Chapter 4
Introduction to Hypothesis Testing

Instructor's Summary of Chapter

This chapter introduces the logic and key terminology of hypothesis testing, illustrating it with the rare-in-practice research situation in which a sample of one person is being compared to a known population.

Hypothesis testing. It is a systematic procedure for deciding whether the results of a study (using a sample) support the researcher's hypothesis and can be generalized (to the population).

Core logic of hypothesis testing. It is an opposite-of-what-you-predict, roundabout type of reasoning. Researchers examine the probability that the outcome of a study could have arisen even if the true situation was that of no difference. (In other words, the question is "What is the probability of getting our research results if the opposite of what we are predicting were true?") If the probability is low, we reject the opposite prediction and accept our prediction.

Research hypothesis and null hypothesis. The expectation of a difference or effect is the research hypothesis (or "alternative hypothesis"). The hypothetical situation in which there is no difference or effect is the null hypothesis.

Possible conclusions. When an obtained result would be extremely unlikely if the null hypothesis were true, then the null hypothesis is rejected and the research hypothesis is supported. If the obtained results are not very extreme, the study is inconclusive.

Conventional significance levels. Psychologists usually consider a result too extreme if it is less likely than the .05 or 5% significance level. Sometimes a more stringent cutoff of .01 or 1% is used. When a result is so extreme that researchers reject the null hypothesis, the result is said to be statistically significant.

Five steps of hypothesis testing:

❶ Restate the question as a research hypothesis and a null hypothesis about the populations.
❷ Determine the characteristics of the comparison distribution.
❸ Determine the cutoff sample score (or critical value) on the comparison distribution at which the null hypothesis should be rejected.
❹ Determine your sample's score on the comparison distribution.
❺ Decide whether to reject the null hypothesis.

One-tailed and two-tailed hypothesis tests. A directional hypothesis, assessed using a one-tailed test, predicts that the populations will differ in a particular direction (e.g., sample has a higher mean than the population). The region of rejection is in one side (tail) of the distribution. A nondirectional hypothesis, assessed using a two-tailed test, predicts that the populations will differ, but does not specify a particular direction. The region of rejection is in both sides (tails) of the distribution. To be cautious, researchers typically use two-tailed tests regardless of whether the hypothesis is directional or nondirectional.

Controversy: Should significance tests be banned? In recent years, there has been much debate about significance testing. The biggest complaint is that significance tests are often interpreted incorrectly and misused. Bayesian analysis has been suggested as an alternative to traditional significance testing.

How the procedures of this chapter are reported in research articles. Research articles typically report the results of hypothesis testing by noting the statistical test computed, whether the result was or was not statistically significant, and the probability level cutoff (usually 5% or 1%) at which the decision was made.

Box 4-1: Jacob Cohen, the ultimate New Yorker: Funny, pushy, brilliant, and kind. Provides a brief biography of Jacob Cohen and his impact on statistical practice (e.g., debate regarding arbitrary cutoffs for significance testing).

List of Transparencies

Lecture 4.1: Introduction to Hypothesis Testing

Materials

Lecture outline
Transparencies 3.5, 3.6, 3.11, 3.13, and 4.1 through 4.6

Outline for Blackboard

I. Review/Last Assignment
II. Inferential Statistics
III. Hypothesis Testing Example
IV. Hypothesis Testing Steps
V. Additional Examples
VI. Review This Class

Instructor's Lecture Outline

I. Review/Last Assignment

A. Idea of descriptive statistics and importance in their own right.

B. The normal curve.

1. Approximations to it are common in distributions used in psychology.
2. Relation between *Z* scores and percentages: Show TRANSPARENCY 3.5.
3. How to use normal curve tables: Show TRANSPARENCY 3.6.

C. Sample and population: Show TRANSPARENCY 3.11.

D. Probability: Show TRANSPARENCY 3.13.

II. Inferential Statistics

A. Involve making inferences about populations based on information from samples (as compared to descriptive statistics, which merely summarize known information).

B. Especially important because we use these statistics to draw conclusions about the world in general (i.e., populations we cannot measure as a whole) based on results from a particular group of people studied (i.e., a sample).

C. This and the next few classes use research examples that would almost never occur in order to introduce the key logic in the simplest possible context. (Later, you learn to apply the logic in more realistic research situations.)

