Errors in Hypothesis Testing

Dr. A. Ramesh
Department of Management Studies
Indian Institute of Technology Roorkee

Example

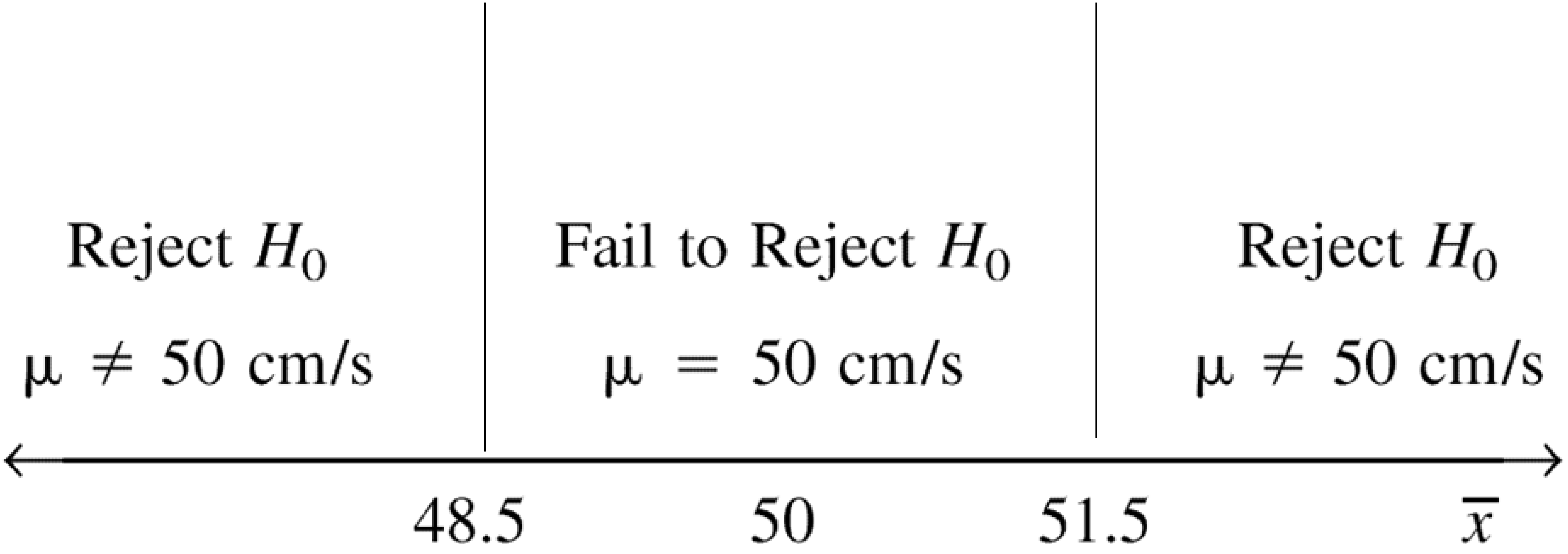
- We are interested in burning rate of a solid propellant used to power aircrew escape systems
- Burning rate is a random variable that can be described by a probability distribution
- Suppose our interest focus on mean burning rate
- Ho: $\mu = 50$ centimeters per second
- H1: $\mu \neq 50$ centimeters per second

Reference: Applied statistics and probability for engineers, Douglas C. Montgomery, George C. Runger, John Wiley & Sons, 2007

Value of the null hypothesis

- The value of the null hypothesis can be obtained by
  - Past experience or knowledge of the process, or even from the previous tests or experiments
  - From some theory or model regarding the process under study
  - From external consideration, such as design or engineering specifications, or from contractual obligations

Note: for this example n=10



Decision criteria for testing H_0: $\mu = 50$ cm/s versus H_1: $\mu \neq 50$ cm/s.

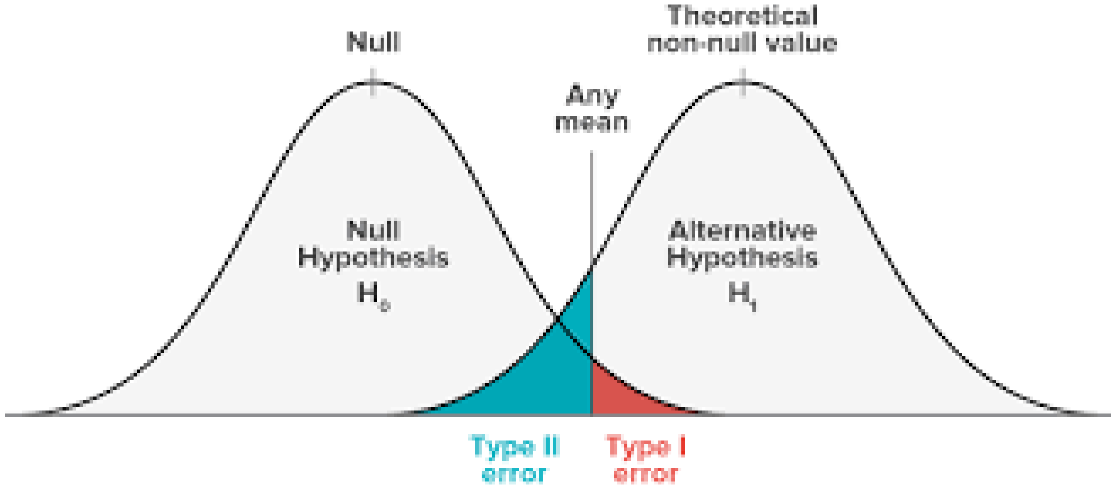
Note: for this example we will assume $\sigma = 2.5$

Type I Error

- The true mean burning rate of the propellant could be equal to 50 centimeters per second
- However randomly selected propellant specimens that are tested, we could observe a value of test statistics $\bar{x}$ that falls into the critical region(rejection region).
- We would then reject the null hypothesis Ho in favor of the alternate H1, in fact, Ho is really true
- This type of wrong conclusion is called a type I error

