Assist. Prof. Dr. Mohanad Faris Abdulhameed
Department of Public Health
College of Veterinary Medicine
University of Basrah
Second Class-Biostatistics
محاضرة رقم 1

Introduction to Statistics

1. What are statistics

- Let's take three examples below in our life.
  1. Heights are varied scales among human population.
  2. Diversity of weights of the watermelon.

❖ Statistics is easy to understand and to learn if you just focus deeper. Statistics involves mathematics and relies upon calculations of numbers. But it also relies heavily on how the numbers are chosen and how the statistics are interpreted. For example, consider the following three scenarios and the interpretations based upon the presented statistics. You will find that the numbers may be right, but the interpretation may be wrong. Try to identify a major flaw with each interpretation before we describe it.

❖ **What is statistics: Statistics is the study and manipulation of data, including ways to collect, review, analyse, and draw conclusions from data. There are two major areas of statistics are descriptive and inferential statistics**

- **Example**: a new advertisement for ice cream introduced in late May of last year resulted in a 30% increase in ice cream sales for the following three months. Thus, the advertisement was effective (because the advertisement is important to make profit from the products like ice cream).

2. Why statistics are so important:

Statistics has become important aspect to our life. You can see the statistics in sport, medicine, business, and law. Statistics is important to reveal on number of the patients with a certain disease, for instance the number of the patients with Covid-19 in 2020 and the number of deaths

among the patients with Covid-19 disease across the world. Any circumstances whether the patients in the hospitals with the diseases or even the veterinarians who treated animals every year that, however, needs to be explained and presents via numbers and written as tables, ratios, and illustrate graphs. After do statistics measure you need to do interpretation of the results or what the number of a percentage mean related to the subject the title or a report.

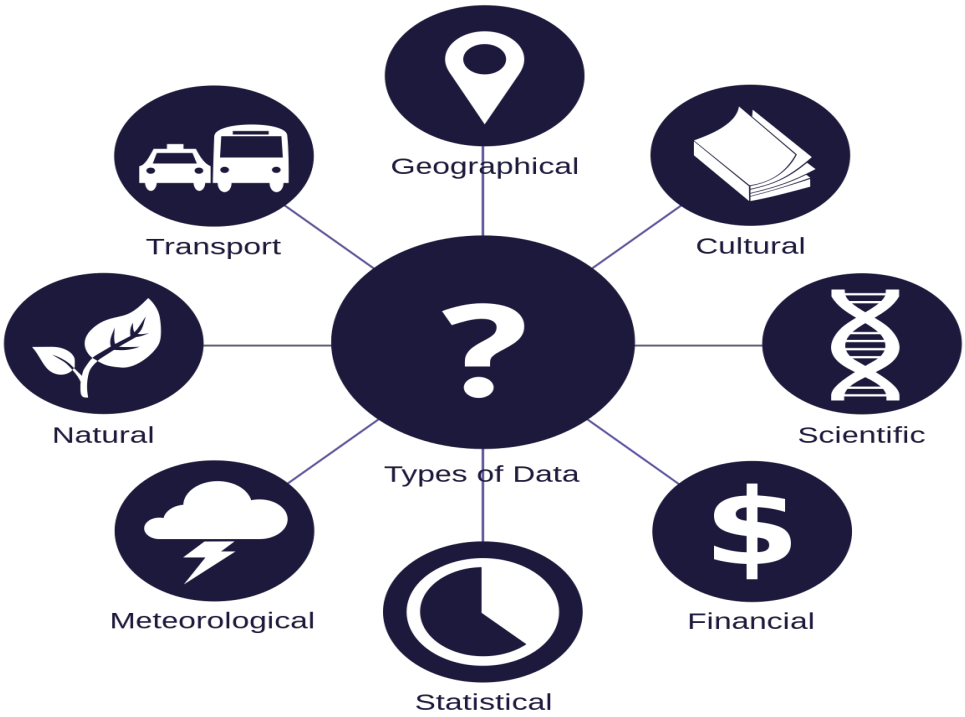
By the way, statistics provides tools that you need in order to react intelligently to information you hear or read. In this sense, statistics is one of the most important things that you can study.

