# Measurement.ppt

## Measurement

## Measurement Theory

* Measurement theory is a branch of applied mathematics that is useful in measurement and data analysis. The fundamental idea of measurement theory is that measurements are not the same as the attribute being measured. Hence, if you want to draw conclusions about the attribute, you must take into account the nature of the correspondence between the attribute and the measurements.

## The map is not the territory

* How we measure something is not the same as the thing being measured.
* For some things the measure seems the same: weight=pounds.
* For others: not so much: Happiness= a high rating on the Happiness Scale.

## Classical Test Theory

* According to classical test theory, all measurement is imperfect.
* All measurement is comprised of a true score + error or Xij= + ij
* This means that any measure is comprised of the thing you are really trying to measure (theta) plus random garbage associated with that particular person at that particular time ( ij )

## So what is measurement?\*

* Measurement is the process of assigning numbers to represent some property, such as the quantity or quality of an object. Examples of measurement would include using inches to represent distance, IQ scores to represent intelligence, or integers to represent gender (e.g., 1 = male, 2 = female).

\* Aside from being imperfect.

## Why should I care about measurement?

* GIGO: Garbage in: Garbage out.
* The adequacy of measurement process, assessed in terms of the reliability and validity of the measure, ultimately impacts the conclusions that are drawn from the statistical analyses. The use of inadequate measures often leads to inappropriate or erroneous conclusions.

## Levels of Measurement

* Level of measurement provides information about the relationship between individual elements to which numbers are assigned.
* For example: Is element A greater than element B? Is the distance between element A and element B greater than the distance between elements B and elements C?
* Level of measurement frequently dictates the appropriate statistical methods for summarizing and/or analyzing the data.

## Level of Measurement:Nominal

* A nominal scale uses numbers to represent some qualitative class or category.
* Examples include using numbers to represent gender (e.g., 1 = male, 2 = female) or marital status (1 = never married, 2 = married, 3 = divorced, 4 = widowed, 5 = other).
* The numbers themselves have no inherent meaning; they connote no particular order or direction.

## Level of Measurement:Ordinal

* An ordinal scale uses numbers to reflect a rank ordering of elements.
* An ordinal scale reflects a relative ranking of elements on some property.
* The numbers in an ordinal scale do not express the true magnitude of the elements on this property.
* Ordinal numbers cannot be used to express the distance between elements, but merely their relative ranking.

## Level of Measurement:Ordinal

* House addresses are an example of an ordinal scale.
* Another common example of ordinal scales is the use of Likert-type items, such as the one below. While a -2 indicates stronger agreement than -1, the difference between -2 and -1 may not be the same as the distance between -1 and 0.
* Strongly Strongly
* Agree Agree Uncertain Disagree Disagree
* -2 -1 0 1 2

## Level of Measurement:Interval

* An interval scale uses numbers to reflect some quantity.
* As with the ordinal scale, the relative order of the numbers in an interval scale is important.
* In addition, however, the relative distance between elements in an interval scale is also important.
* Interval scales typically have a zero-point. The distinguishing feature of the interval scale is that the location of the zero-point is arbitrary and does not reflect the complete absence of some property.

## Level of Measurement:Interval

* Examples of interval scales include the Celsius and Fahrenheit temperature scales.
* Another example is an IQ score or scores on the MMPI.

## Level of Measurement:Ratio

* A ratio scale has all the properties of an interval scale plus a fixed and non-arbitrary zero-point reflecting the complete absence of some property.
* Examples: height, weight, temperature in Kelvin, age in years.
* All of these have a true zero.

# Approaches to research and Data concepts.ppt

## Scientific Method

* Not a method per se but a philosophical outlook characterized by:
* Empirical Reasoning.
* Experimental and non-experimental techniques.
* Methods of external observation and reporting.
* Quantitative and qualitative procedures.

## Research Approaches

* Descriptive.
* Relational.
* Experimental.

## Descriptive Research

* The first step in the research process.
* Mapping out the territory.
* Systematic, careful description of what is happening and where it happens.
* Doesn't really tell us why something is happening or how it happens.
* Example: Ethnographic research.

## Descriptive Research: Example Washington Post/ABC Poll 5/12/06

http://www.washingtonpost.com/wp-srv/politics/polls/postpoll\_nsa\_051206.htm

## Descriptive Research: Example:USA Today Poll 5/14/06

http://www.usatoday.com/news/washington/2006-05-14-nsa-reax-poll\_x.htm

## Systematic Observation

* Systematic observation differs from casual observation in that it uses a methodology that can be evaluated based on certain technical standards.
* Quantitative data recorded in numeric form.
* Qualitative- data recorded in any number of non-numeric ways: recorded conversation, videotape, writing.

## Types of Systematic Observation

* Participant observation.
* Ethnography.
* Secondary Observation and Content analysis.
* Experimental simulations.
* Unobtrusive observation.

## Relational Research

* Often this is the next step in research.
* When two or more things are measured, are they related to one another?
* Relational question: Is presidential approval related to ethnicity?
* Statistics in relational research are both descriptive and inferential (involves making a statement about the likelihood of our observation).
* Often leads to the formation of hypotheses.

## Experimental Research

* Focus of experimental research is on determining how or why something occurs.
* Experimental research often involves the systematic manipulation of conditions and careful measurement of outcomes to determine causes.

## Definitions

* Constructs.
* Variables.
* Independent variables.
* Dependent variables.
* Discrete and continuous data.
* Descriptive Statistics.
* Inferential Statistics.

## Construct

* Theoretical concepts formulated to serve as causal or descriptive explanations.
* Examples:
* Self Esteem
* Depression
* Introversion/extroversion.
* Diffusion of responsibility

## Variable

* A variable is an event or condition that the researcher observes or measures or plans to investigate that is likely to vary (or change).
* Dependent variables (DV) refer to the "effect" or "outcome" in which the researcher is interested.
* Independent variables (IV) refer to the presumed "cause" . Changes in this lead to changes in the dependent variable.
* Note that independent and dependent are very much defined by the research question.

