Computer Science & DA

Probability and Statistics

Sampling Theory & Distribution

Lecture No. 05

By- Dr. Puneet Sharma Sir

Recap of previous lecture

Topic Chi-square Distribution

Topics to be Covered

Topic Miscellaneous

Topic :Miscellaneous

RECAP of CHI-SQUARE :-

Type (1) significance of pop Variance :-

$H_0 : \boxed{\sigma = \sigma_0}$, $H_1 : \sigma \neq \sigma_0$

$$\boxed{\chi^2 = \frac{ns^2}{\sigma^2}}$$

Type (2) Significance of Goodness of fit -

$H_0 : \boxed{O_i = E_i}$, $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

$H_1 : O_i \neq E_i$

Type (III) : Significance of Independence of Attributes -

H_0 : there is No Relationship b/n two Attributes i.e they are Ind

H_1 : there is a relationship b/n two Attributes i.e they are Dep.

$$\boxed{\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}} \quad \nu = (R-1)(C-1)$$

$$E_{ij} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Total}} = \frac{R_i \times C_j}{\text{Total}}$$

#Q. A random sample of size 25 has sample standard derivation = 9. Test the Hypothesis that population standard derivation is 10.5.

Type2

Given $x_{24}^{2}(0.05) = 36.42$

Already Solved YESTERDAY

[NAT]

#Q. (i) A random sample of size 20 has variance 25. If the population standard derivation is 8 calculated x^2

(ii) Also calculated the hypothesis that $\sigma = 8$ for 95% confidence region. Given $x_{19}^2(0.05) = 30.14$

Already Solved YESTERDAY

#Q. The following table shows the numbers of car accidents that occurred during the week. Test whether the accident are uniformly distributed over the week.

Type 2

Given $x_6^2(0.05) = 12.592$

Day	**Mon**	**Tue**	**Wed**	**Thurs**	**Fri**	**Sat**	**Sun**	**Total accident**
Number of accident	14	18	12	11	15	14	14	98

Already solved YESTERDAY

#Q. A die is thrown 276 times and frequencies for various outcomes are $f(1) = 40$, $f(2) = 32$, $f(3) = 29$, $f(4) = 59$, $f(5) = 57$, $f(6) = 59$. Test whether die is Biased or not?

it is given that $x_5^2(0.05) = 11.09$

Already Solved YESTERDAY

#Q. A car manufacture company produces four different color in which 882 are white, 313 Grey, 287 Red and 118 Black. While at the beginning they had decided theoretically that they will produce white, Green, red, Black cars in the ratio 9 : 3 : 3 : 1 respectively. If $x_3^2(0.05) = 7.815$, Establish [redacted]. the correspondence between theoretical and [redacted] experimental result

Note: In a Mixture of 70 ltrs, $M : W = 4 : 3$

$M = \left(\frac{4}{4+3}\right) \times 70 = 40$ ltrs

$W = \left(\frac{3}{4+3}\right) \times 70 = 30$ ltrs

$E_W = \left(\frac{9}{9+3+3+1}\right) \times 1600 = 900$

$\nu = 4 - 1 = 3$

$E_G = \frac{3}{16} \times 1600 = 300$

$E_R = \frac{3}{16} \times 1600 = 300$

$E_B = \frac{1}{16} \times 1600 = 100$

CARS	O_i	E_i	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
White	882	900	324	–
Grey	313	300	169	–
Red	287	300	169	–
Black	118	100	324	–
				$\sum = 4.72$

i.e. $\chi^2 = 4.72$

So H_0 is Accepted

i.e there is a Relationship b/w Th. Value & Exp Value

H_0: there is no difference b/w Exp Values and Theoretical Values

H_1: " is significant " " " "

#Q. What are the expected frequency of 2 × 2 contingency table given below:

Type3

(1)

a	**b**	a+b
C	d	c+d
a+c	b+d	(a+b+c+d)

$$E(a) = \frac{(a+b)(a+c)}{a+b+c+d}$$

$$E(b) = \frac{(a+b)(b+d)}{a+b+c+d}$$

$$E(c) = \frac{(c+d)(a+c)}{a+b+c+d}$$

$$E(d) = \frac{(c+d)(b+d)}{a+b+c+d}$$

(2)

