# $\chi^2$–Test

# &

# Contingency Table

## What is $\chi^2$–**Test**?

---

A $\chi^2$–**Test** is a statistical method used to analyze categorical data to determine if observed frequencies are significantly different from expected frequencies.

## What is the purpose of $\chi^2$–**Test** ?

---

It helps determining if differences between observed and expected data are due to chance or a significant relationship.

---

**What are the types of $\chi^2$–Test?**

There are two main types:

1. **Chi-Square Test of Independence**:

   Determines if categorical variables are dependent or independent.

2. **Chi-Square Goodness-Of-Fit Test**:

   Determines if a sample distribution fits a known distribution.

## How do we perform the **Test of Independence**?

Here are the six steps to perform this test:

1. Set up $H_0$ and $H_1$.

   $H_0$ : The row and column variables are independent.

   $H_1$ : The row and column variables are dependent. RTT

2. Find the **Expected Frequency Table** for each cell of the **Contingency Table** by using the **Augmented Contingency Table** along with the formula below.

$$E = \frac{(\textbf{row total}) \cdot (\textbf{column total})}{(\textbf{grand total})}$$

## How do we perform the **Test of Independence**?

3. Find the computed test statistic using the formula below.

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

where O is the observed value and E is the expected frequency for each corresponding cell.

4. Find the $P$ –value using the degrees of freedom below.

$$df = (r - 1) \cdot (c - 1)$$

where r is the number of rows and c is the number of columns of the contingency table.

## How do we perform the **Test of Independence**?

5. Determine the validity of $H_0$ and $H_1$.

   If $P$ –Value $> \alpha \implies H_0$ valid & $H_1$ invalid

   If $P$ –Value $\leq \alpha \implies H_0$ invalid & $H_1$ valid

6. Draw the final conclusion.

   If $H_0$ valid $\implies$ row and column variables are independent.

   If $H_1$ valid $\implies$ row and column variables are dependent.

# What is a **Contingency Table**?

A **Contingency Table** organizes data into rows and columns, with each cell showing the count that fit both criteria simultaneously.

Suppose I surveyed 100 students about whether they regularly use AI tools (like ChatGPT) for studying and whether they passed or failed a statistics course.

The **Contingency Table** would look like this:

| | Passed | Failed |
|---|---|---|
| Use AI tools | 20 | 10 |
| No AI tools | 40 | 30 |

This **Contingency Table** has 2 rows and 2 columns.

## What does a **Contingency Table** consist of?

---

**Contingency Table** has Rows and Columns. Each row and column define different categorical variables.

The intersection of a row and column contains the frequency of observations for that specific combination of categories.

---

Mathematical representation of a table is called matrix. It is a rectangular array of numbers, symbols, or expressions arranged in rows and columns.

The size of a matrix is defined by the number of rows and columns it contains, expressed as $m \times n$, where $m$ is the number of rows and $n$ is the number of columns.

---

## What is an **Augmented Contingency Table**?

An **Augmented Contingency Table** contains row total and column total. The intersection cell of the row total and column total is called grand total.

Using our earlier example, here is how the **Augmented Contingency Table** would look like:

| | Passed | Failed | Total |
|---|---|---|---|
| Use AI tools | 20 | 10 | 30 |
| No AI tools | 40 | 30 | 70 |
| Total | 60 | 40 | 100 |

## How do we make the **Expected Frequency Table**?

Each cell of an **Expected Frequency Table** is calculated by the formula below.

$$E = \frac{(\textbf{row total}) \cdot (\textbf{column total})}{(\textbf{grand total})}$$

Using our earlier example, here is how the **Expected Frequency Table** would look like:

| | Passed | Failed |
|---|---|---|
| Use AI tools | $\frac{30 \cdot 60}{100} = \boxed{18}$ | $\frac{30 \cdot 40}{100} = \boxed{12}$ |
| No AI tools | $\frac{70 \cdot 60}{100} = \boxed{42}$ | $\frac{70 \cdot 40}{100} = \boxed{28}$ |

## How do we find the **Computed Test Statistic**?

We can find the computed test statistic $\chi^2$ using the formula below.

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

where O is the observed value and E is the expected frequency for each corresponding cell.

---

From our earlier example, Let's have our **Observed** and **Expected** matrices next to each other, that makes it easier for computation.

$$O = \begin{bmatrix} 20 & 10 \\ 40 & 30 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 18 & 12 \\ 42 & 28 \end{bmatrix}$$

---

## How do we find the **Computed Test Statistic**?

Now we can find $\chi^2$ using the formula.

