Lecture notes, lectures 1 - 14 - Introduction to Research Methods in Psychology

Introduction to Research Methods in Psychology (MacEwan University)

Scientific Understanding of Behaviour

ethics are super important

Scientific Research

- Scientific research is a method of inquiry that follows a standard approach to asking and answering questions
- 'Thing' vs. way of thinking
    - Science is not a thing

Uses of research methods

- Informed consumer of info:
    - Read reports critically
        - Evaluate the methods used
            - Ex. Glass of red wine reduces risk of cancer BUT increases liver disease
        - Determine if conclusions are reasonable- are they rational or a leap of faith?
    - Make better decisions: personal & work (ex. Buying a house or car)
        - Influences public policy and judicial decisions
            - They rely on research but don't always fully understand it
        - Evaluate the effectiveness of goal-oriented programs (ex. Safe injection sites for druggies; effectiveness of D.A.R.E.)

Methods for Acquiring Knowledge

- Scientific method is not the only good way to acquire info…
- Intuition
    - Trial & error
    - Observation
    - Experience
    - Rely on anecdotal evidence to draw conclusions about the world around us
    - Rely unquestioningly on personal judgment
    - Involves cognitive and motivational biases
        - Does not protect us from our own biases (Mother of all biases→confirmation bias)
    - Erroneous conclusions about cause and effect
    - **Illusory correlation:**
        - Focus on two events that stand out and occur together
            - Ex. Weirdness on a full moon
        - Assume a relationship… can lead to superstition

- Authority
    - Profs
    - Word of mouth
    - Google (internet)
    - Believing ideas to be true because they were presented by an authority figure:
        - News, media (biased)
        - Books
        - Government officials
        - Religious figures
        - Political pundits
    - Should not unquestioningly believe/accept everything you're told; people can be WRONG
- The Scientific Approach
    - Acknowledges that both intuition and authority are sources of ideas
    - BUT does not accept these ideas without question
    - Ideas must be evaluated on the basis of careful logic and systematic observation
    - **Empiricism**: knowledge based on structured, systematic observations
        - A recipe or set of stages used to answer questions
        - Formulate hypothesis→collect data→analyze results
    - 1. Make **systematic observations** and report them accurately; allows for replication
    - 2. Engage in **theory-building**: involves the development, testing, and refining of theories
    - 3. Maintain an open system where *falsifiable* **ideas are exchanged**, debated and challenged
        - Reduces the likelihood that your biases taint your work
    - 4. Submit findings for ***peer review*** - evaluation by a panel of experts
    - **Objective**: set of rules for gathering, evaluating, an reporting information
    - **Open system** that allows ideas to be supported or refuted
        - Red flags should go up if this is not being done. It is important to be open
    - **Acknowledges value** of intuition & authority as a basis for generating research ideas
        - There is more than one way to acquire info

Question Research!

- What are the person's credentials?
- What is the person's reputation?
- Who is funding the research?
- Beware of pseudoscience! (seems scientific, but isn't)

Pseudoscience

- Can be right, but isn't always scientific
- Hypotheses generated are not typically testable
    - Just because it isn't testable, doesn't mean it's not true
    - You know that it works, just not necessarily HOW it works
- If scientific tests are reported, the methodology is not scientific and validity of data is questionable
- Supportive evidence is anecdotal and does not cite scientific references
    - Sometimes there just isn't enough support for a claim.
- Claims ignore conflicting evidence
- Claims tend to be vague, and appeal to pre-conceived ideas
- Claims are never revised

Goals of Scientific Research

1. *Describe* behavior: careful observation and measurement
    a. Mostly used when behavior is uncommon or rare
2. *Predict* behavior: based on the observation that events are systematically related to one another
    a. More useful, especially in law
3. *Determine* the causes of behavior:
    a. covariation of cause and effect
        i. the effect only happens if the cause is present
        ii. ex. Violence and video games→violence is only present in those who play video games
    b. temporal precedence
        i. the cause ALWAYS precedes the effect
        ii. ex. They start playing video games and then become more violent
    c. elimination of alternative explanations
        i. Does being violent cause people to play video games?
4. *Explain* why behavior occurs: much experimentation is often necessary to shed light on possible explanations of why one event causes another.
    a. Ex. Video games: Is it desensitization, habituation, or arousal/excitement…

Basic and Applied Research

- *Basic research (pure):* attempts to answer fundamental questions about the nature of behavior
    - Attempt to answer a question, generate more knowledge
- *Applied research:* address issues in which there are practical problems and potential solutions; includes program evaluation. Real world problem
    - Trying to apply the research in the real world
    - Sometimes applied research follows basic research, but not always
- Neither is considered superior to the other

- Applied research is often guided by theories and findings of basic research
- Findings in applied settings often require modification of existing theories and spur more basic research

Ethical Research

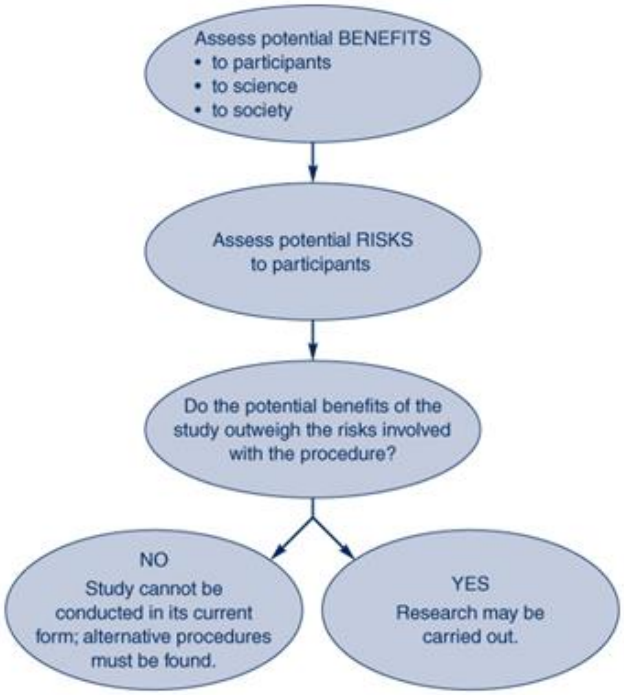
Ethical Principles

- Respect for persons
    - Autonomy
        - Informed consent-have to know what research is about
            - Deception- can't deceive participants about what will happen in experiment
        - Right to withdraw without prejudice
            - Have to dispose of data
        - Protected populations
            - Anyone who could potentially be voluntold (children, animals etc.)
- Concern for participants' welfare
    - Prevent harm:
        - Physical- **any** type of physical harm
            - Sprained ankle
            - Heart attack
        - Psychological
            - Stress
        - Social
            - Any form of social identification that could make them potentially stand out/ be discriminated against- social anonymity helps prevent this
- Justice- treat participants fairly
    - Equitable
        - Equally shared benefits
        - Equally shared risks
        - TUSKEGEE
            - Syphilis experiment

Risks-Benefits Analysis

- Potential Risks in psychological research: (Need a balance between protecting participants and usable data!)
    - Physical harm
    - Psychological stress
    - Loss of privacy and confidentiality

- Potential benefits of psychological research
    - Benefits to the science
    - Benefits to society
    - Benefits to Individual subjects
        - Educational benefits
            - Learn new skill, or access to treatment for a psychological or medical condition
        - Material benefits
        - Personal satisfaction



Ethical Research in Canada

- Tri-Council Policy Statement (TCPS)
    - Canadian Institutes of Health Research (CIHR)
    - Social Sciences and Humanities Research Council of Canada (SSHRC)
    - Natural Sciences and Engineering Research Council of Canada (NSERC)
- Nuremburg code
    - Leaders were put on trial and prosecuted after WWII
    - Biggest emphasis→informed consent
- World Medical Association Helsinki declaration
    - The way research of medical studies is conducted is looked at
- The Belmont Report
    - Related to the use of human subjects in relation to research

- American Psychological Association Ethics Code
    - -Paved the way for ethics in Canada
- Canadian research must:
    - Adhere to the TCPS guidelines
    - Comply with the *Canadian Charter of Rights and Freedoms*, Canadian privacy of information laws, and relevant federal and provincial laws.

