Problem 3

3. A third-grade teacher wants to improve her students' level of engagement during group discussions and instruction. She keeps track of each of the 15 third graders' number of responses every day for 1 week, and the data are available as Chapter 4 Data Set 2. Use SPSS to create a bar chart with one bar for each day (and warning—this may be a toughie).

Class Interval	Frequency
261–280	140
241–260	320
221–240	3380
201–220	600
181–200	500
161–180	410
141–160	315
121–140	300
100–120	200

Problem 5

4. Use the data available as Chapter 4 Data Set 3 on pie preference to create a pie chart ☺ using SPSS.

5. For each of the following, indicate whether you would use a pie, line, or bar chart and why.

   a. The proportion of freshmen, sophomores, juniors, and seniors in a particular university
   b. Change in temperature over a 24-hour period
   c. Number of applicants for four different jobs
   d. Percentage of test takers who passed
   e. Number of scores in each of 10 categories

6. Provide an example of when you might use each of the following types of charts. For example, you would use a pie chart to show the proportion of children in Grades 1 through 6 who receive a reduced-price lunch. When you are done, draw the fictitious chart by hand.

   a. Line
   b. Bar
   c. Scatter/Dot (extra credit)

7. Go to the library and find a journal article in your area of interest that contains empirical data but does not contain any visual representation of them. Use the data to create a chart. Be sure to specify what type of chart you are creating and why you chose the one you did. You can create the chart manually or using SPSS or Excel.
8. Create the worst-looking chart that you can, crowded with chart and font junk. Nothing makes as lasting an impression as a *bad* example.
9. And, finally, what is the purpose of a chart or graph?

STUDENT STUDY SITE

SAGE edge™

Get the tools you need to sharpen your study skills! Visit **edge.sagepub.com/salkind6e** to access practice quizzes, eFlashcards, original and curated videos, data sets, journal articles, and more!

8. Questions 8a through 8d are based on a distribution of scores with $\bar{X}$ = 75 and standard deviation = 6.38. Draw a small picture to help you see what's required.

    a. What is the probability of a score falling between a raw score of 70 and 80?

    b. What is the probability of a score falling above a raw score of 80?

    c. What is the probability of a score falling between a raw score of 81 and 83?

    d. What is the probability of a score falling below a raw score of 63?

9. Identify whether these distributions are negatively skewed, positively skewed, or not skewed at all and explain why you describe them that way.

    a. This talented group of athletes scored very high on the vertical-jump task.

    b. On this incredibly crummy test, everyone received the same score.

    c. On the most difficult spelling test of the year, the third graders wept as the scores were delivered and then their parents complained.

10. Jake needs to score in the top 10% of his class to earn a physical fitness certificate. The class mean is 78, and the standard deviation is 5.5. What raw score does he need to get that valuable piece of paper?

11. Imagine you are in charge of a program in which members are evaluated on five different tests at the end of the program. Why doesn't it make sense to simply compute the average of the five scores as a measure of performance rather than compute a *z* score for each test for each individual and average those?

12. Who is the better student, relative to his or her classmates? Here's all the information you ever wanted to know . . .

Math		
Class Mean	81	
Class Standard Deviation	2	
Reading		
Class Mean	87	
Class Standard Deviation	10	

(Continued)

(Continued)

Raw Scores			
	Math Score	**Reading Score**	**Average**
Noah	85	88	86.5
Talya	87	81	84.0
z Scores			
	Math Score	**Reading Score**	**Average**
Noah	________	________	________
Talya	________	________	________

13. Here's an interesting extra-credit question. As you know, one of the defining characteristics of the normal curve is that the tails do not touch the *x*-axis. Why don't they touch?

STUDENT STUDY SITE

SAGE edge™

Get the tools you need to sharpen your study skills! Visit **edge.sagepub.com/salkind6e** to access practice quizzes, eFlashcards, original and curated videos, data sets, journal articles, and more!

Custom Defining Variables: Using the Variable View Window

To define a variable yourself, you must first go to the Variable View window by clicking the Variable View tab at the bottom of the SPSS screen. Once that is done, you will see the Variable View window, as shown in Figure A.8, and be able to define any one variable as you see fit.

