Correlation & Surveys

Class 5

Confidence Limits/Interval

- Attempts to define range of true population mean based on Standard Error estimate.
- Confidence level
  - 95% chance vs. 99% chance
- Confidence Limits
  - 2 numbers that define the range
- Confidence Intervals
  - The entire range of scores b/w confidence limits
- On surveys
  - [http://www.surveysystem.com/sscalc.htm#two]

Standard Error of the Mean

- Estimate of the average SD for any number of samples of the same size taken from the population.
  - Example: If I tested 30 students on music theory
    - Test 0-100
    - Mean 75; SD 10
    - Standard Error (SE) would estimate average SD among any number of same size samples taken from the population
- SEM = SD/sq root N
- Figure for example on the left.
- 95% Confidence Interval
  - 95% of the area under a normal curve lies within roughly 1.96 SD units (rounded to +/-2)
  - 95% confidence interval = M + or – [SEM X 1.96]
ANCOVA – Analysis of Covariance

- Statistical control for unequal groups
- Adjusts posttest means based on pretest means.
- [example] http://faculty.vassar.edu/lowry/VassarStats.html

The homogeneity of regression assumption is met if within each of the groups there is a linear correlation between the dependent variable and the covariate and the correlations are similar/b/w groups.

Effect Size (Cohen’s $d$)

- [Mean of Experimental group – Mean of Control group/average SD]
- The average percentile standing of the average experimental participant relative to the average control participant.
  - What percentile would the average $X$ of the experimental group fall in the control group.
- Good for showing practical significance
  - When test in non-significant
  - When both groups got significantly better (really effective vs. really really effective!)
- Calculate effect size:
  - Treatment group: $M=24.6; SD=10.70$
  - Control Group: $M=10.8; SD=7.77$

Correlational Research
Correlational Research Basics

- Relationships among two or more variables are investigated
- The extent to which 2 variables move in the same or opposite direction
- The researcher does not manipulate the variables
- Direction (positive [+]) or negative [-]) and degree (how strong) in which two or more variables are related (-1 to +1)

Uses of Correlational Research

- Clarifying and understanding important phenomena (relationship b/w variables—e.g., height and voice range in MS boys)
- Explaining human behaviors (class periods per weeks correlated to practice time)
- Predicting likely outcomes (one test predicts another)

Uses of Correlation Research

- Particularly beneficial when experimental studies are difficult or impossible to design
- Provide preliminary data for experimental research
- Allows for examinations of relationships among variables measured in different units (decibels, pitch; retention numbers and test scores, etc.)
- DOES NOT indicate causation
  - Reciprocal effect (a change in weight may affect body image, but body image does not cause a change in weight)
  - Third (other) variable actually responsible for difference (Tendency of smart kids to persist in music is cause of higher SATs among HS music students rather than music study itself)
Interpreting Correlations

- $r$
  - Correlation coefficient (Pearson, Spearman)
  - Can range from -1.00 to +1.00
- Direction
  - Positive
    - As X increases, so does Y and vice versa
  - Negative
    - As X decreases, Y increases and vice versa
- Degree or Strength (rough indicators)
  - $< 0.30$: small
  - $< 0.65$: moderate
  - $> 0.65$: strong
- $r^2$ (% of shared variance)
  - Overlap b/w two variables
  - Percent of the variation in one variable that is related to the variation in the other.
  - Example: Correlation b/w musical achievement and minutes of instruction is $r = 0.86$. What is the % of shared variance ($r^2$)?
  - Easy to obtain significant results w/ correlation. Strength is most important

Interpreting Correlations (cont.)

- Words typically used to describe correlations
  - Direct (Large values w/ large values or small values w/ small values. Moving parallel, 0 to +1
  - Indirect or inverse (Large values w/small values. Moving in opposite directions. 0 to -1
  - Perfect (exactly 1 or -1)
    - 50 75 9
    - 40 62 14
  - Strong, weak
    - 35 53 20
  - High, moderate, low
    - 24 35 45
    - 15 21 58
  - Positive, Negative
- Correlations vs. Mean Differences
  - Groups of scores that are correlated will not necessarily have similar means. Correlation also works w/ different units of measurement.

