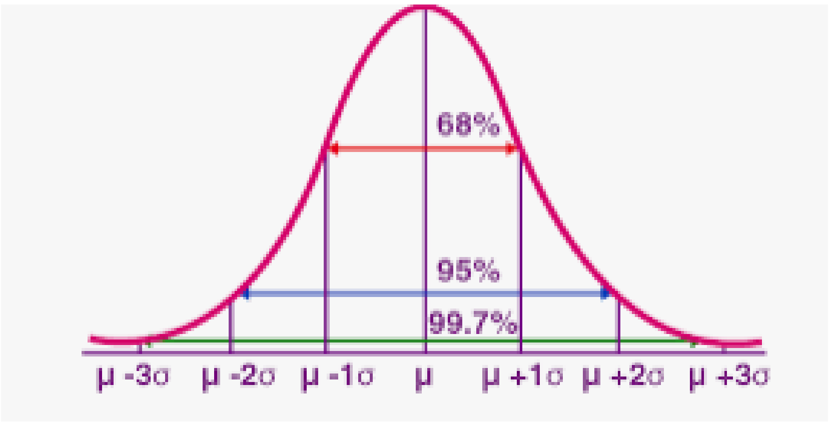
UNIT-4
Describing Data II

Syllabus:

Describing Data II: Normal distributions – z scores – normal curve problems– finding proportions – finding scores –more about z scores – correlation – scatter plots – correlation coefficient for quantitative data –computational formula for correlation coefficient – regression – regression line – least squares regression line – standard error of estimate – interpretation of r2– multiple regression equations – regression toward the mean.

The Normal Distributions

- → A Normal distribution (or Gaussian distribution) is a continuous probability distribution that is symmetrical on both sides of the mean, so that right side of the center is mirror image of the left side.
- → Normal distribution is so important because it accurately describe the distribution of values for many natural phenomena.
- → Many observed frequency distributions approximate the well-documented **normal curve**, an important theoretical curve noted for its symmetrical bell-shaped form.
- → Characteristics that are the sum of many independent processes frequently follow normal distributions. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution.
- → The normal curve is defined in terms of standard deviation and mean.
- → The normal curve can be used to obtain answers to a wide variety of questions.



Properties of the Normal Curve:

Important properties of the normal curve are:

- The **normal curve** is a theoretical curve defined for a continuous variable.
- The normal curve is symmetrical, its lower half is the mirror image of its upper half.
- It is in bell-shaped form

- The normal curve peaks above a point midway along the horizontal spread and then tapers off gradually in either direction from the peak.
- The curve approaches the x-axis, but it never touches, and it extends farther away from the mean.
- The values of the mean, median and mode, located at a point midway along the horizontal spread, are the same for the normal curve.
- The total area under the curve should be equal to 1.
- The normal distribution curve must have only one peak. (i.e., unimodal)

Different Normal Curves

→ When using the normal curve, two bits of information are indispensable: values for the mean and the standard deviation

→ Various types of normal curves are produced by an arbitrary change in the value of either the mean (μ) or the standard deviation (σ)

→ Every normal curve can be interpreted in exactly the same way *once any distance from the mean is expressed in standard deviation units*

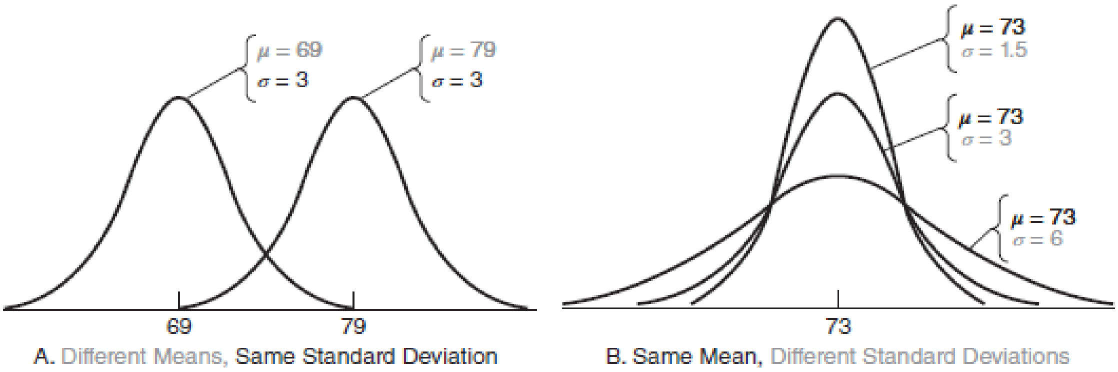


FIGURE 5.3
Different normal curves.

Z Scores

→ A unit-free, standardized score that indicates how many standard deviations a score is above or below the mean of its distribution is called Z Score

→ To obtain a *z* score, express any original score, whether measured in inches, milliseconds, dollars, IQ points, etc., as a deviation from its mean (by subtracting its mean) and then split this deviation into standard deviation units (by dividing by its standard deviation), that is,

z SCORE

$$z = \frac{X - \mu}{\sigma}$$

where *X* is the original score and *μ* and σ are the mean and the standard deviation, respectively, for the normal distribution of the original scores

→ A z score consists of two parts:

1. a positive or negative sign indicating whether it's above or below the mean; and
2. a number indicating the size of its deviation from the mean in standard deviation units.

→ Example: A *z* score of 2.00 always signifies that the original score is exactly two standard deviations above its mean. Similarly, a *z* score of –1.27 signifies that the original score is exactly 1.27 standard deviations below its mean. A *z* score of 0 signifies that the original score coincides with the mean.

→ Problem: Express each of the following scores as a *z* score:

(a) Margaret's IQ of 135, given a mean of 100 and a standard deviation of 15

(b) a score of 470 on the SAT math test, given a mean of 500 and a standard deviation of 100

(c) a daily production of 2100 loaves of bread by a bakery, given a mean of 2180 and a standard deviation of 50

(d) Sam's height of 69 inches, given a mean of 69 and a standard deviation of 3

(e) a thermometer-reading error of –3 degrees, given a mean of 0 degrees and a standard deviation of 2 degrees

Answers:

(a) z = (135-100)/15= 2.33

(b) z = (470-500)/100= 0.30

(c) z = (2100-2180)/50= -1.60

(d) z = (69-69)/3= 0.00

(e) z = (-3-0)/2= -1.50

STANDARD NORMAL CURVE

→ If the original distribution approximates a normal curve, then the shift to standard or *z* scores will always produce a new distribution that approximates the **standard normal curve**.

→ This is the one normal curve for which a table is actually available.

→ The standard normal curve always has a mean of 0 and a standard deviation of 1.

→ Although there is infinite number of different normal curves, each with its own mean and standard deviation, there is only one standard normal curve, with a mean of 0 and a standard deviation of 1.

→ Converting all original observations into *z* scores leaves the normal shape intact but not the units of measurement.

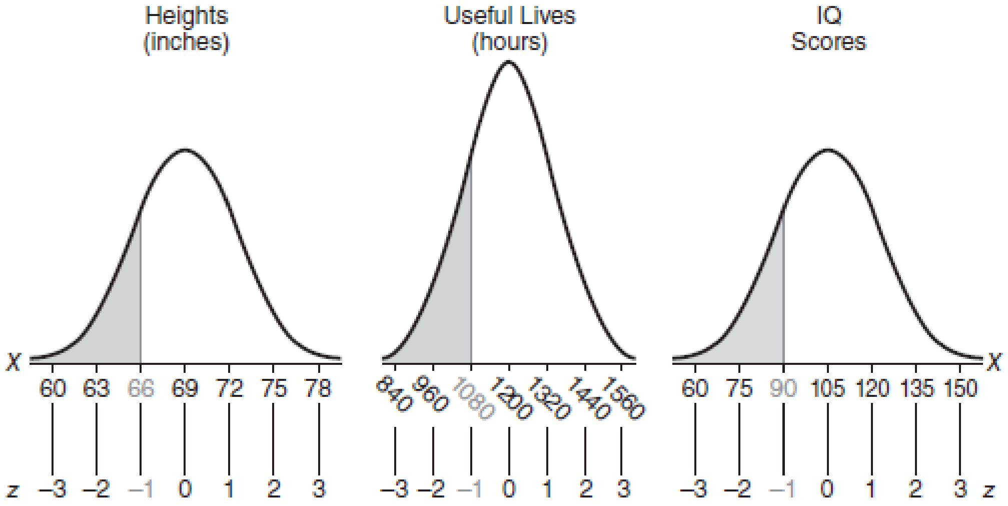


FIGURE 5.4
Converting three normal curves to the standard normal curve.

Standard Normal Table (Z Table)

→ The standard normal table consists of columns of *z* scores coordinated with columns of proportions.

→ In a typical problem, access to the table is gained through a *z* score, such as –1.00, and the answer is read as a proportion

→ Table columns are arranged in sets of three, designated as A, B, and C in the legend at the top of the table. When using the top legend, all entries refer to the upper half of the standard normal curve.

→ The entries in column A are *z* scores, beginning with 0.00 and ending with 4.00. Given a *z* score of zero, column B indicates the proportion of area between the mean and the *z* score, and column C indicates the proportion of area beyond the *z* score, in the upper tail of the standard normal curve.

→ Because of the symmetry of the normal curve, the entries in table also can refer to the lower half of the normal curve. Now the columns are designated as A′, B′, and C′ in the legend at the bottom of the table. When using the bottom legend, all entries refer to the lower half of the standard normal curve.

→ The nonzero entries in column A′ are negative *z* scores, beginning with 0.01 and ending with 4.00.

→ Column B′ indicates the proportion of area between the mean and the negative *z* score, and column C′ indicates the proportion of area beyond the negative *z* score, in the lower tail of the standard normal curve.

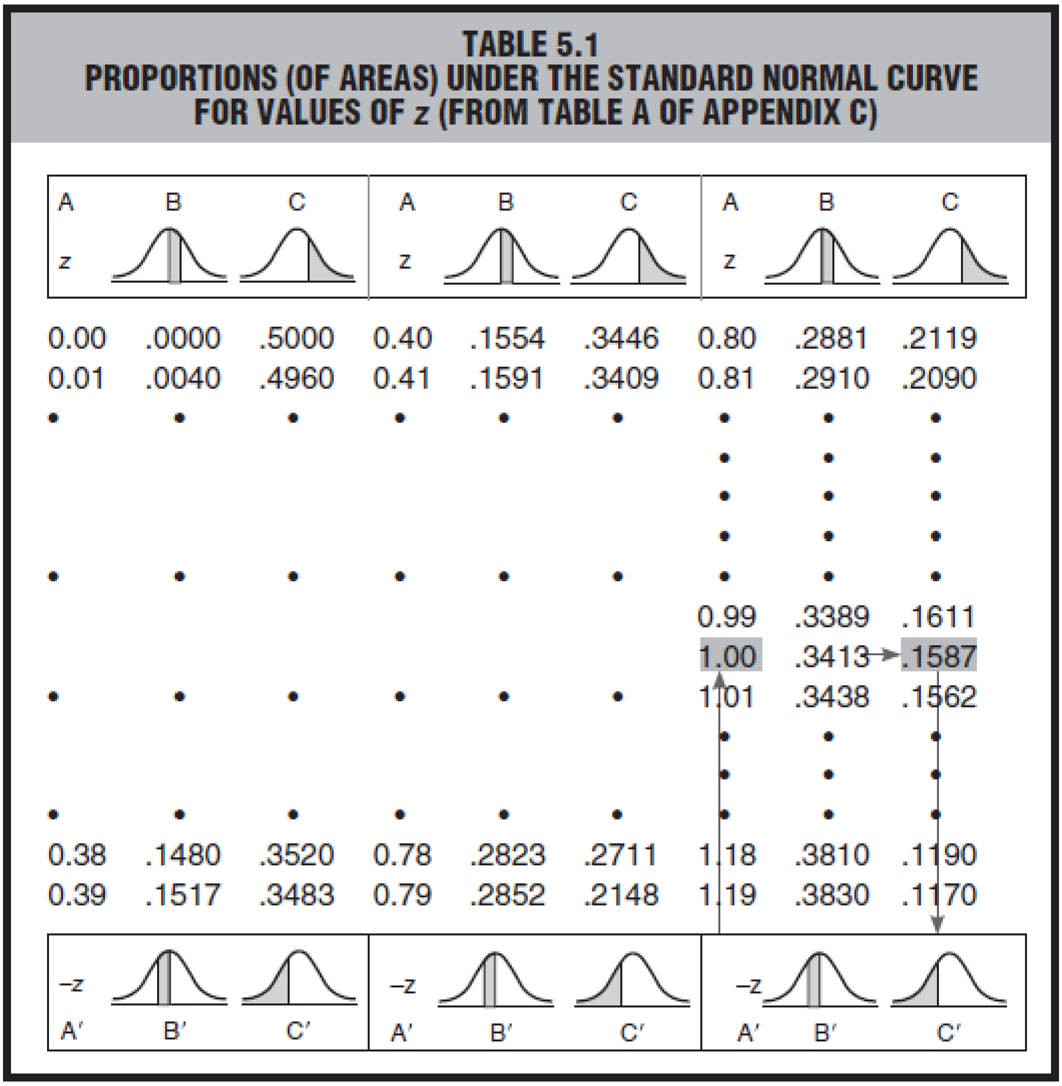


TABLE 5.1
PROPORTIONS (OF AREAS) UNDER THE STANDARD NORMAL CURVE FOR VALUES OF *z* (FROM TABLE A OF APPENDIX C)

A z	B	C	A z	B	C	A z	B	C
0.00	.0000	.5000	0.40	.1554	.3446	0.80	.2881	.2119
0.01	.0040	.4960	0.41	.1591	.3409	0.81	.2910	.2090
•	•	•	•	•	•	•	•	•
						•	•	•
						•	•	•
						•	•	•
•	•	•	•	•	•	•	•	•
						0.99	.3389	.1611
						1.00	.3413	.1587
•	•	•	•	•	•	1.01	.3438	.1562
						•	•	•
						•	•	•
•	•	•	•	•	•	•	•	•
0.38	.1480	.3520	0.78	.2823	.2711	1.18	.3810	.1190
0.39	.1517	.3483	0.79	.2852	.2148	1.19	.3830	.1170
–z A′	B′	C′	–z A′	B′	C′	–z A′	B′	C′

Normal Curve Problems

→ There are two general types of normal curve problems:

(1) **Finding proportions:** these problems require finding the unknown proportion (of area) associated with some score or pair of scores and

(2) **Finding scores:** these problems require finding the unknown score or scores associated with some area.

