Quantitative Methods – II

A.Y. 2021-22

Practice 11

Lorenzo Cavallo

For any clarification/meeting: cavallo@istat.it

THEME #1

—

Hypothesis Testing

Null Hypothesis (H_o)

A null hypothesis is a claim (or statement) about a population parameter that is assumed to be true until it is declared false.

Alternative Hypothesis (H_1)

An alternative hypothesis is a claim about a population parameter that will be true if the null hypothesis is false.

Two Types of Errors:

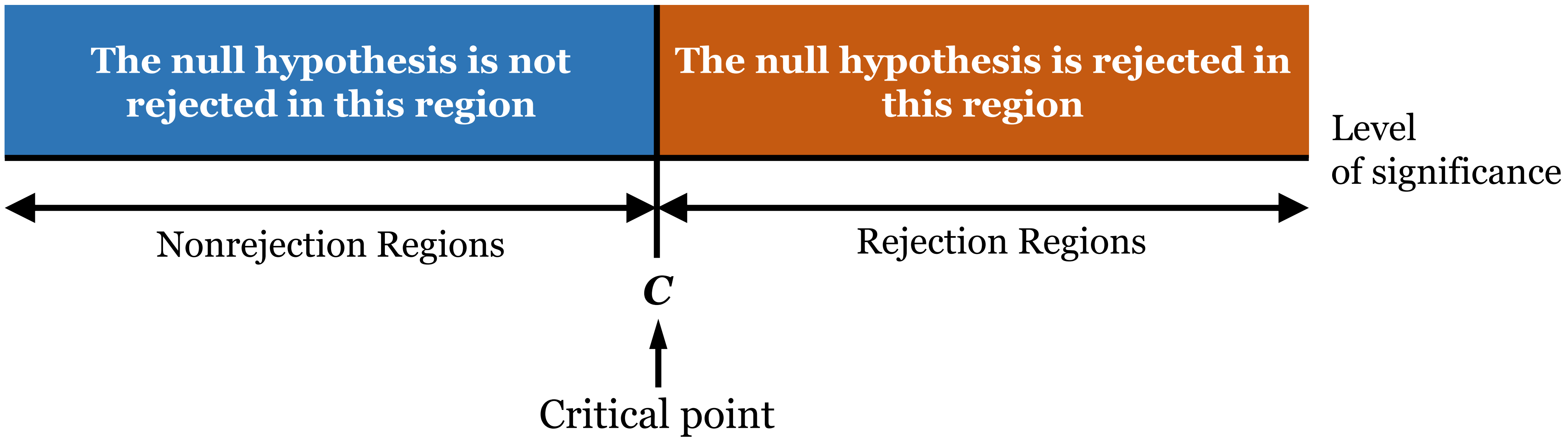
1. **Type I or α error**. A Type I error occurs when a **true** null hypothesis is **rejected**. The value of **α** represents the probability of committing this type of error
2. **Type II or β error.** A Type II error occurs when a **false** null hypothesis is **not rejected**. The value of **β** represents the probability of committing this type of error.

The value **1 – β** is called the **power of the test**.

Two approach to the Hypothesis Test decision:

Critical point: the value(s) from a table (such as the normal distribution table or the t distribution table) to compare with the value of the test statistic for the observed value of the sample statistic.

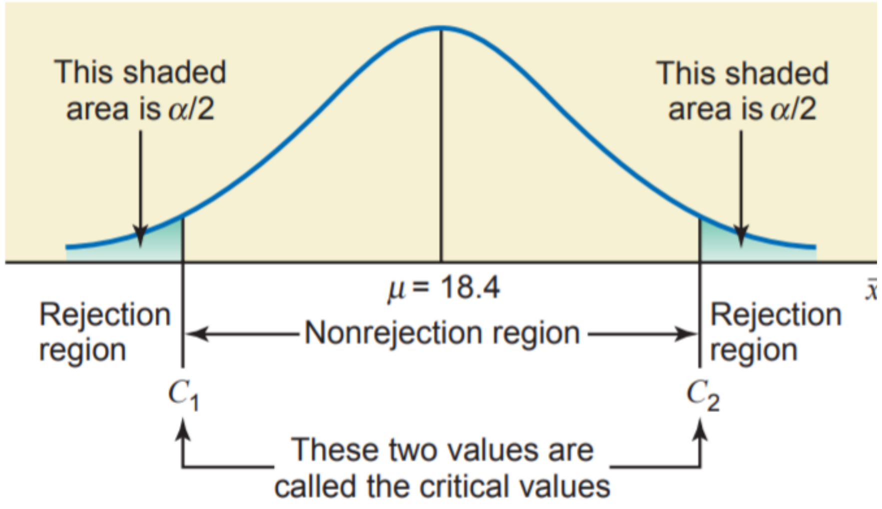
p-Value: the smallest level of significance at which the given null hypothesis is rejected.

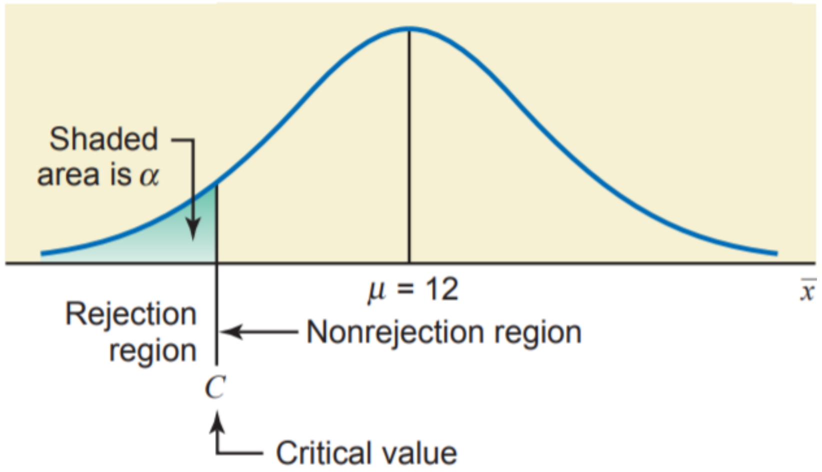


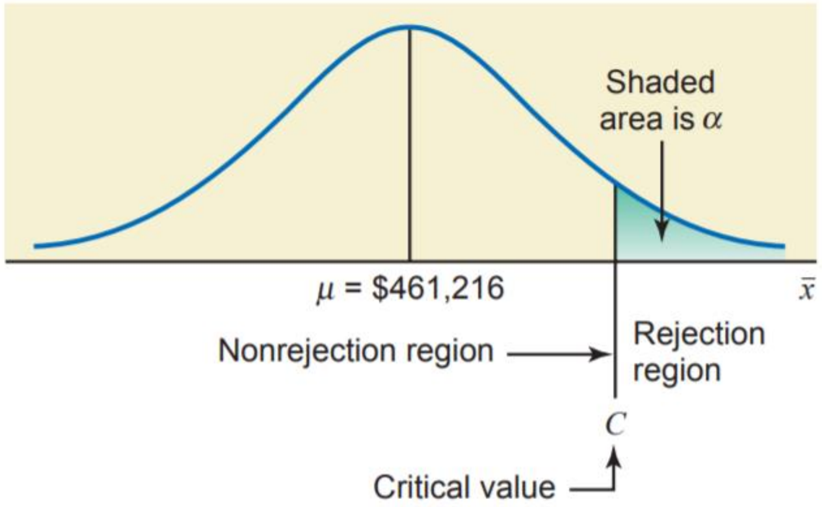
A **two-tailed test** has rejection regions in both tails of the distribution curve.

A **left-tailed test** has the rejection region in the left tail of the distribution curve.

A **right-tailed test** has the rejection region in the right tail of the distribution curve.







A **two-tailed test** has rejection regions in both tails of the distribution curve.

A **left-tailed test** has the rejection region in the left tail of the distribution curve.

A **right-tailed test** has the rejection region in the right tail of the distribution curve.

	Two-Tailed Test	**Left-Tailed Test**	**Right-Tailed Test**
Sign in the null hypothesis H_0	$=$	$=$ or $\geq$	$=$ or $\leq$
Sign in the alternative hypothesis H_1	$\neq$	$<$	$>$
Rejection region	In both tails	In the left tail	In the right tail

IMPORTANT:

We **reject** or **not reject** the Null Hyphotesis

But **NEVER**

We accept the Null Hyphotesis

Or

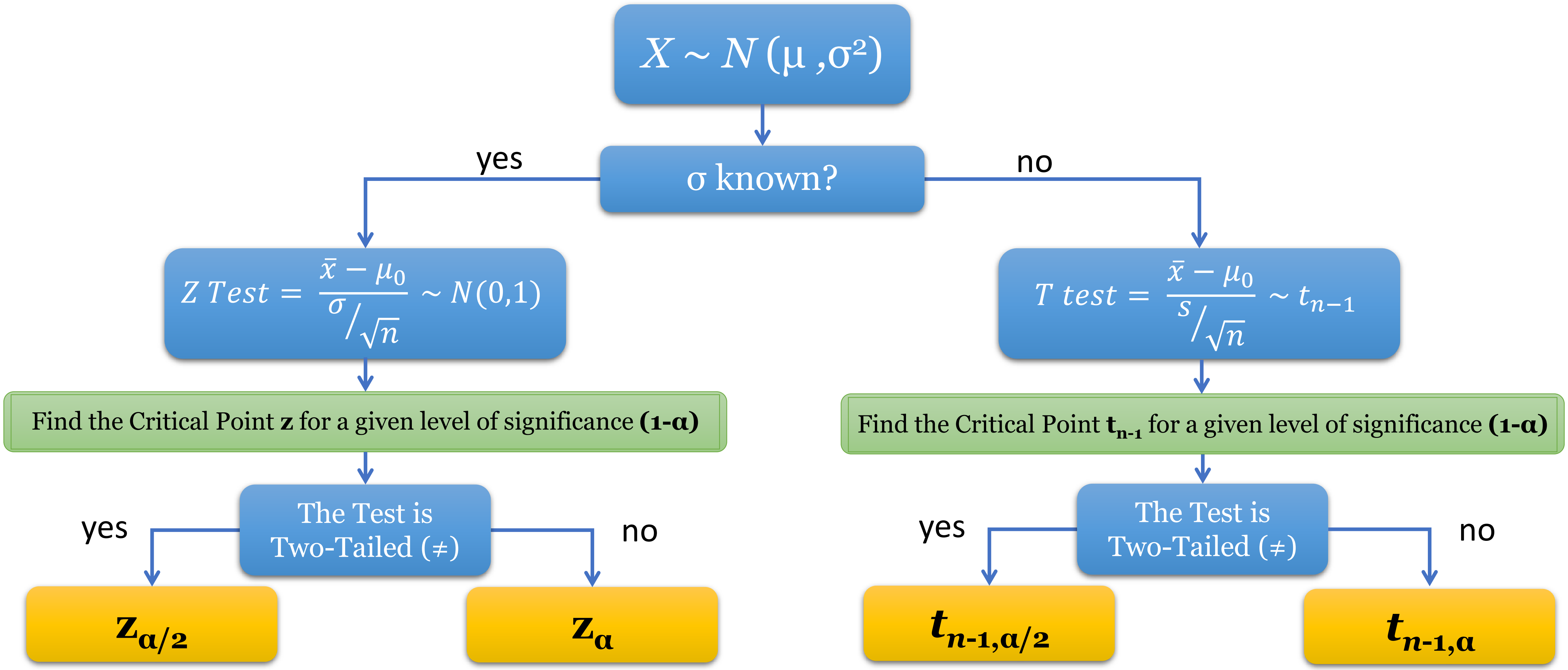
We accept or reject the Alternative Hyphotesis

THEME #2

—

The Critical region approach

Hyphotesis on the mean



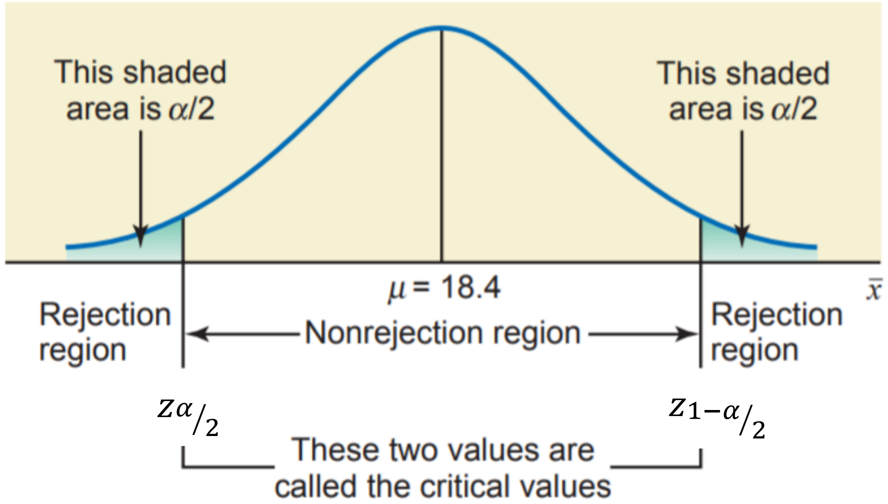
Two-Tailed Test

σ known

Calculate $|z| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right|$ and compare with $|z_{\alpha/2}|$

If $|z|>|z_{\alpha/2}|$ reject H_o

If $|z|\leq|z_{\alpha/2}|$ do not reject H_o

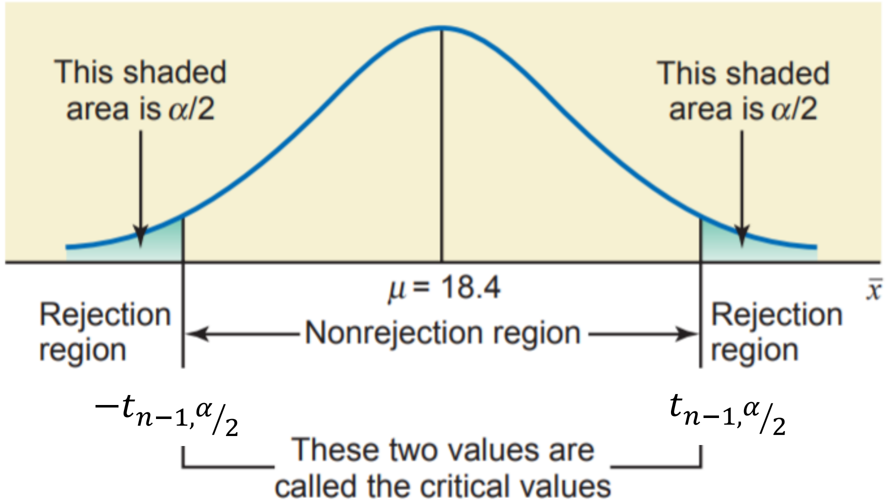


σ unknown

Calculate $|t| = \left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|$ and compare with $|t_{n-1,\alpha/2}|$

If $|t|>|t_{n-1,\alpha/2}|$ reject H_o

If $|t|\leq|t_{n-1,\alpha/2}|$ do not reject H_o



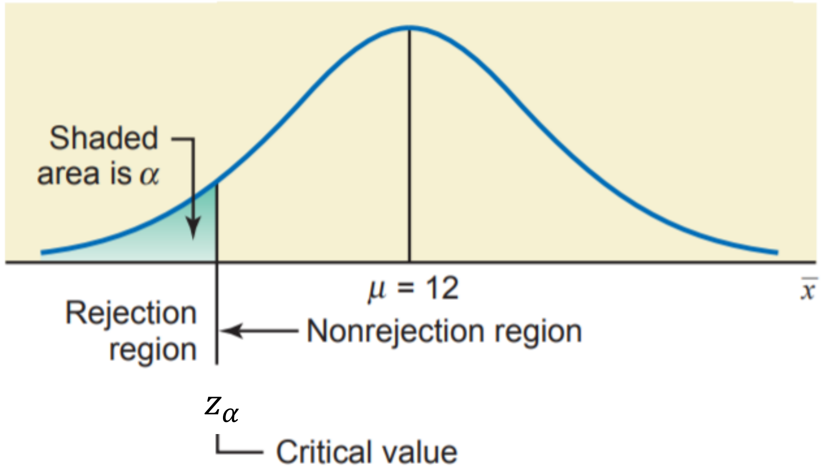
Left-Tailed Test

σ known

Calculate $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ and compare with z_α

If $z < z_\alpha$ reject H_o

If $z \geq z_\alpha$ do not reject H_o

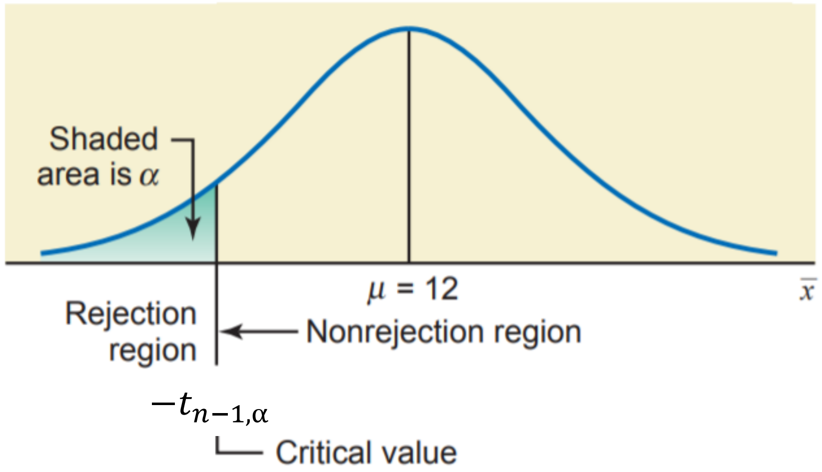


σ unknown

Calculate $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ and compare with $-t_{n-1,\alpha}$

If $t < -t_{n-1,\alpha}$ reject H_o

If $t \geq -t_{n-1,\alpha}$ do not reject H_o



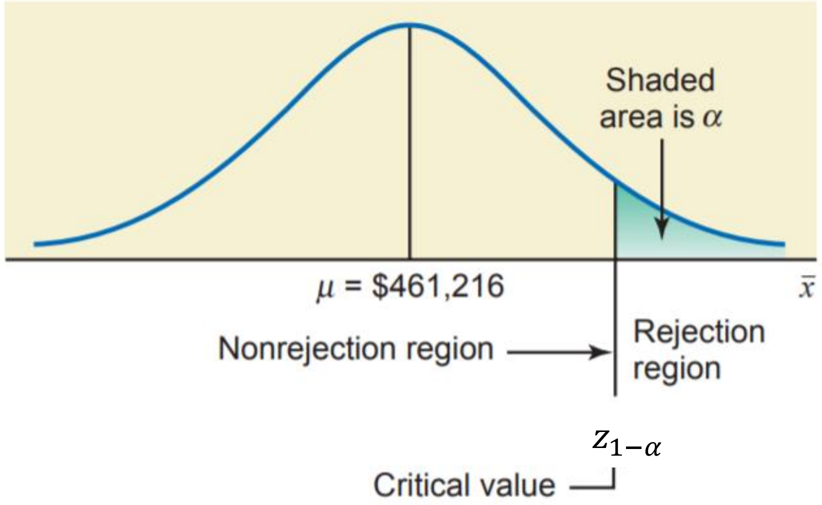
Right-Tailed Test

σ known

Calculate $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ and compare with $z_{1-\alpha}$

