Computer Science & DA

Probability and Statistics

Sampling Theory & Distribution

Lecture No. 04

By- Dr. Puneet Sharma Sir

Recap of previous lecture

Topic t - test

Topics to be Covered

Topic Chi-square Distribution

#Q. A machine produced 20 defective articles in a lot of 400 and after overhauling it produced 20 defective articles in a lot of 300. Has the machine improved at $\alpha = 0.01$.

HW Q

Z-test

For right ~~f~~tailed $Z_\alpha(0.01) = 2.33$

$H_0 : \boxed{p_1 = p_2}$, $H_1 : p_1 > p_2$

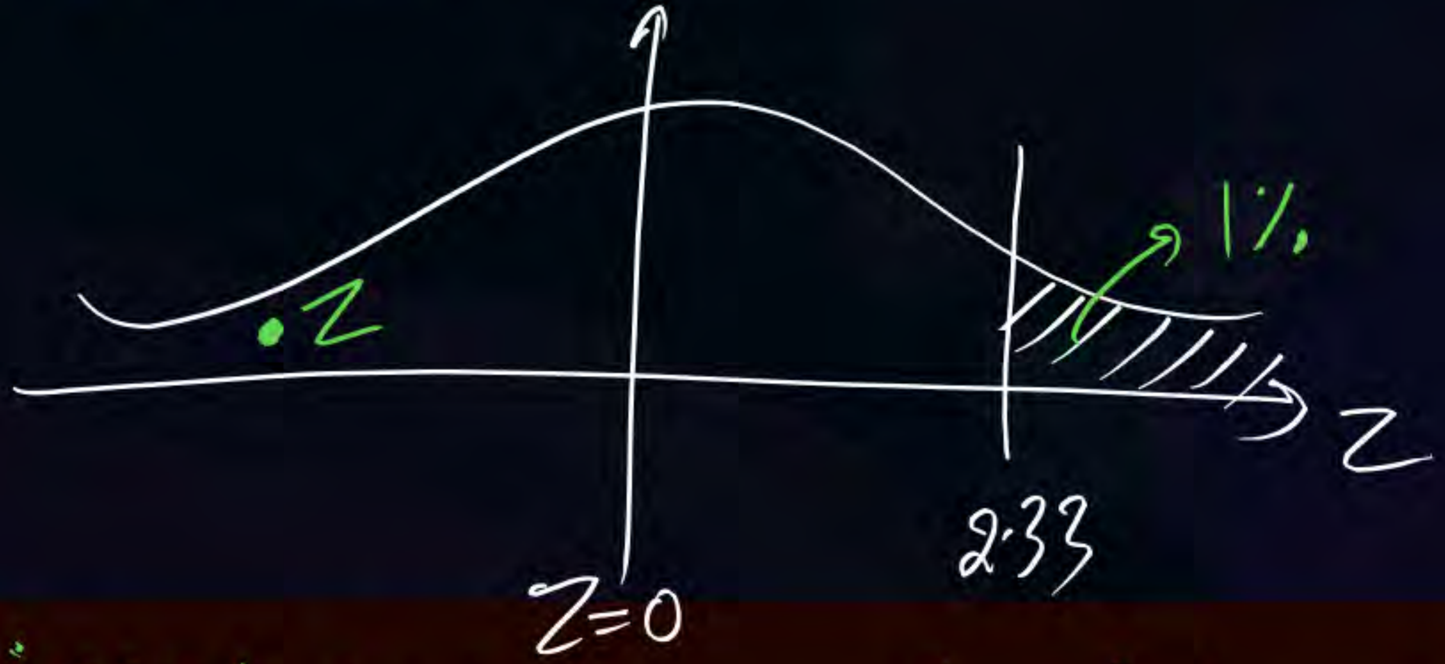
x = { Number of Def. Articles }

$\tilde{p}_1 = \frac{x_1}{n_1} = \frac{20}{400} = 0.050$

$\tilde{p}_2 = \frac{x_2}{n_2} = \frac{20}{300} = 0.067$

$\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{20+20}{400+300} = 0.057$ & $\tilde{q} = 1 - \tilde{p} = 0.943$

$$Z = \frac{\tilde{p}_1 - \tilde{p}_2}{\sqrt{\frac{\tilde{p}\tilde{q}}{n_1} + \frac{\tilde{p}\tilde{q}}{n_2}}} = -0.96$$



$\because |Z| < Z_\alpha(0.01)$ Hence H_0 is accepted $\Rightarrow$ Machine has not been improved

Student's t-Test $\boxed{n<30}$ (it also follows N. Dist)

① for one sample t-test; $\boxed{D.F = n-1}$ & for two sample t-test $\boxed{DF = n-2}$

Type I: Significance of Population Mean →

$H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

$$\boxed{t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}}}$$ where

$$S^2 = \text{sample Variance} = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