Type I Error

- Rejecting the null hypothesis Ho when it is true is defined as a type I error

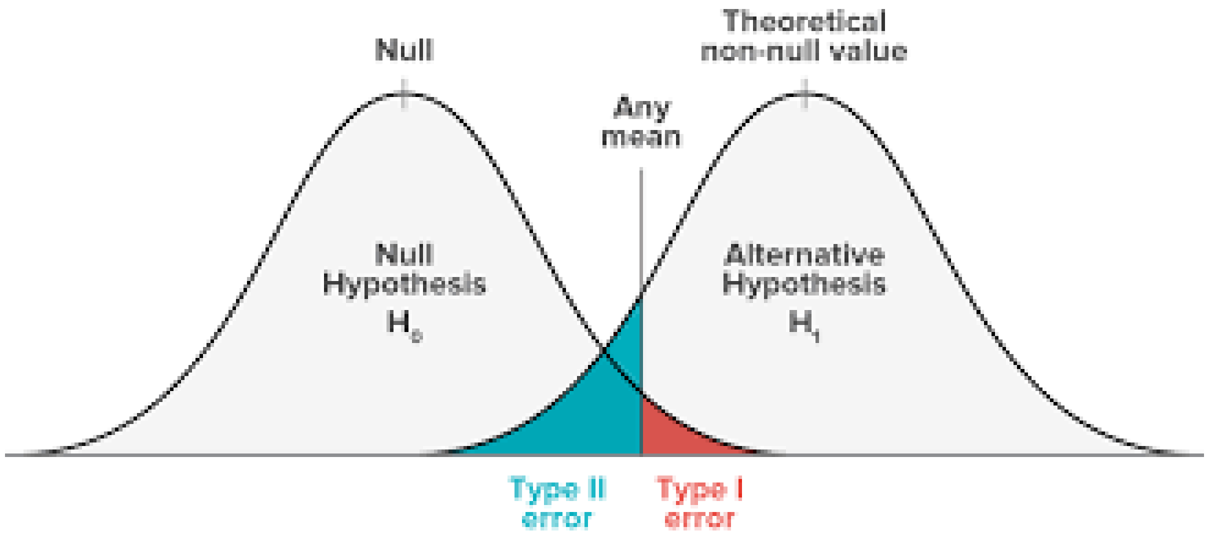


Type II Error

- Now suppose the true mean burning rate is different from 50 centimeters per second, yet the sample mean $\bar{x}$ falls in the acceptance region
- In this case we would fail to reject Ho when it is false
- This type of wrong conclusion is called a type II error

Type II Error

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



Type 1 and Type II Errors

	H_0 is correct	H_0 is incorrect
H_0 is accepted	correct decision	Type II error (β) **Incorrect acceptance**
H_0 is rejected	Type I error (α) **Incorrect rejection**	correct decision

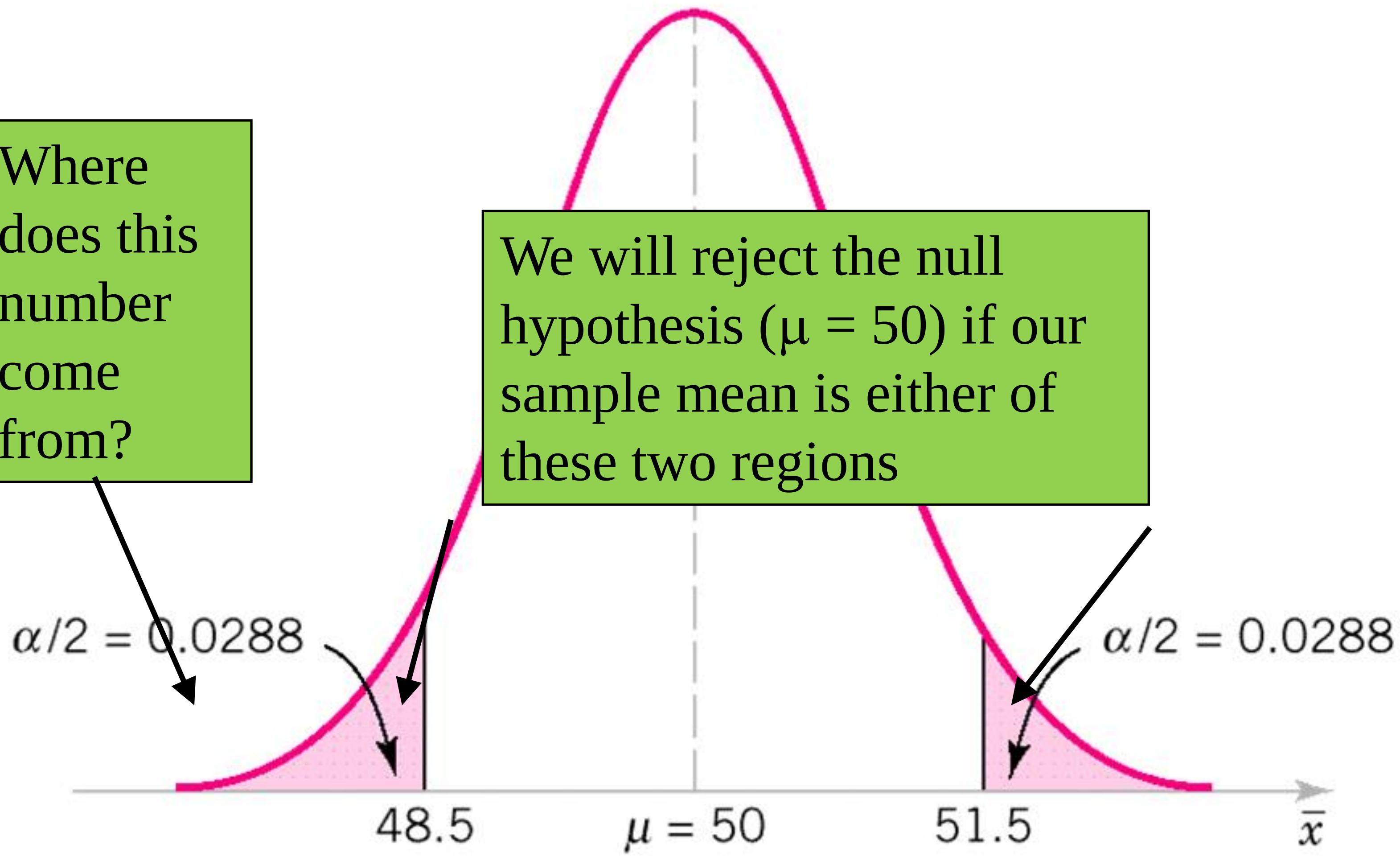
Type I error

- In the propellant burning rate example, a type I error will occur when either $\bar{x} > 51.5\_or\_\bar{x} < 48.5$ when the true mean burning rate is μ = 50 centimeters per second
- Suppose the standard deviation of burning rate is σ = 2.5 centimeters per second and n = 10
- Probability distribution μ = 50,standard error = 0.79.
- Type I error is

$$\alpha = P(\bar{x} < 48.5\_when\_\mu = 50) + P(\bar{x} > 51.5\_when\_\mu = 50)$$

Where does this number come from?

