Lecture 27 (Final) Math 241 12/8/16

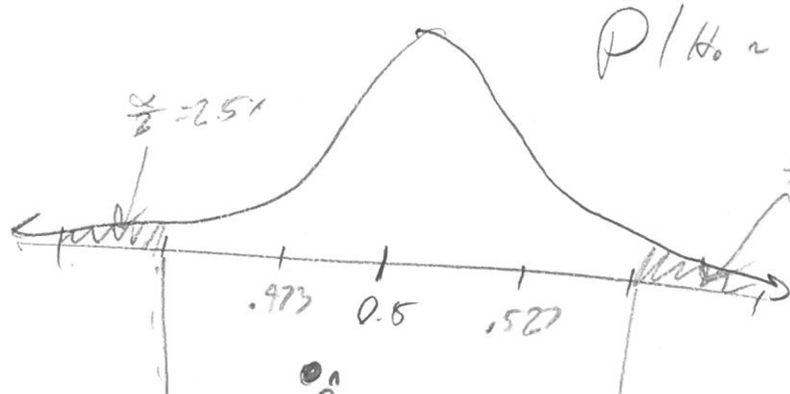
Previously, we ran a hypothesis test for equal human gender proportion $p := P(\text{Male})$

$H_0: p = 0.5$ (default position) ← we don't want to just willy-nilly throw out our theory

$H_1: p \neq 0.5$ (crazy theory)

So we choose $\alpha = 5\%$, sample size $n = 345$

$$\hat{p} \mid H_0 \sim N\left(p, \sqrt{\tfrac{p(1-p)}{n}}^{\,2}\right) = N(0.5, .0269^2)$$



Reject H_0 too extreme less than 0.5

Retain H_0
$= [0.5 \pm 2 \cdot .0269]$
$= [.446, .554]$

Reject H_0 too extreme more than 0.5

Two Sided Test for one proportion

169 male births $\Rightarrow \hat{p} = \frac{169}{345} = .49 \in$ Retainment Region $\Rightarrow$ Fail to reject H_0 — or — Retain H_0

Decision

Truth	Retain H_0	Reject H_0
H_0 true	✓	Type I err
H_0 false	Type II err	✓

$\alpha := P(\text{Type I err})$ you choose this

$P(\text{Type I err}) \uparrow \Rightarrow P(\text{Type II err}) \downarrow$

$P(\text{Type I err}) \downarrow \Rightarrow P(\text{Type II err}) \uparrow$

Could have been here! No way to know!!

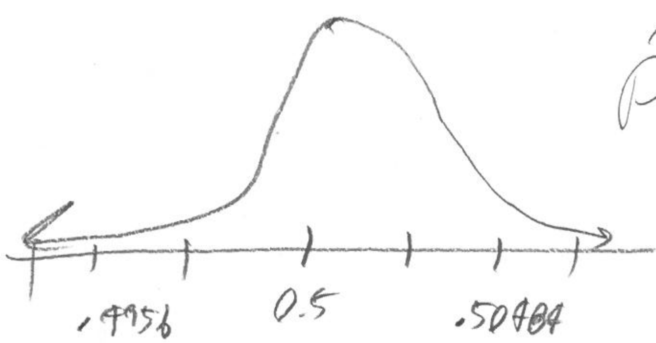
Return to M/F ratio.

$H_0: p = 0.5$

$H_a: p \neq 0.5$

$\alpha = 5\%$

In 2000 in USA $n = 4{,}247{,}000$ babies and 2,173,000 male



$\hat{p} \sim N\left(0.5, \left(\sqrt{\frac{0.5 \cdot 0.5}{n}}\right)^2\right)$

Ret. Region $= [.49516, .50484]$

$\hat{p} = 0.51165 \Rightarrow \hat{p} \notin$ Ret. Region $\Rightarrow$ Rej. H_0. There is sufficient evidence to believe gender ratio unequal. (Scientists can't figure it out)

But I retained H_0 before? Why reject now?? $n \uparrow \Rightarrow$ more change

$n \uparrow \Rightarrow P(\text{Type II err}) \downarrow$ Power $\uparrow$ (ability to find effect / demonstrate false H_0)

$n \uparrow \Rightarrow P(\text{Type I err})$ no change!

Retain H_0 = Accept H_0? No

Before

$H_0: p = 0.5$ or $H_0: p = 0.500001$

$H_a: p \neq 0.5$ $\quad$ $H_a: p \neq 0.500001$

$\alpha = 5\%,\ n = 395$ $\quad$ $\alpha = 5\%,\ n = 395$

H_0 retained $\quad$ H_0 also retained

Accept H_0 for both is contradiction. Accept H_a for both is not a concern.

Not enough evidence to reject $p = 0.5$ or $p = 0.500001$ or so many others

All science works this way. We have a theory until we see enough evidence that it is bunk. We don't throw theories away willy-nilly though. We are conservative. Here α is small.

