Experiment No.: 8
Title: Statistics using spreadsheet

Batch:A2 **Roll No.: 16010421073** **Experiment No.: 8**

Aim: **1.**To generate random numbers and draw samples from the data set using MS Excel
2. Hypothesis testing for mean

Resources needed: MS Excel

Theory

Problem Statement:
Generate random numbers using rand() / randbetween() / Data Analysis Toolpack and draw simple random samples from the dataset.

Concepts

Sample and Sampling

A Sample is a part of the total population. It can be an individual element or a group of elements selected from the population. Although it is a subset, it is representative of the population and suitable for research in terms of cost, convenience, andtime.

A good sample is one which satisfies all or few of the following conditions:

Representativeness: Good samples are those who accurately represent the population. On measurement terms, the sample must be valid. The validity of a sample depends upon its accuracy.

Accuracy: An accurate (unbiased) sample is one which exactly represents the population. It is free from any influence that causes any differences between sample value and population value.

Size: The sample size should be such that the inferences drawn from the sample are accurate to a given level of confidence to represent the entire population under study.

Sampling is the act, process, or technique of selecting a representative part of a population for the purpose of determining the characteristics of the whole population. Sampling is that part of statistical practice concerned with the selection of an unbiased or random subset of individual observations within a population of individuals intended to yield some knowledge about the population of concern, especially for the purposes of making predictions based on statistical inference. Sampling is an important aspect of data collection.

Population OR Universe: The entire aggregation of items from which samples can be drawn is known as a population. Population, contrary to its general notion as a nation's entire population has a much broader meaning in sampling. "N" represents the size of the population.

An operational sampling process can be divided into seven steps as given below:

1. Defining the target population.
2. Specifying the sampling frame.
3. Specifying the sampling unit.
4. Selection of the sampling method.
5. Determination of sample size.

6. Specifying the sampling plan.
7. Selecting the sample.

There are two basic approaches to sampling:

1. Probabilistic Sampling
2. Non-probabilistic sampling.

A Probabilistic sampling scheme is one in which every unit in the population has a chance (greater than zero) of being selected in the sample, and this probability can be accurately determined.

Types of Probabilistic Sampling

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Multistage cluster sampling

Non-probabilistic Sampling It involves the selection of units based on factors other than random chance. It is also known as deliberate sampling and purposive sampling.

Types of Non-Probabilistic Sampling

- Convenience sampling
- Quota sampling
- Judgment sampling
- Snowball sampling

Simple Random Sampling:

A sampling process where each element in the target population has an equal chance or probability of inclusion in the sample is known as Simple Random Sampling. For ex, if a sample of 15000 names is to be drawn from the telephone directory, then there is equal chance for each number in the directory to be selected. These numbers (serial no of name) could be randomly generated by the computer or picked out of a box. These numbers could be later matched with the corresponding names thus fulfilling the list. In small populations random sampling is done without replacement to avoid the instance of a unit being sampled more thanonce.

Hypothesis Testing for mean:

hypothesis test of a mean can be conducted, when the following conditions are met:

- The sampling method is simple random sampling.
- The sampling distribution is normal or nearly normal.

Generally, the sampling distribution will be approximately normally distributed if any of the following conditions apply.

- The population distribution is normal.
- The population distribution is symmetric, unimodal, without outliers, and the sample size is 15 or less.
- The population distribution is moderately skewed, unimodal, without outliers, and the sample size is between 16 and 40.
- The sample size is greater than 40, without outliers.

This approach consists of four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results.

State the Hypotheses

Every hypothesis test requires the analyst to state a null hypothesis and an alternative hypothesis. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

The table below shows three sets of hypotheses. Each makes a statement about how the population mean μ is related to a specified value M. (In the table, the symbol ≠ means " not equal to ".)

