CHAPTER SEVEN: RESEARCH METHODS

Review March 22 ℘ Test March 30

Designing a study and collecting data to test a hypothesis is more difficult than it seems. Furthermore, many conclusions that seem to be supported by the data are in fact unfounded. In particular, variables that are correlated with one another do not necessarily affect one another in the way that is assumed, if at all. Many relationships between variables are actually due to outside variables, called confounds, that simultaneously affect each variable being studied, and if these confounds are not identified, they are likely to lead to faulty conclusions.

7-A Samples and Populations — Monday • 3/6

population • sample • sample size • random selection • sampling bias • population parameter • sample statistic • proportion • mean • standard deviation • probability • statistics

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.

7-B Statistical Conclusions — Wednesday • 3/8

inferential statistics • hypothesis testing • null hypothesis • alternate hypothesis • one-tailed test • two-tailed test • reject the null • statistically significant • type I error • alpha • type II error • beta • power

❶ For a given prediction, identify the alternate hypothesis, the null hypothesis, and the number of tails.
❷ State the meaning of a type I error and a type II error for a test.
❸ Identify factors leading to uncertainties about conclusions from statistical inference.
❹ Calculate the standard deviation σ of a population.
❺ Estimate the standard deviation of a population based on sample data.

7-C Experiments and Quasi-Experiments — Monday • 3/13

independent variable • factor • level • experimental group • control group • placebo • dependent variable • operational definition • experimental design • random assignment • quasi-experimental design • correlational design • factorial design • causal relationship • affect • effect

❶ Identify an experiment or quasi-experiment's independent and dependent variables, and state the levels or operational definitions.
❷ Distinguish between experimental, quasi-experimental, and correlational research designs.
❸ Interpret significant results with or without causation as appropriate.
❹ Distinguish between *affect* and *effect* meaning *influence.*

7-D Confounds — Wednesday • 3/15

extraneous variable • random error • confounding variable • systematic error

❶ Identify possible confounding variables resulting from preexisting differences between participants in different conditions.
❷ Identify possible confounding variables resulting from to differences in the procedure between different conditions.
❸ Identify possible confounding variables resulting from differences in timing or other environmental factors between different conditions.
❹ Identify possible extraneous variables that are not likely to be confounding.
❺ Explain the possible effects of specific extraneous and confounding variables on a study.

7-E Between-Participants and Within-Participants Designs — Tuesday • 3/21

between-participants design • within-participants design • sequence effects • counterbalancing

❶ Outline how a given hypothesis could be tested with either a between-participants design or a within-participants design.
❷ Identify possible order effects with a given within-participants design.
❸ State how a given within-participants design could make use of counterbalancing, and discuss how effective this would be.
❹ Determine whether or not a within-participants design is appropriate.

7-A Samples and Populations

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 Brian surveys 20 random SVHS freshmen and 18 SVHS sophomores to ask whether or not they like having a weekly tutorial period.

1. The samples are the 20 random SVHS freshmen and the 18 SVHS sophomores.
2. The populations are SVHS freshmen and SVHS sophomores 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 still valid but 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 people can hold their breath, Heather times 20 students at break on how long they can hold their breath.

1. She is only testing SVHS teenagers. Compared to other schools and other age groups, SVHS teenagers may have longer times because they are young and tend to be healthy and unlikely to smoke, or shorter times because they are not full grown. Therefore, her sample may accurately represent only SVHS teenagers rather than people in general.
2. If Heather asked her friends, she is likely to have asked friends who are on her basketball team or are otherwise athletic, and who are not likely to use drugs or engage in activities that would decrease lung health. Therefore, her 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 a whole sample or population, resulting when every individual's response consists only of one of two predetermined possibilities, such as *yes* or *no*, or *male* or *female*.

A MEAN is an average, resulting when every individual's overall response consists of a number, such as *25 minutes* or *9 on a scale of 1 to 10*.