If $z > z_{1-\alpha}$ reject H_o

If $z \le z_{1-\alpha}$ do not reject H_o

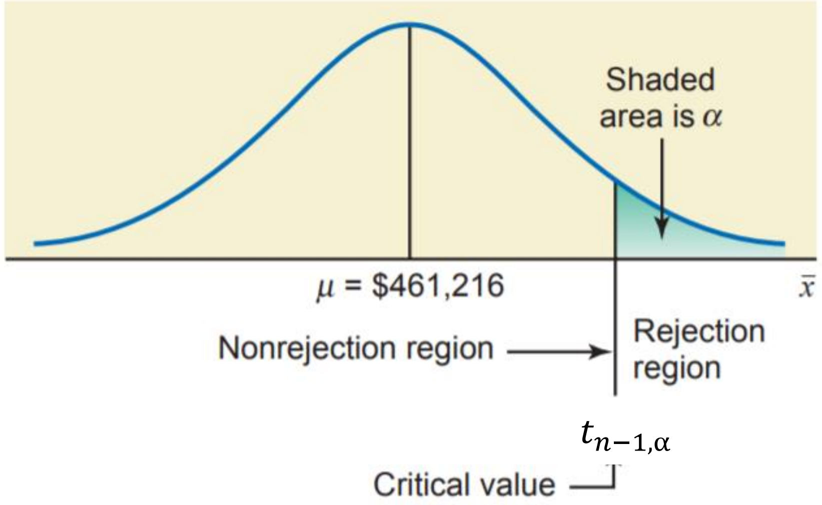


σ unknown

Calculate $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ and compare with $t_{n-1,\alpha}$

If $t > t_{n-1,\alpha}$ reject H_o

If $t \le t_{n-1,\alpha}$ do not reject H_o



Exercise c.1

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 33. Assuming that the population is normally distributed with variance 20:

a. Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$).
b. Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).
c. Can we conclude that the mean is higher than 30 years? ($1-\alpha=0.99$).

Solution

a. **Data:**

Variable X is the mean age of a certain population

$n=10$, $\bar{x}=33$, $\sigma^2=20$, $1-\alpha=0.95$ ($\alpha=0.05$)

Assumptions:

The population is normally distributed with variance 20

Hypothesis:

Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$)

H_0: $\mu_0=30$

H_1: $\mu_0\neq30$

Exercise c.1

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 33. Assuming that the population is normally distributed with variance 20:

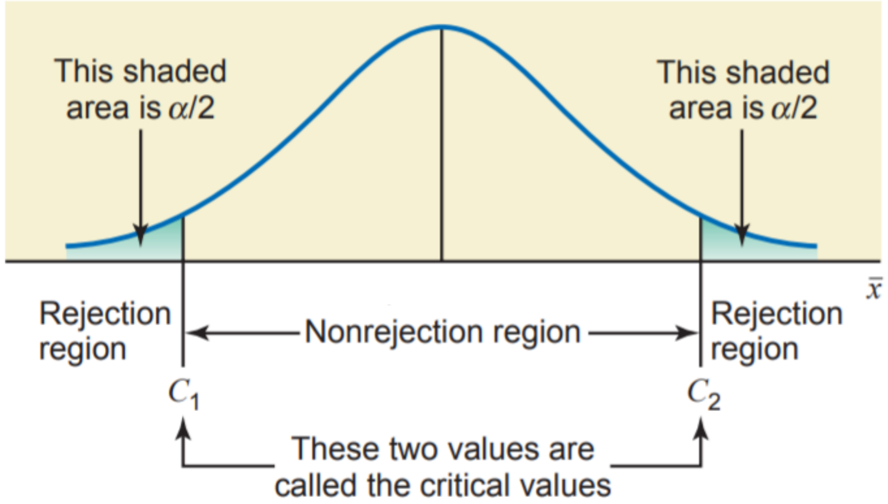
a. Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$).
b. Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).
c. Can we conclude that the mean is higher than 30 years? ($1-\alpha=0.99$).

Solution

Hypothesis:

H_0: $\mu_0=30$

H_1: $\mu_0\neq30$



Test Statistic:

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{33-30}{4.47/3.16} = \frac{3}{1.414} = 2.12$$

Decision Rule:

The alternative hypothesis is H_1: $\mu_0 \neq 30$ so the test is two-tailed test.

Hence we reject H_0 if: $|z| = 2.12 > |z_{\alpha/2}| = |z_{0.025}| = 1.96$ from table of N(0,1).

The rejection regions are $(-\infty; -1.96]$ and from $[1.96; +\infty)$

The test is TRUE, the z falls in the rejection region of the test than we reject H_0 and conclude that the mean is different from 30 years (at the level of evidence of 95%).

Exercise c.1

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 33. Assuming that the population is normally distributed with variance 20:

a. Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$).
b. Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).
c. Can we conclude that the mean is higher than 30 years? ($1-\alpha=0.99$).

Solution

b. **Data:**

Variable X is the mean age of a certain population

$n=10$, $\bar{x}=33$, $\sigma^2=20$, $1-\alpha=0.95$ ($\alpha=0.05$)

Assumptions:

The population is normally distributed with variance 20

Hypothesis:

Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).

H_0: $\mu_0=30$

H_1: $\mu_0<30$

Exercise c.1

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 33. Assuming that the population is normally distributed with variance 20:

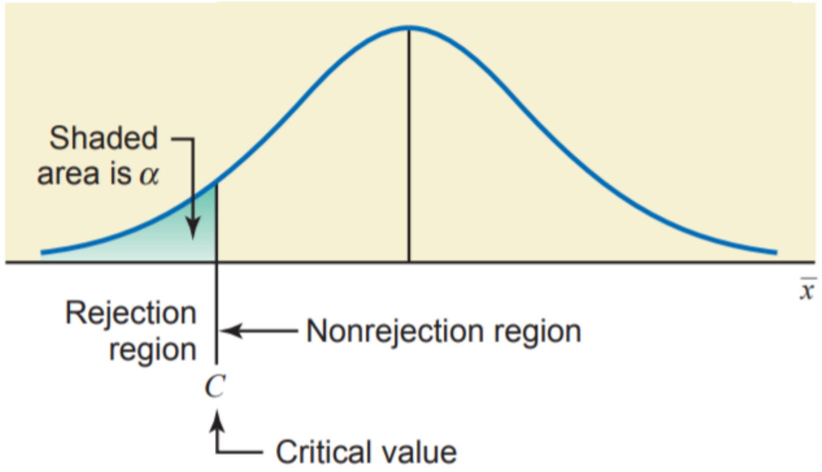
a. Can we conclude that the mean is different from 30 years ? (1-α=0.95).
b. Can we conclude that the mean is lower than 30 years ? (1-α=0.95).
c. Can we conclude that the mean is higher than 30 years? (1-α=0.99).

Solution

Hypothesis:

H_0: μ_0=30

H_1: μ_0<30



Test Statistic:

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{33-30}{4.47/3.16} = \frac{3}{1.414} = 2.12$$

Decision Rule:

The alternative hypothesis is H_1: $\mu_0 < 30$ (left-tailed test)

Hence we reject H_0 if: $z = 2.12 < z_\alpha = z_{0.05} = -1.645$ Critical Value from table of N(0,1)

The test is NOT TRUE, the z falls in the NON rejection region of the test (on the right of the Critical Value), so we do NOT reject the null hypothesis and conclude that the mean is NOT lower than 30 years (at the level of evidence of 95%).

Exercise c.1

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 33. Assuming that the population is normally distributed with variance 20:

a. Can we conclude that the mean is different from 30 years ? (1-α=0.95).
b. Can we conclude that the mean is lower than 30 years ? (1-α=0.95).
c. Can we conclude that the mean is higher than 30 years? (1-α=0.99).

Solution

c. **Data:**

Variable X is the mean age of a certain population

$n=10$, $\bar{x}=33$, $\sigma^2=20$, $1-\alpha=0.99$ ($\alpha=0.01$)

Assumptions:

The population is normally distributed with variance 20

Hypothesis:

Can we conclude that the mean is higher than 30 years? (1-α=0.99).

H_0: $\mu_0=30$

H_1: $\mu_0>30$

Exercise c.1

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 33. Assuming that the population is normally distributed with variance 20:

a. Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$).
b. Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).
c. Can we conclude that the mean is higher than 30 years? ($1-\alpha=0.99$).

