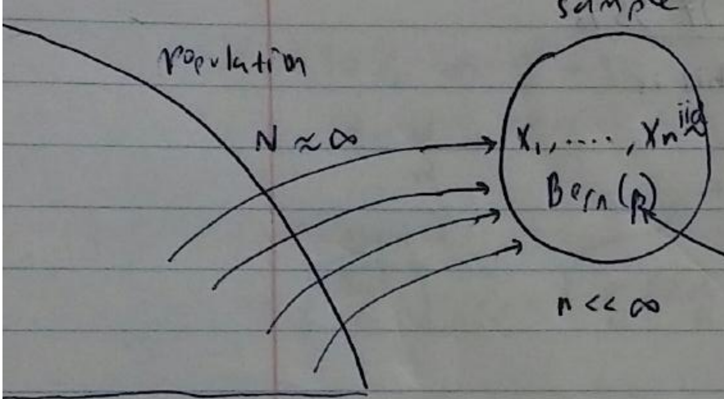
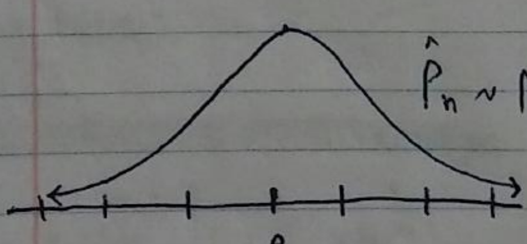
December 6, lecture 22

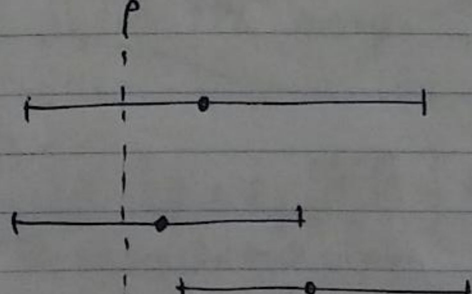


"Inverse Problem"
parameters unknown. Use sample to draw inference about parameters.

- <u>Statistical Inference</u>
  1. Point estimation: Best guess: $\hat{p}$ ← Not gonna do better than $\hat{p}$.
  2. Interval estimation:
     → <u>Confidence Interval</u> : $CI_{1-\alpha, p} := \left[\hat{p} \pm \frac{Z_\alpha}{2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$



$$\hat{P}_n \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$$



If have one experiment, all you see is this:

→ <u>Interpretation of CI (confidence intervals)</u>

① $\frac{\#\{p \in CI\}}{n} \longrightarrow 1-\alpha$ "cover"

② Before experiment,

$$P(p \in (I)) = 1-\alpha$$

③ If you believe in subjective probability (Beysian), then under prior information, you can say $P(p \in (I)) = 1-\alpha$.

$$P\left(p \in \left[\underbrace{\hat{p} \pm \frac{Z_\alpha}{2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{\text{constant}}\right]\right)$$

→ No help after done experiment b/c no r.v.

→ Before, does help because $\hat{p}$ replaced by r.v.

- If want greater coverage, what can I do?

$\alpha \downarrow \Rightarrow 1-\alpha \uparrow \Rightarrow \frac{Z_\alpha}{2} \uparrow$

goes from this...

...to this

catches more p-

However, if interval gets two big, usefulness goes down.

* How can I make interval smaller, but keep usefulness same?

Ex) 95%

95%

How do I do that?

⇓

[This will cost n going up, which means more trials, which means more work]

more useful because the interval expands less, only focusing on areas care about it.

- Statistical Inference continued...

