Statistical Tests

$H_0 \longrightarrow \mu = 20\,gm$

$H_1 \longrightarrow \mu \neq 20\,gm$ —— Two tailed tests

$\bar{x} = 18.8\,gm$

$n = 40$

$SE = 0.8\,gm$

$\alpha = 0.05$

Z - test

$$\boxed{z = \frac{\bar{x} - \mu}{SE}}$$

$$z = \frac{18.8 - 20}{0.8}$$

$$z = -1.5$$

CI $\longrightarrow 1 - \alpha$

$= 1 - 0.05$

$= 0.95$

Critical value = 1.96

p-value = 0.1336

$\boxed{\alpha < \text{p-value}}$ — probability (H_0 true)

Conclude $\longrightarrow$ fail to Reject Ho

OR

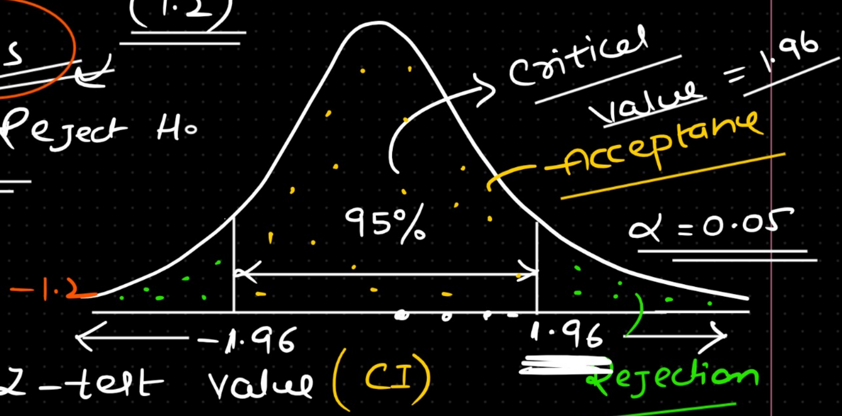
critical value > z-test value

(1.96) (1.2)

abs

* fail to Reject Ho

-1.96



critical value < z-test value (CI)

* Reject Ho

t-test ($n < 30$)

SE = known

↳ z-test

$|$z-test value$|$ < critical value

↳ Acceptance

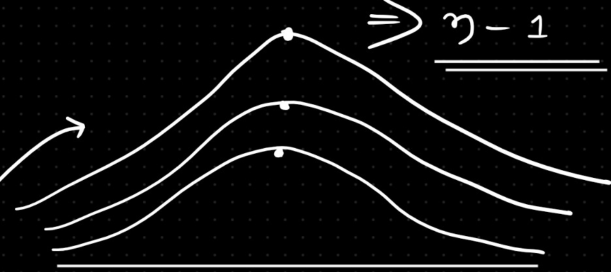
↳ fail to Reject

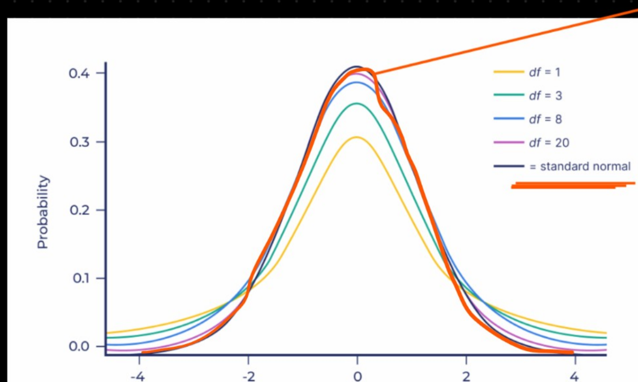
$\sigma = ?$

- Yes → $n > 30$
  - Yes → z-test → $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$
  - No → t-test
- No → t-test → $\boxed{t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}}$

$s \longrightarrow$ sample std deviation

Degree of freedom $\Rightarrow n - 1$

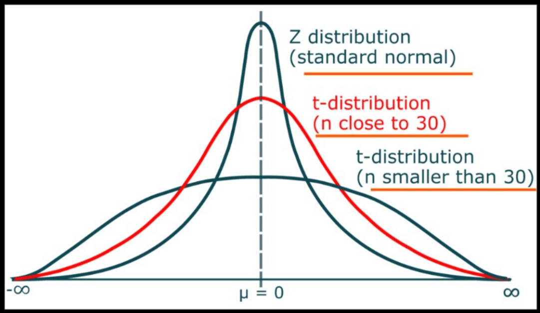




→ z-test

↳ inference

$t \longrightarrow$



A company claims that the average daily sales are \$200,000. A sample of 12 days is taken, and the average daily sales observed is \$210,000 with a standard deviation of \$15,000. Test the company's claim at a 5% significance level.

