

Tishk International University
Faculty of Pharmacy

Introduction to Biostatistics
& Course Book

Dr. Abubakir M. Saleh
Assist.Professor of Community Medicine

Lecture Outline

- Definitions: Statistics & Biostatistics
- Data: sources and types
- Aim and objectives of the course
- Course Syllabus
- Lecture Schedules
- Course outline
- Examinations

Statistics

Statistics is the art and science of data. It deals with:

- Collecting data
- Organizing data
- Describing data
- Summarizing and presenting data
- Analyzing data
- Interpreting results
- Reaching decisions or discovering new knowledge

Biostatistics

Biostatistics is application of statistical methods to a wide range of topics in biology including medicine and health sciences

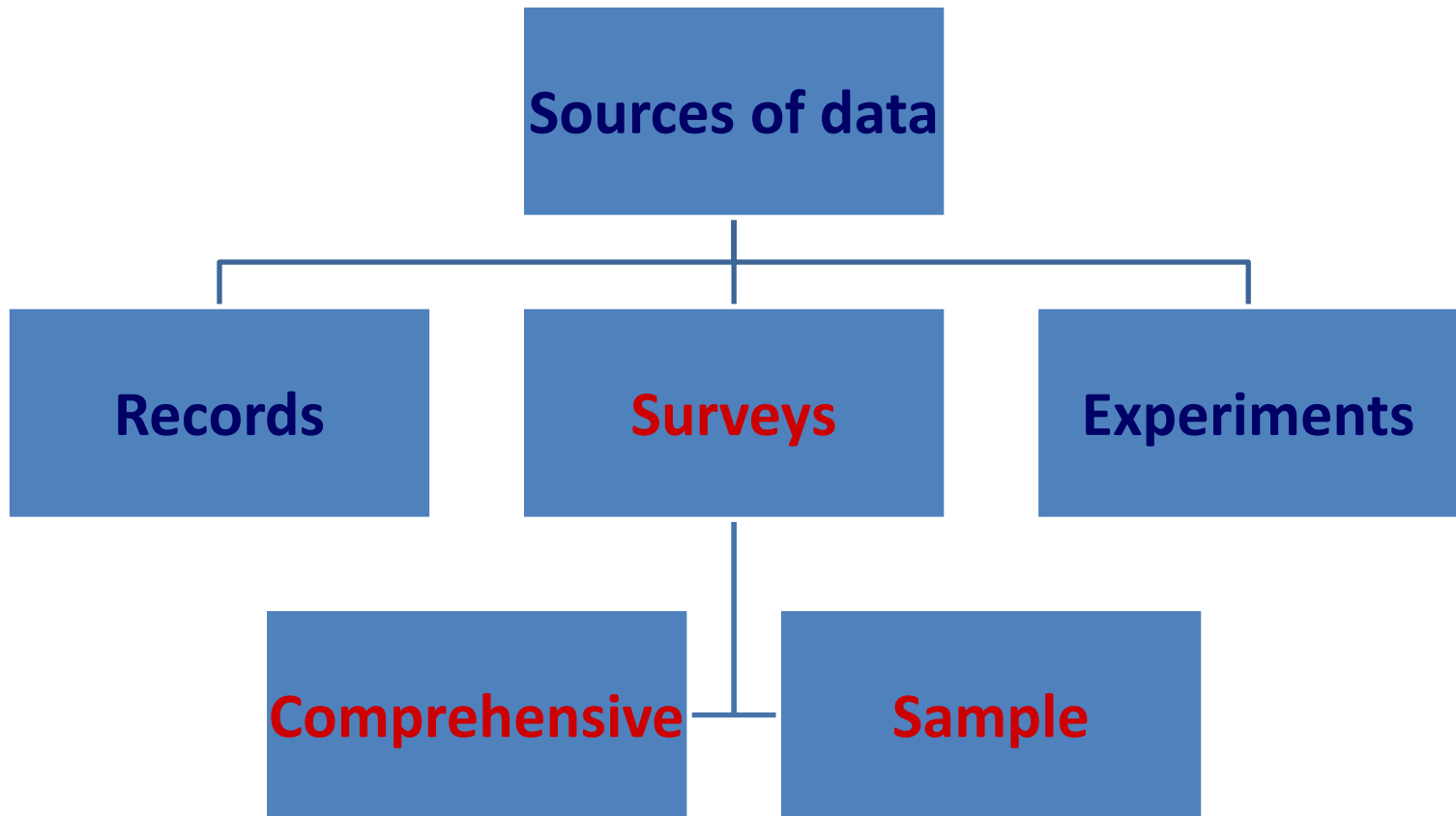
Data

- **Data** is information that has been translated into a form that is efficient for movement or processing

