Meaning of Statistics:

Statistics is concerned with scientific methods for
collecting, organising, summarising, presenting and analysing data
The word ‘ statistic’ is used to refer to
1. Numerical facts, such as the number of people living in particular area.
2. The study of ways of collecting, analysing and interpreting the facts.

Medical Statistics:

The Medical statistics deals with applications of statistics to medicine and the health sciences, including epidemiology, public health, forensic medicine, and clinical research.

* Some Basic Definition :

1- Population :-
A population is the group from which are to be collected.
2- Sample :-
A sample is subset of population A sample is defined as a set of selected individuals, items, or data taken from a population of interest.
3- Variable :-
A variable is a feature characteristic of any member of a population differing in quality or quantity from one member to another.
example , The variables to measure or observe might be the height , weight , gender , etc .

Types of Variables :

1- Quantitative Variable :

A Variable differing in quantity is called quantitative Variable , for example , the weight of a person , number of people in a car.
Quantitative Variable are represented by symbols (X,Y,Z).
Can be sub – classified into :

*- Discrete Variable :
Number of nights spent in hospital
Number of courses of a given drug prescribed during the study period
Age at last birthday
Number of cigarettes smoked in a week

* Continuous Variables :

Blood pressure
Lung function, for example peak expiratory flow rate (PEFR)
2- Qualitative Variable :
A variable differing in quality is called qualitative variable or attribute , for example , color , the degree of damage of a car in accident.

There are two subdivisions of statistical method:
(A) Descriptive Statistics : it deals with the presentation of numerical , or , data , in either table or graphs from , and with the methodology of analyzing the data.
(B) Inferential Statistics : it involves techniques for making inferences about the whole population on the basis of observation obtained from samples.

Scales Used to Measure Variables:

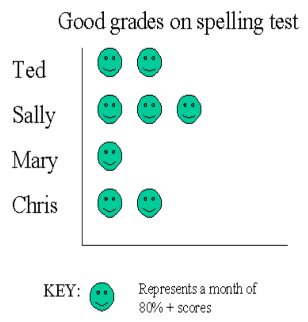
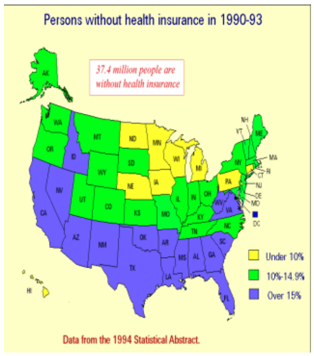
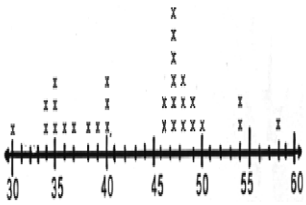
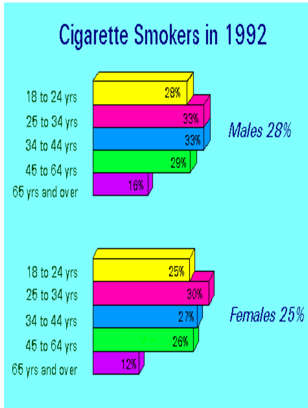
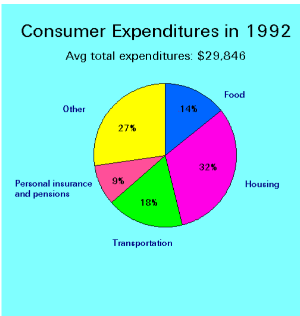
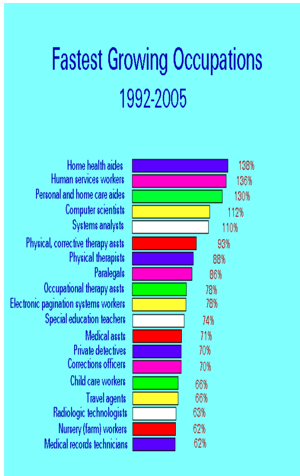
1- Nominal: Gender – respondents can be male or female
2-Ordinal: It is used for the ordered variables (when there is degrees in the variables so we can put them upon each other, in the scale (ex. if the variable is the damage that is caused by a cancer we can put it in categories according to the degree of damages to systems of the body).
3-Interval Scale: Usually not begin from zero (ex. the temperature when we say zero, no mean that there is no temp). i.e. zero, less than, or more than.
4-Ratio Scale: It usually begins from more than zero and it's mainly used for quantitative variables.

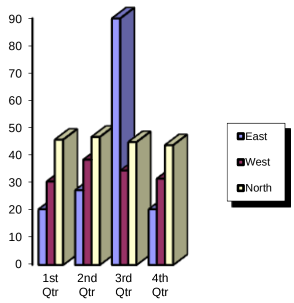
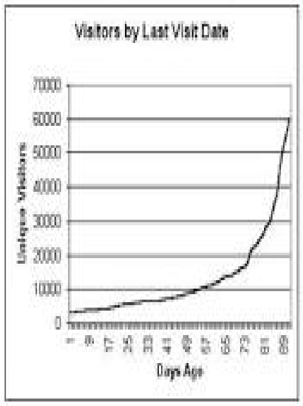
Data type :
1- Grouped Data.

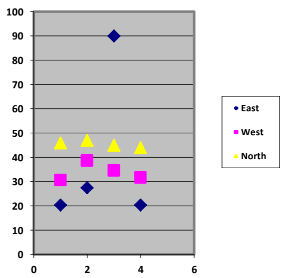
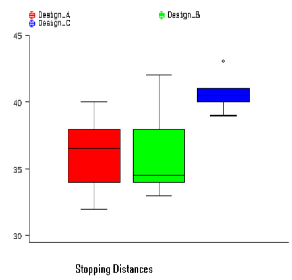
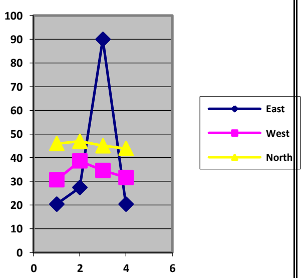
2- Ungrouped Data.

Representing Grouped & Ungrouped Data in Graphs:

Data recorded in experiments or surveys is displayed by a statistical graph. We will discuss eleven types of statistical graphs. Choosing which graph is determined by the type and breadth of the data, the audience it is directed to, and the questions being asked.
Each type of graph has its advantages and disadvantages. Consult the table below when choosing a graph.

Type of Graphs	Example	Type of Graphs	Example
Pictograph A pictograph uses an icon to represent a quantity of data values in order to decrease the size of the graph. A key must be used to explain the icon.		Map Chart A map chart displays data by shading sections of a map, and must include a key. A total data number should be included.	
Line Plot A line plot can be used as an initial record of discrete data values. The range determines a number line which is then plotted with X's for each data value.		Histogram A histogram displays continuous data in ordered columns. Categories are of continuous measure such as time, inches, temperature, etc.	
Pie Chart A pie chart displays data as a percentage of the whole. Each pie section should have a label and percentage. A total data number should be included.		Bar Graph A bar graph displays discrete data in separate columns. A double bar graph can be used to compare two data sets. Categories are considered unordered and can be rearranged alphabetically, by size, etc.	

Frequency Polygon A frequency polygon can be made from a line graph by shading in the area beneath the graph. It can be made from a histogram by joining midpoints of each column.		Line Graph A line graph plots continuous data as points and then joins them with a line. Multiple data sets can be graphed together, but a key must be used.	

Type of Graphs	Example	Type of Graphs	Example
Stem and Leaf Plot Stem and leaf plots record data values in rows, and can easily be made into a histogram. Large data sets can be accommodated by splitting stems.		Box plot A box plot is a concise graph showing the five point summary. Multiple box plots can be drawn side by side to compare more than one data set.	
Scatter plot A scatter plot displays the relationship between two factors of the experiment. A trend line is used to determine positive, negative, or no correlation.			

FREQUENCY DISTRIBUTION

A frequency distribution is constructed for three main reasons:

1. To facilitate the analysis of data.
2. To estimate frequencies of the unknown population distribution from the distribution of sample data and
3. To facilitate the computation of various statistical Measures

Steps of construction a frequency distribution :

1- Number of Classes
value :-
* k (no. of intervals = 1+3.322 logn)
where n : no. of observations .
* the number of classes should be no fewer than six and no more than 10.
A simple formula could be used to find the total number of classes.

2- Find the range. The range in a distance between the lowest and highest values observation.

R (the range) = Xmax – Xmin

3- Compute the width may be determined by dividing the range , by k , the number of class intervals symbolically , the class interval width , given by :

$$W = \frac{R}{K}$$

4 - Class Limits :

The class limits are the lowest and the highest values that can be included in the class. For example, take the class 30-40. The lowest value of the class is 30 and highest class is 40. The two boundaries of class are known as the lower limits and the upper limit of the class. The lower limit of a class is the value below which there can be no item in the class. The upper limit of a class is the value above which there can be no item to that class. Of the class 60-79, 60 is the lower limit and 79 is the upper limit, i.e. in the case there can be no value which is less than 60 or more than 79. The way in which class limits are stated depends upon the nature of the data. In statistical calculations, lower class limit is denoted by L and upper class limit by U.

5- Mid-value or mid-point:

The central point of a class interval is called the mid value or mid-point. It is found out by adding the upper and lower limits of a class and dividing the sum by 2.
(i.e.) Midvalue = L +U/2

For example, if the class interval is 20-30 then the mid-value is

$$\frac{20+30}{2}=25$$

6-find the frequency for each class.

7-true limits.

Note:

Frequency:Certain guidelines should be followed :

If the data values are integer , the lower limit of the first class should be 0.5 less than the lowest data value. The mid point of the class should bean integer.

Cumulative frequency distribution

The cumulative is a progressive total of the frequency. The cumulative frequency of score gives the number of scores equal to or less than , that particular score.

There are two types of cumulative F.D. :

1- Ascending accumulation.
2- Descending accumulation.

Relative Frequency Distribution :

The proportion of value in each class , it is determine by dividing the no. of observation in interval by total no. observation.

Example: the weight of 57 subjects is as the follows:

17	25	37	42	56	68	72
52	62	35	25	32	45	31
45	44	43	67	28	29	21
34	22	26	58	46	54	46
79 max.	48	40	24	27	38	20
49	33	47	42	63	41	30
36	21	23	37	22	28	27
24	15	25	18	26	19	**12 min.**
28						

Solution:

n = 57

k = 1 + 3.322 log(n) = 1 + 3.322 log(57) = 1 + 3.322 (1.7559) **≈ 7**

Range = max. – min. 79 – 12 = **67**

Class length = R / K = 67 / 7 = 9.6 ≈ **10**

Class Interval	Frequency	Mid point	True limit	Ascending Cumulative Frequency	descending Cumulative Frequency	Relative Frequency
10 – 19	5	14.5	9.5-19.5	5	57	8.772
20 – 29	19	24.5	19.5-29.5	24	52	33.333
30 – 39	10	34.5	29.5-39.5	34	33	17.54
40 – 49	13	44.5	39.5-49.5	47	23	22.8
50 – 59	4	54.5	49.5-59.5	51	10	7.0
60 – 69	4	64.5	59.5-69.5	55	6	7.0
70 – 79	2	74.5	69.5-79.5	***57***	2	3.5

EX1 //

The Wight of 40 subject in as the follows :

Class interval	Frequency
118-126	3
127-135	5
136-144	9
145-153	12
154-162	5
163-171	4
172-180	2

Class interval	Frequency	Mid point	True Limits	F.D. ascending	F.D. descending
118-126	3	122	117.5-126.5	3	40
127-135	5	131	126.5-135.5	8	37
136-144	9	140	135.5-144.5	17	32
145-153	12	140	144.5-153.5	29	23
154-162	5	158		34	11
163-171	4	167		38	6
172-180	2	176		40	2

EX2 //

The Wight of 40 subset in as follows , Find Relative Frequency:

Class interval	fi
118-126	3
127-135	5
136-144	9
145-153	12
154-162	5
163-171	4
172-180	2
	∑=40

Class interval	fi	Relative Frequency
118-126	3	7.5
127-135	5	12.5
136-144	9	5.0
145-153	12	30.0
154-162	5	12.5
163-171	4	10.0
172-180	2	5.0
	∑=40	

Mathematical Presentation :

(A) Measures of central Tendency although a frequency distribution serve useful purpose , there are many situations that require other types of data summarization , what we need in many instances is the ability to summarize data by mean of just a few descriptive measures.

