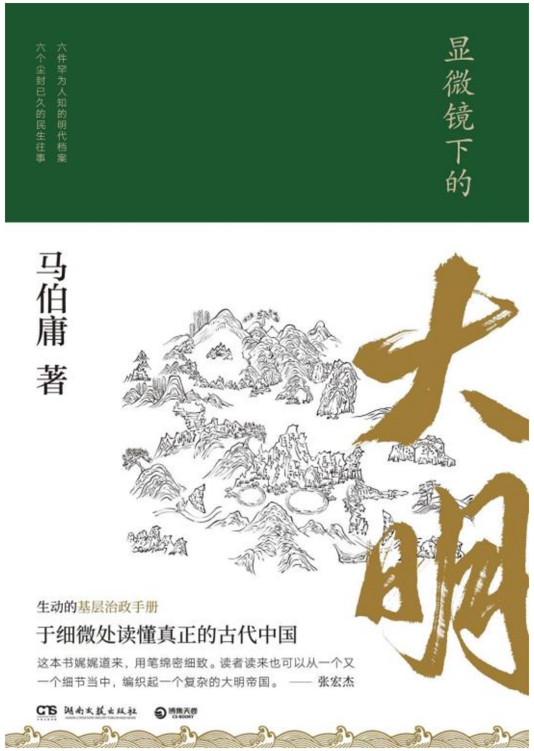
Statistics

PU Zhengning

What is Statistics

Chapter 1



显微镜下的

六件罕为人知的明代档案
六个尘封已久的民生往事

马伯庸 著

大明

生动的基层治政手册

于细微处读懂真正的古代中国

这本书娓娓道来，用笔绵密细致。读者读来也可以从一个又一个细节当中，编织起一个复杂的大明帝国。 —— 张宏杰

湖南文艺出版社 博集天卷

“…这一年，刘邦抢在其他诸侯之前杀入关中，兵临咸阳。秦三世子婴手捧玉玺出降，秦帝国彻底土崩瓦解。这群沛县穷汉进入大秦国都之后，立刻被其繁华富庶迷花了眼，纷纷冲进各处府库去抢金帛财宝。就连刘邦自己，也赖在秦宫里不愿意出来。这里多美好啊，有精致滑顺的帷帐，有名贵的萌犬和骏马，有琳琅满目的宝物，还有不计其数的美女。

在这场狂欢中，只有一个人保持着无比的冷静。他叫萧何。

跟那些出身市井的同僚相比，这位前沛县官员有着丰富的行政经验，他知道，对这个新生政权而言什么才是最重要的。

萧何穿过兴奋的人群和堆积如山的财货，来到大秦丞相、御史专属的档案库房。这里门庭冷落，因为里面没有珠宝金玉，只有天下诸郡县的户口版籍、土地图册、律令等文书，没人对这些写满枯燥数字的竹简有兴趣。

萧何下令将这些资料进行清查、分类，然后一一打包好…”

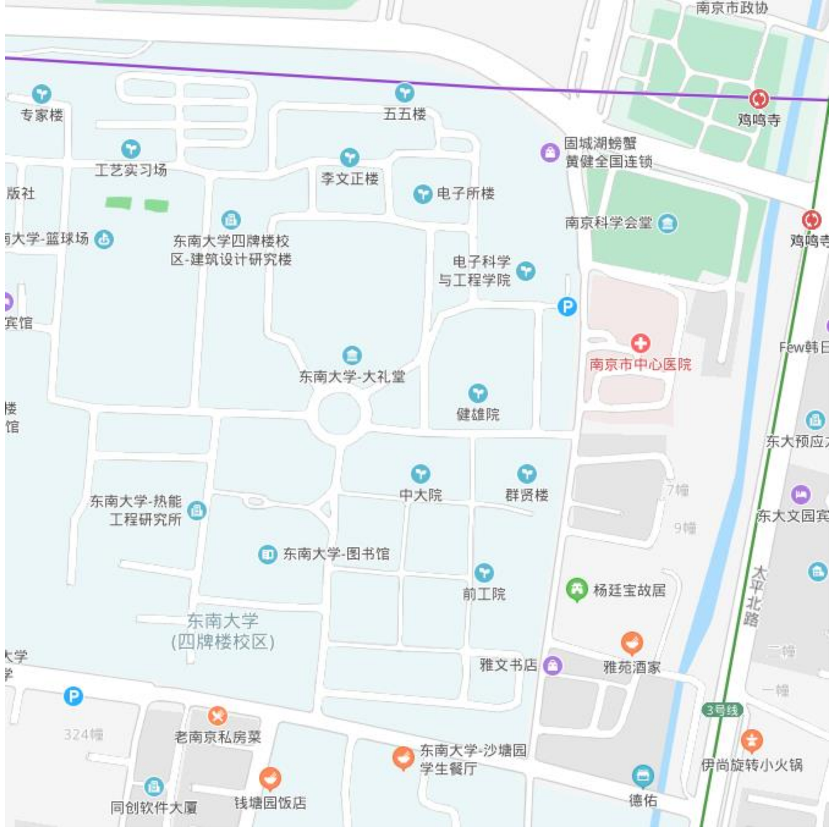
“…自商鞅以来，秦国的行政管理一向以绵密细致而著称，特别热衷于大数据。《商君书》里列举了国家兴盛需要掌握的十三类数据：官营粮仓、金库、壮年男子、女子、老人、儿童、官吏、士、纵横家、商人、马匹、牛，以及牲口草料。其中对于百姓数据的搜集，必须倚重户籍的建设与管理。

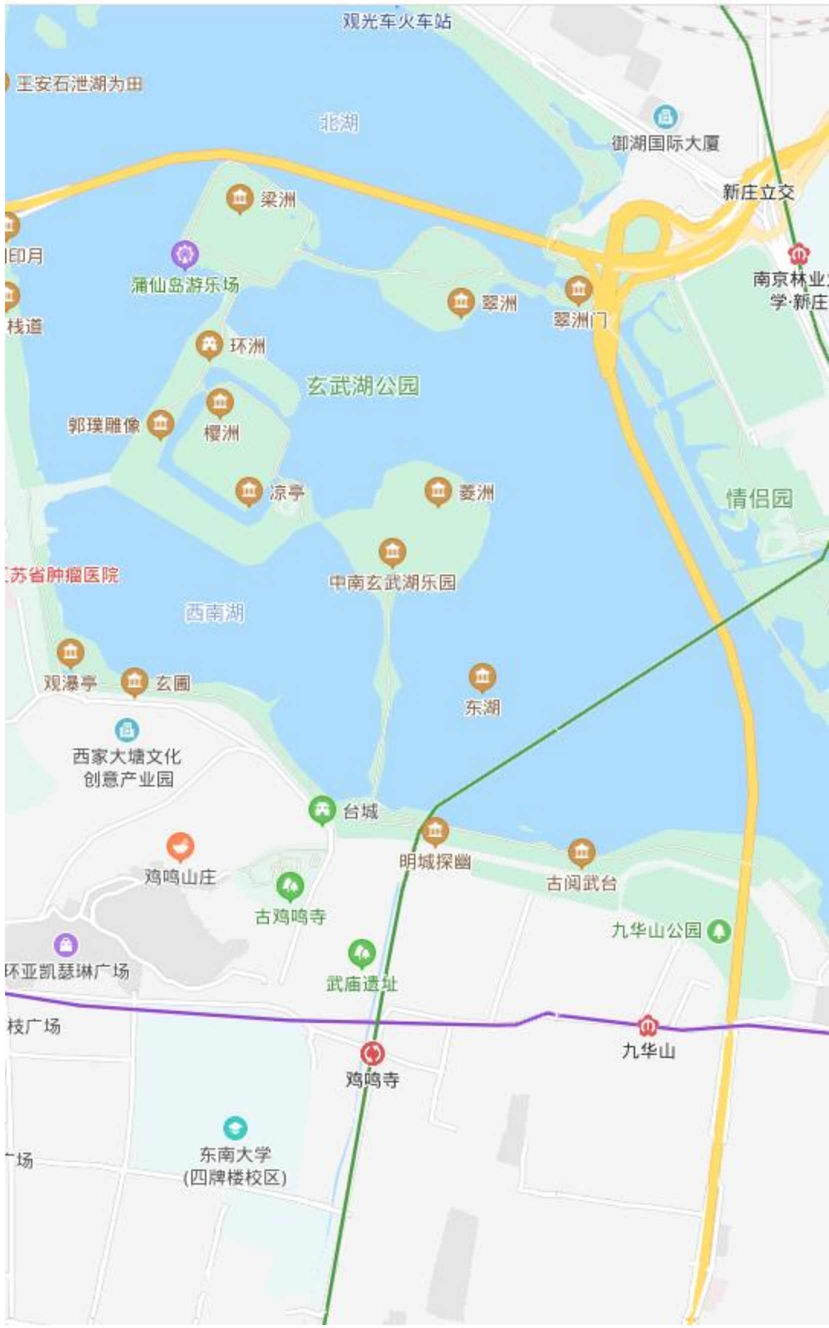
到了秦始皇时代，郡县制推行于全国。从一郡、一县到一乡、一里乃至每一户，官府都有详尽记录。你家里几口人，年纪多大，什么户籍类别，多高的爵位，何年何地迁来，何年傅籍，养几匹马、几头牛，耕种的地在哪儿、多大，种的什么作物，税要交多少，等等，记录得清清楚楚。

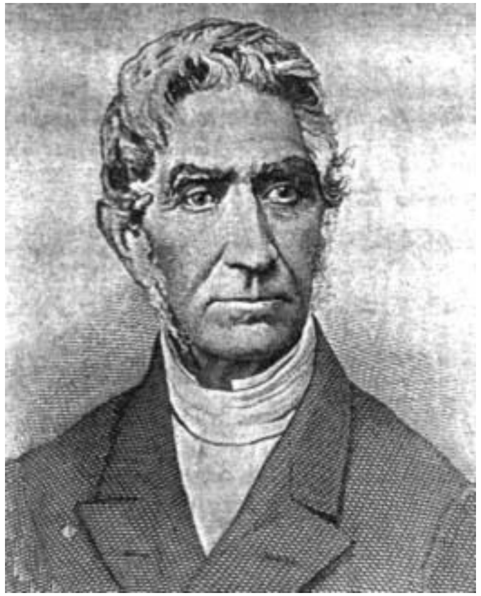
而且相关档案每年还要进行更新，由专门的上计人员送到咸阳留存，以便中央随时掌握地方情况。这套制度，在秦始皇时期一直保持着高效运转，到了秦二世时期，各地官府出于惯性也一直在执行。萧何在官府里当过主吏掾，对这些东西再熟悉不过…”

--马伯庸《显微镜下的大明》第五卷 天下透明 大明第一档案库的前世今生

南
北
©Baidu
©Baidu


南京市政协
专家楼
五五楼
鸡鸣寺
工艺实习场
李文正楼
固城湖螃蟹
黄健全国连锁
电子所楼
版社
南京科学会堂
鸡鸣寺
南大学-篮球场
东南大学四牌楼校
区-建筑设计研究楼
电子科学
与工程学院
宾馆
南京市中心医院
Few韩日
东南大学-大礼堂
健雄院
楼
馆
东大预应
中大院
群贤楼
7幢
东南大学-热能
工程研究所
东大文园宾
9幢
东南大学-图书馆
太平北路
杨廷宝故居
前工院
东南大学
(四牌楼校区)
大学
雅文书店
雅苑酒家
一幢
一幢
3号线
324幢
老南京私房菜
东南大学-沙塘园
学生餐厅
伊尚旋转小火锅
同创软件大厦
钱塘园饭店
德佑


观光车火车站
王安石泄湖为田
北湖
御湖国际大厦
新庄立交
梁洲
印月
南京林业大
学·新庄
蒲仙岛游乐场
翠洲
翠洲门
栈道
环洲
玄武湖公园
郭璞雕像
樱洲
凉亭
菱洲
情侣园
江苏省肿瘤医院
中南玄武湖乐园
西南湖
观瀑亭
玄圃
东湖
西家大塘文化
创意产业园
台城
明城探幽
鸡鸣山庄
古阅武台
古鸡鸣寺
九华山公园
环亚凯瑟琳广场
武庙遗址
枝广场
九华山
鸡鸣寺
场
东南大学
(四牌楼校区)

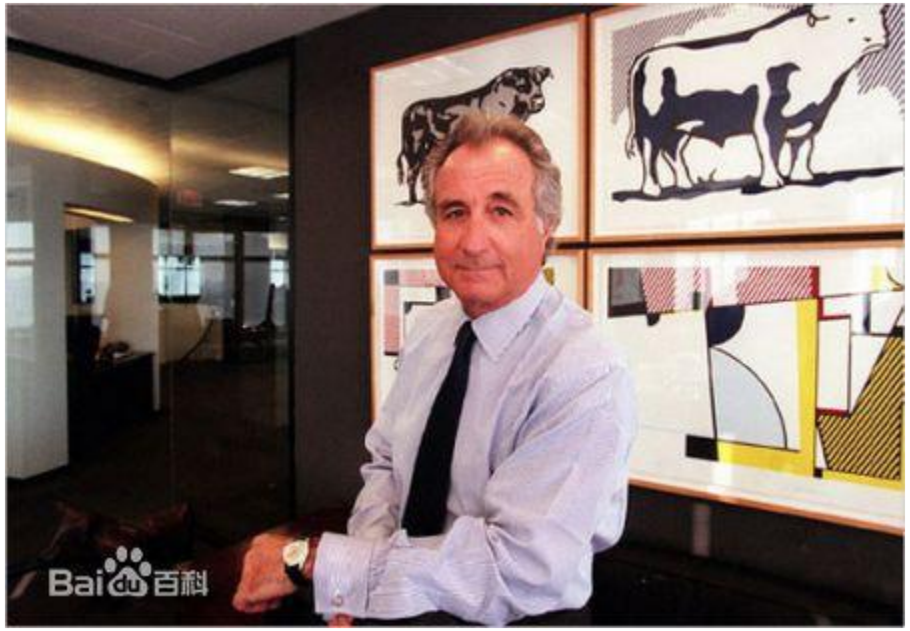


(Quetelet,1796—1874)

（John Snow，1813-1858）

（Karl Prarson,1857-1936）

（Ronald Fisher，1890-1962）



Pyramid scheme

GOALS

- Understand why we study statistics.
- Explain what is meant by descriptive statistics and inferential statistics.
- Distinguish between a qualitative variable and a quantitative variable.
- Describe how a discrete variable is different from a continuous variable.
- Distinguish among the nominal, ordinal, interval, and ratio levels of measurement.

What is Meant by Statistics?

Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting numerical data to assist in making more effective decisions.

Who Uses Statistics?

Statistical techniques are used extensively by marketing, accounting, quality control, consumers, professional sports people, hospital administrators, educators, politicians, physicians, etc...

Types of Statistics – Descriptive Statistics

Descriptive Statistics - methods of organizing, summarizing, and presenting data in an informative way.

EXAMPLE 1: A Gallup poll found that 49% of the people in a survey knew the name of the first book of the Bible. The statistic 49 describes the number out of every 100 persons who knew the answer.

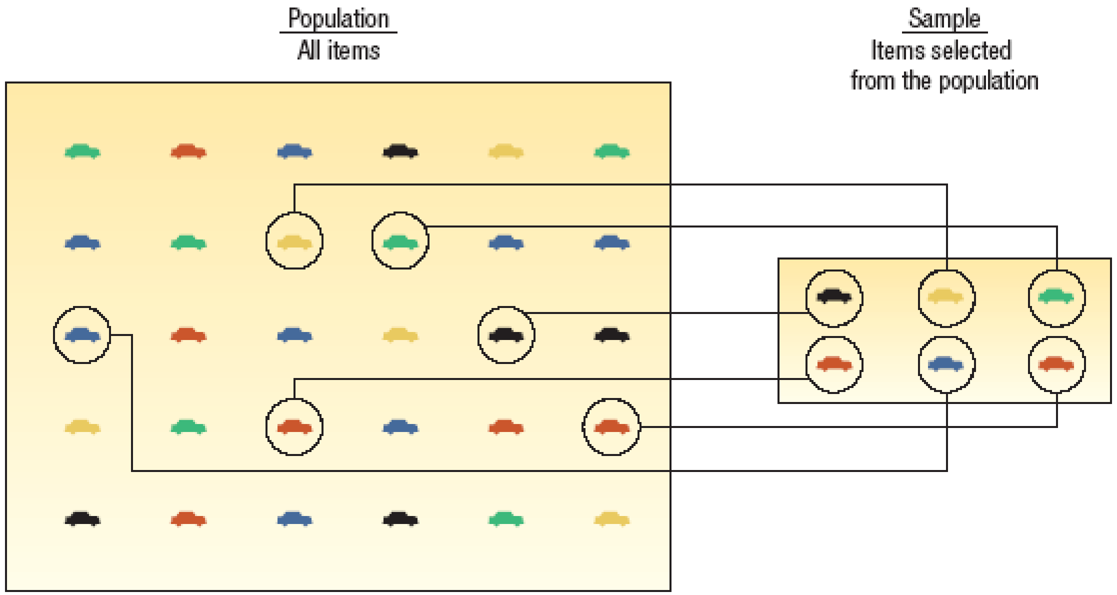
EXAMPLE 2: According to Consumer Reports, General Electric washing machine owners reported 9 problems per 100 machines during 2001. The statistic 9 describes the number of problems out of every 100 machines.

Inferential Statistics: A decision, estimate, prediction, or generalization about a population, based on a sample.

Population versus Sample

A population is a collection of all possible individuals, objects, or measurements of interest.

A sample is a portion, or part, of the population of interest



Types of Variables

A. Qualitative or Attribute variable - the characteristic being studied is nonnumeric.

EXAMPLES: Gender, religious affiliation, type of automobile owned, state of birth, eye color are examples.

B. Quantitative variable - information is reported numerically.

EXAMPLES: balance in your checking account, minutes remaining in class, or number of children in a family.

Quantitative Variables - Classifications

Quantitative variables can be classified as either discrete or continuous.

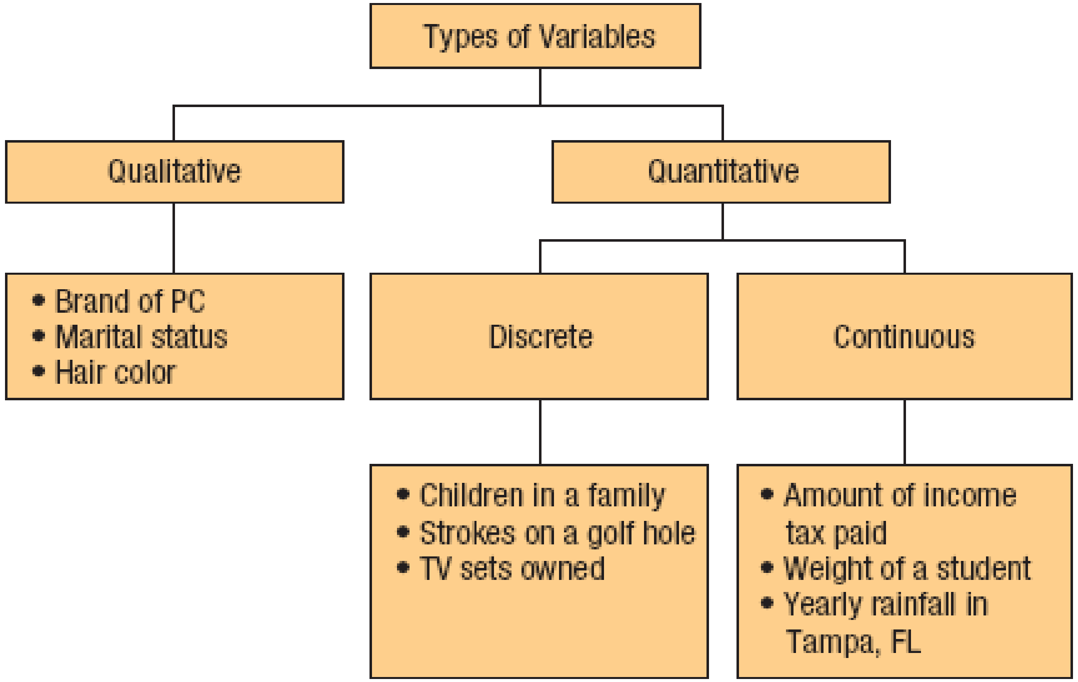
A. Discrete variables: can only assume certain values and there are usually "gaps" between values.

EXAMPLE: the number of bedrooms in a house, or the number of hammers sold at the local Home Depot (1,2,3,…,etc).

B. Continuous variable can assume any value within a specified range.

EXAMPLE: The pressure in a tire, the weight of a pork chop, or the height of students in a class.

Summary of Types of Variables



Types of Variables

Qualitative

- Brand of PC
- Marital status
- Hair color

Quantitative

Discrete

- Children in a family
- Strokes on a golf hole
- TV sets owned

Continuous

- Amount of income tax paid
- Weight of a student
- Yearly rainfall in Tampa, FL

CHART 1–2 Summary of the Types of Variables

Four Levels of Measurement

Nominal level - data that is classified into categories and cannot be arranged in any particular order.

