Semester-VI

B.Sc (Honours) in Physics

DSE 4: **Experimental Techniques**

Lecture

on

Chi-square

Discussed by Dr. K R Sahu

Lecture- IV

Syllabus

- Measurements
  - Accuracy and precision and Significant figures.
  - Error and uncertainty analysis.
  - Types of errors:
    - Gross error,
    - Systematic error,
    - Random error.
  - Statistical analysis of data
    - Arithmetic mean,
    - Deviation from mean,
    - Average deviation,
    - Standard deviation,
    - **Chi-square and**
    - Curve fitting.
  - Guassian distribution.

Goodness of fit

The **goodness of fit** of a statistical model describes how well it fits a set of **observations**. **Measures of goodness** of fit typically summarize the discrepancy between **observed values** and the **values expected under the model in question**. Such measures can be used in **statistical hypothesis testing**, e.g. to test for normality of residuals, to test whether two samples are drawn from identical distributions (**Kolmogorov–Smirnov test**), or whether outcome frequencies follow a specified distribution (**Pearson's chi-squared test**). In the analysis of variance, one of the components into which the variance is partitioned may be a lack-of-fit sum of squares.

- **Fit of distributions**
  - **Bayesian information criterion**
  - **Kolmogorov–Smirnov test**
  - **Cramér–von Mises criterion**
  - **Anderson–Darling test**
  - **Shapiro–Wilk test**
  - **Chi-squared test**
  - **Akaike information criterion**
  - **Hosmer–Lemeshow test**
  - **Kuiper's test**
  - **Kernelized Stein discrepancy**
  - **Zhang's ZK, ZC and ZA tests**
  - **Moran test**

A **chi-squared test**, also written as **χ^2 test**, is a statistical hypothesis test that is valid to perform when the test statistic is **chi-squared distributed** under the **null hypothesis**, specifically **Pearson's chi-squared test** and variants thereof. Pearson's chi-squared test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table.

In the standard applications of this test, the observations are classified into mutually exclusive classes. If the null hypothesis is true, the test statistic computed from the observations follows a χ^2 **frequency distribution**. The purpose of the test is to evaluate how likely the observed frequencies would be assuming the null hypothesis is true.

Test statistics that follow a χ^2 distribution occur when the observations are independent and **normally distributed**, which assumptions are often justified under the **central limit theorem**. There are also χ^2 tests for testing the null hypothesis of independence of a pair of **random variables** based on observations of the pairs.

Chi-squared tests often refers to tests for which the distribution of the test statistic approaches the χ^2 distribution asymptotically, meaning that the sampling distribution (if the null hypothesis is true) of the test statistic approximates a chi-squared distribution more and more closely as sample sizes increase.

Calculating the test-statistic

The value of the test-statistic is

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = N \sum_{i=1}^{n} \frac{(O_i/N - p_i)^2}{p_i}$$

where

χ^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2.

O_i = the number of observations of type *i*.

N = total number of observations

$E_i = Np_i$ = the expected (theoretical) count of type *i*, asserted by the null hypothesis that the fraction of type *i* in the population is p_i

n = the number of cells in the table.

The chi-squared statistic can then be used to calculate a p-value by comparing the value of the statistic to a chi-squared distribution. The number of degrees of freedom is equal to the number of cells n, minus the reduction in degrees of freedom, p.

The result about the numbers of degrees of freedom is valid when the original data are multinomial and hence the estimated parameters are efficient for minimizing the chi-squared statistic. More generally however, when maximum likelihood estimation does not coincide with minimum chi-squared estimation, the distribution will lie somewhere between a chi-squared distribution with n-1-p and n-1 degrees of freedom.

Testing for statistical independence

In this case, an "observation" consists of the values of two outcomes and the null hypothesis is that the occurrence of these outcomes is statistically independent. Each observation is allocated to one cell of a two-dimensional array of cells (called a contingency table) according to the values of the two outcomes. If there are r rows and c columns in the table, the "theoretical frequency" for a cell, given the hypothesis of independence, is

$$E_{i,j} = N p_{i\cdot} p_{\cdot j},$$

where N is the total sample size (the sum of all cells in the table), and

$$p_{i\cdot} = \frac{O_{i\cdot}}{N} = \sum_{j=1}^{c} \frac{O_{i,j}}{N},$$

is the fraction of observations of type *i* ignoring the column attribute (fraction of row totals), and

