EMJ47703 IOT & DATA ANALYTICS

Hypothesis Testing

Hypothesis Testing

Learn about the hypothesis testing in data analytics

Able to describe the hypothesis testing in data analytics and apply in IoT application.

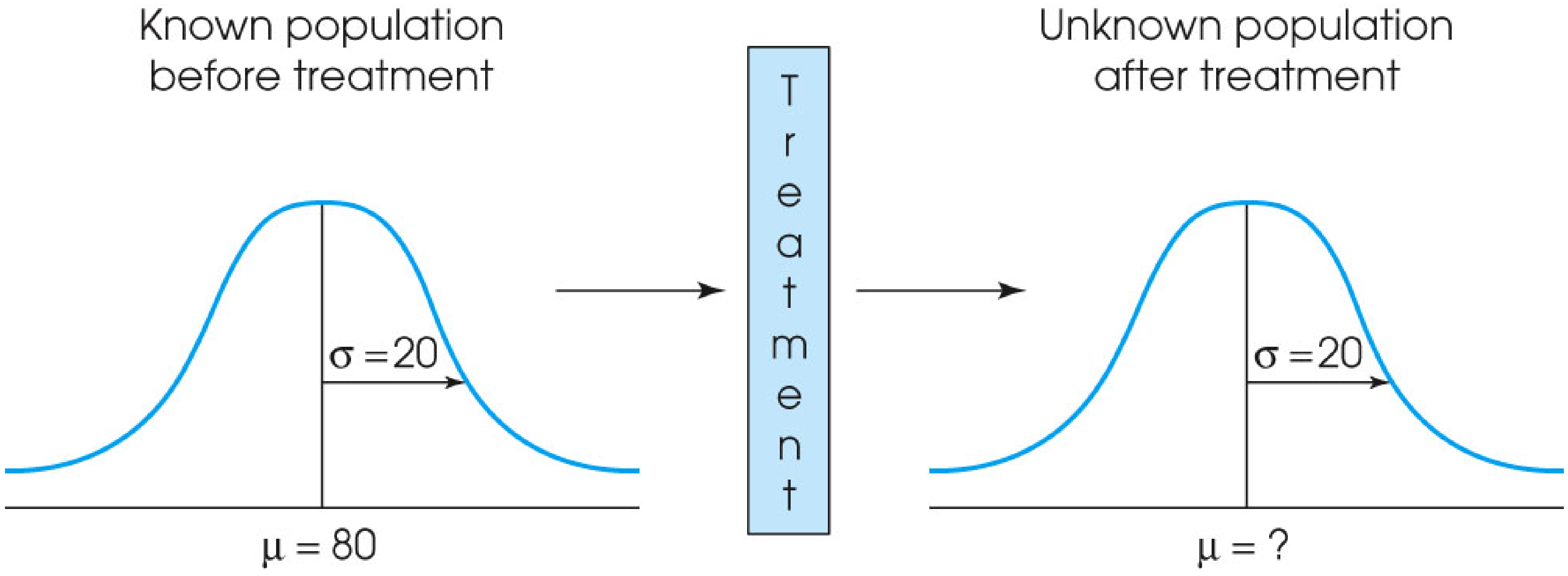
Hypothesis Testing

- 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

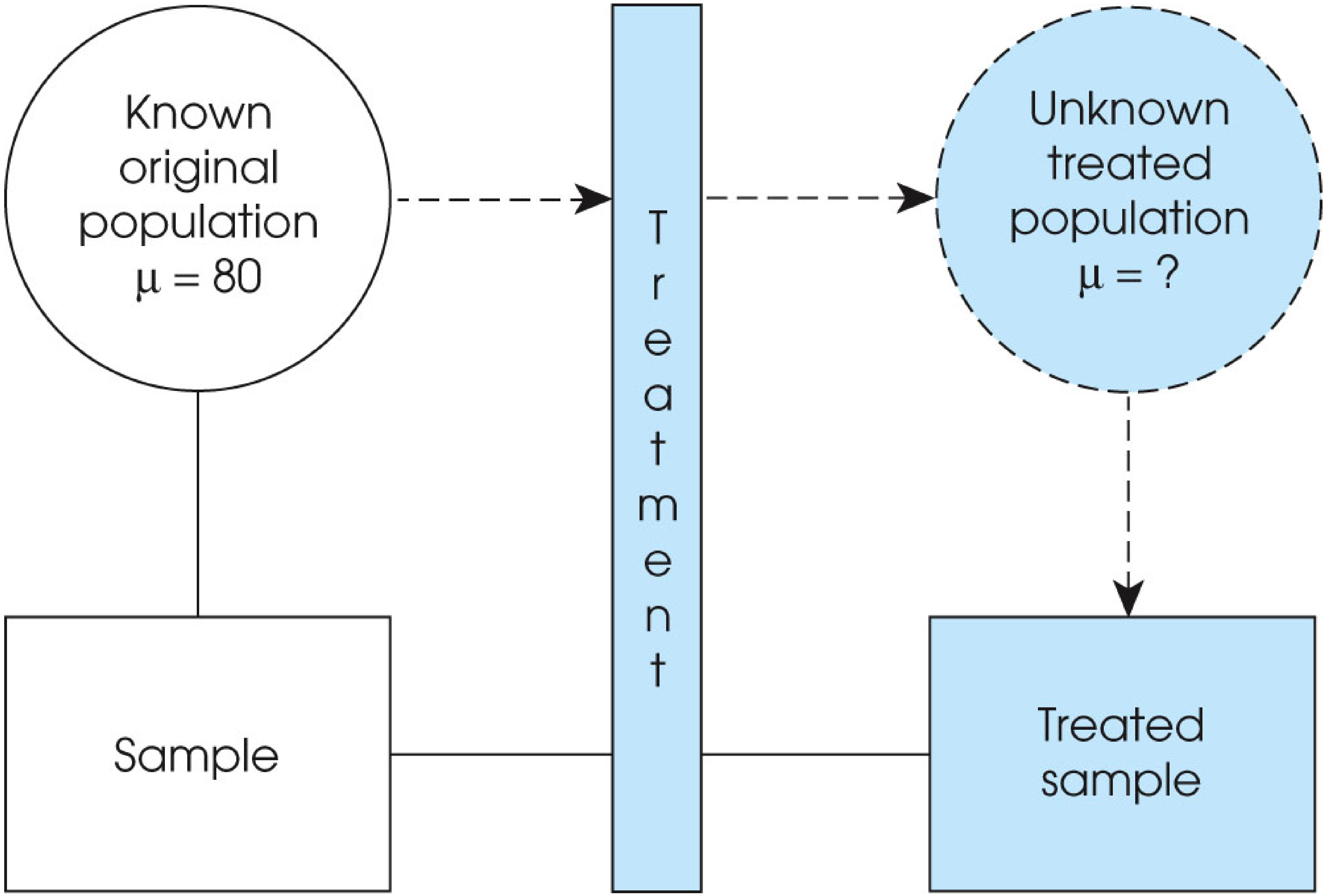
1. A sample is selected from the population.
2. The treatment is administered to the sample.
3. After treatment, the individuals in the sample are measured.



Known population
before treatment
σ = 20
μ = 80
Treatment
Unknown population
after treatment
σ = 20
μ = ?

