Section 6.2 Hypothesis Testing

GIVEN: an unknown parameter μ, and two mutually exclusive statements H_0 and H_1 about θ.

The Statistician must decide either to accept H_0 or to accept H_1.

This kind of problem is a problem of Hypothesis Testing.

A procedure for making a decision is called a **test procedure** or simply a **test**.

H_0 =Null Hypothesis
H_1 = Alternative Hypothesis.

Example To study effectiveness of a gasoline additive on fuel efficiency, 30 cars are sent on a road trip from Boston to LA. Without the additive, the fuel efficiency average is μ= 25.0 mpg with a standard deviation σ=2.4.

The test cars averaged $\overline{y}$=26.3 mpg with the additive. What should the company conclude?

ANSWER: Consider the hypotheses

$$H_0 : \quad \mu = 25.0 \quad \text{Additive is not effective.}$$

$$H_1 : \quad \mu > 25.0 \quad \text{Additive is effective.}$$

It is reasonable to consider a value y^* to compare with the sample mean $\overline{y}$, so that H_0 is accepted or not depending on whether $\overline{y} < y^*$ or not.

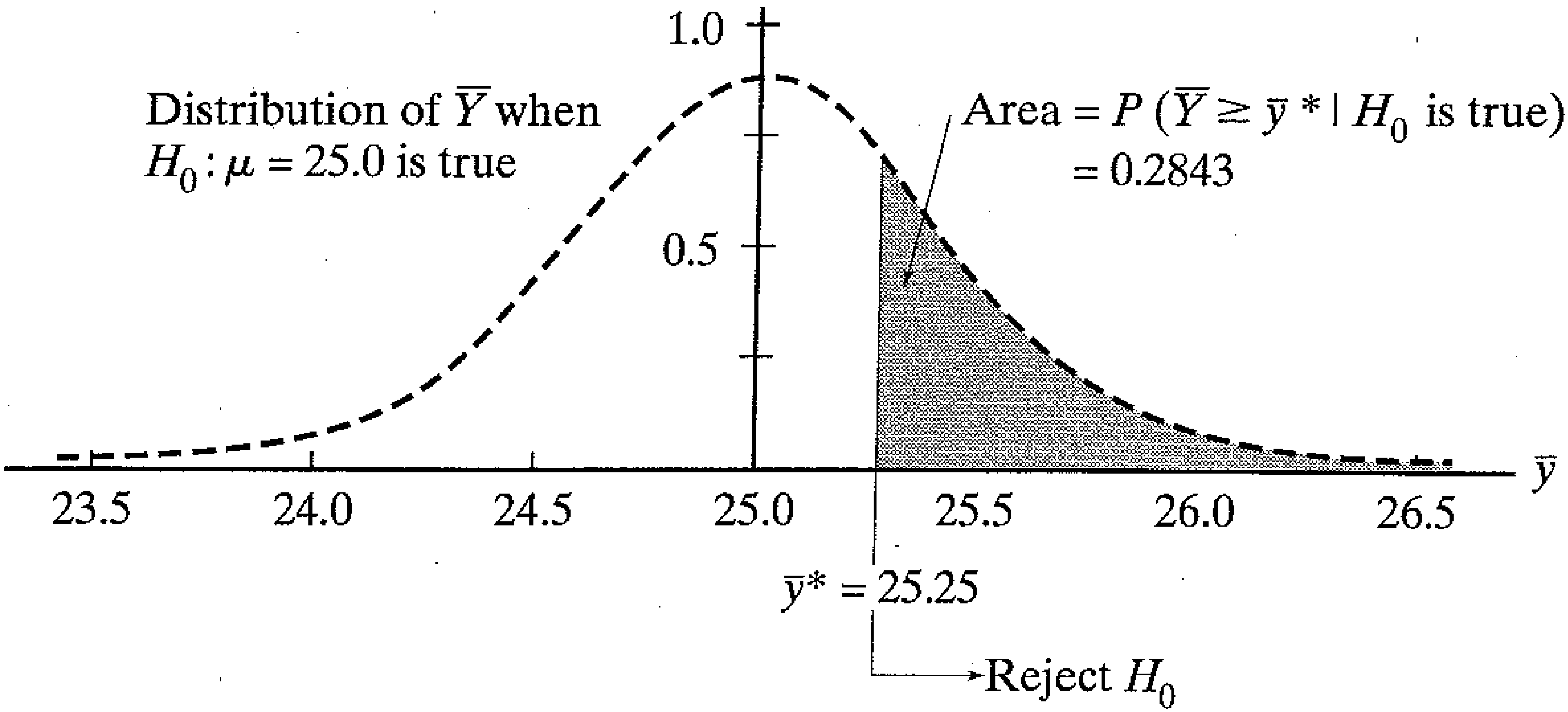
For sake of discussion, suppose $y^* = 25.25$ **is s.t.** H_0 **is rejected if** $\overline{y} > y^*$