III. Hypothesis Testing Example

A. A person claims that she can identify people of above-average intelligence with her eyes closed. A plan is made to take her to a stadium full of randomly selected people from the general population, and ask her to pick someone with her eyes closed.

B. It is known in advance that the distribution of intelligence on the test to be used is normal with $\mu = 100$ and $\sigma = 15$.

C. If she picks someone with:

1. An IQ of 145 ($Z = +3$), it is extremely unlikely to have been by chance. (From normal curve table, probability is .13% or .0013.)
2. An IQ of 130 ($Z = +2$), there is only about a 2% probability of getting this result by chance. We would probably be convinced.
3. An IQ of 115 ($Z = +1$), there is about a 16% probability this was a chance result. We would probably not be convinced.

D. Thus, we set a score by which we will be convinced *in advance*. We might set:

1. A probability level of about 2%... in this case, an IQ of about 130.
2. A probability level of 1%... from the normal curve table, closest *Z* score is 2.33, which is an IQ of about 135.
3. A probability level of 5%... from the normal curve table, closest *Z* score is 1.64, which is equivalent to an IQ of $(1.64)(15) + 100 = 24.6 + 100 = 124.6$.

IV. Hypothesis Testing Steps

A. We conduct the aforementioned experiment described above (in Part III). The woman is blindfolded, taken to the stadium, and chooses someone with an IQ of 140.

1. Show TRANSPARENCY 4.1: Explain logic and language at each step.
2. Show TRANSPARENCY 4.2: Review normal curve illustration.
3. This should convince us that she has done something that would not result from chance. (That is, she has some special ability, is employing a hoax, etc.)

B. Let's consider some alternatives:
   1. Suppose she picks someone with an IQ of 85? We would clearly not be convinced.
   2. Suppose she picks someone with an IQ of 115? This result is inconclusive—she did what she said she could do, but it might have just happened by chance.

V. Additional Examples

A. Lottery winner example: Show and discuss TRANSPARENCIES 4.3 and 4.4.

B. Stress reduction example: Show and discuss TRANSPARENCIES 4.5 and 4.6.

VI. Review This Class: Use Blackboard outline.

Lecture 4.2: Significance Levels and Directional Tests

Materials

Lecture outline
Transparencies 3.5, 4.3, 4.4, and 4.6 through 4.11

Outline for Blackboard

I. Review/Last Assignment
II. Significance Levels
III. One-Tailed and Two-Tailed Tests
IV. Review This Class

Instructor's Lecture Outline

I. Review/Last Assignment

A. Relation between *Z* scores and percentages: Show TRANSPARENCY 3.5.
B. Basic steps of hypothesis testing. Show TRANSPARENCIES 4.3 and 4.4.

II. Significance Levels

A. Significance level represented as cutoff in Step 3 of hypothesis testing process.
B. The lower the level (smaller percentages), the more sure you are about rejecting the null hypothesis; the result is quite extreme. Example: If you used 1%, you would be very sure that with a result that extreme, you are correct in rejecting the null hypothesis.
C. The higher the level (larger percentages), the less sure you are about rejecting the null hypothesis; you may have to deal with inconclusive results. Example: If you used 25%, almost any result you predicted would permit you to reject the null hypothesis.
D. So psychologists usually use 5% as a conventional compromise. Some prefer 1%.

E. How results are expressed:

1. A result more extreme than the set significance level is *statistically significant*.
2. The 5% level is often expressed as .05, and the 1% level as .01.
3. A significant result may be described as $p < .05$.
   a. p is for the probability of getting your result (a result as extreme as you got) if the null hypothesis is true.
   b. $p < .05$ means that the probability of getting a result this extreme if the null hypothesis is true is less than .05 or 5%.
4. Despite the original cutoff, sometimes researchers give more extreme p levels (meaning, lower percentages) to describe results that are very extreme.

III. One-Tailed and Two-Tailed Tests

A. A researcher's prediction is often that those people receiving some experimental treatment will score higher (or lower) than those not receiving it.

1. Show TRANSPARENCY 4.4:
   a. The prediction was that those receiving $10 million would be happier than the general population.
   b. This is a directional hypothesis because the researcher predicts the direction of the result (in this case, more happiness).
   c. It is also called a one-tailed test because the extreme result should be at just one end of the distribution (in this case, the top or high end).
2. Show TRANSPARENCY 4.6:
   a. The prediction was that those receiving the stress-reduction training would be less stressed that the general population.
   b. This is another case of a directional hypothesis (and one-tailed test). The extreme result in this case should be at the bottom or low end of the distribution.

