Glossary

**Systematic empiricism**

Definition: Empiricism refers to learning based on observation. Scientists learn about the world systematically, by carefully planning, making, recording and analyzing observations of it.

Resource (website) title: What is empiricism? – 60 seconds philosophy

Resource (website) link: [https://www.youtube.com/watch?v=FWiqrZhoZkQ](https://www.youtube.com/watch?v=FWiqrZhoZkQ)

Resource (website) description: An explanation of Empiricism as a position within epistemology, a field of philosophy.

**Empirical question**

Definition: Questions about the way the world actually is. Empirical questions can be answered through systematic observation.

Resource (website) title: Developing a research project

Resource (website) link: [https://www.youtube.com/watch?v=QauomrFcrXk](https://www.youtube.com/watch?v=QauomrFcrXk)

Resource (website) description: A video for students on how to develop a research question for their papers.
Public knowledge

Definition: One of the three fundamental features of science, along with systematic empiricism and empirical questions. Scientists publish their work, usually in professional journals. This way their results become public knowledge.

Resource (website) title: “Psychology in the public eye” and “Research is publicly funded - why isn't it publicly available?” – TED Talk by Erica Stone

Resource (website) link: http://www.apa.org/monitor/dec05/eye.aspx
https://www.ted.com/talks/erica_stone_academic_research_is_publicly_funded_why_isn_t_it_publicly_available

Resource (website) description: An interesting article that provides an idea of different aspects of public knowledge. The TED talk is a critical view on how public knowledge actually works in the US today and some problems that might exists. It also gives an overview of related aspects, the publication process etc.

Pseudoscience

Definition: Activities and beliefs that are claimed to be scientific by their proponents – and may appear to be scientific at first glance – but are not.

Resource (website) title: Karl Popper, Science and Pseudoscience - Crash Course Philosophy

Resource (website) link: https://www.youtube.com/watch?v=-X8Xf0JdTQ

Resource (website) description: An interesting video (maybe some debatable points) that uses the theories of Freud and Einstein to illustrate the concept of pseudoscience.

Falsifiability

Definition: All scientific claims must be falsifiable. This means that they need to be expressed in such a way that there are observations that would – if they were made – count as evidence against the claim.

Resource (website) title: Karl Popper’s Falsification

Resource (website) link: https://www.youtube.com/watch?v=wf-sGqBsWv4

Resource (website) description: Another short overview of Popper’s philosophy this time using Marxism as an example of a Pseudoscience.

Doctor of Philosophy

Definition: Also called PhD, a doctoral degree that enables one to conduct research. Psychological research is usually conducted by people with PhDs in Psychology.

Resource (website) title: “The animated guide to a PhD” - YouTube
Basic research

Definition: Research that is conducted primarily for the sake of achieving a more detailed and accurate understanding of human behavior, without necessarily trying to address any particular practical problem.

Resource (website) title: Current issues and new directions in *Psychology and Health*: Bringing basic and applied research together to address underlying mechanisms.


Resource (website) description: This article explains how important it is for both basic and applied research to be conducted, especially in clinical and health-related fields of psychology.

Applied research

Definition: Research conducted in order to address a particular problem (also see basic research).

Resource (website) title: Journal of Applied Psychology


Resource (website) description: To get an idea of how wide the field of applied psychology really is, take a look this APA journal and maybe read a few of the articles available online.

Folk psychology

Definition: Common beliefs about people’s behavior, thoughts and feelings that are based on common sense and intuition. Many of these beliefs run contrary to scientific evidence.

Resource (website) title: “Folk Psychology and Criminal Law: Why we need to replace folk psychology with behavioral science”

Resource (website) link: [http://proxygsu-gsw1.galileo.usg.edu...ive&scope=site](http://proxygsu-gsw1.galileo.usg.edu...ive&scope=site)

Resource (website) description: This article demonstrates how common the use of “folk psychology” still is in various fields of applied psychology such as the criminal justice field. It advocates skepticism towards “common sense” notions on why people commit crimes. The author also offers an explanation of how some “folk psychology” may originate.

Confirmation bias

Definition: A tendency to focus on the evidence that seems to prove our already held beliefs while discounting
evidence that would disprove our own beliefs.

Resource (website) title: New York Times: “You’re not going to change your mind.” And “Why it’s so hard to admit you’re wrong.”


Resource (website) description: These fairly recent (and hopefully unbiased) articles look critically at confirmation bias and the relevance this concept might have gained in today’s political debates.

Skepticism

Definition: Questioning claims, considering the alternatives and searching for evidence in every direction, not just supporting evidence (also see confirmation bias). Focusing on systematically collected empirical evidence.

Resource (website) title: “Skeptics and scepticism”

Resource (website) link: https://www.theguardian.com/science/the-h-word/2012/nov/13/history-science

Resource (website) description: What is true skepticism? A short article from the guardian that considers the differences between skepticism that looks for empirical evidence and other forms of skepticism that are not based on evidence.

Tolerance for uncertainty

Definition: Since there is often not enough evidence to fully evaluate a belief or claim, scientists need to cultivate a tolerance for uncertainty and accept that there are many things they don’t know (yet).

Resource (website) title: Ambiguity tolerance in organizations: definitional clarification and perspectives on future research

Resource (website) link: https://doi.org/10.3389/fpsyg.2015.00344

Resource (website) description: You can download the article under this link. It looks at uncertainty (or ambiguity) tolerance from various perspectives. Uncertainty tolerance is not just something researchers need to work successfully, it is also an important topic in mental health and organizational psychology.

Clinical practice of psychology

Definition: The application of scientific research for the diagnosis and treatment of psychological disorders and related problems.

Resource (website) title: “A Career in Clinical or Counseling Psychology” and “What you need to know to get licensed”

Resource (website) description: Anyone interested in a career as a clinical or counseling psychologist might benefit from reading these. The APA has information on its website on what kinds of jobs you can do as a clinical psychologist, how to get there and how to become a licensed psychologist.

**Empirically supported treatment**

Definition: A treatment that has been studied scientifically and shown to result in greater improvement than no treatment, a placebo, or some alternative treatment.

Resource (website) title: How do I find a good therapist?
Resource (website) description: The APA recommends clients to only use therapists who are familiar with and use evidence-based treatment.

---

**Topic: Getting Started in Research**

**Variable**

Definition: A quantity or quality that varies across people or situations. Research questions in psychology are about variables.

Resource (website) title: Experiments Explained: Clear and Simple! Learn the Basics
Resource (website) link: https://www.youtube.com/watch?v=tK2mBsSb3uw
Resource (website) description: This video explains the basic way experiments work. It uses a biological experiment on plants as an example.

**Quantitative variable**

Definition: A quality, such as height, that is typically measured by assigning a number to each individual.

Resource (website) title: Episode 3: Identifying Qualitative and Quantitative Variables
Resource (website) link: https://www.youtube.com/watch?v=4K9PYYojiA4c
Resource (website) description: This is a video from a business statistics class. It compares qualitative (or categorical) to quantitative variables.
Categorical variable

Definition: A quality, such as sex, occupation or nationality. Typically measured by assigning a category label to each individual.

Resource (website) title: Statistics 101: Describing a Categorical Variable

Resource (website) link: https://www.youtube.com/watch?v=vrWYw8d2830&list=PLIeGtxpvyG-K82r1fgL-DO1xKg133PXKh

Resource (website) description: This is a great youtube channel (Brandon Foltz) with a lot of videos on statistics that might be helpful to students.

Population

Definition: A large group of people, such as American teenagers, professional athletes or simply everyone. Researchers draw conclusions about the population they are interested in by studying a smaller sample of that population.

Resource (website) title: US census: Population Clock

Resource (website) link: https://www.census.gov/popclock/

Resource (website) description: Populations constantly change. You could study the whole world's population or only American females between the ages of 20 and 40. There are endless possibilities of defining a population to be studied.

Sample

Definition: A smaller subset of a population that is easier to study than the whole population but allows the researcher to generalize their findings to the population.

Resource (website) title: Brett Hennig: “What is we replaced politicians with randomly selected people?”

Resource (website) link: https://www.ted.com/talks/brett_hennig_what_if_we_replaced_politicians_with_randomly_selected_people

Resource (website) description: This guy is not a psychologist, but he talks about the idea of a representative random sample of the population. The fact that he talks about this in the context of politics might be a bit far out but is quite interesting.

Operational definition

Definition: A definition of the variable in terms of precisely how it is to be measured.

Resource (website) title: Psychological Research - Crash Course Psychology #2

Resource (website) link: https://www.youtube.com/watch?v=hFV71QPvX2I
Statistical relationship

Definition: There is a statistical relationship between two variables when the average score on one differs systematically across the levels of the other. The two forms of statistical relationships are differences between groups and correlations between quantitative variables.

Resource (website) title: Statistics 101: Understanding Covariance

Resource (website) link: https://www.youtube.com/watch?v=xGbpuFNR1ME&index=6&list=PL1eGxpvyG-JMH5fGDWhNiyET88Mexcw

Resource (website) description: This video is also from Brandon Foltz’s youtube channel (see categorical variable) and focuses on explaining covariance and correlation.

Bar graph

Definition: Information about differences between groups is best presented in a bar graph, where the heights of the bars represent the group means.

Resource (website) title: Reading bar graph examples | Measurement and data | Early Math | Khan Academy

Resource (website) link: https://www.youtube.com/watch?v=OmLl6pkvV-I

Resource (website) description: A simple demonstration of how a bar graph works. The video is for kids but could be a good review.

Scatterplot

Definition: Correlations between quantitative variables are often presented using scatterplots. Each axis represents one of the variables. Each point represents one person’s score on both variables. Taking all the points into account we can see which relationship exists between the two variables.

Resource (website) title: Bivariate relationship linearity, strength and direction

Resource (website) link: https://www.youtube.com/watch?v=30LcZqRfPRY

Resource (website) description: Another Khan Academy video that shows different scatterplots and explains different statistical relationships.

