classmate
Date
Page

CHI-SQUARE (χ^2) Test :-

The chi-square (χ^2) test is a useful measure of comparing experimentally obtained results with those expected, theoretically and based on hypothesis. It is used as a test statistic in testing a hypothesis that provides a set of theoretical frequencies with which observed frequencies are compared.

The magnitude of discrepancy between observed and theoretical frequency is given by the quantity χ^2. If $\chi^2=0$, the observed and expected frequencies completely coincide. As the value of χ^2 increases, the discrepancy between the observed frequencies and $f_{e_1}, f_{e_2}, \ldots f_{e_n}$ be the set of expected frequencies then χ^2 is defined as

$$\chi^2 = \frac{(f_{o_1} - f_{e_1})^2}{f_{e_1}} + \frac{(f_{o_2} - f_{e_2})^2}{f_{e_2}} + \cdots + \frac{(f_{o_n} - f_{e_n})^2}{f_{e_n}}$$

$$= \sum \frac{(f_o - f_e)^2}{f_e}$$

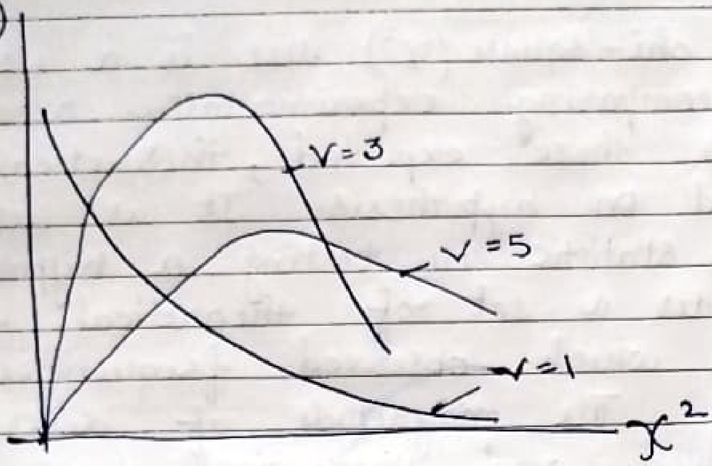
with $(n-1)$ degree of freedom.

In case of bionomial distribution, $v = n-1$

" " " Poission " , $v = n-2$

" " " Normal " , $v = n-3$

Chi-square distribution curve:-



Properties of χ^2 distribution:-

(i) χ^2 test is always positively skewed.

(ii) Mean of χ^2 distribution is no. of degree of freedom.

(iii) Standard deviation $= \sqrt{2v}$

(iv) Chi-square values increases with increases in degree of freedom.

(v) for different value of degree of freedom, the shape of the curve will be different.

CHI-SQUARE TEST: GOODNESS OF FIT :-

~~For~~ The values of x^2 is used to test whether the deviations of the Observed frequencies from the expected frequencies are significant or not. It is also used to fit a set of observations to a given distribution. Hence chi-square test provides a test of goodness of fit and may be used to examine the validity of some hypothesis about an observed freq. distn.

Test of significance :-

Let $f_{o_1}, f_{o_2}, \ldots f_{o_n}$ be a set of observed frequencies and $f_{e_1}, f_{e_2}, \ldots f_{e_n}$ be the corresponding set of expected frequencies. The the χ^2 statistic is given by

$$\chi^2 = \sum \left(\frac{(f_o - f_e)^2}{f_e} \right)$$

Working rule :-

(1) Set up H_0
(2) Set up H_1
(3) Set α
(4) calculate χ^2
(5) Find degree of freedom
(6) Find corresponding value of χ^2 at given level of significance

(7) if $\chi^2_{cal} < \chi^2_{tab}$ then H_0 accepted

if $\chi^2_{cal} > \chi^2_{tab}$ " " rejected.

classmate

Date

Page

Q-1 A dice was thrown 132 times. The following frequencies were observed.

No. obtained :	1	2	3	4	5	6	Total
frequency :	15	20	25	15	29	28	132

Test the hypo. that dice is unbiased.

