Topic #3

VARIABLES AND MEASUREMENT

VARIABLES

Some **definitions** of variables include the following:

1. symbol that can assume a range of numerical values
2. some property of an organism/event that has been measured
3. an aspect of a testing condition that can change or take on different characteristics with different conditions
4. an attribute of a phenomenon

Types Of Variables

1. **Independent and Dependent variables**

   A. An **independent** variable is the condition manipulated or selected by the researcher to determine its effect on behavior.

      ■ The independent variable is the ANTECEDENT variable and has at least 2 forms or levels that defines the variation.

   B. A **dependent** variable is a measure of the behavior of the participant that reflects the effects of the independent variable.

      ■ The dependent variable is the CONSEQUENCE.

   • A specified variable is not ALWAYS an independent or dependent variable. The designation of a variable as an independent or a dependent variable will change (or stay the same) as a function of the particular research study.

2. **Continuous and Discrete variables**

   A. A **continuous** variable is one that falls along a continuum and is not limited to a certain number of values (e.g., distance or time).

      ■ are infinitely divisible into whatever units a researcher may choose.

   B. A **discrete** variable is one that falls into separate categories with no intermediate values possible (e.g., male/female, alive/dead, French/Dutch, flying/walking).

- Discrete variables are also categorical in form.

- Categorical and categorical-ordered variables

- A distinction can be drawn between naturally and artificially discrete variables (e.g., the male/female dichotomy of sex is natural, while the young/old dichotomy of age is artificial)

- It is generally **not** a good idea to create artificially discrete variables because it decreases the variance of the variable (i.e., range restriction). Furthermore, the conceptual basis for the distinction and differences between the various categorical levels can be very tenuous at best (e.g., 39 is young; but 40 is old [ADEA]; 90% is an A; but 89% and 80% are both Bs [typical course grading scheme]).

Editorial: A death to dichotomizing.

Gavan J. Fitzsimons

Journal of Consumer Research. Vol 35(1), Jun 2008, pp. 5-8

The Journal of Consumer Research receives manuscripts on an almost daily basis in which researchers have dichotomized a continuous independent variable. From the *Journal of Consumer Research*'s perspective, the relatively small investment in appropriately analyzing and presenting data involving a continuous independent variable is certainly justified compared to the cost of not doing so. I hope this editorial illustrates how easy it can be to present analyses that are performed appropriately. I hope that this editorial will help hasten the death to dichotomizing continuous independent variables—its day, I hope, is behind us.

So, again recognizing that it is generally bad practice, if one *has* to, then commonly used methods to generate artificially discrete variables (listed from the worst to the best) are:

1. mean split
2. median split
3. **extreme groups**

- Continuous and discrete distinctions are important because this information usually influences the choice of statistical procedures or tests.

  (a) **Pearson's correlation**—assumes that both variables are continuous.

  (b) **Point-biserial**—most appropriate when one variable is measured in the form of a true dichotomy, and we cannot assume a normal distribution.

  (c) **Biserial**—most appropriate when one variable is measured in the form of an artificial dichotomy, and we can assume a normal distribution.

(d) **Phi coefficient (Φ)**—is used when both variables are measured as dichotomies.

3. **Quantitative and Qualitative variables**

   A. A **quantitative** variable is one that varies in amount (e.g., reaction time or speed of response).

   B. A **qualitatve** variable is one that varies in kind (e.g., college major or sex).

      ▪ Qualitative variables are by definition also categorical.

➡ Extraneous variables—any variable other than the IV that influences the DV.

➡ Confounded variables—when an extraneous variable systematically varies with variations or levels of the IV.

MEASUREMENT

"Measuring things better advances everything that we do." Cecil Reynolds (April 2018).

● **Definition of measurement**—the assignment of numbers to events or objects according to rules that permit important properties of the events or objects to be represented by properties of the number system.

● Measurement is closely associated with the concept of operational definitions.

● To scientifically study anything, we must be able to measure it.

