DOCUMENT RESUME

ED 054 936 SE 012 325

AUTHOR	Fuller, Julia M., Ed.
TITLE	Descriptive Statistics for the Health Professions. Guide: Computation, Arithmetic Mean.
INSTITUTION	Public Health Service (DHEW), Atlanta, Ga. Center for Disease Control.
PUB DATE	71
NOTE	76p.
AVAILABLE FROM	Superintendent of Documents, U. S. Government Printing Office, Washington, D.C. 20402, ($0.70-Stock Number 1723-0049)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Health Occupations Education; *Instructional Materials; Mathematical Models; *Mathematics; Professional Training; *Programed Instruction; Statistical Data; *Statistics; Study Guides

ABSTRACT

This booklet is a programmed self-instructional guide for the computation of arithmetic mean. Developed to serve the needs of public health professionals, it is not an exhaustive or technical treatment of statistics. It is limited, first, to descriptive statistics (tables, graphs, descriptive ratios, measures of central tendency, and measures of dispersion) and, second, to those concepts and techniques most needed by health professionals working routinely with basic statistical data. Intended primarily for public health nurses and sanitarians with a college degree or its equivalent, the guide is designed to be used when the need to compute and use the arithmetic mean arises. Several computational procedures for finding the arithmetic mean are provided, each with accompanying step-by-step examples. These include procedures for: ungrouped, discrete data; ungrouped continuous data; grouped (single value), discrete data; grouped (interval), discrete data; grouped (single value), continuous data; grouped (interval), continuous data (rounded); and grouped (interval), continuous data (non-rounded). Also presented is condensed, simplified reference to selecting the correct computational procedure when the proper procedure to use is not known. (BL)

ERIC

ED0 54956

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

H - PHS

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: SE | VT

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

SURPLUS DUPLICATE
4 AUG 25 1971
Copy

DESCRIPTIVE STATISTICS FOR THE HEALTH PROFESSIONS

GUIDE: COMPUTATION

ARITHMETIC MEAN

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service

SE 3325

SPECIFICATIONS

INSTRUCTIONAL OBJECTIVES

Given a list or table of values and following this *Guide,* the student will be able to use the appropriate procedure for computing the arithmetic mean.

PRIMARY TRAINEE POPULATION

Public Health nurses and sanitarians with a college degree or its equivalent.

SECONDARY TRAINEE POPULATION

1. Public Health veterinarians, physicians, dentists, and other similarly related Public Health workers with a college degree or its equivalent should also be able to use this *Guide* except that the examples used in this booklet will not be relevant to this group.
2. With proper motivation and some additional effort, Public Health nurses and sanitarians with a high school education should also be able to use this *Guide.*

INDIVIDUALIZATION PROVIDED

The student may proceed at his own best rate (there is no time limit).

APPROXIMATE TIME

To use a specific computational procedure, 10 to 30 minutes; to use all computational procedures in the *Guide,* 2 to 4 hours.

RESTRICTIONS, LIMITATIONS, AND SPECIAL CHARACTERISTICS

1. The student must be able to perform the basic mathematical functions required.
2. This booklet is a guide and must be used each time the mean is computed until the student is able, without help, to perform the behaviors as indicated in the *Guide.*

DESCRIPTIVE STATISTICS FOR THE HEALTH PROFESSIONS

GUIDE: COMPUTATION

ARITHMETIC MEAN

An Instructive Communication

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
HEALTH SERVICES AND MENTAL HEALTH ADMINISTRATION
Center for Disease Control
Atlanta, Georgia 30333

PRODUCED BY CDC TRAINING PROGRAM
Methods Development Branch

Instructive Communications Activity

Richard E. Lincoln, *Chief*
Virginia H. Eller, *Analyst-Writer*
Julia M. Fuller, *Editor*
Robert L. Reynolds, *Special Consultant*

Technical Advisor

Dr. James C. Terrell, *formerly Chief*
Biostatistics Unit, Georgia Department of Health

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE

Public Health Service Publication No. 2191
U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1971

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price 70 cents
Stock Number 1723-0049

ERIC

CONTENTS

PREFACE

In response to a general need voiced by students and teachers alike, we have developed a self-contained, job-oriented instructional package on *Descriptive Statistics for the Health Professions.* This is not meant to be an exhaustive treatment of statistics in general; it is limited, first, to *descriptive statistics* and, second, to those concepts and techniques most needed by health professionals working routinely with the basic statistical data. This attempt at job relatedness is also reflected in the post-instructional aims—we want the student to be able to put statistics to practical use, not converse in highly theoretical terms.

Because we have sought operational relevancy and technical simplicity, two cautions are in order:

(1) We have used health data in our examples in order to put the health professional in *familiar* surroundings. However, in our eagerness to keep the necessary basic math simple and the text unencumbered, we may have in places stretched the plausibility of certain health phenomena. Therefore, please don't take offense but rather remember that the health data is not intended to be authentic, only familiar.

(2) Also, in keeping with our simple, practical approach, highly complicated, technical concepts, definitions, and techniques have been avoided. Whenever this approach has conflicted with technical completeness, we have decided in favor of simplicity and practicality if technical accuracy is not violated. (Therefore, professional statisticians, please take note and do not hold your fellow professionals—our consulting statisticians—responsible for any instructional liberties.)

Descriptive Statistics for the Health Professions is concerned with only those statistics that are *generally classified* as descriptive statistics:

(1) tables
(2) graphs
(3) descriptive ratios
(4) measures of central tendency
(5) measures of dispersion

The present booklet is a programmed self-instructional *guide* for the computation of the arithmetic mean. This Guide should be used as a supplement to the lesson "Measures of Central Tendency" and on those occasions when there is an actual need to compute a mean. The Guide will be a quick and easy reference which will provide step-by-step directions for computing the mean in most situations. Such a guide was developed because we have recognized:

(1) that statistics is used by many public health professionals on such an occasional basis that details of computation are easily forgotten or made vague through disuse, and
(2) that many statistics textbooks are too general or complex in statistical technique and not specific enough in application to public health.

We feel strongly that this Guide, when properly used, should significantly reduce training time and cost, reduce the public health professional's aversion to using statistics, and increase the effectiveness with which statistics are applied.

HOW TO USE THIS GUIDE

Although this booklet may be used as a study tool in a formal training setting, it has been designed for use as a guide when you are actually confronted with the need to compute and use the arithmetic mean.

Before using the Guide, be sure to refer to page ix, "When to Use the Arithmetic Mean," to make sure you are using the proper statistic.

The Guide also provides a condensed, simple reference on page xi, "Selecting the Correct Computational Procedure," for use when you are unsure of the particular computational procedure to use and therefore cannot use the standard Contents page.

Each computational procedure covers several pages. The general computational PROCEDURE is always given on the left-hand page (A) with an EXAMPLE given on the right-hand page (B) which matches the general procedure step by step.

This Guide will prove effective if you follow the simple suggestions below:

(1) Use the information provided to select the proper computational procedure.

(2) Read a procedure step carefully; study the example given for *that* step carefully; *then* apply *that* step to the computation of your own statistic.

ERIC

WHEN TO USE THE ARITHMETIC MEAN

The arithmetic mean is one of several *measures of central tendency*. It is commonly referred to as the mean or, less precisely, as the average. Like all measures of central tendency, the arithmetic mean is a single most *typical value* that may be used to represent all the individual values in a particular distribution (group of persons, cases, measures, etc.).

The arithmetic mean is one of the more *mathematically* useful measures of central tendency. Because the value of every item or observation in the group is used in its computation, we may say that in the physical sense, the mean is the balancing point of a distribution of values.

However, if your group contains a few values which are much smaller or much larger than most of the values, the mean will be biased unrealistically in their favor and may grossly misrepresent the typical value for the group. (In such instances you should consider using the median or possibly the mode.)

To compute the arithmetic mean using this Guide you need only the *complete* list of values for your group. The next page will tell you how to select the correct computational procedure for your data.

SELECTING THE CORRECT COMPUTATIONAL PROCEDURE

At this point it is assumed that the arithmetic mean is an appropriate statistic for you to use and that you have adequate data. (If you are not sure, see page ix.) Now follow the directions below for the correct procedure and its page number:

I. Are the values in your data discrete or continuous . . .

- If discrete, this means that your values are indivisible units or counts, such as a visit, a person, a pregnancy, an illness, an inspection, etc.—any of these either happened or they didn't. Go to II below.
- If continuous, this means that your values may be measured and stated as portions or fractions. For example, height, weight, age, millimeters of blood pressure, etc.—all of these can be stated in that form (fractional or whole number) which can be most accurately approximated (measured) and most conveniently used. Go to III below.

II. Estimate (or count) the number of values (counts, observations, cases, etc.) in your list of data . . .

- If you estimate less than 50, go to page 1A.
- If you estimate 50 or more, go to IV below.

III. Estimate (or count) the number of values (counts, observations, cases, etc.) in your list of data . . .

- If you estimate less than 50, go to page 3A.
- If you estimate 50 or more, go to V below.

IV. Subtract the smallest value in your list of data from the largest value . . .

- If the difference is less than 15, go to page 5A.
- If the difference is more than 14, go to page 10A.

V. Subtract the smallest value in your list of data from the largest value (Note: Ignore any decimal points in your answer, e.g., treat 7.14 as if it were 714.) . . .

