Ch 8: Confidence Intervals

Often, a desired population parameter is unknown.

In that case, we conduct a survey or take a sample to estimate the unknown parameter.

There are two ways we can do this:

- Point Estimate
- Confidence Interval

Point Estimate – we approximate the unknown parameter with a single value computed from sample data.

Inferential Statistics = using sample data to make generalizations (i.e. to "infer") thing about the pop.

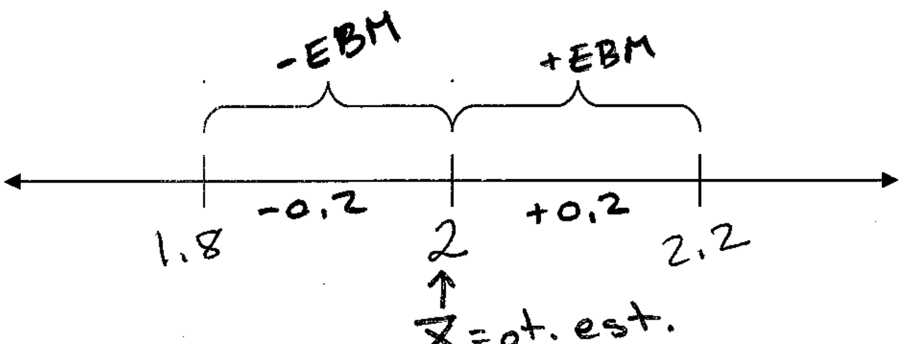
Notation

	Parameter		Point Estimate
μ	pop. mean	$\bar{x}$	sample mean
σ	pop. st. dev.	s_x	sample st. dev.
p	pop proportion	p'	sample proportion
		$\hat{p}$	

Confidence Intervals

A confidence interval (CI) gives a range of values to estimate the parameter, along with a certain level of confidence that the parameter lies within the range.

Form of the Confidence Interval:



$\bar{x}$ - EBM $\quad$ $\bar{x}$ + EBM

*(pt. est – error bound , pt. est. + error bound)

CI = (1.8, 2.2)

So we are "reasonably sure"

$1.8 < \mu < 2.2$

2 Q's:
- How do we find EB?
- What do we mean by reasonably sure?

Math 10 Handout Chapter 8 **Confidence Intervals**

Suppose we are interested in the following question:

- What is the mean number μ of hours of sleep De Anza students got last night?

We take our class as a sample. Suppose the sample average is 7.5 hours per night. Then, we say that $\bar{x} = 7.5$. The value 7.5 is called a **point estimate** of μ.

Using the CLT, we know $\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$.

Another way to give the mean estimate is in the form of a **confidence interval.** A confidence interval has the from ($\bar{x}$ − margin of error, $\bar{x}$ + margin or error).

- Example: Suppose we want to estimate the average number of Giants home games a fan attends per season. A sample of 30 local fans is taken and the average number of games attended is calculated from the sample. We calculate that $\bar{x} = 10$. Suppose we know that the population standard deviation of games attended is $\sigma = 2$. Construct a 95% confidence interval for the true mean μ number of games attended.

$n = 30$
$\bar{x} = 10$
$\sigma = 2$

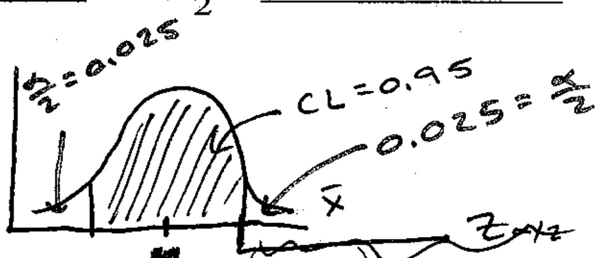
Confidence level = CL = 0.95

Associated with the CL is something called the α **-value**. The α -value is the probability that our confidence interval will **not** contain the true population mean. Thus:

$$\alpha = 1 - CL$$

So: α = 0.05 Thus, $\frac{\alpha}{2}$ = 0.025

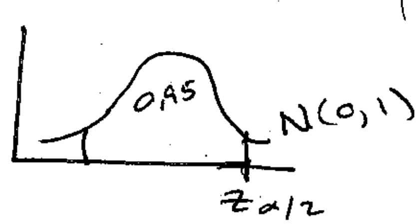
Graphically, we have:



Recall z-scores $z = \frac{\bar{x} - \mu_x}{\sigma_x}$. Since $\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$, a bit of algebra (that we won't do) tells us z-scores for $\bar{X}$ have the normal distribution Z~N(0,1). standard normal distr.

