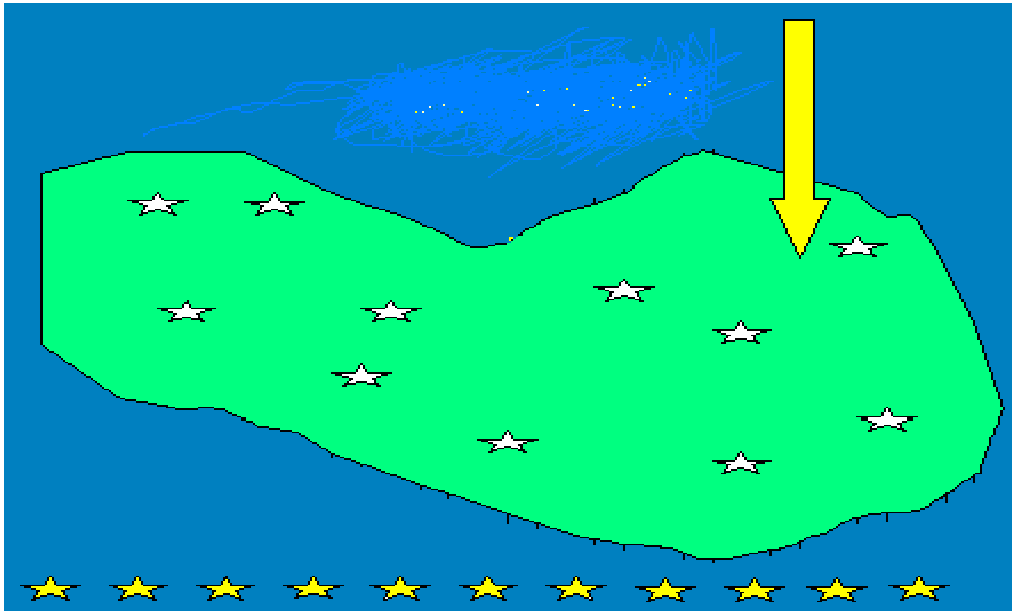
# Hypothesis Test

“ **Hypothesis Test**”: A procedure for deciding between two hypotheses (null hypothesis – alternative hypothesis) on the basis of observations in a random sample

# Sampling models

## A. One sample model



- One sample model usually concerns with an intervention on population: If the intervention should make some change in population?

# One – sample Hypothesis test

- Compare proportion to a given value of rate

- Compare mean value to a given value of expectation

- **Example 1.** If in Ha Dong District less than 90% motorcyclists use helms?

- **Example 2.** If proportion of girl-students equals to 50% ?

- **Example 3.** If in Viet Nam bred feeding is popular among more than 70% women?

## Test 1. *Compare proportion to a given rate*

***Two tails (sides) test:***

(Null) Hypothesis $H_0$: $p = q$

Alternative Hypothesis $H_a$: $p$ *differs from* $q$

***Lower (left) tail test:***

(Null) Hypothesis $H_0$: $p \geq q$

Alternative Hypothesis $H_a$: $p < q$

***Upper (right) tail test:***

(Null) Hypothesis $H_0$: $p \leq q$

Alternative Hypothesis $H_a$: $p > q$

# Test 1. *Compare proportion to a given rate*

$(X_1, X_2, ..., X_n)$ - a sample of ***n*** independent observations collected from a binary variable ***X*** taking value 1 with (unknown) probability *p* *(0 < p < 1)* and value 0 with probability *1 – p* → Given a number *q* , how to have a conclusion comparing *p* with *q* based on information of the sample?

## Solution

By Moivre-Laplace Theorem, for large sample size,

$n \times p \geq 5$ and $n \times (1-p) \geq 5$,

sample proportion $m(p)/n$ of appearance of number 1 has distribution approximate to normal distribution with expectation $p$ and variance $p \times (1-p)/n$. Then a testing procedure can be as follows:

**Step 1.** Estimate a sample proportion by

$$p\hat{} = m(p) / n$$

Calculate **z-test statistic**

$$z_q = (p\hat{}-q)/[p\hat{} \times (1-p\hat{})/n]^{1/2}$$

***Two tails (sides) test:***

(Null) Hypothesis $H_0$: $p = q$

Alternative Hypothesis $H_a$: $p \neq q$

# *P-value Approach*

**Step 2.** (two sides test)

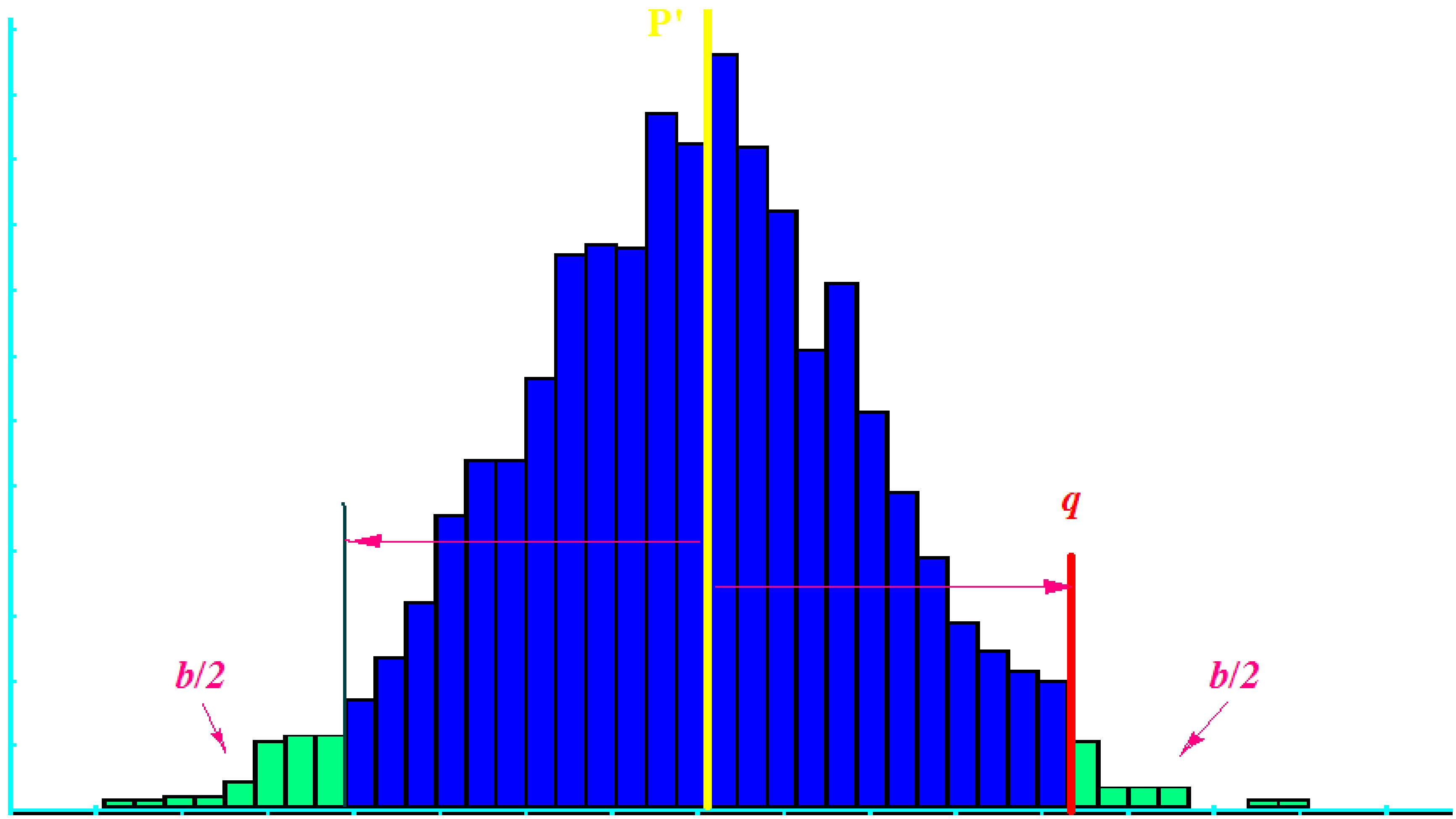
*Approach A (p-value method):* Calculate the probability such that the distance from estimate point to ***p^*** should be more than **|*q* – *p^*|** :

$$b = P\{|Y - q| > |p\hat{} - q|\} = P\{|Z| > |z_q|\}$$

where ***Y*** is a normal variable with mean ***p^*** and variance ***p^ ×(1-p^)/n***,

while ***Z*** is the standard normal variable (the normal variable with mean **0** and variance **1**).

