Probability

Ms. S. R.Valvi
Asst. Prof.
Dept. of Pharmaceutics
JES's College of Pharmacy, Nandurbar

- The chance of occurrence of an event.
- Predictions of an event.
- It is the ratio of number of favourable outcomes of an event to the number of all possible outcomes.
- Many events can’t be predicted with total certainty. So that in that case we can say how likely the events will happen using the idea of probability.

- Probabilities are expressed as fractions (½, ¼) or as decimals (0.41,0.18)
- Probability can range in from 0 to 1 where,
- 0 means that the event can never happen
- 1 indicates something will always happen

Probability

Example

Tossing a coin

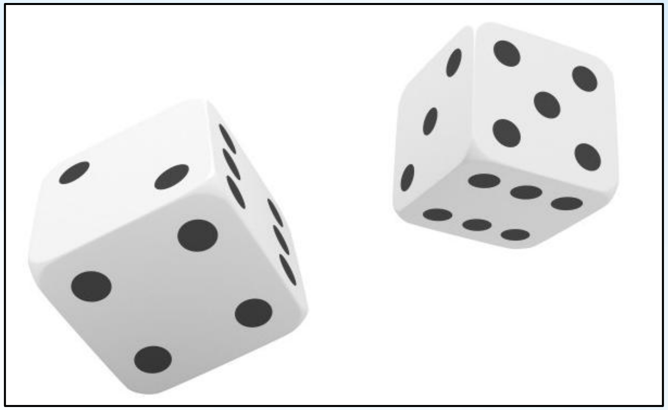
When coin is tossed there are two possible outcomes:

Heads (H) or Tails (T)

We say that the probability of the coin landing H is ½

And the probability of the coin landing T is ½

Probability



Example

Throwing Dice

When a single die is thrown there are six possible outcomes:

1,2,3,4,5,6

The probability of any one of them is 1/6

Formula

$$\text{Probability of event to happen } P(E) = \frac{\text{Number of favourable outcomes}}{\text{Total number of all possible outcomes}}$$

- **Trial & Event:** The experiment is known as trial & outcomes are known as event.
- **Random Experiment:** A series of outcomes where outcomes are always uncertain.
- **Sample space:** It is the set of all possible outcomes of random experiment & it is **denoted by S.**
- **Exhaustive event:** It is the total number of possible outcomes in any trial.

- **Dependent event :** Two events are said to be dependent if the occurrence of one event changes the probability of another event.
- **Independent event :** Two events are said to be independent if the probability of one event that does not affect the probability of another event.

Example:

There are 6 balls in basket, 3 red, 2 yellow, 1 blue. What is the probability of picking a yellow ball

Ans: Yellow : 2 , Total = 6

So, P = 2/6 = 1/3

Probability Distribution

It is a mathematical function that describes the likelihood of obtaining all the possible outcome values that a random variable can assume.

It is of 3 types:

- ✓ Binomial Distribution
- ✓ Poisson’s Distribution
- ✓ Normal Distribution

Binomial Distribution

- It has two possible outcomes, thus it is known as binomial distribution.
- It is the probability of success or failure outcomes in an experiment.

Example :

Adults with allergies might report relief with medication or not.

So, the outcome has two possible ways.

Assuming the coin is tossed once, there can be two possibilities, either head (success or **p**) or tail (failure or **q**)

Probabilities of 2 heads (or 2 successes)

$= p \times p = p^2$

Probabilities of one head and one tail

$= (p \times q) = pq$

Probabilities of one tail and one head

$= (q \times p) = qp$

Probabilities of 2 tails (or 2 failures)

$= q \times q = q^2$

Conditions necessary for Binomial Distribution

- It is necessary that each observation is classified into two categories such as success or failure.

Binomial distribution is denoted by P(r)

$$P(r) = {}^{n}C_{r}\, q^{n-r}\, p^{r}$$

Where,

p = probability of success in a single trial

q = probability of failure = $1-p$

n = number of trials

r = number of successes in n trials

$n-r$ = failures out of n trials

${}^{n}C_{r} = [n! / r!\,(n-r)!]$

Factorial function **(symbol !)** : Multiply all whole numbers from chosen number down to 1.

Examples:

4! = 4 × 3 × 2 × 1 =24

1! = 1

Find the chances of getting 3 successes in 5 trials when the chance of getting a success in one trial in 2/3.

Solution: Here, n= 5, p = 2/3, r = 3 , q = 1-p = 1-2/3 = 1/3

$$P(r) = {}^{n}C_{r}\, q^{n-r}\, p^{r} = {}^{5}C_{3}\, (1/3)^{5-3}\, (2/3)^{3}$$

$$= \frac{5!}{3!\,(5-3)!} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3}$$

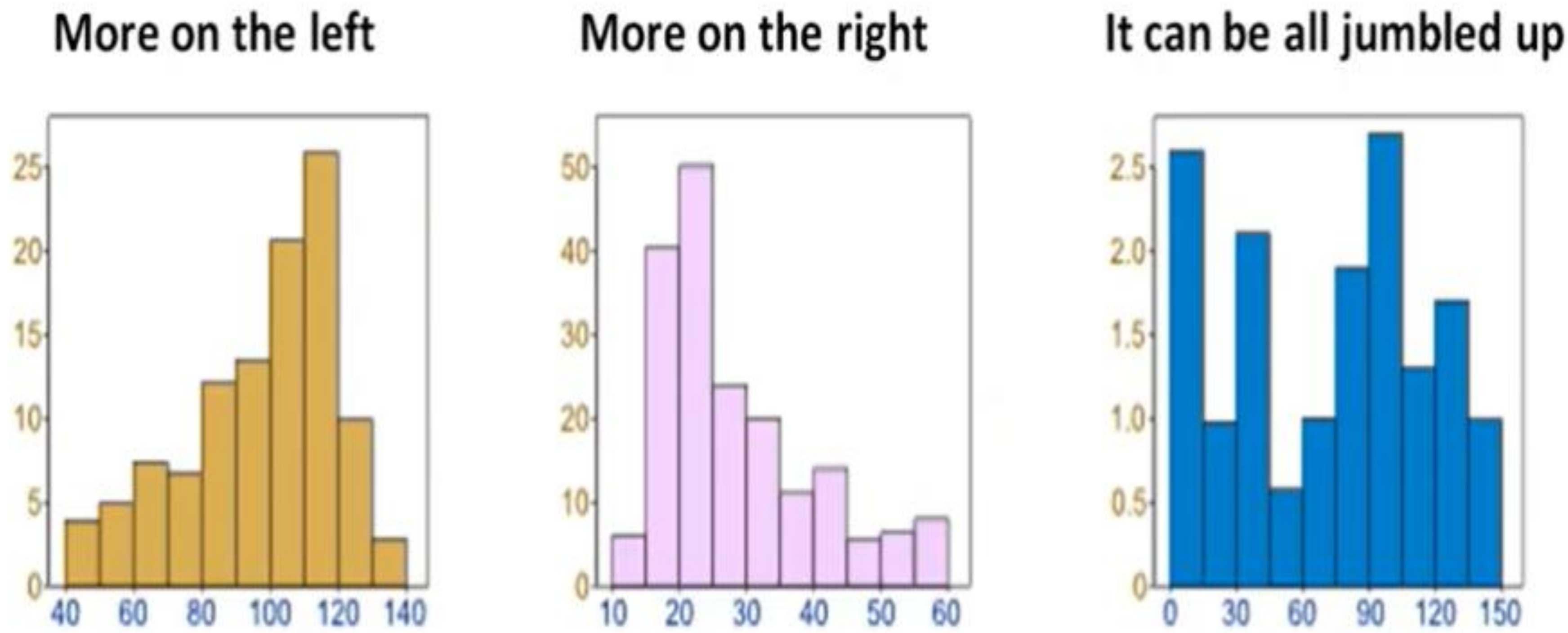
$$= \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3}$$

