

# Introduction To Biostatistics for Post Graduate



**By**

Dr. Lamyia Al-Azzawi PhD. Community Dentistry.

PhD. Diagnosis and oral medicine

Department of Pediatric, Orthodontic and Preventive Dentistry (P.O.P)

# Biostatistics.

can be defined as the application of the mathematical tools used in statistics to the fields of biological sciences and medicine

It is a combination word made from biology and statistics)

The application of statistics to a wide range of topics in biology problems, including:

- ❖ Public health.
- ❖ Medicine.
- ❖ Ecological and environmental.

- Biostatistics is a growing field with applications in many areas of biology including epidemiology, medical sciences, health sciences, educational research and environmental sciences
- Biostatistics is concerned with collection, organization, summarization and analysis of data.

It is the science that provide a way of processing number or analyzing the data that have been collected . The purpose of statistics is to make an implication or an assumption about population which deals with development and application of the most appropriate methods for the:

- ❖ Collection of data.
- ❖ Presentation of the collected data.
- ❖ Analysis and interpretation( understanding) of the results.
- ❖ Making decisions on the basis of such analysis .

# Types of Statistics.

1.Descriptive statistics are used to describe and summarize data. Their objective is to communicate results, without generalizing beyond the sample ,to any population .

Some ways in which results are communicated are through

- a. measures of central tendency (mean, median ,and mode)
- b. measures of dispersion

2. Inferential statistics are used to apply information from the sample to a larger population.

For example, you might stand in a mall and ask a sample of 100 people if they like shopping at [Sears](#). You could make a [bar chart](#) of yes or no answers (that would be [descriptive statistics](#)) or you could use your research (and inferential statistics) to reason that around 75-80% of the population (**all** shoppers in **all** malls) like shopping at Sears.

# Challenge



Much of life is composed of a systematic component (i.e., signal) and a random component (i.e., error or noise)

- Example:
  - Smoking is associated with lung cancer.
  - Yet not everyone that smokes, gets lung cancer, and not everyone that gets lung cancer, smokes
  - Yet we know that there is an association (a systematic component)



- Data is plural of datum which literally means to give or something given. Data is thought to be the lowest unit of information from which other measurements and analysis can be done. Data can be numbers, images, words, figures, facts or ideas.
- Data in itself cannot be understood and to get information from the data one must interpret it into meaningful information. There are various methods of interpreting data. Data sources are broadly classified into primary and secondary data.



- There are two sources of data collection techniques.
- Primary and Secondary data collection techniques, Primary data collection uses surveys, experiments or direct observations.
- Secondary data collection may be conducted by collecting information from a diverse source of documents or electronically stored information, census and market studies are examples of a common sources of secondary data. This is also referred to as "data mining."

## Primary data sources

- **Survey** : is most commonly used method in social sciences, management, marketing and psychology to some extent. Surveys can be conducted in different methods.
- **Questionnaire**

Questionnaire is the most commonly used method in survey.

Questionnaires are a list of questions either an open-ended or close-ended for which the respondent give answers. Questionnaire can be conducted via telephone, mail, live in a public area, or in an institute, through electronic mail or through fax and other methods.

## **.Interview**

- Interview is a face-to-face conversation with the respondent. It is slow, expensive, and they take people away from their regular jobs, but they allow in-depth questioning and follow-up questions

## **• Observations**

- Observations can be done while letting the observing person know that he is being observed or without letting him know. Observation can also be made in natural settings as well as in artificially created environment.

# Advantages of Primary Data

- ❖ Targeted Issues are addressed.
- ❖ Data interpretation is better.
- ❖ Efficient Spending for Information.
- ❖ Decency of Data
- ❖ Addresses Specific Research Issues.
- ❖ Greater Control. Proprietary Issues.

# DISADVANTAGES OF PRIMARY

- High Cost
- Time Consuming
- Inaccurate Feed-backs
- More number of resources is required

# Secondary data

- Secondary data is the data that has been already collected by and readily available from other sources
- Published Printed Sources .There are varieties of published printed sources. Their credibility depends on many factors.
- For example, on the writer, publishing company and time and date when published. New sources are preferred and old sources should be avoided as new technology and researches bring new facts into light

- **Books**

Books are available today on any topic that you want to research. The uses of books start before even you have selected the topic. After selection of topics books provide insight on how much work has already been done on the same topic and you can prepare your literature review. Books are secondary source but most authentic one in secondary sources.