→ Answers to the first type of problem usually require converting original scores into *z* scores and answers to the second type of problem usually require translating a *z* score back into an original score.

→ Rough graphs of normal curves can be used an aid to visualizing the solution. Only after thinking through to a solution, do any calculations and consult the normal tables.

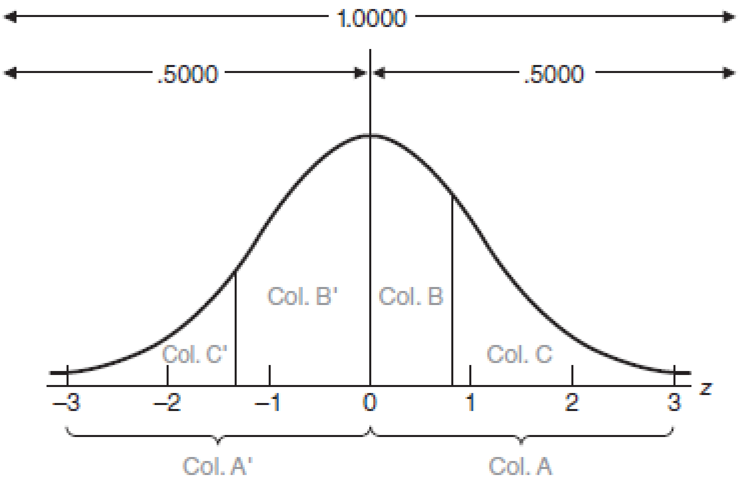


Fig: Interpretation of standard normal table

→ When using the standard normal table, it is important to remember that

- For any *z* score, the corresponding proportions in columns B and C (or columns B′ and C′) always sum to .5000.
- Similarly, the total area under the normal curve always equals 1.0000, the sum of the proportions in the lower and upper halves, that is, .5000 + .5000.
- Finally, although a *z* score can be either positive or negative, the proportions of area under the curve are always positive or zero but *never* negative

Finding Proportions

→ In these Normal curve problems, standard normal table (table A) must be consulted to find the unknown proportion (of area) associated with some known score or pair of known scores.

Finding Proportions for *One* Score

→ Step-by-step procedure:

1. Sketch a normal curve and shade in the target area
2. Plan solution according to the normal table.
3. Convert *X* to *z* using formula, $z = \frac{X - \mu}{\sigma}$
4. Find the target area.

→ **Example:** to find the proportion of all persons who are shorter than exactly 66 inches, given that the distribution of heights approximates a normal curve with a mean of 69 inches and a standard deviation of 3 inches.

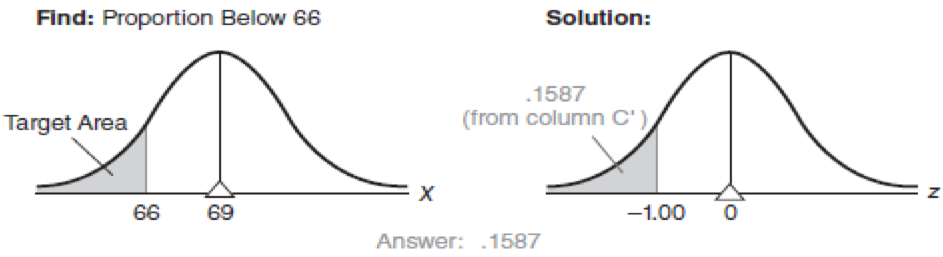


FIGURE 5.6
Finding proportions.

Finding Proportions *between* Two Scores

→ Step-by-step procedure:

1. Sketch a normal curve and shade in the target area
2. Plan solution according to the normal table.
3. Convert X to z using formula, $z = \frac{X - \mu}{\sigma}$
4. Find the target area.

→ Example: Assume that, when not interrupted artificially, the gestation periods for human foetuses approximate a normal curve with a mean of 270 days (9 months) and a standard deviation of 15 days. What proportion of gestation periods will be between 245 and 255 days?

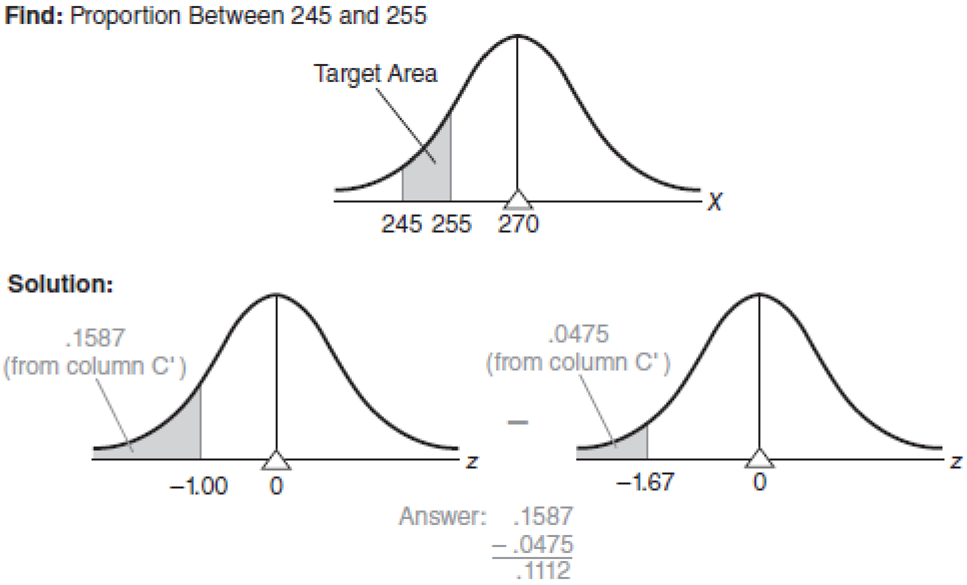


FIGURE 5.7
Finding proportions.

Finding Proportions *beyond* Two Scores

→ Step-by-step procedure:

1. Sketch a normal curve and shade in the two target areas
2. Plan your solution according to the normal table.
3. Convert X *to* z using formula, $z = \frac{X - \mu}{\sigma}$
4. Find the target area.

→ **Problem:** Assume that high school students' IQ scores approximate a normal distribution with a mean of 105 and a standard deviation of 15. What proportion of IQs are more than 30 points either above or below the mean?

Answer:

Expressing IQ scores of 135 and 75 as

$$z = \frac{135-105}{15} = \frac{30}{15} = 2.00$$

$$z = \frac{75-105}{15} = \frac{-30}{15} = -2.00$$

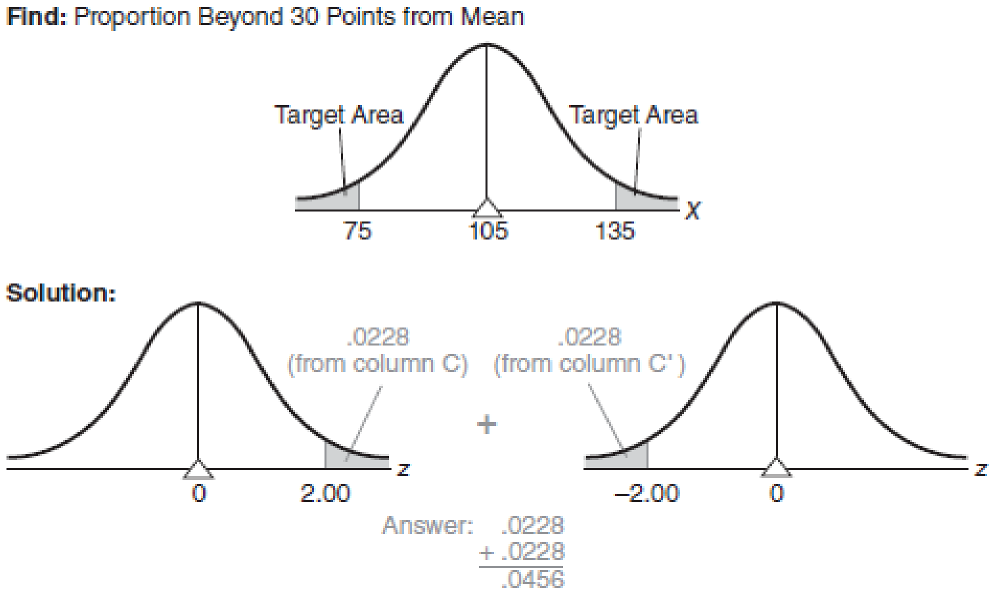


FIGURE 5.8

Finding proportions.

Finding Scores

→ In this type of normal curve problems standard normal table (table A) must be consulted to find the unknown score or scores associated with some known proportion.

→ Essentially, this type of problem requires that the use of table A by entering proportions in columns B, C, B′, or C′ and finding z scores listed in columns A or A′.

Finding One Score

→ Step-by-step procedure:

1. Sketch a normal curve and, on the correct side of the mean, draw a line representing the target score
2. Plan your solution according to the normal table.
3. Find *z*.
4. Convert *z* to the target score using formula, $\mathbf{X} = \mu + (\mathbf{z})(\sigma)$

→ Problem: Exam scores for a large psychology class approximate a normal curve with a mean of 230 and a standard deviation of 50. Furthermore, students are graded "on a curve," with only the upper 20 percent being awarded grades of A. What is the lowest score on the exam that receives an A?

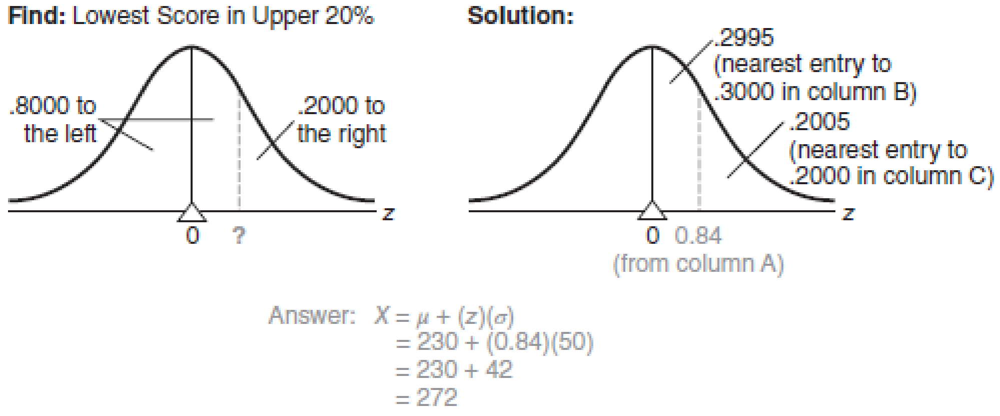


FIGURE 5.9
Finding scores.

Finding *Two* Scores

→ Step-by-step procedure:

1. Sketch a normal curve. On either side of the mean, draw two lines representing the two target scores
2. Plan your solution according to the normal table.
3. Find z.
4. Convert *z* to the target score, using formula X = μ + (z) (σ)

→ Problem: Assume that the annual rainfall in the San Francisco area approximates a normal curve with a mean of 22 inches and a standard deviation of 4 inches. What are the rainfalls for the more atypical years, defined as the driest 2.5 percent of all years and the wettest 2.5 percent of all years?

