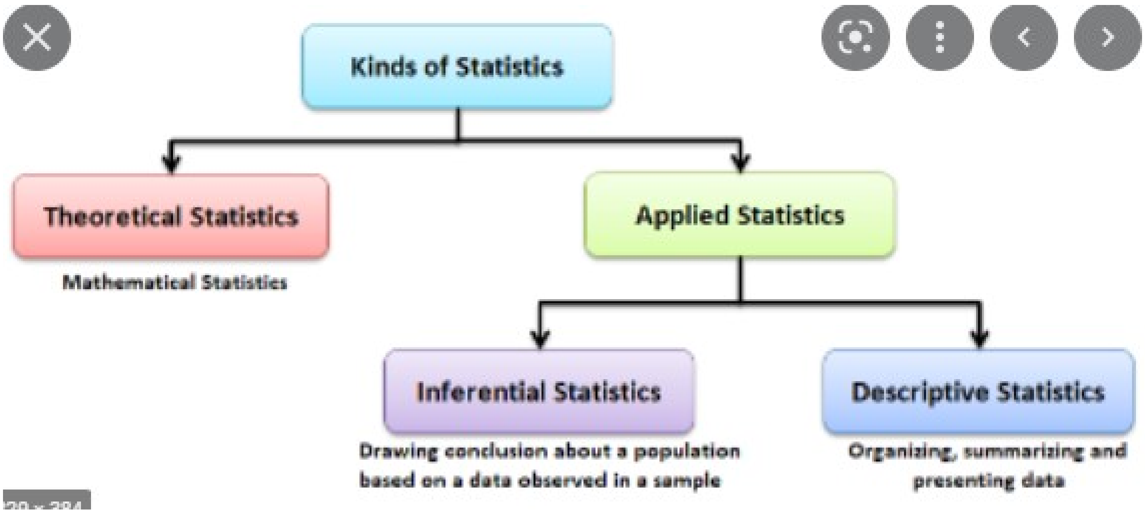
WHAT IS STATISTICS?

In one sense *statistics* refers to a collection of numerical facts. statistics are observations organized into numerical form. प्राचीन काल में स्टैटिसटिक्स का उपयोग आयु (Age)) लिंग(gender) पर आधारित जनसंख्या (population) तथा राष्ट्र के धन एवं सम्पति से संबंधित प्रदत (data) के कलेक्शन के लिए किया जाता था

What is statistics?" is that it is a set of methods for dealing with numerical facts (analysis) for understanding or making sense out of it

Psychologists, like other scientists, refer to numerical facts as *data*. The word *data* is a plural noun and always takes a plural verb, as in "the data *were* analyzed." (The singular form, datum, is rarely used)

This text, however, deals only with *applied statistics*. It describes methods for data analysis that have been worked out by statisticians, but does not show how these methods were derived from more fundamental mathematical principles. For that part of the story, you would need to read a text on *theoretical* or *mathematical statistics*



Parametric test (t test, ANOVA, regression) Central tendency (mean, median, mode)

Non parametric test(chi square test) Variability

One part of applied statistics is concerned only with summarizing and organising .(to bring order to this chaos) the set of data using tables, graphs, and summary measures. so that we can communicate and describe their important characteristic, this is called descriptive statistics. For example College data(no of student, male/ female/, pass/ fail,)

Descriptive procedures are useful because they allow us to quickly and easily get a General understanding of the data without having a look at every single scores .

However, most psychological research involves relatively small groups of people from which inferences are drawn about the larger population; this branch of statistics is called inferential statistics.

Thus inferential statistics are intended to test hypothesis and investigate relationship between **variable** and can be used to make population prediction.

Descriptive statistics is required to demonstrate that there is a difference between the two groups, and inferential statistics is needed to show that if the experiment were repeated, it would be likely that the difference would be in the same direction

Descriptive statistics as the basis for inferential procedures

VARIABLE

Things that can very is called variable
OR variable is condition or characteristics that can have different value.

Things can vary naturally (thins are not identical within group due to some other factor) or due to some intervention (due to some interferences)

In the field of physics there are many important constants (e.g., the speed of light, the mass of a proton), but most human characteristics vary a great deal from person to person

Like stress, happiness, mood, intelligence.

