Week 7: Measuring the Dependent Variable II

Dr. T. Kody Frey

Assistant Professor | School of Information Science

Overview

- BRIEF review: Midterm Study Guide
- Measurement and Dependent Variables: Part II
- Discussion: Validity!
- Workshop: IRB / Measure Building

What's Next?

Week 8 2/27	**Making Sense of Your Data**
	The content this week is designed to help you understand what to do with data once it is collected. Before you can start to analyze, you have to get a better idea of what you are working with. Lecture content includes descriptive vs inferential statistics, parametric statistics, missing data, manipulation checks, and pilot testing.
	Workshop Topic: Survey Building in Qualtrics
	Readings
	*Gliner, J. A., Morgan, G. A., & Leech, N. L. (2017). _Research methods in applied settings: An integrated approach to design and analysis_ (3rd ed.). Routledge. *Read Ch 15, 16, 17*
	Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. _Journal of School Psychology, 48_(1), 5-37. https://doi.org/10.1016/j.jsp.2009.10.001
	Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. _The American Journal of Evaluation, 22_(2), 127-160. https://doi.org/10.1016/S1098-2140(01)00133-3
	Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. In A. E. Kazdin (Ed.), _Methodological issues and strategies in clinical research_ (pp. 547–555). American Psychological Association. https://doi.org/10.1037/14805-033
	Study Codebook Due
	Reaction Paper #6

Midterm Study Guide

You have a maximum of 10 pages for your typed, double-spaced responses; be accurate and concise (like a quantitative researcher).

CI 665-002 **Spring 2023 Midterm Exam STUDY GUIDE**

The exam will begin at 5 pm on Monday, March 6th. You will have until 7:45 to completely answer the questions. You may use your one page of notes, but please do not use any online or textbook resources (there is an example page of notes attached to the Midterm assignment description on Canvas). Exams will be typed and submitted in Canvas when complete. Remember that you have a maximum of 10 pages for your typed, double-spaced responses; be accurate and concise (like a quantitative researcher).

Be familiar with the following topics from the lecture, readings, and discussions:

1. Why is quantitative research important to the communication discipline?
2. What is internal validity? How is it evaluated? What are the potential major threats to internal validity and what can you do to reduce them?
3. What is external validity? How does a theoretical population differ from a sampling frame? An accessible sample? An actual sample? How do you determine the usefulness of a sampling frame?
4. Why is measurement important to quantitative research? How does one go about building a measure that is both reliable and valid? What are the critical steps?
5. What makes a good quantitative study? What criteria would you use to distinguish a good study from a poor study? What are some of the limitations of quantitative work?
6. What is the purpose of experimental research? What are its strengths and weaknesses?
7. What is the purpose of survey research? What are its strengths and weaknesses?
8. What are the characteristics of a good research question / hypothesis? Can you pose at least one for an idea that interests you? How would you design an experimental or survey-based approach to answering this RQ or hypothesis? How would your design address 1) research subjects (e.g., sampling, recruitment strategy, assignment to condition), 2) variables (conceptual and operational definitions for independent/dependent), 3) measures, 4) settings, 5) procedures, and 6) data analysis (not asking you to identify a specific statistical test though you can if you want – instead tell me whether you are comparing groups, identifying associations, or describing data)? Can you articulate the difficulties you might face in pursuing each research approach?

Part 1

Why is quantitative research important to the communication discipline? Among other reasons…

- Allows us to empirically test for patterns, causality, group differences, and theoretical propositions

- May not explain *why* a communicative phenomena occurred, but they do explain human behavior to fullest extent possible

- Allows us to explain, predict, control, and describe communication behavior

- Turn broad ideas in specific theoretical propositions

- Generalize findings to larger populations of people

- Allows us to control variables with a level of objective specificity that simply cannot be achieved through other approaches

- Reduce complex phenomena to measurable variables

Part 2

What is internal validity? How is it evaluated? What are the potential major threats to internal validity and what can you do to reduce them?

