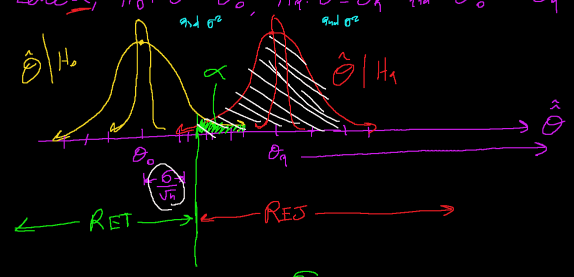
Power in the 1-sample z-test. DGP: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$
where sigsq is known and sample size n.

Level α, $H_0: \theta = \theta_0$ and σ^2, $H_a: \theta = \theta_a$ and σ^2 and $\theta_0 < \theta_a$



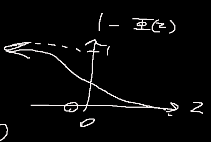
$\underbrace{\theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}}$

$$Pow := P\left(\hat\theta | H_a > \theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\frac{\hat\theta | H_a - \theta_a}{\frac{\sigma}{\sqrt{n}}} > \frac{\theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}} - \theta_a}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P\left(Z > -\frac{\sqrt{n}}{\sigma}(\theta_a - \theta_0) + z_{1-\alpha}\right)$$

$$= 1 - \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta_a - \theta_0) + z_{1-\alpha}\right)$$



If $n \to \infty \Rightarrow Pow \to 1$

If $\alpha \to 0 \Rightarrow z_{1-\alpha} \to \infty \Rightarrow Pow \to 0$

If $\theta_a \to \infty \Rightarrow Pow \to 1$

If $\sigma \to 0 \Rightarrow Pow \to 1$

We assumed sigsq to be known. Is that realistic?? No! What do we do? Sigsq is called a "nuisance parameter" i.e. it's not our target for inference (which is θ) but we have to still estimate it to infer θ. How can we estimate sigsq?

$\theta = E[X]$, $\hat\theta = \frac{1}{n}\sum_{i=1}^n X_i$

$\sigma^2 = E[(X-\theta)^2]$, $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \theta)^2$ but we don't know θ!!!

try again: $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar X)^2$ $\hat\theta = \bar X$ seems reasonable!

Is it unbiased? Let's assume $X_1, \ldots, X_n \overset{iid}{\sim}$ mean θ, variance σ^2

$$E[\hat\sigma^2] = E\left[\frac{1}{n}\sum_{i=1}^n (X_i - \bar X)^2\right] = \frac{1}{n}\sum_{i=1}^n E[(X_i - \bar X)^2]$$

$$\overset{iid}{=} \cancel{\frac{1}{n}}\cancel{n} E[(X_1 - \bar X)^2] = E[X_1^2 - 2X_1\bar X + \bar X^2]$$

$$= E[X_1^2] - 2E\left[X_1\left(\frac{X_1 + \ldots + X_n}{n}\right)\right] + E[\bar X^2]$$

$$= E[X_1^2] - \frac{2}{n}\left(E[X_1^2] + \underbrace{E[X_1X_2]}_{E[X_1]E[X_2]} + \ldots + E[X_1X_n]\right) + E[\bar X^2]$$ due to independence

$\left[\ Var[Y] = E[Y^2] - \mu^2 \Rightarrow E[Y^2] = Var[Y] + \mu^2\right.$

$$= (\sigma^2 + \theta^2) - \frac{2}{n}\left(\sigma^2 + \theta^2 + \overbrace{\theta^2 + \ldots + \theta^2}^{n-1}\right) + \frac{\sigma^2}{n} + \theta^2$$

$$= \sigma^2 + \cancel{\theta^2} - 2\frac{\sigma^2}{n} - \cancel{2\theta^2} + \frac{\sigma^2}{n} + \cancel{\theta^2} = \left(1 - \frac{1}{n}\right)\sigma^2$$

$= \frac{n-1}{n}\sigma^2 < \sigma^2$ i.e. it's biased downward (towards zero)

An estimator is asymptotically unbiased if $\lim_{n \to \infty} E[\hat\theta_n] = \theta$

Can I "fix" this reasonable estimator? Yes! Bessel's Correction

let $S^2 = \boxed{\frac{n}{n-1}}\hat\sigma^2 = \frac{n}{n-1}\frac{1}{n}\sum(X_i - \bar X)^2 = \frac{1}{n-1}\sum(X_i - \bar X)^2$ sample variance estimator

$E[S^2] = E\left[\frac{n}{n-1}\hat\sigma^2\right] = \frac{n}{n-1}E[\hat\sigma^2] = \cancel{\frac{n}{n-1}}\cancel{\frac{n-1}{n}}\sigma^2 = \sigma^2$ unbiased!!!

