STATISTICAL FUNCTIONS and FORMULAS

A SOURCE OF SIMPLIFIED DERIVATIONS BASED ON ELEMENTARY MATHEMATICS

MYERS & ENRICK

Statistical Functions and Formulas

A SOURCE OF SIMPLIFIED DERVIATIONS BASED ON ELEMENTARY MATHEMATICS

Buddy L. Myers
Norbert L. Enrick
Kent State University

ROBERT E. KRIEGER PUBLISHING CO.
Huntington, New York 1981

Original Edition 1981

Based on original title STATISTICAL FUNCTIONS: A Source of Practical Derivations Based on Elementary Mathematics

Printed and Published by
ROBERT E. KRIEGER PUBLISHING COMPANY, INC.
645 New York Avenue
Huntington, New York 11743

Printed in the United States of America

Myers, Buddy L.
Statistical functions and formulas.

First ed. published in 1970 under title:
Statistical functions.
Bibliography: p.
Includes index.
1. Mathematical statistics--Tables.
I. Enrick, Norbert Lloyd, 1920- joint author.
II. Title.
QA276.25.M9 1981 519.5 80-22321
ISBN 0-89874-286-2

STATISTICAL FUNCTIONS and FORMULAS

A Source of Simplified Derivations Based on Elementary Mathematics

Foreword

Increasingly, professional people, students and teachers in the social and managerial sciences are becoming dissatisfied with statistical texts and reference works that give the important functions and formulas without deriving them. Yet, when resort is had to sources that do provide derivations, the amount of mathematical sophistication presumed is hopelessly beyond most people's background.

The present work derives the principal statistical functions and formulas in a methodical, detailed and gap-free manner, tailored for those who have a background of elementary algebra and basic calculus. To this end, derivations are presented in detail, providing the many intermediate steps that are disdained in more mathematical texts; as a result the work can be comprehended readily and quickly. The reader no longer needs to do his own mathematics to fill any gaps.

The advantages inherent in understanding derivations are considerable. Formulas and functions are recognized in terms of their origins, and a deeper appreciation of statistical theory, the rationale of its applications and the nature of limitations in methodologies is gained.

Buddy L. Myers
Norbert L. Enrick

Table of Contents

Chapter 1

Summations

The process of taking a sum is a mathematical one, not specifically related to probability and statistics. Nevertheless, it is fundamental to the derivation of statistical functions.

SUMMATION OF DISCRETE DATA

Summation of discrete data represents the simplest type of addition process. It is denoted by the symbol Σ for "summation of" (a capital Greek sigma). We may sum a number of constants or a number of variables, or a combination of constant times variable. We also bring a few special commonly needed sums which will be needed later in variance and regression analysis.

SUMMATION OF CONTINUOUS DATA

When a continuous variable is summed, we refer to the process as integration. It is indeed from this point of view that the methods of the integral calculus have evolved. There is of course still another aspect of equal importance, which views integration as the process of finding an anti-derivative (the inverse of differentiation). The symbol for integration is an elongated S, hence $\int$.

Integration is a limiting form of summation, as the intervals Δ between discrete values x_i shrink. Thus, for a function $f(x)$ we have:

$$\int_a^b f(x)dx = \lim \sum_{i=1}^{n} f(x_i)\,\Delta x_i$$

n approaches infinity as the term involving delta goes towards zero. Hence, $f(x_i)\Delta x_i$ represents a series of quantities, each of which is a differential expression or element of a total quantity. The relationship shown is the *basic theorem of integral calculus.* The *fundamental theorem of calculus* states that differentiation of an indefinite integral $F(x)$ at x yields $f(x)$ at x.

SPECIAL PRODUCTS

A series of multiplications is shown by the symbol Π, a capital Greek *Pi.* When a product refers to a constant times a variable, the result is an expression containing the constant raised to the appropriate power and multiplied by the product of the variables.

SUMS OF SERIES

Certain sums of series are of special interest, since they are needed in the derivation of some important statistical functions. Particularly significant in this regard are (1) the sum of the first n positive integers and (2) the sum of the squares of these integers. Other, more commonly known sums, will be found in Appendix D.

