SUBCOURSE FI0921

EDITION B

BASIC STATISTICS

AIPD

READINESS/ PROFESSIONALISM THRU GROWTH

THE ARMY INSTITUTE FOR PROFESSIONAL DEVELOPMENT

ARMY CORRESPONDENCE COURSE PROGRAM

BASIC STATISTICS

Subcourse Number FI0921

Edition B
April 1991

United States Army Soldier Support Institute
Fort Jackson, South Carolina

13 Credit Hours

SUBCOURSE OVERVIEW

This subcourse covers comparative measures expressed as ratios (fractions, decimals, percentages, and index numbers); descriptive measures of central tendency (mean, median, and mode), and variability (range, and standard deviation); properties of the normal curve; the use of the Empirical rule; and the use of the normal curve table.

Terminal Learning Objective

Actions: Identify, define, and apply to problem solving situations, basic statistical concepts and indices that include ratios, measures of central tendency, measures of dispersion, and the properties of the normal curve.

Conditions: You will be given various instructional frames contained within this subcourse.

Standards: You will identify and define the concepts contained within this subcourse in accordance with the provisions of standard statistical reference material. Moreover, you must be able to correctly work computations and solve all problems presented in the instructional frames.

There are no prerequisites for this subcourse.

This subcourse reflects the doctrine which was current at the time it was prepared. In your own work situation, always refer to the latest publications.

The words "he," "him," "his," and "men," when used in this publication, represent both the masculine and feminine genders unless otherwise stated.

TABLE OF CONTENTS

PROGRAMED INSTRUCTION

The purpose of this subcourse is to introduce you to basic descriptive statistical methods and their applications. These methods can aid managers in the analysis of problems and in decision making.

Concepts, principles, and methods for working problems and learning techniques of basic statistics are presented in programed format. Graphs, drawings, and charts are included to give emphasis to the applications of statistics.

The text contains frames (units of instruction) each of which presents a concept, principle, or technique. You are asked questions or given problems to work to ensure that you understand the material. Self-examination and appraisal are means by which you may measure your progress and work toward self improvement.

Begin with Frame 1 and work through the programed text in a systematic, step-by-step manner. You have the opportunity to check your work as you proceed from frame to frame. Though not required, a hand calculator with a square root function would be helpful in working problems. The answer to a question in a frame appears on the following page. In this way, you receive instant feedback on whether you are correct or incorrect. If you answer incorrectly, return to the frame and restudy the material for greater understanding.

The frames in this programed instruction text are numbered consecutively from Frame 1 through Frame 72 as follows:

Frame	Lesson	Title
1-19	1	Comparative Measures
20-43	2	Descriptive Measures
44-72	3	The Normal Curve

LESSON 1

COMPARATIVE MEASURES

LEARNING OBJECTIVE

Actions: Identify and compute values for ratios in terms of fractions, decimals, percents, and index numbers.

Conditions: You will be given instructional problems relative to fractions, decimals, percents, and index numbers.

Standards: You must be able to correctly calculate percent change, percent of total, percent of programed, and percent deviation, and to correctly apply a price index to values from several time periods and convert these values into constant dollars.

FRAME 1.

INTRODUCTION

Comparative measures, or ratios, point out existing relationships between factors expressed numerically. They compare two numbers. For example, miles and gallons can he compared--25 miles to each gallon. For each 25 miles we travel, it will take 1 gallon of gasoline. Thus, we have a meaningful relationship, comparing related items.

The gasoline used should be directly and logically related to the miles traveled. The more miles we travel the more gasoline we will use. The ratio, 25 miles to 1 gallon, expresses this relationship exactly. Note two things. First, the two different items compared in ratio form must be logically related--miles and gallons. Second, the related items have to be able to be expressed numerically--25 miles and 1 gallon.

In summary, a ratio compares numerical items that are related.

QUESTION:

From what you have just read, what are two requirements for ratios?

a.____

b.____

ANSWER TO FRAME 1.

a. There must be a logical relationship between the items.

b. They must be able to be expressed as numbers.

FRAME 2.

TYPES OF RATIOS

Now that you know generally what a ratio is, you will learn to calculate and interpret 4 types of ratios:

a. fractions.

b. decimals.

c. percentages.

d. index numbers.

Each of these ratios will be covered in the following sections.

The first and most basic form of ratio is the FRACTION. It compares two numbers by division of one by the other. For instance, in our previous example, 25 miles to 1 gallon would be expressed as 25/1. It still means that 25 miles is divided by 1 gallon. It can also be expressed 25 to 1, 25/1, or 25 4 1. They all mean the same thing but we will work here with the fraction.

QUESTION:

If you traveled 50 miles on 2 gallons of gasoline, what would the fraction be that expressed miles to gallons?________________________________

ANSWER TO FRAME 2.

50/2 (50 miles to 2 gallons) which further reduces to 25/1.

FRAME 3.

During the first half of a fiscal year $40,000 was spent out of $100,000 budgeted for the whole year.

QUESTION:

What fraction relates the actual amount spent to the total amount budgeted?

____

FRAME 4.

The enrollment in a service school is 4,000 students.

QUESTION:

If 1,000 are officers, what fraction expresses the relationship of officers to total students?

____

FRAME 5.

Seventy-five of the 200 AWOLS at your installation last year were returned within 5 days.

QUESTION:

What fraction relates the AWOLS returned within 5 days to the total number of AWOLS?

____

ANSWER TO FRAME 3.

$\frac{\$40,000}{\$1000,000}$ Equivalent expressions are $\frac{\$4}{\$10}$, $\frac{\$2}{\$5}$ ($40,000 to $100,000)

ANSWER TO FRAME 4.

1,000/4,000 An equivalent expressions is 1/4, (1 officer to 4 students)

ANSWER TO FRAME 5.

$\frac{75}{200}$ An equivalent expression is $\frac{3}{8}$, (3 returned for each 8 who went AWOL)

FRAME 6.

Note that in all cases there is a logical relationship and it is expressed in terms of numbers--50 miles to 2 gallons, $2 to $5, 1 officer to 4 students, and 3 returned AWOLS to 8 who went AWOL.

The second form of a ratio is the decimal. A decimal is a fraction where the indicated division has been performed. For example, 1/4 would be 1 divided by 4, which is equal to .25. Note the fraction 1 to 4 is the same as the decimal .25 to 1. However, when expressing a decimal, the "to 1" is understood and left off. So the ratio would be stated as the decimal .25.

QUESTION:

How would you express the following fractions as decimals?

A. $\frac{3}{4}$

B. $\frac{15}{8}$

C. $\frac{5}{6}$

D. $\frac{4}{5}$

ANSWER TO FRAME 6.

A. .75

B. 1.875 or 1.88 (Note that we round our answer to two digits to the right of the decimal point.)

C. .83333 or .83

D. .80

A word about rounding.

As a general rule, carry rounding results just two places beyond or to the right of the decimal when you finish your computations.

If the third digit is 6, 7, 8, or 9, then raise the second digit to the next higher number.

Thus, 1.876 = 1.88

If the third digit is 5, then raise the second digit if it is odd.

Thus, 1.875 = 1.88

Thus, 1.835 = 1.84

If the third digit is 5, then drop the 5 and do not raise the second digit if it is even.

Thus, 1.865 = 1.86

If the third digit is 0, 1, 2, 3, or 4, then drop that digit and leave the second digit as it is.

Thus, 1.874 = 1.87

FRAME 7.

The third form of ratio to be discussed is the PERCENTAGE. It is similar to a decimal. To make a percentage, you would move the decimal point two places to the right by multiplying by 100. For example, the decimal relating 1,000 officer students to 4,000 total students was .25 to 1 or simply .25. The percentage in this case would be 25 percent. Note that the decimal point is simply moved two places to the right (.25.) and the "percent" symbol is added. It is still a ratio: 25% (officers) to 100% (total students). However, as with a decimal, the "to 100l is understood and left off. The symbol "%" may be used instead of writing the word percent.

QUESTION:

How would you express the following fractions and decimals as percentages?

A. $\frac{1}{2}$ =

B. $\frac{3}{4}$=

C. 6.3=

D. $\frac{5}{8}$ =

E. .856 =

F. .1 =

G. 3.46 =

ANSWER TO FRAME 7.

A. $\frac{1}{2}$ = 50%

B. $\frac{3}{4}$ = 75%

C. 6.3 = 630%

D. $\frac{5}{8}$ = 62.5%

E. .856 = 85.6%

F. .1 = 10%

G. 3.46 = 346%

Note that the decimal point was simply moved two places to the right and "%" was added.

FRAME 8.

PERCENTAGE is the relationship of a part to the whole. There are four basic variations of percentage that must be examined before this tool can be properly utilized.

a. Percent of total.

b. Percent change.

c. Percent deviation.

d. Percent of programed.

PERCENT OF TOTAL compares one item or group of items to the total number of items. For example, if the total budget of $400,000 for a service school includes $60,000 for TDY trips, the TDY trips ($60,000) would be 15% ($60,000/$400,000) X (100) of the total budget ($400,000). Note that a part of the budget (TDY trips) has been compared to the total budget. In terms of a formula it would be:

$$\text{Percent of total} = \frac{\text{Part}}{\text{Total}} \times 100$$

$$= \frac{60{,}000}{400{,}000} \times 100$$

$$\text{Percent of total} = 15\%$$

QUESTION:

If civilian pay and benefits amounts to $300,000, what percent of the total would this be of a $400,000 budget? __________

ANSWER TO FRAME 8.

$$\text{Percent of Total} = \frac{\$300{,}000}{\$400{,}000} \times 100$$

$$= 0.75 \times 100$$

Percent of Total = 75%

FRAME 9.

PERCENT CHANGE relates the amount of change between two time periods to the amount in the first time period. For example, if personnel strength was 5,500 men this year and 5,000 men last year, there was an increase of 500. Comparing the amount of change (500) to the amount in the first time period (5000), would give us 500/5000 = 101 change. There was an increase of 101 in personnel this year over last year. In terms of a formula, it would be the following:

$$\text{Percent change} = \frac{\text{Period two amount - Period one amount X 100}}{\text{Period one amount}}$$

$$= \frac{5{,}500-5{,}000}{5{,}000} \text{ X } 100 = \frac{500}{5{,}000} \text{ X } 100$$

$$= 0.10 \text{ X } 100$$

$$\text{Percent change} = 10\%$$

QUESTION:

If the price of an item was $50 last year, what is the percent change if it costs $40 this year? ____

ANSWER TO FRAME 9.

$$\text{Percent Change} = \frac{40-50}{50} \times 100 = \frac{-10}{50} \times 100$$

$$= -0.20 \times 100$$

$$\text{Percent Change} = -20\%$$

FRAME 10.

If the cost last year was $80,000, what is the percent change if this year's cost was $72,000?

FRAME 11.

PERCENT DEVIATION compares the amount by which the actual amount differed from the planned amount. Suppose the actual amount spent was $392,000 while $400,000 had been budgeted. The difference between actual and planned was -$8000. Comparing this -$8000 to the planned amount of $400,000, we get a percent deviation of -2%. In other words, the actual amount was 2% less than the planned amount. In terms of a formula, we have the following:

$$\text{Percent deviation} = \frac{\text{actual amount} - \text{planned amount}}{\text{planned amount}} \times 100$$

$$= \frac{392{,}000 - 400{,}000}{400{,}000} \times 100$$

$$= \frac{-8{,}000}{400{,}000} \times 100$$

Percent deviation a -2%

The percent deviation indicates how much the actual is above or below the planned amount in terms of percent.

QUESTION:

If the amount budgeted for travel is $60,000 and the actual amount spent is $66,000, what is the percent deviation and what does it mean?

ANSWER TO FRAME 10.

$$\text{Percent change} = \frac{-8{,}000}{80{,}000} \times 100$$

$$= -0.10 \times 100$$

Percent change - -10%

(Notice that the percent of change can be positive or negative.)

ANSWER TO FRAME 11.

$$\text{Percent deviation} = \frac{66{,}000 - 60{,}000}{60{,}000} \times 100 \cdot 10\%$$

The amount spent was 10% higher than the amount budgeted.

FRAME 12.

PERCENT OF PROGRAMED compares actual amounts or results to planned or programed amounts. For example, If $400,000 had been budgeted (planned) but only $392,000 (actual) had been spent, the percent of programed would be 98% (392,000/400,000) X (100).

$$\text{Percent of programed} = \frac{\text{actual amount}}{\text{programed amount}} \times 100$$

$$= \frac{392\ 000}{400{,}000} \times 100$$

Percent of programed = 98%

QUESTION:

If the actual amount spent for travel was $66,000, instead of the programed $60,000, what would the percent of programed be? ______________________

FRAME 13.

SUMMARY OF PERCENTAGES: We have examined four types of percentages:

a. $\text{percent of total} = \frac{\text{Part}}{\text{Total}} \times 100$

b. $\text{percent change} = \frac{\text{period two amount - period one amount}}{\text{period one amount}} \times 100$

c. $\text{percent deviation} = \frac{\text{actual amount - planned amount}}{\text{planned amount}} \times 100$

d. $\text{percent programed} = \frac{\text{actual amount}}{\text{planned amount}} \times 100$

Each of these percentages expresses a relation of a part to the whole.

ANSWER TO FRAME 12.

$$\frac{\$66{,}000}{\$60{,}000} \times 100 = 110\%$$

FRAME 14.

Percentages are a useful method for comparing various items. However, to make comparisons over several time periods, the INDEX NUMBER is often a better tool. Percent change measured change over two time periods. Index numbers are used to compare like items in each of several years to one common year which will be called a base year. In terms of a formula, we have the following:

$$\text{Index number} = \frac{\text{Value in any time period}}{\text{Value in base time period}} \times 100$$

Consider the following example involving the cost for a television set during a seven year period where 1985 is the base year.

YEAR	PRICE
1983	$372
1984	388
(base) 1985	400
1986	420
1987	424
1988	460
1989	468

Using the index number formula, the index number for 1985 is as follows:

$$\text{Index Number for 1985} = \frac{400}{400} \times 100$$

$$= 1 \times 100$$

$$= 100$$

Note that we compared the price for 1985 to the price for the base ear which also happened to be 1985. Just as with a percent we multiplied by 10O. In fact, index numbers are sometimes called relative percentages, but index numbers never include the word "percent." The index number for 1985 is simply 100, not 100 percent.

