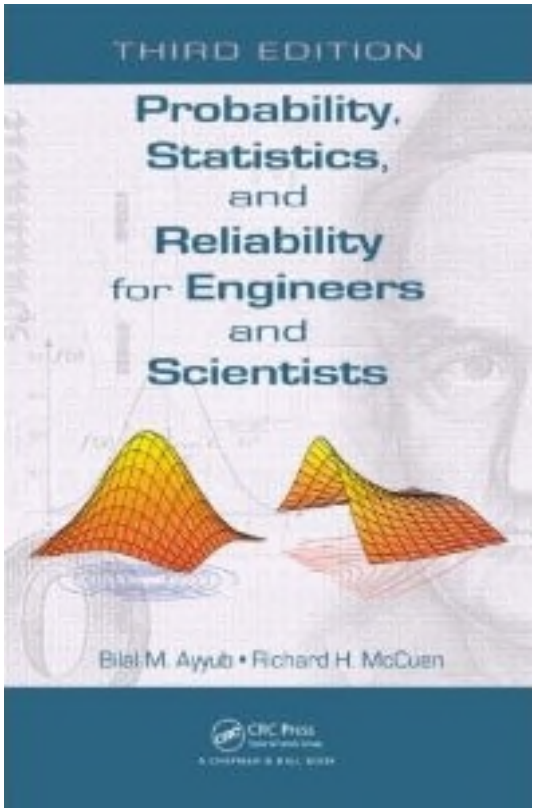
CHAPTER

2

CHAPMAN HALL/CRC

Data Description and Treatment



Probability, Statistics, and Reliability for Engineers and Scientists, Third Edition

Bilal M. Ayyub and Richard H. McCuen

Introduction

- Dealing with *uncertainty* and *dispersion* requires collecting data and information about some variables that are needed to solve an engineering problem.
- The collected data can be used to to establish some understanding about the relationships among the different variables.
- When data are collected, it is necessary to utilize some techniques for:
  - Describing the data, and
  - Treating and analyzing the data.

Classification of Data

- Data can be measured on one of the following scales:
  1. Nominal (lowest level)
  2. Ordinal
  3. Interval
  4. Ratio (highest level)
- Data can be classified based on their dimensionality (the number of axes needed to graphically present the data).

Nominal Scale

- The nominal scale of measurement is at the lowest level (no order to the data).
- Nominal scales are both discrete and qualitative.
- Measurements consists of identifying the sample as belonging to one of several categories.

Nominal Scale

- Examples:
  - Gender: female or male
  - Political affiliation: Republican (Bush), Democrat (Gore), or Independent (Ventura)
  - College major: engineering, science,history, etc.
  - Project failed or not failed
  - Fatal and non-fatal accidents

Ordinal Scale

- Ordinal scale is a higher than the nominal scale.
- There is order among the group, but the magnitude of the difference between groups is not meaningful.
  - Military ranks are measured on an ordinal scale
    - The major is above the sergeant and the sergeant is above the private, but we cannot say that the major is two or three times higher than a sergeant.

Ordinal Scale

- Examples:
  - Infiltration potential of soil texture classes
  - Hazard classifications for dam design
    - High hazard
    - Moderate hazard
    - Low hazard

Interval Scale

- Interval scale is a higher scale than the ordinal scale.
- Interval scale is similar to the ordinal scale with the exception that the difference between the groups is meaningful.
  - A difference in temperature of 10^0 F is less than a difference of 20^0 F.
- Values on an interval scale may be treated with arithmetic operators (+, -, *, /).

Interval Scale

- Examples
  - Yield strength of concrete
  - Compression strength of concrete
  - Shear stress of soil
  - Annual number of traffic fatalities
  - Number of lost worker-hours on construction sites due to accidents

Ratio Scale

- Ratio scale is a higher scale than the interval scale.
- Ratio scale is similar to the interval scale with the exception that the zero is absolute rather than by convention.
  - Absolute temperature
  - Variance or error.
- Values on an ratio scale may be treated with arithmetic operators (+, -, *, /).

Dimensionality of Data

- Definition:

  “The dimensionality of data is defined as the number of axes needed to represent the data”

- Examples:
  - One-dimensional Data
    - Fatal traffic accidents for each state in 1996
  - Two-dimensional Data
    - Corrosion rate of steel as a function of the length of time that the steel has been exposed to a corrosive environment

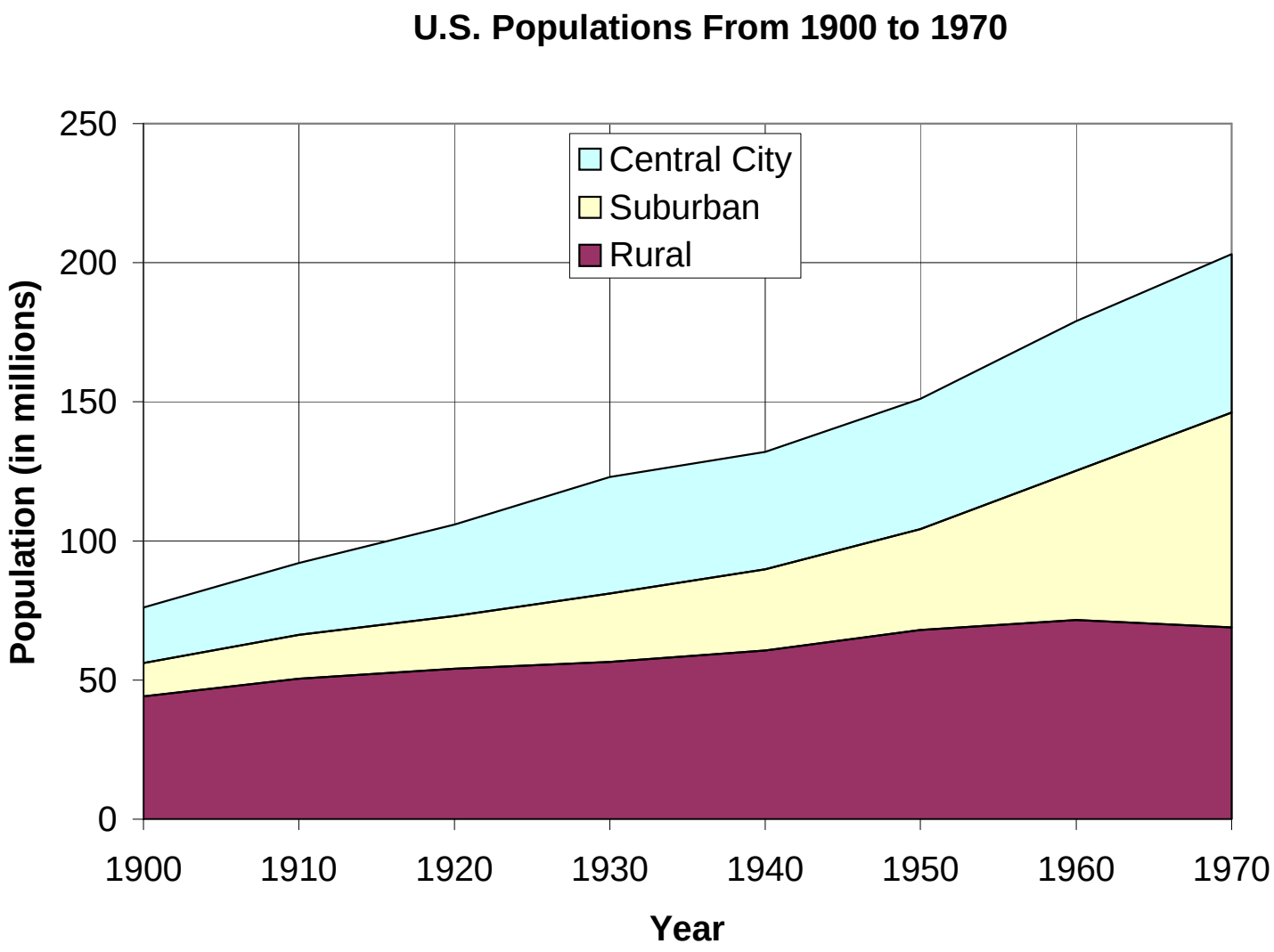
Graphical Description of Data

- Types of Graphical Descriptors
  - Area Charts
  - Pie Charts
  - Column Charts
  - Scatter Diagrams
  - Line charts
  - Combination Charts
  - Three Dimensional Charts

Area Charts

- Area charts are useful for three-dimensional data that include both nominal and interval-independent variables.
  - The value of the dependent variable measured on an interval scale and cumulated over all values of the nominal variable.

Area Charts

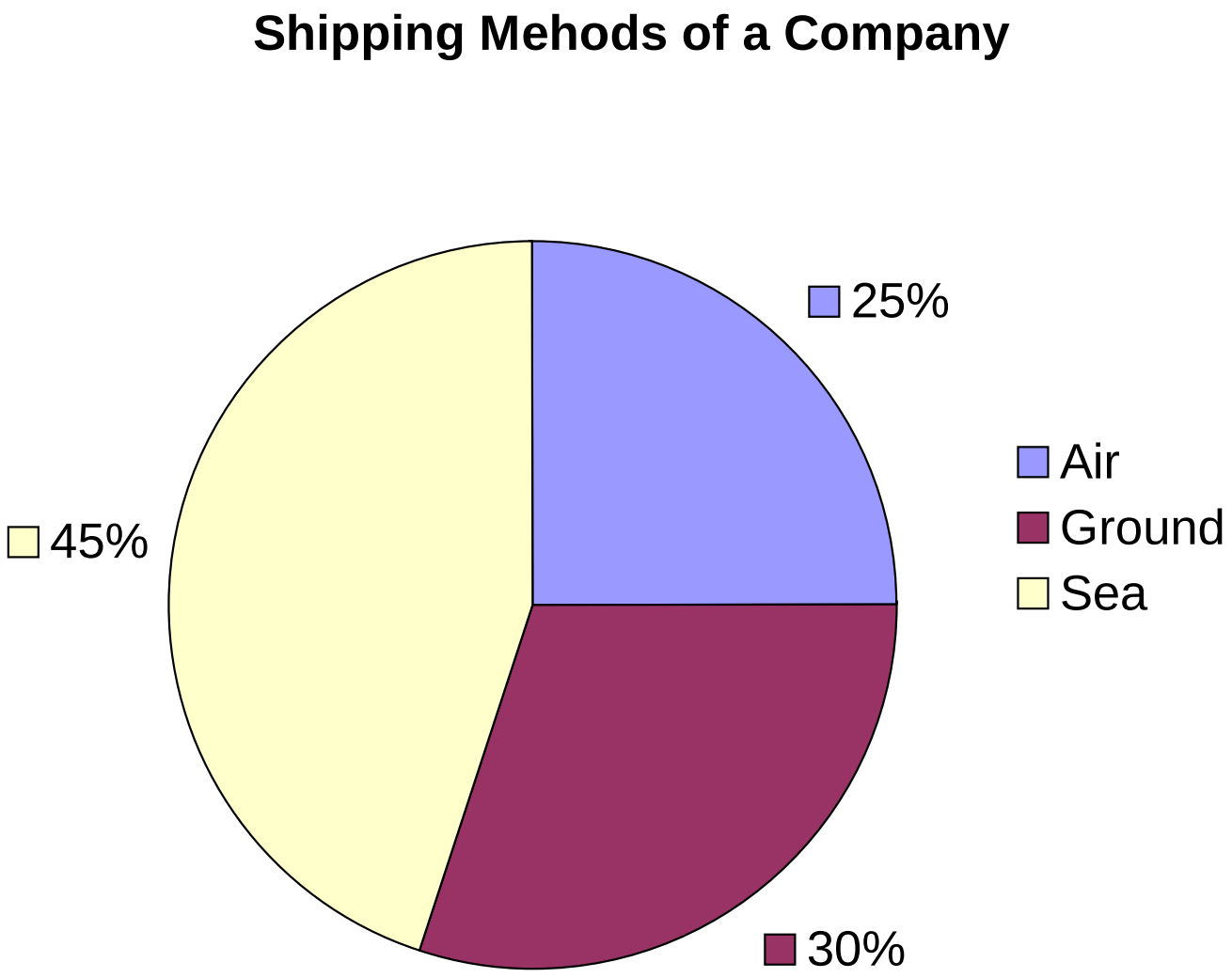


Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Pie Charts

- A pie chart is generally used to show how a whole is divided among several categories.
- The amount in each category is expressed in percentages, fractions, or proportions.
- A circle is divided into segments (slices of pie) proportional to the percentages of each category.
- The variable presented in a pie chart is measured on an interval scale.

Pie Charts



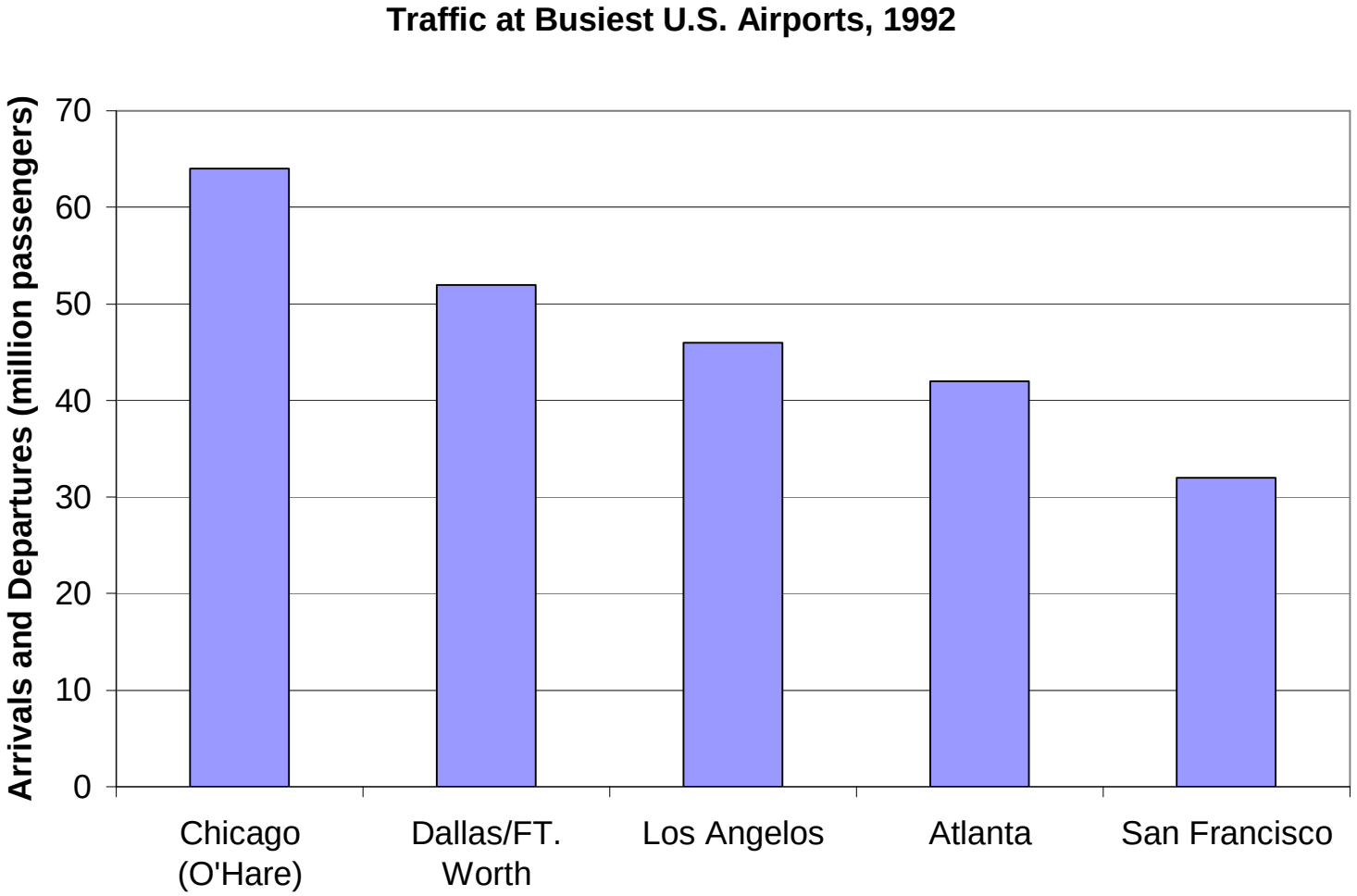
Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Bar Charts

- Bar charts are widely used because they are easy to construct and easy to read.
- They are effective in presenting visual interpretations or comparisons of data.
- Bar charts are useful for data recorded on an interval scale with one or more independent variables recorded on nominal or ordinal scales.
  - Bar charts can be divided into two broad categories:
    1. Vertical Bar Charts
    2. Horizontal Bar Charts

