Stat 100
MATHEMATICAL STATISTICS

Lecturer: Areeg Abdalla

TA: Ms. Walaa

Text Book: Elementary Statistics
by Mario F. Triola, 11th edition

Lectures, Hws, …etc will be available on
http://scholar.cu.edu.eg/?q=areegsaid/node/25645
Go to classes then find your course code

1

Grading System

60 % Final

10 % Midterm

30 % Quizzes

Quizzes will be in Sections or in Class.

Some quizzes will be announced before either in class or in lab. Some quizzes will be pop up quizzes

Hws will be assigned to help you study. Some questions will be picked up to be solved in practice sections

2

11

MATHEMATICAL STATISTICS Syllabus

- Collection, preparation and tabulation of data, frequency distributions,
- central tendency measures: mean, median and mode,
- dispersion measures: range, semi-quartile difference, standard deviation,
- comparison between distributions, linear regression, correlation (Pearson and Spearman) time series analysis,
- introduction to probability: sample space, event, calculus of events, conditional probability and independence, random variables,
- probability distribution, some important distributions (binomial, Poisson, normal), sampling and sampling distribution, ratio distribution (small sample, large sample), confidence interval, tests of statistical hypothesis, index number.

3

Population

- **Population**

the complete collection of ALL individuals (scores, people, measurements, etc.) to be studied

the population is usually too big to be studied directly, then statistics is used

4

Parameter

a numerical measurement (value) describing some characteristic of a **population**

population

parameter

Example:

US population (about 300 million)

$N = 300{,}000{,}000$

Examples of parameters:

1. Proportion of people supporting Health Care Reform

   One can denote it by p $(0<p<1)$

2. Average weight of Americans

   One can denote it by μ $(\mu>0)$

2

Census versus Sample

Census

Collection of data from ***every*** member of a population

(must include all N measurements)

Sample

Subcollection of members selected from a population

(n measurements; $n<N$; n is sample size)

Statistic

Statistic

a numerical measurement describing some characteristic of a **sample**.

sample

statistic

Data

❖ **Data**

collections of observations (such as measurements, records, survey responses, etc.)

Next slides will describe types of data

Quantitative Data

❖ **Quantitative (or numerical) data**

consists of *numbers* representing counts or measurements.

Example: The weights of selected people

Example: The ages of respondents

3

Categorical Data

❖ **Categorical (or qualitative) data**

consists of names or "labels" (representing categories)

Example: The genders (male/female) of professional athletes

Example: Answers in a poll (yes/no)

Example: Students' grades (A,B,C,D,F)

Working with Quantitative Data

Quantitative (numerical) data can further be described by distinguishing between

discrete and continuous

types

Discrete Data

- **Discrete data**

  these are numbers whose possible values are either a finite list of values or a 'countable' list of values

  (for instance, possible values are

  0, 1, 2, 3, . . .)

  Example: The number of children in a family

Continuous Data

- **Continuous (numerical) data**

  these are numbers with infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps

  Example: The weight of a person

  Example: The annual income of a person

Working with Categorical Data

Categorical data can further be described by distinguishing between

nominal and ordinal

types

- **Nominal data types**

  these are data types that consist of names, labels, or categories only, i.e. the data cannot be arranged in any ordering scheme (such as low to high)

  Example: Survey responses (yes, no, undecided)

  Example: Genders (male, female)

❖ Ordinal data

these are data types that can be arranged in some order, but there are no numerical differences between data values

Example: Course grades (A, B, C, D, or F)

Example: Survey responses such as

(highly satisfied/satisfied/unsatisfied/very unsatisfied)

Simple Random Sample

❖ Simple Random Sample

of n subjects selected in such a way that every possible sample of the same size n has the same chance of being chosen

5

Example: Pulse Rates of Females and Males

Original data:

Table 2-1 Pulse Rates (beats per minute) of Females and Males

Females																			
76	72	88	60	72	68	80	64	68	68	80	76	68	72	96	72	68	72	64	80
64	80	76	76	76	80	104	88	60	76	72	72	88	80	60	72	88	88	124	64
Males																			
68	64	88	72	64	72	60	88	76	60	96	72	56	64	60	64	84	76	84	88
72	56	68	64	60	68	60	60	56	84	72	84	88	56	64	56	56	60	64	72

Frequency Distribution Pulse Rates of Females

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60–69	12
70–79	14
80–89	11
90–99	1
100–109	1
110–119	0
120–129	1

Table 2-3 Relative Frequency Distribution of Pulse Rates of Females

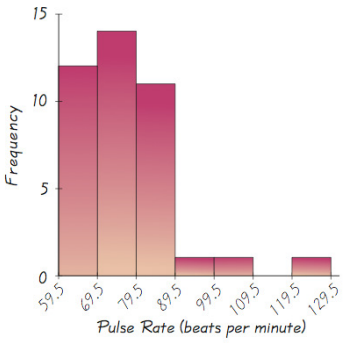
Pulse Rate	Relative Frequency
60–69	30%
70–79	35%
80–89	27.5%
90–99	2.5%
100–109	2.5%
110–119	0
120–129	2.5%

Histogram

Basically a graphic version of a frequency distribution.

Table 2-2 Pulse Rates of Females

Pulse Rate	Frequency
60–69	12
70–79	14
80–89	11
90–99	1
100–109	1
110–119	0
120–129	1



Another graph describing pulse rates of female:

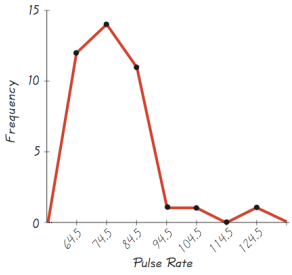


Figure 2-5 Frequency Polygon: Pulse Rates of Women

66

Two graphs allow us to compare two distributions:

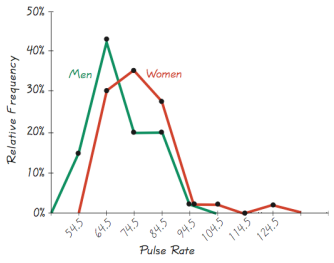
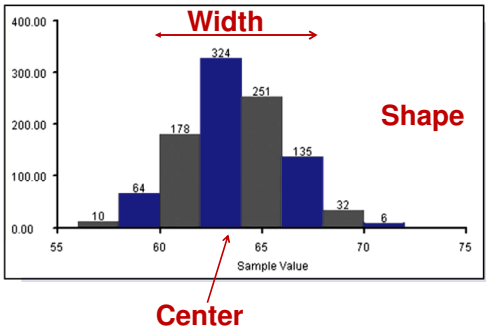


Figure 2-6 Relative Frequency Polygons: Pulse Rates of Women and Men

Characteristics of a distribution



The most common (typical) shape is a "bell" shape, looks like a hill. It is called a normal distribution. It has the following properties:

- **The frequencies start low on the left, then increase to one or two highest frequencies in the middle, then decrease again to low frequencies on the right.**
- **The distribution is approximately symmetric, with frequencies preceding the maximum being roughly a mirror image of those that follow the maximum.**

Normal distribution is characterized by two numbers (parameters):

1. **Center (most typical data value)**

2. **Width (degree of variation of data values); also called spread or variability of data values**