CHAPTER ONE: INTRODUCTION TO RESEARCH METHODS — Review September 24 ⊱ Test October 6

After introducing some terminology and symbols that will be used throughout the year, this chapter explores why the most straightforward conclusion from data is not always an accurate one. Of crucial importance is the concept that correlation does not by itself imply causation. In other words, if two variables are related to each other, it does not necessarily mean that one affects the other. Instead, they may be correlated because of a third variable, called a confound, that affects both. For example, average test scores are higher in Scotts Valley schools than in Watsonville schools, but that does not mean that Scotts Valley schools must be better—the difference could be due to confounding variables such as English language fluency. The best way to reduce the possibility of confounding variables is to randomly assign which condition or order of conditions each participant takes part in, although this is not always feasible.

1-A Populations and Samples — Tuesday • 9/8

population • sample • sample size n • random selection • sampling bias • parameter • statistic • mean • proportion • μ • $\bar{x}$ • σ • s • p • $\hat{p}$

❶ Identify samples and populations.
❷ Identify possible sources and consequences of sampling bias.
❸ Label a mean, standard deviation, or proportion with the correct symbol.
❹ Distinguish between probability and statistics.

1-B Measuring Data — Monday • 9/14

operational definition • Likert scale • level of measurement • data • nominal • ordinal • interval • ratio

❶ Develop an operational definition for a variable.
❷ Classify data by level of measurement.

1-C Relationships Between Variables — Thursday • 9/17

independent variable • level • experimental group • control group • placebo • dependent variable • experiment • random assignment • quasi-experiment • correlational • observational study • causal relationship • affect • effect

❶ Identify an experiment or quasi-experiment's independent variables and their levels, and identify its dependent variables.
❷ Identify a possible mediator variable in an experiment.
❸ Identify a possible moderator variable in an experiment or quasi-experiment.
❹ Identify a main effect, simple effects, and an interaction in an experiment.
❺ Distinguish between experimental, quasi-experimental, and correlational research designs.
❻ Distinguish between *affect* and *effect* meaning *influence*.

1-D Extraneous and Confounding Variables — Tuesday • 9/22

extraneous variable • confounding variable • random error • systematic error • socioeconomic status

❶ Interpret significant results with or without causation as appropriate.
❷ Identify possible confounding variables due to preexisting differences between participants in different conditions.
❸ Identify possible confounding variables due to differences in the procedure between different conditions.
❹ Identify possible confounding variables resulting from differences in time or other environmental factors between different conditions.
❺ Identify possible extraneous variables that are not likely to be confounding.
❻ Explain the possible effects of specific extraneous and confounding variables on a study.

1-A 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. It should be representative of the population.

SAMPLE SIZE, labeled n, is the number of participants in a sample.

❶ Identify samples and populations.

1. Each specific group that provides data is a sample.
2. The overall groups that the samples are intended to represent are the populations.

❶ Logan surveys 20 random SVHS juniors and 18 SVHS seniors to ask whether or not they are happy for school to start.

1. The samples are the 20 random SVHS juniors and the 18 SVHS seniors.
2. The populations are SVHS juniors and SVHS seniors in general.

RANDOM SELECTION means that every member of the population has an equal chance of being part of the sample. A lack of random selection can lead to SAMPLING BIAS, in which the sample is not representative of the population it is intended to represent.

❷ Identify possible sources and consequences of sampling bias.

1. Identify differences between the type of people sampled and the intended population.
   If these difference may cause a difference in the dependent variable, then the study is valid only for the population actually represented by the sample.
2. Identify characteristics about the people actually sampled that are not representative of the group from which they were sampled.
   If these characteristics may cause a difference in the dependent variable, then the study may not provide valid conclusions.

❷ To see how long teenagers can hold their breath, Svea times 20 of her friends at SVHS at break.

1. She is only testing SVHS teenagers. Compared to other students, SVHS teenagers may have longer times because they tend to be healthy and less likely to smoke. Therefore, her sample may accurately represent only SVHS teenagers rather than people in general.
2. Svea's friends are athletes and are less likely than average SVHS teenagers to be smokers. Therefore, his sample may not even accurately represent SVHS teenagers.

A Population PARAMETER is a value calculated from data from an entire population.

