DESCRIPTIVE STATISTICS FOR THE HEALTH PROFESSIONS

LESSON: INTERPRETATION

MEASURES OF CENTRAL TENDENCY

An Instructive Communication

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

SPECIFICATIONS

Instructional Objectives

After taking this *Lesson* as directed the student:

1. Can verbally define *Measure of Central Tendency, Arithmetic Mean, Median,* and *Mode.*
2. Can verbally describe when it is particularly appropriate to use the *Arithmetic Mean,* the *Median,* and the *Mode.*
3. Can verbally describe the data needed to compute the *Arithmetic Mean* or the *Median.*
4. Can verbally describe how an *Arithmetic Mean* or a *Median* may be used.
5. Given actual or verbal description of situations and/or data, can name from memory the *Measure of Central Tendency (Arithmetic Mean, Median* or *Mode)* most appropriate for use.
6. Given data (raw or in tabular form), can match it with certain descriptive factors: continuous, discrete, $N < 50$, $N \geq 50$, and value range of > 14 or < 15.

[See *Limitations, Restrictions,* and *Special Characteristics* below.]

Primary Trainee Population

Public Health nurses and sanitarians with college degrees or equivalent.

Secondary Trainee Population

A. Public Health veterinarians, physicians, dentists, and other similarly related Public Health workers with college degrees or the equivalent should also be able to use this *Lesson;* however, the examples used in this booklet will not be relevant to this group.

B. With proper motivation and some additional effort, Public Health nurses and sanitarians with a high school education should also be able to use this *Lesson.*

Student Study Time

This *Lesson* should require from 2-4 hours, exclusive of breaks. We suggest that the student take a break at least every 1 to 1½ hours. The student should make every effort to complete the *Lesson* within a two-day period.

Individualization

At least 20-25% of the frames may be skipped by the student, depending on his own needs. Of course, there is no time limit imposed–the student may proceed at his own best rate.

Limitations, Restrictions, and Special Characteristics

A. The verbal definitions required of the student (see *Instructional Objectives* above) are brief and nontechnical.

B. The *Lesson* does not teach the student the procedures and techniques for computing the Measures of Central Tendency presented. However, the student should be able to use the companion computational guides more efficiently (less time–fewer errors).

C. For maximum effectiveness for both this *Lesson* and its companion computational guides, we suggest you follow the study of this *Lesson* with the use of the guides as soon as possible.

DESCRIPTIVE STATISTICS FOR THE HEALTH PROFESSIONS

LESSON: INTERPRETATION

MEASURES OF CENTRAL TENDENCY

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. 2192
U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1971

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price 75 cents
Stock Number 1723-0047

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—resonsible 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 *Lesson* on the *selection* and use of the appropriate measure of central tendency. The *Lesson* should be taken prior to the use of its companion *Guides, Arithmetic Mean: Computational Guide* and *Median: Computational Guide.* A unique characteristic of this *Lesson* is that *computational techniques,* easily forgotten or made vague through disuse, are not taught. Such detailed techniques are covered in the *Guides* which are to be used when an actual need arises. Techniques are mentioned in the *Lesson* only as is necessary to make more meaningful the definitions of the specific measure of central tendency.

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

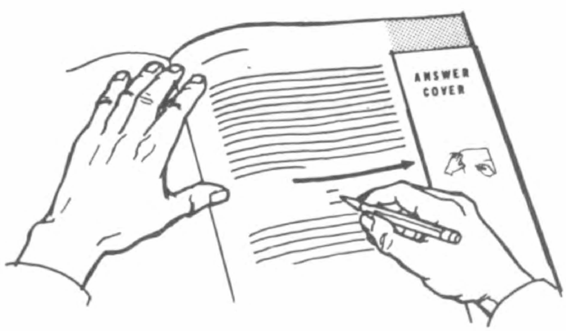
HOW TO USE THIS LESSON

This *Lesson* is probably different from any you have ever used before–it is certainly different from the usual textbook or study manual.

On almost every page of the *Lesson* you will be asked to answer questions about what you are studying. Because of this "question and answer" way of teaching, many people confuse this type of lesson with a test. But *it is not a test;* it is a lesson that asks questions often and at the best times to help you learn. The questions help you think about what you are learning; the correct answers are given so that you can immediately see that you answered correctly and that you are learning.

1 In the front of this booklet you saw a small strip of cardboard called the ANSWER COVER. If you have not already done so, remove it....*Now, place the ANSWER COVER over the gray part of the page on the right* ⟶
The ANSWER COVER should now cover the entire gray area so this page looks like the one pictured below; does it? **yes / no** (Draw a circle around the correct answer.)

Place ANSWER COVER over this area.



You should have circled one of the "yes" or "no" answers with your pencil. If you did not, do so now....That was an easy question to answer, but to see that you are correct move the ANSWER COVER down the page until its top is even with the line below

If you have been following instructions, you should have circled the yes answer.

---

2 Have you read what is written in the gray area? If not, do so now.

You see from what you just did that when a slash (/) is used to separate two or more words you must circle the correct answer. You may also be asked to check (√) the right answer or write your answer in a blank. For example, answer the questions below:

1. This **is / is not** a test. (Circle the correct answer.)
2. Is this a lesson to help you learn? (Check the correct answer.)

   ______ yes

   ______ no
3. This *Lesson* is part of course on Descriptive ________________ for the Health Professions. (Write your answer in the blank.)

Now see if you have given the correct answers ⟶

1. is not
2. √ yes
3. Statistics

BEFORE YOU READ ANY FURTHER BE SURE TO COVER THE GRAY AREA WITH THE ANSWER COVER. DO THIS EACH TIME YOU START A NEW PAGE.

3 You will notice that what you are studying is divided into parts containing various amounts of information and questions. These parts are called "frames."

Most frames have (1) a certain amount of information and
(2) questions for you to answer about the information in that frame or about other information you have studied before.

Is what you are now reading part of a frame? (Check one.)

_______ yes
_______ no

Don't forget to see if your answer is correct after you have written it; move the ANSWER COVER down to the next line

√ yes–what you were just reading is a frame; it gives information *and* also asks a question.

---

4 You will be able to answer many of the questions correctly. However, when you are wrong you should do which one of the following:

_______ 1. Change your written answer; then go to the next frame.
_______ 2. Try to to see why you were wrong; *then* change your written answer and go to the next frame.
_______ 3. Go to the next frame.
_______ 4. Start over again.

√ 2. Don't just copy a correct answer if you are wrong; try to see why you are wrong: then change the answer and continue.

---

5 The correct answer to the last question is very important.

Just copying the correct answer when you are wrong will not help you learn as you should.

Looking at the correct answer before you write your own answer to the question will not help you learn as you should.

Copying your answers will make a difference *only to you* since it will keep you from learning as well as you might otherwise. This failure to learn the material will show up later in post-lesson testing or in on-the-job performance.

To learn as you should you must:

_______ 1. read everything carefully.
_______ 2. follow instructions.
_______ 3. write your answers before looking at the correct answers given.
_______ 4. try to see why you were wrong–don't just copy the correct answer when you make a mistake.
_______ 5. take all the time you need–this lesson was written so that you can set your own pace.

√ 1.
√ 2.
√ 3.
√ 4.
√ 5.

You should have checked all the answers.

INTRODUCTION

6 Because portions of the next several pages appear in part in other booklets of the course on *Descriptive Statistics for the Health Professions,* you may encounter some slight repetition if you have already studied one or more of the other booklets. However, we strongly recommend you give the entire Introduction your full attention.

No Answer Needed

---

7 Professionals in Public Health frequently use statistical methods to *describe* or predict (*infer*). However, these two classifications of statistics–descriptive and inferential–are not mutually exclusive; we must describe before we can infer. For example, descriptive statistics may be used to show that more men than women died from Disease "D," but without inferential statistics we could not infer that there was a real rather than a chance difference between men and women with regard to Disease "D," nor would we be able to predict that there would continue to be such a difference in the future.

Of the five statistics we have classified as "descriptive," this *Lesson* is concerned with (circle one) . . .

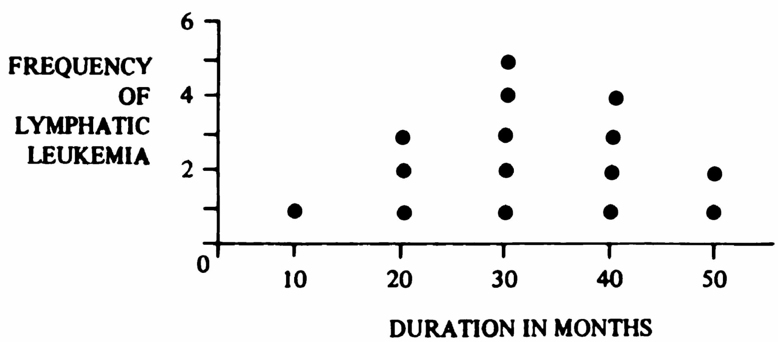
1. tables
2. graphs
3. descriptive ratios
4. measures of central tendency
5. measures of dispersion

(4) measures of central tendency

8 Before we go any further in our specific discussion of "measures of central tendency," let's consider the basic working materials of descriptive statistics.

