06. Estimation

Adapted From :

Probability & Statistics for Engineers & Scientists, 9th Ed.

Walpole/Myers/Myers/Ye (c)2010

Introduction to Business Statistics, 5e

Kvanli/Guynes/Pavur (c)2000

South-Western College Publishing

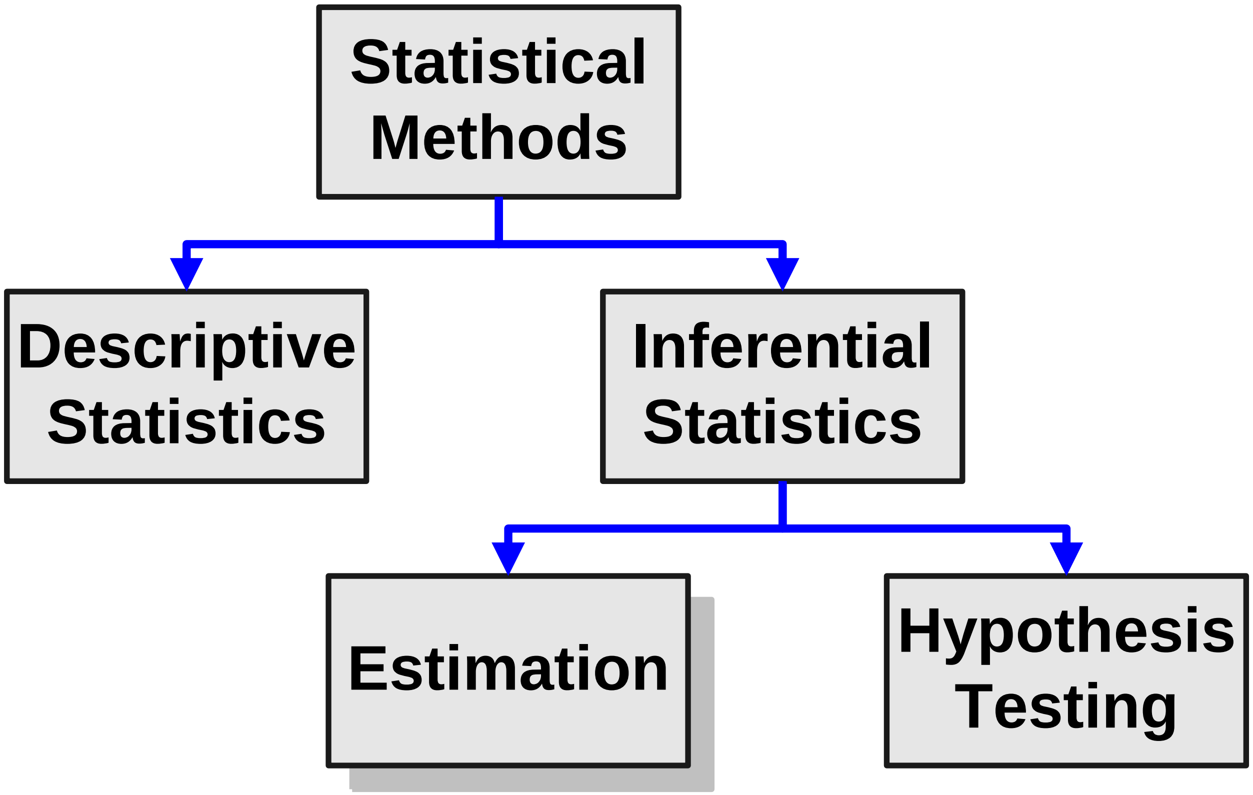
Statistics for Managers

Using Microsoft® Excel 4th Edition

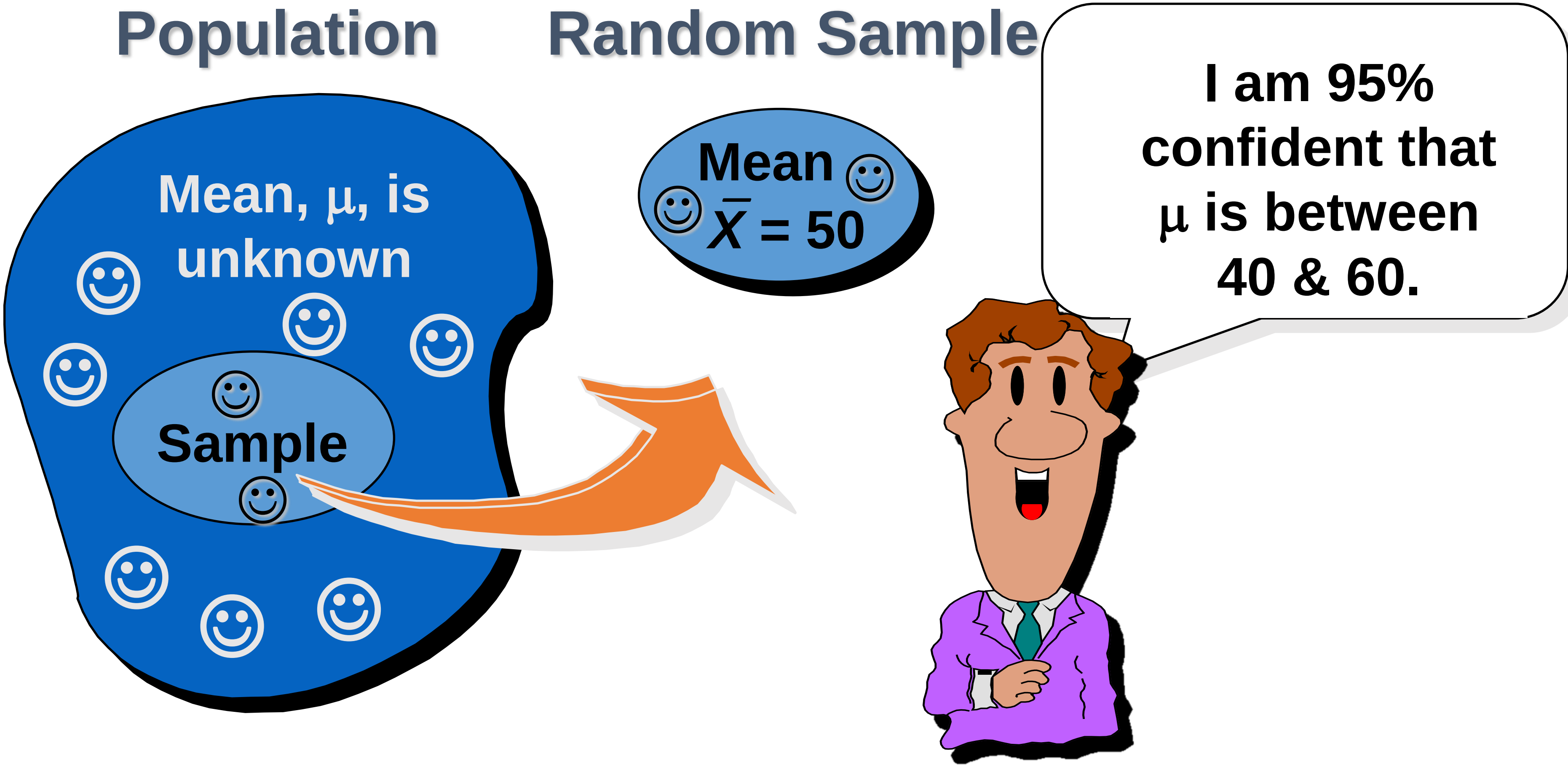
Thinking Challenge

Suppose you're interested in the average amount of money that students in this class (the population) have in their possession. How would you find out?

Statistical Method



Estimation Process



Population Parameter are estimated

Estimate population parameter...		with sample statistic
Mean	μ	$\bar{x}$
Proportion	p	$\hat{p}$
Variance	σ^2	s^2
Differences	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$

Unbiased Estimators

- A statistic Θ_{hat} is an unbiased estimator of the parameter θ if

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta$$

  - Note that in calculating S^2, the reason we divide by n-1 rather than n is so that S^2 will be an unbiased estimator of σ^2.