A STANDARD DEVIATION is a measure of variability, to quantify the extent to which values being averaged vary from one another. The lowest possible standard deviation is 0, representing no variation in the data such as in { 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 in { 8, 8, 12, 12 }.

Parameters and statistics use different symbols. Parameters usually use Greek letters, and statistics usually use Roman (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 calculated from data only 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 value shows how much variation there is in the responses, it is a standard deviation.

❸ Out of 25 juniors, 21 say they would rather go a week without a car than a week without a smartphone.

1. Only 25 juniors were surveyed. This is a sample, not all juniors.
2. Each participant only had two options—car or smartphone—so the result is a proportion.
   $\hat{p} = \frac{21}{25}$

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 lands on heads 10 times in a row, is it reasonable to conclude that $p \neq 50\%$?

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 known population parameters, such as the average IQ is 100 or the proportion of tails in coin flips is 50%, are being used to make predictions about specific outcomes, such as 10 random people having an IQ over 110 or a coin landing on tails 5 times out of 20, this is a probability problem.
   If known sample statistics, such as 10 random people having an IQ over 110 or a coin landing on tails 5 times out of 20, are being used to make a general conclusion about the population, such as average IQ is actually higher than 100 or the coin is weighted toward heads, this is a statistics problem.
   Probability is used to make predictions about specific outcomes based on known parameters.
   Statistics is used to make conclusions about unknown parameters based on specific outcomes.

❹ As of last week, President Trump had a 43% approval rating among 1500 American adults.

1. The known information is only for the 1500 American adults in the survey, not for the whole population of American adults.
2. The sample statistic $\hat{p} = 43\%$ is being used as an estimate for the proportion of all American adults who approve of Trump, so this is statistics.

7-B Statistical Conclusions

The goal of INFERENTIAL Statistics is to identify if a population parameter is different from an established value or from another given parameter. This is called HYPOTHESIS TESTING.

A NULL Hypothesis, abbreviated H_0, states (roughly speaking) that there is nothing to find with respect to the population parameter being studied. For example, "average body temperature upon wakening is not lower than 98.6°" or "boys and girls are equally likely to apply for college" are null hypotheses because they are stating that there is no difference to be found.

An ALTERNATE Hypothesis, abbreviated H_1, states that there is a difference to be found in the population. This is the prediction of the researcher.

A ONE-Tailed H_1 states that a given population parameter is different from a specified value or another given parameter in a specified direction. There are two possible outcomes from a one-tailed test: Either do or do not conclude that the population parameter is different from the specified value or other parameter in the predicted direction. For example, a one-tailed test to see if girls are heavier than boys on average would not find that boys are heavier than girls on average (even though this is clearly true), because this was not being tested.

A TWO-Tailed H_1 states that a given population parameter is different from a specified value or another given parameter in either direction. There are three possible outcomes from a two-tailed test: Conclude that the population parameter is higher than the specified value or other parameter, conclude that it is lower, or do not conclude that it is different.

Two-tailed tests are not common in scientific research because they require the researcher to simultaneously have two opposite predictions.

❶ For a given prediction, identify the alternate hypothesis, the null hypothesis, and the number of tails.

1. The alternate hypothesis is the prediction.
2. The null hypothesis is the complement of the alternate hypothesis, and typically can be stated by adding "not" to the alternate hypothesis.
3. Identify how many tails are to be used for the test:
   A one-tailed test is used when the researchers have justification for a prediction in a specific direction (specifically higher, or specifically lower).
   A two-tailed test is used when the researchers wish to consider the possibility of either direction (either higher or lower, without predicting either).

❶ Solange hypothesizes that doing an additional version of the chapter review for this class increases test scores.

1. H_1: Doing an additional version of the chapter review of this class increases test scores.
2. H_0: Doing an additional version of the chapter review of this class does not increase test scores.
3. She will do a one-tailed test. Although an additional review may not help, there is no reason to test if it would cause lower test scores.

In hypothesis testing, researchers can either REJECT or Fail to Reject the Null hypothesis. If the null hypothesis is rejected, the data are considered STATISTICALLY SIGNIFICANT.