Look take two examples to find out more about importance of the statistics below:

1. Almost 85% of lung cancers in men and 45% in women are tobacco-related.
2. A surprising new study shows that eating egg whites can increase one's life span.

3. Descriptive statistics

What are the descriptive statistics means: it refers to numbers that are used to summarize and describe data. The word "**data**" refers to the information that has been collected from an experiment, a survey, an historical record, etc.



If we are analysing birth certificates, for example, a descriptive statistic might be the percentage of certificates issued in Basrah, or the average age of the mother. Any other number we choose to compute also counts as a descriptive statistic for the data from which the statistic is computed.

The descriptive statistics are shown in Table 1 that illustrate the global prices of oil barrel (crude oil) from 2015 to 2020.

Year	Price per barrel $ on average
2016	60$
2017	55$
2018	50$
2019	50$
2020	20$

Based on the descriptive statistics and from the question is given that why the prices oil has been decreased since 2016 and significantly plunged down in 2020. Just will take 2020 the prices of oil become around 20$ because of the outbreak of Covid-19. So, it is important to look to the data and interpret logically. The conclusion is that statistics can help to make reasonable answer.

4. Variables

Variable is a characteristic that can take values which *vary* from individual to individual or group to group, e.g. height, weight, litter size, blood count, enzyme activity, coat colour, percentage of the flock which are pregnant.

Biological variation

The causes of biological variation, which makes one individual differ from the next or from one time to another, may be obvious or subtle. For example, variations in any characteristic may be attributable to:

- Genetics – e.g. greater variability in the whole cow population compared with just Friesians.
- Environment – e.g. body weight varies with diet, housing, intercurrent disease, etc.
- Gender – sexual dimorphism is common.
- Age – many biological data are influenced by age and maturity, e.g. the quantity of body fat.

Independent and dependent variables

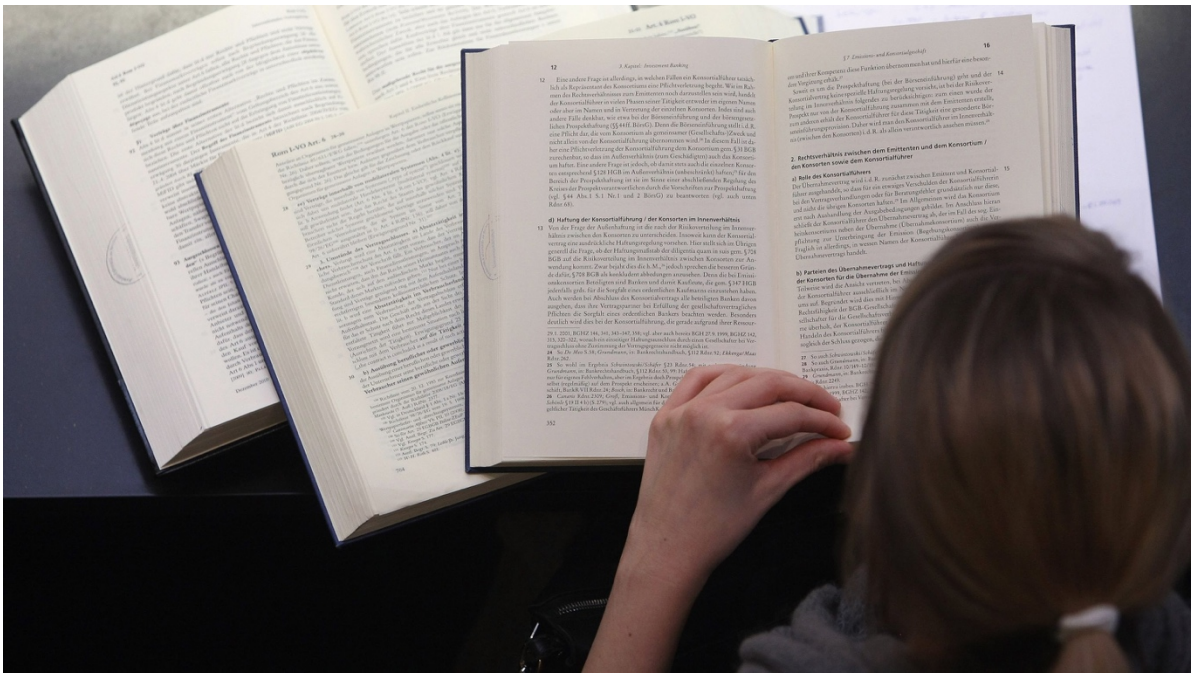
Independent: the "cause" or input. It is the factor that is varied or controlled (e.g., amount of sunlight, dosage of medicine, time spent studying).