Statistical Assumptions

- The mathematical equations used to determine various correlation coefficients carry with them certain assumptions about the nature of the data used...
  - Level of data (types of correlation for different levels)
  - Normal curve (Pearson, if not-Spearman)
  - Linearity (relationships move parallel or inverse)
    - Young students initially have a low level of performance anxiety, but it rises with each performance as they realize the pressure and potential rewards that come with performance. However, once they have several performances under their belts, the anxiety subsides. (non-linear relationship of # of performances & anxiety scores)
  - Presence of outliers
  - Homoscedasticity – relationship consistent throughout
    - Performance anxiety levels off after several performances and remains static (relationship lacks Homoscedasticity)
  - Subjects have only one score for each variable
  - Minimum sample size needed for significance
Correlation Approaches for Assessing Measurement Reliability

- Consistency over time
  - test-retest (Pearson, Spearman)
- Consistency within the measure
  - internal consistency (split-half, KR-20, Cronbach’s alpha)
    - [do example]
    - Spearman Brown Prophecy formula
      - $2r/(1 + r)$
- Among judges
  - Interjudge (Cronbach’s Alpha)
- Consistency b/w one measure and another
  - (Pearson, Spearman)

Multiple Regression

- Can be used to predict the extent to which combinations of independent variables can predict a dependent variable
- Dependent (criterion) Var. & Predictor Vars.
  - Tests needed to predict music achievement
    - Reading test .65 IQ test .70
    - Math test .74 Interest Survey .35
    - Science test .59 Teacher behavior report .20
    - Social studies test .40 Hearing test .45
  - Subsequent variables should correlate least to those already chosen because they are highly related and explain same variance as previous variables

Descriptive Research I

Class 5

http://faculty.vassar.edu/lowry/VassarStats.html
Designing a Descriptive Study

- A purpose/problem is stated
- A population is selected (to whom do you want to generalize results?)
- A mode of data collection is selected (survey or other measure)
- A sample is selected (sampling methods)
- The instrument is constructed/adapted
- Information related to the purpose/problem is collected from a group of individuals (administer measure)
- The information is summarized and analyzed
- From the results, generalizations are made about the population in question

Survey/Interview Types

- Cross-sectional
  - Information collected at one point in time
    - most common
- Longitudinal
  - Information collected at MORE than one point in time
    - Trend study – different subjects from a changing population measured over time (4th graders studied every year)
    - Cohort study – Same population, different sample every time (Beginning IL music teachers that started in 2011-different sample taken from same group every measure)
- Panel Study – same sample of respondents over time (track group of 1st graders through HS)
- Interview
  - Standardized/structured, semi-structured, open-ended

Uses of Survey Research

- Usually Descriptive
- May also be associational
  - Correlations among items (i.e., self reported ratings of performance ability and practice time)
  - Comparisons between groups on items (differences in responses b/w males & females, novice & experienced teachers, musical experience vs. non-musical experience, etc.)
  - Comparisons within group on items (compare all participants on preferences for Ren, Bar., Class, etc.)
- Rarely experimental – but can be if treatment is intended to alter attitudes
Defining the Problem/Purpose

- Questions need to be important and interesting enough to merit response
- Consider a hierarchical approach to question selection
- Also avoid asking for information to be reported when you can find it elsewhere (i.e., unobtrusively)
- In instructions, make sure they understand that data will be shared (IRB procedures)

ID the Population and Sample

- Individuals vs. Units of Analysis
  - Not just people… things, places, events, etc.
- Define the population so that is clear who may or may not be considered in the sample
- Some form of random sampling is best once population is defined
  - Consider sending to those in authority rather than simply the respondents (e.g., teachers to administer to their students)
  - In the case of teachers, perhaps send to them directly if possible

Choose Mode of Data Collection

- Direct administration
  - Pro
    - When researcher has access
    - Response rate often excellent
    - Can clarify on the spot
  - Con
    - Intact groups may not be representative of population
- Mailed/emailed survey
  - Pro
    - Access to individuals who are hard to reach
  - Con
    - Response rate is often poor
- Telephone interview
  - Pro
    - Cheaper and quicker than a personal interview
  - Con
    - Poor response rate
    - Compromises anonymity
- Personal interview
  - Pro
    - Good for encouraging participation
    - Can clarify on the spot
    - Can probe for more info. or detail
  - Con
    - Very time consuming
    - Very costly
    - May require assistants – who then need extensive training
    - Compromises anonymity
Types of Survey Items

• Close Ended
  – Easier to score
  – Harder to write
  – May not include subjects’ desired responses
• Open Ended
  – Harder to score
  – Easier to write
  – Subjects can say whatever they want

• Gateway/Contingency/Filter - good when items may only apply to some of the subjects (“If you do not play in the band, skip to question X)
• May be best to consider a combination (Use more closed than open questions. Only use gateway when appropriate)

Other forms of Questions

• Nominal Questions
  – responses are assigned a number with no meaning. (i.e., indicate level of education)
• Single Choice Question
  – Respondent may only choose one response (Likert scales)
• Rating Questions (Likert scales, etc.)
• Dichotomous Questions
  – Questions with two possible responses (i.e., yes, no)
• Multiple option question
  – Respondent can select more than one option (check all that apply)
• Ordinal Questions
  – Rank a list of items (i.e., rank the following songs from your favorite (1) to your least favorite (5).

