CHAPTER 5

Moments, Skewness, and Kurtosis

MOMENTS

If $X_1, X_2, \ldots, X_N$ are the N values assumed by the variable X, we define the quantity

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

called the rth *moment*. The first moment with $r = 1$ is the arithmetic mean $\bar{X}$.

The rth *moment about the mean* $\bar{X}$ is defined as

$$m_r = \frac{\sum_{j=1}^{N} (X_j - \bar{X})^r}{N} = \frac{\sum (X - \bar{X})^r}{N} = \overline{(X - \bar{X})^r} \tag{2}$$

If $r = 1$, then $m_1 = 0$ (see Problem 3.16). If $r = 2$, then $m_2 = s^2$, the variance.

The rth *moment about any origin* A is defined as

$$m_r' = \frac{\sum_{j=1}^{N} (X_j - A)^r}{N} = \frac{\sum (X - A)^r}{N} = \frac{\sum d^r}{N} = \overline{(X - A)^r} \tag{3}$$

where $d = X - A$ are the deviations of X from A. If $A = 0$, equation (*3*) reduces to equation (*1*). For this reason, equation (*1*) is often called the rth *moment about zero*.

MOMENTS FOR GROUPED DATA

If $X_1, X_2, \ldots, X_K$ occur with frequencies $f_1, f_2, \ldots, f_K$, respectively, the above moments are given by

$$\overline{X^r} = \frac{f_1 X_1^r + f_2 X_2^r + \cdots + f_K X_K^r}{N} = \frac{\sum_{j=1}^{K} f_j X_j^r}{N} = \frac{\sum f X^r}{N} \tag{4}$$

$$m_r = \frac{\sum_{j=1}^{K} f_j (X_j - \bar{X})^r}{N} = \frac{\sum f(X - \bar{X})^r}{N} = \overline{(X - \bar{X})^r} \tag{5}$$

$$m'_r = \frac{\sum_{j=1}^{K} f_j (X_j - A)^r}{N} = \frac{\sum f\,(X - A)^r}{N} = \overline{(X - A)^r} \tag{6}$$

where $N = \sum_{j=1}^{K} f_j = \sum f$. The formulas are suitable for calculating moments from grouped data.

RELATIONS BETWEEN MOMENTS

The following relations exist between moments about the mean m_r and moments about an arbitrary origin m'_r:

$$\begin{aligned} m_2 &= m'_2 - m'^2_1 \\ m_3 &= m'_3 - 3m'_1 m'_2 + 2m'^3_1 \\ m_4 &= m'_4 - 4m'_1 m'_3 + 6m'^2_1 m'_2 - 3m'^4_1 \end{aligned} \tag{7}$$

etc. (see Problem 5.5). Note that $m'_1 = \bar{X} - A$.

COMPUTATION OF MOMENTS FOR GROUPED DATA

The coding method given in previous chapters for computing the mean and standard deviation can also be used to provide a short method for computing moments. This method uses the fact that $X_j = A + cu_j$ (or briefly, $X = A + cu$), so that from equation (6) we have

$$m'_r = c^r \frac{\sum fu^r}{N} = c^r \overline{u^r} \tag{8}$$

which can be used to find m_r by applying equations (7).

CHARLIER'S CHECK AND SHEPPARD'S CORRECTIONS

Charlier's check in computing moments by the coding method uses the identities:

$$\begin{aligned} \sum f(u+1) &= \sum fu + N \\ \sum f(u+1)^2 &= \sum fu^2 + 2\sum fu + N \\ \sum f(u+1)^3 &= \sum fu^3 + 3\sum fu^2 + 3\sum fu + N \\ \sum f(u+1)^4 &= \sum fu^4 + 4\sum fu^3 + 6\sum fu^2 + 4\sum fu + N \end{aligned} \tag{9}$$

Sheppard's corrections for moments are as follows:

$$\text{Corrected } m_2 = m_2 - \tfrac{1}{12}c^2 \qquad \text{Corrected } m_4 = m_4 - \tfrac{1}{2}c^2 m_2 + \tfrac{7}{240}c^4$$

The moments m_1 and m_3 need no correction.

MOMENTS IN DIMENSIONLESS FORM

To avoid particular units, we can define the *dimensionless moments* about the mean as

$$a_r = \frac{m_r}{s^r} = \frac{m_r}{(\sqrt{m_2})^r} = \frac{m_r}{\sqrt{m_2^r}} \tag{10}$$

where $s = \sqrt{m_2}$ is the standard deviation. Since $m_1 = 0$ and $m_2 = s^2$, we have $a_1 = 0$ and $a_2 = 1$.

SKEWNESS

Skewness is the degree of asymmetry, or departure from symmetry, of a distribution. If the frequency curve (smoothed frequency polygon) of a distribution has a longer tail to the right of the central maximum than to the left, the distribution is said to be *skewed to the right*, or to have *positive skewness*. If the reverse is true, it is said to be *skewed to the left*, or to have *negative skewness*.

For skewed distributions, the mean tends to lie on the same side of the mode as the longer tail (see Figs. 3-1 and 3-2). Thus a measure of the asymmetry is supplied by the difference: mean–mode. This can be made dimensionless if we divide it by a measure of dispersion, such as the standard deviation, leading to the definition

$$\text{Skewness} = \frac{\text{mean} - \text{mode}}{\text{standard deviation}} = \frac{\bar{X} - \text{mode}}{s} \tag{11}$$

To avoid using the mode, we can employ the empirical formula (*10*) of Chapter 3 and define

$$\text{Skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}} = \frac{3(\bar{X} - \text{median})}{s} \tag{12}$$

Equations (*11*) and (*12*) are called, respectively, *Pearson's first and second coefficients of skewness*.

