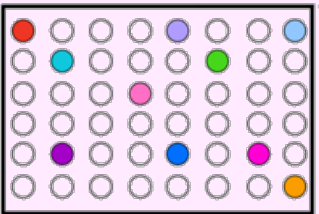

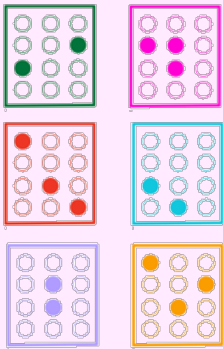
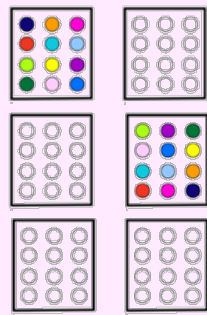
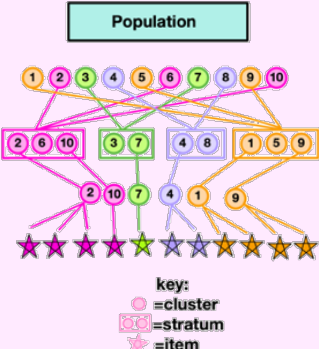
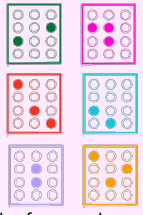
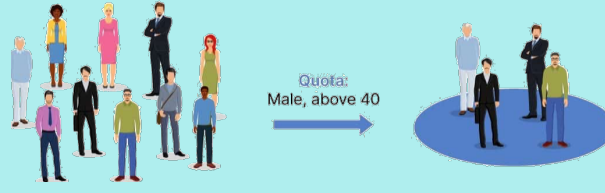

