

Statistical Hypothesis Testing

Part I

Statistical Hypothesis

A hypothesis is a statement about a population.

θ : parameter of interest

The hypothesis takes the form of a prediction that the parameter θ takes a particular numerical value or falls in a certain range of values.

Example

A large pharmaceutical company selects employees to receive management training.

A group of women employees claimed that the company selects males at a disproportionately high rate for such training.

The company denied this claim.

Suppose the employee pool for selection is half male and half female.

The company's claim of a lack of gender bias is a hypothesis i.e. other things being equal, at each choice the probability of selecting a female (male) equals $\frac{1}{2}$.

The women's claim is an alternative hypothesis that the probability of selecting a male exceeds $\frac{1}{2}$.

Null and alternative hypotheses

The **Null hypothesis**, denoted by the symbol H_0 , is the statement we want to test.

H_0 states that the parameter takes a particular value.

The **alternative hypothesis**, denoted by the symbol H_1 , states that the parameter takes a different value or falls in some alternative range of values.

Usually the value in H_0 corresponds, in a certain sense, to no effect. The values in H_1 represent an effect of some type.

Example. If θ is the probability of selecting a male then

$$H_0: \theta = 0.5; \quad H_1: \theta > 0.5$$

No effect in this case refers to a lack of gender bias.

Statistical test

Let $X \sim f(x, \theta)$ be the random variable describing the population and

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta > \theta_0$$

where θ_0 is a possible value of θ

A **statistical test** is a procedure based on a function of the available sample data, called the **test statistics**, that allows one to take a decision as to whether the statistical hypothesis must be rejected or not

The **test statistics** summarizes how far the estimate of θ falls from the parameter value in H_0 . If the distance is **too large** the data do not give evidence that H_0 is true, hence H_0 is rejected

Errors and sample variability

A statistical test is based on a sample statistics and assumes that the data gathering employed randomization, such as a random sample

Being a function of (random) data the value of the test statistics can be accidentally large or small

Because of sampling variability, decisions in tests always have some uncertainty, hence the procedure can lead to a wrong decision

Errors and sample variability

Two types of potential errors, conventionally called Type I and Type II errors

Type I error: H_0 is true and H_0 is rejected.

Type II error: H_0 is false and H_0 is not rejected

		Population	
		H_0 true	H_0 false
Null hypothesis is...	Rejected	Type I error	Correct decision
	Not rejected	Correct decision	Type II error

Hypotheses on the mean of a Normal population with a known variance

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

$$H_0: \mu = \mu_0$$

The alternative hypothesis can take one of the two forms:

$$\left. \begin{array}{l} H_1: \mu < \mu_0 \\ H_1: \mu > \mu_0 \end{array} \right\}$$

One-sided alternative

$$H_1: \mu \neq \mu_0$$

Two-sided alternative

Example

X : weight of adult male patients who have been on a weight reduction program

Research question: the mean weight is less than 200lb.

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

Hypotheses to test: $H_0 : \mu = 200$ versus $H_1 : \mu > 200$

Suppose to have a **random sample** of $n = 10$ men on this program, and have calculated from these data the average weight $\bar{x} = 184lb$

We need a **statistics** to measure the distance between our hypothesis in H_0 and the data

i.e. $\bar{x} - 200 = 184 - 200 = -16$

Rationale: being \bar{X} a good estimator of μ we expect the difference $\bar{X} - 200$ small if H_0 is true

Test on the mean of a Normal population with a known variance

$$H_0 : \mu = \mu_0 \quad H_1 : \mu > \mu_0$$

Sample Data x_1, x_2, \dots, x_n is the observed random sample and \bar{x} is the sample mean

The decision to reject $H_0 : \mu = \mu_0$ is taken comparing \bar{x} to μ_0 according to an appropriate function

Test statistics: $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

Observed value: the value of the test statistics on the observed sample $z_{obs} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Test on the mean of a Normal population with a known variance

- The test statistics $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ summarizes how far the data fall from H_0 , hence the larger z_{obs} is the less reasonable H_0 is for the dataset at hand
Note that If H_0 is true, Z is normal with 0 mean and unit variance.
- Large values of Z are inconsistent with the statement on the population in H_0 , therefore the null hypothesis should be rejected on the base of the available data.

But... how much large the values of Z must be to reject H_0 ?

A threshold is necessary to quantify the idea of “large value”.

One way to do this is the *p-value*.

