STA291
Spring 2009

1

LECTURE 17

THURSDAY, 2 APRIL

Preview & Administrative Notes

- **10 Estimation**
  - **10.1 Concepts of Estimation**
- *Next online homework due next Sat*
- *Suggested* Reading
  - Study Tools Chapter 10.1, 10.2
  - OR: Sections 10.1, 10.2 in the textbook
- *Suggested* problems from the textbook:

  10.1, 10.2, 10.6, 10.10, 10.12, 10.14, 10.16

  10.41, 10.42, 10.51, 12.54, 12.55, 12.58, 12.65
- Exam 2 next Tueday (7 April)

Le Menu

- **10 Estimation**
  - 10.1 Concepts of Estimation
  - **10.2 Estimating the Population Mean**
  - **10.3 Selecting the Sample Size**
  - **(12.3) Confidence Interval for a Proportion**

Confidence Intervals

- A large-sample 95% confidence interval for the population mean is $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

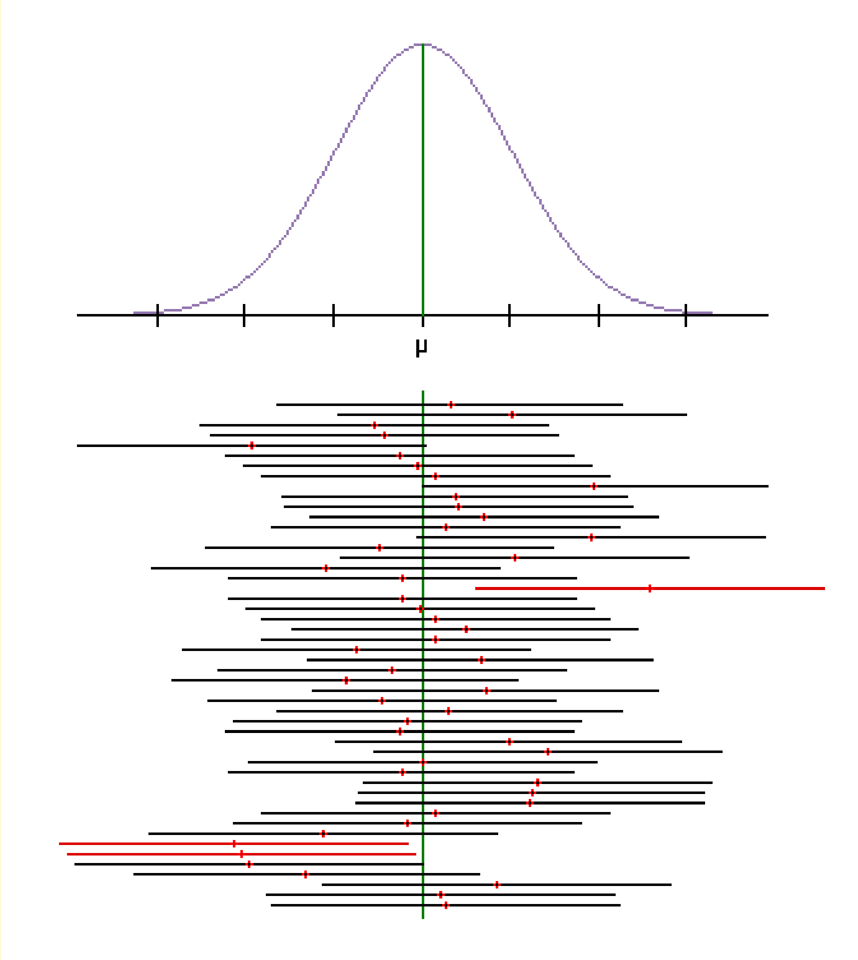
where $\bar{X}$ is the sample mean and

σ = population standard deviation

Confidence Intervals—Interpretation

- “Probability” means that “in the long run, 95% of these intervals would contain the parameter”
- If we repeatedly took random samples using the same method, then, in the long run, in 95% of the cases, the confidence interval will cover (include) the true unknown parameter
- For one given sample, we do not know whether the confidence interval covers the true parameter
- The ***95% probability*** only refers to the ***method*** that we use, but not to the individual sample

