CHAPTER 3

The Mean, Median, Mode, and Other Measures of Central Tendency

INDEX, OR SUBSCRIPT, NOTATION

Let the symbol X_j (read "X sub j") denote any of the N *values* $X_1, X_2, X_3, \ldots, X_N$ assumed by a variable X. The letter j in X_j, which can stand for any of the numbers $1, 2, 3, \ldots, N$ is called a *subscript*, or *index*. Clearly any letter other than j, such as i, k, p, q, or s, could have been used as well.

SUMMATION NOTATION

The symbol $\sum_{j=1}^{N} X_j$ is used to denote the sum of all the X_j's from $j = 1$ to $j = N$; by definition,

$$\sum_{j=1}^{N} X_j = X_1 + X_2 + X_3 + \cdots + X_N$$

When no confusion can result, we often denote this sum simply by $\sum X$, $\sum X_j$, or $\sum_j X_j$. The symbol $\sum$ is the Greek capital letter *sigma*, denoting sum.

EXAMPLE 1. $\displaystyle\sum_{j=1}^{N} X_j Y_j = X_1 Y_1 + X_2 Y_2 + X_3 Y_3 + \cdots + X_N Y_N$

EXAMPLE 2. $\displaystyle\sum_{j=1}^{N} aX_j = aX_1 + aX_2 + \cdots + aX_N = a(X_1 + X_2 + \cdots + X_N) = a\sum_{j=1}^{N} X_j$

where a is a constant. More simply, $\sum aX = a\sum X$.

EXAMPLE 3. If a, b, and c are any constants, then $\sum (aX + bY - cZ) = a\sum X + b\sum Y - c\sum Z$. See Problem 3.3.

AVERAGES, OR MEASURES OF CENTRAL TENDENCY

An *average* is a value that is typical, or representative, of a set of data. Since such typical values tend to lie centrally within a set of data arranged according to magnitude, averages are also called *measures of central tendency*.

Several types of averages can be defined, the most common being the *arithmetic mean*, the *median*, the *mode*, the *geometric mean*, and the *harmonic mean*. Each has advantages and disadvantages, depending on the data and the intended purpose.

THE ARITHMETIC MEAN

The *arithmetic mean*, or briefly the *mean*, of a set of N numbers $X_1, X_2, X_3, \ldots, X_N$ is denoted by $\bar{X}$ (read "X bar") and is defined as

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum_{j=1}^{N} X_j}{N} = \frac{\sum X}{N} \tag{1}$$

EXAMPLE 4. The arithmetic mean of the numbers 8, 3, 5, 12, and 10 is

$$\bar{X} = \frac{8 + 3 + 5 + 12 + 10}{5} = \frac{38}{5} = 7.6$$

If the numbers $X_1, X_2, \ldots, X_K$ occur $f_1, f_2, \ldots, f_K$ times, respectively (i.e., occur with frequencies $f_1, f_2, \ldots, f_K$), the arithmetic mean is

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \cdots + f_K X_K}{f_1 + f_2 + \cdots + f_K} = \frac{\sum_{j=1}^{K} f_j X_j}{\sum_{j=1}^{K} f_j} = \frac{\sum fX}{\sum f} = \frac{\sum f_X}{N} \tag{2}$$

where $N = \sum f$ is the *total frequency* (i.e., the total number of cases).

EXAMPLE 5. If 5, 8, 6, and 2 occur with frequencies 3, 2, 4, and 1, respectively, the arithmetic mean is

$$\bar{X} = \frac{(3)(5) + (2)(8) + (4)(6) + (1)(2)}{3 + 2 + 4 + 1} = \frac{15 + 16 + 24 + 2}{10} = 5.7$$

THE WEIGHTED ARITHMETIC MEAN

Sometimes we associate with the numbers $X_1, X_2, \ldots, X_K$ certain *weighting factors* (or *weights*) $w_1, w_2, \ldots, w_K$, depending on the significance or importance attached to the numbers. In this case,

$$\bar{X} = \frac{w_1 X_1 + w_2 X_2 + \cdots + w_K X_k}{w_1 + w_2 + \cdots + w_K} = \frac{\sum wX}{\sum w} \tag{3}$$

is called the *weighted arithmetic mean*. Note the similarity to equation (*2*), which can be considered a weighted arithmetic mean with weights $f_1, f_2, \ldots, f_K$.

EXAMPLE 6. If a final examination in a course is weighted 3 times as much as a quiz and a student has a final examination grade of 85 and quiz grades of 70 and 90, the mean grade is

$$\bar{X} = \frac{(1)(70) + (1)(90) + (3)(85)}{1 + 1 + 3} = \frac{415}{5} = 83$$

PROPERTIES OF THE ARITHMETIC MEAN

1. The algebraic sum of the deviations of a set of numbers from their arithmetic mean is zero.

EXAMPLE 7. The deviations of the numbers 8, 3, 5, 12, and 10 from their arithmetic mean 7.6 are $8 - 7.6$, $3 - 7.6$, $5 - 7.6$, $12 - 7.6$, and $10 - 7.6$, or 0.4, −4.6, −2.6, 4.4, and 2.4, with algebraic sum $0.4 - 4.6 - 2.6 + 4.4 + 2.4 = 0$.

2. The sum of the squares of the deviations of a set of numbers X_j from any number a is a minimum if and only if $a = \bar{X}$ (see Problem 4.27).
3. If f_1 numbers have mean m_1, f_2 numbers have mean $m_2, \ldots, f_K$ numbers have mean m_K, then the mean of all the numbers is

$$\bar{X} = \frac{f_1 m_1 + f_2 m_2 + \cdots + f_K m_K}{f_1 + f_2 + \cdots + f_K} \tag{4}$$

that is, a weighted arithmetic mean of all the means (see Problem 3.12).

4. If A is any *guessed* or *assumed arithmetic mean* (which may be any number) and if $d_j = X_j - A$ are the deviations of X_j from A, then equations (*1*) and (*2*) become, respectively,

$$\bar{X} = A + \frac{\sum_{j=1}^{N} d_j}{N} = A + \frac{\sum d}{N} \tag{5}$$

$$\bar{X} = A + \frac{\sum_{j=1}^{K} f_j d_j}{\sum_{j=1}^{K} f_j} = A + \frac{\sum fd}{N} \tag{6}$$

where $N = \sum_{j=1}^{K} f_j = \sum f$. Note that formulas (*5*) and (*6*) are summarized in the equation $\bar{X} = A + \bar{d}$ (see Problem 3.18).

THE ARITHMETIC MEAN COMPUTED FROM GROUPED DATA

When data are presented in a frequency distribution, all values falling within a given class interval are considerd to be coincident with the class mark, or midpoint, of the interval. Formulas (*2*) and (*6*) are valid for such grouped data if we interpret X_j as the class mark, f_j as its corresponding class frequency, A as any guessed or assumed class mark, and $d_j = X_j - A$ as the deviations of X_j from A.

Computations using formulas (*2*) and (*6*) are sometimes called the *long* and *short methods*, respectively (see Problems 3.15 and 3.20).

If class intervals all have equal size c, the deviations $d_j = X_j - A$ can all be expressed as cu_j, where u_j can be positive or negative integers or zero (i.e., 0, ±1, ±2, ±3, ...), and formula (*6*) becomes

$$\bar{X} = A + \left(\frac{\sum_{j=1}^{K} f_j u_j}{N} \right) = A + \left(\frac{\sum fu}{N} \right) c \tag{7}$$

which is equivalent to the equation $\bar{X} = A + c\bar{u}$ (see Problem 3.21). This is called the *coding method* for computing the mean. It is a very short method and should always be used for grouped data where the class-interval sizes are equal (see Problems 3.22 and 3.23). Note that in the coding method the values of the variable X are *transformed* into the values of the variable u according to $X = A + cu$.

THE MEDIAN

The *median* of a set of numbers arranged in order of magnitude (i.e., in an array) is either the middle value or the arithmetic mean of the two middle values.

EXAMPLE 8. The set of numbers 3, 4, 4, 5, 6, 8, 8, 8, and 10 has median 6.

EXAMPLE 9. The set of numbers 5, 5, 7, 9, 11, 12, 15, and 18 has median $\frac{1}{2}(9+11) = 10$.

For grouped data, the median, obtained by interpolation, is given by

$$\text{Median} = L_1 + \left(\frac{\frac{N}{2} - (\sum f)_1}{f_{\text{median}}} \right) c \tag{8}$$

where L_1 = lower class boundary of the median class (i.e., the class containing the median)
N = number of items in the data (i.e., total frequency)
$(\sum f)_1$ = sum of frequencies of all classes lower than the median class
f_{median} = frequency of the median class
c = size of the median class interval

Geometrically the median is the value of X (abscissa) corresponding to the vertical line which divides a histogram into two parts having equal areas. This value of X is sometimes denoted by $\tilde{X}$.

THE MODE

The *mode* of a set of numbers is that value which occurs with the greatest frequency; that is, it is the most common value. The mode may not exist, and even if it does exist it may not be unique.

EXAMPLE 10. The set 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, and 18 has mode 9.

EXAMPLE 11. The set 3, 5, 8, 10, 12, 15, and 16 has no mode.

EXAMPLE 12. The set 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, and 9 has two modes, 4 and 7, and is called *bimodal*.

A distribution having only one mode is called *unimodal*.

In the case of grouped data where a frequency curve has been constructed to fit the data, the mode will be the value (or values) of X corresponding to the maximum point (or points) on the curve. This value of X is sometimes denoted by $\hat{X}$.

From a frequency distribution or histogram the mode can be obtained from the formula

$$\text{Mode} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c \tag{9}$$

where L_1 = lower class boundary of the modal class (i.e., the class containing the mode)
Δ_1 = excess of modal frequency over frequency of next-lower class
Δ_2 = excess of modal frequency over frequency of next-higher class
c = size of the modal class interval

THE EMPIRICAL RELATION BETWEEN THE MEAN, MEDIAN, AND MODE

For unimodal frequency curves that are moderately skewed (asymmetrical), we have the empirical relation

$$\text{Mean} - \text{mode} = 3(\text{mean} - \text{median}) \tag{10}$$

Figures 3-1 and 3-2 show the relative positions of the mean, median, and mode for frequency curves skewed to the right and left, respectively. For symmetrical curves, the mean, mode, and median all coincide.