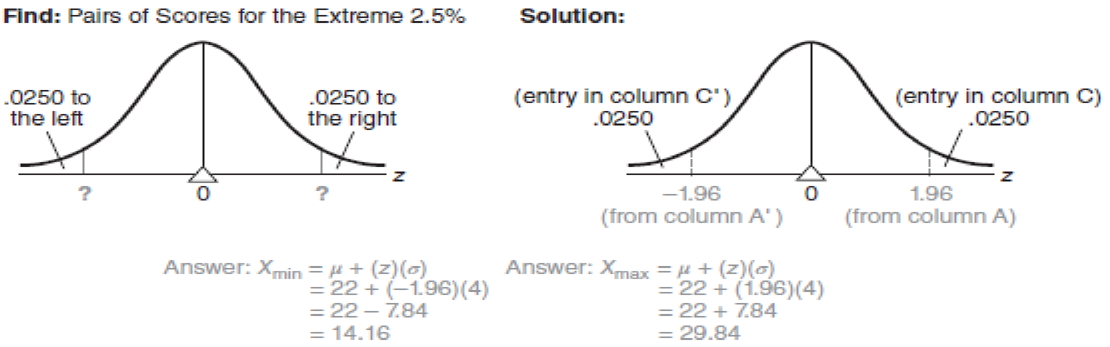
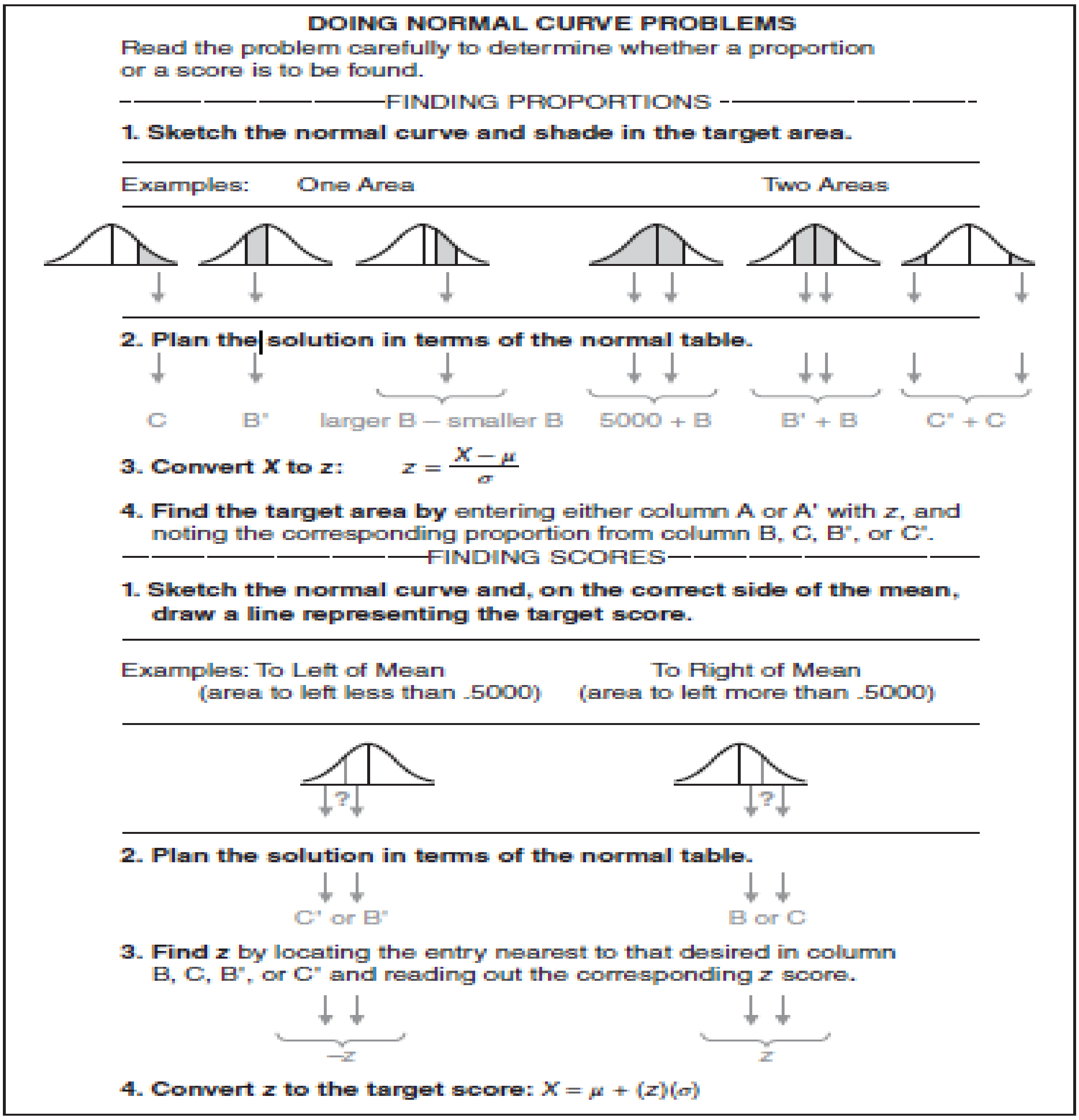


FIGURE 5.10
Finding scores.



DOING NORMAL CURVE PROBLEMS

Read the problem carefully to determine whether a proportion or a score is to be found.

FINDING PROPORTIONS

1. Sketch the normal curve and shade in the target area.

Examples: One Area Two Areas

2. Plan the solution in terms of the normal table.

C B' larger B − smaller B .5000 + B B' + B C' + C

3. Convert X to z: $z = \frac{X - \mu}{\sigma}$

4. Find the target area by entering either column A or A' with z, and noting the corresponding proportion from column B, C, B', or C'.

FINDING SCORES

1. Sketch the normal curve and, on the correct side of the mean, draw a line representing the target score.

Examples: To Left of Mean (area to left less than .5000) To Right of Mean (area to left more than .5000)

2. Plan the solution in terms of the normal table.

C' or B' B or C

3. Find z by locating the entry nearest to that desired in column B, C, B', or C' and reading out the corresponding z score.

$-z$ z

4. Convert z to the target score: $X = \mu + (z)(\sigma)$

More About Z Scores

Z Scores for Non-normal Distributions

→ z scores are not limited to normal distributions.

→ Non-normal distributions also can be transformed into sets of unit-free, standardized z scores.

→ In this case, the standard normal table cannot be consulted, since the shape of the distribution of z scores is the same as that for the original non-normal distribution.

→ Regardless of the shape of the distribution, the shift to z scores always produces a distribution of standard scores with a mean of 0 and a standard deviation of 1.

→ Z scores can provide efficient descriptions of relative performance on one or more tests.

→ The use of *z* scores can help to identify a person's relative strengths and weaknesses on several different tests.

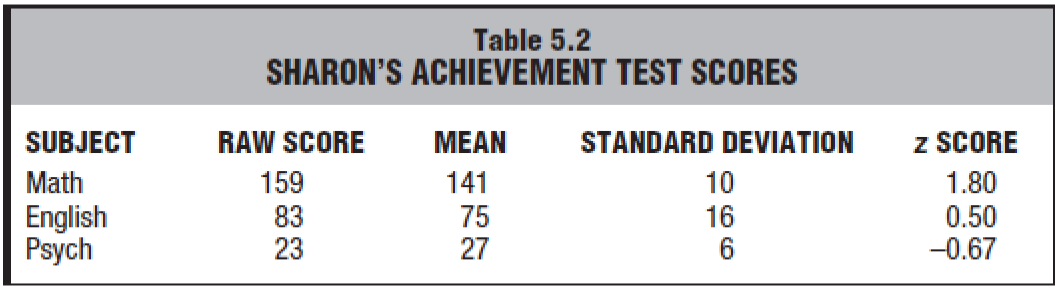


Table 5.2
SHARON'S ACHIEVEMENT TEST SCORES

SUBJECT	RAW SCORE	MEAN	STANDARD DEVIATION	*z* SCORE
Math	159	141	10	1.80
English	83	75	16	0.50
Psych	23	27	6	–0.67

→ For example, above table shows Sharon's scores on college achievement tests in three different subjects. The evaluation of her test performance is greatly facilitated by converting her raw scores into the *z* scores listed in the final column of above table. A glance at the *z* scores suggests that although she did relatively well on the math test, her performance on the English test was only slightly above average, as indicated by a *z* score of 0.50, and her performance on the psychology test was slightly below average, as indicated by a *z* score of –0.67.

Standard Score

→ Any unit-free scores expressed relative to a known mean and a known standard deviation is called standard score.

→ Although *z* scores qualify as standard scores because they are unit-free and expressed relative to a known mean of 0 and a known standard deviation of 1, other scores also qualify as standard scores.

Transformed Standard Scores

→ z scores can be changed to **transformed standard scores,** other types of unit-free standard scores that lack negative signs and decimal points.

→ These transformations change neither the shape of the original distribution nor the relative standing of any test score within the distribution.

→ For example, a test score located one standard deviation below the mean might be reported not as a z score of –1.00 but as a T score of 40 in a distribution of T scores with a mean of 50 and a standard deviation of 10.

→ Following figure shows the values of some of the more common types of transformed standard scores relative to the various portions of the area under the normal curve.

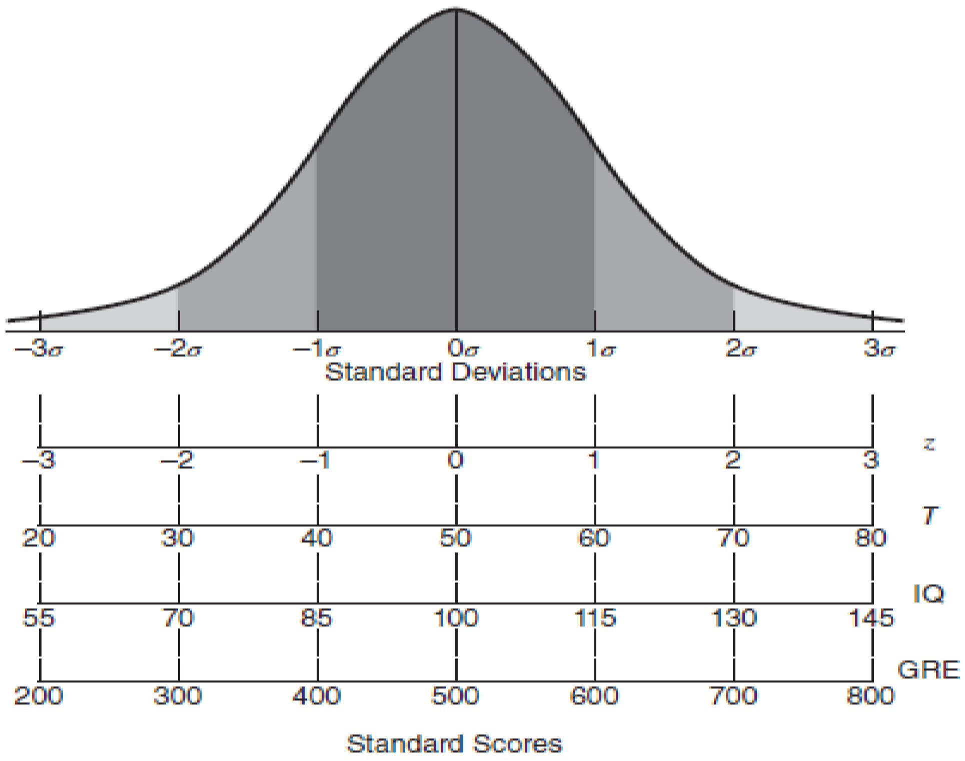


FIGURE 5.11

Common transformed standard scores associated with normal curves.

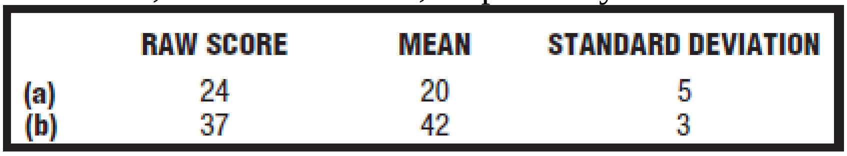
Converting to Transformed Standard Scores

→ Following formula can be used to convert any original standard score, z, into a transformed standard score, z′, having a distribution with any desired mean and standard deviation.

z' = desired mean + (z) (desired standard deviation)

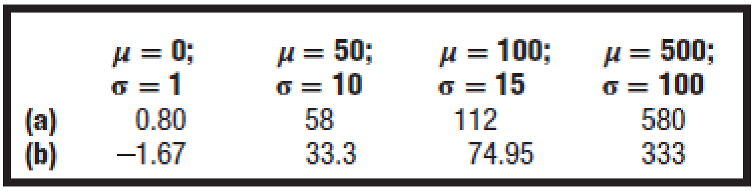
where z' (called *z prime*) is the transformed standard score and z is the original standard score.