Set	Null hypothesis	Alternative hypothesis	Number of tails
1	$\mu = M$	$\mu \neq M$	2
2	$\mu >= M$	$\mu < M$	1
3	$\mu \leq M$	$\mu > M$	1

The first set of hypotheses (Set 1) is an example of a two-tailed test, since an extreme value on either side of the sampling distribution would cause a researcher to reject the null hypothesis. The other two sets of hypotheses (Sets 2 and 3) are one-tailed tests, since an extreme value on only one side of the sampling distribution would cause a researcher to reject the null hypothesis.

Formulate an Analysis Plan

The analysis plan describes how to use sample data to accept or reject the null hypothesis. It should specify the following elements.

- Significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.
- Test method. Use the one-sample t-test to determine whether the hypothesized mean differs significantly from the observed sample mean.

Analyze Sample Data

Using sample data, conduct a one-sample t-test. This involves finding the standard error, degrees of freedom, test statistic, and the P-value associated with the test statistic.

- Standard error. Compute the standard error (SE) of the samplingdistribution.

  $$SE = s * sqrt\{ (1/n) * [(N - n) / (N - 1)] \}$$

  where s is the standard deviation of the sample, N is the population size, and n is the sample size. When the population size is much larger (at least 20 times larger) than the sample size, the standard error can be approximated by:

  $$SE = s / sqrt(n)$$

- Degrees of freedom. The degrees of freedom (DF) is equal to the sample size (n) minus one. Thus, DF = n - 1.

- Test statistic. The test statistic is a t statistic (t) defined by the following equation.

  $$t = (x - \mu) / SE$$

  where x is the sample mean, μ is the hypothesized population mean in the null hypothesis, and SE is the standard error.

- P-value. The P-value is the probability of observing a sample statistic as extreme as the test statistic. Since the test statistic is a t statistic, use the t Distribution Calculator to assess the probability associated with the t statistic, given the degrees of freedom computedabove.

Interpret Results

If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the significance level, and rejecting the null hypothesis when the P-value is less than the significance level.

Test Your Understanding

Two sample problems illustrate how to conduct a hypothesis test of a mean score. The first problem involves a two-tailed test; the second problem, a one-tailed test.

Problem 1: Two-Tailed Test

An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously for 5 hours (300 minutes) on a single gallon of regular gasoline. From his stock of 2000 engines, the inventor selects a simple random sample of 50 engines for testing. The engines run for an average of 295 minutes, with a standard deviation of 20 minutes. Test the null hypothesis that the mean run time is 300 minutes against the alternative hypothesis that the mean run time is not 300 minutes. Use a 0.05 level of significance. (Assume that run times for the population of engines are normally distributed.)

Solution: The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

- State the hypotheses. The first step is to state the null hypothesis and an alternative hypothesis.

  Nul hypothesis: $\mu = 300$

  Alternative hypothesis: $\mu \neq 300$

  Note that these hypotheses constitute a two-tailed test. The null hypothesis will be rejected if the sample mean is too big or if it is too small.

- Formulate an analysis plan. For this analysis, the significance level is 0.05. The test method is a one-sample t-test.

- Analyze sample data. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

  $$SE = s / sqrt(n) = 20 / sqrt(50) = 20/7.07 = 2.83$$

  $$DF = n - 1 = 50 - 1 = 49$$

  $$t = (x - \mu) / SE = (295 - 300)/2.83 = -1.77$$

  where s is the standard deviation of the sample, x is the sample mean, μ is the hypothesized population mean, and n is the sample size.

  Since we have a two-tailed test, the P-value is the probability that the t statistic having 49 degrees of freedom is less than -1.77 or greater than 1.77.

  We use the t Distribution Calculator to find $P(t < -1.77) = 0.04$, and $P(t > 1.77) = 0.04$. Thus, the P-value = 0.04 + 0.04 = 0.08.

- Interpret results. Since the P-value (0.08) is greater than the significance level (0.05), we cannot reject the null hypothesis.

Note: If you use this approach on an exam, you may also want to mention why this approach is appropriate. Specifically, the approach is appropriate because the sampling method was simple random sampling, the population was normally distributed, and the sample size was small relative to the population size (less than 5%).