SUMMARY

Although summation processes are in the domain of mathematics, certain aspects of their application are of special interest in statistical derivations. We examine summations of both discrete and continuous data.

SUMMATION OF A CONSTANT, DISCRETE DATA

Formula: $\Sigma k = nk$

where k is a constant, n represents the number of k's and Σ represents the simple addition of all the k's.

In words: The sum of a number of constants equals the number of constants times the value of the constant.

Derivation: We first assume that the n values of k_i, with $i = 1, 2, \ldots, n$, are defined as shown below.

$$\sum_{i=1}^{n} k_i = k_1 + k_2 + \cdots + k_n \qquad (1)$$

But if these values are the same, k is a constant, and

$$k_1 = k_2 = \cdots = k_n = k \qquad (2)$$

With all k's alike, we no longer use subscripts to distinguish them. Substitution in (1) now yields:

$$\Sigma k = k + k + \cdots + k \qquad (3)$$

Factoring out k, we obtain

$$\sum_{i=1}^{n} k = k(1 + 1 + \cdots + 1) \qquad (4)$$

There are n such values of unity in the parentheses. Hence,

$$\Sigma k = nk \qquad (5)$$

SUMMATION OF A CONSTANT TIMES A VARIABLE, DISCRETE DATA

Formula: $\sum_{i=1}^{n} kx_i = k \sum_{i=1}^{n} x_i$

where k is a constant and n represents the number of values of the variable x.

In words: The sum of a constant times a variable equals the constant times the sum of the values of the variable.

Derivation: In terms of generally used definitions, the left-hand member of the formula can be expressed as shown below:

$$\sum_{i=1}^{n} kx_i = kx_1 + kx_2 + \cdots + kx_n \qquad (1)$$

The common element, k, of each term is now factored out.

$$= k(x_1 + x_2 + \cdots + x_n) \qquad (2)$$

Re-applying the aforementioned general definition

$$= k \sum_{i=1}^{n} x_i \qquad (3)$$

PRODUCT OF A SERIES OF CONSTANTS

Formula: $\Pi k = k^n$

where k is a constant and n represents the number of constants.

In words: The product of a number of n constants equals the constant raised to the nth power.

Derivation: The left-hand side of the formula is quite generally defined as shown below:

$$\prod_{i=1}^{n} k_i = k_1 \cdot k_2 \ldots k_n \qquad (1)$$

But if the various values are the same, k is a constant such that:

$$k = k_1 = k_2 = \cdots = k_n \qquad (2)$$

Substitution in (1) yields:

$$\Pi k = k \cdot k \cdot k \ldots k \tag{3}$$

Next, by a common rule of algebra,

$$\Pi k = k^n \tag{4}$$

PRODUCT OF A CONSTANT TIMES A VARIABLE

Formula: $\prod_{i=1}^{n} kx_i = k^n \prod_{i=1}^{n} x_i$

where k is a constant, x a variable and n the number of values of the variable.

In words: The product of a number, n, of values, each multiplied by the same constant, is the constant raised to the nth power multiplied by the product of the variables.

Derivation: The left-hand side of the formula is commonly defined as shown below:

$$\prod_{i=1}^{n} kx_i = kx_1 \cdot kx_2 \cdot \ldots \cdot kx_n \tag{1}$$

Since there are n constants, k, we can factor out k^n giving:

$$\prod_{i=1}^{n} kx_i = k^n \, (x_1 \cdot x_2 \cdot \ldots \cdot x_n) \tag{2}$$

Next, re-applying the definition in (1),

$$\prod_{i=1}^{n} kx_i = k^n \cdot \prod_{i=1}^{n} x_i$$

DOUBLE AND TRIPLE SUMS

Formula: Double and triple sums occur in the examples below:

$$\sum_{i}^{a}\sum_{j}^{b}(x_{ij} - \overline{x}\,..) = 0$$

$$\sum_{i}^{a}\sum_{j}^{b}\sum_{k}^{c}(x_{ijk} - \overline{x}\,...) = 0$$

General: The formulas state that the sum of the deviations, of individual values, such as x_{ij}, from their mean, $\overline{x}\,..$, is zero. We will check the statement and, in the process, demonstrate operations involving double and triple sums.

