Quiz

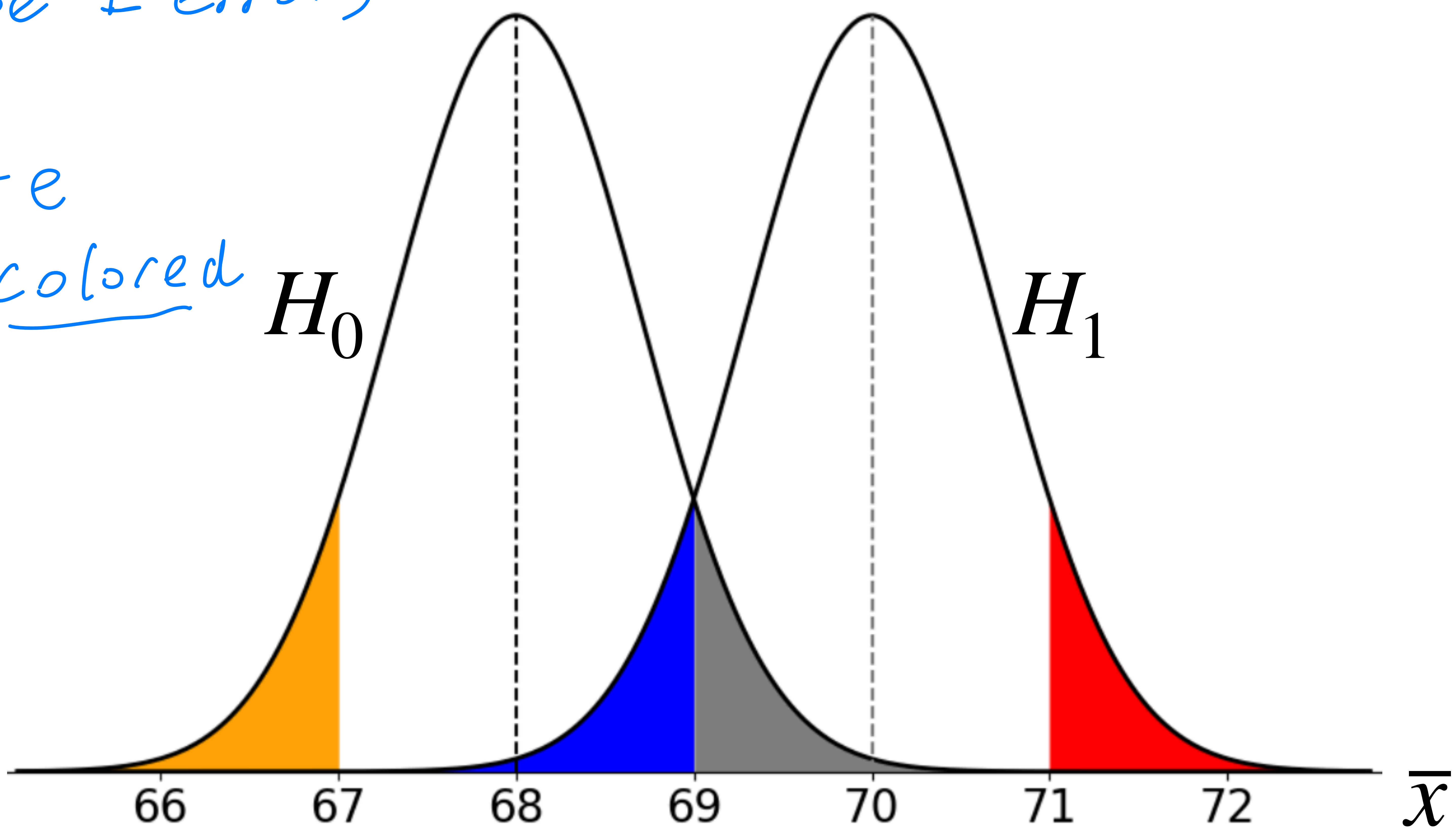
Let $H_0 : \mu = 68$, $H_1 : \mu \neq 68$. Assume the critical region is $\bar{x} < 67$ or $\bar{x} > 69$. The plots corresponding to H_0, H_1 are shown below.

Which area gives α and which area gives β?

P(Type I error)

P(Type II error)

indicate which colored areas?



Quiz

Let $H_0 : \mu = 68$, $H_1 : \mu \neq 68$. Assume the critical region is $\bar{x} < 67$ or $\bar{x} > 69$. The plots corresponding to H_0, H_1 are shown below.

Which area gives α and which area gives β?

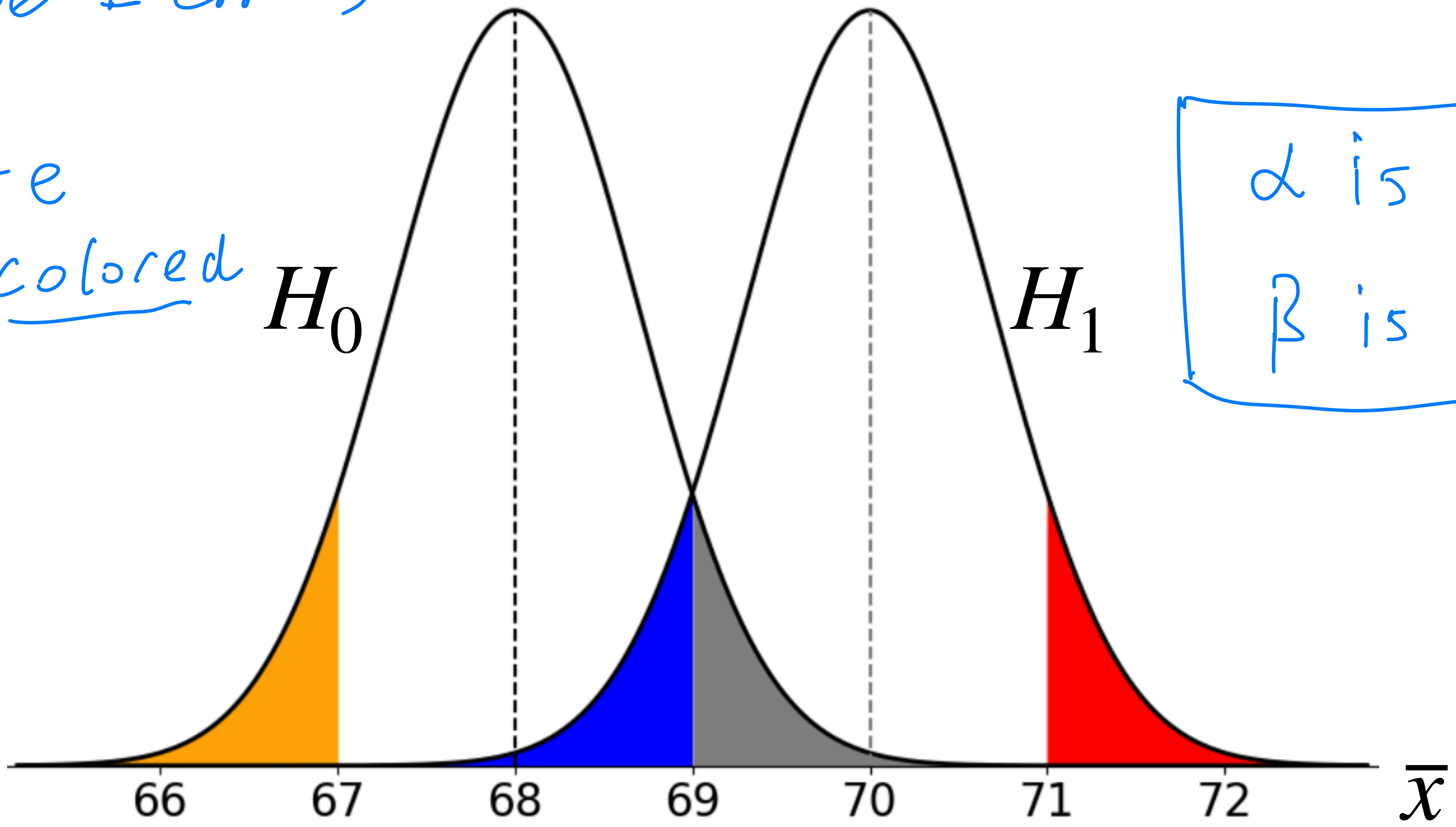
P(Type I error)

P(Type II error)

indicate which colored areas?

α is yellow + gray

β is blue



How to Present Conclusions?

Two ways of presenting conclusions in hypothesis testing:

Fixed significance level vs P-value

Relevant concepts

Test statistic

Critical region/value

Significance level (Size of the test)

Type I error, Type II error

Power of the test

We covered the relevant concepts, but skipped a detail.