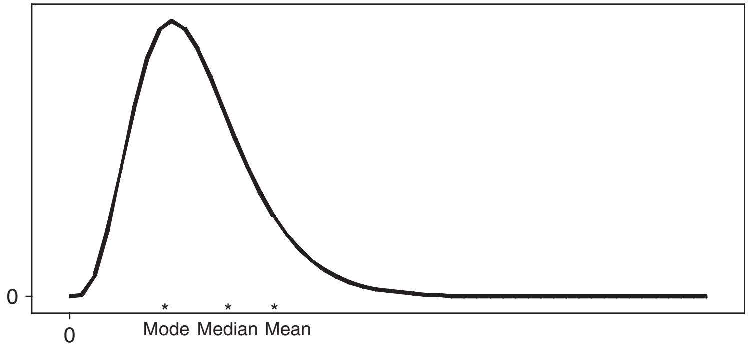

Fig. 3-1 Relative positions of mode, median, and mean for a right-skewed frequency curve.

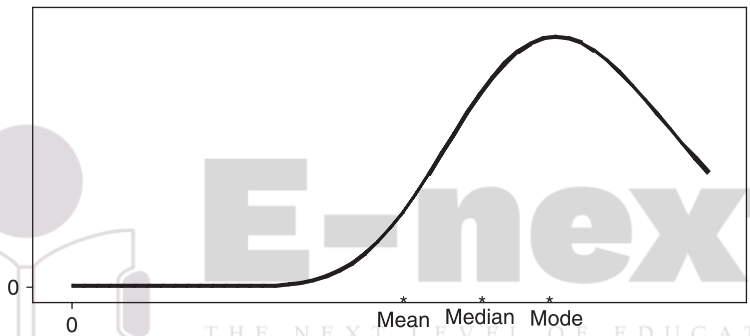

Fig. 3-2 Relative positions of mode, median, and mean for a left-skewed frequency curve.

THE GEOMETRIC MEAN G

The geometric mean G of a set of N positive numbers $X_1, X_2, X_3, \ldots, X_N$ is the Nth root of the product of the numbers:

$$G = \sqrt[N]{X_1 X_2 X_3 \cdots X_N} \tag{11}$$

EXAMPLE 13. The geometric mean of the numbers 2, 4, and 8 is $G = \sqrt[3]{(2)(4)(8)} = \sqrt[3]{64} = 4$.

We can compute G by logarithms (see Problem 3.35) or by using a calculator. For the geometric mean from grouped data, see Problems 3.36 and 3.91.

THE HARMONIC MEAN H

The harmonic mean H of a set of N numbers $X_1, X_2, X_3, \ldots, X_N$ is the reciprocal of the arithmetic mean of the reciprocals of the numbers:

$$H = \frac{1}{\dfrac{1}{N}\displaystyle\sum_{j=1}^{N}\frac{1}{X_j}} = \frac{N}{\sum \dfrac{1}{X}} \tag{12}$$

In practice it may be easier to remember that

$$\frac{1}{H} = \frac{\sum \frac{1}{X}}{N} = \frac{1}{N}\sum \frac{1}{X} \tag{13}$$

EXAMPLE 14. The harmonic mean of the numbers 2, 4, and 8 is

$$H = \frac{3}{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}} = \frac{3}{\frac{7}{8}} = 3.43$$

For the harmonic mean from grouped data, see Problems 3.99 and 3.100.

THE RELATION BETWEEN THE ARITHMETIC, GEOMETRIC, AND HARMONIC MEANS

The geometric mean of a set of positive numbers $X_1, X_2, \ldots, X_N$ is less than or equal to their arithmetic mean but is greater than or equal to their harmonic mean. In symbols,

$$H \le G \le \bar{X} \tag{14}$$

The equality signs hold only if all the numbers $X_1, X_2, \ldots, X_N$ are identical.

EXAMPLE 15. The set 2, 4, 8 has arithmetic mean 4.67, geometric mean 4, and harmonic mean 3.43.

THE ROOT MEAN SQUARE

The root mean square (RMS), or *quadratic mean*, of a set of numbers $X_1, X_2, \ldots, X_N$ is sometimes denoted by $\sqrt{\overline{X^2}}$ and is defined by

$$\text{RMS} = \sqrt{\overline{X^2}} = \sqrt{\frac{\sum_{j=1}^{N} X_j^2}{N}} = \sqrt{\frac{\sum X^2}{N}} \tag{15}$$

This type of average is frequently used in physical applications.

EXAMPLE 16. The RMS of the set 1, 3, 4, 5, and 7 is

$$\sqrt{\frac{1^2+3^2+4^2+5^2+7^2}{5}} = \sqrt{20} = 4.47$$

QUARTILES, DECILES, AND PERCENTILES

If a set of data is arranged in order of magnitude, the middle value (or arithmetic mean of the two middle values) that divides the set into two equal parts is the median. By extending this idea, we can think of those values which divide the set into four equal parts. These values, denoted by Q_1, Q_2, and Q_3, are called the first, second, and third *quartiles*, respectively, the value Q_2 being equal to the median.

Similarly, the values that divide the data into 10 equal parts are called *deciles* and are denoted by $D_1, D_2, \ldots, D_9$, while the values dividing the data into 100 equal parts are called *percentiles* and are denoted by $P_1, P_2, \ldots, P_{99}$. The fifth decile and the 50th percentile correspond to the median. The 25th and 75th percentiles correspond to the first and third quartiles, respectively.

Collectively, quartiles, deciles, percentiles, and other values obtained by equal subdivisions of the data are called *quantiles*. For computations of these from grouped data, see Problems 3.44 to 3.46.

EXAMPLE 17. Use EXCEL to find Q_1, Q_2, Q_3, D_9, and P_{95} for the following sample of test scores:

88	45	53	86	33	86	85	30	89	53	41	96	56	38	62
71	51	86	68	29	28	47	33	37	25	36	33	94	73	46
42	34	79	72	88	99	82	62	57	42	28	55	67	62	60
96	61	57	75	93	34	75	53	32	28	73	51	69	91	35

To find the first quartile, put the data in the first 60 rows of column A of the EXCEL worksheet. Then give the command =PERCENTILE(A1:A60,0.25). EXCEL returns the value 37.75. We find that 15 out of the 60 values are less than 37.75 or 25% of the scores are less than 37.75. Similarly, we find =PERCENTILE(A1:A60,0.5) gives 57, =PERCENTILE(A1:A60,0.75) gives 76, =PERCENTILE(A1:A60,0.9) gives 89.2, and =PERCENTILE(A1:A60,0.95) gives 94.1. EXCEL gives quartiles, deciles, and percentiles all expressed as percentiles.

An algorithm that is often used to find quartiles, deciles, and percentiles by hand is described as follows. The data in example 17 is first sorted and the result is:

test	scores													
25	28	28	28	29	30	32	33	33	33	34	34	35	36	37
38	41	42	42	45	46	47	51	51	53	53	53	55	56	57
57	60	61	62	62	62	67	68	69	71	72	73	73	75	75
79	82	85	86	86	86	88	88	89	91	93	94	96	96	99

Suppose we wish to find the first quartile (25th percentile). Calculate $i = np/100 = 60(25)/100 = 15$. Since 15 is a whole number, go to the 15th and 16th numbers in the sorted array and average the 15th and 16th numbers. That is average 37 and 38 to get 37.5 as the first quartile ($Q_1 = 37.5$). To find the 93rd percentile, calculate $np/100 = 60(93)/100$ to get 55.8. Since this number is not a whole number, always round it up to get 56. The 56th number in the sorted array is 93 and $P_{93} = 93$. The EXCEL command =PERCENTILE(A1:A60,0.93) gives 92.74. Note that EXCEL does not give the same values for the percentiles, but the values are close. As the data set becomes larger, the values tend to equal each other.

SOFTWARE AND MEASURES OF CENTRAL TENDENCY

All the software packages utilized in this book give the descriptive statistics discussed in this section. The output for all five packages is now given for the test scores in Example 17.

EXCEL

If the pull-down "Tools ⇒ Data Analysis ⇒ Descriptive Statistics" is given, the measures of central tendency median, mean, and mode as well as several measures of dispersion are obtained:

Mean	59.16667
Standard Error	2.867425
Median	57
Mode	28
Standard Deviation	22.21098
Sample Variance	493.3277
Kurtosis	−1.24413
Skewness	0.167175
Range	74
Minimum	25
Maximum	99
Sum	3550
Count	60

MINITAB

If the pull-down "Stat ⇒ Basic Statistics ⇒ Display Descriptive Statistics" is given, the following output is obtained:

Descriptive Statistics: testscore

```
Variable     N  N*   Mean  SE Mean  St Dev  Minimum     Q1  Median     Q3
testscore   60   0  59.17     2.87   22.21    25.00  37.25   57.00  78.00

Variable   Maximum
testscore    99.00
```

SPSS

If the pull-down "Analyze ⇒ Descriptive Statistics ⇒ Descriptives" is given, the following output is obtained:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Testscore	60	25.00	99.00	59.1667	22.21098
Valid *N* (listwise)	60				

SAS

If the pull-down "**Solutions** ⇒ **Analysis** ⇒ **Analyst**" is given and the data are read in as a file, the pull-down "**Statistics** ⇒ **Descriptive** ⇒ **Summary Statistics**" gives the following output:

```
                    The MEANS Procedure
                  Analysis Variable : Testscr

       Mean          Std Dev         N          Minimum       Maximum
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
  59.1666667      22.2109811        60       25.0000000    99.0000000
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
```

STATISTIX

If the pull-down "Statistics ⇒ Summary Statistics ⇒ Descriptive Statistics" is given in the software package STATISTIX, the following output is obtained:

```
Statistix 8.0
Descriptive Staistics

                Testscore
N                      60
Mean               59.167
SD                 22.211
Minimum            25.000
1st Quarti         37.250
3rd Quarti         78.000
Maximum            99.000
```

Solved Problems

SUMMATION NOTATION

3.1 Write out the terms in each of the following indicated sums:

(a) $\sum_{j=1}^{6} X_j$ (c) $\sum_{j=1}^{N} a$ (e) $\sum_{j=1}^{3} (X_j - a)$

(b) $\sum_{j=1}^{4} (Y_j - 3)^2$ (d) $\sum_{k=1}^{5} f_k X_k$

SOLUTION

(a) $X_1 + X_2 + X_3 + X_4 + X_5 + X_6$

(b) $(Y_1 - 3)^2 + (Y_2 - 3)^2 + (Y_3 - 3)^2 + (Y_4 - 3)^2$

(c) $a + a + a + \cdots + a = Na$

(d) $f_1 X_1 + f_2 X_2 + f_3 X_3 + f_4 X_4 + f_5 X_5$

(e) $(X_1 - a) + (X_2 - a) + (X_3 - a) = X_1 + X_2 + X_3 - 3a$

3.2 Express each of the following by using the summation notation:

(a) $X_1^2 + X_2^2 + X_3^2 + \cdots + X_{10}^2$

(b) $(X_1 + Y_1) + (X_2 + Y_2) + \cdots + (X_8 + Y_8)$

(c) $f_1 X_1^3 + f_2 X_2^3 + \cdots + f_{20} X_{20}^3$

(d) $a_1 b_1 + a_2 b_2 + a_3 b_3 + \cdots + a_N b_N$

(e) $f_1 X_1 Y_1 + f_2 X_2 Y_2 + f_3 X_3 Y_3 + f_4 X_4 Y_4$

SOLUTION

(a) $\sum_{j=1}^{10} X_j^2$ (c) $\sum_{j=1}^{20} f_j X_j^3$ (e) $\sum_{j=1}^{4} f_j X_j Y_j$

(b) $\sum_{j=1}^{8} (X_j + Y_j)$ (d) $\sum_{j=1}^{N} a_j b_j$

3.3 Prove that $\sum_{j=1}^{N} (aX_j + bY_j - cZ_j) = a\sum_{j=1}^{N} X_j + b\sum_{j=1}^{N} Y_j - c\sum_{j=1}^{N} Z_j$, where a, b, and c are any constants.