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


Known original population μ = 80
Treatment
Unknown treated population μ = ?
Sample
Treated sample

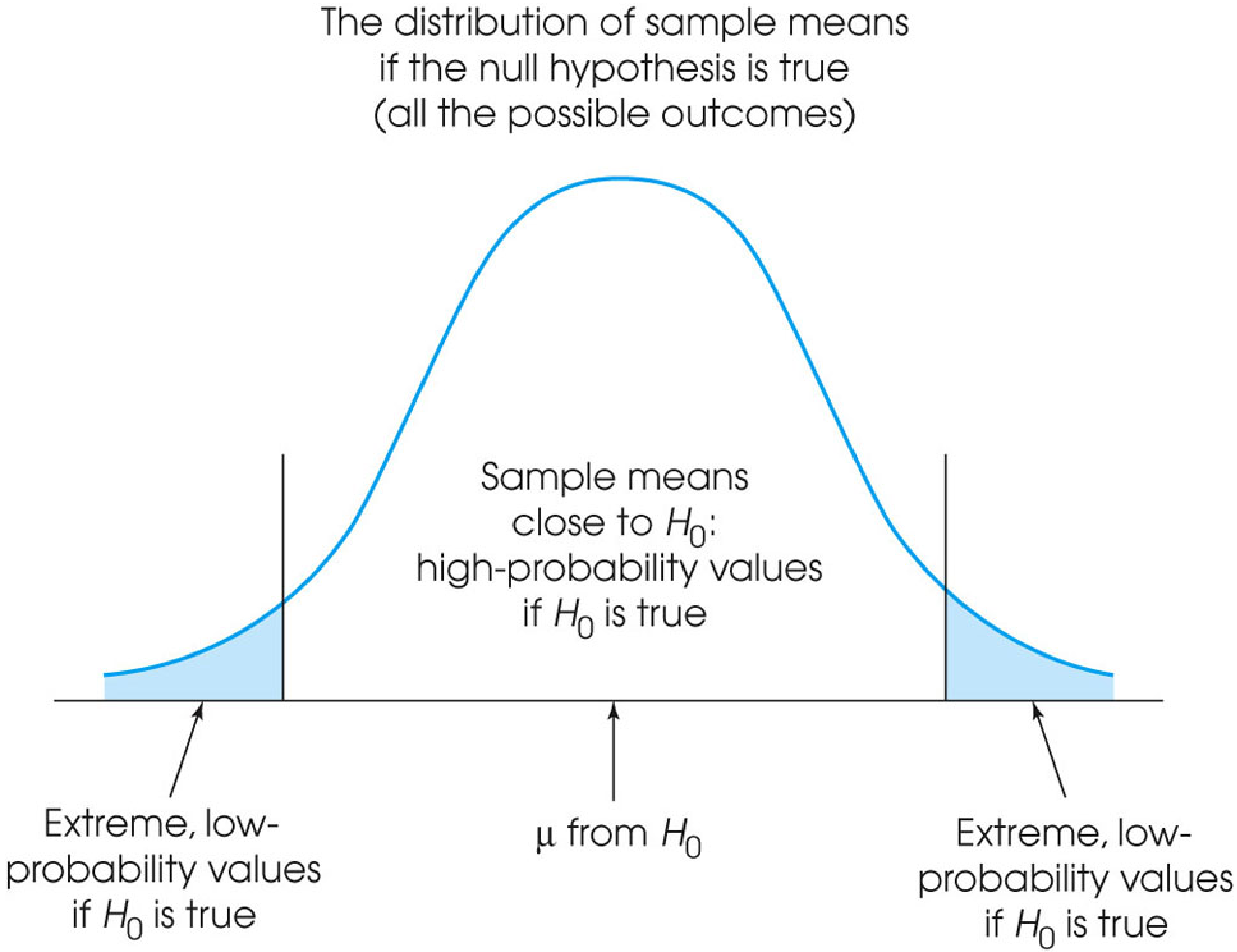
Hypothesis Testing (cont.)

- 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).

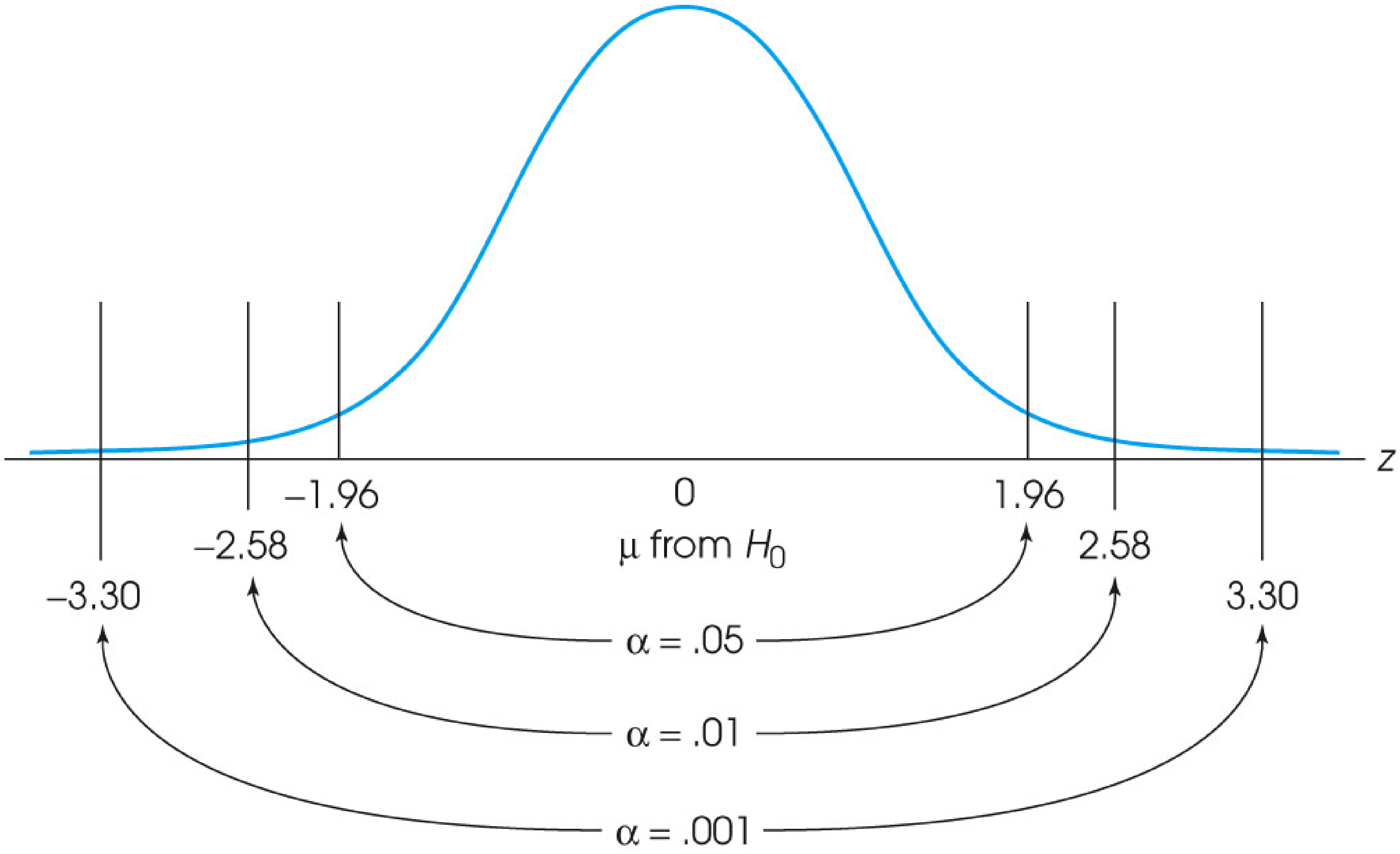

The distribution of sample means
if the null hypothesis is true
(all the possible outcomes)
Sample means
close to H_0:
high-probability values
if H_0 is true
Extreme, low-
probability values
if H_0 is true
μ from H_0
Extreme, low-
probability values
if H_0 is true

