Measurement in Educational Research

Kari Lock
Harvard University
Department of Statistics
Rigorous Research in Engineering Education
12/3/09
Measurement

• It is important to separate the concept you want to measure from the way you decide to measure it.

• Measurement is often not perfect.
The **outcome variable** is the variable you are interested in improving or explaining.

- **Step 1:** *State your outcome variable in general terms.*
  - **Examples:**
    - “Improvement in conceptual understanding”
    - “Meeting a specific learning objective”

- **Step 2:** *Find an effective way to measure this outcome.*
Outcome Variable

• Why not have multiple outcome variables?
  
  • Multiple outcome variables indicate a lack of focus in your research
  
  • Multiple hypotheses require much more work (literature review for each, measurement for each, etc.)
  
  • The analysis for each outcome variable is either done independently (meaning multiple analyses to conduct) or together (meaning much more complex statistical analysis is needed).
  
  • Each outcome variable needs it’s own hypothesis, and testing multiple hypotheses runs into the multiple testing problem...
Multiple Testing

• Significance tests are based on the idea that the results you observe are more extreme than would be expected just by random chance.

• However, once you start doing many tests, the chance that *at least one* of these tests is extreme just by random chance goes up rapidly.

• If you are conducting multiple tests, the criteria to prove each individual test becomes much more stringent (to make sure you don’t “prove” a hypothesis that really was just random chance).
Explanatory Variables

• **Explanatory variables** are variables that you think may explain your outcome variable

• If you are doing a randomized experiment, you probably have only one explanatory variable that want to test. This is also referred to as the “treatment” or “intervention”

• If you are doing an observational study, you can test as many explanatory variables as you like, but recall the issue of multiple testing

• It’s usually best to focus your study on a couple explanatory variables that you think might be important
Choosing Variables

• Sometimes it isn’t clear which should be your outcome variable and which should be your explanatory variable if you only have one of each

• For many statistical analyses, this is okay. The word “association” is symmetric, and you can test for a significant relationship between two variables without specifying which is explanatory and which is outcome

• If you have more than 2 variables, you should choose one in particular to be the outcome variable
Covariates

• **Covariates** are variables that you aren’t directly interested in, but that you still need to measure
  • Potential confounding factors you want to control for
  • Demographics of your sample to relate to the population
  • Age, gender, major, etc.

• You can collect information on as many covariates as you think are important.

• Covariate data is typically collected in the form of a survey, regardless of how you collect the data on your variables of interest
Measuring Variables

• How you choose to measure variables depends on whether it is an outcome variable, an explanatory variable, or a covariate.

• Typically, the most thought and effort is put into the measurement of the outcome variable.

• Explanatory variables are usually either controlled by you or out of your control and simply measured as they are. Often explanatory variables are binary (only two possibilities).

• Covariate measurement is usually pretty straightforward.
Choosing Variables

• What is your outcome variable? (generally)

• What is/are your explanatory variable(s)?

• What other covariates might you want to measure?
Nominal Variables

• **Nominal variables** are categorical variables in which the categories have no meaningful order
  • Gender
  • Major

• The proportion in each category can be expressed, but taking averages or other numerical summaries doesn’t make sense
Ordinal Variables

- **Ordinal variables** are categorical variables in which the categories *do* have a meaningful order
  - Strongly dislike, dislike, neutral, like, strongly like
  - First year, sophomore, junior, senior

- We are interested in the proportion in each category

- We can also now talk about **percentiles**: what percent of students are below/above a certain category.

- The median (the middle value out of a list of ordered values) is now a meaningful measurement, but not the mean
Measuring Categorical Variables

• **Categorical variables** include both nominal and ordinal variables

• Usually (but not always) categorical variables are relatively easy to measure

• Too many categories leads to very small sample sizes in some categories, which can make inference very hard. Too few categories may lose valuable information.
  • Levels can always be aggregated after the fact

• If measuring with a survey, is there a suitable answer for everyone? (yes, no, maybe/i’m not sure)
Categorical Variables Visualization

Nominal Only:

Pie Preferences

- Pumpkin 30%
- Cherry 23%
- Pecan 8%
- Coconut 3%
- Cream 3%
- Apple 36%

Nominal and Ordinal:

Birthdays of Students by Month

- Number of Students
- Month

Goal: display the number or percent of people in each category
Numeric Variables

• **Numeric variables** are variables coded as numbers, and the numbers have *numeric meaning* (coding 1 = Male, 2 = Female does NOT count as numeric)

• **Interval variables**: the distance between values is meaningful

• **Ratio variables**: interval variables, but the value of 0 is also meaningful
Likert Scales

• **Ex//** How much do you agree with (a statement):
  1) Strongly disagree
  2) Disagree
  3) Neither agree or disagree
  4) Agree
  5) Strongly agree

• These can be treated either as ordinal or as numeric data, and you don’t need to decide until you analyze the data.

• If you think the difference between level 1 and level 3 is roughly the same as the difference between level 3 and level 5, then you can treat it as numeric. Otherwise, it is ordinal.