The Research Ethics Board (REB)

- Each institution that receives federal funds must have an REB
- Responsible for reviewing research at the institution
- REB approval must be obtained *prior* to starting research

Types of Research and the REB

- Exempt Research: no risk, or potential for risk. No need for approval by REB
    - Employs publicly available information
    - Involves observing people in public places without any interaction (between researcher and subjects)
    - Uses already collected data that are completely anonymous.
- Minimal Risk Research
    - Risk of harm is no greater than risk encountered in daily life or routine physical or psychological tests
    - Questionnaires or interviews on non-sensitive topics
- Greater than Minimal Risk
    - Subjected to thorough review
    - Additional safeguards beyond informed consent
        - Could offer therapy after etc.

The Research Ethics Board

- Impact on Research
    - Extended time for approval of study
    - Submissions often need to be revised or clarified
    - Very cautious about approval

Ethics and Animal Research

- Reasons for using animals
    - Full experimental control
        - During and out of the study
    - Used for research not possible or ethically acceptable with humans
    - Procedures that are irreversible

- Standards
    - Proper handling & care of animals
    - Minimize pain & avoid cruelty
        - Have to justify the cost
- Canadian Council on Animal Care
    - Replacement: where possible avoid using an animal model
        - Can you use something else?
    - Reduction: minimize the number of animals used
    - Refinement: minimize pain & distress

Professional Ethics

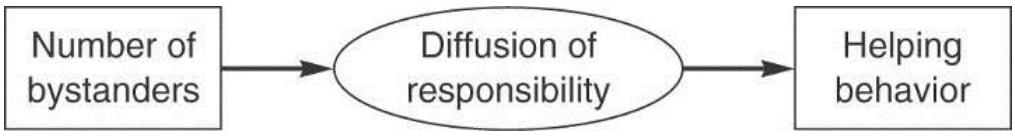
- Research Fraud
    - Fabrication of data
    - Results difficult to replicate
    - Reporting by colleagues
- Plagiarism
    - Misrepresenting someone else's work as your own
    - Failure to cite sources
    - Safeguard = peer review

Studying Behaviour

Variables (anything that changes)

- Four general categories
  - Situational variables
    - Anything to do with the situation
      - Ex. Temperature of room, what color is the paper?
  - Response variables
    - The things the subject does (their behaviours)
    - Most often what we're measuring
      - Ex. Reaction time
  - Participant or subject variables
    - Characteristics within the person
      - Ex. Height, eye color, participants' actual intelligence level etc.
    - Who they are, not what they do
  - Mediating variables
    - Variable that get in the way- sometimes helpful, sometimes not
    - Psychological processes that play an intermediary role between what we are measuring and the response
    - Have an impact on our results

Situational Variable Response variability



Mediating Variable

→Diffusion of responsibility: the more bystanders, the less likely you will help or feel the need to help

Operational Definitions of Variables

- Defines the operations or techniques the researcher will use to measure or manipulate a variable (simplify: Description of how you will create or measure the behaviour)
- Determines the actual question
- Necessary for empirical study (we want consistency in methods)
- Help communicate ideas to others
- Construct: an abstract concept that must be translated into concrete forms of observation or manipulation
- Many ways to operationalize the same construct

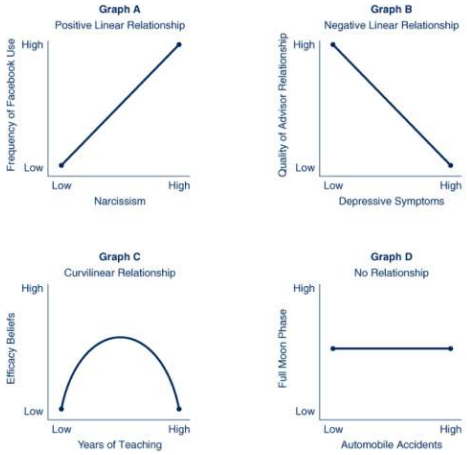
- Example: hunger, intelligence, aggression
    - Possible measures for aggression:
        - Instances of aggressive language/ profanity
            - Tone of voice may come into play
        - Physiological – muscle contraction or heart rate
        - Hitting frequency
        - Crime statistics

Non-Experimental vs. Experimental Methods

- Non-Experimental Method
    - No manipulation of variables
    - Description of behaviour
    - Correlational method
        - Measure covariation between two variables
        - No manipulation, just observation
            - No causation
        - Direction (positive-variables vary together or negative- vary in opposition[one goes up, the other goes down]) & strength (the stronger the relationship, the quicker they go up or down)
- Experimental Methods
    - Manipulation of variables
        - Guy who falls down stairs is dressed in a suit vs. hobo
    - Explanation of behaviour

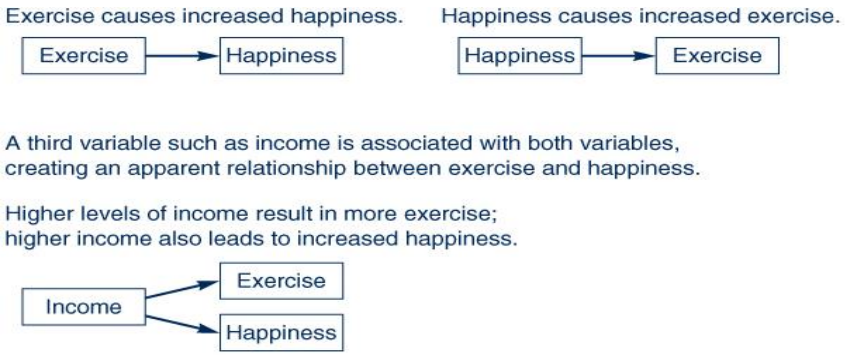
Relationship between variables

- Positive linear relationship: increase in one variable relate to increases in another
- Negative linear relationship: increases in one variable relate to decreases in another
- Curvilinear relationship: increase in one variable relate to both increases and decrease in another …as long as it is curved
- No relationship: changes on one variable do not relate to changes in another



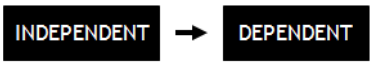
Interpreting Non-Experimental Results

- Cannot make causal statements- not enough info to say which comes first (fails 2 of 3 causal tests)
    - ~~Temporal precedence~~
    - Covariation of cause and effect (the only one we can show)
    - ~~Eliminate alternative explanations~~

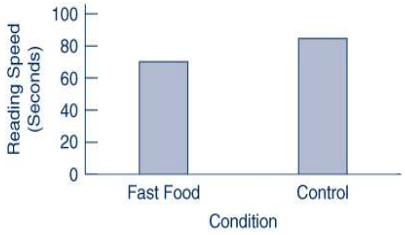


Experimental Methods

- **True experiments have TWO key features:**
    - Random assignment of subjects to groups
    - Manipulation of one variable to measure its impact on another
- Cause and effect relationships between independent and dependent variables
    - Independent variable always goes on the x-axis…



• Cause or effect?



