Here's another example. In a set of scores with a mean of 100 and a standard deviation of 10, a raw score of 117 has a corresponding z score of 1.70. This z score of 1.70 corresponds to an area under the curve between a z score of 0 and a z score of 1.70 of 45.54%. The probability of a raw score of 117 occurring is 95.54% (or 50% + 45.54%) or 95.5 out of 100 or .955.

First, even though we are focusing on *z* scores, there are other types of standard scores. For example, a *t* score is a type of standard score that is computed by multiplying the *z* score by 10 and adding 50. One advantage of this type of score is that you rarely have a negative *t* score. As with *z* scores, *t* scores allow you to compare standard scores from different distributions.

Second, a *standard score* is a whole different animal from a *standardized score*. A standardized score is one that comes from a distribution with a predefined mean and standard deviation. Standardized scores from tests such as the SAT and GRE (Graduate Record Exam) are used so that comparisons can be made easily between scores where the same mean and standard deviation are being used.

What z Scores Really Represent

The name of the statistics game is being able to estimate the probability of an outcome. If we take what we have talked about and done so far in this chapter one step further, statistics is about deciding the probability of some event occurring. Then we use some criterion to judge whether we think that event is as likely, more likely, or less likely than what we would expect by chance. The research hypothesis presents a statement of the expected event, and we use our statistical tools to evaluate how likely that event is.

That's the 20-second version of what statistics is, but that's a lot. So let's take everything from the previous paragraph and go through it again with an example.

Let's say that your lifelong friend, trusty Lew, gives you a coin and asks you to determine whether it is a "fair" one—that is, if you flip it 10 times, you should come up with 5 heads and 5 tails. We would expect 5 heads (or 5 tails) because the probability is .5 of either one head or one tail on any one flip. On 10 independent flips (meaning that one flip does not affect another), we should get 5 heads, and so on. Now the question is, how many heads would disqualify the coin as being fake or rigged?

This Type I error, or level of significance, has certain values associated with it that define the risk you are willing to take in any test of the null hypothesis. The conventional levels set are between .01 and .05.

For example, if the level of significance is .01, then on any one test of the null hypothesis, there is a 1% chance you will reject the null hypothesis when the null is true and conclude that there is a group difference when there really is no group difference at all.

If the level of significance is .05, it means that on any one test of the null hypothesis, there is a 5% chance you will reject it when the null is true (and conclude that there is a group difference) when there really is no group difference at all. Notice that the level of significance is associated with an independent test of the null. Therefore, it is not appropriate to say that "on 100 tests of the null hypothesis, I will make an error on only 5, or 5% of the time."

In a research report, statistical significance is usually represented as $p < .05$, read as "the probability of observing that outcome is less than .05," often expressed in a report or journal article simply as "significant at the .05 level."

With the introduction of fancy-schmancy software such as SPSS and Excel that can do statistical analysis, there's no longer the worry about the imprecision of such statements as "$p < .05$" or "$p < .01$." For example, $p < .05$ can mean anything from .000 to .049999, right? Instead, software such as Excel gives you the *exact* probability, such as .013 or .158, of the risk you are willing to take that you will commit a Type I error. So, when you see in a research article the statement that "$p < .05$," it means that the value of p is equal to anything from .00 to .049999999999 . . . (you get the picture). Likewise, when you see "$p > .05$" or "$p = $ n.s." (for nonsignificant), it means that the probability of rejecting a true null exceeds .05 and, in fact, can range from .0500001 to 1.00. So, it's actually terrific when we know the exact probability of an outcome because we can measure more precisely the risk we are willing to take.

But what to do if the p value is exactly .05? Well, given what you've already read, if you want to play by the rules, then the outcome is not significant. A result either is, or is not. So, .04999999999 is and .05 is not. Now, if Excel (or any other program) generates a value of .05, extend the number of decimal places—it may really be .04999999999.
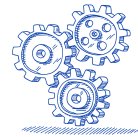
There is another kind of error you can make, which, along with the Type I error, is shown in Table 9.1. A Type II error (Cell 3 in the chart) occurs when you inadvertently accept a false null hypothesis.

taught using visuals, and those in Group 2 were taught using visuals and intense verbal rehearsal. We'll use the data to compute the test statistic step-by-step.

Group 1			Group 2		
7	5	5	5	3	4
3	4	7	4	2	3
3	6	1	4	5	2
2	10	9	5	4	7
3	10	2	5	4	6
8	5	5	7	6	2
8	1	2	8	7	8
5	1	12	8	7	9
8	4	15	9	5	7
5	3	4	8	6	6

Here are the famous eight steps in the computation of the *t*-test statistic:

1. *State the null and research hypotheses.* As represented by Formula 11.2 below, the null hypothesis states that there is no difference between the means for Group 1 and Group 2. For our purposes, the research hypothesis, shown in Formula 11.3, states that there is a difference between the means of the two groups. The research hypothesis is a two-tailed, nondirectional research hypothesis because it posits a difference, but in no particular direction.

   The null hypothesis is as follows:

   $$H_0 : \mu_1 = \mu_2 \qquad\qquad (11.2)$$

   And the research hypothesis is this:

   $$H_1 : \overline{X}_1 \neq \overline{X}_2 \qquad\qquad (11.3)$$

2. *Set the level of risk (or the level of significance or Type I error) associated with the null hypothesis.* The level of risk or Type I error or level of significance (any other names?) is .05. This number is completely up to the researcher.

Pretest (Before)	Posttest (After)	Difference	D^2
5	6	1	1
3	7	4	16
6	8	2	4
7	8	1	1
8	7	−1	1
7	9	2	4
6	10	4	16
7	9	2	4
8	9	1	1
8	8	0	0
9	8	−1	1
9	4	−5	25
8	4	−4	16
7	5	−2	4
7	6	−1	1
6	9	3	9
7	8	1	1
8	12	4	16
Sum 158	188	30	180
Mean 6.32	7.52	1.2	7.2

Here are the famous eight steps in the computation of the *t*-test statistic:

1. *State the null and research hypotheses.* The null hypothesis states that there is no difference between the means for the pretest and the posttest scores on reading achievement. The research hypothesis is a one-tailed, nondirectional research hypothesis because it posits that the posttest score will be higher than the pretest score.

The null hypothesis is as follows:

$$H_0: \mu_{posttest} = \mu_{pretest} \qquad (12.2)$$

The research hypothesis is this:

$$H_1: \bar{X}_{posttest} > \bar{X}_{pretest} \qquad (12.3)$$

8. −1

9. −3159.09

10. 126

Want some more help and more practice? Take a look at these sites:

www.webmath.com

www.math.com/homeworkhelp/BasicMath.html

www.purplemath.com

www.khanacademy.org

There's nothing worse than starting a course and being so anxious that any meaningful learning just can't take place. Thousands of students less well prepared than you have succeeded, and you can as well. Reread the Chapter 1 tips on how to approach the material in this course—and good luck!