DATA SUMMARY AND PRESENTATION

Probability and Statistics

3-1 Data Summary and Display

Sample Mean

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

x bar

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (3\text{-}1)$$

3-1 Data Summary and Display

Example 3-1: O-Ring Strength: Sample Mean

Consider the O-ring tensile strength experiment described in Chapter 1. The data from the modified rubber compound are shown in the **dot diagram** (Fig. 2-2). The sample mean strength (psi) for the eight observations on strength is

$$\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{n} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{1037 + 1047 + \cdots + 1040}{8}$$

$$= \frac{8440}{8} = 1055.0 \text{ psi}$$

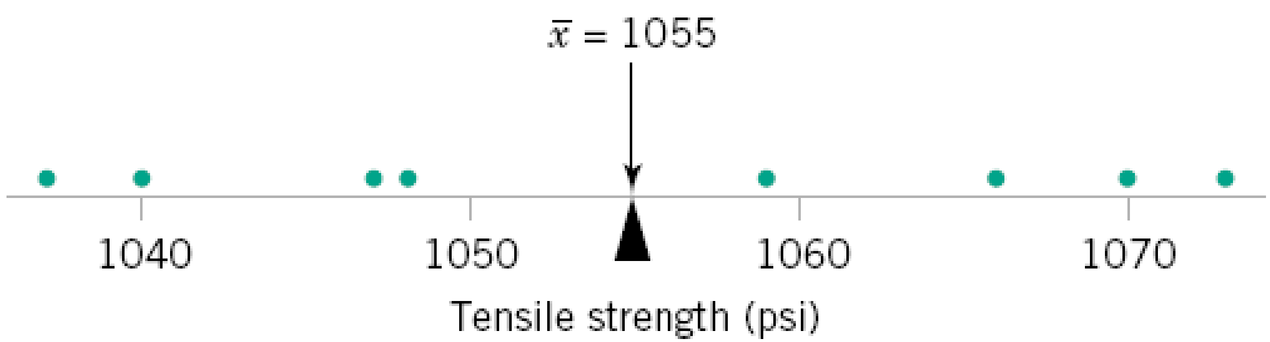
A physical interpretation of the sample mean as a measure of location is shown in Fig. 2-2. Note that the sample mean $\bar{x} = 1055$ can be thought of as a "balance point." That is, if each observation represents 1 pound of mass placed at the point on the x-axis, a fulcrum located at $\bar{x}$ would exactly balance this system of weights.

mean crate to use this number represent value of population / sample

3-1 Data Summary and Display

Example 3-1: O-Ring Strength: Sample Mean

Figure 3-1 Dot diagram of O-ring tensile strength. The sample mean is shown as a balance point for a system of weights.



3-1 Data Summary and Display

Population Mean

For a finite population with N measurements, the mean is

$$\mu = \frac{\sum_{i=1}^{N} x_i}{\underbrace{N}_{\text{population size}}}$$

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

Sample Variance and Sample Standard Deviation

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

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

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

Variance Defined

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

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

For the N observations in a population denoted by $x_1, x_2, \ldots, x_N,$ the population variance, analogous to the variance of a probability distribution, is

$$\sigma^2 = \sum_{i=1}^{N}\left(x_i - \mu\right)^2 \cdot f\left(x\right) = \frac{\sum_{i=1}^{N}\left(x_i - \mu\right)^2}{N} \tag{3-4}$$

What is this “n–1”?

- The population variance is calculated with *N*, the population size. Why isn’t the sample variance calculated with *n*, the sample size?
- The true variance is based on data deviations from the true mean, μ.
- The sample calculation is based on the data deviations from *x-bar*, not μ. *X-bar* is an estimator of μ; close but not the same. So the n-1 divisor is used to compensate for the error in the mean estimation.

Degrees of Freedom

- The sample variance is calculated with the quantity *n*-1.
- This quantity is called the "degrees of freedom".
- Origin of the term:
  - There are *n* deviations from *x-bar* in the sample.
  - The sum of the deviations is zero. (Balance point)
  - *n*-1 of the observations can be freely determined, but the n^{th} observation is fixed to maintain the zero sum.

3-1 Data Summary and Display

0 = x1+x2+…+xn
xn = -(x1+x2+..+xn-1)

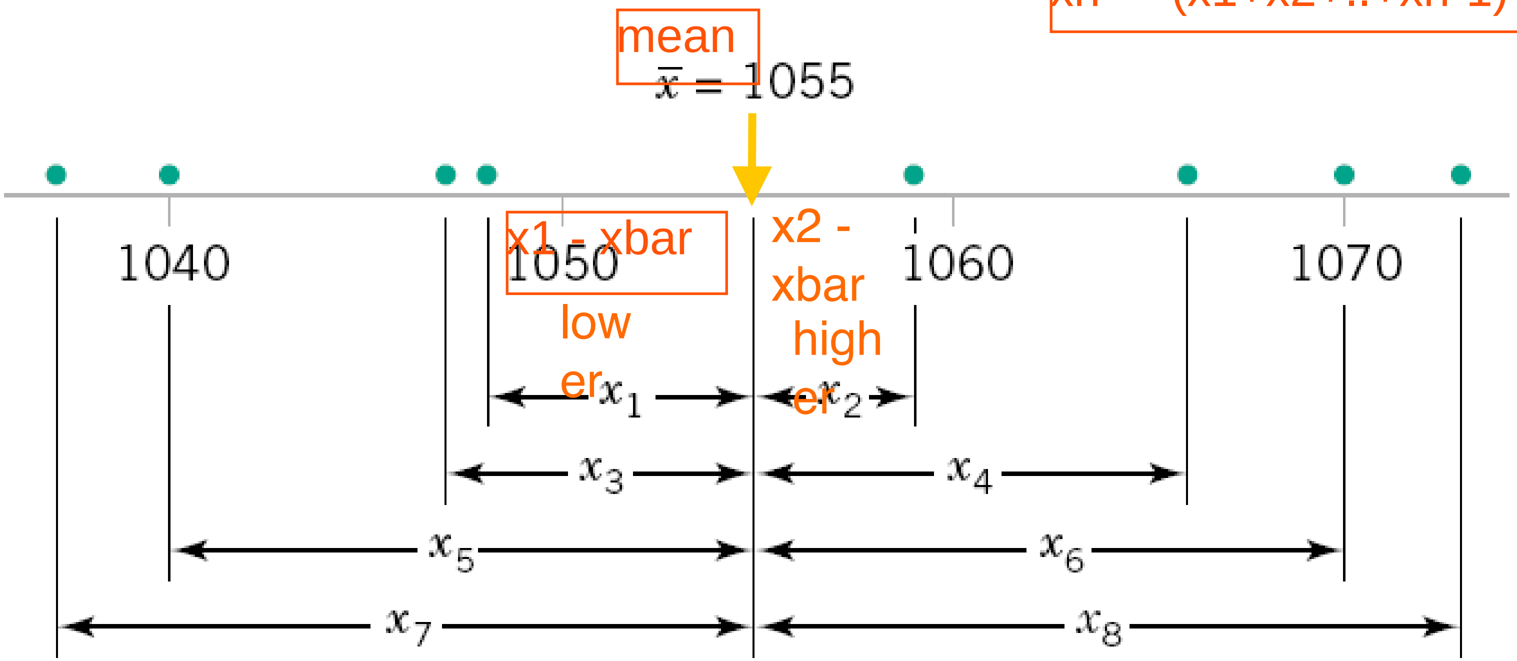


Figure 3-2 How the sample variance measures variability through the deviations $x_i - \bar{x}$.

Example 3-2: O-Ring Strength: Sample Variance

Table 3-1 Calculation of Terms for the Sample Variance and Sample Standard Deviation

i	x_i	$x_i - \overline{x}$ (1055)	$(x_i - \overline{x})^2$
1	1048	−7	49
2	1059	4	16
3	1047	−8	64
4	1066	11	121
5	1040	−15	225
6	1070	15	225
7	1037	−18	324
8	1073	18	324
	8440	0.0	1348

Example 3-3: O-Ring Strength: Alternative Variance Calculation

The sample variance is

$\boxed{S^2 = 1348/(8-1) = 1348/7 = 192.57 \text{ psi}^2}$

The sample standard deviation is

sqrt(192.57) =

Computational formula for s^2

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

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

$$= \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \boxed{\frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \Big/ n}{n-1}} \qquad (3\text{-}5)$$

3-1 Data Summary and Display

Population Variance

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

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} \qquad (3\text{-}6)$$

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

3-2 Stem-and-Leaf Diagram

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.

3-2 Stem-and-Leaf Diagram

Example 3-4: Compressive Strength

To illustrate the construction of a stem-and-leaf diagram, consider the alloy compressive strength data in Table 2-2. We will select as stem values the numbers 7, 8, 9, . . . , 24. The resulting stem-and-leaf diagram is presented in Fig. 2-4. The last column in the diagram is a frequency count of the number of leaves associated with each stem.

Practical interpretation: Inspection of this display immediately reveals that most of the compressive strengths lie between 110 and 200 psi and that a central value is somewhere between 150 and 160 psi. Furthermore, the strengths are distributed approximately symmetrically about the central value. The stem-and-leaf diagram enables us to determine quickly some important features of the data that were not immediately obvious in the original display in the table. ■

3-2 Stem-and-Leaf Diagram

Example 3-4: Compressive Strength (cont.)

Table 3-2 Compressive Strength 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

Example 3-4: Compressive Strength (cont.)

and split first
digit and other digit.

Table 3-2 Compressive Strength (psi) of Aluminum-Lithium 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

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

It is a origin histogram

Figure 3-3 Stem-and-leaf diagram for the compressive strength data in

3-2 Stem-and-Leaf Diagram

Table 3-2 Compressive Strength (psi) of Aluminum-Lithium 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

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

Figure 3-5 A stem-and-leaf diagram from Minitab.

3-2 Stem-and-Leaf Diagram

Table 3-3 Summary Statistics for the Compressive Strength Data from Minitab Text

Variable	N	Mean	Median	StDev	SE Mean
	80	162.66	161.50	33.77	3.78
	Min	Max	Q1	Q3	
	76.00	245.00	143.50	181.00	

Frequency Distributions

- A frequency distribution is a compact summary of data, expressed as a table, graph, or function.
- The data is gathered into bins or cells, defined by class intervals.
- The number of classes, multiplied by the class interval, should exceed the range of the data. The square root of the sample size is a guide.
- The boundaries of the class intervals should be convenient values, as should the class width.

Frequency Distribution Table

Relative Feq
=2/80

Cumulative Relative Freq
=zigma (Relative)

Considerations:

Range = 245-76 = 169

Sqrt(80) = 8.9

Trial class width = 18.9

Decisions:

Number of classes = 9

Class width = 20

Range of classes = 20*9 = 180

Starting point = 70

Table 3-4 Frequency Distribution of Table 3-2 Data

Class	Frequency	Relative Frequency	Cumulative Relative Frequency
$70 \le x < 90$	2	0.0250	0.0250
$90 \le x < 110$	3	0.0375	0.0625
$110 \le x < 130$	6	0.0750	0.1375
$130 \le x < 150$	14	0.1750	0.3125
$150 \le x < 170$	22	0.2750	0.5875
$170 \le x < 190$	17	0.2125	0.8000
$190 \le x < 210$	10	0.1250	0.9250
$210 \le x < 230$	4	0.0500	0.9750
$230 \le x < 250$	2	0.0250	1.0000
	80	1.0000	

3-3 Histograms

A **histogram** is a more compact summary of data than a stem-and-leaf diagram. To construct a histogram for continuous data, we must divide the range of the data into intervals, which are usually called **class intervals**, **cells**, or **bins**. If possible, the bins should be of equal width to enhance the visual information in the histogram.

- Steps to build one with equal bin widths:
  1) Label the bin boundaries on the horizontal scale.
  2) Mark & label the vertical scale with the frequencies or relative frequencies.
  3) Above each bin, draw a rectangle whose height is equal to the frequency or relative frequency.