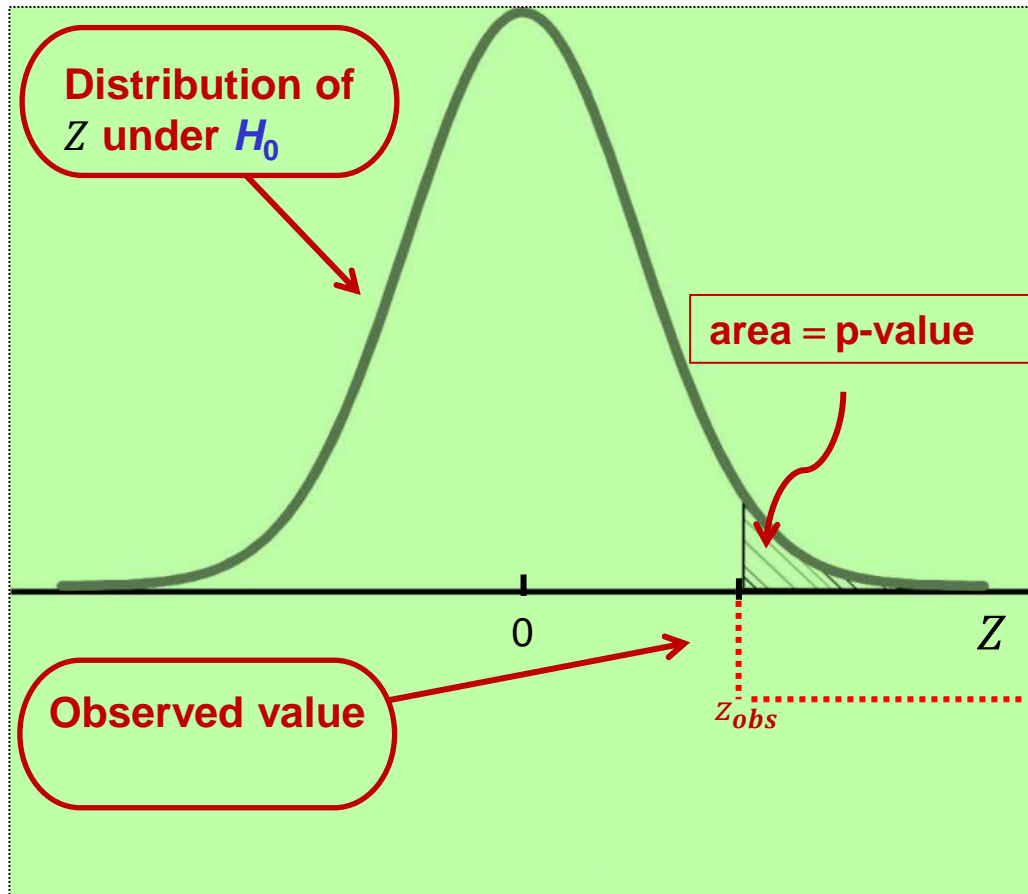
Test on the mean of a Normal population with a known variance: p-value (con't)

- A simple way is to transform Z to the probability scale of 0 to 1.
- This probability value is called the **P-value** of the test.

*The p-value measures the probability to find a sample result (a value of the test statistics) which is **more unusual** of the result observed on the sample (the actual value of the test statistics) if the statement in H_0 holds true.*

Test on the mean of a Normal population with a known variance: p-value

$$H_0 : \mu = \mu_0 \quad H_1 : \mu > \mu_0$$



$$p\text{-value} = P(Z > Z_{obs}; H_0 \text{ true})$$
$$= 1 - \Phi(Z_{obs})$$

(Z_{obs} is Z calculated on the sample data)