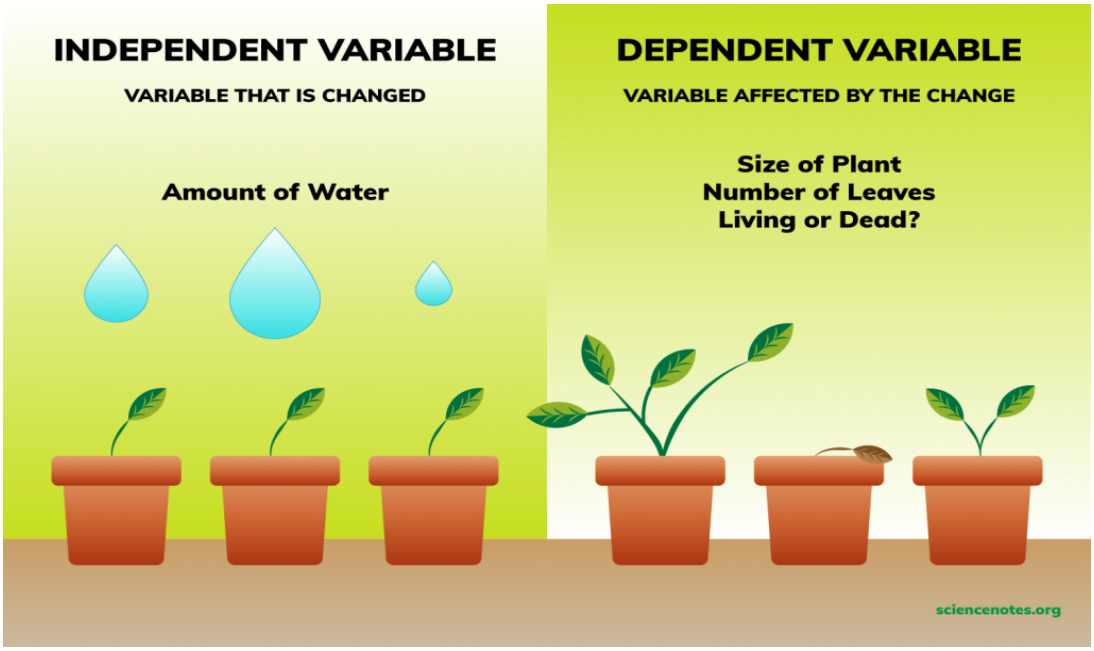
A dependent variable is the factor being measured, tested, or observed in an experiment or study, which changes in response to manipulations of the independent variable. It represents the "effect" or outcome of interest. test scores (affected by study time), plant growth (affected by fertilizer), and blood pressure (affected by medication dosage)

Table: examples for dependent and independent variables

Research Question	**Independent variable(s)**	**Dependent variable(s)**
Do tomatoes grow fastest under fluorescent, incandescent, or natural light?	The type of light the tomato plant is grown under	The rate of growth of the tomato plant
What is the effect of diet and regular soda on blood sugar levels?	The type of soda you drink (diet or regular)	Your blood sugar levels
How well do different plant species tolerate salt water?	The amount of salt added to the plants' water.	• Plant growth • Plant wilting • Plant survival rate

(Time Spent Studying) causes a change in (Test Score) and it isn't possible that (Test Score) could cause a change in (Time Spent Studying).





5. Qualitative and Quantitative Variables

An important distinction between variables is between qualitative variables and quantitative variables. **Qualitative variables** are those that express a qualitative attribute such as hair colour, eye colour, religion, favourite movie, gender, etc... The values of a qualitative variable do not imply a numerical ordering. Qualitative variables are sometimes referred to as categorical variables. **Quantitative variables** are those variables that are measured in terms of numbers. Some examples of quantitative variables are height, weight, and shoe size.

