CH.6 Random Sampling and Descriptive Statistics

- Population vs Sample
- Random sampling
- Numerical summaries :
  - sample mean, sample variance, sample range
- Stem-and-Leaf Diagrams
  - Median, quartiles, percentiles, mode, interquartile range (IQR)
- Frequency distributions and histograms
- Box plots
  - Whisker, outlier
- Time-sequence plots
- Probability plots

Population

- The collection of things (parts, people, services) -- called “members” -- under study
- The letter N is usually defined to be the number of members in the population

Examples of Populations

- Students in INE2002 (N ~ 100)
- Users of a software package (N ~ ?)
- Angioplasty procedures during 2009 at a spesific hospital (N = 1523)
- A week’s (April 6 - 12, 2009) stampings of part #ZG76 at autobody plant (N = 4501)

Sample

- Measurement of only a subset of the population . These will be used to say something about the variables of the entire population.
- The letter n (the "*sample size*") is usually used to represent the number of items in in this subset

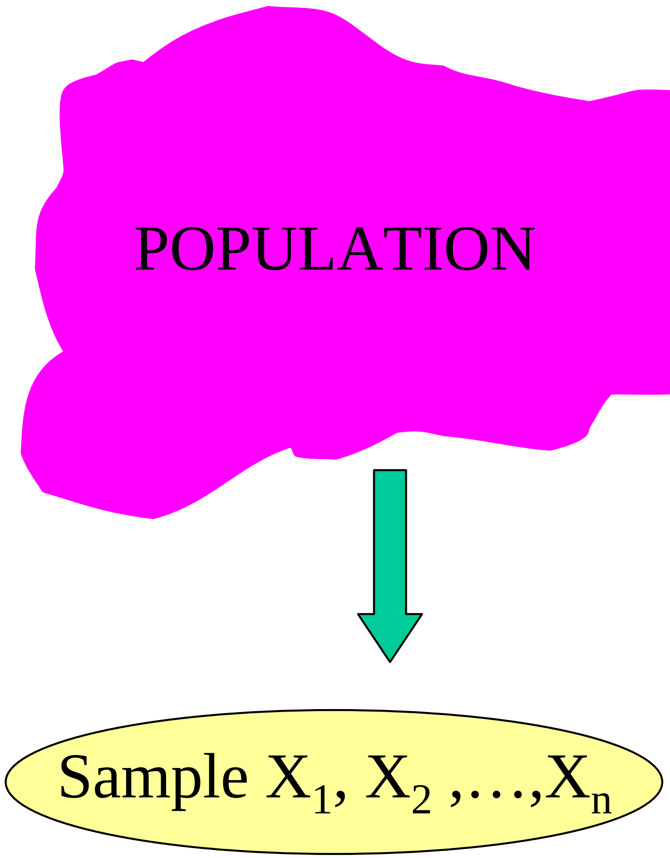
Examples of Samples

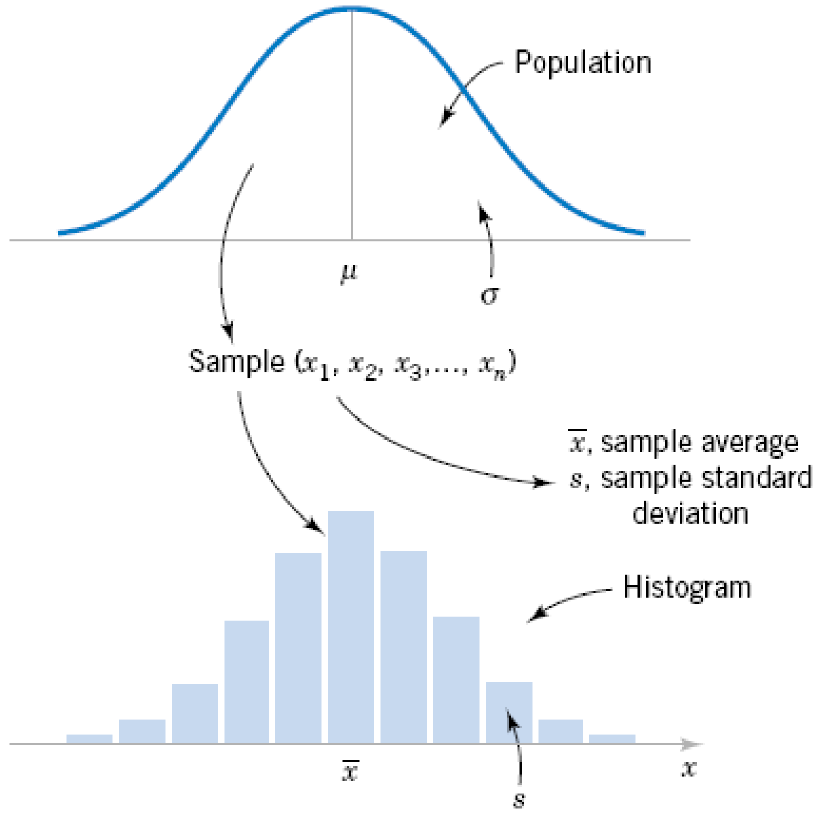
- Asking only 20 (out of 100) INE2002 students the current value of their GPA’s
- Surveying only some of a software package’s users
- Getting detailed angioplasty data only for procedures done on Mondays
- Measuring one auto panel out of every 100 produced

Why Use Samples?

- In most situations, it is impossible or impractical to observe the entire population.
- Impractical: it would be time consuming and expensive
- Impossible: some (perhaps many) of the members of the population do not yet exist at the time a decision is to be made,
- Ex: we could not test the tensile strength of all the chassis structural elements
- So generally, we must view the population as **conceptual.**
- Therefore, we depend on a subset of observations from the population to help make decisions about the population.

Population vs Sample





Random Sampling

- For statistical methods to be valid, the sample must be representative of the population. It is often tempting to select the observations that are most convenient as the sample.
- Otherwise, the parameter of interest will be consistently underestimated (or overestimated). Furthermore, the behavior of a judgment sample cannot be statistically described.
- To avoid these difficulties, it is desirable to select a **random sample** as the result of some chance mechanism:
- The selection of a sample is a random experiment and each observation in the sample is the observed value of a random variable.
- The observations in the population determine the probability distribution of the random variable.
- To define a random sample, let X be a random variable that represents the result of one selection of an observation from the population.

Random Sampling

The random variables $X_1, X_2, \ldots, X_n$ are a random sample of size *n* if

(a) the X_i's are independent random variables, and

(b) every X_i has the same probability distribution.

Example:

- Suppose, we are investigating the effective service life of an electronic component used in a cardiac pacemaker (kalp pili) and that component life is normally distributed.
- Then we would expect each of the observations on component life in a random sample of *n* components to be independent random variables with exactly the same normal distribution.

6-1 Numerical Summaries

Describe data features numerically

Ex: characterize the central tendency in the data by arithmetic average which is refered as sample mean

Other examples: sample variance, sample standard deviation, sample range

Definition: Sample Mean

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

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

6-1 Numerical Summaries

Example 6-1

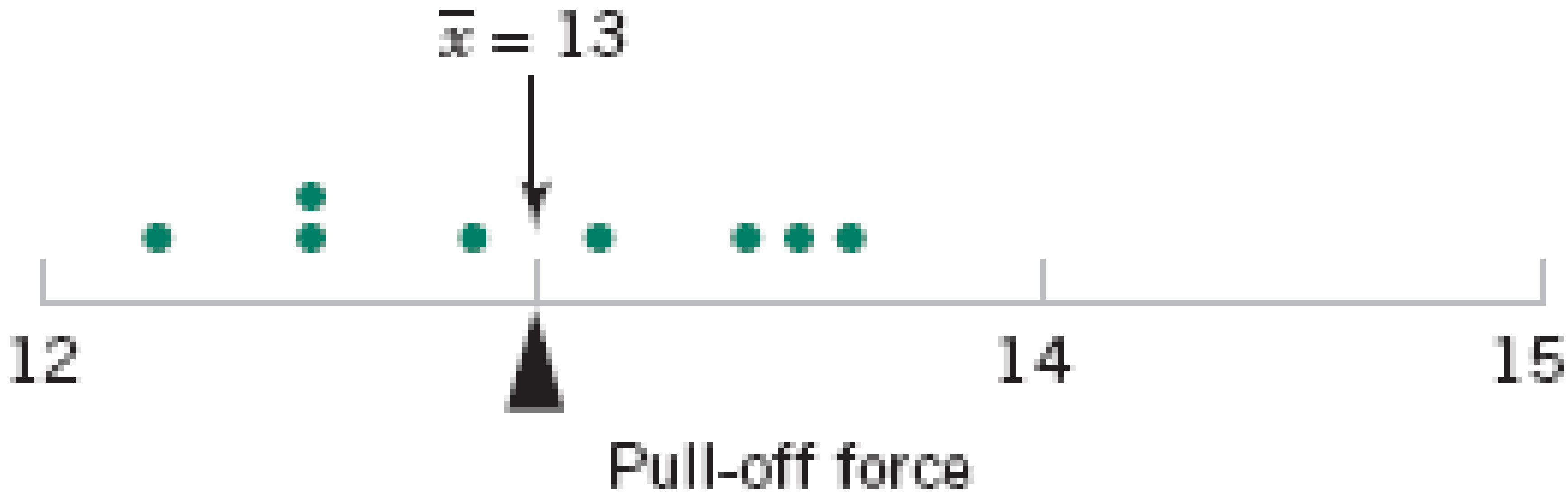
Let's consider the eight observations collected from the prototype engine connectors from Chapter 1. The eight observations are $x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. The sample mean is

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

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

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.

6-1 Numerical Summaries



The sample mean as a balance point for a system of weights.

6-1 Numerical Summaries

Population Mean

For a finite population with *N* measurements, the mean is

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

The sample mean is a reasonable estimate of the population mean.