③ Parameter Value Testing (Hypothesis testing)

* Gender ratio in human births. → $p := P(\text{Male})$

- His theory: $p \neq 0.5$ ie unequal gender ration. Crazy? → Yes
- Default/"null" hypothesis denoted: $H_0: p = 0.5$

  The crazy story is the alternate hypothesis
- H_a: H_0 is false; $p \neq 0.5$

♥ Assume H_0 is true

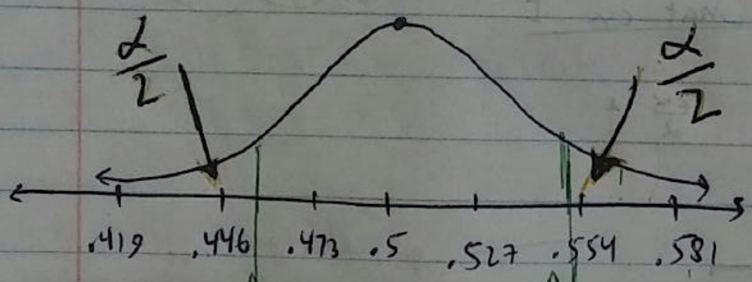
$$\hat{p} \sim N\left(\underset{.5}{p}, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$$

.5 .5

$n = 345$ would calculate to be $.0269$.

We take a sample of size $n = 345$.

So we have $\hat{p} \sim N(0.5, 0.0269^2)$



$\hat{p}_{SM}$ (too small)

$\hat{p}_{LA}$ (too large)

$\hat{p}$ — is this reasonable? Sure why not.

Rejected region H_0

$\hat{p}$ — Most going to come from ~~that~~ fat region.

Rejected region H_0 rejected

rejected

$\hat{p}$ — still reasonable – come from fattish part.

Not likely.

retainment region "H_0 retained"

$\hat{p}$ — way out. Very unlikely to be in that distribution.

Let $\alpha := P(\text{too rare})$.

$$1-\alpha := P(H_0 \text{ retained})$$

$$= P(\hat{p} \in [\hat{p}_{small}, \hat{p}_{large}])$$

$$= P(\hat{p} \in [p \pm \text{margin}])$$

$$= P\left(\hat{p} \in \left[p \pm \frac{z_\alpha}{2}\sqrt{\frac{p(1-p)}{n}}\right]\right)$$

$$\text{Retainment region} = \left[p \pm \frac{z_\alpha}{2} \sqrt{\frac{p(1-p)}{n}} \right]$$

$$\text{Rejection region} = \left[p \pm \frac{z_\alpha}{2} \sqrt{\frac{p(1-p)}{n}} \right]^C$$

We take a sample size of $n = 345$. Calculate $\hat{p}$.

(I) If $\hat{p} \in$ Retainment Region $\Rightarrow$ Retain H_0. We do not have sufficient evidence to reject null hypothesis.

(II) If $\hat{p} \in$ Rejection Region $\Rightarrow$ Reject H_0. Also accept H_a. We have sufficient evidence to reject the null hypothesis.

* Experiment

$n = 345$, $\alpha = 5\%$

$$\text{Retainment region} = \left[0.5 \pm 2 \underbrace{\sqrt{\frac{0.5(1-.05)}{345}}}_{.0269} \right] = [.446, .554]$$

If 169 babies were male $\Rightarrow$ $\hat{p} = \frac{169}{345} = .48 \in$ Retainment Region $\Rightarrow$

thus we do not have sufficent evidence to reject ~~the null hypothesis.~~ human gender ratio equality.

- 3 scenarios
  - ① Flip a coin 100 times. You want to know if coin is fair. $p = P(H)$
    - * Scenario I: Get 51 heads. Fair? $\Rightarrow$ Yes
    - * Scenario II: Get 98 heads. Fair? $\Rightarrow$ No because too far away from what you'd expect.
    - * Scenario III: 61 heads. Fair? $\Rightarrow$

H_0: $p = 0.5$

H_a: $p \neq 0.5$

$\alpha := 5\%$

$$\text{Retainment range} := \left[p \pm \frac{z_\alpha}{2} \sqrt{\frac{p(1-p)}{n}} \right] = \left[0.5 \pm 2 \sqrt{\frac{0.5(1-.5)}{100}} \right] = [0.4, 0.6]$$

cheat sheet

$\boxed{\begin{array}{l} \alpha = 5\% = 2 \\ \alpha = 1\% = 2.84 \end{array}}$

Under scenario III: $\hat{p} = \frac{61}{100} = .61$. Is it in retainment region? $\Rightarrow$ No, not in retainment region. Thus, reject H_0 $\Rightarrow$ coin is not fair.

