Lecture 16 (Ch. 7)

Review:

The transition from the 1st half of the course (descriptive stats) to the latter half inference) is based on the concept of the (empirical) sampling distr. and the CLT.

The thought experiment of taking multiple samples (of size n) from a dist./pop., computing some sample statistic (e.g. $\bar{x}$) for each, and then looking at their hist (technically, dist).

For any population/dist. with mean and var. μ_x, σ_x^2, the dist. of $\bar{x}$ is $N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$.

sample size.

Then, we can compute probs, e.g.

$$pr(a < \bar{x} < b) = pr\left(\frac{a-\mu_x}{\sigma_x/\sqrt{n}} < \underbrace{\frac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}}}_{=z} < \frac{b-\mu_x}{\sigma_x/\sqrt{n}}\right) = \text{Table I.}$$

The example in the last lecture shows that 2 types of probs are useful (and common) in Statistics & data analysis:

$pr(\bar{x} \lessgtr \bar{x}_{obs})$. $\Rightarrow$ p-value (Ch. 8)

$pr(-1.96 < z < 1.96) = 0.95$ $\Rightarrow$ Confidence Int. (Ch. 7) Now.

→ Meaning of "prob"

prob(A) = "long-run proportion of A."

= Out of 10^8 samples of size n taken from the pop., the proportion for which A happens.

→ Prob works on random variables:

I.e. prob($a < \bar{x} < b$) is computable. ✓

prob($a < \mu < b$) is Not ✗

prob($a < \bar{x}_{obs} < b$) is Not ✗

prob($\bar{x} = \bar{x}_{obs}$) = 0 useless

The 1^{st} way is to build a confidence Interval (CI) for μ_x:

The procedure is to start with $pr(a<z<b) = \text{blah}$, with specific values of a, b, and blah. E.g.

self-evident fact

$$pr\left(-1.96 < z < 1.96\right) = 0.95$$

$$pr\left(-1.96 < \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} < 1.96\right) = 0.95$$

$\mu_{\bar{x}} \leftarrow \mu_x$, $\sigma_{\bar{x}} \leftarrow \sigma_x/\sqrt{n}$

$$pr\left(-1.96\frac{\sigma_x}{\sqrt{n}} < \bar{x}-\mu_x < +1.96\,\sigma_x/\sqrt{n}\right) = 0.95$$

"Confidence level"

$$\bar{x} - 1.96\frac{\sigma_x}{\sqrt{n}} < \mu_x < \bar{x} + 1.96\frac{\sigma_x}{\sqrt{n}}$$

NO Prob.

fixed

random

$$\boxed{\therefore\ 95\%\ \text{C.I. for } \mu_x:\quad \bar{x} \pm 1.96\frac{\sigma_x}{\sqrt{n}}}$$

pop. mean

This is a random C.I, because $\bar{x}$ is random (how else would it have a sampling dist?!)

See example, below, about what to do with This unknown.

The (observed) 95% C.I. for μ_x is $\boxed{\bar{x}_{obs} \pm 1.96\frac{\sigma_x}{\sqrt{n}}}$

1 Interpretation: We are 95% Confident That μ_x is in here.

2nd " : Below.

→ Often we forget saying "observed".

It's up to you to find out if we're talking about a random CI or The observed CI.

E.g. A sample of size 25 yields $\bar{x}_{obs} = 3$, $s_{obs} = 1.5$.

Last time: Suppose pop is Normal ($\mu_x = 2$, $\sigma_x = 1$). What's the prob of getting an even larger sample mean?

This time: What's the (observed) 95% C.I for μ_x?

$\boxed{\text{First attempt}}$ We don't use/know $\mu_x = 2$, but we know/use $\sigma_x = 1$

(observed) 95% C.I. for μ_x : $\bar{x}_{obs} \pm 1.96 \frac{\sigma_x}{\sqrt{n}}$

$$3 \pm 1.96 \frac{1}{\sqrt{25}} = 3 \pm .392 = [2.6, 3.4]$$

$\boxed{\text{Second attempt}}$ We don't use/know either μ_x or σ_x.
(Approximate σ_x with sample std. dev.)

$$3 \pm 1.96 \frac{1.5}{\sqrt{25}} = 3 \pm 0.588 = [2.4, 3.6]$$

For now (until we learn about The t-distribution), we will usually follow The second attempt (ie. approximate σ_x with s).

Eitherway, Here is one (of 2) interpretation of The CI:

We are 95% Confident That The True mean is in here.

It's important To note That The word probability does not appear in This interpretation, even Though we started with a probability. It happened in The last step of The derivation of The C.I for μ_x, when I dropped The pr. That is because $pr(\dots < \mu_x < \dots)$ does not exist, because μ_x is a fixed population param, not random.

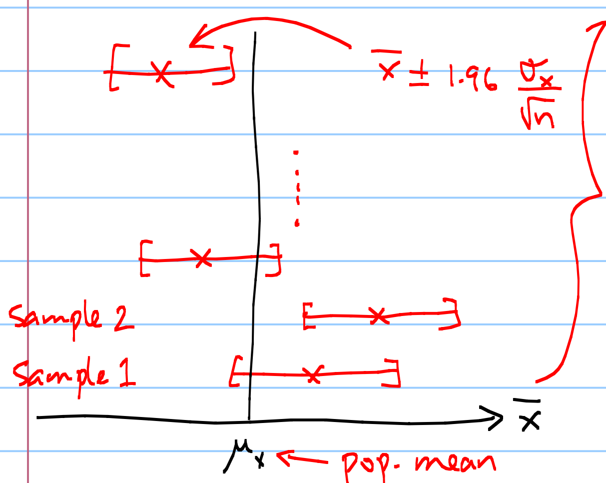
BUT, There **is** a way of squeezing "probability" into The conclusions, but it has to pertain to The random C.I.

