Types of Data	5
1.1	Rating Scale	7
1.2	Properties of Data	10
1.2.1	About The Normal Distribution (Bell Curve)	10
1.2.2	Excess Kurtosis	11
1.2.3	Skew	12
1.3	Databases	13
1.3.1	Public Databases	14
1.3.2	Using Public Databases	15
1.4	Statistical Test Selection	15
1.4.1	Summary Statistics	15
1.4.2	Parametric vs. Nonparametric Tests	16
1.4.3	Hypothesis Test Selection	17
1.5	Statistical Test Sampler	18
1.5.1	Central Measures	18
1.5.2	Spread	19
1.5.3	Frequency Tables	20
1.5.4	Correlation	20
1.5.5	Parametric Tests	20
1.5.6	Nonparametric Hypothesis Tests	21
1.6	Data Mining	22
1.7	Summary	26
III	QUALITATIVE METHODS	27
IV	MIXED METHODS	29
V	REPORTING	31
VI	APPENDIX	33
2	APPENDIX	35
	GLOSSARY	37
	BIBLIOGRAPHY	45

LIST OF FIGURES

Figure 1	Normal Distribution	10
Figure 2	Kurtosis in a Normal Distribution	12
Figure 3	Skew in a Normal Distribution	13
Figure 4	Classifying Variables	15
Figure 5	Central Measure and Spread	16
Figure 6	Parametric vs. Nonparametric Selection	17
Figure 7	Example of Clustering	23
Figure 8	Decision Tree	24
Figure 9	Market Basket Graph	25
Figure 10	Market Basket Rules	25
Figure 11	Data types	26

LIST OF TABLES

Table 1	Political activism binary scale	8
Table 2	Likert scale for employee self-esteem	8
Table 3	Semantic differential scale	9
Table 4	Guttman scale	10
Table 5	Tests to Compare Two or More Samples	18
Table 6	Tests to Association Between Two Samples	18
Table 7	2016 Exit Poll.	20
Table 8	Selected Automobile Correlations.	20

LISTINGS

Part I

BACKGROUND

Research methods are grounded in philosophy, statistics, sociology, and many other disciplines. The chapters in this section introduce these background concepts.

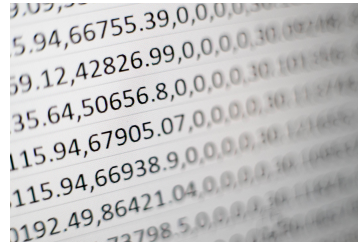
Part II

QUANTITATIVE METHODS

Quantitative methods are based in the measurement of concepts and the statistical analysis of those measures. Quantitative methods include activities like sampling, surveys, and experimental research.

DATA

Data are¹ a collection of facts about some topic. As examples, a “customer loyalty” program gathers data from customers on how often they shop, what they purchase on each trip, what time of day they typically shop, and all sorts of other data. When data are interpreted in some way they become information. The types of analyses that can be done with data are limited by the types of data involved. The purpose of this chapter is to introduce various concepts about data and show how they can be analyzed.²



1.0.1 Types of Data

Psychologist Stanley Smith Stevens defined four generic types of data[9].

- **Qualitative data** groups observations into a limited number of categories; for example, type of pet (cat, dog, bird, etc.) or place of residence (Arizona, California, etc.). Because qualitative data do not have characteristics like means or standard deviations, they are analyzed using non-parametric tests, like Kruskal-Wallis H and Mann-Whitney U. Qualitative data can be further divided into two sub-types, nominal and ordinal.
 - **Nominal data** are categories that do not overlap and have no meaningful order, they are merely labels for attributes. Examples of nominal data include occupations (custodial, accounting, sales, etc.) and blood type (A, B, AB, O). A special subcategory of nominal data is binary, or dichotomous, where there are only two possible responses, like “yes” and “no”. Nominal data are sometimes stored in a database using numbers but they cannot be treated like numeric data. For example, binary data, like “Do you rent or own your home?” can be stored as “1 = rent, 2 = own” but the numbers in this case have no numeric significance and could be replaced by words like “Rent” and “Own.”

¹ The word “data” is plural and is treated that way in this text even though modern usage seems to be trending toward a singular form.

² Photo by Mika Baumeister on Unsplash

- **Ordinal data** are categorical data but, unlike nominal, the categories imply some sort of order (which is why it is called “ordinal” data). One example of ordinal data is the “star” rating system for movies. It is clear that a five-star movie is somehow better than a four-star movie but there is no way to quantify the difference between those two categories. As another example, it is common for hospital staff members to ask patients to rate their pain level on a scale of one to ten. If a patient reports a pain level of “seven” but after some sort of treatment later reports a pain level of “five” then the pain has clearly decreased but it would be impossible to somehow quantify the exact difference in those two levels. Ordinal scales are most commonly used with Likert-type survey questions where the responses are selections like “Strongly Agree”, “Agree”, “Neutral”, “Disagree”, “Strongly Disagree”. Ordinal data are also used when numeric data are grouped. For example, if a dataset included respondents’ ages then those numbers could be grouped into categories like “20 – 29” and “30 – 39.” Those groups would typically be stored in the dataset as a single number so maybe “2” would represent the ages “20 – 29,” which would be ordinal data.
- **Quantitative data** are numbers, typically counts or measures, like a person’s age, a tree’s height, or a truck’s weight. Quantitative data are measured with scales that have equal divisions so the difference between any two values can be calculated. Quantitative data are discrete if they are represented by integers, like the count of words in a document, or continuous if they are represented by fractional numbers, like a person’s height. Because quantitative data includes characteristics like means and standard deviations, they are analyzed using parametric tests, like T-tests and Analysis of Variance (ANOVA). Quantitative data can be further divided into two sub-types, interval and ratio.
 - **Interval data** use numbers to represent quantities where the distance between any two quantities can be calculated but there is no true zero point on the scale. One example is a temperature scale where the difference between 80° and 90° is the same as the difference between 60° and 70°. It is important to note that interval data do not include any sort of true zero point, thus zero degrees Celsius does not mean “no temperature,” and without a zero point it is not reasonable to make a statement like 20° is twice as hot as 10°.
 - **Ratio data** use numbers to describe a specific measurable distance between two quantities; however, unlike interval

data, ratio data have a true zero point. A good example of ratio data is the sales report for an automobile dealership. Because the data are a simple count of the number of automobiles sold it is possible to compare one month with another. Also, since the scale has a true zero point (it is possible to have zero sales) it is possible to state that one month had twice the sales of another.

1.1 RATING SCALE

When working with qualitative data, it is important for researchers to determine a rating scale, also called levels of measure, to record data gathered about any one attribute. For example, male-female-other, M-F-O, and 1-2-3 are three potential rating scales for the attribute *gender*. A researcher could use any of these scales, or devise a completely different one, as long as the scale is used consistently throughout the entire research project. It is easy to imagine that many rating scales exist but the most common ones are *binary*, *Likert*, *semantic differential*, and *Guttman*.

BINARY Binary scales are nominal scales consisting of binary items that assume one of only two possible values, such as yes or no, true or false, and so on. For example, a typical binary scale for a “political activism” construct may consist of the six binary items shown in Table 7. Each item in this scale is a binary item, and the total number of “yes” indicated by a respondent (a value from 0 to 6) can be used as an overall measure of that person’s political activism. Binary scales can also employ other values, such as male or female for gender, full-time or part-time for employment status, and so forth. If an employment status item is modified to allow for more than two possible values (e.g., unemployed, full-time, part-time, and retired), it is no longer binary, but still remains a nominal scaled item.

