Chi-Square (χ^2) Distribution [Basics & Properties]

Standard Normal Variate

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_{d.f.=1}$$

where $X \sim N(\mu, \sigma^2)$

$X_1, X_2, \ldots X_\nu$

$$\chi^2 = \sum_{i=1}^{\nu} \left(\frac{X_i-\mu}{\sigma}\right)^2$$

$d.f. = \nu$

Pdf of χ^2-distribution

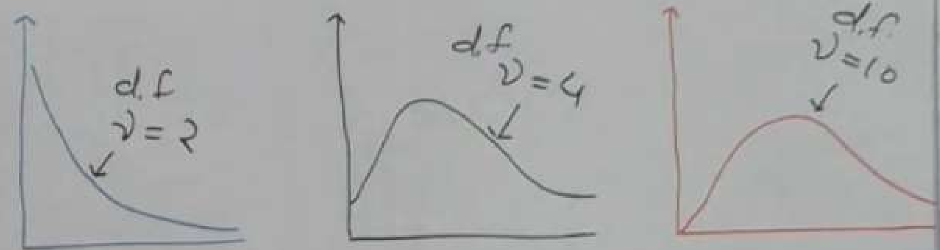
$$f(\chi^2) = C \cdot (\chi^2)^{(\nu/2-1)} \cdot e^{-\chi^2/2}$$

$\chi^2 \geq 0$

C = constant depend on d.f.

ν = d.f.

Properties of χ^2-Distribution



1. Total area under curve is 1.
2. (ν) d.f. is only parameter of χ^2 dist.
3. Mean = d.f (ν), Variance = 2ν
4. χ^2-value lies between 0 to ∞ ($0 \leq \chi^2 < \infty$)
5. For small number of degree of freedom, χ^2 distribution is postively skewed.
6. For different d.f., shape of χ^2-distribution is different.
7. Sum of two independent χ^2 with ν_1 and ν_2 d.f. is also a χ^2-dist with $(\nu_1+\nu_2)$ d.f.

Chi-Square (χ^2) Test [Formula, Conditions, Applications]

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

f_o = observed frequencies
f_e = expected frequencies
d.f. = ν

Conditions

1. Total frequency should be very large. $\Sigma f_o = N > 50$
2. $\Sigma f_o = \Sigma f_e$
3. No expected frequencies should not be less than 5.
4. No Information about parent population
   Non-parametric.

Applications of χ^2-test

1. χ^2- test of Independence of Attributes
2. χ^2- test of
3. χ^2- test of
4. χ^2 test of

$\boxed{1}$ χ^2- test of Independence of Attributes
whethether two attributes are Independent or not.

	Cinegoers	not cinegoers
Literate		
Illiterate		

$\boxed{2}$ χ^2- test of
test of judging whether
with

$\boxed{3}$ χ^2 test of
whether more than
are

Chi-Square (χ^2) Test [Formula, Conditions, Applications]

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

f_o = observed frequencies
f_e = expected frequencies
d.f. = ν

Conditions

1. Total frequency should be very large. $\sum f_o = N > 50$
2. $\sum f_o = \sum f_e$
3. No expected frequencies should not be less than 5.
4. No Information about parent population
   Non-parametric.

Applications of χ^2-test

1. χ^2-test of Independence of Attributes
2. χ^2-test of goodness of fit.
3. χ^2-test of
4. χ^2 test of

$\boxed{1}$ χ^2-test of Independence of Attributes
whethether two attributes are Independent or not

$\boxed{2}$ χ^2-test of goodness fit
is test of judging whether observed data fit or match with expected data.

$\boxed{3}$ χ^2 test of
whether more than

Chi-Square (χ^2) Test [Formula, Conditions, Applications]

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

f_o = observed frequencies
f_e = expected frequencies
d.f. = ν

Conditions

1. Total frequency should be very large. $\sum f_o = N > 50$
2. $\sum f_o = \sum f_e$
3. No expected frequencies should not be less than 5.
4. No Information about parent population
   Non-parametric.