We will reject the null hypothesis ($\mu = 50$) if our sample mean is either of these two regions



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

Defing function for calculating alpha value

In [6]:
```
def z_value(x,mu,SEM):
 z = (x - mu)/SEM
 if(z < 0):
  alfa = stats.norm.cdf(z)
 else:
  alfa = 1 - stats.norm.cdf(z)
 print (alfa)
```

calculating aplha for different values of x,mu, and SEM

In [8]:
```
x =48.5
mu = 50
SEM = 0.79
```

In [9]:
```
z_value(x,mu,SEM)
```

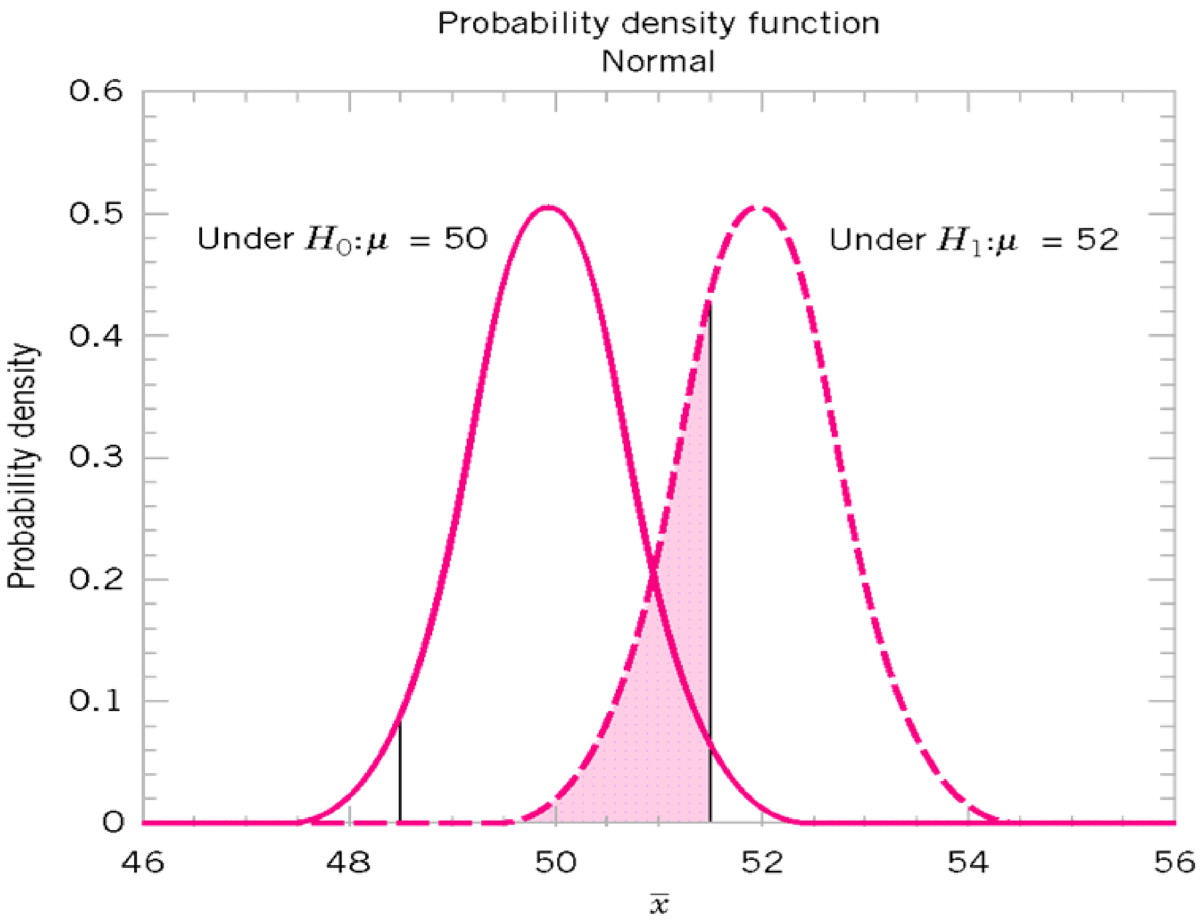
```
0.02879971774715278
```

Type I error

- Type I error = 0.057434
- This implies that 5.7 % of all random samples would lead to rejection of the hypothesis Ho: μ=50 centimeters per second.
- We can reduce the type I error by widening the acceptance region. If we make critical value 48 and 52, the value of alpha is 0.0114 (adding 0.0057 and 0.0057).
- Change sample size to 16 then alpha is 0.0164.

```
In [40]: z_value(48,mu,SEM)
0.005676434117424844

In [41]: z_value(52,mu,SEM)
0.0056764341174248
```

TYPE II ERROR



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

The pink area is the probability

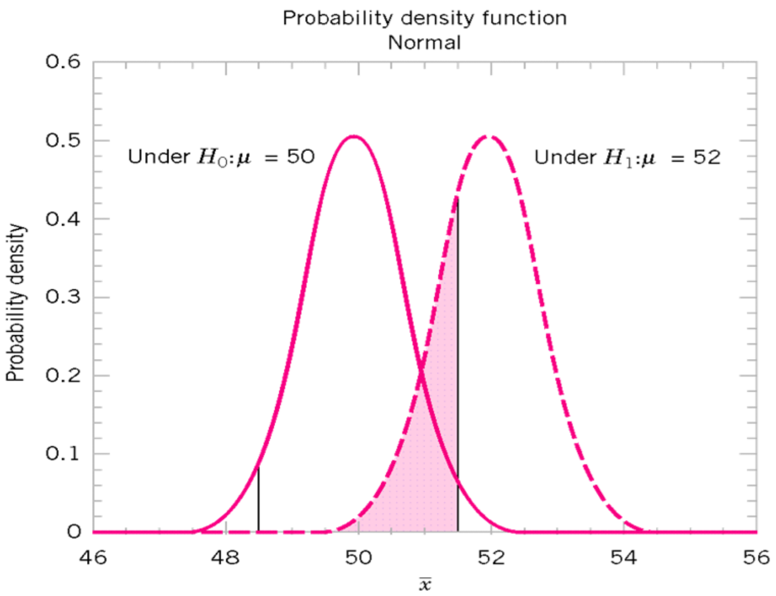
of a Type II error if the actual mean is 52.

