Elementary Statistics

A Step by Step Approach
Sixth Edition

by
Allan G. Bluman
http://www.mhhe.com/math/stat/blumanbrief

SLIDES PREPARED
BY
LLOYD R. JAISINGH
MOREHEAD STATE UNIVERSITY
MOREHEAD KY

Updated by
Dr. Saeed Alghamdi
King Abdulaziz University

Chapter 2

Frequency Distributions and Graphs

Objectives

1-1

- Organize data using frequency distributions.
- Represent data in frequency distributions graphically using histograms, frequency polygons and ogives.
- Represent data using bar chart, Pareto chart, pie graph and time series graph.
- Draw and interpret a stem and leaf plot.

Notes

- …….....
- …….....
- …….....
- …….....
- …….....

Organizing Data

2-2

- When data are collected in original form, they are called *raw data*.
- When the *raw data* are organized into a table which called *frequency distribution*, the frequency will be the number of values in a specific class of the distribution.

Notes

- …….....
- …….....
- …….....
- …….....
- …….....

Organizing Data

2-3

- A *frequency distribution* is the organization of raw data in a table form, using classes and frequencies.
- Types of frequency distributions are *categorical frequency distribution*, *ungrouped frequency distribution* and *grouped frequency distribution*.

Notes

- ……
- ……
- ……
- ……
- ……

Why Construct Frequency Distributions?

2-4

To organize the data in a meaningful, intelligible way.

To enable the reader to make comparisons among different data sets.

To facilitate computational procedures for measures of average and spread.

To enable the reader to determine the nature or shape of the distribution.

To enable the researcher to draw charts and graphs for the presentation of data.

Notes

- ……
- ……
- ……
- ……
- ……

Categorical Frequency Distributions

2-5

- When the sample size (n) is large, the data must be grouped into categories.
- Categorical Frequency Distributions are used for data that can be placed in specific categories, such as *nominal* or *ordinal* level data.

Notes

- ……
- ……
- ……
- ……
- ……

Categorical Frequency Distributions

2-6

- Example: Blood Type Frequency Distribution

A	B	B	AB	O	O	A	O	O	B	A	B	O	AB
O	AB	B	B	A	A	O	B	B	O	O	O	A	O

Class	Frequency	Percent
A	6	21%
B	8	29%
O	11	39%
AB	3	11%
Total	**28**	**100%**

$$\frac{frequency}{Total} = \frac{f}{\sum f} = \frac{f}{n}\%$$

Sample Size

Notes

- .....
- .....
- .....
- .....
- .....

Ungrouped Frequency Distributions

2-7

- Ungrouped frequency distributions are used for data that can be enumerated and when the range of values in the data set is small and the sample size (n) is large.
- Example: number of children in a family.

Notes

- .....
- .....
- .....
- .....
- .....

Ungrouped Frequency Distributions

2-8

- Example: Number of miles traveled by 16 instructors from their houses to university.

10	4	4	20
20	20	10	4
4	4	20	4
10	20	4	4

Class	Frequency	Cumulative Frequency	%
4	8	8	50%
10	3	8+3=11	19%
20	5	11+5=16	31%
Total	**16**	-	**100%**

Notes

- .....
- .....
- .....
- .....
- .....

Grouped Frequency Distributions

2-9

- When the range of the data is large, the data must be grouped into classes that are more than one unit in width, e.g., 24 – 30.
- The *lower class limit* represents the smallest data value that can be included in a class, e.g., 24 in the class limit 24 – 30.
- The *upper class limit* represents the largest value that can be included in the class, e.g., 30 in the class limit 24 – 30.

Notes

- ……
- ……
- ……
- ……
- ……

Grouped Frequency Distributions

2-10

- The *class boundaries* are used to separate the classes so that there are no gaps in the frequency distribution. It can be found by subtracting 0.5 from the last digit in the lower class limit and adding 0.5 to the last digit in the upper class limit e.g., 23.5 – 30.5 for the class limit 24 – 30 and 2.65 – 6.85 for the class limit 2.7 – 6.8.
- Rule of Thumb: Class limits should have the same decimal place value as the data, but the class boundaries have one additional place value and end in a 5.

Notes

- ……
- ……
- ……
- ……
- ……

Finding Class Boundaries

2-11

- The *class width* for a class in a frequency distribution is found by subtracting the lower (or upper) class limit of one class from the lower (or upper) class limit of the next class.
- The *class midpoint* is found by adding the lower and upper boundaries (or limits) and dividing by 2.

Notes

- ……
- ……
- ……
- ……
- ……

Class Rules

2-12

- There should be between 5 and 20 classes.
  As a guide line, the number of classes can be found using $\text{Number of Classes} \approx 1 + 3.3 \times \log(n)$
- The class width should be an odd number.
- The classes must be mutually exclusive.
- The classes must be continuous.
- The classes must be exhaustive.
- The classes must be equal in width.

Notes

- .....
- .....
- .....
- .....
- .....

