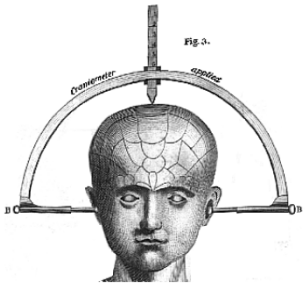
Statistical Methods in Behavioral Science~1



Introductory Matter

February 10, 2012 Edition

$$S=\sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}$$

Introduction~1

- *Statistics*, to most people, simply means NUMBERS…
  - …any numbers…
- …numbers such as unemployment figures, grades, batting averages, disaster casualties, *etc.*
- In this sense the word is being used to refer to either
  - raw numerical data or
  - data that has been "processed" in some way.

Introduction~2

- But, the word *statistics* also refers to strict procedures …
  - mathematical,
  - tabular, or
  - graphic...
    - …designed to analyze raw unprocessed data in order to get it to reveal hidden information not apparent in its raw form.
- The tools of the statistician range from simple procedures for finding *averages* to more complex ones to find relationships, such as *correlation*.

Introduction~3

- We are confronted with statistics constantly in our daily lives.
  - Some are boring,
    - *e.g.*, the volume of codfish caught in RI in a given year;
  - Some can be disturbing,
    - e.g., the number dead or left homeless from a natural disaster;
  - While some may be fascinating and annoying at the same time,
    - *e.g.*, the endless estimates of Bill Gates' wealth.

That's $72,000,000,000.00

Introduction~4

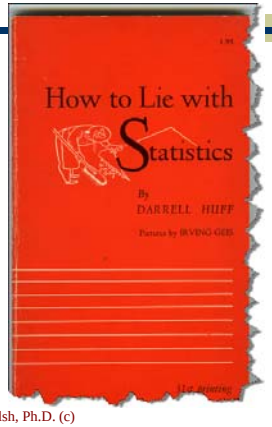
- When the statistics have personal meaning to the consumer , however, they take on special importance.
- For example, people take a very personal interest in statistics such as…
  - their own GPA, or
  - the number of pounds lost per week on a new diet, or
  - the gains and losses of their retirement account.

Introduction~5

- Benjamin Disraeli (1804-1881), former British *Prime Minister*, once wrote that "*There are three kinds of lies:*
  - *lies,*
  - *damned lies, and*
  - *statistics*."
- Although humorous, there is a large "grain of truth" in what he said!
  - …statistics can be manipulated for personal purposes to mislead; and, yes, sometimes not tell the whole story or the "truth."
- As an intelligent consumer of statistics, therefore, you must think critically about what you hear on television, or read in the paper and
  - *…consider the source* as well as…
  - *…the form of analysis used…*
  - *…and what you may NOT have been told…*
  - …before accepting them.

Introduction~6

- Entire books have been written, in fact, exposing just how statistics can lie.
- One of the most famous is Huff's humorous *How to Lie with Statistics* …still very much in print.



Introduction~7

- After the Bush vs. Gore presidential election years ago an editorial in the *New York Times* stated:
  - "*President-elect Bush is going to have an uphill battle when in office since he will be obligated to seriously address those issues of concern to the millions of voters who gave Al Gore the 500,000 vote margin and ultimate win in the popular vote.*"
  - What's wrong with this statement?

Populations vs. Samples

- The terms POPULATION and SAMPLE will be shown to have special relevance in several statistical procedures we will examine this semester.
- This will be the case especially in the context of the distinction made between DESCRIPTIVE and INFERENTIAL statistics.

Population

- POPULATION ~ this term refers to the entire collection or set of objects, people, events, *etc.*, of interest in a particular context, or…
  - …the set of measurements on the members of a population, as in the term "*population data*."
- For example, we can speak of the *population* of people who reside in the U.S., or
  - …the *population* of IQ Scores for those people.