## Examples: Name the IV and DV

* Exercise decreases depression.
* Girls who diet at an early age are more likely to develop eating disorders.
* Schizophrenics with tardive dyskenia are more likely to die than those without the disorder.
* CBT is more effective than non-directive supportive therapy in reducing anxiety symptoms in adolescents.

## Discrete and Continuous Data

* Discrete data have values that can only assume whole numbers.
* For example, children only come in whole units, there aren't really 2.2 children. Besides, have you ever tried to shop for a .2 child?
* Continuous data can take any value within a defined range.
* For example, height , weight and time are really continuous; we impose arbitrary measuring limits on them.

## Statistics: primary types

* Descriptive statistics are concerned with the presentation, organization, and summarization of data.
* Inferential statistics allow us to generalize from our sample of data to a larger group of subjects.

# PY 620 Class 2.ppt

## PY 620 Class 2

* Looking at data.
* Bar Charts and Histograms.
* Graphing data.
* Tables.
* Measures of Central Tendency.
* Why outliers are important.

## Looking at Data

* Bar Charts and Histograms
* Stem and Leaf plots.
* Frequency Polygrams.
* Tables.

## USA Today Snapshot

## Onion Statshot

## Some of our data: raw numbers

## Our data: in a table

## Our data: In a bar chart with counts

## Our data: In a bar chart with percentages

## A Histogram with a Normal Curve

## Stem and Leaf Plots

* Stem and Leaf plots combine elements of both a table and a graph.

The idea of taking statistics gives me intestinal distress Stem-and-Leaf Plot
Frequency Stem & Leaf
4.00 2 . 0000
4.00 3 . 0000
8.00 4 . 00000000
4.00 5 . 0000
Stem width: 1
Each leaf: 1 case(s)

## A box plot

## Box Plots

Developed by John Tukey as a way of easily visualizing data distributions.
Box encloses the middle 50% of data. "Fences" enclose 95% of data. Outliers are beyond these limits.

## More data in a bar chart

## Sometimes a pie chart works best

## Sometimes a pie chart doesn't makes sense at all

## Histogram of Years of Education

## Education as frequency polygram

## Multiple Frequency Polygrams

## Looking at Data: MNNG Age

## Looking Closely at Data: MNNG Age

## Measures of Central Tendency:Mean

* arithmetic average
* sum of the observations divided by the number of observations (e.g. SCARED1 72/20 = 3.60)
* balance point of the distribution > changing any single number changes the mean of the sample
* sum of the deviations about the mean is zero
* appropriate for ordinal, interval & ratio data
* independent of sample size

## Measures of Central Tendency:Median

* median is the middle-most point of the distribution when ordered from lowest to highest score.
* the point at or below which 50% of the cases fall
* order the cases
* for odd number cases, it is the middle most score (i.e., the score at [(N + 1)/2] (in this case 11)
* for even number scores, it is the score midway between N/2 and N/2+1
* less sensitive to extreme scores
* does not take into account all scores > changing one score may not change median
* appropriate for skewed distributions, e.g., reaction time, income
* appropriate for ordinal, interval & ratio data
* independent of sample size

## Measures of Central Tendency:Mode

* most frequently occurring score
* appropriate for any scale of measurement
* least sensitive to changes in scores
* distributions may be bimodal
* not informative if used for continuous measures with infinite number of values
* independent of sample size

## Measures of Dispersion:Range

* difference between the largest and smallest scores
* ease of computation and interpretation
* determined by only 2 scores (minimum & maximum)
* changing any other score will have no effect on range

## Skewness

* index of symmetry
* calculated as 3 times difference between mean & median divided by SD
* sign and magnitude indicate direction (positive & negative) and degree of asymmetry
* normal curve has skew of 0
* informative about relative location of mean, median & mode

Examples of Skew

## Interquartile Range

* Problems with Range
* can change wildly with outlyers.
* As sample grows, range is likely to grow.
* Interquartile range is the difference between the upper quarter and lower quarter of the data- the middle 50%.
* Less affected by extreme scores.

## Variance and Standard Deviation

* We need to measure the "spread" of data around the mean value. One way to do this is to take the difference between an individual value and the mean value and square it. Why square it? To get rid of negative values.
* In other words, take the difference from the mean, square it, and add them all up and divide by the total sample size.
* For example in the scared data, scared1 has a variance of 1.095

## Formula for Variance

This is the formula for calculating the variance of a population. The X represents an individual value, the x with a line over it represents the mean. To summarize, this formula tells you to subtract each individual value from the mean, square that difference, and then add them all up and divide by the size of your sample.

Since we almost always have a sample instead of a population, we use a slightly different formula. Instead of dividing by the number in the sample, you divide by that number, minus 1.

## Standard Deviation

* To make the nonsense number from variance make more sense, you use the following formula.
* This puts the whole thing back into original units. In the case of Scared1, 1.046.

# PY620 -Class 3.ppt

## PY620 -Class 3

* Normal distribution
* Central Limit Theorem
* Standard scores
* Probability: definitions
* Mutually exclusive
* Conditional probability
* Independent events
* Law of "at least one"
* Binomial distribution
* Exercises.
* Coin Jar Calculator.
* SPSS

## Inferential Statistics: a Definition Revisited.

* Inferential Statistics are used to determine the probability (or likelihood) that a conclusion based on analysis from a sample is true.
* All measurement contains random error.
* We need to take this random error into account when we are trying to say something about a population based on a sample.
* High School graduating class. Would a random sample of 10 students characterize your HS class as well as a random sample of 50? Maybe, but probably not.

## Difference between Samples and Populations.

* The sample describes those individuals who are in a study; the population describes the hypothetical (and usually) infinite number of people to whom you wish to generalize.
* Defining the population of interest is very important. Only if the population is well defined can you successfully sample from it.

## Normal distribution

* Also known as Gaussian distribution.
* Provides the basis for many inferential statistics.
* Theoretical distribution that is defined in terms of a mathematical function referred to as the normal probability density function.

