Dougherty

OXFORD UNIVERSITY PRESS

Introduction to Econometrics, 5th edition

Review: Random Variables, Sampling, Estimation, and Inference

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Assumption: $X \sim N(\mu, \sigma^2)$

Null hypothesis: $H_0 : \mu = \mu_0$

Alternative hypothesis: $H_1 : \mu \neq \mu_0$

This sequence describes the testing of a hypothesis at the 5% and 1% significance levels. It also defines what is meant by a Type I error.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Assumption: $X \sim N(\mu, \sigma^2)$

Null hypothesis: $H_0 : \mu = \mu_0$

Alternative hypothesis: $H_1 : \mu \neq \mu_0$

We will suppose that we have observations on a random variable with a normal distribution with unknown mean μ and that we wish to test the hypothesis that the mean is equal to some specific value μ_0.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Assumption: $X \sim N(\mu, \sigma^2)$

Null hypothesis: $H_0 : \mu = \mu_0$

Alternative hypothesis: $H_1 : \mu \neq \mu_0$

The hypothesis being tested is described as the null hypothesis. We test it against the alternative hypothesis H_1, which is simply that μ is not equal to μ_0.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Assumption: $X \sim N(\mu, \sigma^2)$

Null hypothesis: $H_0 : \mu = \mu_0$

Alternative hypothesis: $H_1 : \mu \neq \mu_0$

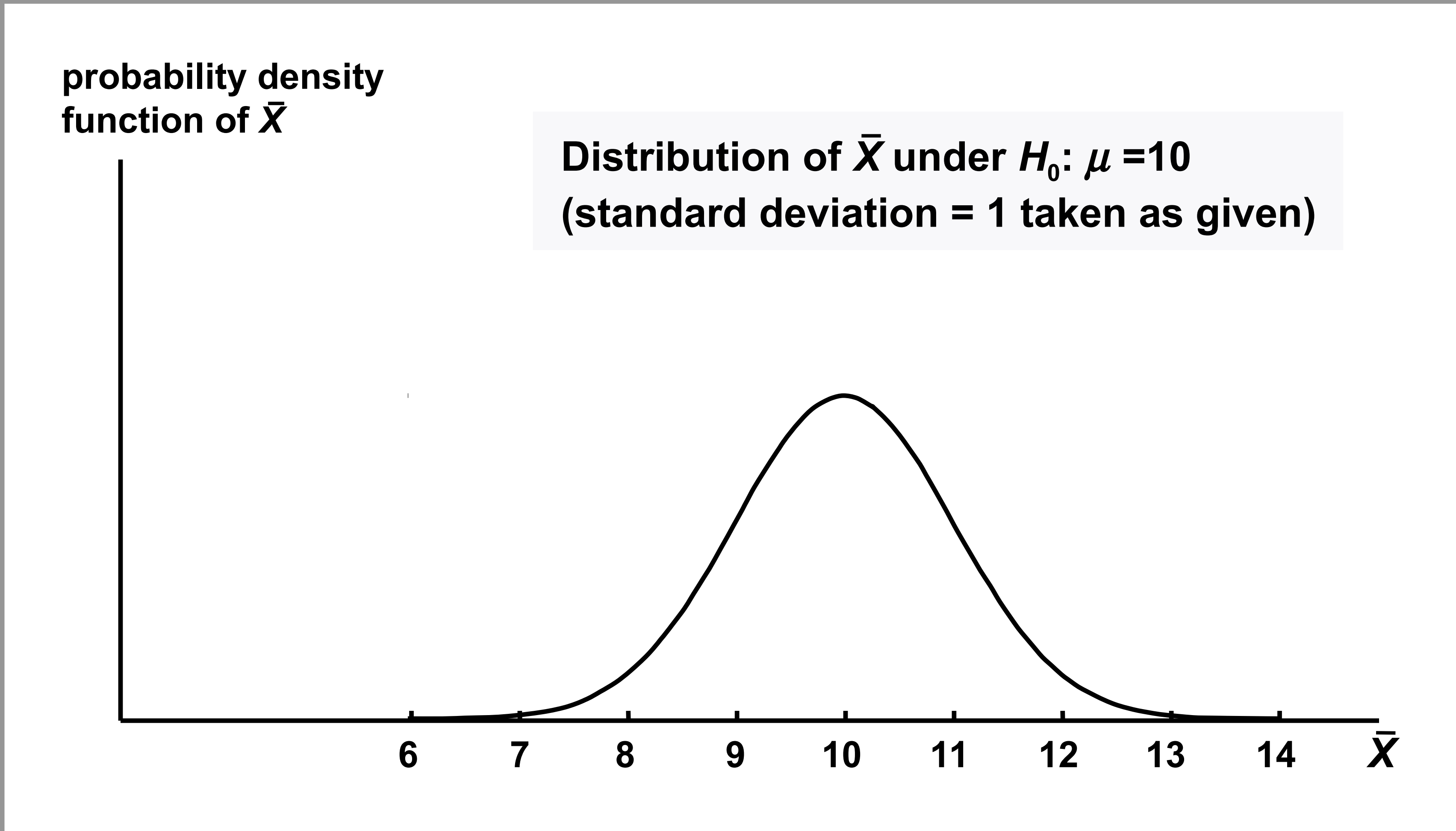
Example

Null hypothesis: $H_0 : \mu = 10$

Alternative hypothesis: $H_1 : \mu \neq 10$

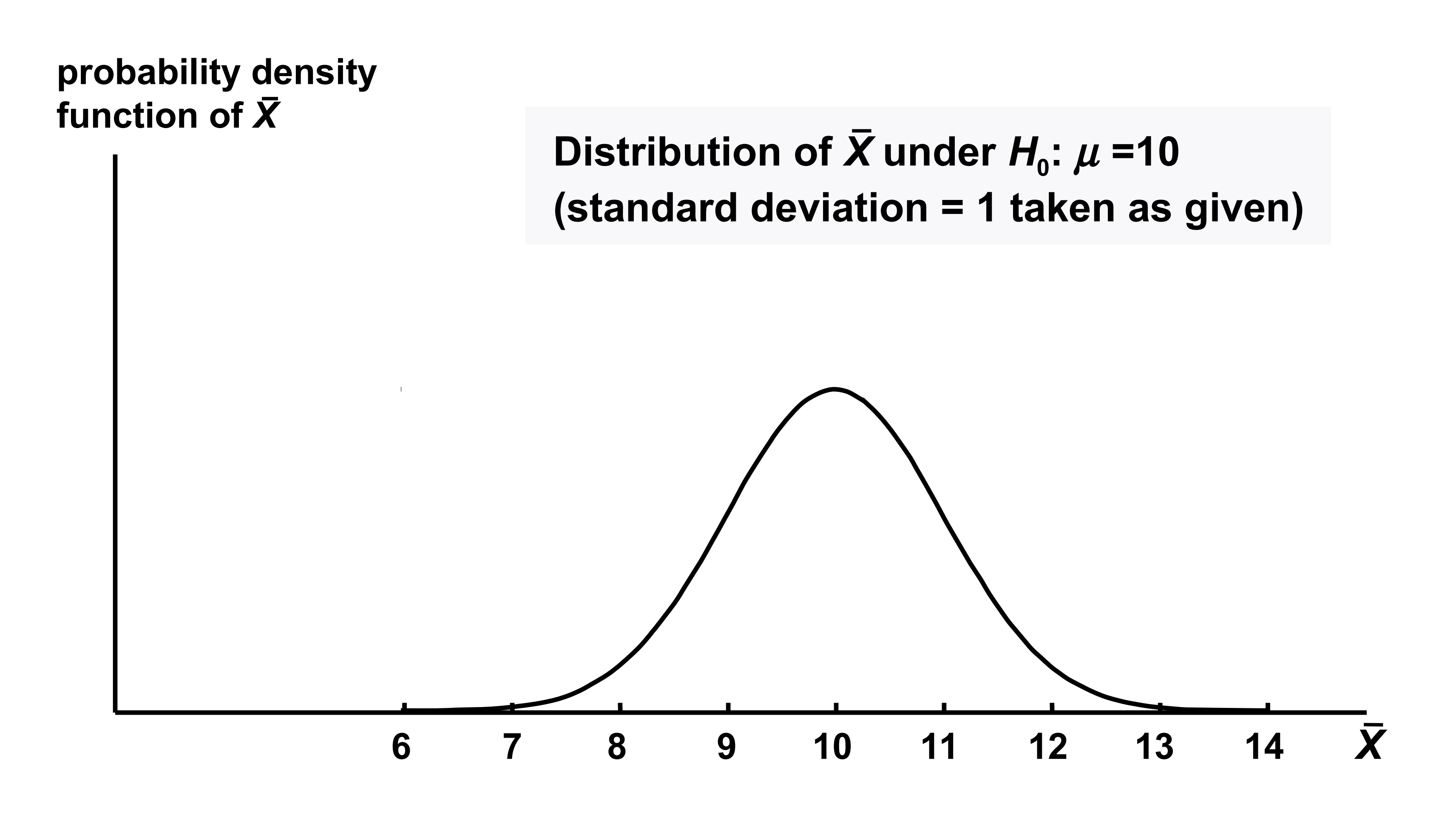
As an illustration, we suppose that the null hypothesis is that the mean is equal to 10 and the alternative hypothesis that it is not equal to 10.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



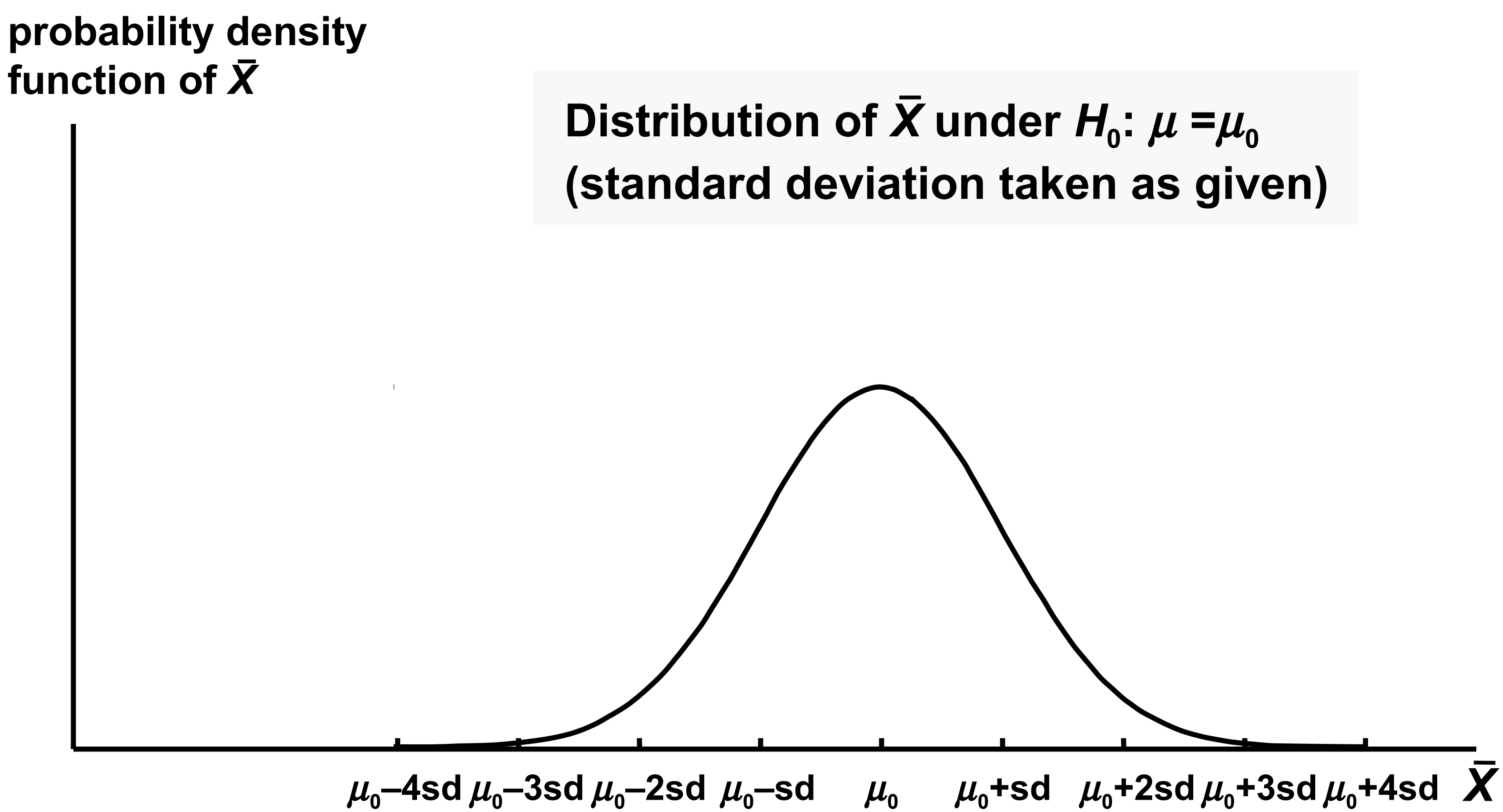
If this null hypothesis is true, $\bar{X}$ will have a distribution with mean 10. To draw the distribution, we must know its standard deviation.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



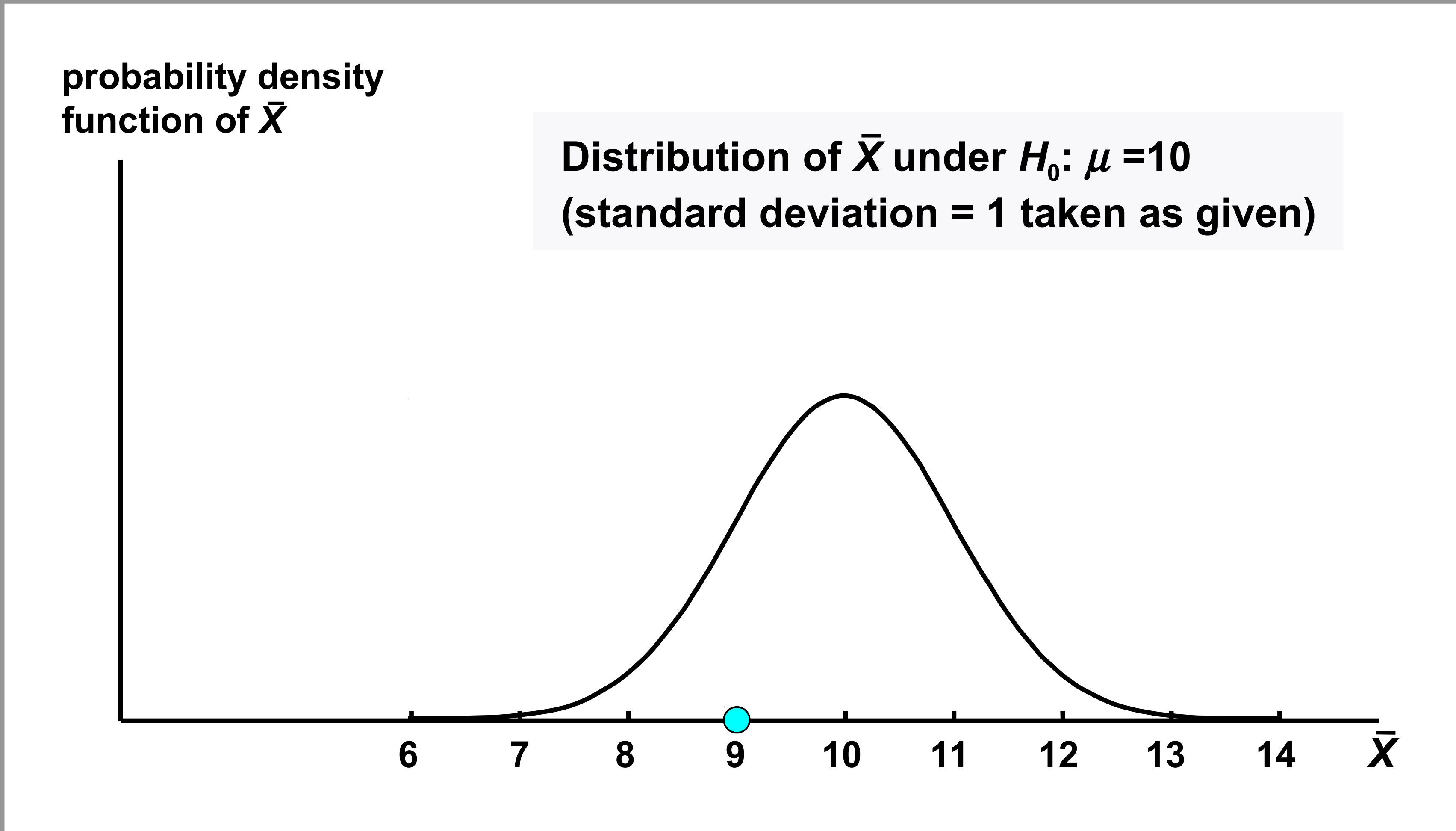
We will assume that we know the standard deviation of $\bar{X}$ and that it is equal to 1. This is a very unrealistic assumption. In practice you have to estimate it.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



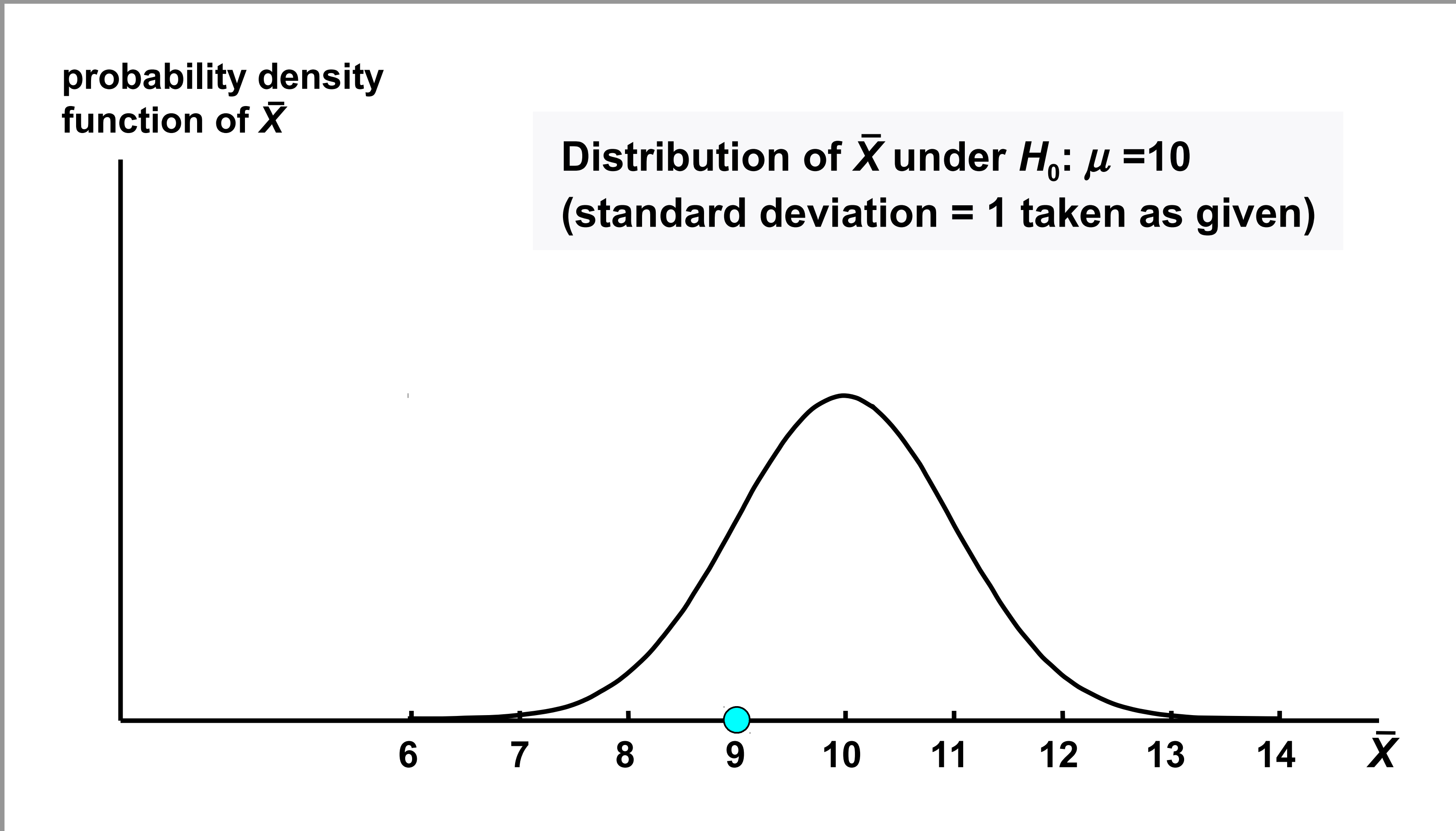
Here is the distribution of $\bar{X}$ for the general case. Again, for the time being we are assuming that we know its standard deviation (sd).