Type II Testing the significance of Difference b/n two population Mean →

$H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$

$$\boxed{t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}}$$

where $S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1}$ & $S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1}$

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

Note → Population Variance $(\sigma^2) = \frac{\sum(x-\mu)^2}{N}$, Sample Variance $(s^2) = \frac{\sum(x-\bar{x})^2}{n-1}$

Reason :→ General Reason :→ While calculating $(x-\bar{x})^2$, we take $\bar{x}$ as one of the values of x then we have only $(n-1)$ differences of the type $(x-\bar{x})$ Hence we should divide with $(n-1)$ instead of n

Mathematical Reason :→

✗ $$s^2 = \frac{\sum(x_i-\bar{x})^2}{n} = \frac{\overbrace{\sigma^2+\sigma^2+\sigma^2+\cdots+\sigma^2}^{n-1\text{ times}}}{n} = \frac{(n-1)\sigma^2}{n} \neq \sigma^2$$ ☹

✓ $$s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{\overbrace{\sigma^2+\sigma^2+\sigma^2+\cdots+\sigma^2}^{(n-1)\text{ times}}}{(n-1)} = \frac{(n-1)\sigma^2}{(n-1)} = \sigma^2$$ ☺

#Q. A random sample of 16 values from normal population has mean of 41.5 cm and sum of squares of deviation from mean is 135 cm^2. can we say that the popular mean is 43.5 cm? can we say that the population mean is 43.5 cm? with 5% level of significance.

(ii) Also find the 95% and 99% confidence for μ it is given that t_{15} (0.05) = 2.131 and t_{15}(0.01) = 2.94

(1) Already solved yesterday

$DF = n-1 = 16-1 = 15$

② $SE(\bar{x}) = \frac{S}{\sqrt{n}}$

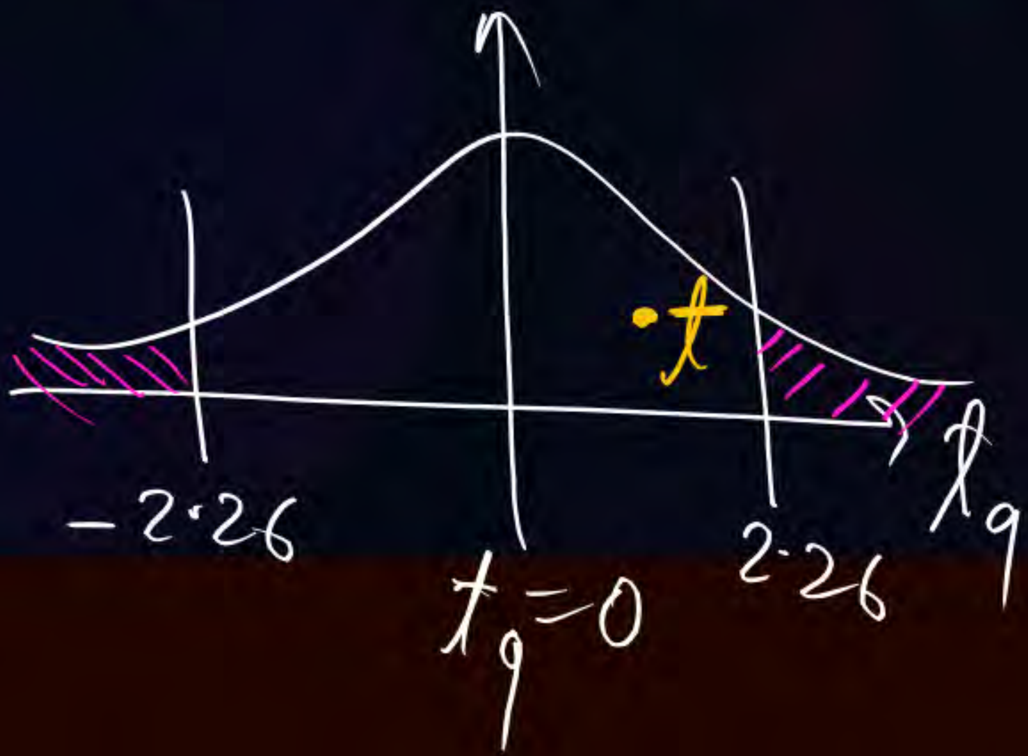
$$\bar{x} - t_{15}(0.05) SE(\bar{x}) \leq \mu \leq \bar{x} + t_{15}(0.05) SE(\bar{x})$$

(ii) $\bar{x} - t_{15}(0.01) SE(\bar{x}) \leq \mu \leq \bar{x} + t_{15}(0.01) SE(\bar{x})$

#Q. Average height of 10 student is a school is observed as 67 inches with sum of the squares of deviations from central value is 88. can we say that average height of student in a school is 65 inches. It is given that, value of t at 5% level of significance with 9 degree of freedom is 2.262.