EXAMPLES: eye color, gender, religious affiliation.

Ordinal level – involves data arranged in some order, but the differences between data values cannot be determined or are meaningless.

EXAMPLE: During a taste test of 4 soft drinks, Mellow Yellow was ranked number 1, Sprite number 2, Seven-up number 3, and Orange Crush number 4.

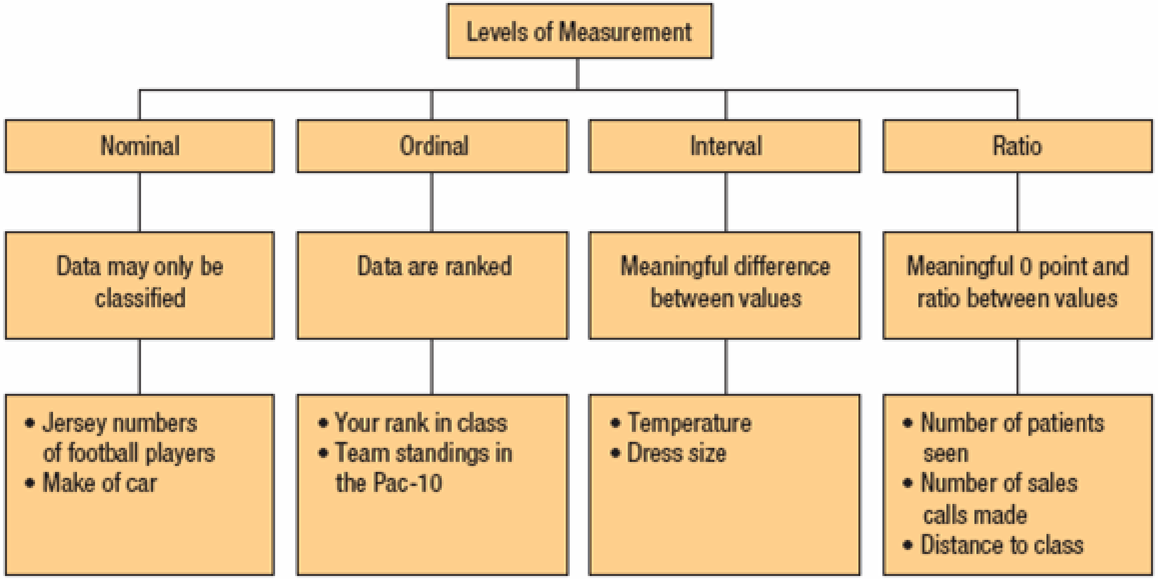
Interval level - similar to the ordinal level, with the additional property that meaningful amounts of differences between data values can be determined. There is no natural zero point.

EXAMPLE: Temperature on the Fahrenheit scale.

Ratio level - the interval level with an inherent zero starting point. Differences and ratios are meaningful for this level of measurement.

EXAMPLES: Monthly income of surgeons, or distance traveled by manufacturer's representatives per month.

Summary of the Characteristics for Levels of Measurement



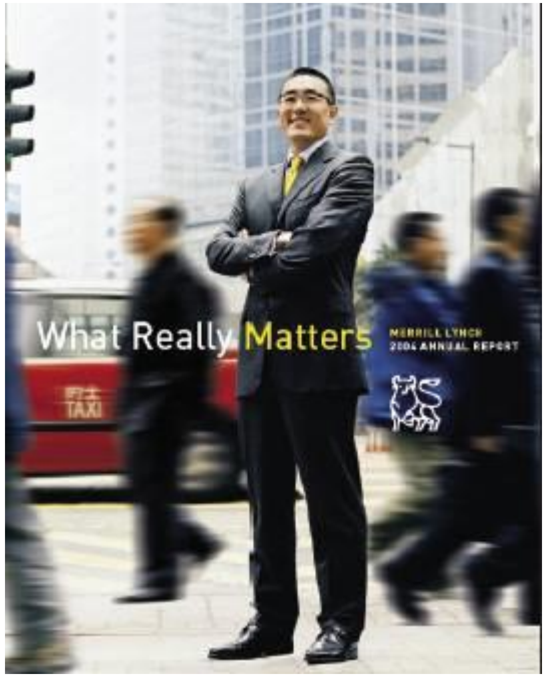
Levels of Measurement

Nominal	Ordinal	Interval	Ratio
Data may only be classified	Data are ranked	Meaningful difference between values	Meaningful 0 point and ratio between values
• Jersey numbers of football players • Make of car	• Your rank in class • Team standings in the Pac-10	• Temperature • Dress size	• Number of patients seen • Number of sales calls made • Distance to class

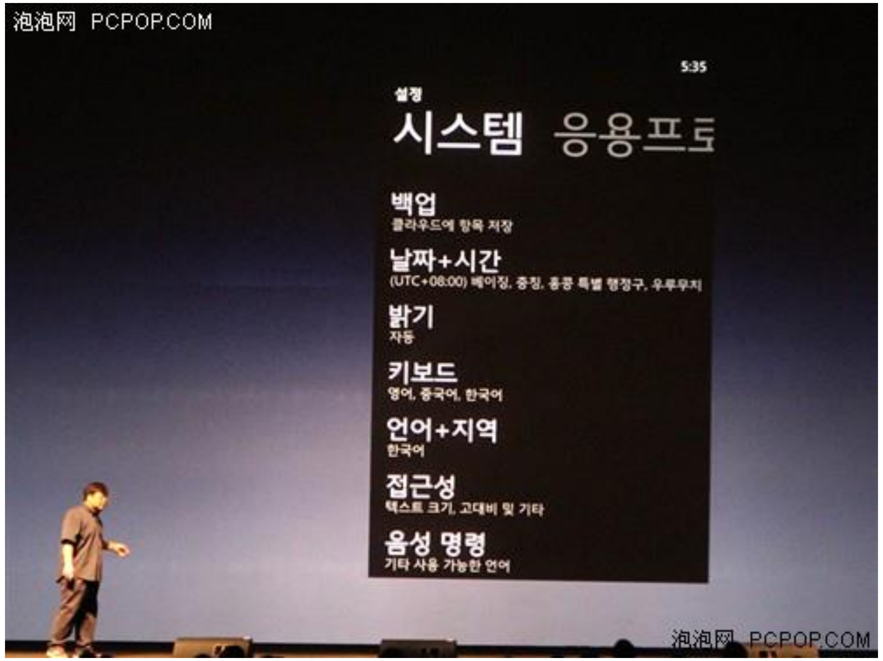
CHART 1–3 Summary of the Characteristics for Levels of Measurement

End of Chapter 1

Describing Data:
Frequency Tables, Frequency Distributions, and Graphic Presentation



Chapter 2



5:35

설정

시스템 응용프로

백업
클라우드에 항목 저장

날짜+시간
(UTC+08:00) 베이징, 충칭, 홍콩 특별 행정구, 우루무치

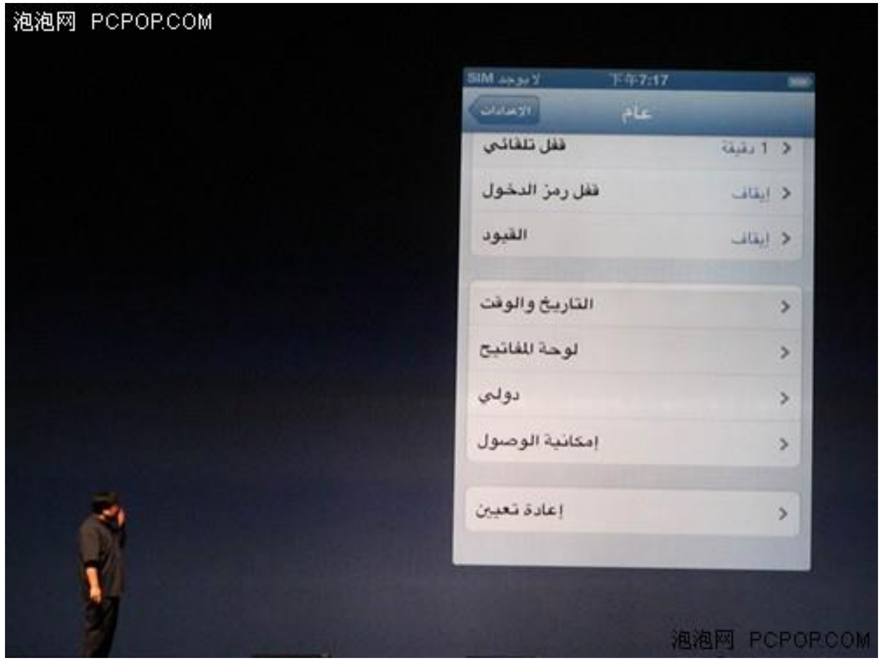
밝기
자동

키보드
영어, 중국어, 한국어

언어+지역
한국어

접근성
텍스트 크기, 고대비 및 기타

음성 명령
기타 사용 가능한 언어



لا يوجد SIM 下午7:17

الإعدادات عام

قفل تلقائي 1 دقيقة

قفل رمز الدخول إيقاف

القيود إيقاف

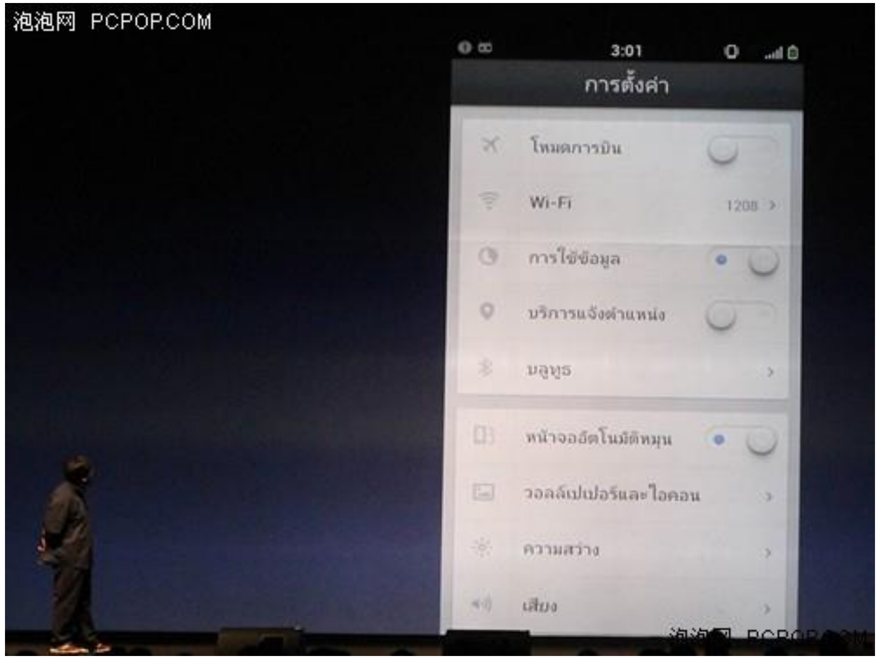
التاريخ والوقت

لوحة المفاتيح

دولي

إمكانية الوصول

إعادة تعيين



3:01

การตั้งค่า

โหมดการบิน

Wi-Fi 1208

การใช้ข้อมูล

บริการแจ้งตำแหน่ง

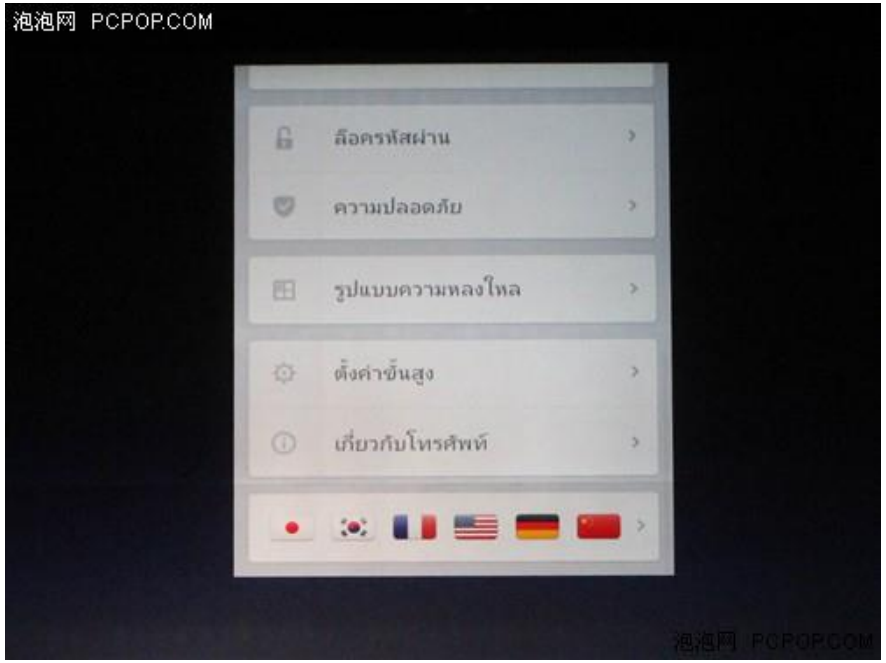
บลูทูธ

หน้าจออัตโนมัติหมุน

วอลล์เปเปอร์และไอคอน

ความสว่าง

เสียง



ล็อครหัสผ่าน

ความปลอดภัย

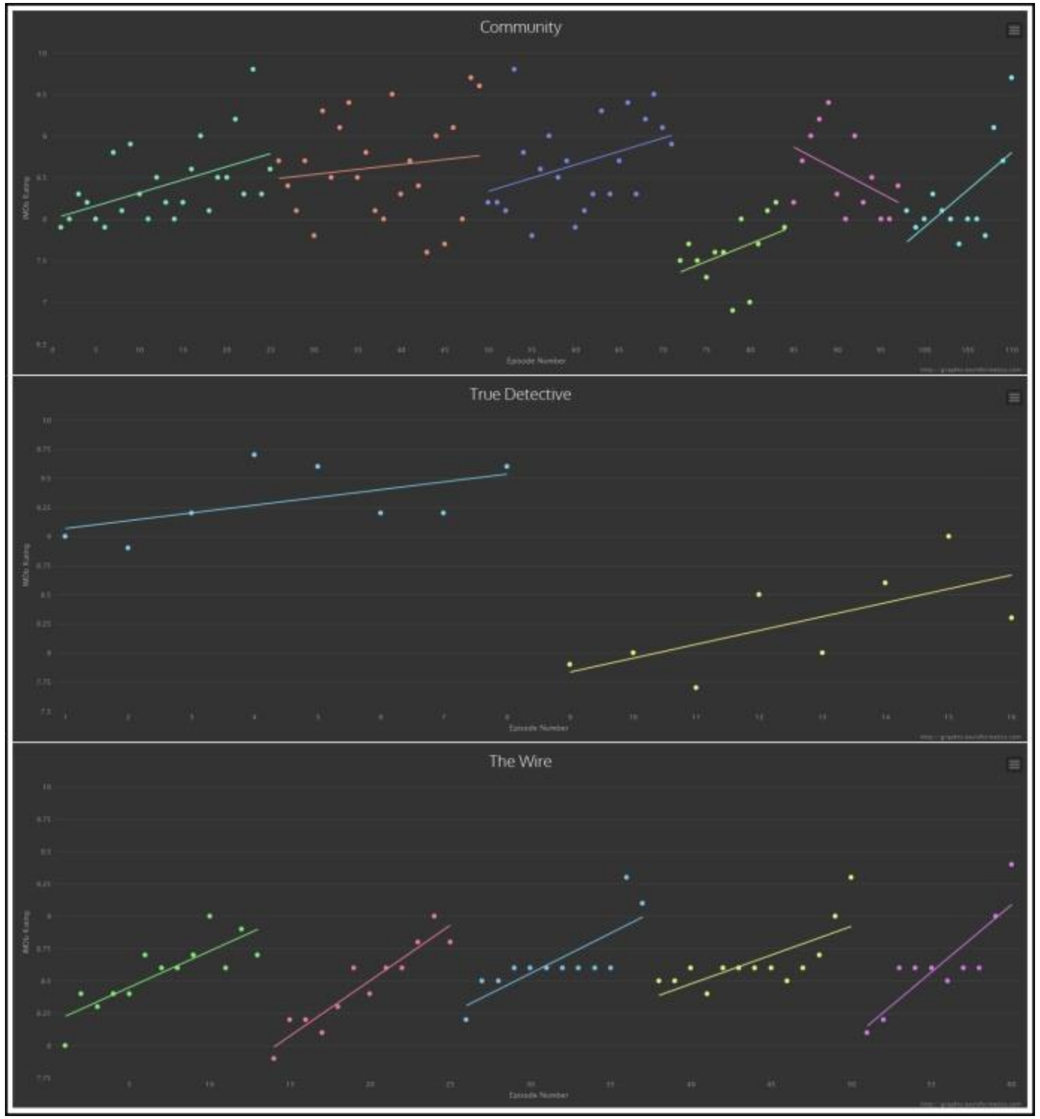
รูปแบบความหลงใหล

ตั้งค่าขั้นสูง

เกี่ยวกับโทรศัพท์

GOALS

- Organize qualitative data into a frequency table.
- Present a frequency table as a bar chart or a pie chart.
- Organize quantitative data into a frequency distribution.
- Present a frequency distribution for quantitative data using histograms, frequency polygons, and cumulative frequency polygons.



http://graphtv.kevinformatics.com/

Ms. Kathryn Ball of AutoUSA wants to develop tables, charts, and graphs to show the typical selling price on various dealer lots. The table on the right reports only the price of the 80 vehicles sold last month at Whitner Autoplex.

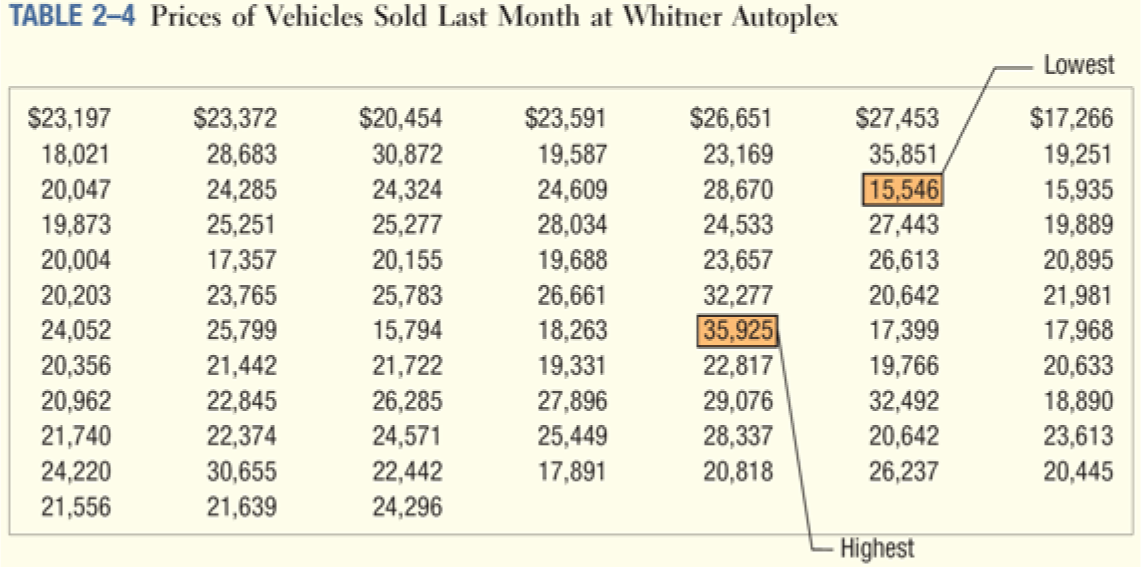


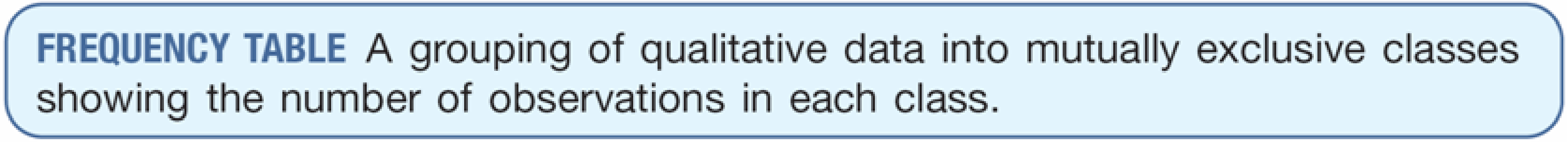
TABLE 2–4 Prices of Vehicles Sold Last Month at Whitner Autoplex

$23,197	$23,372	$20,454	$23,591	$26,651	$27,453	$17,266
18,021	28,683	30,872	19,587	23,169	35,851	19,251
20,047	24,285	24,324	24,609	28,670	15,546	15,935
19,873	25,251	25,277	28,034	24,533	27,443	19,889
20,004	17,357	20,155	19,688	23,657	26,613	20,895
20,203	23,765	25,783	26,661	32,277	20,642	21,981
24,052	25,799	15,794	18,263	35,925	17,399	17,968
20,356	21,442	21,722	19,331	22,817	19,766	20,633
20,962	22,845	26,285	27,896	29,076	32,492	18,890
21,740	22,374	24,571	25,449	28,337	20,642	23,613
24,220	30,655	22,442	17,891	20,818	26,237	20,445
21,556	21,639	24,296				

Lowest

Highest

Frequency Table



FREQUENCY TABLE A grouping of qualitative data into mutually exclusive classes showing the number of observations in each class.