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

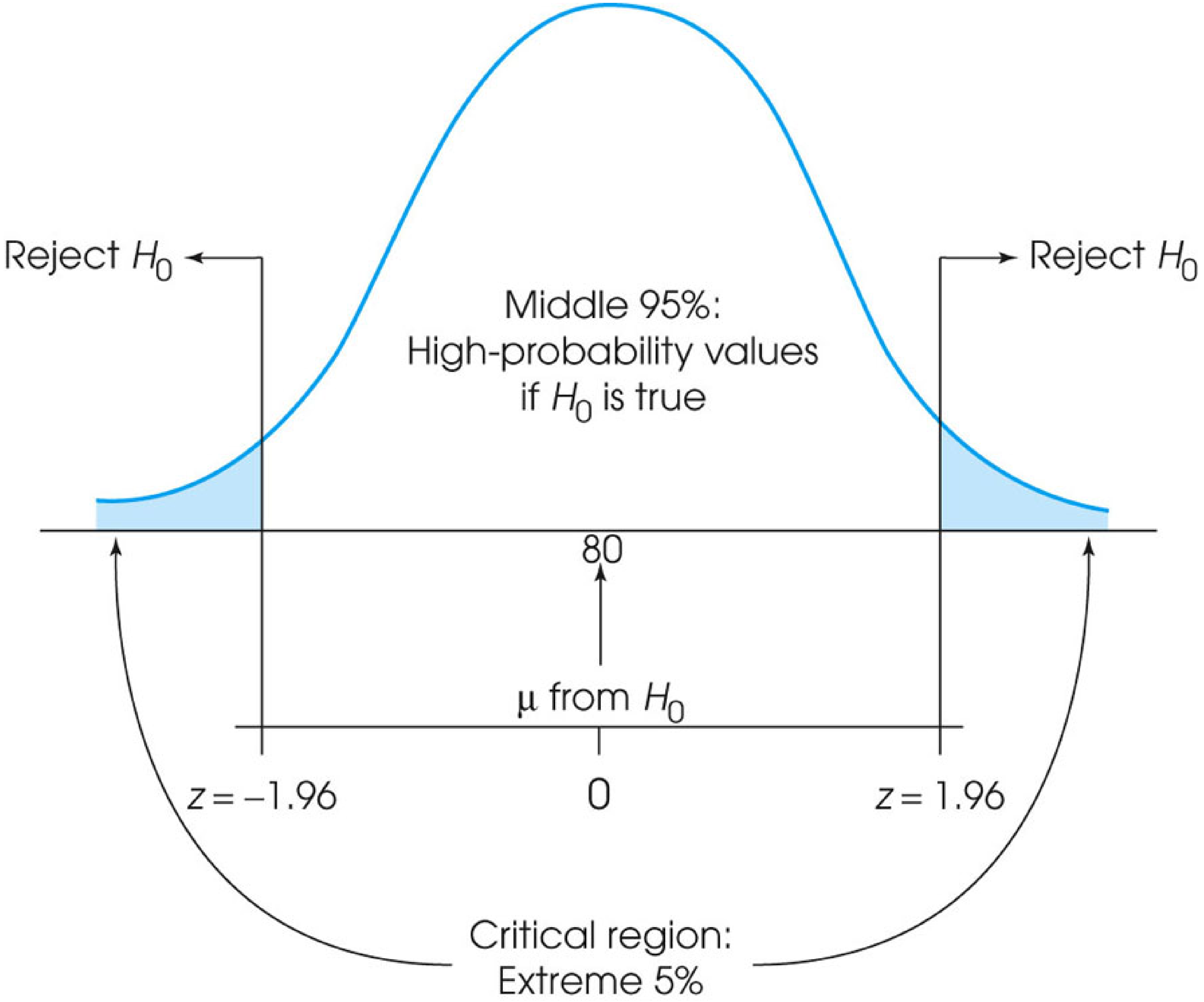
State the hypotheses and select an α level. The **null hypothesis**, H**0**, always states that the treatment has no effect (no change, no difference). According to the null hypothesis, the population mean after treatment is the same is it was before treatment. The **α level** establishes a criterion, or "cut-off", for making a decision about the null hypothesis. The alpha level also determines the risk of a Type I error.



z
−1.96
0
1.96
−2.58
μ from H_0
2.58
−3.30
3.30
α = .05
α = .01
α = .001

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.

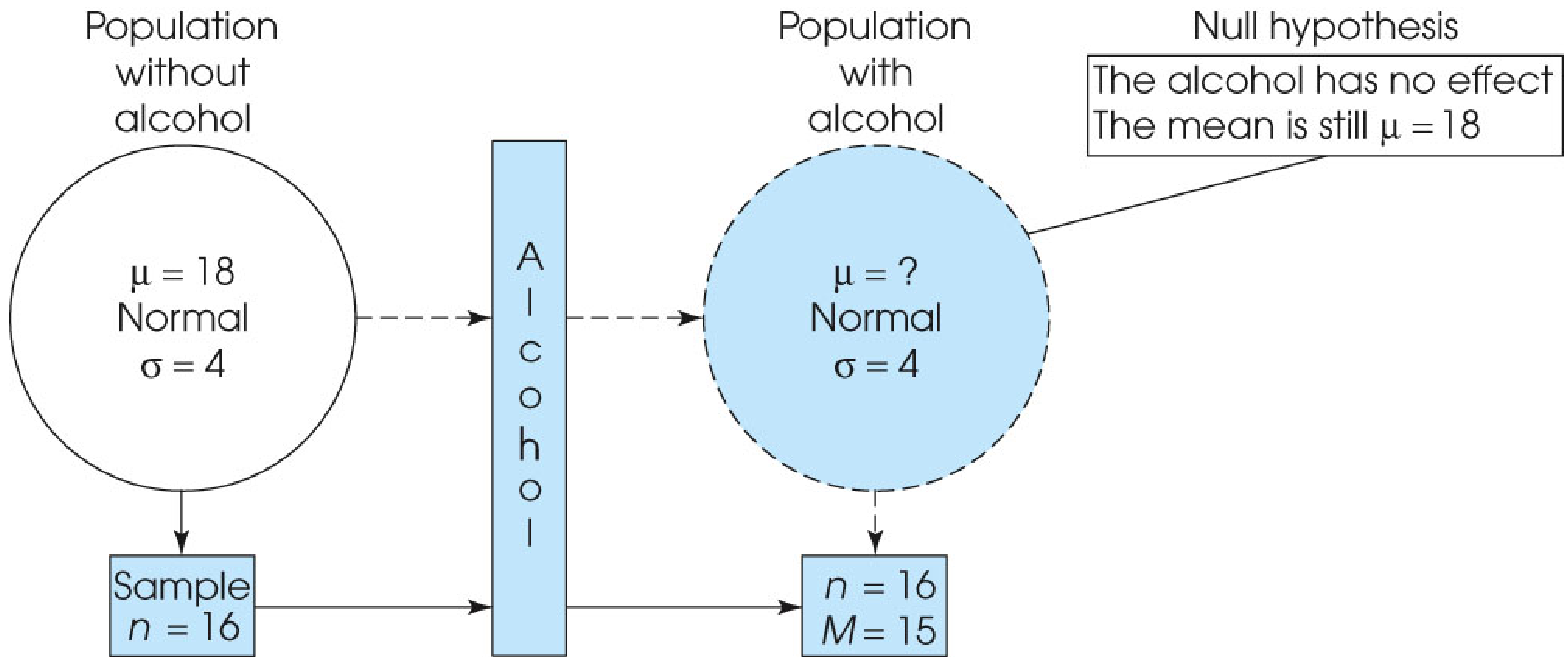

Reject H_0
Reject H_0
Middle 95%:
High-probability values
if H_0 is true
80
μ from H_0
$z = -1.96$
0
$z = 1.96$
Critical region:
Extreme 5%

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
μ = 18
Normal
σ = 4
Sample
n = 16
Alcohol
Population with alcohol
μ = ?
Normal
σ = 4
n = 16
M = 15
Null hypothesis
The alcohol has no effect
The mean is still μ = 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.

