HYPOTHESIS TESTING - 3

T-Test

Hypothesis Testing Framework

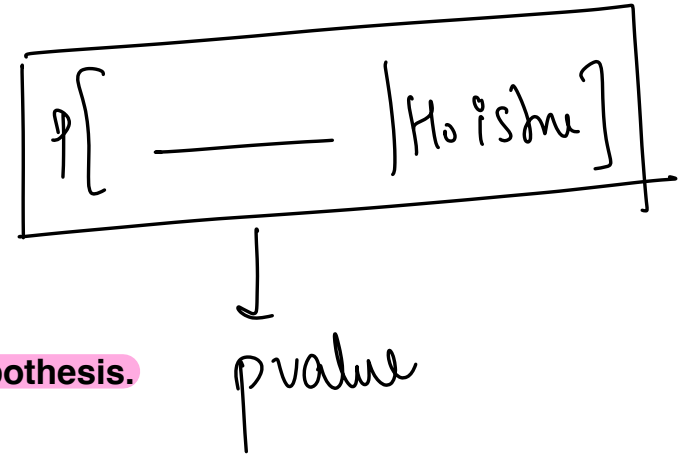
1) Setup the Null and Alternate Hypothesis

2) Choose the right test statistic distribution

3) Left tailed vs Right tailed vs Two-Tailed

4) Compute P-value

5) If P- value is less than alpha , then reject the null hypothesis.



pvalue $< \alpha$ Reject H_0

pvalue $> \alpha$ Fail to Reject H_0

A french cake shop claims that the average number of pastries they can produce in a day exceeds 500. The average number of pastries produced per day over a 70 day period was found to be 530. Assume that the population standard deviation for the pastries produced per day is 125. Test the claim using a z-test with the critical z-value = 1.64 at the alpha (significance level) = 0.05, and state your interpretation.

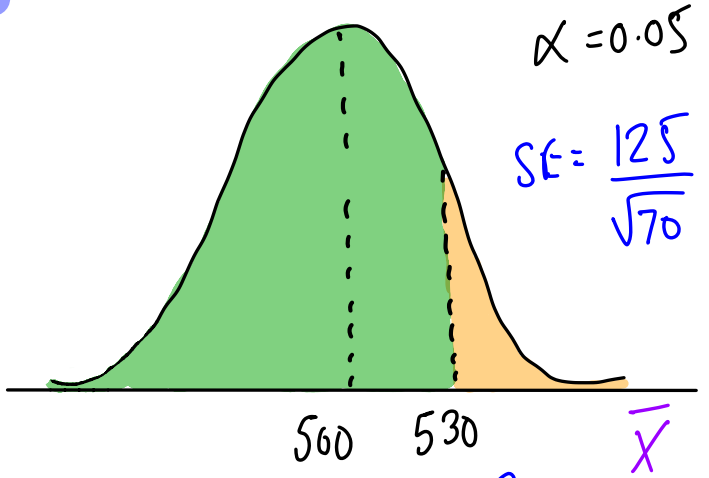
$H_0: \mu = 500$

$H_a: \mu > 500$ Right tailed test

Test statistic: Sample mean of 70 days

Distribution: Gaussian

Observed value: 530



n

$p\text{value} = P[T \geq 530 \mid H_0 \text{ is true}]$

$Z = \frac{530 - 500}{125/\sqrt{70}}$

T Test Student's t distribution

Improve IQ with a pill.

population mean^{IQ} = 100

Try on a few people

110, 95, 98, 115, 112, 108 .... v/s 100

H_0: $\mu = 100$ (pill had no effect)

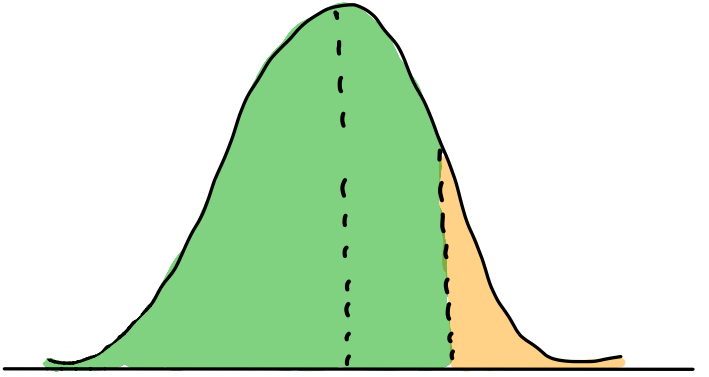
H_a: $\mu > 100$ (pill had an effect)

Right tailed test.