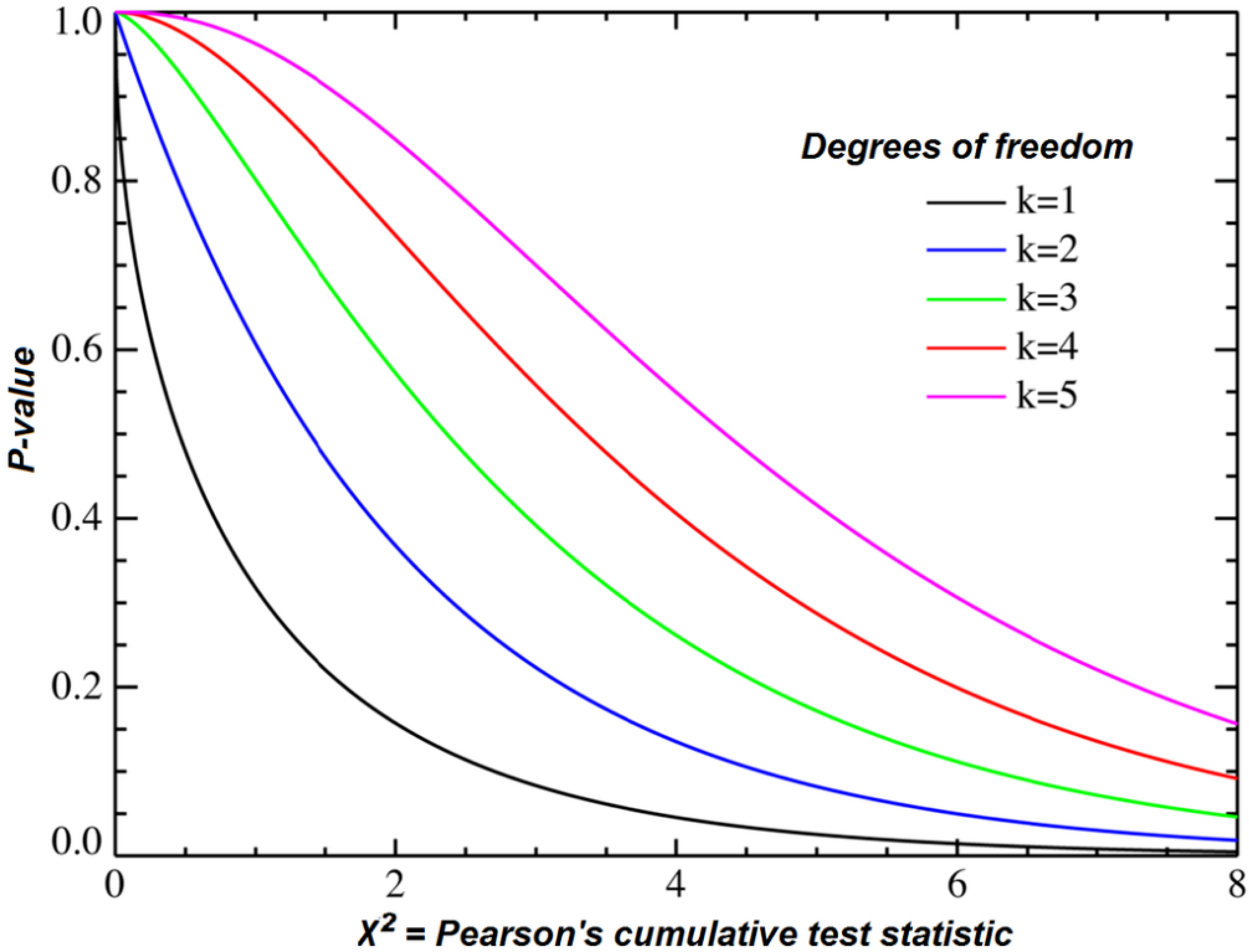
$$p_{\cdot j} = \frac{O_{\cdot j}}{N} = \sum_{i=1}^{r} \frac{O_{i,j}}{N}$$

is the fraction of observations of type j ignoring the row attribute (fraction of column totals). The term "frequencies" refers to absolute numbers rather than already normalized values.

The value of the test-statistic is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$
$$= N \sum_{i,j} p_{i\cdot} p_{\cdot j} \left(\frac{(O_{i,j}/N) - p_{i\cdot} p_{\cdot j}}{p_{i\cdot} p_{\cdot j}} \right)^2$$

Note that χ^2 is 0 if and only if $O_{i,j} = E_{i,j} \forall i, j$, i.e. only if the expected and true number of observations are equal in all cells.