Positive relationship

Definition: In a positive relationship higher scores on one variable tend to be associated with higher scores on the other.

Resource (website) title: Statistics 101: Understanding Correlation – Brandon Foltz
Negative relationship

Definition: Higher scores on one variable tend to be associated with lower scores on the other.

Resource: refer to Scatterplot, positive relationship and Pearson’s r video

Concept: Pearson’s r

Definition: The statistic that is typically used to measure the strength of a correlation between quantitative variables. $r = 0$ means no correlation, $r = -1$ stands for the strongest possible negative relationship, $r = 1$ for the strongest possible positive relationship.

Independent variable

Definition: When there is a causal relationship between two variables, the variable that is thought to be the cause is called the independent variable (often called X).

Resource (website) title: Independent and dependent variables - Intro to Psychology

Resource (website) link: https://www.youtube.com/watch?v=DaEvBlrvTmw

Resource (website) description: The video explains how experimental studies work and uses a mnemonic to explain the difference between the independent and the dependent variable.

Dependent variable

Definition: The variable (often called Y for short) that is thought to be caused by the independent variable.

Resource (website): see videos on variables and independent variables

Directionality problem

Definition: Two variables, X and Y, can be statistically related because X causes Y or because Y causes X. Without experimental evidence there is no way to tell what causes what.

Resource (website) title: Correlation and causality | Statistical studies | Probability and Statistics | Khan Academy

Resource (website) link: https://www.youtube.com/watch?v=ROpbdO-gRUo
Resource (website) description: The video first explains the difference between causation and correlation, then illustrates the directionality and the third-variable problems. It uses an article on web-md as an example of how a study can be misinterpreted to imply causation.

Third-variable problem

Definition: Two variables, X and Y, can be statistically related not because X causes Y, or Y causes X, but because a third variable, Z, causes both X and Y.

Resource (website) title: Correlation and causality | Statistical studies | Probability and Statistics | Khan Academy

Resource (website) link: https://www.youtube.com/watch?v=ROpdbO-gRVo

Resource (website) description: see directionality problem

Experiment

Definition: A study in which the researcher manipulates the independent variable. Most effective way to address the directionality and the third-variable problem.

Resource (website) title: Psychological Research - Crash Course Psychology #2

Resource (website) link: https://www.youtube.com/watch?v=hFV71QPvX2I&t=284s

Resource (website) description: This is a more entertaining, less serious video than others which uses some strange examples. It talks about psychological research in general, the differences between causation and correlation and of course, experiments.

Interestingness

Definition: Is a research question interesting to the general public and the scientific community rather than just the researcher herself? This is the case when the answer is in doubt, has important practical implications and fills a gap in the research literature.

Resource (website) title: Evaluating research questions

Resource (website) link: https://course.oeru.org/research-methods/modules-1-3/module-1-introduction/evaluating-research-questions/

Resource (website) description: What makes a research question interesting as well as feasible?

Feasibility

Definition: The feasibility of successfully answering the question is an important criterion for evaluating research questions. Feasibility is impacted by the time, money, equipment and expertise available. Basically, an experiment that requires a space ship or millions of participants would be very unfeasible.

Resource: see interestingness

Powered by
Professional journals

Definition: Periodicals that publish original research articles. They are usually published monthly or quarterly, with each issue containing several articles.

Resource (website) title: Browse journals by title


Resource (website) description: There is a huge diversity of professional journals in Psychology. This website can give you a first impression of what's out there.

Empirical research report

Definition: One of the two basic types of articles that are typically published in professional journals. They describe one or more new empirical studies conducted by the author. They introduce a research question, explain why it is interesting, review previous research, describe their method and results and draw their conclusions.

Resource (website) title:
Mental condition and specificity of mental disorders in a group of workers from southern Poland: A research report


Resource (website) description: You can pick any psychological topic (for example a group of clinical disorders such as psychotic disorders) and find an example of each type of scholarly article about it. The link above is an empirical research report, the following links will be other types of articles or books also dealing with the topic of psychosis.

Review article

Definition: Along with research reports, one of the two basic types of articles typically published in professional journals. They summarize previously published research on a topic and usually present new ways to organize or explain the results.


Resource (website) description: This review summarizes the literature that points towards adverse life events such as trauma, loss and stress as being some of the primary causes of psychosis. It is a typical review article as they commonly appear in scholarly journals. It summarizes the evidence and draws its own conclusions that contradict some traditionally held views, for example the idea that schizophrenia is a genetically-caused neurological disease.

Theoretical article

Definition: A type of review article that is primarily devoted to presenting a new theory is often called a theoretical article.
Resource (website) title: “Psychosis, Trauma and Ordinary Mental Life”


Resource (website) description: An interesting article from the same 2016 issue of the American Journal of Psychotherapy that the review article was found in. Apart from the interesting content it can serve as an example of a theoretical article in a professional journal. The theory presented here is heavily influenced by psychoanalysis so may not provide the most empirical example of all but does present a reasonable theory nonetheless.

Scholarly book

Definition: Books written by researchers and practitioners mainly for use by other researchers and practitioners.

Resource (website) title: How to find scholarly books

Resource (website) link: https://www.youtube.com/watch?v=Z0fBYNW32pE

Resource (website) description: This is really helpful because it explains how to determine if a book is a scholarly source or not.

Monograph

Definition: A type of scholarly book that is written by a single author or a small group of authors. Usually gives a coherent presentation of a topic much like an extended review article.

Resource (website) title: What is a Monograph?

Resource (website) link: https://www.youtube.com/watch?v=JEip9_ulE18

Resource (website) description: A publisher explains what a monograph is. He specifically distinguishes monographs from textbooks.

Edited volumes

Definition: Another type of scholarly book. Edited volumes have an editor or a small group of editors who recruit many authors to write separate chapters on different aspects of the same topic. It is not unusual for each chapter to take a different perspective or even for the authors to disagree with each other.

PsycINFO

Definition: A comprehensive database produced by the APA that covers the research literature in Psychology. It is available in most university libraries and covers thousands of professional journals and scholarly books going back more than 100 years.

Resource (website) title: PsycINFO

Resource (website) description: Take a closer look at the millions of items PsycINFO has to offer.

Topic: Research Ethics

Ethics

Resource (website) title: “Ethics in Psychology research” and “Trust in research -- the ethics of knowledge production | Garry Gray | TEDxVictoria”

Resource (website) link: https://www.youtube.com/watch?v=PkaJbMH6ZpQ

https://www.youtube.com/watch?v=JSV4VZ8gdUQ

Resource (website) description: This is a talk given by a professor from India about Research ethics in Psychology. Her international perspective on the subject might be valuable. The second video is a TEDx talk that offers a critical view of ethics in contemporary science.

Confederate

Resource (website) title: “The dangers of using study confederates”

Resource (website) link: https://psmag.com/social-justice/youre-a-bad-actor-phd-student

Resource (website) description: In this article, Rick Paulas offers some critical thought on the use of confederates in Psychology experiments.

Autonomy

Resource (website) title: Research Ethics

Resource (website) link: https://depts.washington.edu/bioethx/topics/resrch.html

Resource (website) description: On the website of the University of Washington School of Medicine you can read about autonomy and other ethical principles of research ethics. The site also gives a quick review of the important historical ethics codes such as the Nuremberg code and the Belmont Report.

Informed consent

Resource (website) title: “What is confirmed consent. Adult”

Resource (website) link: https://www.youtube.com/watch?v=OrAA9Lq9xKU

Resource (website) description: This video goes through all the steps of the informed consent process.

Privacy

Resource (website) title: Privacy and Confidentiality

Resource (website) description: The University of California, Irvine, has some information on privacy and confidentiality in research on its websites.

Confidentiality

Resource: see privacy

Nuremberg code

Resource (website) title: “The Nuremberg code” and
“The Nuremberg Trials | PBS American Experience | Documentary 2018”

https://www.youtube.com/watch?v=k-b14ZyU1vl

Resource (website) description: Read the full (1 page) Nuremberg code here.
On youtube you can find a reasonably accurate PBS documentary to provide some historical context.

Declaration of Helsinki

Resource (website) title: Ethical Principles for Medical Research involving Human Subjects

Resource (website) link: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Resource (website) description: Take a look at the full declaration of Helsinki here.

Protocol

Resource (website) title: Recommended Format for a Research Protocol

Resource (website) link: http://www.who.int/rpc/research_ethics/format_rp/en/

Resource (website) description: The WHO (World Health Organization) recommends a certain format to be used for research protocol. This link can give an idea of what belongs in a research protocol.

Belmont Report

Resource (website) title: Read the Belmont Report

Federal Policy for the Protection of Human Subjects

Resource (website) title: Basic HHS Policy for Protection of Human Research Subjects


Resource (website) description: Read the full policy on the website of the US department for Health & Human services.

Institutional review board (IRB)

Resource (website) title: “Balancing risk and research” and


Resource (website) description: An article on the challenges of maintaining the necessary standards to protect subjects without stifling research.

Exempt research

Resource (website) title: Researchers and regulators hammer out guidelines for research risk

Resource (website) link: http://www.apa.org/monitor/julaug05/risk.aspx

Resource (website) description: The APA has official guidelines on what constitutes at-risk, minimal risk or exempt research.

Minimal risk research

Resource (website) title: Researchers and regulators hammer out guidelines for research risk

Resource (website) link: http://www.apa.org/monitor/julaug05/risk.aspx

Resource (website) description: see exempt research

At-risk research

Resource: see exempt research. For examples of what is considered at-risk research, review experiments like the Stanford Prison Experiment, or Stanley Milgram’s obedience study (see deception), which would most likely not be approved by any IRB today.

