Introduction to Quantitative Analysis

MANSUREH KEBRITCHI, PH.D.
EDUCATIONALTECHNOLOGY@PHOENIX.EDU
CENTER FOR EDUCATIONAL AND INSTRUCTIONAL TECHNOLOGY RESEARCH (CEITR)
Agenda

Quantitative Analysis Steps
- Selecting statistical analysis
- Data entry

Definition of Basic Concepts
- Sample size

Basic analysis
- Descriptive and inferential

Resources
What is Quantitative Data?

Quantitative data can be expressed as a number or can be quantified.

Quantitative data can be measured by numerical variables.

- It can be quantified, verified, and counted.
- Can be collected from tests, surveys, reports.

Examples

- Scores on tests and exams e.g. 85, 67, 90.
- The weight of a person or a subject.
- The number of hours of study.
- The square feet of an apartment.
- The temperature in a room.
- The volume of a gas.
Quantitative Analysis Selection Criteria

- Identify focus of your study, research questions, design
- Identify your variables
- Identify level of measurements of your variables
Variables

A **variable** is defined as anything that has a quantity or quality that varies.

The dependent **variable** is the **variable** a researcher is interested in.

An independent **variable** is a **variable** believed to affect the dependent **variable**.

Confounding **variables** are defined as interference caused by another **variable**
Levels of Measurements

Nominal (categorical)
- Categories with no order, e.g., gender

Ordinal (categorical)
- Ordered categories, e.g., students’ grades

Interval (continuous)
- Ordered responses, equally spaced, no starting point, e.g., temperature,

Ratio (Continuous)
- Variable has an absolute zero, e.g., length, weight

Note. In SPSS, interval & ratio data are considered “scale”, and nominal and ordinal data have their own categories.
## Quantitative Research Designs

<table>
<thead>
<tr>
<th>Designs</th>
<th>Focus of the study</th>
<th>Hypothesis</th>
<th>Intervention</th>
<th>Statistical analysis</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-experimental Survey</strong></td>
<td>Observational; Describe the variables</td>
<td>No</td>
<td>No</td>
<td>Continuous V: Mean, STD, Categorical V: Mode</td>
<td>Description of teenagers’ attitudes toward smoking</td>
</tr>
<tr>
<td><em>(no hypothesis)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-experimental Correlation</strong></td>
<td>Test relationships; Prediction (criterion and predictor variables)</td>
<td>Yes</td>
<td>No</td>
<td>Interval/ratio V: Pearson r, Regression</td>
<td>Relationship between IQ and clinical depression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ordinal V: Spearman</td>
<td></td>
</tr>
<tr>
<td>**Non-experimental Causal-</td>
<td>Test cause and effect after the effect occurred (independent &amp; dependent variables)</td>
<td>Yes</td>
<td>No</td>
<td>Interval/ratio DV: T-test, Anova,</td>
<td>Effect of free/reduced lunch on students’ performance</td>
</tr>
<tr>
<td>comparative (ex post facto)</td>
<td></td>
<td></td>
<td></td>
<td>Nominal DV: Chi Square</td>
<td></td>
</tr>
<tr>
<td><strong>Quasi experimental</strong></td>
<td>Test cause and effect; Not a true random sampling (independent &amp; dependent variables)</td>
<td>Yes</td>
<td>Yes</td>
<td>Interval/ratio DV: T-test, Anova,</td>
<td>Effect of after school program on childhood obesity rate.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nominal DV: Chi Square</td>
<td></td>
</tr>
<tr>
<td><strong>Experimental</strong></td>
<td>Test Cause and effect; Random sampling (independent &amp; dependent variables)</td>
<td>Yes</td>
<td>Yes</td>
<td>Interval/ratio DV: T-test, Anova,</td>
<td>Effect of a new diet treatment on insulin levels in diabetics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nominal DV: Chi Square</td>
<td></td>
</tr>
</tbody>
</table>

- **Continuous V:** Mean, STD
- **Categorical V:** Mode
- **Interval/ratio V:** Pearson r
- **Ordinal V:** Spearman
- **Nominal DV:** Chi Square
Quantitative Sample Size Depends on