Solution

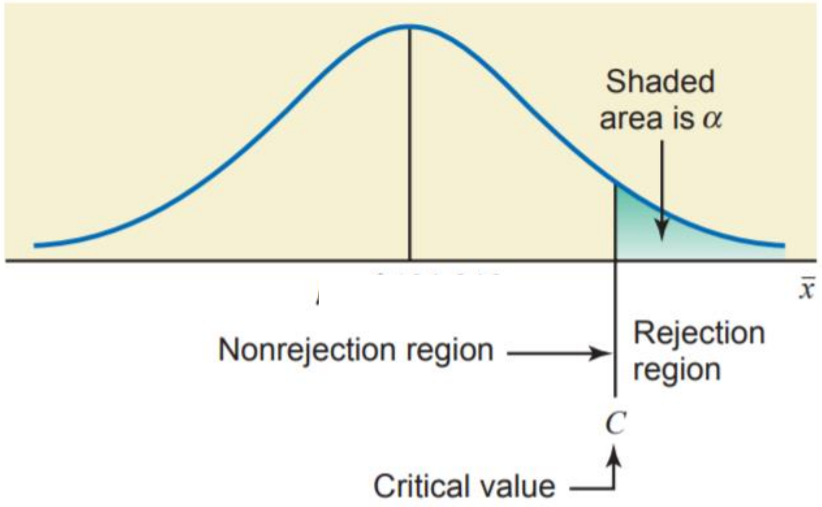
Hypothesis:

H_0: $\mu_0=30$

H_1: $\mu_0>30$

Test Statistic:

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{33-30}{4.47/3.16} = \frac{3}{1.414} = 2.12$$



Decision Rule:

The alternative hypothesis is H_1: $\mu_0 > 30$ (right-tailed test)

Hence we reject H_0 if: $z = 2.12 > z_{1-\alpha} = z_{0.99} = 2.33$ Critical value from table of N(0,1)

The test is TRUE and the z falls in the NON rejection region of the test (the z is on the left of the Critical Value), so we do NOT reject the null hypothesis and conclude that the mean is NOT higher than 30 years (at the level of evidence of 99%).

Exercise c.2

Test the following hypothesis for α=0.05 and 0.01.

a. H_0: μ=23, H_1: μ≠23, n=50, $\bar{x}$ =24.75, σ=5,

b. H_0: μ=15, H_1: μ>15, n=80, $\bar{x}$ =16.75, σ=5.5

c. H_0: μ=38, H_1: μ<38, n=35, $\bar{x}$ =40.25, σ=7.2

Solution

a. It's a two tailed test.

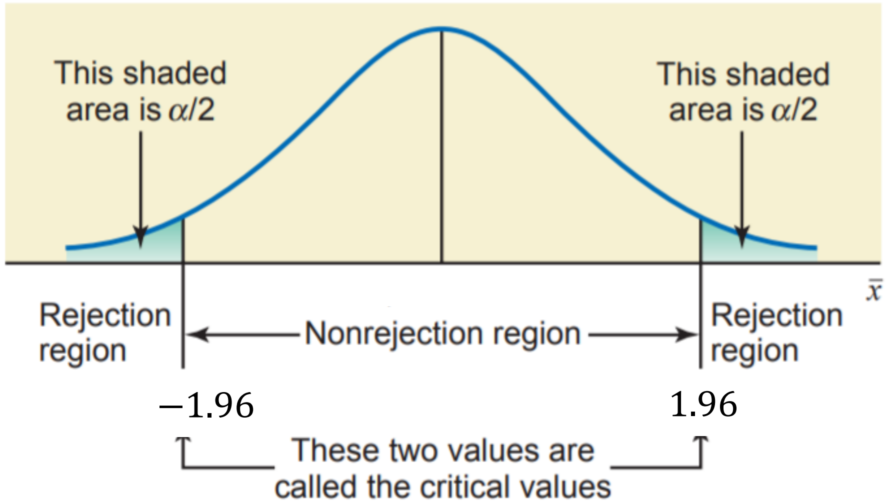
The rejection regions for α=0.05 are (-∞; -1.96] and from [1.96; +∞)

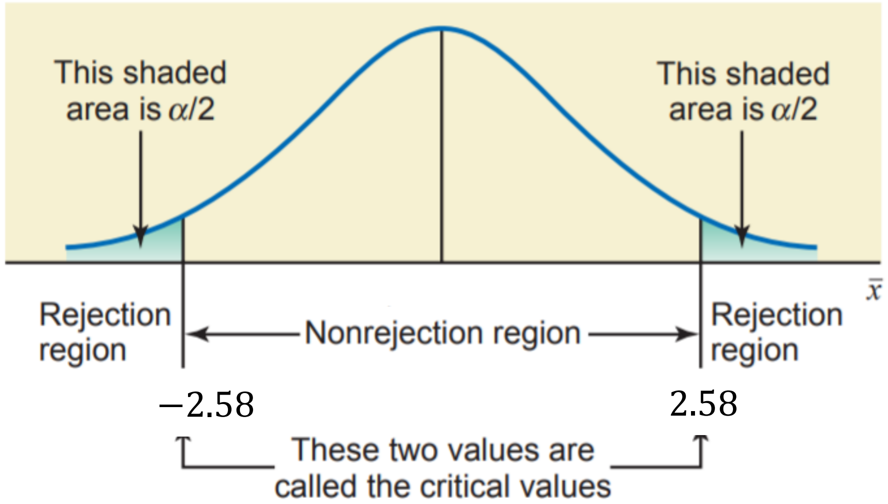
The rejection regions for α=0.01 are (-∞; -2.58] and from [2.58; +∞)

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = 2.47$$

z falls is in the rejection region for α=0.05 and in the NOT rejection region for α =0.01

We reject H_0 for α=0.05 (μ≠23) and do NOT reject H_0 for α =0.01 (μ=23)





Exercise c.2

Test the following hypothesis for α=0.05 and 0.01.

a. H_0: μ=23, H_1: μ≠23, n=50, $\bar{x}$ =24.75, σ=5,

b. H_0: μ=15, H_1: μ>15, n=80, $\bar{x}$ =16.75, σ=5.5

c. H_0: μ=38, H_1: μ<38, n=35, $\bar{x}$ =40.25, σ=7.2

Solution

b. It's a right tailed test.

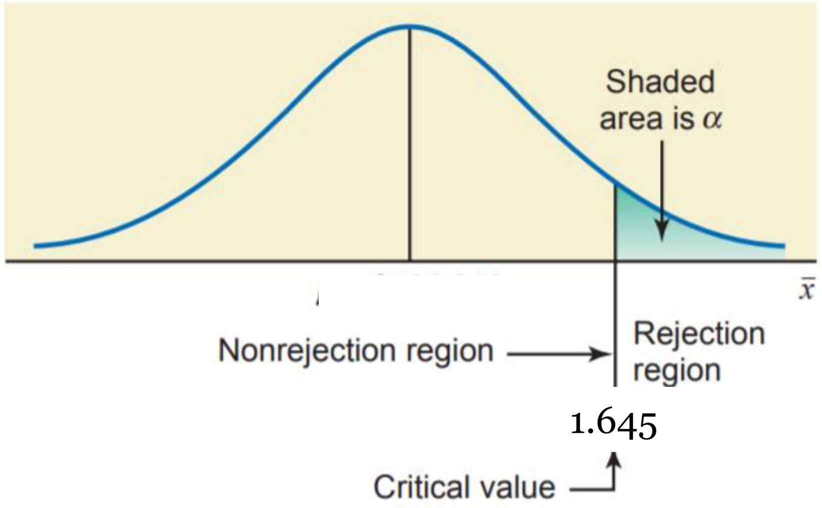
The rejection region for α=0.05 is [1.645; +∞)

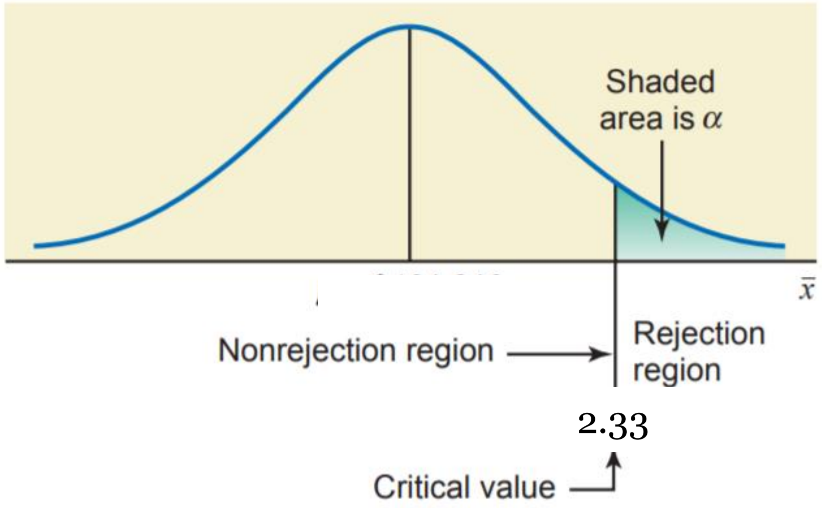
The rejection region for α=0.01 is [2.33; +∞)

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = 2.85$$

z falls is in the rejection region for α=0.05 and for α =0.01

We reject H_0 (so, μ>15) for α=0.05 and for α =0.01





Exercise c.2

Test the following hypothesis for α=0.05 and 0.01.

a. H_0: μ=23, H_1: μ≠23, n=50, $\bar{x}$ =24.75, σ=5,

b. H_0: μ=15, H_1: μ>15, n=80, $\bar{x}$ =16.75, σ=5.5

c. H_0: μ=38, H_1: μ<38, n=35, $\bar{x}$ =40.25, σ=7.2

Solution

b. It's a left tailed test.