Question: $P($ reject $H_0 \mid H_0$ is true $) =?$

We have,

$$
\begin{aligned}
& P(\text{ reject } H_0 \mid H_0 \text{ is true }) \\
& = P(\overline{Y} \geq 25.25 \mid \mu = 25.0) \\
& = P\left(\frac{\overline{Y}-25.0}{2.4/\sqrt{30}} \geq \frac{25.25-25.0}{2.4/\sqrt{30}}\right) \\
& = P(Z \geq 0.57) \\
& = 0.2843
\end{aligned}
$$

FIGURE 6.2.2



1.0
0.5
Distribution of $\overline{Y}$ when
$H_0: \mu = 25.0$ is true
Area = $P(\overline{Y} \geq \bar{y}^* \mid H_0$ is true)
= 0.2843
23.5
24.0
24.5
25.0
25.5
26.0
26.5
$\bar{y}$
$\bar{y}^* = 25.25$
Reject H_0

Let us make y^* larger,say $Y^* = 26.25$

Question: $P($ reject $H_0 \mid H_0$ is true $) =?$

We have,

$$P(\text{ reject } H_0 \mid H_0 \text{ is true })$$

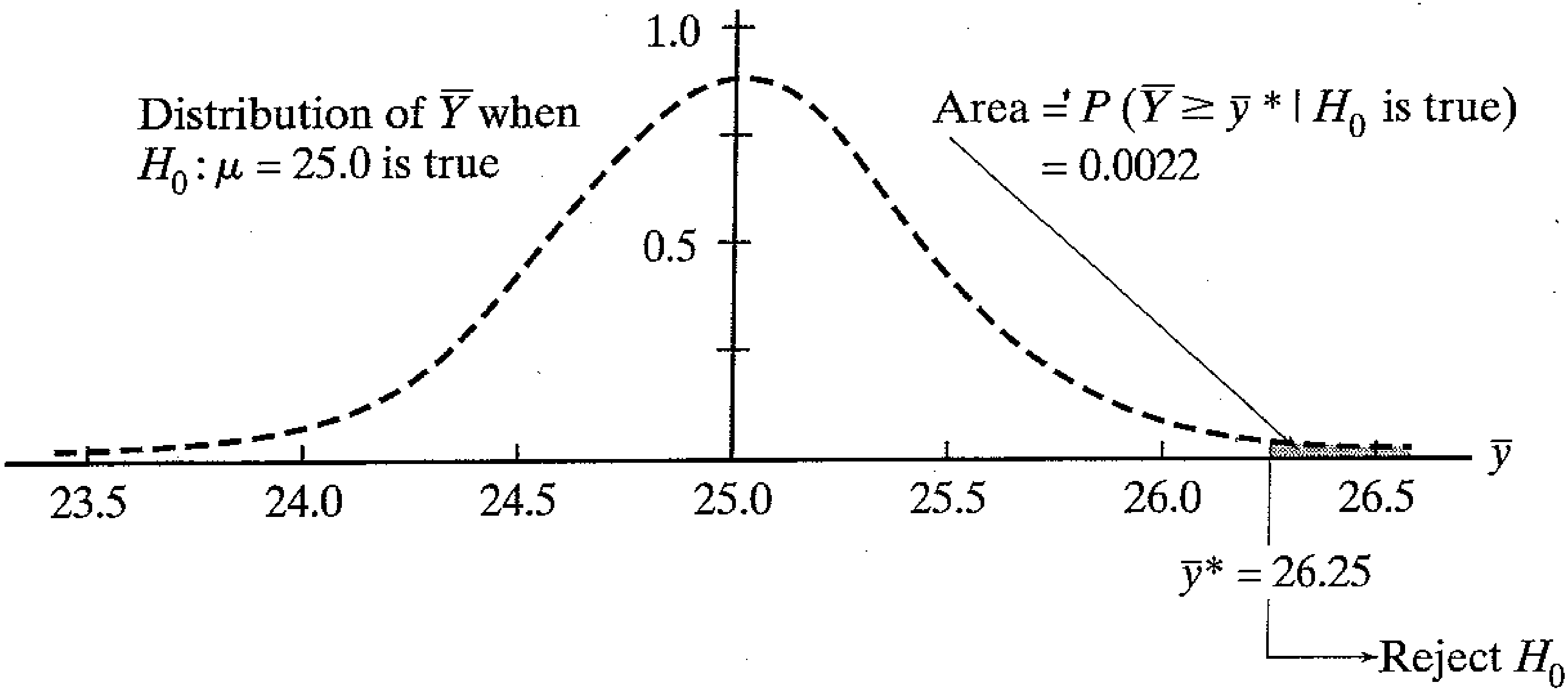
$$= P(\overline{Y} \geq 26.25 \mid \mu = 25.0)$$

