بسم الله الرحمن الرحيم

السلام عليكم ورحمة الله وبركاته

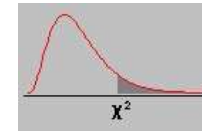
L X1II

Chi Square (χ^2) test

PART 3

PROF. DR. WAQAR AL-KUBAISY

AUGUST 2-2023



Application of χ2

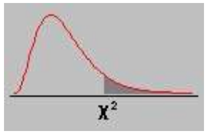
2×2 **table .**

$r \times c$ **table .**

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Application of χ2

- **2 × 2 table**
- **r × c table.**



Application of χ2

I- 2 × 2 table

The application of χ2 is to test the significance association between outcome and certain factor that we are interested in .

Here we have

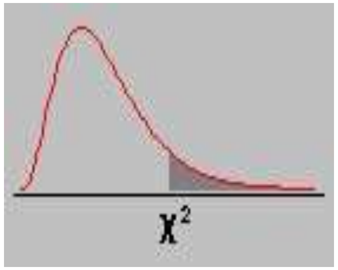
two groups with

two outcomes for each group .

In this case we use what we call it 2 × 2 table .

In this case we are going to compare between two proportion of two groups of population .

II- r × c



- other application of χ^2 is r× c. , ax b

We have two or more than two groups and or with two or more A contingency table also used **than two outcome**

These large table we call it r x c, ax b

r denotes the numbers of **rows** in the table and

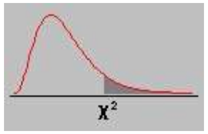
c the numbers of **columns** .

- **more than two rows** and or **more than two columns .**
- In another word **more than four cells,**

we could have 6, 8, 10, ….

- Here we have **more than two rows or two columns** .
- We have **two or more groups**
- with **more than two outcome**
- In another word we have more than four cells, we could have 6, 8, 10,

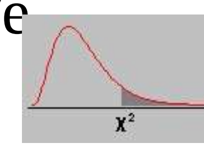
Example



Sample of 273 tuberculosis cases were collected . given three types of treatment either by **PAS alone** **or** **streptomycin alone** **or** **combination of PAS** Para-aminosalicylic acid (**PAS**) **and streptomycin .**

The outcome of treatment was **categorized depen**ding **on the result of sputum exam** either **positive smear positive culture**, **negative smear positive culture** or **negative smear negative culture** .
99 given PAS alone**, 65** of them showed smear +ve &,culture +ve, while only **13 cure**. Of the group (90 patients) who were treated by combination of streptomycin & PAS, **35** **were shows** **negative smear and negative** culture, while **18 of** combined R patients demonstrated **negative smear & positive** culture .For those treated by streptomycin, 46 smear +ve &,culture +ve ,and **18** demonstrated **negative smear & positive culture**

99 given PAS alone, 65 of them showed smear+ve &,culture +Ve while only 13 cure. Of the group (90 patients) who were treated by combination of streptomycin & PAS 35 were shows negative smear and negative culture while 18 of combined R patients demonstrated negative smear & positive culture .for those treated by streptomycine46 smear +ve &,culture +ve and 18 demonstrated negative smear & positive culture



Type R	+S +C	-S +C	-S -C	Total
PAS	65	21	13	99
Stre.	46	18	20	84
Com.	37	18	35	90
Total	148	57	68	273

Type R	+S +C	-S +C	-S -C	Total
PAS	65	21	13	99
Stre.	46	18	20	84
Com.	37	18	35	90
Total	148	57	68	273

cure rate

PAS $\frac{13}{99} \times 100 = 13\%$

Streptomycin $\frac{20}{84}$ X 100 =23.81%

Combine $\frac{35}{90} \times 100 = 39\%$

Failure rat

PAS $\frac{65}{99}$ X 100 =65,7%

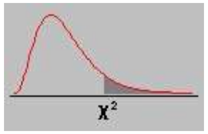
Streptomycin $\frac{46}{84}$ X 100 =54.8%

Combine $\frac{37}{90} \times 100 = 41\%$

Total cure rate $\frac{68}{273} \times 100 = 25\%$

Data

Qualitative data, No. Of T.B patients, treated by 3 different regime (PAS alone, Streptomycin alone or combine both) . Outcome of treatment categorized into 3 group (Failure, not cure and cure) .



Assumption

Independent random sample chosen from normal distribution population

Formulation of Hypothesis

Ho

There is no significance difference in cure rate among the three different treated group .

P1¬ = P2 = P3 = P0 .

The difference observed is due to chance factor, sampling error and sampling variability .

