

**Title:** Implementation of Chi-square Test for Uniformity testing

**Problem Statement:**

Write a program in java or macros in Excel to conduct frequency test & independence test., Generate 5 sets of random numbers using random number generator developed in previous experiment. Each set consisting of 100 random numbers perform chi square test of each set of random numbers.

**Expected Outcome of Experiment:**

| **Index** | **Outcome** |
| --- | --- |
| CO3 | Generate pseudorandom numbers and perform statistical tests to measure the quality of a pseudorandom number generator. |

**Books/ Journals/ Websites referred:**

1. Jerry Banks, John Carson, Barry Nelson, and David M. Nichol, “Discrete Event System Simulation”; Fifth Edition, Prentice-Hall.

2. Averill M Law, “System Modelling &amp; Analysis”; 4th Edition TMH.

3. Banks C M, Sokolowski J A, “Principles of Modelling and Simulation”, Wiley

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

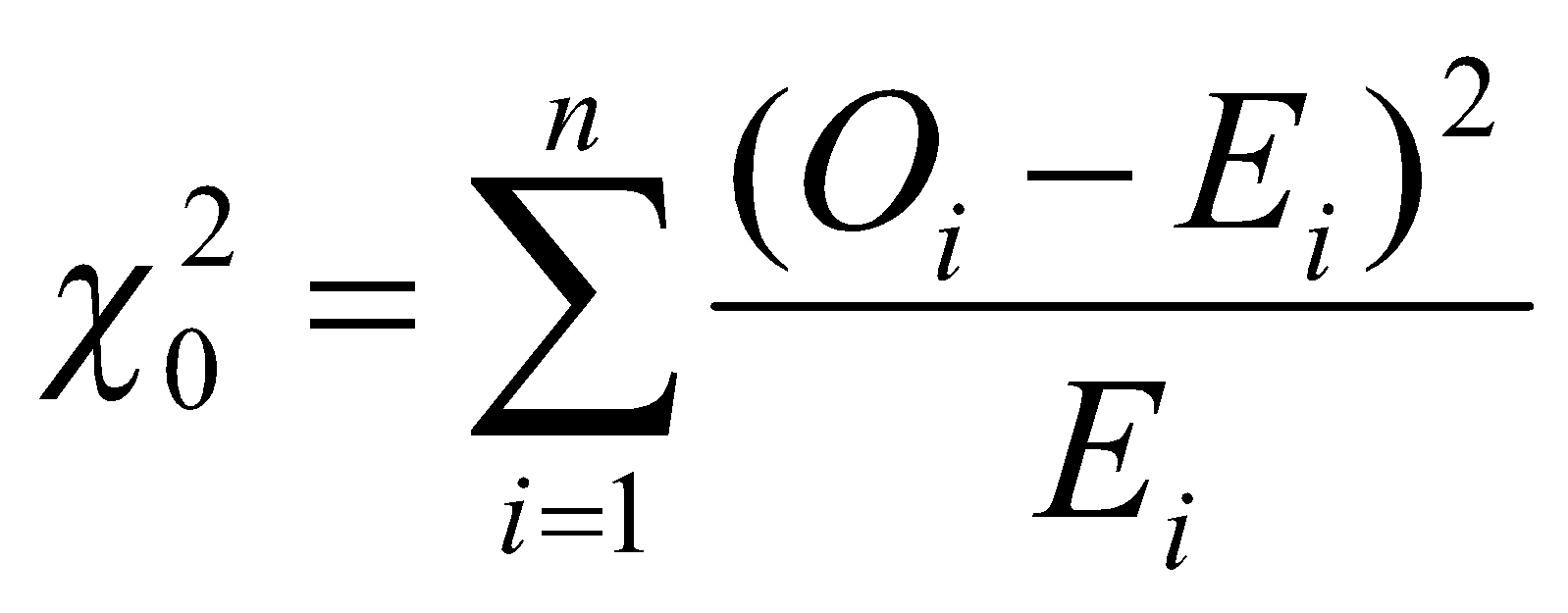
**Pre Lab/ Prior Concepts:**

Meaning of Uniformity and Independence Tests for Random Numbers:

**Frequency Tests** A basic test that should always be performed to validate a new generator is the test of uniformity. Two different methods of testing are available. They are the **Kolmogorov-Smirnov** and the **chi-square test**. Both of these tests measure the degree of agreement between the distribution of a sample of generated random numbers and the theoretical uniform distribution. Both tests are on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution*.*

