Medical Statistics 101

Donald E. Cutlip, MD
Beth Israel Deaconess Medical Center
Harvard Clinical Research Institute
Harvard Medical School

Presenter Disclosure Information

Donald E. Cutlip

The following relationships exist related to this presentation:

No Industry Relationships to Disclose

Salary support from and consultant to Harvard Clinical Research Institute, an academic research organization performing contracted research activities for stent and drug manufacturers

Outline

- **Data Measurement and Comparison**
  - Parametric versus non-parametric testing
- **Hypothesis Testing**
  - Alpha error
  - Beta error and power
  - Sample size calculations
  - Confidence intervals

Data Types

- **Continuous**
  - Data can assume all possible numerical value along a continuum within a specified range
  - Most quantitative measures are continuous
- **Categorical or Discrete**
  - Values can only fall into a limited number of separate categories with no possible intermediate levels
  - Dichotomous = Yes or no responses such as baseline characteristics or event status
  - Nominal = no inherent order (gender, blood type)
  - Ordinal = numerical ranking

Measuring Continuous Data

Suppose you want to assess late loss for a new DES

- Population of interest = all patients with coronary artery disease who may be candidates for DES
- Sample of this population is chosen for study
- Late loss is measured for each subject in sample

Two Measures of Interest for Analysis

- Central tendencies (signal of interest)
  - Mean
  - Median
- Data Dispersion (noise)
  - Variance

Measuring Continuous Data II

- **Example: Late loss values obtained for 200 patients ranging from -0.2 mm – 3.2 mm**
- **Central tendency**
  - Mean = Sum of all late loss values/N subjects
  - Median = 50th percentile (middle) value
- **Dispersion – Spread of data around the central tendency**