E.g. of data:

Patient code	Age	Sex	Height	Weight	Marital status	Education level
1	24	Male	170	80	Married	Primary
2	32	Female	155	60	Single	Secondary
3	41	Male	165	75	Married	College
4	26	Female	160	55	Single	Primary
5	37	Male	172	78	Single	College
6	45	Male	168	87	Single	Primary
7	38	Female	164	68	Married	College
8	56	Male	175	64	Married	Secondary
9	21	Female	159	57	Single	College

Sources of data



Types of Data

```
graph TD; A[Types of Data] --> B[Quantitative data]; A --> C[Qualitative data]; B --> D[Quantitative continuous]; B --> E[Quantitative discrete]; C --> F[Qualitative nominal]; C --> G[Qualitative ordinal];
```

Quantitative data

Quantitative
continuous

Quantitative
discrete

Qualitative data

Qualitative
nominal

Qualitative
ordinal

Biostatistics

Basic tasks of statistics

- To describe or present data
 - Tabular, graphical, numerical
- To draw inference concerning the underlying population
 - Statistical investigations carried on samples
 - Conclusions about population
 - Probability theory and inferential statistics

Presentation of data

- To sort and classify data into groups or classification.
- Objective :
 - to make data simple, concise, meaningful, interesting & helpful for further analysis.
- 2 main methods:
 - i. Tabulations
 - ii. Charts and diagrams

Table 1: Distribution of 50 patients at the hospital according to their age

Age (years)	Frequency
20-29	12
30-39	18
40-49	5
50+	15
Total	50

Table 1: Distribution of 50 patients at the hospital according to their age

Age (years)	Frequency	%
20-29	12	24
30-39	18	36
40-49	5	10
50+	15	30
Total	50	100

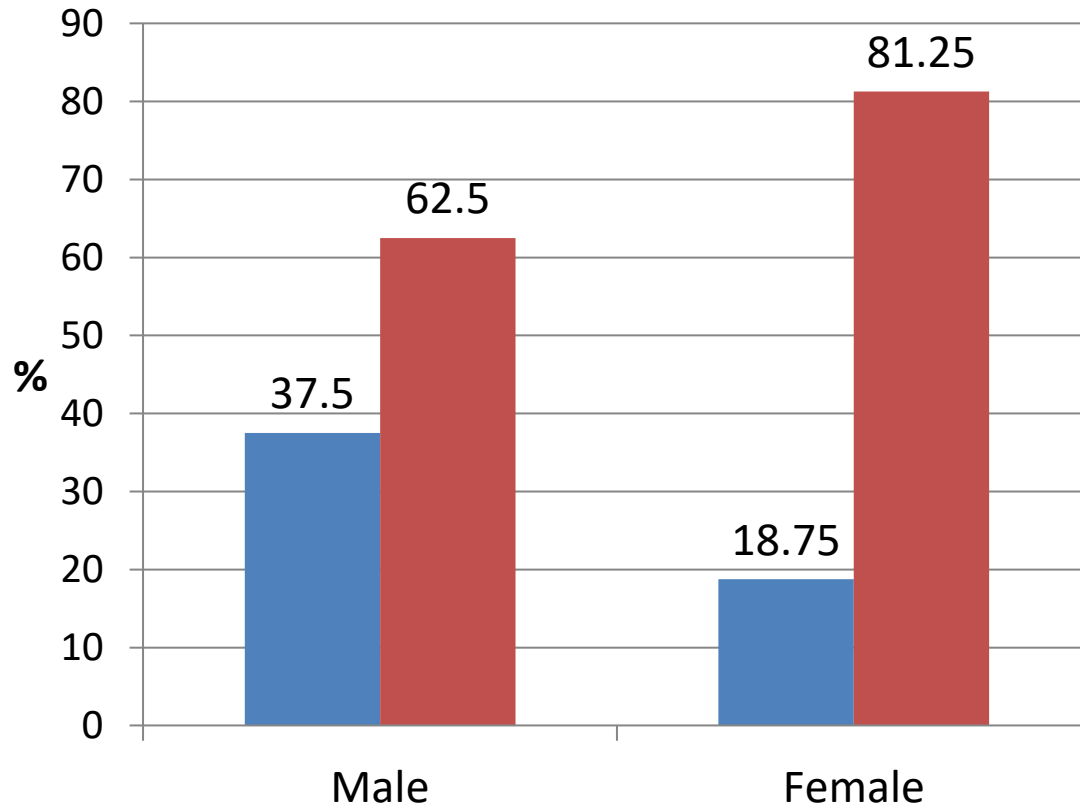
Table 2: Distribution of the sample according to smoking status and developing lung cancer