- If the difference is less than 15, go to page 16A.
- If the difference is more than 14, go to VI below.

VI. Are the values of your data rounded to the nearest unit or fraction of a unit . . .

- If so (and most values are), go to page 21A.
- If not (age at *last* birthday is the most common such "non-rounded" value), go to page 27A.

ERIC

1A | *Arithmetic Mean*

FOR UNGROUPED, DISCRETE DATA — THE PROCEDURE

1. You should be using this procedure only . . .
   (a) if your data consists of *less than* 50 values (cases, observations, readings, etc.).

   SEE EXAMPLE ⟶

   (b) if the values of your data are discrete, i.e., cannot exist as portions or fractions.

   SEE EXAMPLE ⟶

2. The basic formula used for the computation of an arithmetic mean is:

$$\overline{X} = \frac{\Sigma X}{N}$$

For easy reference you may write this formula at the top of the paper on which you are to do your computation.

SEE EXAMPLE ⟶
(For basic definition of symbols)

FOR UNGROUPED, DISCRETE DATA — AN EXAMPLE

1\.

(a) The following is a list of the number of clinic visits made by each woman admitted to prenatal service in Walker County who delivered during 1960: 2, 5, 1, 3, 2, 4, 5, 7, 3, 6, 1, 3, 4, 2, 5, 4, 3, 6. [Note that there are only 18 values (observations) in the list of data.]

RETURN TO THE NEXT STEP IN THE PROCEDURE

(b) In the above data the values are discrete because a visit is indivisible—a woman either made a visit or she didn't. Other types of discrete values are births, deaths, fetal deaths, and inspections.

2\. $\bar{X}$ refers to the arithmetic mean.
X refers to a single value.
ΣX refers to the sum of (Σ) all the values (X).
N refers to the number of values.

GO TO NEXT PAGE

THE PROCEDURE (continued)

3. Find the total for all the values of your data; substitute this total for the sum of values—ΣX—in the formula.

   SEE EXAMPLE ⟶

4. Count the number of values in your data and substitute this count for N in the formula.

   SEE EXAMPLE ⟶

5. To complete the computation, perform the division function indicated by the formula. Because the values in your data are discrete you may round off your answer (quotient) to as many decimal places as desired.

   SEE EXAMPLE ⟶

6. Turn now to page 33A for a general guide to the use of the arithmetic mean you have just computed.

THE END

ARITHMETIC MEAN FOR UNGROUPED, DISCRETE DATA

AN EXAMPLE (continued)

3. The *sum of* X, or total, for the values in our example [2, 5, 1, 3, 2, 4, 5, 7, 3, 6, 1, 3, 4, 2, 5, 4, 3, 6] is 66. That is, $\Sigma X = 66$.

$$\text{Therefore } \overline{X} = \frac{\Sigma X}{N} = \frac{66}{N}$$

4. In our example there are 18 values listed.

$$\text{Therefore } \overline{X} = \frac{\Sigma X}{N} = \frac{66}{18}$$

5. In our example:

$$\overline{X} = \frac{\Sigma X}{N} = \frac{66}{18} = 3.6666\text{------} \text{ (for as many places as we wish)}$$
$$= 3.7 \text{ visits (We think one place is enough.)}$$

6. Turn now to page 33B for an example of the use of an arithmetic mean.

THE END

ARITHMETIC MEAN FOR UNGROUPED, DISCRETE DATA

ERIC

3A | *Arithmetic Mean*

FOR UNGROUPED, CONTINUOUS DATA — THE PROCEDURE

1. You should be using this procedure only . . .
   (a) if your data consists of *less than* 50 values (cases, observations, readings, etc.).

   SEE EXAMPLE ⟶

   (b) if the values of your data are continuous; i.e., can be measured in fractional form on a continuous scale.

   SEE EXAMPLE ⟶

2. The basic formula used for the computation of an arithmetic mean is:

$$\bar{X} = \frac{\Sigma X}{N}$$

For easy reference you may write this formula at the top of the paper on which you are to do your computation.

SEE EXAMPLE ⟶
(For basic definition of symbols)

FOR UNGROUPED, CONTINUOUS DATA — THE EXAMPLE

1.

(a) The following is a list of weights (*to the nearest tenth of a pound*) of two-year-old males attending well-child conferences in Upton County during the month of April, 1960: 30.2, 24.6, 28.7, 33.4, 2[illegible].3, 22.2, 37.8, 31.9, 29.1, 21.1, 26.7, 32.3, 30.6, 28.5, 31.7, 32.4, 29.2, 30.3, 3[illegible].7, 31.7, 28.6, 26.2, 30.2, 29.1. [Note that there are only 24 values (weights) in the list of data.]

← RETURN TO THE NEXT STEP IN THE PROCEDURE

(b) In the above data the values are continuous because weight is measured on a continuous scale. Weight is stated in that form (fractional or whole number) which can be most accurately approximated (measured) and most conveniently used. We have chosen to measure (and state) our values to the nearest tenth of a pound. Other types of continuous values are years of age, millimeters of blood pressure, height, and length of gestation.

2. $\bar{X}$ refers to the arithmetic mean.
   X refers to a single value.
   ΣX refers to the sum of (Σ) all the values (X).
   N refers to the number of values.

GO TO NEXT PAGE →

THE PROCEDURE (continued)

3. Find the total for all the values of your data; substitute this total for the sum of values—ΣX—in the formula.

SEE EXAMPLE →

4. Count the number of values in your data and substitute the count for N in the formula.

SEE EXAMPLE →

5. To complete the computation, perform the division function indicated by the formula. Because the values in your data are continuous, you *must* round off your answer (quotient) to *no more than* the same number of decimal places in the original values. If desirable, your answer could be rounded to have *fewer* decimal places than the original values.

SEE EXAMPLE →

6. Turn now to page 33A for a general guide to the use of the arithmetic mean you have just computed.

THE END

ARITHMETIC MEAN FOR UNGROUPED, CONTINUOUS DATA

AN EXAMPLE (continued)

3. The *sum of X*, or total, for the values in our example [30.2, 24.6, 28.7, 33.4, 27.3, 22.2, 37.8, 31.9, 29.1, 21.1, 26.7, 32.3, 30.6, 28.5, 31.7, 32.4, 29.2, 30.3, 33.7, 31.7, 28.6, 26.2, 30.2, 29.1] is 707.2. That is, $\Sigma X = 707.2$.

$$\text{Therefore } \overline{X} = \frac{\Sigma X}{N} = \frac{707.2}{N}$$

4. In our example there are 24 values listed.

$$\text{Therefore } \overline{X} = \frac{\Sigma X}{N} = \frac{707.2}{N} = \frac{707.2}{24}$$

5. In our example:

$$\overline{X} = \frac{\Sigma X}{N} = \frac{707.2}{24} = 29.46666\text{-----}$$

$= 29.5$ lb. (We round off at least *to the nearest tenth of a pound* because our original values were stated to the nearest tenth; or we could round to the nearest pound if desired, that is, 29.46666----- = 29 lbs.)

6. Turn now to page 33B for an example of the use of an arithmetic mean.

THE END

ARITHMETIC MEAN FOR UNGROUPED, CONTINUOUS DATA

FOR GROUPED (SINGLE VALUE), DISCRETE DATA — THE PROCEDURE

1. You should be using this procedure only . . .
   (a) if your data consists of 50 or more values (cases, observations, readings, etc.).

   SEE EXAMPLE ⟶

   (b) if the values of your data are discrete, i.e., cannot exist as portions or fractions.

   SEE EXAMPLE ⟶

   (c) if the difference between the smallest value and the largest value is less than 15 (this means that the number of different values possible is not more than 15).

   SEE EXAMPLE ⟶

2. The formula to be used to compute the arithmetic mean is:

$$\bar{X} = \frac{\Sigma fX}{N}$$

   For easy reference you may write this formula at the top of the paper on which you are to do your computation.

   SEE EXAMPLE ⟶
   (For basic definition of symbols)

ERIC

FOR GROUPED (SINGLE VALUE), DISCRETE DATA — AN EXAMPLE

1\.

(a) The following is a list of the number of previous pregnancies for women admitted to prenatal services in Jasper County during the calendar year 1960: 0, 3, 1, 0, 2, 1, 0, 0, 2, 1, 0, 5, 1, 0, 1, 2, 4, 0, 1, 0, 1, 1, 3, 5, 0, 1, 4, 9, 2, 3, 0, 0, 1, 2, 2, 3, 0, 1, 4, 3, 4, 3, 0, 1, 0, 1, 0, 2, 3, 0, 2, 1, 0, 4, 3, 2, 0, 1, 2, 2. [Note that there are 60 values (observations) in the list of data.]

RETURN TO THE NEXT STEP IN THE PROCEDURE

(b) In the above data the values are discrete because a pregnancy is indivisible—a woman is either pregnant or she's not. Other types of discrete values are births, deaths, fetal deaths, cases, visits, and inspections.

(c) In our above list of data, the smallest value is 0 and the largest value is 9; the difference between these values is 9. Therefore, only 10 different values are possible in the list ($9 - 0 + 1 = 10$); actually only 0, 1, 2, 3, 4, 5, and 9 (seven values) occur and are repeated one or more times.

