Chi-Square "Goodness of Fit" Test

Ansar Shahzadi
School of Electrical Engineering & Computer Science
National University of Science and Technology(NUST)

Example: Normal Distribution

Test the Hypothesis that the following frequency distribution of weights follows a normal distribution at $\alpha=0.05$.

Weights	28-31	32-35	36-39	40-43	44-47	48-51	52-55	56-59	60-63	64-67
f	1	14	56	172	245	263	156	67	23	3

Fitting of Normal Distribution

To fit a normal distribution to an observed frequency distribution when neither the mean nor the variance is known, we follow following steps.

- We estimate mean and standard deviation by calculating $\bar{x}$ and S from the observed frequency distribution.
- Then we calculate z= $\frac{u.c.b-\bar{x}}{S}$, As the normal distribution is defined from -∞ to ∞, we therefore extend the lower class -∞ to 27 and the last class 68 to ∞.
- We find cumulative probability P(Z<z) for each value of z-value.
- We the obtain probability $\hat{p}$ of each class by successive subtraction of cumulative probability.
- We finally get the expected frequencies by multiplying each of the class probabilities by the total frequencies of given distribution.

Chi-Square Goodness of Fit Test For Normal Distribution

- **Hypothesis**

H_o: The distribution of Weights is Normal.

H_1: The distribution of Weights is not Normal.

- **Level of Significance**

We choose the α=5%

- **Test Statistics**

We use the test statistics $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

Chi-Square Goodness of Fit Test For Normal Distribution

- **Computations**

Firstly we find the value of Mean and Standard deviation, After Calculations we have

$$\bar{x} = \frac{\sum fx}{\sum f} = 47.71$$

$$S = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{\sum f}} = 5.88$$

Now we find expected frequencies

Weights	Upper class boundary	z= $\frac{u.c.b-\bar{x}}{s}$	P(Z<z)	$\hat{p}$	$f_e = \hat{p}\sum f$
-∞ to 27	27.5	-3.44	0.0003	0.0003	
28-31	31.5	-2.76	0.0029	0.0026	2.9
32-35	35.5	-2.08	0.0188	0.0159	15.9
36-39	39.5	-1.40	0.0808	0.0620	62
40-43	43.5	-0.71	0.2389	0.1581	158.1
44-47	47.5	-0.03	0.4880	0.2491	249.1
48-51	51.5	0.64	0.7389	0.2509	250.9
52-55	55.5	1.32	0.9066	0.1677	167.7
56-59	59.5	2.01	0.9776	0.0712	71.2
60-63	63.5	2.68	0.9963	0.0185	18.5
64-67	67.5	3.37	0.9996	0.0033	3.7
68 to ∞	∞	∞	1	0.0004	
$\sum$					1000

Chi-Square Goodness of Fit Test For Normal Distribution

Weights	f_o	f_e	f_o-f_e	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
28-31	1	2.9			
32-35	14	15.9	-3.8	14.44	0.7681
36-39	56	62.0	-6	36	0.5806
40-43	172	158.1	13.9	193.21	1.2221
44-47	245	249.1	-4.1	16.81	0.0675
48-51	263	250.9	12.1	146.41	0.5835
52-55	156	167.7	-11.7	136.89	0.8163
56-59	67	71.2	-4.2	17.64	0.2478
60-63	23	18.5	3.8	14.44	0.6505
64-67	3	3.7			
Σ					χ^2= 4.94

Chi-Square Goodness of Fit Test For Normal Distribution

To meet the requirement that expected frequency of any of the class should be at least 5, here we combine first two classes and last two classes expected frequencies, the corresponding observed frequencies are also combined.

- **Critical Value**

The critical region is $\chi^2 \geq {\chi^2}_{0.05,(5)}$=11.07

Where degree of freedom= 8-1-2=5

Chi-Square Goodness of Fit Test For Normal Distribution

- **Conculsion**

Since the ${\chi^2}_{cal}$=4.94 does not fall in the critical region, we are therefore unable to reject our null hypothesis. Here we may conclude that the normal distribution Provides a good fit for the given distribution of weights .