Histogram of the Table 3-4 Data

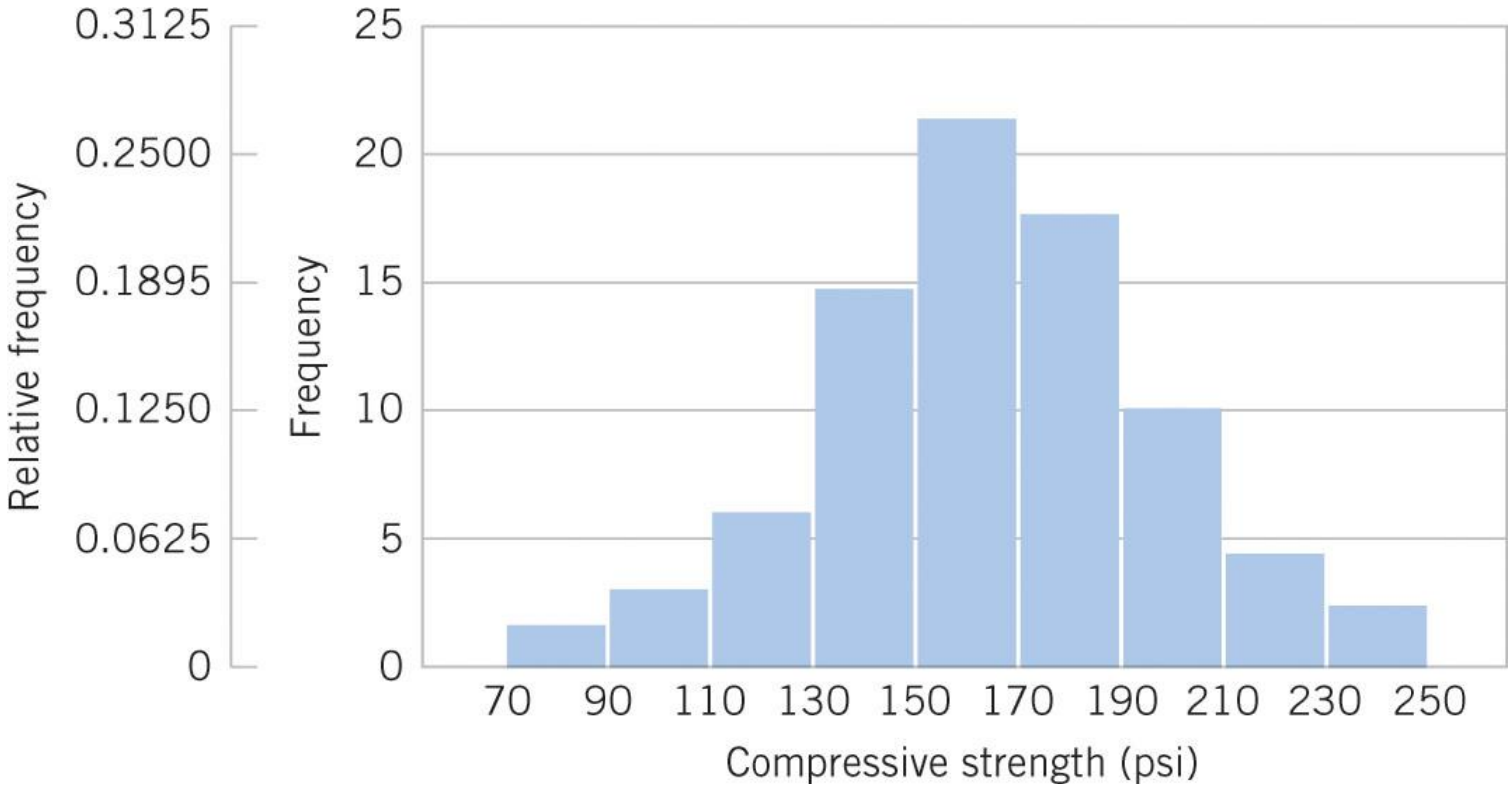


Figure 3-6 Histogram of compressive strength of 80 aluminum-lithium alloy specimens. Note these features – (1) horizontal scale bin boundaries & labels with units, (2) vertical scale measurements and labels, (3) histogram title at top or in legend.

Histograms with Unequal Bin Widths

- If the data is tightly clustered in some regions and scattered in others, it is visually helpful to use narrow class widths in the clustered region and wide class widths in the scattered areas.
- In this approach, the rectangle area, not the height, must be proportional to the class frequency.

$$\text{Rectangle height} = \frac{\text{bin frequency}}{\text{bin width}}$$

Poor Choices in Drawing Histograms-1

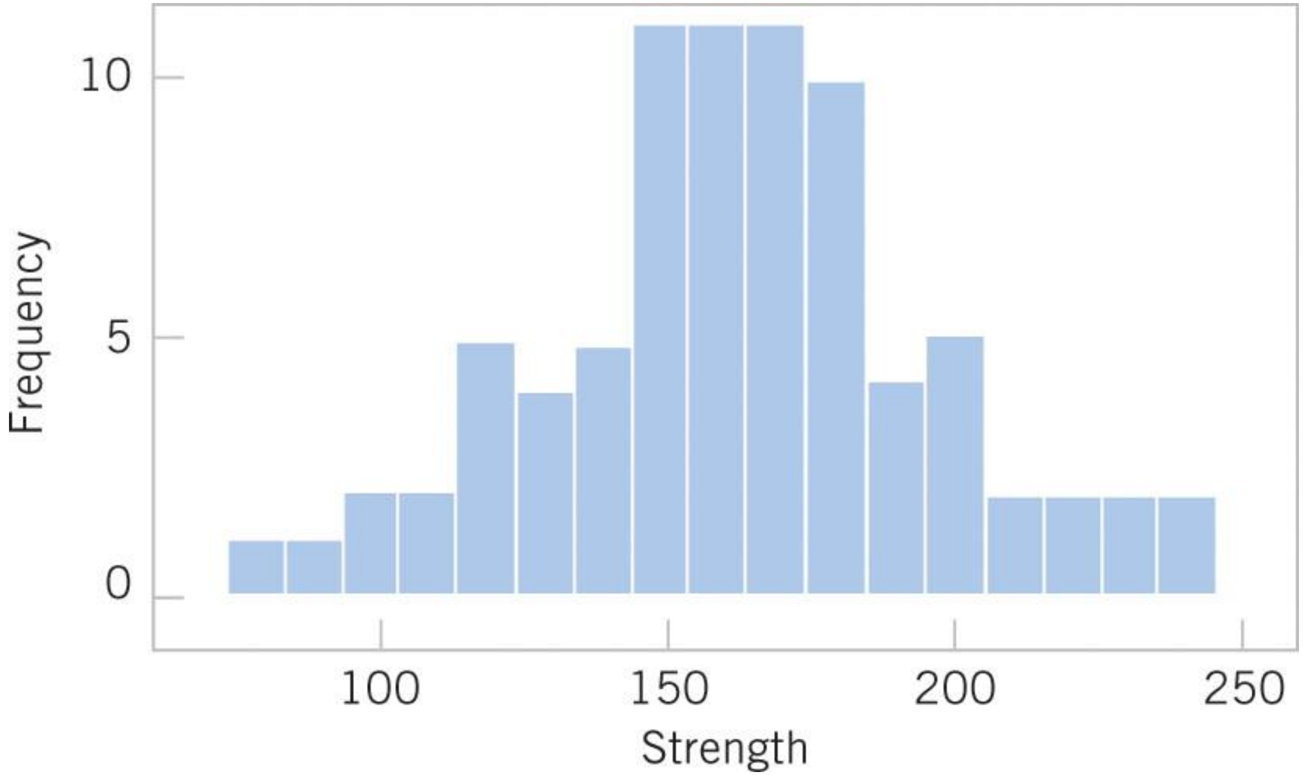


Figure 3-7 Histogram of compressive strength of 80 aluminum-lithium alloy specimens. Errors: too many bins (17) create jagged shape, horizontal scale not at class boundaries, horizontal axis label does not include units.

Poor Choices in Drawing Histograms-2

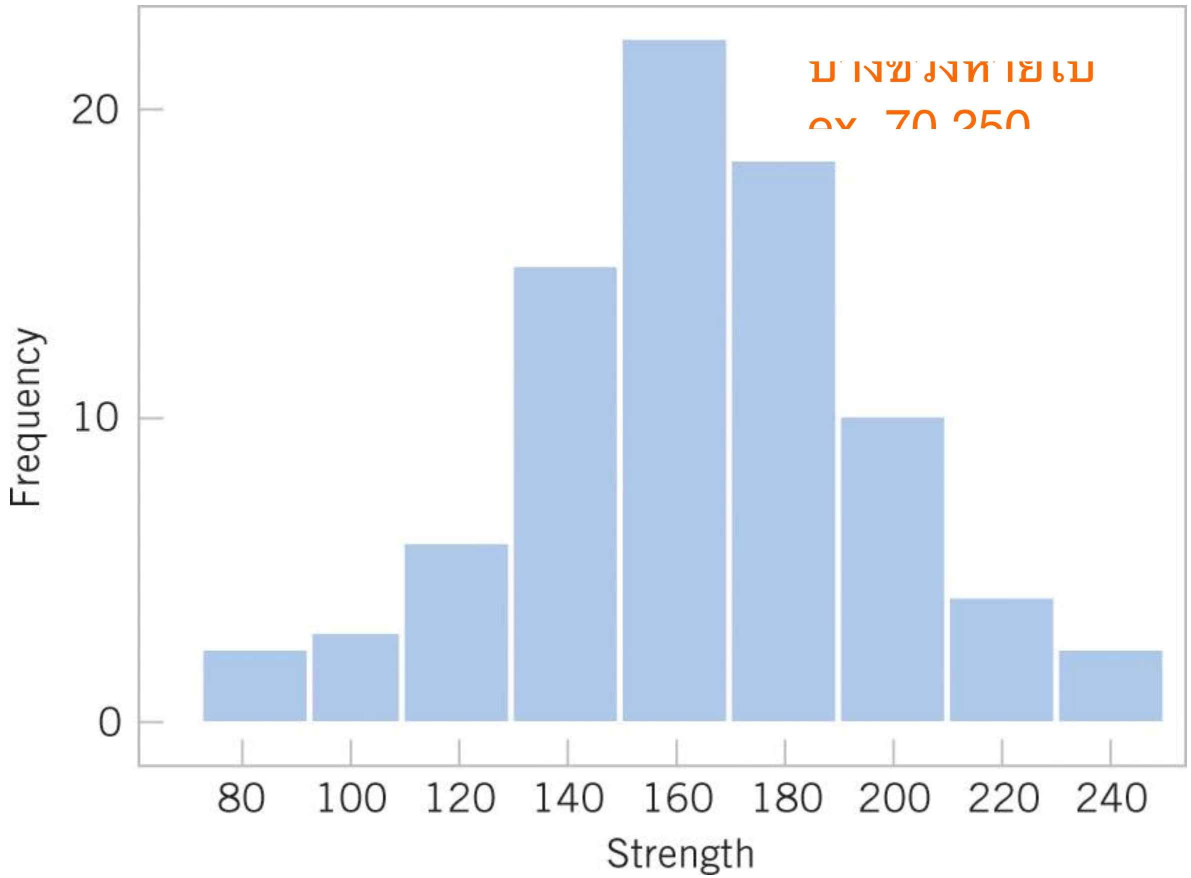


Figure 3-8 Histogram of compressive strength of 80 aluminum-lithium alloy specimens. Errors: horizontal scale not at class boundaries (cutpoints), horizontal axis label does not include units.

Cumulative Frequency Plot

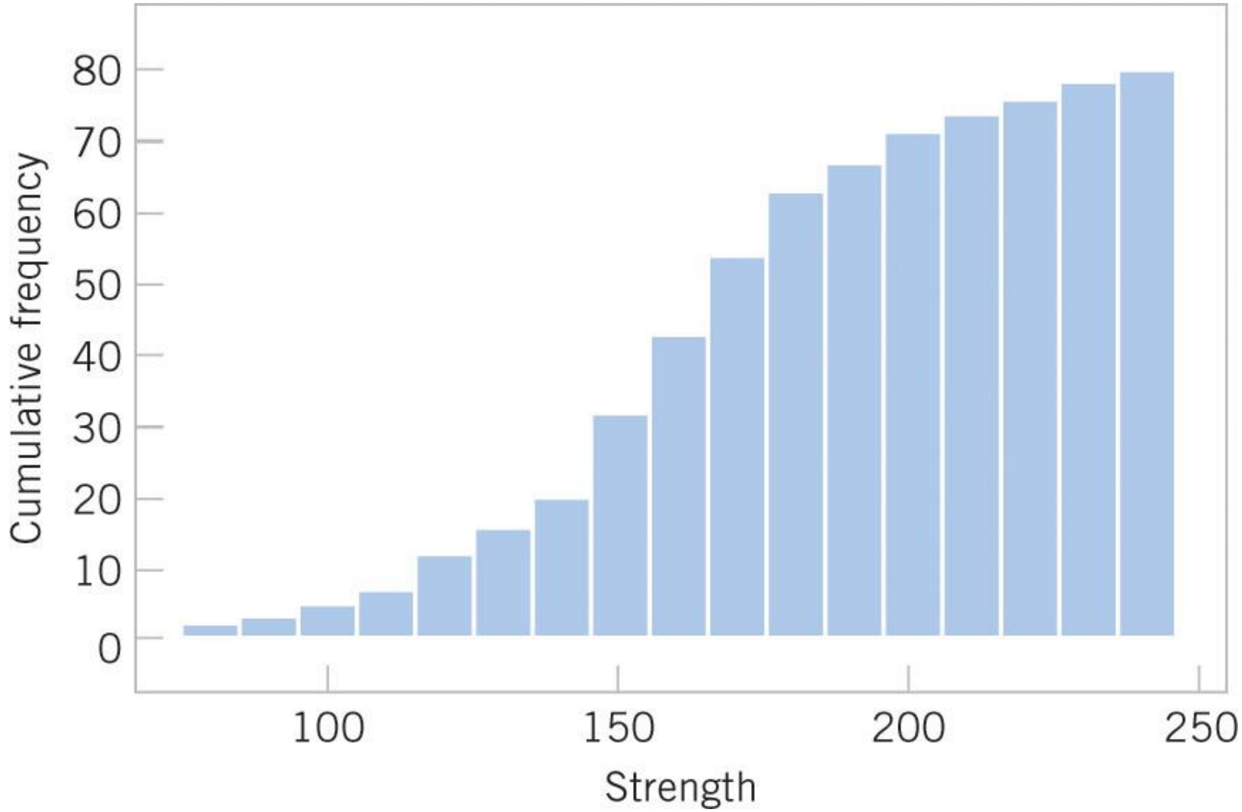


Figure 3-9 Cumulative histogram of compressive strength of 80 aluminum-lithium alloy specimens. Comment: Easy to see cumulative probabilities, hard to see distribution shape.