The confidence interval has the from ($\bar{x}$ − EBM, $\bar{x}$ + EBM) where EBM = $z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

The value $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ -critical value for the standard normal distribution. (Huh?) In other words, the area to the **right** of $z_{\frac{\alpha}{2}}$ is $\frac{\alpha}{2}$. We find this by using invNorm $\left(CL + \frac{\alpha}{2}, 0, 1\right)$ on our calculator.



$CL + \frac{\alpha}{2} = 0.95 + 0.025$

0.025

So $z_{\frac{\alpha}{2}} = z_{0.025}$ = invNorm(0.975, 0, 1) = 1.96

So EB = 0.72 and the confidence interval is (9.28, 10.72)

$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96\left(\frac{2}{\sqrt{30}}\right)$ $\quad (\bar{x} - EB, \bar{x} + EB)$

Let's draw a picture:

We are 95% confident that the true population mean is between 9.28 and 10.72.

* In other words, we are 95% confident that the average number of Giants home games a fan attends per season is between 9.28 and 10.72.

Here's another (slightly easier) way to calculate of this:

Even though $\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$, to create the confidence interval, we will use the distribution

$\bar{X} \sim N\left(\bar{x}, \frac{\sigma_x}{\sqrt{n}}\right)$ (with $\bar{x}$ instead on μ_x since we don't know μ_x). To find the 95% confidence interval, we need to find the values that correspond to the percentiles for $\frac{\alpha}{2}$ and CL+$\frac{\alpha}{2}$.

In this example, $\bar{X} \sim N(10, \frac{2}{\sqrt{30}})$, CL = 0.95 and $\frac{\alpha}{2}$ = 0.025.

So we use the calculator commands invNorm(0.025, 10, $\frac{2}{\sqrt{30}}$) and invNorm(0.975, 10, $\frac{2}{\sqrt{30}}$)

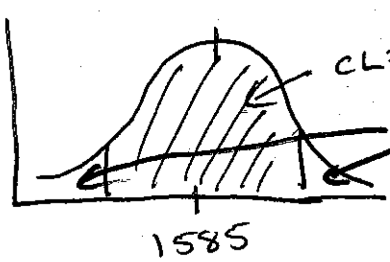
to get the same confidence interval (9.28, 10.72) and the picture:

- Example: Suppose we are interested in finding the mean price of an HDTV that is currently on the market. A random sample of 34 HDTV prices is taken that gives a sample mean of \$1585.00. If it is known that the population standard deviation of HDTV prices is \$390.00, find the 95% confidence interval for the true mean price of an HDTV.

$\bar{X} \sim N(1585, 390/\sqrt{34})$

CI = (1453.9, 1716.1)

We are 95% conf. that the true avg. price of an HDTV is between \$1453.90 and \$1716.10.



$\frac{\alpha}{2} = \frac{1-0.95}{2} = 0.025$

How to Construct a Confidence Interval

I. Conf. Int. for Means

A. σ known (pop. st. dev.)

We will use the distribution for averages:

$$\overline{X} \sim N\left(\overline{x}, \frac{\sigma}{\sqrt{n}}\right)$$

Steps:

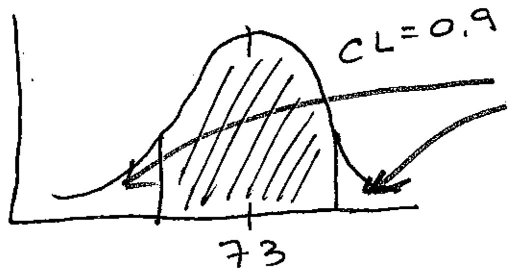
1. Write the distribution for the problem
2. Draw a graph
3. Find the Confidence Interval using the Calculator ← shortcut
4. Interpret your interval in terms of the problem

Example: Statistics exam scores are normally distributed with a population standard deviation of 20. Suppose we do a random sample of 36 students' exams and find that the average score is 73. Find a 90% confidence interval for the population mean exam score.

