

Introduction to Hypothesis Testing

Hypothesis Testing

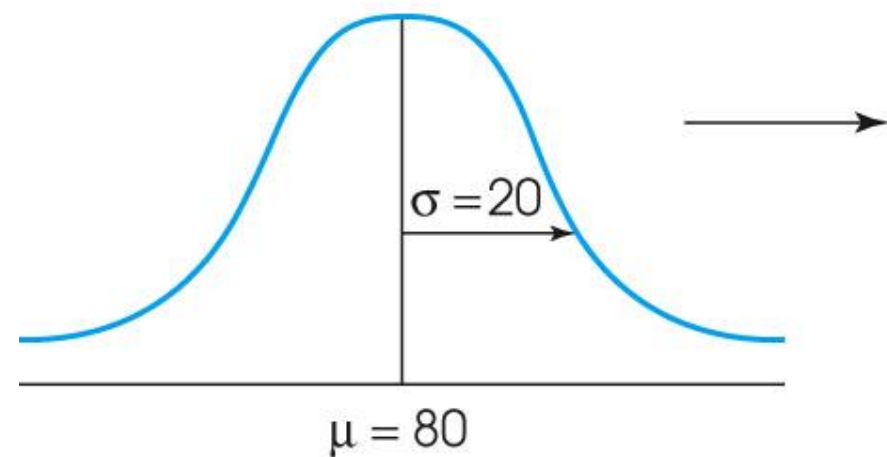
- The general goal of a hypothesis test is to rule out chance (sampling error) as a plausible explanation for the results from a research study.
- Hypothesis testing is a technique to help determine whether a specific treatment has an effect on the individuals in a population.

Hypothesis Testing

The hypothesis test is used to evaluate the results from a research study in which

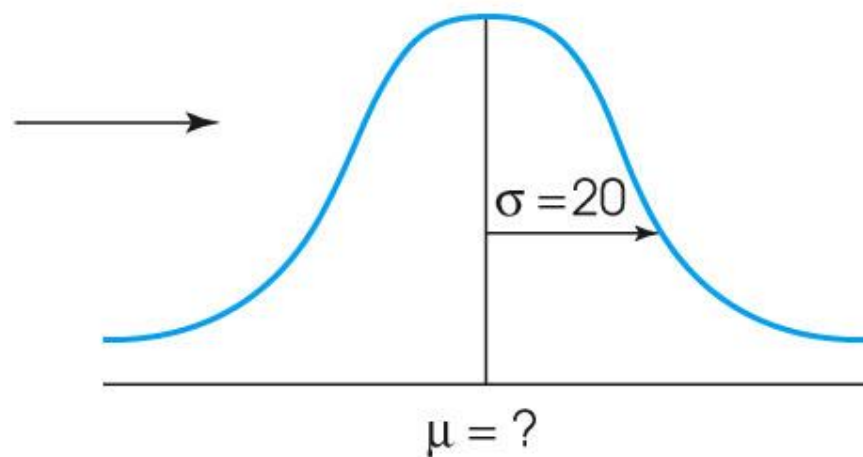
1. A sample is selected from the population.
2. The treatment is administered to the sample.
3. After treatment, the individuals in the sample are measured.