APA Ethics Code

Resource (website) title: The APA code of ethics


Resource (website) description: Get the full ethics code and related articles under this link.
Consent form

Resource (website) title: Informed Consent Form (sample)

Resource (website) link: https://psychology.illinoisstate.edu/Research/files/informedconsent.pdf

Resource (website) description: This is the template that Illinois State University recommends researchers use for their studies.

Deception

Resource (website) title: “Milgram Obedience Study”

Resource (website) link: https://www.youtube.com/watch?v=W147ybOdgpE

Resource (website) description: Psychology’s most infamous example of deception in an experiment. The video may be a bit disturbing for some as it contains original footage of the experiment.

Debriefing

Resource (website) title: ETHICS ROUNDS: Reading the Ethics Code more deeply


Resource (website) description: Read this article also regarding the APA ethics code and deception. It goes into some interesting details regarding research ethics and explains the principles behind the APA ethics code.

Prescreening

$2

---

**Topic: Theory in Psychology**

**Phenomenon**

A phenomenon is a general result that has been observed reliably in systematic empirical research. In essence, it is an established answer to a research question.

**Replication**

To replicate a study means conducting it again – either exactly as it was originally conducted or with modifications – to make sure that it produces the same results.

Resource link: https://www.youtube.com/watch?v=lvw4HBl8Lyo

Resource title: Science 101: The Basics of Reproducibility/Replicability

Resource description: Brian Nosek, a psychology professor at the University of Virginia explains the terms
replicability and reproducibility.

Theory

A theory is a coherent explanation or interpretation of one or more phenomena. Although theories can take a variety of forms, one thing they have in common is that they go beyond the phenomena they explain by including variables, structures, processes, functions, or organizing principles that have not been observed directly.

Resource link: https://www.youtube.com/watch?v=XsmPqtBgKOY

Resource title: Scientific Literacy: Theory vs Hypothesis

Resource description: This video helps to understand the difference between theories and hypotheses in under two minutes.

Perspective

More general than a theory, a perspective is a broad approach to explaining and interpreting phenomena.

Model

A model is a precise explanation or interpretation of a specific phenomenon. It is often expressed in terms of equations, computer programs, or biological structures and processes.

Hypothesis

This term most commonly refers to a prediction about a new phenomenon based on a theory.

Resource link: https://www.youtube.com/watch?v=DRDhDo2dDtq

Resource title: Hypothesis vs Theory – Intro to Psychology

Resource description: This is another really short (30 seconds) video on hypotheses and theories. It’s still worth watching as it explains things differently from the previous video.

Parsimony

The principle of parsimony holds that a theory should include only as many concepts as are necessary to explain or interpret the phenomena of interest. More parsimonious theories organize phenomena more efficiently than more complex, less parsimonious theories. Scientists generally follow this principle when choosing between theories.

Resource link: https://www.youtube.com/watch?v=YQ1qvuioRlk

Resource title: What is Occam's Razor? (Law of Parsimony!)

Resource description: This video actually explains parsimony in the context of physics and astronomy. It provides some perspective on the importance of parsimony in all science, not just psychology. It uses as an example, how the heliocentric model of the solar system was adopted in place of the geocentric model.
Formality

The extent to which the components of the theory and the relationships among them are specified clearly and in detail. Psychological theories vary widely in their formality.

Scope

The number and diversity of the phenomena a given psychological theory attempts to explain or interpret. Psychological theories vary widely in their scope.

Theoretical approach

In addition to varying in formality and scope, psychological theories vary in the kinds of theoretical approaches they are constructed from. Some approaches include, functional or mechanistic theories, state theories and typologies.

Functional theory

Functional theories explain psychological phenomena in terms of their function or purpose.

Mechanistic theory

Mechanistic theories focus on specific variables, structures, and processes, and how they interact to produce the phenomena being studied.

Resource link: https://www.youtube.com/watch?v=wP3Sa_evVxU

Resource title: Episode 6 Descartes

Resource description: This is part of a video series mentioned above on “The Philosophical Roots of Psychology”. The videos are all interesting to watch (and use very nice music) but this one specifically focuses on mechanistic theories.

State theory

Stage theories identify a series of stages that people pass through as they develop or adapt to their environment. Famous theories of this type include Abraham Maslow’s hierarchy of needs and Jean Piaget’s theory of cognitive development.

Resource link: https://www.youtube.com/watch?v=Jt3-PIC2nCs

Resource title: Piaget's stages of cognitive development | Processing the Environment | MCAT | Khan Academy

Resource link: https://www.youtube.com/watch?v=aYCBdZLCDBQ

Resource title: 8 Stages of Development by Erik Erikson

Resource description: The videos above represent two examples of state theories; the developmental theories of Jean Piaget and Erik Erikson.
Typology

Typologies provide organization by categorizing people or behavior into distinct types. Examples include theories categorizing types of emotions, personalities or intelligence.

Hypothetico-deductive method

This is the primary way that scientific researchers use theories. They begin with a set of phenomena and either construct a theory to explain or interpret them or they choose an existing theory to work with. They then generate a hypothesis that should be confirmed if that theory is true. They conduct an empirical study to test their hypothesis and reevaluate the theory in light of the results.

Resource link: https://www.youtube.com/watch?v=ATU27c4crq4

Resource title: Scientific Methodology: The hypothetico-deductive method

Resource description: A fairly challenging half-hour lecture on the hypothetico-deductive method, it's historical origins and surrounding concepts.

---

Topic: Psychological Measurement

Measurement

The assignment of scores to individuals so that the scores represent some characteristic of the individuals.

Psychometrics

Psychological measurement is called psychometrics. Psychological measurement can be achieved in a wide variety of ways, including self-report, behavioral and physiological measures.

Resource link: https://www.youtube.com/watch?v=8Qv0L76iCl0&index=1&list=PLfJYfW-oDYY3j-jM4r6L2KK_q_QGKvj-W

Resource title: Lecture 1a: introduction, uses of testing

Resource description: If anyone is feeling particularly ambitious, there is an entire North Dakota State University class on Psychometrics available on youtube. The introduction at least is worth watching (you can skip ahead to minute 22).

Construct

Psychological constructs (pronounced CON-structs) are variables that cannot be observed directly. One reason is that they often represent tendencies to think, feel, or act in certain ways. This includes personality traits, emotional states, attitudes, and abilities. An important goal of scientific research is to conceptually define psychological constructs in ways that accurately describe them.
Conceptual definition

The conceptual definition of a psychological construct describes the behaviors and internal processes that make up that construct, along with how it relates to other variables.

Resource link: https://www.youtube.com/watch?v=8HyshxsdRb4

Resource title: Abraham Feinberg: Research Methods – Chapter 3 – Conceptual definitions

Resource description: This is the first of several videos by Abraham Feinberg used here. Also watch the videos on operational definitions, levels of measurement, reliability and validity.

Operational definition

An operational definition is a definition of a variable in terms of precisely how it is to be measured. These measures generally fall into one of three broad categories. Self-report measures, behavioral measures and physiological measures.

Resource link: https://www.youtube.com/watch?v=7aHZrXvHsuw&t=24s

Resource title: Abraham Feinberg: Research Methods – Chapter 3 – Operational definitions

Self-report measure

Participants report on their own thoughts, feelings, and actions, as with the Rosenberg Self-Esteem Scale.

Behavioral measure

Some aspect of participants’ behavior is observed and recorded. This is an extremely broad category that includes the observation of people’s behavior both in highly structured laboratory tasks and in more natural settings.

Physiological measure

Physiological measures involve recording any of a wide variety of physiological processes, including heart rate and blood pressure, galvanic skin response, hormone levels, and electrical activity and blood flow in the brain.

Converging operations

When psychologists use multiple operational definitions of the same construct—either within a study or across studies—they are using converging operations. The idea is that the various operational definitions are “converging” on the same construct. When scores based on several different operational definitions are closely related to each other and produce similar patterns of results, this constitutes good evidence that the construct is being measured effectively and that it is useful.

Levels of measurement

S.S. Stevens suggested four different levels of measurement (which he called “scales of measurement”): The nominal (qualitative), ordinal, interval and ratio levels of measurement.
Resource title: Research Methods - Chapter 03 - Levels of Measurement 1/2 and 2/2

Resource description: The first video goes over the nominal and ordinal levels of measurement; the second video covers the other two levels.

Nominal level

The nominal level of measurement is used for categorical variables and involves assigning scores that are category labels. Category labels communicate whether any two individuals are the same or different in terms of the variable being measured. This is the only level of measurement reserved for qualitative data, whereas the remaining three represent quantitative information.

Ordinal level

The ordinal level of measurement involves assigning scores so that they represent the rank order of the individuals. Ranks communicate not only whether any two individuals are the same or different in terms of the variable being measured but also whether one individual is higher or lower on that variable.

Interval level

The interval level of measurement involves assigning scores so that they represent the precise magnitude of the difference between individuals, but a score of zero does not actually represent the complete absence of the characteristic (e.g. temperature).

Ratio level

Finally, the ratio level of measurement involves assigning scores in such a way that there is a true zero point that represents the complete absence of the quantity. Height measured in meters and weight measured in kilograms are good examples.

Reliability

Reliability refers to the consistency of a measure. Psychologists consider three types of consistency: over time (test-retest reliability), across items (internal consistency), and across different researchers (interrater reliability).

Test-retest reliability

When researchers measure a construct that they assume to be consistent across time, then the scores they obtain should also be consistent across time. Test-retest reliability is the extent to which this is the case. For example, intelligence is generally thought to be consistent across time.
Test-retest correlation

Assessing test-retest reliability requires using the measure on a group of people at one time, using it again on the same group of people at a later time, and then looking at test-retest correlation between the two sets of scores. This is typically done by graphing the data in a scatterplot and computing Pearson’s $r$.

Resource link: https://www.youtube.com/watch?v=Z6uAF79KPbo

Internal consistency

A second kind of reliability is internal consistency, which is the consistency of people’s responses across the items on a multiple-item measure. In general, all the items on such measures are supposed to reflect the same underlying construct, so people’s scores on those items should be correlated with each other.

