**Wilcoxon Signed Rank test (one-sample)**

**Introduction:** The sign test is based on the signs ignoring their magnitude. But Wilcoxon Signed Rank test takes sign as well as its magnitude into consideration.This test is more powerful and sensitive than ordinary sign test.

**Procedure:**

Let x1,x2,…..xn be a random sample drawn from a population of size n and let M be the median of the given distribution (or) for the given observations

**H0**: M = M0 (The median divides the distribution equally)

**H1**: M ≠ M0 (or) M > M0 (or) M< M0

[Any one of these 3 conditions depends on the given problem]

**Level of significance**:

Appropriate level of significance is α% (given/chosen)

**Test Statistic:**

*Step1: Calculate di=xi-M0, i=1,2,…n*

*Step2: Arrange calculated di’s in ascending order irrespective of the sign.*

*If any value of di is zero, leave or omit that observation and then decide n.*

*Step3: Calculate the absolute value of dii.e |di|*

*Step4: Rank the values of |di| after discarding zero differences*

*Step5: Compute T+=Sum of the ranks of the +ve differences*

*Compute T‑=Sum of the ranks of the -ve differences*

*Step6: Find, T=minimum(T+,T-)*

**Inference**:

1. If it is a two tailed test (≠): Accept H0, if T>Tα/2,Otherwise reject H0.
2. If it is a right tailed test (>): Accept H0, if T>Tα,Otherwisereject H0.
3. If it is a left tailed test (<):Accept H0, if T<-Tα,Otherwisereject H0.

**Problems on Wilcoxon Signed Rank test**

1. If a random sample taken at public play ground, the time required to play a tennis match by players are: 36, 29, 44, 28, 40, 50, 39, 47, 33mins.

Use Wilcoxon signed rank test and test at 5% level of significance whether the average time required to play a match is 35 mins at this public play ground.

**H0**: µ = 35

**H1**: µ≠35

**Level of significance**:

Appropriate level of significance is α=5% (given)

**Test Statistic:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | di=xi-35 | Ascending order of di irrespective of sign | |di| | Ranks given to  |di| | T+ | T- |
| 36 | 1 | ***1*** | 1 | 1 | 1 |  |
| 29 | -6 | ***-2*** | 2 | 2 |  | 2 |
| 44 | 9 | ***4*** | 4 | 3 | 3 |  |
| 28 | -7 | ***5*** | 5 | 4 | 4 |  |
| 40 | 5 | ***-6*** | 6 | 5 |  | 5 |
| 50 | 15 | ***-7*** | 7 | 6 |  | 6 |
| 39 | 4 | ***9*** | 9 | 7 | 7 |  |
| 47 | 12 | ***12*** | 12 | 8 | 8 |  |
| 33 | -2 | ***15*** | 15 | 9 | 9 |  |
|  |  |  |  |  | **32** | **13** |

Therefore n=9, T=minimum(T+,T-)= minimum(32,13) = **13**

**Inference**:

Being a two tailed test (≠) ,the table value of T from Wilcoxon signed rank test at 5% level of significance for sample number n=9 is Tα/2= 5.

Here, T > Tα/2, i.e 13>5, We accept H0.

Hence we conclude that the average time required to play a match is 35mins.

1. The following data gives a random sample of 15 measurements of Octane ratings of certain kind of Gasoline

97.5, 95.2, 97.3, 96, 96.8, 100.3, 97.4, 95.3, 93.2, 99.1, 96.1, 97.6, 98.2, 98.5, 94.9.

Test whether the median is 98.5 or not using Wilcoxon signed rank test.

**H0**: M0 = 98.5

**H1**: M0≠98.5

**Level of significance**:

Appropriate level of significance is α=5% (chosen)

**Test Statistic:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | di=xi-98.5 | Ascending order of di irrespective of sign | |di| | Ranks given to  |di| | T+ | T- |
| 97.5 | -1 | ***-0.3*** | 0.3 | 1 |  | 1 |
| 95.2 | -3.3 | ***0.6*** | 0.6 | 2 | 2 |  |
| 97.3 | -1.2 | ***-0.9*** | 0.9 | 3 |  | 3 |
| 96 | -2.5 | ***-1*** | 1 | 4 |  | 4 |
| 96.8 | -1.7 | ***-1.1*** | 1.1 | 5 |  | 5 |
| 100.3 | 1.8 | ***-1.2*** | 1.2 | 6 |  | 6 |
| 97.4 | -1.1 | ***-1.7*** | 1.7 | 7 |  | 7 |
| 95.3 | -3.2 | ***1.8*** | 1.8 | 8 | 8 |  |
| 93.2 | -5.3 | ***-2.4*** | 2.4 | 9 |  | 9 |
| 99.1 | 0.6 | ***-2.5*** | 2.5 | 10 |  | 10 |
| 96.1 | -2.4 | ***-3.2*** | 3.2 | 11 |  | 11 |
| 97.6 | -0.9 | ***-3.3*** | 3.3 | 12 |  | 12 |
| 98.2 | -0.3 | ***-3.6*** | 3.6 | 13 |  | 13 |
| 98.5 | 0 | ***-5.3*** | 5.3 | 14 |  | 14 |
| 94.9 | -3.6 |  |  |  |  |  |
|  |  |  |  |  | **10** | **95** |

Therefore n=14 (after omitting one observation because one di=0),

T=minimum(T+,T-)= minimum(10,95) = **10**

**Inference**:

Being a two tailed test (≠) ,the table value of T from Wilcoxon signed rank test at 5% level of significance for sample number n=14 is Tα/2 = **21**.

Here, T < Tα/2, i.e 10<21, We reject H0.

Hence we conclude that the median is not equal to 98.5.

1. A random sample of size 15 infants of one month shows the following pulse rates (beats/min):**119, 120, 125, 122, 118, 117, 126, 114, 115, 123, 121, 120, 124, 127, 126**

Is there an evidence to suggest that the median pulse rate of one month infants is different from 120 beats/min.