(B) Descriptive measures may be computed from the data of the sample (called statistic) or the data of population (called parameter)

The most commonly used measures are :

1- Arithmetic Mean.

2- Median.

3- Mode.

4- Geometric Mean.

1- Arithmetic mean:

The arithmetic mean – or simply mean is computed summing all the observation in sample and dividing the sum by the number of observations symbolically , the mean is represented by :

For Population		*For Sample*	
Ungrouped Data	Grouped Data	Ungrouped Data	Grouped Data
$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\mu = \frac{\sum_{i=1}^{k} (f_i x_i)}{\sum_{i=1}^{k} f_i}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^{k} (f_i x_i)}{\sum_{i=1}^{k} f_i}$

Where:

μ : Symbol of population mean (Parameter).
$\bar{x}$: Symbol of sample mean (statistics).
x_i : Value of every unit.
N : Population size.
n : Sample size.
f_i : Normal frequency.
k : No. of classes.

Advantages of the mean:

1. It is a unique (single no.): that converts all data into a single no.
2. Simplicity: Simple & easy
3. Take in consideration all the values.

Disadvantages of the mean:
It is greatly affected by extreme values (the biggest and the smallest).

EX1 //

The mark obtained in 10 class test is are :

25 – 20 – 20 – 9 – 16 – 10 – 21 – 12 – 8 – 13

Solution ///

The mean = $\bar{X} = \sum_{i=1}^{n} x/n$

$$= \frac{25+20+20+9+16+10+21+12+8+13}{10}$$

$$\bar{X} = \frac{154}{10} = 15.4$$

EX2 //
The following data represents income distribution of 100 families , calculate mean income of 100 families.

class	fi
30 – 39	8

40 – 49	12
50 – 59	25
60 – 69	22
70 – 79	16
80 – 89	11
90 – 99	6

Find the Mean..

Solution //

class	fi	Mid Point	fi xi
30 – 39	8	34.5	276
40 – 49	12	44.5	534
50 – 59	25	54.5	1362.5
60 – 69	22	64.5	1419
70 – 79	16	74.5	1192
80 – 89	11	84.5	929.5
90 – 99	6	94.5	567

We yet $N = \sum fi = 100$

$\sum fi\ xi = 6280$

$$\bar{X} = \frac{\sum_{i=1}^{n} fi\ xi}{N}$$

$$= \frac{6280}{100} = 62.8$$

2- The Median :

Median is defined as the middle item of all given observation arranged in order. For ungrouped data, the median is obvious.
In case of the number of measurement is even , the median is obtained by taking the average of the middle.
* If there are an odd number of scores the median is the middle score:

$$\text{Median} = \frac{N+1}{2}$$

EX// 1 , 3 ,5, 7 , 8 , 13 , 15 , 17 , 18 , 21 , 23

Find the Median..

Solution //

$$\bar{M} = \frac{n+1}{2} = 11+1 / 2 = 6$$

$\bar{M}$ = 13

* If there are an Even number of scores , the Median is the Midpoint between the two middle scores :

$$\text{Median} = \left(\frac{n}{2}, \frac{n}{2}+1\right)$$

EX// 1 , 3 , 6 , 7 , 8 , 13 , 15 , 17 , 18 , 23

Solution //

$$\text{Median} = \left(\frac{n}{2}, \frac{n}{2}+1\right) = (5,6)$$

$$= (8 + 13) / 2 = 10.5$$

Note !
When determining the Median , you must arrange the scores in ascending or descending order first.

For Grouped data :

$$\bar{Me} = L1 + \left[\frac{(\sum fi/2) - Fi}{fi}\right] * W$$

The 1st steps is compute the ascending cumulative frequency table , and :

L1 = The lower actual limit for median class.
∑ fi = The summation of the normal frequency.
Fi = The cumulative frequency at the beginning of the median class.
fi = The normal frequency of the median class , and can compute by:

fi = cumulative freq. at the end of Median class _
cumulative freq. at the begin of Median class

w = The class length.

Advantages of the median:

1. It is a unique (single no.): that convert lots of data into a single no.
2. Simplicity: Simple & easy to be calculated and understood.
3. Not effected by extreme data.

Disadvantages of the median:
It is neglect all the values and take only the median one (the central value only).

EX //

Find The median for the following frequency table is :

Classes	fi
8 – 10	5
11 – 13	7
14 – 16	10
17 – 19	15
20 – 22	8
23 – 25	3
26 – 28	2
	50

Solution //

Classes	fi	Ascending cumulative	True limits
8 – 10	5	5	7.5 – 10.5
11 – 13	7	12	10.5 – 13.5
14 – 16	10	22	13.5 – 16.5
17 – 19	15	27	16.5 – 19.5
20 – 22	8	45	19.5 – 22.5
23 – 25	3	48	22.5 – 25.5
26 – 28	2	50	25.5 – 28.5
	50		

L1 = 16.5
F = 22
fi = 15
w = 3

$$\overline{Me} = L1 + \left[\frac{(\sum fi/2) - Fi}{Fi} \right] * W$$

$$\overline{Me} = 16.5 + \left(\frac{25 - 22}{15} \right) * 3$$

$$= 17.1$$

3- The Mode :

The mode is the observation that occurs most frequently if all values are different there is no mode.
On the other hand, a set of values are may have more than on mode.
The mode is not the greatest variable is the variable that is repeated the greatest no. of times.
The mode for grouped data is :

$$\overline{Mo} = L1 + \left(\frac{d1}{d1 + d2} \right) * W$$

The 1^{st} step: the mode class = the class that is the most frequent.
L1 = upper actual limit of the class that come before the mode class.

d1 = the distance between the frequency of the mod and the frequency of the class that come before the mode class.
w = the class length or width.
EX// Find the mood
10, 10, 2, 2, 5, 7, 9, 8, 9, 9
Solution :

$\overline{Mo}$ = 9

EX// find the mood

Class	frequency
150 – 154	3
155 – 159	7
160 – 164	10
165 – 169	15
170 – 174	8
175 – 179	5
180 – 184	2

Solution //
L1 = 164.5
d1 = 5
d2 = 7
w = 5

$$\overline{Mo} = 164.5 + \left(\frac{5}{5 + 7} \right) * 5$$

$$= 166.6$$

Advantages of Mode :
1- It is easy to understand & easy to calculate.
2- It is not effected by extreme values or sampling fluctuations.
3- It is always present within the data.
4- It can be located graphically.

disadvantages of mode :
1- It is not rigidly defined.
2- It is not based upon all values of the given data.
3- It is not capable of further treatment.

* Relation between mean and median and mood :

$$\overline{X} - \overline{Mo} = 3 (\overline{X} - \overline{Me})$$

Measures of dispersion

Knowing a distributions central tendency is helpful , but it is not enougf.it is also important to know whether the observations tend. To be quite similar (homogenous) or vary considerably heterogeneous.
To describe variability,measures of variation have been devised. The most common of these are the
1 – the range
2 – the variance
3 – the standard deviation
4 – the standard error of the mean
5 – the mean divition.

1 – range : the range is defind as the difference in value between the highest (maximum) and lowest (minimum) observation:
Range = $x_{max} - x_{min}$
The range can be computed quickly but it is not very usful because it considers only the extremes and does not take into consideration the bulk of the observations.

Example.
x_i= 9 – 7 – 12 – 15 – 3 – 4
R= 15 – 3 = 12

2- The varians.
When the values of observations lie close to their mean, the dispersion is less than when they are scaitered over a wide range. If we could measure dispersion relative to the scalter of the values about theire mean. Such a measure is realized in what is known as the variance.

$$\left.\begin{array}{c} \sigma^2 = \dfrac{\sum (x_i - \mu)^2}{N} \\ or \\ \sigma^2 = \dfrac{N \sum x_i^2 - (\sum x_i)^2}{N^2} \end{array}\right\rVert \; un\ grouped\ data$$

Where σ^2 is the variance of population

$$\left.\begin{array}{c}\sigma^2 = \dfrac{\sum f_i (x_i-\mu)^2}{\sum f_i} \\ or \\ \sigma^2 = \dfrac{\sum f_i x_i^2 - \dfrac{(\sum f_i x_i)^2}{\sum f_i}}{\sum f_i}\end{array}\right] grouped\ data$$

And the formula for sample

$$\left.\begin{array}{ll} s^2 = \dfrac{\sum (x_i-\bar{x})^2}{n-1} & n < 60 \\ or & \\ s^2 = \dfrac{\sum (x_i-\bar{x})^2}{n} & n \geq 60\end{array}\right] ungrouped\ data$$

And

$$\left.\begin{array}{c} s^2 = \dfrac{\sum f_i (x_i-\bar{x})^2}{(\sum f_i)-1} \\ or \\ s^2 = \dfrac{\sum x_i^2 - \dfrac{(\sum f_i x_i)^2}{\sum f_i}}{(\sum f_i)-1}\end{array}\right] grouped\ data$$

Example 1

Find the variance

7.1 , 2.5 , 2.5 , 5.4 , 8.3

Solution

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
7.1	1.94	3.7636
2.5	-2.66	7.0756
2.5	-2.66	7.0756
5.4	0.24	0.0576
8.3	3.`14	9.8596
Total 25.8	0.00	27.832

n=5

$$\bar{X} = \frac{\sum x_i}{n} = \frac{25.8}{5} = 5.16$$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$= \frac{27.832}{5-1} = \frac{27.832}{4} = 6.958$$

b) Find the variance for the following frequency table (grouped data):

Classes	Normal frequency (f_i)
60-62 63-65 66-68 69-71 72-74	5 18 42 27 8
Total	100

Solution

Classes	Normal frequency (f_i)	Class mid-point (X_i)	$(f_i X_i)$	$X_i - X$	$(x_i - \bar{x})^2$	$fi(x_i - \bar{x})^2$
60-62 63-65 66-68 69-71 72-74	5 18 42 27 8	61 64 67 70 73	305 1152 2814 1890 584	6.45 3.45 0.45 2.55 5.55	41.60 11.90 0.20 6.50 30.50	208.01 214.245 8.505 175.56 246.42
Total	100		6745		226.5	852.75

$$s^2 = \frac{\sum f_i \, (x_i - \bar{x})^2}{(\sum f_i) - 1}$$

The arithmetic mean = (6745 / 100) = 67.45

$$s^2 = \frac{852.75}{100 - 1} = 8.6$$

Stander deviation

It is the square root of the mean of the squared deviations from the arithmetic mean. It is denoted by the small greek letter σ (sigma).

$S.D = s = \sqrt[2]{s^2}$

Example

Find S.D of data

4 , 6 ,5 ,8 ,7

Solution

$$\sum x_i = 30$$

$\bar{X} = \frac{\sum x_i}{n} = 6$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{4+0+1+4+1}{4} = \frac{10}{4} = 2.5$$

$S.D = \sqrt[2]{s^2}$

$= \sqrt[2]{2.5}$

=1.58

4- The standard error of the mean

standard error of the mean is measures the variabThe standility of the mean of asample as an estimation of the true value of the mean of population from which this sample has been (i.e to indicate the degree in which the sample mean reflect the population mean).

