6-1 Numerical Summaries of Data

- Well-constructed data summaries and displays are essential to good statistical thinking, because they can focus engineer on important features of the data or provide insight about the type of model that should be used in solving problem.
- We often find it useful to describe data features numerically.
- For example, we can characterize location or central tendency in data by the ordinary arithmetic average or mean.
- Because we almost always think of our data as a sample, we will refer to arithmetic mean as *sample mean*.

Definition: Sample Mean

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample mean** is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{6-1}$$

3

6-1 Numerical Summaries of Data

Example 6-1

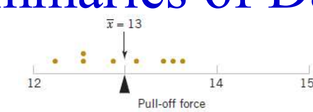


FIGURE 6-1 Dot diagram showing the sample mean as a balance point for a system of weights.

Let's consider the eight observations collected from the prototype engine connectors from Chapter 1. The eight observations are $x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. The sample mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{12.6 + 12.9 + \cdots + 13.1}{8}$$

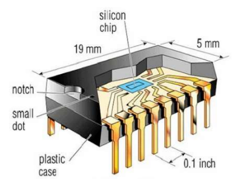
$$= \frac{104}{8} = 13.0 \text{ pounds}$$

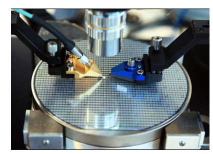
A physical interpretation of the sample mean as a measure of location is shown in the dot diagram of the pull-off force data. See Figure 6-1. Notice that the sample mean $\bar{x} = 13.0$ can be thought of as a "balance point." That is, if each observation represents 1 pound of mass placed at the point on the x-axis, a fulcrum located at $\bar{x}$ would exactly balance this system of weights. 4

4

6-1 Numerical Summaries of Data

- **Sample mean** is average value of all observations in data set.
- Usually, these data are **sample** of observations that have been selected from some larger **population** of observations.
- Here population might consist of all connectors that will be manufactured and sold to customers.
- Recall that this type of population is called conceptual or hypothetical population because it does not physically exist.
- Sometimes there is actual physical population,
- such as a lot of silicon wafers produced in a semiconductor factory.



5

6-1 Numerical Summaries of Data

Population Mean

If we think of **probability distribution** as model for population, one way to think of **mean** is as **average of all measurements in population.**

For a finite population with *N equally likely values,* ***Probability Mass Function*** *(pmf) is $f(x_i) = 1/N$ and mean is*

Mean of probability distribution

$$\mu = \sum_{i=1}^{N} x_i f(x_i) = \frac{\sum_{i=1}^{N} x_i}{N} \qquad (6\text{-}2)$$

μ

Sample mean, $\bar{x}$, is a reasonable estimate of population mean.

6

6

6-1 Numerical Summaries of Data

- Although sample mean is useful, it does not convey all of the information about sample of data.
- Variability or scatter in data may be described by
  - **sample Variance** or
  - **sample Standard Deviation (SD)**.

Sample Variance and Standard Deviation

If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the **sample variance** is

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} \tag{6-3}$$

The **sample standard deviation**, s, is the positive square root of the sample variance.

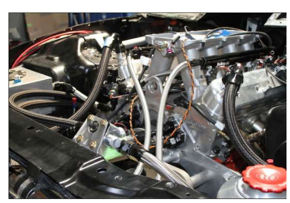
- **Units** of measurement for sample variance are **square** of original units of variable.
- Thus, if *x is measured in pounds, the units for sample variance* are *(pounds)*^{2}.
- **Standard deviation** has desirable property of measuring variability in original units of variable of interest, *x*.

7

7

How Does the Sample Variance Measure Variability?

- Eight prototype units are produced and their pull-off forces are measured (in pounds):
  - 12.6,
  - 12.9,
  - 13.4,
  - 12.3,
  - 13.6,
  - 13.5,
  - 12.6,
  - 13.1.



Nylon connector to be used in automotive engine

Dot Diagram

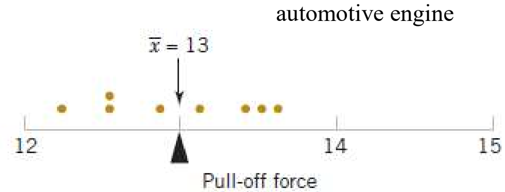
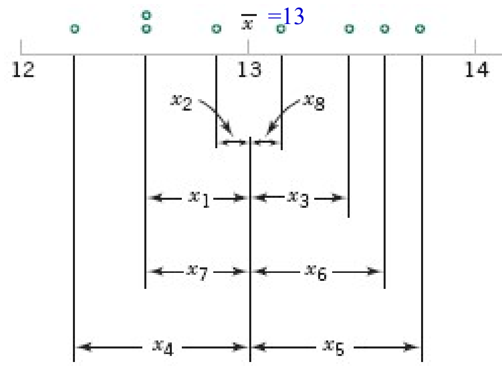


FIGURE 6-1 Dot diagram showing the sample mean as a balance point for a system of weights.

