Chapter One

Intro to Research Methods

Samples and Populations

Measuring Data

Relationships Bewteen Variables

Causation

Populations and Samples

A **population** is a group that is being researched.

A **sample** is a subset of the population from which data are actually collected.

Population values are called **parameters**. Using them to predict sample values is called **probability**.

Sample values are called **statistics**. Using them to estimate population values is called **statistics**.

Value	Population	Sample	Sample example
Size	N	n	$n = 20$ high school students were surveyed.
Mean	μ ("mu")	$\bar{x}$ ("x-bar")	The average age was $\bar{x} = 17.2$ years.
Standard Deviation	σ ("sigma")	s	The standard deviation was $s = 0.28$ years.
Proportion	p	$\hat{p}$ ("p-hat")	$\hat{p} = 30\%$ of the students were taking Statistics.

Descriptive Statistics

Means and standard deviations are used to summarize numerical data sets. Proportions are used to summarize nonnumerical data sets.

Statistic	Description	When used	Example
Proportion	portion of a whole	Each value does or does not meet a specific criterion.	*What is your favorite flavor of ice cream?* 16% of respondents say chocolate.
Mean	average	Each value is numerical.	*How much ice cream do you eat each year?* The average of all the responses is 5 gallons.
Standard Deviation	amount of variation	Each value is numerical.	*How much ice cream do you eat each year?* The standard deviation of all the responses is 1.5 gallons.

Levels of Measurement

Data can be considered at one or more levels.

Level	Description	Examples for *When did they arrive?*	Explanation
Nominal	The data can be categorized.	Saturday Tuesday	Saturday could be before or after Tuesday.
Ordinal	The data can be put in order.	1st 2nd 8th	2nd comes after first, but it is unknown how long after.
Interval	Differences between data values are meaningful.	12:00 1:00 1:10	12:00 is an hour before 1:00, but 12:00 is not 12 times as much as 1:00, and 0:00 does not mean there is no time.
Ratio	Ratios between data values are meaningful. A value of zero means there is none of what is being measured.	5 minutes late 15 minutes late	15 minutes is three times as much as 5 minutes. Zero minutes late means not late at all.

Operational Definitions

An **operational definition** states exactly how a variable will be measured.

Variable	Operational definition example I	Operational definition example II
Age	number of birthdays	years and days since birth
GPA	unweighted overall GPA last semester	weighted academic GPA for 11th grade
Athleticism	number of pull-ups	mile time

For conceptual variables such as athleticism, researchers often mathematically combine multiple measures into a single value called an **index**.

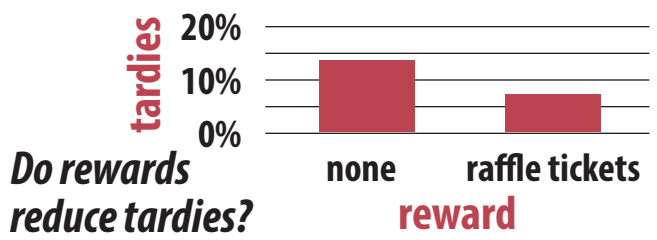
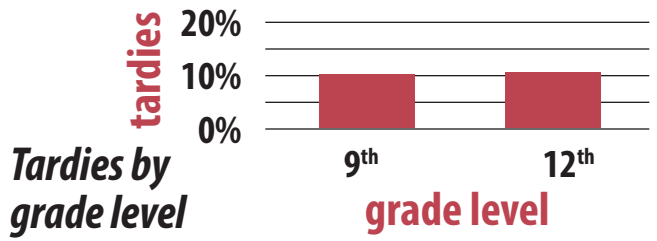
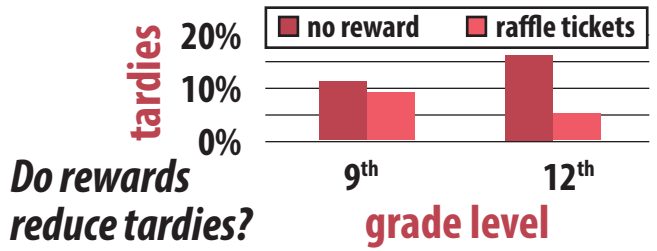
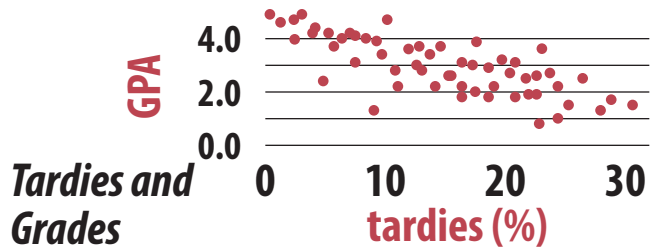
Variables