**H0**: M0 = 120

**H1**: M0≠120

**Level of significance**:

Appropriate level of significance is α=5% (chosen)

**Test Statistic:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | di=xi-120 | Ascending order of di irrespective of sign | |di| | Ranks given to  |di| | T+ | T- |
| 119 | -1 | ***-1*** | 1 | 1.5 |  | 1.5 |
| 120 | 0 | ***1*** | 1 | 1.5 | 1.5 |  |
| 125 | 5 | ***-2*** | 2 | 3.5 |  | 3.5 |
| 122 | 2 | ***2*** | 2 | 3.5 | 3.5 |  |
| 118 | -2 | ***-3*** | 3 | 5.5 |  | 5.5 |
| 117 | -3 | ***3*** | 3 | 5.5 | 5.5 |  |
| 126 | 6 | ***4*** | 4 | 7 | 7 |  |
| 114 | -6 | ***-5*** | 5 | 8.5 |  | 8.5 |
| 115 | -5 | ***5*** | 5 | 8.5 | 8.5 |  |
| 123 | 3 | ***-6*** | 6 | 11 |  | 11 |
| 121 | 1 | ***6*** | 6 | 11 | 11 |  |
| 120 | 0 | ***6*** | 6 | 11 | 11 |  |
| 124 | 4 | ***7*** | 7 | 13 | 13 |  |
| 127 | 7 |  |  |  |  |  |
| 126 | 6 |  |  |  |  |  |
|  |  |  |  |  | **61** | **30** |

Therefore n=13 (after omitting two observations because two di’s are 0),

T=minimum(T+,T-)= minimum(61,30) = 30

**Inference**:

Being a two tailed test (≠) ,the table value of T from Wilcoxon signed rank test at 5% level of significance for sample number n=13 is Tα/2 = **17**.

Here, T >Tα/2, i.e30>17, We acceptH0.

Hence we conclude that the median pulse rate of one month infants is different

from 120 beats/min.

**Problems on Wilcoxon Signed Rank test (two samples/paired sample test)**

1. In order to determine if smoking results increased heart activity, a random sample of size 20 smokers was taken. Their pulse rate of size 20 was taken. Their pulse rate(beats/min) before smoking and after smoking a certain brand of cigarettes are being measured and the result obtained are :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Before | 70 | 69 | 72 | 74 | 66 | 68 | 69 | 70 | 71 | 69 | 73 | 72 | 68 | 72 | 67 | 70 | 68 | 69 | 70 | 71 |
| After | 69 | 72 | 71 | 74 | 68 | 67 | 72 | 72 | 70 | 75 | 75 | 73 | 71 | 72 | 69 | 71 | 72 | 70 | 71 | 71 |

Test whether thereis a change in heart activity among smokers

**H0**: µx =µy

**H1**: µx≠ µy

**Level of significance**:

Appropriate level of significance is α=5% (chosen)

**Test Statistic:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| xi | yi | di=xi-yi | Ascending order of di irrespective of sign | |di| | Ranks given to  |di| | T+ | T- |
| 70 | 69 | 1 | ***-1*** | 1 | 4.5 |  | 4.5 |
| 69 | 72 | -3 | ***-1*** | 1 | 4.5 |  | 4.5 |
| 72 | 71 | 1 | ***-1*** | 1 | 4.5 |  | 4.5 |
| 74 | 74 | 0 | ***-1*** | 1 | 4.5 |  | 4.5 |
| 66 | 68 | -2 | ***1*** | 1 | 4.5 | 4.5 |  |
| 68 | 67 | 1 | ***1*** | 1 | 4.5 | 4.5 |  |
| 69 | 72 | -3 | ***1*** | 1 | 4.5 | 4.5 |  |
| 70 | 72 | -2 | ***1*** | 1 | 4.5 | 4.5 |  |
| 71 | 70 | 1 | ***-2*** | 2 | 10.5 |  | 10.5 |
| 69 | 75 | -6 | ***-2*** | 2 | 10.5 |  | 10.5 |
| 73 | 75 | -2 | ***-2*** | 2 | 10.5 |  | 10.5 |
| 72 | 73 | -1 | ***-2*** | 2 | 10.5 |  | 10.5 |
| 68 | 71 | -3 | ***-3*** | 3 | 14 |  | 14 |
| 72 | 72 | 0 | ***-3*** | 3 | 14 |  | 14 |
| 67 | 69 | -2 | ***-3*** | 3 | 14 |  | 14 |
| 70 | 71 | -1 | ***-4*** | 4 | 16 |  | 16 |
| 68 | 72 | -4 | ***-6*** | 6 | 17 |  | 17 |
| 69 | 70 | -1 |  |  |  |  |  |
| 70 | 71 | -1 |  |  |  |  |  |
| 71 | 71 | 0 |  |  |  |  |  |
|  |  |  |  |  |  | **18** | **135** |

Therefore n=17 (after omitting three observations because three di’s are 0),

T=minimum(T+,T-) = minimum(18,135) = **18**

**Inference**:

Being a two tailed test (≠) , the table value of T from Wilcoxon signed rank test at 5%

level of significance for sample number n=17 is T =35

Here, T <Tα/2, i.e 18<35, We reject H0.

Hence we conclude that there will be a change in the heart activity among smokers.

1. In a random sample of 10 days in two cities of Andhra Pradesh, the following temperatures are recorded:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| City A | 108 | 104 | 109 | 116 | 111 | 113 | 114 | 112 | 102 | 112 |
| City B | 105 | 109 | 110 | 109 | 106 | 112 | 108 | 107 | 104 | 108 |

Test whether the average of temperatures of two cities are same.