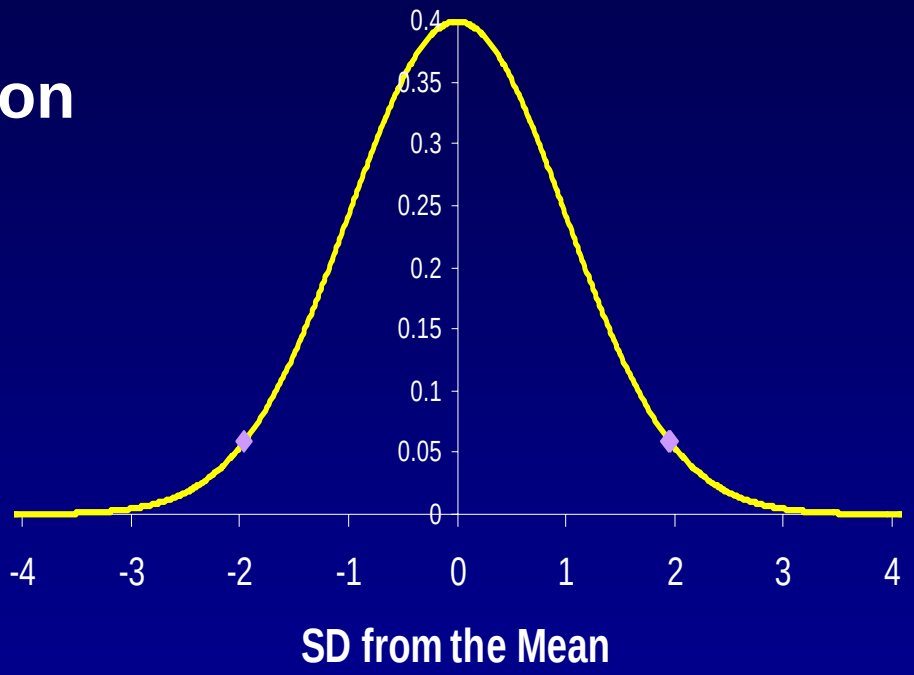
Dispersion for Means

$$\textit{Variance} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$$

Standard deviation = $\sqrt{\text{variance}}$

Distribution of Continuous Data

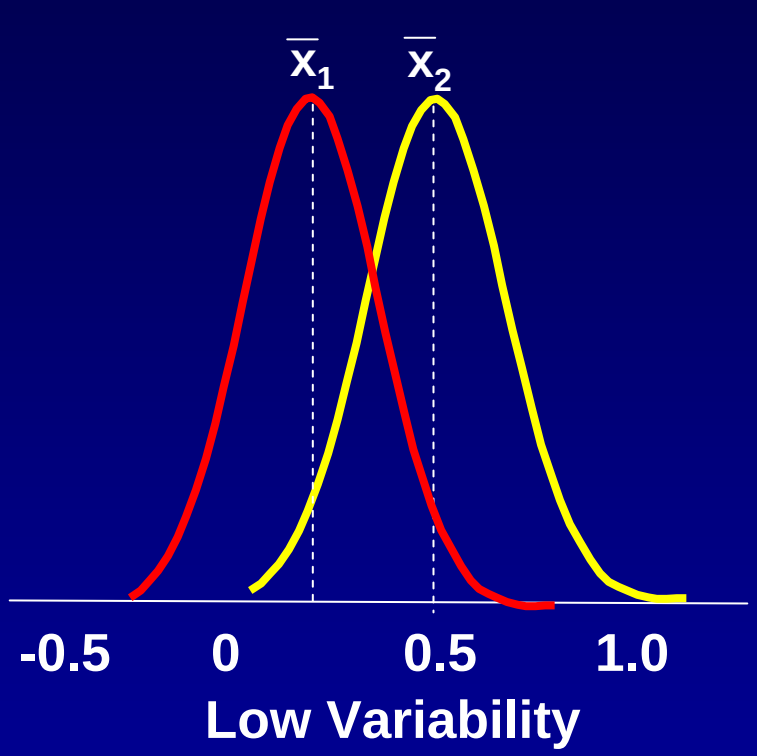
- Normal/Gaussian distribution
- Described by μ & σ
- Mean = Median
- Symmetrical
- Area under the curve sums up to 100%

