Research in the Behavioral Sciences

The Goals of Behavioral research

Description

Necessary first step but very limited in what you can deduce from a description. Descriptive research needs follow-up.

Predicting

Attempt to identify the predictable relationships that underlie behaviors so as to effectively predict an outcome given some data.

E.g., predicting recidivism or class performance or likelihood to commit suicide or successful employee.

Explaining

Identification of the **causal** relationships between behaviors, environmental factors, physiological states, etc.

Controlling behavior

Once you know the causes, you might be able to control behavior. This is important in situations like trying to control the occurrence of suicide, or classroom performance, or self-injurious behaviors.

The Scientific Approach - three criteria for a "scientific "study

Systematic Empiricism - objectivity and observability

Direct vs. indirect observation

Systematic => Structured observation

Public verification

reproducible - the importance of replication

others must have access to your data, your analyses, your methods

Testability

reproducible (can't answer the question - did we evolve from apes? but can answer - are their fossils that could be ancestors of humans that are more apelike?)

measurable (does God exist?)

falsifiable (Freud)

Measurement - defining the things you're measuring

operationalization of terms = operational definitions

e.g. hunger = # of french fries consumed in 5 minutes, or amplitude of stomach growling, or amount of saliva produced while watching an ad for Red Lobster, or subject's rating of hunger on a scale from 1 to 7.

IMPORTANT NOTE!

CLASS EXERCISE

The Role of Theory in Science

A theory is an educated guess

'educated' because it relies on observations, i.e. data

'guess' because it's not guaranteed to be correct - not 'True'

Theory follows from data and data is predicted by theory

Deduction vs. Induction

A good theory is…

Falsifiable

Parsimonious

The Use of theory - Explaining vs. Predicting behavior - a priori vs. post hoc explanations

Hindsight is 20/20

The Limits of Science - proof and disproof, truth vs. Truth.

Several years ago some bones were discovered in South America and it was claimed that they were the bones of the notorious Nazi Doctor Mengele.

Ask students what they would do to determine if the bones were indeed those of Josef Mengele.

The logical impossibility of proof

It doesn't matter how many 'positive' instances of a truth you find, all it takes is one previously unknown negative instance to refute your 'truth.'

Also, you can't prove the antecedent (a theory) from it's consequents, but you can disprove a theory from failed consequences.

The practical impossibility of disproof

Let's say you have evidence that a prediction of a theory doesn't hold true. Have you disproved the theory? Not necessarily.

truth vs. Truth

So, science does not provide us with Truth answers. Good theories are 'true' most of the time (i.e. make correct predictions). Ex: Newtonian vs. quantum mechanics. Newton has now been shown to be 'literally' wrong, but his theories still have great validity.

'facts' are not facts but rather instances of 'lots of evidence show that'... Is it a fact that you were born on X? No, but lots of reliable sources provide evidence of its truth.

AVOID: always, never, prove, disprove

support vs. lack of support = weighing the evidence.

Strategies of Behavioral Research

Descriptive research

Doesn't answer "why" or "how" only "what"

e.g. public opinion polls

Correlational research

Study of factors associated with a behavior

Answers whether there is a *relationship* between factors

Doesn't address what the "causes" are

Poster - drinking and grades

Experimental research

- Attempt to determine causes
- Experimenter manipulates one or more IVs to see whether changes in a DV occur.
- Ethical considerations of 'true experiments'

Quasi-experimental research

- Can't manipulate IV (e.g. for ethical reasons), so you study naturally occuring changes
- Natural changes often examined within a subject

Behavioral variability

Behavior varies across individuals, that's a fact of life.

Our task is to understand that variability

Does it systematically vary with other characteristics?

- personality characteristics
- personal attributes (weight, intelligence, head size)
- gender
- age
- situation (high stress, soft background music, at home vs. in lab)
- task difficulty

Research should be designed in a manner that best allows the researcher to answer questions about behavioral variability

The measurement of behavior involves assessment of variability

Statistical analyses are used to describe and account for observed variability

- descriptive statistics - summarize (i.e. describe) the data, e.g. using means, standard deviations, graphs, figures.
- inferential statistics - make conclusions or inferences about the generality of your results, or whether something happened 'by chance' or because of a true underlying difference.