Question	Yes	No
Have you ever written a letter to a public official?	<input type="radio"/>	<input type="radio"/>
Have you ever signed a political petition?	<input type="radio"/>	<input type="radio"/>
Have you ever donated money to a political cause?	<input type="radio"/>	<input type="radio"/>
Have you ever donated money to a candidate running for public office?	<input type="radio"/>	<input type="radio"/>
Have you ever written a political letter to the editor of a newspaper?	<input type="radio"/>	<input type="radio"/>
Have you ever persuaded someone to change his/her voting plans?	<input type="radio"/>	<input type="radio"/>

Table 1: Political activism binary scale

LIKERT Designed by Rensis Likert, this is a very popular rating scale for measuring ordinal data in business research. This scale includes Likert items that are simply-worded statements to which respondents can indicate their extent of agreement or disagreement on a five or seven-point scale ranging from “strongly disagree” to “strongly agree.” A typical example of a six-item Likert scale for the “employment self-esteem” construct is shown in table 2. Likert scales are summated scales, that is, the overall scale score may be a summation of the attribute values of each item as selected by a respondent.

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I feel good about my job	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I get along well with others at work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I’m proud of my relationship with my supervisor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel like I’m making a contribution at work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can tell that my coworkers respect me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table 2: Likert scale for employee self-esteem

Likert items allow for more granularity (more finely tuned response) than binary items, including whether respondents are neutral to the statement. Three or nine values (often called “anchors”) may also be used, but it is important to use an odd

number of values to allow for a “neutral” (or “neither agree nor disagree”) anchor. Some studies have used a “forced choice approach” to force respondents to agree or disagree with the Likert statement by dropping the neutral mid-point and using even number of values, but this is not a good strategy because some people may indeed be neutral to a given statement and the forced choice approach does not provide them the opportunity to record their neutral stance. A key characteristic of a Likert scale is that even though the statements vary in different items or indicators, the anchors (“strongly disagree” to “strongly agree”) remain the same. Likert scales are ordinal scales because the anchors are not necessarily equidistant, even though sometimes we treat them like interval scales.

SEMANTIC DIFFERENTIAL This is a composite (multi-item) scale where respondents are asked to indicate their opinions or feelings toward a single statement using different pairs of adjectives framed as polar opposites. For instance, the construct “attitude toward health insurance” can be measured using three items shown in Table 3. As in the Likert scale, the overall scale score may be a summation of individual item scores. Notice that in Likert scales, the statement changes but the anchors remain the same across items. However, in semantic differential scales, the statement remains constant, while the anchors (adjective pairs) change across items. Semantic differential is believed to be an excellent technique for measuring people’s attitude or feelings toward objects, events, or behaviors.

How would you rate your opinion on health insurance?						
	Very Much	Much	Neutral	Much	Very Much	
Good	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Bad
Useful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Useless
Caring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Uncaring

Table 3: Semantic differential scale

GUTTMAN Designed by Louis Guttman, this composite scale uses a series of items arranged in increasing order of intensity of the construct of interest, from least intense to most intense. As an example, the construct “attitude toward immigrants” can be measured using five items shown in Table 4. Each item in the Guttman scale has a weight (not indicated above) which varies with the intensity of that item, and the weighted combination of each response is used as aggregate measure of an observation.

How will you rate your opinion on the following statements about immigrants?		
Do you mind immigrants being citizens of your country?	Yes	No
Do you mind immigrants living in your own neighborhood?	Yes	No
Would you mind living next door to an immigrant?	Yes	No
Would you mind having an immigrant as your close friend?	Yes	No
Would you mind if someone in your family married an immigrant?	Yes	No

Table 4: Guttman scale

1.2 PROPERTIES OF DATA

1.2.1 About The Normal Distribution (Bell Curve)

When the quantitative data gathered from some statistical project are plotted on a graph they often form a [normal distribution](#) (sometimes called a “bell curve” due to its shape). As an example, consider the Scholastic Aptitude Test (SAT) which is administered to more than 1.5 million high school students every year. Figure 1 was created with fake data but illustrates the results expected of a typical SAT administration.

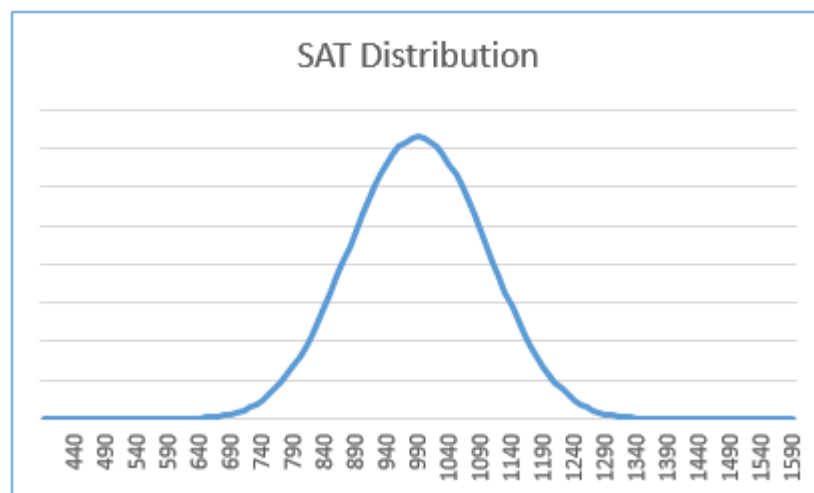


Figure 1: Normal Distribution

SAT scores lie between 400 and 1600 as listed across the X-Axis and the number of students who earn each score is plotted. Since

the most common score is 1000 that score is at the peak of the curve. Very few students scored above 1300 or below 650 and the curve is near the lower bound beyond those points. This illustrates a normal distribution where most scores are bunched near the center of the graph with only a few at either extreme.

The normal distribution is important because it permits researchers to use specific techniques to test a hypothesis about the sample. For example, perhaps a researcher hypothesized that the graduation rate at university “A” was higher than at university “B” because students’ SAT scores were higher. Since SAT scores have a normal distribution, the researcher could use specific tests, like a t-test, to support or refute the hypothesis. However, if the data were not normally distributed then the researcher would need to use a different group of tests.

1.2.2 *Excess Kurtosis*

One way to mathematically describe a normal distribution is to calculate the length of the tails of a bell curve, and that is called its [excess kurtosis](#). For a normal distribution the excess kurtosis is 0.00, a positive excess kurtosis would indicate longer tails while a negative excess kurtosis would indicate shorter tails. Intuitively, many people believe the excess kurtosis represents the “peaked-ness” of the curve since longer tails would tend to lead to a more peaked graph; however, excess kurtosis is a measure of the data outliers, which would be only present in the tails of the graph; so excess kurtosis is not directly indicative of the “sharpness” of the peak. It is difficult to categorically state that some level of excess kurtosis is good or bad. In some cases, data that form a graph with longer tails are desired but in other cases they would be a problem.

Following are three examples of excess kurtosis. Notice that as the excess kurtosis increases the tails become longer.