Smoking	Lung cancer		Total
	Yes	No	
Smoker	15	8	23
Non smoker	5	32	37
Total	20	40	60

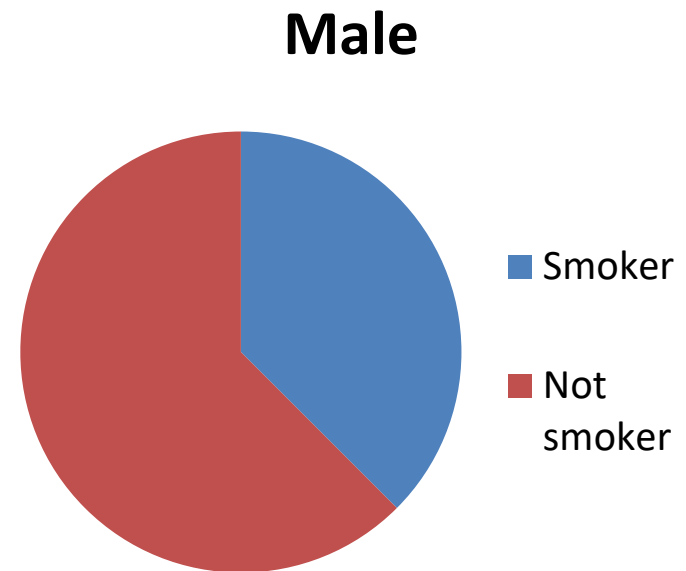
Table 2: Distribution of the sample according to smoking status and developing lung cancer

Smoking	Lung cancer				Total	
	Yes		No			
	No.	%	No.	%	No.	%
Smoker	15	65%	8	35%	23	100
Non smoker	5	14%	32	86%	37	100

Charts and diagrams



■ Smoker
■ Not smoker



Aim of the course

To introduce basic descriptive and inferential statistical concepts and provide an understanding of the relevance and importance of statistics to health research.

Biostatistics

- The course will particularly emphasize the practical applications of basic statistical concepts
- The mathematical formulae will be kept to a minimum.
- The theory and the methods of analysis will be extensively illustrated with examples relating to a wide variety of relevant fields.

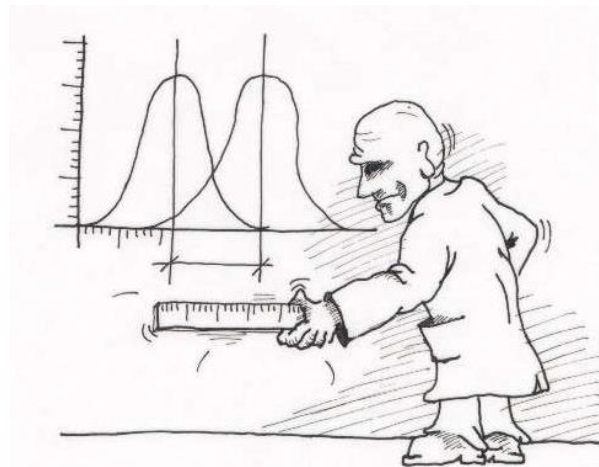
Course Objectives:

By the end of this course the students should be able to:

- Describe the role of statistical methods in health research;
- Present results effectively by making appropriate displays, summaries, and tables of data;
- Appreciate the problem of sampling variation and the role of statistical methods in quantifying it
- Select an appropriate statistical method for the analysis of simple data sets;

Objectives – cont.

