Designation: D 4686 – 91 (Reapproved 2003)

Standard Guide for Identification and Transformation of Frequency Distributions[1]

This standard is issued under the fixed designation D 4686; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide gives the rudiments of identification of some of the most common and useful frequency distributions. It does not give rigorous identification. To achieve exactitude, the procedures similar to those given by Shapiro[2] should be used.

1.2 This guide provides a key to identify frequency distributions.

1.3 This guide gives ways to select the proper transformation to use to transform a particular set of data to one which can be modeled by the normal distribution, if such a transformation can be found at all.

1.4 This guide includes the following topics:

Topic Title	Section
Scope	1
Referenced Documents	2
Terminology	3
Significance and Use	4
Key to Distributions	5
Binomial Frequency Distribution	6
Poisson Frequency Distribution	7
Normal Frequency Distribution	8
Sample Average Distribution	9
Other Distributions	10
Transformations of Data	11

2. Referenced Documents

2.1 *ASTM Standards:* [3]

D 123 Terminology Relating to Textiles
D 4392 Terminology for Statistically Related Terms
E 456 Terminology Relating to Quality and Statistics

2.2 *ASTM Adjuncts:*

TEX-PAC[4]

NOTE 1—Tex-Pac is a group of PC programs on floppy disks, available through ASTM Headquarters, 100 Barr Harbor Drive, Conshohocken, PA 19428, USA. Many of the transformations described in this Guide can be made using a program in this adjunct.

3. Terminology

3.1 *Definitions:*

3.1.1 *Bernoulli distribution*—see *binomial distribution.*

3.1.2 *binomial distribution*, *n*—the frequency distribution which has the probability function:

$$P(r) = (n!/[r!\,(n-r)!])p^r q^{n-r} \qquad (1)$$

where:
$P(r)$ = the probability of obtaining exactly r “successes” in n independent trials,
p = the probability, constant from trial to trial, of obtaining a “success” in a single trial, and
q = $1 - p$.

(Syn. *Bernoulli distribution*)

3.1.3 *distribution*—see *frequency distribution of a sample* and *frequency distribution of a population.*

3.1.4 *frequency distribution*, *n—of a population,* a function that, for a specific type of distribution, gives for each value of a random discrete variate, or each group of values of a random continuous variate, the corresponding probability of occurrence.

3.1.5 *frequency distribution*, *n—of a sample,* a table giving for each value of a discrete variate, or for each group of values of a continuous variate, the corresponding number of observations.

3.1.6 *normal distribution*, *n*—the distribution that has the probability function:

$$f(x) = (1/\sigma)(2\pi)^{1/2}\exp[-(x-\mu)^2/2\sigma^2] \qquad (2)$$

where:
x = a random variate,
μ = the mean of the distribution, and
σ = the standard deviation of the distribution.

(Syn. *Gaussian distribution*, *law of error*)

3.1.7 *Poisson distribution*, *n*—the distribution which has as its probability function:

$$P(r) = e^{-\mu}\mu^r/r! \qquad (3)$$

[1] This guide is under the jurisdiction of ASTM Committee D13 on Textiles and is the direct responsibility of Subcommittee D13.93 on Statistics.

Current edition approved July 15, 1991. Published September 1991. Originally published as D 4686 – 87. Last previous edition D 4686 – 90.

[2] Shapiro, Samuel S., *How to Test Normality and Other Distribution Assumptions*, American Society for Quality Control, Milwaukee, WI, 1980. Vol. 3 of the series, *The ASQC Basic References in Quality Control: Statistical Techniques*, Edward J. Dudewitz, ed.

[3] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For*Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[4] PC programs on floppy disks are available through ASTM. For a 3½ inch disk request PCN:12-429040-18, for a 5¼ inch disk request PCN:12-429041-18.

where:

$P(r)$ = probability of obtaining exactly r occurrences of an event in one unit, such as a unit of time or area,
μ = both mean and variance of distribution, and
e = base of natural logarithms.

3.1.8 *probability function*, *n—of a continuous variate,* the mathematical expression whose definite integral gives the probability that a variate will take a value within the two limits of integration.

3.1.9 *probability function*, *n—of a discrete variate,* the mathematical expression which gives the probability that a variate will take a particular value.

3.1.10 *transformation*, *n*—the change from one set of variables, x, to another set, y, by the use of a function, $y = f(x)$.