Example:

A frequency distribution based on laboratory data from Warren County Hospital in 1966 is represented in crude graph form below:



Now see if you can answer the following questions without spending too much time on the ones you don't know:

1. Give a proper verbal description of the *group* being considered in the graph.

 ____

 ____

2. What is the total frequency of the group? ________________
3. What is the factor (variable) being studied (allowed to vary)?

 ____

4. What is being distributed? ______________________
5. How many cases are there for each value as represented in the graph?

 ____

6. Could male cases be included in the group?

 ________ yes

 ________ no

7. Could cases discovered in 1967 while the report was being prepared be included?

 ________ yes

 ________ no

Check your answers.

If you answered all the questions correctly, skip to Frame 26.

If you could not answer all the questions, go on to the next frame.

1. Cases of laboratory-confirmed lymphatic leukemia at Warren County Hospital in 1966.
2. 15
3. duration (of leukemia) in months
4. The total frequency (15) is being distributed among the various months of duration.
5. 1-10's, 3-20's, 5-30's, 4-40's, 2-50's.
6. √ yes
7. √ no

9 Descriptive statistics may be thought of as a way of describing, in numerical terms, something about GROUPS of "cases" (people or events) having common characteristics. That is, *all* cases of the groups are matched (identical) with regard to certain characteristics. For example . . .

THIS GROUP	HAS THESE CHARACTERISTICS IN COMMON FOR ALL ITS CASES
Cases of laboratory-confirmed canicola fever in Columbus, Georgia, 1962.	(1) *all* were diagnosed-laboratory-confirmed–as canicola fever (2) *all* occurred in Columbus, Georgia (3) *all* occurred during 1962

Although the common characteristics that are made explicit restrict and control the group, certain other characteristics may be true of the group and may be allowed to vary. For example, check (√) the cases below that *may* be included in the group described above:

_______ cases not laboratory confirmed
_______ cases laboratory confirmed
_______ male cases
_______ female cases
_______ cases in Columbus, Ga.
_______ cases not in Columbus, Ga.
_______ cases over 21 years old
_______ cases under 21 years old
_______ cases in high socioeconomic setting
_______ cases in low socioeconomic setting
_______ cases not in 1962
_______ cases in 1962

___ cases not laboratory confirmed
√ cases laboratory confirmed
√ male cases
√ female cases
√ cases Columbus, Ga.
___ cases not in Columbus, Ga.
√ cases over 21 years old
√ cases under 21 years old
√ cases in high socio-economic setting
√ cases in low socio-economic setting
___ cases not in 1962
√ cases in 1962

---

10 Sometimes characteristics that apply to only part of a group are used in a statement about the group. For example . . .

READ THIS STATEMENT	AND LIST ONLY THE CHARACTERISTICS ALL MEMBERS OF THE GROUP HAVE IN COMMON
Of a group of 185 11-year-old boys, many of whom weigh about 60 pounds, each is 56 inches tall, three-fourths are in the 5th grade, and the rest are in other grades.	*All* ____________________ *All* ____________________ *All* ____________________

All are 11 years old
All are boys
All are 56 inches tall

11 Sometimes we tend to think of the size (count) of a group as a common characteristic. This is obviously not correct, as you can easily see by trying to describe just one case in the group in Frame 10; for example, "The case must be a boy, 11 years old, 56 inches tall (but certainly not '185')." In descriptive statistics the terms "total frequency" or "number" are usually used to refer to the ________________ of a group.

size (count)

---

12 We have already pointed out that for any group, many characteristics are *not* held constant but rather are allowed to vary. In fact, a group is often defined in order to see how certain characteristics VARY with respect to the group. In the statement below . . .

> "Of 185 11-year-old boys, each is 56 inches tall, many weigh about 60 pounds, 3/4 s are in the 5th grade, and the rest are in the other grades."

We can see that for the group implicit in the statement, weight is certainly not a common characteristic that is being held constant. If we wish to study more exactly how weight varies, we might formalize the statement thusly:

> "A group of 185 56-inch-tall, 11-year-old boys, *by weight in pounds.*"

Now, you list the following about the above proposed study . . .

THESE ARE THE COMMON CHARACTERISTICS THAT DEFINE THE GROUP	AND THIS IS THE COMMON CHARACTERISTIC THAT WILL VARY
________________	________________
________________	
________________	

The number 185 is the ________________________ for the group.

Common characteristics of the group:
height (56 inches)
age (11 years old)
sex (boys)
Varying characteristic:
weight in pounds
185 is the number or total frequency

13 In most instances the characteristic that is being studied (allowed to vary) is preceded by a certain preposition; circle this preposition in the following observation . . .

"Twenty-five cases of laboratory-confirmed canicola fever, by age, in Columbus, Georgia, 1962.

. . . fever, by age, ...

---

14 In a study, the common characteristics of a group are held constant while one or more common characteristics are allowed to ________________.

vary

---

15 The common characteristic that is allowed to vary may be referred to as the study ____________ able.

vari able

---

16 A number of cases having certain constant common characteristics is called a ________.
Its size or count is called the ________.
The common characteristic that is allowed to vary is called the study ____________________ and is usually preceded by the preposition ________________ in a formal statement of the study.

group
total frequency or number
variable
by

17 In the two observations below . . .

underline the common characteristics of the group

double underline the total frequency of the group

circle the study variable

Example 1

"Distribution of 25 males, ages 25-50 years, by grams (g.) of hemoglobin per 100 milliliters (ml.) of blood, Washoo County, 1960.

Example 2

"Distribution of 25 males with 16-17 grams (g.) of hemoglobin per 100 milliliters (ml.) of blood, by age, Washoo County, 1960.

Example 1: "Distribution of 25 males, ages 25-50 years, by grams (g.) of hemoglobin per 100 milliliters (ml.) of blood, Washoo County, 1960."

Example 2: "Distribution of 25 males with 16-17 grams (g.) of hemoglobin per 100 milliliters (ml.) of blood, by age, Washoo County, 1960."

---

18 In only one of the two examples above is the unit by which the study variable is to be measured made explicit. This is not unusual when such information is assumed to be implicit in the name of the variable itself, i.e., in Example 2 the missing unit of measure is

________________________.

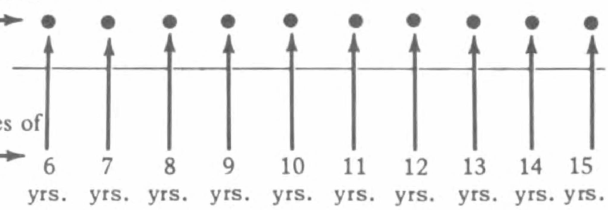
years at last birthday

19 We are now ready to discuss "frequency distributions." Notice that the two statements in the last frame begin with the word "distribution." Actually, this is a standard way of saying that we are going to look at the particular way in which *individual frequencies* of the group are *distributed* among the various values of the study variable. A frequency distribution is often represented graphically. We see this in oversimplified form as follows:

"Distribution of 10 boys with measles by age . . . ."

Each dot represents a boy with measles; there are 10 boys (total frequency) ⟶

and these are the various values of the study variable (age) ⟶

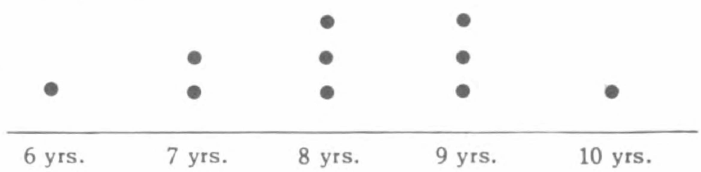


In the above representation we can see that the individual frequencies (totaling 10) are distributed among the various values of the study variable so that no age value is represented by more than ______ case(s) of measles. one

---

20 Often, individual frequencies are distributed so that a particular age value is represented by more than one case. For example . . .

Dots represent the frequency ⟶ distributed among the values of the study variable ⟶



According to the frequency distribution, how many cases are

6 years old? ______ 8 years old? ______ 1 3

7 years old? ______ 9 years old? ______ 2 3

10 years old ? ______ 1

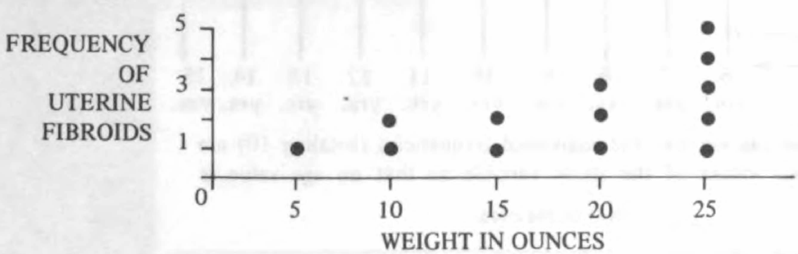
21 The relationship of the frequency of a group to the various values of the study variable is referred to as the ______________________________.

frequency distribution

22 In the last several frames we have referred to frequency distributions used in examples by a single word, as in "______________ of 185 11-year-old boys by weight . . . ."