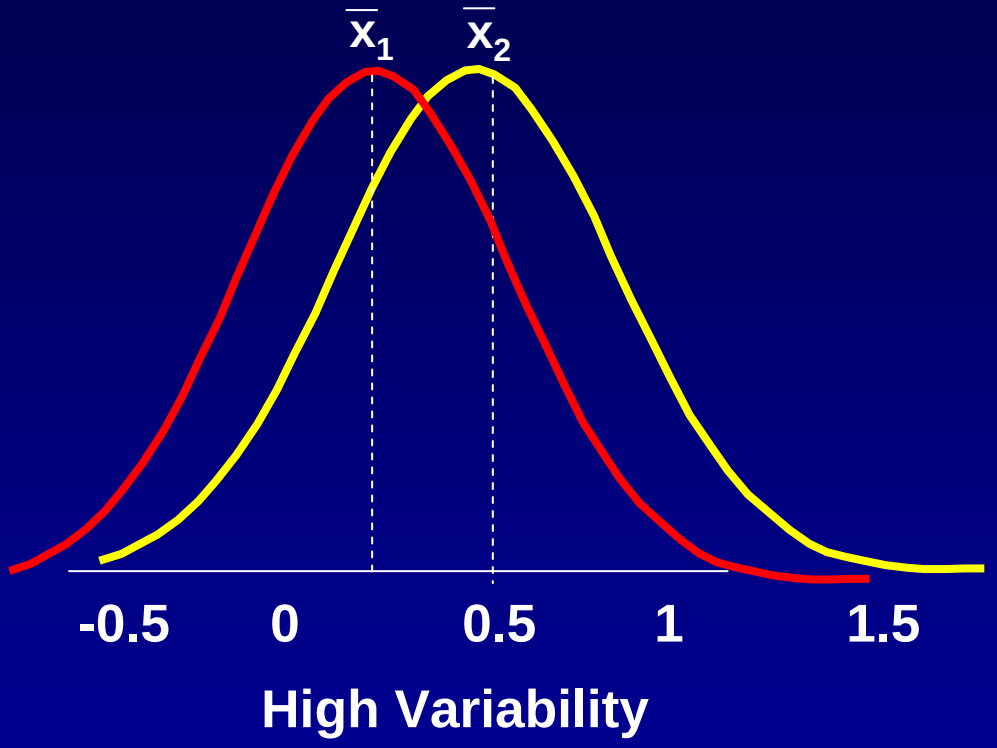


Standard Deviation =

- Zone from -1SD to 1SD covers 68.3% of the sample
- Zone from -1.96SD to 1.96SD covers 95% of the sample
- Zone from -2.58SD to 2.58SD covers 99% of the sample

Comparing Continuous Data





For data that are normally distributed, comparison of means as a function of variance
= Parametric Tests

Parametric Tests

Signal/Noise ratio = Differences in Means/Variability

- **t-Test – Useful for comparing 2 groups; Variability = SE of difference; calculates t statistic > look-up p value**

$$t = \frac{x1 - x2}{\sqrt{var1/n1 + var2/n2}}$$

Variances may be equal or unequal in the 2 groups

- **ANOVA - Analysis of Variance (test for differences in populations means based on variance); useful for comparing >2 groups; calculates F statistic> lookup p value**
- **Linear regression – tests for associations with variable of interest and other variables**

Non-Parametric Tests

If data are not normally distributed

- **Makes no assumptions about distribution of data**
- **Represented by median and interquartile range (25th and 75th percentile)**
- **Statistical tests based on ranking of data points and rank value per group**
- **Examples:**
  - Wilcoxon sign rank (2 correlated measures)
  - Mann Whitney (2 independent groups)
  - Kruskal-Wallis (multiple groups)
  - Spearman Correlation
- **Why not play it safe and just use non-parametric tests?**
  - Less powerful - more difficult to show difference

Measuring Discrete Data

- **Represented as number, frequency, or rates**
- **Rate or proportion = *probability* of success or failure**
- **Statistical tests for proportions assume a binomial distribution with mean = the proportion and variance directly related to p * (1-p) and inversely related to sample size**
- **Tests (based on observed vs expected rates)**
  - Chi-square
  - Fisher's exact (if any cell <5)

2 x 2 Table

	Yes	No	
Group 1	**a**	**b**	**a + b**
Group 2	**c**	**d**	**c + d**
	a + c	**b + d**	

Hypothesis Testing

- Null Hypothesis – H0: PA = PB
- Alternate Hypothesis – H1: PA ≠ PB
- Objective is to reject or fail to reject the null hypothesis. If null is rejected then accept the alternative hypothesis
- Alternative hypothesis implies difference in either direction (A better or worse than B); Two-sided test

Hypothesis Testing 2

- After comparison using statistical tests based on data type and distribution etc, obtain probability that observed result is due to random chance and consistent with the null hypothesis.
- By convention, probability of 5% is usually accepted as convincing evidence.
- The probability of this chance – rejecting the null hypothesis (claiming difference) when it is true (no difference) is termed alpha error
- Since 2-sided test, there is 2.5% chance for error in either direction

Hypothesis Testing 3

- **Since it is not possible to prove the null hypothesis there is some probability that an experiment will fail to reject the null hypothesis even though it is false**
- **This is termed beta error**
- **The power of a study (reject the null hypothesis when it is false and claim a difference) = 1 – beta**
- **Generally study designs include power of at least 80% or *20% probability of failure to reject even though false***

Sample Size

- **Often the most important factor to the success of a trial design**
- **Benefits of smallest possible sample size due to costs and time for enrollment**
- **Sample size dependent on:**
  - alpha error (usually fixed by convention)
  - treatment effect (should be fixed and reasonable)
  - variance (fixed if known)
  - beta error or power (becomes the major driver of sample size in design variations)

