

Sheth L.U.J. & Sir M.V. College

Aim :-

**Conducting Chi-square Test of Independence using `chisq.test()` in R
(Practical 9)**

Title

Chi-square Test – Association between Gender and COVID-19 Test Result

Dataset

COVID-19 Patient Symptoms and Diagnosis Dataset.

<https://www.kaggle.com/datasets/miadul/covid-19-patient-symptoms-and-diagnosis-dataset>

1. Objective

To examine whether there is a statistically significant association between **gender** (Male/Female) and **COVID-19 test result** (Positive/Negative).

2. Hypotheses

Null Hypothesis (H_0):

There is **no association** between gender and COVID-19 test result.

H_0 : Gender and COVID_Result are independent

Alternative Hypothesis (H_1):

There is **an association** between gender and COVID-19 test result.

H_1 : Gender and COVID_Result are not independent

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

3. Assumption Checks

3.1 Type of Data

- Both variables are categorical in nature
- Data is expressed in the form of frequencies
- Observations are independent

Hence, the dataset is suitable for applying the Chi-square test.

3.2 Expected Frequencies Condition

All expected cell frequencies are greater than 5, satisfying the Chi-square test requirement.

Expected Frequency Table

| Gender | Negative | Positive |
|--------|----------|----------|
| Female | 1206.72 | 1307.28 |
| Male | 1193.28 | 1292.72 |

```
>
> # =====
> # STEP 6: Expected Frequencies
> # =====
> print(chi_result$expected)

      Negative Positive
Female 1206.72 1307.28
Male   1193.28 1292.72
>
```

4. Contingency Table (Observed Data)

| Gender | Negative | Positive |
|--------|----------|----------|
| Female | 1220 | 1294 |
| Male | 1180 | 1306 |

Name :- Priya Gupta

Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

```
>
> # =====
> # STEP 3: Create Contingency Table
> # =====
> table_data <- table(df$gender, df$covid_result)
> print(table_data)
```

| | Negative | Positive |
|--------|----------|----------|
| Female | 1220 | 1294 |
| Male | 1180 | 1306 |

```
>
```

5. Chi-square Test Results

| Statistic | Value |
|--------------------|--------|
| Chi-square value | 0.5235 |
| Degrees of Freedom | 1 |
| p-value | 0.4694 |

Interpretation:

The p-value obtained from the Chi-square test is **0.4694**, which is greater than the significance level of 0.05. This indicates that the difference between observed and expected frequencies is not statistically significant.

```
>
> # =====
> # STEP 5: Perform Chi-square Test
> # =====
> chi_result <- chisq.test(table_data)
> print(chi_result)
```

Pearson's Chi-squared test with Yates' continuity correction

data: table_data
X-squared = 0.5235, df = 1, p-value = 0.4694

```
>
```

6. Decision Rule

- If $p\text{-value} < 0.05 \rightarrow \text{Reject } H_0$
- If $p\text{-value} \geq 0.05 \rightarrow \text{Accept } H_0$

Since **0.4694** \geq **0.05**, the null hypothesis is **accepted**.

