|  |
| --- |
| **Chi square Test** |
| **Chi – Square Distribution**  The sum of squares of n standard normal variables is defined to be a chi – square variable with *n* d.f.    **C:\Users\shabnam.ansari\Desktop\chi-square-graph.png**  The variable can be used to perform a variety of tests.  **Uses of**  **test / distribution**   1. **To test the independence of two or more attributes**:   test can be used to determine whether there is any association between two or more  attributes (characteristics) like color of eyes of mothers and daughters, heights of fathers  and sons etc.  In these tests, we proceed on the null hypothesis that the attributes are independent i.e.  there is *no association* between the attributes.    It follows - distribution with n-1 degrees of freedoms.   1. **Test Of Independence Of Attributes-Contingency Tables**   Consider two attributes A & B. Let A be divided into r classes A1, A2…  , Ar  and B be divided into s classes B1, B2…  , Bs.   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  | **A** | **A1** | **A2** | **A3** | **…..** | **Ar** | **Total** | | **B** |  |  |  |  |  |  |  | | **B1** |  | **(A1,B1)** | **(A2,B1)** | **(A3,B1)** |  | **(Ar,B1)** | **(B1)** | | **B2** |  | **(A1,B2)** | **(A2,B2)** | **(A3,B2)** |  | **(Ar,B2)** | **(B2)** | | **B3** |  |  |  |  |  |  |  | | **…** |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  | | **…** |  | **…** | **…** | **…** | **…** | **…** | **…** | | **Bs** |  | **(A1,Bs)** | **(A2,Bs)** | **(A3,Bs)** |  | **(Ar,Bs)** | **(Bs)** | | **Total** |  | **(A1)** | **(A2)** | **(A3)** |  | **(Ar)** | **N** |   Such a classification in which attributes are divided into more than two classes is known as *manifold classification*.  The various cell frequencies can be expressed in a table known as *r x s* manifold contingency table where (Ai) is the number of persons possessing attribute Ai, I = 1,2,…,r and (Bj) is the number of persons possessing attribute Bj, j = 1,2,…,s  Also  Where N is total frequency.  Now, the expected number of persons possessing both the attributes Ai Bj is given by:  (under H0) 𝐸𝑥𝑝𝑒𝑐𝑡𝑒𝑑 𝑓𝑟𝑒𝑞𝑢𝑒𝑛𝑐𝑦 = 𝑅𝑜𝑤 𝑡𝑜𝑡𝑎𝑙 ×𝑐𝑜𝑙𝑢𝑚𝑛 𝑡𝑜𝑡𝑎𝑙  𝑜𝑣𝑒𝑟𝑎𝑙𝑙 𝑡𝑜𝑡𝑎𝑙 (𝑁)  [The RHS can be obtained from the given table]  For large N,    Where  are the observed and expected frequencies respectively in the  cell. (r and s are the number of rows and columns in the table respectively.  If at  LOS, then we reject H0 and conclude that the attributes are not independent. Otherwise we have no reason to reject H0. We obtain  at  LOSfrom the chi-squared table, and it is obtained in the same manner as reading the t distribution. All the areas given in the table are the right tail. Since the question is about the attributes being dependent or independent, we don’t have to worry about one-tailed and two-tailed tests.  While testing the independence of attributes, the Null Hypothesis H0 will always be that the Attributes are independent. The Alternative Hypothesis will suggest that they are dependent, i.e. there is some relationship between them |
| 1. The following table gives the number of accidents in a city during a week. Find whether the accidents are uniformly distributed over a week.  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Day | Sun | Mon | Tues | Wed | Thu | Fri | Sat | Total | | No of accident | 13 | 15 | 9 | 11 | 12 | 10 | 14 | 84 | |
| 1. A die was thrown 132 times and the following frequencies were observed:  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | No obtained | 1 | 2 | 3 | 4 | 5 | 6 | Total | | Freq | 15 | 20 | 25 | 15 | 29 | 28 | 132 | |
| 1. The number of car accidents in a city was found to be 20,17,12,6,7,15,8,5,16 and 14 per month resp. Use chi square test to check whether these frequencies are in agreement with the belief that occurrence of accidents was the same during 10 months period. |
| 1. 300 digits were chosen at random from a table of random numbers. The frequencies was as follows:  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Digit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | | Freq | 28 | 29 | 33 | 31 | 26 | 35 | 32 | 30 | 31 | 25 |   Use chi square test to examine the hypothesis that the digit were distributed in equal number of tables. |
| 1. Theory predicts that the proportion of beans in the four groups A,B,C and D should be 9:3:3:1. In an experiment among 1600 beans the numbers in the four groups were 882,313,287 and 118. Does the experimental results support the theory? |
| 1. Investigate the association between the darkness of eye color in father and son from the following data:  |  |  |  |  |  | | --- | --- | --- | --- | --- | | Color of father’s eyes | | | | | | Color of Son’s eyes |  | Dark | Not dark | Total | | Dark | 48 | 90 | 138 | | Not dark | 80 | 782 | 862 | | Total | 128 | 872 | 1000 | |  |  |  |  |  | |
| 1. The figure given below A are the observed frequencies of a distribution and B the frequencies of the normal distribution having the same mean, SD and the total frequency as in A  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | A | 1 | 12 | 66 | 220 | 495 | 792 | 924 | 792 | 495 | 220 | 66 | 12 | 1 | | B | 2 | 15 | 66 | 210 | 484 | 799 | 943 | 799 | 484 | 210 | 66 | 15 | 2 |   Apply Chi square test of goodness of fit. |
| 1. The following data is collected on two characters. Based on this can you say that there is no relation between smoking and literacy?  |  |  |  | | --- | --- | --- | |  | Smokers | Nonsmokers | | Literates | 83 | 57 | | Illiterates | 45 | 68 | |