Shape of a Frequency Distribution

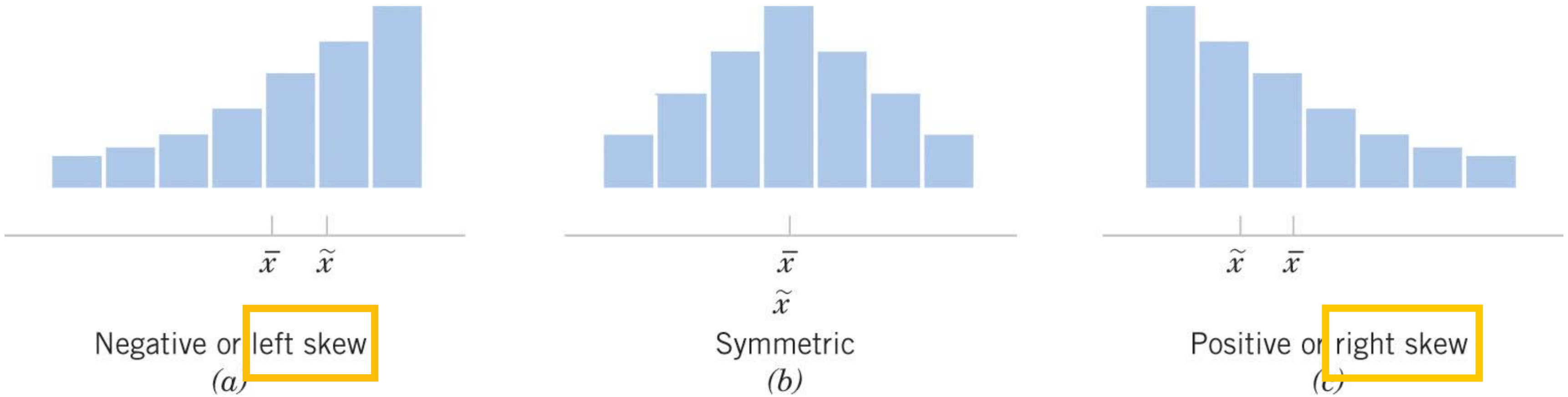


Figure 3-10 Histograms of symmetric and skewed distributions.
(b) Symmetric distribution has identical mean, median and mode measures.
(a & c) Skewed distributions are positive or negative, depending on the direction of the long tail. Their measures occur in alphabetical order as the distribution is approached from the long tail.☺

Histograms for Categorical Data

- Categorical data is of two types:
  - Ordinal: categories have a natural order, e.g., year in college, military rank.
  - Nominal: Categories are simply different, e.g., gender, colors.
- Histogram bars are for each category, are of equal width, and have a height equal to the category's frequency or relative frequency.
- A Pareto chart is a histogram in which the categories are sequenced in decreasing order. This approach emphasizes the most and least important categories.

3-3 Histograms – Pareto Chart

An important variation of the histogram is the **Pareto chart.** This chart is widely used in quality and process improvement studies where the data usually represent different types of defects, failure modes, or other categories of interest to the analyst. The categories are ordered so that the category with the largest number of frequencies is on the left, followed by the category with the second largest number of frequencies, and so forth.

3-3 Histograms – Pareto Chart

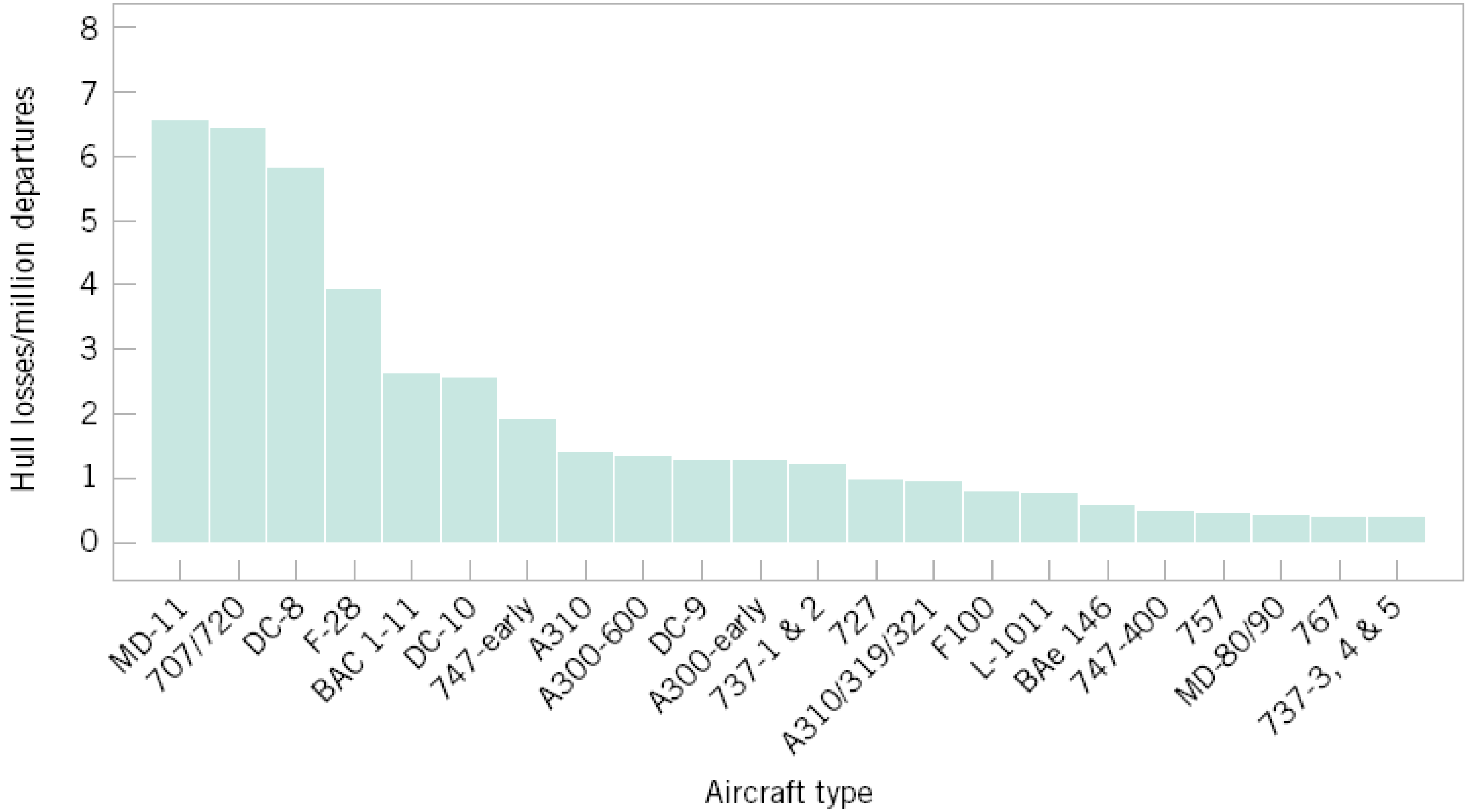


Figure 3-13 Pareto chart for the aircraft accident data.

3-4 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.

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

3-4 Box Plots

this is a right skew because Mean is a littile bit left

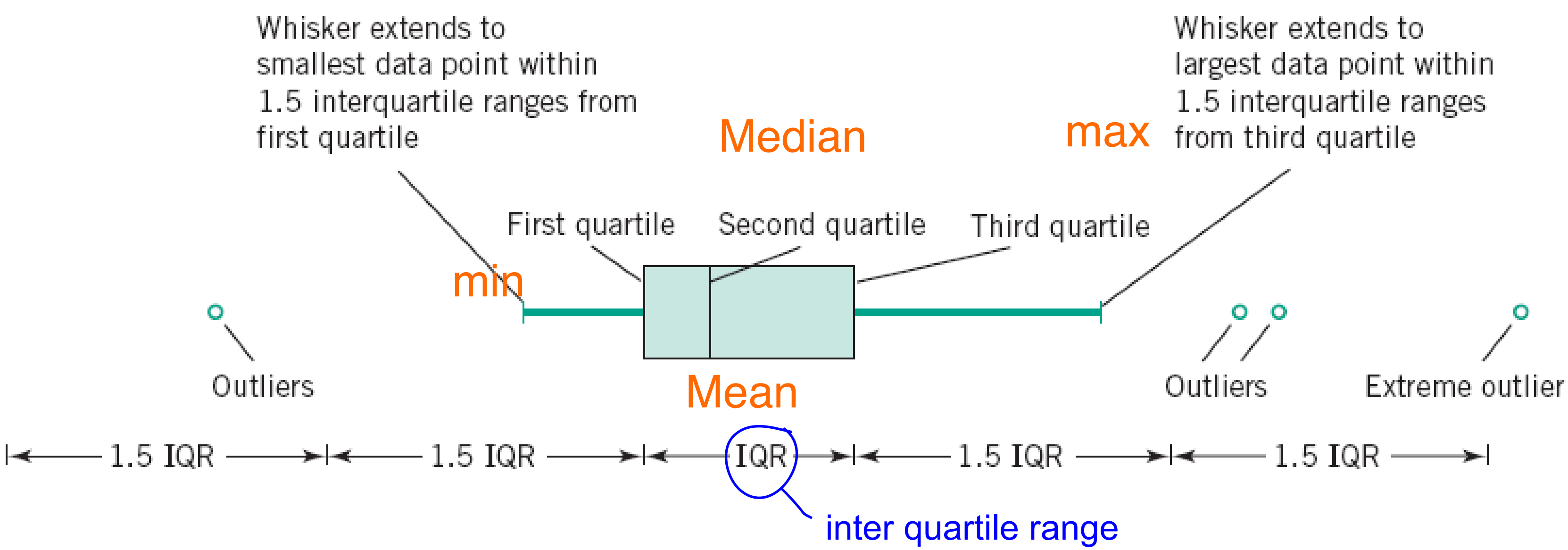


Figure 3-14
Description of a box plot.

Quartiles

Partition amount of data (NOT value)

- The three quartiles partition the data into four equally sized counts or segments.
  - 25% of the data is less than q_1.
  - 50% of the data is less than q_2, the median.
  - 75% of the data is less than q_3.
- Calculated as $Index = f(n+1)$ where:
  - *Index* (I) is the I^{th} item (interpolated) of the sorted data list.
  - f is the fraction associated with the quartile.
  - n is the sample size.
- For the Table 3-2 data:

		Value of indexed item		
f	*Index*	I^{th}	$(I+1)^{th}$	quartile
0.25	20.25	143	145	143.50
0.50	40.50	160	163	161.50
0.75	60.75	181	181	181.00

Percentiles

- Percentiles are a special case of the quartiles.
- Percentiles partition the data into 100 segments.
- The *Index* = $f(n+1)$ methodology is the same.
- The 37%ile is calculated as follows:
  - Refer to the Table 6-2 stem-and-leaf diagram.
  - *Index* = 0.37(81) = 29.97
  - 37%ile = 153 + 0.97(154 – 153) = 153.97

Interquartile Range

- The interquartile range (IQR) is defined as:

$$\mathrm{IQR} = q_3 - q_1.$$

- From Table 3-2:

$$\mathrm{IQR} = 181.00 - 143.25 = 37.75 = 37.8$$

- Impact of outlier data: ส่งผลกระทบ
  - IQR is not affected
  - Range is directly affected.