We are 95% confident That The pop. mean is in The interval $\boxed{\bar{x}_{obs} \pm 1.96 \frac{\sigma_x}{\sqrt{n}}}$.

$\Updownarrow$ Equivalent interpretations of C.I.

There is a 95% prob That a random sample will yield a C.I. $\left(\bar{x} \pm 1.96 \frac{\sigma_x}{\sqrt{n}}\right)$ That covers μ_x.

Look at The derivation of CI; This is obvious.



→ 95% of These intervals cover μ_x.

→ I.e. The prob. That a random C.I. $\left(\bar{x} \pm 1.96 \frac{\sigma_x}{\sqrt{n}}\right)$ will include μ_x is 0.95.

→ If you want to say something directly about μ_x, use "confidence" Not prob.

CI's are all about coverage;
a 95% C.I. for μ_x is designed to cover μ_x in 95% of samples.

---

above Example Data ⟹ 95% observed C.I : [2.6, 3.4]

Conclusion(s):

1) We are 95% confident That μ_x is in [2.6, 3.4]

2) There is a 95% prob. That a random CI covers μ_x. ←

No mention of anything observed!

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

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

$$\text{prob}(-2.575 < \underset{\frac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}}}{\underset{\shortparallel}{z}} < 2.575) = 0.99 \qquad \text{Table I}$$

$$\cdots \implies \text{C.I. for } \mu_x : \; \bar{x} \pm 2.575 \frac{\sigma_x}{\sqrt{n}}$$

The only thing that changes when we change the conf. level is .

So, in general, the self-evident fact becomes

from Table I.

$$\text{pr}(-z^* < z < z^*) = \text{Confidence level.}$$

and the CI formula becomes

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

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

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

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

you can either "derive" these z^* values from Table I

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	} sample / data.
Type 2	61	3.08	1.71	

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

$\bar{x}_{obs} \pm S$

But now, and we have an interpretation to go with our C.I formulas

Type 1

$\bar{x}_{obs} \pm 1.96 \frac{S}{\sqrt{n}}$

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

9.15 ± 0.333

$(8.82, 9.48)$

Type 2

$\bar{x}_{obs} \pm 2.575 \frac{S}{\sqrt{n}}$

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

3.08 ± 0.564

$(2.52, 3.64)$

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 <u>wider</u>
(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.

WARNING!

The math is trivial.

It's the correct jargon & interpretations that are tricky:

⇒ There is random vs. observed vs. fixed pop. params.

↳ $\bar{x}$, CI, ... ↳ $\bar{x}_{obs}$, obs. CI, ... ↳ μ_x, σ_x, ...

E.g.

There are 3 means: $\bar{x}$, $\bar{x}_{obs}$, μ_x

There is random (e.g. $\bar{x}$) vs. not (e.g. $\bar{x}_{obs}$, μ_x)

⇒ There is also random CI vs. observed CI

⇒ Then, there is confidence vs. probability

↳ for pop. params e.g. μ_x, σ_x, ... ↳ for random thing e.g. $\bar{x}$

$pr(\mu_x > 3)$ ✗ $pr(\bar{x} > 3)$ ✓

C.I. for μ_x ✓ C.I. for $\bar{x}$ ✗

As a result of all these "moving parts," there are more ways of saying things incorrectly than correctly!

we have reached our goal of being able to say something about a pop. from a sample!

Celebrate a little! But there is more (a lot more).

hw_lect16-1: Do part e of hw_lect 15_3.

hw_lect16-2

A sample of size 25 yields $\bar{x}_{obs}=3$, $s_{obs}=1.5$. Suppose we know $\sigma_x=1$. Then, as shown in the example, we can be 95% Confident That μ_x is in The interval $(2.6, 3.4)$.

a) Based on This information, Can you find $pr(2.6<\bar{x}<3.4)$? If yes, find it. If not, why not?

b) Show $pr(2.6<\bar{x}<3.4) = pr(z_{obs}-1.96<z<z_{obs}+1.96)$, where $z_{obs}=(\bar{x}_{obs}-\mu_x)/(\sigma_x/\sqrt{n})$. Hint: $2.6=(\bar{x}_{obs}-1.96\frac{\sigma_x}{\sqrt{n}})$ from our data.

c) Now, forget all observed information, and suppose we know $\mu_x=2$, $\sigma_x=1$. What is The numerical value of $pr(2.6<\bar{x}<3.4)$ when $n=25$?

d) If $\mu_x=2$, What is The numerical value of " ?

e) If $\mu_x=2.8$ and $\sigma_x=0.608$, " " ?

hw_lect16-3

One can build C.I for any pop. param (not just The mean μ_x). For example, if The population is Unif(0, b), we can build a CI for The b param. Do it!

zero

Assume $\bar{x}$ is (approximately) Normal: Recall That μ_x and σ_x of Unif(a,b) are $\frac{1}{2}(a+b)$, $\frac{(b-a)}{\sqrt{12}}$, respectively.

Hints: what is The quantity That has a normal dist?

what is " " " " " standard norm. dist?

Use it in a self-evident fact, and solve for b.

hw_optional

Suppose $\sigma_x/\sqrt{n}=1$. Find The possible $\bar{x}_{obs}$ values such That The 95% obs. CI for μ_x includes $\mu_x=0$.