Test Statistics → Sample mean of 8 people → 103.5

$$Z\ stat = \frac{103.5 - 100}{\sigma/\sqrt{8}}$$

σ → Missing



$$T\ stat = \frac{103.5 - 100}{S/\sqrt{8}}$$

S → Sample standard deviation

"This is not SE"

Degree of freedom → [Chi square test]

↳ $n-1$

1. Test of Independance
2. Goodness of fit

1 2 3 1.5 2.5 3.5 4 5 6 7

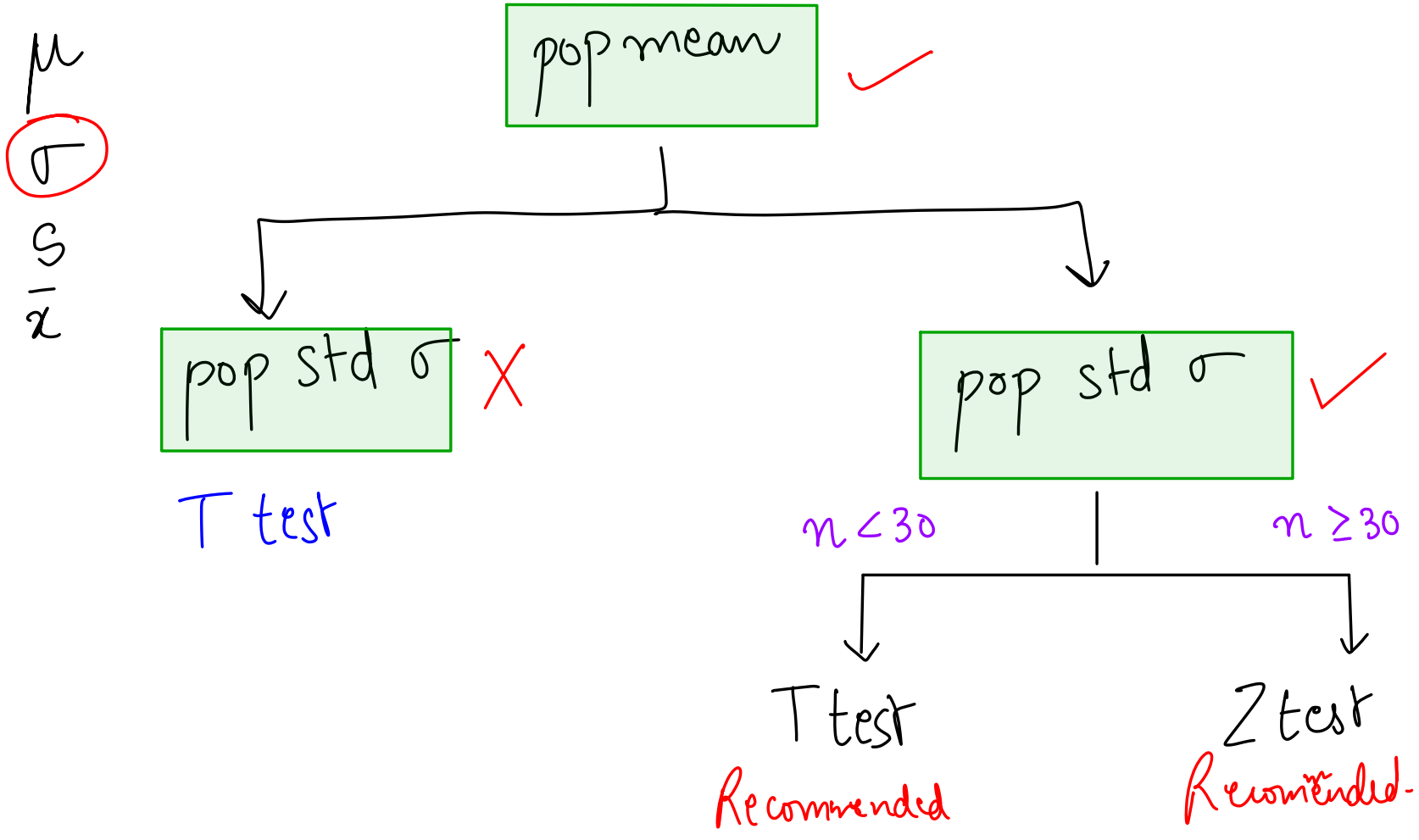
* Amount of Rawmatrials
* Price of Rawmaterials