Bar Charts

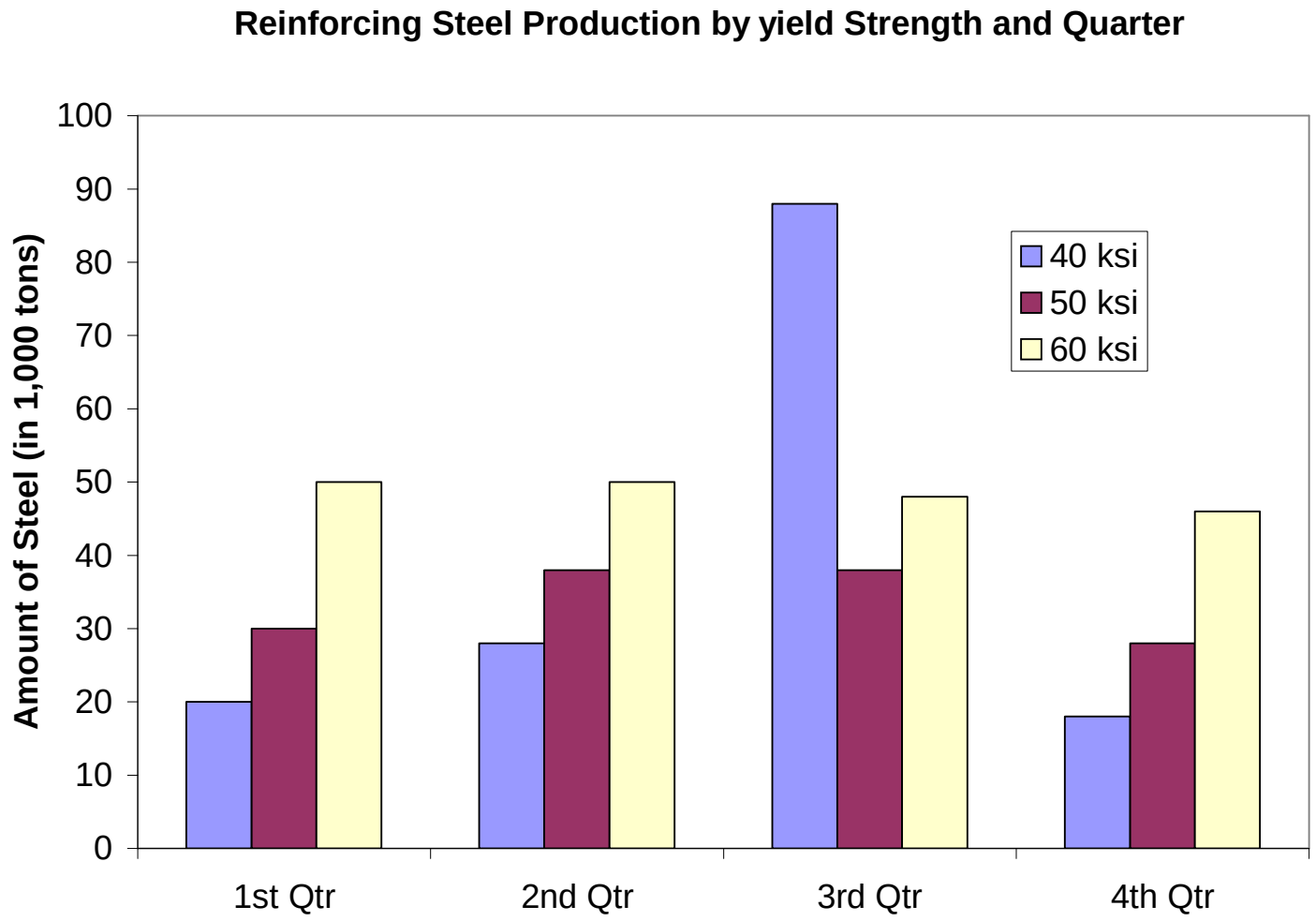
Example 1a: Vertical Bar Chart



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Bar Charts

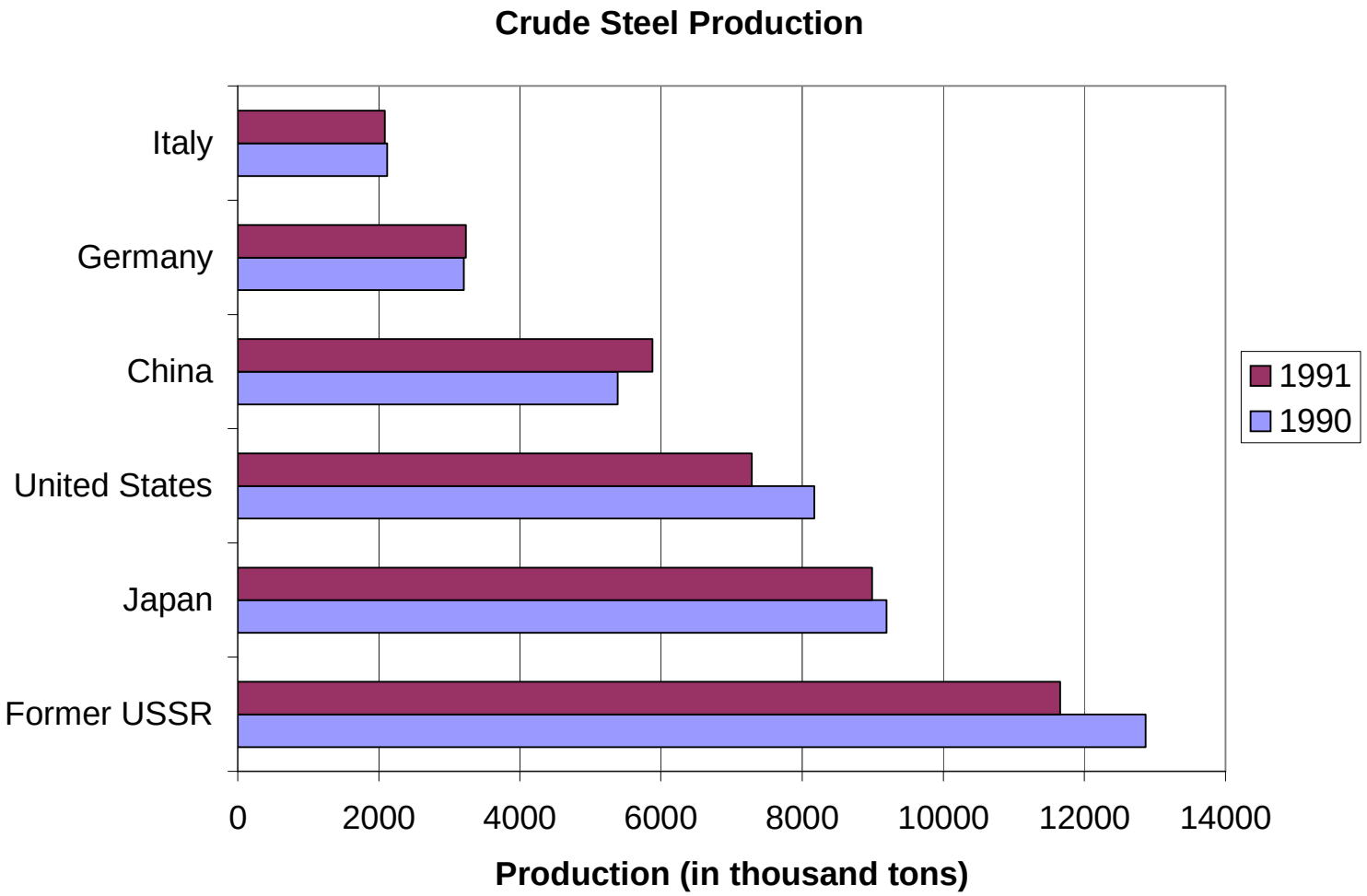
Example 1b: Vertical Bar Chart



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Bar Charts

Example 2: Horizontal Bar Chart



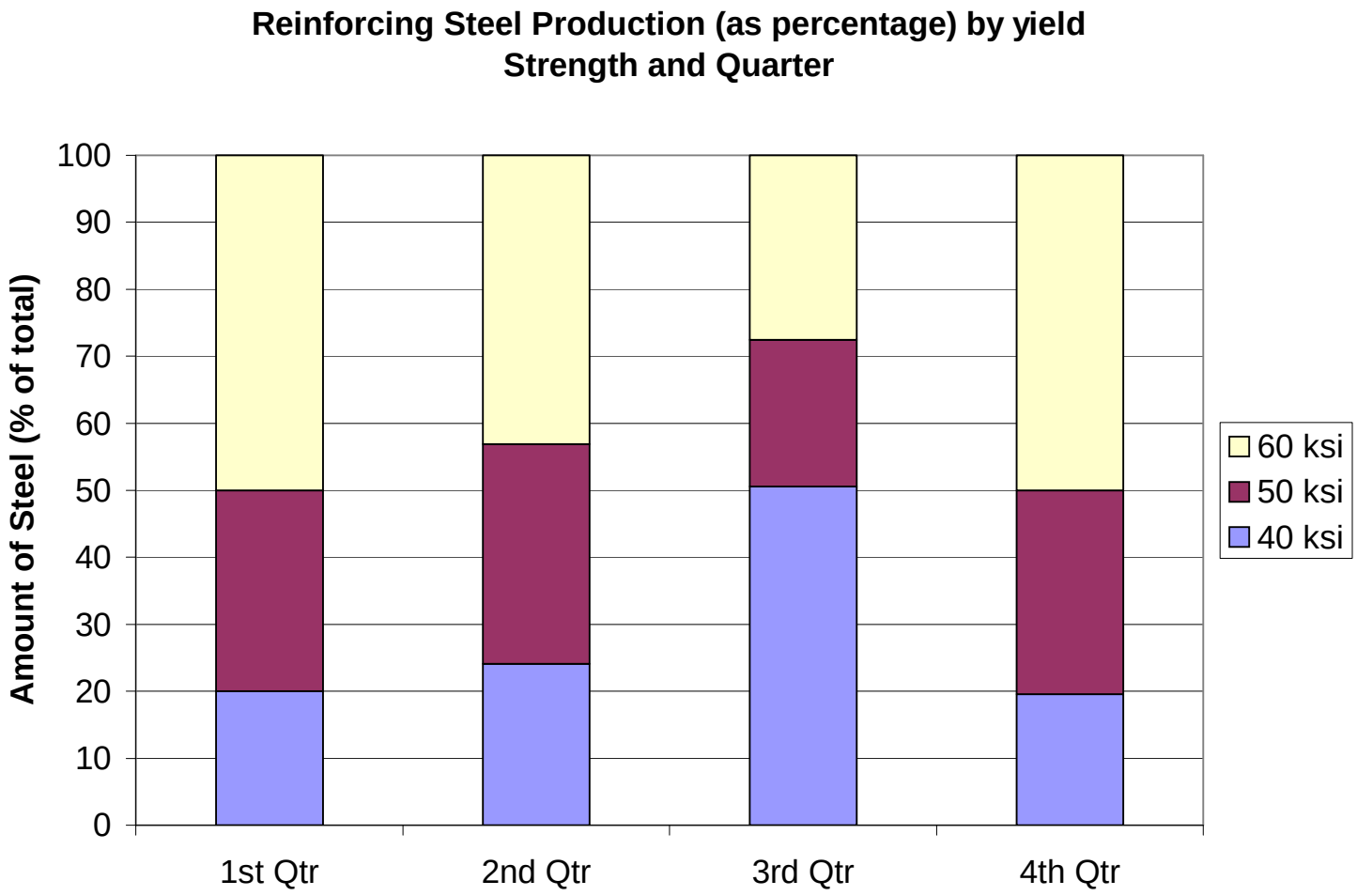
Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Column Charts

- Column charts are very similar to bar charts.
- Unlike the bar chart, the dependent variable is expressed as a percentage (or fraction) of a total.
- In this case, one of the independent variables is used for the abscissa, while the dependent variable is shown as percentages (or fractions) of the second independent variable.

Column Charts

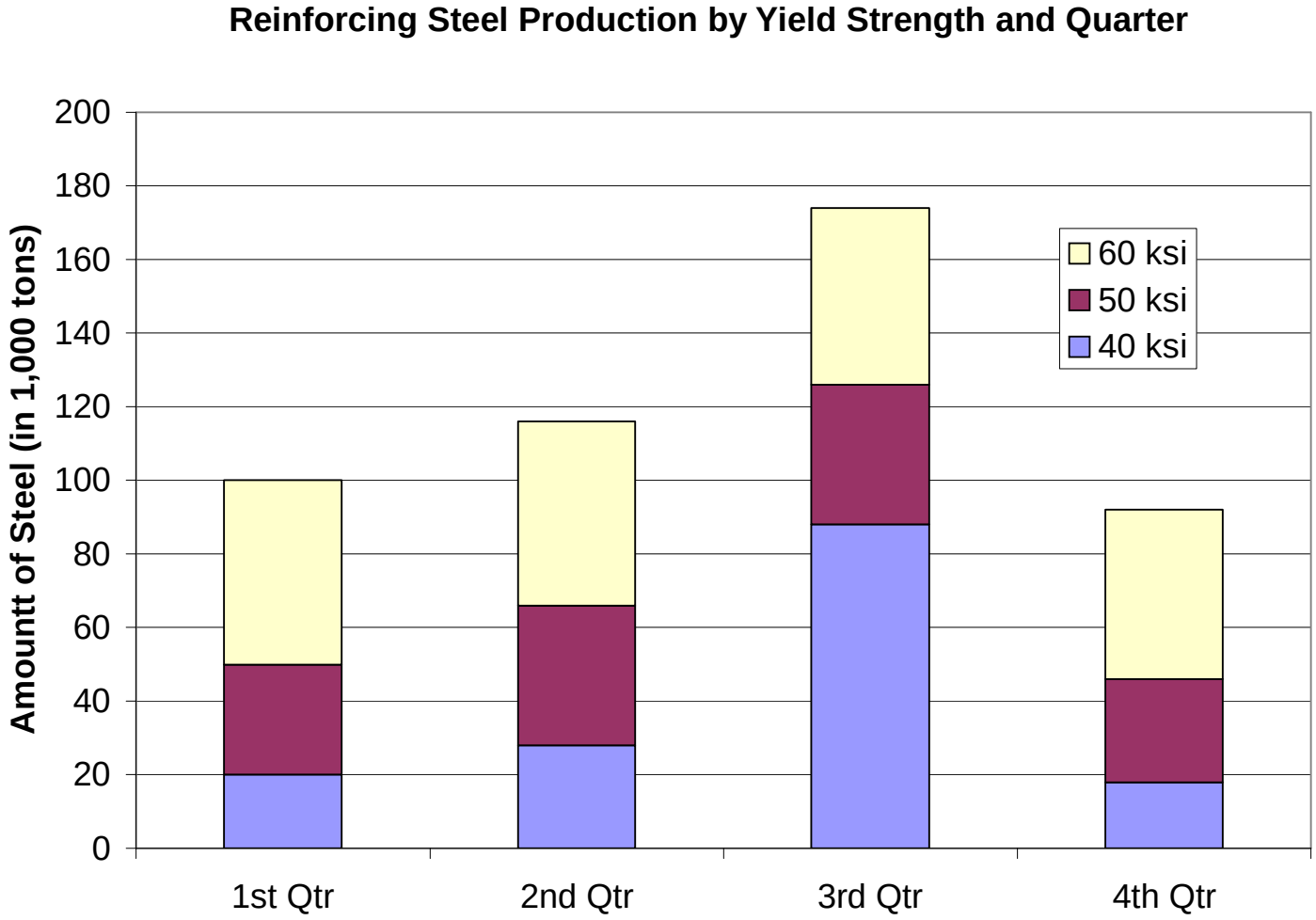
Example 1: Column Chart



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Column Charts

Example 2: Column Chart



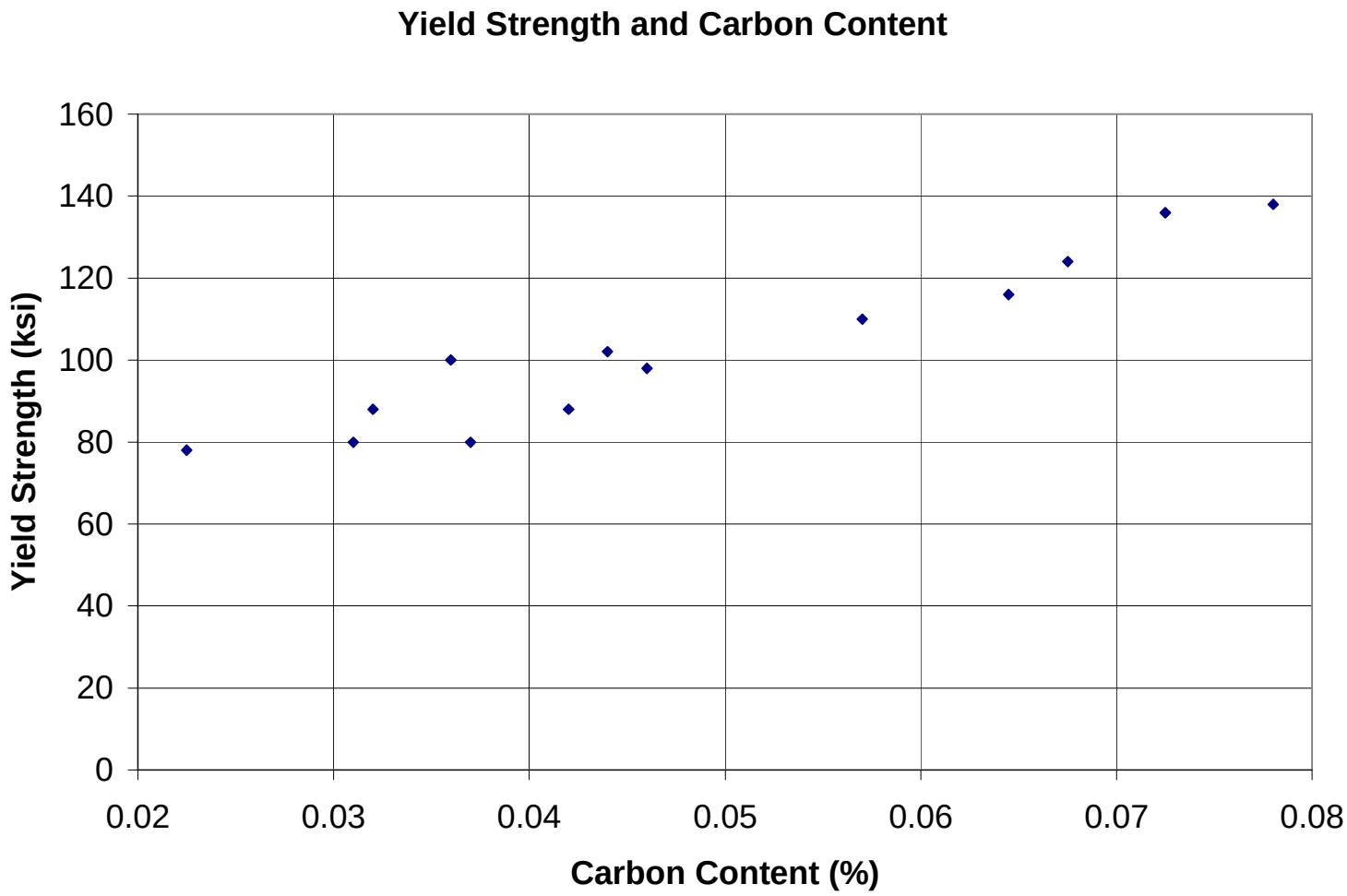
Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Scatter Diagrams

