Data Science and Artificial Intelligence

Probability and Statistics

Introduction to Sampling Distribution

Lecture No.- 04

By- Rahul Sir

Topics to be Covered

Topic Questions based on chi-square Distribution

Chi-square Distribution:

In Gamma Distribution (λ, α)

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma\alpha} e^{-\lambda x} x^{\alpha-1}$$

$\alpha = \nu/2 \quad \lambda = \frac{1}{2}$

$$\chi^2 \quad f(x) = \frac{\left(\frac{1}{2}\right)^{\nu/2}}{\Gamma\frac{\nu}{2}} e^{-\frac{x}{2}} x^{\nu/2 - 1}$$

ν = degree of freedom

If n sample $\longrightarrow$ deg(f) = (n − 1)

$$E[X] = \frac{\alpha}{\lambda} = \frac{\frac{\nu}{2}}{\frac{1}{2}} = \nu$$

$\mu = E[X]$ = degree of Freedom

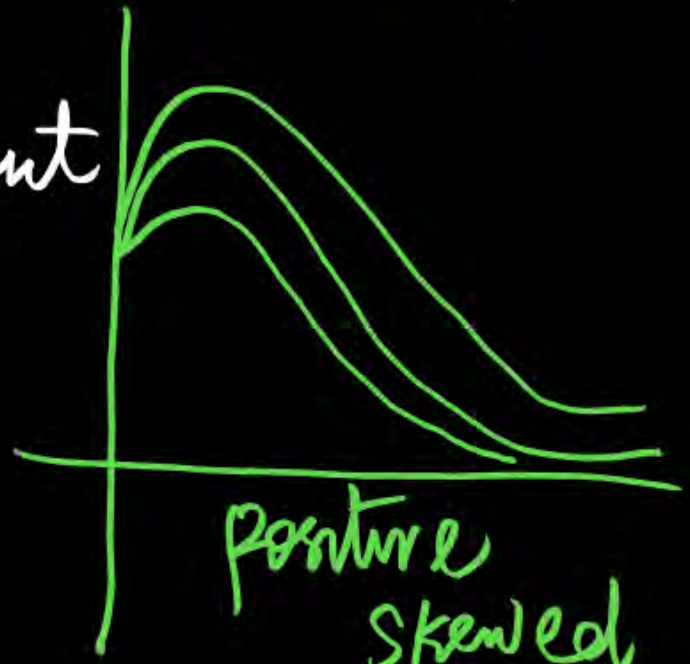
$$V(X) = \frac{\alpha}{\lambda^2} = \frac{\nu/2}{(1/2)^2} = 2\nu$$

$= 2 \times df$

$$\chi^2_{(1)} = \sum_{i=1}^{n} z_i^2$$

Sum of n Independent random var.

$$\chi^2_{(1)} = \sum_{i=1}^{n} z_i^2$$



Moment generating function

$$M_X(s) = \left(\frac{\lambda}{\lambda - s}\right)^{\alpha} = (1-2s)^{-\frac{\nu}{2}}$$

$\lambda = \frac{1}{2}$ $\alpha = \frac{\nu}{2}$

$$\boxed{\text{Moment generating function} = (1-2s)^{-\frac{\nu}{2}}}$$

→ Chi-square Distribution $\chi^2_{(1)}$

Q1. Write down the pdf of chi-square distribution in each of the following cases:

(i) 6 degrees of freedom

(ii) 10 degrees of freedom

Chi-square Distribution

$f(x) = \frac{\lambda^{\alpha}}{\Gamma\alpha} e^{-\lambda x} x^{\alpha-1}$

$\alpha = \frac{\nu}{2} \quad \lambda = \frac{1}{2}$

$f(x) = \frac{\left(\frac{1}{2}\right)^{\alpha}}{\Gamma \nu/2} e^{-\frac{1}{2}x} x^{\frac{\nu}{2}-1}$