2\. $\bar{X}$ refers to the arithmetic mean.
X refers to a single value.
f refers to the frequencies with which values occur.
N refers to the total number of values.
Σf refers to the sum of (Σ) all the frequencies (f)—Note: $N = \Sigma f$.
ΣfX refers to the sum of (Σ) the frequency (f) times the single value (X) to which it refers.

GO TO NEXT PAGE

THE PROCEDURE (continued)

3. For convenience in computation, prepare a worktable as follows:
   (a) Though not always essential, a title that clearly describes the content is sometimes desired. If the table is to be used for general communication (published, etc.) then it should contain the What (the group being studied), How (the characteristics of the group being allowed to vary), Where (the area in which the data was gathered), and When (the time period in which the data was gathered) *in that order.*

   SEE EXAMPLE ⟶

   (b) Draw two parallel lines (the head) about one inch apart directly below the worktable title; then divide the head and the rest of the page into three equal columns.

   SEE EXAMPLE ⟶

   (c) In the head of the first column write the name of the value being considered and the single value symbol "X."

   SEE EXAMPLE ⟶

   (d) In the head of the second column write the name of that which reflects the frequency with which a single value occurs; include the symbol "f."

   SEE EXAMPLE ⟶

   (e) In the head of the third column write only the symbol "fX."

   SEE EXAMPLE ⟶

ERIC

AN EXAMPLE (continued)

3\.

(a) We have decided to use a formal (publishable) title for our worktable below; compare it with the original statement given in Step 1(a), page 5B . . . . Notice that the information is listed *in order:* What, How, Where, When.

WHAT HOW

WORKTABLE: Women Admitted to Prenatal Service, by Number of Previous Pregnancies, Jasper County, 1960

HOW WHERE WHEN

(b) Heads and Columns

Number of Previous Pregnancies X	Number of Women f	fX

(c) We have written in the name of the value being considered and the single value symbol "X."

(d) The number of women having a certain number of previous pregnancies reflects the frequencies (f) with which values (previous pregnancies) occur in our list of data.

(e) We have labeled the third column simply "fX."

THE PROCEDURE (continued)

3. (Worktable construction continued)

(f) In the X column, starting with the lowest value in your data, list consecutively all the *possible* single values up to, and including, your data's highest value. Important: write each possible single value only once.

SEE EXAMPLE →

(g) In the f column, for each single value listed in the X column, write the number of times (frequency) that the single value occurs in your list of data.

SEE EXAMPLE →

(h) In the fX column, for each single value listed, write the products of the f column times the X column.

SEE EXAMPLE →

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Women Admitted to Prenatal Service, by Number of Previous Pregnancies, Jasper County, 1960		
Number of Previous Pregnancies X	Number of Women f	fX
0	18	0
1	15	15
2	11	22
3	8	24
4	5	20
5	2	10
6		
7		
8		
9	1	9

(f) We have listed all the *possible* single values between our lowest (0) and our highest (9). (Each value is listed only once.)

(g) The number of women having a particular number of previous pregnancies is listed as the frequency with which that number of previous pregnancies occurs; that is, 18 women had no previous pregnancy, 15 women had one previous pregnancy, etc.

(h) We have multiplied each single value (X) by its frequency (f) and entered the product in the fX column; that is, 18 times 0 is 0; 15 times 1 is 15; 11 times 2 is 22; etc.

ERIC

THE PROCEDURE (continued)

3. (Worktable construction continued)

(i) Draw two parallel lines (for totals) about ½ inch apart directly below the last single values listed.

SEE EXAMPLE →

(j) Between the parallel lines and beneath the last single value listed in your X column, write "TOTAL."

SEE EXAMPLE →

(k) Add the frequencies listed in your f column and write the total in the TOTAL s row; label the total "N."

SEE EXAMPLE →

(l) Add the products listed in the fX column and enter the total in the TOTAL s row; label the total "ΣfX."

SEE EXAMPLE →

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Women Admitted to Prenatal Service, by Number of Previous Pregnancies, Jasper County, 1960		
Number of Previous Pregnancies X	Number of Women f	fX
0	18	0
1	15	15
2	11	22
3	8	24
4	5	20
5	2	10
6		
7		
8		
9	1	9
TOTAL	N = 60	$\Sigma fX = 100$

(i) Totals

(j) We have labeled the totals row.

(k) We have added all our individual frequencies (Σf) and written the total with its appropriate symbol. You could get the same number (60) if you counted the number of values in our original list of data (see Step 1a, page 5B).

(l) We have added all the individual products and written the total with its appropriate symbol.

THE PROCEDURE (continued)

4. Substitute the totals from your worktable into the computational formula for the mean:

$$\overline{X} = \frac{\Sigma fX}{N}$$

SEE EXAMPLE →

5. To complete the computation, perform the division function indicated by the formula. Because the values in your data are discrete, you may round off your answer (quotient) to as many decimal places as desired.

SEE EXAMPLE →

6. Turn now to page 33A for a general guide to the use of the arithmetic mean you have just computed.

THE END

ARITHMETIC MEAN FOR GROUPED (SINGLE VALUE), DISCRETE DATA

AN EXAMPLE (continued)

4.

WORKTABLE: Women Admitted to Prenatal Service, by Number of Previous Pregnancies, Jasper County, 1960		
Number of Previous Pregnancies X	Number of Women f	fX
0	18	0
1	15	15
9	1	9
TOTAL	N=60	ΣfX=100

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{100}{60}$$

5. In our example:

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{100}{60} = 1.6666\ldots\ldots \text{ (for as many places as we wish)}$$
$$= 1.7 \text{ pregnancies (We think one place is enough.)}$$

6. Turn now to page 33B for an example of the use of an arithmetic mean.

THE END

ARITHMETIC MEAN FOR GROUPED (SINGLE VALUE), DISCRETE DATA

FOR GROUPED (INTERVAL), DISCRETE DATA — THE PROCEDURE

1. You should be using this procedure only . . .
   (a) if your data consists of 50 or more values (cases, observations, readings, etc.).

   SEE EXAMPLE →

   (b) if the values of your data are discrete; i.e., cannot exist as portions or fractions.

   SEE EXAMPLE →

   (c) if the difference between the smallest value and the largest value is more than 14 (this means that the number of different values possible is more than 15).

   SEE EXAMPLE →

2. The formula to be used to compute the arithmetic mean is:

$$\bar{X} = \frac{\Sigma fX}{N}$$

For easy reference you may write this formula at the top of the paper on which you are to do your computation.

SEE EXAMPLE →
(For basic definition of symbols)

ERIC

FOR GROUPED (INTERVAL), DISCRETE DATA — AN EXAMPLE

1.
(a) The following is a list of the attendance (number of children) at each well-child clinic held in Jones County during the fiscal year ending June 30, 1960: 14, 16, 18, 20, 17, 19, 21, 24, 22, 27, 25, 28, 26, 33, 24, 42, 34, 39, 42, 37, 34, 36, 48, 35, 38, 32, 39, 44, 36, 34, 37, 32, 31, 33, 30, 32, 33, 31, 32, 30, 28, 27, 26, 29, 23, 19, 24, 20, 22, 21, 20, 15, 18, 15, 15, 11, 13, 11, 12, 11. [Note that there are 60 values (observations) in the list of data.]

RETURN TO THE NEXT STEP IN THE PROCEDURE.

(b) In the above data the values are discrete because an attendance at the clinic is indivisible—a child either attended or he didn't. Other types of discrete values are births, deaths, fetal deaths, cases, visits, inspections, and pregnancies.

(c) In our above list of data, the smallest value is 11 and the largest value is 48; the difference between these values is 37. Therefore, more than 15 different values are possible in the list; actually 38 different values are possible ($48 - 11 + 1 = 38$).

2. $\overline{X}$ refers to the arithmetic mean.
X refers to a single value (in the present case it will be an interval midpoint).
f refers to the frequencies with which values occur.
N refers to the total number of values.
Σf refers to the sum of (Σ) all the frequencies (f)—Note: $N = \Sigma f$.
ΣfX refers to the sum of (Σ) the frequencies (f) times the single value (X) to which it refers.

GO TO NEXT PAGE

THE PROCEDURE (continued)

3. For convenience in computation, prepare a worktable as follows:
   (a) Though not always essential, a title that clearly describes the content is sometimes desired. If the table is to be used for general communication (published, etc.) then it should contain the What (the group being studied), How (the characteristics of the group being allowed to vary), Where (the area in which the data was gathered), and When (the time period in which the data was gathered) *in that order*.

   SEE EXAMPLE ⟶

   (b) Draw two parallel lines (the head) about one inch apart directly below the worktable title; then divide the head and the rest of the page into four equal columns.

   SEE EXAMPLE ⟶

   (c) In the head of the first column write the name of the value being considered.

   SEE EXAMPLE ⟶

   (d) In the head of the second column write the name of that which reflects the frequencies with which values occur; include the symbol "f."

   SEE EXAMPLE ⟶

   (e) In the head of the third column, write the word "Midpoint" and the single value symbol "X," which in this case will be represented by the interval midpoint.

   SEE EXAMPLE ⟶

   (f) In the head of the fourth column write only the symbol "fX."