Other measures of skewness, defined in terms of quartiles and percentiles, are as follows:

$$\text{Quartile coefficient of skewness} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \tag{13}$$

$$\text{10–90 percentile coefficient of skewness} = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{10}} = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} \tag{14}$$

An important measure of skewness uses the third moment about the mean expressed in dimensionless form and is given by

$$\text{Moment coefficient of skewness} = a_3 = \frac{m_3}{s^3} = \frac{m_3}{(\sqrt{m_2})^3} = \frac{m_3}{\sqrt{m_2^3}} \tag{15}$$

Another measure of skewness is sometimes given by $b_1 = a_3^2$. For perfectly symmetrical curves, such as the normal curve, a_3 and b_1 are zero.

KURTOSIS

Kurtosis is the degree of peakedness of a distribution, usually taken relative to a normal distribution. A distribution having a relatively high peak is called *leptokurtic*, while one which is flat-topped is called *platykurtic*. A normal distribution, which is not very peaked or very flat-topped, is called *mesokurtic*.

One measure of kurtosis uses the fourth moment about the mean expressed in dimensionless form and is given by

$$\text{Moment coefficient of kurtosis} = a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2} \tag{16}$$

which is often denoted by b_2. For the normal distribution, $b_2 = a_4 = 3$. For this reason, the kurtosis is sometimes defined by $(b_2 - 3)$, which is positive for a leptokurtic distribution, negative for a platykurtic distribution, and zero for the normal distribution.

Another measure of kurtosis is based on both quartiles and percentiles and is given by

$$\kappa = \frac{Q}{P_{90} - P_{10}} \tag{17}$$

where $Q = \frac{1}{2}(Q_3 - Q_1)$ is the semi-interquartile range. We refer to κ (the lowercase Greek letter *kappa*) as the *percentile coefficient of kurtosis*; for the normal distribution, κ has the value 0.263.

POPULATION MOMENTS, SKEWNESS, AND KURTOSIS

When it is necessary to distinguish a sample's moments, measures of skewness, and measures of kurtosis from those corresponding to a population of which the sample is a part, it is often the custom to use Latin symbols for the former and Greek symbols for the latter. Thus if the sample's moments are denoted by m_r and m_r', the corresponding Greek symbols would be μ_r and μ_r' (μ is the Greek letter *mu*). Subscripts are always denoted by Latin symbols.

Similarly, if the sample's measures of skewness and kurtosis are denoted by a_3 and a_4, respectively, the population's skewness and kurtosis would be α_3 and α_4 (α is the Greek letter *alpha*).

We already know from Chapter 4 that the standard deviation of a sample and of a population are denoted by s and σ, respectively.

SOFTWARE COMPUTATION OF SKEWNESS AND KURTOSIS

The software that we have discussed in the text so far may be used to compute skewness and kurtosis measures for sample data. The data in Table 5.1 samples of size 50 from a normal distribution, a skewed-right distribution, a skewed-left distribution, and a uniform distribution.

The normal data are female height measurements, the skewed-right data are age at marriage for females, the skewed-left data are obituary data that give the age at death for females, and the uniform data are the amount of cola put into a 12 ounce container by a soft drinks machine. The distribution of

Table 5.1

Normal		Skewed-right		Skewed-left		Uniform	
67	69	31	40	102	87	12.1	11.6
70	62	43	24	55	104	12.1	11.6
63	67	30	29	70	75	12.4	12.0
65	59	30	24	95	80	12.1	11.6
68	66	38	27	73	66	12.1	11.6
60	65	26	35	79	93	12.2	11.7
70	63	29	33	60	90	12.2	12.3
64	65	55	75	73	84	12.2	11.7
69	60	46	38	89	73	11.9	11.7
61	67	26	34	85	98	12.2	11.7
66	64	29	85	72	79	12.3	11.8
65	68	57	29	92	35	12.3	12.5
71	61	34	40	76	71	11.7	11.8
62	69	34	41	93	90	12.3	11.8
66	65	36	35	76	71	12.3	11.8
68	62	40	26	97	63	12.4	11.9
64	67	28	34	10	58	12.4	11.9
67	70	26	19	70	82	12.1	11.9
62	64	66	23	85	72	12.4	12.2
66	63	63	28	25	93	12.4	11.9
65	68	30	26	83	44	12.5	12.0
63	64	33	31	58	65	11.8	11.9
66	65	24	25	10	77	12.5	12.0
65	61	35	22	92	81	12.5	12.0
63	66	34	28	82	77	12.5	12.0

the four sets of sample data is shown in Fig. 5-1. The distributions of the four samples are illustrated by dotplots.

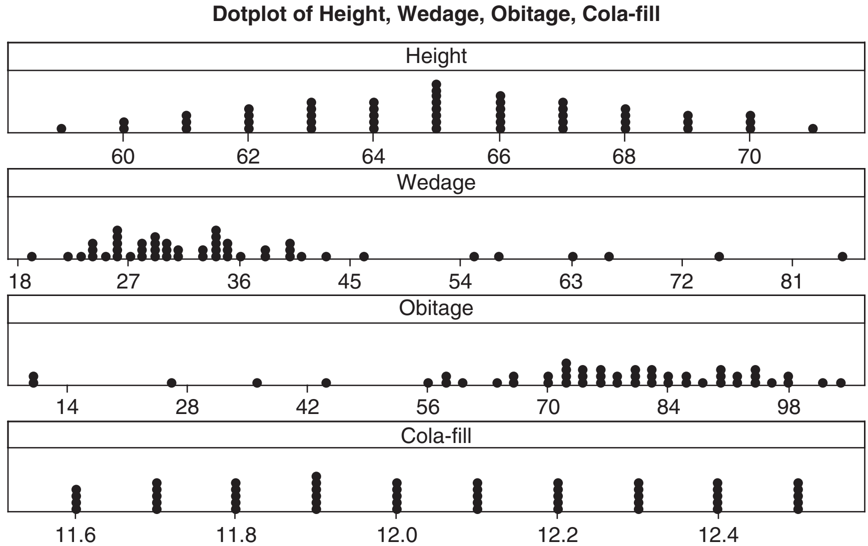

Fig. 5-1 MINITAB plot of four distributions: normal, right-skewed, left-skewed, and uniform.

