Lecture 15 (Ch. 5–7)

We arrived at The Central Limit Theorem (CLT):

weak version: If $x \sim N(\mu, \sigma)$, Then $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

strong version: If $x \sim$ any dist. with mean $= \mu_x$, var. $= \sigma_x^2$

Then $\bar{x} \sim N(\underbrace{\mu_x}_{=\mu_{\bar{x}}}, \underbrace{\frac{\sigma_x}{\sqrt{n}}}_{=\sigma_{\bar{x}}})$ for large n.

In English: For any pop. with mean μ_x and variance σ_x^2, The sampling dist. of The sample means is Normal with $\mu = \mu_x$, $\sigma = \sigma_x / \sqrt{n}$.

$\mu = \mu_{\bar{x}}$, $\sigma = \sigma_{\bar{x}}$

So, if we know The pop. (ie. $f(x)$, $p(x)$), Then we can compute The prob. That a random sample mean will be somewhere.

E.g. $pr(a < \bar{x} < b)$. This is how:

1) Compute μ_x, σ_x: $\mu_x \equiv E[x] = \sum_x x\, p(x)$, $\sigma_x^2 \equiv V[x] = \sum_x (x - \mu_x)^2 p(x)$. (or $\int f(x) \ldots$)

2) From CLT we know $\bar{x} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$

3) Then standardize: $z \equiv \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_x}{\sigma_x / \sqrt{n}} \sim N(0,1)$

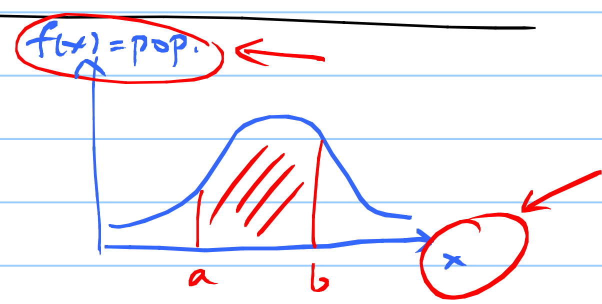
4) Finally $pr(a < \bar{x} < b) = pr(a - \mu_x < \bar{x} - \mu_x < b - \mu_x)$

$$= pr\left(\frac{a - \mu_x}{\sigma_x/\sqrt{n}} < \frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} < \frac{b - \mu_x}{\sigma_x/\sqrt{n}} \right)$$

$$= pr(\text{"} < z < \text{"}) \Rightarrow \text{Table 1}$$

Compare with what we did in Ch. 1, 2:

$$pr(a < x < b) = pr\left(\frac{a - \mu_x}{\sigma_x} < z < \frac{b - \mu_x}{\sigma_x}\right)$$



E.g. Suppose a sample of size 25 yields $\bar{x}_{obs} = 3$, $s_{obs} = 1$.
If the population is $N(\mu = 2, \sigma = 1)$, what's the prob.(*) of getting an even larger sample mean? → $\mu_x = 2$, $\sigma_x = 1$

$\text{prob}(\bar{x} > \bar{x}_{obs}) =$ $\qquad \frac{\bar{x} - \mu_x}{\sigma_x / \sqrt{n}}$

$\text{prob}(\bar{x} > 3) = \text{prob}\left(z > \frac{3-2}{1/\sqrt{25}}\right) = \text{prob}(z > 5) \approx 0$!

This small prob suggests that $\mu = 2$ is a bad assumption. In fact, we may even guess that μ is greater than 2 (closer to 3)! We will formalize these qualitative conclusions, below.

(*) Recall "prob" = proportion of samples (of size n) taken from the population, in the long-run (e.g. out of 10^8 samples)

prob works on random variables:

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

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

Note: in these calculations we are assuming we know the pop. But we don't. Intuitively, these probs give us a sense of how likely it would be to get a random sample mean somewhere, IF the pop. is given. In ch. 7, 8 we will come-up with 2 ways of turning things around to say something about pop. from data.

Recall our symbols:

statistics		pop. parameters
$\bar{x}$ (sample mean)	is a point estimate of	μ_x (pop. mean)
s (" std. dev.)	" " " "	σ_x (" std. dev.)
p (" prop.)	" " " "	π_x (" prop.)