- **Can** make causal statements
    - Establish temporal precedence
    - Attempt to eliminate alternative hypotheses
    - Experimental control & random assignment = eliminate confounds

Alternative explanations:

- Americans who have a glass of wine a day are healthier than those who have wine (or who have a lot of wine or other alcohol).
    - No random assignment… not conclusive
    - Things that may affect study in participants: lifestyle, income, social habits, exercise etc.

Causality & internal validity

- Internal validity: the ability to determine causation
    - Controls within the study
- Necessary & sufficient conditions (to show causation); rare in social psychology that all three points to show causality will be present
    - Reading the material and doing well on exam

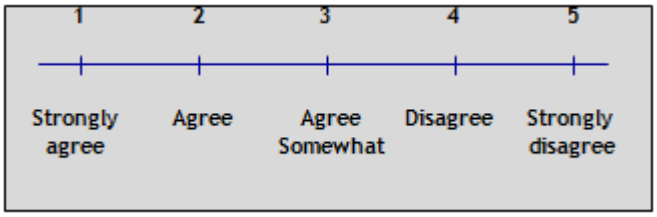
Experimental Method: Issues to consider. What is needed to decide between a non-experimental procedure and an experimental procedure:

- Artificiality of experiment (external validity- how realistic is it?)
    - Generally, the more the study is controlled, the less it relates to the real world
- Ethical and practical considerations
    - Some studies may be ethical, but not practical
- Participant variables cannot be manipulated
    - Ex. Can't manipulate intelligence
- Choice between causation or description or prediction when designing studies. What is your goal of the study?

Measurement Concepts

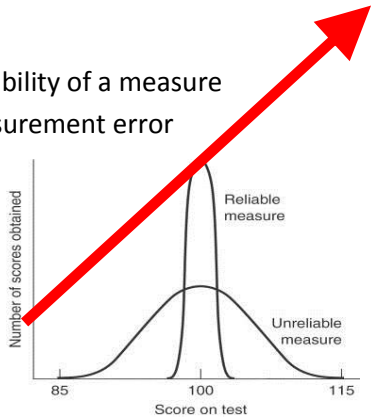
Self-Report Measures

- Typically questionnaires or scales
- Answer on a continuum anchored by polar opposites (e.g., Likert scale)



Reliability of Measures

- Consistency or stability of a measure
- True score + measurement error



Less error

- If it consistently reports the same score, we have a high confidence level that is accurate and reliable

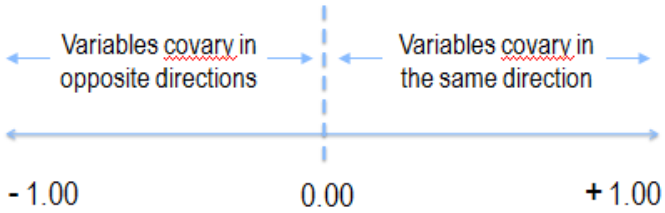
Achieving Reliability

- Training of observers
  - Aggression in Hockey
  - Making a checklist for observers- makes measures more reliable
- Wording of questions and instructions
  - Wording of question or task can change outcome
  - Maintain visual attention on the computer-generated image that will appear on the digital display in front of you and depress the appropriate key on the keyboard connected to the computer in front of you in accordance with the location of the computer-generated image.
    - Not a good way to word it. Too wordy and don't know which keys to press. Make it simple

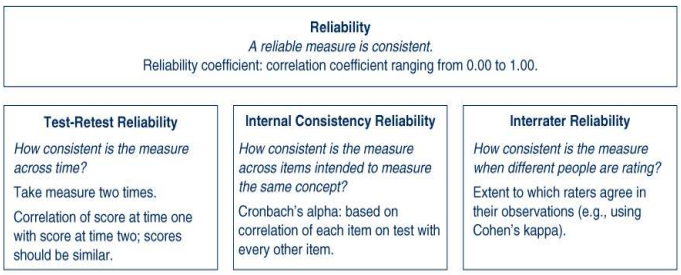
- Calibration & placement of equipment
    - Photo radar
- Multiple observations
    - Possibly have more than 1 observer
    - Multiple questions on a questionnaire that deal with one topic
    - Ex. Scales of the MMPI (personality test)

Assessing Reliability

- Pearson's product moment correlation coefficient (*r)*



- Test-related reliability- test once, test again in future, compare results
    - Assessed by measuring the same individuals two points in time
    - Vulnerable to artificiality
        - Practice effects
    - Alternate forms reliability
        - Another form of test-retest. Create 2 alternate forms of the test, of equal difficulty at different times
- Internal Consistency Reliability
    - Split-half reliability
        - not creating two forms of the test, just comparing different parts of **one** test
    - Cronbach's alpha
        - Takes all correlations and calculates them as a single score
    - Item-total correlations
        - Similar to cronbach's alpha, but create a total score for all items then compare each individual item to that score
- Interrater Reliability
    - Correlation between the observations of raters
    - *Cohen's Kappa*
    - Want positive correlation

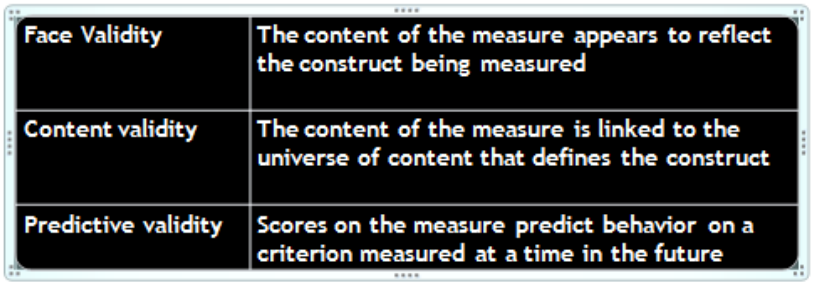


Reliability of measures

- Reliability and Accuracy of Measures
    - Reliability indexes indicate amount of error but not accuracy
    - A measure can be highly reliable but not accurate

Construct Validity

- The degree to which the operational definition of the measure reflects the construct
- Are you measuring what you say you are measuring
- Indicators of construct validity:

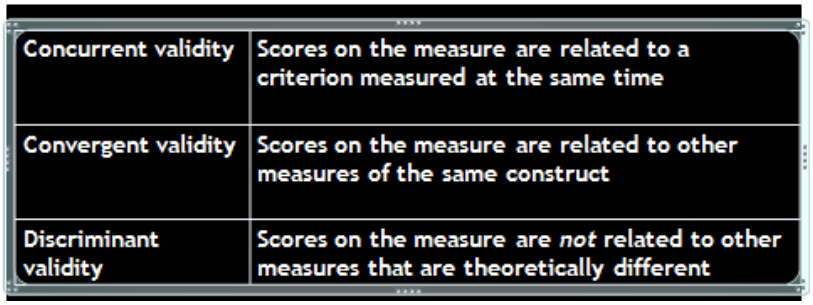


Face Validity	The content of the measure appears to reflect the construct being measured
Content validity	The content of the measure is linked to the universe of content that defines the construct
Predictive validity	Scores on the measure predict behavior on a criterion measured at a time in the future

-Face validity: Weakest form. Do you think it's valid?

