Chi-Square as a Statistical Test

- ***Chi-square test:*** an **inferential statistics** technique designed to test for **significant relationships** between two variables organized in a bivariate table.
- **Chi-square requires no assumptions** about the shape of the population distribution from which a sample is drawn.
- A statistical method used to determine **goodness of fit**
  - Goodness of fit refers to how close the observed data are to those predicted from a hypothesis
- Note:
  - **The chi square test does not prove that a hypothesis is correct**
  - It evaluates to what extent the data and the hypothesis have a good fit.

Limitations of the Chi-Square Test:

- **The chi-square test** does **<u>not</u>** give us much information about the ***strength*** of the relationship or its ***substantive significance*** in the population.
- **The chi-square test** is **sensitive** to ***sample size***. The size of the calculated chi-square is **directly proportional** to the size of the sample, independent of the strength of the relationship between the variables.
- **The chi-square test** is also **sensitive** to **small expected frequencies** in one or more of the cells in the table.

Statistical Independence:

Independence (statistical): the **absence of association** between two cross-tabulated variables. The percentage distributions of the dependent variable within each category of the independent variable are **identical**.

Hypothesis Testing with Chi-Square:
<u>Chi-square follows five steps</u>:

1. Making assumptions (**random sampling**)
2. Stating the research and null hypotheses
3. Selecting the sampling distribution and specifying the test statistic
4. Computing the test statistic
5. Making a decision and interpreting the results

The Assumptions:

- The chi-square test requires **no assumptions** about the **shape of the population distribution** from which the sample was drawn.
- However, like all inferential techniques it assumes **random sampling**.

H_1: The two variables are **related** in the population. Gender and fear of walking alone at night are ***statistically dependent***.

Afraid	Men	Women	Total
No	83.3%	57.2%	**71.1%**
Yes	16.7%	42.8%	**28.9%**
Total	**100%**	**100%**	**100%**

H_0: There is **no association** between the two variables. Gender and fear of walking alone at night are ***statistically independent***.

The Concept of Expected Frequencies:

Expected frequencies f_e : the cell frequencies that would be **expected** in a bivariate table **if** the two tables were **statistically independent**.

Observed frequencies f_o: the cell frequencies **<u>actually observed</u>** in a bivariate table.

Calculating Expected Frequencies:

f_e = <u>(column marginal)(row marginal) / N</u>

To obtain the expected frequencies for any cell in any cross-tabulation in which the two variables are assumed independent, multiply the row and column totals for that cell and divide the product by the total number of cases in the table.

Chi-Square (obtained):

The test statistic that summarizes the differences between the observed (*fo*) and the expected (*fe*) frequencies in a bivariate table.

Calculating the Obtained Chi-Square:

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

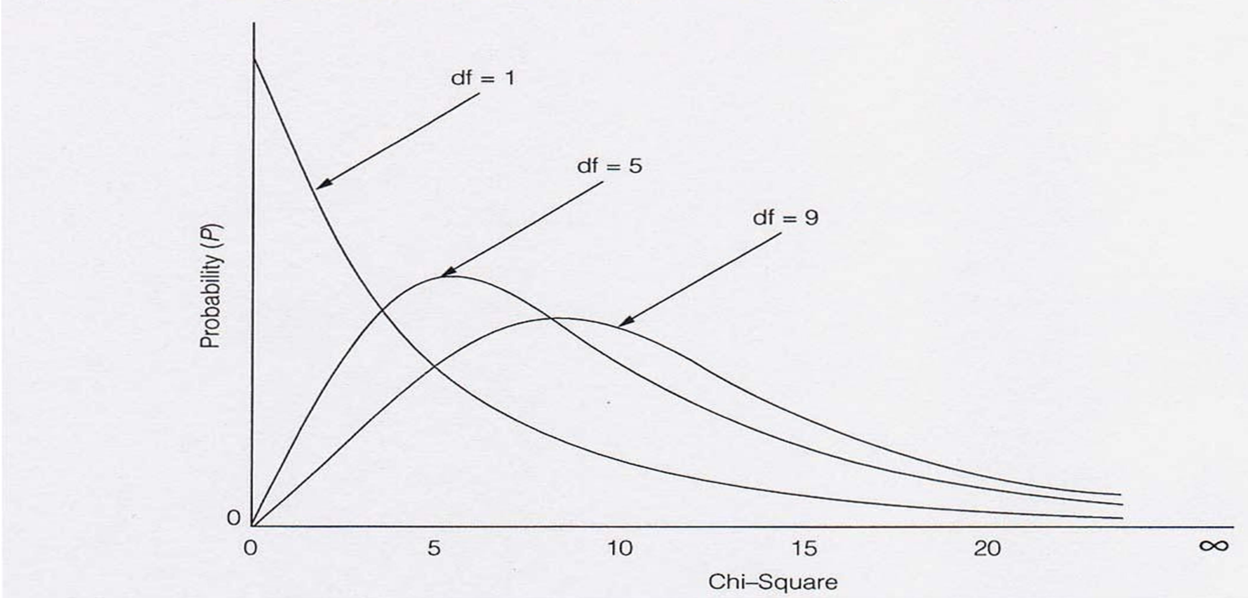
f_e = expected frequencies

f_o = observed frequencies

The Sampling Distribution of Chi-Square:

- The distributions are **positively skewed**. The research hypothesis for the chi-square is **always** a **one-tailed test**.
  - Chi-square values are **always positive**. The minimum possible value is zero, with **no upper limit** to its maximum value.
  - As the number of degrees of freedom increases, the **c^2** distribution becomes **more symmetrical**.

Figure 14.1 **Chi-Square Distributions for 1, 5, and 9 Degrees of Freedom**



df = (r – 1)(c – 1)

where r = the number of rows ; c = the number of columns

(3 – 1)(2 – 1) = **2 degrees of freedom**