$SE = \frac{SD}{\sqrt{n}}$

5 – the mean divition.

the mean divition is also known as the average deviation. It is the average difference between the items in a distribution and the median or mean of that series.

M.D = $\left.\frac{\sum|x_i - \bar{X}|}{n}\right\}$ ungrouped data

And sam formula as discrete series

$M.D = \frac{\sum f_i |x_i - \bar{X}|}{\sum f_i} \}$ grouped data

Where

f_i = normal frequence

x_i= the class mid point

Example 1

Find M.D for data

2 , 8 , 3 , 7 , 6 , 4

M.D = $\frac{1}{6}[|2-5|+|8-5|+|3-5|+|7-5|+|6-5|+|4-5|]$

$=\frac{1}{6}[3+3+2+2+1+1]=2$

Exampla 2

Find M.D for data

class	f_i	Mid point	$\lvert x_i - \bar{X} \rvert$	$f_i \lvert x_i - \bar{X} \rvert$
15-19	3	17	16.5	49.5
20-24	5	22	11.5	57.5
25-29	7	27	6.5	45.5
30-34	15	32	1.5	22.5
35-39	16	37	3.5	56
90-44	8	42	8.5	68
15-49	4	47	13.5	54
	58			353

$\bar{X} = 33.55$

M.D=$\frac{\sum f_i |x_i - \bar{X}|}{\sum f_i}$

$= \frac{3(16.5)+5(11.5)+\cdots+4(13.5)}{58} =$

M.D=$\frac{353}{58}$

=6.086

Relative variation measure

Coefficien of variation

Standard deviation isusful as a measure ofv variation with in agiven set of data . When one desire to compare the dispersion in two set of data , however comparing the two standard deviation may lead to fallacious results.

$$C.V = \frac{SD}{\bar{X}} * 100$$

Example

Suppose that two sample of human males tield the following results.

	Sample 1	Sample 2
Age	25 years	11 years
Mean wight	145 pounds	80 pounds
S.D	10 pounds	10 pounds

Compute C.V for two sample .

$C.V = \frac{SD}{\bar{X}} * 100$

$C.V1= \frac{10}{145} * 100=6.9$

$C.V2= \frac{10}{80} * 100=12.5$

C.V of sample 2 larger than C.V of sample 1

Dispersion of sample 2 larger than C.V of sample 1

(B) Inferential Statistics

And was included two main parts :
1- Estimation :
That a meaning find the estimated values of inference from it to exact values from the sources of information (These estimation values are , either point estimation or interval estimation).
2- Test of Hypothesis :
(Including testing the NULL Hypothesis CHO) to making the primary inference for the study to resaved the aim of reject or accept the NULL hypothesis.

The steps of testing the Hypothesis:

1. Specification of the form of the population distribution:
   a. Parametric Statistical Methods.
   b. Non-Parametric Statistical Methods.
2. Formulation of the Null & Alternative Hypotheses, H_0 & H_1.
3. Selecting the level of Significance or probability level or error level.
4. Selected the Statistical test.
5. Summation the data from the sample and calculate the *test-statistic*.
6. Conclusion.

a. Parametric Statistical Methods:

.t-test:

this test is used to compar a sample mean to the average community.
Also called student test, and is applied in the following situations:
1- one sample T Test
2-independent sample T Test
3-paired sample T Test

I) $t = \frac{y-k}{\frac{s}{\sqrt{n}}}$ for one sample.

$\bar{y}$ =mean of sample
k =assumed population mean
s=sample standerd divition
n=size of sample

Example

Compare the average diabetic healing of patients in hospital with average for patients whose estimated 2.5 for the same disease,when the level of significance of 0.05

Sample: 3.8 4.1 3.4 2.9 3.3 2.5

Solution:

H_0:k_0=2.5

H_0:k_0 ≠2.5

$\bar{y}$=3.33

K=2.5

$\bar{y}$-k=0.83

S^2= 0.3386

Sd=0.5818

Se= 0.2375

$$t=\frac{y-k}{\frac{s}{\sqrt{n}}}$$

t= 0.83/0.2375=3.494

t(0.05) = 1.832

t(0.05) < t

∴ reject zero hypothesis

2-independent sample T Test

Statistical method used to detect significant difference between medium separate sets the continues data like detecting the difference between arthimetic of male and arthimetic of female.

II) $t=\frac{(y_1-y_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ if $\sigma_1^2=\sigma_2^2$ and unknown

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad , \quad d.f. = n_1 + n_2 - 2$$

$\overline{y_1}$ =mean of first sample
$\overline{y_2}$ =mean of second sample
s_1^2= variance of first sample
s_2^2= variance of first sample
s_p^2= A common variation samples
$d.f.$ =degree freedom

Example

Took two sample of male and female thalassemia pattients and spotted the following results

	female	male
n	10	12
$\bar{y}$	71.5	68
s	6	8

average female Question can you find that healing is greeater than the average male healing when this α=0.1

Solution:

H_0:$\mu_1 - \mu_2$=0

H_1:$\mu_1 - \mu_2$>0

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

s_1^2=36
s_2^2=64
s_p^2=51.4

$$t = \frac{(y_1 - y_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{(71.5 - 68) - 0}{\sqrt{\frac{51.4}{10} + \frac{51.4}{12}}}$$

t=1.14

$t=1.14<t(0.1)=1.28$

So iwill accept the zero hypothesis

3-paired sample T Test

Used to compre the average two sets associated with any of the same group its called a test (pre , post).

Like comparing the average cure diabetes before and after taking the course treatment.

$$t = \frac{\sum D}{\frac{\sqrt{n\sum D^2 - (\sum D)^2}}{n-1}}$$

Example:

If we have a group of (8) students and their academic achievement scores were recoreded before and after applying the performance improvement program, required are statistically significant differences exist between the midde two applications α=0.05 .note creitical t = 2.365

Solution:

$H_0:\mu_1 = \mu_2$

$H_1:\mu_1 \neq \mu_2$

Pre.	post	D	D^2
10	12	2	4
50	52	2	4
20	25	5	25
8	10	2	4
115	120	5	25
75	80	5	25
45	50	5	25
170	175	5	25
		$\sum D = 31$	$\sum D^2 = 137$

$$t = \frac{\sum D}{\frac{\sqrt{n\sum D^2 - (\sum D)^2}}{n-1}}$$

$$t = \frac{31}{\frac{\sqrt{8(137) - (31)^2}}{8 - 1}} = 7.059$$

Since calculated t = 7.06
creitical t = 2.365
reject H_0 & accept H_1
conclution there is a significant difference degree pre. Post.

2- *Z- test:*

Used when the standard deviation is known unlike t test.

I) $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ for one sample.

$\bar{x}$: sample maen

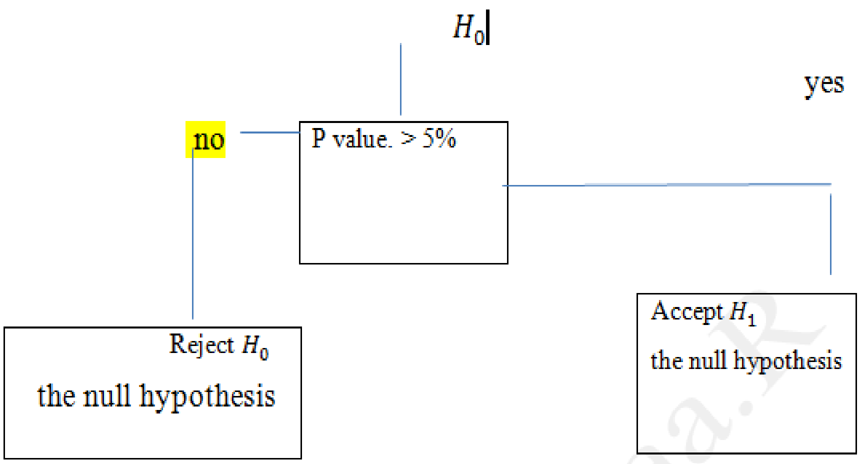
μ_0 : sample population

σ : standard deviation of population

n : sample size

P. value

More statistical programs used P. value for the purpose of imposing accept the zero hypothesis.



casses of the P. value

1- H_0:$\mu \geq \mu_0$ H_1:$\mu < \mu_0$	P. value = $[z < z_{cal}]$
H_0:$\mu \leq \mu_0$ H_1:$\mu > \mu_0$	P. value = $[z < z_{cal}]$
H_0:$\mu = \mu_0$ H_1:$\mu \neq \mu_0$	P. value = $2P[z > z_{cal}]$

Example

$\bar{x}$	μ_0	σ	n
16	20	11	36

Sloution:

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

P(Z<-2.18)=0.0146

reject the zero hypothesis.

H_0:$\mu \geq \mu_0$

H_1:$\mu < \mu_0$

P. value = $[z < z_{cal}]$

a. Non-Parametric Statistical Methods:

In many of the nonparametric statistical tests, the data are changed from scores to ranks or even to signs. Such methods may arouse the criticism that they “do not use all of the information in the sample” or that they “throw away information”.

Types of Nonparametric Statistical Tests:
There are many types of nonparametric statistical tests, some of them as follows:

1. The Chi-Square Family, Tests.
2. The Kruskal-Wallis One-Way Analysis of Variance by Rank.
3. The Friedman Two-way Analysis of Variance by Ranks.
4. The Wilcoxon-Mann-Whitney Family, tests.
5. The Durbin test for Incomplete Block Designs.
6. The Cochrans Test for Related Observations.
7. The Binomial Test (the single-sample case).
8. The Kolmogorov-Smirnov One-Sample Test.
9. Test For Distributional Symmetry.
10. The One-Sample Runs Test Of Randomness.
11. The Change-Point Test.
12. The McNemar Change Test.

1) The Family of Chi-Square Tests:

1. The Chi-Square Goodness-Of-Fit Test: in this test we use absolute numbers. Discrete variable in one sample is compared with another discrete variable in another sample. Which can call the test of goodness of fit (χ^2) or test of independences?

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \quad , \quad \text{d.f.} = (r-1)(c-1)$$

Where: O_{ij}= observed value.

And E_{ij} = expected value. And equal $(TR_i*TC_j)/TT$.

The table is:

	C_1	C_2	C_3	C_4	C_5	TR_i
R_1	O_{11}	O_{12}	O_{13}	O_{14}	O_{15}	TR_1
R_2	O_{21}	O_{22}	O_{23}	O_{24}	O_{25}	TR_2
R_3	.				.	.
R_4	.				.	.
R_5	O_{51}	.	.	.	O_{55}	.
TC_j	TC_1	TC_2	.	.	TC_5	TT

Example:

Distribution by socioeconomic class of patients admitted to self-poisoning (sample A), and gastro-enterological (sample B), units. Here we are going to see if the distribution according to social class differs among the two samples.

Socioeconomic class	Samples		Total
	A	B	
I	17	5	22
II	25	21	46
III	39	34	73
IV	42	49	91
V	32	25	57
Total	155	134	289

First we put forward the *null hypothesis* of no difference, with an *alternative hypothesis* that there is a real difference in the distribution according to social classes among the two samples.
The first step is to calculate the expected value for each cell. This we can do as follows:
E(expected value) = (Row total * column total) / (grand total of the value)
For instance if we want to calculate the expected value of 17=(22*155)/289=11.8, and so on with each cell. Then we will construct a table and as follows:

Class	Expected number		O-E		$(O-E)^2/E$	
	A	B	A	B	A	B
I	11.8	10.20	5.20	-5.20	2.292	2.651
II	24.67	21.33	0.33	-0.33	0.004	0.005
III	39.15	33.85	-0.15	0.15	0.001	0.001
IV	48.81	42.19	-6.81	96.81	0.950	1.000
V	30.57	26.43	1.43	-1.43	0.067	0.077
Total	155	134	0	0	3.314	3.833

We have to notice the following:

i. The sum of the expected numbers is always equal to the sum of the observed.
ii. The sum of the differences between the observed and expected numbers always comes to zero.
iii. Each difference for sample A is matched by the same figure but with opposite sign for sample B.

