APPENDIX D

PRESENTATION AND ANALYSIS OF SOLID WASTE MANAGEMENT DATA

The purpose of this appendix is to introduce (1) techniques used to graphically present and analyze solid waste management data, (2) statistical measures that are commonly used to characterize solid waste collection rates, and (3) graphical procedures that can be used to determine the nature of a distribution. For a more detailed presentation of the fundamentals of statistical analysis, the reader is referred to Refs. 1 to 4.

D-1 GRAPHICAL PRESENTATION OF FIELD DATA

The graphical presentation of observed field data can be used to depict and identify trends in the data. Time series and histogram or frequency plots are used extensively for the presentation and analysis of such data.

Time Series

Observations that are arranged in the order of their occurrence in time are often called time series. By plotting the observed values versus time, it is often possible to establish trends, cycles or periodicities, and fluctuations that may be of value in understanding the basic nature of the phenomenon under evaluation.

Description of Trends, Cycles, and Fluctuations. The word *trend* is used to describe a relatively long-term tendency for field observations to increase or decrease in some orderly manner (see Fig. D-1*a*). The change in the magnitude of the observations may be simple or complex.

Observations that are cyclic tend to form successive peaks and valleys (see Fig. D-1*b*). Like trends, cycles may be of simple periodicity or may be defined by long-term repeating periodicities.

The term *fluctuating time series* is often used to describe observations that change significantly from one time interval to the next with no apparent pattern (see Fig. D-1*c*).

Analysis of Trends, Cycles, and Fluctuations. Mathematically, linear trends are described with straight lines of the form $y = a + bx$. Curvilinear trends are described most commonly with polynomials of the general form $y = a+bx+cx^2$, although a variety of other equations are also used [1]. The method of moving averages is often used to suppress random irregularities in plotted time series data so that short- and long-term trends can be discovered. Moving averages may be simple or weighted. The moving average is obtained by averaging an odd number of successive observations. Thus, the average will correspond to the middle item, which can be weighted for emphasis. Two examples of weighted moving averages are given by Eqs. (D-1) and (D-2).