$f(x) = \frac{\left(\frac{1}{2}\right)^{3}}{2} e^{-\frac{x}{2}} x^{2}$

✓ Chi-square dist → cumulant frequency

$\underline{\underline{\nu = 6}}$

$f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{6}{2}}}{\Gamma\frac{6}{2}} e^{-\frac{x}{2}} x^{\frac{6}{2}-1}$

$$f(x) = \frac{\left(\frac{1}{2}\right)^3}{\overline{|2}} e^{-\frac{x}{2}} x^2$$

$$\boxed{f_{\chi^2_{(1)}} = \frac{1}{16} e^{-\frac{x}{2}} (x^2)^2}$$

$x \rightsquigarrow \chi^2_{(1)}$

$$\boxed{f(\chi^2) = \frac{1}{16} e^{-\frac{x}{2}} (x^2)^2}$$

$$\boxed{f(\chi^2) = \frac{1}{16} e^{-\frac{\chi^2}{2}} (\chi^2)^2}$$

d f = 10 degree

$$f(x) = \frac{\left(\frac{1}{2}\right)^5}{\overline{|5}} e^{-\chi^2/2} (\chi^2)^4$$

Do yourself

Q2. Below, in each case, the pdf of chi-square distribution is given. Obtain the degrees of freedom of each chi-square distribution:

(i) $f\left(\chi^2\right)=\frac{1}{96}e^{-x^2/2}\left(\chi^2\right)^3; \quad 0<\chi^2<\infty$

(ii) $f\left(\chi^2\right)=\frac{1}{2}e^{-x^2/2}; \quad 0<\chi^2<\infty$

$$f(x)=\frac{\left(\frac{1}{2}\right)^{\nu/2}}{\Gamma\frac{\nu}{2}}e^{-\frac{x}{2}}(\chi^2)^{\frac{\nu}{2}-1}$$

Chi square Distribution (1) df = 8 ✓

degree of freedom: (2) df = 1

$\boxed{n \text{ SAMPLE}}$ $\boxed{\nu=(n-1)}$

df = 8 ✓
df = 1 ✓

Q3. What are the mean and variance of chi-square distribution with 10 degrees of freedom?

Chi-square Distribution $\nu = 10$ degree

$$\text{MEAN} = \frac{\alpha}{\lambda} = \frac{\frac{10}{2}}{\frac{1}{2}} = 10$$

$$\text{variance} = \frac{\alpha}{\lambda^2} = \frac{\nu/2}{\left(\frac{1}{2}\right)^2} = 20$$

$$\alpha = \frac{\nu}{2}$$
$$\lambda = \frac{1}{2}$$

Introduction to Sampling Distribution

Q4. What are the mean and variance of chi-square distribution with pdf given below

$$f\left(\chi^2\right) = \frac{1}{96} e^{-x^2/2} \left(\chi^2\right)^3; \qquad 0 < \chi^2 < \infty$$

$$f(\chi^2) = \frac{1}{96} e^{-\chi^2/2} (\chi^2)^3$$

mean = 8

variance = 16

CHI-Square Distribution $\chi^2_{(1)} = \sum_{i=1}^{n} z_i^2 =$ $Z = \frac{\bar{X}-\mu}{\sigma}$

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

✓ Observed Frequency

	Right	Wrong	Total
	27	73	100
Expected Frequency	$\boxed{20}$	$\boxed{80}$	100
Prob.	$\frac{1}{5}$	$\frac{4}{5}$	1

$\boxed{\text{Man}}$ — 27 Right, 73 Wrong

100 SAMPLE

	Right	Wrong	
Observed	207	793	1000
Expected Frequency	200	800	1000
	$\frac{1}{5}$	$\frac{4}{5}$	1

If SAMPLE size Increases

$n = 1000$

$\frac{O}{E}$

Observed Frequency	27	73	100
Expected Frequency	20	80	100
Prob.	$\frac{1}{5}$	$\frac{4}{5}$	1
deviation	+7	−7	

O	27	73	100
E	20	80	100
O − E	$\boxed{+7}$	$\boxed{-7}$	0
$\frac{(O-E)^2}{E}$	$\frac{(7)^2}{20}$	$\frac{(-7)^2}{80}$	