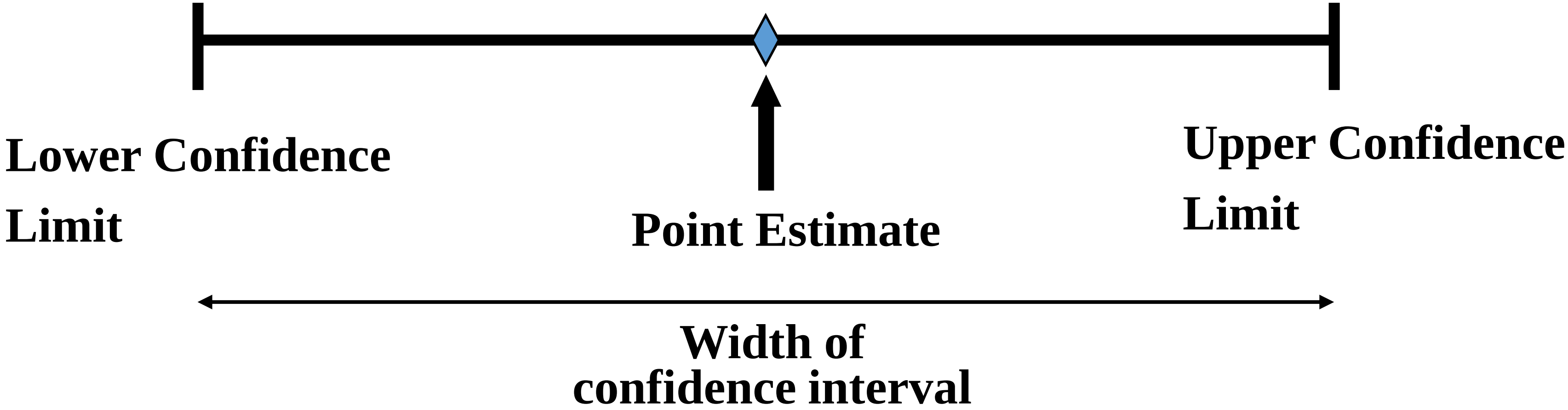
- Of all unbiased estimators of a parameter θ, the one with the smallest variance is called the most efficient estimator of θ.
  - X_{bar} is the most efficient estimator of μ. And,

$$\sigma / \sqrt{n}$$

is called the standard error of the estimator X_{bar}

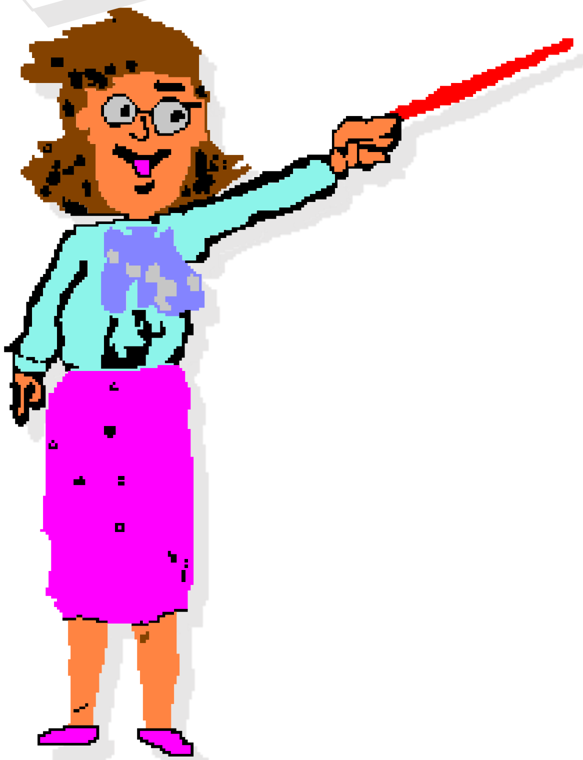
Point Estimates

- A **point estimate** is a single number. For the population mean (and population standard deviation), a point estimate is the sample mean (and sample standard deviation).
- A confidence interval provides additional information about variability



Confidence Limit for population mean

Parameter =
Statistic ± Error



(1) $\mu = \bar{X} \pm Error$

(2) $Error = \bar{X} - \mu$ or $\bar{X} + \mu$

(3) $Z = \dfrac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \dfrac{Error}{\sigma_{\bar{x}}}$

(4) $Error = Z\sigma_{\bar{x}}$

(5) $\mu = \bar{X} \pm Z\sigma_{\bar{x}}$

Confidence Interval Estimates

- A confidence interval gives a range estimate of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on all the observations from 1 sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - Ex. 95% confidence, 99% confidence
    - Can never be 100% confident

