Statistics for Otolaryngologists

Maie A. St. John, MD, PhD

Division of Head & Neck Surgery

Department of Surgery

UCLA School of Medicine

What is Biostatistics?

- *Statistics*: is the art and science of making decisions in the face of uncertainty

- Biostatistics: statistics as applied to the life and health sciences

GENERAL APPROACH

- Concepts, not equations
- Goal is to increase awareness of statistical considerations
  - Organize data
  - Extract meaning form our data
  - Communicating our results
- Statistical software widely available to do basic calculations
- Good basic reference: Altman DG, *Practical Statistics for Medical Research*, Chapman & Hall/CRC

Excellent health statistics - smokers are less likely to die of age related illnesses.'
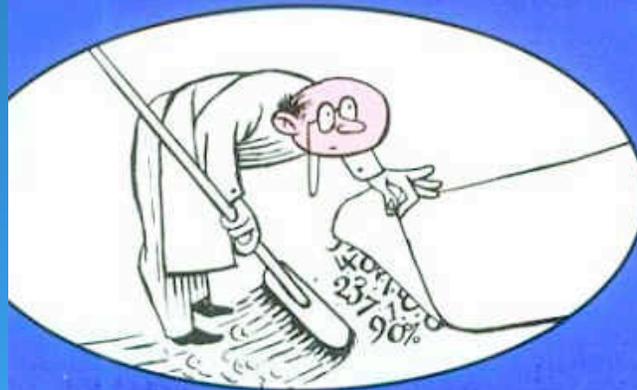
Sample versus Population

♦ The population defines normal.
   – Did anybody see the Twilight Zone episode where everybody has pig noses and the "beautiful girl" is ugly??
   – Normal is defined by its surroundings
♦ Best approach is to compare your patient to all similar individuals in the population.
   – This is practically impossible
     • Too many
     • Too expensive
     • Too much work!
   – Create a Sample population to compare

Sample

- ◆ Create a smaller version of the population to compare with your patient.
- ◆ Be sure it truly represents the population
  - – Best approach is a big sample drawn at random
  - – Usual approach is a small group and hope it is close to representative.

The population is important!

Back to terms

Variable: the event or characteristic under study.

– Dependent: the outcome you are interested in
  - The normal range for hemoglobin
    - Hemoglobin on the X axis
  - The range of **cholesterol** seen in army recruits

– Independent: the agent/event you believe might effect the dependent variable. There can be more than one.
  - Does smoking effect the serum cholesterol
    - Cholesterol on the X axis, **Smoking in Packs/Day on the Y**
  - Does salt intake effect systolic blood pressure
    - Systolic BP on the X, **Na intake in mg on the Y**

Types of variables

- The type of variable you wish to study
  - Determines the Statistical Test used to differentiate normal from abnormal
  - Defines the study type needed to answer the clinical question
- Frames the question
  - Yes/No
  - Honors/High Pass/Pass/Fail
  - 0,1,2,3,4… 100

Still can't answer what is normal...yet.

♦ A sample is used to estimate what one would expect in the vast majority of the population

  – You can't measure the entire country's hemoglobin to reach the normal curve

  – Estimate with census data

♦ Even more

  – Biologic variables vary among and within individuals: there is not one normal hemoglobin value

  – So is normal is fuzzy, what does that make abnormal?
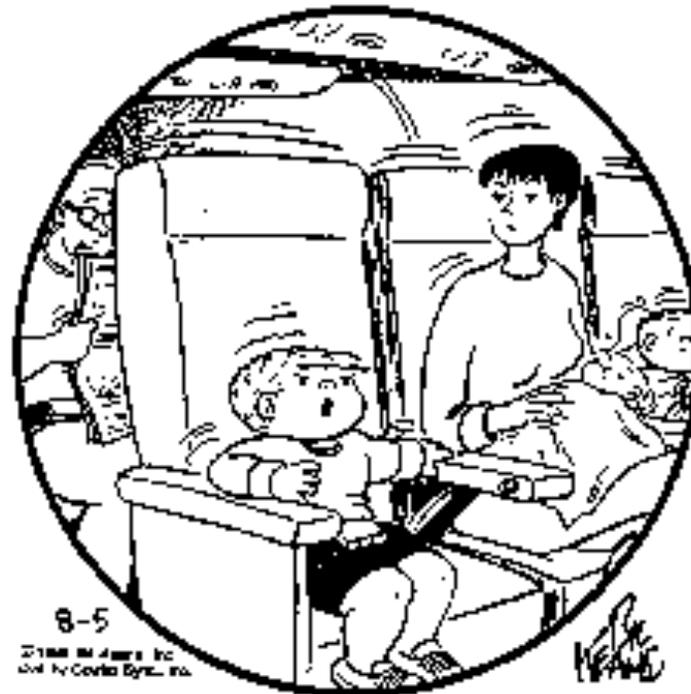
Statistics to the Rescue

Statistics to the Rescue!