$$= P\left(\frac{\overline{Y}-25.0}{2.4/\sqrt{30}} \geq \frac{26.25-25.0}{2.4/\sqrt{30}}\right)$$

$$= P(Z \geq 2.85)$$

$$= 0.0022$$

FIGURE 6.2.3



1.0
0.5
Distribution of $\overline{Y}$ when
$H_0: \mu = 25.0$ is true
Area $\doteq P(\overline{Y} \geq \overline{y}^* \mid H_0$ is true)
$= 0.0022$
23.5
24.0
24.5
25.0
25.5
26.0
26.5
$\overline{y}$
$\overline{y}^* = 26.25$
Reject H_0

WHAT TO USE FOR y^*?

In practice, people often use

$$P(\text{ reject } H_0 \mid H_0 \text{ is true }) = 0.05$$

In our case, we may write

$$P(\overline{Y} \geq y^* \mid H_0 \text{ is true }) = 0.05$$

$$\Longrightarrow P\left(\frac{\overline{Y} - 25.0}{2.4/\sqrt{30}} \geq \frac{y^* - 25.0}{2.4/\sqrt{30}}\right) = 0.05$$

$$\Longrightarrow P\left(Z \geq \frac{y^* - 25.0}{2.4/\sqrt{30}}\right) = 0.05$$

From the Std. Normal table: $P(Z \geq 1.64) = 0.05$. Then,

$$\frac{y^* - 25.0}{2.4/\sqrt{30}} = 0.05 \Longrightarrow y^* = 25.718$$

Simulation p. 369 , TABLE 6.2.1

Some Definitions

The random variable

$$\frac{\overline{Y} - 25.0}{2.4/\sqrt{30}}$$

has a standard normal distribution.

The observed z-value is what you get when a particular $\overline{y}$ is substituted for $\overline{Y}$:

$$\frac{\overline{y} - 25.0}{2.4/\sqrt{30}} = \text{ observed z-value}$$

A Test Statistic is any function of the observed data that dictates whether H_0 is accepted or rejected.

The Critical Region is the set of values for the test statistic that result in H_0 being rejected.

The Critical Value is a number that separates the rejection region from the acceptance region.

Example In our gas mileage example, both

$$\overline{Y} \quad \text{and} \quad \frac{\overline{Y} - 25.0}{2.4/\sqrt{30}}$$

are test statistics, with corresponding critical regions (respectively)

$$C = \{\overline{y} : \overline{y} \geq 25.718\}$$

and

$$C = \{z : z \geq 1.64\}$$

and critical values 25.718 and 1.64.

Definition 6.2.2 The Level of Significance is the probability that the test statistic lies in the critical region when H_0 is true.

In previous slide we used 0.05 as level of significance.

One Sided vs. Two Sided Alternatives

In our fuel efficiency example, we had a one sided alternative, specifically, one sided to the right ($H_1 : \mu > \mu_0$).

In some situations the alternative hypothesis could be taken as one-sided to the left ($H_1 : \mu < \mu_0$) or as two sided ($H_1 : \mu \neq \mu_0$).

Note that, in two sided alternative hypothesis, the level of significance α must be split into two parts corresponding to each one of the two pieces of the critical region.

