Math 341/641

Review of 241. We will prove these again in 340...

$X \sim N(\mu, \sigma^2) \Rightarrow a + bX \sim N(\mu+a, b^2\sigma^2)$

$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow X_1 + \ldots + X_n \sim N(n\mu, n\sigma^2)$

$\Rightarrow \bar{X} = \frac{1}{n}(X_1 + \ldots + X_n) \sim N(\mu, \frac{\sigma^2}{n})$

How to run a $\alpha = 5\%$ test with Binomial? You can't get size exactly = to α. So pick the min size s.t. size $\geq \alpha$. This is only a problem with discrete null distr's. Not all levels are "attainable"!

New type of survey? For men, how tall are you?

$X_1 = \quad , x_2 = \quad , \ldots \ldots$ $\quad \hat{\theta} = \bar{x} = \frac{825}{17} = 68.85''$

Let's pretend we've done an SRS of all American men.

Height is known biologically to be iid normal. And the American $\sigma^2 = 16 in^2$ which we will assume. Thus... $X_1, \ldots X_n \overset{iid}{\sim} N(\theta, 4)$. Tests:

$H_a: \theta \neq \theta_0 \Longleftrightarrow H_0: \theta = \theta_0$ "two sided one sample exact Z test of means"

$H_a: \theta > \theta_0 \Longleftrightarrow H_0: \theta \leq \theta_0$ "right-sided " " " " " " " "

$H_a: \theta < \theta_0 \Longleftrightarrow H_0: \theta \geq \theta$ "left sided " " " " " " " "

"Exact" since we know $\hat{\theta} \sim N(\theta, \frac{\sigma^2}{n})$ exactly. "Z" since that's a synonym for "normal distr."

Let's test if this class's pop mean is different than the American pop mean of $\theta = 70 in$.

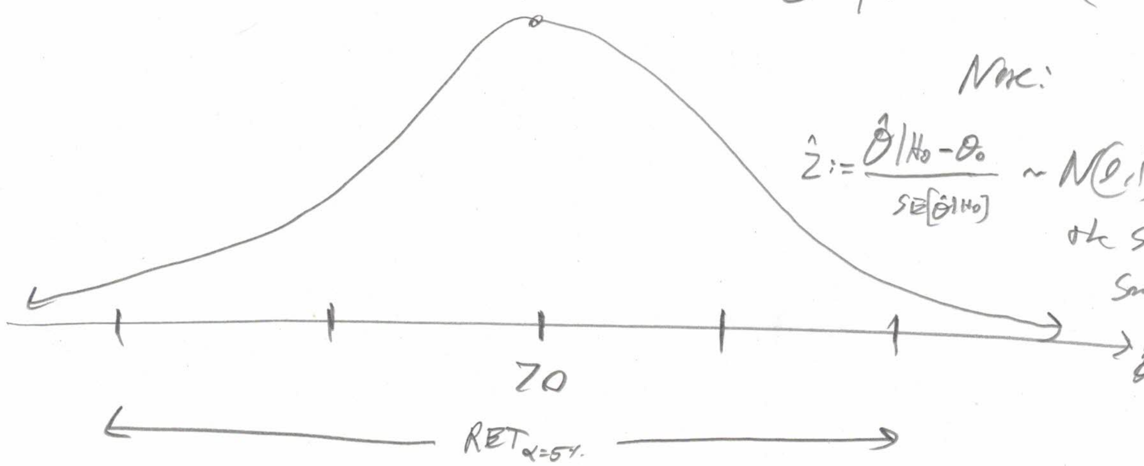
This test doesn't have a name since it is not common to have iid N DGP.

$SE[\hat{\theta}]: \sqrt{Var[\hat{\theta}]} = \frac{\sigma}{\sqrt{n}} = \frac{4in}{\sqrt{13}} =$

What is sampling distr?

