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**

Conversion Rate

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

Cases

Some % of population / Total population

① No impact on ⟶

② Some - Impact ⇒ Two Tailed Side
- → Conversion Rate Increased
- → Conversion Rate Decreased

→ 1 tailed

e Test of proportions'

e A proportion is a way to express a part of a whole Unit.'

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 Sample Size should be large. Typically >30

② Data should ~ follow Normal Dist

Test Statistic for One Sample Z-proportion

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

$Z \sim N(0,1)$

$\hat{p}$: Observed (Sample) Value of Prop

p : Specified Proportion Under H_0 (Hypothized/population/Target proportion)

N : Sample-Size

Null Hypothesis : Target level of proportions

Alternate Hypothesis : S proportions ≠ Target level

H_0) $\hat{p} = p$, $\hat{p} \geq p$, $\hat{p} \leq p$

H_a) $\hat{p} \neq p$, $\hat{p} < p$, $\hat{p} > p$

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%. ⇒ p
- You collected a random sample of 150 customer reviews, and 115 of them expressed satisfaction with the product. ⇒ 115/150 Sample Prop

Step 1 : H_0 and H_a

H_0 : Satisfied Customer proportion = Target level → $p = 0.7$

H_a : $p \neq 0.7$

Step 2 : Distribution and Significance Level

(Normal Dist) (0.05)

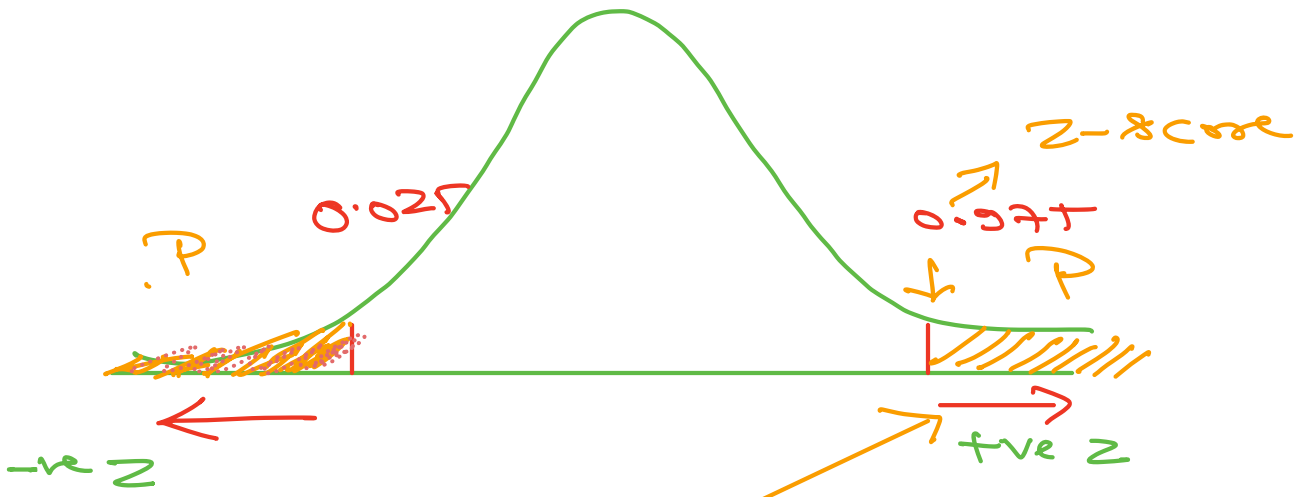
Step 3 : Tailed Test ?

