

* The target population of a survey is the population you wish to study.
* The sampled population is the population which you are **able to observe( means sample of people you were able to get and not what you actually wanted as sampling is a very slow process and the time till which you think you have completed many things might would have changed)** in a sample .
* In an ideal world the target population and the sampled population would be the same, but often they are different.

**Example**, governments around the world survey active businesses to determine estimates for labour force, productivity, innovation, capital expenditure, etc.

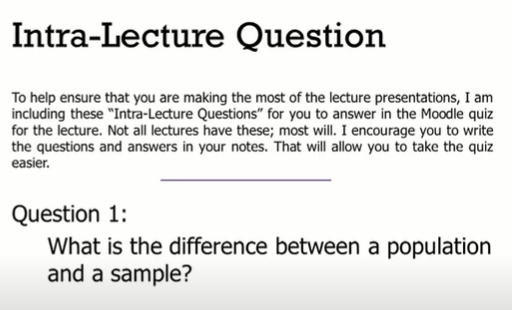
Typically, this is done by maintaining a register of businesses and sampling from this register. However, there is a gap of weeks/months between when this sample was created to when the survey goes to field.

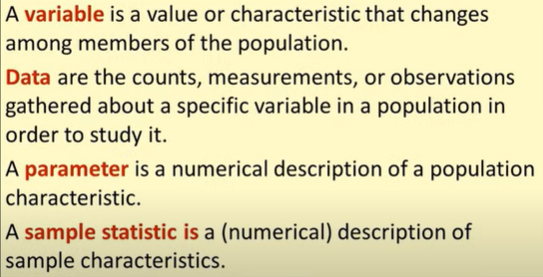
This means there are businesses who have closed after being selected for the survey (ie these entities are no longer in the target population because they have become inactive) and there are new businesses who have started and cannot be surveyed (ie these entities are not in the survey population even though they are active and in the target population).

This difference needs to be accounted for, or else we would never be able to include new businesses in these economic measures.

Differences

* Your target population is theoretical, your sample population is practical, based on the frame you have access to or are able to construct.
* A target population should be more specific than what you've described. Specify exactly the units you mean: for example, full time students studying courses at universities located in the UK on 1st September 2018, excluding those studying by distance not residing in the UK. Note that the specification should be
  + - **CUTE**:
      * **content** (who the survey is about), 🡪 full time students studying courses at universities located in the UK
      * **units** (who you are approaching), 🡪Students
      * **time** (when the survey is about), 🡪 1st September 2018
      * **extent** (where the survey is about) 🡪 not full-time students, those studying by distance not residing in the UK
* The survey and target populations refer to the populations for the entire survey, not individual questions. If you want to ask general questions to all university students, but have some more specific questions for first year students then your survey and target populations are both still all university students, but you may want to control how many first years you survey through stratification to achieve the required accuracy for those questions that specifically concern them. If however you are only interested in first year students but you're sampling other students too, then your target population is just a subset of your sample population and you should try to fix this.





**Value vs Characteristics**

Value 🡪4 Characteristic 🡪 blue

**Population parameter vs. sample statistic**

A parameter is a number describing a whole population (e.g., population mean).

A statistic is a number describing a sample (e.g., sample mean).

You can use estimation or [hypothesis testing](https://www.scribbr.com/statistics/hypothesis-testing/) to estimate how likely it is that a sample statistic differs from the population parameter.

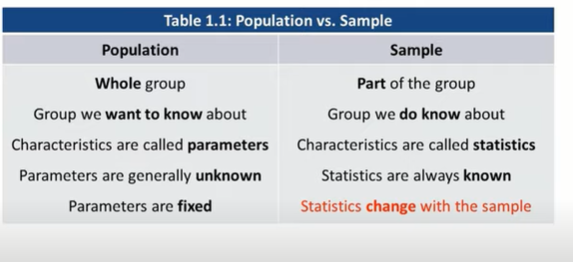
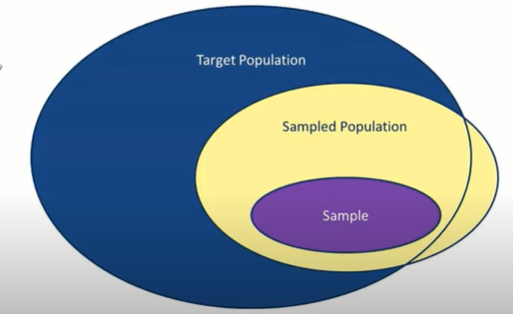
Research example: Parameters and statistics In your study of students’ political attitudes, you ask your survey participants to rate themselves on a scale from 1, very liberal, to 7, very conservative. You find that most of your sample identifies as liberal – the mean rating on the political attitudes scale is 3.2.

