

## Chapter 2 Descriptive Statistics: Tabular and Graphical Methods

---

---

---

---

---

---

---

---



### Ordered Array (nib)

- Organizes a data set by **sorting** it in either ascending or descending order
- Advantages & Disadvantages
  
- Useful in preparing tables and charts in Chapter 2 and numerical summary measures in Chapter 3
  
- What scale (level) of data is required?

2

---

---

---

---

---

---

---



### Frequency Distribution

- A grouping of data into mutually exclusive categories that shows the number of observations in each category
- Can be prepared for either qualitative or quantitative variables
- Class Frequency
- Relative Frequency (RF)
- Percent Frequency

3

---

---

---

---

---

---

---

## Frequency Distribution: Qualitative Variables

- Example: A sample of 200 runners were asked to indicate their favorite brand of running shoe.

Adidas		Brand	# of Runners	RF	PF
Nike		Nike	92	.460	46.0
Nike		Adidas	49	.245	24.5
Asics		Reebok	37	.185	18.5
Reebok		Asics	13	.065	6.5
Adidas		Other	9	.045	4.5
Nike		<b>Total</b>	<b>200</b>	<b>1.000</b>	<b>100.0</b>
Reebok					
Converse					
...					

- Dewey, Cheatham and Howe assignment

4

---

---

---

---

---

---

---

---

---

---

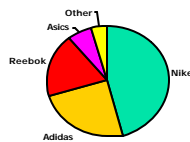
---

---

## Pie Chart

- Effective at showing the **proportion** of the total each category contributes
- How construct for the Running Shoe example?

Brand	# of Runners	RF	PF
Nike	92	.460	46.0
Adidas	49	.245	24.5
Reebok	37	.185	18.5
Asics	13	.065	6.5
Other	9	.045	4.5
<b>Total</b>	<b>200</b>	<b>1.000</b>	<b>100.0</b>



5

---

---

---

---

---

---

---

---

---

---

---

---

## Frequency Distribution: Quantitative Variables

- Construction of:
  - how many classes?  $2^c > n$
  - classes must be **mutually exclusive** and **all inclusive**
  - class interval width =  $(\text{Max} - \text{Min})/c$  then round up
- Gas Price assignment -
- Advantages & Disadvantages

6

---

---

---

---

---

---

---

---

---

---

---

---

## Frequency Distribution: Quantitative Variable

- Class Limits
  - upper and lower
- Class Frequency
- Class Midpoint
- Relative Frequency (RF)
- Cumulative Frequency (CF)
  - less-than orientation
  - must have at least ordinal-level data
- Cumulative Relative Frequency (CRF)
- Gas Price assignment

7

---

---

---

---

---

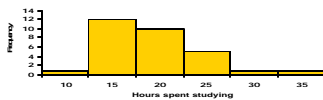
---

---

---

## Pictorial Displays

- Advantages & Disadvantages of graphs



8

---

---

---

---

---

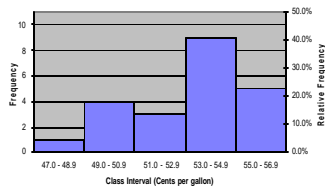
---

---

---

## Histogram

- Bar graph of a frequency distribution
- Can plot frequencies and/or relative frequencies
- Gas Price assignment



9

---

---

---

---

---

---

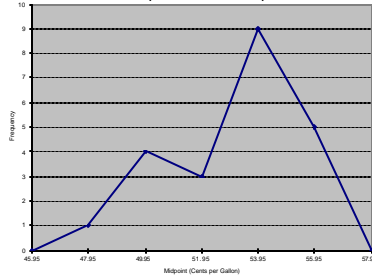
---

---



## Frequency Polygon (nlb)

- Plot of class midpoints vs. frequencies



10

---

---

---

---

---

---

---

---

---

---



## Skew

- Characterizes a distribution's shape by describing its degree of horizontal balance
- Figure 2.8, page 47
- 3 general cases
  - Symmetrical
  - Negative Skew
  - Positive Skew
- Gas Price assignment

