5.4 Chi square distribution

Test for goodness of fit

Conditions to apply χ^2 test for goodness of fit:

- The observations should be independent.
- The total frequency N should be large.
- If any E_i is less than 5, it should be pooled with the adjacent frequency.
- If any parameter is estimated, corresponding to every such estimation, one degree of freedom should be lessened.

Working rule:

Assume:

Expected frequency distribution is a good fit to the observed frequency distribution.

Calculated value:

Under H_0, $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$ with $n - c$ degrees of freedom.

Where $O_i -$ Observed frequency or tabulated frequency

$E_i -$ Expected frequency or theoretical frequency

$n -$ number of terms, $c -$ number of constraints

Critical value:

Level of significance $\alpha = 0.05\ or\ 0.01$ (Always upper tailed)

Degrees of freedom $\gamma = n - c$. Where $c = \begin{cases} 1, & In\ general \\ 2, & For\ Poisson\ distribution \\ 3, & For\ normal\ distribution \end{cases}$

Cont.

Conclusion:

If calculated value < critical value, the expected frequency distribution is **<u>a good fit</u>** to the observed frequency distribution.

If calculated value > critical value, the expected frequency distribution **<u>is not a good</u>** fit to the observed frequency distribution.

Critical values of chi square distribution

d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

1. A die is thrown 60 times and the frequency distribution for the number appearing on the face x is given by the following table:

x:	1	2	3	4	5	6
f:	15	6	4	7	11	17

Test the hypothesis that the die is unbiased. $[\chi_{0.05}(5) = 11.07$

$x =$ Number appearing on the face

$= \{1, 2, 3, 4, 5, 6\}$

$P(x) = \frac{1}{6}$, Total frequency $N = 60.$

$f(x) = N \times P(x) = 60 \times \left(\frac{1}{6}\right) = 10.$

Put $x = 1, 2, 3, 4, 5, 6$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	15	10	25	2.5
2	6	10	16	1.6
3	4	10	36	3.6
4	7	10	9	0.9
5	11	10	1	0.1
6	17	10	49	4.9
				13.6

Cont.

H_0: The die is unbiased.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 13.6$$

Calculated value = 13.6

Degrees of freedom= $n - 1 = 6 - 1 = 5$

Critical value = $\chi_{0.05}(5) = 11.07$

$\therefore$ Calculated value > Critical Value.

Reject H_o.

Therefore, the die is not unbiased.

2. **The following table gives the number of road accidents that occurred in a large city during the various days of a week. Test the hypothesis that the accidents are uniformly distributed over all the days of a week.** $[\chi_{0.05}(6) = 12.59]$

Day:	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
No. of accidents:	14	16	8	12	11	9	14	84

$x =$ Various days of a week

$= \{Sun, Mon, Tue, Wed, Thu, Fri, Sat\}$

$P(x) = \frac{1}{7}$, Total frequency $N = 84$

$f(x) = N \times P(x) = 84 \times \left(\frac{1}{7}\right) = 12$

For $x = Sun, Mon, \ldots$ we get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
Sun	14	12	4	4/12
Mon	16	12	16	16/12
Tue	8	12	16	16/12
Wed	12	12	0	0
Thu	11	12	1	1/12
Fri	9	12	9	9/12
Sat	14	12	4	4/12
				50/12

Cont.

H_o: Accidents are uniformly distributed

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{50}{12}$$

Calculated value = 4.165

Degrees of freedom= $n - 1 = 7 - 1 = 6$,

Critical value= $\chi_{0.05}(6) = 12.59$

$\therefore$ Calculated value < Critical Value.

Accept H_o.

Therefore, the accidents are uniformly distributed.

3. **A set of 5 similar coins is tossed 320 times and the result is**

Number of heads: 0 1 2 3 4 5

Frequency: 6 27 72 112 71 32

Test the hypothesis that the data follows a binomial distribution. $[\chi_{0.05}(5) = 11.07]$

By data, $n = 5, p = \frac{1}{2}, q = \frac{1}{2}, N = 320$

$x =$ Number of heads$= \{0, 1, 2, 3, 4, 5\}$

$$f(x) = \mathrm{N} \times P(x)$$

$$= 320 \times nC_x p^x q^{n-x}$$

$$= 320 \times 5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$$

$$= 320 \times \frac{1}{32} \times 5C_x$$

$$= 10 \times 5C_x$$

Put $x = 0, 1, 2, 3, 4, 5$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0	6	10	16	1.6
1	27	50	529	10.58
2	72	100	784	7.84
3	112	100	144	1.44
4	71	50	441	8.82
5	32	10	484	48.4
				78.68

Cont.