$\sigma = 20, \quad n = 36, \quad \bar{x} = 73, \quad CL = 0.9$

① $\bar{X} \sim N\left(73, \frac{20}{\sqrt{36}}\right) = N\left(73, \frac{10}{3}\right)$

②



$\frac{\alpha}{2} = \frac{0.1}{2} = 0.05$

③ Calculator Instructions

STAT → TESTS → ZInterval → STATS

ZInterval($\sigma, \bar{x}, n$, CL)

ZInterval(20, 73, 36, 0.9)

CI = (67.517, 78.483)

[or invNorm $\left(0.05, 73, \frac{10}{3}\right) = 67.517$

invNorm $(0.95, 73, 10/3) = 78.483$

④ We are 90% confident that the true avg. exam score is between 67.517 and 78.483

Math 10 Handout Chapter 8 **Student-t Distribution & Confidence Intervals for Proportions**

Consider the previous problem:

- Test scores for statistics classes are normally distributed with unknown population mean but <u>population</u> standard deviation 3. A sample of 36 scores is taken that gives a sample mean of 68. Find a 90% confidence interval for the true population mean of test scores.

To solve this problem, we used: $EB = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

If the population standard deviation σ is **<u>not</u>** given, we will use the sample standard deviation s_x but will not use the $z_{\frac{\alpha}{2}}$-score. Instead, we will use a ***t*-score** from a different distribution called the **Student-t distribution.** The Student-t distribution is denoted by $T \sim t_{df}$ where $df = n - 1$.

<u>5 Facts about the Student-t distribution:</u>

1. The graph of a Student-t distribution is similar to a normal curve.
2. Student-t tails have more probability under them because their spread is a little greater.
3. The Student-t distribution can only be used if the underlying population distribution is close to normal (or at least large and bell shaped). Also, we'll only use the Student-t distribution if the population mean and standard deviation are unknown.
4. $EB = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$
5. The Student-t distribution was discovered by a beer maker named William Gosset who worked for the Guinness brewery in Dublin, Ireland.

EXAMPLE

- Suppose we do a study on acupuncture to determine how effective it is in relieving pain. We measure sensory rates for 15 subjects. The results are given below. Use the sample data to construct a 95% confidence interval for the population mean sensory rate. Assume that the distribution of sensory rates is normal.

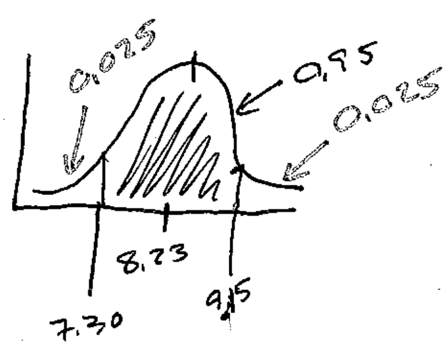
8.6, 9.4, 7.9, 6.8, 8.3, 7.3, 9.2, 9.6, 8.7, 11.4, 10.3, 5.4, 8.1, 5.5, 6.9

$n = 15$

$CL = 0.95$

Find 1st: $\bar{x} = 8.23$ *Use 1-var stats

$s_x = 1.67$



$CI = (7.30, 9.15)$

We are 95% sure that the true pop. mean sensory rate is between 7.30 & 9.15.

Example: From a stack of IEEE Spectrum magazines, announcements for 84 upcoming engineering conferences were randomly picked. The average length of the conferences was 3.94 days, with a standard deviation of 1.28 days. Assume the underlying population is normal.

$n=84$

$\bar{x}=3.94$

$s_x=1.28$

a. Define the Random Variables X and $\overline{X}$, in words.

X = length of 1 conference

$\bar{X}$ = avg. length of 84 conferences

b. Which distribution should you use for this problem? Explain your choice.

Student t-distribution

$T \sim t_{83}$ because we don't know σ

c. Construct a 95% confidence interval for the population average length of engineering conferences.

i. State the confidence interval.

ii. Sketch the graph.

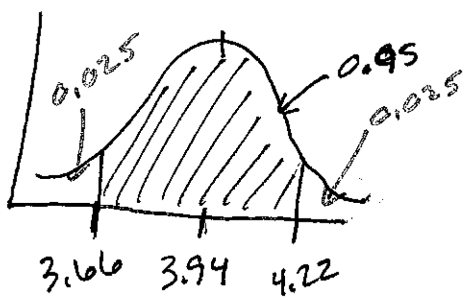
iii. Calculate the error bound.

iv. Write a statement interpreting your interval.

STAT → TESTS → TInterval

i (3.66, 4.22)

ii



$\alpha = 1 - CL$
$= 0.05$

$\frac{\alpha}{2} = 0.025$

iii $4.22 - 3.94 = 0.28$

ok: $\frac{4.22 - 3.66}{2} = \boxed{0.28}$

iv We are 95% confindent that the true pop. avg. is b/t 3.66 + 4.22

Sometimes we are not happy with the confidence interval we obtain. We may want a smaller error bound or a higher confidence level.

