Lecture 16 (Ch. 7)

We started with the CLT: $\bar{x} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$ — pop. mean; pop. std. dev.; sample size

Then we can compute things like $pr(\bar{x} > \text{something})$.

But, we want to know μ_x. So, we turn things around

"self-evident fact"

$pr(-1.96 < z < 1.96) = 0.95 \implies$ 95% CI for μ_x: $\bar{x} \pm 1.96 \frac{\sigma_x}{\sqrt{n}}$.

This CI is a random CI.

The observed CI uses $\bar{x}_{obs}$ in place of $\bar{x}$.

We don't know σ_x; for now, just approximate it with s. (Sample std. dev.)

The math is trivial!

It's the interpretations of CI that are really difficult.

For the prev. example: (Observed) 95% CI: (2.6, 3.4)

1) We can be 95% confident that the true mean is in here.

2) There is a 95% prob. that a random 95% CI will cover μ_x.

Note that the 2nd interp. makes no reference to (2.6, 3.4)!

Relationship between prob. and confidence:

→ $\boxed{\text{prob}}$ acts on random things, like sample means.

e.g. prob($\bar{x} > 3$) is perfectly meaningful.

prob($\mu > 3$) makes no sense!

→ $\boxed{\text{Confidence}}$ acts on fixed things, like pop. means.

e.g. C.I. for μ_x is perfectly meaningful.

C.I. for $\bar{x}$ makes no sense!

important.

Q What about other confidence levels ($\neq 0.95$)?

A E.g. 99% conf. level: "self-evident fact."

$$\text{prob}(-2.575 < z < 2.575) = 0.99 \qquad \boxed{\text{Table I}}$$

$$z = \frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} \quad \cdots \Rightarrow \quad \text{C.I. for } \mu_x : \ \bar{x} \pm 2.575 \frac{\sigma_x}{\sqrt{n}}$$

In general: $\boxed{\text{C.I. for } \mu_x : \ \bar{x} \pm z^* \frac{\sigma_x}{\sqrt{n}}}$ "multiplier"

where $z^* = 1.645,\ 1.96,\ 2.575,\ \ldots$

for conf. level $=$ 90%, 95%, 99%, ... $= 1 - \alpha$

or α-level $= 0.1,\ 0.05,\ 0.01,\ \ldots$

you can either "derive" these z^* values from Table I (just like we did for the above examples), or look them up on the last line of Table IV.

The formula for C.I. can be used to decide what minimum sample size is necessary, even before taking any sample!

But you need to specify what is meant by necessary.

For example, say, you want your estimate of μ_x to be within some range $\pm B$ (for Bound). Then

for now, approximate with s.

$$\frac{z^* \sigma_x}{\sqrt{n}} = B \Rightarrow n_{min} = \left(\frac{z^* \sigma_x}{B}\right)^2$$

Note that B is different from conf. level, or z^*. It has the dimensions of μ_x itself.

Example problem 7.12??

Concentration of zinc in 2 types of fish

	n	$\bar{x}$	s
Type 1	56	9.15	1.27
Type 2	61	3.08	1.71

} sample / data.

What's the True/pop. mean for Type 1 fish, at 95% conf. level?
" " " " 2 " " 99% " " ?

In the old days all we could write was

$\bar{x} \pm S$

But now

Type 1

$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$

$9.15 \pm 1.96 \frac{1.27}{\sqrt{56}}$

9.15 ± 0.333

$(8.82, 9.48)$

Type 2

$\bar{x} \pm 2.575 \frac{s}{\sqrt{n}}$

$3.08 \pm 2.575 \frac{1.71}{\sqrt{61}}$

3.08 ± 0.564

$(2.52, 3.64)$

it's common practice to leave off the "obs" on $\bar{x}_{obs}$ in $\bar{x}_{obs} \pm z^* \frac{s}{\sqrt{n}}$

$\boxed{\text{Interpretation}}$ ⟸ $\boxed{\text{IMPORTANT}}$

→ We are 95% confident that the true pop. mean of Zinc Concentration for Type I fish is between 8.8 and 9.5.

→ There is a 95% prob. that a random sample will yield a C.I. that covers the true mean of zinc concentration.

Note that the 2nd interpretation makes no reference to the observed C.I. (8.8, 9.5) at all!

Note: C.I. for μ_x of Type 2 fish is wider
(ie. our estimate for μ_x is less reliable/precise) Why?

→ The conf. level is higher

→ Sample std. dev. (s) is larger.

→ Eventhough n is larger (which shrinks the C.I.), the increase in n is not enough to compensate for the increase in conf. level and s.

What min. sample size is required for a margin of error of 0.03 $\frac{\mu g}{g}$?

B — units of $\bar{x}$.

$$n = \left(\frac{z^* \sigma_x}{B}\right)^2 \approx \left(\frac{1.96\ (1.27)}{.03}\right)^2 = 6,885 \text{ type I Fish.}$$

$$\approx \left(\frac{2.575\ (1.71)}{.03}\right)^2 = 21,543 \text{ type II Fish.}$$

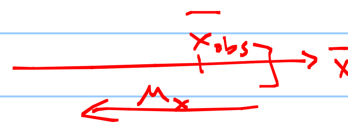
If you have no sample to provide an estimate of σ_x, Then you guess it! It's not hard. For example, if we're dealing with people's height, Then $\sigma_x \sim$ a few inches.