3-4 Box Plots

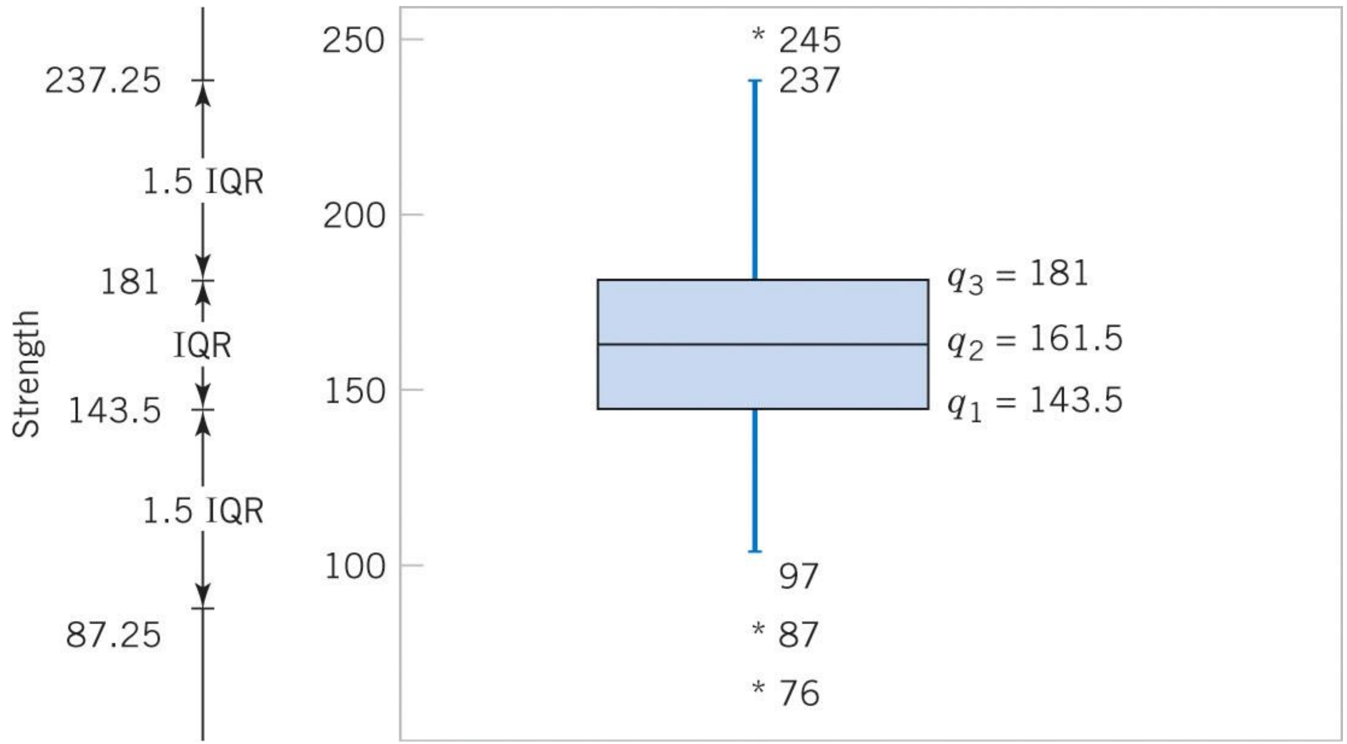


Figure 3-15 Box plot of compressive strength of 80 aluminum-lithium alloy specimens. <u>Comment</u>: Box plot may be shown vertically or horizontally, data reveals three outliers and no extreme outliers. Lower outlier limit is: 143.5 – 1.5*(181.0-143.5) = 87.25.

3-4 Box Plots

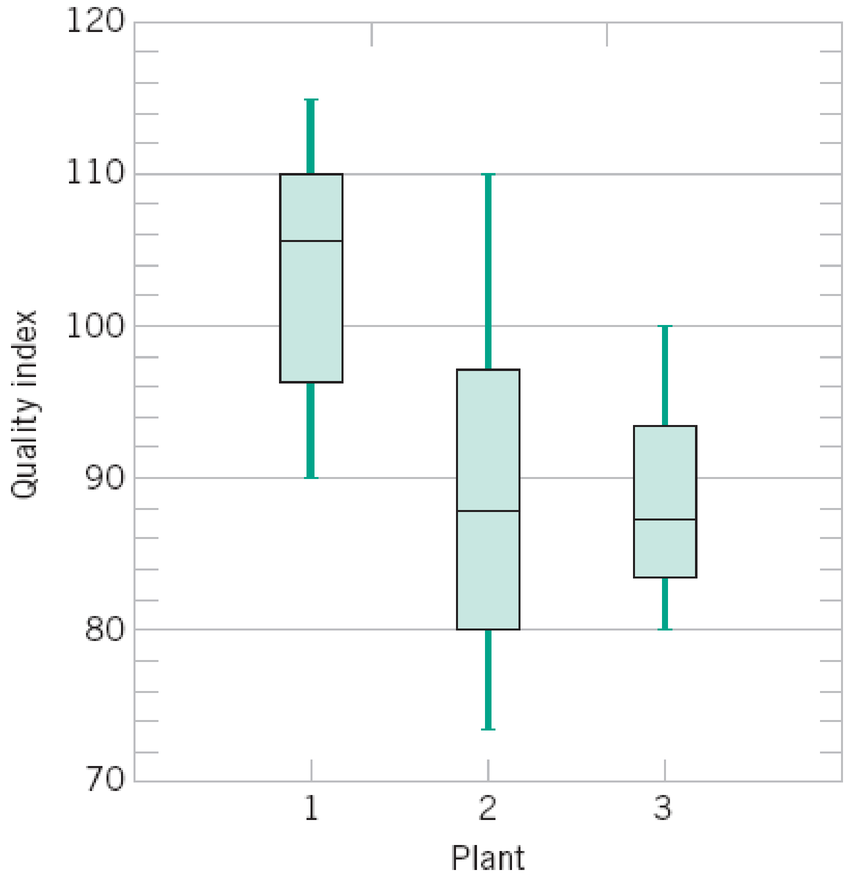


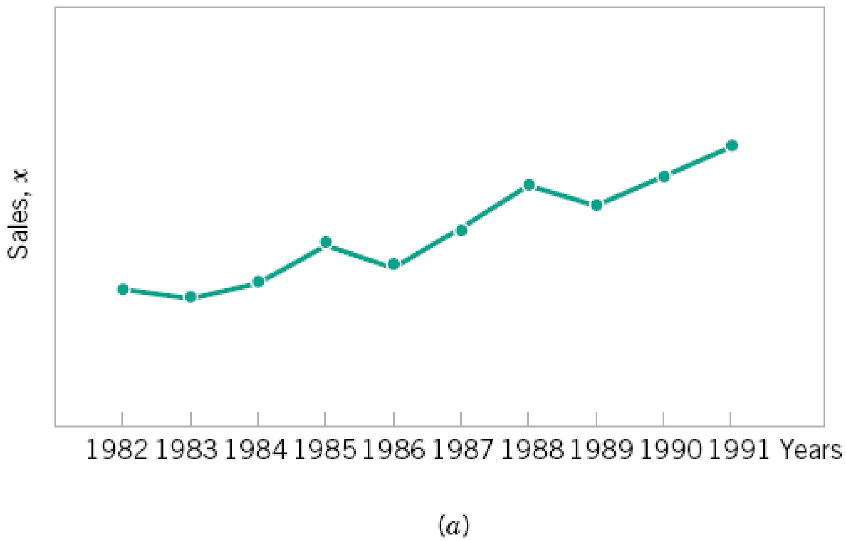
Figure 3-16 Comparative box plots of a quality index at three plants.

3-5 Time Series 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**

3-5 Time Series Plots

Non stable



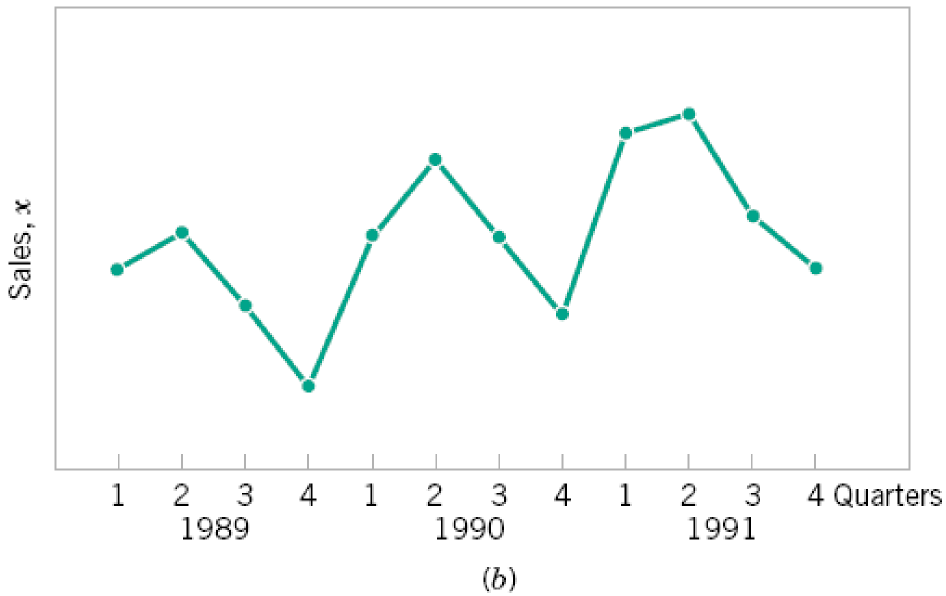


Figure 3-17 Company sales by year (a) and by quarter (b).

3-5 Time Series Plots

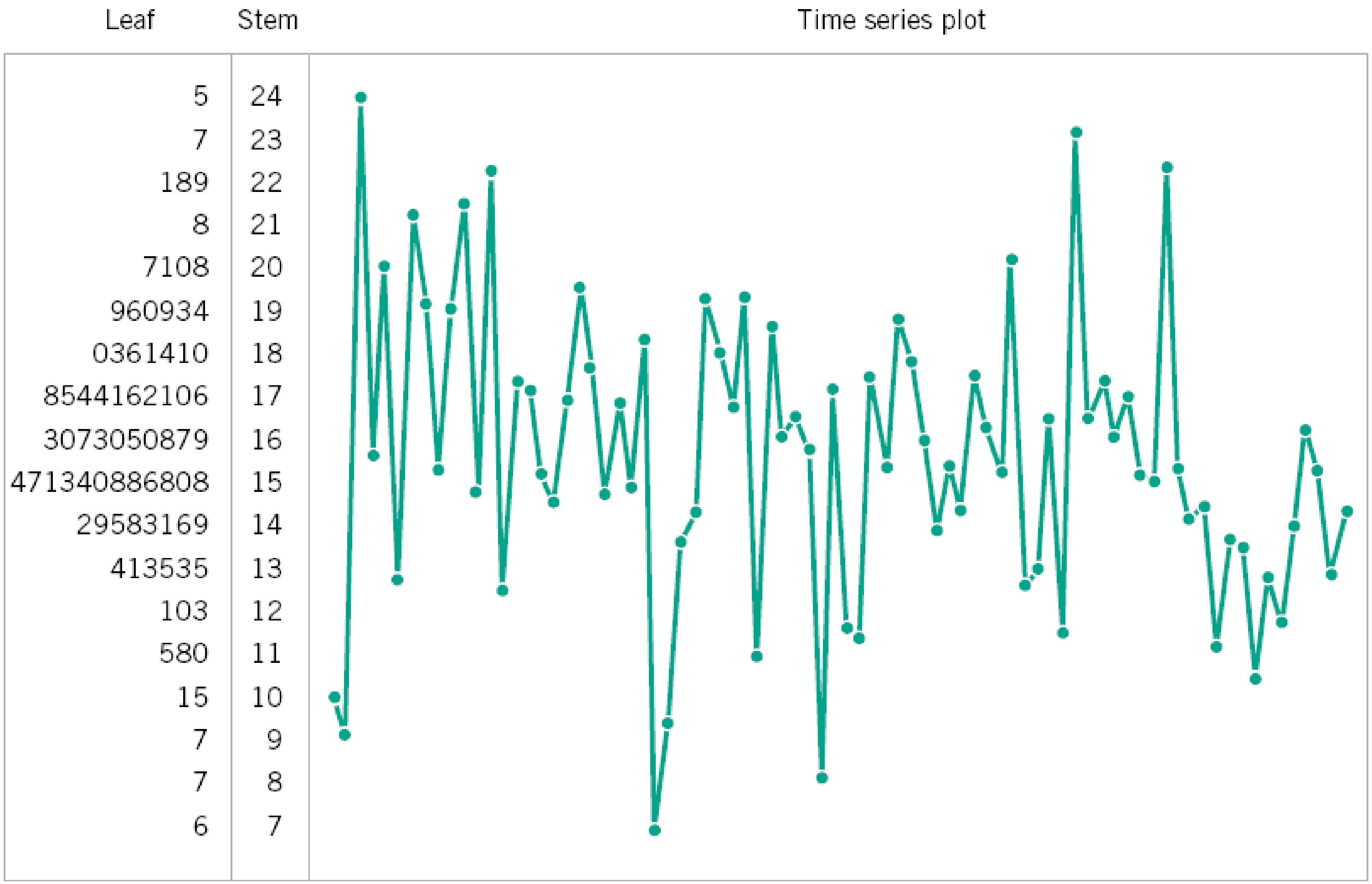


Figure 3-18 A digidot plot of the compressive strength data in Table 2-2.