Distribution

23 You should now be able to answer all the questions about the following distribution based on laboratory data from the Lowin County Clinic in 1964:



1. What is a proper verbal description of the group being considered in the crude graph?
____
2. What is the total frequency for the group? ______________
3. What is the study variable? ______________
4. What is being distributed? ______________
5. How many cases are there for each value (weight)? ______________
6. Could cases discovered in 1965 while the report was being prepared be included?
   ______ yes
   ______ no

1. Laboratory-confirmed uterine fibroids diagnosed at Lowin County Clinic in 1964.
2. 13
3. weight in ounces
4. The frequency (13) is being distributed among the fibroid weights (in ounces).
5. 1-5, 2-10, 2-15, 3-20, 5-25
6. √ no

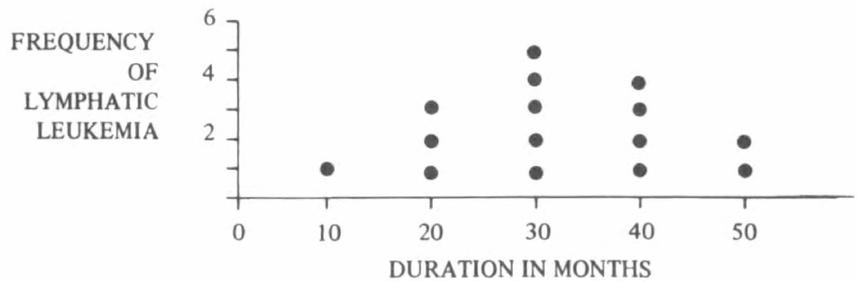
24 If you made no errors in the last frame, go directly to Frame 26.

If you made errors, study for a moment why and then go to the next frame.

No Answer Needed

---

25 A frequency distribution based on laboratory data from Warren County Hospital in 1966 is represented in crude graph form below.



1. Describe the group. ____________________
2. What is the group's total frequency? ____________________
3. What is the study variable? ____________________
4. What is being distributed? ____________________
5. How many cases are there for each value? ____________________
6. Is the group restricted by sex?
   ______ yes
   ______ no
7. Is the group restricted by time of occurrence?
   ______ yes
   ______ no

1. Cases of laboratory-confirmed lymphatic leukemia at Warren County Hospital in 1966.
2. 15
3. duration (of leukemia) in months
4. The frequency (15) is being distributed among the various months of duration.
5. 1-10, 3-20, 5-30, 4-40, and 2-50
6. √ no
7. √ yes

A MEASURE OF CENTRAL TENDENCY

26 If this lesson had been called "averages" rather than "Measures of Central Tendency," more people would know from its title what is being taught. How do you now define "Measures of Central Tendency"–write your definition below:

____

____

____

____

____ See Next Frame

---

27 Does your definition match in meaning, if not stated word-for-word, the one below:

> "A measure of central tendency is a value that is used to represent the center of a distribution of values. It is considered to be a representative value which can be used in place of numerous individual values."

If your definition matches the one above, go now directly to Frame 33.

Or, if not, does your definition match the one below better:

> "A value obtained by adding all the individual values of a distribution and dividing by the number of values. It is considered to be a representative value which can be used in place of numerous individual values."

If so, go to the next frame. No Answer Needed

---

28 The second definition given in the last frame is probably what most people think of as a measure of central tendency or average. However, it more appropriately identifies a particular *type* of average known technically as the "arithmetic mean." Therefore, your answer, though not correct, is not wholly incorrect–just too specific as you will see as you continue reading. No Answer Needed

29 The definition for measure of central tendency (average) we expect you to learn is: "A value that is used to *represent* the *center* of a distribution of values. It is considered to be a *representative* value which can be used in place of numerous *individual values.*"

Although the definition you used originally may suggest the same meaning as the one above, a key word that states the function of a Measure of Central Tendency and that should be used in its definition is ______________________.

represent (representative)

---

30 The average may be thought of as representing (a) the ______________ of the distribution and (b) the individual ______________ of the distribution.

center
values

---

31 A definition of "Measure of Central Tendency" must indicate that its function is to ______________ the ______________ of a distribution and that it may be used to ______________ the ______________ values of the ______________.

represent
center
represent
individual
distribution

---

32 See if you can now state from memory the definition of Measure of Central Tendency we will use in this *Lesson:*

____

____

____

____

In your own words you should have said that it is a value which *represents* the *center* of a distribution and can be used to *represent* the *individual values* of the distribution.

SPECIFIC MEASURES OF CENTRAL TENDENCY

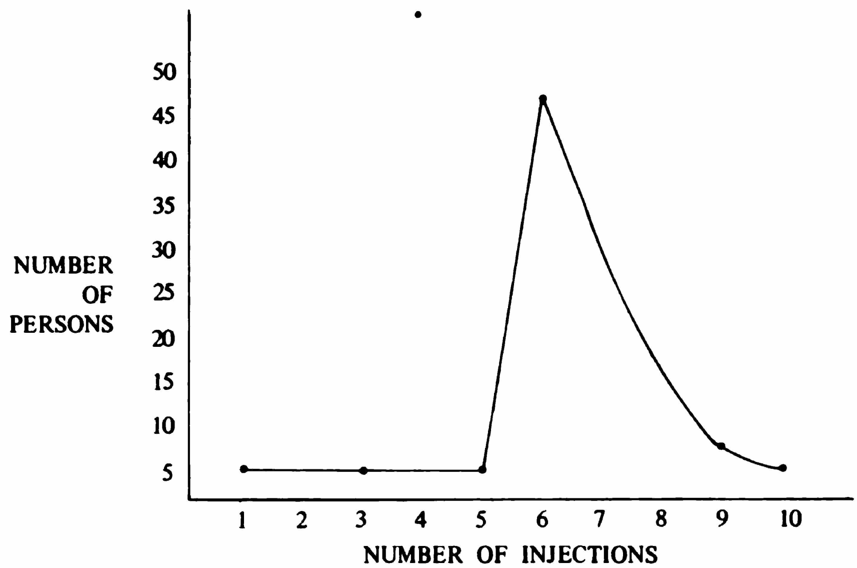
33 We will now discuss the three most commonly used Measures of Central Tendency: the arithmetic mean, the median, and the mode.

No Answer Needed

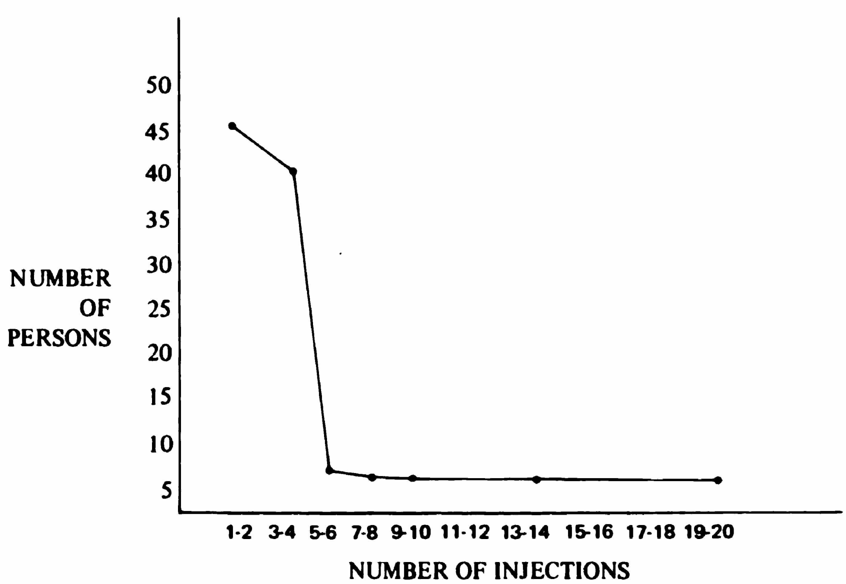
SPECIFIC MEASURES OF CENTRAL TENDENCY: MODE

34 In this *Lesson* we are defining mode as: "A Measure of Central Tendency which, for any list of values, is the single value or group of values which occurs most often." The mode in each of the two simple distributions shown graphically below is the number of injections occurring most often. What are the modes in the two distributions:

1. MODE: ________



2. MODE: ________



1. 6 injections per person
2. 1-2 injections per person

35 In part 1. of the last frame six injections was the mode because it was the ______________ value that occurred most often; in part 2., 1-2 injections was the mode because it was the ______________ ______ ______________ that occurred most often.

single

group of values

36 What is the mode for each of the worktables shown below:

1. MODE: __________

WORKTABLE: Patients Dying From Heart Rupture By Age In Years, Los Angeles County Hospital, July 1941-Oct. 1951.

Age in Years*	Number of Patients
50-54	2
55-59	5
60-69	27
70-79	33
80-89	13
	80

*Age at last birthday

2. MODE: __________

WORKTABLE: Distribution of 75 Restaurants By Number Of Inspections During The Year, Center County, 1965

Number of Inspections	Number of Restaurants
2	6
4	12
6	22
8	19
10	11
12	4
	74

1. 70-79 years
2. 6 inspections

37 In part 2. of the last frame, six inspections (the mode) was the ________________ that occurred ________________; in part 1., 70-79 years was the ________________ that occurred ________________

single value
most often
group of values
most often

38 Now state from memory the definition of mode we are using in this *Lesson:*

____

____

____

____

____

In your own words you should have said that the mode is a *Measure of Central Tendency* which, for any list of values, is that *single value* or *group of values* which *occurs most often.*

39 Before you continue, see if you can recall from memory the definition of A Measure of Central Tendency; think your answer.