8

6-1 Numerical Summaries

How Does Sample Variance Measure Variability?



x1 = 12.6,
x2 = 12.9,
x3 = 13.4,
x4 = 12.3,
x5 = 13.6,
x6 = 13.5,
x7 = 12.6, and
x8 = 13.1.

Figure 6-2 How the sample variance measures variability through deviations. $x_i - \bar{x}$

9

9

6-1 Numerical Summaries

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Degree of freedom

Table 6-1 Calculation of Terms for the Sample Variance and Sample Standard Deviation

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	12.6	−0.4	0.16
2	12.9	−0.1	0.01
3	13.4	0.4	0.16
4	12.3	−0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	−0.4	0.16
8	13.1	0.1	0.01
	104.0	0.0	1.60

10

10

6-1 Numerical Summaries

Example 6-2

Table 6-1 displays the quantities needed for calculating the sample variance and sample standard deviation for the pull-off force data. These data are plotted in Fig. 6-2. The numerator of s^2 is

$$\sum_{i=1}^{8} (x_i - \overline{x})^2 = 1.60$$

Table 6-1 Calculation of Terms for the Sample Variance and Sample Standard Deviation

i	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1	12.6	−0.4	0.16
2	12.9	−0.1	0.01
3	13.4	0.4	0.16
4	12.3	−0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	−0.4	0.16
8	13.1	0.1	0.01
	104.0	0.0	1.60

so the sample variance is

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

$$s^2 = \frac{1.60}{8-1} = \frac{1.60}{7} = 0.2286 \text{ (pounds)}^2$$

and the sample standard deviation is

$$s = \sqrt{0.2286} = 0.48 \text{ pounds}$$

11

11

Computation of s^2

The computation of s^2 requires calculation of $\overline{x}$, n subtractions, and n squaring and adding operations. If the original observations or the deviations $x_i - \overline{x}$ are not integers, the deviations $x_i - \overline{x}$ may be tedious to work with, and several decimals may have to be carried to ensure numerical accuracy. A more efficient computational formula for the sample variance is obtained as follows:

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} \left(x_i^2 + \overline{x}^2 - 2\overline{x}x_i\right)}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 + n\overline{x}^2 - 2\overline{x}\sum_{i=1}^{n} x_i}{n-1}$$

and because $\overline{x} = (1/n)\sum_{i=1}^{n} x_i$, this last equation reduces to

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \qquad (6\text{-}4)$$

Sample Standard Deviation (s) = $\sqrt{s^2}$

12

12

Computation of s^2

Table 6-1 Calculation of Terms for the Sample Variance and Sample Standard Deviation

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	12.6	−0.4	0.16
2	12.9	−0.1	0.01
3	13.4	0.4	0.16
4	12.3	−0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	−0.4	0.16
8	13.1	0.1	0.01
	104.0	0.0	1.60

i	xi	xi^2
1	12.6	158.8
2	12.9	166.4
3	13.4	179.6
4	12.3	151.3
5	13.6	185.0
6	13.5	182.3
7	12.6	158.8
8	13.1	171.6
	104	1353.6

Example 6-3 We will calculate the sample variance and standard deviation using the shortcut method, Equation 6-4. The formula gives

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{1353.6 - \frac{(104)^2}{8}}{7} = \frac{1.60}{7} = 0.2286\,(\text{pounds})^2$$

and

$$s = \sqrt{0.2286} = 0.48 \text{ pounds}$$

These results agree exactly with those obtained previously.

13

Population Variance

- When population is finite and consists of N values, we may define **population variance** as

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} \qquad (6\text{-}5)$$

Sample variance is a reasonable estimate of Population variance.

Population Standard Deviation $\sigma = \sqrt{\sigma^2}$

14

14

Sample Range

- In addition to sample variance and sample standard deviation, **sample Range**, or the difference between the largest and smallest observations, is often a useful measure of variability.
- The **sample Range** is defined as follows.

If the n observations in a sample are denoted by $x_1, x_2, \ldots, x_n$, the **sample range** is

$$r = \max(x_i) - \min(x_i) \qquad (6\text{-}6)$$

x1 = 12.6,
x2 = 12.9,
x3 = 13.4,
x4 = 12.3,
x5 = 13.6,
x6 = 13.5,
x7 = 12.6, and
x8 = 13.1.

r = 13.6 – 12.3
= 1.3

15

15

Computation of s^2

- In most statistics problems, we work with a sample of observations selected from the population that we are interested in studying.
- Figure 6-3 illustrates relationship between Population and Sample.