$\frac{O}{E}$

This is also important No

This No

Control The variance

$$\left.\begin{aligned}\mu &= 0\\ \sigma &= 1\end{aligned}\right] [N(0,1)]^2 = \chi^2_1$$

$O - E = \text{Deviation}$

Frequency $\frac{(O-E)^2}{E_i} = \text{SVM}$

$$\sum \frac{D_i}{E_i} = \sum_{i=1}^{n} \frac{(\text{Observed} - \text{Expected})}{\text{Expected}}$$

O	$\boxed{27}$	$\boxed{73}$	100
E	20	80	100
$\sum_{i=1}^{n} \frac{D_i^2}{E_i}$	$\left(\frac{7^2}{20}\right)$	$\frac{(-7)^2}{80}$	$\boxed{3.0625}$

χ^2 (No)

degree of freedom $= 2-1 = 1$

very important No

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{n} \frac{O_i^2}{E_i}$$

choice	Right	wrong	
O	27	73	100
P	$\frac{1}{5}$	$\frac{4}{5}$	

O	O_1	O_2	O_3	O_4	O_n
E	E_1	E_2	E_3	E_4	E_n

$$\frac{D^2}{E_i} = \frac{(O_i - E_i)^2}{E_i} \quad \frac{(O_1 - E_1)^2}{E_1} \quad \frac{(O_2 - E_2)^2}{E_2} \quad \frac{(O_3 - E_3)^2}{E_3} \cdots$$

Bernoulli Trials $P(S) = \frac{1}{5}$ $P(F) = \frac{4}{5}$

for large SAMPle of data $n = 100$ SAMPle.

$$B(n, p) = B\left(100, \frac{1}{5}\right)$$

$\underline{27}$

$B\left(100, \frac{1}{5}\right)$ is consistent or Not

$$B\left(100, \frac{1}{5}\right) \xrightarrow[n=100]{\text{large}} N(\mu, \sigma^2) = N(np, npq) = N\left(100 \times \frac{1}{5}, 100 \times \frac{1}{5} \times \frac{4}{5}\right)$$