- The data are presented with scatter plots when both the independent (i.e., *x*) and dependent (i.e., *y*) variables are measured on *interval* or *ratio* scales.
- Usually, the predicted (depended) variable is shown on the ordinate and the independent variable on the abscissa.

Scatter Diagrams

Example: Scatter Plot



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Line Graphs

- Line graphs are used to present mathematical expressions.
- The independent and dependent variables are measured on interval or ratio scale
- The predicted variable (dependent) variable is usually shown on the ordinate.

Line Graphs

- Example:

  For one section of the state of Maryland, peak discharge rates Q (ft^3/sec) can be estimated as a function of drainage area A (mi^2) by the following equations:

$$Q_2 = 55.1A^{0.672}$$

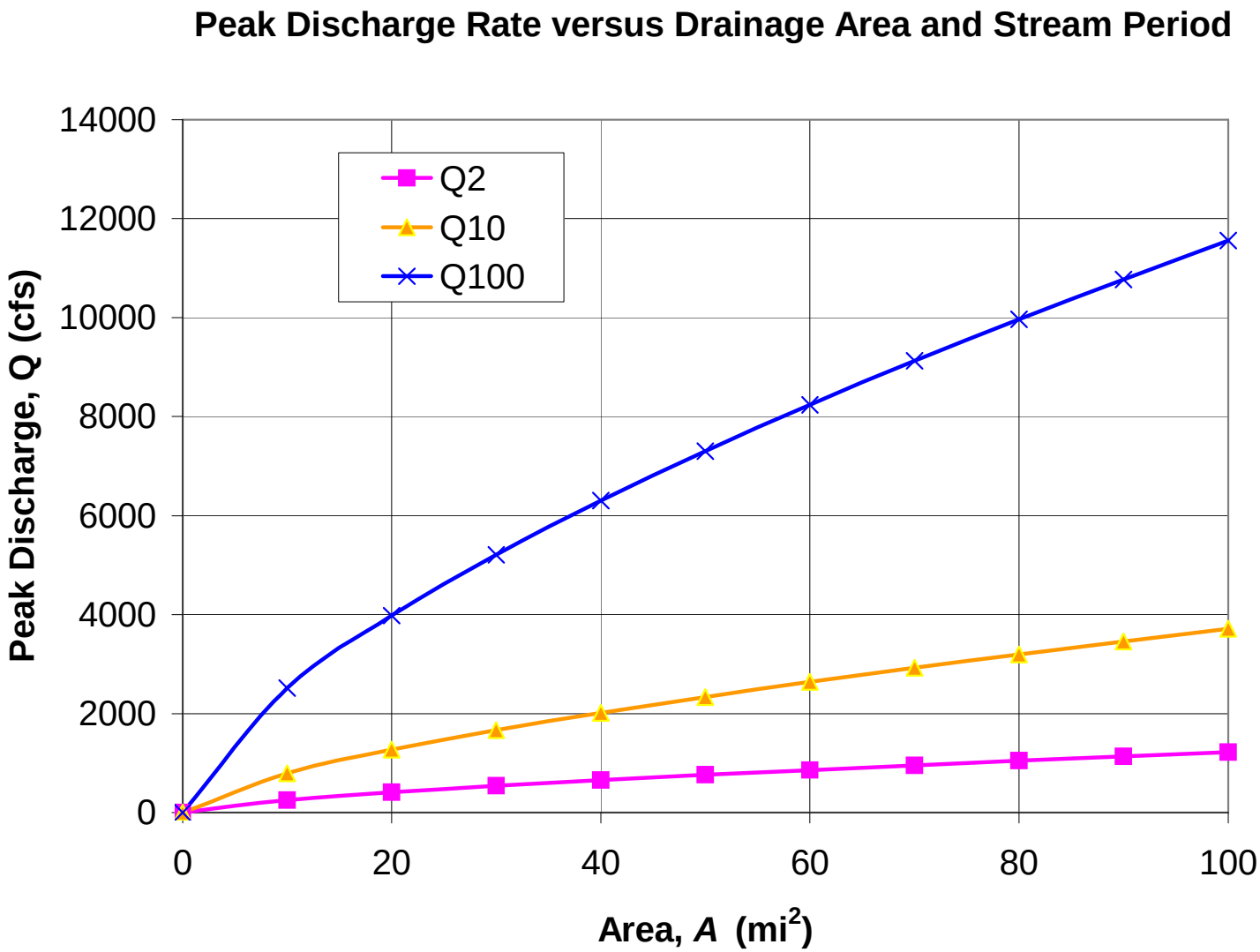
$$Q_{10} = 172A^{0.667}$$

$$Q_{100} = 548A^{0.662}$$

  Where Q_n = n^{th}-year peak discharge.

  Graphically, this can be presented as

Line Graphs



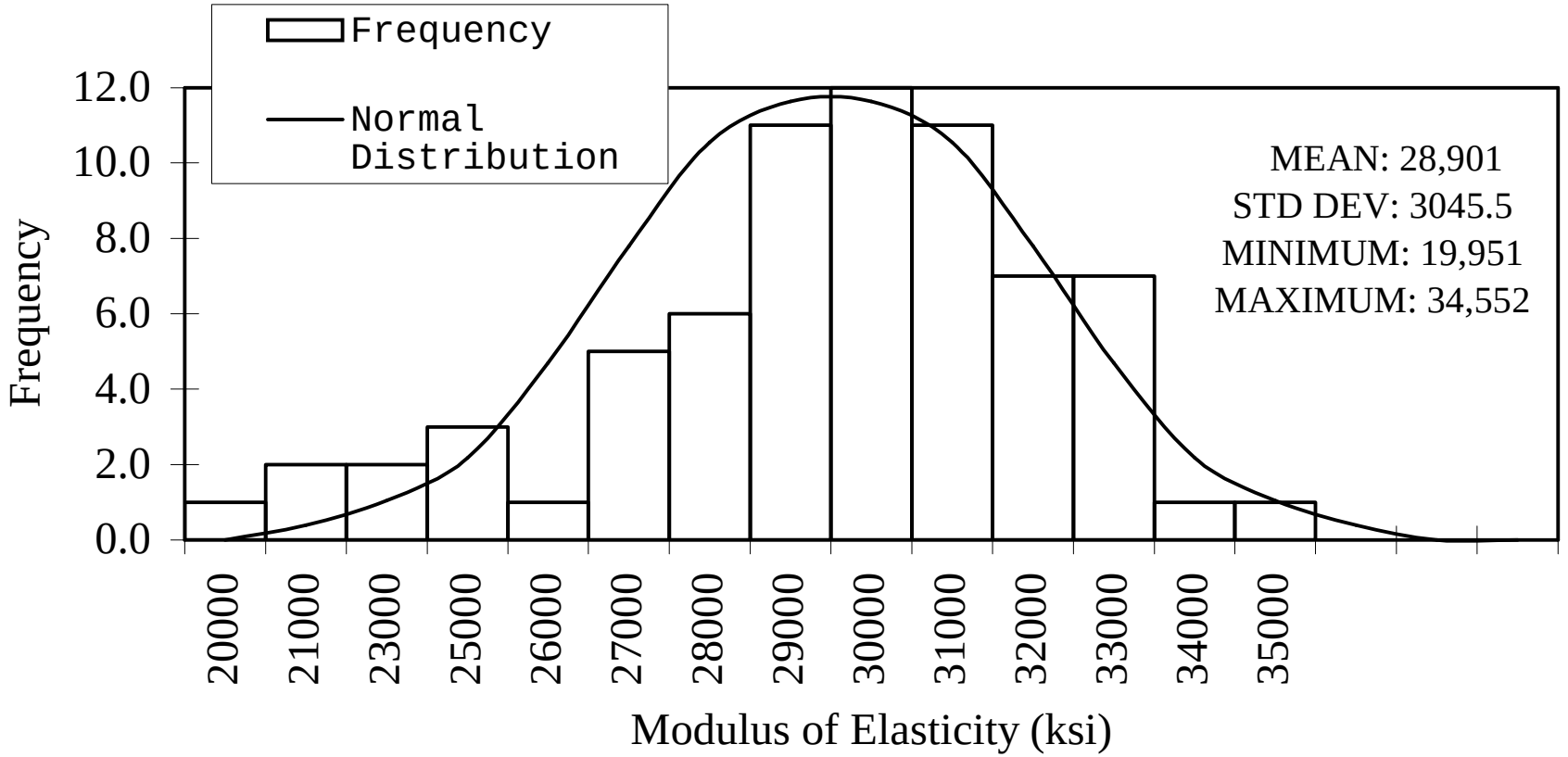
Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Combination Charts

- Combination charts can include two or more different types of charts to present data.
  - Example:
    - A line graph and a bar chart can be combined in the same plot.
    - A combination chart that includes a scatter plot and a line graph to present experimental data and theoretical (or fitted) prediction equation.

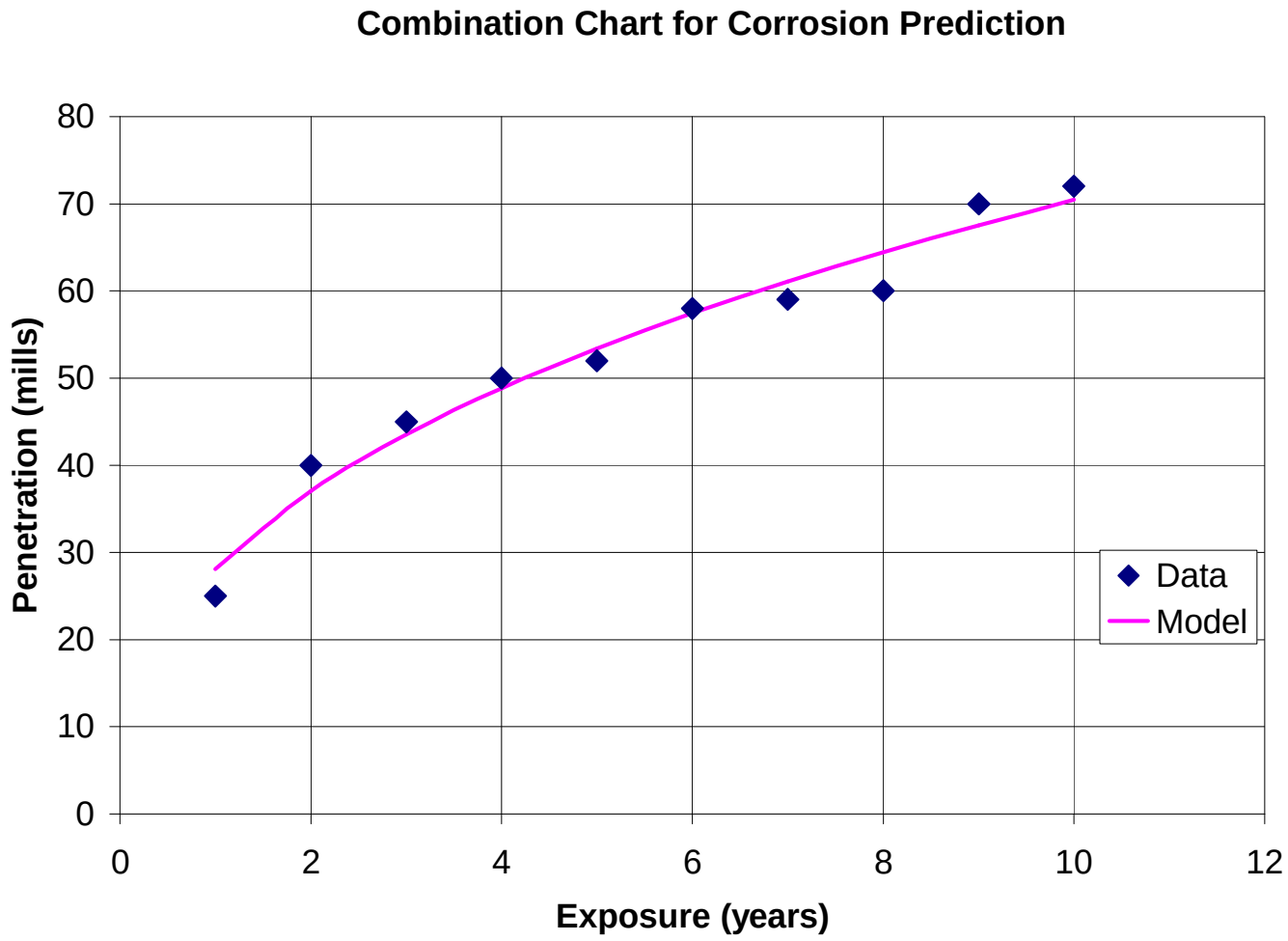
Combination Charts

Example: Experimental and Theoretical Distribution of Modulus of Elasticity of Steel



Combination Charts

Example:

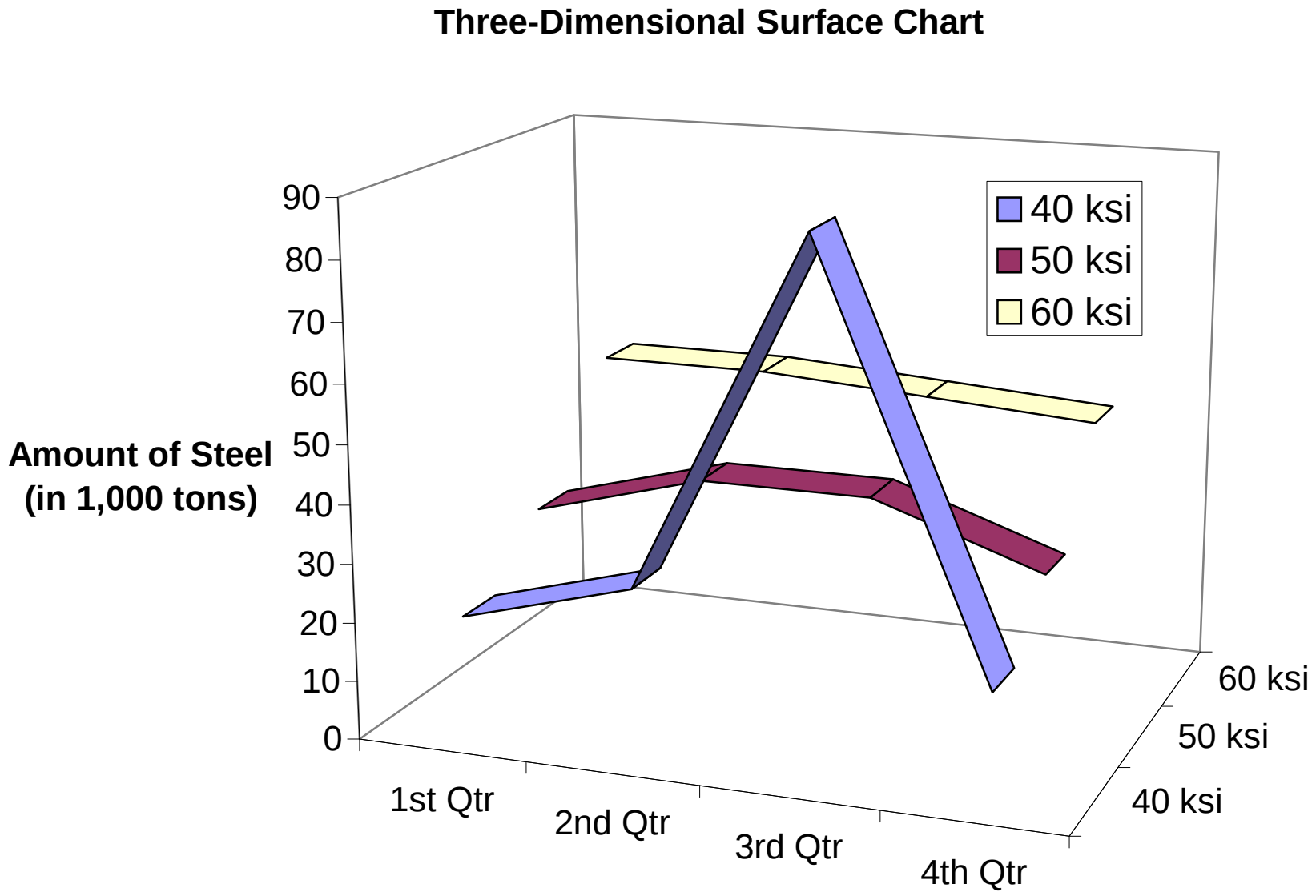


Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Three-Dimensional Charts

- Three dimensional charts are used to describe the relationships among three variables.
- Any types of charts described previously can be displayed in three dimension.