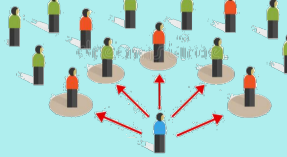

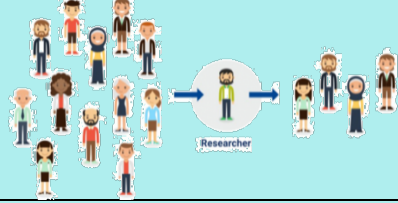

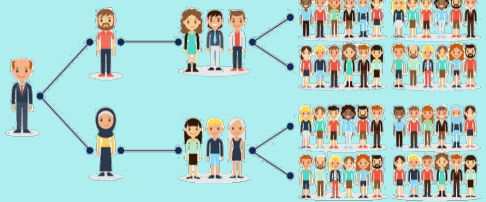

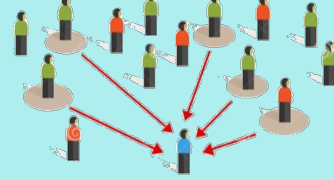


Sampling Methods

Name	Picture	Explanation	Example
Random			
Simple Random Sample (SRS)		Everyone in the population (i.e. sampling frame) has an equal opportunity to be selected and every sample of certain size has the same chance of being selected (as every other sample of the same size) (it is important you realise that not every sample of certain size has the same chance of being selected for stratified sampling)	e.g. 150 cars in a parking lot. Want to pick 3 of them Label cars 001-150 since want blocks of 3 <ul style="list-style-type: none"> If using a number line table: 456124321150240984312102364219814 Split the number line up into blocks of 3 and select 3 cars 456 124 321 150 240 984 312 102 364 219 814 If using calculator: Randint (1, 150, 3) will select 3 cars 60 BMW cars at garage, want to pick 7 (use a calculator to do this) Label cars 01-60 since want blocks of 2 (Note: We could label 00-59) Using calculator: Math, PRB, randint (1,60,7) This gives us 7 numbers 44,1,26,19,59,2,5
Systematic		Choose subjects in a systematic, orderly/logical way by sampling every k^{th} element/value	e.g. Select a sample of 15 people from a business survey of 1200 people Label the people 0001 to 1200 $\frac{1200}{15} = 80$ i.e. we'll take every 80 th value <ul style="list-style-type: none"> If using number line: 12436289461234560124 Split this up into blocks of 4: 1243 6289 4612 3456 0124 124 is our starting point 124+80=204, 204+80=284, 284+80=364 etc 124,204,284,364,444, 524, 604,684, 764,844,924,1004,1084, 1164, 044 If using a calculator: Randint (1,1200,1) gives 24 so start at 24 24+80=104, 104+ 80 = 184, 184+80=264 etc 24,104,184,264,344,424,504,584,664,744,824,904,984,1064,1144
Stratified Random selection within each strata	 Alike/homogeneous WITHIN strata/groups Mutually exclusive/heterogeneous/dissimilar BETWEEN groups i.e. the stratum are non-overlapping	Split population into smaller groups called strata (subgroups) based on common characteristics/shared attributes. Everyone in the population has an equal opportunity to be selected but NOT EVERY SAMPLE OF A CERTAIN SIZE HAS THE SAME CHANCE OF BEING SELECTED, so this is not like SRS. Stratified sampling ensures the sample group is representative of the population's characteristics. A random sample of each stratum is taken in a number proportional to the stratum's size when compared to the population. We stratify based on what outside factor we think may affect what we are testing.	e.g. A company has 320 drivers, 80 office workers and 40 mechanics. We want to select a committee of 11 to represent all the employees. number for each category = $\frac{\text{category total}}{\text{complete total}} \times \text{sample size required}$ Drivers = $\frac{320}{440} \times 11 = 8$, Office = $\frac{80}{440} \times 11 = 2$, Mech = $\frac{40}{440} \times 11 = 1$ Label: Drivers 001-320, Office 321-400, Mechanics 401-440 Drivers: Randint (1,320,8) will select 8 drivers Office workers: Randint (321,400,2) will select 2 office workers Mechanics: Randint (1,320,1) will select 1 mechanic
Cluster random selection between strata stratified vs cluster: Select A few from EVERY category vs all from SOME categories. Randomly select elements within each group versus randomly selecting whole groups	 Homogeneous/alike BETWEEN groups. Dissimilar/ heterogeneous/mutually exclusive WITHIN groups i.e. clusters are non-overlapping	Split population into smaller groups called clusters and sample EVERYONE from randomly chosen subgroups. Normally we split up based on area/geographical location. This is similar to stratified sampling, but with stratified sampling the randomness is applied when selecting within each subgroup whereas with cluster sampling the randomness is applied when selecting between each subgroup. When we select the subgroup, we choose EVERY number inside. Therefore not all subgroups will be chosen and therefore won't be represented in the sample unlike with stratified.	An airline that wants to assess customer satisfaction chooses a random sample of 10 of its flights during a single month and asks all of the passengers on those flights to fill out a survey. Why is this a cluster sample and not stratified?
Multistage	 key: ○ = cluster □ = stratum ★ = item	This is a complex form of cluster sampling. It utilizes characteristics of stratified and cluster sampling. Step 1: The population is divided into multiple clusters Step 2: These clusters are further divided and grouped into various strata based on similarity. Step 3: A SRS is taken from each strata. Step 4: Divide into multiple clusters again Step 5: Divide and group into various strata again This process continues until the cluster can't be divided anymore (sample size gets reduced at each stage).	Country ⇒ States ⇒ Counties ⇒ Neighbourhoods etc Step 1: Start with a population Step 2: Break down into several clusters (e.g. countries) Step 3: Then group/merge the similar clusters into strata (based on similar characteristics of whole clusters) Step 4: Take a SRS of clusters from each strata We will be at the point of having clusters again now but less than originally (e.g. states) REPEAT steps 1-4 i.e. group clusters into strata again based on similar characteristics of whole clusters We will have even less clusters now that before (e.g. counties) We use a combination of 2 or more SRS's If these clusters are now small enough we pick individual items/elements from each cluster and are then done