3-5 Time Series Plots

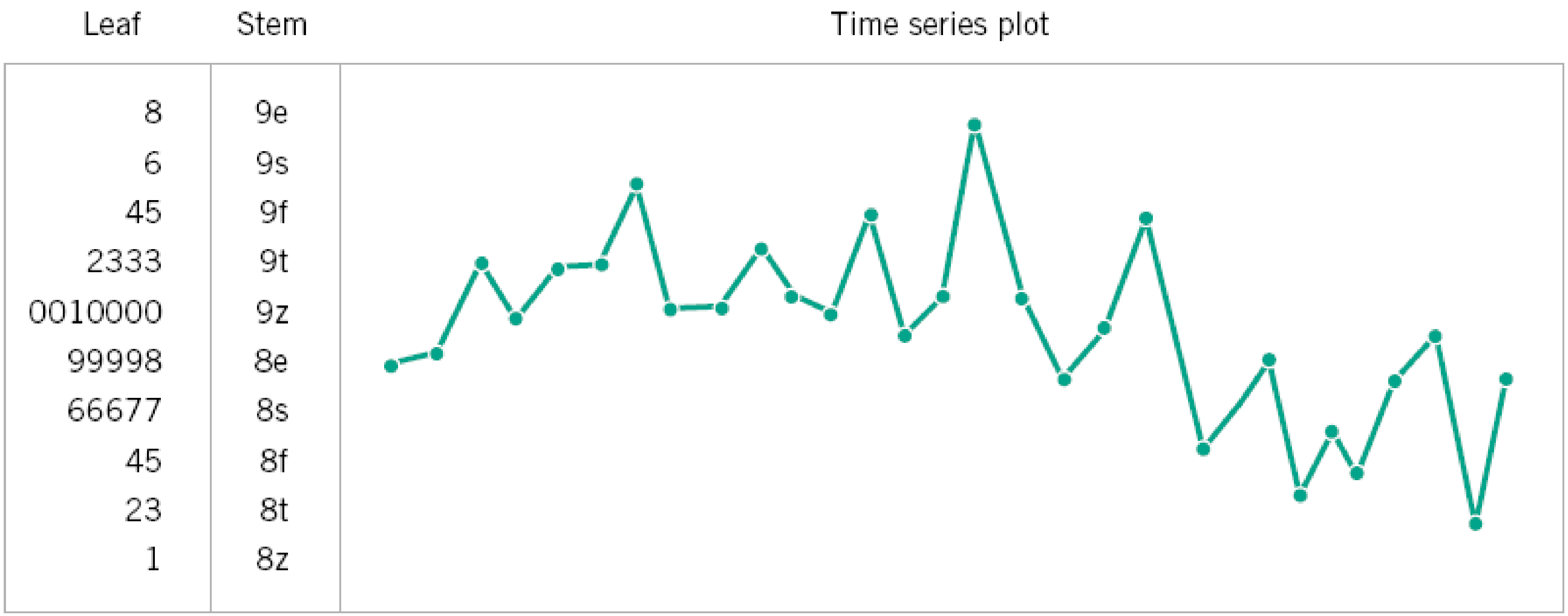


Figure 3-19 A digidot plot of chemical process concentration readings, observed hourly.

Probability Plots

- How do we know if a particular probability distribution is a reasonable model for a data set?
- We use a probability plot to verify such an assumption using a subjective visual examination.
- A histogram of a large data set reveals the shape of a distribution. The histogram of a small data set would not provide such a clear picture.
- A probability plot is helpful for all data set sizes.

How To Build a Probability Plot

- To construct a probability plot:
  - Sort the data observations in ascending order: $x_{(1)}$, $x_{(2)}, \ldots, x_{(n)}$.
  - The observed value $x_{(j)}$ is plotted against the cumulative distribution $(j - 0.5)/n$.
  - The paired numbers are plotted on the probability paper of the proposed distribution.
  - If the paired numbers form a straight line, it is reasonable to assume that the data follows the proposed distribution.

Example 6-7: Battery Life

The effective service life (minutes) of batteries used in a laptop are given in the table. We hypothesize that battery life is adequately modeled by a normal distribution. The probability plot is shown on normal probability vertical scale.

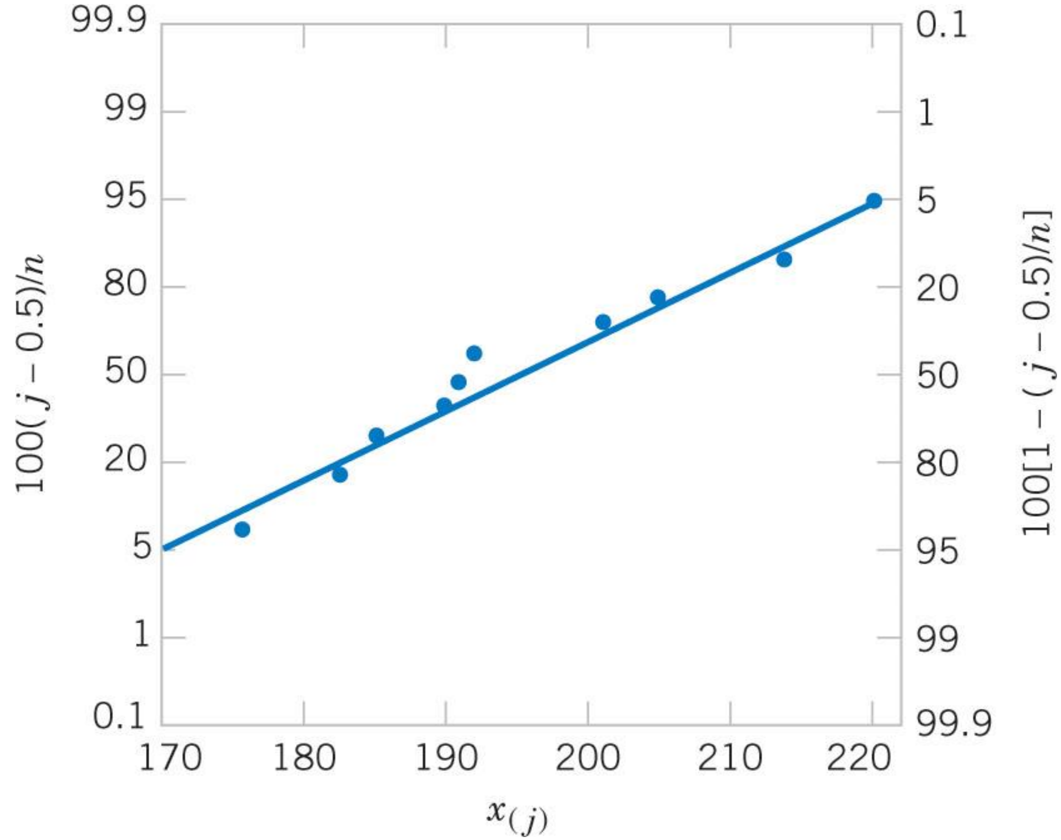


Figure 3-20 Normal probability plot for battery life.

Table 3-6 Calculations for Constructing a Normal Probability Plot

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

Probability Plot on Ordinary Axes

A normal probability plot can be plotted on ordinary axes using z-values. The normal probability scale is not used.

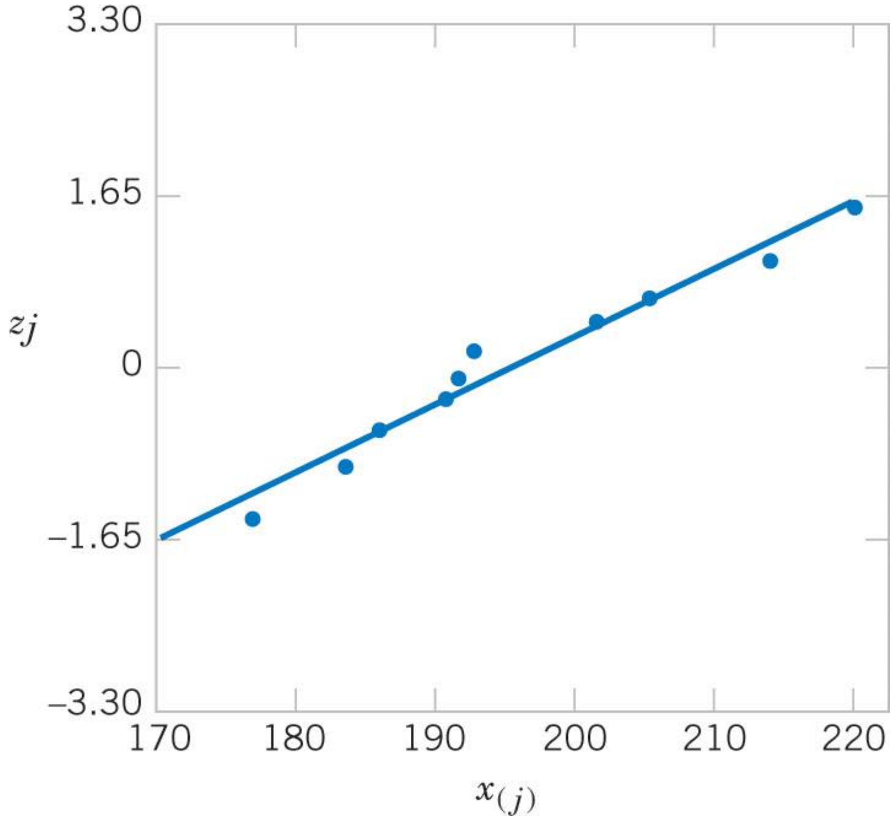


Figure 3-21 Normal Probability plot obtained from standardized normal scores. This is equivalent to Figure 3-20.

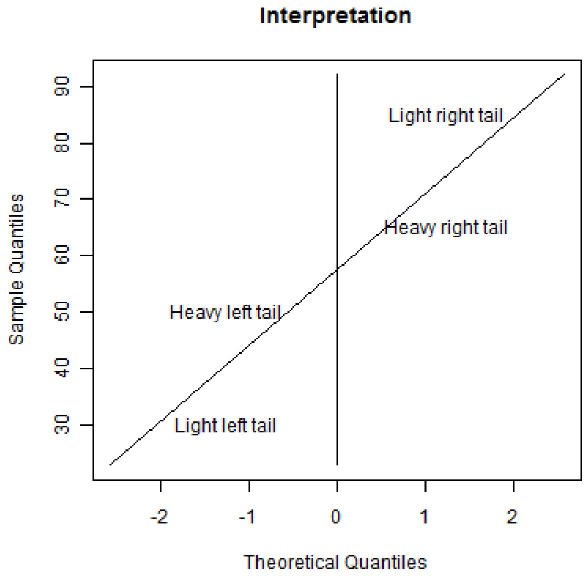
Table 3-6 Calculations for Constructing a Normal Probability Plot

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

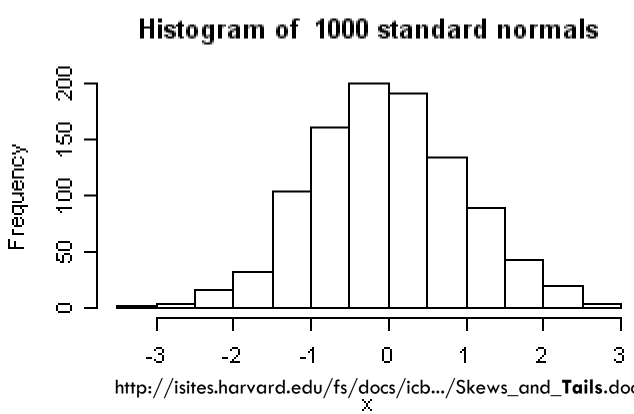
Use of the Probability Plot

- The probability plot can identify variations from a normal distribution shape.
  - Light tails of the distribution – more peaked.
  - Heavy tails of the distribution – less peaked.
  - Skewed distributions.
- Larger samples increase the clarity of the conclusions reached.

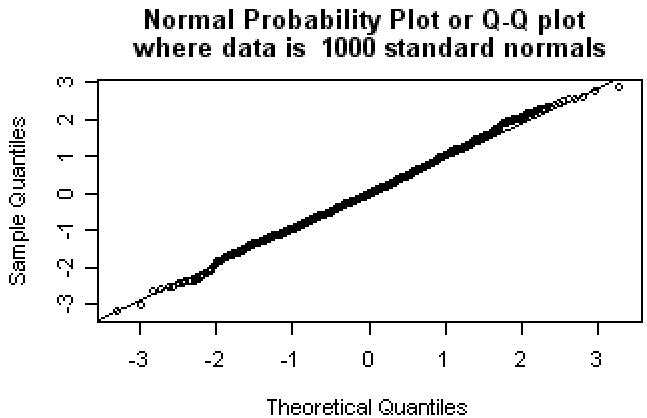
Probability Plots (Standard Normal)



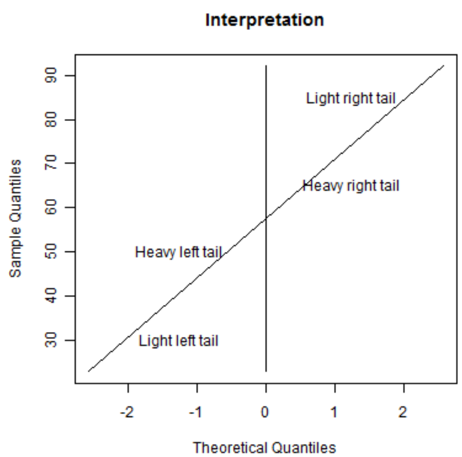
http://www.pmean.com/09/NormalPlot.html

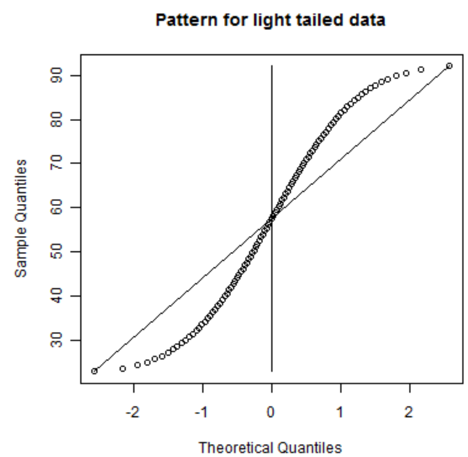


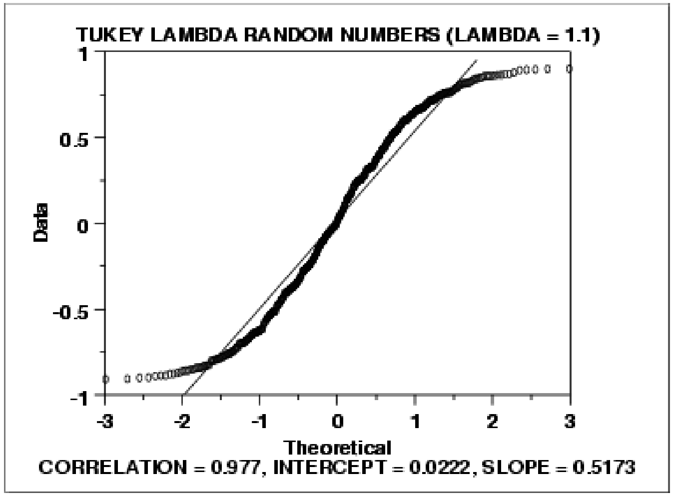
http://isites.harvard.edu/fs/docs/icb.../Skews_and_**Tails**.doc