- **Effect Size (ES):** measures the strength of the relationship between two variables on a numeric scale; small, medium, large

- **Significance Criterion, α (P-Value):** The risk of mistakenly rejecting a true null hypothesis (H); 0.05
  - type I error is the rejection of a true null hypothesis

- **Power:** the probability that it will reject a false null hypothesis; .80
  - type II error is the non-rejection of a false null hypothesis

- **Analysis Test**
### Quantitative Sample Size, Cohen (1991)

**N for Small, Medium, and Large ES at Power = .80 for α = .01, .05, and .10**

<table>
<thead>
<tr>
<th>Test</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sm</td>
<td>Med</td>
<td>Lg</td>
</tr>
<tr>
<td>1. Mean dif</td>
<td>586</td>
<td>95</td>
<td>38</td>
</tr>
<tr>
<td>2. Sig r</td>
<td>1,163</td>
<td>125</td>
<td>41</td>
</tr>
<tr>
<td>3. r dif</td>
<td>2,339</td>
<td>263</td>
<td>96</td>
</tr>
<tr>
<td>4. P = .5</td>
<td>1,165</td>
<td>127</td>
<td>44</td>
</tr>
<tr>
<td>5. P dif</td>
<td>584</td>
<td>93</td>
<td>36</td>
</tr>
<tr>
<td>6. χ²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 df</td>
<td>1,168</td>
<td>130</td>
<td>38</td>
</tr>
<tr>
<td>2 df</td>
<td>1,388</td>
<td>154</td>
<td>56</td>
</tr>
<tr>
<td>3 df</td>
<td>1,546</td>
<td>172</td>
<td>62</td>
</tr>
<tr>
<td>4 df</td>
<td>1,675</td>
<td>186</td>
<td>67</td>
</tr>
<tr>
<td>5 df</td>
<td>1,787</td>
<td>199</td>
<td>71</td>
</tr>
<tr>
<td>6 df</td>
<td>1,887</td>
<td>210</td>
<td>75</td>
</tr>
<tr>
<td>7. ANOVA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2g^{a}</td>
<td>586</td>
<td>95</td>
<td>38</td>
</tr>
<tr>
<td>3g^{a}</td>
<td>464</td>
<td>76</td>
<td>30</td>
</tr>
<tr>
<td>4g^{a}</td>
<td>388</td>
<td>63</td>
<td>25</td>
</tr>
<tr>
<td>5g^{a}</td>
<td>336</td>
<td>55</td>
<td>22</td>
</tr>
<tr>
<td>6g^{a}</td>
<td>299</td>
<td>49</td>
<td>20</td>
</tr>
<tr>
<td>7g^{a}</td>
<td>271</td>
<td>44</td>
<td>18</td>
</tr>
<tr>
<td>8. Mult R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2k^{b}</td>
<td>698</td>
<td>97</td>
<td>45</td>
</tr>
<tr>
<td>3k^{b}</td>
<td>780</td>
<td>108</td>
<td>50</td>
</tr>
<tr>
<td>4k^{b}</td>
<td>841</td>
<td>118</td>
<td>55</td>
</tr>
<tr>
<td>5k^{b}</td>
<td>901</td>
<td>126</td>
<td>59</td>
</tr>
<tr>
<td>6k^{b}</td>
<td>953</td>
<td>134</td>
<td>63</td>
</tr>
<tr>
<td>7k^{b}</td>
<td>998</td>
<td>141</td>
<td>66</td>
</tr>
<tr>
<td>8k^{b}</td>
<td>1,039</td>
<td>147</td>
<td>69</td>
</tr>
</tbody>
</table>

**Note.** ES = population effect size, Sm = small, Med = medium, Lg = large, diff = difference, ANOVA = analysis of variance. Tests numbered as in Table 1.  
^{a} Number of groups.  
^{b} Number of independent variables.
Before we can do any kind of analysis, we need to quantify our data.

“Quantification” is the process of converting data to a numeric format.

- Convert social science data into a “machine-readable” form, a form that can be read & manipulated by computer programs.