Continuous Versus Discrete Variables

some variables vary continuouslyor it can take infinite set of values. whereas others have only a finite (or countable, or whole number) number of levels with **no** intermediate values possible. The latter variables are said to be discrete (see Figure 1.3). A simple example of a *Continuous variable* is height, Weight, temperature . An example of a *discrete variable* is the size of a family

Choosing a scale is just one part of *operationalizing* a variable, which also includes specifying the method by which an object will be measured

In the case of a simple physical measurement such as height, there is little room for confusion or controversy However, for many important psychological variables, the exact operationalization of the variable is critical, as there may among be plenty of room for disagreement researchers studying the same ostensible phenomenon.

Let us reconsider the example of generosity (charity, kindness.दानी karn, dadhuchi, bali) Unlike height, the term generosity does not refer to some obvious(clear, observable, understable) variable that can be measure an easily agreed-upon way..

To operationalize the measurement of generosity--- 1) asking the individual to report all of his or her charitable donations, including those that might not qualify as a tax deduction. Here some people may not donate for charity but they donate for religious work, or for own community

2) ask a participant in a study to donate some proportion (whatever they are comfortable with) of the amount they were paid for the experiment back to the experimenter so more participants could be run. Most of the people may give back as amount is small

3) A completely different variable for measuring generosity would involve asking participants to donate their time to helping a charitable cause. However, some people are very generous with their time in helping friends and family, but not strangers.

As you can see, whatever variable is chosen as a measure of generosity will capture only an aspect of the underlying construct, and whatever statistical results are based on that variable can only contribute partially and indirectly to the understanding of that construct. This is a humbling reality for many areas of psychological research .

Thus generosity like other psychological phenomena is an *underlying construct* that is understood intuitively, but is hard to define exactly. In some contexts, generosity can be viewed as a *latent variable*, as opposed to a manifest or observed variable

Scales of Measurement

Because the types of statistical procedures that can be used to analyze the data from a research study depend in part on the way the variables involved were measured, we turn to this topic next.

psychologists have become accustomed to thinking in terms of levels of measurement that range from the merely categorical to the numerically precise. The four-scale system devised by Stevens (1946) is presented next

Nominal scale; the term *nominal* refers to the fact that the values are simply named, rather than assigned numbers. If numbers are assigned to the values of a nominal scale, they are assigned arbitrarily and therefore cannot be used for mathematical operations.

Nominal scale are often referred to as *categorical scales* because the different levels of the scale represent distinct categories; each object measured is assigned to one and only one category. A nominal scale is also referred to as a *qualitative* level of measurement because each level has a different quality. Examples of nominal scales include gender, , college major, and blood type. Urban/rural, marital status (married/ unmarried), religion,
Personality type(extro/ intro)

some data possess rank or designations which have a meaningful order therefore constitute an *ordinal scale*. Often the levels of an ordinal scale are given numbers. These numbers are not arbitrary like the numbers that may be assigned to the categories of a nominal scale; the gymnast ranked number 2 *is* better than rank four.
However, the rankings cannot be treated as real numbers; that is, it cannot be assumed that the third-ranked gymnast is midway between the second and the fourth.(much better, slightly better,

- Education level (primary, secondary, post-secondary), Income (low, middle, and high).

Interval and Ratio Scales

Because we know that the space, or interval, between 2 and 3 inches is the same as that between 3 and 4 inches, we can say that this measurement scale possesses the *interval property* (see Figure 1.2a)
.
Such scales are based on *units* of measurement (e.g., the inch);

This kind of scale as actual numbers and to assume that a measurement of three units is exactly halfway between two and four units.

But it lacks the value of real zero. An absolute zero point means that the value of zero on the variable indicates a complete absence of the variable.