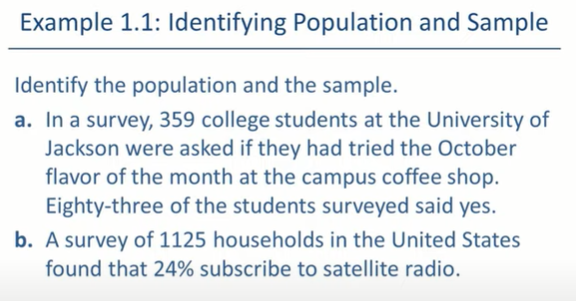
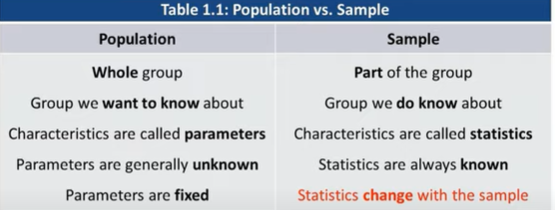
You can use this **statistic**, the sample mean of 3.2, to make a scientific guess about the **population parameter** – that is, to infer the mean political attitude rating of all undergraduate students in the Netherlands.

**Variable vs parameter**

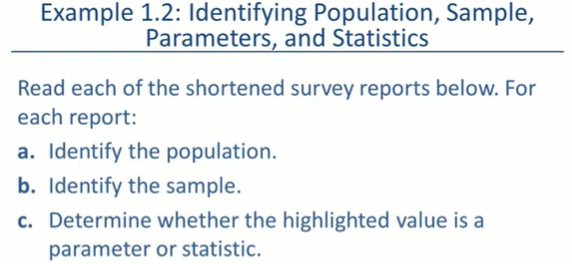
A variable(ROWS) represents a model state, and may change during simulation.

A parameter(COLUMNS) is commonly used to describe objects statically. A parameter is normally a constant in a single simulation, and is changed only when you need to adjust your model behaviour.



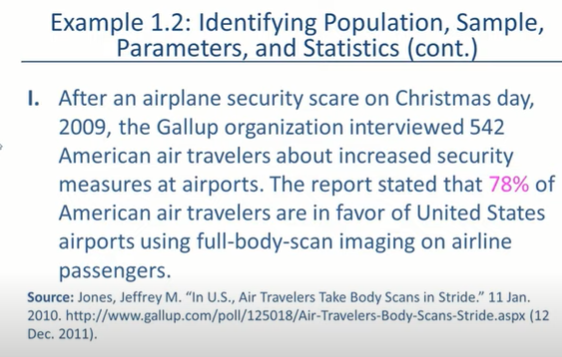
We want sample to be representative of the target population but it’s not always necessary means Sample can go out of the line of the circle of Target population

1. Population 🡪 university of Jackson students, Sample 🡪 359 students whom we contacted
2. Population 🡪 households in US Sample🡪 1125 households in us



Population Sample parameters (depends on population) and

Statistics (depends on Sample)

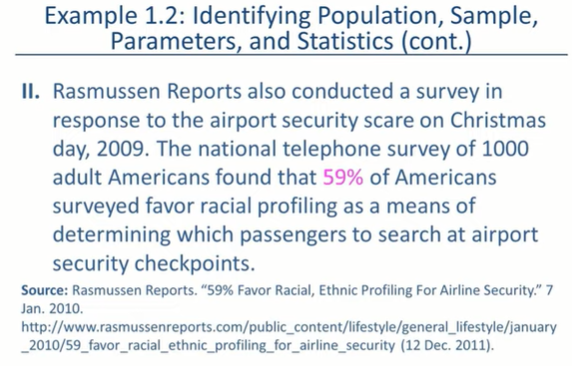


Population 🡪 American air travellers

Sample 🡪 542 American air travellers

Parameter 🡪

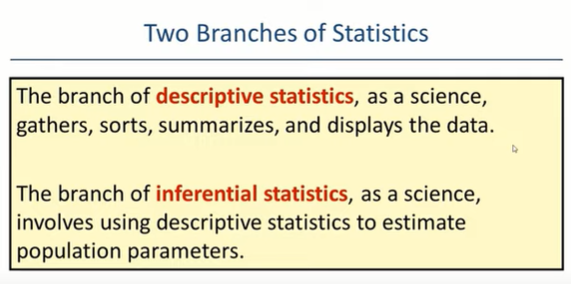
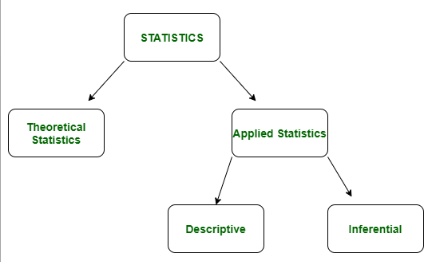
Statistics 🡪78%(as measure out of that Sample)