Definition | **Evaluation** | **Major Threats** | **Mitigating Threat**

Internal validity depends on the strength or soundness of the design and influences whether one can conclude that the independent variable or intervention caused the dependent variable to change
We evaluate it based on…

- Equivalence of groups on participant characteristics

- Control of extraneous experiences and environment variables

- How the research is conducted (e.g., the tools or instruments used)

- The research participants (e.g., were they randomly assigned)

- The researchers themselves (e.g., how do independent raters judge behavior)

- Random assignment

- Instrument selection

- Rater training

- Attend to needs and maintain contact

- Pre-testing

- Controlled environments

- Control groups

Part 3

What is external validity? How does a theoretical population differ from a sampling frame? An accessible sample? An actual sample? How do you determine the usefulness of a sampling frame?

Definition   **Levels**   **Usefulness**   **Big Picture**

External validity is the extent to which samples, settings, and variables can be generalized beyond the study.

Theoretical population: All the participants of theoretical interest to the research and to which he or she would like to generalize

Sampling frame: The group of participants you actually have access to, perhaps through a list or directory

Selected sample: Smaller group selected from the larger accessible population by the researcher and asked to participate in the study.

Actual sample: Participants that complete the study and whose data are actually used in the analysis and in the report of the results.

The sampling frame represents an *exhaustive list* of the participants that a researcher could realistically access for a study.

- Is the frame representative of the theoretical population?

- Does the frame include an exhaustive list of potential participants?

- How was the frame obtained?
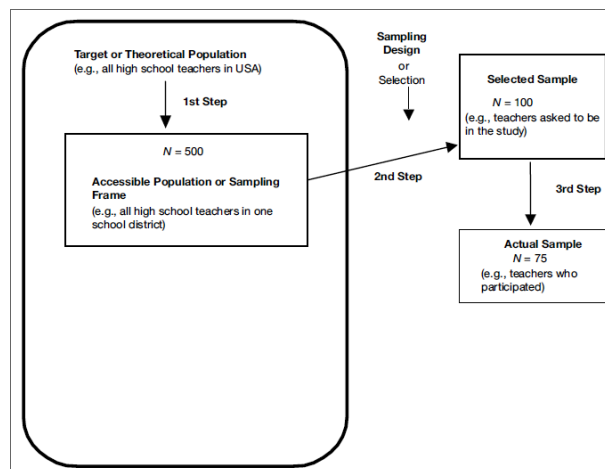
Part 4

Why is measurement important to quantitative research? How does one go about building a measure that is both reliable and valid? What are the critical steps?

| **Definition** | **Importance** | **Reliability and Validity** | **Steps** |

Measurement is the assignment of numbers or symbols to the different levels or values of variables according to rules.

In social science, we often want to assess ideas that we cannot directly observe. Effective measurement aligns conceptualization with operationalization and enhances internal validity.

Measurement also influences which statistics we use to draw conclusions and generalizations.

"Research results are no more valid than the measures used to collect the data" (Levine, 2005, p. 335).

"Poor measurement misrepresents the social phenomena being studied, and a lack of construct validity could lend support to inaccurate theories and distort accurate theories" (Bowman & Goodboy, 2020, p. 232).

Stay tuned!!

Reliability: Is the measuring performing consistently across respondents?

Validity: Whether the scores provide evidence for the use of a measure in a specific setting.

Reliability is necessary for validity, but one can have consistent data that is not valid.

Scale development is useful for capturing not directly observable concepts

- Need to *define* what you are interested in

- Need to create *multiple items* to capture abstract constructs

- Will use *EFA* to reduce multiple items down to meaningful few

Carpenter's Steps

Table 1. 10 steps in scale development and reporting.