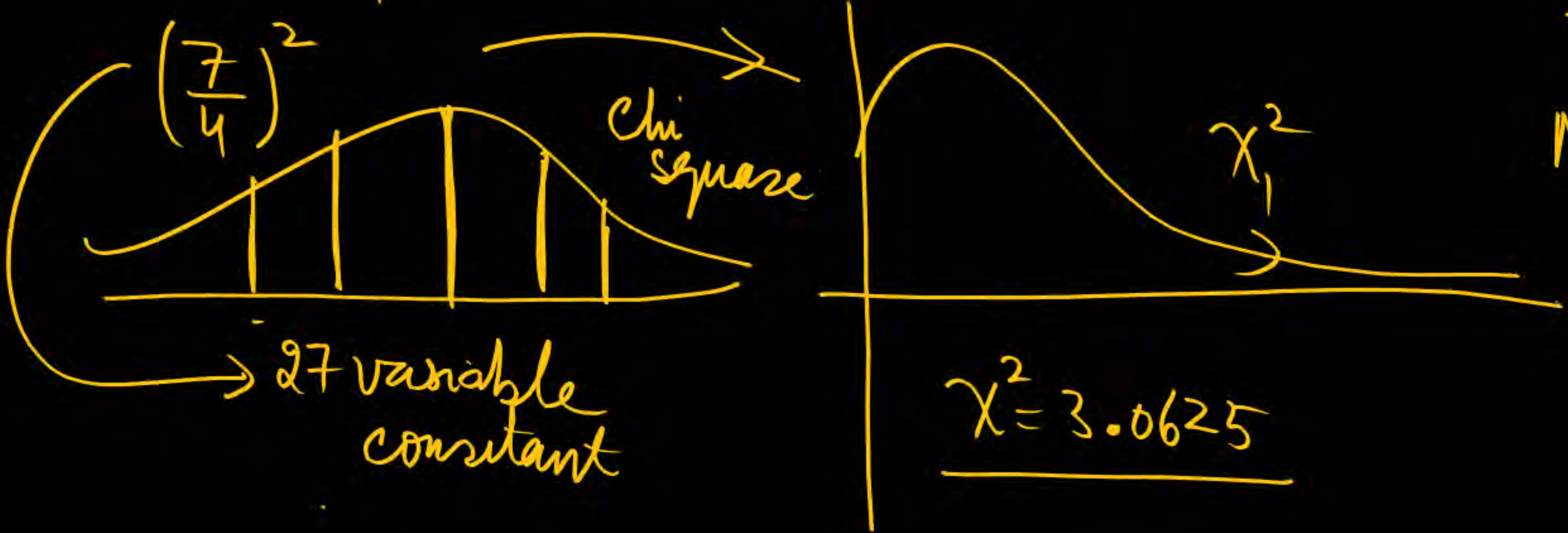
$N(20, 16)$ is constant

If X is consistent $N(\mu, \sigma^2)$

If Z is consistent $N(0,1)$

$$N(20,16) = \frac{X-\mu}{\sigma} = \frac{27-20}{\sqrt{16}} = \frac{7}{4}$$

$\frac{7}{4}$ is consistant $N(0,1)$



sum of n

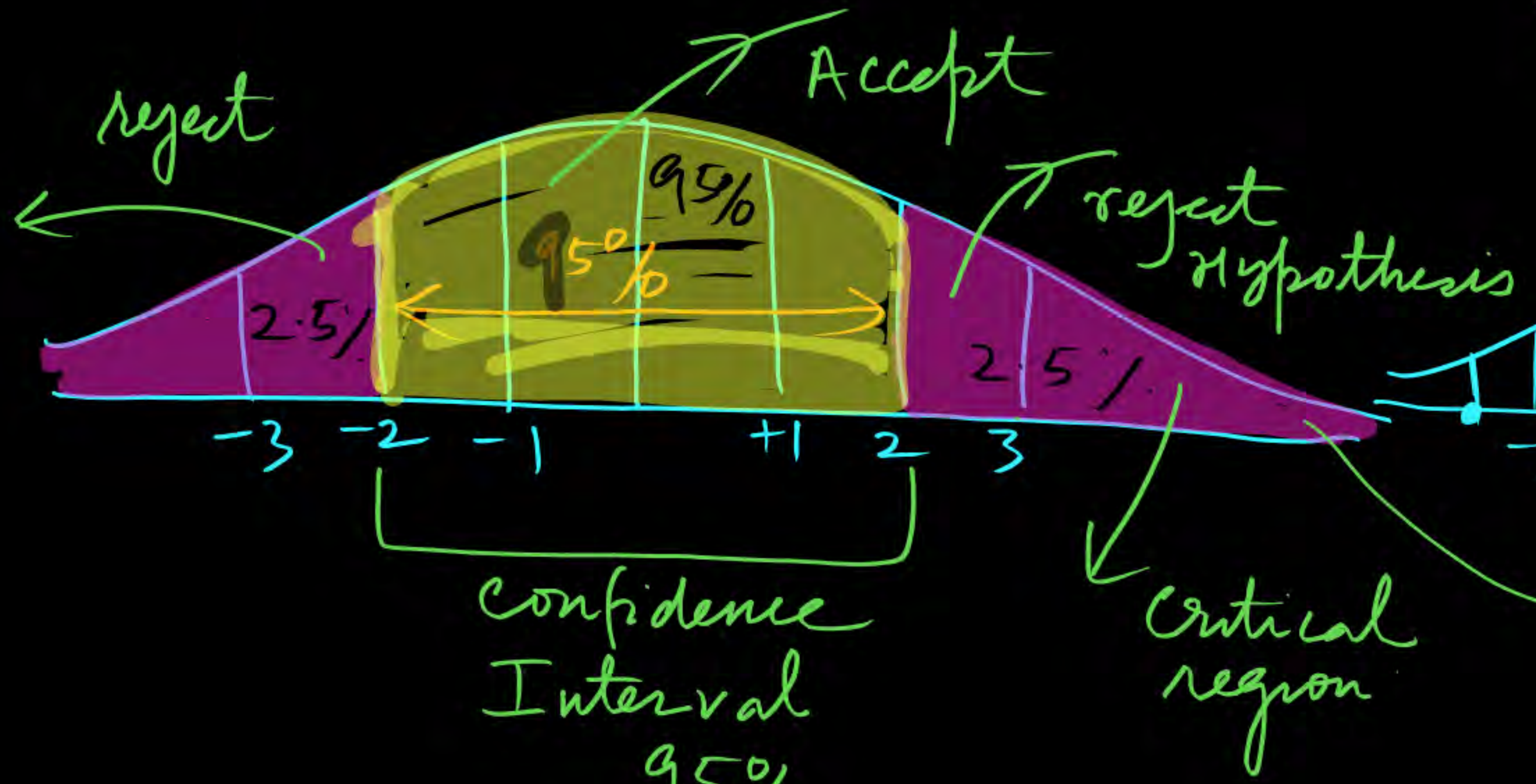
Independ.