$n=10, \bar{x}=67, \sum(x-\bar{x})^2 = 88$

$H_0 : \boxed{\mu=65}, H_1 : \mu \neq 65$



Now, $S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{88}{10-1} = \frac{88}{9} = 9.77$

$$\boxed{t = \frac{\bar{x}-\mu_0}{\sqrt{S^2/n}}} = \frac{67-65}{\sqrt{\frac{9.77}{10}}} = 2.04$$

∵ t lies in Confidence Region so H_0 is Accepted.

i.e Av Height of Student $(\mu) = 65$

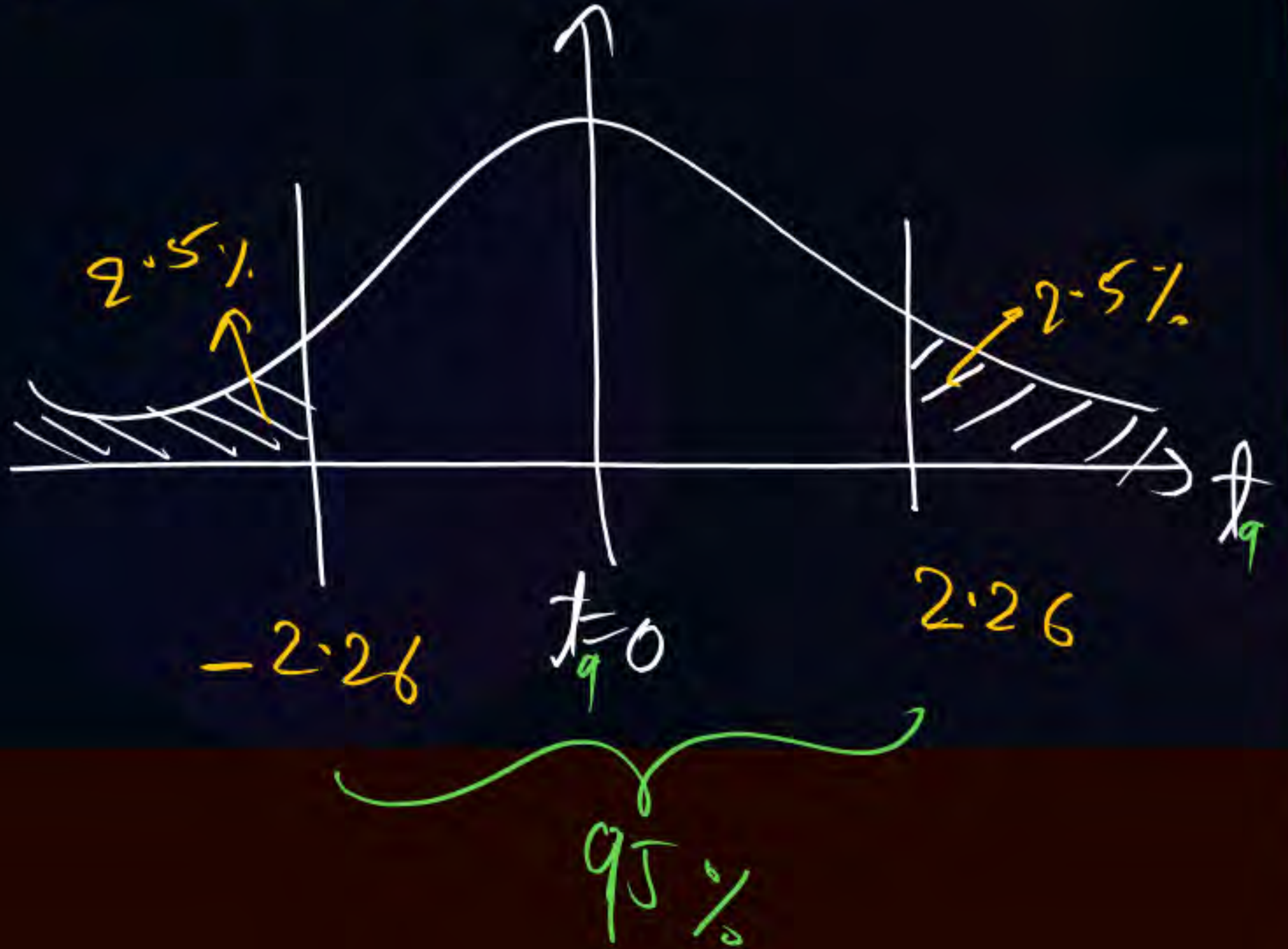
[NAT]

#Q. A machine produces washers of thickness 10mm. A sample of 10 washers has an average thickness of 9.52 mm with SD of 0.6 mm. Find out ‘t’.

