Q INTERVIEW DAY –

Chi-square test (α 0.05)

- A **chi-square** (χ^2) is a statistic that check for patterns or **relationship** in a **categorical** variable.

- For **example**, Is there any relationship between Gender & investing in FD.

FIXED DEPOSIT

	INVESTED	NOT INVESTED
Female	42	12
Male	38	6

For 95% **confidence interval**,
if $p < 0.05$, indicates **significant relationship** between groups.
$p \geq 0.5$, indicates **insignificant relationship** between groups.

Q) why not just look at the number count?

Ans – We can't be sure if it is reliable or created by chance. If we do again can we get the same exact numbers.
So we use **inferential statistics** and chi-square.

Q) Give another example of Chi-square test?

Ans – Does gender vary across educational measures.

	ENGINEERING	BUSINESS	PSYCHOLOGY
Female	20	20	30
Male	30	20	20

$$\chi^2 = \frac{(\text{Expected} - \text{Observed})^2}{\text{Expected}}$$

- If there is a **relationship**, we would expect gender to be **unevenly spread** across majors.
- If there is **no relationship**, we expect gender to be **evenly spread** across majors.
- Tables which we create (categorical vs categorical) known as **contingency table**.

UNEVEN SPREAD

	ENGINEERING	BUSINESS	PSYCHOLOGY
Female	20	40	20
Male	30	60	10

EVEN SPREAD

	ENGINEERING	BUSINESS	PSYCHOLOGY
Female	20	20	20
Male	20	20	20

Q) How to know Chi-square result will hold next time if new data comes?

Ans – So each time chi-square will give **p-values**, the **p value tells us the likelihood is real difference.**
- If p value = 0.05, 5% chance we would get the result with random state. Accepted range is normally, p value ≤ 0.05 or else reject.

Q) Chi-square test limitations and assumptions?

Ans – i) Each group should have **minimum count of 5**. ii) Does **not need normal distribution.**
iii) It doesn't tell which level of **variable** are **driving the effect**.
iv) All **data** should be **independent**. One variable should not affect other variable

Q) What is Interaction effect?

Ans - Happens when one independent variable interact with other independent variable affects the dependent variable.

Example,

	Drinking Diet Juice	
	Yes	No
Eating pills Yes	5 gram	2 gram
No	2 gram	1 gram

- Highest weight loss is happening when we take both pill and drink together.
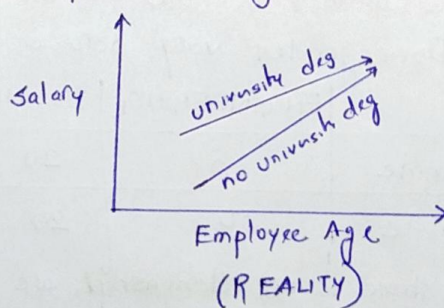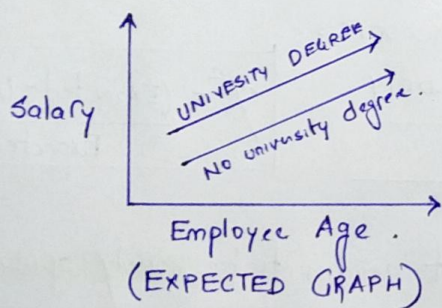- We can find this effect through Factor Analysis.
- Factor analysis extract maximum common variance from all variable and put them into common score.

Q) Give another example of Interaction effect.

Ans - Suppose we build a model to explain salary of an IT company employees

Dependent Var → Salary.   Independent Variable → i) University degree (Y/N)

ii) Employee Age.

Suppose equation is $Salary = \beta_0 + \beta_1 (Employee\ Age) + \beta_2 (University\ degree) + Error.$

Here β_2 is define how much salary difference if we have degree with no degree.

(EXPECTED GRAPH)   (REALITY)

Normally, university degree earn more initially compare to no university degree. As the experience increase, university degree doesn't matter, its about experience.

So create another variable = (Employee Age) * (University degree).

This additional term is called Interaction term. Requires when x_1 affects the relationship x_2 and y.

Common misconception → An interaction term is required when x_1 and x_2 are correlated.