$$x_b = \frac{a + 2b + c}{4} \qquad \text{(D-1)}$$

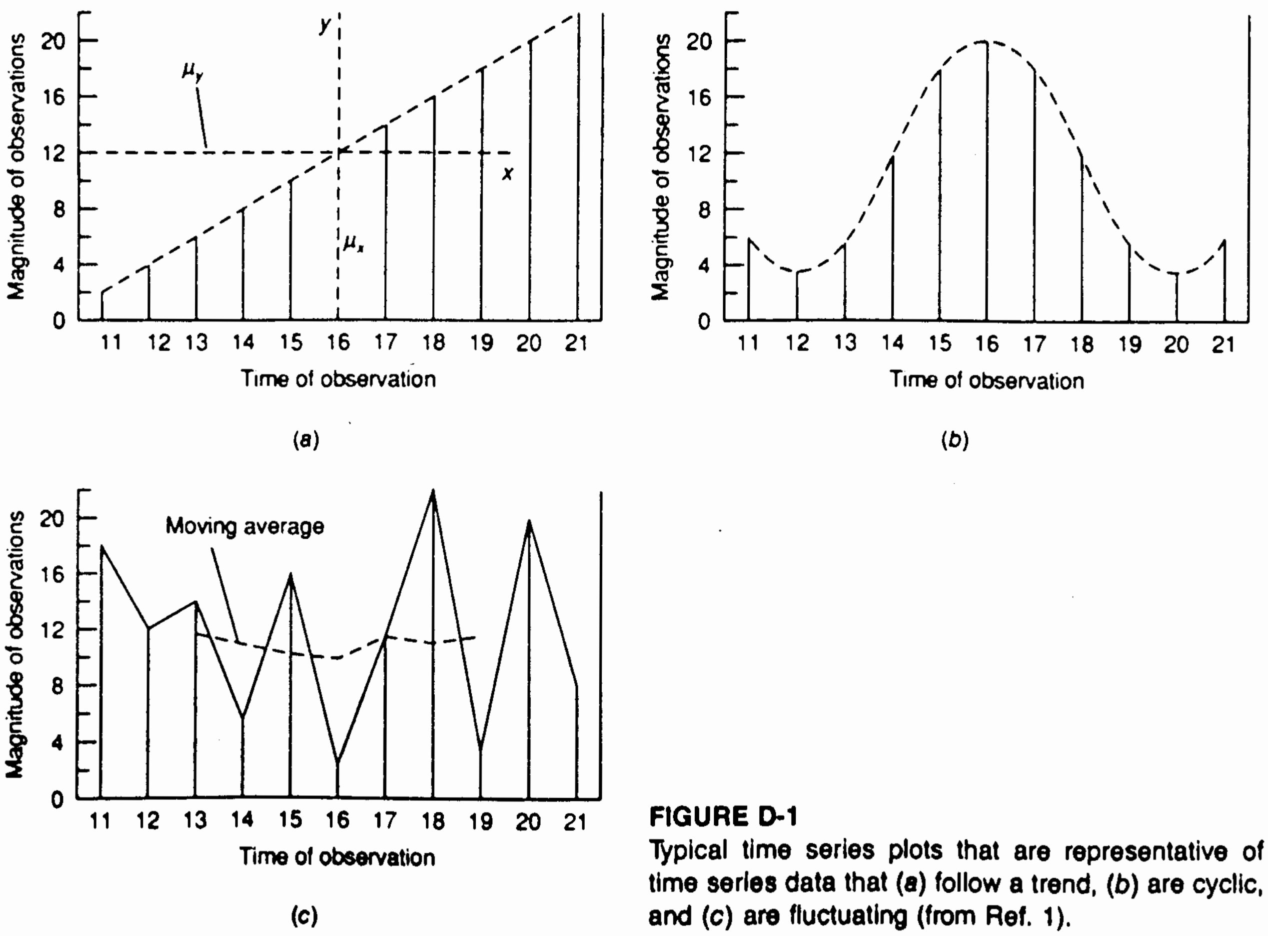


FIGURE D-1
Typical time series plots that are representative of time series data that (*a*) follow a trend, (*b*) are cyclic, and (*c*) are fluctuating (from Ref. 1).

where x_b = average value at point b
a, b, c = observed magnitudes at points a, b, c

$$x_c = \frac{a + 4b + 6c + 4d + e}{16} \tag{D-2}$$

The coefficients in the moving average given by Eq. (D-2) are obtained from the binomial expansion. The weighted moving average for the fluctuating time series shown in Fig. D-1*c* was computed using Eq. (D-1).

Frequency Distributions

Observations arranged in order of magnitude form an array. The same data can be grouped together in a series of data ranges. If whole numbers are assigned to the number of observations that occur in each data range, then the frequency of occurrence of whole numbers can be plotted against the magnitude of the data ranges. The resulting plots are called histograms (see Fig. D-2). As shown, histograms can be symmetrical (Fig. D-2*a*), asymmetrical (Fig. D-2*b*), rectangular

