Population vs Sample

Time consumption

Cost consumption

GPUs

Population

Sample

Exp

$\bar{x} \approx \mu$ (Population mean)

$s \approx \sigma$ (Population std)

Sample mean $\rightarrow \bar{x}$

Sample std $\rightarrow s$

$N = 10{,}000$

$n = 100$

Population std	Sample std
$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$	$s = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$

Degree of freedom

unbiased parameter

Bessel's correction

$$\sigma \not\approx \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$$

(Biased result)

Type 1 vs Type 2 Error

H_0 → Patient is not diabetic

H_1 → Patient is diabetic

H_0 / Exp	Reality: 1	Reality: 0
Reject	Correct	~~Type 2~~ Error (Type 1)
Fail to Reject	Error (Type 2) ~~(Type 1)~~ → Talk 2	Correct

↓↓ Recall ← Healthcare

→ Evaluation metric

Key takeaway :-

Type-1 Error: The patient is falsely diagnosed as diabetic (when they're not).
Type-2 Error: The patient is falsely diagnosed as not diabetic (when they actually are).

→ Critical

Statistical Analysis

Z-test & t-test

σ is available

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \qquad \Big| \qquad \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

critical value (table)

Calculated statistical value $<$ critical value

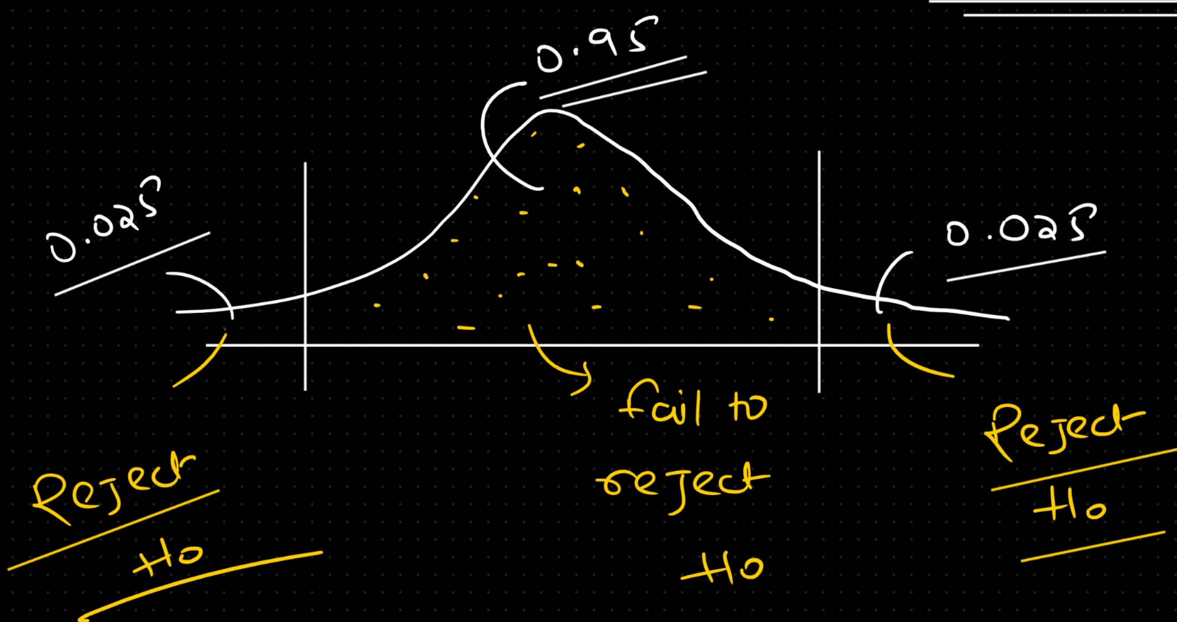
Conclusion $\longrightarrow$ fail to reject the H_0

Calculated statistical value $>$ critical value

Conclusion $\longrightarrow$ Reject the H_0

$0.05 = \alpha \rightarrow$ significance value

$CI \rightarrow 1 - \alpha = 1 - 0.05$
$= 0.95$



The average height of adults in a certain country is known to be 168 cm. A scientist believes that the average height of adult residents in a certain city in that country is different. The scientist measures the heights of 40 randomly selected adult residents of the city and finds an average of 171 cm with a standard deviation of 7 cm. Test the scientist's hypothesis at the 0.10 significance level.