Confidence Interval Estimates

- The general formula for all confidence intervals is:

Point Estimate ± (Critical Value) (Standard Error)

Confidence Level

- Confidence Level
  - Confidence in which the interval will contain the unknown population parameter
- A percentage (less than 100%)

Confidence Level

- Suppose confidence level = 95%
- Also written $(1 - \alpha) = .95$
- A relative frequency interpretation:
  - In the long run, 95% of all the confidence intervals that can be constructed will contain the unknown true parameter
  - α is probability that parameter is *not* within interval
- A specific interval either will contain or will not contain the true parameter

Confidence Interval for μ (σ Known)

Assumptions

- Population standard deviation σ is known
- Population is normally distributed
- If population is not normal, use large sample

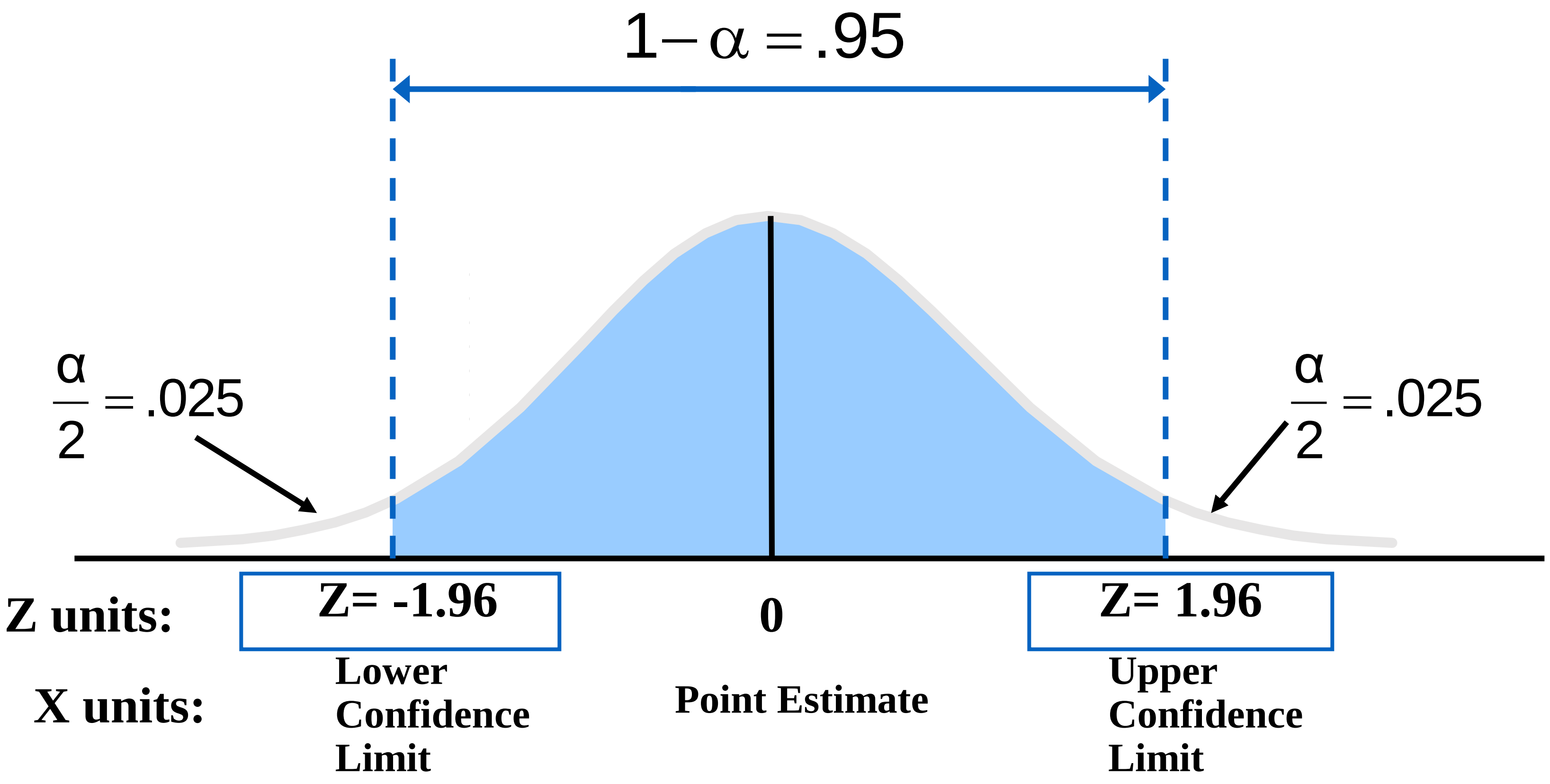
Confidence interval estimate:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