Revisit the Weights Example and α

Example: Two-sided test with a continuous random variable

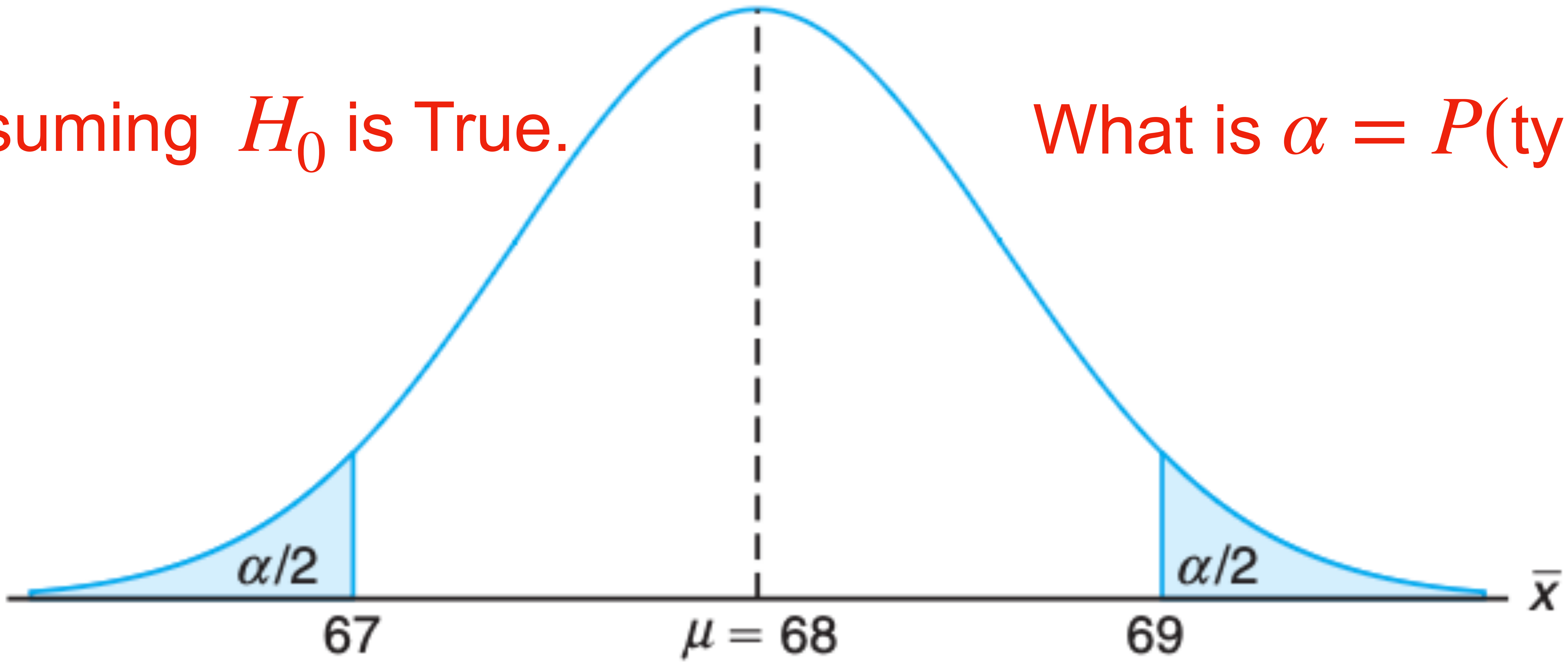
Testing whether average weight of male students is 68kg.

$$H_0 : \ \mu = 68 \qquad\qquad H_1 : \ \mu \neq 68$$

Assume $\sigma = 3.6$ and that we have a sample of size 36.

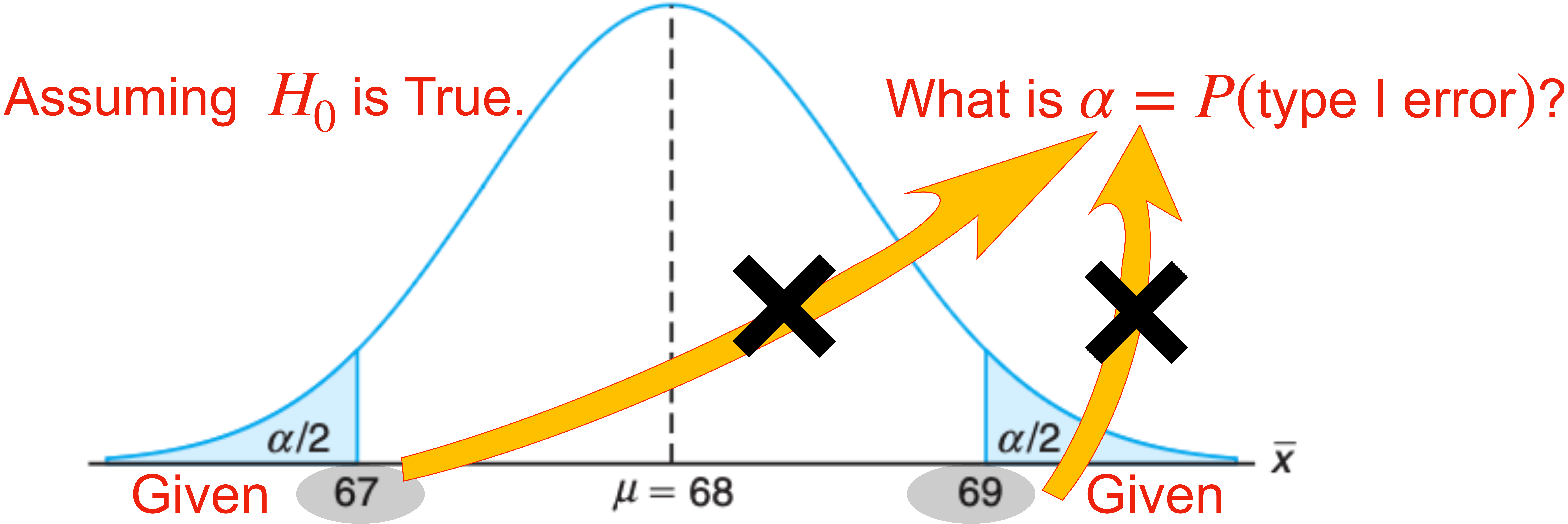
Assuming H_0 is True.

What is $\alpha = P(\text{type I error})$?



Revisit the Weights Example and α

Note: Only for illustrating the concept of significance level. Usually, it is not the case that we have the critical region then find significance.



Revisit the Weights Example and α

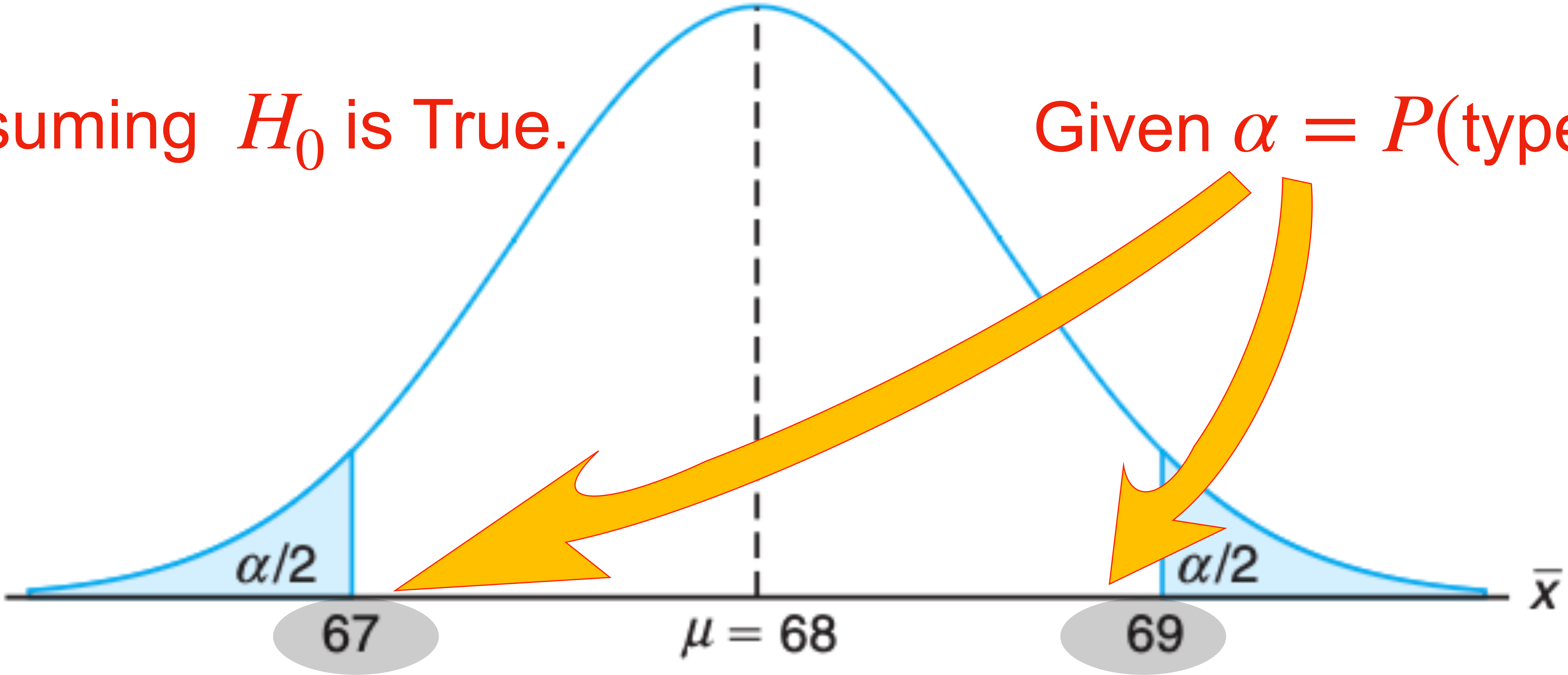
Note: Only for illustrating the concept of significance level. Usually, it is not the case that we have the critical region then find significance.