3.1.11 For definitions of textile terms used in this guide, refer to Terminology D 123. For definitions of other statistical terms used in this guide, refer to Terminology D 4392, Terminology E 456, or appropriate textbooks on statistics.

4. Significance and Use

4.1 In measuring, testing, and experimenting, statistical tests are made to determine whether the observed effect of the introduction of a factor is real or simply due to chance. The appropriate statistical test to use depends on the kind of distribution used to model the data. Distribution identification is useful in selecting the most powerful statistical test.

NOTE 2—There are statistical tests which can be used for data for which a parametric distribution cannot be selected. But these non-parametric tests do not discriminate as well as the distribution-dependent tests.

4.2 For certain types of data, a transformation can be made which will make it possible to use the hypothesis that the normal distribution is a suitable model for the transformed data. When this hypothesis can be made, the analysis of the data is made much easier.

5. Key to Distributions

5.1 Table 1 is a key to the identification of frequency distributions. The table consists of a series of pairs of statements. The statement in the pair that is true either (*1*) directs the user to another pair of statements to solicit additional information or (*2*) identifies the distribution.

6. Binomial Frequency Distribution

6.1 A binomial distribution is the probability distribution of a binomial experiment. This fact provides a means for its identification. Such an experiment has the following four characteristics:

6.1.1 The experiment consists of n independent, identical trials. Each trial is conducted independently.

6.1.2 There are only two possible outcomes on each trial. These may be success and failure, good or bad, pass or fail.

6.1.3 The probability of occurrence of one of the two possible outcomes remains the same from trial to trial. This probability is usually denoted by p and the probability of nonoccurrence by $q = (1 - p)$.

6.1.4 The experimenter is interested in r, the number of successes observed during the n trials.

6.2 Counting the number of successes when trying to ignite eight specimens of carpet is one trial from a binomial experiment. Data from a series of such trials will produce a binomial distribution, provided all of the trials are made on essentially the same kind of material and in the same manner.

TABLE 1 Key to Frequency Distributions[A,B]

	Section
1a. Variates are discrete. (2)	
1b. Variates are continuous. (8)	
2a. Variates are counts of events. (3)	
2b. Variates are on an arbitrary scale . . . . .UNIDENTIFIED	10, 11
3a. Counts are of one of a pair of mutually exclusive events per unit. (4)	
3b. Counts are of other events per unit. (6)	
4a. Experiment consists of n identical trials, each conducted independently. (5)	
4b. Experiment otherwise . . . . .UNIDENTIFIED	10, 11
5a. Probability of occurrence of counted event constant from trial to trial . . . . BINOMIAL	6
5b. Probability otherwise . . . . .UNIDENTIFIED	10, 11
6a. Probability of event occurrence constant from unit to unit. (7)	
6b. Probability otherwise . . . . .UNIDENTIFIED	10, 11
7a. Number of events independent from unit to unit . . . . POISSON	7
7b. Number of events otherwise . . . . .UNIDENTIFIED	10, 11
8a. Variates individual observations. (9)	
8b. Variates sample averages . . . . NORMAL	9
9a. Distribution symmetrical. (10)	
9b. Distribution otherwise . . . . .UNIDENTIFIED	10, 11
10a. Distribution unimodal. (11)	
10b. Distribution multimodal . . . . .UNIDENTIFIED	10, 11
11a. Distribution bell shaped . . . . NORMAL	8
11b. Distribution shape otherwise . . . . .UNIDENTIFIED	10, 11

[A] See Section 5 for instructions for using this table.
[B] See Note 1 concerning unidentified distributions.

6.3 *Approximation of the Binomial Distribution by a Normal Distribution*—Binomial data can be transformed to a near normal distribution. (See Section 11.) Under a certain condition, the binomial distribution can be approximated by the normal distribution without transformation. This condition is met when the interval, $p \pm 3\sigma$, or $p \pm 3\sqrt{(p(1 - p))}$ does not contain zero or one.

7. Poisson Frequency Distribution

7.1 A Poisson distribution is the probability distribution of a Poisson experiment. This fact provides a means for its identification. Such an experiment has the following four characteristics:

7.1.1 The experiment consists of counting the number of times a particular event occurs in a given unit. The unit may be time, area, volume, a single item, or any other unit of measure.

7.1.2 The probability that an event occurs in a given unit is the same for all units.

7.1.3 The number of events occurring in one unit is independent of the numbers occurring in other units.

7.1.4 The number of potential events per unit is essentially unlimited.