**NON-PARAMETRIC TEST**

**Kolmogorov-Smirnov two sample test**

Kolmogorov-Smirnov two sample test is a non-parametric test. It is used to test whether the two independent samples under consideration have been drawn from the similar or same population.

**Case-I:** Small sample case when n1 = n2 40 OR n1 and n2 20.

**Problem:** To test,

**Null hypothesis H0:** F(x) = F(y); the two independent random samples have been drawn from the same or identical populations.

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

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

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

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

D 0 = Maximum of

Where, F n (x) = = Observed cumulative frequency distribution function; k1 = observed cumulative frequencies of first sample and n1 = Sum of frequency of first sample.

F n (y) = = Expected cumulative frequency distribution function; k2 = expected cumulative frequencies of second sample and n2 = Sum of frequency of second sample.

**Critical region:** Next for a pre-assigned level of significance and corresponding sample size n1 and n2, we obtain from Kolmogorov-Smirnov table, the critical value of D is for two-tailed test and for one-tailed test.

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

**Case-II:** Large sample case when n1 = n2 40 OR n1 and n2 20.

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

(i) D 0= Maximum of; for two-tailed test.

(ii) ; for one-tailed test.

**Critical region:** Nest for a pre-assigned level of significance, we obtained critical value from.

(i) Kolmogorov-Smirnov table, = 1.36 ; for two-tailed.

(ii) Chi-square table, for one-tailed.

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

For one-tailed test, if; we reject H0. Otherwise accept H0.

**Example 1:** A weekly record of the National Safety Council reveals that the accidental deaths in certain geographical region had the following distribution according to the principal types of accidents.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Geographical | No. of deaths | | | |
| Area | Motor Vehicle | Poisons | Drowning Burns | Total |
| Remote | 5 | 4 | 1 | 10 |
| Urban | 3 | 2 | 5 | 10 |

Do these data show a significantly different pattern in the distribution of accidental deaths in the geographical regions? Test at = 0.05.

**Solution:** Here,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No. of death | Area | | K1 | K2 | F (x) | F (y) | |F (x) - F (y)| |
| Remote f 1 | Urban f 2 |
| Motor Vehicle | 5 | 3 | 5 | 3 | 0.5 | 0.3 | 0.2 |
| Poisons | 4 | 2 | 9 | 5 | 0.9 | 0.5 | 0.4 |
| Drowning Burns | 1 | 5 | 10 | 10 | 1 | 1 | 0 |
| Total | 10 | 10 |  |  |  |  |  |

= = n = 10

Now,

**Problem:** To test,

**Null hypothesis H0:** F(x) = F(y); the two independent random samples have been drawn from the same or identical populations.

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

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

D 0 = Maximum of = 0.4

**Critical region:** The critical value of D at = 0.05 and sample size n1 = 10 and n2 = 10 is = = = = 0.6

**Decision:** Since, D 0 < , we accept H0.

**Conclusion:**

**Example 2:** Tooth paste “Everest” and “Colgate” were tested by a group of chemicals and the method content in the brands was found to be as under:

|  |  |  |
| --- | --- | --- |
|  | Tooth paste | |
| Menthol contains | Everest | Colgate |
| Below 1 | 11 | 15 |
| 1 – 3 | 15 | 36 |
| 3 – 5 | 10 | 20 |
| 5 – 7 | 24 | 19 |

Do tooth paste “Colgate” contains higher level of menthol as compared to tooth paste “Everest”.

Solution: Here,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Tooth paste | |  |  |  |  |  |
| Menthol contains | Everest | Colgate | K1 | K2 | F(x) = | F (y) = | |F (x) - F (y)| |
| Below 1 | 11 | 15 | 11 | 15 | 0.1833 | 0.1667 | 0.0166 |
| 1 – 3 | 15 | 36 | 26 | 51 | 0.4333 | 0.5667 | 0.1334 |
| 3 – 5 | 10 | 20 | 36 | 71 | 0.6 | 0.7889 | 0.1889 |
| 7-May | 24 | 19 | 60 | 90 | 1 | 1 | 0 |
| Total | 60 | 90 |  |  |  |  |  |

= 60, = 90

D 0 = Maximum of = 0.1889

Now,

**Problem:** To test,

**Null hypothesis H0:** F(x) = F(y); there is no significance difference between tooth paste “Colgate” and tooth paste “Everest”.

