WEEK 6: MEASURING THE DEPENDENT VARIABLE I

Dr. T. Kody Frey

Assistant Professor | School of Information Science

OVERVIEW

- BRIEF review
- Measurement and Dependent Variables: Part 1
- Discussion: Building Meaningful Measures
- Workshop: Let's Measure
  - Operationalizing Variables
  - Applying to your study

WHAT'S NEXT?

Week 7 2/20	**Measuring the Dependent Variable II**
	Our second lesson on measurement will help us determine what our measures mean. We will think about the ways we can offer increased support for measures going forward, as well as the ways in which attempts to objectively capture subjective ideas should be challenged and critiqued.
	Workshop Topic: Construct Validation
	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 Chs 11, 12*
	Bowman, N. D., & Goodboy, A. K. (2020). Evolving considerations and empirical approaches to construct validity in communication science. *Annals of the International Communication Association*, *44*(3), 219-234. https://doi.org/10.1080/23808985.2020.1792791
	Frey, T. K., & Tatum, N. T. (2022). Instructor strictness: instrument development and validation. *Communication Education*, *71*(4), 327-354. https://doi.org/10.1080/03634523.2022.2096246
	Reynolds-Tylus, T., Bigsby, E., & Quick, B. L. (2021). A comparison of three approaches for measuring negative cognitions for psychological reactance. *Communication Methods and Measures*, *15*(1), 43-59. https://doi.org/10.1080/19312458.2020.1810647
	Reaction Paper #5

OLD CONTENT: REVIEW!

INTERNAL VALIDITY

WHAT IS IT?

Internal validity is the extent to which we can infer that that IV caused the change in the DV.

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

CHARACTERISTICS

What are the two characteristics used to evaluate internal validity?
Internal validity refers to the extent that the independent variable, treatment, or intervention caused the change in the dependent variable. We evaluate it based on...

- Equivalence of groups on participant characteristics

- Control of extraneous experiences and environment variables

THREATS

What are the broad threats to internal validity?

- 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)

Remember the sheet I gave you!

THREATS TO EQUIVALENCE OF GROUPS

Regression | **Attrition** | **Selection**

Measurement is often unreliable. Participants who score low on a measure may score higher (closer to the mean) the next time.

Participants drop out for a number of reasons, but it influences group composition.

Problems are created when groups are assigned based on similarities - not randomness.

THREATS TO CONTROL OF EXTRANEOUS VARIABLES

- Maturation
- History
- Testing
- Instrumentation
- Interactive Threats
- Ambiguous Temporal Precedence (Time)

THREATS RA DOES NOT ELIMINATE?

- People talk
- Expectation effects (Hawthorne)
- Observer bias

EXTERNAL VALIDITY

WHAT IS IT?

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

SAMPLING LEVELS

Theoretical population   **Accesible Population**   **Selected Sample**

Actual Sample

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

Also called *sampling frame*

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

The smaller group of participants selected from the larger accessible population by the researcher and asked to participate in the study.

The participants that complete the study and whose data are actually used in the data analysis and in the report of the study's results.

SAMPLING

What are the characteristics of a sample frame researchers should evaluate in determining its usefulness?
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?

ENSURING REPRESENTATIVENESS

How do we ensure that a sample is representative of the target theoretical population?
Random selection is important for high external validity.
Random assignment is important for high internal validity.

TYPES OF SAMPLING

Probability Sampling   **Nonprobability Sampling**

Everyone has a known, nonzero change of being chosen:

- Simple Random Sample
- Systematic Random Sampling
- Stratified Random Sample
- Stratified with Diff Probs of Being Selected
- Cluster (Random) Sampling

No method to estimate probability of being included:

- Quota Sampling
- Purposive Sampling
- Purposeful Sampling
- Convenience Sampling
- Snowball Sampling