***b*** = Probability (***p-value***) of wrong decision of excluding estimation value ***q*** (saying that ***q*** differs from true value of ***p***) when this value should be a “good” value of estimation


P'
q
b/2
b/2

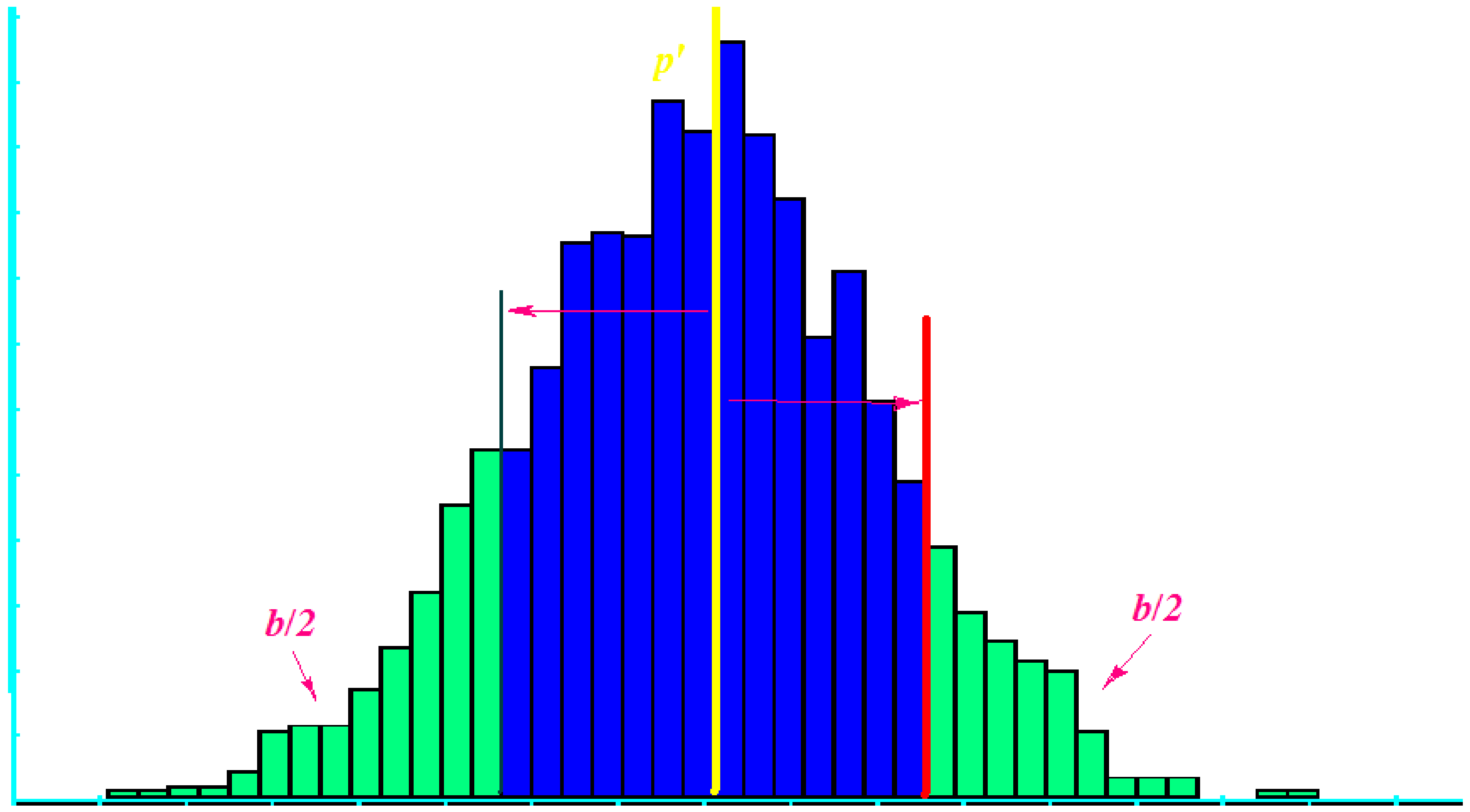
***Step 3***. (two sides test)

Compare **b** to the significance level **α**

(**α = 10%, 5%, 1%, …**)

- If **b** ≤ **α** → ***reject the hypothesis*** $H_0$, conclude that ***q*** differs from ***p***, because possibility of getting mistake in decision is "very small"

* If **b** **>** **α** → ***accept the hypothesis*** $H_0$, confirm ***q* = *p***, because possibility of having mistake by rejecting the hypothesis is too high.