$\alpha = 0.05$

$\mu = \$2{,}00{,}000$

$\bar{x} = \$2{,}10{,}000$

$n = 12$

$s = \$15{,}000$

$H_0 \longrightarrow \mu = \$2{,}00{,}000$

$H_1 \rightarrow \mu \neq \$2{,}00{,}000$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{2{,}10{,}000 - 2{,}00{,}000}{15{,}000/\sqrt{12}}$$

$\Rightarrow 2.31$

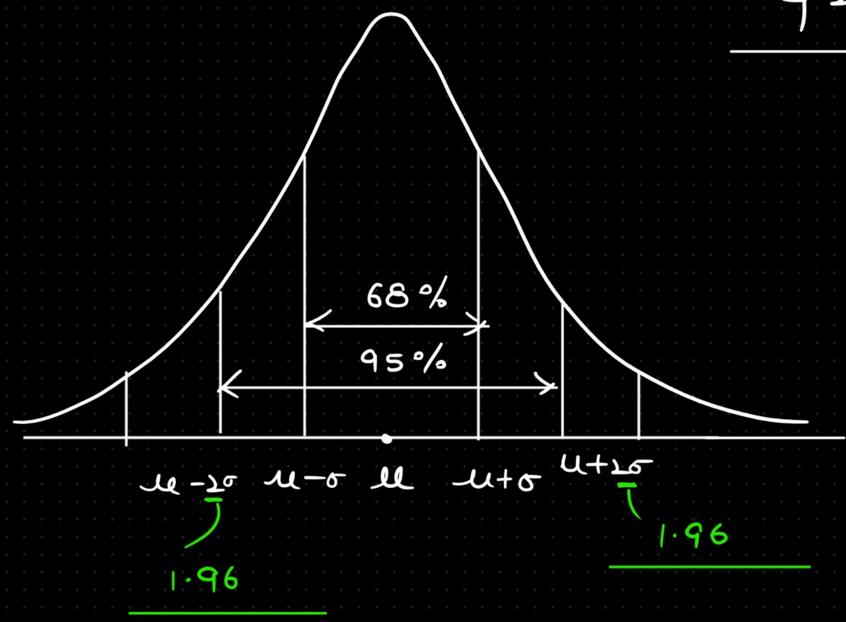
$Df = n - 1 = 11$

p-value = 0.0413

p-value < α $\longrightarrow$ Reject H_0

Average daily sales are significantly different from

$ 2,00,000



ANOVA → Analysis of Variance

Teaching style → ① ; Gender → ② → (M, F)

(A) Physical sessions →	70	68	75	72	74	$\mu_A = 71.8$
(B) Virtual sessions →	65	66	70	60	68	$\mu_B = 65.8$
(C) Hybrid sessions →	80	78	85	82	84	$\mu_C = 81.8$

$H_0 \longrightarrow \mu_A = \mu_B = \mu_C$

$H_1 \longrightarrow$ Atleast one group's test score is different

ANOVA

- The samples are independent
- The data within each group is roughly normally distributed

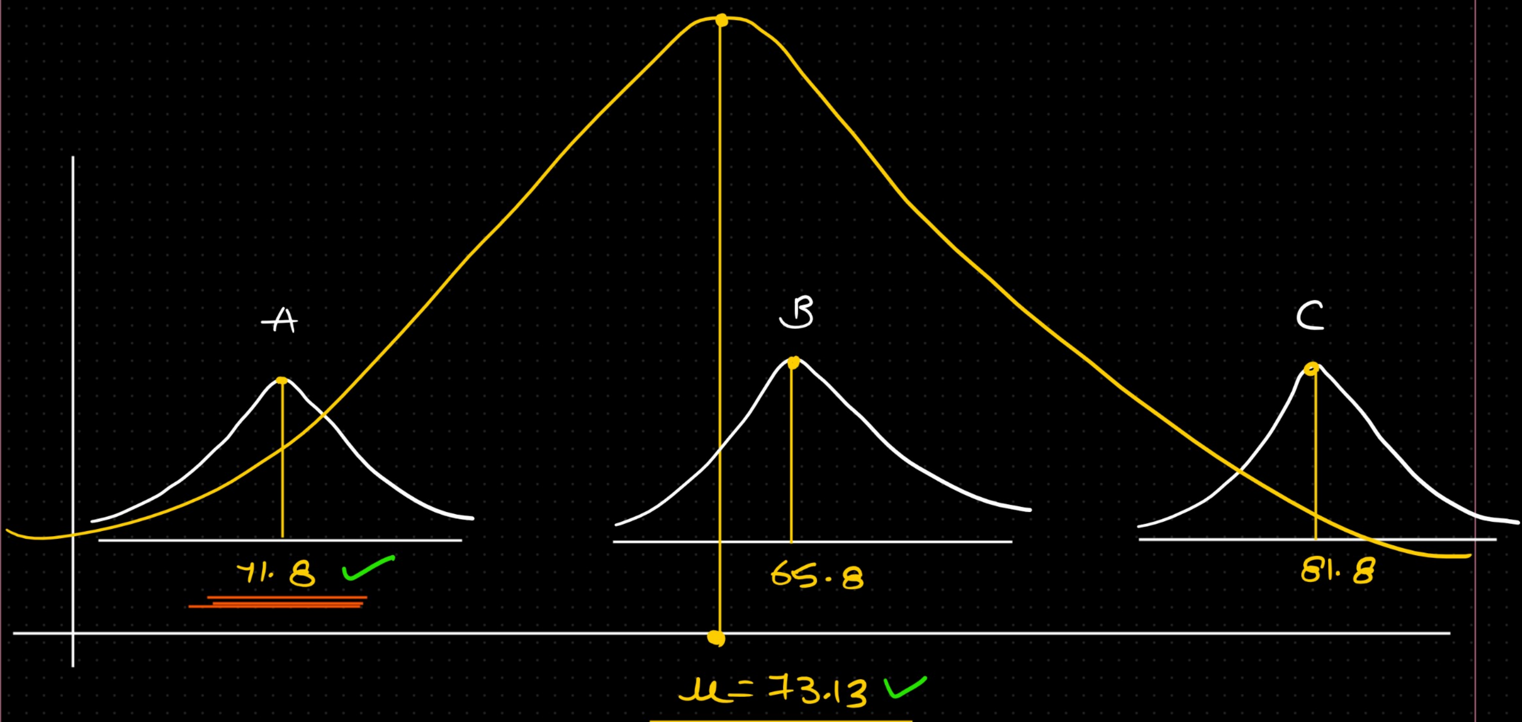
↳ Variance across the group are almost similar