Derivations: We will examine the simple case:

$$\sum_{i=1}^{a}(x_i - \overline{x}.) = 0 \qquad (1)$$

Where $\overline{x}\,.$ is the arithmetic mean, $\sum_{i}^{a} x_i/a$, of the x_i's.

From the definition of a sum of a constant, and transposing,

$$\sum_{i=1}^{a} x_i = a\overline{x}. \qquad (2)$$

Dividing both sides by a,

$$\sum_{i=1}^{a} x_i/a = \overline{x} \qquad (3)$$

But by our definition of a mean, both sides of this equation are equal, thus proving equation (1).

For a double sum,

$$\sum_{i}^{a}\sum_{j}^{b}(x_{ij} - \overline{x}\,..) = 0 \qquad (4)$$

$$= \sum_{i}^{a}\sum_{j}^{b} x_{ij} - ab\overline{x}. \qquad (5)$$

where $ab\overline{x}\,..$ equals the double sum, thus demonstrating the validity of the formula shown above.

Divide through by b and transpose.

$$\sum_i^a \left(\sum_j^b x_{ij}/b \right) = a\overline{x}.. \tag{6}$$

But the expression in parentheses is $\overline{x}_i$ . . , so that (6) becomes:

$$\sum_i^a (\overline{x}_i .) = a\overline{x} .. \tag{7}$$

Now divide through by a, to obtain

$$\sum_i^a (\overline{x}_i ./a) = \overline{x} .. \tag{8}$$

But the left-hand side of (8) is, by definition, $\overline{x}$ . . , thus both sides of the equation are shown equal, thereby proving (4).

For a triple sum, we will merely indicate the sequence of equations. The steps run parallel to the two prior cases, and no further explanations are needed:

$$\sum_i^a \sum_j^b \sum_k^c (x_{ijk} - \overline{x} \ldots) = 0 \tag{9}$$

$$\sum \sum \sum x_{ijk} = abc\, \overline{x} \ldots \tag{10}$$

$$\sum_i \sum_j \left(\sum_k x_{ijk}/c \right) = ab\, \overline{x} \ldots \tag{11}$$

$$\sum_i \sum_j (\overline{x}_{ij} .) = ab\overline{x} \ldots \tag{12}$$

$$\sum_i \left(\sum_j \overline{x}_{ij}/b \right) = a\, \overline{x} \ldots \tag{13}$$

$$\sum_i \overline{x}_i ../a = \overline{x} \ldots \tag{14}$$

By definition of a mean, the left-hand side of (14) equals the right-hand side. Equation (9) is thus proved out.

THE SUM OF A SPECIAL PRODUCT

Formula: $\sum_{i<j}^{n} \sum_{j}^{n} x_i x_j = \frac{1}{2} \left[\left(\sum_{i}^{n} x_i \right)^2 - \sum_{i}^{n} x_i^2 \right] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} x_i x_j$

General: The formula shown will be quite valuable in future derivations, particularly for variance and regression analysis.

Derivations: We will use a simple case of $i = 1, 2, 3$ and $j = 1, 2, 3$, from which extensions to any magnitude ($i = 1, 2, 3, \ldots, n$; and $j = 1, 2, 3, \ldots, n$) are apparent. For the three terms above, separated by equal signs, we will demonstrate that each term gives identical results. Thus:

$$\sum_{i<j}^{3} \sum_{j}^{3} x_i x_j = \sum_{i=1}^{2} \sum_{j=2}^{3} x_i x_j \qquad (1)$$

$$= x_1 x_2 + x_1 x_3 + x_2 x_3 \qquad (2)$$

Next,

$$\left(\frac{1}{2}\right) \left[\left(\sum_{i=1}^{3} x_i \right)^2 - \sum_{i=1}^{3} x_i^2 \right]$$

$$= (^1/_2)\ [(x_1 + x_2 + x_3)^2 - (x_1^2 + x_2^2 + x_3^2)] \qquad (3)$$

$$= (^1/_2)\ [(x_1^2 + x_2^2 + x_3^2 + 2x_1 x_2 + 2x_1 x_3 + 2x_2 x_3) - (x_1^2 + x_2^2 + x_3^2)] \qquad (4)$$

$$= x_1 x_2 + x_1 x_3 + x_2 x_3 \qquad (5)$$

which agrees with (2). Finally,

$$\sum_{i=1}^{2} \sum_{j=2}^{3} x_i x_j = x_1 x_2 + x_1 x_3 + x_2 x_3 \qquad (6)$$