4-3 Hypothesis Testing

4-3.2 Testing Statistical Hypotheses

Rejecting the null hypothesis H_0 when it is true is defined as a **type I error.**

Failing to reject the null hypothesis when it is false is defined as a **type II error.**

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

4-3 Hypothesis Testing

4-3.2 Testing Statistical Hypotheses

Table 4-1 Decisions in Hypothesis Testing

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

Sometimes the type I error probability is called the **significance level**, or the **α-error**, or the **size** of the test.

4-3 Hypothesis Testing

4-3.2 Testing Statistical Hypotheses

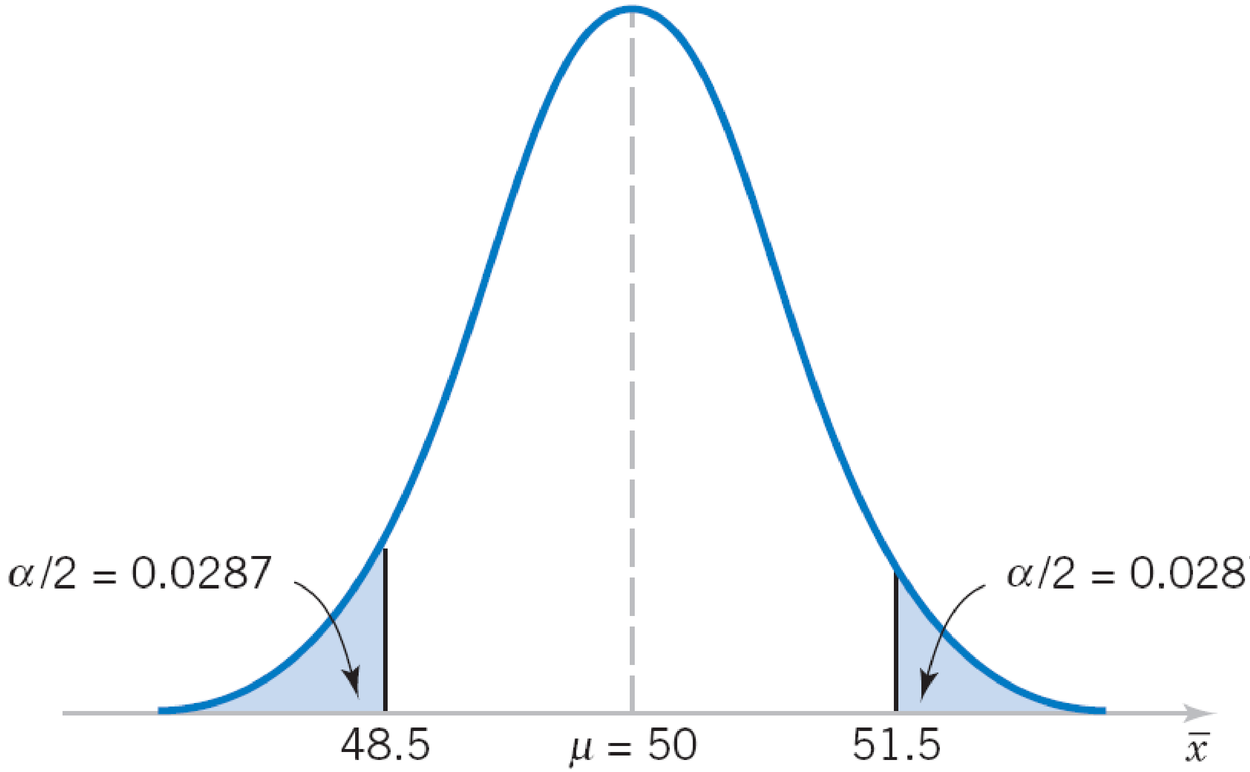


Figure 4-4 The critical region for H_0: $\mu = 50$ versus H_1: $\mu \neq 50$ and $n = 10$.

4-3 Hypothesis Testing

4-3.2 Testing Statistical Hypotheses

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

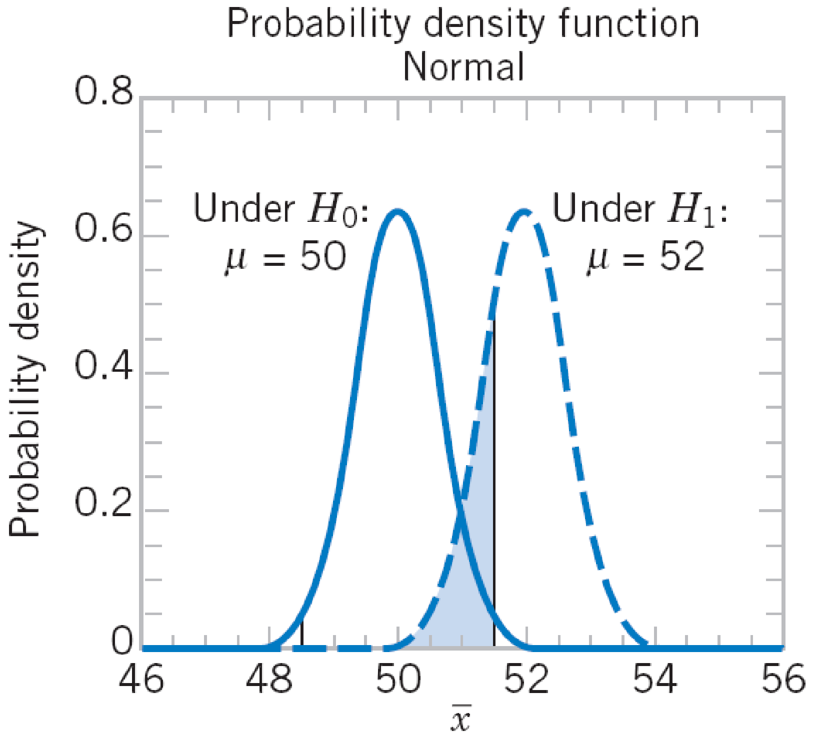


Figure 4-5 The probability of type II error when $\mu = 52$ and $n = 10$.

4-3 Hypothesis Testing

4-3.2 Testing Statistical Hypotheses

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

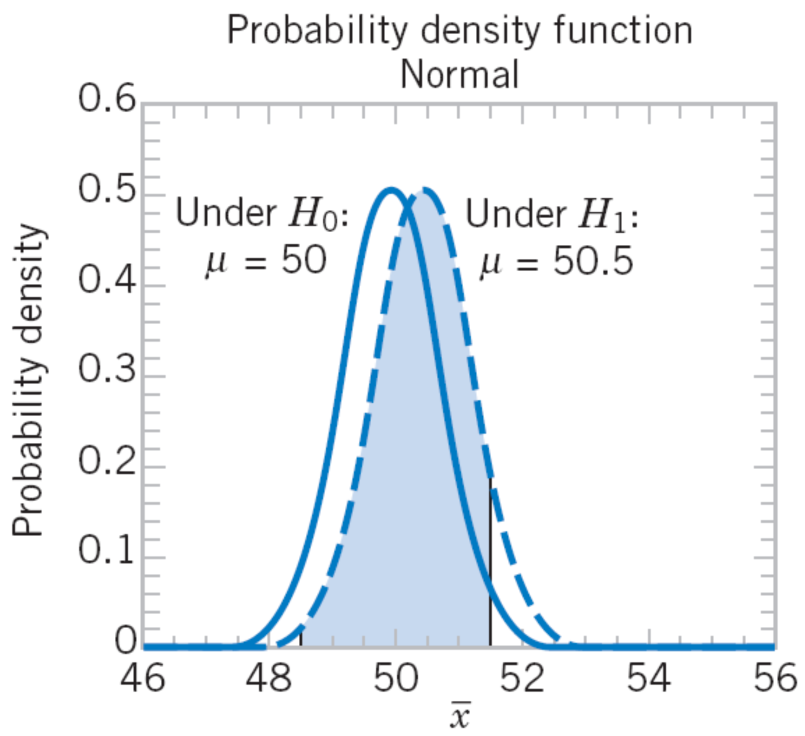


Figure 4-6 The probability of type II error when $\mu = 50.5$ and $n = 10$.