B. Sometimes, however, a researcher predicts that an experimental treatment will make a difference, but does not know whether it will create higher or lower scores.

1. Show TRANSPARENCY 4.7:
   a. This is a nondirectional hypothesis because no direction is specified.
   b. This is also called a two-tailed test because extreme results at either tail (positive or negative) would make you want to reject the null hypothesis (of no difference).
   c. In setting the cutoff for rejecting the null hypothesis, the overall probability (say 5%) must be divided between the two tails (say 2.5% at each).
   d. This dividing means that, to be significant, a score must be more extreme than if a one-tailed test was used.
2. Depression example: Show and discuss TRANSPARENCIES 4.8 and 4.9.

3. Polluted region example: Show and discuss TRANSPARENCIES 4.10 and 4.11.

C. When to use one-tailed and two-tailed tests:

1. If you use a one-tailed test and the result comes out in an unexpected direction, it cannot be considered significant no matter how extreme.
2. To avoid this problem, researchers generally prefer to use two-tailed tests except where it is very clear that only one direction would be of interest. Example: test of new procedure where if it did not work or made things worse, the procedure would not be used.
3. Some researchers, however, use one-tailed tests whenever there is any basis for making the prediction.
4. The whole topic is very controversial.

IV. **Review This Class:** Use Blackboard outline.

TRANSPARENCY 4.1

Intelligence Example
(fictional data)

A person claims that she can identify people of above-average intelligence with her eyes closed. It is known in advance that the distribution of intelligence in the general population is normal with $\mu = 100$ and $\sigma = 15$. The woman is blindfolded, taken to a stadium full of randomly selected people, and chooses someone with an IQ of 140. Could this result have occurred by chance?

❶ **Restate question as a research hypothesis and a null hypothesis about the populations.**

Population 1:	People chosen by the woman with her eyes closed.
Population 2:	The general population (consisting of people who were not chosen by the woman with her eyes closed).
Research hypothesis:	Population 1 will be more intelligent than Population 2. ($\mu_1 > \mu_2$)
Null hypothesis:	Population 1 will not be more intelligent than Population 2. ($\mu_1 \leq \mu_2$)

❷ **Determine the characteristics of the comparison distribution.**

Known normal distribution, with $\mu = 100$ and $\sigma = 15$.

❸ **Determine the cutoff sample score (critical value) on the comparison distribution at which the null hypothesis should be rejected.**

For a .05 probability (top 5% of comparison distribution), Z needed is +1.64.
$X = (Z)(SD) + M \rightarrow X = (1.64)(15) + 100 = 24.6 + 100 = 124.6$, so X needed is 124.6.

❹ **Determine your sample's score on the comparison distribution.**

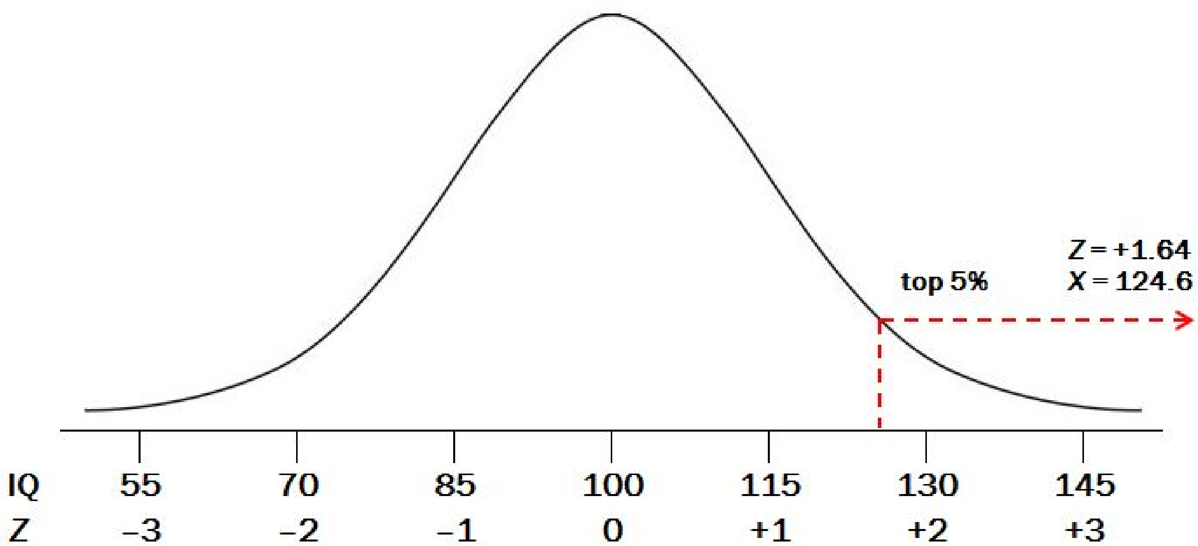
An individual with an IQ of 140 is selected. ($X = 140$)
$Z = (X - M) / SD \rightarrow Z = (140 - 100) / 15 = 40 / 15 = 2.67$. ($Z = +2.67$)