Known population
before treatment



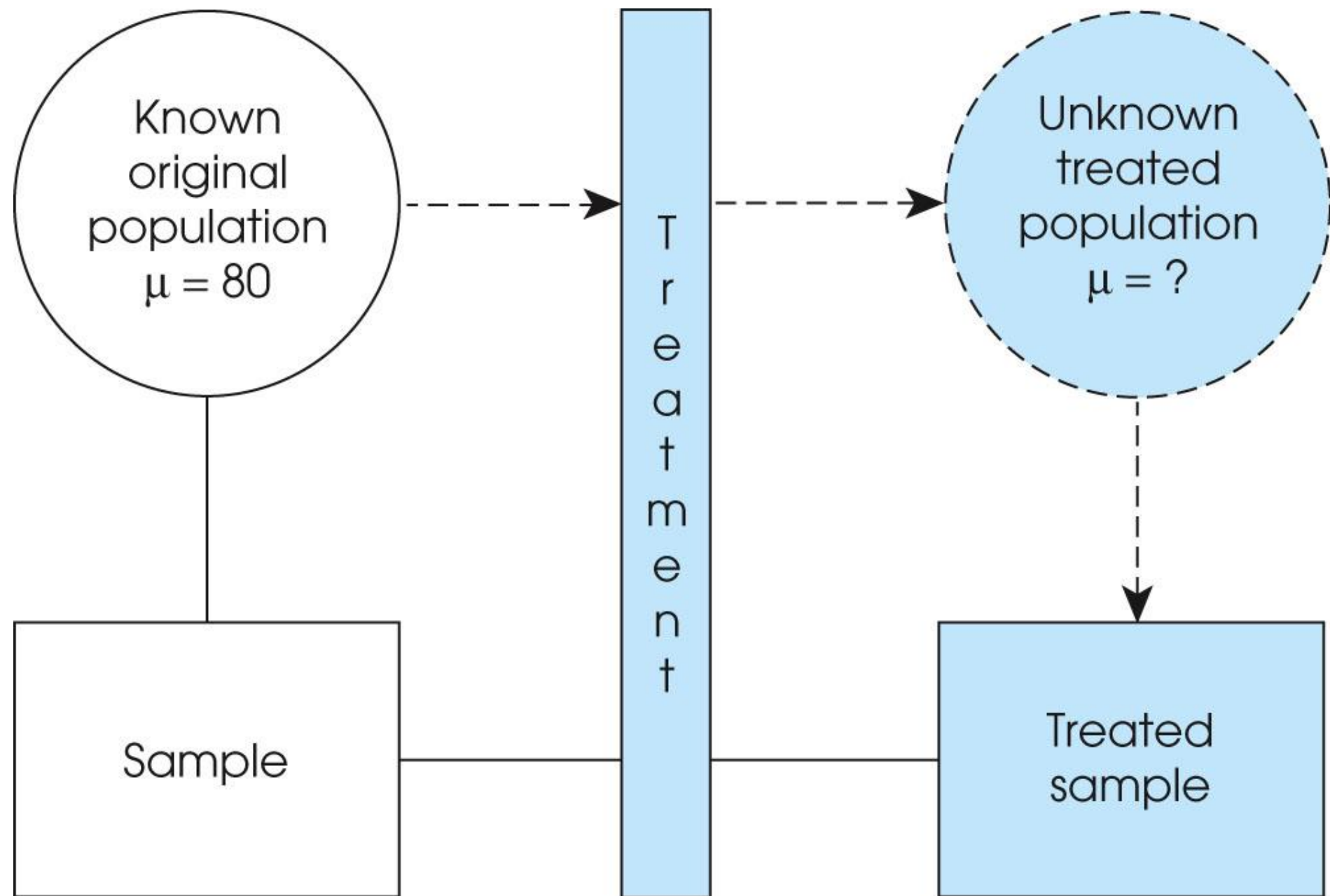
T
r
e
a
t
m
e
n
t

Unknown population
after treatment



Hypothesis Testing (cont.)