Semantic Differential

The customer care representative was...

<table>
<thead>
<tr>
<th></th>
<th>Very Much</th>
<th>Somewhat</th>
<th>Neither</th>
<th>Somewhat</th>
<th>Very Much</th>
</tr>
</thead>
<tbody>
<tr>
<td>helpful</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>friendly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>polite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some general tips when constructing items:

- Avoid ambiguity – Clear, direct statements
- Focus on one issue at a time (no double barreled questions)
- Shorter is better
- Use common language rather than jargon
- Avoid triggers for bias (key words)
- Avoid ‘leading language’
- Avoid double negatives

More Questions Designing Tips

- Make sure your survey questions match your research objectives
  - Validity: are you measuring what you intend to measure. How do you know?
  - Know exactly why you are asking a particular question
- Understand your research participants
  - Take age into account
  - Test reading level (Flesch reading ease score [1-100: higher the score, easier to read; elementary students need 90+]; Flesch-Kincaid Grade Level Score)
- Use natural and familiar language (not academic)
- Write Likert scale questions in one direction (i.e., respond to positive statements about music class anchored by “strongly disagree” and “strongly agree”)
  - Avoid having to transpose numbers for statistical analysis
  - Will avoid confusing respondents (but may make them read questions less carefully)

Response Issues

- Problem: those who do not respond may differ from those who do on some critical issue in some systematic way
- Return rate matters!
- Ideas for increasing response rate
  - Face-to-face = best, Telephone = 2nd, Mail = 3rd
  - Confidentiality, anonymity
  - Organized survey/interview
    - Business-like, conservative interviewer characteristics
    - Short as possible while still getting good data
  - Multiple mailings
    - Postcard ahead of time, survey and cover letter, reminder, 2nd mailing, reminder, 3rd mailing, etc.
  - Call specific individuals
  - Call-backs, appointments
  - Tangible rewards (?)
Examples & Practice

• Look at my examples
• Write 2 questions for X topic.
• They don’t all have to be Likert questions

Online Surveys

  – Free up to 50 responses in 10 days
  – $9.95 per month for students
  – Free and unlimited
  – Only 10 free questions
  – Must pay annually

Non-Parametric Statistics
Chi-Squared

- Measure statistical significance b/w frequency counts (nominal/categorical data)
- Compare 2 numbers
- Can compare w/ you have with what is expected
  - Proportions of contest ratings (I, II, III or I & non Is)
  - Agree vs. Disagree
- Weak statistical test

Compare 2 means

- Independent
  - Mann Whitney U

- Pairs or dependent samples
  - Wilcoxon signed ranks

Non-Parametric ANOVAs

- Friedman – Related (correlated) Samples
- Kruskal-Wallis – Independent Samples
Reliability of Survey

- What broad single dimension is being studied?
  - e.g. = attitudes towards elementary music
  - Preference for Western art music
  - “People who answered a on #3 answered c on #5”
- Use Cronbach’s alpha
  - Measure of internal consistency
  - Extent to which responses on individual items correspond to each other

Spearman Brown
Prophesy Formula

- Reliability = \( \frac{n \times r}{1+(n-1)r} \)
- n=number of times we multiply items to get new test length (10 item to 20 item – n=2)
- For a test of 10 items w/ reliability (α) of .60
  - (15 items) \( 1.5 \times .60/1+(1.5 - 1).60 \) = reliability for test 1.5x size
  - (20 items) \( 2 \times .60/1+(2-1).60 \) = reliability for a test 2x size
  - (25 items) \( 2.5 \times .60/1+(2.5 - 1).60 \) = reliability for test 2.5x size
  - (5 items) \( .5 \times .60/1+(.5 - 1).60 \) = reliability for test .5 size

For Tuesday/Wednesday

- Revise Introduction and submit
- Make a list of questions related to your study or class content in general
- Read two articles online. Complete worksheet for Hash (2011) “Effect of Pullout Lessons on the Academic Achievement of Eighth Grade Band Students”.
- Revise lit review and turn in methodology section (pt. 3.)
- Find one additional research article related to your study and present to the class (5 mins.)
- (Formal presentations Friday!!)