**H0**: µx =µy

**H1**: µx≠ µy

**Level of significance**:

Appropriate level of significance is α=5% (chosen)

**Test Statistic:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| xi | yi | di=xi- yi | Ascending order of di irrespective of sign | |di| | Ranks given to  |di| | T+ | T- |
| 108 | 105 | -3 | ***-1*** | 1 | 1.5 |  | 1.5 |
| 104 | 109 | -5 | ***1*** | 1 | 1.5 | 1.5 |  |
| 109 | 110 | -1 | ***-2*** | 2 | 3 |  | 3 |
| 116 | 109 | 7 | ***-3*** | 3 | 4 |  | 4 |
| 111 | 106 | 5 | ***4*** | 4 | 5 | 5 |  |
| 113 | 112 | 1 | ***-5*** | 5 | 7 |  | 7 |
| 114 | 108 | 6 | ***5*** | 5 | 7 | 7 |  |
| 112 | 107 | 5 | ***5*** | 5 | 7 | 7 |  |
| 102 | 104 | -2 | ***6*** | 6 | 9 | 9 |  |
| 112 | 108 | 4 | ***7*** | 7 | 10 | 10 |  |
|  |  |  |  |  |  | **39.5** | **15.5** |

Therefore n=10, T=minimum(T+,T-) = minimum(39.5,15.5) = **15.5**

**Inference**:

Being a two tailed test (≠) , the table value of T from Wilcoxon signed rank test at 5%

level of significance for sample number n=10 is T =8

Here, T >Tα/2, i.e 15.5> 8, We accept H0.

Hence we conclude that there is no significant different between the average temperatures in both the cities.

**Mann-Whitney Test (or) U-test**

**Procedure:**

**H0**: µx = µy

**H1**: µx≠µy (or) µx> µy (or) µx<µy

[Any one of these 3 conditions depends on the given problem]

**Level of significance**:

Appropriate level of significance is α% (given/chosen)

**Test Statistic:**

*Step1: Combine the samples and arrange all the values in ascending order*

*Step2: Assign ranks to all the arranged values.*

*Step3: i) Find the sum of the ranks for each of the original values in the samples*

*ii) Let us denote the sum of the ranks of the first sample as* R1

*iii) Let us denote the sum of the ranks of the second sample as* R2

*vi) Let us denote the number of observations in the first sample as* n1

*v) Let us denote the number of observations in the second sample as* n2

*Step4: Calculate*

*Calculate*

U=minimum{U1,U2}

**Inference**:

1. If it is a two tailed test (≠): Accept H0, if U>Uα/2,Otherwise reject H0.
2. If it is a right tailed test (>): Accept H0, if U>Uα,Otherwisereject H0.
3. If it is a left tailed test (<):Accept H0, if U<-Uα, Otherwisereject H0.

**Problems on Mann-Whitney test**

1. Nicotine content of two brands of cigarettes was found as follows:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BrandA | 2.1 | 4.0 | 6.3 | 5.4 | 4.8 | 3.7 | 6.1 | 3.3 |  |  |
| BrandB | 4.1 | 0.6 | 3.1 | 2.5 | 4.0 | 6.2 | 1.6 | 2.2 | 1.9 | 5.4 |

Test whether the distribution of nicotine in two brands is equal using U-test at 5% level of significance.

**H0: µx = µy**

**H1: µx≠ µy**

**Level of significance**:

Appropriate level of significance is α=5% (given)

**Test Statistic:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | yi | Combined sample | Ascending order of the values in the combined sample | Ranks | Ranks of the first sample | Ranks of the second sample |
| 2.1 | 4.1 | 2.1 | 0.6 | 1 | 4 | 12 |
| 4.0 | 0.6 | 4.0 | 1.6 | 2 | 10.5 | 1 |
| 6.3 | 3.1 | 6.3 | 1..9 | 3 | 18 | 7 |
| 5.4 | 2.5 | 5.4 | 2.1 | 4 | 14.5 | 6 |
| 4.8 | 4.0 | 4.8 | 2.2 | 5 | 13 | 10.5 |
| 3.7 | 6.2 | 3.7 | 2.5 | 6 | 9 | 17 |
| 6.1 | 1.6 | 6.1 | 3.1 | 7 | 16 | 2 |
| 3.3 | 2.2 | 3.3 | 3.3 | 8 | 8 | 5 |
|  | 1.9 | 4.1 | 3.7 | 9 |  | 3 |
|  | 5.4 | 0.6 | 4.0 | 10.5 |  | 14.5 |
|  |  | 3.1 | 4.0 | 10.5 |  |  |
|  |  | 2.5 | 4.1 | 12 |  |  |
|  |  | 4.0 | 4.8 | 13 |  |  |
|  |  | 6.2 | 5.4 | 14.5 |  |  |
|  |  | 1.6 | 5.4 | 14.5 |  |  |
|  |  | 2.2 | 6.1 | 16 |  |  |
|  |  | 1.9 | 6.2 | 17 |  |  |
|  |  | 5.4 | 6.3 | 18 |  |  |
|  |  |  |  |  | **93** | **78** |

Here, n1=8, n2=10, R1=93, R2=78

*Calculate*

*Calculate*

U=minimum{U1,U2}=minimum{23,57}=23

**Inference**:

Being a two tailed test (≠) ,the table value of U from Mann-Whitney test tables for n1=8, n2=10 at 5% level of significance is 17 i.eUα/2 = 17.

Here, U>Uα/2, i.e23>17, We accept H0.

Hence we conclude that the average nicotine in both the brands are same.

1. The following data represents the weight in Kgs of a personal luggage carried in an aircraft by the members of two baseball clubs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Club A | 34 | 39 | 41 | 28 | 33 |  |
| Club B | 36 | 40 | 35 | 31 | 39 | 36 |

Use α=0.05 to test the hypothesis that the two clubs carry same amount of the luggage.