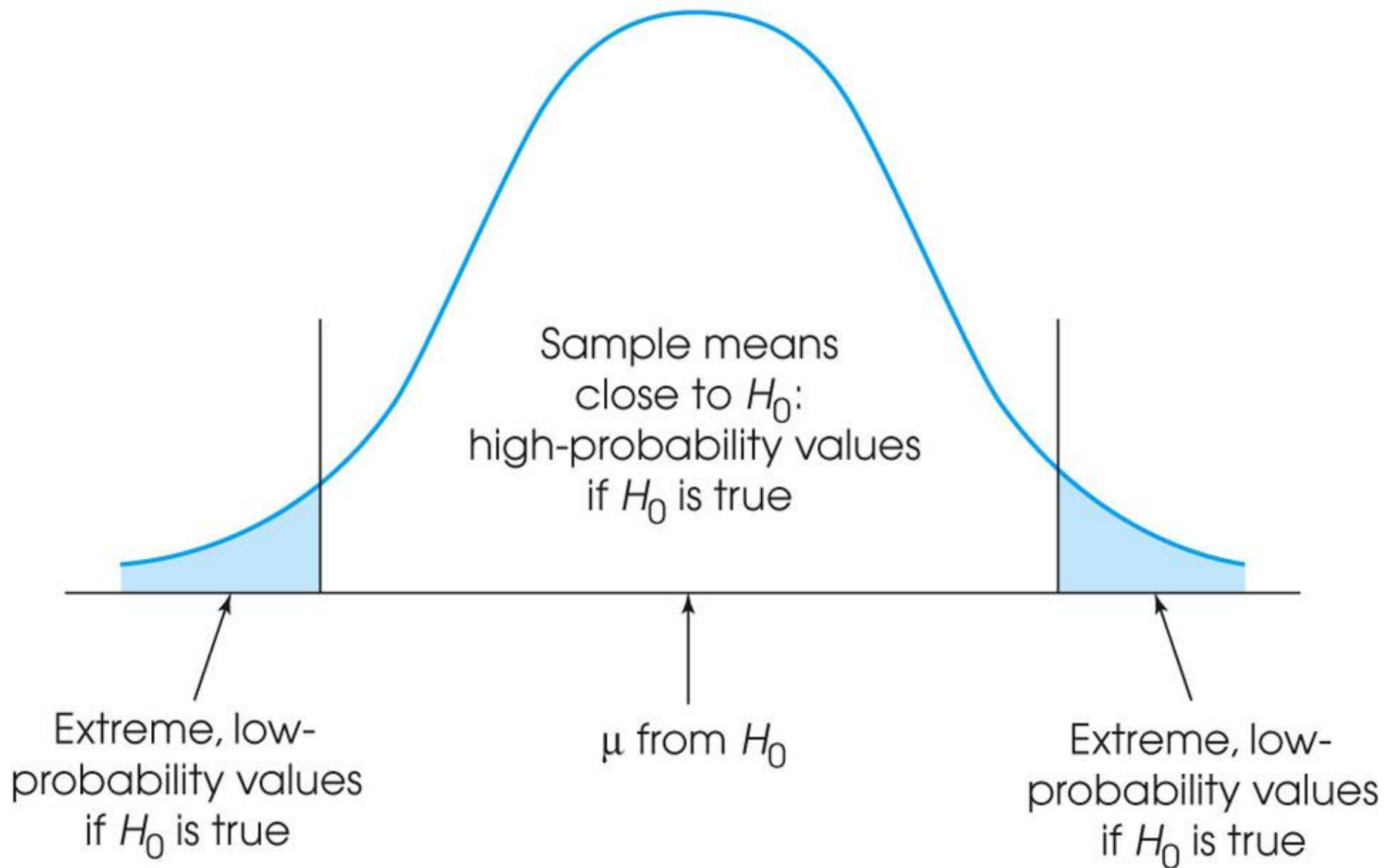
- If the individuals in the sample are noticeably different from the individuals in the original population, we have evidence that the treatment has an effect.
- However, it is also possible that the difference between the sample and the population is simply sampling error



Hypothesis Testing (cont.)

- The purpose of the hypothesis test is to decide between two explanations:
 1. The difference between the sample and the population can be explained by sampling error (there does not appear to be a treatment effect)
 2. The difference between the sample and the population is too large to be explained by sampling error (there does appear to be a treatment effect).