→ Problem: Assume that each of the raw scores listed originates from a distribution with the specified mean and standard deviation. After converting each raw score into a z score, transform each z score into a series of new standard scores with means and standard deviations of 50 and 10, 100 and 15, and 500 and 100, respectively.



	RAW SCORE	MEAN	STANDARD DEVIATION
(a)	24	20	5
(b)	37	42	3

Answers:



	$\mu = 0$; $\sigma = 1$	$\mu = 50$; $\sigma = 10$	$\mu = 100$; $\sigma = 15$	$\mu = 500$; $\sigma = 100$
(a)	0.80	58	112	580
(b)	−1.67	33.3	74.95	333

Correlation

→ Two variables are related if pairs of scores show an orderliness that can be depicted graphically with a **scatter plot** and numerically with a **correlation coefficient.**

→ The data in following table represent a very simple observational study with two dependent variables.

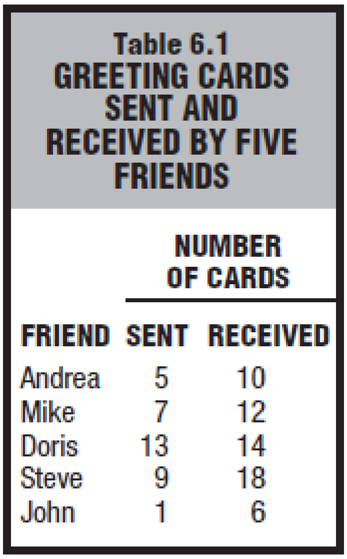


Table 6.1
GREETING CARDS SENT AND RECEIVED BY FIVE FRIENDS

	NUMBER OF CARDS	
FRIEND	SENT	RECEIVED
Andrea	5	10
Mike	7	12
Doris	13	14
Steve	9	18
John	1	6

Three Types of Relationships (Types of correlation)

- Positive Relationship
- Negative Relationship
- Little or No Relationship

Positive Relationship

→ Two variables are *positively* related if pairs of scores tend to occupy similar relative positions (relatively low values are paired with relatively low values, and relatively high values are paired with relatively high values,) in their respective distributions.

→ Example: (Height, Weight)
(Temperature, Ice cream sales)

Negative Relationship

→ Two variables are *negatively* related if pairs of scores tend to occupy dissimilar relative positions (relatively low values are paired with relatively high values, and relatively high values are paired with relatively low values,) in their respective distributions.

→ Example: (Exercise, Body Fat)
(Watching Movies, Exam scores)

Little or No Relationship

→ No regularity is apparent among the pairs of scores

→ Example: (Shoe Size, Movies Watched)
(Coffee Consumption, Intelligence)

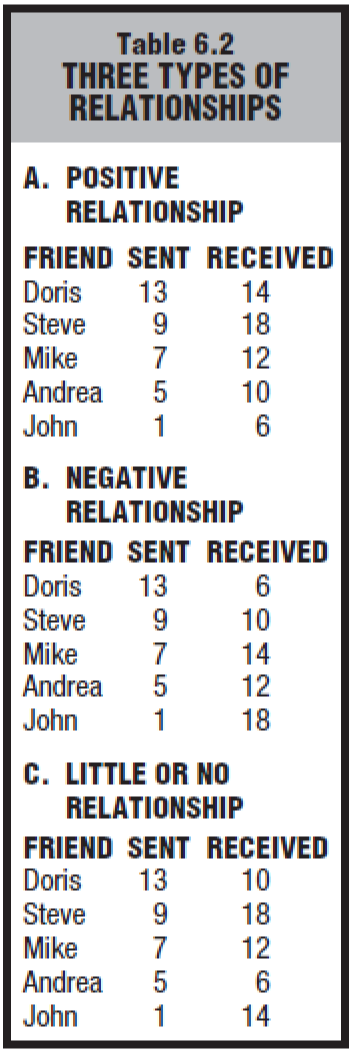


Table 6.2
THREE TYPES OF RELATIONSHIPS

A. POSITIVE RELATIONSHIP

FRIEND	SENT	RECEIVED
Doris	13	14
Steve	9	18
Mike	7	12
Andrea	5	10
John	1	6

B. NEGATIVE RELATIONSHIP

FRIEND	SENT	RECEIVED
Doris	13	6
Steve	9	10
Mike	7	14
Andrea	5	12
John	1	18

C. LITTLE OR NO RELATIONSHIP

FRIEND	SENT	RECEIVED
Doris	13	10
Steve	9	18
Mike	7	12
Andrea	5	6
John	1	14

Describing relationship between pairs of variables

→ There are two more efficient and exact statistical techniques for describing relationship between two variables, namely, a special graph known as a *scatter plot* and a measure known as a *correlation coefficient.*

Scatter Plots

→ A **scatter plot** is a graph containing a cluster of dots that represents all pairs of scores.

→ We can use any dot cluster as a preview of a fully measured relationship.

Construction

→ To construct a scatter plot scale each of the two variables along the horizontal (X) and vertical (Y) axes, and use each pair of scores to locate a dot within he scatter plot.

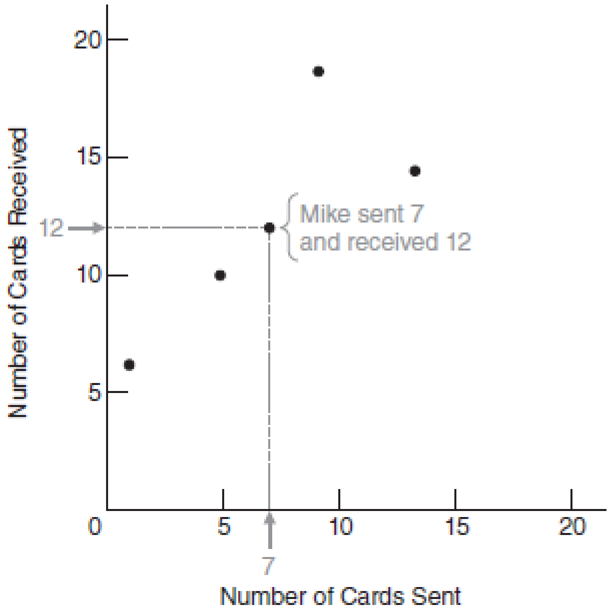


FIGURE 6.1
Scatterplot for greeting card exchange.

Table 6.1
GREETING CARDS SENT AND RECEIVED BY FIVE FRIENDS

	NUMBER OF CARDS	
FRIEND	**SENT**	**RECEIVED**
Andrea	5	10
Mike	7	12
Doris	13	14
Steve	9	18
John	1	6

Categorizing relationship using scatter plot

(Positive, Negative, or Little or No Relationship?)

→ A dot cluster that has a slope from the lower left to the upper right reflects a positive relationship. Small values of one variable are paired with small values of the other variable, and large values are paired with large values.

→ Example: In panel A of below figure, short people tend to be light, and tall people tend to be heavy.

→ A dot cluster that has a slope from the upper left to the lower right reflects a negative relationship. Small values of one variable tend to be paired with large values of the other variable, and vice versa.

→ Example: In panel B of below figure, people who have smoked heavily for few years or not at all tend to have longer lives, and people who have smoked heavily for many years tend to have shorter lives

→ A dot cluster that lacks any apparent slope reflects little or no relationship. Small values of one variable are just as likely to be paired with small, medium, or large values of the other variable.

→ Example: In panel C of below figure, notice that the dots are strewn about in an irregular shotgun fashion, suggesting that there is little or no relationship between the height of young adults and their life expectancies.

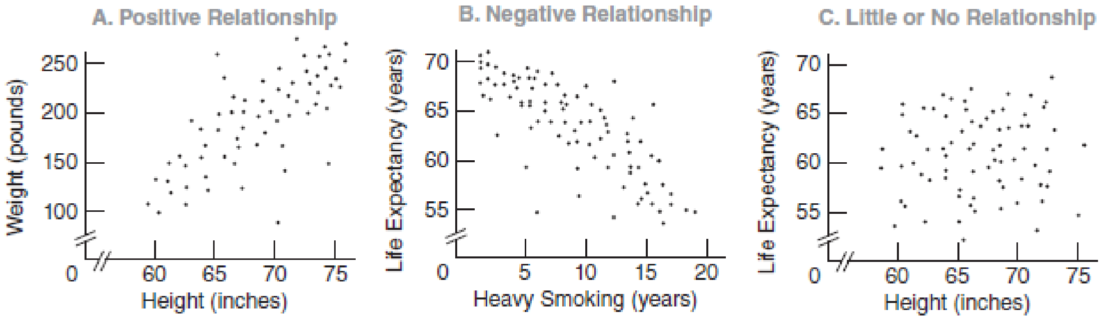


FIGURE 6.2
Three types of relationships.

Perfect Relationship

→ A dot cluster that equals (rather than merely approximates) a straight line reflects a perfect relationship between two variables.

Linear Relationship

→ A relationship that can be described best with a straight line.

Curvilinear Relationship

→ A relationship that can be described best with a curved line.

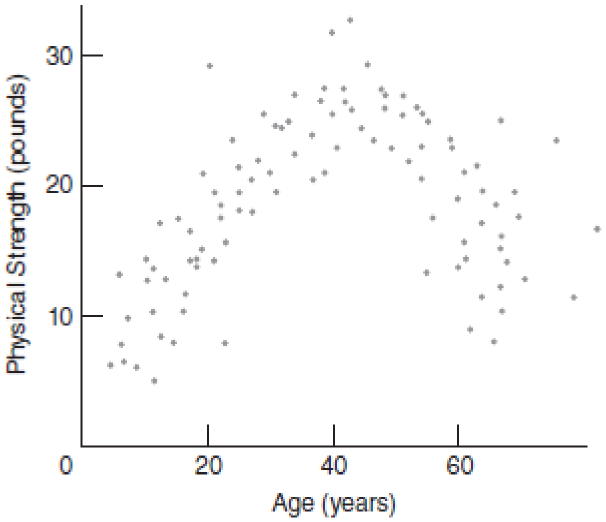


FIGURE 6.4
Curvilinear relationship.

A Correlation Coefficient For Quantitative Data : *r*

→ A **correlation coefficient** is a number between –1 and 1 that describes the relationship between pairs of variables.

→ The type of correlation coefficient, designated as *r*, that *describes the linear relationship between pairs of variables for quantitative data* is called the **Pearson correlation coefficient**, ***r*****,** can equal any value between –1.00 and +1.00.

→ Furthermore, the following two properties apply:

1. The sign of **r** indicates the type of linear relationship, whether positive or negative.
2. The numerical value of **r**, without regard to sign, indicates the strength of the linear relationship.

→ A number with a plus sign (or no sign) indicates a positive relationship, and a number with a minus sign indicates a negative relationship. For example, an *r* with a plus sign describes the positive relationship between height and weight, and an *r* with a minus sign describes the negative relationship between heavy smoking and life expectancy.

→ The more closely a value of *r* approaches either –1.00 or +1.00, the stronger (more regular) the relationship. Conversely, the more closely the value of *r* approaches 0, the weaker (less regular) the relationship.

→ For example, an *r* of –.90 indicates a stronger relationship than does an *r* of –.70, and an *r* of –.70 indicates a stronger relationship than does an *r* of .50

→ A correlation coefficient, regardless of size, never provides information about whether an observed relationship reflects a simple cause-effect relationship or some more complex state of affairs.