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

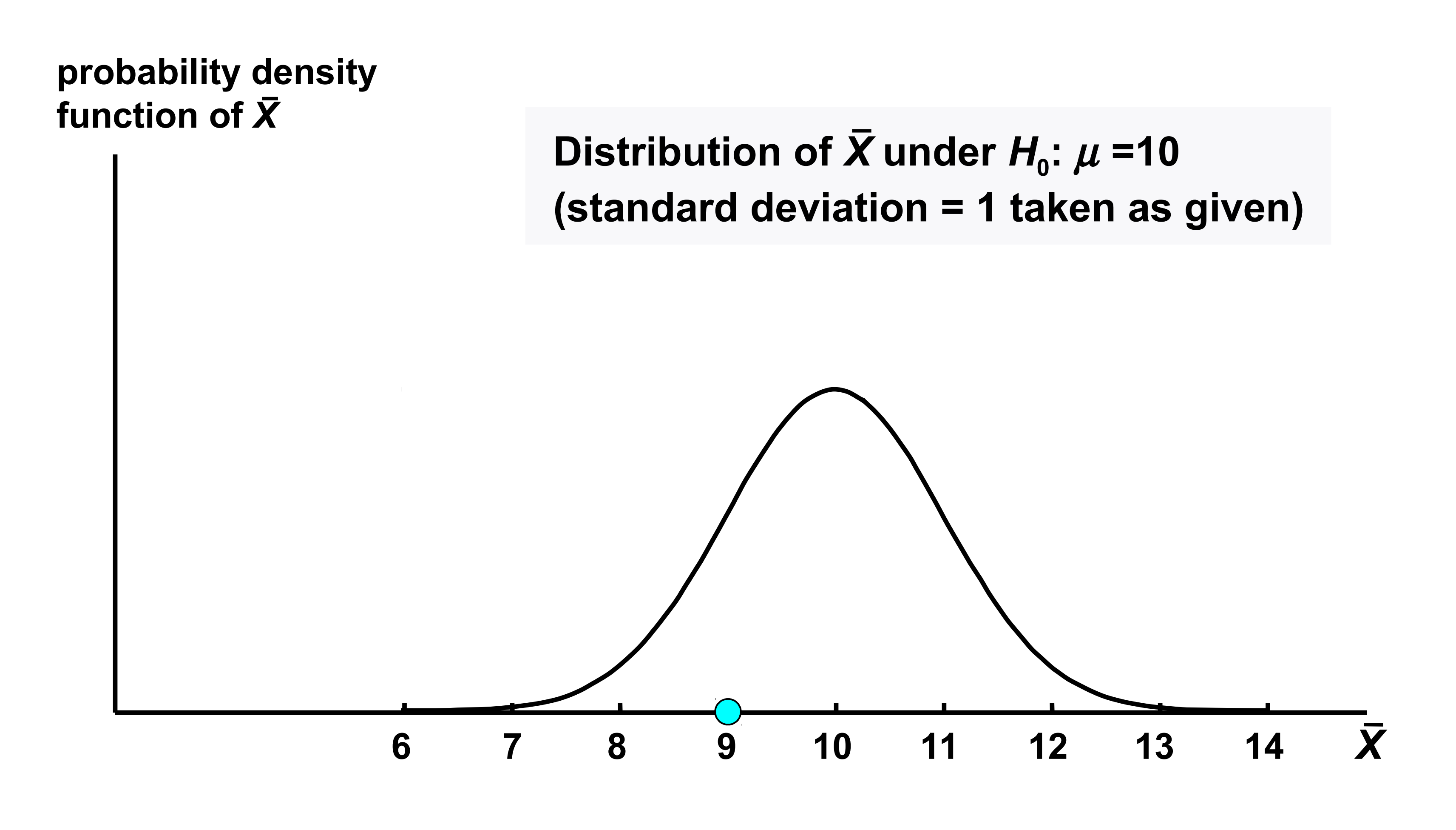


Suppose that we have a sample of data for the example model and the sample mean $\bar{X}$ is 9. Would this be evidence against the null hypothesis μ = 10?

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

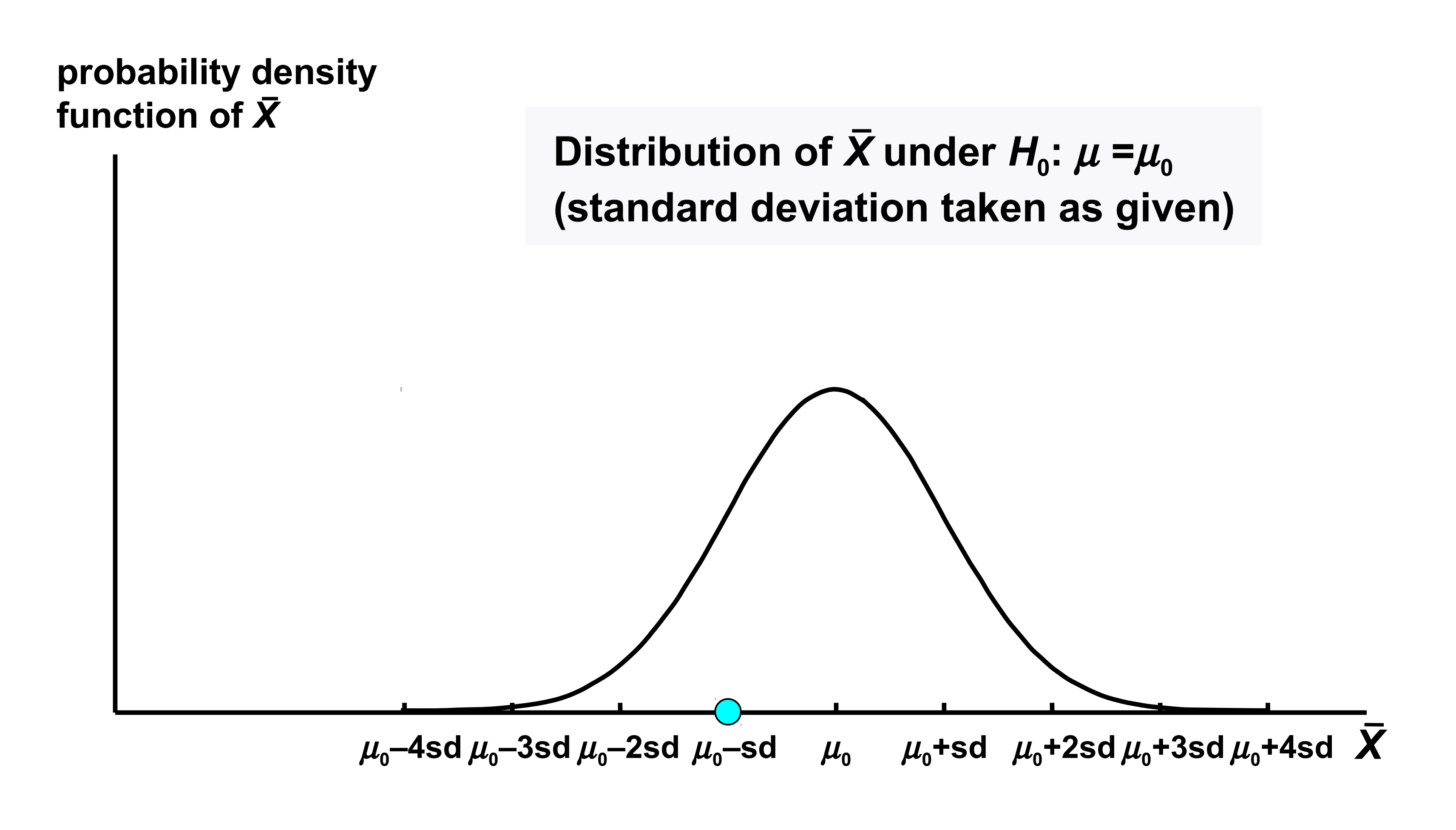


No, it is not. It is lower than 10, but we would not expect to be exactly equal to 10 because the sample mean has a random component.

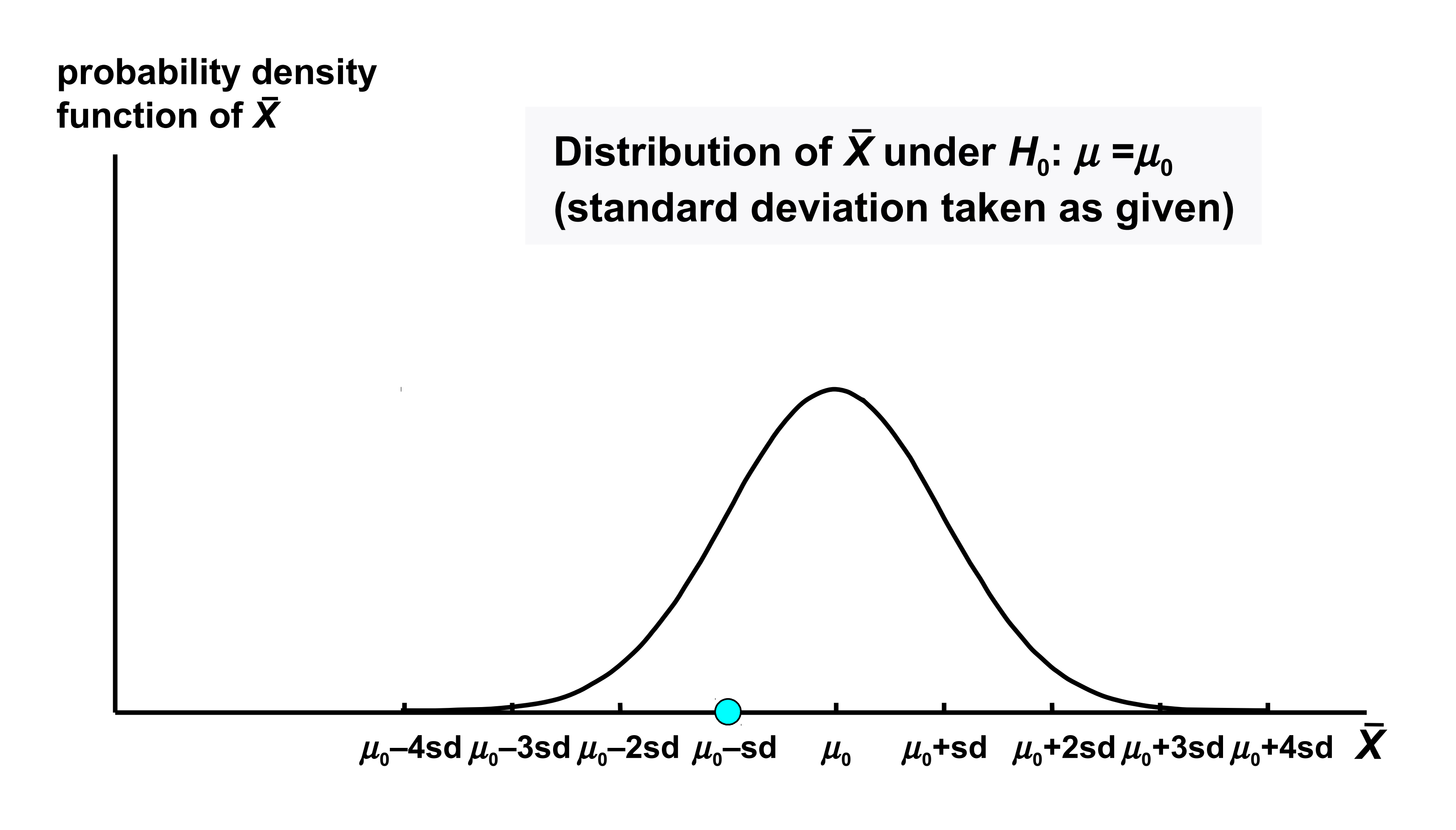


If the null hypothesis is true, we should frequently get estimates as low as 9, so there is no real conflict.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

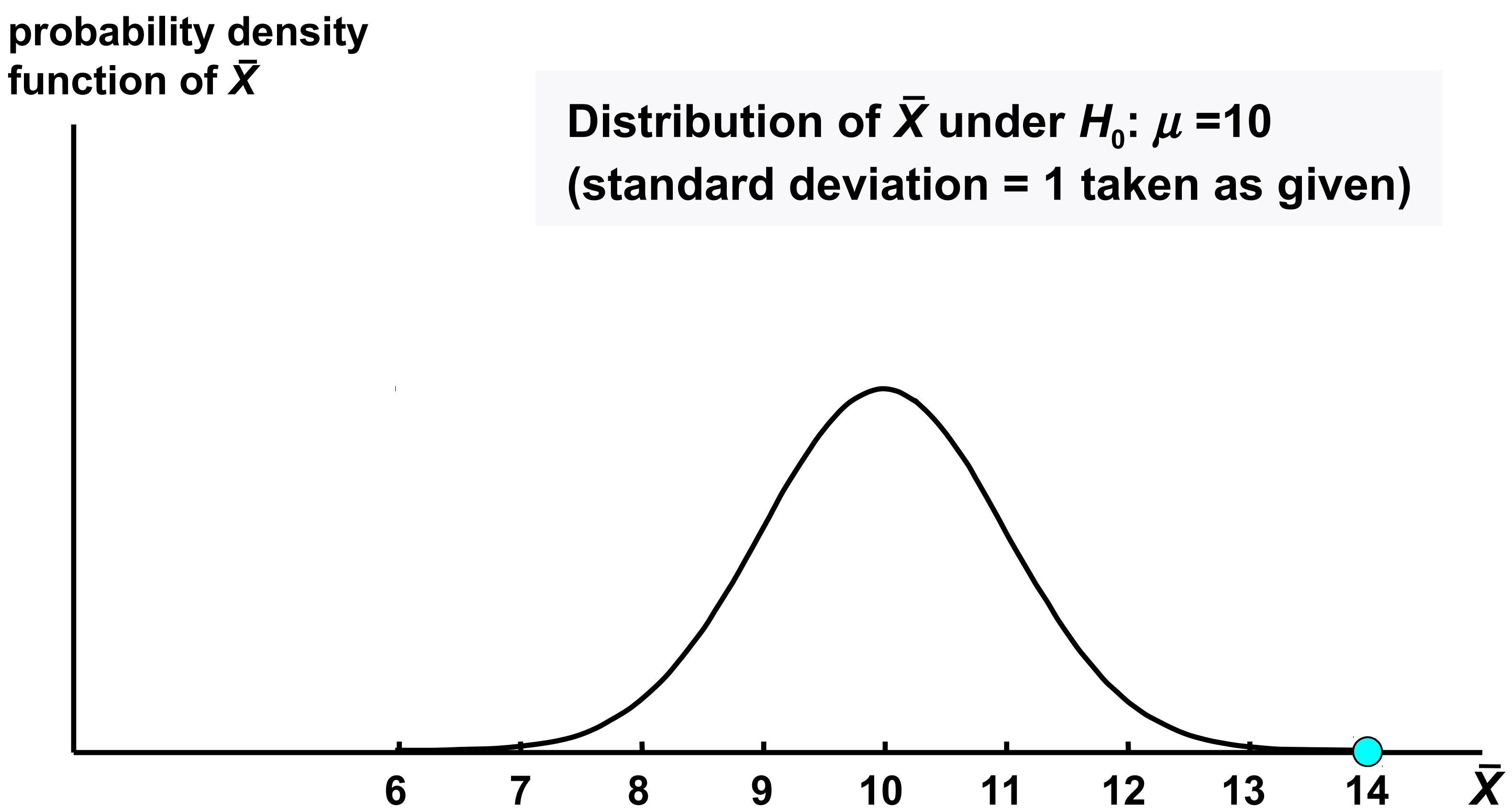


In terms of the general case, the sample mean is one standard deviation below the hypothetical population mean.



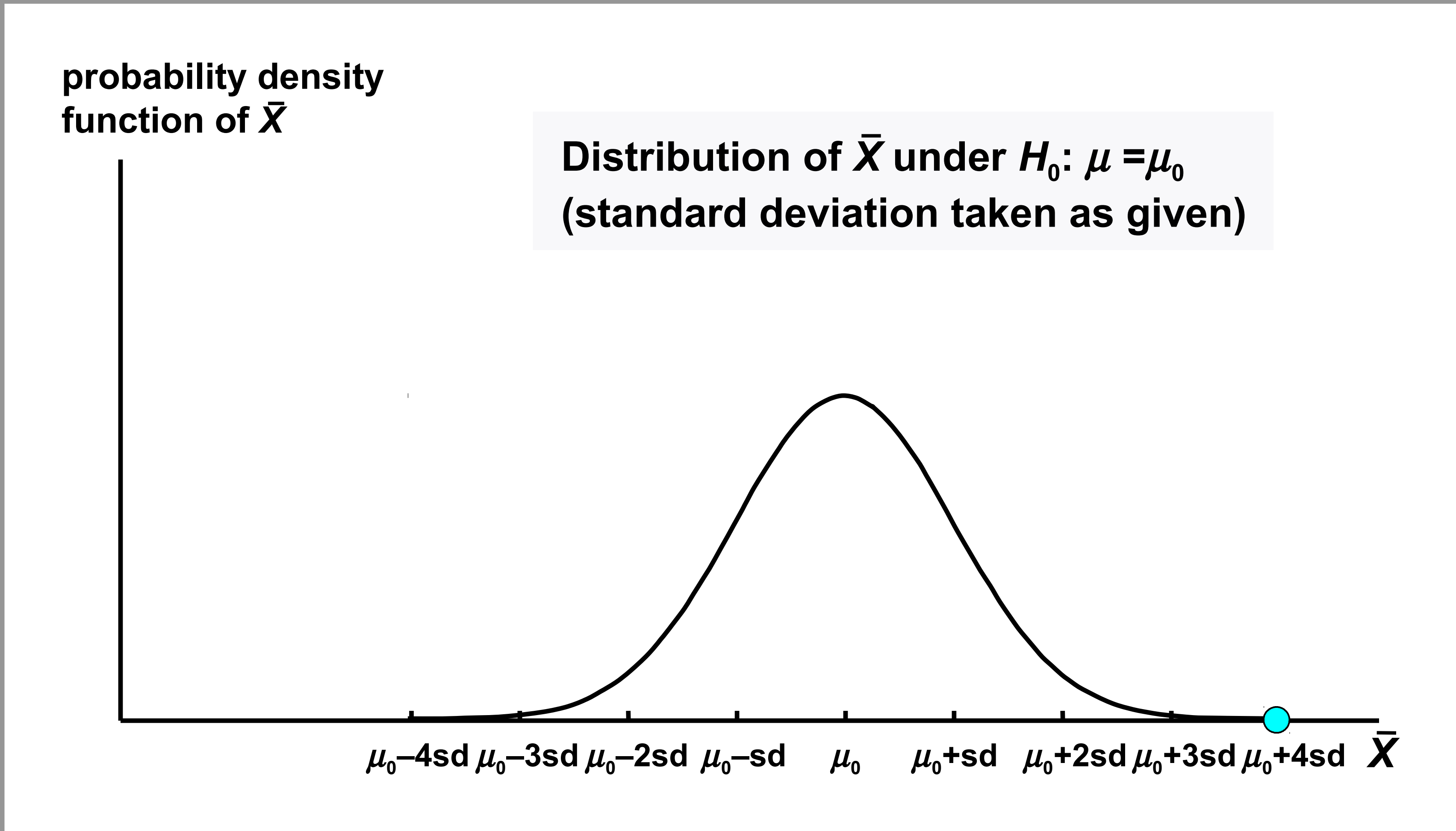
If the null hypothesis is true, the probability of the sample mean being one standard deviation or more above or below the population mean is 31.7%.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



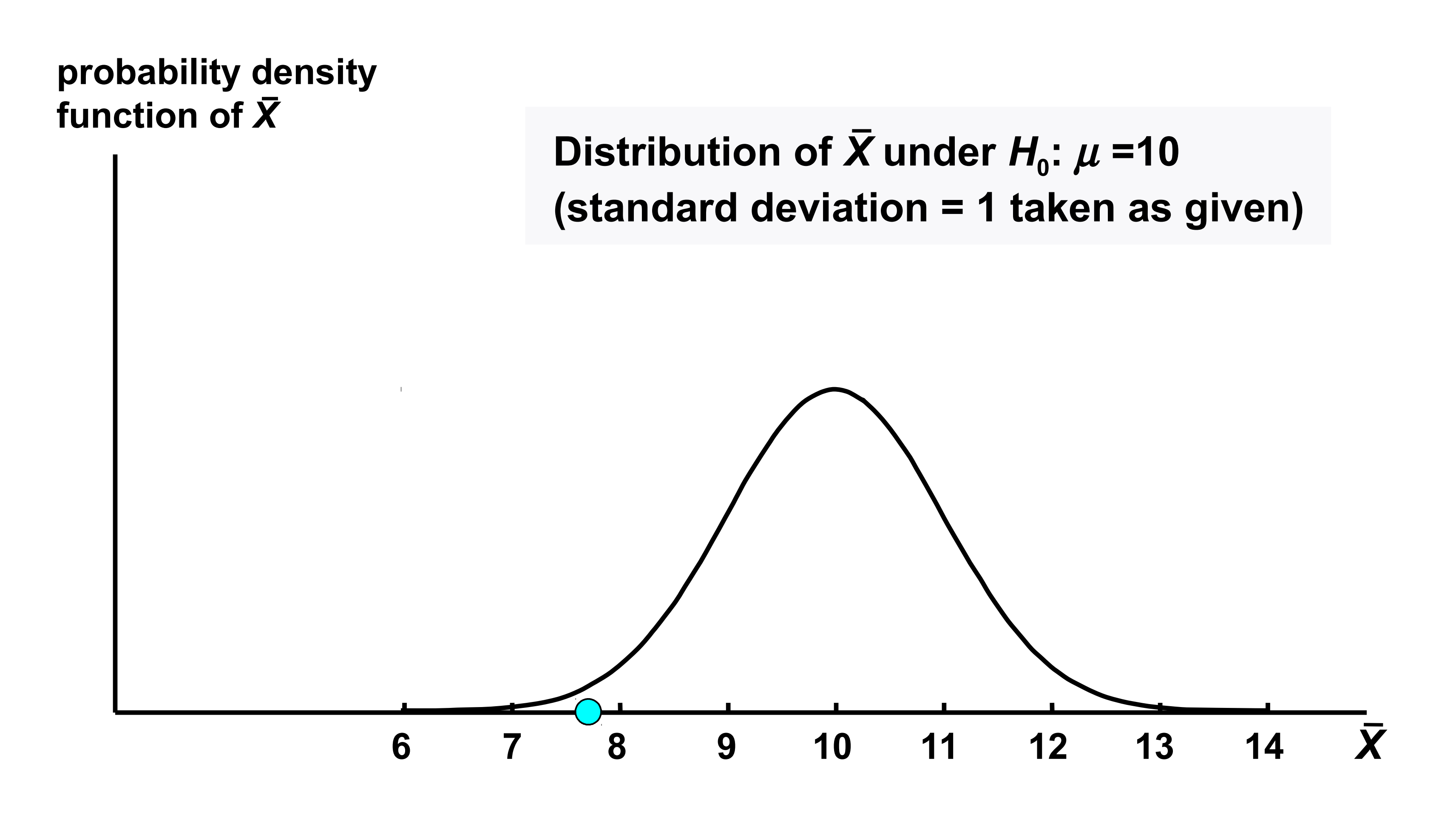
Now suppose that in the example model the sample mean is equal to 14. This clearly conflicts with the null hypothesis.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