2	**10**	12
6	6	12
8	16	24

$$E(2) = \frac{12 \times 8}{24}$$

$$E(10) = \frac{12 \times 16}{24}$$

$$E(6) = \frac{12 \times 8}{24}$$

$$E(6) = \frac{12 \times 16}{24}$$

#Q. In an engineering college of Delhi, the IQ and economic condition at home of 1000 students are given in the table;

I.Q / Eco-condition	High	Low	Total
Rich	100	300	400
Poor	350	250	600
Total	450	550	1000 = Total

$E(100) = \frac{400 \times 450}{1000} = 180$

$E(300) = \frac{400 \times 550}{1000} = 220$

$E(350) = \frac{600 \times 450}{1000} = 270$

$E(250) = \frac{600 \times 550}{1000} = 330$

It is 2 × 2 contingency table $\Rightarrow \nu = (2-1)(2-1) = 1$

can we conclude any association between economic condition at home and IQ of the students. Given $x_1(0.05) = 3.84$

H_0: Two Attributes are Ind

H_1: " " " Dep.

O_i	E_i	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
100	180	$(-80)^2=6400$	—
300	220	$(80)^2=6400$	—
350	270	$(80)^2=6400$	—
250	330	$(-80)^2=6400$	—
			$\sum=107.73$

is $\chi^2=\frac{\sum(O_i-E_i)^2}{E_i}=107.73$

So we will Reject H_0 & Accept H_1

is there is a Relationship b/w E.C & I.Q

#Q. From the following data, Find out whether there is any relationship between sex and preference of color;

Color \ Sex	Males	Female	Total
Red	10	40	50
White	70	30	100
Green	30	20	50
Total	110	90	200 = N

It is 3 × 2 contingency table so difference = (3 − 1) (2 − 1) = 2 and $x_2^2(0.05) = 5.991$