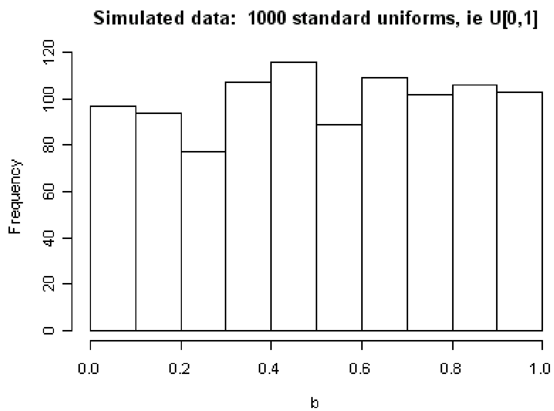


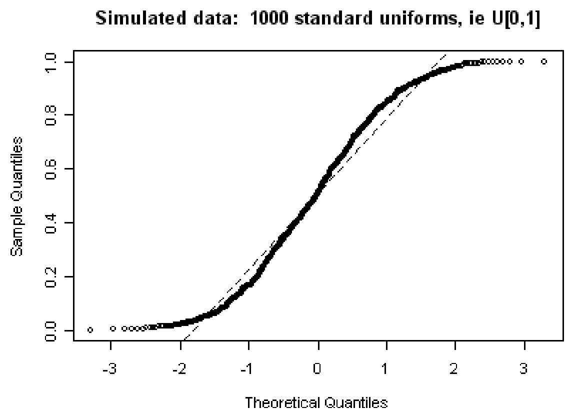
Probability Plots (Light Tailed)



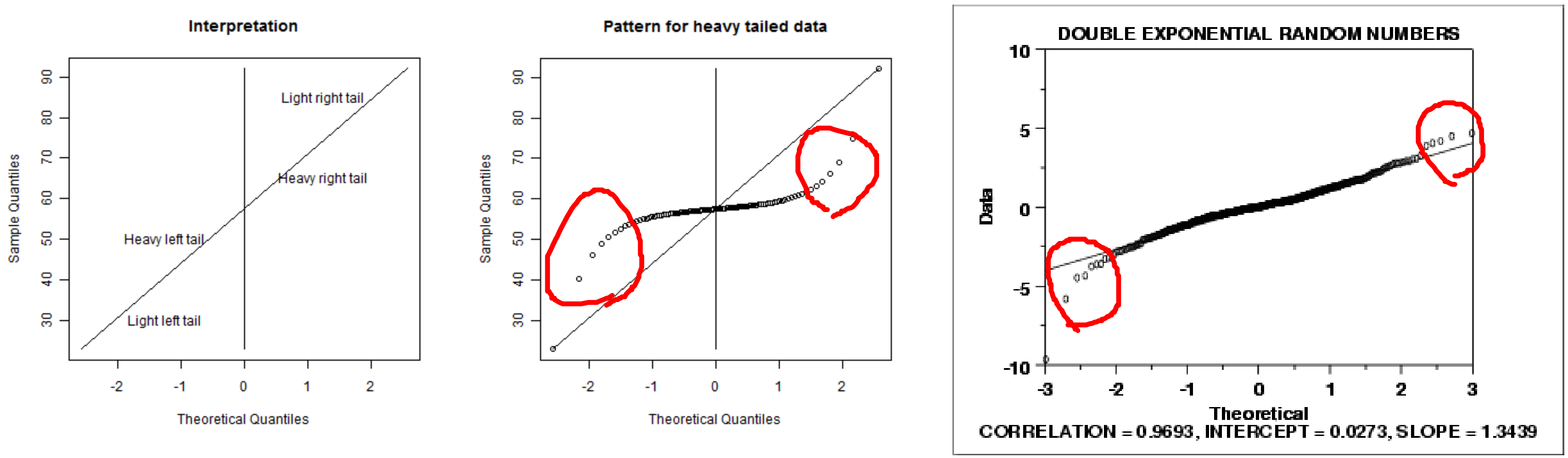


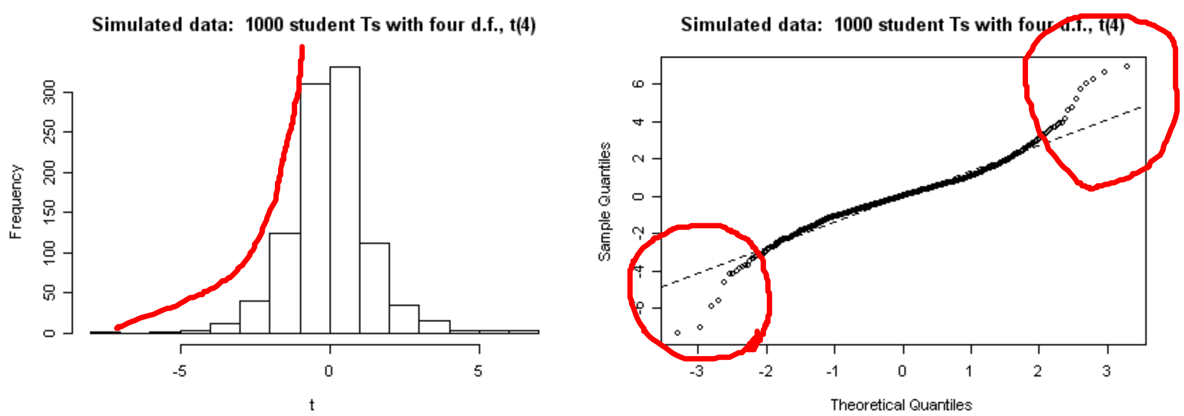




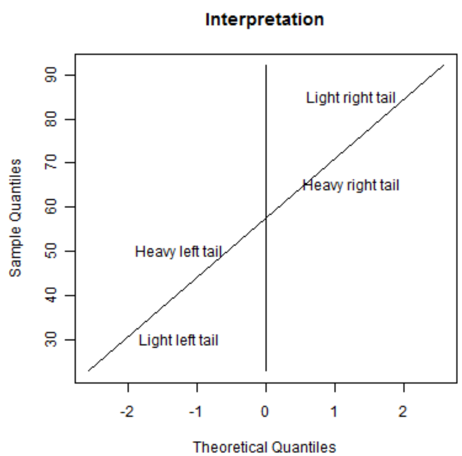


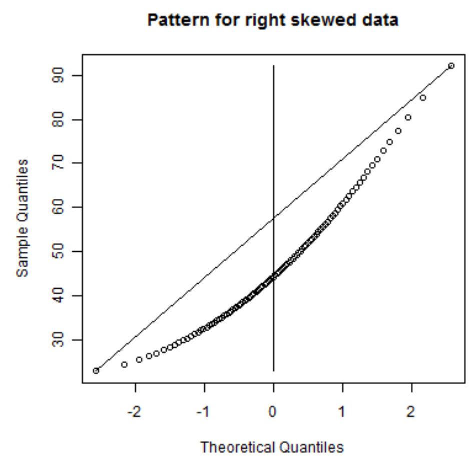
Probability Plots (Heavy Tailed)

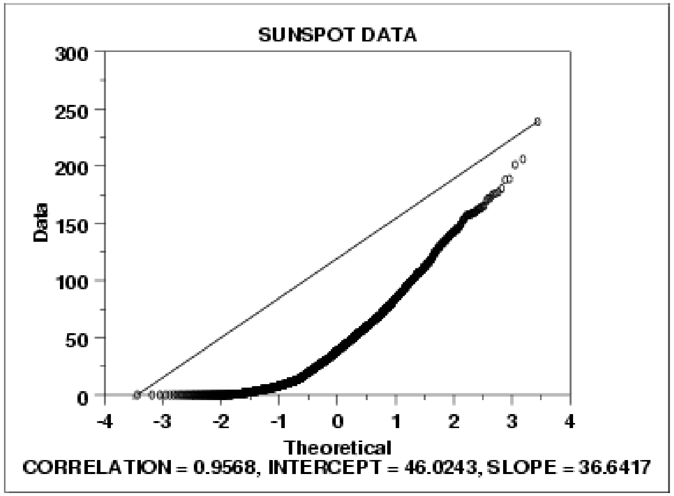


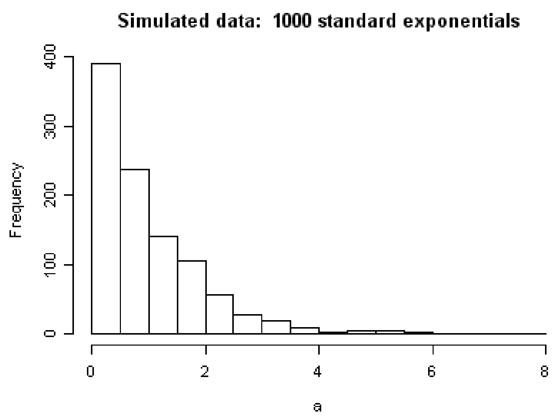


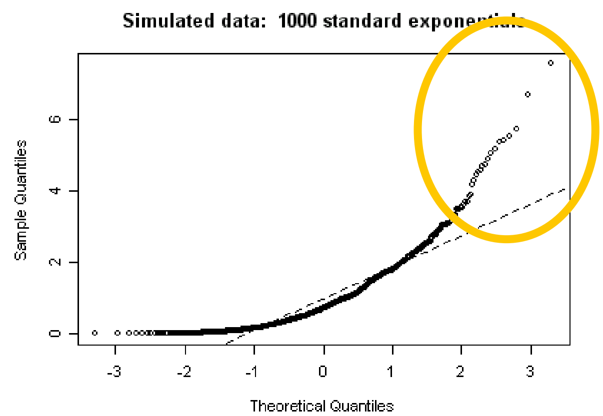
Probability Plots (Right Skewed)





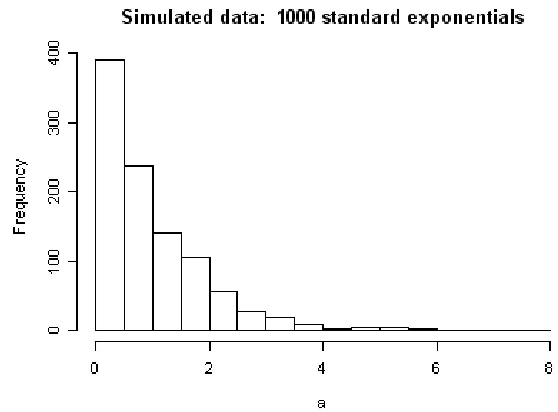


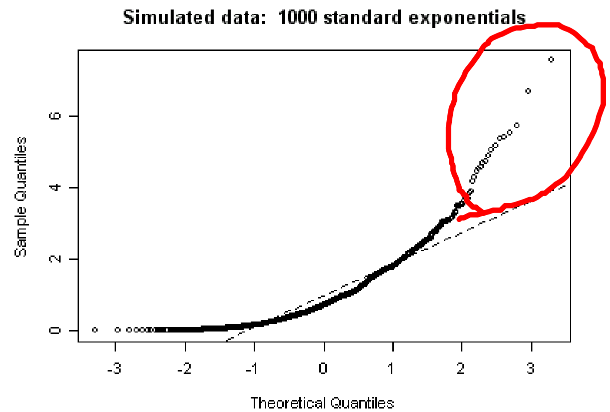




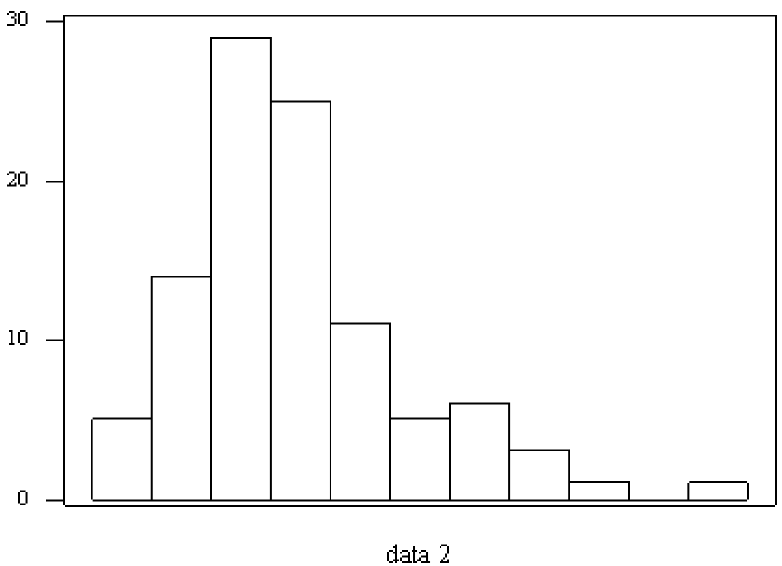
Heavy tail to the right

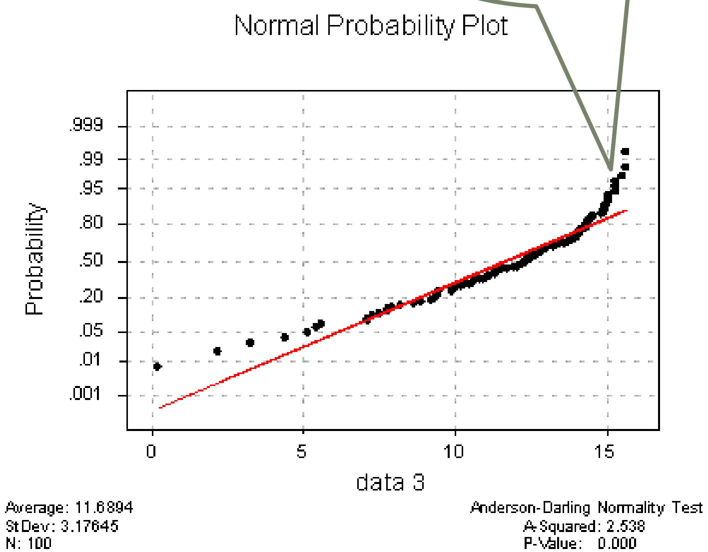
Probability Plots (Right Skewed)



