What exactly is a chi-square test doing?

What is a Chi-Square test?

A Chi-Square test is used to determine association between two categorical variables. This test is used to examine whether two categorical variables are independent of each other. This test is also known as the ***Chi-Square test of independence***

Assumptions

The following assumptions should be met before applying Chi-Square test:

1. Data must be categorical. It cannot be used to analyze differences in scores or their means
2. Expected frequencies in each cell should not be less than 5
3. No subjects can be count more than once
4. Categories should be defined prior to data collection and analysis

Application of the Chi-Square test

The following data was collected in a clinical trial evaluating a new compound designed to improve wound healing in trauma patients. The new compound is compared against a placebo. After treatment for 5 days with the new compound or placebo, the extent of wound healing is measured and the data are shown below. The number in each cell is a count.

	Percent Wound Healing			
Treatment	**0-25%**	**26-50%**	**51-75%**	**76-100%**
New Compound (n=125)	15	37	32	41
Placebo (n=125)	36	45	34	10

Data source: Essentials of Biostatistics for Public Health, 2nd Edition, Lisa M. Sullivan, PhD

Is there a difference in the extent of wound healing by treatment? (Hint: Are treatment and the percent wound healing independent?) Run the appropriate test at a 5% level of significance.

This scenario is a good fit for the Chi-Square test because both the 'treatment' and 'percent wound healing' are categorical variables. We are interested in evaluating whether these two variables are associated.

So, let's follow these steps to apply chi-square test to solve the above problem.

Step 1. **Set up hypotheses and determine level of significance.**

Null hypothesis (H_0): Treatment and percent wound healing are independent (not associated)
Alternative hypothesis (H_1): H_0 is false.
Alpha=0.05

Step 2. **Select the appropriate test statistic formula for Chi-Square test.**

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

Where,
O=Observed count
E=Expected count

The condition for appropriate use of the above test statistic is that each expected frequency is at least 5. In Step 4 we will compute the expected frequencies and we will ensure that the condition is met.

Step 3. **Set up your decision rule.**

Degrees of freedom (df) = (number of rows – 1) (number of columns – 1)

df=(2-1)(4-1)=3 and the decision rule is Reject H_0 if $\chi^2 \geq$ **7.81**.

Step 4. **Compute the test statistic.**

We now compute the expected frequencies using the formula,

Expected Frequency = (Row Total * Column Total)/N.

The top number in each cell of the table is the observed frequency and the bottom number is the expected frequency, shown in parentheses.

	Percent Wound Healing				**Total**
Treatment	**0-25%**	**26-50%**	**51-75%**	**76-100%**	
New Compound	15 (25.5)	37 (41)	32 (33)	41 (25.5)	125
Placebo	36 (25.5)	45 (41)	34 (33)	10 (25.5)	125
Total	51	82	66	51	250

The test statistic is computed as follows:

$$\chi^2 = \frac{(15-25.5)^2}{25.5} + \frac{(37-41)^2}{41} + \frac{(32-33)^2}{33} + \frac{(41-25.5)^2}{25.5}$$
$$+ \frac{(36-25.5)^2}{25.5} + \frac{(45-41)^2}{41} + \frac{(34-33)^2}{33} + \frac{(10-25.5)^2}{25.5}$$

$\chi^2 =$ 4.32 + 0.39 + 0.03 + 9.42 + 4.32 + 0.39 + 0.03 + 9.42 = **28.32**

Step 5. **Conclusion**

Reject the null hypothesis (H_0) because **28.32 ≥ 7.81**. We have statistically significant evidence at 5% level of significance (α=0.05) to show that the null hypothesis (H_0) is false; treatment and percent wound healing are not independent. The p-value is $p < 0.001$.

How to report results from a Chi-Square test:

A chi-square test was performed to determine whether the treatment (new compound, placebo) was related to percent wound healing in trauma patients. The extent of wound healing rates are significantly different between treatment groups; χ^2 (3) = 28.32, p<.001.