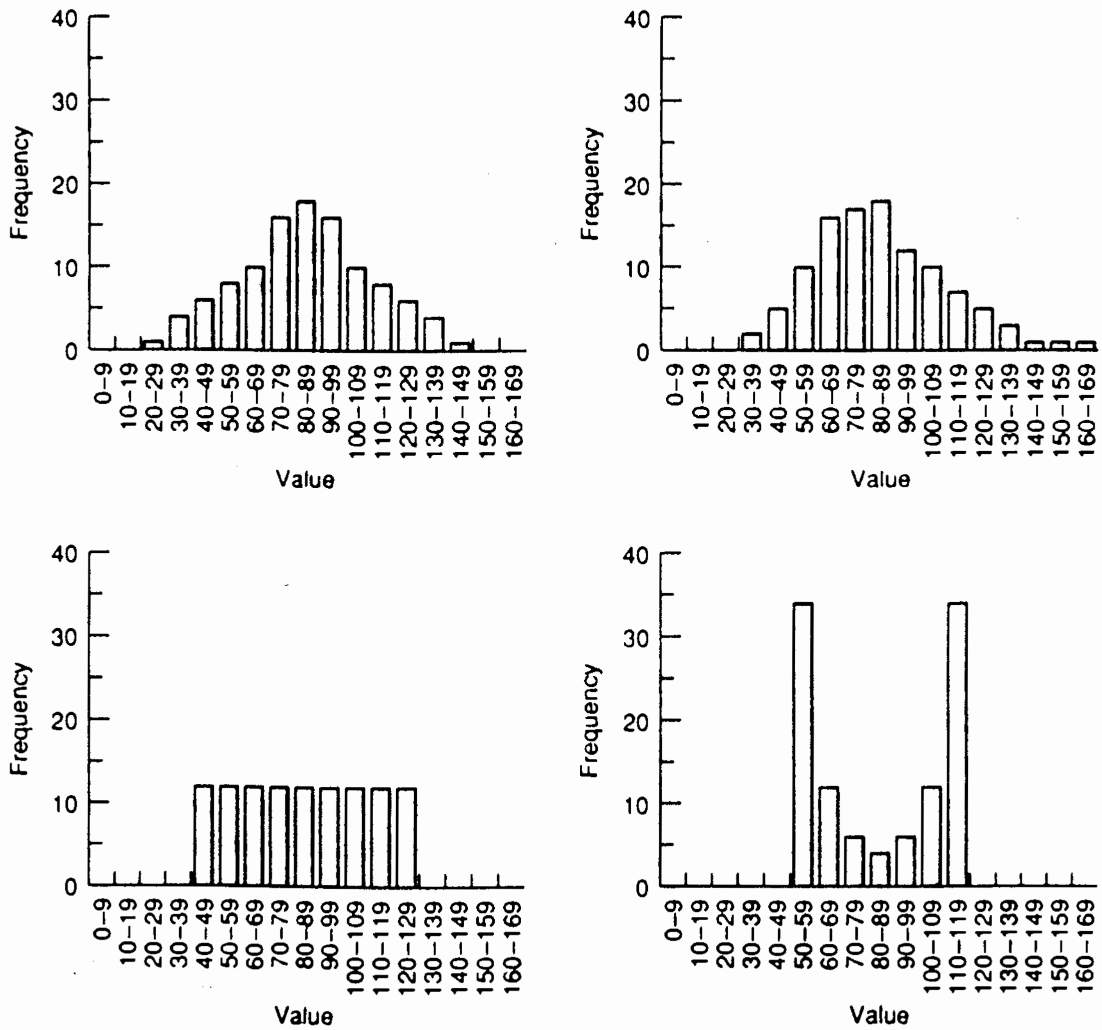


FIGURE D-2
Histogram of four different distributions which have the same arithmetic mean and standard deviation (adapted from Ref. 4).

TABLE D-1
Grouped weekly waste collection volumes obtained from a small municipality for a period of one year

Collection volume, yd^3/wk	Data range mid-point	Frequency, f_i
1000–1100	1050	1
1100–1200	1150	3
1200–1300	1250	4
1300–1400	1350	9
1400–1500	1450	11
1500–1600	1550	10
1600–1700	1650	7
1700–1800	1750	4
1800–1900	1850	2
1900–2000	1950	1
Total		52

(Fig. D-2*c*), or U-shaped (Fig. D-2*d*). The data forming symmetrical histograms are said to be *normally* distributed, whereas the data in asymmetrical histograms are said to be *skewed*.