## Normal Distribution: more facts

* It is a continuous distribution that contains values ranging from negative infinity to positive infinity.
* The normal distribution is bell-shaped, unimodal, and symmetrical.
* The mean, median, and mode all have the same value.

## Normal Distribution

* The percentage of observations within any given interval can be determined. Approximately 68% of the observations fall within one standard deviation above and below the mean, approximately 95% of the observations fall within two standard deviations above and below the mean, and over 99% of all observations fall within 3 standard deviations above and below the mean.

## Many naturally occurring distributions are normally distributed

* A number of naturally occurring distributions are approximately normally distributed.
* The distribution of certain human characteristics, such as height, weight, IQ, and serum cholesterol is approximately normally distributed.
* The distribution of errors, such as errors of measurement or discrimination.
* It should be pointed out that the normal distribution cannot be considered natures rule, since many naturally occurring distributions do not resemble a normal distribution.

## Good approximation for other distributions

* A second reason that the normal distribution is important is that it serves as a good approximation for a number of other theoretical distributions that are difficult or impossible to define precisely.
* For example, the binomial distribution, which is used to determine the likelihood of a given outcome over a series of successive trials, is frequently approximated by the normal distribution. This significantly reduces the calculations required to perform the test.

## Central limit theorem.

* A third reason that the normal distribution is important has to do with sampling and a principle known as the central limit theorem.
* Consider a case where a large number of samples of size N were drawn from the same finite population, the mean of each sample was calculated, and these means were used to form a new distribution of means (referred to as the sampling distribution of means).
* The central limit theorem states that, regardless of the form of the original distribution, the distribution of means will be approximately normal when N is large. This principle provides the foundation for interval estimation and a variety of inferential statistics including t-tests and analysis of variance.

## Law of Large Numbers

* Basically, the average of a sample is likely to be close to the average of the population from which it is drawn.

## NG- Age distribution

## Population vs. 1% sample

## Population vs. 5% sample

## Population vs. 10% sample

## Population vs. 20% sample

## Direct links to probability.

* A fourth reason that the normal distribution is important has to do with its direct links to probability.
* Recall that the percentage of observations can be determined in a normal curve for any given interval. Also, keep in mind that probability is defined in terms of the long-run relative frequency of occurrence of an event. Given these considerations, then what is the probability of randomly drawing an observation from a normal curve that is within 1 standard deviation above or below the mean? The answer is given in terms of the relative frequency of that event if we were to draw a random observation over a large number of trials. Since approximately 68% of all observations in a normal curve fall within this interval, then the probability of that event is approximately .68.

## Standard Scores

* Standard scores ,usually abbreviated as z or Z, is a way of expressing any raw score in terms of standard deviation units.

## Standard Scores: Class Height

## Z distribution

* You can check the z scores against the z distribution to see where they fall.
* For instance a z score of 1.39 means that 90% of the class are shorter than that person.

Reading the Z-distribution table.

For a Z value of 1.25, we can find the distribution by going to the 1.2 row and then reading down the column marked .05.

A value of .8925 means that 89.25% of the distribution has a lower score and only 10.75 percent of the distribution as a higher score.

## Probability

* Probability deals with the relative likelihood that a certain event will or will not occur, relative to some other events.
* Empirically derived (many trials).
* Theoretically derived (dice rolls, card hands, etc).

## Mutually exclusive events

* Two events , X and Y, are mutually exclusive if the occurrence of one precludes the occurrence of the other.
* Examples: Coin flip, vote.
* Additive rule: if X and Y are mutually exclusive events, then the probability of X or Y is the probability of X plus the probability of Y.

## Conditionally Dependent Events

* Two events, X and Y, are conditionally dependent if the outcome of Y depends on X, or X depends on Y.
* Multiplicative Law
* If X and Y are conditionally probable, then the probability that both will occur is the probability of X times the probability of Y, given X has occurred.

## Conditional probability: Loss of mother in childhood and traumatic brain injury

What is the probability of having a brain injury and having lost your mother during childhood?

## What is the probability of having a head injury?

In this sample (folks with substance abuse problems) it is 39.6%

## What is the probability of having a head injury if you lost your mother in childhood?

The probability of having lost your mother as a child if you have had a brain injury is 32.5%.

## Bringing it together

* Therefore, in this sample, the probability of having a head injury and having lost your mother as a child is:
* .396 (probability of head injury) times
* .325 (probability of head injury if mother was lost in childhood).
* =.129 of 12.9% or 66/513

## Independent Events

* Many events are independent . Each occurrence is independent of previous occurrences.
* Conditional events depend on the occurrence of a previous event.
* If nine coin flips are heads, what is the probability that the tenth will be heads?
* Of these types of gambling which are independent and which are conditional?
* Lottery
* Blackjack
* Roulette
* Texas Hold-em

## Law of at least one

* Every test has a false-positive rate. This chance of a false test can add up.
* If a test has a false positive rate of 5%, what is the likelihood that at least one test will be false after 20 tests.
* Formula= 1-Pr(none)number of tests. 1-.9520
* 1-.358=64.2%.
* More than 50% chance of positive test.
* This function in Excel tells us the likelihood of having NO false positives in 20 trials of a test with an accuracy rate of .95 or 95%.

## Binomial Distribution

* If continuous measures get the normal distribution, is there a distribution for binomial (yes/no) outcomes?
* The binomial distribution shows the probabilities of different outcomes for a series of random events, each which can have only one of two values.

## Binomial Distribution

* Applicable when:
* Series of independent trials (e.g., flips of coins) where the outcome of 1 trial does not influence the outcome of another
* Each trial can result in only 1 of 2 possible outcomes (e.g., heads or tails, six or not six)
* The probability of success is constant from trial to trial.

## Binomial expansion

* N = # trials
* r = # successes
* p = probability of success on a single trial
* q = probability of failure on a single trial
* Pr(r successes on N trials) =
* Note that p + q = 1 and that q = 1 - p

## Binomial Expansion:Not as scary as it looks.

* What is the likelihood of a professor getting to work on time exactly 7 times out of 10 if, on any given day, he has a 60% likelihood.
* Number of tries (n)=10.
* Number of favorable outcomes (r) =7.
* Probability on each try (p) =.6
* q=1-p (1-.6 =.4).