Now we apply the equation of the Chi-square:
Chi square = 3.314 + 3.833 = *7.147*
Then we calculate the degrees of freedom as follows:
d.f. = (number of columns -1) * (number of rows -1)
= (2-1)*(5-1) = 4
They we use the chi square distribution Table, and we enter at the degree of freedom of 4 and see where our calculated chi square is located. It is located between 3.357 and 7.779, we follow the columns up to get the probability, which we find it lies between 0.05 and 0.01, we express our result as $0.05 > p > 0.01$. the value is much more above 0.05 level, so it is found that there is no significant difference in the distribution among the two samples, so null hypothesis is not disprove.

Analysis of variance (ANOVA)
Use the test to compare the averages of three samples to more than one independent variable. Like comparing of thalassemia patients healing from three different hospitals.

1- one way anova
2- two way anova
3- three way anova
4- multy way anova

Example:

	a	b	c	
	10	2	3	
	20	4	6	
	30	6	9	
	40	8	11	
	50	10	13	
total	150	30	42	222(*T..)*

1-

H_0:$\mu_A = \mu_B = \mu_C$

H_0: one of the averages at least different

F(CAL)>F(TAB)
reject zero hypothesis

N= n1+n2+n3
N= 5+5+5=15

3-
C.F
$C.F = \frac{T^2}{N}$ = $\frac{(222)^2}{15}$ =3285.6
4-
SST= $\sum Y_i^2 - C.F$

SST= $[10^2 + 20^2 + \cdots + 13^2] - 32856$

SST= 2850.4

5-

$$SSA = \frac{\sum (Y_i)^2}{n} - C.F$$

SSA= $\left[\frac{(150)^2}{5} + \frac{(30)^2}{5} + \frac{(42)^2}{5}\right] - 3285.6$

SSA= 1747.2

6- SSE
SST=SSA+SSE
SSA=1747.2
SST=2850.4
SSE=2850.4 - 1747.2=1103.2

7-
N=15 , K=3
df(A) =K-1
=3-1=2

df(A) =N-K
=15-3=12

df(A) =N-1
=15-1=14

8-

F	MS	df	SS
$F = \frac{MSA}{MSE}$	$MSA = \frac{SSA}{K-1}$	K-1	SSA
	$MSE = \frac{SSE}{N-K}$	N-K	SSE
		N-1	SST

F	MS	df	SS
9.5	873.6	2	1747.2
	91.93	12	1103.2
		14	2850.4

F(CAL)>F(TAB)

9.5>3.89

reject zero hypothesis

Correlation and regression

Introduction:

In practical appications, we might come across certain set of data,where each item of the set may comprise of the values of two or more variables. Suppose we have a set of 30 student in a class and we want to measure the heights and weights of all the students.we observe that each individual (unit) of the set assumes two values , one relating to the height and the other to the weight. Such a distribution in which each individual or unit of the set is made up of two values is called a bivariate distribution. The following example will illustrate clearly the meaning of bivariate distribution:

1 -in a class of 60 students the series of marks obtained in two subjects by all of them .
2 -the series of sales revenue and advertising expenditure of two companies in a particulare year.

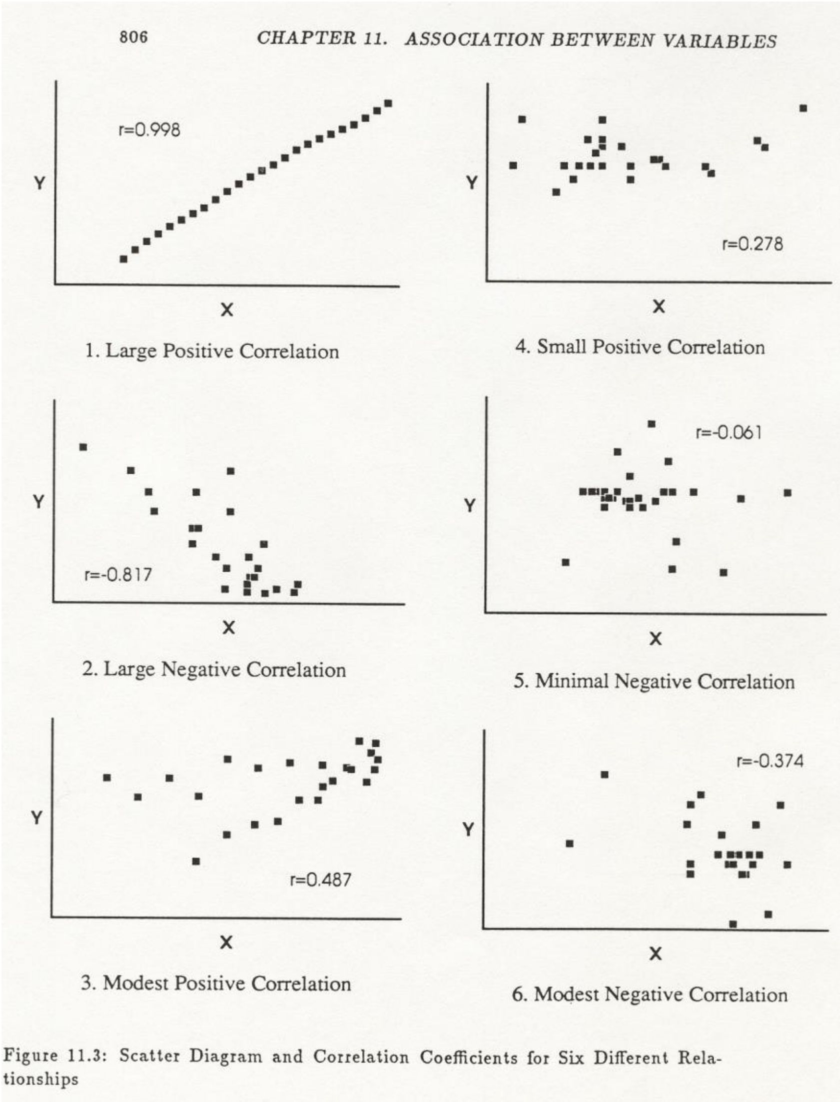
3 -The series of ages of husbands and wives in sample of selected married

Example:
The folloing data are the height and weights of 15 students of a class. Draw a graph to indicate whether the correlation is negative or positive:

Heights (cms.)	Weights (kgs.)
170	65
172	66
181	69
157	55
150	51
168	63
166	61
175	75
177	72
165	64
163	61
152	52
161	60
173	70
175	72

2-linear and non – linear correlation.
The correlation between two variable is said to be lineare, if the change of one unit in one variable result in the corresponding change in the other variable over the entire range of values.

The coefficient of correlation:



Pearson's r

Pearson's r summarized the relationshib between two variables that have a straight line or linear relationship with each other.

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

Example

What is the Pearson's correlation coefficient for MDI and linguistic skills?

x	y	xy	x^2	y^2
3	2	6	9	4
4	2	8	16	4
2	2	4	4	4
2	1	2	4	1
2	1	2	4	1
2	1	2	4	1
$\sum$ 15	9	24	41	15

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

$$r_p = \frac{6(24) - (15)(9)}{\sqrt{(6 \times 41 - (15)^2)(6 \times 15 - (9)^2)}}$$

$$= \frac{144-135}{\sqrt{(246-225)(90-81)}} = \frac{9}{\sqrt{189}} = \frac{9}{13.75} = 0.65$$

Spearman's correlation coefficient

The Spearman's rank correlation coefficient is very useful , and r Latively easy, statistic to calculate, When at least one of the variables has no more than an ordinal level of measurement, the Spearman correlation is often used.The Spearman's rank correlation coefficient between x and y is defind as

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$$

Example

The data given below are obtaind from student records.

Calculate the rank correlation r_s from the data.

number	x	y	Rank of x	Rank of y	D	D_i^2
1	8.5	2300	7	5	*2*	*4*
2	8.6	2250	5	7	-2	4
3	9.2	2380	3	2	1	1
4	9.8	2400	1	1	0	0
5	8.0	2000	8	9	-1	1
6	7.8	2100	9	8	1	1
7	9.4	2360	2	*3*	-1	1
8	9.0	2350	4	4	0	0
9	7.2	2000	10	*10*	0	0
10	8.6	2260	6	6	0	0

N = 10

D_i^2 =12

$$r_s = 1 - 6\frac{12}{10(100-1)}$$

1-0.0727 =0.9273

Linear regression

Linear regression gives the equation of the straight line that describes how the y variable increases (or decreases) with an increase in the x .
The equation of the regression line is

$$\hat{y} = a + bx$$

$$b = \frac{n\sum xy-(\sum x)(\sum y)}{n(\sum x^2)-(\sum x)^2}$$

$$a = \frac{\sum y - b\sum x}{n}$$

Example

number	*x*	*y*	x^2	*xy*
1	2	7	4	14
2	4	5	*16*	*20*
3	3	4	*9*	*12*
4	6	3	*36*	*18*
5	8	2	*64*	*16*
6	1	9	1	9
	24	30	130	89

$b= \frac{n\sum xy-(\sum x)(\sum y)}{n(\sum x^2)-(\sum x)^2}$

$$\sum x_i y_{i=89}$$

$b= \frac{6(89)-(24)(30)}{6(130)-576} = -0.91$

$a=\frac{\sum y-b\sum x}{n}$

$a=\frac{30-(-0.91)24}{6}$

a=8.64

$$\hat{y} = a + bx$$

= 8.64 +(- 0.91) x

= 8.64 – 0.91 × 1

=7.73

المقدمة

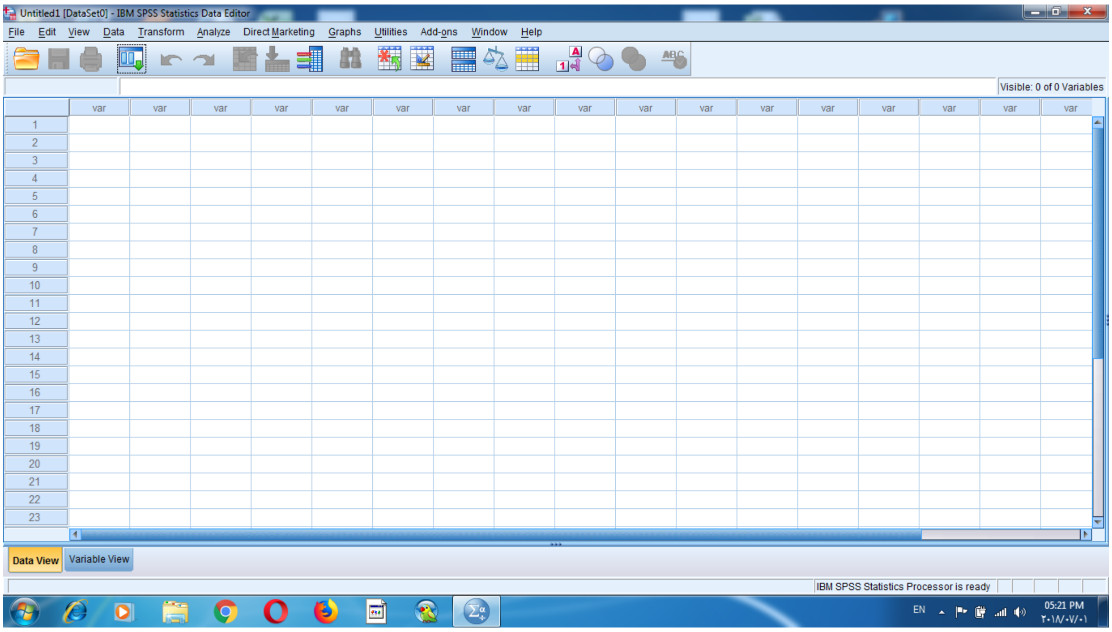
(Statistical package for social sciences) SPSS

يستخدم البرنامج عادة في جميع البحوث العلمية التي تشمل على العديد من البيانات الرقمية ولا تقتصر على البحوث الاجتماعية فقط بالرغم من أنها أنشأت أصلا لهذا الغرض ، ولكن اشتماله على معظم الاختبارات الإحصائية تقريبًا وقدرته الفائقة في معالجة البيانات وتوافقه مع معظم البرمجيات المشهورة جعل منه أداة فاعلة لتحليل شتى أنواع البحوث العلمية.
وتستطيع SPSS قراءة البيانات من معظم أنواع الملفات لتستخدمها لاستخراج النتائج على هيئة تقارير إحصائية أو أشكال بيانية أو بشكل توزيع اعتدالي أو إحصاءًا وصفيًا بسيطًا أو مركبًا وتستطيع الحزم جعل التحليل الإحصائي مناسبًا للباحث المبتدئ والخبير على حد سواء.

