Definitions/Terminology

- Qualitative Research: Nature of the 'data' the distinguishing characteristic. With qualitative research no summary or reduction to a numerical representation of the data is made.
- Quantitative Research: With quantitative research, descriptive and/or inferential statistics are used to summarize data and infer from a sample something about the population the sample represents.
- Note: Good research begins with a good and well-articulated question. This will help you decide what type of research and data you need to examine.

Kinesiology 251

Analysis of Research & Issues in Kinesiology
Fall 2009

Definitions/Terminology

Scientific Method: Process through which the systematic and replicable investigation of a question/problem can be conducted. Steps:

Define and delimit the problem

Formulate the hypothesis

Gather Data

Analyze and interpret findings

Getting Started

- Web Site
- Syllabus
- Course Split
- Research Project & Set up Groups
- Course Web site &
- e-campus online resource

Research Overview

- Data Collection Methods
  - Participant Observation
  - Unobtrusive Observation
  - Content analysis
  - Historical Assessment
  - Interviews
  - Surveys
  - Active Interaction

Definitions/Terminology

- INDEPENDENT VARIABLE: The variable manipulated by the experimenter. Or a broader definition would be - any variable that is assumed to produce an effect on, or be related to, a behavior of interest.
- LEVELS OF AN INDEPENDENT VARIABLE: The various values or groupings of values of an independent variable. Ex: a study is conducted to determine the effect of room temperature (70, 80, and 90 degrees) on performance. There is one independent variable - room temperature - with three levels.

Research Design & Analysis Issues - Ethics

- Before enrolling participants in an experiment, the investigator should be genuinely uncertain of the outcome. In other words, a true null hypothesis should exist at the onset.
- The investigator must consider how adverse events will be handled; who will provide care for a participant injured in a study and who will pay for that care are important considerations.
- Government/institutions typically have definitions around misconduct. In addition, there are many activities commonly considered unethical.
- Check article on website and course website

Definitions/Terminology

- DEPENDENT VARIABLE: The behavior or characteristic observed or analyzed by the researcher, generally in regards to how the independent variable(s) affected or were related to it.
- TYPE OF DEPENDENT VARIABLE: In empirical research, the dependent variable is quantified in some way. Statistical analysis is carried out on the numerical values of the dependent variable. The three basic types are score data (ratio, interval), ordered data (ordinal), and frequency data (categorical).

Research Design & Analysis Issues - Ethics

Treatment of participants

- Informed Consent: Federal regulations state: "no investigator may involve a human being as a subject in research covered by these regulations unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative."
- For informed consent to be valid, these principles apply:
  - Disclosure: The potential participant must be informed as fully as possible of the nature and purpose of the research, the procedures to be used, the expected benefits to the participant and/or society, the potential of reasonably foreseeable risks, stresses, and discomforts, and alternatives to participating in the research. Document must make clear who to contact with questions/concerns.

Research Design & Analysis Issues - Ethics

Treatment of participants

- IRB review: primary concern safety of the research participant. Institutional Review Boards for the Protection of Human Subjects (IRB's): committees made up of scientists, clinical faculty, and administrators who review research according to the procedures set out in Federal Regulations.
  - If your research is part of a routine educational experience, or if your participants will remain completely anonymous you may apply to the IRB for a certificate of exemption.
  - A study may also qualify for "expedited review" if an IRB reviewer determines that it meets assessment criteria for minimal risk, and involves only procedures that are commonly done.

Research Design & Analysis Issues - Ethics

Treatment of participants - Informed Consent Principles

- Understanding: The participant must understand what has been explained and must be given the opportunity to ask questions and have them answered by one of the investigators.
- The participant's consent to participate in the research must be voluntary, free of any coercion or promises of benefits unlikely to result from participation.
- Competence: The participant must be competent to give consent. In some cases a surrogate is acceptable.
- Consent: The potential human subject must authorize his/her participation in the research study, preferably in writing.

Check U of Washington web site Examples

Research Design & Analysis Issues - Ethics

Treatment of participants

- Deception
  - Occasionally exploring your area of interest fully may require misleading your participants about the subject of your study. For example, home plate strike zone study. The IRB will review any proposal that suggests using deception or misrepresentation very carefully. They will require an in-depth justification of why the deception is necessary for the study and the steps you will take to safeguard your participants.
- Privacy and confidentiality for subjects is critical.

- Design and implementation of protocols should take into consideration threats to internal validity.
- Accuracy in reporting
  - Care should be taken not to compromise external validity of the research.
  - Conclusions must be grounded in data.

Manipulation of data

- Complexity for no legitimate reason should not be done.
- The manipulation of raw data to other levels of measurement acceptable provided sound reasons apply for example:
  - Age to age groups
  - Collapse when cells have too few cases.
- Never acceptable: fabricating, falsifying, or misrepresenting research data.

Evaluation of Research - General Guidelines

Problem too narrowly defined to make an important contribution to the advancement of knowledge?