There is no significance difference in cure rate among the three different treated group .

P1¬ = P2 = P3 = P0 .

The difference observed is due to chance factor, sampling error and sampling variability .

There is no significance association between cure rate level and type of treatment .

HA

There is a significance difference in cure rate between three group .

This difference due to effect of different treatment . There is no or minimum effect of chance factor .

P1¬ ≠ P2 ≠ P3 ≠ P0

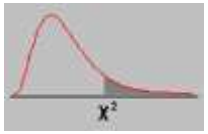
Critical region

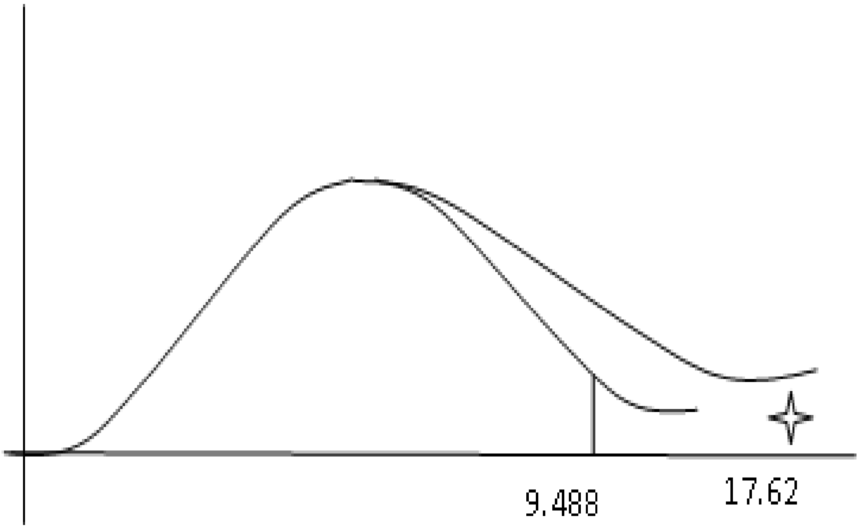
d.F = (C – 1) (r – 1)

= (3 – 1) (3 – 1) = 4

$\alpha = 0.05$

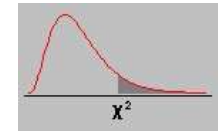
tabulated $\chi 2$ = 9.488





$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

E expected (E) $= \dfrac{\text{total column X total row}}{\text{Grand total}}$



E18= $\dfrac{90 \text{ X}57}{273}$ = 18.8

E35= $\dfrac{90 \text{ X}68}{273}$ = 22.42

E65 = $\dfrac{99 \text{ X}148}{273}$ =53.67

E21 == $\dfrac{99 \text{ X}57}{273}$ =20.67

E13= = $\dfrac{99 \text{ X}68}{273}$ = 24.66

E46= $\dfrac{84 \text{ X}148}{273}$ =45.54

E18= $\dfrac{84 \text{ X}57}{273}$ = 17.54

E20= $\dfrac{84 \text{ X}68}{273}$ =20.9

E37= $\dfrac{90 \text{ X}148}{273}$ = 48.8

Type R	+S +C	-S +C	-S -C	Total
PAS	65	21	13	99
Stre.	46	18	20	84
Com.	37	18	35	90
Total	148	57	68	273

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\frac{(65-53.67)^2}{53.67} + \frac{(21-20.67)^2}{20.67} + \frac{(13-24.66)^2}{24.66} + \frac{(46-45.54)^2}{45.54} \; \frac{(18-17.54)^2}{17.54} +$$

$$\frac{(20-20.9)^2}{20.9} + \frac{(37-48.8)^2}{48.8} + \frac{(18-18.8)^2}{18.8} + \frac{(35-22.42)^2}{22.42}$$

=2.4+0.005+5.513+0.005+0.012+0.047+2.85+0.034+7.067=17.978

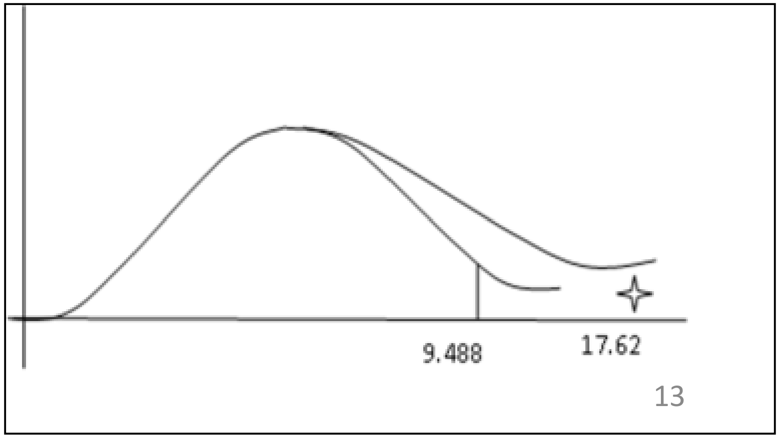
Calculated χ^2 greater than **tabulated χ^2 .**

