

## Large Sample Confidence Interval for Means and Proportions

Dr. Tom Ilvento  
FREC 408

### Thus far

- We established that the mean has a sampling distribution that is known
  - A sampling distribution is the theoretical distribution of taking repeated samples of similar size  $n$
- The mean and variance of the sampling distribution is equal to
  - $\mu$  the population parameter
  - $\sigma^2/n$

### Thus far

- I can take a random sample of the population
- And compare it to a theoretical sampling distribution
  - Based on the population
  - or a hypothesized population
- I will place the comparison in a probabilistic framework – due to sampling error

### Thus far

- I place it in a probabilistic framework to
  - Place a bound of error or confidence interval around my estimate
  - Or to see how likely or unlikely it is my sample estimate compares to a hypothesized value

### Confidence Interval

- Suppose I am concerned about the quality of drinking water for people who use wells in a particular geographic area
- I will test for nitrogen, as Nitrate+Nitrite
- The U.S. EPA sets a MCL of 10 mg/l of Nitrate/Nitrite (MCL=Maximum contaminant level)
  - Below the threshold is considered safe

### Water Quality Example

- Let's say there are 2,500 households in the area
- I could try to test them all, but at \$50 a test it would cost \$125,000 and weeks of work
- So, I decide to take 50 well water samples, and test for the presence of nitrogen

### Water Quality Example

- My sample
  - $n = 50$
  - Mean = 7 mg/l
  - $s = 3$  mg/l
  - Standard error =  $3/(50)^{.5} = .424$
- The sample provides an estimate – **Point Estimate**, a single value computed from a sample and used to estimate the value of the target population. (Def7.1 p339)
  - $\bar{x}$  and  $s^2$  are point estimates of population mean  $\mu$  and population variance  $\sigma^2$  respectively.

### Water Quality Example

- But I know other possible samples would have yielded a slightly different mean level
- I would like to place a bound of error around the estimate

### Water Quality Example

- I need to think of my sample as one of many possible samples
- I know from our work on the Normal curve that a z-value of  $\pm 1.96$  corresponds to 95 percent of the values
  - A z-value of 1.96 is associated with a probability of .475 on one side of the normal curve
  - 2 times that value yields 95%

### Water Quality Example

- If I think of my sample as part of the sampling distribution
- I can place a  $\pm 1.96$ (standard error) around my estimate
- Like this:
  - $7 \pm 1.96(.424)$
  - $7 \pm .832$
  - 6.168 to 7.832

### Why did I use the standard error in my formula?

- I am asking the question about the mean level of nitrate-nitrite in the wells in the area
- I want some sense of how well my sample estimates the population
  - If it is drawn randomly it will represent the population
  - Plus some sampling error

### Why did I use the standard error in my formula?

- I need to think of my sample as one of the many possible samples I could have taken
- A 95% confidence interval means that
  - If I would have taken all possible samples
  - And calculated a confidence interval for each one
  - 95% of them would have contained the true population parameter

### To construct a confidence interval we need

- An point estimator
- A sample and a sample estimate using the estimator
- Knowledge of the Sampling Distribution of the point estimator
  - Standard Error
  - The form of the sampling distribution

### To construct a confidence interval we need

- A probability level we are comfortable with – how much certainty. It's also called "Confidence Coefficient" in book on page 343.
  - $\alpha$  refers to the combined area to the right and left of our interval
  - A 95% C.I. Has
    - $\alpha = 1 - .95 = .05$
    - $\alpha$  refers to the probability of being wrong in our confidence interval

### To construct a Confidence Interval

- An point estimator
  - The mean, given as
    - $\sum x/n$
- A sample and a sample estimate using the estimator
  - A particular sample which is drawn randomly, where we calculate  $\bar{x}$

### To construct a Confidence Interval

- Knowledge of the Sampling Distribution of the estimator
  - Standard Error
  - The form of the sampling distribution
- A probability level we are comfortable with – how much certainty
- Sampling Distribution of the mean
  - $s/\sqrt{n}$
  - Distributed normally
  - 95% C.I. or  $\alpha = .05$

### What is a Confidence Interval?

- It is an interval estimate of a population parameter
- The plus or minus part is also known as a Bound of Error
- Placed in a probability framework

### What is a Confidence Interval?

- We calculate the probability that the estimation process will result in an interval that contains the true value of the population mean
  - If we had repeated samples
  - Most of the C.I.s would contain the population parameter
  - But not all of them will