Applications of χ^2-test

1. χ^2- test of Independence of Attributes
2. χ^2- test of goodness of fit
3. χ^2- test of homogenity
4. χ^2 test of

$\boxed{1}$ χ^2- test of Independence of Attributes
whethether two attributes are Independent or not

$\boxed{2}$ χ^2- test of goodness fit
is test of judging whether observed data fit or match with expected data

$\boxed{3}$ χ^2 test of homogenity
whether more than two proportions are equal or not

Chi-Square (χ^2) Test [Formula, Conditions, Applications]

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

Non-Parametric

f_o = observed frequencies
f_e = expected frequencies
d.f. = ν

Conditions

1. Total frequency should be very large. $\Sigma f_o = N > 50$
2. $\Sigma f_o = \Sigma f_e$
3. No expected frequencies should not be less than 5.
4. No Information about parent population
   Non-parametric.

Applications of χ^2-test

1. χ^2-test of Independence of Attributes
2. χ^2-test of goodness of fit
3. χ^2-test of homogenity
4. χ^2 test of inference about population variance → [Popln is Normal] Assume

[1] χ^2-test of Independence of Attributes
whethether two attributes are Independent or not

[2] χ^2-test of goodness fit
is test of judging whether observed data fit or match with expected data

[3] χ^2-test of homogenity
whether more than two proportions are equal or not

Chi-Square (χ^2) Test [Formula, Conditions, Applications]

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

f_o = observed frequencies

f_e = expected frequencies

d.f. = ν

Applications of χ^2-test

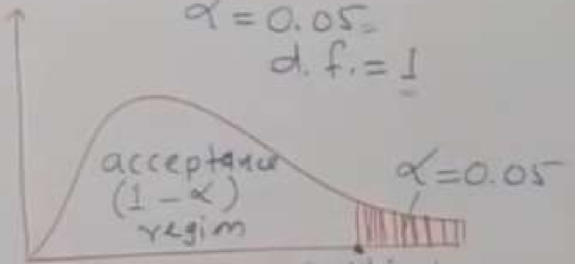
Non-Parametric χ^2

1. χ^2-test of Independence of Attributes
2. χ^2-test of goodness of fit
3. χ^2-test of homogenity
4. χ^2 test of inference about population variance → [popl^n is Normal] Assume

Parametric χ^2

$\alpha = 0.05$

d.f. = 1



critical value

$\chi^2_{0.05,1} = 3.84$ (tabulated)

Cal $\chi^2_{\alpha, d.f.}$ tab. $\chi^2_{\alpha, d.f}$

Cal χ^2 < tab. χ^2 Ho accept

cal. χ^2 > tab χ^2 Ho reject

Demo table (χ^2)

Level of Significance (α)

d.f. ↓	0.99	0.95	0.05 ✓	0.01
✓1	–	0.004	$\boxed{3.84}$	6.63
2	0.02	0.10	5.99	9.21
3	–	–	–	–
4	–	–	–	–
5	0.55	1.15	11.07	15.09

Chi-Square (χ^2) Test of Independence of Attributes

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

f_o = observed frequencies
f_e = expected frequencies

Independence of Attributes

2×2 Contigency Table

			RT
	a	b	(a+b)
	c	d	(c+d)
CT	(a+c)	(b+d)	N = a+b+c+d

Expected frequencies (f_e)

$$E(a) = \frac{(a+b)(a+c)}{N} \quad \left[E = \frac{RT \times CT}{N}\right]$$

$$E(b) = \frac{(a+b) \times (b+d)}{N}$$

$$E(c) = \frac{(c+d)(a+c)}{N}$$

$$E(d) = \frac{(c+d)(b+d)}{N}$$

H_0: No Association b/w two attributes

$$d.f. = \nu = (\underset{\text{no. of row}}{r}-1)(\underset{\text{no. of column}}{c}-1) = (2-1)(2-1) = 1$$

f_o	f_e
a	E(a)
b	E(b)
c	E(c)
d	E(d)

cal. $\chi^2_{\alpha,\nu}$ < tab. $\chi^2_{\alpha,\nu}$ Ho accept

cal χ^2 > tab χ^2 Ho reject

Illustration 1. In an anti malarial campaign in a certain area, quinine was administered to 812 persons out of a total population of 3,248. The number of fever cases is shown below :

Treatment	*Fever*	*No Fever*	*Total*
Quinine	20	792	812
No quinine	220	2,216	2,436
Total	240	3,008	3,248

Discuss the usefulness of quinine in checking malaria.