Example#. In the study on the effect of diet, the independent variable was type of supplement: none, strawberry, blueberry, and spinach. The variable "type of supplement" is a qualitative variable; there is nothing quantitative about it. In contrast, the dependent variable "memory test" is a quantitative variable since memory performance was measured on a quantitative scale (number correct).

Quantitative Variables	Qualitative Variables
Take on numeric values	*Take on names or labels*
Number of students in a class	Eye color
Number of square feet in a house	Gender
Population size of a city	Breed of dog
Age of an individual	Level of Education
Height of an individual	Marital status

6. Discrete and Continuous Variables

Variables such as number of children in a household are called discrete variables since the possible scores are discrete points on the scale. For example, a household could have three children or six children, but not 4.53 children. Other variables such as "time to respond to a

question" are continuous variables since the scale is continuous and not made up of discrete steps. The response time could be 1.64 seconds, or it could be 1.64237123922121 seconds.

7. Levels of Measurement

Is very important to know the level of measurement in order to see the Tendency and Variance of the data and then estimate the Mode, Mean, and Median (will be explained in another class).

Four types of Scales

1) **Nominal scales**

When measuring using a nominal scale, one simply name or categorizes responses (no calculation). Examples: Gender, yes or no, handedness, favourite colour, and religion are examples of variables measured on a nominal scale.

2) **Ordinal scales**

A researcher wishing to measure consumers' satisfaction with their microwave ovens might ask them to specify their feelings as either "very dissatisfied," "somewhat dissatisfied," "somewhat satisfied," or "very satisfied." The items in this scale are ordered, ranging from least to most satisfied. This is what distinguishes ordinal from nominal scales.