SOLUTION

$$\begin{aligned}
\sum_{j=1}^{N} (aX_j + bY_j - cZ_j) &= (aX_1 + bY_1 - cZ_1) + (aX_2 + bY_2 - cZ_2) + \cdots + (aX_N + bY_N - cZ_N) \\
&= (aX_1 + aX_2 + \cdots + aX_N) + (bY_1 + bY_2 + \cdots + bY_N) - (cZ_1 + cZ_2 + \cdots + cZ_N) \\
&= a(X_1 + X_2 + \cdots + X_N) + b(Y_1 + Y_2 + \cdots + Y_N) - c(Z_1 + Z_2 + \cdots + Z_N) \\
&= a\sum_{j=1}^{N} X_j + b\sum_{j=1}^{N} Y_j - c\sum_{j=1}^{N} Z_j
\end{aligned}$$

or briefly, $\sum (aX + bY - cZ) = a\sum X + b\sum Y - c\sum Z$.

3.4 Two variables, X and Y, assume the values $X_1 = 2$, $X_2 = -5$, $X_3 = 4$, $X_4 = -8$ and $Y_1 = -3$, $Y_2 = -8$, $Y_3 = 10$, $Y_4 = 6$, respectively. Calculate (*a*) $\sum X$, (*b*) $\sum Y$, (*c*) $\sum XY$, (*d*) $\sum X^2$, (*e*) $\sum Y^2$, (*f*) $(\sum X)(\sum Y)$, (*g*) $\sum XY^2$, and (*h*) $\sum (X+Y)(X-Y)$.

SOLUTION

Note that in each case the subscript j on X and Y has been omitted and $\sum$ is understood as $\sum_{j=1}^{4}$. Thus $\sum X$, for example, is short for $\sum_{j=1}^{4} X_j$.

(*a*) $\sum X = (2) + (-5) + (4) + (-8) = 2 - 5 + 4 - 8 = -7$

(*b*) $\sum Y = (-3) + (-8) + (10) + (6) = -3 - 8 + 10 + 6 = 5$

(*c*) $\sum XY = (2)(-3) + (-5)(-8) + (4)(10) + (-8)(6) = -6 + 40 + 40 - 48 = 26$

(*d*) $\sum X^2 = (2)^2 + (-5)^2 + (4)^2 + (-8)^2 = 4 + 25 + 16 + 64 = 109$

(*e*) $\sum Y^2 = (-3)^2 + (-8)^2 + (10)^2 + (6)^2 = 9 + 64 + 100 + 36 = 209$

(*f*) $(\sum X)(\sum Y) = (-7)(5) = -35$, using parts (*a*) and (*b*). Note that $(\sum X)(\sum Y) \neq \sum XY$.

(*g*) $\sum XY^2 = (2)(-3)^2 + (-5)(-8)^2 + (4)(10)^2 + (-8)(6)^2 = -190$

(*h*) $\sum (X+Y)(X-Y) = \sum (X^2 - Y^2) = \sum X^2 - \sum Y^2 = 109 - 209 = -100$, using parts (*d*) and (*e*).

3.5 In a *USA Today* snapshot, it was reported that per capita state taxes collected in 2005 averaged \$2189.84 across the United States. The break down is: Sales and receipts \$1051.42, Income \$875.23, Licenses \$144.33, Others \$80.49, and Property \$38.36. Use EXCEL to show that the sum equals \$2189.84.

SOLUTION

Note that the expression = sum(A1:A5) is equivalent to $\sum_{j=1}^{5} X_j$.

1051.42	sales and receipts
875.23	income
144.33	licenses
80.49	others
38.36	property
2189.83	=sum(A1:A5)

THE ARITHMETIC MEAN

3.6 The grades of a student on six examinations were 84, 91, 72, 68, 87, and 78. Find the arithmetic mean of the grades.

SOLUTION

$$\bar{X} = \frac{\sum X}{N} = \frac{84 + 91 + 72 + 68 + 87 + 78}{6} = \frac{480}{6} = 80$$

Frequently one uses the term *average* synonymously with *arithmetic mean*. Strictly speaking, however, this is incorrect since there are averages other than the arithmetic mean.

3.7 Ten measurements of the diameter of a cylinder were recorded by a scientist as 3.88, 4.09, 3.92, 3.97, 4.02, 3.95, 4.03, 3.92, 3.98, and 4.06 centimeters (cm). Find the arithmetic mean of the measurements.

SOLUTION

$$\bar{X} = \frac{\sum X}{N} = \frac{3.88 + 4.09 + 3.92 + 3.97 + 4.02 + 3.95 + 4.03 + 3.92 + 3.98 + 4.06}{10} = \frac{39.82}{10} = 3.98\,\text{cm}$$

3.8 The following MINITAB output shows the time spent per week on line for 30 Internet users as well as the mean of the 30 times. Would you say this average is typical of the 30 times?

```
MTB > print cl
```

Data Display

```
time
3   4   4   5   5   5   5   5   5   6
6   6   6   7   7   7   7   7   8   8
9  10  10  10  10  10  10  12  55  60
```

```
MTB > mean cl
```

Column Mean

```
Mean of time = 10.400
```

SOLUTION

The mean 10.4 hours is not typical of the times. Note that 21 of the 30 times are in the single digits, but the mean is 10.4 hours. A great disadvantage of the mean is that it is strongly affected by extreme values.

3.9 Find the arithmetic mean of the numbers 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, and 4.

SOLUTION

First method

$$\bar{X} = \frac{\sum X}{N} = \frac{5+3+6+5+4+5+2+8+6+5+4+8+3+4+5+4+8+2+5+4}{20} = \frac{96}{20} = 4.8$$

Second method

There are six 5's, two 3's, two 6's, five 4's, two 2's and three 8's. Thus

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{N} = \frac{(6)(5)+(2)(3)+(2)(6)+(5)(4)+(2)(2)+(3)(8)}{6+2+2+5+2+3} = \frac{96}{20} = 4.8$$

3.10 Out of 100 numbers, 20 were 4's, 40 were 5's, 30 were 6's and the remainder were 7's. Find the arithmetic mean of the numbers.

SOLUTION

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{N} = \frac{(20)(4)+(40)(5)+(30)(6)+(10)(7)}{100} = \frac{530}{100} = 5.30$$

3.11 A student's final grades in mathematics, physics, English and hygiene are, respectively, 82, 86, 90, and 70. If the respective credits received for these courses are 3, 5, 3, and 1, determine an appropriate average grade.

SOLUTION

We use a weighted arithmetic mean, the weights associated with each grade being taken as the number of credits received. Thus

$$\bar{X} = \frac{\sum wX}{\sum w} = \frac{(3)(82) + (5)(86) + (3)(90) + (1)(70)}{3 + 5 + 3 + 1} = 85$$

3.12 In a company having 80 employees, 60 earn \$10.00 per hour and 20 earn \$13.00 per hour.

(*a*) Determine the mean earnings per hour.

(*b*) Would the answer in part (*a*) be the same if the 60 employees earn a mean hourly wage of \$10.00 per hour? Prove your answer.

(*c*) Do you believe the mean hourly wage to be typical?

SOLUTION

(*a*)

$$\bar{X} = \frac{\sum fX}{N} = \frac{(60)(\$10.00) + (20)(\$13.00)}{60 + 20} = \$10.75$$

(*b*) Yes, the result is the same. To prove this, suppose that f_1 numbers have mean m_1 and that f_2 numbers have mean m_2. We must show that the mean of all the numbers is

$$\bar{X} = \frac{f_1 m_1 + f_2 m_2}{f_1 + f_2}$$

Let the f_1 numbers add up to M_1 and the f_2 numbers add up to M_2. Then by definition of the arithmetic mean,

$$m_1 = \frac{M_1}{f_1} \qquad \text{and} \qquad m_2 = \frac{M_2}{f_2}$$

or $M_1 = f_1 m_1$ and $M_2 = f_2 m_2$. Since all $(f_1 + f_2)$ numbers add up to $(M_1 + M_2)$, the arithmetic mean of all numbers is

$$\bar{X} = \frac{M_1 + M_2}{f_1 + f_2} = \frac{f_1 m_1 + f_2 m_2}{f_1 + f_2}$$

as required. The result is easily extended.

(*c*) We can say that \$10.75 is a "typical" hourly wage in the sense that most of the employees earn \$10.00, which is not too far from \$10.75 per hour. It must be remembered that, whenever we summarize numerical data into a single number (as is true in an average), we are likely to make some error. Certainly, however, the result is not as misleading as that in Problem 3.8.

Actually, to be on safe ground, some estimate of the "spread," or "variation," of the data about the mean (or other average) should be given. This is called the *dispersion* of the data. Various measures of dispersion are given in Chapter 4.

3.13 Four groups of students, consisting of 15, 20, 10, and 18 individuals, reported mean weights of 162, 148, 153, and 140 pounds (lb), respectively. Find the mean weight of all the students.

SOLUTION

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{(15)(162) + (20)(148) + (10)(153) + (18)(140)}{15 + 20 + 10 + 18} = 150\,\text{lb}$$

3.14 If the mean annual incomes of agricultural and nonagricultural workers are \$25,000 and \$35,000, respectively, would the mean annual income of both groups together be \$30,000?

SOLUTION

It would be \$30,000 only if the numbers of agricultural and nonagricultural workers were the same. To determine the true mean annual income, we would have to know the relative numbers of workers in each group. Suppose that 10% of all workers are agricultural workers. Then the mean would be $(0.10)(25,000) + (0.90)(35,000) = \$34,000$. If there were equal numbers of both types of workers, the mean would be $(0.50)(25,000) + (0.50)(35,000) = \$30,000$.

3.15 Use the frequency distribution of heights in Table 2.1 to find the mean height of the 100 male students at XYZ university.

