STA291
Fall 2009

LECTURE 25

THURSDAY, 19 NOVEMBER

Confidence Interval

- An inferential statement about a parameter should always provide the probable accuracy of the estimate
- How close is the estimate likely to fall to the true parameter value?
- Within 1 unit? 2 units? 10 units?
- This can be determined using the sampling distribution of the estimator/ sample statistic
- In particular, we need the standard error to make a statement about accuracy of the estimator

Confidence Interval—Example

- With sample size n = 64, then with 95% probability, the sample mean falls between

$$\mu - 1.96\frac{\sigma}{\sqrt{64}} = \mu - 0.245\sigma \quad \& \quad \mu + 1.96\frac{\sigma}{\sqrt{64}} = \mu + 0.245\sigma$$

Where μ = population mean and

σ = population standard deviation

Confidence Interval

- A confidence interval for a parameter is a range of numbers within which the true parameter likely falls

- The probability that the confidence interval contains the true parameter is called the *confidence coefficient*

- The confidence coefficient is a chosen number close to 1, usually 0.95 or 0.99

Confidence Intervals

- The sampling distribution of the sample mean $\bar{X}$ has mean μ and standard error $\dfrac{\sigma}{\sqrt{n}}$

- If n is large enough, then the sampling distribution of $\bar{X}$ is approximately normal/bell-shaped (Central Limit Theorem)

Confidence Intervals

- To calculate the confidence interval, we use the Central Limit Theorem

- Therefore, we need sample sizes of at least, say,

  $n = 30$

- Also, we need a z–score that is determined by the confidence coefficient

- If we choose 0.95, say, then $z = 1.96$

Confidence Intervals

- With 95% probability, the *sample mean* falls in the interval

$$\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}}$$

- Whenever the sample mean falls within 1.96 standard errors from the population mean, the following interval contains the *population mean*

$$\bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}}$$

Confidence Intervals

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

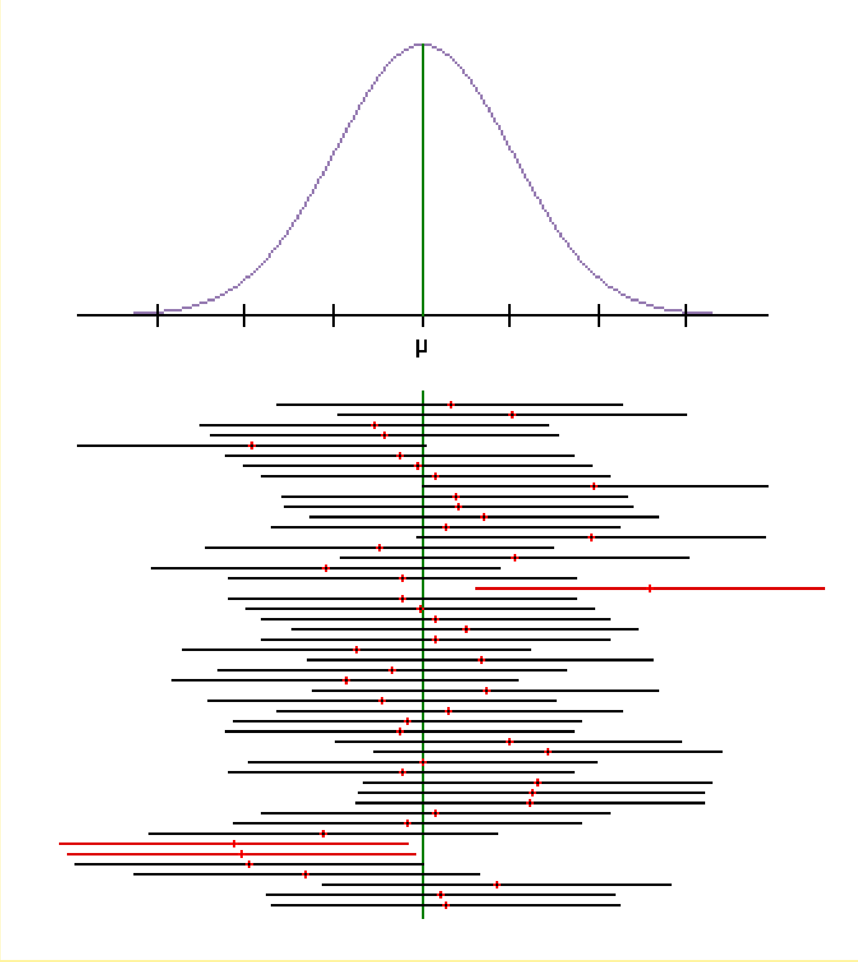
- where $\overline{X}$ is the sample mean and

- s is the sample 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

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

- 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*

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

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: $\bar{X} \pm z \frac{s}{\sqrt{n}}$
- These three numbers determine the margin of error of the confidence interval: $z \frac{s}{\sqrt{n}}$
- 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 $\lceil \quad \rceil$ 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?