4-3 Hypothesis Testing

4-3.2 Testing Statistical Hypotheses

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

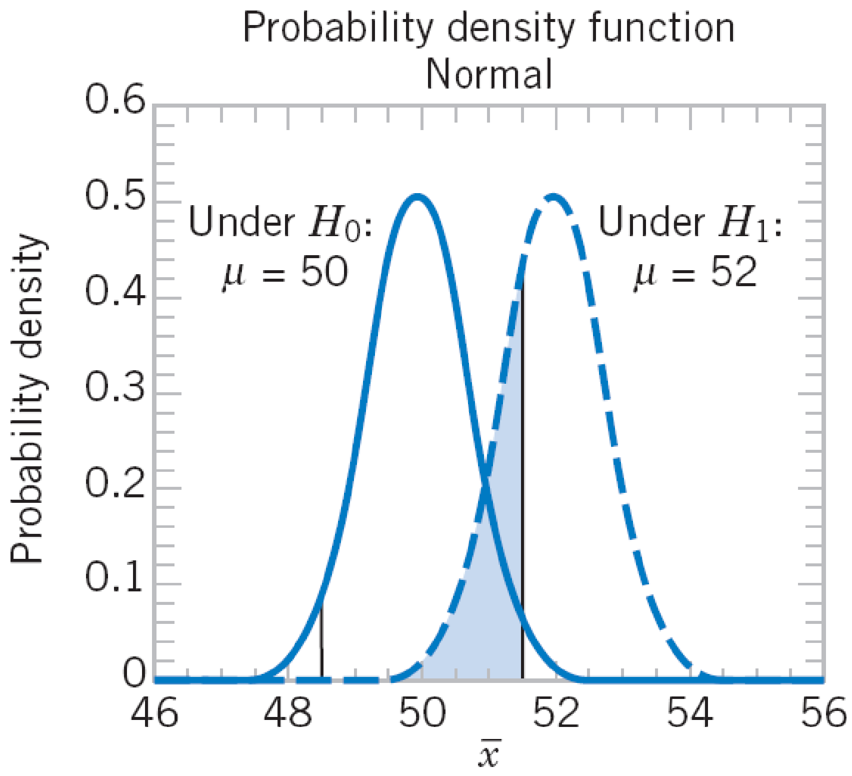


Figure 4-7 The probability of type II error when $\mu = 52$ and $n = 16$.

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:

$$H0: \mu \leq 80 \quad \text{(there is no increase)}$$

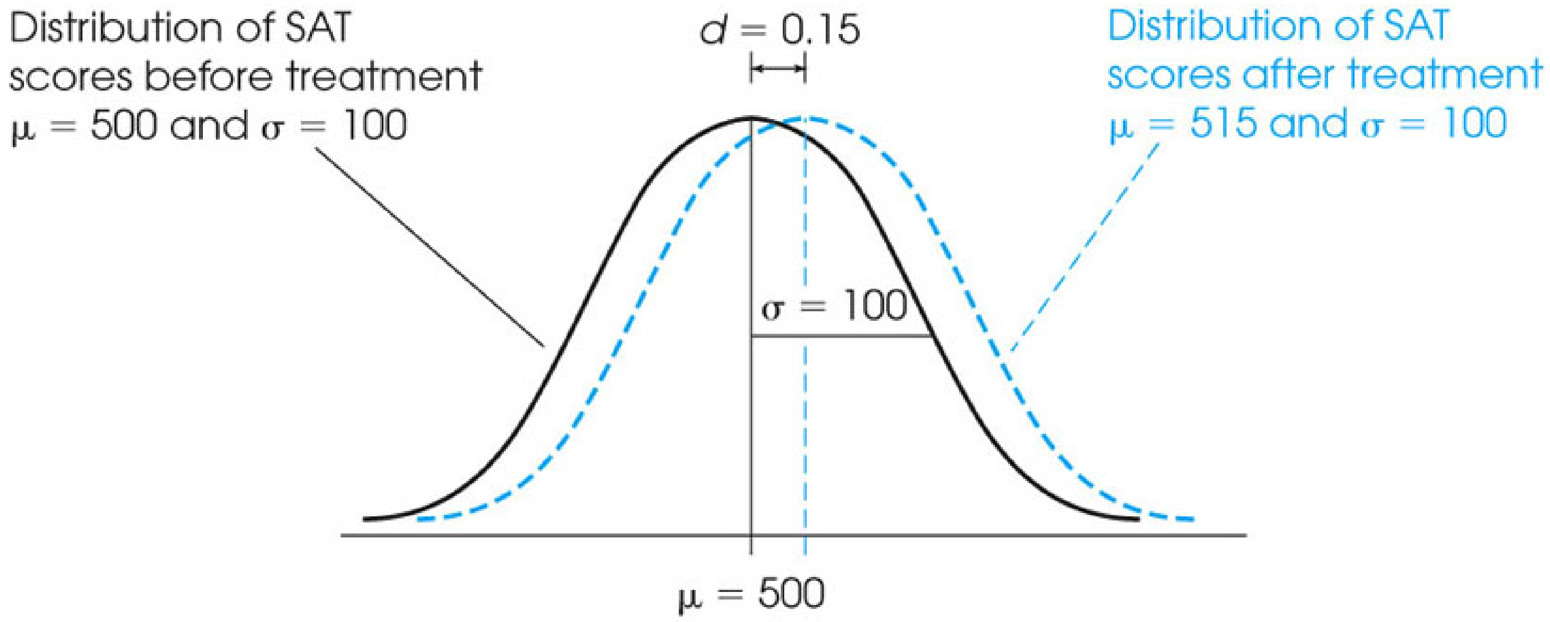
- In this case, the entire critical region would be located in the right-hand tail of the distribution because large values for M would demonstrate that there is an increase and would tend to reject the null hypothesis.