A Measure of Central Tendency is a value that is used to *represent* the *center* of a distribution of values. It is considered to be a *representative* value which can be used in place of numerous *individual values.*

SPECIFIC MEASURES OF CENTRAL TENDENCY: ARITHMETIC MEAN

40 In this *Lesson* we are defining "arithmetic mean" (or simply "mean") as: "A Measure of Central Tendency obtained by adding all the individual values and dividing by the number of values." Actually when most people use the word "average" (a term roughly synonymous with Measure of Central Tendency), they are talking about the mean.

The college student who computes his quality-point "average" is actually finding his *mean* quality point value. To find the mean he . . .

1. *lists* the quality points he has received for each hour of credit,
2. *adds* up the list of values (quality points), and
3. *divides* the total value by the total number of credit hours.

The ages of a group of children who have the measles are 1, 3, 7, 9, 9, 12, and 15.

What is their mean age? ____________

8 years

Skip to frame 42 if you were correct.

---

41 To find the mean age in the last frame . . .

1. list all ages 1, 3, 7, 9, 9, 12, 15
2. add all values 56
3. divide by number of children $7\overline{)56}$ = 8 years is the group's mean age

No Answer Needed

---

42 Recall (and write) from memory the definition of the mean:

____

____

____

____

In your own words you should have said that the mean is a *Measure of Central Tendency* obtained by *adding* all the *individual values* and *dividing* by the *number of values.*

43 The mean is often referred to as a "weighted" average since the size (weight) of each individual value is mathematically reflected in the mean value for the group. Because the mean is so mathematically sensitive to the size of *all* the individual cases, atypical values–extremely high or low–will tend to bias the mean in their ____________

direction (favor, etc.)

---

44 In the following distribution of ages . . .

5, 6, 20, 22, 23, 24, 26, 27, 30, 31, 32, 34

the mean will be: (check one or more)

_______ (a) lower than it should be to represent the distribution "operationally"

_______ (b) higher than it should be to represent the distribution "operationally"

_______ (c) a mathematically correct Measure of Central Tendency

_______ (d) none of the above

√ (a) (It is biased toward the 5 and 6.)

√ (c) (It is still mathematically correct.)

SPECIFIC MEASURES OF CENTRAL TENDENCY: MEDIAN

45 In this *Lesson* we are defining "median" as: "A Measure of Central Tendency which, for any distribution of values ranked from smallest to largest, is above one half and below the other half of the values."

Below, seven cases have been *ranked* in order from youngest to oldest. The *middle position* is occupied by a case whose value is 11 years:

MIDDLE POSITION

$$\text{AGES IN YEARS} \Rightarrow \overbrace{2\quad 9\quad 10}^{3\text{ Cases}} \quad \overset{\Downarrow}{\underset{\Uparrow}{11}} \quad \overbrace{12\quad 13\quad 15}^{3\text{ Cases}}$$

MEDIAN age = 11 years

The seven cases in the distribution were *ranked* in order from ____________________

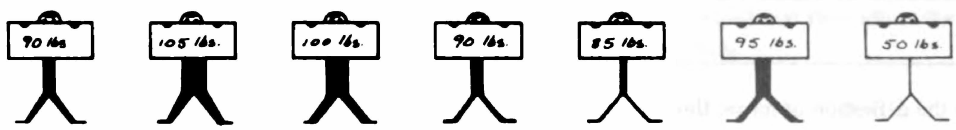
to ____________________ . youngest (2) oldest (15)

Is there an atypical age among the ranked cases–if so, what is it? 2 years

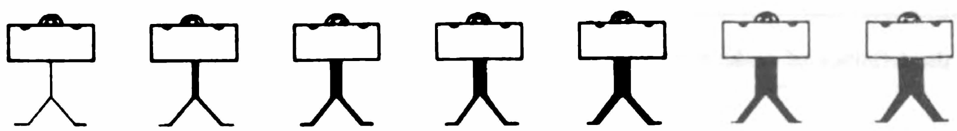
____

46 **The great advantage of the median as a Measure of Central Tendency is that it is *not* a "weighted" average as is the mean. Therefore, it is *not* affected by the extreme value (size) of any case in the distribution. Values are used only to assign the rank *position* to the cases; the value of the case in the middle position is the median value.**

A group of seven cases by weight is shown below . . .



Now, you rearrange the cases in rank order by filling in the *values* below:



Draw an arrow to the middle position in the rank order . . . therefore, what is the median? ________________________

Are there any atypical values (cases) in the distribution?

_______ yes
_______ no

If so, list the value(s). ______________________________

Was the selection of the median affected by the atypical size of the case *value(s)?*

_______ yes
_______ no

90 lbs.

√ yes

50 lbs.

√ no The presence and size of the case obviously affected the median, but the *atypical small size* of the value did not.

47 **Recall and write from** memory the definition of the median: ______________________

In your own words you should have said that the median is a *Measure of Central Tendency* which, for a distribution of values *ranked* from *smallest to largest*, is *above* one half and *below* the other half of the values.

48 **From memory, recall** the definition of mean; think your answer.

The arithmetic mean is a *Measure of Central Tendency* obtained by *adding* all the *individual values* and *dividing by the number* of *values*.

49 **From memory, recall** the definition of mode; think your answer.

The mode is a *Measure of Central Tendency* which, for any list of values, is the *single value* or *group of values* which occurs *most often*.

50 **From memory, recall** the definition of Measure of Central Tendency; think your answer.

A Measure of Central Tendency is a value that is used to represent the *center* of a distribution of values. It is considered to be a *representative* value which can be used in place of numerous *individual values*.

MEAN AND MEDIAN COMPARED

51 Complete the two statements below by filling in the letter a or b:

1. The mean ____________ a. is more mathematically sensitive to the sizes (weights) of the values of a distribution.

2. The median ____________ b. is sensitive to the size of the values of a distribution only to the extent that they affect the ranked position of the values.

1. a
2. b

52 Let's see more specifically how the size (weight) of the individual values affects the mean and median in the following distribution.

The values in this list are in a *random order* with *no* regard to rank.

The *same* values are listed below in *rank order*—from smallest to largest.

4	What is the mean value of the list of values—use the column on the left ______ (mean)	2
6		3
2		4
8		4
7		5
6	What is the median value of the list of values—use the column on the right ______ (median)	6
3		6
4		7
5		8

mean: 5
median: 5
If correct, you may skip Frame 53.

---

53 You should notice that the mean value of 5 could be determined by simply adding the *random* list of values and dividing this total (45) by the total number of values (9).

However, to determine the median of 5, it was necessary to *rank* the values from smallest to largest to find which value represents the middle position below one half the values and above the other half.

No Answer Needed

---

54 For the values in Frame 52, both mean and median are 5. What if the value 8 were changed to 17—would the mean and median still be equal?

4, 6, 2, ~~8~~ 17, 7, 6, 3, 4, 5

_______ yes
_______ no
_______ don't know

√ no (skip to Frame 56 if correct)

_______ yes
_______ don't know
go to next frame if you checked either of these

55 *No,* the new mean and median are not equal when the value 8 is changed to 17. Let's see what it would be–we will demonstrate with only the list that is *ranked* even though it is not needed for the mean . . .

2
3
4
4

We see that 5 still occupies the middle position in the *ranked* data and, therefore, is still the median value → MEDIAN 5 ◄ MIDDLE POSITION

6
6
7
~~8~~ 17

However, the change in size (weight) of one of the values has such a direct effect on the mean that it is changed → 54 ÷ 9 = 6 MEAN

No Answer Needed

---

56 We see that although the mean and median for a distribution may at times be the same, the median is affected simply by relative *position* of values, whereas, the mean is much more sensitive to individual size (weight) of values. Because of the greater mathematical sensitivity of the mean, we pointed out earlier how it is "biased" ("weighted") in the direction of an extreme atypical value–did this happen in the example above?

_______ yes
_______ no
_______ don't know

√ yes (skip to Frame 58 if correct)

_______ no
_______ don't know

go to next frame if you checked either of these

57 *Yes*, by increasing the 8 to 17, an atypical value was introduced to the list and the mean was weighted unrealistically in the direction of the larger values. Study the illustration below:

$$\begin{array}{rr} & 2 \\ & 3 \\ & 4 \\ & 4 \\ \text{MIDDLE POSITION: MEDIAN} \longrightarrow & 5 \\ & 6 \\ & 6 \\ & 7 \\ & \underline{17} \end{array}\left.\vphantom{\begin{array}{r}2\\3\\4\\4\\5\\6\\6\\7\\17\end{array}}\right\}\text{Most values are under 10}$$

$$54 \div 9 = 6 \text{ MEAN}$$

We see that the median value of 5 is more representative of most values than is the mean (6) which increased in order to take into account the extreme value 17.

No Answer Needed

58 A list of ages can also be modified by using open-ended intervals. An open-ended interval is an interval in which only one limit, or "end," is known. Both ">55" (greater than 55) and "<8" (less than 8) are ______________________ intervals.

open-ended

---

59 Now let's add an open-ended interval to our list of ages to see the effect it has on the mean and median:

2
3
4
4
5
6
6 and two values >6 (greater than 6)

Can the median be computed in the above example–check and complete the answer below:

_______ yes, the median value is ______
_______ no, because ______________________
____
_______ don't know

✓ yes, the median value is 5 (skip to Frame 61 if correct)

___ no
___ don't know

go to next frame if you checked either of these

---

60 Yes, the median can be computed even when the data contain an open-ended interval–if you know the frequency involved, and if the median does not fall in the open-ended interval.