QUESTIONS:

What is the index number for 1983? _______________

What is the index number for 1987? _______________

ANSWER TO FRAME 14.

Index number for 1983 = $\frac{372}{400}$ X 100 93

Index number for 1987 = $\frac{424}{400}$ X 100 106

FRAME 15.

Note that when the price is greater than the base year price (as in 1987) the index Is over 100, and when the price Is less than the base year price (as in 1983) the index is less than 100.

QUESTION:

What are the Index numbers for 1984, 1986, 1988, and 1989?

YEAR	PRICE	INDEX
1983	$372	93
1984	388	_____
(base) 1985	400	100
1986	420	_____
1987	424	106
1988	460	_____
1989	468	_____

ANSWER TO FRAME 15.

1983	372/400 X 100	=	93
1984	388/400 X 100	=	97
1985	400/400 X 100	=	100
1986	420/400 X 100	=	105
1987	424/400 X 100	=	106
1988	460/400 X 100	=	115
1989	468/400 X 100	=	117

FRAME 16.

USES OF INDICES

Note that in every case we compared the price for each year to the base year price and multiplied by 100. Index numbers are used to compare prices for like items over time. The purpose of any ratio is to compare values. It is much easier to relate the index each year to a common base year index (100) than to compute the percentage change in price between each of the seven years. For instance, it is easier to look at the 117 for 1989 and compare it to the 100 for 1985 than to compare the price $468 in 1989 to $400 in 1985. The index immediately tells us that the price in 1989 is 17% greater than in the base year. It is important that the base year selected be a "typical" year.

How are index numbers actually used? The most common area of application is price indices. Two examples of price indices are the Simple Price Index (SPI) and the Consumer Price Index (CPI). The Simple Price Index includes just one item or commodity. If you combine Simple Price Indices, then you have an aggregate price index. An example of an aggregate index is the CPI. The CPI is a statistical measure of change in prices of goods and services bought by urban wage earners and clerical workers. There are some 400 food, housing, and transportation items in this index and they are grouped by type of commodity. These are then weighted to form the overall index. Let us look at an income index. Your income has probably increased in dollars over the past 5 years. What about its real purchasing power? Has it increased or decreased? To find out, we must convert your income for each of the 5 years into constant dollars so they can be compared. Thus,

YEAR	INCOME	CPI
1985	$10,000	100
1986	10,500	108
1987	11,500	120
1988	12,000	125
1989	13,500	150

To find the purchasing power of your 1989 income in 1985 dollars, use the following ratio:

$$\frac{\text{Base Year Value}}{\text{Base Year Index}} = \frac{\text{Current Year Value}}{\text{Current Year Index}}$$

But since the base year index is always 100, we can rewrite the ratio as follows:

$$\text{Base Year Value} = \frac{\text{Current Year Value}}{\text{Current Year Index}} \times 100$$

Substituting values from the table, we have the following:

$$\text{Base Year Value} = \frac{13{,}500}{150} \times 100 = 90 \times 100 = 9{,}000$$

1989 income is $9,000 vice 1985 income. You lost $1,000 in purchasing power.

QUESTION: What is your 1988 income in 1985 dollars? ____________

ANSWER TO FRAME 16.

$$\text{The ratio} = \frac{\text{Current Year Value}}{\text{Current Index}} \times 100$$

$$= \frac{12{,}000}{125} \times 100$$

$$= 9.60 \times 100$$

$$= \$9{,}600$$

Thus, your 1988 income is $9,600. In terms of 1985 dollars, you lost $400 in purchasing power.

FRAME 17.

Now complete the rest of the table.

YEAR	INCOME	CPI	INCOME IN 1985 COLLARS
1985	$10,000	100	______________
1986	10,500	108	______________
1987	11,500	120	______________
1988	12,000	125	$ 9,600
1989	13,500	150	$ 9,000

FRAME 18.

The expenditures for maintenance during the past 5 years and the price Indices for this time period are shown below. Compute the "real" cost of maintenance in terms of 1987 dollars for all 5 years.

YEAR	MAINTENANCE EXPENSE	PRICE INDEX
1985	350	82
1986	475	93
1987	560	100
1988	720	111
1989	840	134

ANSWER TO FRAME 17.

YEAR	INCOME IN 1985 DOLLARS
1984	$10{,}000 = \frac{10{,}000}{100} \times 100$
1985	$9{,}722 = \frac{10{,}500}{108} \times 100$
1986	$9{,}583 = \frac{11{,}500}{120} \times 100$

ANSWER TO FRAME 18.

YEAR	=	REAL COST OF MAINTENANCE
1985	=	$427 = \frac{350}{82} \times 100$
1986	=	$511 = \frac{475}{93} \times 100$
1987	=	$560 = \frac{560}{100} \times 100$
1988	=	$649 = \frac{720}{111} \times 100$
1989	=	$627 = \frac{840}{134} \times 100$

FRAME 19.

As demonstrated above, a commonly used index is a price index used to compute the real value, or value in constant dollars, of incomes and expenditures over various time periods. This allows the analyst to compare like items over time, all measured with a constant standard.

Proceed now to the Practice Exercise for Lesson 1.

LESSON 1

PRACTICE EXERCISE

PART A - PERCENTAGES

SITUATION I. You are given the following budget information. Refer to this information to answer Questions 1 through 6.

	Current Year	Next Year
Civilian Pay	$ 150,000	$ 200,000
Equipment Purchase	65,000	75,000
Maintenance	35,000	35,000
Supplies	100,000	90,000
Construction	150,000	200,000
	$ 500,000	$ 600,000

1. What percent of the Current Year's total budget is Civilian Pay?

2. What percent of Next Year's total budget is Civilian Pay?

3. What percent of the Current Year's total budget is Maintenance?

4. What is the percent change in Supplies from the Current Year's Budget to Next Year's Budget?

5. What is the percent change in Construction from the Current Year's Budget to Next Year's Budget?

6. What is the percent change in Maintenance from the Current Year's Budget to Next Year's Budget?

SITUATION II. You are given the following budget information. Refer to this information to answer Questions 7 through 10.

	Current Year Budget	Current Year Actual Expenses
Civilian Pay	$ 150,000	$ 100,000
Equipment Purchase	65,000	75,000
Maintenance	35,000	35,000
Supplies	100,000	90,000
Construction	150,000	175,000
	$ 500,000	$ 475,000

7. What is the percent deviation of the Civilian Pay expenses from the programed (budgeted) amount?

8. What is the percent deviation of the Construction expenses from the programed amount?

9. What percent of programed funds was actually used in the category of Equipment Purchase?

10. What percent of programed funds was actually used in the category of Construction?

PART B - INDEX NUMBERS

Fill In the following chart. Make sure you round to the nearest whole number.

YEAR	CURRENT DOLLAR BUDGET	INDEX	BUDGET IN 1986 CONSTANT DOLLARS
1985	$ 516	95	$__________(11)
1986	520	100	$__________(12)
1987	525	104	$__________(13)
1988	527	109	$__________(14)
1989	530	116	$__________(15)

LESSON 1

PRACTICE EXERCISE

ANSWER KEY AND FEEDBACK

PART A - PERCENTAGES

1. $\frac{150,000}{500,000} \times 100 = \frac{15}{50} \times 100 = .30 \times 100 = 30\%$ (See frame 8.)

2. $\frac{200,000}{600,000} \times 100 = \frac{2}{6} \times 100 = .33 \times 100 = 33\%$ (See frame 8.)

3. $\frac{35,000}{500,000} \times 100 = \frac{35}{500} \times 100 = .07 \times 100 = 7\%$ (See frame 8.)

4. $\frac{-10,000}{100,000} \times 100 = \frac{-10}{100} \times 100 = -10 \times 100 = -10\%$ (See frame 9.)

5. $\frac{50,000}{150,000} \times 100 = \frac{5}{15} \times 100 = .333 \times 100 = 33\%$ (See frame 9.)

6. $\frac{0}{35,000} \times 100 = 0\%$ (See frame 9.)

7. $\frac{100,000-150,000}{150,000} \times 100 = \frac{-50\ 000}{150,000} \times 100 = -.33 \times 100 = -33\%$

(See frame 11.)

8. $\frac{175,000-150,000}{150,000} \times 100 = \frac{25,000}{150,000} \times 100 = .17 \times 100 = 17\%$

(See frame 11.)

9. $\frac{75,000}{65,000} \times 100 = \frac{15}{13} \times 100 = 1.15 \times 100 = 115\%$ (See frame 12.)

10. $\frac{175,000}{150,000} \times 100 = \frac{7}{6} \times 100 = 1.17 \times 100 = 117\%$ (See frame 12.)

If you missed any of the exercises return to the indicated frame.

Part B - INDEX NUMBERS

YEAR		BUDGET IN 1986 CONSTANT DOLLARS
11.	1985	$\frac{16}{95}$ X 100 = 5.43 X 100 = 543
12.	1986	$\frac{520}{100}$ X 100 = 5.20 X 100 = 520
13.	1987	$\frac{525}{104}$ X 100.X 5.05 X 100 = 505
14.	1988	$\frac{527}{109}$ X 100 X 4.83 X 100 = 483
15.	1989	$\frac{530}{116}$ X 100 X 4.57 X 100 = 457

If you missed any part of this exercise for reasons other than arithmetic errors, return to frame 14 and rework the entire INDEX NUMBER SECTION.

LESSON 2

DESCRIPTIVE MEASURES

LEARNING OBJECTIVE

Actions: Identify and compute measures of central tendency and dispersion.

Conditions: In this lesson you will be given questions and problems that pertain to descriptive measures.

Standards: Your explanations and computations will be In accordance with the descriptions provided In this lesson.

FRAME 20.

INTRODUCTION

The advances in the size and complexity of organizations have given rise to the development of a new type of manager. With commercial Industries and the military establishment growing more complex, it becomes necessary to devise tools whereby the details and results of operations can be simplified and presented to the manager in a way that will assist him to make better decisions.

FRAME 20. (Continued)

One of the tools available to the manager is STATISTICS. Increasing numbers of executives desire the guidance of statistical analysis in making decisions. This growing use of statistical methods in business has been paralleled by a new emphasis on statistics by Government agencies. Government and business generate, manipulate, store and print incredible quantities of raw data. Statistics is a science that concerns itself with the measurable properties of a "pile of data" or group of objectives and thus can be helpful in bringing some order to the chaos.

We might separate statistics into closely related categories of quantitative techniques. These two categories are commonly called DESCRIPTIVE and INFERENTIAL statistics.

We use DESCRIPTIVE statistics to marshall and put data into some organized structure that has meaning to the manager or decision maker. We attempt to identify those measurable properties mentioned earlier, and to attach a number or value to them. We also try to identify a "shape" or form for the data, and relate that shape to some known shape that has thoroughly understood and documented characteristics.

We use INFERENTIAL statistical techniques to analyze the data and draw a conclusion or conclusions from the results of the analysis. In doing so, we, as analysts, go beyond the data collected, and predict, with some measure of accuracy, quantities to be realized.

There are terms in statistics that we will need to know.

The POPULATION is a set of data which consists of all possible observations of a certain event. Another definition might be the totality of persons, items, time periods, equipment, and so forth, with a common characteristic. Perhaps the simplest definition would be to call the population the WHOLE, whatever we define it to be.

Examples would include such disparate populations as all the people in the world who are US citizens, all the automobiles produced by General Motors last year, all the tanks currently in the Army inventory, all the members of a specified platoon, or all the people in your family.

Should we decide to look at each item in a population, we would be performing a CENSUS. Usually we can not perform a census. Some reasons which may deter us from doing a census include lack of time, prohibitive expense, destructive nature of the census, and changing characteristics of the population. In these situations, we examine a SAMPLE drawn from the population in question.

The SAMPLE is a set of data containing only a part of all the observations found in the population. It is a part or a subset of the population. More simply, a sample is just a part of the whole. We sample quite frequently. For example, when

FRAME 20. (Continued)

you went to dinner at that fancy restaurant and pronounced their food 'excellent," you were asking a statement based on a sample of only one meal about the population of all meals that had keen, or will be, prepared at that restaurant.

If we have a small population, say an infantry platoon and we are interested in average age, we could take a census. The value we compute as the average age is called a PARAMETER. In general, a parameter is a value that can computed fro a population if the entire population is available.