Sampling Methods

		When the clusters become small enough we pick elements rather than divide into strata and are done	
Non-Random (no sampling frame) aka non-probability			
Quota	 <p>Except samples from each are not random</p>	This is similar to stratified sampling except samples chosen from each strata are not random. We use this when strata are present and stratified sampling is not possible i.e. can't get sampling frame.	 <p>Quota: Male, above 40</p>
Convenience/ Opportunity		It involves sampling being drawn from part of the population that is close to hand. In other words, this sampling method involves getting participants wherever you can find them and typically wherever is convenient.	<p style="text-align: center;">Convenience sampling</p> 
Judgemental/ Purposeful		The researcher selects units to be sampled based on their knowledge and professional judgment	
Snowball		Existing study subjects recruit future subjects from among their acquaintances. It is used where potential participants are hard to find. Thus, the sample group is said to grow like a rolling snowball.	
Voluntary		A voluntary sample is made up of people who self-select into the survey. Often, these folks have a strong interest in the main topic of the survey	

Understanding The Difference between SRS and Stratified Examples

With stratified sampling, one stratifies the whole population before applying random sampling methods to ensure each subgroup within the population receives proper representation (proportional representation) within the sample, which SRS does not. Therefore the sample should be highly representative of the target population. Thus stratified sampling provides better coverage of the population. A SRS might over/under represent a subgroup unlike stratified.

Reasons for using SRS over stratified

- i. When population is homogeneous (too similar to be divided into groups)
- ii. Every sample of certain size has the same chance of being selected
- iii. Easier and less expensive
- iv. Use when no reason to stratify, i.e. no reason to suspect that an outside factor is affecting the response variable. If we suspect that there is an outside factor, we stratify based on that factor

Reasons for using stratified over SRS

- i. When the population is heterogeneous (dissimilar) and certain homogeneous (similar) strata can be isolated i.e. population can easily be divided into strata
- ii. Each subgroup within population will receive proper representation so sample will be highly representative of population (this provides better coverage of the population)
- iii. If there is large variation in the values of the population (i.e. very high values and also very low values in our crop example above), SRS should balance these out. However, this is not always the case since all possible combinations are possible, all the selected values might be low or all selected values might be high. Stratification reduces the variation in the sample values calculated since the values chosen for a particular stratum vary very little from sample to sample (i.e. between strata) relative to the variability in the population.
- iv. When one suspects that there is some outside factor that can influence or affect the response one must choose the strata correctly though (i.e. homo within and hetero between). One can identify how the outside factor is affecting what you're testing
- v. highly representative of the strata or layers in population. Useful when comparing e.g. age, gender, grade, ethnicity. Sample accurately reflects population structure
- vi. When a SRS of an entire population will likely not generate enough analysable cases for a given group of particular interest

A hotel has 30 floors with 40 rooms per floor. The rooms on one side of the hotel face the water while rooms on the other side face a golf course. There is also an extra charge for rooms with a water view. The hotel manager wants to survey 120 guests who stayed at the hotel during a convention about their overall satisfaction with the property. Explain why a stratified random sample would be preferable to just taking a simple random sample, in this case. Also, indicate what you would use for the strata.

Ans. 1200 rooms. Stratify based on the views. **One strata would be rooms with water view and the other strata would be rooms with golf course view and a SRS proportional to the size be picked from each of these strata.** SRS may not give an accurate picture of customer satisfaction as satisfaction may differ for customers with room with either a water view or golf course view. SRS tests satisfaction levels purely, it doesn't link it directly to which view the customer has. Although random, SRS can under/over represent particular subgroups. SRS doesn't always balance everything out i.e. balance out in terms of all possible values being represented. Unlikely to be a problem here though since tiny variation in values i.e. only values here are pretty much only very satisfied, fairly satisfied and dissatisfied or some scale from 0-10.

In other words, SRS would require one to select 120 guests from 1200. A stratified sample would require one to take small samples from independent subgroups which would allow one to take into account a different variable on which the subgroups are based (water views) thus allowing one to take into account the opinion based on the type of room if the strata are types of rooms

You decide to take a survey of high school students at ACS school to see what their opinions are regarding the current high school daily timetable schedule. There are currently 598 students in the high school and you decide to survey 50 of them. In general, why might a researcher prefer a stratified random sample over a simple random sample?