In the above example we know that there are 9 values in all (the total frequency or number), we can make an *approximate* ranking to find the value of the middle position, and we know from what's given that the middle position does *not* fall in the open-ended interval. With this in mind, what is the median for the distribution below:

46, 40, 40, 48, 47, 45, 44, 43, 41, 42, 39, 42, 45, 49, 42, 2 values under 39, and 4 values over 50.

MEDIAN is __________

HINT: Rank your values and then find the middle position.

44 (21 is the total frequency; therefore, the median value occupies the 11th position of the ranked values.)

61 When all specific values of a distribution are not known because there is an open-ended interval, for example . . .

2
3
4
4
5
6
6 and two values > 6

can the mean be computed; check and complete the answer below:

_______yes, the mean value is _______

_______no, because ____

_______don't know

√ no, because to compute the mean the sum of *all* specific values must be known. (skip to Frame 63 if correct)

____yes
____don't know

go to next frame if you checked either of these

---

62 The mean is found by dividing the sum of the values by the number (total frequency) of values. In the example above we know the total frequency (9) but we cannot determine the sum of values because we do not know the two values in the open-ended interval.

No Answer Needed

---

63 A comparison of the mean and median indicates the following characteristics (circle the correct answer) :

1. Extremely high or low (atypical) values in a distribution will unrealistically (and impractically) bias the mean / median in their direction.
2. When a distribution contains an open-ended interval, only the mean / median can be computed.
3. With respect to atypical values and open-ended intervals, the mode is most like the mean / median .

1. mean
2. median
3. median

64 Is it enough to say that all that is needed to compute the mean, median, or mode is the total number of values in a distribution and a listing of each value that occurs in that distribution?

_______ yes
_______ no
_______ don't know

√ no (skip to Frame 66 if correct)

____ yes
____ don't know
go to next frame if you checked either of these

---

65 No, we had hoped you would know that such an unqualified statement is not valid. Although the median can sometimes be computed when all values are not known, the mean requires that all values and their frequencies be reported so that we can compute the sum of all values and the total number of values.

No Answer Needed

66 We have already touched on how the completeness of your data affects the computability of the Measure of Central Tendency. More explicitly, to compute the mean we use a list of *all* the individual values or a table in which *all* the values are listed according to their frequency of occurrence—no open-ended intervals are permissible since we need to find the sum of *all* values and the total number of values. Can the mean number of visits be computed for the following examples:

_______ 1. Thirty patients visited a clinic 2, 3, 4, 5, 6, 8, 10, or 20 times apiece during its first quarter of operation.

_______ 2. During its first quarter of operation a clinic had patients make the following number of repeated visits: 2, 4, 8, 10, 6, 5, 4, 3, 2, 4, 3, 2, 6, 8, 8, 10, 20, 20, 2, 3, 4, 6, 5, 3, 4, 5, 4, 3, 2, 8.

no 1.
yes 2.

If you missed any or are unsure, go to next frame; otherwise skip to Frame 68.

---

67 We have said that to compute the mean we use a list of *all the individual values* . . . . In example one above we have the total number of cases (30) given, but only the *types* of values indicated. What is needed is how often 2 visits were made, how often 3 visits were made, etc.

In example two above we are given a list with each individual value listed as often as it occurs. With this information we can find the sum of the visits and the number of cases involved.

No Answer Needed

---

68 To compute the MEAN we use a list of *all* the ______________________

______________ or a ______________ in which the values are listed according to their ______________ of occurrence—*no open-ended intervals are permissible.*

individual values
table
frequency

69 Can the mean be computed for any of the examples to follow:

_______ 1. Distribution of ages of patients at a special clinic during January: 46, 40, 40, 48, 47, 45, 44, 43, 41, 42, 39, 42, 45, 49, 42, 2 ages under 39, and 4 ages over 50.

_______ 2.

WORKTABLE: Distribution Of 175 Preschool Children By Number Of Immunizations, Center County, 1963

Number of Immunizations	Number of Children
1	15
3	23
4	40
6	38
7	30
8	29
Total	N* = 175

*N is a symbol used for *N*umber (total frequency)

_______ 3.

WORKTABLE: Distribution Of Cases Of Poliomyelitis By Age In Years, Center County, January-June, 1962

Age in Years	Cases of Poliomyelitis
Under 4	25
5	10
6	12
7	9
8	16
9	13
10	14
Total	N = 99

Continue on next page.

69 (continued)

_______ 4.

WORKTABLE: Number Of Longshoremen Covered By Medical Care Plan For 12 Months, By Age In Years On July 1, 1955, Stockton, Calif., July 1955-June 1956

Age in Years	Number of Longshoremen
16-19	5
20-24	22
25-29	47
30-34	43
35-39	55
40-44	91
45-49	78
50-54	68
55-59	34
60-64	30
65-69	16
> 69	3
Total	N = 492

_______ 5.

WORKTABLE: Distribution Of 100 One-Year-Old Babies, By Weight In Pounds, Center County Baby Clinic, 1965

Weight in Pounds	One-year-Old Babies
20-21	4
22-23	15
24-25	31
26-27	35
28-29	12
30-31	3
Total	N = 100

no 1.
yes 2.
no 3.
no 4.
yes 5.

If you missed any or are unsure, go to next frame; otherwise, go to Frame 71.

70 The mean can be computed from data in list or in tabular form–but *no open-ended intervals are permissible.*

In example one on page 34, though you can determine that 22 patients are involved, the sum of ages cannot be found when you are not given the *particular* ages less than 39 and more than 50.

In example two, the number of cases is 175, and you are given *all* individual values even though they are grouped; therefore, the sum of values can be found.

In example three, the particular values under 4 years are not given.

In example four, the particular values over 69 years are not given.

In example five, enough information is given so that you can find the midpoint values to represent all individual values–this will allow the mean to be computed.

We see that the mean could not be computed for examples 1, 3, and 4 because the data given contained ________________ intervals.

open-ended

---

71 To compute the MEAN we use ________________________________

In your own words you should have said we use *a list of individual values* or *a table* in which the values are listed according to their *frequency* of occurrence–*the data should contain no open-ended intervals.*

72 The data used to compute the median is similar to that used to compute the mean. However, does the median require that no open-ended intervals be present?

_______ no
_______ yes
_______ don't know

√ no (skip to Frame 75 if correct)
___ yes
___ don't know
If you checked either of these or are unsure of *your* answer, go to the next frame.

---

73 Recall that the median is that value occupying the "middle position" among values ranked from smallest to largest. For example, consider the ranked data below:

3 values $< 2, 2, 2, 3, 4, 4, 4, 5, 6, 6, 7, 8, 9, 4 > 9$

Although both ends of the ranked values are "open" we know that there is a total of 19 values and that the middle position is occupied by __________ .

5

---

74 We see from the last example that open-ended data does not prevent us from finding the median value ***as long as the inclusive individual frequencies of the open-ended intervals are given,*** and the median does not fall in an open-ended interval.

No Answer Needed

---

75 To compute the mean *or* median we use a list of the __________ values or a __________ in which the values are listed according to their __________ of occurrence. To compute the *mean* the data may contain __________ intervals. To compute the median the data may contain __________ intervals if inclusive frequencies of open-ended intervals are given and the median does not fall in an __________ interval.

individual
table
frequency
no open-ended
open-ended
open-ended

76 In addition to open-ended intervals, data may also contain *missing values*, for example:

2
3
4
5
6
and 2 unknown values

Can the mean be computed in the above example–check and complete the answer below:

_______ yes, the mean value is _______
_______ no, because ____
____
____
_______ don't know

√ no, because to compute the mean the sum of *all* specific values must be known. *(skip to Frame 78 if correct)*

_____ yes
_____ don't know
} go to next frame if you checked either of these.

---

77 The mean is found by dividing the sum of the values by the number (total frequency) of values. In the example above we know the total frequency (7) but we cannot determine the sum of values because we do not know the missing (unknown) values.

No Answer Needed

78 **When** the data contains *missing values,* for example . . .

2
3
4
5
6
and 2 unknown values

can the median be computed; check and complete the answer below:

_______ yes, the median value is
_______ no, because ____
____
____
_______ don't know

√ no, because to compute the median the values must be ranked from smallest to largest and we cannot rank values which are unknown. *(skip to Frame 80 if all your answers were correct)*

_____ yes
_____ don't know
go to next frame if you checked either of these.

---

79 **The m**edian cannot be computed when the data contains missing values, even if we know **the fr**equency. In the above example we do not know what position the unknown values **would** occupy in the ranking; therefore, we cannot find the middle position, or median.

No Answer Needed

80 Place the appropriate letter beside the descriptions of data below to show that it can be used to compute . . .

a. the mean
b. the median
c. either
d. neither

______ 1. Distribution of ages of patients at a special clinic during January: 46, 40, 40, 48, 47, 45, 44, 43, 41, 42, 39, 42, 45, 49, 42, 2 ages under 39, and 4 ages over 50.

______ 2.

WORKTABLE: Distribution Of Intensive Care Patients By Age In Years, General Hospital, 1960

Age in Years	Number of Patients
11-20	5
21-30	9
31-40	20
41-50	41
51-60	39
$>$ 60	143
Total	N = 257

______ 3.

WORKTABLE: Distribution Of 175 Preschool Children By Number Of Immunizations, Center County, 1963

Number of Immunizations	Number of Children
1	15
3	23
4	40
6	38
7	30
8	29
Total	N = 175

Continued on next page.

