WED OCT 25

NOTE: HW DUE TUESDAY IS ALWAYS DISCUSSED MONDAY CLASS
IF YOU ARE NOT UNDERSTANDING ASSIGNMENT, YOU SHOULD BE ATTENDING CLASS

TODAY: T-test, confidence intervals, example when distribution is not Normal
POWER OF TESTS

REVIEW:

DISTRIBUTION NOT SPECIFIED

$X_1, X_2, \ldots, X_n$ IID $\mu = E X_i$ $\sigma^2 = Var(X_i)$ unknown

$\mu = \int_{-\infty}^{\infty} x f_X(x)\,dx$ ← pdf of X_i

DIST'N OF $\bar{X}$ IS approx.

USE $\boxed{}$ to est. μ, $\boxed{}$ to estimate σ^2 $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Questions

(1) How do we evaluate evidence against specific null-hypotheses?

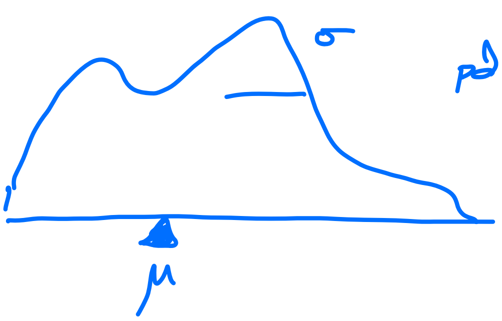
(2) How do we quantify uncertainty in estimate of μ?

(3) What is role of simulation vs theory?
Why simulate / how?

Review:

DATA numbers

MODEL: formulate a probability model for data
context dependent
library distribution — geometric, Binomial, Poisson, Normal
unknown parameters ⟷ pick parameter values so distribution approx data best
maximum likelihood estimation is algorithm for finding good parameter values for data



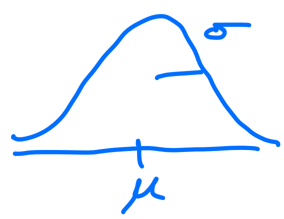
pdf of X

$\mu = E(X)$

$\sigma^2 = Var(X)$

$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

$\sigma = \sqrt{\sigma^2}$



μ, σ numerical quantities (not random)
property of pdf/pmf of X_i

Guess μ, σ from data: $x_1, x_2, \ldots, x_n$

unobserved parameters

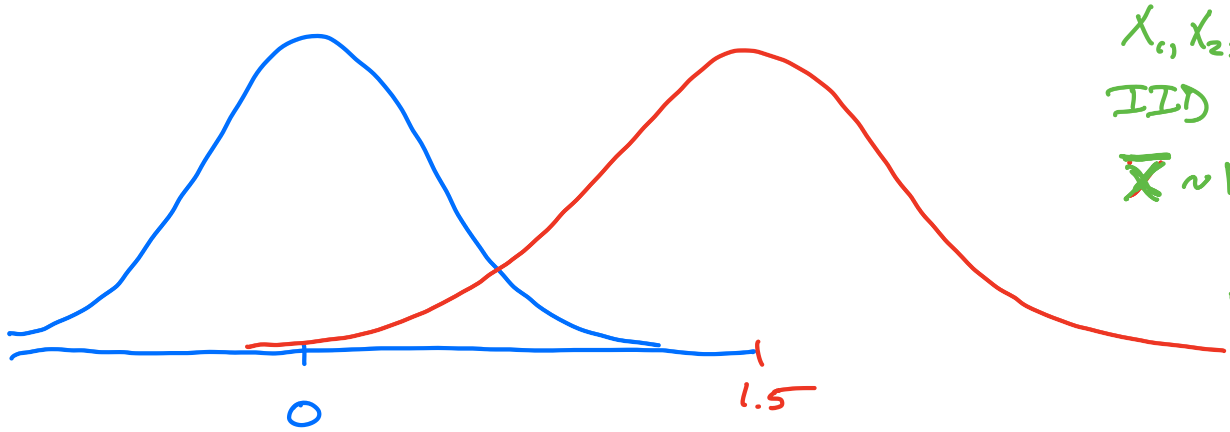
$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ sample average

