University of Jordan
Faculty of Medicine
Batch 2013-2019

Medical Committee
The University of Jordan

Epidemiology & Biostatistics

○ Slides ○ Sheet ● Handout ○ other.....

Number#1

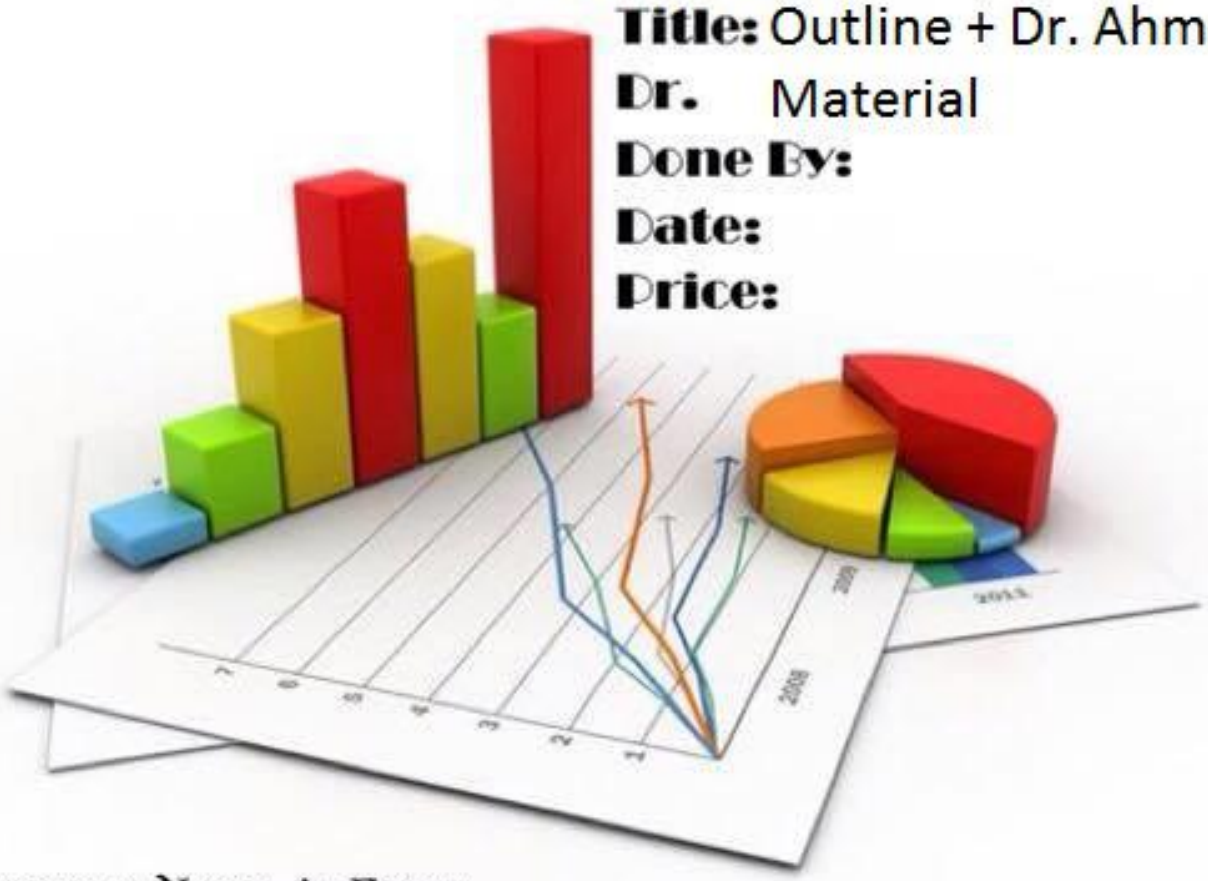
Title: Outline + Dr. Ahmad Material

Dr.

Done By:

Date:

Price:



Designed by Nadeen Al-Freihat

Biostatistics

Introduction

Some Basic Concepts: Statistics, Biostatistics, Variable, Quantitative Variables, Qualitative Variables, Random Variable, Discrete Random Variable vs Continuous Random Variable, Population, Sample.