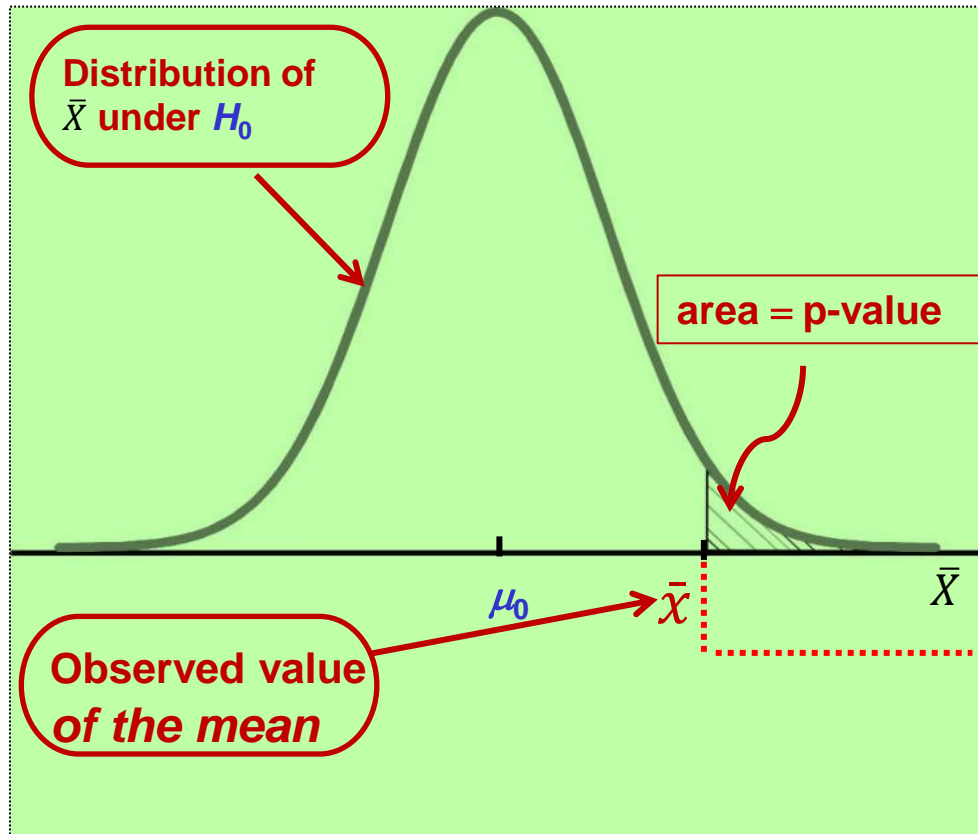
Small p-values means

large values of $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ i.e. $Z \gg 0$ and $\bar{X} \gg \mu_0$.

Hence the data suggest a value of the true μ larger than μ_0

Test on the mean of a Normal population with a known variance: p-value

$$H_0 : \mu = \mu_0 \quad H_1 : \mu > \mu_0$$



In other words, a large p-value implies a large value of \bar{X} which is unlikely if the true value of μ is μ_0 (remember that \bar{X} is a «good» estimator of the true mean of X)

In practice H_0 is rejected if the p-value is smaller than a fixed (small) threshold α (i.e. 0.05 or 0.01) called the **level of the test** or **significance level**

Example

X : arsenic concentration of underground water.

In normal situations it is expected that X is distributed $N(19.5, 10^2)$.

Hence, the expected concentration in the considered aquifer is 19.5mg/l.

In a random sample of size $n=12$ the following concentration values are observed:

29.506, 29.493, 29.511, 29.503, 29.487, 29.507, 29.495, 29.512, 29.510,
29.502, 29.497, 30.797

The sample mean is $\bar{x} = 29.61$.

We want to check if the water quality conforms to the required standard i.e.

$H_0 : \mu = 19.5$ versus $H_1 : \mu > 19.5$

Example

$$H_0: \mu = 19.5 \text{ versus } H_1: \mu > 19.5$$

$$\text{Test statistics: } Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 19.5}{10/\sqrt{12}}$$

$$\text{Value of } Z \text{ observed on the sample } z = \frac{29.61 - 19.5}{10/\sqrt{12}} = 3.5$$

$$p\text{-value} = P(Z > 3.5 | \mu = 19.5) = 1 - \Phi(3.5) = 1 - 0.99977 = 0.00023$$

Assuming a significance level $\alpha = 0.05$, the NULL hypothesis is rejected being $0.00023 < 0.05$.

Hence, the data confirm that the arsenic concentration does not conform to the required standard. The terminology often used is that the sample mean is statistically significant.

Test on the mean of a Normal population with a known variance: interpretation of the significance level

$$H_0 : \mu = \mu_0 \quad H_1 : \mu > \mu_0$$

Decision of the test: H_0 is rejected if the p-value is smaller than the **significance level α**

Remember, we can obtain an extreme observed value by chance and simply as a consequence of the sampling variability.