Sample

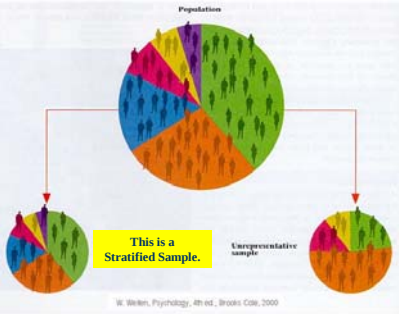
- SAMPLE ~ this term refers to a subset of objects, people, observations, *etc.*, selected from a *population*.
- By studying samples, researchers hope to obtain valid data about the populations from which they were drawn.
- One of two common procedures are used to produce either a …
  - …random sample; or, in special cases,
  - a stratified sample.

Population/Sample Graphic ~ 1

Working with samples is necessary since populations are often way too large to be studied or surveyed in their entirety.

Researchers must be cautious, however, to insure that the samples correctly represent the populations from which they were drawn.

Population/Sample Graphic ~ 2



Stratified Samples insure Representativeness.

President Harry S. Truman(1884 -1972) after winning the 1948 Presidential election. Notice that the Chicago paper got it wrong BIG TIME!

Descriptive Statistics

- DESCRIPTIVE STATISTICS ~ the area of statistics concerned with…
  - …organizing and
  - …summarizing information
    - …about a collection of actual observations, *e.g.*, the arithmetic mean or the standard deviation.
- They employ…
  - graphical,
  - tabular, and
  - numerical
- …tools to look for patterns, to summarize, and to extract information from *population* or *sample* data.

Inferential Statistics

- This is the area of statistics concerned with generalizing beyond actual observations.
- It employs *sample data only* to produce estimates, to …
  - …draw inferences, make predictions, draw conclusions, or make other generalizations…
  - …about the *populations* from which the *sample data* was drawn.
- Since it involves speculation, *there is always a risk that the inferences may be inaccurate*.

Parametric vs. Nonparametric Statistics

- PARAMETRIC STATISTIC ~
  - This is a statistic performed on a *sample* in order to infer the comparable value in a *population*.
  - Certain "restrictions" about the population from which the sample was drawn limit the circumstances under which this statistic can be used.
- NON-PARAMETRIC STATISTIC ~
  - This is a statistic which does not specify conditions about the *population parameters* from which the sample was drawn.

Levels of Measurement~1

- There are four *levels of measurement*, sometimes called "*scales*," of interest to statisticians. They are named:
  - *nominal*
  - *ordinal*
  - *interval* &
  - *ratio levels or scales.*
- A statistician may speak of these levels in different ways such as:
  - *The data I obtained in that survey is mostly nominal.* Or,
  - *My dependent variable data is at the interval level.* Or,
  - *To calculate the Spearman's rho statistic, you need ordinal level data.*

Levels of Measurement~2

- The four levels range in complexity from the least sophisticated
  - *nominal scale* which possesses only one property
  - to the most sophisticated *ratio scale* which possesses four properties.
- Progressively more complex levels possess all of the characteristics of lesser levels plus one new property.

Nominal Measurement~1

- NOMINAL DATA ~ nominal data is created by assigning observations to *categories* or *classifications* and then counting the frequency of occurrence within those *categories* or *classifications*.
- It is sometimes, therefore, simply called, "frequency data."
- Observations here reflect differences in kind.

Nominal Measurement~2

- This is the least sophisticated of the measurement scales but this type of data is often obtained in psychology.
  - This data type is also said to be QUALITATIVE.
- Examples include data such as those classifying:
  - people as male or female;
  - psychologists as behaviorists, humanists, or psychoanalysts;
  - voters as republicans, democrats, or independents;
  - racial groups as Chicano, Asian-American, Caucasian, or African American; or,
  - recording the frequency of occurrence of a variety of diseases in a particular population.