تشغيل البرنامج

يعمل البرنامج الاحصائي spss في بيئة النوافذ ويتم تشغيله باختيار الامر start وبعد ذلك نحدد البرنامج spss.

الشاشة الرئيسية للبرنامج



يتجزأ البرنامج إلى 4 أقسام:

1 -COMMAND FUNCTIONS (لائحة الأوامر)

2 -DATA VIEW (شاشة البيانات)

3- VARIABLE VIEW (شاشة تعريف المتغيرات)

4- OUTPUT NAVIGATOR (لائحة التقارير والمخرجات)

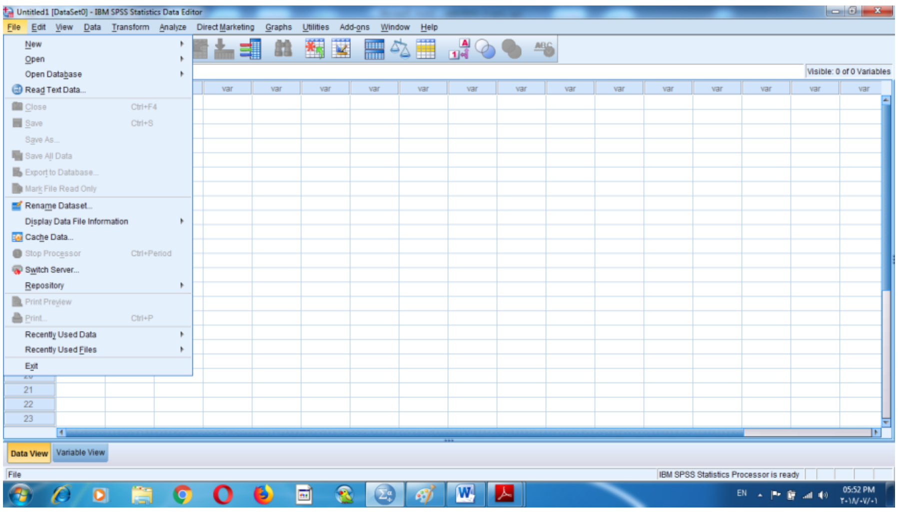
1 -COMMAND FUNCTIONS **(لائحة الاوامر)**

يحتوي محرر البيانات على صفوف وأعمدة، فالأعمدة عبارة عن متغيرات Variables ويعين لكل متغير عمود معين، أما الصفوف فتمثل الحالاتCases ويعين لكل حالة صف معين برقم.
ومحرر البيانات يعرض البيانات بشكلين :عرض البيانات :ويعرض البيانات الحقيقية، وعرض المتغيرات :ويعرض معلومات عن المتغيرات ، ويشمل هذا تعريف المتغيرات وأسماء القيم ونوع البيانات مثلا حروف، أرقام، أسماء.
وكذلك المقياس المختبر(اسمي، رتبي، مقياس).

2 -DATA VIEW (شاشة البيانات)

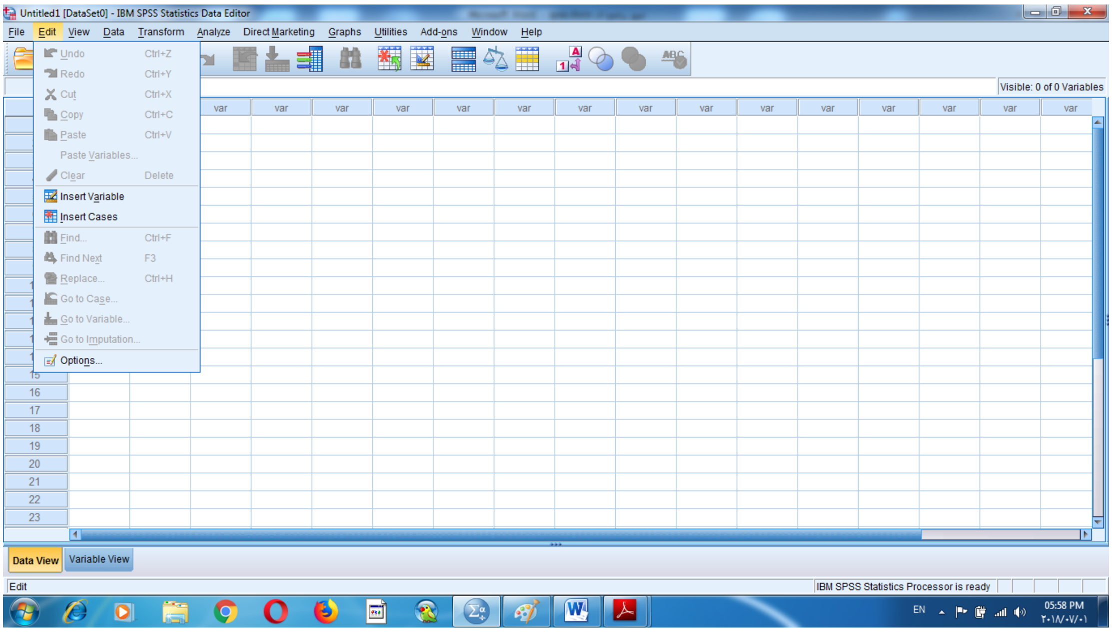
وتشمل الاوامر التالية:

ملف: file لفتح وحفظ الملفات وقراءة وطباعة البيانات بالاضافة الا ايعازات اخرى كما في الشكل:

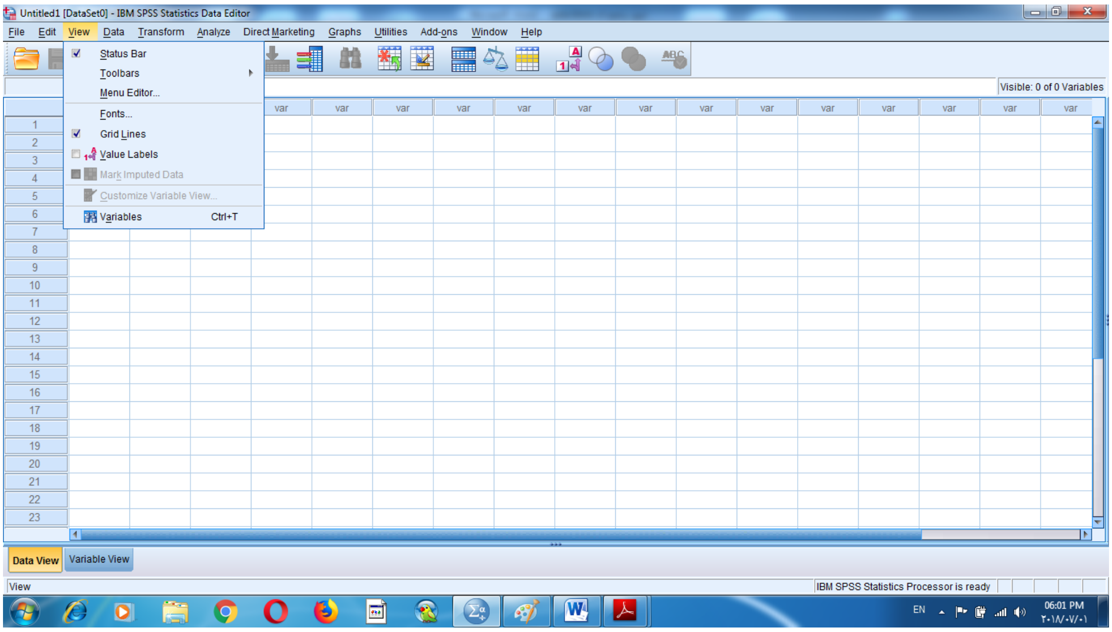


.

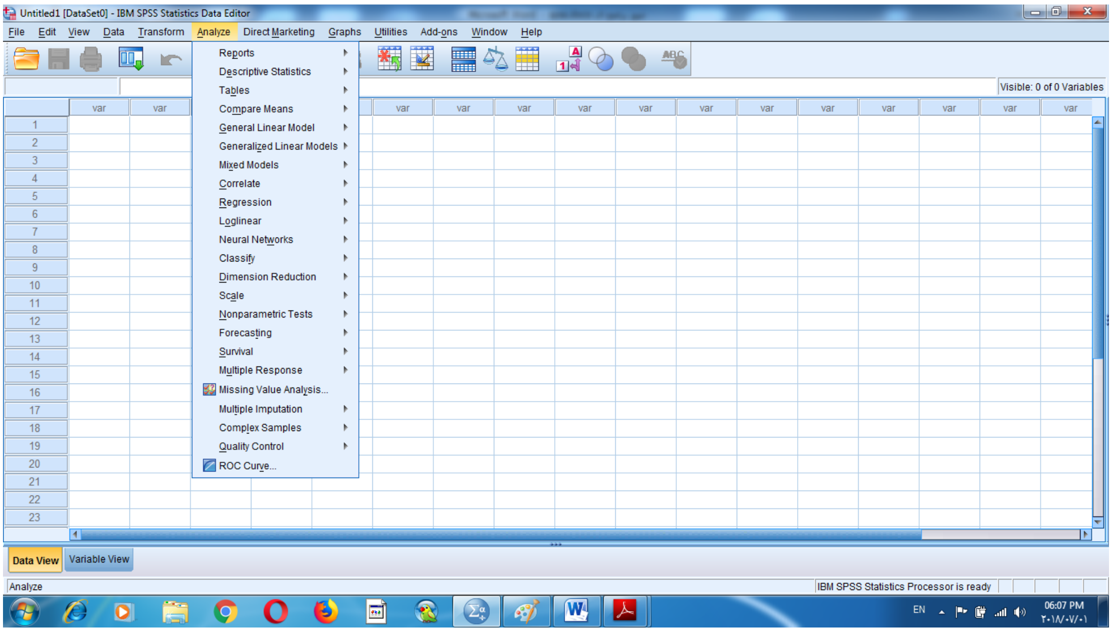
تحرير: edit يقص وينسخ ويلصق القيم وللحصول على قيم بيانات ولتغير الخيارات كما موضح في الشكل:



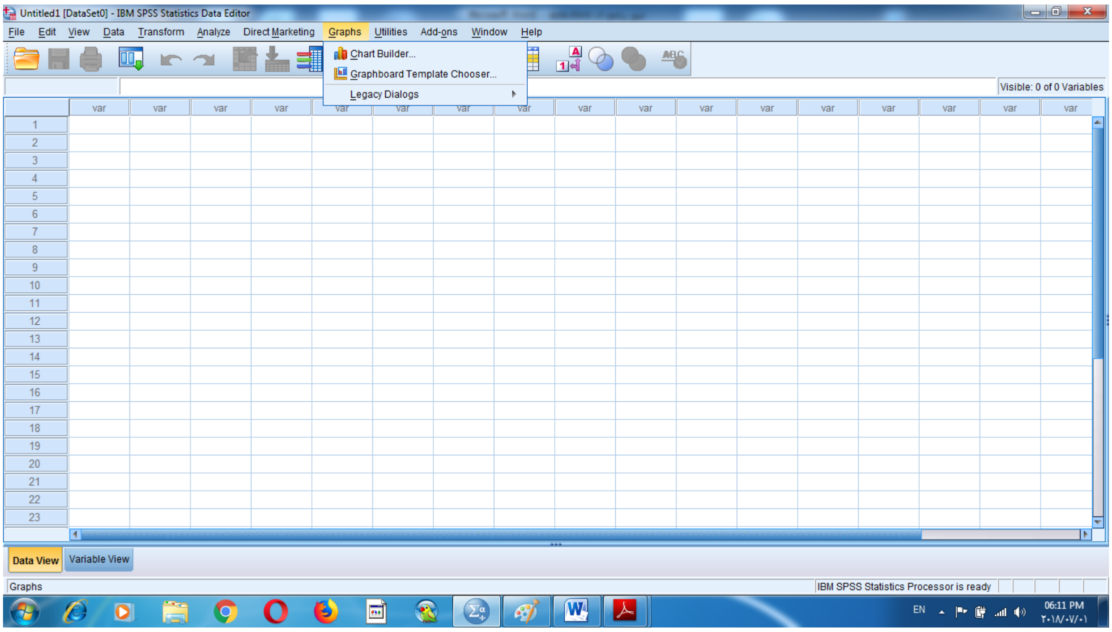
عرض: view للتحكم في شكل القيم وشرحها كما في الشكل:



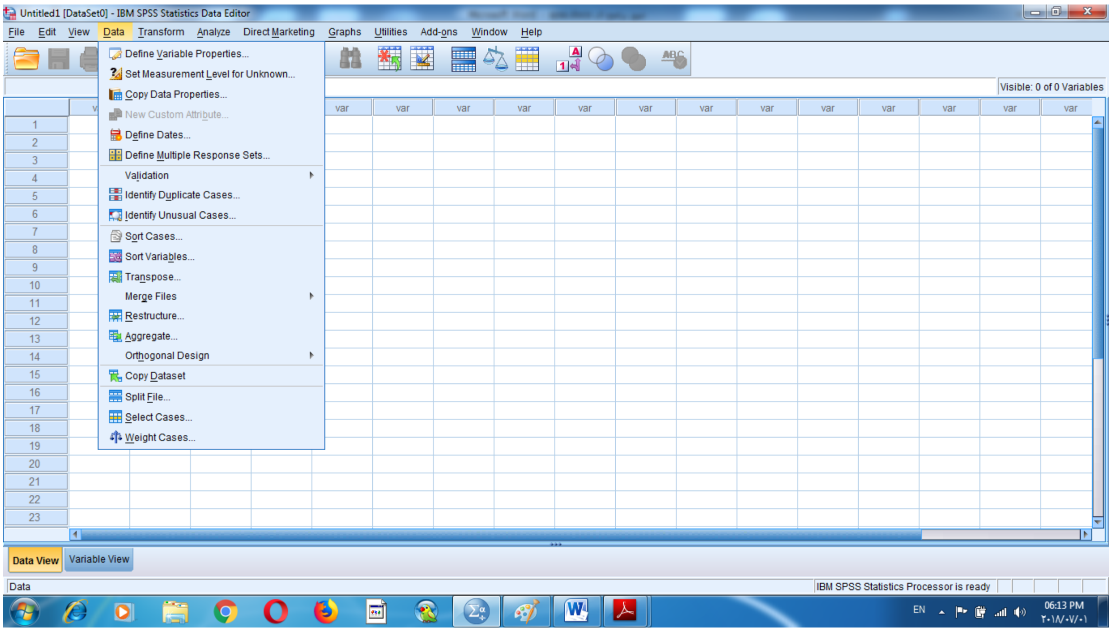
الإحصاء: analyze لاختيار مجموعة كبيرة ومتباينة من العمليات والاختبارات الإحصائية مثل اختبارات وتحليل التباين والاختبارات اللامعلمية. ويعتبر هذا الخيار بيت القصيد من الحزم كلها ويشمل أكبر كمية من الخيارات الضمنية كما في الشكل:



الاشكال : Graphs لإعداد رسوم بيانية بأنواعها : طولي ، دائري ، نقطي..والخ كما في الشكل:



البيانات : data لعمل تغير على ملف البيانات كما في الشكل:



3- VARIABLE VIEW(شاشة تعريف المتغيرات)

تحوي هذه الصفحة شرح ووصف لكل من المتغيرات الموجودة في محرر البيانات، و يجب ملاحظة أن الصفوف تحوي المتغيرات، بينما الأعمدة تبين وصف لهذه المتغيرات، ويشمل ذلك:

Name:اسم المتغير:

1- أسماء المتغيرات يجب أن تبدأ بحرف أما الباقي فيمكن أن يكون حروف، أو أرقام، أو نقطة، أو@ ، أو# ، أو- ، أو$

2-أسماء المتغيرات يجب أن لاتنتهي بنقطة.

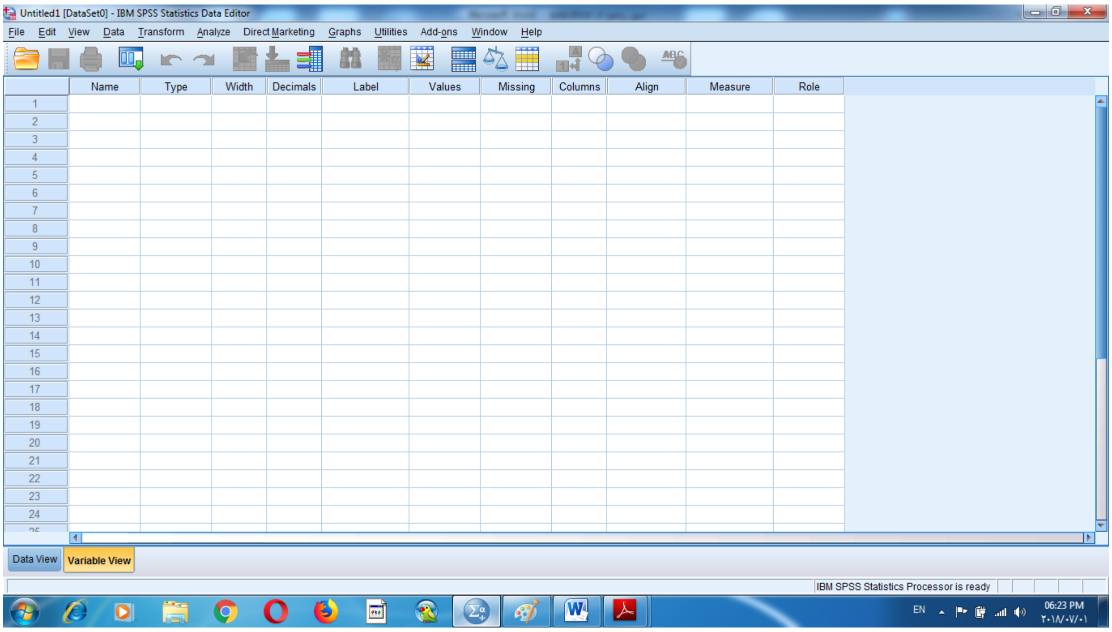
3-يجب أن لا يتعدى الاسم ثمان خانات.

4-يجب أن لايوجد ضمن الاسم فراغ أو أي من الاشارات الخاصة مثل (! ؛: ؟،*)

Type & Width نوع المتغير و عرض المتغير : في الأصل أن جميع البيانات رقمية،ولكن يمكن إدخال القيم على هيئة حروف أو نقط أو عمله أو خلافه، أما عرض المتغير فإنه يعتمد على نوعه.

تسمية المتغير:Labels عبارة عن وصف كامل للمتغير، يمكن أن يصل الى 256 خانة.

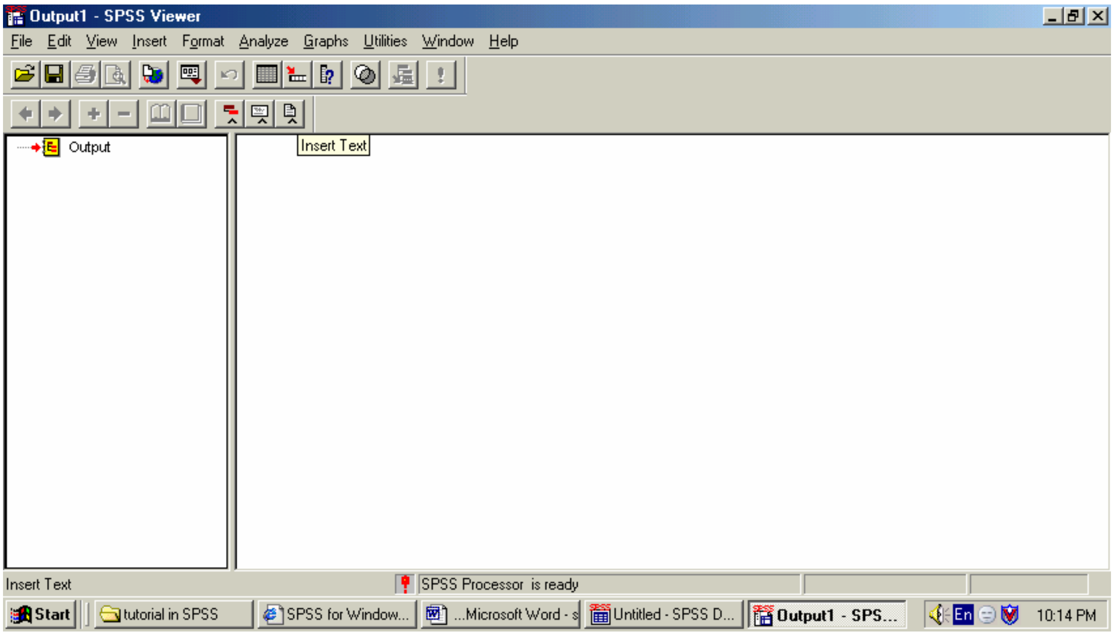
و الشكل الذي امامنا يوضح شاشة تعريف المتغيرات VARIABLE VIEW.



4- OUTPUT NAVIGATOR(لائحة التقارير والمخرجات)

ملف: File فتح و حفظ وطباعة المخرجات.
تحرير:Edit قطع ونسخ ولصق المخرجات ، ولتحريك المخرجات ولتغير إعدادات الخيارات
عرض: View للتحكم في مسطرة الأوامر.
إدراج: Insert لإدراج فاصل صفحة أو عنوان أو شكل أو نص أو أي هدف من برنامج اخر.
تشكيل: Format لتغير حدود مخرجات محددة.
إحصاء: Statistics لاختبار أي من العمليات أو الاختبارات الإحصائية.
أدوات: Utilities للحصول على معلومات عن متغير وللتحكم في المتغيرات التي تظهر في الصندوق الحواري.
نافذة: Window للتحول بين نوافذ SPSS أو لتصغير جميع نوافذ SPSS المفتوحة.
المساعدة: Help للحصول على الصفحة الأساسية للبرنامج .

و الشكل التالي يوضح لنا لائحة التقارير و المخرجات



•تسمية المتغيرات:

لتسمية المتغيرات يجب استبدال الاسم الافتراضي باسم مناسب) خاص بك مث: Sex, Subject, Attitude1, Attitude2، etc ولعمل ذلك:

1- انقر مرتين على اسم المتغير var00001 في أعلى العمود الأول أوضع المؤشر في أي خلية في الصف الأول ، ثم من data اختار define variable .

2- امسح الاسم الافتراضي var00001 واستبداله ب age .