80 (continued)

_______ 4.

WORKTABLE: Distribution Of Cases Of Poliomyelitis By Age In Years, Center County, January-June, 1962

Age in Years	Cases of Poliomyelitis
Under 4	25
5	10
6	12
7	9
8	16
9	13
10	14
Total	N = 99

Continued on next page.

80 (continued)

_______ 5.

WORKTABLE: Number of Longshoremen Covered By Medical Care Plan for 12 Months, By Age In Years On July 1, 1955, Stockton, Calif., July 1955-June 1956

Age in Years	Number of Longshoremen
16-19	5
20-24	22
25-29	47
30-34	43
35-39	55
40-44	91
45-49	78
50-54	68
55-59	34
60-64	30
65-69	16
Unknown	3
Total	N = 492

_______ 6.

WORKTABLE: Distribution Of 100 One-Year-Old Babies, By Weight In Pounds, Center County Baby Clinic, 1965

Weight in Pounds	One-Year-Old Babies
20-21	4
22-23	15
24-25	31
26-27	35
28-29	12
30-31	3
Total	N = 100

Continued on next page.

80 **(continued)**

________7.

WORKTABLE: Distribution of Cases of Diabetes By Age In Years, Center County, 1965

Age in Years	Number of Cases
5-24	15
25-44	44
45-54	61
55-64	79
65-94	97
> 94	Unknown

b 1
d 2
c 3
b 4
d 5
c 6
d 7

If you missed any or are unsure, go to next frame; otherwise go to Frame 82.

81 In example one, because the data is open at each end, you cannot compute the mean; however, because the frequencies of the open-ended intervals are given, we will be able to find the median.

In example two, the number of patients > 60 years old is more than half of the total number of patients; therefore, the median would fall within the open-ended interval and neither mean nor median can be computed.

In example three, all values and their frequency are given; therefore, either mean or median can be computed

In example four, the open-ended interval (under 4) allows for the computation of only the median.

In example five, the missing values (Unknown) allow for the computation of neither the median nor the mean.

In example six, all values and their frequency are given; therefore, either mean or median can be computed.

In example seven, the frequency for the open-ended interval (> 94) is unknown; therefore, neither the mean nor the median can be computed.

We can see from the above examples that the **mean / median** can be computed anytime the **mean / median** can, but not the reverse.

median

mean

82 What is the similarity and the difference in data that can be used to compute the mean and the median? ______

In your own words you should have said that *both* can be computed from a list of the *individual values* or a *table* in which values are listed with their particular *frequency*. In computing the mean *open-ended intervals* are *not* permissible; in computing the median *open-ended intervals are* permissible if inclusive *frequencies are provided* for the open-ended intervals, and the median does not fall in an open-ended interval.

---

83 From memory, define Measure of Central Tendency (think your answer):

In your own words you should have said that it is a value which *represents* the *center* of a distribution and can be used to *represent* the *individual values* of the distribution.

---

84 From memory, define mean (think your answer):

In your own words you should have said that the arithmetic mean is a *Measure of Central Tendency* obtained by *adding* all the *individual values* and *dividing* by the *number of values.*

---

85 From memory, define median (think your answer):

In your own words you should have said that the median is a *Measure of Central Tendency* which, for a distribution of values *ranked* from *smallest to largest,* is *above* one half and *below* the other half of the values.

86 From memory, define mode (think your answer):

In your own words you should have said that the mode is a *Measure of Central Tendency* which, for any list of values, is that *single value* or *group of values* which *occurs most often*.

---

87 On which of the Measures of Central Tendency will extreme, atypical values have the most undesirable effect? ______

What is that effect? (Think your answer.)

mean

The *mean* will be biased unrealistically toward extreme atypical values so that it will not be relevant operationally.

---

88 Which Measure of Central Tendency cannot be computed from data having "open-ended intervals?" ______

mean

MEETING THE CONDITIONS FOR USE

89 If the median is used as the reporting statistic in a major reference paper on family income and health, the Measure of Central Tendency to use for any subsequent related reporting should be the ____________________.

median

---

90 The particular Measure of Central Tendency you use should be the one ***generally accepted*** for the data or the situation to which it is applied. This is true in order to ensure ____________________.

comparability
(or comparable answers)

MEAN

91 The mean should *always* be the Measure of Central Tendency of choice when not prohibited by the characteristics of the data or the eventual use of the statistics. List the three conditions which prohibit the use of the mean:

1. ______________________________
2. ______________________________
3. ______________________________

1. Another Measure of Central Tendency is the one generally accepted for the data or the situation involved.
2. The data contain extremely large or extremely small atypical values.
3. The data contain open-ended intervals.

Skip to Frame 95 if all your answers were correct.

---

92 In any particular situation the mean may be the Measure of Central Tendency of choice except that it is not the one ______________ for the situation or the data; in this instance we might have to forego certain particular preferences in order to achieve ______________ in our Measure of Central Tendency.

generally accepted

comparability

---

93 In any particular situation the mean may be the Measure of Central Tendency of choice except that the data contains ______________ large or ______________ small atypical values. In this instance the mean would not be used because it would be biased unrealistically toward the ______________ values.

extremely large

extremely small

atypical values

94 In any particular situation the mean may be the Measure of Central Tendency of choice except that the data contain ______________ intervals. This condition would prevent you from being able to compute the sum of ______________ needed.

open-ended
values

---

95 Unless prohibited by the characteristics of the ______________ or the eventual ______________ of the statistics, the ______________ is always the Measure of Central Tendency of choice.

data
use
mean

MEDIAN

96 The first condition you should consider when you are evaluating the merits of using the median is the same as that for any other Measure of Central Tendency; namely,

1. ______________________________

____

The other two conditions that *favor* the use of the median are those that *prohibit* the use of the mean; namely,

2. ______________________________

____

3. ______________________________

____

1. It is the Measure of Central Tendency generally accepted for the data or the situation involved.
2. The data contain extremely large or extremely small atypical values.
3. The median is used when the data contain open-ended intervals.

Skip to Frame 98 if all your answers were correct.

97

1. The median is used when it is the Measure of Central Tendency generally accepted for the data or the situation involved.
2. The median is used when the data contain extremely large or extremely small atypical values.
3. The median is used when the data contain open-ended intervals.

The median is probably used most often when the first choice Measure of Central Tendency cannot be used. Therefore, if you remember the specific conditions that prohibit the use of the ________ you will know when to use the median.

mean

MODE

98 We cannot compute either the mean or the median when the data contain *missing values*, as in this list of ages:

2
3
3
3
4
and 1 unknown value.

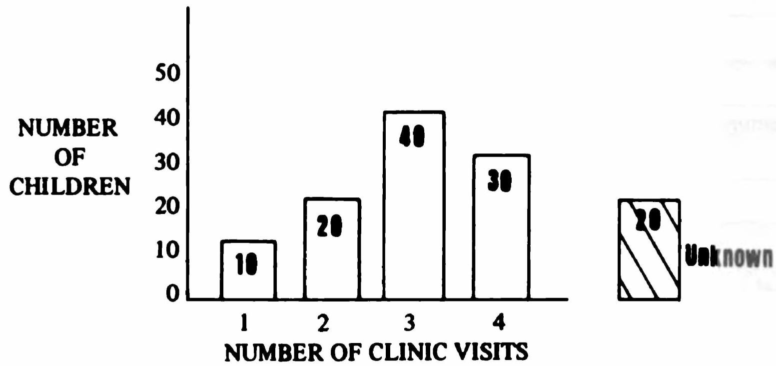
However, can we compute the mode for the above example?

_______ yes, the mode value is _______

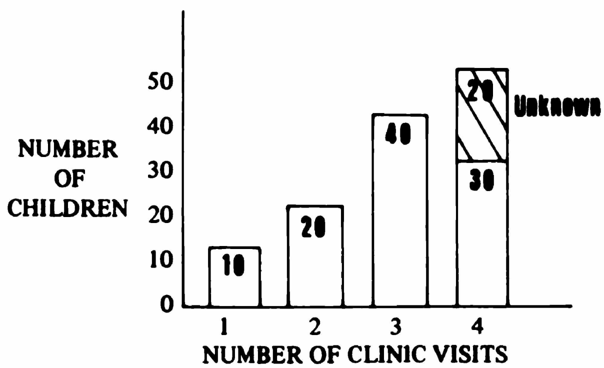
_______ no, because ____
____

__√__ yes, the mode value is __3__

99 We *can* compute the mode in the preceding frame because we can see that 3 is the most frequently occurring value. However, sometimes missing values prevent the use of the mode. Look at the following example:



Here it is possible that the 20 "unknown" values could all be the same value and that this value could be 4 visits. Since we cannot prove that this is not the case, we add the frequency of missing values to the second highest frequency of a known value:



The possible frequency of 4 visits is now: ________ 50

REMEMBER: When the data contain missing values, find the sum of the frequency of missing values and the second highest frequency of a known value. If this sum is greater than the highest frequency of a known value, the mode cannot be computed.

The new frequency of 50 is greater than the frequency of 40 for the known value which occurs most often. Therefore, we **can** / **cannot** compute the mode. cannot

100 Although the mode can be used with atypical values, open-ended intervals, and many cases of missing values, it is still not generally used unless:

1. ______________________________

____

____

(Hint: This is the one common for all.)