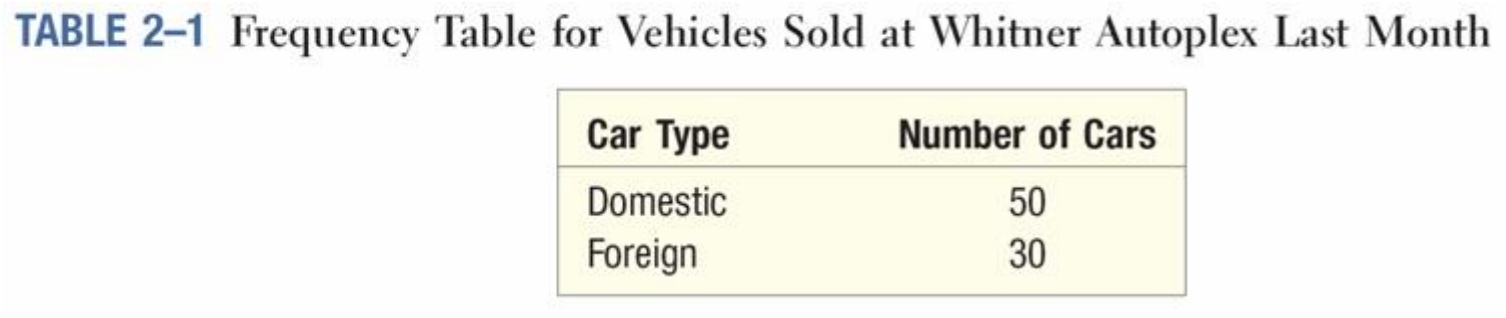


TABLE 2–1 Frequency Table for Vehicles Sold at Whitner Autoplex Last Month

Car Type	Number of Cars
Domestic	50
Foreign	30

Relative Class Frequencies

- Class frequencies can be converted to **relative class frequencies** to show the fraction of the total number of observations in each class.
- A relative frequency captures the relationship between a class total and the total number of observations.

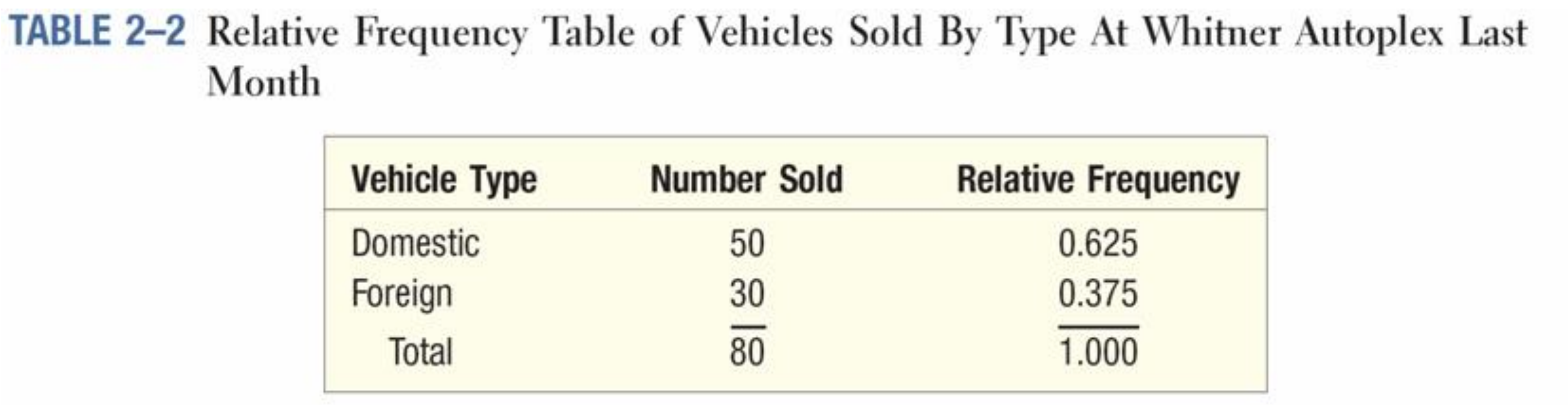


TABLE 2–2 Relative Frequency Table of Vehicles Sold By Type At Whitner Autoplex Last Month

Vehicle Type	Number Sold	Relative Frequency
Domestic	50	0.625
Foreign	30	0.375
Total	80	1.000

Bar Charts

BAR CHART A graph in which the classes are reported on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are proportional to the heights of the bars.

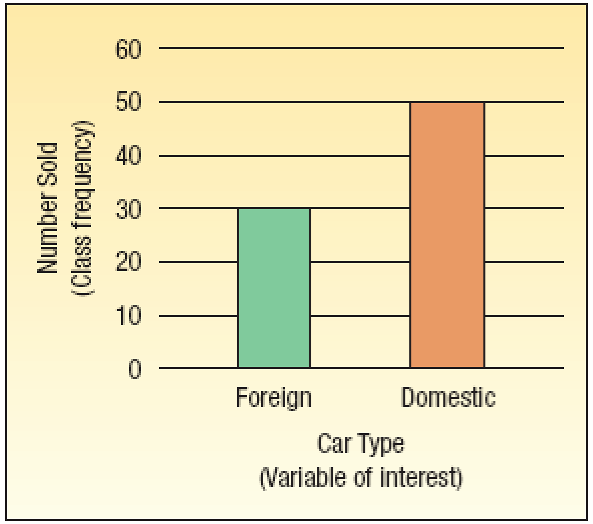


CHART 2–1 Vehicle Sold by Type Last Month At Whitner Autoplex

Pie Charts

PIE CHART A chart that shows the proportion or percent that each class represents of the total number of frequencies.

Use of Sales	Amount ($ million)	Percent of Share
Prizes	1,276.0	59
Payments to Education	648.1	30
Bonuses/Commissions	132.8	6
Operating Expenses	97.7	5
Total	2,154.6	100

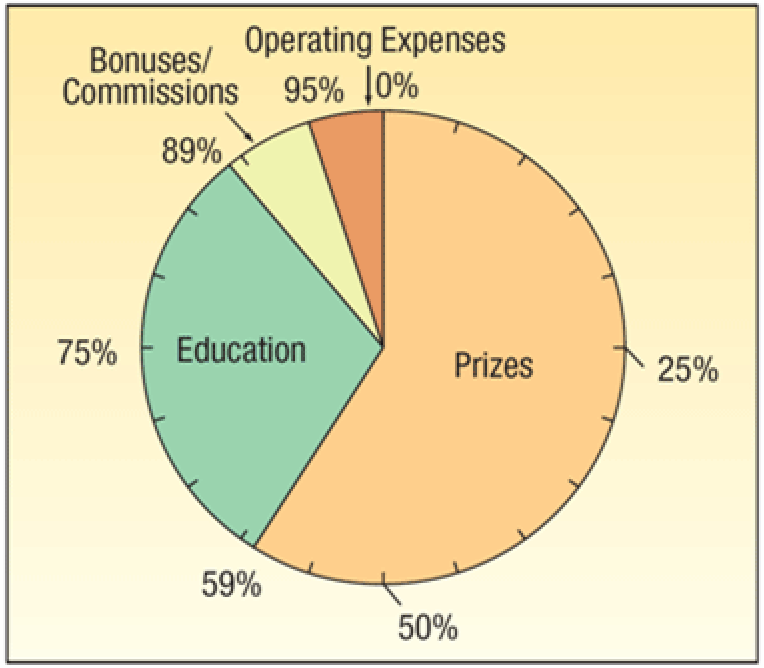
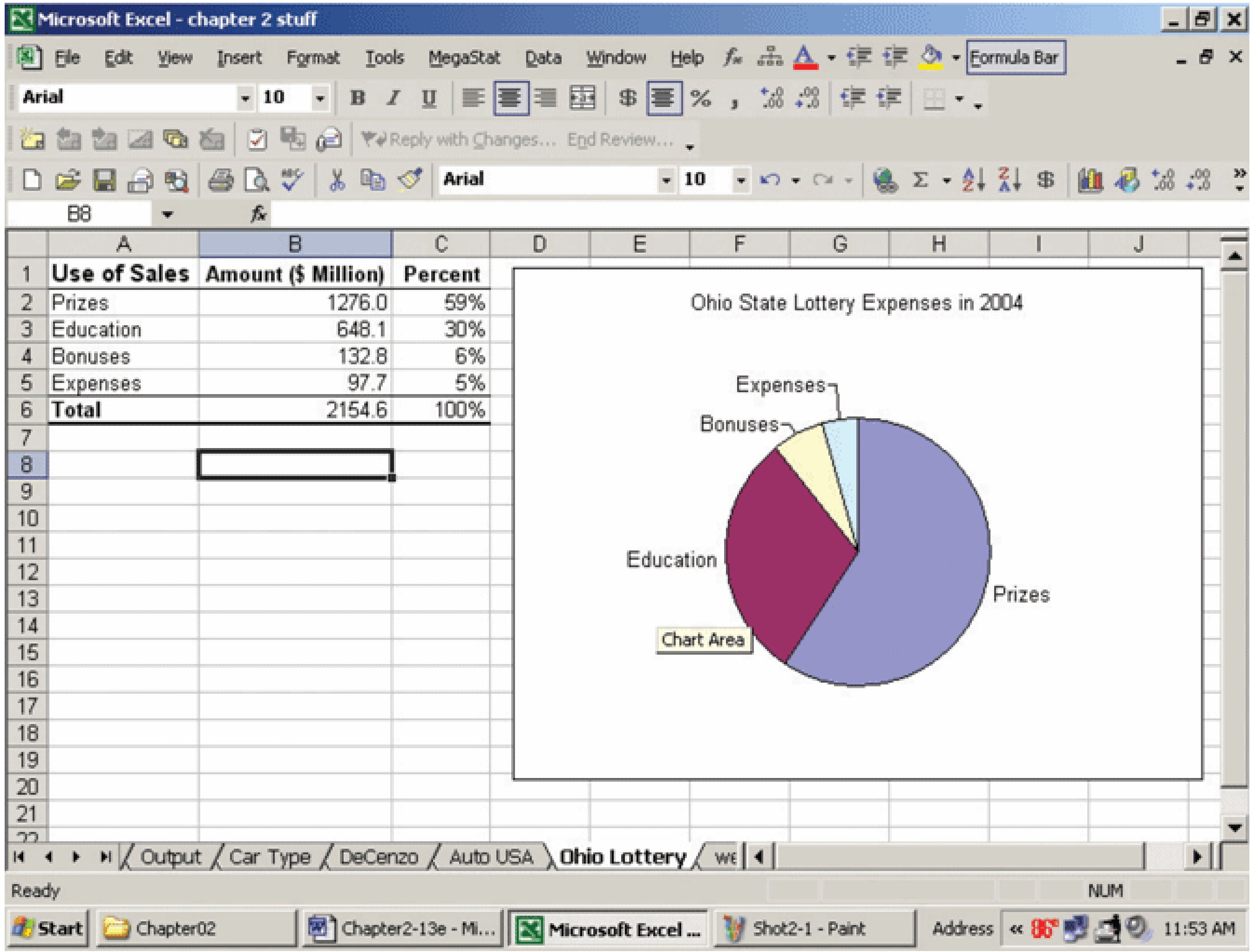


CHART 2–2 Pie Chart of Ohio Lottery Expenses in 2004

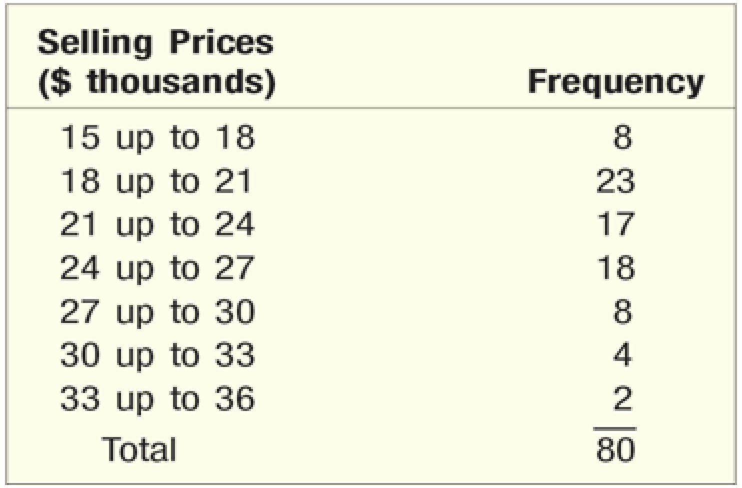
Pie Chart Using Excel



Use of Sales	Amount ($ Million)	Percent
Prizes	1276.0	59%
Education	648.1	30%
Bonuses	132.8	6%
Expenses	97.7	5%
Total	2154.6	100%

Microsoft Excel - chapter 2 stuff
File Edit View Insert Format Tools MegaStat Data Window Help
Formula Bar
Arial 10
Reply with Changes... End Review...
B8
Ohio State Lottery Expenses in 2004
Expenses
Bonuses
Education
Prizes
Chart Area
Output Car Type DeCenzo Auto USA Ohio Lottery
Ready
NUM
Start Chapter02 Chapter2-13e - Mi... Microsoft Excel ... Shot2-1 - Paint
Address
11:53 AM

Frequency Distribution



Selling Prices ($ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	80

A Frequency distribution is a grouping of data into mutually exclusive categories showing the number of observations in each class.

Frequency Distribution

Class midpoint: A point that divides a class into two equal parts. This is the average of the upper and lower class limits.

Class frequency: The number of observations in each class.

Class interval: The class interval is obtained by subtracting the lower limit of a class from the lower limit of the next class.

EXAMPLE – Creating a Frequency Distribution Table

Ms. Kathryn Ball of AutoUSA wants to develop tables, charts, and graphs to show the typical selling price on various dealer lots. The table on the right reports only the price of the 80 vehicles sold last month at Whitner Autoplex.

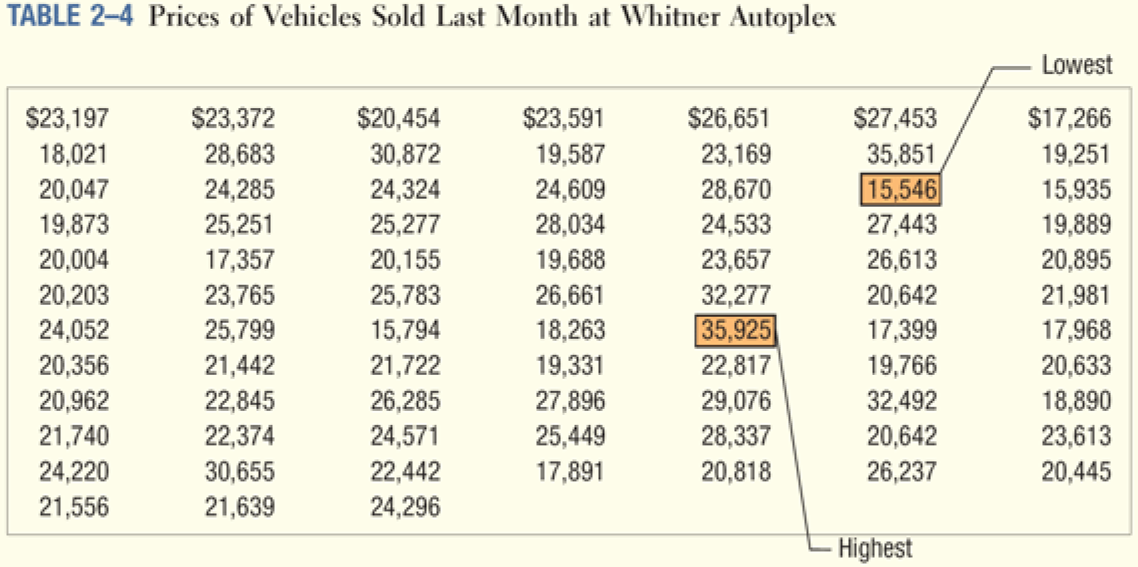


TABLE 2–4 Prices of Vehicles Sold Last Month at Whitner Autoplex

Lowest

$23,197	$23,372	$20,454	$23,591	$26,651	$27,453	$17,266
18,021	28,683	30,872	19,587	23,169	35,851	19,251
20,047	24,285	24,324	24,609	28,670	15,546	15,935
19,873	25,251	25,277	28,034	24,533	27,443	19,889
20,004	17,357	20,155	19,688	23,657	26,613	20,895
20,203	23,765	25,783	26,661	32,277	20,642	21,981
24,052	25,799	15,794	18,263	35,925	17,399	17,968
20,356	21,442	21,722	19,331	22,817	19,766	20,633
20,962	22,845	26,285	27,896	29,076	32,492	18,890
21,740	22,374	24,571	25,449	28,337	20,642	23,613
24,220	30,655	22,442	17,891	20,818	26,237	20,445
21,556	21,639	24,296				

Highest

Constructing a Frequency Table - Example

- **Step 1: Decide on the number of classes.**

  A useful recipe to determine the number of classes (k) is the "2 to the k rule." such that $2^k > n$.

  There were 80 vehicles sold. So $n = 80$. If we try $k = 6$, which means we would use 6 classes, then $2^6 = 64$, somewhat less than 80. Hence, 6 is not enough classes. If we let $k = 7$, then 2^7 128, which is greater than 80. So the recommended number of classes is 7.

- **Step 2: Determine the class interval or width.**

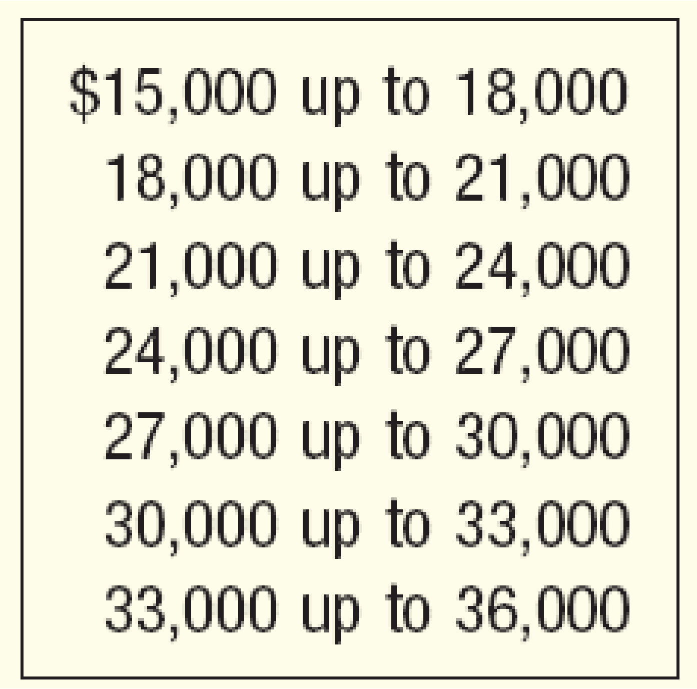
  The formula is: $i \geq (H-L)/k$ where i is the class interval, H is the highest observed value, L is the lowest observed value, and k is the number of classes.

  ($35,925 - $15,546)/7 = $2,911

  Round up to some convenient number, such as a multiple of 10 or 100. Use a class width of $3,000

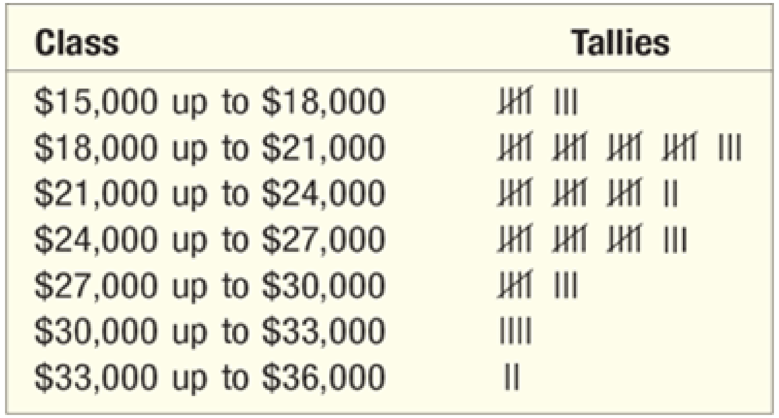
Constructing a Frequency Table - Example

- **Step 3: Set the individual class limits**

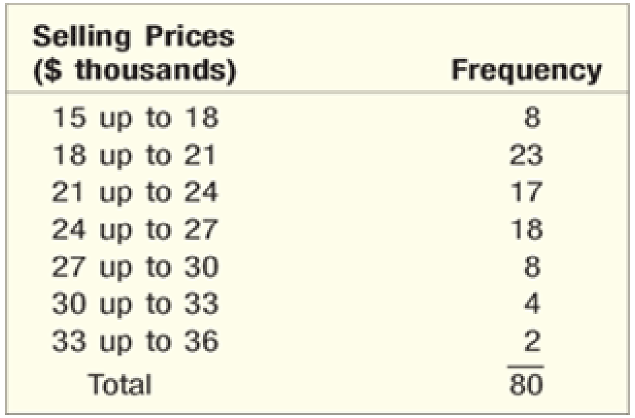


$15,000 up to 18,000
18,000 up to 21,000
21,000 up to 24,000
24,000 up to 27,000
27,000 up to 30,000
30,000 up to 33,000
33,000 up to 36,000

Constructing a Frequency Table



Class	Tallies
\$15,000 up to \$18,000	卌 III
\$18,000 up to \$21,000	卌 卌 卌 卌 III
\$21,000 up to \$24,000	卌 卌 卌 II
\$24,000 up to \$27,000	卌 卌 卌 III
\$27,000 up to \$30,000	卌 III
\$30,000 up to \$33,000	IIII
\$33,000 up to \$36,000	II



Selling Prices (\$ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	80

- Step 4: Tally the vehicle selling prices into the classes.