Three-Dimensional Charts



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Histograms and Frequency Diagrams

- Definitions:

  “A *histogram* is a plot (or tabulation) of the number of data points versus selected intervals or values for a parameter.”

  “ A *frequency diagram* (or frequency histogram) is a plot (or tabulation) of the frequency of occurrence versus selected intervals or values of the parameter.”

Histograms and Frequency Diagrams

- The number of intervals (k) can be subjectively selected depending on the sample size, n.
- The number of intervals can be approximately determined as

$$k = 1 + 3.3 \log_{10}(n)$$

Histograms and Frequency Diagrams

- Also, the number of interval can depend on the level of dispersion in the data.
- The frequency diagrams (or histograms) can be derived from the histograms by dividing the number of data points that correspond to each interval by the sample size.

Histograms and Frequency Diagrams

- Example:

  The starting salaries (in thousands of dollars) of 20 graduates, chosen at random from the graduating class of an urban university, were determined and recorded in the following table:

34	29	27	39	41
28	32	37	35	36
23	31	33	34	29
27	35	29	30	32

  Draw a histogram and frequency diagram of the graduate salaries.

Histograms and Frequency Diagrams

Sorted Data

23	32
27	33
27	34
28	34
29	35
29	35
29	36
30	37
31	39
32	41

Number of Data Points = 20,

Min = 23, Max = 41

Data Range = 41 – 23 =18

$$k = 1 + 3.3\log_{10}(n) = 1 + 3.3\log_{10}(20) = 5.29$$

Take $k = 5$

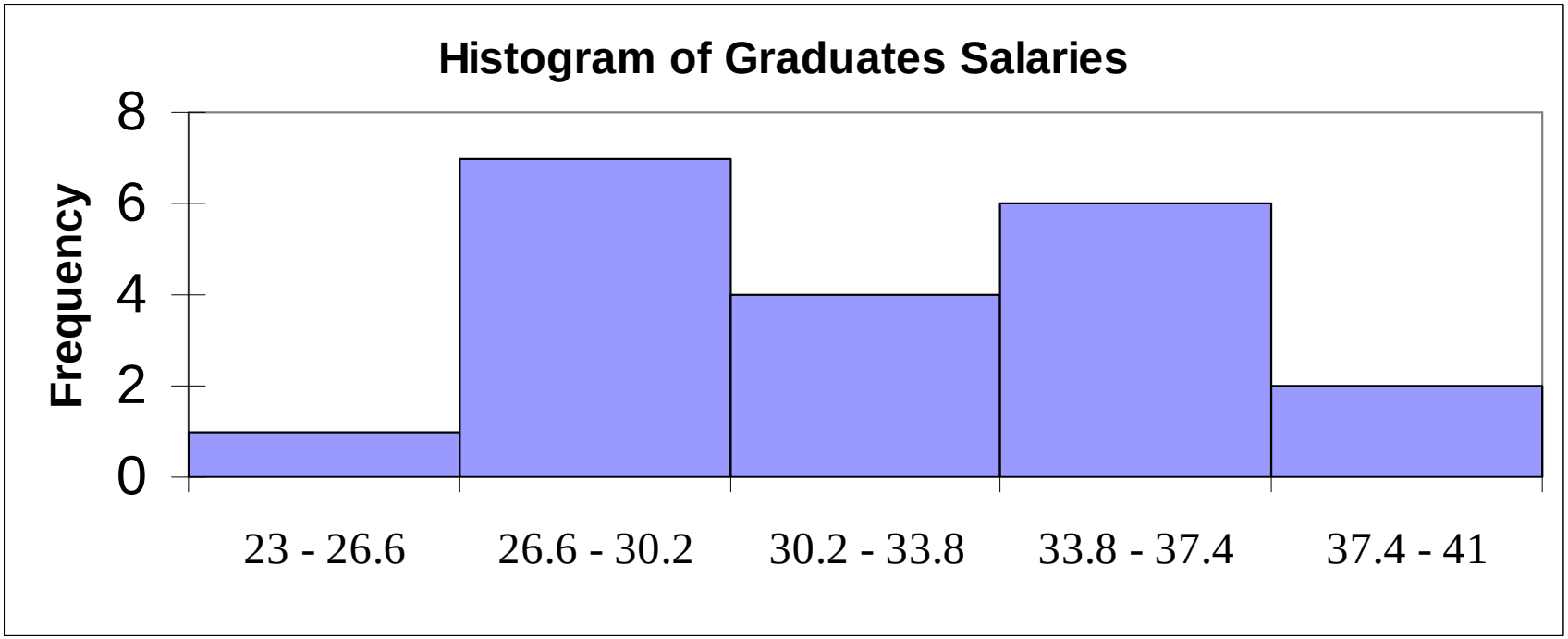
Hence, interval width = 18/5 = 3.6

The following histogram table and graphs can be constructed:

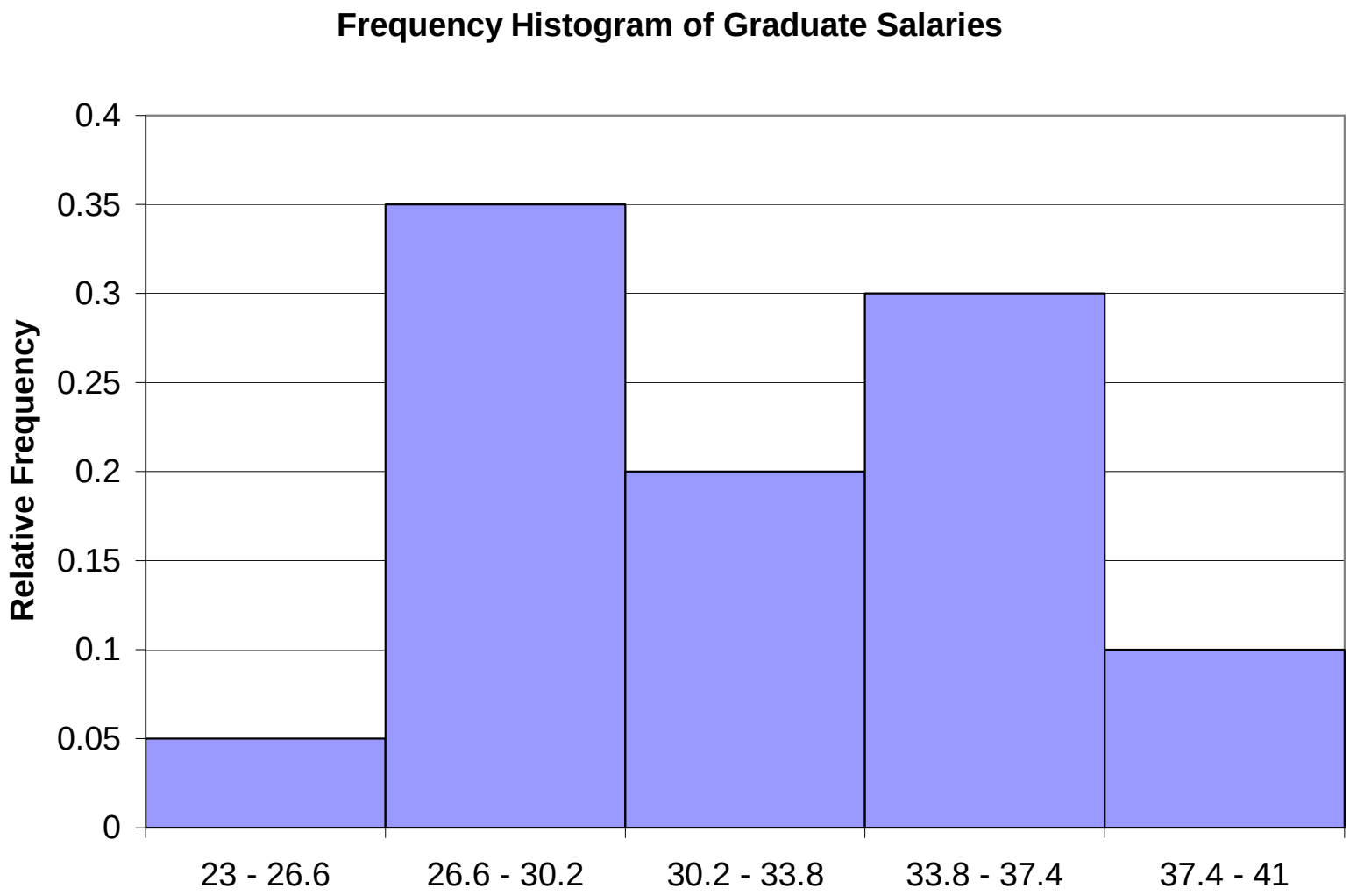
Histograms and Frequency Diagrams

Interval	**Frequency**	**Relative Frequency**
23.0 - 26.6	1	0.05
26.6 - 30.2	7	0.35
30.2 - 33.8	4	0.2
33.8 - 37.4	6	0.3
37.4 - 41.0	2	0.1
Total =	20	1

Histograms and Frequency Diagrams



Histograms and Frequency Diagrams



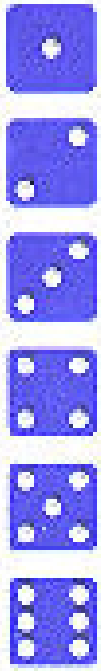
Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Histograms and Frequency Diagrams

Example: Rolling of a Pair of Dice

SECOND DIE

FIRST DIE



(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

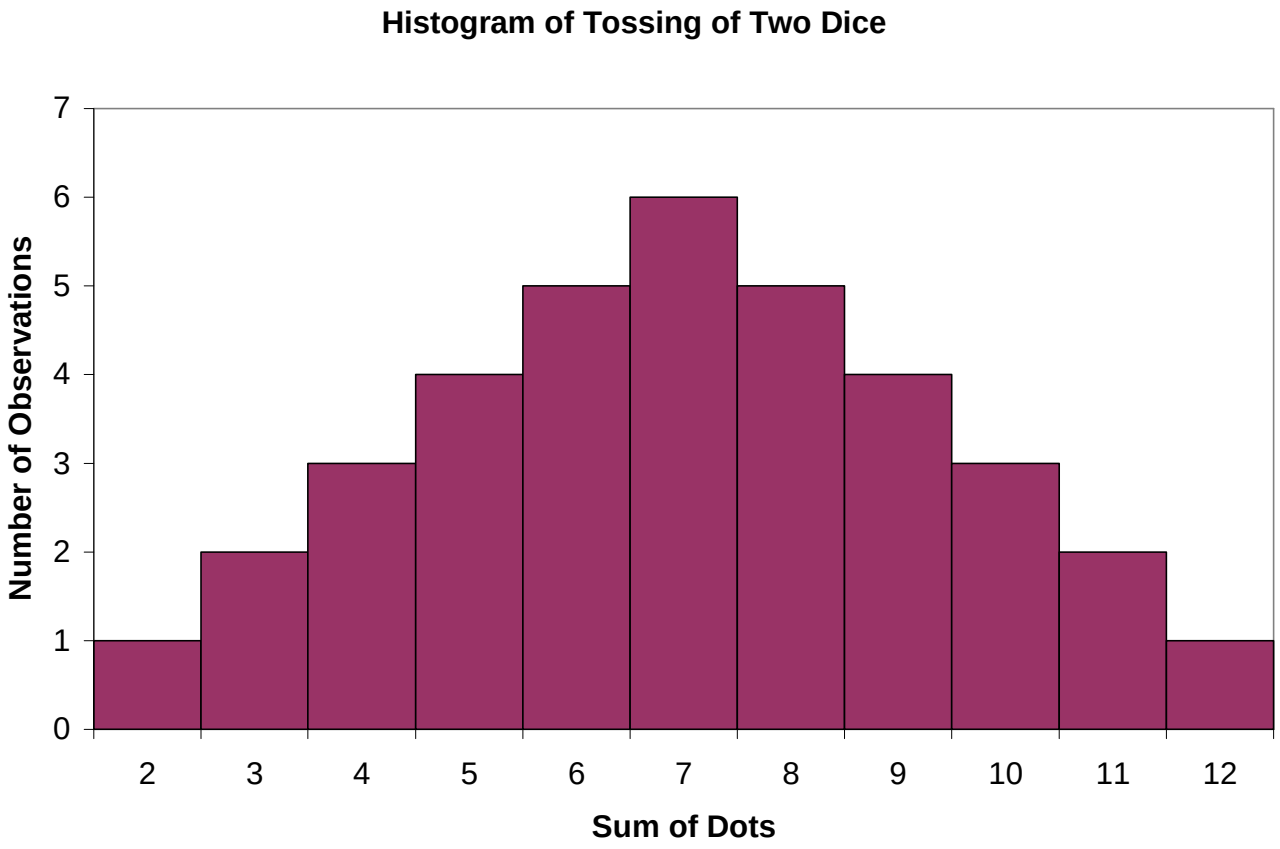
Histograms and Frequency Diagrams

- Example: Rolling of a Pair of Dice

  If a pair of dice rolled simultaneously, the histograms for the sum of dots from the two dice would appear as shown in the following tables and figures:

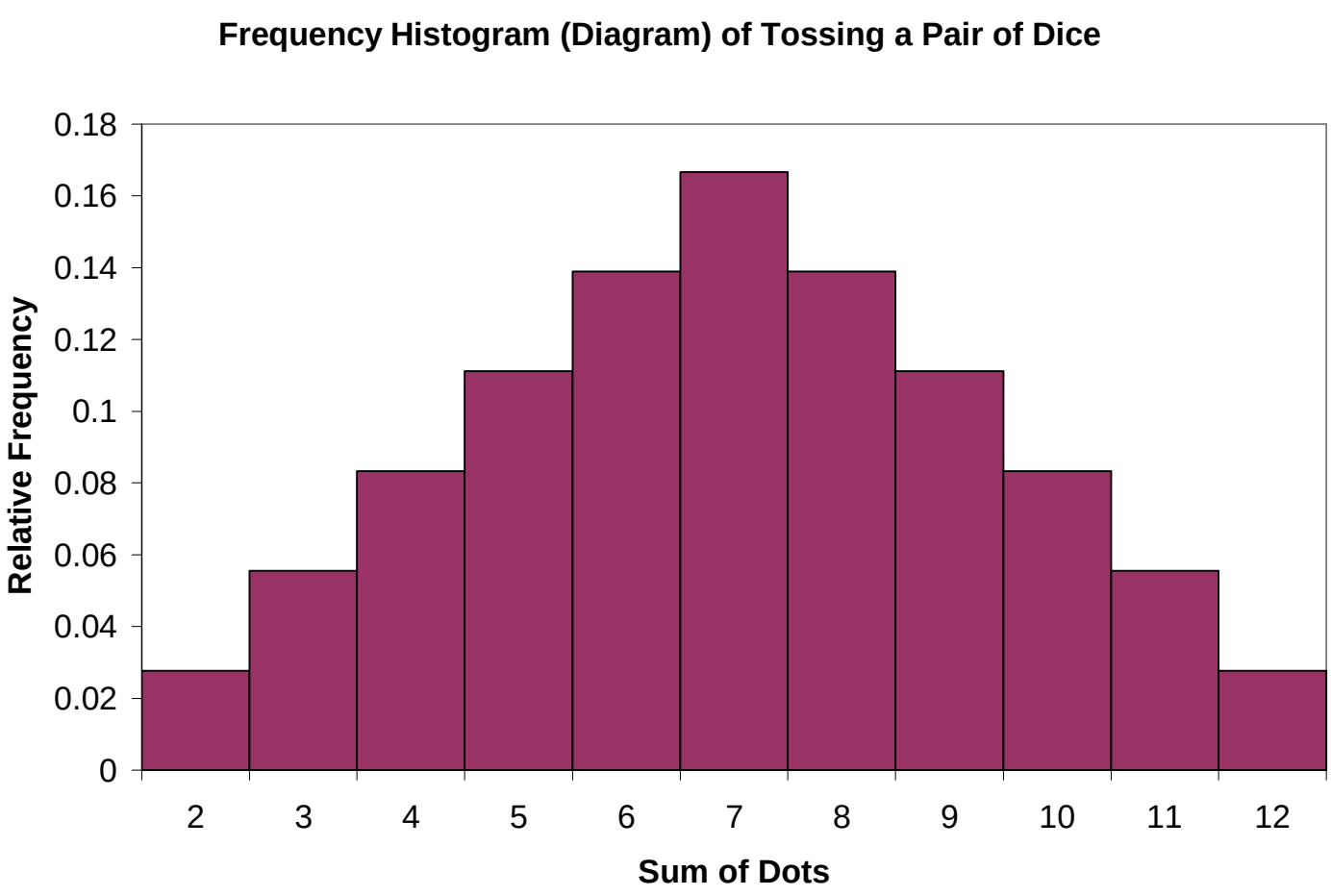
Histograms and Frequency Diagrams

Sum of Dots	No. of Observations
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1