(where Z is the standardized normal distribution critical value for a probability of α/2 in each tail)

Finding the Critical Value, Z

Consider a 95% confidence interval:



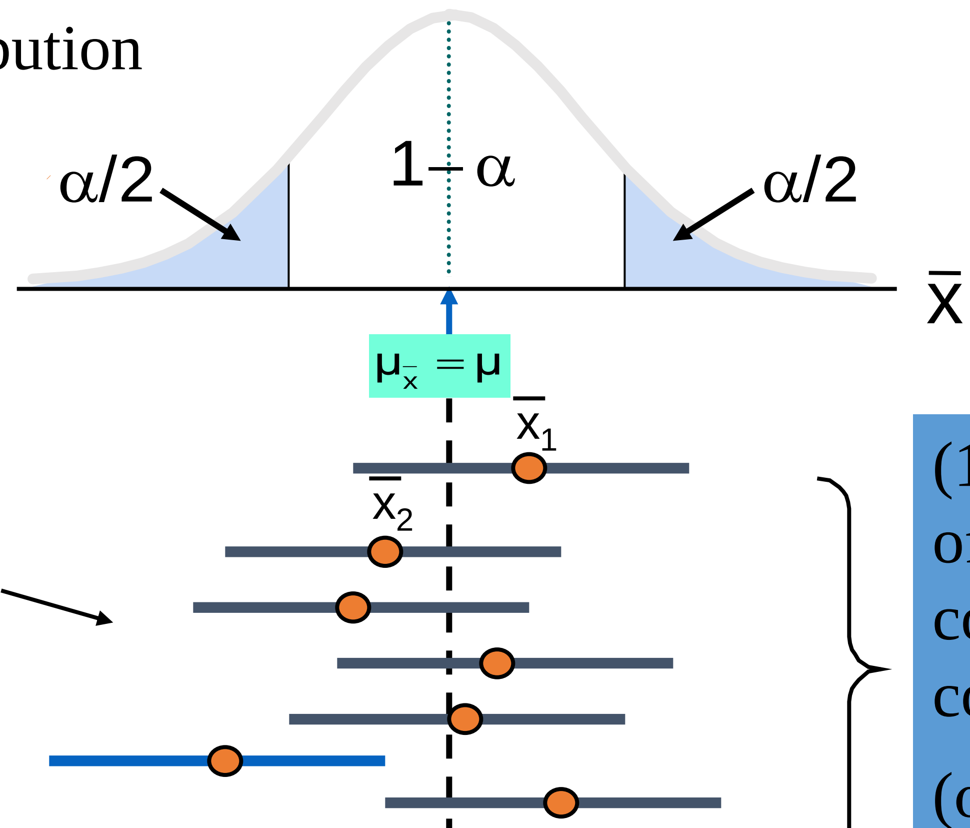
Finding the Critical Value, $Z_{\alpha/2}$

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

Confidence Level	***Confidence Coefficient***	***$Z_{\alpha/2}$ 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

Sampling Distribution of the Mean



Confidence Intervals

Intervals extend from

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

to

$$\overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

(1-α)x100% of intervals constructed contain μ;

(α)x100% do not.

Confidence Interval for μ (σ Known) 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 .35 ohms.
- Determine a 95% confidence interval for the true mean resistance of the population.