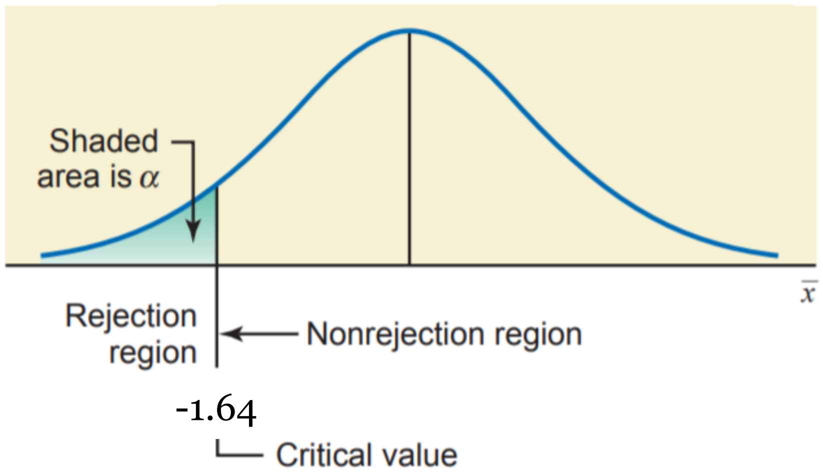
The rejection region for α=0.05 is (-∞ ; -1.645]

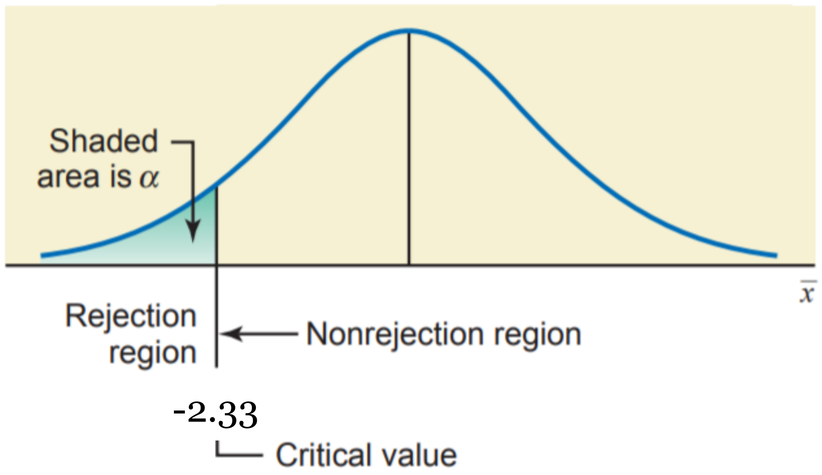
The rejection region for α=0.01 is (-∞ ; -2.33]

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = 1.84$$

z falls is in the NOT rejection region for α=0.05 and for α =0.01

We do NOT reject H_0 for α=0.05 and α =0.01 (μ=38)





Exercise c.3

A random sample of 18 observations produced a sample mean of 9.24. Find the critical and observed values of z for each of the following tests of hypothesis using α=0.05. The population standard deviation is known to be 5.40 and the population distribution is normal.

a. $H_0 : \mu_0 = 8.5$ versus H1: $\mu_0 \neq 8.5$
b. $H_0 : \mu_0 = 8.5$ versus H1: $\mu_0 > 8.5$

Solution

a. Two tailed test

Observed value of z is = $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ = 0.74/1.273 = 0.58

Critical values of z is ± 1.96

The rejection regions for α=0.05 are (-∞; -1.96] and from [1.96; +∞)

Observed value of z is in the NON rejection region of the test, then do NOT reject H_0

b. Right tailed test

Observed value of z is = $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ = 0.74/1.273 = 0.58

Critical value of z is 1.64

The rejection region for α=0.05 is [1.645; +∞)

Observed value of z is in the NON rejection region of the test (z<Critical Value), then do NOT reject H_0

THEME #3

—

The p-value

p-Value approach

σ known

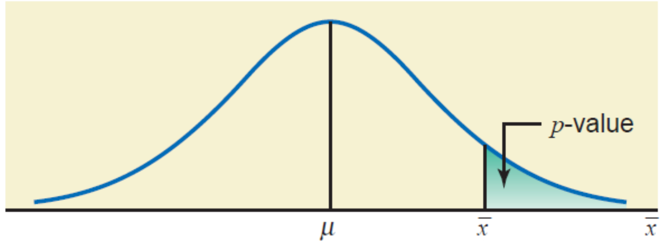
Calculate $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

σ unknown

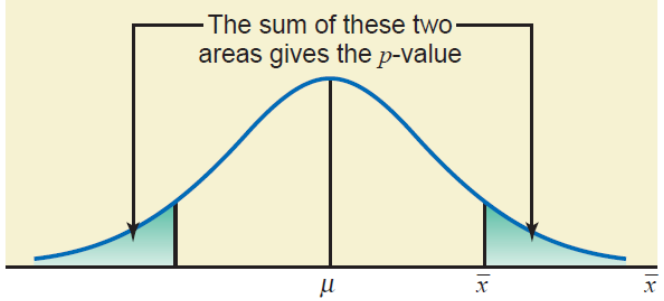
Calculate $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

From the Table (of Normal Standard or Student's T) find the corresponding area:

- For a one-tailed test, the **p-value** is given by the **area in the tail** of the sampling distribution curve beyond the observed value of the sample statistic.



- For a two-tailed test, the **p-value** is <u>**twice**</u> the **area in the tail** of the sampling distribution curve beyond the observed value of the sample statistic.

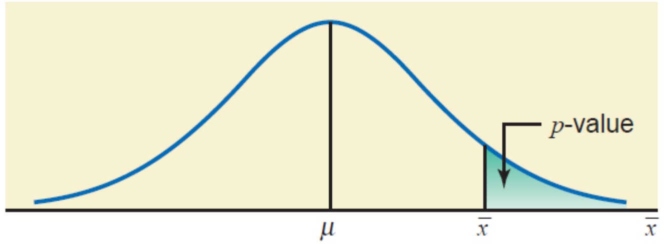


Using the p-value approach, we **reject** the **null hypothesis** if: **p-value < α**
and we **do not reject** the **null hypothesis** if: **p-value ≥ α**

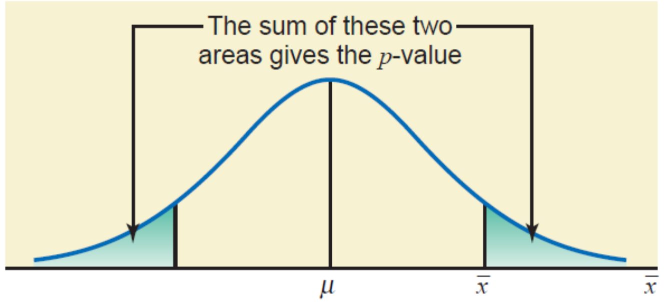
Where **α is the level of confidence given in the exercise.**

p-Value approach

- In a one-tailed test, the **p-value** is given by the **area in the tail** of the sampling distribution curve beyond the observed value of the sample statistic.



- For a two-tailed test, the **p-value** is <u>**twice**</u> the **area in the tail** of the sampling distribution curve beyond the observed value of the sample statistic.



If you **do not have** the table of the Standard Normal Distribution for the negative values of **z** to find the p-value:

for a one-tailed test the p-value is: **1-α** where α is shaded area for |z|

for a two-tailed test the p-value is: **2 · (1-α)** where α is shaded area for |z|

*If you have the table of the Standard Normal Distribution for the negative values of **z** the p-value in a left-tailed test for a negative z is the value in the table.*

How to use the p-Value approach in hypothesis testing

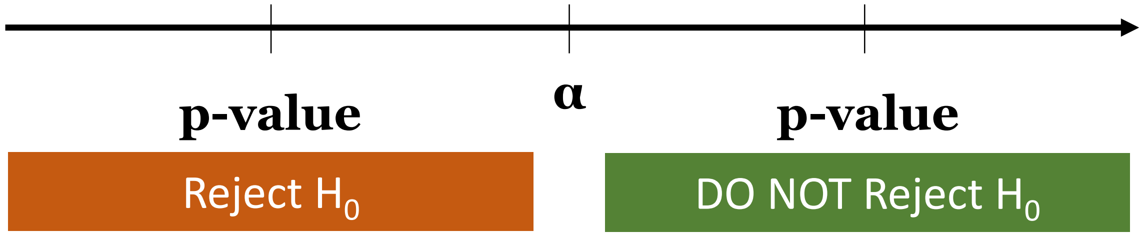
Using the p-value approach, we **reject** the **null hypothesis** if:

$$\textbf{p-value} < \alpha$$

and we **do not reject** the **null hypothesis** if:

$$\textbf{p-value} \geq \alpha$$

where **α is the level of confidence given in the exercise.**



Example (Exercise c.8)

Find the p-value for each of the following hypothesis tests.

a. $H_0: \mu = 23, \quad H_1: \mu \neq 23, \quad n = 50, \quad \bar{x} = 21.25, \quad \sigma = 5$

b. $H_0: \mu = 15, \quad H_1: \mu < 15, \quad n = 80, \quad \bar{x} = 13.25, \quad \sigma = 5.5$

z	.00	.01	.02	.03	.04	.05	.06	.07
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932

Solution

a.