$E(10) = \frac{50 \times 110}{200} = 27.5$

$E(40) = \frac{50 \times 90}{200} = 22.5$

$E(70) = \frac{100 \times 110}{200} = 55$

$E(30) = \frac{100 \times 90}{200} = 45$

$E(30) = \frac{50 \times 110}{200} = 27.5$

$E(20) = \frac{50 \times 90}{200} = 22.5$

~~$E(110) =$~~ Blunder

~~$E(90) =$~~

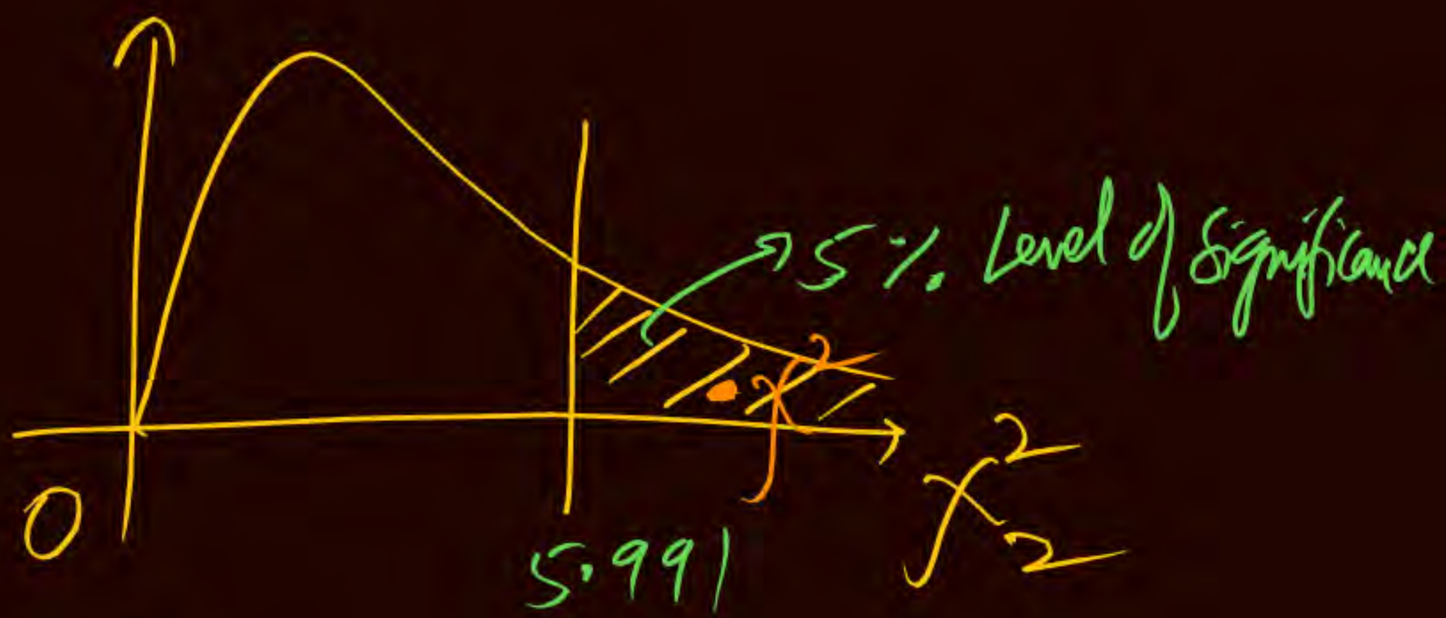
Sol: $d.f(\nu) = (3-1)(2-1) = 2$

H_0: there is No Relationship b/w SEX & Preference of Colour i.e. Ind

H_1: " is a " " " " i.e. Dep

O_i	E_i	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
10	27.5	$(-17.5)^2 = 306.25$	—
40	22.5	$(+17.5)^2 = 306.25$	—
70	55	225	—
30	45	225	—
30	27.5	6.25	—
20	22.5	6.25	—

$$\Sigma = 34.34 = \chi^2$$



$$\because \chi^2_{Cal} > \chi^2_{Tabulated}$$

So H_0 is Rejected & H_1 is Accepted

#Q. In a survey of 200 boys of which 75 were intelligent, and remaining were unintelligent out of intelligent boys 40 had educated father while out of unintelligent boys 85 had uneducated fathers.

Can we conclude that educated fathers have intelligent boys given that $x_1^2(0.05) = 3.841$

Father \ Boy	Intelligent	unintelligent	
Educated father	40	40	80
uneducated father	35	85	120
	75	125	Total = 200

$\nu = (R-1)(C-1) = (2-1)(2-1) = 1$

$E(40) = \frac{80 \times 75}{200} = 30$

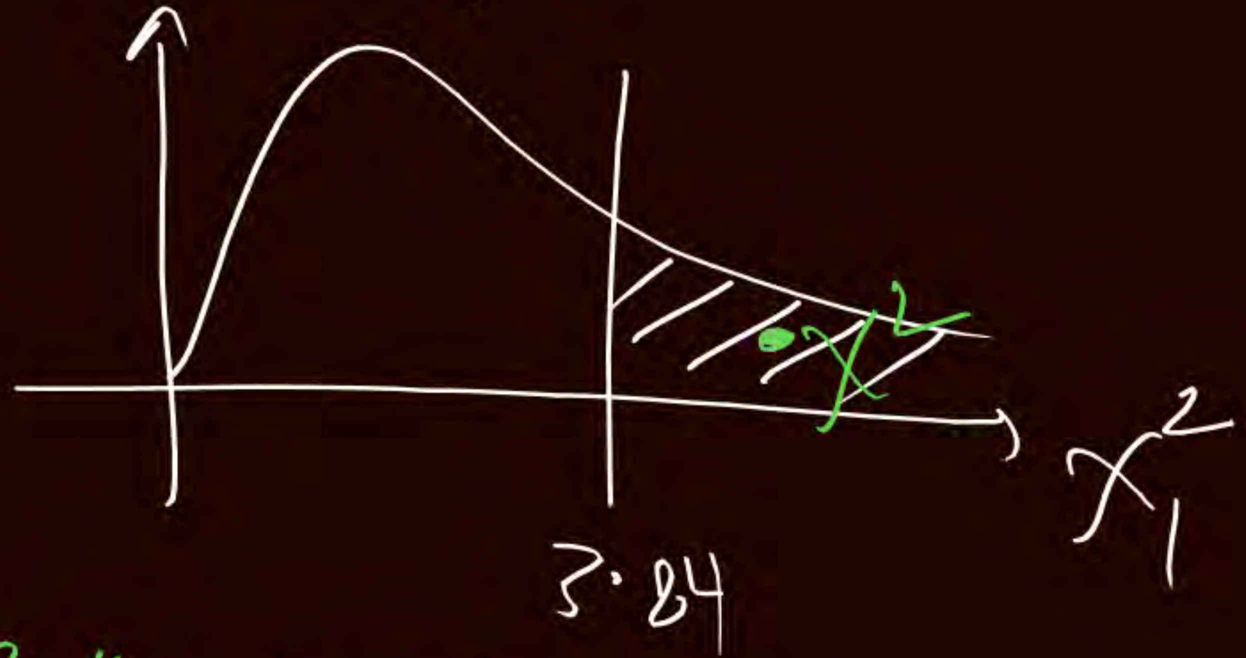
$E(40) = \frac{80 \times 125}{200} = \frac{80 \times 5}{8} = 50$