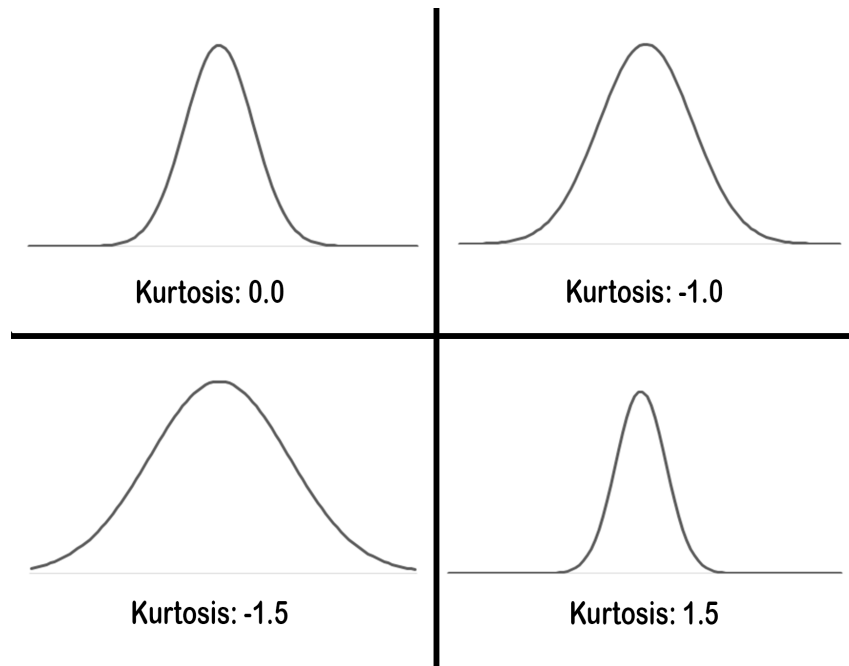


Figure 2: Kurtosis in a Normal Distribution

1.2.3 Skew

The second numerical measure of a normal distribution that is frequently reported is its [skew](#), which is a measure of the symmetry of the curve about the mean of the data. The normal distribution in [Figure 3](#) has a skew of 0.00. A positive skew indicates that the tail on the right side is longer, which means that there are several data points on the far right side of the graph “pulling” the tail out that direction. A negative skew indicates that the tail on the left side of the graph is longer. Following are three examples of skew:

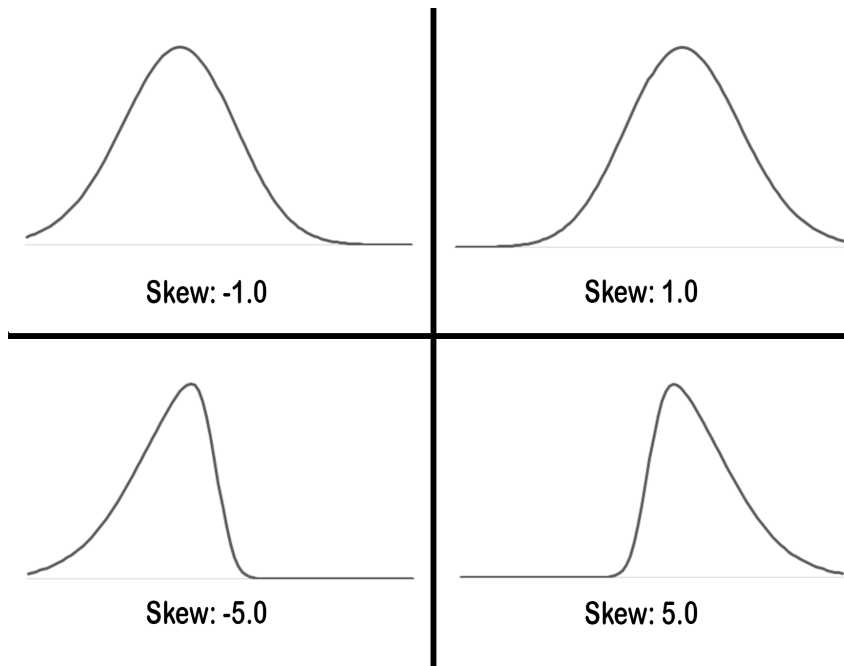


Figure 3: Skew in a Normal Distribution

1.3 DATABASES

When a lot of data are gathered into a single location they are referred to as a [database](#). This is not a new concept. Fifty years ago a library would have a series of 3X5 cards that contained information about all of the books in the library (title, author, subject, etc.). Those cards were stored in a wooden cabinet called the “card catalog” and customers could find information about and the location of whatever book they wanted. Today, databases are often contained in electronic form on the internet where they can be accessed from customers’ home computers, tablets, or even phones.

Data in a database is typically stored in tables that resemble spreadsheets, that is, rows and columns where each row is one record (or observation) about some phenomenon and each column is one descriptor of that record. For example, a database that contains information about the people who work at a company would be organized such that each row contained data about just one person and each column would contain a single aspect of that person’s employment, like name, employee number, date of birth, etc.³ A database is designed to deliver answers to questions through a lookup process so, for example, if the CEO of a company wanted to know the birth date for someone in the accounting department that data could be easily found.

³ Of course, databases are much more complex than described in this paragraph but this is not a database text so the simple explanation offered is adequate for this context.

One common problem with any database is “dirty data.” These are data that contain errors or are missing. For example, it is easy for a data entry clerk to enter something like “1000000” instead of “100000” (count the zeros) for a person’s salary and create an “outlier” in the data. Another common problem are missing data. For example, if employees are asked to update their personal information but someone could not remember their ZIP code then they would simply leave that field blank.

Dirty data makes it difficult to analyze the database. For example, if a researcher wanted to report the median salary for the workers in a factory but ten percent of the salaries were missing from the database then the median would not be accurate. There are several methods statisticians use to mitigate the problems caused by dirty data but those are beyond the scope of this text.

1.3.1 *Public Databases*

There are hundreds of publicly available databases that can be used for research. As one example, the United States Census Bureau maintains a huge database that contains information about the people of the United States.⁴ The data at that site is freely available to anyone who wants to use it, and the site is organized so information is fairly quick and easy to find. As an example, it is not difficult to discover that among adults in the United States, 29% have a high school diploma, 20% have a Bachelor’s degree, 8% have a Master’s degree, 1.7% have a Doctoral degree, and the rest fall elsewhere on the education spectrum. The US Census Bureau has more advanced tools available that permit researchers to focus on a single county or even smaller region.

When using a public database, a researcher must be concerned with bias. Since the researcher did not gather the data first-hand it is possible that the data are biased. For example, if the database includes people’s attitude towards work how is the researcher going to know if the data gathered was from a well-designed, neutral survey or if it was just gathered using some sort of convenience sample? In general, databases found at governmental websites (with urls that end with .gov) or education websites (with urls that end with .edu) would more likely be bias-free while databases from .com sites would need to have extra attention paid to bias.

Sometimes, students will find a website with a list of links to journal articles or chapters from books. While these are valuable resources for a researcher, they are not the same as a database that contains raw data from a survey, experiment, or other activity. Journal articles provide good information for a literature review but would not be appropriate for an online database source.

⁴ The US Census Bureau’s website is at <https://www.census.gov/>.

1.3.2 Using Public Databases

As an example of using a public database, imagine that the CEO of “BASVFOODS” is interested in opening a neighborhood market in a small town they have never serviced before. In order to gather information about that location it is possible to use US Census Bureau data to find out things like the median household size, income, and education level. The CEO could then compare that data with similar data from a town that has a successful store to help inform a decision to open a new store.

