SHODH PRAKASHAN

An International Open Access, Peer Reviewed, Refereed Journal

ISSN : 2583-9861

Vol. 3, Issue 8, February 2026

Available Online : http://shodhprakashan.in/

Application of chi-square test in Pollution research

Dr. Kavita shrivastav[1], Dr. Jyoti Nema[1]Akanksha Kushwaha**

[1]HOD, Department of Mathematics,

[1]Guest faculty, Department of Mathematics,*

**B.Sc. 4th year (honours)*

Sarojini Naidu Govt. Girls P.G. (Autonomous) College, Bhopal (M.P.)

Abstract: The present study employed the Chi-square test to analyse the relationship between pollution levels and selected categorical environmental factors. Observed frequencies of pollution indicators were compared with expected frequencies across different locations and time periods. The analysis revealed statistically significant differences, indicating non-random variation in pollution distribution. These findings suggest that pollution levels are influenced by environmental and anthropogenic factors. The Chi-square test proved to be an effective statistical tool for identifying associations in pollution data and supporting evidence-based environmental assessment and management strategies.

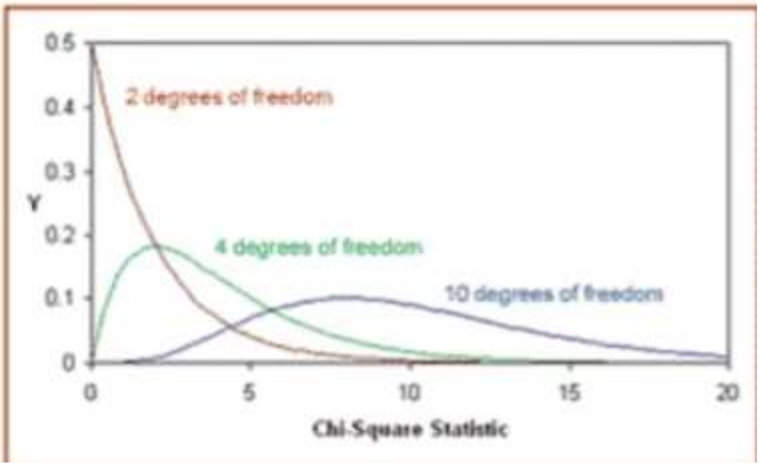
Key word: *Chi-square test, pollution levels, categorical data, observed frequency, expected frequency, association, independence, environmental monitoring, significance level, null hypothesis.*

1. Introduction

[1] The chi-square test is a widely used statistical tool in environmental studies to analyse pollution-related data. [2] It helps determine whether there is a significant association between categorical variables such as pollution levels and locations, seasons, or sources of pollutants. By comparing observed frequencies with expected frequencies,[3] the chi-square test assesses whether differences in pollution patterns occur by chance or due to underlying environmental factors. [4] This test is especially useful in pollution monitoring, impact assessment, and environmental management, as it supports data-driven decision-making and helps identify relationships that influence environmental quality and public health. [5], [6], [7]

2. CHI-SQUARE ($\chi 2$) DISTRIBUTION & $\chi 2$ STATISTIC

The chi-square (χ^2) test and distribution are used in pollution statistics to analyse frequency data such as counts of air-quality levels (good, moderate, poor). It helps determine whether observed pollution levels differ significantly from expected values or whether pollution is associated with factors like location or season. The chi-square statistic is calculated by comparing observed and expected frequencies. The chi-square distribution is positively skewed and depends on degrees of freedom. In environmental studies, it is useful because it does not require normally distributed data and is effective for categorical pollution analysis.



2.1 Application of Chi-square test;

- To test the hypothesis of no association between two or more groups, population or criteria (i.e. to check independence between two variables) contingency table; and
- To test how likely the observed distribution of data fits with the distribution that is expected (i.e., to test the goodness-of-fit) consistency of data recorded with the theoretical values.