Chi-squared distribution, showing χ2 on the x-axis and P-value on the y-axis.

Fitting the model of "independence" reduces the number of degrees of freedom by $p = r + c - 1$. The number of degrees of freedom is equal to the number of cells rc, minus the reduction in degrees of freedom, p, which reduces to $(r - 1)(c - 1)$.
For the test of independence, also known as the test of homogeneity, a chi-squared probability of less than or equal to 0.05 (or the chi-squared statistic being at or larger than the 0.05 critical point) is commonly interpreted by applied workers as justification for rejecting the null hypothesis that the row variable is independent of the column variable. The alternative hypothesis corresponds to the variables having an association or relationship where the structure of this relationship is not specified.

Assumptions

The chi-squared test, when used with the standard approximation that a chi-squared distribution is applicable, has the following assumptions

Simple random sample

The sample data is a random sampling from a fixed distribution or population where every collection of members of the population of the given sample size has an equal probability of selection. Variants of the test have been developed for complex samples, such as where the data is weighted. Other forms can be used such as purposive sampling.

Sample size (whole table)

A sample with a sufficiently large size is assumed. If a chi squared test is conducted on a sample with a smaller size, then the chi squared test will yield an inaccurate inference. The researcher, by using chi squared test on small samples, might end up committing a Type II error.

Expected cell count

Adequate expected cell counts. Some require 5 or more, and others require 10 or more. A common rule is 5 or more in all cells of a 2-by-2 table, and 5 or more in 80% of cells in larger tables, but no cells with zero expected count. When this assumption is not met, Yates's correction is applied.

Independence

The observations are always assumed to be independent of each other. This means chi-squared cannot be used to test correlated data (like matched pairs or panel data). In those cases, McNemar's test may be more appropriate. A test that relies on different assumptions is Fisher's exact test; if its assumption of fixed marginal distributions is met it is substantially more accurate in obtaining a significance level, especially with few observations. In the vast majority of applications this assumption will not be met, and Fisher's exact test will be over conservative and not have correct coverage.

Example 1: Researchers have conducted a survey of 1600 coffee drinkers asking how much coffee they drink in order to confirm previous studies. Previous studies have indicated that 72% of Americans drink coffee. The results of previous studies (left) and the survey (right) are below. At $\alpha = 0.05$, is there enough evidence to conclude that the distributions are the same?

Response	% of Coffee Drinkers
2 cups per week	15%
1 cup per week	13%
1 cup per day	27%
2+ cups per day	45%

Response	Frequency
2 cups per week	206
1 cup per week	193
1 cup per day	462
2+ cups per day	739

(i) The null hypothesis H0:the population frequencies are equal to the expected frequencies (to be calculated below).
(ii) The alternative hypothesis, Ha: the null hypothesis is false.
(iii) $\alpha = 0.05$.
(iv) The degrees of freedom: $k - 1 = 4 - 1 = 3$.
(v) The test statistic can be calculated using a table:

Response	% of Coffee Drinkers	E	O	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2 cups per week	15%	$0.15 \times 1600 = 240$	206	-34	1156	4.817
1 cup per week	13%	$0.13 \times 1600 = 208$	193	-15	225	1.082
1 cup per day	27%	$0.27 \times 1600 = 432$	462	30	900	2.083
2+ cups per day	45%	$0.45 \times 1600 = 720$	739	19	361	0.5014

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \sum \frac{(O-E)^2}{E} = \underline{8.483}.$$

(vi) From $\alpha = 0.05$ and $k - 1 = 3$, the *critical value* is $\underline{7.815}$.

(vii) Is there enough evidence to reject H_0? Since $\chi^2 \approx 8.483 > 7.815$, there is enough statistical evidence to reject the null hypothesis and to believe that the old percentages no longer hold.

Example 2: A department store, A, has four competitors: B,C,D, and E. Store A hires a consultant to determine if the percentage of shoppers who prefer each of the five stores is the same. A survey of 1100 randomly selected shoppers is conducted, and the results about which one of the stores shoppers prefer are below. Is there enough evidence using a significance level $\alpha = 0.05$ to conclude that the proportions are really the same?