1.4 STATISTICAL TEST SELECTION

Once the data are gathered it is important to run appropriate statistical processes to see if the data contain anything of interest. There are hundreds of tests that can be used and researchers must consider both the goal of the analysis and the type of data being analyzed. The first step is to classify the variables being evaluated.

- **Qualitative data** are numeric and are generated through measurement or counting. **continuous data** can be any value, including fractions or decimals, like a person’s height or the length of time it takes to complete some task. **Discrete data** are integers normally found by counting, like the number of people at an event or the age of a respondent.
- **Qualitative data** are non-numeric and are often generated by people checking boxes on a survey. **Ordinal data** have some implied order, like a student’s class (senior, junior, etc.) and **nominal data** have no order, like a respondent’s gender.

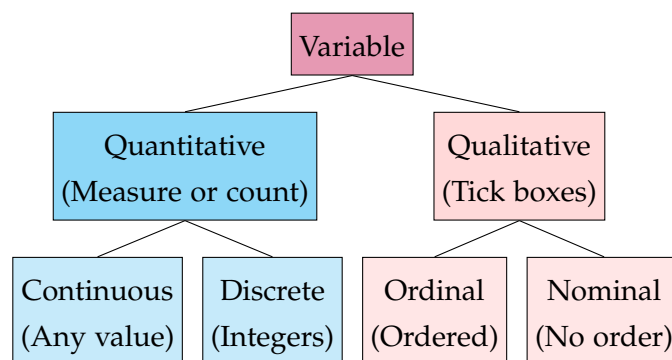


Figure 4: Classifying Variables

1.4.1 Summary Statistics

One of the easiest type of analysis to complete is determining the center and spread of a data set. This analysis is also one of the first

that research readers expect to find. Here is a guide for what sort of central measure and spread to report.

- For normally distributed continuous data use the mean and standard deviation.
- For skewed continuous or discrete data use the median and Interquartile Range (IQR).
- For ordinal data use the median and Interquartile Range (IQR).
- For nominal data use the mode as a central measure but there is no spread.

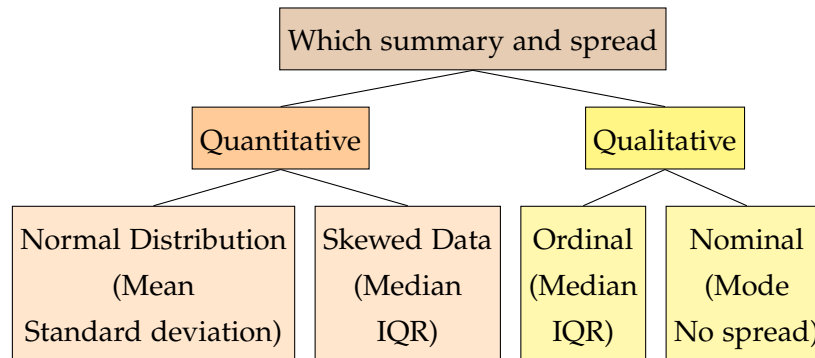


Figure 5: Central Measure and Spread

1.4.2 Parametric vs. Nonparametric Tests

Normally, a researcher posits a hypothesis and then conducts research to either support or refute that hypothesis. The type of hypothesis test needed depends on the type of the variables being tested. There are two broad categories of variables in a hypothesis test: independent and dependent.

- **Independent variables** (explanatory variables) are those that cause something to happen; they explain why some outcome was observed. For example, if the hypothesis is that women spend more on groceries than men then the independent variable is the sex of the shopper. If the hypothesis is that old people are more dangerous drivers than young people then the independent variable is the age of the driver.
- **Dependent variables** (outcome variables) are those that are the outcome for whatever is happening. For example, if the hypothesis is that women spend more on groceries than men then the dependent variable is amount of money spent. If the hypothesis is that elderly drivers are more dangerous than younger drivers then the dependent variable is the number of reported accidents.

The chart in Figure 6 is used to determine if the hypothesis test should be parametric or nonparametric.

- **Parametric** tests assume that the data follow a particular distribution, usually normally distributed. Parametric tests are more powerful than nonparametric tests and are more likely to detect true differences or relationships that exist.
- **Nonparametric** tests are used when the dependent variable is not normally distributed; that is, the data are skewed or are qualitative. Nonparametric techniques are usually based on ranks or signs rather than the actual data and are usually less powerful than parametric tests.
- Chi-square is a test that is only used when the dependent variable is nominal in nature. A chi-square test determines if there is a significant difference in the actual observed frequencies and those hypothetically expected. For example, if a coin were tossed 100 times it would be expected to land “heads” half of the time but if the actual observation was that “heads” came up 75% of the time then the chi-square statistic would indicate that this was a significant difference.

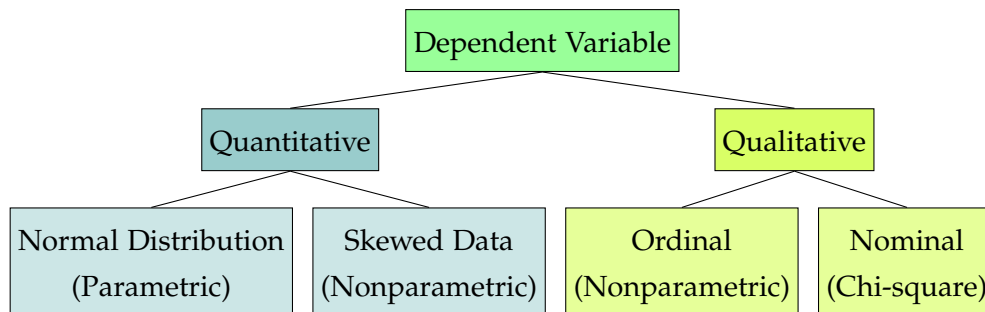


Figure 6: Parametric vs. Nonparametric Selection

1.4.3 Hypothesis Test Selection

Researchers often hypothesize that some treatment will lead to an outcome. In order to test that, they will apply the treatment to one group but not another and then compare the two groups to see if there is any difference. Table 5 lists the statistical tests that are commonly used to compare the averages of two or more groups. In the table, “independent” groups are those that have no overlapping members while “matched” groups test the same members two times. As an example, if researchers wanted to know how if a movie affected people in some way they could survey people leaving two different theaters (independent groups) or survey the same people before and after seeing the movie (matched groups).

Comparing	Dep Var	Ind Var	Para	NonPara
2 independent groups	Quant	Binary	Indep t-test	Mann-Whitney
3+ independent groups	Quant	Nom	ANOVA	Kruskal-Wallis
2 matched groups	Quant	Time	Paired t-test	Wilcoxon
3+ measures same subj	Quant	Time	Rep ANOVA	Friedman

Table 5: Tests to Compare Two or More Samples

Another common research goal is to see if there is any association (or “correlation”) between two or more variables. Further, a correlation may be able to predict an outcome for a new observation. For example, a researcher may run an experiment where students work with a tutor once a week and then see if there is an improvement in test scores. In this case, “tutoring time” would be correlated with “test scores” and that may be used to predict a new student’s test score based on the amount of tutoring time. Table 6 lists the hypothesis tests that are commonly used to find correlations and predictions between two groups.