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

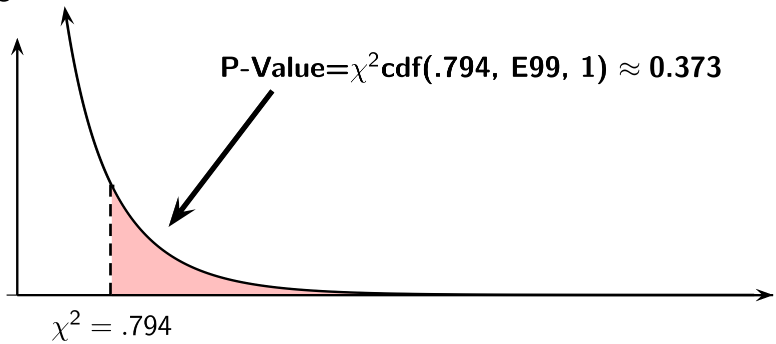
$$= \frac{(20-18)^2}{18} + \frac{(10-12)^2}{12} + \frac{(40-42)^2}{42} + \frac{(30-28)^2}{28}$$

$$\approx .794$$

## How do we find the corresponding *P*–**Value**?

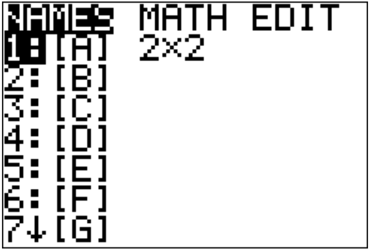
$\chi^2$–Test is always Right-Tail-Test.

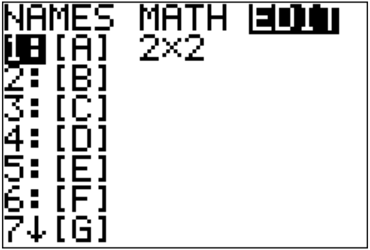
From our example, we draw the Chi–Square distribution curve with $df = (r-1)\cdot(c-1) = (2-1)\cdot(2-1) = 1$, clearly label the computed test statistic $\chi^2 = .794$, and shade the right tail for the right tail test.



# $\chi^2$–Test & TI

First we store the **Contingency Table** in a matrix. Press **2ND** followed by $\boldsymbol{x^{-1}}$ for the **MATRIX** menu, then use the → to go to the **EDIT** menu option.



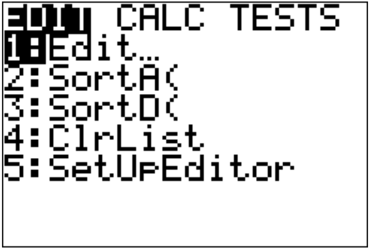



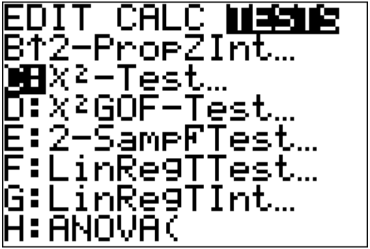
## $\chi^2$–**Test** & **TI**

By choosing option 1, we can enter the matrix size, and the values of the contingency table.

# $\chi^2$–Test & TI

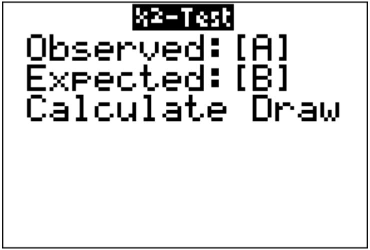
Now press **STAT** followed by $\rightarrow$ to get to the **TESTS** menu, then use the $\downarrow$ to go to **$\chi^2$–Test** option, and press **ENTER** to execute.

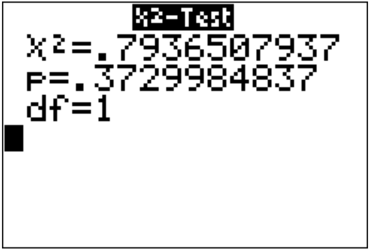


## $\chi^2$–**Test** & **TI**

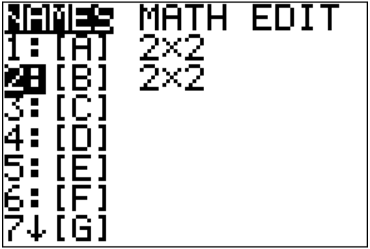
Now use the $\boxed{\downarrow}$ to go to **Calculate** option, press **ENTER** to execute. The display screen verifies the same answer we had earlier.

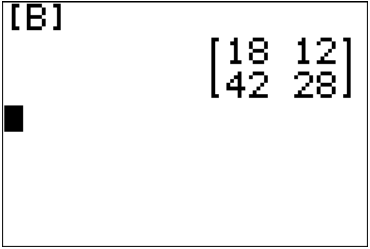


## What about **Matrix B**?

**Matrix B** is the **Expected Frequency Table** and it is updated by TI automatically. To view **Matrix B**, press **2ND** followed by $\boldsymbol{x^{-1}}$ for the **MATRIX** menu, then choose option 2 on the default menu followed by **ENTER**.



