MODULE 2 (PART 2)

χ^2 (chi-square) distribution: χ^2, t test, F-test are used for small samples.

$n < 30 \rightarrow$ small, $n > 30 \rightarrow$ large

Properties:

* χ^2 Distribution curve is not symmetrical. It lies entirely in 1st Quadrant & is not a Normal Distribution.
* χ^2 varies from 0 to ∞. It depends only on degrees of freedom (df).
* Mean $= n$
* Variance $= 2n$

Applications:

* To test the goodness of fit
* To test the independence of attributes.

χ^2 - Goodness of fit:

$$\chi^2 = \Sigma \left[\frac{(O_i - E_i)^2}{E_i}\right] \sim \chi^2_{n-1}$$

where $O_i \rightarrow$ observed frequency

$E_i \rightarrow$ Expected "

Expected frequency:

$$E_i = \frac{\text{Total of obs freq}}{\text{No. of observations}}$$

Frequency given in question & sample is small → Chi Square.

How to read Chi-square table?

* The rows represent degrees of freedom.
* degrees of freedom is given by: no. of observations -1.

  Example: df for 10 obs $= 10 - 1 = 9$ df.
* Column represents Level of significance.
* Value of χ^2 for 9 df at 5% significance:

→ 5% LOS

df \ α	0.100	0.05	0.025	0.010	...
1					
2					
3					
4					
5					
6					
7					
8					
df → 9		16.9			

⇒ χ^2 for 9 df at 5% LOS = 16.9

χ^2 for independence of Attributes:

H_0: The attributes are independent

H_1: The " " not independent.

(or)

H_0: The attributes are associated

H_1: " " " not associated.

$$\boxed{\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right] \sim \chi^2_{(m-1)(n-1)}}$$

↓

Use the formula when sample size is small and the frequency of two attributes is recorded. (Keyword → independent).
↳ association

Problems:

100 workers (gender & Nature of Work).

$$E_{ij} = \frac{\text{Corresponding Row Total} \times \text{Corresponding Column total}}{\text{Grand total}}$$

	Stable	Unstable	Total
Male	40	20	60 → (40+20)
Female	10	30	40 → (10+30).
Total	50 (40+10)	50 (20+30)	100 [(60+40) or (50+50)]

O_{ij}	E_{ij}	$O_{ij}-E_{ij}$	$(O_{ij}-E_{ij})^2$	$\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$
40	30	10	100	3·33
20	30	−10	100	3·33
10	20	−10	100	5
30	20	10	100	5
				$x^2 = 16·66$

$$E_{ij}(40) = \frac{1^{st}\text{ row total} \times 1^{st}\text{ Column total}}{\text{Grand total}}$$

$$= \frac{60 \times 50}{100} = 30$$

$$E_{ij}(20) = \frac{1^{st}\text{ row total} \times 2^{nd}\text{ Column total}}{\text{Grand total}}$$

$$= \frac{60 \times 50}{100} = 30$$

$$E_{ij}(10) = \frac{2^{nd}\text{ row total} \times 1^{st}\text{ Column total}}{\text{Grand total}}$$

$$= \frac{40 \times 50}{100} = 20.$$

$$E_{ij}(30) = \frac{2^{nd}\text{ row total} \times 2^{nd}\text{ Column total}}{\text{Grand total}}$$

$$= \frac{40 \times 50}{100} = 20$$

$df \rightarrow (m-1)(n-1)$

$= (2-1)(2-1) = 1$ df for 5% LOS

$= 3.84.$

Cal Val > Tab Val

Reject H_0.

Student's t distribution:

Properties:

* Mean = 0
* Variance = $\frac{n}{n-2}$, $n > 2$
* Mgf does not exist

Applications:

* To test significance of mean of a small random sample.
* " " " " difference of means of 2 samples.

t-test → df = 10

	1%	5%
2-tailed	3.169	2.228
left-tailed	-2.764	-1.812
Right-tailed	2.764	1.812

T-test for Single Mean (sample is small)

①
$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$\mu \rightarrow$ population mean

$\bar{x} \rightarrow$ Sample "

$$S = \sqrt{\frac{1}{n-1}\Sigma(x_i - \bar{x})^2}$$

②
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$s = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$$

~~Note~~ B/w ① & ②, remember any one of the 2 formulae.

Confidence Interval

$$\bar{x} \pm t_{\alpha/2}\, S/\sqrt{n}$$

(or)

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n-1}}$$

T-test for difference of Means (sample is small)

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\bar{x} \rightarrow$ I Sample Mean

$\bar{y} \rightarrow$ II " "

$n_1 \rightarrow$ I " size

$n_2 \rightarrow$ II " "

$$S = \sqrt{\frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{m}(x-\bar{x})^2 + \sum_{j=1}^{n}(y-\bar{y})^2\right]}$$

$n_1 + n_2 - 2 \rightarrow df.$

Paired t-test: (Dependency b/w Attributes). (Reactions, before & after)

$$t = \frac{\bar{d}}{s/\sqrt{n}} \sim t_{n-1}$$

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$d_i = x_i - y_i$$

$$s = \sqrt{\frac{1}{n-1}\sum(d_i - \bar{d})^2}$$

$$= \sqrt{\frac{1}{n-1}\left[\sum d_i^2 - \frac{(\sum d_i)^2}{n}\right]}$$

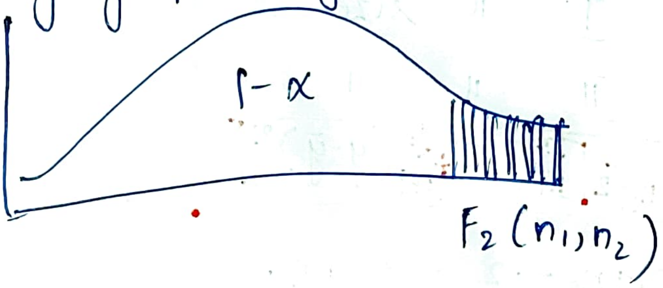
F- Distribution:

$$\boxed{F = \frac{X/\nu_1}{Y/\nu_2}}$$

Def: F is defined as the ratio of two independent χ^2 variables divided by corresponding degrees of freedom.

Properties:

1. Mean $\rightarrow \dfrac{\nu_2}{\nu_2 - 2}$, $\nu_2 > 2$
2. highly positively skewed graph $\rightarrow$ F dist



$\boxed{F \rightarrow \text{test}}$

population variance of small samples.

$$F = \frac{S_1^2}{S_2^2}$$

$$S_1^2 = \frac{1}{n_1 - 1} \Sigma (x_i - \bar{x})^2$$

$$S_2^2 = \frac{1}{n_2 - 1} \Sigma (y_i - \bar{y})^2$$

Chi Square	T-test	F-test
Used when sample is small. Frequencies, independence, association are words to look for in problems	Used when sample is small. Significance of Means or difference of means. **Paired t-test**: Used for checking dependency. (Reactions, b4 & After).	Used for small samples. Variance problems.