2.2 Underlying Assumptions for a Chi-square Test

- The data are randomly drawn from a population
- The values in the cells are considered adequate when expected counts are not <5 and there are no cells with zero count
- The sample size is sufficiently large. The application of the Chi-square test to a smaller sample could lead to type II error (i.e. accepting the null hypothesis when it is actually false). There is no expected cut-off for the sample size; however, the minimum sample size varies from 20 to 50
- The variables under consideration must be mutually exclusive. It means that each variable must only be counted once in a particular category and should not be allowed to appear in other category. In other, words no item shall be counted twice.

2.3 Calculation of Chi-square static:

Chi-square statistic:

It is calculated by the formula, $\chi^2 = \sum \frac{(O-E)^2}{E}$, where, χ2 … the test statistic that asymptotically approaches a chi-square distribution,

O … the observed frequency or recorded observation

E … the expected (theoretical) frequency calculated by the formula, E= a x b/N

a= Row total, b= Column total & N= Gross total

r……the number of rows in the contingency table,

c … the number of columns in the contingency table.

Degrees of freedom (d.f.): In a chi-square test for pollution data, the degrees of freedom (df) depend on the type of test used. For a chi-square goodness-of-fit test, the degrees of freedom are calculated as

df = k − 1, where k is the number of pollution categories (such as air quality levels). If any parameters (like the mean) are estimated from the data, subtract those as well.

For a chi-square test of independence, commonly used to study pollution versus another factor (e.g., location or time),

df = (r − 1) (c − 1), where r is the number of rows and c is the number of columns in the contingency table.

Level of significance(α): In a chi-square test for pollution data, the level of significance (α) is usually set at 0.05(5%) or 0.01(1%). The calculated chi-square value is compared with the critical value at the chosen α and degrees of freedom. If the calculated value exceeds the critical value, the result is statistically significant.

Contingency table: In a chi-square contingency table for pollution, data are grouped by categories such as pollution level and location. Expected frequencies are calculated from totals, and the chi-square statistic checks whether a significant association exists.

1.4 Steps in applying the chi-square test:

1.Define the hypothesis

- **Null hypothesis (H_0):** There is no association or difference in pollution levels.
- **Alternative hypothesis (H_1):** There is an association or difference.

2.Collect data

- Gather pollution measurements (e.g., air quality levels) for different locations, times, or categories.

3.Create a contingency table

- Arrange observed frequencies of pollution across categories (rows = locations, columns = pollution levels).

4.Calculate expected frequencies

- Formula: Expected = (row total × column total) / grand total

5.Compute chi-square statistic

- Formula: $\chi^2 = \Sigma\ [(\text{Observed} - \text{Expected})^2 / \text{Expected}]$

6.Determine degrees of freedom (df)

- df = (rows − 1) (columns − 1)

7.Choose significance level (α)

- Usually, 0.05 or 0.01

8.Compare χ^2 with critical value

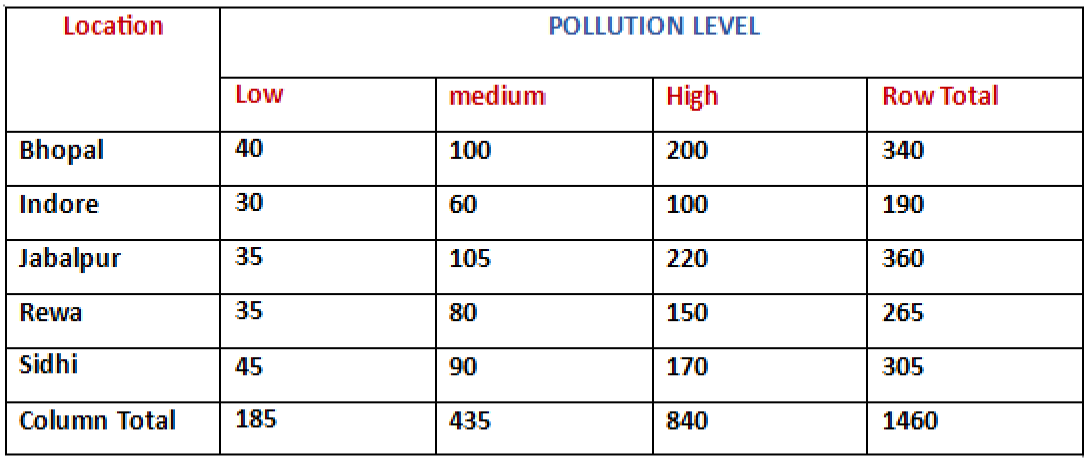
- If $\chi^2 > \chi^2$ critical or p-value < α → reject H_0