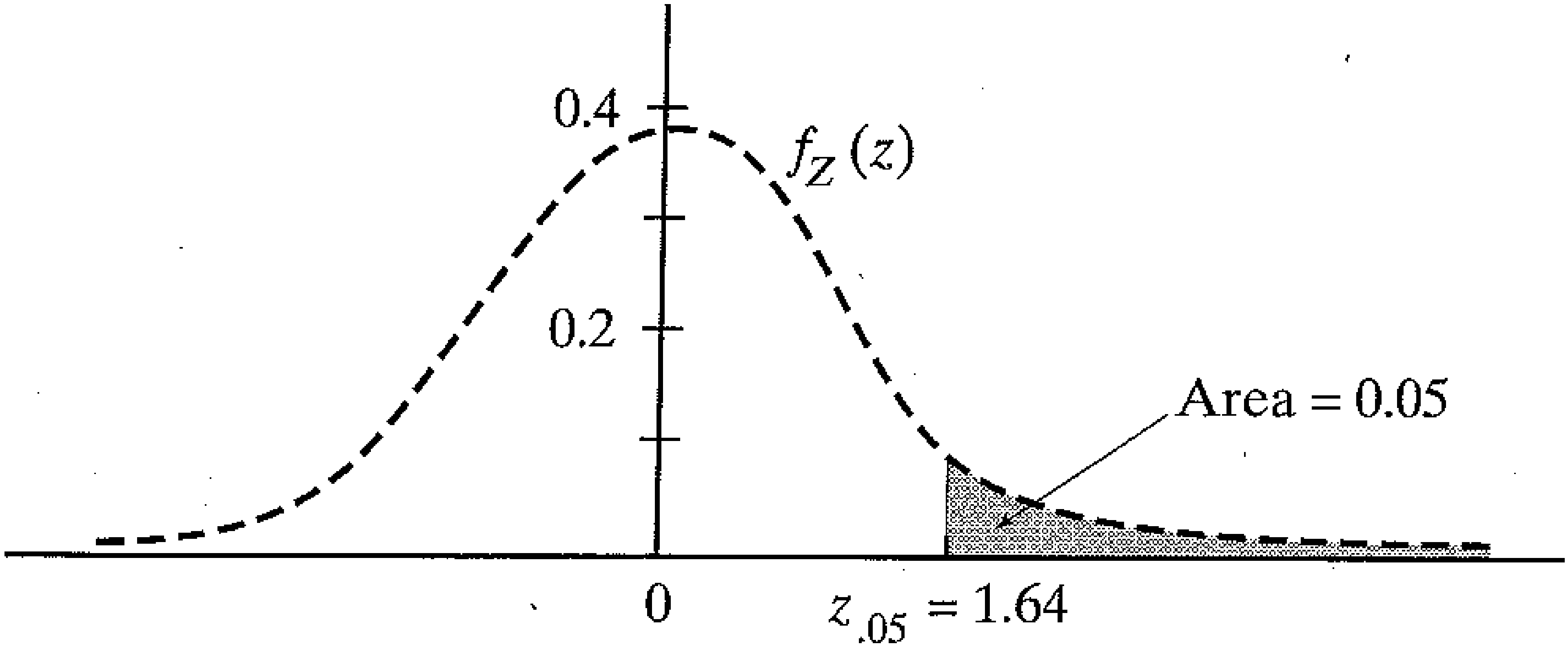
In our fuel example, if we had used a two sided H_1, then each half of the critical region has 0.025 associated probability, with $P(Z \leq -1.96) = 0.025$. This leads to $H_0 : \mu = \mu_0$ to be rejected if the observed z satisfies $z \geq 1.96$ or $z \leq 1.96$.

Testing μ_0 with known σ: Theorem 6.2.1

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n taken from a normal distribution where σ is known, and let $z = \frac{\overline{Y} - \mu_0}{\sigma/\sqrt{n}}$

Test	Signif. level	Action
$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$	α	Reject H_0 if $z \geq z_\alpha$
$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$	α	Reject H_0 if $z \leq z_\alpha$
$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$	α	Reject H_0 if $z \geq z_{\alpha/2}$ or $z \leq z_{\alpha/2}$

FIGURE 6.2.4



Example 6.2.1 Bayview HS has a new Algebra curriculum. In the past, Bayview students would be considered "typical", earning SAT scores consistent with past and current national averages (national averages are mean = 494 and standard deviation 124).

Two years ago a cohort of 86 student were assigned to classes with the new curriculum. Those students averaged 502 points on the SAT. Can it be claimed that at the $\alpha = 0.05$ level of significance that the new curriculum had an effect?