## We plug in all of the numbers.

10!=3628800

7! \* 3! =
5040\*6=30240

.67=0.0279936

.43=0.064

=21.5% Likelihood of said professor getting to work on time exactly 7 days out of 10.

* Software commands: Excel: =BINOMDIST (number\_s, trials, probability\_s, cumulative)
* Number\_s:
* number of successes in trials.
* Trials:
* number of independent trials.
* Probability\_s:
* probability of success on each trial.
* Cumulative:
* a logical value that determines the form of the function.
* TRUE, or 1, for the cumulativeP(X Number\_s)
* FALSE, or 0, for the probability function P(X = Number\_s).

## Binomial Distribution Properties

* Mean=np
* Variance=npq
* SD=
* Mean for last problem = 10 x .6 =6.
* Variance=10 x .6 x .4=2.4
* SD=sqrt 2.4=1.55
* As sample size gets larger, the binomial distribution approximates the normal distribution. As Martha says, "it's a good thing".

# Belmont Report July 2003.ppt

## IRB Member Discussion Module 2003 irb@umn.edu http://www.research.umn.edu/subjects/

The Belmont Report

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research
April 18, 1979

## History of Protection of Human Subjects

* Nuremberg Code (1947)
* Declaration of Helsinki (1964,75,83,89,2000)
* Beecher Ethics and Clinical Research (1966) [NEJM 274 (1966) 1354-60]
* Willowbrook (195601965)
* Jewish Chronic Disease (1963)
* USPHS Syphilis Study (aka Tuskegee Study) (1932-72)
* Radiation experiments (1940s - 60s)
* Modern day scandals.

## Response to Research Abuses

* National Research Act 1974
* May 1974 45 CFR 46
* June 1974- National Commission
* April 1979 Belmont Report
* 1981 45 CFR 46 revised

## National research act 1974

* The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (i.e., The National Commission)
* Initially met over a four-day period in 1976
* Met at the Smithsonian Institute's Belmont Conference Center (thus The Belmont Report)
* Monthly deliberations over the next four years

## Belmont Principles:

* Are general prescriptive judgments;
* (Other principles may also be relevant)
* they cannot always be applied so as to resolve beyond dispute particular ethical problems;
* their objective is to provide an analytical framework that will guide the resolution of ethical problems arising from research involving human subjects.

## Primary Focus

* Boundaries Between Practice and Research
* Basic Ethical Principles
* Applications (of Principles)

## The Belmont Report

* Boundaries Between Practice and Research
* IRB must determine that the researcher (and through informed consent, the subject) distinguishes practice from experiment in both social science and medical science research

## The Belmont Report

* Basic Ethical Principles:
* Respect for Persons
* Individual autonomy
* Protection of individuals with reduced autonomy
* Beneficence
* Maximize benefits and minimize harms
* Justice
* Equitable distribution of research costs and benefits

## Respect for Persons

* Treat individuals as autonomous agents
* Do not use people as means to an end
* Allow people to choose for themselves
* Give extra protection to those with limited autonomy

## Beneficence

* Acts of kindness or charity that go beyond duty
* Obligations derived from beneficence
* Do no harm
* Prevent harm
* Prevent evil
* Promote good

## Justice

* Treat people fairly
* Fair sharing of burdens and benefits of research
* Distinguish procedural justice from distributive justice

## The Belmont ReportApplication:

* Equal moral force
* and ,
* Ethical conflict expected

Principles have:

## IRB Members should consider

## Protocol Design: Respect for Persons

* How can the consent process maximize autonomy?
* How can the protocol maximize autonomy?
* What additional protections can be in place for protected populations?
* How can this study maximally protect subject privacy?

## Protocol Design: Beneficence

* Can the research design be improved?
* (to enhance safety and benefit)
* What are the risks? How can they be minimized?
* What are the benefits? How can they be maximized?

## Protocol Design: Justice

* How can you ensure that recruitment targets the population that will benefit from the research?
* How can you ensure that recruitment will not unfairly target a population?
* (avoid exploitation of population of convenience)
* How can the inclusion/exclusion criteria be made fair?

## The Belmont Report

* All principles are essential to sound ethical research (regulatory compliance)
* Principles carry equal moral weight
* Cannot approve research that violates or diminishes principles
* Some IRBs consider beneficence first in deliberations

## For Further information:

Email UMN RSPP Office: irb@umn.edu
UMN IRB Web Site:
http://www.research.umn.edu/subjects/
OHRP Web Site for Report:
http://ohrp.osophs.dhhs.gov/humansubjects/guidance/belmont.htm
July 2003

# PY620-Class 4 Sampling.ppt

## PY620-Class 4

* Inferential stats.
* Samples and Populations
* Sample statistics and Population parameters.
* Hypothesis testing.
* Standard deviation and standard error.
* Z-test
* Rationale behind significance testing.
* Signal to Noise.
* Errors, errors, and more errors.
* Type I error.
* Type II error.
* Power.
* Effect Size.
* Confidence Intervals.

## A Helpful Website on Research Methods

* Bill Trochim's Center for Social Research Methods website is a very helpful source of information on research methods.

## Inferential Statistics: a Definition Revisited.

* Inferential Statistics are used to determine the probability (or likelihood) that a conclusion based on analysis from a sample is true.
* All measurement contains random error.
* We need to take this random error into account when we are trying to say something about a population based on a sample.
* High School graduating class. Would a random sample of 10 students characterize your HS class as well as a random sample of 50? Maybe, but probably not.

## Difference between Samples and Populations.

* The sample describes those individuals who are in a study; the population describes the hypothetical (and usually) infinite number of people to whom you wish to generalize.
* Defining the population of interest is very important. Only if the population is well defined can you successfully sample from it.

## Selecting the Research Participants

* Typically use opportunity samples if goal is simply to learn about human nature.
* But what if you want to generalize results from the sample to a population?
* Need way of selecting a sample that is representative of the population.

## Why is sampling important?

This is why.