Split-half correlation

Like test-retest reliability, internal consistency can only be assessed by collecting and analyzing data. One approach is to look at a split-half correlation. This involves splitting the items into two sets, such as the first and second halves of the items or the even- and odd-numbered items. Then a score is computed for each set of items, and the relationship between the two sets of scores is examined.

Cronbach’s alpha

Perhaps the most common measure of internal consistency used by researchers in psychology is a statistic called Cronbach’s $\alpha$ (the Greek letter alpha). Conceptually, $\alpha$ is the mean of all possible split-half correlations for a set of items.

Interrater reliability

Many behavioral measures involve significant judgment on the part of an observer or a rater. Interrater reliability is the extent to which different observers are consistent in their judgments.

Cohen’s kappa

Interrater reliability is often assessed using Cronbach’s $\alpha$ when the judgments are quantitative, or an analogous statistic called Cohen’s $\kappa$ (the Greek letter kappa) when they are categorical.

Resource link: https://www.youtube.com/watch?v=LYuogmzs4

Validity

Validity is the extent to which the scores from a measure represent the variable they are intended to. Validity is a judgment based on various types of evidence. The relevant evidence includes the measure’s reliability, whether it
covers the construct of interest, and whether the scores it produces are correlated with other variables they are expected to be correlated with and not correlated with variables that are conceptually distinct.

Resource link: https://www.youtube.com/watch?v=7fiCycqMFug&list=PLLfSdO6kkX-YxoedI5fCQgDUIe3DSW-z&index=26

Resource title: Research Methods - Chapter 03 - Validity (1/5)

Resource description: The link leads you to a playlist of videos starting with the video on validity in general. Keep watching through the next four videos on the different types of validity.

Face validity

Face validity is the extent to which a measurement method appears "on its face" to measure the construct of interest. Most people would expect a self-esteem questionnaire to include items about whether they see themselves as a person of worth and whether they think they have good qualities. So, a questionnaire that included these kinds of items would have good face validity.

Content validity

Content validity is the extent to which a measure “covers” the construct of interest. For example, if a researcher conceptually defines test anxiety as involving both sympathetic nervous system activation (leading to nervous feelings) and negative thoughts, then his measure of test anxiety should include items about both nervous feelings and negative thoughts.

Criterion validity

Criterion validity is the extent to which people’s scores on a measure are correlated with other variables (known as criteria) that one would expect them to be correlated with. For example, people’s scores on a new measure of test anxiety should be negatively correlated with their performance on an important school exam.

Criterion

A criterion can be any variable that one has reason to think should be correlated with the construct being measured, and there will usually be many of them. For example, one would expect test anxiety scores to be negatively correlated with exam performance and course grades and positively correlated with general anxiety and with blood pressure during an exam.

Discriminant validity

Discriminant validity is the extent to which scores on a measure are not correlated with measures of variables that are conceptually distinct. For example, self-esteem is a general attitude toward the self that is fairly stable over time. It is not the same as mood, which is how good or bad one happens to be feeling right now. So, people’s scores on a new measure of self-esteem should not be very highly correlated with their moods. If the new measure of self-esteem was highly correlated with a measure of mood, it could be argued that the new measure is not really measuring self-esteem; it is measuring mood instead.
People can react in a variety of ways to being measured that reduce the reliability and validity of the scores. Although some disagreeable participants might intentionally respond in ways meant to “mess up” a study, participant reactivity is more likely to take the opposite form. Agreeable participants might respond in ways they believe they are expected to.

Socially desirable responding

Participants might engage in socially desirable responding. They might give certain answers not because they really feel this way but because they believe this is the socially appropriate response or they do not want to look bad in the eyes of the researcher.

Demand characteristics

Research studies can have built-in demand characteristics: cues to how the researcher expects participants to behave. These expectations can bias participants’ behaviors in unintended ways and lead to socially desirable responding.

Topic: Experimental Research

Experiment

An experiment is a type of study designed specifically to answer the question of whether there is a causal relationship between two variables. Experiments feature the manipulation of an independent variable, the measurement of a dependent variable, and control of extraneous variables. The different levels of the independent variable are called conditions.

Internal validity

Studies are high in internal validity to the extent that the way they are conducted supports the conclusion that the independent variable caused any observed differences in the dependent variable. Experiments are generally high in internal validity because of the manipulation of the independent variable and control of extraneous variables.
External validity

Studies are high in external validity to the extent that the result can be generalized to people and situations beyond those actually studied. Generally, studies are higher in external validity when the participants and the situation studied are similar to those that the researchers want to generalize to.

Resource link: https://www.youtube.com/watch?v=4NQHeI8GD54

Resource title: Reliability, validity, generalizability and credibility. Pt. 1 of 3: Research Quality

Resource description: A full lecture on these topics. A good overview/review of what constitutes quality research.

Field experiment

Experiments conducted entirely outside the laboratory – in “the field”.

Manipulation

To manipulate an independent variable means to change its level systematically so that different groups of participants are exposed to different levels of that variable, or the same group of participants is exposed to different levels at different times.

Manipulation check

A manipulation check is a separate measure of the construct the researcher is trying to manipulate.

Extraneous variable

An extraneous variable is anything that varies in the context of a study other than the independent and dependent variables. Extraneous variables pose a problem because many of them are likely to have some effect on the dependent variable. This can make it difficult to separate the effect of the independent variable from the effects of the extraneous variables, which is why it is important to control extraneous variables by holding them constant.

Confounding variable

One way that extraneous variables can make it difficult to detect the effect of the independent variable is by becoming confounding variables. A confounding variable is an extraneous variable that differs on average across levels of the independent variable.

Resource link: https://www.youtube.com/watch?v=Qe7wN8qT8ic

Resource title: Research Methods: Extraneous and Confounding Variables

Resource Description: A short, 2-minute explanation

Resource link: https://www.youtube.com/watch?v=KfRQ7CYbGHM

Resource title: Research methods - Chapter 06 - Extraneous and confounding variables
Between-subjects experiment

In a between-subjects experiment, each participant is tested in only one condition.

Within-subjects experiment

In a within-subjects experiment, each participant is tested under all conditions. This leads to less noise in the date but runs a risk of carryover effects.

Resource link: https://www.youtube.com/watch?v=d-25eGNNCLq

Resource title: Between and Within Subject Designs – Brooke Miller

Resource description: A very helpful, simple 4-minute explanation. It uses an example of an experiment on the effect of smoking on lung capacity.

Random assignment

The primary way that researchers accomplish this control of extraneous variables across conditions is called random assignment, which means using a random process to decide which participants are tested in which conditions.

Resource link: https://www.youtube.com/watch?v=ULmR6ZbsGTs

Resource title: Randomization in Clinical Trials - University of Miami


Block randomization

One problem with coin flipping and other strict procedures for random assignment is that they are likely to result in unequal sample sizes in the different conditions. A standard approach is block randomization. In block randomization, all the conditions occur once in the sequence before any of them is repeated.

Treatment

Between-subjects experiments are often used to determine whether a treatment works. In psychological research, a treatment is any intervention meant to change people’s behavior for the better. This includes psychotherapies and medical treatments for psychological disorders but also interventions designed to improve learning, promote conservation, reduce prejudice, and so on.

Treatment and Control conditions

Participants are randomly assigned to either a treatment condition, in which they receive the treatment, or a control condition, in which they do not receive the treatment. In research on the effectiveness of psychotherapies and medical treatments, this type of experiment is often called a randomized clinical trial.

Resource link: https://www.youtube.com/watch?v=fkOCYov1p-o
No-treatment control condition

There are different types of control conditions. In a no-treatment control condition, participants receive no treatment whatsoever. One problem with this approach, however, is the existence of placebo effects.

Placebo

A placebo is a simulated treatment that lacks any active ingredient or element that should make it effective, and a placebo effect is a positive effect of such a treatment.

Placebo control condition

Fortunately, there are several solutions to the problem of placebos. One is to include a placebo control condition, in which participants receive a placebo that looks much like the treatment but lacks the active ingredient or element thought to be responsible for the treatment’s effectiveness.

Waitlist control condition

An alternative solution to the problem of placebos is the use of a waitlist. In a waitlist control condition, participants are told that they will receive the treatment but must wait until the participants in the treatment condition have already received it.

Carryover effect

A carryover effect is an effect of being tested in one condition on participants’ behavior in later conditions.

Practice effect

One type of carryover effect is a practice effect, where participants perform a task better in later conditions because they have had a chance to practice it. Another type is a fatigue effect, where participants perform a task worse in later conditions because they become tired or bored.

Context effect

Being tested in one condition can change how participants perceive stimuli or interpret their task in later conditions.
This is called a context effect.

**Counterbalancing**

This is a solution to order effects (context/carryover), that can be used in many situations. It is counterbalancing, which means testing different participants in different orders.

Resource link: [https://www.youtube.com/watch?v=dFdeEw8kIWE](https://www.youtube.com/watch?v=dFdeEw8kIWE)

Resource title: Wk 12 - Order Effects and counterbalancing

Resource Description: This video gives a detailed explanation (including examples) of counterbalancing.

**Subject pool**

There are several approaches to recruiting participants. One is to use participants from a formal subject pool—an established group of people who have agreed to be contacted about participating in research studies. For example, at many colleges and universities, there is a subject pool consisting of students enrolled in introductory psychology courses who must participate in a certain number of studies to meet a course requirement.

**Experimenter expectancy effect**

One important source of such variation is the experimenter’s expectations about how participants “should” behave in the experiment. This is referred to as an experimenter expectancy effect.

**Double-blind design**

Good practice is to arrange for the experimenters to be “blind” to the research question or to the condition that each participant is tested in. The idea is to minimize experimenter expectancy effects by minimizing the experimenters’ expectations. Because both the participants and the experimenters are blind to the condition, this is referred to as a double-blind study.