❺ **Decide whether to reject the null hypothesis.**

Raw score at Step 4 ($X = 140$) is more extreme than raw score at Step 3 ($X = 124.6$).
Z score at Step 4 ($Z = +2.67$) is more extreme than Z score at Step 3 ($Z = +1.64$).
Therefore, **reject** null hypothesis; the research hypothesis is supported.

TRANSPARENCY 4.2

Intelligence Example
(fictional data)



TRANSPARENCY 4.3

Lottery Winner Example
(from text)

A study is done in which a randomly selected person is given $10 million. This person's happiness, measured 6 months later, is 80. It is known in advance that happiness in the general population is normally distributed with $\mu = 70$ and $\sigma = 10$. Could this result have occurred by chance?

❶ **Restate question as a research hypothesis and a null hypothesis about the populations.**

Population 1:	People who 6 months ago received $10 million.
Population 2:	The general population (consisting of people who 6 months ago did not receive $10 million).

Research hypothesis:	Population 1 will be happier than Population 2. ($\mu_1 > \mu_2$)
Null hypothesis:	Population 1 will not be happier than Population 2. ($\mu_1 \leq \mu_2$)

❷ **Determine the characteristics of the comparison distribution.**

Known normal distribution, with $\mu = 70$ and $\sigma = 10$.

❸ **Determine the cutoff sample score (critical value) on the comparison distribution at which the null hypothesis should be rejected.**

For a .05 probability (top 5% of comparison distribution), Z needed is +1.64.
$X = (Z)(SD) + M \rightarrow X = (1.64)(10) + 70 = 16.4 + 70 = 86.4$, so X needed is 86.4.

❹ **Determine your sample's score on the comparison distribution.**

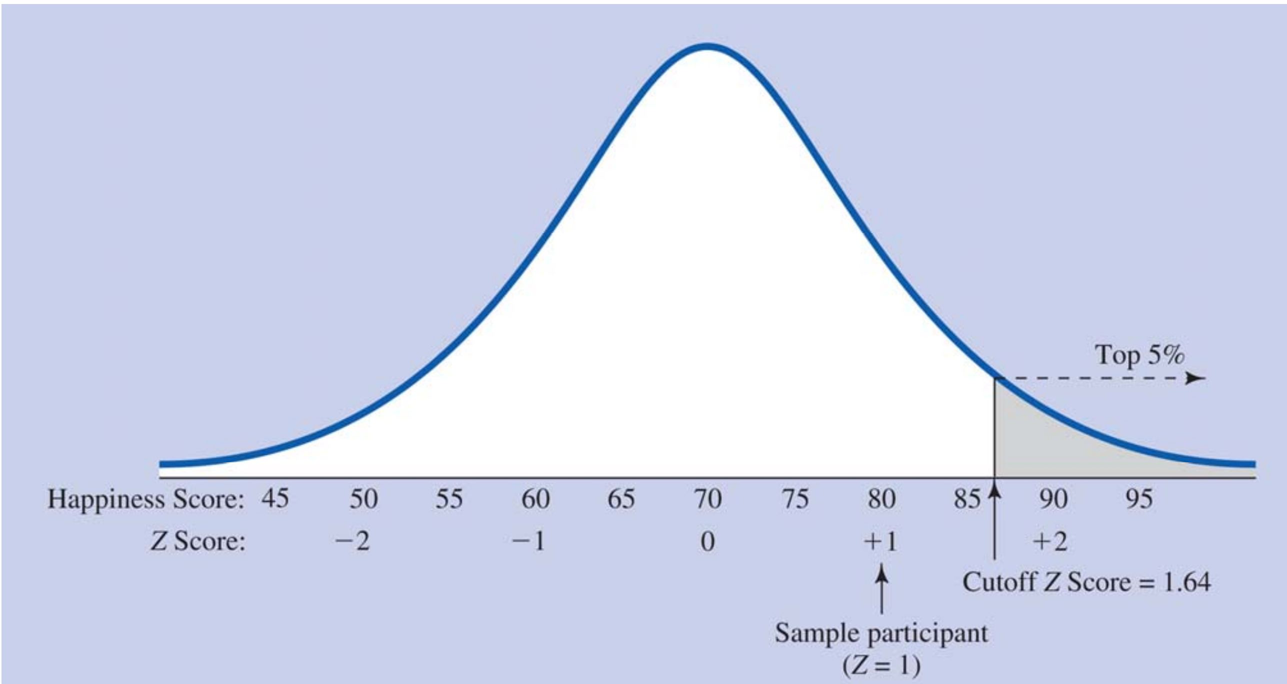
The person's score is 80 on the happiness test. ($X = 80$)
$Z = (X - M) / SD \rightarrow Z = (80 - 70) / 10 = 10 / 10 = 1$. ($Z = +1.00$)