$N(\mu, \sigma^2)$

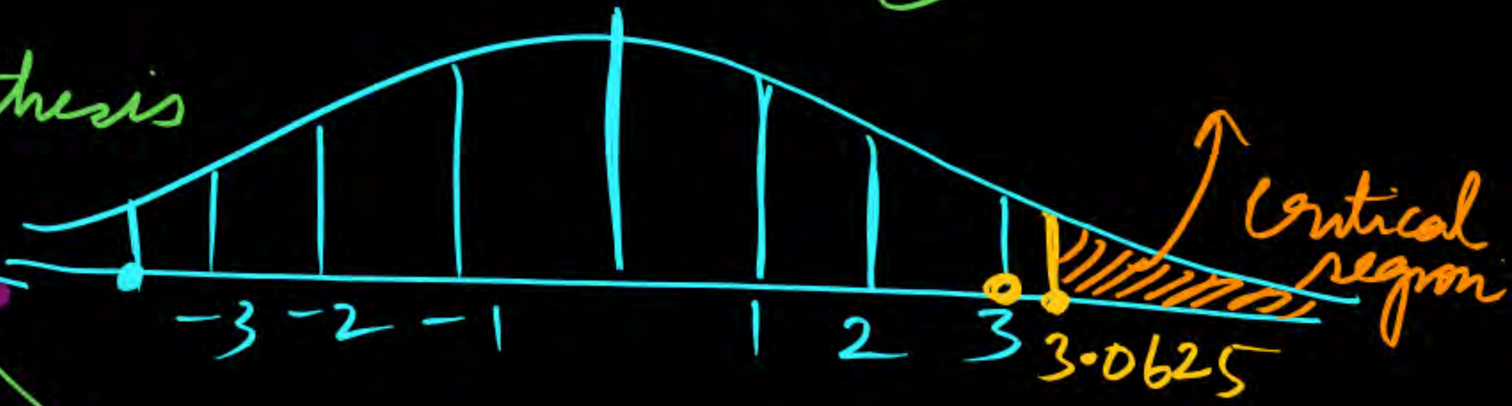
$$z_1^2 + z_2^2 + \dots + z_n^2 = \chi_1^2$$