Null Sampling Distr / Null distr:

$$\hat{\theta} \mid H_0 \sim N(70in, 1.109^2)$$



Note:

$$\hat{z} := \frac{\hat{\theta}\mid H_0 - \theta_0}{SE[\hat{\theta}\mid H_0]} \sim N(0,1)$$

the standardized sampling distr.

Note: $\hat{\theta} \mid H_0$ is continuous $\Rightarrow$ all sizes $\in (0,1)$ are possible $\Rightarrow$ all levels are "attainable".

$\Rightarrow$ size = level, so set $\alpha = 5\%$ which is the scientific convention.

Recall. If $\hat{z} \sim N(0,1) \Rightarrow P(\hat{z} > 1) \approx 16\%$, $P(\hat{z} > 2) \approx 2.5\%$

$$\Rightarrow P(\hat{z} < -2) + P(\hat{z} > 2) \approx 5\% = \alpha$$

$$\Rightarrow RET^{std}_{\alpha=5\%} = \{\hat{z} : \hat{z} \in [-2, 2]\} = \{\hat{z} : \hat{z}\, SE[\hat{\theta}] + \theta_0 \in [-2\,SE[\hat{\theta}] + \theta_0, 2\,SE[\hat{\theta}] + \theta_0]\}$$

$$\Rightarrow RET_{\alpha=5\%} = \{\hat{\theta} : \hat{\theta} \in [-2\,SE[\hat{\theta}] + \theta_0, 2\,SE[\hat{\theta}] + \theta_0]\}$$

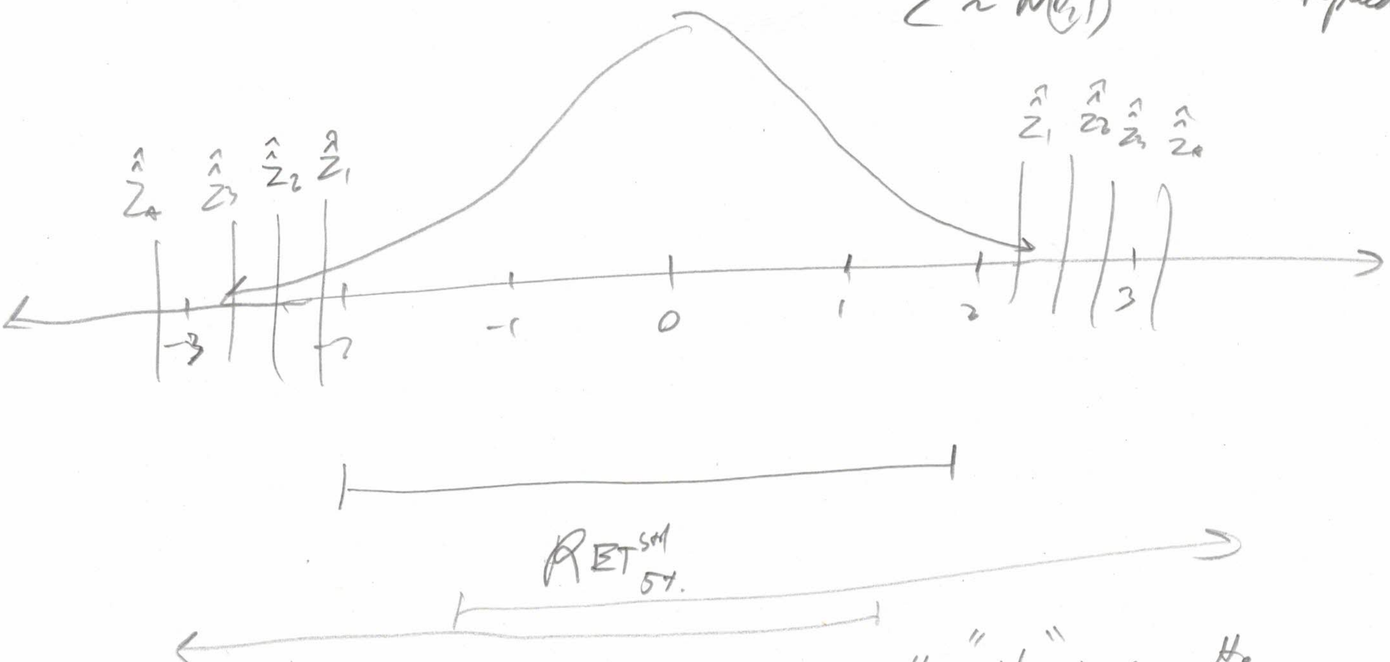
Simply check if $\hat{\theta}$ is in the RET region defined on the original scale