Only the null hypothesis is tested, not the alternate hypothesis. Failing to reject the null hypothesis does not mean the alternate hypothesis is accepted; it simply means there is not enough evidence to claim otherwise. For example, if the first two people you meet in Brazil are very nice, this is not enough information for you to conclude that Brazilians are nicer than Americans, but this lack of information should not be interpreted as indicating that Brazilians are not nicer than Americans.

Researchers never know whether their statistical conclusion is in fact correct. If it were known, the experiment would not have been conducted in the first place.

A TYPE I ERROR is rejecting the null hypothesis when it is actually true. The probability of a type I error (given H_0 is true) is called α, the Greek letter ALPHA.

A TYPE II ERROR is not rejecting the null hypothesis when it is actually false. The probability of a type II error (given a specified H_1 is true) is called β, the Greek letter BETA.

The POWER of a Test, $1 - \beta$, is the probability of rejecting H_0 (given a specified H_1 is true).

Since statistics are values calculated from samples, not from populations, statistics can never *prove* anything about populations. Any findings could be coincidental. Do not use the word *prove* with statistics.

❷ State the meaning of a type I error and a type II error for a test.

1. In a type I error, researchers claim their alternate hypothesis is correct, but actually it is not, causing them to report a result that was actually a coincidence and thus make a false claim.
2. In a type II error, researchers do not claim their alternate hypothesis is incorrect but actually it is, causing them to not report their results and thus disregard a correct claim.

❷ Solange hypothesizes that doing an additional version of the chapter review for this class increases test scores.

1. Type I error: Doing an extra review does not increase test scores, but based on her data Solange concludes that it does.
2. Type II error: Doing an extra review does increase test scores, but based on her data Solange does not conclude that it does.

❸ Identify factors leading to uncertainties about conclusions from statistical inference.

1. If there is not much difference in the results, we have little evidence.
2. If the sample size is small, we less confident that it fairly represents the population.
3. If the standard deviation is high, this indicates that other samples may be very different from the one collected.

❸ Solange hypothesizes that doing an additional version of the chapter review for this class increases test scores.

1. if the test averages are similar, such as 83% for not doing the extra review and 84% for doing it
2. if the sample size is small, such as only four people doing the extra review and only four people not doing it
3. If there is a lot of variation within the groups, such as s_1 and s_2 being higher than 15%

In statistics, Σx means the sum of all x values.

Standard deviation is the square root of the average squared difference between each data value in the population and the population mean: $\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n}}$.

❹ Calculate the standard deviation σ of a population.

1. Find the population mean μ.
2. Subtract μ from each data value.
3. Square each difference.
4. Find the average of the squared differences.
5. Take the square root of the average.

❹ Four random scores from a test taken by hundreds of people are 80, 92, 71, and 93.

x	$x - \overline{x}$	$(x - \overline{x})^2$
80	-4	16
92	8	64
71	-13	169
93	9	81
$\Sigma x = 336$		$\Sigma(x - \overline{x})^2 = 330$
$\mu = 336 \div 4 = 84$		$\sigma = \sqrt{\frac{330}{4}} \approx 9.1$

Typically not all values in the population are known, making it impossible to calculate the standard deviation σ of the population. Instead it can be estimated by $\sigma \approx s = \sqrt{\frac{\Sigma(x-\mu)^2}{n-1}}$. Dividing by a smaller number yields a larger result, which accounts for additional variance in the population that is not in the sample.

❺ Estimate the standard deviation of a population based on sample data.

1. Follow the steps above to calculate σ, except divide by $n - 1$ instead of by n when finding the average squared difference in step 4.
2. The resulting value is the sample standard deviation s. This is not the standard deviation of the sample, but rather an estimate of the standard deviation of the population based on sample data.

❺ Four random test scores on a test were 80, 92, 71, and 93.