If $Z = N(0,1)$

$$N(0,1) = Z^2 \approx \chi^2_{(1)}$$



$\chi^2 = \alpha \qquad \chi^2 > 0,\ < 0,\ = 0$



Reject The Hypothesis

Significance Level

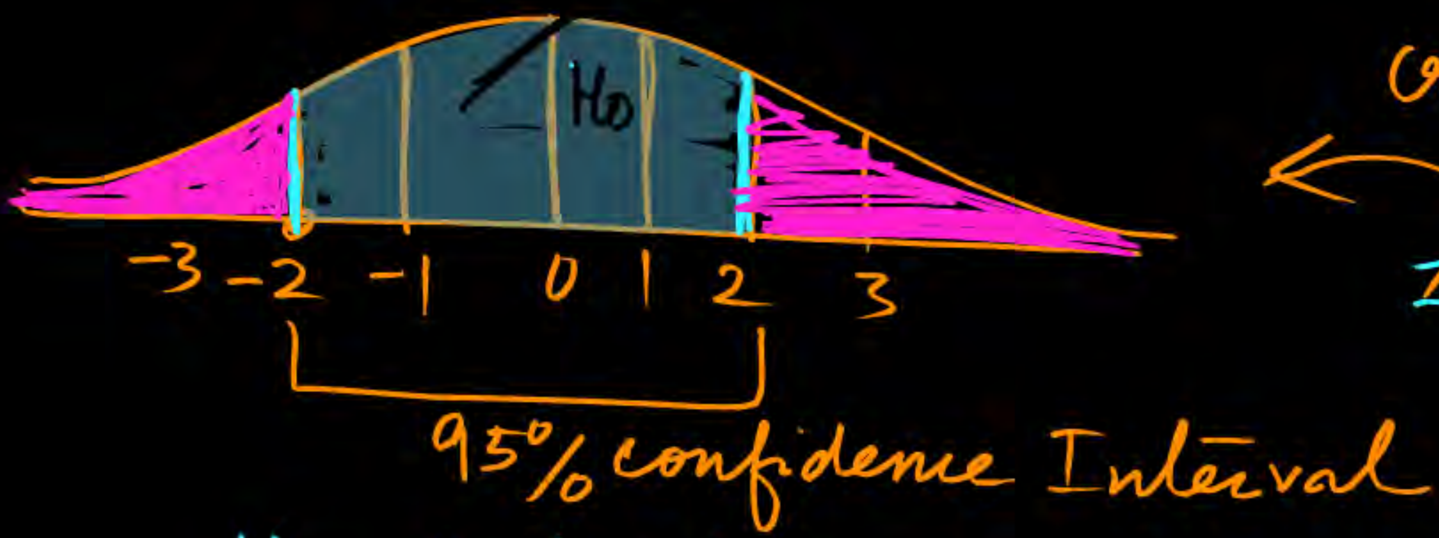
$$\chi^2 = (\text{Number})$$

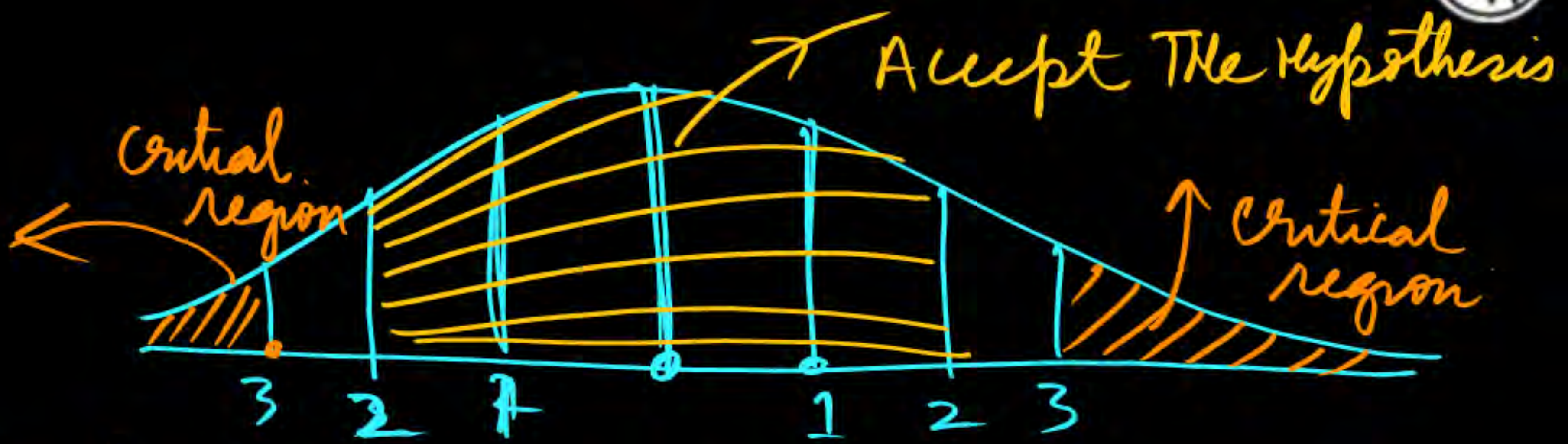
$$\chi^2 = \sum_{i=1}^{n} \frac{D_i^2}{E_i}$$