Assign numeric representations to variables:

- Turning *male* into “1” and *female* into “2”
- Assigning “3” to *Very Interested*, “2” to *Somewhat Interested*, “1” to *Not Interested*
Goal – reduce a wide variety of information to a more limited set of variable attributes:

• Example, “What is your occupation?”
• Use pre-established scheme: Professional, Managerial, Clerical, Semi-skilled, etc.
• Create a scheme after reviewing the data
• Assign value to each category in the scheme: Professional = 1, Managerial = 2, etc.
Select a software and enter the data

Excel Spreadsheet, SPSS,

Create variables/column headings & enter each case
Type in variable name, type of variable (e.g., numeric, date, string), width of character (usually 8), decimals, Label (it is advisable to fill in this section with label because the variable name doesn’t always clearly Identify the variable, Missing (what to do if there are any missing values), Type of measure (scale, ordinal, nominal)
Entering Data to SPSS

Double click on var in top row, then type in the name of variable. You can start to type in data below each variable.
SPSS videos
https://ecampus.phoenix.edu/secure/aapd/cme/spss/tutorial/index.asp

Complete the following Activity
1. Click on [this link](https://ecampus.phoenix.edu/secure/aapd/cme/spss/tutorial/index.asp) to access the site for downloading SPSS in a new window. Note that you may be asked to log in the eCampus with your account.
2. Download the installer file according to your operating system.
Statistics Analysis Videos
https://ecampus.phoenix.edu/secure/aapd/AO/SPSS/

Statistics Help

Need more help on statistics? These short videos explain basic statistics concepts. Using lightboard technology and powerful examples, the videos walk you through each step. Use the practice problems to apply and evaluate your learning.

- Scales of Measurement
- Frequency
- Percentile Table
- Central Tendency
- Mode
- Median
- Mean
- Dispersion
- Range
- Variance
- Standard Deviation
- Probability
- Binomial Distribution
- Normal Distribution
- Z-Scores
- Sampling Methods
- Independent & Dependent Variables
- Central Limit Theorem
- Sampling Distribution
- Standard Error
- Confidence Intervals
- Hypothesis Testing
- Statistical Significance
- One and Two Tailed Tests
- Type I Errors
- Type II Errors
- One Sample T-Test
- Two Sample T-Test
Basic Analysis

Descriptive statistics

- Describe what the data on hand show, such as:
  - Distribution (frequency), Measures of central tendency (Mean, Mode), Measure of dispersion (Range, standard deviation)

Inferential statistics

- Allow you to make inferences about the population based on the data from your sample, such as:
  - Paired-Samples t-test
  - Independent Samples t-test
  - More complex: Correlation, Anova, Ancova, Regression
Descriptive Statistics

1. Measures of Distribution
   - Frequency distribution
   - Listing every value Vs. grouping data into categories
   - Counts Vs. percentages

2. Measures of Central Tendency
   - Mean (average)
   - Median (middle)
   - Mode (most common)
Descriptive Statistics

3. Measure of dispersion/spread of values

- Range: highest minus lowest
- Standard deviation: distance from mean
Inferential Statistics, Paired Sample t-test

Paired Sample t-test

- Samples are related
- Compare two sample means
- Same sample has been tested twice (pre/post)

Example

- In study on students’ math grades, all students are tested before receiving a treatment and after the treatment
Inferential Statistics, Independent Sample t-test

- Two samples drawn independently of each other
- Test for difference in sample mean from two groups
Inferential Statistics
Independent Sample t-test

Example

• Suppose the students are randomly assigned to a control and treatment groups
• Participants in the treatment group are given the treatment (e.g., mathematics game), control group is not given any treatment.
• After 4 months the two-sample t-test is used to compare math scores between the two groups. Each participant is measured once and belongs to one group.
Tips for data analysis

Leave enough time for data entry and data formatting

- Can take much longer than you expect

In your codebook – note the TYPE of variable for each measurement/question

This will allow you to plan the proper levels and types of analysis

https://statistics.laerd.com/ (Free membership for UOPX Faculty)
Thank you

Center for Educational and Instructional Technology Research

EDUCATIONALTECHNOLOGY@PHOENIX.EDU