$$= \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{80}{243} = 0.33$$

Normal Distribution

- It describes how the values of variable are distributed.
- It is also known as the Gaussian distribution.
- In graph normal distribution appears as a bell shape curve, thus it is also known as bell curve.
- It is a symmetric distribution where most of the values cluster around the central peak.

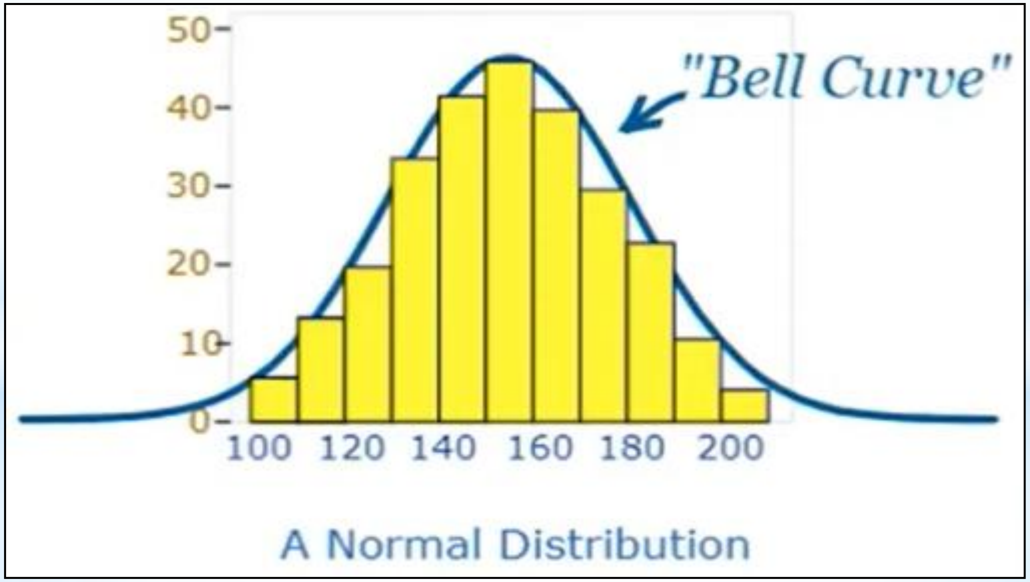
➢ Data can be "Distributed" (spread out) in different ways



Graph of Normal Distribution:

It depends on two factor;

- **Mean :** Determines the location of the center of the graph.
- **Standard Deviation :** Determines the height & width of the graph.



Properties of Normal Distribution:

- The mean, mode and median are equal.
- The curve is symmetric at the center.
- Half of the values fall on the left of center and half of the values on the right side of the center.
- Total AUC is equal to 1.

Standard units or Standard scores or Z scores

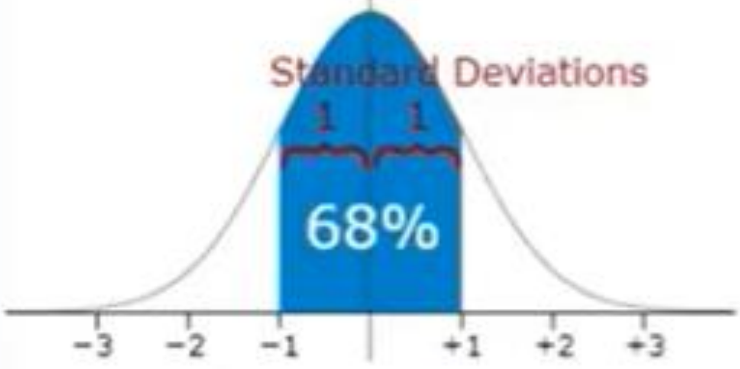
The number of standard deviations from the mean is called the **standard units** or **standard score** or **z-score.**

The units are denoted by **Z** & are called as **z values** or **z scores.**

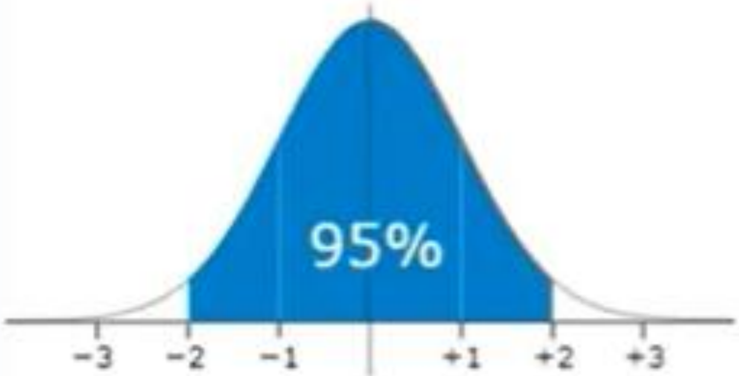
$$z = \frac{x - \mu}{\sigma}$$

Where,

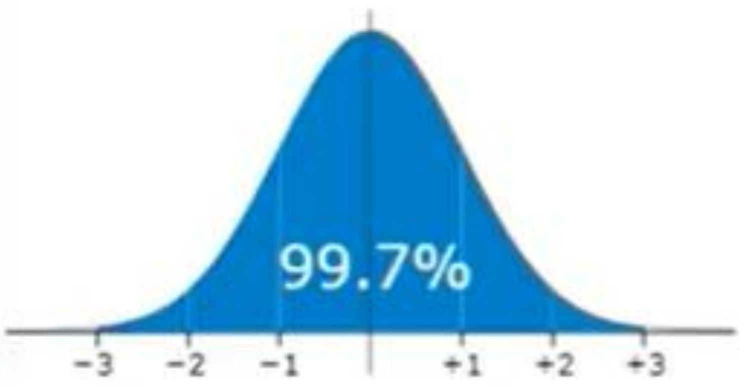
X = Value to be standardized



68% of values are within
1 standard deviation of the mean



95% of values are within
2 standard deviations of the mean



99.7% of values are within
3 standard deviations of the mean

- Example: The test score is 190. The test has a mean of 130 and a standard deviation of 30. Find the z score. (Assume it is a normal distribution)