Confidence Intervals—Interpretation



Confidence Intervals—Interpretation

- To avoid misleading use of the word "probability", we say:

  "We are 95% confident that the true population mean is in this interval"

- Wrong statement:

  "With 95% probability, the population mean is in the interval from 3.5 to 5.2"

Confidence Intervals

- If we change the confidence coefficient from 0.95 to 0.99 (or .90, or .98, or ...), the confidence interval changes
- Increasing the probability that the interval contains the true parameter requires increasing the length of the interval
- In order to achieve 100% probability to cover the true parameter, we would have to take the whole range of possible parameter values, but that would not be informative
- There is a tradeoff between precision and coverage probability
- *More coverage probability = less precision*

Example

- Find and interpret the 95% confidence interval for the population mean, if the sample mean is 70 and the sample standard deviation is 10, based on a sample of size

  1. $n = 25$
  2. $n = 100$

Confidence Intervals

- In general, a large sample confidence interval for the mean μ has the form

$$\left[\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}} \right]$$

- Where z is chosen such that the probability under a normal curve within z standard deviations equals the confidence coefficient

Different Confidence Coefficients

- We can use Table B3 to construct confidence intervals for other confidence coefficients
- For example, there is 99% probability that a normal distribution is within 2.575 standard deviations of the mean

  (z = 2.575, tail probability = 0.005)
- A 99% confidence interval for μ *is*

$$\left[\bar{X} - 2.575\frac{\sigma}{\sqrt{n}}, \bar{X} + 2.575\frac{\sigma}{\sqrt{n}}\right]$$

Error Probability

- The error probability (α) is the probability that a confidence interval does **not contain the** population parameter
- For a 95% confidence interval, the error probability $\alpha = 0.05$
- α = 1 – confidence coefficient, or
- confidence coefficient = 1 – α
- The error probability is the probability that the sample mean $\bar{X}$ falls more than *z* standard errors from μ (in both directions)
- The confidence interval uses the *z*-value corresponding to a one-sided tail probability of $\alpha/2$

Different Confidence Coefficients

Confidence Coefficient	α	$\alpha/2$	$z_{\alpha/2}$
.90	.10		
.95			1.96
.98			
.99			2.58
			3.00

Facts about Confidence Intervals

- The width of a confidence interval
  - __________ as the confidence coefficient increases
  - __________ as the error probability decreases
  - __________ as the standard error increases
  - __________ as the sample size increases

Facts about Confidence Intervals II

- If you calculate a 95% confidence interval, say from 10 to 14, there is ***no probability associated with the true unknown parameter*** being in the interval or not
- The true parameter is either in the interval from 10 to 14, or not – we just don't know it
- The 95% refers to the method: If you repeatedly calculate confidence intervals with the same method, then 95% of them will contain the true parameter

Choice of Sample Size

- So far, we have calculated confidence intervals starting with z, s, n: $\bar{X} \pm z\frac{\sigma}{\sqrt{n}}$
- These three numbers determine the margin of error of the confidence interval: $z\frac{\sigma}{\sqrt{n}}$
- What if we reverse the equation: we specify a desired precision B (bound on the margin of error)???
- Given z and σ, we can find the minimal sample size needed for this precision

Choice of Sample Size

- We start with the version of the margin of error that includes the population standard deviation, σ, setting that equal to B:

$$B = z\frac{\sigma}{\sqrt{n}}$$

- We then solve this for n:

$$n = \left\lceil \sigma^2 \left(\frac{z^2}{B^2} \right) \right\rceil$$, where $\lceil \ \rceil$ means "round up".

Example

- For a random sample of 100 UK employees, the mean distance to work is 3.3 miles and the standard deviation is 2.0 miles.
- Find and interpret a 90% confidence interval for the mean residential distance from work of all UK employees.
- About how large a sample would have been adequate if we needed to estimate the mean to within 0.1, with 90% confidence?