6-1 Numerical Summaries

Definition: Sample Variance

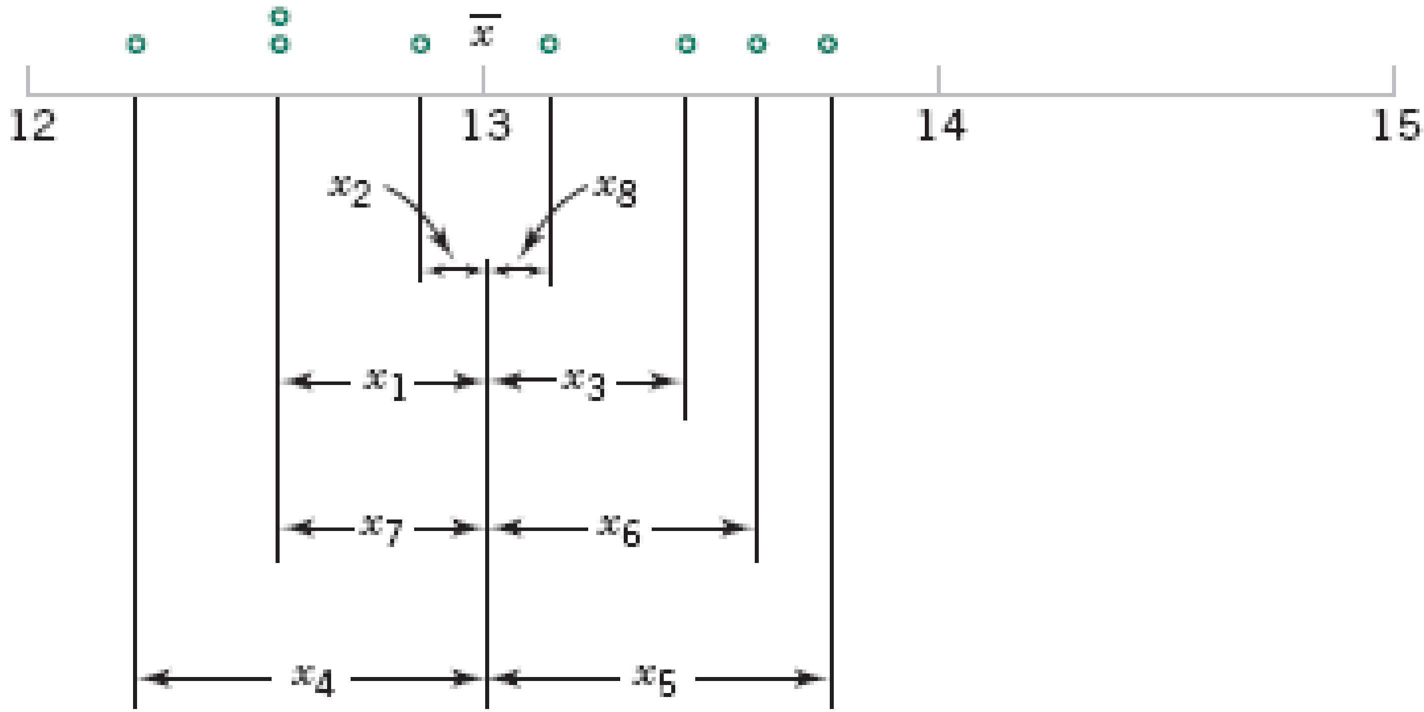
If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the **sample variance** is

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

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

6-1 Numerical Summaries

How does the Sample Variance Measure Variability through the deviations $x_i - \overline{x}$ **?**



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

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

so the sample variance is

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

and the sample standard deviation is

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

6-1 Numerical Summaries

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

6-1 Numerical Summaries

Computation of s^2

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

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

Remember

$$\overline{x} = (1/n)\sum_{i=1}^{n} x_i$$

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

Shortcut method to compute *s*

6-1 Numerical Summaries

Population Variance

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

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

The sample variance is a reasonable estimate of the population variance.

6-1 Numerical Summaries

Definition

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

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

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.

6-2 Stem-and-Leaf Diagrams

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 max	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 min	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

psi: pounds per square inch

6-2 Stem-and-Leaf Diagrams

From the diagram

- Most of the data lie between 110 and 200 psi
- A central value is somewhere between 150 and 160 psi
- The data are distributed approximately symmetrically about the central value

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)

6-2 Stem-and-Leaf Diagrams

25 observations on batch yields from a chemical process

Stem: Tens digits.

Leaf: Ones digits.

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)

6-2 Stem-and-Leaf Diagrams - ordered

Easier to find

- percentiles
- quartiles
- median

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

6-2 Stem-and-Leaf Diagrams

Data Features : median, range, quartiles

The **median,** $\tilde{x}$, is a measure of central tendency that divides the data into two equal parts, half below the median and half above. If the number of observations is even, the median is halfway between the two central values.

In the 80 compressive strength data, the 40th and 41st values of strength are 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.

The **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 the figure, the range is 245 - 76 = 169.

6-2 Stem-and-Leaf Diagrams

Data Features : median, range, quartiles, interquartile range, mode

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

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

The **second quartile**, q_2, has approximately one-half (50%) of the observations below its value. The second quartile is *exactly* equal to the **median**.

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

6-2 Stem-and-Leaf Diagrams

Data Features : median, range, quartiles, interquartile range, mode

- The compressive strength data contains n = 80 observations. The first and third quartiles (q_1 and q_3) are calculated as the

  (n + 1)/4 and 3(n + 1)/4 ordered observations and interpolated as needed.

- For example, (80 + 1)/4 = 20.25 and 3(80 + 1)/4 = 60.75.

- q_1 is interpolated between the 20th and 21st ordered observation
  q_1 = [(145-143)/(21-20)]*(20.25-20)+143 = 143.50

- q_3 is interpolated between the 60th and 61st ordered observation
  q_3 = [(181-181)/(61-60)]*(60.75-60)+181 = 181.00

6-2 Stem-and-Leaf Diagrams

Data Features : median, range, quartiles, interquartile range, mode

• The **interquartile range** is the difference between the upper and lower quartiles, and it is sometimes used as a measure of variability.
IQR=$q_3 - q_1$= 181-143.5 = 37.5

• In general, the 100*k*th **percentile** is a 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.

•The **sample mode** is the most frequently occuring data value.
Mode is 158 in the compressive strength data.

Stem-and-Leaf Exercise 6.15 (6.23)

70 data: Numbers of cycles to failure of aluminum test coupons subjected to repeated alternating stress at 21000 psi, 18 cycles per second

1115	2130	1674	2265	1260	1730	1535
1310	1421	1016	1910	1888	1102	1781
1540	1109	1102	1018	1782	1578	1750
1502	1481	1605	1452	1522	758	1501
1258	1567	706	1890	1792	1416	1238
1315	1883	2215	2100	1000	1560	990
1085	1203	785	1594	1820	1055	1468
798	1270	885	2023	1940	1764	1512
1020	1015	1223	1315	1120	1330	1750
865	845	375	1269	910	1608	1642

Stem-and-Leaf Exercise 6.15 (6.23)

- Median = 1436.5
- Q1 = 1097.8
- Q3 = 1735.0

unit = 100   1|2 represents 1200

```
  1     0T|3
  1     0F|
  5     0S|7777
 10     0o|88899
 22     1*|000000011111
 33     1T|22222223333
(15)    1F|444445555555555
 22     1S|66667777777
 11     1o|888899
  5     2*|011
  2     2T|22
```

Stem-and-Leaf Exercise 6.16 (6.24)

64 data: The percentage of cotton in material used to manufacture men's shirts

34,2	37,8	33,6	32,6	33,8	35,8	34,7	34,6
33,1	36,6	34,7	33,1	34,2	37,6	33,6	33,6
34,5	35,4	35	34,6	33,4	37,3	32,5	34,1
35,6	34,6	35,4	35,9	34,7	34,6	34,1	34,7
36,3	33,8	36,2	34,7	34,6	35,5	35,1	35,7
35,1	37,1	36,8	33,6	35,2	32,8	36,8	36,8
34,7	34	35,1	32,9	35	32,1	37,9	34,3
33,6	34,1	35,3	33,5	34,9	34,5	36,4	32,7

Stem-and-Leaf Exercise 6.16 (6.24)

- Median = 34.7
- Q1 = 33.8
- Q3 = 35.575

Leaf Unit = 0.10   32|1 represents 32.1%

```
   1    32|1
   6    32|56789
   9    33|114
  17    33|56666688
  24    34|0111223
 (14)   34|55666667777779
  26    35|001112344
  17    35|56789
  12    36|234
   9    36|6888
   5    37|13
   3    37|689
```

6-3 Frequency Distributions and Histograms

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

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

•In practice # bin= $\sqrt{n}$ where n is the sample size

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.

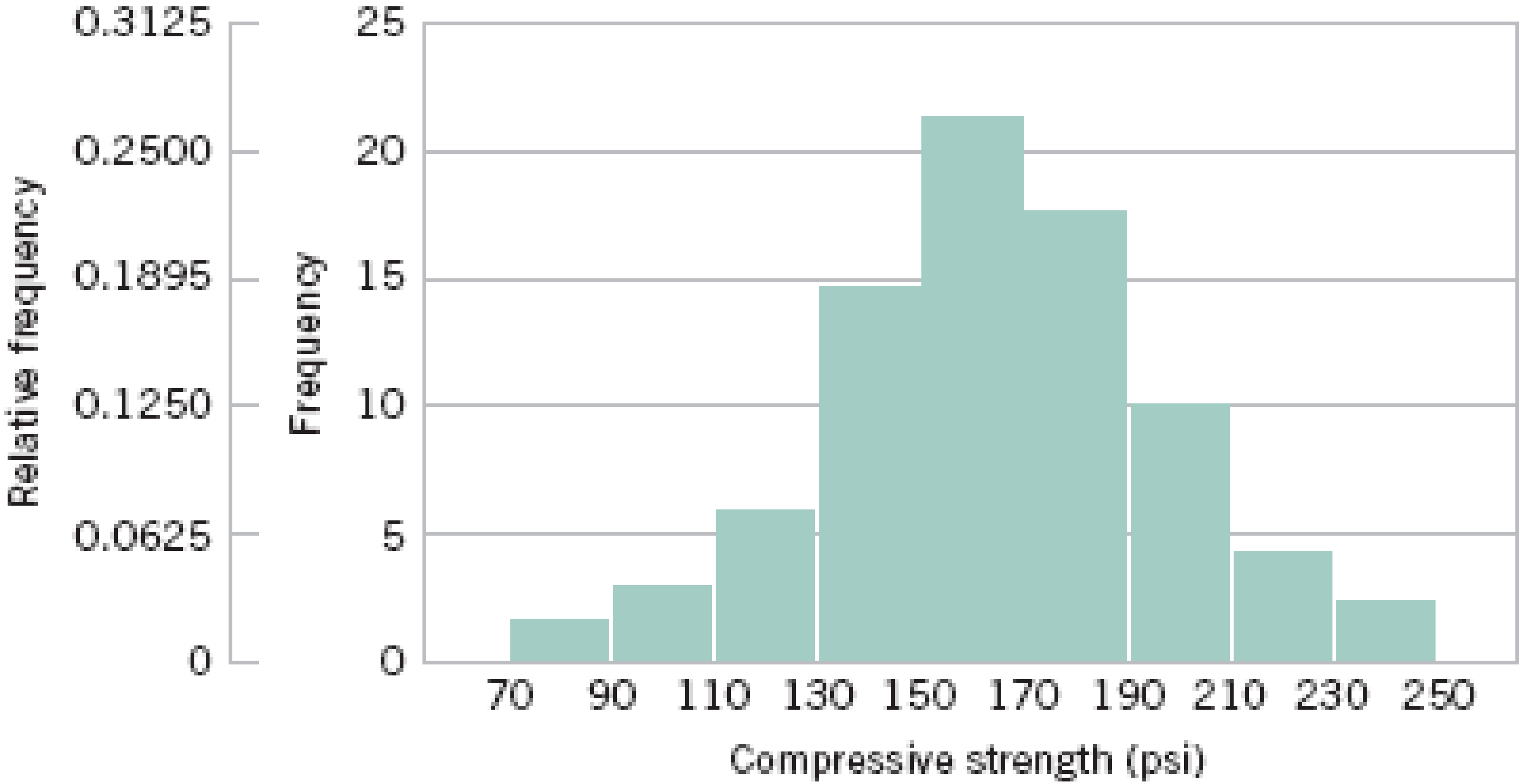
6-3 Frequency Distributions and Histograms