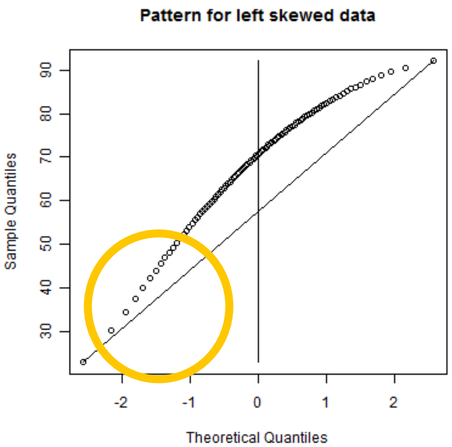


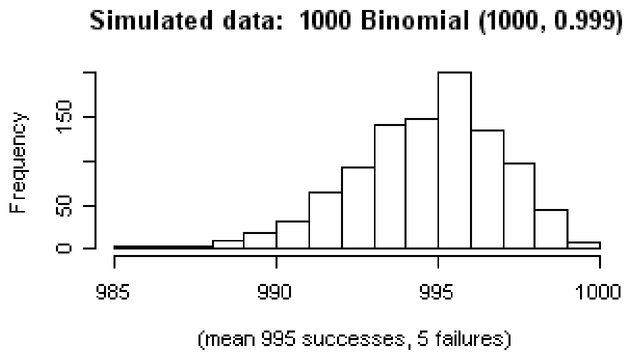
Heavy tailed to the right

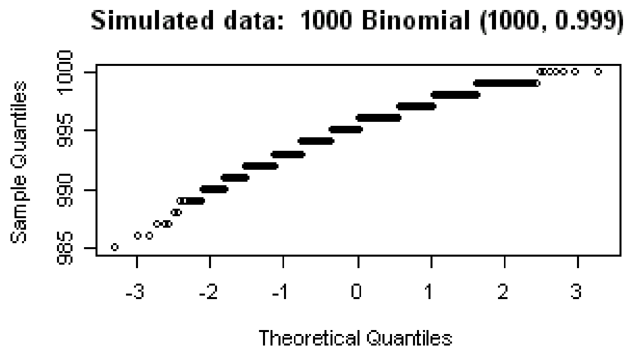


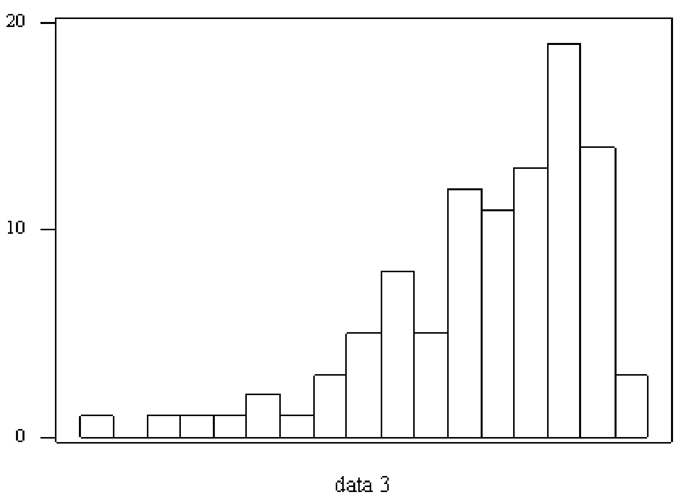


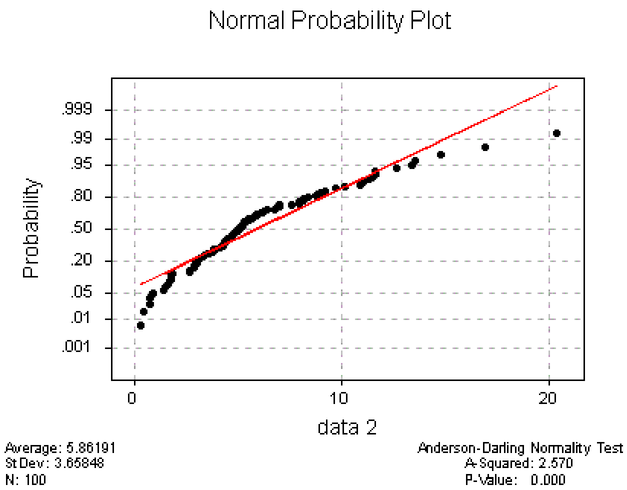
Probability Plots (Left Skewed)











Probability Plots with Minitab

- Obtained using Minitab menu: Graphics > Probability Plot. 14 different distributions can be used.
- The curved bands provide guidance whether the proposed distribution is acceptable – all observations within the bands is good.

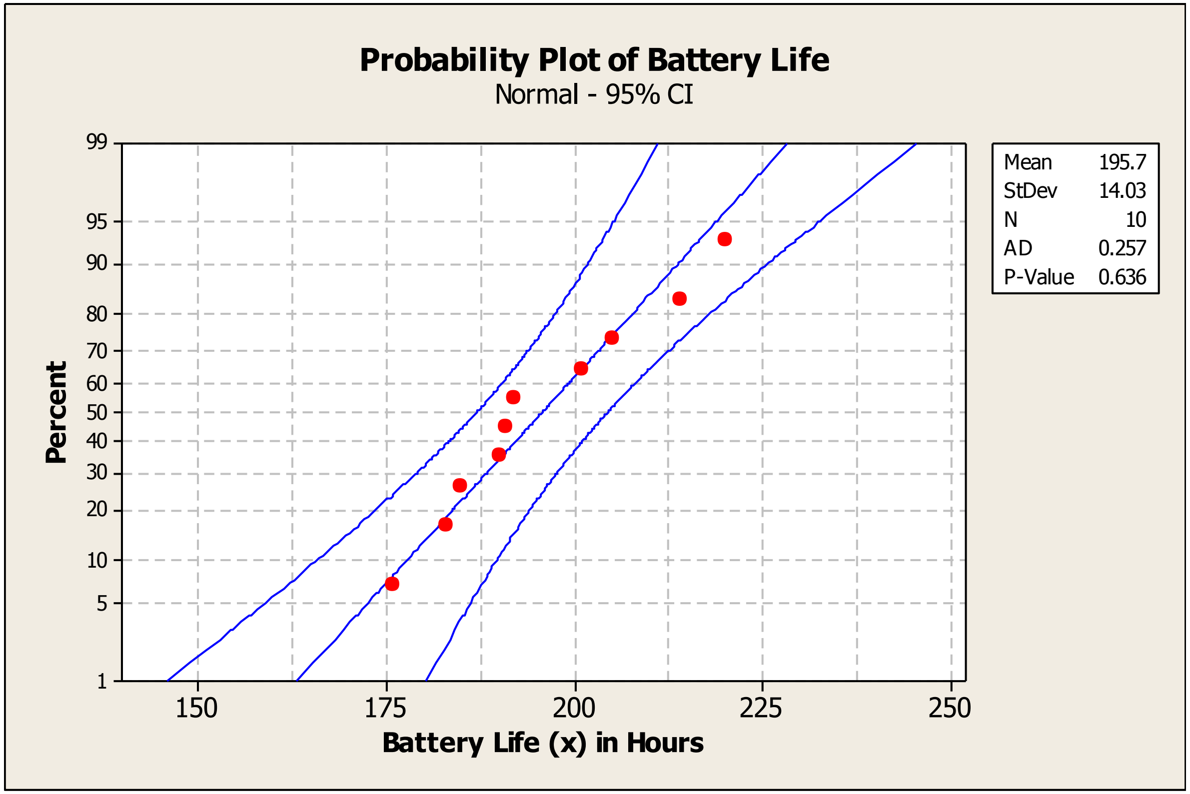


Figure 3-23

3-6 Multivariate Data

- The dot diagram, stem-and-leaf diagram, histogram, and box plot are descriptive displays for **univariate** data; that is, they convey descriptive information about a single variable.
- Many engineering problems involve collecting and analyzing **multivariate data**, or data on several different variables.
- In engineering studies involving multivariate data, often the objective is to determine the relationships among the variables or to build an empirical model.

3-6 Multivariate Data

Output effect (Dependence) independence variable

input cause

Table 3-7 Wire Bond Data

Observation Number	Pull Strength, y	Wire Length, x_1	Die Height, x_2	Observation Number	Pull Strength, y	Wire Length, x_1	Die Height, x_2
1	9.95	2	50	14	11.66	2	360
2	24.45	8	110	15	21.65	4	205
3	31.75	11	120	16	17.89	4	400
4	35.00	10	550	17	69.00	20	600
5	25.02	8	295	18	10.30	1	585
6	16.86	4	200	19	34.93	10	540
7	14.38	2	375	20	46.59	15	250
8	9.60	2	52	21	44.88	15	290
9	24.35	9	100	22	54.12	16	510
10	27.50	8	300	23	56.63	17	590
11	17.08	4	412	24	22.13	6	100
12	37.00	11	400	25	21.15	5	400
13	41.95	12	500				

3-6 Multivariate Data

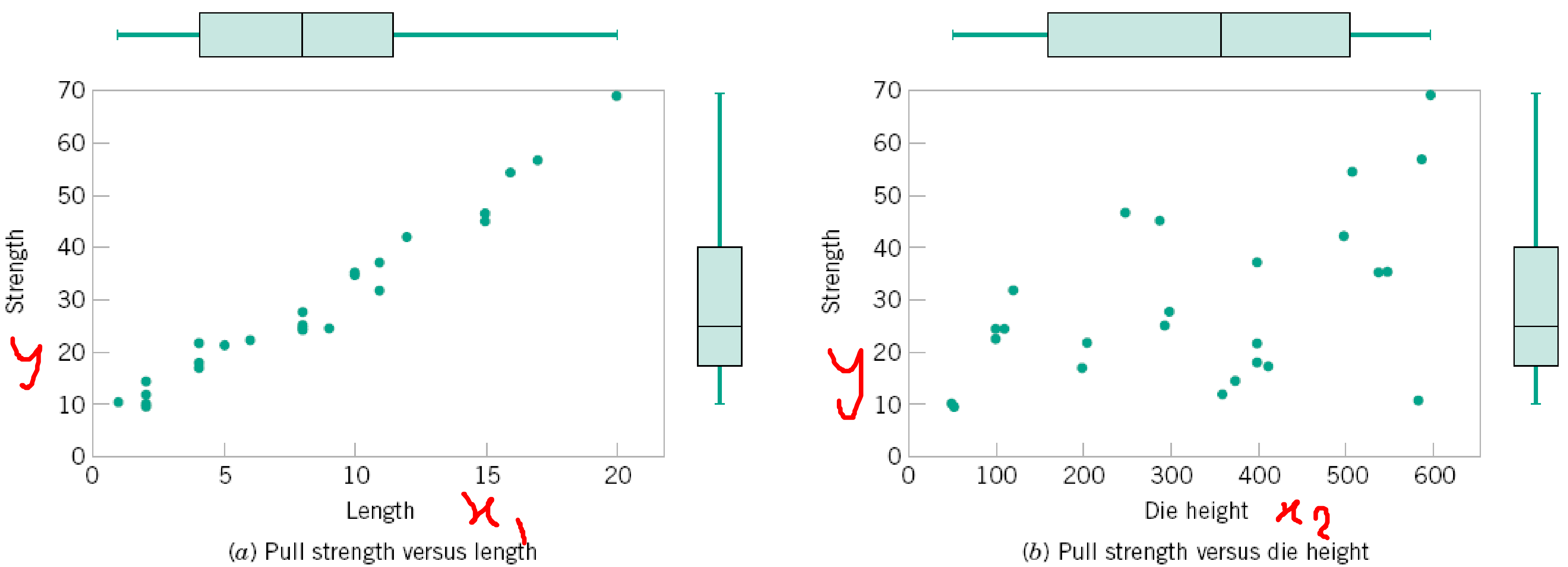


Figure 3-24 Scatter diagrams and box plots for the wire bond pull strength data in Table 1-1. (*a*) Pull strength versus length. (*b*) Pull strength versus die height.