-Content Validity: looking at the content of the questions and comparing that to the construct

-Predictive Validity: the ability of your test to predict the construct that you're interested in (ex. Can I predict your intelligence based on your IQ score?).



Concurrent validity	Scores on the measure are related to a criterion measured at the same time
Convergent validity	Scores on the measure are related to other measures of the same construct
Discriminant validity	Scores on the measure are *not* related to other measures that are theoretically different

-Concurrent Validity- given one test immediately following the other. Then comparing these two tests

-Convergent Validity: relating your measure to other measures

-Discriminant Validity- refers to the fact that scores on one of your tests SHOULD NOT relate to another test (ex. Anger and depression)

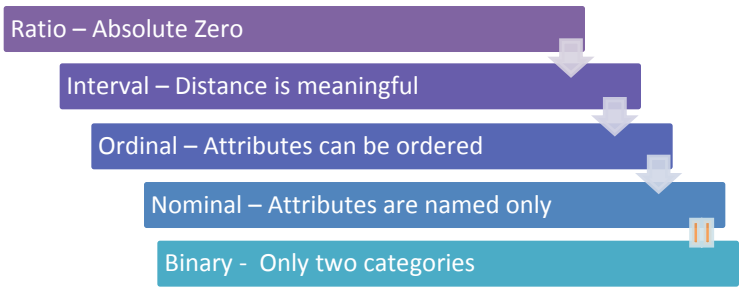
- Evidence for construct validity
    - The more methods used, the better
    - Sometimes a measure is valid in one context but not another
    - Must be tested and reassessed
- Fae- appears valid
- Content- content matches definition
- Predictive- predicts **future** behaviour
- Concurrent- relates to **current** behaviour

Reactivity of Measures

- Measure is reactive if awareness of being measured changes an individual's behavior
    - Does awareness of being measured change their behaviour?
    - Highly reactive is not good
- Measures of behavior vary in terms of their potential reactivity
- Can allow for familiarization (habituation) & use unobtrusive measures (ex. Hiding a camera)

Variables and Measurement Scales (NOIR)

- Nominal Scales (can't correlate)
- Ordinal Scales
- Interval Scales
- Ratio Scales



*As you move up the scale, everything above has the characteristics of the one below it

Nominal scale

- Categories with no meaningful numeric value
- Ex.: males/females; experimental condition/control condition
- Impossible to define any quantitative values or differences across categories

Ordinal Scale

- Rank ordering with numeric values
- Ex.: restaurant ratings; birth order; Olympic medals

- Has magnitude; values are smaller or larger than the next
- Interval between items not known

Interval Scale (ex. Thermometer)

- Has magnitude; values are smaller or larger than the next
- Interval between items is known and is meaningful
- No true zero point
- Ex.: Intelligence score, temperature

Ratio Scale

- Has magnitude; values are smaller or larger than the next
- Interval between items is known and is meaningful
- Has a true zero point
- Ex.: reaction time, duration of response, age

Measure vs. Dimension

- Has magnitude; values are smaller or larger than the next
- Interval between items is known and is meaningful
- Has a true zero point
- Ex.: reaction time, duration of response
- Could have a ratio scale (IQ test with a 0) but be measuring a construct that is actually interval (can't have 0 intelligence)

HOW something is measured can change the scale of what you're measuring

Observational Methods

Quantitative Approaches: Empirical

- Test theories in a controlled setting
- Collect numerical data
- Uses large samples
- **Analyze data via statistical methods**

Qualitative Approaches: Interpretive

- In-depth description of individuals in their natural setting
- Data = words rather than numbers
- Small number of participants
- Data are content analyzed for major themes

Naturalistic Observation

- Observations made in a natural setting over a period of time, using a variety of techniques
    - Ex. Want to see what children are like at recess, so go watch them play at recess
- Observe EVERYTHING, Change NOTHING
- Describe and understand how people live, work, and experience their particular setting
- A good start for hypothesis development
- Researchers become immersed in the situation (can be good or bad)
    - Goal = accurate description & interpretation
    - Strategies = observation, interviewing, and surveying documents
- Useful in complex and novel settings (rare/not common behaviour)
- Can gather data in real-life settings to inform quantitative studies
- BUT cannot study well-defined hypotheses with specific conditions (because there is no manipulation)
- Disadvantages
    - Time-consuming
    - Environment is ever-changing
    - Data interpretation is complex
- Issues:
    - **Participation** & loss of objectivity
        - Participant observation: Participating while observing, might create biases
    - **Concealment** & ethical concerns
        - As long as it's public, doesn't interfere with people, or contain sensitive information (although there are always ethical concerns with this)
    - Limiting the scope of the study- you can't analyze EVERYTHING, limit what you're looking for

Systematic Observation

- Careful observation of specific behaviors in a particular setting
- Less global than naturalistic observation
- Observations are quantifiable
    - Ex. Checklist
    - Ex. What percentage of boys compared to girls prefer this type of play?
- Testing specific a priori hypothesis- Have hypothesis already developed before starting observation
- Observations are summarized using a coding system
- Developed a priori- based on thoughts or expectations but being refined as the study goes
- Data will be more accurate if simple & easy to apply, and if you hire coders and don't tell them what the study is
    - You could potentially miss stuff happening if your coding system isn't simple because as you're writing down you're not observing
- Pre-established coding systems available
- Issues:
    - Reliability
        - Inter-rater reliability
            - 80% accuracy is high enough to trust data
    - Reactivity- will react the fact that they are being observed
        - Concealment
        - Habituation
    - Sampling
        - Method
        - Length of time
            - Time you need to get accurate data/ time for people to get used to camera being there so they act naturally

Case studies

- Description of an individual or event
    - Most often one person or event that is followed in depth
- *Psychobiography* –the use of psychological theory to explain the life of an individual
    - Often a historical individual
    - Growing in popularity
- Methods include library search, interviews, and sometimes direct observation
- Rich in information
- Valuable in informing us of conditions that are rare or unusual
- Can lead to the development of hypotheses to be tested
- Generalizability is limited
    - Can't generalize to the population if you only look at one subject

Archival Research

- Usually do not need any ethics approval because the research has already been done
- Involves using previously compiled information to answer research questions
- Data can be qualitative or quantitative
- Three major sources of data (and examples):
    - Statistical records
        - Public & private organizations
        - Public records
        - Major sports leagues
    - Survey Archives
        - Consortium for Political & Social Research
        - World Values Survey
        - General Social Survey
    - Written & Mass Communication records
        - Written: diaries, letters, speeches
        - Mass communication records: books, newspapers, magazine articles, TV programs, Facebook (but is it really private?)
- Content analysis of documents using a coding system- improves accuracy of data
- Can address questions that cannot be addressed otherwise
- Issues: Access to records & accuracy of information
- How do you access the info?

EXAMPLE

50 shades of grey impact on North American sexual practices archival research

Rope sales- 1000% increase in soft cotton rope

Riding crop sales- 100% increase

Handcuff sales- unspecified increase

Sex toy sales

- 30% overall increase

-Up to 200% on specific toys

-81% increase traffic to BDSM section of online stores

Asking People about Themselves: Survey Research

Qualitative: generally open ended questions (can still be sent out as questionnaires), but generally done as an interview as there are often follow up questions

Why conduct surveys?