Procedure for Constructing a Grouped Frequency Distribution

2-13

- Find the highest (H) and lowest (L) value.
- Find the range (R). R=H – L
- Select the number of classes desired, usually between 5 and 20.
- Find the width by dividing the range by the number of classes and rounding up.

Notes

- .....
- .....
- .....
- .....
- .....

Procedure for Constructing a Grouped Frequency Distribution

2-14

- Select a starting point (usually the lowest value); add the width to get the lower limits.
- Find the upper class limits.
- Find the boundaries.
- Tally the data.
- Find the numerical frequencies from the tallies.
- Find the cumulative frequencies.

Notes

- .....
- .....
- .....
- .....
- .....

Grouped Frequency Distributions

2-15

- Example: Added cost per vehicle use due to bad roads.

165	186	122	172	140
153	208	169	156	114
113	135	131	125	177
136	136	127	112	188
171	179	152	155	116
90	187	136	159	97
141	85	91	170	111
147	165	163	159	150

Class Limits	Class Boundary	Tally
85—103	84.5—103.5	////
104—122	103.5—122.5	/// /
123—141	122.5—141.5	/// ////
142—160	141.5—160.5	/// ///
161—179	160.5—179.5	/// ////
180—198	179.5—198.5	///
199—217	198.5—217.5	/
Total		40

Notes

- .....
- .....
- .....
- .....
- .....

Grouped Frequency Distributions

2-16

- Example: Added cost per vehicle use due to bad roads frequency distribution

$\frac{Upper + Lower}{2}$

Class Limits	Class Boundary	Class Midpoint	Frequency	Cumulative Frequency	%
85—103	84.5—103.5	94	4	4	10%
104—122	103.5—122.5	113	6	10	15%
123—141	122.5—141.5	132	9	19	22.5%
142—160	141.5—160.5	151	8	27	20%
161—179	160.5—179.5	170	9	36	22.5%
180—198	179.5—198.5	189	3	39	7.5%
199—217	198.5—217.5	208	1	40	2.5%
Total	-	-	**40**	-	**100%**

Notes

- .....
- .....
- .....
- .....
- .....

The Role of Graphs

2-17

- The purpose of graphs in statistics is to represent the data to the viewer in pictorial form.
- Graphs are useful in getting the audience's attention in a publication or a presentation.

Notes

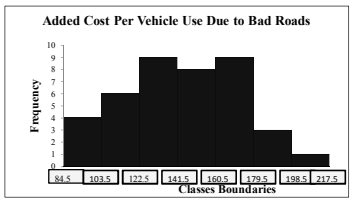
- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-18

- The *histogram* displays the continuous data that are organized in a grouped frequency distribution by using vertical bars of various heights to represent the frequencies.

Class Boundary	Frequency
84.5—103.5	4
103.5—122.5	6
122.5—141.5	9
141.5—160.5	8
160.5—179.5	9
179.5—198.5	3
198.5—217.5	1
Total	40



Notes

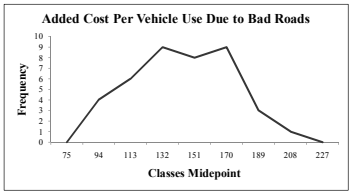
- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-19

- The *frequency polygon* displays the continuous data that are organized in a grouped frequency distribution by using lines that connect points plotted for the frequencies at the midpoints of the classes.

Class Midpoint	Frequency
94	4
113	6
132	9
151	8
170	9
189	3
208	1



Notes

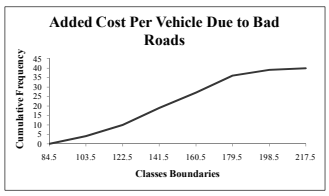
- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-20

- The *cumulative frequency graph* or *ogive* represents the cumulative frequencies for the classes in a grouped frequency distribution.

Class Boundary	Cumulative Frequency
84.5—103.5	4
103.5—122.5	10
122.5—141.5	19
141.5—160.5	27
160.5—179.5	36
179.5—198.5	39
198.5—217.5	40



Notes

- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-21

- *Relative frequency graph*

  Graphs of relative frequencies used instead of frequencies when the proportion of data values that fall into a given class is more important than the actual number of data values that fall into that class.