A typical example of a year's worth of weekly solid waste collection data that has been arranged in data ranges is reported in Table D-1. As shown, the 52 individual weekly data points have been grouped in data ranges varying from 1000–1100 to 1900–2000. The midpoint of each data range is given in column two. The frequency values reported in column three correspond to the number

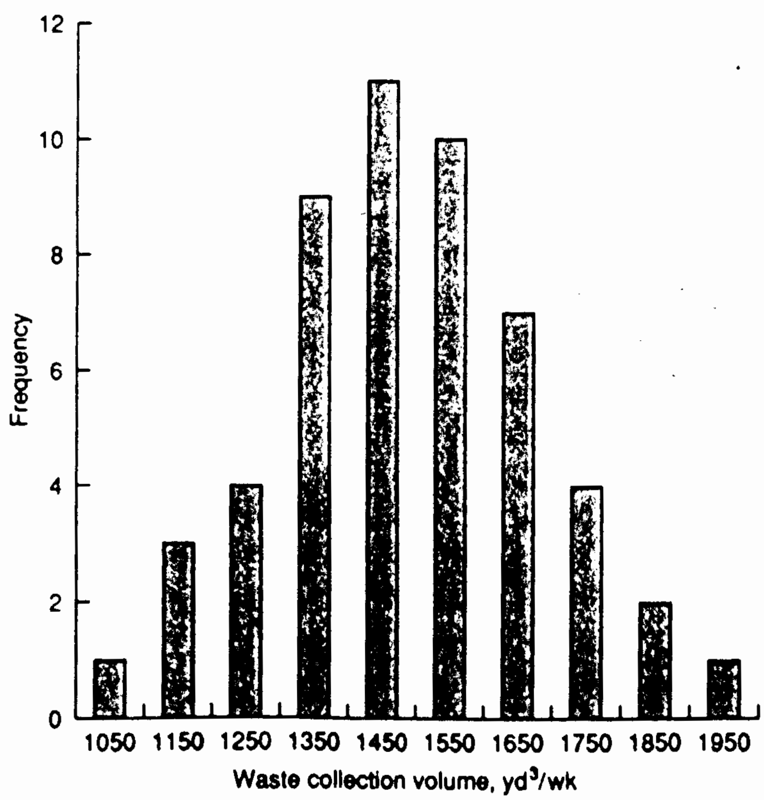


FIGURE D-3
Histogram of yearly solid waste collection data from a small municipality.

of weekly values that were in a given data range. A histogram of the data reported in Table D-1 is presented in Fig. D-3. The statistical characteristics of the grouped solid waste collection data are determined in the following section in Example D-1.

D-2 STATISTICAL MEASURES

Commonly used statistical measures include the mean, median, mode, standard deviation, and coefficient of variation. Although the terms just cited are the most commonly used statistical measures, it is instructive to review the distributions plotted in Fig. D-2 and to note that they all have the same arithmetic mean and standard deviation. Thus, two additional statistical measures are needed to quantify the nature of a given distribution. The two additional measures are the coefficient of skewness, and coefficient of kurtosis. All of the statistical measures are defined below. Determination of these statistical measures from field data is illustrated in Example D-1 presented following the discussion of the coefficient of kurtosis.

Mean

The mean is defined as the arithmetic average of a number of individual or grouped observations. The mean for grouped observations is given by

$$\bar{x} = \frac{\sum f_i x_i}{n} \tag{D-3}$$

where $\bar{x}$ = mean value
f_i = frequency (for ungrouped data $f_i = 1$)
x_i = the midpoint of the ith data range (for ungrouped data x_i = the ith observation)
n = number of observations (note $\sum f_i = n$)

Median

If a series of observations are arranged in order of increasing value, the midmost observation, or the arithmetic mean of the two midmost observations, in a series is known as the median. For example, in a set of 15 measurements, the median will be the 8th value, whereas in a set of 16 measurements, the median will be the average of the 8th and 9th values. In a symmetrical distribution, the median will equal the mean.