Hence, we can erroneously reject the null hypothesis, i.e., we can reject H_0 when it is actually true. However, it can be proved that the probability of making this error is smaller or at most equal to the significance level α . So, **α represents the probability of Type I error.**

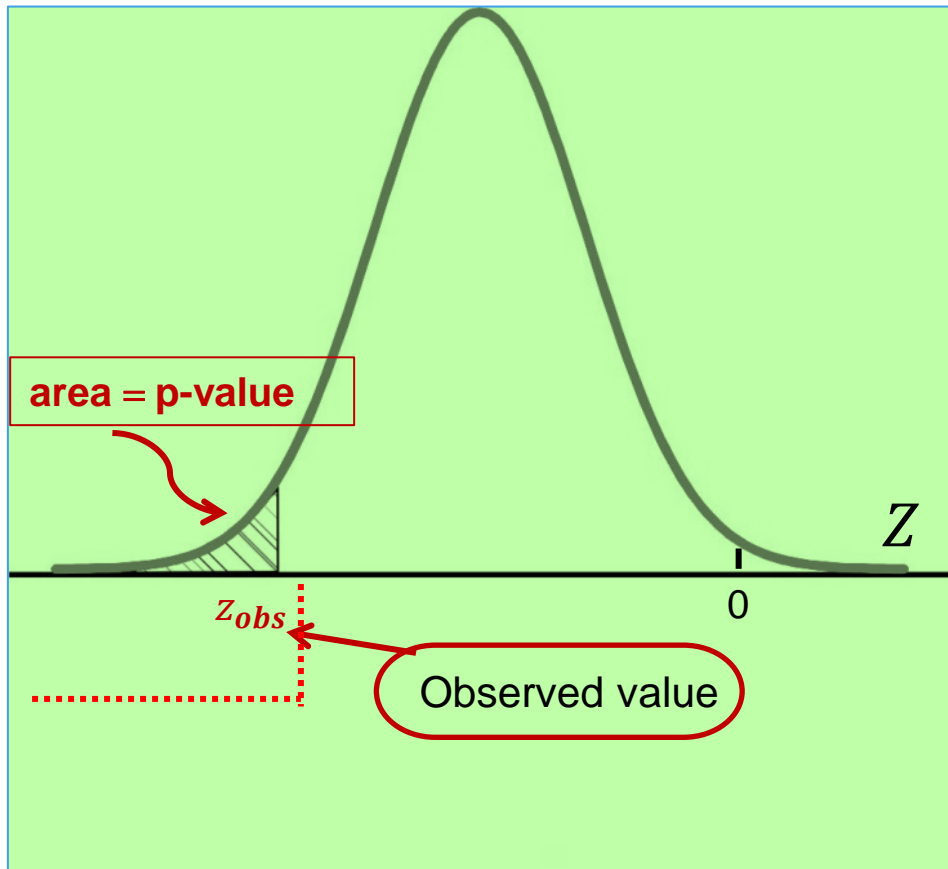
For this reason it is typically chosen small i.e. 0.1, 0.05, 0.001

Any statistical test allows us to control this probability and guarantees a procedure that makes a Type I error with a probability not larger than α .

Test on the mean of a Normal population - known variance

p-value

$$H_0 : \mu = \mu_0 \quad H_1 : \mu < \mu_0$$



$$p - value = P(Z < \mathbf{z_{obs}}; H_0 \text{ true}) = \Phi(\mathbf{z_{obs}})$$

$\mathbf{z_{obs}}$: the observed value of Z

Small p-values means small values (negative) of

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad \text{i.e. } Z \ll 0 \text{ and } \bar{X} \ll \mu_0.$$