p'
b/2
b/2

***Two tails (sides) test:***

(Null) Hypothesis $H_0$: $p = q$

Alternative Hypothesis $H_a$: $p \neq q$

# *Critical value Approach*

**Step 2.** (two sides test)

*Approach B (critical value method):* With the given significance level **$\alpha$** (**$\alpha$ = 10%, 5%, 1%, …**), calculate the critical value $\mathbf{z_{1-\alpha/2}}$, the number point such that **$1-\alpha/2$** is the probability of the standard normal variable ***Z*** taking values smaller or equal to $\mathbf{z_{1-\alpha/2}}$:

$$1-\alpha/2 = P\{Z \le z_{1-\alpha/2}\}$$

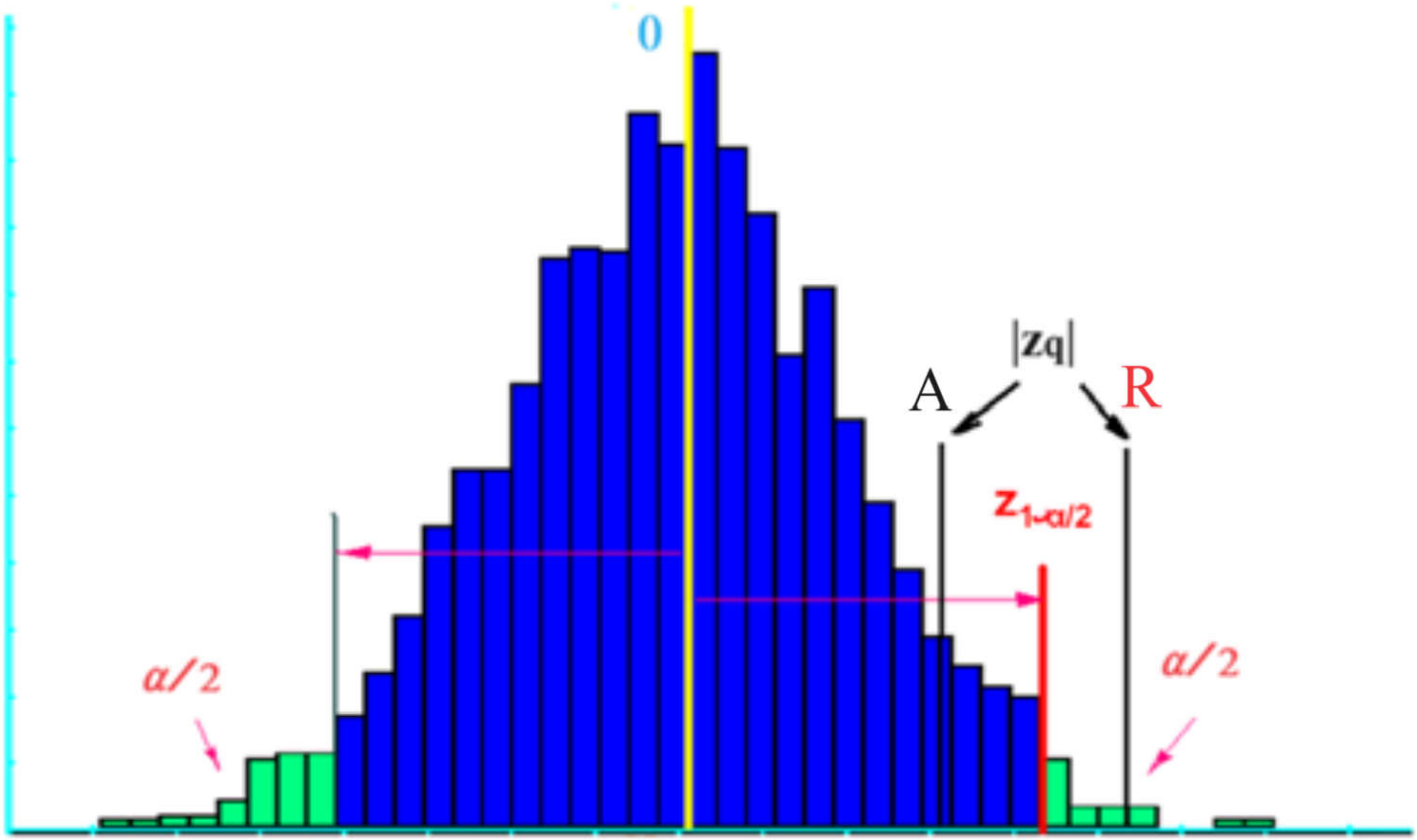
***Step 3***. (two sides test)

Compare **absolute value** of the **z-test statistic** ($| z_q |$) too the *critical value* $z_{1-\alpha/2}$

(**$\alpha$ = 10%, 5%, 1%, …**)

- If $| z_q | \geq z_{1-\alpha/2}$ → ***reject the hypothesis*** $H_0$, conclude that ***q*** differs from ***p***.