Given test score x = 190
Mean, μ = 130
Standard deviation, σ = 30

$$z = (x - \mu)/\sigma$$
$$= (190 - 130)/30$$
$$= 60/30$$
$$= 2$$

Hence, the required z score is 2.

Example : 2

- On a final examination of pharmacy student, the mean and S.D. was 50 and 11 respectively. Obtain the standard scores of students receiving grades (a) 60, (b) 75, (c) 90.

We have $\mu = 50$ and $\sigma = 11$

Poisson's Distribution

- It was developed by a French Mathematician **S. D. Poisson,** in 1837, hence name after it.
- It is based on the same assumptions of the binomial distribution.
- It counts the average number of success in given unit of time or space.
- It is used in,

- Quality control statistics to count the number of defective items.
- The number of patients arriving to consult a doctor in a given time period.

Formula

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Where,

P(x) = Probability of x occurrences

λ^x = Lambda (i.e. the mean number of occurrences per interval of time) raised to the x power

$e^{-\lambda}$ = 2.71828 (being the base of the natural logarithm system), raised to the negative lambda power

X! = x factorial

Example : 1

- Suppose, we have a production process of some item that is manufactured in large quantities.
- We find that, in general, the proportion of defective items is p = 0.01, A random sample of 100 items is selected. What is the probability that there are **2 defective items** in this sample?

λ (Mean number of occurrence)
= np = 100 X 0.01 = 1.0,
x = 2

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$P(x) = \frac{(1)^2(2.71828)^{-1}}{2\ X\ 1}$$

$$P(2) = \frac{(1)^2 X\ 0.36788}{2}$$

$$P(2) = 0.18394$$

Example : 1

- Suppose, we have a production process of some item that is manufactured in large quantities.
- We find that, in general, the proportion of defective items is p = 0.01, A random sample of 100 items is selected. What is the probability that there are **2 defective items** in this sample?

λ (Mean number of occurrence)
= np = 100 X 0.01 = 1.0,
x = 2

Important terms

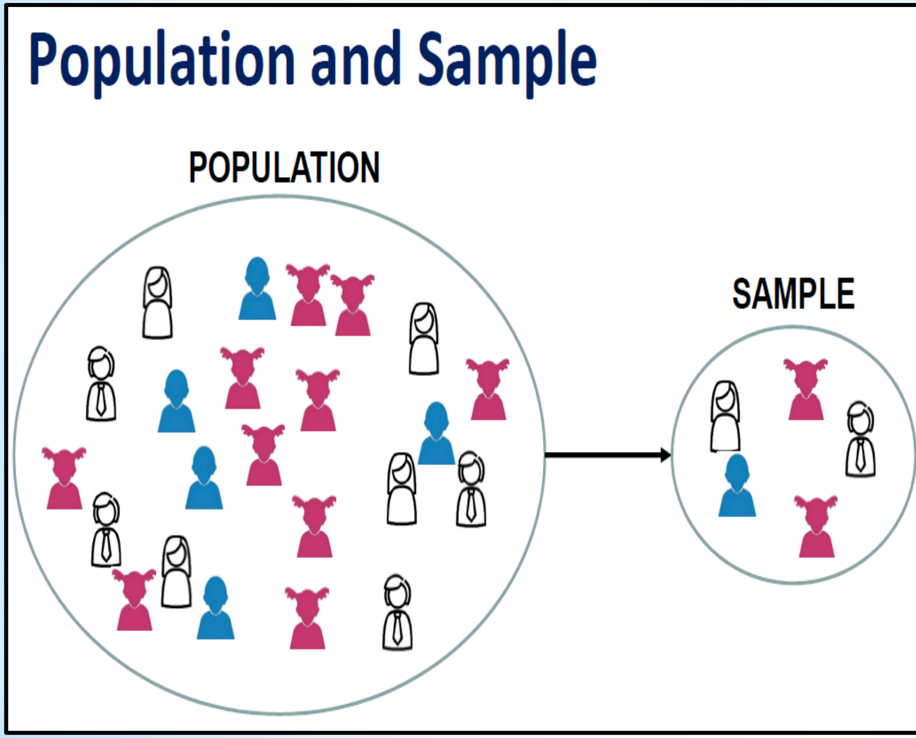
Population

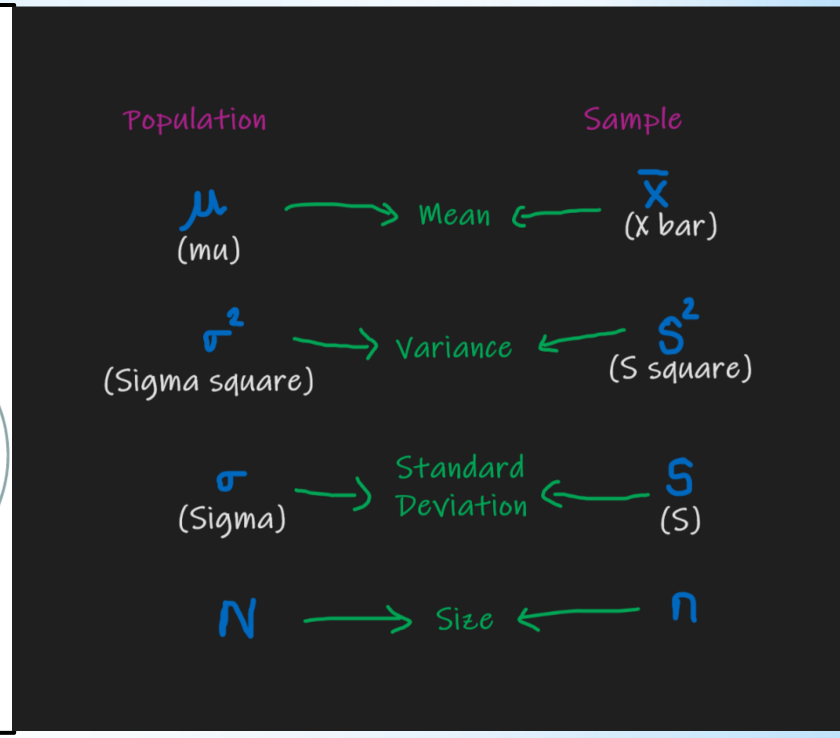
- A complete set of elements that possess some common characteristics.
- It is a group phenomenon that have something in common.