We have control over 2 things:

n	sample size
CL	confidence level

What happens if we change n or CL?

Bigger n ⇒ either bigger CL or smaller EB
Smaller n ⇒ either smaller CL or bigger EB
Bigger CL ⇒ either bigger n or ~~smaller~~ bigger EB
Smaller CL ⇒ either smaller n or smaller EB

Summary

If n is increased, the error bound gets smaller

If CL is increased, the error bound gets larger

To Decrease the error bound:

- Increase n (i.e. take a larger sample)

or

- Decrease the confidence level

Ch 8.
II. Confidence Intervals for Proportions

The underlying distribution is Binomial

X = # of successes out of n trials $X \sim B(n, p)$

$\frac{X}{n} = P'$ random variable for proportions

To construct the confidence interval, use

$$P' \sim N\left(p', \sqrt{\frac{p'q'}{n}}\right)$$

p' = sample proportion of successes
q' = sample proportion of failures
$p' + q' = 1$

ex) Will you vote for Hillary?
Do you own a PC?

X = number of successes
p' = proportion of successes

Example: I am interested in the proportion of De Anza students who feel that the economic outlook for California will improve in the next two years.

\# of successes

x = 3 n = 15

← proportion of successes

p' = x/n = $\frac{3}{15} = 0.2$

X = number of DA students (out of 15) who think the econ. in CA will improve in 2 yrs

P' = proportion of DA students who think the econ. in CA will improve in 2 yrs

$P' \sim N\left(0.2, \sqrt{\frac{(0.2)(0.8)}{15}}\right)$

$p' = 0.2$

$q' = 1 - 0.2 = 0.8$

Construct the confidence interval: for 90% CL

STAT → TESTS → A: 1-PropZInt

CI = (0.0301, 0.3699)

Interpret your interval

We are 90% confident that the actual percent of DA student that think the econ. will improve is between 3.01% and 36.99%

8.8 Practice 4: Confidence Intervals for Proportions[9]

8.8.1 Student Learning Outcomes

- The student will calculate confidence intervals for proportions.

8.8.2 Given

The Ice Chalet offers dozens of different beginning ice-skating classes. All of the class names are put into a bucket. The 5 P.M., Monday night, ages 8 - 12, beginning ice-skating class was picked. In that class were 64 girls and 16 boys. Suppose that we are interested in the true proportion of girls, ages 8 - 12, in all beginning ice-skating classes at the Ice Chalet. Assume that the children in the selected class is a random sample of the population.

8.8.3 Estimated Distribution

Exercise 8.8.1
What is being counted? number of girls in the class

Exercise 8.8.2 *(Solution on p. 372.)*
In words, define the Random Variable X. $X =$ number of girls (out of 80) in the class

Exercise 8.8.3 *(Solution on p. 372.)*
Calculate the following:

a. $x = 64$
b. $n = 80$
c. $p' = \frac{64}{80} = 0.8$

Exercise 8.8.4 *(Solution on p. 372.)*
State the estimated distribution of X. $X \sim$

Exercise 8.8.5 *(Solution on p. 372.)*
Define a new Random Variable P'. What is p' estimating? overall prop. of girls in all the classes

Exercise 8.8.6 *(Solution on p. 372.)*
In words, define the Random Variable P'. $P' =$ proportion of girls in the classes

Exercise 8.8.7
State the estimated distribution of P'. $P' \sim$

8.8.4 Explaining the Confidence Interval

Construct a 92% Confidence Interval for the true proportion of girls in the age 8 - 12 beginning ice-skating classes at the Ice Chalet.

Exercise 8.8.8 *(Solution on p. 372.)*
How much area is in both tails (combined)? $\alpha = 1 - 0.92 = 0.08$

Exercise 8.8.9 *(Solution on p. 372.)*
How much area is in each tail? $\frac{\alpha}{2} = 0.04$

Exercise 8.8.10 *(Solution on p. 372.)*
Calculate the following:

[9] This content is available online at <http://cnx.org/content/m16968/1.13/>.

a. lower limit = 0.72
b. upper limit = 0.88
c. error bound = 0.08

Exercise 8.8.11 *(Solution on p. 372.)*

The 92% Confidence Interval is:

Exercise 8.8.12 1-PropZInt: (0.72, 0.88)

Fill in the blanks on the graph with the areas, upper and lower limits of the Confidence Interval, and the sample proportion.

