**Level of measurement** is a classification that describes the nature of information within the numbers assigned to variables.

**Data Level of measurements:**

1. nominal

a)nominal dichotomous

b)nominal categorical

1. Ordinal
2. Ratio (continuous or scale)
3. Interval

The first step for understanding statistics starts by understanding the level of measurements.

The **dependent variable** is the most important variable that you have to determine in your study or case that you're working on.

For example ;smoking & lung cancer ,where is the outcome? Lung cancer.

-lung cancer is the dependent variable.

Lung cancer is a nominal variable ,which means the pt has lung cancer or does not have lung cancer.

0 or 1

1: having the disease

0:not having the disease

That's what we call nominal dichotomous level of measurement.

There is another kind that we call categorical level of measurement (nominal ) used for medical diagnosis for ex:

1.medical

2.surgical

3.ortho

4.onco

5.rehab

-In this way we give an arbitrary number for each diagnosis ,just for coding.

Numbers here mean nothing & can not be used for addition ,subtraction or anything.

2)ordinal :

Here numbers do mean something, like Pain intensity.

We usually measure pain by :

-visual analogue scale. From (0-100)

In some Rcts ,we use pain intensity as **a RATIO** level of measurement.

We mentioned before that in medical research we don't have RATIO level of measurement because there is no Absolute zero.

Weight ,heart rate ,body temperature,height can't be zero!

Because if those measurements were zero ,the patient is dead or He doesn't exist.

But metaphorically we do use ratio level of measurement (at interval or scale ,continuous e) in medical studies just for the sake of making Rcts for analytical purposes.

So we converge those variables  like (oxygenation level ,hemoglobin levels,BMi) to ratio level of measurement so to analyze them in a robust way using parametric statistics.

The most used software for analysis is **SPSS**.

**SPSS** :Statistical package for social sciences.

This software has hundreds of thousands of programmed mathematical equations ,all you have to do is good interpretation.

You can download this software online if you're really interested.

You only have to know how to deal with outputs.

One of the most important analysis in SPSS is **Univariate analysis** :

It’s analysis that is used to clean the data.

For example :the sample age that we have from (20-100) ,then we realized that we have one age of 1000 ,here we have a problem that later when we analyze the data , the average of the age & it's association with lung cancer will give false outcome.

So that's why we need to clean the data by **Univariate analysis**.

Demographic studies of the patients

Clinical characteristics of the patients

Outcome of the patients related to lung cancer for example

Ways to present the data:

**1)tables** (the best way of data presentation)

-Easy to read

-contains unlimited numbers of information.

**2)figures**

**3)charts**

-Bar chart : usually used to represent data for layperson (ex.school students)

-Vertical

-horizontal

-clustered : when we want to describe **two variables together**

Each color of the chart represent a different variable.

Pie chart : used in economics,companies & stock markets.

4)histogram:

Very important to represent data on the continuous level of measurements

Because it has "the normal curve" that is related to continuous data.

Age ,waiting time at ER ,patients satisfaction,oxygen saturation are all examples of continuous data that the best way to represent it is using histogram.

By this you can see whether your data is deviated or not.

Deviation or skewness is really important when we doفحوص تشخيصية إحصائية.

If you insisted on doing فحص تشخيصي إحصائي on a skewed ,deviated ,shifted data ,you'll definitely get false results.

So we have to normalize the data first.

**5) polygon**

Before we start analysis we do the following steps:

1)data organization

2)data cleaning

3)data categorization

4)data analysis

5)data interpretation

You have to follow those steps ,otherwise if you skip one of them you'll get false interpretation and outcome .

Type 1 error

This error is due to interrupting true data wrongly & it's your fault. It's like executing an innocent person.

Type 2 error

When you generalize results that are not able to be generalized upon population.

**1)Descriptive measures:**

1-central tendency

2-variablity

3-skewess

4-kurtosis

We use descriptive measures to **describe properties of the sample or population** we are working on. And we can’t use them to compare between samples of populations.

Once we want to describe the difference between samples we ought to use ***inferential statistics*** not descriptive ones.

2) **inferential statistics**

This depend on the situation of your research

What we should do to prove or disprove if there is a statistically significant relationship between for ex. Smoking and lung cancer.

So descriptive is all about numbers for ex. We have 100 lung cancer patients 60 smokers ,40 non smokers this should not be enough for you to take a decision,to prove or disprove that smoking causes lung cancer,there's many factors that affect this relation.

You need inferential statistics to know if there is a relation 100%

To take decision upon inference ,you have to do X^2 test (chi square test)

**Main statistical measures:**

1. central tendency:

1)mean:
M = ∑ X/N

  Mean only works to describe continuous data like waiting time in ER,HR

 .but for gender for ex. It doesn't work.

Mean is intended mainly for interval and ratio variables “**continuous values**” .

\*but not in ordinal variables or nominal ones. We cannot say, for instance, the mean of gender = 0.75

 2) median

The **middle value** of a set of ordered numbers “**50thpercentile**”.

* Used in symmetrical and a symmetrical distributions
* The median is **not sensitive to extreme scores**.
* It is useful when the data are **skewed**
* Appropriate in ratio, interval and ordinal variables, but not for nominal data.
* It is useful when giving grades “A,B,C,D” in the university

 3)mode=x bar=M mode.

* The **most frequent value or category** in a distribution.
* Nominal data are best described using mode
* Not calculated, but spotted
* It is appropriate for all variables including the nominal ones.

In a perfect world ,theoretically where data is collected & distrputed in a right way ,then mean=median =mode which doesn’t happen in our world

1. Variability
2. Skewness
* numbers in descriptive measures are usually misleading for ex.tawjihi results but you make it leading through inferential s.

Every central tendency measure has an analogue from **the measures of variability (dispersion).**

1)mean accompanies Standard deviation

2)median accompanies percentile

3)mode accompanies range

In any study ,you should have those 6 descriptive statistics for analysis & each of which represent a different indication.

If your working on continous data ,you should represent it in this way.

If any of those measures were missing ,the study is considered misleading.

For ex. If SD was more than 2,it means there is a high variability .

Percentile ,whenever the 25th percentile is higher the marks are worse.

**P25 =60** that means 25% of the marks is 60 or less.

**P75=90** is good indicator ,75% of the marks is 90 or less.

**Range** :minimum compared to maximum

-HR heart rate for ex,60-66 close together

-Age groups 20-120 :its a wide range that may affect other factors in your study.

For ex, age is another factor that may affect lung cancer not only smoking.

So when the range is wider there is a bigger chance for compounding variables to affect your study.

So those measures should be available for any descriptive study to see if your data is applicable or we can do further modifications or further inferential statistics .