I) UFO's &

H_0: Aliens Don't Exist

H_a: Aliens Do Exist

α low

II

H_0: Aliens Don't Exist ← "traditional"

H_a: Aliens Do exist

α high

III

H_0: Aliens Do Exist

H_a: Aliens Don't Exist

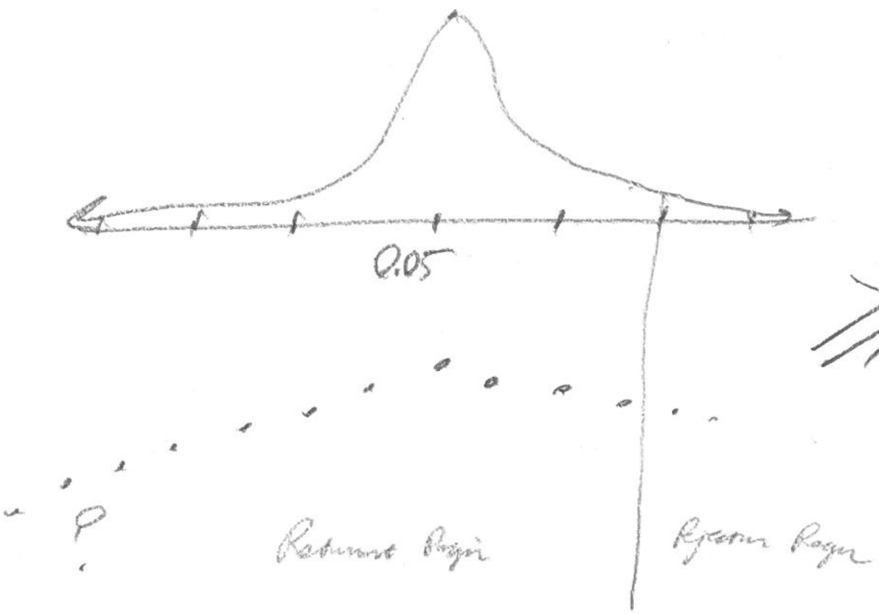
α low

H_0: Aliens Do Exist

H_a: Aliens Don't Exist

α high

If more than 5% of customers don't like a driver, Uber fires them. Decision is made after 1,000 rides.



$$H_0: p \le 0.05 \quad \text{(Good driver)}$$

$$H_a: p > 0.05 \quad \text{(BAD driver)}$$

		Keep	Fire
Driver	Good	✓	Type I err
	Bad	Type II err	✓

Since Rejection Region is only on "one side", this is a "Right Sided / One sided test for one proportion."

$$\alpha = P(\text{Rejection}) = P(\text{Rejection Reg})$$

$$1-\alpha = P(\text{Retention Region}) = P(Z \le z_\alpha) = P\left(p + Z\sqrt{\tfrac{p(1-p)}{n}} \le p + z_\alpha\sqrt{\tfrac{p(1-p)}{n}}\right) = P\left(\hat{p} \le p + z_\alpha\sqrt{\tfrac{p(1-p)}{n}}\right)$$

Recall $z_{\frac{\alpha}{2}} := F_Z^{-1}\left(1-\frac{\alpha}{2}\right) \Rightarrow z_\alpha := F_Z^{-1}(1-\alpha)$

$$\Rightarrow \text{Retention Region} = \left(-\infty, p + z_\alpha\sqrt{\tfrac{p(1-p)}{n}}\right]$$

Run experiment

$\Rightarrow$ if $\hat{p} \in$ Ret Region $\Rightarrow$ Retain H_0 $\Rightarrow$ Do not fire driver

if $\hat{p} \notin$ Ret Region $\Rightarrow$ Reject H_0 $\Rightarrow$ Fire driver

.00689

let $\alpha = 2.5\%$ $z_\alpha = 2$ $\Rightarrow$ Ret Reg $= \left(-\infty, 0.05 + 2\sqrt{\frac{0.05 \cdot .95}{1000}}\right] = (-\infty, .0638]$

$n = 1000$

Why doesn't $.0638 = .05$?

Due to randomness. And we want to be sure we don't commit a Type I error.