The variable Height is the height of 50 adult females, the variable Wedage is the age at the wedding of 50 females, the variable Obitage is the age at death of 50 females, and the variable Cola-fill is the amount of cola put into 12 ounce containers. Each sample is of size 50. Using the terminology that we have learned in this chapter: the height distribution is mesokurtic, the cola-fill distribution is platykurtic, the wedage distribution is skewed to the right, and the obitage distribution is skewed to the left.

EXAMPLE 1. If MINITAB is used to find skewness and kurtosis values for the four variables, the following results are obtained by using the pull-down menu "**Stat ⇒ Basic statistics ⇒ Display descriptive statistics.**"

Descriptive Statistics: Height, Wedage, Obitage, Cola-fill

Variable	N	N*	Mean	StDev	Skewness	Kurtosis
Height	50	0	65.120	2.911	−0.02	−0.61
Wedage	50	0	35.48	13.51	1.98	4.10
Obitage	50	0	74.20	20.70	−1.50	2.64
Cola-fill	50	0	12.056	0.284	0.02	−1.19

The skewness values for the uniform and normal distributions are seen to be near 0. The skewness measure is positive for a distribution that is skewed to the right and negative for a distribution that is skewed to the left.

EXAMPLE 2. Use EXCEL to find the skewness and kurtosis values for the data in Fig. 5-1. If the variable names are entered into A1:D1 and the sample data are entered into A2:D51 and any open cell is used to enter =SKEW(A2:A51) the value −0.0203 is returned. The function =SKEW(B2:B51) returns 1.9774, the function =SKEW(C2:C51) returns −1.4986, and =SKEW(D2:D51) returns 0.0156. The kurtosis values are obtained by =KURT(A2:A51) which gives −0.6083, =KURT(B2:B51) which gives 4.0985, =KURT(C2:C51) which gives 2.6368, and =KURT(D2:D51) which gives −1.1889. It can be seen that MINITAB and EXCEL give the same values for kurtosis and skewness.

EXAMPLE 3. When STATISTIX is used to analyze the data shown in Fig. 5-1, the pull-down "**Statistics ⇒ Summary Statistics ⇒ Descriptive Statistics**" gives the dialog box in Fig. 5-2.

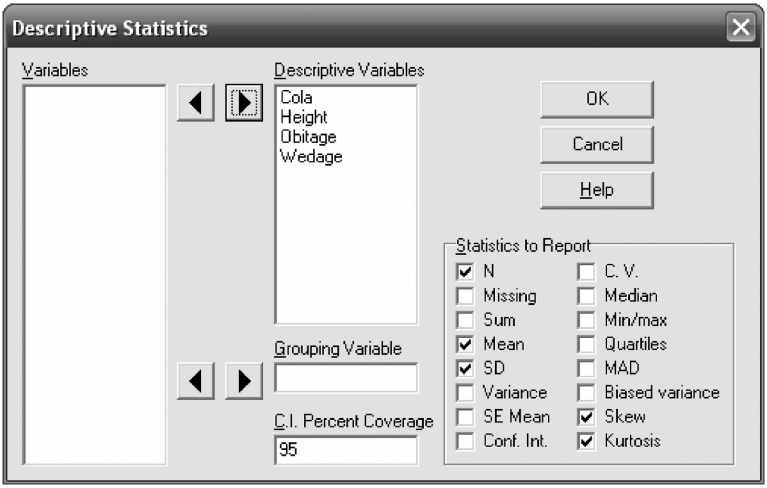

Fig. 5-2 Dialogue box for STATISTIX.

Note that N, Mean, SD, Skew, and Kurtosis are checked as the statistics to report. The STATISTIX output is as follows.

Descriptive Statistics

Variable	N	Mean	SD	Skew	Kurtosis
Cola	50	12.056	0.2837	0.0151	−1.1910
Height	50	65.120	2.9112	−0.0197	−0.6668
Obitage	50	74.200	20.696	−1.4533	2.2628
Wedage	50	35.480	13.511	1.9176	3.5823

Since the numerical values differ slightly from MINITAB and EXCEL, it is obvious that slightly different measures of skewness and kurtosis are used by the software.

EXAMPLE 4. The SPSS pull-down menu "**Analyze ⇒ Descriptive Statistics ⇒ Descriptives**" gives the dialog box in Fig. 5-3 and the routines Mean, Std. deviation, Kurtosis, and Skewness are chosen. SPSS gives the same measures of skewness and kurtosis as EXCEL and MINITAB.

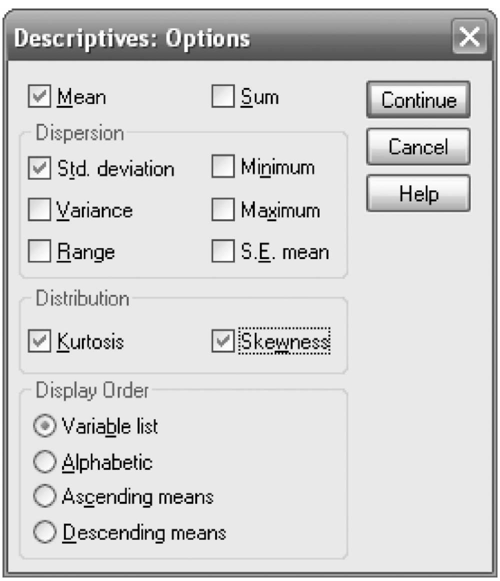

Fig. 5-3 Dialogue box for SPSS.

The following SPSS output is given.