Store	A	B	C	D	E
Number of Shoppers	262	234	204	190	210

(i) The null hypothesis H0:the population frequencies are equal to the expected frequencies (to be calculated below).
(ii) The alternative hypothesis, Ha: the null hypothesis is false.
(iii) $\alpha = 0.05$.
(iv) The degrees of freedom: $k - 1 = 5 - 1 = 4$.
(v) The test statistic can be calculated using a table:

Preference	% of Shoppers	E	O	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
A	20%	$0.2 \times 1100 = 220$	262	42	1764	8.018
B	20%	$0.2 \times 1100 = 220$	234	14	196	0.891
C	20%	$0.2 \times 1100 = 220$	204	-16	256	1.163
D	20%	$0.2 \times 1100 = 220$	190	-30	900	4.091
E	20%	$0.2 \times 1100 = 220$	210	-10	100	0.455

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \sum \frac{(O-E)^2}{E} = \underline{14.618}.$$

vi) From α = 0.05 and k − 1 = 4, the critical value is 9.488.

(vii) Is there enough evidence to reject H0? Since χ 2 ≈ 14.618 > 9.488, there is enough statistical evidence to reject the null hypothesis and to believe that customers do not prefer each of the five stores equally.

Independence

Recall that two events are independent if the occurrence of one of the events has no effect on the occurrence of the other event.

A chi-square independence test is used to test whether or not two variables are independent.

An experiment is conducted in which the frequencies for two variables are determined. To use the test, the same assumptions must be satisfied: the observed frequencies are obtained through a simple random sample, and each expected frequency is at least 5. The frequencies are written down in a table: the columns contain outcomes for one variable, and the rows contain outcomes for the other variable.

The procedure for the hypothesis test is essentially the same. The differences are that:

(i) H0 is that the two variables are independent.

(ii) Ha is that the two variables are not independent (they are dependent).

(iii) The expected frequency Er,c for the entry in row r, column c is calculated using:

Er,c = (Sum of row **r**) × (Sum of column **c**)/ Sample size

(iv) The degrees of freedom: (**number of rows - 1**)×(**number of columns - 1**).

Example 3: The results of a random sample of children with pain from musculoskeletal injuries treated with acetaminophen, ibuprofen, or codeine are shown in the table. At $\alpha = 0.10$, is there enough evidence to conclude that the treatment and result are independent?

	Acetaminophen (c. 1)	Ibuprofen (c. 2)	Codeine (c. 3)	Total
(r. 1) Significant Improvement	58 (66.7)	81 (66.7)	61 (66.7)	200
(r. 2) Slight Improvement	42 (33.3)	19 (33.3)	39 (33.3)	100
Total	100	100	100	300

First, calculate the column and row totals. Then find the expected frequency for each item and write it in the parenthesis next to the observed frequency.

Now perform the hypothesis test.

(i) The null hypothesis H0: the treatment and response are independent.

(ii) The alternative hypothesis, Ha: the treatment and response are dependent.

(iii) $\alpha = 0.10$.

(iv) The degrees of freedom: (number of rows - 1)×(number of columns - 1) $= (2-1) \times (3-1) = 1 \times 2 = 2$.

(v) The test statistic can be calculated using a table:

Row, Column	E	O	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1,1	$\frac{200 \cdot 100}{300} = 66.7$	58	-8.7	75.69	1.135
1,2	$\frac{200 \cdot 100}{300} = 66.7$	81	14.3	204.49	3.067
1,3	$\frac{200 \cdot 100}{300} = 66.7$	61	-5.7	32.49	0.487
2,1	$\frac{100 \cdot 100}{300} = 33.3$	42	8.7	75.69	2.271
2,2	$\frac{100 \cdot 100}{300} = 33.3$	19	-14.3	204.49	6.135
2,3	$\frac{100 \cdot 100}{300} = 33.3$	39	5.7	32.49	0.975

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \sum \frac{(O-E)^2}{E} = 14.07.$$

(vi) From $\alpha = 0.10$ and d.f = 2, the critical value is 4.605.

(vii) Is there enough evidence to reject H0? Since $\chi^2 \approx 14.07 > 4.605$, there is enough statistical evidence to reject the null hypothesis and to believe that there is a relationship between the treatment and response.

Practice Problem 1: A doctor believes that the proportions of births in this country on each day of the week are equal. A simple random sample of 700 births from a recent year is selected, and the results are below. At a significance level of 0.01, is there enough evidence to support the doctor's claim?