Once in the Variable View window, you can define variables along the following parameters:

- *Name* allows you to give a variable a name of up to 64 characters.
- *Type* defines the type of variable, such as text, numeric, string, scientific notation, and so on.
- *Width* defines the number of characters of the column housing the variable.
- *Decimals* defines the number of decimals that will appear in the Data View.
- *Label* defines a label of up to 256 characters for the variable.
- *Values* defines the labels that correspond to certain numerical values (such as 1 for male and 2 for female).
- *Missing* indicates how any missing data will be dealt with.
- *Columns* defines the number of spaces allocated for the variable in the Data View window.
- *Align* defines how the data are to appear in the cell (right, left, or center aligned).
- *Measure* defines the scale of measurement that best characterizes the variable (nominal, ordinal, or interval).
- *Role* defines the part that the variable plays in the overall analysis (input, target, etc.).

If you place the cursor in the first cell under the "Name" column, enter any name, and press the Enter key, then SPSS will automatically provide you with the default values for all the variable characteristics. Even if you are not in the Data View screen (click the tab on the bottom of the window), SPSS will automatically name the variables var0001, var0002, and so on.

Figure A.8 The Variable View Window

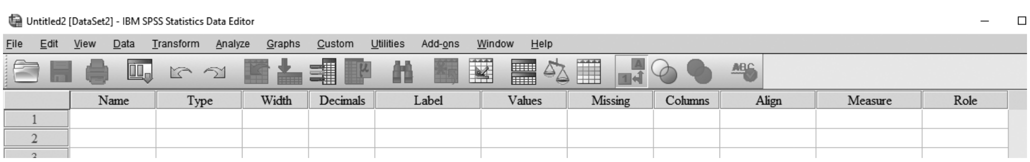


Table B.1 Areas Beneath the Normal Curve

z-Score	Area Between the Mean and the *z*-Score	*z*-Score	Area Between the Mean and the *z*-Score	*z*-Score	Area Between the Mean and the *z*-Score	*z*-Score	Area Between the Mean and the *z*-Score	*z*-Score	Area Between the Mean and the *z*-Score	*z*-Score	Area Between the Mean and the *z*-Score	*z*-Score	Area Between the Mean and the *z*-Score	*z*-Score	Area Between the Mean and the *z*-Score
0.00	0.00	0.50	19.15	1.00	34.13	1.50	43.42	2.00	47.72	2.50	49.38	3.00	49.87	3.50	49.98
0.01	0.40	0.51	19.50	1.01	34.38	1.51	43.45	2.01	47.78	2.51	49.40	3.01	49.87	3.51	49.98
0.02	0.80	0.52	19.85	1.02	34.61	1.52	43.57	2.02	47.83	2.52	49.41	3.02	49.87	3.52	49.98
0.03	1.20	0.53	20.19	1.03	34.85	1.53	43.70	2.03	47.88	2.53	49.43	3.03	49.88	3.53	49.98
0.04	1.60	0.54	20.54	1.04	35.08	1.54	43.82	2.04	47.93	2.54	49.45	3.04	49.88	3.54	49.98
0.05	1.99	0.55	20.88	1.05	35.31	1.55	43.94	2.05	47.98	2.55	49.46	3.05	49.89	3.55	49.98
0.06	2.39	0.56	21.23	1.06	35.54	1.56	44.06	2.06	48.03	2.56	49.48	3.06	49.89	3.56	49.98
0.07	2.79	0.57	21.57	1.07	35.77	1.57	44.18	2.07	48.08	2.57	49.49	3.07	49.89	3.57	49.98
0.08	3.19	0.58	21.90	1.08	35.99	1.58	44.29	2.08	48.12	2.58	49.51	3.08	49.90	3.58	49.98
0.10	3.98	0.60	22.57	1.09	36.21	1.59	44.41	2.09	48.17	2.59	49.52	3.09	49.90	3.59	49.98
0.11	4.38	0.61	22.91	1.10	36.43	1.60	44.52	2.10	48.21	2.60	49.53	3.10	49.90	3.60	49.98
0.12	4.78	0.62	23.24	1.11	36.65	1.61	44.63	2.11	48.26	2.61	49.55	3.11	49.91	3.61	49.98
0.13	5.17	0.63	23.57	1.12	36.86	1.62	44.74	2.12	48.30	2.62	49.56	3.12	49.91	3.62	49.98
0.14	5.57	0.64	23.89	1.13	37.08	1.63	44.84	2.13	48.34	2.63	49.57	3.13	49.91	3.63	49.99
0.15	5.96	0.65	24.22	1.14	37.29	1.64	44.95	2.14	48.38	2.64	49.59	3.14	49.92	3.64	49.99
0.16	6.36	0.66	24.54	1.15	37.49	1.65	45.05	2.15	48.42	2.65	49.60	3.15	49.92	3.65	49.99
0.17	6.75	0.67	24.86	1.16	37.70	1.66	45.15	2.16	48.46	2.66	49.61	3.16	49.92	3.66	49.99
0.18	7.14	0.68	25.17	1.17	37.90	1.67	45.25	2.17	48.50	2.67	49.62	3.17	49.92	3.67	49.99
0.19	7.53	0.69	25.49	1.18	38.10	1.68	45.35	2.18	48.54	2.68	49.63	3.18	49.93	3.68	49.99
0.20	7.93	0.70	25.80	1.19	38.30	1.69	45.45	2.19	48.57	2.69	49.64	3.19	49.93	3.69	49.99
0.21	8.32	0.71	26.11	1.20	38.49	1.70	45.54	2.20	48.61	2.70	49.65	3.20	49.93	3.70	49.99
0.22	8.71	0.72	26.42	1.21	38.69	1.71	45.64	2.21	48.64	2.71	49.66	3.21	49.93	3.71	49.99
0.23	9.10	0.73	26.73	1.23	39.07	1.73	45.82	2.23	48.71	2.73	49.68	3.23	49.94	3.73	49.99
0.24	9.48	0.74	27.04	1.24	39.25	1.74	45.91	2.24	48.75	2.74	49.69	3.24	49.94	3.74	49.99