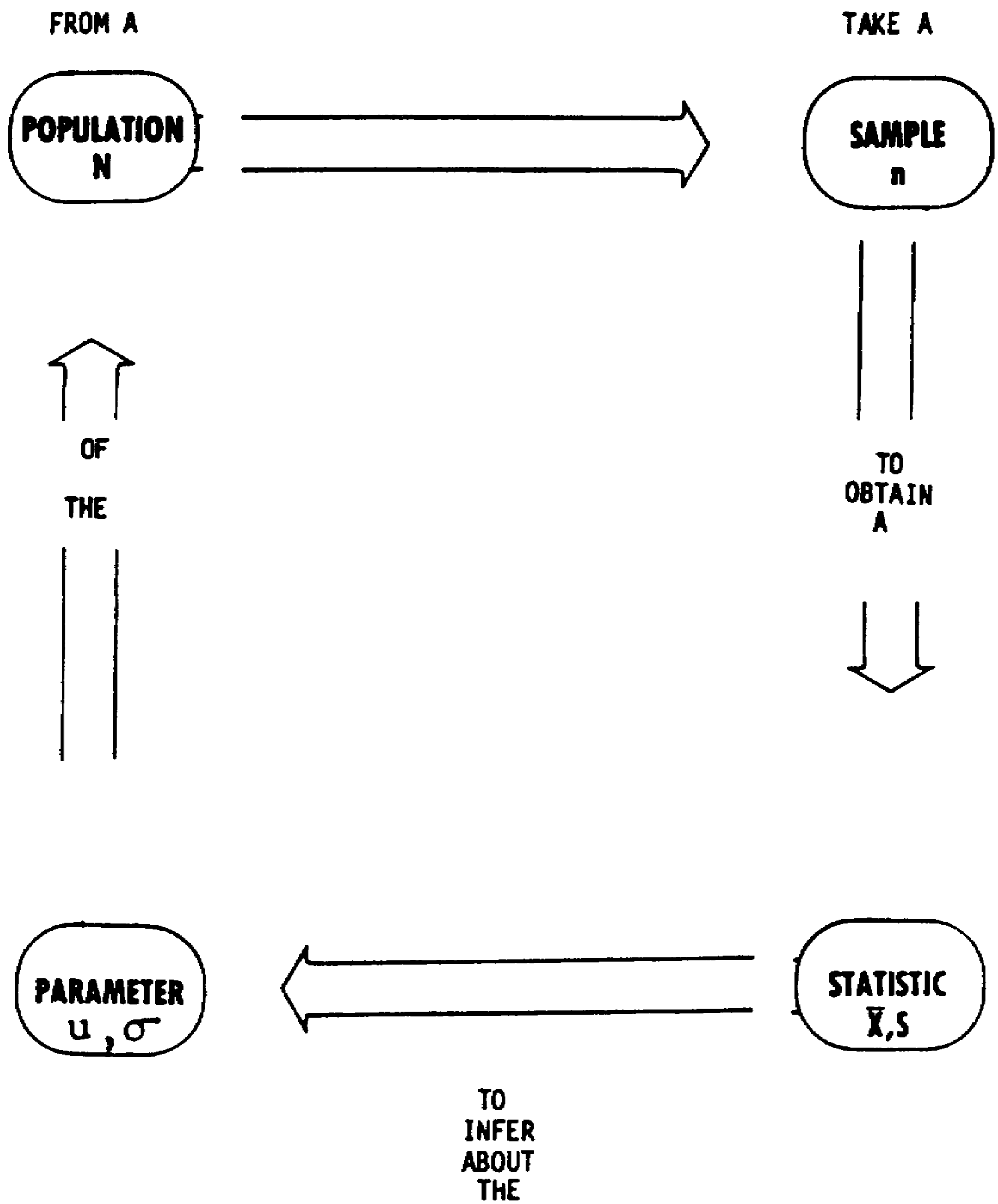
However, the entire population is often not available. So, we use a port of the population (sample) to compute an ESTIMATE of the population value. This estimate, computed from a sample, is called a STATISTIC.

QUESTION:

Now would you summarize this relationship? The__________________________________ is the whole; the ________________________ is a part of the whole.
The __________________________________ is the true value computed from the population; the __________________________________is the estimate of this value computed from the sample.

ANSWER TO FRAME 20.

Summarizing the relationship, the population is the whole; the sample a part of the whole. The parameter is the true value computed from the population; the statistic is the estimate of this value computed from the sample. Symbols such as those depicted in the diagram that follows are explained more fully in the notations found in lesson 2.



FRAME 21.

There are two characteristics of a group of data that we will concern ourselves with in this text. The first characteristic answers the question, "do the data tend to cluster or group around some particular value or values?"

This first characteristic is CENTRAL TENDENCY. The three MEASURES OF CENTRAL TENDENCY, discussed in the following pages, are used to identify the value around which the data tend to cluster.

The second characteristic answers the question, "do the observations cluster very tightly around the measure of central tendency, or, are they spread out?" The second characteristic is called DISPERSION. Two measures of dispersion will be discussed in this text.

QUESTION:

What are the two characteristics of a group of data that we concern ourselves with in this text?

a. ____________

b. ____________

ANSWER TO FRAME 21.

a. Central Tendency

b. Dispersion

FRAME 22.

MEASURES OF CENTRAL TENDENCY

The first of these characteristics that we will discuss is the tendency for data to cluster around some particular value or values.

This characteristic is often referred to as CENTRAL TENDENCY.

There are three measures of CENTRAL TENDENCY:

MEAN, MEDIAN, AND MODE

The ARITHMETIC MEAN is found by adding the values of a group of data and dividing the sum by the number of items.

The term MEAN will be used in place of arithmetic mean in this text. For example, the mean of the five test scores:

98, 100, 82, 75, 95, is

$$\text{mean score} = \frac{\text{sum of the scores}}{\text{number of scores}} = \frac{450}{5} = 90$$

QUESTION:

What is the mean of the following ages?

24, 25, 32, 28, 26? _______________

ANSWER TO FRAME 22.

$$\text{mean age} = \frac{\text{sum of the ages}}{\text{number of ages}} = \frac{135}{5} = 27$$

If we wanted one number to describe all the ages in our group of data, It would be 27.

FRAME 23.

In the example above, we indicated the total of several measurements by using the term "sum of." We use this term so frequently that we have accepted a symbol to replace the words. The symbol is the Greek letter Σ (pronounced Sigma). Thus, we could have expressed the problem as follows:

$$\text{mean score} = \frac{\Sigma\ (\text{test scores})}{\text{number of test scores}}$$

Since we do not limit ourselves to finding Just average test scores, we need to develop a more general statement for the relationship among the sum of the items we are interested in, the number of those items, and the mean value of those items. Such a general statement is called a formula. The formula is like a recipe, or set of general directions. It is composed of general symbols that we attach specific values to for specific problems. For example, the formula for the sample mean is:

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

Where,

$\overline{X}$ is the symbol for the sample mean,

X is the symbol for each observation of interest, and

n is the symbol for the number of observations.

In the example above, we are interested in 5 ages, thus X in this example represents each of the given ages. ΣX represents the sum of the five ages. For each problem we do, or situation we encounter, what X represents will be specified.

FRAME 24.

Answer the following questions below using the following five dollar amounts, $35, $27, $24, $26, $28 taken from a group of 100 dollar amounts.

a. X represents ___, ___, ___, ___, ___

b. ΣX = ___

c. n = ___

d. $\frac{\Sigma X}{n}$ = ___

ANSWER TO FRAME 24.

a. X represents 35, 27, 24, 26, 28, (the five observations of interest)

b. $\Sigma X = 140$ (this is the sum of these 5 observations)

c. n represents 5 (the number of observations)

d. $\frac{\Sigma X}{n} = \frac{140}{5} = 28$ (This is the sample mean for the 5 observations)

FRAME 25.

The preceding problem demonstrated the calculation of a sample mean. We knew it was a sample because the five values of interest were taken from a population of 100. We could have computed the mean of all 100 dollar amounts. In such a case we would have had a population mean. The difference between a sample mean and a population mean is that the population mean includes every value in the population; whereas the sample mean includes only a portion of the values in the population. The population mean is still the sum of the values of interest divided by the number of values of interest. Now we are interested in the entire population.

Even though they are computed in the same manner they are different concepts and we will need a different symbol for each. The symbol for population mean is the Greek letter u (pronounced mu). The formula for the population mean is

$$u = \frac{\Sigma X}{N}$$ where,

X still represents each value of interest. However, we are now interested in all values in the population)

N represents the number of items of interest in the population. Note this is a capital N rather than the small n used in the sample mean formula.

FRAME 26.

For example, suppose the number of calculator batteries used during all ten 3-week courses held this year is 37, 62, 28, 31, 58, 29, 35, 47, 52, and 25. We need to know the average number of batteries used in a 3-week course this year.

a. Find the mean. ___________________

b. Is this a sample or a population mean? ___________________

ANSWER TO FRAME 26.

a. mean = u = $\frac{\Sigma X}{N} = \frac{404}{10} = 40.4$

b. We defined the population to be the batteries used in each of the ten 3-week courses this year. Since we used all 10 observations, we have computed a population mean.

FRAME 27.

Suppose gathering all this data were very expensive. An alternative approach would be to take a sample from this population. Assume the observations 37, 58, 35, 54 were chosen as a sample from the population above.

a. Compute the mean. _______________

b. Is this a sample or population mean? _______________

FRAME 28.

Compute the mean for the following samples.

Sample 1: 22, 29, 22, 7, 34. _________________

Sample 2: 22, 29, 22, 7, 10,000. ________________

FRAME 29.

The second measure of central tendency is called the MEDIAN. The median is a positional average. It is determined by arranging the data in ascending or descending order and locating the middle value.

If there is an even number of observations in our data, we find the median by taking the mean of the MIDDLE two values. If there is an odd number of observations, there will always be one value in the data set that is the median.

For example, suppose we have the following nine observations: 1, 22, 29, 23, 7, 34, 14, 17, 31. We would find the median by arranging the data in ascending (or descending) order and selecting the middle value. Putting the values in ascending order gives: 1, 7, 14, 17, 22, 23, 29, 31, 34. The median or middle value in this ordered list is 22. There are four values larger than 22 and four values smaller than 22. If we arrange the data in descending order, the result would be the same. Try it.

QUESTION:

What is the median of the following eight observations:

8, 24, 28, 19, 7, 31, 2, 22 ________________

ANSWER TO FRAME 27.

a. mean = $\bar{X} = \frac{\Sigma X}{n} = \frac{184}{4} = 46.0$

b. This is a sample mean because only four observations from the population were used in computing it.

In a later text, we will use the sample mean to make a statement about the population mean. For the time being, note that the sample mean and the population mean DO NOT USUALLY GIVE THE SAME VALUE.

ANSWER TO FRAME 28.

Sample 1 mean = $\bar{X} = \frac{\Sigma X}{n} = \frac{114}{5} = 22.80$

Sample 2 mean = $\bar{X} = \frac{\Sigma X}{n} = \frac{10{,}080}{5} = 2016$

Note the effect of extreme values on the sample mean. Although only one value changed, the mean was changed significantly. This points out the key characteristic that the mean is affected by extreme values and it may not be representative of all values in the sample (or population).

ANSWER TO FRAME 29.

Ordering the data gives: 2, 7, 8, 19, 22, 24, 28, 31. Since there are an even number of observations we must find the mean of the middle two values. This would be (19 + 22) /2 = 20.5. The median is then 20.5.

Note that, if you forgot to order the data, you would have computed the median to be (19 + 7) /2 = 13. This is incorrect. The data must be ordered.

FRAME 30.

Find the median of the following two samples:

Sample A: 17, 22, 29, 23, 7, 34 _______________

Sample B: 22, 17, 10,000, 29, 7, 23 __________________

FRAME 31.

The median is a more appropriate measure of central tendency when there are extreme values in the data. For example, average Income is usually a median, rather than a mean, because of the extremely large incomes of a few individuals.

The third measure of central tendency is called the MODE. The mode is the MOST FREQUENTLY occurring observation in a group of data. For example, given the set of data: 38, 19, 27, 25, 27, 26, 19, 27, 22, we see that the observation 27 occurs 3 times. The mode for this set of data is 27. Find the mode for the following three samples:

Sample A: 22, 29, 22, 7, 34, 14, 17 _________

Sample B: 31, 2, 8. 22, 19, 28, 13 _________

Sample C: 22, 7, 29, 34, 22, 14, 31, 7 _________

ANSWER TO FRAME 30.

Ordering both samples gives:

Sample A: 7, 17, 22, 23, 29, 34

Sample B: 7, 17, 22, 23, 29, 10,000

Since there are an even number of observations we find the mean of the middle two which gives:

Sample A: (22 + 23) /2 · 22.5

Sample B: (22 + 23) /2 - 22.5

Note that the median is not affected by extreme values as the mean was.

ANSWER TO FRAME 31.

Sample A: The value 22 occurs twice and all other values occur only once, so the mode is 22.

Sample B: Each value occurs the same number of times, once. The mode is the value or values that occur most often in the data set.

Sample C: The values 7 and 22 both occur twice. Both are modes of this group of data. Since two values occur twice, the data are bi-modal (two modes).

Notice there may be none, one, or more than one mode for a group of observations. There is, however, only one mean or median for any group of data.

FRAME 32.

A typical application of the mode would be to determine the most frequent cause or reason for some event. For example, what is the most frequent cause of traffic accidents given the following data?

CAUSES	NUMBER OF ACCIDENTS
Mechanical Failure	100
Reckless Driving	260
Drunken Driving	130
Speeding	380
Weather Conditions	130
	1000

Clearly the mean or median would not be appropriate in this case because the causes cannot be numerically measured. The modal cause is speeding because more accidents are attributed to it than any other factor.

QUESTION:

The following data represents the number of hits on target by 10 tank crews selected at large from a tank battalion: 37, 43, 31, 40, 37, 45, 40, 37, 40, 38. Find the mean, median, and mode.

ANSWER TO FRAME 32.

First order the data low to high:

31, 37, 37, 37, 38, 40, 40, 40, 43, 45

The mean is $\bar{X} = \frac{\Sigma X}{n} = \frac{388}{10} = 38.8$ (This is a sample, not a population).

The median is (38 + 40) /2 * 39. (There are an even number of observations.)

The modes are 37 and 40 because both occur three times.

FRAME 33.

MEASURES OF DISPERSION

Now we will discuss two measures of dispersion. As indicated earlier, dispersion measures the spread or variability of the data about the mean value. The measures of central tendency provide significant information about a group of data. In most cases, however, more information is needed. It is the combination of measures of central tendency and measures of dispersion that provide a clearer description or summary of the data.

The following example illustrates that a measure of dispersion provides important additional information about a group of observations.

ANNUAL WAGES OF TWO GROUPS OF WORKERS

Group A		Group B
$ 1,000		$ 2,500
1,275		2,775
1,325		2,800
1,350		2,850
1,475		2,875
1,525		3,025
2,325		3,175
2,475		3,225
2,575	**Median**	3,275
3,450	◄── **and** ──►	3,450
3,825	**Mean**	3,700
4,025		3,725
4,425		3,775
4,475		3,875
4,525		3,950
4,550		4,000
4,575		4,025
6,375		4,050
10,000		4,500

QUESTION:

How do the two groups differ?

ANSWER TO FRAME 33.

Groups A and B have the same mean and median of $3,450. However, the observations in group A are spread out over a greater range of values than the observations in group B. Note that the $10,000 extreme value in Group A markedly drives up the value of the arithmetic mean. Without it, the average annual wage for Group A workers drops by ten percent.

FRAME 34.

From this example, it can be seen that to reach valid conclusions it is important to measure how data are dispersed around the measure of central tendency.

Dispersion is important, not merely as a supplement to the mean, but also as a significant item in itself. The performance of a student is judged not only by his average but also by his consistency. When a measure of central tendency and a measure of dispersion have been computed for a series, generally the two most important characteristics have been summarized.