- Step 5: Count the number of items in each class.

Relative Frequency Distribution

To convert a frequency distribution to a *relative* frequency distribution, each of the class frequencies is divided by the total number of observations.

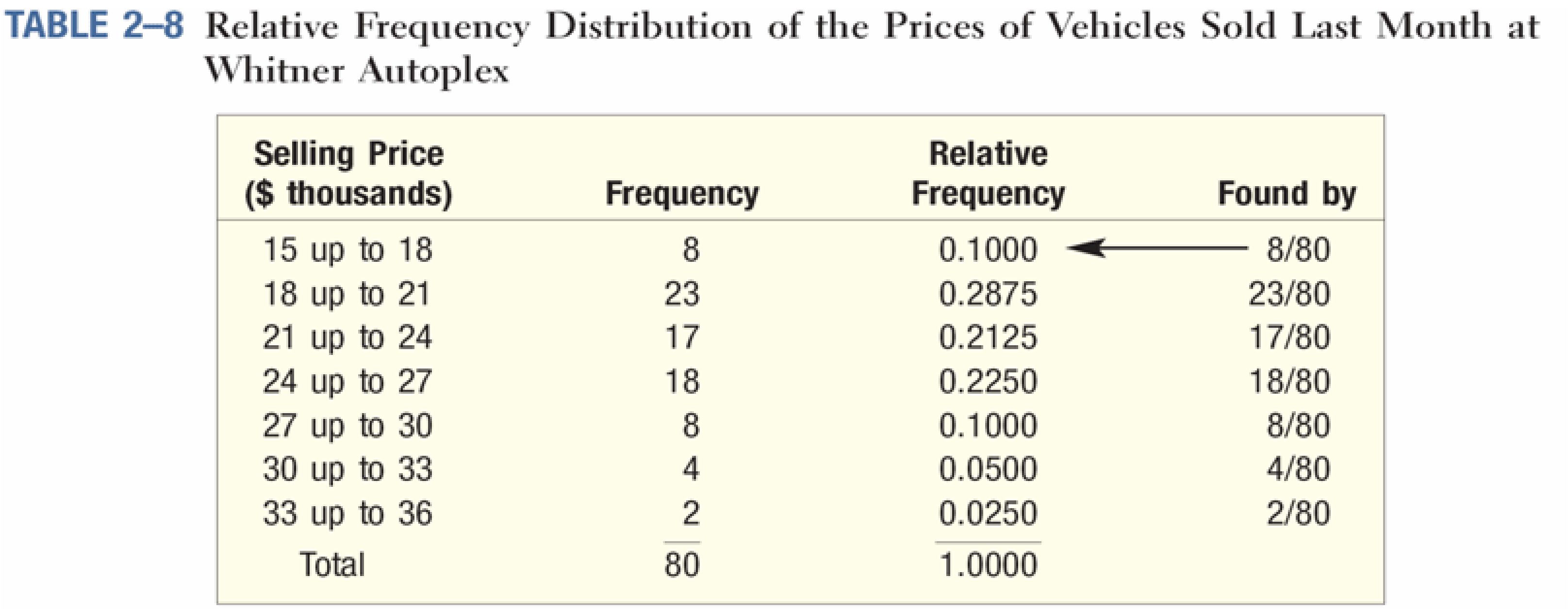


TABLE 2–8 Relative Frequency Distribution of the Prices of Vehicles Sold Last Month at Whitner Autoplex

Selling Price ($ thousands)	Frequency	Relative Frequency	Found by
15 up to 18	8	0.1000	8/80
18 up to 21	23	0.2875	23/80
21 up to 24	17	0.2125	17/80
24 up to 27	18	0.2250	18/80
27 up to 30	8	0.1000	8/80
30 up to 33	4	0.0500	4/80
33 up to 36	2	0.0250	2/80
Total	80	1.0000	

Graphic Presentation of a Frequency Distribution

The three commonly used graphic forms are:

- Histograms
- Frequency polygons
- Cumulative frequency distributions

Histogram

Histogram for a frequency distribution based on quantitative data is very similar to the bar chart showing the distribution of qualitative data. The classes are marked on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are represented by the heights of the bars.

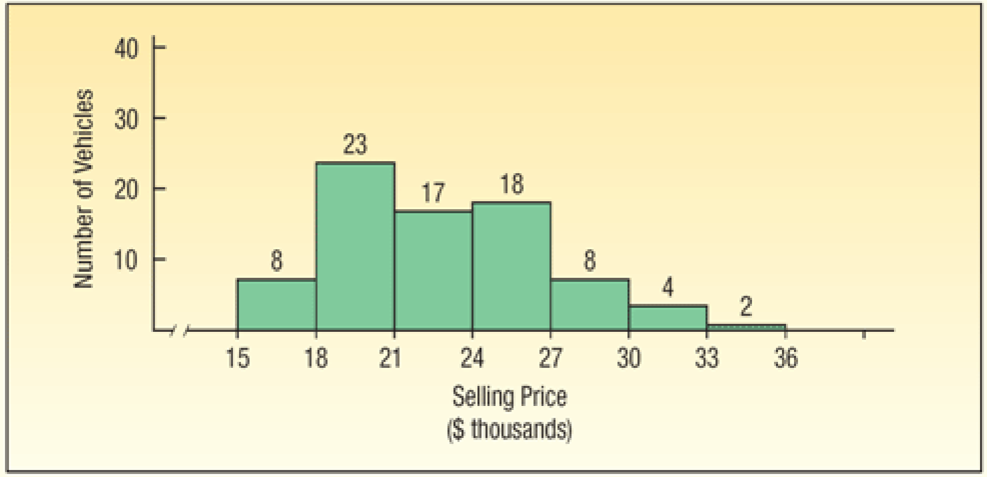
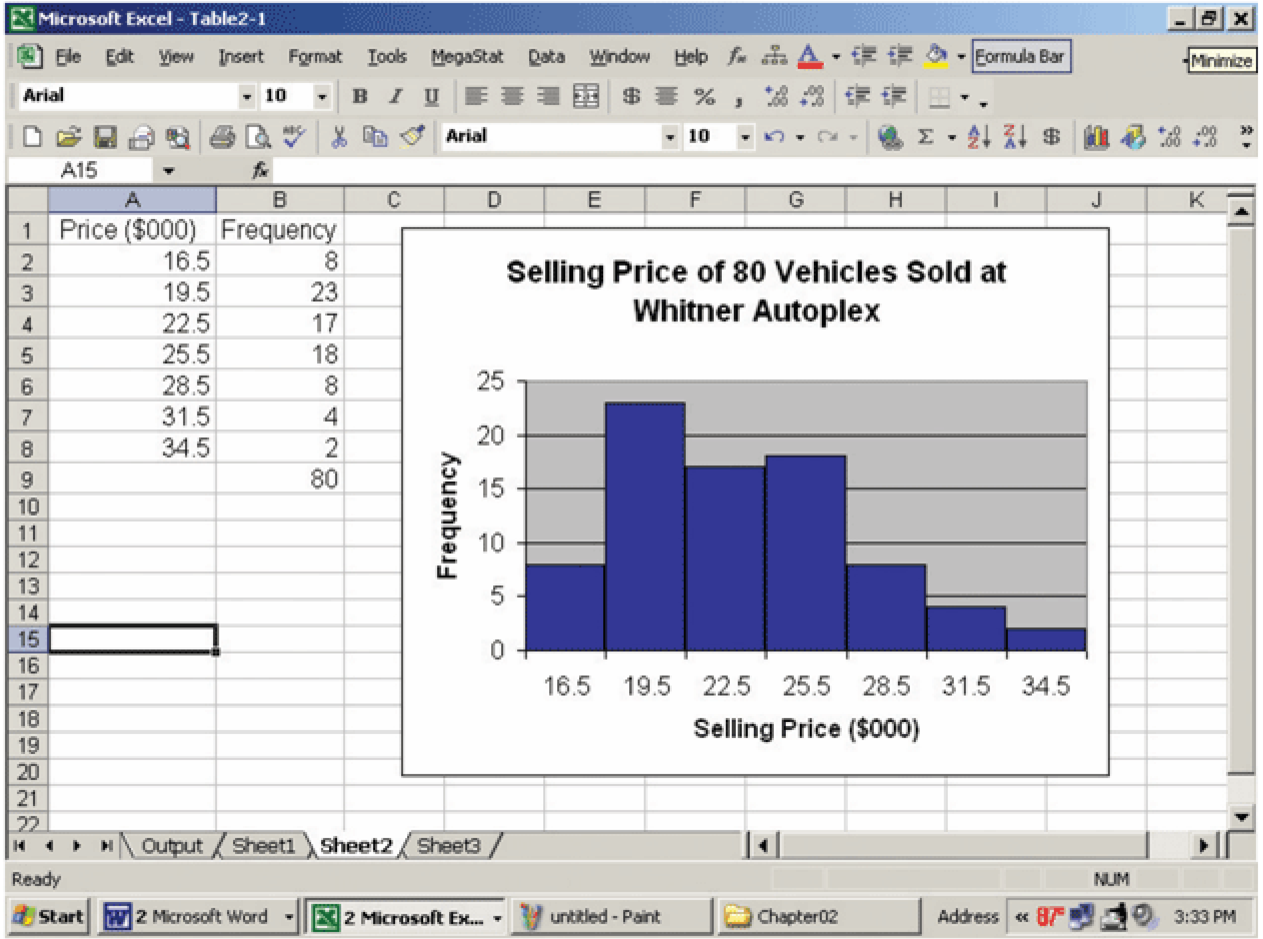
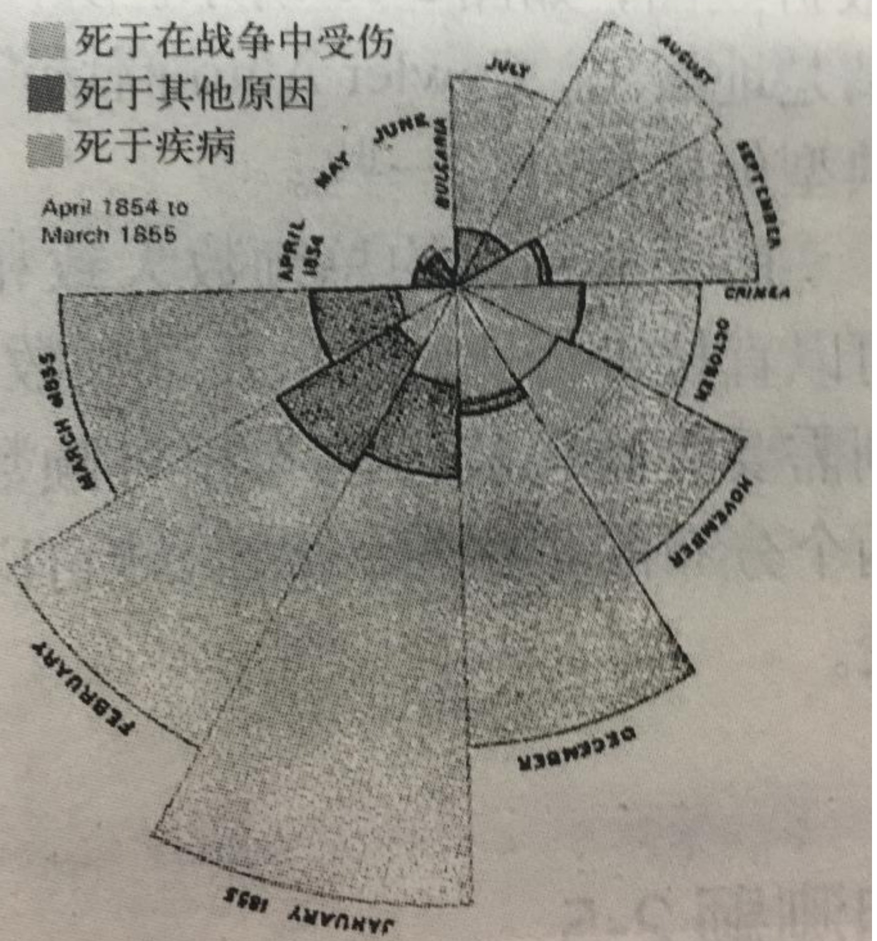


CHART 2–4 Histogram of the Selling Prices of 80 Vehicles at Whitner Autoplex

Histogram Using Excel

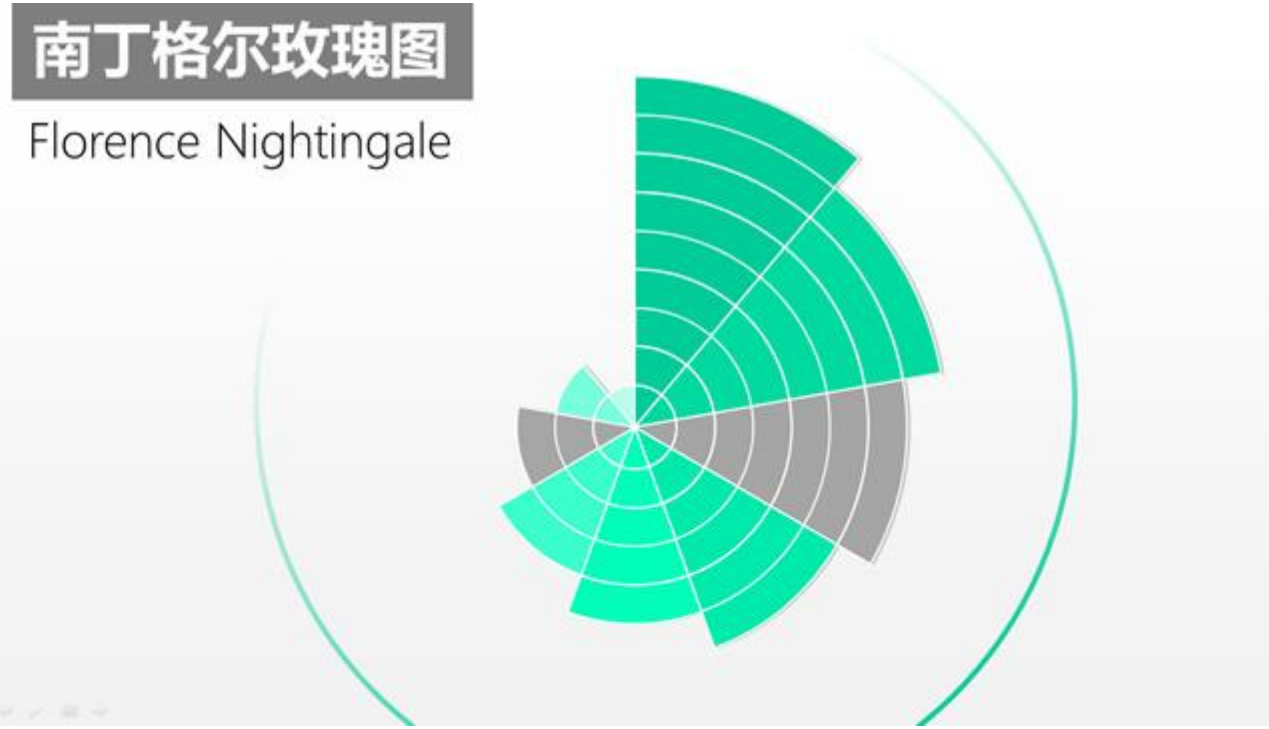


Price ($000)	Frequency
16.5	8
19.5	23
22.5	17
25.5	18
28.5	8
31.5	4
34.5	2
	80



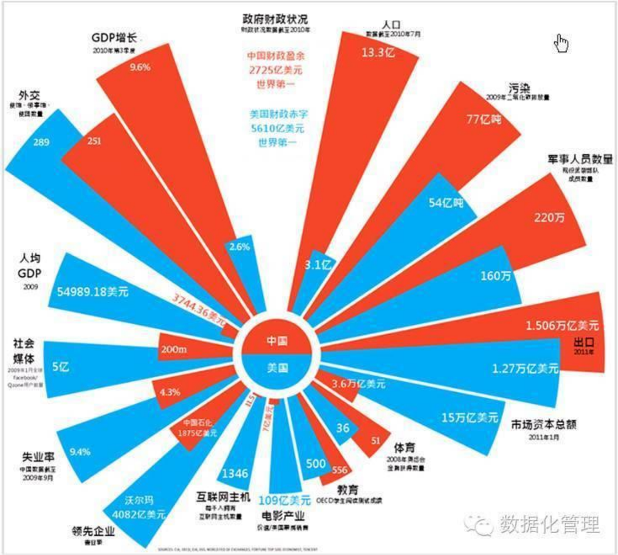
死于在战争中受伤
死于其他原因
死于疾病
April 1854 to
March 1855
APRIL 1854
MAY
JUNE
BULGARIA
JULY
AUGUST
SEPTEMBER
CRIMEA
OCTOBER
NOVEMBER
DECEMBER
JANUARY 1855
FEBRUARY
MARCH 1855

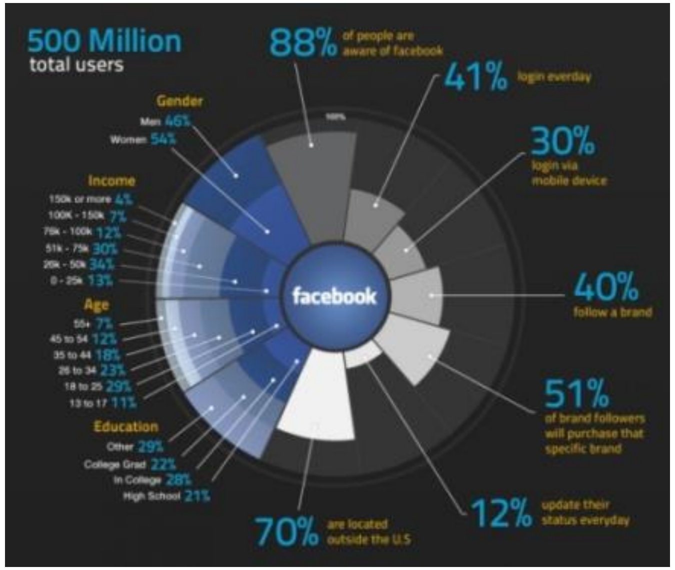
图 2-11



南丁格尔玫瑰图

Florence Nightingale

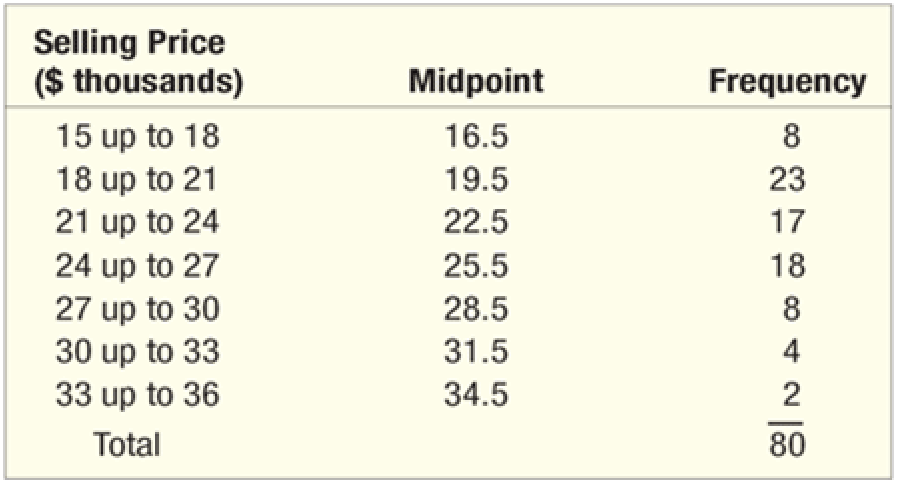




Frequency Polygon

- A **frequency polygon** also shows the shape of a distribution and is similar to a histogram.

- It consists of line segments connecting the points formed by the intersections of the class midpoints and the class frequencies.