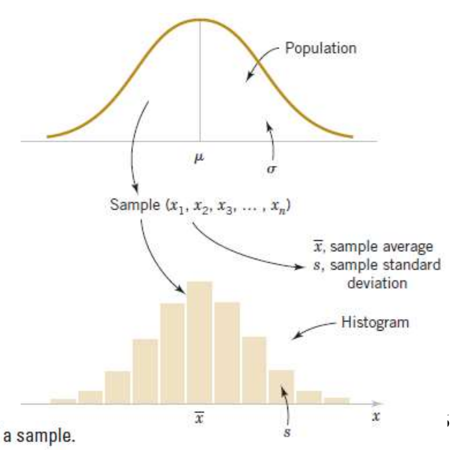


FIGURE 6-3 Relationship between a population and a sample.

16

DotPlot Dot Diagram

- The production of Bidri is a traditional craft of India.
- Bidri wares (bowls, vessels, and so on) are cast from an alloy containing primarily zinc along with some copper.
- Consider the following observations on copper content (%) for a sample of Bidri artifacts in London's Victoria and Albert Museum ("Enigmas of Bidri," *Surface Engr.*, 2005: 333–339), listed in increasing order:

2.0 2.4 2.5 2.6 2.6 2.7 2.7 2.8 3.0 3.1 3.2 3.3 3.3
3.4 3.4 3.6 3.6 3.6 3.6 3.7 4.4 4.6 4.7 4.8 5.3 10.1

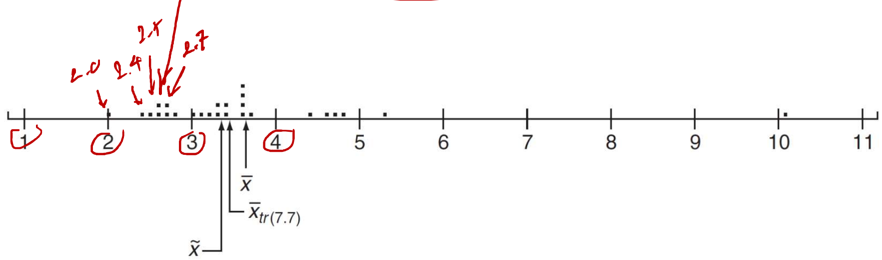


17

17

DotPlot

2.0 2.4 2.5 2.6 2.6 2.7 2.7 2.8 3.0 3.1 3.2 3.3 3.3
3.4 3.4 3.6 3.6 3.6 3.6 3.7 4.4 4.6 4.7 4.8 5.3 10.1



Dotplot of copper contents from Example 16

Figure 1.18

18

18

6-2 Stem-and-Leaf Diagrams

- **Dot diagram** is a useful data display for **small samples** up to about **20** observations.
- However, when the number of observations is moderately large, other graphical displays may be more useful.

TABLE • 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

19

19

7, 8, 10 stem

$x_1 = 7|0$ leaf
$x_2 = 7|2$
$x_3 = 7|3$
$x_4 = 8|2$
$x_5 = 8|3$
$x_6 = 10|2$
$x_7 = 10|5$

6-2 Stem-and-Leaf Diagrams

A **stem-and-leaf diagram** is a good way to obtain an informative visual display of a data set $x_1, x_2, \ldots, x_n$, where each number x_i consists of at least two digits. To construct a stem-and-leaf diagram, use the following steps.

Steps for Constructing a Stem-and-Leaf Diagram

(1) Divide each number x_i into two parts: a **stem**, consisting of one or more of the leading digits and a **leaf**, consisting of the remaining digit.
(2) List the stem values in a vertical column.
(3) Record the leaf for each observation beside its stem.
(4) Write the units for stems and leaves on the display.

Stem	Leaf	frequency
7	0 2 5	3
8	2 3	2
10	2 5	2

20

20

Example 6-4 Alloy Strength

- To illustrate the construction of a stem-and-leaf diagram, consider the alloy compressive strength data in Table 6-2.

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

- We will select as stem values the numbers *7, 8, 9, … , 24*.

22

22

Example 6-4 Alloy Strength

- The resulting stem-and-leaf diagram is presented in Fig. 6-4.

Figure 6-4 Stem-and-leaf diagram for compressive strength data in Table 6-2.

frequency count of the number of leaves associated with each stem

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Stem : Tens and hundreds digits (psi); Leaf: Ones digits (psi)

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

23

Example 6-4 Alloy Strength

- ❑ Inspection of this display immediately reveals that most of compressive strengths lie between 110 and 200 psi and that central value is somewhere between 150 and 160 psi.
- ❑ Furthermore, strengths are distributed approximately symmetrically about central value.
- ❑ Stem-and-leaf diagram enables us to determine quickly some important features of data that were not immediately obvious in original display in Table 6-2.