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

7. Conclusion

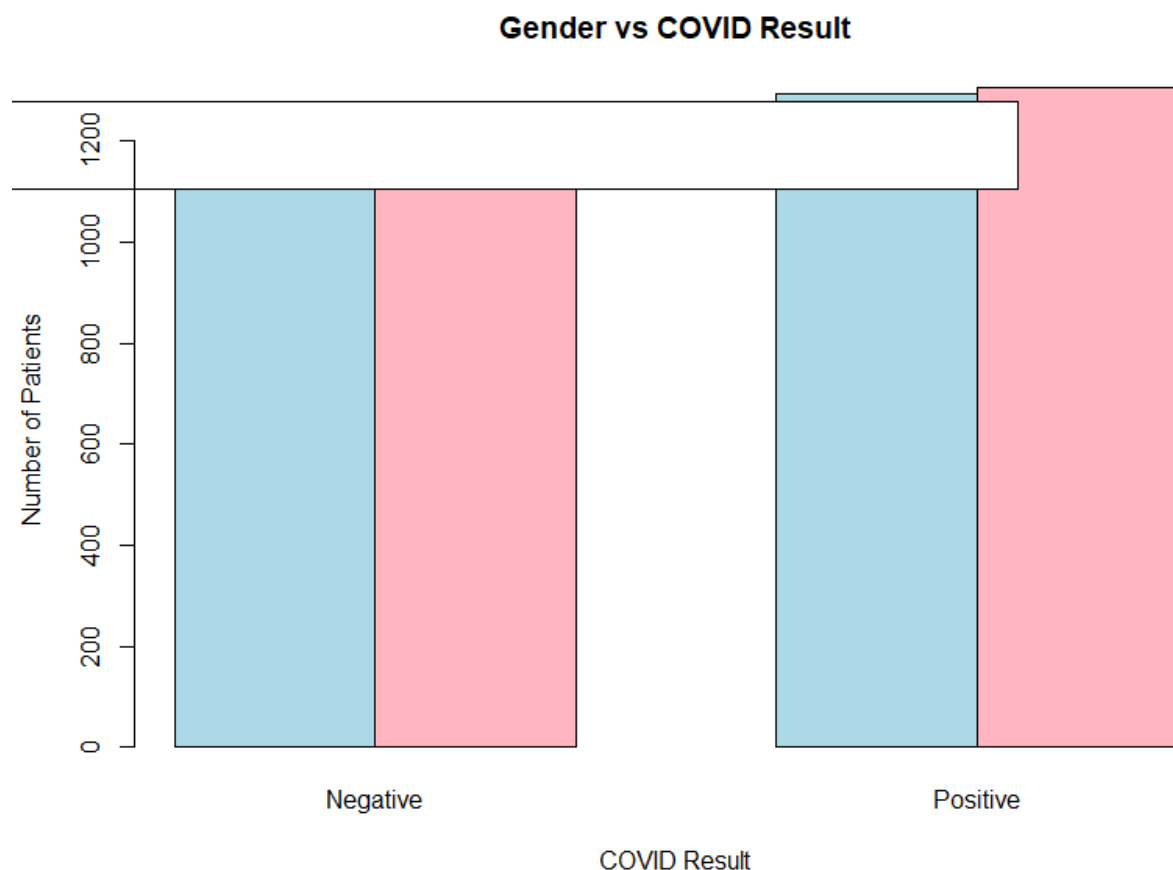
The Chi-square test of independence reveals that there is **no significant association** between gender and COVID-19 test result in the given dataset.

This implies that COVID-19 positivity or negativity does not depend on gender.

8. Visualization

8.1 Bar Plot – Gender vs COVID Result

The bar plot visually represents the distribution of COVID-19 test results across male and female patients. The bars show similar frequencies for both genders, indicating no major difference.