$E(35) = \frac{120 \times 75}{200} = \frac{120 \times 3}{8} = 45$

$E(85) = \frac{120 \times 125}{200} = \frac{120 \times 5}{8} = 75$

O_i	E_i	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
40	30	100	—
40	50	100	—
35	45	100	—
85	75	100	—
			$\sum =$

$$\chi^2 = \sum \frac{(O_i-E_i)^2}{E_i} = 8.88$$



So H_0 is Rejected & H_1 is Accepted

H_0: there is No Relationship b/w Education of father & Intelligence of Boys.

H_1: " " a " " " " "

#Q. The number of screw declared fit or unfit by three inspectors x, y and z is shown in the following table.

~~Inspectors~~ Insp. SCREW	X	Y	z	Total
Fit screws	50	47	56	153
Unfit screw	5	14	8	27
Total	55	61	64	80

$$d.f = (\nu) = (2-1)(3-1) = 2$$

Given $x_1^2(0.05) = 3.81$ and $\chi_2^2(0.05) = 5.99$ test the hypothesis that proportion of screws declare unfit by three inspectors are same.

H_0 : there is No difference b/w the prop of screws declared fit or unfit by 3 Inspectors (Ind)

H_1 : there is a significant diff b/w ,, ,, ,, ,, ,, ,, ,, (Dep)

- - - - - - - - H_0: Accepted

Qs If we are using Chi-Square test to fit Binomial, Poisson, Normal Dist. then degree of freedom will be respectively

(a) 1, 2, 3

(b) $(n-1), (n-2), (n-3)$

(c) $(n-1)$ in each case

(d) Can't say anything

For B. Dist : $df = n-1$ & Constraint is $\boxed{\sum O_i = \sum E_i}$

For P. Dist : $df = n-2$ & " are $\sum O_i = \sum E_i$ & Mean $= \lambda$

For N. Dist : $df = n-3$ & " are $\sum O_i = \sum E_i$, Mean $= \mu$ & SD $= \sigma$

Note: (1) for large d.f, Chi sq. dist converts into Normal Dist.

(2) " " " t-dist " " " " "

Type I $H_0 : \boxed{\sigma = \sigma_0}$

$$\boxed{\chi^2 = \frac{ns^2}{\sigma^2}}$$

Type II $H_0 : \boxed{\sum O_i = \sum E_i}$

$$\chi^2 = \frac{\sum (O_i - E_i)^2}{E_i}$$

Type III : (Independence of Attributes)

H_0 : Two Attributes are Ind.

$$\chi^2 = \frac{\sum (O_i - E_i)^2}{E_i}$$

Note – (1) $SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ (2) $SE(\tilde{p}) = \sqrt{\frac{\tilde{p}\tilde{q}}{n}}$ (3) Most Probable limits : $\bar{x} - 3SE(\bar{x}) \le \mu \le \bar{x} + 3SE(\bar{x})$

& $\tilde{p} - 3SE(\tilde{p}) \le p_0 \le \tilde{p} + 3SE(\tilde{p})$

Z-test → Type I $H_0 : \boxed{\mu = \mu_0}$

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Type II $H_0 : \boxed{\mu_x = \mu_y}$

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Type III $H_0 : \boxed{p = p_0}$

$$Z = \frac{\tilde{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Type IV $H_0 : \boxed{p_1 = p_2}$

$$Z = \frac{\tilde{p}_1 - \tilde{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

where $\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\tilde{p}_2}{n_1 + n_2}$