- Correctly interpret and critically appraise the results of statistical analyses reported in the health literature;
- Present and discuss the findings from statistical analyses in a clear, concise and logical manner



Syllabus

- Course length: 15 weeks (including mid-course examination)
- Course language: English
- Assessment (grading):
 - Mid course exam : 40%
 - Attendance & Class activity:10%
 - Quizz: 10%
 - nd of course exam: 40%

- **Course Reading List:**

Comprehensive notes are provided

Extra reading

- [Essential Medical Statistics](#), by Betty Kirkwood & Jonathan Sterne
(Published by Blackwell)
- [Statistics Without Tears](#), a Primer for Non-mathematicians, by Derek Rowntree
(Published by Penguin)
- All the statistical tables you need for this course are provided.

Lecture Schedule

Week No.	Topic
1	Introduction
2	Population and sample
3	Types of variables
	Data presentation
4	Descriptive measures
5	Confidence intervals
6	Mid-course examination
7	Hypothesis testing
9	Comparison of two means (t-test)
9	Comparison of two proportions(Chi-square)
10	Contingency tables
11	Sample size calculation

Lecture Schedule

Week No.	Topic
12	Using computer software in data analysis (SPSS).
13	Using computer software in data analysis(SPSS).
14	Using computer software in data analysis(SPSS).
15	Using computer software in data analysis(SPSS).

Course outline:

Subject

Introduction

- Course description
- Course objectives
- Course syllabus
- Students evaluation (examination)

Population and sample

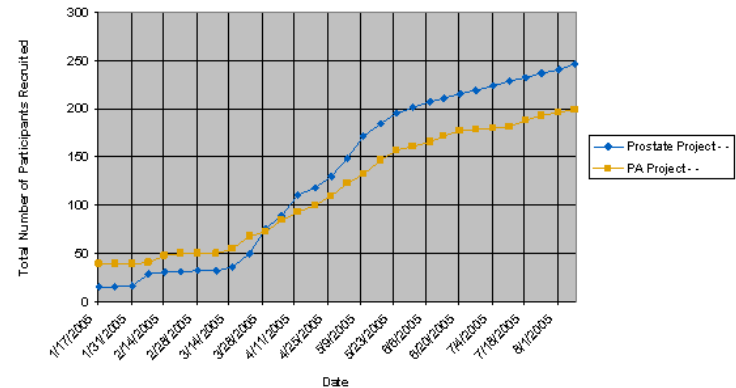
- Study population
- Sample
- Reasons for sampling
- Probability (random) sampling
- Probability sampling methods
 - Simple random sampling
 - Systematic sampling
 - Stratified sampling
 - Cluster sampling
- Non-probability sampling

Types of variables

- Qualitative (categorical) variables
 - Nominal
 - Ordinal
- Quantitative variables
 - Continuous
 - Discrete
- Dependent (outcome) variable
- Independent (exposure/risk factor) variable

Data presentation

- Types of data
- Frequency distributions
- Graphical methods
 - Bar graph
 - Histogram
 - Pie chart
 - Line graph

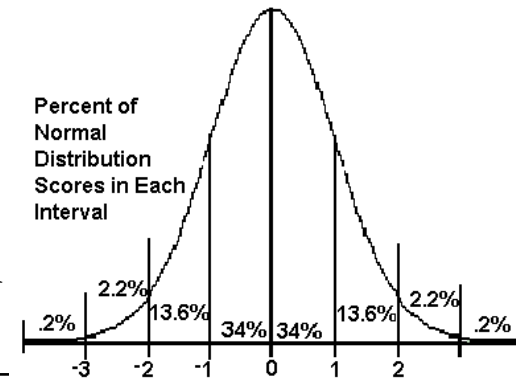


Descriptive measures

- Measure of central tendency
 - Mean/Median/Mode
- Measure of dispersion
 - Standard deviation/Range

Confidence intervals

- Statistical inference
- Estimation
- Confidence interval for a population mean
- Confidence interval for a population proportion
- Standard error and standard deviation