   SEE EXAMPLE ⟶

AN EXAMPLE (continued)

3\.

(a) We have decided to use a formal (publishable) title for our worktable below; compare it with the original statement given in Step 1(a), page 10B. . . . Notice that the information is listed *in order:* What, How, Where, When

WHAT HOW

WORKTABLE: Distribution of Well-Child Clinics, by Number of Children Attending, Jones County, Year Ending June 30, 1960

HOW WHERE WHEN

Number of Children	Number of Clinics f	Midpoint X	fX

(b) Heads and Columns

(c) We have written in the name of the value being considered.

(d) The number of clinics having a certain number of children attending reflects the frequency (f) with which values (attendance number) occur in our list of data.

(e) We have labeled the third column "Midpoint" and included the symbol "X."

(f) We have labeled the fourth column simply "fX."

THE PROCEDURE (continued)

3. (Worktable construction continued)

(g) List your values in the first column of the worktable so that you will have from 7 - 15 intervals. To determine interval *size* subtract your lowest value from the highest, divide by 10, and round off to the nearest *whole number.* Your smallest value should be included in your first interval as either the lower limit of the interval or as some value within the interval; your largest value should be included in the last interval.

[NOTE: Some latitude is allowed in deciding the actual interval size. The above computational method gives you an approximate interval size that can be used. However, you may increase or decrease the actual interval size *as long as you do not exceed the 7 - 15 limit on the number of intervals.* This flexibility permits you to select an interval size that is more readable (e.g., 5) for tables that are to be published. *IMPORTANT:* This practice should not be abused—computationally, it is best to have the smallest possible interval size.]

SEE EXAMPLE ⟶

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of Well-Child Clinics, by Number of Children Attending, Jones County, Year Ending June 30, 1960

Number of Children	Number of Clinics f	Midpoint X	fX
10 - 14			
15 - 19			
20 - 24			
25 - 29			
30 - 34			
35 - 39			
40 - 44			
45 - 49			

(g) The smallest value in our data is 11 children; the largest is 48 children; therefore (48 — 11) ÷ 10 = 3.7 or 4 children is the computed interval size. However, we decided to increase the actual interval size to 5 children since it would make the table more readable and still produce at least 7 intervals (and less than 16). Again for readability, we decided to make the lower limit of the first interval 10 children. To actually determine the interval, the value 10 was listed to the left in the first column; then the interval size "5" was successively added until the interval was produced that would include our largest value 48 children. Once the lower limit of each interval was written, we wrote in the upper limits.

[NOTE: To make our table more appropriate computationally, we would have used the 4 unit interval size as follows:

11 - 14
15 - 18
19 - 22
23 - 26
27 - 30
31 - 34 — 10 intervals — Notice that we would still have had less than the maximum (15) number of intervals recommended.]
35 - 38
39 - 42
43 - 46
47 - 50

THE PROCEDURE (continued)

3. (Worktable construction continued)

(h) In the f column, for each interval listed in the first column, write the number of values in your data that fall within the interval; this will give you your *interval* frequency.

SEE EXAMPLE ⟶

(i) In the X column enter the midpoint of each interval; this single value (midpoint) is found by adding the lower limit and upper limit of each stated interval and dividing the sum by 2.

SEE EXAMPLE ⟶

(j) In the fX column, for each midpoint value listed, enter the products of the f column times the X column.

SEE EXAMPLE ⟶

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of Well-Child Clinics, by Number of Children Attending, Jones County, Year Ending June 30, 1960

Number of Children	Number of Clinics f	Midpoint X	fX
10 - 14	6	12	72
15 - 19	9	17	153
20 - 24	11	22	242
25 - 29	8	27	216
30 - 34	14	32	448
35 - 39	8	37	296
40 - 44	3	42	126
45 - 49	1	47	47

(h) The number of clinics whose attendance count falls within each interval is written beside that interval in the f column. That is, 6 clinics had between 10-14 children attend during the year, 9 clinics had 15-19 children attend, etc.

(i) To find the midpoint of each interval we added the lower limit and upper limit of each stated interval and divided by 2. That is,

$\frac{10 + 14}{2} = 12$; $\frac{15 + 19}{2} = 17$; etc.

(j) We have multiplied each single value X (midpoint) by its *interval* frequency and entered the products in the fX column. That is, 12 times 6 is 72; 17 times 9 is 153; etc.

THE PROCEDURE (continued)

3. (Worktable construction continued)

(k) Draw two parallel lines (for totals) about ½ inch apart directly below the last interval listed.

SEE EXAMPLE ⟶

(l) Between the parallel lines and beneath the last interval listed in your first column, write "TOTAL."

SEE EXAMPLE ⟶

(m) Add the frequencies listed in your f column and write the total in the TOTALs row; label the total "N."

SEE EXAMPLE ⟶

(n) Add the products listed in the fX column and enter the total in the TOTALs row; label the total "ΣfX."

SEE EXAMPLE ⟶

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of Well-Child Clinics, by Number of Children Attending, Jones County, Year Ending June 30, 1960			
Number of Children	Number of Clinics f	Midpoint X	fX
10 - 14	6	12	72
15 - 19	9	17	153
20 - 24	11	22	242
25 - 29	8	27	216
30 - 34	14	32	448
35 - 39	8	37	296
40 - 44	3	42	126
45 - 49	1	47	47
TOTAL	N=60		ΣfX=1600

(k) Totals

(l) We have labeled the totals row.

(m) We have added all our individual frequencies (Σf) and written the total with its appropriate symbol. You could get the same total (60) if you counted the number of values in our original list of data (see Step 1(a), page 10B).

(n) We have added all the individual products and written the total with its appropriate symbol.

THE PROCEDURE (continued)

4. Substitute the totals from your worktable into the computational formula for the mean:

$$\overline{X} = \frac{\Sigma fX}{N}$$

SEE EXAMPLE →

5. To complete the computation, perform the division functions indicated by the formula. Because the values in your data are discrete, you may round off your answer (quotient) to as many decimal places as desired.

SEE EXAMPLE →

6. Turn now to page 33A for a general guide to the use of the arithmetic mean you have just computed.

THE END

ARITHMETIC MEAN FOR GROUPED (INTERVAL), DISCRETE DATA

AN EXAMPLE (continued)

4.

WORKTABLE: Distribution of Well-Child Clinics, by Number of Children Attending, Jones County, Year Ending June 30, 1960			
Number of Children	Number of Clinics f	Midpoint X	fX
10 - 14	6	12	72
15 - 19	9	17	153
20 - 24	11	22	242
45 - 49	1	47	47
TOTAL	N=60		ΣfX=1600

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{1600}{60}$$

5. In our example:

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{1600}{60} = 26.6666\ldots\ldots \text{ (for as many places as we wish)}$$
$$= 26.7 \text{ children per clinic (We think one place is enough.)}$$

6. Turn now to page 33B for an example of the use of an arithmetic mean.

THE END

ARITHMETIC MEAN FOR GROUPED (INTERVAL), DISCRETE DATA

FOR GROUPED (SINGLE VALUE), CONTINUOUS DATA — THE PROCEDURE

1. You should be using this procedure only . . .
   (a) if your data consists of 50 or more values (cases, observations, readings, counts, etc.).

   SEE EXAMPLE ⟶

   (b) if the values of your data are continuous, i.e., can be measured in fractional form on a continuous scale.

   SEE EXAMPLE ⟶

   (c) if, when assuming for the moment that the continuous values of your data are discrete and ignoring all decimal points, the difference between the smallest value given and the largest value given is less than 15 (this means that the number of different *major* values possible is not more than 15).

   SEE EXAMPLE ⟶

2. The formula used to compute the arithmetic mean is:

$$\bar{X} = \frac{\Sigma fX}{N}$$

For easy reference you may write this formula at the top of the paper on which you are to do your computation.

SEE EXAMPLE ⟶
(For basic definition of symbols)

FOR GROUPED (SINGLE VALUE), CONTINUOUS DATA—AN EXAMPLE

1.

(a) The following is a list of the heights *to the nearest inch* of 2-year-old children attending well-child clinics in Jones County during the months April-June, 1960: 34, 38, 35, 36, 40, 39, 37, 35, 38, 36, 37, 39, 42, 33, 37, 34, 40, 32, 35, 34, 36, 38, 35, 39, 41, 37, 39, 38, 36, 38, 33, 37, 36, 39, 36, 35, 37, 34, 35, 37, 36, 38, 35, 36, 37, 34, 36, 38, 39, 36, 35, 34, 36, 33, 34, 36, 37, 35, 36, 33. [Note that there are 60 values (measurements) in the list of data.]

RETURN TO THE NEXT STEP IN THE PROCEDURE

(b) In the above data the values are continuous because height is measured on a continuous scale. Height is stated in that form (fractional or whole number) which can be most accurately approximated (measured) and most conveniently used. We have chosen to measure (and state) our values to the nearest whole inch. Other types of continuous values are years of age, millimeters of blood pressure, weight, and length of gestation.

(c) Although the values in our above list of data are continuous, if we assume for the moment they are discrete then the difference between the smallest value given (32 inches) and the largest value given (42 inches) is "10." Therefore, only 11 different major values are possible ($42 - 32 + 1 = 11$). [NOTE: In this type of computation decimal points are ignored, e.g., 1.042 and 1.032 would be treated as whole numbers 1042 and 1032.]