$$\text{Degree of freedom} = (n-1) = 2-1 = \underline{\underline{2}}$$

Degree of Freedom

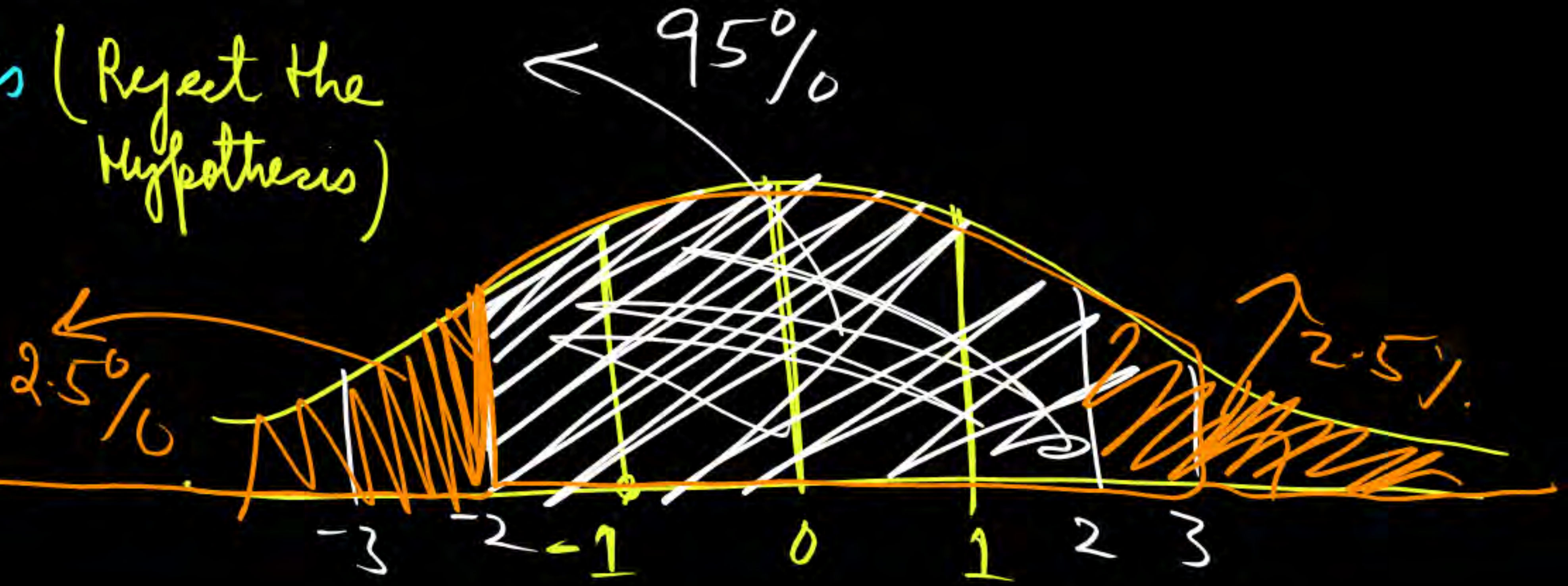
1
2
3
4
5
6
7





H_0 = Null Hypothesis

H_1 = Alternate Hypothesis (Reject the Hypothesis)



✓ SAMPLE 50

✓ Cell - 5

If $\chi^2 = 0$ [Null Hypothesis Accepted)