Two words of caution

Which of the strings HHTHTTHHTH, and HTTTTHTHH looks more like it happened due to chance?

- People tend to thing that random strings of events must be 'locally' random
- e.g. How many times have you played a night of Uno, Bridge, Gin, or some other game and lost almost all night only to turn around a few weeks later and win almost every hand? This is very likely to occur, even when all opponents are of equal skill.
- *Chance is lumpy*

People have a tendency to underestimate the chance factors involved.

- People prefer false certitude to the recognition of chance variability.
- *Overconfidence abhors uncertainty*

Measuring variability - a review

Show some various types of frequency distributions (normal, skewed, bimodal, multimodal, amodal).

Variance is a statistic that summarizes the amount of variance but not the type

Like all statistics, information is lost when it is summarized.

variance

Relative to a mean or average value

Steps to calculate:

find the mean $\bar{x} = \frac{\sum_i x_i}{n}$ "x-bar"

find deviation scores (for all X, find X - mean)

Square the deviation scores to get a direction independent measure of deviance

Add the squared deviations (sum of squares) and divide by n-1

$$s^2 = \frac{\sum_i \left(x_i - \bar{x}\right)^2}{n-1}$$

RECALL: standard deviation is square root of s^2

Take some class heights and measure variance and means.

Raw scores on board marked male/female.

LIKE ALL STATISTICS, THIS IS A SUMMARY - INFO IS LOST.

Total variance = systematic variance + error variance

Systematic variance = that part of the total variability related in a predictable fashion to the VARIABLES UNDER STUDY.

"Error" variance = that part due to factors not being investigated or controlled for.

Recalculate above but do means/variance by gender - note how much lower the variance is within each group.

total var = that calculated earlier

remaining, unaccounted for variance (err. var) = *weighted* average of the var. for the remaining groups.

Systematic = total - error

Controlling sources of variability to decrease the amount of 'error' variance

e.g. gender may have an effect on cognitive style, but if we're interested in SES's effect, we want to "control for" gender

holding other variables constant (e.g. use all of one gender - this can create generalizability problems) = reduces variation

choosing equal amounts of each value of the other variables (e.g. half male, half female) = this does not reduce variability randomly choosing variable values (increases generalizability to the entire population from which sample is taken) = also does not reduce variability

However - in the latter two cases, you can statistically control for group differences

Reduce Errors

Assessing the strength of relationships

Proportion of total variance "accounted for"

Proportion = systematic/total variance

values are always between 0 to 1 inclusive (or 0% and 100%)

Behavioral research is notorious for small amounts of variance accounted for.

Cohen's criteria: small = .01, medium = .06, large >.15

Meta-analysis

Metacognitition - thinking about your thinking

Metamemory - remembering what you remember

Metaanalysis - analyzing analyses

Quantify the effect sizes observed in various studies = dv

Note differences between studies = ivs.

Analyze dv = ivs.

Psychology in perspective among the sciences

There is more variability, more sources of error variance, i.e. more uncontrolled for factors.

LOTS of things impact human behavior making it extremely difficult to predict.

This means that the effect sizes, e.g. correlation values, that are acceptable in psychology are often much lower than those considered acceptable in other sciences.

Measuring Behavior

Types of measures

Behavioral measures

WHAT SUBJECTS **DO**

these are direct measures

Physiological measures

WHAT SUBJECTS' **BODY STATES** ARE

EEG, GSR, ERP, blood hormones, spit hormones, MRI, PET

Self-reports

WHAT SUBJECTS **SAY** THEY **DO/THINK/FEEL**

cognitive self reports (incl. verbal protocols)

behavioral self reports (how often do you read the newspaper, lose something, forget names)

affective self reports (how do you feel - stressed, depressed, tense, happy)

IS THIS INTROSPECTION? Not if these reports are treated as 'data' rather than 'fact'

Converging operations: use multiple measurements to make the strongest arguments.