4. Here's what your one paragraph might look like.

   As usual, the Chicken Littles (the mode) led the way in sales. The total value of food sold was $303, for an average of $2.55 for each special.

5. There's nothing too dramatic here in terms of values that are really big or small or look weird (all reasons to use the median), so we'll just use the mean. You can see the mean for the three stores in the last column. Seems like these might be the numbers you want the new store to approximate for it to be around the average of all the stores you manage.

Average	Store 1	Store 2	Store 3	Mean
Sales (in thousands of dollars)	$323	$234	$308	$288.83
Number of items purchased	3,454	5,645	4,565	4554.67
Number of visitors	4,534	6,765	6,654	5984.33

6. Well, it seems like beer chicken is the winner and the fruit cup comes in last. We computed the mean since this is a rating of values across a scale that is interval (at least in its intent). Here are the data . . .

Snack Food	North Fans	East Fans	South Fans	West Fans	Mean Rating
Loaded Nachos	4	4	5	4	4.25
Fruit Cup	2	1	2	1	1.5
Spicy Wings	4	3	3	3	3.25
Gargantuan Overstuffed Pizza	3	4	4	5	4
Beer Chicken	5	5	5	4	4.75

7. You use the median when you have extreme scores that would disproportionately bias the mean. One situation in which the median is preferable to the mean is when income is reported. Because it varies so much, you want a measure of central tendency that is insensitive to extreme scores. As another example, say you are studying the speed with which a group of adolescents can run 100 yards when one or two individuals are exceptionally fast.

8. You would use the median because it is insensitive to extreme scores.

9. The median is the best measure of central tendency, and it is the one score that best represents the entire set of scores. Why? Because it is relatively unaffected by the (somewhat) extreme data point $199,000. As you can see in the table below, the mean is affected (it is more than $83,000 when the highest score is included).

Before Removal of Highest Score	Mean	$83,111
	Median	$77,153
After Removal of Highest Score	Mean	$75,318
	Median	$76,564

10. The averages are as follows: Group 1 = 5.88, Group 2 = 5.00, and Group 3 = 7.00.

11. This should be an easy one. Anytime values are expressed as categories, the only type of average that makes sense is the mode. So, who likes pie? Well, the first week, the winner is apple, and week 2 finds apple once again the leader. For week 3, it's Douglas County Pi, and the month is finished up in week 5 with lots of helpings of chocolate silk.