Confidence Interval for μ (σ Known) Example

$$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

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

$$= 2.20 \pm .2068$$

$$(1.9932\,,\ \ 2.4068)$$

- 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

One-Sided Confidence Bounds

- Sometimes, instead of a confidence interval, we're only interested in a bound in a single direction.
  - In this case, a $(1-\alpha)100\%$ confidence bound uses z_α in the appropriate direction rather $z_{\alpha/2}$ in either direction.
  - So the $(1-\alpha)100\%$ confidence bound would be either

$$\mu > \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \qquad or \qquad \mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

  depending upon the direction of interest.
  - Look at page 278, Example 9.4

Confidence Interval for μ (σ Unknown)

- If σ is unknown, the calculations are the same, using $t_{\alpha/2}$ with $\nu = n-1$ degrees of freedom, instead of $z_{\alpha/2}$, and using s calculated from the sample rather than σ.
  - As before, use of the t-distribution requires that the original population be normally distributed.
- The standard error of the estimate (i.e., the standard deviation of the estimator) in this case is

$$s/\sqrt{n}$$

- Note that if σ is unknown, but $n \geq 30$, s is still used instead of σ, but the normal distribution is used instead of the t-distribution
  - This is called a large sample confidence interval.

Confidence Interval for μ (σ Unknown)

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:

$$\overline{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

(where t is the critical value of the t distribution with n-1 d.f. and an area of α/2 in each tail)

Student’s t Distribution

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

$$d.f. = n - 1$$

Degrees of Freedom

Idea: Number of observations that are free to vary after sample mean has been calculated

Example: Suppose the mean of 3 numbers is 8.0

- Let $X_1 = 7$
- Let $X_2 = 8$
- What is X_3?

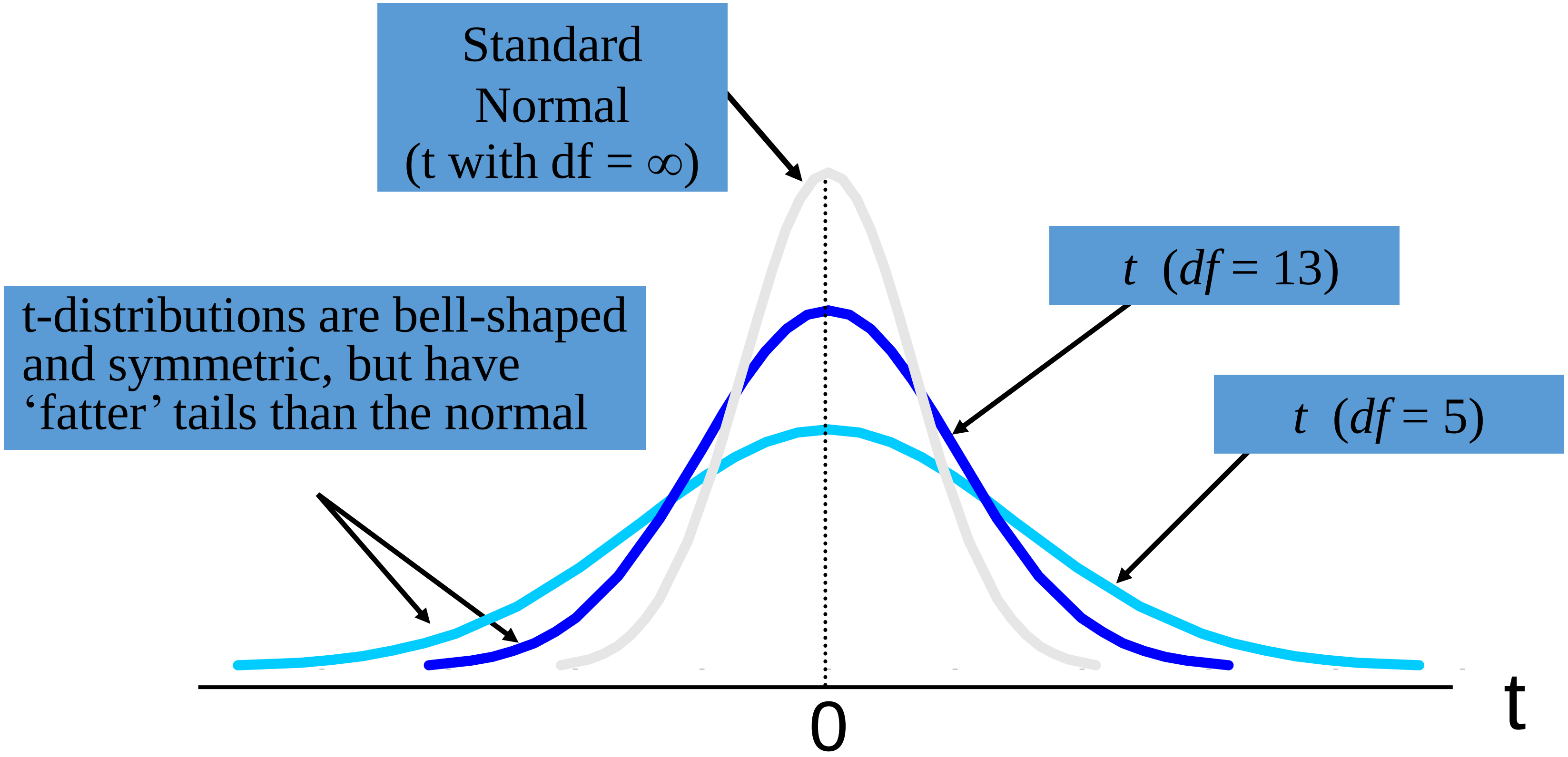
If the mean of these three values is 8.0,
then X_3 must be 9
(i.e., X_3 is not free to vary)

