Ch. 7 - Estimation: Single Population

Ercan Karadas

New York University

ercan@nyu.edu

Summer, 2016

Confidence Intervals

Contents of this chapter:

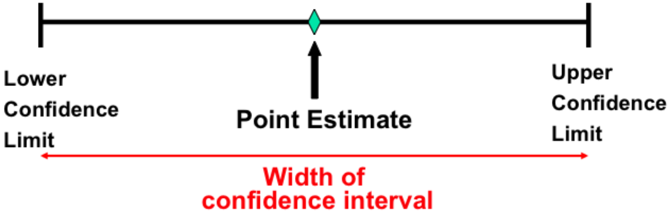
- Confidence Intervals for the Population Mean, μ
  - when Population Variance σ^2 is Known
  - when Population Variance σ^2 is Unknown
- Confidence interval estimates for the variance of a normal population
- Sample-size determination

Properties of Point Estimators

- An estimator of a population parameter is
  - a random variable that depends on sample information . . .
  - whose value provides an approximation to this unknown parameter
- A specific value of that random variable is called an estimate

Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about variability



Unbiasedness

- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if its expected value is equal to that parameter:

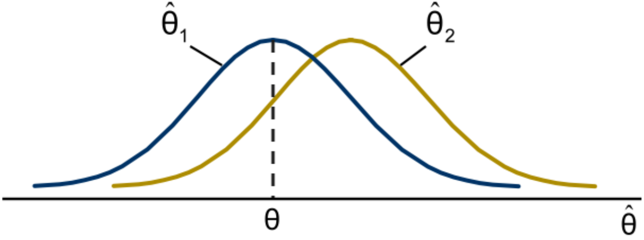
$$\boxed{E(\hat{\theta}) = \theta}$$

- Examples:
  - The sample mean $\bar{x}$ is an unbiased estimator of μ
  - The sample variance s^2 is an unbiased estimator of σ^2
  - In this course, population parameter θ is either μ or σ^2 and the estimator $\hat{\theta}$ is either $\bar{x}$ or s^2.

Unbiasedness

(continued)

- $\hat{\theta}_1$ is an unbiased estimator, $\hat{\theta}_2$ is biased:



Bias

- Let $\hat{\Theta}$ be an estimator of θ
- The bias in $\hat{\theta}$ is defined as the difference between its mean and θ

$$\boxed{\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta}$$

- The bias of an unbiased estimator is 0

Most Efficient Estimator

- Suppose there are several unbiased estimators of θ
- The most efficient estimator or the minimum variance unbiased estimator of θ is the unbiased estimator with the smallest variance
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ, based on the same number of sample observations. Then,
  - $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if $\boxed{\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)}$
  - The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is the ratio of their variances:

$$\boxed{\text{Relative Efficiency} = \frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)}}$$

Confidence Interval Estimation

- How much uncertainty is associated with a point estimate of a population parameter?
- An interval estimate provides more information about a population characteristic than does a point estimate
- Such interval estimates are called confidence interval estimates

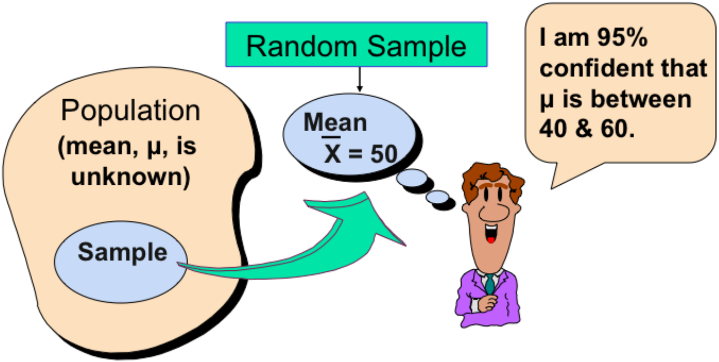
Confidence Interval Estimate

- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observation from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - Can never be 100% confident

Confidence Interval and Confidence Level

- If $P(a < \theta < b) = 1 - \alpha$ then the interval from a to b is called a $100(1 - \alpha)\%$ confidence interval of θ.
- The quantity $100(1 - \alpha)\%$ is called the confidence level of the interval
  - α is between 0 and 1
  - In repeated samples of the population, the true value of the parameter θ would be contained in $100(1 - \alpha)\%$ of intervals calculated this way.
  - The confidence interval calculated in this manner is written as $a < \theta < b$ with $100(1 - \alpha)\%$ confidence

Estimation Process



Confidence Level, $(1-\alpha)$

(continued)

- Suppose confidence level = 95%
- Also written $(1 - \alpha) = 0.95$
- A relative frequency interpretation:
  - From repeated samples, 95% of all the confidence intervals that can be constructed of size n will contain the unknown true parameter
- A specific interval either will contain or will not contain the true parameter
  - No probability involved in a specific interval