- **Journals/periodicals Journals** and periodicals are becoming more important as far as data collection is concerned. The reason is that journals provide up-to-date information which at times books cannot and secondly, journals can give information on the very specific topic on which you are researching rather talking about more general topics.



# Published Electronic Sources

- As internet is becoming more advance, fast and reachable to the masses; it has been seen that much information that is not available in printed form is available on internet. In the past the credibility of internet was questionable but today it is not. The reason is that in the past journals and books were seldom published on internet but today almost every journal and book is available online
- .Some are free and for others you have to pay the price. E-journals: e-journals are more commonly available than printed journals. Latest journals are difficult to retrieve without subscription but if your university has an e-library you can view any journal, print it and those that are not available you can make an order for them

# ADVANTAGES OF SECONDARY DATA

- Inexpensive
- Easily accessible
- Immediately available.
- Will provide essential background and help to clarify or refine research problem – essential for literature review
- Secondary data sources will provide research method alternatives. Will also alert the researcher to any potential difficulties.

# DISADVANTAGES OF SECONDARY DATA

- ❖ Expensive
- ❖ Not immediately available –
- ❖ takes time to define problem, sampling frame, method and analysis.
- ❖ Not as readily accessible
- ❖ Incomplete Information

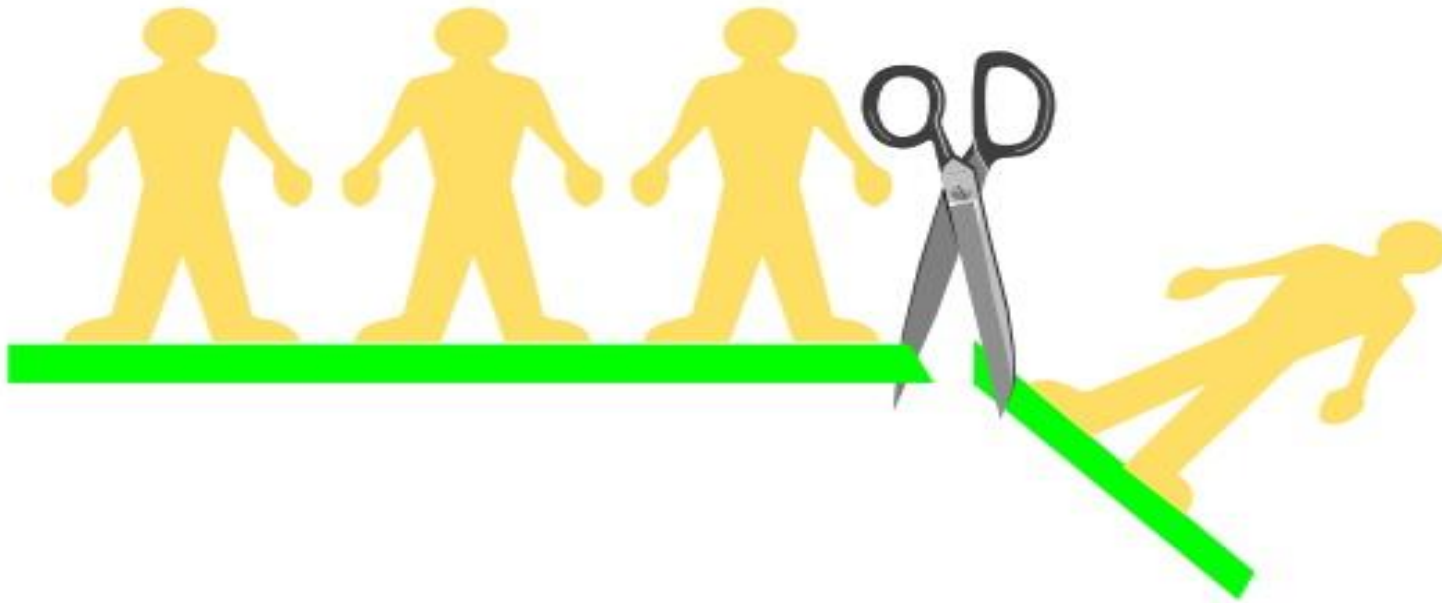
# Populations and Samples

Studying populations is too expensive and time-consuming, and thus impractical.