I. **Measures of Central Tendency**: the mean, the median, the mode

II. **Measures of dispersion**: the range, the variance, standard deviation (SD), coefficient of variation (CV).

Probability

Some basic probability concepts

Elementary properties of probability
Factorials: permutations, combinations
The union of two conjoint sets
The union of two disjoints sets
Venn diagram
Marginal probabilities, Joint probabilities, Additive Law, Multiplicative Law, Independent events, Conditional probability, Bayes' Theorem.

Probability Distributions

The probability distribution of discrete variables

Cumulative probability distribution

The Binomial distribution
Bernoulli Trials

The Poisson distribution

The Normal Distribution

Approximately Normally Distributed Data

Continuous Probability Distribution

The t Distribution: Student's t tests

Student's t test for a single small sample
Student's t test for independent samples
Student's t test for Paired samples

Contingency Table and Degrees of Freedom

The Chi-Square Distribution: The Chi-Square tests
Test of goodness of fit
Test of Independence
Test of Homogeneity

Hypothesis Testing
Nine Steps of Hypothesis testing: Data, Assumptions, Hypothesis (the null and the alternative hypothesis), Test Statistic (normal Z, Normal t, Chi-Square Test), Distribution of the test statistic, Decision Rule (acceptance region, and rejection region)
Types of errors in hypothesis testing (Alpha error, Beta error), Power of the test (1-Beta), Level of significance (Alpha), P- value
Testing the difference between two population means
Sampling from normally distributed populations, population variances known (M=0, G=1)
Sampling from normally distributed populations, population variances unknown
Sampling from a population that is not normally distributed (Central Limit Theorem)

Correlation analysis: correlation definition, correlation coefficients
(Pearson correlation coefficient, Spearman correlation coefficient)

Regression Analysis: regression definition, regression coefficients, regression model, linear regression, the best fit regression line, R-Square, Dependent variable and independent variable.

N 1

Introduction

Some Basic Concepts

Statistics: is a scientific field concerned with
① the collection, organization, and summarization of data and,
② the drawing of inferences about a body of data when only a part of the data are observed.

Biostatistics: The application of statistics to biomedical problems.
When the data being analyzed are derived from the biological sciences and medicine, we use the term biostatistics to distinguish this particular application of statistical tools and concepts.

Variable: If, as we observe a characteristic, we find that it takes on different values in different persons, places, or things, we label the characteristic a variable.

Examples: Heights of adult males.
Weights of preschool children.
Ages of patients seen in a dental clinic.

Quantitative variables: one that can be measured in the usual sense. Measurements made on quantitative variables convey the concept of amount. Examples: heights, weights, ages, etc.

Qualitative variables: measurements made on qualitative variables convey the concept of attribute. e.g: ethnic group, person, place (categorizing).

om Variable: whenever we determine the
nt, weight, or age of an individual, the result
equently referred to as a value of the
ective variable. When the values obtained
e as a result of chance factors, the
able is called a random variable.
ues resulting from measurement procedures
often referred to as observations or,
nply, measurements.

(ete Random Variable vs continuous Random Variable

ete Random Variable: is characterized by gaps
rerruptions in the values that it can assume.
the # of daily admissions to a general hospital
(0, 1, 2, 3 ..... It cannot be 1.5, 2.9, 3.3 ... etc).
the # of decayed, missing, or filled teeth
per child in an elementary school.
inuous Random Variable: doesnot possess
gaps or interruptions characteristic of a
rete random variable. It can assume any
e within a specified interval of values
umed by the variable.
: measurements made on individuals such
as height, weight, skull circumference.
opulation vs a Sample

Population: we define a population of entities as the largest collection of entities for which we have an interest at a particular time.

a population of values: The largest collection of values of a random variable for which we have an interest at a particular time.

e.g: - Weights of all children enrolled in a certain County elementary school system.

- Weights of first grade students enrolled in the school system.

Sample: may be defined as "a part of a population".

e.g: weights of only a fraction of children in the above mentioned example.