The distribution of sample means
if the null hypothesis is true
(all the possible outcomes)

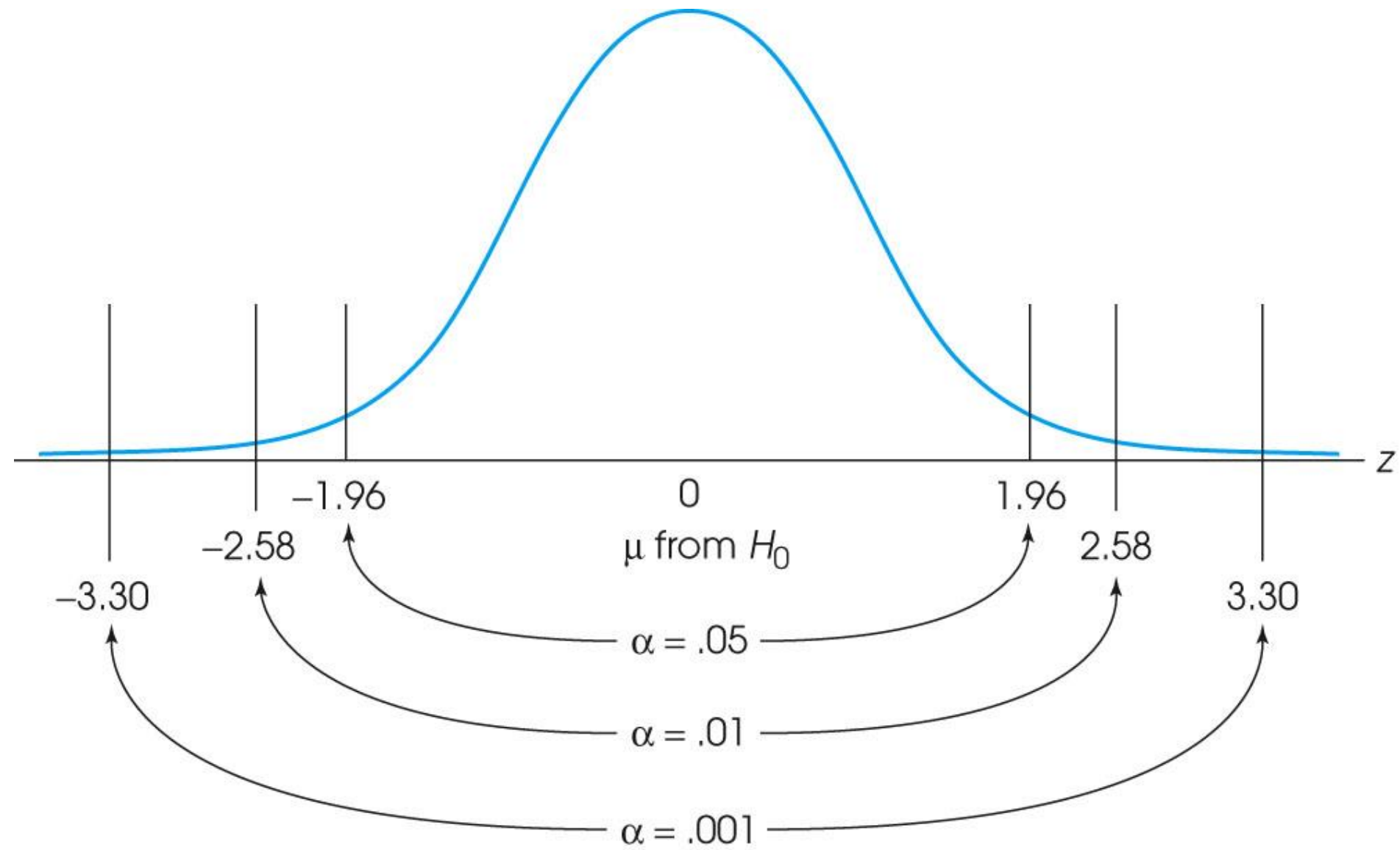


The Null Hypothesis, the Alpha Level, the Critical Region, and the Test Statistic

- The following four steps outline the process of hypothesis testing and introduce some of the new terminology:

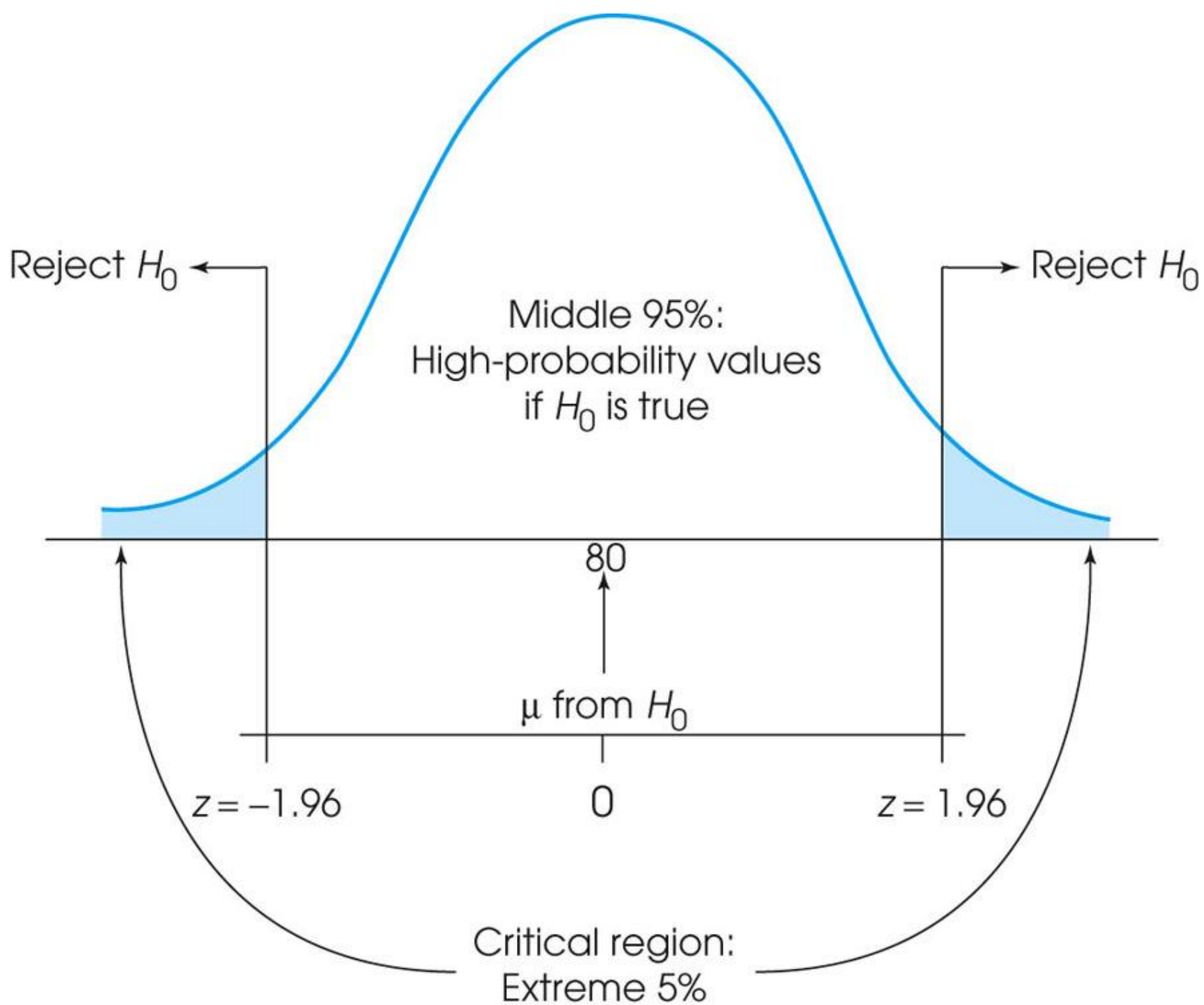
Step 1

State the hypotheses and select an α level. The **null hypothesis**, H_0 , always states that the treatment has no effect (no change, no difference). According to the null hypothesis, the population mean after treatment is the same as it was before treatment. The **α level** establishes a criterion, or "cut-off", for making a decision about the null hypothesis. The alpha level also determines the risk of a Type I error.



Step 2

Locate the critical region. The **critical region** consists of outcomes that are very unlikely to occur if the null hypothesis is true. That is, the critical region is defined by sample means that are almost impossible to obtain if the treatment has no effect. The phrase “almost impossible” means that these samples have a probability (p) that is less than the alpha level.