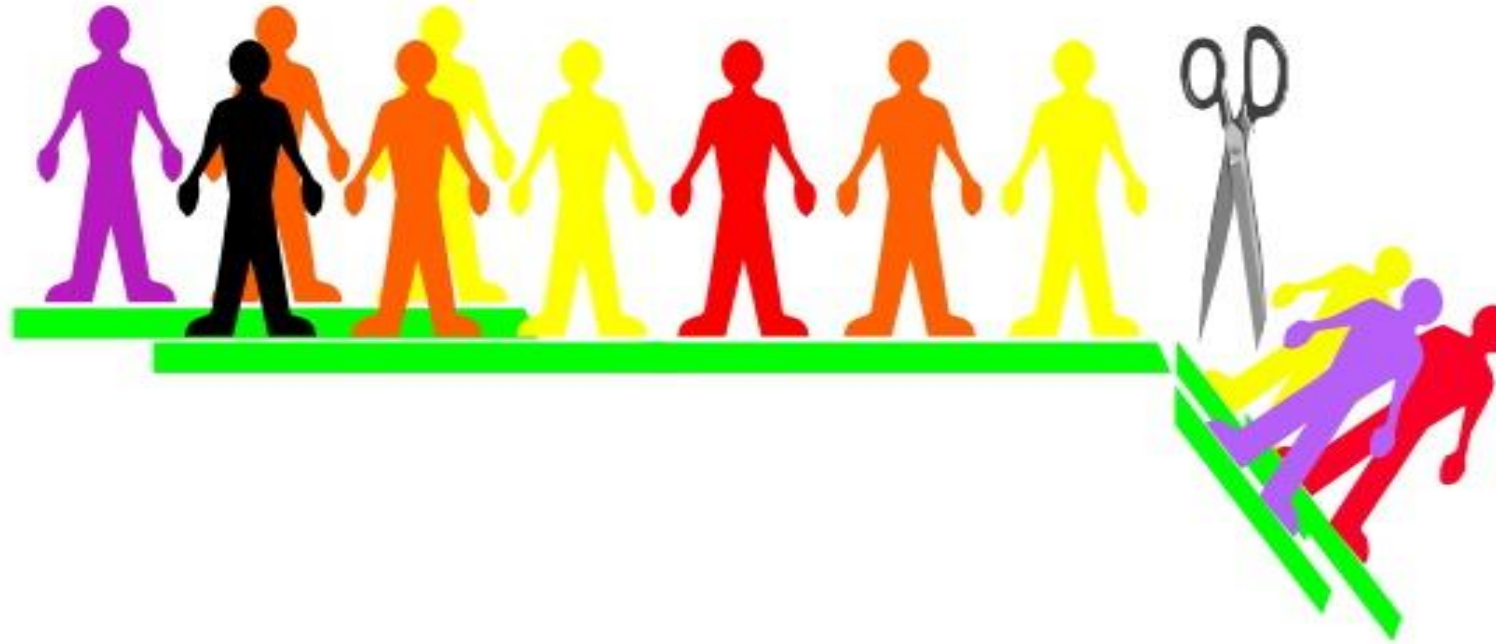
A sample is a portion of the population that, if properly selected can provide meaningful information about the entire population.

And thus by looking at the characteristics of the sample (statistics), we may learn something about the characteristics of the population (parameters)

*IF THE POPULATION IS  
HOMOGENEOUS*



*IF THE POPULATION IS  
HETEROGENEOUS*



# The advantages of sampling over the census enumeration

## Census

1. Information is collected from all the individuals in the population
2. Cost of organization and execution will be more
3. Require more time and personnel for collection and analysis .
4. Lesser accuracy and completeness

## Sampling

1. Information is collected from the units in the sample.
2. Cost is less since sample is smaller than the population.
3. Requires lesser time and personnel for collection and analysis .
4. More accuracy and completeness

# Types of Samples

## 1. Purposive (Judgmental) selection

This type of sample provide a sample through personal judgment ,of subjects who would be most representative of population, it is easy to carry out and does not need the preparation of sampling frame.

For example:

Studying oral cancer in men forty representative patients may be picked ,Examined and assessed for this disease.



## 2.Random selection.

A sample in which each individual in the population has an equal chance of appearing is a random sample .Random ,here does not mean haphazard ,but it indicates the chance of the population unit being selected in the sample. So ,it is also called probability sampling.

# Sampling Designs

There are different sample designs depending on the type and the nature of population and the objectives of the investigation. Some of the commonly used design are:

- 1.Simple random sampling
- 2.Systematic random sampling
- 3.Stratified random sampling
- 4.Cluster sampling
- 5.Multistage sampling

# Simple Random Samples

## Advantages

- Minimal of the knowledge of the population needed
- Easy to analyze data

## Dis advantages

- Low frequency of use.
- Dose not use researchers expertise.
- Larger risk of random error.

