

Dr. Mohanad Faris Abdulhameed

Department of Public Health

College of Veterinary Medicine

University of Basrah

Second Class-Statistics

محاضرة رقم 2

Data Collection

- **In research field**, the question is how data can be obtained or generate good data? We can test two way of hypothesis: by observation naturally what is happening or manipulate some aspect of environment and observe the effects on variable interest us in the study.
- Therefore, data can be defined as the process of gathering and measure information on variables of interest using three ways: **censuses, sample surveys, and administrative data**.

1- Census (counts): a census can be referred to data collection of every unit in a group or population. If a researcher collected data about the heights of the students in the veterinary medicine school that would be all students from different stages regardless gender.

2- Sample survey: Example, collection the blood samples from the cattle and buffalo to identify presence of the blood parasites or collection milk samples from dairy cattle with mastitis to identify bacteria or may be fungi causes.

3- Administrative data: Administrative data are collected as a result of an organization's day-to-day operations. Examples include data on births, deaths, marriages, divorces and car registrations.

The data collection component of research is common to all fields of study including physical and social sciences, humanities, business. Data collection is a very demanding job which needs thorough planning, hard work, patience, perseverance and more to be able to complete the task successfully. Data collection starts with determining what kind of data required followed by the selection of a sample from a certain population. After that, you need to use a certain instrument to collect the data from the selected sample. All data need to be managed very well based on the date collected, sufficient details of a subject, provide computerised copy (in excel).

Table 1. shows numbers and frequency of persons with categorical age.

Age in years	No. of persons	Frequency
0-12	1124	98
12-24	1026	217
24-36	809	382
36-48	427	269
48-60	158	138
60-72	20	15
72-84	5	2
84-96	3	2
96-108	1	1
108-120	0	0

- **Classification of data**

The data are classified in to two types:

A- Primary data

Data that has been collected from first-hand-experience is known as primary data. Primary data has not been published yet and is more reliable, authentic and objective. Primary data has not been changed or altered by researchers; therefore, its likely to be valid than secondary data.

Importance of Primary Data: in statistical surveys it is necessary to get information from primary sources and work on primary data. **For example**, the statistical records of number of farmers in rural areas cannot be based on newspaper, magazine and other printed sources. A research can be conducted without secondary data but a research based on only secondary data is least reliable and may have biases because secondary data has already been manipulated by human beings. One of such sources is old and secondly, they contain limited information as well as they can be misleading and biased.

Sources of Primary Data: Sources for primary data are limited and at times it becomes difficult to obtain data from primary source because of either scarcity of population or lack of cooperation. Following are some of the sources of primary data:

1-Data originated from experiments: experiments require an artificial or natural setting in which to perform logical study to collect data. Experiments are more suitable for medicine, psychological studies, nutrition and for other scientific studies.

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

3- Questionnaire: It is the most commonly used method in survey. Questionnaires are a list of questions either open-ended or close-ended for which the respondents 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.

4- Interview: Interview is a face-to-face conversation with the respondent.

5-Observations: Observation can be done while letting the observing person know that she/he is being observed or without letting him know.

B- Secondary data

Data collected from a source that has already been published in any form is called as secondary data. The review of literature in any research is based on secondary data. It is collected by someone else for some other purpose (but being utilized by the investigator for another purpose). For examples, Census data being used to analyse the impact of education on career choice and earning. Common sources of secondary data for social science include censuses, organizational records and data collected through qualitative methodologies or qualitative research. Secondary data is essential, since it is impossible to conduct a new survey that can adequately capture past change and/or developments.

Sources of Secondary Data: The following are some ways of collecting secondary data –

- Books.
- Records.
- Biographies.
- Newspapers.
- Published censuses or other statistical data.
- Data archives.
- Internet articles.
- Research articles by other researchers (journals).
- Databases.

Types of data

1. **Quantitative Data:** Quantitative data is numerical in nature and can be mathematically computed. Quantitative data measure uses different scales, which can be classified as nominal scale, ordinal scale, interval scale and ratio scale. Often (not always), such data includes measurements of something. Quantitative approaches address the ‘what’ of the program. They use a systematic standardized approach and employ methods such as surveys and ask questions. Quantitative approaches have the advantage that they are cheaper to implement, are standardized so comparisons can be easily made and the size of the effect can usually be measured.

