Chapter(17)
Nonparmetric Methods
(Examples)

The Chi-Square distribution

1- Test of Goodness of Fit:

Goodness of fit test describes how well it fits a set of observations. Measures of goodness of fit typically summarize the difference between observed values and the values expected under the model in question. It is used to test whether outcome frequencies follow a specified distribution. The test is applied when you have one categorical variable from a single population.

H_0: The data are consistent with a specified distribution.

H_1: The data are *not* consistent with a specified distribution.

Example (1):

Suppose we hypothesize that we have an unbiased six-sided die. To test this hypothesis, we roll the die 300 times and observe the frequency of occurrence of each of the faces as shown in the table below. For α=0.05 Is the die fair?

face	1	2	3	4	5	6
Observed frequency	42	55	38	57	64	44

Solution:

1. State the hypothesis based on your basic question

   H_0: The die is fair and the proportion of each face is equal.

   H_1: The die is biased and the proportion of each face is significantly different.
2. The expected frequencies E_i = 300/6 =50, i=1,2,…,6
3.

face	1	2	3	4	5	6
Observed frequency	42	55	38	57	64	44
Expected frequency	50	50	50	50	50	50
O-E	-8	5	-12	7	14	-6
$(O-E)^2$	64	25	144	49	196	36
$(O-E)^2/E$	1.28	0.5	2.88	0.98	3.92	0.72

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 10.28$$

4. Find the degrees of freedom: df =6-1= 5 (one category).
5. Find the chi-square statistic from the Chi-Square Distribution table $\chi^2_{5,0.05} = 11.07$

The calculated chi-square value (10.28) <tabular chi-square statistic (11.07), we cannot reject the null hypothesis and conclude that the differences between the observed and expected frequencies in each of the six categories are attributable to chance fluctuation.

2- Test of independence:

The test is applied when you have two categorical variables from a single population, and it is used to determine whether there is a significant association between the two variables.

H_0: The two categorical variables are independent.

H_1: The two categorical variables are related.

Example (2):

Suppose you conducted a drug trial on a group of animals and you hypothesized that the animals receiving the drug would show increased heart rates compared to those that did not receive the drug. You conduct the study and collect the following data:

	Heart Rate Increased	No Heart Rate Increase	Total
Treated	36	14	50
Not treated	30	25	55
Total	66	39	105

Solution:

1. H_0: The proportion of animals whose heart rate increased is independent of drug treatment.
   H_1: The proportion of animals whose heart rate increased is associated with drug treatment.
2. The expected value:

$$E_{ij} = \frac{\sum_i O_{ij} \times \sum_j O_{ij}}{\sum_j \sum_i O_{ij}}$$

For example: $E_{11} = \frac{50 \times 66}{105} = 31.4$, and so on.

3.

	Heart Rate Increased		No Heart Rate Increase		Total
Treated	36	31.4	14	18.6	50
Not treated	30	35.6	25	20.4	55
Total	66		39		105

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \frac{(36 - 31.4)^2}{31.4} + \frac{(30 - 35.6)^2}{35.6} + \frac{(14 - 18.6)^2}{18.6} + \frac{(25 - 20.4)^2}{20.4} = 3.42$$

4. Degree of freedom= (2-1)(2-1)=1
5. The chi-square statistic (for α=0.05) $\chi^2_{1,0.05}$ equal 3.841
6. Our calculated value (3.42) is less than the tabular value so we cannot reject the null hypothesis and we conclude that; the proportion of animals whose heart rate increased is independent of drug treatment.

3- Test of Homogeneity:

The Goodness of Fit test can be used to decide whether a population fits a given distribution, but the Goodness of Fit test will not suffice to compare whether two populations follow the same unknown distribution. A different test, called the Test for Homogeneity, can be used to make a conclusion about whether two populations have the same distribution. The test is applied to a single categorical variable from two different populations. It is used to determine whether frequency counts are distributed identically across different populations.

H_0: The distributions of the two populations are the same.

H_1: The distributions of the two populations are not the same.

Example (3):

Do male and female college students have the same distribution of living conditions? Use a level of significance of 0.05. Suppose that 250 randomly selected male college students and 300 randomly selected female college students were asked about their living conditions: Dorm, Apartment, With Parents, Other. The results are shown in the table below:

	Dorm	Apartment	With Parents	Other
Male	72	84	49	45
Female	91	86	88	35

Solution:

1. The null and alternative hypotheses are:
   H_0: The distribution of living conditions for male college students is the same as the distribution of living conditions for female college students.
   H_1: The distribution of living conditions for male college students is the not the same as the distribution of living conditions for female college students.
2. The expected value:

$$E_{ij} = \frac{\sum_i O_{ij} \times \sum_j O_{ij}}{\sum_j \sum_i O_{ij}}$$

3.

	Dorm		Apartment		With Parents		Other		Total
	O	E	O	E	O	E	O	E	
Male	72	74.1	84	77.3	49	62.3	45	36.4	250
Female	91	88.9	86	92.7	88	74.7	35	43.6	300
Total	163		170		137		80		550

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 10.13$$

4. Degree of freedom= (4-1)(2-1)=3
5. The chi-square statistic (for α=0.05) $\chi^2_{3,0.05}$equal 7.815

6. Our calculated value (10.13)> the tabular value (7.815) therefore, *reject* the null hypothesis and accept the alternative hypothesis. You can conclude that the distributions of living conditions for male and female college students are not the same. Notice that the conclusion is only that the distributions are not the same. One cannot use the test for homogeneity to make any conclusions about how they differ.