# Stratified Random Sampling

- Population is divided into two or more groups called strata
- Subsamples are randomly selected from each strata

**For example:** if it is decide to know the prevalence of caries in different age groups, then the age groups from the strata and a random sample is to be chosen from each stratum.

## Advantages

- ❖ Assures representation of all groups in sample population.
- ❖ Characteristics of each stratum can be estimated and comparison made

## Disadvantages

- ❖ Requires accurate information on proportion of each stratum
- ❖ Stratified lists costly to prepare

# CLUSTER SAMPLING

- The population is divided into subgroups (clusters) like families.
- A simple random sample is taken from each cluster



# Cluster Samples

## Advantages

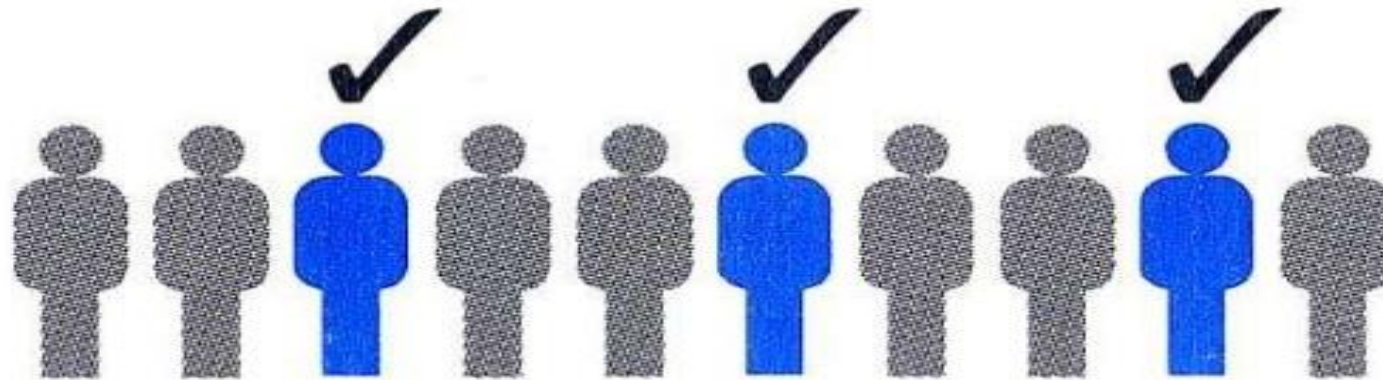
- ❖ Can estimate characteristics of both cluster and population

## Disadvantages

- ❖ The cost to reach an element to sample is very high
- ❖ Each stage in cluster sampling introduces sampling error-the more Stages there are, the more error tend to be

# SYSTEMATIC RANDOM SAMPLING

- Order all units in the sampling frame
- Then every  $n$ th number on the list is selected
- $N$  = Sampling Interval





# Systematic Random Sampling

## Advantages

- ❖ Moderate cost, moderate usage.
- ❖ Simple to draw sample
- ❖ Easy to verify

## Disadvantages

- ❖ Periodic ordering is required

# MULTISTAGE SAMPLING

- Carried out in stages
- Using smaller and smaller sampling units at each stage

Primary  
Clusters

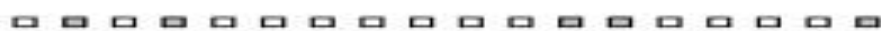


Secondary  
Clusters



1

Simple Random Sampling within 5



# Multistage Sampling

## Advantages

- ❖ More accurate
- ❖ More effective

## Disadvantages

- ❖ Costly
- ❖ Each stage in sampling introduce sampling error. The more stages there are the more error tend to be

# Sample Size

The sample size need to be decided and calculated before selection and starting the study .Size of the sample is dependent on the statistical characteristics of the data to be collected.

- Larger the sample less the sampling variation i.e. Less chance of misleading results.

**Most important point that the sample should be representative.**

# SAMPLING ERRORS




# There are two types of errors that arise in sampling investigation:

1. The sampling errors which occur due to the sampling process, including:

- a. Faulty sampling design
- b. small size of the sample

2. Non sampling error arise due:

- a. Coverage error due to  non-response of the informant  
non cooperation of the informant

b. Observational errors due to :

- ★ interviewers bias.
- ★ imperfect experimental technique.
- ★ interaction of both.

c. processing errors:

 Due to errors in statistical analysis



A decorative arrangement featuring three lit pink candles in a glass dish, surrounded by white cherry blossoms and petals on a blue background. The candles are lit, casting a warm glow. The text "Thank you for your kind attention" is overlaid in the center.

**Thank you for your kind attention**