The first measure of dispersion that we will examine is the range. The range is the difference between the highest and lowest values in a group of observations. In the example above, the range of Group A is 10,000 -1,000 9,000, and the range of Group B is 4,500 -2,500 = 2,000. This quickly illustrates that Group B's wages vary much less than Group A's wages.

FRAME 35.

Generally it is desirable to express the range in terms of the upper and lower limits; thus, we would say A's range is $1,000 to $10,000 and B's is $2,500 to $4,500. This gives the reader an idea of the general location of the data.

Although it has the merit of being simple, the range is a rather unsatisfactory measure because it is determined only by the two extreme values in a group of observations, the high and the low. Since these two figures are only "boundaries" of the rest of the data, they are insensitive to the behavior or location of figures between them. The range should be used only in cases where a quick, cursory look at the data is desired.

QUESTION:

Find the range of the following costs:

$2600
2750
2400
3800
4500
3100
2800

Range = __________________

ANSWER TO FRAME 35.

Ranges = $4,500 -$2,400 = $2,100. Note that the middle five values could have been anything between $2,400 and $4,500 and the range would be the same. The range considers only two values - the high and the low. All other values in between are not considered.

FRAME 36.

The second measure of dispersion is the standard deviation. The standard deviation is a much better measure of dispersion than the range because it considers all observations, rather than just the two extremes. It measures an "average" dispersion around the mean.

As an example, let's find the standard deviation of the following seven observations: (15, 20, 25, 30, 35, 40, 45).

The first step is to compute the mean. In this case the mean is

$$u = \frac{\Sigma X}{N} = \frac{210}{7} = 30.$$

One approach to measuring dispersion is to measure the distance between each observation and the mean. The formula for this relationship is X - u where X is an observation and u is the mean. Consider the following table:

X	X - u	
15	15-30=-15	Notice that if an observation is
20	20-30=-10	smaller than the mean, X - u is
25	25-30=-5	negative.
30	30-30= 0	
35	35-30= 5	
40	40-30= 10	
45	45-30= 15	
	Σ (X-u) = 0	

In this example, Σ(X-u) equals zero. This is always true. Consequently, Σ(X-u) is useless in helping us compute a measure of dispersion. The combination of positive and negative values always offset each other. To avoid this problem we will square the result of each X-u. (To square a number means to multiply it by itself.)

FRAME 37.

For example:

X	X-u	$(X-u)^2$	
15	-15	(-15)(-15) = 225	Note that a minus times a minus always equals a plus.
20	-10	(-10)(-10) = 100	
25	- 5	(-5)(-5) = 25	
30	0	(0)(0) = 0	
35	5	(5)(5) ·= 25	
40	10	(10)(10) = 100	
45	15	(15)(15) = 225	
	Σ(X-u) =0,	$\Sigma(X-u)^2$ · 700	

By squaring the differences, we have eliminated the problem of a total being equal to zero. But this total of the squares does not consider the number of observations that contributed to the dispersion. We can correct for this by dividing the $(X-u)^2$ by the number of observations. The result is a value called variance. For our example:

$$\text{variance} = \frac{\Sigma(X-u)^2}{N} = \frac{700}{7} = 100$$

FRAME 38.

Squaring the individual differences increased their magnitude. We must adjust or correct for this increased magnitude or our measure of dispersion will be far too large and thus not representative. The method for correcting this is to perform the opposite operation. Since we squared the individual differences, the opposite operation would be to take the square root of the variance. The square root of the variance is called standard deviation. Standard deviation is a measure of dispersion Just like the centimeter is a measure of distance. Just as 5 centimeters represents a greater distance than 2 centimeters, a standard deviation of 4 represents a greater dispersion than a standard deviation of 2. The more variability or scatter in a group of observations, the larger the standard deviation will be. To find the standard deviations when the variance is 100, the following equation is used.:

$$\text{standard deviation} = \sqrt{\text{variance}} = \sqrt{10} = 10$$

FRAME 39.

SUMMARY

1. Compute the population mean. The formula for the mean is $u = \frac{\Sigma X}{N}$. The procedure is adding the values of the data and dividing the total by the number of observations.

2. Subtract the mean from each value of the data. The formula is $(X - u)$.

3. Square the difference between each value and the mean. The formula is

$(X - u)^2 = (X - u) \bullet (X - u)$.

4. Add together the squared difference between each value and the mean.

The formula is $\Sigma (X - u)^2$.

5. Divide the total (of the squared difference between each value and the mean) by the number of observations. The formula is $\frac{\Sigma (X-u)^2}{N}$

This value is called the variance.

6. Take the square root of the variance which will result in the standard deviation.

The formula is $\sqrt{\frac{\Sigma (X - U)^2}{N}}$

The standard deviation can thus be computed using the formula:

$$\sigma = \sqrt{\frac{\Sigma (X - u)^2}{N}}$$

where σ (pronounced sigma) is the symbol used to represent population standard deviation.

FRAME 40.

Just as we make a distinction between population mean (u) and sample mean ($\bar{X}$), we also make a distinction between population standard deviation (s) and sample standard deviation (s). Sample standard deviation is computed using the formula:

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}}$$

where s represents sample standard deviation, n the number of observations in the sample, and $\bar{X}$ the sample mean. The sum of the squared differences, $\Sigma(X - \bar{X})^2$, is divided by n-1, rather than by n alone. This adjustment is called degrees of freedom. Allowing for degrees of freedom (df) also gives a better estimate of the sample standard deviation since not all data points were analyzed.

As a rule of thumb, you should do the following: When dealing with a population, use the population mean and population standard deviation. When dealing with a sample, use the sample mean and the sample standard deviation.

QUESTION:

From the following data, what is the mean and standard deviation of the days of accumulated leave for all ten people in an office: 6, 6, 7, 8,

9, 11, 13, 13, 13, 14?

mean = _________

standard deviation ·__________

ANSWER TO FRAME 40.

mean = $u = \frac{\Sigma X}{N} = \frac{100}{10} = 10$

X	X - u	(X-u)²
6	-4	16
6	-4	16
7	-3	9
8	-2	4
9	-1	1
11	1	1
13	3	9
13	3	9
13	3	9
14	4	16
$\Sigma X = 100$	$\Sigma (X-u) = 0$	$\Sigma (x-U)^2 = 90$

Note: We are interested in a population of all ten people in an office. Thus we compute the population mean and standard deviation.

standard deviation = $\sigma = \sqrt{\frac{\Sigma (X - u)^2}{N}} = \sqrt{\frac{90}{10}} = \sqrt{\frac{9}{1}} = 3$

FRAME 41.

About 100 Government Property Lost or Damaged (GPLD) vouchers were processed this month. Five of these vouchers showed dollar amounts of $25, $30, $29, $23, and $33.

QUESTION:

Find the mean and standard deviation of these five dollar amounts.

mean = ______________________

standard deviation ________________________

ANSWER TO FRAME 41.

$$\text{mean} = \bar{X} = \frac{\Sigma X}{n} = \frac{140}{5} = 28$$

X	$X - \bar{X}$	$(X - \bar{X})^2$
25	-3	9
30	2	4
29	1	1
23	-5	25
33	5	25
115	0	64

Note that since we have a sample of 5 taken from the total population of 100 we must compute a sample mean and sample standard deviation.

$$\text{standard deviation} = s = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N-1}} = \sqrt{\frac{64}{5-1}} = \sqrt{16} = 4$$

FRAME 42.

SUMMARY OF THE MEASURES OF CENTRAL TENDENCY

MEAN

The mean is the sum of all observations divided by the number of observations. The mean considers all of the observations and it is greatly affected by extreme values. The mean can be computed as shown below.

$$\text{population mean} = u = \frac{\Sigma X}{N}$$

$$\text{sample mean} = \bar{X} = \frac{\Sigma X}{n}$$

MEDIAN

The median is the middle observation in a group of observations after arranging them in ascending or descending order. If there are an even number of observations, the median is found by taking the mean of the middle two values. Since the median is not affected by extreme values, it is the more appropriate measure of central tendency when there are extreme values in the data.

MODE

The mode is the most frequently occurring observation in a group of data. A typical application of the mode would be to determine the most frequent cause or reason for some event when the causes cannot be numerically measured.

FRAME 43.

SUMMARY OF THE MEASURES OF DISPERSION

RANGE

The range is the difference between the highest and lowest values in a group of observations. The range is simple to compute but only considers the two extreme values. The range should be used only in cases where a quick, cursory look at the data is required.

STANDARD DEVIATION

The standard deviation is a unit of measure of dispersion. The more variability or scatter in a group of data, the larger the standard deviation will be. Standard deviation is computed as shown.

Population Standard Deviation $= \sigma = \sqrt{\dfrac{\Sigma (X - u)^2}{N}}$

Sample Standard Deviation $= s = \sqrt{\dfrac{\Sigma (X - X)^2}{n-1}}$

Note that In the denominator, we are using n-1 rather than n. This expression, n-1, allows for degrees of freedom (df). Df will not be discussed here because it is beyond the scope of this subcourse. It is treated in inferential rather than descriptive statistics.

Proceed to the Practice Exercise for Lesson 2.

LESSON 2

PRACTICE EXERCISE

SITUATION I: A review of the ten TDY travel vouchers filed by members of the Fort Flatland Comptroller Office during the last quarter was conducted recently. It was found the following amounts (in dollars) were claimed from the airport: 14, 13, 9, 13, 11, 7, 6, 13, 6, 8. Note: Space to work the problems is provided at the end of this exercise.

Refer to Situation I to answer questions 1 through 5.

1. What is the MEAN cab fare?

   A. $100.
   B. $10.
   C. $90.
   D. $13.

2. What is the MEDIAN cab fare?

   A. $10.
   B. $13.
   C. $9.
   D. $18.

3. What is the MODE of this data set?

   A. $3.
   B. $6.
   C. $10.
   D. $13.

4. What is the RANGE of this data set?

   A. $7.
   B. $6.
   C. $9.
   D. $8.

5. What is the STANDARD DEVIATION of the cab fares for this data set?

   A. $9.
   B. $90.
   C. $3.
   D. $3.16.

SITUATION II: You are in charge of 28 keypunchers in the data conversion branch of the installation data processing center. You wonder how accurate your keypunchers are. A sample of the work done by the keypunch section indicates the following number of errors per hundred cards punched: 2, 5, 9, 3, 5, 6, 2, 6, 5, 7. Note: Space to work the problems is provided at the end of this exercise.

Refer to SITUATION II to answer questions 6 through 10.

6. What is the MEAN of your sample?

   A. 5.
   B. 6.56
   C. 50.
   D. 4.55

7. What is the MODE of your sample?

   A. 6.
   B. 5.
   C. 5 and 6; it is bimodal.
   D. There is no mode.

8. What is the MEDIAN of your sample?

   A. 4.
   B. 5.
   C. 6.
   D. 7.

9. What is the RANGE of your sample?

   A. 4.
   B. 5.
   C. 6.
   D. 7.

10. What is the STANDARD DEVIATION of your sample?

   A. 4.40
   B. 4.89
   C. 2.10
   D. 2.21

USE THIS SPACE TO WORK YOUR PROBLEMS FOR LESSON 2.

SITUATION I and Questions 1-5.

SITUATION II and Questions 6-10.

LESSON 2

ANSWER KEY AND FEEDBACK

Situation I

Construction of the table below facilitates computation of the measures of central tendency and dispersion. Note that the observations have been listed in ascending order.

X	(X - u)	$(X-u)^2$
6	-4	16
6	-4	16
7	-3	9
8	-2	4
9	-1	1
11	1	1
13	3	9
13	3	9
13	3	9
14	4	16
$\Sigma X = 100$	$\Sigma (X - u) = 0$	$\Sigma (X-u)^2 = 90$

1. B Mean: $= \Sigma X = 100 = 10$ (Frame 25)

2. A Median: $\frac{9 + 11}{2} = \frac{20}{2} = 10$. If you did not put the data in an ordered array, you would have obtained the wrong answer. (Frame 29)

3. D Mode: 13 (occurs 3 times) (Frame 31)

4. D Range = largest - smallest - 14 - 6 = 8 (Frame 34)

5. C
   Standard deviation $\sigma = \sqrt{\frac{(X-u)^2}{N}} = \sqrt{\frac{90}{10}} = \sqrt{9} = 3$

   (Frame 36)

Situation II

As was the case in I, a similar table for situation II is constructed below. Note that the $\bar{X}$ is used instead of u.

X	$(X - \bar{X})$	$(X-\bar{X})^2$
2	-3	9
2	-3	9
3	-2	4
5	0	0
5	0	0
5	0	0
6	1	1
6	1	1
7	2	4
9	4	16
$\Sigma X = 50$	$\Sigma (X-\bar{X}) = 0$	$\Sigma (X-\bar{X})2 = 44$

6. A Mean = $\bar{X} = \frac{\Sigma X}{n} = \frac{50}{10} = 5$ (Frame 23)

7. B Mode = 5 (occurs 3 times) (Frame 31)

8. B Median = $\frac{5 + 5}{2} = \frac{10}{2} = 5$ (Frame 29)

9. B Range = 9 - 2 = 7 (Frame 34)

10. D Standard Deviation = s = $\sqrt{\frac{\Sigma (\bar{X} X)^2}{n-1}}$ (Frame 40)

$$= \sqrt{\frac{44}{9}} = \sqrt{4.89} = 2.21$$

LESSON 3

THE NORMAL CURVE

LEARNING OBJECTIVE

When you have completed this lesson, you should be able to perform the following task:

Actions: Identify and define the six characteristics of the normal curve and apply the normal curve to decision making situations.

Conditions: You will be given several questions and problems dealing with the normal curve. You may use Lesson 3, The Normal Curve, which explains the topic.

Standards: Your explanations and solutions will be in accordance with the instructions in this subcourse.

FRAME 44.

INTRODUCTION

The distribution which occurs most frequently in nature is called the NORMAL DISTRIBUTION.

It occurs so often and in such diverse situations that mathematicians have carefully documented and quantified its characteristics. Examples of situations which are closely approximated by the normal distribution include: human characteristics, such as height, weight, and IQ; the outputs of many physical processes, such as the actual amount of paint put into a 'gallon' can; and the results of human activity, such as consumer demand or a product.

FRAME 45.

CHARACTERISTICS OF THE NORMAL CURVE

There are six characteristics of the normal curve which, when combined, make it different from any other type of distribution.