Scales of measurement

Categorical/discrete vs. Continuous variables

nominal scale - simple 'naming': no specified order

religious preference, key choice, NBA conference, computer used

ordinal scale - labels specify an order ('ordinal').

e.g. star rating system on restaurants

e.g. class rank in HS graduating class

e.g. star rating on some movies

e.g. degrees of dryness in wines (doux, demi-sec, dry, extra-dry, brut.

will one part doux, one part brut give you 'dry'???

interval scale - specifies the relative distance among ratings, i.e. 'how much' is specified.

e.g. IQ, Fahrenheit temperature, shoe size, Likert scale. Makes sense to say it's 10 degrees warmer than yesterday, but not to say 'it's twice as hot as yesterday'

e.g. Time-of-*day*

ratio scale - able to multiply and divide the quantities

e.g. weight, height, Reaction Times, error rates.

e.g. time elapsed (NOT time-of-day)

Permissible statements:

nominal: $A \neq B$

ordinal: $A > B$

interval: $(A-B) = (C-D)$

ratio: $A = 2*B$

MOST STAT MEASURES you will use (e.g. ANOVA, t-test) require at least the use of interval scales.

Chi-square, loglinear, etc. for categorical data.

Reliability vs. Validity

"watch" example - reliable, always shows the same time when a certain event occurs (e.g. the rising of the sun). But could be an hour fast = invalid.

reliable and valid - always shows the "right" time

Reliability of measurements

Defn: Consistency or dependability of a measure

Observed score = True score + Measurement error

ASSESSING RELIABILITY, you must assess the amount of variability in observed scores due to true score differences and what amount is due to measurement variability.

TEST-RETEST RELIABILITY: Take same individual and test multiple times. Amount of variability between scores is due to error variance.

INTERITEM RELIABILITY: One way: split-half reliability. If total measure (e.g. IQ or RT or error rate) matches well when done on different halves of the data (randomly chosen), you have good interitem reliability.

INTERRATER RELIABILITY: If measurement is via observation, two different observers' measurements should correlate well.

Validity of measurements

Defn: generally refers to the extent to which a variabile has been adequately measured AND manipulated

Examples

To assess knowledge of course material, use a "test" that includes exams on research methods, not "head size"

Head size as a valid measure of intelligence? NOT! Consider male/female differences

Shoe size as a measure of knowledge? (consider children)

Various forms of validity

EXAMPLES (there are many more ways to assess validity, I'll talk about a couple)

construct validity - Measurement must assess the thing you believe you are measuring – THESE FORMS FOCUS ON MEASUREMENT

Internal and External validity – will return to later…

Task performance range

So, you have a valid and reliable measure. Are you set? No... you must take care that your task performance range is appropriate.

ceiling effects

floor effects

Bottom line - choose a task for which you will see a range of performance for ALL of your subjects.

Describing and Presenting Data

Graphical representations

Tables

provide more detail than graphs

careful not to provide too much detail

4 rules for table construction

stem and leaf diagrams

A quick way to get a feel for the distribution of the data

1. Scan the data to get a feel for the range and level of stems required.

2. Write down the stems in order - use a consistent level of stems

3. Go through each datum and place it as a leaf on the stem - you may have empty stems

NO RIGHT OR WRONG WAY TO DO IT, ONLY GOOD AND BAD WAYS

Back-to-back stem and leaf diagrams

frequency distributions

Summary of the distribution of values

simple

grouped

frequency histograms

bar graphs vs. line graphs

Histograms are bar graphs

Can also use bar graphs to show average value of dv. by each iv (as most did on their projects)

Use bar graphs to illustrate relative sizes

Can also represent anything in a bar graph as a **line-graph**

Use line graphs to emphasize trends

Scatter plots and enhanced scatter plots

Applicable when you have two interval/ratio variables

May be informative when one or both variables are ordinal

They emphasize mathematical relations between the variables.