Chi-Square (χ^2) Test of Independence of Attributes

Q 1

Treatment	Fever	No Fever	(RT) Row Total
Quinine	20	792	812
Not Quinine	220	2216	2436
Column Total (CT)	240	3008	N = 3248

Null Hypothesis, H_0: There is no association between Quinine & Malaria.

Under H_0, Expected frequencies

$$E(20) = \frac{812 \times 240}{3248} = 60$$

$$E(792) = \frac{812 \times 3008}{3248} = 752$$

$$E(220) = \frac{2436 \times 240}{3248} = 180$$

$$E(2216) = \frac{2436 \times 3008}{3248} = 2256$$

Calculation of χ^2

f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o-f_e)^2}{f_e}$
20	60	-40	1600	1600/60 = 26.6
792	752	40	1600	2.12
220	180	40	1600	8.88
2216	2256	-40	1600	0.70
$\Sigma f_o = 3248$	Σf_e 3248			$\Sigma \frac{(f_o-f_e)^2}{f_e} = 38.37$

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] = 38.37$$

d.f. $= \nu = (r-1)(c-1) = (2-1)(2-1) = 1$

LOS, $\alpha = 0.05$

$\boxed{\text{Calculated } \chi^2_{0.05,1} = 38.37}$

Tabulated $\chi^2_{0.05,1} = 3.84$

Decision:- Cal χ^2 > tab χ^2 H_0 reject

38.37 > 3.84

Two researchers adopted different sampling techniques while investigating the same group of students to find the number of students falling in different intelligence level. The results are as follows :

No. of students in each level

Researcher	Below average	Average	Above average	Genius	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Would you say that the sampling techniques adopted by the two researchers are significantly different ?

Chi-Square (χ^2) Test of Independence of Attributes

Q.2

Researchers	Below Average	Average	Above Average	Genius	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	N = 300

Null Hypothesis, H_0: Sampling Techniques adopted by two reachers are not Significantly Differ.

Under H_0

$E(86) = \frac{2\cancel{00} \times 126}{3\cancel{00}} = 84$

$E(60) = \frac{2\cancel{00} \times 93}{3\cancel{00}} = 62$

$E(44) = \frac{2\cancel{00} \times 69}{3\cancel{00}} = 46$

$E(10) = \frac{2\cancel{00} \times 12}{3\cancel{00}} = 8$

$E(40) = \frac{1\cancel{00} \times 126}{3\cancel{00}} = 42$

$E(33) = \frac{1\cancel{00} \times 93}{3\cancel{00}} = 31$

$E(25) = \frac{1\cancel{00} \times 69}{3\cancel{00}} = 23$

$E(2) = \frac{1\cancel{00} \times 12}{3\cancel{00}} = 4$

f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
86	84	2	4	0.046
60	62	-2	4	0.064
44	46	-2	4	0.086
10	08	2	4	0.5
40	42	-2	4	0.095
33	31	2	4	0.012
$\left.\begin{matrix}25\\02\end{matrix}\right\} 27$	$\left.\begin{matrix}23\\04\end{matrix}\right\} 27$	0	0	0
$\Sigma f_o = 300$	$\Sigma f_e = 300$			$\Sigma \frac{(f_o - f_e)^2}{f_e} = 0.911$

$d.f. = \nu = (r-1)(c-1) - 1 = (2-1)(4-1) - 1$

$d.f. = 3 - 1 = 2$

Chi Square test for Independence of Attributes| Concept and Solved Numerical |

Chi-Square (χ^2) Test of Independence of Attributes

Q.2

Researchers	Below Average	Average	Above Average	Genius	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	N = 300

Null Hypothesis, H_0: Sampling Techniques adopted by two reachers are not Significantly Differ.