♦ Central Tendency can help.

- The fact that variables focus around an average value and we can define this and describe the pattern of values around the average

- Mean, Median, Mode, Variance, Deviation…

Are you looking at the data backwards?

Two Kinds of Statistics

- Descriptive
  - Describe a set of data
  - Reporting measures (observations)

- Inferential
  - Make predictions based on a collection of observations (tools for generalizing beyond observations)

Definitions

- Parameter: usually a numerical value that describes a population

- Statistic: usually a numerical value that describes a sample

- Data: (plural) are measurements or observations. A datum (singular) is a single measurement or observation and is more commonly called a score or a raw score.

Data

- Collection of observations from a survey or experiment

- Two types of data:
  - Qualitative
  - Quantitative

Types of Data

- Qualitative Data
  - A single observation represents a class or category (marital status, type of car)

- Quantitative Data
  - A single observation is an amount or count (reaction time, weight, distance from campus)

- *Look at a single observation to help you decide if it is qualitative or quantitative*

Quantitative or Qualitative?

- Political Party
- Blood Pressure
- Body Temp
- Gender
- Place of Residence

Two Types of Data

- Discrete:
  - Consist of a countable number of possible values; countable with integers (i.e. number of residents in this room)
- Continuous:
  - Consist of an infinite number of possible values on a scale in which there are no gaps or interruptions (i.e. height or weight)
- Both can be infinite; continuous data have a higher degree of infinity

Scales of Measurement

- The methods of assigning numbers to objects or events

- There are 4 scales of measurement:
  - Nominal
  - Ordinal
  - Interval
  - Ratio

Nominal

- Think Nominal: names
- Labels
- Identify different categories
- No concept of more, less; no magnitude
- Data cannot be meaningfully arranged in order

- Examples: Gender, Ice cream flavors, fruits

Ordinal

- Think Ordinal: Order
- Ordered set of observations
- Different categories, but the categories have a logical order & magnitude
- Distances between categories varies

- Examples: Class rank, Sports Rankings

Interval

- Think: Interval has constant intervals

- Different categories, logical order

- Distance between categories is constant

- No meaningful zero point – no ratio comparisons (zero point may be lacking or arbitrary)

- Examples: Temperature in F or C, Pain sensitivity?

Ratio

- Think: ratio allows ratio comparisons
- Different categories, logical order, constant distances, meaningful zero
- Interval scale with a true zero
- Zero: absence of the quantity being measured

- Examples: Height, Weight, temperature in K

# Three things we want to know about a set of Test Data

- Shape
- "Typical" value
  - Measurement of central tendency
- Spread of Scores
  - Measure of variability

DESCRIBING DATA

- Two basic aspects of data
  - Centrality
  - Variability
- Different measures for each
- Optimal measure depends on type of data being described

Things to Remember: Bar Graphs

- Height of bar = frequency of that category
- Width of bar: irrelevant but should be constant
- If width is irrelevant so is area
- Bars do not touch!
- Nominal or Ordinal Data


Types of Graphs: Bar Graph
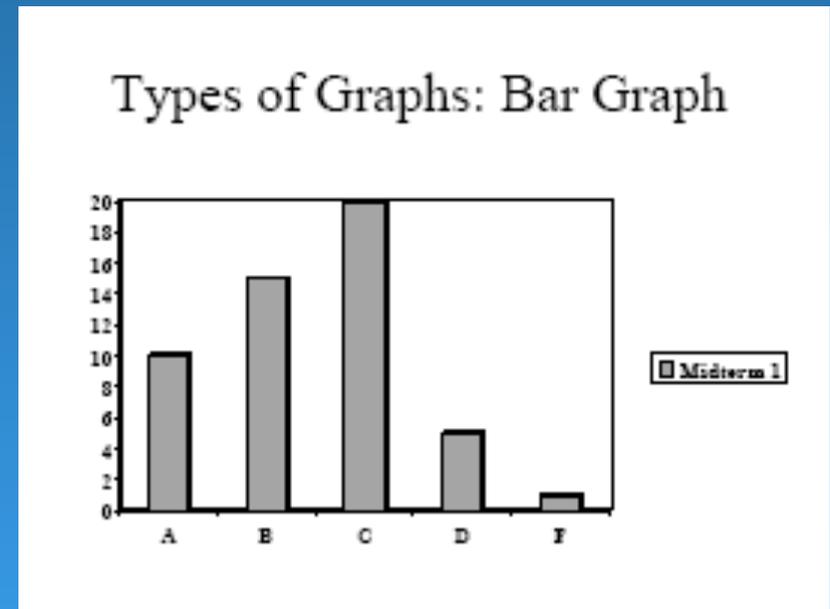
Things to Remember: Histograms