Comparing	Dep Var	Ind Var	Para	NonPara
2 Continuous	Quant	Quant	Pearson’s r	Spearman’s rho
Prediction	Quant	Any	Regression	None
Prediction	Nominal	Any	Log regression	None
2 Qual vars	Qual	Qual	None	Chi-square

Table 6: Tests to Association Between Two Samples

1.5 STATISTICAL TEST SAMPLER

While there are hundreds of statistical procedures available, this section covers those that are commonly used.

1.5.1 Central Measures

Three different central measures are commonly used, depending on the type of data being summarized.

- **Mean.** An arithmetic mean is calculated by adding together all of the terms and then dividing by the number of terms. This process is taught in elementary school as calculating the “average.” However, if the terms have wildly different values then a *geometric mean* is a better choice. In a geometric mean the values are multiplied together and then the n -root is taken, where “ n ” is the number of terms in the data. Finally, if the mean of a series of rates is needed, then a *harmonic mean* is used. For a harmonic mean all of the terms are reciprocated, a mean is found of those reciprocated terms, and then the mean is reciprocated.
- **Median.** A median is found by putting all of the terms in numeric order and then selecting the middle term. This is useful if the data set includes outliers, that is, a few values far outside the other terms. Medians are frequently used to report house values since a few houses may be worth far more than the other houses in an area. If the data set has an even number of terms, so there is no middle term, then the median is found by taking the mean of the middle two terms.
- **Mode.** The mode is used for nominal data and is the count of the most common term. For example, if a count of the number of undergraduate students in each class (senior, junior, etc.) is made and it turns out that there are more seniors than any other group then the mode would be “senior.”

1.5.2 Spread

Two measures of spread are commonly used.

- **Range.** The range, or “dispersion,” is the difference between the highest value and lowest value.
- **Standard deviation.** This often misunderstood value is nothing more than an indicator of how much variation exists in the data, or how “scattered out” the data are. In general, the greater the standard deviation the more variation there are in the data. As an example, imagine that a professor administered an examination to a group of students and found the mean to be 70% with a standard deviation of 15. Then the professor changed something about the class the next semester and administered the same examination to that group of students and found the mean was 85% with a standard deviation of 5. This result would indicate that the scores the second semester had much less variation, they were much “tighter,” and that would likely be good news for the professor.

1.5.3 Frequency Tables

Discrete or qualitative data items are normally reported in frequency tables where the counts for a particular item are displayed. When a frequency table has two dimensions it is usually called a “cross-tab” or “pivot table.” As an example, Table 7 contains the results of an exit poll from 2016 presidential election. [6]

Party	Clinton	Trump	Other
Democrats	89%	8%	3%
Republicans	8%	88%	4%
Independents	42%	46%	12%

Table 7: 2016 Exit Poll.

1.5.4 Correlation

Correlation is a method used to describe a relationship between the independent (or x-axis) and dependent (or y-axis) variables in a research project. A correlation is expressed as a number between -1.0 and $+1.0$ where the closer the correlation gets to the extremes the more meaningful it becomes. Thus, two variables with a correlation of $+0.65$ have a closer association than two variables with a correlation of $+0.23$. Correlations are often displayed in a matrix. Table 8 shows the Miles per Gallon, Displacement, and Horsepower for a group of automobiles. In this table, notice that there is a negative correlation between horsepower and miles per gallon, which means that cars with greater horsepower get fewer miles per gallon. On the other hand, there is a positive correlation between displacement and horsepower so cars with greater engine displacement have more horsepower.

	mpg	disp	hp
mpg	+1.00	-0.85	-0.78
disp	-0.85	+1.00	+0.79
hp	-0.78	+0.79	+1.00

Table 8: Selected Automobile Correlations.

1.5.5 Parametric Tests

Some analysis techniques are only useful with [parametric](#) data. While there are dozens of statistical processes that will work with parametric data, two are most commonly seen.

- **T-test.** A t-test is used to analyze the difference in two groups of samples that are normally distributed. For example, a researcher may hypothesize that there is a significant difference in the ages of people in two towns. Once people's ages are recorded a t-test could be used to see if there is a significant difference in the mean age for the people in those two towns.

There are two varieties of t-test commonly used, depending on the type of data being analyzed. If the two groups being compared are independent of each other then an "independent t-test" would be used. However, researchers often use the same group but test them at two different times. For example, a medical trial may measure some factor (like blood pressure), apply some treatment, and then measure the factor a second time. In that case, a "paired t-test" would be used where the values from the first trial would be compared to those in the second.

- **Analysis of Variance (ANOVA).** An ANOVA is similar to a t-test but is used to analyze three or more groups to see if there is a significant difference in any of the groups.

The result of either a t-test or ANOVA is a *p-value* ("probability value"). A p-value describes the probability that some result was caused by pure chance and not from some applied treatment. P-values are expressed as percentages and, normally, researchers expect a p-value under 0.05 (5%) in order to declare the result to be significant. If the calculated p-value is above 0.05 then the researcher declares that no significant result was found.

1.5.6 Nonparametric Hypothesis Tests

Some analysis techniques are only useful with [nonparametric](#) data. While there are dozens of statistical processes that will work with nonparametric data, two are most commonly seen.

- **Mann-Whitney U.** This test is used to determine if there are any significant differences in two groups of data that are not normally-distributed, often categorical. As an example of using a Mann-Whitney U test, Gluck et al. used it in *How Short Is Too Short? Implications of Length and Framing on the Effectiveness of Privacy Notices*[3]. This study postulated that privacy notices were so long people did not often read, or understand, them. They conducted an experiment where they presented groups of people with both short and long privacy notices and then assessed their understanding. A Mann-Whitney U test was used to compare the groups and, not surprisingly, people who used the short form seemed to understand their privacy rights better

than those who used a long form. This experiment also varied the framing of the policy (positive vs. negative) and used a Kruskal-Wallis H test to compare the results but they found no effect provided by different framing. This study is a good example of using both Kruskal-Wallis H and Mann-Whitney U in a single analysis.

- **Kruskal-Wallis H.** This test is used to determine if there are any significant differences in three or more groups of data that are not normally-distributed, often categorical. As an example of using a Kruskal-Wallis H test, Titlebaum and Lawrence published *Perceived Motivations for Corporate Suite Ownership in the “Big Four” Leagues*[10]. This study looked at 29 different reasons (entertain clients, support the community, and business-to-business networking, etc.) corporations purchase luxury suites in any of the four major sports leagues (NFL, MLB, NBA, and NHL). They surveyed corporate leadership and then applied a Kruskal-Wallis H test to the survey results to see which reason was most important. They found four that were significant: 1) Entertaining Employees, 2) Supporting the Community, 3) Perception of the Company in the Community, and 4) Customized Gifts for Suite Owners. They also compared the leagues, two at a time, to see if the reasons were different between the leagues and used a Mann-Whitney U test for this part of the analysis. They found that MLB had a significantly higher rank for “Entertaining Employees” than all of the other sports. This study is a good example of using both Kruskal-Wallis H and Mann-Whitney U in a single analysis.

The result of either a Mann-Whitney U or Kruskal-Wallis H is a *p-value* (“probability value”). A *p-value* describes the probability that some result was caused by pure chance and not from some applied treatment. *P-values* are expressed as percentages and, normally, researchers expect a *p-value* under 0.05 (5%) in order to declare the result to be significant. If the calculated *p-value* is above 0.05 then the researcher declares that no significant result was found.