* If $| z_q | < z_{1-\alpha/2}$ → ***accept the hypothesis*** $H_0$, confirm ***q*** = ***p***.


0
|zq|
A
R
z1-α/2
α/2
α/2

***Two tails (sides) test:***

(Null) Hypothesis $H_0$: $p = q$

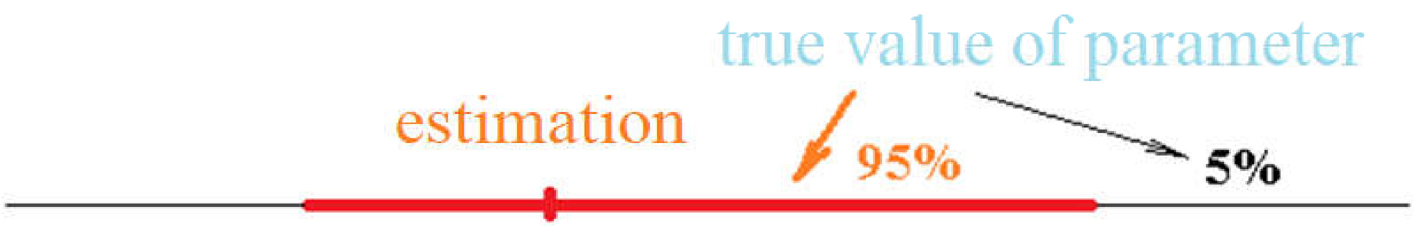
Alternative Hypothesis $H_a$: $p \neq q$

# *Confidence Interval Approach*

# **Approach C**. Using confidence interval

With ***confidence level*** of ***95%***, we can use confidence interval for hypothesis testing:

$$\left[p^\wedge - 1.96 * \underbrace{\sqrt{p^\wedge.(1-p^\wedge)/n}}_{\text{standard error}} ; p^\wedge + \underbrace{1.96 * \sqrt{p^\wedge.(1-p^\wedge)/n}}_{\text{margin of error}}\right]$$




# Decide

- ***Reject the Hypothesis*** H if the confidence interval does not contain the point ***q***
- ***Accept the Hypothesis*** H if the confidence interval contains the point ***q***

# Using Excel to Compute Normal P-value and Critical Value

- Excel has two functions for computing P-value and critical respect to normal distribution:

  - NORMDIST, NORMSDIST are used to compute the P-values.

  - NORMINV and NORMSINV are used to compute the Critical values.

# Compare mean value to a given value of expectation

# Test 2. Compare mean value to a given value of expectation

**Problem**: Taking a sample from a variable X with normal distribution, we need to compare the **expectation** $\mu$ of X to a given value **u**.

