HANDBOOK OF
PROBABILITY AND STATISTICS WITH TABLES

BY

RICHARD STEVENS BURINGTON, Ph. D.

AND

DONALD CURTIS MAY, Jr., Ph. D.

HANDBOOK OF

PROBABILITY AND STATISTICS WITH TABLES

BY

RICHARD STEVENS BURINGTON, Ph.D.

AND

DONALD CURTIS MAY, Jr., Ph.D.

HANDBOOK PUBLISHERS, INC.

SANDUSKY, OHIO

PREFACE

The rapid growth of the applications of the theory of probability and statistics to a vast variety of fields has resulted in the publication of many books and articles on the subject. It is the purpose of this book to provide in a small, convenient size a pocket handbook of probability and statistics, sufficiently comprehensive to fill a broad variety of needs, and yet simple enough in its structure to permit use by people having a wide difference of training, background and experience.

This book is intended as a convenient summary of theory, working rules and tabular material useful in practical problems in probability and statistics. It brings together information which is not otherwise readily available in simple form except by reference to numerous journals, tables and treatises on the subject. The book is divided into two main parts. The first part includes a summary of the more important formulas and definitions of elementary statistics and probability theory. The second part consists of tables of distributions and other quantities of frequent use in statistical work.

The handbook has been compiled to meet the needs of students and workers in statistics, probability, engineering, physics, chemistry, the various natural and social sciences, operations research and analysis, education, business, and other fields in which statistical calculations and methods can be used. Readers without detailed statistical training should find this volume a sufficient guide for the more commonly met statistical aspects of their studies. Those with statistical training should find it a convenient summary of the material most often needed.

This book, devoted primarily to probability and statistics, is a companion to Burington's *Handbook of Mathematical Tables and Formulas*, which contains general mathematical reference material. The two together should serve the needs of students and others using statistical and mathematical procedures in a wide variety of fields.

The authors are indebted to Professor Sir Ronald A. Fisher, Cambridge, and to Messrs. Oliver and Boyd, Ltd., Edinburgh, for permission to reprint Tables XI, XII, XIV from their book *Statistical Methods for Research Workers;* to Professor N. Arley of the Universitetets Institut for Teoretisk Fysik, Copenhagen, for Table 13.76.1; to the British Standards Institution for Table 13.58.1; to Professor G. W. Snedecor of the Iowa State College, and the Iowa State College Press, for Tables X, 14.57.1

and 14.57.2; to Professor E. S. Pearson and *Biometrika* for Table 13.86.1; and to the American Society for Testing Materials for Table 17.43.2.

The authors also wish to acknowledge the extensive help of Jennet Burington in typing the final draft of the manuscript and carrying through the many arduous tasks necessary to the successful completion of this volume, and the help of Helen May in typing the tabular material which forms the latter part of the book.

R. S. B.
D. C. M.

Arlington, Virginia, 2 November 1952.

TABLE OF CONTENTS

Chapter I

INTRODUCTION

1.1. **Introduction.** Modern mathematical and statistical methods are playing a most important and growing role in many fields of endeavor, such as: engineering; agriculture; business, insurance, economics; the physical sciences such as chemistry, physics, astronomy; biological sciences, medical research; sociology and psychology; transportation; industrial studies, etc. In nearly all fields there arise problems where precise measurements or observations are impossible to make, or where events are not exactly reproducible or predictable. The analysis of such situations leads to considerable uncertainty. Statistical methods lend themselves to the analysis of such situations and furnish a means of describing and indicating trends, or expectations, often with an associated degree of reliability. Statistics makes use of scientific methods of collecting, analyzing, and interpreting data.

1.2. Modern *statistical method* is a science which deals with such problems as: (1) How to plan a program for obtaining data so that reliable conclusions can be made from the data so obtained. (2) How to analyze the data obtained. (3) What valid conclusions can be drawn properly from the data? (4) To what extent are the conclusions reliable?

To outline all known statistical methods in use today would lead deeply into theory, and into the subject matter of many fields, and would require several large volumes. However, the principal elementary fundamental concepts and methods can be outlined in a reasonably small space. It is the purpose of this Handbook to furnish such an outline, in elementary and compact form, and in a form that can be used by workers in many fields.