ANSWER: we have the hypotheses

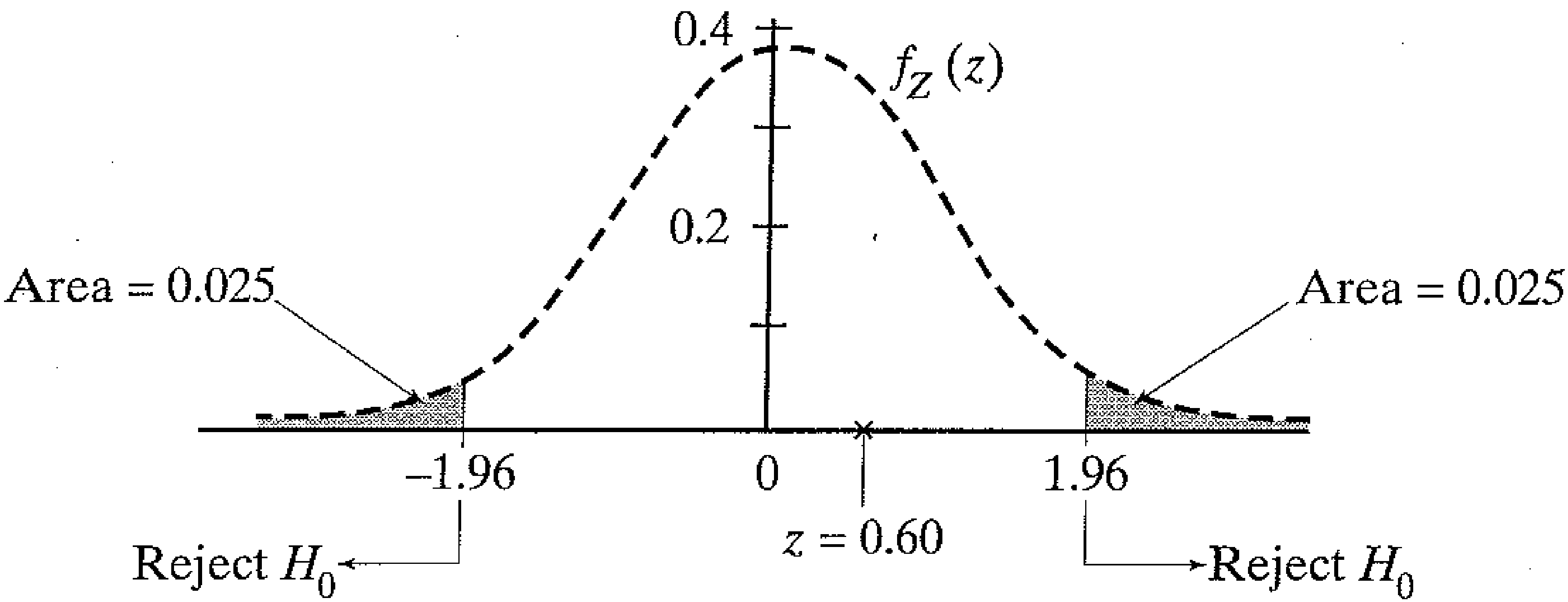
$$\begin{cases} H_0 & : \ \mu = 494 \\ H_1 & : \ \mu \neq 494 \end{cases}$$

Since $z_{\alpha/2} = z_{0.025} = 1.96$, and

$$z = \frac{502 - 494}{124/\sqrt{86}} = 0.60,$$

the conclusion is "FAIL TO REJECT H_0".

FIGURE 6.2.5


0.4
0.2
$f_Z(z)$
Area = 0.025
Area = 0.025
–1.96
0
1.96
$z = 0.60$
Reject H_0
Reject H_0

The P-Value, Definition 6.2.3.

Two methods to quantify evidence against H_0:

(a) The statistician selects a value for α before any data is collected, and a critical region is identified. If the test statistic falls in the critical region, H_0 is rejected at the α level of significance.

