**NON-PARAMETRIC TEST**

**Kolmogorov-Smirnov one sample test**

Kolmogorov-Smirnov test is a non-parametric test. It is an alternative to chi-square test for goodness of fit between observed frequency and expected frequency. It is especially used in small sample size. It is more powerful than chi-square test.

**Problem:** To test,

**Null hypothesis H0:** F(x) = F0(x); the sample values have come from the population having theoretical distribution F0(x).

**Alternative hypothesis H1:** F(x) F0(x); the sample values have not come from the population having theoretical distribution F0(x). Two-tailed test.

**H1:** F(x) > F0(x); Right-tailed test.

**H1:** F(x) < F0(x); Left-tailed test.

**Test statistic:** Under H0, test statistic is

D 0= Maximum of

Where, F 0 (x) = = Observed cumulative frequency distribution function; k1 = observed cumulative frequencies.

F e (x) = = Expected cumulative frequency distribution function; k2 = expected cumulative frequencies.

**Critical region:** Next for a pre-assigned level of significance and corresponding sample size n, we obtain from Kolmogorov-Smirnov table, the critical value of for one tailed and for two tailed test.

**Decision:** For two-tailed test, if D 0 we reject H0. Otherwise accept H0.

Note: For n>35, the critical value is given by = 1.36/ for = 0.05

**Numerical problem:**

**Example 1:** A game consists matching of four pairs of color cards. Twenty chimpanzees of same age were taught the matching game of color cards for a specified period of time. At the end of the training 4 pairs of color cards are given to each of the chimpanzee for matching. The results were as shown below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Matched set X | 0 | 1 | 2 | 3 | 4 | Total |
| Frequency f | 1 | 1 | 5 | 7 | 6 | 20 |

Does the training an effective means? Use = 0.01.

**Solution:** Here,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Matched set X | Frequency f 0 | f e = | K1 | K2 | F 0 (X) | F e(X) | |F 0 (X) – F e (X)| |
| 0 | 1 | 4 | 1 | 4 | 0.05 | 0.2 | 0.15 |
| 1 | 1 | 4 | 2 | 8 | 0.1 | 0.4 | 0.3 |
| 2 | 5 | 4 | 7 | 12 | 0.35 | 0.6 | 0.25 |
| 3 | 7 | 4 | 14 | 16 | 0.7 | 0.8 | 0.1 |
| 4 | 6 | 4 | 20 | 20 | 1 | 1 | 0 |
| Total | 20 |  |  |  |  |  |  |

n = 5

= = = 4

Now,

**Problem:** To test,

**Null hypothesis H0:** F(x) = F0(x); the sample values have come from the population having theoretical distribution F0(x).

**Alternative hypothesis H1:** F(x) F0(x); the sample values have not come from the population having theoretical distribution F0(x). Two-tailed test.

**Test statistic:** Under H0, test statistic is

D 0= Maximum of = 0.3

**Critical region:** The tabulated value of D at level of significance = 0.01 and corresponding sample size n = 20, we obtain from Kolmogorov-Smirnov table is = = = 0.352

**Decision:** For two-tailed test, if D 0 < we accept H0.

**Conclusion:** Therefore sample values have come from the population having theoretical distribution F0(x).

**Example 2:** An emergency ward of a newly constructed hospital in a village consists of 20 beds. An analyst in due course of study in between 2 pm to 4 pm found the distribution of bed occupants during the seasons is as follows.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Occupants X | Summer | Fall | Winter | Spring | Total |
| Frequency f | 8 | 5 | 20 | 15 | 48 |

Do the data provide sufficient evidence to indicate that the bed occupants of the hospital are uniform for the week period?

Solution: Here,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Occupants X | Frequency f 0 | f e | K1 | K2 | F 0(X) | F e(X) | |F 0(X) - F e(X)| |
| Summer | 8 | 12 | 8 | 12 | 0.166667 | 0.25 | 0.083333 |
| Fall | 5 | 12 | 13 | 24 | 0.270833 | 0.5 | 0.229167 |
| Winter | 20 | 12 | 33 | 36 | 0.6875 | 0.75 | 0.0625 |
| Spring | 15 | 12 | 48 | 48 | 1 | 1 | 0 |
|  | 48 |  |  |  |  |  |  |

n = 4

Expected frequency () = = = 12

Now,

**Problem:** To test,

**Null hypothesis H0:** F(x) = F0(x); the bed occupants of the hospital are uniform distributed for the week period.

**Alternative hypothesis H1:** F(x) F0(x); Two-tailed test.

**Test statistic:** Under H0, test statistic is

D 0= Maximum of = 0.229

**Critical region:** For n>35, the critical value D for = 0.05 is = = 1.36/ = 1.36/ = 0.196

**Decision:** For two-tailed test, if D 0 > we reject H0.

**Conclusion:** Therefore the bed occupants of the hospital are not uniform distributed for the week period.

**Two independent sample test:**

**Two sample Median test**

Median test is a non-parametric test. It is used to test whether the two independent groups have been drawn from two populations with same median or not. Also, median test is used to test whether the treatments applied in an experiment are equally effective or not.