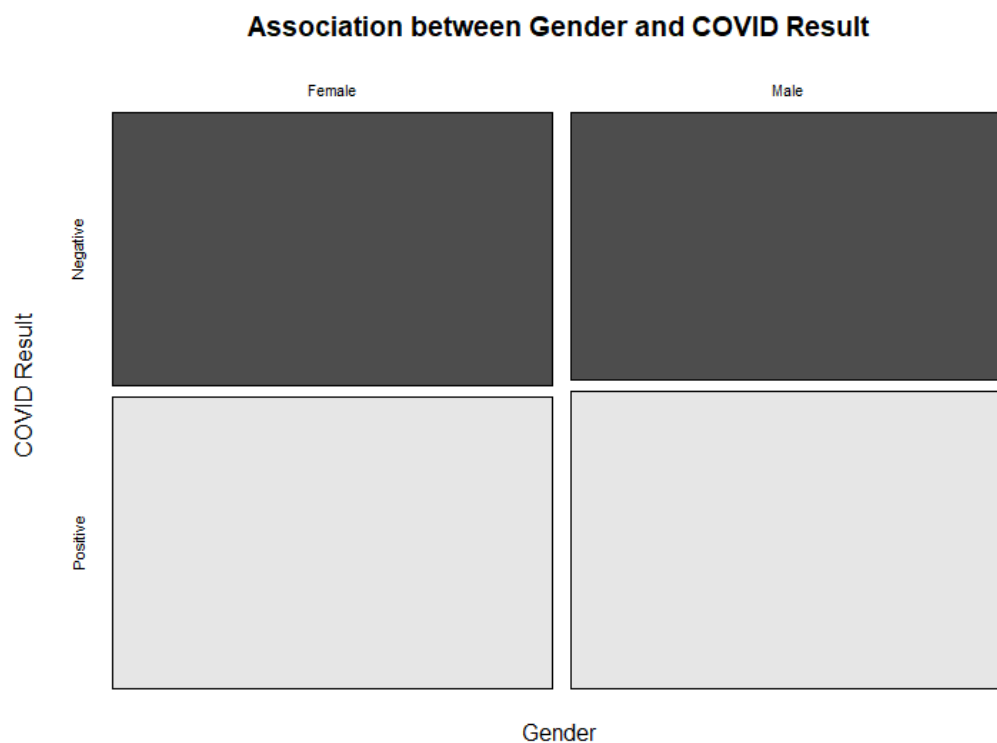


Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

8.2 Mosaic Plot – Association Visualization

The mosaic plot provides a proportional representation of the relationship between gender and COVID result. The plot does not show strong deviation, supporting the Chi-square test findings.



Interpretation from Graphs

- Both genders show nearly equal proportions of COVID-19 positive and negative cases
 - No strong visual association is observed
 - Graphical analysis supports the statistical conclusion
-

9. Final Remark

All assumptions required for the Chi-square test are satisfied.

Although minor graphical warnings were encountered during plotting, they do not affect the validity of the statistical results.

Given the large sample size ($n = 5000$), the findings are considered reliable.

Name :- Priya Gupta

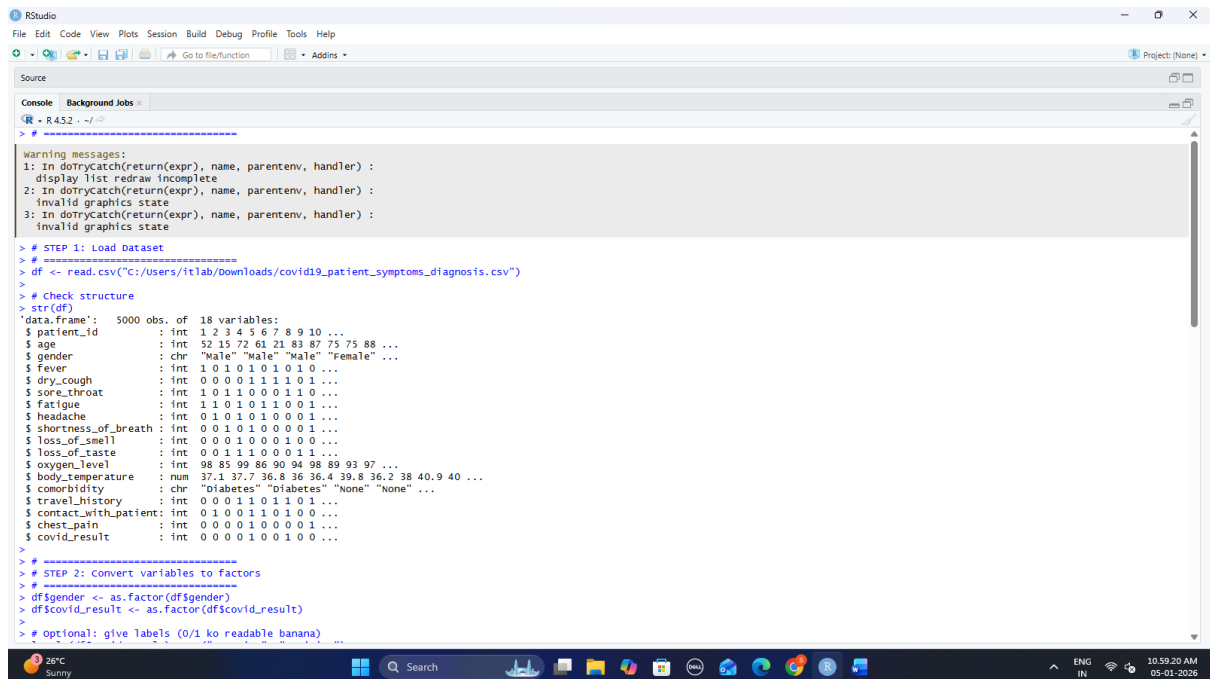
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

10. One-line Practical Conclusion

The Chi-square test confirms that gender and COVID-19 test result are independent, as the p-value is greater than 0.05.

