population distribution

normal distribution → parametric test → such as: T, Z tests

Assumptions: (two levels)

1) independent "dichotomous" or dependent "continuous"
2) same variances" homogeneity"
3) random sampling.
4) no Overlap!!!
في dependent t test بال overlap.

-more Poweful

non normal or unknown distribution → non parametric → such as: chi square

1) no assumptions about data distribution. "distribution-free"
2) can be with small sample.
3) ONLY: nominal and ordinal data.
4) good with outliers.
5) no overlap

Level of Measurement	Sample Characteristics					Correlation
	1 Sample	2 Sample		K Sample (i.e., >2)		
		Independent	Dependent	Independent	Dependent	
Categorical or Nominal	X^2	X^2	Macnamar's X^2	X^2	Cochran's Q	
Rank or Ordinal		Mann Whitney U	Wilcoxin Matched Pairs Signed Ranks	Kruskal Wallis H	Friendman's ANOVA	Spearman's rho
Parametric (Interval & Ratio)	z test or t test	t test between groups	t test within groups	1 way ANOVA between groups	1 way ANOVA (within or repeated measure)	Pearson's r
		Factorial (2 way) ANOVA				

Test of normality

we assume that our data is non-normally distributed "H_1" ← H_0 ⇒ normally distributed

⇒ α > P-value ⇒ H_0 is rejected, H_1 is accepted.

* Chi square :-

1) nominal data, sample >20.
2) establish the level of significance "α"
3) formulate the "statical" hypothesis = H_A or H_1 alternative = statical hypothesis and null hypothesis H_0
4) calculate the statical: "χ^2_{calc}"
5) determine the degree of freedom.
6) compare the static against critical "χ^2_{cv}"
7) not Prove causality.
8) Actual count not percentage data.

The chi square test can only be used on data that has the following characteristics:

The data must be in the form of frequencies

The frequency data must have a precise numerical value and must be organised into categories or groups.

The expected frequency in any one cell of the table must be greater than 5.

The total number of observations must be greater than 20.

Degree	Frequency	Relative Frequency	Percentage
High School	2	0.050	5.0
Bachelor's	7	0.175	17.5
MBA	20	0.500	50.0
Master's	3	0.075	7.5
Law	4	0.100	10.0
PhD	4	0.100	10.0
	40		

H_1 ⇒ dependent ⇒ مثلا أنا بفرض إنه الرجل من مالك بصه المقاطعة إله علاقة بقلة الأخلاق، فهاي هي الـ H_1 ← أنا افترضت إنهم مرتبطين ببعض.

* الـ H_0 رح تكون عكس H_1، يعني رح يكون بالـ H_0 إنه قلة الأخلاق غير مرتبطة بالرجل من مالك بصه المقاطعة.

- إذا كانت الـ P-value > α ← معناه إنه H_1 خطأ ← H_0 صحيحة ← بقبل H_0
- إذا كانت الـ P-value < α ← معناه إنه H_1 صحيحة ← برفض H_0

"تمثيل ال data"   degree of freedom ↑

* بدنا α ، H_0 ، r ، F_e ، F_o ، χ^2 ، p.value ⇐ عشان نقرر

F_e ⇒ expected frequency $= \frac{row \times column}{total} = \frac{(a+b)(a+c)}{a+b+c+d}$

a	b	a+b
c	d	
a+c		a+b+c+d

χ^2 ⇒ chi square $= \sum \frac{(F_o - F_e)^2}{F_e}$

$r = df = (R-1)(C-1)$

* $\chi^2_{calc} > \chi^2$ ⇒ reject H_0 ⇒ يعني على مثالنا السابق إنه الي بوطل من مالك بعد المقاطعة هو شخص ماعنده أخلاق.

T test:

"William Gosset test"

- T test is used for comparing mean value of two groups.

* بال H_0 بنفرض انه ال data لل 2 groups انها كانت نفس ال distribution فا ن الهم نفس ال means ⇐ كل ما زاد ال t test كل ما زاد الفرق بين ال means كل ما بنرفض ال H_0 اكثر

independent t test

- صرنا بنقارن بين 2 groups

- أرباح منتج لشركة غير داعمة للإحتلال مقابل أرباح شركة داعمة.

dependent t test

- نفس ال group بس 2 variables.

- مثلا أرباح مالك قبل وبعد المقاطعة

$H_0 \Rightarrow M_1 - M_2 = 0$

1) independent t test:

$H_0 \Rightarrow M_1 - M_2 = 0$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{SD_1^2}{n_1} - \frac{SD_2^2}{n_2}}}$$

$df = n_1 + n_2 - 2$

* lavene's test of equals of varience

* non sig → reject H_0

2) dependent t test:

$$t_d = \frac{\bar{D}}{\sqrt{\frac{SD_d^2}{n}}}$$

$df = n - 1$

$\alpha > P$ reject H_0