❺ **Decide whether to reject the null hypothesis.**

Score at Step 4 ($Z = +1.00$) is not more extreme than score at Step 3 ($Z = +1.64$). Therefore, **do not reject** null hypothesis; the result is inconclusive.

TRANSPARENCY 4.4

Lottery Winner Example
(from text)



TRANSPARENCY 4.5

Stress Reduction Example
(fictional data)

A health psychologist wants to test the effectiveness of a new stress-reduction method. In the general population, stress level is normally distributed with $\mu = 40$ and $\sigma = 10$. A randomly selected person is trained in the method; the person's score afterward is 25.

❶ **Restate question as a research hypothesis and a null hypothesis about the populations.**

Population 1: People trained in the stress-reduction method.
Population 2: The general population (consisting of people who are not trained in the stress-reduction method).

Research hypothesis: Population 1 will have less stress than Population 2. ($\mu_1 < \mu_2$)
Null hypothesis: Population 1 will not have less stress than Population 2. ($\mu_1 \geq \mu_2$)

❷ **Determine the characteristics of the comparison distribution.**

Known normal distribution, with $\mu = 40$ and $\sigma = 10$.

❸ **Determine the cutoff sample score (critical value) on the comparison distribution at which the null hypothesis should be rejected.**

For a .01 probability (bottom 1% of comparison distribution), Z needed is –2.33.
$X = (Z)(SD) + M \rightarrow X = (-2.33)(10) + 40 = -23.3 + 40 = 16.7$, so X needed is 16.7.