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Stem : Tens and hundreds digits (psi); Leaf: Ones digits (psi)

24

24

Stem-and-Leaf Diagrams

5 < 5L 0,1,2,3,4
5U 5,6,7,8,9

- ❑ In some data sets, providing more classes or stems may be desirable.
- ❑ One way to do this would be to modify original stems as follows: Divide stem 5 into two new stems, 5L and 5U.
  - ❑ Stem 5L has leaves 0, 1, 2, 3, and 4, and
  - ❑ stem 5U has leaves 5, 6, 7, 8, and 9.
- ❑ This will double the number of original stems.
- ❑ We could increase the number of original stems by four by defining five new stems:
  - ❑ 5z with leaves 0 and 1,
  - ❑ 5t (for twos and three) with leaves 2 and 3,
  - ❑ 5f (for fours and fives) with leaves 4 and 5,
  - ❑ 5s (for six and seven) with leaves 6 and 7, and
  - ❑ 5e with leaves 8 and 9.

5: z 0,1; t 2,3; f 4,5; s 6,7; e 8,9

25

25

6-2 Stem-and-Leaf Diagrams

Example 6-5 Chemical Yield

Figure 6-5 illustrates the stem-and-leaf diagram for 25 observations on batch yields from a chemical process. In Fig. 6-5(a) we have used 6, 7, 8, and 9 as the stems. This results in too few stems, and the stem-and-leaf diagram does not provide much information about the data. In Fig. 6-5(b) we have divided each stem into two parts, resulting in a display that more

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

(a)

26

26

6-2 Stem-and-Leaf Diagrams

❑Stem 5L has leaves 0, 1, 2, 3, and 4, and
❑stem 5U has leaves 5, 6, 7, 8, and 9.

❑5z with leaves 0 and 1,
❑5t (for twos and three) with leaves 2 and 3,
❑5f (for fours and fives) with leaves 4 and 5,
❑5s (for six and seven) with leaves 6 and 7, and
❑5e with leaves 8 and 9.

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

(a)

Stem	Leaf
6L	1 3 4
6U	5 5 6
7L	0 1 1 3
7U	5 7 8 8 9
8L	1 3 4 4
8U	7 8 8
9L	2 3
9U	5

(b)

Stem	Leaf
6z	1
6t	3
6f	4 5 5
6s	6
6e	
7z	0 1 1
7t	3
7f	5
7s	7
7e	8 8 9
8z	1
8t	3
8f	4 4
8s	7
8e	8 8
9z	
9t	2 3
9f	5
9s	
9e	

(c)

Figure 6-5 Stem-and leaf displays for Example 6-5.

Stem: Tens digits.

Leaf: Ones digits.

too many stems in plot, resulting in display that does not tell us muc about shape of data

27

27

Ordered stem-and-leaf diagram

มัธยฐาน (Median) = ตำแหน่ง $\frac{(n+1)}{2}$

$= \frac{80+1}{2} = 40.5$

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Stem : Tens and hundreds digits (psi); Leaf: Ones digits (psi)

เรียงลำดับ

แล้วจะมา

ค่ากลาง

37

33

$\frac{37+33+10}{80}$

Median = $\frac{160+163}{2}$ (*)

ordered stem-and-leaf diagram

Character Stem-and-Leaf Display

Stem-and-leaf of Strength

N = 80 Leaf Unit = 1.0

1	7	6
2	8	7
3	9	7
5	10	1 5
8	11	0 5 8
11	12	0 1 3
17	13	1 3 3 4 5 5
25	14	1 2 3 5 6 8 9 9
37	15	0 0 1 3 4 4 6 7 8 8 8 8
(10)	16	0 0 0 3 3 5 7 7 8 9
33	17	0 1 1 2 4 4 5 6 6 8
23	18	0 0 1 1 3 4 6
16	19	0 3 4 6 9 9
10	20	0 1 7 8
6	21	8
5	22	1 8 9
2	23	7
1	24	5

1 2 3 4 5

the 40th and 41st values of strength

sample median

Figure 6-6 Stem-and-leaf diagram from Minitab.

29

Ordered stem-and-leaf diagram

- Ordered stem-and-leaf display makes it relatively easy to find **data features** such as
  - percentiles,
  - quartiles, and the
  - median.

30

Ordered stem-and-leaf diagram

Data Features

- **Sample Median** is measure of central tendency that divides data into two equal parts, half below median and half above.

If number of observations is even, the median is halfway between the two central values. ตำแหน่ง 40.5

From Fig. 6-6, the 40th and 41st values of strength as 160 and 163, so the median is (160 + 163)/2 = 161.5.

If the number of observations is odd, the median is the *central* value.

พิสัย — Max - Min.