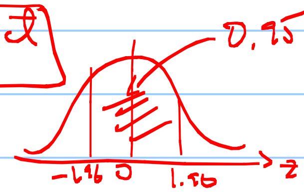
n (" size) is ~~NOT~~ NOT related to pop. size. ← for us $= \infty$

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

$$pr(-1.96 < z < 1.96) = 0.95$$



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

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

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

"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?!)

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

For now, approximate this with sample std. dev.

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. Suppose a sample of size 25 yields $\bar{x}_{obs} = 3$, $s_{obs} = 1$.
What can we say about the pop. mean?

prev. eg.

Suppose pop is Normal ($\mu_x = 2$, $\sigma_x = 1$). What's the prob of getting an even larger sample mean? ch.8

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

prob($\bar{x} > 3$) = prob($z > \frac{3-2}{1/\sqrt{25}}$) = prob($z > 5$) ≈ 0!

$\bar{x} > 3$ is unlikely, if $\mu_x = 2$.

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

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

1 Interp : We can be 95% Confident That The True mean is in here.

Note That we have actually made it to our goal of being able to say something about a pop. mean, from a sample.
Review how we needed everything we've done since Ch.1.
Go and celebrate!

Q1: For The above e.g. which of The following is correct.

A) The prob That $2.6 < \mu_x < 3.4$ is 95% $\quad \mu_x$ = fixed

B) " $2.6 < \bar{x}_{obs} < 3.4$ " $\quad \bar{x}_{obs}$ = fixed

C) " $2.6 < \bar{x} < 3.4$ " $\quad pr(\cdots < \bar{x} < \mu_x + 1.96\frac{\sigma_x}{\sqrt{n}}) = 0.95$

$\boxed{D)}$ none of The above. $\quad$ not $\bar{x}_{obs}$.

See Below for an interpretation That does involve prob.

Note that in the last step of the derivation of the C.I. for μ_x, I dropped the pr. That is because $pr(\cdots > \mu_x > \cdots)$ does not exist, because μ_x is fixed, not random.

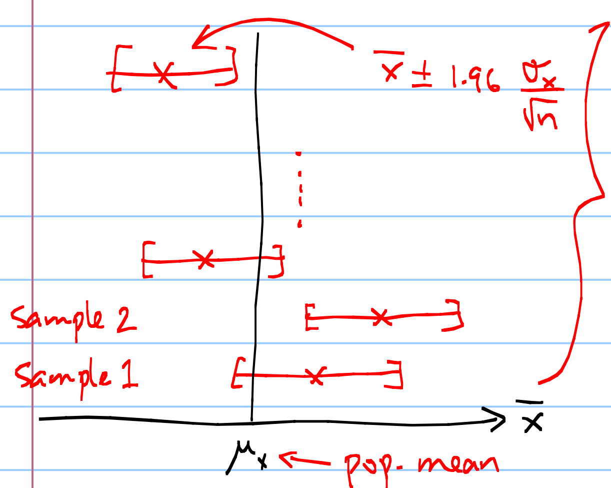
There is a way of squeezing "probability" into the conclusions, but it has to pertain to the random C.I.

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

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.

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

For the above example: (Observed) 95% CI $(2.6, 3.4)$

2nd interp.: There is 95% prob that a random CI will cover μ_x.

hw-lect 15-1

A sample of size 36 from a Normal pop. yields $\bar{x}=3, s=1$.

a) Under the assumption that $\mu_x=2.5, \sigma_x=2$, what's the prob of a sample mean larger than the one observed.

b) Under the assumption that $\mu_x=2.5, \sigma_x=2$, what's the prob of a sample mean smaller than the one observed.

c) Under the assumption that $\mu_x=3.5, \sigma_x=2$, what's the prob of a sample mean larger than the one observed.

d) Under the assumption that $\mu_x=3.5, \sigma_x=2$, what's the prob of a sample mean smaller than the one observed.

---

hw-lect 15-2

a) It turns out that the sample std. dev., s, has a Normal distr. with parameters σ_x and $\sigma_x/\sqrt{2n}$, where σ_x is the pop. std. dev. Now, follow the procedure we have developed, starting from a "self-evident fact" to develop a C.I. formula for σ_x.

b) Suppose for a specific data set based on a sample of size 169, we have found the sample std. dev. of 3.73. Compute the 95% CI for the pop. std. dev.

c) provide 2 interpretations.