Ordinal Measurement~1

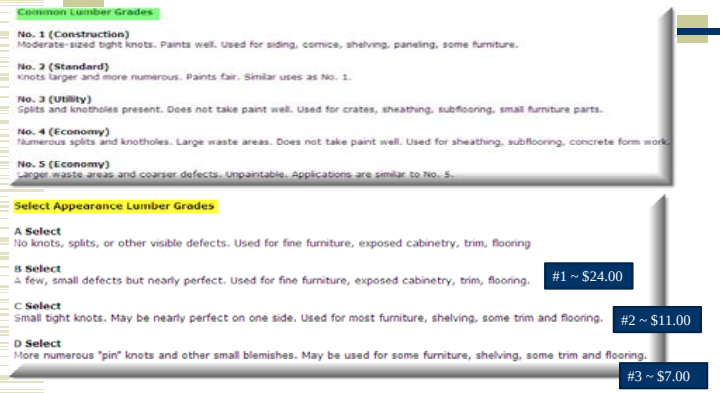
- ORDINAL DATA ~ in ordinal data, not only do we know that X or Y is present, we also know which one is greater than or less than another;
- *i.e.*, we know how they are *ranked*…from that observed measure representing the *most* of what we are assessing to that showing the *least* amount of it.
- The distinctive property of *ordinal data* is RANK order.
- The distances between the points on an ordinal scale are not known or needed.

Ordinal Measurement~2a

- The distances between the points on an ordinal scale are not needed since observations on this scale reflect differences in degree only. This data is also QUALITATIVE.
- Examples include:
  - academic letter grades;
  - "Best Colleges" ratings;
  - graduating class position;
  - military title; and,
  - the grading systems used to rate the hardness of minerals or the "appearance or utility grade" of lumber.

All of these have the common feature of being expressed as a kind of RANK.

Ordinal Measurement~2b



Interval Measurement~1

- INTERVAL DATA ~ with *interval data* we not only have the properties of *classification* & *order*…
  - …but we have the added knowledge of how much greater or less one observation is when compared to another in the ordered array.
- Our observations reflect measurable differences in amount and the distances between successive points on the scale are equal.
- But, zero on an interval scale is arbitrary, and ratios between observations cannot be calculated.
  - Interval data can not be multiplied or divided.

Interval Measurement~2

- *Interval data* is also common in the behavioral & other sciences. The data is truly QUANTITATIVE.
- Examples include:
  - IQ scores
  - grade point averages
  - SAT scores
  - Temperature ~ Fahrenheit vs. Celsius

Also, **0** does not mean no temperature!

But, while the difference between a temperature of 100 degrees and 90 degrees is the same difference as between 90 degrees and 80 degrees, we can not say that a temperature of 80 degrees is *2 times as hot as a temperature of 40 degrees because of the arbitrary zero.*

Ratio Measurement~1

- RATIO DATA ~ in *ratio data* we have the highest form of measurement.
- Often the product of familiar measuring devices
  - rulers, clocks, scales, meter sticks…
- …it permits us to not only interpret one observation as exceeding another by a certain amount as in *interval measurement*, but by a certain *ratio, e.g.*,
  - twice as much, four times more, *etc.*
- This data is absolutely QUANTITATIVE.

Ratio Measurement~2

- *Ratio data* not only possess all of the properties of the other measurement levels,
  - …it has the added advantage of an *absolute zero*.
- With a true *absolute zero*, meaningful statements about proportion can be made.
  - … as in A being half as large as B, and B being three times larger than C.
- Examples include height, weight, volume, length, density, pitch, reaction time, income, family size, *etc.*

Levels of Measurement Compared (Table B.1)

Table B.1
LEVELS OF MEASUREMENT

LEVEL	PROPERTIES	OBSERVATIONS REFLECT . . .	EXAMPLES	TYPE OF DATA
Ratio	true zero equal intervals order classification	measurable differences in *total* amount	weight income reaction time family size	quantitative
Interval	equal intervals order classification	measurable differences in amount	Fahrenheit temperature IQ score* grade point average* verbal aptitude score*	quantitative
Ordinal	order classification	differences in degree	graded attitude toward abortion developmental stages academic letter grade movie ratings	qualitative
Nominal	classification	differences in kind	sex gender ethnic background political affiliation major in college	qualitative

*Approximates interval measurement.

Copyright Notice

References

- Witte, Robert S. and John S. Witte. Statistics. New York: Harcourt, 9th edition, 2010.