**H0: µx = µy**

**H1: µx≠ µy**

**Level of significance**:

Appropriate level of significance is α=0.05 (given)

**Test Statistic:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | yi | Combined sample | Ascending order of the values in the combined sample | Ranks | Ranks of the first sample | Ranks of the second sample |
| 34 | 36 | 34 | 28 | 1 | 4 | 6.5 |
| 39 | 40 | 39 | 31 | 2 | 8.5 | 10 |
| 41 | 35 | 41 | 33 | 3 | 11 | 5 |
| 28 | 31 | 28 | 34 | 4 | 1 | 2 |
| 33 | 39 | 33 | 35 | 5 | 3 | 8.5 |
|  | 36 | 36 | 36 | 6.5 |  | 6.5 |
|  |  | 40 | 36 | 6.5 |  |  |
|  |  | 35 | 39 | 8.5 |  |  |
|  |  | 31 | 39 | 8.5 |  |  |
|  |  | 39 | 40 | 10 |  |  |
|  |  | 36 | 41 | 11 |  |  |
|  |  |  |  |  | **27.5** | 38.5 |

Here, n1=5, n2=6, R1=27.5, R2=38.5

*Calculate*

*Calculate*

U=minimum{U1,U2}=minimum{17.5,12.5}=12.5

**Inference**:

Being a two tailed test (≠) ,the table value of U from Mann-Whitney test tables for n1=5, n2=6 at 5% level of significance is 3i.e Uα/2 = 3.

Here, U>Uα/2, i.e12.5>3, We accept H0.

Hence we conclude that the average luggage carried by the members of the two baseball clubs are same.

1. Suppose that two drugs A and B are being compared. Suppose 3 people took drug A and 4 people took drug B, the number of minutes that took to relieve from the pain are given below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Drug A | 9 | 11 | 15 |  |
| Drug B | 6 | 8 | 10 | 13 |

Are the mean number of minutes that took to get relieved from the pain are same for both the drugs? Test at 1% level of significance.

**H0: µx = µy**

**H1: µx≠ µy**

**Level of significance**:

Appropriate level of significance is α=1% (given)

**Test Statistic:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | yi | Combined sample | Ascending order of the values in the combined sample | Ranks | Ranks of the first sample | Ranks of the second sample |
| 9 | 6 | 9 | 6 | 1 | 3 | 1 |
| 11 | 8 | 11 | 8 | 2 | 5 | 2 |
| 15 | 10 | 15 | 9 | 3 | 7 | 4 |
|  | 13 | 6 | 10 | 4 |  | 6 |
|  |  | 8 | 11 | 5 |  |  |
|  |  | 10 | 13 | 6 |  |  |
|  |  | 13 | 15 | 7 |  |  |
|  |  |  |  |  | **15** | **13** |

Here, n1=3, n2=4, R1=15, R2=13

*Calculate*

*Calculate*

U=minimum{U1,U2}=minimum{3,9}=3

**Inference**:

Being a two tailed test (≠) ,the table value of U from Mann-Whitney test tables for n1=3, n2=4 at 1% level of significance is 1i.e Uα/2 = 1.

Here, U>Uα/2, i.e3>1, We accept H0.

Hence we conclude that the mean number of minutes that took to get relieved from the pain are same for both the drugs.

1. Students in X-class were divided into two groups according to the methods of instruction. The following data gives the scores obtained by the two groups of students in the final examination. Test whether the new method is more advantageous at 5% level of significance

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| New method | 431 | 225 | 372 | 681 | 576 | 628 | 605 | 689 | 476 | 359 | 269 | 654 |  |
| Old method | 655 | 603 | 466 | 860 | 382 | 580 | 750 | 855 | 776 | 640 | 532 | 630 | 670 |

**Solution :H0: µx = µy H1: µx> µyLevel of significance**: α=5% (given)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | yi | Combined sample | Ascending order of the values in the combined sample | Ranks | Ranks of the first sample | Ranks of the second sample |
| 431 | 655 | 431 | 225 | 1 | 6 | 18 |
| 225 | 603 | 225 | 269 | 2 | 1 | 12 |
| 372 | 466 | 372 | 359 | 3 | 4 | 7 |
| 681 | 860 | 681 | 372 | 4 | 20 | 25 |
| 576 | 382 | 576 | 382 | 5 | 10 | 5 |
| 628 | 580 | 628 | 431 | 6 | 14 | 11 |
| 605 | 750 | 605 | 466 | 7 | 13 | 22 |
| 689 | 855 | 689 | 476 | 8 | 21 | 24 |
| 476 | 776 | 476 | 532 | 9 | 8 | 23 |
| 359 | 640 | 359 | 576 | 10 | 3 | 16 |
| 269 | 532 | 269 | 580 | 11 | 2 | 9 |
| 654 | 630 | 654 | 603 | 12 | 17 | 15 |
|  | 670 | 655 | 605 | 13 |  | 19 |
|  |  | 603 | 628 | 14 |  |  |
|  |  | 466 | 630 | 15 |  |  |
|  |  | 860 | 640 | 16 |  |  |
|  |  | 382 | 654 | 17 |  |  |
|  |  | 580 | 655 | 18 |  |  |
|  |  | 750 | 670 | 19 |  |  |
|  |  | 855 | 681 | 20 |  |  |
|  |  | 776 | 689 | 21 |  |  |
|  |  | 640 | 750 | 22 |  |  |
|  |  | 532 | 776 | 23 |  |  |
|  |  | 630 | 855 | 24 |  |  |
|  |  | 670 | 860 | 25 |  |  |
|  |  |  |  |  | **119** | **206** |

Here, n1=12, n2=13, R1=119, R2=206

*Calculate*

*Calculate*

U=minimum{U1,U2}=minimum{115,41}=41

**Inference**:

Being a two tailed test (≠) , the table value of U from Mann-Whitney test tables for n1=12, n2=13 at 1% level of significance is 1 i.e Uα =47.

Here, U < Uα, i.e41<47, We reject H0.Hence we conclude that the new method is more advantageous than the old method of instruction.