The Quantitative data collection methods rely on random sampling and structured data collection instruments that fit diverse experiences into predetermined response categories. They produce results that are easy to summarize, compare, and generalize. If the intent is to generalize from the research participants to a larger population, the researcher will employ probability sampling to select participants. **Typical quantitative data gathering strategies include:**

- Experiments/clinical trials.
- Observing and recording well-defined events (e.g., counting the number of patients waiting in emergency at specified times of the day).
- Obtaining relevant data from management information systems.
- Administering surveys with closed-ended questions (e.g., face-to face and telephone interviews, questionnaires etc).
- In quantitative research (survey research), interviews are more structured than in Qualitative research. Face to face interviews have a distinct advantage of enabling the researcher to establish rapport with potential participants and therefore gain their cooperation.
- Paper-pencil-questionnaires can be sent to a large number of people and saves the researcher time and money.

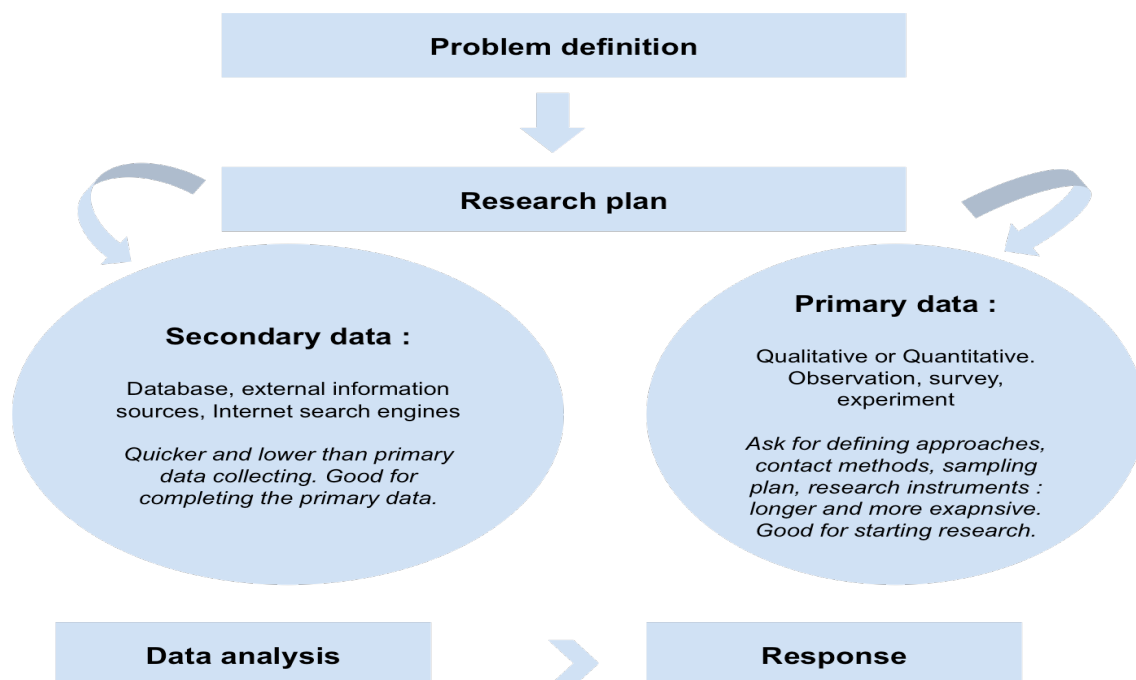
2- Qualitative Data: Qualitative data are mostly non-numerical and usually descriptive or nominal in nature. This means the data collected are in the form of words and sentences. Qualitative approaches aim to address the ‘how’ and ‘why’ of a program and tend to use unstructured methods of data collection to fully explore the topic. Qualitative questions are

open-ended. Qualitative methods include focus groups, group discussions and interviews. Qualitative approaches are good for further exploring the effects and unintended consequences of a program. They are, however, **expensive and time consuming to implement**.

Qualitative data collection methods play an important role in impact evaluation of a subject by providing information useful to understand the processes behind observed results and assess changes in people’s perceptions of their well-being. **These methods are characterized by the following attributes -**

- they tend to be open-ended and have less structured protocols (i.e., researchers may change the data collection strategy by adding, refining, or dropping techniques or informants);
- they rely more heavily on interactive interviews; respondents may be interviewed several times to follow up on a particular issue;
- they use triangulation (**triangulation mean determine important points**) to increase the credibility of their findings (i.e., researchers rely on multiple data collection methods to check the authenticity of their results).

Typical example, evaluation health care system in Iraq; evaluation role of veterinary department in Basrah towards farmers or introduce vaccine against prevailing diseases.