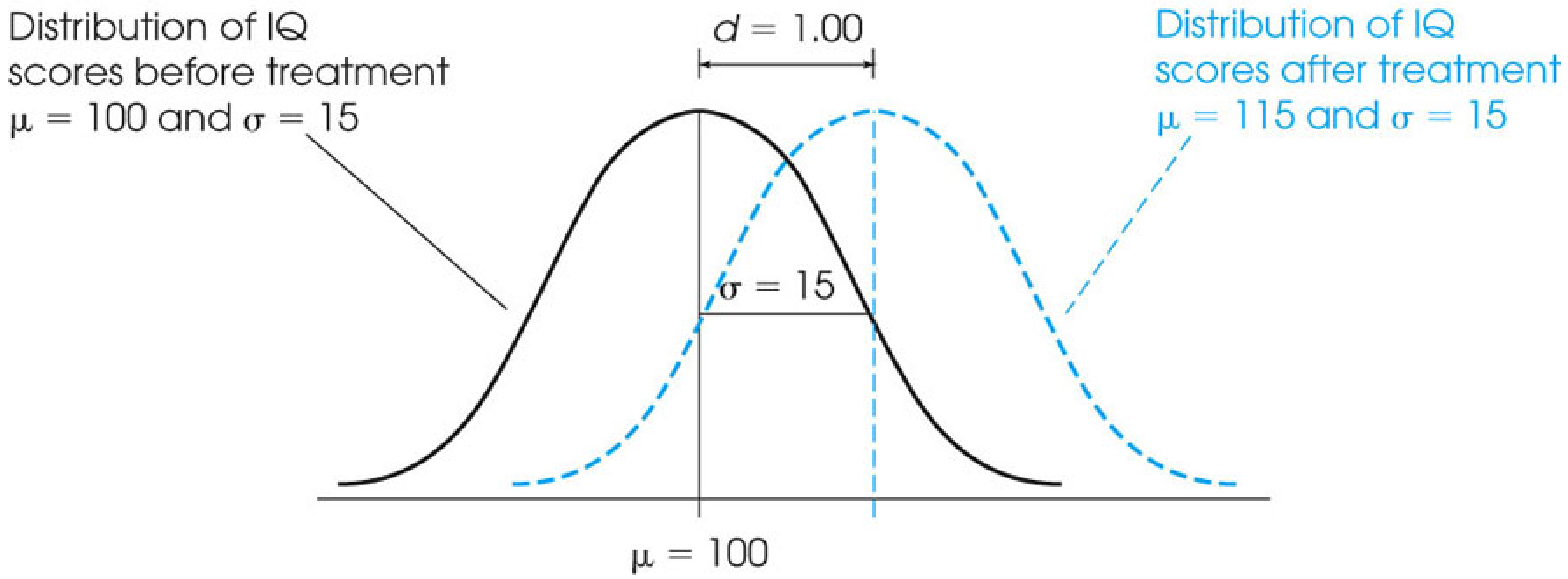
Measuring Effect Size

- 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

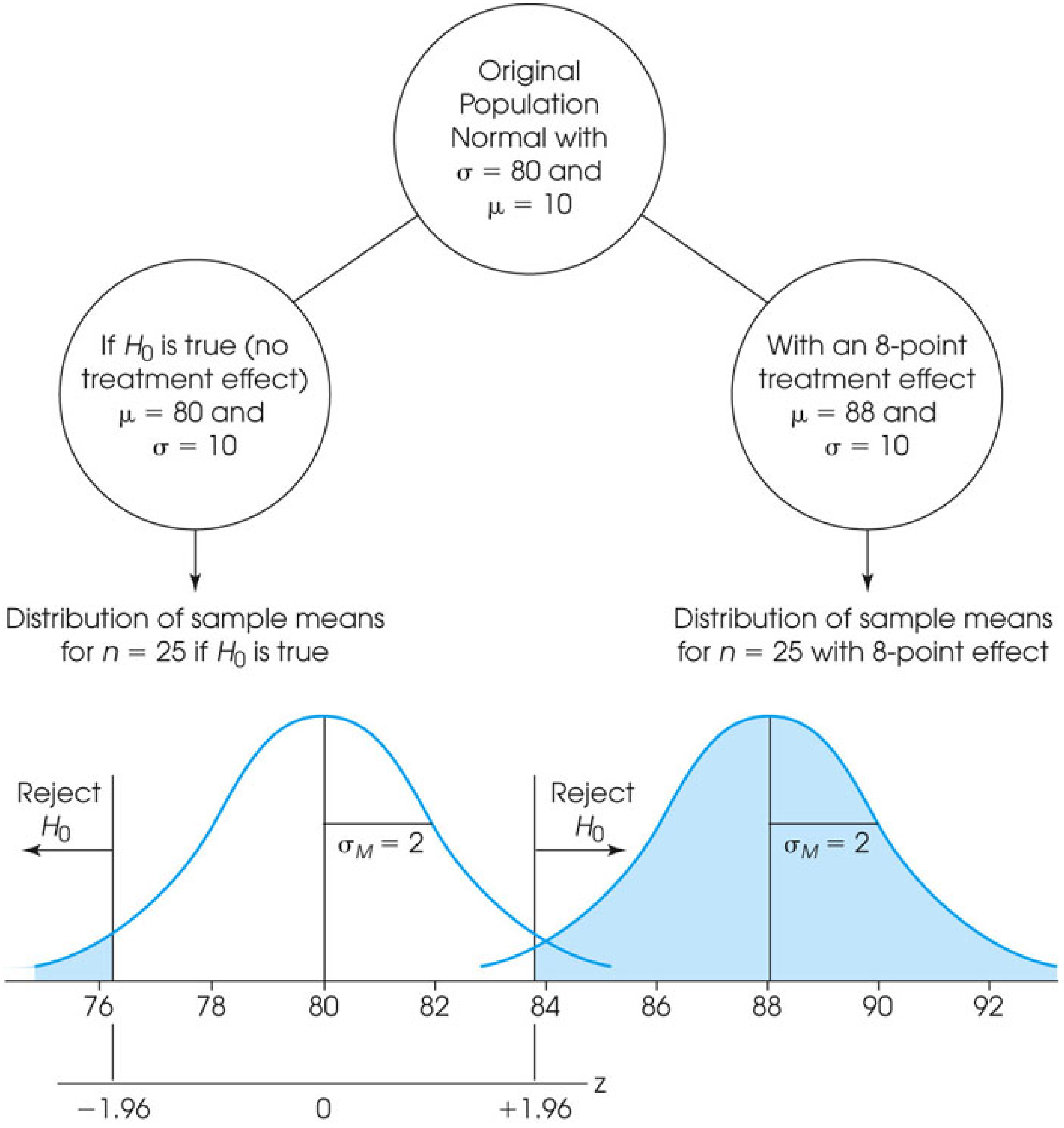
- Because a significant effect does not necessarily mean a large effect, it is recommended that the hypothesis test be accompanied by a measure of the **effect size**.
- We use Cohen=s d as a standardized measure of effect size.
- Much like a z-score, **Cohen=s d** measures the size of the mean difference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


Distribution of SAT
scores before treatment
μ = 500 and σ = 100
d = 0.15
Distribution of SAT
scores after treatment
μ = 515 and σ = 100
σ = 100
μ = 500


Distribution of IQ
scores before treatment
μ = 100 and σ = 15
d = 1.00
Distribution of IQ
scores after treatment
μ = 115 and σ = 15
σ = 15
μ = 100

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.



Original Population Normal with σ = 80 and μ = 10
If H_0 is true (no treatment effect) μ = 80 and σ = 10
With an 8-point treatment effect μ = 88 and σ = 10
Distribution of sample means for n = 25 if H_0 is true
Distribution of sample means for n = 25 with 8-point effect
Reject H_0
Reject H_0
$\sigma_M = 2$
$\sigma_M = 2$
76
78
80
82
84
86
88
90
92
z
−1.96
0
+1.96

Illustrative Example: "Body Weight"

- **The problem:** In the 1970s, 20–29 year old men in the U.S. had a mean μ body weight of 170 pounds. Standard deviation σ was 40 pounds. We test whether mean body weight in the population now differs.
- **Null hypothesis** $H_{0:}$ μ = 170 ("no difference")
- The **alternative hypothesis** can be either
  $H_{a:}$ μ > 170 (**one-sided test**) or
  $H_{a:}$ μ ≠ 170 (**two-sided test**)

§9.2 Test Statistic

This is an example of a one-sample test of a mean when σ is known. Use this statistic to test the problem:

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}}$$

$$\text{where } \mu_0 \equiv \text{population mean assuming } H_0 \text{ is true}$$

$$\text{and } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Illustrative Example: *z* statistic

- For the illustrative example, μ_0 = 170
- We know σ = 40
- Take an SRS of *n* = 64. Therefore