Histograms and Frequency Diagrams

Sum of Dots	Relative Frequency
2	0.028
3	0.056
4	0.083
5	0.111
6	0.139
7	0.167
8	0.139
9	0.111
10	0.083
11	0.056
12	0.028



Histograms and Frequency Diagrams

- Example: Tossing of 3 Coins

  Suppose that we are interested in the number of heads (0, 1, 2, or 3) appearing on each toss of the three coins, then the outcome would be the following:

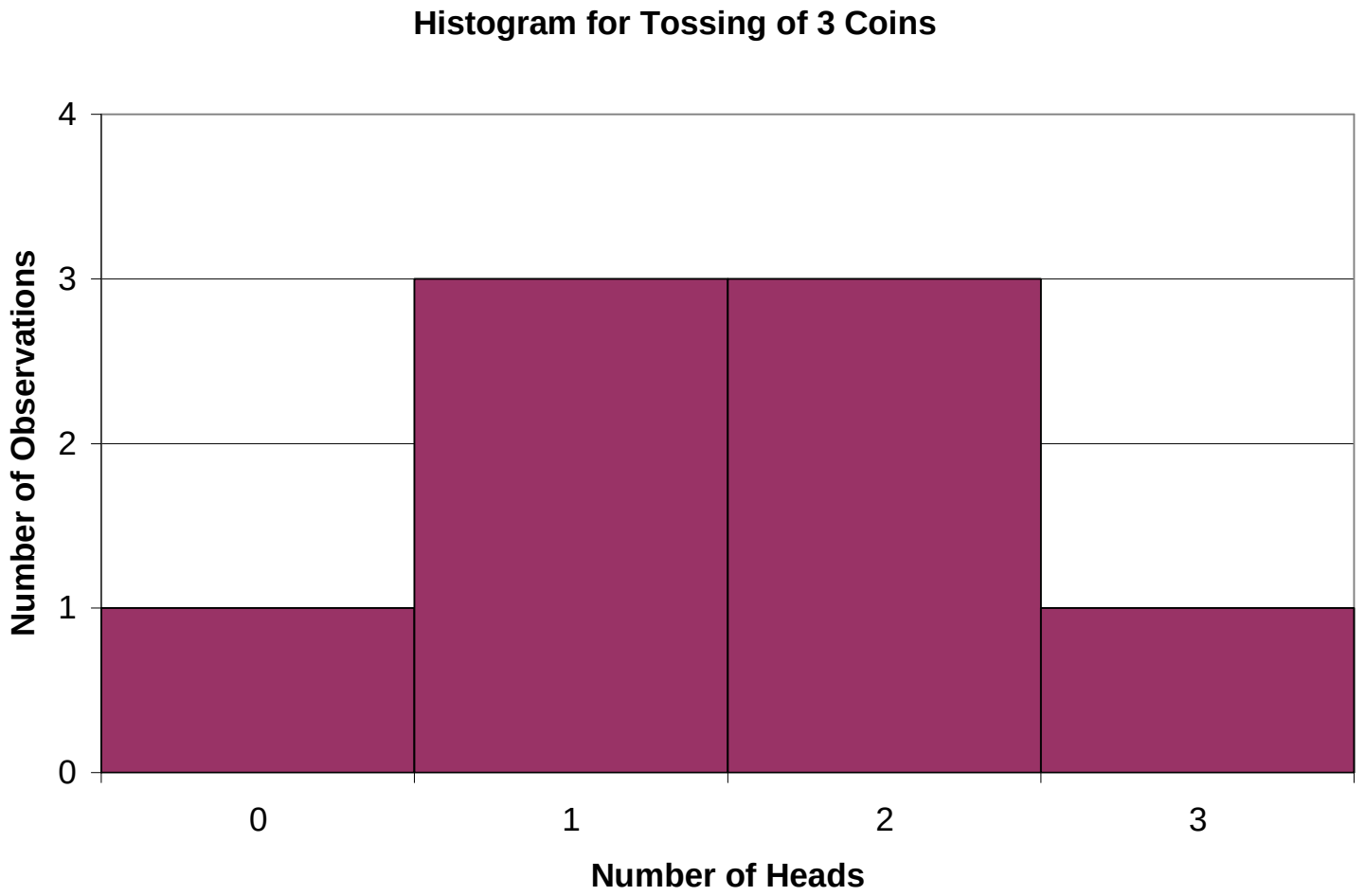
Histograms and Frequency Diagrams

Example: Tossing of 3 Coins

Outcome		Number of Heads	Frequency
TTT		0	1
(TTH), (THT), and (HTT)		1	3
(THH), (HTH), and (HHT)		2	3
(HHH)		3	1

Histograms and Frequency Diagrams

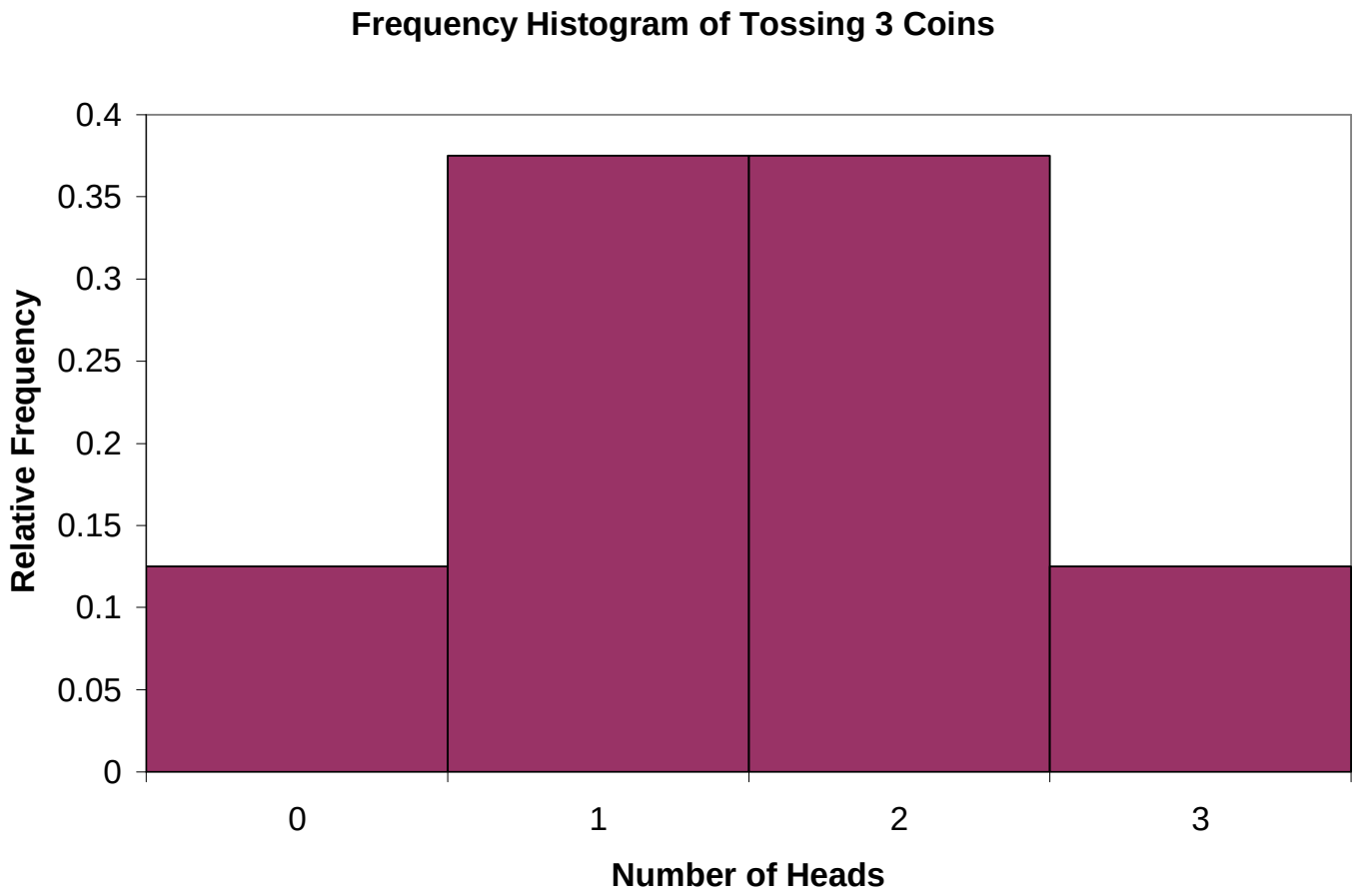
Example: Tossing of 3 Coins



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Histograms and Frequency Diagrams

Example:

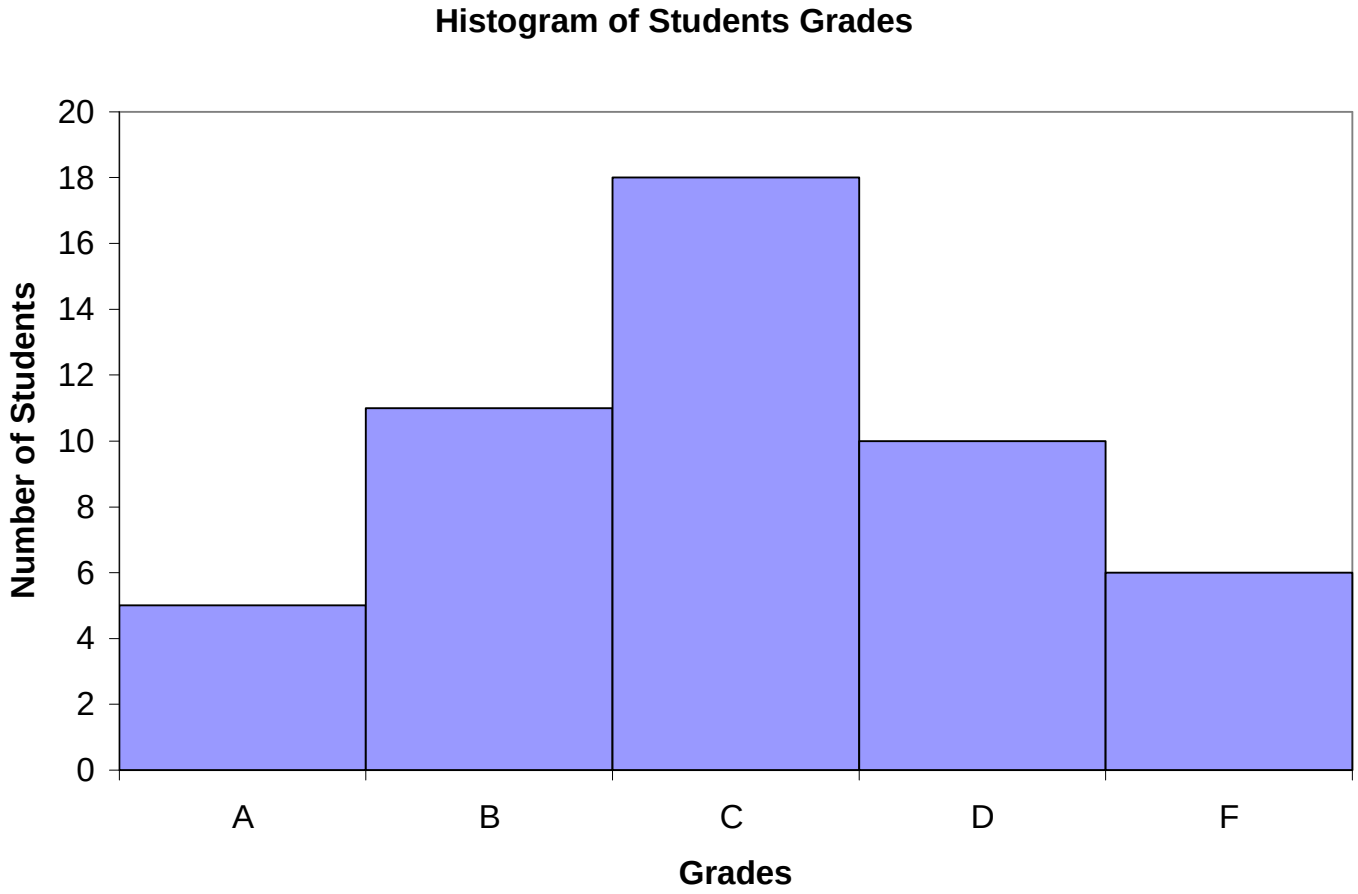


Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Histograms and Frequency Diagrams

Example: Grades of Students (a total of 50 students)

Grades	Number of Students
A	5
B	11
C	18
D	10
F	6

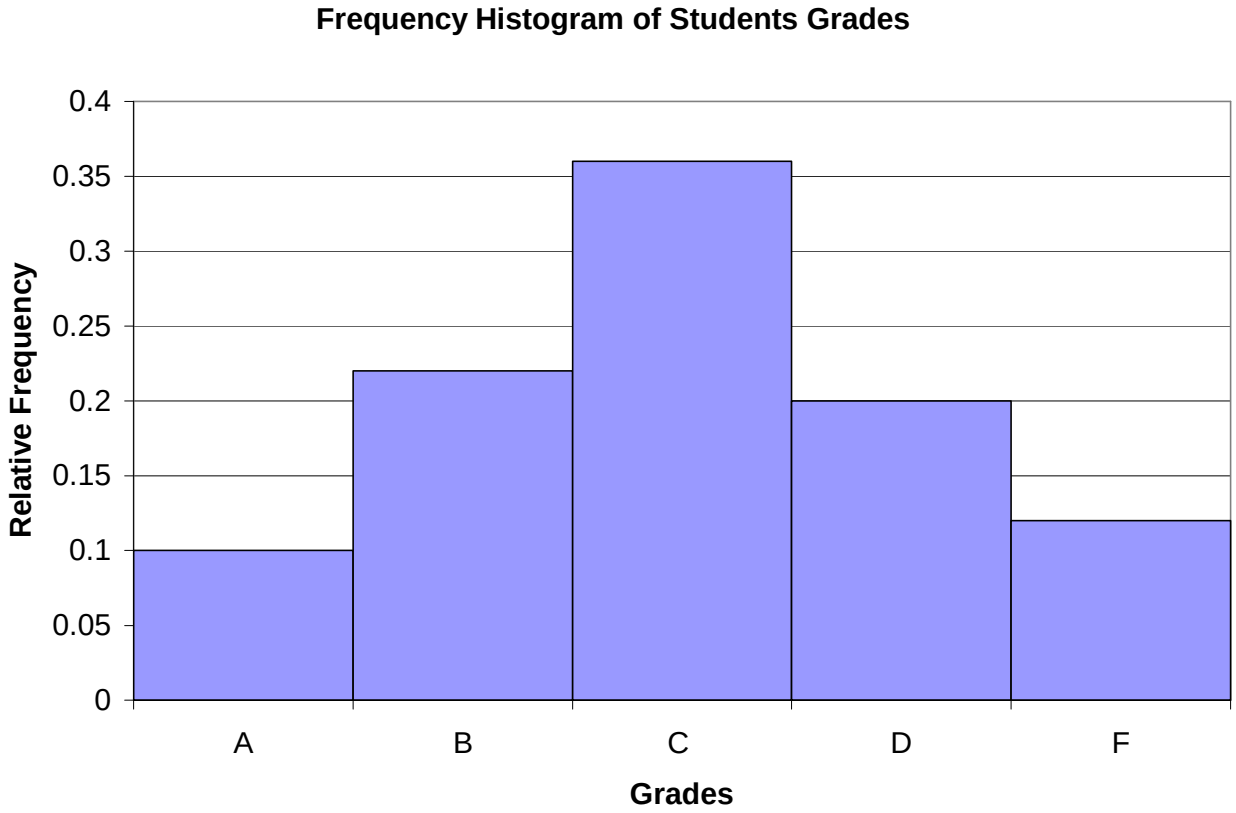


Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Histograms and Frequency Diagrams

