Victoria Lombardi

9/16/20 ①

Math 369
Lecture 6

Standardize the estimator

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

$SE[\bar{X}]$

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

What if sigma is unknown?

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

Does $\dfrac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \overset{?}{\sim} N(0,1)$ NO! But close!

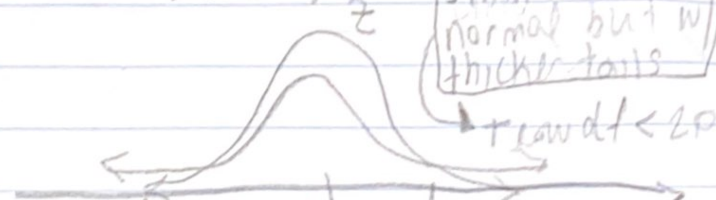
In 1907 Gosset proved:

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

exactly

student's "standard" T distribution with n-1 degrees of freedom (the parameter for the standard T distr.

Similar to normal but w/ thicker tails



data from n=10 male student heights

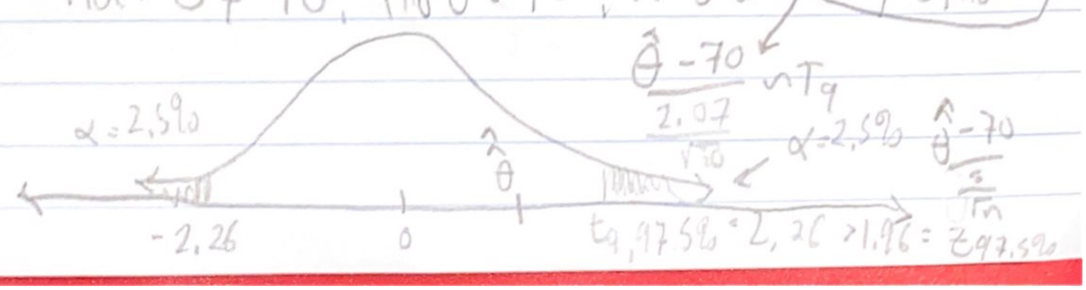
$\bar{x} = 70.5 \quad S = 2.07$

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

the standardized dist. of $\hat{\theta} \mid H_0$

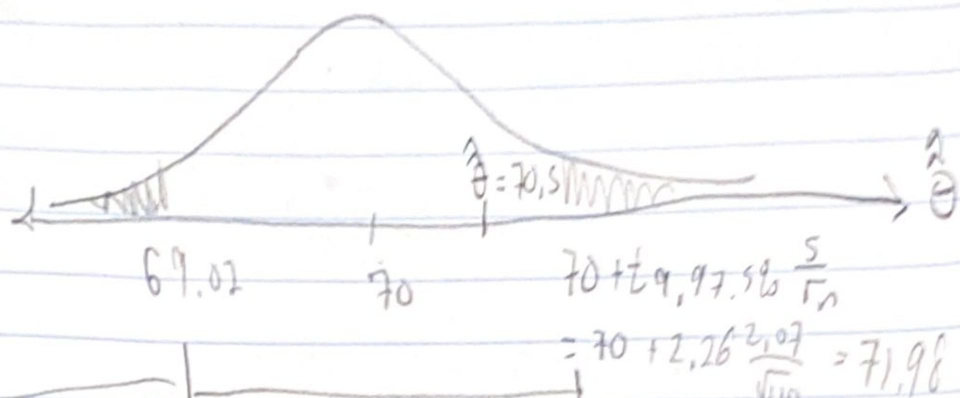
$$\frac{\hat{\theta} - 70}{\frac{2.07}{\sqrt{10}}} \sim T_9$$

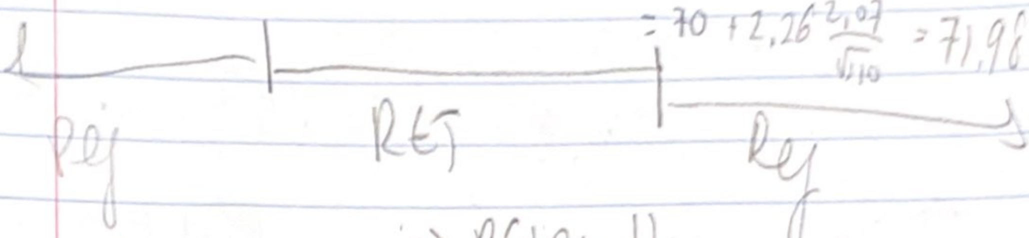
$\dfrac{\hat{\theta} - 70}{\frac{S}{\sqrt{n}}}$