Type II Error

- Type II error will be committed if the sample mean x-bar falls between 48.5 and 51.5 (critical region boundaries) when μ = 52.

$$\beta = P(48.5 \le \bar{x} \le 51.5\_when\_\mu = 52)$$

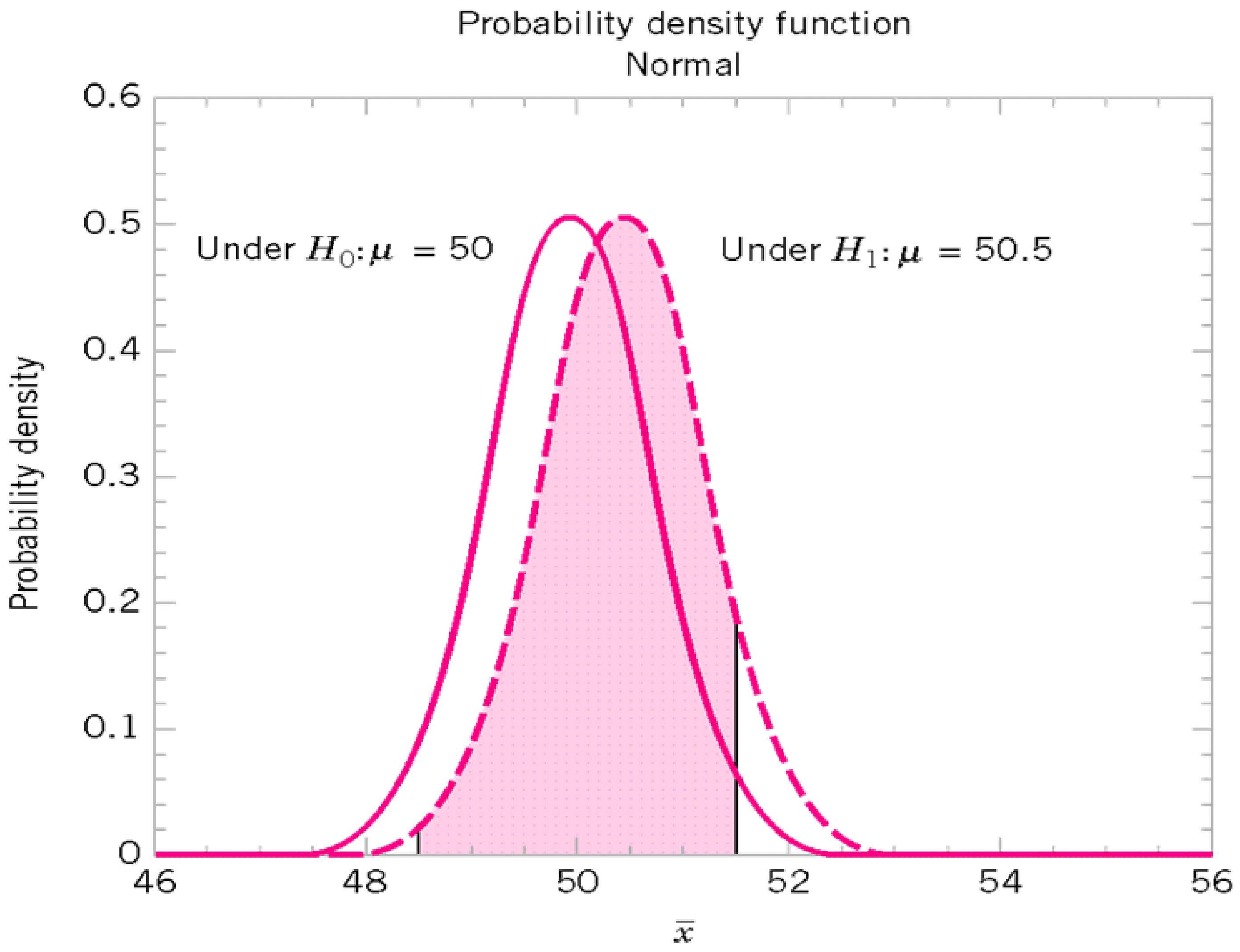
- 0.2643
- When μ = 50.5
- 0.8923



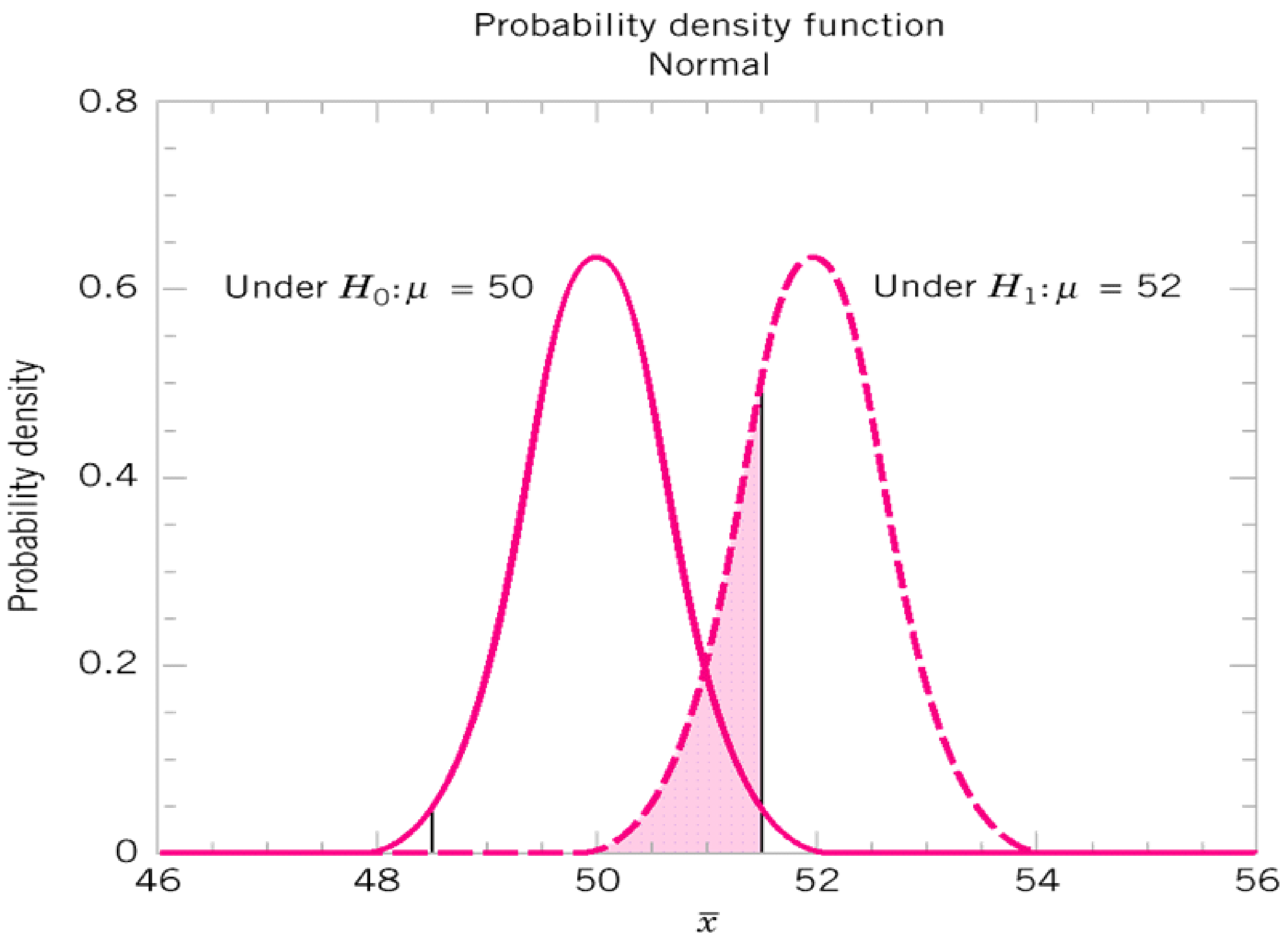
The probability of type II error when $\mu = 52$ and $n = 10$.

```
In [4]: beta = stats.norm.cdf((51.5-52)/0.79) #

In [5]: beta
Out[5]: 0.26339575390741593
```

```
In [8]: beta = stats.norm.cdf((51.5-50.5)/0.79) #

In [9]: beta
Out[9]: 0.8972117321157791
```