Here, n = 3, so degrees of freedom $= n - 1 = 3 - 1 = 2$

(2 values can be any numbers, but the third is not free to vary for a given mean)

Student's t Distribution

Note: $t \longrightarrow Z$ as n increases

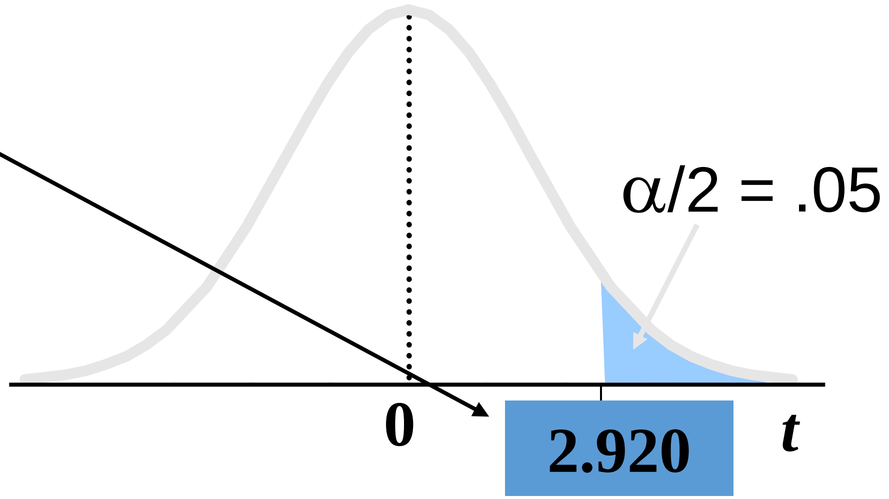


Student's t Table

	Upper Tail Area		
df	.25	.10	**.05**
1	1.000	3.078	6.314
2	0.817	1.886	**2.920**
3	0.765	1.638	2.353

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

The body of the table contains t values, not probabilities



Confidence Interval for μ (σ Unknown) Example

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

- d.f. = n – 1 = 24, so
- The confidence interval is

$$\overline{X} \pm t_{\alpha/2,\,\text{n-1}} \frac{\text{S}}{\sqrt{\text{n}}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

$$(46.698\ ,\ 53.302)$$

Difference Between Two Means

- If x_{bar1} and x_{bar2} are the means of independent random samples of size n_1 and n_2, drawn from two populations with variances σ_1^2 and σ_2^2, then, if $z_{\alpha/2}$ is the z-value with area $\alpha/2$ to the right of it, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

  - Requires a reasonable sample size or a normal-like population for the central limit theorem to apply.
  - It is important that the two samples be <u>randomly selected</u> (and <u>independent</u> of each other).
  - Can be used if σ unknown as long as sample sizes are large.

Example

An experiment was conducted in which two types of engines, A and B, were compared. Gas mileage, in miles per gallon, was measured. Fifty experiments were conducted using engine type A and 75 experiments were done for engine type B. The gasoline used and other conditions were held constant. The average gas mileage for engine A was 36 miles per gallon and the average for machine B was 42 miles per gallon. Find a 96% confidence interval on $\mu_B - \mu_A$, where μ_A and μ_B are population mean gas mileage for machine A and B, respectively. Assume that the population standard deviations are 6 and 8 for machine A and B, resepectively.