1.3. **Quantitative statistical observations.** A sequence or set of measurements, or observations, made on a set of objects in a specified set, or population, of objects is known as *quantitative statistical observations*. When the observations are made on only some of the objects of the population the set of observations is called a *sample*. Populations, or samples, may be *finite*, or *indefinitely large*.

To condense and describe samples of quantitative data, *frequency* *distributions* are used, and certain parameters, called *statistics*, are calculated from the samples used to describe the frequency distributions. Likewise, populations and their distributions are described by *population parameters*.

As a rule, one only has a sample from a population and consequently does not have data for the entire population. A common procedure is to use statistics calculated from the sample frequency distribution to estimate likely values of the population parameters. For very large samples the statistics of properly chosen samples will have values quite close to those of the corresponding population parameters. For small samples the discrepancies are apt to be larger, and the business of predicting values of the population parameters from the sample statistics becomes involved and has to be resolved by the use of probability theory.

1.4. **Qualitative statistical observations.** A sequence or set of observations in which each observation in the set (and population) belongs to one of several *mutually exclusive* classes (perhaps non-numerical) is called *qualitative statistical observations.* (For example, an inspector at the exit of a railway car equipped with radio might ask disembarking passengers whether or not they like the radio (just heard), might get a sequence of 50 answers, such as: yes, no, yes, yes, no comment, yes, no, · · · . The answers are qualitative.)

The problems of how to gather, analyze, and draw conclusions from qualitative observations are similar to those for quantitative observations.

In some cases the order of the observations is of principal interest; and time is often a leading variable. This is true both in qualitative and quantitative observations.

1.5. **Probability theory and statistics.** The theories of probability and statistics have much in common. In one sense the theory of statistics rests on the theory of probability, the latter often furnishing much of the basic or ideal structure for the former.

1.6. **On the applications of statistics.** The most useful and successful applications of statistics require knowledge of both statistical methods and the subject matter to which the methods are applied. Unless the methods used are well anchored in the field of application the results of the statistical analyses may prove to be misleading. In the applications of statistics and the theory of probability it is often wise to draw upon what might be termed the "dynamical approach" (e.g., the appropriate engineering, economical, biological, or physical approach) appropriate to the fields under study. This is particularly true where small samples are involved, and where to gain a sizable sample may require undue expense and effort.

1.7. **Graphical presentations of data.** There are many ways of presenting graphically various types of data. Fig. 1.7.1 is an example of a *horizontal bar chart*; Fig. 1.7.2, a *vertical bar chart*; Fig. 1.7.3, a *pie chart*; Fig. 1.7.4, a *silhouette chart*; Fig. 1.7.5, a *semi-logarithmic chart*; Fig. 1.7.6, a *logarithmic chart*; Fig. 1.7.7, a *curve graph*; an *area diagram* (e.g., Fig. 1.7.3); Fig. 1.7.8, a *solid diagram*; Fig. 3.3.2, a *histogram*. The reader is familiar with many other types of graphs; some of which make use of color, dots, cross hatchings, shadings, and the like.