1.6 DATA MINING

Data mining is a relatively new statistical technique that attempts to extract (“mine”) valuable intelligence from massive databases. Raval[8] published an excellent overview of data mining techniques. As but one example, a grocery store chain mines customer purchases, shopping habits, and other data to plan sales, send coupons, and make other suggestions that will, hopefully, lead to increased spending. While there are more than a dozen data mining techniques, these three are most common.

1.6.0.1 Clustering

It is beneficial to “cluster” customers in some way so advertising (and sales) can be more effective. For example, if a large grocery store chain can determine that most of the customers in one region share some common trait then it becomes easier to market to that region.

Figure 7 shows a scatter plot with three clusters. This plot was generated from dummy data but it shows the type of clustering that a researcher may be able to find in a data set. The location of a specific cluster in relationship to the others would perhaps drive a marketing campaign focused on the specific properties the customers in that one cluster share.

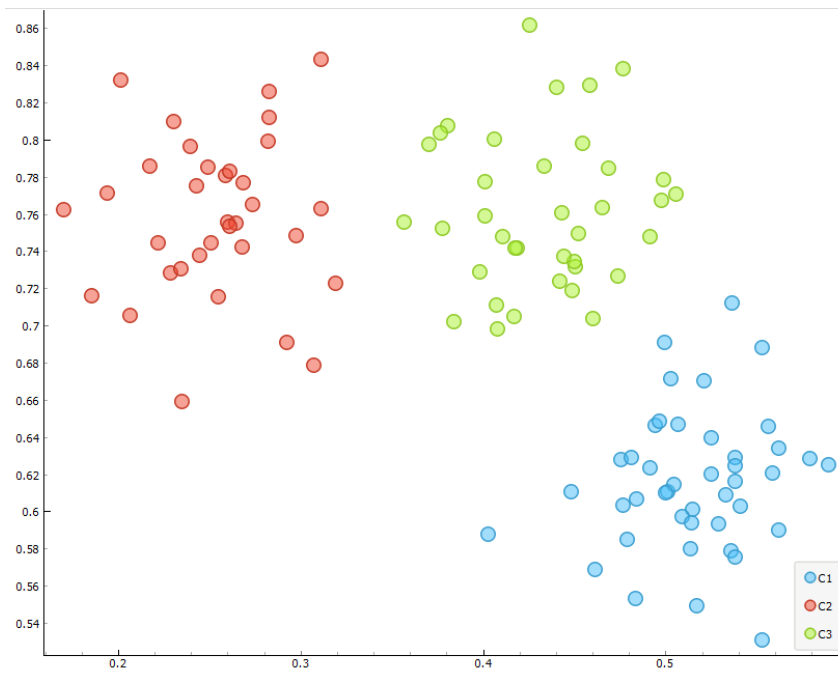


Figure 7: Example of Clustering

As an example of using clustering, Vogel, Greiser, and Mattfeld published *Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns*[11] in 2011. They analyzed more than 760,000 pickups and returns at bike stations in Vienna from 2008 – 2009. Their goal was to determine if the stations could be clustered in some way by usage patterns. They used cluster analysis to divide the city’s stations into five groups, for example, the “returns morning pickups evening” group showed especially high number of morning returns and evening pickups. They provided the city with information that was designed to help in future station placement.

1.6.0.2 Decision Tree

A decision tree organizes known data so predictions can be made on new data. Figure 8 is taken from a data set that shows the fate

of all the passengers on the Titanic. Starting at the top, few males survived. Of the females, there were no survivors who had more than 4.5 siblings or spouses on board, so large families perished. Notice, though, that near the bottom of the tree nearly all first class female passengers with few family members on board survived. A decision tree like this could be used to predict whether a given passenger survived.

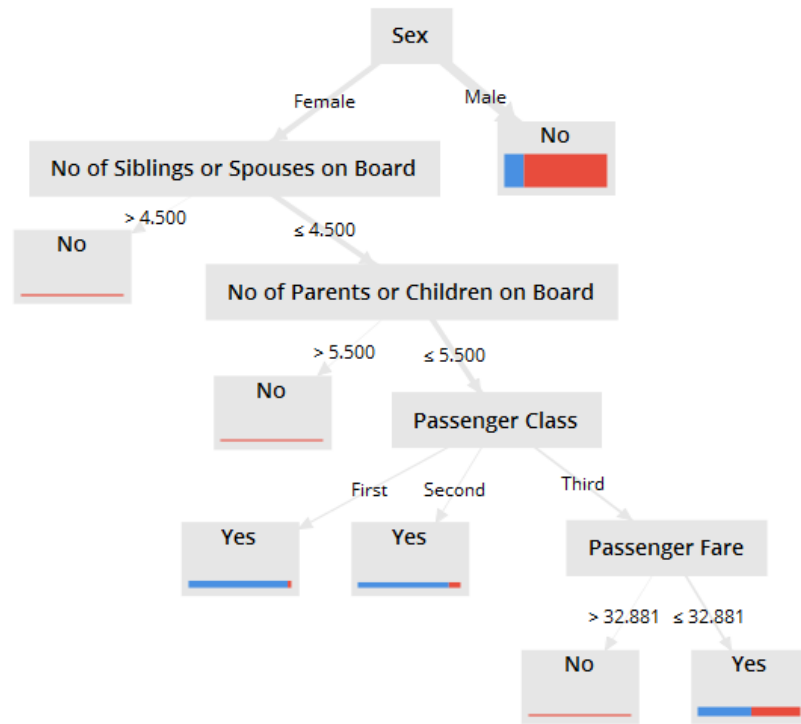


Figure 8: Decision Tree

As an example of using a decision tree, Ona, Ona and Lopez completed a study of the Grenada, Spain, public transportation system[5]. They data mined 3,664 passenger surveys conducted from 2008 until 2011. The surveys asked passengers some demographic questions (age, etc.) and about various aspects of the bus service. The researchers first created four clusters of passenger types using cluster analysis, “Young Students,” “Working Women,” “Sporadic Users,” and “Elderly Passengers.” Then they generated five different decision trees, one for each group and one overall, to define the difference between “poor” and “good” service ratings from the passengers. For example, in the Overall tree, service frequency was the first, most important, branch on the tree; but for Young Students the first branch was punctuality. This is an excellent study that is easy to read and understand and shows how two different data mining techniques, cluster analysis and decision tree building, can be used in a single study.

1.6.0.3 Market Basket

A market basket analysis evaluates the products that customers purchase at the same time (they are in the same “market basket”) and then uses that data to drive decisions on things like store organization and sales. For example, it is well known that customers who purchase beer also tend to purchase potato chips. What may be surprising is that customers who purchase beer also tend to purchase baby diapers. A market basket analysis does not attempt to determine why these relationships exist, just that they do exist. A grocery store owner may choose to put beer and chips on sale at the same time or create a display near the front of the store with both beer and chips.

Figure 9 is a graphic representation of the rule set found in Figure 6, created from a market basket analysis of a bakery. The goal was to determine what sorts of products customers purchased with cherry tarts. Rule 53 indicates that apricot danish and cherry tarts tend to be purchased at the same time, and that would be valuable information for a bakery owner. The explanation of the numbers displayed (“Support,” “Confidence,” etc.) is beyond the scope of this class, but business owners would want to work with a researcher who could complete this sort of analysis.

