**NON-PARAMETRIC TEST**

A statistical test does not say anything about parameter of population is called non-parametric test.

Difference between parametric and non-parametric test:

|  |  |
| --- | --- |
| Parametric test | Non-parametric test |
| 1. It specifies certain condition about parameter of the population from which sample is selected. | 1. It does not specify certain condition about parameter of the population from which sample is selected. |
| 2. It is used in testing of hypothesis and estimation of parameters. | 2. It is used in testing of hypothesis but not estimation of parameters. |
| 3. Mostly it is used in data measured in interval and ration scale. | 3. It is used in data measured in nominal and ordinal scale. |
| 4. It is most powerful test. | 4. It is less powerful test. |
| 5. It requires complicated sampling technique. | 5. It does not require complicated sampling technique. |

**One-Sample Test:**

**Run Test**

A run is defined as a sequence of letters of one kind bound by letters of the other kind. The number of run is denoted by r. A run test is used for testing the randomness of sequence of sample events on the basis of the order of sample events. The sequence of sample events may be defective (D) and non-defective (N), rise (R) and fall (F) of stream, high (H) and low (L) human blood pressure, success (S) and failure (F) etc.

**Case-I:** Small sample case when n1 and n2 20.

**Null hypothesis H0:** The sample observations are in random order.

**Alternative hypothesis H1:** The sample observations are not in random order. Two-tailed test)

**H1:** The sample observations are in few runs. (Left-tailed test)

**H1:** The sample observations are in much runs. (Right-tailed test)

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

r = No. of runs.

**Critical region:** Next for a pre-assigned level of significance and (n1, n2) we obtain from run table, two values of lower run () and upper run ().

**Decision:** (i) For two tailed test: If r lies between (), we accept H0. Otherwise reject H0.

(ii) For one tailed test: If r or r , we reject H0. Otherwise accept H0.

**Case-II:** Large sample case when n1 or n2 > 20.

In case of large sample size is approximately normally distributed with mean and variance.

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

Z =

In small sample case; other than = 5% then test statistic is given by

Z =

Where,

=

=

**Critical region:** Next for a pre-assigned level of significance, the probability (p0) is associated with the values as extreme as z, we obtained from z- table is

p0 = p(z ).

**Decision:** For one-tailed test, if p0 , we reject H0. Otherwise accept H0.

For two-tailed test, if2 p0 , we reject H0. Otherwise accept H0.

**Example (1):** A professional baseball team had the following sequence of wins (W) and losses (L) for the month of April:

W L WW L W LLL W LL WW L W LLL W L W L W LL W L W LL W

Does the win-loss record appear to be random?

**Solution:** Given sample observation is

W L WW L W LLL W LL WW L W LLL W L W L W LL W L W LL W

n1 = n (W) = 14

n2 = n (L) = 17

r = 23

Now,

**Null hypothesis H0:** The sample observations are in random order.

**Alternative hypothesis H1:** The sample observations are not in random order. Two-tailed test)

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

r = No. of runs = 23

**Critical region:** Next for a pre-assigned level of significance = 0.05 and (n1, n2) = (14, 17), we obtain from run table, two values of lower run = 10 and upper run = 23

**Decision:** Since r does not lies between (), we reject H0.

**Conclusion:** Therefore sample observations win-loss is not in random order.

**Example (2):** Records of depth of water in a “Alpha Lake” was regularity kept for 22 days during winter.

1.6, 1.5, 1.4, 1.5, 1.5, 1.5, 1.5, 1.6, 1.6, 1.7, 1.7, 1.5, 1.5, 1.3, 1.3, 1.6, 1.7, 1.8, 1.9, 1.9, 1.9, 2.0

Do the data provide sufficient evidence to conclude that the water level of the lake falls for the first half period and then rises for the second half period (too few runs are produced?

**Solution:** Given data are arranged in ascending order

1.3, 1.3, 1.4, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.6, 1.6, 1.6, 1.6, 1.7, 1.7, 1.7, 1.8, 1.9, 1.9, 1.9, 2.0

n = 22

Median (M d) = item value = item value = item value = = 1.6

Then, symbolic sequence is

1.6, 1.5, 1.4, 1.5, 1.5, 1.5, 1.5, 1.6, 1.6, 1.7, 1.7, 1.5, 1.5, 1.3, 1.3, 1.6, 1.7, 1.8, 1.9, 1.9, 1.9, 2.0

Effective symbols is

+ +

No. of + sign (n1) = n (-) = 10

No. of – sign (n2) = n (+) = 8

No. of runs (r) = 4

Now,

**Null hypothesis H0:** The rise and fall of water level are in random order.

**Alternative hypothesis H1:** The water level falls for the first half and rises for the second half period. (One-tailed test)