- A graphical way to display frequencies using a bar graph
- Height of bar = frequency of scores for the interval
- Width of bar = width of range
- Height is meaningful; width is meaningful so therefore area is meaningful
- Area= total number of observations in a given "range"
- Bars touch
- Interval or ratio data

Types of Graphs: Histogram
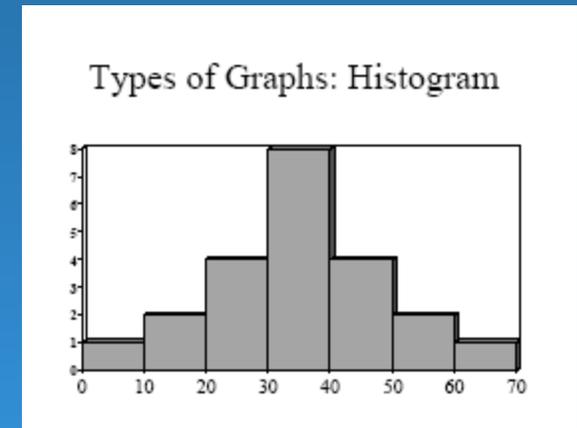
Sample Histogram
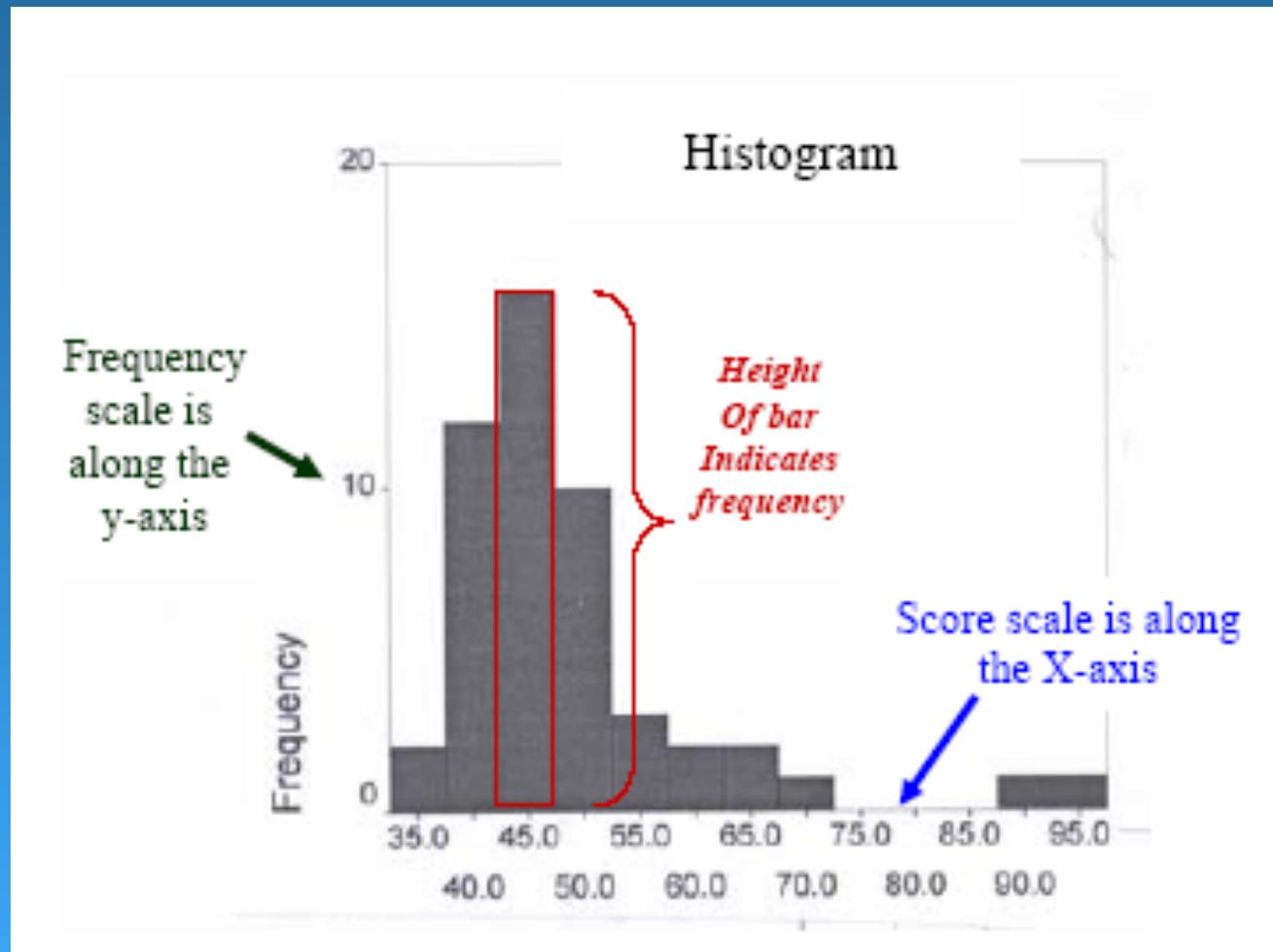
Stem and Leaf Display