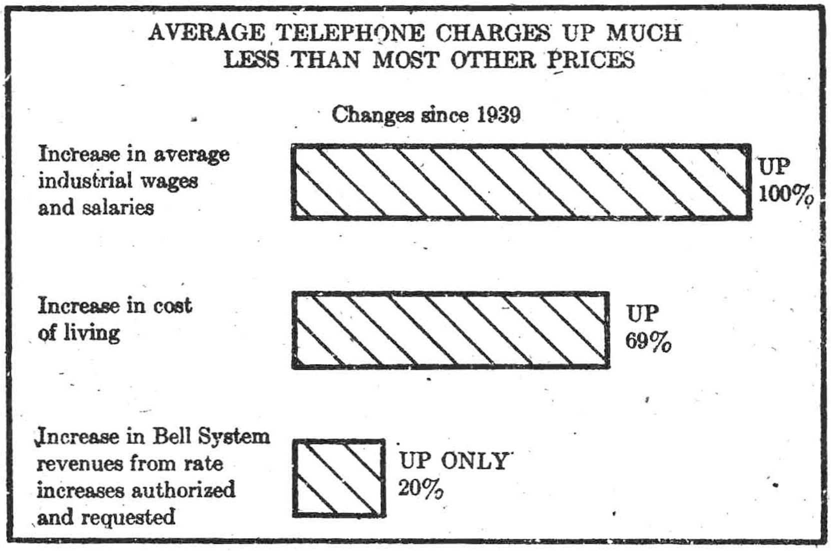


FIGURE 1.7.1

RELATIVE INCREASE IN EMPLOYMENT COSTS AND NUMBER OF EMPLOYEES

1939 = 100%

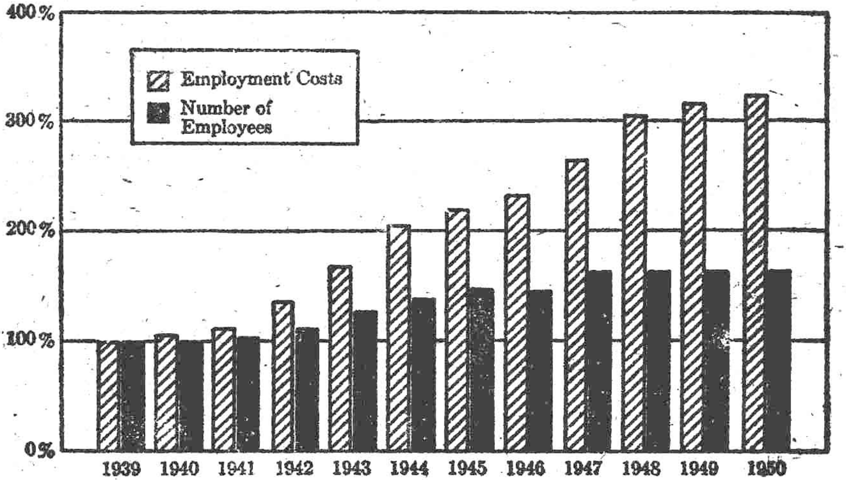


FIGURE 1.7.2

Income and Outgo for 1951

Total Revenues

$370,000,000

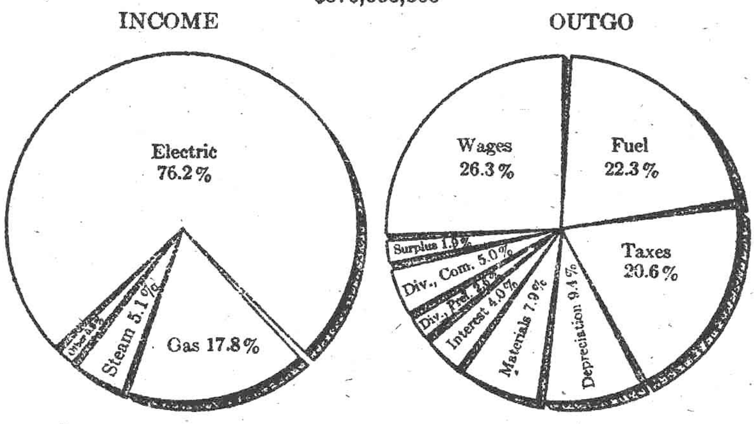
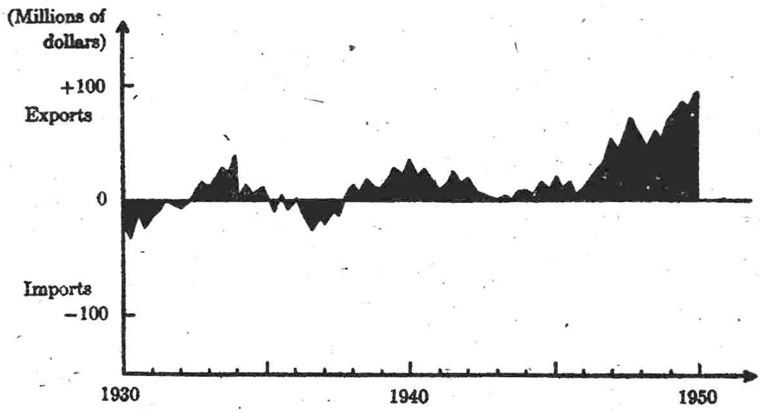
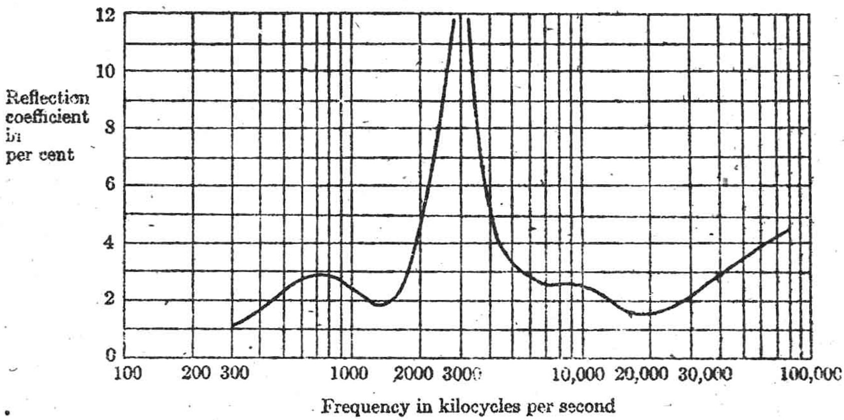