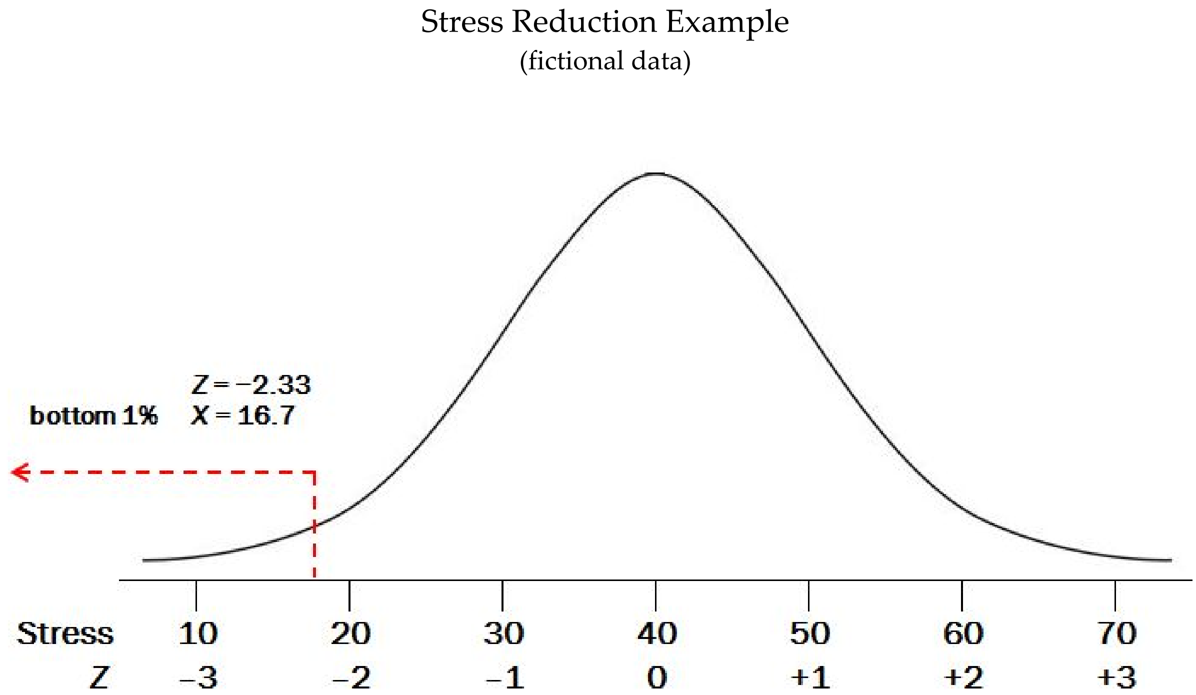
❹ **Determine your sample's score on the comparison distribution.**

The person's score is 25 on the stress test. ($X = 25$)
$Z = (X - M) / SD \rightarrow Z = (25 - 40) / 10 = -15 / 10 = -1.5$. ($Z = -1.50$)

❺ **Decide whether to reject the null hypothesis.**

Score at Step 4 ($Z = -1.50$) is not more extreme than score at Step 3 ($Z = -2.33$). Therefore, **do not reject** null hypothesis; the result is inconclusive.

