Session 12

Sampling

Prof. Jigar M. Shah

Session 12

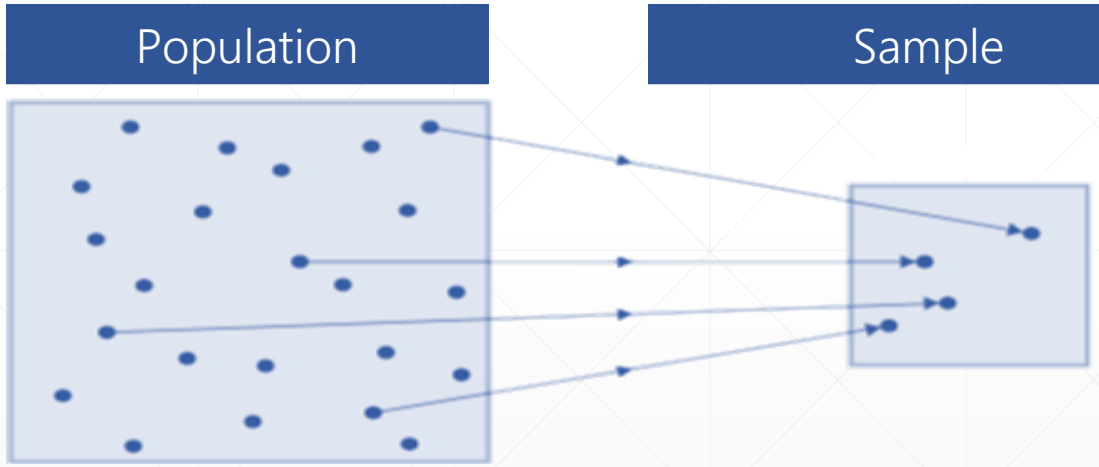
Sampling

- Sampling
  - Sampling Theory
  - Principles of Sampling
  - Sampling Methods
  - Simple Random Sampling

Sampling

Sampling

- Sampling Theory



Set of all elements of interest in a particular study

Subset of the population from whose elements data are collected

Population Parameters	Measures computed for data from a population
Sample Statistics	Measures computed for data from a sample
Census	Survey to collect data for entire population
Sample Survey	Survey to collect data for a sample

Sampling

Sampling

- Sampling Theory

Point Estimation	Estimating the value of a population parameter based on the value of the corresponding sample statistic
Point Estimator	Sample statistic corresponding to a population parameter
Point Estimate	Value of the sample statistic
Parameter Value	Value of the population parameter

Measure of the Characteristic	Population Parameter	Sample Statistic
Mean	μ	$\bar{x}$
Standard Deviation	σ	s
Proportion of Success	p	$\bar{p}$

A sample statistic is the point estimator of the corresponding population parameter

Sampling

Sampling

- Sampling Theory

A sample can be taken from a finite or an infinite population	
Finite Population	E.g. - an app developer introduced some changes in its app only for 5,000 of its 1,00,000 users. 3,895 users liked the changes. Based of this, the developed inferred that 77.9% of its 1,00,000 users will like the changes and hence, it introduced the changes for all users.
Infinite Population	E.g. - A tire manufacturer is considering producing a new tire designed to provide an increase in the life over the firm's current line of tires. To estimate the mean useful life of the new tires, the manufacturer produced a sample of 120 tires for testing. The test results provided a sample mean of 36,500 km. Hence, an estimate of the mean useful life for the population of new tires was 36,500 km.