"I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; It may be the beginning of knowledge, but you have scarcely, in your thoughts advanced to the stage of science, whatever the matter may be" **Lord Kelvin** (19th century physicist)	"We mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations" **Percy Bridgman**, (the "father" of operationism), 1927, The logic of modern physics, p. 5).

● To scientifically study anything, we must first operationally define the construct of interest, and then use measurement rules to quantify it. This permits us to study the construct. The key is that properties of the events are represented by properties of the number system.

Levels of measurement

A *scale* is a measuring device to assess a person's score or status on a variable.

The five basic types of scales (levels of measurement) are:

1. **Labels**

   When numbers are used as a way of keeping track of things without any suggestion that the numbers can be subjected to mathematical analyses.

   - Examples include participant ID, university identity number (UIN), and social security numbers.

2. **Nominal scale**

   Grouping objects or people without any specified quantitative relationships among the categories.

   - Examples include coding all men as 1; and women as 2. Or cats as 1 and dogs as 2.

3. **Ordinal scale**

   People or objects are ordered from "most" to "least" with respect to an attribute.

   There is no indication of "how much" in an absolute sense, any of the objects possess the attribute.

   There is no indication of how far apart the objects are with respect to the attribute.

   Rank ordering is basic to all higher forms of measurement and conveys only meager information.

   - Examples include college football pools, top 5 contestants in a beauty pageant.

4. **Interval scale**

   Most common level of measurement in psychology.

   Measures how much of a variable or attribute is present.

   Rank order of persons or objects is known with respect to an attribute.

   How far apart the persons or objects are from one another with respect to the attribute is known (i.e., intervals between persons or objects is known).

   Provides information about the magnitude of the attribute for any object or person.

   - Examples include how well you like this course, where 1 = do not like at all, and 5 = like very much.

5. **Ratio scale**

   Has properties of preceding 4 levels of measurement in addition to a true zero-point.

   Rank order of persons or objects is known in respect to an attribute.

   How far apart the persons or objects are from one another with respect to the attribute is known (i.e., intervals between persons or objects is known).

   The distance from a true zero-point (or rational zero) is known for at least one of the objects or persons.

   - Examples include speed (no motion).

Evaluation of Measurement Methods and Instruments

- The extent to which data obtained from a measurement method fit a mathematical model.

   A. **Reliability**—presence of/susceptibility to measurement error. To the extent that the construct is stable, then would expect consistency over time, place, occasion, etc.

   B. **Validity**—extent to which a method measures what it is supposed to measure.

Reliability

- **Reliability**—refers to the presence of measurement error. To the extent that the measured construct is stable, then could also speak to the consistency of scores obtained by the same person when examined with the same test (or equivalent forms) on different occasions, times, places, etc.

- For a measurement system or method to be of any use in science, its scores must be both reliable and valid.

- Reliability, like validity, is based on correlations.

- **Correlation** (reliability [r_{xx}] and validity [r_{xy}]) **coefficients** [r_{xy}] can be computed by the formula:

$$\frac{N(\Sigma XY)-(\Sigma X)(\Sigma Y)}{\sqrt{\left[(N)(\Sigma X^2)-(\Sigma X)^2\right]\left[(N)(\Sigma Y^2)-(\Sigma Y)^2\right]}}$$

- Correlation coefficients measure the degree of relationship or association between two variables.

● Correlation coefficients can assume values of -1.00 to +1.00. The closer this value is to either of these limits, the stronger the relationship between the two variables. And the sign denotes whether the relationship is negative or positive.

Methods for Assessing The Reliability of a Test's Scores

All things being equal, the more items a test has, the more reliable its scores will be.

1. **Test-retest reliability** (temporal consistency or stability)—Involves the repeated administration of the same test to the same sample.

2. **Alternate-form reliability** (temporal consistency or stability, and inter-form consistency or equivalence)—A measure of the extent to which 2 separate forms of the same test are equivalent.

3. **Split-half, odd-even (or random split) reliability** (internal consistency)—The primary issue here is one of obtaining comparable halves.