Example: Grades of Students (a total of 50 students)

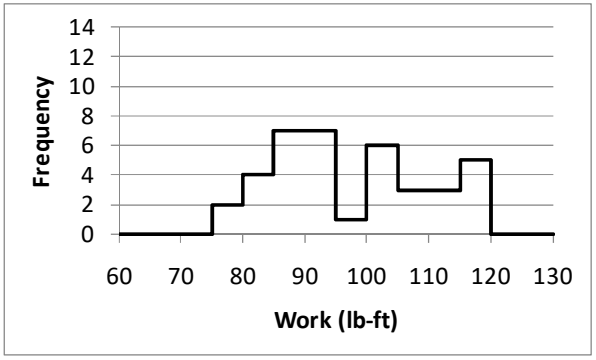
Grades	Relative Frequency
A	5/50
B	11/50
C	18/50
D	10/50
F	6/50



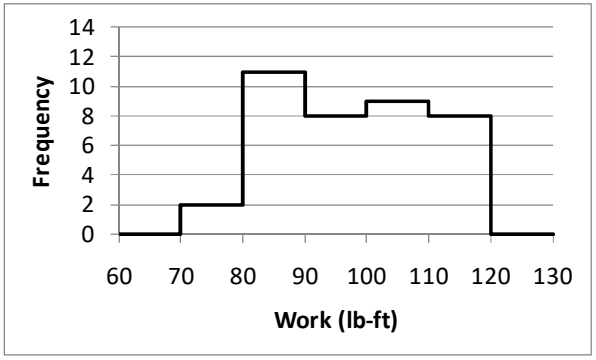
Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Histograms: Bin Sizes

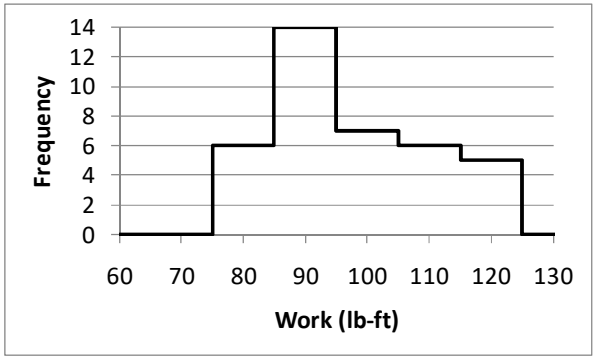
Data: 76, 78, 81, 82, 84, 84, 86, 86, 87, 88, 88, 88, 89, 91, 91, 92, 92, 92, 94, 94, 98, 101, 103, 103, 103, 104, 104, 106, 108, 109, 112, 113, 114, 116, 116, 118, 118, 119



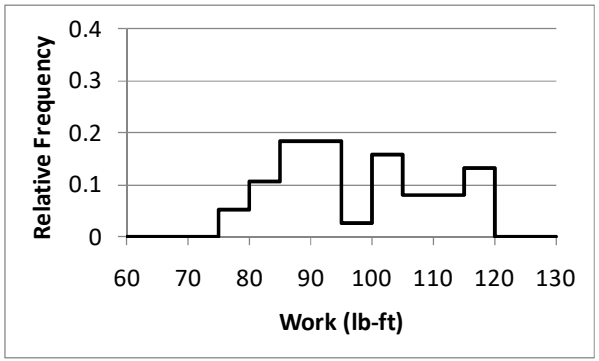
(a) Cell width 5 lb-ft



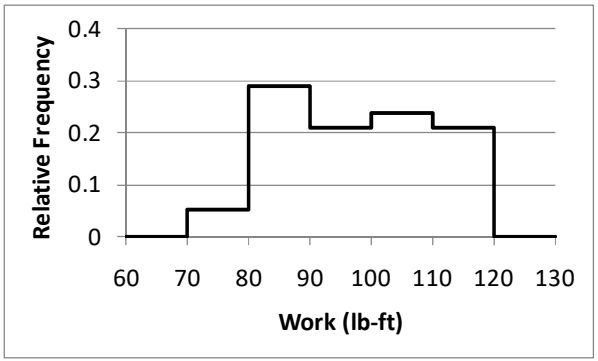
(b) Cell width 10 lb-ft



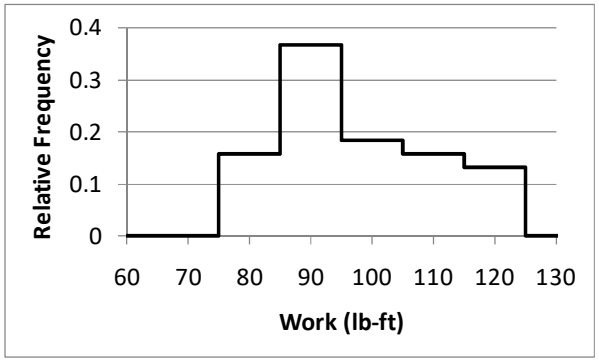
(c) Cell width 10 lb-ft



(a) Cell width 5 lb-ft



(b) Cell width 10 lb-ft



(c) Cell width 10 lb-ft

Descriptive Measures

- There are three fundamental types of measures for data analysis:
  1. Central Tendency Measures
     a. Average Value
     b. Median Value
     c. Mode Value
  2. Dispersion Measures
  3. Percentile Measures

Central Tendency Measures

- These measures are very important descriptors of data.
- The following three types can be used:
  1. Average (Mean) Value
  2. Median Value
  3. Mode Value

Central Tendency Measures

- Average (Mean) Value
  - For *n* observations, if all observations are given equal weights, the average value can be given by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where x_i = a sample point, and i =1, 2, …,n

Central Tendency Measures

- Example: Average or Mean Value

  Find the average (mean) for the sample measurements 3, 5, 1, 8, 6, 5, 4, and 6.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(3+5+1+8+6+5+4+6) = \frac{38}{8} = 4.75$$

Central Tendency Measures

- Median Value
  - The median value x_m is defined as the point that divides the data into two equal parts.
  - 50% of data are above x_m and 50% are below x_m.
  - The median value can be determined by ranking the n values in the sample in increasing or decreasing order.

Central Tendency Measures

- Steps for Computing the Median, x_m
  1. If n is an odd number, the median is the value with a rank of $(n + 1)/2$.
  2. If n is an even number, the median equals the average of the two middle values, that is, those with ranks $n/2$ and $(n/2) + 1$.

Central Tendency Measures

- Example: Median Value

Find the the median for the following sets of measurements:

Set 1: 5, 21, 8, 7, and 13

Set 2: 10, 9, 23, 15, 20, and 34

Sorted Data

Set 1	Set 2
5	9
7	10
8	15
13	20
21	23
-	24
$n =$ 5	6

$$\text{Median}_{\text{Set 1}} = 8$$

$$\text{Median}_{\text{Set 2}} = \frac{15+20}{2} = 17.5$$

Central Tendency Measures

- Mode Value
  - The mode value x_d is defined as the point of highest percent for the frequency of occurrence.
  - This point can be determined with the aid of Histogram or frequency histogram (diagram)

Central Tendency Measures

- Example 1: Mode Value

Find the mode for the following set of data:

2, 1, 2, 1, 1, 5, 1, 9, and 4

The data can be arranged in ascending order as follows:

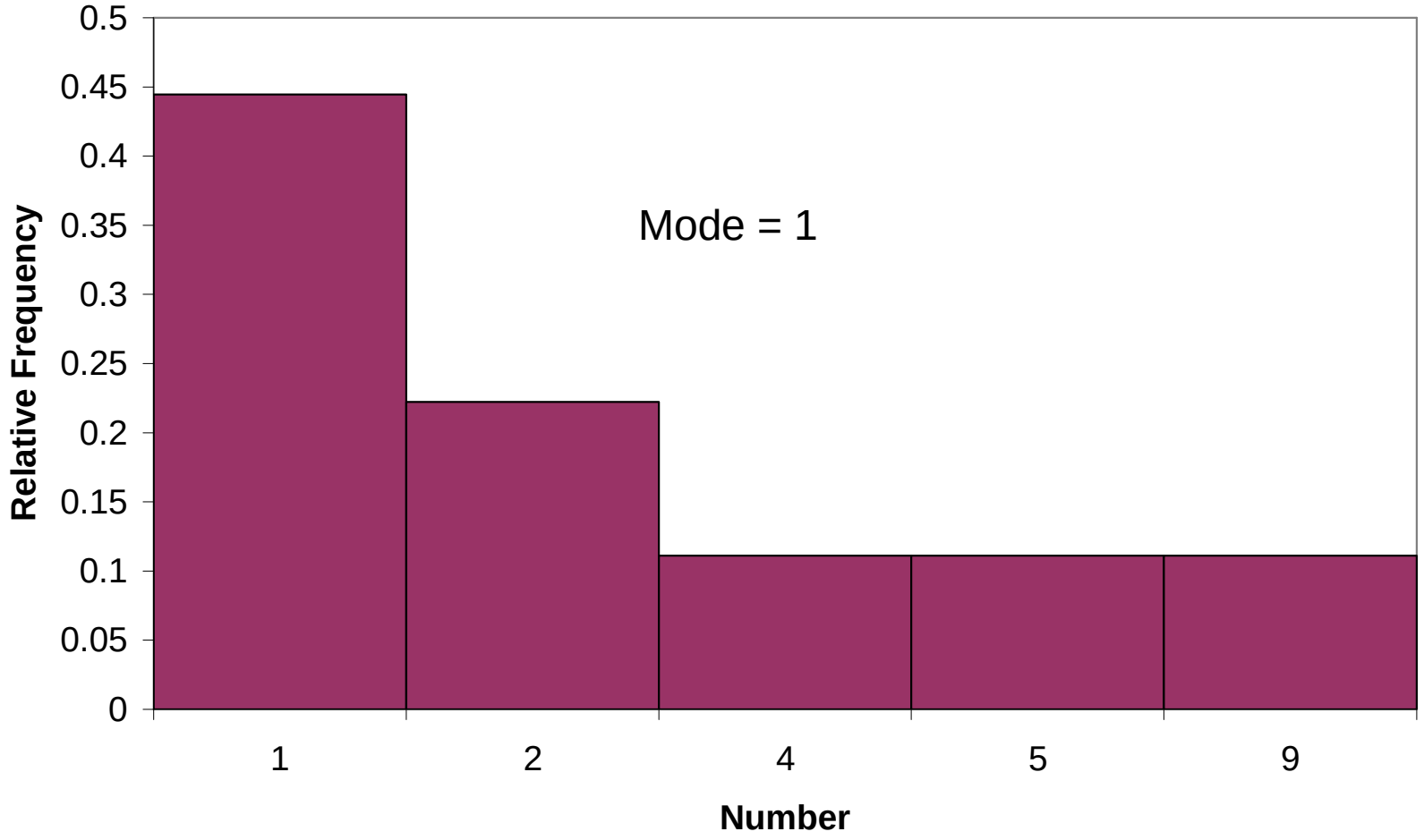
1, 1, 1, 1, 2, 2, 4, 5, 9

Hence, the mode = 1

Number	Frequency	Relative Frequency
1	4	0.44
2	2	0.22
4	1	0.11
5	1	0.11
9	1	0.11
Total =	9	1

Central Tendency Measures

Example 1: Mode Value

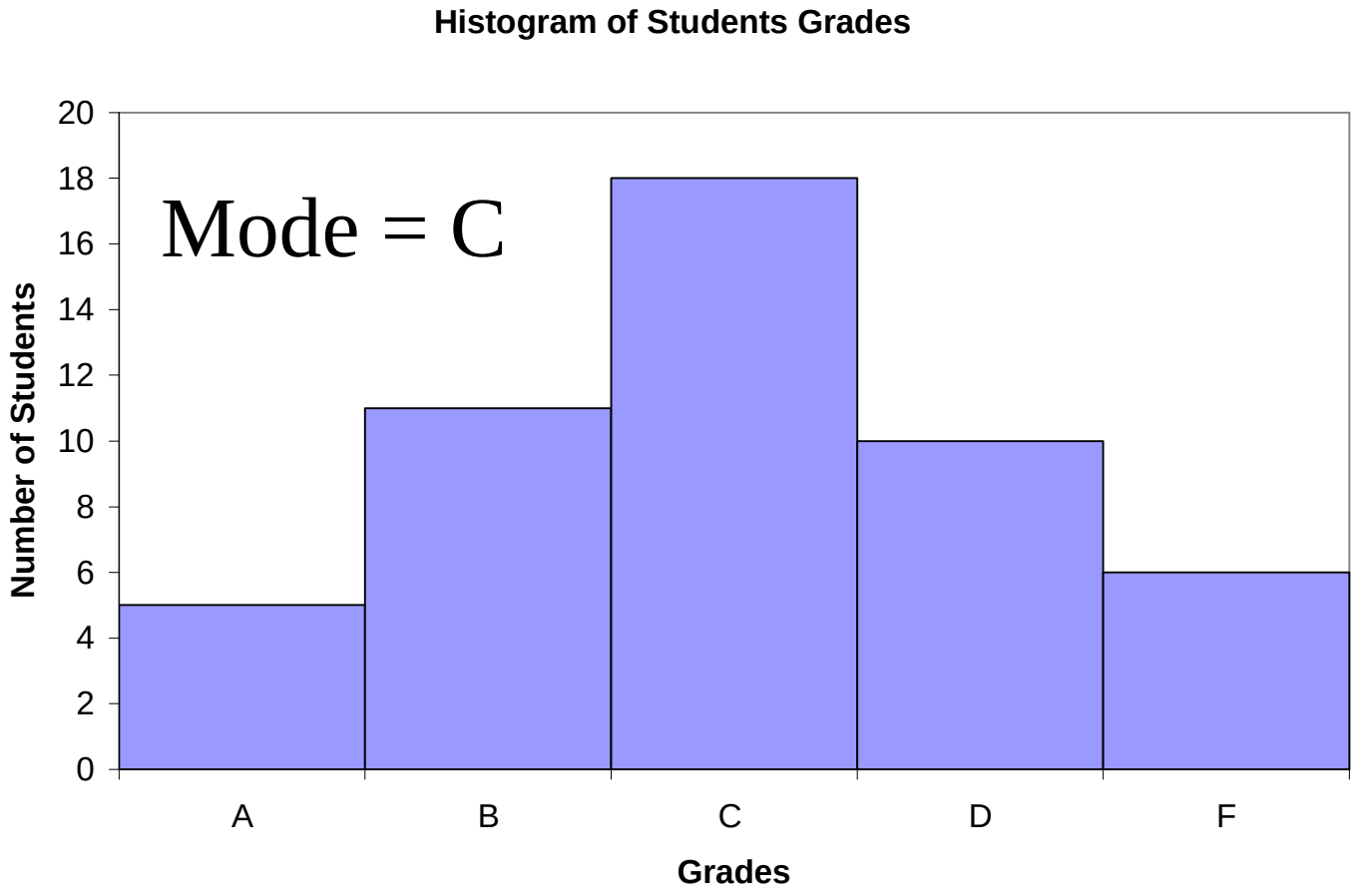


Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Central Tendency Measures

- Example 2: Mode Value

Grades	Number of Students
A	5
B	11
C	18
D	10
F	6



Dispersion Measures

- A measure of central tendency gives us a typical value that can be used to describe a whole set of data, but it does not tell us whether the data are tightly clustered or widely dispersed.
- Therefore, the dispersion measures describe the level in the data about the central tendency location

Dispersion Measures

- The dispersion measures include:
  - Range
  - Variance, S^2
  - Standard Deviation, S
  - Coefficient of Variation, COV or δ
  - Percentiles
  - Box-and-Whisker Plots

Dispersion Measures

- Range:

"The range of a set of data is the difference between the largest and smallest (extreme values) values in the data set."

