Goodness of fit and contingency tables

Method for testing goodness of fit:

1. Determine which distribution would conceptually be most appropriate
2. Set significance level
3. Estimate parameters (if necessary) from observed data
4. Form hypotheses H_0 and H_1
5. Calculate expected frequencies
6. Combine expected frequencies so that none are < 5
7. Find degrees of freedom
8. Calculate critical value of χ^2 from the table
9. Calculate $\sum \frac{(O_i - E_i)^2}{E_i}$
10. See if the value is significant and draw conclusion

X^2 is distributed with a chi squared distribution χ^2_ν
Where ν = degrees of freedom

$$\text{The number of degrees of freedom} = \text{Number of classes (after combining)} - 1$$

0.1 Example

The data in the table is thought to be modelled by a binomial $B(10, 0.2)$. Use the table for the binomial cumulative distribution function to find expected values, and conduct a test to see if this is a good model. Use a 5% significance level.

x	0	1	2	3	4	5	6	7	8
Freq of x	12	28	28	17	7	4	2	2	0

Define Hypotheses

$$H_0 : \text{A B(10,0.2) distribution is suitable for the results}$$

$$H_1 : \text{The distribution is not suitable for the results}$$

Calculate the sum of frequencies

$$N = 100$$

Complete the table of probabilities and expected frequencies, expected frequency=probability×N

x	0	1	2	3	4	5	6	7	8
$p(x)$	0.1074	0.2684	0.3020	0.2013	0.0881	0.0264	0.0055	0.0008	0.0001
Expected freq	10.74	26.84	30.20	20.13	8.81	2.64	0.55	0.08	0.01

As expected frequencies need to be greater than or equal to five, combine all probabilities greater than or equal to four

x	0	1	2	3	$\geqslant 4$
O_i	12	28	28	17	15
E_i	10.74	26.84	30.20	20.13	12.09
$\frac{(O_i - E_i)^2}{E_i}$	0.1478	0.0501	0.1603	0.4867	0.7004

Find the value of ν

$$\nu = 5 - 1 = 4$$

Find the value of X^2

$$X^2 = 0.1478 + 0.0501 + 0.1603 + 0.4867 + 0.7004 = 1.5453$$

Compare the value of X^2 to the value on the tables corresponding to the 5% significance level and $\nu = 4$

$$9.488 > 1.5453$$

Write conclusion

Not in critical region so insufficient evidence to reject H_0, binomial is a possible model