Range is a measure of variability that can be easily computed from the ordered stem-and-leaf display. It is the maximum minus the minimum measurement.
From Fig.6-6 the range is 245 - 76 = 169.

31

Ordered stem-and-leaf diagram

Data Features

25% q_1 — 50% q_2 — q_3 75%

- When **ordered** set of data is **divided** into **four** equal parts, division points are called **quartiles.**

ตำแหน่ง $(n+1)(\frac{1}{4})$

First or **lower quartile**, q_1, is value that has approximately one-fourth (25%) of observations **below** it and approximately 75% of the observations **above**.

ตำแหน่ง $(n+1)(\frac{1}{2})$

Second quartile, q_2, has approximately one-half (50%) of observations below its value.

The second quartile is *exactly* equal to **median**.

ตำแหน่ง $(n+1)(3/4)$

Third or **upper quartile**, q_3, has approximately three-fourths (75%) of observations below its value. As in the case of the median, the quartiles may not be unique.

32

32

6-2 Stem-and-Leaf Diagrams

$q_1 =$ مرتبة $(80+1)(\frac{1}{4}) = 20.25$

$\boxed{143}$ — 145

$143 + \frac{(145-143)}{4}$

$143 + 0.5$

$\boxed{143.5}$

- compressive strength data in Figure 6-6 contains
- $n = 80$ observations.
- Minitab software calculates the first and third quartiles as the $(n + 1)/4$ and $3(n + 1)/4$ ordered observations and interpolates as needed.
- For example, $(80 + 1)/4 = 20.25$ and $3(80 + 1)/4 = 60.75$.
- Therefore, Minitab interpolates between the 20th and 21st ordered observation to obtain $q_1 = 143.50$ and between the 60th and 61st observation to obtain $q_3 = 181.00$.

$q_3 =$ مرتبة $(80+1)(\frac{3}{4}) = 60.75$

181 ↑ 181

Stem-and-leaf of Strength

$N = 80$	Leaf	Unit = 1.0
1	7	6
2	8	7
3	9	7
5	10	1 5
8	11	0 5 8
11	12	0 1 3
17	13	1 3 3 4 5 5
25	14	1 2 3 5 6 8 9 9
37	15	0 0 1 3 4 4 6 7 8 8 8 8
(10)	16	0 0 0 3 3 5 7 7 8 9
33	17	0 1 1 2 4 4 5 6 6 8
23	18	0 0 1 1 3 4 6
16	19	0 3 4 6 9 9
10	20	0 1 7 8
6	21	8
5	22	1 8 9
2	23	7
1	24	5

33

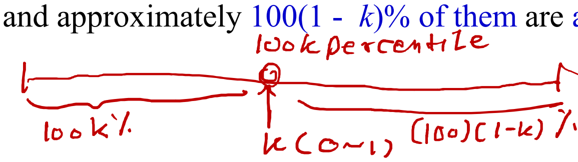
6-2 Stem-and-Leaf Diagrams

Data Features

• **Interquartile range (IQR)** is the difference between the upper and lower quartiles, and it is sometimes used as a measure of variability.

$$\boxed{IQR = q3 - q1}$$

• In general, the 100*k*th **percentile** is data value such that approximately 100*k*% of the observations are at or below this value and approximately 100(1 - *k*)% of them are above it.



34

34

6-3 Frequency Distributions and Histograms

- **Frequency distribution** is a more compact summary of data than a stem-and-leaf diagram.

- To construct a frequency distribution, we must **divide range** of the data into **intervals**, which are usually called **class intervals**, **cells**, or **bins**.

อันตรภาคชั้น ความถี่ ความถี่สัมพัทธ์

Constructing a Histogram (Equal Bin Widths):

(1) Label the bin (class interval) boundaries on a horizontal scale.
(2) Mark and label the vertical scale with the frequencies or the relative frequencies.
(3) Above each bin, draw a rectangle where height is equal to the frequency (or relative frequency) corresponding to that bin.

35

35

Frequency Distributions and Histograms

- **Number of bins** depends on the number of observations and the amount of scatter or dispersion in data.
- Frequency distribution that uses either too few or too many bins will not be informative.
- We usually find that between 5 and 20 bins is satisfactory in most cases and that number of bins should increase with *n*.
- Several sets of rules can be used to determine the number of bins in **Histogram.**
- However, choosing **number of bins approximately equal to square root of the number of observations** often works well in practice.

จำนวนอันตรภาคชั้น = $\sqrt{n}$ ⊛

36

36

จำนวน Class Interval = $\sqrt{80} \approx 8.944 \Rightarrow 8, 9$; พิสัย (Range) = 245 - 76 = 169

TABLE • 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

$\frac{180}{9} = 20$ ความกว้างของอันตรภาคชั้น

180

❑ **Relative frequencies** are found by dividing the **observed frequency in each bin** by total number of observations.