FIGURE 1.7.3



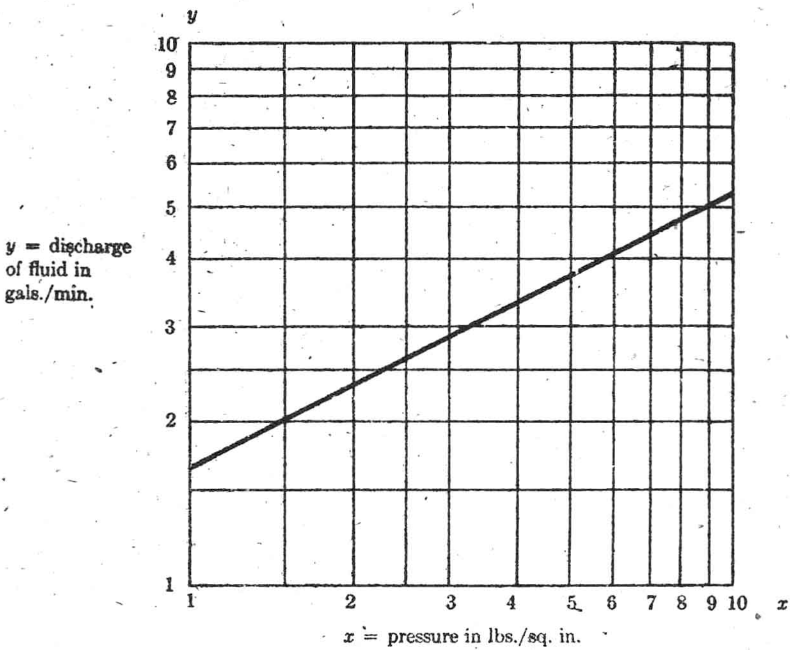
Movements of Commodity A from the U. S.
1930–1950

FIGURE 1.7.4



Difference Between Filter and Line Impedance Expressed as a Reflection Coefficient