A Sample STATISTIC is a value calculated from data from a sample of the population.

The most commonly used parameters and statistics are means, standard deviations, and proportions.

A PROPORTION is a percentage of the group.

A MEAN is an average.

A STANDARD DEVIATION is a measure of variability, to quantify the extent to which values vary from one another.

A proportion can be calculated when every individual's response either does or does not fall into a specific catagory, such as *child*, *chocolate*, or *public school*. Proportion questions can always be given as yes-or-no questions, such as *are you under 18* or *is your current school a public school*?

A mean and standard deviation can be calculated when every individual's overall response consists of a number, such as *25 minutes* or *8 on a scale of 1 to 10*.

The lowest possible standard deviation is 0, representing no variation in the data set, such as { 8, 8, 8, 8 }. The highest possible standard deviation is half the range, which occurs when half of the values are one value and the other half are another value, such as { 8, 8, 12, 12 }.

Parameters and statistics use different symbols. Parameters usually use Greek letters, and statistics usually use English letters.

Measure:	mean	standard deviation	proportion
population parameter:	μ (mu)	σ (sigma)	p
sample statistic:	$\bar{x}$ (x-bar)	s	$\hat{p}$ (p-hat)

If there are multiple samples, each statistic is labeled with a subscript number, such as $\bar{x}_1$ and $\bar{x}_2$.

❸ Label a mean, standard deviation, or proportion with the correct symbol.

1. If the value is calculated from data from the whole population, it is a parameter.
   If the value is intended to represent a population but is calculated from data from a sample, it is a statistic.
2. If each participant's response either is or is not a specific value, such as *yes* (versus no), or *vanilla* (versus any other flavor), the value is a proportion.
   If each participant's response is a number and these numbers are averaged, the value is a mean.
   If each participant's response is a number and the calculated value shows how much variation there is in the responses, it is a standard deviation.

❸ Out of 60 seniors, 51 say they would rather go a week without a car than a week without a smartphone.

1. Only 60 seniors were surveyed. This is a sample, not all seniors.
2. Each participant only had two options—car or smartphone—so the result is a proportion.
   $\hat{p} = \frac{51}{60}$

The field of PROBABILITY uses parameters to make predictions about samples. For example, given a coin lands on heads $p = 50\%$ of the time, what is the probability that it will land on heads 10 times in a row?

The field of STATISTICS uses statistics to make estimates about populations. For example, given a coin landed on heads 10 times in a row, is it reasonable to conclude that $p \neq 50\%$ for this coin?

❹ Distinguish between probability and statistics.

1. Identify whether the known information is for the whole population or for a specified sample. Keep in mind that the whole population is not necessarily people, but instead could be all possible occurrences of an event, such as an infinite number of coin flips.
2. If the population parameters are already known, such as the average IQ of all people is 100 or the overall proportion of heads in coin flips is 50%, probability will be used.
   If only sample statistics are known instead of population parameters, statistics will be used.

❹ Out of 60 seniors, 51 say they would rather go a week without a car than a week without a smartphone.

1. The known information is only for the 60 seniors in the survey, not for the whole population of seniors.
2. The sample statistic $\hat{p} = \frac{51}{60} = 85\%$ is being used as an estimate for the proportion of all seniors, so this is statistics.

1-B Measuring Data

Not all concepts have straightforward, universal definitions. This is especially a concern for conceptual variables such as anxiety or opinions, but it is also an issue for variables that can be measured in different ways, such as earnings (annual salary, hourly wages, etc.) or growth (centimeters, kilograms, etc.). An OPERATIONAL DEFINITION states how a variable will actually be observed and measured for a given research design.

A common operational definition for opinions is response on a LIKERT Scale, in which people rate their agreement with items on a range, such as 1 to 5 with 1 meaning *strongly disagree* and 5 meaning *strongly agree*.

More abstract variables are often measured by an INDEX, which uses an operational definition that mathematically combines multiple measurements into one.

❶ Develop an operational definition for a variable.

1. Decide on a measurable variable or collection of measurable variables that validly reflect the concept.
2. Decide precisely how the measurable variables will be measured, and, if applicable, combined.

❶ List some possible operational definitions for the variable "academic achievement" for seniors.