Soln:-

$n=6$

Expected frequency of each no $= \frac{132}{6} = 22$

$$\boxed{f_e = 22}$$

(1) Null Hypothesis :- H_0 : Dice is unbiased.

(2) Alternate Hypo :- H_1 : Dice is biased

(3) $\alpha = .05$

(4) $$\chi^2 = \Sigma\left(\frac{(f_o - f_e)^2}{f_e}\right)$$

No. obtained	Observed freq f_o	Expected freq f_e	$(f_o-f_e)^2/f_e$
1	15	22	2.23
2	20	22	0.18
3	25	22	0.41
4	15	22	2.23
5	29	22	2.23
6	28	22	1.64

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

$\chi^2 = 8.92$

Critical value:-

$\boxed{v = n-1 = 5}$

$\chi^2_{.05}(v=5) = 11.07$

Decision:- Since $\chi^2_{cal} < \chi^2_{.05}$, the null hypothesis is accepted at 5% level of significance ie dice is unbaised.

Q2 The no. of car accidents in a metropolitan city was found to be 20, 17, 12, 6, 7, 15, 8, 5, 16, and 14 per month resp. Use χ^2 test to check whether these frequencies are in agreement with the belief that occurence of accident was same during 10 month period. Test at 5% level of significance.

Soln:- $n = 10$

Null Hypo:- Occurence of accident was same during 10 month period.

Alternate: Occurence of accident was not same during 10 month period.

$\alpha = .05$

If occurence of accident is same, the expected frequency of accident per month

$$f_e = \frac{20+17+12+6+7+15+8+5+16+14}{10} = 12$$

classmate

Date

Page

Observed freq. f_o	Expected freq. f_e	$\frac{(f_o - f_e)^2}{f_e}$
20	12	5.33
17	12	2.08
12	12	0
6	12	3
7	12	2.08
15	12	0.75
8	12	1.33
5	12	4.08
16	12	1.33
14	12	0.33

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2_{cal} = 20.31$$

Critical value :- $\quad v = n - 1 = 9.$

$$\chi^2_{.05}(v=9) = 16.92$$

as $\chi^2_{cal} > \chi^2_{.05}$

ie H_0 is rejected. ie occurence of accident was not same during 10 month period.

classmate
Date
Page

Ex-3 200 digits were chosen at random from a set of tables. The frequency of digits are shown below

Digits:	0	1	2	3	4	5	6	7	8	9
freq:	18	19	23	21	16	25	22	20	21	15

Use χ^2 test to access the correctness of the hypothesis that the digits were distributed in equal no. in the tables from which these were chosen.

Solⁿ:- $n = 10$

Null Hypo:- The digits are distributed in equal no. in tables.

Alternate Hypo:- The digits are not distributed in equal no.

$\alpha = .05$

expected freq. of each digit $= \frac{200}{10}$

$$\boxed{f_e = 20}$$

f_e	f_o	$\frac{(f_o - f_e)^2}{f_e}$
18	20	.2
19	20	0.05
23	20	0.45
21	20	0.05
16	20	0.8
25	20	1.25
22	20	0.2
20	20	0
21	20	0.05
15	20	1.25.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\boxed{\chi^2 = 4.3}$$

$v = 10 - 1 = 9$

$\chi^2_{.05}(v=9) = 16.92$.

as $\chi^2_{cal} < \chi^2_{0.05}$

ie Ho accepted ie digits are distributed in equal no. in table

(1) Theory predicts that the proportion of beans in four groups A, B, C, D should be 9 : 3 : 3 : 1. In an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287, 118. Does the experimental result support the theory.

Solⁿ. $n = 4$

Null Hypo:- H_0: The proportion of bean in four groups is 9 : 3 : 3 : 1.