- A graphical technique to display all data
- The stem: the leading (most significant) digits
- The leaves: the trailing (less significant) digits
- Allows for a manageable frequency count of individual items)

- The leaf is the digit in the place farthest to the right in the number, and the stem is the digit, or digits, in the number that remain when the leaf is dropped.
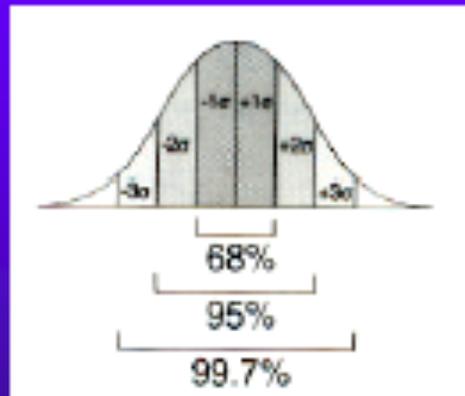
Stem-and-Leaf Plot	
Ages of People at Family Reunion	
Stem	Leaf
0	1, 8, 9
3	2, 4, 7
4	5
5	1, 5
8	1

(Ages: 1,8,9,32,34,37 etc)

"Statistics say that religious people live longer, so I practice a different religion every day of the week to be sure I'm covered."

Characteristics of the Normal Curve

68%
95%
99.7%

- ◆ Unimodal
  - – One mode only
- ◆ Symmetric
  - – 50% of the values occur above and below the mean and median
- ◆ Asymptotic
  - – Never reaches 0 or 1
  - – The curve is a probability not a value. (not real numbers)
- ◆ Mean, Median and Mode are the same

CENTRALITY

- Mean
  - Sum of observed values divided by number of observations
  - Most common measure of centrality
  - Most informative when data follow normal distribution (bell-shaped curve)
- Median
  - "middle" value: half of all observed values are smaller, half are larger (the 50th percentile); split the difference if even number in the sample)
  - Best centrality measure when data are skewed
- Mode
  - Most frequently observed value
  - There can be more than one mode

MEAN CAN MISLEAD

- Group 1 data: 1,1,1,2,3,3,5,8,20
  - Mean: 4.9   Median: 3   Mode: 1
- Group 2 data: 1,1,1,2,3,3,5,8,10
  - Mean: 3.8   Median: 3   Mode: 1
- When data sets are small, a single extreme observation will have great influence on mean, little or no influence on median
- In such cases, median is usually a more informative measure of centrality

Measures of Dispersion

♦ How does the data occur around the mean, median and mode.

- Min-max: lowest and highest values
- Range: difference between the highest and lowest values
- Variance: the average value of the squared deviations of each value from the mean; (s^2)
- Standard Deviation: the square root of the variance; (s)

VARIABILITY

- Most commonly used measure to describe variability is standard deviation (SD)
- SD is a function of the squared differences of each observation from the mean
- If the mean is influenced by a single extreme observation, the SD will overstate the actual variability
- SEM: standard method of calculating the stability of the mean
  - SEM = SD $\div \sqrt{n}$