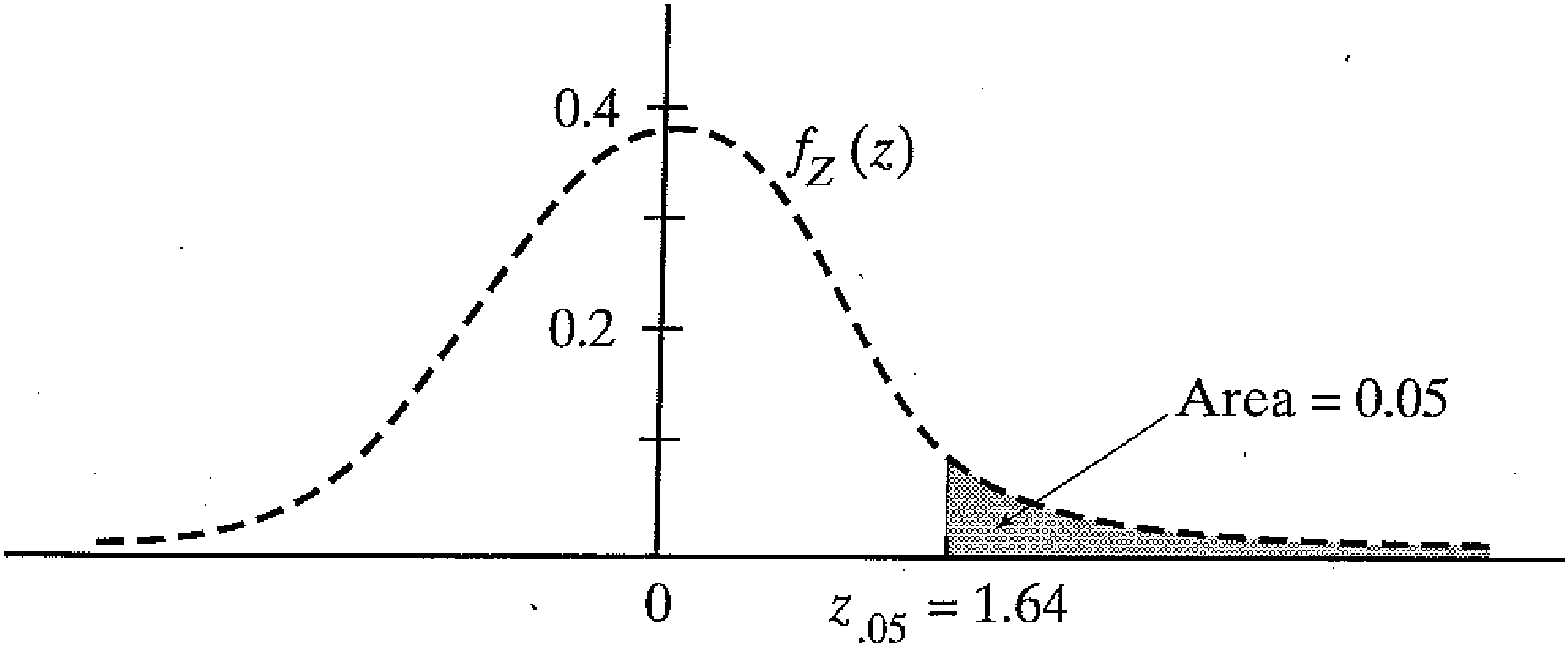
(b) The statistician reports a P-value, which is the probability of getting a value of that test statistic as extreme or more extreme than what was actually observed (relative to H_1), given that H_0 is true.

Example 6.2.2 Recall Example 6.2.1. Given that $H_0 : \mu = 494$ is being tested against $H_1 : \mu \neq 494$, what P-value is associated with the calculated test statistic, $z = 0.60$, and how should it be interpreted?

ANSWER: If $H_0 : \mu = 494$ is true, then $Z =$ has a standard normal pdf. Relative to the two sided H_1, any value of $Z \geq$ 0.60 or $\leq$ -0.60 is as extreme or more extreme than the observed z, Then,

$$
\begin{aligned}
& P - \text{value} \\
& \quad = P(Z \geq 0.60) + P(Z \leq 0.60) \\
& \quad = 0.2743 + 0.2743 \\
& \quad = 0.5486
\end{aligned}
$$

FIGURE 6.2.4



Section 6.3: Testing Binomial Data Suppose $X_1, \ldots X_n$ are outcomes in independent trials, with $P(X_\ell = 1) = p$ and $P(X_\ell = 0) = 1 - p$, with p unknown.

A test with null hypothesis $H_0 : \quad p = p_0$ is called binomial hypothesis test.

We consider two cases: large n and small n.