$\frac{\hat{\theta} - 70}{\frac{s}{\sqrt{n}}} = \frac{70.5 - 70}{\frac{2.07}{\sqrt{10}}} = .76 \Rightarrow$ Retain Ho

2

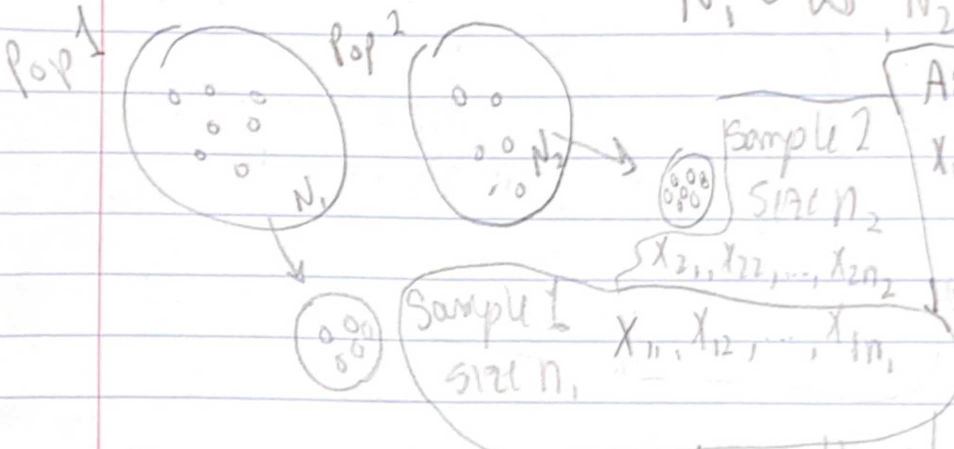




$\Rightarrow$ RETAIN Ho

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

$N_1 \approx \infty$, $N_2 \approx \infty \Rightarrow$ iid



Assume
$X_{11}, \dots, X_{1n_1} \overset{iid}{\sim} N(\theta_1, \sigma_1^2)$
independent of
$X_{21}, \dots, X_{2n_2} \overset{iid}{\sim} N(\theta_2, \sigma_2^2)$
σ_1^2, σ_2^2 known
θ_1, θ_2 unknown

There are 3 types of tests that are usually done.

I) Ha: $\theta_1 \neq \theta_2 \Rightarrow$ Ho: $\theta_1 = \theta_2$ equivalently
Ha: $\theta_1 - \theta_2 \neq 0 \Rightarrow$ Ho: $\theta_1 - \theta_2 = 0$

③

(II) $H_a: \theta_1 < \theta_2 \Rightarrow H_0: \theta_1 \geq \theta_2$ equivalently

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

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

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

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

$\Rightarrow SE[\hat{\theta}_1 - \hat{\theta}_2] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Recall:

$\hat{\theta}_1 \sim N\left(\theta_1, \frac{\sigma_1^2}{n_1}\right)$

indep $\hat{\theta}_2 \sim N\left(\theta_2, \frac{\sigma_2^2}{n_2}\right)$

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

Under H_0 (all 3), $\theta_1 - \theta_2 = 0$

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

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

Lets test if male mean height is different than female mean height.