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

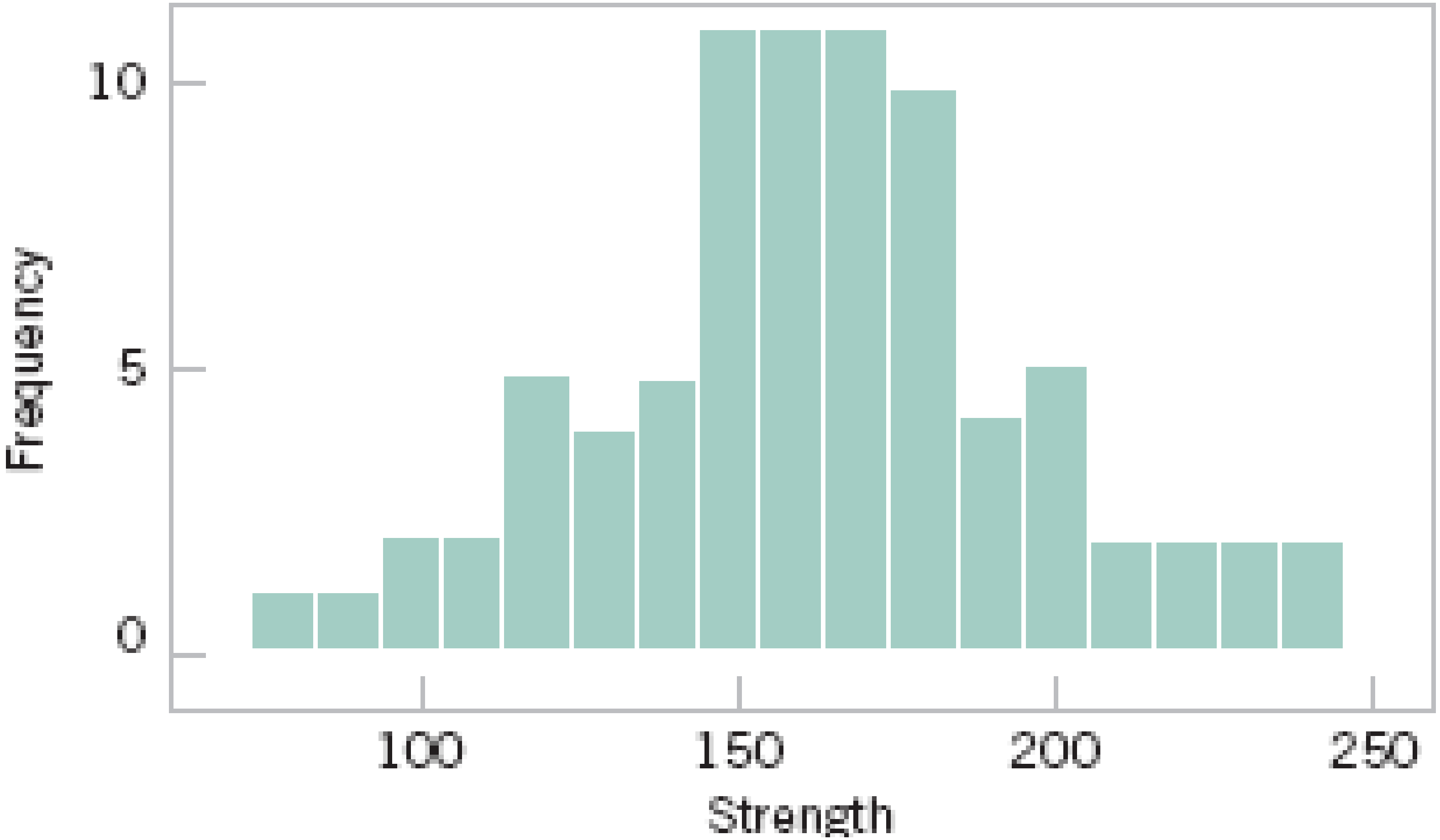
Frequency Distribution of compressive strength for 80 aluminum-lithium alloy specimens.

6-3 Frequency Distributions and Histograms



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

6-3 Frequency Distributions and Histograms



A histogram of the compressive strength data from Minitab with 17 bins.

6-3 Frequency Distributions and Histograms

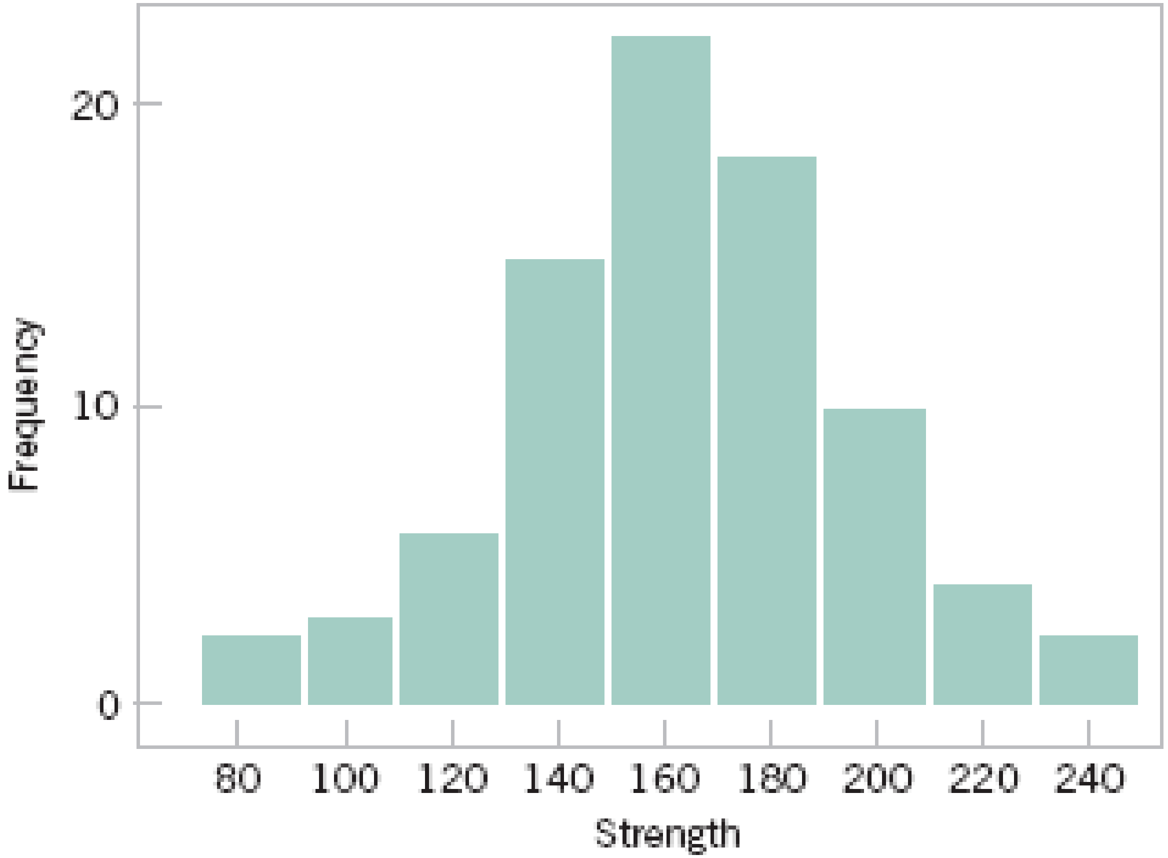
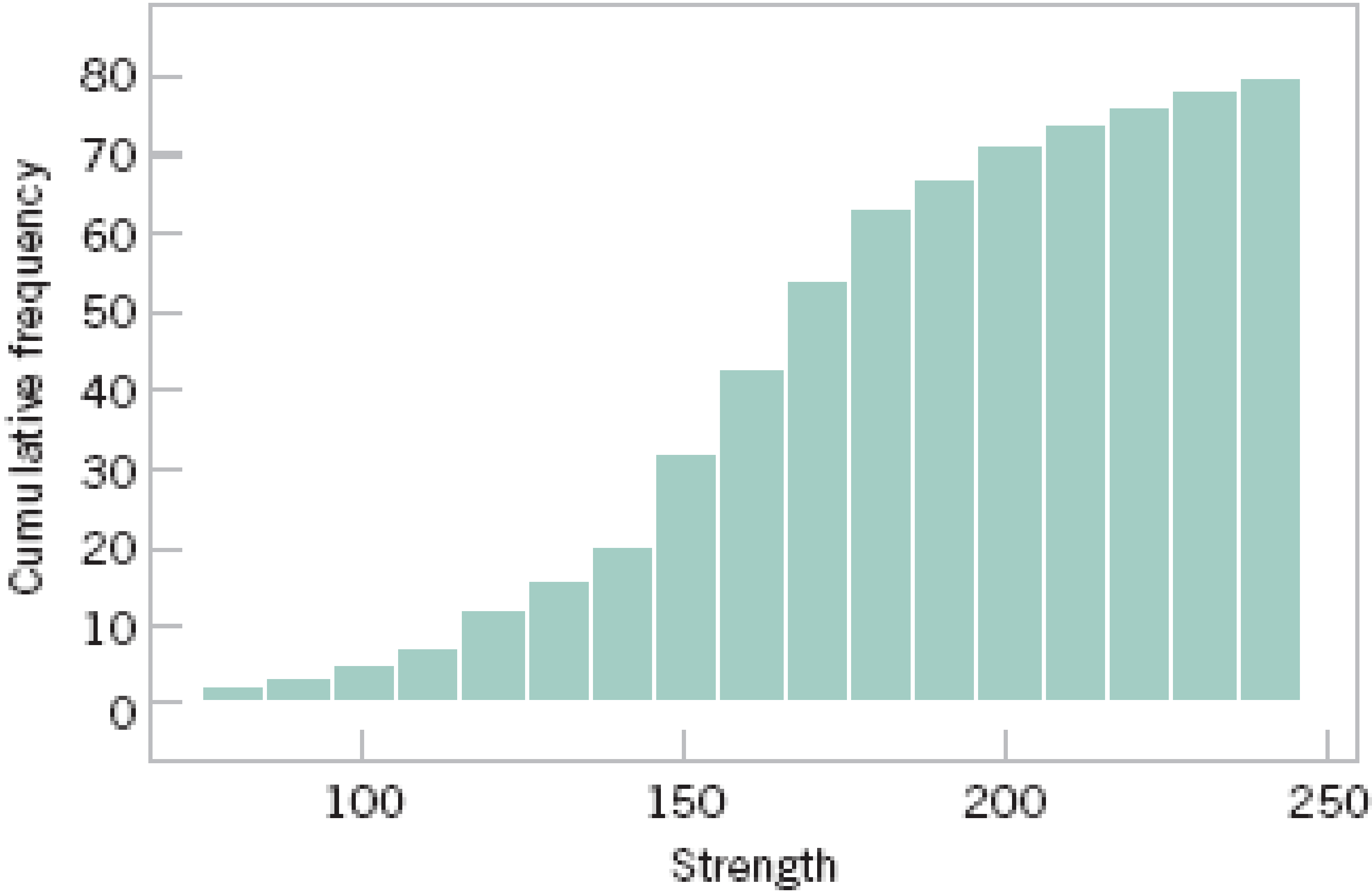