Enough detail on protocol?

To what extent is the method likely to produce valid and reliable data?

Participants a convenience sample? Where self selection is an issue look 1st for the author's acknowledgment of the problem and 2nd consider whether or not the problem is great enough to invalidate or obscure findings.

Analyses appropriate? Validity/reliability info? Match between analysis and research question? Appropriate analyses for data type & assumptions.

Author(s) claim research 'proves'.... ?

Rigorous Error Checking

- Simple tools available to detect many errors
  - Frequency distribution tables
  - Crosstab tables
  - Graphs – for outliers

Evaluation of Research - Abstract

- Purpose of the study should be stated or clearly implied.
- A snapshot of the methodology should be given.
- Full titles of instruments should not be used unless the purpose of the study is to evaluate the reliability and validity of data from the instrument(s).
- Highlights of results should be include, but, the brevity should not result in a misrepresentation of findings.
- References to implications or future research do not belong in the abstract.

Evaluation of Research - Titles

- Titles should be sufficiently specific but, should not describe results.
- Titles should not pose a simple yes/no question.
- If two part titles are employed BOTH parts should contain specific/important information about the study.
- If the title framed around the main analytical question it is desirable to have the IV and DV in the title. e.g. The relationship between cholesterol level and exercise frequency.

Evaluation of Research - Introduction

- Intro should lead in by identifying one or more problems without a lot of extraneous verbage. Ideally, the 1st sentence provides a concise statement of the problem and a reference to support the statement.
- The importance of the problem should be made clear. Keep personal opinions out of the mix. The point is - has the author made the case?
- Unless chronology (eg historical research) is of over riding importance the intro should be developed around topics (not references).

Evaluation of Research - Titles

- If a narrowly delimited sample used it is desirable to include a reference to the population in the title.
- Titles should not infer causality unless the analytical techniques employed appropriate for drawing this type of inference.
- Title should not use acronyms or jargon.

Evaluation of Research - Literature Review

- Current research must be included, however not at the expense of relevant work - regardless of publication date.
- Opinion, when it surfaces should be clearly communicated to be such.
- Should use primary sources predominantly.

Evaluation of Research - Introduction

- Key terms should be defined as they come up in the intro.
- While the author's opinion may be brought into the intro it must be clear that's what it is. Any factual statement requires a source for support. e.g. incidence of injury has increased in recent years....
- The intro should lead the reader smoothly into a wrap up paragraph with the study's purpose, research questions, and reason study undertaken.

Quantitative Design & Analysis Issues; Research Methods; Reliability & Validity of Research & Data; Descriptive and Inferential Statistics

The methods section of a research paper needs to be meticulously inclusive. Someone not connected with the study should be able to replicate your work just by reading your methods section.

- Instrument development (including reliability & validity information)
  - Pilot Study
- Sampling
- Data collection protocol(s)
  - Pilot & Main Study
  - Consider sources of invalidity
- Statistical analysis
  - Descriptive Statistics
  - Inferential Statistics - main question
  - Descriptive/Inferential Statistics - related questions

Evaluation of Research - Literature Review

- Structure: Depending on the nature of the end product (thesis vs. journal article) the breadth of the opening will vary. In either case, the structure is that of a funnel and the entire section should take the reader down a logical path that ends with a restatement of the purpose, but, now in the context of previous work.
- Researcher should be selective in lit review. Long lists suggests work has not been scrutinized.
- When results vary across studies, the author should identify for the reader which they deem more dependable and why.

Research Design & Analysis Issues - Sampling

4 major sampling techniques

- Simple random sampling
- Simple stratified sampling
- Proportional stratified sampling
- Cluster sampling

One step is common in each of the 4 techniques above and that is randomization.

A random sample is one where each observation possible in the population has an equal chance of being included in the sample. In practice, samples seldom meet this criterion for randomness, but they are treated as random if no systematic bias exists that might be expected to invalidate the generalizations based on the sample.

Instrument Development

Issues in Construction of Likert-Scaled Survey

- Clearly define issue under study
- Determine what factors important to the issue under investigation
- Proceed to develop items (at least 5) under factors
- Avoid having extraneous/irrelevant items included
- Have experts review survey and give feedback on clarity and factor structure
- Revised survey based on experts' review
- Administered to a pilot group
- Analyze responses from pilot group. Need to conduct a factor analysis conducted to assess content validity & an intraclass R done on each factor to assess reliability.
- Revise survey based on review of pilot data

Research Design & Analysis Issues - Data Collection

- The quality of the methods employed to collect data is another key to limiting the problems of internal and external validity and reliability of the research.
- For the thesis format methods section must be meticulously detailed. Absolutely every piece of information related to collecting the data must be included. In an article format it typically must be tighter since there will be page limits. Specificity particularly with respect to the dependent variable is necessary.
- When data is collected via survey, information on the development of the instrument (including reliability & validity information) is critical.
- Good practice to conduct a pilot study.