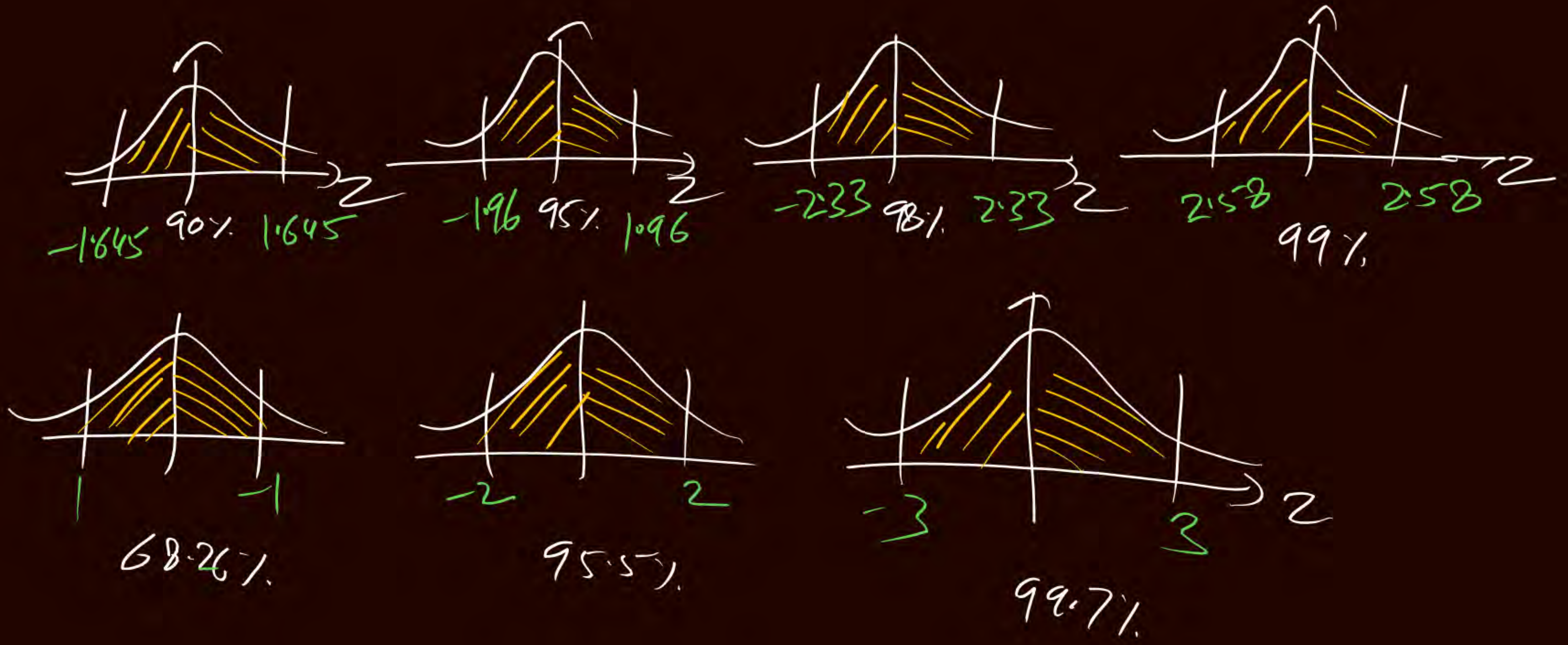
& $\hat{q} = 1 - \tilde{p}$

t-test → Type I $H_0 : \boxed{\mu = \mu_0}$,

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{where } S^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

Type II → $H_0 : \boxed{\mu_x = \mu_y}$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} \quad \text{where } S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$



z
-1.645 90% 1.645
z
-1.96 95% 1.96
z
-2.33 98% 2.33
z
2.58 2.58
99%
1 -1
68.26%
-2 2
95.5%
z
-3 3
99.7%

GATE (2024) → Two fair coins are tossed independently. X is a random variable that takes a value of 1 if both tosses are head and 0 otherwise. Y is a random variable that takes a value of 1 if at least one of tosses is head and 0 otherwise then Cov(X, Y) = ?

$S = \{\underbrace{(HH), (HT)\ (TH)}, (TT)\}$

$P(X=1) = \frac{1}{4}$, $P(X=0) = \frac{3}{4}$

$P(Y=1) = \frac{3}{4}$, $P(Y=0) = \frac{1}{4}$

X \ Y	0	1	P(X)
0	1/4	2/4	3/4
1	0	1/4	1/4
P(Y)	1/4	3/4	1

$P(X=1, Y=1) = P\{(\text{Both head}) \cap (\text{At least one } H)\}$

$= P\left[(HH) \cap \{(HH), (HT)\ (TH)\}\right] = P(HH)$

$= \frac{1}{4}$

$Cov(X, Y) = E(XY) - E(X).E(Y) = ?$

$= \frac{1}{4} - \left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{4-3}{16} = \frac{1}{16} = ?$

