Lecture 6:

standardize the estimator.

DGP: $X_1, \dots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$

$$\hat{\theta} = \bar{X} \underset{\text{Math 241}}{\sim} N\left(\theta, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right) \iff \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$\frac{\sigma}{\sqrt{n}} \to SE[\bar{X}]$

What if σ is unknown?

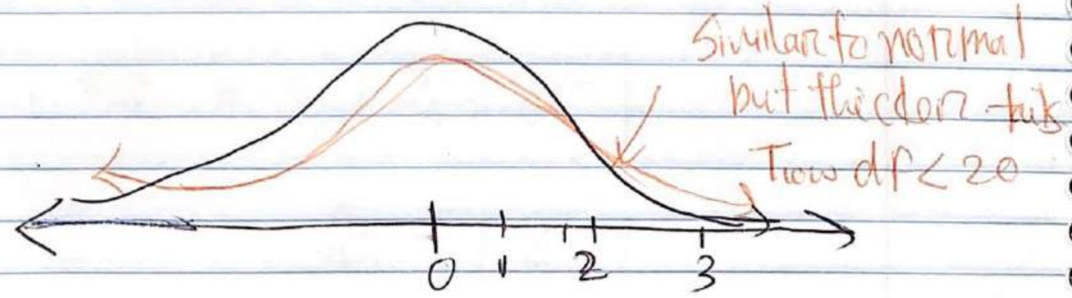
S^2 estimates $\sigma^2 \Rightarrow S = \sqrt{S^2}$ estimates σ.

In 1907 Gosset Proved:

$$\frac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \sim T_{n-1}$$

Student's standard T distribution with n-1 "degree of freedom"

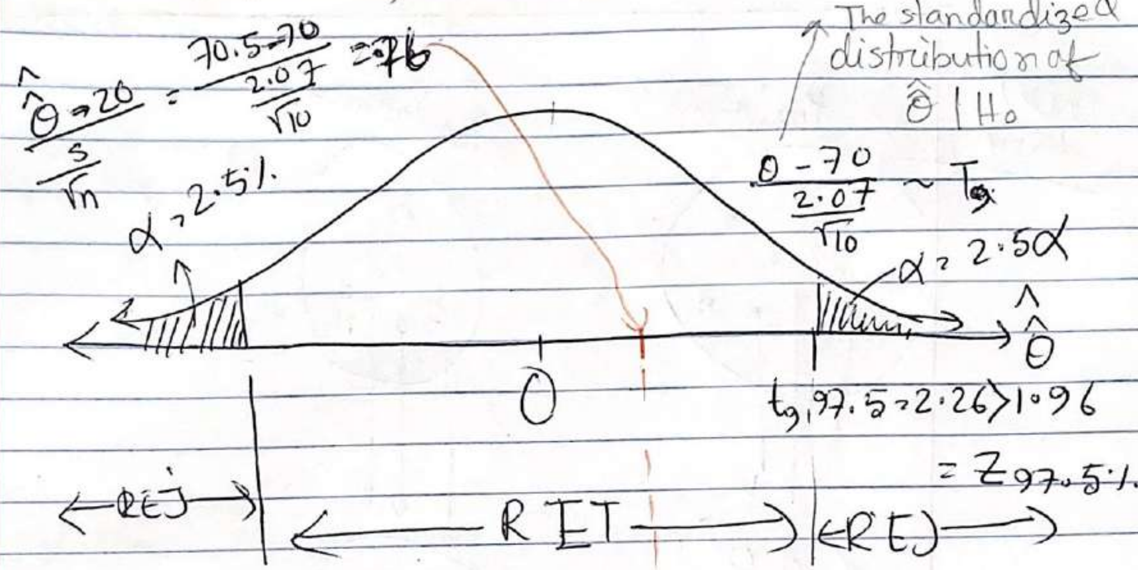
(the parameter for the standard T distn)