Figure 6-9 A histogram of the compressive strength data from Minitab with nine bins.

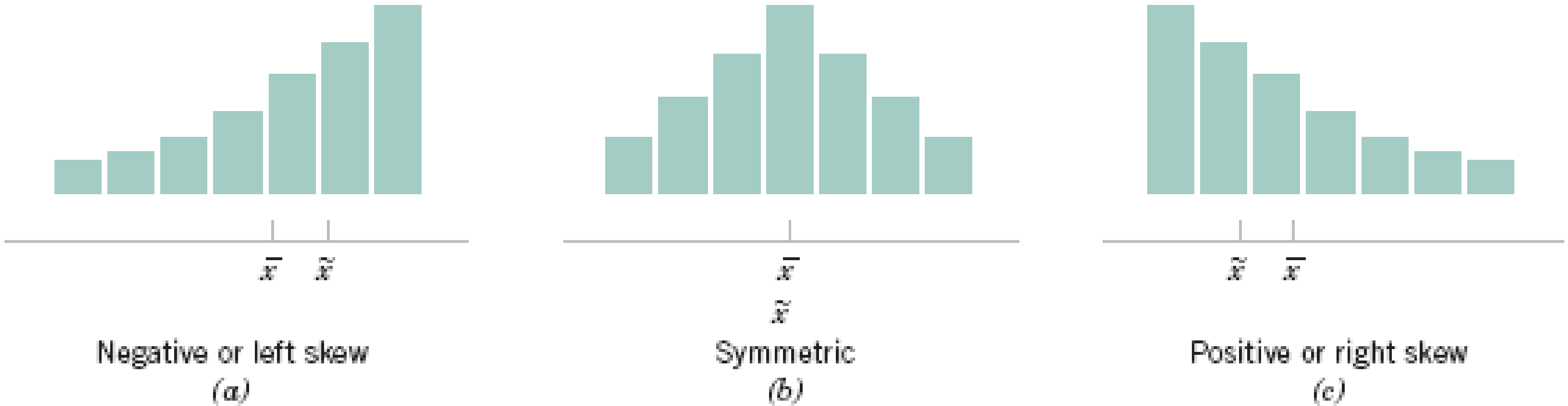
A histogram of the compressive strength data from Minitab with nine bins.

6-3 Frequency Distributions and Histograms



A cumulative distribution plot of the compressive strength data from Minitab.

6-3 Frequency Distributions and Histograms



Histograms for symmetric and skewed distributions.

6-3 Frequency Distributions and Histograms

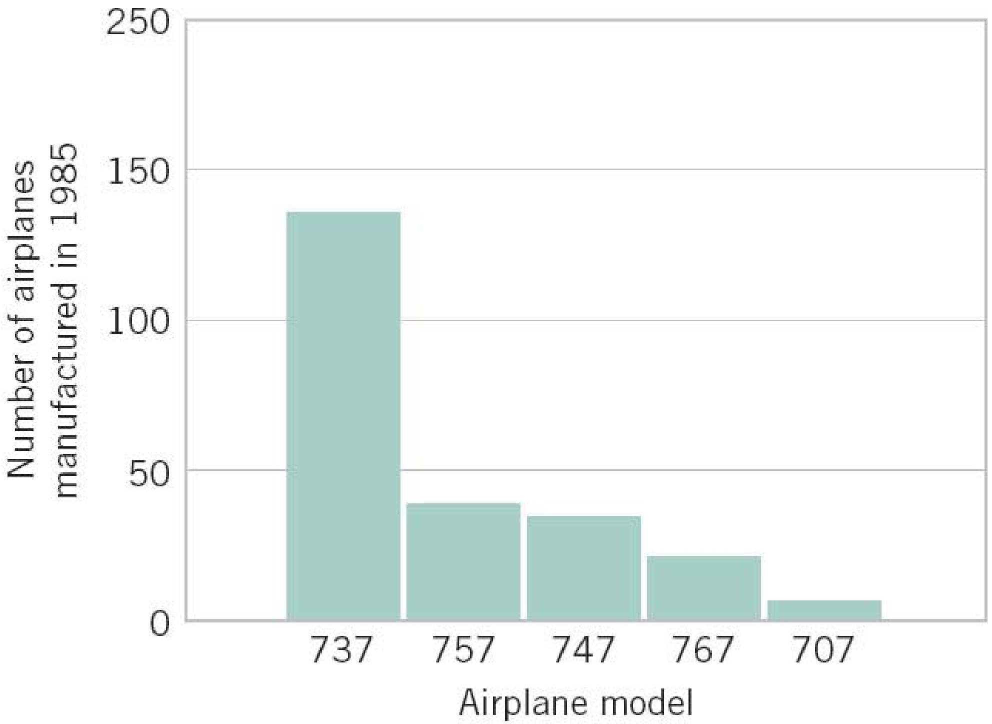


Figure 6-12
Airplane production in 1985. (*Source:* Boeing Company.)

Histograms for categorical data

Pareto charts can also be used

Exercise 6.32 (6.40)

64 data: The percentage of cotton in material used to manufacture men's shirts

34,2	37,8	33,6	32,6	33,8	35,8	34,7	34,6
33,1	36,6	34,7	33,1	34,2	37,6	33,6	33,6
34,5	35,4	35	34,6	33,4	37,3	32,5	34,1
35,6	34,6	35,4	35,9	34,7	34,6	34,1	34,7
36,3	33,8	36,2	34,7	34,6	35,5	35,1	35,7
35,1	37,1	36,8	33,6	35,2	32,8	36,8	36,8
34,7	34	35,1	32,9	35	32,1	37,9	34,3
33,6	34,1	35,3	33,5	34,9	34,5	36,4	32,7

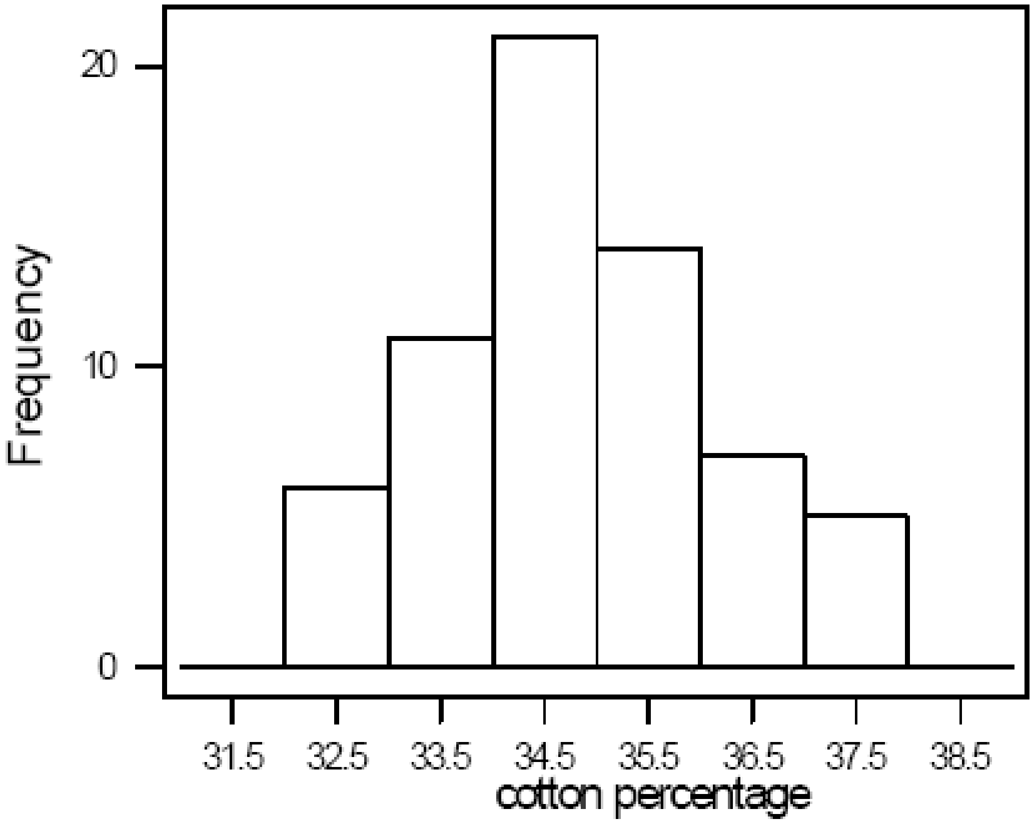
Frequency Distributions Exercise 6.32 (6.40)

Frequency Tabulation for Exercise 6-16.Cotton content

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		31.0		0	.0000	0	.0000
1	31.0	32.0	31.5	0	.0000	0	.0000
2	32.0	33.0	32.5	6	.0938	6	.0938
3	33.0	34.0	33.5	11	.1719	17	.2656
4	34.0	35.0	34.5	21	.3281	38	.5938
5	35.0	36.0	35.5	14	.2188	52	.8125
6	36.0	37.0	36.5	7	.1094	59	.9219
7	37.0	38.0	37.5	5	.0781	64	1.0000
8	38.0	39.0	38.5	0	.0000	64	1.0000
above	39.0			0	.0000	64	1.0000