Research Design & Analysis Issues - Sampling

- The selection of a sample is one of the keys to limiting the problems of internal and external validity and reliability of the research.
- Sample Size: is directly related to 'power' - probability of correctly rejecting the null hypothesis. Therefore, it is important that researchers determine sample size from the perspective of power. Software is available to help determine sample size.
- Being representative is as important as being large.
- Bias: Any influence, condition, or set of conditions which singly or together cause distortion of the data from what would have been obtained by pure chance. Bias due to inadequate sampling impairs external validity.

Research Design & Analysis Issues
Validity & Reliability of Research

Sources of Invalidity

- Instrumentation effect: Changes in instruments can be mistaken for changes in subjects.
- Pre-testing effect: Subjects can be changed or learning can take place during a pre-test which could affect results.
- Time: Over a length of time, maturation may have more of an impact than the independent variable. Also, major events can affect subjects' behaviors and/or opinions.
- Hawthorne effect: When the giving of attention rather than the independent variable is the cause of observed differences/relationships.

Example: Success Oriented Curriculum Article

Research Design & Analysis Issues
Validity & Reliability of Research

The data collection protocol(s) and selection of a sample are the keys to limiting the problems of internal and external validity and reliability of the research.

- Validity of Research:
  - Internal validity: Extent to which results can be attributed to "treatment".
  - External validity: Extent to which results can be generalized. External validity is examined qualitatively by scrutinizing the sampling scheme employed.

Research Design & Analysis Issues
Validity & Reliability of Research

Reliability of Research

- Essentially this refers to the replicability of the research. The reliability of the research is assessed qualitatively by scrutinizing the design and methodology employed in the research.
- Reliability of the research hinges on the thoroughness of the data collection protocol in addition to obtaining a representative sample.
- An estimate of power (probability of correctly rejecting the null hypothesis) is also useful in examining the stability of the research.

Research Design & Analysis Issues
Validity & Reliability of Research

Sources of Invalidity

- Rosenthal effect: Self fulfilling prophecy - you get what you expect. Best to do a double-blind study when this is a potential source of invalidity.
- Halo effect: General effect of good or bad feeling you have about a person. In observational designs this may be a particular problem. Best to use a check-list a verify reliability of the instrument and those collecting data.
- Demand characteristics: Allowing subjects to know what the goals are. Deception (of an ethical nature) may be needed to avoid this source of invalidity.
- Volunteer effect: Volunteers may be fundamentally different from the overall population you are trying to generalize to.

Psychometric Characteristics of Data - Validity

Content validity (assessed quantitatively) Ex: survey research - expect authors to:

- Pilot test the survey
- Conduct a factor analysis of survey results
- Revise based on analysis
- Administer survey and conduct another factor analysis

Psychometric Characteristics of Data

- Validity, objectivity & reliability of data from dependent variable
- Validity of Data; The instrument used to quantify the dependent variable should be examined for it's ability to produce valid data (ability to truly measure what it's supposed to). Valid data is clean and relevant. If the instrument is a well known one with work already in place establishing the validity of data produced by it, it may be enough to site a reference where validity was examined and show that the same protocol was followed in your study on similar subjects.
- Depending on the type and purpose of data collection, validity can be examined from one or more of several perspectives. Content; Concurrent; Predictive; Construct Validity. Can be assessed using an interclass coefficient: e.g. Pearson Product Moment Correlation (r).

Psychometric Characteristics of Data - Validity

Criterion-related validity (predictive and concurrent) - Compare measures from your dependent variable with measures from a criterion (expert, another test, etc.) of the same skill/knowledge.

Concurrent validity (assessed quantitatively) - expect authors to

- Gather x [dependent variable] and y measures [criterion] from a large group
- Compute an appropriate correlation coefficient
- If correlation strong, the data is said to have good concurrent validity

Psychometric Characteristics of Data - Validity

When measures are found to be valid for one purpose they will not necessarily be valid for another purpose. Validity also may not be generalizable across groups with varying characteristics.

For content/logical validity (assessed qualitatively) - expect authors to

- Clearly define what was measured.
- State all procedures used to gather measures.
- Have had an "expert" assess whether or not instrument/test producing clean & relevant data.

Psychometric Characteristics of Data - Reliability

Reliability of Data:

Concerned primarily with the dependent variable. The instrument used to quantify the dependent variable should be examined for it's ability to produce reliable data. (accuracy of measures reflected in consistency)

Reliability of the dependent variable can be assessed quantitatively using and intraclass coefficient:

- Intraclass R
- Coefficient alpha

Psychometric Characteristics of Data - Validity

Predictive validity (assessed quantitatively) - expect authors to

- Gather measures using their instrument (x) and measures on the variable(s) they are trying to predict (y)
- Compute an appropriate correlation coefficient
- If correlation strong, the data is said to have good predictive validity
- Follow up with estimation of SEE - band place around predicted score to quantify prediction error.

