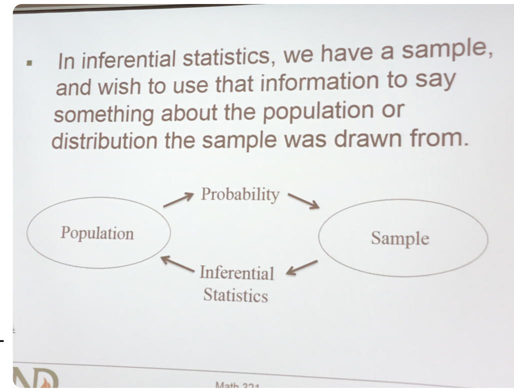
→ Test claims about population by sampling

parameter: unknown quantity
(ex: μ)

statistic: quantity calc. w/ dataset

estimation: stat used to say something about unknown params.



sample → population

Point Estimation

of θ, is a stat, $\hat{\theta}$, which is estimate for θ

$X \sim f(x)$ unknown

$\hat{\mu} = \bar{X}$

ex: $V(X) = \sigma^2$ w/ $\hat{\sigma}^2 = S^2$

if $X \sim \text{Binomial}(n, p)$ - discrete random

n known, p unknown

est. p w/ $\hat{p} = X/n$

ex: $X \sim \text{Bin}(100, p)$

X = num of students from Minnesota out of 100

suppose sample = 100 + 34 from Minnesota

$\hat{p} = 34 : 100$

or $\boxed{.34}$ estimate

Properties of Estimates

estimate gives correct value for param

$E(\hat{\theta}) = \theta$, $\therefore \hat{\theta}$ is unbiased estimate

av. of all estimates = θ hopefully? $\left(E(\bar{X}) = \mu\right)$

$\sim E(X)$

mean of sample means = pop. mean

also

$\boxed{E(S^2) = \sigma^2 \; + \; E(\hat{p}) = p}$

$\Rightarrow E(\hat{p}) = E(X/n) = \frac{1}{n}E(X) = \frac{1}{n}(n \cdot p) = p$

sample var + estimate are unbiased estimators

standard dev. is not unbiased

~~$E(S^2) = \sigma^2$~~

~~$E\sqrt{X} \neq \sqrt{E(X)}$~~

~~$E(S) = \sigma$~~

$X \sim Bin(25, 0.3)$

$X = 7$, $\hat{p}$ $7/25 \approx 0.28$

$X = 8$, $\hat{p}$ $8/25 \approx .32$

if you do enough samples, you will get close to pop. #'s

$\boxed{\text{WANT:}}$ low bias, low variance :)

mean squared error:

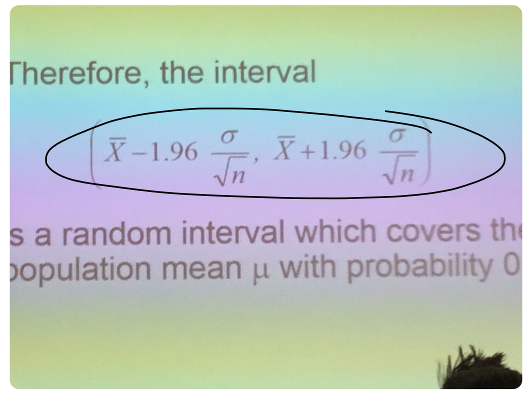
small mse = both bias & var. are small

$$mse(\hat{\theta}) = bias^2(\theta) + variance(\hat{\theta})$$

Central limit theorem

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

normal dist. table ☺



95% confidence interval

consub. $\bar{x}$ for random $\bar{X}$

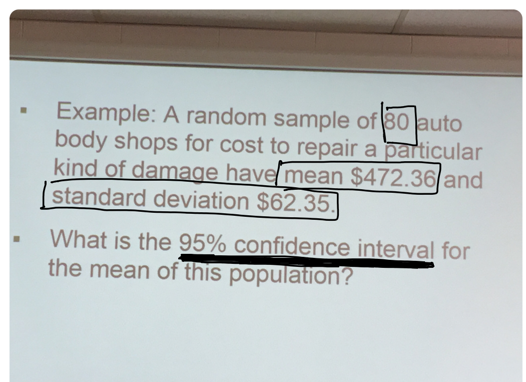
s for σ → if unknown

↓ sample standard deviation

need $\bar{x} = 472.36$

$s = 62.35$

$n = 80$



$\bar{x} - 1.96\left(\frac{s}{\sqrt{n}}\right)$

$= 472.36 - 1.96\left(\frac{62.35}{\sqrt{80}}\right)$

$= 458.70$ lower bound

$472.36 + 1.96\left(\frac{62.35}{\sqrt{80}}\right)$

$= 486.02$

95% CI $= (458.7, 486.02)$

or 458.7 ± 13.66

or $458.7 < \mu < 486.02$

$\cancel{P(458.70 < \mu < 486.02) = 0.95?}$ nothing is random

random are sample statistics
no sample stats in

* if process was done many times, ~95% intervals contain true parameter

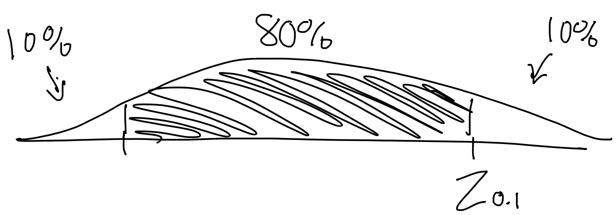
- randomness is sample, not population

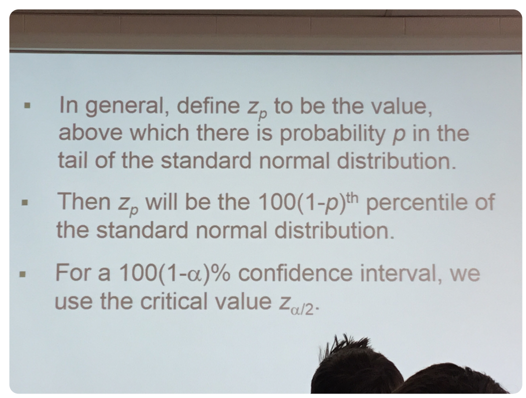
- 95% confident pop mean lies in this range

  - other confidence levels can be done by changing 1.96 but 95% conf. is standard

---

80% conf. interval





$1 - 0.8 = 0.2$

$Z_{\alpha/2} = Z_{0.1} \rightarrow$ go to table + look for .9 in middle

$Z_{0.1} = 1.28$

---

what effects size (precision of interval?)

$$\left(\bar{X} - 1.96\frac{s}{\sqrt{n}},\ \bar{X} + 1.96\frac{s}{\sqrt{n}} \right)$$

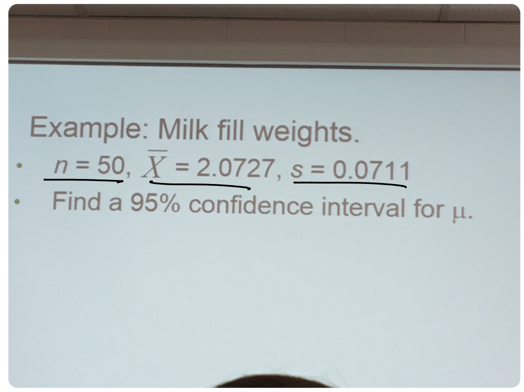
↑ standard dev = interval wider

smaller conf level = smaller interval

$z^* = 1.96$

$2.0727 - 1.96 \frac{0.0711}{\sqrt{50}} = \sim 2.053$

$2.0727 + 1.96 \frac{0.0711}{\sqrt{50}} = \sim 2.092$



$\boxed{CI = (2.053, 2.092)}$

w=? length? $2 * 1.96\left(\frac{0.0711}{\sqrt{50}}\right)$

$= \boxed{0.0394}$

---

want $w \leq 0.02$

$n = \left(\frac{2 * z^* (s)}{w}\right)^2$

$n = \left(\frac{2 * (1.96)(0.0711)}{0.02}\right)^2$

$\approx \boxed{194.2}$ ← sample size requirement

$\boxed{195}$

↳ s will change w/ new n

Confidence Bounds

only ~~want to~~ know upper or lower bound

95% lower = $\mu > \bar{x} - 1.645 \frac{s}{\sqrt{n}}$

1.645

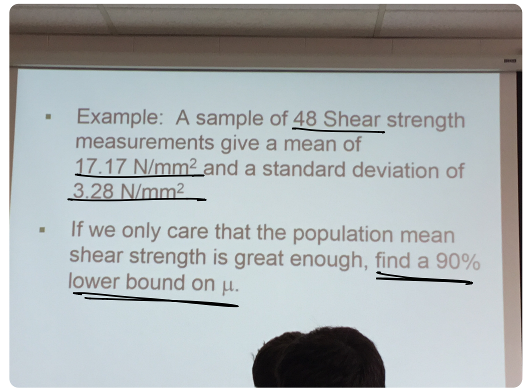
upper = $\mu < \bar{x} + 1.645 \frac{s}{\sqrt{n}}$

90% = $z^* = 1.28$

99% = $z^* = 2.33$ etc

Z_α

1 $s = 3.28$
$n = 48$
$\bar{x} = 17.17$
$z^* = -1.28$

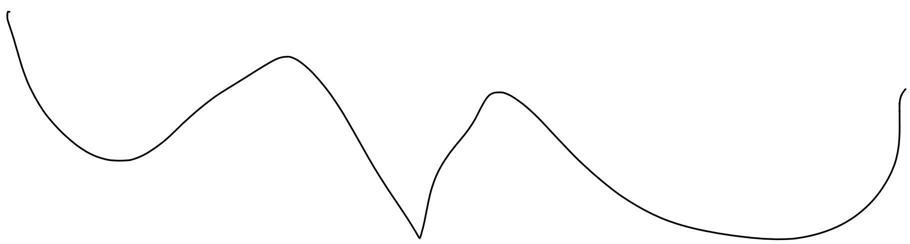


$z_{.10}$ → search for .9 in table = -1.28

$17.17 - (-1.28)\left(\frac{3.28}{\sqrt{48}}\right)$

$= 16.56$

90% conf true mean (μ) > 16.56



all based on CLT

- must be random draw
- pop is normal or sample size > 30
- need to know σ or n large enough that s is good for σ
- about $n \geq 40$