Mode

The value occurring with the greatest frequency in a set of observations is known as the mode. If a continuous graph of the frequency distribution is drawn, the mode is the value of the high point, or hump, of the curve. In a symmetrical set of observations, the mean, median, and mode will be the same value. The mode can be approximated with reasonable accuracy using the following expression:

$$\text{Mod} = 3(\text{Med}) - 2(\bar{x}) \tag{D-4}$$

where Mod = mode
Med = median
$\bar{x}$ = mean

Standard Deviation

Because of the laws of chance, there is uncertainty in any set of measurements. The precision of a set of measurements can be assessed in a number of different ways. Most commonly, the error of an individual measurement in a set is defined as the difference between the arithmetic mean and the value of the measurement. The standard deviation for grouped data is defined as follows:

$$s = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{n - 1}} \tag{D-5}$$

where s = standard deviation
f_i = frequency (for ungrouped data $f_i = 1$)
x_i = the midpoint of the ith data range (for ungrouped data x_i = the ith observation)
$\bar{x}$ = mean value
n = number of observations

From the form of the equation, it can be concluded that the larger the scatter in a set of measurements is, the larger the value of s will be. Conversely, as the precision of a set of measurement improves, the value of the standard deviation will decrease. From theoretical considerations, it can be shown that if the measurements are distributed normally, then 68.27 percent of the observations will fall within plus or minus one standard of deviation of the mean ($\bar{x} \pm s$) [4, 5].

Coefficient of Variation

Although the standard deviation can be used as an indication of the absolute dispersion of a set of measured values, it provides little or no information about whether the value is large or small. To overcome this difficulty, the coefficient of variation, as defined in Eq. (D-4), is used as a relative measure of dispersion [4].

$$CV = \frac{100\, s}{\bar{x}} \tag{D-6}$$

where CV = coefficient of variation, percent
s = standard deviation (see Eq. D-5)
$\bar{x}$ = mean value (see Eq. D-3)

Typically, the coefficient of variation for solid waste generation rates will vary from 10 to 60 percent. To judge whether this percentage represents a large or small scatter, it can be compared to values obtained from measurements in other fields. For measurements in the biological field, the coefficient of variation will vary from about 10 to 30 percent. The coefficient of variation for chemical analyses varies from 2 to 10 percent. Clearly, the scatter in solid waste generation data is significant.

Coefficient of Skewness

When a frequency distribution is asymmetrical (see Fig. D-2*b*), it is usually defined as being a skewed frequency distribution. Over the years, a number of measures of skewness have been proposed, but none is accepted universally. For the purposes of this textbook, skewness is defined by the following relationship:

$$\alpha_3 = \frac{\sum f_i(x_i - \bar{x})^3/n - 1}{s^3} \qquad \text{(D-7)}$$

where α_3 = coefficient of skewness
f_i = frequency (for ungrouped data $f_i = 1$)
x_i = the midpoint of the ith data range (for ungrouped data x_i = the ith observation)
$\bar{x}$ = mean value
n = number of observations
s = standard deviation

The coefficient of skewness has also been computed using the following relationship.

$$\alpha_3{}' = \frac{(\bar{x} - \text{Mod})}{s} \qquad \text{(D-8)}$$

where $\alpha_3{}'$ = coefficient of skewness
$\bar{x}$ = mean value
Mod = mode
s = standard deviation

Coefficient of Kurtosis

The extent to which a distribution is more peaked (see Fig. D-2*b*) or more flat-topped (see Fig. D-2*c*) than the normal distribution is defined by the kurtosis of the distribution. The coefficient of kurtosis can be computed using the following equation:

$$\alpha_4 = \frac{\sum f_i(x_i - \bar{x})^4/n - 1}{s^4} \qquad \text{(D-9)}$$

where α_4 = coefficient of kurtosis
f_i = frequency (for ungrouped data $f_i = 1$)
x_i = the midpoint of the ith data range (for ungrouped data x_i = the ith observation)
$\bar{x}$ = mean value
n = number of observations
s = standard deviation

The value of the kurtosis for a normal distribution is 3. A peaked curve will have a value greater than 3 where as a flatter curve it will have a value less than 3. The value α_4 that separates mound-shaped curves from rectangular or U-shaped curves is in the range from 1.75 to 1.8. Values of α_4 for U-shaped distributions are less than 1.75.