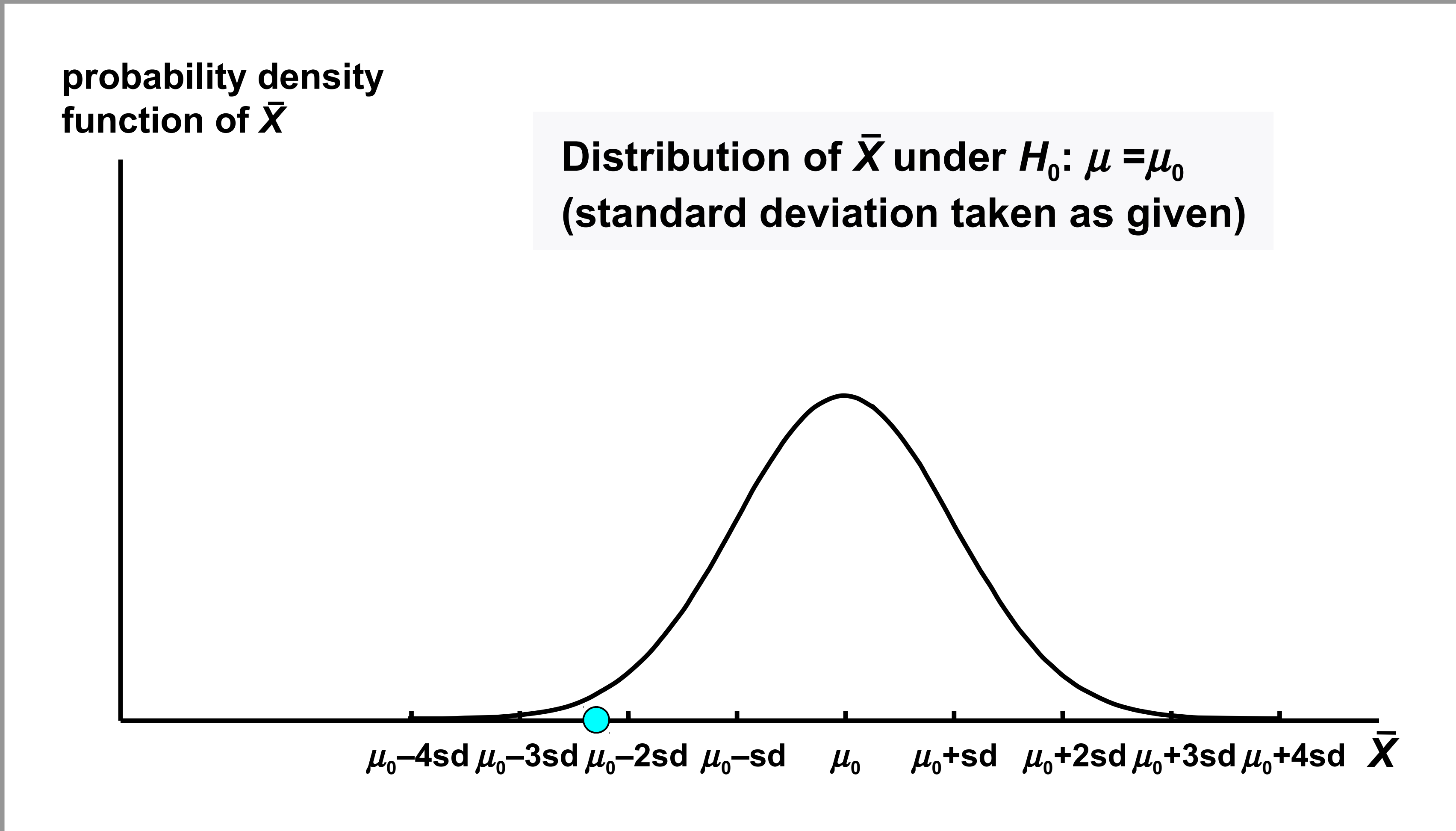
1.4 is four standard deviations above the hypothetical mean and the chance of getting such an extreme estimate is only 0.006%. We would reject the null hypothesis.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



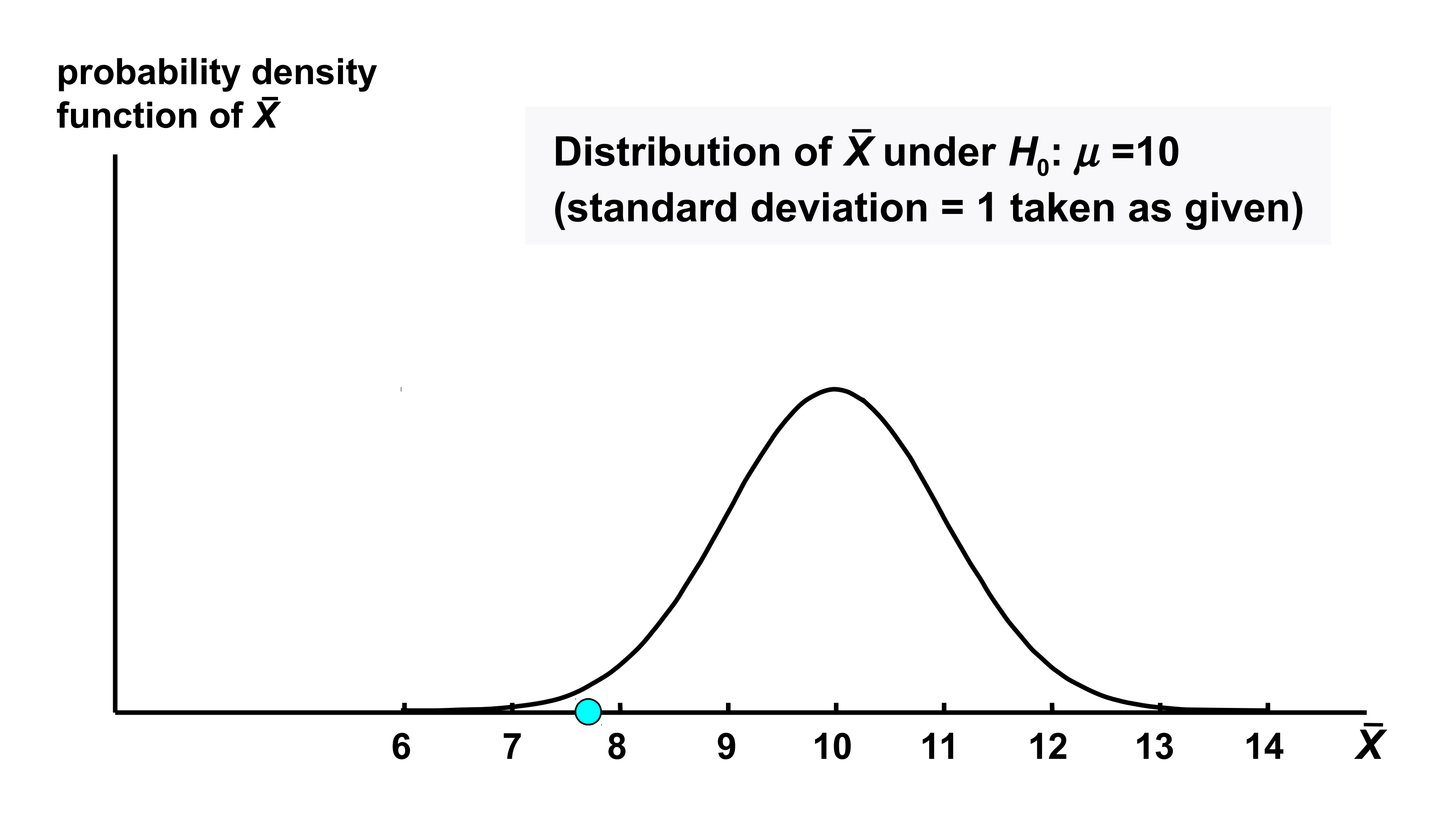
Now suppose that in the example model the sample mean is equal to 7.7. This is an awkward result.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

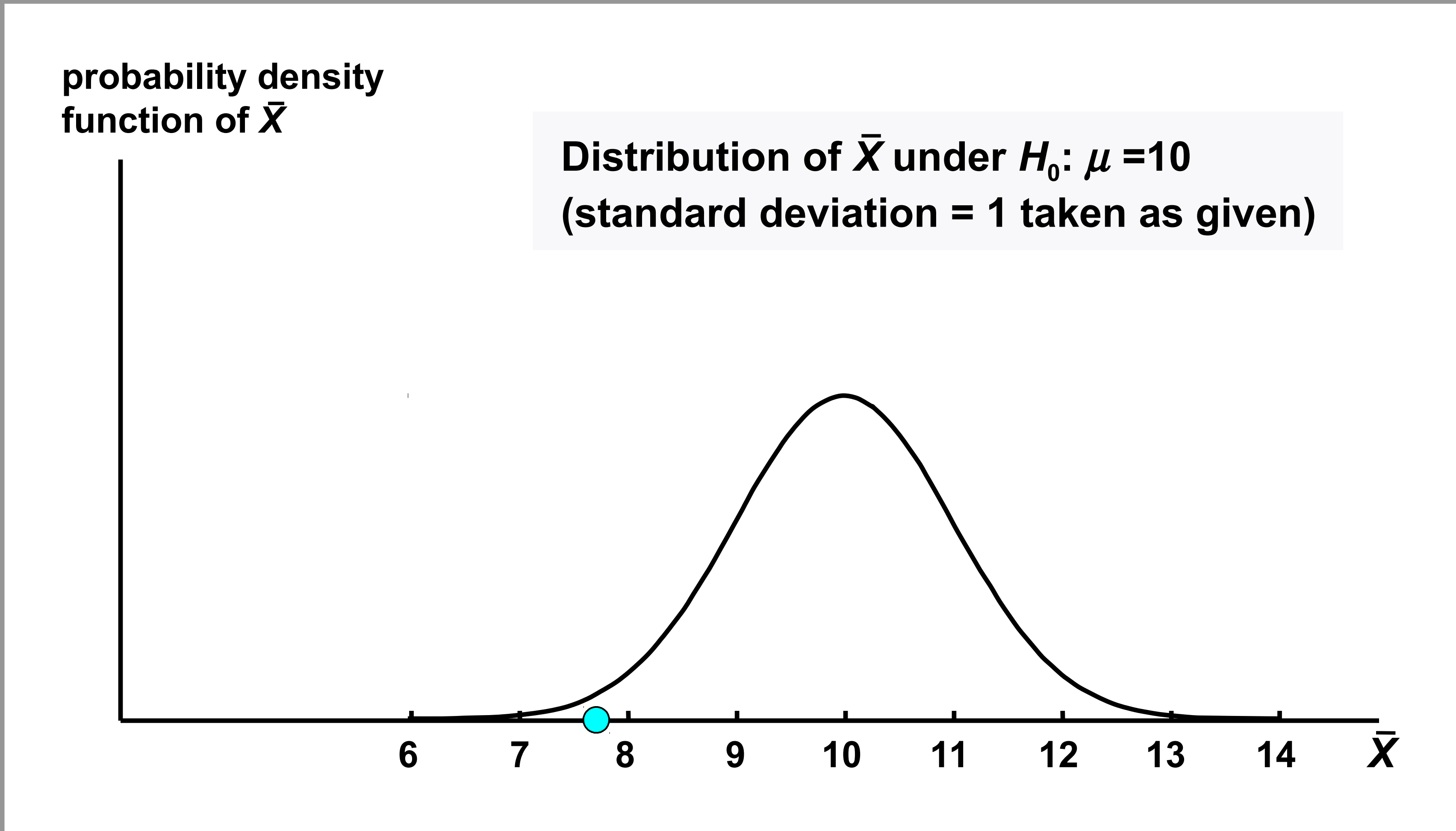


Under the null hypothesis, the estimate is between 2 and 3 standard deviations below the mean.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

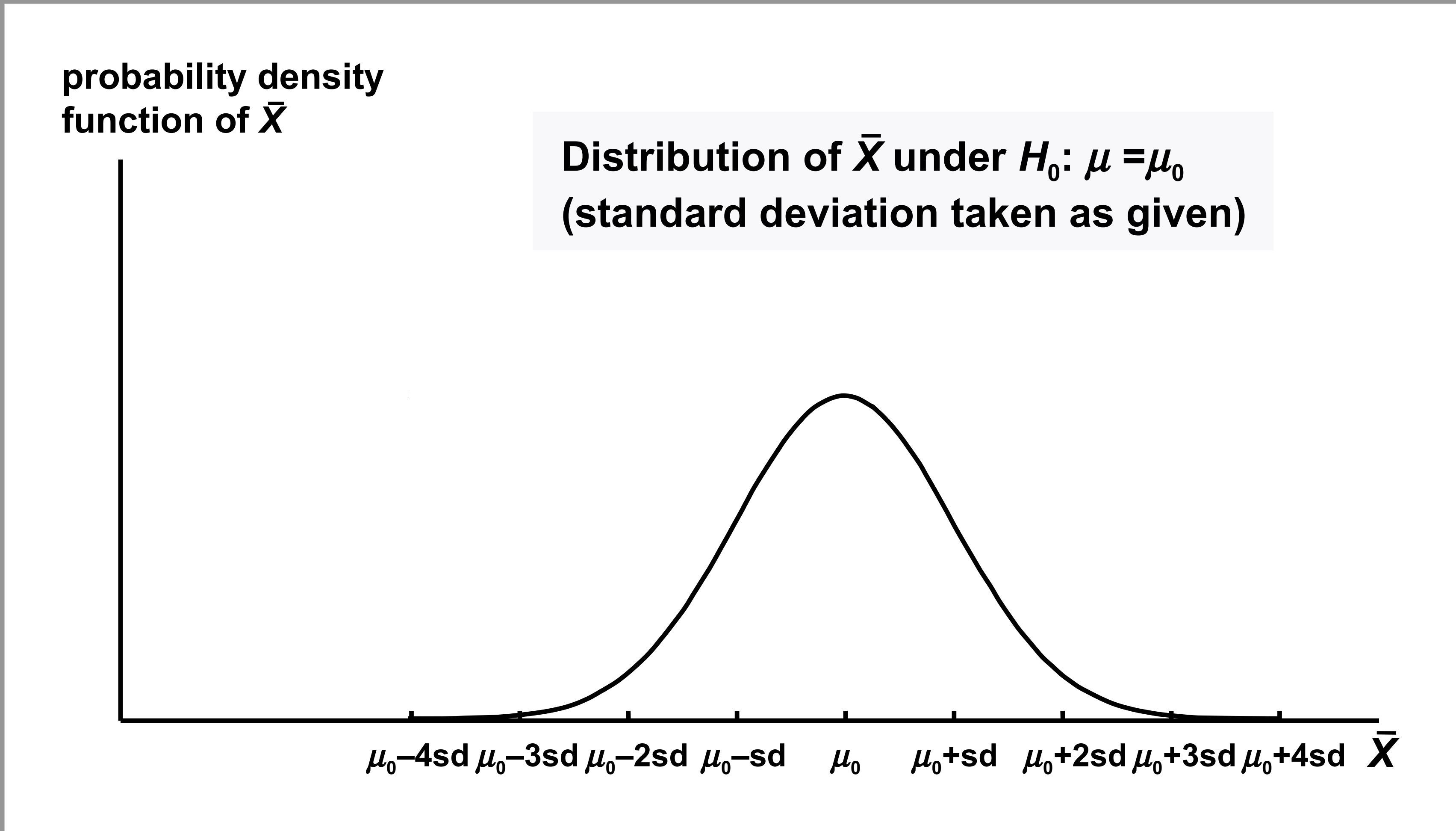


There are two possibilities. One is that the null hypothesis is true, and we have a slightly freaky sample mean.



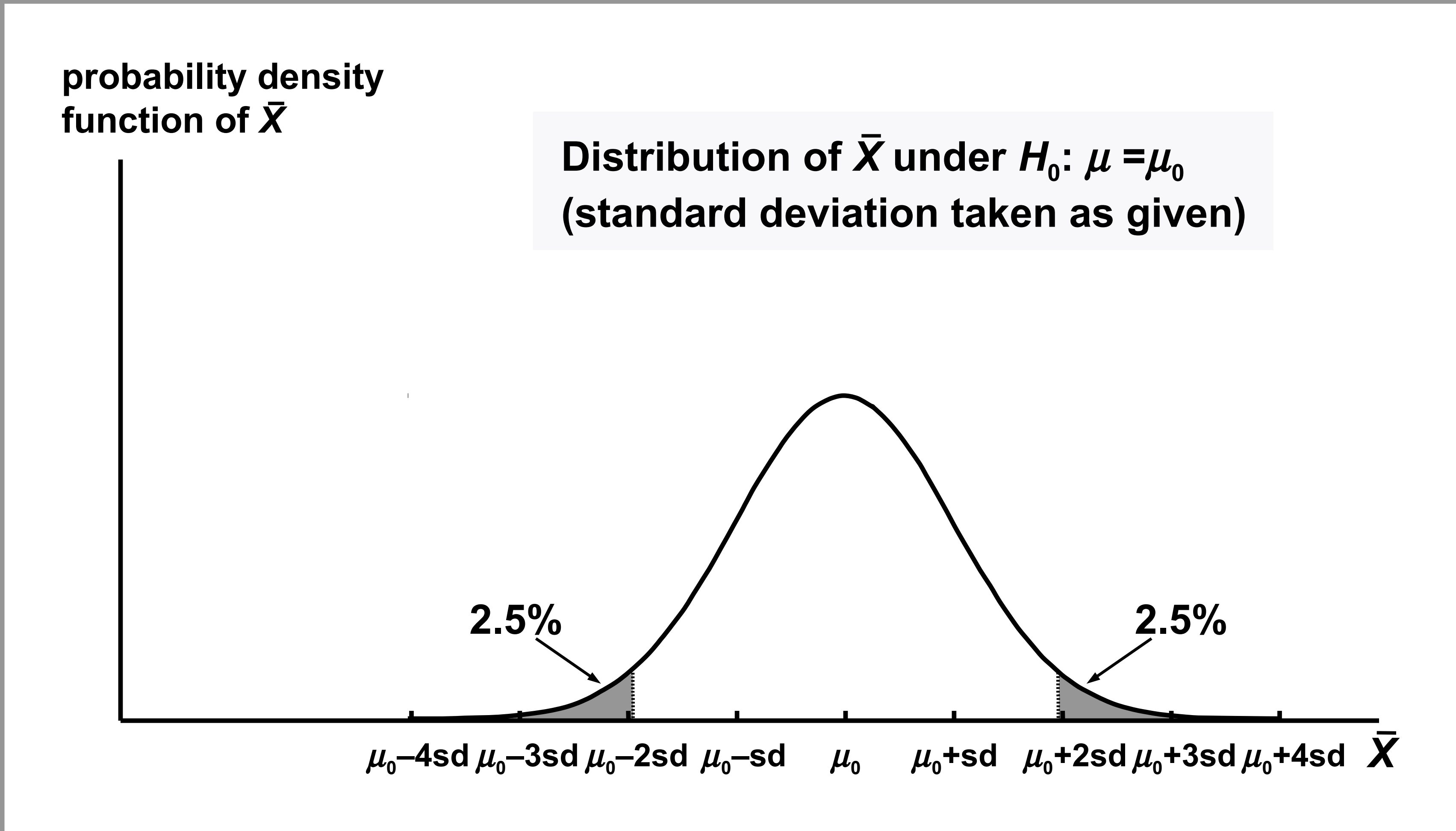
The other is that the null hypothesis is false. The population mean is not equal to 10.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



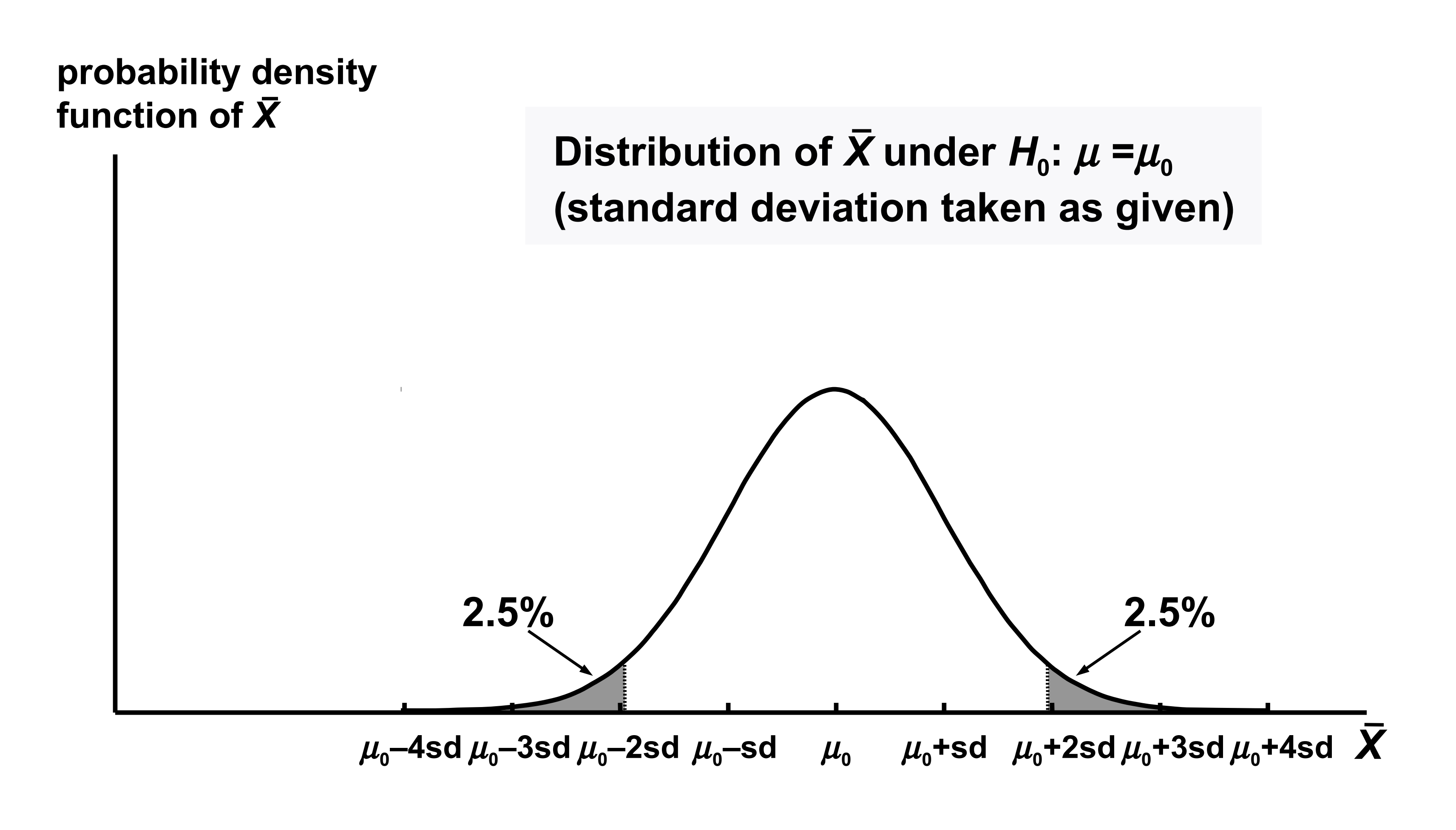
The usual procedure for making decisions is to reject the null hypothesis if it implies that the probability of getting such an extreme sample mean is less than some (small) probability *p*.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



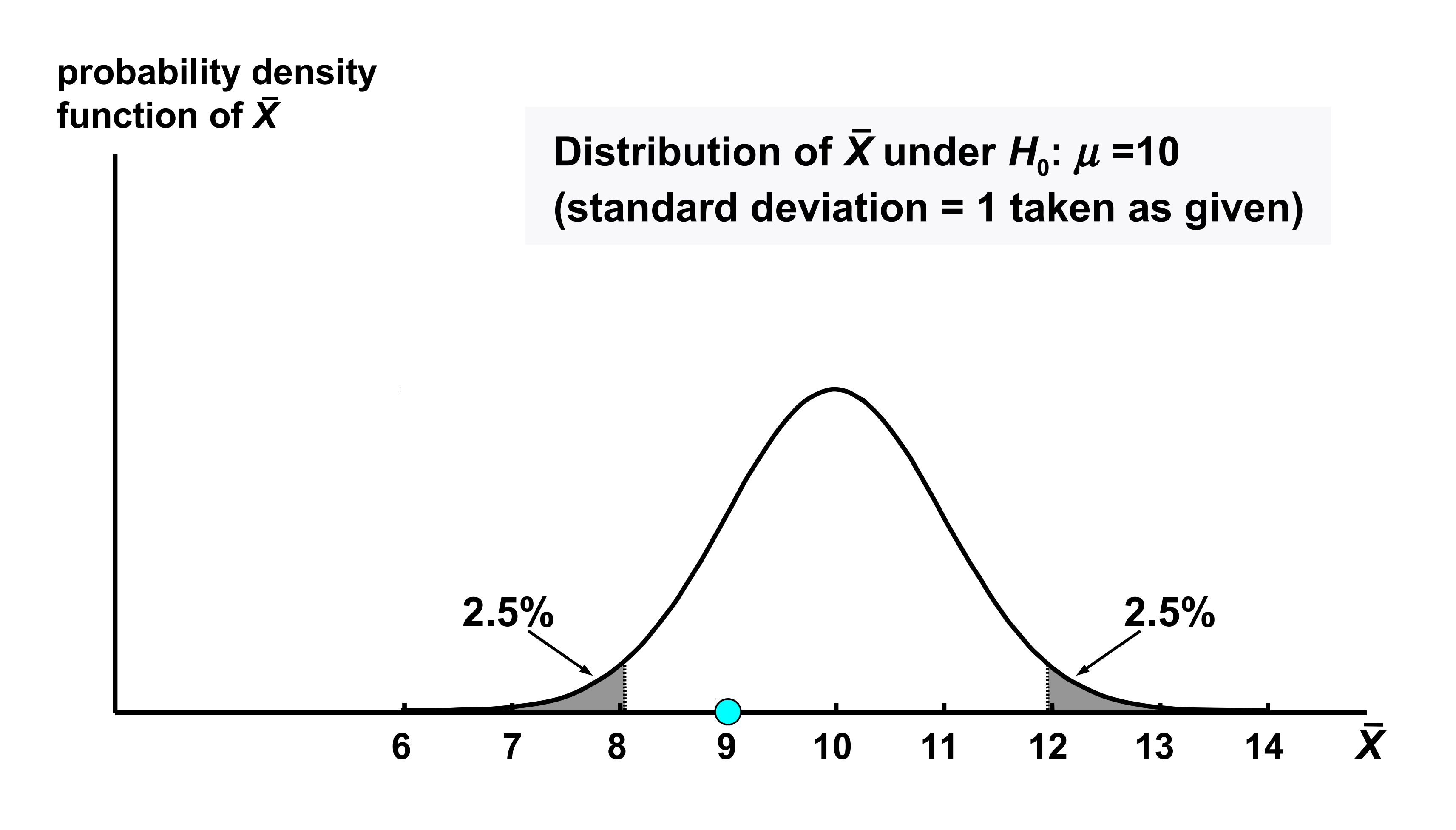
For example, we might choose to reject the null hypothesis if it implies that the probability of getting such an extreme sample mean is less than 0.05 (5%).