Sampling Methods

Ans. Opinions about timetable may differ for students in different grades or perhaps differ by gender. SRS would show opinions only, it doesn't link to grades/gender/subjects taken to allow a comparison to be made. One can however stratify based on these categories. A researcher might prefer a stratified sample when the population can easily be divided into strata. Stratified sampling ensures each subgroup receives proper representation within the sample therefore sample will be highly representative of the population. A SRS might over/underestimate a subgroup unlike stratified when there is large variation in the values. Stratification reduces the variation in the sample values tested since the values chosen for a particular stratum vary very little from sample relative to the variability in the population and thus all categories chosen would be included in the sample. This is very important when there is large variability in the values of the population. So, when one suspects that some outside factor might affect the response, one can identify this by stratifying. However, one must choose the strata correctly!

A farmer has just cleared a new field for corn. It is a unique plot of land in that a river runs along one side. The corn looks good in some areas of the field but not others. The farmer doesn't know whether harvesting all the plots would be worth the expense. So, he decides to pick a sample of 10 to harvest and use the info to estimate the total yield and make a prediction based on the mean yield per plot and total yield. Based on this info, he will decide whether to harvest the remaining plots. Which would be the best method of sampling for him to choose? Mean yield per plot are displayed in diagram below:

6	17	20	38	47	55	69	76	82	97
7	14	23	34	43	56	63	75	81	92
2	14	28	30	50	50	62	80	85	96
9	15	27	34	43	51	65	72	88	91
4	15	28	32	44	50	64	76	82	97
5	16	27	31	48	59	69	72	86	99
5	18	28	34	50	60	62	75	90	90
8	15	20	38	40	54	62	77	88	93
7	17	29	39	44	53	61	77	80	90
7	19	22	33	49	53	67	76	86	97



Hint: We want to stratify based on an outside factor you think might be affecting what you're testing such as distance to river so want the groups to be similar based on this, so stratify based on similarity, closeness to river

Ans. Choose vertical strata, not horizontal. Vertical strata will have homogeneity within the groups based on proximity to the river. If we choose horizontal strata then we will not have homogeneity within groups based on proximity since the elements will be different distances from the river i.e. heterogeneous within. We're stratifying based on an outside factor that might affect what we're testing. Sample will be highly representative of the population since we are sure to include high and low yield plots due to having each subgroup based on proximity to river represented

You want to study income differences between black and white in UK. Want sample size of 500. Should we use SRS or stratified?

Ans. If we use SRS we might get mostly white people since only 10% of population is black. So SRS will likely generate 50 black and 450 white. If divide into strata we can make sure we pick a suitable size from both white and black people and therefore make sure both are well represented

A corporation employs 2000 male and 500 female engineers. A stratified random sample of 200 male and 50 female engineers give each engineer 1 chance in 10 to be chosen. This sample design gives every individual in the population the same chance to be chosen for the sample. Is it an SRS? Explain your answer

Ans. No, not every sample of a certain size has the same chance of being selected

Your school will send a delegation of 35 seniors to a student life convention. 200 girls and 150 boys are to be chosen. If a sample of 20 girls and a separate sample of 15 boys are each selected randomly, it gives each senior the same chance to be chosen to attend the convention. Is this an SRS?

Ans. No because not every sample/group of 35 people has the same chance of being selected. Every person might have the same chance, but every sample doesn't!

Fill in The Blank Questions

_____ You label all the freshmen 001-160, sophomores 161-310, juniors 311-430 and seniors 431-600. Then you use your calculator to randomly select 100 students. Which sampling method did you employ?

_____ You label all the freshmen 001-160, sophomores 161-310, juniors 311-430 and seniors 431-600. Then you use your calculator to randomly select 25 students from each grade because you think options may vary based on grade level. Which sampling method did you employ?

_____ You stand in the marble hall one morning before school and randomly pick 25 students to get their opinions about the food in the canteen. Which type of sampling method did you employ?

_____ You wait until lunch, number the tables in the canteen 1 through 40 and then randomly select 5 of them. You then ask everyone at each table about their opinion regarding the canteen's food. Which type of sampling method did you employ?

_____ You want to get a good idea of what students at the high school think of the food in the canteen. You label the freshmen 001-120, sophomores 121-280, juniors 281-460 and the seniors 461-600. The you use your calculator to randomly select 50 students.