Already Solved YESTERDAY

#Q. If the mean and variance of sample of 10 observation are 66 and 14.67 respectively then find the 95% confidence limits for population mean (μ). Given that upper 2.5% point of t distribution with 9 degree of freedom is 2.26

Sol: $n = 10$, $\bar{x} = 66$, $s^2 = 14.67$



$$SE(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{\sqrt{14.67}}{\sqrt{10}} = 1.21$$

$$\bar{x} - t_9(0.05)\,SE(\bar{x}) \le \mu \le \bar{x} + t_9(0.05)\,SE(\bar{x})$$

$$66 - 2.26(1.21) \le \mu \le 66 + 2.26(1.21)$$

$$63.26 \le \mu \le 68.73$$

#Q. The means of two random samples of size 9 and 7 are 196.42 and 198.82 respective variance are 26.94 and 18.73. can we say that the sample are drawn from same normal population?

Type II

$t_{14}(0.05) = 2.145$

$n_1 = 9, \quad n_2 = 7$ & $Df = (n_1+n_2)-2 = 14$

$\bar{x} = 196.42, \quad \bar{y} = 198.82$

$S_1^2 = 26.94, \quad S_2^2 = 18.73$

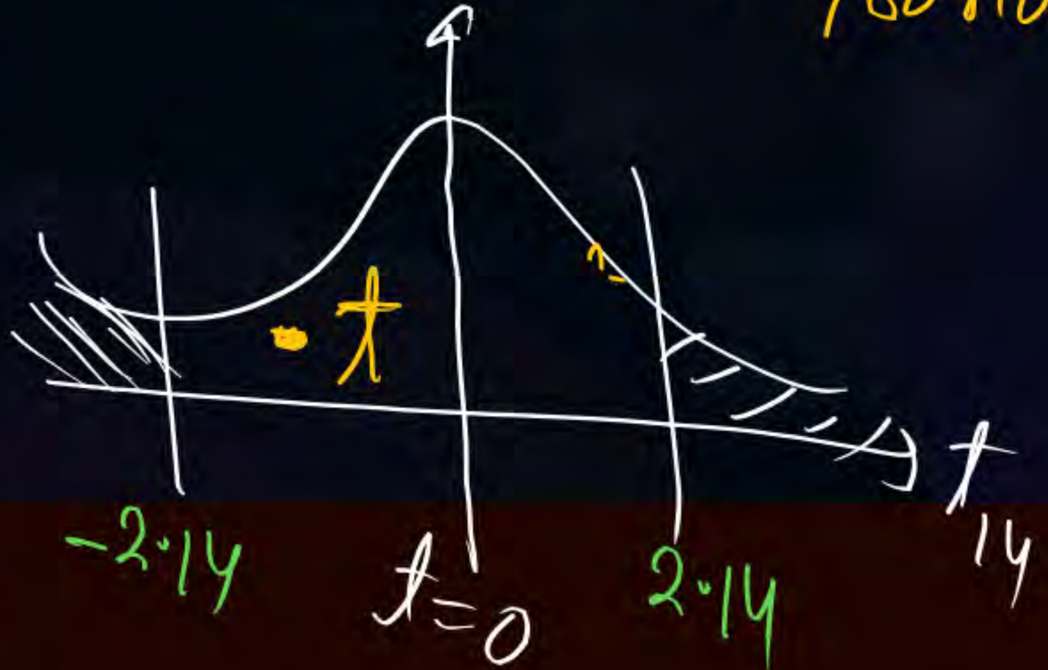
$H_0 : \boxed{\mu_1 = \mu_2}, \quad H_1 : \mu_1 \neq \mu_2$

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)} = 23.42$$

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{S^2}{n_1}+\frac{S^2}{n_2}}} = \frac{196.42-198.82}{\sqrt{\frac{23.42}{9}+\frac{23.42}{7}}}$$

$$= -0.984$$

So H_0 is Accepted



CHI-SQUARE TEST (Large sample test)

Let $x_1, x_2, x_3 \dots x_n$ are Ind Normal R. Variables with mean μ & Var $= \sigma^2$