CHAPTER 3

1. The range is the most convenient measure of dispersion, because it requires only that you subtract one number (the lowest value) from another number (the highest value). It's imprecise because it does not take into account the values that fall between the highest and the lowest values in a distribution. Use the range when you want a very gross (not very precise) estimate of the variability in a distribution.

2.

High Score	Low Score	Inclusive Range	Exclusive Range
12.1	3	10.1	9.1
92	51	42	41
42	42	1	0
7.5	6	2.5	1.5
27	26	2	24

know what the range is by knowing only the standard deviation or the variance. You can't even tell whether the range is large or small, because you don't know what's being measured and you don't know the scale of the measurement (itsy-bitsy bugs or output from steam engines).

10.
   a. Range = 6, standard deviation = 2.58, variance = 6.67.
   b. Range = 0.6, standard deviation = 0.25, variance = 0.06.
   c. Range = 3.5, standard deviation = 1.58, variance = 2.49.
   d. Range = 123, standard deviation = 48.23, variance = 2326.5.

11. Here's a chart that summarizes the results. Look familiar? It should—it's exactly what the SPSS output looks like.

Statistics

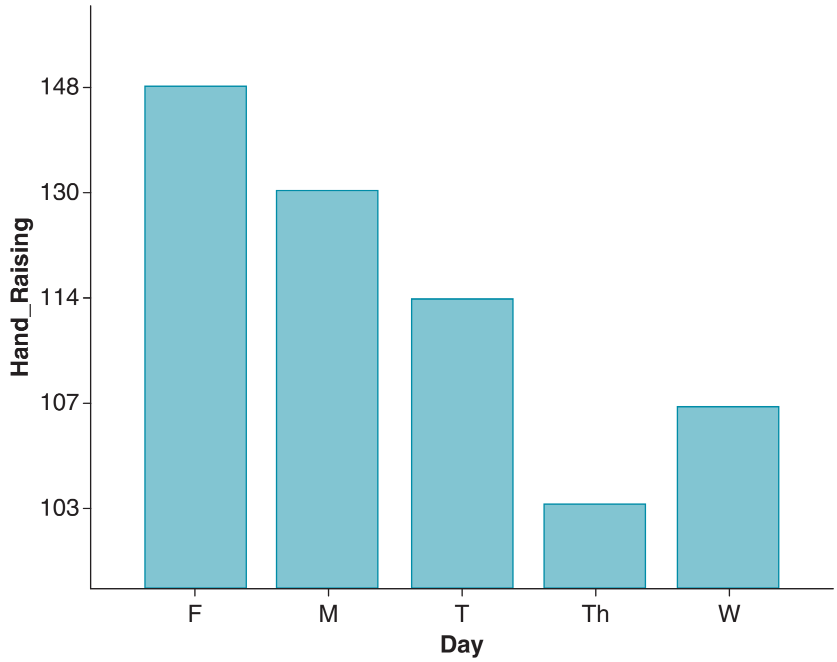
		Height	**Weight**
N	Valid	20	20
	Missing	0	0
Std. Deviation		11.44	15.66
Variance		130.78	245.00
Range		43	59

12. Okay, here's how you answer it. Compute the standard deviation for any set of 10 or so numbers by hand using the formula for the standard deviation that has $n - 1$ in the denominator (the unbiased estimate). Then compare it, using the same numbers, with the SPSS output. As you will see, they are the same, indicating that SPSS produces an unbiased estimate. If you got this one correct, you know what you're doing—go to the head of the class. Another way would be to write to the SPSS folks and ask them. No kidding—see what happens.

13. The unbiased estimate of the standard deviation is 6.49, and the biased estimate is 6.15. The unbiased estimate of the variance is 42.1, and the biased estimate is 37.89. The biased estimates are (always) smaller because they are based on a larger n and are a more conservative estimate.

14. A standard deviation of 0.94 means that on average, each score in the set of scores is a distance of 0.94 correct words from the average of all scores.

Figure D.3 A Simple Bar Chart



4. We used SPSS to create a simple pie chart and then changed the fill in each section of the pie, as shown in Figure D.4. What kind of pie do you like?

5.

   a. Pie because you are interested in looking at proportions.

   b. Line because you are looking at a trend (over time).

   c. Bar because you are looking at a number of discrete categories.

   d. Pie because you are looking at categories defined as proportions.

