5. MEASUREMENT TECHNIQUES & SAMPLING METHODS

1. STEVEN'S FOUR SCALES OF MEASUREMENT

- NOMINAL

Used to name, categorize or classify, e.g. gender, marital status, personality type

- ORDINAL

Used to rank individuals or objects, e.g. social class, ranking of need for therapy

- INTERVAL

Used to rank order, plus has equal intervals or distances between adjacent numbers, e.g. Celcius/Fahrenheit temperature, IQ scores

- RATIO

Fully quantitative, includes rank order, equal intervals, and has an absolute zero point, e.g. Kelvin temperature, response time, height, weight, annual income

2. RELIABILITY

Reliability refers to consistency or stability of scores. Frequently, reliability coefficients are obtained as quantitative indexes – should be strong and positive (i.e. >.70) to indicate a strong consistency relationship. Four primary types of reliability:

- TEST-RETEST RELIABILITY

Consistency of a group of individuals' scores on a test over time. Generally, the longer the time interval between two testing occasions, the lower the reliability coefficient will be

- EQUIVALENT-FORMS RELIABILITY

Consistency of scores obtained on two equivalent versions of the same test or research instrument designed to measure the same thing, e.g. college entrance exams. Success hinges on the equivalence of the two forms of the test.

- INTERNAL CONSISTENCY RELIABILITY

Consistency with which items on a test measure a single construct, e.g. various personality dimensions. Multiple items on a test must be used to provide a sufficient measurement of the construct, and reliability generally increases with the length of the test (goal is to obtain high reliability with few items). Most commonly reported index of internal consistency is the *coefficient alpha* (*Cronbach's alpha*), which should be >.70.

- ### INTERRATER RELIABILITY

The degree of consistency or agreement between two or more scorers, judges, observers or raters, e.g. markers of essay papers. The *interobserver agreement* is the percentage of times the raters agree.

3. VALIDITY

Accuracy of inferences, interpretations, or actions made on the basis of test scores. Because tests and research instruments always involve the measurement of constructs (e.g. intelligence, personality), psychometricians agree all validity types are part of *construct validity*.

- **Operationalization:** (*Operational Definition*) The way a construct is represented and measured in a particular research study. The important issue is whether the operations produce a correct or appropriate representation of the intended construct.

A summary of methods for obtaining evidence of validity:

- ### EVIDENCE BASED ON CONTENT

Experts on the construct examine the test/scale content and determine whether the content adequately represents the construct.

- ### EVIDENCE BASED ON INTERNAL STRUCTURE

Some tests/instruments measure one general construct (unidimensional), but others measure several dimensions of a *multidimensional construct* (two or more factors).

- **Factor Analysis:** is used to determine the number of dimensions in a set of items – a statistical program may indicate whether items are interrelated or whether there are subsets of items closely related.
- **Homogeneity:** is the degree to which a set of items measures a single construct of trait, calculated by *item-to-total correlation* and *coefficient alpha*.

- ### EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

Determine whether the scores are related to known criterion by collecting concurrent and predictive validity evidence. Also determine if the test/scale scores are strongly correlated with participants' scores from another measure of the same construct (*convergent validity evidence*) and NOT correlated with scores and measures of different constructs (*discriminant validity evidence*). Lastly, determine if groups that are known to differ on the construct are accurately classified by the scale under consideration (*known groups validity evidence*).

4. SAMPLING

The sample is the set of elements selected from a population; sampling refers to the process of choosing that group. To obtain a sample with high validity and reliability, the sample must be a sufficient size, minimize the potential for bias or error and representative of the larger population (*equal probability selection method: EPSEM*).

- ### RANDOM SAMPLING TECHNIQUES

- **Simple Random Sampling:** Popular basic equal probability selection method
- **Stratified Random Sampling:** Division of population into mutually exclusive groups and then selection of random sample from each group

- **Proportional Stratified Sampling:** Stratified sampling where the sample proportions are made to be the same as the population proportions on the stratification variable (opposite to disproportional stratified sampling).
- **Cluster Random Sampling:** Where clusters are randomly selected, e.g. neighborhoods, schools, families
- **Systematic Sampling:** Uses a sampling interval (k) and randomly selects an element between 1-k, selecting every kth element. *Periodicity* is a rare potential problem in systematic sampling if there is a cyclical pattern in the sampling frame.

- NON-RANDOM SAMPLING TECHNIQUES
  - **Convenience Sampling:** Use of people readily available, volunteer, or easily recruited, e.g. intro psych students.
  - **Quota Sampling:** Researcher sets quotas (number of kinds of people desired) and locates (using convenience sampling) to meet quotas
  - **Purposive Sampling:** Researchers specifies characteristics of the population and locates individuals who have those characteristics
  - **Snowball Sampling:** Each sampled person is asked to identify another potential participant with the inclusion characteristic, useful to select from a hard-to-find population.

5. RANDOM SELECTION VS RANDOM ASSIGNMENT

- RANDOM SELECTION

Selection of participants using a random sampling method. The purpose is to obtain a sample that represents the population. ***Note: Simple random sampling = Representative sample***

- RANDOM ASSIGNMENT

Placement of participants into experimental conditions on the basis of a chance process, used in experimental research to produce treatment and control groups (or comparison groups) that are similar on all possible characteristics – *a key element in producing the strongest experimental designs available for the study of causation.*

6. DETERMINING SAMPLE SIZE

A *sample size calculator*, a statistical program, is used to provide a recommended sample size. You will need a larger sample size when:

- Your population is heterogeneous (i.e. composed of widely different kinds of people)
- If you want to break down data into multiple categories
- If you want to obtain a relatively precise confidence interval
- If you expect a weak relationship or small effect
- If you use a less efficient technique of random sampling (e.g. cluster sampling is less efficient than proportional stratified sampling)
- If you expect a low response rate