## What is sampling?

* The process of selecting units from a population so that by studying the sample we can generalize to that population.

## Sampling Terminology

* Population-in theory, the group to whom you want to generalize.
* Study population- the group that you can actually get access to.
* Sampling frame-how you access that population.
* Sample- those you select for your study (not necessarily the same as the people in your study).

## Sampling types

* Probability sampling.
* Simple random samples.
* Stratified random samples.
* Area probability sampling.
* Non Probability sampling.
* Convenience sampling.
* Purposive sampling.
* Snowball sampling.

## Simple Random Sampling

* Sample is selected from an undivided population.
* Sample is chosen using a process that gives every sampling unit an equal chance of being selected.
* Requires knowledge of the existence of all the units in the population.

## Examples of Random Sampling Strategies

* Throwing dice
* Using a table of random digits
* Spinning a roulette wheel
* Drawing capsules from an urn
* Using random digit dialing for telephone surveys.

## Random Sampling Options

* Sampling with replacement
* Selected names are placed back in the selection pool.
* Exactly the same odds of selection for each unit but a unit may be chosen more than once.
* Sampling without replacement
* Selected names are discarded and cannot be reselected.
* All selection can be done at once, but odds will change as population decreases with selection.

## Other Probability Sampling Plans

* Stratified random sampling
* Create homogeneous subpopulations based on a known characteristic
* Area probability sampling
* Divide the population into geographic areas

## Point Estimates and Confidence Intervals

* Point Estimate: Describes some particular characteristic of the population.
* Interval Estimate: Indicates how much the point estimates are likely to be in error.
* Confidence Interval: Probability that a point estimate is within plus-or-minus some specified interval.

Benefits of Stratification: If Simple Random Sampling Is Used

## Benefits of Stratification: If Stratified Random Sampling Is Used

## Nonresponse Bias

* Error due to nonparticipation or nonresponse.
* Possible consequences of this bias:
* Smaller effective sample size
* Inaccurate estimates of population values

## Non-Probability sampling

* Convenience. Whoever is convenient and available for your study. Any study which recruits volunteers. Most studies involving college students.
* Purposive. Seeks out particular types of people for the sample (e.g. adult female students over 40).Used in marketing surveys.
* Snowball. (does not actually involve snow).
* A type of purposive sampling that asks people in the sample to name other people like them that they know. Then those people are contacted and asked to name people they know, etc.

## Example of a Nonresponse Bias

## Characteristics of Typical Volunteer Subjects

* Better educated
* Higher social class
* Higher IQ scores
* Higher need for social approval
* More sociable
* More arousal-seeking
* More unconventional
* More often women
* Less authoritarian

## Excerpt: CBT paper.

* Participants
* "Participants were 78 clinicians who volunteered to participate in the trial and who provided written informed consent. The participants were required to be currently employed full-time as a clinician treating a predominantly substance-using population. Clinicians were recruited through newsletters and direct contact with clinics. A total of 100 clinicians were initially contacted, 2 were excluded because they were not currently treating substance users, and 20 elected not to participate."

## Sampling: CBT paper

* What was the population being sampled in the CBT paper?
* How did they sample?
* Was it random?
* How would you draw a random sample from the population of interest?

## What the authors said.

* Finally, these findings reflect effects of training on real world clinicians who volunteered to participate in an evaluation of training strategies, and it is not known whether these findings generalize to other, possibly less motivated, groups of clinicians or to other types of treatment.

## When is randomization used?

* Random selection is how you draw the sample of people for your study from a population. Random selection is related to sampling which is related to external validity.
* Random assignment is how you assign the sample that you draw to different groups or treatments in your study. Random assignment is related to internal validity.

# PY620-Class 4 standard error-z-tests-types of error.ppt

## PY620-Class 4

* Inferential stats.
* Samples and Populations
* Sample statistics and Population parameters.
* Hypothesis testing.
* Standard deviation and standard error.
* Z-test
* Rationale behind significance testing.
* Signal to Noise.
* Errors, errors, and more errors.
* Type I error.
* Type II error.
* Power.
* Effect Size.
* Confidence Intervals.

## Its all Greek to me: What those funny letters mean.

## Standard Error of the Mean

* A standard deviation is a measure of the dispersion (spread) of individual values around a sample mean.
* What if we want to know what the dispersion (spread) of sample means were around a population mean?
* If we took an infinite number of samples from a population, there would be a normal distribution of those sample means around the population mean.

## Red dot: mean value of individual sample

95% of all sample means will be within roughly 2 standard deviations (2\*s/n) of the population parameter m.
Because distances are symmetrical, this implies that the population parameter m must be within roughly 2 standard deviations from the sample average , in 95% of all samples.

This reasoning is the essence of statistical inference.

## Standard Error of the Mean and Standard Deviation

* The Standard Deviation (SD) reflects how close individual scores cluster around their mean, whereas the Standard Error of the Mean (SE) shows how close mean scores from repeated samples will be to the true (population) mean.

## Standard Error of the Mean (SEM): Formula

* The SE is related directly to the sample mean and the sample size.

## Standard Error of the Mean: an example

* What if we know that the sleep time for our class was 7.19 hours and that the SD of that mean is .973. What if we were interested in the mean hours of sleep for St. Mary University students (our population of interest).
* We can work from our sample standard deviation and estimate the SE for the population it came from.

## Standard Error of the Mean: an example

The Standard Unit of dispersion of sample means (Standard Error) around the population mean is .1004 hours.

## Now we have the SE, why did we bother?

* So we have the Standard Error- big deal! What's it good for?
* The SE lets us compare our mean to the population mean.
* Suppose that the "true" mean hours of sleep of the St. Mary's University students was 7.5. How would our class sleep time compare to the St. Mary's population sleep time?

## Enter the z-test

* Calculating z test will tell us where the sample lies on the normal distribution of population means.
* This tell us that our class mean lies 3.09 deviation units below the population mean.

## Z-scores and Z-tests

* I'm so confused? Z-scores, z-tests, what z hell is going on here?
* The z-score is a z-test for a single number. Just as the z-score tells how far an individual score is from the sample mean, the z-test tells us how far a sample mean is from the population mean.