Mean = 34.798 Standard Deviation = 1.364 Median = 34.700

Histogram for Exercise 6.32 (6.40)

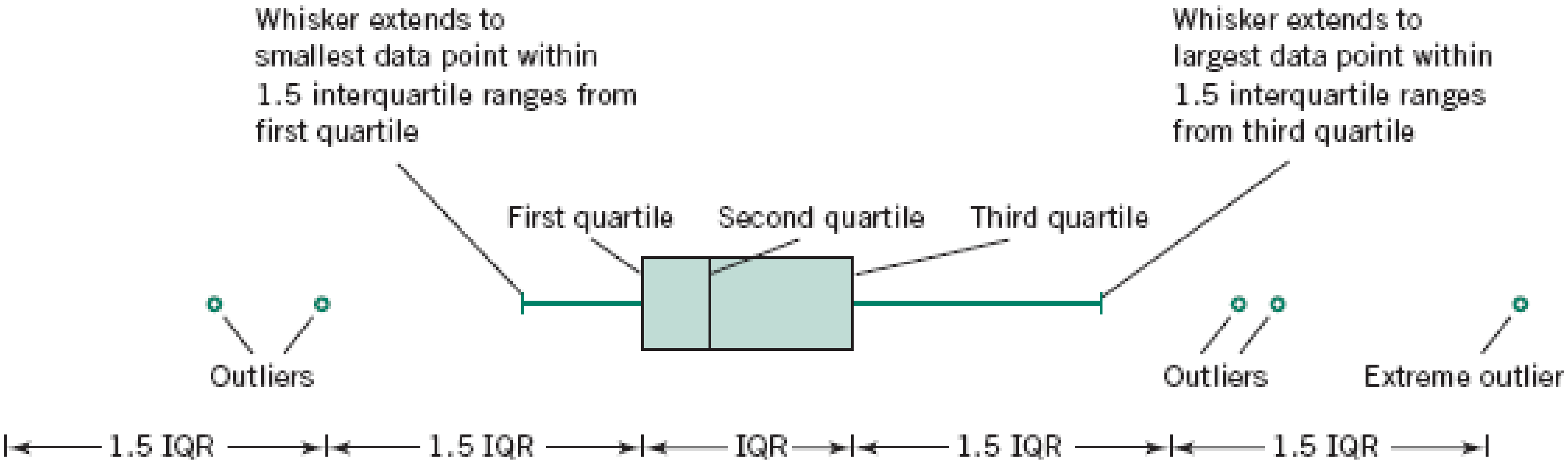


6-4 Box Plots

- The **box plot** is a 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 (*outliers*).

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

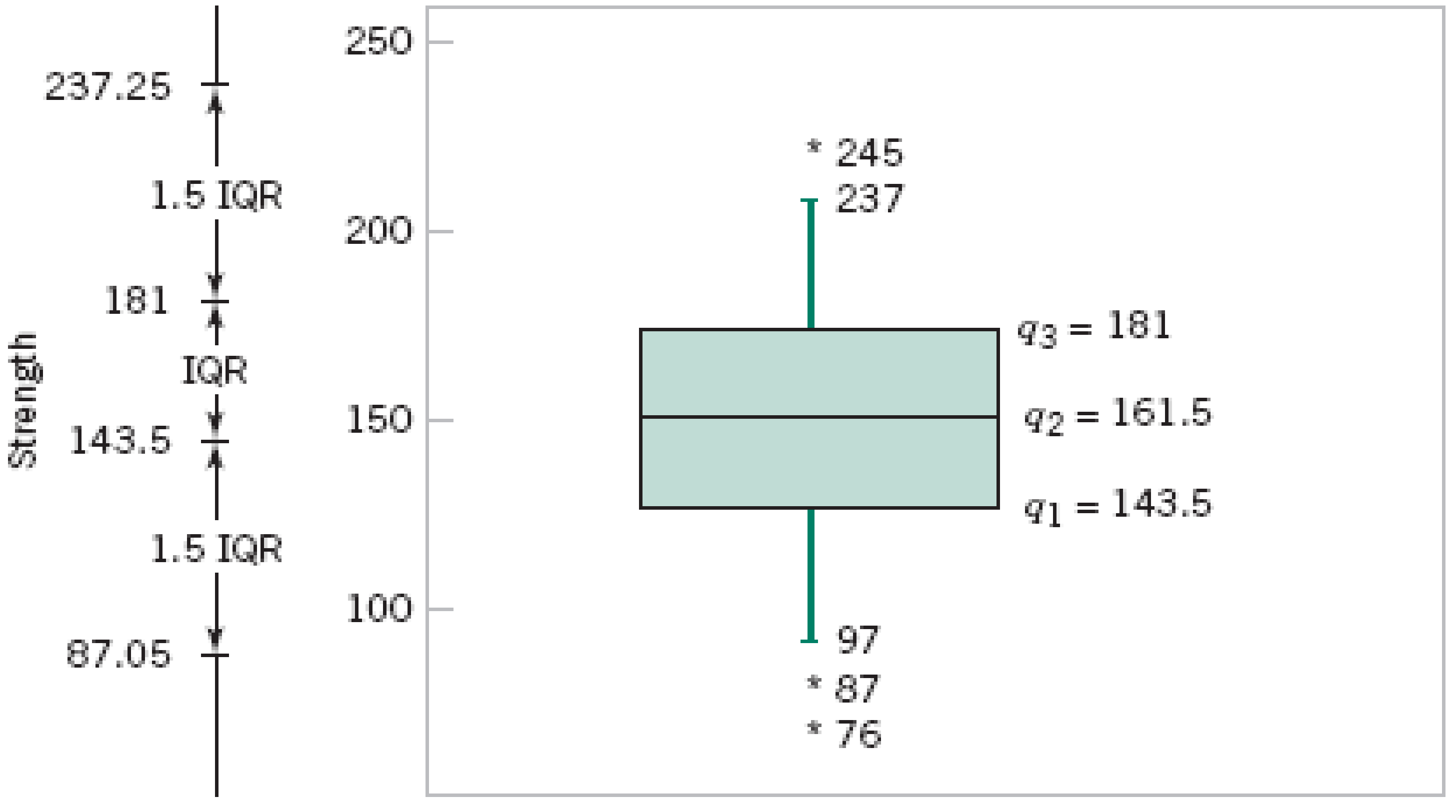
6-4 Box Plots



Outlier : a point beyond whisker but less than 3 IQR from the box edge

Extreme outlier: a point more than 3 IQR from the box edge

6-4 Box Plots

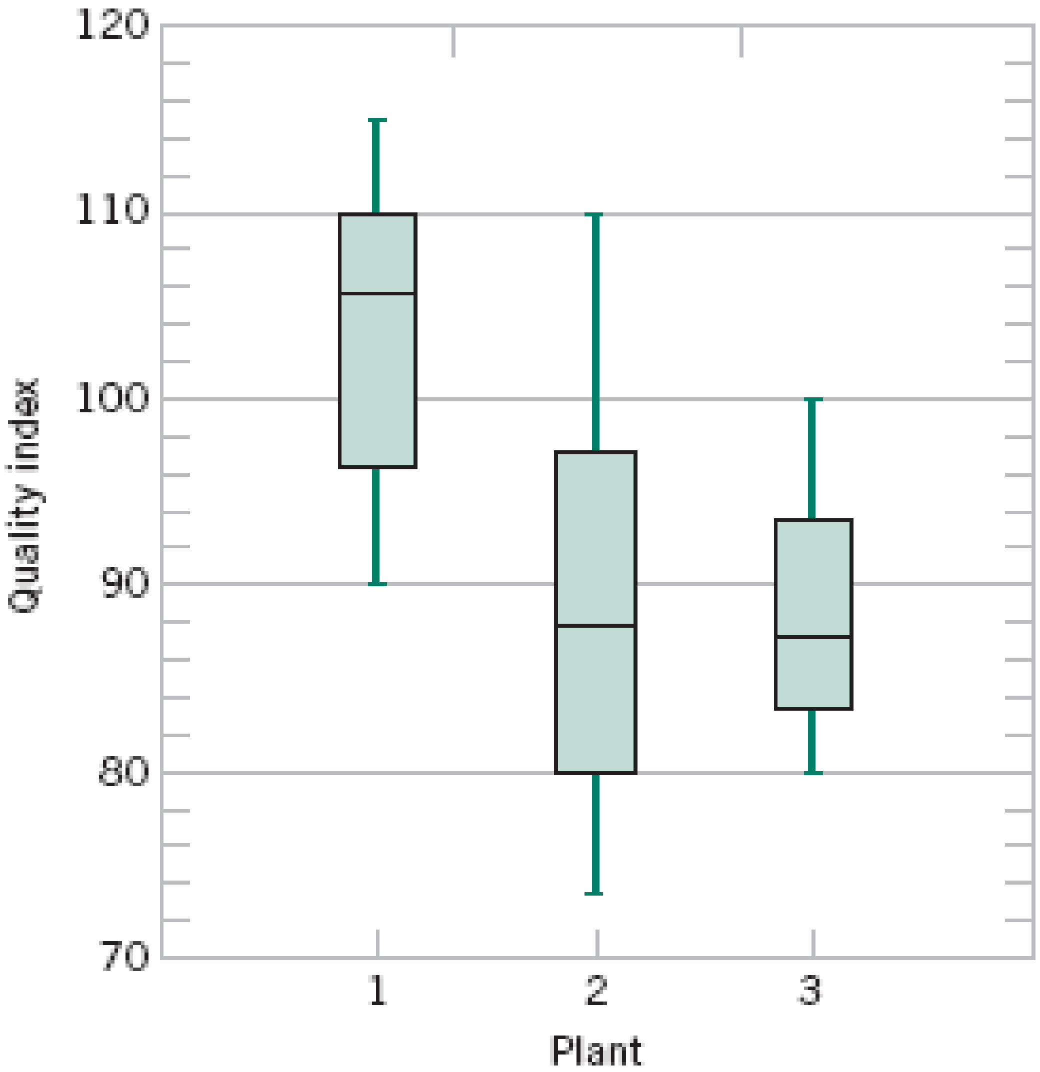


Box plot for compressive strength data

6-4 Box Plots

Box plots are useful for graphical comparisons among data sets.

Comparative box plots of a quality index at three plants.