2. $\bar{X}$ refers to the arithmetic mean.
X refers to a single value.
f refers to the frequencies with which values occur.
N refers to the total number of values.
Σf refers to the sum of (Σ) the frequencies (f)—Note: $N = \Sigma f$.
ΣfX refers to the sum of (Σ) the frequencies (f) times the single value (X) to which it refers.

GO TO NEXT PAGE

THE PROCEDURE (continued)

3. For convenience in computation, prepare a worktable as follows:
   (a) Though not always essential, a title that clearly describes the content is sometimes desired. If the table is to be used for general communication (published, etc.) then it should contain the What (the group being studied), How (the characteristics of the group being allowed to vary), Where (the area in which the data was gathered), and When (the time period in which the data was gathered) *in that order*.

   SEE EXAMPLE →

   (b) Draw two parallel lines (the head) about one inch apart directly below the worktable title; then divide the head and the rest of the page into three equal columns.

   SEE EXAMPLE →

   (c) In the head of the first column write the name of the value being considered, the unit by which it is being measured, and the single value symbol "X."

   SEE EXAMPLE →

   (d) In the head of the second column write the name of that which reflects the frequencies with which values occur; include the symbol "f."

   SEE EXAMPLE →

   (e) In the head of the third column write only the symbol "fX."

   SEE EXAMPLE →

ERIC

THE PROCEDURE (continued)

3. (Worktable construction continued)

(f) In the X column, starting with the lowest value in your data, list consecutively all the possible major values up to, and including, your data's highest value. [NOTE: By using only the *major* values, you are giving your continuous data a discrete treatment.]

SEE EXAMPLE →

(g) In the f column, for each single value listed in the X column, write the number of times (frequency) that the single value occurs in your list of data.

SEE EXAMPLE →

(h) In the fX column, for each single value listed, enter the products of the f column times the X column.

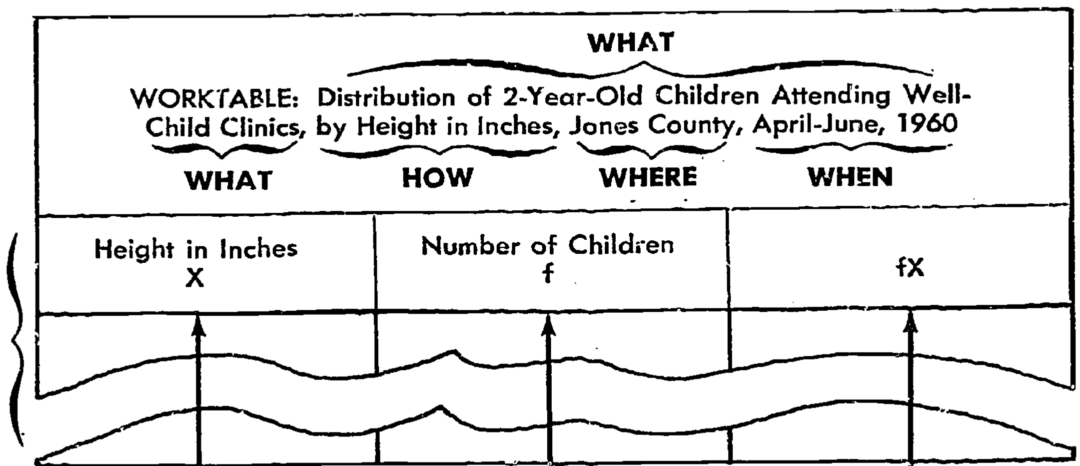
SEE EXAMPLE →

AN EXAMPLE (continued)

3\.

(a) We have decided to use a formal (publishable) title for our worktable below; compare it with the original statement given in Step 1(a), page 16B. . . . Notice that the information is listed *in order:* What, How, Where, When.

(b) Heads and Columns



(c) We have written in the name of the value being considered (height), the unit by which it is being measured (inches), and the single value symbol "X."

(d) The number of children having a certain height reflects the frequencies (f) with which values (height) occur in our list of data.

(e) We have labeled the third column simply "fX."

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of 2-Year-Old Children Attending Well-Child Clinics, by Height in Inches, Jones County, April-June, 1960		
Height in Inches X	Number of Children f	fX
32	1	32
33	4	132
34	7	238
35	9	315
36	13	468
37	9	333
38	7	266
39	6	234
40	2	80
41	1	41
42	1	42

(f) We have listed all the possible *major* single values between our lowest (32) and our highest (42). [There are, of course, an *unlimited* number of possible values between 32 and 42 *more fractio l* than those listed.]

(g) The number of children having a particular height is listed as the frequency with which that height occurs; that is, 1 child was 32 inches tall, 4 children were 33 inches tall, etc.

(h) We have multiplied each single value (X) by its frequency (f) and entered the product in the fX column; that is, 1 times 32 is 32; 4 times 33 is 132, etc.

THE PROCEDURE (continued)

3. (Worktable construction continued)

(i) Draw two parallel lines (for totals) about ½ inch apart directly below the last single values listed.

SEE EXAMPLE ⟶

(j) Between the parallel lines and beneath the last single value listed in your X column, write "TOTAL."

SEE EXAMPLE ⟶

(k) Add the frequencies listed in your f column and write the total in the TOTALs row; label the total "N."

SEE EXAMPLE ⟶

(l) Add the products listed in the fX column and enter the total in the TOTALs row; label the total "ΣfX."

SEE EXAMPLE ⟶

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of 2-Year-Old Children Attending Well-Child Clinics, by Height in Inches, Jones County, April-June, 1960		
Height in Inches X	Number of Children f	fX
32	1	32
33	4	132
34	7	238
35	9	315
36	13	468
37	9	333
38	7	266
39	6	234
40	2	80
41	1	41
42	1	42
TOTAL	N=60	ΣfX=2181

(i) Totals

(j) We have labeled the totals row.

(k) We have added all our individual frequencies (Σf) and written the total with its appropriate symbol. You could get the same total (60) if you counted the number of values in our original list of data (see Step 1(a), page 16B).

(l) We have added all the individual products and written the total with its appropriate symbol.

THE PROCEDURE (continued)

4. Substitute the totals from your worktable into the computational formula for the mean:

$$\overline{X} = \frac{\Sigma fX}{N}$$

SEE EXAMPLE →

5. To complete the computation, perform the division function indicated by the formula. Because the values in your data are continuous, you must round off your answer (quotient) to *no more than* the same number of decimal places in the original values. If desirable, your answer could be rounded to have *fewer* decimal places than the original values.

SEE EXAMPLE →

6. Turn now to page 33A for a general guide to the use of the arithmetic mean you have just computed.

THE END

ARITHMETIC MEAN FOR GROUPED (SINGLE VALUE), CONTINUOUS DATA

ERIC

AN EXAMPLE (continued)

4.

WORKTABLE: Distribution of 2-Year-Old Children Attending Well-Child Clinics, by Height in Inches, Jones County, April-June, 1960		
Height in Inches X	Number of Children f	fX
32	1	32
33	4	132
42	1	42
TOTAL	N=60	$\Sigma fX = 2181$

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{2181}{60}$$

5. In our example:

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{2181}{60} = 36.35\ldots$$

$= 36.$ inches (We round off at least *to the nearest whole number* because our original values were stated to the nearest whole inch.)

6. Turn now to page 33B for an example of the use of an arithmetic mean.

THE END

ARITHMETIC MEAN FOR GROUPED (SINGLE VALUE), CONTINUOUS DATA

21A | *Arithmetic Mean*

FOR GROUPED (INTERVAL), CONTINUOUS DATA (ROUNDED) — THE PROCEDURE

1. You should be using this procedure only . . .
   (a) if your data consists of 50 or more values (cases, observations, readings, etc.).

   SEE EXAMPLE ⟶

   (b) if the values of your data are *rounded* to the nearest unit or fraction of a unit (most values are of this type).

   SEE EXAMPLE ⟶

   (c) if the values of your data are continuous, i.e., can be measured in fractional form on a continuous scale.

   SEE EXAMPLE ⟶

   (d) if, when assuming for the moment that the continuous values of the data are discrete and ignoring all decimal points, the difference between the smallest value given and the largest value given is more than 14 (this means that the number of different *major* values possible is more than 15).

   SEE EXAMPLE ⟶

2. The formula used to compute the arithmetic mean is:

$$\overline{X} = \frac{\Sigma fX}{N}$$

For easy reference you may write this formula at the top of the paper on which you are to do your computation.

SEE EXAMPLE ⟶
(For basic definition of symbols)

FOR GROUPED (INTERVAL), CONTINUOUS DATA (ROUNDED) — AN EXAMPLE

1.

(a) The following is a list of weights *to the nearest tenth* at birth for live births occurring during 1960 to parents who are residents of Jones County: 3.4, 4.9, 5.6, 11.6, 8.5, 9.1, 7.6, 8.2, 6.7, 7.4, 6.0, 6.5, 9.6, 9.8, 10.0, 7.5, 8.3, 7.7, 8.1, 7.6, 8.2, 7.9, 8.0, 6.8, 7.4, 6.9, 7.2, 5.0, 5.9, 6.2, 10.9, 9.7, 8.4, 9.2, 8.8., 8.0, 7.8, 8.2, 7.6, 7.5, 9.2, 6.6, 7.4, 7.1, 8.3, 8.1, 7.5, 7.7, 8.2, 9.1, 8.5, 4.9, 6.3, 5.9, 7.8, 8.1, 7.9, 8.0, 7.6, 6.8, 7.2, 10.5, 9.4, 8.7, 9.2, 6.8, 7.0, 7.2, 6.3, 5.9. [Note that there are 70 values (observations) in the list of data.]