Notes

- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-22

- Relative frequency example: $\frac{frequency}{Total} = \frac{f}{\sum f} = \frac{f}{n}$

Class Boundary	Frequency	Relative Frequency	Cumulative Frequency	Relative Cumulative Frequency
84.5—103.5	4	0.1	4	0.1
103.5—122.5	6	0.15	10	0.25
122.5—141.5	9	0.225	19	0.475
141.5—160.5	8	0.2	27	0.675
160.5—179.5	9	0.225	36	0.90
179.5—198.5	3	0.075	39	0.975
198.5—217.5	1	0.025	40	1
Total	**40**	**1**	-	-

$n = \sum f$

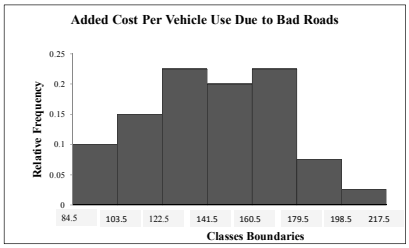
Notes

- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-23

- The *histogram*



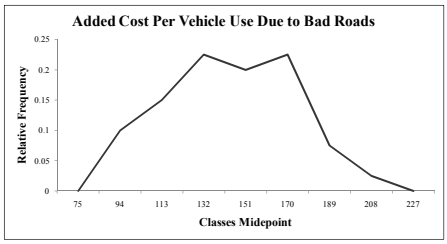
Notes

- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-24

- The *frequency polygon*



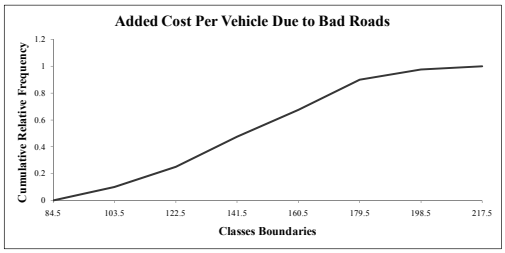
Notes

- .....
- .....
- .....
- .....
- .....

The Most Common Graphs

2-25

- The *cumulative frequency graph* or *ogive*



Notes

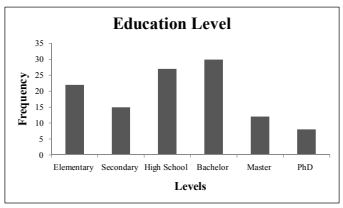
- .....
- .....
- .....
- .....
- .....

Other Types of Graphs

2-26

- The *bar charts* displays the data by using vertical bars of various heights to represent the frequencies of discrete or categorical variables.

Education Levels	Frequency
Elementary	22
Secondary	15
High School	27
Bachelor	30
Master	12
PhD	8
Total	**114**



Notes

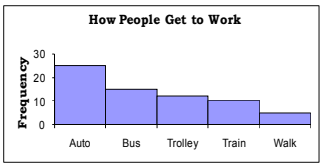
- .....
- .....
- .....
- .....
- .....

Other Types of Graphs

2-27

- A *Pareto chart* is used to represent a frequency distribution for categorical variable. The frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

Method	Frequency
Walk	5
Auto	25
Bus	15
Train	11
Trolley	13
Total	**69**



Notes

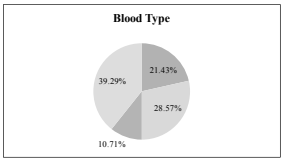
- .....
- .....
- .....
- .....
- .....

Other Types of Graphs

2-28

- The *pie graph* is a circle that is divided into sections according to the percentage of frequencies in each category of the distribution, $Degree = \frac{f}{n} \times 360$.

Class	Frequency	Percentage	Degree
A	6	21.43%	77.14
B	8	28.57%	102.86
O	11	39.29%	141.43
AB	3	10.71%	38.57
Total	**28**	**100%**	**360**



Notes

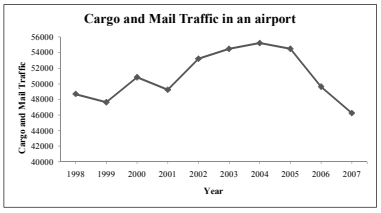
- .....
- .....
- .....
- .....
- .....

Other Types of Graphs

2-29

- The *time series graph* represents data that occur over a specific period of time.

Year	Domestic
1998	48685
1999	47621
2000	50827
2001	49229
2002	53224
2003	54476
2004	55221
2005	54483
2006	49622
2007	46240



Notes

- .....
- .....
- .....
- .....
- .....

Other Types of Graphs

2-30

- A *stem-and-leaf plot* is a data plot that uses part of a data value as the stem, the most significant digit (i.e. the 'tens'), and the other part of the data value as the leaf, the less significant digits (the 'units'), to form groups or classes.
- It has the advantage over grouped frequency distribution of retaining the actual data while showing them in a graphic form.

Notes

- .....
- .....
- .....
- .....
- .....

Other Types of Graphs

2-31

The ages of a sample of a university instructors are shown below. Construct a stem and leaf plot and analyze the data.

57	54	52	55	51	56
61	68	56	55	54	61
57	51	46	54	51	52
57	49	54	42	60	69
58	64	49	51	62	64
57	48	50	56	43	46
61	65	47	55	55	54

Notes

- .....
- .....
- .....
- .....
- .....

Other Types of Graphs

2-32

4	2 3
4	6 6 7 8 9 9
5	0 1 1 1 1 2 2 4 4 4 4 4
5	5 5 5 5 6 6 6 7 7 7 7 8
6	0 1 1 1 2 4 4
6	5 8 9

The majority of the university instructors were in their 50's.

Notes

- .....
- .....
- .....
- .....
- .....