Agenda

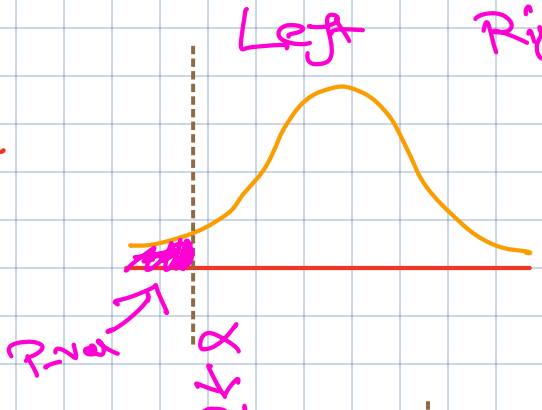
* Recap : Hypothesis testing Framework
* Z- test proportions
    * One Sample
    * Two Sample
* T- test using 1 Sample
* T- test using 2 Samples

Steps involved in Hypothesis Testing

1. Setup Null and Alternate Hypothesis
2. Choose Distribution
   Test Statistic and α (Significance level)
3. Select Left / Right / Two-Tailed
4. Compute P- Value
5. Compare P-Value to α:
   if P-val $< \alpha$; Reject Null Hypo
   else : Fail to Reject Null Hypothesis

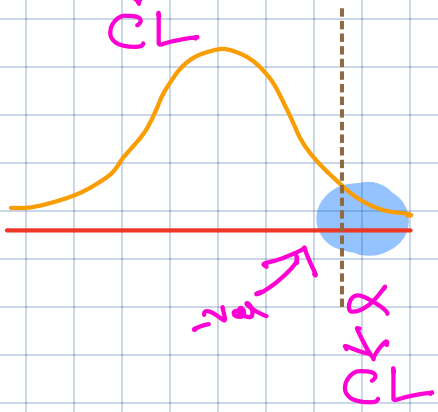
① $H_0 : \mu \geq \text{Value}$

$H_a : \mu < \text{Value}$



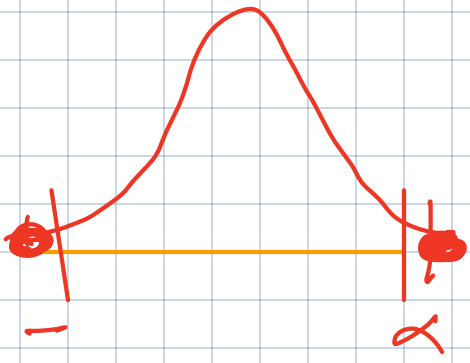
② $H_0 : \mu \leq \text{Value}$

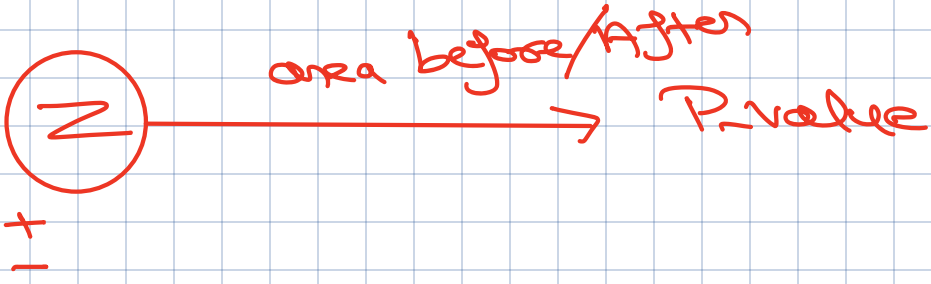
$H_a : \mu > \text{Value}$



③ $H_0 : \mu = \text{Value}$

$H_a : \mu \neq \text{Value}$





Z- proportion Test

Let's say you are a Data Scientist working for a Web Application

- Let's say that the website wants to add a new feature to make "more" customers buy their product, and increase their Proportion of Sales
- **Proportion of Sales = No.of customers buying the product / No of customers visiting the web page**

> What do you think can be the impact of adding the new feature?