- Sampling Theory - sampling theory provides the tools and techniques for data collection keeping in mind the objectives to be fulfilled and nature of population

Sampling

Sampling

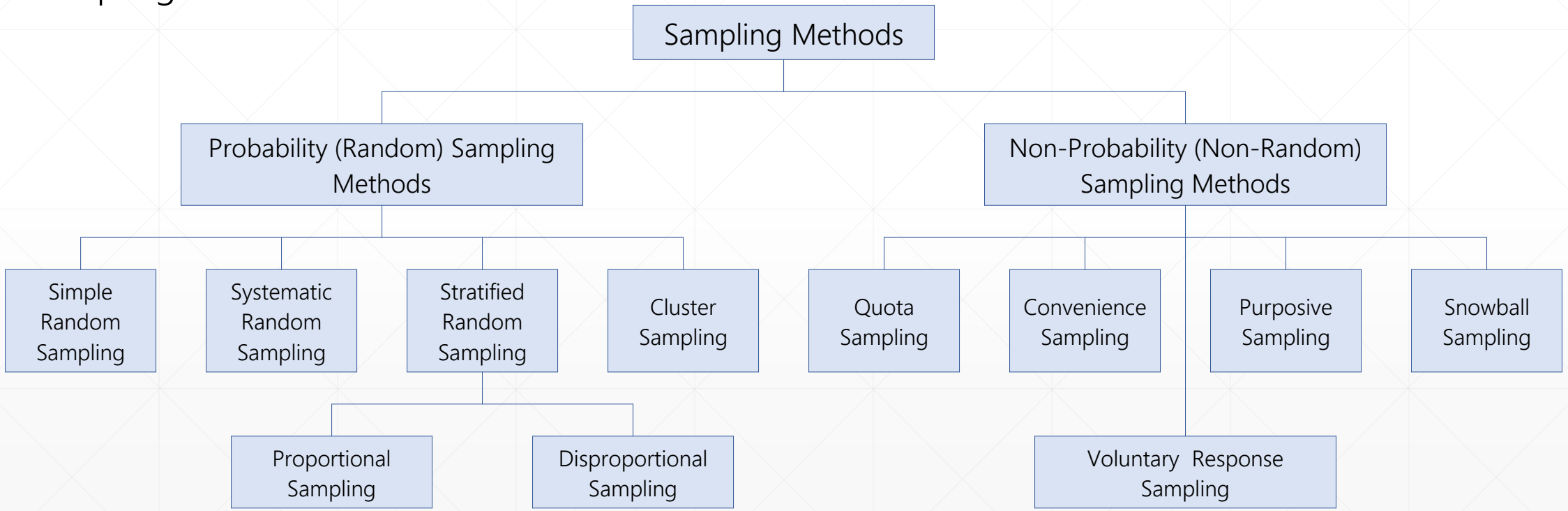
- Principles of Sampling

Principles of Sampling	
Principle of Statistical Regularity	A moderately large no. of items chosen at random from a large group are almost sure on the average to possess the characteristic of the large group
Large Sample Size	As the sample size increases, it becomes more & more representative of the parent population & shows its characteristics
Random Selection	A sample selected randomly (each unit having equal chance of inclusion in the sample) is more likely to possess almost the same characteristics & qualities of the population
Principle of Inertia of Large Numbers	As the sample size (no. of observations in a sample) get large enough, the statistical inference is likely to be more accurate & stable

Sampling

Sampling

- Sampling Methods



Sampling

Sampling

- Sampling Methods
  - Probability Sampling Methods
    - Methods of (randomly) selecting samples according to certain law of probability such that each unit of population has some definite probability (greater than zero) of being included in the sample
    - A number of samples of specified types, e.g., S_1, S_2, S_3, …, S_k can be formed by grouping units of the population such that each sample S_i has an associated probability of selection p_i; i = 1, 2, 3, …, k, with p_i denoting the probability of selection of the i^{th} sample
    - Increase the sample's representativeness of the population
    - Decrease the sample error & sample bias
    - Also known as Random Sampling Methods