4. **Coefficient alpha (Cronbach's alpha)** (inter-item consistency)—This is a measure of inter-item consistency (i.e., the consistency of responses to all items on the test).

   ▪ This is an indication of the extent to which each item on the test measures the same thing as every other item on the test.

   ▪ The more homogeneous the domain (test), the higher the inter-item consistency.

5. **Scorer reliability or inter-rater reliability and agreement**—the extent to which 2 or more raters are consistent, or agree.

   ▪ reliability = rank order

   ▪ agreement = magnitude (**and** rank order)

Test and Measurement Validity

● The **validity** of a test's scores concerns WHAT it measures and HOW WELL it does so.

● It tells us what can be inferred from test scores (i.e., the inferences that are made from a test's scores).

● The validity of a test cannot be reported in general terms.

● Validity depends on the USE of the test; no test can be said to have "high" or "low" validity in the abstract.

● Test validity must be established with reference to the particular use for which the test is being considered (i.e., the appropriateness of inferences drawn from data).

● For example, the SAT may be valid for predicting performance in college but will it validly predict aggressive behavior?

● Validity is a key—maybe *the* key criterion in the evaluation of a test or measure. The validity of a test or measure is the extent to which inferences drawn from the test scores are appropriate.

Strategies for Assessing Test Score Validity (i.e., Validation Techniques or Strategies)

Several, but for purposes of this course will limit to only the following:

▪ Criterion-related
▪ Content-related
▪ Construct-related
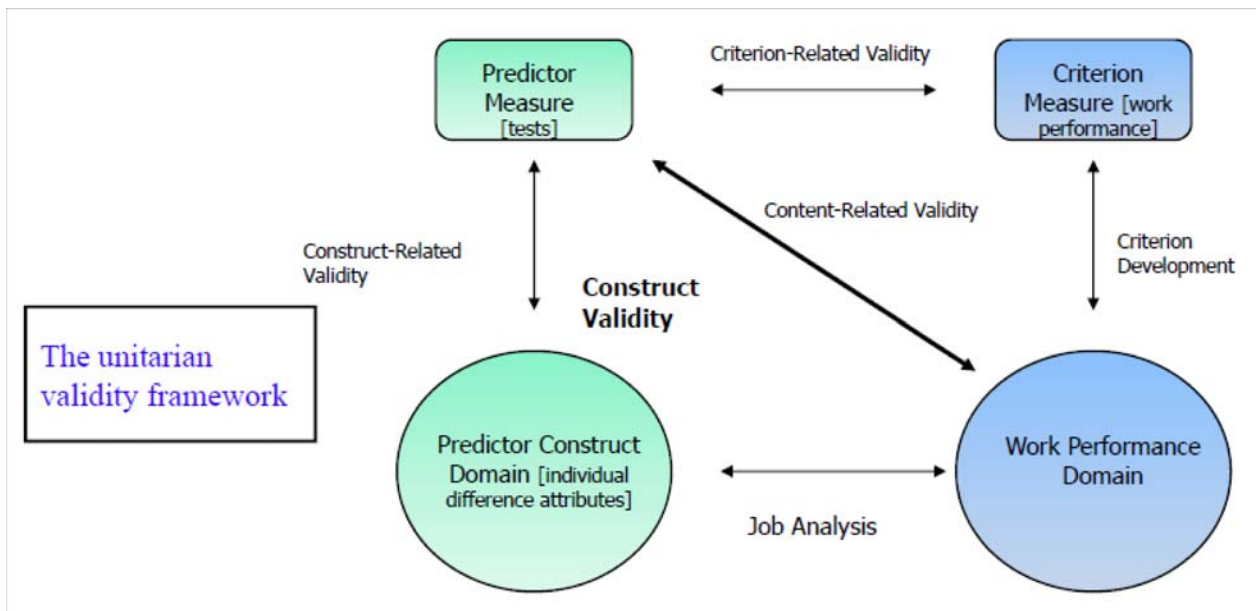
1. **Criterion-related validity**—effectiveness of a test in predicting an individual's behavior in specific situations.

