Basic Statistics

Chapter One
Introduction

1

Agenda

- What is statistics? History
- Broad categories
- Introductory material – including lots of new terms
- Levels of measurement
- Types of Variables
- SPSS software - opening files and running basic procedures

2

Definitions (2):

- The mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling.
- Numerical data.
- History
  - state numbers
  - governments generate volumes of statistical data

3

Two broad categories of statistics

- Descriptive and inferential
- Descriptive: summarizes or describes large data set
- The mean is a descriptive statistic
- Graphical techniques

4

Two categories (cont.)

- Inferential: working with a sample we introduce some unknown amount of error due to the effects of chance
- Inferential statistics allow conclusions about a population based on data from a sample
- find probability that a difference we have observed between two groups would occur if there was no difference in the population

5

Overview of process

- **1. phenomenon, event, process want to understand better**
- **2. translate into numbers somehow (operationalize)**
- **3. run statistics on the numbers**
- **4. translate numbers back into the phenomenon**
- **5. write the story of the phenomenon**

6

Why is this important?

- researcher must understand
- important to be able to understand and appropriately question the data we are confronted with in our day to day lives
- real world examples
- stronger ability to understand the writing in scientific journals
- able to communicate with others who use statistics
- able to persuade others with evidence

7

A new language

- Population: all members of a certain group
- Sample: subset of elements from a population
- Research is almost always aimed at answering questions about a population
- population too large to measure in entirety: use sample and quantify precision of our estimates
- random sampling and random assignment to groups important to control for confounding and extraneous variables

8

Samples and writing

- samples usually described in great detail
- population being generalized to often is not
- generalize results from this group?

9

More terms

- parameter: a characteristic (numerical or nominal) of a population
- statistic: a numerical or nominal characteristic of a sample
- parameters don't change
- parameter values can generally not be known
- sample statistics are best estimator of the population parameter

10

Basic concepts (cont.)

- if we draw multiple samples from some population, statistics likely to differ slightly
- Variable is a quality or concept that exists in more than one amount or form
- Quantitative variables: the number that is assigned provides information about the amount or magnitude of the variable

11

Basic concepts (cont.)

- larger scores indicate "more" than smaller scores
- upper and lower limits for each value, for height in inches, weight in pounds, and IQ scores are generally reported in whole numbers
- individuals who obtain these scores may not be exactly the same
- could measure more precisely, but the same idea would apply for figuring out the limits
- add and subtract half the unit of measurement from the last digit of a score

12

Basics (cont.)

- Qualitative variables are not continuous: gender, political party are not simply different in terms of the amount of some "thing"
- Qualitative variables can have order
- Research study design
- Independent, dependent, and extraneous variables

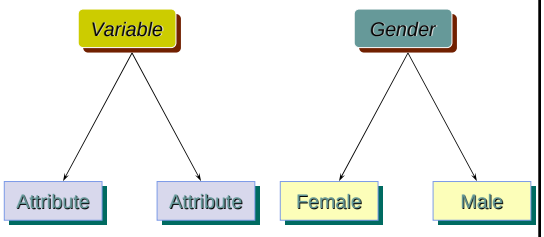
13

The Levels of Measurement

- Nominal
- Ordinal
- Interval
- Ratio

14

Some Definitions



15

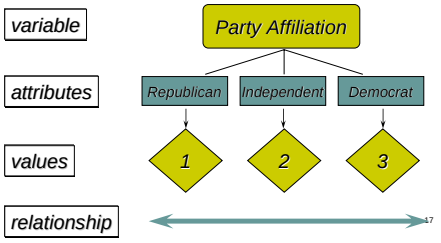
Qualities of Variables

- exhaustive - should include all possible answerable responses
- mutually exclusive - no respondent should be able to have two attributes simultaneously (e.g., employed vs. unemployed - it is possible to be both if looking for a second job while employed)

16

What is Level of Measurement?

the relationship between the values that are assigned to the attributes for a variable



17

Why is Level of Measurement Important?

- helps you decide what statistical analysis is appropriate on the values that were assigned
- helps you decide how to interpret the data from that variable

18

Nominal Measurement

- the values just "name" the attribute uniquely
- does not imply any ordering of the cases
- for example, jersey numbers in football
- even though player 32 has a higher number than player 19, you can't say from the data that he's greater than or more than the other

19

Other Nominal Measures

- gender
- colors
- makes/models of cars

20

Ordinal Measurement

- when attributes can be rank-ordered
- distances between attributes do not have any meaning
- for example, code Educational Attainment as 0=less than H.S.; 1=some H.S.; 2=H.S. degree; 3=some college; 4=college degree; 5=post college

- *is distance from 0 to 1 same as 3 to 4?*

21

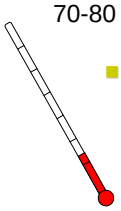
Other Ordinal Measures

- finishing place in a race
- likert-type scale items (Strongly Agree-Strongly Disagree)
- class rank

22

Interval Measurement

- when distance between attributes has meaning
- for example, temperature (in Fahrenheit) - distance from 30-40 is same as distance from 70-80
  - *but note that ratios don't make any sense - 80 degrees is not twice as hot as 40 degrees (although the attribute values are)*



23

Other Interval Measures

- SAT/GRE scores
- IQ
- scale scores from likert-type items

24

Ratio Measurement

- has an absolute zero that is meaningful
- can construct a meaningful ratio (fraction)
- for example,number of clients in past six months
- it is meaningful to say that "...we had twice as many clients in this period as we did in the previous six months

25

Other Ratio Measures

- kelvin temperature
- income in $
- time in a race

26

The Hierarchy of Levels