- Provides a methodology for asking people to tell about themselves
- Provides useful information for making public policy decisions
    - E.g. MacEwan and the Govn't
- Influence decision as well
    - Lobbyists often cite surveys for policy change
- *Way easier to mail out 100 questionnaires opposed to wandering around to find data
- *Generally want surveys to reflect 'the truth' so we learn how to influence decisions to avoid ruining our data but some people can use this to create surveys that reflect the answers they want
- To study relationships between and/or among variables
    - Eg. Wealth and happiness
- Can serve as an important complement to experiment research findings

Response Sets

- Tendency to answer all questions in a particular manner
    - Yeah or Nay saying (more common when the person is unsure as to their answer)
    - Social desirability: "faking good" (tendency in surveys, ESPECIALLY interviews)
        - E.g. sexuality interviews of students; often lacks accurate info as they tend to mirror what they think the surveyor thinks
    - Malingering: "Faking bad"
        - E.g. if I do poorly on this test maybe I'll get more money, support, ect.
- Scales available to detect response sets
    - Flipping how questions are asked
- *Questions asked early on have an impact on the responses to later questions; remember priming
- *Can limit the response set by how the question is asked

Avoiding Response Sets

- Clear wording
- Avoid loaded (or offensive) terms (push you toward a certain answer because of emotional responses)
- Avoid leading questions; either purposefully or accidentally leading you toward a certain response
    - E.g. do you still beat your wife?
- Avoid "double barreled" questions;
    - E.g. do you think the govn't should increase spending on bombs and hospitals for children; yes or no?

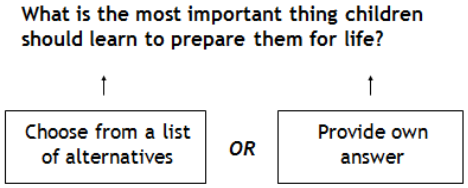
- Avoid double negatives; relates to clear wording
- Use mutually exclusive categories
- Reverse wording in some questions
- *Aim for a junior high reading level

Constructing Questions to ask

- Define the research objectives (e.g. guns) :
    - Attitudes and beliefs
    - Facts and demographics?
    - Behaviors?
    - *First option is to use a survey that has already been created
    - Attitude: How do you feel about gun ownership?
    - Fact: Do you own a gun?
    - Behaviour: do you hunt?
- Problems with question wording:
    - Unfamiliar technical terminology
    - Vague or imprecise terms
    - Ungrammatical sentence structure
    - K.I.S.S.- Keep it Simple, Stupid
    - Phrasing that overloads working memory
    - Embedding the question with misleading information
- Ensure simplicity & avoid:
    - Double-barreled questions
    - Loaded questions
    - Negative wording
    - Yea-saying and nay-saying
        - Reverse the question wording
    - *The first 3 generally have the biggest impact

Response to Questions

- Close-ended versus open-ended questions



- You have to initially choose what you want; can use both and/or switch back and forth, just means you are using mixed methods and effect data interpretation

Rating Scales (work better if you 'anchor' them)

University students should be required to pass a comprehensive examination to graduate.

1	2	3	4	5	6	7
Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

How confident are you that the defendant is guilty of attempted murder?

1	2	3	4	5	6	7
Not at all confident						Very confident

Have you ever smoked marijuana?

☐ Yes

☐ No

If yes, about how many times have you smoked marijuana?

☐ Once

☐ 2 to 5 times

☐ 6 to 10 times

☐ 11 to 20 times

☐ more than 20 times

Graphic Rating Scale

How would you rate the movie you just saw?

Not very enjoyable _____ Very enjoyable

- Allows a much more detailed score; but the weakness is that people generally feel less comfortable
- To score you can simply measure distance of the mark from the end of the line

Semantic Differential Scale

- **Evaluation (good/bad)**
- **Activity (low/high)**
- **Potency (weak/strong)**
- Based on the two anchor words these questions can assessed E, A, or P
- Have ONE open-ended question, and then asked to rate it based on a series of answers.
    - Ex. How do you feel about police? Good……..bad/ cruel……kind etc.

Non-Verbal scale

Finalizing the Questionnaire

- Should appear attractive and professional
    - Increases face validity
- Neatly typed and free from errors
- Use point scales consistently
- Ask interesting questions first

Administering Questionnaires

- In person to groups or individuals
- Mail surveys
    - Risk- low response rate
- Internet surveys
- Other technologies
- Less costly than interviews
- Ensures anonymity

  BUT…
- Understanding questions?- no way to clarify
- Boredom & distraction?

Interview Surveys

- Face-to-face interviews
- Telephone interviews
- Focus group interviews
- Problem: Interviewer bias
    - Sometimes given a script to eliminate biases

Survey Designs to Study Changes over time (longitudinal study)

- Questions are the same each time surveyed
- Tracks changes over time
- Panel Study: “two wave” (twice) or “three wave” (three times)
- Problem: Drop out

Sampling from a Population

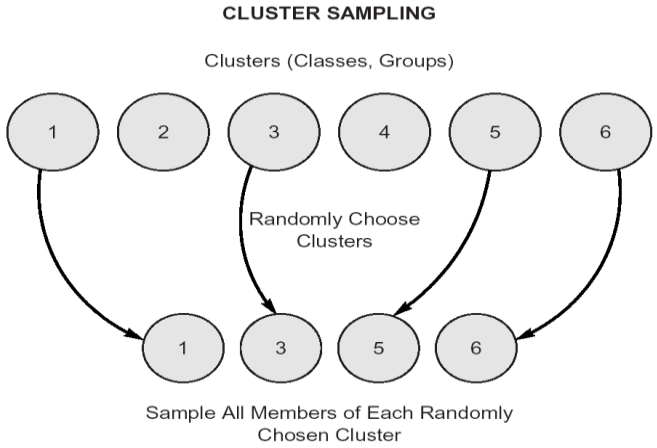
- Confidence intervals
    - Level of confidence that the true population value lies within an interval of the obtained sample (how confident you are that the value you got is actually in the population)
    - **Sampling error or margin of error**
- Sample size
    - A larger sample size reduces the size of the confidence interval
    - Must consider the cost / benefit of increasing sample size

Evaluating Samples

- Sources of bias
    - The sampling frame
    - Response rate

Sampling Techniques

- Probability sampling
  - each member of population has a specifiable probability of being chosen
  - **Simple random sampling**: every member of the population has an equal probability of being selected
  - **Stratified random sampling**: population divided into subgroups (strata) and random samples taken from each strata
  - **Cluster sampling** – identify clusters and sample from the clusters



- Non-probability Sampling
  - The probability of a member of the population being chosen is unknown
  - **Haphazard sampling** – convenience sampling (take people where you find them. ex. mall)
    - Less expensive; less time-consuming
    - Suitable for measuring relationships between variables.
    - BUT findings may not be generalizable – further replication with different samples is required.
  - **Purposive sampling** – sample meets predetermined criterion (ex. Women who have post-partum depression)
  - **Quota sampling** – sample reflects the numerical composition of various subgroups in the population

Theory

- Organize & explain
- Generate new knowledge
- General and abstract
- Parsimonious (simple)
- Falsifiable