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



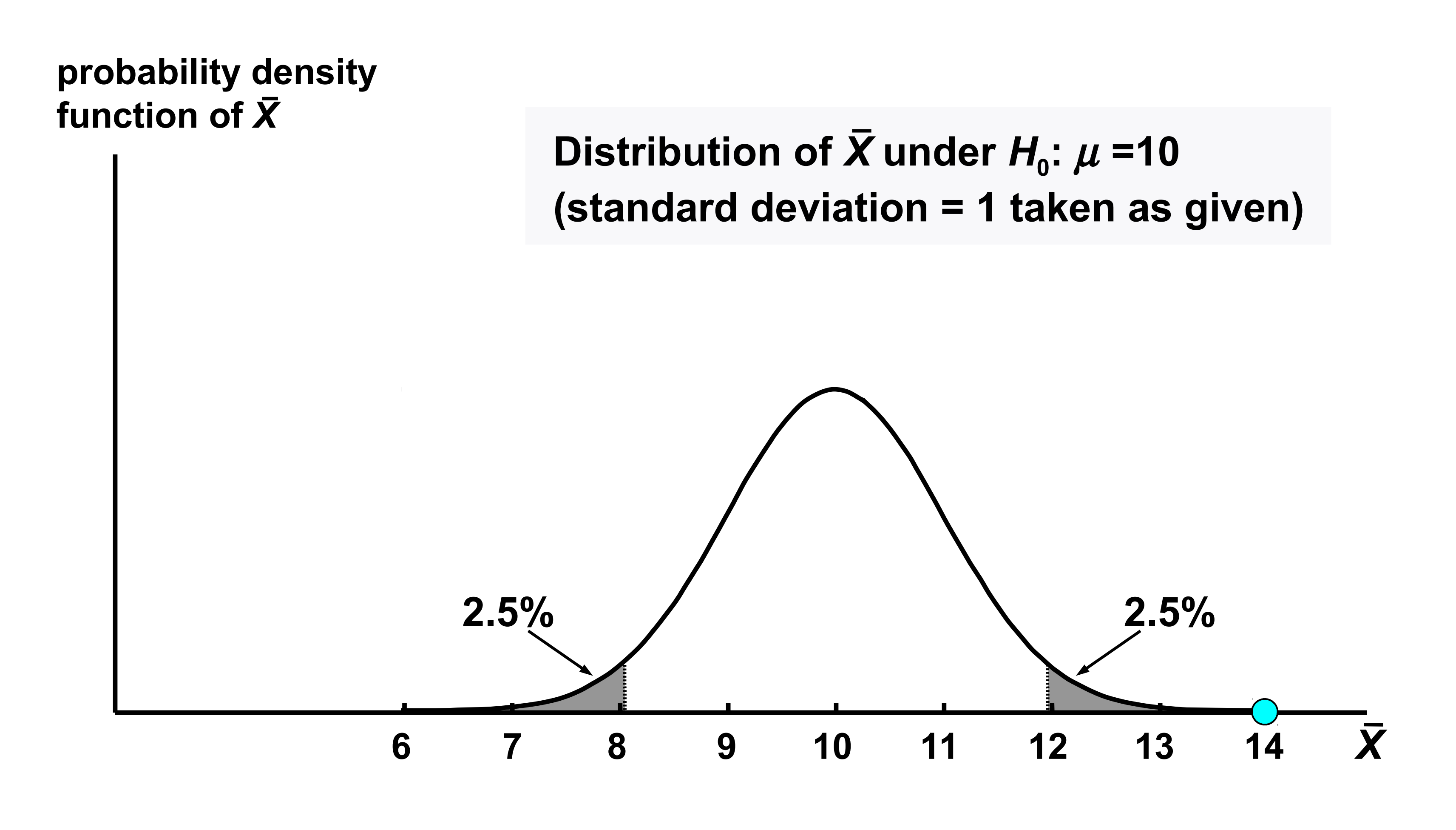
According to this decision rule, we would reject the null hypothesis if the sample mean fell in the upper or lower 2.5% tails.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



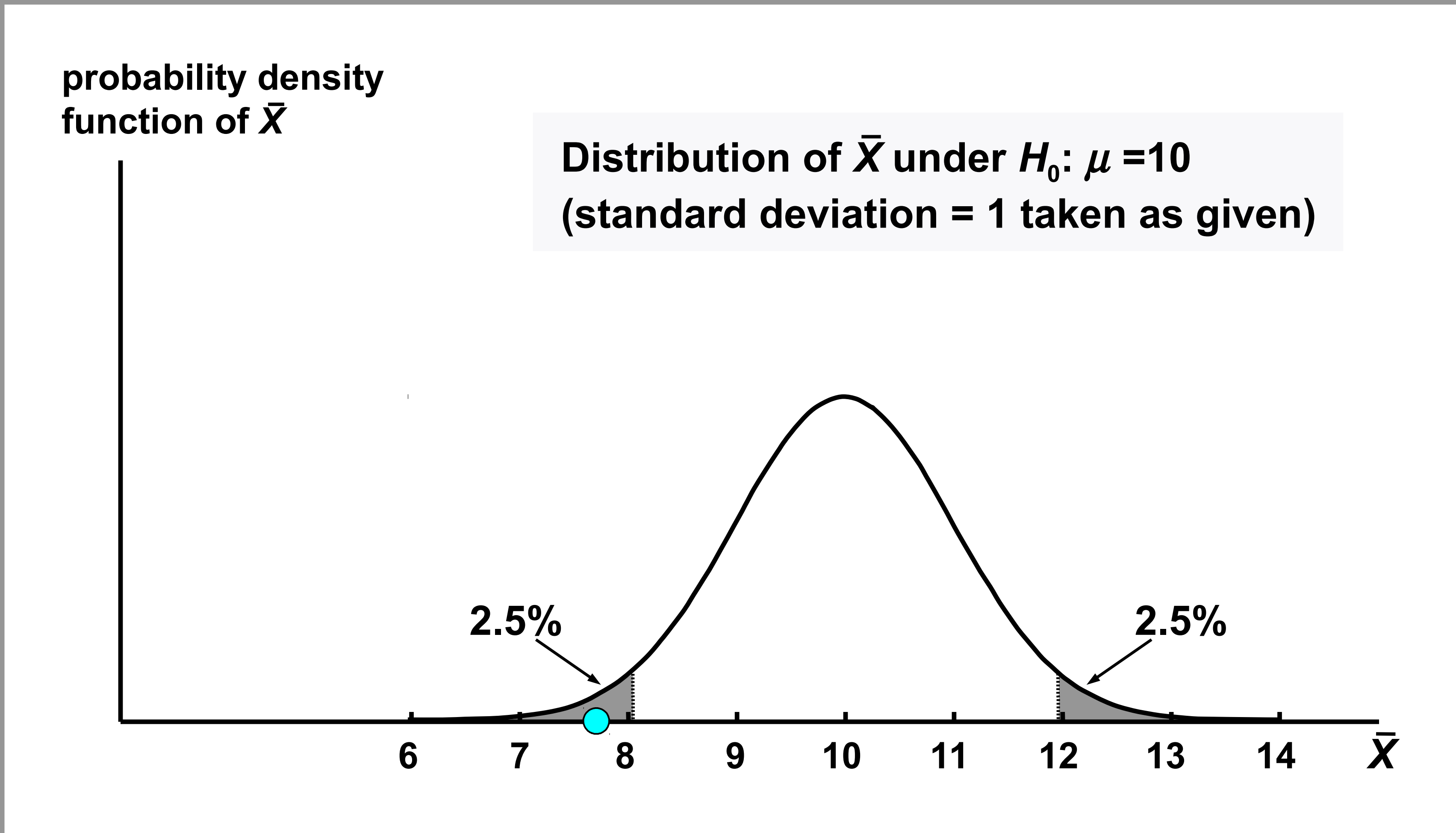
If we apply this decision rule to the example model, a sample mean of 9 would not lead to a rejection of the null hypothesis.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



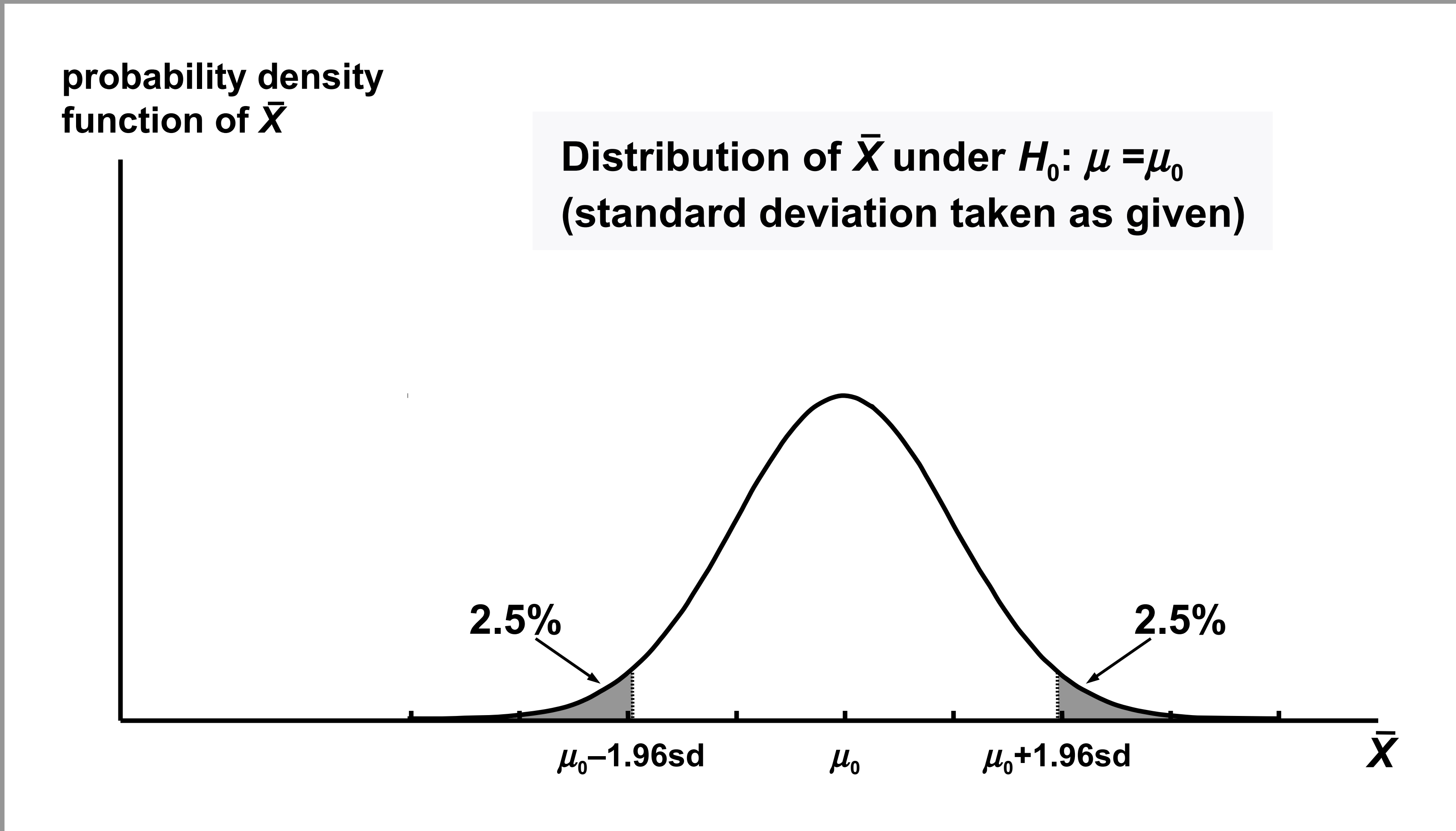
A sample mean of 14 definitely would lead us to reject H_0.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



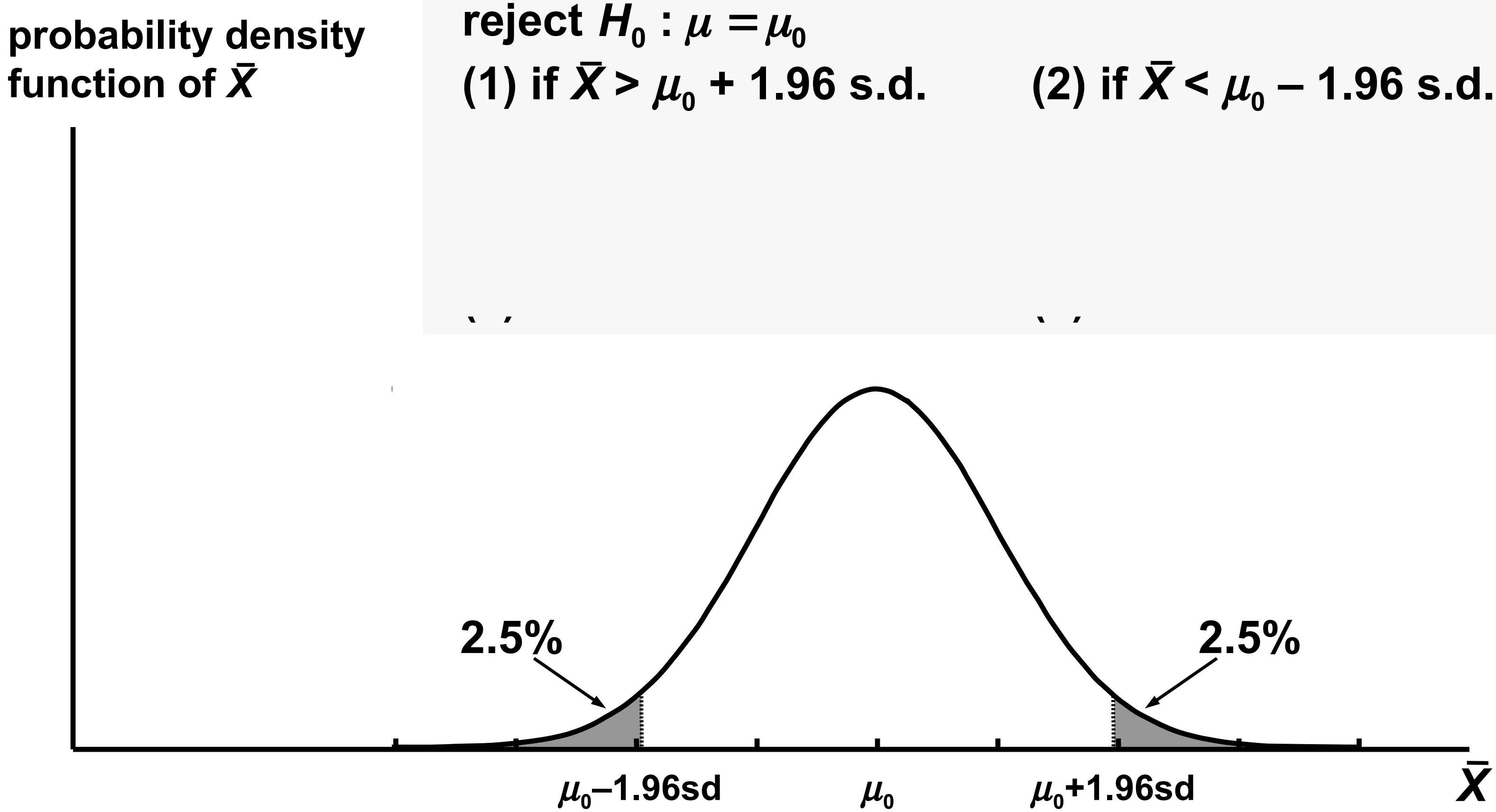
A sample mean of 7.7 also would lead to rejection..

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



The 2.5% tails of a normal distribution always begin 1.96 standard deviations from its mean.

Decision rule (5% significance level):
reject $H_0 : \mu = \mu_0$
(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

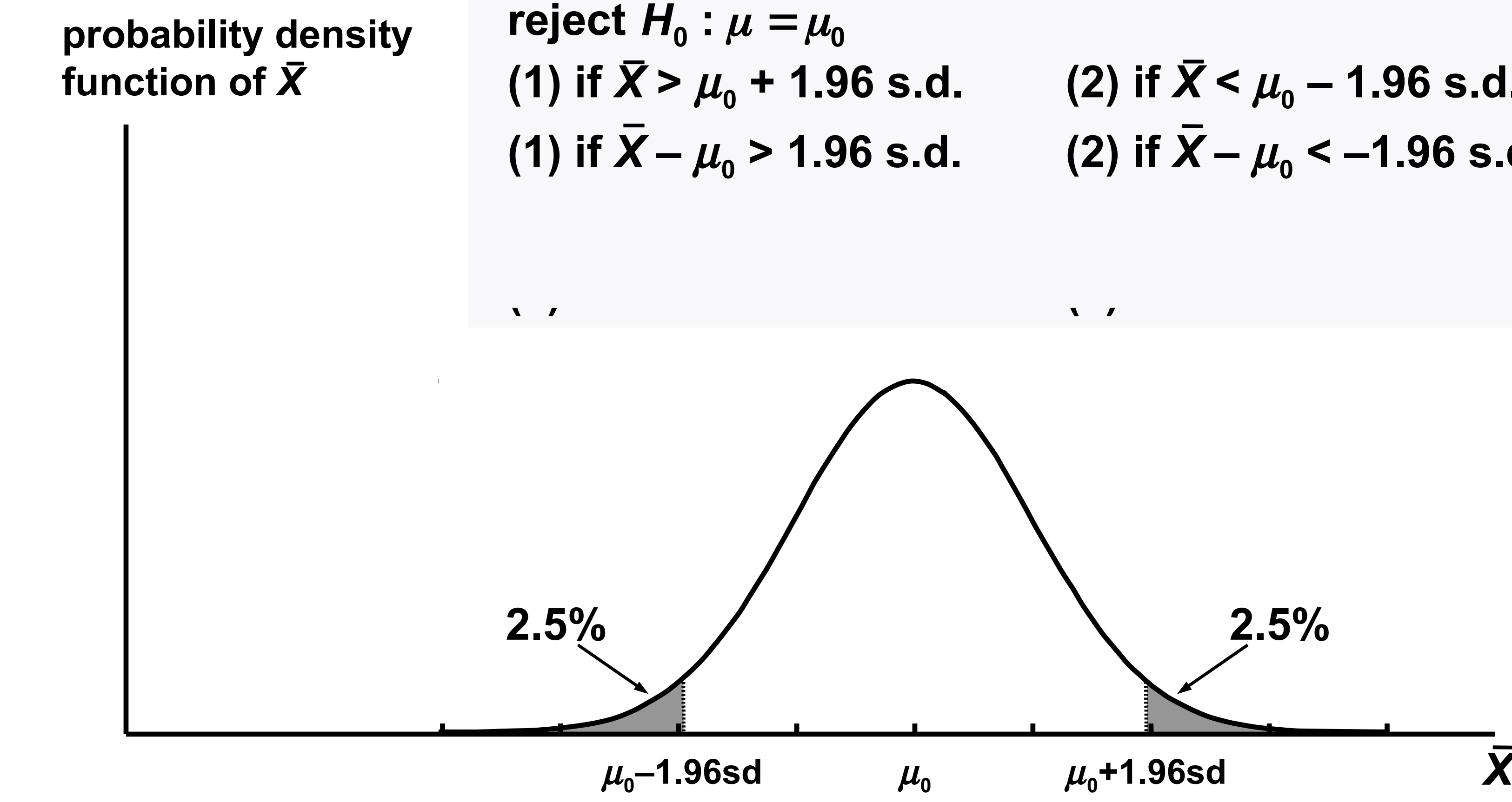


Thus we would reject H_0 if the sample mean were 1.96 standard deviations (or more) above or below the hypothetical population mean.

Decision rule (5% significance level):
reject $H_0 : \mu = \mu_0$

(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

(1) if $\bar{X} - \mu_0 > 1.96$ s.d. **(2) if $\bar{X} - \mu_0 < -1.96$ s.d.**



We would reject H_0 if the difference between the sample mean and hypothetical population mean were more than 1.96 standard deviations.