Population 🡪 Americans

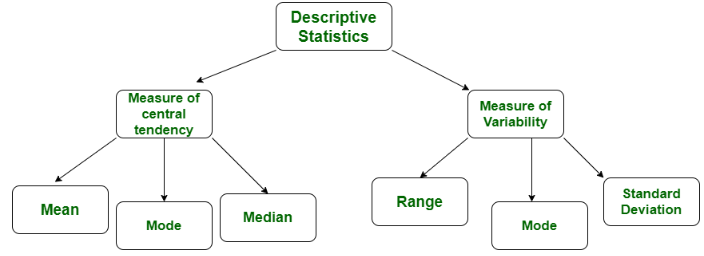
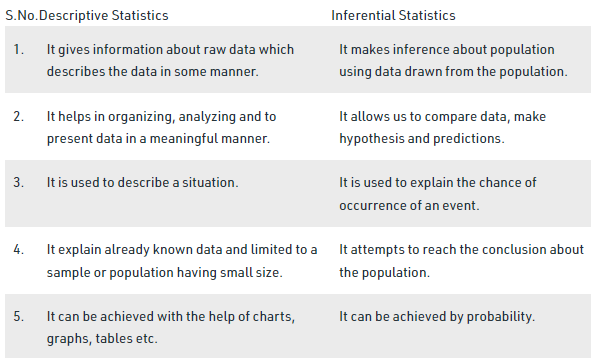
Sample 🡪 1000 Americans

Statistics 🡪59% of the surveyed(as measure out of that Sample)



**Types of Descriptive Statistics –**

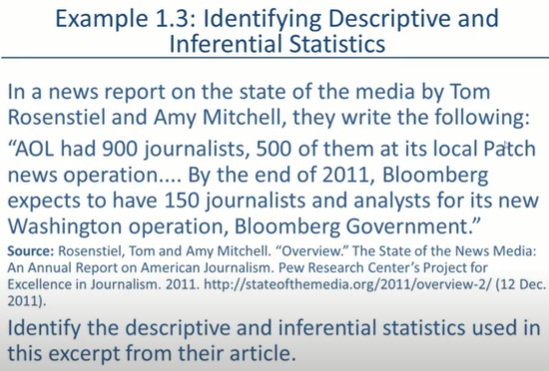
* Measure of Central Tendency
* Measure of Variability

[2. Inferential Statistics](https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/)**:**  
In inferential statistics predictions are made by taking any group of data in which you are interested. It can be defined as a random sample of data taken from a population to describe and make inference about the population. 

Exploratory data analysis (EDA) and confirmatory data analysis (CDA) operate most effectively when they proceed side-by-side. 

<https://online.ndm.edu/news/analytics/exploratory-analysis-vs-confirmatory-analysis/>

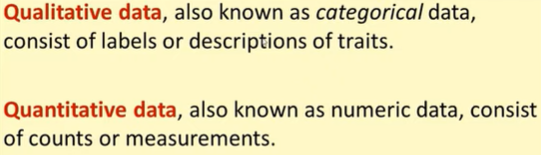
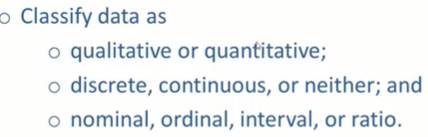
<https://stats.stackexchange.com/questions/89401/what-are-some-differences-between-confirmatory-analysis-and-exploratory-analysis>

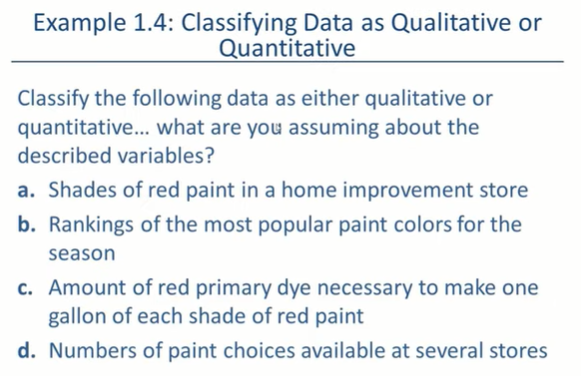


Descriptive 🡪 AOL 900 journalist (fact from data)