**Pilot test**

A pilot test is a small-scale study conducted to make sure that a new procedure works as planned.

---

**Topic: Nonexperimental Research**

**Nonexperimental research**

Nonexperimental research is research that lacks the manipulation of an independent variable, random assignment of participants to conditions or orders of conditions, or both. Nonexperimental research falls into three broad categories: single-variable research, correlational and quasi experimental research, and qualitative research.

**Single-variable research**

First, research can be nonexperimental because it focuses on a single variable rather than a statistical relationship
between two variables. Although there is no widely shared term for this kind of research, we will call it single-variable research. Milgram’s original obedience study was nonexperimental in this way.

**Correlational research**

Research can also be nonexperimental because it focuses on a statistical relationship between two variables but does not include the manipulation of an independent variable, random assignment of participants to conditions or orders of conditions, or both. This kind of research takes two basic forms: correlational research and quasi experimental research. In correlational research, the researcher measures the two variables of interest with little or no attempt to control extraneous variables and then assesses the relationship between them.

**Quasi-experimental research**

In quasi-experimental research, the researcher manipulates an independent variable but does not randomly assign participants to conditions or orders of conditions.

Resource link: https://www.youtube.com/watch?v=vm-7k6unuLo

Resource title: Quasi-experimental designs

Resource description: A video by Fred Allcotte, providing a definition and an example.

**Naturalistic observation**

Naturalistic observation is an approach to data collection that involves observing people’s behavior in the environment in which it typically occurs. Thus, naturalistic observation is a type of field research (as opposed to a type of laboratory research). Researchers engaged in naturalistic observation usually make their observations as unobtrusively as possible so that participants are often not aware that they are being studied. Ethically, this is considered to be acceptable if the participants remain anonymous and the behavior occurs in a public setting where people would not normally have an expectation of privacy.

Resource link: https://www.youtube.com/watch?v=Gljv31Sumsw

Resource title: Naturalistic Observation Method

Resource description: A 2.30-minute long, animated video by Josh Knapp.

**Coding**

When the observations require a judgment on the part of the observers—as in Kraut and Johnston’s study—this process is often described as coding. Coding generally requires clearly defining a set of target behaviors. The observers then categorize participants individually in terms of which behavior they have engaged in and the number of times they engaged in each behavior. The observers might even record the duration of each behavior. The target behaviors must be defined in such a way that different observers code them in the same way.

Resource link: https://www.youtube.com/watch?v=eT-EDgwRvRU

Resource title: Qualitative Data Analysis – Coding and Developing Themes
Archival data

Another approach to correlational research is the use of archival data, which are data that have already been collected for some other purpose.

Content analysis

A family of systematic approaches to measurement using complex archival data. Just as naturalistic observation requires specifying the behaviors of interest and then noting them as they occur, content analysis requires specifying keywords, phrases, or ideas and then finding all occurrences of them in the data. These occurrences can then be counted, timed (e.g., the amount of time devoted to entertainment topics on the nightly news show) or analyzed in a variety of other ways.

Nonequivalent groups design

When participants are not randomly assigned to conditions, the resulting groups are likely to be dissimilar in some ways. For this reason, researchers consider them to be nonequivalent. A nonequivalent groups design, then, is a between-subjects design in which participants have not been randomly assigned to conditions.

Pretest-posttest design

In a pretest-posttest design, the dependent variable is measured once before the treatment is implemented and once after it is implemented. The pretest-posttest design is much like a within-subjects experiment in which each participant is tested first under the control condition and then under the treatment condition.

Resource link: https://dictionary.apa.org/one-group-pretest-posttest-design

Resource title: The APA dictionary of Psychology

History

If the average posttest score is better than the average pretest score, then it makes sense to conclude that the treatment might be responsible for the improvement. Unfortunately, one often cannot conclude this with a high degree of certainty because there may be other explanations for why the posttest scores are better. One category of alternative explanations goes under the name of history. Other things might have happened between the pretest and the posttest.

Maturation

Another category of alternative explanations goes under the name of maturation. Participants might have changed between the pretest and the posttest in ways that they were going to anyway because they are growing and learning.

Regression to the mean

Another alternative explanation for a change in the dependent variable in a pretest-posttest design is regression to
the mean. This refers to the statistical fact that an individual who scores extremely on a variable on one occasion will tend to score less extremely on the next occasion. For example, a bowler with a long-term average of 150 who suddenly bowls a 220 will almost certainly score lower in the next game. Her score will “regress” toward her mean score of 150. Regression to the mean can be a problem when participants are selected for further study because of their extreme scores.

Spontaneous remission

A closely related concept—and an extremely important one in psychological research—is spontaneous remission. This is the tendency for many medical and psychological problems to improve over time without any form of treatment. The common cold is a good example.

Interrupted time-series design

A variant of the pretest-posttest design is the interrupted time-series design. A time series is a set of measurements taken at intervals over a period of time.

Quantitative research

Quantitative researchers typically start with a focused research question or hypothesis, collect a small amount of data from each of a large number of individuals, describe the resulting data using statistical techniques, and draw general conclusions about some large population.

Interviews

As with correlational research, data collection approaches in qualitative research are quite varied and can involve naturalistic observation, archival data, artwork, and many other things. But one of the most common approaches, especially for psychological research, is to conduct interviews. Interviews in qualitative research tend to be unstructured—consisting of a small number of general questions or prompts that allow participants to talk about what is of interest to them. The researcher can follow up by asking more detailed questions about the topics that do come up. Such interviews can be lengthy and detailed, but they are usually conducted with a relatively small sample.

Resource link: https://www.youtube.com/watch?v=6PhcglOGFq8

Resource title: Fundamentals of Qualitative Research Methods: Interviews

Resource description: Part of a lecture series by Dr Leslie Curry of Yale University

Focus group

Small groups of people who participate together in interviews focused on a particular topic or issue are often referred to as focus groups. The interaction among participants in a focus group can sometimes bring out more information than can be learned in a one-on-one interview. The use of focus groups has become a standard technique in business and industry among those who want to understand consumer tastes and preferences. The content of all focus group interviews is usually recorded and transcribed to facilitate later analyses.

Resource link: https://www.youtube.com/watch?v=cCAPz14yjd4
Participant observation

In participant observation, researchers become active participants in the group or situation they are studying. The data they collect can include interviews (usually unstructured), their own notes based on their observations and interactions, documents, photographs, and other artifacts. The basic rationale for participant observation is that there may be important information that is only accessible to, or can be interpreted only by, someone who is an active participant in the group or situation.

Grounded theory

Just as there are many ways to collect data in qualitative research, there are many ways to analyze data. Here we focus on one general approach called grounded theory (Glaser & Strauss, 1967). This approach was developed within the field of sociology in the 1960s and has gradually gained popularity in psychology. Remember that in quantitative research, it is typical for the researcher to start with a theory, derive a hypothesis from that theory, and then collect data to test that specific hypothesis. In qualitative research using grounded theory, researchers start with the data and develop a theory or an interpretation that is “grounded in” those data. They do this in stages. First, they identify ideas that are repeated throughout the data. Then they organize these ideas into a smaller number of broader themes. Finally, they write a theoretical narrative—an interpretation—of the data in terms of the themes that they have identified. This theoretical narrative focuses on the subjective experience of the participants and is usually supported by many direct quotations from the participants themselves.

Resource link: https://www.youtube.com/watch?v=D5AHmHS6WQ

Resource title: A Discussion with Prof Kathy Charmaz on Grounded Theory

Mixed-methods research

Many researchers from both the quantitative and qualitative camps now agree that the two approaches can and should be combined into what has come to be called mixed-methods research. One approach to combining quantitative and qualitative research is to use qualitative research for hypothesis generation and quantitative research for hypothesis testing.


Resource title: Qualitative and mixed methods research - Better understanding, better science.

Resource description: In this article, Michael Sladek advocates the use of mixed methods in psychological research and gives tips on using qualitative techniques.
Triangulation

A second approach to combining quantitative and qualitative research is referred to as triangulation. The idea is to use both quantitative and qualitative methods simultaneously to study the same general questions and to compare the results. If the results of the quantitative and qualitative methods converge on the same general conclusion, they reinforce and enrich each other. If the results diverge, then they suggest an interesting new question: Why do the results diverge and how can they be reconciled?

---

Topic: Complex Research Designs

Multiple dependent variables

Researchers in psychology often include multiple dependent variables in their studies. The primary reason is that this easily allows them to answer more research questions with minimal additional effort.

Manipulation check

When an independent variable is a construct that is manipulated indirectly, it is a good idea to include a manipulation check. This is a measure of the independent variable typically given at the end of the procedure to confirm that it was successfully manipulated.

Factorial design

By far the most common approach to including multiple independent variables in an experiment is the factorial design. In a factorial design, each level of one independent variable (which can also be called a factor) is combined with each level of the others to produce all possible combinations. Each combination, then, becomes a condition in the experiment. A factorial design table represents all possible conditions of the experiment.

Resource link: https://www.youtube.com/watch?v=EgVmfryCAkU

Resource title: Factorial Research Design – An Example

Resource description: Michael Britt uses an example to explain factorial designs.

Between-subjects factorial design

In a between-subjects factorial design, all of the independent variables are manipulated between subjects. This means that each participant is tested in one and only one condition.

Within-subjects factorial design

In a within-subjects factorial design, all the independent variables are manipulated within subjects. This means that each participant is tested in all conditions.

Mixed factorial design

It is also possible to manipulate one independent variable between subjects and another within subjects. Each
participant in a mixed design is tested in some of the conditions.

**Nonmanipulated independent variable**

In many factorial designs, one of the independent variables is a nonmanipulated independent variable. The researcher measures it but does not manipulate it. Nonmanipulated independent variables are usually participant variables (private body consciousness, hypochondriasis, self-esteem, and so on), and as such they are by definition between-subjects factors.