Sample Size 2

If you hold all the other variables constant,

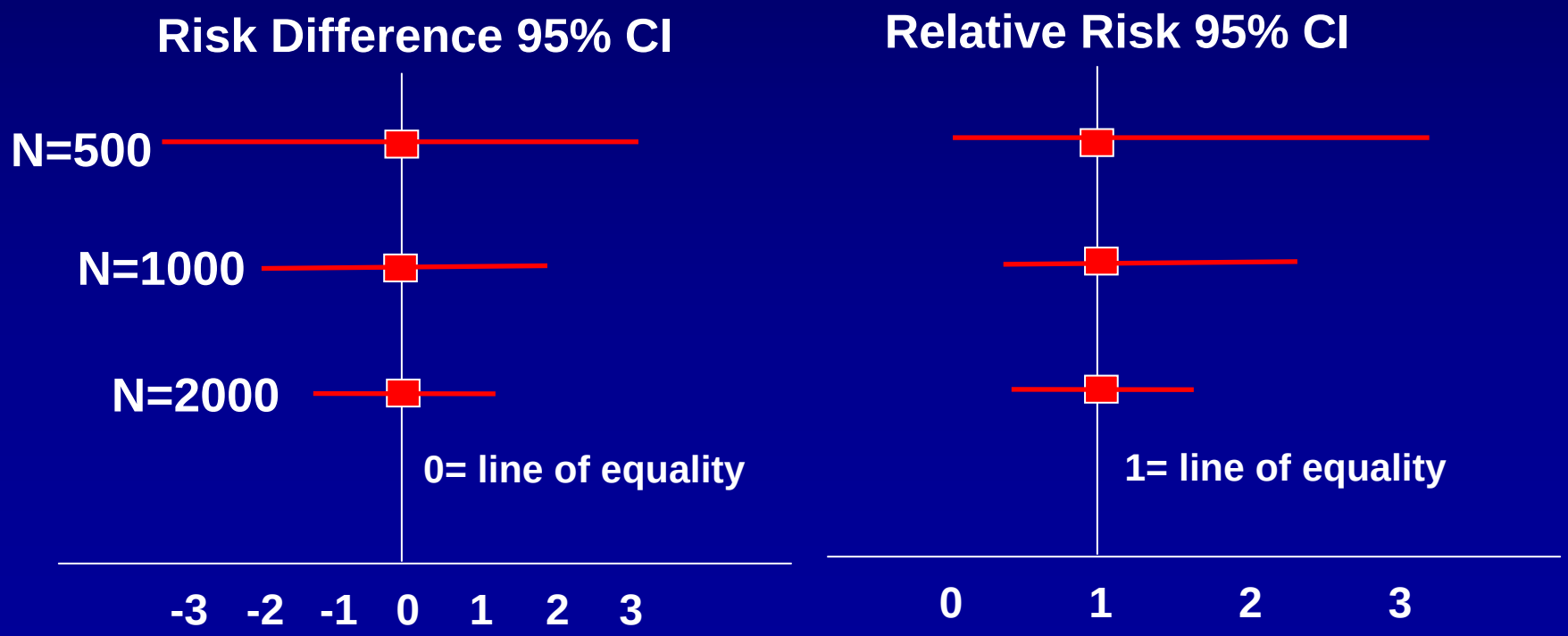
- If alpha decreases, then power decreases (i.e., it's harder to find a difference if you use a very stringent p-value).
- If the magnitude of difference sought increases, power increases (i.e., it's easier to detect a big difference, harder to detect a small difference).
- If the sample size increases, then power increases (i.e., it's easier to find a difference if you have a large sample).
- If the standard deviation of the variable in question increases, the power decreases (i.e., more noise, less power).