2. ______________________________

____

____

(Hint: This condition has to do with the definition of the mode.)

1. It is the Measure of Central Tendency generally accepted for the data or the situation involved.

2. Our interest is in the most frequently occurring value.

Skip to Frame 102 if all your answers were correct.

---

101 First, the mode is used when it is the Measure of Central Tendency generally accepted for the data or the situation involved. Second, the mode is not necessarily prohibited if the data have atypical or missing values or open-ended intervals–it can be used. However, the mode is still not usually used unless the investigator has a particular interest in the most ______________ ______________ value(s).

frequently occurring

102 Write the name(s) of the particular Measure of Central Tendency to which each of the conditions below applies:

_______ 1. Always the Measure of Central Tendency of choice unless prohibited by specific circumstances.

_______ 2. The Measure of Central Tendency generally accepted for the data or situation involved.

_______ 3. The data contain extremely large or extremely small atypical values.

_______ 4. The data contain open-ended intervals.

_______ 5. The interest of the investigator is in the most frequently occurring values.

_______ 6. Conditions 3 or 4 and the decision to use the Measure of Central Tendency of "second choice."

mean 1.
mean, median, mode 2.
median, mode 3.
median, mode 4.
mode 5.
median 6.

---

103 From memory, recall the conditions for using the mean: (Think your answer.)

The mean is always the Measure of Central Tendency of choice unless another Measure of Central Tendency is the one generally accepted for use for the data or situation involved; unless the data contain extremely large or extremely small atypical values; or unless the data contain open-ended intervals.

---

104 From memory, recall the conditions for using the median: (Think your answer.)

It is the Measure of Central Tendency generally accepted for use for the data situation involved; the data contain extremely large or extremely small atypical values; or the data contain open-ended intervals.

105 From memory, recall the conditions for using the mode: (Think your answer.)

It is the Measure of Central Tendency generally accepted for use for the data or situation involved, or the interest of the investigator is in the most frequently occurring value.

USING THE MEASURES OF CENTRAL TENDENCY

106 From our earlier discussion you can see that "representativeness" and "comparability" are important characteristics when considering a Measure of Central Tendency. To answer the question, "How can a mean or median be used?", we simply expand the two characteristics as follows:

A mean or median can be used to ______________________ all values of its distribution, they can be ______________________ to the mean or median for other distributions, and/or they can be used as a "normal" value against which individual values of their distribution can be ______________________ .

represent
compared
compared

---

107 A sanitarian reporting on his activities for the past 6 months says that he "averaged" 3.5 inspections per day. How is he using the Measure of Central Tendency? ______________________

He is using the Measure of Central Tendency (mean) to represent all values (visits per day).

---

108 The median income of Solka City West Clinic patients is $3300 per year; the median income of Solka City East Clinic patients is $5500 per year.

How is the Measure of Central Tendency used in this instance? ______________________

The Measure of Central Tendency (median) of one distribution of values (income) is being compared to the Measure of Central Tendency (median) of another distribution of values (income).

---

109 All mothers having delivery complications had less than the mean number of clinic visits during their first six months of pregnancy. How is the Measure of Central Tendency being used in this example? ______________________

The Measure of Central Tendency (mean) is being used as a "normal" value (number of visits) against which individual values (visits of problem delivery) are compared.

110 *Both* the mean and median can be used to *represent* all values of their distribution: *both* mean and median can be *compared* to means or medians of other distributions; and *both* mean and median can be used as a "normal" value against which individual values of their distribution can be *compared.*

However, because *it can also be used in further statistical computation and applications,*

the **mean / median** is the Measure of Central Tendency of choice.

mean

111 Though it will not be covered further in this *Lesson,* an example of a more complex statistic requiring the use of the mean is the "standard deviation"–a commonly used measure of dispersion within a distribution.

No Answer Needed

112 From memory recall the uses of the mean and median: ______

In your own words you should have written that *both* the mean and median can be used to *represent* all values of their distribution; *both* mean and median can be *compared* to means or medians of other distributions; and *both* mean and median can be used as a "normal" value against which individual values of their distribution can be *compared.* In addition the *mean* can be used in further statistical computation and application.

113 Use the worktable below to answer the questions:

WORKTABLE: Hypothetical Distribution Of 74 Restaurants By Number Of Inspections During One Year.

Number of Inspections	Number of Restaurants
1-2	6
3-4	12
5-6	22
7-8	19
9-10	11
11-12	4
Total	N = 74

1. Which Measure(s) of Central Tendency *could* be used on the above data:

____

2. Which Measure(s) of Central Tendency *should* be used on the above data if not otherwise prohibited? _______________

1. mean, median, mode
2. mean

114 Use the worktable below to answer the questions:

WORKTABLE: Distribution Of Paralytic Polio Cases, By Age, Texas, January-October, 1962

Age in Years	Number of Cases
0-4	77
5-9	22
10-14	8
15-19	5
20-29	4
30-39	3
40-49	5
Total	N = 124

1. Which Measure(s) of Central Tendency *could* be used on the above data?

____

2. Which Measure(s) of Central Tendency should *not* be used in view of the characteristics of the data? _______________

Why? _______________

____

____

1. mean, median, mode
2. mean

there are extreme atypical values (older cases) that would bias the mean unrealistically high

115 Use the worktable below to answer the questions:

WORKTABLE: Distribution Of 185 Boys By Weight.

Weight In Pounds	Number of Boys
60-64	3
65-69	23
70-74	50
75-79	51
80-84	31
85-89	11
90-94	4
95-99	3
* $\geq$ 100	4
Total	N = 180

* $\geq$ is a symbol meaning "greater than or equal to."

1. Which Measure(s) of Central Tendency *could* be used on the above data?

____

2. Which Measure(s) of Central Tendency could *not* be used on the above data?

____

Why? ______________________________

____

____

1. median, mode
2. mean

all individual values are not given (open-ended interval) and therefore, the sum of all values needed to compute the mean cannot be determined.

116 A health administrator planning for peak periods of activity would be interested in which of the three types of Measures of Central Tendency?

_______________ Why? _______________

mode because it will indicate the value or group of values that occurs most frequently and would therefore involve the great "peak activity"

117 An investigator wishes to compare income in a particular state with that of other states. How will his methods of reporting "average" income be affected by the way other states report their average income? _______________

He must use the same Measure of Central Tendency that is the generally accepted one for reporting income in order to insure comparability.

118 An investigator in the incidence of VD will be doing exhaustive statistical analysis on his data. Which is the Measure of Central Tendency he will use?

_______________ Why? _______________

mean because it is the one of choice especially when additional statistical computation and application will be done

119 From memory, recall the data used to compute the mean or the median: (Think your answer.)

A list of the individual values or a table in which the values are listed according to their frequency of occurrence; the mean *cannot* be computed when data contain open-ended intervals; the median can.

120 From memory, recall how the mean and the median can be used: (Think your answer.)

They can be used to represent all values of a distribution, to compare with the same type of Measure of Central Tendency of other distributions, and to compare as a "normal" value against individual values of a distribution; the mean can also be used in further statistical computations and applications.

---

121 What are the conditions for using the mean, median, or mode? (Think your answer.)

1. The mean is the Measure of Central Tendency of choice unless prohibited by the conditions to follow.
2. Use the Measure of Central Tendency *generally accepted* for the situation.
3. Use the median or mode if the data contain extremely large or small atypical values.
4. Use the median or mode if some of the values of the data are contained in open-ended intervals.
5. Use the mode if the interest of the investigator is in the most frequently occurring value.

RECOGNIZING CHARACTERISTICS OF THE DATA

122 As we mentioned at the beginning of this *Lesson,* detailed techniques of computation are not covered. Rather, they are included in *Guides* which you will use whenever the need arises. However, you should learn how to recognize certain characteristics of the data that will affect the selection of the correct computational techniques.

No Answer Needed

DATA WITH N OF < 50 or ≥ 50

123 **The technique** you will use to compute a Measure of Central Tendency will often depend **on the *N*umber** of values involved. Identify the random lists of values below as having an ***N* (count** of individual values or total of frequencies) of . . .

a. less than 50 (< 50)

b. greater than or equal to 50 (≥ 50)

_______ 1. A random list of values: 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6.

_______ 2. A random list of values: 5, 5, 10, 10, 15, 15, 20, 20, 20, 25, 25, 25, 30, 30, 30, 35, 35, 40, 40, 45, 45, 45, 45, 45, 50, 50, 50, 50, 55, 55, 55, 55, 55, 55, 55, 60, 60, 60, 60, 60, 65, 65, 70, 75, 80.

_______ 3. A random list of values: 1, 2, 3, 3, 7, 11, 11, 11, 18, 20, 24, 24, 29, 30, 34, 37, 42, 43, 48, 50, 50, 55.

_______ 4.

WORKTABLE: A Random List of Values.

Values	Frequency of Value
2	6
4	18
6	24
8	10
10	8

_______ 5.

WORKTABLE: A Random List Of Values.

Values	Frequency of Value
10-19	14
20-29	23
30-39	20
40-49	15
50-59	4

b 1.
a 2.
a 3.
b 4.
b 5.

Go to next frame if you missed any or were unsure of your answers; otherwise skip to Frame 128.