Screenshots



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Background Jobs x
R - R4.5.2 - ~/ /
> # =====
warning messages:
1: In dotrycatch(return(expr), name, parentenv, handler) :
  display list redraw incomplete
2: In dotrycatch(return(expr), name, parentenv, handler) :
  invalid graphics state
3: In dotrycatch(return(expr), name, parentenv, handler) :
  invalid graphics state
> # STEP 1: Load Dataset
> # =====
> df <- read.csv("~/Downloads/covid19_patient_symptoms_diagnosis.csv")
>
> # Check structure
> str(df)
'data.frame': 5000 obs. of  18 variables:
 $ patient_id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ age             : int  52 15 72 61 21 83 87 75 75 88 ...
 $ gender          : chr  "Male" "Male" "Male" "Female" ...
 $ fever           : int  1 0 1 0 1 0 1 0 1 0 ...
 $ dry_cough       : int  0 0 0 0 1 1 1 1 0 1 ...
 $ sore_throat     : int  1 0 1 1 0 0 0 1 1 0 ...
 $ fatigue         : int  1 1 0 1 0 1 1 0 0 1 ...
 $ headache        : int  0 1 0 1 0 1 0 0 0 1 ...
 $ shortness_of_breath : int  0 0 1 0 1 0 0 0 0 1 ...
 $ loss_of_smell   : int  0 0 0 1 0 0 0 1 0 0 ...
 $ loss_of_taste   : int  0 0 1 1 1 0 0 0 1 1 ...
 $ oxygen_level    : int  98 85 99 86 90 94 98 89 93 97 ...
 $ body_temperature : num  37.1 37.7 36.8 36 36.4 39.8 36.2 38 40.9 40 ...
 $ comorbidity     : chr  "Diabetes" "Diabetes" "None" "None" ...
 $ travel_history  : int  0 0 1 1 0 1 1 0 1 ...
 $ contact_with_patient: int  0 1 0 0 1 1 0 1 0 0 ...
 $ chest_pain      : int  0 0 0 0 1 0 0 0 0 1 ...
 $ covid_result    : int  0 0 0 0 1 0 0 1 0 0 ...
>
> # =====
> # STEP 2: Convert variables to factors
> # =====
> df$gender <- as.factor(df$gender)
> df$covid_result <- as.factor(df$covid_result)
>
> # optional: give labels (0/1 ko readable banana)
```

Name :- Priya Gupta
Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Background Jobs x
R - R4.5.2 ~ /
> # =====
> df$gender <- as.factor(df$gender)
> df$covid_result <- as.factor(df$covid_result)
> # optional: give labels (0/1 ko readable banana)
> levels(df$covid_result) <- c("Negative", "Positive")
> # =====
> # STEP 3: Create contingency Table
> # =====
> table_data <- table(df$gender, df$covid_result)
> print(table_data)
      Negative Positive
Female    1220     1294
Male      1180     1306
> # =====
> # STEP 4: Hypothesis
> # H0: gender and covid_result are independent
> # H1: gender and covid_result are associated
> # =====
> # STEP 5: Perform chi-square Test
> # =====
> chi_result <- chisq.test(table_data)
> print(chi_result)

Pearson's Chi-squared test with Yates' continuity correction

data:  table_data
X-squared = 0.5235, df = 1, p-value = 0.4694
> # =====
> # STEP 6: Expected Frequencies
> # =====
> print(chi_result$expected)
      Negative Positive
Female 1206.72 1307.28
Male   1193.28 1292.72
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Background Jobs x
R - R4.5.2 ~ /
X-squared = 0.5235, df = 1, p-value = 0.4694
> # =====
> # STEP 6: Expected Frequencies
> # =====
> print(chi_result$expected)
      Negative Positive
Female 1206.72 1307.28
Male   1193.28 1292.72
> # =====
> # STEP 7: Decision (Accept / Reject H0)
> # =====
> if (chi_result$p.value < 0.05) {
+   print("Reject H0: Gender and COVID result are significantly associated")
+ } else {
+   print("Accept H0: Gender and COVID result are not significantly associated")
+ }
[1] "Accept H0: Gender and COVID result are not significantly associated"
> # =====
> # STEP 8: Bar Plot
> # =====
> barplot(table_data,
+         beside = TRUE,
+         col = c("lightblue", "lightpink"),
+         legend = TRUE,
+         main = "Gender vs COVID Result",
+         xlab = "COVID Result",
+         ylab = "Number of Patients")
> # =====
> # STEP 9: Mosaic Plot
> # =====
> mosaicplot(table_data,
+           main = "Association between Gender and COVID Result",
+           color = TRUE,
+           xlab = "Gender",
+           ylab = "COVID Result")
> |
```

Name :- Priya Gupta
Roll no. :- S081