Selling Price ($ thousands)	Midpoint	Frequency
15 up to 18	16.5	8
18 up to 21	19.5	23
21 up to 24	22.5	17
24 up to 27	25.5	18
27 up to 30	28.5	8
30 up to 33	31.5	4
33 up to 36	34.5	2
Total		80

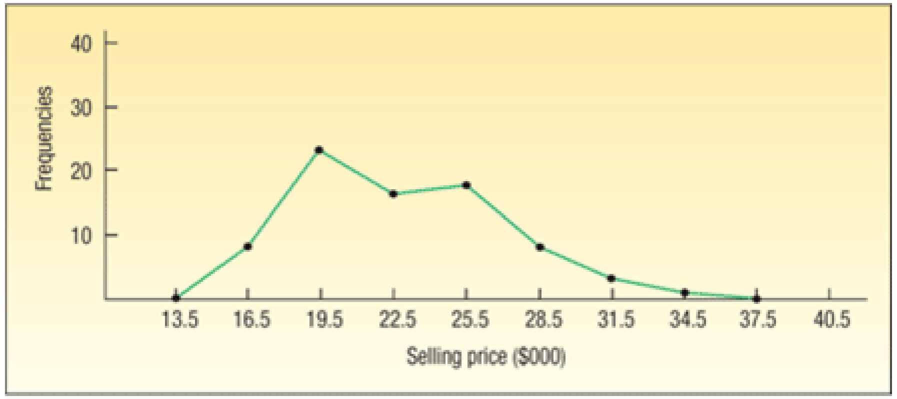


CHART 2–5 Frequency Polygon of the Selling Prices of 80 Vehicles at Whitner Autoplex

Cumulative Frequency Distribution

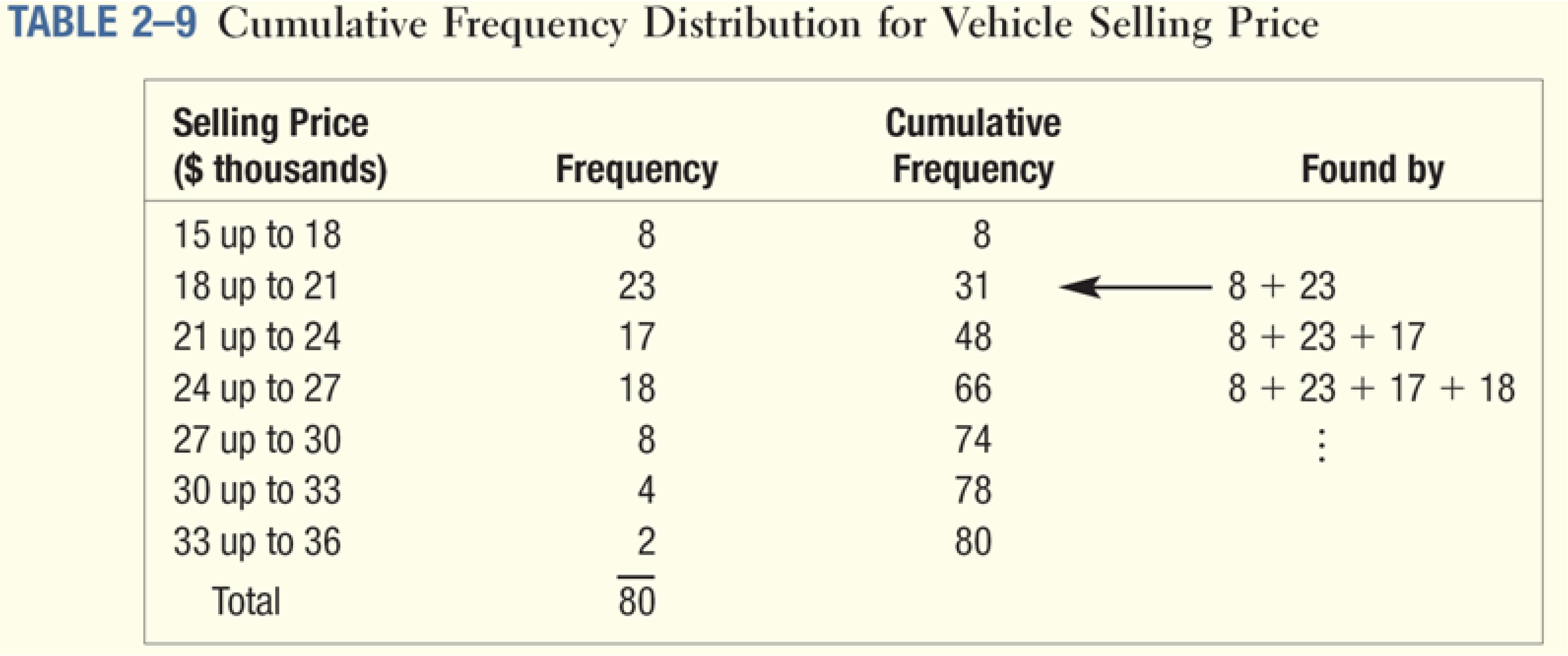


TABLE 2–9 Cumulative Frequency Distribution for Vehicle Selling Price

Selling Price ($ thousands)	Frequency	Cumulative Frequency	Found by
15 up to 18	8	8	
18 up to 21	23	31	← 8 + 23
21 up to 24	17	48	8 + 23 + 17
24 up to 27	18	66	8 + 23 + 17 + 18
27 up to 30	8	74	⋮
30 up to 33	4	78	
33 up to 36	2	80	
Total	80		

Cumulative Frequency Distribution

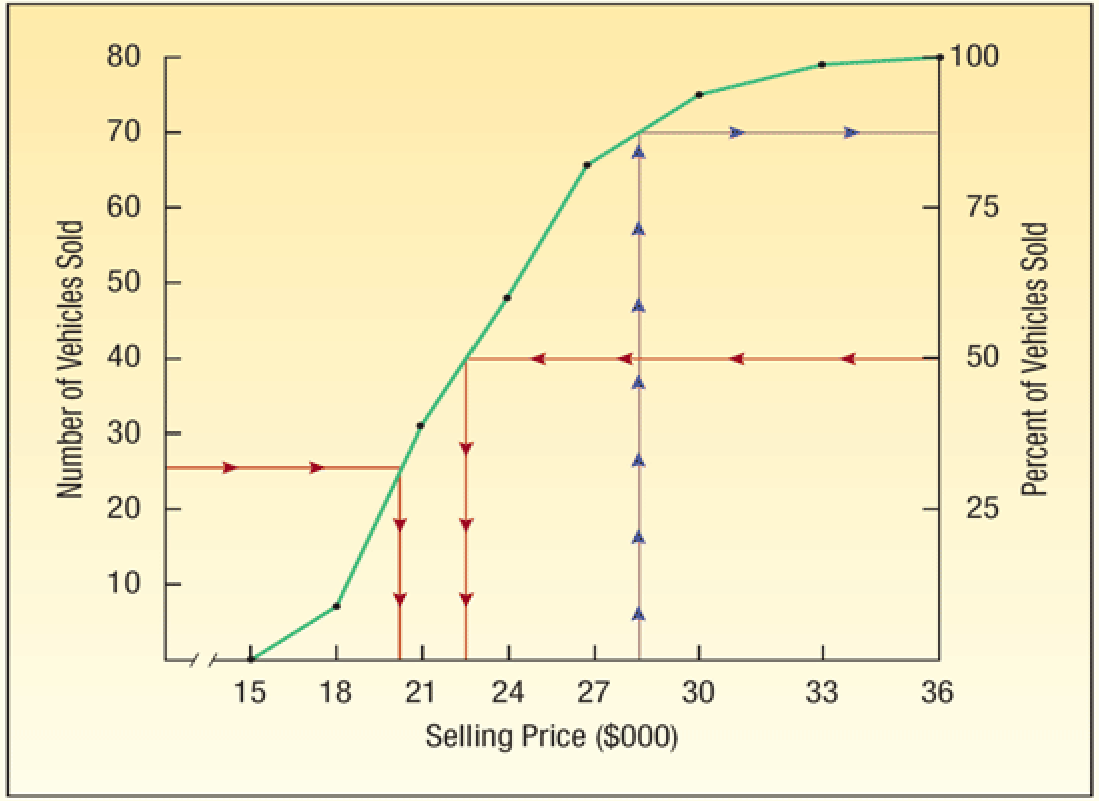


CHART 2–7 Cumulative Frequency Distribution for Vehicle Selling Price

End of Chapter 2

Describing Data:
Numerical Measures

Chapter 3

GOALS

- Calculate the arithmetic mean, weighted mean, median, mode, and geometric mean.
- Explain the characteristics, uses, advantages, and disadvantages of each measure of location.
- Identify the position of the mean, median, and mode for both symmetric and skewed distributions.
- Compute and interpret the range, mean deviation, variance, and standard deviation.
- Understand the characteristics, uses, advantages, and disadvantages of each measure of dispersion.
- Understand Chebyshev’s theorem and the Empirical Rule as they relate to a set of observations.

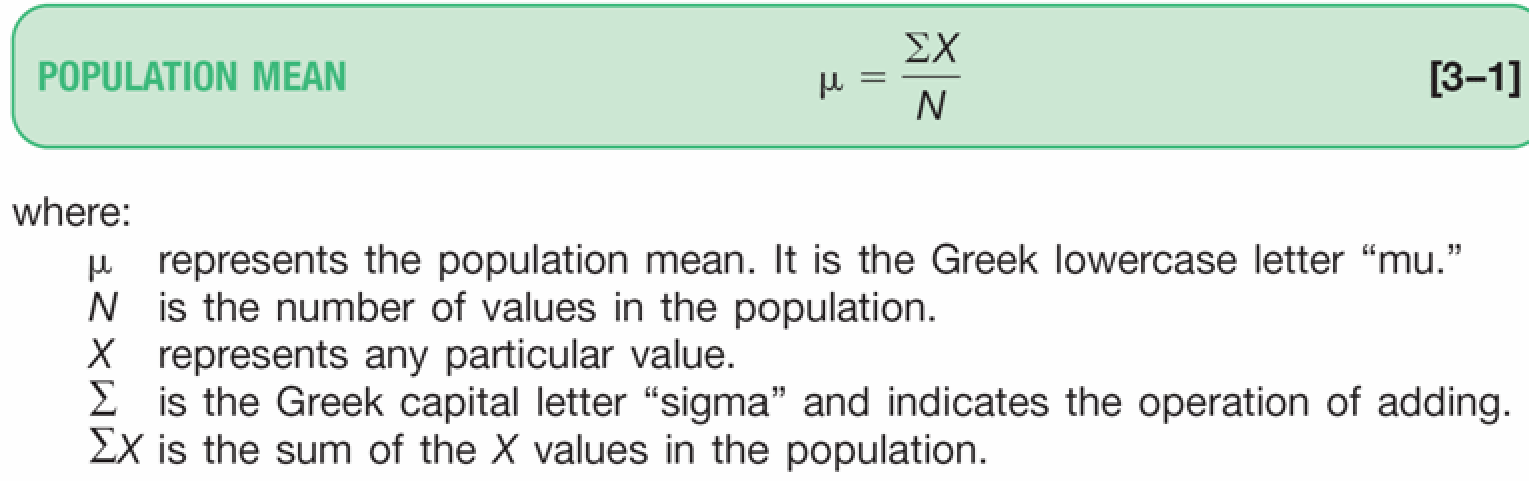
Characteristics of the Mean

The arithmetic mean is the most widely used measure of location. It requires the interval scale. Its major characteristics are:

- All values are used.
- It is unique.
- The sum of the deviations from the mean is 0.
- It is calculated by summing the values and dividing by the number of values.

Population Mean

For ungrouped data, the population mean is the sum of all the population values divided by the total number of population values:

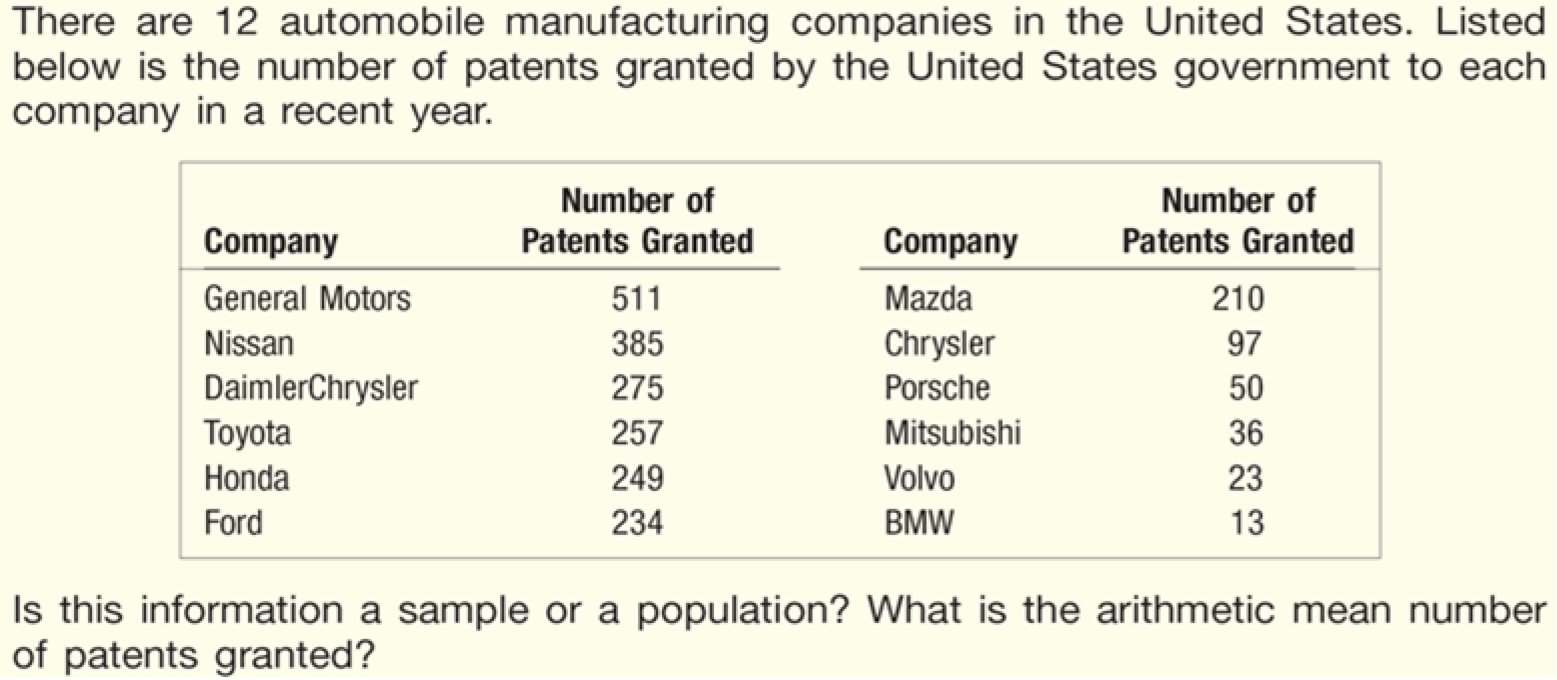


POPULATION MEAN $$\mu = \frac{\Sigma X}{N}$$ **[3–1]**

where:

μ represents the population mean. It is the Greek lowercase letter "mu."
N is the number of values in the population.
X represents any particular value.
Σ is the Greek capital letter "sigma" and indicates the operation of adding.
ΣX is the sum of the X values in the population.

EXAMPLE – Population Mean



There are 12 automobile manufacturing companies in the United States. Listed below is the number of patents granted by the United States government to each company in a recent year.

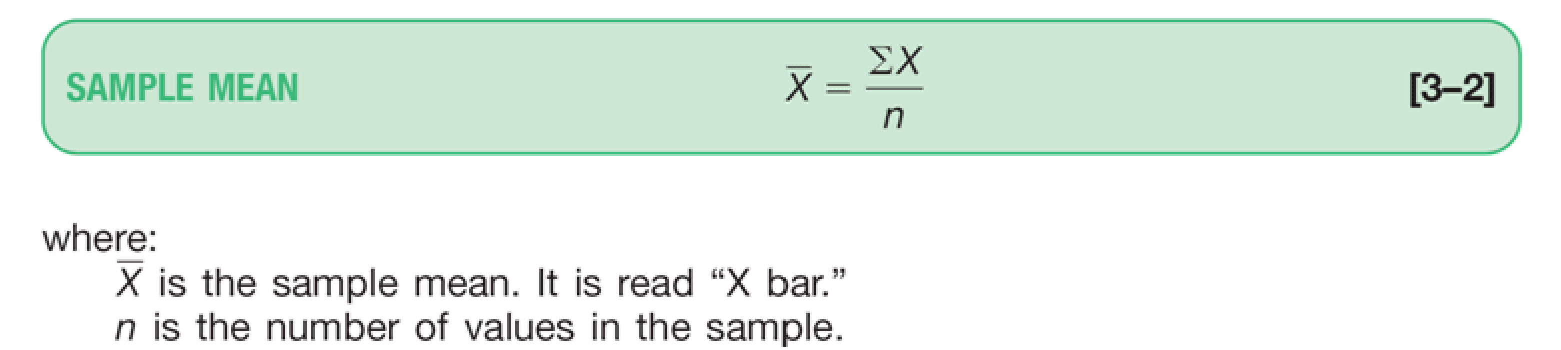
Company	Number of Patents Granted	Company	Number of Patents Granted
General Motors	511	Mazda	210
Nissan	385	Chrysler	97
DaimlerChrysler	275	Porsche	50
Toyota	257	Mitsubishi	36
Honda	249	Volvo	23
Ford	234	BMW	13

Is this information a sample or a population? What is the arithmetic mean number of patents granted?

$$\mu = \frac{\sum X}{N} = \frac{511+385+275+\ldots+36+23+13}{12} = \frac{2340}{12} = 195$$

Sample Mean

- For ungrouped data, the sample mean is the sum of all the sample values divided by the number of sample values:



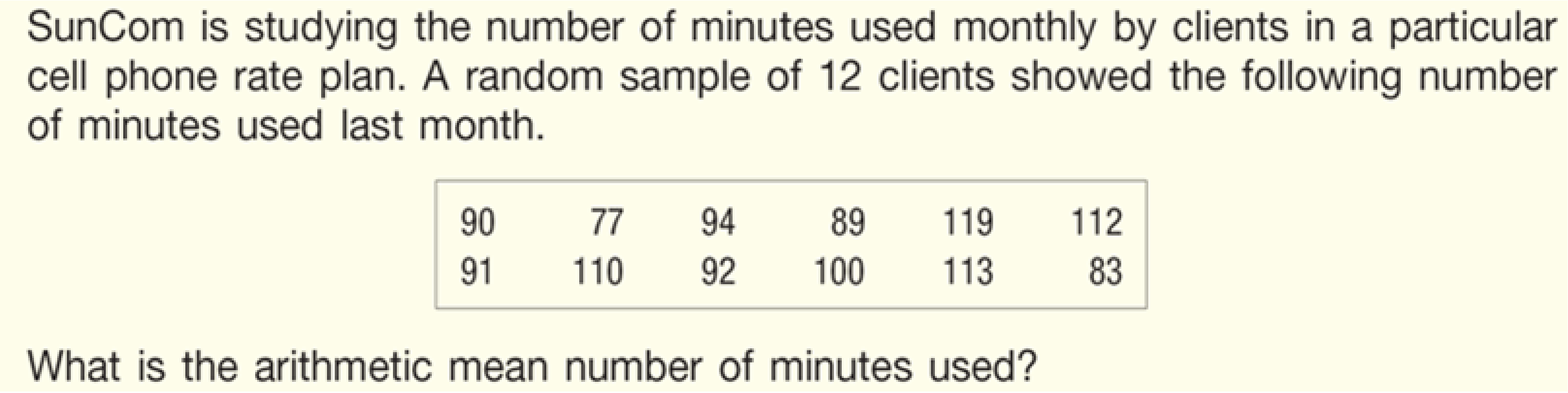
SAMPLE MEAN $$\overline{X} = \frac{\Sigma X}{n}$$ **[3–2]**

where:

$\overline{X}$ is the sample mean. It is read “X bar.”

n is the number of values in the sample.