Step 3

Compute the test statistic. The **test statistic** (in this chapter a z-score) forms a ratio comparing the obtained difference between the sample mean and the hypothesized population mean versus the amount of difference we would expect without any treatment effect (the standard error).

Step 4

A large value for the test statistic shows that the obtained mean difference is more than would be expected if there is no treatment effect. If it is large enough to be in the critical region, we conclude that the difference is **significant** or that the treatment has a significant effect. In this case we reject the null hypothesis. If the mean difference is relatively small, then the test statistic will have a low value. In this case, we conclude that the evidence from the sample is not sufficient, and the decision is fail to reject the null hypothesis.

Population
without
alcohol

$\mu = 18$
Normal
 $\sigma = 4$

Sample
 $n = 16$

A
l
c
o
h
o
l

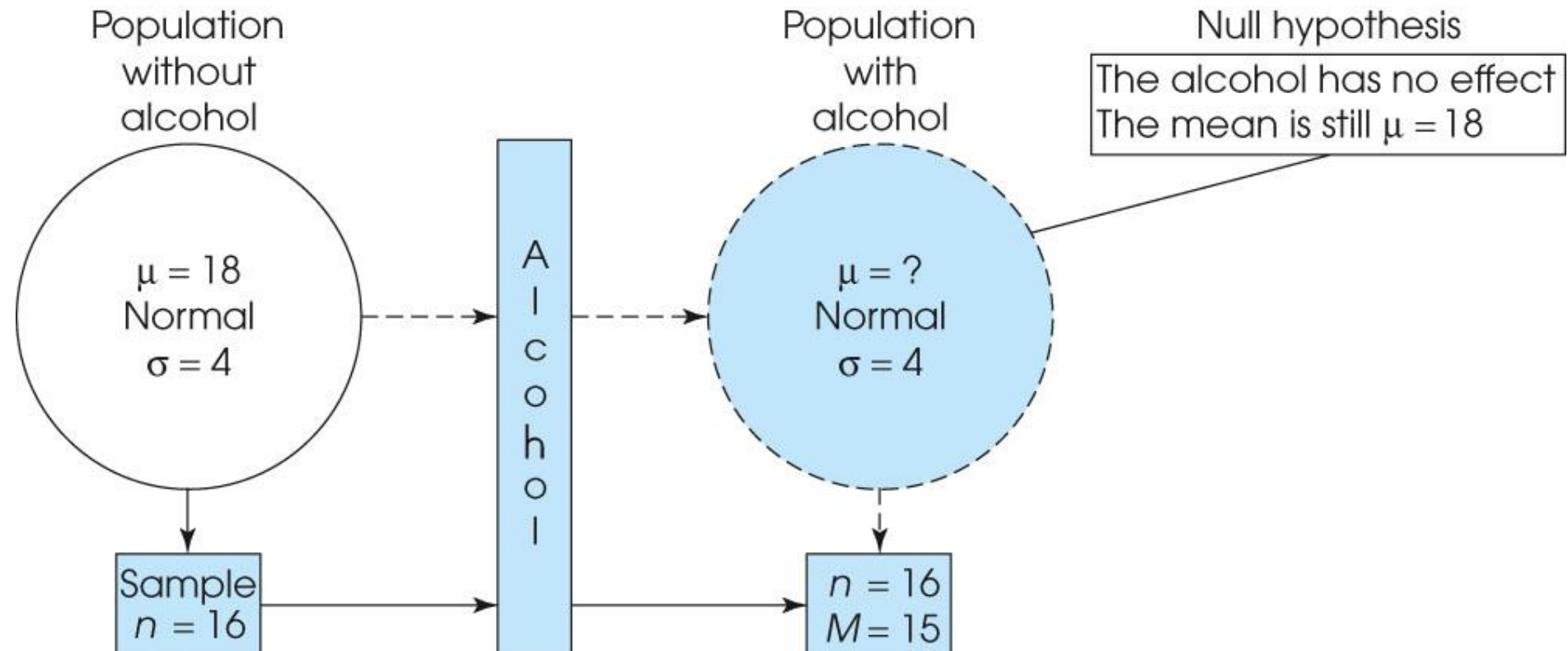
Population
with
alcohol

$\mu = ?$
Normal
 $\sigma = 4$

$n = 16$
 $M = 15$

Null hypothesis

The alcohol has no effect
The mean is still $\mu = 18$



Errors in Hypothesis Tests

- Just because the sample mean (following treatment) is different from the original population mean does not necessarily indicate that the treatment has caused a change.
- You should recall that there usually is some discrepancy between a sample mean and the population mean simply as a result of sampling error.

Errors in Hypothesis Tests (cont.)

- Because the hypothesis test relies on sample data, and because sample data are not completely reliable, there is always the risk that misleading data will cause the hypothesis test to reach a wrong conclusion.
- Two types of error are possible.

Type I Errors

- A **Type I error** occurs when the sample data appear to show a treatment effect when, in fact, there is none.
- In this case the researcher will reject the null hypothesis and falsely conclude that the treatment has an effect.