- **Target population: e.g.** All school age children with asthma.

- **Accessible population: e.g.** All school age children with asthma in Gujarat.

- **Sample & Sample size:** A finite subset of statistic individual in a population is called as a sample & the total number of an individual in sample is called as a sample size.
- **Sampling:** A process of selecting a portion of population to represent the entire population.
- **The process of selecting a group of** people or other elements with which to conduct a study.


Population and Sample
POPULATION
SAMPLE


Population
Sample
μ (mu) → Mean ← x̄ (x bar)
σ² (Sigma square) → Variance ← S² (S square)
σ (Sigma) → Standard Deviation ← S (S)
N → Size ← n

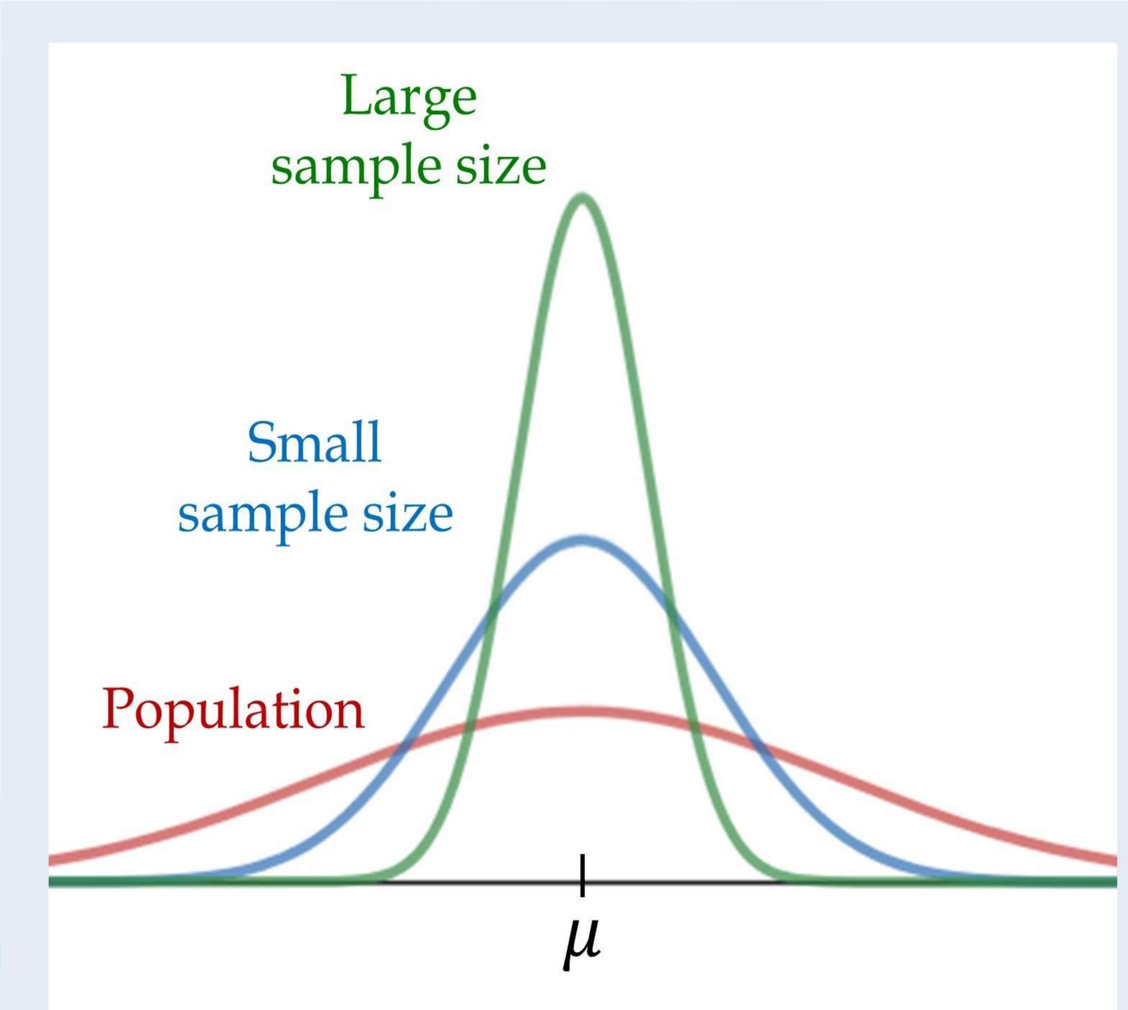
- **Large Sample:** When sample size n is greater than 30 ($n \geq 30$), it is known as large sample.

Generally, selecting larger sample is good, because larger sample provide more accurate mean values & identify outliners.

While the study of sampling distribution of statistics for larger sample is known as large sample theory.

- **Small Sample:** When sample size n is smaller than 30 ($n \leq 30$), it is known as small sample.

While the study of sampling distribution of statistics for small sample is known as small sample theory.


Large
sample size
Small
sample size
Population
μ

Objectives of Sampling

- To study selected units instead of the whole universe.
- To get results in a short period of time.
- To get essential information at low cost.
- To minimize sampling variance.
- To get real & error free conclusion.

Essence of Sampling

It is the selection of a part of sample from the whole population in order to make inference about the whole.

- It must possess same characteristics of the original population.
- For result reliability sample should be enough in number.

Advantages of Sampling

1. Economy
2. Time-lag
3. Scope
4. Accuracy of data
5. Convenience
6. Reliability

Limitations of Sampling

- Errors due to sampling may be high for small sample size.
- Not feasible for problems that require high accuracy.
- When sampling units of population are not homogenous, the sampling technique will be unscientific.

Thank You!!!!