EXAMPLE – Sample Mean



SunCom is studying the number of minutes used monthly by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

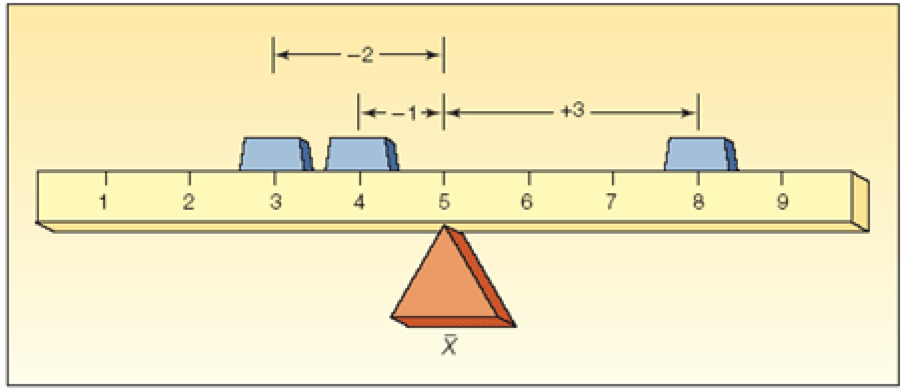
90	77	94	89	119	112
91	110	92	100	113	83

What is the arithmetic mean number of minutes used?

$$\bar{X} = \frac{\Sigma X}{n} = \frac{99+77+94+\ldots+100+113+83}{12} = \frac{1{,}170}{12} = 97.5$$

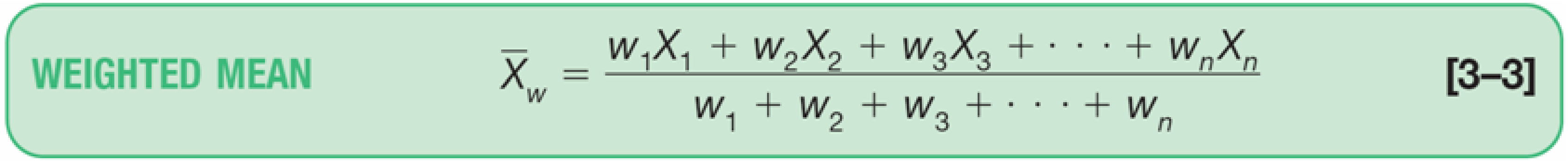
Properties of the Arithmetic Mean

- Every set of interval-level and ratio-level data has a mean.
- All the values are included in computing the mean.
- A set of data has a unique mean.
- The mean is affected by unusually large or small data values.
- The arithmetic mean is the only measure of central tendency where the sum of the deviations of each value from the mean is zero.



Weighted Mean

- The weighted mean of a set of numbers X_1, X_2, ..., X_n, with corresponding weights w_1, w_2, ...,w_n, is computed from the following formula:



WEIGHTED MEAN

$$\bar{X}_w = \frac{w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_nX_n}{w_1 + w_2 + w_3 + \cdots + w_n} \qquad [3\text{–}3]$$

EXAMPLE – Weighted Mean

The Carter Construction Company pays its hourly employees \$16.50, \$19.00, or \$25.00 per hour. There are 26 hourly employees, 14 of which are paid at the \$16.50 rate, 10 at the \$19.00 rate, and 2 at the \$25.00 rate. What is the mean hourly rate paid the 26 employees?

$$\begin{aligned} \bar{X}_w &= \frac{14(\$16.50)+10(\$19.00)+2(\$26.00)}{14+10+2} \\ &= \frac{\$471.00}{26} = \$18.1154 \end{aligned}$$

The Median

- The Median is the midpoint of the values after they have been ordered from the smallest to the largest.
  - There are as many values above the median as below it in the data array.
  - For an even set of values, the median will be the arithmetic average of the two middle numbers.

Properties of the Median

- There is a unique median for each data set.
- It is not affected by extremely large or small values and is therefore a valuable measure of central tendency when such values occur.
- It can be computed for ratio-level, interval-level, and ordinal-level data.
- It can be computed for an open-ended frequency distribution if the median does not lie in an open-ended class.

EXAMPLES - Median

The ages for a sample of five college students are:

21, 25, 19, 20, 22

Arranging the data in ascending order gives:

19, 20, 21, 22, 25.

Thus the median is 21.

The heights of four basketball players, in inches, are:

76, 73, 80, 75

Arranging the data in ascending order gives:

73, 75, 76, 80.

Thus the median is 75.5

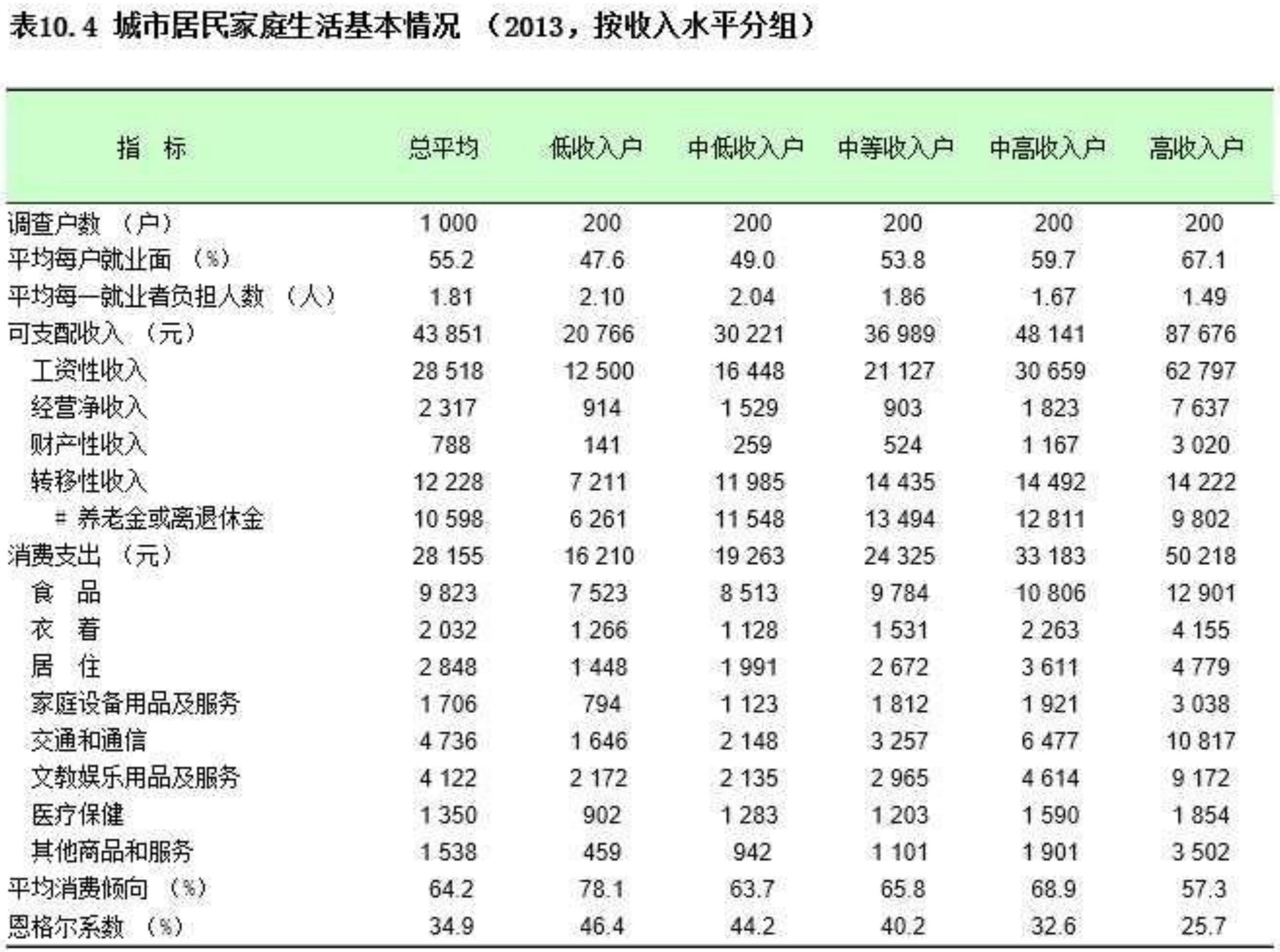


表10.4 城市居民家庭生活基本情况 （2013，按收入水平分组）

指　标	总平均	低收入户	中低收入户	中等收入户	中高收入户	高收入户
调查户数 （户）	1 000	200	200	200	200	200
平均每户就业面 （%）	55.2	47.6	49.0	53.8	59.7	67.1
平均每一就业者负担人数 （人）	1.81	2.10	2.04	1.86	1.67	1.49
可支配收入 （元）	43 851	20 766	30 221	36 989	48 141	87 676
工资性收入	28 518	12 500	16 448	21 127	30 659	62 797
经营净收入	2 317	914	1 529	903	1 823	7 637
财产性收入	788	141	259	524	1 167	3 020
转移性收入	12 228	7 211	11 985	14 435	14 492	14 222
# 养老金或离退休金	10 598	6 261	11 548	13 494	12 811	9 802
消费支出 （元）	28 155	16 210	19 263	24 325	33 183	50 218
食　品	9 823	7 523	8 513	9 784	10 806	12 901
衣　着	2 032	1 266	1 128	1 531	2 263	4 155
居　住	2 848	1 448	1 991	2 672	3 611	4 779
家庭设备用品及服务	1 706	794	1 123	1 812	1 921	3 038
交通和通信	4 736	1 646	2 148	3 257	6 477	10 817
文教娱乐用品及服务	4 122	2 172	2 135	2 965	4 614	9 172
医疗保健	1 350	902	1 283	1 203	1 590	1 854
其他商品和服务	1 538	459	942	1 101	1 901	3 502
平均消费倾向 （%）	64.2	78.1	63.7	65.8	68.9	57.3
恩格尔系数 （%）	34.9	46.4	44.2	40.2	32.6	25.7
]

The Mode

- The mode is the value of the observation that appears most frequently.

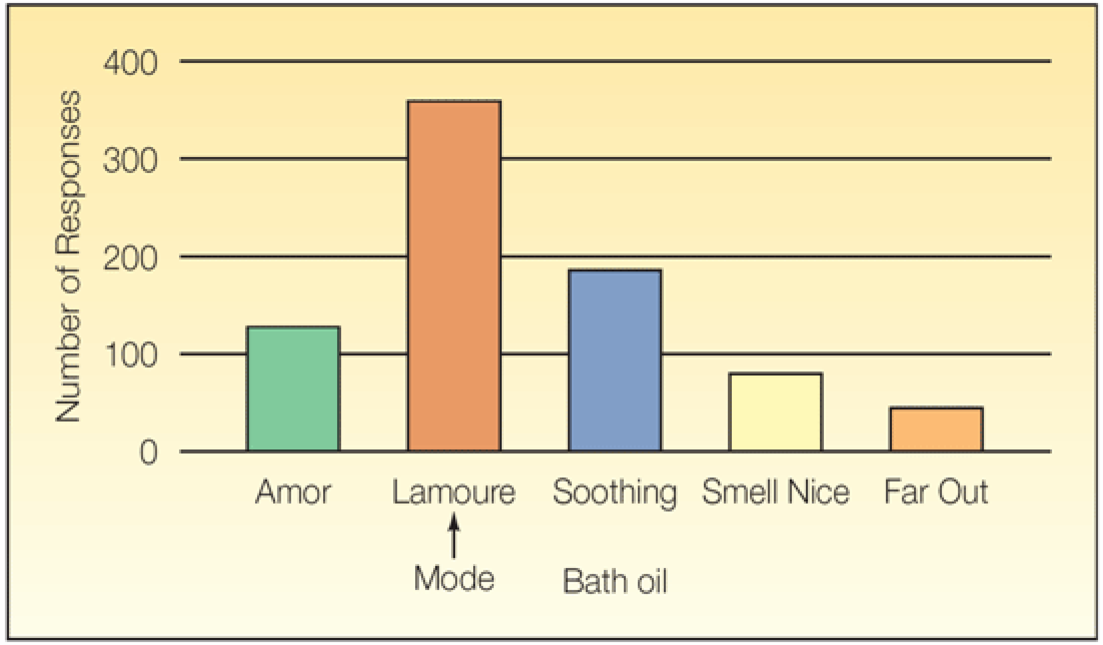
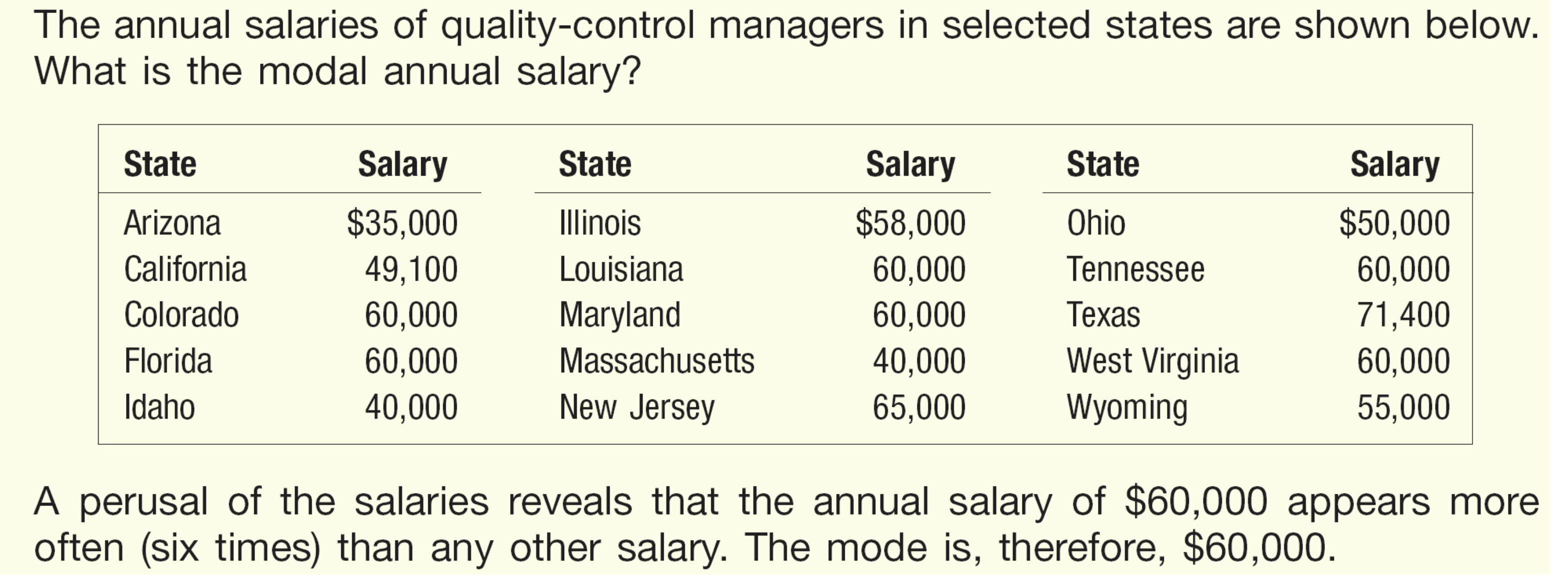


CHART 3–1 Number of Respondents Favoring Various Bath Oils

Example - Mode



The annual salaries of quality-control managers in selected states are shown below. What is the modal annual salary?

State	Salary	State	Salary	State	Salary
Arizona	$35,000	Illinois	$58,000	Ohio	$50,000
California	49,100	Louisiana	60,000	Tennessee	60,000
Colorado	60,000	Maryland	60,000	Texas	71,400
Florida	60,000	Massachusetts	40,000	West Virginia	60,000
Idaho	40,000	New Jersey	65,000	Wyoming	55,000

A perusal of the salaries reveals that the annual salary of $60,000 appears more often (six times) than any other salary. The mode is, therefore, $60,000.

Mean, Median, Mode Using Excel

Table 2–4 in Chapter 2 shows the prices of the 80 vehicles sold last month at Whitner Autoplex in Raytown, Missouri. Determine the mean and the median selling price. The mean and the median selling prices are reported in the following Excel output. There are 80 vehicles in the study. So the calculations with a calculator would be tedious and prone to error.

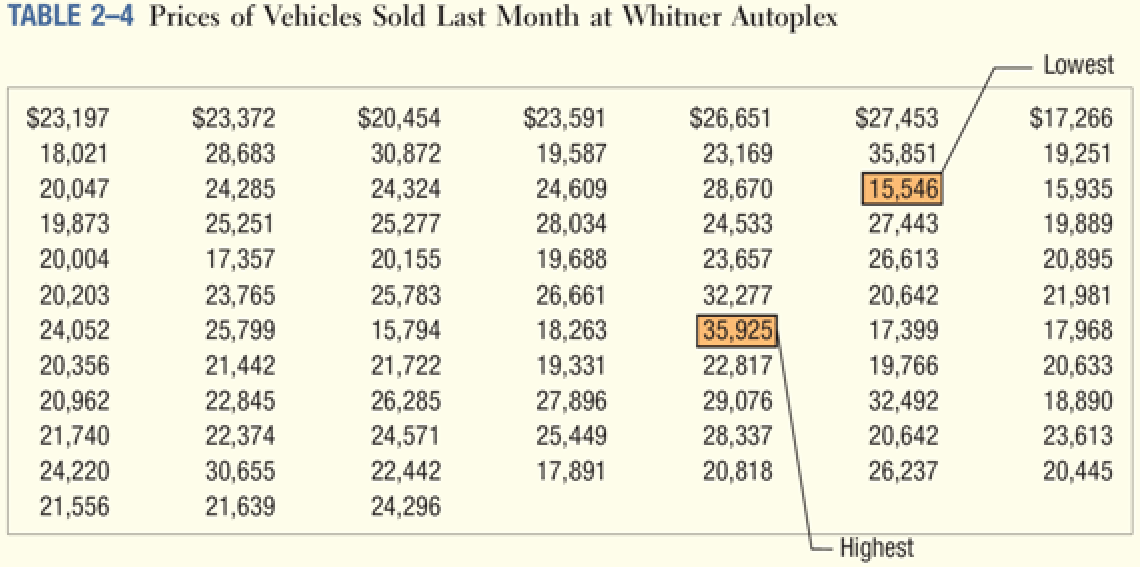


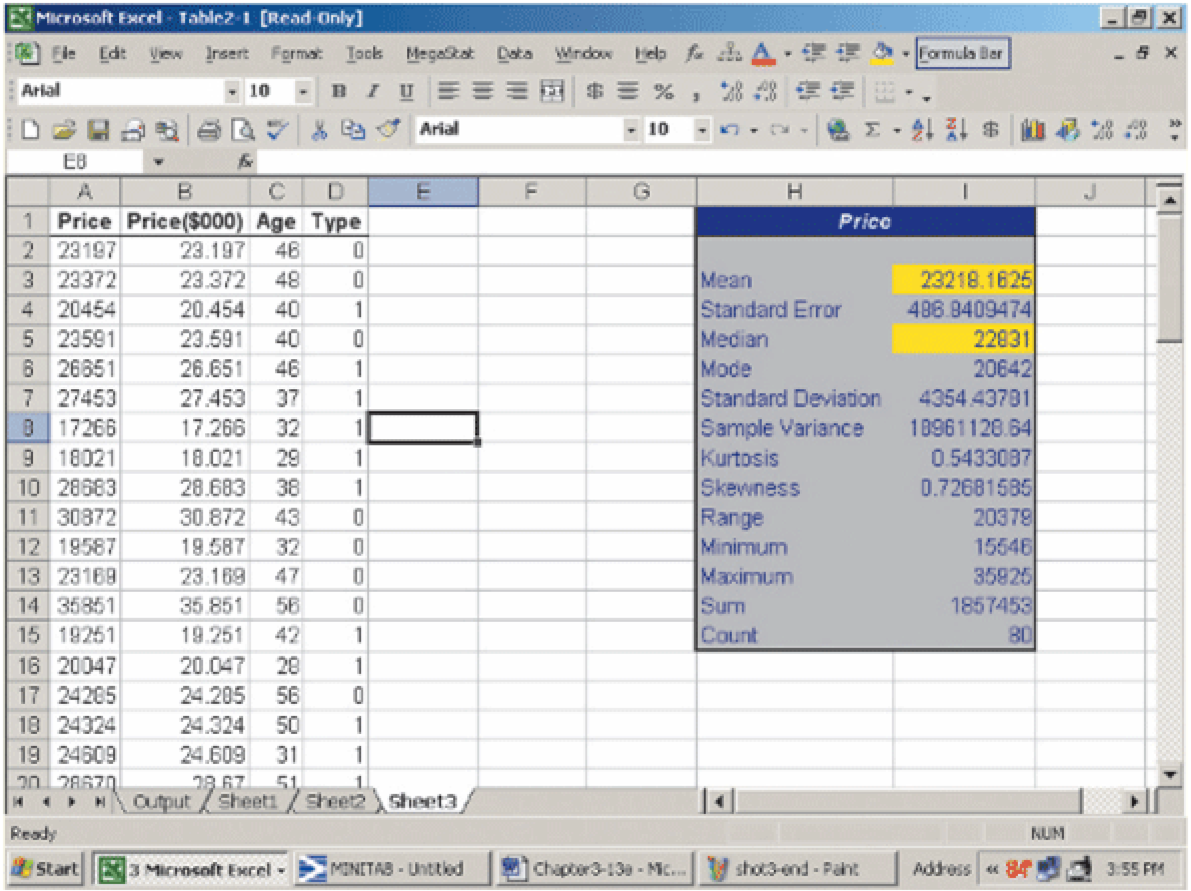
TABLE 2–4 Prices of Vehicles Sold Last Month at Whitner Autoplex

$23,197	$23,372	$20,454	$23,591	$26,651	$27,453	$17,266
18,021	28,683	30,872	19,587	23,169	35,851	19,251
20,047	24,285	24,324	24,609	28,670	15,546	15,935
19,873	25,251	25,277	28,034	24,533	27,443	19,889
20,004	17,357	20,155	19,688	23,657	26,613	20,895
20,203	23,765	25,783	26,661	32,277	20,642	21,981
24,052	25,799	15,794	18,263	35,925	17,399	17,968
20,356	21,442	21,722	19,331	22,817	19,766	20,633
20,962	22,845	26,285	27,896	29,076	32,492	18,890
21,740	22,374	24,571	25,449	28,337	20,642	23,613
24,220	30,655	22,442	17,891	20,818	26,237	20,445
21,556	21,639	24,296				