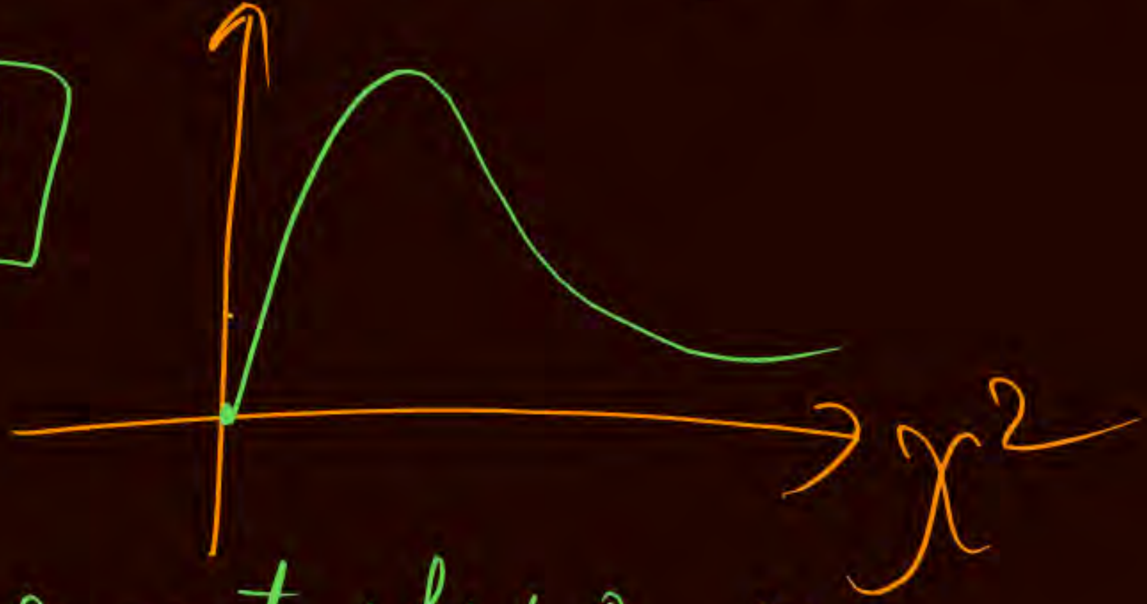
then $Z_i = \boxed{\frac{x_i - \mu}{\sigma}}$ is called Standard Normal Variable with mean 0 & Var = 1

then the variables $Z_1^2, Z_2^2, Z_3^2, \dots, Z_n^2$ are called Chi Square Variables

& the Distribution $Z^2 = Z_1^2 + Z_2^2 + Z_3^2 + \dots + Z_n^2$ is called Chi Sq. Distribution

ie $\chi^2 = \sum Z_i^2$ (ie χ^2 test is Right tailed)

or $\boxed{\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2}$, $0 \le \chi^2 < \infty$



it has only <u>one</u> parameter which is <u>n</u>

⊛ x has two parameters μ & σ^2 i.e $x \sim N\{\mu, \sigma^2\}$

Z has No parameter i.e $Z \sim N\{0, 1\}$

& χ^2 has only one parameter which is n

⊛ if n values are given then $\boxed{d.f\ (\nu) = n-1}$

⊛ if Data is given in terms of contingency table then $\boxed{d.f\ (\nu) = (R-1)(C-1)}$

Application of χ^2 Test → ① To find the significance of population Variance →

H_0 : $\boxed{\sigma^2 = \sigma_0^2}$

H_1 : $\sigma^2 \neq \sigma_0^2$

$$\chi^2 = \sum\left(\frac{x-\bar{x}}{\sigma}\right)^2 = \frac{\sum(x-\bar{x})^2}{\sigma^2} = \boxed{\frac{ns^2}{\sigma^2}}$$

Note ① Population Value ≃ Theoretical Value = Expected Value (E_i) = Approx Value

② Sample Value ≃ Experimental Value = Observed Value (O_i) = Exact Value

Type ② Significance of Goodness of Fit

if we want to check that whether there exist any significant difference b/w O_i & E_i then we will use this test.

