Chi-Square (χ^2) Goodness of Fit Test

The Chi-Square Goodness of Fit is used to determine whether a frequency distribution (observations) follows a specific distribution (expected values).

H_0: Follows the specific distribution
H_1: Does **not** follow the specific distribution

The Chi-Square test is a right tail test. The Chi-Square distribution is right skewed. It is not symmetric. The shape of the Chi-Square distribution is dependent on the degrees of freedom (k-1, where k is the number of categories).

The χ^2 Distribution
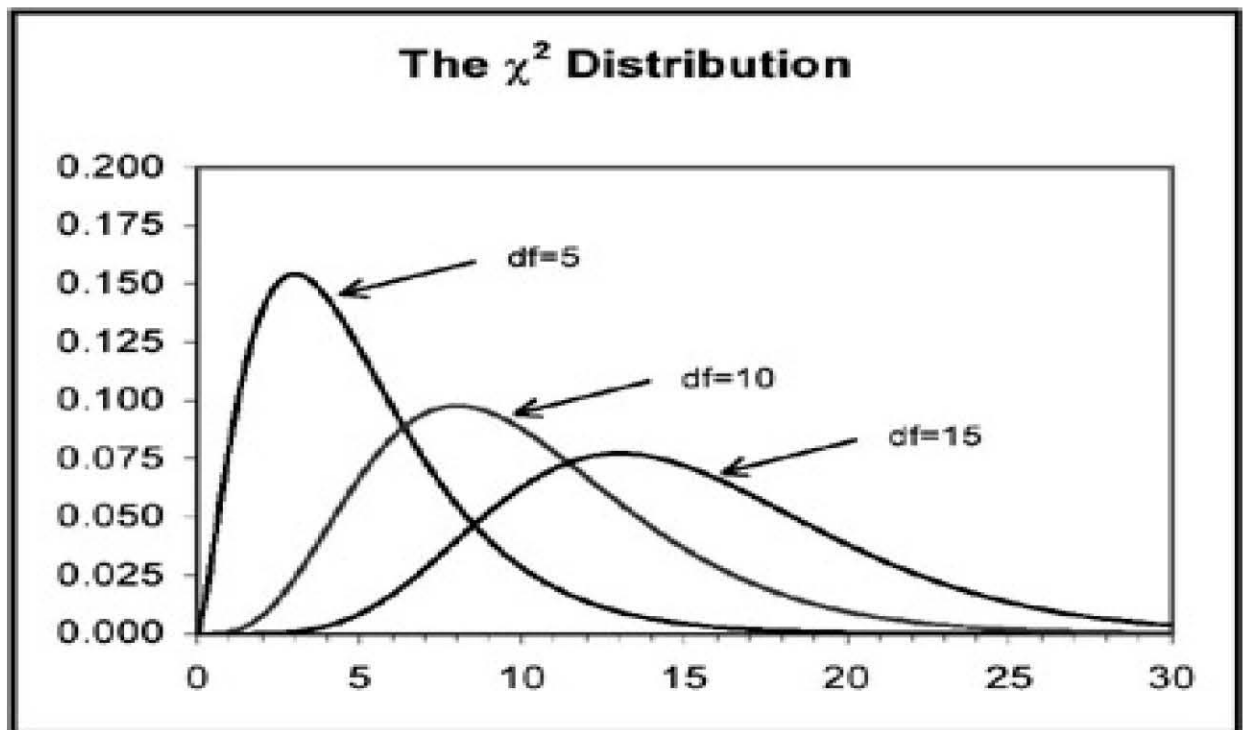
1. <u>Set up the hypothesis test</u>

   H_0: Follows Distribution   (assumed truth)

   H_1: Does not follow Distribution   (suspicion)

2. <u>Choose α</u>

   Once the level of significance has been chosen, the Chi-Square critical value must be looked up in the table for the Area to the Right of Critical Value (TABLE VII for Sullivan Fundamentals of statistics).

3. <u>Calculate the test statistic:</u>

   Expected counts are calculated as: $E_i = np_i$

   Where n is the total number of observations and p_i is the expected probability.

$$\chi^2_{test} = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$\text{for } i = 1,2,\dots,k$$

   and degrees of freedom $df = k - 1$

   - *all expected frequencies* ≥ 1
   - no more than 20% of frequencies < 5

4. Use the Chi-Square Test Statistic χ^2_{test} and

   the degrees of freedom to estimate the P-value using TABLE VII.

5. <u>Make your decision:</u>

   Reject H_0 if

P-value $< \alpha$	OR	$\chi^2_{test} > \chi^2_{critical}$ (ie., the test value lies within the critical area of the graph)

6. <u>Interpret your decision</u>

   There is sufficient (H_0 $rejected$)/ insufficient (fail to reject H_0) evidence that the frequency distribution does not follow the specified distribution.

Example:

Is a six sided die a fair die?

We roll the die 60 times and observe the following frequency distribution.

Roll	Observed (O)	Expected (E)	Difference (O − E)	$(O - E)^2$	$\dfrac{(O - E)^2}{E}$
1	3	10	-7	49	$\dfrac{49}{10}$
2	7	10	-3	9	$\dfrac{9}{10}$
3	9	10	-1	1	$\dfrac{1}{10}$
4	11	10	1	1	$\dfrac{1}{10}$
5	13	10	3	9	$\dfrac{9}{10}$
6	17	10	7	49	$\dfrac{49}{10}$

1. H_0: The outcomes from rolling the die follow the uniform distribution

   (or $p = \dfrac{1}{6}$ or "the die is fair")

   H_1: The die does not follow the uniform distribution

   (or $p \neq \dfrac{1}{6}$ or "the die is unfair")

2. Decide on $\alpha = 0.05$ level of significance.

3. Calculate the test statistic.

$$\chi^2_{test} = \frac{49}{10} + \frac{9}{10} + \frac{1}{10} + \frac{1}{10} + \frac{9}{10} + \frac{49}{10}$$

$$\chi^2_{test} = 11.8$$

4. From TABLE VII, the $\chi^2_{critical}$ value for an area of 0.05 with 5 degrees of freedom is 11.070.

5. Since $\chi^2_{test} = 11.8 > 11.070 = \chi^2_{critical}$

   Therefore, $\chi^2_{test} > \chi^2_{critical}$, so we reject H_0

6. There is sufficient evidence at the $\alpha = 0.05$ level of confidence to conclude that the die is not fair.

The TI-84 calculator can do this:
Hit the **STAT** key
Choose **1:Edit**
Enter the Observed values into a list and the Expected values into another list.
Hit the **STAT** key
Arrow right or left to **TESTS**
Choose χ^2 **GOF-TEST**
Enter List Name for the Observed Values
Enter List Name for the Expected Values
Then enter the degrees of freedom (Number of categories subtract 1)
Choose **Calculate**
Make your decision based on the P-value

Use the TI-83 calculator to do it this way:
Hit the **STAT** key
Choose **1:Edit**
Enter the Expected values into **L1**
Enter the observed values into **L2**
Move the cursor up to highlight **L3**

Enter (2nd **L1** − 2nd **L2**) $\boxed{x^2}$ ÷ 2nd **L2**
Hit the **STAT** key and arrow to **CALC**
Choose **1: 1-Var Stats** 2nd **L3** then the **ENTER** key
The value for ∑x is the Chi-Square test statistic
Compare this value to the Chi-Square critical value from TABLE VII to make your decision.