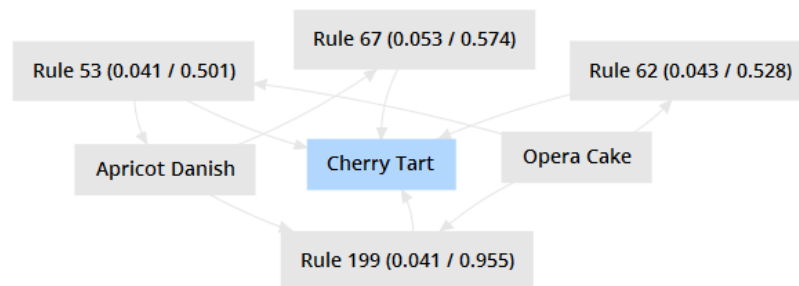


Figure 9: Market Basket Graph

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
53	Opera Cake	Cherry Tart, Apricot Danish	0.041	0.501	0.962	-0.123	0.037	9.431	1.897
62	Opera Cake	Cherry Tart	0.043	0.528	0.964	-0.121	0.036	5.671	1.923
67	Apricot Danish	Cherry Tart	0.053	0.574	0.964	-0.132	0.044	6.156	2.126
199	Apricot Danish, Opera Cake	Cherry Tart	0.041	0.955	0.998	-0.045	0.037	10.255	20.322

Figure 10: Market Basket Rules

As an example of a market basket analysis, Musalem, Aburto, and Bosch completed a study of a mid-sized supermarket in Latin America[4]. That market sells approximately 7,000 products and the average basket size is products from 3.6 categories. They reviewed the register receipts from the month of July 2000 and recorded the products sold, units sold, date, and time. They categorized the products into

four categories: non-perishable (cereals, flour, noodles, etc.), immediate consumption (meat, milk, cheese, etc.), hygiene (shampoo, conditioner, diapers, etc.), and hedonic (ice cream, beer, candy, etc.). They found that items in the hygiene basket were associated with larger transactions but there were not many of them. The non-perishable basket had a high number of different categories and a high transaction size, so these buyers are extremely important to the market. This study is a good example of how market basket analysis can be used to improve sales.

1.7 SUMMARY

Figure 11 illustrates the relationship between the various types of data and the rating scales commonly used to work with those data types. Researchers with a positivism philosophy tend to use parametric statistical analysis and gather interval and ratio data. Researchers with an interpretivism philosophy tend to use nonparametric statistical analysis and gather nominal and ordinal data. Nominal data are typically gathered with binary and semantic difference rating scales while ordinal data are typically gathered with Likert and Guttman rating scales.

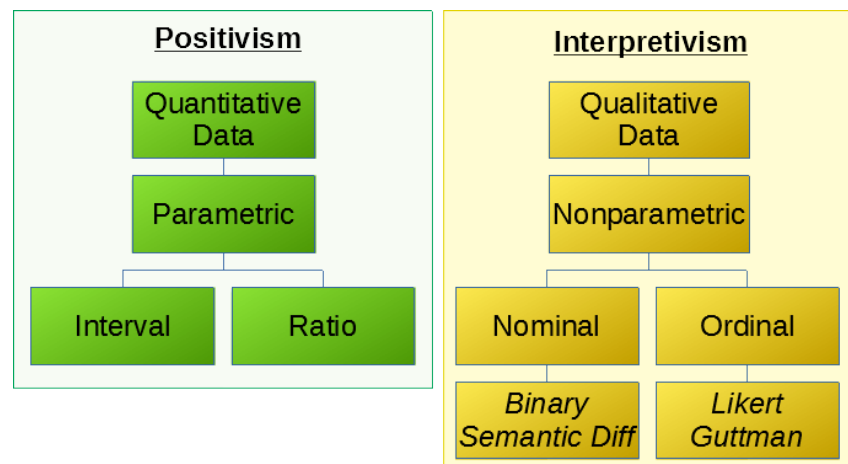


Figure 11: Data types

Part III

QUALITATIVE METHODS

Qualitative methods are based in the evaluation of non-numeric data, like photographs and text documents. These methods include activities like field work, unobtrusive, and interpretive research methods.

Part IV

MIXED METHODS

All quantitative and qualitative research methods have certain strengths and weaknesses. Mixed methods are an attempt to use more than one research method on a given project to utilize the strengths of each method while mitigating their weaknesses.

Part V

REPORTING

After a research project is completed, the investigator must report the results of the project, often in both written and oral forms. This chapter concerns the reporting process.

Part VI

APPENDIX

APPENDIX

No appendix content yet.

GLOSSARY

applied research	Research that is intended to be applied to a situation rather than further the knowledge of some topic. For example, if a researcher completes a project designed to increase the sales of bottled water in a small town it would be considered applied research. <i>see</i>
basic research	Research that is intended to be further the knowledge of some topic rather than be applied to a specific situation. For example, if a researcher completes a project designed to refine some aspect of the Law of Supply and Demand it would be considered basic research. <i>see</i>
concurrent validity	The degree that a measure relates to an outcome that is presumed to occur simultaneously. For example, the results of a new employee attitude test would be the same as an older test if those tests have high concurrent validity. <i>see</i> predictive validity
construct validity	The degree to which a test measures what it claims to measure. For example, if a research project purports to investigate some aspect of local farmers' markets, does the project actually research that aspect? Construct validity is sometimes thought to be the overarching type of validity since research projects that do not address the construct of interest can have no other validity. <i>see</i> validity

content validity	A determination of whether a measure correctly assesses the construct's content. For example, if a research project is attempting to determine the drivers for total sales in a store but only measured the price of the merchandise being sold then ignoring factors like advertising, competition, and even the general economy of the region would call into question the content validity of the study. <i>see</i> validity
continuous data	Continuous data are a type of quantitative data that can represent any measured value, including fractions and decimals. In mathematics terms, continuous data are members of the real number system. 15 , <i>see</i> quantitative data
convergent validity	The closeness that two measures relate to, "converge on," a single construct. For example, if a research project measures the amount of sales of carbonated drinks, fruit juices, and bottled water in a store it would be expected that those would converge on a construct of "drink sales." <i>see</i> discriminant validity
criterion validity	The degree to which a measure is related to an outcome. <i>see</i> validity
database	A database is a collection of data that is organized in such a way that the data can be easily managed. While the internal structure of a database can be complex, it is typically represented as tables with data in rows and columns, like a spread sheet. 13
dependent variable	Dependent variables are the outcomes for an observation. For example, if a medical researcher conducts an experiment where a drug is administered and then the patient's blood pressure is measured, the blood pressure reading is the dependent variable; that is, the blood pressure depends on the drug being administered. 16 , <i>see</i> independent variable

