Lecture 20 (Ch. 7)

or not, if n = large.

We have The CLT : $\bar{x} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$ if $x \sim N(\mu_x, \sigma_x)$

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 : $\bar{x} \pm 1.96 \frac{\sigma_x}{\sqrt{n}}$.

→ $\boxed{\text{prob}}$ 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."

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

$z = \frac{\bar{x} - \mu_x}{\sigma_x / \sqrt{n}}$ ... $\implies$ C.I. for μ_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"

$\boxed{\text{where}}$ $z^* = 1.645, 1.96, 2.575, \ldots$

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

or α-level = 0.1, 0.05, 0.01, ...

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.

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 $\bar{x}_{obs} \pm z^* \frac{s}{\sqrt{n}}$

Interpretation ⟸ 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.

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

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

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

Example Consider the fish example from prev. lect. What min. sample size is required for a margin of error of 0.03 $\frac{\mu g}{g}$?

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

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

B units of $\bar{x}$.

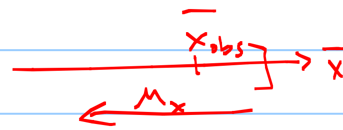
If you have no sample to provide an estimate of σ_x, Then you guess it! It's not hard. For example, if we're dealing with people's height, Then $\sigma_x \sim$ a few inches.

This above C.I. is called 2-sided.

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

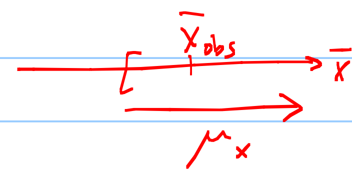
These are called 1-sided C.I. (or Conf. Bound): different from 8, above.

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 upper conf. bound will be greater than μ_x.

Q1: In here ↗ which statement below is FALSE.

A) There is a 95% prob that $\bar{x}_{obs}$ is less than 0.3. no prob

B) We are 95% confident that $\mu_x \leq 0.3$ same as $\mu_x < 0.3$

C) There is a 95% prob. that (a random) $\bar{x}$ is $> \mu_x - 1.645 \frac{\sigma_x}{\sqrt{n}}$ same as $\mu_x < \bar{x} + z^* \frac{\sigma}{\sqrt{n}}$

In Summary:

next time.

1) $E[\bar{x}] \equiv \mu_{\bar{x}} = \mu_x$

$V[\bar{x}] = \sigma^2_{\bar{x}} = \frac{\sigma_x^2}{n}$

$\boxed{E[p] \equiv \mu_p = \pi_x \quad V[p] \equiv \sigma_p^2 = \frac{\pi_x(1-\pi_x)}{n}}$

quite independently of The distr. of x (i.e. The pop.)

The Right-hand sides are all pop. parameters.

Q what to do if you don't know what they are?

A1 Assume the pop, and calculate probs of $\bar{x}$. ← ch. 8.

A2 Build CI's for Them. ← ch. 7.

2) If pop = normal with params μ, σ, ie. $f(x) = N(\mu, \sigma)$

$x \sim N(\mu, \sigma)$

then $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

3) even if pop is not normal, as long as n = large

ch. 8

4) Then, we can compute $\text{prob}(a < \bar{x} < b)$ and $\text{prob}(a < p < b)$

5) We can build C.Is for μ_x and π_x, both 1-sided and 2-sided.

ch. 7.

We used CLT to derive The sampl. distr. of $\bar{x}$ and p and from That, The C.I. for μ_x and π_x.

But even if we have no idea what The sampl. distr. of our statistic is, we can still build it empirically, using The bootstrap idea (in Lab).

hw-lect20-1

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-lect20-2

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-lect20-3

Show that the z^* for a 99% lower conf. bound for a pop. mean is 2.33. In other words, derive the formula for the 99% Lower conf. bound for μ_x, and show that it's $\bar{x} - 2.33 \frac{\sigma_x}{\sqrt{n}}$.

hw-lect20-4

Write R code to take 1000 samples of size 100 from a standard Normal, compute the 95% upper confidence bound for each (of the 1000 samples) using the formula given in lect, and count how many (of the 1000 1-sided confidence intervals) cover the pop. mean. Hint: see prelab 6.