I "Measures of Central Tendency"

Descriptive measures may be computed from the data of a sample or the data of a population. To distinguish between them we have the following definitions:

a) A descriptive measure computed from the data of a sample is called a Statistic.

b) A descriptive measure computed from the data of a population is called a parameter.

Descriptive measures are divided into:

I) Measures of central Tendency.

II) Measures of Dispersion.

I "Measures of Central Tendency"

The three most commonly used measures of central tendency are:

A) the mean

B) the median

C) the mode

A) Arithmetic mean (mean). The mean is obtained by adding all the values in a population or a sample and dividing by the number of values that are added.

e.g: Table 1.4.1 "Ages (in years) of patients admitted to a chronic disease hospital during a certain month".

TABLE 1.4.1

Ages (in Years) of Patients Admitted to a Chronic Disease Hospital During a Certain Month

Number	*Age*	*Number*	*Age*	*Number*	*Age*	*Number*	*Age*
1	10	26	48	51	63	76	53
2	22	27	39	52	53	77	33
3	24	28	6	53	88	78	3
4	42	29	72	54	48	79	85
5	37	30	14	55	52	80	8
6	77	31	36	56	87	81	51
7	89	32	69	57	71	82	60
8	85	33	40	58	51	83	58
9	28	34	61	59	52	84	9
10	63	35	12	60	33	85	14
11	9	36	21	61	46	86	74
12	10	37	54	62	33	87	24
13	7	38	53	63	85	88	87
14	51	39	58	64	22	89	7
15	2	40	32	65	5	90	81
16	1	41	27	66	87	91	30
17	52	42	33	67	28	92	76
18	7	43	1	68	2	93	7
19	48	44	25	69	85	94	6
20	54	45	22	70	61	95	27
21	32	46	6	71	16	96	18
22	29	47	81	72	42	97	17
23	2	48	11	73	69	98	53
24	15	49	56	74	7	99	70
25	46	50	5	75	10	100	49

$$\text{Mean age} = \frac{10+22+\cdots+70+49}{100} = \frac{3920}{100} = 39.20 \text{ (years)}$$

Calculation

* The general formula for a population mean (μ) is determined as:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

The symbol $\sum_{i=1}^{N}$ means Summation of all values of the variable from the first to the last.

$\Sigma \rightarrow$ called summation sign.

* The general formula for the sample mean ($\bar{X}$) is determined as:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

where Σ = "the sum of"

X_i = each of the values in the series (sample)

n = the number of the values in the series (sample).

* e.g.:

The ages (yrs) of seven children seen in an (ER) after a house fire are: 1, 1, 1, 2, 4, 6, 6.

$$\bar{X} = \frac{1+1+1+2+4+6+6}{7} = \frac{21}{7} = 3 \text{ years.}$$

The arithmetic mean has the following properties:

a) Uniqueness - for a given set of data there is one and only one mean.

b) Simplicity. Easy to compute and easy to understand.

c) Extreme values have an influence on the mean and in some cases can so distort it. It becomes undesirable as a measure of central tendency.

B. The median. The value which divides the set into two equal parts such that the # of values equal to or greater than the median is equal to the # of values equal to or less than the median.

Calculation

a) In a series with an odd number of values, the values in the series are arranged from lowest to highest, and the value that divides the series in half is the median.

Steps:

* arrange the values in order of magnitude.
* $\boxed{\frac{n+1}{2}}$, e.g $\frac{7+1}{2} = \frac{8}{2} = 4 \Rightarrow$ The value #4 is the median

e.g The ages (yrs) of seven children seen in an (ER) after a house fire are:

2, 4, 1, 1, 6, 1, 6

What is the median age?

Steps:

* arrange the values in order of magnitude.

  1, 1, 1, 2, 4, 6, 6
* odd # $\rightarrow$ the median order $= \boxed{\frac{n+1}{2}} = \frac{7+1}{2} = \frac{8}{2}$

  $\Rightarrow = 4$, the observation #4

  $\Rightarrow$ The median age is 2 years.

b) In a series with an even number of values, the two values that divide the series in half are determined, and the arithmetic mean of these two values is the median.