6-5 Time Sequence Plots

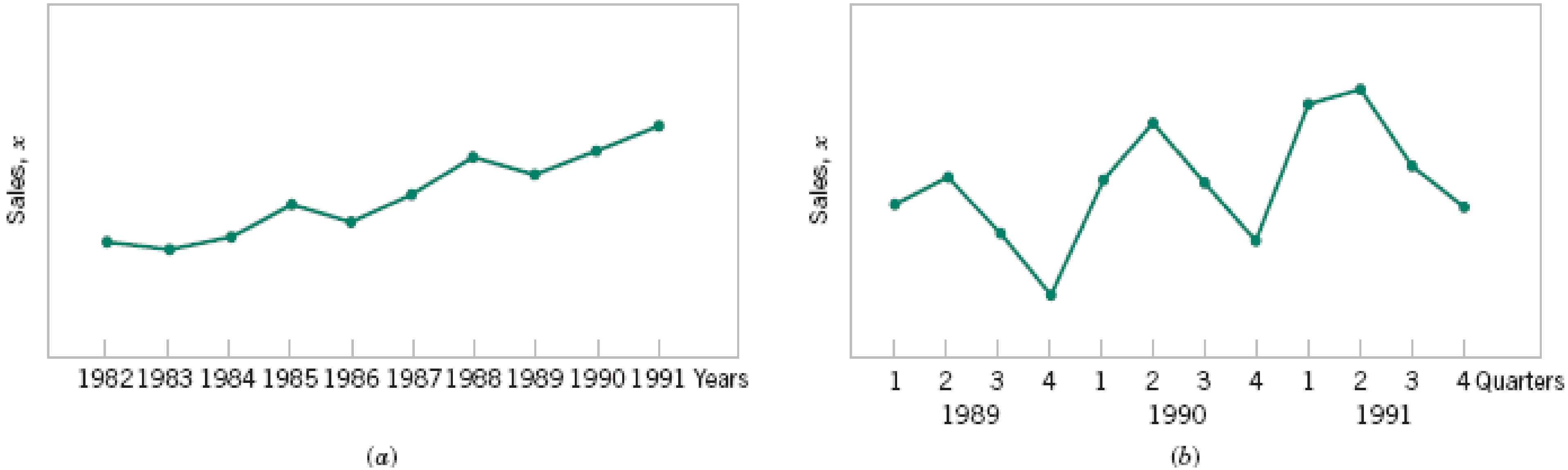
• A **time series** or **time sequence** is a data set in which the observations are recorded in the order in which they occur.

• A **time series plot** is a graph in which the vertical axis denotes the observed value of the variable (say x) and the horizontal axis denotes the time (which could be minutes, days, years, etc.).

• When measurements are plotted as a time series, we often see

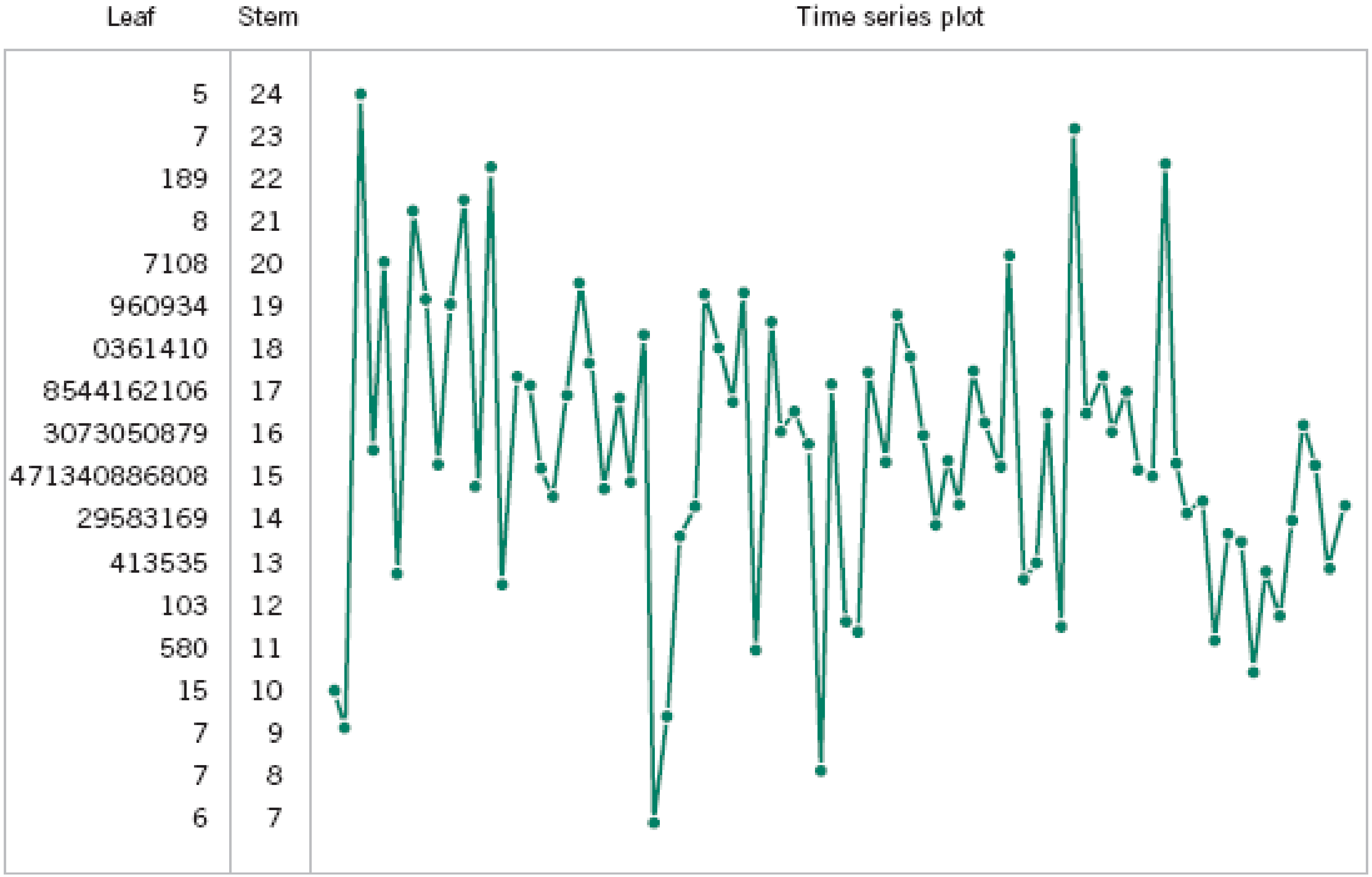
- **trends,**
- **cycles, or**
- **other broad features of the data**

6-5 Time Sequence Plots



Company sales by year (a) and by quarter (b).

6-5 Time Sequence Plots

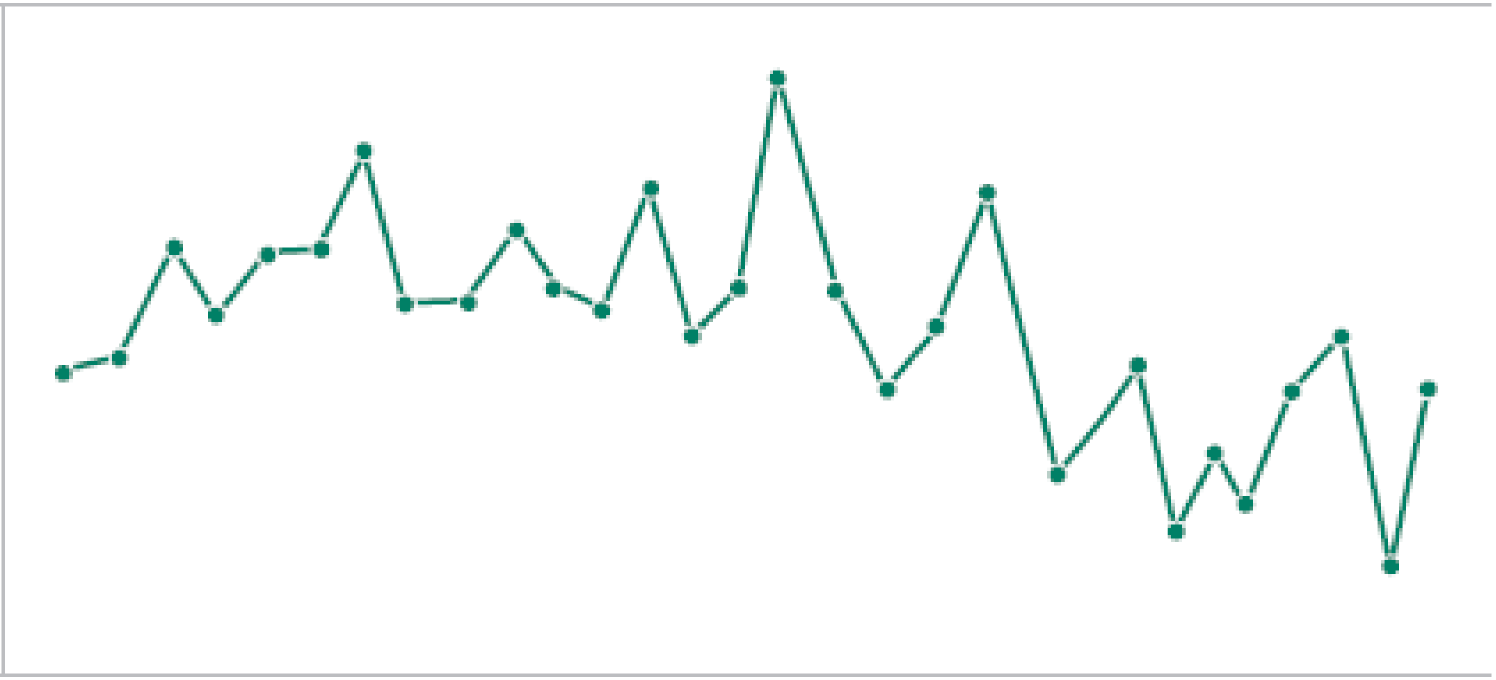


A digidot *(stem-and-leaf + time series)* plot of the compressive strength data

6-5 Time Sequence Plots

Leaf	Stem
8	9e
6	9s
45	9f
2333	9t
0010000	9z
99998	8e
66677	8s
45	8f
23	8t
1	8z

Time series plot



A digidot plot of chemical process concentration readings, observed hourly.

After 20 hours, lower concentrations begin to occur.

6-6 Probability Plots

- **Probability plotting** is a graphical method for determining whether sample data conform to a hypothesized distribution based on a subjective visual examination of the data.

- Probability plotting typically uses special graph paper, known as **probability paper,** that has been designed for the hypothesized distribution. Probability paper is widely available for the normal, lognormal, Weibull, and various chi-square and gamma distributions.