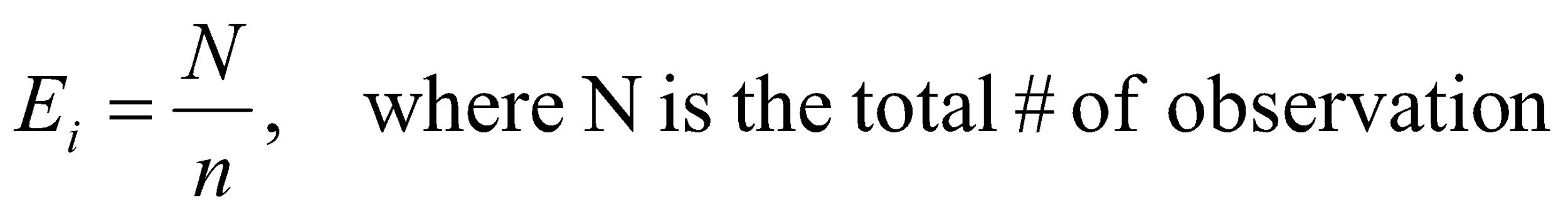
**The Chi Square test**:

Chi-square test uses the sample statistic:



Approximately the chi-square distribution with *n-1* degrees of freedom (where the critical values are tabulated in Table A.6)

For the uniform distribution, *Ei*, the expected number in the each class is:



Valid only for large samples, e.g. N >= 50

Reject *H0* if χ02 > χα,N-12

**Implementation details & Example**:

***Code***

import numpy as np

from random import random

import sortednp as snp

n=input("enter the value:")

n=int(n)

arr = [random() for \_ in range(int(n))]

print('The numbers are as follows\n')

for i in arr:

print(i,end="\n")

arr.sort()

arr= np.array(arr)

f = [None]\*n

for i in range(n):

f[i] = len(snp.intersect(arr[arr>i/n],arr[arr<(i+1)/n]))

print('Number Frequencies: ')

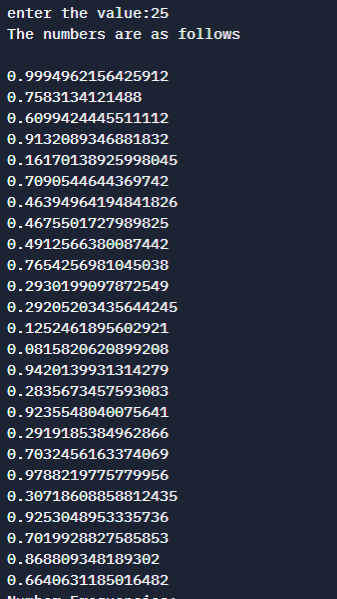
for i in range(n):

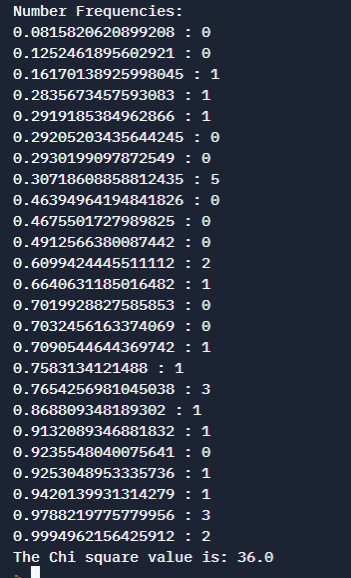
print(arr[i],":",f[i])

chi2 = sum([(f[i]-len(arr)/n)\*\*2/(len(arr)/n) for i in range(len(f))])