On the contrary, we find the critical region from fixed α. Then we check:

Does the test statistic lie in critical region? If so reject H_0.

Assuming H_0 is True.

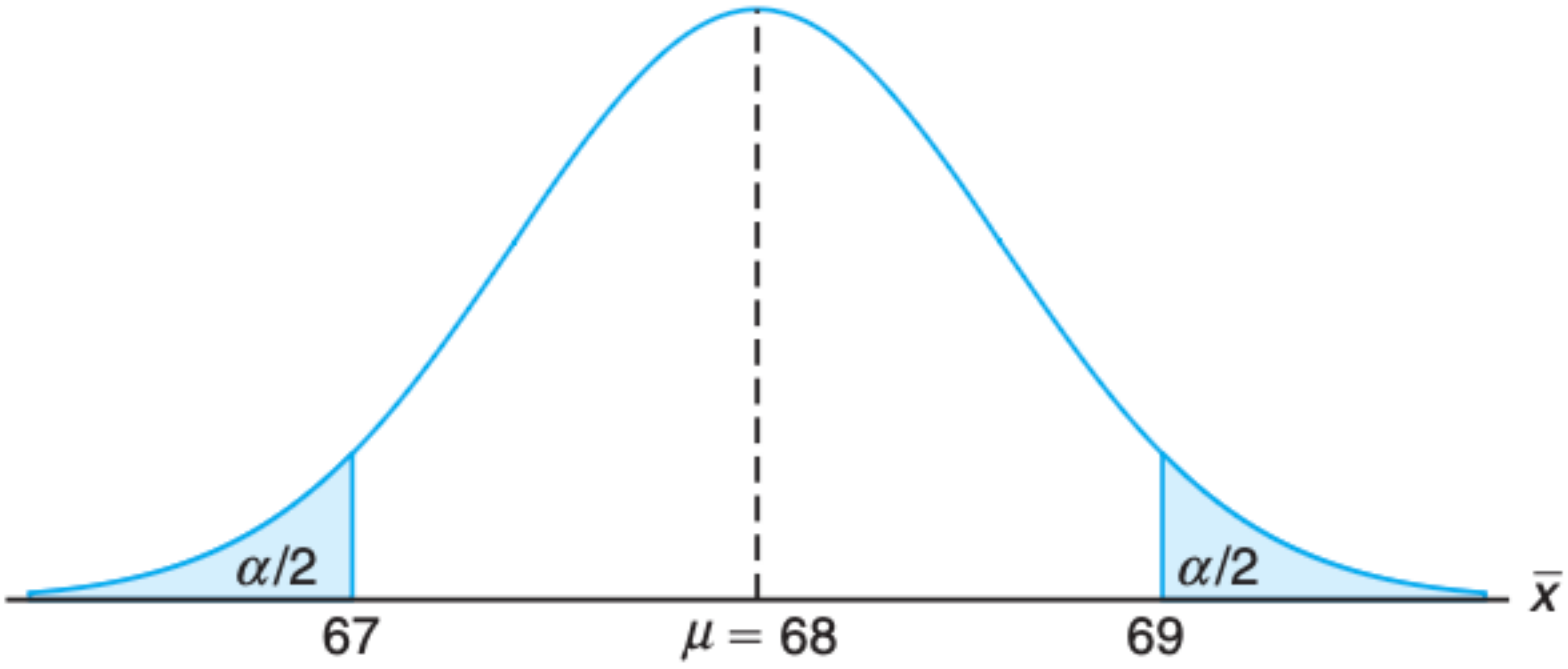
Given $\alpha = P(\text{type I error})$.



Fixed Significance Level

This is called reporting results with fixed significance level.

α is usually set to 0.05.

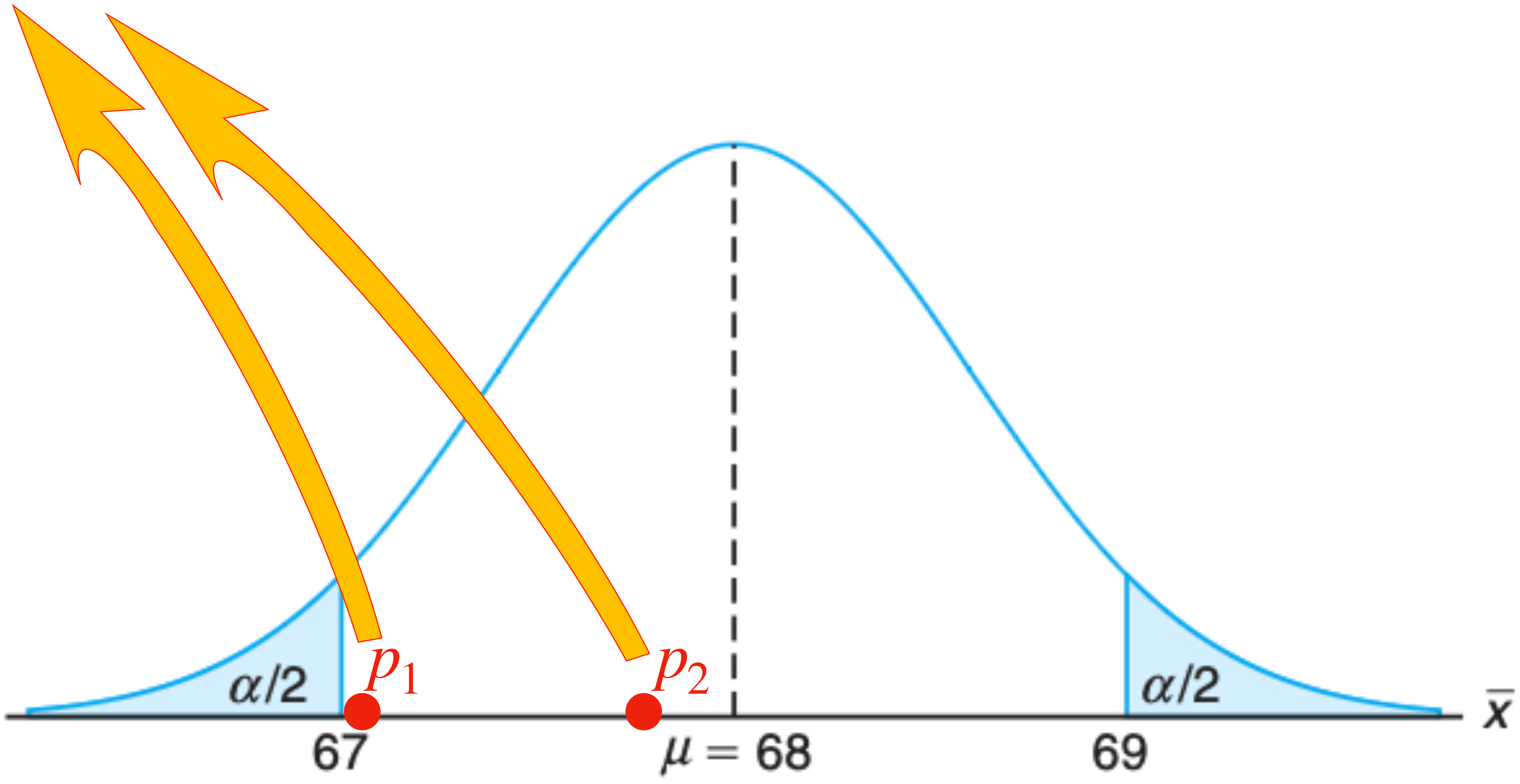


Fixed Significance Level

Cases where such reporting has disadvantages:

Say some fixed α. Both these instances would be reported the same way:

"We don't reject H_0 at significance α" (although p_1 is so close to rejection)