## The Language of Hypothesis Testing: Stating the null hypothesis

* In the language of hypothesis testing, we always start by assuming no effect or no difference. The null (meaning nothing, nil, zippo) hypothesis for the comparison of our class sleep time to the mean sleep time of all St. Mary's students is:
* H0: There is no difference between the sleep time of class PY620C and the St. Mary University student population.

## The Language of Hypothesis Testing: Stating the alternative hypothesis

* H1: There is a difference between the sleep time of class PY620C and the St. Mary University student population.
* Strictly speaking there MUST be a difference between the mean sleep time of students in PY620C and the mean sleep time of St. Mary's students overall. It might be very tiny but there will be a difference.
* The question really is, how unlikely does it have to be that there is no difference before you say that there is.
* The null hypothesis is a very specific statement about a parameter of the population(s). It is labeled H0.
* The alternative hypothesis is a more general statement about a parameter of the population(s) that is exclusive of the null hypothesis. It is labeled Ha.
* Weight of cherry tomato packs:
* H0 : µ = 227 g (µ is the average weight of the population of packs)
* Ha : µ 227 g (µ is either larger or smaller)

## One Tailed and Two Tailed Tests

* A one-tailed test specifies the direction of the difference in advance.
* A two-tailed test is a test of any difference between groups, regardless of the direction of the difference.
* We usually use two-tailed tests unless a difference is only possible in one direction. We do this because if we posit that a difference is in one direction and it winds up being in the other direction, we have a dilemma. In essence we have evidence but it is inadmissible.

How to choose?

What determines the choice of a one-sided versus a two-sided test is what we know about the problem before we perform a test of statistical significance.
A health advocacy group tests whether the mean nicotine content of a brand of cigarettes is greater than the advertised value of 1.4 mg.

Here, the health advocacy group suspects that cigarette manufacturers sell cigarettes with a nicotine content higher than what they advertise in order to better addict consumers to their products and maintain revenues.
Thus, this is a one-sided test: H0 : µ = 1.4 mg Ha : µ > 1.4 mg

It is important to make that choice before performing the test or else you could make a choice of convenience or fall in circular logic.

## The Rationale of Null Hypothesis Testing

* In Null Hypothesis in Significance Testing (NHST) we make a bet. It could be that the difference in heights is simply by chance. This is the bet: If , given the SD and N for the sample, the difference observed can occur more than 5 times out of 100, we can't reject the null hypothesis, we can't conclude the difference is anything more than chance. If, on the other hand, the difference observed is unlikely to be observed more than 5 times out of 100, than we can reject the null hypothesis and say that the class is different from the overall school in height.

## What does rejecting the Null hypothesis tell us?

* Does it tell us how big the difference was? NO.
* Does it tell us how rare the sample mean would be in the population if there was no difference? Strictly speaking NO. The probability level that we choose is a decision rule. We either make the cut or we don't.
* We decide ahead of time what risk we will take in saying that something is different when in fact it is not.

## More NHST stuff

* If our rule is that we will reject the null hypothesis if the probability of the observed sample is less than 5 in 100 (p<.05) than the following is true:
* P=.049999 reject the null.
* P=.000001 reject the null.
* P=.050001 fail to reject the null.
* P=.99999 fail to reject the null.

## It's Fisher's fault (sort of)

* Who the hell came up with this crazy way of thinking?
* Sir Ronald Fisher. The father of modern statistics.

## What was Fisher thinking?

* Fisher came out of agronomy and was looking for a yes/no decision rule based on probability. Did this plant strain grow better than that plant strain? Was this manure better than that manure? (in science lingo we sometimes say " Those data aren't worth manure !" or something like that- its really just a tribute to Fisher.)

## Fisher vs. Pearson

* Karl Pearson was also one of the towering figures of modern statistics. Pearson felt that , instead of rejecting the null, we should choose between two alternative hypotheses.
* These guys (both Sirs) did not get along at all. In fact, Pearson (who founded Biometrika, wouldn't publish Fisher's papers. Fisher, turned down a prestigious position because he would have had to work for Pearson.

Vs.

## Interpreting a P-value

* Could random variation alone account for the difference between the null hypothesis and observations from a random sample?
* A small P-value implies that random variation because of the sampling process alone is not likely to account for the observed difference.
* With a small p-value we reject H0. The true property of the population is significantly different from what was stated in H0.
* Thus, small P-values are strong evidence AGAINST H0.
* But how small is small?

P = 0.1711

P = 0.2758

P = 0.0892

P = 0.0735

P = 0.01

P = 0.05

When the shaded area becomes very small, the probability of drawing such a sample at random gets very slim. Oftentimes, a P-value of 0.05 or less is considered significant: The phenomenon observed is unlikely to be entirely due to chance event from the random sampling.

Significant P-value???

## Where did p<.05 come from?

* Apparently Fisher once said, while taking tea, that " If the probability of such an event were sufficiently small say, 1 chance in 20- than one might regard the result as significant."
* In other words, it is somewhat arbitrary.
* Rosnow and Rosenthal have famously said " Surely God loves the .06 nearly as much as the .05".

## The Signal to Noise Ratio

* The z-test that we performed and all other statistical tests are signal to noise ratios.
* The signal is the observed difference between groups.
* The noise is the variability in measures between individuals within each group.
* It's like trying to listen to a conversation in a restaurant. If the background noise is too great, we can't distinguish the conversation from the noise that surrounds it.

## Signal to Noise

## Four outcomes of a decision

* There was a difference (signal) when there wasn't one.
* There was a difference (signal) when there really was one.
* There was no difference (signal) when there really was one.
* There was no difference (signal) when there really wasn't one.

Possible Decision Outcomes in Hypothesis Testing

## Type I and II errors

* A Type I error is made when we reject the null hypothesis and the null hypothesis is actually true (incorrectly reject a true H0).
* The probability of making a Type I error is the significance level
* A Type II error is made when we fail to reject the null hypothesis and the null hypothesis is false (incorrectly keep a false H0).
* The probability of making a Type II error is labeled .The power of a test is 1 b.