1. Research the intended meaning and breadth of the theoretical concept
   a. Select appropriate conceptual labels
   b. Select conceptual definitions
   c. Identify potential dimensions and items
   d. Conduct qualitative research to generate dimensions and items
   i. Use feedback to refine scale
   i. Expert feedback, pre-tests, cognitive interviews, or pilot tests can be employed to evaluate item wording, item validity, questionnaire design, and model structure
2. Determine sampling procedure (5:1)
3. Examine data quality
4. Verify the factorability of the data
   a. Bartlett's Test of Sphericity (≤.05)
   b. Kaiser-Meyer-Olkin test of sampling adequacy (≥.60)
   c. Inspect correlation matrix (≥.30)
5. Conduct Common Factor Analysis
6. Select factor extraction method
   a. Principal Factors Analysis
   b. Maximum Likelihood
7. Determine number of factors
   a. Theoretical convergence and parsimony
   b. Scree test
   c. Parallel Analysis (PA)
   d. Minimum Average Partials (MAP)
8. Rotate factors
   a. Oblique rotation (Direct Oblimin, Promax)
9. Evaluate items based on a priori criteria
   a. Theoretical convergence
   b. Parsimony
   c. Weak loadings (≥.32)
   d. Cross loadings
   e. Inter-item correlations
   f. At least three-item factors
   g. Communalities of items (≥.40)
10. Present results
    a. Scale and subscale naming logic, conceptual definitions, sample size logic, methods for determining factor numbers, Bartlett's test of sphericity, Kaiser-Meyer-Olkin test of sampling adequacy results, factor extraction method, rotational method, strategies for deciding on items, eigenvalues for all factors, pattern matrix, computer program package, communalities for each variable, descriptive statistics, subscale reliabilities, and percentage of variance accounted for by each factor.

Part 5

What makes a good quant study? What criteria would you use to distinguish a good study from a poor study? What are some of the limitations of quantitative work?

Quant Research   **Criteria**   **Evaluation**   **Limitations**

- Is *systematic*. It is intentional, replicable, and valid

- Observes, explains, and predicts

- Tests theories that describe human behavior

- Answers questions of **group differences** or **variable relationships**

- External Validity

- Internal Validity

- Measurement Validity

- Probability vs. Nonprobability samples
  - Increasing (1) ecological and (2) population validity
  - Random *selection*

- Research Design
  - Ensuring (1) equivalence of groups and (2) control of extraneous variables
  - Random *assignment*

- Instrumentation
  - Increasing (1) reliability and (2) validity
  - Evidence for selection

- Look to the readings for specific examples (e.g., sampling, design, measurement, funding)

- Consider metatheory
  - Ontology, epistemology, axiology. Is absolute truth possible?
  - Is reality fixed and measurable?

- Consider ethical objections

- Consider characteristics needed to be good at it
  - Curious, knowledgeable, analytic, careful, meticulous, and resourceful

- What else?

Part 6

What is the purpose of experimental research? What are its strengths and weaknesses?

Experiments | **Strengths** **Weaknesses**

- Purpose is to discover *causal* relationships between variables
- *Active* independent variables
- Central characteristic is *control*
- Isolates specific relationships between variables of interest
- Better suited for theory testing
- Conclusions better reflect true, meaningful, and observed relationships within a physical, tangible world
- Complex
- Require careful attention to participants
- Increased control = decreased ecological validity
- Others?

Part 7

What is the purpose of survey research? What are its strengths and weaknesses?

| **Surveys** | **Strengths** | **Weaknesses** |

- Purpose is to discover how large groups of people think and act.
- Describes the characteristics of the respondents and the populations they were chosen to represent.
- Suited to answer questions about preexisting attributes of persons or their ongoing environment that do not change
- Flexibility and efficiency in data collection
- Sets the stage for later examining causality
- Lack of control
- Non-experimental by nature
- Relies on *attribute* independent variables
- MANY conditions necessary to trust generalizations
- Others?