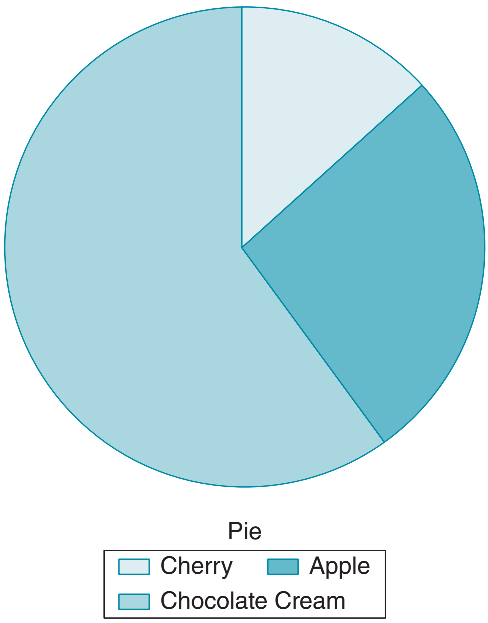
   e. Bar because you are looking at a number of discrete categories.

6. You will come up with examples of your own, but here are some of ours. Draw your own examples.

   a. The number of words that a child knows as a function of ages from 12 months to 36 months

   b. The percentage of senior citizens who belong to AARP as a function of gender and ethnicity

Figure D.4 A Simple Pie Chart



c. The plotting of paired scores such as height and weight for each individual. We haven't discussed this type of chart much, but take a few glances at the Graph menu on SPSS and you'll get it.

7. On your own!

8. We created Figure D.5 using SPSS and the chart editor, and it's as uninformative as it is ugly. Major chart junk attack.

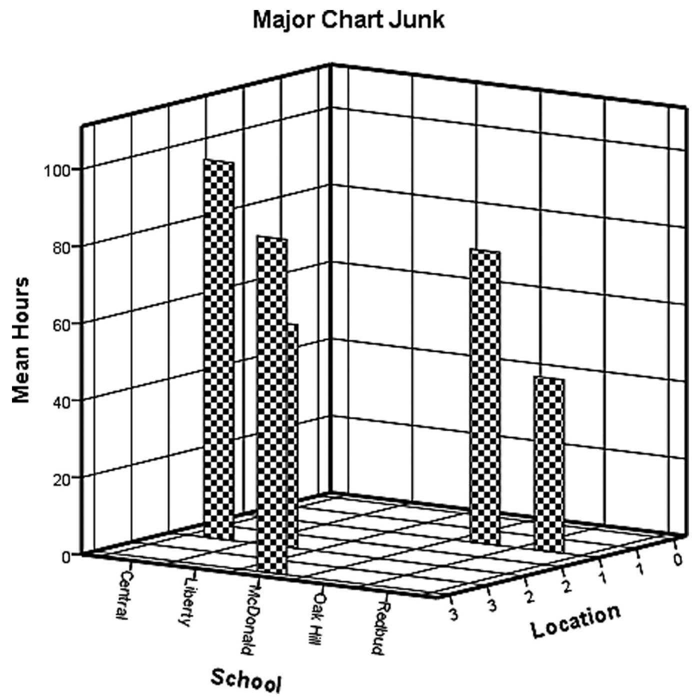
9. A picture (a chart or graph) is worth more than 1,000 words. In other words, there are many possible answers to such a question, but in general, the purpose of a chart or a graph is to visually illustrate information in a way that is as simple as possible while communicating a central message that is clear and direct.

   a. This is negatively skewed because the majority of athletes scored in the upper range.

   b. This is not skewed at all. In fact, the distribution is like a rectangle because everyone scored exactly the same.

   c. This distribution is positively skewed, because most of the spellers scored very low.

Figure D.5 A Really, Really Ugly Chart



CHAPTER 5

1.

   a. $r = .596$.