P_1 → before

P_2 → after

Case 1 : No impact on Sales

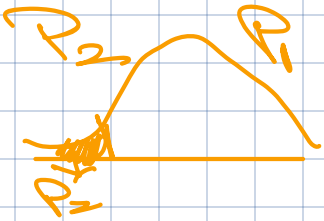
Case 2 : Some impact $H_a : P_1 \neq P_2$

Sales increase

$H_a : P_1 < P_2$

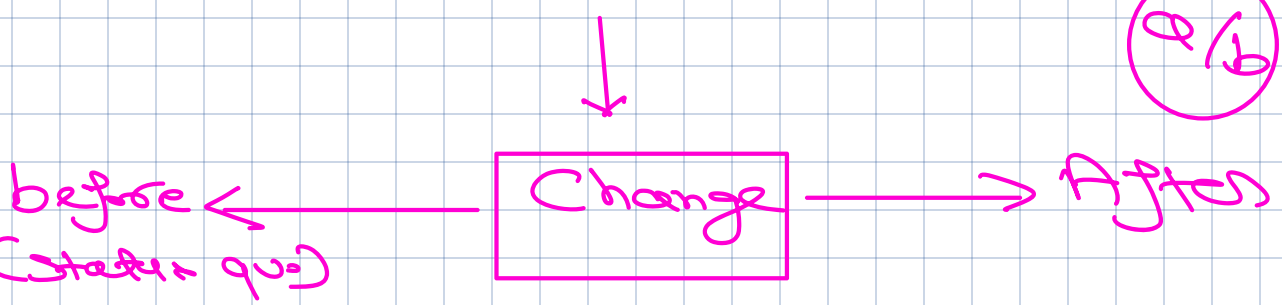
Sales decrease

$H_a : P_1 > P_2$



Test of proportions

Q/b



Types of proportions Test :

- One Sample Z-proportion
- Two Sample Z-proportion

One Sample Z-proportion

Imagine you are a product manager in a company, and you want to determine the satisfaction rate of customers with a new product. [5-7 mins]

- A proportion is a way to express a part of a whole. It's often used to measure the percentage of a specific outcome within a larger population.
- In our case, it's the **proportion of satisfied customers.**

Conditions for One Sample proportions Test:

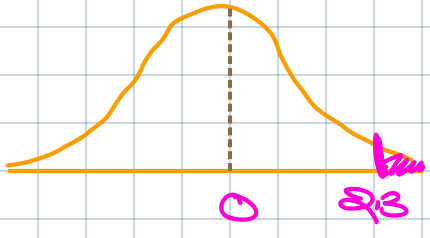
① The sample size should be sufficiently large (≥ 30)

② The data must follow Normal Dist.

Test Statistic for One Sample Z-proportion

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

$Z \sim N(0,1)$



Where $\hat{P}$: Prop of Satisfied Customers (Observed)

P : Specified prop under Null Hypothesis

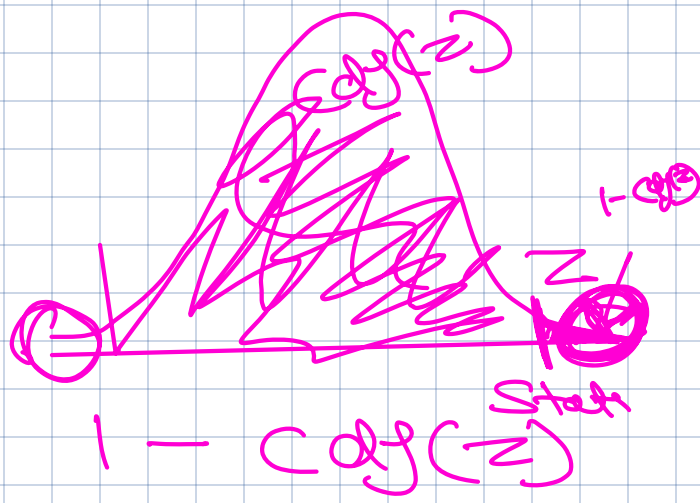
n : Sample Size

Null Hypothesis :

Alternate Hypothesis :

H_0 :)

H_a :(



Questions

- You are a product manager for a company that has recently launched a new product.
- Customer satisfaction is a critical metric, and you want to determine if the proportion of satisfied customers with the new product meets your target satisfaction level of 70%.
- You collected a random sample of 150 customer reviews, and 115 of them expressed satisfaction with the product.