TYPES

What are the *two* types of external validity?
External validity is the extent to which samples, settings, and variables can be generalized bey
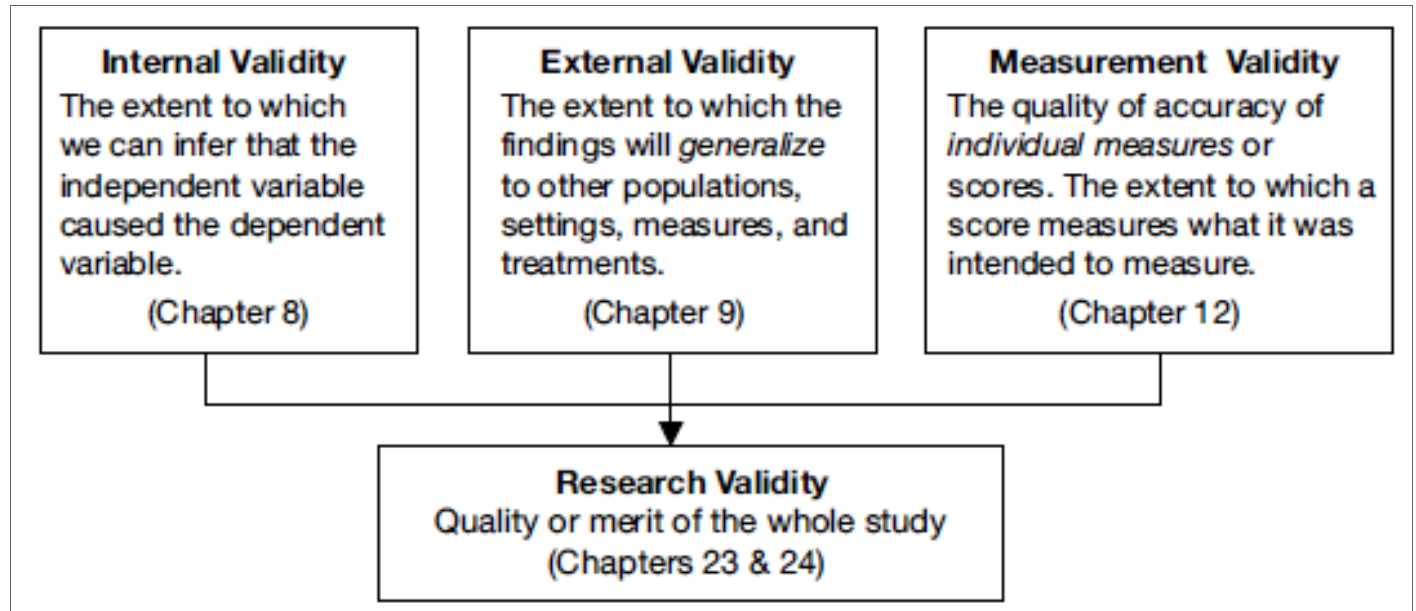
Population external validity	**Ecological external validity**

- Was the *sampling frame* representative of the theoretical pop?
- Was the *selected sample* representative of the population?
- Was the *actual sample* representative of the population?

  Whether the conditions, settings, times, testers, or procedures are rep of natural conditions
  AKA is the research environment similar to the natural environment? Does the manipulation

OTHER TYPES OF VALIDITY?

Internal Validity
The extent to which we can infer that the independent variable caused the dependent variable.
(Chapter 8)

External Validity
The extent to which the findings will *generalize* to other populations, settings, measures, and treatments.
(Chapter 9)

Measurement Validity
The quality of accuracy of *individual measures* or scores. The extent to which a score measures what it was intended to measure.
(Chapter 12)

Research Validity
Quality or merit of the whole study
(Chapters 23 & 24)

POWER

Power is the probability of detecting an effect, given that the effect is really there.
In other words, it is the *probability of rejecting the null hypothesis when it is in fact false.*

Alpha	**Beta**

The probability of a Type I error (reject a true null).
For a test with a level of significance of 0.05 = 1/20, a true null hypothesis will be rejected on
We are willing to live with a 5% chance that we will conclude that there is a difference when
The probability of a Type II error (fail to reject a false null)
The probability that we would accept the null hypothesis even if the alternative hypothesis is
If power is .80 or 80%, then beta is .2 or 20%

KEY TERMS

What is the difference between a conceptual definition and an operational definition?