**Alternative hypothesis H1:** F(x) > F(y); “Colgate” contains higher level of “Everest” tooth paste, one-tailed test.

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

= = 5.1384

**Critical region:** The critical value of chi-square at = 0.05 and 2 df. is = = 5.991

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

**Conclusion:** there is no significance difference between level of tooth paste “Colgate” and “Everest”.

**Chi-square test**

**Definition:** Chi-square test is used to test the significance difference between observed frequency () and expected frequency ().

**Applications:**

1. Chi-square test for goodness of fit.

2. Chi-square test for independence of attributes.

**Chi-Square Test for Goodness of Fit:**

Chi-square test is used to test the significance difference between observed frequency () and expected frequency ().

Steps of Chi-square test for goodness of fit

**Problem: T o test**

H0: = Ʃ; there is no significance difference between observed frequency and expected frequency i.e. Fitted (expected) frequency are good fit.

H1: Ʃ; there is significance difference between observed frequency and expected frequency i.e. Fitted (expected) frequency are not good fit.

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

=

Where,

= = Observed frequency

= = Expected frequency =

In case of probability distribution, then = Expected frequency = N P (X = x)

n = No. of classified class of group of frequency.

P (X = x) = for binomial distribution.

= for Poisson distribution.

**Critical region:** The tabulated value of chi-square at % level of significance and n – 1 – k 1 – k 2 degree of freedom is.

Where, n = No. of categorical frequency of class frequency.

= No. of expected frequency less than 5

= No. of calculated parameters of the probability distribution itself.

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

**Numerical problem**

**Example (1):** Suppose that after losing a large amount of money, an unlucky gambler questions whether the game was fair and the die was really unbiased. The last 90 times thrown of this die give the following results:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No. of dots on the die | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| No. of times is occurred | 20 | 15 | 12 | 17 | 9 | 17 | 90 |

**Solution:** Given,

n = No. of classified frequency = 6

|  |  |  |  |
| --- | --- | --- | --- |
| No. of dots on the die | No. of times occurred () | Expected frequency () |  |
| 1 | 20 | 15 | 1.667 |
| 2 | 15 | 15 | 0 |
| 3 | 12 | 15 | 0.6 |
| 4 | 17 | 15 | 0.267 |
| 5 | 9 | 15 | 2.4 |
| 6 | 17 | 15 | 0.267 |
|  | = 90 |  | = 5.201 |

Expected frequency (Ei) = = = 15

Now,

**Problem: T o test**

H0: = Ʃ; there is no significance difference between observed frequency () and expected frequency () i.e. Fitted (expected) frequency are good fit.

H1: Ʃ; there is significance difference between observed frequency () and expected frequency () i.e. Fitted (expected) frequency are not good fit.

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

= = 5.201

**Critical region:** The tabulated value of chi-square at = 0.05 level of significance and n – 1 = 6 – 1 = 5 degree of freedom is = = 11.070

**Decision:** If; we accept H0.

**Conclusion:** Therefore fitted frequency is good fit.

**Example (2):** Among 64 offspring’s of a certain cross between guinea pigs, 34 were red, 10 were black and 20 were white. According to the genetic model, these numbers should be in the ration 9: 3: 4. Are the data consistent with the model at 5 percent level?

**Solution:** Given,

Total ratio = 9 + 3 + 4 = 16

Expected frequency () =

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Colors | Observed frequency () | Expected frequency () | ( |  |  |
| Red | 34 | = 36 | -2 | 4 | 0.111 |
| Black | 10 | = 12 | -2 | 4 | 0.333 |
| White | 20 | = 16 | 4 | 16 | 1 |
|  | = 64 |  |  |  | = 1.444 |

n = 3

Now,

**Problem: T o test**

H0: = Ʃ; there is no significance difference between observed frequency () and expected frequency () i.e. Fitted (expected) frequency are good fit.

H1: Ʃ; there is significance difference between observed frequency () and expected frequency () i.e. Fitted (expected) frequency are not good fit.

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

= = 1.444

**Critical region:** The tabulated value of chi-square at = 0.05 level of significance and n – 1 = 3 – 1 = 2 degree of freedom is = = 5.991

**Decision:** If; we accept H0.

**Conclusion:** Therefore fitted frequency is good fit.

**Example (3):** Fit a binomial distribution to the following data, assuming that the coin is unbiased.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X = x | 0 | 1 | 2 | 3 | 4 |
| f | 28 | 62 | 46 | 10 | 4 |

Test the goodness of fit for above data at 1% level of significance.