   • That is, the test or measure is intended as an indicator or predictor of some other behavior (that typically will not be observed until some future date).

   • With criterion-related procedures, performance on the test, predictor, or measure is correlated with a criterion (i.e., a direct and independent measure of that which the test is designed to predict).

   • As mentioned earlier, validity is assessed using a correlation coefficient. As such, validity coefficients can range from -1.0 to +1.0. The absolute value is used to compare different validity coefficients in terms of magnitude.

• **Specific criterion-related validation designs**

   (a) concurrent
   (b) predictive
   (c) postdictive

   ▪ Differences between these criterion-related validation designs have to do with differences in time-frames in the collection of criterion and predictor data.

2. **Content-related validity**—For some tests and measures, validity depends primarily on the adequacy with which a specified content domain is sampled.

   • Content-related validity involves the degree to which a predictor covers a *representative sample* of the behavior being assessed (e.g., classroom tests).

   • Content-related validity involves a systematic examination of test content to determine whether it covers a representative sample of the behavior domain being measured.

   • Content-related validity is typically rational and nonempirical, in contrast to criterion-related validity which is empirical.

   • The content domain to be tested should be fully described in advance in very specific terms.

3. **Construct-related validity**—The construct-related validity of a test or measure is the extent to which the test may be said to measure a theoretical construct or trait.

   • A **construct** is a label for a theoretical dimension on which people are thought to differ.

   • A construct represents a hypothesis (usually only half-formed) that a variety of behaviors will correlate with one another in studies of individual differences or will be similarly affected by experimental treatments.

   • **Sources of construct-related validity evidence**

      (a) **Convergent validity**—different measures of the same construct should be correlated or related to each other.

(b) **Discriminant validity**—different measures of different constructs should **not** be correlated or related to each other.

● The **MULTI-TRAIT/MULTI-METHOD MATRIX** has historically been the most common approach to assessing the construct-related validity of a test or measure's scores.

  ▪ In the example below, the Bio-Feedback measure is the measure being validated.

METHOD

		Paper-and-pencil	Physiological
T **R** **A** **I**	**Anxiety**	MMPI (Scale 7) A	Bio-Feedback (to measure anxiety) B
T	**Verbal ability**	C Wesman Personnel Classification Test (verbal subscale)	D Physiological indicators of brain neural information processing efficiency (EEG—brain wave patterns) as measure of verbal ability

A and B = converge (convergent validity)
A and C = diverge (discriminant validity)
C and D = converge (convergent validity)
B and D = diverge (discriminant validity)

4. **Face validity**—Face validity has to do with the extent to which a test or measure *looks* like it measures what it is supposed to. Face validity is relevant because in can influence test takers' reactions and attitudes towards the test (e.g., test-taking motivation) and ultimately, their test performance.

SUMMARY

- A test's scores can be reliable but not valid.

- However, a test's scores **cannot** be valid but not reliable—a test that does not correlate with itself cannot be expected to correlate with anything else.

- Thus, reliability is a necessary but not sufficient condition for validity.

- A test with unknown reliability and validity is to be avoided.

- Finally, reliability and validity are properties or characteristics of test **scores** and not inherent, imbued properties of tests or measures.

Statistical (Empirical) Analyses of Data

Statistical tests are means, tools, or procedures that are used to:

(a) describe data and
(b) analyze relationships between variables (i.e., make inferences).

Statistical Tests and Procedures					
Descriptive Statistics				Inferential Statistics	
Measures of Central Tendency	**Measures of Variability (Dispersion)**	**Frequency Distributions**	**Correlation (Association)**	**Parametric Statistical Procedures**	**Nonparametric Statistical Procedures**
mean mode median	variance standard deviation range min max	normal skew kurtosis	strength direction	correlations *t*-tests (independent & dependent) ANOVA (*F* tests) regression	chi-square

Chapters 14 and 15 are a good review source for material and concepts that were covered in PSYC 301. Consequently, you are very strongly encouraged to continuously review this material and if there is anything contained therein that is not clear to you, please do not hesitate to ask me to explain it in class.