_____ We divide a school into four groups based on class (freshmen, sophomore, juniors and seniors). Then we ask random samples of 25 from each class what their opinion regarding whether it's better to have first semester exams in December or January.

_____ To gather information about the validity of a new maths tests for students in the UK, a random sample of 15 state schools was selected. The new test was then administered to every student in the selected schools.

_____ In order to select a sample of undergraduate students in the United Kingdom, I select a simple random sample of ten countries. From each of these countries, I select a simple random sample of five councils. Finally, from each of these countries, I select a simple random sample of five councils. Finally, from each of these 5 councils, I select a simple random sample of 25 residents. My final sample has 1250 respondents

_____ An ACS board of directors want to get feedback from parents on the quality of teachers at the Cobham campus however they feel opinions may vary based on which school is surveyed. Therefore, they randomly select parents of students from early childhood, the lower school. Middle school and high school separately.

_____ A London bus tour company wants customer feedback, so they randomly select 10 of their buses on a particular Saturday and ask every customer on board to fill out a quick customer satisfaction form before they get off the bus.

_____ A sampling technique which gives every individual the same likelihood of being selected as well as every sample of a certain size the same likelihood of being selected

Answers: SRS, Stratified, Convenience, Cluster, SRS, Stratified, Cluster, Multi-stage, Stratified, Cluster, SRS

Multiple Choice Examples

A public opinion poll in England wants to determine whether or not the country approves of the measure to ban smoking in all public areas. They select a simple random sample of 50 residents from each county of the country to ask whether they approve or disapprove of the measure. This is an example of a

- Systematic Random Sample
- Stratified Random Sample
- Multistage Sample
- Cluster Sample
- Convenience Sample

A sample of students will be selected from all the students at a high school. Which of the following sampling methods is least likely to produce a representative sample of the students?

- From a list of all student names, randomly select 50 names from the list for the sample.
- Randomly choose five classrooms in the school and use all the students in those classrooms for the sample.
- Divide the students into the four class years (freshman, sophomore, junior, senior) and randomly select students from each year in proportion to the number of students in each year.
- From a numbered list of all student names, randomly choose a starting point and then choose every tenth student on the list
- From a randomly chosen home basketball game, choose every tenth student who enters the gymnasium

A certain store has 3,000 employees working at its main location in a city and 100 employees working at a smaller location outside the city. The store manager will select a sample of 50 employees from all the employees to ask their opinions about extending store hours during the holidays. What is the advantage of selecting a stratified random sample, with location as strata, instead of a simple random sample?

- Divide the town into nonoverlapping regions. Randomly select from the nonoverlapping regions, and select all housing units in those regions.
- The stratified sample assures the most random selection from among all selection methods.
- The stratified sample will be less expensive to implement than a simple random sample and will save money over time.
- The stratified sample assures employees that store hours will not be extended until their opinions are recorded.

Sampling Methods

- e) There is no real advantage to stratifying because a simple random sample is always the best

A certain company has five departments, A, B, C, D, and E. The number of employees in each department is 10, 10, 20, 30, and 30, respectively. A sample of 10 employees from all employees will be selected. Of the following descriptions of sampling procedures, which is most likely to introduce a potential source of bias?

- Randomly select 10 employees from department c
- Randomly select 10 employees from all 100 employees
- Randomly select 2 employees from each of the five departments
- Randomly select 1 employee from each of A and B, 2 employees from C, and 3 employees from each of D and E
- Randomly select every tenth employee from a randomized list of names of all employees

A large elementary school has 15 classrooms, with 24 children in each classroom. A sample of 30 children is chosen by the following procedure. Each of the 15 teachers selects 2 children from his or her classroom to be in the sample by numbering the children from 1 to 24, then using a random digit table to select two different random numbers between 01 and 24. The two children with those numbers are in the sample. Did this procedure give a simple random sample of 30 children from the elementary school?