Example D-1 Statistical analysis of solid waste collection data. Determine the statistical characteristics of the weekly solid waste collection data presented in Table D-1.

Solution

1. Determine the statistical characteristics of the solid waste collection data.
   (*a*) Set up a data analysis table to obtain the quantities needed to determine the statistical characteristics.

Data range, yd^3/wk	x_i[a]	Freq., f_i	$f_i x_i$	$(x_i - \bar{x})$	$f_i(x_i - \bar{x})^2 \times 10^{-3}$	$f_i(x_i - \bar{x})^3 \times 10^{-5}$	$f_i(x_i - \bar{x})^4 \times 10^{-10}$
1,000–1,100	1,050	1	1,050	−437	191.0	−834.5	3.647
1,100–1,200	1,150	3	3,450	−337	340.7	−1.148.2	3.870
1,200–1,300	1,250	4	5,000	−237	224.7	−532.5	1.262
1,300–1,400	1,350	9	12,150	−137	168.9	−231.4	0.317
1,400–1,500	1,450	11	15,950	−37	15.1	−5.6	0.002
1,500–1,600	1,550	10	15,500	63	39.7	25.0	0.016
1,600–1,700	1,650	7	11,550	163	186.0	303.2	0.494
1,700–1,800	1,750	4	7,000	263	276.7	727.7	1.914
1,800–1,900	1,850	2	3,700	363	263.5	1,435.0	5.208
1,900–2,000	1,950	1	1,950	463	214.4	1,985.1	9.190
Total		52	77,300		1,920.6	1,723.8	25.920

$\bar{x}$ = 77,300/52 = 1,487

[a] Midpoint of data range.

   (*b*) Determine these statistical characteristics.
      i. Mean

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

$$\bar{x} = \frac{77{,}300}{52} = 1487 \text{ yd}^3\text{/wk (see data table above)}$$

ii. Median (the midmost value)

$$\text{Med} = 1450 \text{ yd}^3/\text{wk (see data table above)}$$

iii. Mode

$$\text{Mod} = 1450 \text{ yd}^3/\text{wk (see data table above)}$$

iv. Standard deviation

$$s = \sqrt{\frac{\sum f_i\left(x_i - \bar{x}\right)^2}{n - 1}}$$

$$s = \sqrt{\frac{1920.7 \times 10^5}{51}} = 194.1$$

v. Coefficient of variation

$$CV = \frac{100\ s}{\bar{x}}$$

$$CV = \frac{100\ (194.1)}{1487} = 13.1\%$$

vi. Coefficient of skewness

$$\alpha_3 = \frac{\sum f_i\left(x_i - \bar{x}\right)^3 / n - 1}{s^3}$$

$$\alpha_3 = \frac{1723.8 \times 10^5/51}{194.1^3} = 0.462$$

vii. Coefficient of kurtosis

$$\alpha_4 = \frac{\sum f_i\left(x_i - \bar{x}\right)^4 / n - 1}{s^4}$$

$$\alpha_4 = \frac{25.920 \times 10^{10}/51}{(194.1)^4} = 3.6$$

Comment. The solid waste collection data have a slight negative skewness (normal bell-shaped curve is distorted slightly to the right). Further, the field data are more peaked than the normal curve (3.58 versus 3 for a normal curve), which means the data are bunched together more closely than for a normal curve.

Comparison of Statistical Measures for Various Distributions

The statistical measures for the distributions shown in Fig. D-2 are reported below [4]. As shown, the mean and standard deviation are the same for all of the distributions. If only these measures had been computed, it would be concluded that all the distributions are the same when in reality they are quite different.

Clearly, having information on the values of α_3 and α_4 is important in assessing the nature of the distributions.

Statistical measure	Fig. D-2*a*	Fig. D-2*b*	Fig. D-2*c*	Fig. D-2*d*
n	108	108	108	108
$\bar{x}$	85	85	85	85
s	25.8	25.8	25.8	25.8
α_3	0	+0.57	0	0
α_4	2.565	3.188	1.770	1.23

From Ref. 4.

