EMBLEM
High Quality, High Impact Research

Cancer Research Methods Workshop

Mayur M. Desai, PhD, MPH

Mwanza, Tanzania

September 18-19, 2012

Yale SCHOOL OF PUBLIC HEALTH

Hypothesis Testing and Making Inferences

Drawing inferences from a study

- In general, we are taking measurements on a sample of individuals and using those individuals to make inferences about the population from which those individuals were drawn
- A *parameter* is a numerical measure of the population, while a *sample statistic* or *parameter estimate* is the measure obtained from the sample

Estimates of error

- We normally cannot perform repeated sampling schemes, but instead use measurements only on a single sample
- As a result, there is a possibility that our sample estimates are different from the population parameters – the question is how different?
- Every sample estimate thus has an element of error associated with it – this is a measure of the uncertainty associated with a single estimate

Confidence intervals

- Given a sample statistic (e.g., a mean), a confidence interval around the statistic is a range of values that we are confident (95% confident) include the "true" population value

Confidence intervals (cont.)

- To calculate a CI, we need 3 pieces of information:
  - the sample statistic
  - it's standard error (SE)
  - the level of desired confidence

95% CI = estimate ± 1.96 (SE of the estimate)

- A typical level of desired confidence is 95%, which corresponds to 1.96 times the SE on either side of the statistic
- The 1.96 corresponds to the number of standard deviations on either side of the mean that cover 95% of the population distribution

What does the CI mean?

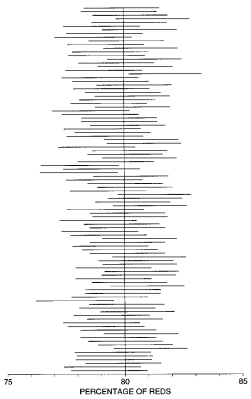
- A CI means that you are 95% sure (or any other level of confidence set) that the true value of the measure falls within the range of the interval
- The larger the SE, the wider the CI, and the less precise the estimate

95% CI = estimate ± 1.96 (SE of the estimate)

A simple example

- Suppose that 100 students are asked to estimate the percentage of red marbles in a large box with thousands of marbles
- Each student takes a simple random sample of 250 marbles and computes a 95% CI for the percentage of reds in the box, using the formula:

➢ percentage of reds in sample ± 1.96*SE

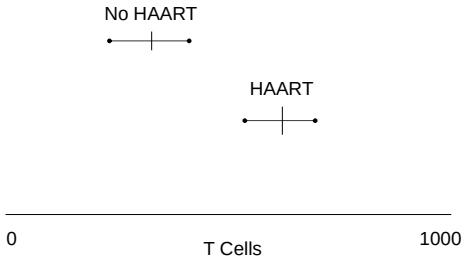


The 95% CI is shown for each of the 100 different samples. The interval changes from sample to sample. For about 95% of the samples, the interval covers the true population percentage, marked by a vertical line.

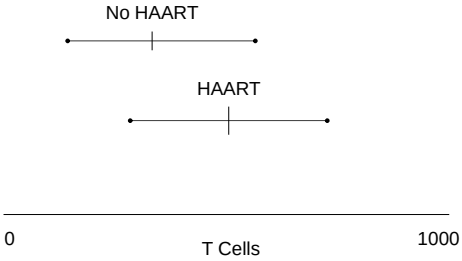
How can we use the CI to make comparisons?

- Confidence intervals can be very useful in comparing groups with one another in a very intuitive way
- Consider a study that takes a sample of HIV+ patients and divides them into those on HAART and those not, and let's compare the mean CD4 cell count of the 2 groups

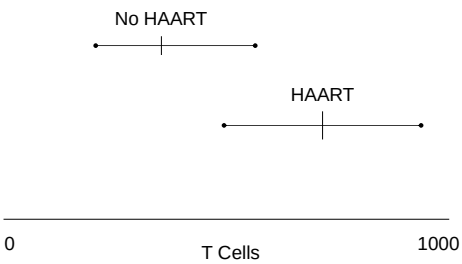
Scenario 1



Scenario 2



Scenario 3



Underlying the confidence interval calculation

- Assumptions:
  - the only thing that would differ in hypothetical replications of the study would be the statistical, or chance, element in the data
  - the variability in the data can be described adequately by statistical methods, and biases are nonexistent
- These assumptions may not be realistic in health care studies

Hypothesis testing

- Usually, we want to know not only whether groups are different, but also by how much they are different, and whether that difference could have been found by chance
- This is the basic aim of hypothesis testing
- The effect is the numerical value of the comparison across groups (e.g., difference in means, odds ratio, etc.)

You can never "prove" that something is true…

- …only demonstrate that it is not true
- Therefore, in hypothesis testing, we hypothesize that there is no difference between groups (called the null hypothesis)
- Then, we try to "reject" that hypothesis

Rejection

- We reject the null hypothesis when it is highly improbable that we could have obtained our observed data if the null hypothesis were true
- The p-value represents that probability
- Thus, the smaller the p-value, the more likely we are to reject the null hypothesis

How do we figure out the p-value?

- The p-value is linked to the probability distribution of the test statistic
- We (1) calculate the value of the test statistic, (2) place that value on the known distribution of that test statistic under the null hypothesis, and then (3) calculate the area under the curve (AUC) to the right/left of that statistic
- The AUC (p-value) is the probability of observing a test statistic that size or larger purely by chance

Interpreting p-values

- After calculating the p-value, we evaluate that value against an arbitrary cutoff (in most instances, 0.05 or 5%) to determine whether a test statistic is “statistically significant”
- This simply means that there is a small (but non-zero) chance that the null hypothesis is true