SOLUTION

The work is outlined in Table 3.1. Note that all students having heights 60 to 62 inches (in), 63 to 65 in, etc., are considered as having heights 61 in, 64 in, etc. The problem then reduces to finding the mean height of 100 students if 5 students have height 61 in, 18 have height 64 in, etc.

The computations involved can become tedious, especially for cases in which the numbers are large and many classes are present. Short techniques are available for lessening the labor in such cases; for example, see Problems 3.20 and 3.22.

Table 3.1

Height (in)	Class Mark (X)	Frequency (f)	fX
60–62	61	5	305
63–65	64	18	1152
66–68	67	42	2814
69–71	70	27	1890
72–74	73	8	584
		$N = \sum f = 100$	$\sum fX = 6745$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{\sum fX}{N} = \frac{6745}{100} = 67.45 \text{ in}$$

PROPERTIES OF THE ARITHMETIC MEAN

3.16 Prove that the sum of the deviations of $X_1, X_2, \ldots, X_N$ from their mean $\bar{X}$ is equal to zero.

SOLUTION

Let $d_1 = X_1 - \bar{X}$, $d_2 = X_2 - \bar{X}, \ldots, d_N = X_N - \bar{X}$ be the deviations of $X_1, X_2, \ldots, X_N$ from their mean $\bar{X}$. Then

$$\text{Sum of deviations} = \sum d_j = \sum (X_j - \bar{X}) = \sum X_j - N\bar{X}$$
$$= \sum X_j - N\left(\frac{\sum X_j}{N}\right) = \sum X_j - \sum X_j = 0$$

where we have used $\sum$ in place of $\sum_{j=1}^{N}$. We could, if desired, have omitted the subscript j in X_j, provided that it is *understood*.

3.17 If $Z_1 = X_1 + Y_1$, $Z_2 = X_2 + Y_2, \ldots, Z_N = X_N + Y_N$, prove that $\bar{Z} = \bar{X} + \bar{Y}$.

SOLUTION

By definition,

$$\bar{X} = \frac{\sum X}{N} \qquad \bar{Y} = \frac{\sum Y}{N} \qquad \bar{Z} = \frac{\sum Z}{N}$$

Thus
$$\bar{Z} = \frac{\sum Z}{N} = \frac{\sum (X+Y)}{N} = \frac{\sum X + \sum Y}{N} = \frac{\sum X}{N} + \frac{\sum Y}{N} = \bar{X} + \bar{Y}$$

where the subscripts j on X, Y, and Z have been omitted and where $\sum$ means $\sum_{j=1}^{N}$.

3.18 (*a*) If N numbers $X_1, X_2, \ldots, X_N$ have deviations from any number A given by $d_1 = X_1 - A$, $d_2 = X_2 - A, \ldots, d_N = X_N - A$, respectively, prove that

$$\bar{X} = A + \frac{\sum_{j=1}^{N} d_j}{N} = A + \frac{\sum d}{N}$$

(*b*) In case $X_1, X_2, \ldots, X_K$ have respective frequencies $f_1, f_2, \ldots, f_K$ and $d_1 = X_1 - A, \ldots, d_K = X_K - A$, show that the result in part (*a*) is replaced with

$$\bar{X} = A + \frac{\sum_{j=1}^{K} f_j d_j}{\sum_{j=1}^{K} f_j} = A + \frac{\sum fd}{N} \qquad \text{where} \qquad \sum_{j=1}^{K} f_j = \sum f = N$$

SOLUTION

(*a*) **First method**

Since $d_j = X_j - A$ and $X_j = A + d_j$, we have

$$\bar{X} = \frac{\sum X_j}{N} = \frac{\sum (A + d_j)}{N} = \frac{\sum A + \sum d_j}{N} = \frac{NA + \sum d_j}{N} = A + \frac{\sum d_j}{N}$$

where we have used $\sum$ in place of $\sum_{j=1}^{N}$ for brevity.

Second method

We have $d = X - A$, or $X = A + d$, omitting the subscripts on d and X. Thus, by Problem 3.17,

$$\bar{X} = \bar{A} + \bar{d} = A + \frac{\sum d}{N}$$

since the mean of a number of constants all equal to A is A.

(*b*)
$$\bar{X} = \frac{\sum_{j=1}^{K} f_j X_j}{\sum_{j=1}^{K} f_j} = \frac{\sum f_j X_j}{N} = \frac{\sum f_j (A + d_j)}{N} = \frac{\sum A f_j + \sum f_j d_j}{N} = \frac{A \sum f_j + \sum f_j d_j}{N}$$

$$= \frac{AN + \sum f_j d_j}{N} = A + \frac{\sum f_j d_j}{N} = A + \frac{\sum fd}{N}$$

Note that *formally* the result is obtained from part (*a*) by replacing d_j with $f_j\,d_j$ and summing from $j = 1$ to K instead of from $j = 1$ to N. The result is equivalent to $\bar{X} = A + \bar{d}$, where $\bar{d} = (\sum fd)/N$.

THE ARITHMETIC MEAN COMPUTED FROM GROUPED DATA

3.19 Use the method of Problem 3.18(*a*) to find the arithmetic mean of the numbers 5, 8, 11, 9, 12, 6, 14, and 10, choosing as the "guessed mean" A the values (*a*) 9 and (*b*) 20.

SOLUTION

(*a*) The deviations of the given numbers from 9 are -4, -1, 2, 0, 3, -3, 5, and 1, and the sum of the deviations is $\sum d = -4 - 1 + 2 + 0 + 3 - 3 + 5 + 1 = 3$. Thus

$$\bar{X} = A + \frac{\sum d}{N} = 9 + \frac{3}{8} = 9.375$$

(*b*) The deviations of the given numbers from 20 are -15, -12, -9, -11, -8, -14, -6, and -10, and $\sum d = -85$. Thus

$$\bar{X} = A + \frac{\sum d}{N} = 20 + \frac{(-85)}{8} = 9.375$$

3.20 Use the method of Problem 3.18(*b*) to find the arithmetic mean of the heights of the 100 male students at XYZ University (see Problem 3.15).

SOLUTION

The work may be arranged as in Table 3.2. We take the guessed mean A to be the class mark 67 (which has the largest frequency), although any class mark can be used for A. Note that the computations are simpler than those in Problem 3.15. To shorten the labor even more, we can proceed as in Problem 3.22, where use is made of the fact that the deviations (column 2 of Table 3.2) are all integer multiples of the class-interval size.

Table 3.2

Class Mark (X)	Deviation $d = X - A$	Frequency (f)	fd
61	-6	5	-30
64	-3	18	-54
$A \longrightarrow 67$	0	42	0
70	3	27	81
73	6	8	48
		$N = \sum f = 100$	$\sum fd = 45$

$$\bar{X} = A + \frac{\sum fd}{N} = 67 + \frac{45}{100} = 67.45 \text{ in}$$

3.21 Let $d_j = X_j - A$ denote the deviations of any class mark X_j in a frequency distribution from a given class mark A. Show that if all class intervals have equal size c, then (*a*) the deviations are all multiples of c (i.e., $d_j = cu_j$, where $u_j = 0, \pm 1, \pm 2, \ldots$) and (*b*) the arithmetic mean can be computed from the formula

$$\bar{X} = A + \left(\frac{\sum fu}{N}\right)c$$

SOLUTION

(*a*) The result is illustrated in Table 3.2 of Problem 3.20, where it is observed that the deviations in column 2 are all multiples of the class-interval size $c = 3$ in.

To see that the result is true in general, note that if $X_1, X_2, X_3, \ldots$ are successive class marks, their common difference will for this case be equal to c, so that $X_2 = X_1 + c$, $X_3 = X_1 + 2c$, and in general $X_j = X_1 + (j-1)c$. Then any two class marks X_p and X_q, for example, will differ by

$$X_p - X_q = [X_1 + (p-1)c] - [X_1 + (q-1)c] = (p-q)c$$

which is a multiple of c.

(*b*) By part (*a*), the deviations of all the class marks from any given one are multiples of c (i.e., $d_j = cu_j$). Then, using Problem 3.18(*b*), we have

$$\bar{X} = A + \frac{\sum f_j d_j}{N} = A + \frac{\sum f_j(cu_j)}{N} = A + c\frac{\sum f_j u_j}{N} = A + \left(\frac{\sum fu}{N}\right)c$$

Note that this is equivalent to the result $\bar{X} = A + c\bar{u}$, which can be obtained from $\bar{X} = A + \bar{d}$ by placing $d = cu$ and observing that $\bar{d} = c\bar{u}$ (see Problem 3.18).

3.22 Use the result of Problem 3.21(*b*) to find the mean height of the 100 male students at XYZ University (see Problem 3.20).

SOLUTION

The work may be arranged as in Table 3.3. The method is called the *coding method* and should be employed whenever possible.

Table 3.3

	X	u	f	fu
	61	−2	5	−10
	64	−1	18	−18
$A \longrightarrow$	67	0	42	0
	70	1	27	27
	73	2	8	16
			$N = 100$	$\sum fu = 15$

$$\bar{X} = A + \left(\frac{\sum fu}{N}\right)c = 67 + \left(\frac{15}{100}\right)(3) = 67.45\,\text{in}$$

3.23 Compute the mean weekly wage of the 65 employees at the P&R Company from the frequency distribution in Table 2.5, using (*a*) the long method and (*b*) the coding method.

SOLUTION

Tables 3.4 and 3.5 show the solutions to (*a*) and (*b*), respectively.

Table 3.4

X	f	fX
\$255.00	8	\$2040.00
265.00	10	2650.00
275.00	16	4400.00
285.00	14	3990.00
295.00	10	2950.00
305.00	5	1525.00
315.00	2	630.00
	$N = 65$	$\sum fX = \$18{,}185.00$

Table 3.5

	X	u	f	fu
	\$255.00	−2	8	−16
	265.00	−1	10	−10
$A \longrightarrow$	275.00	0	16	0
	285.00	1	14	14
	295.00	2	10	20
	305.00	3	5	15
	315.00	4	2	8
			$N = 65$	$\sum fu = 31$

It might be supposed that error would be introduced into these tables since the class marks are actually \$254.995, \$264.995, etc., instead of \$255.00, \$265.00, etc. If in Table 3.4 these true class marks are used instead, then $\bar{X}$ turns out to be \$279.76 instead of \$279.77, and the difference is negligible.

$$\bar{X} = \frac{\sum fX}{N} = \frac{\$18{,}185.00}{65} = \$279.77 \qquad \bar{X} = A + \left(\frac{\sum fu}{N}\right)c = \$275.00 + \frac{31}{65}(\$10.00) = \$279.77$$

3.24 Using Table 2.9(*d*), find the mean wage of the 70 employees at the P&R Company.

SOLUTION

In this case the class intervals do not have equal size and we must use the long method, as shown in Table 3.6.