## Impact of sample size

* The spread in the sampling distribution of the mean is a function of the number of individuals per sample.
* The larger the sample size, the smaller the standard deviation (spread) of the sample mean distribution.
* But the spread only decreases at a rate equal to n.

Sample size n

Standard error n

## Type I error and sample size: comparing two means.

## Type II error and sample size: comparing two means.

## Power

* Power is the probability of concluding that there was a difference, when, in fact, there really was one (power=1-).
* If probability of Type II error is 20%, power =80% (1-.2=.8).

## Effect Size

* The effect size is a measure of the size of the difference between two samples or the degree of association between two variables.

## Putting it all together

* Type I error, Type II error, sample size, and effect size are all tied together.

## Decreasing Error

* How can we decrease the likelihood of making a Type I (saying it was different when it wasn't) or Type II error (saying it wasn't different when it was)?
* Increase sample size (decreases SE)
* Increase effect size (a bigger difference requires less precision).

## Confidence Intervals

* Confidence intervals are the area around a sample mean that we are sure that the true population mean lies- within a certain probability.
* 95%CI=Sample Mean +1.96\*SE
* The 95% confidence interval around the class mean sleep time is 7.19 + 1.96(.1004)=7.19 + .197 or 6.93 to 7.39.

## Confidence Interval

* The 95% confidence interval states that 95 times out of 100 the true mean lies within this range.

## Confidence intervals to test hypotheses

Because a two-sided test is symmetrical, you can also use a confidence interval to test a two-sided hypothesis.

/2

/2

In a two-sided test, C = 1 .
C =confidence level
=significance level

Ex: Your sample gives a 99% confidence interval of .
With 99% confidence, could samples be from populations with µ = 0.86? µ = 0.85?

99% C.I.

Logic of confidence interval test

Cannot reject
H0: m = 0.85

Reject H0 : m = 0.86

A confidence interval gives a black and white answer: Reject or don't reject H0. But it also estimates a range of likely values for the true population mean µ.
A P-value quantifies how strong the evidence is against the H0. But if you reject H0, it doesnt provide any information about the true population mean µ.

## Running a test of significance is a balancing act between the chance of making a Type I error and the chance of making a Type II error. Reducing reduces the power of a test and thus increases .

* It might be tempting to emphasize greater power (the more the better).
* However, with too much power trivial effects become highly significant.
* A type II error is not definitive since a failure to reject the null hypothesis does not imply that the null hypothesis is wrong.

# Placebo effect.ppt

## Placebo: Some Definitions

* an inert substance given to a patient as if it were a drug in order to placate or to serve as a control in an experiment.
* Placebo effects are the effects produced by placebos.
* in the Roman Catholic Church, vespers for the dead. Placebo in Latin means I shall please and is the first word in these vespers.
* By the 1300s the term had become pejorative and suggested a sycophant. Probably linked to the singing of paid mourners who sang placebos.

## Placebo: An expanded definition

* Physically inert substances and medical procedures that are identical in appearance to an active pharmacological or (less commonly) other medical treatment being investigated.
* Placebos can be active substances that mimic the side effects of the drug being investigated but do not possess the physical properties hypothesized to produce the beneficial treatment effect.
* From Kirsch Prevention and Treatment V.5, article 22, 2002

## Placebo history

* A landmark study, in the 1950s, by Henry Beecher of Harvard suggested that 30% to 40% of patients (for a wide range of afflictions) received relief from placebos.
* Beecher, H.K. The Powerful Placebo. JAMA, 1955; 159:1602-6.

## Evidence for the placebo effect

* Asthmatic patients show bronchoconstriction after inhaling a placebo bronchoconstrictor and bronchodilation after inhaling a placebo bronchodilator.
* Placebo morphine is more effective than placebo Darvon which is more effective than placebo aspirin.

## Placebo evidence (cont)

* Brand name placebos are more effective than generic placebos.
* Placebo injections are more effective than placebo pills.
* Higher doses of placebos produce a greater placebo response than lower doses of placebos.

## Even Placebo Surgery!

* A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee. J. Bruce Moseley, M.D., Kimberly O'Malley, Ph.D., Nancy J. Petersen, Ph.D., Terri J. Menke, Ph.D., Baruch A. Brody, Ph.D., David H. Kuykendall, Ph.D., John C. Hollingsworth, Dr.P.H., Carol M. Ashton, M.D., M.P.H., and Nelda P. Wray, M.D., M.P.H.
* N Engl J Med 2002; 347:81-88July 11, 2002
* Is Placebo Surgery Unethical? Sam Horng, B.A., and Franklin G. Miller, Ph.D. N Engl J Med 2002; 347:137-139July 11, 2002

## A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee.

* A total of 180 patients with osteoarthritis of
* the knee were randomly assigned to receive arthroscopic débridement, arthroscopic lavage, or placebo surgery.
* Results showed no difference in pain or functioning between the placebo group and the two surgery groups.

## Does the placebo effect mean I only have to think about working out-to get a work out?

* Mind-set matters: exercise and the placebo effect. Crum AJ, Langer EJ. Psychol Sci. 2007 Feb;18(2):165-71.

In a study testing whether the relationship
between exercise and health is moderated by ones mindset,
84 female room attendants working in seven different hotels were measured on physiological health variables affected by exercise.
Those in the informed condition
were told that the work they do (cleaning hotel rooms) is
good exercise and satisfies the Surgeon Generals recommendations for an active lifestyle. Examples of how their
work was exercise were provided.
Subjects in the control
group were not given this information

## Although actual behavior did not change, 4 weeks after the intervention, the informed group perceived themselves to be getting significantly more exercise than before. As a result, compared with the control group, they showed a decrease in weight, blood pressure, body fat, waist-to-hip ratio, and body mass index. These results support the hypothesis that exercise affects health in part or in whole via the placebo

Langer study Results.