First, it is SYMMETRICAL around the arithmetic mean. In other words, the half on each side of the mean will exactly fit the other half when 'flipped over." Both identically shaped halves of the curve together take on the general shape of a bell.

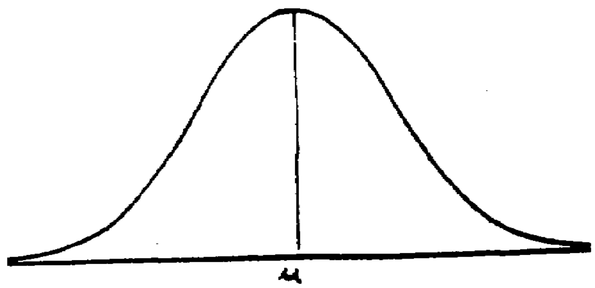
FRAME 46.

Second, the tails at each end of the normal distribution get closer and closer to the base line, the farther they are from the mean. The tails never touch the base line, however. We say the tails are asymptotic to the horizontal axis. This allows for extreme values - the three foot midget, the eight foot giant, very low expenses, or very high expenses, whichever the case may be.

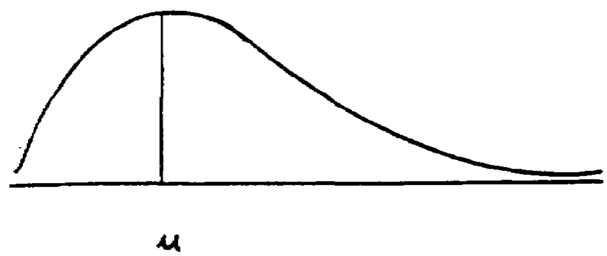
QUESTION A:

Which of the following distributions is symmetrical?

CURVE A:



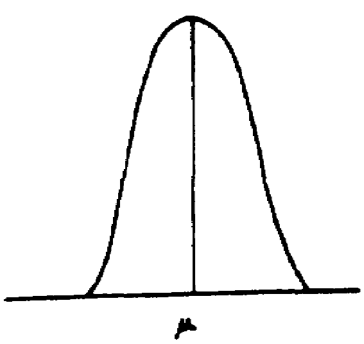
CURVE B:



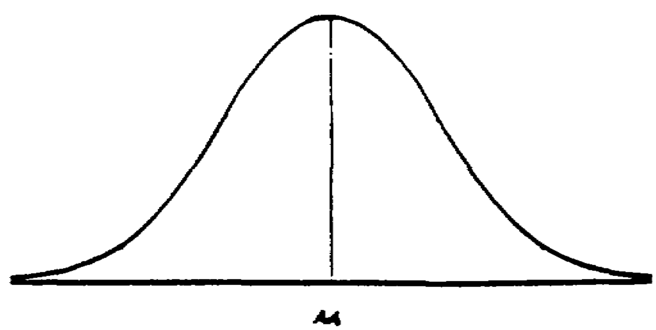
QUESTION B:

Which of the following appears to be a normal curve?

CURVE A



CURVE B



ANSWER TO A.

Curve A is symmetrical because the left half is identical to the right half. This is not true for curve b.

ANSWER TO B.

Curve B appears to be a normal curve because the tails get closer to but do not touch the base line. The tails on curve A are definitely touching the horizontal axis or base line.

FRAME 47.

Third, the mean, median and mode all lie at the center of the normal distribution. The mean is an arithmetic average that is computed from all the values, and reasonably lies in the middle. The median divides the distribution into two equal parts: half greater, half smaller than the median. The mode is the most frequently occurring observation. By symmetry (the first characteristic) half the observations must lie on either side of the mean. So the mean, and the median are the same. The height of the curve at any point is the frequency (number of occurrences), so the mode must occur at the highest point on the curve. This must be the point that divides the curve into two equal parts. So the mean, median, and mode all lie at the same place - the center - of the normal curve.

QUESTION:

A. If the curve was not normal, which measure of central tendency would indicate the highest

point? ______________________________

B. If the curve was not normal, which measure of central tendency would divide the area under

the curve in half? ________________________

ANSWER TO FRAME 47.

A. The mode would indicate the highest point because the height represents the frequency.

B. The median divides the distribution in half. Half the items in the distribution are greater than the median, half are less than the median.

FRAME 48.

Fourth, the total area under the curve is one or 100%. All the items in the distribution are represented under the curve. The normal curve represents the shape of all the observations in the population.

QUESTION:

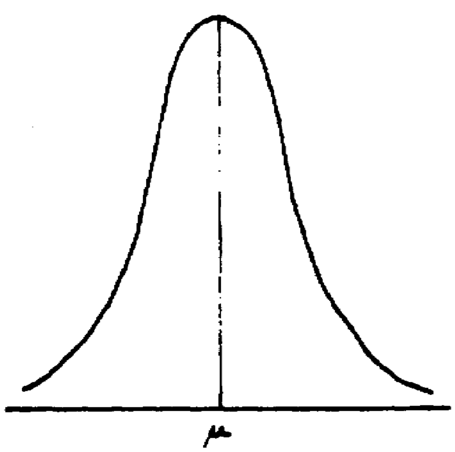
Are all the elements of a normal population represented by the normal

curve? _____________________________

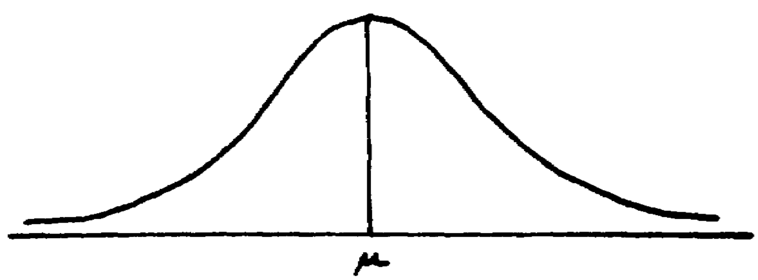
FRAME 49.

Fifth, the mean and standard deviation completely determine a normal curve. The normal distribution, is a two-parameter distribution. In order to draw any normal distribution, all we need are the mean and standard deviation. All normal curves with the SAME mean and standard deviation are identical. We previously defined standard deviation as a unit of measure of dispersion. The more spread out the data, the greater the dispersion, and thus the larger the standard deviation. As the standard deviation gets larger, the normal curve tends to flatten out. Which of the following normal curves has the larger standard deviation?

CURVE A



CURVE B



ANSWER TO FRAME 48.

Yes. By definition, 100% (or all) of the observations of a normal population are included under

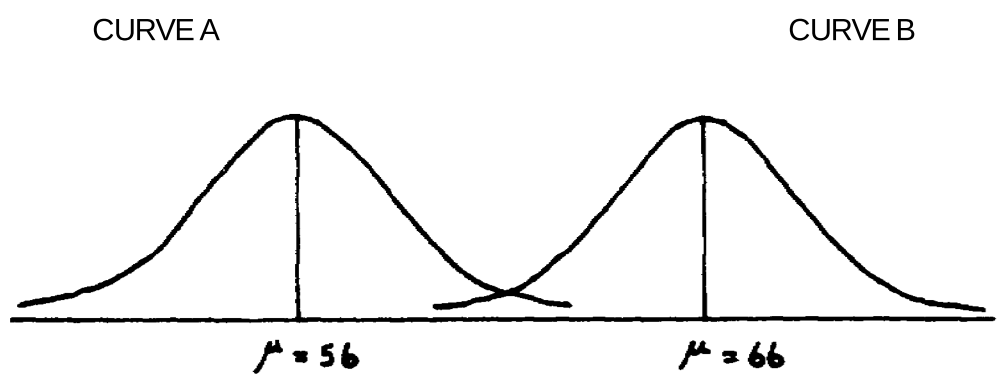
the normal curve.

ANSWER TO FRAME 49.

Curve B because it appears flatter or more spread out than Curve A.

FRAME 50.

We might say that standard deviation specifies the shape (flat, peaked) of the normal curve. The mean, then, determines the location of the normal curve along the horizontal base line. Two normal curves may have exactly the same standard deviation (and thus the same shape) but have different means and thus be centered at different locations along the baseline. For example, let's look at two curves which both have a standard deviation of 2 inches. Curve A has a mean of 56 inches and Curve B has a mean of 66 inches. The curves would appear as follows along the same continuum or baseline:



QUESTION:

How would you specify the shape and the location of a normal curve?

ANSWER TO FRAME 50.

The STANDARD DEVIATION specifies the shape of a normal curve and the MEAN specifies the location of a normal curve along the base line.

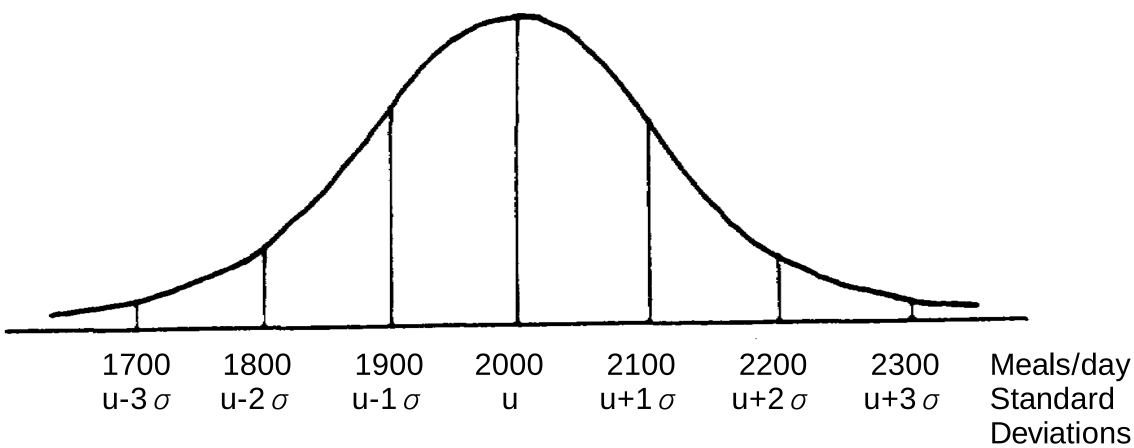
FRAME 51.

The sixth characteristic relates horizontal distances on the base line to areas under the normal curve.

Thus far we have measured the horizontal base line in terms of the unit of measure we are dealing with in the specific problem: dollars, inches, etc. It would be more convenient if we could use the same unit of measure for all problems. Our unit of measure for dispersion, the standard deviation, is used to measure this horizontal distance. When using standard deviations, we always measure from the mean of the distribution. Thus we can find a relationship between the area under the normal curve and the number of standard deviations measured from the mean.

Suppose we collected data about the number of meals served per day in the fort Flatland Consolidated Mess for the last 6 months. The mean was computed to be 2000 and the standard deviation was 100.

The normal curve for the population would appear as:



FRAME 52.

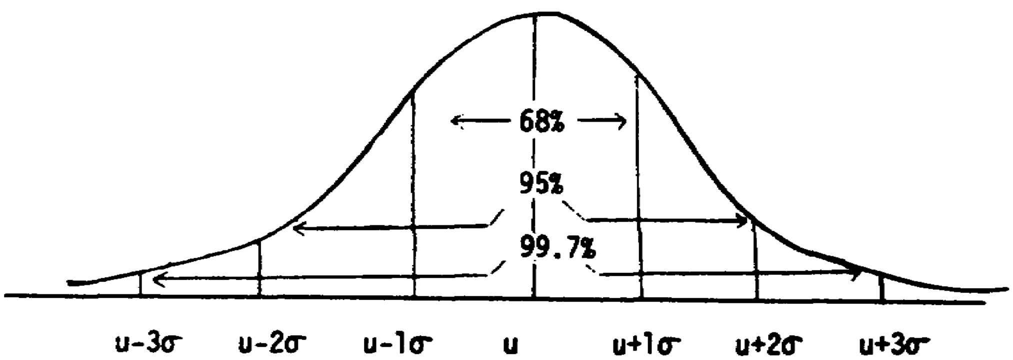
THE EMPIRICAL RULE

Notice the curve has two horizontal scales; one measured in meals/day and the other measured in standard deviations from the mean. The distance 2000 to 2200 is the same as the distance u to u +2σ A special relationship exists between the area under the normal curve and the horizontal distance from the mean measured in standard deviations.

This relationship is the sixth characteristic of a normal curve and is commonly referred to as the Empirical rule.

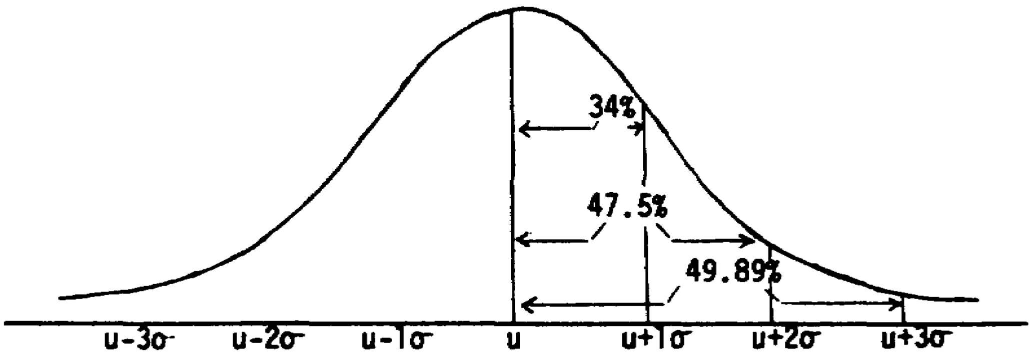
The rule states that: a) the area under the normal curve from u -a to u + σ is approximately 68% of the total area under the curve. b) the area under the normal curve from u -2σ to u + 2σ is approximately 95%. c) the area under the normal curve from u -3σ to u + 3σ is approximately 99.7%.

The following diagram illustrates this Empirical rule:



FRAME 53.

Since the normal curve is symmetric, we know that half of these areas will lie on either side of the mean. The areas for the right side are shown in the diagram below:



The areas for the left side are exactly the same as for the right side.