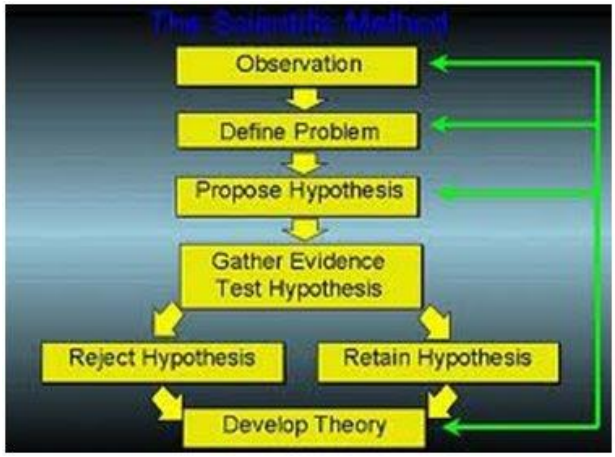
Hypothesis

- General statement about how things may be
- Describes relations between variables
- More concrete than a theory
    - Specific, but still general
- Testable
- Parsimonious (simple) & falsifiable

Prediction

- Specific prediction about what will happen here

The Scientific Method



Experimental Design

Confounding and Internal Validity

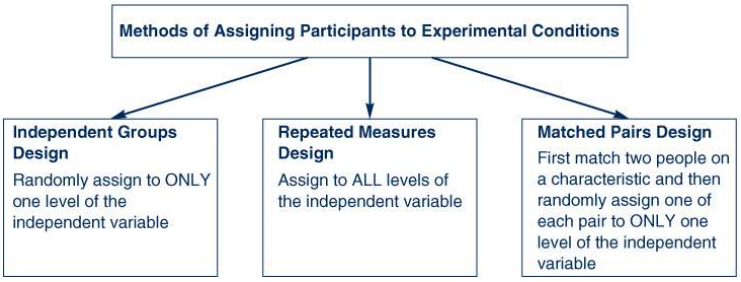
- Internal validity
  - Ability to make causal inferences
  - Ensure that only the independent variable can be cause of the results
  - True experiments have the HIGHEST degree of internal validity
    - They control for confounding variables
- Confounding variable
  - occurs along with the independent variable
  - Uncontrolled / unaccounted for
  - **Cannot determine which variable is responsible for the effect**

Basic Experiments

- 2 **EQUIVALENT** GROUPS
- CHANGE **ONE** THING
- **MEASURE** THE RESULT

Planning Basic Experiments

- Three basic steps:
  - Choose participants and assign them to the levels of the independent variable
  - Operationally define the independent variable
    - At least two levels (zero IS a level)
  - Operationally define the dependent variable to allow for comparisons
    - Has to be the same for both groups
- Three basic ways:



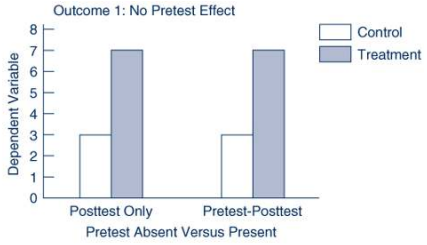
  - Independent- the group is exposed to the SAME thing and then compared to other groups
  - Repeated- same group under different conditions
  - Matched pairs- match 2 people on a specific characteristic, then split them randomly among the 2 conditions

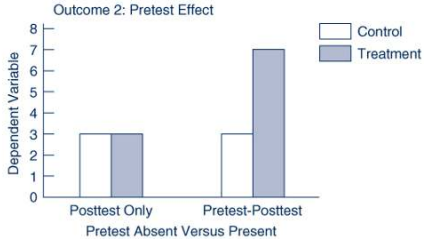
Independent groups design (between subjects)

- Different participants are assigned to each level of the independent variable
- Assignment to conditions is random

Pre-Test/ Post-Test Design

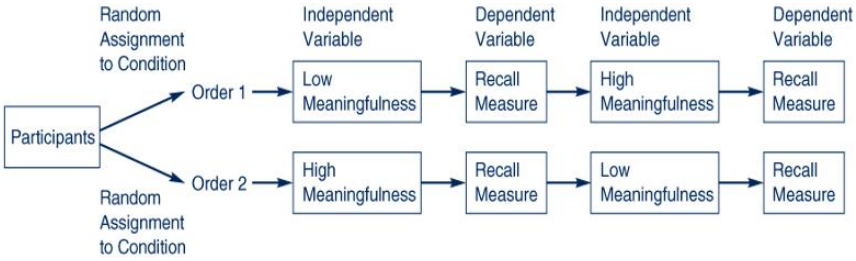
- Ensure equivalent groups
- A pretest is given to each group prior to introduction of the experimental manipulation
- Test to ensure that groups are equivalent at the beginning of the experiment
    - Vital to determine causation
- Is useful when:
    - Sample size is small; allows you to assess the equivalency of groups
    - When you need specific participants for the experiment
    - When the likelihood of mortality (dropout) is high
        - Know who dropped out
- Disadvantages
    - Can be time-consuming & awkward to administer
    - Can sensitize participants to what is being studied, thus creating demand characteristics
    - Can reduce external validity
        - Can be hard to generalize to the population, because your groups are very similar but not everyone else will be
- Solutions: for demand characteristics
    - Use deception
    - Embed the test with other irrelevant measures
- Solomon four-group design:
    - Assessing the impact of the pretest
        - Treat the pretest as a second independent variable
            - Half of the participants receive the pre-test and the post-test
            - Half receive only the post-test





Repeated Measures Design (within subjects)

- The same participants experience all levels of the independent variable
- One group experiences all conditions at different times
- Advantages:
    - Fewer participants needed
    - Extremely sensitive to statistical differences
    - Error variance is reduced because people serve as their own control group
- Disadvantages
    - The order of presenting the treatments may affect the dependent variable
        - Practice effects
            - Individual gets used to conditions over time
        - Fatigue effects
            - Individual gets tired of the tests over time
        - Contrast effects
            - Ex. Loud music may not affect you in one condition because you were already exposed to louder music beforehand
    - Solutions to order effects:
        - Complete (literally expose participants to EVERY possible order) or partial counterbalancing (don't expose to every order, but account for everything)
            - Complete: becomes complex



                - 2 conditions: 2x1= 2 orders
                - 3 conditions: 3x2x1= 6 orders
                - 4 conditions: 4x3x2x1= 24 orders etc.
            - Partial: Create the smallest number of orders that will allow each independent variable to occur first, last and to follow each other.
                - ABCD
                - BDAC
                - DCBA
                - CADB
        - Spacing time intervals
            - A rest period may counteract fatigue effects and contrast effects
            - May result in greater dropout
            - May add confounding variables

Choosing a Design

- One is not superior to the other
- Both designs have advantages and disadvantages
- Consider generalizability of results
- Consider reversibility of effects

Matched Pairs Design

- Match people on a participant characteristic
    - The dependent measure

      Or
    - A variable strongly related to the dependent variable
- Ensures participants are equivalent on the matching variable prior to introduction of the independent variable
- Useful when:
    - **Few available** participants
    - **High cost** per subject

Conducting Experiments

Conducting experiments

- Select Participants
- Choose Manipulations for independent variable
- Choose Measurements for dependent variable
- Establish control
- Collect
- Analyze
- Communicate

Selecting Research Participants

- Draw samples from the population
    - probability sampling
    - nonprobability sampling
- Determine sample size: larger samples provide more accurate estimates of population values

Manipulating the Independent Variable

- Setting the stage
    - Obtain informed consent
        - Debrief?
    - Decide on straightforward or staged manipulation of independent variable
        - Relates to operational definition
    - Decide on mundane (real world) versus experimental (lab) realism
    - Ensure manipulation is strong enough to detect differences
    - **Maximize the differences** between the experimental and control groups
    - Consider the cost of the manipulation
- Three types
    - Self-report measures (scales, questionnaires)
    - Behavioural measures (observation)
    - Physiological measure- completely unbiased
        - Galvanic skin response (GSR)
        - Electromyogram (EMG)
        - Electrocardiogram(ECG)
        - Electroencephalogram (EEG)
        - Functional MRI (fMRI)
- Issues to consider
    - Sensitivity of the dependent variable
    - ceiling effects- test isn't sensitive enough
    - floor effects – test it too sensitive