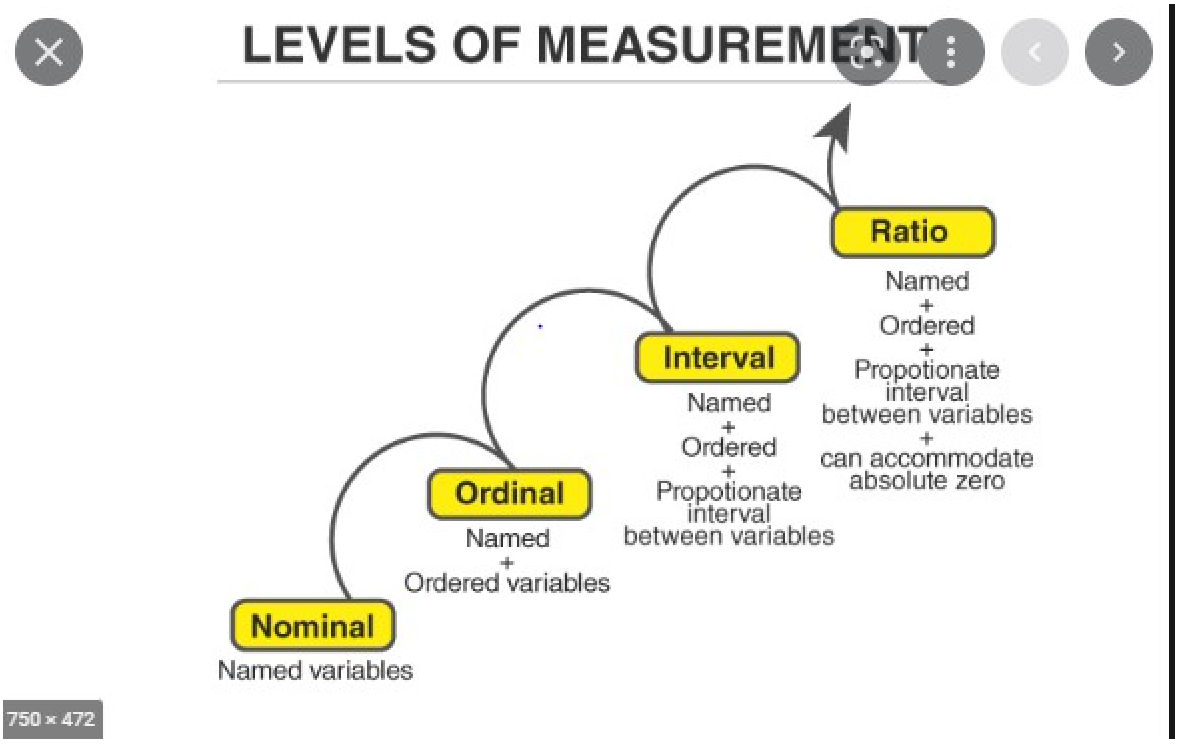
Examples of interval scales include temperature in Celsius and Fahrenheit, credit scores (300-850), SAT scores (200-800), and dates on a calendar.

If we grant that IQ scores have the interval property (which is open to debate), we still would not consider IQ a ratio scale. It doesn't make sense to say that someone who scores a zero on a particular IQ test has no intelligence at all,

Scales that have the ratio property in addition to the interval property are called *ratio scales* (see Figure 1.2b).
it has got real zero value.

Ratio scales are the top level of measurement. Like interval scales, they let you order observations and know the difference between any two values. Additionally, they allow you to assess ratios. A height of 4m is twice as tall as 2m. A period of 10 minutes is twice as long as 5 minutes.

For example height, weight, speed, and time periods.