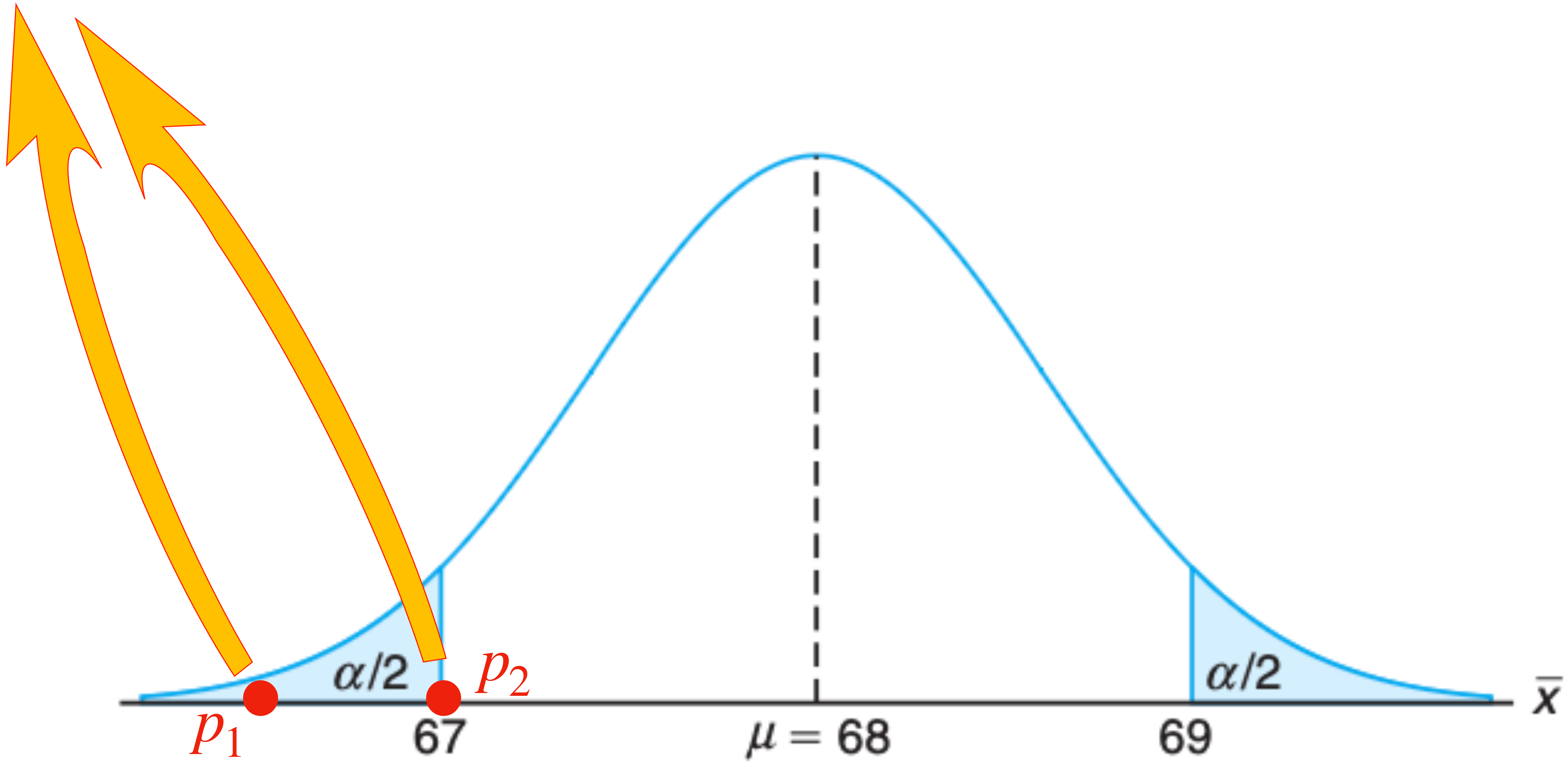


Fixed Significance Level

Cases where such reporting has disadvantages:

Say some fixed α. Both these instances would be reported the same way:

"We reject H_0 at significance α" (although for p_1 rejection is much stronger)



P-value

Reporting the P-value takes care of this:

Probability of observing data at least as extreme as observed test statistic.

(Assuming H_0 is true.)

A ***P*-value** is the lowest level (of significance) at which the observed value of the test statistic is significant.

Revisit the Weights Example

Example: Two-sided test with a continuous random variable

Testing whether average weight of male students is 68kg.

$$H_0 : \ \mu = 68 \qquad\qquad H_1 : \ \mu \neq 68$$

Assume $\sigma = 3.6$ and that we have a sample of size 36. Say $\overline{X} = 67$.

Revisit the Weights Example

Example: Two-sided test with a continuous random variable

Testing whether average weight of male students is 68kg.

$$H_0 : \mu = 68 \qquad H_1 : \mu \neq 68$$

Assume $\sigma = 3.6$ and that we have a sample of size 36. Say $\bar{X} = 67$.

$P(\bar{X} < 67$ when $\mu = 68) + P(\bar{X} > 69$ when $\mu = 68)$ Already found as 0.095.

Revisit the Weights Example

Example: Two-sided test with a continuous random variable

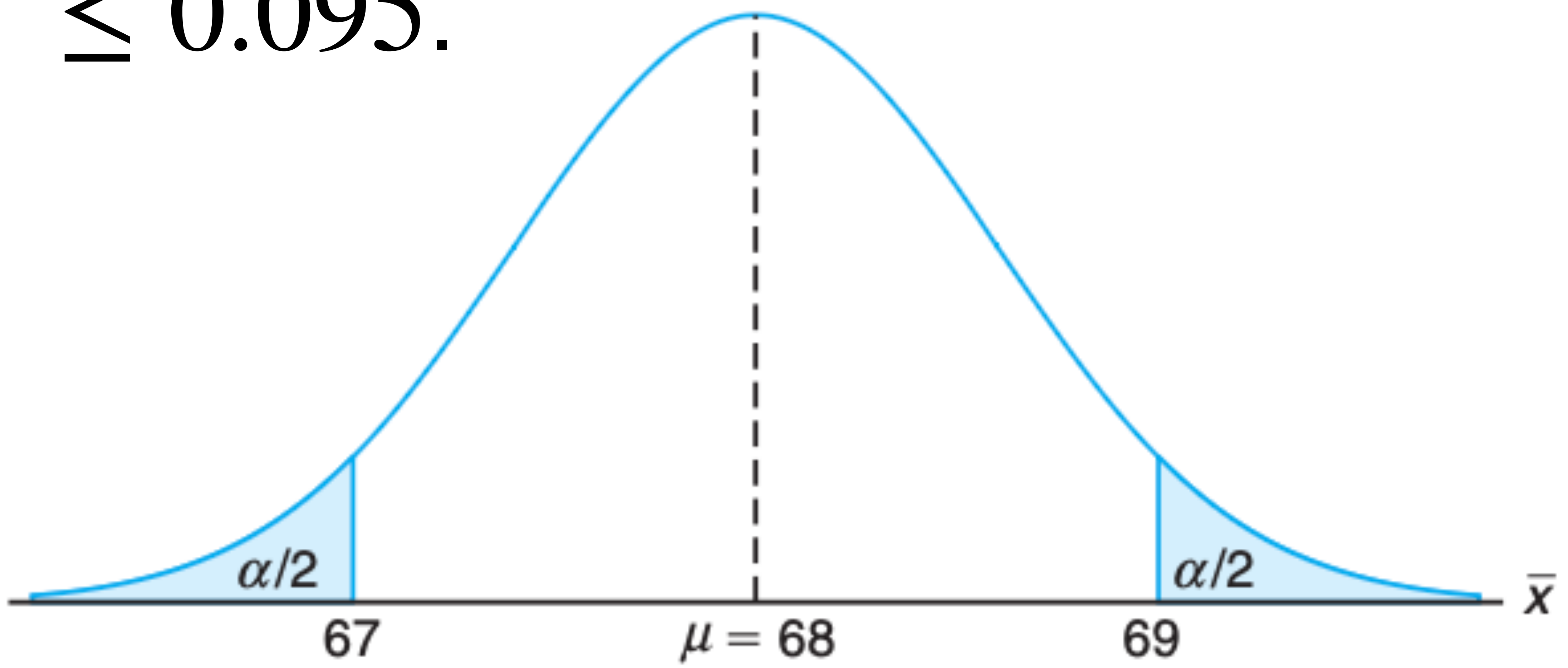
Testing whether average weight of male students is 68kg.

$$H_0 : \mu = 68 \qquad H_1 : \mu \neq 68$$

Assume $\sigma = 3.6$ and that we have a sample of size 36. Say $\overline{X} = 67$.

$P(\overline{X} < 67 \text{ when } \mu = 68) + P(\overline{X} > 69 \text{ when } \mu = 68)$ Already found as 0.095.

In terms of P-value: This probability is the P-value. Observed value would not be considered significant at any significance level ≤ 0.095.



Revisit the Weights Example

Example: Two-sided test with a continuous random variable

Testing whether average weight of male students is 68kg.

$$H_0 : \mu = 68 \qquad H_1 : \mu \neq 68$$

Assume $\sigma = 3.6$ and that we have a sample of size 36. Say $\overline{X} = 67$.