Descriptive Statistics

	N	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Height	50	65.1200	2.91120	−.020	.337	−.608	.662
Wedage	50	35.4800	13.51075	1.977	.337	4.098	.662
Obitage	50	74.2000	20.69605	−1.499	.337	2.637	.662
Colafill	50	12.0560	.28368	.016	.337	−1.189	.662
Valid N (listwise)	50						

EXAMPLE 5. When SAS is used to compute the skewness and kurtosis values, the following output results. It is basically the same as that given by EXCEL, MINITAB, and SPSS.

```
                         The MEANS Procedure
 Variable       Mean           Std Dev        N    Skewness      Kurtosis
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
 Height       65.1200000     2.9112029     50   -0.0203232    -0.6083437
 Wedage       35.4800000    13.5107516     50    1.9774237     4.0984607
 Obitage      74.2000000    20.6960511     50   -1.4986145     2.6368045
 Cola_fill    12.0560000     0.2836785     50    0.0156088    -1.1889600
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
```


Solved Problems

MOMENTS

5.1 Find the (*a*) first, (*b*) second, (*c*) third, and (*d*) fourth moments of the set 2, 3, 7, 8, 10.

SOLUTION

(*a*) The first moment, or arithmetic mean, is

$$\bar{X} = \frac{\sum X}{N} = \frac{2+3+7+8+10}{5} = \frac{30}{5} = 6$$

(*b*) The second moment is

$$\overline{X^2} = \frac{\sum X^2}{N} = \frac{2^2+3^2+7^2+8^2+10^2}{5} = \frac{226}{5} = 45.2$$

(*c*) The third moment is

$$\overline{X^3} = \frac{\sum X^3}{N} = \frac{2^3+3^3+7^3+8^3+10^3}{5} = \frac{1890}{5} = 378$$

(*d*) The fourth moment is

$$\overline{X^4} = \frac{\sum X^4}{N} = \frac{2^4+3^4+7^4+8^4+10^4}{5} = \frac{16{,}594}{5} = 3318.8$$

5.2 Find the (*a*) first, (*b*) second, (*c*) third, and (*d*) fourth moments about the mean for the set of numbers in Problem 5.1.

SOLUTION

(*a*) $m_1 = \overline{(X - \bar{X})} = \dfrac{\sum (X - \bar{X})}{N} = \dfrac{(2-6)+(3-6)+(7-6)+(8-6)+(10-6)}{5} = \dfrac{0}{5} = 0$

m_1 is always equal to zero since $\overline{X - \bar{X}} = \bar{X} - \bar{X} = 0$ (see Problem 3.16).

(*b*) $m_2 = \overline{(X - \bar{X})^2} = \dfrac{\sum (X - \bar{X})^2}{N} = \dfrac{(2-6)^2+(3-6)^2+(7-6)^2+(8-6)^2+(10-6)^2}{6} = \dfrac{46}{5} = 9.2$

Note that m_2 is the variance s^2.

(*c*) $m_3 = \overline{(X - \bar{X})^3} = \dfrac{\sum (X - \bar{X})^3}{N} = \dfrac{(2-6)^3+(3-6)^3+(7-6)^3+(8-6)^3+(10-6)^3}{5} = \dfrac{-18}{5} = -3.6$

(*d*) $m_4 = \overline{(X - \bar{X})^4} = \dfrac{\sum (X - \bar{X})^4}{N} = \dfrac{(2-6)^4+(3-6)^4+(7-6)^4+(8-6)^4+(10-6)^4}{5} = \dfrac{610}{5} = 122$

5.3 Find the (*a*) first, (*b*) second, (*c*) third, and (*d*) fourth moments about the origin 4 for the set of numbers in Problem 5.1.

SOLUTION

(*a*) $m_1' = \overline{(X - 4)} = \dfrac{\sum (X - 4)}{N} = \dfrac{(2-4)+(3-4)+(7-4)+(8-4)+(10-4)}{5} = 2$

(*b*) $m_2' = \overline{(X - 4)^2} = \dfrac{\sum (X - 4)^2}{N} = \dfrac{(2-4)^2+(3-4)^2+(7-4)^2+(8-4)^2+(10-4)^2}{5} = \dfrac{66}{5} = 13.2$

(*c*) $m_3' = \overline{(X - 4)^3} = \dfrac{\sum (X - 4)^3}{N} = \dfrac{(2-4)^3+(3-4)^3+(7-4)^3+(8-4)^3+(10-4)^3}{5} = \dfrac{298}{5} = 59.6$

(*d*) $m_4' = \overline{(X - 4)^4} = \dfrac{\sum (X - 4)^4}{N} = \dfrac{(2-4)^4+(3-4)^4+(7-4)^4+(8-4)^4+(10-4)^4}{5} = \dfrac{1650}{5} = 330$

5.4 Using the results of Problems 5.2 and 5.3, verify the relations between the moments (*a*) $m_2 = m_2' - m_1'^2$, (*b*) $m_3 = m_3' - 3m_1'm_2' + 2m_1'^3$, and (*c*) $m_4 = m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4$.

SOLUTION

From Problem 5.3 we have $m_1' = 2$, $m_2' = 13.2$, $m_3' = 59.6$, and $m_4' = 330$. Thus:

(*a*) $m_2 = m_2' - m_1'^2 = 13.2 - (2)^2 = 13.2 - 4 = 9.2$

(*b*) $m_3 = m_3' - 3m_1'm_2' + 2m_1'^3 = 59.6 - (3)(2)(13.2) + 2(2)^3 = 59.6 - 79.2 + 16 = -3.6$

(*c*) $m_4 = m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4 = 330 - 4(2)(59.6) + 6(2)^2(13.2) - 3(2)^4 = 122$

in agreement with Problem 5.2.

5.5 Prove that (*a*) $m_2 = m_2' - m_1'^2$, (*b*) $m_3 = m_3' - 3m_1'm_2' + 2m_1'^3$, and (*c*) $m_4 = m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4$.