**Solution:** Here

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | Observed frequency () or f | f X | P (X = x)= | Expected frequency = NP(X = x) |  |
| 0 | 28 | 0 | 0.2015 | 30 | 0.1333 |
| 1 | 62 | 62 | 0.3970 | 60 | 0.0667 |
| 2 | 46 | 92 | 0.2933 | 44 | 0.0909 |
| 3 | 10 | 30 | 0.0963 | 14 | = 0.25 |
| 4 | 4 | 16 | 0.0119 | 2 |  |
|  | N = 150 | = 200 | =1 | N = 150 | = 0.5409 |

Mean (x̅) = = = 1.33

In binomial distribution, we have

X = 0, 1, 2, 3, 4

n = 4

Mean (x̅) = 1.33

= 1.33

= 1.33

p = = 0.33

q = 1 – p = 1 – 0.33 = 0.67

**Now for chi-square test**

No. of categorical frequency (n) = 5

No. expected frequency less than 5 (k1) = 1

No. calculated parameter (k 2) = 1

Level of significance () = 1% = 0.01

Then,

**Problem: T o test**

H0: = Ʃ; there is no significance difference between observed frequency () and expected frequency () i.e. Binomial fitted (expected) frequency are good fit.

H1: Ʃ; there is significance difference between observed frequency () and expected frequency () i.e. Binomial fitted (expected) frequency are not good fit.

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

= = 0.5409

**Critical region:** The tabulated value of chi-square at = 0.05 level of significance and n – 1 – k1 – k 2 = 5 – 1 – 1 – 1 = 2 degree of freedom is = = 5.991

**Decision:** If; we accept H0.

**Conclusion:** Therefore binomial fitted frequency is good fit.

**Example (4):** The number of telephone calls received during the month of Frequency is summarized in the following table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. of calls received per days | 0 | 1 | 2 | 3 | 4 |
| No. of days | 9 | 12 | 5 | 4 | 1 |

Fit the Poisson distribution and test the no. of calls are equally received per days.

Solution: Here,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | f = | f x | P (X = x) = | Expected frequency () = NP (X =x) |  |
| 0 | 9 | 0 | = 0.2935 | 9.0985 = 9 | 0 |
| 1 | 12 | 12 | = 0.3598 | 11.1538 = 11 | 0.0909 |
| 2 | 5 | 10 | = 0.2205 | 6.8355 = 7 | 20.4546 |
| 3 | 4 | 12 | = 0.0901 | 2.7931 = 3 | - |
| 4 | 1 | 4 | = 0.0276 | 0.8556 = 1 | - |
|  | N = 31 | = 38 | = 0.9915 | N = 30.7365 = 31 | = 20.5455 |

Mean (x̅) = = = 1.2258

In Poisson distribution, we have

X = 0, 1, 2, 3, ----------, .

Mean = x̅

= 1.2258

**Now for chi-square test**

No. of categorical frequency (n) = 5

No. frequency less than 5 (k1) = 2

No. calculated parameter (k 2) = 1

Level of significance () = 5% = 0.05

Then,

**Problem: T o test**

H0: = Ʃ; there is no significance difference between observed frequency () and expected frequency () i.e. Poisson fitted (expected) frequency are good fit.

H1: Ʃ; there is significance difference between observed frequency () and expected frequency () i.e. Poisson fitted (expected) frequency are not good fit.

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

= = 20.5455

**Critical region:** The tabulated value of chi-square at = 0.05 level of significance and n – 1 – k1 – k 2 = 5 – 1 – 2 – 1 = 1 degree of freedom is = = 3.841

**Decision:** If; we reject H0.

**Conclusion:** Therefore Poisson fitted frequency is not good fit i.e. expected frequency are not equally distribution.

**Chi-square test for independence of attributes**

It is used to test the significance difference between two criteria of classification of attributes are independent or not i.e. two categorical variables are independent or not. The categorical variables are classified into some rows or some columns.

Steps of Chi-square test for independence of attributes

**Problem: T o test**

H0: = Ʃ; there is no significance difference between observed frequency and expected frequency i.e. Two categorical variables are independent.

H1: Ʃ; there is significance difference between observed frequency and expected frequency i.e. Two categorical variables are not independent.

**Test statistics:** (i)Under H0 test statistics is given by

=

**(ii) Alternative test statistic only in 2 2 contingency table is**

If any of cell frequency less than **5** in **2 2** contingency table, then use chi-square test statistics is given by

Where,

= Observed frequency for rows and columns.

= Expected frequency = =

r = No. of rows.

c = No. of columns.