1 2 3 ☐ = 4

4

= 4 × 4

$x_1 + x_2 + x_3 + x_4 = 16$


μ
σ
S
x̄
pop mean
pop std σ
X
T test
pop std σ
n < 30
n ≥ 30
T test
Recommended
Z test
Recommended

T test

1. 1 Sample μ v/s pop. μ — t test_1samp

Independant

2. 1st Sample μ_1 v/s 2nd Sample μ_2 — t test_ind

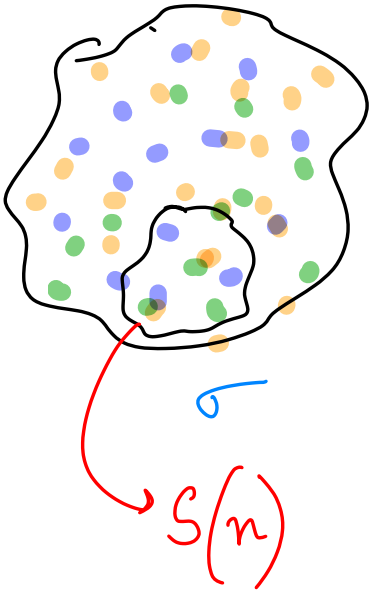
Relative

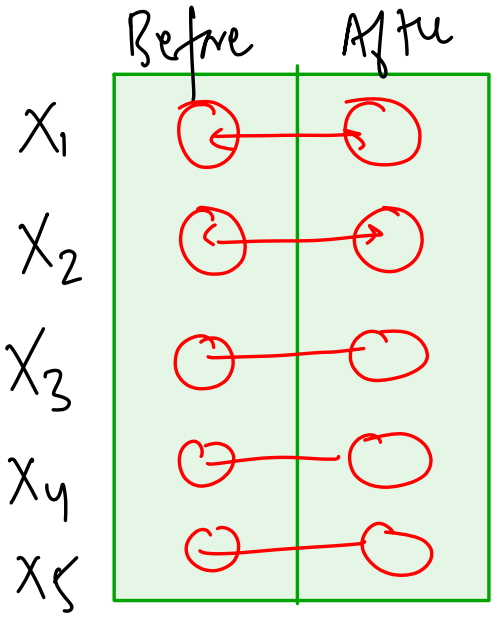
3. 1st Sample μ_1 (before) v/s 2nd Sample μ_2 (After) — t test_rel

$$x_1, x_2, x_3, x_4 \; - \; - \; - \; x_n$$

$$\text{Var}(\text{Sample}) = \frac{\sum (x - \bar{x})^2}{n-1}$$

$$\text{Std}(\text{Sample}) = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$