Table 3.6

X	f	fX
\$255.00	8	\$2040.00
265.00	10	2650.00
275.00	16	4400.00
285.00	15	4275.00
295.00	10	2950.00
310.00	8	2480.00
350.00	3	1050.00
	$N = 70$	$\sum fX = \$19{,}845.00$

$$\bar{X} = \frac{\sum fX}{N} = \frac{\$19{,}845.00}{70} = \$283.50$$

THE MEDIAN

3.25 The following MINITAB output shows the time spent per week searching on line for 30 Internet users as well as the median of the 30 times. Verify the median. Would you say this average is typical of the 30 times? Compare your results with those found in Problem 3.8.

```
MTB > print c1

Data Display

time
3   4   4   5   5   5   5   5   5   6
6   6   6   7   7   7   7   7   8   8
9  10  10  10  10  10  10  12  55  60

MTB > median c1

Column Median

Median of time = 7.0000
```

SOLUTION

Note that the two middle values are both 7 and the mean of the two middle values is 7. The mean was found to be 10.4 hours in Problem 3.8. The median is more typical of the times than the mean.

3.26 The number of ATM transactions per day were recorded at 15 locations in a large city. The data were: 35, 49, 225, 50, 30, 65, 40, 55, 52, 76, 48, 325, 47, 32, and 60. Find (*a*) the median number of transactions and (*b*) the mean number of transactions.

SOLUTION

(*a*) Arranged in order, the data are: 30, 32, 35, 40, 47, 48, 49, 50, 52, 55, 60, 65, 76, 225, and 325. Since there is an odd number of items, there is only one middle value, 50, which is the required median.

(*b*) The sum of the 15 values is 1189. The mean is $1189/15 = 79.267$.

Note that the median is not affected by the two extreme values 225 and 325, while the mean is affected by it. In this case, the median gives a better indication of the average number of daily ATM transactions.

3.27 If (*a*) 85 and (*b*) 150 numbers are arranged in an array, how would you find the median of the numbers?

SOLUTION

(*a*) Since there are 85 items, an odd number, there is only one middle value with 42 numbers below and 42 numbers above it. Thus the median is the 43rd number in the array.

(*b*) Since there are 150 items, an even number, there are two middle values with 74 numbers below them and 74 numbers above them. The two middle values are the 75th and 76th numbers in the array, and their arithmetic mean is the required median.

3.28 From Problem 2.8, find the median weight of the 40 male college students at State University by using (*a*) the frequency distribution of Table 2.7 (reproduced here as Table 3.7) and (*b*) the original data.

SOLUTION

(*a*) **First method** (using interpolation)

The weights in the frequency distribution of Table 3.7 are assumed to be continuously distributed. In such case the median is that weight for which half the total frequency ($40/2 = 20$) lies above it and half lies below it.

Table 3.7

Weight (lb)	Frequency
118–126	3
127–135	5
136–144	9
145–153	12
154–162	5
163–171	4
172–180	2
	Total 40

Now the sum of the first three class frequencies is $3 + 5 + 9 = 17$. Thus to give the desired 20, we require three more of the 12 cases in the fourth class. Since the fourth class interval, 145–153, actually corresponds to weights 144.5 to 153.5, the median must lie 3/12 of the way between 144.5 and 153.5; that is, the median is

$$144.5 + \frac{3}{12}(153.5 - 144.5) = 144.5 + \frac{3}{12}(9) = 146.8 \text{ lb}$$

Second method (using formula)

Since the sum of the first three and first four class frequencies are $3 + 5 + 9 = 17$ and $3 + 5 + 9 + 12 = 29$, respectively, it is clear that the median lies in the fourth class, which is therefore the median class. Then

$$\begin{aligned}
L_1 &= \text{lower class boundary of median class} = 144.5 \\
N &= \text{number of items in the data} = 40 \\
(\textstyle\sum f)_1 &= \text{sum of all classes lower than the median class} = 3 + 5 + 9 = 17 \\
f_{\text{median}} &= \text{frequency of median class} = 12 \\
c &= \text{size of median class interval} = 9
\end{aligned}$$

and thus

$$\text{Median} = L_1 + \left(\frac{N/2 - (\sum f)_1}{f_{\text{median}}}\right)c = 144.5 + \left(\frac{40/2 - 17}{12}\right)(9) = 146.8 \text{ lb}$$

(*b*) Arranged in an array, the original weights are

119, 125, 126, 128, 132, 135, 135, 135, 136, 138, 138, 140, 140, 142, 142, 144, 144, 145, 145, 146,

146, 147, 147, 148, 149, 150, 150, 152, 153, 154, 156, 157, 158, 161, 163, 164, 165, 168, 173, 176

The median is the arithmetic mean of the 20th and 2lst weights in this array and is equal to 146 lb.

3.29 Figure 3-3 gives the stem-and-leaf display for the number of 2005 alcohol-related traffic deaths for the 50 states and Washington D.C.
Stem-and-Leaf Display: Deaths

Stem–and–Leaf Display: Deaths

Stem–and–leaf of Deaths N = 51
Leaf Unit = 10

14	0	22334556667889
23	1	122255778
(7)	2	0334689
21	3	124679
15	4	22669
10	5	012448
4	6	3
3	7	
3	8	
3	9	
3	10	
3	11	
3	12	
3	13	
3	14	7
2	15	6
1	16	
1	17	1

Fig. 3-3 MINITAB stem-and-leaf display for 2005 alcohol-related traffic deaths.

Find the mean, median, and mode of the alcohol-related deaths in Fig. 3-3.

SOLUTION

The number of deaths range from 20 to 1710. The distribution is bimodal. The two modes are 60 and 120, both of which occur three times.

The class `(7) 2 0334689 is the median class.` That is, the median is contained in this class. The median is the middle of the data or the 26th value in the ordered array. The 24th value is 200, the 25th value is 230, and the 26th value is also 230. Therefore the median is 230.

The sum of the 51 values is 16,660 and the mean is 16,660/51 = 326.67.

3.30 Find the median wage of the 65 employees at the P&R Company (see Problem 2.3).

SOLUTION

Here $N = 65$ and $N/2 = 32.5$. Since the sum of the first two and first three class frequencies are $8 + 10 = 18$ and $8 + 10 + 16 = 34$, respectively, the median class is the third class. Using the formula,

$$\text{Median} = L_1 + \left(\frac{N/2 - (\sum f)_1}{f_{\text{median}}}\right)c = \$269.995 + \left(\frac{32.5 - 18}{16}\right)(\$10.00) = \$279.06$$

THE MODE

3.31 Find the mean, median, and mode for the sets (*a*) 3, 5, 2, 6, 5, 9, 5, 2, 8, 6 and (*b*) 51.6, 48.7, 50.3, 49.5, 48.9.

SOLUTION

(*a*) Arranged in an array, the numbers are 2, 2, 3, 5, 5, 5, 6, 6, 8, and 9.

$$\text{Mean} = \tfrac{1}{10}(2 + 2 + 3 + 5 + 5 + 5 + 6 + 6 + 8 + 9) = 5.1$$

$$\text{Median} = \text{arithmetic mean of two middle numbers} = \tfrac{1}{2}(5 + 5) = 5$$

$$\text{Mode} = \text{number occurring most frequently} = 5$$

(*b*) Arranged in an array, the numbers are 48.7, 48.9, 49.5, 50.3, and 51.6.

$$\text{Mean} = \tfrac{1}{5}(48.7 + 48.9 + 49.5 + 50.3 + 51.6) = 49.8$$

$$\text{Median} = \text{middle number} = 49.5$$

$$\text{Mode} = \text{number occurring most frequently (nonexistent here)}$$

3.32 Suppose you desired to find the mode for the data in Problem 3.29. You could use frequencies procedure of SAS to obtain the following output. By considering the output from FREQ Procedure (Fig. 3-4), what are the modes of the alcohol-linked deaths?

The FREQ Procedure

Deaths

deaths	Frequency	Percent	Cumulative Frequency	Cumulative Percent
20	2	3.92	2	3.92
30	2	3.92	4	7.84
40	1	1.96	5	9.80
50	2	3.92	7	13.73
60	3	5.88	10	19.61
70	1	1.96	11	21.57
80	2	3.92	13	25.49
90	1	1.96	14	27.45
110	1	1.96	15	29.41
120	3	5.88	18	35.29
150	2	3.92	20	39.22
170	2	3.92	22	43.14
180	1	1.96	23	45.10
200	1	1.96	24	47.06
230	2	3.92	26	50.98
240	1	1.96	27	52.94
260	1	1.96	28	54.90
280	1	1.96	29	56.86
290	1	1.96	30	58.82
310	1	1.96	31	60.78
320	1	1.96	32	62.75
340	1	1.96	33	64.71
360	1	1.96	34	66.67
370	1	1.96	35	68.63
390	1	1.96	36	70.59
420	2	3.92	38	74.51
460	2	3.92	40	78.43
490	1	1.96	41	80.39
500	1	1.96	42	82.35
510	1	1.96	43	84.31
520	1	1.96	44	86.27
540	2	3.92	46	90.20
580	1	1.96	47	92.16
630	1	1.96	48	94.12
1470	1	1.96	49	96.08
1560	1	1.96	50	98.04
1710	1	1.96	51	100.00

Fig. 3-4 SAS output for FREQ procedure and the alcohol-related deaths.

SOLUTION

The data is bimodal and the modes are 60 and 120. This is seen by considering the SAS output and noting that both 60 and 120 have a frequency of 3 which is greater than any other of the values.

3.33 Some statistical software packages have a mode routine built in but do not return all modes in the case that the data is multimodal. Consider the output from SPSS in Fig. 3-5.

What does SPSS do when asked to find modes?

Deaths

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20.00	2	3.9	3.9	3.9
	30.00	2	3.9	3.9	7.8
	40.00	1	2.0	2.0	9.8
	50.00	2	3.9	3.9	13.7
	60.00	3	5.9	5.9	19.6
	70.00	1	2.0	2.0	21.6
	80.00	2	3.9	3.9	25.5
	90.00	1	2.0	2.0	27.5
	110.00	1	2.0	2.0	29.4
	120.00	3	5.9	5.9	35.3
	150.00	2	3.9	3.9	39.2
	170.00	2	3.9	3.9	43.1
	180.00	1	2.0	2.0	45.1
	200.00	1	2.0	2.0	47.1
	230.00	2	3.9	3.9	51.0
	240.00	1	2.0	2.0	52.9
	260.00	1	2.0	2.0	54.9
	280.00	1	2.0	2.0	56.9
	290.00	1	2.0	2.0	58.8
	310.00	1	2.0	2.0	60.8
	320.00	1	2.0	2.0	62.7
	340.00	1	2.0	2.0	64.7
	360.00	1	2.0	2.0	66.7
	370.00	1	2.0	2.0	68.6
	390.00	1	2.0	2.0	70.6
	420.00	2	3.9	3.9	74.5
	460.00	2	3.9	3.9	78.4
	490.00	1	2.0	2.0	80.4
	500.00	1	2.0	2.0	82.4
	510.00	1	2.0	2.0	84.3
	520.00	1	2.0	2.0	86.3
	540.00	2	3.9	3.9	90.2
	580.00	1	2.0	2.0	92.2
	630.00	1	2.0	2.0	94.1
	1470.00	1	2.0	2.0	96.1
	1560.00	1	2.0	2.0	98.0
	1710.00	1	2.0	2.0	100.0
	Total	51	100.0	100.0	

Statistics

Deaths

N	Valid	51
	Missing	0
Mode		60.00[a]

[a]Multiple modes exist. The smallest value is shown.