70 76 169 245 250

TABLE • 6-4 Frequency Distribution for the Compressive Strength Data in Table 6-2

Class	$70 \le x < 90$	$90 \le x < 110$	$110 \le x < 130$	$130 \le x < 150$	$150 \le x < 170$	$170 \le x < 190$	$190 \le x < 210$	$210 \le x < 230$	$230 \le x < 250$
Frequency	2	3	6	14	22	17	10	4	2
Relative frequency	0.0250	0.0375	0.0750	0.1750	0.2750	0.2125	0.1250	0.0500	0.0250
Cumulative relative frequency	0.0250	0.0625	0.1375	0.3125	0.5875	0.8000	0.9250	0.9750	1.0000

2/80 3/80

ความถี่สัมพัทธ์สะสม

38

6-3 Frequency Distributions and Histograms

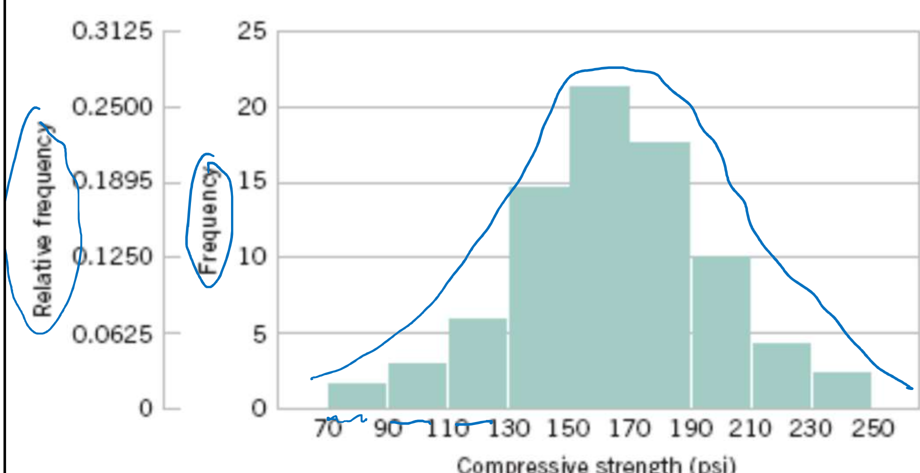


Figure 6-7 Histogram of compressive strength for 80 aluminum-lithium alloy specimens.

39

39

6-3 Frequency Distributions and Histograms

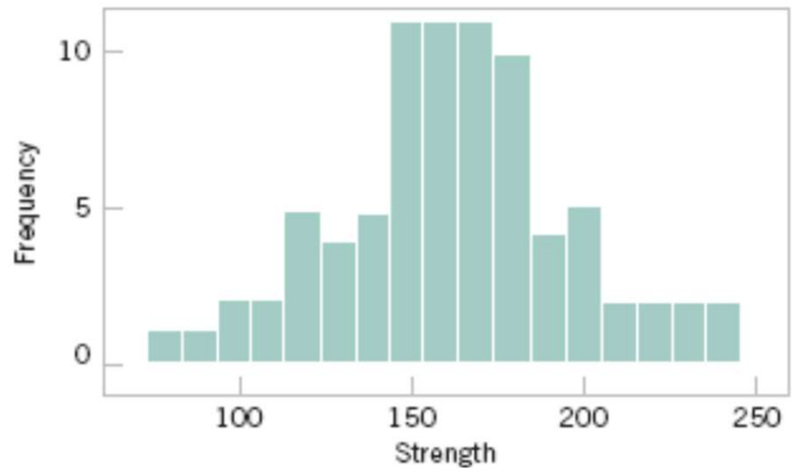


Figure 6-8 A histogram of the compressive strength data from Minitab with 17 bins.

40

40

6-3 Frequency Distributions and Histograms

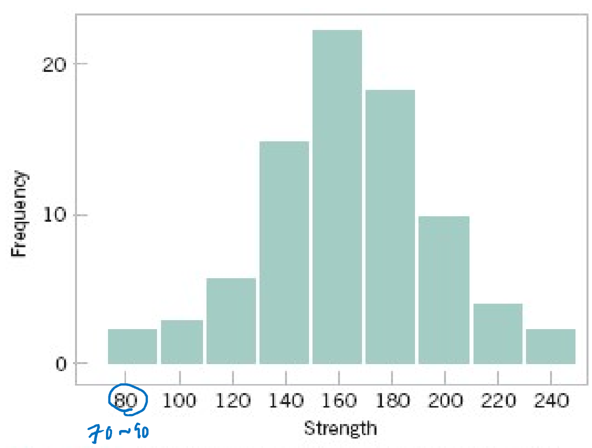


Figure 6-9 A histogram of the compressive strength data from Minitab with nine bins.

