

Week 9

Measurement

Overview of Research Process

1. Problem formulation
2. Methodology
 - **Operationalization and measurement**
 - Study population and sampling
 - Research design
 - Data collection
 - Data analysis plan
3. Implementation- Data collection
4. Data analysis
5. Dissemination

Today's class

- Operationalization of Variables
- Ways to Measure/Categories
- Sources of Measurement Error
- Reliability
- Validity
- Avoid Measurement Errors
- Using Existing Scale
- Designing Questions

Operational Definitions

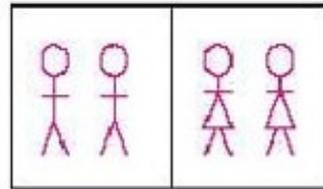
- In quantitative research we must first translate variables into operational definitions (observable terms) before a study is implemented
- Operational definitions differ from nominal definitions. For example, consider the concept *level of social adjustment*:
 - *Nominal definition*: “How well people perform their major roles in life”
 - *Operational definition*: “Score on a scale that measures social adjustment”

Ways of Measuring

SCALE

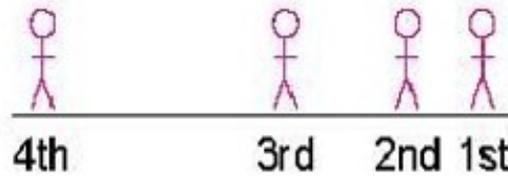
EXAMPLE

Nominal:



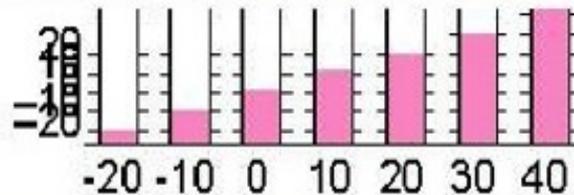
Gender

Ordinal:



Happiness A semantic differential scale question

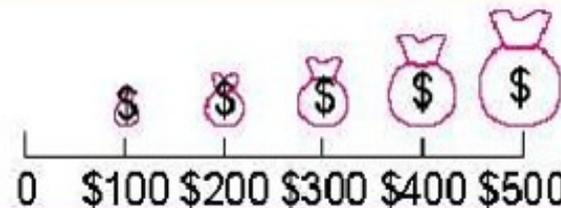
Interval:



Temperature (in Fahrenheit) IQ score

Ordered, Equal differences

Ratio:



Money

Zero

Nominal Level of Measurement

- Describes a variable in terms of the number of cases in each category of that variable.
- Examples
 - gender
 - ethnicity
 - religious affiliation

Ordinal Level of Measurement

- Describes a variable whose categories can be rank-ordered according to how much of that variable they are.
 - We know only whether one case has more or less of something than another case, but we don't know precisely *how much* more.
- Examples
 - Likert scale and a semantic differential scale
 - Happiness
 - level of client satisfaction
 - brief rating scale:

Amount of anxiety you feel:				
1	2	3	4	5
Little or none	Some	Moderate amount	Strong	Intense

Interval Level of Measurement

- **Interval:** differences between different levels have the same meanings.
 - Example: IQ. The difference between an IQ score of 95 and 100 is the same in magnitude as the difference between 100 and 105.
 - “**The Likert scale** is widely used in social work research and is commonly constructed with four to seven points. It is usually treated as an **interval scale**, but strictly speaking it is an **ordinal scale**, where arithmetic operations cannot be conducted.” (Wu & Keung, 2017)
 - Likert scale, Semantic Differential scale.

Ratio Level of Measurement

- **Ratio:** have the same attribute as interval measures, but in addition have a true zero point.
 - Example: # of arrests- possible to have no arrests, one arrest... Because there is a true zero point, we know that the persons with 4 arrests has been arrested exactly twice as many time as person with 2 arrests.
 - Zero of IQ test differ across the types of test: Stanford-Binet Intelligence Scale, Universal Nonverbal Intelligence, Differential Ability Scales. (Not true 0)

Illustration of levels of Measurement



LEVELS OF MEASUREMENT

01

NOMINAL

Named variables

ORDINAL

Named + ordered variables

02

03

INTERVAL

Named + ordered + proportionate interval between variables

RATIO

Named + ordered + proportionate interval between variables
+ Can accommodate absolute zero

04

More than one indicator?

Measuring life satisfaction

- Single item (one question)

e.g., How many days are you satisfied during the last 30 days?

- Indexes: Composite/cumulative measures

e.g., Choose yes or no

When I think about myself, I feel downhearted and blue.

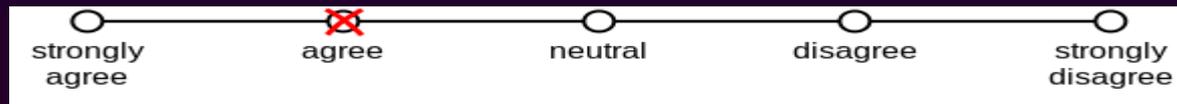
I often get tired for no reason

- Scales: composite measure that is composed of several items

e.g., Likert scale 0-7

In most ways my life is close to my ideal

The conditions of my life are excellent



Example

- Variable- Life satisfaction
- Operationalization- Scores on the Satisfaction With Life Scale (SWLs), which is a short 5 item instrument to measure global cognitive judgements of satisfaction with one's life ranging 5-31.

The Satisfaction with Life Scale

DIRECTIONS: Below are five statements with which you may agree or disagree. Using the 1 -7 scale below, indicate your agreement with each item by placing the appropriate number in the line preceding that item. Please be open and honest in your responding.