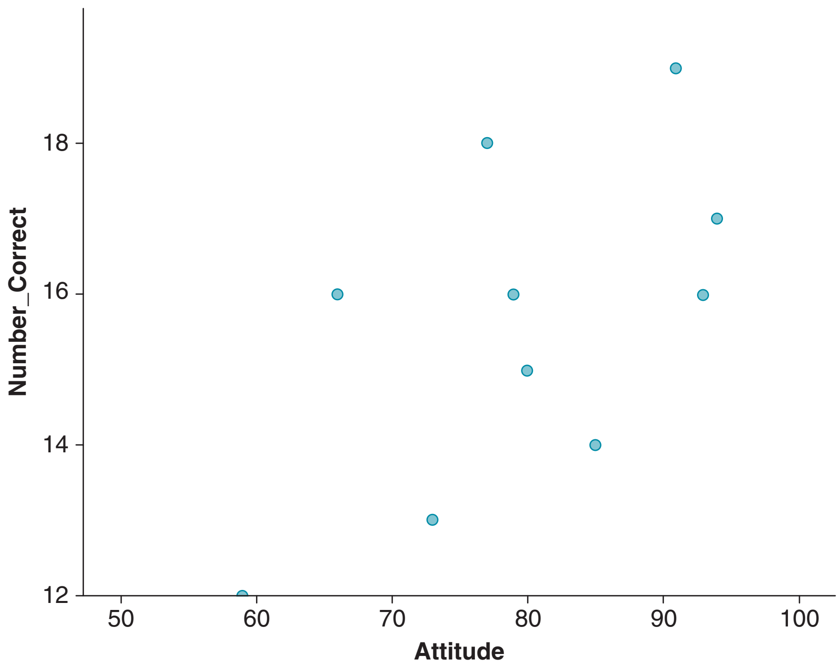
   b. From the answer to 1a, you already know that the correlation is direct. But from the scatterplot shown in Figure D.6 (we used SPSS, but you should do it by hand), you can predict it to be such (without actually knowing the sign of the coefficient) because the data points group themselves from the lower left corner of the graph to the upper right corner and assume a positive slope.

2.

   a. $r =\ \ .269$.

   b. According to the table presented earlier in the chapter, the general strength of the correlation of this magnitude is weak. The coefficient of determination is $.269^2$, so .072 or 7.2% of the variance is accounted for. The subjective analysis (weak) and the objective one (7.2% of the variance accounted for) are consistent with one another.

Figure D.6 Scatterplot of Data From Chapter 5 Data Set 2



3. Note that the .71 and the .47 have no sign, and in that case, we always assume that (like any other number) the values are positive.

   +.36

   −.45

   +.47

   −.62

   +.71

4. The correlation is .64, meaning that increases in budget and increases in classroom achievement are positively related to one another (and note that we have to test for significance). From a descriptive perspective, a bit more than 40% of the variance is shared between the two variables.

5. The correlation between number of minutes of exercise per day and GPA is .49, showing that as exercise increases, so does GPA. And of course, as exercise decreases, so does GPA. And want to be buff and get good grades? Work out, study, read broadly (and shower often). Keep in mind that GPA has nothing to do with

how good you look at the gym. That would be causality and we're only dealing with associations here.

6. The correlation is .14 and is so low because both the hours of studying and the set of GPA scores has very little variability. When there is so little variability, there is nothing to share, and the two sets of scores have little in common—hence, the low correlation.

7.
    a. .8
    b. Very strong
    c. 1.00 – .64, or .36 (36%)

8. Here's the matrix . . .

	Age at Injury	Level of Treatment	12-Month Treatment Score
Age at Injury	1		
Level of Treatment	0.0557	1	
12-Month Treatment Score	–0.154	0.389	1

9. To examine the relationship between sex (defined as male or female) and political affiliation, you would use the phi coefficient because both variables are nominal in nature. To examine the relationship between family configuration and high school GPA, you would use the point biserial correlation because one variable is nominal (family configuration) and the other is interval (GPA).

10. Just because two things are related does not mean that one causes the other. There are plenty of runners with average strength who can run fast, and plenty of very strong people who run slowly. Strength can help people run faster—but technique is more important (and, by the way, accounts for more of the variance).