H_0 : there is no difference b/w O_i & E_i i.e $\boxed{O_i = E_i}$

H_1 : there is a significant diff b/w O_i & E_i i.e $O_i \neq E_i$

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

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

Type 1

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

$n = 25$, $s^2 = 81$, $\sigma^2 = (10.5)^2$

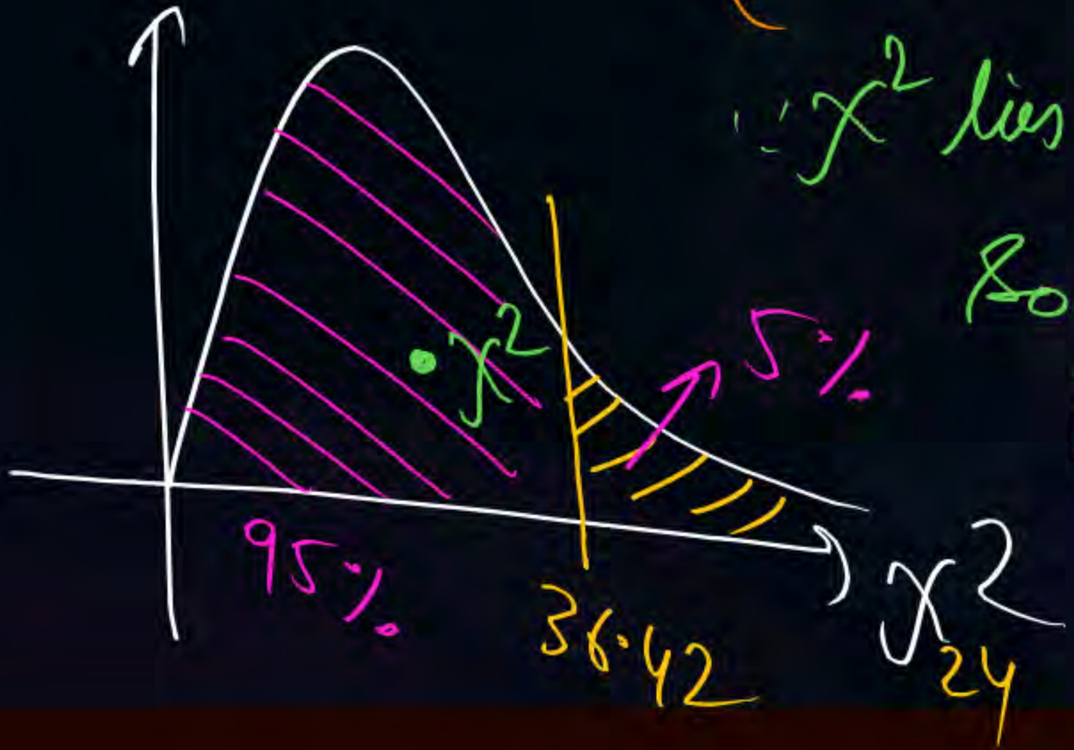
$H_0 : \sigma^2 = (10.5)^2$

$H_1 : \sigma^2 \neq (10.5)^2$

$Df(\nu) = 25 - 1 = 24$

Hence Chi-Square statistic is

$$\chi^2 = \frac{ns^2}{\sigma^2} = \frac{25 \times 81}{(10.5)^2} = 18.367$$



$\because \chi^2$ lies in Confidence Region

So H_0 is Accepted

ie

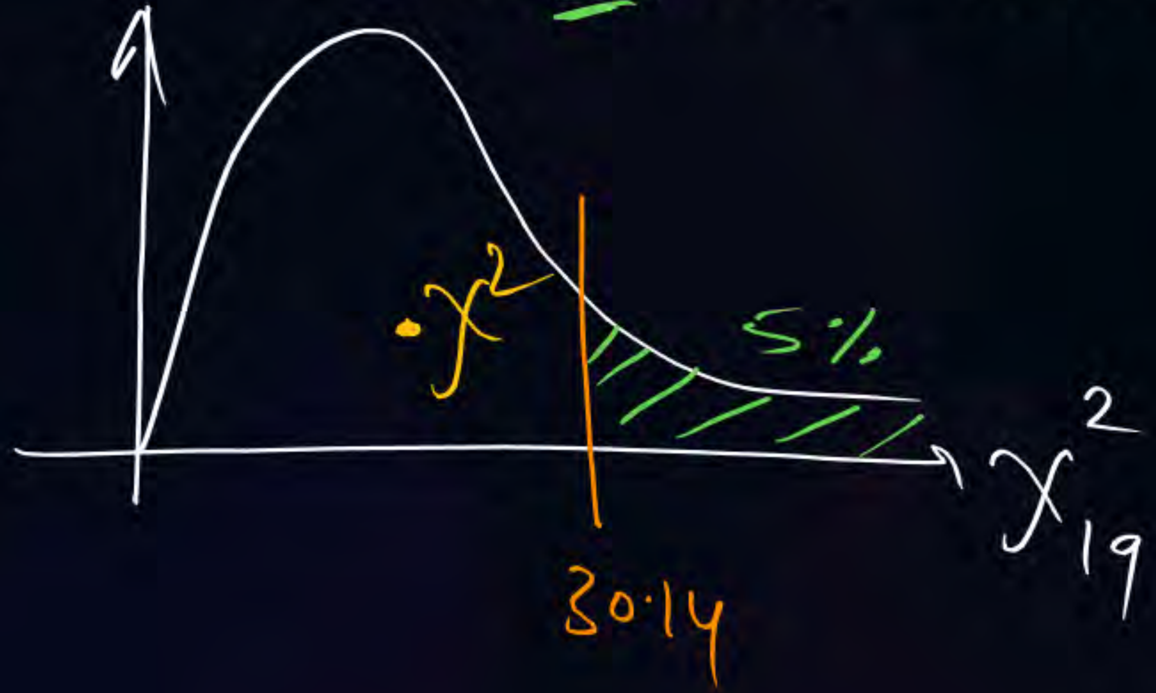
$\sigma = 10.5$

#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 $x_{19}^{2}(0.05) = 30.14$

Sol:



$n=20, \; s^2=25, \; \sigma^2=64$

(i) $\chi^2 = \boxed{\frac{ns^2}{\sigma^2}} = \frac{20\times 25}{64} = 7.8125$

(ii) $H_0 : \boxed{\sigma=8}, \; H_1 : \sigma \neq 8$

Hence H_0 is Accepted

Type 2

#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.