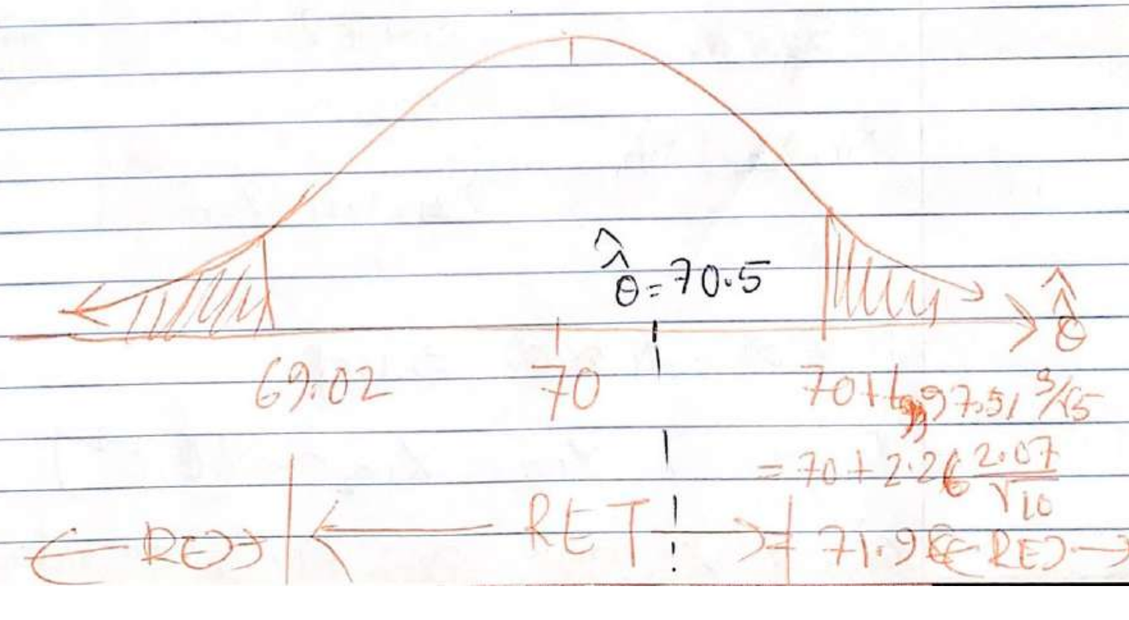
data from n=10 males student heights:

$\bar{X} = 70.5$, $s = 2.07$

$H_a: \theta \neq 70$, $H_0: \theta = 70$, $\alpha = 5\%$



$\Rightarrow$ REJ H_0

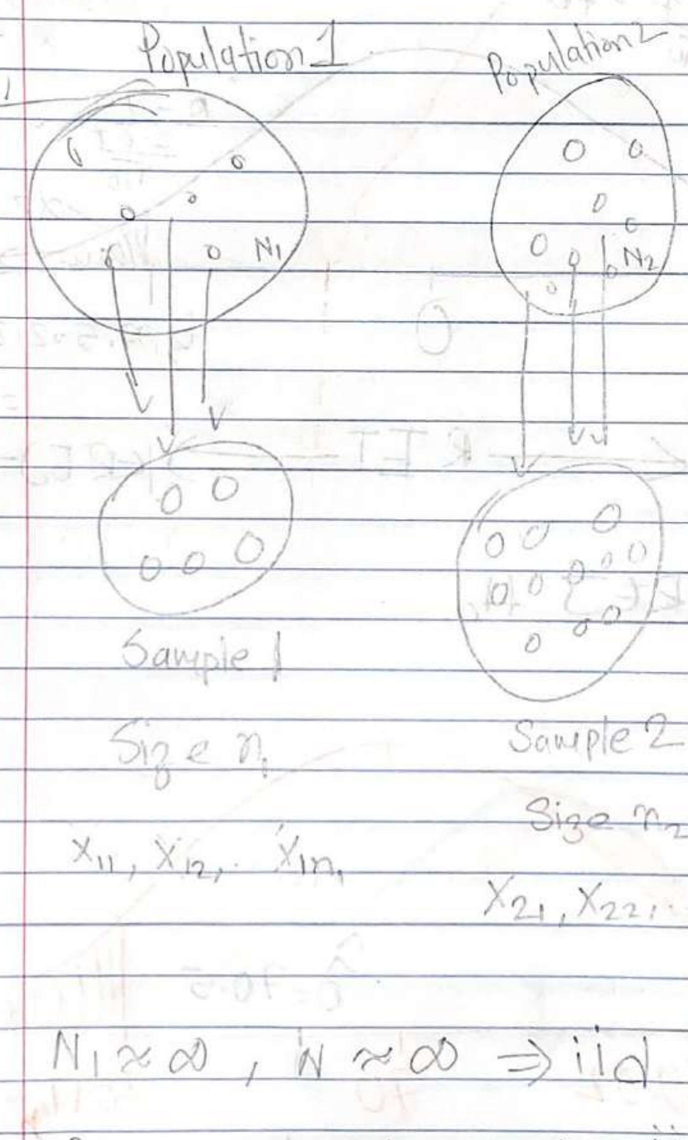


$\Rightarrow$ Retain H_0

We just did our first "One sample two-sided t test" (of a mean).