- Mars, a candy company, says that the proportion of blue M&Ms is 20%. You think otherwise. (typical hypothesis test)

Let $p := P(\text{Blue})$

$H_0: p = 0.2$
$H_a: p \neq 0.2$ } "two-sided test"

$n = 615$ M&Ms

$\alpha = 1\%$ ← want to be sure Mars company is wrong

Retainment region $:= \left[p \pm \frac{z\alpha}{2} \sqrt{\frac{p(1-p)}{n}} \right]$

$= \left[0.2 \pm 2.84 \underbrace{\sqrt{\frac{(0.2)(1-0.2)}{615}}}_{.0458} \right] = [.1542, .2458]$

* 158 blue m&m's

$\hat{p} = \frac{158}{615} = .2569 \Rightarrow$ $\hat{p}$ is not inside the region. Not in the retainment region.
Reject H_0. The probability of blue M&Ms is not 20%.

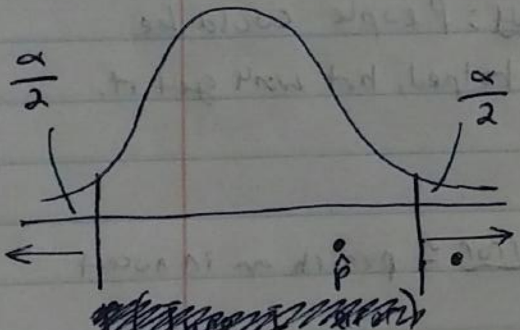
- $\hat{p}$ is drawn from $\hat{p} \overset{\text{given}}{\mid} H_0 \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$

→ Is little $\hat{p}$ a random distribution? : Yes

* Mistakes could have Made

Decision

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



* There are two different mistakes you can make. *

* $P(\text{Type I error}) = \alpha$

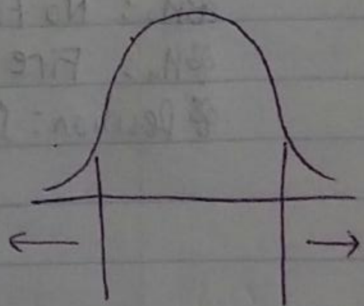
→ Type I error : reject even if true.

α is your choice. If you are skeptical person, will make α really small.

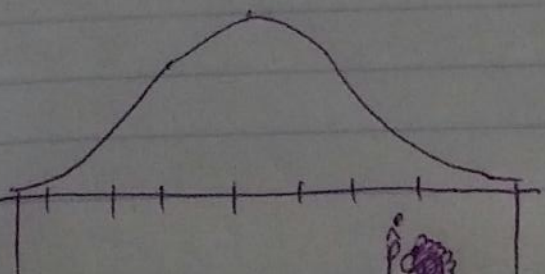
* $P(\text{Type II error})$ = beyond the scope of this class.
* $P(\text{Reject } H_0 \mid H_0 \text{ false})$ = POWER.

- $\alpha \uparrow \Rightarrow P(\text{Type I error}) \uparrow \Rightarrow P(\text{Type II error}) \downarrow$

going to make more rejections, more likely to think ~~is wrong~~ Hypothesis is wrong.



- $\alpha \downarrow \Rightarrow P(\text{Type I error}) \downarrow \Rightarrow P(\text{Type II error}) \uparrow$



- 3 scenarios
  - Clinical trial
    - H_0 : drug doesn't work
    - H_a : drug does work
    - Decision release drug to market

    * Type I error: release a drug that does not work
      Cost : possible death

    * Type II error : Not releasing a drug that works.
      Cost : People could be helped, but won't get it.

  - Court Case
    - H_0 : Not guilty
    - H_a : Guilty
    - Decision : punish or not

    * Type I error : punish an innocent person

    * Type II error : let a guilty person go free.

  - Fire Alarm
    - H_0 : No fire
    - H_a : Fire
    - Decision : Set off alarm

    * Type I error : False alarm
      cost : waste time
    * Type II error : There's a fire, but no alarm.
      cost : someone dying
      ↳ α should be high.