# How to Perform a Chi-Square Goodness of Fit Test in Google Sheets

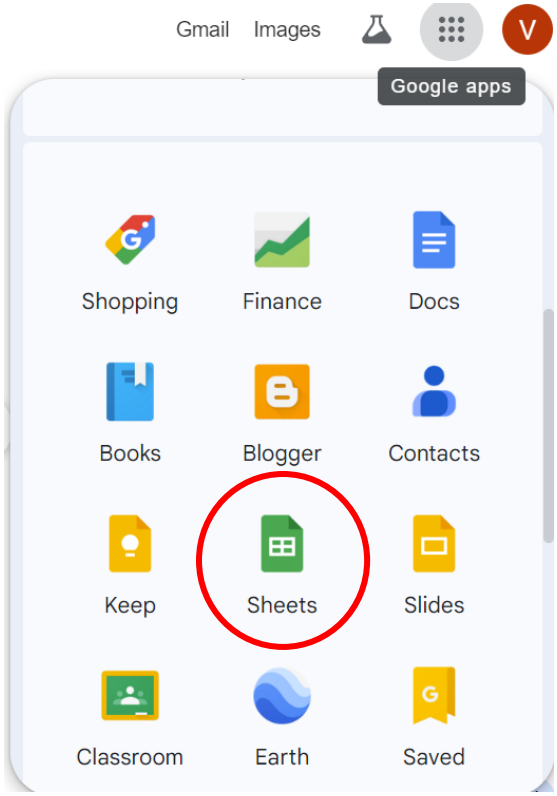
The Chi-Square Goodness of Fit test is used to determine if there is a significant difference between the observed frequencies and the expected frequencies in one or more categories. This handout will guide you through the steps to perform this test using Google Sheets.

Step-by-Step Instructions:

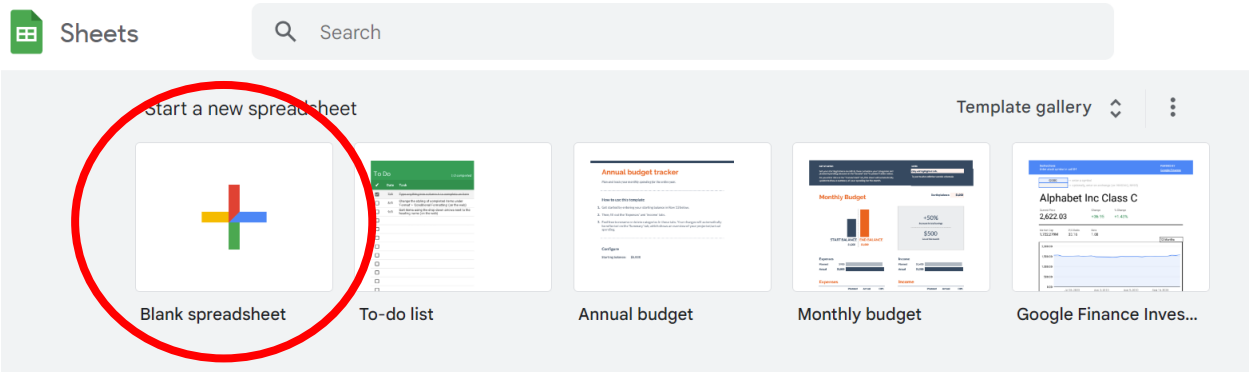
Steps to Open Google Sheets

1. Sign In:
   - Ensure you are signed into your Google account.

2. Open Google Sheets:
   - Open the Google Web Browser.
   - Click on the 'Google Apps' dropdown menu in the upper right corner.
   - Scroll down and click on the 'Sheets' app icon.



3. Create a New Spreadsheet:
    - Click on the “+” icon to create a new spreadsheet.
    - Alternatively, you can access Google Sheets through your Google Drive by clicking on "New" and then selecting "Google Sheets."