Two Tailed Test

Step 4: Test-Statistic

$$Z \Rightarrow \frac{(\hat{p} - p)}{\sqrt{\frac{p \times (1-p)}{n}}}$$

Step 5: Compare P-value with α

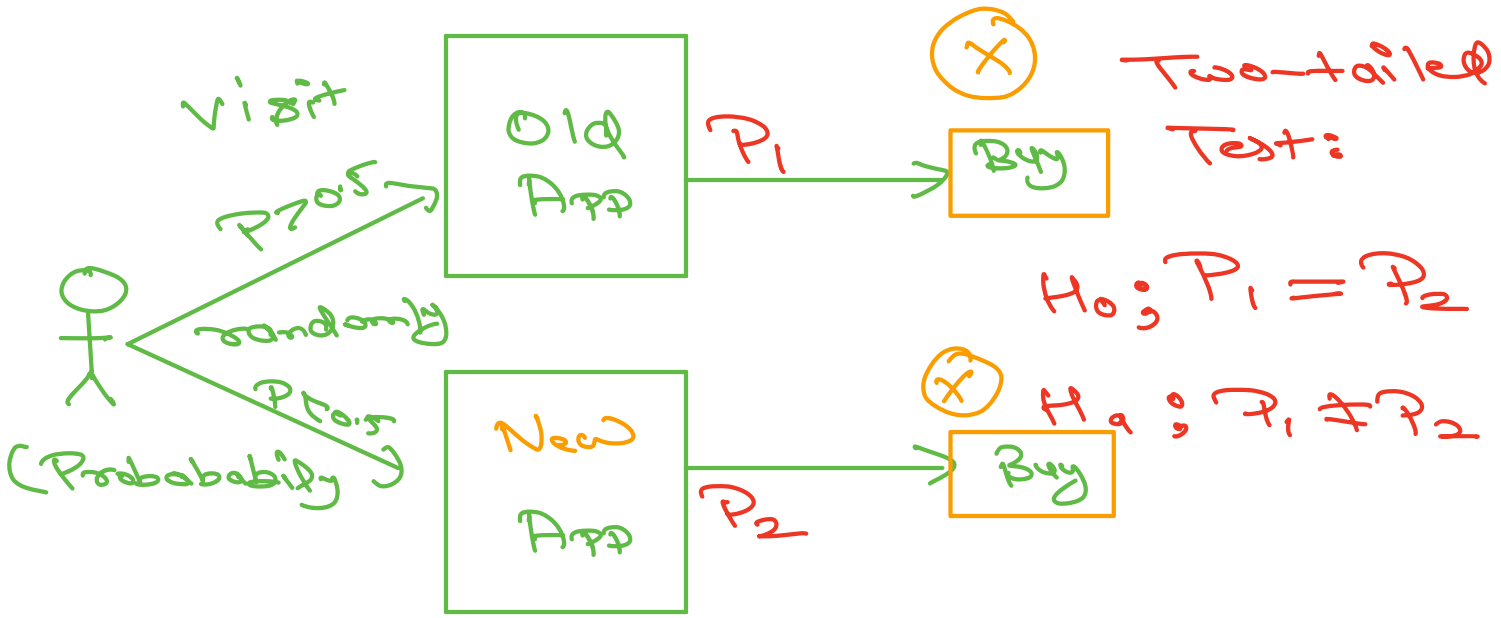


abs(Z)

$\alpha/2$

2x(1- cdf(abs(Z)))

Two Sample Z-proportion



Conditions for Two Sample proportions Test:

1. Sample Size should be large. Typically > 30 (Both Samples)
2. Data should be follow Normal Dist

Null Hypothesis : $\hat{P}_1 = \hat{P}_2$

Alternate Hypothesis : $\hat{P} \neq \hat{P}_2$

Test Statistic for Two Sample Z-proportion

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

$\sim N(0,1)$

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

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

n_1 : Sample Size 1

n_2 : Sample Size 2

$\hat{p} : \frac{(x_1 + x_2)}{n_1 + n_2}$ (Combine proportion)

$\hat{p}_1, \hat{p}_2, \hat{p} \rightarrow [0,1]$

Questions

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

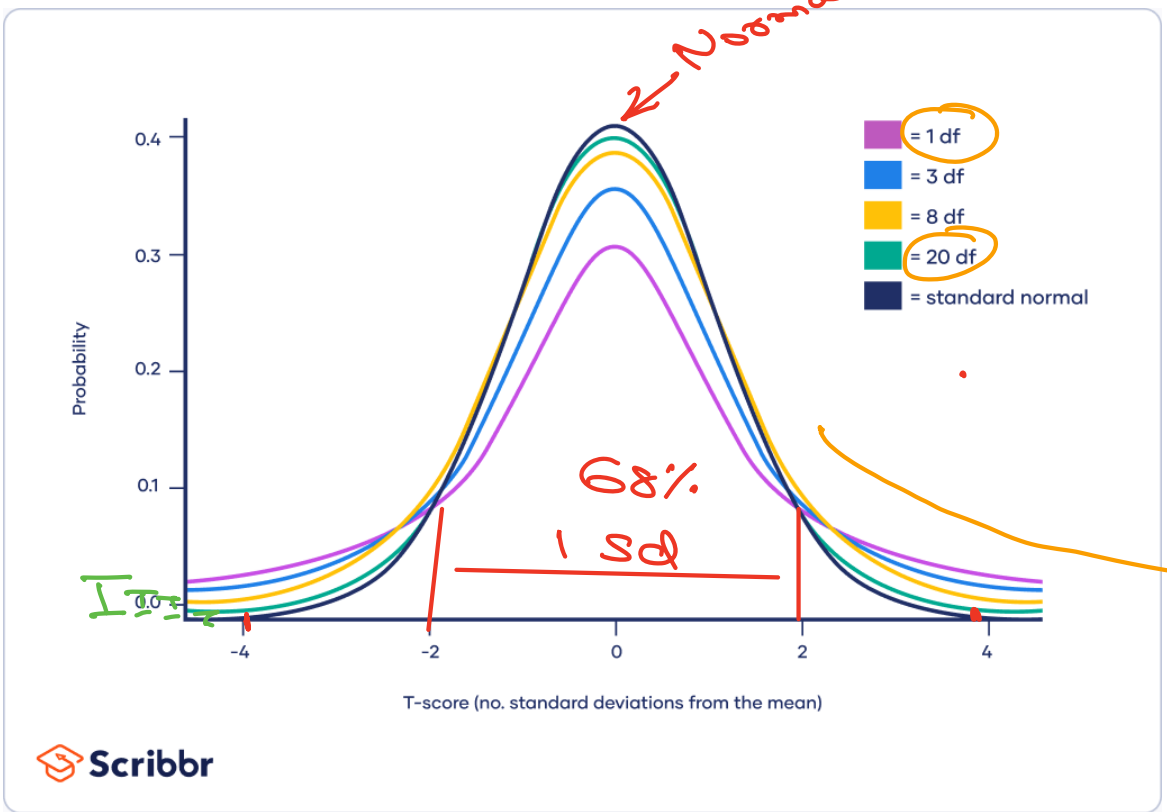
To evaluate the effectiveness of the new page, you collected data on the conversion rates for both the old and new web pages.