Type	Description	Example
Independent	hypothesized to affect the dependent variable directly or through mediator variables	**Reading the textbook** causes higher test scores.
Dependent	hypothesized to be affected by the independent variable directly or through mediator variables	**Test scores** are improved if the textbook is read.
Mediator	explains how the independent variable affects the dependent variable	Reading the textbook gives students **questions to ask in class**, which causes higher test scores.
Moderator	influences the strength of the relationship between the independent variable and dependent variable	Reading the textbook affects test scores differently depending on **how conceptual the chapter is**.
Extraneous	affects the dependent variable, but does not fit into any category above	Amount of **extracurricular activities** affects test scores.
Confounding	extraneous variable that shows how the independent variable is linked to the dependent variable without directly or indirectly affecting it	**Better students** are more likely to read the textbook, and are more likely to do well on tests whether or not they read the textbook.

Extraneous and Confounding Variables

Variable	Extraneous but not confounding	Confounding
Type of error created	Random error: All conditions are affected randomly, and thus approximately equally.	Systematic error: Some conditions are systematically affected differently than others.
Problem created	Due to the random noise, the data may not show the link between the independent variable and the dependent variable, or, less commonly, may indicate a relationship when there is none.	The data may show the hypothesized link between the independent variable and the dependent variable, but it is not known if this is due to the independent variable or the confounding variable.
Severity of problem	Moderate: The researchers are more likely to fail to reach a conclusion, but are not likely to reach a conclusion that is not valid.	Major: The researchers are likely to reach a conclusion that is not valid.
How to avoid	Using a large sample size averages out random variations.	Confounds from participant differences can be eliminated by random assignment. Confounds from procedural or environmental differences can be reduced by pilot studies, standardization of procedure, and careful critical analysis of method.

Research Designs

Design	Description	Example
Experimental	The independent variable has two or more conditions, and each participant is randomly assigned to one condition or one order of conditions.	
Quasi-Experimental	The independent variable has two or more conditions, but there is no random assignment.	
Factorial	There is more than one independent variable. Each can be either experimental or quasi-experimental.	
Correlational	The independent variable and the dependent variable are both numerical (not categorical).	
Observational	The participants are not influenced by the study.	The studies above that do not involve rewards may be observational.

Correlation and Causation

Correlation does not imply causation: Two variables being related does not necessarily mean that one affects the other. Causation does imply correlation, however.

Relationship:	Correlation	Causation
Summary	The dependent variable can be predicted by the independent variable.	The dependent variable is affected by the independent variable.
What it explains	*what* relationship exists between the variables	*why* the relationship exists between the variables
How it can be established	any study, including quasi-experimental designs and correlational designs	only true experiments (that is, with random assignment)
Confounding variables	may be the primary or only reason for the relationship—the independent variable itself may have little or no effect on the dependent variable	may be eliminated, because random assignment can make the groups initially exactly identical other than random fluctuations
Example: college degree and salary	People with college degrees have higher salaries on average. This could be due to the degrees themselves, but it also could be due to confounding variables such as socioeconomic status and motivation.	Sending out identical resumes, except that some have a college degree and some do not, could determine whether or not degrees actually cause people to get higher salaries.

Affect and Effect

Discussions of causation frequently use forms of the words *affect* and *effect*.

Word	Part of speech	Clarification	Examples
Affect(s)	verb	has a subject, which is usually one of the following: • an independent variable such as *age* • a confounding variable such as *socioeconomic status*	Smoking affects health. Childhood experiences affect adult personality.
Effect(s)	noun	usually preceded by one of the following: • the article *the* or *an* • an adjective such as *significant* or *two* • a possessive such as *religion's* or *its*	Alcohol has multiple effects. The data demonstrate music's effect on concentration.