Confidence Intervals

- **Level of assurance that the results are reproducible and not due to chance**
- **Most common level is 95% (correlates with alpha error of 5%) interpreted as: If the same trial were conducted 100 times then 95 times the estimate would be within the specified interval**
- **Confidence intervals are calculated based on actual result, the desired level of assurance, and a measure of variance (standard error)**
- **Confidence intervals can be based on:**
  - Proportion or mean estimate (actual study result)
  - Risk difference (Group 1 – Group 2)
  - Relative risk

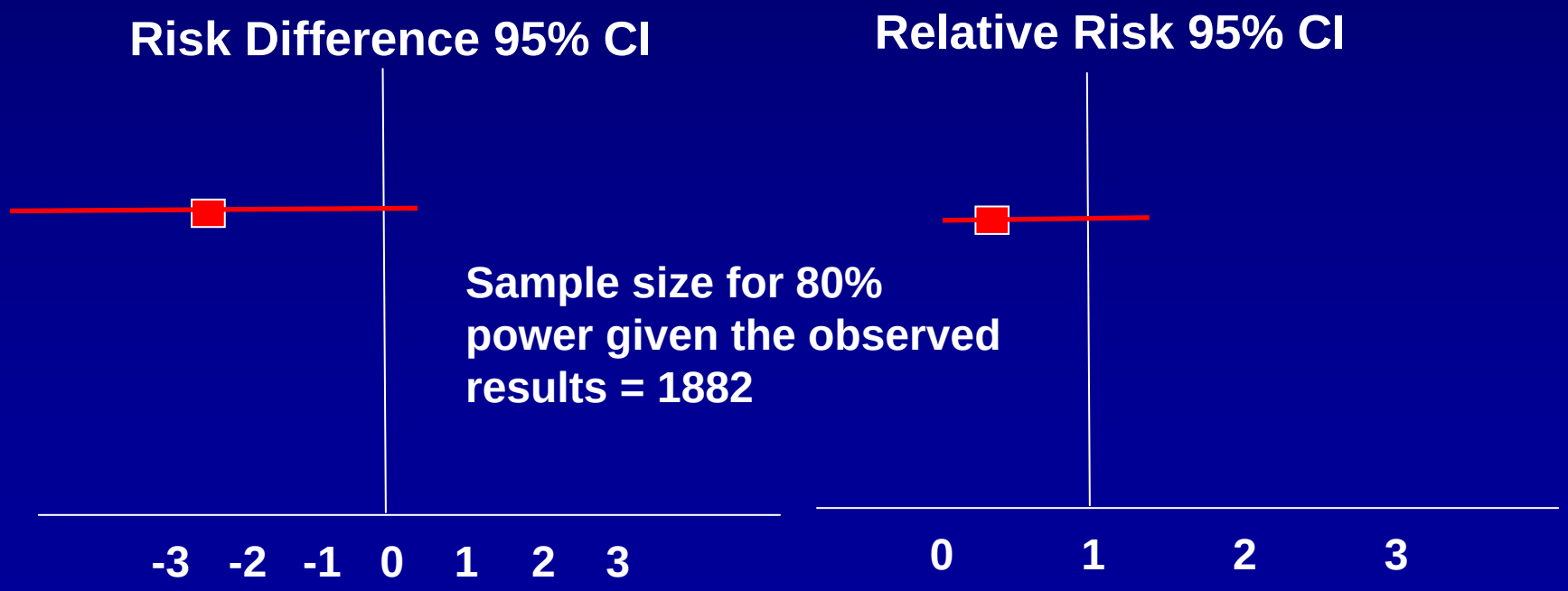
Confidence Intervals & Interpretation of Results

- Suppose a clinical trial with 250 pts in each of 2 arms has death or MI rate of 2% in each arm. The authors report no differences in safety – Is this true?



Confidence Intervals & Impact of Sample Size

- Suppose a clinical trial sponsor has new DES with expected 0% TLR; assume control DES = 5%, new DES =0.5%; sample size = 500
- Actual Results at 1 year: 3.2% vs. 1.2% (observed 62% relative risk reduction)



Beware of Underpowered Trials

Claims of "No Difference"

- **Failure to reject the null hypothesis does not equate equivalence or mean "not different"**
- **Possibilities**
  - Play of chance (80% power means 20% chance of failure to reject null when in fact is false, i.e. there is a difference)
  - Overly optimistic treatment effect in trial design
  - Failure to attain event rate assumptions
- **Confidence intervals help determine what difference can be excluded with specified assurance**
- **Requires formal non-inferiority trial design**