Given $x_6^2(0.05) = 12.592$

$O_i =$

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

$\because \Sigma O_i = \Sigma E_i = \boxed{98 = N}$

$H_0 : \boxed{O_i = E_i} \Rightarrow$ ie Accidents are U-Distributed

$H_1 : O_i \neq E_i \Rightarrow$ " " not U- "

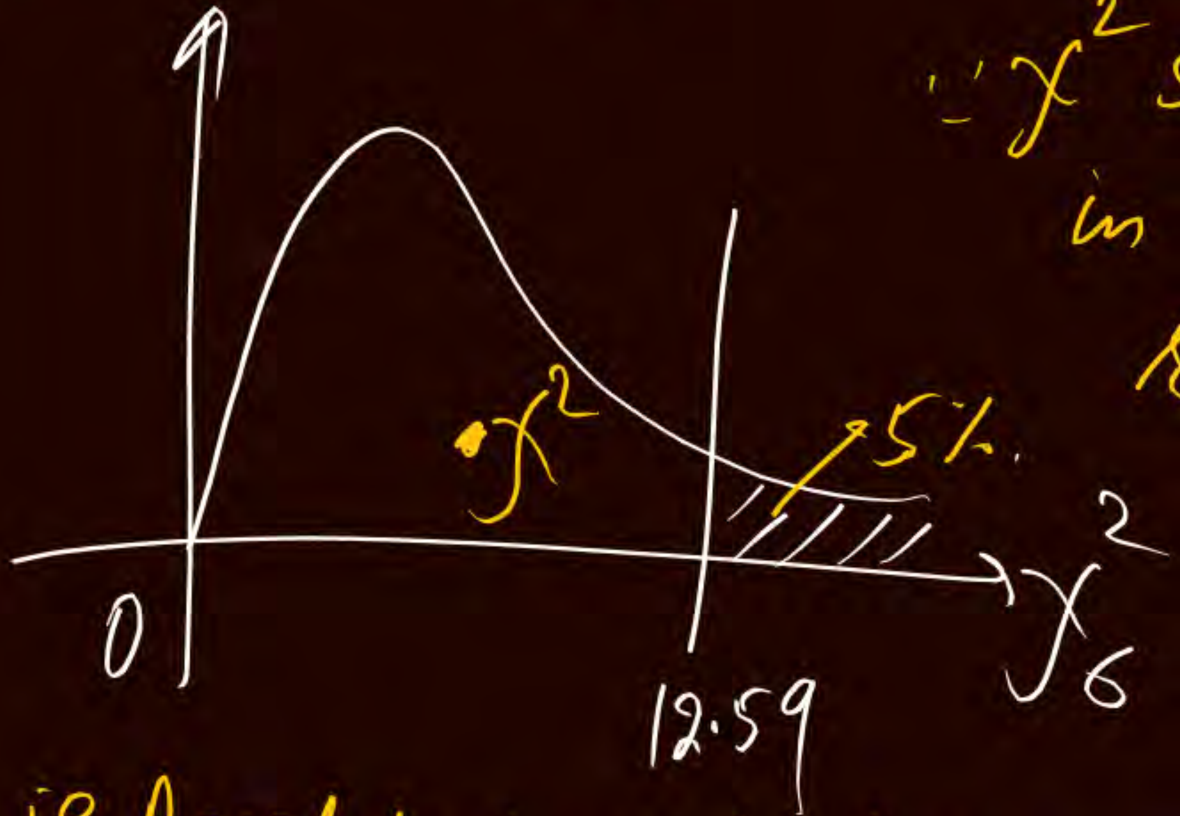
$E_i = \frac{\text{Total No of Accidents}}{\text{No. of Days}} = \frac{98}{7} = 14$

$\nu = n-1 = 7-1 = 6$

	O_i	E_i	(O_i-E_i)	$(O_i-E_i)^2$
M	14	14	0	0
T	18	14	4	16
W	12	14	-2	4
Th	11	14	-3	9
F	15	14	1	1
S	14	14	0	0
S	14	14	0	0
				$\Sigma = 30$

Now $\sum(O_i-E_i)^2 = 30$

$$\chi^2 = \frac{\sum(O_i-E_i)^2}{E_i} = \frac{30}{14} = 2.14$$



∵ χ^2 statistic falls in Confidence Region so H_0 is Accepted

i.e. Accidents are uniformly Dist over the week.