1. measurable variable	2. operational definition
a) grades	9-11 weighted academic GPA
b) standardized test scores	highest SAT total score
c) teacher ratings	1-5 Likert item: *This student learned a lot in my class last year.*
d) index using all of the above	$(1000 \times \text{GPA}) + \text{SAT} + (200 \times \text{average teacher rating})$

DATA is a plural word that essentially means *observed values*.

Data can be classified into four LEVELS OF MEASUREMENT, listed below from lowest to highest:

NOMINAL Data can be sorted into categories, such as red, green, and white, but there is not necessarily any order to the categories. For example, white does not necessarily come before or after red.

ORDINAL Data can be put in order, such as first to last or strongest to weakest, but the difference between positions may not be known. For example, a race's first place finisher may have beaten the second place finisher by five seconds or by two minutes.

INTERVAL Data can be subtracted to find a meaningful difference, such as 60° being 40° warmer than 20°, but may not be able to be divided to find a meaningful ratio. In this case, it would not make sense to say that 60° is three times as warm as 20°.

RATIO Data can be divided to find a meaningful ratio, such as $60 per hour being three times as much as $20 per hour. For ratio data, a value of zero means there is none of what is being measured.

Any variable can be treated as a lower level of measurement. For example, temperatures are interval data, but also can but put in order from coldest to warmest.

❷ Classify data by level of measurement.

1. All data can be considered nominal.
2. If the data can be put in order based on what they are measuring, they are ordinal.
3. If the data have specific numeric values independent of the other data, they are interval. For example, *9 minutes* is a specific amount, but *slower* or *9th place* are not because they could represent different values depending on the other data.
4. If a value of zero would mean there were none of the variable (even if such a value is not possible in reality), then meaningful ratios can be found and the data are ratio.

❷ Show how calorie consumption meets the criteria for different levels of measurement.

2. Ordinal: The data could be sorted as *low, medium,* or *high*.
3. Interval: Numbers of calories are specific numeric values.
4. Ratio: Zero calories means no caloric content.

❷ For each of the following variables, give a possible way it could be measured and what level of measurement applies.

a) color	b) temperature	c) distance	d) iPhone capacity	e) agree with *Trump is an effective president.*
color name	degrees Celsius	meters	gigabytes	strongly disagree, disagree, agree, strongly agree
nominal	interval	ratio	ratio	ordinal

1-C Relationships Between Variables

An INDEPENDENT Variable (IV), or FACTOR, has two or more different conditions, called LEVELs, with the prediction that the dependent variable is not the same across all levels.

A DEPENDENT Variable (DV) is the variable whose values the researcher finds out, often by measuring, to learn if the prediction was correct.

❶ Identify the independent and dependent variables of an experiment or quasi-experiment, including their levels or operational definitions.

1. The independent variable comes first and is hypothesized to influence or predict the dependent variable. On a graph, it is on the x-axis.
2. The dependent variable comes second and is hypothesized to be influenced by or predictable based on the independent variable. On a graph, it is on the y-axis.
3. The levels of the independent variable are its different conditions. On a graph, each bar represents a level.
4. The dependent variable may need an operational definition (see 1-B ❶).

❶ Giselle predicts that starting each class with five minutes of yoga increases math achievement.

1. She hypothesizes that yoga affects math achievement.
2. She hypothesizes that math achievement is affected by yoga.
3. In the experimental condition, students will start each class with five minutes of yoga. In the control condition, they will do no yoga.
4. She could use overall semester math grade as the dependent variable.

A MEDIATOR Variable is affected by the independent variable and in turn affects the dependent variable, showing how the independent variable indirectly affects the dependent variable.

❷ Identify a possible mediator variable in an experiment.

1. Identify a possible direct effect of the independent variable that could influence the dependent variable in the manner hypothesized.

❷ Identify a possible reason why yoga may improve math grades.

1. Yoga could help students relax, which may help them learn math.

A MODERATOR Variable is an additional factor that shows how the independent variable can have different effectiveness in different conditions.

❸ Identify a possible moderator variable in an experiment or quasi-experiment.

1. Consider how different conditions could lead to different results of the study.

❸ Identify a variable that could affect the effect of yoga on math achievement.