$P(\overline{X} < 67$ when $\mu = 68) + P(\overline{X} > 69$ when $\mu = 68)$ Already found as 0.095.

In terms of P-value: This probability is the P-value. Observed value would not be considered significant at any significance level ≤ 0.095.

In terms of fixed significance level:

"At $\alpha = 0.1$ we reject H_0." ⇝ since $0.095 \leq 0.1$

"At $\alpha = 0.05$ we fail to reject H_0." ⇝ since $0.095 > 0.05$

Procedure for Hypothesis Testing

Approach to Hypothesis Testing with Fixed Probability of Type I Error	**1.** State the null and alternative hypotheses. **2.** Choose a fixed significance level α. **3.** Choose an appropriate test statistic and establish the critical region based on α. **4.** Reject H_0 if the computed test statistic is in the critical region. Otherwise, do not reject. **5.** Draw scientific or engineering conclusions.

Significance Testing (P-Value Approach)	**1.** State null and alternative hypotheses. **2.** Choose an appropriate test statistic. **3.** Compute the P-value based on the computed value of the test statistic. **4.** Use judgment based on the P-value and knowledge of the scientific system.

Hypothesis Testing Concerning a Single Mean

Mean of a distribution with known variance: Two-sided

$X_1, \cdots, X_n$ with unknown μ and known variance σ^2. Assume large sample.

Test the hypotheses: $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$

so we can apply CLT: $\bar{X}$ is normal.

Hypothesis Testing Concerning a Single Mean

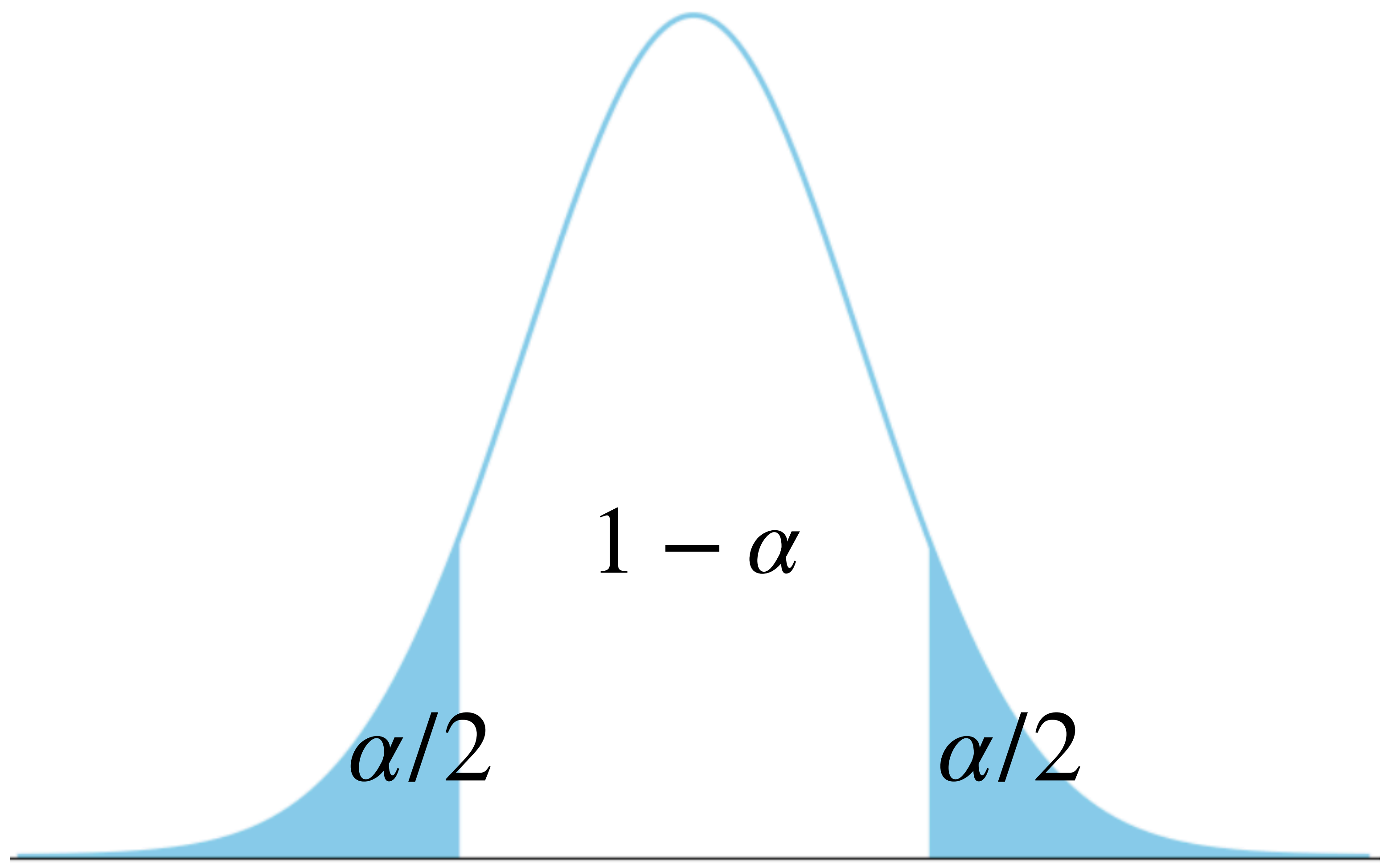
Mean of a distribution with known variance: Two-sided

$X_1, \cdots, X_n$ with unknown μ and known variance σ^2. Assume large sample.

Test the hypotheses: $H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$

Reject H_0 if:

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \text{ or } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}$$



Hypothesis Testing Concerning a Single Mean

Mean of a distribution with known variance: Two-sided

$X_1, \cdots, X_n$ with unknown μ and known variance σ^2. Assume large sample.

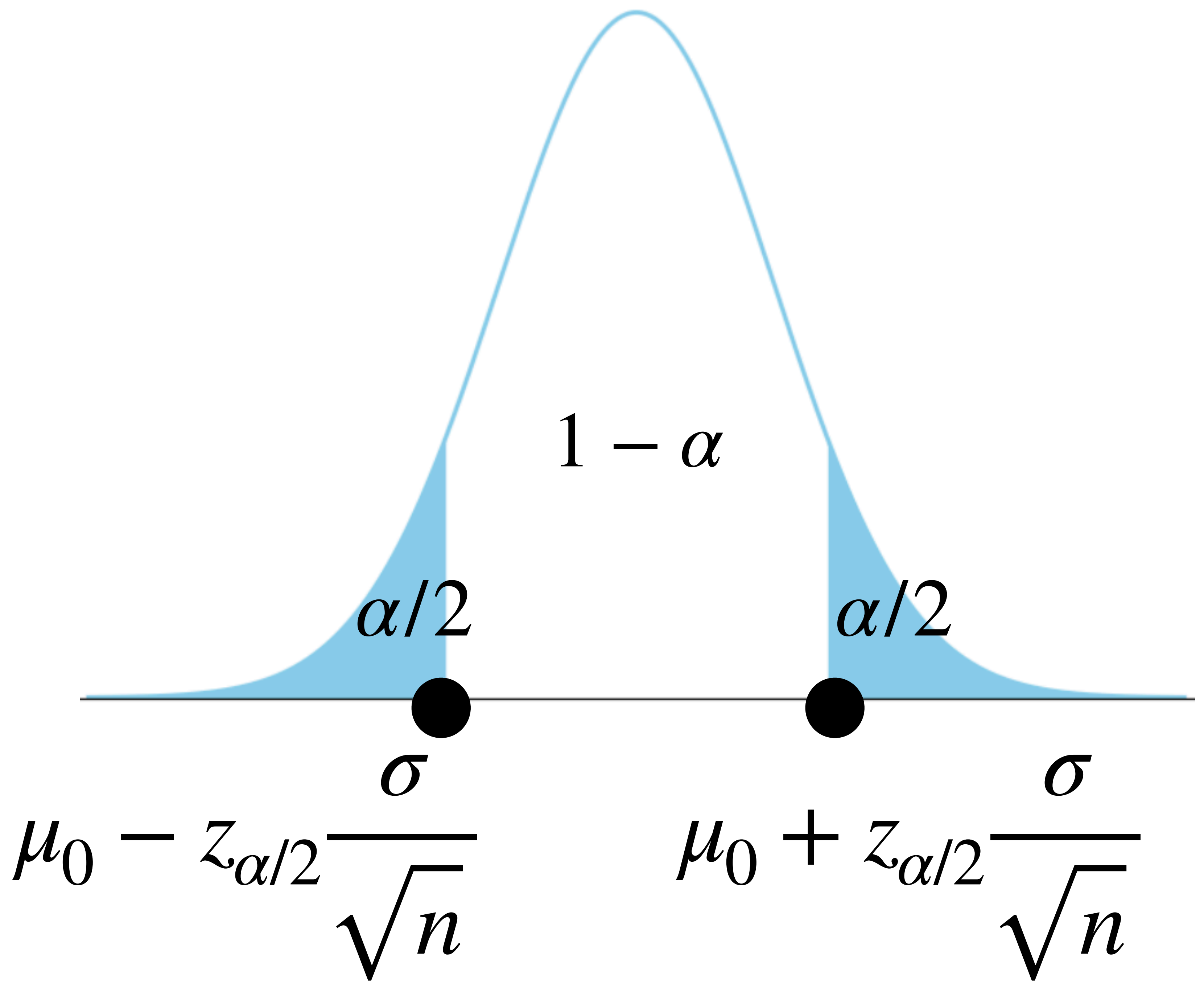
Test the hypotheses: $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$

Reject H_0 if:

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \text{ or } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}$$

or equivalently

Reject H_0 if $\bar{x}$ lies in:



Hypothesis Testing Concerning a Single Mean

Mean of a distribution with known variance: Two-sided

$X_1, \cdots, X_n$ with unknown μ and known variance σ^2. Assume large sample.