Steps:

* arrange the values in order of magnitude.
* $\boxed{\frac{n}{2}}$, e.g $\frac{100}{2} = 50 \Rightarrow$ The median will be the mean of observation #50 and 51.

$$\text{Median} = \frac{\#50 + \#51}{2}$$

e.g. Table 1.42

$$\text{Median age of patients} = \frac{\#50 + \#51}{2} = \frac{36+37}{2} = 36.5 \text{ years}$$

The median has the following properties:

a) Uniqueness. there is only one median for a given set of data.

b) Simplicity. easy to calculate.

c) It is not as drastically affected by extreme values as is the mean.

TABLE 1.4.2

An Ordered Array of the Ages of Patients Admitted to a Chronic Disease Hospital During a Certain Month

1	14	37	58
1	15	39	60
2	16	40	61
2	17	42	61
2	18	42	63
3	21	46	63
5	22	46	69
5	22	48	69
6	22	48	70
6	24	48	71
6	24	49	72
7	25	51	74
7	27	51	76
7	27	51	77
7	28	52	81
7	28	52	81
8	29	52	85
9	30	53	85
9	32	53	85
10	32	53	85
10	33	53	87
10	33	54	87
11	33	54	87
12	33	56	88
14	36	58	89

c) The mode. The value of most frequent occurrence.

* the mode of a set of values is that value which occurs most frequently.

Examples:

* The sample consisting of the following values:
  15, 20, 25, 16, 25, 17, 25.
  The mode is 25

* If all the values are different ⇒ there is no mode.
  e.g: The following sample:
  10, 21, 33, 53, 54 ⇒ The sample has no mode since all the values are different.

* A set of values may have more than one mode.
  e.g: a laboratory has 10 employees whose ages are: 20, 21, 20, 20, 34, 22, 24, 27, 27, 27.
  The data have two modes 20 and 27.

II. Measures of Dispersion.

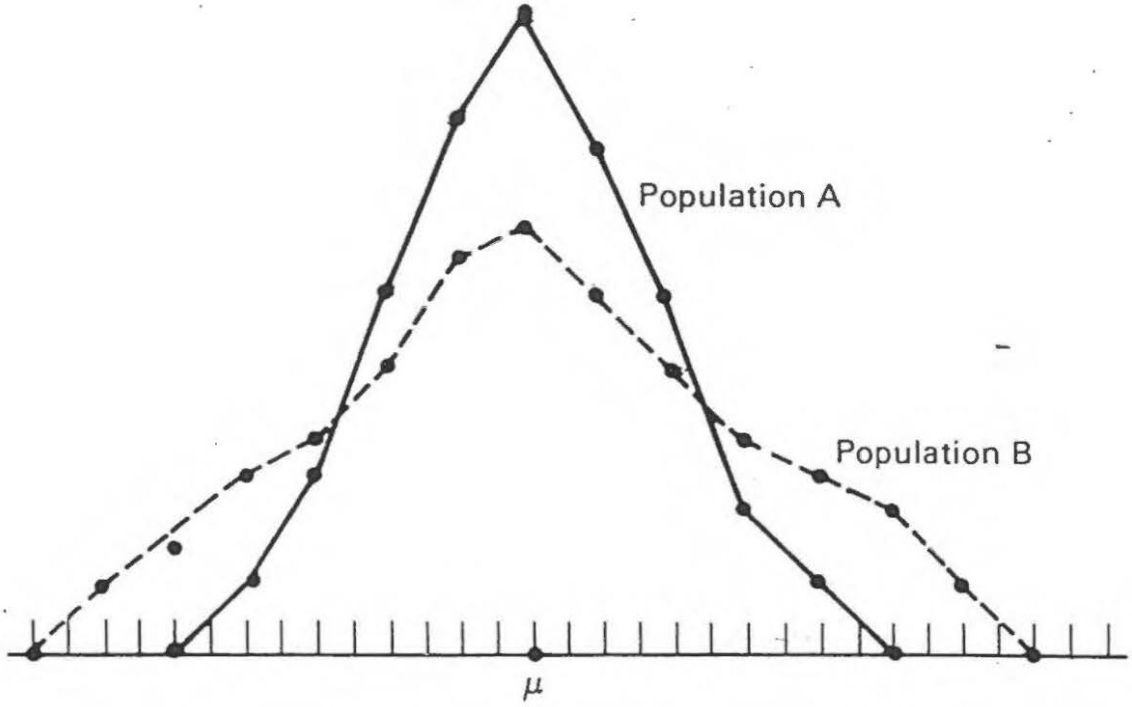
* characteristics that are used to describe the spread, variation, and scatter of a series of values.