Steps to Prepare Your Data

1. Enter Your Data:

- In column A, list all the names you are analyzing.
- In column B, enter the observed frequencies (the number of times each name was selected) for each corresponding name.
- In column C, enter the expected frequencies for each name. To calculate the expected frequency, divide the total number of observations by the total number of names.
    - In the example below, the total observed count is 40 and there are 4 names, the expected frequency for each name is 40 divided by 4, which equals 10.

Example:

| | A | B | C |
|---|---|---|---|
| 1 | **Names** | **Observed** | **Expected** |
| 2 | Name 1 | 5 | 10 |
| 3 | Name 2 | 17 | 10 |
| 4 | Name 3 | 8 | 10 |
| 5 | Name 4 | 10 | 10 |

Step 3: Calculate the Chi-Square Statistic

1. Calculate the Differences: In column D, calculate the difference between observed and expected frequencies for each category.

- In cell D2, enter: `=B2 - C2`

- Drag the fill handle down to apply the formula to other cells in column D.

2. Square the Differences: In column E, square the differences calculated in column D.

- In cell E2, enter: `=D2^2`

- Drag the fill handle down to apply the formula to other cells in column E.

3. Divide by Expected Frequency: In column F, divide each squared difference by the expected frequency for that category.

- In cell F2, enter: `=E2 / C2`

- Drag the fill handle down to apply the formula to other cells in column F.

4. Sum the Results: Sum the values in column F to get the Chi-Square statistic.

- In cell F7, enter: `=SUM(F2:F5)`

Example:

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **Names** | **Observed** | **Expected** | **Difference** | **Squared** | **Chi-Square Component** |
| 2 | Name 1 | 5 | 10 | -5 | 25 | 2.5 |
| 3 | Name 2 | 17 | 10 | 7 | 49 | 4.9 |
| 4 | Name 3 | 8 | 10 | -2 | 4 | 0.4 |
| 5 | Name 4 | 10 | 10 | 0 | 0 | 0 |
| 6 | | | | | | |
| 7 | | | | | **Chi-Square =** | 7.8 |

Step 4: Determine the Degrees of Freedom

1. Calculate Degrees of Freedom (df):

- The formula for degrees of freedom is `df = (number of categories - 1)`.
- If you have 4 categories, df = 4 - 1 = 3.

Step 5: Interpret the Chi-Square Statistic

1. Choose a Critical Value:

- Select a desired significance level (e.g., 0.05).

2. Evaluate the Result:

- In cell F8, enter the following formula to find the p-value: `=CHISQ.DIST.RT(F7,3)`.
- This formula calculates the right-tailed probability of the Chi-Square statistic (F7) with 3 degrees of freedom.

3. Interpret the Chi-Square Statistic at the Chosen Critical Value:

- Compare the p-value in F8 with your chosen significance level.
- If the p-value is greater than the significance level (e.g., 0.05), fail to reject the null hypothesis (i.e., there is no significant difference between observed and expected frequencies).
- If the p-value is less than or equal to the significance level, reject the null hypothesis (i.e., there is a significant difference between observed and expected frequencies).

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **Names** | **Observed** | **Expected** | **Difference** | **Squared** | **Chi-Square Component** |
| 2 | Name 1 | 5 | 10 | -5 | 25 | 2.5 |
| 3 | Name 2 | 17 | 10 | 7 | 49 | 4.9 |
| 4 | Name 3 | 8 | 10 | -2 | 4 | 0.4 |
| 5 | Name 4 | 10 | 10 | 0 | 0 | 0 |
| 6 | | | | | | |
| 7 | | | | | **Chi-Square =** | 7.8 |
| 8 | | | | | **p-value =** | 0.05033109786 |

Example Conclusion

Based on the Chi-Square Goodness of Fit test, we obtained a p-value of 0.05033109786. Since this p-value is slightly greater than our chosen significance level of 0.05, we fail to reject the null hypothesis. This indicates that there is no statistically significant difference between the observed frequencies of the name selections and the expected frequencies under a uniform distribution. Therefore, we conclude that the observed distribution of name selections could reasonably be due to random selection.

**Tip:**

- Use the COUNTIF function in Google Sheets to help tally observed frequencies if you are working with raw data.