General Formula

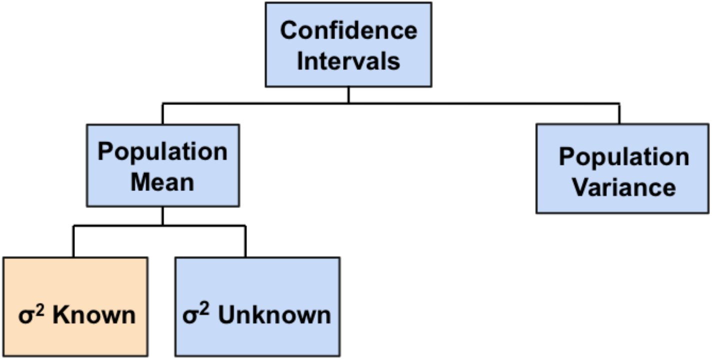
- The general form for all confidence intervals is:

$$\boxed{\hat{\theta} \pm ME}$$

Point Estimate ± Margin of Error

- The value of the margin of error depends on the desired level of confidence

Confidence Intervals



(From normally distributed populations)

Confidence Interval Estimation for the Mean (σ^2 Known)

- Assumptions
  - Population variance σ^2 is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

$$\boxed{\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

(where $z_{\alpha/2}$ is the normal distribution value for a probability of $\alpha/2$ in each tail)

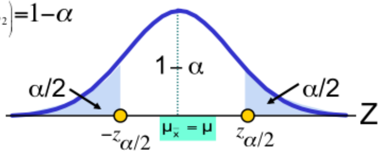
How do we get the interval estimate:

Want: Find a and b such that P(a < μ < b) = 1 - α

- Inside the probability, multiply both sides by (-1) to get P(-b <- μ <- a)
- Then add $\overline{X}$ and divide by $\sigma/\sqrt{n}$ to get

$$P\left(\frac{\overline{X}-b}{\sigma/\sqrt{n}}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<\frac{\overline{X}-a}{\sigma/\sqrt{n}}\right)=1-\alpha$$

Or equivalently $P\left(-z_{\alpha/2}<Z<z_{\alpha/2}\right)=1-\alpha$



where

$$-z_{\alpha/2}=\frac{\overline{X}-b}{\sigma/\sqrt{n}} \quad \text{and} \quad z_{\alpha/2}=\frac{\overline{X}-a}{\sigma/\sqrt{n}}$$

Finally solve for a and b to get

$$a=\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad b=\overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

a = LCL and b = UCL

Confidence Limits

- The confidence interval is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- The endpoints of the interval are

$$UCL = \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$ Upper confidence limit

$$LCL = \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$ Lower confidence limit

Margin of Error

- The confidence interval,

$$\boxed{\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

- Can also be written as $\boxed{\bar{x} \pm ME}$
  where ME is called the margin of error

$$\boxed{ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

- The interval width, w, is equal to twice the margin of error

Reducing the Margin of Error

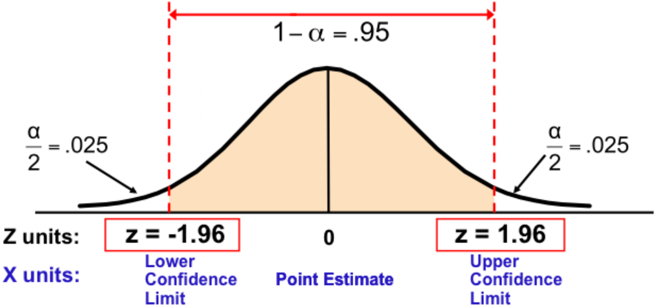
$$\boxed{ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

The margin of error can be reduced if

- the population standard deviation can be reduced ($\sigma \downarrow$)
- The sample size is increased ($n \uparrow$)
- The confidence level is decreased, $(1 - \alpha) \downarrow$

Finding $z_{\alpha/2}$

- Consider a 95% confidence interval:



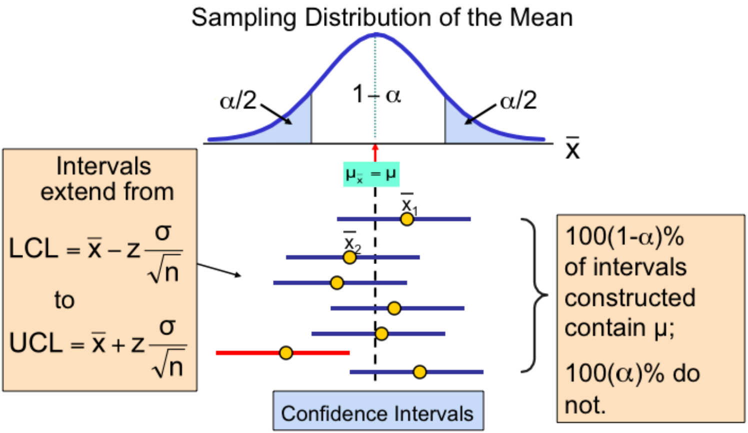
- Find $z_{.025} = \pm 1.96$ from the standard normal distribution table