Populations Versus Samples

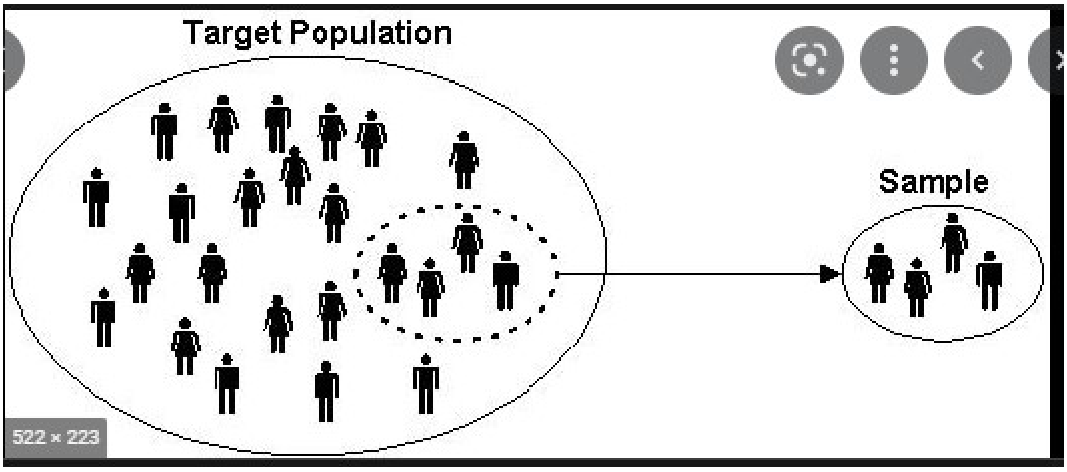
The collection of all individual items (data) under consideration in statistical study (The collection of all people who could be measured) which possesses the same characteristics.

Or complete set of element being studied.

Individual or items drawn from population is called sample.

A subset of a population is called a *sample*, especially if the subset is considerably smaller than the population and is chosen at random

It is less time consuming, economical,



A number that describes a characteristic of sample is called a **statistic and** symbolized by a letter from the English alphabet (a, b,c)**.** On the other hand, a number that describes a characteristic of a population of scores is called a **parameter.** The symbols for different parameters are letters from the Greek alphabet α, β, γ, δ

Although technically descriptive statistics are used to describe samples, their logic is also applied to populations. We want to answer the same questions for the population being represented by the sample.

Because we usually cannot measure the scores in the population, however, we must *estimate* the description of the population, based on the sample data.

Essentially, inferential procedures are for deciding whether to *believe* what the sample data seem to indicate about the scores and relationship that would be found in the population

Example --- People learn more when they study in a quiet versus noisy place

Parametric Versus Nonparametric Statistics

Statistical procedures based on distributions (some pattern) and their parameters are called parametric statistics

With interval/ratio data it is often (but not always) appropriate to use parametric statistics .

If all of your variables have been measured on nominal scales, or your interval/ratio data do not even come close to meeting the distributional assumptions of parametric statistics (which will be explained at the appropriate time), you should be using nonparametric statistics, as described in Part VII.

If at least one of your variables has been measured on an interval or ratio scale, and certain additional assumptions have been met, it may be appropriate to use parametric statistics to draw inferences about population parameters from your sample statistics. If all of your variables have been measured on ordinal or nominal scales, or the assumptions of parametric statistics have not been met, it may be necessary to use nonparametric statistics

Experimental Versus Observational Research

It is important to realize that not all research involves experiments; much of the research in some areas of psychology involves measuring differences between groups that were not created by the researcher. For instance,

Examining whether length of daily exposure to a sun lamp (15 minutes versus 60 minutes) alters self-reported depression

For Practice

1. Whether you are ahead or behind when running involves a _____ scale.
2. The number of hours you slept last night involves a _____ scale.
3. Your blood type involves a _____ scale.
4. Whether you are a lieutenant or major in the army involves a _____ scale.
5. A _____ scale allows fractions; a _____ scale allows only whole amounts

1. What type of scale is being used for each of the following measurements? 2. Which of the following variables are discrete and which are continuous?

a. Number of arithmetic problems correctly solved
b. Class standing (i.e., one's rank in the graduating class)
c. Type of phobia
d. Body temperature (in ◦F)
e. Self-esteem, as measured by self-report questionnaire
f. Annual income in dollars
g. Theoretical orientation toward psychotherapy
h. type of psychological disorder
i. Heart rate in beats per minute
j. rating the movie(*,**,***,)
k. children of men
i Dollars in your pocket
k. The number of people in one's social network
l. volume of a jar
m. Size of vocabulary
n Blood pressure
o. no of face book like
p wind speed
q. temperature of body
r Clothing size
s Work absences
t Words recalled
u number of siblings a person has

3. A psychologist records how many words participants recall from a list under three different conditions: large reward for each word recalled, small reward for each word recalled, and no reward.