Example:

2.3, 1.2, 4.6, 10.4, 8.0

Therefore:

Range = Max – Min = 10.4 – 1.2 = 9.2

Dispersion Measures

- Variance, S^2
  - The sample variance of a set of n sample measurements $x_1, x_2, \ldots, x_n$ with mean $\overline{X}$ is given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2 \qquad (2\text{-}1)$$

NOTE: the factor $(n - 1)$ is used instead of n to obtain an unbiased estimate of S^2

Dispersion Measures

- Example: Variance

  Find the variance for the following sample measurements:

  1, 3, 5, 4, and 3

$$n = 5$$

$$\overline{X} = \frac{1+3+5+4+3}{5} = 3.2$$

$$S^2 = \frac{(1-3.2)^2 + (3-3.2)^2 + (5-3.2)^2 + (4-3.2)^2 + (3-3.2)^2}{5-1} = 2.20$$

Dispersion Measures

– For computational purposes, the following alternative set of equations can be used to compute the variance:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

OR

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\overline{X}^2\right] \qquad (2\text{-}2)$$

Dispersion Measures

Example: Find the variance for the sample measurements 1, 3, 5, 4, and 3 using Eq. 2-2.

$$S^2 = \frac{1}{5-1}\left[1^2+3^2+5^2+4^2+3^2-\frac{1}{5}(1+3+5+4+3)^2\right] = \frac{8.8}{4} = 2.2$$

OR

$$S^2 = \frac{1}{5-1}\left[1^2+3^2+5^2+4^2+3^2-5(3.2)^2\right] = \frac{8.8}{4} = 2.2$$

Dispersion Measures

- Standard Deviation, *S*

  The standard deviation by definition is the square root of the variance. It is given by

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]}$$

OR

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\overline{X}^2\right]} \qquad (2\text{-}3)$$

Dispersion Measures

- Coefficient of Variation, *COV* or δ

  The coefficient of variation is a normalized quantity based on the standard deviation and the mean. It is a dimensionless quantity. The *COV* is defined as

$$COV = \frac{\text{standard deviation}}{\text{mean (or average)}} = \frac{S}{\overline{X}} \qquad (2\text{-}4)$$

Dispersion Measures

- Example: Dispersion Measures of Concrete Strength

  A sample of five tests was taken to determine the compression strength (ksi) of concrete. Tests results are 2.5, 3.5, 2.2, 3.2, and 2.9 ksi. Compute the variance, standard deviation, and coefficient of variation of concrete strength.

Dispersion Measures

- Example (cont'd): Concrete Strength

$$\overline{X} = \frac{2.5 + 3.5 + 2.2 + 3.2 + 2.9}{5} = 2.86 \text{ ksi}$$

$$S^2 = \frac{2.5^2 + 3.5^2 + 2.2^2 + 3.2^2 + 2.9^2 - \frac{(2.5 + 3.5 + 2.2 + 3.2 + 2.9)^2}{5}}{5 - 1} = 0.273 \text{ ksi}^2$$

$$S = \sqrt{0.273} = 0.5225 \text{ ksi}$$

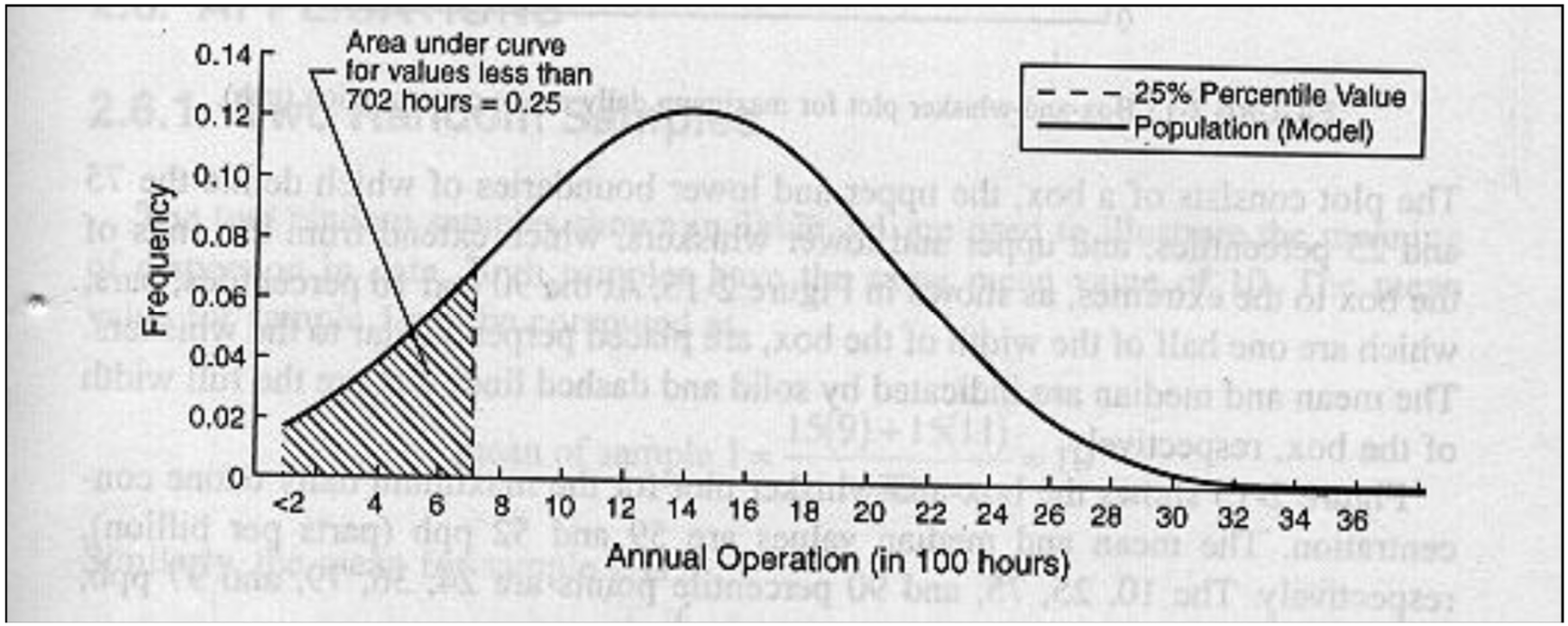
$$\delta \text{ or } COV(X) = \frac{S}{\overline{X}} = \frac{0.5225}{2.86} = 0.183$$

Dispersion Measures

- Percentiles
  - A p percentile value (x_p) for a variable based on a sample is the value of the variable such that p% of the data is less or equal to x_p.
  - On the basis of this definition, the median value is considered to be the 50 percentile value.
  - It is common in engineering to have interest in the 10, 25, 50, 75, and 90 percentile values.

Dispersion Measures

- Example: Operation of a Marine Vessel



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Dispersion Measures

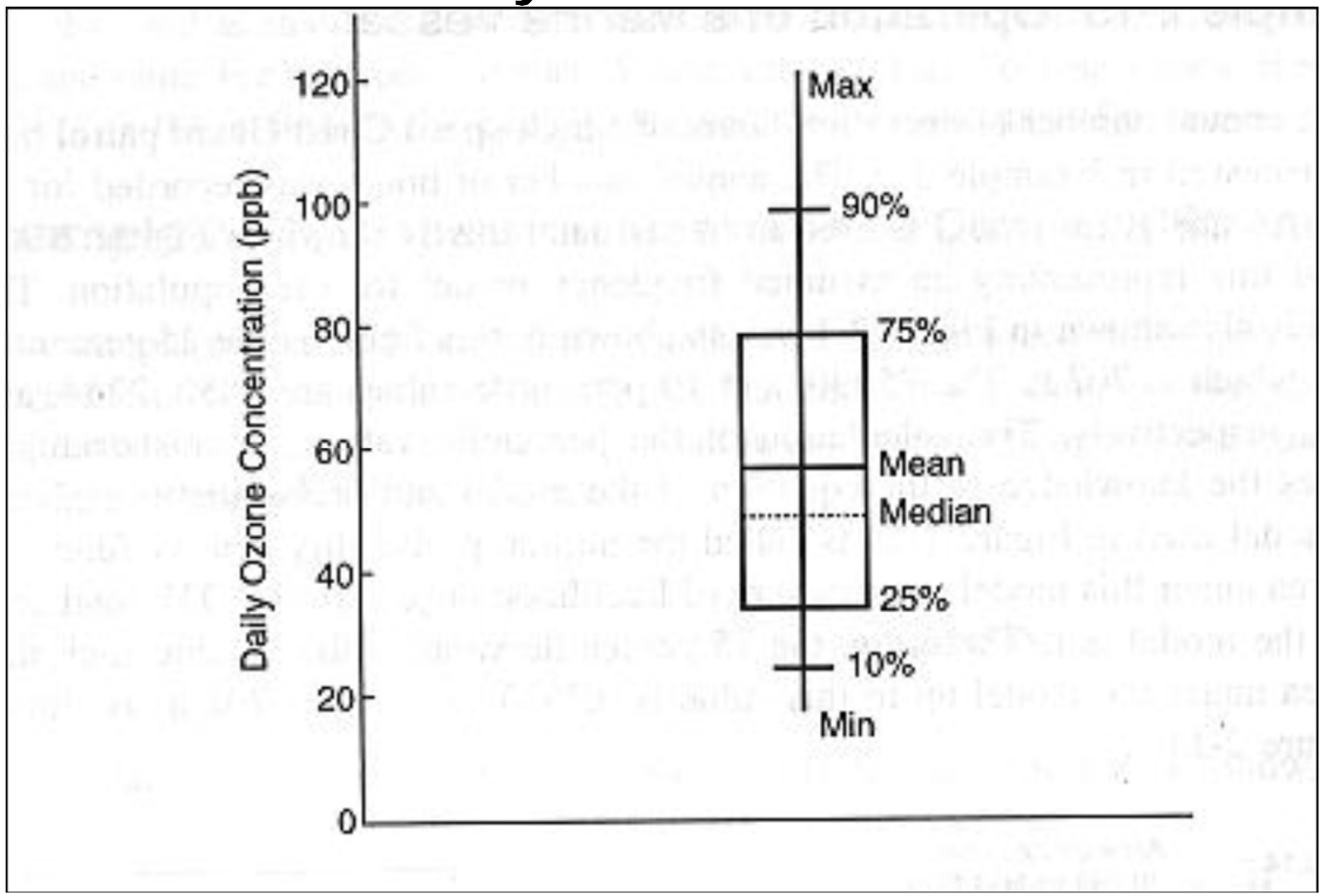
- Box-and-Whisker Plots
  - They are graphical methods for showing the distribution of sampled data, including:
    - Central tendency (mean and median)
    - Dispersion (variance, standard deviation, *COV*)
    - Percentiles (10, 25, 75, and 90 percentiles)
    - Extremes (minimum and maximum)

Dispersion Measures

- To construct a box-and-whisker plot, the following characteristic of a data set need to be computed:
  1. Mean and median of the sample.
  2. Minimum and maximum of the sample.
  3. 90, 75, 25, and 10 percentile values.

Dispersion Measures

- Example 1: Box-and Whisker Plot for Maximum Daily Ozone Concentration



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Dispersion Measures

- Example 2: Box-and Whisker Plot for Maximum Daily Ozone Concentration

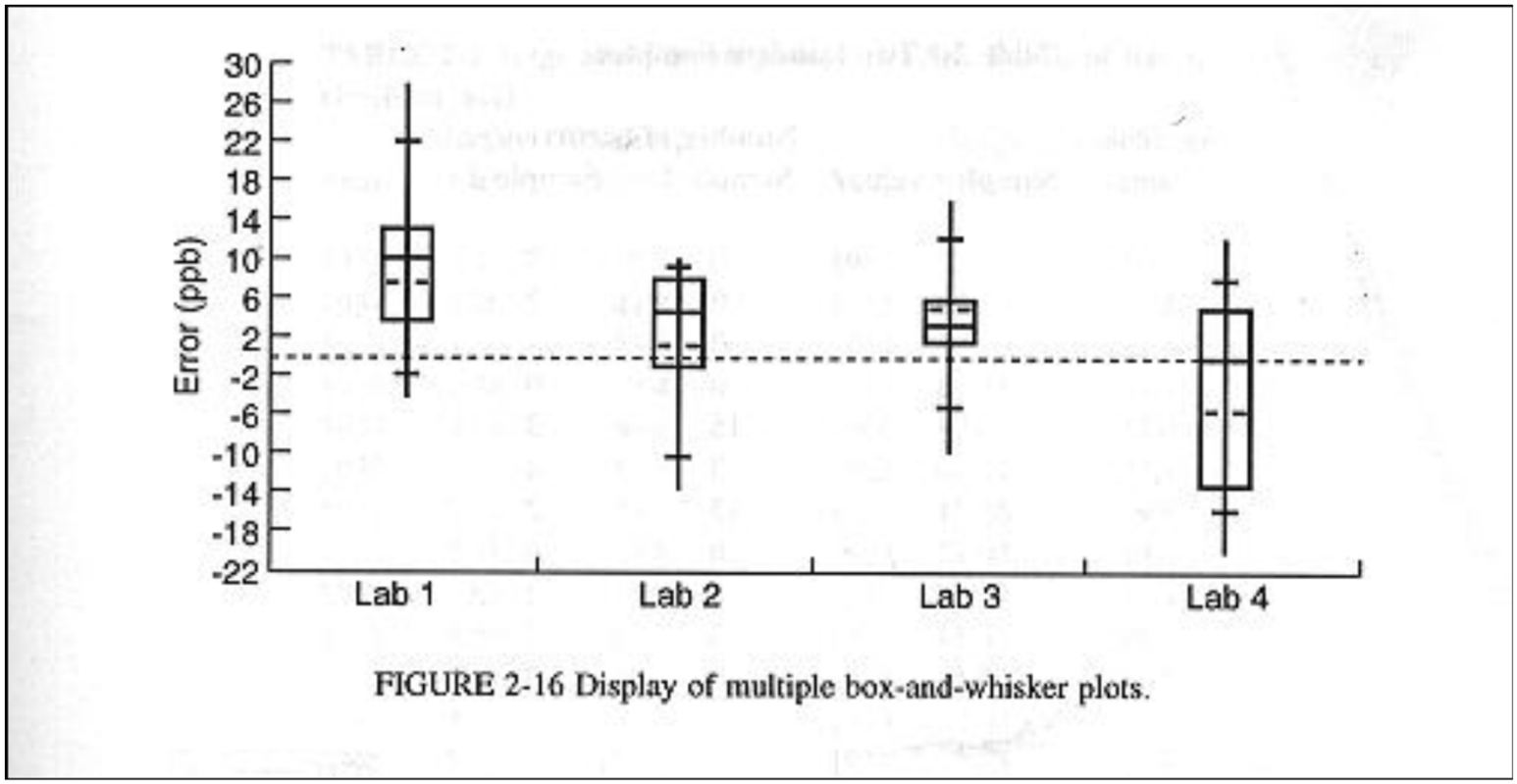


FIGURE 2-16 Display of multiple box-and-whisker plots.

Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Analysis of Simulated Data

- It is a common practice in engineering both to use *simulated data* in decision making and to compare *simulated* and *actual* data.
- Before simulated data are used in decision making, they should be analyzed in much the way as the measured data.

Analysis of Simulated Data

- When analyzing simulated data, descriptive measures (i.e., mean, COV, etc.) are computed, and graphical analysis are made.
- The descriptive measures should be compared with the descriptive measures and graphs of the actual data.
- These analyses can be used as an indication of the reasonableness of the simulation.

Analysis of Simulated Data

- Example: Discharge of a Little Patuxent River in Guilford, Maryland
  - For the Little Patuxent River near Guilford, MD, the discharge in cubic meters per second (cms) was obtained as shown in the following table. The table shows the discharge for the years 1933 to 1989. Using a seed of 8765, generate a sample of 57 discharges using the midsquare method. Compute the descriptive measures of the simulated discharges and compare them with that of the actual data. Also, compare the histograms and the frequency diagrams of the actual (measured) and simulated data.