#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 $x_5^2(0.05) = 11.09$

$H_0 : O_i = E_i \Rightarrow$ there is no Diff b/w O_i & E_i or Die is Unbiased

$H_1 : O_i \neq E_i \Rightarrow$ there is significant diff b/w O_i & E_i or Die is Biased

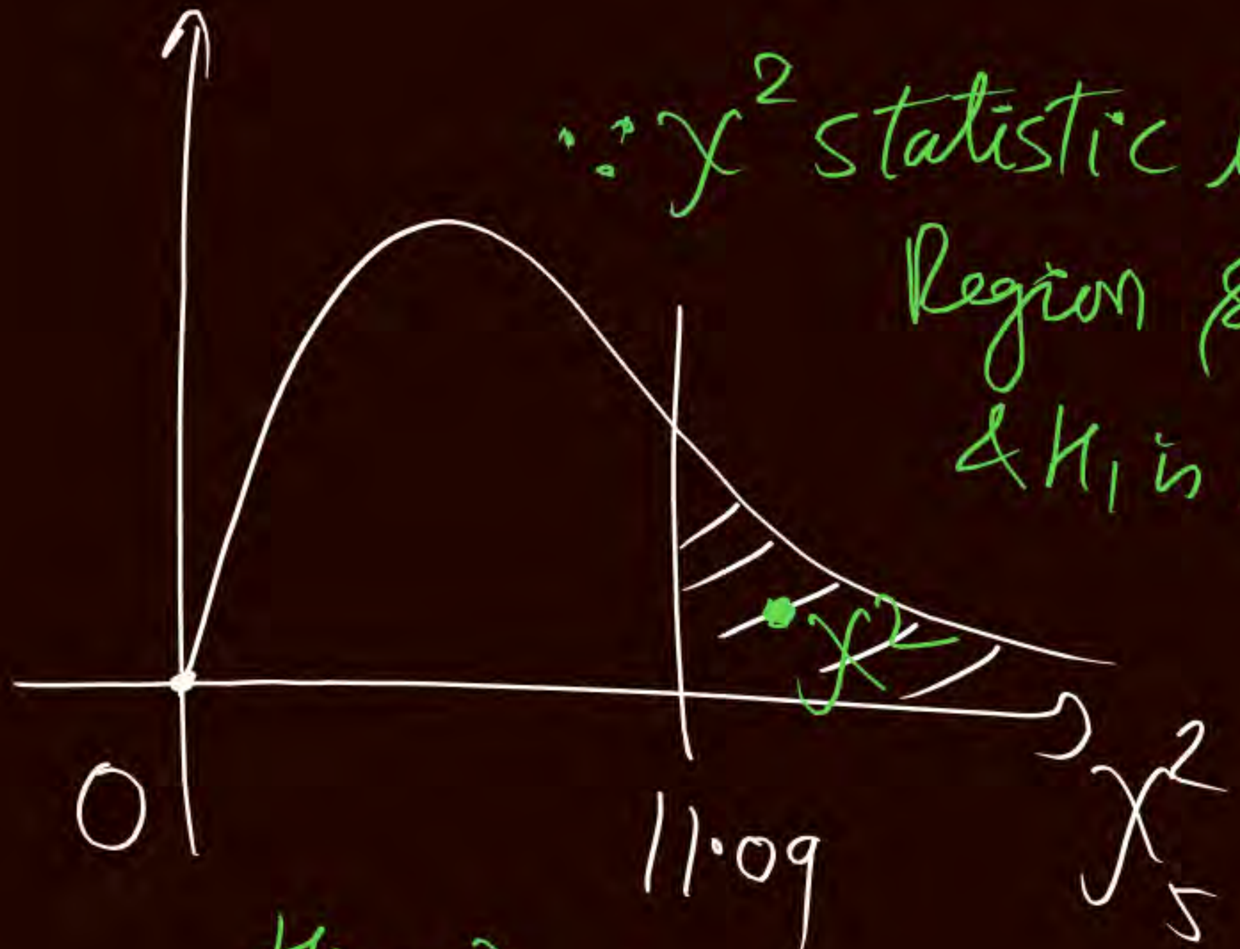
$df(\nu) = n-1 = 5$, $\boxed{\sum O_i = \sum E_i}$ & $N = 276$

$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$

$\Rightarrow E(1) = E(2) = E(3) = E(4) = E(5) = E(6) = \frac{276}{6} = 46$

	O_i	E_i	$(O_i-E_i)^2$	$(O_i-E_i)^2/E_i$
1	40	46	36	—
2	32	46	196	—
3	29	46	289	—
4	59	46	169	—
5	57	46	121	—
6	59	46	169	—
			$\Sigma = 980$	

$$\chi^2 = \frac{\sum (O_i - E_i)^2}{E_i} = \frac{980}{46} = 21.30$$



$\because \chi^2$ statistic lies Critical Region so H_0 is Rejected & H_1 is Accepted

Hence Die is BIASED.

[NAT]

#Q. A car manufacture company produces four different color in which 882 are while, 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 $x_3^2(0.05) = 7.815$, Enemies the correspondence between and experiment does the experimental result support the theory?

HW

H_0 = Accepted //

THANK - YOU