

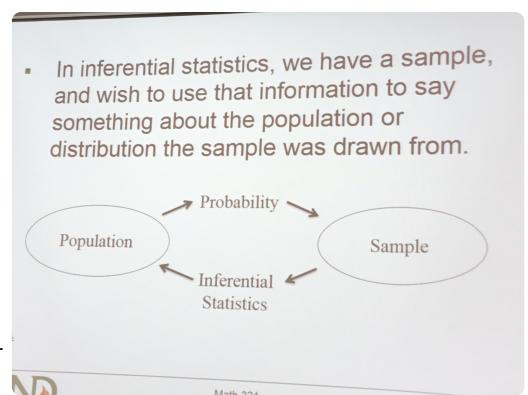
→ Test claims about population by sampling

parameter: unknown quantity
(ex: μ)

statistic: quantity calc. w/ dataset

estimation: stat used to say something about unknown params.

Sample → population



Point Estimation

If θ , is a stat, $\hat{\theta}$, what is estimator for

$X \sim f(x)$ 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 points

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

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

X = num of students from Minnesota out of 160

Suppose sample = 100 + 34 from Minnesota

$$\hat{p} = 34 : 100$$

or

$$\frac{34}{100} \text{ estimate}$$

Properties of Estimates

estimate gives correct value for param

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

Avg. of all estimates = $\hat{\theta}$ hopefully? ($E(\bar{X}) = \mu$)
 $n E(X)$

also

$$\begin{aligned} E(S^2) &= \sigma^2 \\ E(\hat{P}) &= p \end{aligned}$$

$$E(\hat{P}) = E(\bar{X}_n) = \frac{1}{n} E(X) = \frac{1}{n}(n \cdot p) = p$$

mean of sample
means =
pop. mean

Sample Var + estimate are unbiased estimators
standard dev. is not unbiased

$$\begin{aligned} E(S^2) &= \sigma^2 \\ E\sqrt{X} &\neq \sqrt{E(X)} \\ E(S) &= \sigma \end{aligned}$$

$$X \sim \text{Bin}(25, 0.3)$$

$$X = \tau, \hat{P} \frac{7}{25} \approx 0.28$$

$$X = \rho, \hat{P} \frac{8}{25} \approx 0.32$$

If you do enough samples,
you will get close to
pop. #

WANT: low bias, low variance \Downarrow

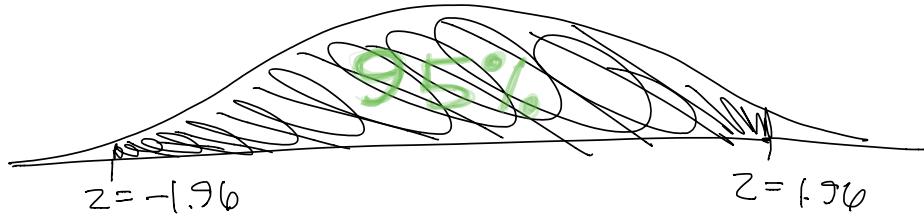
mean squared error:

small mse = both bias & var. are small

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

Central limit theorem

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



normal dist. table

95% confidence interval

comes from \bar{X} for X
 s for σ
 ↓ → if unknown
 sample standard deviation

need $\bar{X} = 472.36$
 $s = 62.35$
 $n = 80$

$$\begin{aligned}\bar{X} - 1.96 \left(\frac{s}{\sqrt{n}} \right) &= 472.36 - 1.96 \left(\frac{62.35}{\sqrt{80}} \right) \\ &= 458.70 \text{ lower bound} \\ 472.36 + 1.96 \left(\frac{62.35}{\sqrt{80}} \right) &= 486.02\end{aligned}$$