H_o: The data follows Binomial distribution.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 78.68$$

Calculated value = 78.68

Degrees of freedom= $n - 1 = 6 - 1 = 5$

Critical value= $\chi_{0.05}(5) = 11.07$

∴ Calculated value > Critical Value.

Reject H_o.

Therefore, the data does not follow Binomial distribution.

4. Fit a Poisson distribution to the following data and test for its goodness of fit at level of significance 0.05.

x:	0	1	2	3	4
f:	419	352	154	56	19

$[\chi_{0.05}(3) = 7.82]$

$x = \{0, 1, 2, 3, 4\}$

$$m = mean = \frac{\Sigma fx}{\Sigma f} = \frac{904}{1000} = 0.904$$

$$e^{-m} = e^{-0.904} = 0.4049$$

$$P(x) = \frac{e^{-m} m^x}{x!}$$

$$= \frac{(0.4049)(0.904)^x}{x!}$$

$$f(x) = 1000 \times P(x)$$

$$= 404.9 \frac{(0.904)^x}{x!}$$

Put $x = \{0, 1, 2, 3, 4\}$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0	419	405+1	169	0.4033
1	352	366	196	0.5355
2	154	165	121	0.7333
3	56	50	36	0.7200
4	19	11+2	36	2.7692
				5.1613

Numbers added in E_i only to preserve totality.

Cont.

H_o: The data follows Poisson distribution.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 5.1613$$

Calculated value = 5.1613

Degrees of freedom= $n - 2 = 5 - 2 = 3$

($\because$ It follows Poisson distribution)

Critical value = $\chi_{0.05}(3) = 7.82$

$\therefore$ Calculated value < Critical Value.

Accept H_o.

Therefore, the data follows Poisson distribution.

5. Genetic theory states that children having one parent of blood type M and other blood type N will always be one of the three types M, MN, N and that the proportion of these types will on average 1:2:1. A report says that out of 300 children having one M parent and one N parent, 30% were found to be type M, 45% of type MN and remainder of type N. Test the hypothesis by χ^2 test. $[\chi_{0.05}(2) = 5.99]$

---

x is a set of types of blood

$x = \{M, MN, N\}$

$P(M) = \frac{1}{4}$

$P(MN) = \frac{2}{4}$

$P(N) = \frac{1}{4}$

$f(x) = 300 \times P(x)$

Put $x = M, MN, N$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
M	30% of 300 $= 90$	$300 \times \frac{1}{4} = 75$	225	3
MN	45% of 300 $= 135$	$300 \times \frac{2}{4} = 150$	225	1.5
M	25% of 300 $= 75$	$300 \times \frac{1}{4} = 75$	0	0
				4.5

Cont.

H_0: The proportion of these types is on average 1:2:1

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 4.5$$

Calculated value = 4.5

Degrees of freedom= $n - 1 = 3 - 1 = 2$,

Critical value= $\chi_{0.05}(2) = 5.99$

$\therefore$ Calculated value < Critical Value.

Accept H_0.

The proportion of these types is on average 1:2:1

6. In experiments on Pea breading, the following frequencies of seeds were obtained:

Round and yellow	**Wrinkled and yellow**	**Round and green**	**Wrinkled and green**	**Total**
315	**101**	**108**	**32**	**556**

Theory predicts that the frequencies should be in proportions $9:3:3:1$. Examine the correspondence between theory and experiment. $[\chi_{0.05}(3) = 7.82]$

$x = Set\ of\ types\ of\ seeds$

$= \{RY, WY, RG, WG\}$

$P(RY) = \frac{9}{16}, P(WY) = \frac{3}{16},$

$P(RG) = \frac{3}{16}, P(WG) = \frac{1}{16}$

$f(x) = 556 \times P(x)$

Put $x = RY, WY, RG, RY$ to get E_i

x (i)	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
RY	315	313	4	0.0128
WY	101	104	9	0.0865
RG	108	104	16	0.1538
WG	32	35	9	0.2571
				0.5102

Cont.

H_o: The frequencies are in proportions $9:3:3:1$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 0.5102$$

Calculated value = 0.5102

Degrees of freedom= $n - 1 = 4 - 1 = 3$

Critical value= $\chi_{0.05}(3) = 7.82$

$\therefore$ Calculated value < Critical Value.

Accept H_o.

The frequencies should be in proportions $9:3:3:1$