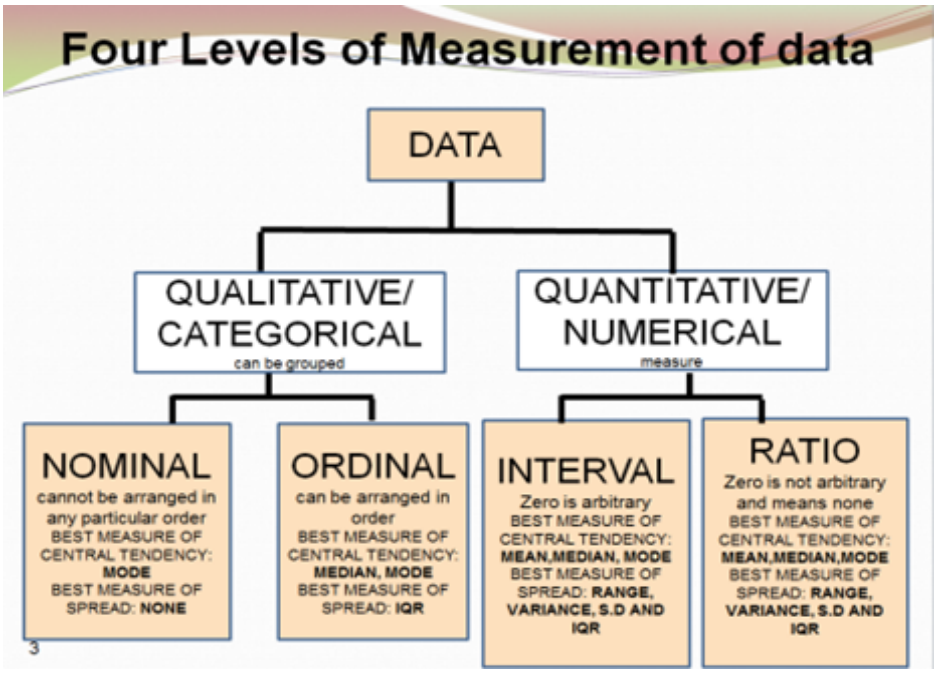
3) **Interval scales**

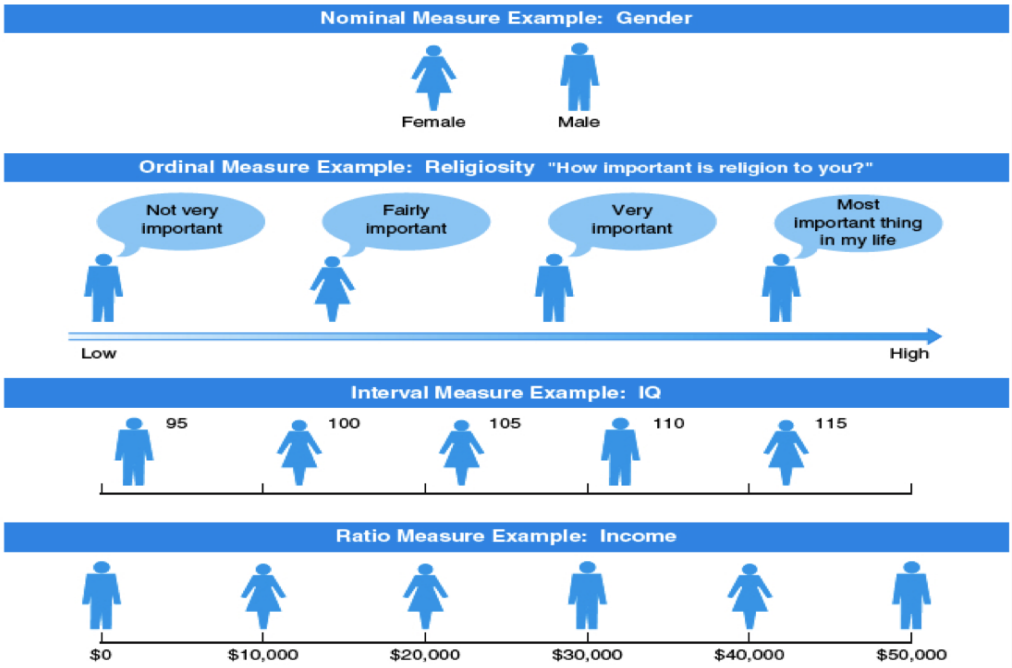
Interval scales are numerical scales in which intervals have the same interpretation throughout. As an example, consider the Fahrenheit scale of temperature. The difference between 30 degrees and 40 degrees represents the same temperature difference as the difference between 80 degrees and 90 degrees. This is because each 10-degree interval has the same physical meaning (in terms of the kinetic energy of molecules).

4) **Ratio scales**

The ratio scale of measurement is the most informative scale. It is an interval scale with the additional property that its zero position indicates the absence of the quantity being measured. You can think of a ratio scale as the three earlier scales rolled up in one. Like a nominal scale, it provides a name or category for each object (the numbers serve as labels). Like an ordinal scale, the objects are ordered (in terms of the ordering of the numbers). Like an interval scale, the same difference at two places on the scale has the same meaning. And in addition, the same ratio at two places on the scale also carries the same meaning.

Typical example age and weight both has zero point. We can say also that 2 meters is twice as long as 1 meter, and at a 40 age-years old is twice as old as 20-years old.


Four Levels of Measurement of data
DATA
QUALITATIVE/ CATEGORICAL
can be grouped
QUANTITATIVE/ NUMERICAL
measure
NOMINAL
cannot be arranged in any particular order
BEST MEASURE OF CENTRAL TENDENCY: MODE
BEST MEASURE OF SPREAD: NONE
ORDINAL
can be arranged in order
BEST MEASURE OF CENTRAL TENDENCY: MEDIAN, MODE
BEST MEASURE OF SPREAD: IQR
INTERVAL
Zero is arbitrary
BEST MEASURE OF CENTRAL TENDENCY: MEAN,MEDIAN, MODE
BEST MEASURE OF SPREAD: RANGE, VARIANCE, S.D AND IQR
RATIO
Zero is not arbitrary and means none
BEST MEASURE OF CENTRAL TENDENCY: MEAN,MEDIAN,MODE
BEST MEASURE OF SPREAD: RANGE, VARIANCE, S.D AND IQR
3


Nominal Measure Example: Gender
Female
Male
Ordinal Measure Example: Religiosity "How important is religion to you?"
Not very important
Fairly important
Very important
Most important thing in my life
Low
High
Interval Measure Example: IQ
95
100
105
110
115
Ratio Measure Example: Income
$0
$10,000
$20,000
$30,000
$40,000
$50,000