Common Levels of Confidence

- Commonly used confidence levels are 90%, 95%, 98%, and 99%

Confidence Level	Confidence Coefficient, $1-\alpha$	$Z_{\alpha/2}$ value
80%	.80	1.28
90%	.90	1.645
95%	.95	1.96
98%	.98	2.33
99%	.99	2.58
99.8%	.998	3.08
99.9%	.999	3.27

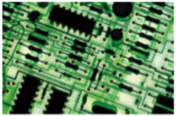
Intervals and Level of Confidence



Example

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.

- Determine a 95% confidence interval for the true mean resistance of the population.



Example

(continued)

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is .35 ohms.
- Solution:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96\,(.35/\sqrt{11})$$

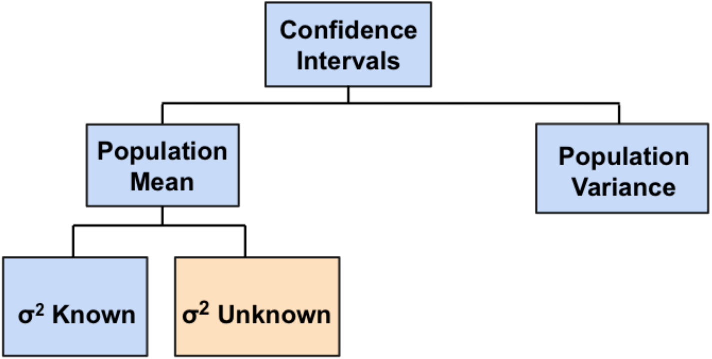
$$= 2.20 \pm .2068$$

$$\boxed{1.9932 < \mu < 2.4068}$$

Interpretation

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean

Confidence Interval Estimation for the Mean (σ^2 Unknown)



(From normally distributed populations)

Student's t Distribution

- Consider a random sample of n observations
  - with mean $\bar{x}$ and standard deviation s
  - from a normally distributed population with mean μ
- Then the variable

$$\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$$

follows the Student's t distribution with (n - 1) degrees of freedom

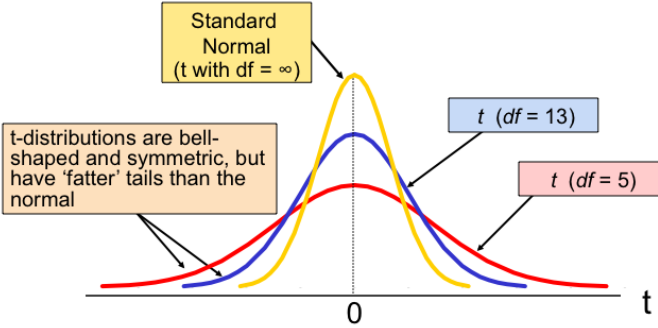
Student's t Distribution

- The t is a family of distributions
- The t value depends on degrees of freedom (d.f.)
  - Number of observations that are free to vary after sample mean has been calculated

$$\boxed{d.f. = n - 1}$$

Student's t Distribution

Note: t ⟶ Z as n increases

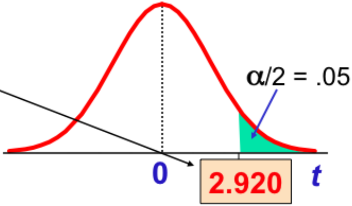


Student's t Table

	Upper Tail Area		
df	.10	**.05**	.025
1	3.078	6.314	12.706
2	1.886	**2.920**	4.303
3	1.638	2.353	3.182

The body of the table contains t values, not probabilities

Let: n = 3
df = n - 1 = 2
α = .10
$\alpha/2$ = .05



t distribution values

With comparison to the Z value

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	Z
.80	1.372	1.325	1.310	1.282
.90	1.812	1.725	1.697	1.645
.95	2.228	2.086	2.042	1.960
.99	3.169	2.845	2.750	2.576

Note: $t \longrightarrow Z$ as n increases

Confidence Interval Estimation for the Mean (σ^2 Unknown)

- If the population standard deviation σ is unknown, we can substitute the sample standard deviation, s
- This introduces extra uncertainty, since s is variable from sample to sample
- So we use the t distribution instead of the normal distribution