**Kolmogorov-Smirnov test (one-sample)**

**Procedure:**

**H0**: The data follows given distribution

**H1**: The data do not follow given distribution

**Level of significance**:

Appropriate level of significance is α% (given/chosen)

**Test Statistic:**

*Step1: Count the number of given observations n (say)*

*Step2: Arrange the given observations in ascending order without any repetitions say x(i)*

*Step3: Generate the following table*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **i** | **x(i)** | **F(x(i))** |  |  |  |  |
|  |  |  |  |  |  |  |

*Step4:*

*Now D= Max{D+,D-}*

**Inference**:

If calculated D < The table value of D at α% level of significance,

we accept H0. Otherwise reject H0.

1. The sequence of numbers 0.63, 0.49, 0.24, 0.57, 0.76, 0.89 have been generated. Use the Kolmogorov-Smirnov test to determine if the hypothesis that the numbers are uniformly distributed in the interval [0,1].

**H0**: The numbers are uniformly distributed

**H1**: The numbers are not uniformly distributed

**Level of significance**:

Appropriate level of significance is 5% (chosen)

**Test Statistic:**

*Step1: Count the number of given observations n=6*

*Step2: Arrange the given observations in ascending order without any repetitions say x(i)*

0.24, 0.49, 0.57, 0.63, 0.76, 0.89

*Step3: Generate the following table*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **i** | **x(i)** | **F(x(i))** |  |  |  |  |
| 1 | 0.24 | 0.24 | 1/6=0.1666 | 0/6=0.0000 | -0.0734 | 0.2400 |
| 2 | 0.49 | 0.49 | 2/6=0.3333 | 1/6=0.1666 | -0.1567 | 0.3234 |
| 3 | 0.57 | 0.57 | 3/6=0.5000 | 2/6=0.3333 | -0.0700 | 0.2367 |
| 4 | 0.63 | 0.63 | 4/6=0.6666 | 3/6=0.5000 | 0.0366 | 0.1300 |
| 5 | 0.76 | 0.76 | 5/6=0.8333 | 4/6=0.6666 | 0.0733 | 0.0934 |
| 6 | 0.89 | 0.89 | 6/6=1.0000 | 5/6=0.8333 | 0.1100 | 0.0567 |

*Step4: =****0.1100***

*=****0.3234***

*Now D= Max{D+,D-}=****0.3234***

**Inference**:

The table value of D for n=6 at 5% level of significance isD0.05=0.521.

Here, D<Dαwe accept H0

1. Ten points are taken in an interval of length one metre. The distance of each point from the start of the interval is (in meters) as follows:

0.414, 0.523, 0.229, 0.942, 0.097, 0.394, 0.572, 0.486, 0.273, 0.358

The points may be supposed to be chosen at random and independently of each other if and only if the 10 observations form a random sample from the uniform distribution over the interval [0,1]. Examine whether this is borne out by the data by using Kolmogorov-Smirnov test.

**H0**: The numbers are uniformly distributed

**H1**: The numbers are not uniformly distributed

**Level of significance**:

Appropriate level of significance is 5% (chosen)

**Test Statistic:**

*Step1: Count the number of given observations n=10*

*Step2: Arrange the given observations in ascending order without any repetitions say x(i)*

0.097, 0.229, 0.273, 0.358, 0.394, 0.414, 0.486, 0.523, 0.572, 0.942

*Step3: Generate the following table*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **i** | **x(i)** | **F(x(i))** |  |  |  |  |
| 1 | 0.097 | 0.097 | 0.1 | 0 | 0.003 | 0.097 |
| 2 | 0.229 | 0.229 | 0.2 | 0.1 | -0.029 | 0.129 |
| 3 | 0.273 | 0.273 | 0.3 | 0.2 | 0.027 | 0.073 |
| 4 | 0.358 | 0.358 | 0.4 | 0.3 | 0.042 | 0.058 |
| 5 | 0.394 | 0.394 | 0.5 | 0.4 | 0.106 | -0.006 |
| 6 | 0.414 | 0.414 | 0.6 | 0.5 | 0.186 | -0.086 |
| 7 | 0.486 | 0.486 | 0.7 | 0.6 | 0.214 | -0.114 |
| 8 | 0.523 | 0.523 | 0.8 | 0.7 | 0.277 | -0.177 |
| 9 | 0.572 | 0.572 | 0.9 | 0.8 | 0.328 | -0.228 |
| 10 | 0.942 | 0.942 | 1.0 | 0.9 | 0.058 | 0.042 |

*Step4: =0.328*

*=****0.129***

*Now D= Max{D+,D-}=****0.328***

**Inference**:

The table value of D for n=10 at 5% level of significance isD0.05 =0.410.

Here, D<Dα we accept H0.

1. It is desired to check whether pinholes in electrolytic tin plate are uniformly distributed across a plated coil on the basis of the following distances in inches of 10 pin holes form one edge of a long strip of tin plate 30 inches wide: 4.8, 14.8, 28.2, 23.1, 4.4, 28.7, 19.5, 2.4, 25.0, 6.2

Test the null hypothesis at the 0.05 level of significance

**H0**: The pinholes are uniformly distributed

**H1**: The pinholes are not uniformly distributed

**Level of significance**:

Appropriate level of significance is 0.05 (given)

**Test Statistic:**

*Step1: Count the number of given observations n=10*

*Step2: Arrange the given observations in ascending order without any repetitions say x(i)*