Computation Formula for Correlation Coefficient

→ Correlation Coefficient can be calculated by using following Computation Formula

CORRELATION COEFFICIENT (COMPUTATION FORMULA)

$$r = \frac{SP_{xy}}{\sqrt{SS_x SS_y}}$$

where the two sum of squares terms in the denominator are defined as

$$SS_x = \Sigma\left(X - \bar{X}\right)^2 = \Sigma X^2 - \frac{\left(\Sigma X\right)^2}{n}$$

$$SS_y = \Sigma\left(Y - \bar{Y}\right)^2 = \Sigma Y^2 - \frac{\left(\Sigma Y\right)^2}{n}$$

and the sum of the products term in the numerator, SP_{xy}, is defined as

$$SP_{xy} = \Sigma\left(X - \bar{X}\right)\left(Y - \bar{Y}\right) = \Sigma XY - \frac{\left(\Sigma X\right)\left(\Sigma Y\right)}{n}$$

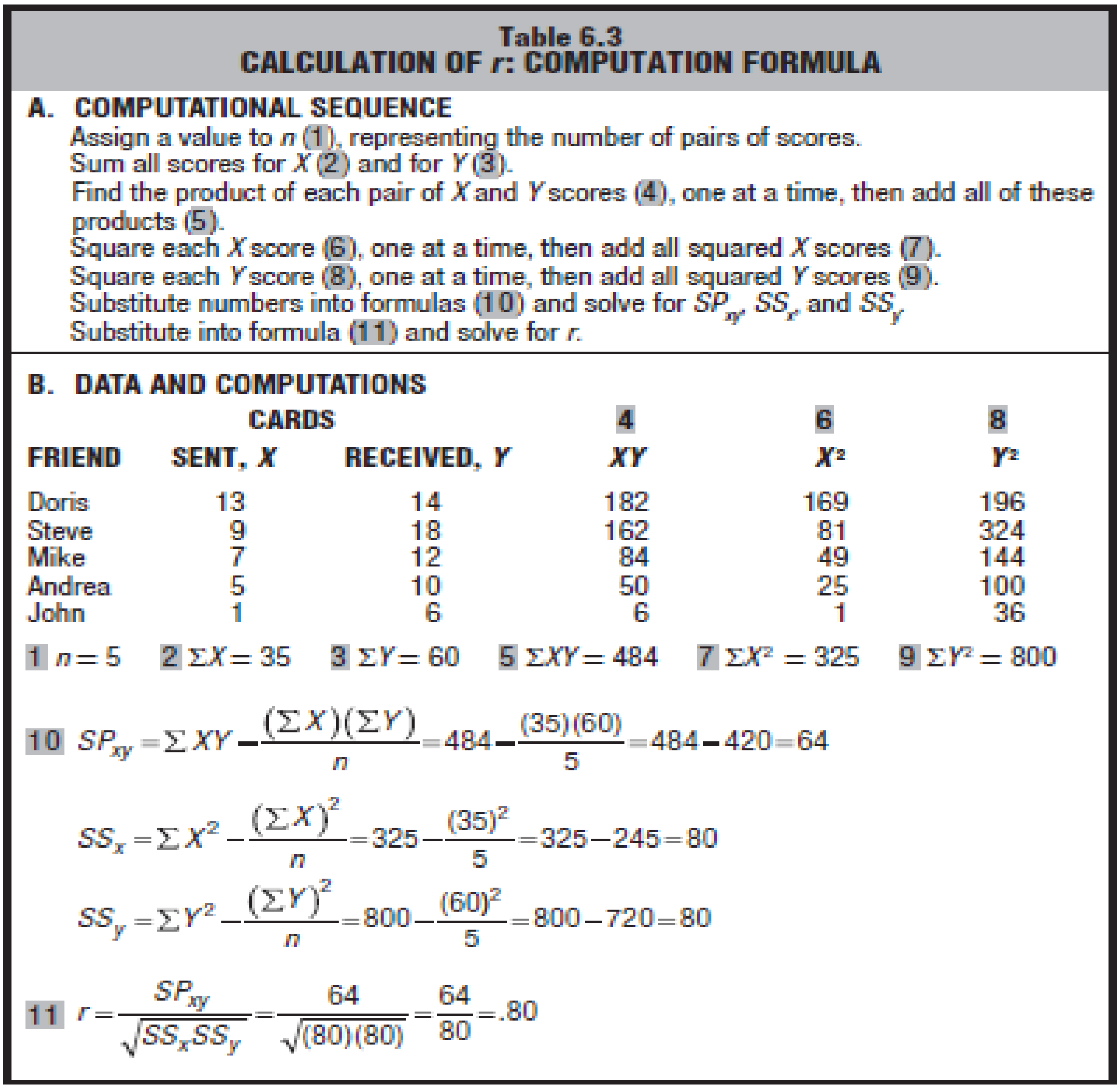


Table 6.3
CALCULATION OF *r*: COMPUTATION FORMULA

A. COMPUTATIONAL SEQUENCE

Assign a value to *n* (1), representing the number of pairs of scores.
Sum all scores for *X* (2) and for *Y* (3).
Find the product of each pair of *X* and *Y* scores (4), one at a time, then add all of these products (5).
Square each *X* score (6), one at a time, then add all squared *X* scores (7).
Square each *Y* score (8), one at a time, then add all squared *Y* scores (9).
Substitute numbers into formulas (10) and solve for SP_{xy}, SS_x, and SS_y
Substitute into formula (11) and solve for *r*.

B. DATA AND COMPUTATIONS

	CARDS		4	6	8
FRIEND	**SENT, *X***	**RECEIVED, *Y***	***XY***	***X*²**	***Y*²**
Doris	13	14	182	169	196
Steve	9	18	162	81	324
Mike	7	12	84	49	144
Andrea	5	10	50	25	100
John	1	6	6	1	36

1 $n = 5$ 2 $\Sigma X = 35$ 3 $\Sigma Y = 60$ 5 $\Sigma XY = 484$ 7 $\Sigma X^2 = 325$ 9 $\Sigma Y^2 = 800$

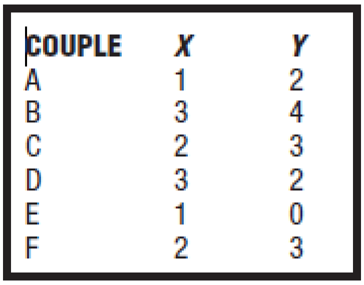
10 $$SP_{xy} = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n} = 484 - \frac{(35)(60)}{5} = 484 - 420 = 64$$

$$SS_x = \Sigma X^2 - \frac{(\Sigma X)^2}{n} = 325 - \frac{(35)^2}{5} = 325 - 245 = 80$$

$$SS_y = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = 800 - \frac{(60)^2}{5} = 800 - 720 = 80$$

11 $$r = \frac{SP_{xy}}{\sqrt{SS_x SS_y}} = \frac{64}{\sqrt{(80)(80)}} = \frac{64}{80} = .80$$

Problem: Couples who attend a clinic for first pregnancies are asked to estimate (independently of each other) the ideal number of children. Given that *X* and *Y* represent the estimates of females and males, respectively, the results are as follows:



COUPLE	*X*	*Y*
A	1	2
B	3	4
C	2	3
D	3	2
E	1	0
F	2	3

Calculate a value for *r*, using the computation formula

Answer:

$$r = \frac{4}{\sqrt{(4)(9.33)}} = .65$$

Regression

→ A regression is a statistical technique that relates a dependent variable to one or more independent (explanatory) variables.

→ A regression model is able to show whether changes observed in the dependent variable are associated with changes in one or more of the explanatory variables.

→ Regression captures the correlation between variables observed in a data set, and quantifies whether those correlations are statistically significant or not.

Regression Line

→ A regression line is a line that best describes the behaviour of a set of data. In other words, it's a line that best fits the trend of a given data.

→ The purpose of the line is to describe the interrelation of a dependent variable (Y variable) with one or many independent variables (X variable).

→ By using the equation obtained from the regression line an analyst can forecast future behaviours of the dependent variable by inputting different values for the independent ones.

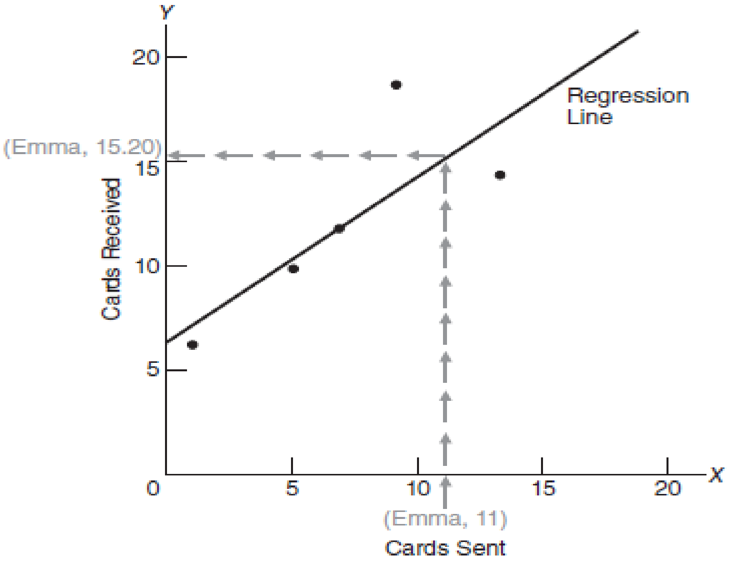


FIGURE 7.2
Prediction of 15.20 for Emma (using the regression line).

Types of regression

The two basic types of regression are

- **Simple linear regression**: Simple linear regression uses one independent variable to explain or predict the outcome of the dependent variable Y
- **Multiple linear regression:** Multiple linear regressions use two or more independent variables to predict the outcome

Predictive Errors

→ Prediction error refers to the difference between the predicted values made by some model and the actual values.

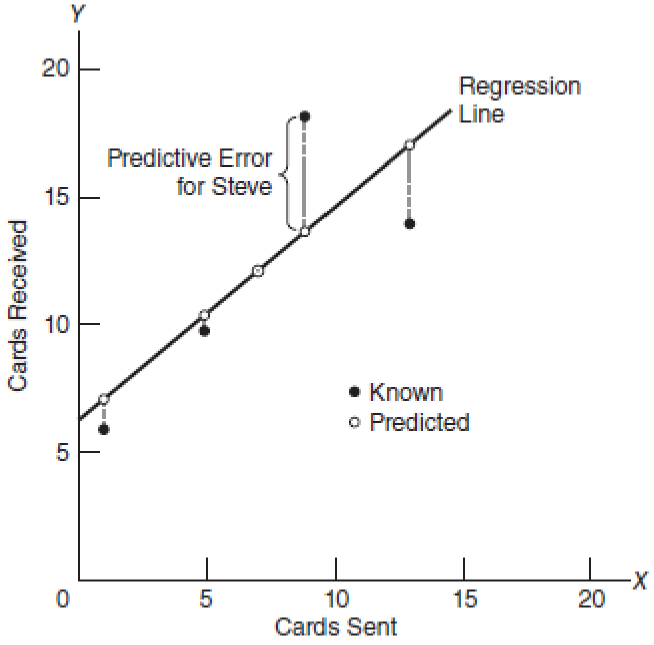


FIGURE 7.3
Predictive errors.

Least Squares Regression Line

→ The placement of the regression line minimizes not the total predictive error but the total squared predictive error, that is, the total for all squared predictive errors. When located in this fashion, the regression line is often referred to as the least squares regression line.

→ The Least Squares Regression Line is the line that minimizes the sum of the residuals squared. The residual is the vertical distance between the observed point and the predicted point, and it is calculated by subtracting ˆy from y.

→ Least Squares Regression Equation: an equation pinpoints the exact least squares regression line for any scatter plot. Most generally, this equation reads:

$$Y' = bX + a$$

where Y' represents the predicted value
X represents the known
b and a represent numbers calculated from the original correlation analysis, described by

$$b = r\sqrt{\frac{SS_y}{SS_x}}$$

$$a = \overline{Y} - b\overline{X}$$

→ The regression equation can be used to predict the Y' value for given X value by simply substituting X value in equation.

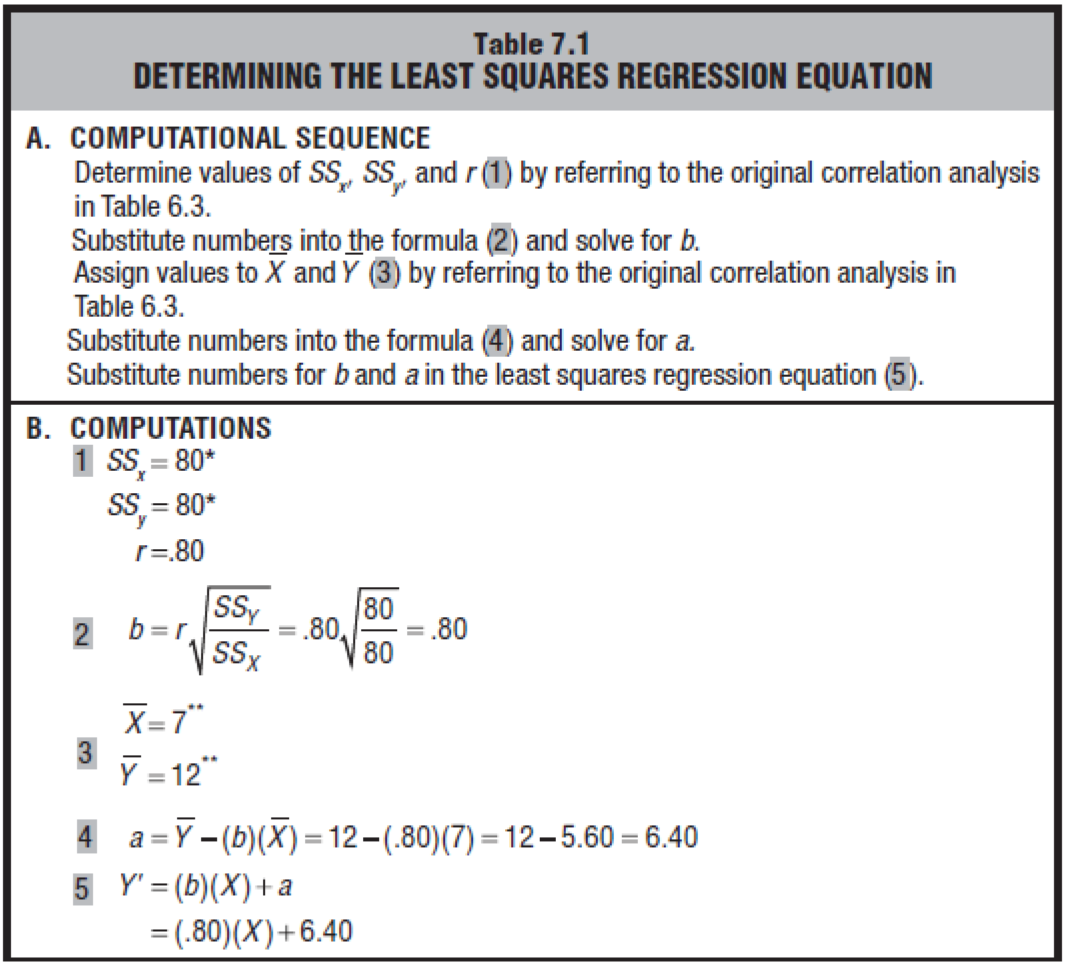


Table 7.1
DETERMINING THE LEAST SQUARES REGRESSION EQUATION

A. COMPUTATIONAL SEQUENCE

Determine values of SS_x, SS_y, and r (1) by referring to the original correlation analysis in Table 6.3.