$\alpha = 0.10$

$H_0 \rightarrow \mu = 168\,cm$

$H_1 \longrightarrow \mu \neq 168\,cm$

$$\frac{171 - 168}{7/\sqrt{40}}$$

$n = 40$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$\bar{x} = 171\,cm$

$s = 7\,cm$

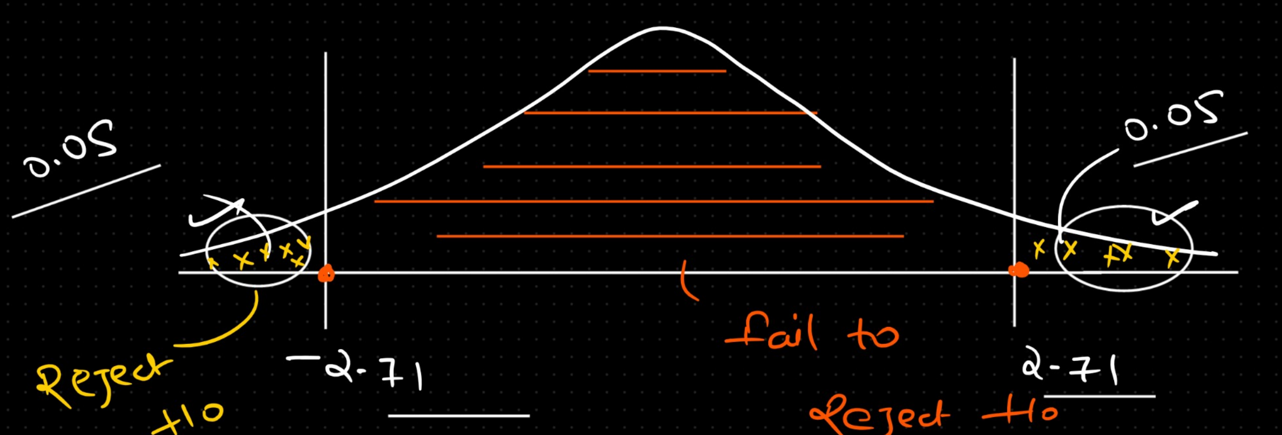
$= 2.71$

critical value $\Rightarrow 1.68$

t-value > critical value

$\hookrightarrow$ Reject the H_0

$\alpha = 0.10$



$t = -2.71$

Critical value

$abs(t) = 2.71$

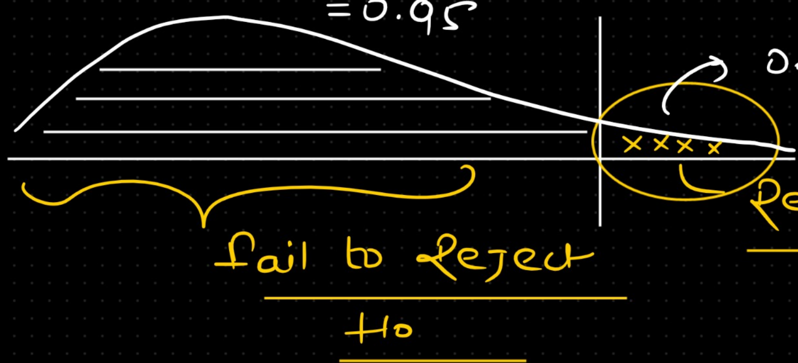
Chi-square Test

→ Categorical data points

Right skewed

$\alpha = 0.05$

$CI = 1 - 0.05$
$= 0.95$



Use chi-square test to determine b/w observed & expected weight distribution in the population at 5% α value.

Weight Category (kg)	Observed Frequency	Expected Frequency
1 50-60	40	45
2 60-70	55	50
3 70-80	30	35
4 80-90	25	20
5 90-100	15	15

num of categories = 5

Null Hypothesis (H_0) → No significant difference b/w observed & expected frequency

H_1 ⇒ There is a significant difference b/w observed & expected frequency

Step-1

Chi-square test ⇒ $\sum \frac{(O_i - E_i)^2}{E_i}$

$$\frac{(0-E)^2}{E}$$

$$\frac{(40-45)^2}{45} + \frac{(55-50)^2}{50} + \frac{(30-35)^2}{35} + \frac{(25-20)^2}{20} + \cancel{\frac{(15-15)^2}{15}}$$

$$\frac{25}{45} + \frac{25}{50} + \frac{25}{35} + \frac{25}{20}$$

$$\Rightarrow \underline{3.02}$$

step 2 — Critical value

$\alpha = 0.05$

$CI = 0.95$

$df = \underbrace{\text{num of}}_{\text{categories}} - 1$

$= 4$

Critical value $= 9.488$

≈ 9.49

step 3 — Calculated value $\Rightarrow$ 3.02

Critical value $\rightarrow$ 9.49

Calc. $\longrightarrow$ fail to reject the H_0

ANOVA Test

→ Analysis of Variance

f - table → Critical value

Researcher wants to determine if there is a significant difference b/w the avg weight of individuals based on their dietary habits.

Group 1 (Low-carb diet)	Group 2 (High-protein diet)	Group 3 (Balanced diet)
65	70	68
60	72	66
58	69	65
63	71	67
62	73	66

$\alpha = 0.05$

ANOVA Test

$H_0 \rightarrow$ Mean weight of individuals in three diet groups all are same

$H_1 \rightarrow$ Atleast one group have different mean weight

F - table