Psychometric Characteristics of Data - Reliability

The reliability of measures is typically assessed in one of two ways:

Internal consistency - By examining precision and consistency of test scores throughout one administration of a test.

Stability - By examining precision and consistency of test scores over time. (test-retest)

To estimate reliability you need 2 or more scores per person. If a test is given just once the most common way of getting 2 scores per person is to split the test in half - usually by odd/even trials or items.

Psychometric Characteristics of Data - Validity

Construct validity (assessed quantitatively)

A construct is an intangible characteristic. When you want to measure a construct such as anxiety, competitiveness, etc., you have no direct means to do so. Therefore indirect methods need to be employed. To then estimate the validity of the indirect measures (as reflections of the construct you're interested in) you record a pattern of correlations between the indirect measure(s) and other similar and dissimilar measures. Your hope is that the pattern reveals high correlations with similar measures (convergent validity) and low correlations with different measures (divergent/discriminant validity).

Expect authors to employ one of two techniques used to quantitatively assess construct validity: Multi-trait multi-method matrix; factor analysis.

Statistics - Measurement Issues

The place to start is with how to classify data - the scale the data appropriately belongs on will affect analysis decisions.

Categorical/nominal scale: Used to measure discrete variables that can be classified by two or more mutually exclusive categories.

Ex: Gender is a categorically scaled variable with two categories: male & female. the scale scores (0,1) have no meaning.

Data at this level of measurement can be summarized by:

- Frequency distribution tables
- Crosstabulation tables
- Charts/graphs

Statistics - Measurement Issues

Ordinal scale: Used to measure discrete variables that are categorical in nature and can be ordered (meaningfully).

Ex: Undergraduate class is an ordinally scaled variable with four meaningfully ordered categories: freshman, sophomore, Junior, Senior. The scale scores (1,2,3,4) have meaning in that Juniors have complete more units than sophomores who have completed more than freshman . . .

Another example is likert scaled items: eg strong agree ---- strong disagree

Data at this level of measurement can be summarized by:

- Frequency distribution tables
- Crosstabulation tables
- Charts/graphs

Psychometric Characteristics of Data - Reliability

Once you have 2 scores per person the question is how consistent overall were the scores. The inference here is that if two sets of scores are consistent there likely is little measurement error and so the scores are likely to be accurate reflections of true scores and so the observed scores are considered reliable.

What statistic to use?

Reliability has been estimated using the Pearson Product Moment Correlation coefficient. This is not appropriate since (1) the PPMC is meant to show the relationship between two different variables - not consistency of two measures of the same variable, and (2) the PPMC is not sensitive to fluctuations in test scores. Need an intraclass coefficient: Intraclass R or coefficient alpha.

Summary:Validity & Reliability

You are at all times interested in the reliability & validity of both the research and the data when analyzing the quality of research.

Examining the reliability & validity of the research is done by scrutinizing the design, sampling and data collection protocols. As a reader you should not assume that if no mention is made by the author(s) no threats to internal/external validity or reliability existed. When mention is made that should not necessarily cause you to question the quality.

Examining the reliability & validity of the data is done by scrutinizing the process and statistics used to assess validity and reliability of the data representing the dependent variable.

> Note: It is possible to have reliable data that is invalid. Data that is valid is typically expected to also be reliable. So, reliability does not insure validity.

Analysis of Data - Descriptive Stats

Summary information should be provided on:

- Participant Demographics
  - Whole Group
  - Relevant sub groups
- All variables relevant to the question under study
  - Whole group
  - Relevant sub groups

Statistics - Measurement Issues

Interval scale: Used to measure continuous variables that are ordinal in nature and result in values that represent actual and equal differences in the variable measured.

Ex: Temperature is an interval scaled variable with meaningfully ordered categories (hot, cold) that can be measured (scale has a constant unit of measurement) to finer and finer degrees given appropriate instrumentation.

Data at this level of measurement can be summarized by:

- Charts/graphs
- Central Tendency & Variability (excellent way to summarize descriptive information on subjects' characteristics at this level of measurement
- Correlation

Analysis of Data - Descriptive Stats

Typical choices for summary of descriptive data:

Frequency Distribution Tables - Percentages
Crosstabulation Tables - Percentages
Central Tendency - Mean, Median, Mode
Variability - Standard Deviation, Range
Correlation - Pearson Product Moment Correlation

Statistics - Measurement Issues

Ratio scale: Used to measure continuous variables that have a true zero, implying total lack of the attribute/property being measured.

Ex: Weight is a ratio scaled variable with meaningfully ordered categories (heavy, light) that can be measured to finer and finer degrees that also has a true rather than arbitrary zero.

Data at this level of measurement can be summarized by:

- Charts/graphs
- Central Tendency & Variability
- Correlation

Analysis of Data - Descriptive Stats

Correlation

Provides a measure of the strength of the relationship between two variables. Selection of a correlation coefficient depends on the variable type.