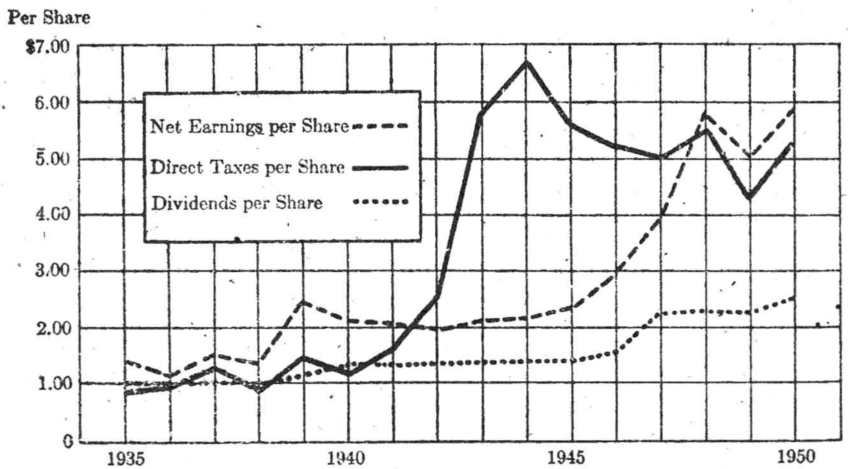
FIGURE 1.7.5



Fluid Discharge as Function of Pressure

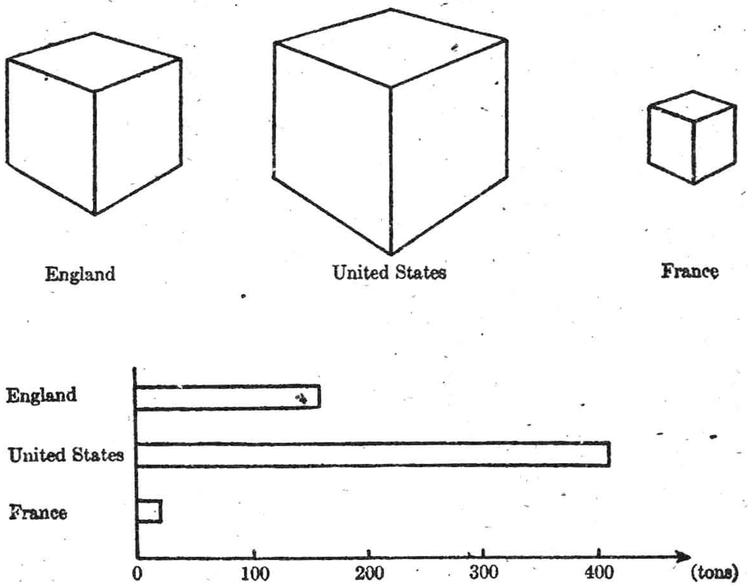
$y = 1.65x^{0.505}$ (Plotted on logarithmic scales)

FIGURE 1.7.6



Earnings, Dividends and Direct Taxes per Share of Common Stock

FIGURE 1.7.7



Stock of Mineral M on Hand, in Tons, December 31, 1939
Shown in bar chart form and solid diagram. Volume of solid shown corresponds to length of bar.

FIGURE 1.7.8

Chapter II

CERTAIN DEFINITIONS USED IN STATISTICS

2.1. The term *variate* is often used in statistics and probability theory to denote a *variable*. Variates may be discrete or continuous.

Measures of location

2.2. Let $x_1, x_2, \cdots, x_k$ be a set of numbers or variates. An *arithmetic mean*, or *average*, $\bar{x}$ of this set is

$$(2.2.1) \qquad \bar{x} = (1/k) \sum_{i=1}^{k} x_i = (x_1 + x_2 + \cdots + x_k) \div k.$$

If the values of $x_1, \cdots, x_k$ occur with corresponding frequencies $f_1, f_2, \cdots, f_k$, respectively, the *weighted arithmetic mean*, or *weighted average*, $\bar{x}$ is

$$(2.2.2) \qquad \bar{x} = (1/n) \sum_{i=1}^{k} f_i x_i,$$

where $n = \sum_{i=1}^{k} f_i$. Each $f_i \geqq 0$. The values $f_1, \cdots, f_k$ are called *weighting factors*; n is the *total weight*.

The quantity $x_i - \bar{x}$ is called the *deviation*, or *error*, of x_i with respect to $\bar{x}$; $| x_i - \bar{x} |$ is the *absolute deviation*, or *absolute error* of x_i.

The *least*, L, and *greatest*, G, of the variates $x_1, \cdots, x_k$ define the *range* $(G - L)$ of the variate.

2.3. The *root-mean-square* (R.M.S.) value of $x_1, \cdots, x_k$ is

$$(2.3.1) \qquad \left[\left(\sum_{i=1}^{k} f_i x_i^2\right) \div n\right]^{1/2}.$$

The *geometric mean* (G.M.) of n positive numbers $x_1, \cdots, x_n$ is the n^{th} root of their product, that is,

$$(2.3.2) \qquad (x_1 \, x_2 \cdots x_n)^{1/n}.$$

The *harmonic mean* (H.M.) of n positive numbers $x_1, \cdots, x_n$ is the reciprocal of the arithmetic mean of the reciprocals of the values, that is,

$$(2.3.3) \qquad n \div \sum_{i=1}^{n} (1/x_i).$$

If the positive number x_i appears f_i times, $i = 1, \cdots, k$, and $n = \sum_{i=1}^{k} f_i$, then the *geometric mean* is

(2.3.4) $$[x_1^{f_1} x_2^{f_2} \cdots x_k^{f_k}]^{1/n}.$$

2.4. **Mode** (Mo). A value of the variable x which occurs most frequently is called a *mode*. A mode (Mo) for $x_1, x_2, \cdots, x_n$ is a value of x_i having a maximum frequency f_i. If $x_1, \cdots, x_n$ all have the same frequencies, a mode is a value about which $x_1, \cdots, x_n$ "cluster most densely". A frequency distribution may have one or more than one mode. In certain cases it is difficult to give a completely satisfactory definition of mode.