**Main effect**

In factorial designs, there are two kinds of results that are of interest: main effects and interaction effects (which are also called just “interactions”). A main effect is the statistical relationship between one independent variable and a dependent variable—averaging across the levels of the other independent variable. Thus, there is one main effect to consider for each independent variable in the study.

**Interaction**

There is an interaction effect (or just “interaction”) when the effect of one independent variable depends on the level of another.

Resource link: [https://www.youtube.com/watch?v=OE46w0RqmQA](https://www.youtube.com/watch?v=OE46w0RqmQA)

Resource title: Main effects & interactions

Resource description: The video uses an example of a study on sleep deprivation, caffeine, and memory, to illustrate main effects and interactions.

**Crossover interaction**

This is the strongest form of interaction between independent variables. One example of a crossover interaction comes from a study by Kathy Gilliland on the effect of caffeine on the verbal test scores of introverts and extroverts (Gilliland, 1980). Introverts perform better than extroverts when they have not ingested any caffeine. But extroverts perform better than introverts when they have ingested 4 mg of caffeine per kilogram of body weight.

**Correlation matrix**

A correlation matrix is a table showing the correlation (Pearson’s r) between every possible pair of variables in a study.

**Factor analysis**

When researchers study relationships among a large number of conceptually similar variables, they often use a complex statistical technique called factor analysis. In essence, factor analysis organizes the variables into a smaller number of clusters, such that they are strongly correlated within each cluster but weakly correlated between clusters. Each cluster is then interpreted as multiple measures of the same underlying construct. These underlying constructs are also called “factors.” The Big Five personality factors have been identified through factor analyses of people’s scores on a large number of more specific traits.
Statistical control

It is true that correlational research cannot unambiguously establish that one variable causes another. Complex correlational research, however, can often be used to rule out other plausible interpretations. The primary way of doing this is through the statistical control of potential third variables. Instead of controlling these variables by random assignment or by holding them constant as in an experiment, the researcher measures them and includes them in the statistical analysis.

Multiple regression

Many studies of this type use a statistical technique called multiple regression. This involves measuring several independent variables \( (X_1, X_2, X_3, \ldots, X_i) \), all of which are possible causes of a single dependent variable \( (Y) \). The result of a multiple regression analysis is an equation that expresses the dependent variable as an additive combination of the independent variables. This regression equation has the following general form:

\[
b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_iX_i = Y.
\]

The quantities \( b_1, b_2, \text{ and so on} \) are regression weights that indicate how large a contribution an independent variable makes, on average, to the dependent variable. Specifically, they indicate how much the dependent variable changes for each one-unit change in the independent variable.

The advantage of multiple regression is that it can show whether an independent variable makes a contribution to a dependent variable over and above the contributions made by other independent variables.

Resource link: https://www.youtube.com/watch?v=oVxY57pfMOo

Resource title: Multiple Regression: 1 - Multiple regression and multicollinearity

Resource description: Ross Avilla (who has also made other useful videos on research topics) here talks about multiple regression.

---

Survey research

Survey research is a quantitative approach that has two important characteristics. First, the variables of interest are measured using self-reports. In essence, survey researchers ask their participants (who are often called respondents in survey research) to report directly on their own thoughts, feelings, and behaviors. Second, considerable attention is paid to the issue of sampling. In particular, survey researchers have a strong preference for large random samples because they provide the most accurate estimates of what is true in the population. In fact, survey research may be the only approach in psychology in which random sampling is routinely used.
Context effect

Survey questionnaire responses are subject to numerous context effects due to question wording, item order, response options, and other factors. Researchers should be sensitive to such effects when constructing surveys and interpreting survey results.

Item-order effect

The order in which the items are presented may affect people’s responses. One item can change how participants interpret a later item or change the information that they retrieve to respond to later items.

Open-ended item

Survey questionnaire items are either open-ended or closed-ended. Open-ended items simply ask a question and allow respondents to answer in whatever way they want.

Closed-ended item

Closed-ended items ask a question and provide several response options that respondents must choose from. All closed-ended items include a set of response options from which a participant must choose. For categorical variables like sex, race, or political party preference, the categories are usually listed and participants choose the one (or ones) that they belong to. For quantitative variables, a rating scale is typically provided.

Rating scale

A rating scale is an ordered set of responses that participants must choose from.

BRUSO

BRUSO stands for "brief," “relevant,” “unambiguous,” “specific,” and “objective.” Effective questionnaire items are brief and to the point. They avoid long, overly technical, or unnecessary words. Effective questionnaire items are also relevant to the research question. This makes the questionnaire faster to complete, but it also avoids annoying respondents with what they will rightly perceive as irrelevant or even “nosy” questions. Effective questionnaire items are also unambiguous; they can be interpreted in only one way. Effective questionnaire items are also specific, so that it is clear to respondents what their response should be about and clear to researchers what it is about. Finally, effective questionnaire items are objective in the sense that they do not reveal the researcher’s own opinions or lead participants to answer in a particular way.

Resource link: https://www.youtube.com/watch?v=Iq_fhTuY1hw

Resource title: 7 tips for good survey questions

Resource description: This video features a political science professor who talks about surveys. Since surveys are extremely prevalent in political science, he might have some useful input for psychologists.

Probability sampling

Survey research usually involves probability sampling, in which each member of the population has a known
probability of being selected for the sample. Types of probability sampling include simple random sampling, stratified random sampling, and cluster sampling.

Resource link: https://www.youtube.com/watch?v=be9e-Q-jC-0

Resource title:

Sampling: Simple Random, Convenience, systematic, cluster, stratified - Statistics Help

Resource description: A woman with a great Kiwi accent explains the different types of sampling in an easy way.

Nonprobability sampling

Nonprobability sampling occurs when the researcher cannot specify these probabilities. Most psychological research involves nonprobability sampling. Convenience sampling—studying individuals who happen to be nearby and willing to participate—is a very common form of nonprobability sampling used in psychological research.

Sampling frame

Once the population has been specified, probability sampling requires a sampling frame. This is essentially a list of all the members of the population from which to select the respondents. Sampling frames can come from a variety of sources, including telephone directories, lists of registered voters, and hospital or insurance records. In some cases, a map can serve as a sampling frame, allowing for the selection of cities, streets, or households.

Simple random sampling

Simple random sampling is done in such a way that each individual in the population has an equal probability of being selected for the sample. This could involve putting the names of all individuals in the sampling frame into a hat, mixing them up, and then drawing out the number needed for the sample. Given that most sampling frames take the form of computer files, random sampling is more likely to involve computerized sorting or selection of respondents.


Resource title: Techniques for generating a simple random sample

Resource description: A helpful Khan academy video. This is part of a playlist of several potentially useful explanations.

Stratified random sampling

A common alternative to simple random sampling is stratified random sampling, in which the population is divided into different subgroups or “strata” (usually based on demographic characteristics) and then a random sample is taken from each “stratum.” Stratified random sampling can be used to select a sample in which the proportion of respondents in each of various subgroups matches the proportion in the population.
Cluster sampling

Yet another type of probability sampling is cluster sampling, in which larger clusters of individuals are randomly sampled and then individuals within each cluster are randomly sampled. For example, to select a sample of small-town residents in the United States, a researcher might randomly select several small towns and then randomly select several individuals within each town. Cluster sampling is especially useful for surveys that involve face-to-face interviewing because it minimizes the amount of traveling that the interviewers must do.

Sampling bias

Sampling bias occurs when a sample is selected in such a way that it is not representative of the population and therefore produces inaccurate results.

Nonresponse bias

The most pervasive form of sampling bias is nonresponse bias, which occurs when people who do not respond to the survey differ in important ways from people who do respond. The best way to minimize nonresponse bias is to maximize the response rate by prenotifying respondents, sending them reminders, constructing questionnaires that are short and easy to complete, and offering incentives.

Resource link: https://www.youtube.com/watch?v=6kxajhKWDII

Resource title: How to Reduce Survey Non-Response

Resource description: Steven Litt gives tips on how to reduce nonresponse bias. He uses several examples of surveys he or students of his have done.

Topic: Single-Subject Research

Single-subject research

Single-subject research—which involves testing a small number of participants and focusing intensively on the behavior of each individual—is an important alternative to group research in psychology.

Group research

Single-subject research can be contrasted with group research, which typically involves studying large numbers of participants and examining their behavior primarily in terms of group means, standard deviations, and so on.

Case study

Single-subject studies must be distinguished from case studies, in which an individual case is described in detail. Case studies can be useful for generating new research questions, for studying rare phenomena, and for illustrating general principles.
Social validity

An assumption of single-subject research is that it is important to study strong and consistent effects that have biological or social importance. Applied researchers, in particular, are interested in treatments that have substantial effects on important behaviors and that can be implemented reliably in the real-world contexts in which they occur. This is sometimes referred to as social validity.

Experimental analysis of behavior

In the middle of the 20th century, B. F. Skinner clarified many of the assumptions underlying single-subject research and refined many of its techniques (Skinner, 1938). He and other researchers then used it to describe how rewards, punishments, and other external factors affect behavior over time. This work was carried out primarily using nonhuman subjects—mostly rats and pigeons. This approach, which Skinner called the experimental analysis of behavior—remains an important subfield of psychology and continues to rely almost exclusively on single-subject research.

Applied behavior analysis

By the 1960s, many researchers were interested in using the behavioristic approach to conduct applied research primarily with humans—a subfield now called applied behavior analysis (Baer, Wolf, & Risley, 1968). Applied behavior analysis plays an especially important role in contemporary research on developmental disabilities, education, organizational behavior, and health, among many other areas.

Steady state strategy

Single-subject research designs typically involve measuring the dependent variable repeatedly over time and changing conditions (e.g., from baseline to treatment) when the dependent variable has reached a steady state. This approach allows the researcher to see whether changes in the independent variable are causing changes in the dependent variable.