These are *not* interchangeable

- **Concept** (Theoretical concept) - a mental representation (more abstract than a construct)

- **Construct** - a set of operational measures that allow for the study of a theoretical concept (less abstract than a concept)

EXPERIMENTS VS. SURVEYS

Experiments	**Surveys**

- *Active* independent variables
- Tests *causal* relationships
- Isolates specific relationships between variables of interest
- Central characteristic is *control*
- Better suited for theory testing
- Conclusions better reflect true, meaningful, and observed relationships within a physical, tangible world
- Non-experimental by nature
- Relies on attribute independent variables
- Examines the *presumed* effect of IV on DV
- Suited to answer questions about preexisting attributes of persons or their ongoing environment that do not change
- Discovers how larger populations think and act without central component of control
- Sets the stage for later examining causality

IVS AND DVS

When we discussed the IV, we discussed designing experiments that allow us to *control* the v
In other cases, the variables we are interested in might be *continuous.*
This is most often the *dependent* variable (i.e., what is thought to be changed by the IV)

MEASUREMENT: DEPENDENT VARIABLES PART I

MEASUREMENT THEORY (STEVENS, 1958)

4 levels of measurement used to describe the range and the relationship among the values a variable can take.

.

.

WHY IS THIS IMPORTANT?

Depending on the level, the data can mean different things.
For example, the number 2 might indicate a score of two; it might indicate that the participant was a male; or it might indicate that the participant was ranked second in the class.

NORMALITY

The normal curve provides a model for the fit of the distributions of many of the dependent variables used in the behavioral sciences.

DEFINITIONS

X axis: scores or responses on an ordered variable from very low to very high
Y axis: number of participants with a particular response

APPLYING TO THE CURVE

MORE ON THE NORMAL CURVE

Think of it as a *probability* distribution
What is the probability of a participant's typical response?
If externally valid, should align with probability distribution of the theoretical population.

PROPERTIES

5 properties that are ALWAYS present:

- Its unimodal
  - Only has 1 hump in the middle
- Mean, median, mode are equal
- Curve is symmetric; it is not skewed
- Range is infinite
  - Extremes approach but never touch the x
- Neither too peaked or too flat
  - It has no kurtosis
  - leptokurtic
  - platokurtic

NORMALLY DISTRIBUTED VARIABLES

Ordered from low to high, responses are at least approximately normally distributed in the population from which the sample as selected

A number of statistical tests rely on the assumption that the distribution is normal

DESCRIPTIVE STATISTICS: CENTRAL TENDENCY

| Mean | Median | Mode |

- Sum of raw scores divided by number of observations in the sample
- Arithmetic average
- Appropriate for normal data

- The mid-point of the raw scores in a sample
- Appropriate for ordinal or skewed data

- Equal to the raw score that appears most frequently
- Least precise
- Useful for nominal or dichotomous data or few categories

WHICH DO I USE?

When data are normally distributed, the mean, median, and mode are all the same and in the center of the distribution.
If data are normal, mean is the stat to use.

DESCRIPTIVE STATISTICS: VARIABILITY

Variability describes the spread or dispersion of the scores.
If all scores are the same, there is no variation.

APPLICATION!

In your research, you test to see if an IV can explain the variation in the DV.
Is the IV the reason that the **values of the DV varied from person to person?**

DESCRIPTIVE STATISTICS: VARIABILITY

Standard Deviation is the most common. It is a measure of how scores vary about the mean.
Gives you a sense of the spread of the data

- Are they close to the mean or all apart?

- How much variability is in the sample?

.

Remember the normal curve?