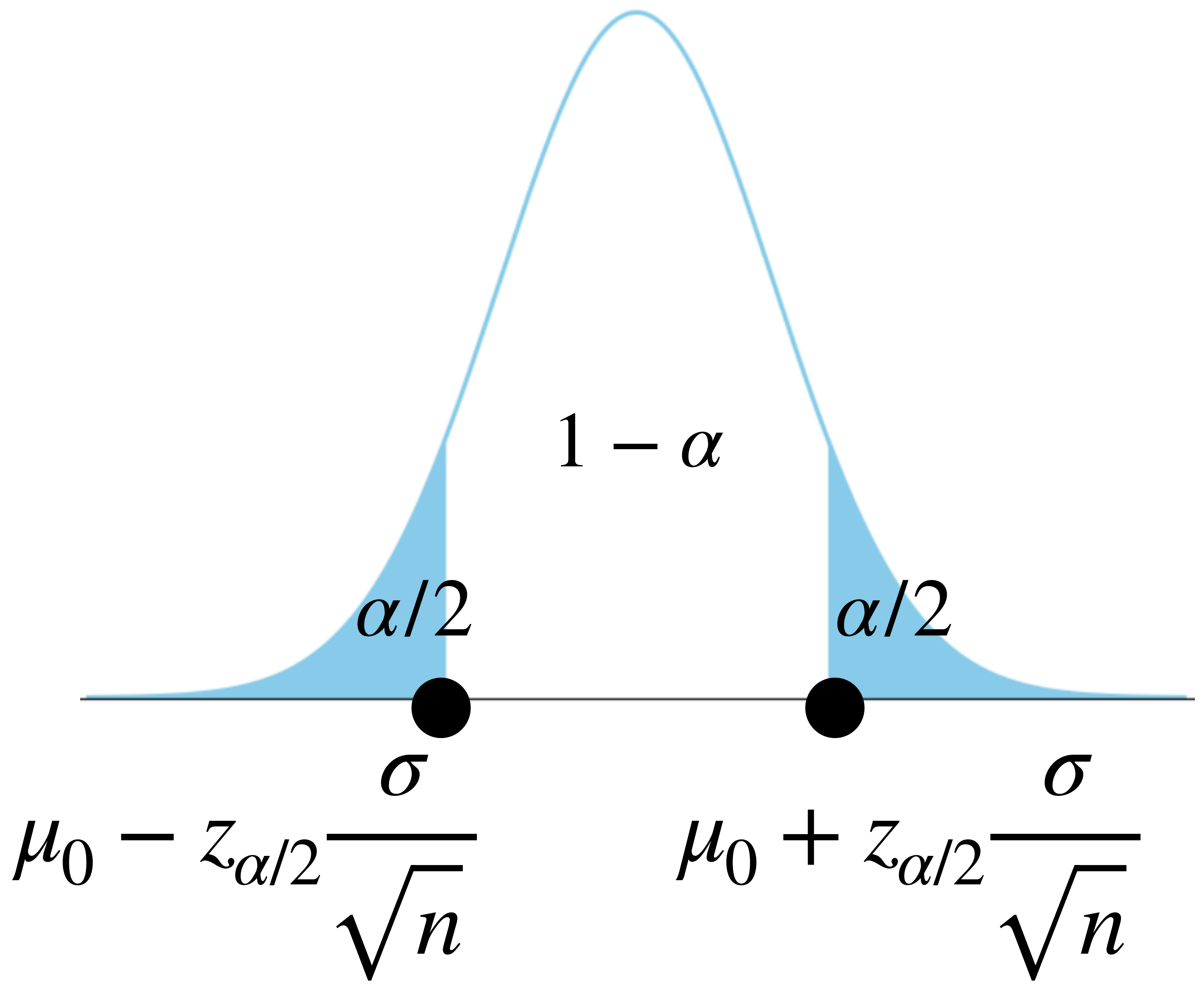
Test the hypotheses: $H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$

Reject H_0 if:

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \text{ or } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}$$

or equivalently

Reject H_0 if $\bar{x}$ lies in:



Note: The test is similar to computing $100(1-\alpha)$ confidence interval.

Hypothesis Testing Concerning a Single Mean

Mean of a distribution with known variance: One-sided

$X_1, \cdots, X_n$ with unknown μ and known variance σ^2. Assume large sample.

Test the hypotheses: $H_0: \ \mu < \mu_0 \quad H_1: \ \mu \geq \mu_0$

Hypothesis Testing Concerning a Single Mean

Mean of a distribution with known variance: One-sided

$X_1, \cdots, X_n$ with unknown μ and known variance σ^2. Assume large sample.

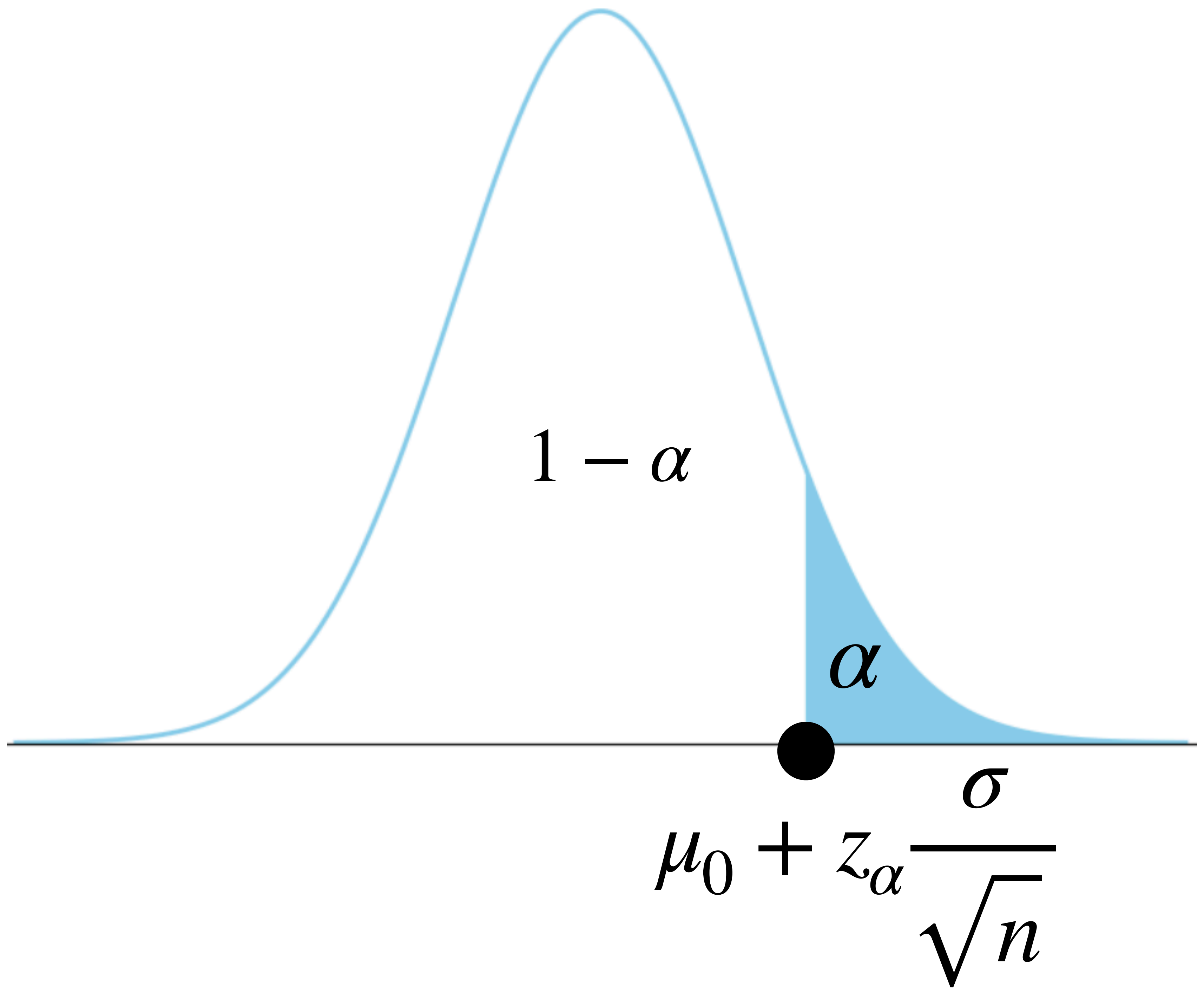
Test the hypotheses: $H_0: \ \mu < \mu_0$ $H_1: \ \mu \geq \mu_0$

Reject H_0 if:

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$$

or equivalently

Reject H_0 if $\bar{x}$ lies in:



Hypothesis Testing Concerning a Single Mean

Mean of a distribution with known variance: One-sided

$X_1, \cdots, X_n$ with unknown μ and known variance σ^2. Assume large sample.