or if $\hat{z} := \frac{\hat{\theta} - \theta_0}{SE[\hat{\theta}]}$ " " " " " RET^{std} " " " " " " standardized scale

Both conventions are employed.

Convention not to write $\hat{Z} | H_0$ as its implied

$\hat{Z} \sim N(0,1)$



As $\hat{\hat{z}}$ decreases or increases, you still "only" reject H_0 but do these rejections seem more unambiguous? Fisher thought so and defined the "p-value".

$$\text{pval} := P(\text{estimate is more "extreme" than the estimate observed} \mid H_0)$$

$$= \underset{\alpha}{\arg\max} \left\{ \alpha : \hat{\hat{z}} \in RET^{sig}_{\alpha} \right\} = \underset{\alpha}{\arg\max} \left\{ \alpha : \hat{\hat{\theta}} \in RET_{\alpha} \right\}$$

for 2-sided test

$$= \underset{\alpha}{\arg\max} \begin{cases} 2P(\hat{Z} > \hat{\hat{z}}) & \text{if } \hat{\hat{z}} > 0 \\ 2P\,\hat{Z} < \hat{\hat{z}} & \text{if } \hat{\hat{z}} < 0 \end{cases} = \underset{\alpha}{\arg\max} \begin{cases} 2P(\hat{\theta} > \hat{\hat{\theta}} \mid H_0) & \text{if } \hat{\hat{\theta}} > \theta_0 \\ 2P(\hat{\theta} < \hat{\hat{\theta}} \mid H_0) & \text{if } \hat{\hat{\theta}} < \theta_0 \end{cases}$$

for left sided i.e. $H_0: \theta \geq \theta_0$

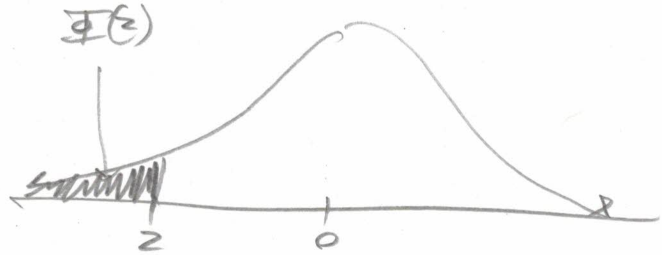
$$= P(\hat{Z} < \hat{\hat{z}}) = P(\hat{\theta} < \hat{\hat{\theta}} \mid H_0)$$

for right-side i.e. $H_0: \theta \leq \theta_0$

$$= P(\hat{Z} > \hat{\hat{z}}) = P(\hat{\theta} > \hat{\hat{\theta}} \mid H_0)$$

To calculate p-values for this DGP, use a "Z-table"
or a the standard normal CDF

$$\Phi(z) := \int_{-\infty}^{z} f_Z(u)\,du$$



Note $pval \geq size \Rightarrow$ Retain H_0, $pval < size \Rightarrow$ Reject H_0
is equivalent to being in the RET and not respectively.

If H_0 is rejected, the estimate is called "statistically significant".

$\Updownarrow$

If estimate < size 〃 〃 〃 〃 〃 〃 〃 〃 〃.

