Lab 10: Item Analyses

Item Analyses – Introduction

- We perform item analyses to evaluate individual items.
- The basic idea is to compute a series of specific descriptive statistics for each item and then evaluate those statistics.

When Can You Do Item Analyses?

When responses to items are scored as correct or incorrect.

- Achievement tests
- Exams
- Written driving tests
- NOT on most attitude or personality measures. Why?
Key Aspects of an Item Analysis

- Item Difficulty
- Item Discrimination

Item Difficulty

- Item responses are recoded:
  - Wrong answers are coded as 0
  - Right answers are coded as 1
- The mean of each recoded item is the percentage of people who got the item correct.
- It ranges from 0 (no one got the item right) to 1 (everybody got it right).

Example

- Imagine that you ask 100 people. 56 people give the right answer and 44 people give the wrong answer.
- What is the item mean?
Item Difficulty II

- This is the traditional way item difficulty is calculated.
- The *higher* the value, the *easier* the item (i.e., more people got it correct)
- Letter used to represent difficulty: \( p \)
- \( p = .77 \) means that 77% of individuals got the item right

The Scoring Office Report

- The scoring office tells you the Scoring Office Difficulty (SOD) of each item.
- This is the opposite of \( p \).
- \( SOD = (1 - p) \times 100 \)

- If \( p = .80 \), what is the SOD?
- Why does the scoring office do this?

An Easier Item

- Which of these is the name of a hit song by Katy Perry?
  A. *I Kissed a Boy*
  B. *Shake It*
  C. *I Kissed a Girl *** (142 out of 163)*
  D. *Pocketful of Sunshine*

- What is the \( p \)?
- What is the SOD?
A More Difficult Item

- What is the name of the woman with whom John Edwards had an affair that ended in 2006?
  A. Rielle Hunter *** (29 of 164)
  B. Elizabeth Rielle
  C. Rachel Hunter
  D. Nicole Rielle
- What is the $p$?
- What is the SOD?

Summary: Item Difficulty

- When few people get an item correct, the item is difficult, $p$ is low, and SOD is high.

- When a lot of people get an item correct, the item is easy, $p$ is high, and SOD is low.

SOD Rules of thumb

- The goal is to have moderately difficult items.

- Ideally, all items would be between 20 and 80 and most would be between 30 and 50.
The Scoring Office Report about an Exam

Look at items 1-5 on the report from the scoring office.

- Where can you find the item difficulty for each item?
- Which item is the easiest? Which is the most difficult?

Item Discrimination I

- An index of how well an item differentiates between people who did well on the test (and presumably knew the material) and those who did not do well (and presumably did not know the material).
- Calculation:
  \[ p \text{ for upper range of people (e.g., top 27%)} - p \text{ for lower range of people (e.g., bottom 27%) on an item} \]

Item Discrimination II

- Item discrimination values should always be positive.
- The higher the value, the better the item predicts how well people did on the test.
- Rule of thumb:
  
  .40 and above = very good
  
  .30 - .39 = ok
Point Biserial Correlation

• Another indicator of item discrimination.
• This is the relation between overall test score and correctly answering a given item.
• 0.4 and above is considered good.
• For example, if the point biserial correlation for an item is .43, we know that people who did well on the test were much more likely than low scorers to answer the question correctly.

The Scoring Office Report about an Exam

Look at items 1-5 on the report from the scoring office.

• Where can you find the item discrimination for each item?
• How did the scoring office compute the item discrimination?
• Which item discriminates best? Which worst?

Relation between Item Difficulty and Item Discrimination

• Consider a very easy item, e.g. item 1
• What percentage of the upper 27% of people on the test got it right?
• What percentage of the lower 27%?

→ This item does not discriminate cleanly between those who did well on the test and those who did not. This is called a ceiling effect.
Relation between Item Difficulty and Item Discrimination II

- Consider a moderately difficult item, e.g., item 3
- What percentage of the upper 27% of people on the test got it right? What percentage of the lower 27%?
- Items with moderate difficulty tend to have higher discrimination.
- If an item discriminates well, we are able to make a judgment about a person's ability based on whether or not s/he answers the item correctly.

Your Papers

- Part of your paper assignment is to learn how to write in APA format
- In both lecture and lab we have gone over several important areas
- You Will be graded on the following formatting issues:

Formatting Your Papers
Review

- Technical Details: Double-spaced, 12 point Font, 1 inch margins.
- Main Text: 10-12 pages text (10 pts off if paper is less than 10 pages or greater than 13 pages)
- Main text refers to the body of the paper. It does NOT include the References section, Cover page, and any Tables/Figures.
- If you have questions please ask your TA – you will lose points for formatting errors! (Such as using larger fonts or wider margins).
Summary of requirements for PSY 395 Papers
(SEE LAB 6 FOR DETAILS)

Introduction (30 points)
- Big Idea in First Paragraph
- Summary of Existing Literature
- Justification for Proposed Study
- Clearly Stated Null and Alternative Hypotheses
- Identify IV and DV

References Page Example
Method (30 points)
• Identify Sample and Sampling Plan
• Describe Procedure
• Describe and Justify Measures
• Plan of Analyses

Discussion (30 points)
• Discussion if Results were Consistent with Hypotheses
  – Summary of findings, implications, & future directions
• Discussion if Results were Not Consistent with Hypotheses
  – Summary of findings, implications, & future directions

References (10 points) & Writing/APA Style (20 points)
• Are there at least 3 journal articles?
• Are references in appropriate APA style?
• Are the in-text citations correct?
• Are there no violations of APA Writing Conventions?
• Is the paper edited and clearly written?
More info

- One of the main articles must be from 2009-2010.
- The paper should not be less than 10 pages or longer than 13 pages.
- Copies of the first page for all cited articles should be attached.
- There should be no direct quotations in the paper.
- Early bird bonus = 10 extra points (but total score CANNOT exceed 120 points).

Future Lab Sessions

- This is the last lab session in the room and with learning content.
- The final sessions are optional. They are project consultations held in TA offices NOT the computer lab. TAs will be available during the regularly scheduled lab sections.
- If you have project questions I can answer them during those sessions.

Extra Credit Opportunity: Homework 10

- This homework is optional but will give you 5 additional points on your homework grade.
- These points can be used to bring your total HW score up to the MAXIMUM score of 80.
- It is due next week at the regular time for homework assignments.
Optional Homework Part I

Use Attached Report

1. Report and interpret the item difficulty for items 6 – 10. (1 point)
2. Which of those 5 items is the most difficult? Which is the easiest? (1 point)
3. Report and interpret the item discrimination for items 6 – 10. (1 point)
4. Which of those 5 items has the highest discrimination? Which has the lowest? (1 point)

Optional Homework Part 2

5. Imagine the scoring office gives you these statistics for four different items. For each item, report whether
   – you would keep the item on the test or
   – revise the item in future versions of the test and explain your decisions! (1 point)
   
   Item 1: SOD = 5; discrimination index = 0
   Item 2: SOD = 90; discrimination index = 40
   Item 3: SOD = 90; discrimination index = 0
   Item 4: SOD = 50; discrimination index = 60