Enhanced scatterplots

How to mislead with graphs

non-zero y-origins

split one graph into two or more using different scales on each

plot average values rather than raw data (reduces perceived variability)

transformed scales (logarithmic)

Rules for constructing graphs (all methods above)

1. Make the data stand out!

2. Graphing should be an iterative and experimental process.

3. Avoid putting notes and keys in the data region (keeps the data from standing out).

4. Choose the scales so that the data fill up as much of the data region as possible (be careful with non-zero origins and misleading the reader, though).

5. Use a consistent scale when graphs must be compared

Central tendency

Methods for summarizing data using a single number

Mean

You know how to compute this

What is it? Imagine a balance where you hang weights at various points along the scale. The mean is the place to put the fulcrum so that everything balances out

- Note that the mean is sensitive to outliers
- Means can be very misleading when the data is skewed or has a few severe outliers

Median

- The midpoint where half of the values are above this value and half below
- How to find median
- To avoid sorting...
- Conceptually

Mode

- Find peak in the data - the most commonly occuring data element (or groups of elements - e.g. in a grouped histogram)
- Computing - just look for the dv. value with the greatest frequency
- THIS IS THE ONLY MEASURE OF CENTRAL TENDENCY THAT IS APPROPRIATE IF A NOMINAL SCALE WAS USED!
- This measure is rarely used.

Mean vs. Median

- In the summer of 1994 there was a baseball strike; some of the salary data appearing in the papers provide a nice example of how to select statistics favorable to support your case. The owners kept referring to the typical player salary of $1.2 million per year, clearly implying that the amount was more than enough for survival. The players, however, pointed out that the typical annual salary was actually a mere (mere?) $415,000.
- Be wary of this with tax information, reported salary levels, reaction times, etc. anything with a highly skewed distribution.

Variability

Range

- Max - Min
- Very susceptible to outliers
- Of limited use

Interquartile range

- 3Q - 1Q
- 75%ile - 25%ile
- Insurance against outliers
- Tells you about the middle 50%
- Demonstration using sorted data and using stem & leaf diagram

Variance

- Functions like an average of squared deviations
- Conceptual issues

Standard Deviation

To make the measure of variability understandable - consumable - take the square root to make it in the original units.

Variance vs. Standard Deviation

St. Dev. is measured in same units as the data themselves (rather than inches^2, for example)

St. deviation is a measure of 'average' distance data values are from the mean.

Can be used to determine how unusual a particular value is (e.g. data values more than 3 s.d.s away from the mean are very unlikely - see discussion below on z-scores).

HOWEVER, Variance has the nice property of additivity noted in the last project - you can apportion sources of variance in an additive fashion (e.g. systematic + error = total).

Normal distributions

The standard deviation is especially helpful when one knows that the data is "normally distributed"

Standard Error

This tells you how much variability there is in your sample mean.

This measure recognizes that your estimate of a sample mean is affected by the size of your sample.

std. error of the mean = $s/\sqrt{n}$

Standard error bars

95% Confidence Intervals for a mean

Comparing apples and apples (z-scores)

e.g. how do you determine if gender salary differences are larger or smaller in the US, UK, Sudan, or Columbia?

AKA standardizing the data

Calculation:

$z = (x-\mu)/\sigma$

Note the use of population statistics, not sample statistics…

Large z-scores will crop up when you have a very large sample size; should see transition toward them, though.

The 5 point summary and box plots

The 5 point summary is a nice method for summarizing your data without going to a single value.

Min, 25%ile, 50%ile (median), 75%ile, Max

Easy to do once you have a stem and leaf diagram.

Box Plots are produced from 5 point summaries

the 25%ile and 75%ile are called 'hinges'

Eases comparisons of groups

Skeletal box plot

Box plot

GOAL FOR THIS SECTION - Appreciate the richness of data and how much information is lost in a simple bar graph of averages.