General Interpretation:

zero to ± .2 Weak
± .2 - .4 Low
± .4 - .7 Moderate
± .7 - 1.0 Strong

Analysis of Data - Descriptive Stats

Frequency Distribution & Crosstabulation Tables

When reporting percentages, author should report the underlying frequencies because percentages alone can be misleading.

	College A	College B
Total # Students	150	350
# Sport Studies Majors	12 (8%)	15 (4%)

Experimental Research Designs - Main Question (Differences)

- When interested in differences or change over time for one group or between groups a number of designs are applicable. The most frequently used designs can be collapsed into two broad types: true experimental and quasi-experimental.
- True experimental designs: these designs all have in common the fact that the groups are randomly formed. The advantage associated with this feature is it permits the assumption to be made that the groups were equivalent at the beginning of the research which would provide control over sources of invalidity based on non-equivalency of groups.
- The control is of course not inherent in the design. The researcher must still work with the groups in such a way that nothing happens to one group (other that the treatment) that does not happen to the other and that scores on the dependent measure do not vary as a result of instrumentation problems, or that the loss of subjects is not different between the groups.

Analysis of Data - Descriptive Stats

Central Tendency - Mean, Median, Mode
Variability - Range, Standard Deviation

Central Tendency: Provides a measure of where scores tend to center. Most commonly reported is the mean; however, it is NOT a representation of the center when the distribution is skewed. The median should be reported in that instance.

Data may be severely misrepresented when an inappropriate measure of central tendency is reported.

Data should be at least interval scaled when using the median or mean.

Variability: The companion to central tendency. Provides a measure of the spread of scores. Should always be reported with measures of central tendency.

Experimental Research Designs - Main Question (Differences)

Pretest-Posttest Randomized-groups design

- In its simplest form, this design requires the formation of two groups. One group will receive the 'experimental treatment' the other will not. The group not receiving the treatment is still referred to as the control group. Consider a dietary seminar intended to change eating habits particularly with respect to consumption of fat.

Group 1 (control)	Pre test		Post test
Group 2	Pre test	Seminar	Post test

- The first independent variable is seminar condition (two levels - experimental & control). The second independent variable is test (two levels - pretest & posttest). The dependent variable is grams of fat consumed.

Experimental Research Designs - Main Question (Differences)

Randomized Groups Design

- This design requires the formation of at least two groups. One group will receive the 'experimental treatment' the other will not. The group not receiving the treatment is commonly referred to as the control group.
- This design allows the researcher to test for significant differences between the control and experimental group after the experimental group has received the treatment. An independent t-test or one-way analysis of variance (ANOVA) may be used to statistically test the null hypothesis that H0: $\mu 1 = \mu 2$.
- When there are 3 or more levels of the independent variable then the one-way ANOVA must be used. For example, when there are 3 levels of the independent variable the null hypothesis is: H0: $\mu 1 = \mu 2 = \mu 3$.

Experimental Research Designs - Main Question (Differences)

Repeated Measures Design

- The repeated measures design is a variation of the completely randomized design though not considered a true experimental design. Instead of using different groups of subjects, only one group of subjects is formed and all subjects are measured/tested multiple times. There is no control group.
- The major advantage of this design over the completely randomized design is that fewer subjects are required. In addition, often increased statistical power is gained because the random variability of a single subject from one measure to the next is usually much less than the variability introduced by measuring and comparing different subjects. The major disadvantage is that there may be carry-over effects from one treatment/ testing to the next. In addition, subjects might become progressively more proficient at performing the criterion task and show an improvement in performance more attributable to learning than the treatment.

Experimental Research Designs - Main Question (Differences)

Factorial design:

- Essentially an extension of the randomized-groups design, this design has more than one independent variable and just one dependent variable.
- Allows the researcher to test for significant differences as a function of each independent variable separately (main effects) and in combination (interaction). Example: Do method of training and frequency interact to impact strength gains?
- The 'jargon' commonly associated with a factorial design: A 2X2 ANOVA ..... Conveys that there are two levels of the first independent variable and two levels of the 2nd independent variable. The language used to talk about the results would be the main effect for the first IV, the main effect for the 2nd IV, and the interaction.
- Variation: When one or more of the independent variables is a categorical variable, such as gender. Not true experimental design.

Analysis of Data - Inferential Stats

β = The level of risk (not under direct control of an experimenter) of failing to reject a false null hypothesis. It is also common to define beta as the probability of making a type II error.

Power = 1 - β. The probability of correctly rejecting a false null hypothesis.

Ideally, power should be considered when planning a study, not after it is over. If power is not considered at the start of a study it should be estimated at the end, particularly when non-significant results arise.

Sample size is closely tied to power. True differences/relationships go unnoticed without enough subjects. On the other hand, trivial differences/relationships can be statistically significant with large sample sizes.