Sampling

Sampling

- Sampling Methods
  - Probability Sampling Methods

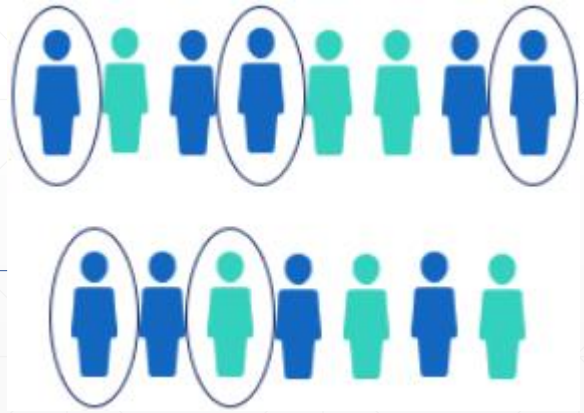
E.g. Total income of adults living in a given street can be estimated by visiting each household in that street, identifying all adults living there, randomly selecting an adult from each household, (for e.g., by allocating each person a random number, generated from a uniform distribution between 0 & 1, and selecting the person with the highest number in each household), interviewing the selected person and finding their income. People living on their own will certainly be selected, so their incomes will simply be added to the estimate of total income. But a person living in a household of two adults would have only a one-in-two chance of selection, and to reflect this, the selected person's income would be counted twice towards the total. (The person selected from that household can be loosely viewed as also representing the person not selected.) Here, not every person has the same probability of being selected, but each person's probability of selection is known & hence the sampling is probability sampling.

Sampling

Sampling

- Sampling Methods
  - Probability Sampling Methods
    - Simple Random Sampling
      - A probability sampling method where every unit of population has an equal chance of being included in the sample
      - Requires the sampling frame to include the whole population
      - Tools like random number generators can be used that are based entirely on chance

        E.g. Selecting a sample of 10 employees of an organization having a total of 100 employees by assigning a number from 1 to 100 to every employee in the employee database and using a random number generator to the generate 10 random numbers between 1 & 100 and selecting the employees that are assigned those random numbers

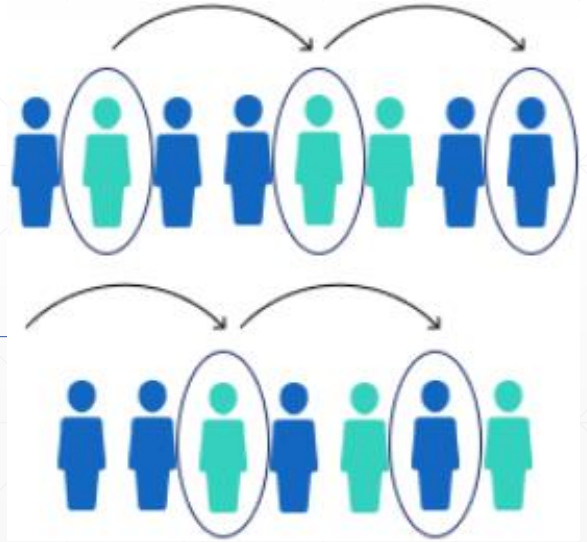


Sampling

Sampling

- Sampling Methods
  - Probability Sampling Methods
    - Systematic Random Sampling
      - A probability sampling method similar to simple random sampling but where, after a random start, units of population are chosen at regular intervals such that every kth unit of the population is chosen to be included in the sample
      - Slightly easier to conduct than simple random sampling

        E.g. After listing all employees of the organization in alphabetical order and randomly selecting a starting point from the first 10 numbers, say 6, selecting every 10th person on the list from that number (6) onwards such that the 6th, 16th, 26th, 36th, 46th, 56th, 66th, 76th, 86th & 96th employees on the list are included in the sample