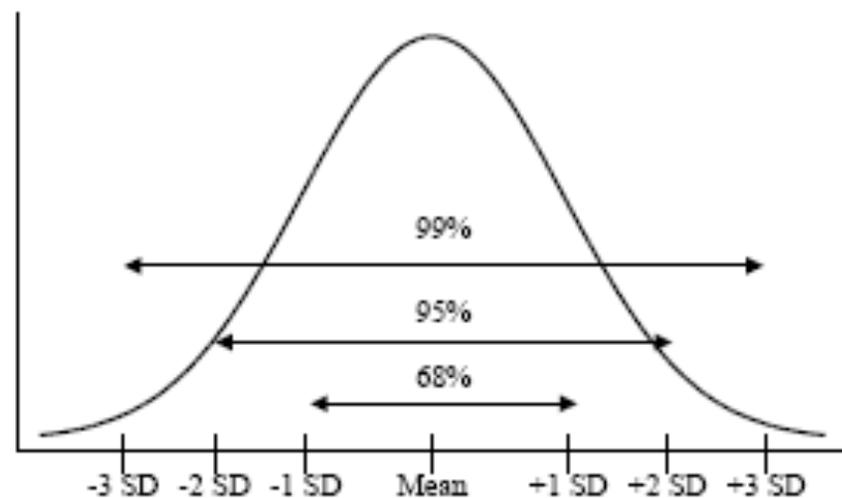# Characteristics of the Normal Curve

68%
95%
99.7%
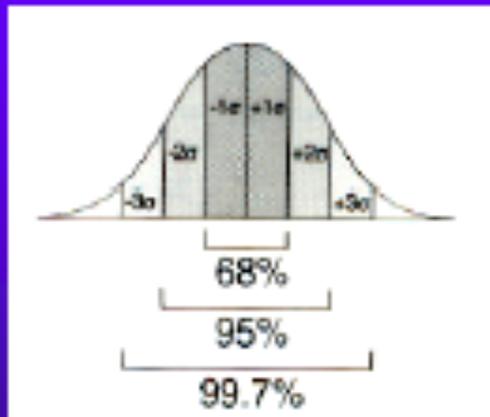
◆ The SD or σ is the amount of the population encompassed by those values

◆ 1 SD around the mean is 68% of individuals have a value inside while 32% lie outside

   – Half of the 68% is above the mean/median, half below.

   – You can define normal in this curve.

Normal Curve Characteristics

The Normal Curve

Normal Curve: Characteristics

In any normal curve:

68.26% of the area falls within 1SD of the mean

95.44% of the area falls within 2SD of the mean

99.74% of the area falls within 3SD of the mean

Guidelines:

2/3 of the area falls within 1SD of the mean

95% of the area falls within 2SD of the mean

99% of the area falls within 3SD of the mean

To obtain exactly 95% of the observation frequencies, 1.96 standard deviations on either side of the mean are required=inner 95th percentile range or ipr_{95}

CONFIDENCE INTERVALS

- A confidence interval is intended to provide a sense of the variability of an estimated mean

- Can be defined as the set of possible values that includes, with specified probability, the true mean

- Confidence intervals can be constructed for any type of variable, but here we consider the most common case of a normally distributed variable

- (based on SEM; ipr_{95} based on SD)
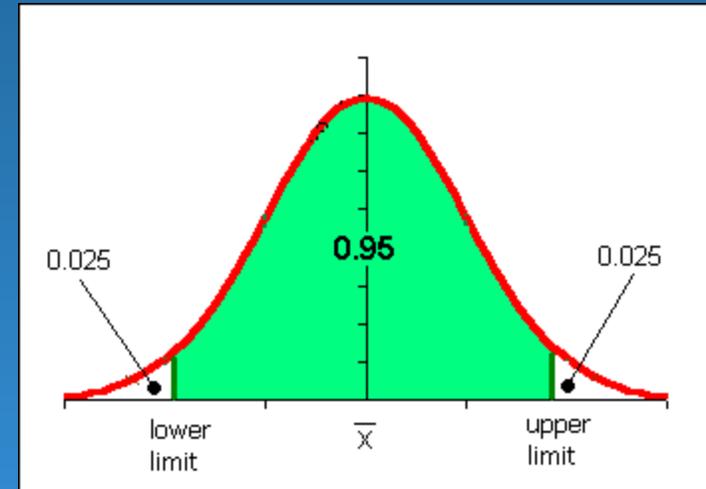
VALUE OF CONFIDENCE INTERVALS

- Two data sets may have the same mean; but if one data set has 5 observations and the second has 500 observations, the two means convey very different amounts of information

