

POLI30: Inference

Week 2: Variables & Measurement

Scott Desposato D. Alex Hughes

September 29, 2011



Contact Information

D. Alex Hughes

Office: SSB 349

Email: dhughes@ucsd.edu

TA Website: <http://polisci2.ucsd.edu/dhughes>

Office Hours: Tentatively Thursday 12:00pm - 1:50pm

Clarifying a Concept

To be valid, or even useful, a concept must be two things:

- ▶ **Concrete** – Must be built of things that we can put our analytic “hands” on.
- ▶ **Variable** – Do some people have more of the concept, and others less?

Should we define one concept in terms of another concept?

Conceptual Definition

- ▶ After identifying the concrete, measurable attributes, we need to **define** how, and for whom those attributes describe a concept.
- ▶ The concept of [concept's label] is defined as the extent to which [the unit of analysis] exhibit the characteristic of [concept's measurement]

Operational Definition

- ▶ Translation of an idea or a construct into something real, concrete, and measurable
- ▶ It describes explicitly how the concept is to be measured empirically
- ▶ A theory of measurement

Measurement Error

- ▶ Distortions of the linkage between a concept and its empirical measure
- ▶ Two types of measurement error:
 - ▶ **Systematic Measurement Error** – consistent, chronic distortion of an empirical measurement
 - ▶ **Random Measurement Error** – distortion caused by any factors that randomly affect measurement of the variable across the sample

Reliability and Validity

- ▶ **Reliability** – the extent to which it is a consistent measure of a concept
- ▶ **Validity** – the extent to which it records the true value of the intended characteristic and does not measure any unintended characteristics

Variables

- ▶ “A **Variable** is an empirical measurement of a characteristic.” (Tetlock, p. 26).
- ▶ More precisely, a variable is an object that holds the empirical measurements of a characteristic.

So, what are some variables?

- ▶ Age of Students in Class
- ▶ Greek Society Membership
- ▶ Orange-ness of Snookie

Levels of Measurement

There are four levels of measurement.

- ▶ Nominal
- ▶ Ordinal
- ▶ Interval
- ▶ Ratio

Nominal Variables

- ▶ Communicate Differences between Units Being Compared
- ▶ No Order to the Variables
- ▶ “Binning”

Examples of Nominal Variables

- ▶ Fruit
- ▶ Marriage Status
- ▶ Greek Status
- ▶ Sexual Orientation
- ▶ Gender in Mad Men

Ordinal Variables

- ▶ Communicate Differences between Units Being Compared
- ▶ Order to the Comparison – One Bin is Larger than another Bin
- ▶ However, either the size of the differences between bins is unknown, or it varies.

Examples of Ordinal Variables

- ▶ Nutrition of Fast Food
- ▶ Number of Towns North of La Jolla
- ▶ Skeezyiness of Dude at Bar
- ▶ Hella, Hecka, Grip
- ▶ Job Title in Mad Men

Interval Variables

- ▶ Communicate Differences between Units Being Compared
- ▶ Order to the Comparison
- ▶ Size of difference between Units is **known & constant**

Examples of Interval Variables

- ▶ Grade in School
- ▶ Happiness Index
- ▶ Score on Election Fraud Index
- ▶ ? in Mad Men

Ratio Variables

- ▶ Interval Variables, but with a *meaningful Zero*
- ▶ “Does it make sense to compare this to Zero?”
- ▶ Could I divide this by a number or another variable, and would it still make sense?

Examples of Ratio Level Variables

- ▶ Age
- ▶ Miles from School
- ▶ Number of Whiskeys in Mad Men

Central Tendency

There are three measures of central tendency we will use. Each have different strengths and weaknesses, and summarize different amounts of data.

- ▶ Mode – What is the most frequently occurring variable level?
- ▶ Median – When arranging a variable according to its levels, what value falls in the middle?
- ▶ Mean – The average. When summing the all the values of a variable, and dividing by the number of entries, what is the value?

Which to Use?

The level of measurement dictates which measure of central tendency you may use

- ▶ **Nominal** – Mode
- ▶ **Ordinal** – Mode, Median
- ▶ **Interval** – Mode, Median, Mean
- ▶ **Ratio** – Median, Mean

Dispersion

- ▶ How *spread out* around the central moment is the data?
- ▶ Does it seem to be evenly spread out around the central moment?
- ▶ **Positive Skew** – The distribution has a longer/skinnier right-hand tail.
- ▶ **Negative Skew** – The distribution has a longer/skinnier left-hand tail.