$$z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{21.25-23}{\frac{5}{\sqrt{50}}} = \frac{-1.75}{\frac{5}{7.0711}} = \frac{-1.75}{0.707} = -2.47$$

$|z|=2.47$

So, $\alpha = 0.9932$

and $(1-\alpha) = 1 - 0.9932 = 0.0068$

It's a Two-tailed test and for this reason the p-Value is twice $2 \cdot (1-\alpha)$

p-value $= 2 \cdot (1-\alpha) = 2 \cdot 0.0068 \sim 0.014$

Example (Exercise c.8)

Find the p-value for each of the following hypothesis tests.

a. $H_0: \mu = 23, \quad H_1: \mu \neq 23, \quad n = 50, \quad \bar{x} = 21.25, \quad \sigma = 5$

b. $H_0: \mu = 15, \quad H_1: \mu < 15, \quad n = 80, \quad \bar{x} = 13.25, \quad \sigma = 5.5$

z	.00	.01	.02	.03	.04	.05	.06	.07
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979

Solution

b.

$$z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{13.25-15}{\frac{5.5}{\sqrt{80}}} = \frac{-1.75}{\frac{5.5}{8.94}} = -2.85$$

$|z|=2.85$

$\alpha = 0.9978$

and $(1-\alpha)=0.0022$

It's a One-tailed test and for this reason the p-Value is $(1-\alpha)$

p-value = 0.0022

Exercise c.9

Find the p-value for each of the following hypothesis tests.

a. H_0: $\mu = 46$, H_1: $\mu \neq 46$, $n = 40$, $\bar{x} = 49.60$, $\sigma = 9.7$
b. H_0: $\mu = 26$, H_1: $\mu < 26$, $n = 33$, $\bar{x} = 24.30$, $\sigma = 4.3$
c. H_0: $\mu = 18$, H_1: $\mu > 18$, $n = 55$, $\bar{x} = 20.50$, $\sigma = 7.8$

Solution c.9

a. $|z|=2.35$, $\alpha = 0.9906$ and $(1-\alpha) = 0.0094$, it is a two tailed test so, p-value = $2 \cdot (1-\alpha) = 2 \cdot 0.0094 = 0.0188$
b. $|z|=2.27$, $\alpha = 0.9884$ and $(1-\alpha) = 0.0116$, it is a one-tailed test so, p-value = $(1-\alpha) = 0.0116$
c. $|z|=2.37$, $\alpha = 0.9913$ and $(1-\alpha) = 0.0087$, it is a one-tailed test so, p-value = $(1-\alpha) = 0.0087$

Exercise c.10

Consider H_0 : $\mu=72$ versus H_1: $\mu> 72$. A random sample of 16 observations taken from this population produced a sample mean of 75.2. The population is normally distributed with a $\sigma=6$.
Calculate the p-value.

Solution c.10

$|z|=2.13$, , $\alpha=0.9836$, it is a one-tailed test so, p-value = $(1- \alpha) = 0.0164$

THEME #4

—

Hyphotesis Testing using the Critical value and the p-value

Example (Exercise c.1)

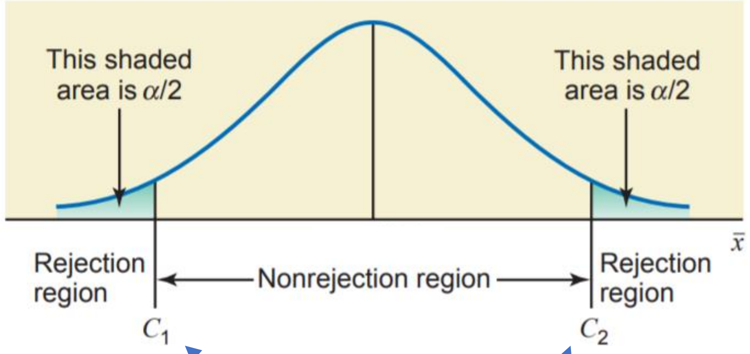
Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 27. Assuming that the population is approximately normally distributed with variance 20:

a. Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$).

b. Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).

Solution (point a.)

Data: variable is age $n=10$, $\bar{x}=27$, $\sigma^2=20$, $1-\alpha=0.95$ ($\alpha=0.05$)



Hypotheses: H_0: $\mu=30$ versus H_1: $\mu\neq30$

Decision Rule: The alternative hypothesis is H_1: $\mu \neq 30$ (Two-tailed Test)

Test Statistic: $|z| = \left|\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right| = \left|\frac{27-30}{\frac{\sqrt{20}}{\sqrt{10}}}\right| = \left|\frac{-3}{\frac{4.47}{3.16}}\right| = \frac{3}{1.414} = 2.12$

The p-Value for $|z|=2.12$ in a two-tailed is $2 \bullet (1-\alpha)$ where $\alpha = 0.9830$:

So, **p-Value** = $2 \bullet (1-0.9830) = 2 \bullet 0.0170 =$ **0.0340**

Critical Values: $z_{\alpha/2} = z_{0.025} = \pm1.96$ (from table of $N(0, 1)$) Hence, the Non-rejection region is $[-1.96; +1.96]$

$z = 2.12 > z_{\alpha/2} = 1.96$ and p-Value $= 0.0340 < \alpha = 0.05$

Hence, we reject H_0. We conclude that the mean is different from 30 years (at the 95% evidence level)

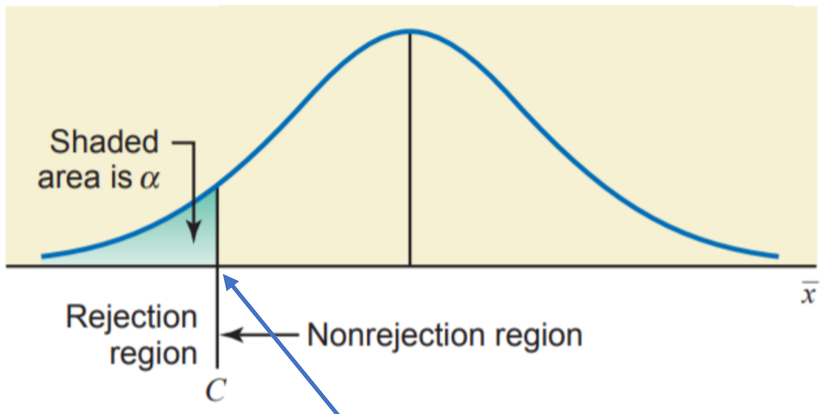
The p-Value gives us the same information

Example (Exercise c.1)

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 27. Assuming that the population is approximately normally distributed with variance 20:

a. Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$).

b. Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).



Solution (point b.)

Data: variable is age $n=10$, $\bar{x}=27$, $\sigma^2=20$, $1-\alpha=0.95$ ($\alpha=0.05$)

Hypotheses: H_0: $\mu=30$ versus H_1: $\mu<30$

Decision Rule: The alternative hypothesis is H_1: $\mu < 30$ (Left-tailed Test)

Critical Value: $z_\alpha = z_{0.05} = -1{,}645$ (from table of N(0 , 1)) Hence, the Non-rejection region is $[-1{,}645; +\infty]$

Test Statistic: $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{27-30}{\frac{\sqrt{20}}{\sqrt{10}}} = \frac{-3}{\frac{4.47}{3.16}} = \frac{-3}{1.414} = -2.12$

For the p-Value in a one tailed test we have to find the value $(1-\alpha)$ for $|z|=2.12$ from the Z table, so: **p-Value** = $(1-\alpha)=0.0170$

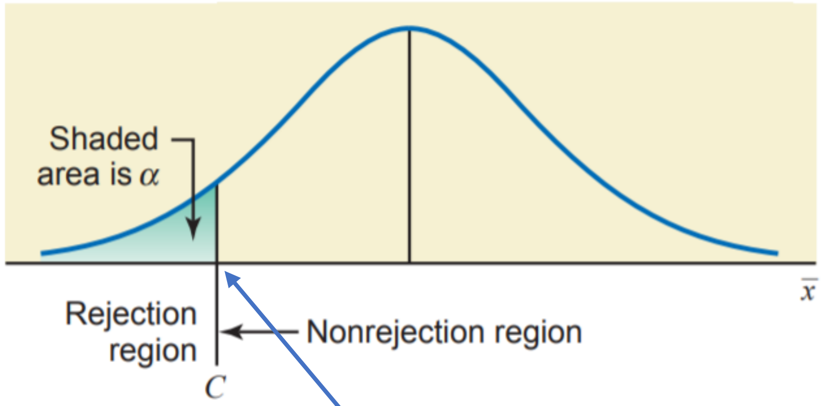
$z = -2.12 < z_\alpha = -1.645$ or p-Value $= 0.0170 < \alpha = 0.05$

Hence, we reject H_0. We conclude that the mean is lower than 30 years (at the 95% evidence level)

Example (Exercise c.1)

Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 27. Assuming that the population is approximately normally distributed with variance 20:

a. Can we conclude that the mean is different from 30 years ? ($1-\alpha=0.95$).
b. Can we conclude that the mean is lower than 30 years ? ($1-\alpha=0.95$).
c. Can we conclude that the mean is lower than 30 years? ($1-\alpha=0.99$).



Solution (point c.)