D-3 SKEWED DISTRIBUTIONS

In general, some degree of positive skewness (normal bell-shaped curve is distorted to the left) is common in solid waste generation data. Fortunately, most statistical tests based on the normal distribution are robust, and small amounts of skewness can be tolerated. If the field data are skewed severely, they may have to be rescaled by taking the logarithm or the square root to make them more normal. The most common statistical measures for skewed distributions are the geometric mean and standard deviation. These measures are computed as follows [1].

Geometric Mean

The geometric mean is defined as the log average of a number of individual measurements and is given by

$$\log M_g = \frac{\sum f_i(\log x_i)}{n} \qquad \text{(D-10)}$$

where M_g = geometric mean value
f_i = frequency (for ungrouped data $f_i = 1$)
x_i = the midpoint of the ith data range (for ungrouped data x_i = the ith observation)
n = number of observations

Geometric Standard Deviation

The geometric standard deviation is defined as follows:

$$\log s_g = \sqrt{\frac{\sum (f_i \log^2 x_g)}{n-1}} \qquad \text{(D-11)}$$

where s_g = geometric standard deviation
f_i = frequency (for ungrouped data $f_i = 1$)
$x_g = x_i/M_g$
x_i = the midpoint of the ith data range (for ungrouped data x_i = the ith observation)
n = number of observations

D-4 GRAPHICAL ANALYSIS OF FIELD DATA

For most practical purposes, the type of the distribution can be determined by plotting the data on both arithmetic and logarithmic probability paper and noting whether the data can be fitted with a straight line. The use of arithmetic and logarithmic probability paper is discussed and illustrated in the following paragraphs [1, 2, and 3].

Probability Paper

Although it is possible to express the summation or probability curve in equation form, it has been found more useful to develop a type of graph paper, called "probability paper," with special coordinates on which data that are normal or logarithmically normal will plot as a straight line. Three steps are involved in the use of arithmetic and logarithmic probability paper.

1. The measurements in a data set are first arranged in order of increasing magnitude and assigned a rank serial number.
2. Next, a corresponding plotting position is determined for each data point using Eq. (D-12).
3. The data are then plotted on arithmetic and logarithmic probability paper. The probability scale is labeled "Percent of values equal to or less than the indicated value."

The plotting position is computed using the following equation:

$$\text{Plotting position (\%)} = \left(\frac{m}{n+1}\right) 100 \qquad \text{(D-12)}$$

where m = rank serial number
n = number of observations

The term $(n + 1)$ is used to account for the fact that there may be an observation that is either larger or smaller than the largest or smallest in the data set. In effect, the plotting position represents the percent or frequency of observations that are equal to or less than the indicated value.

Arithmetic Probability Paper. By plotting data on arithmetic probability paper, it is possible to determine:

1. Whether the data are normally distributed by noting if the data can be fit with a straight line. Significant departure from a straight line can be taken as an indication of skewness.
2. The approximate magnitude of the arithmetic mean. Usually it will be best to compute the mean and to pass the straight line plotted by eye through the

4. The expected frequency of any observation of a given magnitude.

Logarithmic Probability Paper. When the data are skewed (see Fig. D-2*b*), logarithmic probability paper can be used. The implication here is that the logarithm of the observed values is normally distributed. On logarithmic probability paper, the straight line of best fit passes through the geometric mean and through the intersection of $M_g \times s_g$ at a value of 84.1 percent and M_g / s_g at a value of 15.9 percent. The geometric standard deviation can be determined from the following equation:

$$s_g = \frac{P_{84.1}}{M_g} = \frac{M_g}{P_{15.9}} \qquad \text{(D-13)}$$

where s_g = geometric standard deviation
$P_{84.1}$ = value from curve at 84.1 percent
M_g = geometric mean
$P_{15.9}$ = value from curve at 15.9 percent