QUESTION:

What percentage of the area under the curve lies between:

a. u - 1σ and u + 1σ ? ______

b. u - 2σ and u + 2σ ? ______

c. u - 3σ and u + 3σ ? ______

d. u - 1σ and u ? ______

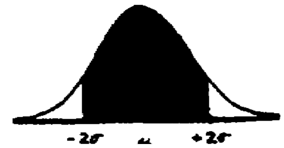
e. u and u + 1σ ? ______

f. u - 1σ and u + 2σ ? ______

ANSWER TO FRAME 53.

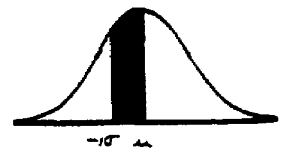
a. 68%

b. 95% by the EMPIRICAL rule



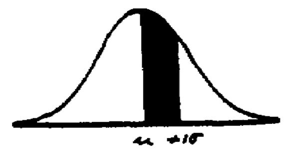
c. 99.7%

d. 34%

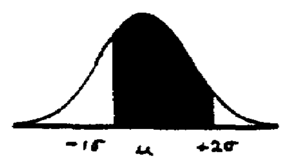


using the symmetry characteristic

e. 34%

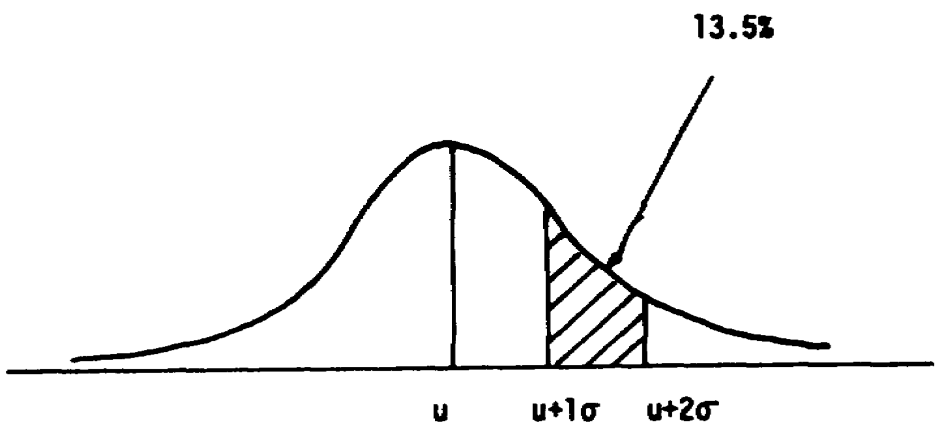


f. 34% + 47.5$% = 81.5% In this case we identify the area under the curve on each side of the man and add the areas.



FRAME 54.

Suppose we want to find the area under the curve between u + 1σ and u + 2σ . This area cannot be measured directly because one edge doesn't lie on the mean. Since the relationship between the area under the curve and the distance on the horizontal base line is measured from the mean, this area must be found indirectly. we can find the area from u to u + 2σ (47.5%) and the area from u + 1σ (345). The area from u to u + 2σ includes the area from u to u + 1σ. 8y subtracting the smaller area from the larger area we eliminate this common area. what's left is the area from u + 1σ to u + 2σ, or 47.5% - 34% - 13.5%.

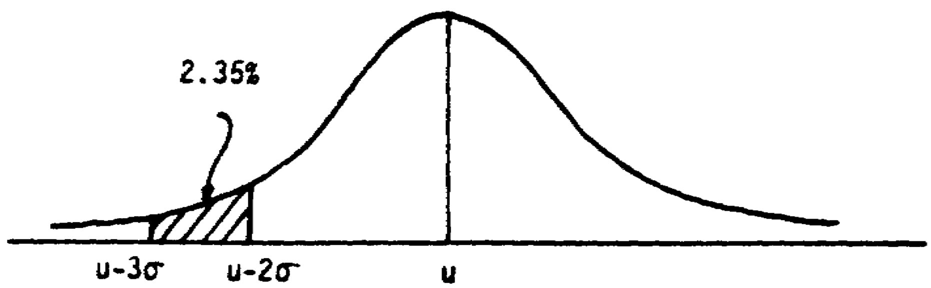


QUESTION:

What is the area under the curve from u - 2σ to u - 3σ? Draw a diagram and shade the area in question.

ANSWER TO FRAME 54.

The area from u - 3 σ to u is 49.85% (half of 99.7%). The area from u - 2 σ u is 47.5%. The difference is 49.85 - 47.50 - 2.35%.



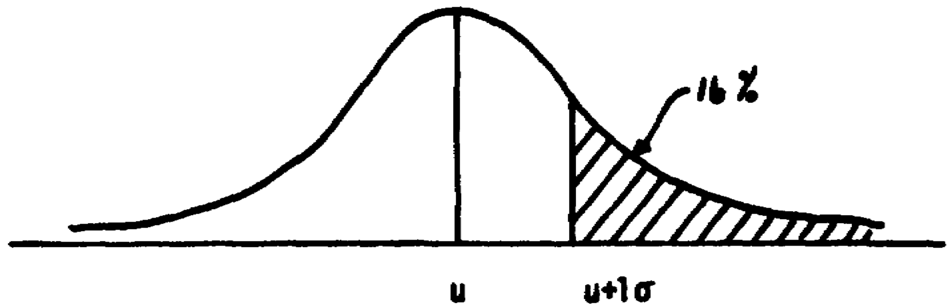
FRAME 55.

One of the characteristics of the normal curve is that the area under the curve is 100%. By symmetry, 50% of the area lies above the mean and 50% lies below the mean. We need this characteristic to answer the following type of question:

What area lies under the normal curve from u + 1σ to as far to the right as the curve goes? Since the tails of the curve approach, but never touch the horizontal base line, the curve extends infinitely far in both directions and we have to answer the question indirectly. We do know that half of the area lies to the right of the mean. The area from u to 1σ we know is 34%. The difference 50% - 34% = 16% is that area in the right half of the curve that lies above u + 1σ. The shaded area is referred to as the right tail of the curve.

ANSWER TO FRAME 55.

The diagram would be:

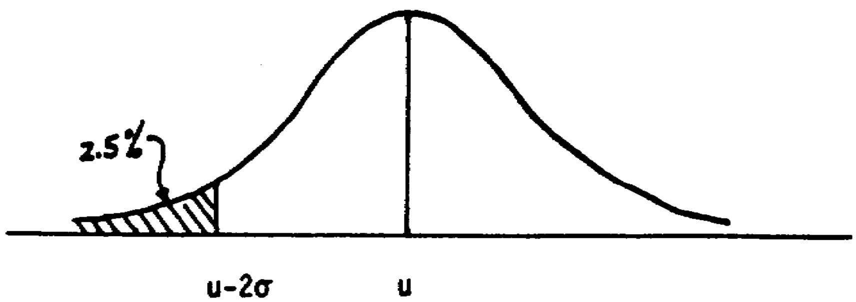


FRAME 56.

What is the area under the curve in the left tail below u - 2σ ? Include a diagram and shade the area in question.

ANSWER TO FRAME 56.

The area to the left of the mean is 50%. The area from u - 2σ to u is 47.5%. The difference is 50% - 47.5%1 = 2.5% and is the left tail area below u - 2σ. The diagram would appear as follows:



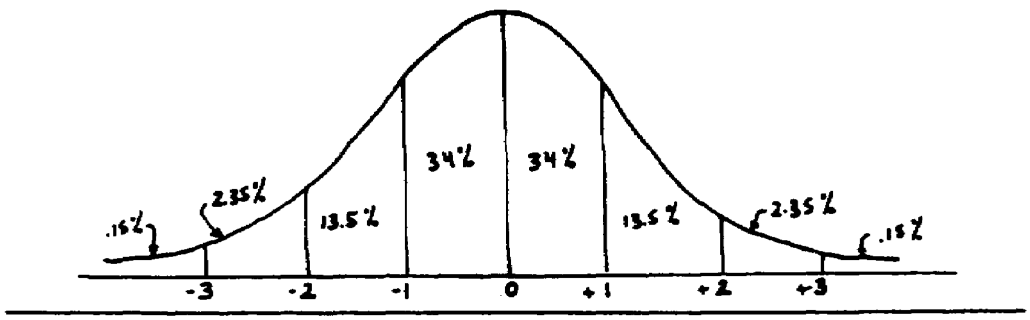
FRAME 57.

THE z SCORE

The following diagram shows the areas under the curve for whole standard deviations. You should be able to compute each of them given only the six characteristics of the normal curve.

Approximate Areas under the

Normal Curve



FRAME 58.

Notice that only the number of standard deviations above (or below) the mean are shown as the horizontal base line. The symbols u and σ have been deleted because we know that -2 means 2 standard deviations below the mean or u -2σ. To avoid confusion, we will label this new scale "z" and specify that z = the number of standard deviations from the mean. So, if we say z = -3 we know we are talking about a distance of 3 standard deviations to the LEFT of the mean.

QUESTION:

What is the value of z if we are interested in a distance:

a. 1 standard deviation above the mean?

b. 2 standard deviations below the mean?

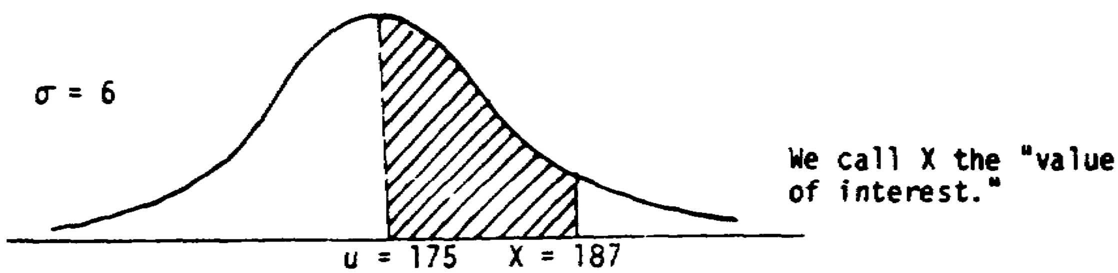
ANSWER TO FRAME 58.

a. = +1

b. = -2

FRAME 59.

Suppose we had evidence that the heights of recruits followed a normal distribution with a mean of 175cm and a standard deviation of 6cm. We are interested in finding what percent of the population of all recruits have heights between 175cm and 187cm. A diagram would look like:



To answer the question we need to find the shaded area. We know there is a relationship between the area under the normal curve and the distance along the horizontal base line measured in standard deviations from the mean (or z"s).

The distance from u = 175cm to X = 187cm is 12cm. The value of one standard deviation is σ = 6cm. Then the distance in question along the horizontal axis must be 12/6 = 2 standard deviations from the mean, or z -2. We can formalize this thought process with the following formula:

$$z = \frac{X = u}{\sigma}$$ where,

z = number of standard deviations from the mean.

u = mean of the population.

σ = standard deviation of the population

X = value of interest.

FRAME 60.

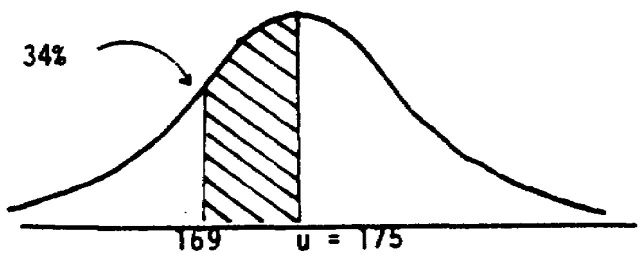
Further, we know that the area under the curve for z = 2 is about 47.5% by the Empirical rule.

In a similar manner, find the percentage of recruits that have heights between 169 cm and 175 cm.

Draw a diagram and shade the area of interest.

ANSWER TO FRAME 60.

The diagram would appear as:



Convert the distance from X = 169 to u - 175 into a number of z's by using the relationship:

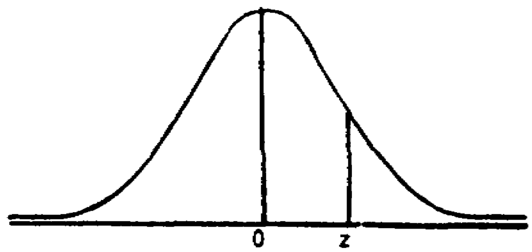
$$z = \frac{X - u}{\sigma} = \frac{169 - 175}{6} = \frac{-6}{6} = -1$$

Thus X = 169cm and lies 1 standard deviation below the mean. This corresponds to an area of about 34%.

FRAME 61.

When z is a whole number, the Empirical rule can be used to estimate the appropriate area under the normal curve. What happens if z is not a whole number? Suppose we needed to know the percent of recruits with heights between 175cm and 185cm. Using the z equation, we would get $z = \frac{X - u}{\sigma} = \frac{185 - 175}{6} = \frac{10}{6} = 1.67$. There is no value in the Empirical rule for 1.67 standard deviations above the mean. Interpolating (finding the value 2/3 of the way between 1 and 2 standard deviations) would cause a large error because the normal curve is not a straight line. Computing the area using a pocket calculator would be a lengthy and tedious task. The easiest and most common method is to look up the area in a table. Such a table, titled AREAS UNDER THE STANDARD NORMAL CURVE, is on the next page.

TABLE FOR AREAS UNDER THE STANDARD NORMAL CURVE



z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

FRAME 62.

To find the area for z = 1.67, find 1.67 in the perimeter of the table. The units and tenths digits (1.6) determine the row and the hundredths digit (.07) determines the column. The value at the intersection (.4525) is the area under the curve for a distance 1.67 standard deviations above the mean. The diagram below demonstrates how the area was found:

z	0	1	2	3	4	5	6	7
0.0	*	*	*	*	*	*	*	*
*								
*								
*								
1.6	.4452	.4463	.4474	.4484	*	*	*	.4525

In a similar manner we would find the area under the normal curve from the mean to z = 2.51 to be .4940. Find 2.51 in the perimeter of the table and, at the intersection of the 2.5 row and .01 column, read the area .4940. If the value of z is negative, it simply means the area is to the left of the mean. We can ignore the sign and find the area under the curve as before. To preclude interpolating between two z values, always round to two decimal points. The z value will be rounded up if it is 5 or greater, or rounded down if it is 4 or less. Indicated below are examples of rounding the z value.

Original z Value	Rounded	Area
1.2340	1.23	.3907
1.6667	1.67	.4525
2.4934	2.49	.4936

The normal curve table gives areas under the curve for a number of standard deviations (z's) measured from the mean.