6-6 Probability Plots

Example 6-7

Ten observations on the effective service life in minutes of batteries used in a portable personal computer are as follows: 176, 191, 214, 220, 205, 192, 201, 190, 183, 185. We hypothesize that battery life is adequately modeled by a normal distribution. To use probability plotting to investigate this hypothesis, first arrange the observations in ascending order and calculate their cumulative frequencies $(j - 0.5)/10$ as shown in Table 6-6.

Table 6-6 Calculation for Constructing a Normal Probability Plot

j	$x_{(j)}$	$(j - 0.5)/10$	z_j
1	176	0.05	−1.64
2	183	0.15	−1.04
3	185	0.25	−0.67
4	190	0.35	−0.39
5	191	0.45	−0.13
6	192	0.55	0.13
7	201	0.65	0.39
8	205	0.75	0.67
9	214	0.85	1.04
10	220	0.95	1.64

6-6 Probability Plots

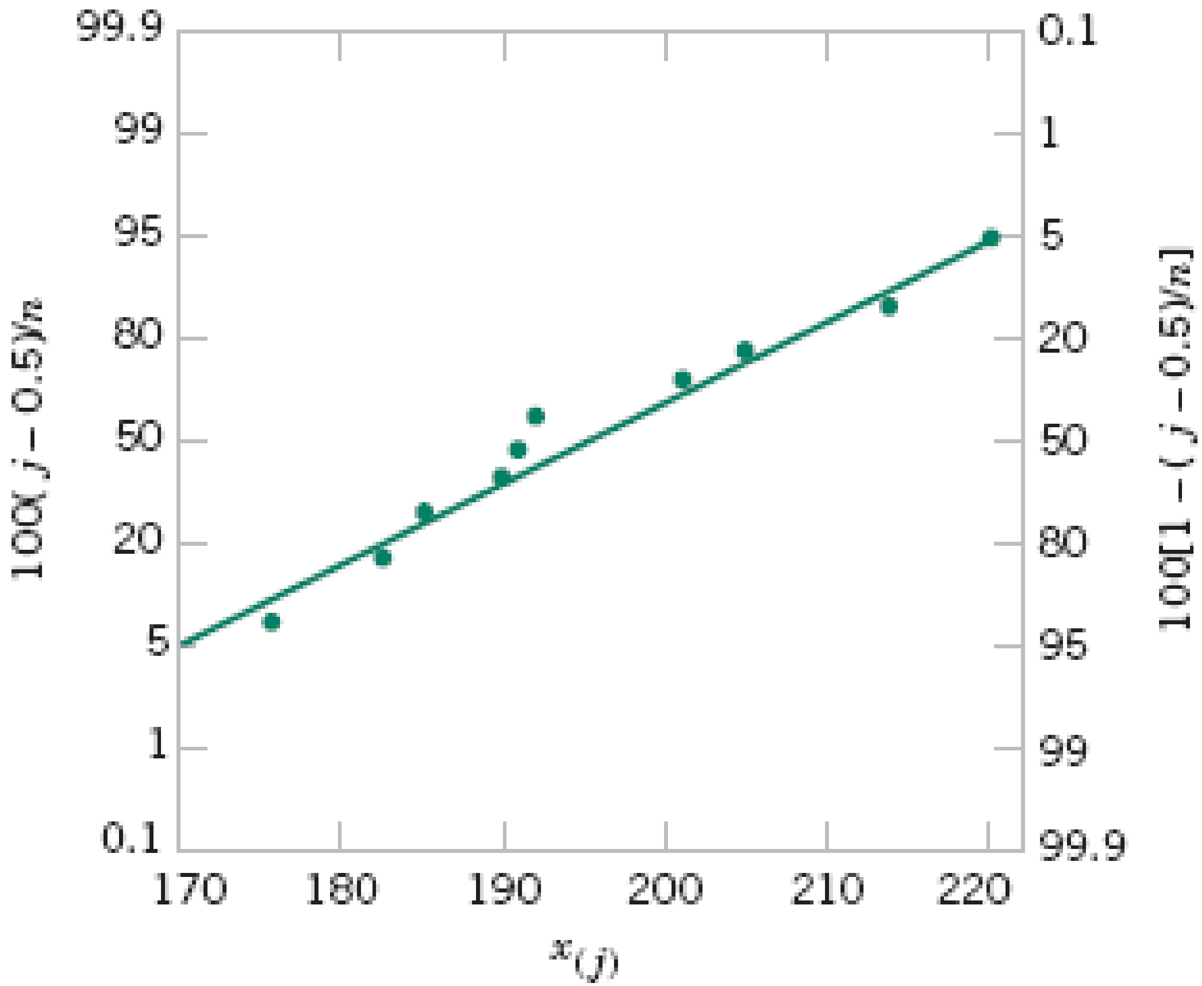
Example 6-7 (continued)

- A straight line, chosen subjectively, is drawn through the plotted points.
- In drawing the straight line, you should be influenced more by the points near the middle of the plot than by the extreme points.
- A good rule of thumb is to draw the line approximately between the 25th and 75th percentile points.
- Imagine a **fat pencil** lying along the line. If all the points are covered by this imaginary pencil, a normal distribution adequately describes the data.

6-6 Probability Plots

Normal probability plot
for battery life

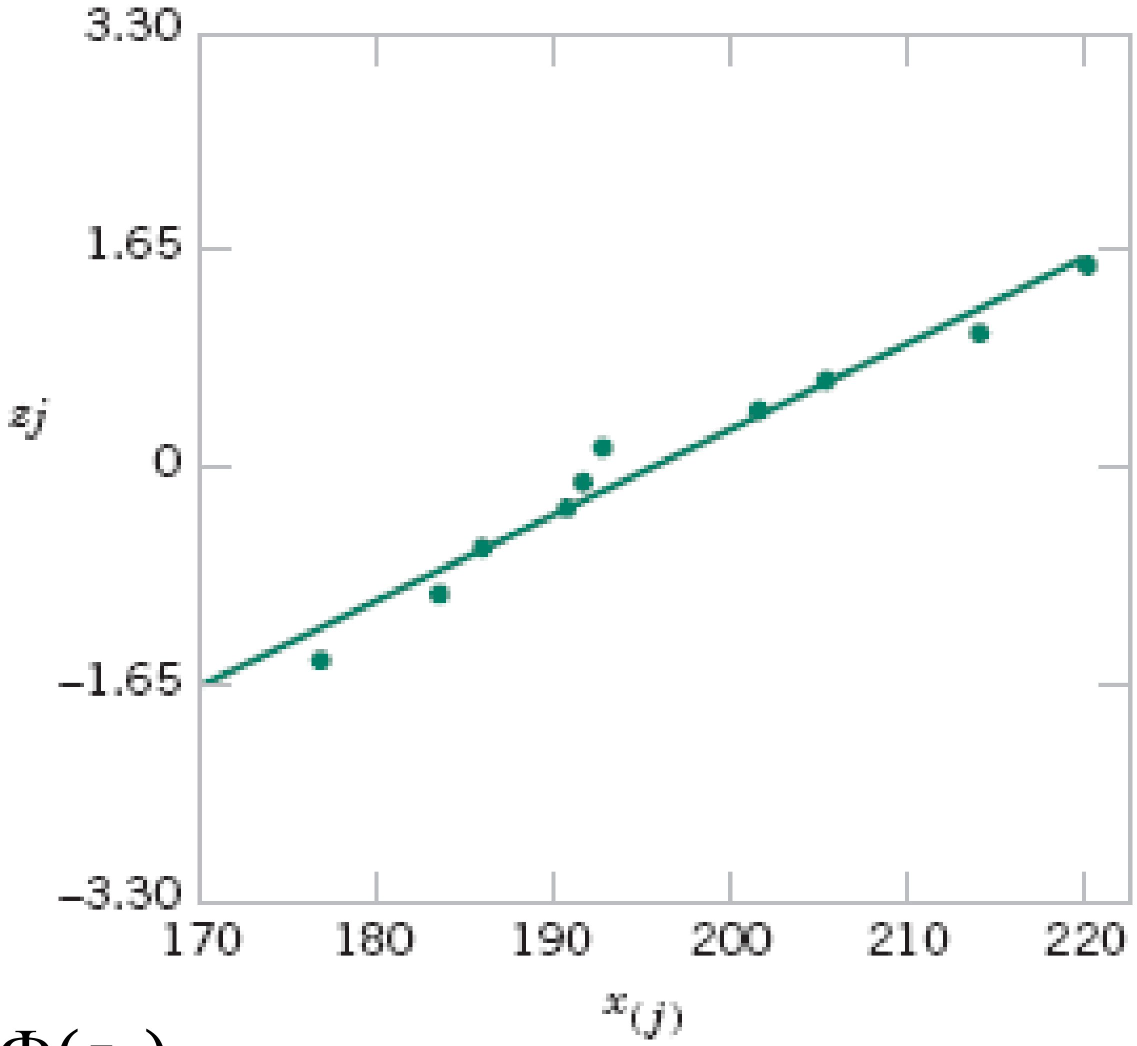
obtained from
cumulative frequencies



The points pass the
“fat pencil” test,
So, the normal distribution
is an appropriate model.

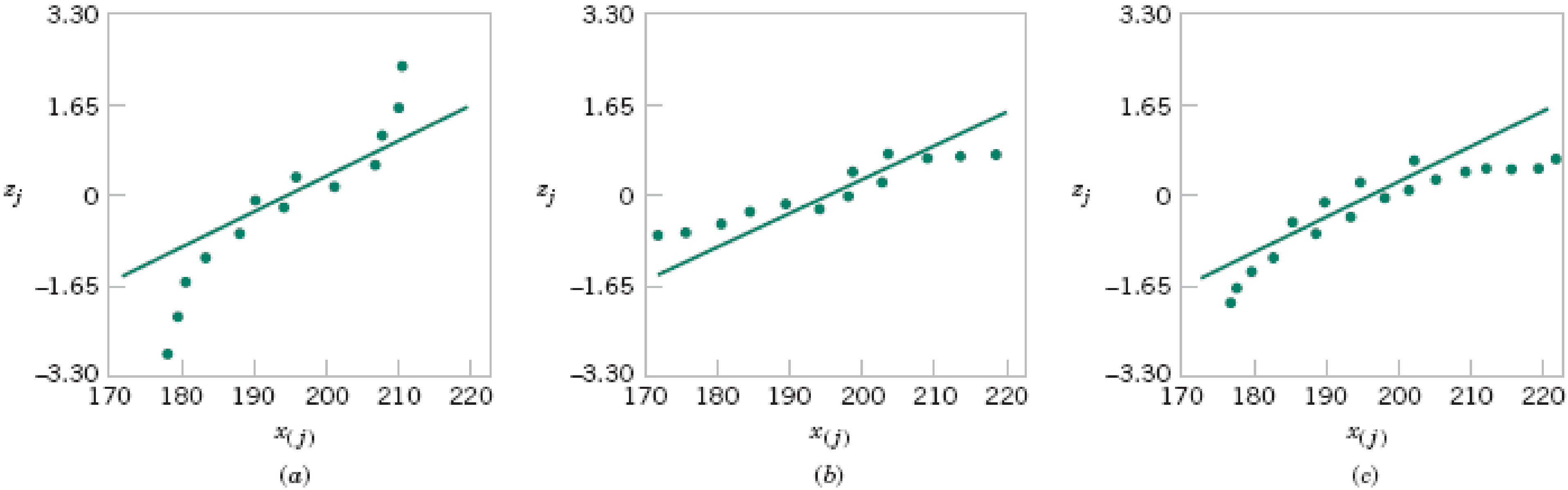
6-6 Probability Plots

Normal probability plot obtained from standardized normal scores.



$$\frac{j-0.5}{n} = P(Z \le z_j) = \Phi(z_j)$$

6-6 Probability Plots



Normal probability plots indicating a nonnormal distribution. (a) Light-tailed distribution. (b) Heavy-tailed distribution. (c) A distribution with positive (or right) skew.