How can we estimate sigma? $S = \sqrt{S^2}$ to estimate σ?

Yes! This is what we do. However, it is biased... $= \sqrt{\frac{1}{n-1}\sum(X_i - \bar X)^2}$

Back to the main story... our DGP is iid normal with both mean and variance unknown. The variance is a nuisance parameter.

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

Is $\frac{\hat\theta - \theta}{\frac{S}{\sqrt{n}}} \overset{?}{\sim} N(0,1)$ No.... but $\frac{\hat\theta - \theta}{\frac{S}{\sqrt{n}}} \sim T_{n-1}$

This is called the standard T-distribution also called "Student's T distribution" with "n-1 degrees of freedom".

Note: the derivation of this distribution is in MATH 368 and it literally takes the whole semester.

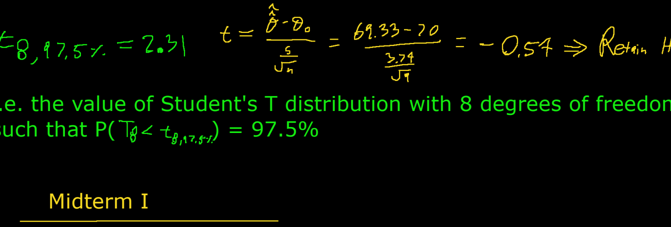
This test will be called the 1-sample t-test. Let's do our first t-test.

Heights of male students: 67, 75, 64, 66, 74, 72, 67, 69, 70",
n = 9, xbar = 69.33, s = 3.74

We want to test if the class's male students are "different" than the American male population whose mean is 70".

$H_a: \theta \neq 70$, $H_0: \theta = 70$, $\alpha = 5\%$
$S = 3.74$

We'll do this test on the standardized scale



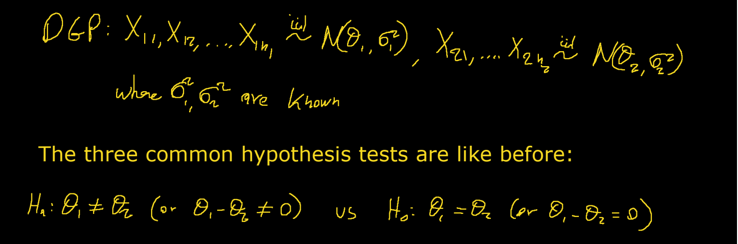
$t_{8, 97.5\%} = 2.31$ $t = \frac{\hat\theta - \theta_0}{\frac{s}{\sqrt{n}}} = \frac{69.33 - 70}{\frac{3.74}{\sqrt{9}}} = -0.54 \Rightarrow$ Retain H_0.

i.e. the value of Student's T distribution with 8 degrees of freedom such that P($T_8 < t_{8, 97.5\%}$) = 97.5%

Midterm I

Midterm II

Let's consider sampling from two possibly different populations



DGP: $X_{11}, X_{12}, \ldots, X_{1n_1} \overset{iid}{\sim} N(\theta_1, \sigma_1^2)$, $X_{21}, \ldots, X_{2n_2} \overset{iid}{\sim} N(\theta_2, \sigma_2^2)$

where σ_1^2, σ_2^2 are known

The three common hypothesis tests are like before:

$H_a: \theta_1 \neq \theta_2$ (or $\theta_1 - \theta_2 \neq 0$) vs $H_0: \theta_1 = \theta_2$ (or $\theta_1 - \theta_2 = 0$)

two-sided, two-sample test

$H_a: \theta_1 < \theta_2$ (or $\theta_1 - \theta_2 < 0$) vs. $H_0: \theta_1 \geq \theta_2$ (or $\theta_1 - \theta_2 \geq 0$)

left-sided, two-sample test

$H_a: \theta_1 > \theta_2$ (or $\theta_1 - \theta_2 > 0$) vs. $H_0: \theta_1 \leq \theta_2$ (or $\theta_1 - \theta_2 \leq 0$)

right-sided, two-sample test