a. What is the independent variable?
b. What is the dependent variable?
c. What kind of scale is being used to measure the dependent variable

4. Which of the following studies are experimental and which are observational?

a. Comparing pet owners with those who don't own pets on an empathy measure
b. Comparing men and women with respect to performance on a video game that simulates landing a space shuttle
c. Comparing participants run by a male experimenter with participants run by a female experimenter with respect to the number of tasks completed in 1 hour
d. Comparing the solution times of participants given a hint with those not given a hint

5. Using the terms *sample, population, variable, statistics,* and *parameter,* summarize the steps a researcher follows, starting with a hypothesis and ending with a conclusion about a nature

Mathematical notation

variable whose value is unknown is most commonly represented by the letter X.

For example we want average temperature in July, we can say x=35*c. The usual way to find the average temperature for the entire month of July is to take the average temperature for each day in July and then average these averages.it would be awkward to use the same letter, X, for each day of the month if we then want to write a formula. On the other hand, we certainly cannot use a different letter of the alphabet for each day. The solution is to use subscripts. average temperature for July 1 can be written X1, X2,so on up to X31.

We now have a compact way of referring

The Summation Sign

the subscripts introduced above, the average temperature for July can be expressed as (X1+ X2+ X3+............ X31) /31

symbol that indicates that a string of variables is to be added is called the summation sign, and it is symbolized by the uppercase Greek letter sigma ((Σ).

First, you write i = 1 under the summation sign to indicate that the summing should start with the variable that has the subscript 1.

On top of the summation sign you indicate the subscript of the last variable to be added

next to the summation sign you write the letter that stands for the collection of variables to be added, using the subscript i

$$\sum_{i=1}^{31} X_i$$

So the sum of the average temperatures for each day in July can be symbolized as follows:

To be more general, you could use the letter N to stand for the number of days in an, which leads to the following expression:

$$\sum_{i=1}^{N} X_i$$

Summation Rule 1A

$$\sum (X_i + Y_i) = \sum X_i + \sum Y_i$$

The rule works in exactly the same way for subtraction.

Summation Rule 1B

$$\sum (X_i - Y_i) = \sum X_i - \sum Y_i$$

Rule 1A works because if all you're doing is adding, it doesn't matter what order you use. Note that $\Sigma(X_i + Y_i)$ can be written as: $(X_1 + Y_1) + (X_2 + Y_2) + (X_3 + Y_3) + \cdot \cdot \cdot + (X_N + Y_N)$

Summation Rule 2

$$\sum C = NC$$

Summation Rule 3

$$\sum CX_i = C \sum X_i$$

Summation Rule 4

$$\sum (X_i Y_i) \neq \left(\sum X_i\right) \left(\sum Y_i\right)$$

The first two exercises are based on the following values for two variables: $X_1 = 2$, $X_2 = 4$, $X_3 = 6$, $X_4 = 8$, $X_5 = 10$; $Y_1 = 3$, $Y_2 = 5$, $Y_3 = 7$, $Y_4 = 9$, $Y_5 = 11$.

*1. Find the value of each of the following expressions:

a. $\sum_{i=2}^{5} X_i$ b. $\sum_{i=1}^{4} Y_i$ c. $\sum 5X_i$

d. $\sum 3Y_i$ e. $\sum X_i^2$ f. $\left(\sum 5X_i\right)^2$

g. $\sum Y_i^2$ h. $\left(\sum Y_i\right)^2$

*2. Find the value of each of the following expressions:

a. $\sum (X + Y)$ b. $\sum XY$ c. $\left(\sum X\right)\left(\sum Y\right)$

d. $\sum (X^2 + Y^2)$ e. $\sum (X - Y)$ f. $\sum (X + Y)^2$

g. $\sum (X + 7)$ h. $\sum (Y - 2)$

*4. Use the appropriate summation rule(s) to simplify each of the following expressions (assume all letters represent variables rather than constants):

a. $\Sigma(9)$, b. $\Sigma(A - B)$, c. $\Sigma(3D)$, d. $\Sigma(5G + 8H)$, e. $\Sigma(Z^2 + 4)$