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.

Normal Distribution



T - Distribution with an addition parameter DF (degree of Freedom)

DF controls how Fat the tail would be

DF ≥ 30

T-Distibution is Normal Distribution

Research Scientist working on a medicine to Enhance IQ-score

iq-score = [110, 105, 98, 102, 99, 104, 115, 95]

Population-IQ = 100

$H_0 = 100$ $\quad$ $H_0 = 100$

$H_a \neq 100$ $\quad$ $H_a > 100$

$\mu = 100$

$\bar{x} = 103.5$

Z-score $\Rightarrow$ $\dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ $\sigma \rightarrow$ Pop-SD $\rightarrow$ Sample-SD

SE $\Rightarrow$ $\dfrac{\sigma}{\sqrt{n}} \rightarrow$ SD of population

Population SD σ

Sample SD $\Rightarrow$ $\sqrt{\dfrac{\sum (\bar{x} - x_i)^2}{n-1}}$ $\leftarrow$ Bessel's Correction

Z-test

↓

Pop-sd

Sample-SD with Bessel Correction

↓

T-test

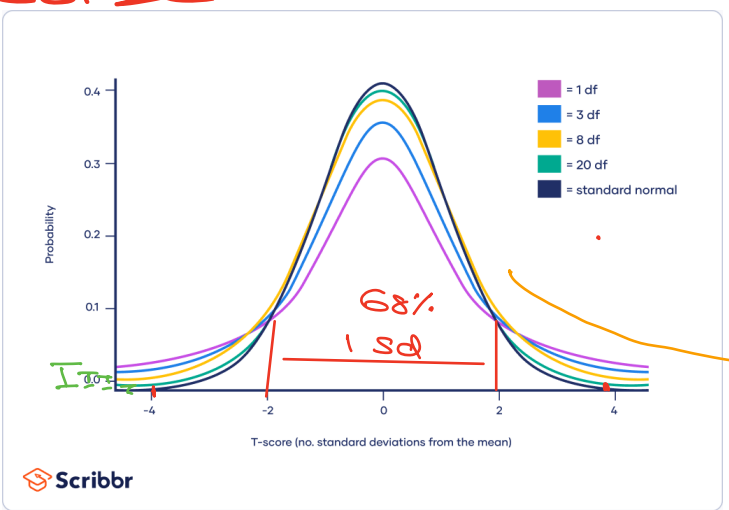
DF = (n-1)

iq-score = [110, 105, 98, 102, 99, 104, 115, 95]

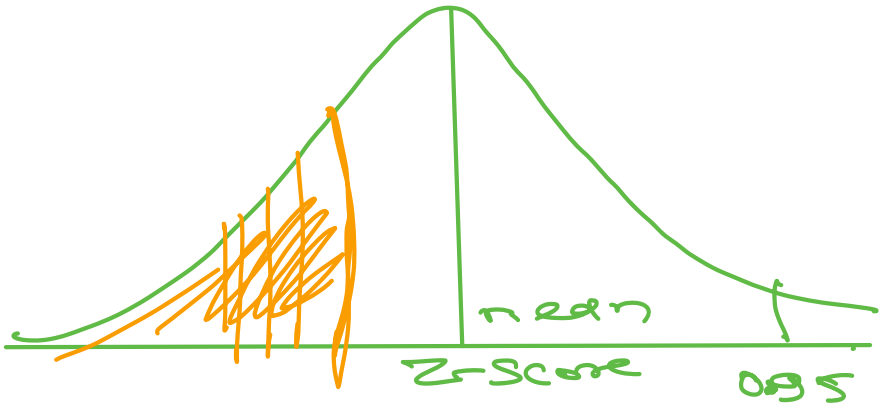
DF = 8-1 = 7

T-test

- Pop-sd is not known
- Sampksize < 30



T-dist (N)
≥ 30
≈ Normal Dist



Critical Z-score

↓

Number → Critical Level

cdf (b-a)