1. $\Sigma(x - \overline{x})^2 = 330$ (see ❹)
   $s = \sqrt{\frac{330}{4-1}} \approx 10.5$
2. Although 9.1 is the standard deviation of these four test scores (see ❹), the best estimate of the standard deviation of all the test scores is 10.5.

7-C Experiments and Quasi-Experiments

An INDEPENDENT Variable, or FACTOR, is a variable whose values are known (often chosen) before the values of the dependent variables are known.

The LEVELS of a Variable are the different possible conditions. In medical studies, the levels are commonly the EXPERIMENTAL Group, which takes the drug, and the CONTROL Group, which takes a fake drug called a PLACEBO.

A DEPENDENT Variable is a variable whose values are known (often measured) after the values of the independent variables are known. It is predicted to be related to the independent variables.

Not all variables have straightforward definitions. Many variables are conceptual in nature and must be assigned a specific, measurable definition, called an OPERATIONAL DEFINITION, in order to be used.

❶ Identify an experiment or quasi-experiment's independent and dependent variables, and state the levels or operational definitions.

1. An independent variable has two or more different conditions that exist before data are collected and that are being compared to each other. Often one is the control condition and the others are experimental conditions. Each condition is a level of the variable.
2. A dependent variable is a variable for which the researcher finds values by collecting data, often by measuring. For conceptual variables, an operational definition is needed to specify exactly what is meant by the variable and how it will be quantified or categorized.

❶ Shannon predicts that starting each day with five minutes of yoga increases math grades and also causes students to be more interested in school, especially among girls.

1. Amount of yoga (with two levels, 5 minutes per day and none) and sex (with two levels, male and female) are the independent variables.
2. Math grades and interest in school are the dependent variables. Possible operational definitions could be semester grade and agreement with the statement "I am interested in school" on a scale of 0 to 10.

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 sex or religion, not practical, such as wealth or school attended, or not ethical, such as drug use or exposure to toxins.

A CORRELATIONAL Design uses a numerical independent variable instead of a categorical independent variable. Correlational studies do not have distinct levels or random assignment.

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

❷ 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.

❷ Shannon predicts that starting each day with five minutes of yoga will increase math grades and also cause students to be more interested in school, especially among girls.

1. Shannon will randomly assign who does yoga and who does not, so this is experimental.
2. Shannon cannot randomly assign sex, so this is quasi-experimental.
3. There are two independent variables, so the study is using a factorial design. In this case, one factor is experimental and one is quasi-experimental.

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

❸ Interpret significant results with or without causation as appropriate.

1. For statistically signficant quasi-experimental or correlational results, 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. For statistically significant experimental results, 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.

❸ Wang and Kenny (2013) 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 cause of this correlation:

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

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

❹ 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. Only *affected* (not *effected*) means *influenced,* and only *affecting* (not *effecting*) means *influencing.*

❹ 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 preceeded 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

7-D Confounds

An EXTRANEOUS Variable is any variable, other than the independent 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 to 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.

Last year, I hypothesized that if I gave a surprise pizza party it would result in higher grades for the students. I brought in pizza, candy and drinks for everyone in fifth period Statistics & Research Methods but not for third period Statistics & Research Methods. The following week I compared scores on the chapter test between the two periods.

❶ Identify possible confounding variables resulting from 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.)

❶ There is an advanced English class third period, preventing a lot of stronger students from being in third period Statistics & Research Methods. Therefore, students in fifth period may have already been stronger academically.

❷ Identify possible confounding variables resulting from 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.

❷ Fifth period is the second time I teach the class for the day. It could be that I do a better job explaining to fifth period than to third period.

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

1. Identify differences between conditions in surroundings or other environmental factors, especially those based on differences in timing.

❸ Fifth period may have had an advantage on the test if students from third period told them questions on it during break.
Fifth period is later in the day, allowing more time for students to study and to be fully awake.

❹ 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.

❹ Students in both classes vary in terms of studiousness, motivation, intelligence, preparedness, etc.

❺ 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 the link between the two variables may be due to the confound instead.