Problem 2: One-Tailed Test

Bon Air Elementary School has 1000 students. The principal of the school thinks that the average IQ of students at Bon Air is at least 110. To prove her point, she administers an IQ test to 20 randomly selected students. Among the sampled students, the average IQ is 108 with a standard deviation of 10. Based on these results, should the principal accept or reject her original hypothesis? Assume a significance level of 0.01. (Assume that test scores in the population of engines are normally distributed.)

Solution: The solution to this problem takes four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results. We work through those steps below:

- State the hypotheses. The first step is to state the null hypothesis and an alternative hypothesis.

  Nul**l** hypothesis: $\mu >= 110$

  Alternative hypothesis: $\mu < 110$

  Note that these hypotheses constitute a one-tailed test. The null hypothesis will be rejected if the sample mean is too small.

- Formulate an analysis plan. For this analysis, the significance level is 0.01. The test method is a one-sample t-test.

- Analyze sample data. Using sample data, we compute the standard error (SE), degrees of freedom (DF), and the t statistic test statistic (t).

  SE = s / sqrt(n) = 10 / sqrt(20) = 10/4.472 = 2.236

  DF = n - 1 = 20 - 1 = 19

  $t = (x - \mu) / SE = (108 - 110)/2.236 = -0.894$

  where s is the standard deviation of the sample, x is the sample mean, μ is the hypothesized population mean, and n is the sample size.

  Here is the logic of the analysis: Given the alternative hypothesis ($\mu < 110$), we want to know whether the observed sample mean is small enough to cause us to reject the null hypothesis.

  The observed sample mean produced a t statistic test statistic of -0.894. We use the t Distribution Calculator to find $P(t < -0.894) = 0.19$. This means we would expect to find a sample mean of 108 or smaller in 19 percent of our samples, if the true population IQ were 110. Thus the P-value in this analysis is 0.19.

- Interpret results.

- Since the P-value (0.19) is greater than the significance level (0.01), we cannot reject the null hypothesis.

Note: If you use this approach on an exam, you may also want to mention why this approach is appropriate. Specifically, the approach is appropriate because the sampling method was simple random sampling, the population was normally distributed, and the sample size was small relative to the population size (less than5%)

Procedure:

- Draw random numbers in MS Excel using Rand() / Rand between() and using Data Analysis Tool pack draw N(100, 15) random numbers
- Generate 4 (2 each) sample sets (Each set consisting of 10 random numbers) from the data set generated in the previous step using the Sampling feature of the Data Analysis Tool pack
- Use the Rank and Percentile feature of the Data Analysis Tool Pack
- Compute the mean and standard deviation of the samples from the Normal random number set and compare it with the given mean and standarddeviation
- Consider your performance in last 6 semesters
- Make an hypothesis regarding mean score
- Use the t.test () function in excel to compute p value
- Compare the p value with the level of significance
- Take a decision.
- Use the tdist() function of excel to compute p value and compare it with the p value computed using the t.test () function

Results: (Screen shot of the excel sheet)

1. Draw random numbers in MS Excel using Rand() / Rand between() and using Data Analysis Tool pack draw N(100, 15) random numbers

Random Numbers		
Data Analysis Tool Pack	RANDBETWEEN()	RAND()
104.2126089	51	0.648449938
96.72626245	151	0.381332835
108.5215675	143	0.085814815
108.3870702	123	0.332324966
96.96207169	24	0.687979307
99.80549319	200	0.441476787
91.73874698	87	0.196425338
99.57135515	51	0.780268111
107.3571186	188	0.010773068
102.3010841	122	0.793005538
99.28198235	90	0.882560032
110.0020429	75	0.515477158
97.80916369	87	0.230303185
105.9674335	166	0.178106568
91.8067715	129	0.17063097
102.2906306	106	0.806485387
107.135327	12	0.303355957
104.5575234	15	0.753154586
84.47891585	80	0.951580291
91.56189006	75	0.968648749
122.6498287	41	0.400948597
103.1751028	25	0.328843915
77.85057531	147	0.089847995
86.14332526	104	0.268770571
75.25435447	158	0.352959855
106.4399387	87	0.5874931
85.53593059	48	0.004379057
128.3545432	97	0.130716724
76.33804014	139	0.941905363
102.1131598	4	0.26467584
101.5328283	69	0.485503018
95.13023567	29	0.902894644
93.57643219	62	0.171178237
83.04837163	149	0.350344109