← RETURN TO THE NEXT STEP IN THE PROCEDURE

(b) The values in our data are weights and are rounded to the nearest tenth of a pound. Other such values are age at *nearest* birthday, millimeters of blood pressure, heights, etc.

(c) In the above data the values are continuous because weight is measured on a continuous scale. Weight is stated in that form (fractional or whole number) which can be most accurately approximated (measured) and most conveniently used. We have chosen to state our values to the nearest tenth of a pound. Other types of continuous values are age, millimeters of blood pressure, heights, length of gestation, etc.

(d) Although the values in our list of data are continuous, if we assume for the moment they are discrete, then the difference between the smallest value given (3.4 lbs.) and the largest value given (11.6 lbs.) is 82. [NOTE: In this type of computation decimal points are ignored, i.e., whereas 11.6 — 3.4 would ordinarily be said to be 8.2, we treat the values as though they were 116 — 34 to get 82.] Therefore, more than 15 different major values are possible; actually, 83 different major values are possible ($116 - 34 + 1 = 83$).

2. $\overline{X}$ refers to the arithmetic mean.
   X refers to a single value (in the present case it will be an interval midpoint).
   f refers to the frequencies with which values occur.
   N refers to the total number of values.
   Σf refers to the sum of (Σ) the frequencies (f)—Note: $N = \Sigma f$.
   ΣfX refers to the sum of (Σ) the frequencies (f) times the single value (X) to which it refers.

GO TO NEXT PAGE

THE PROCEDURE (continued)

3. For convenience in computation, prepare a worktable as follows:
   (a) Though not always essential, a title that clearly describes the content is sometimes desired. If the table is to be used for general communication (published, etc.) then it should contain the What (the group being studied), How (the characteristics of the group being allowed to vary), Where (the area in which the data was gathered), and When (the time period in which the data was gathered) *in that order.*

   SEE EXAMPLE ⟶

   (b) Draw two parallel lines (the head) about one inch apart directly below the worktable title; then divide the head and the rest of the page into four equal columns.

   SEE EXAMPLE ⟶

   (c) In the head of the first column write the name of the value being considered and the unit by which it is being measured.

   SEE EXAMPLE ⟶

   (d) In the head of the second column, write the name of that which reflects the frequencies with which values occur; include the symbol "f."

   SEE EXAMPLE ⟶

   (e) In the head of the third column, write the word "Midpoint" and the single value symbol "X," which in this case will be represented by the interval midpoint.

   SEE EXAMPLE ⟶

   (f) In the head of the fourth column write only the symbol "fX."

   SEE EXAMPLE ⟶

AN EXAMPLE (continued)

3.

(a) We have decided to use a formal (publishable) title for our worktable below; compare it with the original statement given in Step 1(a), page 21B . . . . Notice that the information is listed *in order:* What, How, Where, When.

WORKTABLE: $\overbrace{\text{Resident Live Births,}}^{\text{WHAT}}$ $\overbrace{\text{by Birth Weight,}}^{\text{HOW}}$ $\underbrace{\text{Jones County,}}_{\text{WHERE}}$ $\underbrace{\text{1960}}_{\text{WHEN}}$

(b) Heads and Columns

Birth Weight in Pounds	Number of Live Births f	Midpoint X	fX

(c) We have written in the name of the value being considered (birth weight) and the unit by which it is being measured (pounds).

(d) The number of live births having a certain birth weight reflects the frequencies (f) with which values (birth weight) occur in our list of data.

(e) We have labeled the third column "Midpoint" and included the symbol "X."

(f) We have labeled the fourth column "fX."

THE PROCEDURE (continued)

3. (Worktable construction continued)

(g) List your values in the first column of the worktable so that you will have from 7 - 15 intervals. To determine interval *size* subtract your lowest value from the highest, divide by 10, and round off to the *same number* of decimal places as in your original values. Your smallest value should be included in your first interval as either the lower limit of the interval or as some value within the interval; your largest value should be included in the last interval.

[NOTE: Some latitude is allowed in deciding the actual interval size. The above computational method gives you an approximate interval size that can be used. However, you may increase or decrease the actual interval size *as long as you do not exceed the 7 - 15 limit on the number of intervals*. This flexibility permits you to select an interval size that is more readable (e.g., 5) for tables that are to be published. *IMPORTANT:* This practice should not be abused—computationally, it is best to have the smallest possible interval size.]

SEE EXAMPLE ⟶

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Resident Live Births, by Birth Weight, Jones County, 1960			
Birth Weight in Pounds	Number of Live Births f	Midpoint X	fX
3.0 - 3.9			
4.0 - 4.9			
5.0 - 5.9			
6.0 - 6.9			
7.0 - 7.9			
8.0 - 8.9			
9.0 - 9.9			
10.0 - 10.9			
11.0 - 11.9			

(g) The smallest value in our data is 3.4 lbs.; the largest is 11.6 [illegible] Therefore, (11.6 — 3.4) ÷ 10 = 0.82 or 0.8 lbs. is the computed interval size. However, we decided to increase the actual interval size to 1.0 lbs. since this would make the table more readable and still produce at least 7 intervals (and less than 16). Again for readability, we decided to make the lower limit of the first interval 3.0 lbs. To actually determine the interval, the value 3.0 was listed to the left in the first column; then the interval size "1.0" was successively added until the interval was produced that would include our largest value 11.6 lbs. Once the lower limit of each interval was written, we wrote in the upper limits.

[NOTE: To make our table more appropriate computationally, we would have used the 0.8 lbs. interval size as follows:

3.0 - 3.7
3.8 - 4.5
4.6 - 5.3
5.4 - 6.1
6.2 - 6.9
7.0 - 7.7
7.8 - 8.5
8.6 - 9.3
9.4 - 10.1
10.2 - 10.9
11.0 - 11.7

} 11 intervals

Notice that we would still have had less than the maximum (15) number of intervals recommended.]

THE PROCEDURE (continued)

3. (Worktable construction continued)

(h) In the f column, for each interval listed in the first column, write the number of values in your data that fall within the interval; this will give you your interval frequency. [NOTE: For a *stated* interval 5 - 9, values 4.5 up to (but *not* through) 9.5 would be considered to fall therein; therefore, the *actual* interval would be 4.5 to 9.5.]

SEE EXAMPLE →

(i) In the X column, enter the midpoint of each interval; this single value is found by averaging the lower and upper limits of the *actual* interval. For example, from (h) above we see that the *stated* interval 5 - 9 is the *actual* interval 4.5 - 9.5; therefore, $\frac{4.5 + 9.5}{2} = 7.0$, the midpoint.

SEE EXAMPLE →

(j) In the fX column, for each midpoint value listed, enter the products of the f column times the X column.

SEE EXAMPLE →

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Resident Live Births, by Birth Weight, Jones County, 1960			
Birth Weight in Pounds	Number of Live Births f	Midpoint X	fX
3.0 - 3.9	1	3.45	3.45
4.0 - 4.9	2	4.45	8.90
5.0 - 5.9	5	5.45	27.25
6.0 - 6.9	11	6.45	70.95
7.0 - 7.9	21	7.45	156.45
8.0 - 8.9	17	8.45	143.65
9.0 - 9.9	9	9.45	85.05
10.0 - 10.9	3	10.45	31.35
11.0 - 11.9	1	11.45	11.45

(h) The number of live births whose birth weight falls within each interval is written beside that interval in the f column. That is, for the *stated* interval 3.0 - 3.9, one live birth had a birth weight of 2.95 lbs. to (but *not* through) 3.95 lbs.; for the interval 4.0 - 4.9, two live births had a birth weight of 3.95 lbs. to (but *not* through) 4.95 lbs., etc.

(i) To find the midpoint of each *stated* interval, we average the lower and upper limit of the *actual* interval. That is, for the *stated* interval 3.0 - 3.9, the midpoint is $\frac{2.95 + 3.95}{2} = 3.45$ lbs.; for the *stated* interval 4.0 - 4.9, the midpoint is $\frac{3.95 + 4.95}{2} = 4.45$ lbs.; etc.

(j) We have multiplied each single value X (midpoint) by its *interval* frequency and entered the products in the fX column. That is, one times 3.45 is 3.45; two times 4.45 is 8.90, etc.

THE PROCEDURE (continued)

3. (Worktable construction continued)

(k) Draw two parallel lines (for totals) about ½ inch apart directly below the last interval listed.

SEE EXAMPLE →

(l) Between the parallel lines and beneath the last interval listed in your first column, write "TOTAL."

SEE EXAMPLE →

(m) Add the frequencies listed in your f column and write the total in the TOTALs row; label the total "N."

SEE EXAMPLE →

(n) Add the products listed in the fX column and enter the total in the TOTALs row; label the total "ΣfX."