11

---

---

---

---

---

---

---

---

---

---

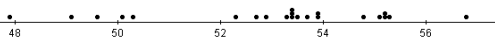


## Dotplot (nlb)

- Just like it sounds!
- No arbitrary decisions to make regarding the number of classes and class limits
- {2, 2, 4, 5, 6}



- Gas Price



12

---

---

---

---

---

---

---

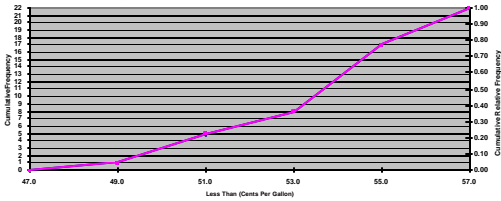
---

---

---

## Ogive

- Line graph of cumulative frequencies or cumulative relative frequencies



13

---

---

---

---

---

---

---

---

## Percentile (Chapter 3)

- Indicates the **proportion** of observations whose value is less than or equal to a specified value
- Easy to estimate from an ogive
  - in the Gas Price data set, 55 cents per gallon is approximately \_\_\_\_ percentile

14

---

---

---

---

---

---

---

---

## Exploratory Data Analysis

- Techniques that employ simple arithmetic and easy-to-draw pictures to summarize data quickly
- Stem-and-Leaf Display
- 5-Number Summary (Ch 3)
- Box Plot (Ch 3)

15

---

---

---

---

---

---

---

---

## Stem-and-Leaf Display

- A hybrid of an Ordered Data Array and Histogram
- Advantages & Disadvantages

Stem	Leaf
47	9
48	
49	1 6
50	1 3
51	
52	3 7 9
53	3 4 4 4 5 7 9 9
54	8
55	1 2 2 3
56	8

- Gas Price assignment
  - note the leaves are sorted and aligned

16

---

---

---

---

---

---

---

---

---

---

---

---

## Crosstabulation

- aka Contingency Table
- Cross-classifies two variables
- What are the two **variables**?

Brand	Male	Female	All Runners
Nike	52	40	92
Adidas	35	14	49
Reebok	11	26	37
Asics	8	5	13
Other	6	3	9
Total	112	88	200

- Column percentages or row percentages can reveal **relationships (interactions)** between two variables

column percentages

Brand	Male	Female	All Runners
Nike	46.4	45.5	46.0
Adidas	31.3	45.5	33.5
Reebok	9.8	29.5	18.5
Asics	7.1	5.7	6.5
Other	5.4	3.4	4.5
Total	100.0	100.0	100.0

- Does market share vary by gender?
- Hormel Anti-Spam assignment
- Textbooks America assignment

17

---

---

---

---

---

---

---

---

---

---

---

---

## Scatter Plot

- Graphs the value of one **quantitative** variable vs. value of another **quantitative** variable
- Bivariate data
  - Y-axis plots the **dependent** (response) variable
  - X-axis plots the **independent** (explanatory) variable
- Can reveal relationships (interactions) between two **quantitative** variables
- A trend line can estimate the relation between variables
  - direct relation vs. inverse relation
- Amount of **scatter around the trend line** indicates **strength** of relationship

18

---

---

---

---

---

---

---

---

---

---

---

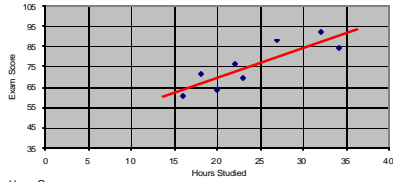
---



## Scatter Plot

hours score

20	64
16	61
34	84
23	70
27	88
32	92
18	72
22	77



- Do you detect a pattern?
- A scatter plot and trend line can help
- Absences and Score assignment

19

---

---

---

---

---

---

---

---