Data: variable is age $n=10$, $\bar{x}=27$, $\sigma^2=20$, $1-\alpha=0.99$ ($\alpha=0.01$)

Hypotheses: H_0: $\mu=30$ versus H_1: $\mu<30$

Decision Rule: The alternative hypothesis is H_1: $\mu < 30$ (Left-tailed Test)

Critical Value: $z_\alpha = z_{0.01}$ =-2,33 (from table of N(0 , 1)) Hence, the Non-rejection region is [+∞ ; -2,33]

Test Statistic: $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{27-30}{\frac{\sqrt{20}}{\sqrt{10}}} = \frac{-3}{\frac{4.47}{3.16}} = \frac{-3}{1.414} = -2.12$

The p-Value is the same of the previous point: **p-Value** = 0.0170

$z = -2.12 > z_\alpha = -2.33$ or p-Value = 0.0170 > α = 0.01

Hence, we DO NOT reject H_0. We conclude that the mean is equal to 30 years (at the 99% evidence level)

THEME #5

—

Hyphotesis Testing with σ unknown

Exercise c.5

Among 30 Italian men, the mean systolic blood pressure was 146 mm/Hg with a standard deviation of 27. We wish to know if on the basis of these data, assuming that the population is normally distributed, we may conclude that the mean systolic blood pressure for a population of Italians is greater than 140. Use α=0.01.

Solution

Data: Variable is systolic blood pressure, n=30, $\bar{x}$ =146, $\hat{\sigma}$ =27, α=0.01.

Assumption: population is normal, σ is unknown

Hypotheses: H_0: μ=140 versus H_1: μ>140

Test Statistic: $t = \frac{\bar{x}-\mu_0}{\hat{\sigma}/\sqrt{n}} = \frac{146-140}{\frac{27}{\sqrt{30}}} = \frac{6}{\frac{27}{5.48}} = \frac{6}{4.93} = 1.22$

ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659

Decision Rule:
The test is Right-Tailed then we reject H_0 if $t > t_{n-1,\alpha} = t_{29,0.01} = 2.462$ (from table of Student's T)

Decision: t is **not greater** than the critical value so, we **do not reject** H_0.
We may conclude that the mean systolic blood pressure for a population of Italian is not greater than 140.

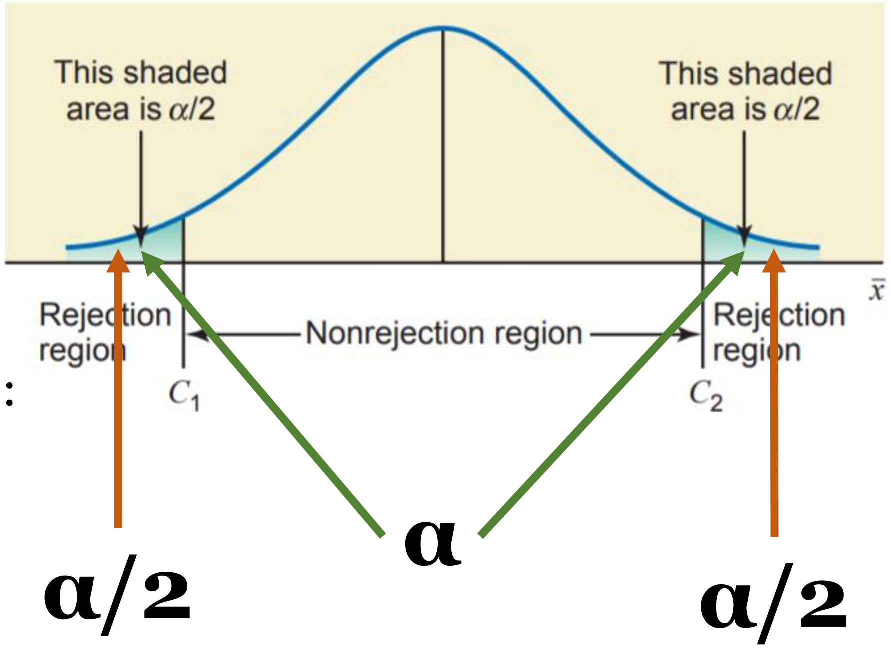
Hypothesis on the mean

		Two-Tailed Test	Left-Tailed Test	Right-Tailed Test
Null hypothesis H_o		$\mu = \mu_o$	$\mu =$ or $\geq \mu_o$	$\mu =$ or $\leq \mu_o$
Alternative hypotesis H_1		$\mu \neq \mu_o$	$\mu < \mu_o$	$\mu > \mu_o$
Critical point	σ known	$z_{\alpha/2}$ and $z_{1-\alpha/2}$	z_{α}	$z_{1-\alpha}$
	σ not known	$t_{n-1, \alpha/2}$ and $-t_{n-1, \alpha/2}$	$-t_{n-1, \alpha}$	$t_{n-1, \alpha}$

Hypothesis on proportion

	Two-Tailed Test	Left-Tailed Test	Right-Tailed Test
Null hypothesis H_o	$p = p_o$	$p =$ or $\geq p_o$	$p =$ or $\leq p_o$
Alternative hypotesis H_1	$p \neq p_o$	$p < \mu_o$	$p > p_o$
Critical point	$z_{\alpha/2}$ and $z_{1-\alpha/2}$	z_{α}	$z_{1-\alpha}$

How to find the critical value(s)?



<u>Two tailed test</u>

For a two tailed test with a level of significance (1- α) we have to find the critical point that left in the two tail of the distribution an area of α

The total area is α so the two area in the two tails are α/2: α/2 in the right tail and α/2 in the left tail

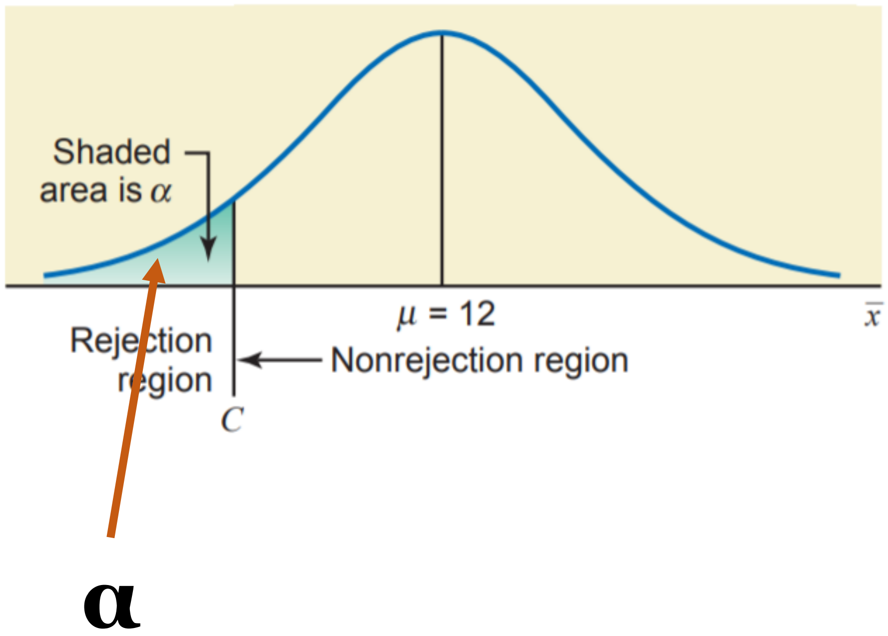
Common critical values for a two tailed test:

(1- α)=0.95 (95%) the Critical values are:
-1.96 (left tail) and **1.96** (right tail)
(in the Z tables you have to find the z that left on the two tails α/2=0.025)

(1- α)=0.90 (90%) the Critical values are:
-1.645 (left tail) and **1.645** (right tail)
(in the Z tables you have to find the z that left on the two tails α/2=0.05)

(1- α)=0.99 (99%) the Critical values are:
-2.58 (left tail) and **2.58** (right tail)
(in the Z tables you have to find the z that left on the two tails α/2=0.005)

How to find the critical value(s)?



<u>Left-tailed test</u>

For a left-tailed test with a level of significance (1- α) we have to find the critical point that left in the left tail of the distribution an area of α

The area on the left tail is α

Common critical values for a left-tailed test:

(1- α)=0.95 (95%) the Critical value is:
-1.645 (in the left tail)
(in the Z tables you have to find the z that left on the left tail α=0.05)

(1- α)=0.90 (90%) the Critical value is:
-1.28 (in the left tail)
(in the Z tables you have to find the z that left on the left tail α=0.1)

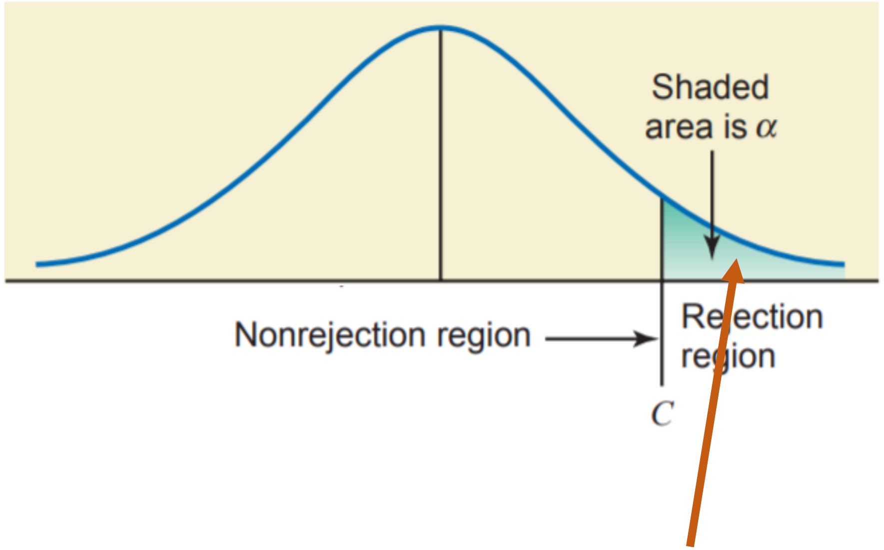
(1- α)=0.99 (99%) the Critical value is:
-2.33 (in the left tail)
(in the Z tables you have to find the z that left on the left tails α=0.01)

How to find the critical value(s)?