Year	Discharge (cms)	Year	Discharge (cms)	Year	Discharge (cms)	Year	Discharge (cms)
1933	119.2	**1948**	43.9	**1963**	23.2	**1978**	103.9
1934	41.9	**1949**	19.9	**1964**	32.8	**1979**	132.5
1935	25.9	**1950**	15.9	**1965**	24.8	**1980**	31.4
1936	37.4	**1951**	55.8	**1966**	34.0	**1981**	28.0
1937	56.6	**1952**	150.1	**1967**	33.4	**1982**	22.0
1938	51.5	**1953**	56.6	**1968**	39.9	**1983**	78.7
1939	27.4	**1954**	18.3	**1969**	19.4	**1984**	42.2
1940	77.6	**1955**	107.3	**1970**	19.4	**1985**	40.5
1941	15.9	**1956**	28.6	**1971**	86.9	**1986**	19.3
1942	61.7	**1957**	17.3	**1972**	351.1	**1987**	34.8
1943	39.6	**1958**	34.8	**1973**	53.2	**1988**	34.8
1944	58.3	**1959**	21.4	**1974**	31.4	**1989**	143.5
1945	107.9	**1960**	28.3	**1975**	152.0		
1946	30.6	**1961**	27.4	**1976**	54.1		
1947	23.1	**1962**	36.2	**1977**	21.3		

Analysis of Simulated Data

Descriptive Statistics of Discharge for the Little Patuxent River, Guilford, MD (actual data)

Parameter	Discharge (cms)
Average (Mean)	54.82
Median	34.80
Mode	34.80
Standard Deviation	53.78
Sample Variance	2892.52
COV	0.98
Range	335.20
Minimum	15.90
Maximum	351.10
Count	57

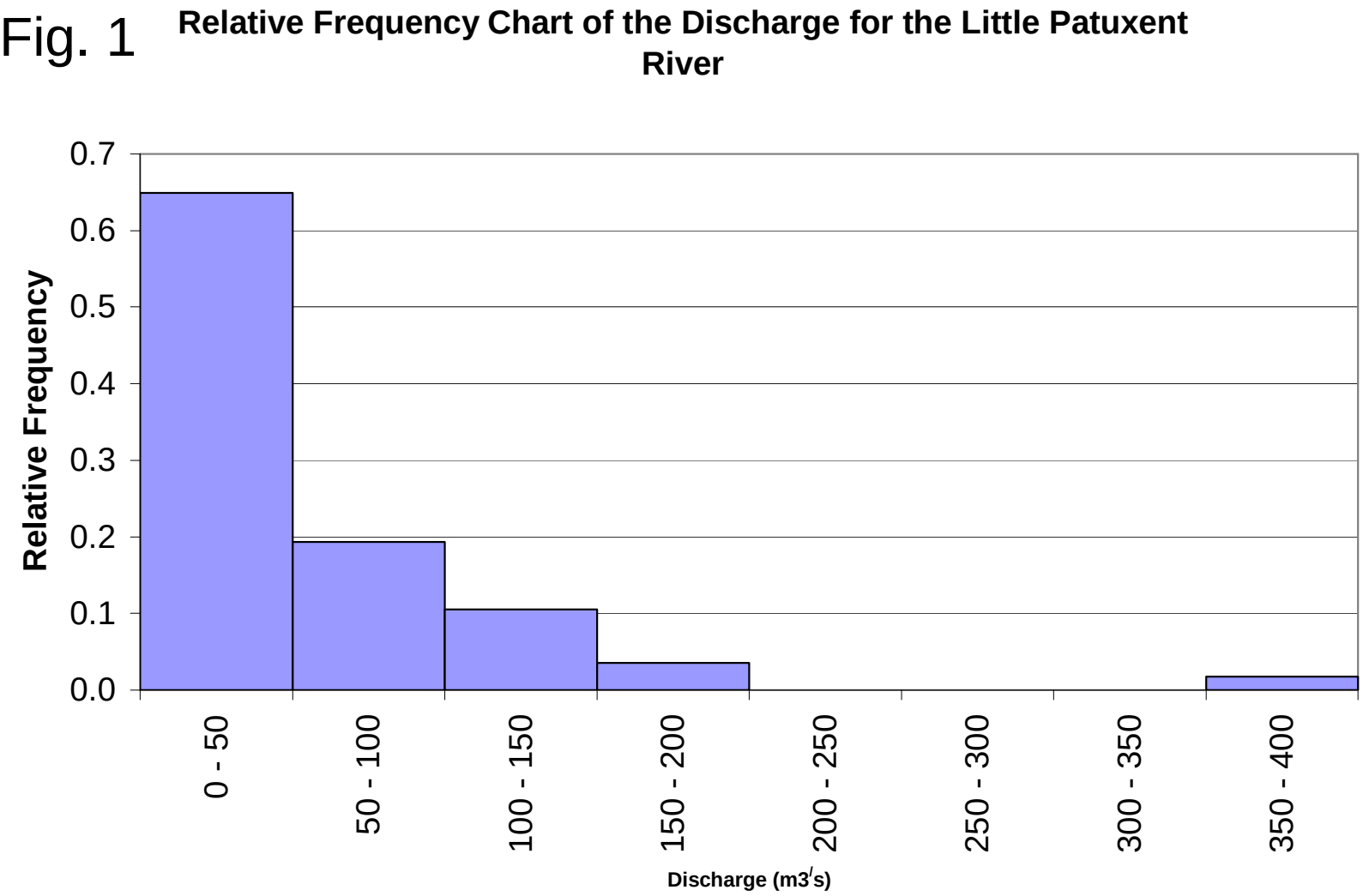
Analysis of Simulated Data

Histogram and Frequency Histogram Table of Discharge for the Little Patuxent River, Guilford, MD

Interval	Frequency	Relative Frequency
0 - 50	37	0.6491
50 - 100	11	0.1930
100 - 150	6	0.1053
150 - 200	2	0.0351
200 - 250	0	0.0000
250 - 300	0	0.0000
300 - 350	0	0.0000
350 - 400	1	0.0175
Total =	57	1

Analysis of Simulated Data

Fig. 1



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Analysis of Simulated Data

The histogram of Figure 1 suggests that the data follow an exponential decay, which is given mathematically as

$$f(x) = \frac{1}{b} e^{-x/b}$$

where b is a parameter. For data that follow this distribution, the sample mean can be used as an estimate of b

Analysis of Simulated Data

To simulate the discharges, the cumulative function $F_X(x)$ for the density function of the previous equation is given as

$$F(x) = 1 - e^{-x/b}$$

It can be shown (see Textbook) that the above equation can be rewritten in the following form:

$$x = -b \ln\left[F_X'(x)\right]$$

Analysis of Simulated Data

- Simulation of River Discharge Rates

$$x_i = -54.82 \ln(u_i)$$

Where u_i is the i^{th} uniform variate (i.e., a random number between 0 and 1) and x_i is the i^{th} simulated discharge.

The above equation can be used to simulate the River discharge.

NOTE: b = mean of actual data = 54.82

Analysis of Simulated Data

Simulated Discharge Rates for Little Patuxent River

r^2	r	Random number (u)	Discharge, x (cms)	Rank	r^2	r	Random number (u)	Discharge, x (cms)	Rank
	8,765				06,594,624	5,946	0.5946	28.5	36
76,825,225	8,252	0.8252	10.5	47	35,354,916	3,549	0.3549	56.8	25
68,095,504	955	0.0955	128.8	9	12,595,401	5,954	0.5954	28.4	38
00,912,025	9,120	0.912	5.0	53	35,450,116	4,501	0.4501	43.8	32
83,174,400	1,744	0.1744	95.7	17	20,259,001	2,590	0.259	74.1	23
03,041,536	415	0.0415	174.4	5	06,708,100	7,081	0.7081	18.9	43
00,172,225	1,722	0.1722	96.4	16	50,140,561	1,405	0.1405	107.6	12
02,965,284	9,652	0.9652	1.9	56	01,974,025	9,740	0.974	1.4	57
93,161,104	1,611	0.1611	100.1	15	94,867,600	8,676	0.8676	7.8	50
02,595,321	5,953	0.5953	28.4	37	75,272,976	2,729	0.2729	71.2	24
35,438,209	4,382	0.4382	45.2	29	07,447,441	4,474	0.4474	44.1	31
19,201,924	2,019	0.2019	87.7	18	20,016,676	166	0.0166	224.7	1
04,076,361	763	0.0763	141.1	8	00,027,556	275	0.0275	197.0	2
00,582,169	5,821	0.5821	29.7	35	00,075,625	756	0.0756	141.6	7
33,884,041	8,840	0.884	6.8	51	00,571,536	5,715	0.5715	30.7	34
78,145,600	1,456	0.1456	105.6	13	32,661,225	6,612	0.6612	22.7	41
02,119,936	1,199	0.1199	116.3	10	43,718,544	7,185	0.7185	18.1	44
01,437,601	4,376	0.4376	45.3	28	51,624,225	6,242	0.6242	25.8	39
19,149,376	1,493	0.1493	104.3	14	38,962,564	9,625	0.9625	2.1	55
02,229,049	2,290	0.229	80.8	19	92,640,625	6,406	0.6406	24.4	40
05,244,100	2,441	0.2441	77.3	20	41,036,836	368	0.0368	181.0	4
05,958,481	9,584	0.9584	2.3	54	00,135,424	1,354	0.1354	109.6	11
91,853,056	8,530	0.853	8.7	49	01,833,316	8,333	0.8333	10.0	48
72,760,900	7,609	0.7609	15.0	46	69,438,889	4,388	0.4388	45.2	29
57,896,881	8,968	0.8968	6.0	52	19,254,544	2,545	0.2545	75.0	21
80,425,024	4,250	0.425	46.9	27	06,477,025	4,770	0.477	40.6	33
18,062,500	625	0.0625	152.0	6	22,752,900	7,529	0.7529	15.6	45
00,390,625	3,906	0.3906	51.5	26	56,685,841	6,858	0.6858	20.7	42
15,256,836	2,568	0.2568	74.5	22	47,032,164	321	0.0321	188.5	3

Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Analysis of Simulated Data

Comparison of Statistics between the Actual and Simulated Discharge Rates for Little Patuxent River

Parameter	Discharge (cms)	
	Actual	Simulated
Average (Mean)	54.82	64.81
Median	34.80	45.20
Mode	34.80	28.40
Standard Deviation	53.78	57.57
Sample Variance	2892.52	3314.35
COV	0.98	0.89
Range	335.20	223.30
Minimum	15.90	1.40
Maximum	351.10	224.70
Count	57	57

Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

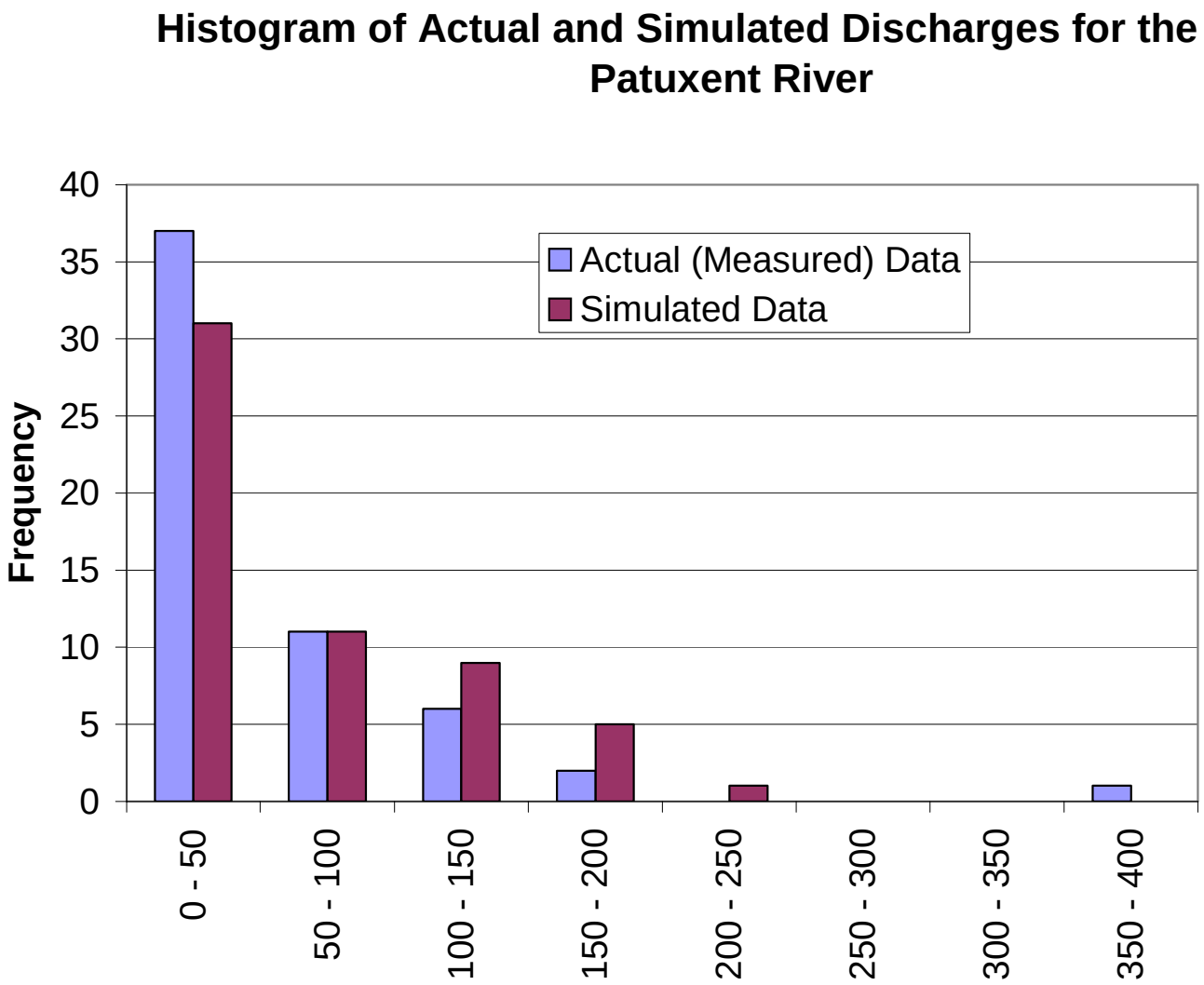
Analysis of Simulated Data

Comparison of Tabulated Frequency Histograms between the Actual and Simulated Discharge Rates for Little Patuxent River

Interval	Frequency	
	Actual	Simulated
0 - 50	37	31
50 - 100	11	11
100 - 150	6	9
150 - 200	2	5
200 - 250	0	1
250 - 300	0	0
300 - 350	0	0
350 - 400	1	0
Total =	57	57

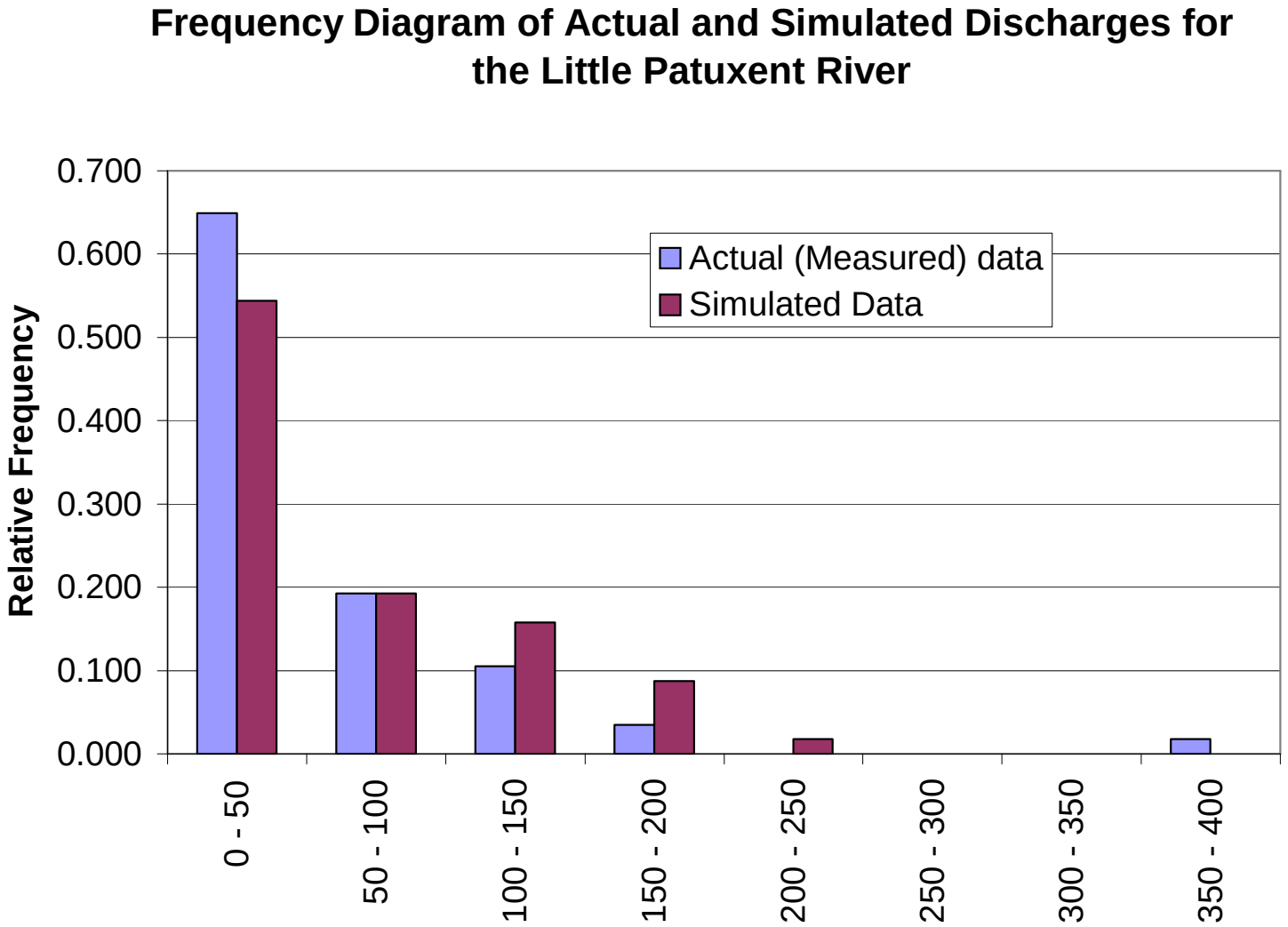
Interval	Relative Frequency	
	Actual	Simulated
0 - 50	0.649	0.544
50 - 100	0.193	0.193
100 - 150	0.105	0.158
150 - 200	0.035	0.088
200 - 250	0.000	0.018
250 - 300	0.000	0.000
300 - 350	0.000	0.000
350 - 400	0.018	0.000
Total =	1	1

Analysis of Simulated Data



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen

Analysis of Simulated Data



Probability, Statistics and Reliability for Engineers and Scientists, Third Edition, Bilal M. Ayyub and Richard H. McCuen