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Slightly Disagree
- 4 = Neither Agree or Disagree
- 5 = Slightly Agree
- 6 = Agree
- 7 = Strongly Agree

- ___ 1. In most ways my life is close to ideal.
- ___ 2. The conditions of my life are excellent.
- ___ 3. I am satisfied with life.
- ___ 4. So far I have gotten the important things I want in life.
- ___ 5. If I could live my life over, I would change almost nothing.

How Do We Measure?

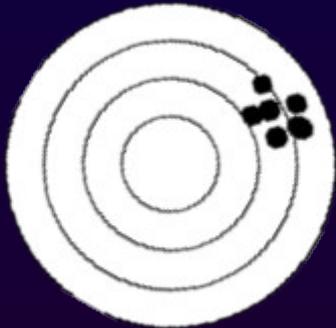
1. Verbal report
2. Observation
3. Archival records
 - Triangulation
4. Using existing scales and indexes via self-report
 - Popular way to operationally define variables
 - Consider the quality of existing scales and indexes (Reliability- Alphas run from 0 to 1)

Reliability and Validity

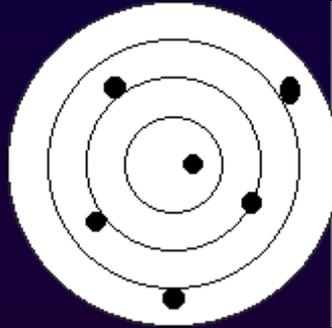
- Reliability- a particular measurement technique, when applied repeatedly to the same object, would yield the same result each time (Consistency or stability)
- Validity- the degree to which a measure covers the range of meanings included within the concept (Accuracy)

Relationship between Reliability and Validity

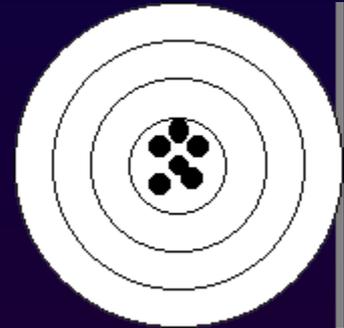
- Reliability does not ensure validity



Reliable but
not valid



Neither reliable
nor valid



Valid and
reliable

Types of validity

Construct validity....when we are able to generalize about our construct of interest because we are accurately measuring that construct. Think of construct validity as being truthful in how we are labeling our construct.

Sources of Measurement Error

- Measurement Error: Data do not accurately reflect the concept we attempt to measure
 - Systematic error: when the information we collect consistently reflects a false picture; related to validity (e.g. Social desirability bias; Cultural bias)
 - Random error: no consistent patterns; errors by chance; related to reliability (e.g. Cumbersome, complex, boring measurement procedures; Measure uses professional jargon which respondents are not familiar with)

Errors in Alternative Forms of Measurement

- Written Self-reports
 - Item wording
 - Words vs. deeds
- Direct Behavioral Observation
 - Social desirability bias
 - Observers might be biased
- Interviews
 - Social desirability bias
 - Different interviewers
- Examining Available Records
 - Practitioners might exaggerate their records
 - Improper documenting

Avoiding Measurement Error

- Use unbiased wording in questionnaires and survey
- Use understandable terms
- Obtain collegial feedback
- Pilot testing
- Triangulation
- Thorough training interviewer/observer

Designing Questions Guidelines

1. Closed VS. open-ended questions

- Exhaustive and mutually exclusive

2. Wording of questions

- Simple, short, direct and clear
- Never use slang terminology

3. Avoid double-barreled questions

- “Should the state abandon its community-based services and spend the money on improving institutional care?”

Designing Questions

4. Culturally sensitive wording choice

5. Respondents Must be Willing to Answer

6. Questions Should be Relevant -

Study results are not useful if questions are not relevant to respondents

7. Short Items are Best

8. Avoid Words like No or not- easy misinterpretation

Designing Questions

9. Structure and Design of Questionnaires

- Be easy on the eyes and easy to complete
- Directions: clear & Precise
- Order: Begin with easy, interesting items, opinion first, factual later; Put sensitive questions at the end

Critically Appraising Quantitative Instruments: Questionnaires

- Spread out and uncluttered
- Format for Respondents
 - Use genuine boxes
- Contingency Questions
 - Respondents will only answer questions that are relevant to them

Have you ever smoked marijuana?

Yes

No

Filter question

Contingency question

If yes, about how many times have you smoked marijuana?

Once

2 to 5 times

6 to 10 times

11 to 20 times

more than 20 times

Circle one response for each of the following ten items.

	<i>Strongly Agree</i>	<i>Agree</i>	<i>Disagree</i>	<i>Strongly Disagree</i>
1. I feel that I am a person of worth, at least on an equal basis with others.	1	2	3	4
2. I feel that I have a number of good qualities.	1	2	3	4
*3. All in all, I am inclined to feel that I am a failure.	1	2	3	4
4. I am able to do things as well as most other people.	1	2	3	4
*5. I feel I do not have much to be proud of.	1	2	3	4
6. I take a positive attitude toward myself.	1	2	3	4
7. On the whole, I am satisfied with myself.	1	2	3	4
*8. I wish I could have more respect for myself.	1	2	3	4
*9. I certainly feel useless at times.	1	2	3	4
*10. At times I think I am no good at all.	1	2	3	4

Sensitivity to Diversity in Measurement

- Refine measurement, if necessary
- Use culturally sensitive measures
 - Use key informants
 - Translation-back translation
 - Pilot testing