1. Students' flexibility could influence how effective yoga is at improving math grades.

A MAIN Effect is an overall effect of the independent variable on the dependent variable.

A SIMPLE Effect is an effect of the independent variable on the dependent variable under specific circumstances (that is, within a single level of a moderator variable).

An INTERACTION occurs when the simple effects are not the same as each other.

❹ Identify a main effect, simple effects, and an interaction in an experiment.

1. There is a main effect if the independent variable has an effect on the dependent variable overall, regardless of moderator variables.
2. There is a simple effect in each condition of a moderator variable in which the independent variable affects the dependent variable.
3. There is an interaction if one simple effect is different from another, that is, if the independent variable is more effective in one condition than another.

❹ Giselle finds that students in the daily yoga condition do end up with higher math grades overall, and that flexible students get the most benefit.

1. The positive effect of yoga on math grades is a main effect.
2. The large positive effect of yoga on math grades among flexible students is a simple effect.
   The small positive effect of yoga on math grades among unflexible students is another simple effect.
3. The effect of yoga on math grades being greater for flexible students than for unflexible students is an interaction.

An EXPERIMENTAL Design is a setup in which the independent variable has different levels and the experimenter randomly chooses which condition each participant is in or which order each participant takes part in each of the different conditions. This is called RANDOM ASSIGNMENT.

A QUASI-EXPERIMENTAL Design is the same as an experimental design except without random assignment. Quasi-experiments are used when the independent variable is not randomly assigned because it is not possible such as for sex or religion, not practical such as for wealth or school attended, or not ethical such as for drug use or exposure to toxins.

A CORRELATIONAL Design uses a numerical independent variable on a scale instead of a categorical independent variable with a small number of possible values. Correlational studies do not have distinct levels or random assignment.

A FACTORIAL Design involves more than one factor, each of which can be experimental or quasi-experimental.

In an OBSERVATIONAL Study, data are collected without influencing the participants. Experimental designs cannot be observational.

❺ Distinguish between experimental, quasi-experimental, and correlational research designs.

1. If there is random assignment of participants to the different conditions or order of conditions, the research is experimental.
2. If there are different conditions that are being compared but participants are not randomly assigned, the research is quasi-experimental.
3. If the independent variable has many possible different numerical values, the research is correlational.

❺ What type of study is Giselle's study on yoga?

1. Giselle will randomly assign who does yoga and who does not, so this is experimental.
2. Giselle cannot randomly assign flexibility, so this is quasi-experimental.
3. There are two factors (one independent variable and one moderator variable), so the study is using a factorial design. In this case, one factor is experimental and one is quasi-experimental.

The words AFFECT and EFFECT both mean influence. In this case, *affect* is a verb and *effect* is a noun.

❻ Distinguish between *affect* and *effect* meaning *influence*.

1. If it is preceded by an adjective, article, or possessive pronoun (e.g., *the, an, its, some, three, significant,* etc.), it is a noun: effect.
2. If it has a subject, it is a verb: affect. The subject will usually be an independent variable or a confounding variable (e.g., *caffeine, study habits, socioeconomic status*).
3. *Affected* (not *effected*) means *influenced*.
4. *Effective* (not *affective*) means *influential*.

❻ Even though people may seem to be un__ffected, the things we say can have big __ffects on others. Likewise, the way we say things can __ffect others, whether or not we realize the __ffect we are having.

a) an adjective meaning *uninfluenced*: unaffected
b) a noun preceded by the adjective *big*: effects
c) a verb with the subject *the way we say things*: affect
d) a noun preceded by the article *the*: effect

1-D Causation

A CAUSAL Relationship is one in which the independent variable affects, rather than is simply correlated with, the dependent variable. Causation cannot be determined without random assignment.

❶ Interpret statistically significant results with or without causation as appropriate.

1. In quasi-experimental or correlational research, conclude that the variables are associated with each other, but do not conclude that the independent variable necessarily affects the dependent variable because any one or more of the following could be true:
   a) Causation: The independent variable affects the dependent variable (as predicted).
   b) Reverse causation: The dependent variable affects the independent variable.
   c) Confounding: One or more other variables affect the independent variable and the dependent variable together.
2. In experimental research, conclude that the independent variable does in fact affect the dependent variable, unless there is a possibility that the way the experiment was carried out may have led to confounding.