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

r = No. of runs = 4

**Critical region:** Next for a pre-assigned level of significance = 0.05/2 = 0.025 and (n1, n2) = (10, 8), we obtain from run table, two values of lower run = 5 and upper run = 15

**Decision:** Since r does not lies between (), we reject H0.

**Conclusion:** Therefore water level falls for the first half and rises for the second half period.

**Example (2):** Records of depth of water in a “Alpha Lake” was regularity kept for 22 days during winter.

1.6, 1.5, 1.4, 1.5, 1.5, 1.5, 1.5, 1.6, 1.6, 1.7, 1.7, 1.5, 1.5, 1.3, 1.3, 1.6, 1.7, 1.8, 1.9, 1.9, 1.9, 2.0

Do the data provide sufficient evidence to conclude that the water level of the lake falls for the first half period and then rises for the second half period at 10% level of significance (too few runs are produced?

**Solution:** Given data are arranged in ascending order

1.3, 1.3, 1.4, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.6, 1.6, 1.6, 1.6, 1.7, 1.7, 1.7, 1.8, 1.9, 1.9, 1.9, 2.0

n = 22

Median (M d) = item value = item value = item value = = 1.6

Then, symbolic sequence is

1.6, 1.5, 1.4, 1.5, 1.5, 1.5, 1.5, 1.6, 1.6, 1.7, 1.7, 1.5, 1.5, 1.3, 1.3, 1.6, 1.7, 1.8, 1.9, 1.9, 1.9, 2.0

Effective symbols is

+ +

No. of + sign (n1) = n (-) = 10

No. of – sign (n2) = n (+) = 8

No. of runs (r) = 4

= = = 9.89

= = = 2.03

Now,

**Null hypothesis H0:** The rise and fall of water level are in random order.

**Alternative hypothesis H1:** The water level falls for the first half and rises for the second half period. (One-tailed test)

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

**Z = = = 2.65**

**Critical region:** Next for a pre-assigned level of significance = 0.10, the probability (p0) is associated with the values as extreme as z, we obtained from z- table is

p0 = p (z ) = p (z ) = p (z ) = 0.0040

**Decision:** Since p0 < , we reject H0.

**Conclusion:** Therefore the sample observations are not in random order.

**Example (4):** A house officer of emergency department of a hospital wanted to test the hypothesis that emergency patients arrived randomly in his department. A record of the times (in minutes) between successive arrivals of patients during one evening was as follows.

3.5, 3.6, 1.1, 2.4, 14.3, 4.3, 4.2, 12.2, 10.5, 8.3, 11.1, 8.7, 10.8, 36.2, 12.9, 18.2, 10.0, 10.0, 10.0, 10.0, 10.0, 10.2, 5.0, 4.8, 8.3, 3.6, 3.8, 16.1, 2.5, 13.2, 5.6, 3.0, 10.2, 12.6, 6.8, 10.3, 15.4, 17.5, 17.6, 17.8, 10.2, 10.2

Does the data highlight random arrival of the emergency patients?

**Solution:** Given observations are arranged in ascending order

1.1, 2.4, 2.5, 3.0, 3.5, 3.6, 3.6, 3.8, 4.2, 4.3, 4.8, 5.6, 6.8, 8.3, 8.3, 8.7, 10.0, 10.0, 10.0, 10.0, 10.0, 10.2, 10.2, 10.2, 10.2, 10.3, 10.5, 10.8, 11.1, 12.2, 12.6, 12.9, 13.2, 14.3, 15.4, 16.1, 17.5, 17.6, 17.8, 18.2, 25.0, 36.2

n = 42

Median (M d) = item value = item value = item value = = 10.1

Then, symbolic sequence is

3.5, 3.6, 1.1, 2.4, 14.3, 4.3, 4.2, 12.2, 10.5, 8.3, 11.1, 8.7, 10.8, 36.2, 12.9, 18.2, 10.0, 10.0, 10.0, 10.0, 10.0,

10.2, 5.0, 4.8, 8.3, 3.6, 3.8, 16.1, 2.5, 13.2, 5.6, 3.0, 10.2, 12.6, 6.8, 10.3, 15.4, 17.5, 17.6, 17.8, 10.2, 10.2

No. of + sign (n1) = 20

No. of – sign (n2) = 22

No. of runs (r) = 18

= = = 21.95

= = = = = 3.19

Now,

**Null hypothesis H0:** The sample observations are in random order.

**Alternative hypothesis H1:** The sample observations are not in random order. (Two-tailed test)

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

Z = = = - 1.24

**Critical region:** Next for a pre-assigned level of significance = 0.05, the probability (p0) is associated with the values as extreme as z, we obtained from z- table is

p0 = p (z ) = p (z ) = p (z ) = 0.1075

2p0 = 20.1075 = 0.215

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

**Conclusion:** Therefore the sample observations are in random order.

**Binomial Test**

A binomial test is a non-parametric test. It is used to test whether the binomial population has two distinct groups of equal no. of outcomes or not. It is preferred when the population under consideration is of dichotomous feature such as: smokers and non-smokers; male and female; vegetarians and non-vegetarians etc. It is used data measured in nominal and ordinal scales.