descriptive research	Research that is designed to describe observed phenomena. The goal is to improve understanding rather than explore new ideas. <i>see</i> exploratory research
discrete data	Discrete data are a type of quantitative data that can be counted with integers. In mathematics terms, discrete data are integers, though negative values are rather rare. 15 , <i>see</i> quantitative data
discriminant validity	The degree that a measure does not measure, “discriminates between,” one of two competing constructs. For example, a measure of the sale of toiletries in a department store would not be related to the construct of “drink sales.” <i>see</i> convergent validity
excess kurtosis	Excess kurtosis is a measure of the “tailedness” of a normal distribution. Greater excess kurtosis values indicate longer “tails” (and a “sharper” appearance) on the graph of the distribution. 11
explanatory research	Research that is designed to explain an observed phenomena or process. <i>see</i> exploratory research
external validity	The degree to which a research project’s results can be applied outside the context of the study. For example, if the results of a research project that studied manufacturing firms in the mid-west could be applied to firms in the south then that study would have high external validity. <i>see</i> validity
face validity	A determination of whether an indicator is a reasonable measure of an underlying construct “on its face.” For example, is the amount of money spent on live theater tickets a measure social class? <i>see</i> validity

independent variable	Independent variables are those that create an observed effect. For example, if a farmer conducts an experiment where different types of fertilizer are applied to two fields in order to see which is more effective then the type of fertilizer is the independent variable; that is, the type of fertilizer is the variable that is creating the observed effect. 16, <i>see</i> dependent variable
internal validity	The degree to which a research project avoids confounding multiple variables within the study. A project with high internal validity facilitates selecting one explanation over an alternate since the number of confounding variables are controlled. <i>see</i> validity
interval data	Interval data are a type of quantitative data that are measured along a scale where each point is an equal distance from the next. It is possible to compare the distance between two points on an interval scale; for example, the difference between 90 and 100 degrees is the same as the difference between 40 and 50 degrees. However, since an interval scale does not have a zero point, stating 100 degrees is twice as hot as 50 is not possible. 6, <i>see</i> quantitative data
nominal data	Nominal data are a type of qualitative data that are grouped but with no order implied in the grouping. As an example, the gender of survey respondents is nominal data. 5, 15, <i>see</i> qualitative data
nonparametric	Nonparametric data are data that do not conform to a distribution, are skewed, or are qualitative in nature. Statistical tests that work with nonparametric data are generally less powerful and predictive than tests that work with parametric data. 17, 21, <i>see</i> nonparametric

normal distribution	A listing of all the possible values in a data set along with the number of times each value actually appears is called a “distribution.” In many, perhaps most, business research projects a distribution exhibits a bell shape when plotted on a graph where the values in the middle of the range are more frequent than values at the extremes of the range. This is called a “normal distribution” since it is so common. 10
ordinal data	Ordinal data are a type of qualitative data that are grouped where the groupings have an implied order. As an example, the “satisfaction” rating on a customer survey typically permits respondents to choose from several levels where one level is somehow better than another. 6 , 15 , see qualitative data
paradigm	A pattern or model of how things work in the world. see theory
parametric	Parametric data are data that conform to a distribution, usually a normal distribution. Statistical tests that work with parametric data are generally much more powerful and predictive than tests that work with nonparametric data. 17 , 20 , see nonparametric
positivist	A researcher who uses positivist techniques on research projects. see positivism
predictive validity	The degree to which a measure predicts an outcome. For example, does increasing beer sales (a measure) predict increasing potato chip sales? see concurrent validity
qualitative data	Qualitative data approximates or describes attributes that cannot be directly measured, like employee morale, customer relationships, and management effectiveness. Typically, qualitative data attempt to answer questions like “why” and “how come.” 5 , 15 , see quantitative data

qualitative research	Qualitative research typically intends to explore observed phenomena with a goal of developing hypotheses and dive deep into a problem. Qualitative data collection involves semi-structured activities like focus groups and ethnographies. <i>see</i> quantitative research
quantitative data	Quantitative data are numeric measurements of attributes, like the number of employees, the median value of housing, and total revenue. Quantitative data are gathered and analyzed using statistical methods. 6 , <i>see</i> qualitative data
quantitative research	Quantitative research typically uses numerical data and statistical analysis to find patterns and generalize results to a large population. Quantitative data collection involves structured activities like surveys, interviews, and systematic observations. <i>see</i> qualitative research
ratio data	Ratio data are a type of quantitative data that are measured along a scale where each point is an equal distance from the next and there is a zero point. An example of ratio data is people's heights, which is measured along a uniform scale, e.g. inches or centimeters. Because there is a true zero point, it is possible to determine that one person is twice as tall as another. 6 , <i>see</i> quantitative data
reliability	A descriptor for the consistency of a concept's measure. It is desirable to achieve the same, or nearly same, values for each sampling. For example, if the mean age of the people in one sample is 30 while in another is 50 then this would indicate a problem with reliability of the data. <i>see</i> validity
skew	Skew is asymmetry in a distribution, so a graph appears distorted. A positive skew creates a longer tail on the right side of the graph. 12

statistical conclusion validity	The degree to which the conclusions found in a research project are correct. Studies with high statistical conclusion validity minimize the two types of statistical errors: Type I (finding a correlation when there is none) and Type II (failing to find a correlation when one exists). <i>see</i> validity
theory	A system of ideas that is intended to explain phenomena. Theories that are accepted by scientists have been repeatedly tested and can be used to make accurate predictions. Unlike common usage, a scientific theory is a tested, falsifiable explanation for phenomena. <i>see</i> paradigm
translational validity	The degree to which a construct has been measured by a research project. Translational validity is divided into two types: face and content. <i>see</i> validity
validity	A descriptor of whether a research project is measuring the variable under question. For example, if a project hypothesis is that older men tend to tip more than younger men then the validity of the study would be in question if the researcher only sampled men under the age of 40. <i>see</i> reliability

BIBLIOGRAPHY

- [1] Anol Bhattacharjee. "Social science research: Principles, methods, and practices." In: (2012).
- [2] Amy Blackstone. *Principles of sociological inquiry—qualitative and quantitative methods*. The Saylor Foundation, 2012.
- [3] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj Agarwal. "How short is too short? Implications of length and framing on the effectiveness of privacy notices." In: *12th Symposium on Usable Privacy and Security (SOUPS)*. 2016, pp. 321–340.
- [4] Andres Musalem, Luis Aburto, and Maximo Bosch. "Market basket analysis insights to support category management." In: *European Journal of Marketing* (2018).
- [5] Juan de Oña, Rocío de Oña, and Griselda López. "Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation." In: *Transportation* 43.5 (2016), pp. 725–747.
- [6] CNN Politics. *Election 2016*. 2016. URL: <https://www.cnn.com/election/2016/results/exit-polls> (visited on 12/23/2018).
- [7] Paul Christopher Price, I-Chant A Chiang, and Rajiv Jhangiani. *Research methods in psychology*. BCcampus, BC Open Textbook Project, 2015.
- [8] Kalyani M Raval. "Data Mining Techniques." In: *International Journal of Advanced Research in Computer Science and Software Engineering* 2.10 (2012).
- [9] Stanley Smith Stevens et al. "On the theory of scales of measurement." In: (1946).
- [10] Peter Titlebaum and Heather Lawrence. "Perceived motivations for corporate suite ownership in the 'big four' leagues." In: (2016).
- [11] Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. "Understanding bike-sharing systems using data mining: Exploring activity patterns." In: *Procedia-Social and Behavioral Sciences* 20 (2011), pp. 514–523.

COLOPHON

This document was typeset using the typographical look-and-feel classicthesis developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". classicthesis is available for both L^AT_EX and L^yX:

<https://bitbucket.org/amiede/classicthesis/>

Happy users of classicthesis usually send a real postcard to the author, a collection of postcards received so far is featured here:

<http://postcards.miede.de/>