If $\bar{x}$ = 70 and s = 15.22
For population: μ = 70 and σ^2 = 15.22

STANDARDIZING THE NORMAL CURVE

Can convert a normal curve into a standard curve by setting mean equal to zero and SD equal to 1

Calculated by subtracting the mean from each data point, and then dividing the difference by the standard deviation of the population

Proportions are always the same. This allows comparisons for curves with different means
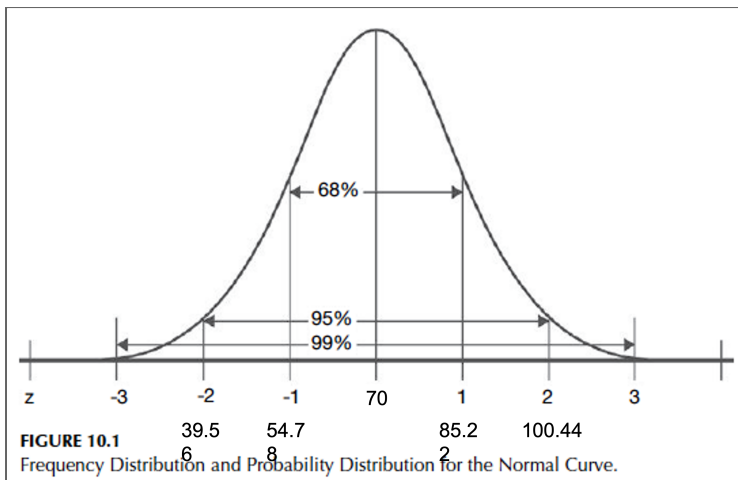
Z-SCORES

A standard score that indicates the number of standard deviation units that a person's score deviates from the group mean

AN EXAMPLE

Let's say $\bar{x}$ = 70 and s = 15.22 represent values for student test scores.

- A student who had 93 on the test would have a z score of **+1.51** **(93–70 ÷ 15.22)**

- A student who scored 43 had a **z of −1.77**.

FIGURE 10.1
Frequency Distribution and Probability Distribution for the Normal Curve.

WHAT MAKES THIS USEFUL?

Allows us to **categorize outliers** (if z > 3.29 or 3 standard deviations)

- If a variable is perfectly normally distributed, only 0.1% of values fall outside this range.

Allows you to **compare scores** on different tests.

- Student with an exam score of 80 (in a class whose mean was 70 and **SD was 5**) has a z score of +2.0 and did relatively better than the student (in a class whose mean was 70 and SD was 15.22) who had a test score of 93 and z = 1.51.

BACK TO CH. 13: DATA COLLECTION TECHNIQUES

What techniques and instruments do we use to collect data?
Major Types of Data Collection Techniques:

- Researcher observed measures

- Tests and Documents

- Self-report measures

OVERVIEW

RESEARCHER OBSERVED MEASURES

Direct Observation

- Trains observes to observe and record behaviors of participants in study
- Differs based on:
  - Naturalness of the setting
  - Degree of Observer Participation
  - Amount of Detail
  - Breadth of Coverage

TESTS AND DOCUMENTS

A set of problems with right or wrong answers

- Standardized tests
- Achievement Tests
- Performance and Authentic Assessments
- Aptitude tests
- Documents

SELF-REPORT MEASURES

- Standardized Personality
  Inventories
- Attitude scales
  - Summated Likert
    scale
  - Semantic Differential

1. Every person should give some of his time for the good of his town or country.	SD D U A (SA)
2. Letting your friends down is not so bad because we can't do good all the time.	(SD) D U A SA
3. It is the duty of each person to do his job the very best he can.	SD D U (A) SA

QUESTIONNAIRES AND INTERVIEWS

Survey research methods!

| **Open ended questions** | **Partially open-ended unordered items** |

Closed ended unordered items   **Closed ended questions**

- Formulate answers in own words
  - Demanding for participants
  - Especially useful to build knowledge for closed ended questions!
- Give possible answers then have space for response or comments
  - People usually don't use the spaces or give useful info
- Used when answers to a question fit nominal categories that do not fall on a continuum
- Ordered choices
  - Probably what you see most often in measures of attitude
  - All allowable responses given

FINAL THOUGHTS:

WHY is the measurement theory important in quantitative statistical analysis??
BOTTOM LINE: Levels of measurement influence the appropriate use of statistics!

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.

If those aren't available…

DEVELOPING A MEASURE

Scale development is useful for capturing not directly observable concepts

- This is what is meant by 'latent'

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**