$\bar{X}_2 = \langle 60, 59, 64, 64, 63 \rangle$ $n_2 = 6$ $\bar{X}_2 = 62.3$

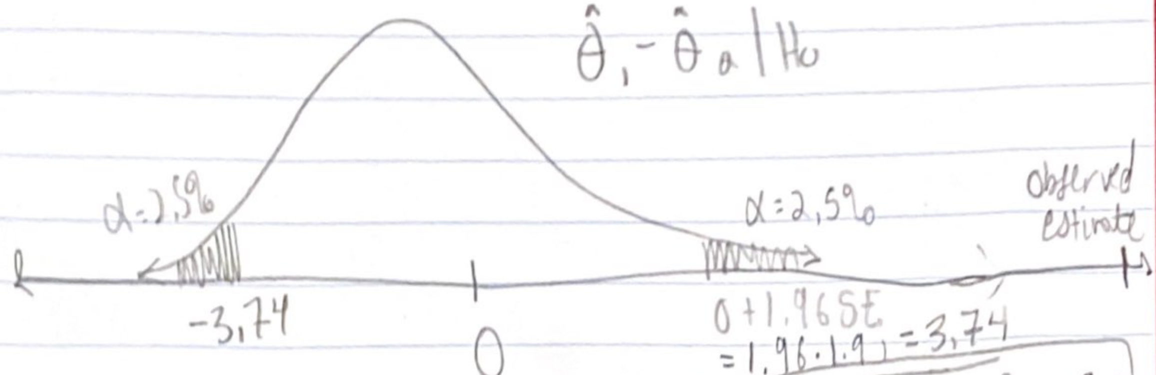
$n_1 = 10$ $\bar{X}_1 = 70.5$

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

④

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, $\sigma^2_1 = 4^2$, $\sigma^2_2 = 3.5^2$, $\alpha = 5\%$

We can now do our 2-sample 2sided z test



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

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

Lets sample from two populations again however, this time we have the same variance, σ^2 which we still assume know

$$\boxed{\sigma^2 = \sigma^2_1 = \sigma^2_2}$$

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

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

Under H_0, $\hat{\theta}_1 - \hat{\theta}_2 \sim N\left(0, \sqrt{\sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}^2\right)$

also $\dfrac{\hat{\theta}_1 - \hat{\theta}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$ 2 sample 2-sided Z test of equal variance

This test can be run again, you can prob assume $\sigma = 3.75$

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

(5)

S^2_1, S^2_2 are the sample variances in both samples 1 + 2.

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

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

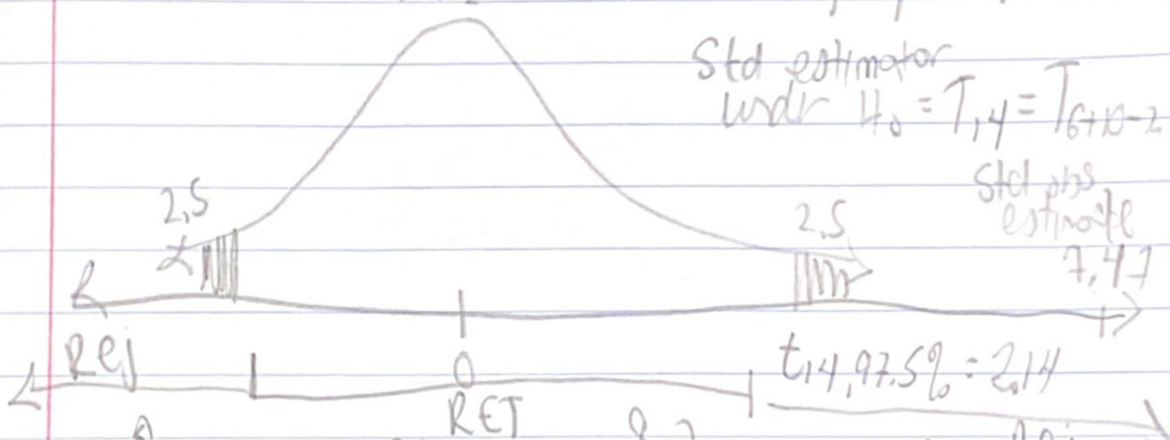
$$S^2_{pooled} := \frac{(n_1-1)S^2_1+(n_2-1)S^2_2}{n_1+n_2-2}$$ weighted average

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"

Std estimator under $H_0 = T_{14} = T_{6+10-2}$



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

$$S^2_{pooled} = \frac{9\cdot 2.07^2+5\cdot 2.25^2}{14} = 4.56$$

$\Rightarrow$ Reject H_0