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$$

- If we found a sample mean of 173, then

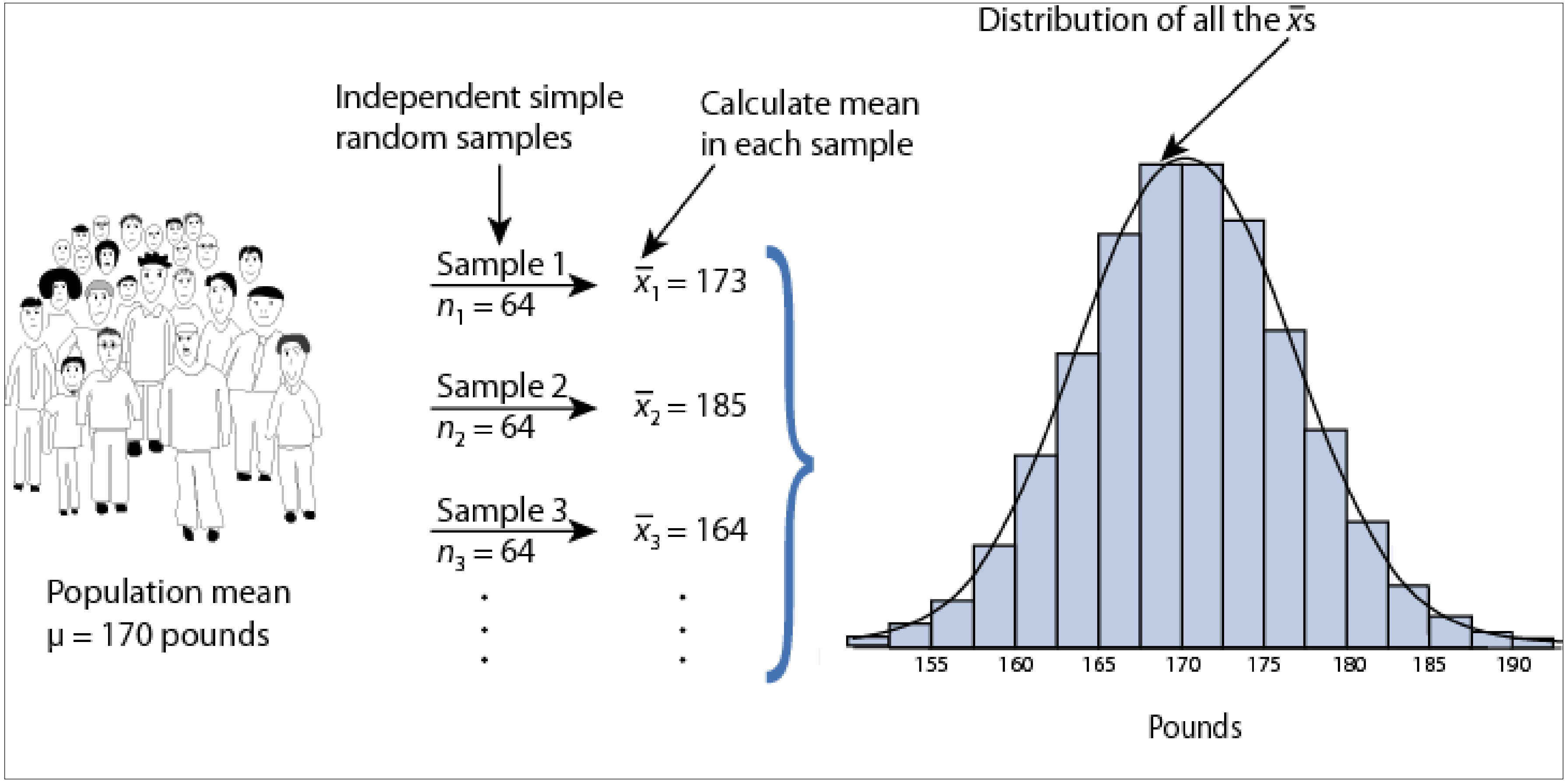
$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{173 - 170}{5} = 0.60$$

Illustrative Example: *z* statistic

If we found a sample mean of 185, then

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{185 - 170}{5} = 3.00$$

Reasoning Behinµz_{stat}



Sampling distribution of xbar under H_0: µ = 170 for n = 64 ⇒ $\bar{x} \sim N(170,5)$

§9.3 *P*-value

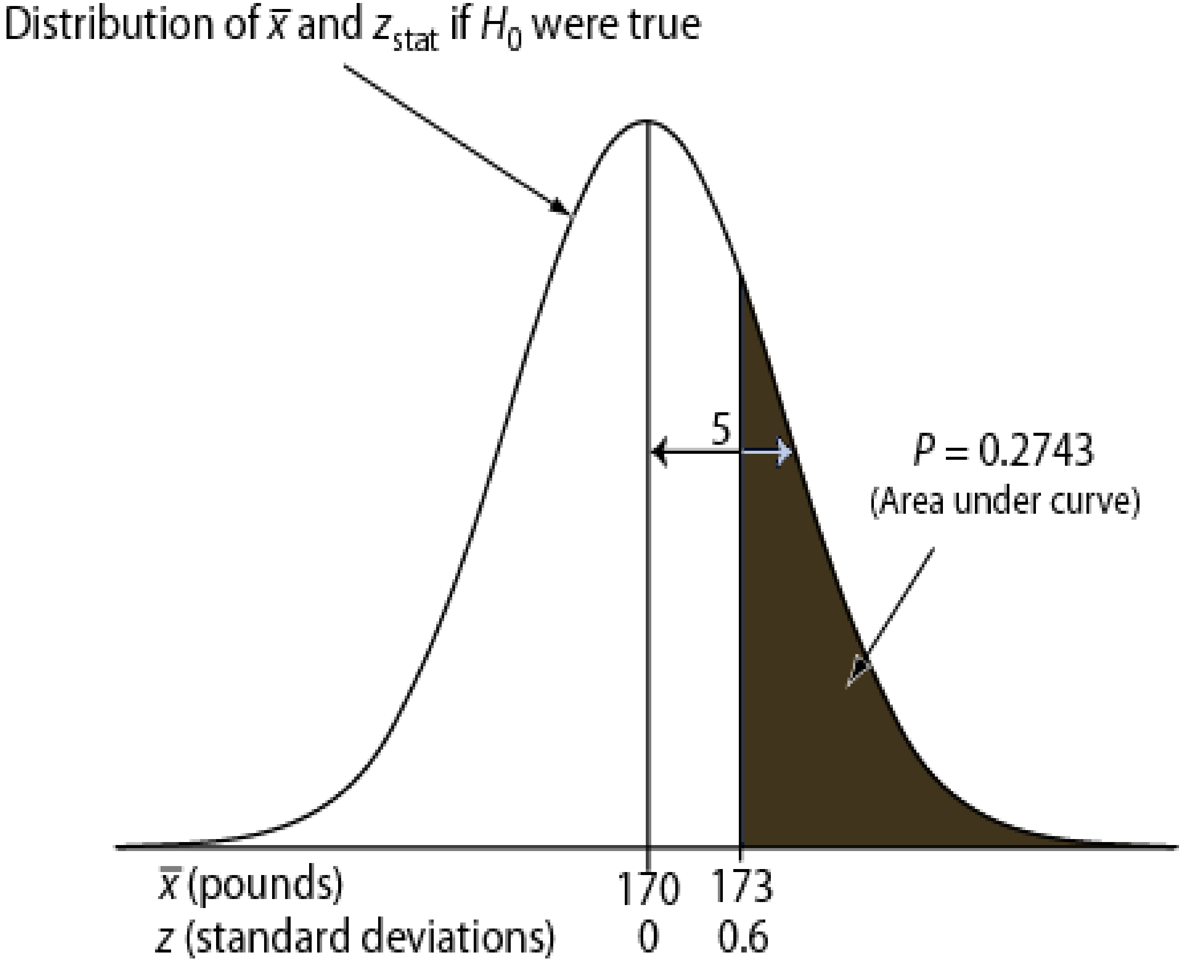
- The *P*-value answer the question: What is the probability of the observed test statistic or one more extreme **when H_0 *is true*?**
- This corresponds to the AUC in the tail of the Standard Normal distribution beyond the $z_{stat.}$
- Convert *z* statistics to *P*-value :

  For H_a: $\mu > \mu_0 \Rightarrow P = \Pr(Z > z_{stat})$ = right-tail beyond z_{stat}