Ms. Saroj. R. Valvi

Asst. Professor

Dept. of Pharmaceutics

JES's College of Pharmacy, Nandurbar

- It is used to estimate unknown value of one variable from known variable of the related variable is called as regression.
- It is used to determine the relationship between two or more variables.
- **It measures the nature and extent of relation between two or more variables, thus enables us to make predictions.**

Correlation	Regression
Relationship	One affects other
Variables move together	Cause & effects
x & y can be interchanged	x & y cannot be interchanged

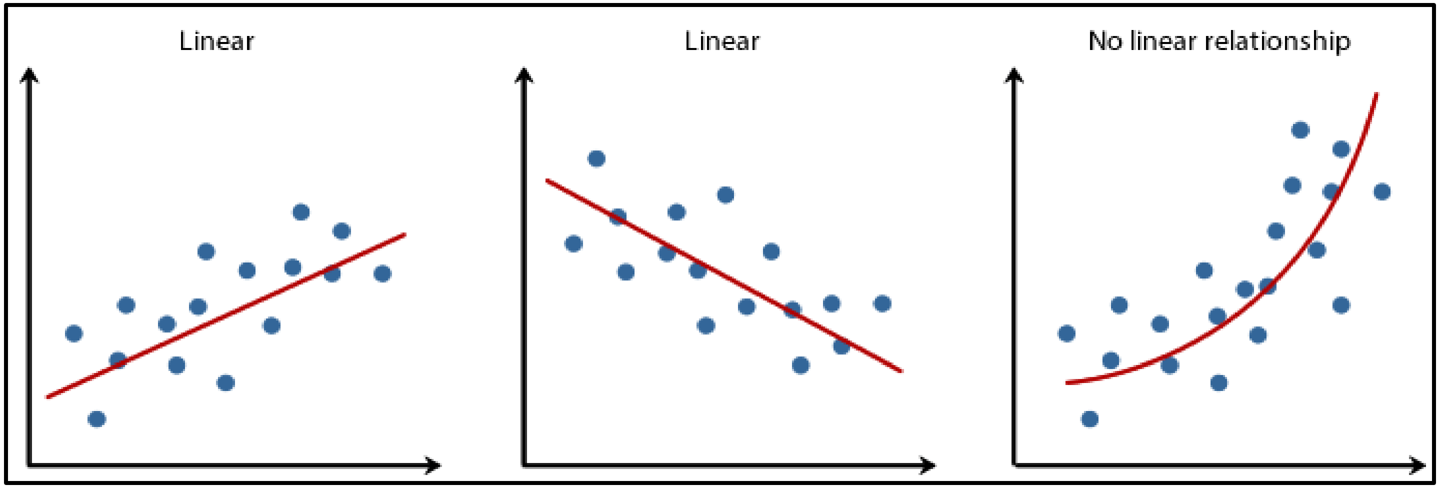
Types of Regression

1) Simple linear regression:- Uses one independent variable to predict the outcome.

2) Multiple linear regression:- Uses two or more independent variables to predict the outcome.

3) Linear regression:-

4) Non-linear regression (Curve linear regression):-



Properties of Regression

1) Is illustration of response variable as a function of predictor value.
2) It predicts only a probable value of the response on a known value of the predictor.
3) Regression coefficient is used to work out regression equation.
4) It is estimated when there is significant correlation between the response & predictor value.

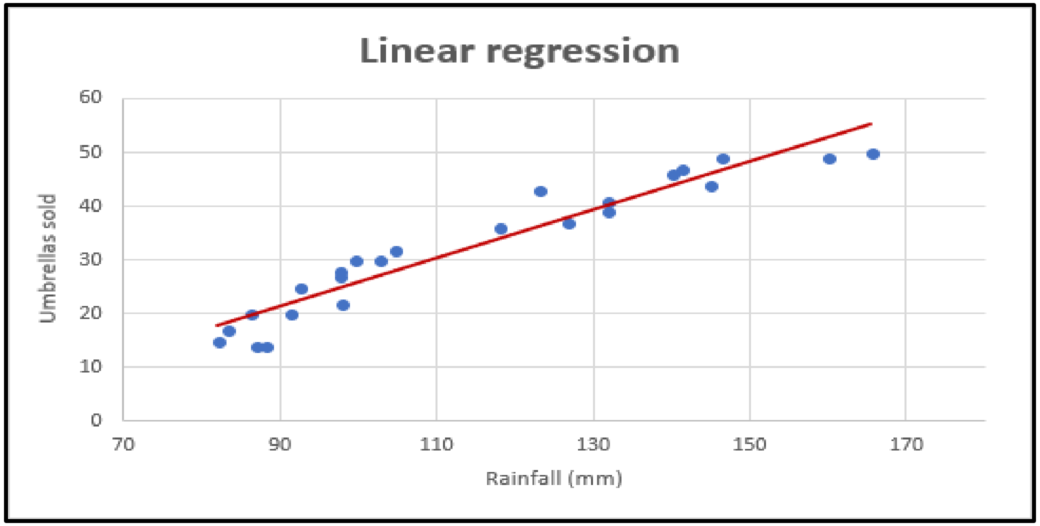
APPLICATION OF REGRESSION IN PHARMACY

1) Study of drug dose & response relationship over time.
2) In clinical trials for fitting linear portions of pharmacokinetics data.
3) In development of biochemical & chemical assay.
4) Calibration of analytical data for the quantitative analysis.
5) To study stability predictions.

Regression Analysis

- It is the mathematical measure of average relationship between two or more variables in terms of the original units of the data.
- In regression analysis there are two types of variables.
- The variable whose value is influenced is called **dependent variable (y) and is also known as regressed or explained variable.**
- The variable which influences the values is called as **independent variable (x) and is also known as regressor or predictor or explanatory variable.**

- **Regression line**
- **Regression equation**



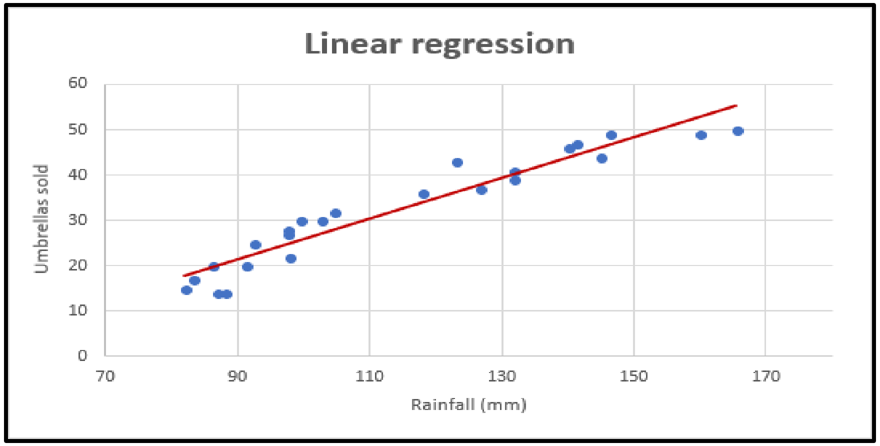
y = a + bx

Here, the aim is to predict the regression of the relation of above equation i.e. regression of y on x where y is treated as dependent & x is independent variable.