$\bar{x}$ guess at μ

s^2 guess at σ^2

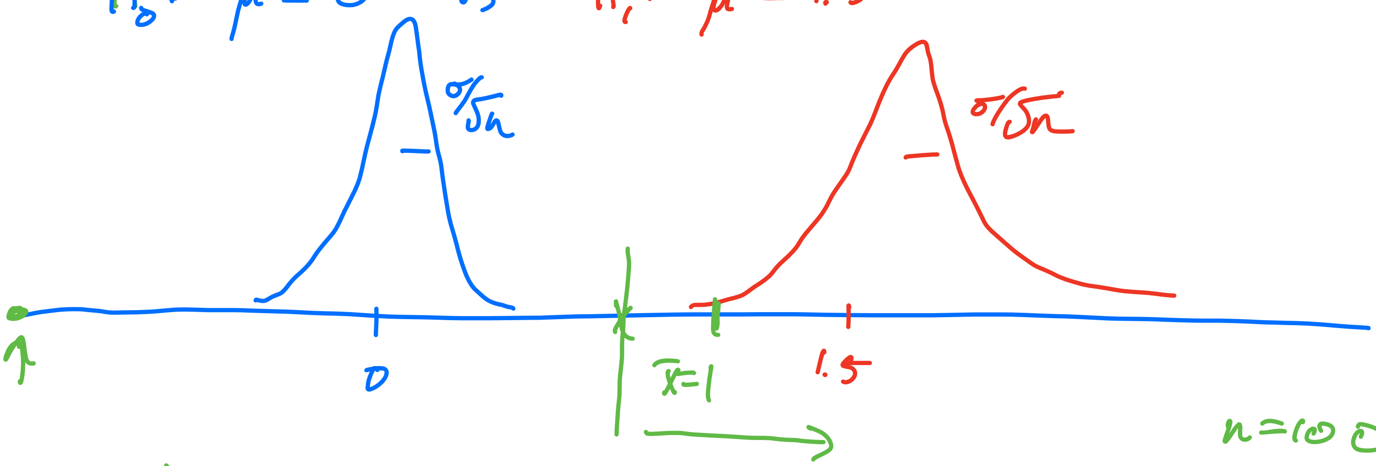
$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$

data dependent statistics



$X_1, X_2, \ldots, X_n$
IID $N(\mu, \sigma^2)$
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
$E(\bar{X})$ $Var(\bar{X})$
$SD = \sigma/\sqrt{n}$

$H_0: \mu = 0$ vs $H_1: \mu = 1.5$



$n = 100$

Compute p-value: Assume null hypothesis

"Prob of observing data at least as extreme as observed data, assuming NULL is true"

$\underline{\mathbb{P}(\bar{X} \geq 1)}$

Normal $\left(0, \tau \right.$ scale = σ estimated $\left.\right)$

SD = ?

$$\sigma = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2} = \text{np.std}(x)$$

$$\frac{\sigma}{\sqrt{100}}$$ (If $n=100$)

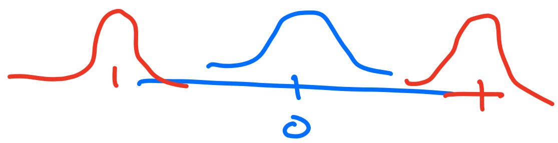
NOT QUITE CORRECT BUT CLOSE

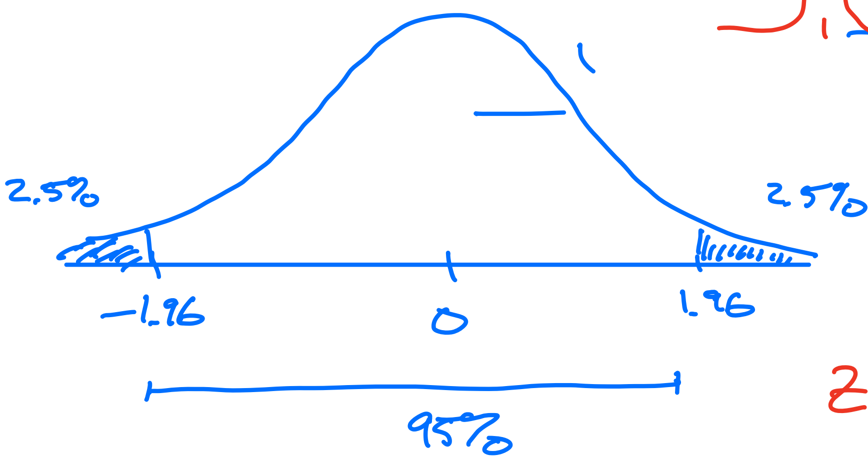
---

z-test μ_0 expected value under null

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$ ← SD of $\bar{x}$

$Z \sim N(0,1)$





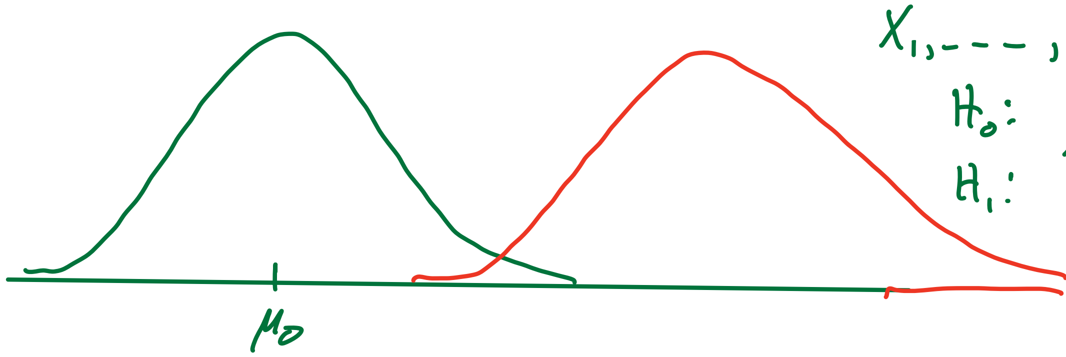
$Z < -2$ OR $Z > 2$ occurs w/ prob 0.05 If null is correct