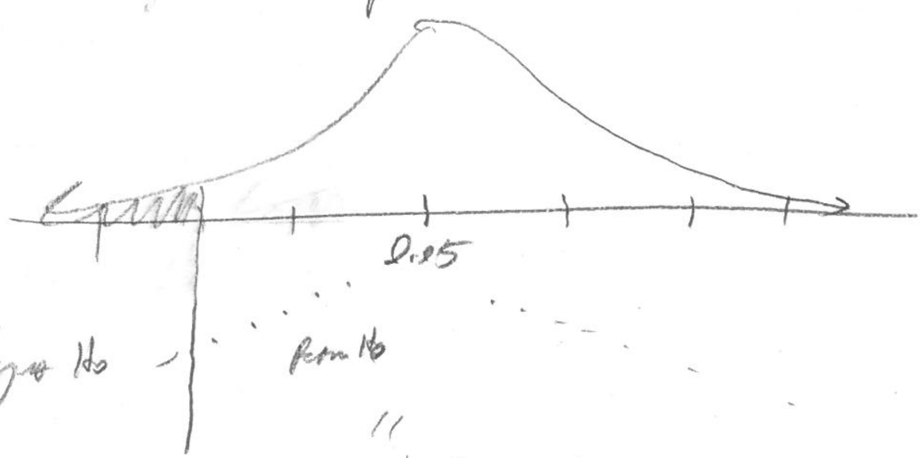
Run experiment # complaints = 71 $\Rightarrow \hat{p} = \frac{71}{1000} = .071 \notin$ Ret Region $\Rightarrow$ Fire Driver

---

There's another way to do this:

H_0: Bad driver $p \geq 0.05$

H_a: Good driver $p < 0.05$



"Left-sided Test of one proportion"

$$\text{Ret. Region} = \left(p - z_\alpha \sqrt{\frac{p(1-p)}{n}}, \infty\right)$$

if $\alpha = 2.5\%$

$$= \left[.05 - 2\sqrt{\frac{.05 \cdot .95}{1000}}, \infty\right) = \left[.0362, \infty\right)$$

if $\hat{p} \in$ Ret Reg $\Rightarrow$ Retain H_0 $\Rightarrow$ Fire driver

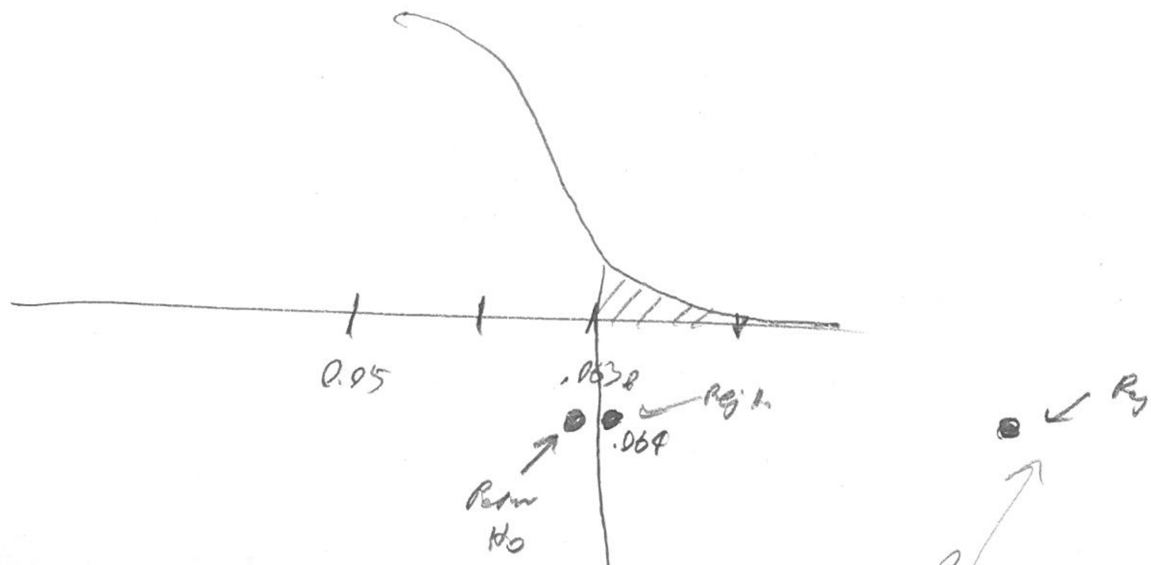
if $\hat{p} \notin$ Ret. Region $\Rightarrow$ Reject H_0 $\Rightarrow$ Keep driver

—— END of Math 241 ——

p-values (not covered on final)

H_0: Good drive $p \le 0.05$

H_a: Bad drive $p > 0.05$



But a much more "emphatic" rejection!

How to distinguish between the two?

$$pval := P(\text{seeing the data "or more extreme"} \mid H_0 \text{ true})$$

$$= P(\hat{P} \ge \hat{p} \mid H_0)$$

$$= P(\hat{P} \ge \hat{p} \mid p = 0.05)$$

$$= P(\hat{P} \ge .071 \mid p = 0.05)$$

$$= P\left(\frac{\hat{P} - .05}{.00689} \ge \frac{.071 - .05}{.00689}\right)$$

$$= P(Z \ge 3.05) \approx .0018 \approx 0.1\% < \alpha = 2.5\% \Rightarrow \text{Reject}$$

Highly contested! But in use all the time!