• Treating this type of data as numeric allows for simple summaries (means) and easier analysis.
Numeric Variables
Meaningful summary statistics

**Mean** – the “average”: add everything up and divide by the sample size.

**Median** – the “middle” point at which 50% of the data are below and 50% are above.

**Standard Deviation** – A measure of the average distance from the mean. Indicates how “spread out” the data are.

**Variance** = (Standard Deviation)^2
• Goal: Visually display the entire distribution of observed numbers. Location, spread, shape and outliers are all visible.
Categorical Explanatory Variable
Numeric Outcome Variable

Side-by-Side Boxplots

- Compare the numerical distributions of two (or more) groups
- t-test (two groups) or ANOVA (>2 groups) for difference in means
• Correlation measures the strength of the linear association between two numeric variables
• Linear regression models this relationship
Categorical Explanatory Variable
Categorical Outcome Variable

Table:

<table>
<thead>
<tr>
<th></th>
<th>Like</th>
<th>Dislike</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Control</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

- Test for difference in proportions
- Chi-square test for independence
Categorical vs Numeric

• Will your outcome variable be nominal, ordinal, or numeric?

• Will your explanatory variable(s) be nominal, ordinal, numeric, or some of each?

• If categorical, what will the categories be?
Measurement

• You have a general idea of what you want to measure (outcome, explanatory, covariates)

• You have decided whether you want categorical or numeric measurements

• What makes a good measurement???
• **Reliability** refers to the *consistency* of a measurement

  • Test-Retest Reliability
    • If a person were to be measured again, you should get the same measurement

  • Internal Reliability
    • If items on a questionnaire are supposedly measuring the same construct, they should give consistent measurements
Reliability

• Measurements are often associated with some random error

• Think of exam scores: student’s have an underlying knowledge of the subject matter, but their actual score may depend on many other factors (the amount of sleep they got the night before, the time of day, food, caffeine, distractions, etc.)

• A reliable measurement has a small amount of random error
Internal Consistency

- Measured by Cronbach’s alpha (α)

<table>
<thead>
<tr>
<th>Alpha Value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 to 0.90</td>
<td>Excellent Measure</td>
</tr>
<tr>
<td>0.89 to 0.80</td>
<td>Good Measure</td>
</tr>
<tr>
<td>0.79 to 0.70</td>
<td>Acceptable Measure</td>
</tr>
<tr>
<td>0.69 to 0.60</td>
<td>Questionable Measure</td>
</tr>
<tr>
<td>0.59 to 0.50</td>
<td>Poor Measure</td>
</tr>
<tr>
<td>0.49 to 0.00</td>
<td>Unacceptable Measure</td>
</tr>
</tbody>
</table>
Validity

- **Validity** refers to whether a measurement is measuring the quality it is supposed to be measuring.

- If an exam is supposed to be measuring conceptual understanding but is really only measuring memorization ability, that is bad.

- While reliable measurements minimize random error, valid measurements minimize *systematic error*. 
Reliability and Validity

Our “measurement” the dart, is trying to hit the bullseye:

High Reliability, Low Validity

High Validity, Moderate Reliability
Assessing reliability and validity

• If developing your own test or form of measurement, it is very important that you assess your measurement for reliability and validity

• This should include conducting a **pilot study** of your measurement: measuring a small group of people before conducting your experiment and checking reliability and validity
Assessing Reliability

• Measure the same people twice

• If a test, ask essentially the same general question twice, but slightly tweak the numbers or context

• If not testing short-term knowledge or measurements that you’d expect to change over time, measure the same people at different points in time

• This should occur before your experiment, and doesn’t have to be a very large sample
Assessing Validity

• This is harder, particularly if you don’t have any proven way to measure the abstract quality you want data on.

• If you have any way to assess the abstract quality you would like to measure (perhaps through conversation?), perform this form of measurement along with the measurement you plan to use in your research to see if they agree.

• If you can’t see the bullseye and have no way to find it besides your proposed measurement, it will be very hard to determine validity.
Reliability and Validity

• *Much* easier: use an existing form of measurement that has already been proven to have both high reliability and validity

• Where can you find good existing measures?
  
  • online
  • colleagues, experts on the topic you are studying
  • see “Locating Scales” word document
Reliability and Validity

• Is your outcome variable something that may be able to be measured by existing measures or scales?

• If not, how might you go about creating a way of measuring your outcome variable that has both high reliability and high validity?

• How will you assess reliability and validity?
Questions

To consider when writing a question:

• Make sure everyone will interpret a question the same

• Don’t ask two questions in one ("Did you enjoy and learn from the experience?")

• Keep time in mind: don’t ask unnecessarily long questions, or a lot of probably irrelevant questions
Measurement

• What do you want to measure?
  • What is your outcome variable of interest?
  • What are your explanatory variables?
  • What covariates should you measure?

• What type of variables do you want these to be? (nominal, ordinal, or numeric)

• How do you plan on getting measurements of these variables with high reliability and high validity?