Alternate Hypo:

H_1: The proportion of beans in four groups A, B, C, D is not 9 : 3 : 3 : 1

$\alpha = .05$ $\quad v = 3$

Test statistic

	f_o	f_e	$\frac{(f_o - f_e)^2}{f_e}$
A	882	$\frac{9}{16} \times 1600 = 900$	0.36
B	313	$\frac{3}{16} \times 1600 = 300$	0.56
C	287	$\frac{3}{16} \times 1600 = 300$	0.56
D	118	$\frac{1}{16} \times 1600 = 100$	3.24

classmate

Date

Page

$$x^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

$$\boxed{\chi^2 = 4.72}$$

$$\chi^2_{0.5}(r=3) = 7.81$$

as $\chi^2_{cal} < \chi^2_{.05}(r=3)$

ie Null Hypo is accepted. ie experimental result support the theory. and proportion of the bean is 9:3:3:1.

classmate

Date ____
Page ____

More:- Question on Goodness of fit:- (fitting a distribution to given data)

Q:- The following mistakes per page were observed in a book:

No. of mistakes Per Page:	0	1	2	3	4
No. of Pages:-	211	90	19	5	0

Fit a Poission distribution and test the goodness of fit.

Sol^n:- H_0: The mistakes follow poission distribution and Poission distribution can be fitted to the data.

H_1: The mistakes do not follow poisson distribution

$\alpha = .05$

Test statistic:-

The expected frequencies by poisson distribution are given by.

$$f_e = NP = N\cdot\left(\frac{e^{-\lambda}\lambda^{x}}{x!}\right), \quad x = 0, 1, 2, 3, --$$

$$\lambda = \text{mean} = \frac{\Sigma f x}{N} = \frac{211\times0 + 90\times1 + 19\times2 + 5\times3 + 0\times4}{5}$$

$$\boxed{\lambda = 0.44}$$

$$f_e = NP = 325\left(\frac{e^{-0.44}(.44)^{x}}{x!}\right)$$

$x = 0, 1, 2, 3, ---$

ie expected or Theoretical frequencies

x	0	1	2	3	4
f_e	209.31	92.1	20.26	2.97	0.33

No. of mistakes	f_o	f_e	$f_o - f_e$	$\frac{(f_o - f_e)^2}{f_e}$
0	211	209.31	1.69	.014
1	90	92.10	-2.1	.048
2	19	20.26	-1.26	.078
3	5	2.97	2.03	1.38
4	0	0.33	-0.33	0.33

$$x^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 1.85$$

degree of freedom $v = (n-2)$
$v = 3.$

$$\chi^2_{.05}(v=3) = 7.815$$

as $\chi^2_{cal} < \chi^2_{.05}(v=4)$

ie Null Hypothesis is accepted at 5% level of significance ie the mistakes follow Poisson's Distribution.

Ex 2 A set of five similar coin is tossed 320 times and result is obtained as follows:

No. of Heads:	0	1	2	3	4	5
frequencies:	6	27	72	112	71	32

Test the hypothesis that the data follow a binomial distribution.

Solⁿ:- Null Hypothesis:-

H_0: The data follow a binomial distⁿ

Alternate Hypothesis:-

H_1: The data donot follow binomial distribution.

$\alpha = .05$ (assumption)

Test statistic:

probability of getting head $p = 1/2$

" " " tail $q = 1/2$.

By binomial distribution.

$$P(X=x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$

$$= {}^{5}C_{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{5-x}$$

$N = 320$

Expected frequency $= NP = 320 \times {}^{5}C_{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{5-x}$

$x = 0, 1, 2, 3, 4, 5$

x :	0	1	2	3	5
f_e :	10	50	100	50	10

No of heads	f_o	f_e	$\frac{(f_o-f_e)^2}{f_e}$
0	6	10	1.6
1	27	50	10.58
2	72	100	7.84
3	112	100	1.44
4	71	50	8.82
5	32	10	48.4

$$\chi^2 = \sum \frac{(f_o-f_e)^2}{f_e}$$

$$\boxed{\chi^2 = 78.68}$$

critical value :- $v = n-1 = 5$

$$\chi^2_{.05}(v=5) = 11.07$$

as $\chi^2_{cal} > \chi^2_{.05}$ ie null Hypothesis is rejected ie data do not follow the bionomial distribution.