- Confidence intervals remind us how uncertain our estimate really is
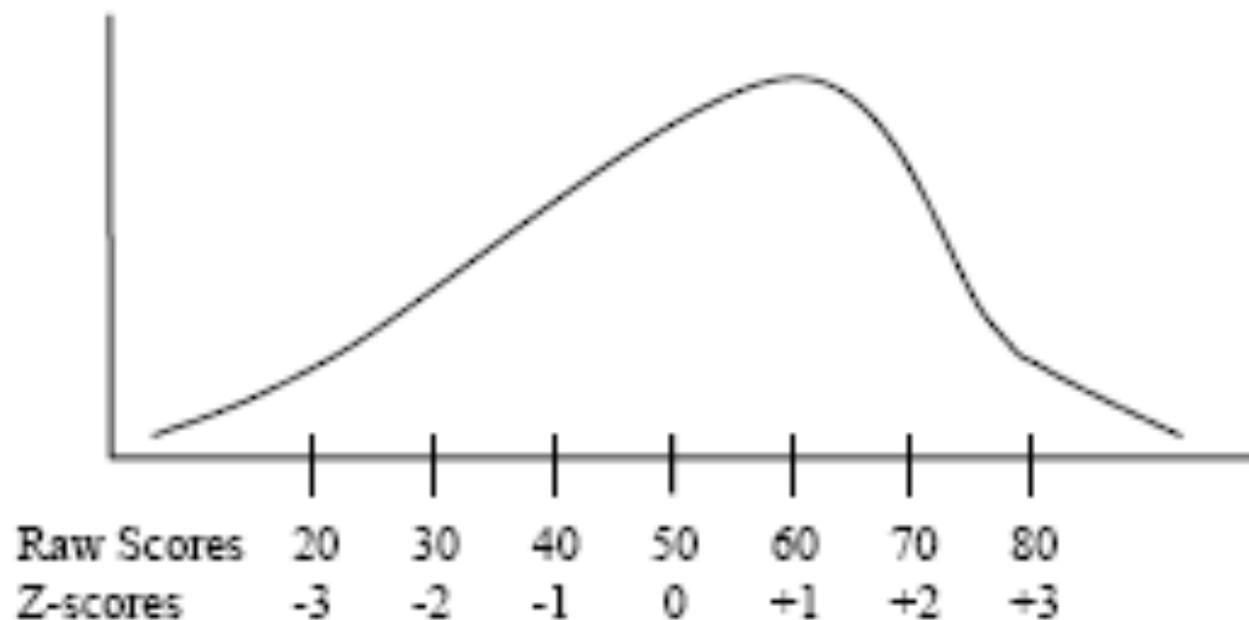
Characteristics of the Normal Curve
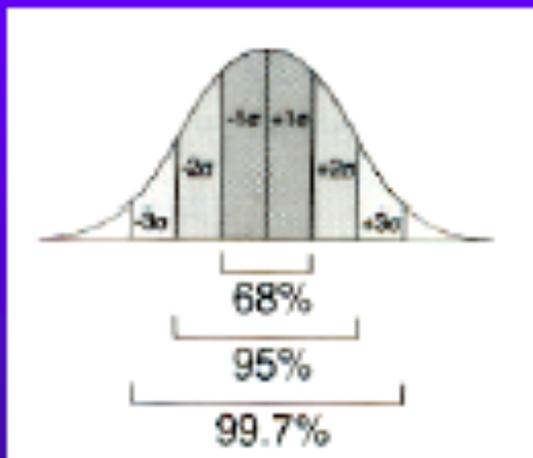
- ◆ This curve can represent variation in a sample
  - − Standard deviation
    ### OR
- ◆ Variation of samples around the population
  - − Standard Error
- ◆ A single group compared within itself
  - − Standard Deviation
- ◆ Multiple Groups compared to each other
  - − Standard Error

Z-scores

- Indicate how many sd a score is away from the mean
- Two components:
  - Sign: positive (above the mean) or negative (below the mean
  - Magnitude: how far form the mean the score falls
- Can be used to compare scores within a distribution and across different distributions provided that the different distributions have the same shape

Properties of Z-Scores

- If you transform a set of raw scores into z-scores, the shape of the distribution does not change!

Raw Scores	20	30	40	50	60	70	80
Z-scores	-3	-2	-1	0	+1	+2	+3

Normal Curve: Characteristics

- A normal curve expressed as z-scores is called a **standard normal curve.**
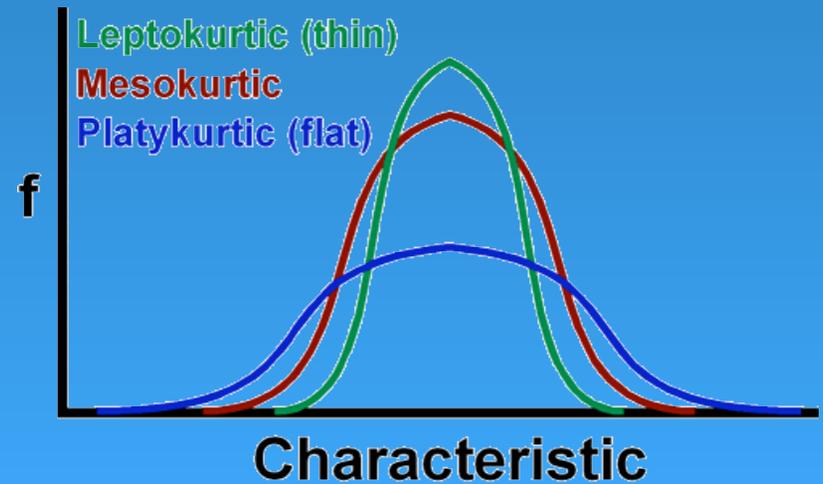
Skewness