Step 1 : H_0 and H_a

H_0 : $P = 70\%$

H_a : $P \neq 70\%$

Step 2 : Distribution and Significance Level

$Z \Rightarrow N(0,1)$ $\alpha = 5\%$

Step 3 : Tailed Test ?

Two tailed Test

Step 4: Test-Statistic and p-value

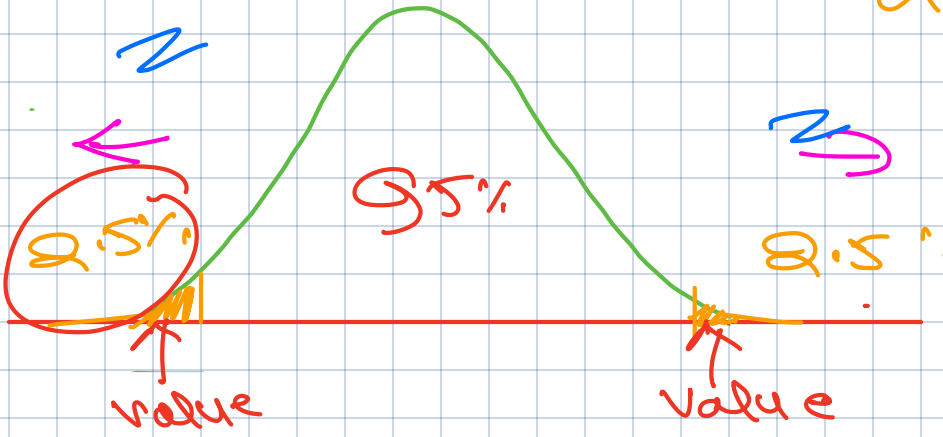
$$Z = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

$\xrightarrow{\text{P-value}}$ 0.03 $\times 2$

Step 5: Compare P-value with α

$H_a: \hat{P} \neq P$ $\alpha \Rightarrow 5\%$

$\alpha \Rightarrow 5\%$



For two tailed Tests:

or

$P < \alpha/2$

$Q = 3 \times 2 = 6$

$b = 5$

$a = 3$

$b = 5/2$

$2 \times P < \alpha$

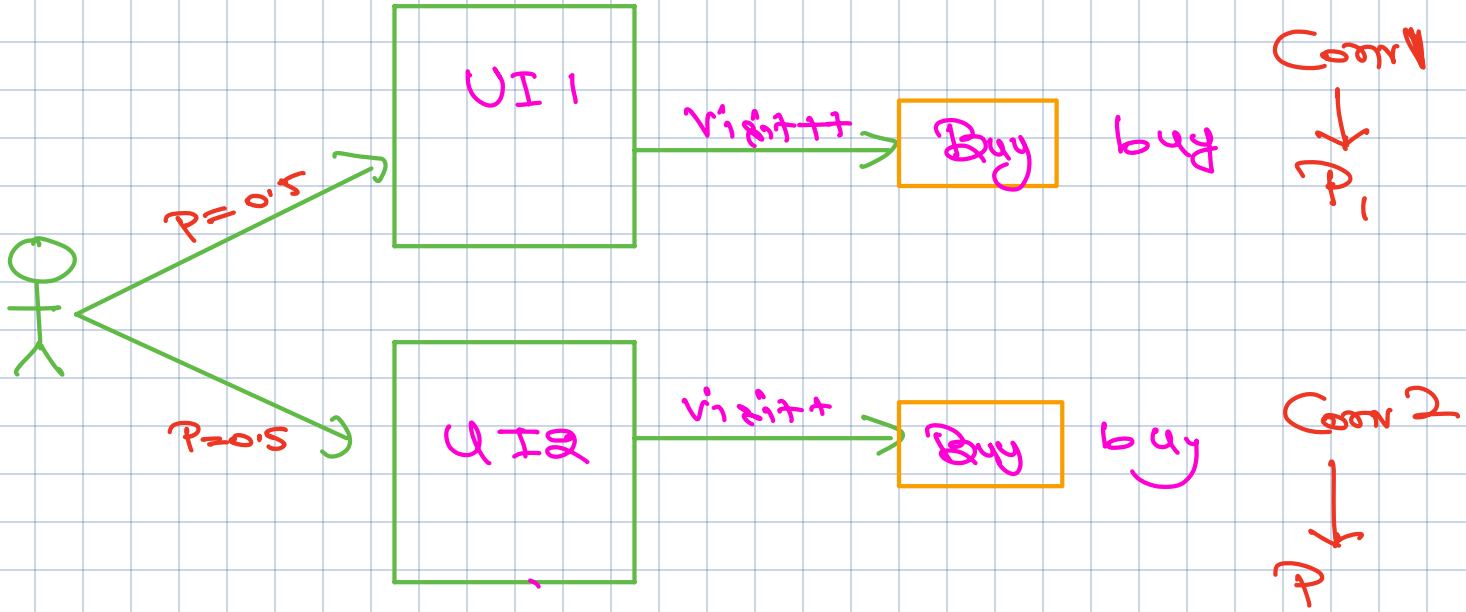
A fast-food restaurant claims that 80% of their customers prefer their new burger over the old one.
In a random sample of 100 customers, 85 said they preferred the new burger.
What is the null and alternative hypothesis?