2.4, 4.4, 4.8, 6.2, 14.8, 19.5, 23.1, 25.0, 28.2, 28.7

*Step3: Generate the following table*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **i** | **x(i)** | **F(x(i))=** |  |  |  |  |
| 1 | 2.4 | 0.08 | 0.1 | 0 | 0.02 | 0.08 |
| 2 | 4.4 | 0.1466 | 0.2 | 0.1 | 0.0534 | 0.0466 |
| 3 | 4.8 | 0.16 | 0.3 | 0.2 | 0.14 | -0.04 |
| 4 | 6.2 | 0.2066 | 0.4 | 0.3 | 0.1934 | -0.0934 |
| 5 | 14.8 | 0.4933 | 0.5 | 0.4 | 0.0067 | 0.0933 |
| 6 | 19.5 | 0.65 | 0.6 | 0.5 | -0.05 | 0.15 |
| 7 | 23.1 | 0.77 | 0.7 | 0.6 | -0.07 | 0.17 |
| 8 | 25.0 | 0.8333 | 0.8 | 0.7 | -0.0333 | 0.1333 |
| 9 | 28.2 | 0.94 | 0.9 | 0.8 | -0.04 | 0.14 |
| 10 | 28.7 | 0.9566 | 1.0 | 0.9 | 0.0434 | 0.0566 |

*Step4: =0.1934*

*=****0.17***

*Now D= Max{D+,D-}=****0.1934***

**Inference**:

The table value of D for n=10 at 5% level of significance isD0.05 =0.410.

Here, D<Dα we accept H0 .

1. The following service times ( in mins) of a random sample of 15 customers were recorded at a bank. 7, 3, 2, 7, 5, 4, 9, 15, 6, 6, 1, 6, 8, 10, 12.

Test the hypothesis that the service times are exponentially distributed with mean time of 5 minutes.

**H0**: The data follows exponential distribution with mean 5 minutes i.e θ=5

**H1**: The data do not follow exponential distribution with mean 5 minutes θ≠5

**Level of significance**:

Appropriate level of significance is 5% (chosen)

**Test Statistic:**

*Step1: Count the number of given observations n=15*

*Step2: Arrange the given observations in ascending order without any repetitions say x(i)*

1,2,3,4,5,6,7,8,9,10,12,15

*Step3: Generate the following table*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **i** | **x(i)** | **F(x(i))=** |  |  |  |  |
| 1 | 1 | 0.9936 | 0.0666 | 0 | -0.9266 | 0.9936 |
| 2 | 2 | 0.9995 | 0.1333 | 0.0666 | -0.8662 | 0.9333 |
| 3 | 3 | 0.9999 | 0.2000 | 0.1333 | -0.7999 | 0.8666 |
| 4 | 4 | 0.9999 | 0.2666 | 0.2000 | -0.7333 | 0.7999 |
| 5 | 5 | 1 | 0.3333 | 0.2666 | -0.6667 | 0.7333 |
| 6 | 6 | 1 | 0.4000 | 0.3333 | -0.6000 | 0.6666 |
| 7 | 7 | 1 | 0.4666 | 0.4000 | -0.5334 | 0.6000 |
| 8 | 8 | 1 | 0.5333 | 0.4666 | -0.4667 | 0.5333 |
| 9 | 9 | 1 | 0.6000 | 0.5333 | -0.4000 | 0.4666 |
| 10 | 10 | 1 | 0.6666 | 0.6000 | -0.3334 | 0.4000 |
| 11 | 12 | 1 | 0.7333 | 0.6666 | -0.2666 | 0.3334 |
| 12 | 15 | 1 | 0.8000 | 0.7333 | -0.2000 | 0.2666 |

*Step4: =-0.2000*

*=****0.9936***

*Now D= Max{D+,D-}=****0.9936***

**Inference**:

The table value of D for n=15 at 5% level of significance isD0.05 =0.338

Here, D>Dα we reject H0 .

1. The following are 15 measurements of the boiling point of a silicon compound ( in degrees celsius) :

166, 141, 136, 153, 170, 162, 155, 146, 183, 157, 148, 132, 160, 175, 150.

Use the Kolmogorov-Smirnov test at the 0.01 level of significance to test the null hypothesis that the boiling points come from a normal population with µ=160 degrees celsius and σ=10 degrees celsius.

**H0**: The data follows normal distribution with µ=160 and σ=10

**H1**: The data do not follow normal distribution with µ=160 and σ=10

**Level of significance**:

Appropriate level of significance is 0.01 (given)

**Test Statistic:**

*Step1: Count the number of given observations n=15*

*Step2: Arrange the given observations in ascending order without any repetitions say x(i)*

132, 136, 141, 146, 148, 150, 153, 155, 157, 160, 162, 166, 170, 175, 183

*Step3: Generate the following table*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **i** | **x(i)** | **F(z=)** |  |  |  |  |
| 1 | 132 | F(132-160/10)=F(-2.8)=0.0026 | 0.0666 | 0.0000 | 0.0640 | 0.0026 |
| 2 | 136 | F(136-160/10)=F(-2.4)=0.0082 | 0.1333 | 0.0666 | 0.1251 | -0.0584 |
| 3 | 141 | F(141-160/10)=F(-1.9)=0.0287 | 0.2000 | 0.1333 | 0.1713 | -0.1046 |
| 4 | 146 | F(146-160/10)=F(-1.4)=0.0808 | 0.2666 | 0.2000 | 0.1858 | -0.1192 |
| 5 | 148 | F(148-160/10)=F(-1.2)=0.1151 | 0.3333 | 0.2666 | 0.2182 | -0.1515 |
| 6 | 150 | F(150-160/10)=F(-1.0)=0.1587 | 0.4000 | 0.3333 | 0.2413 | -0.1746 |
| 7 | 153 | F(153-160/10)=F(-0.7)=0.2420 | 0.4666 | 0.4000 | 0.2246 | -0.1580 |
| 8 | 155 | F(155-160/10)=F(-0.5)=0.3085 | 0.5333 | 0.4666 | 0.2248 | -0.1581 |
| 9 | 157 | F(157-160/10)=F(-0.3)=0.3821 | 0.6000 | 0.5333 | 0.2179 | -0.1512 |
| 10 | 160 | F(160-160/10)=F(0)=0.5000 | 0.6666 | 0.6000 | 0.1666 | -0.1000 |
| 11 | 162 | F(162-160/10)=F(0.2)=0.5793 | 0.7333 | 0.6666 | 0.1540 | -0.0873 |
| 12 | 166 | F(166-160/10)=F(0.6)=0.7257 | 0.8000 | 0.7333 | 0.0743 | -0.0076 |
| 13 | 170 | F(170-160/10)=F(1.0)=0.8413 | 0.8666 | 0.8000 | 0.0253 | 0.0413 |
| 14 | 175 | F(175-160/10)=F(1.5)=0.9332 | 0.9333 | 0.8666 | 0.0001 | 0.0666 |
| 15 | 183 | F(183-160/10)=F(2.3)=0.9893 | 1.0000 | 0.9333 | 0.0107 | 0.0560 |

*Step4: =-0.2413*

*=0.0666*

*Now D= Max{D+,D-}=****0.2413***

**Inference**:

The table value of D for n=15 at 5% level of significance is D0.01 =0.404

Here, D<Dα we accept H0 .