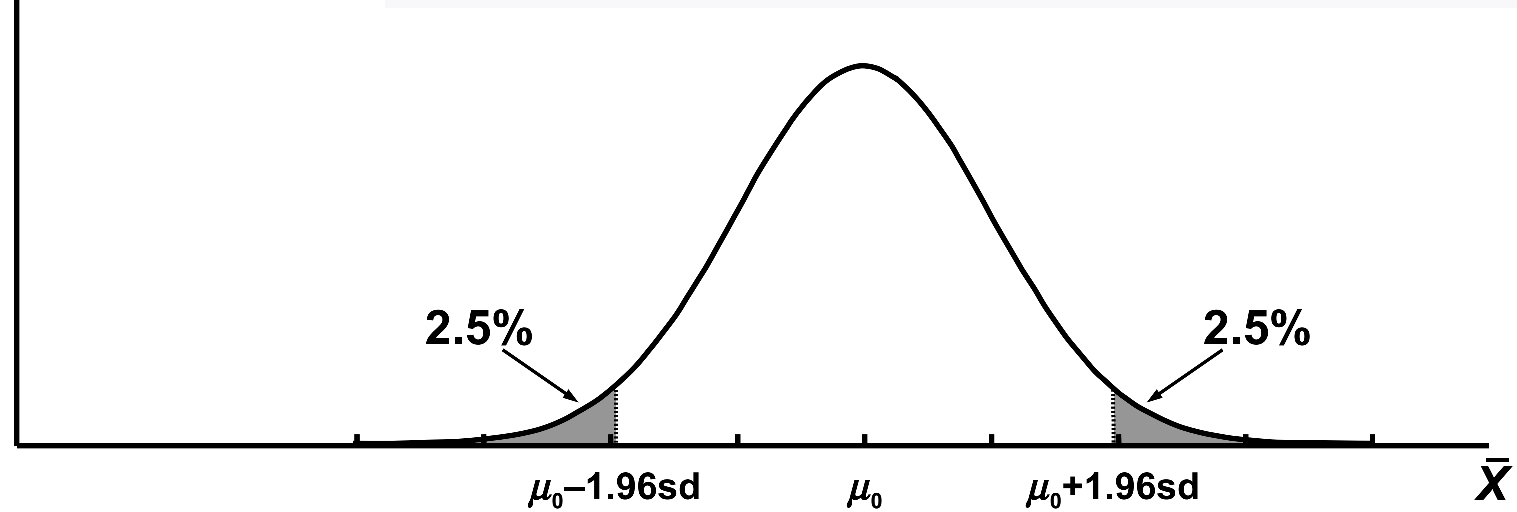
Decision rule (5% significance level):
reject $H_0: \mu = \mu_0$

(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

(1) if $\bar{X} - \mu_0 > 1.96$ s.d. **(2) if $\bar{X} - \mu_0 < -1.96$ s.d.**

(1) if $(\bar{X} - \mu_0)$ / s.d. > 1.96 **(2) if $(\bar{X} - \mu_0)$ / s.d. < -1.96**

probability density function of $\bar{X}$



We would reject H_0 if the difference, expressed in terms of standard deviations, were more than 1.96 in absolute terms (positive or negative).

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (5% significance level):

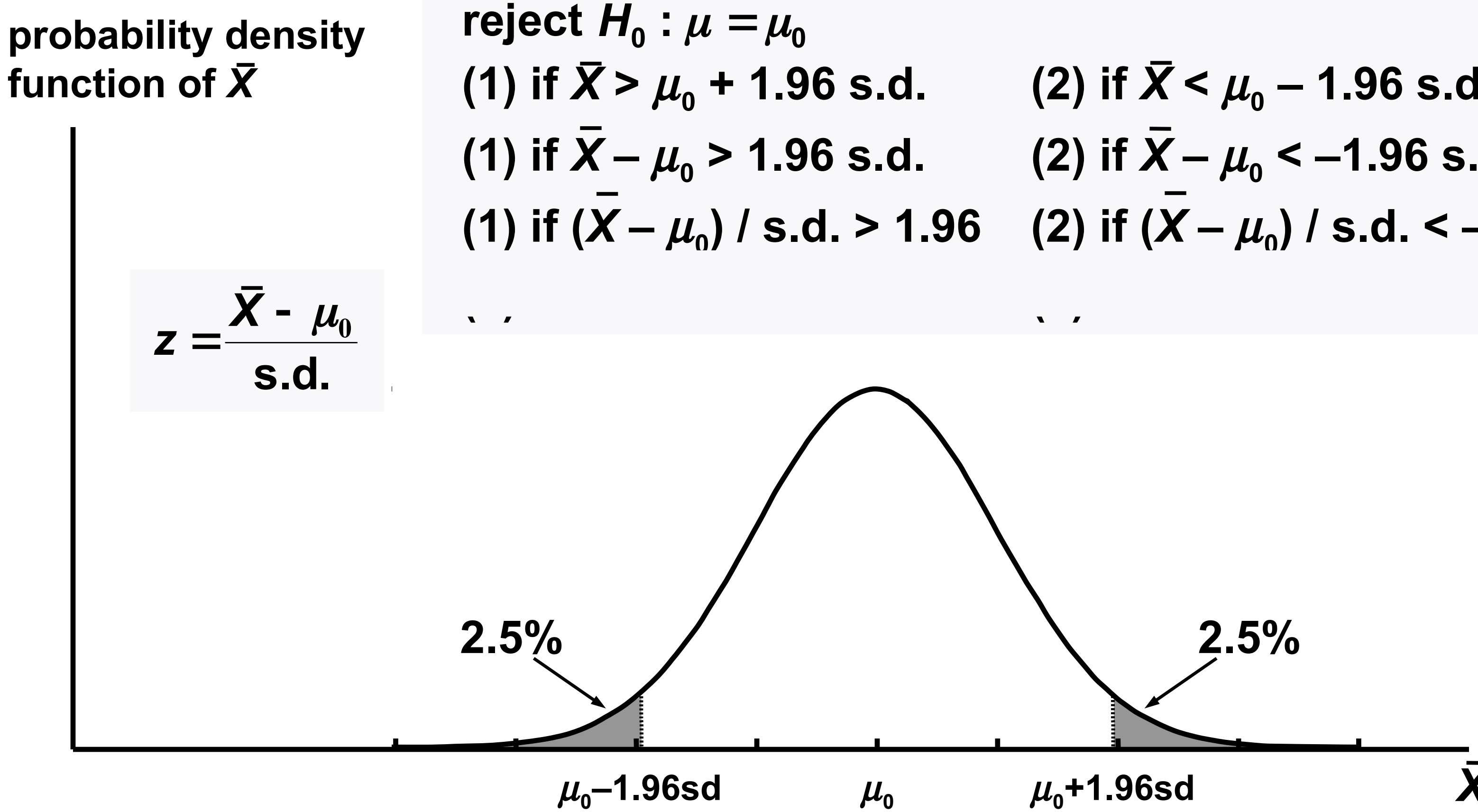
reject $H_0 : \mu = \mu_0$

(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

(1) if $\bar{X} - \mu_0 > 1.96$ s.d. **(2) if $\bar{X} - \mu_0 < -1.96$ s.d.**

(1) if $(\bar{X} - \mu_0)$ / s.d. > 1.96 **(2) if $(\bar{X} - \mu_0)$ / s.d. < –1.96**

$$z = \frac{\bar{X} - \mu_0}{\text{s.d.}}$$



We will denote the difference, expressed in terms of standard deviations, as z.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (5% significance level):
reject $H_0 : \mu = \mu_0$

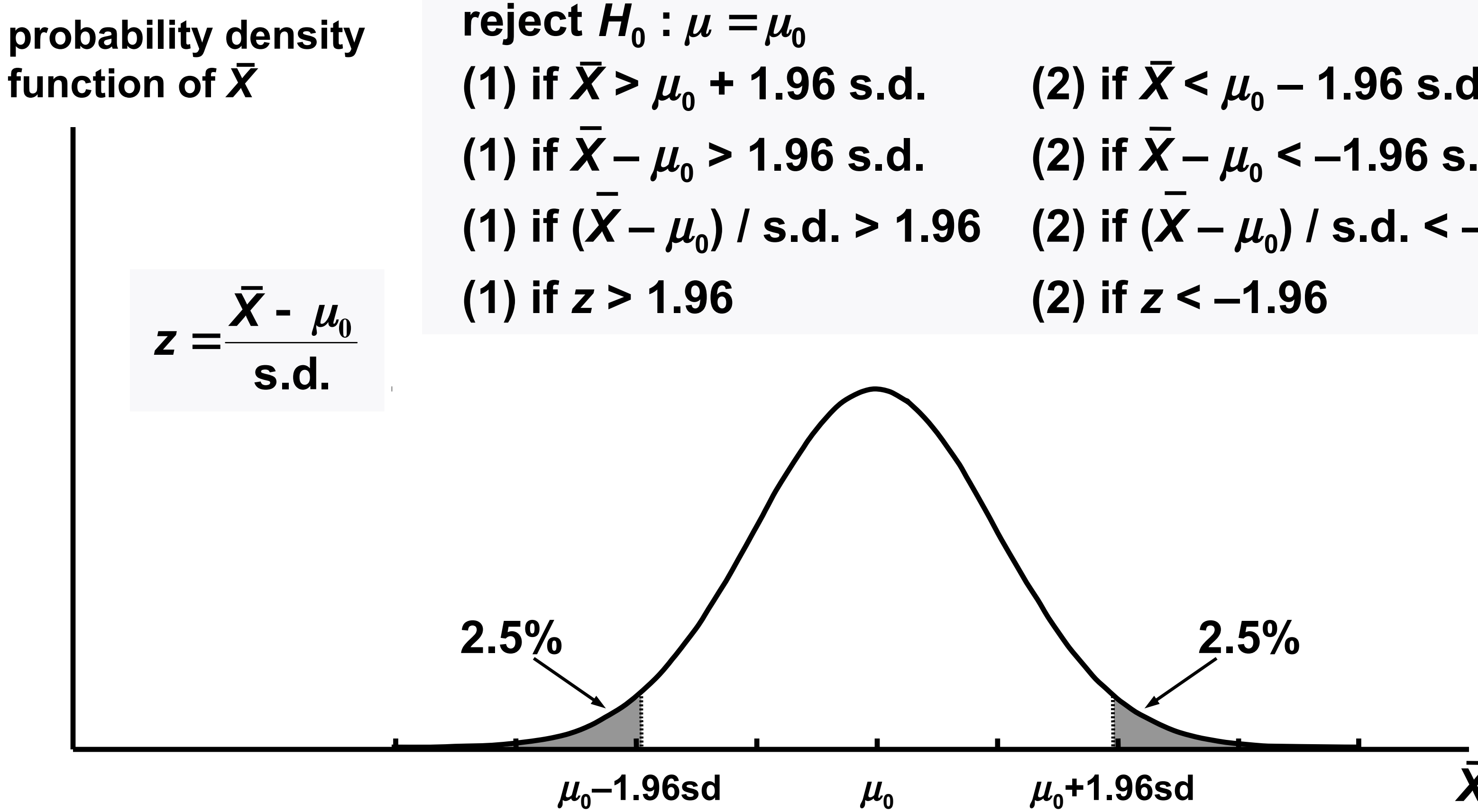
(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

(1) if $\bar{X} - \mu_0 > 1.96$ s.d. **(2) if $\bar{X} - \mu_0 < -1.96$ s.d.**

(1) if $(\bar{X} - \mu_0)$ / s.d. > 1.96 **(2) if $(\bar{X} - \mu_0)$ / s.d. < -1.96**

(1) if $z > 1.96$ **(2) if $z < -1.96$**

$$z = \frac{\bar{X} - \mu_0}{\text{s.d.}}$$



Then the decision rule is to reject the null hypothesis if z is greater than 1.96 in absolute terms.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (5% significance level):
reject $H_0 : \mu = \mu_0$

(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

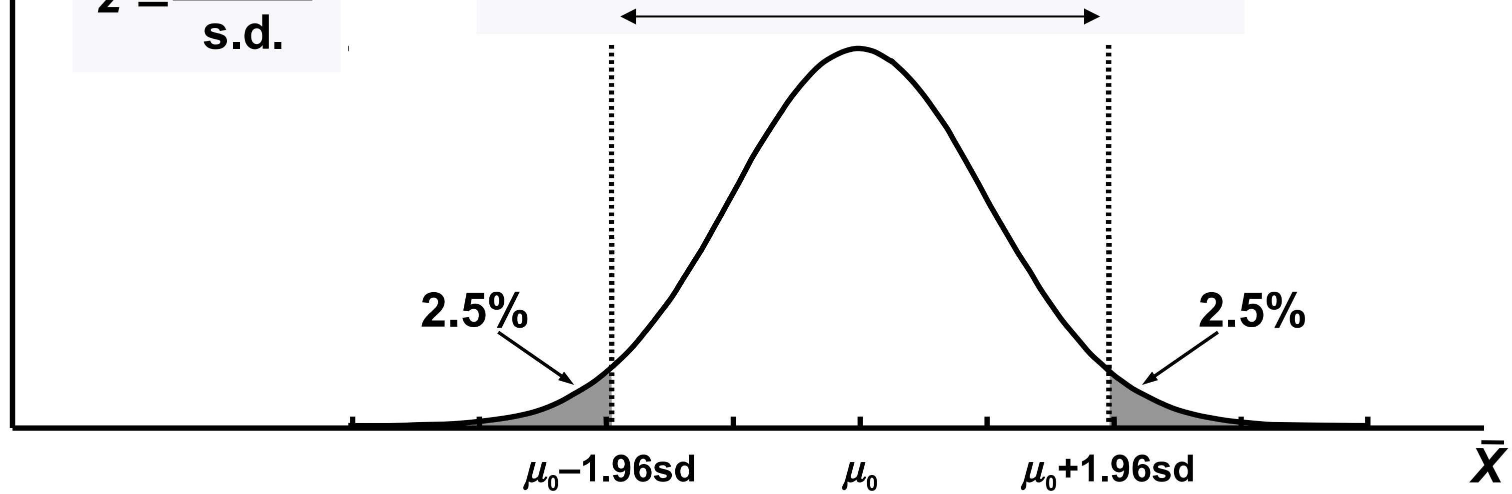
(1) if $z > 1.96$ **(2) if $z < -1.96$**

probability density function of $\bar{X}$

$$z = \frac{\bar{X} - \mu_0}{\text{s.d.}}$$

Acceptance region for $\bar{X}$:

$$\mu_0 - 1.96\ \text{s.d.} \le \bar{X} \le \mu_0 + 1.96\ \text{s.d.}$$



The range of values of $\bar{X}$ that do not lead to the rejection of the null hypothesis is known as the acceptance region.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (5% significance level):
reject $H_0 : \mu = \mu_0$

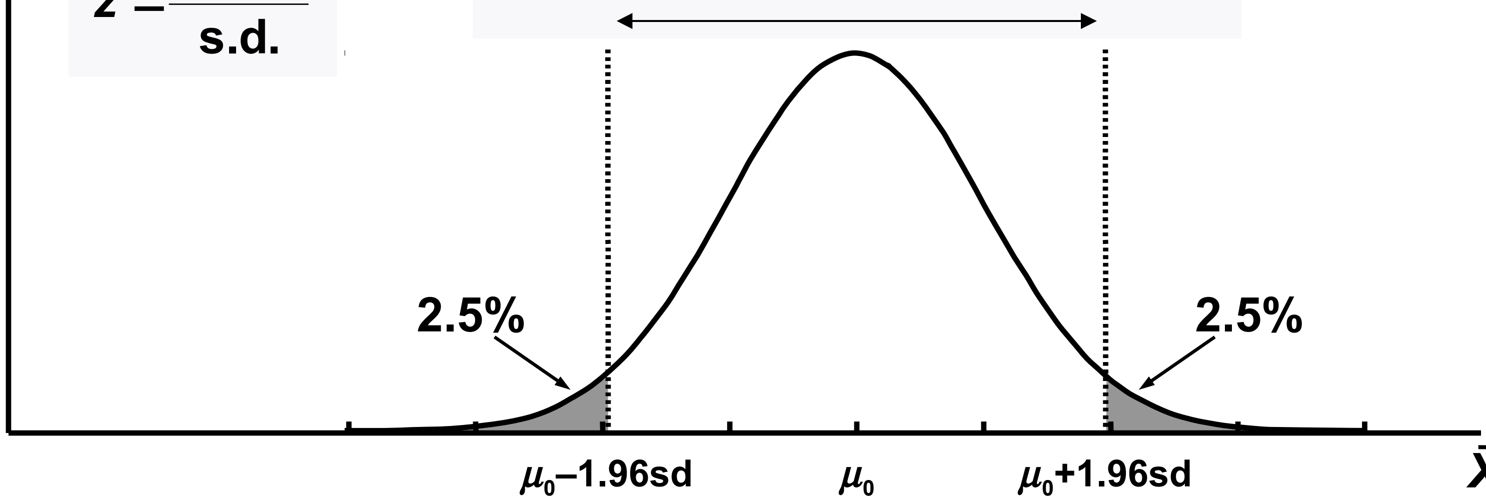
(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

(1) if $z > 1.96$ **(2) if $z < -1.96$**

probability density function of $\bar{X}$

$$z = \frac{\bar{X} - \mu_0}{\text{s.d.}}$$

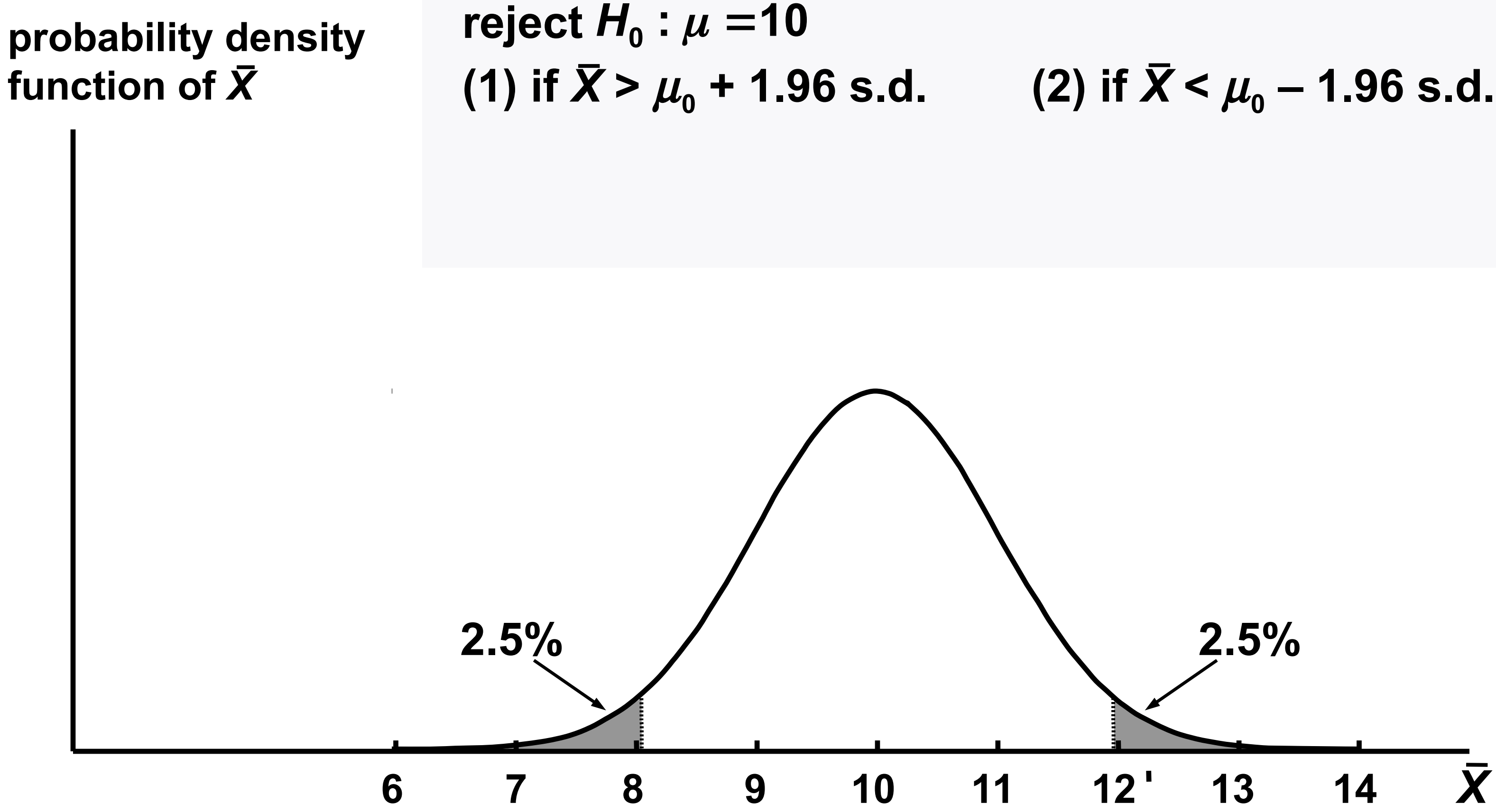
Acceptance region for $\bar{X}$:
$-1.96 \leq z \leq 1.96$



The limiting values of z for the acceptance region are 1.96 and –1.96 (for a 5% significance test).