Estimating a Proportion

- An estimator of p in a binomial experiment is P_{hat} = X / n , where X is a binomial random variable indicating the number of successes in n trials. The sample proportion, p_{hat} = x / n is a point estimator of p.
- What is the mean and variance of a binomial random variable X?
- To find a confidence interval for p, first find the mean and variance of P_{hat}:

$$\mu_{\hat{P}} = E(\hat{P}) = E\left[\frac{X}{n}\right] = \frac{np}{n} = p$$

$$\sigma^2_{\hat{P}} = \sigma^2_{X/n} = \frac{\sigma^2_X}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

Confidence Interval for a Proportion

- If p_{hat} is the proportion of successes in a random sample of size n, and $q_{hat} = 1 - p_{hat}$, then a $(1-\alpha)100\%$ confidence interval for the binomial parameter p is given by

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- Note that n must be reasonably large and p not too close to 0 or 1.
- Rule of thumb: both np and nq must be ≥ 5.
- This also works if the binomial is used to approximate the hypergeometric distribution (when n is small relative to N).

Confidence Intervals for the Population Proportion, Example

A random sample of 100 people shows that 25 have opened IRA's this year. Form a 95% confidence interval for the true proportion of the population who have opened IRA's.

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$= 25/100 \pm 1.96\sqrt{.25(.75)/100}$$

$$= .25 \pm 1.96\ (.0433)$$

$$(0.1651\ ,\ 0.3349)$$

Confidence Intervals for the Population Proportion, Example

- We are 95% confident that the true percentage of left-handers in the population is between 16.51% and 33.49%. Although the interval from .1651 to .3349 may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

The Difference of Two Proportions

- If p_{1hat} and p_{2hat} are the proportion of successes in random samples of size n_1 and n_2, an approximate $(1-\alpha)100\%$ confidence interval for the difference of two binomial parameters is

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2$$

$$< (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Determining Sample Size

- The required sample size can be found to reach a desired margin of error (e) with a specified level of confidence (1 - α)
- The margin of error is also called sampling error
  - the amount of imprecision in the estimate of the population parameter
  - the amount added and subtracted to the point estimate to form the confidence interval

Determining Sample Size

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence (1 - α), which determines the critical Z value
  - The acceptable sampling error (margin of error), e
  - The standard deviation, σ

$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Now solve for n to get

$$n = \left(\frac{Z_{\alpha/2}\,\sigma}{e} \right)^2$$

Finding Sample Sizes For Estimating μ

(1) $$Z_{\alpha/2} = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{Error}{\sigma_{\bar{x}}}$$

(2) $$Error = Z_{\alpha/2}\sigma_{\bar{x}} = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

(3) $$n = \frac{(Z_{\alpha/2})^2\sigma^2}{Error^2}$$

Determining Sample Size

If σ = 45, what sample size is needed to estimate the mean within ± 5 with 90% confidence?

$$n = \frac{(Z_{\alpha/2})^2\,\sigma^2}{e^2} = \frac{(1.645)^2(45)^2}{5^2} = 219.19$$

So the required sample size is **n = 220**

Determining Sample Size

- If unknown, σ can be estimated when using the required sample size formula
  - Use a value for σ that is expected to be at least as large as the true σ
  - Select a pilot sample and estimate σ with the sample standard deviation, S

Error of Estimate for a Proportion & Determining Sample Size

- If p_{hat} is used to estimate p, we can be $(1-\alpha)100\%$ confident that the error of estimate will not exceed

$$z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- Then, to achieve an error of e, the sample size must be at least

$$\lceil n \rceil = \frac{z_{\alpha/2}^{2}\hat{p}\hat{q}}{e^{2}}$$

- If p_{hat} is unknown, we can be <u>at least</u> $100(1-\alpha)\%$ confident using an upper limit on the sample size of

$$\lceil n \rceil = \frac{z_{\alpha/2}^{2}}{4e^{2}}$$

Determining Sample Size

- How large a sample would be necessary to estimate the true proportion defective in a large population within ±3%, with 95% confidence?
- (Assume a pilot sample yields $p_{hat} = .12$)

Determining Sample Size

Solution:

For 95% confidence, use Z = 1.96

e = .03

p = .12, so use this to estimate π

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2} = \frac{(1.96)^2(.12)(1-.12)}{(.03)^2} = 450.74$$

So use n = 451

Thank You

"We trust in GOD, all others must bring data"