SOLUTION

If $d = X - A$, then $X = A + d$, $\bar{X} = A + \bar{d}$, and $X - \bar{X} = d - \bar{d}$. Thus:

(*a*)
$$\begin{aligned} m_2 &= \overline{(X - \bar{X})^2} = \overline{(d - \bar{d})^2} = \overline{d^2 - 2\bar{d}d + \bar{d}^2} \\ &= \overline{d^2} - 2\bar{d}^2 + \bar{d}^2 = \overline{d^2} - \bar{d}^2 = m_2' - m_1'^2 \end{aligned}$$

(b) $$m_3 = \overline{(X - \bar{X})^3} = \overline{(d - \bar{d})^3} = \overline{(d^3 - 3d^2\bar{d} + 3d\bar{d}^2 - \bar{d}^3)}$$
$$= \overline{d^3} - 3\bar{d}\overline{d^2} + 3\bar{d}^3 - \bar{d}^3 = \overline{d^3} - 3\bar{d}\overline{d^2} + 2\bar{d}^3 = m_3' - 3m_1'm_2' + 2m_1'^3$$

(c) $$m_4 = \overline{(X - \bar{X})^4} = \overline{(d - \bar{d})^4} = \overline{(d^4 - 4d^3\bar{d} + 6d^2\bar{d}^2 - 4d\bar{d}^3 + \bar{d}^4)}$$
$$= \overline{d^4} - 4\bar{d}\overline{d^3} + 6\bar{d}^2\overline{d^2} - 4\bar{d}^4 + \bar{d}^4 = \overline{d^4} - 4\bar{d}\overline{d^3} + 6\bar{d}^2\overline{d^2} - 3\bar{d}^4$$
$$= m_4' - 4m_1'm_3' + 6m_1'^2 m_2' - 3m_1'^4$$

By extension of this method, we can derive similar results for m_5, m_6, etc.

COMPUTATION OF MOMENTS FROM GROUPED DATA

5.6 Find the first four moments about the mean for the height distribution of Problem 3.22.

SOLUTION

The work can be arranged as in Table 5.2, from which we have

$$m_1' = c\frac{\sum fu}{N} = (3)\left(\frac{15}{100}\right) = 0.45 \qquad m_3' = c^3\frac{\sum fu^3}{N} = (3)^3\left(\frac{33}{100}\right) = 8.91$$

$$m_2' = c^2\frac{\sum fu^2}{N} = (3)^2\left(\frac{97}{100}\right) = 8.73 \qquad m_4' = c^4\frac{\sum fu^4}{N} = (3)^4\left(\frac{253}{100}\right) = 204.93$$

Thus
$$m_1 = 0$$
$$m_2 = m_2' - m_1'^2 = 8.73 - (0.45)^2 = 8.5275$$
$$m_3 = m_3' - 3m_1'm_2' + m_1'^3 = 8.91 - 3(0.45)(8.73) + 2(0.45)^3 = -2.6932$$
$$m_4 = m_4' - 4m_1'm_3' + 6m_1'^2 m_2' - 3m_1'^4$$
$$= 204.93 - 4(0.45)(8.91) + 6(0.45)^2(8.73) - 3(0.45)^4 = 199.3759$$

Table 5.2

X	u	f	fu	fu^2	fu^3	fu^4
61	-2	5	-10	20	-40	80
64	-1	18	-18	18	-18	18
67	0	42	0	0	0	0
70	1	27	27	27	27	27
73	2	8	16	32	64	128
		$N = \sum f = 10$	$\sum fu = 15$	$\sum fu^2 = 97$	$\sum fu^3 = 33$	$\sum fu^4 = 253$

5.7 Find (a) m_1', (b) m_2', (c) m_3', (d) m_4', (e) m_1, (f) m_2, (g) m_3, (h) m_4, (i) $\bar{X}$, (j) s, (k) $\overline{X^2}$, and (l) $\overline{X^3}$ for the distribution in Table 4.7 of Problem 4.19.

SOLUTION

The work can be arranged as in Table 5.3.

Table 5.3

X	u	f	fu	fu^2	fu^3	fu^4
70	-6	4	-24	144	-864	5184
74	-5	9	-45	225	-1125	5625
78	-4	16	-64	256	-1024	4096
82	-3	28	-84	252	-756	2268
86	-2	45	-90	180	-360	720
90	-1	66	-66	66	-66	66
$A \longrightarrow$ 94	0	85	0	0	0	0
98	1	72	72	72	72	72
102	2	54	108	216	432	864
106	3	38	114	342	1026	3078
110	4	27	108	432	1728	6912
114	5	18	90	450	2250	11250
118	6	11	66	396	2376	14256
122	7	5	34	245	1715	12005
126	8	2	16	128	1024	8192
		$N = \sum f = 480$	$\sum fu = 236$	$\sum fu^2 = 3404$	$\sum fu^3 = 6428$	$\sum fu^4 = 74{,}588$

(a) $m_1' = c\dfrac{\sum fu}{N} = (4)\left(\dfrac{236}{480}\right) = 1.9667$

(b) $m_2' = c^2\dfrac{\sum fu^2}{N} = (4)^2\left(\dfrac{3404}{480}\right) = 113.4667$

(c) $m_3' = c^3\dfrac{\sum fu^3}{N} = (4)^3\left(\dfrac{6428}{480}\right) = 857.0667$

(d) $m_4' = c^4\dfrac{\sum fu^4}{N} = (4)^4\left(\dfrac{74{,}588}{480}\right) = 39{,}780.2667$

(e) $m_1 = 0$

(f) $m_2 = m_2' - m_1'^2 = 113.4667 - (1.9667)^2 = 109.5988$

(g) $m_3 = m_3' - 3m_1'm_2' + 2m_1'^3 = 857.0667 - 3(1.9667)(113.4667) + 2(1.9667)^3 = 202.8158$

(h) $m_4 = m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4 = 35{,}627.2853$

(i) $\bar{X} = \overline{(A+d)} = A + m_1' = A + c\dfrac{\sum fu}{N} = 94 + 1.9667 = 95.97$

(j) $s = \sqrt{m_2} = \sqrt{109.5988} = 10.47$

(k) $\overline{X^2} = \overline{(A+d)^2} = \overline{(A^2 + 2Ad + d^2)} = A^2 + 2A\bar{d} + \overline{d^2} = A^2 + 2Am_1' + m_2'$

$= (94)^2 + 2(94)(1.9667) + 113.4667 = 9319.2063$, or 9319 to four significant figures

(l) $\overline{X^3} = \overline{(A+d)^3} = \overline{(A^3 + 3A^2d + 3Ad^2 + d^3)} = A^3 + 3A^2\bar{d} + 3A\overline{d^2} + \overline{d^3}$

$= A^3 + 3A^2m_1' + 3Am_2' + m_3' = 915{,}571.9597$, or 915,600 to four significant figures

CHARLIER'S CHECK

5.8 Illustrate the use of Charlier's check for the computations in Problem 5.7.

SOLUTION

To supply the required check, we add to Table 5.3 the columns shown in Table 5.4 (with the exception of column 2, which is repeated in Table 5.3 for convenience).