    - Multiple dependent measures
        - Order effects?
        - Valuable but not always feasible
    - Cost considerations

Additional Controls

- Participant expectations
- Experimenter expectations
- Controlling for participant expectations
    - Demand characteristics
        - Deception
        - Camouflage- advantage- no ethical concerns
    - Placebo/ nocebo effects
        - Use a single blind placebo group
- Controlling for experimenter expectations
    - Experimenter bias or expectancy effects
        - Treat participants differently
        - Record or interpret behaviors differently
    - Solutions:
        - Training→consistency
        - Automated procedures
        - Use of double-blind experiments

Additional considerations

- Pilot studies
    - Trial run→test the process
- Manipulation checks
    - Directly measure the effect of the independent variable
- Debriefing

Analyzing and interpreting results

- Use statistics to analyze data
- Examine and interpret the pattern of results
- Decide whether the relationship between the independent and dependent variables is significant

Communicating research to others

- Journal articles
- Professional meetings
    - Presentations, Seminars

Increasing the levels of the independent variable (more points of data, perhaps smaller scale on graph)

- Provides more information about the relationship between the independent variable and the dependent variable
- Allows you to check for curvilinearity
- Allows you to compare the effects of more than two groups

Development research designs

- Longitudinal
    - Same group is observed at different times (as they age)
    - Can be days, weeks, or years
- Cross-sectional method
    - People of different ages are measured at the same point in time
- Sequential method
    - Combination of both designs

Choosing a Design

- One is not superior to the other
- Both designs have advantages and disadvantages
- Consider generalizability of results
- Consider reversibility of effects

Matched Pairs Design

- Match people on a participant characteristic
    - The dependent measure

      Or
    - A variable strongly related to the dependent variable
- Ensures participants are equivalent on the matching variable prior to introduction of the independent variable
- Useful when:
    - **Few available** participants
    - **High cost** per subject

Factorial Designs

- **More** than one independent variable (or factor)
    - Simplest factorial design: 2x2 factorial design
        - Has **two** independent variables, with two levels each

Why Factorial Designs?

- Build on previous research
- Control for threats to validity
- Enhances information

2x2 Factorial Design

- CHART ON CANDIES…know how to read the graphs and do the math

Factorial Designs

- Interpretation of factorial designs (essentially, yes or no, for each)
    - Main effects of an independent variable
    - Interaction between the independent variables
        - If you draw it out, the lines will cross or not be parallel
- Factorial designs with manipulated variables and non-manipulated variables
    - IV x PV: independent variable by participant variable
- Moderator variable
    - Moderates (influence) the relationship between 2 other variables

- Interactions and moderator variables
- A 2x2 factorial design can yield one of eight combinations of main effects and interactions
- Assignment procedures
    - Independent groups design
        - Use different people for each variable
    - Repeated measure design
        - Use the same people for each variable
    - Mixed factorial design using combined assignment
        - Use different people, but have to keep the groups for certain variables the same.
- 2x3 factorial design (2 variables; one has 2 levels, & one has 3 levels)
    - chart
- 2x2x2 factorial design (3 variables; they each have 2 levels)

Complex Factorial Designs

- Higher order factorials: more than 2 independent variables
- The more complex the design, the more difficult it is to interpret

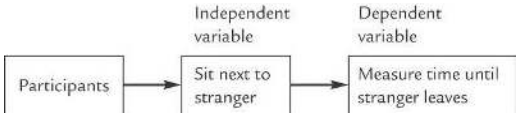
Penultimate

Single case

- Measured from baseline period to treatment periods
- ONE person each time
- Difficult to account for confounding variables
    - Reversal (ABA) or withdrawal design; baseline [remove treatment]
    - A=baseline
    - B=treatment
- Causation
    - Can't say anything about causation
- Generalizability
    - Hard to generalize to the population
- ABA designs can be improved by extending them to ABAB or ABABAB designs
- Benefits
    - Increased power
    - More ethical- leaving them with the treatment instead of how you found them if it does work
- Multiple Baseline Designs
    - Multiple measures taken before and after the manipulation
    - Treatment is effective if change in behavior occurs following manipulation, under multiple circumstances
    - Introduce treatment at different times for each subject. Excludes external reasons for changes
    - Can be measured across:
        - [multiple] Subjects
        - [multiple] Behaviours
        - [multiple] Situations
- Replications
    - That it is possible to replicate procedures
        - Strengthens support for effectiveness
        - Improves generalizability
        - Report individual rather than group data (ex. don't average results)
    - Cannot test hypotheses using statistical procedures, however

Quasi-Experimental Designs

- Used when the control (e.g., random assignment) of true experiments cannot be achieved
- Lower Internal validity than true experiments
    - Less ability to make casual statements
- Use only when true experimentation is not possible

- One-group PostTest only design
    - Compare the responses of a number of individuals exposed to the same event
    - No control variable
    - 

    - Bias: Frequently used to show something worked instead of exploring the fact that it might not
- One-group PreTest-PostTest design
    - Add a baseline measure to provide a basis for comparison
- Threats to internal validity
    - History effects
        - Impact of anything in the environment outside of your study
    - Maturation effects
        - Subjects naturally mature and change as time goes on
        - Internal changes in subject
    - Testing effects
        - Impact of the tests themselves
        - The simple act of measuring causes a change in behaviour
    - Instrument decay
        - Often, as people use the specific testing instrument, they get sloppy over time
    - Regression toward the mean
        - A statistical principal that is true
- Overcoming threats to internal validity
    - Control group
        - An equivalent control group is preferable
        - A non-equivalent group is frequently used
- Non-equivalent control group design
    - Separate control group introduced
        - Not equivalent due to lack of random assignment
        - Selection differences
- Non-equivalent control group PreTest-PostTest design
    - An improvement on the nonequivalent control group design
    - Addition of pre-test improves internal validity
    - Can assess for differences prior to testing
- Interrupted Time Series Design
    - Examines the dependent variable over an extended period of time, both before and after the IV is implemented
    - Longitudinal

- Control series Design
    - Improves interrupted time series design by finding an appropriate “control group”
    - Involves finding a similar population that did not receive a particular manipulation
        - Ex. Find a province that doesn’t have the blood-alcohol level law and compare it to one who does and look at traffic fatalities

Understanding Research Results: Description and Correlation

Analyzing research Results

- Three approaches
    - Comparing group percentages
    - Correlating individual scores
        - Relationship between variables
    - Comparing group means
        - Is the difference between the means statistically significant?