75.17345491	79	0.30914971
115.2781467	164	0.294566583
117.3756007	91	0.220223421
110.2209356	170	0.100993298
70.66156538	129	0.757361121
103.5823462	24	0.377942025
126.2267349	175	0.214097762
97.42451564	134	0.97897598
98.15769343	90	0.479067403
85.47921223	180	0.504715321
126.588873	85	0.895403377
98.78292556	130	0.630829187
104.8117499	90	0.099826458
102.6896885	112	0.393860794
105.2012069	194	0.5337642
110.6891093	184	0.940824157
116.2550123	74	0.064570614
92.98138391	87	0.420082522
104.1362796	43	0.112306326
97.89610001	96	0.751670095
88.93927113	56	0.924044782
92.09169346	71	0.579159328
95.93043185	107	0.516303818
127.5410457	162	0.224800348
120.2761157	153	0.786556264
117.3935746	40	0.499799199
107.2950456	157	0.114685337
93.27704927	12	0.054362967
83.9172915	171	0.365211955
84.77574081	176	0.393215631
89.67606416	167	0.352054249
102.7445139	74	0.067464386
93.26183797	66	0.801850247
110.9675511	53	0.124890861
81.75360361	127	0.826652875
89.72688218	143	0.649586464
93.45295123	114	0.388022071
83.49710495	73	0.540044986

99.01992851	2	0.531643618
121.9969706	174	0.379597468
99.35089363	184	0.445155954
103.1375521	72	0.460291923
123.9326255	35	0.029497887
100.9984717	190	0.44383182
92.88008894	164	0.175239203
88.38965212	171	0.46316107
101.3564318	191	0.648362443
99.15672788	82	0.808594129
103.9458996	145	0.671683027
147.0499799	85	0.882419014
81.76560894	142	0.739882457
100.5308266	175	0.697491973
106.6304551	34	0.550584438
115.9568572	146	0.604240662
103.8912788	144	0.073730431
84.3884666	155	0.026722731
102.8753107	190	0.66055162
93.39989927	185	0.399301355
110.3836555	5	0.100171557
101.4601824	126	0.416290904
83.89075699	70	0.536455873
103.022717	182	0.766185231
108.572863	25	0.837063024
117.6020308	112	0.718767586
95.81244197	102	0.611434284
87.45497578	89	0.850910448

2. Generate 4 (2 each) sample sets (Each set consisting of 10 random numbers) from the data set generated in the previous step using the Sampling feature of the Data Analysis Tool pack

Data Analysis Sample	
Sample-1	Sample-2
110.967551	122.6498287
89.6760642	88.93927113
116.255012	104.5575234
147.04998	84.3884666
110.967551	103.9458996
97.8961	83.04837163
126.588873	85.53593059
104.557523	105.2012069
103.137552	126.2267349
99.0199285	102.8753107

Random Between Sample	
Sample-1	Sample-2
155	109
168	101
166	84
84	24
23	5
162	175
87	58
73	84
108	5
130	83

Random Sample	
Sample-1	Sample-2
0.055325509	0.643017841
0.116396659	0.657104557
0.960617372	0.184474956
0.118647124	0.166233959
0.300064479	0.006648229
0.397446371	0.142059731
0.708849516	0.692855844
0.246294301	0.859929583
0.147883226	0.69363874
0.847245231	0.54230117