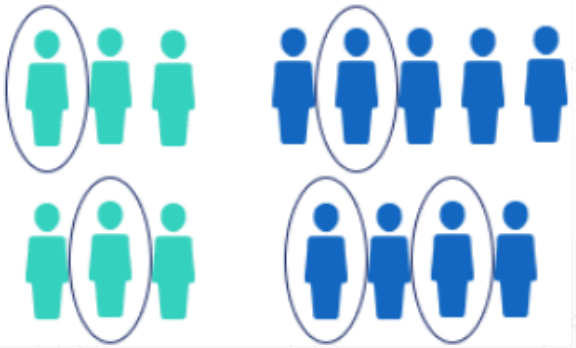


Sampling

Sampling

- Sampling Methods
  - Probability Sampling Methods
    - Stratified Random Sampling
      - A probability sampling method that involves dividing the population into sub-populations, known as strata, based on relevant characteristics such as gender, age group, income range, job role, etc., and then using simple random sampling or systematic random sampling to choose units from each stratum to be included in the sample

        E.g. If of the 100 employees of the organization, 80 are female & 20 male, the population can be divided into two strata: female employees (having size 80) & male employees (having size 20), and then using simple or systematic random sampling to choose 8 units from the female employees stratum & 2 units from the male employees stratum to be included in the sample



Sampling

Sampling

- Sampling Methods
  - Probability Sampling Methods
    - Stratified Random Sampling
      - Proportional Sampling - stratified random sampling with the size of subpopulations i.e., strata, proportional to the sizes of the subpopulations in the population such that the sample has the same proportions as the population

        E.g. If population of 100 employees consists 80 females & 20 males, it can be divided into 2 strata: female employees (having size 80) & male employees (having size 20) & then choosing units from these 2 strata such that female-to-male ratio in sample is 4:1 as in the population

      - Disproportional Sampling - stratified random sampling with the size of subpopulations i.e., strata, not proportional to the sizes of the subpopulations in the population such that the sample does not have the same proportions as the population

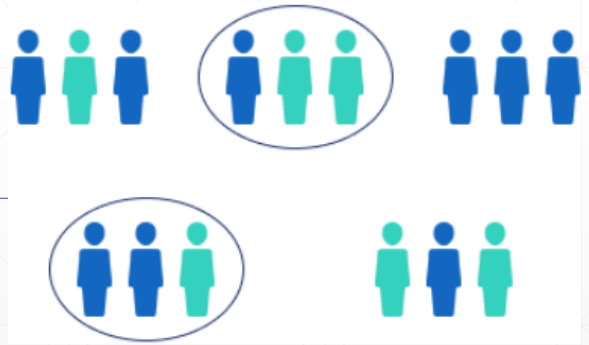
        E.g. If population of 100 employees consists 80 females & 20 males, it can be divided into 2 equal-sized strata (each having size 50) and then choosing units from these 2 strata such that female-to-male ratio in sample may not be 4:1

Sampling

Sampling

- Sampling Methods
  - Probability Sampling Methods
    - Cluster Sampling
      - A probability sampling method that involves dividing the population into sub-populations, known as clusters, such that each cluster has the same characteristics as the population and then randomly choosing some clusters to either survey every unit or to draw samples from using any sampling method

        E.g. If the organization has offices in 10 cities across the country (all with roughly the same number of employees in similar roles), random sampling can be used to select 3 offices as clusters, and from those clusters either every employee could be surveyed, or samples could be drawn to survey some employees



Sampling

Sampling

- Sampling Methods
  - Non-Probability Sampling Methods
    - Methods of (non-randomly) selecting samples based on the discretion of sampler / researcher such that not all units have a chance of being included in the sample (some may have no chance, zero probability of being included in the sample), or where the probability of selection can't be accurately determined
    - Do not involve use of any sampling frame
    - Decrease the sample's representativeness of the population, thereby restricting the generalization of the results
    - Easier & cheaper to implement, but increase the sample bias
    - Also known as Non-Random Sampling Methods