SEE EXAMPLE →

ERIC

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Resident Live Births, by Birth Weight, Jones County, 1960

Birth Weight in Pounds	Number of Live Births f	Midpoint X	fX
3.0 - 3.9	1	3.45	3.45
4.0 - 4.9	2	4.45	8.90
5.0 - 5.9	5	5.45	27.25
6.0 - 6.9	11	6.45	70.95
7.0 - 7.9	21	7.45	156.45
8.0 - 8.9	17	8.45	143.65
9.0 - 9.9	9	9.45	85.05
10.0 - 10.9	3	10.45	31.35
11.0 - 11.9	1	11.45	11.45
TOTAL	N=70		$\Sigma fX = 538.50$

(k) Totals

(l) We have labeled the totals row.

(m) We have added all our individual frequencies (Σf) and written the total with its appropriate symbol. You could get the same number (70) if you counted the number of values in our original list of data (see Step 1(a), page 21B).

(n) We have added all the individual products and written the total with its appropriate symbol.

THE PROCEDURE (continued)

4. Substitute the totals from your worktable into the computational formula for the mean:

$$\overline{X} = \frac{\Sigma fX}{N}$$

SEE EXAMPLE →

5. To complete the computation, perform the division function indicated by the formula. Because the values in your data are continuous, you must round off your answer (quotient) to *no more than* the same number of decimal places in the original values. If desirable, your answer could be rounded to have *fewer* decimal places than the original values.

SEE EXAMPLE →

6. Turn now to page 33A for a general guide to the use of the arithmetic mean you have just computed.

THE END

ARITHMETIC MEAN FOR GROUPED (INTERVAL), CONTINUOUS DATA (ROUNDED)

ERIC

AN, EXAMPLE (continued)

4.

WORKTABLE: Resident Live Births, by Birth Weight, Jones County, 1960			
Birth Weight in Pounds	Number of Live Births f	Midpoint X	fX
3.0 - 3.9	1	3.45	3.45
4.0 - 4.9	2	4.45	8.90
11.0 - 11.9	1	11.45	11.45
TOTAL	$N = 70$		$\Sigma fX = 538.50$

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{538.50}{70}$$

5. In our example:

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{538.50}{70} \approx 7.692\ldots$$

$\approx$ 7.7 lbs. (We round off at least *to the nearest tenth of a pound* because our original values were stated to the nearest tenth; or we could round to the nearest pound if desired, that is, $7.692\ldots \approx 8$ lbs.)

6. Turn now to page 33B for an example of the use of an arithmetic mean.

THE END

ARITHMETIC MEAN FOR GROUPED (INTERVAL), CONTINUOUS DATA (ROUNDED)

ERIC

FOR GROUPED (INTERVAL), CONTINUOUS DATA (NON-ROUNDED)—THE PROCEDURE

1. You should be using this procedure only . . .
   (a) if your data consists of 50 or more values (cases, observations, readings, etc.).

   SEE EXAMPLE ⟶

   (b) if the values of your data are *not rounded* to the nearest unit or fraction of a unit, but rather all fractions of a unit are dropped entirely (age at *last* birthday is the most common such value).

   SEE EXAMPLE ⟶

   (c) if the values of your data are continuous, i.e., can be measured in fractional form on a continuous scale.

   SEE EXAMPLE ⟶

   (d) if, when assuming for the moment th the continuous values of the data are discrete and ignoring all decimal poi the difference between the smallest value given and the largest value giv more than 14 (this means that the number of different *major* values pc e is more than 15).

   SEE EXA PLE ⟶

2. The formula used to compute the arithmetic mean is:

$$\bar{X} = \frac{\Sigma fX}{N}$$

For easy reference you may write this formula at the top of the paper on which you are to do your computation.

SEE EXAMPLE ⟶
(For basic definition of symbols)

FOR GROUPED (INTERVAL), CONTINUOUS DATA (NON-ROUNDED)—AN EXAMPLE

1.

(a) The following is a list of ages *in whole years* of women admitted to prenatal service in Jones County during the calendar year of 1960: 31, 23, 26, 16, 21, 18, 37, 41, 22, 29, 20, 19, 21, 32, 24, 27, 22, 20, 21, 30, 23, 43, 17, 22, 19, 38, 30, 20, 18, 22, 25, 19, 16, 28, 20, 21, 17, 23, 37, 33, 28, 23, 39, 24, 22, 23, 17, 27, 42, 16, 18, 21, 31, 26, 21, 37, 22, 20, 28, 19, 29, 47, 18, 16, 23. [Note that there are 65 values (cases) in the list of data.]

RETURN TO THE NEXT STEP IN THE PROCEDURE

(b) The values in our data are ages at *last* birthday. Another such value is months of gestation, e.g., both 3 months 5 days and 3 months 25 days are actually stated as 3 months gestation.

(c) In the above data the values are continuous because age is measured on a continuous scale. Age is stated in that form (fractional or whole number) which can be most accurately approximated (measured) and most conveniently used. We have chosen to state our values in whole years by using age at last birthday. Other types of continuous values are weights, millimeters of blood pressure, height, and length of gestation.

(d) Although the values in our above list of data are continuous, if we assume for the moment they are discrete, then the difference between the smallest value given (15 years) and the largest value given (47 years) is 32. Therefore, more than 15 different major values are possible; actually, 33 different major values are possible $(47 - 15 + 1 = 33)$. [NOTE: In this type of computation decimal points are ignored; e.g., 1.047 and 1.015 would be treated as whole numbers 1047 and 1015.]

2. $\overline{X}$ refers to the arithmetic mean.
X refers to a single value (in the present case it will be an interval midpoint).
f refers to the frequencies with which values occur.
N refers to the total number of values.
Σf refers to the sum of (Σ) the frequencies (f)—Note: $N = \Sigma f$.
ΣfX refers to the sum of (Σ) the frequencies (f) times the single value (X) to which it refers.

GO TO NEXT PAGE

THE PROCEDURE (continued)

3. For convenience in computation, prepare a worktable as follows:
   (a) Though not always essential, a title that clearly describes the content is sometimes desired. If the table is to be used for general communication (published, etc.) then it should contain the What (the group being studied), How (the characteristics of the group being allowed to vary), Where (the area in which the data was gathered), and When (the time period in which the data was gathered) *in that order*.

   SEE EXAMPLE →

   (b) Draw two parallel lines (the head) about one inch apart directly below the worktable title; then divide the head and the rest of the page into four equal columns.

   SEE EXAMPLE →

   (c) In the head of the first column write the name of the value being considered and the unit by which it is being measured.

   SEE EXAMPLE →

   (d) In the head of the second column, write the name of that which reflects the frequencies with which values occur; include the symbol "f."

   SEE EXAMPLE →

   (e) In the head of the third column, write the word "Midpoint" and the single value symbol "X," which in this case will be represented by the interval midpoint.

   SEE EXAMPLE →

   (f) In the head of the fourth column write only the symbol "fX."

   SEE EXAMPLE →

AN EXAMPLE (continued)

3.

(a) We have decided to use a formal (publishable) title for our worktable below; compare it with the original statement given in Step 1 (a), page 27B. . . . Notice that the information is listed *in order:* What, How, Where, When.

WHAT

WORKTABLE: Distribution of Women Admitted to Prenatal Service, by Age in Years, Jones County, 1960

HOW WHERE WHEN

(b) Heads and Columns

Age in Years	Number of Women f	Midpoint X	fX

(c) We have written in the name of the value being considered (age) and the unit by which it is being measured (years).

(d) The number of women in a certain age group (interval) reflects the frequency (f) with which this age group occurs in our list of data.

(e) We have labeled the third column "Midpoint" and included the symbol "X."

(f) We have labeled the fourth column "fX."

THE PROCEDURE (continued)

3. (Worktable construction continued)

(g) List your values in the first column of the worktable so that you will have from 7 - 15 intervals. To determine interval *size* subtract your lowest value from the highest, divide by 10, and round off to the *same number* of decimal places as in your original values. Your smallest value should be included in your first interval as either the lower limit of the interval or as some value within the interval; your largest value should be included in the last interval.

[NOTE: Some latitude is allowed in deciding the actual interval size. The above computational method gives you an approximate interval size that can be used. However, you may increase or decrease the actual interval size *as long as you do not exceed the 7 - 15 limit on the number of intervals.* This flexibility permits you to select an interval size that is more readable (e.g., 5) for tables that are to be published. *IMPORTANT:* This practice should not be abused—computationally, it is best to have the smallest possible interval size.]

SEE EXAMPLE →

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of Women Admitted to Prenatal Service, by Age in Years, Jones County, 1960

Age in Years	Number of Women f	Midpoint X	fX
15 - 19	15	17.5	262.5
20 - 24	25	22.5	562.5
25 - 29	10	27.5	275.0
30 - 34	6	32.5	195.0
35 - 39	5	37.5	187.5
40 - 44	3	42.5	127.5
45 - 49	1	47.5	47.5

(g) The smallest value in our data is 16 years; the largest is 47 years; therefore, (47 — 16) ÷ 10 = 3.1 or 3 years is the computed interval size. However, we decided to increase the actual interval size to 5 years since this would make the table more readable and still produce at least 7 intervals (and less than 16). Again for readability, we decided to make the lower limit of the first interval 15 years. To actually determine the interval, the value 15 was listed to the left in the first column; then the interval size "5" was successively added until the interval was produced that would include our largest value 47 years. Once the lower limit of each interval was written, we wrote in the upper limits.