**Kolmogorov-Smirnov test (two-samples)**

**Procedure:**

**H0**: f1(x)=f2(y) i.e the distributions are same

**H1**: f1(x)≠f2(y)

**Level of significance**:

Appropriate level of significance is α% (given/chosen)

**Test Statistic:**

*Step1: Combine the samples*

*Step2: Arrange all the values in ascending order without repetitions x(i) then count the observations say n*

*Step3: Calculate*

*Step4: Calculate*

*Step5: Generate the following table*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sample 1 | Sample 2 | Combined samples | Ascending order of the combined sample without repetitions  (ordered Statistics)  ***x(i)*** | ***Sm(x)*** | ***Tn(x)*** | ***Sm(x)- Tn(x)*** | ***Tn(x)-Sm(x)*** |
|  |  |  |  |  |  |  |  |

*Step6:*

**Inference**:

If calculated D < The table value of D at α% level of significance,

we accept H0. Otherwise reject H0.

1. The following data represent the life times of batteries of two different brands in hours.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Brand A | 40 | 30 | 40 | 45 | 55 | 30 |
| Brand B | 50 | 55 | 45 | 55 | 60 | 40 |

Test whether the two brands of the batteries are same by Kolmogorov -Smirnov test at 5% level of significance.

**H0**: µx=µy

**H1**: µxµy

**Level of significance**:

Appropriate level of significance is 5% (given)

**Test Statistic:**

*Step1: Combine the samples*

*Step2: Arrange all the values in ascending order without repetitions x(i) then count the observations, n=6*

*Step3: Calculate*

*Step4: Calculate*

*Step5: Generate the following table*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Brand A | Brand B | Combined samples | Ascending order of the combined sample without repetitions  (ordered Statistics)  ***x(i)*** | ***Sm(x)*** | ***Tn(x)*** | ***Sm(x)- Tn(x)*** | ***Tn(x)-Sm(x)*** |
| 40 | 50 | 40 | 30 | 2/6 | 0/6 | 2/6 | -2/6 |
| 30 | 55 | 30 | 40 | 4/6 | 1/6 | 3/6 | -3/6 |
| 40 | 45 | 40 | 45 | 5/6 | 2/6 | 3/6 | -3/6 |
| 45 | 55 | 45 | 50 | 5/6 | 3/6 | 2/6 | -2/6 |
| 55 | 60 | 55 | 55 | 6/6 | 5/6 | 1/6 | -1/6 |
| 30 | 40 | 30 | 60 | 6/6 | 6/6 | 0 | 0 |
|  |  | 50 |  |  |  |  |  |
|  |  | 55 |  |  |  |  |  |
|  |  | 45 |  |  |  |  |  |
|  |  | 55 |  |  |  |  |  |
|  |  | 60 |  |  |  |  |  |
|  |  | 40 |  |  |  |  |  |

*Step6:*

**Inference**:

The table value of D for n=6 at 5% level of significance isD0.05 =0.521. Here, D<Dα we accept H0

1. As a part of a training program, some trainees are instructed by two different methods and the scores obtained in the tests are :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Method A | 71 | 75 | 65 | 69 | 73 | 66 | 68 | 71 | 74 | 68 |
| Method B | 72 | 77 | 84 | 78 | 69 | 70 | 77 | 73 | 65 | 75 |

Test whether the two different training methods are significant or not by Kolmogorov -Smirnov test at1% level of significance.

**H0**: µx=µy**H1**: µxµy

**Level of significance**: Appropriate level of significance is 1% (given)

**Test Statistic:**

*Step1: Combine the samples*

*Step2: Arrange all the values in ascending order without repetitions x(i) then count the observations, n=13*

*Step3: Calculate*

*Step4: Calculate*

*Step5: Generate the following table*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Brand A | Brand B | Combined samples | Ascending order of the combined sample without repetitions  (ordered Statistics)  ***x(i)*** | ***Sm(x)*** | ***Tn(x)*** | ***Sm(x)- Tn(x)*** | ***Tn(x)-Sm(x)*** |
| 71 | 72 | 71 | 65 | 1/13 | 1/13 | 0 | 0 |
| 75 | 77 | 75 | 66 | 2/13 | 1/13 | 1/13 | -1/13 |
| 65 | 84 | 65 | 68 | 4/13 | 1/13 | 3/13 | -3/13 |
| 69 | 78 | 69 | 69 | 5/13 | 2/13 | 3/13 | -3/13 |
| 73 | 69 | 73 | 70 | 5/13 | 3/13 | 2/13 | -2/13 |
| 66 | 70 | 66 | 71 | 7/13 | 3/13 | 4/13 | -4/13 |
| 68 | 77 | 68 | 72 | 7/13 | 4/13 | 3/13 | -3/13 |
| 71 | 73 | 71 | 73 | 8/13 | 5/13 | 3/13 | -3/13 |
| 74 | 65 | 74 | 74 | 9/13 | 5/13 | 4/13 | -4/13 |
| 68 | 75 | 68 | 75 | 10/13 | 6/13 | 4/13 | -4/13 |
|  |  | 72 | 77 | 10/13 | 8/13 | 2/13 | -2/13 |
|  |  | 77 | 78 | 10/13 | 9/13 | 1/13 | -1/13 |
|  |  | 84 | 84 | 10/13 | 10/13 | 0 | 0 |
|  |  | 78 |  |  |  |  |  |
|  |  | 69 |  |  |  |  |  |
|  |  | 70 |  |  |  |  |  |
|  |  | 77 |  |  |  |  |  |
|  |  | 73 |  |  |  |  |  |
|  |  | 65 |  |  |  |  |  |
|  |  | 75 |  |  |  |  |  |