3. Use the Rank and Percentile feature of the Data Analysis Tool Pack

Data Analysis

Sample-1

Data Analysis Rank and Percentile			
Point	*Column1*	*Rank*	*Percent*
4	147.0499799	1	100.00%
7	126.588873	2	88.80%
3	116.2550123	3	77.70%
1	110.9675511	4	55.50%
5	110.9675511	4	55.50%
8	104.5575234	6	44.40%
9	103.1375521	7	33.30%
10	99.01992851	8	22.20%
6	97.89610001	9	11.10%
2	89.67606416	10	0.00%

Sample-2

Data Analysis Rank and Percentile			
Point	*Column1*	*Rank*	*Percent*
9	126.2267349	1	100.00%
1	122.6498287	2	88.80%
8	105.2012069	3	77.70%
3	104.5575234	4	66.60%
5	103.9458996	5	55.50%
10	102.8753107	6	44.40%
2	88.93927113	7	33.30%
7	85.53593059	8	22.20%
4	84.3884666	9	11.10%
6	83.04837163	10	0.00%

Random Between

Sample-1

Random Between Rank and Percentile			
Point	*Column1*	*Rank*	*Percent*
2	168	1	100.00%
3	166	2	88.80%
6	162	3	77.70%
1	155	4	66.60%
10	130	5	55.50%
9	108	6	44.40%
7	87	7	33.30%
4	84	8	22.20%
8	73	9	11.10%
5	23	10	0.00%

Sample-2

Random Between Rank and Percentile			
Point	*Column1*	*Rank*	*Percent*
6	175	1	100.00%
1	109	2	88.80%
2	101	3	77.70%
3	84	4	55.50%
8	84	4	55.50%
10	83	6	44.40%
7	58	7	33.30%
4	24	8	22.20%
5	5	9	0.00%
9	5	9	0.00%

Random Sample

	Random Rank and Percentile			
	Point	Column1	Rank	Percent
	3	0.960617372	1	100.00%
	10	0.847245231	2	88.80%
	7	0.708849516	3	77.70%
	6	0.397446371	4	66.60%
Sample-1	5	0.300064479	5	55.50%
	8	0.246294301	6	44.40%
	9	0.147883226	7	33.30%
	4	0.118647124	8	22.20%
	2	0.116396659	9	11.10%
	1	0.055325509	10	0.00%

	Random Rank and Percentile			
	Point	Column1	Rank	Percent
	8	0.859929583	1	100.00%
	9	0.69363874	2	88.80%
	7	0.692855844	3	77.70%
	2	0.657104557	4	66.60%
Sample-2	1	0.643017841	5	55.50%
	10	0.54230117	6	44.40%
	3	0.184474956	7	33.30%
	4	0.166233959	8	22.20%
	6	0.142059731	9	11.10%
	5	0.006648229	10	0.00%

4. Compute the mean and standard deviation of the samples from the Normal random number set and compare it with the given mean and standard deviation

Data Analysis Samples Comparison for Mean and SD

	Data Analysis Sample	
	Sample-1	Sample-2
	110.967551	122.6498287
	89.6760642	88.93927113
	116.255012	104.5575234
	147.04998	84.3884666
	110.967551	103.9458996
	97.8961	83.04837163
	126.588873	85.53593059
	104.557523	105.2012069
	103.137552	126.2267349
	99.0199285	102.8753107
Mean	110.611614	100.7368544
Standard Deviation	16.4721683	15.38294378

Random Between Samples Comparison for Mean and SD

	Random Between Sample	
	Sample-1	Sample-2
	155	109
	168	101
	166	84
	84	24
	23	5
	162	175
	87	58
	73	84
	108	5
	130	83
Mean	115.6	72.8
Standard Deviation	48.83350398	52.38702129

Random Samples Comparison for Mean and SD

	Random Sample	
	Sample-1	Sample-2
	0.055325509	0.643017841
	0.116396659	0.657104557
	0.960617372	0.184474956
	0.118647124	0.166233959
	0.300064479	0.006648229
	0.397446371	0.142059731
	0.708849516	0.692855844
	0.246294301	0.859929583
	0.147883226	0.69363874
	0.847245231	0.54230117
Mean	0.389876979	0.458826461
Standard Deviation	0.330698947	0.301235309