Hence the data suggest a value of the **true μ** smaller than μ_0

Test on the mean of a Normal population - known variance

p-value

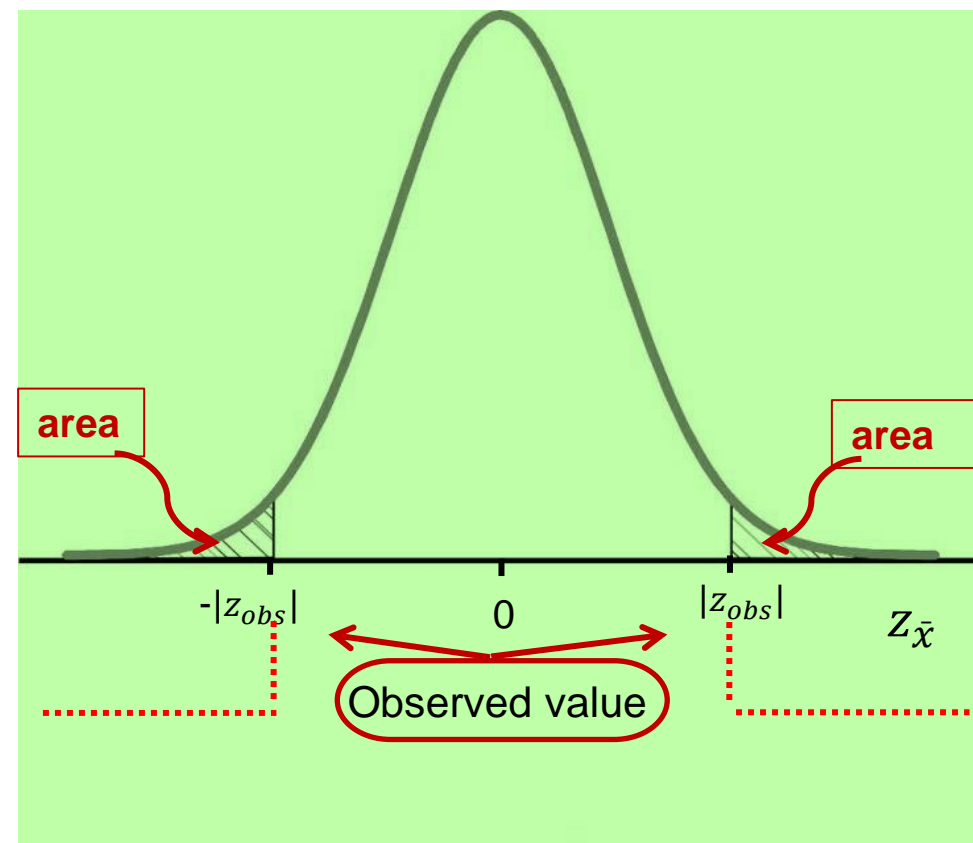
$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$$

$$\begin{aligned} p\text{-value} &= P_{H_0}(Z \leq -z_{obs}) + P_{H_0}(Z \geq z_{obs}) \\ &= 2[1 - \Phi(|z_{obs}|)] \end{aligned}$$

z_{obs} : the observed value of Z

In this case unusual values are those values of z_{obs} much larger or much smaller than 0.

Therefore, H_0 is rejected for those values of \bar{X} much larger or smaller than μ_0 which are not likely under H_0



Test on the mean of a Normal population - unknown variance

$$H_0 : \mu = \mu_0$$

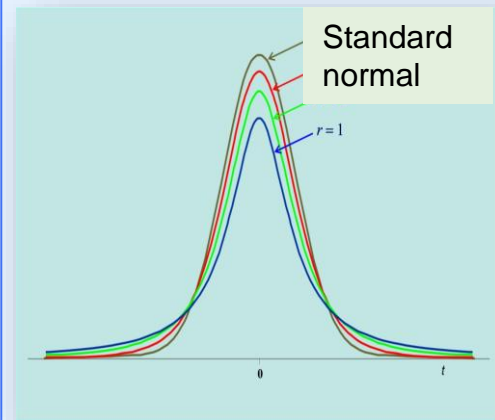
The decision to reject H_0 is taken comparing \bar{x} to μ_0 according to an appropriate statistics

If the variance is unknown the test statistics is $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$.

This is a Student's T with $n - 1$ df.

The statistical test is identical to the procedure seen before using the sample standard deviation S instead of the (unknown) standard deviation σ and the Student's T distribution instead of the standard normal distribution to calculate p-values

- $E(T) = 0$
- T is symmetric about 0
- if $r \gg 0$ then T is approximated by the standard normal rv



Example

X : weight of adult male patients who have been on a weight reduction program

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

Hypotheses to test: $H_0 : \mu = 200$ versus $H_1 : \mu < 200$

Data: random sample of $n = 10$ men where $\bar{x} = 184lb$ and $S = 26.5$

Test statistics: $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T_9$ (df = $n - 1 = 10 - 1 = 9$)

Observed value of T $= \frac{184 - 200}{26.5/\sqrt{10}} = \frac{-16}{8.38} = -1.9$

P-value $= P(T < t_{obs}; H_0 \text{ true}) = F_{t_9}(t_{obs}) = 0.0449$

where F_{t_9} is the cdf of a Student's t with 9 degrees of freedom

Decision: Therefore being $0.0449 < 0.05$ the null hypothesis is rejected at the 0.05 level