- Skewness: the degree of asymmetry of a distribution
  - To the left: negatively skewed
  - To the right: positively skewed
  - The skewness value for a symmetric distribution is zero.
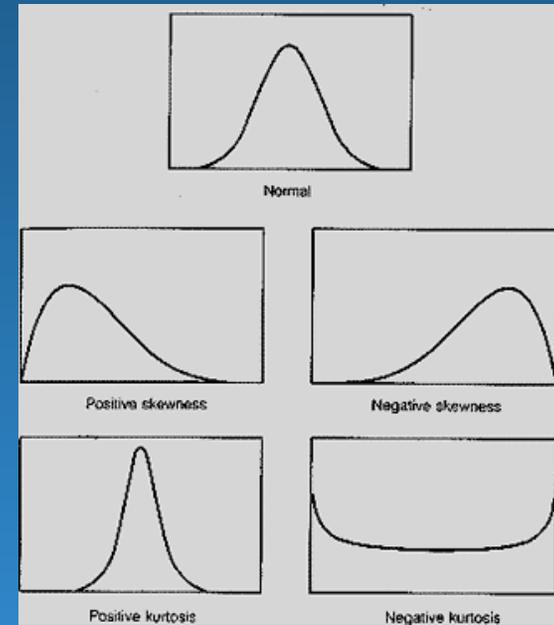
Kurtosis

- Kurtosis: the "peakedness" of a distribution

- The kurtosis value for a normal distribution is zero.

Where to now?

- Now that we can define our data and know how to plot it, where do we go?

Null Hypothesis

- H_0: no difference between test groups really exists
  - Most statistical tests are based on rejection of H_0
- Statistical Hypothesis testing is used to check for one of 2 conditions:
  - Is there a difference between sets of data?
  - Is there a relationship between sets of data
- The test does NOT prove that there is or is not, but allows us to know (and set) the statistical probability on which the decision was made.

P values

♦ The probability of a specific value or a more extreme value in the given population

♦ In clinical medicine a p value of less than or equal to 0.05 (5%) is usually considered statistically significant

  – A P value states chance that the relationship between our variables is a "fluke" or

    • p=0.05, a 5% chance that the values compared are not different statistically (overlap between two distributions)

    • 5% of all values are further from the central point on the Distribution curve.(normal ranges for a single value)

Which test should I use?

- Chi square?
- t-test?
- Mann- Who?

Chi square (χ^2) test of Independence

- Tests for independence between 2 nominal or ordinal variables
- Comparison between f observed in cells (f_o) and the numbers you would expect (f_e) if variables were statistically independent
- If H_0 of no association is true, then f_o and f_e will be close and the χ^2 value small
- If H_0 of no association is false, then f_o and f_e will be farther apart and the χ^2 value larger
- $\chi^2=0$ when $f_o = f_e$

Chi square test of independence

- *Eyeball* differences between percentages: large enough to be "important"
- Better: Are they *statistically significant*?
- *Statistical significance*: are observed differences significantly different from zero that they could not occur by chance?

Statistical association
(statistical dependence)

Percentage believing in life after death by religious affiliation				
Belief in life after death	Religious affiliation			
	Protestant	Catholic	Jewish	Other
Yes	75.0	86.7	10.0	15.0
N	(200)	(150)	(50)	(100)

Statistical independence

Belief in life after death	Religious affiliation (hypothetical)			
	Protestant	Catholic	Jewish	Other
Yes	120 (60%)	90 (60%)	30 (60%)	60 (60%)
No	80	60	20	40
Total	200	150	50	100