---

new topic,



$X_{11}, X_{12}, \dots X_{1n_1}$

$X_{21}, X_{22}, \dots X_{2n}$

$N_1 \approx \infty$, $N \approx \infty \Rightarrow$ iid

Assume, $X_{11}, \dots X_{1n_2} \overset{iid}{\sim} N(\theta, \sigma^2)$

σ_1^2, σ_2^2 is know but θ_1, θ_2 are unknown.

independent of

$X_{21}, \ldots X_{2n_2} \overset{iid}{\sim} N(\theta_2, \sigma_2^2)$

There are 3 types of tests that are usually done:-

① $H_a: \theta_1 \neq \theta_2 \Rightarrow H_0: \theta_1 = \theta_2$ equivalently.

$H_a: \theta_1 - \theta_2 \neq 0 \Rightarrow H_0: \theta_1 - \theta_2 = 0$

II $H_a: \theta_1 < \theta_2 \Rightarrow H_0: \theta_1 \geq \theta_2$ equavalent.

$H_a: \theta_1 - \theta_2 < 0 \Rightarrow H_0\ \theta_1 - \theta_2 \geq 0,$

III $H_a: \theta_1 > \theta_2 \Rightarrow H_0: \theta_1 \leq \theta_2$ equivalently

$H_a: \theta_1 - \theta_2 > 0 \Rightarrow H_0: \theta_1 - \theta_2 \leq 0.$

What is a test statistic? (For $\theta_1 - \theta_2$)

$\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2$ Point estimate.

The estimator that produces these estimates is simply:

$\hat{\theta}_1 \sim N(\theta_1, \frac{\sigma_1^2}{n_1})$ ind. $\hat{\theta}_2 \sim N(\theta_2, \frac{\sigma_2^2}{n_2})$

Math 241 (difference of 2 normal)

$$\hat{\theta}_1 - \hat{\theta}_2 = \bar{X}_1 - \bar{X}_2 \underset{\text{exactly}}{\sim} N\left(\theta_1 - \theta_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\Rightarrow SE[\hat{\theta}_1 - \hat{\theta}_2] = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n_2}}$$

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

under H_0 (all three), $\theta_1 - \theta_2 = 0$

$\hookrightarrow \hat{\theta}_1 - \hat{\theta}_2 \sim N(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

Standardized $\rightarrow$ $$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Let's test if male mean height is different than female mean height.

$\vec{X}_2 = \langle 60, 59, 64, 64, 64, 63 \rangle$

LoV9 → From male

$n_2 = 6$ $\quad \bar{X}_2 = 62.3$ $\quad n_1 = 10, \bar{X}_1 = 70.5$

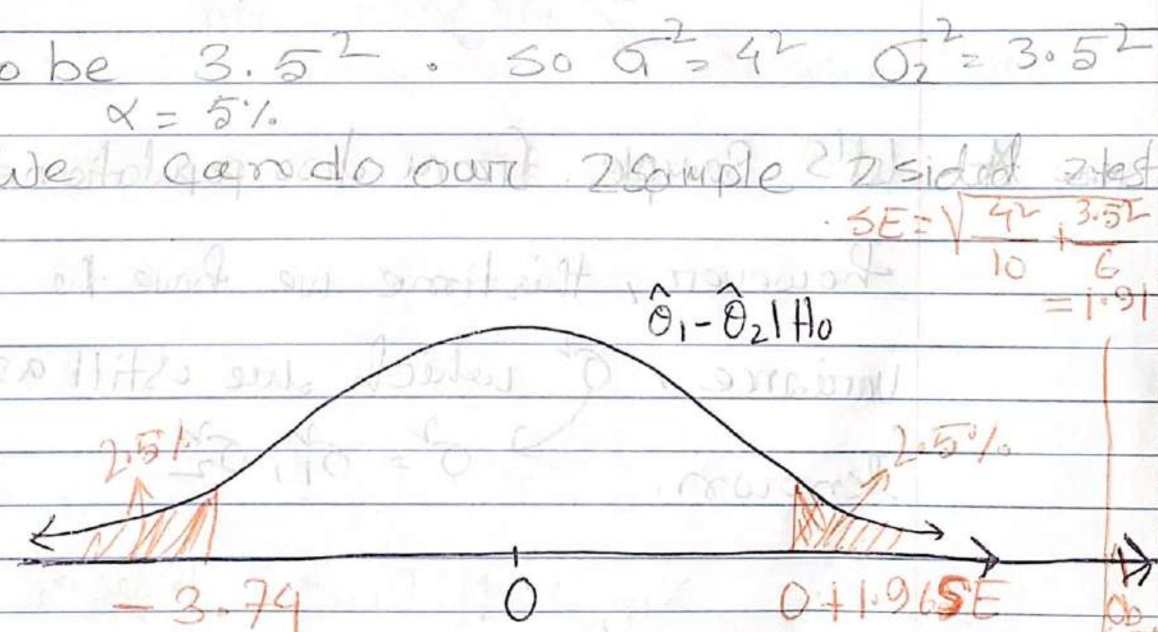
$\hat{\theta}_1 - \hat{\theta}_2 = 70.5 - 62.3 = 8.2$