$\chi^2 > 0$ [don't agree exactly)

data Inaccurate — DATA - discrepancy.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \sum_{i=1}^{n} Z_i^2$$

N(0,1)

✓ Table:

	Right	Wrong
O	207	793
E	200	800
P	$\frac{1}{5}$	$\frac{4}{5}$

Accepted
↓
Large
SAMPLE
of
data

Chi square
TEST
valid

deg

1

2

sig

2) ✓2011

H_0 = Accepted

$\chi^2 = (\quad)$ significance

60% 1 car

28% 2 car

12% more than 3 car

Use a significance level 5%

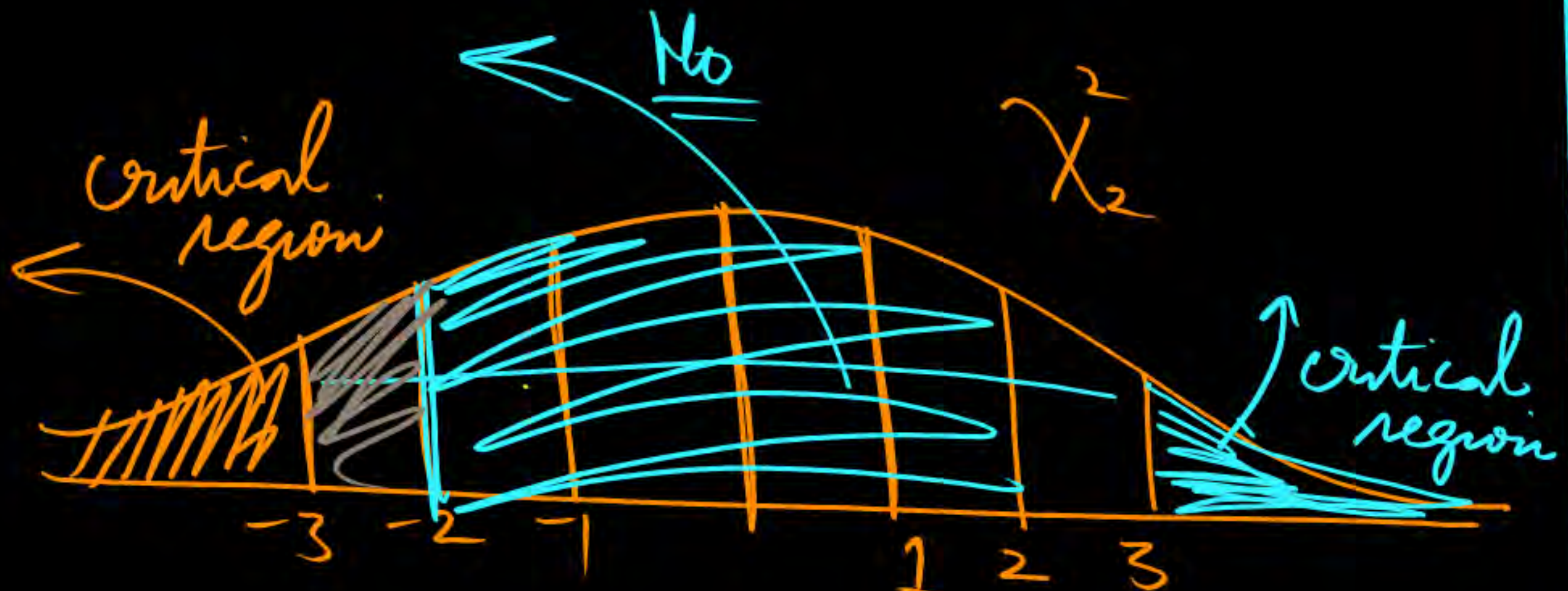
129 car owners

73 — one car

38 — Two car

$\chi^2 = 0.7$

Observed	ExpE	O − E	$\sum \frac{(O-E)^2}{E}$
73	0.6×129 =	$O_1 - E_1$	✓
38	0.28×129 =	$O_2 - E_2$	✓
19	0.12×129 =	$O_3 - E_3$	✓
129			



THANK - YOU