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



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

Computing the probability of a type II error may be the most difficult concept

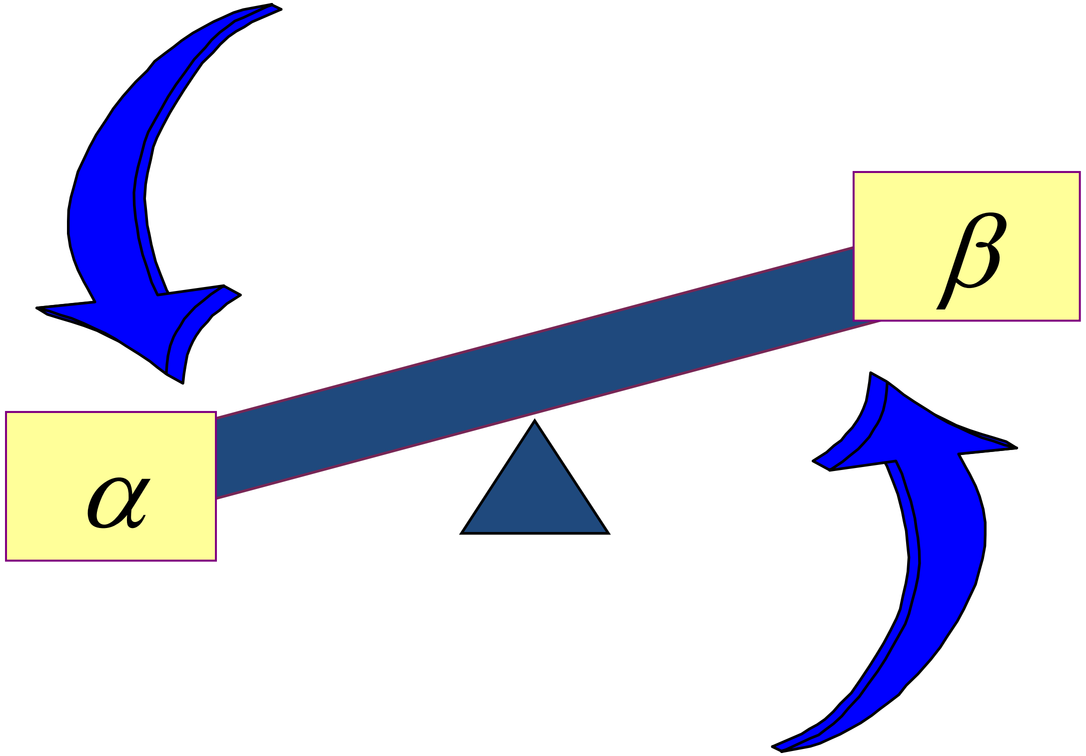
acceptance region	sample size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \bar{x} < 52$	10	0.0114	$\boxed{0.5000}$	$\boxed{0.9705}$
$48.5 < \bar{x} < 51.5$	16	0.0164	0.2119	$\boxed{0.9445}$
$48 < \bar{x} < 52$	16	$\boxed{0.0014}$	$\boxed{0.5000}$	$\boxed{0.9918}$

For constant n, increasing the acceptance region (hence decreasing α) increases β.

Increasing n, can decrease both types of errors.

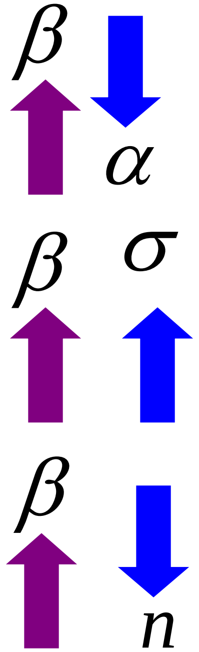
Type I & II Errors Have an Inverse Relationship

If you reduce the probability of one error, the other one increases so that everything else is unchanged.



Factors Affecting Type II Error

- True value of population parameter
  - β Increases when the difference between hypothesized parameter and its true value decrease
- Significance level α
  - Increases when β decreases
- Population standard deviation σ
  - Increases when β increases
- Sample size
  - β Increases when n decreases



How to Choose between Type I and Type II Errors

- Choice depends on the cost of the errors
- Choose smaller Type I Error when the cost of rejecting the maintained hypothesis is high
  - A criminal trial: convicting an innocent person
- Choose larger Type I Error when you have an interest in changing the status quo