11. You will come up with your own explanations, but here are two examples.
    a. We bet you expected this one—consumption of ice cream and number of crimes.
    b. Amount of money spent on political ads and the number of people who vote

b. The probability of a score falling above a raw score of 80 is .2177. A z score for a raw score of 80 is 0.78. The area between the mean and a z score of 0.78 is 28.23%. The area below a z score of 0.78 is .50 + .2823, or .7823. The difference between 1 (the total area under the curve) and .7823 is .2177 or 21.77%.

c. The probability of a score falling between a raw score of 81 and a raw score of 83 is .068. A z score for a raw score of 81 is 0.94, and a z score for a raw score of 83 is 1.25. The area between the mean and a z score of 0.94 is 32.64%. The area between the mean and a z score of 1.25 is 39.44%. The difference between the two is .3944 – .3264 = .068, or 6.8%.

d. The probability of a score falling below a raw score of 63 is .03. A z score for a raw score of 63 is −1.88. The area between the mean and a z score of −1.88 is 46.99%. The area below a z score of 1.88 is 1 – (.50 + .4699) = .03, or 3%.

9.

a. This is negatively skewed because the majority of athletess-cored in the upper range.

b. This is not skewed at all. In fact, the distribution is like arect-angle because everyone scored exactly the same.

c. This distribution is positively skewed, because most of thes-pellers scored very low.

10. A little magic lets us solve for the raw score using the same formulas for computing the z score that you've seen throughout this chapter. Here's the transformed formula . . .

$$X = (s \times z) + \overline{X}$$

And taking this one step further, all we really need to know is the z score of 90% (or 40% in Table B.1), which is 1.29.

So we have the following formula:

$$X = (s \times z) + \overline{X}$$

or

$$X = 78 + (5.5 \times 1.29) = 85.095$$

Jake is home free if he gets that score and, along with it, his certificate.

11. It doesn't make sense because raw scores are not comparable to one another when they belong to different distributions. A raw score of 80 on the math test, where the class mean was 40, is just not comparable to an 80 on the essay-writing skills test, where everyone got the one answer correct. Distributions, like people, are not always comparable to one another. Not everything (or everyone) is comparable to something else.

12. Here's the info with the unknown values in bold.

Math			
Class Mean	81		
Class Standard Deviation	2		
Reading			
Class Mean	87		
Class Standard Deviation	10		
Raw Scores			
	Math Score	**Reading Score**	**Average**
Noah	85	88	86.5
Talya	87	81	84.0
***z* Scores**			
	Math Score	**Reading Score**	**Average**
Noah	**2**	**0.1**	**1.05**
Talya	**3**	**–0.6**	**1.2**

Noah has the higher average raw score (86.5 vs. 84 for Talya), but Talya has the higher average z score (1.2 vs. 1.05 for Noah). Remember that we asked who was the better student relative to the rest, which requires the use of a standard score (we used z scores). But why is Talya the better student relative to Noah? It's because on the tests with the lowest variability (Math with an $SD = 2$), Talya really stands out with a z score of 3. That put her ahead to stay.

13. The fact that the tails do not touch the x-axis indicates that there is always a chance—even though it may be very, very small—that extreme scores (all the way far out in either direction on the x-axis) are possible. If the tails did touch the x-axis, it would mean there is a limit to how improbable an outcome can be. In other words, no matter what the outcome, there's always a chance it will occur.

correct and time. A one-tailed test was used because the research hypothesis was that the relationship was indirect or negative, and approximately 20% of the variance is accounted for.

c. With 48 degrees of freedom at the .05 level, the critical value for rejection of the null hypothesis is 0.273 for a two-tailed test. There is a significant correlation between number of friends a child might have and GPA, and the correlation accounts for 13.69% of the variance.

2.

Correlations

		Motivation	**GPA**
Motivation	Pearson Correlation	1	.434*
	Sig. (2-tailed)		.017
	N	30	30
GPA	Pearson Correlation	.434*	1
	Sig. (2-tailed)	.017	
	N	30	30

*Correlation is significant at the.05 level (two-tailed).

a. and b. We used SPSS to compute the correlation as .434, which is significant at the .017 level using a two-tailed test.

c. True. The more motivated you are, the more you will study; and the more you study, the more you are motivated. But (and this is a big "but") studying more does not cause you to be more highly motivated, nor does being more highly motivated cause you to study more.

3.

		Income	**Level_Education**
Income	Pearson Correlation	1	.629(*)
	Sig. (2-tailed)		.003
	N	20	20
Level_Education	Pearson Correlation	.629(**)	1
	Sig. (2-tailed)	.003	
	N	20	20

**Correlation is significant at the .01 level (2-tailed).