**Hypothesis**

$H_0$: $\mu$ = **u**

**Alternative Hypothesis**

$H_a$: $\mu$ differs ($\neq$, $\leq$, $\geq$) from **u**

***Two tails (sides) test:***

(Null) Hypothesis $H_0$: $\mu = u$

Alternative Hypothesis $H_a$: $\mu$ ***differs from u***

***Lower (left) tail test:***

(Null) Hypothesis $H_0$: $\mu \geq u$

Alternative Hypothesis $H_a$: $\mu < u$

***Upper (right) tail test:***

(Null) Hypothesis $H_0$: $\mu \leq u$

Alternative Hypothesis $H_a$: $\mu > u$

# Compare mean value when variance of variable is unknown

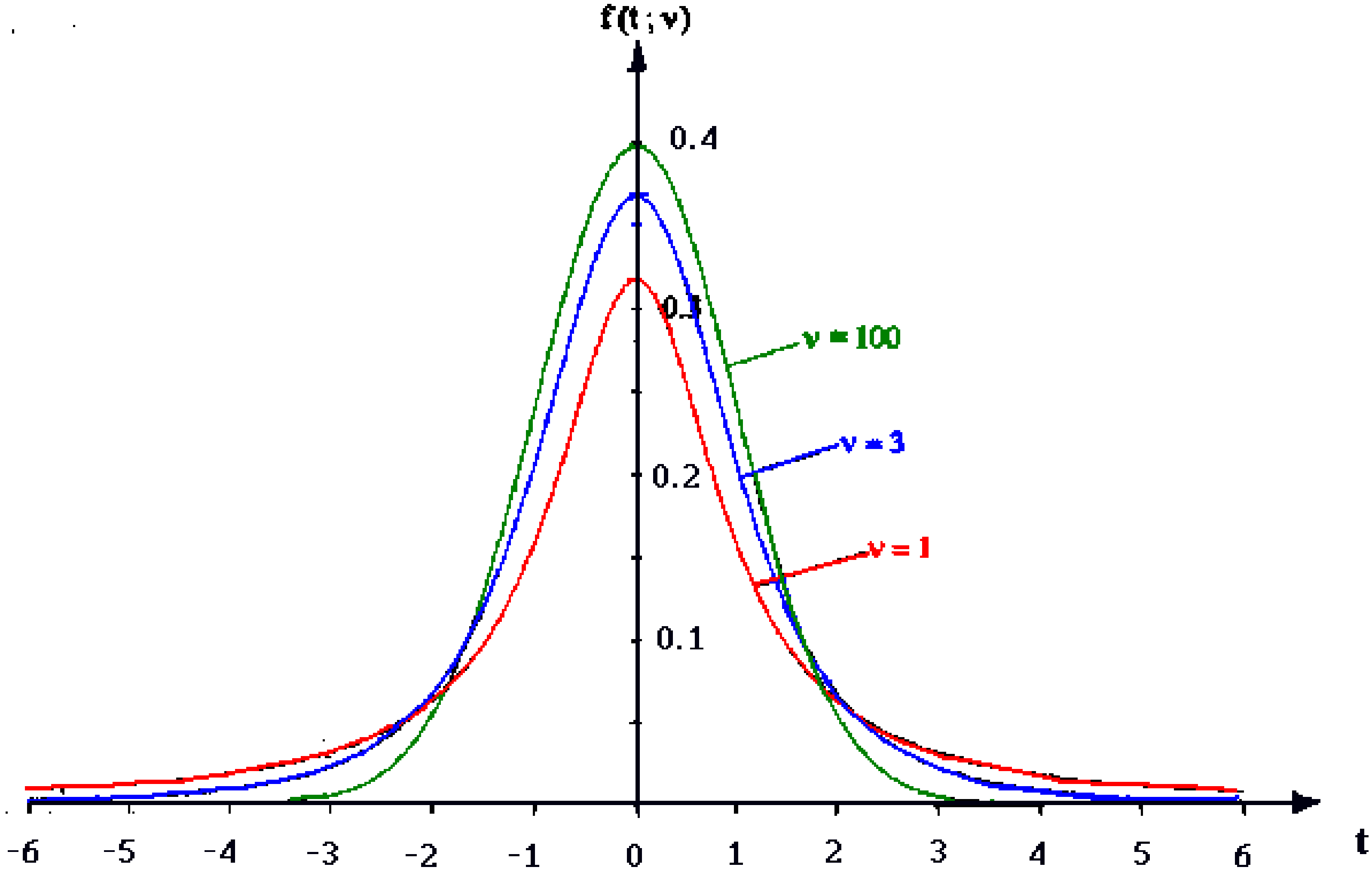
For testing the above hypothesis, the distribution of sample mean value must be known. When **variance of variable is unknown** the following theorem can be applied:

**Theorem.** Let $(X_1, X_2, ..., X_n)$ be a sample of $n$ independent observations taken from a normal distributed variable $X$ with expectation $\mu$, $\bar{X}$ is sample mean value and $S^2$ is sample variance. Then the (new) variable