X:	0	1
$p(x)$:	$\frac{3}{4}$	$\frac{1}{4}$

$E(x) = \sum p_i x_i$

$= \frac{1}{4}$

Y:	0	1
$p(y)$:	$\frac{1}{4}$	$\frac{3}{4}$

$E(Y) = \sum p_i y_i$

$= \frac{3}{4}$

XY:	0	1
$p(XY)$:	$3/4$	$\frac{1}{4}$

$E(XY) = \sum (x_i y_i) p_{ij} = 1 \times \frac{1}{4}$

$X=1, Y=0$ — $XY=0$

$X=0, Y=1$ — $XY=0$

$X=0, Y=0$ — $XY=0$

$X=1, Y=1$ — $XY=1$

M-II $E(XY) = x_0 y_0 \, p(0,0) + x_0 y_1 \, p(0,1)$

$+ x_1 y_0 \, p(1,0) + x_1 y_1 \, p(1,1)$

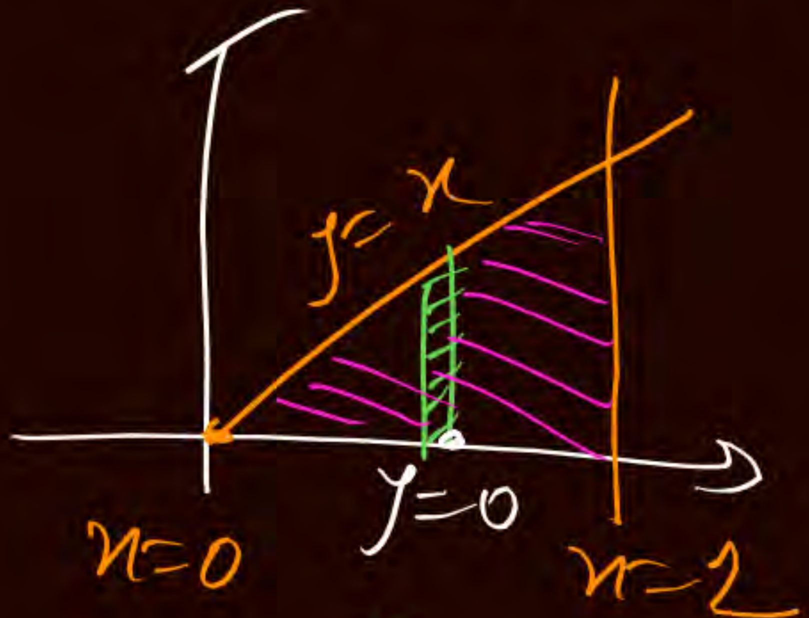
$= 0 \times 0 \left(\frac{1}{4}\right) + 0 \times 1 \left(\frac{2}{4}\right)$

$+ 1 \times 0 (0) + 1 \times 1 \left(\frac{1}{4}\right) = \frac{1}{4}$

GATE (2024) $f(x,y) = \begin{cases} 2xy, & 0<x<2,\ 0<y<x \\ 0, & \text{otw} \end{cases}$ then $E\{Y/X=1.5\} = ?$

M.T.A

<u>Verification</u>: $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{x=0}^{2}\int_{y=0}^{x} (2xy)\,dy\,dx = \int_{x=0}^{2}\left\{\frac{y^2}{2}\right\}_{y=0}^{x}(2x\,dx)$



$$= \int_{x=0}^{2}(x^3)\,dx = \frac{2^4}{4} = \frac{16}{4} = 4 \neq 1$$

M.T.A

(*) Let us assume that Q. is correct then what should be the procedure?

$$E\{Y=y/X=x\} = \int_{-\infty}^{\infty} y \cdot f\left(\frac{y}{x}\right) \cdot dy = \int_{y=0}^{x} y\left(\frac{2y}{x^2}\right) dy = \frac{2}{x^2}\left(\frac{y^3}{3}\right)_{y=0}^{x} = \frac{2}{3}x$$

where $f\left(\frac{y}{x}\right) = \dfrac{f(x,y)}{f(x)} = \dfrac{2xy}{x^3} = \dfrac{2y}{x^2}$

where $f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{y=0}^{x} (2xy)\,dy = 2x\left(\frac{y^2}{2}\right)_0^x = x^3$

$$E\{Y=y/X=1.5\} = \frac{2}{3} \times 1.5 = 1$$

THANK - YOU