We assumed we knew the variances. So the variance for the men was assumed to be 4^2 and now the variance for the women is assumed to be 3.5^2. So $\sigma_1^2 = 4^2$ $\quad \sigma_2^2 = 3.5^2$

$\alpha = 5\%$

We can do our 2 sample 2 sided z test

$SE = \sqrt{\frac{4^2}{10} + \frac{3.5^2}{6}} = 1.91$



$= 1.96 \times 1.91$ ← Observed estimate

$= 3.74$

← Rej → | ← RET → | ← Rej →

Observed estimate

$\hat{\theta}_1 - \hat{\theta}_2 \notin RET \Rightarrow$ Reject H_0

PVal $= 2P(\hat{\theta}_1 - \hat{\theta}_2 > 8.2)$

$= 2P\left(\frac{\hat{\theta}_1 - \hat{\theta}_2}{SE} > \frac{8.2}{1.91}\right)$

$= 2P(Z > 4.29)$

$= 1.8 \times 10^{-5} < \alpha$

$\Rightarrow$ Reject H_0

New ✱ Let's Sample from two populations again however, this time we have the same variance, σ^2 which we still assume known. $\sigma^2 = \sigma_1^2, \sigma_2^2$

$X_{11}, \ldots, X_{1n_1} \overset{iid}{\sim} N(\theta_1, \sigma^2)$ indep of $X_{21}, \ldots,$

$X_{2n_2} \overset{iid}{\sim} N(\theta_2, \sigma^2)$

Under H_0,

$$\hat{\theta}_1 - \hat{\theta}_2 \sim N\left(0, \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}^2\right)$$

also, $$\frac{\theta_1 - \theta_2}{\sigma\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \sim N(0,1)$$

2 Sample 2 sided z test of equal variance

The test can be run again, you can probably assume $\sigma = 3.75$.

Same as above but σ^2 unknown. How can we estimate the standard error?

S_1^2 & S_2^2 are the sample variances in both samples 1 and 2.

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1,i} - \bar{X}_1)^2$$

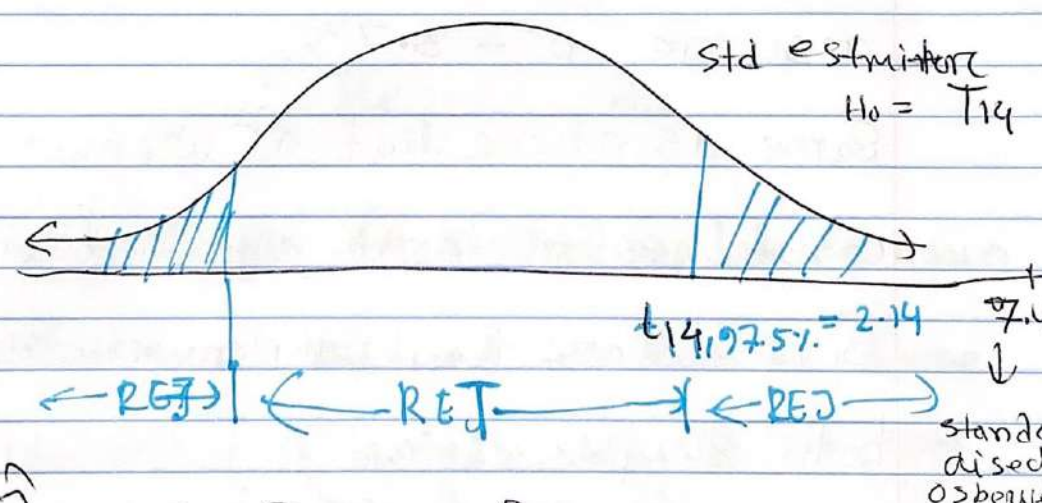
$$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2,i} - \bar{X}_2)^2$$

$$S^2_{pooled} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$ Weighted average.

You can prove that

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{S_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim T_{n_1+n_2-2}.$$

this allows you to do the "2-Sample t test of equal variance"



$$\hat{\theta} = \frac{70.5 - 62.3}{S_{pooled}\sqrt{\frac{1}{10}+\frac{1}{6}}} = \frac{8.2}{2.14 \times 0.51} = 7.42$$

$$S_{pooled} = \frac{9 \times 2.07^2 + 5 \times 2.25^2}{14}$$

$$= 4.56$$