Four types of graphs

- Pie charts
- Bar graphs
- Frequency polygons
- Histograms
    - Interval or ratio data, not mutually exclusive

Frequency distributions

- Useful for several reasons:
    - Visual display of responses
    - Shape of distribution
    - Identification of outliers
    - Group comparisons

Important considerations

- Restriction of the range
    - Sample as much as you can→ bigger sample size

Effect Size

- Refers to the *strength of association* between variables
- Pearson *r* is one indicator of effect size
- Reporting effect size provides a scale of values that is consistent across all types of studies
- Differences in effect sizes
    - Small effects near $r = .15$
    - Medium effects near $r = .30$
    - Large effects above $r = .40$
- Squared correlation coefficient
    - Squared value of the coefficient r^2 - transforms the value of *r* to a percentage

    - Percent of shared variance between the two variables
- Cohen's d

$$d = \frac{M_1 - M_2}{\sqrt{\frac{SD_1^2 + SD_2^2}{2}}}$$

    - Used in experiments with two or more treatment conditions
    - Describes the magnitude of the effect of the IV on the DV
    - Cohen's *d* expresses effect size in terms of standard deviation units

Understanding Research Results: Statistical Inference

Inferential statistics: Samples and Populations

- Inferential statistics are necessary to determine if the results of a given study extend to the larger population
- Do the observed sample mean differences reflect a true differences in population means?
    - How trustworthy are our results?
    - Is it true, or was it just a fluke?

Inferential Statistics: Ruling out Chance

- Differences in sample means reflects the effect of the independent variable plus any random error
- Testing for significance allows us to determine the extent to which results are due to a **true effect** or likely **due to error**

Selecting the appropriate statistical test

IV	DV	Statistical Test
Nominal Male-Female	Nominal Vegetarian - Yes / No	Chi Square
Nominal (2 Groups) Male-Female	Interval / Ratio Grade Point Average	*t*-test
Nominal (3 groups) Study time (Low, Medium, High)	Interval / Ratio Test Score	One-way ANOVA
Interval / Ratio Optimism Score	Interval / Ratio Sick Days Last Year	Pearson's correlation

Inferential Statistics: An Overview

- Significant: mean differences due to true effect
- Non-significant: mean differences due to chance
- Significance level: chances of being wrong about your inference

Null and Research Hypotheses

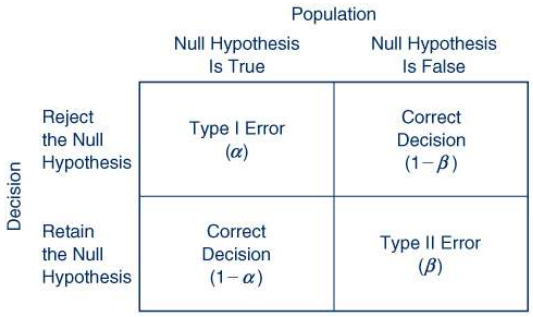
- Null Hypothesis- nothing happened/no difference
    - H_0: Population means are equal
- Research Hypothesis- what you predict
    - H_t: Population means are not equal

Probability and Sampling Distributions

- Probability: the likelihood of some event or outcome
- Statistical significance decisions are based on sampling distributions
    - Predicting what the population mean will be

- Sampling distributions reflect the probability that the null hypothesis is true
    - Based on set probability outcome

Decision Matrix



Type I Errors

- Made when the null hypothesis is rejected but the null hypothesis is actually true
- You claim that population means are likely to be different when in fact, they are not
- Probability of a Type I error is reflected in the alpha level

Type II Errors

- Made when the null hypothesis is accepted when it is in fact false
- You claim that population means are the same when in fact, they are different
- Factors related to making a Type II error:
    - alpha level
        - If you set your p-value lower, the more likely you will make a type II error
    - sample size
        - The larger the sample size, the risk for type II error goes down
    - effect size
        - the bigger the effect size, the easier it is to detect
        - the smaller the effect size, the more likely you are to make a type II error as it is easy to miss
- The more you reduce the likelihood of a type I error, the more likely you are to get a type II error

Interpreting <u>Non-Significant Results</u>

- Accepting the null hypothesis versus failing to reject the null hypothesis (Same thing- nothing happened. Just different ways to say it)
- Non-significant results can be attributed to a number of factors other than the true absence of an effect
    - Procedures
        - How did I measure? Operation definition?
        - Most likely the instructions weren't understood. Can fix with a pilot study!

- Alpha level
  - Maybe it was too low OR high. Typically 5%
- Sample size
  - Too small→get larger sample
- Effect Size
  - May not be a big difference between means, may just need to take a closer look to verify results.

Choosing a Sample Size: Power Analysis

- A method for determining the optimal sample size based on the probability of correctly rejecting the null hypothesis- what is the optimal sample size based on our alpha value?
- Power= % chance that you have of NOT making an error
  - If Power=80, you have an 80% chance of not making an error and a 20% chance of making an error

Power = 1 - *p* (Type II error)

Choosing a Sample Size

- Total sample size needed to detect a significant difference for a *t test*

Population effect size *d*	Power = .80	Power = .90
.20 (a small effect)	788	1054
.50 (a moderate effect)	128	172
.80 (a large effect)	52	68
1.00	34	46
1.50	18	22

Computer Analysis of Data

- Software Programs include: **SPSS**, SAS, R, and Microsoft Excel
- Steps in analysis:
  - input data
  - conduct analysis
  - interpret output

Considering More than the Null Hypothesis

- Hypothesis testing is pass or fail
- This leaves out interesting information about the data, such as **effect sizes** and **confidence intervals**

Importance of Replications

- Scientists attach little importance to results of a single study
- Detailed understanding requires numerous studies examining same variables
- Researchers look at the results of studies that replicate previous investigations

Generalizing Results

Replication and Generalization

- Exact replications
  - An attempt to replicate **precisely** the procedures of a study
  - Ascertain whether the same results are obtained with replication
- Conceptual Replications
  - The use of different procedures to replicate a research finding
  - The Independent Variable is manipulated in different ways than in the original study
  - The Dependent Variable can also be measured differently

Evaluating Generalizations

- Literature Review
  - Summarizes what has been found
  - Tells the reader what findings are strongly/weakly supported
  - Exposes inconsistent findings and areas lacking proper research
  - Discusses future directions for research
- Meta-analysis
  - Method for determining the reliability of a finding by examining the results from many different studies
  - Researcher pools actual results from other studies, which are then analyzed statistically
    - Your subjects are other people's study results

Generalizing to other Populations

- Participants typically not selected from the general population
- They are usually university students
- How generalizable are these results?
- University Students
  - Knowledge of psychology
  - Similar age group: early adulthood
  - High level of intelligence
  - W.E.I.R.D.
    - Western, Educated, Industrialized, Rich, and Democratic
- Volunteers
  - Different from non-volunteers
  - More highly educated
  - High level of conscientiousness
  - Individual difference among volunteers

- Sex Considerations
    - Is the study mixed or single-gender?
    - Are there gender biases in the questions asked?
    - Are both genders interpreting questions in a similar manner?
- Cultural Considerations
    - Racial and ethnic composition of the sample
    - Used to be only white Caucasian; now more culturally diverse
    - Operational definitions can be influenced by culture

Generalization as Statistical Interaction

- The problem of generalization can be thought of as a statistical interaction
- Include the subject variable as another independent variable in the study
- No interaction = generalizability

In Defense of University Students and Rats

- We have learned and continue to learn a great deal from animal models
- University students are increasingly diverse and representative of society
- Replicability of findings

Generalizing to other Experimenters

- Experimenter's influence on participants must remain constant throughout the experiment
- Experimenter personality characteristics and gender can also affect participants
- Solution: use two or more experimenters

Pretests and Generalization

- Should a pretest be given?
    - Helps assess possible mortality effects
    - Can use a Solomon four-group design to assess any interaction between the IV and the pretest variable

Generalizing from Laboratory Settings

- Controlled but artificial environment
- Field experiments can increase external validity
- Lab and field studies contribute to a greater understanding of human behavior