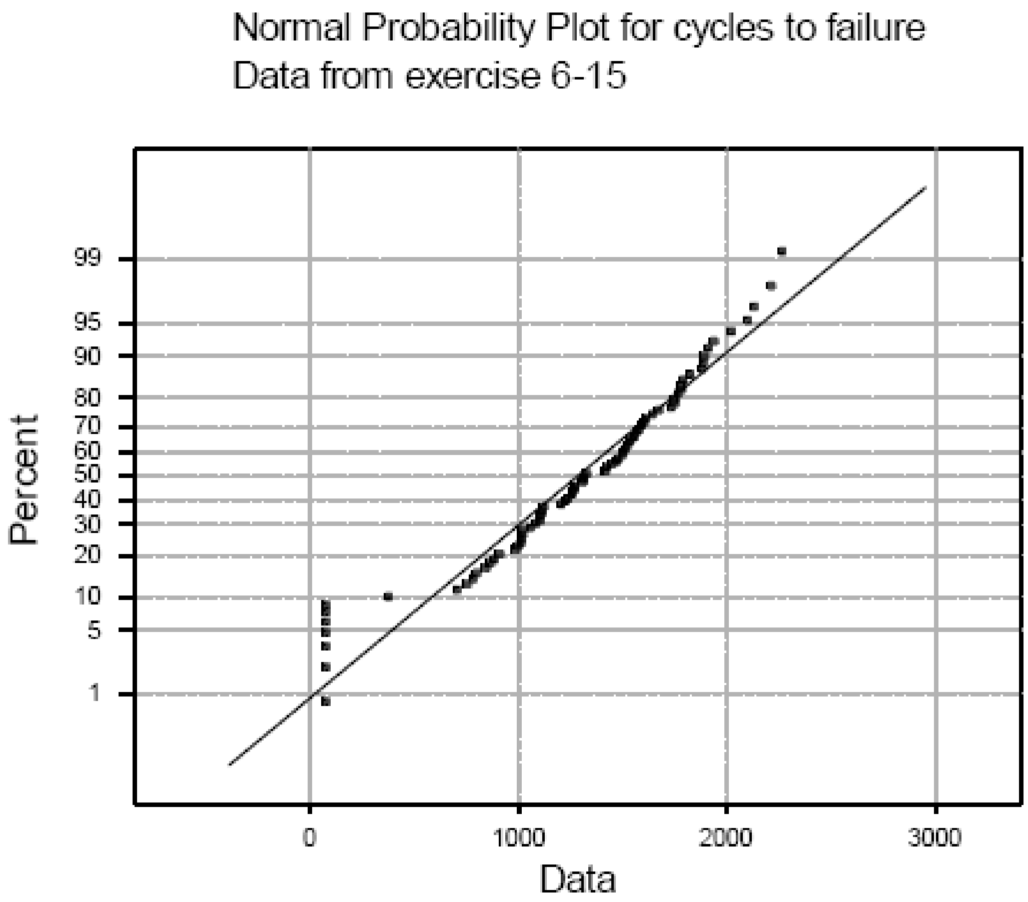
6-6 Excercise 6-71 (6.79)

70 data: Numbers of cycles to failure of aluminum test coupons subjected to repeated alternating stress at 21000 psi, 18 cycles per second

1115	2130	1674	2265	1260	1730	1535
1310	1421	1016	1910	1888	1102	1781
1540	1109	1102	1018	1782	1578	1750
1502	1481	1605	1452	1522	758	1501
1258	1567	706	1890	1792	1416	1238
1315	1883	2215	2100	1000	1560	990
1085	1203	785	1594	1820	1055	1468
798	1270	885	2023	1940	1764	1512
1020	1015	1223	1315	1120	1330	1750
865	845	375	1269	910	1608	1642

6-6 Excercise 6-71 (6.79)

The data appears to be normally distributed although there are some departures at the ends



6-6 Excercise 6-27 (6.35)

40 data: Wine gradings on a 0-100 point scale

94	90	92	91	91	86	89	91	91	90
90	93	87	90	91	92	89	86	89	90
88	95	91	88	89	92	87	89	95	92
85	91	85	89	88	84	85	90	90	83

6-6 Excercise 6-27 (6.35)

Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{40} x_i}{40} = \frac{3578}{40} = 89.45$$

Sample variance:

$$\sum_{i=1}^{40} x_i = 3578 \qquad \sum_{i=1}^{40} x_i^2 = 320366$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{320366 - \frac{(3578)^2}{40}}{40-1} = \frac{313.9}{39}$$

$$= 8.05$$

Sample modes: 90, 91

Leaf unit: 0.1

1|2 represents 1.2

1	83|0
2	84|0
5	85|000
7	86|00
9	87|00
12	88|000
18	89|000000
(7)	90|0000000
15	91|0000000
8	92|0000
4	93|0
3	94|0
2	95|00

6-6 Excercise 6-27 (6.35)

Sample quartiles:

$$for\ q_1 \quad \frac{n+1}{4} = \frac{40+1}{4} = 10.25$$

$$for\ q_2 \quad \frac{2(n+1)}{4} = \frac{40+1}{2} = 20.5 \quad ===> Median$$

$$for\ q_3 \quad \frac{3(n+1)}{4} = \frac{3(40+1)}{4} = 30.75$$

$$q_1 = 88$$

$$q_2 = \tilde{x} = 90$$

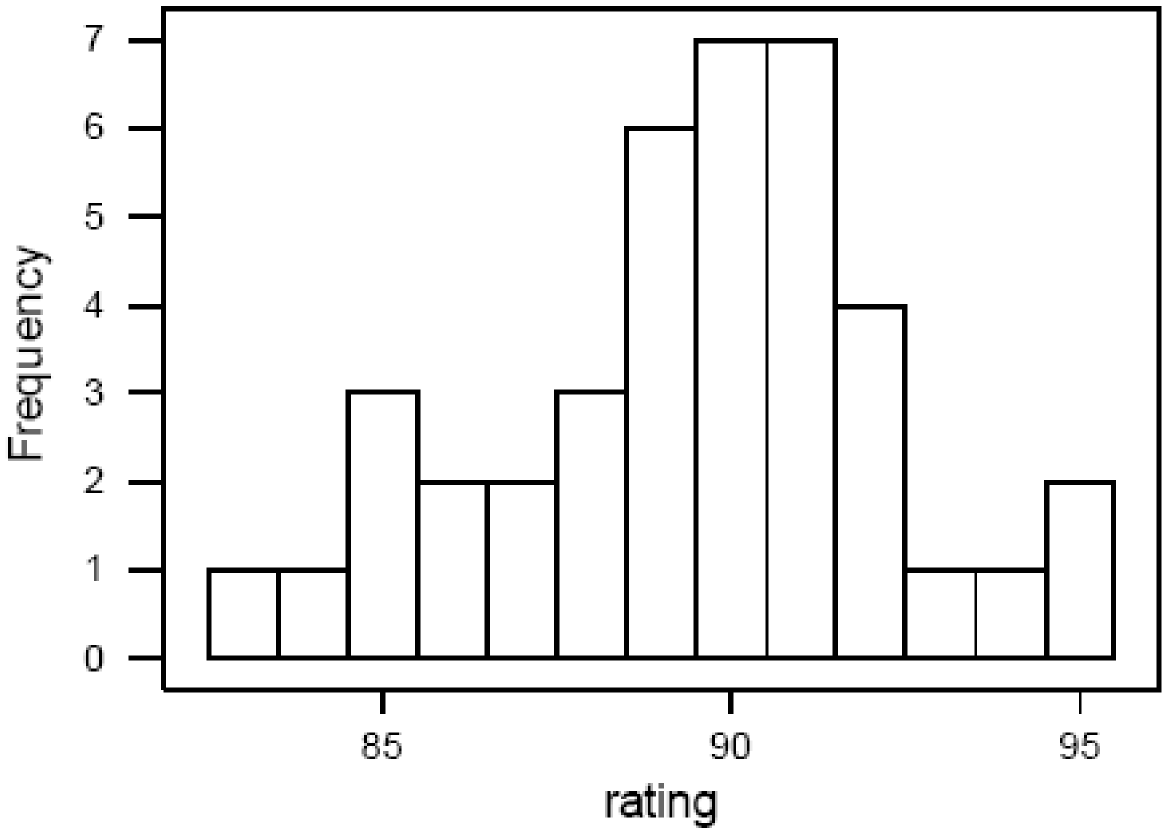
$$q_3 = 91$$

Leaf unit: 0.1

1|2 represents 1.2

1	83|0
2	84|0
5	85|000
7	86|00
9	87|00
12	88|000
18	89|000000
(7)	90|0000000
15	91|0000000
8	92|0000
4	93|0
3	94|0
2	95|00

6-6 Excercise 6-27 (6.35)



Leaf unit: 0.1

1|2 represents 1.2

```
  1    83|0
  2    84|0
  5    85|000
  7    86|00
  9    87|00
 12    88|000
 18    89|000000
(7)    90|0000000
 15    91|0000000
  8    92|0000
  4    93|0
  3    94|0
  2    95|00
```

6-6 Excercise 6-50 (6.58 – data of 6.22)

Q_1=88.6
Q_2=90.4
Q_3=92.2

IQR=3.6
1.5*IQR=5.4

Whiskers
83.4
96.5

Outliers
98.8
100.3