Qualitative vs Quantitative Observations

Qualitative observations use your senses to observe the results. (Sight, smell, touch, taste and hear.)

Quantitative observations are made with instruments such as rulers, balances, graduated cylinders, beakers, and thermometers. These results are measurable.

Read the following examples and then decide if the example is qualitative (1) or quantitative (2).

1. It is light green in color. _____ (1 or 2)
2. It taste sour. _____ (1 or 2)
3. One leaf is 9 cm long. ____ (1 or 2)
4. It makes a loud pop sound. ____ (1 or 2)
5. The mass of the computer is 1 1/2 kg. ____ (1 or 2)
6. It smells sweet. ____ (1 or 2)
7. The temperature of the room increases by 8 degrees C. ____ (1 or 2)
8. It gets darker over a period of time. ____ (1 or 2)
9. The flower clusters in 3 blooms. ____ (1 or 2)
10. Feels very rough. ____ (1 or 2)
11. The plant is short. ____ (1 or 2)
12. Leaves are stiff. ____ (1 or 2)
13. The veins are 3 mm wide. ____ (1 or 2)

What are the definition of **mean, median, mode**?

The mean, mode, median, and range are averages, which are used in statistics to give information about data and help the user to draw conclusions. The mean, mode, median, midrange, and range are defined as:

- **Mean:** the average, which is found by adding up all the values in a set of data and dividing it by the total number of values you added together.
- **Median:** the middle number in the set of values. You find it by putting the numbers in order from the smallest to largest and covering up one number on each end until you get to the middle.
- **Mode:** the number or value, which appears most often in the set. To find the mode, you need to

Normal Distribution | Examples, Formulas, & Uses

In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the centre.

Why do normal distributions matter?

All kinds of variables in natural and social sciences are normally or approximately normally distributed. Height, birth weight, reading ability, job satisfaction, or SAT scores are just a few examples of such variables.

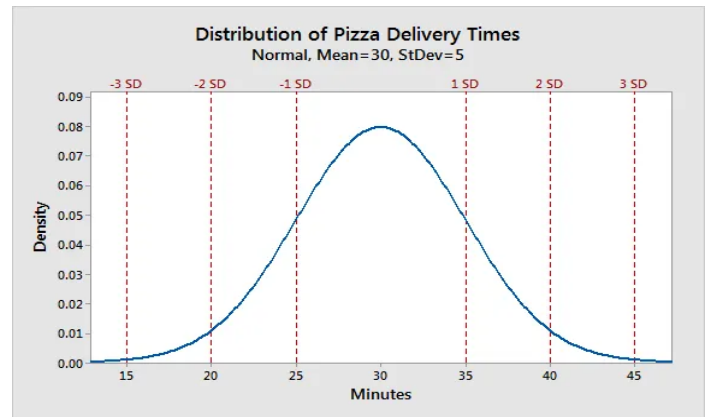
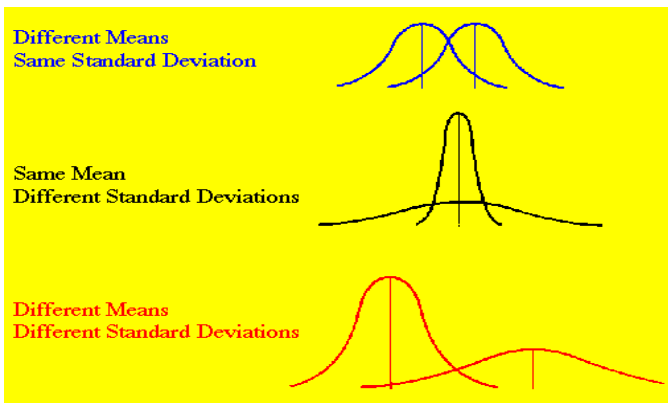
Because normally distributed variables are so common, many statistical tests are designed for normally distributed populations.

Understanding the properties of normal distributions means you can use inferential statistics to compare different groups and make estimates about populations using samples.

What are the properties of normal distributions?

Normal distributions have key characteristics that are easy to spot in graphs:

- The [mean](#), [median](#) and [mode](#) are exactly the same.
- The distribution is symmetric about the mean—half the values fall below the mean and half above the mean.
- The distribution can be described by two values: the mean and the [standard deviation](#).



These distributions are sometimes called asymmetric or asymmetrical distributions as they don't show any kind of [symmetry](#).