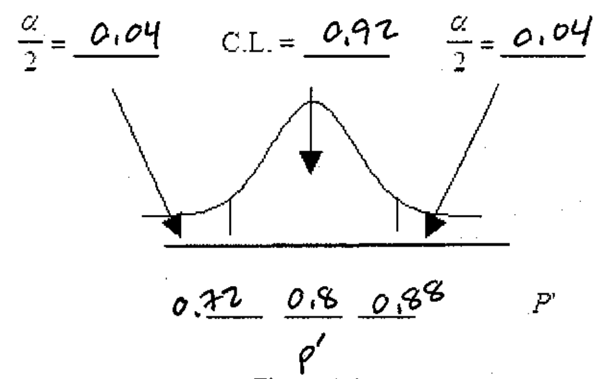


Figure 8.4

Exercise 8.8.13

In one complete sentence, explain what the interval means.

8.8.5 Discussion Questions

Exercise 8.8.14

Using the same p' and level of confidence, suppose that n were increased to 100. Would the error bound become larger or smaller? How do you know?

Exercise 8.8.15

Using the same p' and $n = 80$, how would the error bound change if the confidence level were increased to 98%? Why?

Exercise 8.8.16

If you decreased the allowable error bound, why would the minimum sample size increase (keeping the same level of confidence)?

Calculating the Sample Size n

An organization that does a survey usually has in mind what confidence level and error bound they would like to use. They then do a survey based on those parameters, but they need to determine the sample size to use.

To do this we need to understand how the Error Bound is calculated.

In general, the error bound is calculated using the following general formula:

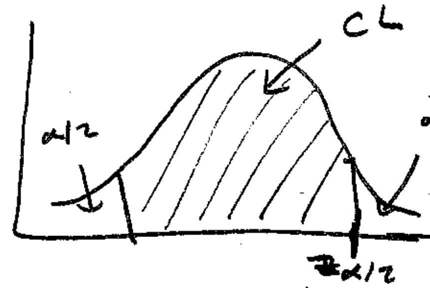
$$EB = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p'q'}{n}} \quad \text{for proportions}$$

We need to solve this equation for n:

$$\left(\frac{EB}{z_{\alpha/2}}\right)^2 = \sqrt{\frac{p'q'}{n}}^{\,2}$$

$$n \cdot \frac{(EB)^2}{(z_{\alpha/2})^2} = \frac{p'q'}{\cancel{n}} \cdot \cancel{n}$$

$$\boxed{n = \frac{p'q'(z_{\alpha/2})^2}{(EB)^2}}$$



$$z_{\alpha/2} = \text{invNorm}\left(CL + \frac{\alpha}{2}, 0, 1\right)$$

Since we don't know p', we use $p' = q' = 0.5$

Example: The Field Corporation wants to determine the percent of California voters who believe that the governor is doing a good job. How many voters should they survey in order to be 95% confident that the estimated sample proportion is within 3% of the true population proportion?

Find n if $CL = 0.95$

$EB = 0.03$

$$n = \frac{(z_{\alpha/2})^2 \cdot (0.5)^2}{(0.03)^2}$$

$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$z_{0.025} = \text{invNorm}(\underbrace{0.975}_{\substack{CL + \frac{\alpha}{2} \\ = 0.95 + 0.025}}, 0, 1) = 1.96$

$$n = \frac{(1.96)^2 \cdot (0.5)^2}{(0.03)^2} = 1067.1$$

*Always round up

We need to survey at least 1,068 people.

Example: Calculating Survey Sample Size n

Insurance companies are interested in knowing the population percent of drivers who always buckle up before riding in a car.

1. When designing a study to determine this population proportion, what is the minimum number you would need to survey to be 95% confident that the population proportion is estimated to within 0.03?

   At least 1068 people

2. If it was later determined that it was important to be more than 95% confident but have the same error bound and a new survey was commissioned, how would that affect the minimum number you would need to survey? Why?

3. Suppose you wanted to be 99% confident with the same error bound of 0.03? What would be the minimum number you would need to survey this time?

$CL = 0.99$

$\alpha = 1 - 0.99$

$= 0.01$

$\frac{\alpha}{2} = 0.005$

$z_{0.005} = invNorm(0.99 + 0.005, 0, 1)$

$= invNorm(0.995, 0, 1) = 2.576$

$$n = \frac{(2.576)^2 (0.5)^2}{(0.03)^2}$$

$\boxed{n = 1843}$

4. Suppose you were OK with the 95% confidence level, but you wanted to decrease the error bound to 0.01. What would be the minimum number you would need to survey in this case?

$CL = 0.95$

$z_{0.025} = 1.96$

$$n = \frac{(1.96)^2 (0.5)^2}{(0.01)^2} = 9604$$