TRANSPARENCY 4.6



TRANSPARENCY 4.7

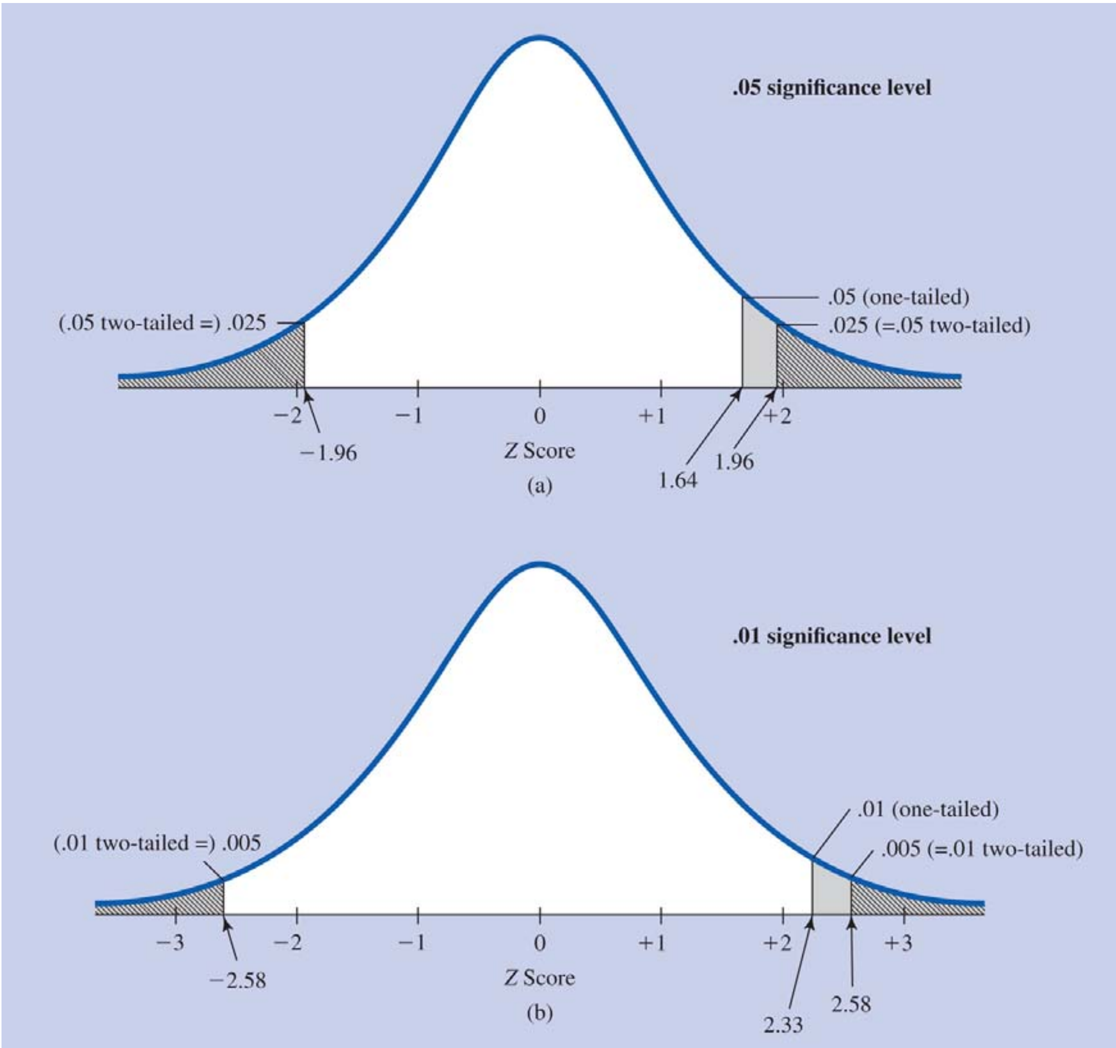


Table 4-2 One-Tailed and Two-Tailed Cutoff *Z* Scores for the .05 and .01 Significance Levels

		Type of Test	
		One-Tailed	**Two-Tailed**
Significance	*.05*	−1.64 *or* 1.64	−1.96 *and* 1.96
Level	*.01*	−2.33 *or* 2.33	−2.58 *and* 2.58

TRANSPARENCY 4.8

Depression Example
(from text)

Clinical psychologists developed a new type of therapy to reduce depression. It could make patients' depression better or worse. The psychologists randomly select an incoming patient to receive the new form of therapy instead of the usual therapy. After 4 weeks, the patient scores 41 on a standard depression scale. For those who receive the usual therapy, their depression after 4 weeks is normally distributed with $\mu = 69.5$ and $\sigma = 14.1$.

❶ **Restate question as a research hypothesis and a null hypothesis about the populations.**

Population 1:	Patients diagnosed as depressed who receive the new therapy.
Population 2:	Patients diagnosed as depressed in general (who receive the usual therapy).

Research hypothesis:	Population 1 will not have the same amount of depression as Population 2. ($\mu_1 \neq \mu_2$)
Null hypothesis:	Population 1 will have the same amount of depression as Population 2. ($\mu_1 = \mu_2$)

❷ **Determine the characteristics of the comparison distribution.**

Known normal distribution, with $\mu = 69.5$ and $\sigma = 14.1$.

❸ **Determine the cutoff sample score (critical value) on the comparison distribution at which the null hypothesis should be rejected.**

For a two-tailed test at the .05 significance level (top 2.5% and bottom 2.5% of comparison distribution), Z needed is above +1.96 or below –1.96.