<u>Right-tailed test</u>

For a right-tailed test with a level of significance (1- α)
we have to find the critical point that left
in the right tail of the distribution an area of α

The area on the right tail is α



α

Common critical values for a right-tailed test:

(1- α)=0.95 (95%) the Critical value is:
1.645 (in the light tail)
(in the Z tables you have to find the z that right on the right tail α=0.05)

(1- α)=0.90 (90%) the Critical value is:
1.28 (in the right tail)
(in the Z tables you have to find the z that right on the right tail α=0.1)

(1- α)=0.99 (99%) the Critical value is:
2.33 (in the right tail)
(in the Z tables you have to find the z that right on the right tails α=0.01)

THEME #6

—

Hypothesis Testing on two means

Testing two means with independent samples

σ known and the same for the two means

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \cancel{(\mu_1 - \mu_2)}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If $(\mu_1 - \mu_2) = 0$

σ known and different for the two means σ_1 and σ_2

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \cancel{(\mu_1 - \mu_2)}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

σ unknown
(σ_1 and σ_2 unequal)

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \cancel{(\mu_1 - \mu_2)}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with df = $\dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$

σ unknown
(σ_1 and σ_2 equal) ***pooled t procedure*** (*df*=n_1+n_2-2)

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \cancel{(\mu_1 - \mu_2)}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

Exercise 1

A chocolate company has 2 kind of candies. The candy A has an expiring date of 3 days longer than B.
The quality control selects a simple random sample of 100 units of the product A and 100 of the product B and verifies that the average shelf-life of A is 15 days and 10 days for B's.
The company knows that the standard deviation of the shelf-life of the candy A is 3 days and 2 for the candy B.
Test the claim that the product A expire 3 days after the product B. Use a 0.95 level of significance.

Solution
State the hypotheses. The claim of the company is that the product A (μ_1) has an average expiring date 3 days longer than the product B (μ_2).
Then the alternative hypothesis is that $\mu_1 - \mu_2 > 3$
The first step is to state the null hypothesis and an alternative hypothesis.
H_0: $\mu_1 - \mu_2 \leq 3$
H_1: $\mu_1 - \mu_2 > 3$
These hypotheses constitute a right-tailed test.

The significance level is (1-α)=0.95 and σ_1 and σ_2 are known. We will conduct a two-sample z-test

Using the data, we compute the standard error (SE) and the z statistic (z).

Compute the standard error:

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{3^2}{100} + \frac{2^2}{100}} = \sqrt{0.09 + 0.04} = 0.11$$

Exercise 1

A chocolate company has 2 kind of candies. The candy A has an expiring date of 3 days longer than B.
The quality control selects a simple random sample of 100 units of the product A and 100 of the product B and verifies that the average shelf-life of A is 15 days and 10 days for B's.
The company knows that the standard deviation of the shelf-life of the candy A is 3 days and 2 for the candy B.
Test the claim that the product A expire 3 days after the product B. Use a 0.95 level of significance.

Solution

From the statement of the test $(\mu_1 - \mu_2) = 3$

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{SE} = \frac{15 - 10\ - 3}{0.11} = 18.18$$

The critical value for the level of significance of 0.95 in a right-tailed test is 1.645, in a left-tailed test is -1.645, so we reject the null hyphotesis in both the cases, because the z test falls in the rejection regions of the tests

$(z > z_\alpha$ right-tailed)

Exercise 1

The Acme Company has developed a new battery. The engineer in charge claims that the new battery will operate continuously for at least 7 minutes longer than the old battery with a standard deviation for the old batteries of 3 minutes and 2.5 for the new batteries. To test the claim, the company selects a simple random sample of 100 new batteries and 100 old batteries. The old batteries run continuously for 190 minutes; the new batteries, 200 minutes.

Test the engineer's claim that the new batteries run at least 7 minutes longer than the old. Use a 0.95 level of significance.

Solution

State the hypotheses. The claim of the company is that the operarative time of the new batteries is, in mean, (μ_1) seven minutes longer than the old ones (μ_2).

The first step is to state the null hypothesis and an alternative hypothesis.

H_0: $\mu_1 - \mu_2 \leq 7$

H_1: $\mu_1 - \mu_2 > 7$

These hypotheses constitute a right-tailed test.

The null hypothesis will be rejected if the mean difference between sample means is too big.

For this analysis, the significance level is (1-α)=0.95 and σ_1 and σ_2 are known.

We will conduct a two-sample z-test of the null hypothesis.

Using the data, we compute the standard error (SE) and the z statistic (z).

Exercise 1

The Acme Company has developed a new battery. The engineer in charge claims that the new battery will operate continuously for at least 7 minutes longer than the old battery with a standard deviation for the old batteries of 3 minutes and 2.5 for the new batteries. To test the claim, the company selects a simple random sample of 100 new batteries and 100 old batteries. The old batteries run continuously for 190 minutes; the new batteries, 200 minutes.

Test the engineer's claim that the new batteries run at least 7 minutes longer than the old. Use a 0.95 level of significance.

Solution

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Compute the standard error (the denominator):

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{3^2}{100} + \frac{2.5^2}{100}} = \sqrt{0.09 + 0.0625} = 0.39$$

From the statement of the test $(\mu_2 - \mu_1) = 7$, so

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{SE} = \frac{200 - 190 \ - 7}{0.39} = 7.69$$

The critical value for the level of significance of 0.95 of 1.645, so we reject the null hyphotesis because the z test falls in the rejection region of the test $(z > z_\alpha)$

Exercise 2

Within a school district, students were randomly assigned to one of two Math teachers - Mrs. Smith and Mrs. Jones. After the assignment, Mrs. Smith had 30 students, and Mrs. Jones had 25 students. At the end of the year, each class took the same standardized test. Mrs. Jones' students had an average test score of 85, with a standard deviation of 15; and Mrs. Smith's students had an average test score of 78, with a standard deviation of 10. Assume that student performance is approximately normal.

Test the hypothesis that Mrs. Smith and Mrs. Jones are equally effective teachers. Use a $(1-\alpha)=0.90$ level of significance.

Solution

The first step is to state the null hypothesis and an alternative hypothesis.

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

These hypotheses constitute a two-tailed test. The null hypothesis will be rejected if the difference between sample means is too big or if it is too small.

For this analysis, the significance level is $(1-\alpha)=0.90$ and the σ of the population is not known (we know the standard deviations of the samples s_1 and s_2).

We have only the sample data; we will conduct a two-sample T-test of the null hypothesis.

Using sample data, we compute the standard error (SE), degrees of freedom (df), and the t statistic (t).

Exercise 2

Within a school district, students were randomly assigned to one of two Math teachers - Mrs. Smith and Mrs. Jones. After the assignment, Mrs. Smith had 30 students, and Mrs. Jones had 25 students. At the end of the year, each class took the same standardized test. Mrs. Jones' students had an average test score of 85, with a standard deviation of 15; and Mrs. Smith's students had an average test score of 78, with a standard deviation of 10. Assume that student performance is approximately normal.

Test the hypothesis that Mrs. Smith and Mrs. Jones are equally effective teachers. Use a (1-α)=0.90 level of significance.

Solution

We don't know if σ_1 and σ_2 are equal so:

$$\text{SE} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{10^2}{30} + \frac{15^2}{25}} = \sqrt{3.33 + 9} = \sqrt{12.33} = 3.51$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{(10^2/30 + 15^2/25)^2}{\frac{(10^2\,/\,30)^2}{29} + \frac{(15^2\,/\,25)^2}{24}} = \frac{(3.33 + 9)^2}{\frac{3.33^2}{29} + \frac{9^2}{24}} = \frac{152.03}{0.382 + 3.375} = \frac{152.03}{3.757} = 40.47 \cong 40$$

Then we can calculate the T test:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Exercise 2

Within a school district, students were randomly assigned to one of two Math teachers - Mrs. Smith and Mrs. Jones. After the assignment, Mrs. Smith had 30 students, and Mrs. Jones had 25 students. At the end of the year, each class took the same standardized test. Mrs. Jones' students had an average test score of 85, with a standard deviation of 15; and Mrs. Smith's students had an average test score of 78, with a standard deviation of 10. Assume that student performance is approximately normal.

Test the hypothesis that Mrs. Smith and Mrs. Jones are equally effective teachers. Use a (1-α)=0.90 level of significance.

Solution

The null hyphotesis is that: $\mu_1 - \mu_2 = 0$, so the test is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE} = (85 - 78) / 3.51 = 7/3.51 = 1.99$$

The test is two-tailed hence, we reject the null hypothesis if *t* is greater of the critical value $t_{df,\alpha/2}$

$$t=1.99>t_{40,0.05} = 1.684$$

Then the t falls in the rejection region of the test, hence we reject the null hypothesis (“the two teachers are not equally effective at the significance level of 90%”).