*Example:*

Suppose a survey asked 200 students which platform they use most (such as TikTok, Instagram, or Snapchat) and how often they check social media during study time.

| Social Media Platform | Rarely | Sometimes | Often |
|---|---|---|---|
| TikTok | 57 | 10 | 13 |
| Instagram | 70 | 5 | 5 |
| Snapchat | 25 | 5 | 10 |

Using the significance level $\alpha = 0.1$ to test whether social media platform and checking frequency are independent

**Solution:**

1. Set up $H_0$ and $H_1$.

   $H_0$ : The row and column variables are independent. Claim

   $H_1$ : The row and column variables are dependent. RTT

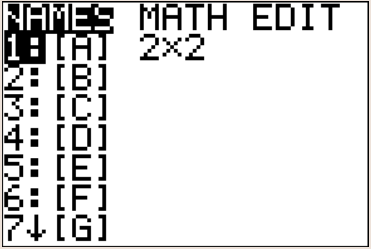
   The row variables are social media platforms while the column variables are the frequency usage divided into rarely, often, and sometimes categories.

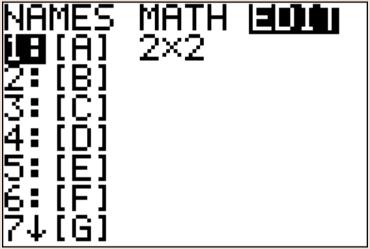
2. Since the **Contingency Table** has 3 rows and 3 columns, we can find the degrees of freedom.

$$df = (r-1)\cdot(c-1) = (3-1)\cdot(3-1) = 4$$

## Solution Continued:

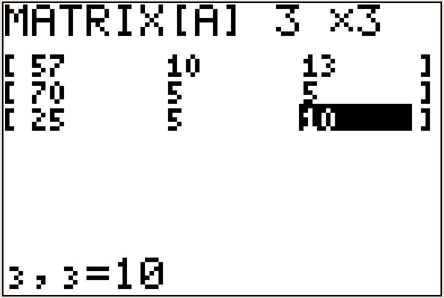
3. Now we store the **Contingency Table** in a matrix. Press **2ND** followed by $\boldsymbol{x^{-1}}$ for the **MATRIX** menu, then use the **→** to go to the **EDIT** menu option, and then press 1 to select matrix A.



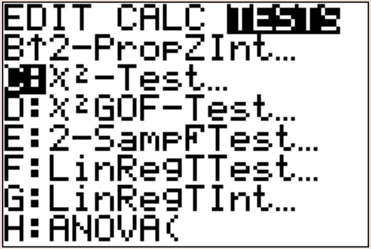
## Solution Continued:

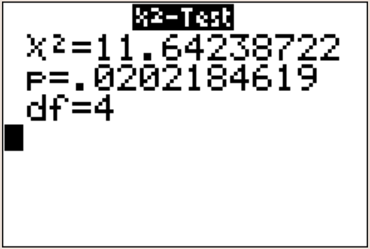
4. Now we can enter the matrix size, and the values of the contingency table.



## Solution Continued:

5. Now press **STAT** followed by $\rightarrow$ to get to the **TESTS** menu, then use the $\downarrow$ to go to **$\chi^2$–Test** option, and press **ENTER** to execute. Follow the steps to get to the result screen.



**Solution Continued:**

6. Since $p-$**value** is less than the given significance level $\alpha = 0.1$, therefore $H_0$ is invalid and $H_1$ is valid.

   We conclude that the social media platforms and frequency usage categories are dependent.

   We are rejecting the claim.

---

If we wish to support the claim that the social media platforms and frequency usage categories are independent, we can simply choose $\alpha = 0.01$.

If you wish to view the values of the **Expected Frequency Table**,

press **2ND** followed by $\boldsymbol{x^{-1}}$ for the **MATRIX** menu, then choose option 2 on the default menu followed by **ENTER**.

[B]

$$\begin{bmatrix} 60.8 & 8 & 11.2 \\ 60.8 & 8 & 11.2 \\ 30.4 & 4 & 5.6 \end{bmatrix}$$

If you wish to calculate the computed test statistic, here is the formula

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

And here are the **Observed** and **Expected** matrices next to each other.

$$O = \begin{bmatrix} 57 & 10 & 13 \\ 70 & 5 & 5 \\ 25 & 5 & 10 \end{bmatrix} \qquad \text{and} \qquad E = \begin{bmatrix} 60.8 & 8 & 11.2 \\ 60.8 & 8 & 11.2 \\ 30.4 & 4 & 5.6 \end{bmatrix}$$

Now let's show the calculation process to find the computed test statistics $\chi^2$ using the formula.

$$\begin{aligned}
\chi^2 &= \sum \frac{(O-E)^2}{E} \\
&= \frac{(57-60.8)^2}{60.8} + \frac{(10-8)^2}{8} + \cdots + \frac{(100-5.6)^2}{5.6} \\
&\approx 11.642
\end{aligned}$$

To find the **P–Value**, we draw the Chi–Square distribution curve with $df = (r-1)\cdot(c-1) = (3-1)\cdot(3-1) = 4$, clearly label the computed test statistic $\chi^2 = 11.642$, and shade the right tail for the right tail test.