Sampling

Sampling

- Sampling Methods
  - Non-Probability Sampling Methods

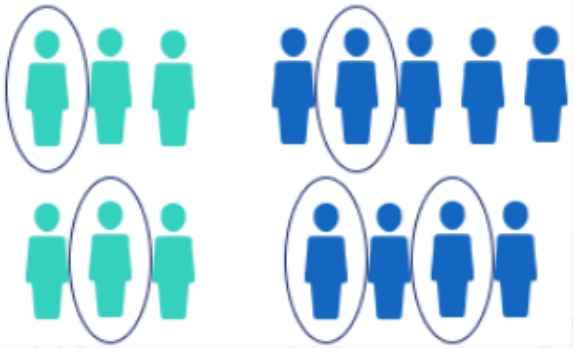
E.g. Total income of adults living in a given street can be estimated by visiting each household in that street and interviewing the first person to answer the door. In any household with more than one occupant, this is a non-probability sample, because some people are more likely to answer the door (e.g., an unemployed person spending most of their time at home is more likely to answer than an employed housemate who might be at work when the interviewer calls) and it's not practical to calculate these probabilities.

Sampling

Sampling

- Sampling Methods
  - Non-Probability Sampling Methods
    - Quota Sampling
      - A non-probability sampling method that involves setting quotas for subgroups based on proportions of certain characteristics of the population so that the sample reflects those characteristics in proportion to their prevalence in the population
      - Requires prior knowledge about the population characteristics
      - Similar to stratified random sampling but does not involve random sampling

        E.g. If the population of customers of a store comprises of 50% male & 50% female customers, then selecting the first 35 male customers & the first 35 female customers (in a non-random manner) meeting the inclusion criterion to obtain a sample of 70 customers

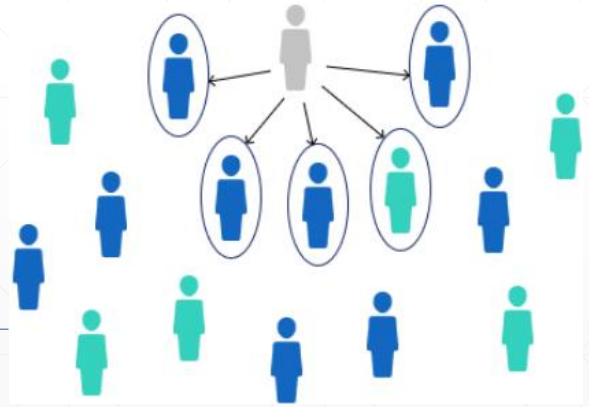


Sampling

Sampling

- Sampling Methods
  - Non-Probability Sampling Methods
    - Convenience Sampling
      - A non-probability sampling method that simply includes the most accessible units in the sample
      - Easy and inexpensive to gather initial data, but no way to determine the sample's representativeness, so can't produce generalizable results
      - Also known as chunk sampling or accidental sampling or incidental sampling

E.g. A professor researching opinions about student support services in the university, asking fellow students after every class to complete a survey on the research topic

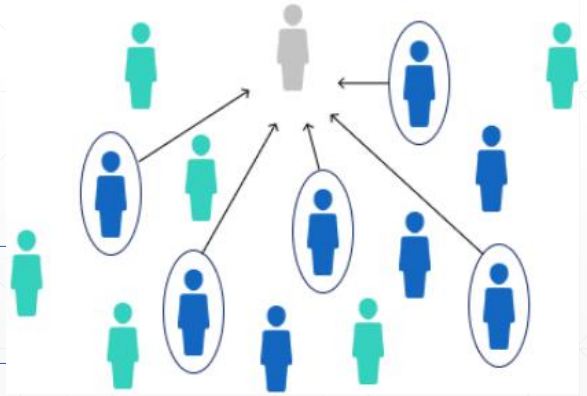


Sampling

Sampling

- Sampling Methods
  - Non-Probability Sampling Methods
    - Voluntary Response Sampling
      - A non-probability sampling method similar to convenience sampling but in which the units volunteer themselves for the survey instead of being chosen by the researcher
      - Always somewhat biased as some units will inherently be more likely than others to volunteer to provide their responses

E.g. Sending out survey to all students at the university and receiving responses only from those who volunteered to complete the survey & share their responses