9.Interpret results

- Conclude whether pollution levels significantly differ or are associated with categories.

3. WORK EXAMPLES

Example-1: A city measures air pollution at 5 locations (Bhopal, Indore, Jabalpur, Rewa, Sidhi) over a week. Pollution levels are categorized as Low, Medium, High. The observed data are:



Location	POLLUTION LEVEL			
	Low	medium	High	Row Total
Bhopal	40	100	200	340
Indore	30	60	100	190
Jabalpur	35	105	220	360
Rewa	35	80	150	265
Sidhi	45	90	170	305
Column Total	185	435	840	1460

Table 1: Location & Pollution level (observed frequency (o))

We Want to know: **Is Pollution level independent of area? – 50.6**

Steps 1: state the hypotheses

- **Null hypothesis (H_0):** Pollution level is independent of area.
- **Alternative hypothesis (H_1):** Pollution level is dependent on area.

Steps 2: Calculate expected frequencies

The expected frequency for each cell is:

$$E = \frac{\text{(Row total) (Column total)}}{\text{Grant total}}$$

Let's calculate a few:

1. • Bhopal & Low pollution E= (340 X 185)/1000 = 62.9 • Bhopal & medium pollution E= (340 X 435)/1000 = 147.9 • Bhopal & High Pollution E = (340 X 840)/1000 = 285.6	2. • Indore & Low Pollution E = (190 X 185)/1000 = 35.15 • Indore & Medium Pollution E = (190 X 435)/1000 = 82.65 • Indore & High Pollution E = (190 X 840)/1000 = 159.6
3. • Jabalpur & Low Pollution E = (360 X 185)/1000 = 66.6 • Jabalpur & Medium Pollution E = (360 X 435)/1000 = 156.6 • Jabalpur & High Pollution E = (360 X 840)/1000 = 302.4	4. • Rewa & Low Pollution E = (265X 185)/1000 = 49.025 • Rewa & Medium Pollution E = (265 X 435)/1000 = 115.275 • Rewa & High Pollution E = (265 X 840)/1000 = 222.6
5. • Sidhi & Low Pollution E = (305 X 185)/1000 = 56.425 • Sidhi & Medium Pollution E = (305 X 435)/1000 = 132.675 • Sidhi & High Pollution E = (305 X 840)/1000 = 256.2	

So, the expected table is:

Table 2: Location & Pollution level (expected frequency)

location	POLLUTION LEVEL			
	Low	Medium	High	Row Total
Bhopal	62.9	147.9	285.6	496.4
Indore	35.15	82.65	159.6	277.4
Jabalpur	66.6	156.6	302.4	525.6
Rewa	49.025	115.275	222.6	386.9
Sidhi	56.425	132.675	256.2	445.3
Column Total	270.1	635.1	1226.4	2131.6