Experimental Research Designs - Main Question (Relationships)

When interested in the relationship between/among variables, there are no design designations like 'factorial'. The design in this situation is equated with the analytical technique to be employed. Even without design names, good research communicate clearly

- What the independent and dependent variables were
- How the strength of the relationship was tested
- Results from an examination of practical significance

The null hypothesis under examination with a relationship question is:

$H0: \rho = 0$

Analysis of Data - Inferential Stats; Differences

Independent t-test Statistical Procedure for testing $H0: \mu1 = \mu2$ when the two levels of the independent variable are not related.

Dependent t-test Statistical Procedure for testing $H0: \mu1 = \mu2$ when the two measures of the dependent variable are related. For example, when one group of subjects is tested twice, the two scores are related.

In addition to assumptions about the level of data (at least interval) There are distributional assumptions associated with parametric statistics such as the t-test and ANOVAs. The most basic are:

Homogeneity of Variance: Are the spread of scores associated with each mean similar.

Normality: Is the shape of the distribution of scores around each mean normal.

Authors should convey results of checking assumptions.

Analysis of Data - Inferential Stats

Hypothesis testing involves examination of a statistically expressed hypothesis. You can think of hypothesis testing as trying to see if your results are unusual enough so that they would not even be expected by chance.

Type I error = Incorrectly deciding to reject a null hypothesis. Incorrectly reject a true null hypothesis.

Type II error = Incorrectly deciding not to reject a null hypothesis. Failing to reject a false null hypothesis.

α = The level of risk an experimenter is willing to take of rejecting a true null hypothesis. Often called the level of significance, this value is used in establishing a critical value around which decisions (reject or not reject null) are made. It is also common to define alpha as the probability of making a type I error.

Analysis of Data - Inferential Stats; Differences

Analysis of Variance (ANOVA). To examine whether or not there is a statistically significant difference in means on some dependent variable (continuous) as a function of some independent variable (categorical) you can use the F test from an ANOVA table when you have two or more levels of the independent variable (ex: 3 training protocols)

One-way ANOVA Statistical Procedure for testing H0: $\mu 1 = \mu 2 = \mu 3, \ldots$ when the two or more levels of the independent variable are not related.

Assumptions - Data interval scaled, Homogeneity of Variance, Normality

Authors should convey to reader results of checking assumptions. If assumptions violated then the non-parametric equivalent should be used.

Analysis of Data - Inferential Stats; Differences

Assessing statistical significance. Following analyses using a t-test you could compare the t statistic to an appropriate table of critical values. Information needed is alpha and degrees of freedom(df):

n1+n2 - 2 (independent t-test) N – 1 (dependent t-test)

If the t statistic > critical value you can reject your null hypothesis.

More often, authors have used software to give them a p value to compare to the alpha they've chosen. If the p value < the alpha you can reject the null hypothesis.

REMEMBER, if multiple tests done, alpha should be modified before comparison done.

Note: p value can be considered the probability that findings due to chance (sampling error).

Analysis of Data - Inferential Stats; Differences

Assessing statistical significance: Following analyses using an F test from a one-way ANOVA you could compare the F statistic to an appropriate table of critical values. Information needed is alpha and df:

K-1; N-K

If your F statistic > critical value you can reject your null hypothesis. More often authors have used software to give them a p value to compare to the alpha they've chosen.

If the p value < the alpha you can reject the null hypothesis.

REMEMBER, if multiple tests done, alpha should be modified before comparison done.

Analysis of Data - Inferential Stats; Differences

Assessing practical significance. A statistical 'test' tells you whether there's a statistically significant difference not whether the difference is of any practical importance. Therefore, it's important for researchers to take the next step and examine practical significance. One useful measure is an effect size.

Effect Size. It conveys the size of the effect observed in a way that permits interpretation of the practical significance of the results.

Interpretation:

.30 Small effect

.50 Moderate effect

.80 Large effect

Two-way ANOVA

Statistical Significance: Take a look at the p values for each of the main effects and interaction. If the p value < the alpha you can reject the null hypothesis. REMEMBER, since multiple tests done, alpha should be divided by 3 before comparison done.

Practical significance Remember the F 'test' tells you whether there's a statistically significant difference not whether the difference is of any practical importance. Therefore, it's important for researchers to take the next step and examine practical significance by calculating a statistic such as eta squared - proportion of total variance that can be explained by the independent variable. Effect size still a good measure as well.

Assessing practical significance. Following the F test which conveys information on statistically significant differences, when there are more than two groups, a post-hoc analysis (Scheffe) is commonly done to determine where the differences are. In addition, it's important for researchers to take the next step and examine practical significance. Again effect size is a useful measure.

Interpretation:

.30 Small effect

.50 Moderate effect

.80 Large effect

Repeated Measures Analysis of Variance (ANOVA). Statistical Procedure for testing H0: $\mu 1 = \mu 2 = \ldots$ when the two or more measures of the dependent variable are related. For example, when one group of subjects is tested two or more times, the two scores are related.