Substitute numbers into the formula (2) and solve for b.

Assign values to X and Y (3) by referring to the original correlation analysis in Table 6.3.

Substitute numbers into the formula (4) and solve for a.

Substitute numbers for b and a in the least squares regression equation (5).

B. COMPUTATIONS

1 $SS_x = 80^*$

$SS_y = 80^*$

$r = .80$

2 $b = r\sqrt{\dfrac{SS_Y}{SS_X}} = .80\sqrt{\dfrac{80}{80}} = .80$

3 $\overline{X} = 7^{**}$

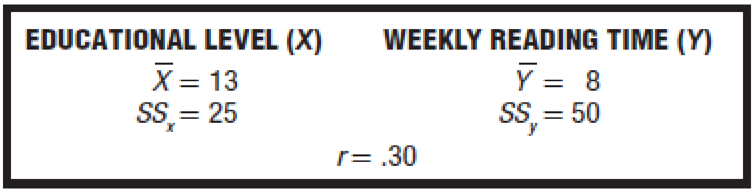
$\overline{Y} = 12^{**}$

4 $a = \overline{Y} - (b)(\overline{X}) = 12 - (.80)(7) = 12 - 5.60 = 6.40$

5 $Y' = (b)(X) + a$

$= (.80)(X) + 6.40$

→ **Problem:** Assume that an *r* of .30 describes the relationship between educational level (highest grade completed) and estimated number of hours spent reading each week. More specifically:



EDUCATIONAL LEVEL (X)	WEEKLY READING TIME (Y)
$\overline{X} = 13$	$\overline{Y} = 8$
$SS_x = 25$	$SS_y = 50$
$r = .30$	

(a) Determine the least squares equation for predicting weekly reading time from educational level.

(b) Faith's education level is 15. What is her predicted reading time?

(c) Keegan's educational level is 11. What is his predicted reading time?

Answer:

(a) $b = \sqrt{\frac{50}{25}}(.30) = .42;\ a = 8 - (.42)(13) = 2.54$

(b) $Y' = (.42)(15) + 2.54 = 8.84$

(c) $Y' = (.42)(11) + 2.54 = 7.16$

Standard Error Of Estimate($S_{y/x}$)

- → The standard error of the estimate is a measure of the accuracy of predictions.
- → The regression line is the line that minimizes the sum of squared deviations of prediction (also called the sum of squares error), and the standard error of the estimate is the square root of the average squared deviation.
- → The standard error of estimate represents a special kind of standard deviation that reflects the magnitude of predictive error.
- → It is a rough measure of the average amount of predictive error—that is, as a rough measure of the average amount by which known *Y* values deviate from their predicted *Y* values.
- → This estimate of predictive error complies with the general format for any sample standard deviation, that is, the square root of a sum of squares term divided by its degrees of freedom.

STANDARD ERROR OF ESTIMATE (DEFINITION FORMULA)

$$s_{y|x} = \sqrt{\frac{SS_{y|x}}{n-2}} = \sqrt{\frac{\sum(Y-Y')^2}{n-2}}$$

- → We can also estimate the overall predictive error by dealing directly with predictive errors, *Y* − *Y′*, it is more efficient to use the following computation formula:

STANDARD ERROR OF ESTIMATE (COMPUTATION FORMULA)

$$s_{y|x} = \sqrt{\frac{SS_y(1-r^2)}{n-2}}$$

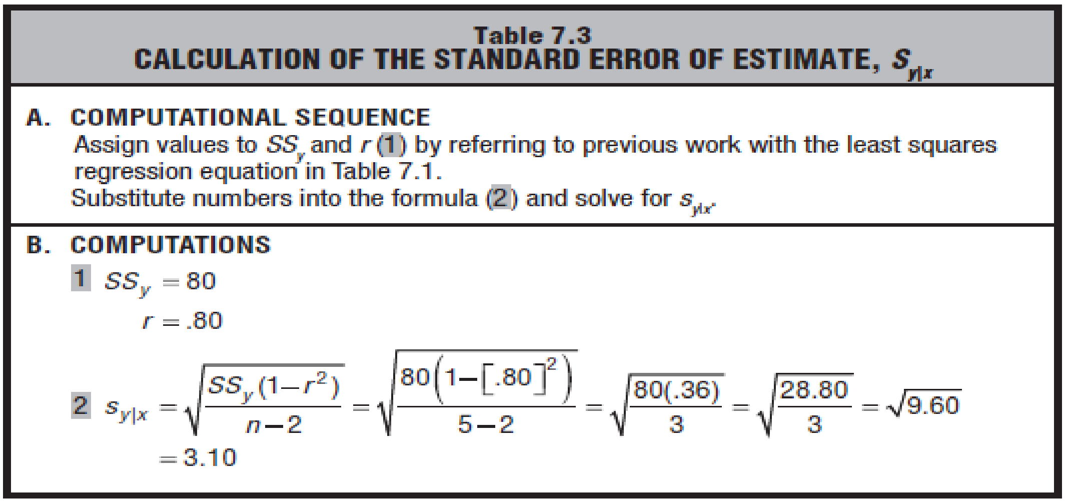


Table 7.3
CALCULATION OF THE STANDARD ERROR OF ESTIMATE, $s_{y|x}$

A. COMPUTATIONAL SEQUENCE

Assign values to SS_y and r (1) by referring to previous work with the least squares regression equation in Table 7.1.

Substitute numbers into the formula (2) and solve for $s_{y|x}$.

B. COMPUTATIONS

1 $SS_y = 80$

$r = .80$

2 $s_{y|x} = \sqrt{\frac{SS_y(1-r^2)}{n-2}} = \sqrt{\frac{80(1-[.80]^2)}{5-2}} = \sqrt{\frac{80(.36)}{3}} = \sqrt{\frac{28.80}{3}} = \sqrt{9.60}$

$= 3.10$

Interpretation of r^2

→ The squared correlation coefficient, r^2, provides a measure of predictive accuracy that supplements the *standard error of estimate*, $S_{y/x}$

→ r^2 indicates the proportion of total variability in one variable that is predictable from its relationship with the other variable**.**

→ It is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression model (the goodness of fit).

→ r-squared can take any values between 0 to 1. Although the statistical measure provides some useful insights regarding the regression model, the user should not rely only on the measure in the assessment of a statistical model.

→ In addition, it does not indicate the correctness of the regression model. Therefore, the user should always draw conclusions about the model by analyzing r-squared together with the other variables in a statistical model.

→ Expressing the equation for *r* in symbols, we have:

$$r^2 = \frac{SS_{Y'}}{SS_Y} = \frac{SS_Y - SS_{Y|X}}{SS_Y}$$

Example: Suppose

$SS_y = 80$ and $SS_{y/x} = 28.8$

then

$$SS_y - SS_{y|x} = 80 - 28.8 = 51.2$$

$$\frac{SS_y - SS_{y|x}}{SS_y} = \frac{80 - 28.8}{80} = \frac{51.2}{80} = .64$$

Multiple Regression Equations

→ Serious predictive efforts usually involve multiple regression equations composed of more than one predictor, or *X*, variable.

→ Most generally, these equations take the form:

$$Y' = b_1(X1) + b_2(X2) + b_3(X3) + a$$

Where Y' is dependent variable and X1, X2 and X3 are independent (predictor or X) variable

→ For instance, a serious effort to predict college GPA might culminate in the following equation: $Y' = .410(X1)+.005(X2\ ) + .001(X3\ ) + 1.03$ where Y' represents predicted college GPA and $X1$, $X2$, and $X3$ refer to high school GPA, IQ score, and SAT score, respectively.

→ By capitalizing on the combined predictive power of several predictor variables, these **multiple regression equations** supply more accurate predictions for *Y′* (often referred to as the *criterion variable*) than could be obtained from a simple regression equation.

→ These multiple regression equations share many common features with the simple regression equations.

Regression toward the Mean

→ **Regression toward the mean** refers to a tendency for scores, particularly extreme scores, to shrink toward the mean.

→ This tendency often appears among subsets of observations whose values are extreme and at least partly due to chance.

→ Regression toward the mean refers to the principle that, over repeated sampling periods, outliers tend to revert to the mean. High performers show disappointing results when they fail to continue delivering; strugglers show sudden improvement.

→ Regression toward the mean occurs when the correlation between two measures is imperfect, and so one data point cannot predict the next data point reliably.

→ **In other words,** when we ignore regression toward the mean, we ***overestimate*** the correlation between the two measures**.**

→ For example, because of regression toward the mean, we would expect that students who made the top five scores on the first mid exam would not make the top five scores on the second mid exam. Although all five students might score above the mean on the second mid exam, some of their scores would regress back toward the mean.

→ Example2: A military commander has two units return, one with 20% casualties and another with 50% casualties. He praises the first and berates the second. The next time, the two units return with the opposite results. From this experience, he “learns” that praise weakens performance and berating increases performance.

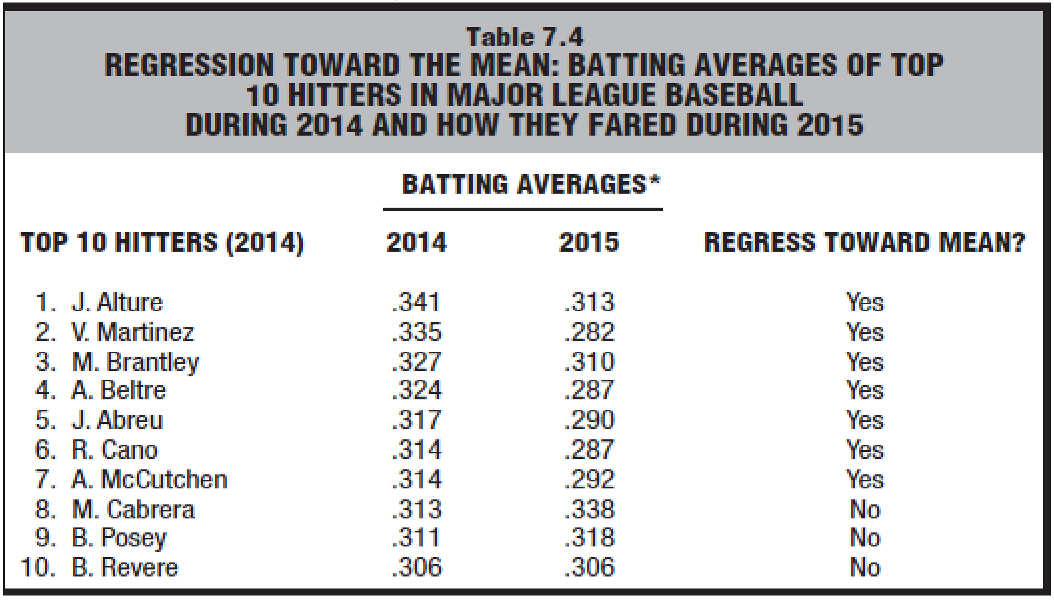


Table 7.4
REGRESSION TOWARD THE MEAN: BATTING AVERAGES OF TOP 10 HITTERS IN MAJOR LEAGUE BASEBALL DURING 2014 AND HOW THEY FARED DURING 2015

	BATTING AVERAGES*		
TOP 10 HITTERS (2014)	2014	2015	REGRESS TOWARD MEAN?
1. J. Alture	.341	.313	Yes
2. V. Martinez	.335	.282	Yes
3. M. Brantley	.327	.310	Yes
4. A. Beltre	.324	.287	Yes
5. J. Abreu	.317	.290	Yes
6. R. Cano	.314	.287	Yes
7. A. McCutchen	.314	.292	Yes
8. M. Cabrera	.313	.338	No
9. B. Posey	.311	.318	No
10. B. Revere	.306	.306	No

The Regression Fallacy

→ The **regression fallacy** is committed whenever regression toward the mean is interpreted as a real, rather than a chance, effect.