2.5. **Median.** The *median* is defined as a value which is greater than half the variates and less than the other half, the median value being selected as follows. Let $x_1, x_2, \cdots, x_n$ be a set of real numbers arranged in order of magnitude algebraically,

$$x_1 \leqq x_2 \leqq \cdots \leqq x_n.$$

When n is odd, the median is x_h, where $n = 2h - 1$. When n is even, the median is not uniquely defined unless $x_h = x_{h+1}$, when $n = 2h$, in which case the median is this common value. In case the median is not uniquely defined, it is conventional to take for the median $(x_h + x_{h+1}) \div 2$, when $n = 2h$.

Moments

2.6. **Moments about the origin.** The r^{th} *moment about the origin* $x = 0$ of the numbers $x_1, \cdots, x_k$, having a weighted mean $\bar{x}$, where f_i is the frequency of x_i and $n = \sum_{i=1}^{k} f_i$, is

(2.6.1) $$\nu_r = (1/n) \sum_{i=1}^{k} f_i x_i^r. \qquad (r = 1, 2, \cdots)$$

To indicate that ν_r is calculated for the variable x, the notation $\nu_{r:x}$ is often used.

2.7. **Moments about the mean $\bar{x}$.** The r^{th} *(central) moment* μ_r *about the weighted mean* $\bar{x}$ of the numbers $x_1, x_2, \cdots, x_k$ is

(2.7.1) $$\mu_r = (1/n) \sum_{i=1}^{k} f_i (x_i - \bar{x})^r.$$

2.8. Relations between μ_r and ν_r.

(2.8.1)
$$\mu_0 = 1, \qquad \mu_1 = 0,$$
$$\mu_2 = \nu_2 - \nu_1^2 = \sigma^2. \qquad (\mu_2 \text{ is called the } \textit{variance})$$
$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3.$$
$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.$$
$$\vdots \qquad \qquad \vdots$$
$$\mu_k = \sum_{r=0}^{k} C_r^k \nu_{k-r}(-\nu_1)^r, \quad \text{where } C_r^k = k!/(k-r)!\,r!.$$

(2.8.2)
$$\nu_0 = 1, \qquad \nu_1 = \bar{x}.$$
$$\nu_2 = \mu_2 + \nu_1^2.$$
$$\nu_3 = \mu_3 + 3\mu_2\nu_1^1 + \nu_1^3.$$
$$\nu_4 = \mu_4 + 4\mu_3\nu_1^1 + 6\mu_2\nu_1^2 + \nu_1^4.$$
$$\vdots \qquad \qquad \vdots$$
$$\nu_k = \sum_{r=0}^{k} C_r^k \mu_{k-r}\nu_1^r.$$

2.9. Moments about arbitrary point. The r^{th} *moment about an arbitrary value* x_0 of x is

(2.9.1)
$$\mu_r(x_0) = (1/n) \sum_{i=1}^{k} f_i(x_i - x_0)^r.$$

$\mu_2(x_0)$ is called the *mean square deviation* from $x = x_0$, and $\sqrt{\mu_2(x_0)}$ is called the *root mean square deviation* from x_0. The *order* of the moment μ_r is r.

The function $\mu_2(x_0)$, treated as a function of x_0, has a minimum value σ^2 when $x_0 = \bar{x}$.

2.10. Absolute moments about x_0. The r^{th} *absolute moment about the arbitrary value* x_0 is

(2.10.1)
$$(1/n) \sum_{i=1}^{k} f_i \mid x_i - x_0 \mid^r.$$

This function is commonly used with $x_0 = \bar{x}$ or $x_0 = 0$.

2.11. Variance. The *variance* of $x_1, \cdots, x_k$ is defined as

(2.11.1) $$\sigma^2 = \mu_2(\bar{x}) = \mu_2 = (1/n) \sum_{i=1}^{k} f_i(x_i - \bar{x})^2.$$

σ is the *standard deviation*. Also,

(2.11.2) $$\sigma^2 = (1/n) \sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2 = \nu_2 - \bar{x}^2.$$

The relation

(2.11.3) $$\sigma^2 = \mu_2(x_0) - (\bar{x} - x_0)^2$$

is often useful since it enables one to calculate σ^2 from the second moment about any convenient value x_0 of x.