$$\underline{\underline{F\text{-value}}} = \frac{MSB}{MSW}$$

MSB → Mean square between

MSW ↳ Mean square within

$$\mu = (70+68+75+72+74+65+66+70+60+68+80+78+85+82+84)/15$$



N → # groups

n → # samples

sum of squared between —— $SSB = \sum_{i=1}^{N} n * (\bar{X}_i - \hat{X})^2$

↳ $8.88 + 268.88 + 375.55 = 653.33$

$$SSB_A = 5 * (71.8 - 73.13)^2$$

$$= 8.88$$

$$SSB_B = 5 * (65.8 - 73.13)^2 = 268.88$$

$$SSB_C = 5 * (81.8 - 73.13)^2$$

$$= 375.55$$

SSW (sum of square within group)

$$SSW_A = (70 - 71.8)^2 + (68 - 71.8)^2 + (75 - 71.8)^2 + (72 - 71.8)^2 + (74 - 71.8)^2$$

$$\Rightarrow 32.8$$

$$SSW_B \Rightarrow (65 - 65.8)^2 + (66 - 65.8)^2 + (70 - 65.8)^2 + (60 - 65.8)^2 + (68 - 65.8)^2$$

$$= 56.8$$

$$SSW_C \Rightarrow 32.8$$

$$SSW = SSW_A + SSW_B + SSW_C = 122.4$$

$$15 - 3 = 12$$

$SSB = 653.33$

$SSW = 122.4$

$K \rightarrow$ # groups

$N \rightarrow$ # observation

$$MSB = SSB / \overset{k-1}{df\_between} \Rightarrow 653.33/2 \Rightarrow 326.67$$

$$MSW = SSW / \overset{N-k}{df\_within} = 122.4/12 \Rightarrow 10.2$$

$$F = \frac{MSB}{MSW} = \frac{326.67}{10.2}$$

$$= 32.03$$

$A \longrightarrow n-1 \Rightarrow 4$

$B \longrightarrow n-1 \Rightarrow 4$

$C \longrightarrow n-1 \Rightarrow 4$

p-value = 0.0001

<

$\alpha = 0.05 \longrightarrow$ Reject H_0

the average mean of A, B & C is significantly different

Talk

You work for a company that produces three types of energy drinks. The company wants to know if the average caffeine content differs significantly among these three types. You collect data on the caffeine content (in milligrams) from five samples of each drink:
Energy Drink A: 105 mg, 110 mg, 98 mg, 107 mg, 103 mg
Energy Drink B: 130 mg, 125 mg, 132 mg, 128 mg, 129 mg
Energy Drink C: 120 mg, 115 mg, 117 mg, 119 mg, 121 mg
Question: Is there a statistically significant difference in the average caffeine content between these three types of energy drinks?

Caffeine Content

A = [105, 110, 98, 107, 103]

B = [130, 125, 132, 128, 129]

C = [120, 115, 117, 119, 121]

$\alpha = 0.05$

$H_0 \rightarrow \mu_A = \mu_B = \mu_C$

$H_1 \rightarrow$ Atleast one of the energy drink have significant difference in the caffeine Consumption

factor (variable)	Level (groups)

one way ANOVA

Real time $\longrightarrow$ Applications of ANOVA Test (Heavily used)

1. Product Testing — V_A, V_B, V_C

2. Clinical Trials — d_A, d_B, d_C

3. Marketing — S_A, S_B, S_C

Chi-square test