In each of the following groupings, the first is taken from Table 5.4 and the second is taken from Table 5.2. Equality of results in each grouping provides the required check.

Table 5.4

$u+1$	f	$f(u+1)$	$f(u+1)^2$	$f(u+1)^3$	$f(u+1)^4$
-5	4	-20	100	-500	2500
-4	9	-36	144	-576	2304
-3	16	-48	144	-432	1296
-2	28	-56	112	-224	448
-1	45	-45	45	-45	45
0	66	0	0	0	0
1	85	85	85	85	85
2	72	144	288	576	1152
3	54	162	486	1458	4374
4	38	152	608	2432	9728
5	27	135	675	3375	16875
6	18	108	648	3888	23328
7	11	77	539	3773	26411
8	5	40	320	2560	20480
9	2	18	162	1458	13122
	$N=\sum f$ $=480$	$\sum f(u+1)$ $=716$	$\sum f(u+1)^2$ $=4356$	$\sum f(u+1)^3$ $=17{,}828$	$\sum f(u+1)^4$ $=122{,}148$

$$\sum f(u+1) = 716$$

$$\sum fu + N = 236 + 480 = 716$$

$$\sum f(u+1)^2 = 4356$$

$$\sum fu^2 + 2\sum fu + N = 3404 + 2(236) + 480 = 4356$$

$$\sum f(u+1)^3 = 17{,}828$$

$$\sum fu^3 + 3\sum fu^2 + 3\sum fu + N = 6428 + 3(3404) + 3(236) + 480 = 17{,}828$$

$$\sum f(u+1)^4 = 122{,}148$$

$$\sum fu^4 + 4\sum fu^3 + 6\sum fu^2 + 4\sum fu + N = 74{,}588 + 4(6428) + 6(3404) + 4(236) + 480 = 122{,}148$$

SHEPPARD'S CORRECTIONS FOR MOMENTS

5.9 Apply Sheppard's corrections to determine the moments about the mean for the data in (*a*) Problem 5.6 and (*b*) Problem 5.7.

SOLUTION

(a) Corrected $m_2 = m_2 - c^2/12 = 8.5275 - 3^2/12 = 7.7775$

$$\text{Corrected } m_4 = m_4 - \tfrac{1}{2}c^2 m_2 + \tfrac{7}{240}c^4$$
$$= 199.3759 - \tfrac{1}{2}(3)^2(8.5275) + \tfrac{7}{240}(3)^4$$
$$= 163.3646$$

m_1 and m_2 need no correction.

(b) Corrected $m_2 = m_2 - c^2/12 = 109.5988 - 4^2/12 = 108.2655$

$$\text{Corrected } m_4 = m_4 - \tfrac{1}{2}c^2 m_2 + \tfrac{7}{240}c^4$$
$$= 35{,}627.2853 - \tfrac{1}{2}(4)^2(109.5988) + \tfrac{7}{240}(4)^4$$
$$= 34{,}757.9616$$

SKEWNESS

5.10 Find Pearson's (a) first and (b) second coefficients of skewness for the wage distribution of the 65 employees at the P&R Company (see Problems 3.44 and 4.18).

SOLUTION

Mean = \$279.76, median = \$279.06, mode = \$277.50, and standard deviation s = \$15.60. Thus:

(a) $$\text{First coefficient of skewness} = \frac{\text{mean} - \text{mode}}{s} = \frac{\$279.76 - \$277.50}{\$15.60} = 0.1448, \text{ or } 0.14$$

(b) $$\text{Second coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{s} = \frac{3(\$279.76 - \$279.06)}{\$15.60} = 0.1346, \text{ or } 0.13$$

If the corrected standard deviation is used [see Problem 4.21(b)], these coefficients become, respectively:

(a) $$\frac{\text{Mean} - \text{mode}}{\text{Corrected } s} = \frac{\$279.76 - \$277.50}{\$15.33} = 0.1474, \text{ or } 0.15$$

(b) $$\frac{3(\text{mean} - \text{median})}{\text{Corrected } s} = \frac{3(\$279.76 - \$279.06)}{\$15.33} = 0.1370, \text{ or } 0.14$$

Since the coefficients are positive, the distribution is skewed positively (i.e., to the right).

5.11 Find the (a) quartile and (b) percentile coefficients of skewness for the distribution of Problem 5.10 (see Problem 3.44).

SOLUTION

Q_1 = \$268.25, $Q_2 = P_{50}$ = \$279.06, Q_3 = \$290.75, $P_{10} = D_1$ = \$258.12, and $P_{90} = D_9$ = \$301.00. Thus:

(a) $$\text{Quartile coefficient of skewness} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{\$290.75 - 2(\$279.06) + \$268.25}{\$290.75 - \$268.25} = 0.0391$$

(b) $$\text{Percentile coefficient of skewness} = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} = \frac{\$301.00 - 2(\$279.06) + \$258.12}{\$301.00 - \$258.12} = 0.0233$$

5.12 Find the moment coefficient of skewness, a_3, for (a) the height distribution of students at XYZ University (see Problem 5.6) and (b) the IQ's of elementary school children (see Problem 5.7).