❶ Wang and Kenny (2014) found that the more teenagers were harshly verbally disciplined by their parents, the more depressed and aggressive they tended to be.

This is correlational research. It can be concluded that depressed and aggressive behavior is correlated with harsh verbal discipline, but it cannot be determined which one or more of the following is the case:

a) Harsh verbal discipline causes teenagers to become depressed or aggressive.
b) Depressive or aggressive behavior causes parents to give harsh verbal discipline.
c) Other factors, such as drug use or poor grades, causes depressed or aggressive behavior and likewise causes parents to give harsh verbal discipline.

An EXTRANEOUS Variable is any variable, other than the variables being studied, that affects the dependent variables.
A CONFOUNDING Variable is an extraneous variable that systematically affects the dependent variables differently in some conditions than in others. Confounding variables provide alternative explanations for why the independent variable and the dependent variable are linked.

	Extraneous but not confounding	**Confounding**
Definition	affects the dependent variable but is not linked with the independent variable	affects the dependent variable and is linked with, but not affected by, the independent variable
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 really 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.
Example: To test the effects of caffeine on attention, 10 boys are given a glass of Coke and 10 girls are given a glass of water.	Some kids are naturally more attentive than others, and the lesson may be more interesting to some students than to others. Because of these random factors, there may not be a clear difference between the average attention levels of the two groups.	If the kids who drank Coke have decreased attention levels, it could be because of the caffeine, but it also could be because they are boys (a participant confound) or because of the sugar in the soda (a procedural confound). Therefore, no conclusion should be made about the original hypothesis that caffeine affects attention.

Understanding how confounding variables limit the validity of conclusions is the most important aspect of this course.

Sean and Linus are testing to see if people have better memory for words presented visually or auditorily. Sean reads a list of 30 words to a seventh period Econ class and then has them write down as many as they can remember. Linus does the same with a fifth period Advanced Physics class, except he shows the words in a PowerPoint presentation instead of reading them.

❷ Identify possible confounding variables due to preexisting differences between participants in different conditions.

1. Identify ways in which the people in one condition are different from the people in other conditions that could cause the different conditions to get different results even if the independent variable has no effect. (This will not happen if there is random assignment.)

❷ 1. They did not randomly choose which condition each person would be in. A possible confound is that people who take Advanced Physics tend to be top students, whereas a full range of students take Econ.

❸ Identify possible confounding variables due to differences in the procedure between different conditions.

1. Identify differences between conditions in the way data were collected that could cause the different conditions to get different results even if the independent variable has no effect.

❸ 1. It is possible that Linus ran the experiment differently than Sean did, in terms of timing, directions, personal interactions, etc.

❹ Identify possible confounding variables resulting from differences in time or other environmental factors between different conditions.

1. Identify differences between conditions in surroundings or other environmental factors, especially those based on differences in when the different conditions took place.

❹ 1. It is possible that people tend to be more alert in the morning than at the end of the school day.

❺ Identify possible extraneous variables that are not likely to be confounding.

1. Identify factors that could cause people to get different results but that would not be more prevalent in one condition than another.

❺ 1. Some people may remember more words because they personally relate to a lot of them, regardless of which group they are in.

❻ Explain the possible effects of specific extraneous and confounding variables on a study.

1. If the sample is small, extraneous variables can have a big effect in a random direction because of the influence of unbalanced outliers.
2. If there is a confounding variable, the researchers cannot conclude that the independent variable affects the dependent variable, because any link between the two variables may be due to the confound instead.

❻ 1. Since they used two full classes of participants, one or two outliers will have minimal effect.
2. They will not be able to conclude that one modality is better than the other. For example, if the average score is higher in the visual condition, this may be because it truly is easier to remember visual information than auditory information (as hypothesized), but it could also be because the type of people who take Advanced Physics tend to have better memories in the first place, or because people tend to be more alert in the morning, or because Linus ran the experiment differently than Sean did.

A common confounding variable in social science research is SOCIOECONOMIC STATUS, which includes social and economic factors such as wealth, income, education, occupation, and residence.