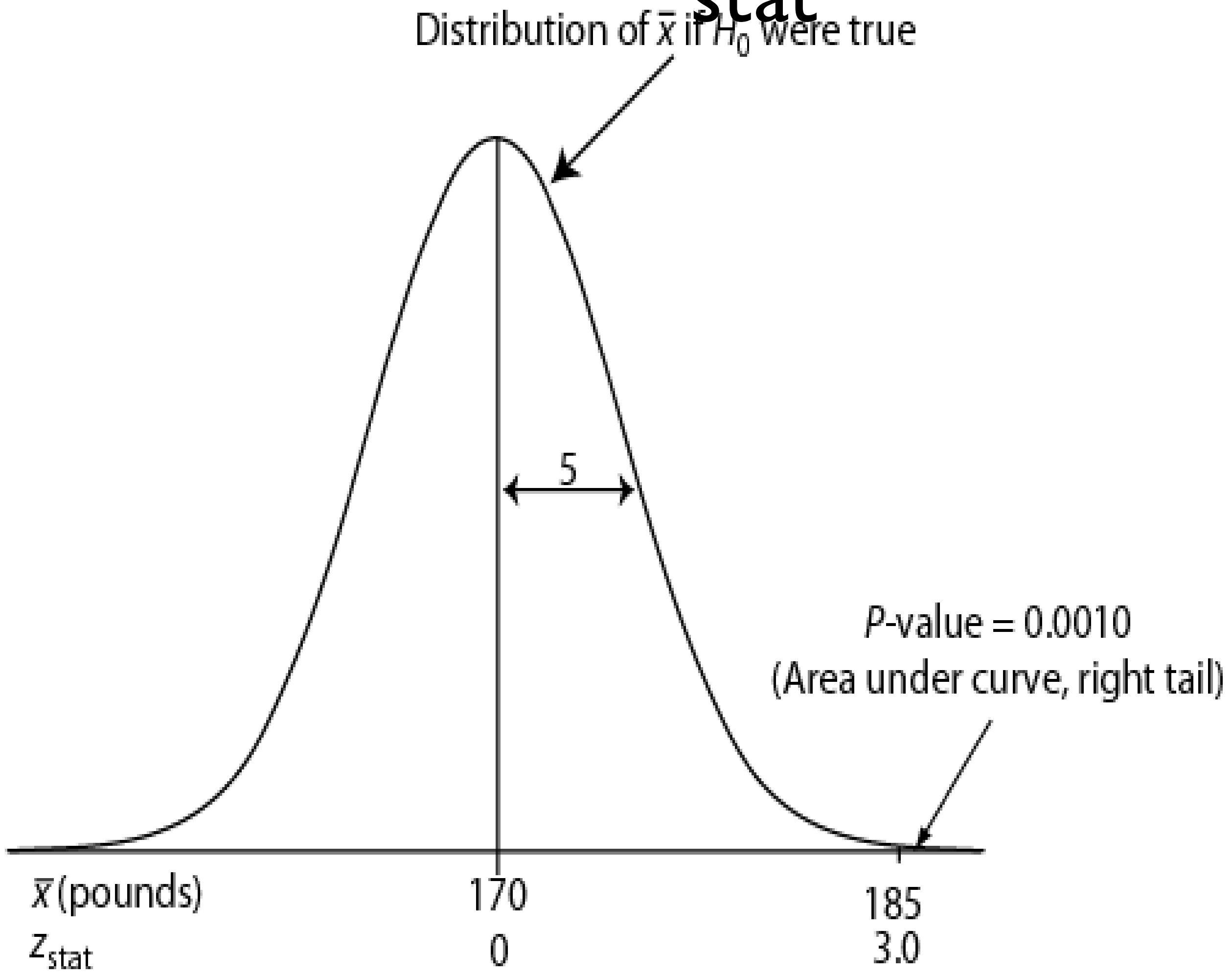
  For H_a: $\mu < \mu_0 \Rightarrow P = \Pr(Z < z_{stat})$ = left tail beyond z_{stat}

  For H_a: $\mu \neq \mu_0 \Rightarrow P = 2 \times$ one-tailed *P*-value
- Use Table B or software to find these probabilities (next two slides).

One-sided *P*-value for z_{stat} of 0.6

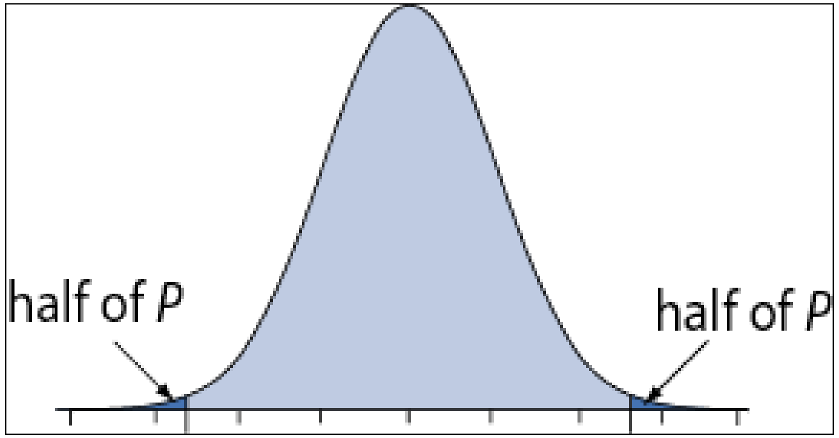


One-sided *P*-value for z_{stat} of 3.0



Two-Sided *P*-Value

- One-sided $H_a \Rightarrow$ AUC in tail beyond z_{stat}
- Two-sided $H_a \Rightarrow$ consider potential deviations in both directions $\Rightarrow$ double the one-sided *P*-value



Examples: If one-sided *P* = 0.0010, then two-sided *P* = 2 × 0.0010 = 0.0020. If one-sided *P* = 0.2743, then two-sided *P* = 2 × 0.2743 = 0.5486.

Interpretation

- P-value answer the question: What is the probability of the observed test statistic … **when H_0 *is true*?**
- Thus, smaller and smaller P-values provide stronger and stronger evidence against H_0
- Small P-value $\Rightarrow$ strong evidence

Interpretation

Conventions*

$P > 0.10 \Rightarrow$ non-significant evidence against H_0

$0.05 < P \leq 0.10 \Rightarrow$ marginally significant evidence

$0.01 < P \leq 0.05 \Rightarrow$ significant evidence against H_0

$P \leq 0.01 \Rightarrow$ highly significant evidence against H_0

Examples

$P = .27 \Rightarrow$ non-significant evidence against H_0

$P = .01 \Rightarrow$ highly significant evidence against H_0

*** It is *unwise* to draw firm borders for "significance"**

α-Level (Used in some situations)

- Let $\alpha \equiv$ probability of erroneously rejecting H_0
- Set α threshold (e.g., let $\alpha = .10, .05,$ *or whatever*)
- Reject H_0 when $P \leq \alpha$
- Retain H_0 when $P > \alpha$
- Example: Set $\alpha = .10$. Find $P = 0.27 \Rightarrow$ retain H_0
- Example: Set $\alpha = .01$. Find $P = .001 \Rightarrow$ reject H_0

(Summary) One-Sample *z* Test

A. Hypothesis statements
H_0: $\mu = \mu_0$ vs.
H_a: $\mu \neq \mu_0$ (two-sided) or
H_a: $\mu < \mu_0$ (left-sided) or
H_a: $\mu > \mu_0$ (right-sided)

B. Test statistic

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} \text{ where } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

C. P-value: convert z_{stat} to P value

D. Significance statement (usually not necessary)

Matlab Example

- https://www.mathworks.com/help/stats/hypothesis-tests-1.html
- https://www.mathworks.com/help/stats/selecting-a-sample-size.html