Lowest

Highest

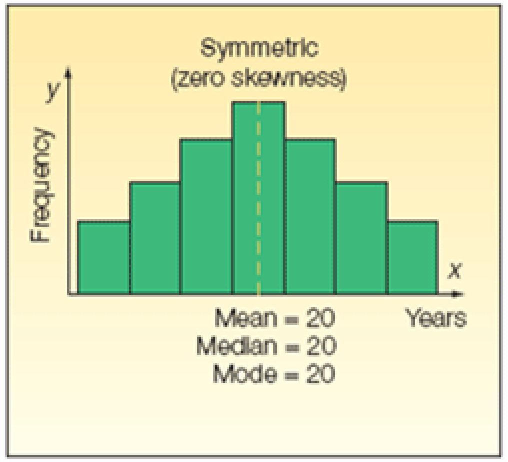
Mean, Median, Mode Using Excel


Microsoft Excel - Table2-1 [Read-Only]
File Edit View Insert Format Tools MegaStat Data Window Help
Formula Bar
Arial 10
E8
Output Sheet1 Sheet2 Sheet3
Ready
NUM
Start
3 Microsoft Excel
MINITAB - Untitled
Chapter3-13e - Mic...
shot3-end - Paint
Address
3:55 PM

	A	B	C	D
1	Price	Price($000)	Age	Type
2	23197	23.197	46	0
3	23372	23.372	48	0
4	20454	20.454	40	1
5	23591	23.591	40	0
6	26651	26.651	46	1
7	27453	27.453	37	1
8	17266	17.266	32	1
9	18021	18.021	29	1
10	28683	28.683	38	1
11	30872	30.872	43	0
12	19587	19.587	32	0
13	23169	23.169	47	0
14	35851	35.851	56	0
15	19251	19.251	42	1
16	20047	20.047	28	1
17	24285	24.285	56	0
18	24324	24.324	50	1
19	24609	24.609	31	1
20	28670	28.67	51	1

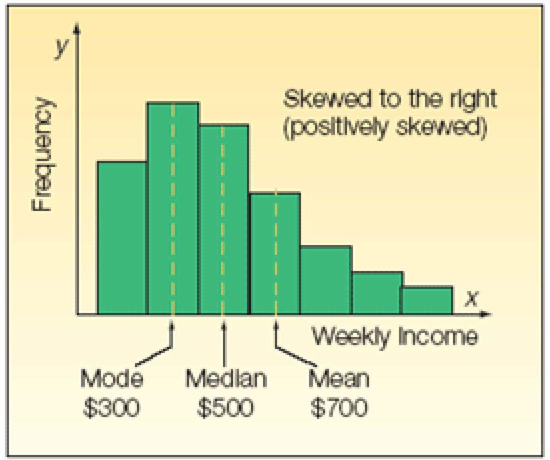
Price	
Mean	23218.1625
Standard Error	486.8409474
Median	22831
Mode	20642
Standard Deviation	4354.43781
Sample Variance	18961128.64
Kurtosis	0.5433087
Skewness	0.72681585
Range	20379
Minimum	15546
Maximum	35925
Sum	1857453
Count	80

The Relative Positions of the Mean, Median and the Mode



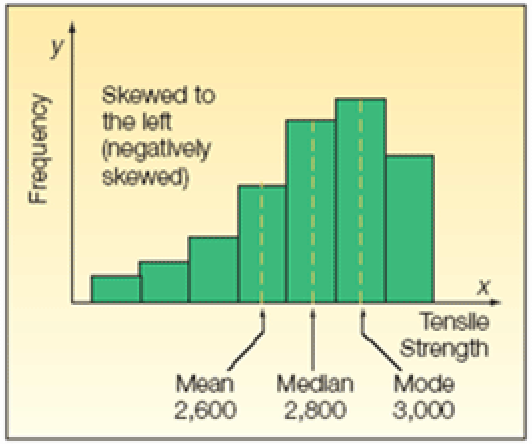
zero skewness

mode = median = mean



positive skewness

mode < median < mean

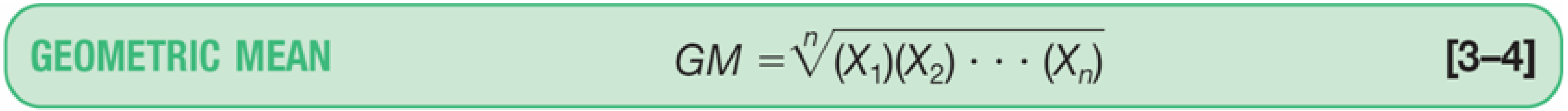


negative skewness

mode > median > mean

The Geometric Mean

- Useful in finding the average change of percentages, ratios, indexes, or growth rates over time.
- It has a wide application in business and economics because we are often interested in finding the percentage changes in sales, salaries, or economic figures, such as the GDP, which compound or build on each other.
- The geometric mean will always be less than or equal to the arithmetic mean.
- The geometric mean of a set of *n* positive numbers is defined as the *n*th root of the product of *n* values.
- The formula for the geometric mean is written:



GEOMETRIC MEAN

$$GM = \sqrt[n]{(X_1)(X_2)\cdots(X_n)} \qquad [3\text{–}4]$$

EXAMPLE – Geometric Mean

Suppose you receive a 5 percent increase in salary this year and a 15 percent increase next year. The average annual percent increase is 9.886, not 10.0. Why is this so? We begin by calculating the geometric mean.

$$GM = \sqrt{(1.05)(1.15)} = 1.09886$$

EXAMPLE – Geometric Mean (2)

The return on investment earned by Atkins construction Company for four successive years was: 30 percent, 20 percent, -40 percent, and 200 percent. What is the geometric mean rate of return on investment?

$$GM = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)} = \sqrt[4]{2.808} = 1.294$$

Dispersion

Why Study Dispersion?

- A measure of location, such as the mean or the median, only describes the center of the data. It is valuable from that standpoint, but it does not tell us anything about the spread of the data.
- For example, if your nature guide told you that the river ahead averaged 3 feet in depth, would you want to wade across on foot without additional information? Probably not. You would want to know something about the variation in the depth.
- A second reason for studying the dispersion in a set of data is to compare the spread in two or more distributions.

Samples of Dispersions

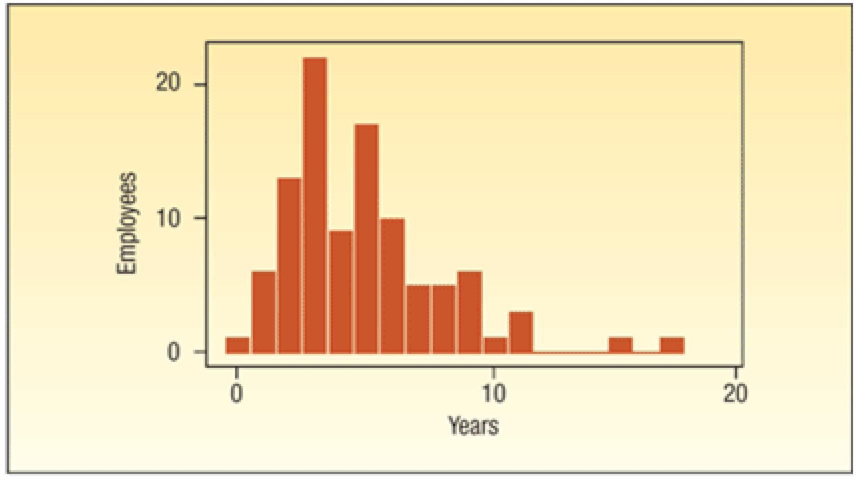


CHART 3–5 Histogram of Years of Employment at Hammond Iron Works, Inc.

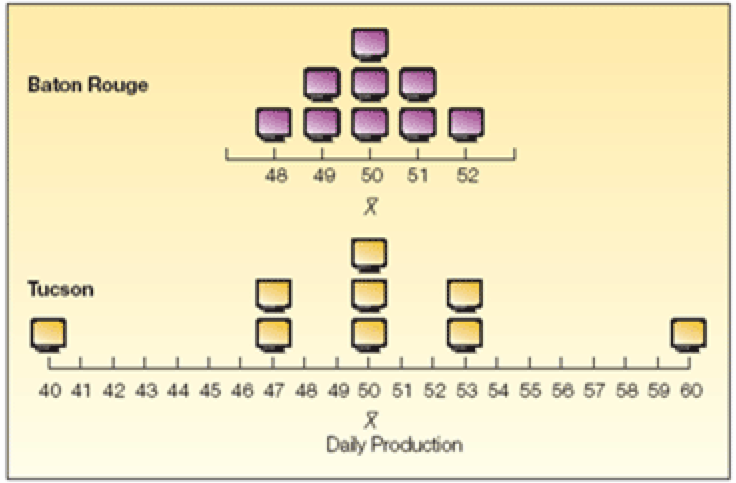
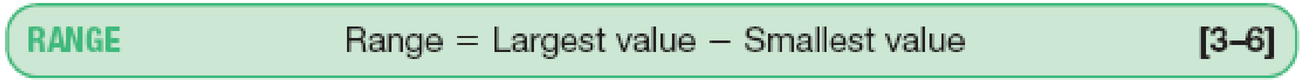


CHART 3–6 Hourly Production of Computer Monitors at the Baton Rouge and Tucson Plants

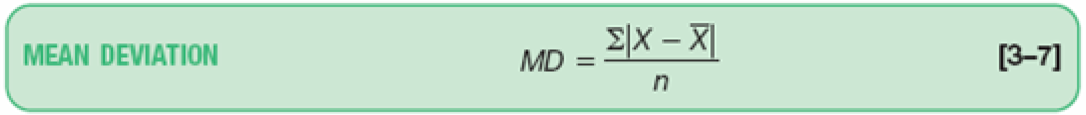
Measures of Dispersion

- Range



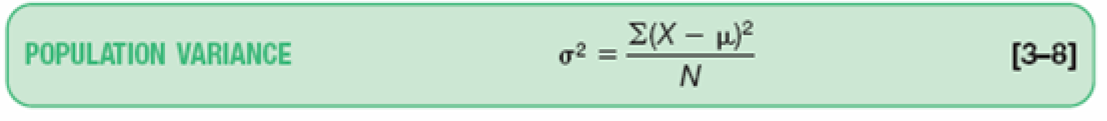
RANGE Range = Largest value − Smallest value [3–6]

- Mean Deviation

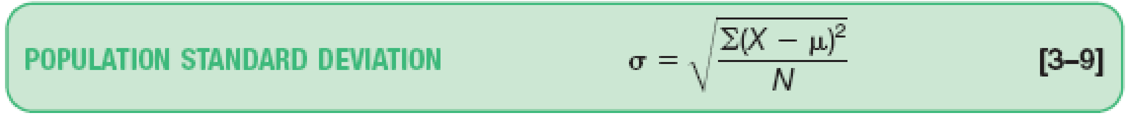


MEAN DEVIATION $MD = \frac{\Sigma|X - \overline{X}|}{n}$ [3–7]

- Variance and Standard Deviation



POPULATION VARIANCE $\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$ [3–8]



POPULATION STANDARD DEVIATION $\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$ [3–9]

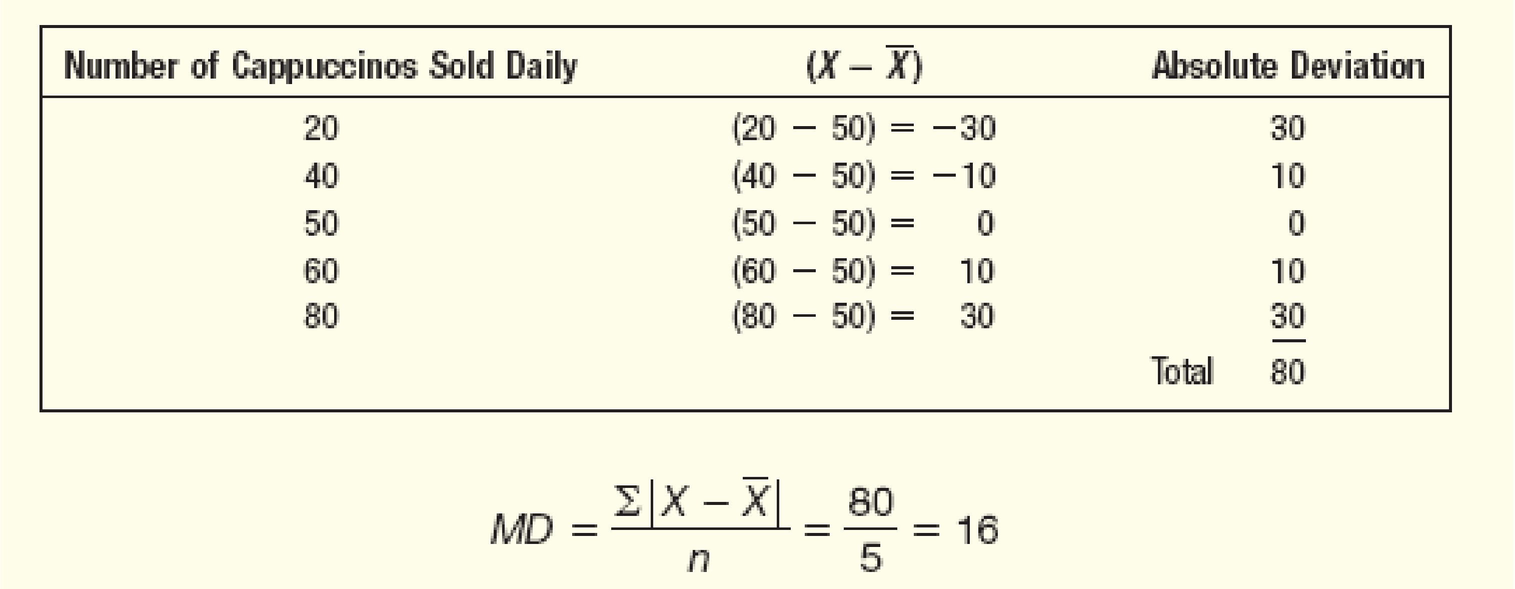
EXAMPLE – Range

The number of cappuccinos sold at the Starbucks location in the Orange Country Airport between 4 and 7 p.m. for a sample of 5 days last year were 20, 40, 50, 60, and 80. Determine the mean deviation for the number of cappuccinos sold.

Range = Largest – Smallest value
= 80 – 20 = 60

EXAMPLE – Mean Deviation

The number of cappuccinos sold at the Starbucks location in the Orange Country Airport between 4 and 7 p.m. for a sample of 5 days last year were 20, 40, 50, 60, and 80. Determine the mean deviation for the number of cappuccinos sold.

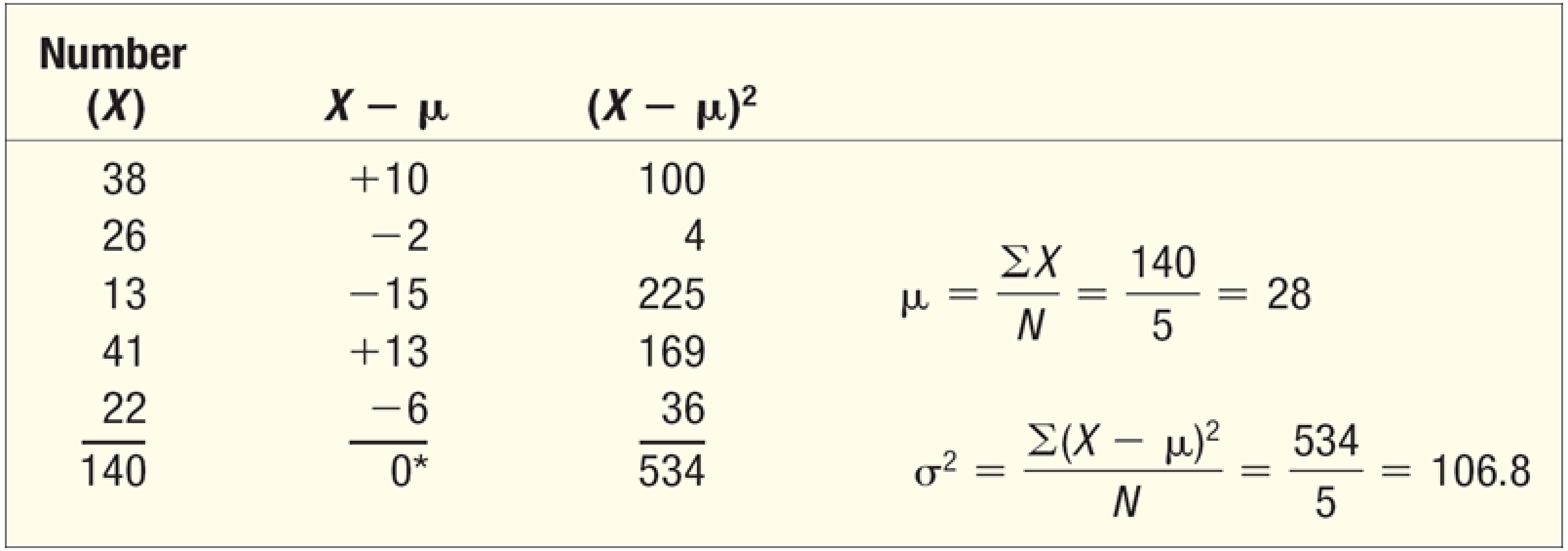


Number of Cappuccinos Sold Daily	$(X - \bar{X})$	Absolute Deviation
20	$(20 - 50) = -30$	30
40	$(40 - 50) = -10$	10
50	$(50 - 50) = 0$	0
60	$(60 - 50) = 10$	10
80	$(80 - 50) = 30$	30
		Total 80

$$MD = \frac{\Sigma|X - \bar{X}|}{n} = \frac{80}{5} = 16$$

EXAMPLE – Variance and Standard Deviation

The number of traffic citations issued during the last five months in Beaufort County, South Carolina, is 38, 26, 13, 41, and 22. What is the population variance?



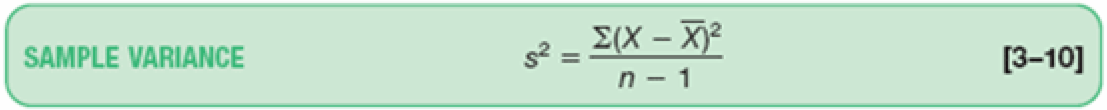
Number (X)	$X - \mu$	$(X - \mu)^2$
38	+10	100
26	−2	4
13	−15	225
41	+13	169
22	−6	36
140	0*	534

$$\mu = \frac{\Sigma X}{N} = \frac{140}{5} = 28$$

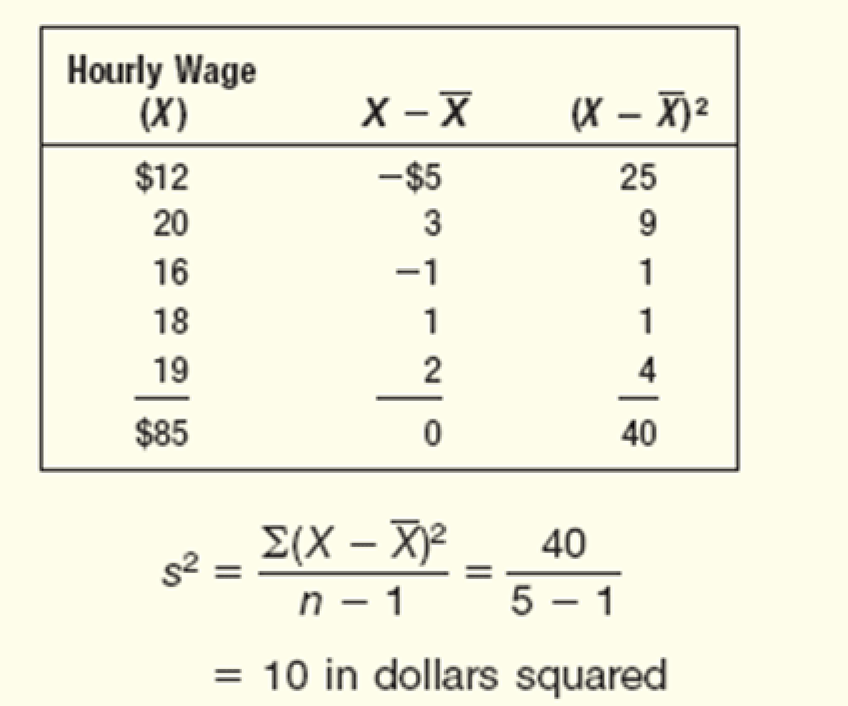
$$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N} = \frac{534}{5} = 106.8$$

EXAMPLE – Sample Variance

The hourly wages for a sample of part-time employees at Home Depot are: $12, $20, $16, $18, and $19. What is the sample variance?