**Case I:** Small sample case i.e. when n1 10 and n2 10 i.e. n1 + n2 = n 20.

**Problem:** To test,

**Null hypothesis H0:** F (X) = F (Y) ⇒ M d(x) = M d (y); the two independent random sample has been drawn from two populations are equal i.e. equally effective.

**Alternative hypothesis H1:** F (X) F (Y) ⇒ M d(x) M d (y); Two-tailed test.

**H1:** F (X) >F (Y) ⇒ M d(x) >M d (y); One-tailed test.

**H1:** F (X) <F (Y) ⇒ M d(x) <M d (y); One-tailed test.

**Test statistic:** Under H0, test statistic is

|  |  |  |  |
| --- | --- | --- | --- |
|  | Xi M d | Xi > M d | Total |
| Sample-I | a | c | a + c |
| Sample-II | b | d | b + d |
| Total | a + b | c + d | N |

Where, the number of observations xiM d and xi>M d are classified into 2x2 contingency table as follows:

**Test statistic:** Under H0, test statistics is given by

If any of cell frequency less than 5, then use chi-square test statistics is given by

**Critical region:** The tabulated value of at% level of significance and 1 df, we obtained from chi-square table

is

**Decision:** If, we accept H0. Otherwise reject H0.

**Example-1:** The following table gives the gain in weight in decagrams in a feeding experiment with pigs on the relative’s value of limestone and bone meal for bone development.

|  |  |
| --- | --- |
| Limestone | 49, 53, 51, 52, 47, 51, 52, 53 |
| Bone meal | 52, 55, 52, 54, 51, 54, 54, 53, 45 |

Use Median test for testing significance of difference between median weight gains of the pigs.

**Solution:** Here,

Arrange the given combined sample observations in ascending order:

45, 47, 49, 51, 51, 51, 52, 52, 52, 52, 53, 53, 53, 54, 54, 54, 55

n1 = 8, n2 = 9 n = n1 + n2 = 8 + 9 = 17

Md = item value = item value = 9th item value = 52

k = = 8.5 = 9

a = No. of observations Md in first sample= 6

Now,

**Problem:** To test,

**Null hypothesis H0:** M d(x) = M d (y); there is no difference between median weight gains of the pigs.

**Alternative hypothesis H1:** M d(x) M d (y); two tailed test.

**Test statistic:** Under H0, test statistic is

P (A = 6) =; a = 0, 1, 2, ------------, 8.

**Critical region:** Next for a pre-assigned level of significance = 0.05, the probability (p0) is associated with the values as extreme as observed values of a, k = 9, n1  = 8and n2 = 9.

P0 = p (A) = p (A) = p (A = 6) + p (A=7) + p(A=8)

= =

= + +

= 0.1089

2P0 = 2X0.1089 = 0.2179

**Decision:** Since, 2p0 > ; we accept H0.

**Conclusion:** There is both samples are equally effective.

**Example-2:** The following table gives the gain in weight in decagrams in a feeding experiment with pigs on the relative’s value of limestone and bone meal for bone development.

|  |  |
| --- | --- |
| Limestone | 49, 53, 51, 52, 47, 51, 52, 53, 60, 45, 40, 61, 56, 42, 59 |
| Bone meal | 52, 55, 52, 54, 51, 54, 54, 53, 45, 65, 57, 50, 48 |

Use Median test for testing significance of difference between median weight gains of the pigs.

**Solution:** Here,

Arrange the given combined sample observations in ascending order:

40, 42, 45, 45,47, 48,49, 50,51, 51, 51,52, 52, 52, 52, 53, 53, 53,54, 54, 54, 55,56, 57,59, 60, 61, 65.

n1 = 15, n2 = 13 n = n1 + n2 = 15 + 13 = 28

Md = item value = item value = 14.5th item value = (52+52)/2 = 52

k = = 8.5 = 9

The number of observations xiM d and xi>M d are classified into 2x2 contingency table as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Xi M d | Xi > M d | Total |
| Sample-I | a = 9 | c = 6 | a + c= 15 |
| Sample-II | b=6 | d = 7 | b + d=13 |
| Total | a + b=15 | c + d=13 | N = 28 |

Now,

**Problem:** To test,

**Null hypothesis H0:** M d(x) = M d (y); there is no difference between median weight gains of the pigs.

**Alternative hypothesis (H1):** M d(x) M d (y); two tailed test.

Test statistic: Under H0, test statistics is given by

= 0.0198

**Critical region:** The tabulated value of at = 0.05 level of significance, we obtained from chi-square table

is = 3.841

**Decision:** Since,, we accept H0.

**Conclusion:** There is both samples are equally effective.