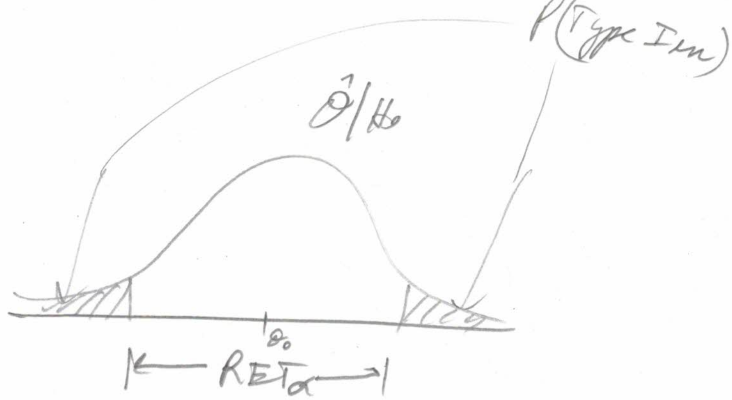
Define $z_p := p$ s.t. $\Phi(z_p) = p$

e.g. $z_{2.5\%} \approx -2$, $z_{8\%} \approx -1$, $z_{50\%} = 0$, $z_{92\%} \approx 1$, $z_{97.5\%} \approx 2$

Recall both types of errors from hypothesis tests

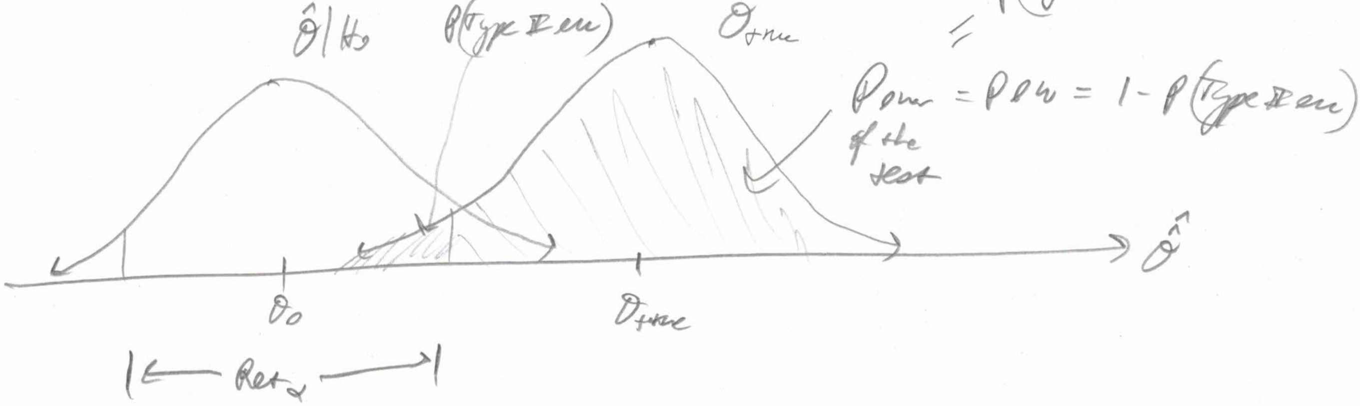
$H_0: \theta = \theta_0$

		Decision	
		Retain H_0	Reject H_0
Reality (unseen)	H_0 true	✓	Type I err.
	H_0 false	Type II err.	✓



What's the P(Type II err)? Impossible to know... unless you know the true value of θ, θ_{true}.

prob. of finding the effect, proving your hypothesis
$P(\text{Reject } H_0 \mid H_0 \text{ is false})$
=