SAMPLE VARIANCE $$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$$ [3–10]



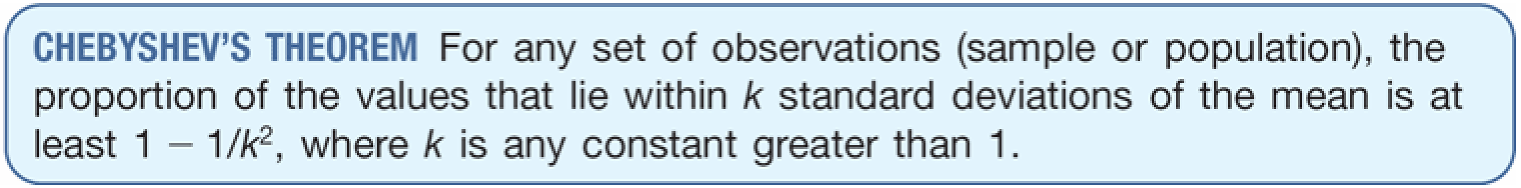
Hourly Wage (*X*)	$X - \overline{X}$	$(X - \overline{X})^2$
$12	−$5	25
20	3	9
16	−1	1
18	1	1
19	2	4
$85	0	40

$$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1} = \frac{40}{5 - 1}$$

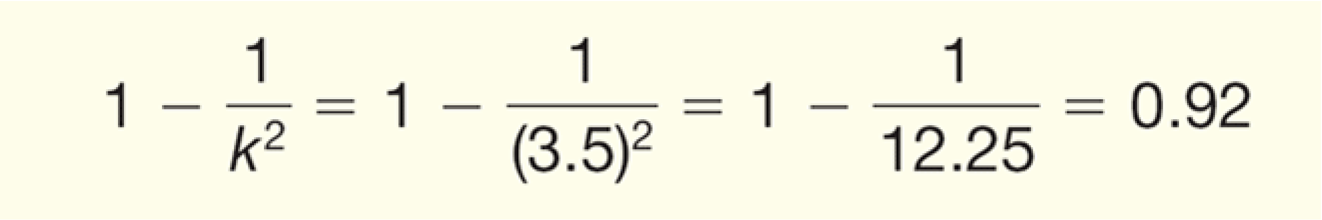
= 10 in dollars squared

Chebyshev’s Theorem

The arithmetic mean biweekly amount contributed by the Dupree Paint employees to the company’s profit-sharing plan is \$51.54, and the standard deviation is \$7.51. At least what percent of the contributions lie within plus 3.5 standard deviations and minus 3.5 standard deviations of the mean?

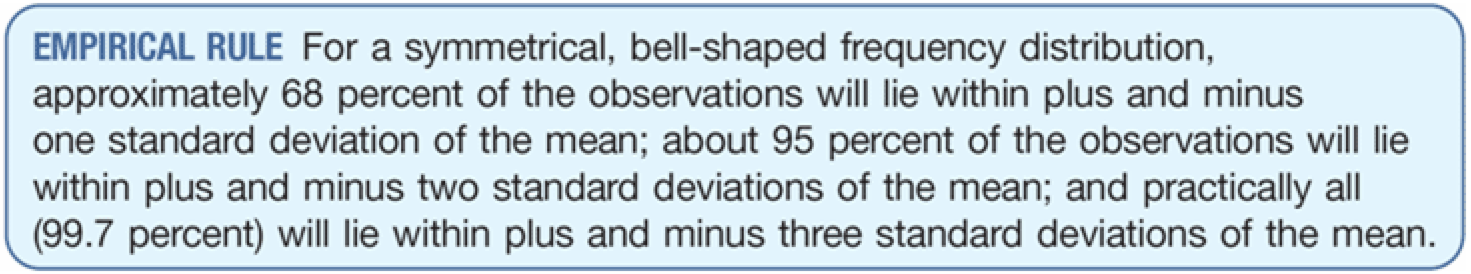


CHEBYSHEV’S THEOREM For any set of observations (sample or population), the proportion of the values that lie within *k* standard deviations of the mean is at least $1 - 1/k^2$, where *k* is any constant greater than 1.



$$1 - \frac{1}{k^2} = 1 - \frac{1}{(3.5)^2} = 1 - \frac{1}{12.25} = 0.92$$

The Empirical Rule



EMPIRICAL RULE For a symmetrical, bell-shaped frequency distribution, approximately 68 percent of the observations will lie within plus and minus one standard deviation of the mean; about 95 percent of the observations will lie within plus and minus two standard deviations of the mean; and practically all (99.7 percent) will lie within plus and minus three standard deviations of the mean.

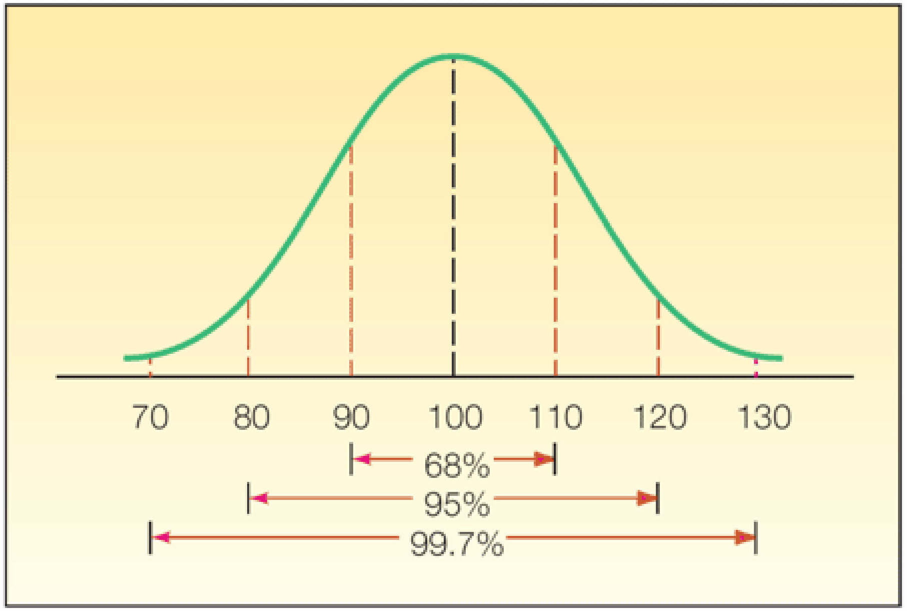
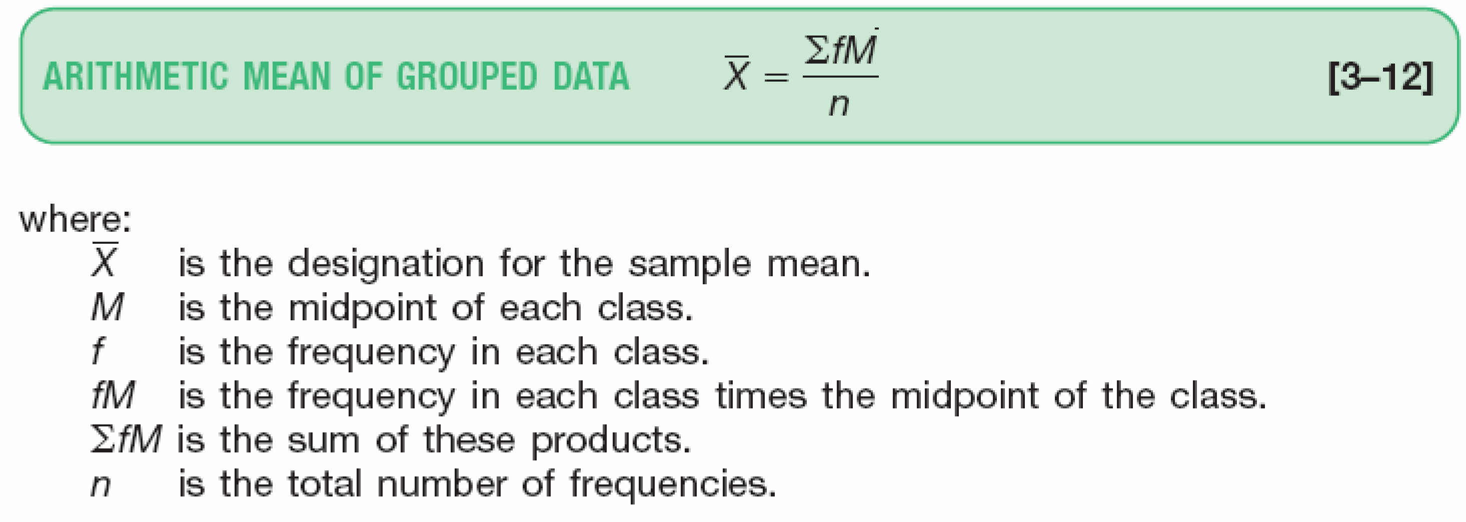


CHART 3–7 A Symmetrical, Bell-Shaped Curve Showing the Relationships between the Standard Deviation and the Observations

The Arithmetic Mean of Grouped Data



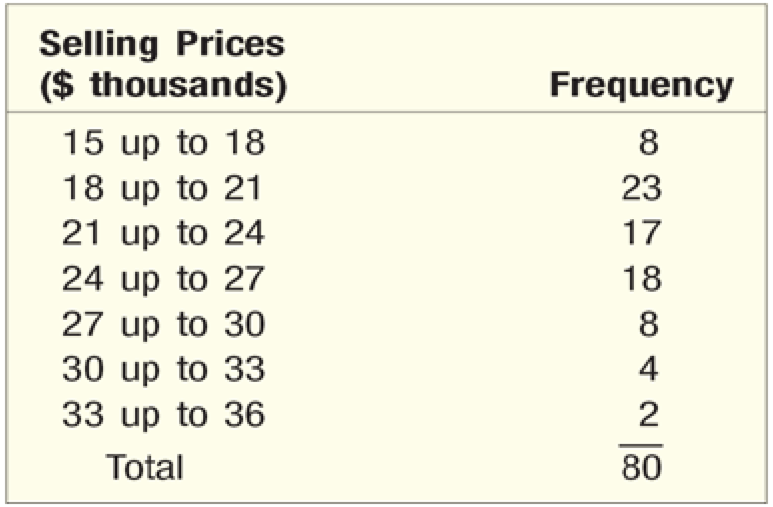
ARITHMETIC MEAN OF GROUPED DATA $\bar{X} = \frac{\Sigma fM}{n}$ **[3–12]**

where:

$\bar{X}$ is the designation for the sample mean.
M is the midpoint of each class.
f is the frequency in each class.
fM is the frequency in each class times the midpoint of the class.
ΣfM is the sum of these products.
n is the total number of frequencies.

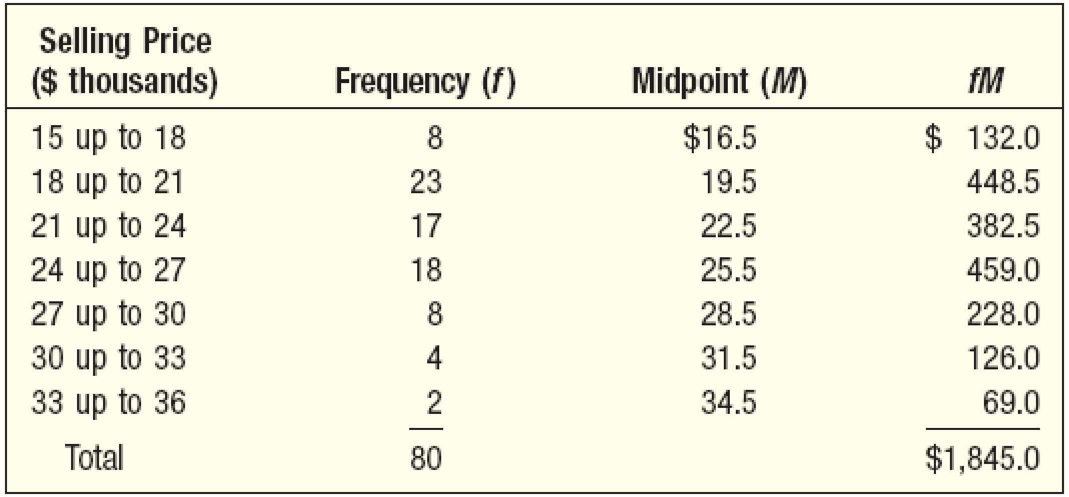
The Arithmetic Mean of Grouped Data - Example

Recall in Chapter 2, we constructed a frequency distribution for the vehicle selling prices. The information is repeated below. Determine the arithmetic mean vehicle selling price.

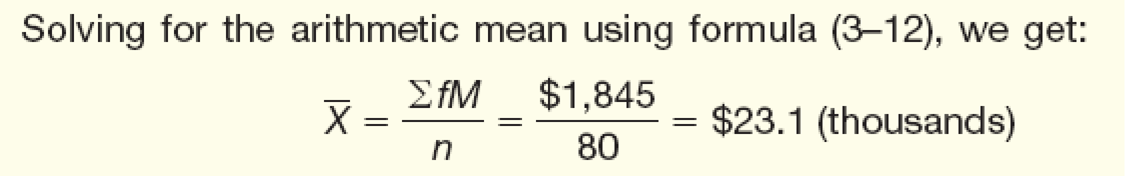


Selling Prices ($ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	80

The Arithmetic Mean of Grouped Data - Example



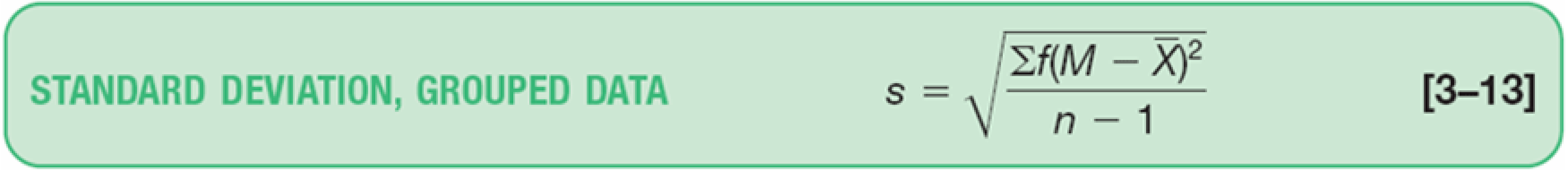
Selling Price ($ thousands)	Frequency (f)	Midpoint (M)	fM
15 up to 18	8	$16.5	$ 132.0
18 up to 21	23	19.5	448.5
21 up to 24	17	22.5	382.5
24 up to 27	18	25.5	459.0
27 up to 30	8	28.5	228.0
30 up to 33	4	31.5	126.0
33 up to 36	2	34.5	69.0
Total	80		$1,845.0



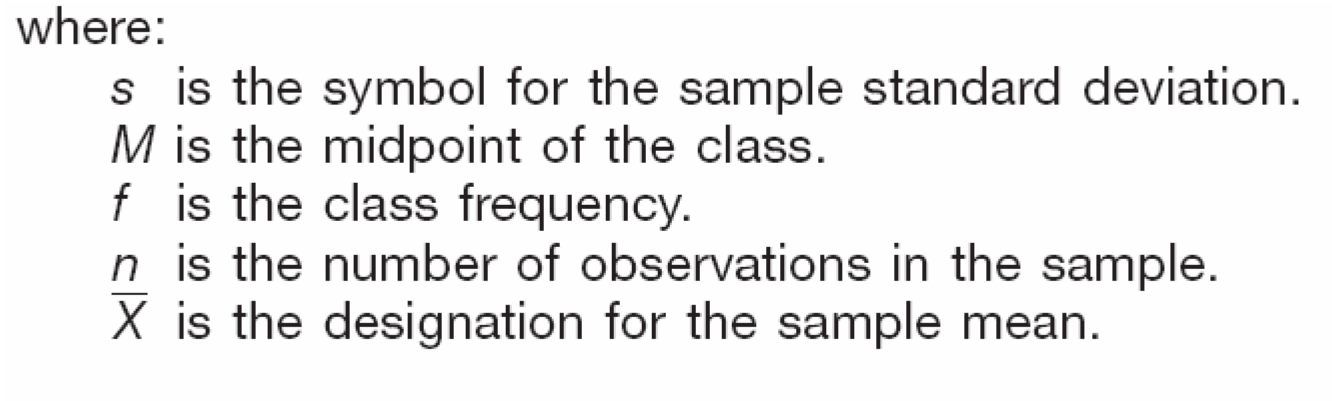
Solving for the arithmetic mean using formula (3–12), we get:

$$\overline{X} = \frac{\Sigma fM}{n} = \frac{\$1,845}{80} = \$23.1 \text{ (thousands)}$$

Standard Deviation of Grouped Data



STANDARD DEVIATION, GROUPED DATA $$s = \sqrt{\frac{\Sigma f(M - \overline{X})^2}{n - 1}}$$ **[3–13]**

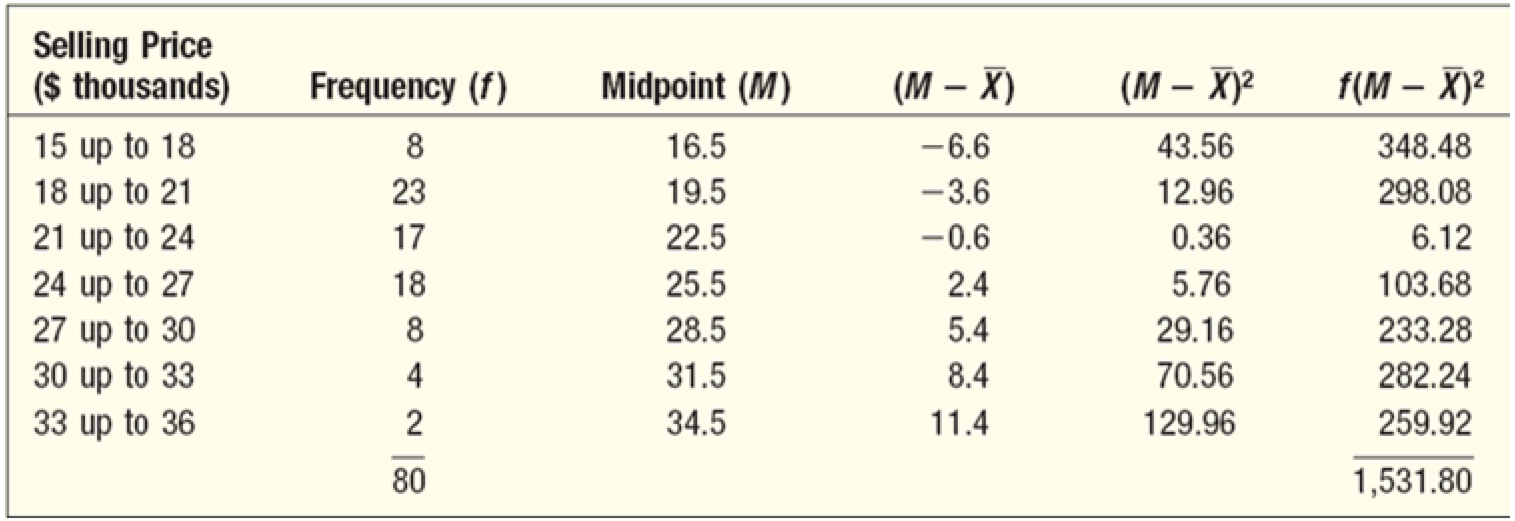


where:

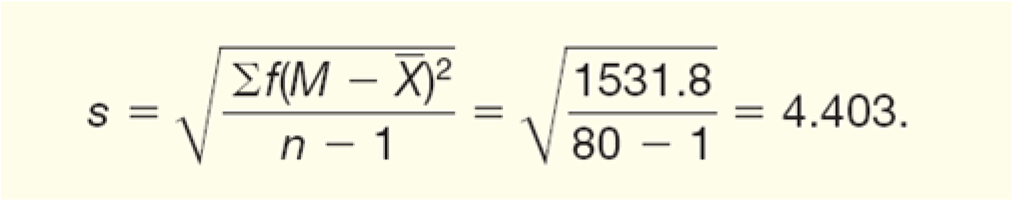
- s is the symbol for the sample standard deviation.
- M is the midpoint of the class.
- f is the class frequency.
- n is the number of observations in the sample.
- $\overline{X}$ is the designation for the sample mean.

Standard Deviation of Grouped Data - Example

Refer to the frequency distribution for the Whitner Autoplex data used earlier. Compute the standard deviation of the vehicle selling prices



Selling Price ($ thousands)	Frequency (f)	Midpoint (M)	$(M - \bar{X})$	$(M - \bar{X})^2$	$f(M - \bar{X})^2$
15 up to 18	8	16.5	−6.6	43.56	348.48
18 up to 21	23	19.5	−3.6	12.96	298.08
21 up to 24	17	22.5	−0.6	0.36	6.12
24 up to 27	18	25.5	2.4	5.76	103.68
27 up to 30	8	28.5	5.4	29.16	233.28
30 up to 33	4	31.5	8.4	70.56	282.24
33 up to 36	2	34.5	11.4	129.96	259.92
	80				1,531.80



$$s = \sqrt{\frac{\Sigma f(M - \bar{X})^2}{n - 1}} = \sqrt{\frac{1531.8}{80 - 1}} = 4.403.$$

End of Chapter 3