Regression line is the straight line that describes how a response variable (y) changes as a explanatory variable(x)

CURVE FITTING – METHOD OF LEAST SQUARE

- Least square method is a mathematical procedure for finding the best fit curve to a given set of data points by minimizing the sum of the squares of the offsets of the points from the sets.
- **Line of best fit**
- Try to have line as close as possible to all points.
- But for better accuracy lets see how to calculate the line using least square regression method.



Least Square Regression

$$\mathbf{y = a + bx}$$

Our aim is to calculate value of **a (intercept)** and **b (slope)**

Where,

y = Dependent variable

x = Independent variable

a = Intercept

b = Slope

Steps

To find the line of best fit for N observations

1) For each x & y variables, calculate x^2 & xy
2) Sum all x^2, x, y & xy i.e Σx, Σy, Σx^2 & Σxy
3) Calculate slope b :

$$\boxed{\frac{n(\Sigma xy)-(\Sigma x)(\Sigma y)}{n(\Sigma x^2)-(\Sigma x)^2}}$$

4) Calculate slope a :

$$\boxed{=\frac{\Sigma y-m(\Sigma x)}{n}}$$

5) Assemble the equation of line

$$y = a + bx$$

Example:

Sam found how many hours of sunshine vs how many ice creams were sold at the shop from Monday to Friday:

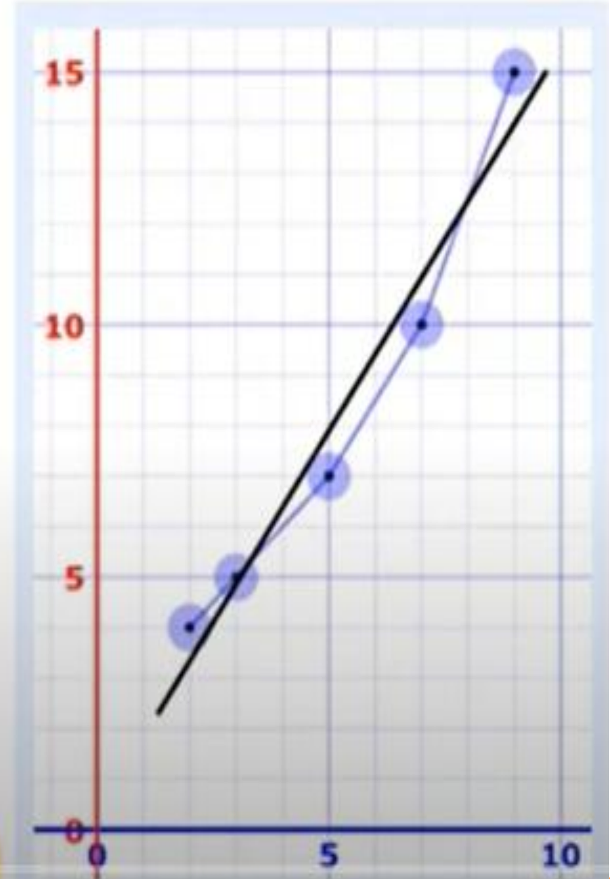
X (hours of ice sunshine)	Y (ice cream sold)
2	4
3	5
5	7
7	10
9	15

Find $y = a + bx$

y = 0.305 + 1.518x

- Here are the (x,y) points and the line **y = 0.305 + 1.518x** on a graph:

x	y	y = 1.518x + 0.305
2	4	3.34
3	5	4.86
5	7	7.89
7	10	10.93
9	15	13.97



am hears the weather forecast which says "we expect 8 urs of sun tomorrow", so he uses the above equation to estimate that he will sell

y = 1.518 x 8 + 0.305 = 12.45 Ice Creams

- Compute the two regression equations on the basis of the following information:

	X	Y
Mean	40	45
Standard deviation	10	9

- Karl Pearson's Correlation coefficient = 0.5
- Also estimate the value of Y for X = 48, using the appropriate regression equation.

Regression equation of Y on X is

$$Y - \bar{Y} = r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

$Y - 45 = 0.5\ (9/10)\ (X - 40)$

$Y = 45 + 0.45\ (X - 40)$

$Y = 45 + 0.45X - 18$

$Y = 27 + 0.45X$

Regression equation of X on Y is

$$X - \bar{X} = r\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$$

$X - 40 = 0.5\ (10/9)\ (Y - 45)$

$X = 40 + 0.556\ (Y - 45)$

$X = 40 + 0.556Y - 25.02$

$X = 14.98 + 0.556Y$

The value of Y for X = 48, we have to use the regression equation of Y on X

$Y = 27 + 0.45X$

$Y = 27 + 0.45(48)$

$Y = 27 + 21.6$

$Y = 48.6$

Standard error of regression or Standard error of estimate (Se)

- It is the measure of the spread of the observed values from the estimated one.
- **SD** measures variation in the set of data from its mean, similarly, **Se** measures variation in actual values of Y from the predicted values of Y on the regression line.
- It is valid for linear as well as non linear regression models.
- It is important in calculation of confidence and prediction intervals.

