

Statistics: the art and science of collecting, analyzing, presenting, and interpreting data.

Data: facts and figures (measurements) that are collected.

Data set: all the data collected in a particular study.

Elements: the entities on which the data are collected. (People, firms, mobiles,...)

Population: the collection of all elements in a study. (**Population size:** number of elements of a population)

Sample: a subset of the population. (**Sample size:** number of elements of a sample)

Census: collecting data on the entire population.

Survey: collecting data on a sample.

Variable: characteristic of interest for the elements.

Observation: the set of measurements for a particular element.

Table 1. The following table represents a study on a **sample** of BZU students.

Student Number	Eye color	Major	Weight (kg)	Birth year	Credit Hours	Rating of BZU
1174586	Brown	Accounting	56	1999	77	Excellent
1102364	Brown	Engineering	80	1992	142	Poor
1130002	Black	Math	68	1995	90	Good
1162548	Blue	Finance	71	1998	65	Excellent
1180210	Green	Accounting	94	2000	32	Neutral
1168794	Brown	Finance	62	1998	59	Good
1156984	Blue	Economic	52	1997	85	Excellent
1112020	Blue	Accounting	102	2001	112	Excellent

- **Population:** All students of BZU.
- **No. of elements:** 8
- **No. of observations:** 8
- **No. of variables:** 7 (Student number, Eye color, ..., Rating of BZU)
- **Type of study:** Survey.

Scales of measurement

Variables (data) have four scales of measurements:

Scale of measurement	Properties	Examples
Nominal	Data are just labels or names Order or rank is not meaningful Data are either numeric or nonnumeric	Gender (male, female) Blood type (A, B, AB, O) Phone numbers Student Numbers
Ordinal	Data are labels or names Order or rank is meaningful Data are either numeric or nonnumeric	Service rating (excellent, good, poor) Clothes size (xlarge, large, medium, small) Satisfaction (Strongly approve, approve, neutral, disapprove, strongly disapprove)
Interval	Data are numeric Order or rank is meaningful The interval (difference) between values exists No true or meaningful zero	Temperature Date (year) of birth SAT scores IQ test
Ratio	Data are numeric Order or rank is meaningful True or meaningful zero Ratios are meaningful	Profit Age Distance Salary Family size

Types of variables (Data):

- 1) **Qualitative (Categorical)**: data are nominal or ordinal.
- 2) **Quantitative**: data are interval or ratio.

Types of quantitative variables:

- 1) **Discrete**: data values can be counted and gaps between values exist. Usually, no decimals. (How many)

Examples: Family size, Number of chairs in a room.

- 2) **Continuous**: Infinite number of values. No gaps. Include decimals. (How much)

Examples: revenue (\$), weight, Amount of calories in a cake.

Types of statistics:

- 1) **Descriptive statistics:** Tabular, graphical, and numerical summaries of data.

Tabular: frequency and cumulative frequency distributions, ...

Graphical: Histogram, pie chart, ogive, ...

Numerical: proportion, percentage, average, ...

- 2) **Statistical inference:** using data from a sample to make estimates about the population.

Example: In a sample of 200 students in BZU, 80 (40 %), have blue eyes. Based on this, The university reported that 40% of all BZU students have blue eyes.

Data sources:

- 1) **Existing sources:** data already exist. (Bank records, customer profiles, ...)
 - 2) **Statistical studies:** (a) **Experimental:** studying how a new drug affects the blood pressure.
(b) **Observational:** Surveys and questionnaires.
-

Cross-sectional data: data collected at the same, or approximately, the same time.

Time series data: data collected over several periods of time.

Parameter: a numerical measure based on the population. (population mean, ...)

Statistic: a numerical measure based on the sample. (sample mean, ...)

Proportion, percentage, average

Example: given the sample of blood types: A, AB, O, A, B, A, B, AB, AB, AB, A, AB

- 1) Find the proportion of the blood of type AB. $\text{Proportion} = \frac{5}{12} = 0.42$
- 2) Find the percentage of blood of type O. $\text{percentage} = \frac{1}{12} \times 100 = 8.33 \%$

Example: Given the sample of ages: 12, 15, 20, 57, 16, 19 Find the average (mean)

$$\text{Average} = \frac{12+15+20+57+16+19}{6} = \frac{139}{6} = 23.17$$

Scales (Levels) of measurements

The table below shows some **examples** of the levels of measurements.

Nominal	Ordinal	Interval	Ratio
Specialization (major)	Hotel service (Very good, good, poor)	Temperature	Salary
Gender	Employee rank (1 - 10)	Date (year) of birth	Years of experience
Employee (student) number	Room rate (\$, \$\$, \$\$\$, ...)	SAT scores	Family size
Phone numbers	Product rate (Excellent, good, average, poor)	IQ scores	Prices
Department (Accounting, Finance, IT, ...)	Interest in Statistics (Low, medium, high)	ACT scores	Tuitions (JD)
Blood type	Sound quality (very good, good, fair)		Age
Type of school (Private, governmental)	T-shirt size (xlarge, large, medium, small)		Income (\$)
Place of birth	Cafeteria Evaluation (Low, moderate, high)		Overtime hours
Mailbox numbers	Grades (A, b, C, D, F)		Weight of cell phones
State of residence	Opinion of government policy (Strongly approve, approve, neutral, disapprove, strongly disapprove)		Area
Religion	Ranking of tennis players		Amount of calories
Mode of transportation			Height
Marital status			Number of students in BZU
Types of cars			
Sector of economy (Public, private, nonprofit)			
Eye color			
Nationality			
Categories of magazines			