Newly Collected Data → 85/100

$H_0 : P = 0.80$

$H_a : P \neq 0.80$

Two Sample Z-proportion



buy / # visit

$H_0 : \quad p_1 = p_2$

$H_a : \quad p_1 \neq p_2$

Two Sample Z proportion two tailed Hypothesis Test

Test Statistic for Two Sample Z-proportion

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\sim N(0,1)$

$\hat{p}_1 = \frac{x_1}{n_1}$

$\hat{p}_2 = \frac{x_2}{n_2}$

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$n_1 \rightarrow$ # in site 1

$n_2 \rightarrow$ # in site 2

Questions

You are the manager of an e-commerce website, and you have recently implemented a new web page in hopes of increasing sales.

To evaluate the effectiveness of the new page, you collected data on the conversion rates for both the old and new web pages.

The conversion rate is defined as the proportion of visitors who make a purchase.

- For the old web page (Web Page A), you had **1000** visitors, resulting in **50** conversions.
- For the new web page (Web Page B), you had **500** visitors, resulting in **30** conversions.

Now, you want to determine if there is a statistically significant difference in the conversion rates between the old and new web pages.

T-test

Let's say you are a Research Scientist working on a new cognitive enhancement pill

- The goal is to develop a pill that can significantly improve IQ scores in individuals.
- The researchers believe that the new pill will lead to a significant increase in average IQ scores for the population.

* Given Data is collected

```
# The pill is given to a few people and their IQ is tested with following results:
iq_scores = [110, 105, 98, 102, 99, 104, 115, 95]
```

Sample

Let's test Hypothesis that new pill actually increases IQ

H_0 : IQ ≤ 100

H_a : IQ > 100

≤ 100

Questions can we apply Z-test Here?

↓ Normal Dist

① Sample-size < 30

② Std-Error → $\frac{\sigma_{pop}}{\sqrt{n}} \rightarrow \frac{\sigma_{sam}}{\sqrt{n}}$ &

t - test → t - distribution

$$t\text{-score} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \left(\text{Z-score} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

s : Sample SD
n : Sample size

* Types of T-test
  * 1 Sample T-test
  * 2 Sample Test
    - Independent
    - Dependent

One-Sample T-test

Two-Sample T-test (Independent Samples)

Suppose we have IQ data samples across 2 schools, and we want to compare and see which school's students have better IQ

* Null and Alternate Hypothesis

H_0 :

H_a :

① H_0 and H_a

trick:

H_0 : $=$ → two H_a $\neq$

, $\leq$ → right H_a $>$

, $\geq$ → Left H_a $<$

→ α or $\alpha/2$

Calculate Test - statistics

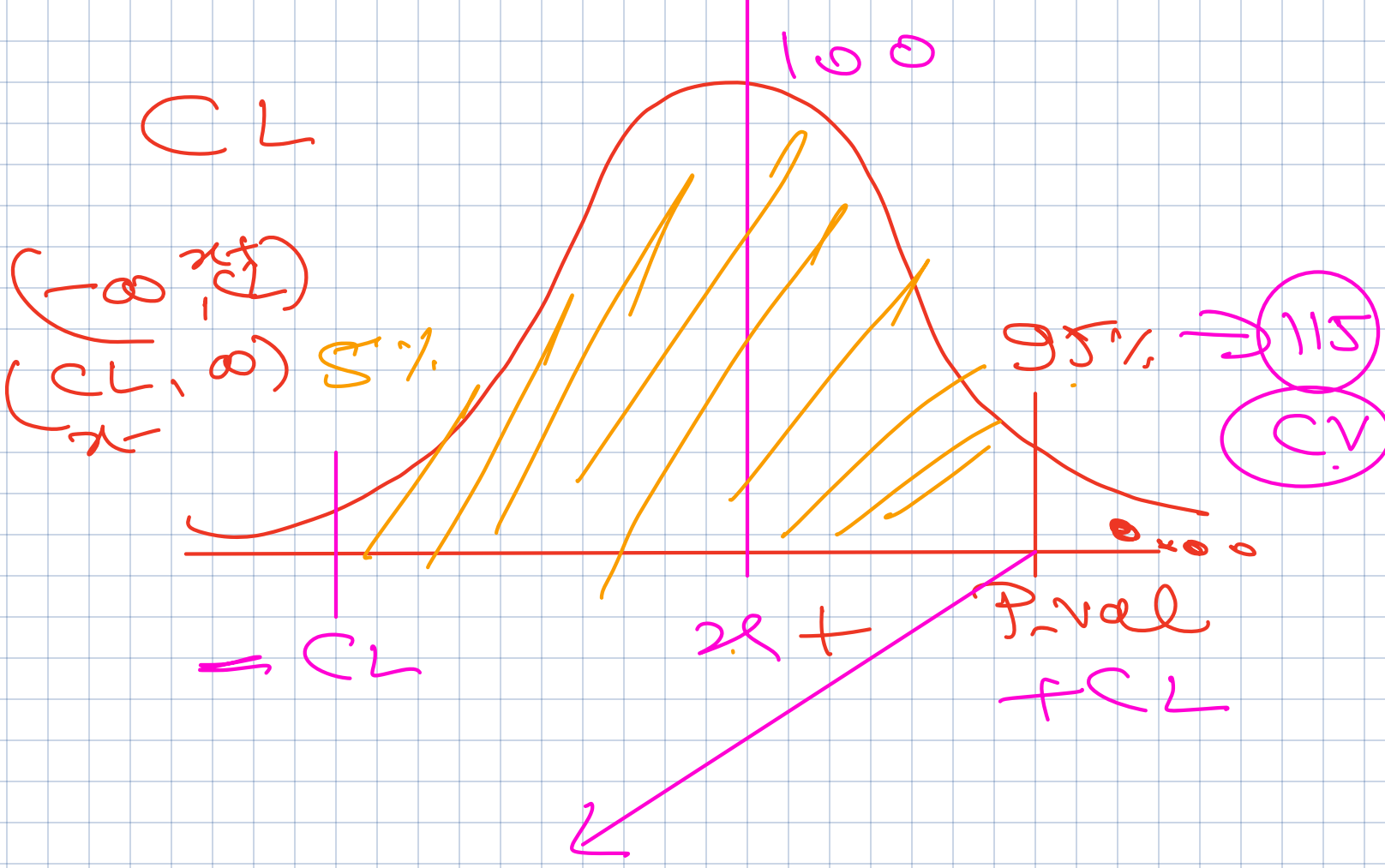
- Z-score [- - - - - -] $n \geq 30$
- Prop-2 → Success / trial
- t-score [- - - -] $n < 30$

Z → P-value → Compare α

P-value $\leq \alpha$

↓

reject H_0



Exact value 구할 때는
Z-score

95% ⇒ 115 IQ
≥ 118 이상

IQ
118 ⟶ Z ⟶ P-value
0.05

X

C2 Value

Value associated with

C2