Chi Square Test

One of the primary tasks involved in any supervised Machine Learning venture is to select the best features from the given dataset to obtain the best results. One way to select these features is the Chi-Square Test.

Mathematically, a Chi-Square test is done on two distributions two determine the level of similarity of their respective variances. In its **null hypothesis**, it assumes that the given distributions are independent. This test thus can be used to determine the best features for a given dataset by determining the features on which the output class label is most dependent on. For each feature in the dataset, the value is calculated and then ordered in descending order according to the value. The higher the value, the more dependent the output label is on the feature and higher the importance the feature has on determining the output.

Code :

```cpp
#include <bits/stdc++.h>
using namespace std;
int main()
{
    float chi;
    int random[20];
    int s1,s2,mod,i=2;
    cout<<"Enter s1 :";
    cin>>s1;
    cout<<"Enter s2 : ";
    cin>>s2;
    cout<<"Enter modulus : ";
    cin>>mod;

    /// Generating Random numbers
    random[0] = (s1+s2)%mod;
    random[1] = (s2+random[0])%mod;
    for(i=2;i<20;i++)
        random[i] = (random[i-1] + random[i-2])%mod;

    int ctsize = ceil(mod/10);
    int ct[20];
    memset(ct,0,sizeof(ct));

    cout<<"Enter value of chi-square at alpha(0.5) : ";
    cin>>chi;
    for(i=0;i<20;i++)
    {
        if(random[i]%ctsize==0)
            random[i] = random[i]-1;
        ct[random[i]/ctsize + 1]++;
    }
    cout<<"The hypothesis value of random numbers generated is :    ";
    for(i=0;i<ctsize;i++)
        cout<<ct[i]<<" ";
    cout<<endl;

    float chisq  = 0;
    for(i=0;i<ctsize;i++)
```
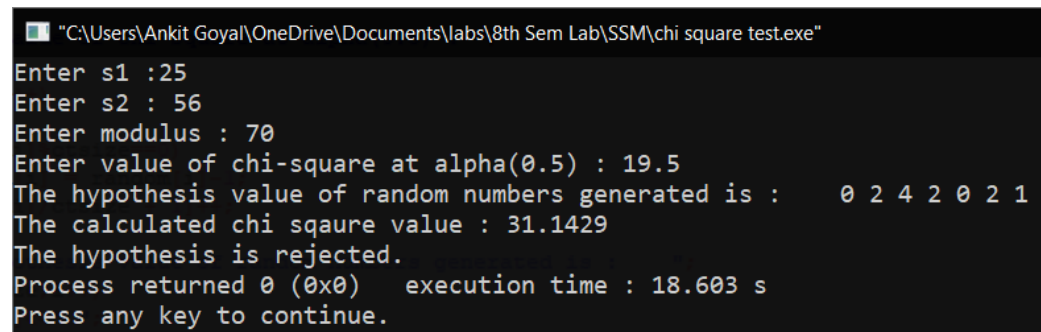
```
        {
            chisq += ((ctsize - (float)ct[i])*(ctsize - (float)ct[i]))/ctsize;
        }
        cout<<"The calculated chi sqaure value : "<<chisq<<"\n";
        if(chisq<chi)
            cout<<"The hypothesis is true: The numbers are uniformly distributed";
        else
            cout<<"The hypothesis is rejected.";

    }
```

```
■ "C:\Users\Ankit Goyal\OneDrive\Documents\labs\8th Sem Lab\SSM\chi square test.exe"
Enter s1 :25
Enter s2 : 56
Enter modulus : 70
Enter value of chi-square at alpha(0.5) : 19.5
The hypothesis value of random numbers generated is :    0 2 4 2 0 2 1
The calculated chi sqaure value : 31.1429
The hypothesis is rejected.
Process returned 0 (0x0)    execution time : 18.603 s
Press any key to continue.
```