→ If misinterpreted as a real effect, regression toward the mean can lead to erroneous conclusions called regression fallacy.

→ The Regression Fallacy occurs when one mistakes regression to the mean, for a causal relationship. For example, if a tall father were to conclude that his tall wife committed adultery because their children were shorter, he would be committing the regression fallacy.

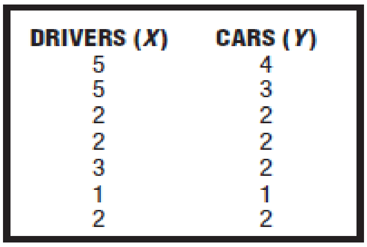
→ The regression fallacy can be avoided by splitting the subset of extreme observations into two groups.

Tutorial Questions:

1. What is normal curve? List out the properties of normal curve
2. Explain in detail about z scores
3. Outline standard normal curve and standard normal table
4. Explain in detail about finding proportions and finding scores.
   (or) What are two types of normal curve problems? How to answer these problems
5. Explain in detail about z scores for non-normal distribution
6. Discuss the three types of relationships with example. How to categories these types of relationships using scatter plot and correlation coefficient.
7. Highlight the significance of correlation coefficient? Outline the procedure for finding correlation coefficient using computational formula with example and corresponding python program.
8. Explain the significance of regression line and least square regression line with examples.
9. Calculate and analyze the correlation coefficient between the number of study hours and the number of sleeping hours of different students.

Number of study Hours	2	4	6	8	10
Number of Sleeping Hours	10	9	8	7	6

10. How standard error of estimate is calculated
11. What is significance of r^2? Give a detailed interpretation of $\mathbf{r^2}$?
12. Elucidate regression towards the mean with example. Explain regression fallacy and state how it can be avoided.
13. Discuss scatter plot with example and corresponding python program. How to interpret scatter plot.
14. Each of the following pairs represents the number of licensed drivers (X) and the number of cars (Y) for seven houses in my neighborhood:



DRIVERS (*X*)	CARS (*Y*)
5	4
5	3
2	2
2	2
3	2
1	1
2	2

i) Calculate correlation coefficient *using* the computation formula

ii) Determine the least squares equation for these data.

iii) Determine the standard error of estimate.

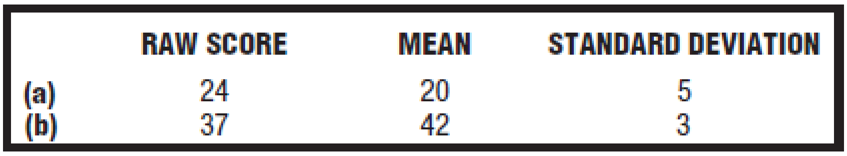
iv) Predict the number of cars for each of two new families with two and five drivers.

v) Compute r^2

Assignment Questions:

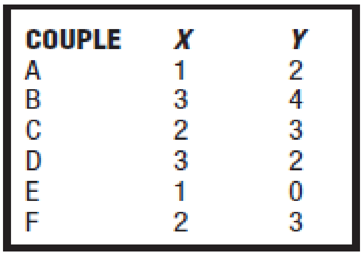
1. Express each of the following scores as a *z* score:
   (a) Margaret's IQ of 135, given a mean of 100 and a standard deviation of 15
   (b) a score of 470 on the SAT math test, given a mean of 500 and a standard deviation of 100
   (c) a daily production of 2100 loaves of bread by a bakery, given a mean of 2180 and a standard deviation of 50
   (d) Sam's height of 69 inches, given a mean of 69 and a standard deviation of 3
   (e) a thermometer-reading error of –3 degrees, given a mean of 0 degrees and a standard deviation of 2 degrees
2. Find the proportion of the total area identified with the following statements:
   (a) above a *z* score of 1.80
   (b) between the mean and a *z* score of –0.43
   (c) below a *z* score of –3.00
   (d) between the mean and a *z* score of 1.65
   (e) between *z* scores of 0 and –1.96
3. Assume that GRE scores approximate a normal curve with a mean of 500 and a standard deviation of 100. Find the proportions that correspond to the target area described by each of the following statements:
   (a) less than 400
   (b) more than 650
   (c) less than 700
4. Assume that SAT math scores approximate a normal curve with a mean of 500 and a standard deviation of 100.Find the target area(s) described by each of the following statements:
   (a) more than 570
   (b) less than 515
   (c) between 520 and 540
   (d) between 470 and 520
   (e) more than 50 points above the mean
   (f) more than 100 points either above or below the mean

5. For the normal distribution of burning times of electric light bulbs, with a mean equal to 1200 hours and a standard deviation equal to 120 hours, what burning time is identified with the
   (a) upper 50 percent?
   (b) lower 75 percent?
   (c) lower 1 percent?
   (d) middle 90 percent?
6. Assume that each of the raw scores listed originates from a distribution with the specified mean and standard deviation. After converting each raw score into a *z* score, transform each *z* score into a series of new standard scores with means and standard deviations of 50 and 10, 100 and 15, and 500 and 100, respectively



	RAW SCORE	MEAN	STANDARD DEVIATION
(a)	24	20	5
(b)	37	42	3

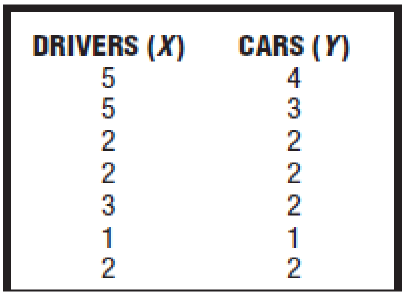
7. Indicate whether the following statements suggest a positive or negative relationship:
   (a) More densely populated areas have higher crime rates.
   (b) Schoolchildren who often watch TV perform more poorly on academic achievement tests.
   (c) Heavier automobiles yield poorer gas mileage.
   (d) Better-educated people have higher incomes.
   (e) More anxious people voluntarily spend more time performing a simple repetitive task.
8. Couples who attend a clinic for first pregnancies are asked to estimate (independently of each other) the ideal number of children. Given that *X* and *Y* represent the estimates of females and males, respectively, the results are as follows:



COUPLE	*X*	*Y*
A	1	2
B	3	4
C	2	3
D	3	2
E	1	0
F	2	3

Calculate a value for correlation coefficient *r*, using the computation formula

9. Each of the following pairs represents the number of licensed drivers (*X*) and the number of cars (*Y*) for seven houses in my neighborhood:



DRIVERS (*X*)	CARS (*Y*)
5	4
5	3
2	2
2	2
3	2
1	1
2	2

(a) Calculate a value for correlation coefficient *r*, using the computation formula

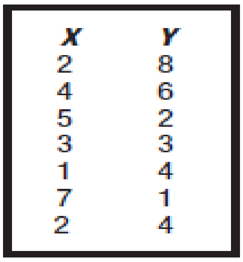
(b) Determine the least squares equation for these data.

(c) Determine the standard error of estimate.

(d) Predict the number of cars for each of two new families with two and five drivers.

(e) Determine the square of the correlation coefficient r^2

10. Consider the following data



X	*Y*
2	8
4	6
5	2
3	3
1	4
7	1
2	4

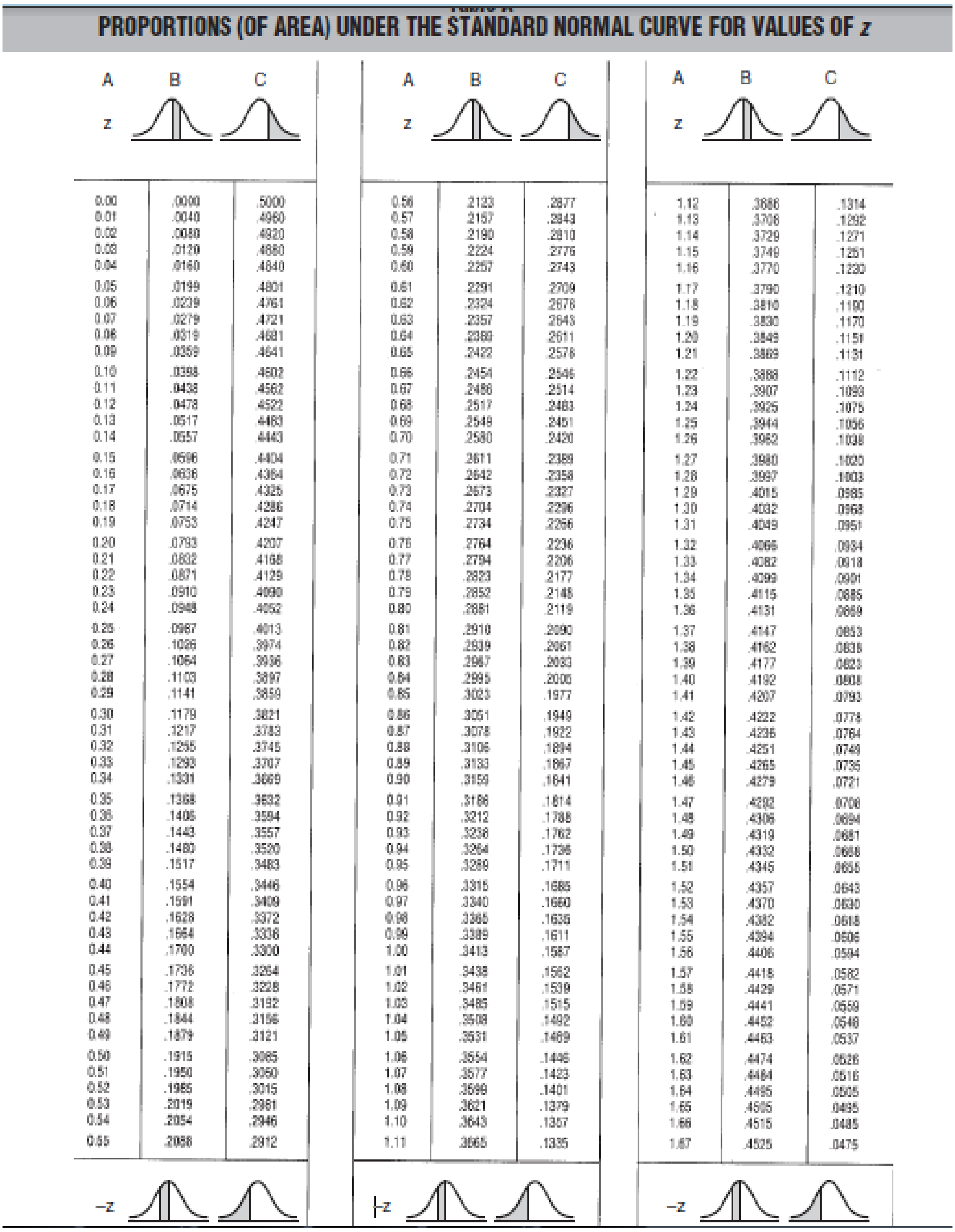
(a) Calculate a value for correlation coefficient *r*, using the computation formula

(b) Determine the least squares equation for these data.

(c) Determine the standard error of estimate.

(d) Determine the square of the correlation coefficient.