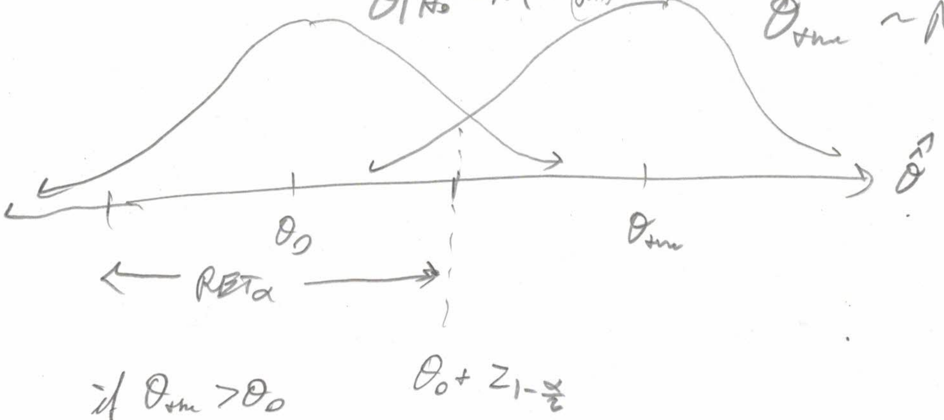


$$P(\text{Type II err}) = P\left(\hat{\theta}_{true} \in RET_\alpha\right)$$

Calc. this for $\theta_{true} = 72''$ in our example

In general... for a 2-sided test $H_0: \theta = \theta_0$ with level α

$\hat{\theta} | H_0 \sim N\left(\theta_0, \left(\frac{\sigma_0}{\sqrt{n}}\right)^2\right)$ $\qquad$ $\hat{\theta}_{true} \sim N\left(\theta_{true}, \left(\frac{\sigma_{true}}{\sqrt{n}}\right)^2\right)$



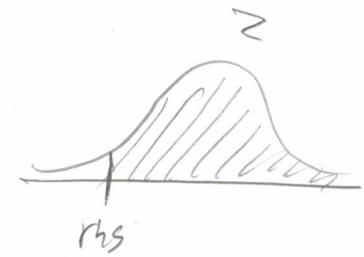
if $\theta_{true} > \theta_0$

$\downarrow$

$$POW = P\left(\hat{\theta}_{true} > \theta_0 + z_{1-\frac{\alpha}{2}} SE[\hat{\theta}|H_0]\right) = P\left(\frac{\hat{\theta}_{true} - \theta_{true}}{SE[\hat{\theta}_{true}]} > \frac{\theta_0 + z_{1-\frac{\alpha}{2}} SE[\hat{\theta}|H_0] - \theta_{true}}{SE[\hat{\theta}_{true}]}\right)$$

$$= P\left(Z > \frac{\theta_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} - \theta_{true}}{\frac{\sigma_{true}}{\sqrt{n}}}\right)$$

$$= P\left(Z > \frac{\sqrt{n}(\theta_0 - \theta_{true}) + z_{1-\frac{\alpha}{2}} \sigma_0}{\sigma_{true}}\right)$$



power function

$$= P\left(Z > \frac{z_{1-\frac{\alpha}{2}} \sigma_0 - \sqrt{n}(\theta_{true} - \theta_0)}{\sigma_{true}}\right) = POW(\theta_0, \theta_{true}, \sigma_0, \sigma_{true}, \alpha, n)$$

$$= 1 - \Phi(\;)$$

How is POW affected by inputs?

If $n \to \infty$, rhs $\to -\infty \Rightarrow POW \to 1 \Rightarrow$ more data $\Rightarrow$ more powerful test

If $\sigma_0 \downarrow \Rightarrow$ rhs $\downarrow \Rightarrow POW \uparrow$

If $\theta_{true} - \theta_0 \uparrow \Rightarrow$ rhs $\downarrow \Rightarrow POW \uparrow$

If $\sigma_{true} \downarrow \Rightarrow$ rhs $\downarrow \Rightarrow POW \uparrow$

If $\alpha \downarrow \Rightarrow$ rhs $\uparrow \Rightarrow POW \downarrow$

How to calc. power in the real world.

The researcher usually sets θ_{true} to a value that would define minimal "clinical significance" or "practical significance" $\theta_{sign.}$ and then compute the appropriate sample size n to achieve 80% power.

i.e.

$$80\% \overset{set}{=} 1 - \Phi\left(\frac{z_{1-\frac{\alpha}{2}}\sigma_0 - \sqrt{n}\,(\theta_{sign} - \theta_0)}{\sigma_{true}}\right)$$

and solve for n using a computer