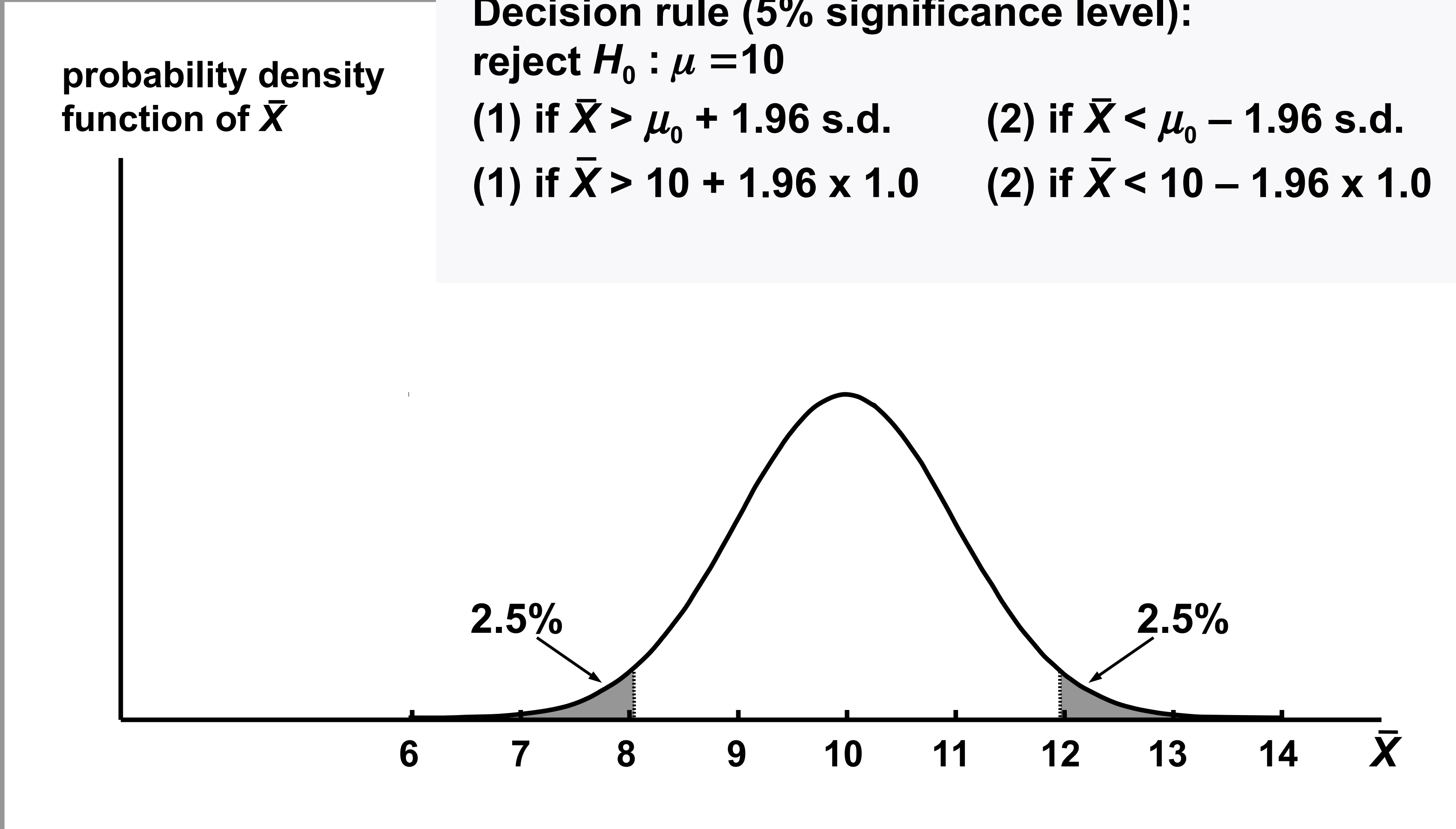
TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (5% significance level):
reject $H_0 : \mu = 10$
(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**



We will look again at the decision process in terms of the example model. The null hypothesis is that the slope coefficient is equal to 10.

Decision rule (5% significance level):
reject $H_0 : \mu = 10$
(1) if $\bar{X} > \mu_0 + 1.96$ s.d.   (2) if $\bar{X} < \mu_0 - 1.96$ s.d.
(1) if $\bar{X} > 10 + 1.96 \times 1.0$ (2) if $\bar{X} < 10 - 1.96 \times 1.0$

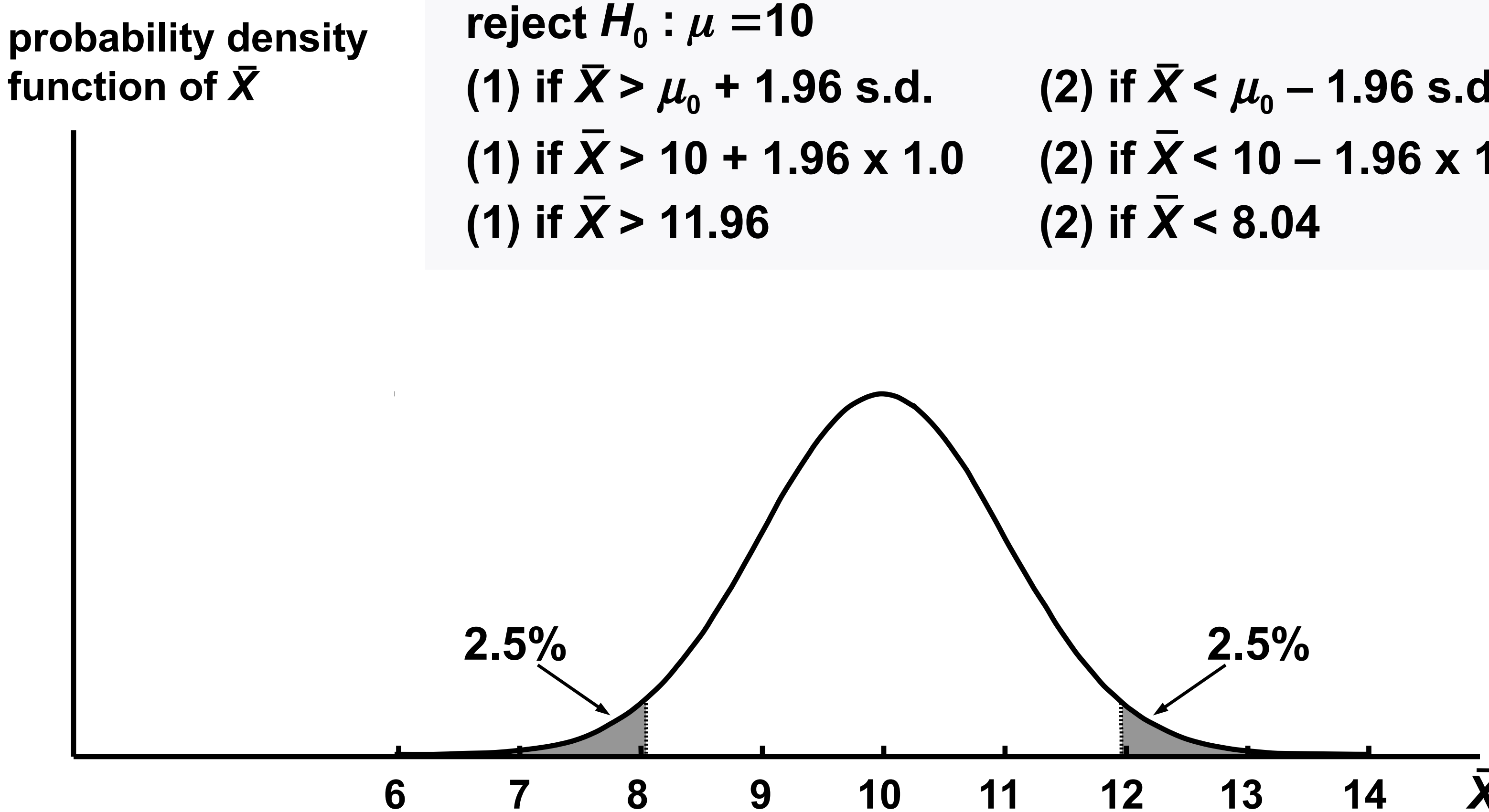


We are assuming that we know the standard deviation and that it is equal to 1.0.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (5% significance level):
reject $H_0 : \mu = 10$

(1) if $\bar{X} > \mu_0 + 1.96$ s.d.	(2) if $\bar{X} < \mu_0 - 1.96$ s.d.
(1) if $\bar{X} > 10 + 1.96 \times 1.0$	(2) if $\bar{X} < 10 - 1.96 \times 1.0$
(1) if $\bar{X} > 11.96$	(2) if $\bar{X} < 8.04$



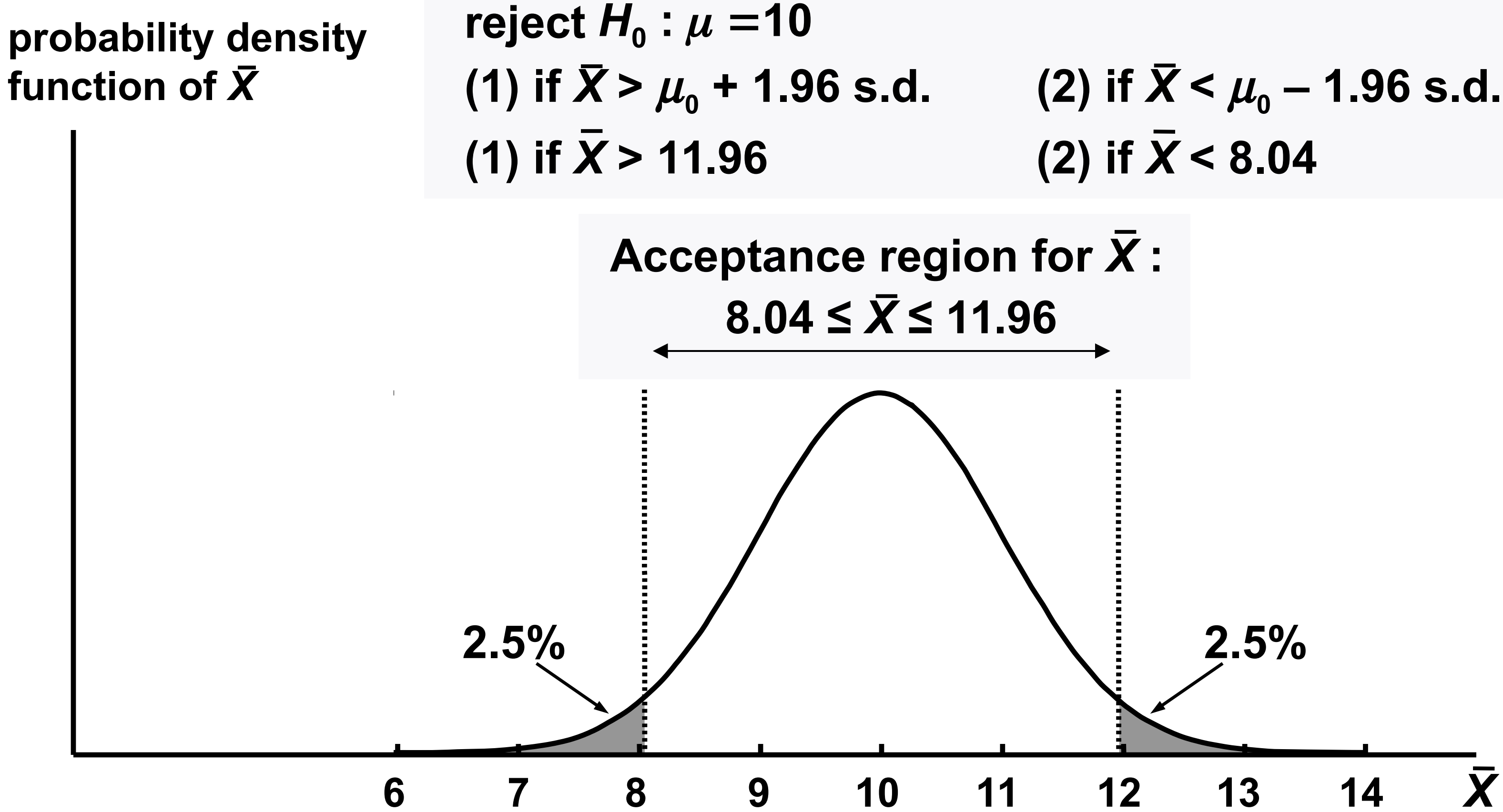
We will reject H_0 if $\bar{X} > 11.96$ or $\bar{X} < 8.04$.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (5% significance level):
reject $H_0 : \mu = 10$

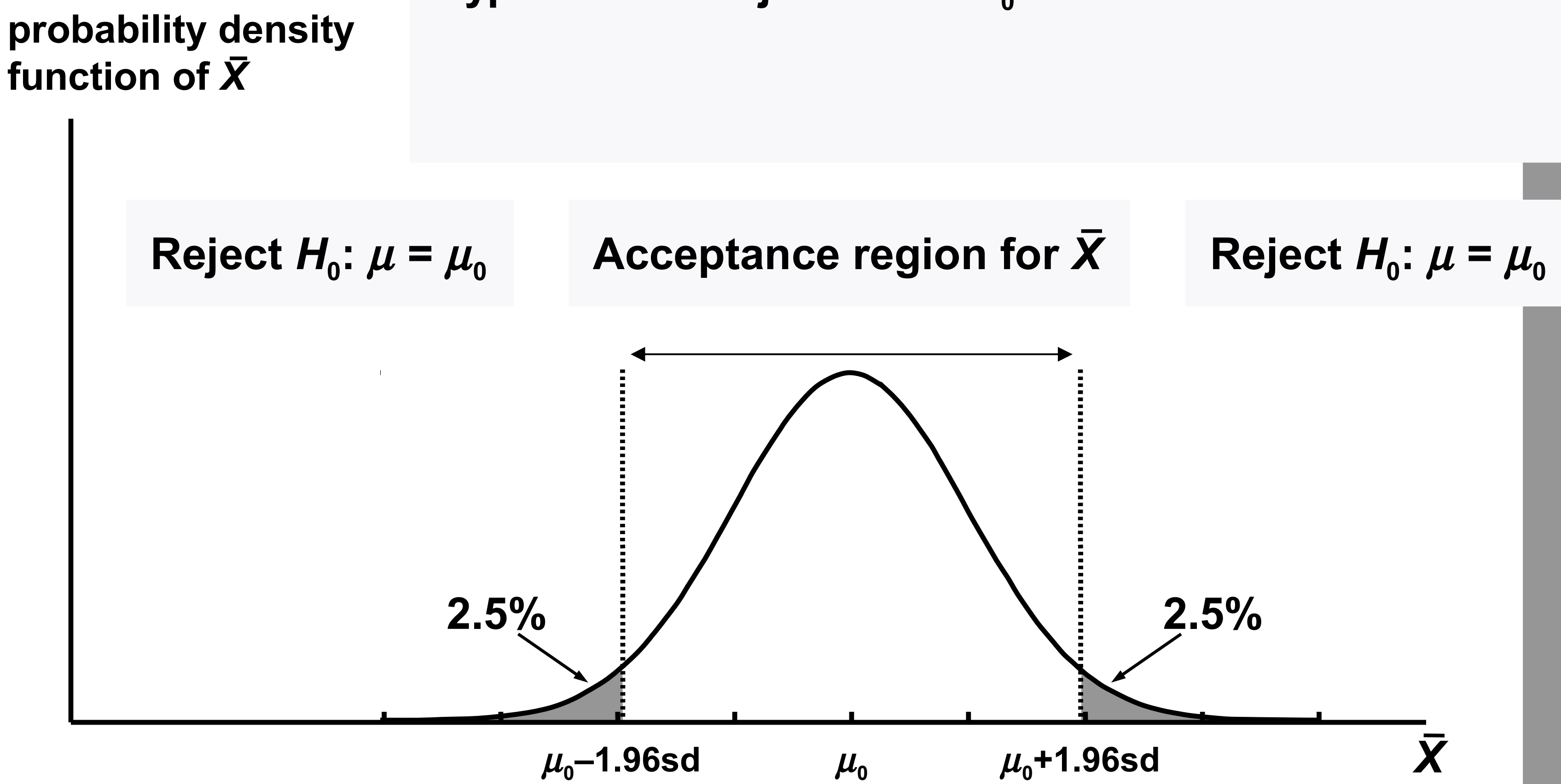
(1) if $\bar{X} > \mu_0 + 1.96$ s.d. **(2) if $\bar{X} < \mu_0 - 1.96$ s.d.**

(1) if $\bar{X} > 11.96$ **(2) if $\bar{X} < 8.04$**



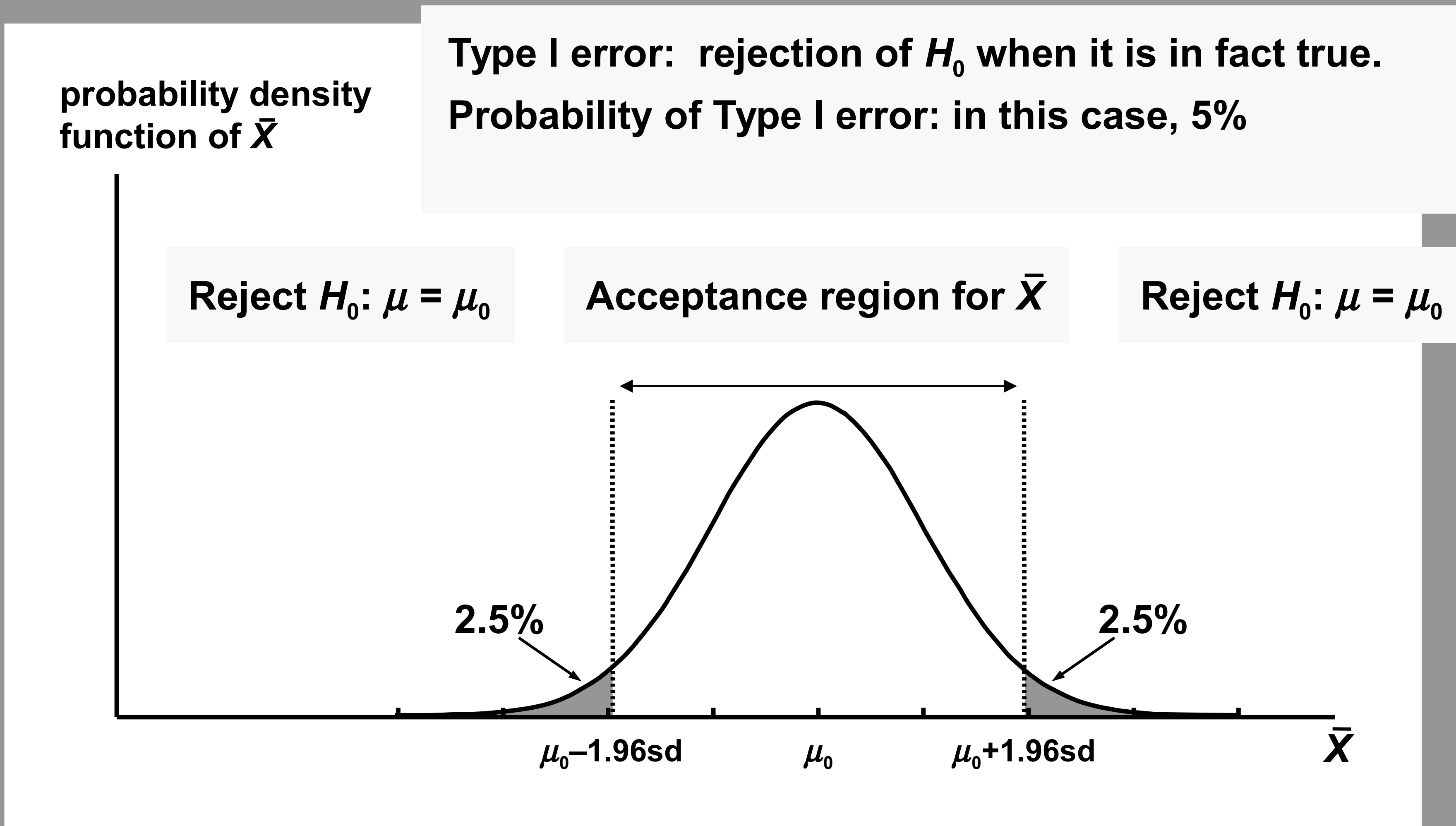
The acceptance region for $\bar{X}$ is therefore the interval 8.04 to 11.96. A sample mean in this range will not lead to the rejection of the null hypothesis.

Type I error: rejection of H_0 when it is in fact true.



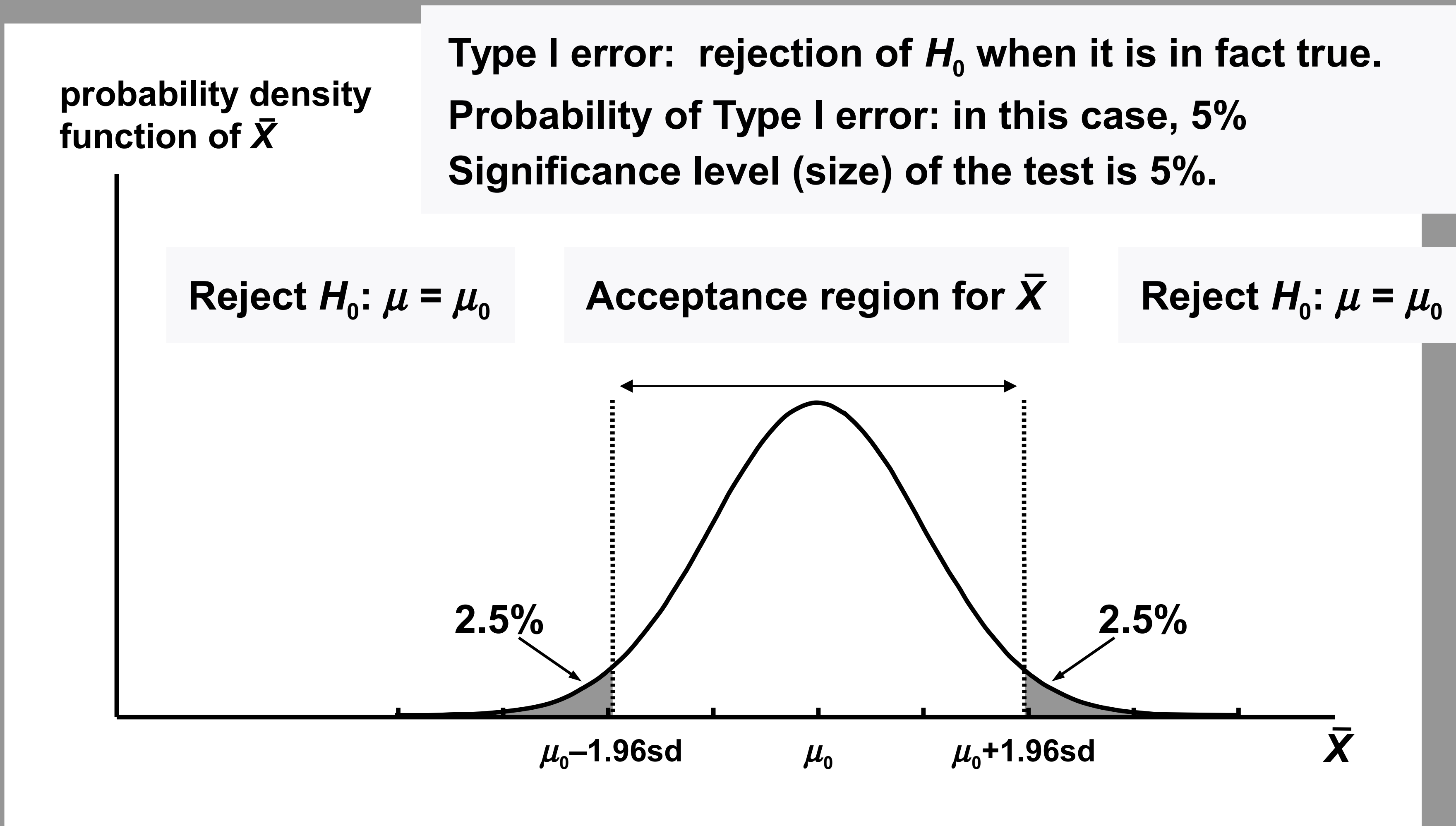
Rejection of the null hypothesis when it is in fact true is described as a Type I error.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