SOLUTION

(*a*) $m_2 = s^2 = 8.5275$, and $m_3 = -2.6932$. Thus:

$$a_3 = \frac{m_3}{s^3} = \frac{m_3}{(\sqrt{m_2})^3} = \frac{-2.6932}{(\sqrt{8.5275})^3} = -0.1081 \qquad \text{or} \qquad -0.11$$

If Sheppard's corrections for grouping are used [see Problem 5.9(*a*)], then

$$\text{Corrected } a_3 = \frac{m_3}{(\sqrt{\text{corrected } m_2})^3} = \frac{-2.6932}{(\sqrt{7.7775})^3} = -0.1242 \qquad \text{or} \qquad -0.12$$

(*b*)

$$a_3 = \frac{m_3}{s^3} = \frac{m_3}{(\sqrt{m_2})^3} = \frac{202.8158}{(\sqrt{109.5988})^3} = 0.1768 \qquad \text{or} \qquad 0.18$$

If Sheppard's corrections for grouping are used [see Problem 5.9(*b*)], then

$$\text{Corrected } a_3 = \frac{m_3}{(\sqrt{\text{corrected } m_2})^3} = \frac{202.8158}{(\sqrt{108.2655})^3} = 0.1800 \qquad \text{or} \qquad 0.18$$

Note that both distributions are moderately skewed, distribution (*a*) to the left (negatively) and distribution (*b*) to the right (positively). Distribution (*b*) is more skewed than (*a*); that is, (*a*) is more symmetrical than (*b*), as is evidenced by the fact that the numerical value (or absolute value) of the skewness coefficient for (*b*) is greater than that for (*a*).

KURTOSIS

5.13 Find the moment coefficient of kurtosis, a_4, for the data of (*a*) Problem 5.6 and (*b*) Problem 5.7.

SOLUTION

(*a*)

$$a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2} = \frac{199.3759}{(8.5275)^2} = 2.7418 \qquad \text{or} \qquad 2.74$$

If Sheppard's corrections are used [see Problem 5.9(*a*)], then

$$\text{Corrected } a_4 = \frac{\text{corrected } m_4}{(\text{corrected } m_2)^2} = \frac{163.36346}{(7.7775)^5} = 2.7007 \qquad \text{or} \qquad 2.70$$

(*b*)

$$a_4 = \frac{m_4}{s^4} = \frac{m_4}{m_2^2} = \frac{35{,}627.2853}{(109.5988)^2} = 2.9660 \qquad \text{or} \qquad 2.97$$

If Sheppard's corrections are used [see Problem 5.9(*b*)], then

$$\text{Corrected } a_4 = \frac{\text{corrected } m_4}{(\text{corrected } m_2)^2} = \frac{34{,}757.9616}{(108.2655)^2} = 2.9653 \qquad \text{or} \quad 2.97$$

Since for a normal distribution $a_4 = 3$, it follows that both distributions (*a*) and (*b*) are *platykurtic* with respect to the normal distribution (i.e., less peaked than the normal distribution).

Insofar as peakedness is concerned, distribution (*b*) approximates the normal distribution much better than does distribution (*a*). However, from Problem 5.12 distribution (*a*) is more symmetrical than (*b*), so that as far as symmetry is concerned, (*a*) approximates the normal distribution better than (*b*) does.

SOFTWARE COMPUTATION OF SKEWNESS AND KURTOSIS

5.14 Sometimes the scores on a test do not follow the normal distribution, although they usually do. We sometimes find that students score low or high with very few in the middle. The distribution shown in Fig. 5-4 is such a distribution. This distribution is referred to as the U distribution. Find the mean, standard deviation, skewness, and kurtosis for the data using EXCEL.

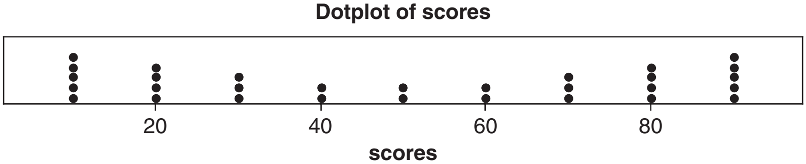

Fig. 5-4 .A MINITAB plot of data that follows a U distribution.

SOLUTION

The data is entered into A1:A30 of an EXCEL worksheet. The command "=**AVERAGE(A1:A30)**" gives 50. The command "=**STDEV(A1:A30)**" gives 29.94. The command "=**SKEW(A1:A30)**" gives 0. The command "=**KURT(A1:A30)**" gives -1.59.

Supplementary Problems

MOMENTS

5.15 Find the (*a*) first, (*b*) second, (*c*) third, and (*d*) fourth moments for the set 4, 7, 5, 9, 8, 3, 6.

5.16 Find the (*a*) first, (*b*) second, (*c*) third, and (*d*) fourth moments about the mean for the set of numbers in Problem 5.15.

5.17 Find the (*a*) first, (*b*) second, (*c*) third, and (*d*) fourth moments about the number 7 for the set of numbers in Problem 5.15.

5.18 Using the results of Problems 5.16 and 5.17, verify the relations between the moments (*a*) $m_2 = m_2' - m_1'^2$, (*b*) $m_3 = m_3' - 3m_1'm_2' + 2m_1'^3$, and (*c*) $m_4 = m_4' - 4m_1'm_3' + 6m_1'^2m_2' - 3m_1'^4$.

5.19 Find the first four moments about the mean of the set of numbers in the arithmetic progression 2, 5, 8, 11, 14, 17.

5.20 Prove that (*a*) $m_2' = m_2 + h^2$, (*b*) $m_3' = m_3 + 3hm_2 + h^3$, and (*c*) $m_4' = m_4 + 4hm_3 + 6h^2m_2 + h^4$, where $h = m_1'$.

5.21 If the first moment about the number 2 is equal to 5, what is the mean?

5.22 If the first four moments of a set of numbers about the number 3 are equal to -2, 10, -25, and 50, determine the corresponding moments (*a*) about the mean, (*b*) about the number 5, and (*c*) about zero.

5.23 Find the first four moments about the mean of the numbers 0, 0, 0, 1, 1, 1, 1, and 1.

5.24 (*a*) Prove that $m_5 = m_5' - 5m_1'm_4' + 10m_1'^2m_3' - 10m_1'^3m_2' + 4m_1'^5$.