Hypothesis testing

- Null hypothesis
- Level of significance
- Test of significance

Comparison of two means

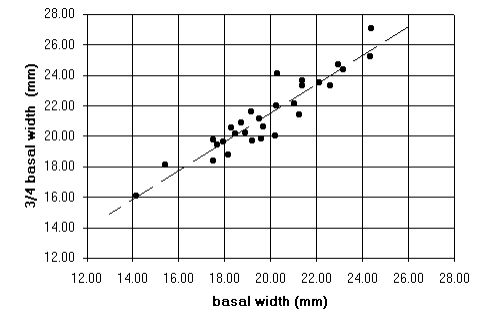
- The t-test for difference between two means (independent samples)
- Paired t-test: difference between paired samples

Comparison of two proportions

- Contingency tables
- Chi-square test: Relationship between two categorical variables

Sample size calculation

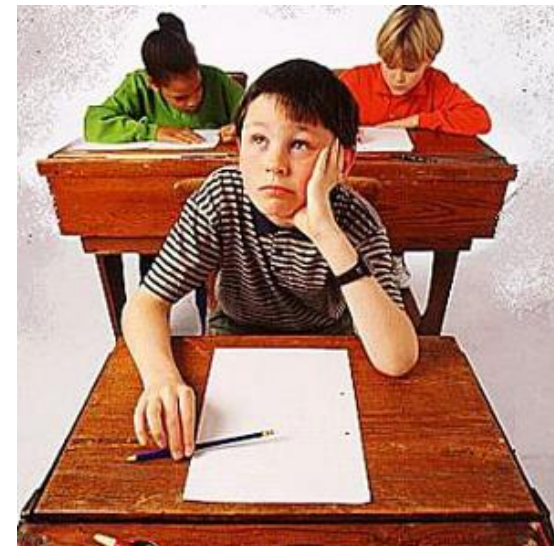
- One sample
 - Mean
 - Proportion
- Two samples
 - Difference between two means
 - Difference between two proportions
- Application of computer software for data analysis (SPSS).



Examinations

The examinations that will be held during the course include:

- Quizzes - during the sessions
- Mid course exam – Theoretical including basic statistical exercises
- End of course exam – Theoretical including basic statistical exercises



Examinations

- Theoretical exams will include different types of question.
- Samples of the expected questions and their answers:

Q1- Mark each of the following by (T) if true or (F) if false, in front of each of them.

- 1. In simple random sampling, each unit has an equal and known probability of being selected.
- 2. If we study the association between eating salty food and having high blood pressure, then high blood pressure is the independent variable.
- 3. A histogram can be used to display frequency distribution but not relative frequency distribution.
- 4. The best way to visualize correlation is by using scatter-plot.
- 5. When the confidence intervals of two values overlap, this means there is evidence of having a real difference between the two values.

Q2- Encircle the most appropriate answer in the following statements:

Examples of multiple choice questions, with notes and solutions

1. A study was carried out to investigate low birth weight in a large city. Pregnant women were recruited early in pregnancy and followed up until the birth. Each baby's birth weight was classified as low if it was 2.5 kg or less and normal otherwise. This variable is:

- a Continuous
- b Ordered categorical
- c Binary
- d Skewed

Q3. Statistical questions

The weights of 120 successive babies born alive to residents of a Health District were recorded to the nearest 100g. The results were as follows:

<u>Wt (cgm)</u>	<u>Number</u>
19	1
20	0
21	0
22	1
23	0
24	1
25	2
26	6
27	2
28	11

<u>Wt (cgm)</u>	<u>Number</u>
29	2
30	7
31	4
32	12
33	11
34	15
35	5
36	5
37	13
38	4

<u>Wt (cgm)</u>	<u>Number</u>
39	6
40	1
41	5
42	3
43	1
44	0
45	0
46	1
47	1

Q3. cont.

- (a) Draw up the cumulative percentage frequency table, and find the quartiles and the median.
- (b) Illustrate the distribution using (i) a histogram, and (ii) a box plot. Were these birth weights symmetrically distributed?
- (c) Compute the mean and standard deviation of the birth weights, and compare the mean and the median. Also compute the mean \pm SD, and the mean \pm 2SD, and see how these relate to the percentiles of the distribution.

Questions?