Calculating the probability of Type II Error

$H_o: \mu = 8.3$
$H_1: \mu < 8.3$

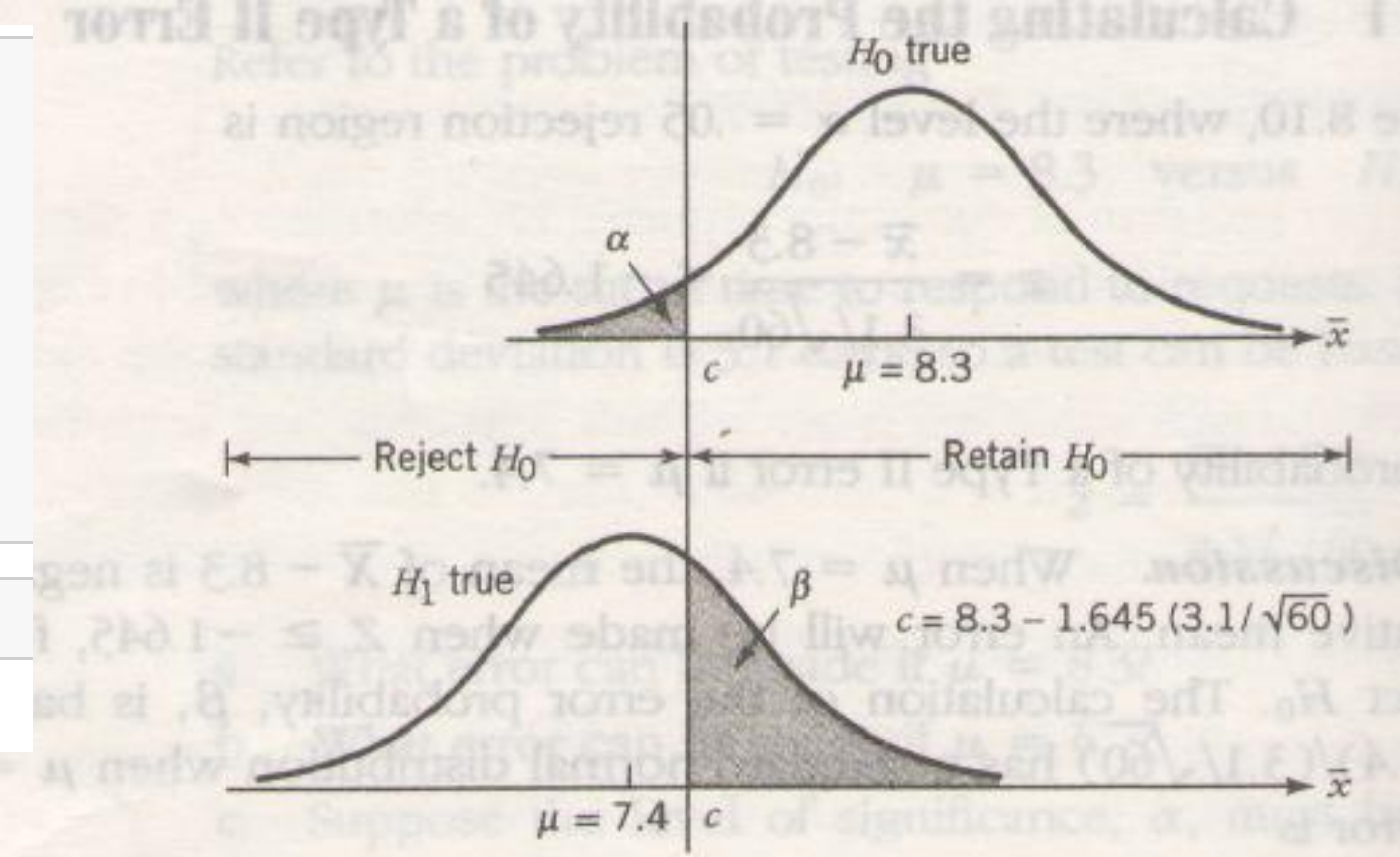
Determine the probability of Type II error if $\mu = 7.4$ at 5% significance level. $\sigma = 3.1$ and $n = 60$.

Solution:

In [48]:
```
def type_2(mu1,mu2,sigma,n,alfa):
 z =stats.norm.ppf(alfa)
 xbar = (mu1)+(z*sigma/np.sqrt(n))
 z2 = (xbar - mu2)/(sigma/np.sqrt(n))
 if(mu1 > mu2):
  beta= 1-stats.norm.cdf(z2)
 else:
  beta = stats.norm.cdf(z2)

 print (beta)
```

In [49]:
```
type_2(8.3,7.4,3.1,60,0.05)
```

```
0.27292999450730004
```



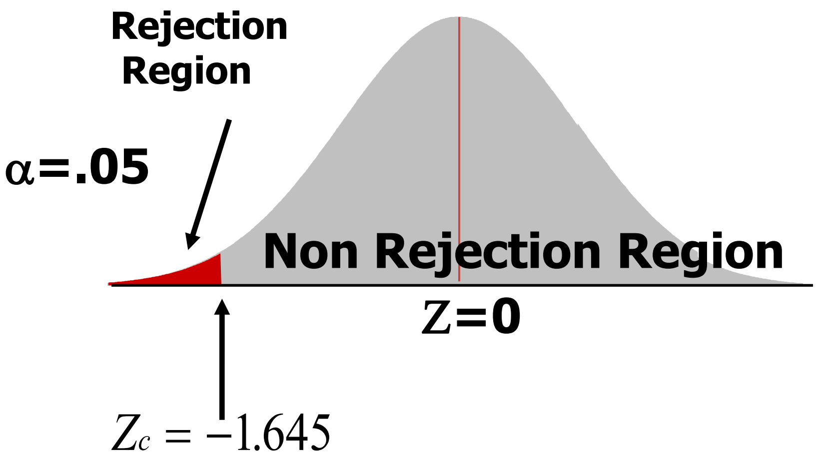
An error will be made when Z ≥ -1.645, for that will fail to reject Ho.