3-6 Multivariate Data

Sample Correlation Coefficient

Population

Given n pairs of data $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$, the **sample correlation coefficient** r is defined by

ρ

$$r = \frac{S_{xy}}{\sqrt{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)\left(\sum_{i=1}^{n}(y_i - \overline{y})^2\right)}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \qquad (3\text{-}7)$$

with $-1 \leq r \leq +1$.

Correlation is to scale it down in order to compare the same boundary.

3-6 Multivariate Data

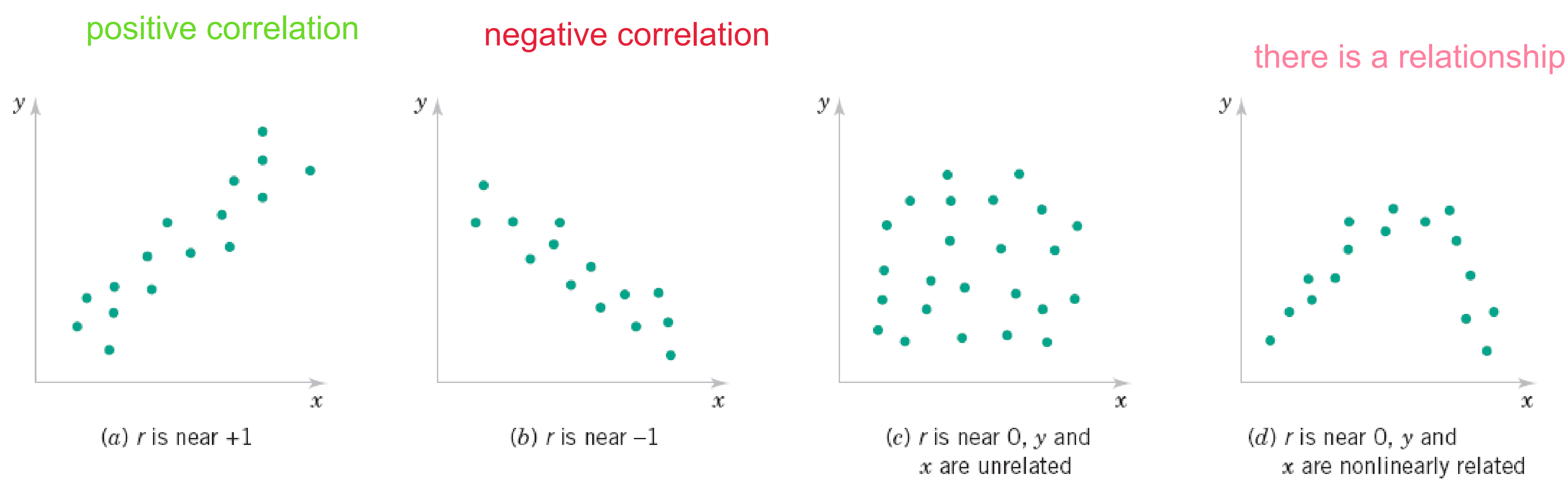


Figure 3-25 Scatter diagrams for different values of the sample correlation coefficient r. (a) r is near $+1$. (b) r is near -1. (c) r is near 0; y and x are unrelated. (d) r is near 0; y and x are nonlinearly related.

3-6 Multivariate Data

Table 3-8 Data on Shampoo

Foam	Scent	Color	Residue	Region	Quality
6.3	5.3	4.8	3.1	1	91
4.4	4.9	3.5	3.9	1	87
3.9	5.3	4.8	4.7	1	82
5.1	4.2	3.1	3.6	1	83
5.6	5.1	5.5	5.1	1	83
4.6	4.7	5.1	4.1	1	84
4.8	4.8	4.8	3.3	1	90
6.5	4.5	4.3	5.2	1	84
8.7	4.3	3.9	2.9	1	97
8.3	3.9	4.7	3.9	1	93
5.1	4.3	4.5	3.6	1	82
3.3	5.4	4.3	3.6	1	84
5.9	5.7	7.2	4.1	2	87
7.7	6.6	6.7	5.6	2	80
7.1	4.4	5.8	4.1	2	84
5.5	5.6	5.6	4.4	2	84
6.3	5.4	4.8	4.6	2	82
4.3	5.5	5.5	4.1	2	79
4.6	4.1	4.3	3.1	2	81
3.4	5.0	3.4	3.4	2	83
6.4	5.4	6.6	4.8	2	81
5.5	5.3	5.3	3.8	2	84
4.7	4.1	5.0	3.7	2	83
4.1	4.0	4.1	4.0	2	80

3-6 Multivariate Data

foam and foam —> 1
scent and scent —> 1
….. because it is the same thing

Table 3-9	Foam	Scent	Color	Residue	Region
Scent	0.002				
Color	0.328	0.599			
Residue	0.193	0.500	0.524		
Region	−0.032	0.278	0.458	0.165	
Quality	0.512	−0.252	−0.194	−0.489	−0.507

3-6 Multivariate Data

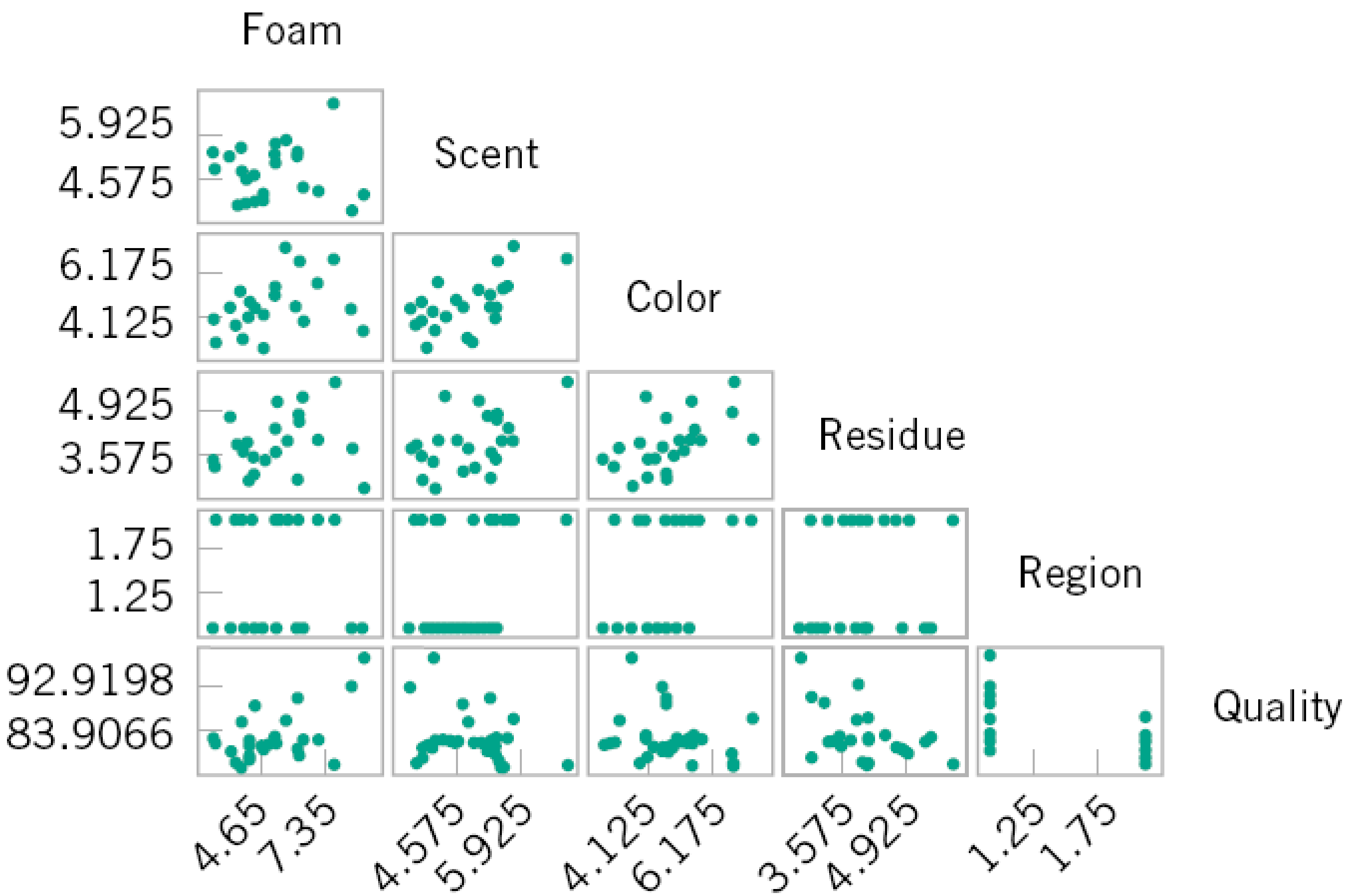


Figure 3-26 Matrix of scatter plots for the shampoo data in Table 3-9

3-6 Multivariate Data

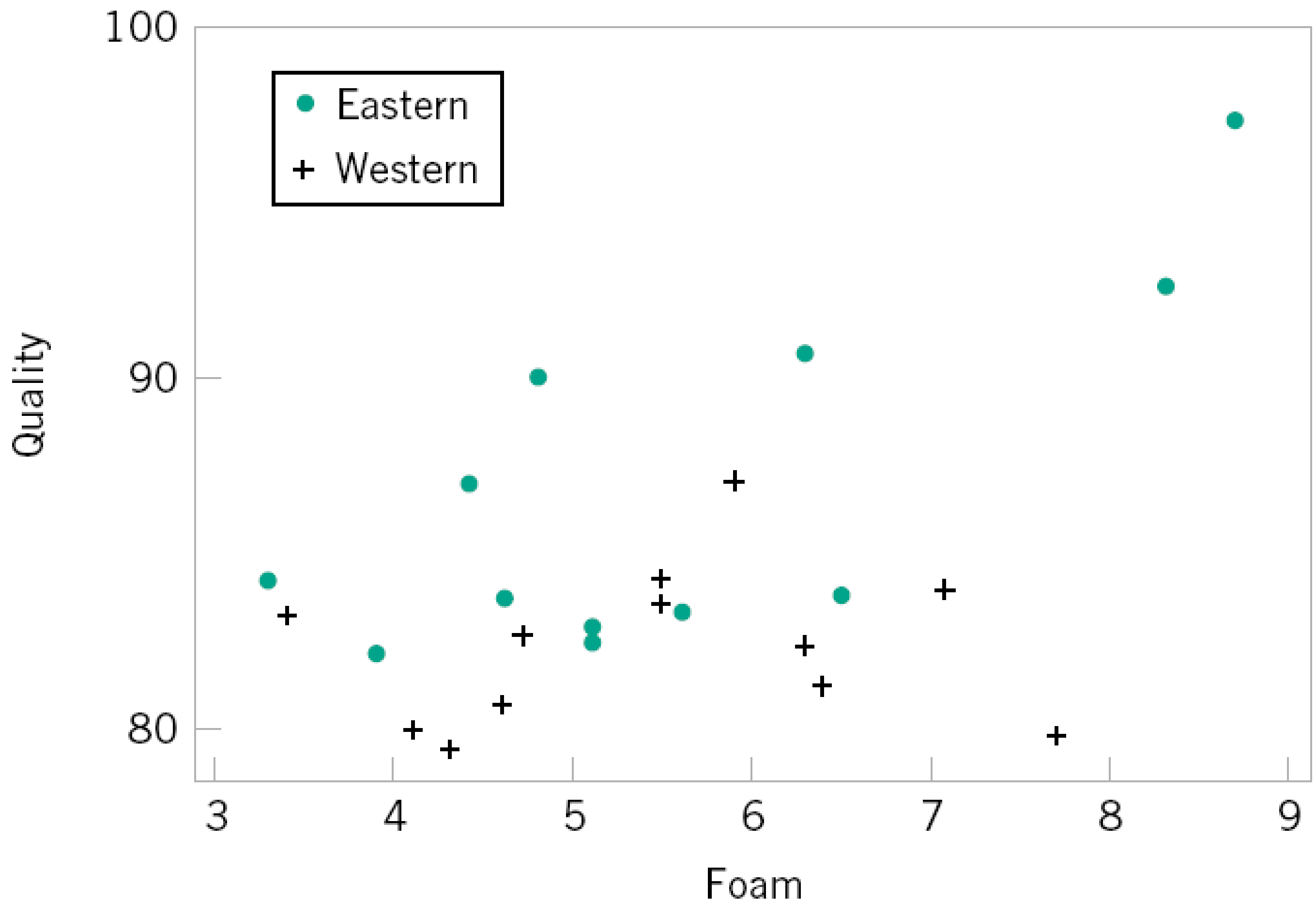


Figure 3-27 Scatter diagram of shampoo quality versus foam.