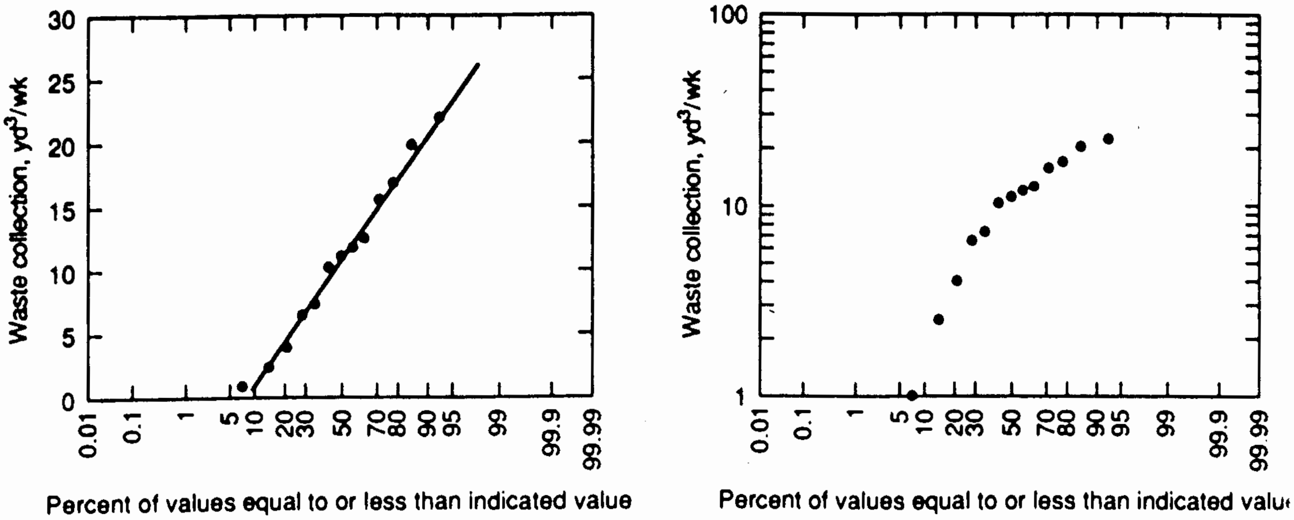
Use of Probability Paper

The use of probability paper to determine the type of distribution is illustrated using the ungrouped solid waste collection data given in the following table.

Rank serial no., m	Waste, yd^3/d	Plotting position,[a] %
1	1.0	7.1
2	2.5	14.3
3	4.0	21.4
4	6.6	28.6
5	7.4	35.7
6	10.4	42.9
7	11.3	50.0
8	12.0	57.1
9	12.6	64.3
10	15.8	71.4
11	17.0	78.6
12	20.0	85.7
13	22.2	92.9

[a] See Eq. (D-12).

As shown, the data have been arranged in order of increasing magnitude, and plotting positions have been computed by using Eq. (D-12). The arithmetic and logarithmic probability plots of these data are given in Figs. D-4*a* and D-4*b*,



(*a*) (*b*)

FIGURE D-4
Probability plots of solid waste collection data: (*a*) arithmetic and (*b*) logarithmic.

respectively. Because the data fall on a straight line on the arithmetic probability paper, it can be concluded that the data are distributed normally and that normal statistics can be applied.

D-5 REFERENCES

1. Fair, G. M. and J. C. Geyer: *Water Supply and Waste-Water Disposal*, John Wiley & Sons, Inc., New York, 1954.
2. Velz, C. J.: "Graphical Approach to Statistics," *Water and Sewage Works*, Reference and Data Issue, Vol. 99, No. 4, 1952.
3. Velz, C. J.: "Applied Stream Sanitation," *Statistical Tools*, Appendix A, Wiley-Interscience, New York, 1970.
4. Waugh, A. E.: *Elements of Statistical Analysis*, 2d ed., McGraw-Hill, New York, 1943.
5. Young, H. D.: *Statistical Treatment of Experimental Data*, McGraw-Hill, New York, 1962.