note
The date on the reference (SAP/QAPP-39) should be 1976

CRC

Standard Mathematical Tables

Twenty-fourth Edition

Editor of Mathematics and Statistics

WILLIAM H. BEYER, Ph.D.

Professor of Mathematics and Statistics
and Head of the Department of Mathematics
and Statistics, University of Akron, Akron, Ohio

SAMUEL M. SELBY

Editor Emeritus

published by:

CRC PRESS

18901 Cranwood Parkway • Cleveland, Ohio 44128

Library of Congress Card No. 30-4052
International Standard Book No. 0-87819-623-4

This book represents information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Every reasonable effort has been made to give reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Direct all inquiries and correspondence to
CRC PRESS, INC., 18901 Cranwood Parkway, Cleveland, Ohio 44128

DESCRIPTIVE STATISTICS

a) Ungrouped Data

The formulas of this section designated as *a*) apply to a random sample of size n, denoted by x_i, $i = 1, 2, \ldots, n$.

b) Grouped Data

The formulas of this section designated as *b*) apply to data grouped into a frequency distribution having class marks x_i, $i = 1, 2, \ldots, k$, and corresponding class frequencies f_i, $i = 1, 2, \ldots, k$. The total number of observations given by

$$n = \sum_{i=1}^{k} f_i$$

In the formulas that follow, c denotes the width of the class interval, x_o denotes one of the class marks taken to be the computing origin, and $u_i = \frac{x_i - x_o}{c}$. Then coded class marks are obtained by replacing the original class marks with the integers $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$ where 0 corresponds to class mark x_o in the original scale.

Mean (Arithmetic Mean)

$$a)\quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$b.1)\quad \bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{f_1x_1 + f_2x_2 + \cdots + f_kx_k}{n}$$

If data is coded

$$b.2)\quad \bar{x} = x_o + c\,\frac{\sum_{i=1}^{k} f_i u_i}{n}$$

Weighted Mean (Weighted Arithmetic Mean)

If with each value x_i is associated a weighting factor $w_i \geq 0$, then $\sum_{i=1}^{n} w_i$ is the total weight, and

$$a)\quad \bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n}$$

Geometric Mean

$$a)\quad \text{G.M.} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

In logarithmic form

$$\log(\text{G.M.}) = \frac{1}{n}\sum_{i=1}^{n} \log x_i = \frac{\log x_1 + \log x_2 + \cdots + \log x_n}{n}$$

b) $\text{G.M.} = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdots x_k^{f_k}}$

In logarithmic form

$$\log(\text{G.M.}) = \frac{1}{n}\sum_{i=1}^{k} f_i \log x_i = \frac{f_1 \log x_1 + f_2 \log x_2 + \cdots + f_k \log x_k}{n}$$

Harmonic Mean

a) $\text{H.M.} = \dfrac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{x_i}} = \dfrac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}}$

b) $\text{H.M.} = \dfrac{n}{\displaystyle\sum_{i=1}^{k} \frac{f_i}{x_i}} = \dfrac{n}{\dfrac{f_1}{x_1} + \dfrac{f_2}{x_2} + \cdots + \dfrac{f_k}{x_k}}$

Relation Between Arithmetic, Geometric, and Harmonic Mean

$\text{H.M.} \le \text{G.M.} \le \bar{x}$, (Equality sign holds only if all sample values are identical.)

Mode

a) A mode M_o of a sample of size n is a value which occurs with greatest frequency, i.e., it is the most common value. A mode may not exist, and even if it does exist it may not be unique.

b) $M_o = L + c\,\dfrac{\Delta_1}{\Delta_1 + \Delta_2}$,

where L is the lower class boundary of the modal class (class containing the mode),
Δ_1 is the excess of modal frequency over frequency of next lower class,
Δ_2 is the excess of modal frequency over frequency of next higher class.

Median

a) If the sample is arranged in ascending order of magnitude, then the median M_d is given by the $\dfrac{n+1}{2}$ *nd* value. When n is odd, the median is the middle value of the set of ordered data; when n is even, the median is usually taken as the mean of the two middle values of the set of ordered data.

b) $M_d = L + c\,\dfrac{\dfrac{n}{2} - F_c}{f_m}$,

where L is lower class boundary of median class (class containing the median),
F_c is the sum of the frequencies of all classes lower than the median class,
f_m is the frequency of the median class.

Empirical Relation Between Mean, Median, and Mode

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$