Assumptions:

- Repeated Measures at least interval scaled
- Sphericity
- Normality
- Homogeneity of variance

Two-way Analysis of Variance (ANOVA). You now have two independent variables and one dependent variable. The two way ANOVA provides information on three null hypotheses:

A difference in the dependent variable due to the 1st independent variable

A difference in the dependent variable due to the 2nd independent variable.

A difference in the dependent variable due to the interaction of the two independent variables.

Assumptions - Data at least interval scaled, Homogeneity of Variance, Normality

Analysis of Data - Inferential Stats; Differences

Non Parametric Statistics

Mann Whitney: This statistic is the non-parametric equivalent to the independent t-test. There are no distributional assumptions to meet. This statistic tests for a difference in two medians.

Kruskal Wallace: This statistic is the non-parametric equivalent to the one-way ANOVA. There are no distributional assumptions to meet. This statistic tests for a difference in two or more medians.

Wilcoxon: This statistic is the non-parametric equivalent to the dependent t-test and repeated measures ANOVA. There are no distributional assumptions to meet. This statistic tests for a difference in two or more medians from repeated measures of the dependent variable.

Analysis of Data - Inferential Stats; Differences

Assessing statistical significance Following a repeated measures ANOVA you could compare the F statistic to an appropriate table of critical values. Information needed is alpha and df:

K-1; N-K

If your F statistic > critical value you can reject your null hypothesis. Could also compare the p value to the alpha they've chosen. If the p value < the alpha you can reject the null hypothesis.

Assessing practical significance Remember the above 'test' tells you whether there's a statistically significant difference not whether the difference is of any practical importance. Therefore, it's important for authors to take the next step and examine practical significance.

Analysis of Data - Inferential Stats; Relationships

Parametric Statistic - Pearson Product Moment Correlation

Assumptions:

Linearity: straight line can be draw through points on scatterplot

Data for both x and y at least interval scaled

Assessing statistical significance: Following analyses using a PPMC you could compare the PPMC statistic to an appropriate table of critical values. Information needed is alpha and df:N-2

If the PPMC statistic > critical value you can reject your null hypothesis. If you work from a p value then you compare to the alpha chosen. If the p value < the alpha you can reject the null hypothesis.

Practical significance: Coefficient of determination. r squared.

Analysis of Data - Inferential Stats; Differences

Non Parametric Statistics

Assumptions:

Samples were drawn at random from the population under consideration.

Variable(s) under study have underlying continuity.

Dependent variable at least ordinal scaled.

Analysis of Data - Inferential Stats; Relationships

Non Parametric Statistics - Chi Squared

This statistic that will test for the presence relationship between two categorical (though can also be used on ordinal data) variables. The null hypothesis under examination is: p= 0.

Assumptions

The expected frequency in all cells is at least 5.

Data must be random samples from multinomial distributions.

Statistical significance: Following analyses using the $\chi 2$ statistic , you could compare the $\chi 2$ statistic to an appropriate table of critical values. Information needed is alpha and df = (R-1)(C-1). Where R = # of rows, and C = # of columns in a cross-tabulation table.

Analysis of Data - Inferential Stats; Relationships

Parametric Statistics - Regression

This is the most common approach to prediction problems when you have one dependent variable and multiple independent variables.

Assumptions

- Linearity: straight line can be draw through points on scatterplot
- Homoscedasticity: Y values at each x similar in variability
- Dependent variable at least interval scaled
- Multicolinearity: relationship among independent variables

Statistical significance: Following analyses to test H0: b=0, you could compare the F statistic to an appropriate table of critical values. Information needed is alpha and df: K; N-k-1

Analysis of Data - Inferential Stats; Relationships

Non Parametric Statistics - Chi Squared

Practical Significance

In this circumstance, the effect size is the correlation coefficient (Phi, Cramer's V). These statistics convey the strength of the relationship between the two categorical variables.

.30 Small

.50 Moderate

.80 Large

Analysis of Data - Inferential Stats; Relationships

Parametric Statistics - Regression

Practical significance: Coefficient of determination - r squared.

Interpretation: Proportion of total variance that can be explained by the independent variable(s).

.30 Small

.50 Moderate

.80 Large

Structure/Content Continued

- Sampling
  - Will the procedure utilized produce a representative sample that inferences can reasonably be made from?
  - To what extent is bias likely to be present? Common techniques: random, pseudo-random, stratified random, proportional stratified random. If optimal sampling not achieved, the author should explicitly identify the implications/limitations.
  - Is the sample size adequate? How was the target established?

Length: Technical reports and journal articles may have page limits and within that authors should convey critical components others would need to replicate work. In contrast, a thesis needs to take as much space as needed and be as meticulously inclusive as possible. Someone not connected with the study should be able to replicate your work just by reading your methods section.