FRAME 63.

Find the areas under the curve associated with the following values for z:

a. 1.34

b. 2.69

c. 1.00

d. 3.00

e. -.09

ANSWER TO FRAME 63.

a. .4099 or 40.99%

b. .4964 or 49.64%. The minus sign indicates this is to the left or "below" the mean.

c. .3413 or 34.135. The Empirical rule gave an estimate of 34% -- very close.

d. .4987 or 49.875. Again very close to the Empirical rule value.

e. .0359 or 3.59%

FRAME 64.

It is good procedure to draw a picture or diagram of the problem that includes all available information. The diagram makes it easier to visualize which part of the normal curve you are interested in. Also, after solving the problem, the diagram serves as a quick check to see if your answers are reasonable."

You are responsible for ordering uniforms for the 10,000 recruits expected to be trained at Fort Flatland next year. Size of the uniforms ordered depends on height in centimeters. There are five sizes available. Please note that our sizes overlap at the extremes (end of one, beginning of the other). We are using numbers here that are easily manipulated when performing the necessary subtractions and divisions. You should assume, then, that we are dealing with five gradations where size A ends at 164.9 and size 8 begins at 165, and so on. In doing the computations, however, use the numbers that are given.

Size A fits recruits having a height of 165cm or less.

Size B fits recruits 165cm to 173cm tall.

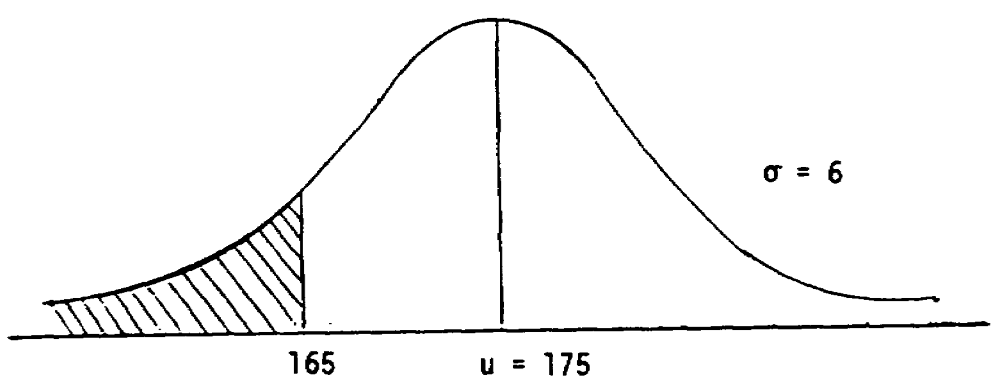
Size C fits recruits 173cm to 180cm tall.

Size D fits recruits 180cm to 191cm tall.

Size E fits recruits 191cm or taller.

Past history has shown that the heights of recruits tends to be normally distributed with u = 175cm and σ = 6cm. How many uniforms of each size should be ordered?

Consider Size A first. A diagram with the area of interest shaded would appear as:



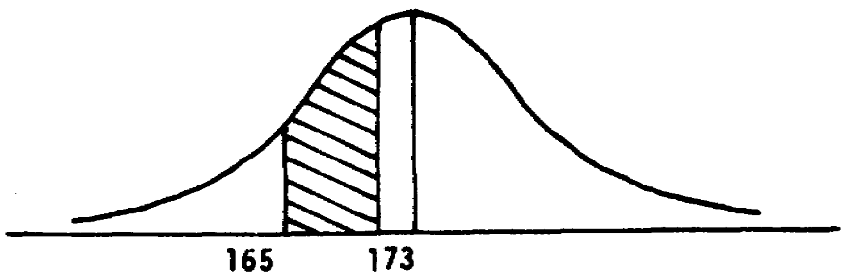
FRAME 65.

To find the shaded area we first find the area from 165cm to 175cm because we always measure standard deviations from the mean. Using the z equation

$$z = \frac{X - u}{\sigma} = \frac{165 - 175}{6} = \frac{-10}{6} = -1.67$$

The table value for the area under the normal curve corresponding to -1.67 standard deviations from the mean is .4525. The total area to the left of the mean is .5000, so the area of interest is .5000-.4525 = .0475 (4.75%). Thus, we can expect 475 (.0475 X 10,000 - 475) recruits to require Size A uniforms.

For Size B the area of interest would be:



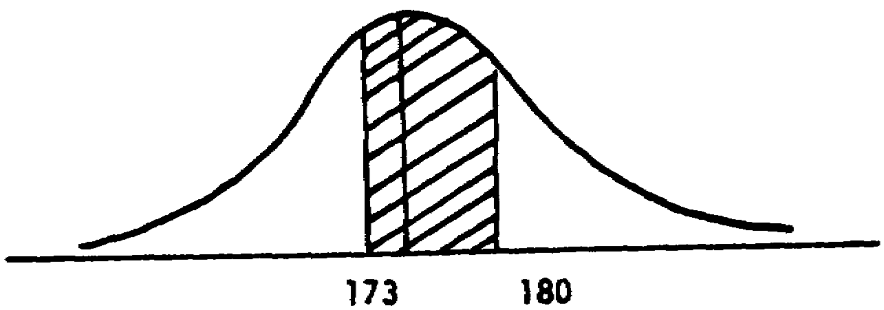
u - 175

We already know that 165cm is -1.67 standard deviations from the mean, so 45.25% of the area under the curve lies between the mean and 165cm. Finding the area under the curve between the mean and 173cm will allow us to solve the problem. Using the z equation:

$$z = \frac{X - u}{\sigma} = \frac{173 - 175}{6} = \frac{-2}{6} = -.33$$

FRAME 66.

The table value for the area under the normal curve corresponding to -.33 standard deviation from the mean is .1293. The area between 165cm and 173cm is .4525 - .1293 = .3232. Thus we can expect 32.32% of the 10,000 recruits to require Size B uniforms, which is 3,232. For Size C the area of interest would be:



u = 175

We already know that 12.93% of the area lies between 173cm and 175cm. Using the z equation we can find the number of standard deviations from u = 175cm to 180cm:

$$z = \frac{X - u}{\sigma} = \frac{180 - 175}{6} = \frac{5}{6} = .83$$

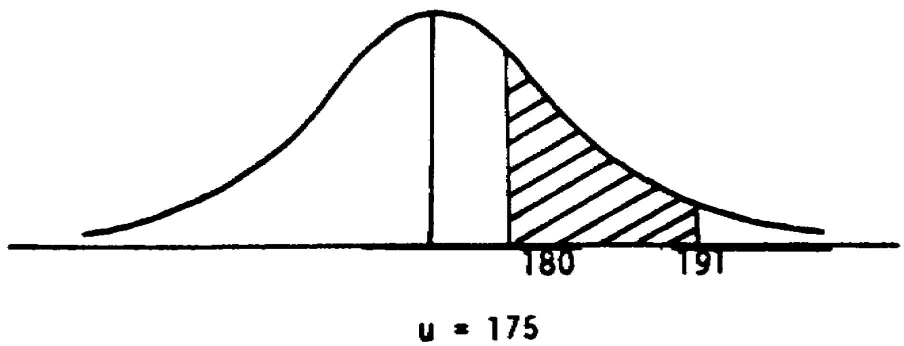
The table value for .83 standard deviations above the mean is .2967. Then the area between 173cm and 180cm is .1293 + .2967 = .4260. We would expect 42.60% of the recruits to require Size C uniforms. Thus, we can expect 4,260 (.426 x 10,000 = 4,260) recruits to require Size C uniforms.

QUESTION:

What percent of the recruits would take a Size D uniform? Draw the diagram.

ANSWER TO FRAME 66.

The area of interest is shaded:



We found the area under the curve from 175cm to 180cm is .2967. The number of standard deviations from the mean to 191cm is

$$z = \frac{X - u}{\sigma} = \frac{191 - 175}{6} = \frac{16}{6} = 2.67$$

This corresponds to an area of .4902. So the area from 180cm to 191cm is .4962 - .2967 = .1995.

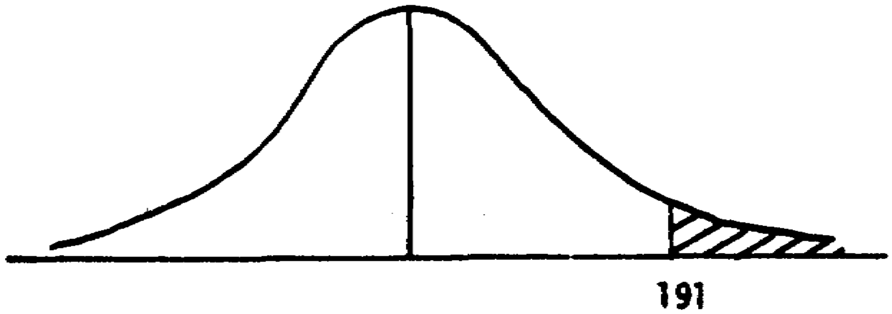
We would expect 19.95% of the recruits to require Size O.

FRAME 67.

How many of the recruits would wear a Size E uniform? Draw the diagram.

ANSWER TO FRAME 67.

The area of interest is shaded:



u = 175

We found the area under the curve from the mean to 191cm was .4962. Since 50% of the area falls to the right of the mean, the area from 191cm or greater is .5000 - .4962 = .0038. There are 10,000 recruits expected, so (10,000) (.0038) · 38 recruits who would require a Size E uniform.

ANSWER TO FRAME 68.

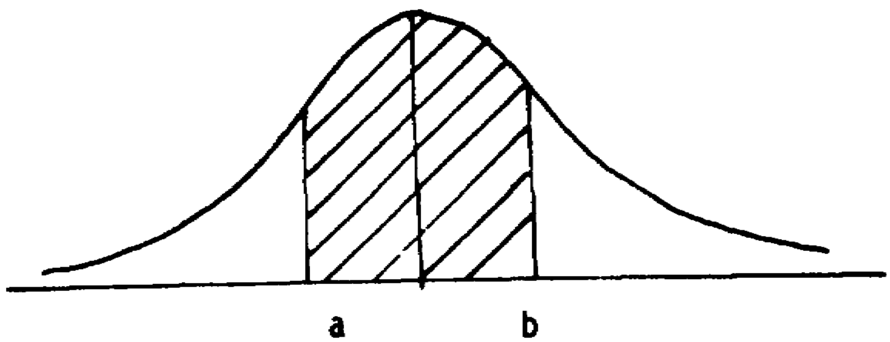
A summary of the number of uniforms for each size is indicated below.

Size	%	# of Uniforms
A	4.75	475
B	32.32	3232
C	42.60	4260
D	19.95	1995
E	.38	38
TOTAL	100.00	10,000

Note that the total % is 100% which is the fourth characteristic of the normal curve as indicated earlier in the lesson (page 73).

FRAME 69.

Often we know the area under the curve and want to find the associated values on the horizontal base line. For example, the middle 80% of all trainees could be expected to have heights between what two values? Unless otherwise specified, it is assumed the given area is centered on the mean. A diagram would appear as follows:



u = 175

The shaded area is 801 (in this case) of the total area and the distance from a to 175cm is the same as the distance from 175cm to b. We want to find values for a and b. By symmetry, the shaded area on each side of the mean is 40% (half of 80%). Recall the construction of the table of areas under the normal curve. The perimeter values give standard deviations and the body of the table gives areas. In this case, we will work in the opposite direction. Instead of knowing standard deviations and looking up the corresponding area, we will know the area and look up the corresponding standard deviation. In our example, we will look up an area of .4000. From the table we see that no entry is exactly .4000. We will use the higher z value if the area (%) falls between two areas (%) on the table. For an area of .4000, the z value would be 1.29.

Initial Area Value	Lower Area (A) and Value	Higher Area (%) and Value	Value Selected
.4500	.4495	.4505	
	1.64	1.65	1.65
.4000	.3997	.4015	
	1.28	1.29	1.29
.3400	.3389	.3413	
	.99	1.00	1.00
.2500	.2486	.2518	
	.67	68	.68
.2075	.2054	.2088	
	.54	55	.55

FRAME 70.

The z equation can be manipulated algebraically. Since a and b are the values of interest, we will use them instead of X:

	Below the Mean	Above the Mean	
1.	$-z = \frac{a - u}{\sigma}$	$z = a\ \frac{b - u}{\sigma}$	
2.	$-z\sigma = a - u$	$z\sigma = b - u$	Multiply both sides by σ
3.	$u - z\sigma = a$	$u + z\sigma = b$	Add u to both sides.

As before, -z is used to denote standard deviations below the mean, and z is used to denote standard deviations above the mean.

We know 3 of the values in equation 3: u = 175, z = 1.29, and $\sigma = 6$. Given this information we can compute a and b.

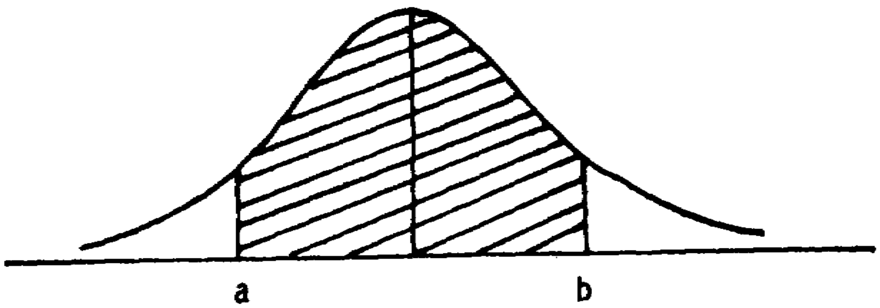
$a = u - z\sigma$	$b = u + z\sigma$
a = 175 - (1.29) (6)	b = 175 + (1.29) (6)
a = 175 - 7.74	b = 175. + 7.74
a = 167.26	b = 182.74

Thus, we can say that the middle 80% of the recruits have heights between 167.26cm and 182.74cm.

The middle 95% of all trainees could be expected to have heights between what two values? Use the table of areas under the normal curve to find the values. Draw the diagram.

ANSWER TO FRAME 70.