$\beta = 0.2729$

Solving for Type II Errors: Example

$$H_o: \mu = 12$$

$$H_a: \mu < 12$$



$$\overline{X}_c = \mu + Z_c \frac{\sigma}{\sqrt{n}}$$

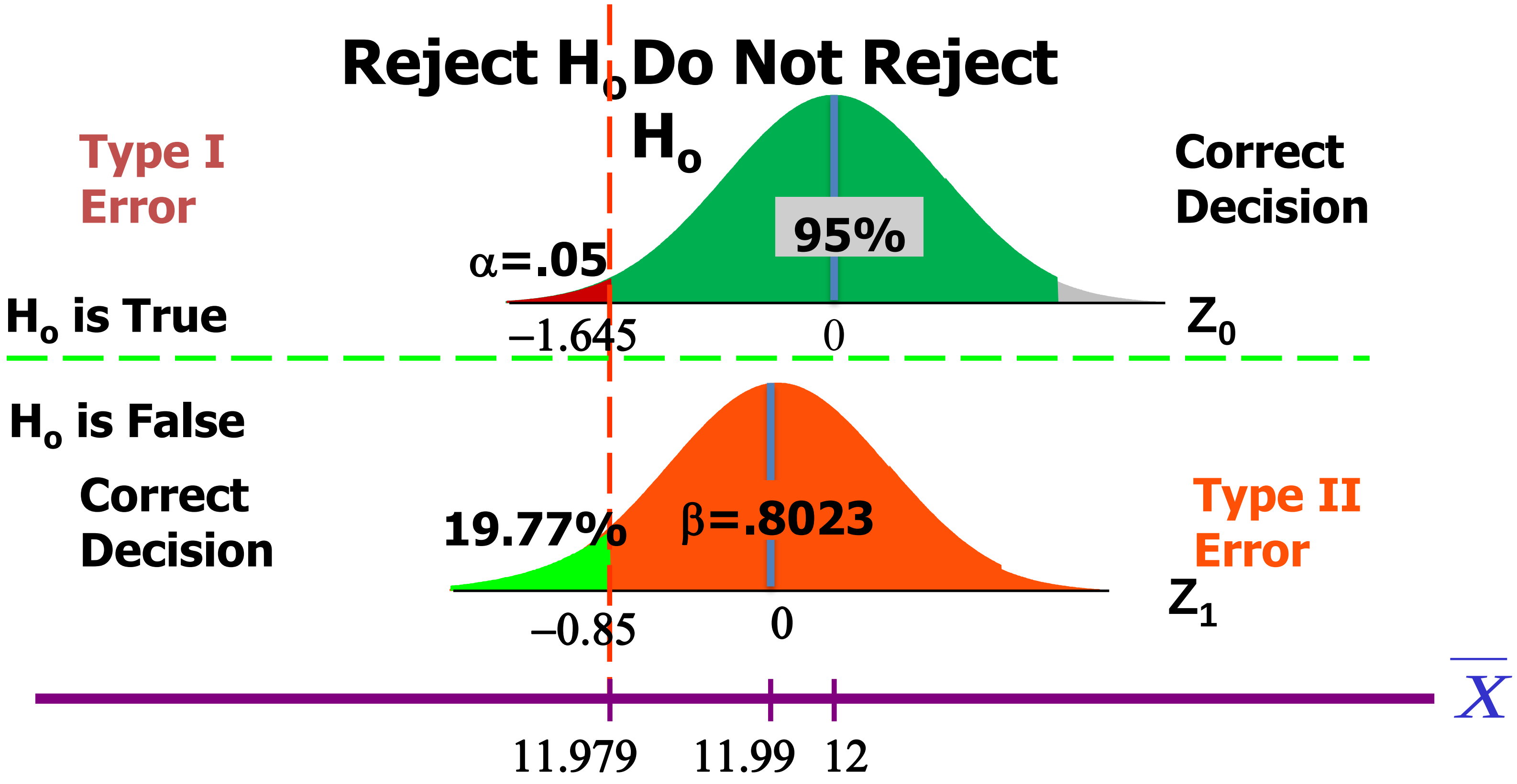
$$= 12 + (-1.645)\frac{0.10}{\sqrt{60}}$$

$$= 11.979$$

If $\overline{X} < 11.979$, reject H_o.

If $\overline{X} \geq 11.979$, do not reject H_o.

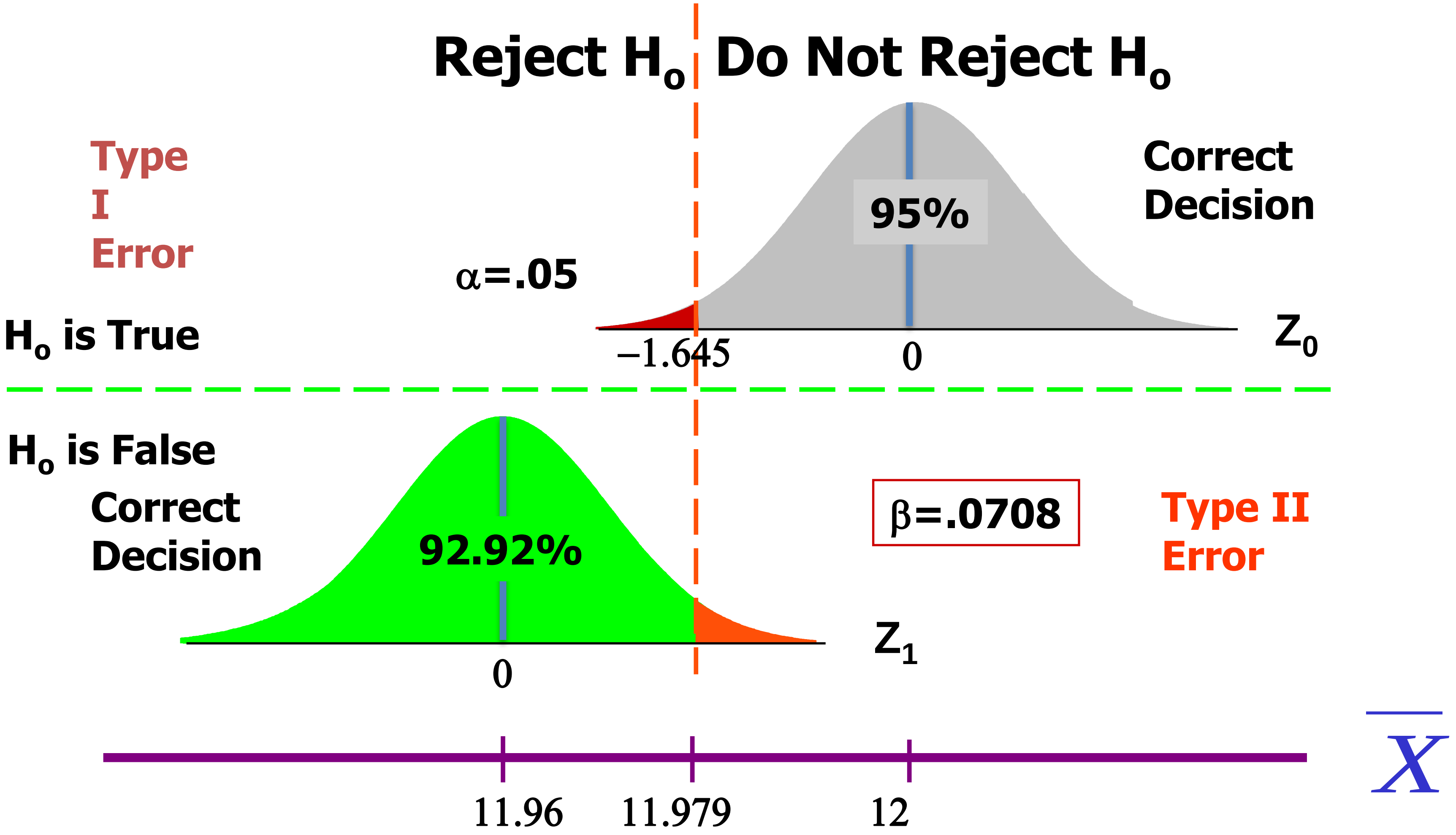
Type II Error for Example with μ =11.99 Kg



In [50]:
```
type_2(12,11.99,0.1,60,0.05)
```

```
0.8079200023112734
```

Type II Error for Demonstration with μ=11.96 Kg



In [51]:
```
type_2(12,11.96,0.1,60,0.05)
```

```
0.07303790512847008
```

Hypothesis Testing and Decision Making

- We have illustrated hypothesis testing applications referred to as significance tests
- In the tests, we compared the *p*-value to a controlled probability of a Type I error, *a*, which is called the level of significance for the test
- With a significance test, we control the probability of making the Type I error, but not the Type II error
- We recommended the conclusion "do not reject H_0" rather than "accept H_0" because the latter puts us at risk of making a Type II error

Hypothesis Testing and Decision Making

- With the conclusion “do not reject H_0”, the statistical evidence is considered inconclusive
- Usually this is an indication to postpone a decision until further research and testing is undertaken
- In many decision-making situations the decision maker may want, and in some cases may be forced, to take action with both the conclusion “do not reject H_0 “and the conclusion “reject H_0.”
- In such situations, it is recommended that the hypothesis-testing procedure be extended to include consideration of making a Type II error

Power of a test

- The mean response time for a random sample of 40 food-order is 13.25 minutes
- The population standard deviation is believed to be 3.2 minutes.
- The restaurant owner wants to perform a hypothesis test, with α =0.05 level of significance, to determine whether the service goal of 12 minutes or less is being achieved.