This above C.I. is called <u>2-sided</u>. **1-sided**

Sometimes, Though, we want to find only an upper confidence bound, or a lower confidence bound, for μ_x.

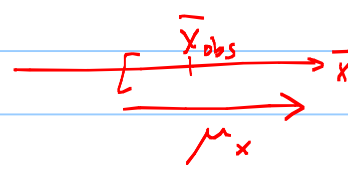
These are called 1-sided C.I. (or Conf. Bound): different from B, above.

upper. Conf. Bound: $\bar{x}_{obs} + z^* \frac{\sigma_x}{\sqrt{n}}$



approx. with s_x

lower " " $\bar{x}_{obs} - z^* \frac{\sigma_x}{\sqrt{n}}$



But z^* is diff. from 2-sided z^*'s:

See hw

90%	95%	99%	etc.
1.28	1.645	2.33	

Table I or last line in Table IV.

1-sided C.I. (or conf. bounds) are useful when we want to see if The True mean is greater (or smaller) Than some value.

Interpretation (IMPORTANT!)

Suppose 95% upper conf. bound for μ_x is 0.3. Then

1) We are 95% confident That $\mu_x < 0.3$

2) There is a 95% prob. That a random (95%) upper conf. bound will be <u>greater</u> Than μ_x.

Example

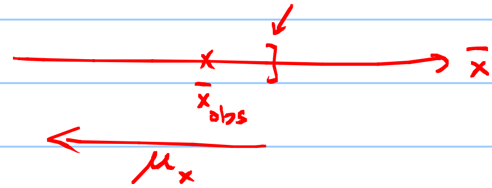
Consider The fish example again, but only The Type 1 fish:

n	$\bar{x}$	s	
56	9.15	1.27	sample / data.

This time, we don't care to put bounds on both sides of μ_x.
What we want to know is how big μ_x can get, with 95% Conf.

95% observed Upper Conf. Bound : $\bar{x}_{obs} + 1.645\,\sigma_x/\sqrt{n} \simeq \overbrace{9.15 + 1.645\,\frac{1.27}{\sqrt{56}}}^{\approx}$

: $9.15 + 0.28 = 9.43$



Note : This is not The upper end of The 2-sided C.I (ie. 9.48)

Interpretation : 1) We are 95% confident That $\mu_x < 9.43$.

2) There is a 95% prob. That a random (95%) upper confidence bound will be greater Than μ_x.

Caution: In addition to interpretation, another difficult Task is to figure out which conf. bound to compute: lower or upper.
The choice depends on The problem.
For reasons That will become more clear in Ch. 8, The question you want to ask yourself is How "bad" can μ_x get?

Sometimes "bad" means small $\Rightarrow$ lower conf. bound.
" " " large $\Rightarrow$ upper " ".

In The above example "bad" means large, because higher level of zinc is a bad Thing.

Q1: For the fish examples, for Type I fish, the 95% 2-sided CI was found to be $9.15 \pm 1.96 \frac{1.27}{\sqrt{56}} = (8.82, 9.48)$.

The 95% upper conf. bound was found to be $9.15 + 1.645 \frac{1.27}{\sqrt{56}} = 9.43$

Which statement is <u>false</u>?

A) There is a 95% prob. that a random $\bar{x} > \mu_x - 1.645\, \sigma_x/\sqrt{n}$.

B) We are 95% conf. that $8.82 < \mu_x < 9.48$

C) " " " " " $\mu_x < 9.43$

$\boxed{\text{D)}}$ " " " " " $\mu_x < 9.48$

A) This is the defn of 95% upper conf. bound; it has to cover μ_x 95% of the time. But if you want math, here it is:

$$pr\left(\bar{x} > \mu_x - 1.645 \frac{\sigma_x}{\sqrt{n}}\right) = pr\left(\frac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}} > \frac{\mu_x - 1.645\,\sigma_x/\sqrt{n} - \mu_x}{\sigma_x/\sqrt{n}}\right) = pr(z > -1.645)$$

$$= pr(z < 1.645) = 0.95$$

B) Defn. of 2-sided C.I.

C) " " upper conf. bound.

D) The interval ——] 9.43 covers μ_x 95% of the time.

So " " ——] 9.48 will cover it more frequently.

hw-lect16-1

Suppose you have computed a 95% C.I. for μ_x based on a sample of size n. Your friend, however, wants to compute a 99% C.I. for μ_x. How big should his sample size (m) be in order for the two CIs to have the same width?

hw-lect-16-2

Suppose we are developing a new composite material for building airplain wings. We take a sample of size 100 of the material and test its breaking strength under a set of standard conditions. The sample mean and sample standard deviation of the breaking strength are 20 and 5, respectively.
a) What type of confidence interval is appropriate for this problem (2-sided interval, an upper confidence bound, or a lower confidence bound)? Explain.
Hint: A small breaking point is a very bad thing! So ask yourself this question: do you want to know how small mu can get in which case you must compute a lower conf bound), or how large it can get (in which case ...)?
b) Compute it for this data, and provide two interpretations. Use a confidence level of 95%.

hw-lect16-3

Starting from a self-evident fact, derive the formula for the a) Lower 95% conf. bound, b) Upper 95% conf. bound

Hint: It may take some trial-and-error to find the correct self-evident fact for each part; but you will learn something from all of the trial-and-error.

As the hint suggested, you may start with the $<$ and $>$ signs switched in parts a and b. But at the end of the calculation you will know if you made that mistake. In other words my starting point is <u>not</u> obvious