With the present test, if the null hypothesis is true, a Type I error will occur 5% of the time because 5% of the time we will get estimates in the upper or lower 2.5% tails.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



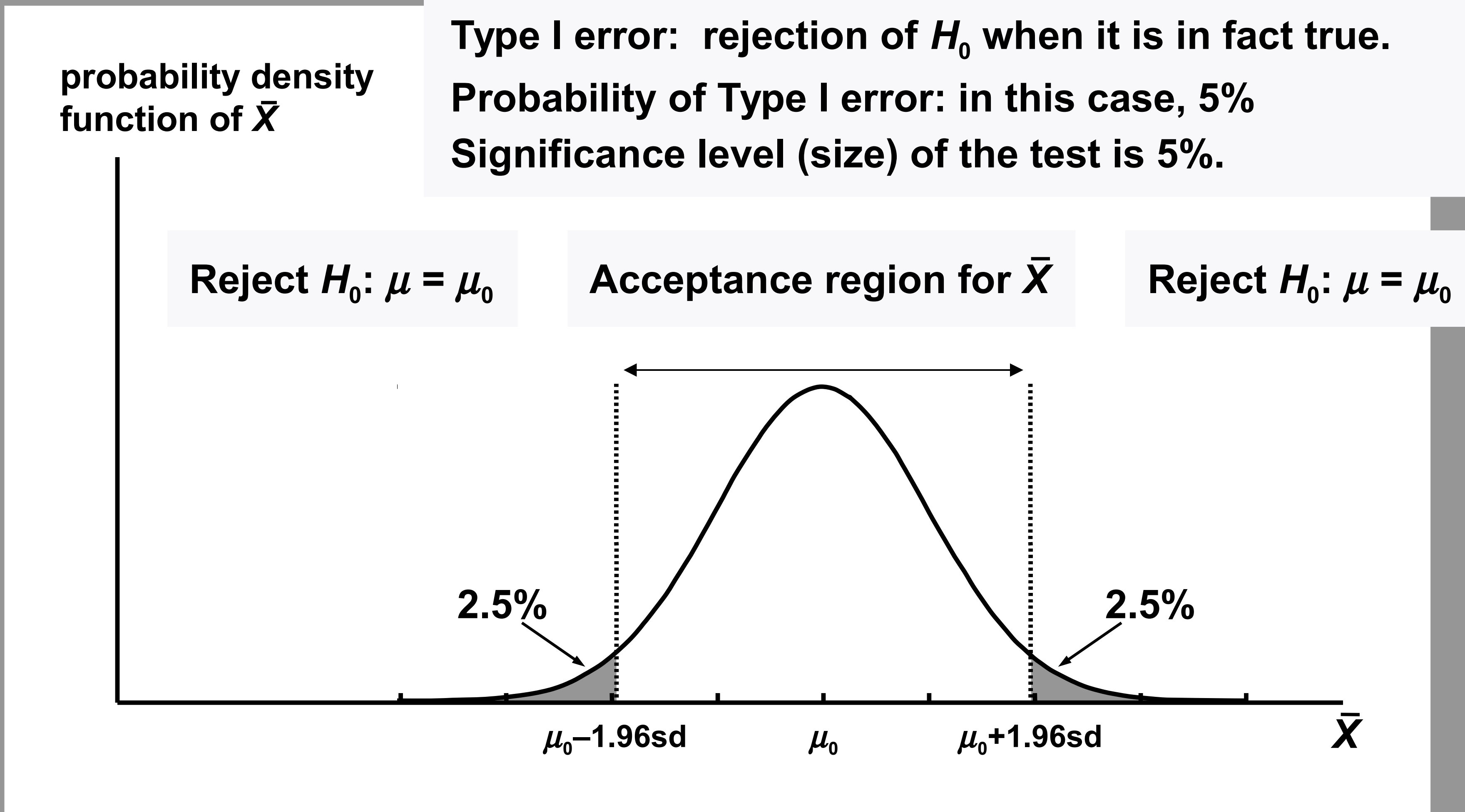
The significance level of a test (often described as the size of a test) is defined to be the probability of making a Type I error if the null hypothesis is true.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Type I error: rejection of H_0 when it is in fact true.

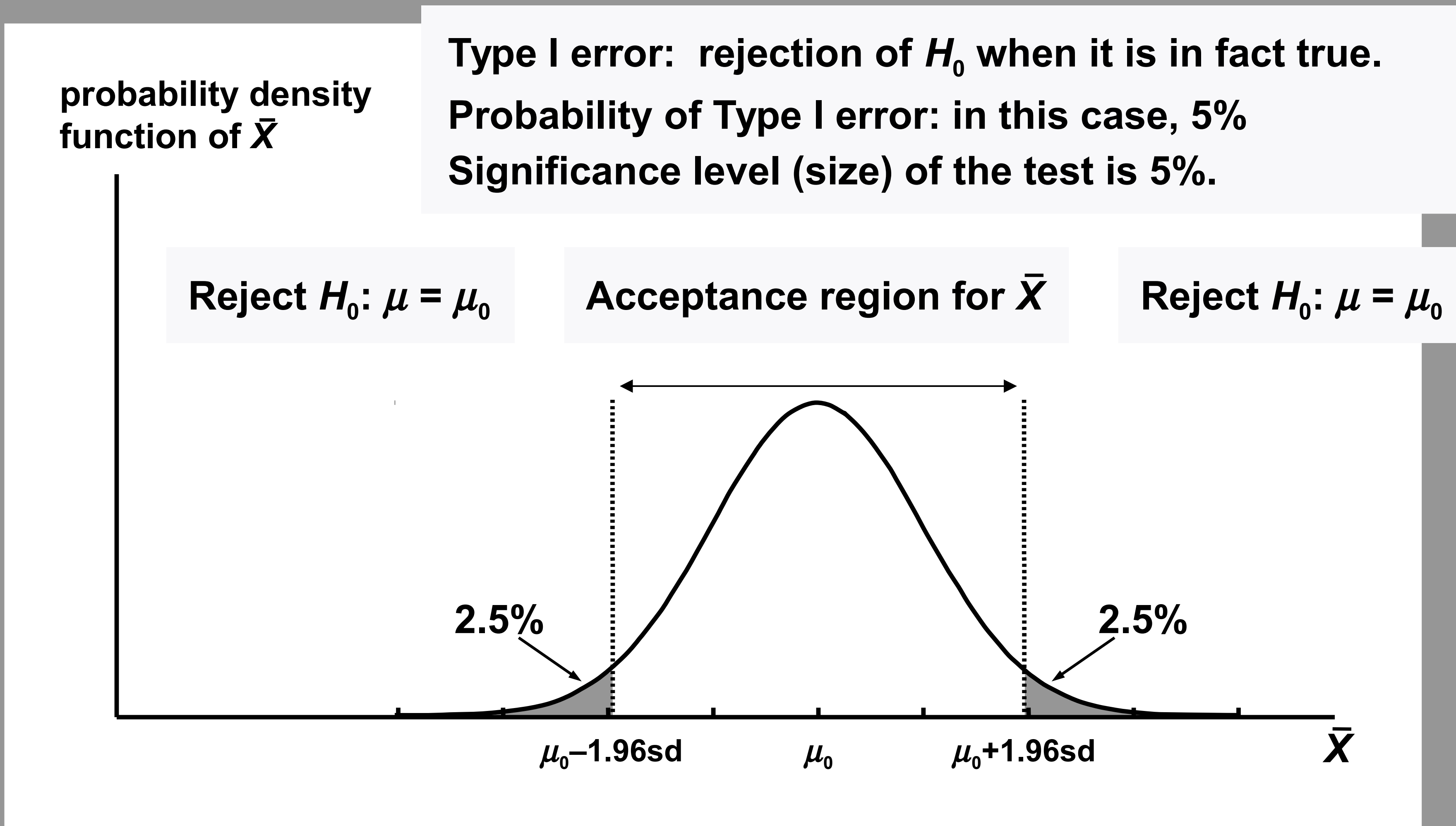
Probability of Type I error: in this case, 5%

Significance level (size) of the test is 5%.



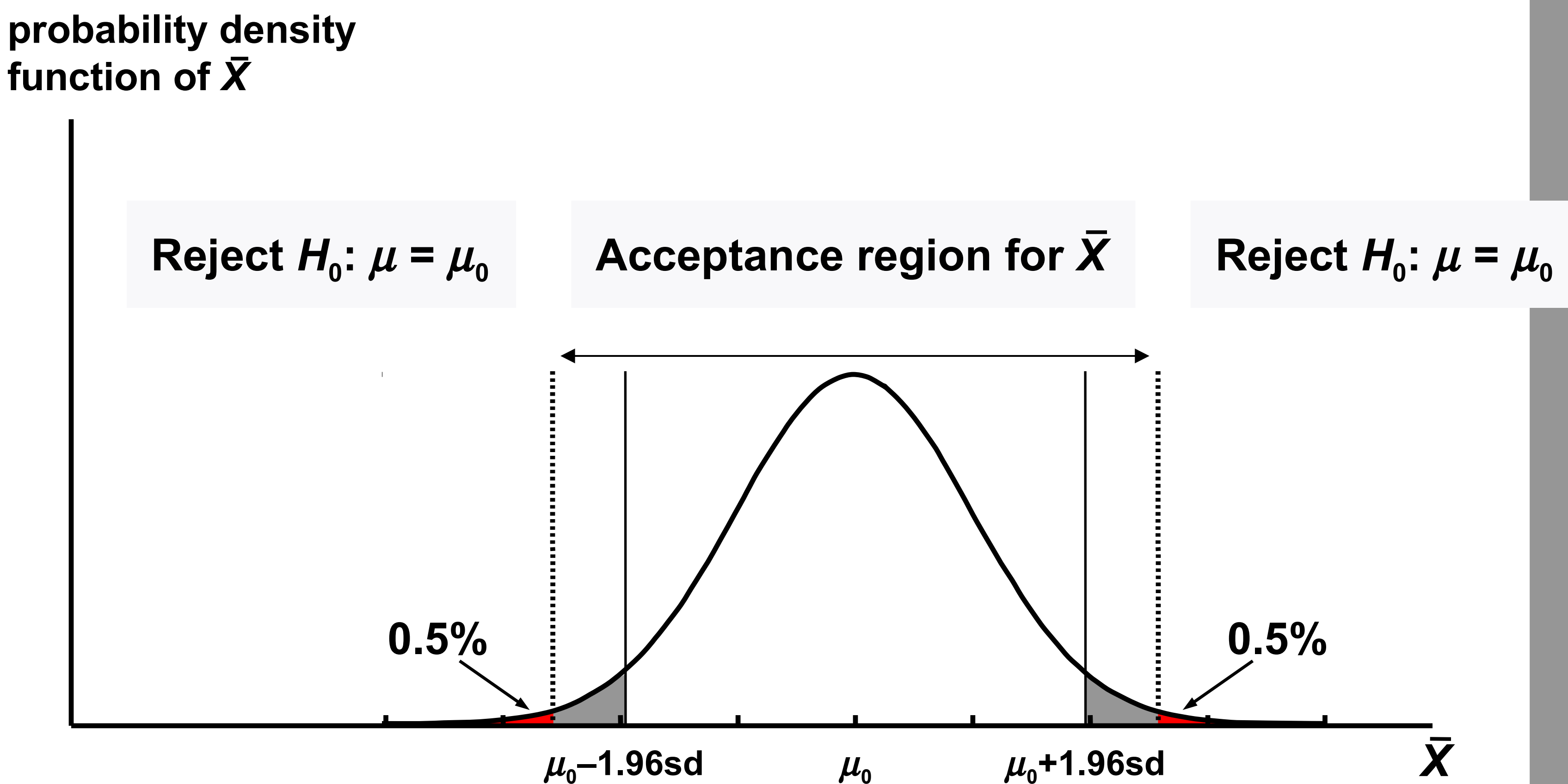
We can of course reduce the risk of making a Type I error by reducing the size of the rejection region.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



For example, we could change the decision rule to “reject the null hypothesis if it implies that the probability of getting the sample estimate is less than 0.01 (1%)”.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



The rejection region now becomes the upper and lower 0.5% tails

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (1% significance level):

reject $H_0: \mu = \mu_0$

(1) if $\bar{X} > \mu_0 + 2.58$ s.d. **(2) if $\bar{X} < \mu_0 - 2.58$ s.d.**

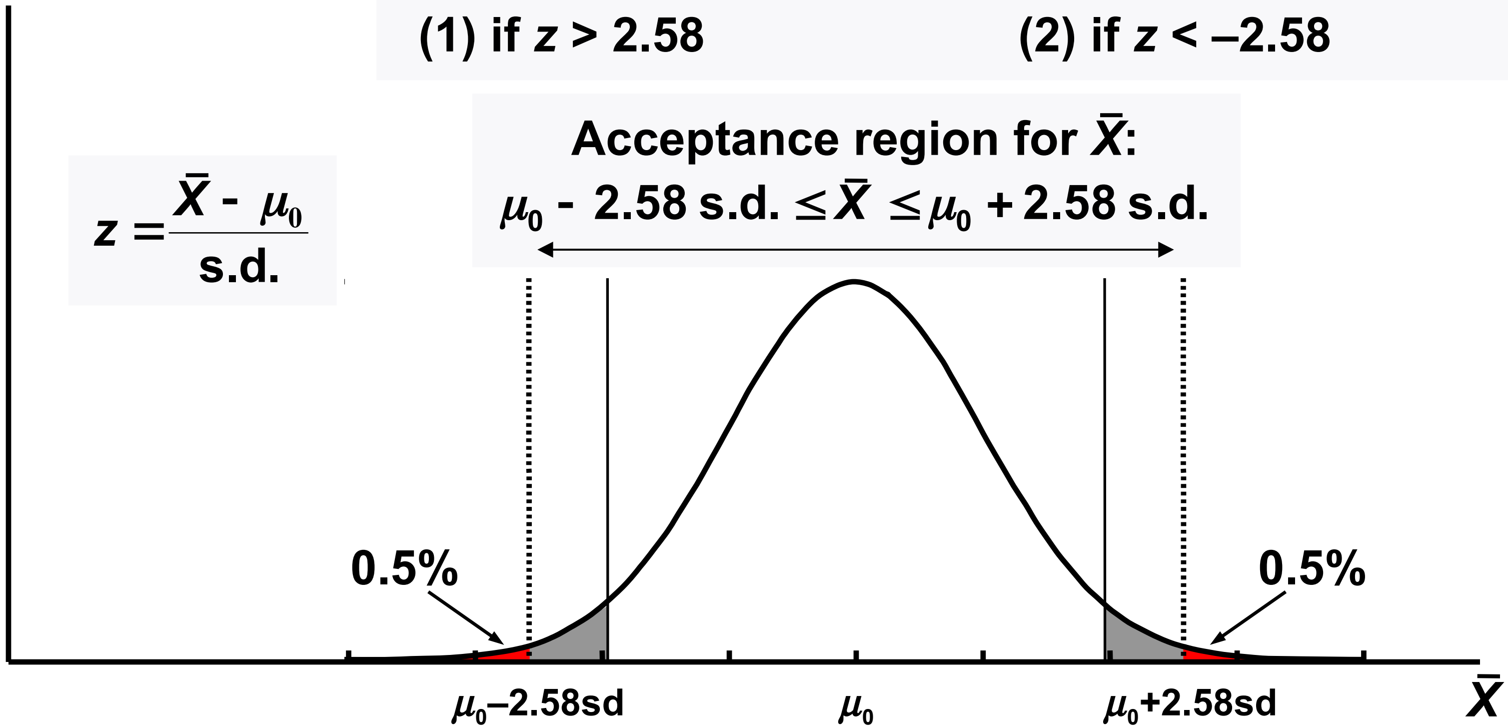
(1) if $z > 2.58$ **(2) if $z < -2.58$**

probability density function of $\bar{X}$

$$z = \frac{\bar{X} - \mu_0}{\text{s.d.}}$$

Acceptance region for $\bar{X}$:

$$\mu_0 - 2.58 \text{ s.d.} \leq \bar{X} \leq \mu_0 + 2.58 \text{ s.d.}$$



The 0.5% tails of a normal distribution start 2.58 standard deviations from the mean, so the acceptance region ranges from 2.58 standard deviations below $\bar{X}$ to 2.58 standard deviations above it.

Decision rule (1% significance level):

reject $H_0: \mu = \mu_0$

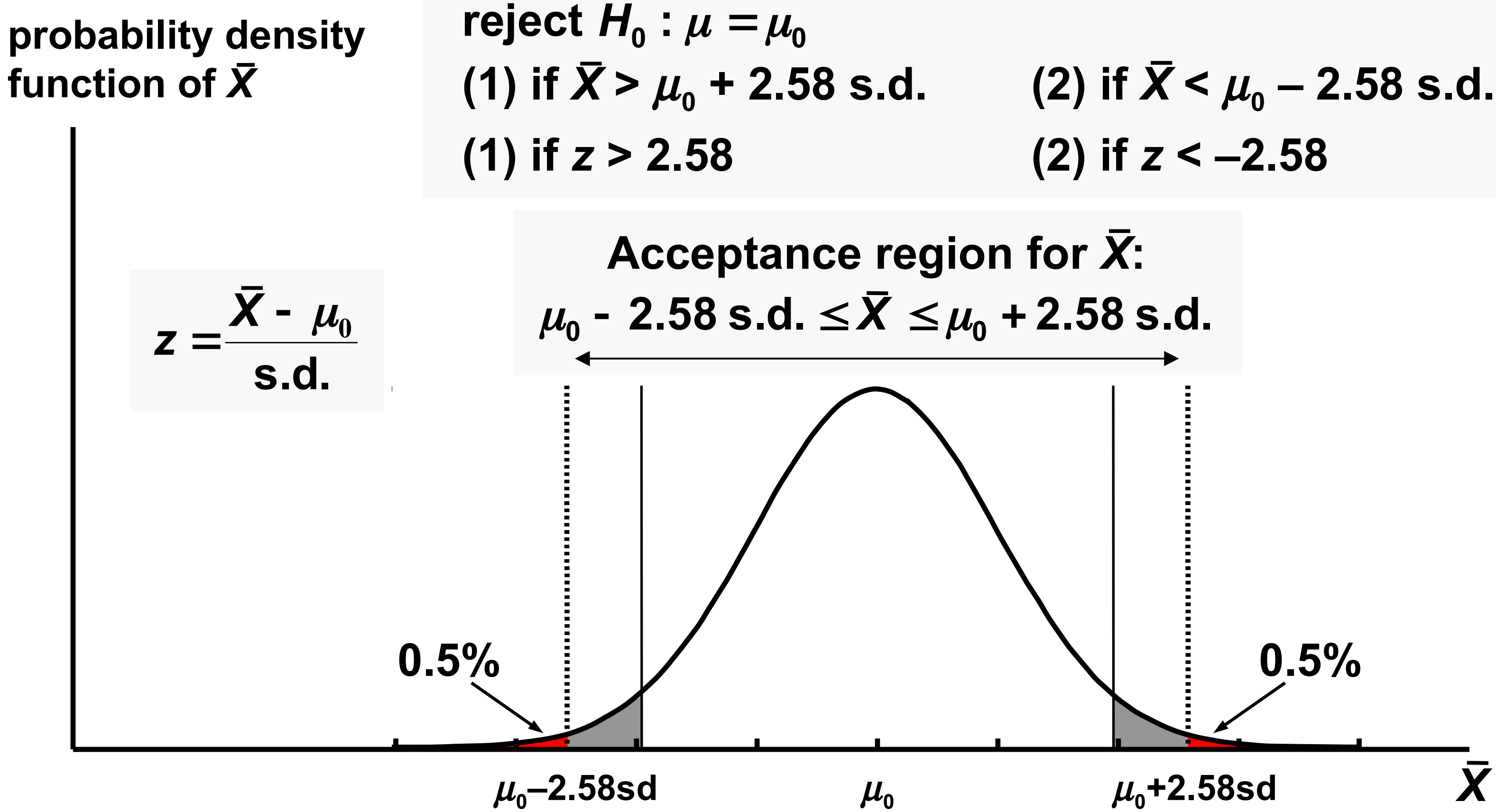
(1) if $\bar{X} > \mu_0 + 2.58$ s.d. **(2) if $\bar{X} < \mu_0 - 2.58$ s.d.**

(1) if $z > 2.58$ **(2) if $z < -2.58$**

Acceptance region for $\bar{X}$:

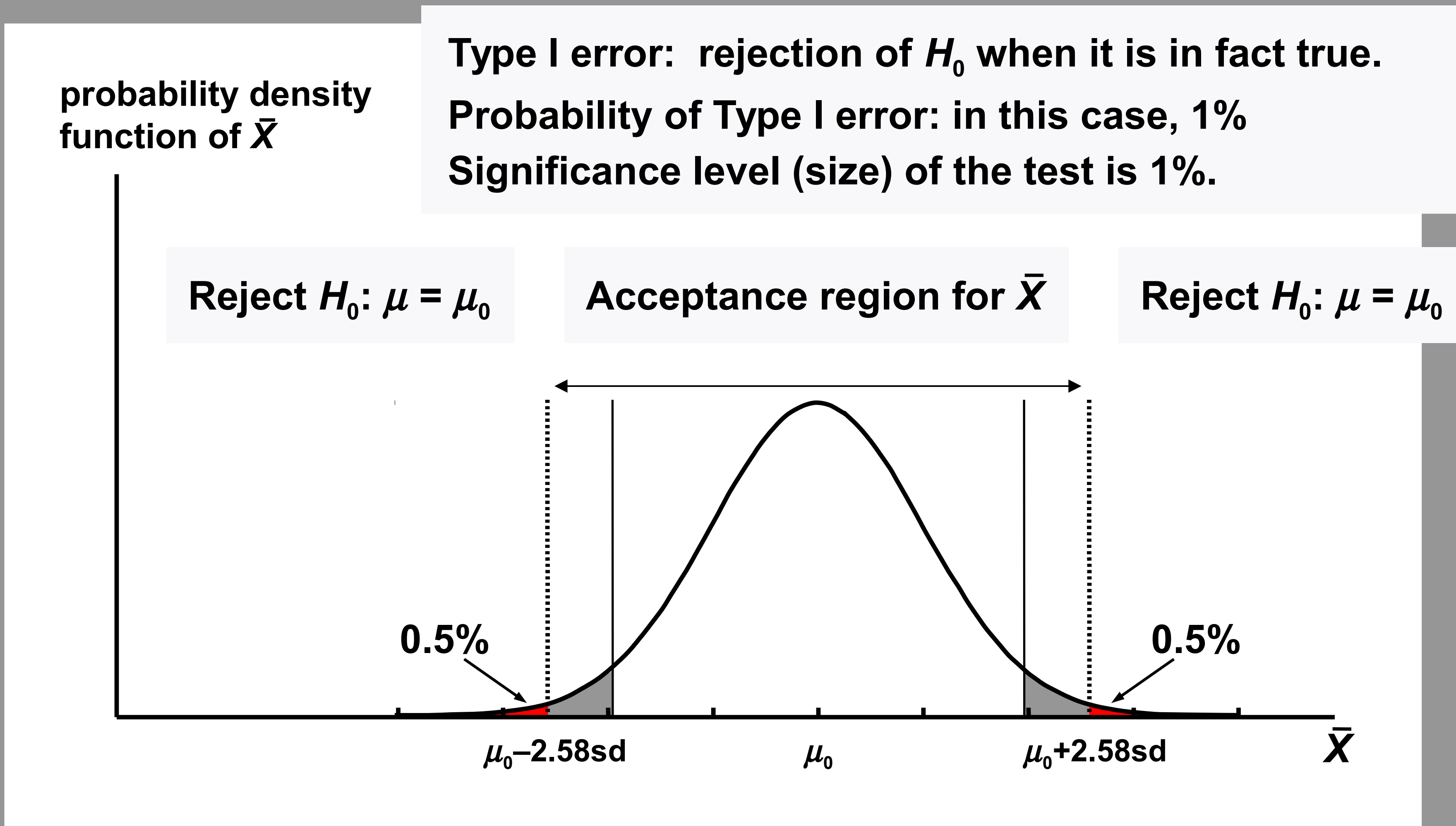
$$\mu_0 - 2.58 \text{ s.d.} \leq \bar{X} \leq \mu_0 + 2.58 \text{ s.d.}$$

$$z = \frac{\bar{X} - \mu_0}{\text{s.d.}}$$



Equivalently, we now reject the null hypothesis if *z* is greater than 2.58, in absolute terms.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN



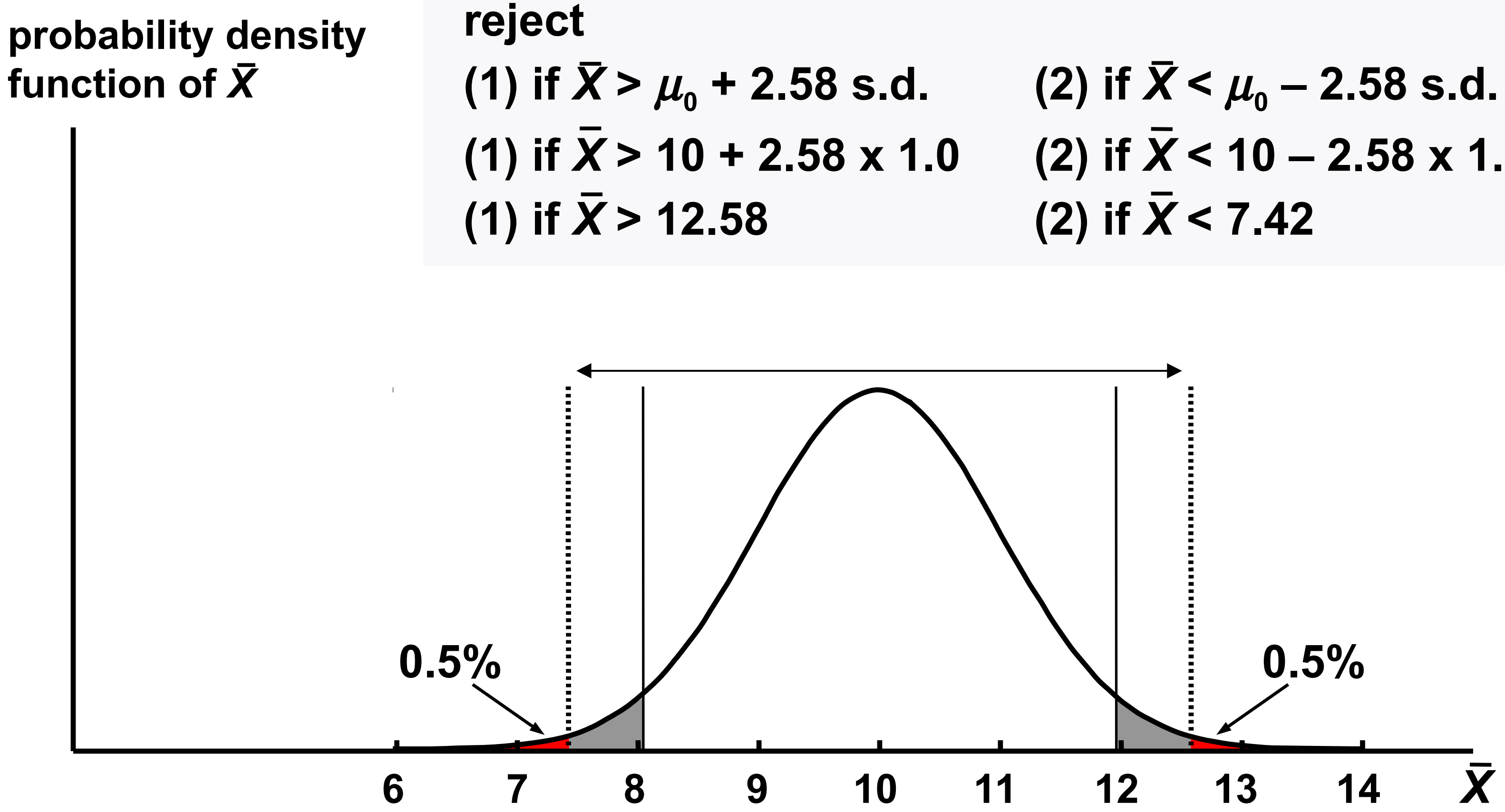
Since the probability of making a Type I error, if the null hypothesis is true, is now only 1%, the test is said to be a 1% significance test.

Decision rule (1% significance level):
reject

(1) if $\bar{X} > \mu_0 + 2.58$ s.d. (2) if $\bar{X} < \mu_0 - 2.58$ s.d.

(1) if $\bar{X} > 10 + 2.58 \times 1.0$ (2) if $\bar{X} < 10 - 2.58 \times 1.0$

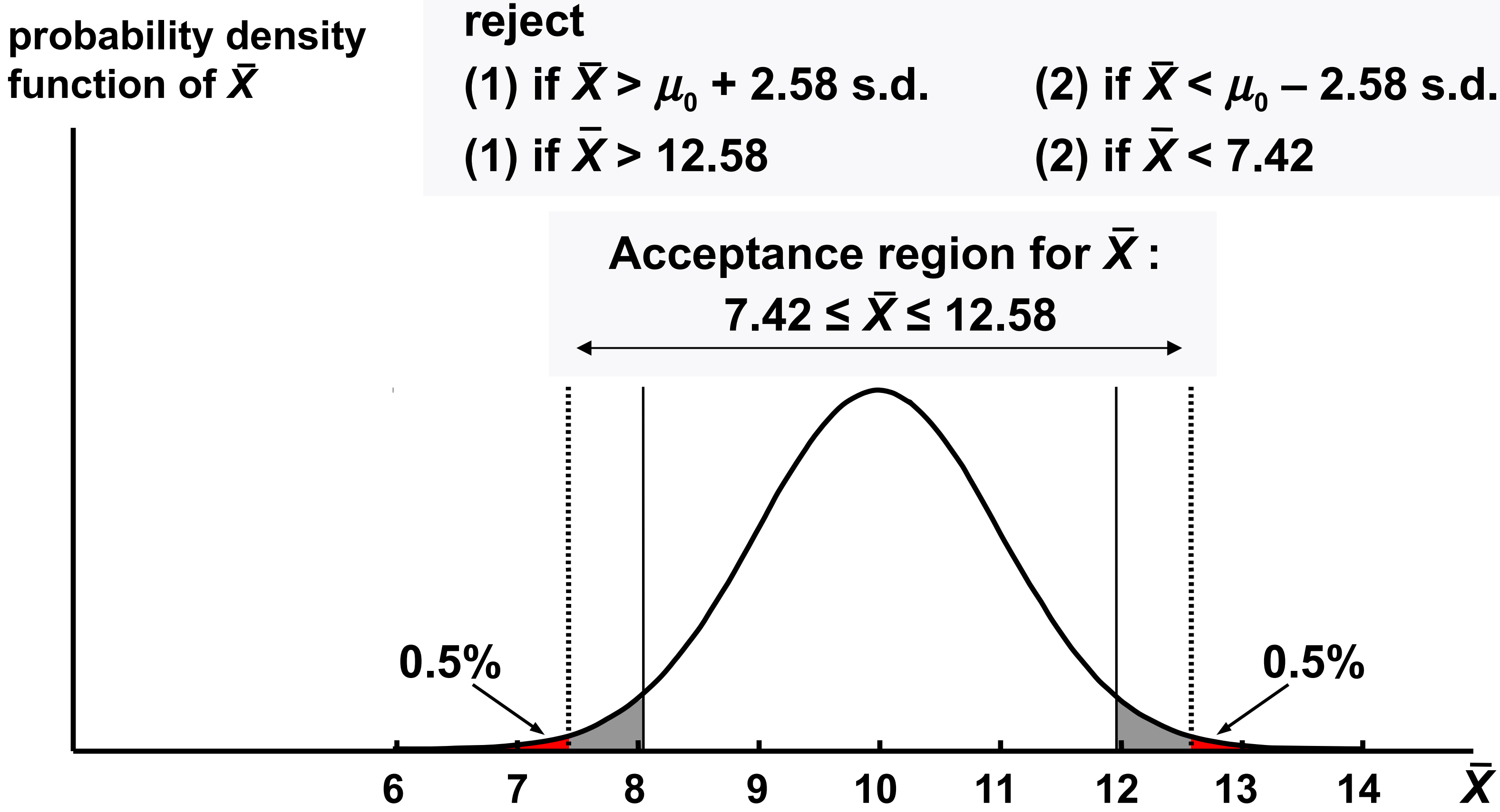
(1) if $\bar{X} > 12.58$ (2) if $\bar{X} < 7.42$



In the case of the example model, given that the standard deviation is 1.0, the 0.5% tails start 2.58 above and below the mean, that is, at 7.42 and 12.58.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

Decision rule (1% significance level):
reject
(1) if $\bar{X} > \mu_0 + 2.58$ s.d. **(2) if $\bar{X} < \mu_0 - 2.58$ s.d.**
(1) if $\bar{X} > 12.58$ **(2) if $\bar{X} < 7.42$**



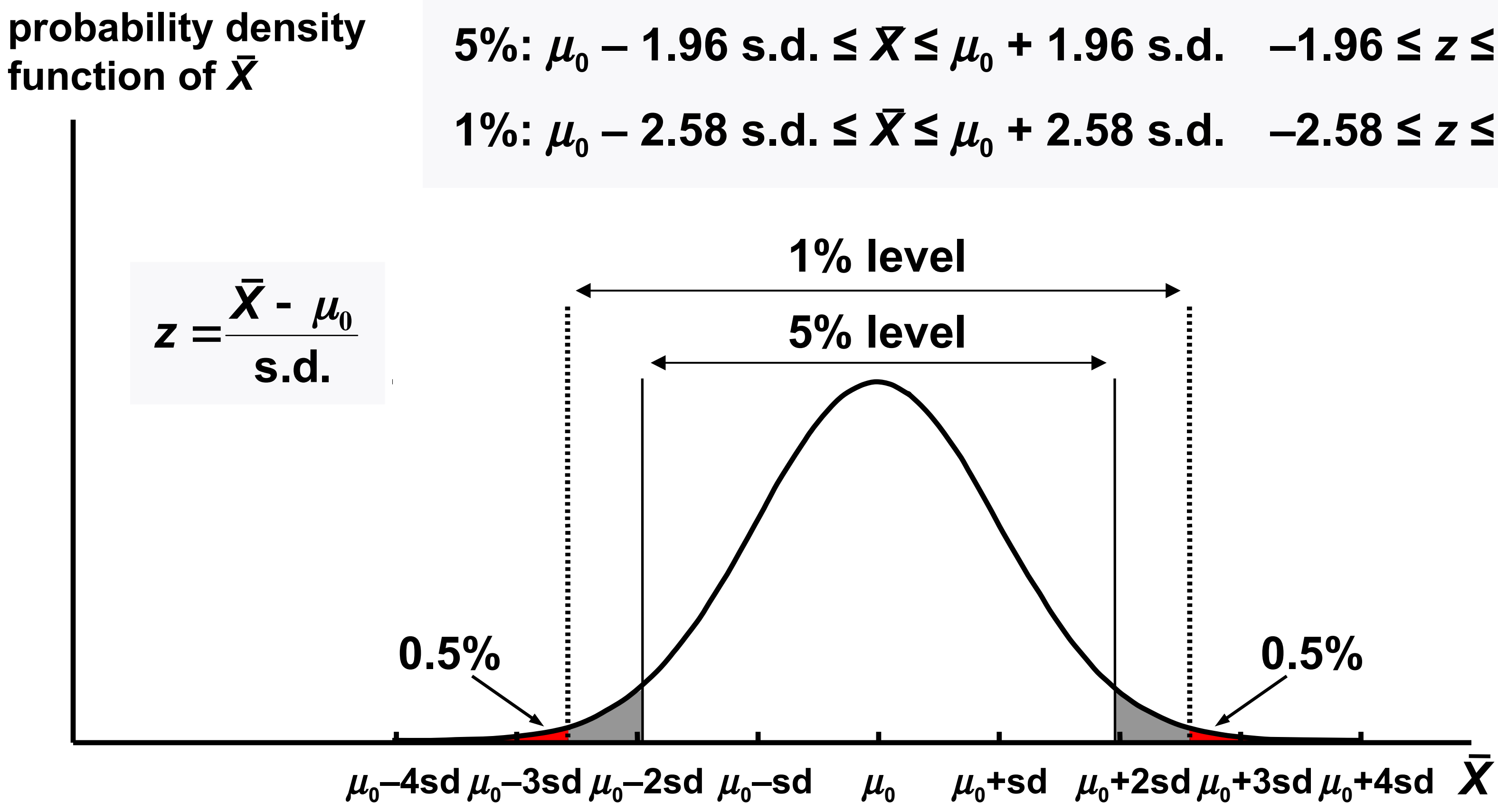
The acceptance region for $\bar{X}$ is therefore the interval 7.42 to 12.58. Because it is wider than that for the 5% test, there is less risk of making a Type I error, if the null hypothesis is true.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

5% and 1% acceptance regions compared

5%: $\mu_0 - 1.96 \text{ s.d.} \leq \bar{X} \leq \mu_0 + 1.96 \text{ s.d.} \quad -1.96 \leq z \leq 1.96$

1%: $\mu_0 - 2.58 \text{ s.d.} \leq \bar{X} \leq \mu_0 + 2.58 \text{ s.d.} \quad -2.58 \leq z \leq 2.58$



This diagram compares the decision-making processes for the 5% and 1% tests. Note that if you reject H_0 at the 1% level, you *must* also reject it at the 5% level.

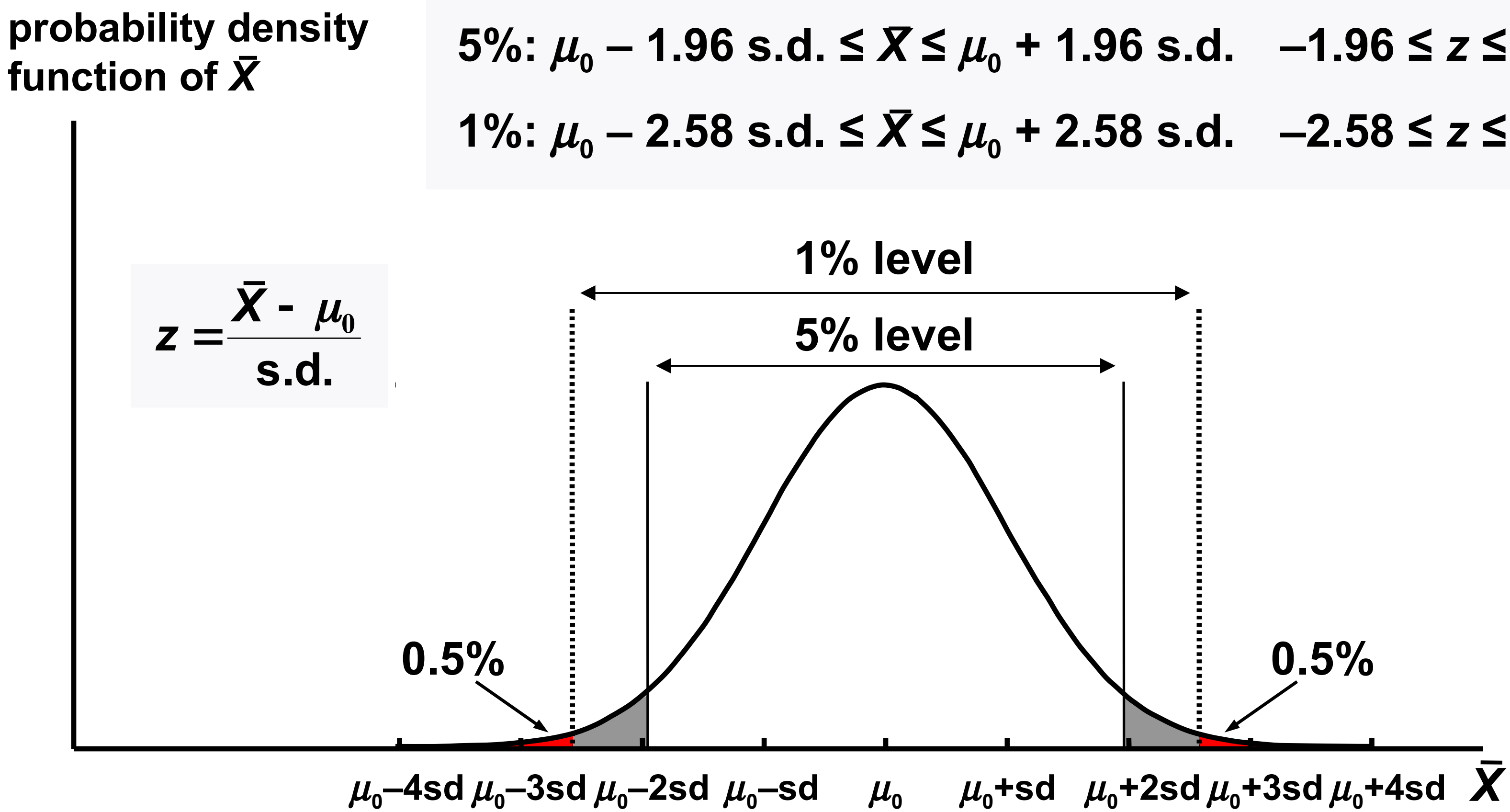
TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

5% and 1% acceptance regions compared

5%: $\mu_0 - 1.96$ s.d. $\leq \bar{X} \leq \mu_0 + 1.96$ s.d. $\quad -1.96 \leq z \leq 1.96$

1%: $\mu_0 - 2.58$ s.d. $\leq \bar{X} \leq \mu_0 + 2.58$ s.d. $\quad -2.58 \leq z \leq 2.58$

$$z = \frac{\bar{X} - \mu_0}{\text{s.d.}}$$



Note also that if $\bar{X}$ lies within the acceptance region for the 5% test, it *must* also fall within it for the 1% test.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

General case	Decision	Example model
	Reject H_0 at 1% level (and also 5% level)	
μ_0 + 2.58 s.d.		12.58
	Reject H_0 at 5% level but not 1% level	
μ_0 + 1.96 s.d.		11.96
μ_0	Do not reject H_0 at 5% level (or at 1% level)	10.00
μ_0 - 1.96 s.d.		8.04
	Reject H_0 at 5% level but not 1% level	
μ_0 - 2.58 s.d.		7.42
	Reject H_0 at 1% level (and also 5% level)	

The diagram summarizes the possible decisions for the 5% and 1% tests, for both the general case and the example model.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

General case	Decision	Example model
	Reject H_0 at 1% level (and also 5% level)	
μ_0 + 2.58 s.d.		12.58
	Reject H_0 at 5% level but not 1% level	
μ_0 + 1.96 s.d.		11.96
μ_0	Do not reject H_0 at 5% level (or at 1% level)	10.00
μ_0 - 1.96 s.d.		8.04
	Reject H_0 at 5% level but not 1% level	
μ_0 - 2.58 s.d.		7.42
	Reject H_0 at 1% level (and also 5% level)	

The middle of the diagram indicates what you would report. You would not report the phrases in parentheses.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

General case	Decision	Example model
	Reject H_0 at 1% level (and also 5% level)	
μ_0 + 2.58 s.d.		12.58
	Reject H_0 at 5% level but not 1% level	
μ_0 + 1.96 s.d.		11.96
μ_0	Do not reject H_0 at 5% level (or at 1% level)	10.00
μ_0 - 1.96 s.d.		8.04
	Reject H_0 at 5% level but not 1% level	
μ_0 - 2.58 s.d.		7.42
	Reject H_0 at 1% level (and also 5% level)	

If you can reject H_0 at the 1% level, it automatically follows that you can reject it at the 5% level and there is no need to say so. Indeed, you would look ignorant if you did.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

General case	Decision	Example model
	Reject H_0 at 1% level (and also 5% level)	
μ_0 + 2.58 s.d.		12.58
	Reject H_0 at 5% level but not 1% level	
μ_0 + 1.96 s.d.		11.96
μ_0	Do not reject H_0 at 5% level (or at 1% level)	10.00
μ_0 - 1.96 s.d.		8.04
	Reject H_0 at 5% level but not 1% level	
μ_0 - 2.58 s.d.		7.42
	Reject H_0 at 1% level (and also 5% level)	

Likewise, if you cannot reject H_0 at the 5% level, that is all you should say. It automatically follows that you cannot reject it at the 1% level and you would look ignorant if you said so.

TESTING A HYPOTHESIS RELATING TO THE POPULATION MEAN

General case	Decision	Example model
	Reject H_0 at 1% level (and also 5% level)	
μ_0 + 2.58 s.d.		12.58
	Reject H_0 at 5% level but not 1% level	
μ_0 + 1.96 s.d.		11.96
μ_0	Do not reject H_0 at 5% level (or at 1% level)	10.00
μ_0 - 1.96 s.d.		8.04
	Reject H_0 at 5% level but not 1% level	
μ_0 - 2.58 s.d.		7.42
	Reject H_0 at 1% level (and also 5% level)	

You should report the results of both tests only when you can reject H_0 at the 5% level but not at the 1% level.

The content of this slideshow comes from Section R.9 of C. Dougherty, *Introduction to Econometrics*, fifth edition 2016, Oxford University Press. Additional (free) resources for both students and instructors may be downloaded from the OUP Online Resource Centre
http://www.oxfordtextbooks.co.uk/orc/dougherty5e/

Individuals studying econometrics on their own who feel that they might benefit from participation in a formal course should consider the London School of Economics summer school course
EC212 Introduction to Econometrics
http://www2.lse.ac.uk/study/summerSchools/summerSchool/Home.aspx
or the University of London International Programmes distance learning course
EC2020 Elements of Econometrics
www.londoninternational.ac.uk/lse.