* Dispersion = variation = Spread = Scatter.

e.g: Figure 1.7.1 shows the frequency polygons for two populations that have equal means but different amounts of variability. Population B, which is more variable than population A, is more spread out.

a) The Range: is the difference between the smallest and largest value in a set of observations.

* Calculation: The range is calculated by subtracting the lowest value in the series from the highest value.



(Figure 1.7.1)

"Two Frequency Distributions with equal Means but Different Amounts of Dispersion".

$$\boxed{R = X_L - X_S}$$

where: R = The Range
X_L = the largest value
X_S = the smallest value

e.g: The ages of five medical workers are:
22, 24, 26, 32, 27 years
What is the Range of ages?

$R = X_L - X_S = 32 - 22 = 10$ years.
The Range is 10 years.

Applications
The Range is used to measure data spread.

Features of the Range (properties)
- It is a poor measure of dispersion
- Simple to calculate.

b) The Variance: is the sum of the squared deviations from the mean divided by the number of values in the series minus 1.

Applications - The principal use of the variance is in calculating the standard deviation.

Calculation:

Definitional formula for the Sample Variance:

$$\boxed{S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$$

Computational formula for the Sample Variance:

$$\boxed{S^2 = \frac{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}}$$

Where: $\sum X_i^2$ = the sum of squares of all X_i.

$(\sum X_i)^2$ = the square of the sum of all X_i.

n = # of values (observations).

$(n-1)$ = degrees of freedom.

Definitional formula for the population Variance:

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

Computational formula for the population Variance:

$$\boxed{\sigma^2 = \frac{N \sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}{N \cdot N}}$$

e.g: The ages of five Individuals are:
17, 18, 18, 21, 26 years

Calculate the Age Variance for these five Individuals (Sample Variance)?

Using Table 3-1

$$S^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

$$S^2 = \frac{5(2054) - (100)^2}{5 \times 4}$$

$$S^2 = \frac{10270 - 10000}{20} = \frac{270}{20} = 13.5 \text{ years}$$

The sample variance $(S^2) = 13.5$ years^2

C) Standard deviation: is the positive square root of the variance.

Applications and characteristics

* The most useful measure of dispersion.
* when (SD) or (S) is small, the sample mean is close to each individual value.
* (SD) or (S) decreases when the sample size (n) increases.

Table 3-1. Calculating the Age Variance for Five Individuals

Case Number	Age (years)	Mean Age (years)	Deviation from the Mean	Deviation from the Mean Squared	Age Squared
1	17	20	−3	9	289
2	18	20	−2	4	324
3	18	20	−2	4	324
4	21	20	+1	1	441
5	26	20	+6	36	676
Total	100	. . .	0	54	2054

Calculation: The standard deviation (SD) or (S) is determined as:

$$\boxed{S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}}$$

e.g: from the previous example (Table 3.1)

$S^2 = 13.5$ years^2

$S = \sqrt{S^2} = \sqrt{13.5} = 3.67 = 3.7$ years

D) Coefficient of Variation (CV): is the ratio of the standard deviation of a series to the arithmetic mean of the series. The (CV) is unitless and is expressed as a percentage.

Applications and characteristics

* CV is used to compare the relative variation or spread, of the distributions of different series, samples, or populations.

Calculation: The CV is calculated as:

$$\boxed{CV = \frac{S}{\bar{X}} \times 100 \quad \text{or} \quad CV = \frac{SD}{\bar{X}} \times 100}$$