To decide if n is considered "small" or "large", we use the relation

$$0 < np_0 - 3\sqrt{np_0(1-p_0)} < np_0 + 3\sqrt{np_0(1-p_0)} < n$$

A large sample test for binomial parameter

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of n Bernoulli RVs for which $0 < np_0 - 3\sqrt{np_0(1-p_0)} < np_0 + 3\sqrt{np_0(1-p_0)} < n$. Let $X = X_1 + \cdots + X_n$, and set $z := \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$

Test	Signif. level	Action
$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$	α	Reject H_0 if $z \geq z_\alpha$
$\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}$	α	Reject H_0 if $z \leq -z_\alpha$
$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$	α	Reject H_0 if $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$

Case Study 6.3.1

A point spread is a hypothetical increment added to the score of the weaker of two teams to make them even.

A study examined records of 124 NFL games; it was found that in 67 of them (or 54%) the favored team beat the spread. Is 54% due to chance, or was the spread set incorrectly?

ANSWER: Set

p = P(favored team beats the spread).

We have the hypotheses

$$H_0: \; p = 0.50 \quad \text{versus} \quad H_1: p \neq 0.50$$

We shall use the 0.05 level of significance.

We have

$$n = 124, \quad p_0 = 0.50$$

and

$$X_\ell = 1 \quad \text{if favored team beats spread in } \ell\text{-th game.}$$

Thus the number of times the favored team beats the spread is $X = X_1 + \cdots + X_n$.

We compute z as follows:

$$z := \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{67 - 1240.50}{\sqrt{124 \cdot 0.50 \cdot 0.50}} = 0.90$$

With $\alpha = 0.05$, we have $z_{\alpha/2} = 1.92$. So z does not fall in the critical region.

The null hypothesis is not rejected, that is, 54% is consistent with the statement that the spread was chosen correctly.

Case Study 6.3.2 Do people postpone death until birthday?

Among 747 obituaries in the newspaper, 60 (or 8%) corresponded to people that died in the three months preceding their birthday.

If people die randomly with respect to their b-days, we would expect 25% of them to die in the three months preceding their b-day.

Is the postponement theory valid?

ANSWER: Let $X_\ell = 1$ if ℓ-th person died during 3 months before b-day, and $X_\ell = 0$ if not. Then $X = X_1 + \cdots + X_n = \#$ of people that died during 3 months before b-day. Let $p = P(X = 1)$, $p_0 = 1/4 = 0.25$, and $n = 747$. A one sided test is

$$H_0 : \ p = 0.25 \quad \text{versus} \quad H_1 : p < 0.25$$

We have,

$$z = \frac{x - np_0}{\sqrt{np_0(1 - p_0)}} = \frac{60 - 747(0.25)}{\sqrt{747(0.25)(1 - (0.25))}} = -10.7$$

With $\alpha = 0.05$, H_0 should be rejected if

$$z \leq -z_\alpha = -1.64$$

Since the last inequality holds, we must reject H_0.

The evidence is overwhelming that the reduction from 25% to 8% is due to something other than chance.

What to do for binomial p with small n?

Suppose that for $\ell = 1, \ldots, 19$,

$$X_\ell = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Let $X = X_1 + \cdots + X_n$ with independent X_ℓ's.

Find the Critical Region for the Test

$$H_0 : \ p = 0.85 \quad \text{versus} \quad H_1 : p \neq 0.85$$

with $\alpha \approx 0.10$.