Standard error of regression measures the variability, or scatter of the observed values around the regression line.

$$s_e = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}}$$

- y = Values of the dependent variable or actual or observed value of variable
- ŷ = estimated values from the estimating equation that correspond to each y value
- n = number of data points used to fit the regression line
- S_e = Standard deviation of regression of Y values on X values

- Now, we will find the Standard Error for ...
- y = 0.305 + 1.518 x

x	y	$\hat{y} = 1.518x + 0.305$	$y - \hat{y}$	$(y - \hat{y})^2$
2	4	3.34	0.66	0.435
3	5	4.86	0.14	0.012
5	7	7.90	-0.9	0.81
7	10	10.93	-0.93	0.865
9	15	13.97	1.03	1.061
				$\Sigma = 3.183$

$$s_e = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}$$

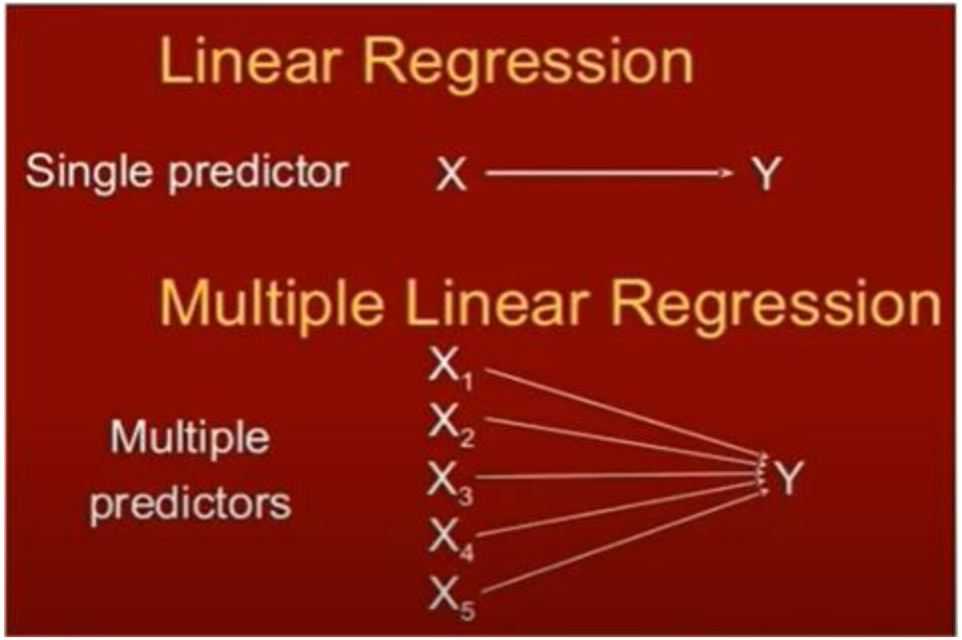
$$s_e = \sqrt{\frac{3.183}{5-2}}$$

$$s_e = \sqrt{1.061}$$

$$s_e = 1.03$$

Multiple Regression

- When there are two or more independent variables, the relationship between them is called as multiple correlations, and the equation that describes such relationship is known as the **multiple regression equation.**



$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_kx_k$$

- Where
  - y = dependent variable which is to be predicted
  - a = intercept
  - x_1, x_2, x_3, x_k = values of the independent variables
  - b_1, b_2, b_3, b_k = slopes associated with x_1, x_2, x_3, x_k respectively
  - k = number of independent variable

Applications of Multiple Linear Regression

- To predict the trends and future values.
- To forecast the effects or impacts of changes.
- Helps to understand how much will the dependent variable change when we will change independent variables.

Thank You!!!!

Sampling Method

Ms. Saroj. R.Valvi
Asst. Prof.
(Department of Pharmaceutics)
JES's College of Pharmacy, Nandurbar

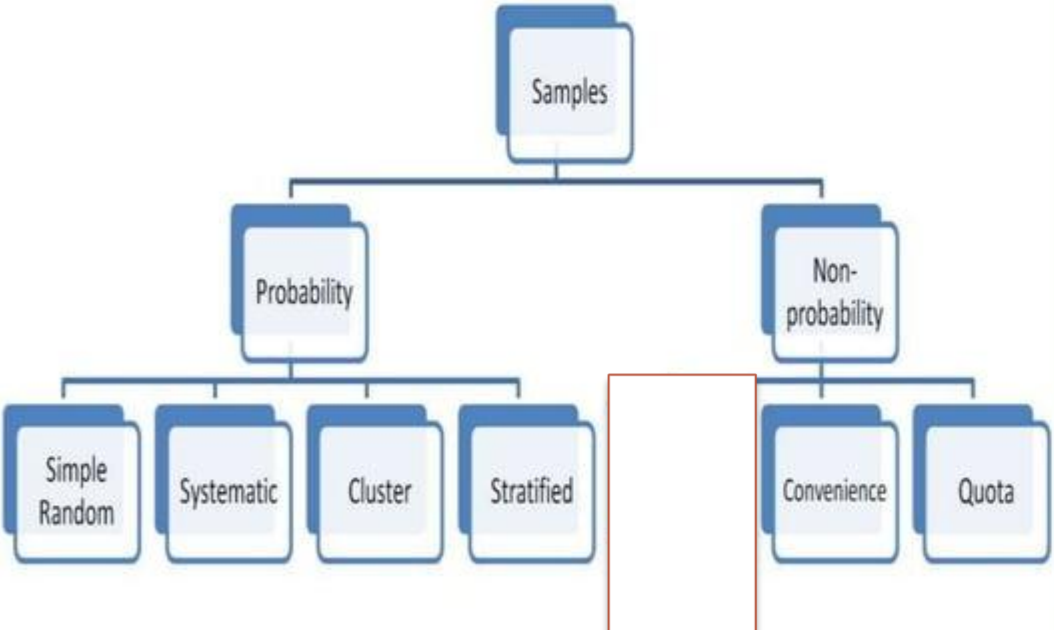
Definition of sample

- A sample is a smaller collection of units from a population used to determine truths and facts about that population.

or

- A small group of people taken from a larger group and used to represent the larger group is called sample.

Classification of Sampling Techniques



Probability Sampling

- A probability sampling is one that have been selected in such away that every element chosen has a known probability of being included.

or

- Probability sampling involves the selection of elements from the population using random in which each element of the population has an equal and independent chance of being chosen.
- It is also called random sampling.

Characteristics of Probability Sampling

1. It refers from the sample as well as the population.
2. Every individual of the population has an equal probability to be taken into the sample.
3. It may be representative of the population.

1. Simple random sampling

- It is, in which each element of the population has an equal independent chance of being included in the sample.
- Thus a sample selected by randomization method is known as simple random sampling.
- In simple random sampling , most commonly used method is the ‘’lottery method”.

Selection process

- By using the lottery method, there is a need for complete listing of the member of population.
- The numbers of all members are written on piece of paper and placed in a container.
- The researcher draws the desired number of sample from container.
- This process is relatively easy for small population but relatively difficult and time consuming for a large population.

2. Stratified random sampling

- The population is divided into two or more groups called strata, on the basis of some characteristics such as geographic location, age, income or status and sub samples are randomly selected from each strata.

Selection process

- Identify and define the population.
- Determine the desired sample size.
- Identify the subgroups i.e. (Strata) for which we want to guarantee appropriate representation.
- Classify all members of the population as members of one of the identified subgroups.