Figure 6-9 A histogram of the compressive strength data from Minitab with nine bins.

41

41

6-3 Frequency Distributions and Histograms

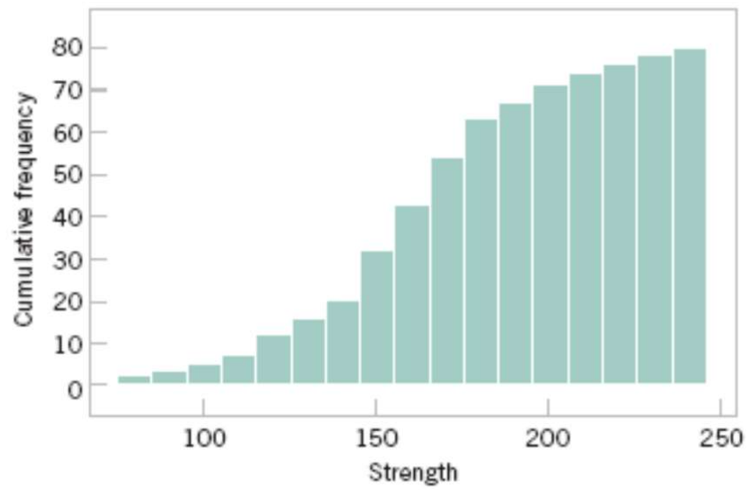


Figure 6-10 A cumulative distribution plot of the compressive strength data from Minitab.

42

42

6-3 Frequency Distributions and Histograms

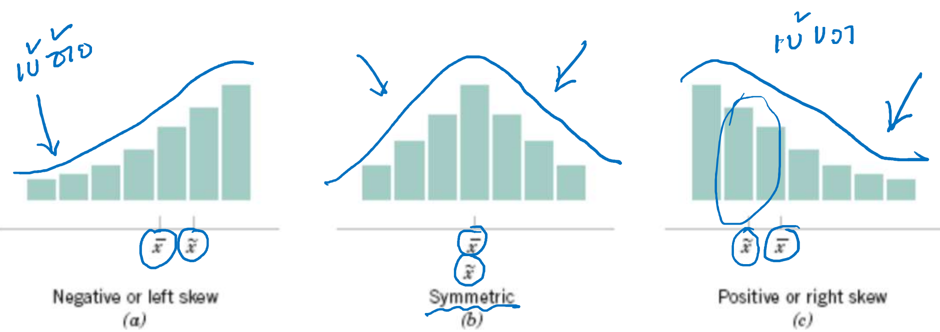


Figure 6-11 Histograms for symmetric and skewed distributions.

43

43

6-4 Box Plots

- Stem-and-leaf display and Histogram provide general visual impressions about data set, but numerical quantities such as x_bar or s provide information about only one feature of data.
- **Box plot** is graphical display that simultaneously describes several important features of a data set, such as
  - center,
  - spread,
  - departure from symmetry, and
  - identification of observations that lie unusually far from the bulk of the data.

q_1 q_2 q_3

ปกติ

- **Whisker**
- **Outlier** ผิดปกติ
- **Extreme outlier**

ผิดปกติมาก

44

44

6-4 Box Plots

q_1, q_3, IQR

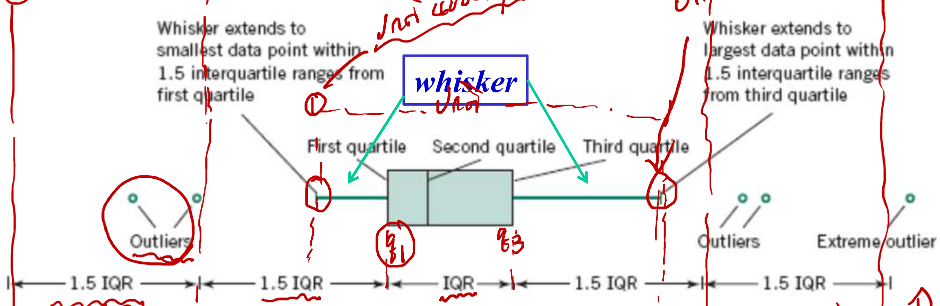
> ต่ำ

> ต่ำ

ปกติ

ปกติ แต่น้อยที่สุด (Data)

ปกติ มากที่สุด (ปกติ)



ผิดปกติ

ผิดปกติ

ผิดปกติมาก

Figure 6-13 Description of a box plot.