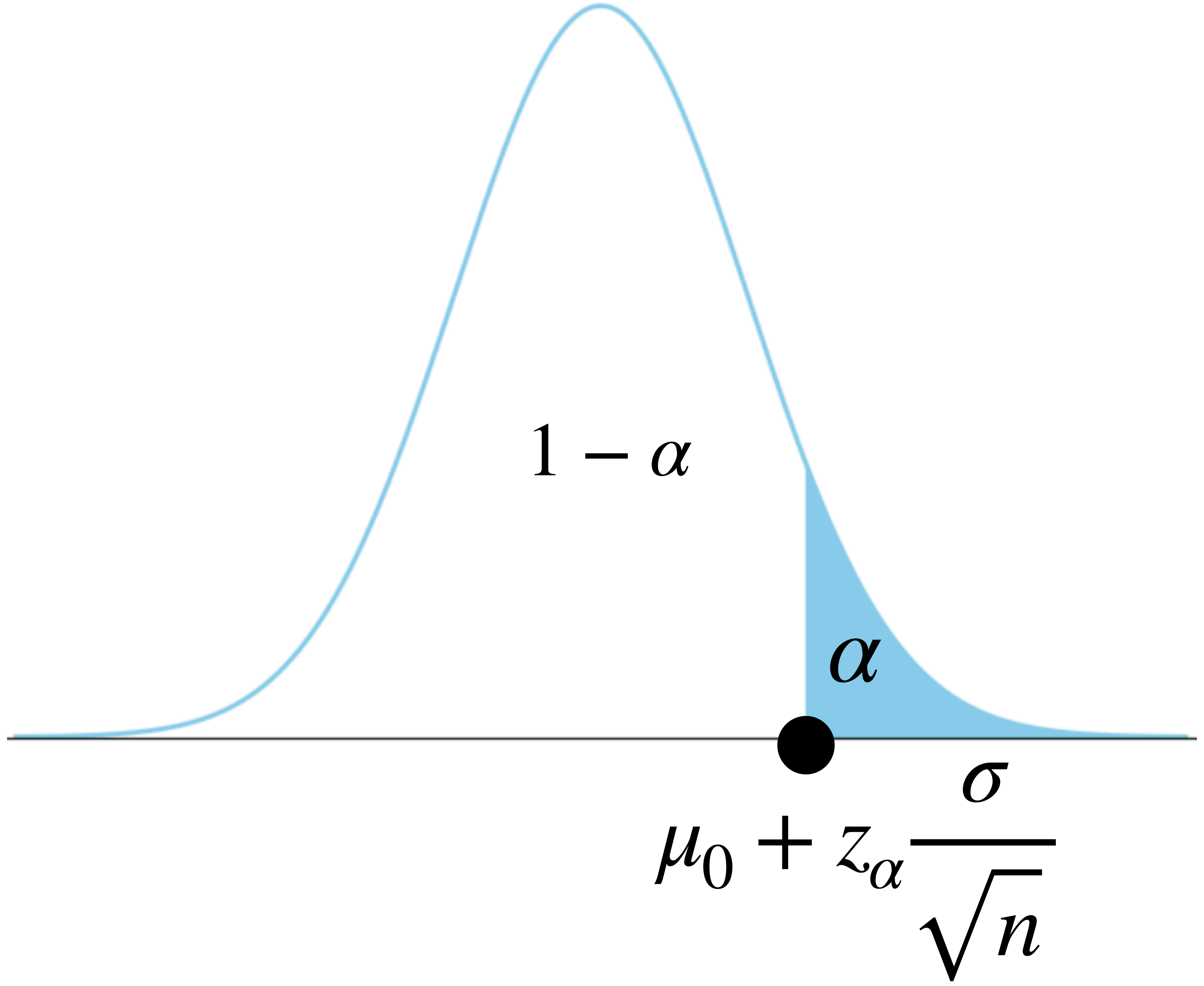
Test the hypotheses: $H_0: \ \mu < \mu_0 \qquad H_1: \ \mu \geq \mu_0$

Reject H_0 if:

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$$

or equivalently

Reject H_0 if $\bar{x}$ lies in:



The case of testing: (left tail testing)

$H_0: \ \mu > \mu_0 \qquad H_1: \ \mu \leq \mu_0$

can be done similarly.

Hypothesis Testing Concerning a Single Mean

Example 10.3: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

n $\bar{x}$ σ α

Use $norm.ppf(0.95) = 1.645$

$norm.cdf(2.02) = 0.9783$

Hypothesis Testing Concerning a Single Mean

Example 10.3: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Use $norm.ppf(0.95) = 1.645$

$norm.cdf(2.02) = 0.9783$

Solution:

1. H_0: $\mu = 70$ years.
2. H_1: $\mu > 70$ years.
3. $\alpha = 0.05$.
4. Critical region: $z > 1.645$, where $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
5. Computations: $\bar{x} = 71.8$ years, $\sigma = 8.9$ years, and hence $z = \frac{71.8-70}{8.9/\sqrt{100}} = 2.02$.
6. Decision: Reject H_0 and conclude that the mean life span today is greater than 70 years.

With P-value reporting: $P = P(Z > 2.02) = 0.0217$.

Hypothesis Testing Concerning a Single Mean

Example 10.3: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

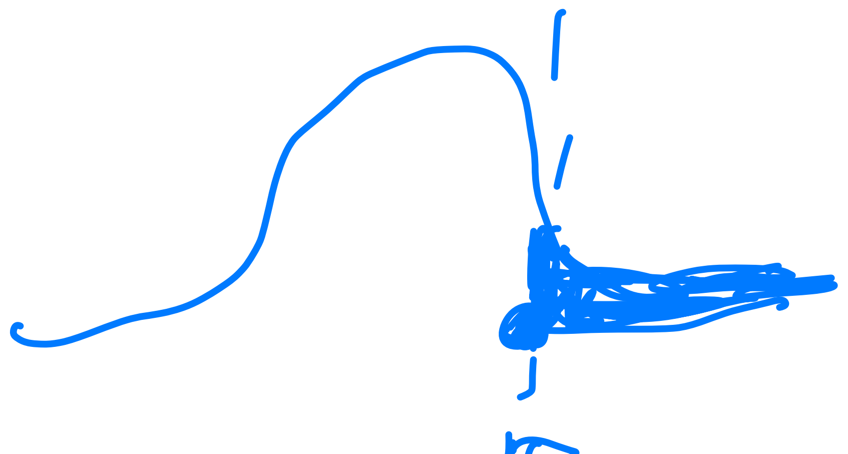
Use $norm.ppf(0.95) = 1.645$

$norm.cdf(2.02) = 0.9783$

Solution:

$\mu \leq 70$

1. H_0: $\mu = 70$ years.
2. H_1: $\mu > 70$ years.
3. $\alpha = 0.05$.
4. Critical region: $z > 1.645$, where $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.
5. Computations: $\bar{x} = 71.8$ years, $\sigma = 8.9$ years, and hence $z = \frac{71.8-70}{8.9/\sqrt{100}} = 2.02$.
6. Decision: Reject H_0 and conclude that the mean life span today is greater than 70 years.



what is z-value such that area to the right = 0.05

With P-value reporting: $P = P(Z > 2.02) = 0.0217$.