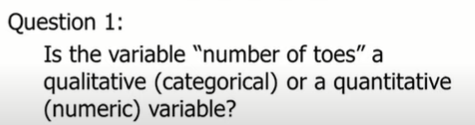
Inferential 🡪By the end of 2011, 150 journalists and analysts (fact that can be derived from data regarding future / regarding population)

DATA CLASSIFICATION

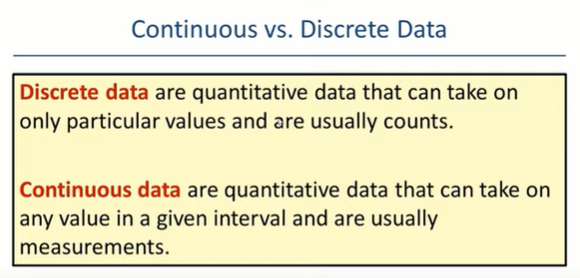


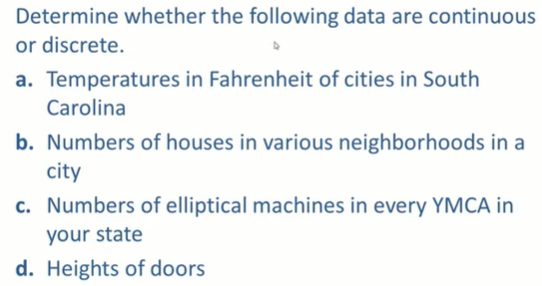
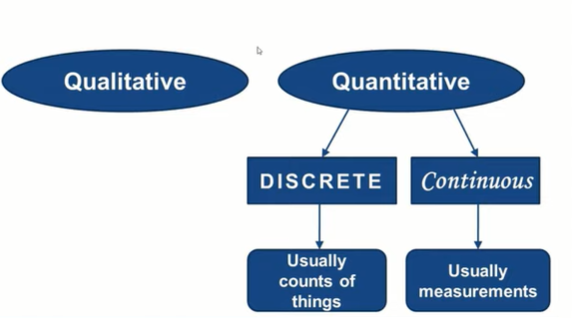


1. Qualitative 🡪 if it’s like baby red,…….etc if it’s like 75% red then quantitative
2. Quantitative(number of stars) Qualitative 🡪 loved…….Hated
3. Quantitative
4. Quantitative



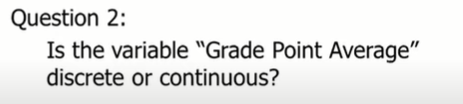
Quantitative

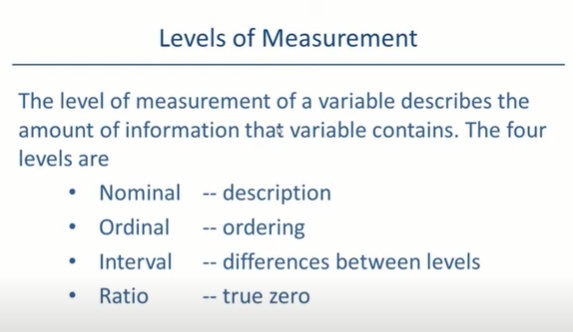
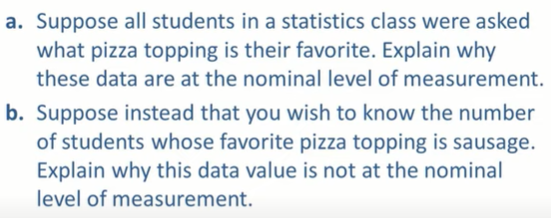




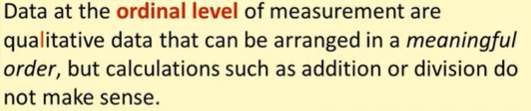
1. continuous
2. Discrete
3. Discrete
4. Continuous

Variable 🡪 What you are measuring , Data 🡪 What you write for that data

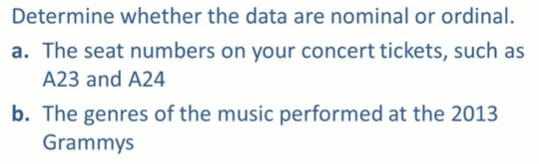


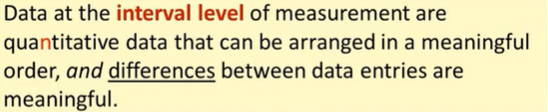
1. As in nominal as it will be a categorical data which can’t be ordered
2. It will not be nominal as number of students is not a actegorical type of data it’s quantitative.