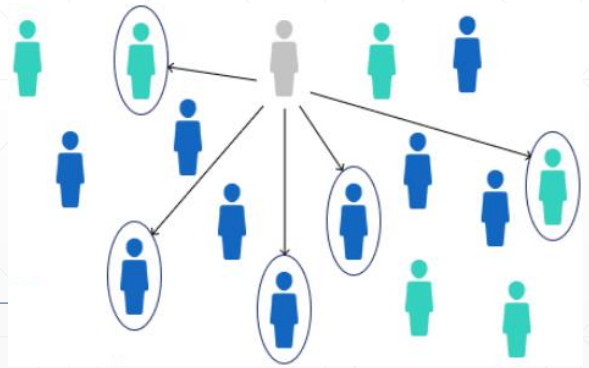


Sampling

Sampling

- Sampling Methods
  - Non-Probability Sampling Methods
    - Purposive Sampling
      - A non-probability sampling method that involves using the researcher's expertise to select a sample that will be most useful to the purpose of the research
      - Useful in qualitative research to gain detailed knowledge about a specific phenomena than to make statistical inferences about the population
      - Also known as judgmental sampling

        E.g. Purposefully selecting a number of university students with different support needs in order to gather a varied range of data on their experiences with student services while wanting to know more about the opinions and experiences of disabled students at the university

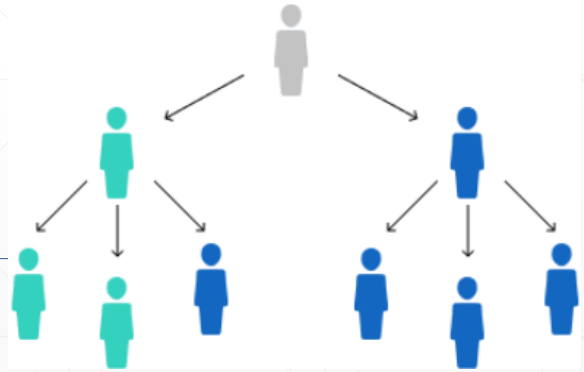


Sampling

Sampling

- Sampling Methods
  - Non-Probability Sampling Methods
    - Snowball Sampling
      - A non-probability sampling method in which the units refer other units to the researcher such that the no. of accessible units snowballs as more units are contacted
      - Useful when the population is hard to access
      - Also known as network sampling

E.g. One homeless person participating in the research and connecting the researcher with other homeless people for researching experiences of homelessness in the city



Sampling

Sampling

- Simple Random Sampling

Simple Random Sample
A simple random sample of size n from a finite population of size N is a sample selected such that every population unit has an equal chance of being included in the sample i.e., A simple random sample of size n from a finite population of size N is a sample selected such that each possible sample of size n has the same probability of being selected
E.g. - using random number table to choose elements for the sample one at a time in such a way that, at each step, each of elements remaining in the population has the same probability of being selected

Sampling

Sampling

- Simple Random Sampling

Simple Random Sample Without Replacement

Ignoring previously used random numbers so that the observations (data) already included in the sample can not be chosen in the sample again

Simple Random Sample With Replacement

Accepting previously used random numbers so that the observations (data) already included in the sample can be chosen in the sample again

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	95436	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	84358	21625	16999	13385	22782
55363	07449	34835	15290	76616	67191	12777	21861	68689	03263
69393	92785	49902	58447	42048	30378	87618	26933	40640	16281
13186	29431	88190	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02421	98063	89641	64953	99337
84649	48968	75215	75498	49539	74240	03466	49292	36401	45525
63291	11618	12613	75055	43915	26488	41116	64531	56827	30825
70502	53225	03655	05915	37140	57051	48393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	67342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09226	64419	29457
10078	28073	85389	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	79607	52244	63303	10413	63839	74762	50289

Random Number Table

Thank You

Prof. Jigar M. Shah