124 In any list of values, the *N*umber of values must *not* be confused with the extent to which different figures representing size or quantity occur. In random list 1. in Frame 123 the N of 53 values is represented by only ____________ different figures; in worktable 4. the N of 66 values is represented by only ____________ different figures. six five

---

125 The size of the figures representing the values of a list in no way indicates the N of values in that list. In list No. 2. of Frame 123, there is an N of < 50 values (45 to be exact) and yet the sizes of the values range up to ____________ (well above 50); in No. 4., the worktable, there is an N of ≥ 50 (66 values) and yet the highest value is only ____________ (well below 50). 80 10

---

126 The difference between the lowest value of a list and the highest in no way indicates the *N* of the list. In list No. 3. of Frame 123, there is an N of only 22 values (< 50) and yet the range represented is **above / below** 50; in No. 4., the worktable, there is an N of 66 values (≥ 50) and yet the range represented is **above / below** 50. above below

---

127 To find the N for values listed in worktables you need only add the figures listed in the frequency column. The fourth and fifth examples (Nos. 4. and 5.) in Frame 123 illustrate the fact that a worktable usually will have an **N < 50 / N ≥ 50** . N ≥ 50

DATA WITH DISCRETE OR CONTINUOUS VALUES

128 The number of inspections made by 100 sanitarians during a particular week ranged from 1 through 5 each. Therefore, each sanitarian made either: _______, ______, ______, ______ or ______ inspections.

1, 2, 3, 4, or 5

---

129 For statistical purposes, a sanitarian either makes an inspection or he doesn't. A sanitarian cannot make .5 or 2.75 actual inspections. Therefore, the values representing the number of inspections made by each sanitarian are said to be . . .

_______ a. discrete values (indivisible units or counts).

_______ b. continuous values (measurable as portions or fractions).

√ a. discrete

---

130 The humidity readings for 75 hospital nurseries ranged from .46 through .54 each. Therefore, can we say that each hospital had a humidity reading of either .46, .47, .48, .49, .50, .51, .52, .53, or .54?

_______ yes

_______ no

√ no

---

131 All values are *not* discrete, i.e., indivisible units or counts. Sometimes values are units or counts that are at best very close approximations. For example, between humidity readings of .46 and .47 there may be actual readings of .461, .462, .463, . . . and .469; or between .461 and .462, there may be actual readings of .4611, .4612, .4613 . . . and .4619. Therefore, the values representing humidity readings are said to be . . .

_______ a. discrete values (indivisible units or counts).

_______ b. continuous values (measurable as portions or fractions) depending on the accuracy of the gauge.

√ b. continuous

132 DISCRETE values are indivisible units or counts that either happen or do not happen. They are usually counted not measured.

CONTINUOUS values are divisible units or counts that are stated in that form (fractional or whole numbers) which can be most accurately approximated (measured) and most conveniently used.

Identify the values named below as either . . .

a. discrete
b. continuous

_______ 1. inspection
_______ 2. millimeters of blood pressure
_______ 3. age
_______ 4. weight
_______ 5. person
_______ 6. height
_______ 7. pregnancy
_______ 8. illness
_______ 9. innoculation

a 1.
b 2.
b 3.
b 4.
a 5.
b 6.
a 7.
a 8.
a 9.

DATA WITH VALUE RANGE OF > 14 or < 15

133 The technique you will use to compute a Measure of Central Tendency will often depend on the difference (range) between the highest and lowest values of your list. Identify the random lists of values below as having a range of . . .

a. greater than 14 (>14)
b. less than 15 (<15)

_______ 1. A random list of discrete values: 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6.

_______ 2. A random list of discrete values: 5, 5, 5, 5, 10, 10, 10, 10, 15, 15, 15, 20, 20, 20, 20, 20, 25, 25, 25, 25, 25, 25, 30, 30, 30, 30, 35, 35, 35, 40, 40, 40, 45, 45, 45, 45, 45, 45, 45, 45, 50, 50, 50, 50, 50, 50, 50, 50, 55, 55, 55, 55, 55, 55, 60, 60, 60, 60, 60, 60, 65, 65, 65, 75, 75, 80, 90.

_______ 3.

WORKTABLE: A Random List Of Discrete Values.

Value	Frequency of Value
2	6
4	18
6	24
8	10
10	8
Total	66

_______ 4.

WORKTABLE: A Random List Of Discrete Values.

Value	Frequency of Value
10-19	14
20-29	23
30-39	20
40-49	15
50-59	4
Total	126

b 1.
a 2.
b 3.
a 4.

Go to next frame if you missed any or were unsure of your answers; otherwise, skip to Frame 136.

134 If you are on this frame because the simplicity of the last one made you suspicious, let's look at each example in turn to see how really simple the frame was.

1. The high value is 6, the low value 1, the difference is ____________ and therefore, > 14 / < 15 .

5 ; <15

---

135 In the list:

2. High value ______ minus low value ______ = ______; > 14 / < 15 .
3. ______ minus ______ = ______; > 14 / < 15 .
4. ______ minus ______ = ______; > 14 / < 15 .

2. 90 - 5 = 85 ; >14
3. 10 - 2 = 8 ; <15
4. 59 - 10 = 49 ; >14

---

136 The four examples in Frame 133 support the fact that the question of > 14 or < 15 arises usually (if not always) when there is an N of < 50 / ≥ 50 .

≥ 50

---

137 Does further inspection of the four examples in Frame 133 support the fact that the list of values are in fact discrete or are *treated* as discrete values?

______ yes
______ no

√ yes

138 Actually the purpose of the >14 or <15 determination is to estimate the approximate number of *different* value sizes (figures) the data contains. This is particularly useful when you have an extremely large N.

However, you saw in our discussion of discrete and continuous values, that only with discrete values can such a determination be made. Therefore, we must employ certain unusual devices and know certain things about our values to make the >14 or <15 determination work with continuous values.

If we knew, for 75 humidity readings ranging from .46 through .54, that all were reported to the nearest hundredth, then how many possible readings are there?__________ 9

---

139 You could have gotten the right answer to the last question by counting the values that could occur (in the hundredth) on your fingers; or you could have done as follows:

1. *Ignoring* the *decimal places*, subtract the lowest value from the highest

$$\begin{array}{r} 54 \\ -\ 46 \\ \hline 8 \end{array}$$

2. Add 1

$$\begin{array}{r} 8 \\ +\ 1 \\ \hline 9 \end{array}$$

(which is how many different values may occur)

What if you know (can observe) that many, if not most, of the 75 humidity readings are reported to the nearest thousandth – would you then have > 14 or < 15 ?________ > 14

How many different values would be possible?__________ 81

---

140 The precision with which values are stated is usually obvious to you at the onset so that you would know in the last problem if the lowest reading should be stated as .46 or .460. If the latter then:

540	(ignore decimal places)
- 460	(ignore decimal places)
80	(this is > 14)
+1	
81	(which is how many different values may occur)

No Answer Needed

141 Check the appropriate descriptions below that best describe the data that follows:

_______ a. $N<50$
_______ b. $N \geq 50$
_______ c. discrete values
_______ d. continuous values

The following is a list 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.

√ a.
√ c.

---

142 Check the appropriate descriptions below that best describe the data that follows:

_______ a. $N<50$
_______ b. $N \geq 50$
_______ c. discrete values
_______ d. continuous values
_______ e. range >14
_______ f. range <15

WORKTABLE: Distribution Of Well-Child Clinics, By Number Of Children Attending, Jones County, Year Ending June 30, 1960.

Number of Children	Number of Clinics
10-14	6
15-19	9
20-24	11
25-29	8
30-34	14
35-39	8
40-44	3
45-49	1

√ b.
√ c.
√ e.

143 Check the appropriate descriptions below that best describe the data that follows:

_______ a. N<50
_______ b. N≥50
_______ c. discrete values
_______ d. continuous values
_______ e. range >14
_______ f. range <15

WORKTABLE: Distribution Of 2-Year-Old Children Attending Well-Child Clinics, By Height In Inches, Jones County, April-June, 1960.

Height in Inches	Number of Children
32	1
33	4
34	7
35	9
36	13
37	9
38	7
39	6
40	2
41	1
42	1

√ b.
√ d.
√ f.

144 Check the appropriate descriptions below that best describe the data that follows:

_______ a. N<50
_______ b. N≥50
_______ c. discrete values
_______ d. continuous values
_______ e. range > 14
_______ f. range < 15

Following is a list of weights to the nearest tenth of a pound 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.

√ b.
√ d.
√ e.

RESULTS OF FIELD DEMONSTRATIONS

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

Some 33 students at CDC took the pretest in a supervised group. Each student was then given a copy of the *Lesson* to complete on a take-home basis. The students met together a week later to take the posttest.

The 61 Los Angeles students worked in a formal classroom setting for three half-day sessions. A total of 4 hours classroom time was allotted each student to work on the *Lesson* after taking the pretest under supervision. If necessary, each student was allowed extra time outside class to complete the *Lesson*. A posttest was administered when all of the students had completed the *Lesson.*

There were specific differences between the two groups. Students at CDC had voluntarily participated, while the Los Angeles participants 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.

RESULTS

	Pretest	Posttest
CDC.	range = 0% – 48% median = 15%	range = 63% – 100% median = 92%
Los Angeles	range = 5% – 45% median = 18%	range = 42% - 100% median = 85%

Public Health Service Publication No. 2192