Thus, equation (2) = equation (5) = equation (6), thereby demonstrating the identity of the three terms under "Formula" above.

THE SUM $\sum_{i}^{a} n_i (\overline{x}_i . - \mu)$

Formula: $\sum_{i}^{a} n_i (\overline{x}_i . - \mu) = n(\overline{x} . . - \mu)$

This formula pertains to analysis of variance, where $(\overline{x}_i .)$ is the within-group sample average, each group having a sample size n_i; such that $n_1 = n_2 = \cdots = n_a$ and hence $(n_i)(a) = n$. Here a represents the number of sample averages used and μ is the population (grand) mean.

Derivation: In the formula above, expand $n_i(\overline{x}_i . - \mu)$ to $(n_i\overline{x}_i . - n_i\mu)$, giving

$$\left[\sum_{i}^{a} n_i\overline{x}_i . - \sum_{i}^{a} n_i\mu\right] \tag{1}$$

But n_i is a constant, as is μ. Hence

$$\sum_{i}^{a} n_i\mu = (an_i)\mu = n\mu \tag{2}$$

Now,

$$\sum_{i}^{a} n_i\overline{x}_i . = (n_1\overline{x}_1 . + n_2\overline{x}_2 . + \cdots + n_a\overline{x}_a .) \tag{3}$$

$$= (n_1 + n_2 + \ldots n_a) \left[\sum_{i=1}^{a} \overline{x}_i ./a\right] \tag{4}$$

But $n = (n_1 + n_2 + \ldots n_a)$, hence (4) becomes $n(\overline{x} . .)$, where the $\overline{x} . .$ denotes an average of a values x_i. Combine $n\mu$ and $n(\overline{x} . .)$ to find:

$$\sum_{i}^{a} n_i(\overline{x} .-\mu) = n(\overline{x} . .) - n\mu = n(\overline{x} . .-\mu) \tag{5}$$

which proves out the summation formula.

INTEGRAL OF A CONSTANT

Formula: $\int kdx = kx + c$

where k and c are different constants, x is the independent variable (a dependent variable y being assumed), the elongated S is an integration sign without limits and dx is the differential of x.

In words: The integral of a constant is the constant times the independent variable plus a further constant.

Derivation: The left-hand side of the formula asks us to find a function of $x, f(x) = y$, with the special property that its first differential is k.

We begin with

$$f(x) = y = kx + c \tag{1}$$

where the $kx + c$ is given in the formula.

The first derivative is:

$$y' = \frac{dy}{dx} = \frac{d}{dx}(kx + c) = \frac{d}{dx}(kx) + \frac{d}{dx}(c) \tag{2}$$

But since the derivative of the constant c is zero and the derivative of kx is k,

$$\frac{d}{dx}(kx + c) = \frac{dy}{dx} = k \tag{3}$$

Multiplying through by dx

$$dy = d(kx + c) = kdx \tag{4}$$

Integrating both sides

$$\int dy = \int d(kx + c) = \int kdx = k\int dx \tag{5}$$

Thus the integral of kdx turns out, indeed, to be $kx + c$.

Also y, being the sum of all differentials, dy, is:

$$y = \int kdx \tag{6}$$

INTEGRAL OF A VARIABLE

Formula: $\int xdx = {}^1\!/_2\, x^2 + c$

where x is the independent variable and c is a constant. The elongated S is an integration sign without limits and dx is the differential of x.

In words: Not applicable.

Derivation: The left-hand side of the formula asks us to find a function of