Reversal design

In a reversal design, the participant is tested in a baseline condition, then tested in a treatment condition, and then returned to baseline. If the dependent variable changes with the introduction of the treatment and then changes back with the return to baseline, this provides strong evidence of a treatment effect.

ABA design

A basic type of reversal design. During the first phase, A, a baseline is established for the dependent variable. This is the level of responding before any treatment is introduced, and therefore the baseline phase is a kind of control condition. When steady state responding is reached, phase B begins as the researcher introduces the treatment. There may be a period of adjustment to the treatment during which the behavior of interest becomes more variable and begins to increase or decrease. Again, the researcher waits until that dependent variable reaches a steady state so that it is clear whether and how much it has changed. Finally, the researcher removes the treatment and again waits until the dependent variable reaches a steady state. This basic reversal design can also be extended with the reintroduction of the treatment (ABAB), another return to baseline (ABABA), and so on.
Multiple-treatment reversal design

There are close relatives of the basic reversal design that allow for the evaluation of more than one treatment. In a multiple-treatment reversal design, a baseline phase is followed by separate phases in which different treatments are introduced. For example, a researcher might establish a baseline of studying behavior for a disruptive student (A), then introduce a treatment involving positive attention from the teacher (B), and then switch to a treatment involving mild punishment for not studying (C). The participant could then be returned to a baseline phase before reintroducing each treatment—perhaps in the reverse order as a way of controlling for carryover effects. This particular multiple-treatment reversal design could also be referred to as an ABCACB design.

Alternating treatments design

In an alternating treatments design, two or more treatments are alternated relatively quickly on a regular schedule. For example, positive attention for studying could be used one day and mild punishment for not studying the next, and so on. Or one treatment could be implemented in the morning and another in the afternoon. The alternating treatments design can be a quick and effective way of comparing treatments, but only when the treatments are fast acting.

Multiple-baseline design

In a multiple-baseline design, baselines are established for different participants, different dependent variables, or different settings—and the treatment is introduced at a different time on each baseline. If the introduction of the treatment is followed by a change in the dependent variable on each baseline, this provides strong evidence of a treatment effect.

Visual inspection

Plotting individual participants’ data, looking carefully at those data, and making judgments about whether and to what extent the independent variable had an effect on the dependent variable. Inferential statistics are typically not used in single-subject research.

Level, Trend and Latency

In visually inspecting their data, single-subject researchers take several factors into account. One of them is changes in the level of the dependent variable from condition to condition. If the dependent variable is much higher or much lower in one condition than another, this suggests that the treatment had an effect. A second factor is trend, which refers to gradual increases or decreases in the dependent variable across observations. If the dependent variable begins increasing or decreasing with a change in conditions, then again this suggests that the treatment had an effect. It can be especially telling when a trend changes direction—for example, when an unwanted behavior is increasing during baseline but then begins to decrease with the introduction of the treatment. A third factor is latency, which is the time it takes for the dependent variable to begin changing after a change in conditions. In general, if a change in the dependent variable begins shortly after a change in conditions, this suggests that the treatment was responsible.

Percentage of nonoverlapping data

This is the percentage of responses in the treatment condition that are more extreme than the most extreme
response in a relevant control condition. The greater the percentage of nonoverlapping data, the stronger the treatment effect.

**Topic: Presenting Your Research**

**APA style**

APA style is a set of guidelines for writing in psychology and related fields. These guidelines are set down in the *Publication Manual of the American Psychological Association* (APA, 2006). The *Publication Manual* originated in 1929 as a short journal article that provided basic standards for preparing manuscripts to be submitted for publication (Bentley et al., 1929). It was later expanded and published as a book by the association and is now in its sixth edition. The primary purpose of APA style is to facilitate scientific communication by promoting clarity of expression and by standardizing the organization and content of research articles and book chapters. APA style can be seen as having three levels. There is the organization of a research article, the high-level style that includes writing in a formal and straightforward way, and the low-level style that consists of many specific rules of grammar, spelling, formatting of references, and so on.

**Reference citation**

When you refer to another researcher’s idea, you must include a reference citation (in the text) to the work in which that idea originally appeared and a full reference to that work in the reference list.

**Empirical research report**

an article that presents the results of one or more new studies.

**Title page**

An APA-style research report begins with a title page. The title is centered in the upper half of the page, with each important word capitalized. The title should clearly and concisely (in about 12 words or fewer) communicate the primary variables and research questions.

**Abstract**

The abstract is a summary of the study. It is the second page of the manuscript and is headed with the word *Abstract*. The first line is not indented. The abstract presents the research question, a summary of the method, the basic results, and the most important conclusions. Because the abstract is usually limited to about 200 words, it can be a challenge to write a good one.

**Introduction**

The introduction begins on the third page of the manuscript. The heading at the top of this page is the full title of the manuscript, with each important word capitalized as on the title page. The introduction includes three distinct subsections, although these are typically not identified by separate headings. The opening introduces the research question and explains why it is interesting, the literature review discusses relevant previous research, and the closing restates the research question and comments on the method used to answer it.
Method section

The method section is where you describe how you conducted your study. An important principle for writing a method section is that it should be clear and detailed enough that other researchers could replicate the study by following your “recipe.” At the same time, it should avoid irrelevant details.

Results section

The results section describes the results in an organized fashion. Each primary result is presented in terms of statistical results but also explained in words.

Discussion

The discussion typically summarizes the study, discusses theoretical and practical implications and limitations of the study, and offers suggestions for further research.

Appendix

An appendix is appropriate for supplemental material that would interrupt the flow of the research report if it were presented within any of the major sections. An appendix could be used to present lists of stimulus words, questionnaire items, detailed descriptions of special equipment or unusual statistical analyses, or references to the studies that are included in a meta-analysis.

Review and theoretical articles

Recall that review articles summarize research on a particular topic without presenting new empirical results. When these articles present a new theory, they are often called theoretical articles. Review and theoretical articles are structured much like empirical research reports, with a title page, an abstract, references, appendixes, tables, and figures, and they are written in the same high-level and low-level style. Because they do not report the results of new empirical research, however, there is no method or results section.

Copy manuscript

manuscripts that will be submitted to a professional journal for publication. Many features of a copy manuscript—consistent double-spacing, the running head, and the placement of tables and figures at the end—are intended to make it easier to edit and typeset on its way to publication.

Final manuscript

Final manuscripts are manuscripts which are prepared by the author in their final form with no intention of submitting them for publication elsewhere. They include dissertations, theses, and other student papers.

Professional conference

One of the ways that researchers in psychology share their research with each other is by presenting it at professional conferences. Professional conferences can range from small-scale events involving a dozen researchers who get together for an afternoon to large-scale events involving thousands of researchers who meet for several days. Although researchers attending a professional conference are likely to discuss their work with each other.
other informally, there are two more formal types of presentation: oral presentations ("talks") and posters.

**Poster session**

A poster is typically presented during a one- to two-hour poster session that takes place in a large room at the conference site. Presenters set up their posters on bulletin boards arranged around the room and stand near them. Other researchers then circulate through the room, read the posters, and talk to the presenters. In essence, poster sessions are a grown-up version of the school science fair. But there is nothing childish about them. Posters are used by professional researchers in all scientific disciplines and they are becoming increasingly common.

---

**Topic: Descriptive Research**

**Descriptive statistics**

Descriptive statistics refers to a set of techniques for summarizing and displaying data.

**Distribution**

Every variable has a distribution, which is the way the scores are distributed across the levels of that variable.

**Frequency table and histograms**

The distribution can be described using a frequency table and histogram. A histogram is a graphical display of a distribution. It presents the same information as a frequency table but in a way that is even quicker and easier to grasp.

**Symmetrical and Skewed**

One characteristic of the shape of a distribution is whether it is symmetrical or skewed. If a distribution is symmetrical, its left and right halves are mirror images of each other. In a negatively skewed distribution, the peak is shifted toward the upper end of its range and the distribution has a relatively long negative tail. A distribution with its peak toward the lower end of its range and a relatively long positive tail is positively skewed.

**Outlier**

An outlier is an extreme score that is much higher or lower than the rest of the scores in the distribution. Sometimes outliers represent truly extreme scores on the variable of interest. However, outliers can also represent errors or misunderstandings on the part of the researcher or participant, equipment malfunctions, or similar problems.

**Central tendency**

The central tendency of a distribution is its middle—the point around which the scores in the distribution tend to cluster. Another term for central tendency is *average*.

**Mean**

The mean of a distribution (symbolized \(\bar{M}\)) is the sum of the scores divided by the number of scores. As a formula,
it looks like this: \[M=\sum X N. \nonumber\]

**Median**

The median is the middle score in the sense that half the scores in the distribution are less than it and half are greater than it. The simplest way to find the median is to organize the scores from lowest to highest and locate the score in the middle.

**Mode**

The mode is the most frequent score in a distribution.

**Variability**

The variability of a distribution is the extent to which the scores vary around their central tendency.

**Range**

One simple measure of variability is the range, which is simply the difference between the highest and lowest scores in the distribution.

**Standard deviation**

By far the most common measure of variability is the standard deviation. The standard deviation of a distribution is, roughly speaking, the average distance between the scores and the mean.

**Variance**

Although the variance is itself a measure of variability, it generally plays a larger role in inferential statistics than in descriptive statistics. The standard deviation is the square root of the variance.

**Percentile rank**

The percentile rank of a score is the percentage of scores in the distribution that are lower than that score.

**z score**

The z score for a particular individual is the difference between that individual’s score and the mean of the distribution, divided by the standard deviation of the distribution: \[z=X-\text{MSD. \nonumber}\]

**Effect size**

It is important to be able to describe the strength of a statistical relationship, which is often referred to as the effect size.