95% CI = (458.7, 486.02)

or 458.7 ± 13.66

or $458.7 < \mu < 486.02$

Therefore, the interval

$$\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right]$$

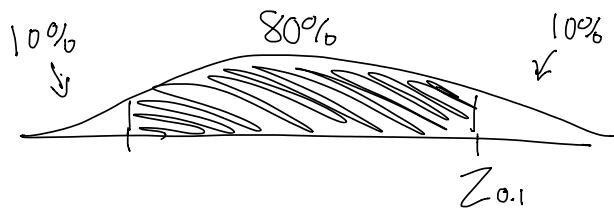
is a random interval which covers the population mean μ with probability 0.95

- Example: A random sample of 80 auto body shops for cost to repair a particular kind of damage have mean \$472.36 and standard deviation \$62.35.
- What is the 95% confidence interval for the mean of this population?

~~$P(458.70 < \bar{X} < 486.02) = 0.95?$~~ 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, ^{top} mean lies in this range
- other confidence levels can be done by changing %
 but 95% Cnf. is standard

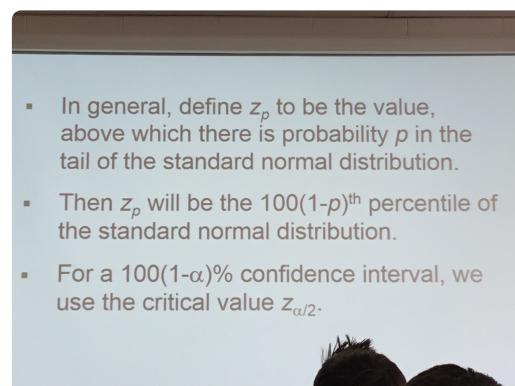
80% Cnf. interval



$$1 - 0.8 = 0.2$$

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

$$z_{0.1} = 1.28$$



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

$$z^* = 1.96$$

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

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

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

$w = ?$
length?

$$2 * 1.96 \left(\frac{0.0711}{\sqrt{50}} \right)$$

$$\boxed{0.0394}$$

want $w \leq 0.02$

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

$$\approx \boxed{194.2}$$

\leftarrow sample size requirement

$$\boxed{195}$$

$\downarrow s \text{ will change w/ new } n$

Confidence Bounds

only want to know upper & lower bound

$$95\% \text{ IWR} = \mu > \bar{x} - 1.96 \frac{s}{\sqrt{n}}$$



$$\text{Upper} = \mu < \bar{x} + 1.96 \frac{s}{\sqrt{n}}$$

$$90\% = z^* = 1.28$$

$$99\% = z^* = 2.33 \quad \text{etc}$$

$$\boxed{z_\alpha}$$

Example: Milk fill weights.

- $n = 50, \bar{X} = 2.0727, s = 0.0711$
- Find a 95% confidence interval for μ .

$$s = 3.28$$

$$n = 48$$

$$\bar{x} = 17.17$$

$$z^* = -1.28$$

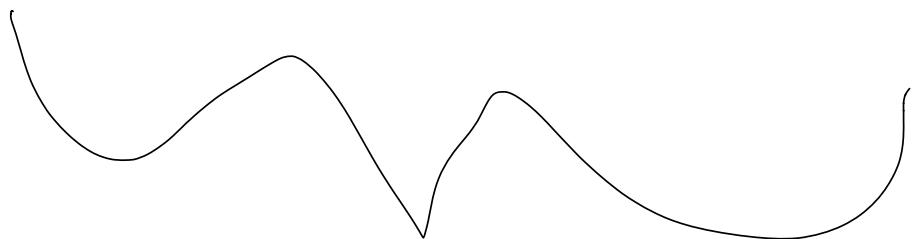


$$17.17 - (-1.28) \left(\frac{3.28}{\sqrt{48}} \right) \xrightarrow{\text{in table}} = -1.28$$

$$= 16.56$$

90% confidence lower bound $\mu > 16.56$

- Example: A sample of 48 Shear strength measurements give a mean of 17.17 N/mm^2 and a standard deviation of 3.28 N/mm^2
- If we only care that the population mean shear strength is great enough, find a 90% lower bound on μ .



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 \approx
- about $n \geq 40$