Structure/Content Continued

- Data collection protocol(s)
  - Should be thorough and 'tight' enough to reduce the likelihood of compromising internal validity. Where the collection and/or recording of data has a subjective component, evidence that objectivity/ reliability assessed essential.

- Structure/Content: The following is recommended. The order may vary, however the content should be present.
  - Instrument development (including reliability & validity information)
    - Process thorough?
    - Instrument pilot tested (with validity & reliability examined appropriately)?

Evaluation of Research - Methods

Structure/Content Continued

- Statistical Analysis of Data
  - Analyses pertaining to related problems.
    - Appropriate for data type?
    - Appropriate for question under examination?

Evaluation of Research - Methods

Structure/Content Continued

- Statistical Analysis of Data
  - Descriptive Statistics.
    - Appropriate for data type?
    - Validity & reliability of data examined? Appropriately?
    - Contributes to understanding of sample or problem?
    - Provides good context (backdrop)?

Results Section; Analysis of Data - Summary

Descriptive Statistics
- Frequency Distribution Tables - Percentages
  - Categorical & Ordinal Data
- Crosstabulation Tables - Percentages
  - Categorical & Ordinal Data
- Central Tendency - Mean, Median, Mode
  - Interval & Ratio Data
- Variability - Standard Deviation, Range
  - Interval & Ratio Data
- Correlation
  - Categorical, Ordinal, Interval, Ratio

Evaluation of Research - Methods

Structure/Content Continued

- Statistical Analysis of Data
  - Inferential Statistics - main question.
    - Appropriate for data type?
    - Appropriate for question under examination?
    - Should be the least complex it can be and still provide insight to the question(s) being examined.
    - Assumptions checked?
    - Both statistical and practical significance reported?

Results Section; Analysis of Data - Summary

Graphs

- Continuous data: Frequency polygon or Histogram
- Discrete Data: Bar Chart; Pie Chart
- General Principles
  - 3/4 Rule
  - Label axis for correct interpretation
  - Begin the vertical axis with the value zero.
  - To show trend, several points along the way have to be depicted if interpretation is to be sound

Results Section; Analysis of Data - Summary

Inferential Statistics

- Differences - parametric
  - t-test - Independent & Dependent; omega squared
  - ANOVA - one-way, two-way, repeated measures; eta squared
  - Assumptions - Homogeneity of Variance, Normality, Interval, ratio scaled data, sample random
- Differences - Non-parametric
  - Mann Whitney; Kruskal Wallace
  - Wilcoxon
  - Assumptions - Sample random, data at least ordinal scaled, underlying continuity.

Evaluation of Research - Results

Structure/Content - Overall

Structure/Content: Entire text should be cohesive and follow a logical path that generates confidence in the findings. The following is one recommended structure. The content should match the information conveyed in the analysis portion(s) of the methods section.

Note: Where relevant, tables should be used to summarize large volumes of data and text should highlight important elements.

Results Section; Analysis of Data - Summary

Inferential Statistics

- Relationships - Parametric
  - Correlation; Coefficient of Determination
    - Assumptions: Linearity, Homoscedasticity, Interval Scaled Data
  - Regression; Coefficient of Determination
    - Assumptions: Linearity, Homoscedasticity, Interval Scaled Data
    - Multicolinearity
- Relationships - Non-parametric
  - Chi-squared; Cramer's V & Phi
  - Assumptions: Cell Frequencies at least 5, data categorical or ordinal

Structure/Content - Related Question(s)

- Statistical examination - not recommended
  - Note: Remember that each statistical test requires that you divide your alpha to adjust for the increased chance of making a type I error. Therefore, it is wise to limit the number of statistical tests conducted.
- Descriptive Statistics
  - Should directly provide information pertaining to related question
  - Statistics reported must be appropriate for the data type summarized.

Structure/Content - Descriptive Stats

On sample: Should summarize the personal and demographic information that helps the reader understand the nature of the research participants

On relevant variables: Should summarize pertinent variables that shown interesting patterns and/or provide insights to main question.

On variables by subgroup(s): Should convey additional insights that contribute to an understanding of the nature of the research participants and/or the problem under examination.

Psychometric properties of data: As appropriate, results of examinations of objectivity, reliability, and validity of data.

- A tight summary of purpose and results should be included.
- Methodological limitations should be clarified.
- Results should be cast in light of literature cited in the intro/review of lit.
- New literature should not be brought into the discussion.
- Implications and/or recommended action in light of findings should be drawn out in this section.
- Next steps likely to extend or clarify research presented should be suggested.
- Any speculations must be clearly identified as such. There should be no doubt as to whether the discussion of results is data based or conjecture.

Structure/Content - Main Problem

- Should provide concise reporting of results from check of assumptions.
- Should provide clear results of hypothesis testing including:
  - Inferential Statistic
  - Degrees of freedom
  - P value
  - Table (e.g. ANOVA) when appropriate
- Examination of practical significance.

Kinesiology 251

Analysis of Research & Issues in Kinesiology
Fall 2009