**Cohen’s d**

The most widely used measure of effect size for differences between group or condition means is called Cohen’s d, which is the difference between the two means divided by the standard deviation: \[d=\frac{M_1-M_2}{SD}.\nonumber\]
Nonlinear relationship

Nonlinear relationships are those in which the points of a scatterplot are better fit by a curved line than a straight one. Nonlinear relationships are quite common in psychology.

Restriction of range

when one or both of the variables have a limited range in the sample relative to the population, this is referred to as restriction of range.

Bar graph

bar graphs are generally used to present and compare the mean scores for two or more groups or conditions

Error bars

smaller vertical bars that extend both upward and downward from the top of each main bar. These are error bars, and they represent the variability in each group or condition. Although they sometimes extend one standard deviation in each direction, they are more likely to extend one standard error in each direction

Standard error

The standard error is the standard deviation of the group divided by the square root of the sample size of the group. The standard error is used because, in general, a difference between group means that is greater than two standard errors is statistically significant.

Line graph

Line graphs are used to present correlations between quantitative variables when the independent variable has, or is organized into, a relatively small number of distinct levels. Each point in a line graph represents the mean score on the dependent variable for participants at one level of the independent variable.

Scatterplot

Scatterplots are used to present relationships between quantitative variables when the variable on the x-axis (typically the independent variable) has a large number of levels. Each point in a scatterplot represents an individual rather than the mean for a group of individuals, and there are no lines connecting the points.

Correlation matrix

Another common use of tables is to present correlations—usually measured by Pearson’s r—among several variables. This is called a correlation matrix.

Raw data

“Raw” data are unanalyzed data.

Data file

You can use a general spreadsheet program like Microsoft Excel or a statistical analysis program like SPSS to
create your data file. The most common format is for each row to represent a participant and for each column to represent a variable.

**Topic: Inferential Statistics**

**Parameter**

Researchers must use sample statistics to draw conclusions about the corresponding values in the population. These corresponding values in the population are called parameters.

**Sampling error**

Sample statistics are not perfect estimates of their corresponding population parameters. This is because there is a certain amount of random variability in any statistic from sample to sample. This random variability in a statistic from sample to sample is called sampling error.

**Null hypothesis testing**

Null hypothesis testing is a formal approach to deciding between two interpretations of a statistical relationship in a sample. One interpretation is called the null hypothesis (often symbolized $H_0$ and read as “H-naught”). The other interpretation is called the alternative hypothesis (often symbolized as $H_1$).

**Null hypothesis**

This is the idea that there is no relationship in the population and that the relationship in the sample reflects only sampling error. Informally, the null hypothesis is that the sample relationship “occurred by chance.”

**Alternative hypothesis**

This is the idea that there is a relationship in the population and that the relationship in the sample reflects this relationship in the population.

**Reject or retain the null hypothesis**

The logic of null hypothesis testing involves assuming that the null hypothesis is true, finding how likely the sample result would be if this assumption were correct, and then making a decision. If the sample result would be unlikely if the null hypothesis were true, then it is rejected in favor of the alternative hypothesis. If it would not be unlikely, then the null hypothesis is retained.

**$p$ value**

The probability of obtaining the sample result if the null hypothesis were true (the $p$ value) is based on two considerations: relationship strength and sample size. Reasonable judgments about whether a sample relationship is statistically significant can often be made by quickly considering these two factors.
**Alpha**

How low must the $p$ value be before the sample result is considered unlikely enough to reject the null hypothesis? In null hypothesis testing, this criterion is called $\alpha$ (alpha) and is almost always set to .05. If there is less than a 5% chance of a result as extreme as the sample result if the null hypothesis were true, then the null hypothesis is rejected.

**Statistically significant**

When the null hypothesis is rejected, the result is said to be statistically significant.

**Practical significance**

It is important to distinguish between the statistical significance of a result and the practical significance of that result. Practical significance refers to the importance or usefulness of the result in some real-world context. Many sex differences are statistically significant—and may even be interesting for purely scientific reasons—but they are not practically significant. In clinical practice, this same concept is often referred to as “clinical significance.”

**t test**

Many studies in psychology focus on the difference between two means. The most common null hypothesis test for this type of statistical relationship is the $t$ test.

**One sample $t$ test**

The one-sample $t$ test is used to compare a sample mean ($M$) with a hypothetical population mean ($\mu_0$) that provides some interesting standard of comparison. The null hypothesis is that the mean for the population ($\mu$) is equal to the hypothetical population mean: $\mu = \mu_0$. The alternative hypothesis is that the mean for the population is different from the hypothetical population mean: $\mu \neq \mu_0$. To decide between these two hypotheses, we need to find the probability of obtaining the sample mean (or one more extreme) if the null hypothesis were true.

**Critical value**

The $t$-score needed to either reject or retain the null hypothesis.

**Two-tailed test**

If the $t$-score we compute is beyond the critical value in either direction, then we reject the null hypothesis. If the $t$-score we compute is between the upper and lower critical values, then we retain the null hypothesis.

**One-tailed test**

In a one-tailed test, we reject the null hypothesis only if the $t$ score for the sample is extreme in one direction that we specify before collecting the data. This makes sense when we have good reason to expect the sample mean will differ from the hypothetical population mean in a particular direction.

**Dependent-samples $t$ test**

The dependent-samples $t$ test (sometimes called the paired-samples $t$ test) is used to compare two means for the
same sample tested at two different times or under two different conditions. This makes it appropriate for pretest-posttest designs or within-subjects experiments. The null hypothesis is that the means at the two times or under the two conditions are the same in the population. The alternative hypothesis is that they are not the same. This test can also be one-tailed if the researcher has good reason to expect the difference goes in a particular direction.

### Difference score

The first step in the dependent-samples t test is to reduce the two scores for each participant to a single difference score by taking the difference between them. At this point, the dependent-samples t test becomes a one-sample t test on the difference scores. The hypothetical population mean (μ₀) of interest is 0 because this is what the mean difference score would be if there were no difference on average between the two times or two conditions.

### Independent-samples t-test

The independent-samples t-test is used to compare the means of two separate samples (M₁ and M₂). The two samples might have been tested under different conditions in a between-subjects experiment, or they could be preexisting groups in a correlational design (e.g., women and men, extroverts and introverts). The null hypothesis is that the means of the two populations are the same: μ₁ = μ₂. The alternative hypothesis is that they are not the same: μ₁ ≠ μ₂. Again, the test can be one-tailed if the researcher has good reason to expect the difference goes in a particular direction.

### Analysis of Variance (ANOVA)

When there are more than two groups or condition means to be compared, the most common null hypothesis test is the analysis of variance (ANOVA).

#### One-way ANOVA

The one-way ANOVA is used to compare the means of more than two samples (M₁, M₂...MG) in a between-subjects design. The null hypothesis is that all the means are equal in the population: μ₁ = μ₂ =...= μ₉. The alternative hypothesis is that not all the means in the population are equal.

#### Mean squares between and within groups

The test statistic for the ANOVA is called F. It is a ratio of two estimates of the population variance based on the sample data. One estimate of the population variance is called the mean squares between groups (MSB) and is based on the differences among the sample means. The other is called the mean squares within groups (MSW) and is based on the differences among the scores within each group. The F statistic is the ratio of the MSB to the MSW and can therefore be expressed as follows: \[ F = \frac{MSB}{MSW}. \]

#### Post hoc comparisons

Statistically significant one-way ANOVA results are typically followed up with a series of post hoc comparisons of selected pairs of group means to determine which are different from which others.

#### Repeated-measures ANOVA

One-way ANOVA is appropriate for between-subjects designs in which the means being compared come from
separate groups of participants. It is not appropriate for within-subjects designs in which the means being compared come from the same participants tested under different conditions or at different times. This requires a slightly different approach, called the repeated-measures ANOVA. The basics of the repeated-measures ANOVA are the same as for the one-way ANOVA. The main difference is that measuring the dependent variable multiple times for each participant allows for a more refined measure of $MS_W$.

**Factorial ANOVA**

When more than one independent variable is included in a factorial design, the appropriate approach is the factorial ANOVA. Again, the basics of the factorial ANOVA are the same as for the one-way and repeated-measures ANOVAs. The main difference is that it produces an $F$ ratio and $p$ value for each main effect and for each interaction.

**Type I error**

Rejecting the null hypothesis when it is true is called a Type I error. This means that we have concluded that there is a relationship in the population when in fact there is not.

**Type II error**

Retaining the null hypothesis when it is false is called a Type II error. This means that we have concluded that there is no relationship in the population when in fact there is. In practice, Type II errors occur primarily because the research design lacks adequate statistical power to detect the relationship (e.g., the sample is too small).

**File drawer problem**

An issue related to Type I errors is the so-called file drawer problem (Rosenthal, 1979). The idea is that when researchers obtain statistically significant results, they tend to submit them for publication, and journal editors and reviewers tend to accept them. But when researchers obtain nonsignificant results, they tend not to submit them for publication, or if they do submit them, journal editors and reviewers tend not to accept them. Researchers end up putting these nonsignificant results away in a file drawer (or nowadays, in a folder on their hard drive). One effect of this is that the published literature probably contains a higher proportion of Type I errors than we might expect on the basis of statistical considerations alone. Even when there is a relationship between two variables in the population, the published research literature is likely to overstate the strength of that relationship.

**Statistical power**

The statistical power of a research design is the probability of rejecting the null hypothesis given the expected relationship strength in the population and the sample size. Researchers should make sure that their studies have adequate statistical power before conducting them.

**Confidence interval**

A confidence interval around a statistic is a range of values that is computed in such a way that some percentage of the time (usually 95%) the population parameter will lie within that range. Confidence intervals are used as an alternative to null hypothesis testing.
Bayesian statistics

There are more radical solutions to the problems of null hypothesis testing that involve using very different approaches to inferential statistics. Bayesian statistics, for example, is an approach in which the researcher specifies the probability that the null hypothesis and any important alternative hypotheses are true before conducting the study, conducts the study, and then updates the probabilities based on the data.