Hypothesis Testing Concerning a Single Mean

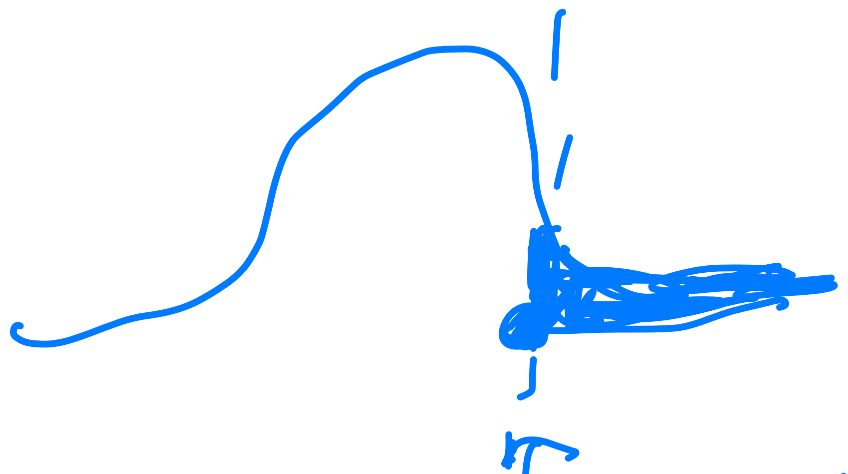
Example 10.3: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Use $norm.ppf(0.95) = 1.645$

$norm.cdf(2.02) = 0.9783$

Solution:

$\mu \le 70$



what is z-value such that area to the right = 0.05

1. H_0: $\mu = 70$ years.
2. H_1: $\mu > 70$ years.
3. $\alpha = 0.05$.
4. Critical region: $z > 1.645$, where $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
5. Computations: $\bar{x} = 71.8$ years, $\sigma = 8.9$ years, and hence $z = \frac{71.8-70}{8.9/\sqrt{100}} = 2.02$.
6. Decision: Reject H_0 and conclude that the mean life span today is greater than 70 years.

With P-value reporting: $P = P(Z > 2.02) = 0.0217$.

Testing a Single Mean with Unknown Variance

The t-Statistic for a Test on a Single Mean (Variance Unknown)

For the two-sided hypothesis

$$H_0\colon\ \mu = \mu_0,$$
$$H_1\colon\ \mu \neq \mu_0,$$

we reject H_0 at significance level α when the computed t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

sample standard deviation ← s

exceeds $t_{\alpha/2,n-1}$ or is less than $-t_{\alpha/2,n-1}$.

Testing a Single Mean with Unknown Variance

Example 10.5: The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

n $\bar{x}$ s α

Use $t.ppf(0.05, 11) = -1.796$

$t.cdf(-1.16, 11) = 0.135$

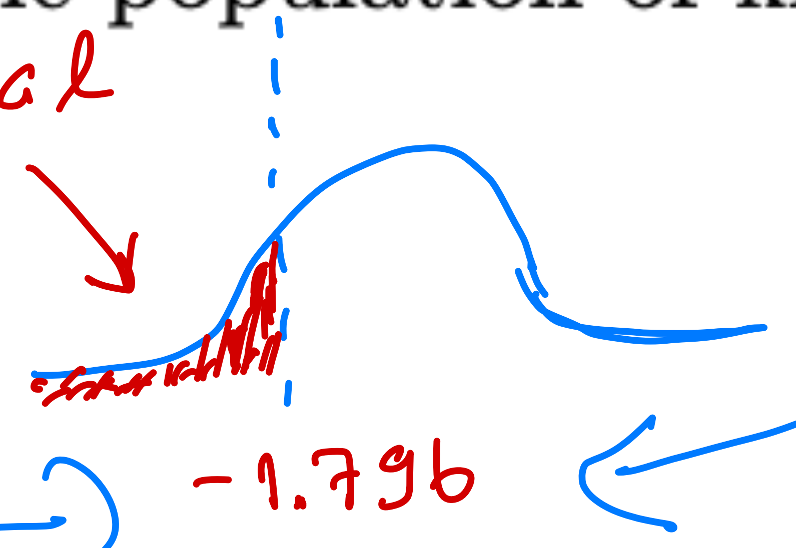
Testing a Single Mean with Unknown Variance

Example 10.5: The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

Solution:

Use $t.ppf(0.05,11) = -1.796$

$t.cdf(-1.16,11) = 0.135$



1. H_0: $\mu = 46$ kilowatt hours.
2. H_1: $\mu < 46$ kilowatt hours.
3. $\alpha = 0.05$.
4. Critical region: $t < -1.796$, where $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ with 11 degrees of freedom.
5. Computations: $\bar{x} = 42$ kilowatt hours, $s = 11.9$ kilowatt hours, and $n = 12$. Hence,

$$t = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16,$$

6. Decision: Do not reject H_0 and conclude that the average number of kilowatt hours used annually by home vacuum cleaners is not significantly less than 46.

Testing a Single Mean with Unknown Variance

Example 10.5: The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

Solution:

Use $t.ppf(0.05,11) = -1.796$

$t.cdf(-1.16,11) = 0.135$

1. H_0: $\mu = 46$ kilowatt hours.
2. H_1: $\mu < 46$ kilowatt hours.
3. $\alpha = 0.05$.
4. Critical region: $t < -1.796$, where $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ with 11 degrees of freedom.
5. Computations: $\bar{x} = 42$ kilowatt hours, $s = 11.9$ kilowatt hours, and $n = 12$. Hence,

$$t = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16,$$

With P-value reporting:

$P = P(T < -1.16) \approx 0.135.$

6. Decision: Do not reject H_0 and conclude that the average number of kilowatt hours used annually by home vacuum cleaners is not significantly less than 46.

Testing a Single Mean with Unknown Variance

Using scipy for single sample t-test

```
from scipy import stats
import numpy as np
sample = np.array([7.07, 7.00, 7.10, 6.97, 7.00, 7.03, 7.01, 7.01, 6.98, 7.08])
null_hypothesis_mean = 7
t_statistic, p_value = stats.ttest_1samp(sample, null_hypothesis_mean)
print(f"T-statistic: {t_statistic:.4f}")
print(f"P-value: {p_value:.4f}")
```

Output

```
T-statistic: 1.7954
P-value: 0.1062
```

Testing a Single Mean with Unknown Variance

Using scipy for single sample t-test

```
from scipy import stats
import numpy as np
sample = np.array([7.07, 7.00, 7.10, 6.97, 7.00, 7.03, 7.01, 7.01, 6.98, 7.08])
null_hypothesis_mean = 7
t_statistic, p_value = stats.ttest_1samp(sample, null_hypothesis_mean)
print(f"T-statistic: {t_statistic:.4f}")
print(f"P-value: {p_value:.4f}")
```

Output

```
T-statistic: 1.7954
P-value: 0.1062
```

Passing the parameter:

alternative='less'

tests whether the mean is less than the null hypothesis mean. (alternative='greater' tests whether it is greater than the null hypothesis mean)

Two Samples Tests on Means

Known Variance Case: First consider the two-sided test.

$$H_0 : \mu_1 - \mu_2 = d_0$$

Two Samples Tests on Means

Known Variance Case: First consider the two-sided test.

$$H_0 : \mu_1 - \mu_2 = d_0$$

Similar to test concerning single mean. Now $\bar{x}_1 - \bar{x}_2$ is standardized:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

(apply standardization of a normal R.V. X: $\frac{X - \mu_X}{\sigma_X}$. In this case X is replaced with $\bar{x}_1 - \bar{x}_2$ and its mean is thus $\mu_1 - \mu_2 = d_0$. Its variance is $\sigma_1^2/n_1 + \sigma_2^2/n_2$.)

Two Samples Tests on Means

Known Variance Case: First consider the two-sided test.

$$H_0 : \mu_1 - \mu_2 = d_0$$

Similar to test concerning single mean. Now $\bar{x}_1 - \bar{x}_2$ is standardized:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

The rest is the same as the two-sided test concerning single mean:

Reject H_0 if $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$ otherwise don't reject H_0

Two Samples Tests on Means

Known Variance Case: First consider the two-sided test.

$$H_0 : \mu_1 - \mu_2 = d_0$$

Similar to test concerning single mean. Now $\bar{x}_1 - \bar{x}_2$ is standardized:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

The rest is the same as the two-sided test concerning single mean:

Reject H_0 if $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$ otherwise don't reject H_0

For one-sided tests, we simply use one-tailed critical regions, as before.

Two Samples Tests on Means

Unknown but Equal Variances Case: Consider the two-sided test.

Two-Sample Pooled t-Test

For the two-sided hypothesis

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 \neq \mu_2,$$

we reject H_0 at significance level α when the computed t-statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}},$$

where

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

exceeds $t_{\alpha/2, n_1+n_2-2}$ or is less than $-t_{\alpha/2, n_1+n_2-2}$.

Two Samples Tests on Means

Unknown but Equal Variances Case: Consider the **two-sided test.**

Two-Sample Pooled t-Test

For the two-sided hypothesis

$$H_0\colon\ \mu_1 = \mu_2,$$
$$H_1\colon\ \mu_1 \neq \mu_2,$$

we reject H_0 at significance level α when the computed t-statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}},$$

where

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

exceeds $t_{\alpha/2, n_1+n_2-2}$ or is less than $-t_{\alpha/2, n_1+n_2-2}$.

s_p^2 for estimating variance: weighted average of sample variances, where weights are the degrees of freedom of the samples.

s_p: pooled standard deviation. It estimates the unknown population standard deviation

Other Hypothesis Testing Cases

There are many other types of tests we do not cover in detail such as,

Two samples test on means (unknown and unequal variance case)

Two samples test on means (with paired observations)

One sample test concerning variance

…

The background we provided here and the conceptual similarity between testing and the confidence intervals should allow you to learn these with independent study, if there is such a need.