## Does the placebo effect work even when we know its a placebo?

* Placebos without deception: a randomized controlled trial in irritable bowel syndrome. Kaptchuk TJ, Friedlander E, Kelley JM, Sanchez MN, Kokkotou E, Singer JP, Kowalczykowski M, Miller FG, Kirsch I, Lembo AJ. PLoS One. 2010 Dec 22;5(12):e15591.
* Two-group, randomized, controlled three week trial (August 2009-April 2010) conducted at a single academic center, involving 80 primarily female (70%) patients, mean age 47 ± 18 with IBS diagnosed by Rome III criteria and with a score 150 on the IBS Symptom Severity Scale (IBS-SSS). Patients were randomized to either open-label placebo pills presented as "placebo pills made of an inert substance, like sugar pills, that have been shown in clinical studies to produce significant improvement in IBS symptoms through mind-body self-healing processes" or no-treatment controls with the same quality of interaction with providers. The primary outcome was IBS Global Improvement Scale (IBS-GIS). Secondary measures were IBS Symptom Severity Scale (IBS-SSS), IBS Adequate Relief (IBS-AR) and IBS Quality of Life (IBS-QoL).

## Apparently it does!

* Open-label placebo produced significantly higher mean (±SD) global improvement scores (IBS-GIS) at both 11-day midpoint (5.2 ± 1.0 vs. 4.0 ± 1.1, p<.001) and at 21-day endpoint (5.0 ± 1.5 vs. 3.9 ± 1.3, p =.002). Significant results were also observed at both time points for reduced symptom severity (IBS-SSS, p =.008 and p =.03) and adequate relief (IBS-AR, p =.02 and p =.03); and a trend favoring open-label placebo was observed for quality of life (IBS-QoL) at the 21-day endpoint (p =.08).

## Placebos and Anti-depressants

* 30 years of placebo controlled trials show placebo response rates of 30% to 50% and drug response rates of 45% to 70%.
* Drug-placebo differences range from 18% to 25%.

## Are Anti-depressants effective?

* What do we make of the small effect size for anti-depressants?
* Is it all just about getting treated and the meds make no difference?
* Are we a victim of thinking like Gee this Drug Company pen is great! Their drugs must be great too!?
* Have we been scammed?

Effect of Hypericum perforatum (St Johns Wort) in Major Depressive Disorder. JAMA (2002) v.287(14):1807-1814.

## What about the placebo effect?

* Is it all in our heads?
* Maybe

## Changes in Brain Function of Depressed SubjectsDuring Treatment With Placebo

* Leuchter et al. (Am J Psychiatry 2002; 159:122129)
* QEEG study double-blind placebo trial of SSRIs.
* Findings suggest that effective placebo treatment induces changes in brain function that are distinct from those associated with antidepressant medication.

## Are Placebo effects getting stronger?

* Placebos Are Getting More Effective. Drugmakers Are Desperate to Know Why. Steve Silberman NY Times Magazine 08.24.09
* Placebo response is all about the power of expectations. If we expect to feel better we will feel better. If we expect to feel worse, we will feel worse (Nocebo effect).

## Has Big Pharma outsold itself?

* 15 years of pharmaceutical advertising has raised our expectations about pharmacological treatment.
* These expectations involve the same higher cortical centers in the brain involved in many of the maladies that drug companies are attempting to treat.
* By raising our expectations for treatment. Big Pharma may have increased placebo response itself. Drugs that once outperformed placebo no longer do.

# PY620C class 6 -t tests.ppt

## t-tests: History

* The Student's t-test was developed to compare the distribution of two independent groups on a continuous measure.
* For example, a group of professors and a group of students comparing beer on a measure of "drinkability".
* William Gossett, a statistician working for Guinness brewery developed the test to compare smaller samples. He published under the pseudonym "student" because Guinness employees were forbidden to publish.

## Why t? Why not s (for stout)?

* Why "t"? -Nobody really knows. Maybe Gossett developed the test during his afternoon tea (t) breaks?

=

## T-test: How does it work?

* The t-test is a signal-to-noise ratio. The "signal" is the difference between two means. The "noise" is variability in scores around the two means.
* The t-test is the ratio of the difference between the means over the standard error of that difference. Unlike the z-test, there is a different t value for each sample size (and significance level). When samples get large, t approaches z.

## T-test how does it work (cont)

* Remember that each sample is representative of a population. The t-test allows us to test whether there is likely to be a sufficient difference between population means to state that the two populations are different.

## The t test as a Signal-to-Noise Ratio

## The Independent t Test

* Used to compare the means of two independent or unrelated groups
* Formula:

Where:

## Degrees of Freedom (df )

* The number of observations minus the number of restrictions limiting the observations freedom to vary
* For the independent t-test, df = N 2
* If you calculate the deviations from the mean they should sum to zero. If you know all but one of the scores, the last score is no longer free to vary but is known in order to allow the deviations to sum to zero. Thus with 5 scores you have 4 degrees of freedom.

t values required for significance at various p levels

Rosnow, Beginning Behavioral Research, 5/e. Copyright 2005 by Prentice Hall

## T-test example data

## Calculating the Effect Size for t

Because the IV has only two groups, this effect size is the equivalent to the point biserial correlation

## T-test: Settings and sub-species

* Where observed: Any time two means are compared.
* T-tests need normally distributed environments in order to thrive. They do not fare well when the measures are badly skewed or groups are very different in size AND variance.
* When the two groups being compared are NOT independent (e.g. two measurements from the same person), a matched pairs t-test is used instead.

## Maximizing the t Test: Options

* Drive the means further apart
* e.g., Use a stronger treatment
* Decrease the variability within groups
* e.g., Standardize the research procedures
* e.g., Use a more homogeneous sample
* Increase the effective size of study

## Maximizing the t test: Conceptualizing the t Test

## Assumptions of the t Test

* Homogeneity of Variance: The population variance of the two groups is assumed to be equal.
* Serious violations can to result in a misleading p value and effect size r.

## Paired t-test

* Used to compare means of two groups that are related to each other in some way
* Formula:

Where MD = Mean of the difference in related measures and SD is the variance of the difference in those measures.

## Other Effect Size Indices