Calculated χ^2 fall in area of **rejection**,

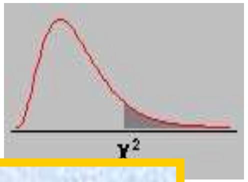
so we **reject Ho** = we **reject no significance difference** .

There is significance difference in cure rate between the three groups .

P < 0.05 .



II- r × c



$$X^2 = \sum \frac{(O - E)^2}{E} \qquad d.f. = (c - 1)(r - 1)$$

The expected No. must be computed for each cell

$$E = \frac{\text{Column total} \times \text{row total}}{\text{Overall total}}$$

There is no continuity correction

- ✓Chi square is only valid if applied to the
- ✓actual numbers in the various categories .
- ✓It must never be applied to table showing just proportions or percentages

Validity of Chi Square

χ^2 is valid

- when the overall total is more than 40 , regardless the expected values and
- when the overall total between 20 and 40 provided that all expected values are at least 5
- Chi square is valid provided that
- less than 20% of expected numbers are less than 5
- And none is less than 1

- When the expected numbers are very small the chi We recommended other test (Exact Test)
- Chi square test is not valid when we have cell zero

This restriction can be overcome by
combining rows or columns with the low expected numbers provide that these combination make biological sense

Fisher's exact test of independence

Sir Ronald Aylmer Fisher

Fisher's exact test

is a statistical significance test used in the analysis of contingency tables where sample sizes are small.

It is named after its inventor, R. A. Fisher **Sir Ronald Aylmer Fisher**

The test is useful for categorical data that result from **classifying** objects in **two different ways**;

it is used to examine the **significance of** the **association** (contingency) between the **two kinds of classification**

Most uses of the Fisher test involve, like this example, a 2 × 2 contingency table.

With large samples, a chi-squared test can be used in this situation

When to use it

Fisher's exact test is used when you have

two nominal variables.

A data set in rows and columns.

Fisher's exact test is more accurate than the chi-squared test of independence **when the expected numbers are small**.

The most common use of **Fisher's** exact test

is for 2×2 tables,

You can do Fisher's exact test for greater than two rows and columns.

- when sample **sizes are small**, or
- the data are very **unequally distributed** among the cells of the table,
- resulting in the cell counts predicted on the null hypothesis
- (the "expected values") being low.

The usual rule for deciding whether the chi-squared approximation is good enough is that,
the chi-squared test is not suitable when;
the **expected values in any of the cells of a contingency table** are **below 5**, or **below 10** when there is only **one degree of freedom** .
Fisher test can therefore be used regardless of the sample characteristics.
It becomes difficult to calculate with large samples or well-balanced tables, but fortunately these are exactly the conditions where the chi-squared test is appropriate.

When some of the expected values are small,
Fisher's exact test is more accurate than the chi-squared of independence.

- **If all of the expected values are very large,**
Fisher's exact test becomes **computationally impractical**;

- fortunately, **the chi-squared will then give an accurate result**
- **If you have a 2x2 frequency table with small numbers of expected frequencies**
- **(in case the total number of observations is less than 20), you should not perform the *Chi-square test* but you should use *Fisher's exact test*.**

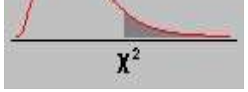
$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

Thank You

	Total	succeeded	%	Not succeeded
Baghdad	220	180	82%	40
UiTM	200	170	85%	30
Syria	320	200	62.5%	120
Mutah	380	220	57.9%	160
	1120	770		350

770/1120 = 0.687

770/1120 X 100 = 68.7%



Data

Qualitative data consist of sample of medical students divided into four groups,.

Variation in the Successful rate was detected

Assumption

.

Formulation of Hypothesis

Ho

HA

	Succeeded		Not succeeded		Total
	O	E	O	E	
Baghdad	180	151.25	40	68.75	220
UiTM	170	137.5	30	62.5	200
Syria	200	220	120	100	320
Mutah	220	261.25	160	118.75	380
Total	770		350		1120