Example 🡪 Socio-economic Status



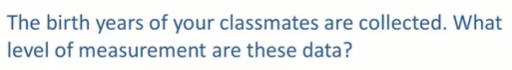
1. Ordinal
2. Nominal 🡪 No inherit ordering

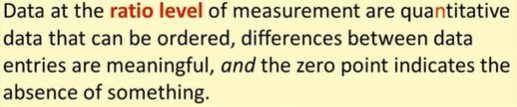


Shoe sizes 🡪

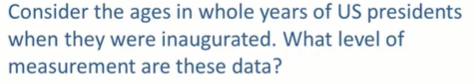
difference between 8 and 9 number of shoe is 1 which is similar to difference between 10 and 9 number of shoe i.e. 1

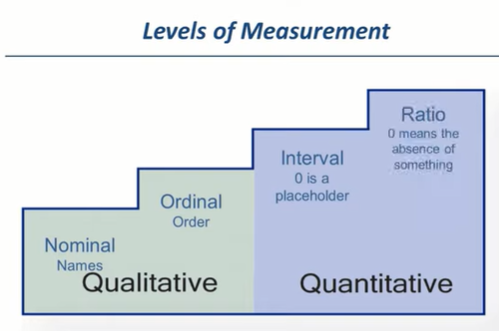
difference between temp i.e 25K and 26K is 1 which is similar to difference between 10K and 9K i.e. 1

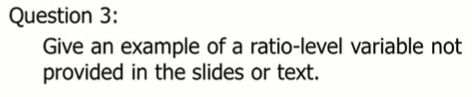
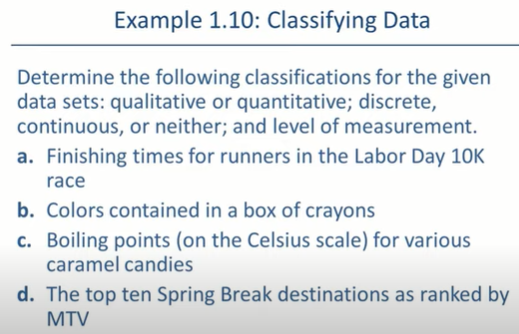




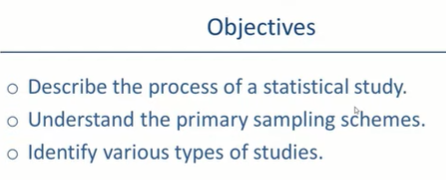
Ratio tell how manys times(multiplication) one thing differes from the another

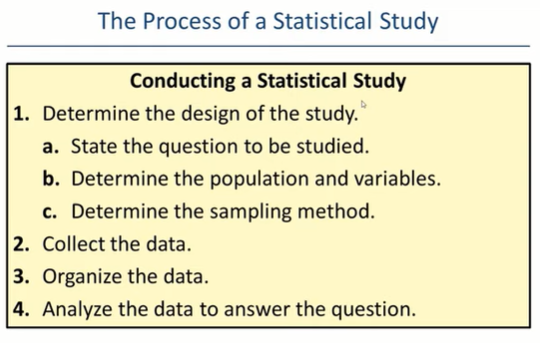
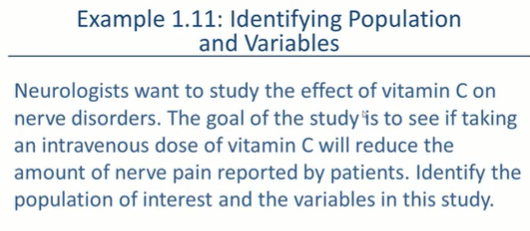




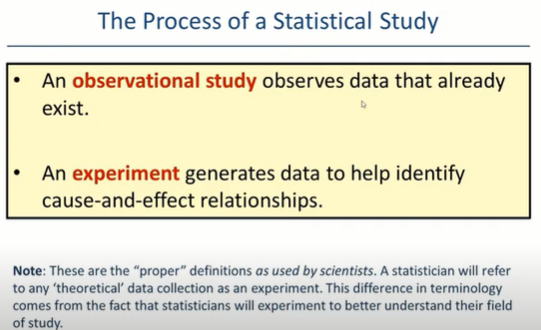
1. Quantitative , Continuous , Ratio
2. Qualitative, Nominal
3. Quantitative , Continuous , Interval
4. Qualitative, Nominal

