Chapter 1

The Role of Statistics

Three Reasons to Study Statistics

1. Being an informed "Information Consumer"
   - Extract information from charts and graphs
   - Follow numerical arguments
   - Know the basics of how data should be gathered, summarized and analyzed to draw statistical conclusions

Three Reasons to Study Statistics

2. Understanding and Making Decisions
   - Decide if existing information is adequate
   - Collect more information in an appropriate way
   - Summarize the available data effectively
   - Analyze the available data
   - Draw conclusions, make decisions, and assess the risks of an incorrect decision

Three Reasons to Study Statistics

3. Evaluate Decisions That Affect Your Life
   - Help understand the validity and appropriateness of processes and decisions that effect your life

What is Statistics?

Statistics is the science of
- Collecting data
- Analyzing data
- Drawing conclusions from data

Major Branches of Statistics

1. Descriptive Statistics
   - Organizing, Summarizing Information
   - Graphical techniques
   - Numerical techniques

2. Inferential Statistics
   - Estimation
   - Decision making

Important Terms

Population – The entire collection of individuals or objects about which information is desired is called the population.

Collection of data from every member of the population,

- allows a question about the population to be definitively answered.
- may be expensive, impractical, or sometimes impossible.

Important Terms

Sample – A sample is a subset of the population, selected for study in some prescribed manner.

Using sample data rather than population data is

- more practical than a census.
- Gives variable results with some possibility of a wrong conclusion being adopted.

Discussion on Important Terms

Generally it not reasonable, feasible or even possible to survey a population so that descriptions and decisions about the population are made based on using a sample.

The study of statistics deals with understanding how to obtain samples and work with the sample data to make statistically justified decisions.

Important Terms

Variable – A variable is any characteristic whose value may change from one individual to another

Examples:

- Brand of television
- Height of a building
- Number of students in a class

Important Terms

Data results from making observations either on a single variable or simultaneously on two or more variables.

A **univariate data set** consists of observations on a single variable made on individuals in a sample or population.

Important Terms

A **bivariate data set** consists of observations on two variables made on individuals in a sample or population.

A **multivariate data set** consists of observations on two or more variables made on individuals in a sample or population.

Data Sets

- A univariate data set is **categorical** (or **qualitative**) if the individual observations are categorical responses.
- A univariate data set is **numerical** (or **quantitative**) if the individual observations are numerical responses *where numerical operations generally have meaning*.

Examples of categorical Data

The brand of TV owned by the six people that work in a small office

RCA Magnavox Zenith Phillips
GE RCA …

The zipcodes of the 6 students in the first row of seats

40208	40202	40241
40207	40059	40245

*Since numerical operations with zipcodes make no sense, the zipcodes are categorical rather than numeric.

Types of Numerical Data

Numerical data is **discrete** if the possible values are isolated points on the number line.

Numerical data is **continuous** if the set of possible values form an entire interval on the number line.

Examples of Discrete Data

- The number of costumers served at a diner lunch counter over a one hour time period is observed for a sample of seven different one hour time periods

  13 22 31 18 41 27 32

- The number of textbooks bought by students at a given school during a semester for a sample of 16 students

  5 3 6 8 6 1 3 6 12
  3 5 7 6 7 5 4

Examples of Continuous Data

- The height of students that are taking a Data Analysis at a local university is studied by measuring the heights of a sample of 10 students.

  72.1” 64.3” 68.2” 74.1” 66.3”
  61.2” 68.3” 71.1” 65.9” 70.8”

  Note: Even though the heights are only measured accurately to 1 tenth of an inch, the actual height could be any value in some reasonable interval.

Examples of Continuous Data

The crushing strength of a sample of four jacks used to support trailers.

7834 lb 8248 lb 9817 lb 8141 lb

Gasoline mileage (miles per gallon) for a brand of car is measured by observing how far each of a sample of seven cars of this brand of car travels on ten gallons of gasoline.

23.1 26.4 29.8 25.0 25.9
22.6 24.3

Frequency Distributions

- A **frequency distribution for categorical data** is a table that displays the possible categories along with the associated frequencies or relative frequencies.
- The **frequency** for a particular category is the number of times the category appears in the data set.

Frequency Distributions

The **relative frequency** for a particular category is the fraction or proportion of the time that the category appears in the data set. It is calculated as

$$\text{relative frequency} = \frac{\text{frequency}}{\text{number of observations in the data set}}$$

When the table includes relative frequencies, it is sometimes referred to as a **relative frequency distribution**.

Classroom Data Example

This slide along with the next contains a data set obtained from a large section of students taking Data Analysis and will be utilized throughout this slide show in the examples.

Code	Age	Weight	Height	Gender	Vision	Smoke
1	21	150	70	Male	None	No
2	19	124	70	Male	Glasses	No
3	19	121	68	Male	None	No
4	23	200	74	Male	Glasses	No
5	24	130	69	Male	Glasses	No
6	20	188	72	Male	None	No
7	19	183	69	Male	None	No
8	20	140	70	Male	None	No
9	19	155	69	Male	None	No
10	19	125	63	Male	Glasses	No
11	18	165	70	Male	None	No
12	19	168	69	Male	Glasses	Yes
13	24	138	67	Male	Glasses	No
14	21	160	69	Male	Glasses	No
15	19	150	71	Male	Glasses	No
16	20	150	74	Male	None	No
17	21	117	66	Male	None	No
18	21	145	70	Male	None	No
19	20	155	68	Male	None	No
20	19	135	69	Male	Contacts	No
21	22	145	68	Male	None	No
22	23	175	70	Male	Glasses	No
23	22	170	72	Male	None	No
24	21	140	66	Male	None	No
25	21	175	70	Male	None	No
26	20	140	71	Male	None	No
27	21	210	73	Male	Glasses	No
28	18	225	76	Male	Glasses	No
29	21	170	67	Male	Glasses	No
30	28	237	70	Male	None	Yes
31	26	175	68	Male	Glasses	No
32	25	140	71	Male	Glasses	No
33	22	160	70	Male	None	No
34	19	130	68	Male	None	No
35	18	160	74	Male	None	No
36	20	135	68	Male	Glasses	No
37	19	145	68	Male	Glasses	No
38	23	170	76	Male	Glasses	No
39	22	140	69	Male	None	Yes