Surveys

To make trusthworhty generalizations from surveys…

- Samples must be **representative**

- Response rates must be **sufficient**

- Questions must be **unbiased**

- Data collections procedures must be **uniform**

- Coding and analysis must be **accurate**

Part 8

What are the characteristics of a good research question / hypothesis? Can you pose at lea
How would you design an experimental or survey-based approach to answering this RQ c
How would your design address:

- Research subjects (e.g., sampling, recruitment strategy, assignment to conditions)

- Variables (conceptual and operational definitions for independent/dependent)

- Measures

- Settings

- Procedures

- Data analysis (not asking you to identify a specific statistical test though you can if you wa

- Can you articulate the difficulties you might face in pursuing each research approach?

An Example

Let's say your group wants to study men and womens' reactions to violent crime shows or

- Can you envision an experimental approach?
- Can you envision a survey approach?

With each:

- What is the IV? DV? Constant?
- Is this an active or attribute IV?
- What would be an appropriate sampling technique for your group?
- What are some problems that could affect *internal* validity?
- What are some problems that could affect *external* validity?

GML Chapter 11

We're finally talking about targets!

Study quality depends on the *consistency* and accuracy of measurement instruments.

Measurement Reliability

Think about what it means to be a reliable person.

Reflects the **consistency of a series of measurements**

Are people responding to a measure consistently over time?

If our outcome measure does not provide reliable data, then we cannot accurately assess the results of our study.

How do we know that a score is due to the intervention or due to some other, unsystematic factor?

Understanding Reliability

Reliability is a *coefficient*

The ratio of the variance of true scores to the variance of observed scores

The higher the reliability of the data, the closer the true scores will be to observed scores

Indicating Reliability

Reliability will be displayed as a correlation

This indicates the strength of the relationship between two variables

A strong positive relationship indicates that people who score high on one test also will score high on a second test. To say that scores from a measure are reliable, one usually would expect a coefficient between +.7 and +1.0. Reliabilities will not be negative (or something is wrong)

Assessing reliability

- Test retest
- Parallel forms
- Split-half
- KR-20
- **ALPHA**
- Percentage agreements
- ICC
- Kappa
- **OMEGA**

Alpha

α

- Measures **internal consistency**
  - The consistency of people's responses across the items on a multiple-item measure
  - Do all items reflect the same underlying construct?
- Only used when one has multiple items being combined for composite
- Mean of all possible split-half correlations for a set of items

Omega

ω

- Measures **internal consistency** but BETTER
- Only used when a measure is *unidimensional*
- The gold standard for reliability of composite scores
- Alpha depends on a variety of restrictive conditions that Omega bypasses

Application

An instrument of support was used to measure perceived support from coworkers in a mental health institution. Participants responded to four items on a seven-point Likert-like scale. Cronbach's alpha for the (support) scale was .79. What does this mean?

Choosing a Measure

- Past reliability needs to be high
  - This ensures participants will respond consistently in your study
- Samples need to be similar

Why is this important?

If we have less reliable measures, we are less confident that our participants' observed scores are close to their true scores.

Observed Scores

Any score that we obtain from an individual on an instrument
Observed score = True Score + Error
We cannot know the true score, but we can estimate where it might fall

Standard Error of Measurement

A range of scores (i.e., confidence interval) within which should lie a performer's true score.

Establishing Confidence

Most of the time, we want to be 95% sure that we capture the true score (2 standard deviations)

Putting it together

Lets say we are measuring *satisfaction*

Calculating SEM	**The Observed Score**	**Z Scores**	**Confidence Interval**

$\sigma^2 = 15\ \alpha = .92$ SEM $= 4.24$

Someone completes the combined measures and scores a 110 total

If the instrument has 20 Likert questions ranging from 1 to 7, possible values are 20 to 140.

The z score for any score that is 2 standard deviations away is 1.96

1.96 [Z score] * 4.24 [SEM] = 8.32

We can conclude that our true score falls within the 95 percent confidence interval of 110 ± 8.32 or between **101.68 and 118.32**

If the test were given to the same person a large number of times, 95 percent of the confidence intervals would contain the true score.