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] = 0.911$$

$d.f = \nu = 2$

$$\boxed{\text{Cal } \chi^2_{0.05,2} = 0.911}$$

Tab. $\chi^2_{0.05,2} = 5.991$

Cal χ^2 < tab χ^2

0.911 < 5.991

f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o-f_e)^2}{f_e}$
86	84	2	4	0.046
60	62	-2	4	0.064
44	46	-2	4	0.086
10	08	2	4	0.5
40	42	-2	4	0.095
33	31	2	4	0.012
25, 02 } 27	23, 04 } 27	0	0	0
Σf_o = 300	Σf_e = 300			$\Sigma\frac{(f_o-f_e)^2}{f_e} = 0.911$

$d.f = \nu = (r-1)(c-1) - 1 = (2-1)(4-1) - 1$

$d.f = 3 - 1 = 2$

1. The demand for a particular spare part in a factory was found to vary from day to day. In a sample study the following information was obtained:-

Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
No. of parts demanded	1124	1125	1110	1120	1126	1115

Test the hypothesis that the number of parts does not depend on the day of the week.($\chi^2_{0.05}$ at 5 d.f. at 0.05% level of significance is 11.07).

Chi-Square (χ^2) Test [goodness of fit for Uniform Distribution]

Q 1

Null Hypothesis, H_0 : The no. of Parts demanded is uniform.

Under H_0 : Expected frequencies are given by

$$f_e = \frac{6720}{6} = 1120$$

Tabulated $\chi^2_{0.05,5} =$

Calculation of χ^2 - test

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
1124	1120	4	16	16/1120 = 0.014
1125	1120	5	25	0.022
1110	1120	−10	100	0.084
1120	1120	0	0	0
1126	1120	6	36	0.032
1115	1120	−5	25	0.022
$\Sigma f_o =$ 6720	$\Sigma f_e =$ 6720			$\Sigma \frac{(f_o - f_e)^2}{f_e} = 0.179$

$$\chi^2 = \Sigma \left[\frac{(f_o - f_e)^2}{f_e} \right] = 0.179$$

d.f., $\nu = n - 1 = 6 - 1 = 5$

$$\boxed{\text{Calculated } \chi^2_{0.05,5} = 0.179}$$

Chi-Square (χ^2) Test [goodness of fit for Uniform Distribution]

Q 1

Null Hypothesis, H_0: The no. of Parts demanded is uniform.

Under H_0: Expected frequencies are given by

$$f_e = \frac{6720}{6} = 1120$$

Tabulated $\chi^2_{0.05,5} = 11.07$

Decision:- Cal χ^2 < tab χ^2

0.179 < 11.07

H_0 accept.

Calculation of χ^2 - test

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
1124	1120	4	16	16/1120 = 0.014
1125	1120	5	25	0.022
1110	1120	-10	100	0.084
1120	1120	0	0	0
1126	1120	6	36	0.032
1115	1120	-5	25	0.022
$\Sigma f_o =$ 6720	$\Sigma f_e =$ 6720			$\Sigma \frac{(f_o - f_e)^2}{f_e} = 0.179$

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] = 0.179$$

d.f., $\nu = n - 1 = 6 - 1 = 5$

$$\boxed{\text{Calculated } \chi^2_{0.05,5} = 0.179}$$

The following figures show the distribution of digits in number chosen at random from a telephone directory:

Digits	0	1	2	3	4	5	6	7	8	9	Total
Frequency	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whether the digit may be taken to occur equally frequently in the directory.

(The 5% value of χ^2 for 9 d.f. is 16.919).

Chi-Square (χ^2) Test [goodness of fit for uniform Distribution]

Q.2

Null Hypothesis, H_0: Digits occurs equally frequently.

Under H_0: Expected frequencies are given by

$$f_e = \frac{10,000}{10} = 1000$$

Calculation of χ^2 test

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
1026	1000	26	676	676/1000 = 0.676
1107	1000	107	11449	11.449
997	1000	-3	9	0.009
966	1000	-34	1156	1.156
1075	1000	75	5675	5.625
933	1000	-67	4489	4.489
1107	1000	107	11449	11.449
972	1000	-28	784	0.784
964	1000	-36	1296	1.296
853	1000	-147	21609	21.609
$\Sigma f_o = 10,000$	$\Sigma f_e = 10000$			$\Sigma \frac{(f_o - f_e)^2}{f_e} = 58.54$

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] = 58.54$$

d.f., $\nu = n - 1 = 10 - 1 = 9$

Calculated $\chi^2_{0.05,9} = 58.54$

Example 15·14. *A survey of 800 families with four children each revealed the following distribution :*

No. of boys	:	0	1	2	3	4
No. of girls	:	4	3	2	1	0
No. of families	:	32	178	290	236	64

Is this result consistent with the hypothesis that male and female births are equally probable ?