3.Systematic random sampling

- It is the type of probability sampling, in which one or two items are selected randomly, but other items are selected by adding the average sampling interval to the item selected randomly.
- It is also called an Nth name selection technique.
- Selecting every nth subject from a list of the member of the population.

Selection process

- Identify and define the population.
- Determine the desired sample size.
- Obtain the list of the population.
- Determine what nth is equal to by dividing the size of population by the desired sample size .
- Start at some random place in population on list .
- Take every nth individual on the list.

4.Cluster random sampling

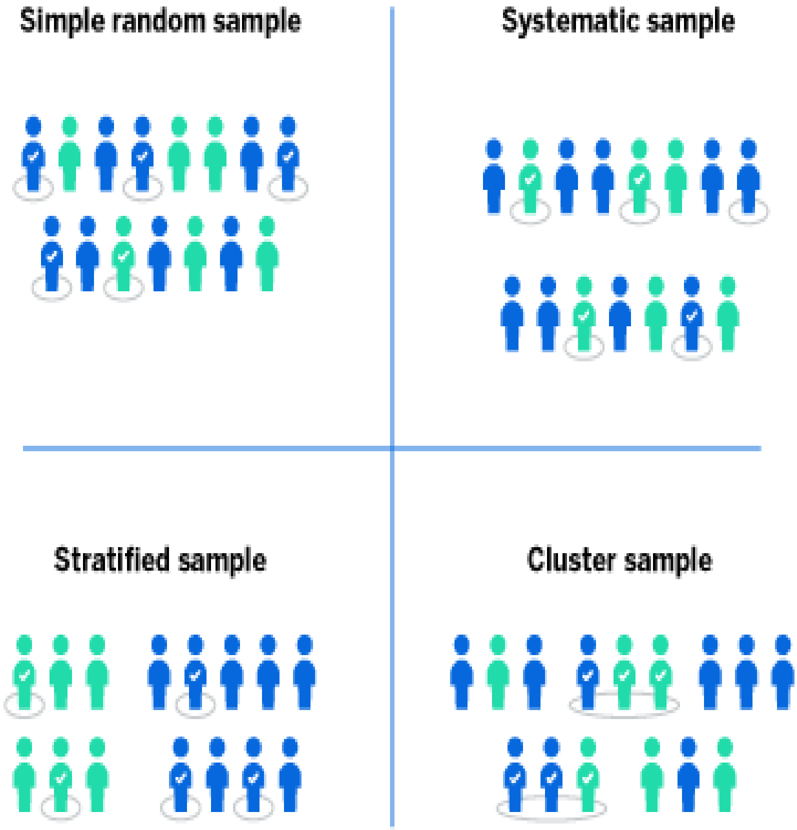
- The process of randomly selecting groups, not individuals, within the define population sharing similar characteristics.
- Cluster are location with in which group of member of the population can be found.
- Example:
- Schools
- Classrooms etc

Cluster Random Sampling

- Cluster random sampling is done when simple random sampling is almost impossible because f the size of the population.
- Just imagine doing simple random sampling when population in question is the entire population of Asia.

Selection process

- Identify and define the population.
- Determine the desired sample size.
- Identify and define the cluster.
- Estimate the average number of population member per cluster.
- Determine the number of cluster s need by dividing the sample size by the estimated size of a cluster.
- Randomly select the needed number of clusters.
- Include in the study all individuals in each selected clusters.


Simple random sample
Systematic sample
Stratified sample
Cluster sample

Non Probability Sampling

- Any sampling method where some elements of population have no chance of selection (these are sometimes referred to as 'out of coverage'/'under covered'), or where the probability of selection can't be accurately determined. It involves the selection of elements based on assumptions regarding the population of interest, which forms the criteria for selection. Hence, because the selection of elements is nonrandom, nonprobability sampling not allows the estimation of sampling errors.

Non Probability Sampling

- **Example:** We visit every household in a given street, and interview the first person to answer the door. In any household with more than one occupant, this is a nonprobability sample, because some people are more likely to answer the door (e.g. an unemployed person who spends 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.

Quota Sampling

- The population is first segmented into mutually exclusive sub-groups, just as in stratified sampling.
- Then judgment used to select subjects or units from each segment based on a specified proportion.
- For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.
- It is this second step which makes the technique one of non-probability sampling.
- In quota sampling the selection of the sample is non-random.

Quota Sampling

- For example interviewers might be tempted to interview those who look most helpful. The problem is that these samples may be biased because not everyone gets a chance of selection. This random element is its greatest weakness and quota versus probability has been a matter of controversy for many years.

Convenience Sampling

- Sometimes known as **grab** or **opportunity sampling or accidental or haphazard sampling.**
- A type of nonprobability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, readily available and convenient.
- The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough.

Hypothesis

Null Hypothesis	Alternative Hypothesis
States that no relationship between 2 variables	States that there is a relationship between 2 variables
Denoted by **Ho**	Denoted by **Ha**
Followed by equals to sign	Followed by less than or greater than sign
Researcher tries to disapprove things	Researcher tries to prove things
Indicates no change in opinions	Indicates change in opinions
If accepted, results become insignificant.	If accepted, results become significant.

Sampling Error

- It is a statistical error that occurs when the researcher does not select a sample that truly represents the whole population.
- Non-sampling error arises due to faulty data collection process.
- It can be reduced by increasing the number of sample or by randomizing the sampling process.

Sampling Error

Type I error: Occurs when statistician rejects a null hypothesis that is actually true in population. This is also called as false-positive.

Type II error: Occurs when statistician accepts a null hypothesis that is actually false in population. This is also called as false-negative.

Non Sampling Error

- **Population specification error:** When we do not understand population.
- **Sample frame error:** When wrong sub population is used to select a sample.
- **Selection error:** When respondents self select their participants.
- **Sampling error:** Occurs because of variation in the number or characteristics of sample that responds.

Sampling Error

Type I	Type II
Caused by rejecting null hypothesis, when it is true.	Caused by accepting null hypothesis, when it is false.
Denoted by **α**	Denoted by **β**
Refers as false positive	Refers as false negative
False rejection of true hypothesis.	False acceptance of incorrect hypothesis.
Caused by luck or chance.	Caused by smaller sample size.

Standard Error of Mean (SEM)

- It is also called as SD of mean.
- It is used to estimate sample mean dispersion from population mean.
- It gives accuracy of sample mean.
- It is always smaller than the standard deviation.

$$SE = \frac{\overset{\text{Standard deviation}}{\sigma}}{\underset{\text{Number of samples}}{\sqrt{n}}}$$

Thank You!!!!