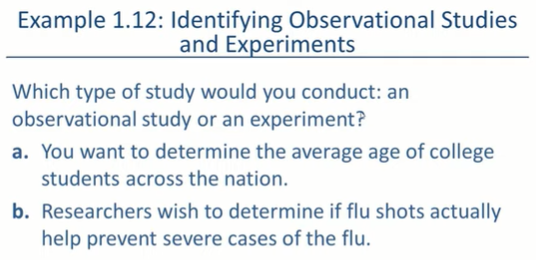
l 

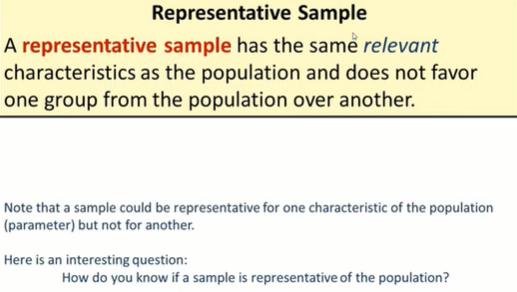
Population of Interest : All patients wit nerve dis orders

Variables in study : amount of Vitamin C and amount of Pain





1. Observational {Here we are observing ages of students}
2. Experimental {Here we need to do something in order to generate data}



Sample could e representative for one characteristic of the population but not for another

**How to check whether your sample is representative or non – representative**



Answer : It’s not representative

### 

**EXAMPLE:**  If we have to select 50 students from washington then posible combination of 50 people has an equal chance of being selected

### **Systematic Sampling**

In the systematic sampling method, the items are selected from the target population by selecting the random selection point and selecting the other methods after a fixed sample interval. It is calculated by dividing the total population size by the desired population size.

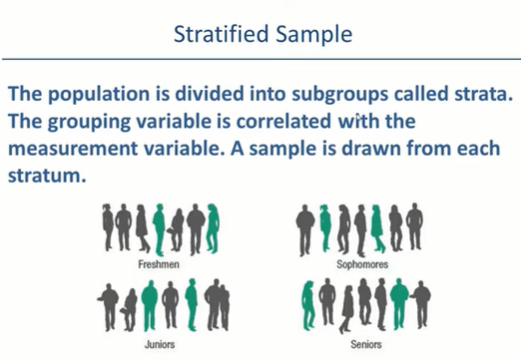
**Example:**

Suppose the names of 300 students of a school are sorted in the reverse alphabetical order. To select a sample in a systematic sampling method, we have to choose some 15 students by randomly selecting a starting number, say 5.  From number 5 onwards, will select every 15th person from the sorted list. Finally, we can end up with a sample of some students.

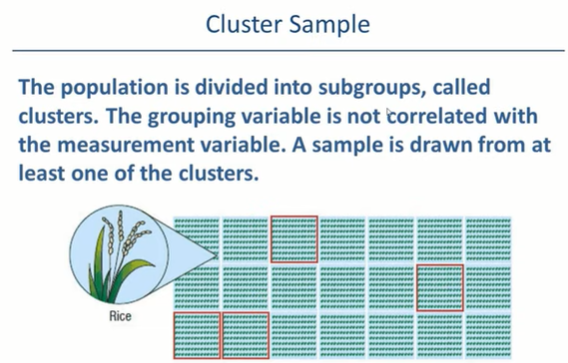
### **Stratified Sampling**

In a stratified sampling method, the total population is divided into smaller groups to complete the sampling process. The small group is formed based on a few characteristics in the population. After separating the population into a smaller group, the statisticians randomly select the sample.

For example, there are three bags (A, B and C), each with different balls. Bag A has 50 balls, bag B has 100 balls, and bag C has 200 balls. We have to choose a sample of balls from each bag proportionally. Suppose 5 balls from bag A, 10 balls from bag B and 20 balls from bag C.



Important 🡪 **grouping variable is correlated with the measurement variable**



Important 🡪 **grouping variable is not correlated with the measurement variable**

In cluster sampling, **researchers divide a population into smaller groups known as clusters.** **They then randomly select among these clusters to form a sample**. Cluster sampling is a method of probability sampling that is often used to study large populations, particularly those that are widely geographically dispersed. Researchers usually use pre-existing units such as schools or cities as their clusters.