47

47

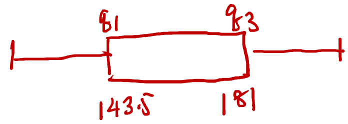
Character Stem-and-Leaf Display

Stem-and-leaf of Strength

N = 80   Leaf Unit = 1.0

1	7	6
2	8	7
3	9	7
5	10	1 5
8	11	0 5 8
11	12	0 1 3
17	13	1 3 3 4 5 5
25	14	1 2 3 5 6 8 9 9
37	15	0 0 1 3 4 4 6 7 8 8 8 8
(10)	16	0 0 0 3 3 5 7 7 8 9
33	17	0 1 1 2 4 4 5 6 6 8
23	18	0 0 1 1 3 4 6
16	19	0 3 4 6 9 9
10	20	0 1 7 8
6	21	8
5	22	1 8 9
2	23	7
1	24	5

142 ~ 145



the first and third quartiles as the$(n + 1)/4$ and $3(n + 1)/4$

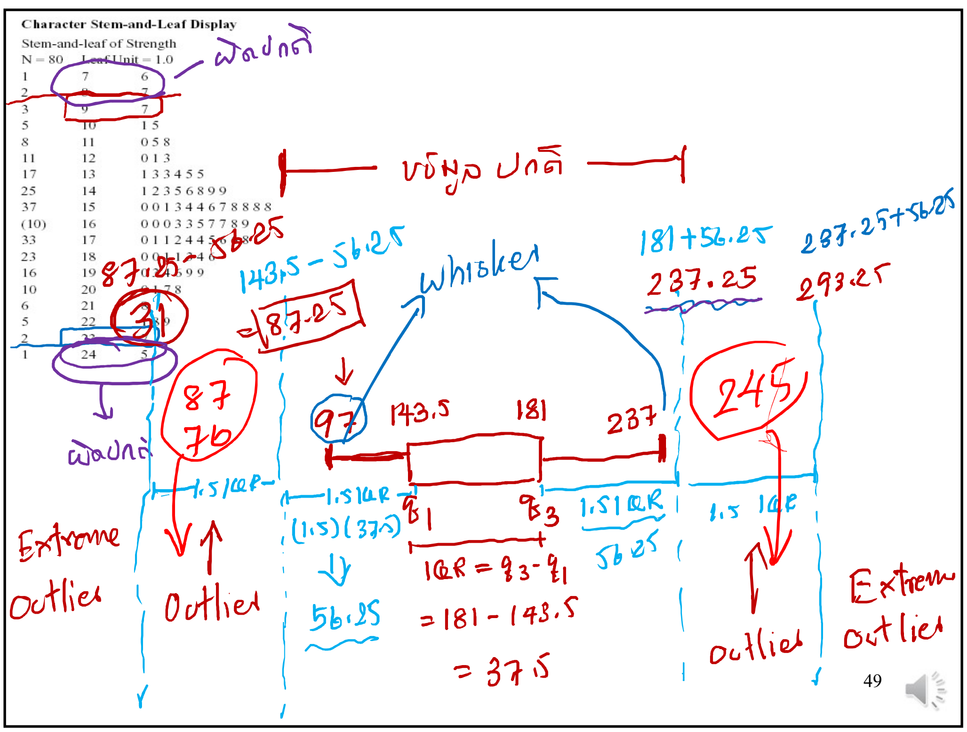
$(80 + 1)/4 = 20.25$ and $3(80 + 1)/4 = 60.75$.

$q_1 = 143.50$, $q_3 = 181.00$.

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

48



49

6-4 Box Plots

the first and third quartiles as the$(n + 1)/4$ and $3(n + 1)/4$

$(80 + 1)/4 = 20.25$ and $3(80 + 1)/4 = 60.75$.

$q_1 = 143.50$, $q_3 = 181.00$.

$q_3 + 1.5\text{IQR} = 181 + 1.5(181 - 143.5) = 237.25.$

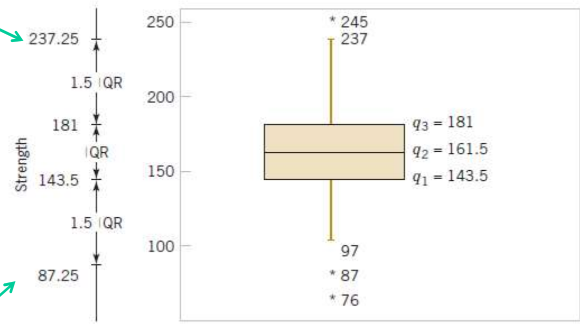
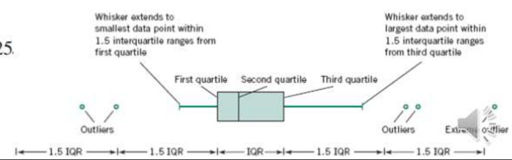


FIGURE 6-14 Box plot for compressive strength data in Table 6-2.

$q_1 - 1.5\text{IQR} = 143.5 - 1.5(181 - 143.5) = 87.25$



50

52

52