2.12. Mean deviation, M.D.$|_0$, from an arbitrary point x_0 is:

$$(1/n) \sum_{i=1}^{k} f_i \mid x_i - x_0 \mid .$$

The value of x_0 for which M.D.$|_0$ is a minimum is the median.

2.13. Mean (absolute) deviation, M.D., from the mean is:

$$(1/n) \sum_{i=1}^{k} f_i \mid x_i - \bar{x} \mid .$$

This is also called the *mean absolute error* (*m.a.e.*).

2.14. Factorial moments. The r^{th} *factorial moment about the origin* $x = 0$ of the numbers $x_1, \cdots, x_k$, is

(2.14.1) $$\nu_{[r]} = (1/n) \sum_{i=1}^{k} f_i x_i^{[r]}, \qquad (r = 1, 2, \cdots),$$

where

(2.14.2) $$x_i^{[r]} = x_i(x_i - 1) \cdots (x_i - r + 1). \qquad (r = 1, 2, 3, \cdots)$$

The r^{th} *central factorial moment about the weighted mean* $\bar{x}$ is

(2.14.3) $$\mu_{[r]} = (1/n) \sum_{i=1}^{k} f_i(x_i - \bar{x})^{[r]},$$

where

(2.14.4) $$(x_i - \bar{x})^{[r]} = (x_i - \bar{x})(x_i - \bar{x} - 1) \cdots (x_i - \bar{x} - r + 1).$$
$$(r = 1, 2, \cdots)$$

The relation

$$(2.14.5)\qquad \sigma^2 = (1/n)\sum_{i=1}^{k} f_i x_i(x_i - 1) - \bar{x}(\bar{x} - 1) = \nu_{[2]} - \bar{x}(\bar{x} - 1)$$

is often more convenient to use than (2.11.2).

Factorial moments are particularly useful in those cases where the variable x takes on discrete values spaced at equal intervals (e.g., when the x_i are positive integers, and when $r \leq x_i$ in (2.14.1), and $r \leqq x_i - \bar{x}$ in case (2.14.4)). Such moments play a role in the theory of interpolation and curve fitting. Some writers define $x^{[0]} = 1$.

Relations between $\nu_{[r]}$ and moments ν_r about $x = 0$.

$$(2.14.6)\qquad \begin{aligned} \nu_1 &= \nu_{[1]}, & \nu_2 &= \nu_{[2]} + \nu_{[1]}, \\ \nu_3 &= \nu_{[3]} + 3\nu_{[2]} + \nu_{[1]}, & \nu_4 &= \nu_{[4]} + 6\nu_{[3]} + 7\nu_{[2]} + \nu_{[1]}. \end{aligned}$$

2.15. Change of units. To indicate that μ_r is calculated for the variable x, the notation $\mu_{r:x}$ is sometimes used for μ_r. Thus,

$$(2.15.1)\qquad \mu_{r:x} = (1/n)\sum_{i=1}^{k} f_i(x_i - \bar{x})^r, \qquad \mu_{r:u} = (1/n)\sum_{i=1}^{k} f_i(u_i - \bar{u})^r.$$

If the change of variables, or units, is given by

$$(2.15.2)\qquad u = (x - x_0)/c,$$

then

$$(2.15.3)\qquad \mu_{r:x} = c^r \mu_{r:u}.$$

If $\sigma_x = \sqrt{\mu_{2:x}}$ is the standard deviation expressed in terms of x, and $\sigma_u = \sqrt{\mu_{2:u}}$ is the standard deviation in terms of u,

$$(2.15.4)\qquad \sigma_x = c\sigma_u.$$

Note: μ_r remains unchanged under any transformation of the form $u = x - x_0$.

2.16. Standard units. Deviations from the mean $\bar{x}$ are sometimes measured in units of the standard deviation σ_x, called *t units*, where

$$(2.16.1)\qquad t = (x - \bar{x})/\sigma_x.$$

The r^{th} *moment in standard t units about* $t = 0$ is

$$(2.16.2)\qquad \alpha_r = (1/n)\sum_{i=1}^{k} f_i t_i^r.$$