3- انقر على ok.

ويمكن اي اسم اخر يناسب اسم المتغير الذي تمثله البيانات.

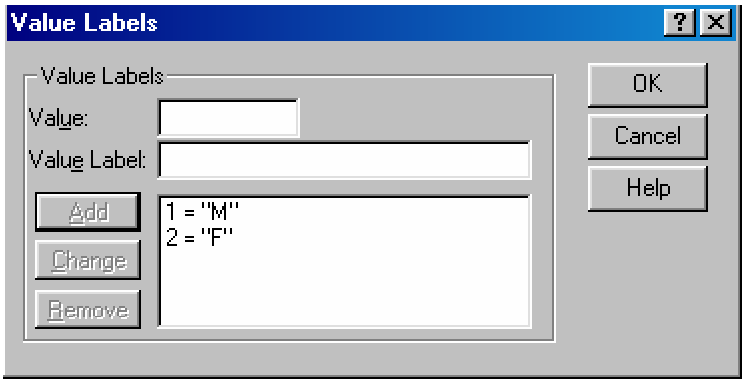
1 : age 34

	age	var	var	var	var	var	var
1	34.00						
2	22.00						
3	23.00						
4	24.00						
5	.						
6	23.00						

تعريف المتغيرات :

يمكن تحديد نوعية البيانات المضافة فالمتغيرات يمكن اضافتها كما هي ،اما البيانات والمتغيرات تحدد من قبل الباحث بطريقة البدائل (ذكر او انثى ،مريض او سليم) ويتم تعريف المتغير بالانتقال الى شاشة تعريف المتغيرات

1) إسم المتغير ، النوع ، حجم المتغير ، عدد النقاط العشرية .
2) (تحديد قيم المتغير) الترميز في خانة VALUES حيث تظهر الشاشة التالية:



3) ادخال في قيمة الرمز في خانة value واسم الرمز في خانة vallu labels والضغط على مفتاح add.

4) بعد اجراء الخطوات السابقة يتم اضافة المتغيرات في شاشة البيانات ولاظهار القيم الكتابية المرادفة بدل القيم الرقمية وذلك باجراء الخطوات التالية:

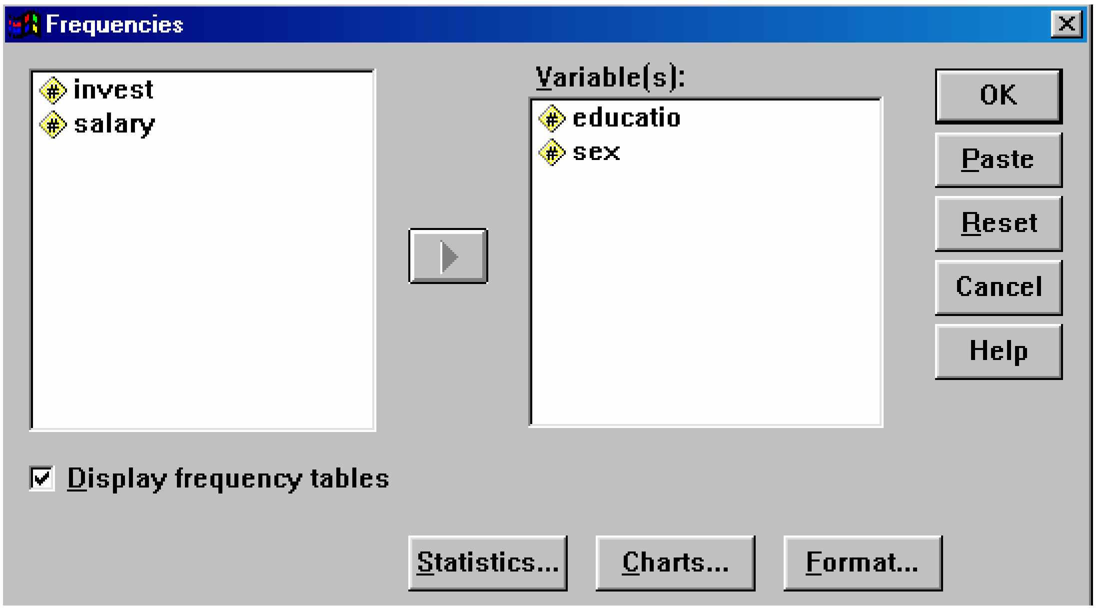
1- اختر الامر view من اللائحة الرئيسية.
2- اختر الامر الفرعي value label .

الاحصاء الوصفي والمدرج التكراري للبيانات

* المدرج التكراري (FREQUENCIES)

اختر من اللائحة الرئيسية ما يلي:

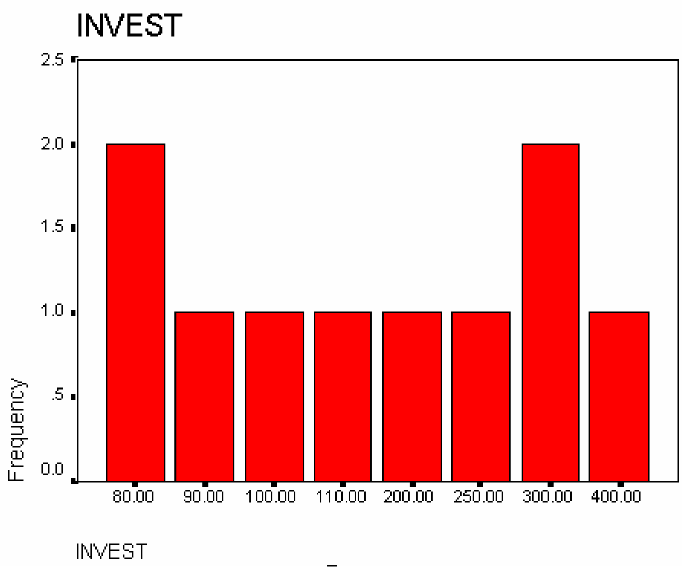
(1) ANALYSE
(2) اختر الأمر DESCRIPTIVE STATISTICS
(3) FREQUENCIES وتستخدم لعرض الجداول التكرارية للمشاهدات قيد الدراسة.



EDUCATIO

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	educated	5	50.0	50.0	50.0
	noneducated	5	50.0	50.0	100.0
	Total	10	100.0	100.0	

يمكن تحديد المطلوب إظهاره بتحديد الاختيارات بالضغط على مفتاح ..statistics والضغط على مفتاح الرسم البياني ..charts .



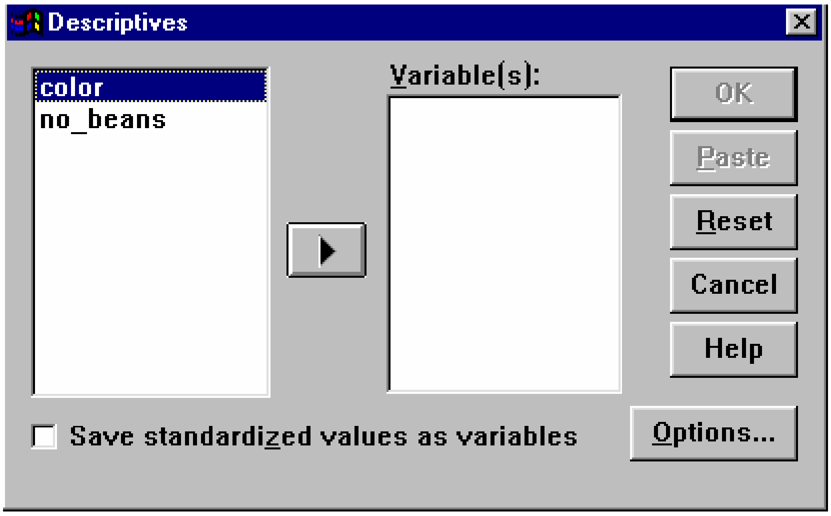
*** الإحصاء الوصفي DESCRIPTIVE ANALYSES :**

اختر من اللائحة الرئيسية ما يلي:

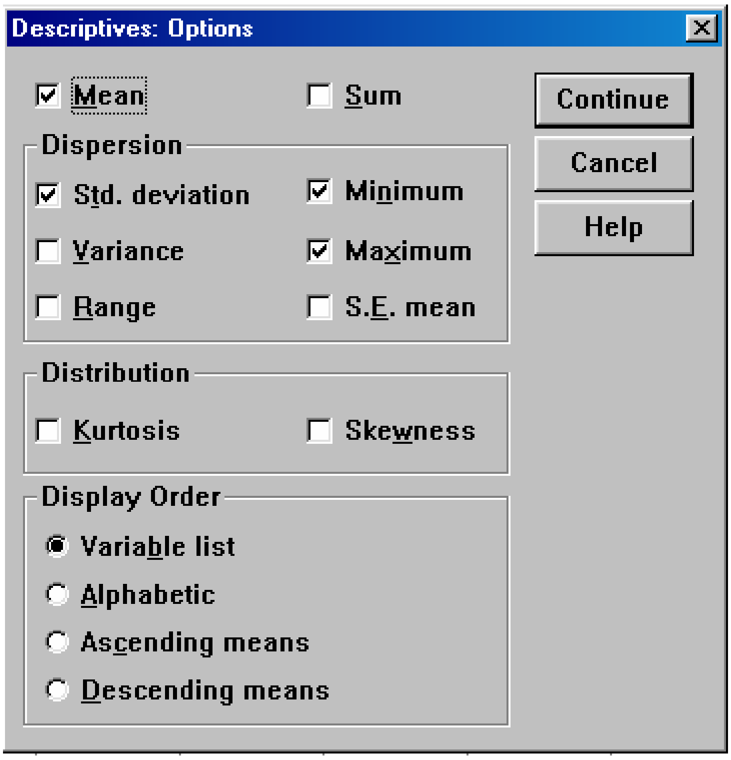
ANALYSE (1)

(2) اختر الأمر DESCRIPTIVE STATISTICS

DESCRIPTIVE (3) وتعني الاحصاء الوصفي



ولتحديد مخرجات الاحصاء الوصفي اختر Option من اللائحة الفرعية، ثم حدد ما هو المطلوب

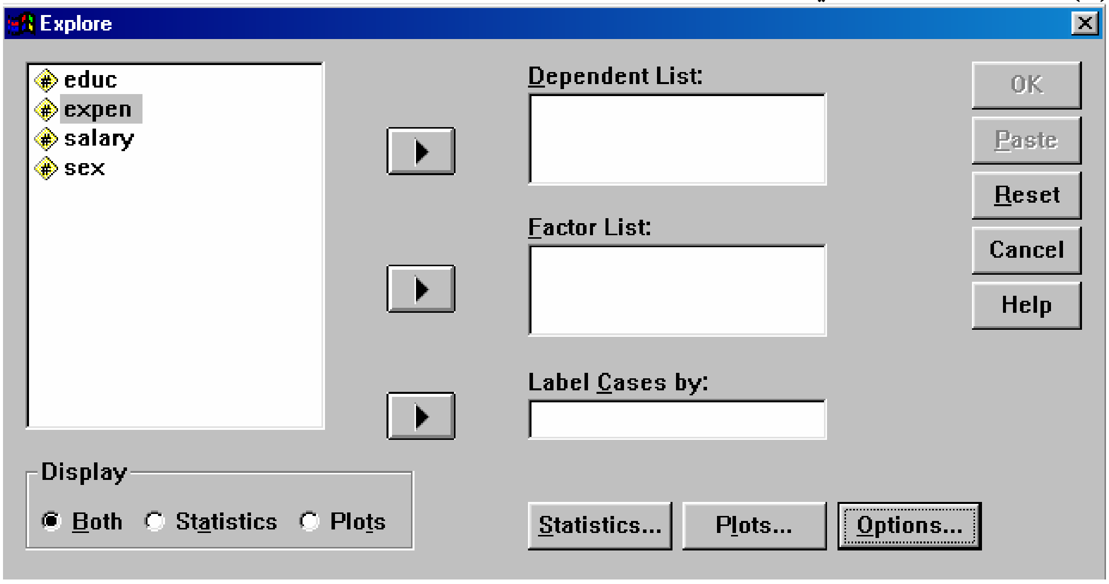


EXPLORE *

اختر من اللائحة الرئيسية ما يلي:

ANALYSE (1)

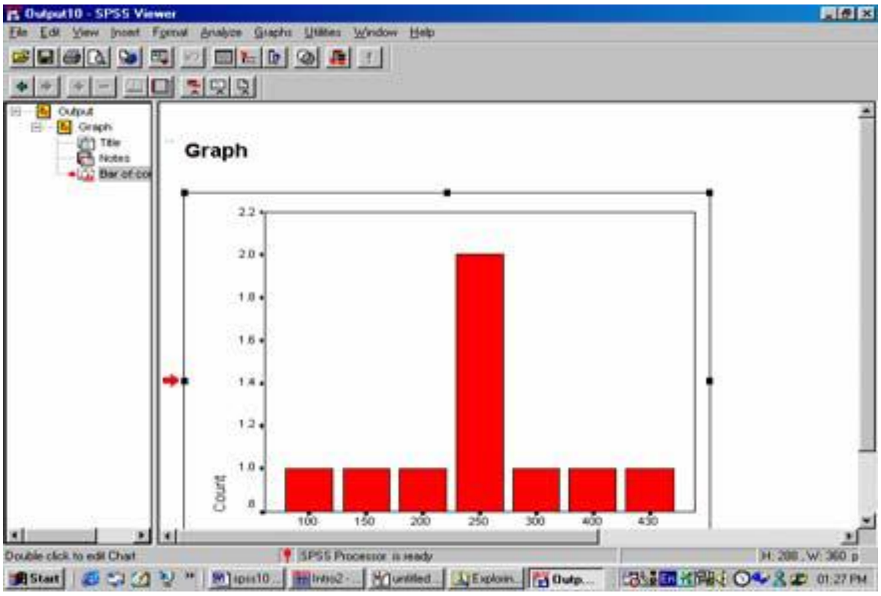
(2) اختر الأمر DESCRIPTIVE STATISTIC
(3) EXPLORE وتعني تبيان أو إظهار الخصائص الاحصائية للمتغير



			Statistic	Std. Error
INVEST	Mean		191.0000	36.6197
	95% Confidence Interval for Mean	Lower Bound	108.1606	
		Upper Bound	273.8394	
	5% Trimmed Mean		185.5556	
	Median		155.0000	
	Variance		13410.000	
	Std. Deviation		115.8016	
	Minimum		80.00	
	Maximum		400.00	
	Range		320.00	
	Interquartile Range		212.5000	
	Skewness		.615	.687
	Kurtosis		-1.036	1.334

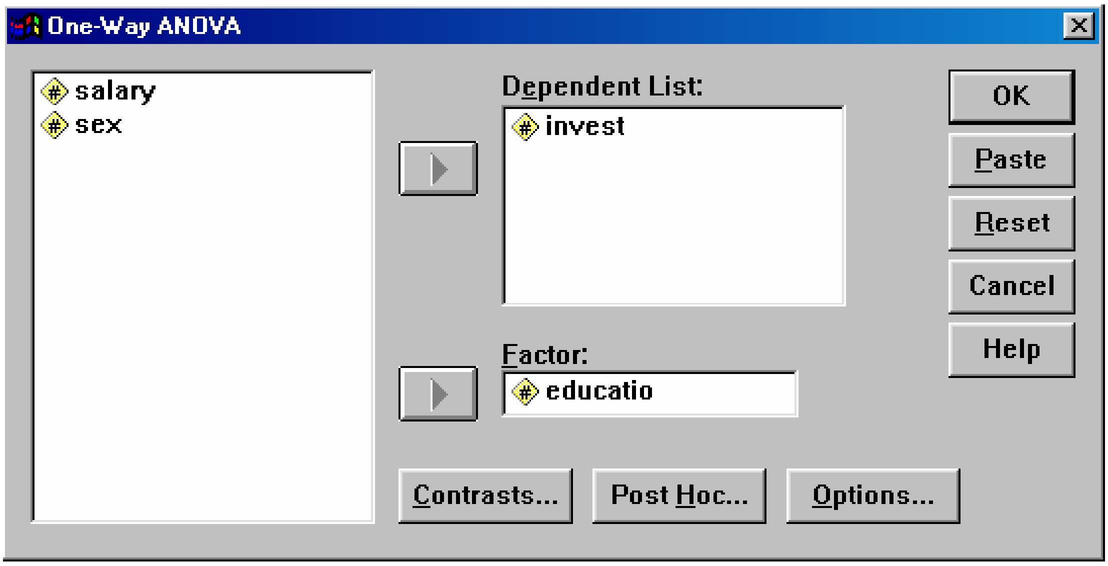
الرسم البياني للمتغيرات

يمكن تمثيل المتغيرات بالرسم البياني وذلك لتحليلها وتفسيرها ويتفرع من الأمر الرئيسي GRAPHS العديد من الأوامرالمتعددة باشكال الرسم البياني ولكل امر فرعي اختيارات حسب رغبة الباحث على سبيل المثال الاختيار line وتعني تمثيل البيانات بالرسم الخطي ، بعد تحديد الرسم البياني واختيار المتغيرات تظهر النتائج في نافذة خاصة
للرسم البياني حيث يمكن إضافة وتعديل العناوين بالضغط على الرسم البياني مرتين بالماوس.



تحليل التباين One Way ANOVA

تستخدم One Way ANOVA في تحليل التباين لتفسير ظاهرة اقتصادية معينة وذلك بتحديد متغير تابع يفسر من قبل متغير آخر (مثال موضح ادخار الفرد بين المتعلمين والغير المتعلمين من خلال الامر الرئيسي ANALYSE اختر COMPARE MEANS ثم امر التحليل One Way ANOVA ويتم تحديد المتغير التابع DEPENDENT والمتغيرFACTOR الذي يفسر الظاهرة الاقتصادية.



نتائج تحليل التباين

ANOVA

INVEST

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	98010.000	1	98010.000	34.571	.000
Within Groups	22680.000	8	2835.000		
Total	120690.0	9			

عند استخدام ANOVA لتفسير سلوك ظاهرة اقتص ادية معينة ، يجب تحديد فرضيات العدم والبديل (NULL & ALTERNATIVE HYPOTHESIS) وتستخدم احصائية (F) المحسوبة اكبر من القيمة الجدولية .

الارتباط CORRELATION

قياس العلاقة بين عدد من المتغيرات (R) هو معامل الارتباط وتتراوح هو معامل الارتباط وتتراوح قيمته بين الواحد الصحيح الموجب والواحد الصحيح السالب، وإذا اقتربت القيمة للواحد فهذا يعني أن العلاقة بين المتغيرات تحت الدرس قوية جدا "، والعكس هو إذا اقتربت القيمة من الصفر وهذا يعني أن العلاقة ضعيفة جدا."من خلال الأمرANALYSE اختر الأمر CORRELATION وتظهر بعد ذلك شاشة لتحديد المتغيرات تحت الدرسه.

Correlations

		INVEST	SALARY
INVEST	Pearson Correlation	1.000	.895**
	Sig. (2-tailed)	.	.000
	N	10	10
SALARY	Pearson Correlation	.895**	1.000
	Sig. (2-tailed)	.000	.
	N	10	10

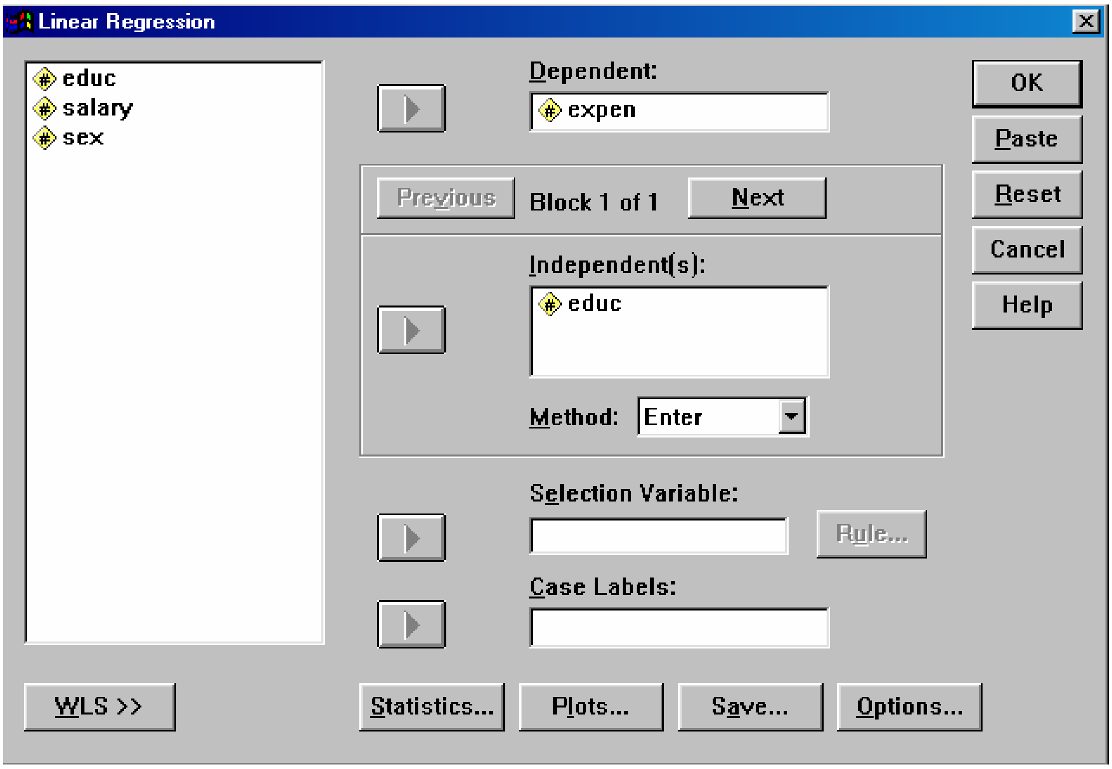
**. Correlation is significant at the 0.01 level

الانحدار الخطي REGRESSION

لايجاد العلاقة بين متغيرين على هيئة علاقة دالة خطية حيث ان y المتغير التابع و x المتغير المستقل(المفسر)والذي يفسر التغير في المتغير التابع، وتكون بالشكل التالي:

$$y = a + b\,x1$$

وذلك باختيار الأمر الفرعي **REGRESSION** من اللائحة الرئيسية ANALYSE ثم تحديد المتغيرات في النافذه الخاصة بذلك آما هو موضح أدناه:



يتم تحديد المتغيرات المستقلة في خانة (INDEPENDENTS) والمتغيرات التابعة في خانة (DEPENDENT) وتظهر النتائج بتحديد معامل الارتباط واحصائية (F) مع معاملات الانحدار كما هو موضح ادناه:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.809[a]	.655	.598	52.52

a. Predictors: (Constant), EDUC

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	31450.909	1	31450.909	11.403	.015[a]
	Residual	16549.091	6	2758.182		
	Total	48000.000	7			

a. Predictors: (Constant), EDUC

b. Dependent Variable: EXPEN

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	78.182	41.895		1.866	.111
	EDUC	67.636	20.030	.809	3.377	.015

a. Dependent Variable: EXPEN

في الشكل السابق تبين أن معادلة الانحدار الخطي هي:
Y=78.182+67.636x ومعامل الارتباط هو 0.809 قريب من الواحد الصحيح أي وجود علاقة قوية للمتغيرات (INVEST & EDUCATION).