❹ **Determine your sample's score on the comparison distribution.**

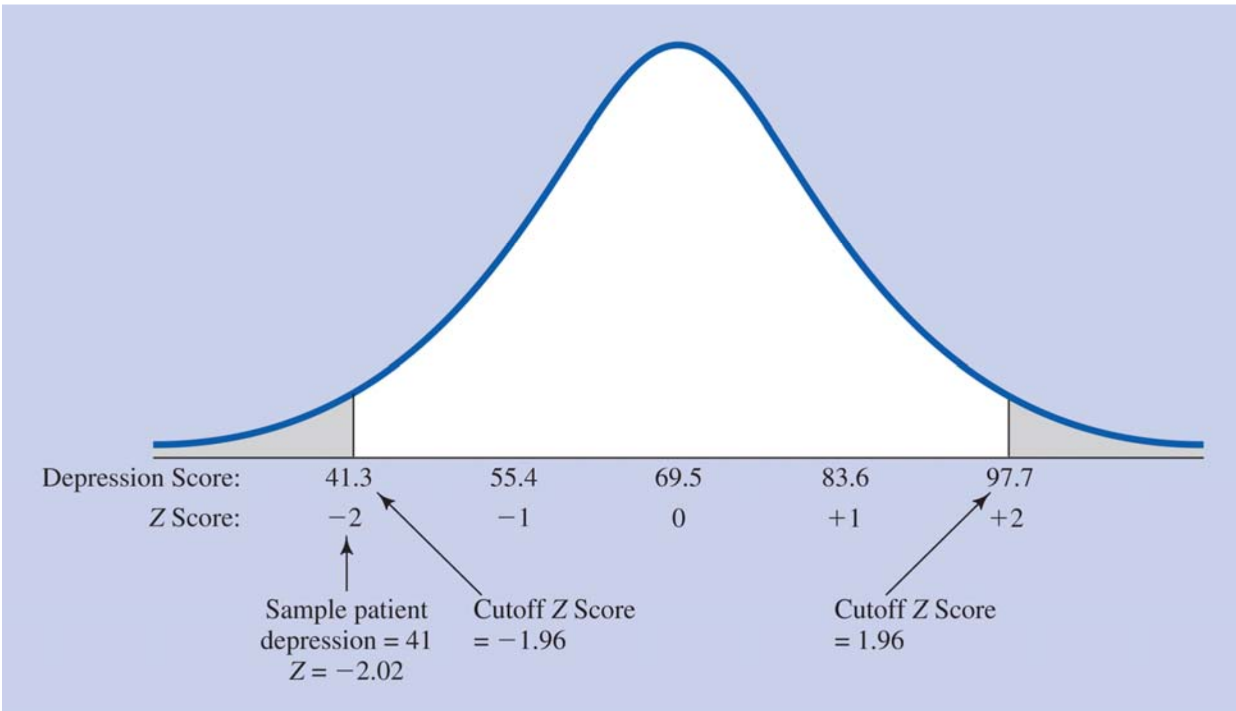
The patient scores 41 on the depression test. ($X = 41$)
$Z = (X - M) / SD \rightarrow Z = (41 - 69.5) / 14.1 = -28.5 / 14.1 = -2.02$. ($Z = -2.02$)

❺ **Decide whether to reject the null hypothesis.**

Score at Step 4 ($Z = -2.02$) is more extreme than score at Step 3 representing the bottom 2.5% of the distribution ($Z = -1.96$). Therefore, **reject** null hypothesis; the research hypothesis is supported. Patients receiving the new therapy have different depression levels than other patients.

TRANSPARENCY 4.9

Depression Example
(from text)



TRANSPARENCY 4.10

Polluted Region Example
(fictional data)

Does pollution affect amount of sleep? A person living in a polluted region was randomly selected; his sleep the night before was 10.2 hours. In the general population, amount of sleep is normally distributed with $\mu = 8$ and $\sigma = 1$.

❶ **Restate question as a research hypothesis and a null hypothesis about the populations.**

Population 1:	People who live in a polluted region.
Population 2:	People in general (who do not live in a polluted region).
Research hypothesis:	Population 1 will not have the same amount of sleep as Population 2. ($\mu_1 \neq \mu_2$)
Null hypothesis:	Population 1 will have the same amount of sleep as Population 2. ($\mu_1 = \mu_2$)

❷ **Determine the characteristics of the comparison distribution.**

Known normal distribution, with $\mu = 8$ and $\sigma = 1$.

❸ **Determine the cutoff sample score (critical value) on the comparison distribution at which the null hypothesis should be rejected.**

For a two-tailed test at the .05 significance level (top 2.5% and bottom 2.5% of comparison distribution), Z needed is above +1.96 or below –1.96.

❹ **Determine your sample's score on the comparison distribution.**

Person sampled from a polluted region slept 10.2 hours. ($X = 10.2$)
$Z = (X - M) / SD$ → $Z = (10.2 - 8) / 1 = 2.2 / 1 = 2.20$. ($Z = +2.20$)

❺ **Decide whether to reject the null hypothesis.**

Score at Step 4 ($Z = +2.20$) is more extreme than score at Step 3 representing the top 2.5% of the distribution ($Z = +1.96$). Therefore, **reject** null hypothesis; the research hypothesis is supported. People living in the polluted area sleep a different amount than people in general.

TRANSPARENCY 4.11

Polluted Region Example
(fictional data)