- Type I errors are caused by unusual, unrepresentative samples. Just by chance the researcher selects an extreme sample with the result that the sample falls in the critical region even though the treatment has no effect.
- The hypothesis test is structured so that Type I errors are very unlikely; specifically, the probability of a Type I error is equal to the alpha level.

Type II Errors

- A **Type II error** occurs when the sample does not appear to have been affected by the treatment when, in fact, the treatment does have an effect.
- In this case, the researcher will fail to reject the null hypothesis and falsely conclude that the treatment does not have an effect.
- Type II errors are commonly the result of a very small treatment effect. Although the treatment does have an effect, it is not large enough to show up in the research study.

TABLE 8.1

Possible outcomes of a statistical decision

		Actual Situation	
		No Effect, H_0 True	Effect Exists, H_0 False
EXPERIMENTER'S DECISION	Reject H_0	Type I error	Decision correct
	Retain H_0	Decision correct	Type II error

Directional Tests

- When a research study predicts a specific direction for the treatment effect (increase or decrease), it is possible to incorporate the directional prediction into the hypothesis test.
- The result is called a **directional test** or a **one-tailed test**. A directional test includes the directional prediction in the statement of the hypotheses and in the location of the critical region.

Directional Tests (cont.)

- For example, if the original population has a mean of $\mu = 80$ and the treatment is predicted to increase the scores, then the null hypothesis would state that after treatment:
$$H_0: \mu \leq 80 \quad (\text{there is no increase})$$
- In this case, the entire critical region would be located in the right-hand tail of the distribution because large values for M would demonstrate that there is an increase and would tend to reject the null hypothesis.