- No, because the teachers were not selected randomly
- No, because not all possible groups of 30 children had the same chance of being chosen
- No, because not all children had the same chance of being chosen
- Yes, because each child had the same chance of being chosen
- Yes, because the numbers were assigned randomly to the children

Answers: b, e since don't have a random starting point, b, a, b

Longer Examples

ACS high school decides to send a delegation of students to a conference on student life of international high school students. They want to send a group representative of all four years of the high school, so they will randomly select 12 of the 120 freshmen, 16 of the 160 sophomores, 18 of the 180 juniors and 14 of the 140 seniors.

- Is this a simple random sample (SRS), explain why or why not as if it is not, what type of sample is it?
- Use the line of random digits below to select to select the first three juniors who will be sent to the conference.
71487 09984 29077 14863 61683 47052 62224 51025
- Choose the subjects using a random number generator instead to select a (total) sample of size 30 this time

Ans.
a) no, not every sample of a certain size has the same chance of being selected. This is stratified sampling.

b) 2 Ways to Label:

Non-Continuous: 001-120 freshman, 001-160 sophomore, 001-180 juniors, 001-140 seniors

OR
Continuous: 001-120 freshman, 121-280 sophomore, 281-460 juniors, 461-600 seniors

Hint: Use the one which will give you a sample size of 3 from the given table which is the latter

We select students with numbers: 077, 148, 168

c) This is harder than part b since we were told the sample size that we wanted from the strata juniors only and just found the first 3. Here we are told the **total** sample size is instead 30 and need to proportion the sizes for each strata

$$\text{freshman: } \frac{120}{600} \times 30 = 6$$

$$\text{sophomores: } \frac{160}{600} \times 30 = 8$$

$$\text{juniors: } \frac{180}{600} \times 30 = 9$$

$$\text{seniors: } \frac{140}{600} \times 30 = 7$$

Label 001-120 freshman, 121-280 sophomore, 281-460 juniors, 461-600 seniors. Ignore repetitions

Pick 6 freshman Randint (001,120,6)

8 sophomores Randint (121,280,8)

9 juniors Randint (281,460,6)

7 seniors Randint (461,600,6)

Mr. Van Nuys needs to pick 5 of his probability & statistics students to go to 10th grade advisory classes so they can describe what the course is like to students who may take the course next year. Use the line of random digits and roster given below to select the five students. Make sure to show how you label the students!

Andrew	Alex	Sophia	Garrett	Jude
Esther	Kyle	Abdulla	Bailey	AJ
Tyler	Matvey	Williams	Mackenzie	Guy
Brandi	Tarek	Madelief	Julian	Ben
Tori	Shuming			

Random Digits: 69051 64817 87174 09517 84534 06489 87201 97245

Ans.

Label students 00-22 going horizontally

We select 05, 16, 17, 20, 19

Jude, Brandi, Tarek, Ben, Julian

An apartment building has nine floors and each floor has four apartments. The building owner wants to install new carpeting in eight apartments to see how well it wears before she decides whether to replace the carpet in the entire building. The figure below shows the floors of apartments in the building with their apartment numbers. Only the nine apartments indicated with an asterisk (*) have children in the apartment.

- For convenience, the apartment building owner wants to use a cluster sampling method, in which the floors are clusters, to select the eight apartments. Describe a process for randomly selecting eight different apartments using this method.
- An alternative sampling method would be to select a stratified random sample of eight apartments, where the strata are apartments with children and apartments with no children. A stratified random sample of size eight might include two randomly selected apartments with children and six randomly selected apartments with no children. In the context of this situation, give one statistical advantage of selecting such a stratified sample as opposed to a cluster sample of eight apartments using the floors as clusters.

Ans.

Step 1: Assign unique random numbers to each of the 9 floors by labelling them 01-09

Step 2: use random number generator to select 2 floors

Step 2: Replace the carpet of the 4 apartments on the 2 floors

Since wear and tear on the carpet may be different in apartments with and without children, you would be guaranteed to have both represented using a stratified sample. In a cluster sample, it is possible to get 8 apartments where no children live.

11*	12	21	22*	31	32				
14	1st Floor	13	24	2nd Floor	23*	34	3rd Floor	33	
41	42	51*	52	61	62				
44	4th Floor	43	54	5th Floor	53	64	6th Floor	63	* = Children in the apartment
71	72	81	82	91	92*				
74*	73*	84*	83	94	9th Floor	93*			