**Case I:** Small case when n25.

**Problem:** To test,

**Null hypothesis H0:** P = P0 = ½, the sample has been drawn from binomial population.

**Alternative hypothesis H1:** P P0. Two-tailed test.

**H1:** P P0. Right-tailed test.

**H1:** P P0. Left-tailed test.

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

X0 = Minimum of (n1 and n2).

**Critical region:** Next for a pre-assigned level of significance, the probability (p0) is associated with the values as extreme as X0, we obtained from binomial table is

p0 = p(X X0)

**Decision:** For one-tailed test, if p0 , we reject H0. Otherwise accept H0.

For two-tailed test, if2 p0 , we reject H0. Otherwise accept H0.

**Example-1:** A marketing manager wishes to determine if consumers equally favor lemon flavor and orange flavor in a ice cream. Out of 20 consumers interviewed, 12 expressed their preference for the lemon flavor and the remaining 8 preferred orange flavor. Do the data provide strong evidence that there is a difference in popularity between two flavors? Test at level of significance = 0.05.

**Solution:** Given,

n = 20

n1 = 12

n2 = 8

Now,

**Problem:** To test,

**Null hypothesis (H0):** P = ½, there is a no significance difference in popularity between two flavors.

**Alternative hypothesis (H1):** P . Two-tailed test.

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

x0 = Minimum of (n1 and n2) = 8

**Critical region:** Next for a pre-assigned level of significance = 0.05 and n = 20, the probability (p0) is associated with the values as extreme as X0, we obtained from binomial table is

p0 = p(X X0) = p(X 8) = 0.252

2p0 = 20.252 = 0.504

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

**Conclusion:** There is no significance difference in popularity between two flavors.

**Example-2:** A coin is tossed 23 times and results obtained are given below

THHHTHHTHHTHHHHTTTHHTHHTTH

Test whether the coin is biased or not, using binomial test.

**Solution:** Given observation is

THHTHHTHHTHHHTTHHTHHTHT

n = 23

n1 = 14

n2 = 9

Now,

**Problem:** To test,

**Null hypothesis (H0):** P = P0 = ½, there is a no significance difference between probable of H and T.

**Alternative hypothesis (H1):** P P0. Two-tailed test.

**Test statistic:** Under H0, test statistic is x0 = Minimum of (n1 and n2) = 9

**Critical region:** Next for a pre-assigned level of significance = 0.05 and n = 23, the probability (p0) is associated with the values as extreme as X0, we obtained from binomial table is

p0 = p(X X0) = p(X 9) = 0.202

2p0 = 20.202 = 0.404

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

**Conclusion:** Therefore equal probable of H and T in the toss of coin.

**Case II:** Large sample case when n 25.

**Test statistic:** When n 25, the sampling distribution of X0 is approximately normal distribution with mean = n p and variance = n p q, then test statistic is

Z =; use +0.5 if X0 and – 0.5 if X0 n P.

**Critical region:** Next for a pre-assigned level of significance, the probability (p0) is associated with the values as extreme as z, we obtained from z- table is

p0 = p (z ).

**Decision:** For one-tailed test, if p0 , reject H0. Otherwise accept H0.

For two-tailed test, if2 p0 , we reject H0. Otherwise accept H0.

**Example-3:** There is a commonly held belief in certain communities that more boys than girls are born on a full moon day. A curious statistician collected births records from several maternity hospitals. He could locate 26 records of births on a full moon day. Of these 17 were boys. Would this evidence be considered enough to suggest that there was “something” to the popular belief?

**Solution:** Given,

n = 26

n1 = 17

n2 = 9

= Minimum of (n1 or n2) = 9

n P = 260.5 = 13

Now,

**Problem:** To test,

**Null hypothesis H0:** P = ½, boys and girls are equal born.

**Alternative hypothesis H1:** P > , boys born more than girls born.

**Test statistic:** Under H0; test statistics is

Z =

=; since x0 < n p.

=

=

= - 1.37

**Critical region:** Next for a pre-assigned level of significance = 0.05, the probability (p0) is associated with the values as extreme as z, we obtained from z- table is

p0 = p (z ) = p (z ) = 0.0853

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

**Conclusion:** Therefore boys and girls are equal probable of born.