Calculating the Probability of a Type II Error

Hypotheses are: H_0: $\mu \leq 12$ and H_a: $\mu > 12$

Rejection rule is: Reject H_0 if $z \geq 1.645$

Value of the sample mean that identifies the rejection region:

$$z = \frac{\bar{x} - 12}{3.2/\sqrt{40}} \geq 1.645$$

$$\bar{x} \geq 12 + 1.645\left(\frac{3.2}{\sqrt{40}}\right) = 12.8323$$

We will accept H_0 when $x \leq 12.8323$

Calculating the Probability of a Type II Error

Probabilities that the sample mean will be in the acceptance region:

Values of μ	$z = \dfrac{12.8323 - \mu}{3.2/\sqrt{40}}$	β	1-β
14.0	-2.31	.0104	.9896
13.6	-1.52	.0643	.9357
13.2	-0.73	.2327	.7673
12.8323	0.00	.5000	.5000
12.8	0.06	.5239	.4761
12.4	0.85	.8023	.1977
12.0001	1.645	.9500	.0500

```
In [20]: type_2(14,12,3.2,40,0.05)
0.010499750448532241

In [21]: type_2(13.6,12,3.2,40,0.05)
0.06457982995225997

In [23]: type_2(13.2,12,3.2,40,0.05)
0.2336575101104159

In [22]: type_2(12.8323,12,3.2,40,0.05)
0.49995065746353273

In [27]: type_2(12.8,12,3.2,40,0.05)
0.5254013387545549

In [24]: type_2(12.4,12,3.2,40,0.05)
0.8035262335707292

In [26]: type_2(12.0001,12,3.2,40,0.05)
0.9499796127157129
```

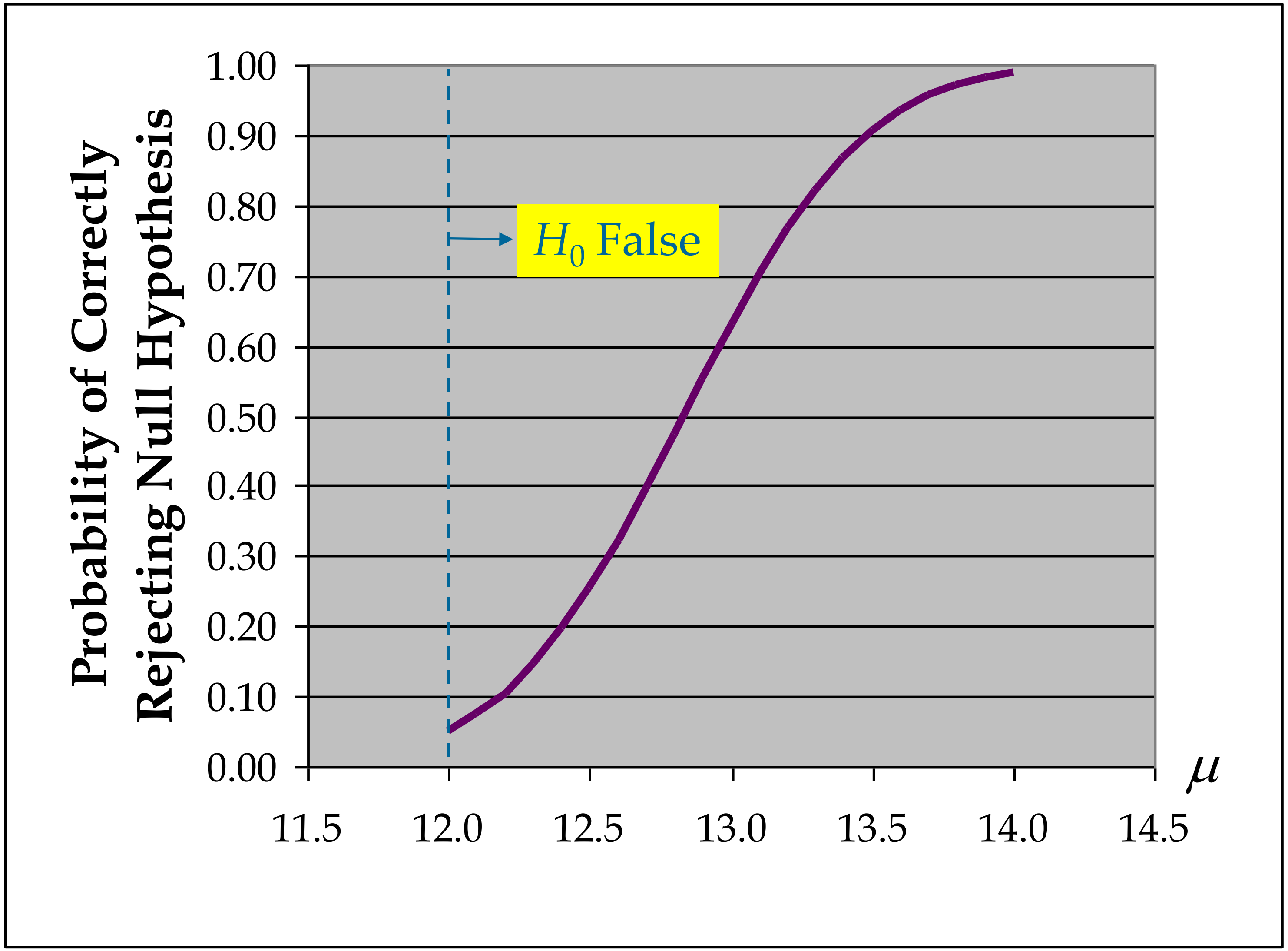
Power of the Test

- The probability of correctly rejecting H_0 when it is false is called the power of the test.

- For any particular value of m, the power is $1 - b$.

- We can show graphically the power associated with each value of μ ; such a graph is called a power curve.

Power Curve



Probability of Correctly Rejecting Null Hypothesis
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
H_0 False
11.5
12.0
12.5
13.0
13.5
14.0
14.5
μ

Thank You