Fig. 3-5 SPSS output for the alcohol-related deaths.

SOLUTION

SPSS prints out the smallest mode. However, it is possible to look at the frequency distribution and find all modes just as in SAS (see the output above).

THE EMPIRICAL RELATION BETWEEN THE MEAN, MEDIAN, AND MODE

3.34 (*a*) Use the empirical formula mean − mode = 3(mean − median) to find the modal wage of the 65 employees at the P&R Company.

(*b*) Compare your result with the mode obtained in Problem 3.33.

SOLUTION

(*a*) From Problems 3.23 and 3.30 we have mean = \$279.77 and median = \$279.06. Thus

$$\text{Mode} = \text{mean} - 3(\text{mean} - \text{median}) = \$279.77 - 3(\$279.77 - \$279.06) = \$277.64$$

(*b*) From Problem 3.33 the modal wage is \$277.50, so there is good agreement with the empirical result in this case.

THE GEOMETRIC MEAN

3.35 Find (*a*) the geometric mean and (*b*) the arithmetic mean of the numbers 3, 5, 6, 6, 7, 10, and 12. Assume that the numbers are exact.

SOLUTION

(*a*) Geometric mean $= G = \sqrt[7]{(3)(5)(6)(6)(7)(10)(12)} = \sqrt[7]{453{,}600}$. Using common logarithms, $\log G = \frac{1}{7}\log 453{,}600 = \frac{1}{7}(5.6567) = 0.8081$, and $G = 6.43$ (to the nearest hundredth). Alternatively, a calculator can be used.

Another method

$$\begin{aligned}\log G &= \tfrac{1}{7}(\log 3 + \log 5 + \log 6 + \log 6 + \log 7 + \log 10 + \log 12)\\ &= \tfrac{1}{7}(0.4771 + 0.6990 + 0.7782 + 0.7782 + 0.8451 + 1.0000 + 1.0792)\\ &= 0.8081\end{aligned}$$

and $$G = 6.43$$

(*b*) Arithmetic mean $= \hat{X} = \frac{1}{7}(3 + 5 + 6 + 6 + 7 + 10 + 12) = 7$. This illustrates the fact that the geometric mean of a set of unequal positive numbers is less than the arithmetic mean.

3.36 The numbers $X_1, X_2, \ldots, X_K$ occur with frequencies $f_1, f_2, \ldots, f_K$, where $f_1 + f_2 + \cdots + f_K = N$ is the total frequency.

(*a*) Find the geometric mean G of the numbers.

(*b*) Derive an expression for $\log G$.

(*c*) How can the results be used to find the geometric mean for data grouped into a frequency distribution?

SOLUTION

(*a*) $$G = \sqrt[N]{\underbrace{X_1 X_1 \cdots X_1}_{f_1 \text{ times}}\ \underbrace{X_2 X_2 \cdots X_2}_{f_2 \text{ times}} \cdots \underbrace{X_K X_K \cdots X_K}_{f_K \text{ times}}} = \sqrt[N]{X_1^{f_1} X_2^{f_2} \cdots X_K^{f_K}}$$

where $N = \sum f$. This is sometimes called the *weighted geometric mean.*

(*b*)
$$\log G = \frac{1}{N}\log(X_1^{f_1}X_2^{f_2}\cdots X_K^{f_K}) = \frac{1}{N}(f_1\log X_1 + f_2\log X_2 + \cdots + f_K\log X_K)$$
$$= \frac{1}{N}\sum_{j=1}^{K} f_j \log X_j = \frac{\sum f \log X}{N}$$

where we assume that all the numbers are positive; otherwise, the logarithms are not defined.

Note that the logarithm of the geometric mean of a set of positive numbers is the arithmetic mean of the logarithms of the numbers.

(*c*) The result can be used to find the geometric mean for grouped data by taking $X_1, X_2, \ldots, X_K$ as class marks and $f_1, f_2, \ldots, f_K$ as the corresponding class frequencies.

3.37 During one year the ratio of milk prices per quart to bread prices per loaf was 3.00, whereas during the next year the ratio was 2.00.

(*a*) Find the arithmetic mean of these ratios for the 2-year period.

(*b*) Find the arithmetic mean of the ratios of bread prices to milk prices for the 2-year period.

(*c*) Discuss the advisability of using the arithmetic mean for averaging ratios.

(*d*) Discuss the suitability of the geometric mean for averaging ratios.

SOLUTION

(*a*)
$$\text{Mean ratio of milk to bread prices} = \tfrac{1}{2}(3.00 + 2.00) = 2.50.$$

(*b*) Since the ratio of milk to bread prices for the first year is 3.00, the ratio of bread to milk prices is $1/3.00 = 0.333$. Similarly, the ratio of bread to milk prices for the second year is $1/2.00 = 0.500$. Thus
$$\text{Mean ratio of bread to milk prices} = \tfrac{1}{2}(0.333 + 0.500) = 0.417$$

(*c*) We would expect the mean ratio of milk to bread prices to be the reciprocal of the mean ratio of bread to milk prices if the mean is an appropriate average. However, $1/0.417 = 2.40 \neq 2.50$. This shows that the arithmetic mean is a poor average to use for ratios.

(*d*)
$$\text{Geometric mean of ratios of milk to bread prices} = \sqrt{(3.00)(2.00)} = \sqrt{6.00}$$
$$\text{Geometric mean of ratios of bread to milk prices} = \sqrt{(0.333)(0.500)} = \sqrt{0.0167} = 1/\sqrt{6.00}$$

Since these averages are reciprocals, our conclusion is that the geometric mean is more suitable than the arithmetic mean for averaging ratios for this type of problem.

3.38 The bacterial count in a certain culture increased from 1000 to 4000 in 3 days. What was the average percentage increase per day?

SOLUTION

Since an increase from 1000 to 4000 is a 300% increase, one might be led to conclude that the average percentage increase per day would be $300\%/3 = 100\%$. This, however, would imply that during the first day the count went from 1000 to 2000, during the second day from 2000 to 4000, and during the third day from 4000 to 8000, which is contrary to the facts.

To determine this average percentage increase, let us denote it by r. Then

$$\text{Total bacterial count after 1 day} = 1000 + 1000r = 1000(1+r)$$
$$\text{Total bacterial count after 2 days} = 1000(1+r) + 1000(1+r)r = 1000(1+r)^2$$
$$\text{Total bacterial count after 3 days} = 1000(1+r)^2 + 1000(1+r)^2 r = 1000(1+r)^3$$

This last expression must equal 4000. Thus $1000(1+r)^3 = 4000$, $(1+r)^3 = 4$, $1 + r = \sqrt[3]{4}$, and $r = \sqrt[3]{4} - 1 = 1.587 - 1 = 0.587$, so that $r = 58.7\%$.

In general, if we start with a quantity P and increase it at a constant rate r per unit of time, we will have after n units of time the amount

$$A = P(1 + r)^n$$

This is called the *compound-interest formula* (see Problems 3.94 and 3.95).

THE HARMONIC MEAN

3.39 Find the harmonic mean H of the numbers 3, 5, 6, 6, 7, 10, and 12.

SOLUTION

$$\frac{1}{H} = \frac{1}{N}\sum\frac{1}{X} = \frac{1}{7}\left(\frac{1}{3}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}+\frac{1}{7}+\frac{1}{10}+\frac{1}{12}\right) = \frac{1}{7}\left(\frac{140+84+70+70+60+42+35}{420}\right)$$
$$= \frac{501}{2940}$$

and $$H = \frac{2940}{501} = 5.87$$

It is often convenient to express the fractions in decimal form first. Thus

$$\frac{1}{H} = \tfrac{1}{7}(0.3333 + 0.2000 + 0.1667 + 0.1667 + 0.1429 + 0.1000 + 0.0833)$$
$$= \frac{1.1929}{7}$$

and $$H = \frac{7}{1.1929} = 5.87$$

Comparison with Problem 3.35 illustrates the fact that the harmonic mean of several positive numbers not all equal is less than their geometric mean, which is in turn less than their arithmetic mean.

3.40 During four successive years, a home owner purchased oil for her furnace at respective costs of \$0.80, \$0.90, \$1.05, and \$1.25 per gallon (gal). What was the average cost of oil over the 4-year period?

SOLUTION

Case 1

Suppose that the home owner purchased the same quantity of oil each year, say 1000 gal. Then

$$\text{Average cost} = \frac{\text{total cost}}{\text{total quantity purchased}} = \frac{\$800 + \$900 + \$1050 + \$1250}{4000\text{ gal}} = \$1.00/\text{gal}$$

This is the same as the arithmetic mean of the costs per gallon; that is, $\frac{1}{4}(\$0.80 + \$0.90 + \$1.05 + \$1.25) = 1.00/\text{gal}$. This result would be the same even if x gallons were used each year.

Case 2

Suppose that the home owner spends the same amount of money each year, say \$1000. Then

$$\text{Average cost} = \frac{\text{total cost}}{\text{total quantity purchased}} = \frac{\$4000}{(1250 + 1111 + 952 + 800)\text{gal}} = \$0.975/\text{gal}$$

This is the same as the harmonic mean of the costs per gallon:

$$\frac{4}{\dfrac{1}{0.80}+\dfrac{1}{0.90}+\dfrac{1}{1.05}+\dfrac{1}{1.25}} = 0.975$$

This result would be the same even if y dollars were spent each year.

Both averaging processes are correct, each average being computed under different prevailing conditions.

It should be noted that in case the number of gallons used changes from one year to another instead of remaining the same, the ordinary arithmetic mean of Case 1 is replaced by a weighted arithmetic mean. Similarly, if the total amount spent changes from one year to another, the ordinary harmonic mean of Case 2 is replaced by a weighted harmonic mean.

3.41 A car travels 25 miles at 25 miles per hour (mi/h), 25 miles at 50 mph, and 25 miles at 75 mph. Find the arithmetic mean of the three velocities and the harmonic mean of the three velocities. Which is correct?