5. Consider your performance in last 6 semesters

Semester	Pointer
1	9.67
2	8.3
3	8.09
4	8.18
5	8.78
6	9.23
Average	8.708333333

6. Make an hypothesis regarding mean score

Column1	
Mean	8.708333333
Standard Error	0.260146326
Median	8.54
Mode	#N/A
Standard Deviation	0.637225758
Sample Variance	0.406056667
Kurtosis	-1.235078346
Skewness	0.672130975
Range	1.58
Minimum	8.09
Maximum	9.67
Sum	52.25
Count	6

Let H0 = 8.7 be Null Hypothesis
Let H0 != 8.7 be Alternate Hypothesis
Alpha = 0.05

Semester	Pointer		
1	9.67	8.7	
2	8.3	8.7	H0 = 8.7
3	8.09	8.7	H1 != 8.7
4	8.18	8.7	
5	8.78	8.7	
6	9.23	8.7	
Average	**8.708333333**		

7. Use the t.test () function in excel to compute p value
8.

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	8.708333333	8.7
Variance	0.406056667	3.78653E-30
Observations	6	6
Hypothesized Mean Difference	0	
df	5	
t Stat	0.032033254	
P(T<=t) one-tail	0.487842457	
t Critical one-tail	2.015048373	
P(T<=t) two-tail	0.975684915	
t Critical two-tail	2.570581836	

9. Compare the p value with the level of significance
   P = 0.975684915 and
   Level of significance = 0.05
   As P > LOS so H0 is accepted.
10. Take a decision.
   Therefore the assumed mean is consistent across 6 semesters.

11. Use the tdist() function of excel to compute p value and compare it with the p valuecomputed using the t.test () function

Using t.test() function	
p-value	0.975684915
alpha	0.05
As p>alpha, H0 is accepted	

Using tdist() function	
Mean - Hypothesis	0.00833333
(Mean - Hypothesis)/Standard error	0.03203325
tdist() function	0.97548454

Questions:

1. Define the term sample and sampling with an example?
   Ans: A sample refers to a subset of individuals or objects selected from a larger population, which is used to represent the entire population in a research study. Sampling is the process of selecting this subset from the population. For example, if a researcher wants to study the average height of students in a school, they might select a sample of 100 students from the entire student population.

2. Why is it necessary to do sampling during any research study?
   Ans: Sampling is necessary during any research study for several reasons:
   Cost-effectiveness: It's often impractical or impossible to study an entire population due to constraints such as time, resources, and logistics.
   Feasibility: Some populations are too large or dispersed to study comprehensively, making sampling the only practical option.
   Accuracy: By carefully selecting a representative sample, researchers can make accurate inferences about the population as a whole.
   Ethical considerations: Sampling allows researchers to minimize the burden on participants while still obtaining meaningful data.

3. What is the significance of p value?
   Ans: The significance of the p-value lies in its role in hypothesis testing. In statistical hypothesis testing, the p-value represents the probability of observing a test statistic as extreme as, or more extreme than, the one observed in the sample data, under the assumption that the null hypothesis is true. A small p-value (typically less than 0.05) suggests that the observed data is unlikely to have occurred if the null hypothesis is true, leading to the rejection of the null hypothesis in favor of the alternative hypothesis.
4. Joe is the third-string quarter back for the university of lower Alatoona. The probability that Joe gets into any game is 0.40.
   (a) *What is the probability that the first game Joe enters is the fourth game of the season?*
   Ans: The probability that Joe enters the first game is (1 - probability that he doesn't enter in the first three games) * probability that he enters the fourth game. So,
   (Joe enters in the fourth game)=(1−0.40)3×0.40=(0.60)3×0.40=0.216×0.40=0.0864P(Joe enters in the fourth game)=(1−0.40)3×0.40=(0.60)3×0.40=0.216×0.40=0.0864.