*Step6:*

**Inference**: The table value of D for n=13 at 1% level of significance is D0.01 =0.433.

Here, D<Dα we accept H0

1. The following are the no. of sales which a sample of 9 sales people of industrial chemicals in California and a sample of 6 sales people of industrial chemicals in Oregon made over a certain fixed period of time

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| California | 59 | 68 | 44 | 71 | 63 | 46 | 69 | 54 | 48 |
| Oregon | 50 | 36 | 62 | 52 | 70 | 41 |  | | |

Test whether the sales in the two states are distributed equally by using Kolmogorov -Smirnov test at 1% level of significance.

**H0**: µx=µy

**H1**: µxµy

**Level of significance**:Appropriate level of significance is 1% (given)

**Test Statistic:**

*Step1: Combine the samples*

*Step2: Arrange all the values in ascending order without repetitions x(i) then count the observations, n=15*

*Step3: Calculate*

*Step4: Calculate*

*Step5: Generate the following table*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| California | Oregon | Combined samples | Ascending order of the combined sample without repetitions  (ordered Statistics)  ***x(i)*** | ***Sm(x)*** | ***Tn(x)*** | ***Sm(x)- Tn(x)*** | ***Tn(x)-Sm(x)*** |
| 59 | 50 | 59 | 36 | 0/15 | 1/15 | -1/15 | 1/15 |
| 68 | 36 | 68 | 41 | 0/15 | 2/15 | -2/15 | 2/15 |
| 44 | 62 | 44 | 44 | 1/15 | 2/15 | -1/15 | 1/15 |
| 71 | 52 | 71 | 46 | 2/15 | 2/15 | 0 | 0 |
| 63 | 70 | 63 | 48 | 3/15 | 2/15 | 1/15 | -1/15 |
| 46 | 41 | 46 | 50 | 3/15 | 3/15 | 0 | 0 |
| 69 |  | 69 | 52 | 3/15 | 4/15 | -1/15 | 1/15 |
| 54 |  | 54 | 54 | 4/15 | 4/15 | 0 | 0 |
| 48 |  | 48 | 59 | 5/15 | 4/15 | 1/15 | -1/15 |
|  |  | 50 | 62 | 5/15 | 5/15 | 0 | 0 |
|  |  | 36 | 63 | 6/15 | 5/15 | 1/15 | -1/15 |
|  |  | 62 | 68 | 7/15 | 5/15 | 2/15 | -2/15 |
|  |  | 52 | 69 | 8/15 | 5/15 | 3/15 | -3/15 |
|  |  | 70 | 70 | 8/15 | 6/15 | 2/15 | -2/15 |
|  |  | 41 | 71 | 9/15 | 6/15 | 3/15 | -3/15 |

*Step6:*

**Inference**: The table value of D for n=15 at 1% level of significance is D0.01 =0.404.

Here, D<Dα we accept H0

1. To compare two kinds of bumper guards 6 of each kind were mounted on a certain kind of compact car . Then each car was run into a concrete wall at 5 miles per hour and the following are the cost of repairs in dollars.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Bumper guard 1 | 107 | 148 | 123 | 165 | 102 | 119 |
| Bumper guard 2 | 134 | 115 | 112 | 151 | 133 | 129 |

Test whether the two bumper gaurds are equally efficient at 5% level of significance by using Kolmogorov -Smirnov test at 5% level of significance.

**H0**: µx=µy

**H1**: µxµy

**Level of significance**:Appropriate level of significance is 1% (given)

**Test Statistic:**

*Step1: Combine the samples*

*Step2: Arrange all the values in ascending order without repetitions x(i) then count the observations, n=125*

*Step3: Calculate*

*Step4: Calculate*

*Step5: Generate the following table*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Bumper  Gaurd1 | Bumper  Gaurd2 | Combined samples | Ascending order of the combined sample without repetitions  (ordered Statistics)  ***x(i)*** | ***Sm(x)*** | ***Tn(x)*** | ***Sm(x)- Tn(x)*** | ***Tn(x)-Sm(x)*** |
| 107 | 134 | 107 | 102 | 1/12 | 0/12 | 1/12 | -1/12 |
| 148 | 115 | 148 | 107 | 2/12 | 0/12 | 2/12 | -2/12 |
| 123 | 112 | 123 | 112 | 2/12 | 1/12 | 1/12 | -1/12 |
| 165 | 151 | 165 | 115 | 2/12 | 2/12 | 0 | 0 |
| 102 | 133 | 102 | 119 | 3/12 | 2/12 | 1/12 | -1/12 |
| 119 | 129 | 119 | 123 | 4/12 | 2/12 | 2/12 | -2/12 |
|  |  | 134 | 129 | 4/12 | 3/12 | 1/12 | -1/12 |
|  |  | 115 | 133 | 4/12 | 4/12 | 0 | 0 |
|  |  | 112 | 134 | 4/12 | 5/12 | -1/12 | 1/12 |
|  |  | 151 | 148 | 5/12 | 5/12 | 0 | 0 |
|  |  | 133 | 151 | 5/12 | 6/12 | -1/12 | 1/12 |
|  |  | 129 | 165 | 6/12 | 6/12 | 0 | 0 |

*Step6:*

**Inference**: The table value of D for n=12 at 5% level of significance is D0.05 =0.375.

Here, D<Dα we accept H0