10 kgs 3 month

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

① Shapiro Wilkin

②

③ 10000 → 100 → $\bar{x}$

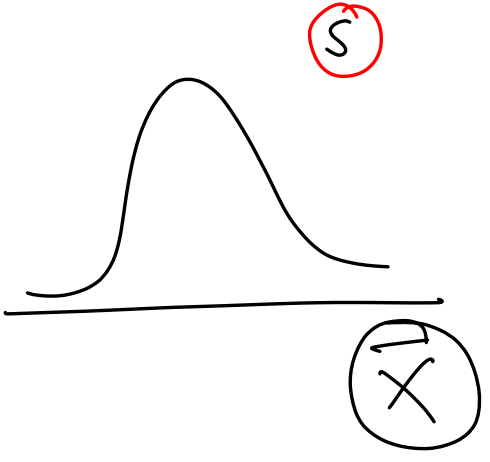
$n = 1000$ $k = 200$

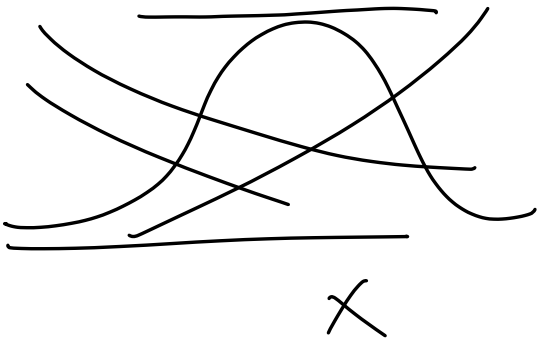
2022 M1 1000 → 200

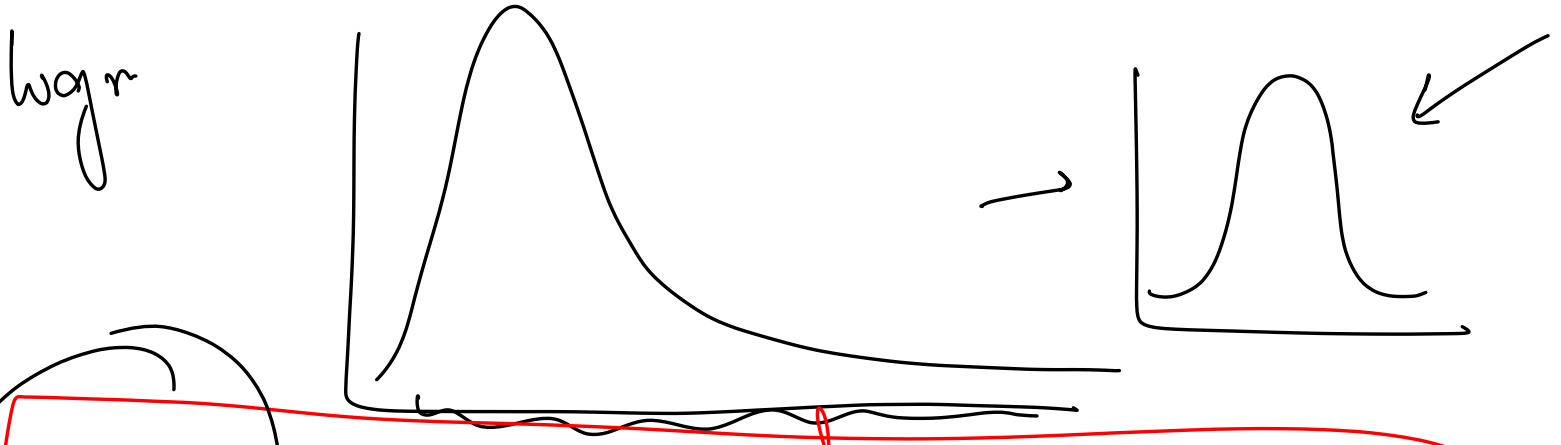
2023 M2 1200 → 220

$n = 1200$ → $k = 220$

P = → Not Normal







product Sector Industries & problem statement

ecommerce

finance

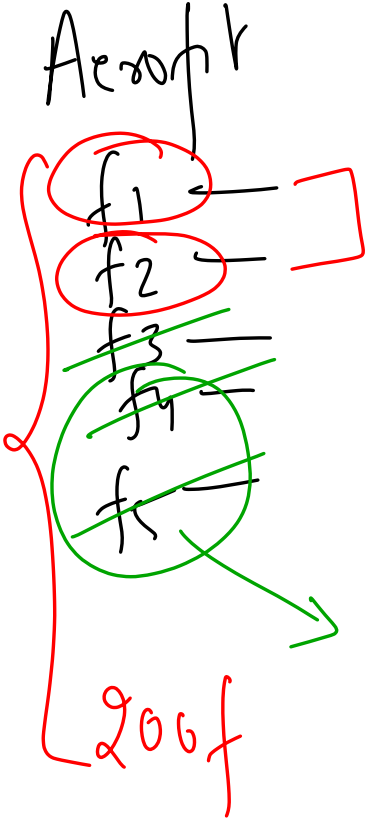
(1) Z test Sample < 5

t test

1-2σ → t test.

>2σ → t test

(2) Z test t test



Numerical Categorical

V/S

Categ

feature engineering

feature Importance

↓

feature Selection