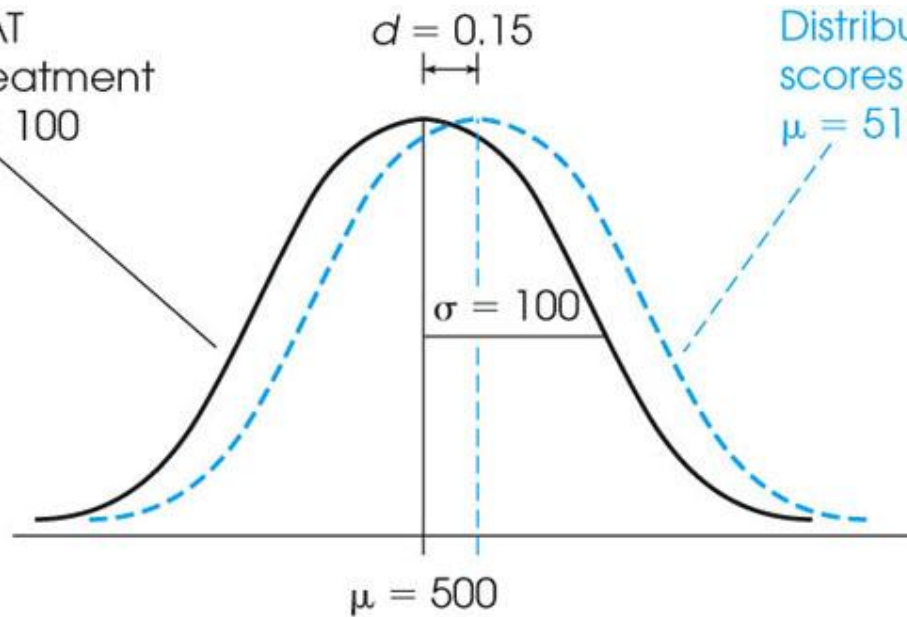
Measuring Effect Size

- A hypothesis test evaluates the *statistical significance* of the results from a research study.
- That is, the test determines whether or not it is likely that the obtained sample mean occurred without any contribution from a treatment effect.
- The hypothesis test is influenced not only by the size of the treatment effect but also by the size of the sample.
- Thus, even a very small effect can be significant if it is observed in a very large sample.

Measuring Effect Size

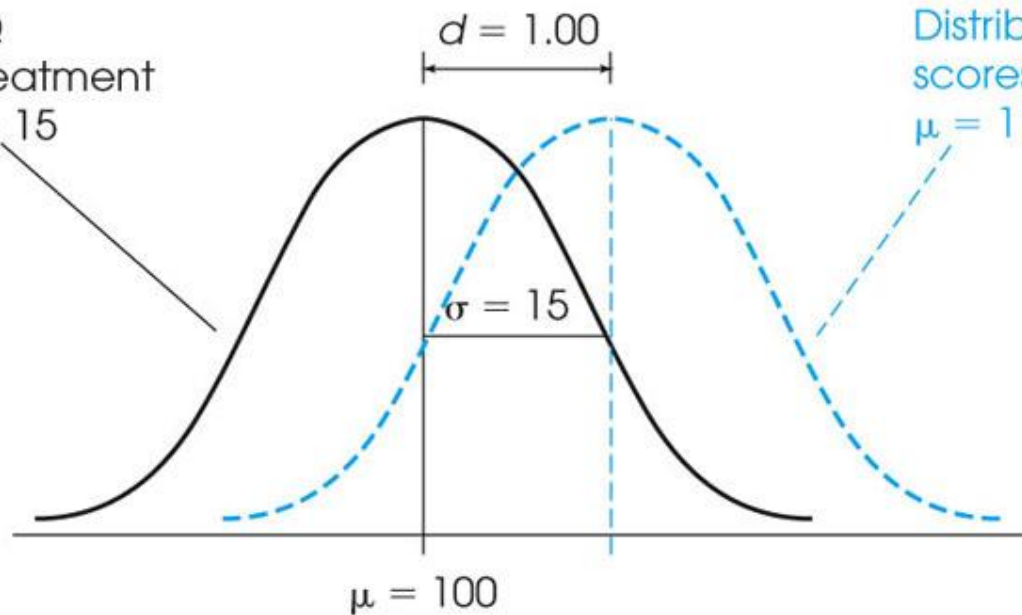
- Because a significant effect does not necessarily mean a large effect, it is recommended that the hypothesis test be accompanied by a measure of the **effect size**.
- We use Cohen's d as a standardized measure of effect size.
- Much like a z-score, **Cohen's d** measures the size of the mean difference in terms of the standard deviation.

Distribution of SAT
scores before treatment
 $\mu = 500$ and $\sigma = 100$



Distribution of SAT
scores after treatment
 $\mu = 515$ and $\sigma = 100$

Distribution of IQ
scores before treatment
 $\mu = 100$ and $\sigma = 15$



Distribution of IQ
scores after treatment
 $\mu = 115$ and $\sigma = 15$

Power of a Hypothesis Test

- The **power** of a hypothesis test is defined is the probability that the test will reject the null hypothesis when the treatment does have an effect.
- The power of a test depends on a variety of factors including the size of the treatment effect and the size of the sample.