N = Total frequency = a + b + c + d.

**Critical region:** The tabulated value of chi-square at % level of significance and (r – 1) (c – 1) degree of freedom is.

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

**Numerical problems**

**Example (1):** For the data in the following table, test for independence between a person’s ability in mathematics and interest in Economics. Use = 5%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Interest in Economics |  | Interest in Mathematics | | |
| Low | Average | High |
| Low | 63 | 42 | 15 |
| Average | 58 | 61 | 31 |
| High | 14 | 47 | 29 |

**Solution:** Here,

No. of row (r) = 3

No. of column (c) = 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Interest in Economics |  | Probability in Mathematics | | | Row total () |
| Low | Average | High |
| Low | 63 | 42 | 15 | 120 |
| Average | 58 | 61 | 31 | 150 |
| High | 14 | 47 | 29 | 90 |
| Column total () | | 135 | 150 | 75 | 360 |

Calculation table for Chi-square value

|  |  |  |
| --- | --- | --- |
| Observed frequency (O) | Expected frequency (E)  = |  |
| 63 | 45 | 7.2 |
| 42 | 50 | 1.28 |
| 15 | 25 | 4 |
| 58 | 56.25 | 0.05 |
| 61 | 62.5 | 0.04 |
| 31 | 31.25 | 0.00 |
| 14 | 33.75 | 11.56 |
| 47 | 37.5 | 2.41 |
| 29 | 18.75 | 5.60 |
| N = 360 | N = 360 | = 32.14 |

Now,

**Problem: T o test**

H0: = ƩE; there is no significance difference between observed frequency and expected frequency i.e. Probability of mathematics and interest in Economics are independent.

H1: ƩE; there is significance difference between observed frequency and expected frequency i.e. Probability of mathematics and interest in Economics are independent.

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

= = 32.14

**Critical region:** The tabulated value of chi-square at = 0.05 level of significance and (r – 1) (c – 1) = (3 – 1) (3 – 1) = 4 degree of freedom is = = 9.488

**Decision:** Since; we reject H0.

**Conclusion:** Therefore probability of mathematics and interest in Economics are not independent.

**Example (2):** Use chi-square test whether the color of sons, eyes is associated with that of the fathers at 5% level of significance using the data available in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sons, eye color | | Row total |
| Fathers, eye color | No light | Light |
| Not light | 230 | 148 | 378 |
| Light | 151 | 471 | 622 |
| Column total | 381 | 619 | 1000 |

**Solution:** Given,

No. of row (r) = 2

No. of column (c) = 2

Now,

**Problem: T o test**

H0: = Ʃ;; Sons, eye color is independent of fathers, eye color.

H1: Ʃ; Sons, eye color is not independent of fathers, eye color.

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

= = = 133.33

**Critical region:** The tabulated value of chi-square at = 0.05 level of significance and (r – 1) (c – 1) = (2 – 1) (2 – 1) = 1 degree of freedom is = = 3.841

**Decision:** Since; we reject H0.

**Conclusion:** Therefore sons, eye color is not independent of fathers, eye color.

**Example (3):** The following table reveals the condition of the house and the condition of the children.

|  |  |  |
| --- | --- | --- |
| Condition of Children | Condition of House | |
| Clean | Not Clean |
| Clean | 76 | 43 |
| Not Clean | 3 | 17 |

Using the Chi-square test, find out whether the condition of house affects the condition of children.

**Solution:** Given,

No. of row (r) = 2

No. of column (c) = 2

|  |  |  |  |
| --- | --- | --- | --- |
| Condition of Children | Condition of House | | Row Total |
| Clean | Not Clean |
| Clean | 76 | 44 | 120 |
| Not Clean | 3 | 17 | 20 |
| Column Total | 79 | 61 | 140 |

Now,

**Problem: T o test**

H0: = Ʃ;; the condition of house not affects the condition of children.

H1: Ʃ; the condition of house affects the condition of children.

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

= = = 14.38

**Critical region:** The tabulated value of chi-square at = 0.05 level of significance and (r – 1) (c – 1) = (2 – 1) (2 – 1) = 1 degree of freedom is = = 3.841

**Decision:** Since; we reject H0.

**Conclusion:** Therefore the condition of house affects the condition of children.

30.

|  |  |  |  |
| --- | --- | --- | --- |
| Influenza | Drug | | Row total |
| Administered | Not administered |
| Attact | 24 |  |  |
| Not attect |  | 12 |  |
| Column total | 76 |  | Total = 120 |