Interpretation: religious affiliation has no effect on whether one believes in life after death

t-test

- A test of the <u>null hypothesis</u> that the <u>means</u> of two <u>normally distributed</u> populations are equal.
- When comparing 2 groups on a continuous variable, significance depends on:
    - The magnitude of the observed difference
    - The amount of spread or variability of the data
    - When comparing > 2 groups use analysis of variance ANOVA
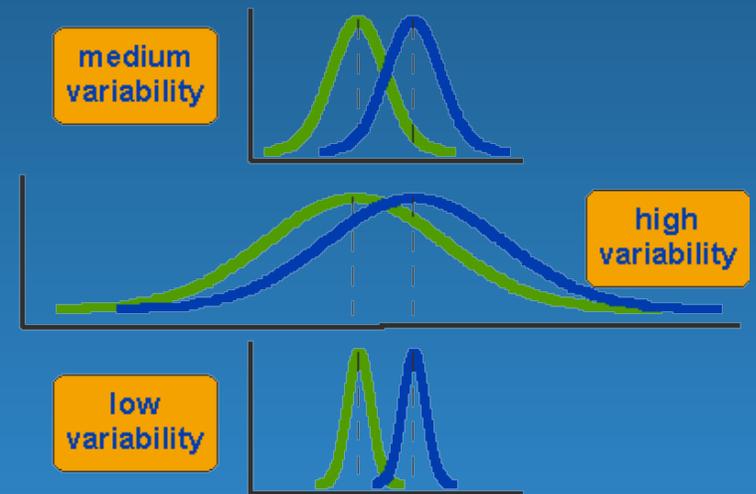
$$\frac{signal}{noise} = \frac{difference\ between\ group\ means}{variability\ of\ groups}$$

$$= \frac{\bar{X}_T - \bar{X}_C}{SE(\bar{X}_T - \bar{X}_C)}$$

$$= t\text{-value}$$

From one μ to two

- Independent samples
  - none of the observations in one group is in any way related to the observations in the other group

- Dependent samples
  - typically consist of a matched sample (or a "paired" sample) or one group that has been tested twice (repeated measures).

Statistical Hypotheses

- $H_0: \mu_1 - \mu_2 = 0$
- $H_1: \mu_1 - \mu_2 \neq 0$ (two-tailed)
- $H_1: \mu_1 - \mu_2 > 0$ (one-tailed, positive difference)
- $H_1: \mu_1 - \mu_2 < 0$ (one-tailed, negative difference)

One-tailed versus Two-tailed Test

Two-tailed   One-tailed

$\frac{\alpha}{2}$ α
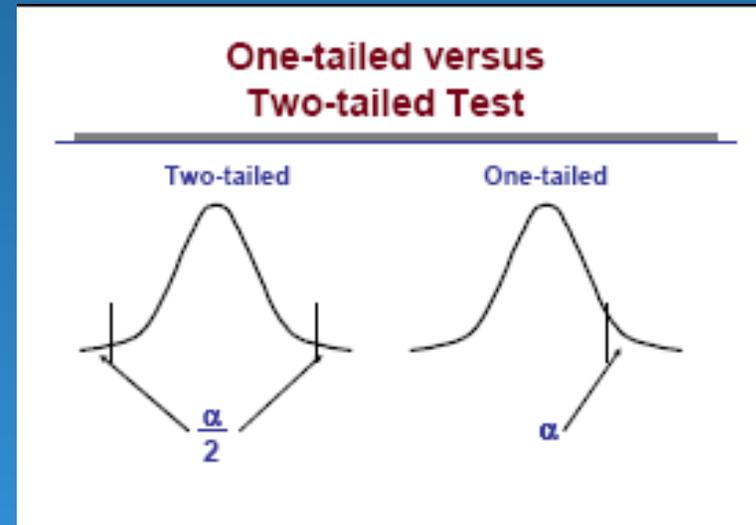
Wilcoxon-Mann-Whitney
aka rank sum test

- Also used to compare 2 independent samples
  - Different from t test b/c it is valid even if the population distributions are not normal
  - Data are form random samples
  - Observations are independent
  - Samples are independent
- Distribution-free type of test
  - Does not focus on any one parameter like the mean
  - Instead examines the distributions of the 2 groups
- The test statistic denoted by U
  - Large U = 2 samples are well separated with little overlap
  - Small U = 2 samples are not well separated with much overlap

Wilcoxon-Mann-Whitney vs. Independent Test

● Both try to answer the same question, but treat data differently.

- W-M-W uses rank ordering
  - ❑ Positive: doesn't depend on normality or population parameters
  - ❑ Negative: distribution free lacks power because it doesn't use all the info in the data
- T-test uses actual Y values
  - ❑ Positive : Incorporates all of the data into calculations
  - ❑ Negative : Must meet normality assumption
- neither is superior

● So…

- If your data are normally distributed use the t-test
- If your data are not normal use the WMW test

What % of residents are asleep now vs at the beginning of this talk?

?