SOLUTION

The average velocity is equal to the distance traveled divided by the total time and is equal to the following:

$$\frac{75}{\left(1+\dfrac{1}{2}+\dfrac{1}{3}\right)} = 40.9 \text{ mi/h}$$

The arithmetic mean of the three velocities is:

$$\frac{25+50+75}{3} = 50 \text{ mi/h}$$

The harmonic mean is found as follows:

$$\frac{1}{H}=\frac{1}{N}\sum\frac{1}{X}=\frac{1}{3}\left(\frac{1}{25}+\frac{1}{50}+\frac{1}{75}\right)=\frac{11}{450} \qquad \text{and} \qquad H=\frac{450}{11}=40.9$$

The harmonic mean is the correct measure of the average velocity.

THE ROOT MEAN SQUARE, OR QUADRATIC MEAN

3.42 Find the quadratic mean of the numbers 3, 5, 6, 6, 7, 10, and 12.

SOLUTION

$$\text{Quadratic mean} = \text{RMS} = \sqrt{\frac{3^2+5^2+6^2+6^2+7^2+10^2+12^2}{7}} = \sqrt{57} = 7.55$$

3.43 Prove that the quadratic mean of two positive unequal numbers, a and b, is greater than their geometric mean.

SOLUTION

We are required to show that $\sqrt{\frac{1}{2}(a^2+b^2)} > \sqrt{ab}$. If this is true, then by squaring both sides, $\frac{1}{2}(a^2+b^2) > ab$, so that $a^2+b^2 > 2ab$, $a^2-2ab+b^2>0$, or $(a-b)^2>0$. But this last inequality is true, since the square of any real number not equal to zero must be positive.

The proof consists in establishing the reversal of the above steps. Thus starting with $(a-b)^2 > 0$, which we know to be true, we can show that $a^2 + b^2 > 2ab$, $\frac{1}{2}(a^2+b^2) > ab$, and finally $\sqrt{\frac{1}{2}(a^2+b^2)} > \sqrt{ab}$, as required.

Note that $\sqrt{\frac{1}{2}(a^2+b^2)} = \sqrt{ab}$ if and only if $a = b$.

QUARTILES, DECILES, AND PERCENTILES

3.44 Find (*a*) the quartiles Q_1, Q_2, and Q_3 and (*b*) the deciles $D_1, D_2, \ldots, D_9$ for the wages of the 65 employees at the P&R Company (see Problem 2.3).

SOLUTION

(*a*) The first quartile Q_1 is that wage obtained by counting $N/4 = 65/4 = 16.25$ of the cases, beginning with the first (lowest) class. Since the first class contains 8 cases, we must take 8.25 $(16.25 - 8)$ of the 10 cases from the second class. Using the method of linear interpolation, we have

$$Q_1 = \$259.995 + \frac{8.25}{10}(\$10.00) = \$268.25$$

The second quartile Q_2 is obtained by counting the first $2N/4 = N/2 = 65/2 = 32.5$ of the cases. Since the first two classes comprise 18 cases, we must take $32.5 - 18 = 14.5$ of the 16 cases from the third class; thus

$$Q_2 = \$269.995 + \frac{14.5}{16}(\$10.00) = \$279.06$$

Note that Q_2 is actually the median.

The third quartile Q_3 is obtained by counting the first $3N/4 = \frac{3}{4}(65) = 48.75$ of the cases. Since the first four classes comprise 48 cases, we must take $48.75 - 48 = 0.75$ of the 10 cases from the fifth class; thus

$$Q_3 = \$289.995 + \frac{0.75}{10}(\$10.00) = \$290.75$$

Hence 25% of the employees earn \$268.25 or less, 50% earn \$279.06 or less, and 75% earn \$290.75 or less.

(*b*) The first, second, ..., ninth deciles are obtained by counting $N/10, 2N/10, \ldots, 9N/10$ of the cases, beginning with the first (lowest) class. Thus

$$D_1 = \$249.995 + \frac{6.5}{8}(\$10.00) = \$258.12 \qquad D_6 = \$279.995 + \frac{5}{14}(\$10.00) = \$283.57$$

$$D_2 = \$259.995 + \frac{5}{10}(\$10.00) = \$265.00 \qquad D_7 = \$279.995 + \frac{11.5}{14}(\$10.00) = \$288.21$$

$$D_3 = \$269.995 + \frac{1.5}{16}(\$10.00) = \$270.94 \qquad D_8 = \$289.995 + \frac{4}{10}(\$10.00) = \$294.00$$

$$D_4 = \$269.995 + \frac{8}{16}(\$10.00) = \$275.00 \qquad D_9 = \$299.995 + \frac{0.5}{5}(\$10.00) = \$301.00$$

$$D_5 = \$269.995 + \frac{14.5}{16}(\$10.00) = \$279.06$$

Hence 10% of the employees earn \$258.12 or less, 20% earn \$265.00 or less, ..., 90% earn \$301.00 or less.

Note that the fifth decile is the median. The second, fourth, sixth, and eighth deciles, which divide the distribution into five equal parts and which are called *quintiles*, are sometimes used in practice.

3.45 Determine (*a*) the 35th and (*b*) the 60th percentiles for the distribution in Problem 3.44.

SOLUTION

(*a*) The 35th percentile, denoted by P_{35}, is obtained by counting the first $35N/100 = 35(65)/100 = 22.75$ of the cases, beginning with the first (lowest) class. Then, as in Problem 3.44,

$$P_{35} = \$269.995 + \frac{4.75}{16}(\$10.00) = \$272.97$$

This means that 35% of the employees earn \$272.97 or less.

(*b*) The 60th percentile is $P_{60} = \$279.995 + \frac{5}{14}(\$10.00) = \$283.57$. Note that this is the same as the sixth decile or third quintile.

3.46 The following EXCEL worksheet is contained in A1:D26. It contains the per capita income for the 50 states. Give the EXCEL commands to find Q_1, Q_2, Q_3, and P_{95}. Also, give the states that are on both sides those quartiles or percentiles.

State	Per capita income	State	Per capita income
Wyoming	36778	Pennsylvania	34897
Montana	29387	Wisconsin	33565
North Dakota	31395	Massachusetts	44289
New Mexico	27664	Missouri	31899
West Virginia	27215	Idaho	28158
Rhode Island	36153	Kentucky	28513
Virginia	38390	Minnesota	37373
South Dakota	31614	Florida	33219
Alabama	29136	South Carolina	28352
Arkansas	26874	New York	40507
Maryland	41760	Indiana	31276
Iowa	32315	Connecticut	47819
Nebraska	33616	Ohio	32478
Hawaii	34539	New Hampshire	38408
Mississippi	25318	Texas	32462
Vermont	33327	Oregon	32103
Maine	31252	New Jersey	43771
Oklahoma	29330	California	37036
Delaware	37065	Colorado	37946
Alaska	35612	North Carolina	30553
Tennessee	31107	Illinois	36120
Kansas	32836	Michigan	33116
Arizona	30267	Washington	35409
Nevada	35883	Georgia	31121
Utah	28061	Louisiana	24820

SOLUTION

		Nearest states
=PERCENTILE(A2:D26,0.25)	\$30338.5	Arizona and North Carolina
=PERCENTILE(A2:D26,0.50)	\$32657	Ohio and Kansas
=PERCENTILE(A2:D26,0.75)	\$36144.75	Illinois and Rhode Island
=PERCENTILE(A2:D26,0.95)	\$42866.05	Maryland and New Jersey

Supplementary Problems

SUMMATION NOTATION

3.47 Write the terms in each of the following indicated sums:

(a) $\sum_{j=1}^{4}(X_j+2)$ (c) $\sum_{j=1}^{3}U_j(U_j+6)$ (e) $\sum_{j=1}^{4}4X_jY_j$

(b) $\sum_{j=1}^{5}f_jX_j^2$ (d) $\sum_{k=1}^{N}(Y_k^2-4)$

3.48 Express each of the following by using the summation notation:

(a) $(X_1+3)^3+(X_2+3)^3+(X_3+3)^3$

(b) $f_1(Y_1-a)^2+f_2(Y_2-a)^2+\cdots+f_{15}(Y_{15}-a)^2$

(c) $(2X_1-3Y_1)+(2X_2-3Y_2)+\cdots+(2X_N-3Y_N)$

(d) $(X_1/Y_1-1)^2+(X_2/Y_2-1)^2+\cdots+(X_8/Y_8-1)^2$

(e) $\dfrac{f_1a_1^2+f_2a_2^2+\cdots+f_{12}a_{12}^2}{f_1+f_2+\cdots+f_{12}}$

3.49 Prove that $\sum_{j=1}^{N}(X_j-1)^2=\sum_{j=1}^{N}X_j^2-2\sum_{j=1}^{N}X_j+N$.

3.50 Prove that $\sum(X+a)(Y+b)=\sum XY+a\sum Y+b\sum X+Nab$, where a and b are constants. What subscript notation is implied?

3.51 Two variables, U and V, assume the values $U_1=3$, $U_2=-2$, $U_3=5$, and $V_1=-4$, $V_2=-1$, $V_3=6$, respectively. Calculate (a) $\sum UV$, (b) $\sum(U+3)(V-4)$, (c) $\sum V^2$, (d) $(\sum U)(\sum V)^2$, (e) $\sum UV^2$, (f) $\sum(U^2-2V^2+2)$, and (g) $\sum(U/V)$.

3.52 Given $\sum_{j=1}^{4}X_j=7$, $\sum_{j=1}^{4}Y_j=-3$, and $\sum_{j=1}^{4}X_jY_j=5$, find (a) $\sum_{j=1}^{4}(2X_j+5Y_j)$ and (b) $\sum_{j=1}^{4}(X_j-3)(2Y_j+1)$.

THE ARITHMETIC MEAN

3.53 A student received grades of 85, 76, 93, 82, and 96 in five subjects. Determine the arithmetic mean of the grades.

3.54 The reaction times of an individual to certain stimuli were measured by a psychologist to be 0.53, 0.46, 0.50, 0.49, 0.52, 0.53, 0.44, and 0.55 seconds respectively. Determine the mean reaction time of the individual to the stimuli.

3.55 A set of numbers consists of six 6's, seven 7's, eight 8's, nine 9's and ten 10's. What is the arithmetic mean of the numbers?

3.56 A student's grades in the laboratory, lecture, and recitation parts of a physics course were 71, 78, and 89, respectively.

(a) If the weights accorded these grades are 2, 4, and 5, respectively, what is an appropriate average grade?

(b) What is the average grade if equal weights are used?

3.57 Three teachers of economics reported mean examination grades of 79, 74, and 82 in their classes, which consisted of 32, 25, and 17 students, respectively. Determine the mean grade for all the classes.

3.58 The mean annual salary paid to all employees in a company is \$36,000. The mean annual salaries paid to male and female employees of the company is \$34,000 and \$40,000 respectively. Determine the percentages of males and females employed by the company.