[NOTE: To make our table more appropriate computationally, we would have used the 3 year interval size as follows:

16 - 18
19 - 21
22 - 24
25 - 27
28 - 30
31 - 33
34 - 36
37 - 39
40 - 42
43 - 45
46 - 48

11 intervals

Notice that we would still have had less than the maximum (15) number of intervals recommended.]

THE PROCEDURE (continued)

3. (Worktable construction continued)

(h) In the f column, for each interval listed in the first column, write the number of values in your data that fall within the interval; this will give you your *interval* frequency. [NOTE: Though not explicitly stated, each interval is assumed to have an *inclusive* upper limit of .999 ....]

SEE EXAMPLE →

(i) In the X column, enter the midpoint of each interval; this single value is found by adding ½ the interval size to the lower limit of each interval.

SEE EXAMPLE →

(j) In the fX column, for each midpoint value listed, enter the products of the f column times the X column.

SEE EXAMPLE →

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of Women Admitted to Prenatal Service, by Age in Years, Jones County, 1960			
Age in Years	Number of Women f	Midpoint X	fX
15 - 19	15	17.5	262.5
20 - 24	25	22.5	562.5
25 - 29	10	27.5	275.0
30 - 34	6	32.5	195.0
35 - 39	5	37.5	187.5
40 - 44	3	42.5	127.5
45 - 49	1	47.5	47.5

([illegible] ber of women whose age falls within each interval is written beside each interval in the f column. That is, 15 women admitted to prenatal service had ages within the 15 - 19.999 ... year interval, 25 women had ages within the 20 - 24.999 ... year interval, etc.

(i) To find the midpoint of each interval we added ½ times 5 (the interval size) or 2.5 years to the lower limit of each interval. That is, 15 + 2.5 is 17.5 years; 20 + 2.5 is 22.5 years; etc.

(j) We have multiplied each single value X (midpoint) by its *interval* frequency and entered the products in the fX column. That is, 15 times 17.5 is 262.5; 25 times 22.5 is 562.5; etc.

THE PROCEDURE (continued)

3. (Worktable construction continued)

(k) Draw two parallel lines (for totals) about ½ inch apart directly below the last interval listed.

SEE EXAMPLE ⟶

(l) Between the parallel lines and beneath the last interval listed in your first column, write "TOTAL."

SEE EXAMPLE ⟶

(m) Add the frequencies listed in your f column and write the total in the TOTAL s row; label the total "N."

SEE EXAMPLE ⟶

(n) Add the products listed in the fX column and enter the total in the TOTAL s row; label the total "ΣfX."

SEE EXAMPLE ⟶

ERIC

AN EXAMPLE (continued)

3. (Worktable construction continued)

WORKTABLE: Distribution of Women Admitted to Prenatal Service, by Age in Years, Jones County, 1960

Age in Years	Number of Women f	Midpoint X	fX
15 - 19	15	17.5	262.5
20 - 24	25	22.5	562.5
25 - 29	10	27.5	275.0
30 - 34	6	32.5	195.0
35 - 39	5	37.5	187.5
40 - 44	3	42.5	127.5
45 - 49	1	47.5	47.5
TOTAL	$N=65$		$\Sigma fX=1657.5$

(k) Totals

(l) We have labeled the totals row.

(m) We have added all our individual frequencies (Σf) and written the total with its appropriate symbol. You could get the same number (65) if you counted the number of values in our original list of data (see Step 1(a), page 27B).

(n) We have added all the individual products and written the total with its appropriate symbol.

THE PROCEDURE (continued)

4. Substitute the totals from your worktable into the computational formula for the mean:

$$\overline{X} = \frac{\Sigma fX}{N}$$

SEE EXAMPLE ⟶

5. To complete the computation, perform the division function indicated by the formula. Because the values in your data are continuous, you must round off your answer (quotient) to *no more than* the same number of decimal places in the original values. If desirable, your answer could be rounded to have *fewer* decimal places than the original values.

SEE EXAMPLE ⟶

6. Turn now to page 33A for a general guide to the use of the arithmetic mean you have just computed.

THE END

ARITHMETIC MEAN FOR GROUPED (INTERVAL), CONTINUOUS DATA (NON-ROUNDED)

AN EXAMPLE (continued)

4.

WORKTABLE: Distribution of Women Admitted to Prenatal Service, by Age in Years, Jones County, 1960			
Age in Years	Number of Women f	Midpoint X	fX
15 - 19	15	17.5	262.5
20 - 24	25	22.5	562.5
40 - 44	3	42.5	127.5
45 - 49	1	47.5	47.5
TOTAL	$N = 65$		$\Sigma fX = 1657.5$

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{1657.5}{65}$$

5. In our example:

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{1657.5}{65} = 25.5\ldots$$

$= 26$ years (We round off at least *to the nearest whole number* because our original values were stated as age at the last birthday, a whole number.)

6. Turn now to page 33B for an example of the use of an arithmetic mean.

THE END

ARITHMETIC MEAN FOR GROUPED (INTERVAL), CONTINUOUS DATA (NON-ROUNDED)

ERIC Full Text Provided by ERIC

USE OF THE ARITHMETIC MEAN — IN GENERAL

Regardless of the particular method of computation, your arithmetic mean may be used in many specific but similar ways. Listed below are the *most commonly accepted* uses of an arithmetic mean. *IMPORTANT:* All means will *not* necessarily be used in every way listed; conversely, means may be used acceptably in some other manner.

The arithmetic mean you have just computed may be used . . .

(a) to project future program requirements (you must also have knowledge of changes in program emphasis and estimates of future number of events—observations, admissions, visits, etc.); and

(b) to compare with other related means, for example . . .

- recommended standards (means) for the *same* time period, population, and geographic area;
- means of different populations or geographic areas but within *same* time period; or
- means of previous (but comparable) time period but for *same* population and geographic area; etc.

SEE EXAMPLE ⟶

USE OF THE ARITHMETIC MEAN—AN EXAMPLE

In Walker County health clinics, 18 women who delivered in 1960 were admitted for prenatal service. These *18 admissions* made a total of *66 visits* for a mean average of 3.7 visits. It was during this year that a new educational program on clinic visits was started.

Matching the general uses listed on the previous page item for item, we can use the computed mean of 3.7 visits in the following *specific* ways:

(a) Because we will be further emphasizing our newly started educational program on clinic visits, we are expecting the average to increase from 3.7 visits per prenatal admission to 4.0 next year. Also, because of the emphasis in education and the general increase in the childbearing female population, we expect admissions to increase from 18 to 28. Therefore, estimated clinic visits for next year are expected to be $28 \times 4 = 112$ (number of estimated admissions *times* average estimated visits per admission). This estimate of 112 (compared to 66 total visits for the past year) requires that the program director decide if an increase in number of clinics and/or staff is required.

(b)

- The director's expectation of 3.4 visits per prenatal admission was somewhat exceeded by this year's average of 3.7.
- Compared to an adjacent county's 4.2 visits, ours was somewhat less for this year; within our own county for this year, an average of 3.0 visits for the lower socioeconomic groups suggests need for increased efforts within this group.
- Compared to last year's average of 2.9 visits, we may assume that our recently introduced educational program is at least partially effective.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1971 O—427-543

RESULTS OF FIELD DEMONSTRATIONS

Field demonstrations of the *Guide: Arithmetic Mean* were held at the Center for Disease Control, Atlanta, Ga., and at the Los Angeles County Health Department, Los Angeles, Calif. The *Guide: Arithmetic Mean* is one part of a three-part course on *Descriptive Statistics for the Health Professions.* Other parts of the course are the *Guide: Median* and the prerequisite for both guides, the *Lesson, Measures of Central Tendency.*

Some 33 students at CDC completed the *Lesson* and were given the *Guide: Arithmetic Mean* with sample problems to be completed on a take-home basis within a week. Each student was evaluated not only on the basis of correct scores, but also on how well he followed the *Guide* in computing the sample problems. Twenty-seven students completed the problems with the following results:

Range = 31% – 100%
Median = 96%

Sixty-one students in Los Angeles worked in a formal classroom setting for three half-day sessions and on a take-home basis. After completing the prerequisite *Lesson,* the class was divided into two groups with one group receiving the *Guide: Arithmetic Mean,* the other, the *Guide: Median.* A total of 4 hours class time was allotted each student. If necessary, he was allowed to complete the sample problems outside class. As each student completed the first guide *(Mean* or *Median),* he was given the other guide to complete. Forty-one students completed the sample problems for the *Guide: Mean* with the following results:

Range = 50% – 100%
Median = 95%

In both groups, reasons for failing to complete the course included students' lacking time due to job responsibilities or thinking the course would be more advanced. There were specific differences between the two test groups. Students at CDC participated voluntarily, while the Los Angeles students had been requested to attend the course. Sixty percent of each group had college degrees, but 33% of the CDC group had post-graduate degrees. In comparison, 8% of the Los Angeles students had post-graduate degrees.

D

HE

ON

AR

ICE

ATIO

IC

ILM

ERIC
Full Text Provided by ERIC