Confidence Interval Estimation for the Mean (σ^2 Unknown)

(continued)

- Assumptions
  - Population standard deviation is unknown
  - Population is normally distributed
  - If population is not normal, use large sample
- Use Student's t Distribution
- Confidence Interval Estimate:

$$\boxed{\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}}$$

where $t_{n-1,\alpha/2}$ is the critical value of the t distribution with n-1 d.f. and an area of $\alpha/2$ in each tail: $\boxed{P(t_{n-1} > t_{n-1,\alpha/2}) = \alpha/2}$

Margin of Error

- The confidence interval,

$$\boxed{\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}}$$

- Can also be written as $\boxed{\bar{x} \pm ME}$

  where ME is called the margin of error:

$$\boxed{ME = t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}}$$

Example

A random sample of n = 25 has $\bar{x}$ = 50 and s = 8. Form a 95% confidence interval for μ

- d.f. = n – 1 = 24, so $t_{n-1,\alpha/2} = t_{24,.025} = 2.0639$

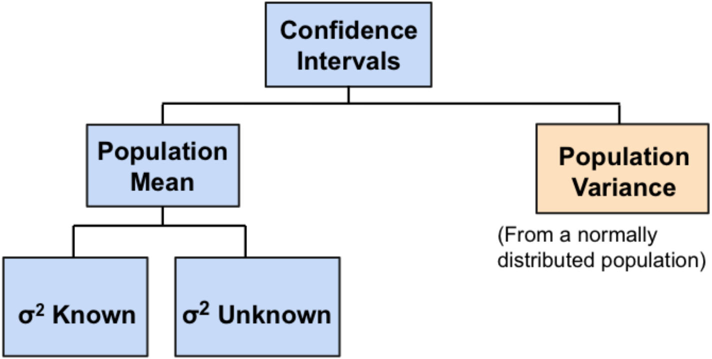
The confidence interval is

$$\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

$$50 \pm (2.0639)\frac{8}{\sqrt{25}}$$

$$46.698 < \mu < 53.302$$

Confidence Interval Estimation for the Variance



Confidence Intervals for the Population Variance

- Goal: Form a confidence interval for the population variance, σ^2
  - The confidence interval is based on the sample variance, s^2
  - Assumed: the population is normally distributed

Confidence Intervals for the Population Variance

(continued)

The random variable

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

follows a chi-square distribution with (n – 1) degrees of freedom

Where the chi-square value $\chi^2_{n-1,\alpha}$ denotes the number for which

$$P(\chi^2_{n-1} > \chi^2_{n-1,\alpha}) = \alpha$$

Confidence Intervals for the Population Variance

(continued)

The 100(1 - α)% confidence interval for the population variance is given by

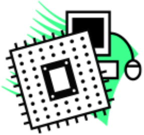
$$\boxed{LCL = \frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}}}$$

$$\boxed{UCL = \frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}}}$$

Example

You are testing the speed of a batch of computer processors. You collect the following data (in Mhz):

Sample size	**17**
Sample mean	**3004**
Sample std dev	**74**



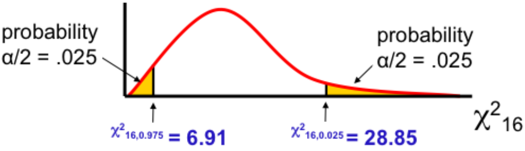
Assume the population is normal.
Determine the 95% confidence interval for σ^2

Finding the Chi-square Values

- n = 17 so the chi-square distribution has (n – 1) = 16 degrees of freedom
- α = 0.05, so use the the chi-square values with area 0.025 in each tail:

$$\chi^2_{n-1,\,\alpha/2} = \chi^2_{16,\,0.025} = 28.85$$

$$\chi^2_{n-1,\,1-\alpha/2} = \chi^2_{16,\,0.975} = 6.91$$



Calculating the Confidence Limits

- The 95% confidence interval is

$$\frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}}$$

$$\frac{(17-1)(74)^2}{28.85} < \sigma^2 < \frac{(17-1)(74)^2}{6.91}$$

$$\boxed{3037 < \sigma^2 < 12680}$$

Converting to standard deviation, we are 95% confident that the population standard deviation of CPU speed is between 55.1 and 112.6 Mhz

Sample-Size Determination: Large Populations

Large Populations

For the Mean

(Known population variance)

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Margin of Error (sampling error)

$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Sample-Size Determination: Large Populations

(continued)

Large Populations

For the Mean

(Known population variance)

$$ME = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Now solve for n to get

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{ME^2}$$

Sample-Size Determination

(continued)

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence (1 - α), which determines the $z_{\alpha/2}$ value
  - The acceptable margin of error (sampling error), ME
  - The population standard deviation, σ