(*b*) Derive a similar formula for m_6.

5.25 Of a total of N numbers, the fraction p are 1's, while the fraction $q = 1 - p$ are 0's. Find (*a*) m_1, (*b*) m_2, (*c*) m_3, and (*d*) m_4 for the set of numbers. Compare with Problem 5.23.

5.26 Prove that the first four moments about the mean of the arithmetic progression $a, a+d, a+2d, \ldots, a+(n-1)d$ are $m_1 = 0$, $m_2 = \frac{1}{12}(n^2-1)d^2$, $m_3 = 0$, and $m_4 = \frac{1}{240}(n^2-1)(3n^2-7)d^4$. Compare with Problem 5.19 (see also Problem 4.69). [*Hint*: $1^4 + 2^4 + 3^4 + \cdots + (n-1)^4 = \frac{1}{30}n(n-1)(2n-1)(3n^2-3n-1)$.]

MOMENTS FOR GROUPED DATA

5.27 Calculate the first four moments about the mean for the distribution of Table 5.5.

Table 5.5

X	f
12	1
14	4
16	6
18	10
20	7
22	2
	Total 30

5.28 Illustrate the use of Charlier's check for the computations in Problem 5.27.

5.29 Apply Sheppard's corrections to the moments obtained in Problem 5.27.

5.30 Calculate the first four moments about the mean for the distribution of Problem 3.59(*a*) without Sheppard's corrections and (*b*) with Sheppard's corrections.

5.31 Find (*a*) m_1, (*b*) m_2, (*c*) m_3, (*d*) m_4, (*e*) $\bar{X}$, (*f*) s, (*g*) $\overline{X^2}$, (*h*) $\overline{X^3}$, (*i*) $\overline{X^4}$, and (*j*) $\overline{(X+1)^3}$ for the distribution of Problem 3.62.

SKEWNESS

5.32 Find the moment coefficient of skewness, a_3, for the distribution of Problem 5.27 (*a*) without and (*b*) with Sheppard's corrections.

5.33 Find the moment coefficient of skewness, a_3, for the distribution of Problem 3.59 (see Problem 5.30).

5.34 The second moments about the mean of two distributions are 9 and 16, while the third moments about the mean are -8.1 and -12.8, respectively. Which distribution is more skewed to the left?

5.35 Find Pearson's (*a*) first and (*b*) second coefficients of skewness for the distribution of Problem 3.59, and account for the difference.

5.36 Find the (*a*) quartile and (*b*) percentile coefficients of skewness for the distribution of Problem 3.59. Compare your results with those of Problem 5.35 and explain.

5.37 Table 5.6 gives three different distributions for the variable X. The frequencies for the three distributions are given by f_1, f_2, and f_3. Find Pearson's first and second coefficients of skewness for the three distributions. Use the corrected standard deviation when computing the coefficients.

Table 5.6

X	f_1	f_2	f_3
0	10	1	1
1	5	2	2
2	2	14	2
3	2	2	5
4	1	1	10

KURTOSIS

5.38 Find the moment coefficient of kurtosis, a_4, for the distribution of Problem 5.27 (*a*) without and (*b*) with Sheppard's corrections.

5.39 Find the moment coefficient of kurtosis for the distribution of Problem 3.59 (*a*) without and (*b*) with Sheppard's corrections (see Problem 5.30).

5.40 The fourth moments about the mean of the two distributions of Problem 5.34 are 230 and 780, respectively. Which distribution more nearly approximates the normal distribution from the viewpoint of (*a*) peakedness and (*b*) skewness?

5.41 Which of the distributions in Problem 5.40 is (*a*) leptokurtic, (*b*) mesokurtic, and (*c*) platykurtic?

5.42 The standard deviation of a symmetrical distribution is 5. What must be the value of the fourth moment about the mean in order that the distribution be (*a*) leptokurtic, (*b*) mesokurtic, and (*c*) platykurtic?

5.43 (*a*) Calculate the percentile coefficient of kurtosis, κ, for the distribution of Problem 3.59.

(*b*) Compare your result with the theoretical value 0.263 for the normal distribution, and interpret.

(*c*) How do you reconcile this result with that of Problem 5.39?

SOFTWARE COMPUTATION OF SKEWNESS AND KURTOSIS

5.44 The data in Fig. 5-5 show a sharp peak at 50. This should show up in the kurtosis measure of the data. Using EXCEL, show that the skewness is basically zero and that the kurtosis is 2.0134.

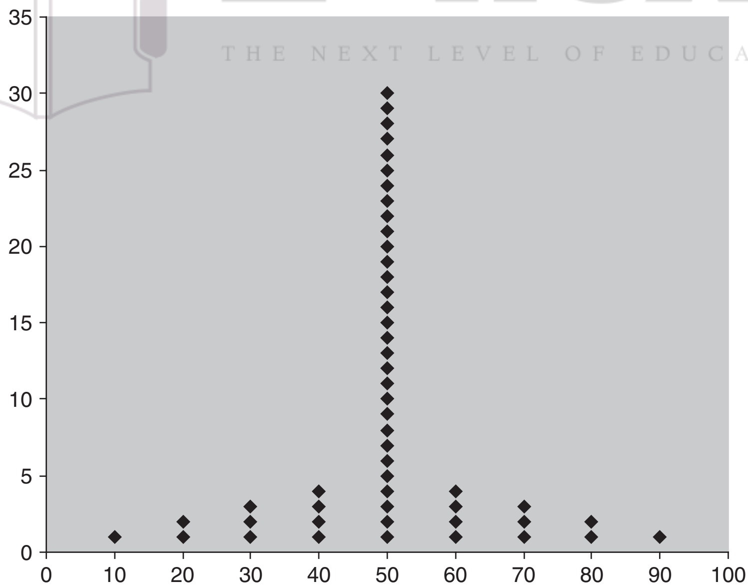

Fig. 5-5 EXCEL plot of test score data.