   (b) What is the probability that Joe plays in no more than two of the first five games?
   Ans: To find the probability that Joe plays in no more than two of the first five games, we can calculate the probabilities of him playing in exactly 0, 1, or 2 games and sum them up.
   Probability that Joe plays in 0 games: (1−0.40)5=0.07776(1−0.40)5=0.07776.
   Probability that Joe plays in 1 game: (51)×(0.40)1×(1−0.40)4=0.2304(15)×(0.40)1×(1−0.40)4=0.2304. Probability that Joe plays in 2 games: (52)×(0.40)2×(1−0.40)3=0.3456(25)×(0.40)2×(1−0.40)3=0.3456. Adding these probabilities together: 0.07776+0.2304+0.3456=0.653760.07776+0.2304+0.3456=0.65376.

Outcomes: CO3 Analyze simulation results to reach an appropriate conclusion.

Conclusion: Thus we created random numbers using data analysis tool pack,randbetween() and rand() functions and compared the mean and standard deviation by taking 2 sample sets for each function.
Also tested for hypothesis mean for pointer in last 6 semesters.

Grade: AA / AB / BB / BC / CC / CD /DD

Signature of faculty in-charge with date

References:

Books/ Journals/ Websites:

1. "Linear Congruential Generators" by Joe Bolte, Wolfram Demonstrations Project.
2. Severance, Frank (2001). *System Modeling and Simulation*. John Wiley & Sons, Ltd. p. 86. ISBN 0-471-49694-4.
3. The GNU C library's *rand()* in stdlib.h uses a simple (single state) linear congruential generator only in case that the state is declared as 8 bytes. If the state is larger (an array), the generator becomes an additive feedback generator and the period increases. See the simplified code that reproduces the random sequence from this library.
4. "A collection of selected pseudorandom number generators with linear structures, K. Entacher, 1997". Retrieved 16 June 2012.
5. "How Visual Basic Generates Pseudo-Random Numbers for the RND Function". *Microsoft Support*. Microsoft. Retrieved 17 June 2011.
6. In spite of documentation on MSDN, RtlUniform uses LCG, and not Lehmer's algorithm, implementations before Windows Vista are flawed, because the result of multiplication is cut to 32 bits, before modulo is applied
7. GNU Scientific Library: Other random number generators
8. Novice Forth library
9. Matsumoto, Makoto, and Takuji Nishimura (1998) ACM Transactions on Modeling and Computer Simulation
10. S.K. Park and K.W. Miller (1988). "Random Number Generators: Good Ones Are Hard To Find". *Communications of the ACM* **31** (10): 1192–1201. doi:10.1145/63039.63042.
11. D. E. Knuth. *The Art of Computer Programming*, Volume 2: *Seminumerical Algorithms*, Third Edition. Addison-Wesley, 1997. ISBN 0-201-89684-2. Section 3.2.1: The Linear Congruential Method, pp. 10–26.
12. P. L'Ecuyer (1999). "Tables of Linear Congruential Generators of Different Sizes and Good Lattice Structure". *Mathematics of Computation* **68** (225): 249–260. doi:10.1090/S0025-5718-99-00996-5.
13. Press, WH; Teukolsky, SA; Vetterling, WT; Flannery, BP (2007), "Section 7.1.1. Some History", *Numerical Recipes: The Art of Scientific Computing* (3rd ed.), New York: Cambridge University Press, ISBN 978-0-521-88068-8
14. Gentle, James E., (2003). *Random Number Generation and Monte Carlo Methods*, 2nd edition, Springer, ISBN 0-387-00178-6.
15. Joan Boyar (1989). "Inferring sequences produced by pseudo-random number generators". *Journal of the ACM* **36** (1): 129–141. doi:10.1145/58562.59305. (in this paper, efficient algorithms are given for inferring sequences produced by certain pseudo-random number generators).