ANSWER: first we must check the inequality

$$0 < np_0 - 3\sqrt{np_0(1-p_0)} < np_0 + 3\sqrt{np_0(1-p_0)} < n$$

With $n = 19$, $p_0 = 0.85$ we get

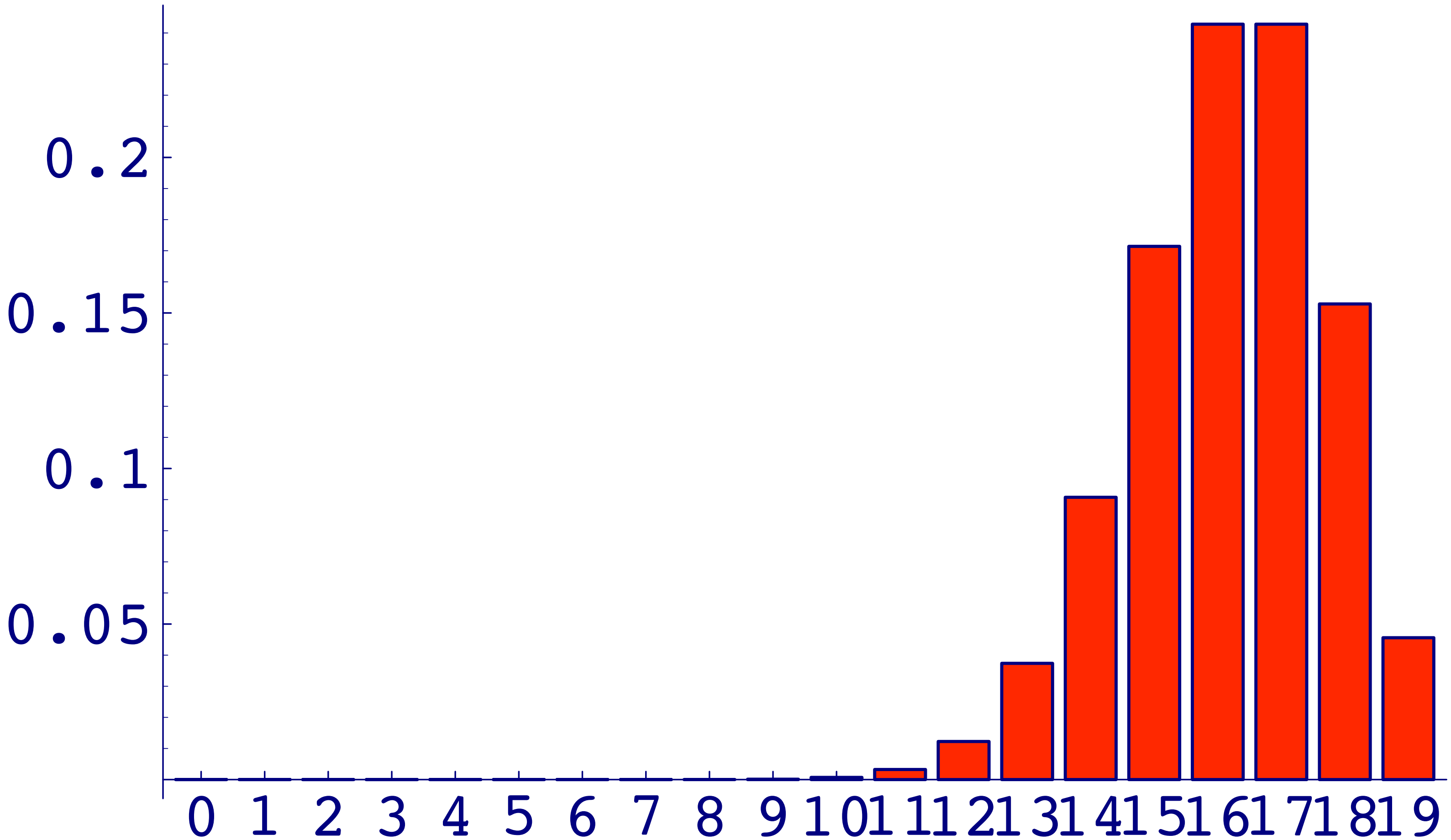
$$19(0.85) + 3\sqrt{19(0.85)(0.15)} = 20.8 \not< 19$$

that is, Theorem 6.3.1 DOES NOT APPLY.

We will use the binomial distribution to define the critical region.

If the null hypothesis is true, the expected value fo X is $19(0.85) = 16.2$. Thus values to the extreme left or right of 16.2 constitute the critical region.

Here is a plot of $p_X(k) = \binom{19}{k}(0.85)^k(0.15)^{19-k}$:



From the table below we get the critical region C:

k	$p_X(k)$	total probability
0	$2.21684\, 10^{-16}$	
1	$2.3868\ \ 10^{-14}$	
2	$1.21727\ \ 10^{-12}$	
3	$3.90878\, 10^{-11}$	
4	$8.85989\ \ 10^{-10}$	
5	$1.50618\, 10^{-8}$	
6	$1.99151\, 10^{-7}$	
7	$2.09582\, 10^{-6}$	$P(X \leq 13) = 0.0536$
8	0.0000178145	
9	0.000123382	
10	0.000699164	
11	0.00324158	
12	0.012246	
13	0.0373659	
14	0.0907457	
15	0.171409	
16	0.242829	
17	0.242829	
18	0.152892	
19	0.0455994	$P(X = 19) = 0.0455994$

$$C = \{x : x \leq 13 \quad \text{or} \quad x = 19\}$$