Summary on the p-value calculation

$$H_1 : \mu < \mu_0$$



$$p - value = P(Z_{\bar{X}} < z_{\bar{x}} | \mu = \mu_0)$$

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

$$H_1 : \mu > \mu_0$$



$$p - value = P(Z_{\bar{X}} > z_{\bar{x}} | \mu = \mu_0)$$

$$H_1 : \mu \neq \mu_0$$



$$p - value = P(|Z_{\bar{X}}| > |z_{\bar{x}}| | \mu = \mu_0)$$

- H_0 is rejected if the p-value is smaller than 0.05 (or other small thresholds)
- If the variance is unknown use the Student's T distribution and the test statistics $T = (\bar{X} - \mu_0) / (S / \sqrt{n})$

Confidence Interval (CI) and hypothesis testing

$$H_0: \mu = 0$$

$$H_1: \mu \neq 0$$

Output 1: One Sample t-test

t = 1.1521, df = 9, p-value = 0.279

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval: -0.2624543 0.8072536

sample estimates: mean of x: 0.2723996

P-value «big» therefore H_0 is not rejected

The CI includes the value specified in H_0

Output 2: One Sample t-test

t = 6.1784, df = 9, p-value = 0.000163

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval: 1.334253 2.875676

sample estimates: mean of x: 2.104964

P-value «small» therefore H_0 is not rejected

The CI does not include the value specified in H_0

An easy test for $H_0: \mu = \mu_0$ is then to build the CI of μ
and check whether it includes μ_0

Confidence Interval (CI) and hypothesis testing

Consider the hypothesis $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$

We know the CI of μ with a confidence level of $1 - \alpha$ is

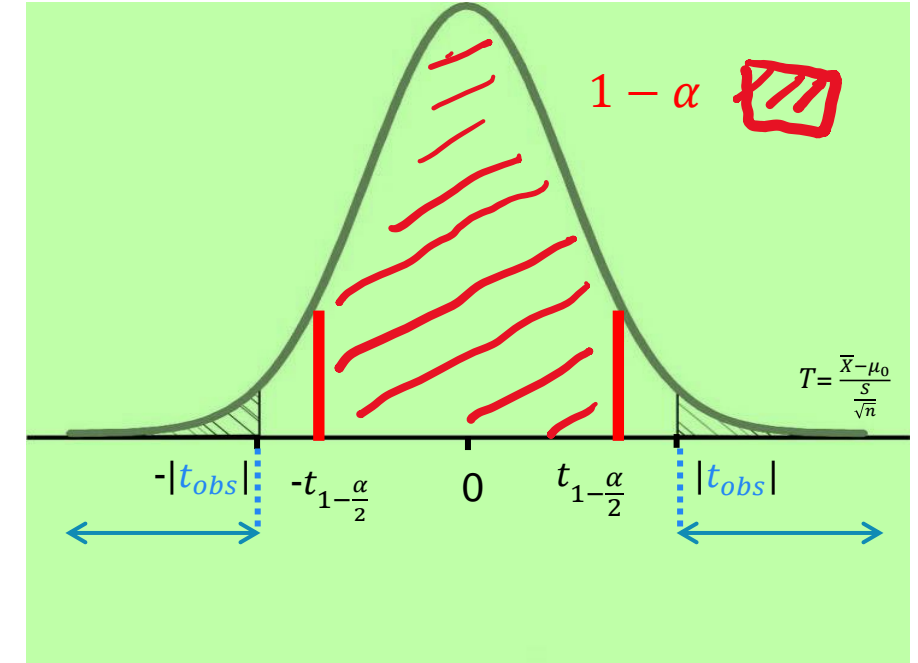
$$\left(\bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right)$$

Assume that $\mu_0 \notin \left(\bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right)$

$$\Rightarrow \mu_0 < \bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \quad \text{or} \quad \mu_0 > \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$\Rightarrow t_{1-\frac{\alpha}{2}} < \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = t_{obs} \quad \text{or} \quad -t_{1-\frac{\alpha}{2}} > \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = -|t_{obs}| \quad (\text{in this case } \bar{x} - \mu_0 < 0)$$

$$\Rightarrow p\text{-value} = 2[1 - F(|z_{obs}|)] = P\left(-t_{obs} < \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < t_{obs}\right) < \alpha$$



An easy test for $\mu = \mu_0$ is then to build the CI of μ and check whether it includes μ_0