Classroom Data Example
continued

Code	Age	Weight	Height	Gender	Vision	Smoke
41	19	130	69	Male	Glasses	No
42	18	170	76	Male	None	No
43	19	155	68	Male	None	No
44	22	165	71	Male	None	No
45	23	185	75	Male	None	No
46	19	160	60	Male	Contacts	No
47	22	225	75	Male	Glasses	No
48	21	180	73	Male	Contacts	No
49	28	239	69.5	Male	None	Yes
50	21	175	74	Male	Contacts	Yes
51	18	140	68	Male	Glasses	No
52	19	165	73	Male	Glasses	No
53	19	170	72	Male	Glasses	Yes
54	19	156	69	Male	Contacts	No
55	38	150	61	Female	Glasses	No
56	17	140	68	Female	Glasses	No
57	19	155	61	Female	Contacts	No
58	44	195	67	Female	Glasses	No
59	24	139	66	Female	Glasses	No
60	37	200	65	Female	Contacts	No
61	21	157	62	Female	None	Yes
62	20	130	63	Female	Glasses	No
63	20	113	60	Female	None	No
64	22	130	64	Female	None	No
65	23	121	65	Female	Contacts	Yes
66	21	140	67	Female	Contacts	No
67	22	140	62	Female	Glasses	No
68	21	150	64	Female	Contacts	No
69	19	125	61	Female	Glasses	Yes
70	22	135	67	Female	None	No
71	20	124	64	Female	None	No
72	21	130	67	Female	None	No
73	30	150	65	Female	None	No
74	23	125	67	Female	None	No
75	22	120	69	Female	None	No
76	22	103	61	Female	None	No
77	47	170	70	Female	Glasses	No
78	19	124	66.5	Female	None	No
79	19	160	69	Female	Glasses	No

Frequency Distribution Example

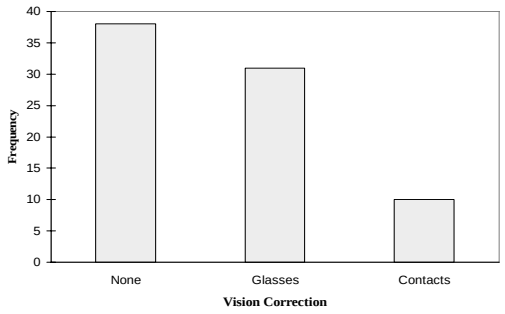
The data in the column labeled vision is the answer to the question, "What is your principle means of correcting your vision?" The results are tabulated below

Vision Correction	Frequency	Relative Frequency
None	38	38/79 = 0.481
Glasses	31	31/79 = 0.392
Contacts	10	10/79 = 0.127
	79	1.000

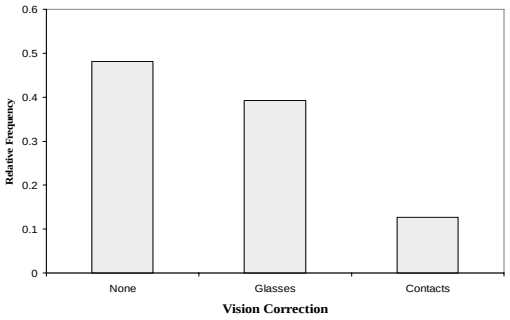
Bar Chart – Procedure

1. Draw a horizontal line, and write the category names or labels below the line at regularly spaced intervals.
2. Draw a vertical line, and label the scale using either frequency (or relative frequency).
3. Place a rectangular bar above each category label. The height is the frequency (or relative frequency) and all bars have the same width.

Bar Chart – Example (frequency)

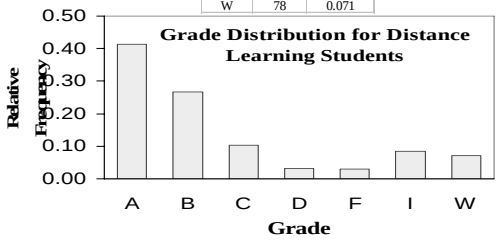


Bar Chart – (Relative Frequency)



Another Example

Grade	Students	Student Proportion
A	454	0.414
B	293	0.267
C	113	0.103
D	35	0.032
F	32	0.029
I	92	0.084
W	78	0.071

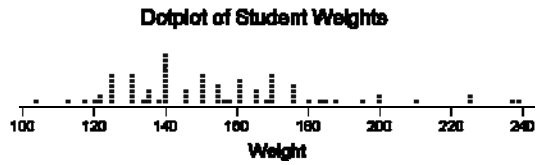


Dotplots - Procedure

1. Draw a horizontal line and mark it with an appropriate measurement scale.
2. Locate each value in the data set along the measurement scale, and represent it by a dot. If there are two or more observations with the same value, stack the dots vertically.

Dotplots - Example

Using the weights of the 79 students



Dotplots – Example continued

To compare the weights of the males and females we put the dotplots on top of each other, using the same scales.