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Frequency	65	103	114	116	115	112	75

(i) The null hypothesis H0: the population frequencies are equal to the expected frequencies (to be calculated below).
(ii) The alternative hypothesis, Ha: the null hypothesis is false.
(iii) $\alpha = 0.01$.
(iv) The degrees of freedom: $k - 1 = 7 - 1 = 6$.
(v) The test statistic can be calculated using a table:

Day	E	O	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Sunday	$700/7=100$	65	-35	1225	12.25
Monday	$700/7=100$	103	3	9	0.09
Tuesday	$700/7=100$	114	14	196	1.96
Wednesday	$700/7=100$	116	16	256	2.56
Thursday	$700/7=100$	115	15	225	2.25
Friday	$700/7=100$	112	12	144	1.44
Saturday	$700/7=100$	75	-25	625	6.25

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \sum \frac{(O-E)^2}{E} = \underline{26.8}.$$

(vi) From $\alpha = 0.01$ and $k - 1 = 6$, the critical value is 16.812.

(vii) Is there enough evidence to reject H0? Since $\chi 2 \approx 26.8 > 16.812$, there is enough statistical evidence to reject the null hypothesis and to believe that the proportion of births is not the same for each day of the week.

Practice Problem 2: The side effects of a new drug are being tested against a placebo. A simple random sample of 565 patients yields the results below. At a significance level of $\alpha = 0.05$, is there enough evidence to conclude that the treatment is independent of the side effect of nausea?

Result	Drug (c.1)	Placebo (c.2)	Total
Nausea (r.1)	36	13	49
No nausea (r.2)	254	262	516
Total	290	275	565

(i) The null hypothesis H0: the treatment and response are independent.

(ii) The alternative hypothesis, Ha: the treatment and response are dependent.

(iii) $\alpha = 0.01$.

(iv) The degrees of freedom: (number of rows - 1)×(number of columns - 1) = $(2-1) \times (2-1) = 1 \times 1 = 1$.

(v) The test statistic can be calculated using a table:

Row, Column	E	O	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1,1	$\frac{49 \cdot 290}{565} = 25.15$	36	10.85	117.72	4.681
1,2	$\frac{49 \cdot 275}{565} = 23.85$	13	−10.85	117.72	4.936
2,1	$\frac{516 \cdot 290}{565} = 264.85$	254	−10.85	117.72	0.444
2,2	$\frac{516 \cdot 275}{565} = 251.15$	262	10.85	117.72	0.469

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \sum \frac{(O-E)^2}{E} = 10.53.$$

(vi) From $\alpha = 0.10$ and d.f = 1, the critical value is 2.706.

(vii) Is there enough evidence to reject H0? Since $\chi 2 \approx 10.53 > 2.706$, there is enough statistical evidence to reject the null hypothesis and to believe that there is a relationship between the treatment and response.

Practice Problem 3: Suppose that we have a 6-sided die. We assume that the die is unbiased (upon rolling the die, each outcome is equally likely). An experiment is conducted in which the die is rolled 240 times. The outcomes are in the table below. At a significance level of $\alpha = 0.05$, is there enough evidence to support the hypothesis that the die is unbiased?

Outcome	**1**	**2**	**3**	**4**	**5**	**6**
Frequency	34	44	30	46	51	35

(i) The null hypothesis H0: each face is equally likely to be the outcome of a single roll.
(ii) The alternative hypothesis, Ha: the null hypothesis is false.
(iii) $\alpha = 0.05$.
(iv) The degrees of freedom: $k - 1 = 6 - 1 = 5$.
(v) The test statistic can be calculated using a table:

Face	E	O	$O - E$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
1	$240/6 = 40$	34	-6	36	0.9
2	$240/6 = 40$	44	4	16	0.4
3	$240/6 = 40$	30	-10	100	2.5
4	$240/6 = 40$	46	6	36	0.9
5	$240/6 = 40$	51	11	121	3.025
6	$240/6 = 40$	35	-5	25	0.625

$$\chi^2 = \sum \frac{(\text{ observed } - \text{ expected })^2}{\text{expected}} = \sum \frac{(O - E)^2}{E} = 8.35.$$

vi) From α = 0.01 and k − 1 = 6, the critical value is 15.086.
(vii) Is there enough evidence to reject H0? Since $\chi^2 \approx 8.35 < 15.086$, we fail to reject the null hypothesis, that the die is fair.

References

https://en.wikipedia.org/wiki/Goodness_of_fit
http://websupport1.citytech.cuny.edu/Faculty/mbessonov/MAT1272/Worksheet%20November%2021%20Solutions.pdf
https://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test