$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$ t-distribution w/ n-1 deg. of freedom

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

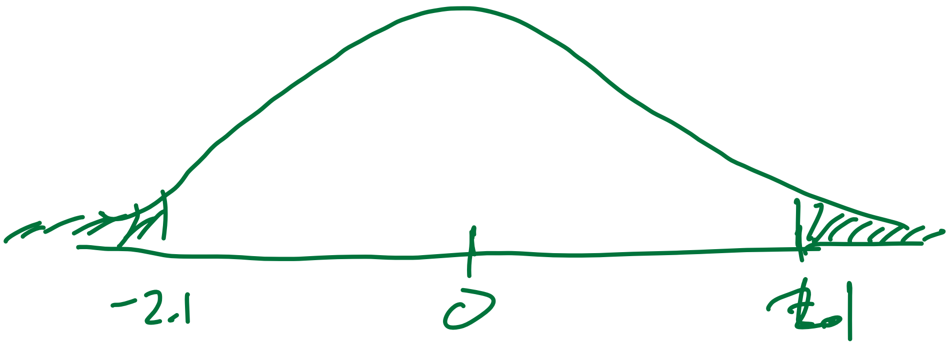
$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$



Known prob distn

$t = 2.1$



Confidence intervals

$X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$

$\mathbb{P}\left(-1.96 \le \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le 1.96\right) = 0.95$

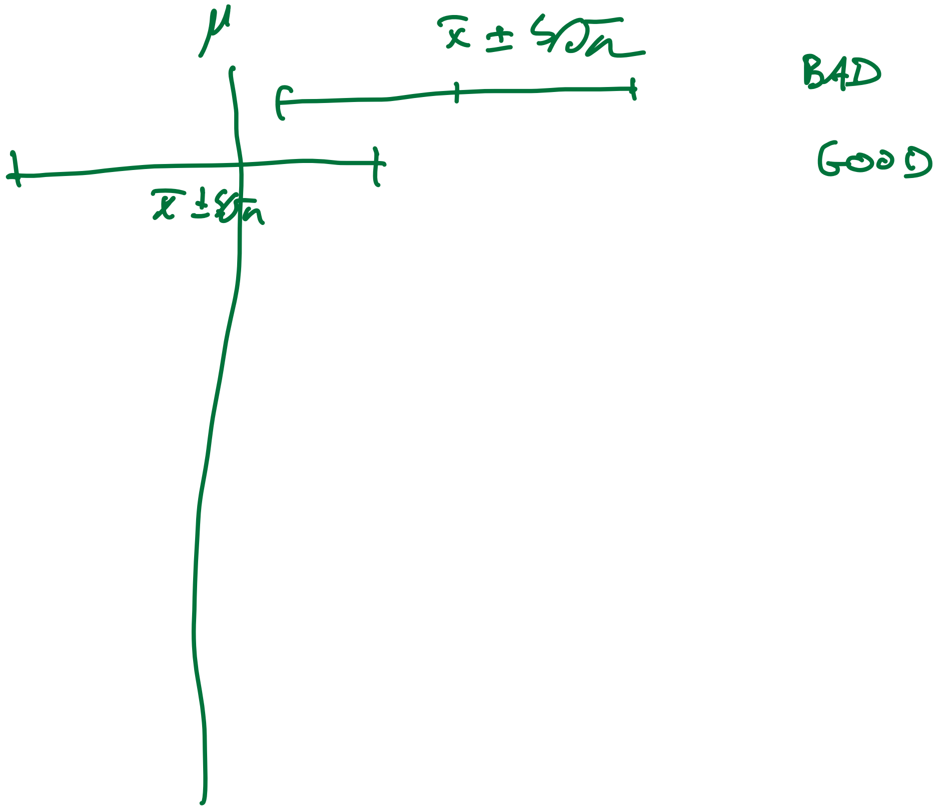
$\hookrightarrow N(0,1)$

$\mathbb{P}\left(\bar{X} - 1.96\,\sigma/\sqrt{n} \le \mu \le \bar{X} + 1.96\,\sigma/\sqrt{n}\right) = 0.95$

$\bar{x} \pm 1.96\,\sigma/\sqrt{n}$ is a 95% conf. interval

$\bar{x} \pm \boxed{2}\, s/\sqrt{n}$ $\qquad s = \hat{\sigma} = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$

↑ t distribution



μ
$\bar{x} \pm s/\sqrt{n}$
BAD
GOOD
$\bar{x} \pm s/\sqrt{n}$