Standard Normal Table (Table A)



PROPORTIONS (OF AREA) UNDER THE STANDARD NORMAL CURVE FOR VALUES OF *z*

A	B	C	A	B	C	A	B	C
z			z			z		
0.00	.0000	.5000	0.56	.2123	.2877	1.12	.3686	.1314
0.01	.0040	.4960	0.57	.2157	.2843	1.13	.3708	.1292
0.02	.0080	.4920	0.58	.2190	.2810	1.14	.3729	.1271
0.03	.0120	.4880	0.59	.2224	.2776	1.15	.3749	.1251
0.04	.0160	.4840	0.60	.2257	.2743	1.16	.3770	.1230
0.05	.0199	.4801	0.61	.2291	.2709	1.17	.3790	.1210
0.06	.0239	.4761	0.62	.2324	.2676	1.18	.3810	.1190
0.07	.0279	.4721	0.63	.2357	.2643	1.19	.3830	.1170
0.08	.0319	.4681	0.64	.2389	.2611	1.20	.3849	.1151
0.09	.0359	.4641	0.65	.2422	.2578	1.21	.3869	.1131
0.10	.0398	.4602	0.66	.2454	.2546	1.22	.3888	.1112
0.11	.0438	.4562	0.67	.2486	.2514	1.23	.3907	.1093
0.12	.0478	.4522	0.68	.2517	.2483	1.24	.3925	.1075
0.13	.0517	.4483	0.69	.2549	.2451	1.25	.3944	.1056
0.14	.0557	.4443	0.70	.2580	.2420	1.26	.3962	.1038
0.15	.0596	.4404	0.71	.2611	.2389	1.27	.3980	.1020
0.16	.0636	.4364	0.72	.2642	.2358	1.28	.3997	.1003
0.17	.0675	.4325	0.73	.2673	.2327	1.29	.4015	.0985
0.18	.0714	.4286	0.74	.2704	.2296	1.30	.4032	.0968
0.19	.0753	.4247	0.75	.2734	.2266	1.31	.4049	.0951
0.20	.0793	.4207	0.76	.2764	.2236	1.32	.4066	.0934
0.21	.0832	.4168	0.77	.2794	.2206	1.33	.4082	.0918
0.22	.0871	.4129	0.78	.2823	.2177	1.34	.4099	.0901
0.23	.0910	.4090	0.79	.2852	.2148	1.35	.4115	.0885
0.24	.0948	.4052	0.80	.2881	.2119	1.36	.4131	.0869
0.25	.0987	.4013	0.81	.2910	.2090	1.37	.4147	.0853
0.26	.1026	.3974	0.82	.2939	.2061	1.38	.4162	.0838
0.27	.1064	.3936	0.83	.2967	.2033	1.39	.4177	.0823
0.28	.1103	.3897	0.84	.2995	.2005	1.40	.4192	.0808
0.29	.1141	.3859	0.85	.3023	.1977	1.41	.4207	.0793
0.30	.1179	.3821	0.86	.3051	.1949	1.42	.4222	.0778
0.31	.1217	.3783	0.87	.3078	.1922	1.43	.4236	.0764
0.32	.1255	.3745	0.88	.3106	.1894	1.44	.4251	.0749
0.33	.1293	.3707	0.89	.3133	.1867	1.45	.4265	.0735
0.34	.1331	.3669	0.90	.3159	.1841	1.46	.4279	.0721
0.35	.1368	.3632	0.91	.3186	.1814	1.47	.4292	.0708
0.36	.1406	.3594	0.92	.3212	.1788	1.48	.4306	.0694
0.37	.1443	.3557	0.93	.3238	.1762	1.49	.4319	.0681
0.38	.1480	.3520	0.94	.3264	.1736	1.50	.4332	.0668
0.39	.1517	.3483	0.95	.3289	.1711	1.51	.4345	.0655
0.40	.1554	.3446	0.96	.3315	.1685	1.52	.4357	.0643
0.41	.1591	.3409	0.97	.3340	.1660	1.53	.4370	.0630
0.42	.1628	.3372	0.98	.3365	.1635	1.54	.4382	.0618
0.43	.1664	.3336	0.99	.3389	.1611	1.55	.4394	.0606
0.44	.1700	.3300	1.00	.3413	.1587	1.56	.4406	.0594
0.45	.1736	.3264	1.01	.3438	.1562	1.57	.4418	.0582
0.46	.1772	.3228	1.02	.3461	.1539	1.58	.4429	.0571
0.47	.1808	.3192	1.03	.3485	.1515	1.59	.4441	.0559
0.48	.1844	.3156	1.04	.3508	.1492	1.60	.4452	.0548
0.49	.1879	.3121	1.05	.3531	.1469	1.61	.4463	.0537
0.50	.1915	.3085	1.06	.3554	.1446	1.62	.4474	.0526
0.51	.1950	.3050	1.07	.3577	.1423	1.63	.4484	.0516
0.52	.1985	.3015	1.08	.3599	.1401	1.64	.4495	.0505
0.53	.2019	.2981	1.09	.3621	.1379	1.65	.4505	.0495
0.54	.2054	.2946	1.10	.3643	.1357	1.66	.4515	.0485
0.55	.2088	.2912	1.11	.3665	.1335	1.67	.4525	.0475
−z			−z			−z		

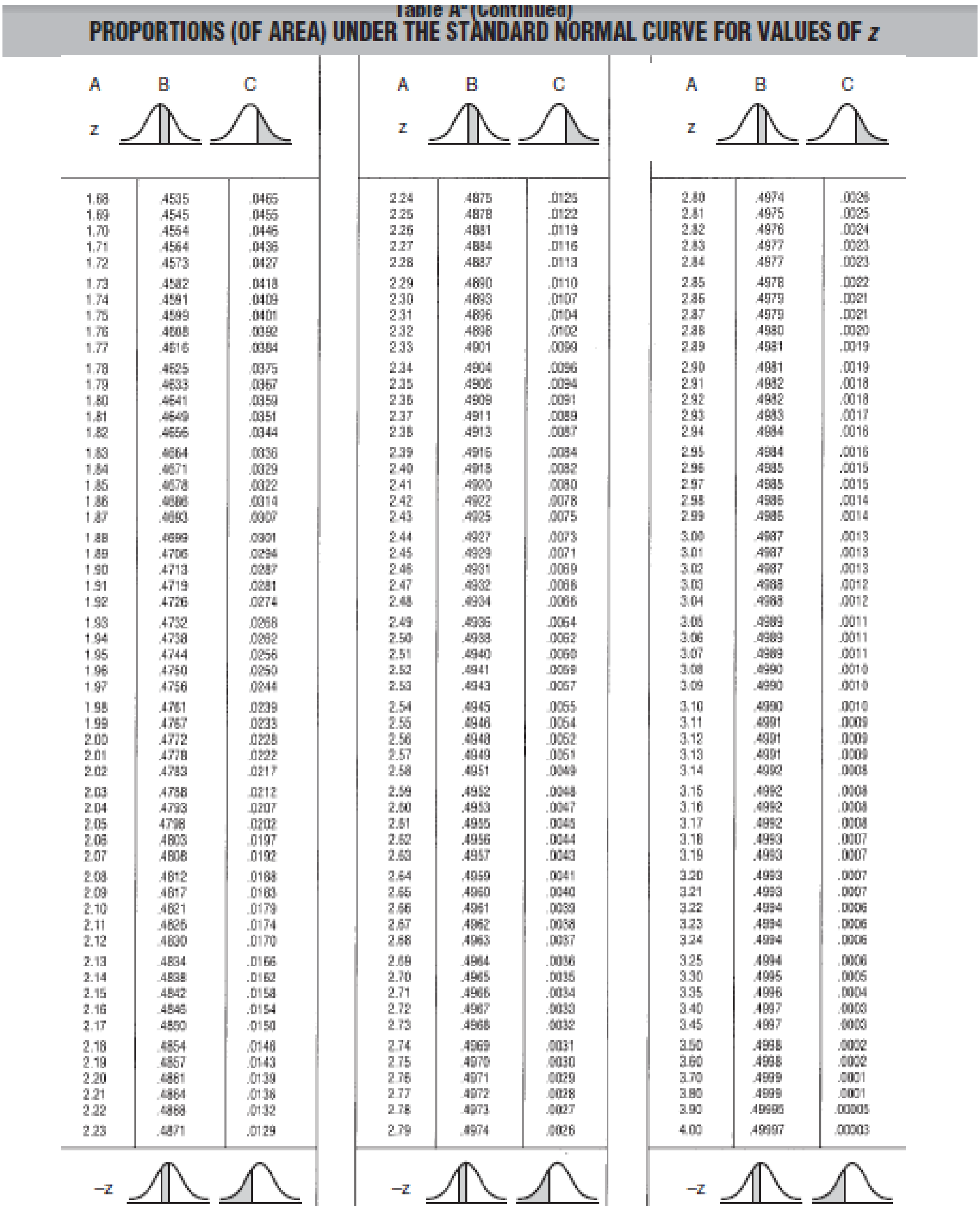


Table A-(Continued)

PROPORTIONS (OF AREA) UNDER THE STANDARD NORMAL CURVE FOR VALUES OF *z*

A	B	C
z		
1.68	.4535	.0465
1.69	.4545	.0455
1.70	.4554	.0446
1.71	.4564	.0436
1.72	.4573	.0427
1.73	.4582	.0418
1.74	.4591	.0409
1.75	.4599	.0401
1.76	.4608	.0392
1.77	.4616	.0384
1.78	.4625	.0375
1.79	.4633	.0367
1.80	.4641	.0359
1.81	.4649	.0351
1.82	.4656	.0344
1.83	.4664	.0336
1.84	.4671	.0329
1.85	.4678	.0322
1.86	.4686	.0314
1.87	.4693	.0307
1.88	.4699	.0301
1.89	.4706	.0294
1.90	.4713	.0287
1.91	.4719	.0281
1.92	.4726	.0274
1.93	.4732	.0268
1.94	.4738	.0262
1.95	.4744	.0256
1.96	.4750	.0250
1.97	.4756	.0244
1.98	.4761	.0239
1.99	.4767	.0233
2.00	.4772	.0228
2.01	.4778	.0222
2.02	.4783	.0217
2.03	.4788	.0212
2.04	.4793	.0207
2.05	.4798	.0202
2.06	.4803	.0197
2.07	.4808	.0192
2.08	.4812	.0188
2.09	.4817	.0183
2.10	.4821	.0179
2.11	.4826	.0174
2.12	.4830	.0170
2.13	.4834	.0166
2.14	.4838	.0162
2.15	.4842	.0158
2.16	.4846	.0154
2.17	.4850	.0150
2.18	.4854	.0146
2.19	.4857	.0143
2.20	.4861	.0139
2.21	.4864	.0136
2.22	.4868	.0132
2.23	.4871	.0129
2.24	.4875	.0125
2.25	.4878	.0122
2.26	.4881	.0119
2.27	.4884	.0116
2.28	.4887	.0113
2.29	.4890	.0110
2.30	.4893	.0107
2.31	.4896	.0104
2.32	.4898	.0102
2.33	.4901	.0099
2.34	.4904	.0096
2.35	.4906	.0094
2.36	.4909	.0091
2.37	.4911	.0089
2.38	.4913	.0087
2.39	.4916	.0084
2.40	.4918	.0082
2.41	.4920	.0080
2.42	.4922	.0078
2.43	.4925	.0075
2.44	.4927	.0073
2.45	.4929	.0071
2.46	.4931	.0069
2.47	.4932	.0068
2.48	.4934	.0066
2.49	.4936	.0064
2.50	.4938	.0062
2.51	.4940	.0060
2.52	.4941	.0059
2.53	.4943	.0057
2.54	.4945	.0055
2.55	.4946	.0054
2.56	.4948	.0052
2.57	.4949	.0051
2.58	.4951	.0049
2.59	.4952	.0048
2.60	.4953	.0047
2.61	.4955	.0045
2.62	.4956	.0044
2.63	.4957	.0043
2.64	.4959	.0041
2.65	.4960	.0040
2.66	.4961	.0039
2.67	.4962	.0038
2.68	.4963	.0037
2.69	.4964	.0036
2.70	.4965	.0035
2.71	.4966	.0034
2.72	.4967	.0033
2.73	.4968	.0032
2.74	.4969	.0031
2.75	.4970	.0030
2.76	.4971	.0029
2.77	.4972	.0028
2.78	.4973	.0027
2.79	.4974	.0026
2.80	.4974	.0026
2.81	.4975	.0025
2.82	.4976	.0024
2.83	.4977	.0023
2.84	.4977	.0023
2.85	.4978	.0022
2.86	.4979	.0021
2.87	.4979	.0021
2.88	.4980	.0020
2.89	.4981	.0019
2.90	.4981	.0019
2.91	.4982	.0018
2.92	.4982	.0018
2.93	.4983	.0017
2.94	.4984	.0016
2.95	.4984	.0016
2.96	.4985	.0015
2.97	.4985	.0015
2.98	.4986	.0014
2.99	.4986	.0014
3.00	.4987	.0013
3.01	.4987	.0013
3.02	.4987	.0013
3.03	.4988	.0012
3.04	.4988	.0012
3.05	.4989	.0011
3.06	.4989	.0011
3.07	.4989	.0011
3.08	.4990	.0010
3.09	.4990	.0010
3.10	.4990	.0010
3.11	.4991	.0009
3.12	.4991	.0009
3.13	.4991	.0009
3.14	.4992	.0008
3.15	.4992	.0008
3.16	.4992	.0008
3.17	.4992	.0008
3.18	.4993	.0007
3.19	.4993	.0007
3.20	.4993	.0007
3.21	.4993	.0007
3.22	.4994	.0006
3.23	.4994	.0006
3.24	.4994	.0006
3.25	.4994	.0006
3.30	.4995	.0005
3.35	.4996	.0004
3.40	.4997	.0003
3.45	.4997	.0003
3.50	.4998	.0002
3.60	.4998	.0002
3.70	.4999	.0001
3.80	.4999	.0001
3.90	.49995	.00005
4.00	.49997	.00003
−z		