Step 3: Compute the chi-square statistic:

$$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$$

Where O = observed frequency, E = expected frequency

1. • Bhopal & Low pollution $(40 – 62.9)^2/62.9 = 8.34$ • Bhopal & medium pollution $(100 -147.9)^2/147.9 = 15.5$ • Bhopal & High Pollution $(200 – 285.6)^2/285.6 = 25.7$	2. • Indore & Low Pollution $(30 – 35.15)^2/35.15 = 0.75$ • Indore & Medium Pollution $(60 – 82.65)^2/82.65 = 6.2$ • Indore & High Pollution $(100 – 159.6)^2/159.6 = 22.26$
3. • Jabalpur & Low Pollution $(35 – 66.6)^2/66.6 = 14.9$ • Jabalpur & Medium Pollution $(105 – 156.6)^2/156.6 = 17.1$ • Jabalpur & High Pollution $(220 – 302.4)^2/302.4 = 22.5$	4. • Rewa & Low Pollution $(35 – 49.025)^2/49.025 = 4.1$ • Rewa & Medium Pollution $(80 – 115.275)^2/115.275 =10.8$ • Rewa & High Pollution $(150 – 222.6)^2/222.6 = 23.7$
5. • Sidhi & Low Pollution $(45 – 56.425)^2/56.425 = 2.31$ • Sidhi & Medium Pollution $(90 – 132.675)^2/132.675 = 13.8$ • Sidhi & High Pollution $(170 – 256.2)^2/256.2 = 29.1$	

Add them up:

χ^2=8.34+15.5+25.7+0.75+6.2+22.26+14.9+17.1+22.5+4.1+10.8+

23.7+2.31+13.8+29.1=217.06

Step 4: Degrees of freedom

df = (rows – 1) X (columns – 1) = (5 – 1) X (4 – 1) = 4 X 3 = 12

Step 5: Compare with chi-square critical value

At α =0.05, the critical value for df = 12 is 21.03.

χ^2 =217.06 > 21.03

Since χ^2 > critical value, we reject the null hypothesis.

Conclusion: Pollution level is not independent of area. Jabalpur, Bhopal, Sidhi areas have higher Pollution, Indore, Rewa areas have lower pollution.

4.CONCLUSION

the chi-square test is a useful statistical method for studying pollution-related problems, especially when the data are in categorical form. It helps in determining whether there is a significant association between pollution levels and different factors such as sources of pollution, area type, seasons, or population density. By comparing observed frequencies with expected frequencies, the chi-square test provides an objective way to test hypotheses related to environmental pollution.

When the calculated chi-square value exceeds the table value at a chosen level of significance (generally 0.05), the null hypothesis is rejected, indicating that pollution levels are significantly influenced by the studied factor. Conversely, if the calculated value is lower than the table value, the null hypothesis is accepted, showing no significant relationship.

Thus, the chi-square test supports scientific decision-making in environmental studies. The conclusions drawn from this test help researchers and policymakers understand pollution patterns and formulate effective pollution control and management strategies.

References

1. Gupta, S. P., & Gupta, M. P. (2019). Business Statistics. Sultan Chand & Sons, New Delhi.

   (Useful for chi-square test theory and applications.)

2. Spiegel, M. R., & Stephens, L. J. (2017). Schaum's Outline of Statistics. McGraw-Hill Education.

(Clear explanation of chi-square test with worked examples.)

3. Kothari, C. R. (2004). Research Methodology: Methods and Techniques. New Age International Publishers, New Delhi.

   (Widely used reference for hypothesis testing in research.)

4. Montgomery, D. C., & Ranger, G. C. (2018). Applied Statistics and Probability for Engineers. Wiley India.

   (Applied use of chi-square test in real-world data analysis.)

5. Trivedi, P. R., & Raj, G. (2011). Environmental Pollution and Control. BS Publications, Hyderabad.

   (Background on pollution variables and environmental data.)

6. Peavy, H. S., Row, D. R., & Tchobanoglous, G. (2014). Environmental Engineering. McGraw-Hill.

   (Environmental data interpretation and pollution studies.)

7. World Health Organization (WHO). Environmental Pollution and Health Reports.

   (For pollution-related categorical data used in statistical analysis.)