Counter Example

Lets say we are measuring *satisfaction*

Calculating SEM	**The Observed Score**	**Z Scores**	**Confidence Interval**

$\sigma^2 = 15\ \alpha = .65\ \text{SEM} = 8.87$

Someone completes the combined measures and scores a 110 total

If the instrument has 20 Likert questions ranging from 1 to 7, possible values are 20 to 140.

The z score for any score that is 2 standard deviations away is 1.96

1.96 [Z score] * 8.87 [SEM] = 17.39

We can conclude that our true score falls within the 95 percent confidence interval of 110 ± 17.39 or between **92.61 and 127.61**

We are *less precise* in our understanding of how participants are responding

GML Chapter 12

We're finally talking about targets!

Study quality depends on the consistency and *accuracy* of measurement instruments.

Measurement Validity

Concerned with *establishing evidence* for the use of a particular measure or instrument in a particular setting with a particular population for a specific purpose.

accumulating evidence to provide a sound scientific basis for the proposed score interpretations.

Are we hitting the target?

Differences from Reliability

Reliability is necessary for validity, but one can have consistent data that is not valid.
Suppose I used measurements of your dart throws to indicate your ability to pass this class.

- If you all hit the target in similar ways, we would have reliability

- Data is not valid because it is not giving information about ability to pass the course

Understanding Validity

Validity is tough to establish and takes multiple studies
One type of evidence is insufficient for validity
Often must demonstrate performance of measure in relation to other, similar measures
There is *no statistic* for validity

Types of Validity

- Face
- Content
- Criterion
  - Predictive
  - Concurrent
- Construct
  - Convergent
  - Discriminant
  - Factorial

Grit Example

Predictive: Do scores on grit predict graduation from West Point?

Concurrent: Does grit relate to amount of time studying for a test?

Convergent: Is grit related to a measure of a theoretically similar variable (e.g., resilience)

Divergent: Is grit different from a theoretically similar variable (e.g., different from coping)

Establishing Support

The goal should be to build evidence that a measure performs in ways one would expect.

- Use experts of focus groups to establish that *content* represents concept

- Establish correlations with other measures we theoretical expect it to

- Establishing the *nomological network*

Frey Facts

In general, it is advisable to select instruments that have been used in other studies if they have been shown to produce *reliable* and *valid* data with the types of participants and for the purpose that you have in mind.

Summary

- Measurement is complex!
- Make sacrifices when you choose a specific type of measure
- Reliability involves the **consistency** of reponses
- Validity reflects the **accuracy** of responses

Workshop

Associated Steps

For your projects…

- Need to *define* what you are interested in

- Need to create *multiple items* to capture abstract constructs

- Will use EFA to reduce multiple items down to meaningful few

Phrasing items and questions

General Guidelines

- Language is simple, straightforward, and appropriate for reading level
- Ask about one (and ONLY one) issue
  - Avoid double-barreled
- Questions should not be LOADED (leading)
- Avoid using emotionally charged language
- Avoid double negatives
- Avoid trendy expressions-
- Avoid items everyone or no one will endorse
- Avoid mixing items that assess behaviors with items that assess affective responses
  - Ex: My boss is hardworking vs I respect my boss

Developing an item pool

Fundamental goal at this stage is to sample systematically all content that is potentially relevant to the target construct.
You will drop weak items but cannot add them back.
Recommend Likert scale but you can use semantic differential if you really want.

This Workshop

- Pick a concept that your group wants to measure for project

- Write a definition

- Come up with 25-30 items based on that definition

- Focus groups!! Use OneDrive doc on Canvas

Reviewers

- Read definition closely
  - What is missing or runs counter to your knowledge?
- Read items closely
  - Does the wording make sense?
  - Do the items reflect the definition?

tkfr222@uky.edu | **@TKFrey** | **tkodyfrey.github.io**

University of
Kentucky