3.59 Table 3.8 shows the distribution of the maximum loads in short tons (1 short ton = 2000 lb) supported by certain cables produced by a company. Determine the mean maximum loading, using (*a*) the long method and (*b*) the coding method.

Table 3.8

Maximum Load (short tons)	Number of Cables
9.3–9.7	2
9.8–10.2	5
10.3–10.7	12
10.8–11.2	17
11.3–11.7	14
11.8–12.2	6
12.3–12.7	3
12.8–13.2	1
	Total 60

3.60 Find $\bar{X}$ for the data in Table 3.9, using (*a*) the long method and (*b*) the coding method.

Table 3.9

X	462	480	498	516	534	552	570	588	606	624
f	98	75	56	42	30	21	15	11	6	2

3.61 Table 3.10 shows the distribution of the diameters of the heads of rivets manufactured by a company. Compute the mean diameter.

3.62 Compute the mean for the data in Table 3.11.

Table 3.10

Diameter (cm)	Frequency
0.7247–0.7249	2
0.7250–0.7252	6
0.7253–0.7255	8
0.7256–0.7258	15
0.7259–0.7261	42
0.7262–0.7264	68
0.7265–0.7267	49
0.7268–0.7270	25
0.7271–0.7273	18
0.7274–0.7276	12
0.7277–0.7279	4
0.7280–0.7282	1
	Total 250

Table 3.11

Class	Frequency
10 to under 15	3
15 to under 20	7
20 to under 25	16
25 to under 30	12
30 to under 35	9
35 to under 40	5
40 to under 45	2
	Total 54

3.63 Compute the mean TV viewing time for the 400 junior high students per week in Problem 2.20.

3.64 (*a*) Use the frequency distribution obtained in Problem 2.27 to compute the mean diameter of the ball bearings.

(*b*) Compute the mean directly from the raw data and compare it with part (*a*), explaining any discrepancy.

THE MEDIAN

3.65 Find the mean and median of these sets of numbers: (*a*) 5, 4, 8, 3, 7, 2, 9 and (*b*) 18.3, 20.6, 19.3, 22.4, 20.2, 18.8, 19.7, 20.0.

3.66 Find the median grade of Problem 3.53.

3.67 Find the median reaction time of Problem 3.54.

3.68 Find the median of the set of numbers in Problem 3.55.

3.69 Find the median of the maximum loads of the cables in Table 3.8 of Problem 3.59.

3.70 Find the median $\tilde{X}$ for the distribution in Table 3.9 of Problem 3.60.

3.71 Find the median diameter of the rivet heads in Table 3.10 of Problem 3.61.

3.72 Find the median of the distribution in Table 3.11 of Problem 3.62.

3.73 Table 3.12 gives the number of deaths in thousands due to heart disease in 1993. Find the median age for individuals dying from heart disease in 1993.

Table 3.12

Age Group	Thousands of Deaths
Total	743.3
Under 1	0.7
1 to 4	0.3
5 to 14	0.3
15 to 24	1.0
25 to 34	3.5
35 to 44	13.1
45 to 54	32.7
55 to 64	72.0
65 to 74	158.1
75 to 84	234.0
85 and over	227.6

Source: U.S. National Center for Health Statistics, Vital Statistics of the U.S., annual.

3.74 Find the median age for the U.S. using the data for Problem 2.31.

3.75 Find the median TV viewing time for the 400 junior high students in Problem 2.20.

THE MODE

3.76 Find the mean, median, and mode for each set of numbers: (*a*) 7, 4, 10, 9, 15, 12, 7, 9, 7 and (*b*) 8, 11, 4, 3, 2, 5, 10, 6, 4, 1, 10, 8, 12, 6, 5, 7.

3.77 Find the modal grade in Problem 3.53.

3.78 Find the modal reaction time in Problem 3.54.

3.79 Find the mode of the set of numbers in Problem 3.55.

3.80 Find the mode of the maximum loads of the cables of Problem 3.59.

3.81 Find the mode $\hat{X}$ for the distribution in Table 3.9 of Problem 3.60.

3.82 Find the modal diameter of the rivet heads in Table 3.10 of Problem 3.61.

3.83 Find the mode of the distribution in Problem 3.62.

3.84 Find the modal TV viewing time for the 400 junior high students in Problem 2.20.

3.85 (*a*) What is the modal age group in Table 2.15?
(*b*) What is the modal age group in Table 3.12?

3.86 Using formulas (*9*) and (*10*) in this chapter, find the mode of the distributions given in the following Problems. Compare your answers obtained in using the two formulas.

(*a*) Problem 3.59 (*b*) Problem 3.61 (*c*) Problem 3.62 (*d*) Problem 2.20.

3.87 A continuous random variable has probability associated with it that is described by the following probability density function. $f(x) = -0.75x^2 + 1.5x$ for $0 < x < 2$ and $f(x) = 0$ for all other x values. The mode occurs where the function attains a maximum. Use your knowledge of quadratic functions to show that the mode occurs when $x = 1$.

THE GEOMETRIC MEAN

3.88 Find the geometric mean of the numbers (*a*) 4.2 and 16.8 and (*b*) 3.00 and 6.00.

3.89 Find (*a*) the geometric mean G and (*b*) the arithmetic mean $\bar{X}$ of the set 2, 4, 8, 16, 32.

3.90 Find the geometric mean of the sets (*a*) 3, 5, 8, 3, 7, 2 and (*b*) 28.5, 73.6, 47.2, 31.5, 64.8.

3.91 Find the geometric mean for the distributions in (*a*) Problem 3.59 and (*b*) Problem 3.60. Verify that the geometric mean is less than or equal to the arithmetic mean for these cases.

3.92 If the price of a commodity doubles in a period of 4 years, what is the average percentage increase per year?

3.93 In 1980 and 1996 the population of the United States was 226.5 million and 266.0 million, respectively. Using the formula given in Problem 3.38, answer the following:

(*a*) What was the average percentage increase per year?

(*b*) Estimate the population in 1985.

(*c*) If the average percentage increase of population per year from 1996 to 2000 is the same as in part (*a*), what would the population be in 2000?

3.94 A principal of \$1000 is invested at an 8% annual rate of interest. What will the total amount be after 6 years if the original principal is not withdrawn?

3.95 If in Problem 3.94 the interest is compounded quarterly (i.e., there is a 2% increase in the money every 3 months), what will the total amount be after 6 years?

3.96 Find two numbers whose arithmetic mean is 9.0 and whose geometric mean is 7.2.

THE HARMONIC MEAN

3.97 Find the harmonic mean of the numbers (*a*) 2, 3, and 6 and (*b*) 3.2, 5.2, 4.8, 6.1, and 4.2.

3.98 Find the (*a*) arithmetic mean, (*b*) geometric mean, and (*c*) harmonic mean of the numbers 0, 2, 4, and 6.

3.99 If $X_1, X_2, X_3, \ldots$ represent the class marks in a frequency distribution with corresponding class frequencies $f_1, f_2, f_3, \ldots$, respectively, prove that the harmonic mean H of the distribution is given by

$$\frac{1}{H} = \frac{1}{N}\left(\frac{f_1}{X_1} + \frac{f_2}{X_2} + \frac{f_3}{X_3} + \cdots\right) = \frac{1}{N}\sum \frac{f}{X}$$

where $N = f_1 + f_2 + \cdots = \sum f$.

3.100 Use Problem 3.99 to find the harmonic mean of the distributions in (*a*) Problem 3.59 and (*b*) Problem 3.60. Compare with Problem 3.91.

3.101 Cities A, B, and C are equidistant from each other. A motorist travels from A to B at 30 mi/h, from B to C at 40 mi/h, and from C to A at 50 mi/h. Determine his average speed for the entire trip.

3.102 (*a*) An airplane travels distances of d_1, d_2, and d_3 miles at speeds v_1, v_2, and v_3 mi/h, respectively. Show that the average speed is given by V, where

$$\frac{d_1 + d_2 + d_3}{V} = \frac{d_1}{v_1} + \frac{d_2}{v_2} + \frac{d_3}{v_3}$$

This is a weighted harmonic mean.

(*b*) Find V if $d_1 = 2500$, $d_2 = 1200$, $d_3 = 500$, $v_1 = 500$, $v_2 = 400$, and $v_3 = 250$.

3.103 Prove that the geometric mean of two positive numbers a and b is (*a*) less than or equal to the arithmetic mean and (*b*) greater than or equal to the harmonic mean of the numbers. Can you extend the proof to more than two numbers?

THE ROOT MEAN SQUARE, OR QUADRATIC MEAN

3.104 Find the RMS (or quadratic mean) of the numbers (*a*) 11, 23, and 35 and (*b*) 2.7, 3.8, 3.2, and 4.3.

3.105 Show that the RMS of two positive numbers, *a* and *b*, is (*a*) greater than or equal to the arithmetic mean and (*b*) greater than or equal to the harmonic mean. Can you extend the proof to more than two numbers?

3.106 Derive a formula that can be used to find the RMS for grouped data and apply it to one of the frequency distributions already considered.

QUARTILES, DECILES, AND PERCENTILES

3.107 Table 3.13 shows a frequency distribution of grades on a final examination in college algebra. (*a*) Find the quartiles of the distribution, and (*b*) interpret clearly the significance of each.

Table 3.13

Grade	Number of Students
90–100	9
80–89	32
70–79	43
60–69	21
50–59	11
40–49	3
30–39	1
	Total 120

3.108 Find the quartiles Q_1, Q_2, and Q_3 for the distributions in (*a*) Problem 3.59 and (*b*) Problem 3.60. Interpret clearly the significance of each.

3.109 Give six different statistical terms for the balance point or central value of a bell-shaped frequency curve.

3.110 Find (*a*) P_{10}, (*b*) P_{90}, (*c*) P_{25}, and (*d*) P_{75} for the data of Problem 3.59, interpreting clearly the significance of each.

3.111 (*a*) Can all quartiles and deciles be expressed as percentiles? Explain.

(*b*) Can all quantiles be expressed as percentiles? Explain.

3.112 For the data of Problem 3.107, determine (*a*) the lowest grade scored by the top 25% of the class and (*b*) the highest grade scored by the lowest 20% of the class. Interpret your answers in terms of percentiles.

3.113 Interpret the results of Problem 3.107 graphically by using (*a*) a percentage histogram, (*b*) a percentage frequency polygon, and (*c*) a percentage ogive.

3.114 Answer Problem 3.113 for the results of Problem 3.108.

3.115 (*a*) Develop a formula, similar to that of equation (*8*) in this chapter, for computing any percentile from a frequency distribution.

(*b*) Illustrate the use of the formula by applying it to obtain the results of Problem 3.110.