The diagram would be:



u = 175

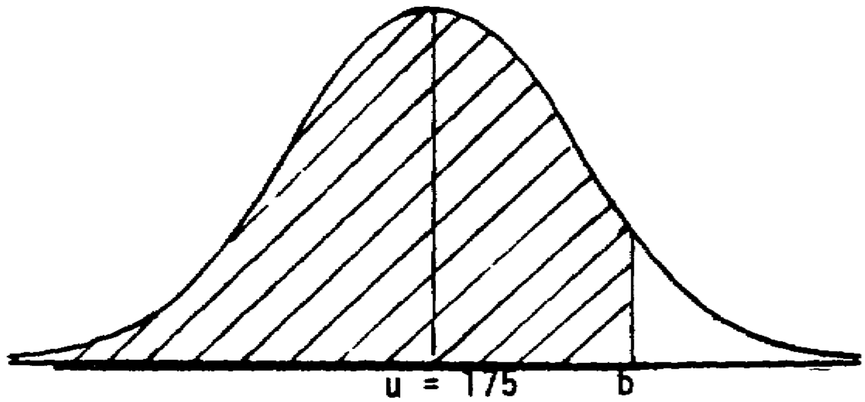
Half of the 95% would fall above and half below the mean. So we want to find the value of z associated with .4750. Searching the table, we find exactly .4750 corresponding to 1.96 standard deviations. Using the modified z equation:

$a = u - z\,\sigma$	$b = u + z\,\sigma$
a = 175 - (1.96) (6)	b = 175 + (1.96) (6)
a = 175 - 11.76	b = 175 + 11.76
a = 163.24	b = 186.76

Thus, we can state the middle 95% of the recruits have heights between 163.24cm and 186.76cm.

FRAME 71.

We can also find the value on the horizontal base line when we are given the areas under the curve less than or greater than that amount. For instance, suppose we wanted to find the height below which 90% of all trainees fell. We would draw a diagram as follows:



find b when the shaded area is 90%, we must first determine the area between u and b. Since the entire area is 90%, and the area below u is 50%, the area between u and b is: 90% - 50% = 40%. Converting to decimal form, we look in the table for .4000. As you might recall, that area falls between .3997 (z= 1.28) and .4015 (z= 1.29). Our rule says that we must use the larger z value if our area falls between two areas on the table, so z = 1.29 for an area of .4000.

Using the formula from Frame 70 for values above the mean:

$$b = u + z\,\sigma$$

$$= 175 + 1.29\,(6)$$

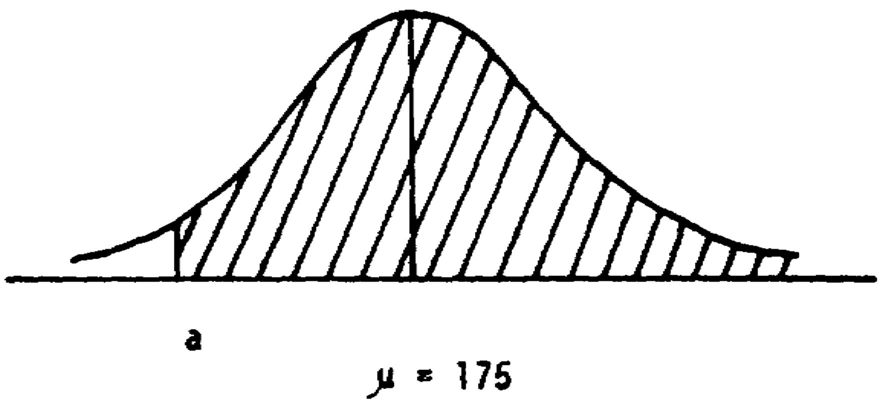
$$= \underline{182.74}$$

In other words, 90% of the time, recruits will be less than or equal to 182.74cm in height.

QUESTION:

Above what height will 95% of all recruits be? ________________________________

ANSWER TO FRAME 71.



The area between a and u is: 955 - 505 = 45.
The z value for .4500 is 1.65. Therefore:

a = 175 - 1.65(6)
= 175 - 9.9
= 165.10

Thus, 95% of the time, recruits will be at or above 165.10 cm in height.

FRAME 72.

In summary, the normal distribution has the following 6 characteristics:

1. It is symmetrical around the mean.

2. The tails of the curve are asymptotic, that is, they approach but never touch the base line.

3. The mean, median, and mode all lie at the center of the curve.

4. The area under the curve is 1.00 or 100%.

5. The mean determines the location of the center of the distribution and the standard deviation determines the shape of the distribution.

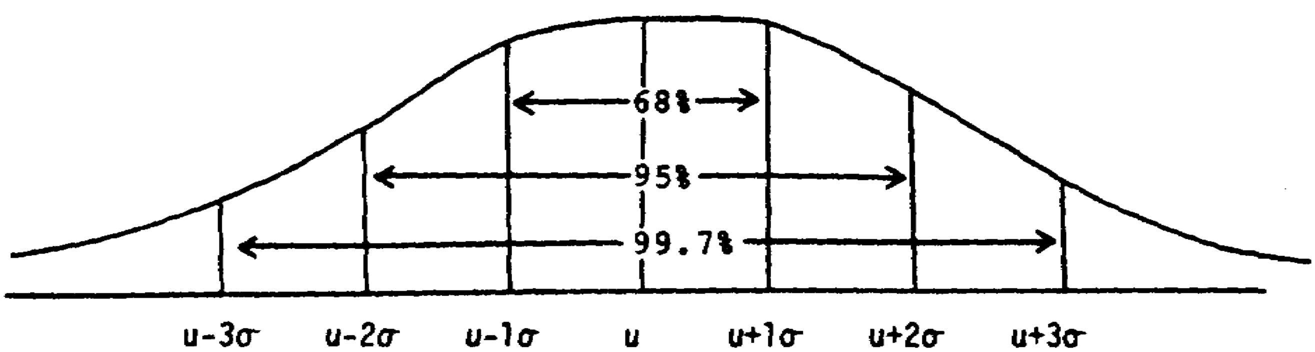
6. The Empirical rule relates areas under the curve to horizontal distances on the baseline by using the following approximations:

a. u + 1σ encompasses about 68%.

b. u + 2σ encompasses about 95%.

c. u + 3σ encompasses about 99.7%.

Schematically, the Empirical rule would appear as follows:



The Empirical rule only reflects approximations whereas the normal curve table is more precise. The Empirical rule provides a quick analytical tool when the normal curve table is not available, such as in a meeting or a review and analysis. The z equation can be used to convert distances from the mean in any unit of measure to our standard unit of measure, standard deviations from the mean. The equation is z = X - u divided by σ, where z is the number of standard deviations from the mean. Expressed symbolically, $z = \frac{X - u}{\sigma}$

We use the table of areas under the standard normal curve, or the Empirical rule, to find the area under the curve between the mean and z. The Normal Curve Table is the more preferred choice.

Proceed to the Practice Exercise for Lesson 3.

LESSON 3

PRACTICE EXERCISE

REQUIREMENT: Answer the following seven (7) problems below.

Situation: An analysis of your organization's self-service supply account expenditures for the last year indicates an average monthly expense of $2370 with a standard deviation of $375. Your experience has been that self-service accounts tend to be normally distributed and there is no reason to doubt this is the case for your current organization.

1. What percent of the time will your monthly expenditures exceed $2500?

2. How often will expenses be less than $1995 in a month?

3. You would expect expenditures to be between $2400 and $2600 what percent of the time?

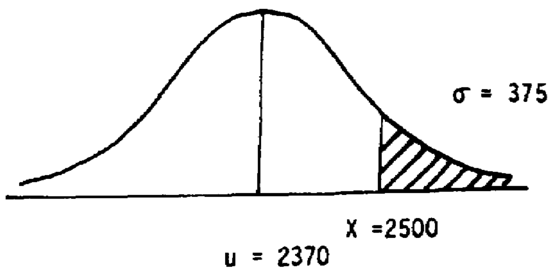
4. Between what two values around the mean would you expect expenditures to be 80% of the time?

5. Between what two values around the mean would you expect expenditures to occur 95% of the time?

6. What should you budget if you wanted to be 90% sure of having enough money to cover your supply expense for a given month?

7. What should you budget if you wanted to be 95% sure of having enough money to cover your supply expense for a given month?

ANSWER KEY AND FEEDBACK



$z = \frac{X - u}{\sigma} = \frac{2500 - 2370}{375} = \frac{130}{775} = .35$

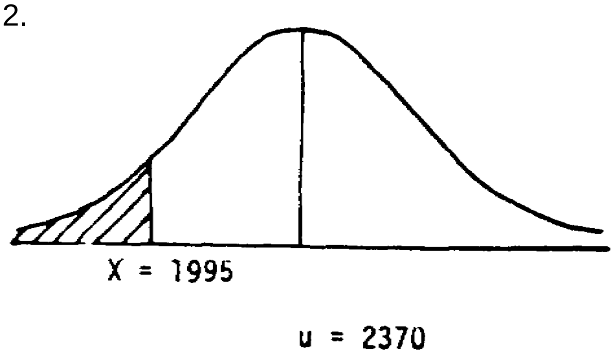
From table: z = .35,
area = .1368
or 13.68%

Since this is the area between u and X, we must subtract it from 50% to get the shaded area:

50% - 13.68% = 36.32%

There is a 36.32% chance of exceeding $2500. (Reference: Frame 65.)

2.

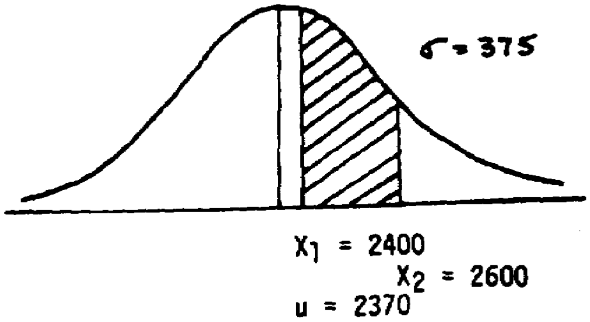


$z = \frac{1995 - 2370}{375} = \frac{-375}{375} = -1.0$

Using the Empirical rule, the area between u and X is 34%. Therefore, we would have a shaded area of 50% - 34% = <u>16%</u>. (Reference Frame 56.) Using the table, the area between u and X is 34.13%, so the shaded area of interest would be 50% - 34.13% = <u>15.87%.</u> (Reference Frames 61-64.)

From the table, there is a 15.87% chance of getting an expense that is less than $1,995.

3.



$$z = \frac{X1 - u}{\sigma} = \frac{2400 - 2370}{375} = .08$$

area = .0319 or 03.19%

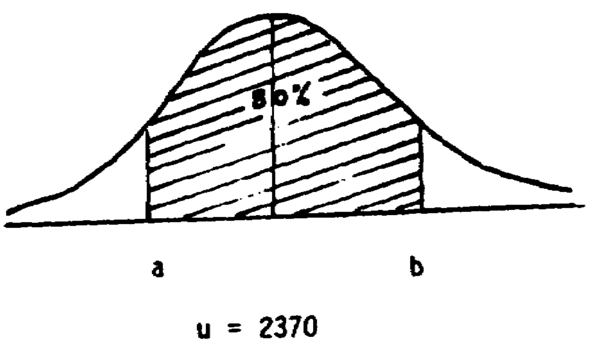
$$Z = \frac{x2 - u}{\sigma} = \frac{2600 - 2370}{375} = .61$$

area = .2291 or 22.91%

In order to find the shaded area, we must subtract the smallest area from the largest: 22.91% - 3.19% = 19.72%.

There is a 19.72% chance that the supply expenditures for a given month are between $2400 and $2600. (Reference: Frames 67-68.)

4.



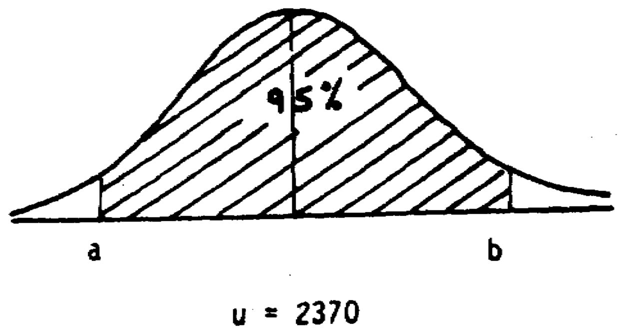
Area is 80%/2 = 40% or .4000, Therefore, z = 1.29.

$b = u + z\sigma = 2370 + 1.29 (375) = \$2853.75$

$a = u - z\sigma = 2370 - 1.29 (375) = \$1886.25$

80% of the time, supply expenses will be between $1886.25 and $2853.75. (Reference: Frames 69-70.)

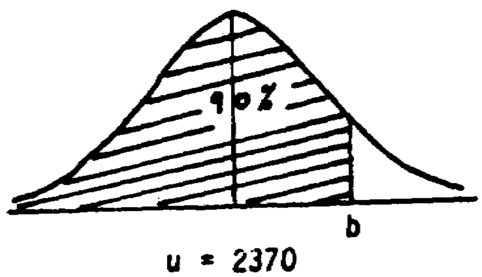
5.



This is the same as problem 4, except z = 1.96
a = u - zσ = 2370 - 1.96(375) = 1635

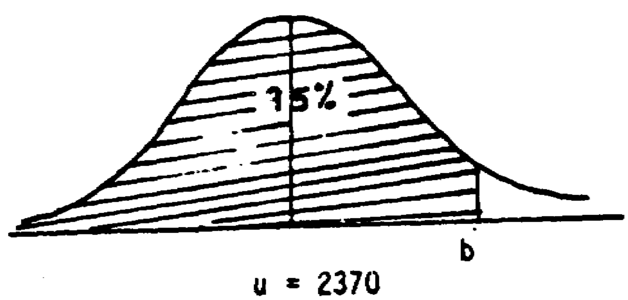
b = u + zσ = 2370 + 1.96(375) = 3105

6.



For this problem, the area between u and b is 90% - 50 = 40%. Therefore, z = 1.29 (see problem 4) and b = $2370 + 1.29 ($375) = $2853.75. Reference: Frame 71.)

7.



Same as problem 6, except the area between u and b is 951 - 50% = 45%. Therefore, z = 1.65 (.4500 or table between .4495 and .4505) and b = $2370 + 1.65 ($375) = $2988.75. (Reference: Frame 71.)