❺

1. Each sample had almost 40 students, so outliers were probably fairly well balanced or evened out and extraneous variables were not likely to have a major effect.
2. The confounding variables made it so that even though fifth period did better than third period, as predicted, I was not able to conclude that the pizza party was the reason.

7-E Between-Participants and Within Participants Designs

In a BETWEEN-PARTICIPANTS Design, each participant takes part in one level of the independent variable.

In a WITHIN-PARTICIPANTS Design, each participant takes part in each level of the independent variable.

In the reverse Stroop test, participants are timed in matching 20 colored words to boxes of the same color, and, separately, in matching 20 colored words to boxes of the color stated.

❶ Outline how a given hypothesis could be tested with either a between-participants design or a within-participants design.

1. In a between-participants design, each participant takes part in a single condition.
2. In a within-participants design, each participant takes part in each condition.

❶

1. Between-participants design: Each participant takes part in either the color-match condition or the meaning-match condition, but not both.
2. Within-participants design: Each participant takes part in both conditions.

In some within-participants designs, taking part in one condition will influence the results of other conditions. These influences are called SEQUENCE EFFECTS. Within-participants designs are further limited because some independent variables, such as race or sex, cannot be manipulated by the researcher at all.

❷ Identify possible sequence effects with a given within-participants design.

1. Participants may act differently once they know what is involved in the experiment.
2. Participants may tend to improve after the first condition due to practice.
3. Participants may tend to do worse after the first condition due to fatigue, boredom, or interference.
4. Outside factors may change over time.

❷

2. Participants may be quicker in the second condition because they understand what to do and how to move quickly as appropriate.
3. Participants may be slower in the second condition because they have interference from the first condition, that is, they could get confused now that they have had one set of directions and moved to a different set.

Some sequence effects can be canceled by COUNTERBALANCING, in which some participants take part in one one condition first, and the rest take part in the other condition first. To be a true experiment rather than a quasi-experiment, a within-participants design must assign the order of conditions randomly.

❸ State how a given within-participants design could make use of counterbalancing, and discuss how effective this would be.

1. In a study with two conditions, counterbalancing can be achieved by having half the participants randomly assigned to take part in condition A first and the other half take part in condition B first.
2. Weak and moderate order effects due to practice, fatigue, boredom, interference, or outside factors will be practically eliminated.

❸

1. The reverse Stroop test randomly assigns each participant to either the color-match condition first or the meaning-match condition first.
2. There should no longer be significant concern that order effects will cause one condition to consistently be easier than the other.

Within-participants designs are more powerful than between-participants designs because they greatly reduce the effects of extraneous variables: A participant who is an outlier for an extraneous variable in one condition will likely be an outlier in the other conditions as well, making the outlying values cancel each other out. Therefore, within-participant designs are superior to between-participant designs in terms of statistical analysis. However, within-participants designs require participants to take part in each condition, which is not always possible, practical, or ethical. In addition, within-participants designs may be prone to order effects to an extent that counterbalancing cannot sufficiently compensate for.

❹ Determine whether or not a within-participants design is appropriate.

1. Identify whether or not the independent variable is possible to manipulate, and, if so, whether or not it is practical and ethical to do so. If not, do not use a within-participants design.
2. If participating in the first condition could help the participant do significantly better in the second condition, such as due to practice or finding out about the experiment, do not use a within-participants design.
3. If participating in the first condition could cause the participant to do significantly worse in the second condition, such as due to tiredness or boredom, do not use a within-participants design.
4. If neither of the above apply, a within-participants design should usually be used, especially if order of conditions can be randomly assigned.

❹ Matt hypothesizes that right-handed people can kick harder with their right leg than with their left leg.

1. Having participants kick with both legs is clearly not impossible, impractical, or unethical.
2. A single kick will not provide significant practice for a kick with the other leg.
3. A single kick will not cause significant disadvantage in any way for a kick with the other leg.
4. Matt should use a within-participants design, but he should randomly assign which leg each person starts with.