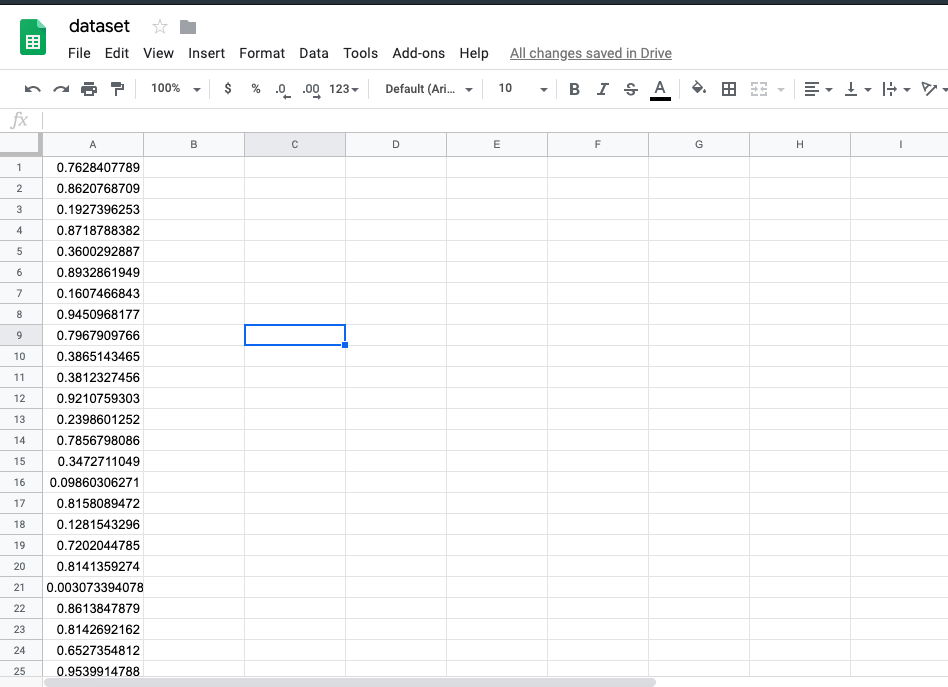
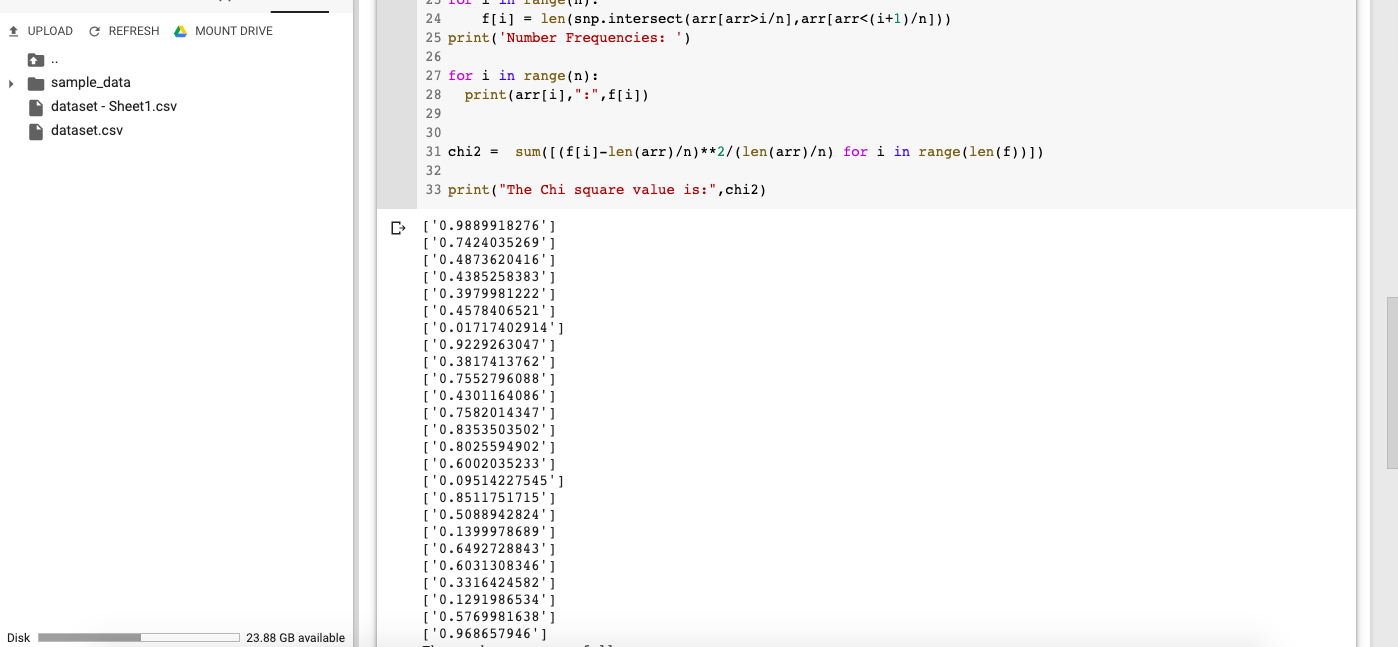
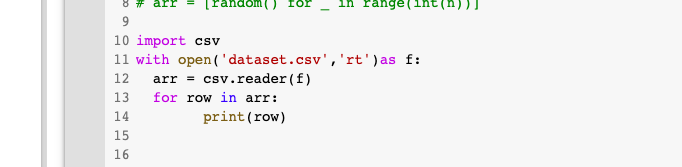
print("The Chi square value is:",chi2)

***Output***





***Dataset:***



**Conclusion:**

We successfully understood and performed the Implementation of Chi-square Test for Uniformity Testing of generated Random numbers in python.

**Post lab questions:**

1. Explain ‘T’ test, its distribution & limitations

* A t-test is an analysis of two populations through the use of statistical examination; a t-test with two samples is commonly used with small sample sizes, testing the difference between the samples when the variances of two normal distributions are not known.
* T-distribution is basically any continuous probability distribution that arises from an estimation of the mean of a normally distributed population using a small sample size and an unknown standard deviation for the population. The null hypothesis is the default assumption that no relationship exists between two different measured phenomena.
* Assumptions include:
* the scale of measurement. The assumption for a t-test is that the scale of measurement applied to the data collected follows a continuous or ordinal scale, such as the scores for an IQ test. (Investopedia)
* random sampling. The data is collected from a representative, randomly selected portion of the total population.
* data is normally distributed.
* the two populations have the same variance. This can be adjusted though.
* Limitations

The t and F tests for independent means only examine means, they have virtually nothing to say about individual scores. It is important to keep track of the fact that our conclusions are about means, not about individuals. So, if we run a study with gender as an independent variable, and get a statistically significant result, then we can say that the mean of males differs from the mean of females in terms of the dependent variable, but we cannot say that any particular male will have a higher or lower score than any particular female.