$$t = \frac{\sqrt{n-1}}{S}.(\bar{X} - \mu)$$

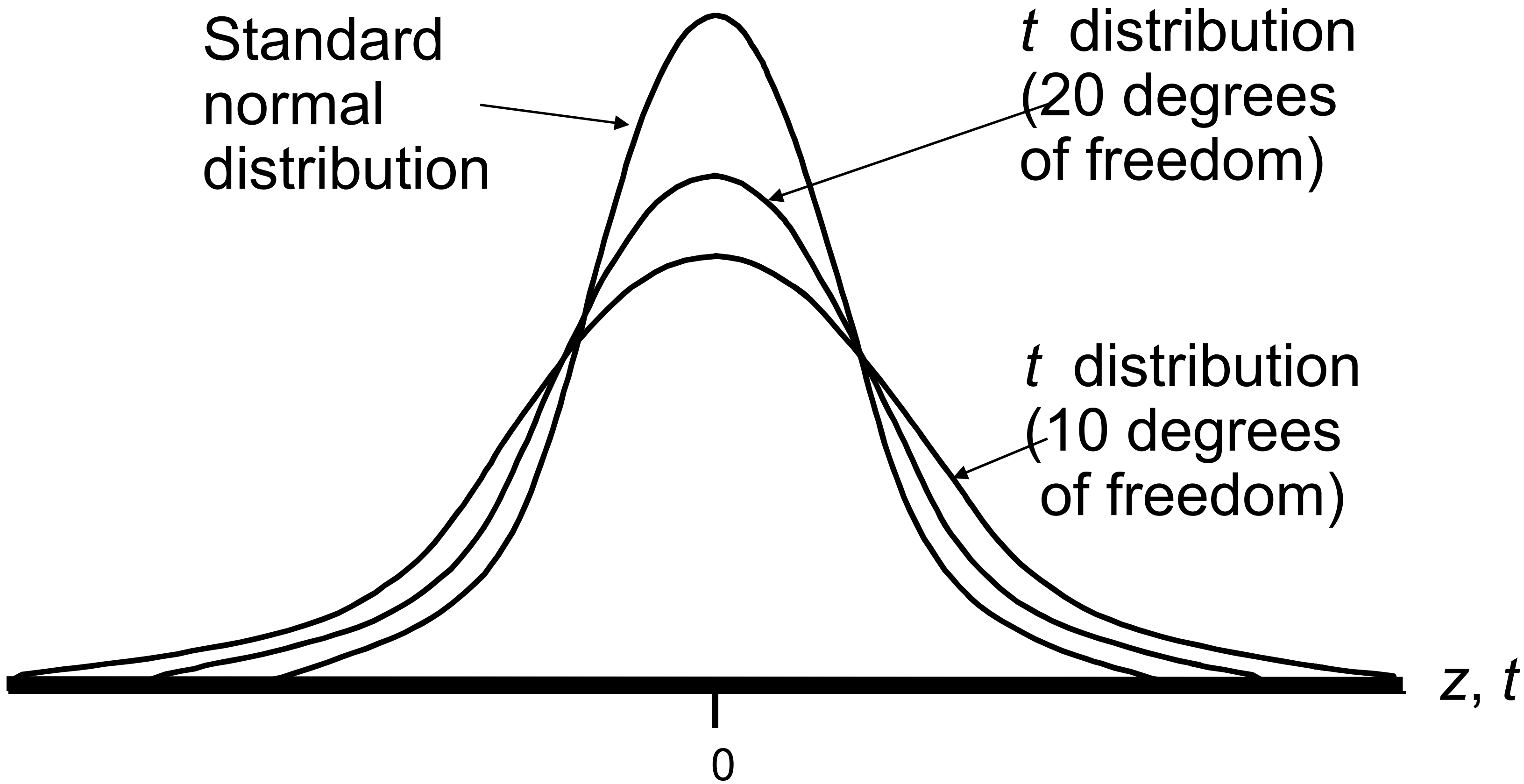
has T-Student distribution with *(n-1)* degrees of freedom.

# Student (T) distribution




Parameter of Student Distribution: “Degree of Fredom” $\nu$

# *t* Distribution



# $t$ Distribution

- For more than 100 degrees of freedom, the standard normal $z$ value provides a good approximation to the $t$ value.

# Steps of Testing

Step 1. Estimate sample Mean Value Mean(X) Standard Deviation SD(X)

Step 2. Calculate the statistic

$$t(u) = \frac{\sqrt{n-1}.(Mean(X) - u)}{SD(X)}$$

where ***n*** is sample size and ***u*** is the given value to which the mean value has be compared

***Two tails (sides) test:***

(Null) Hypothesis $\mathbf{H_0}$: $\mu$ = ***u***

Alternative Hypothesis $\mathbf{H_a}$: $\mu$ ***differs from u***

# *P-value Approach*

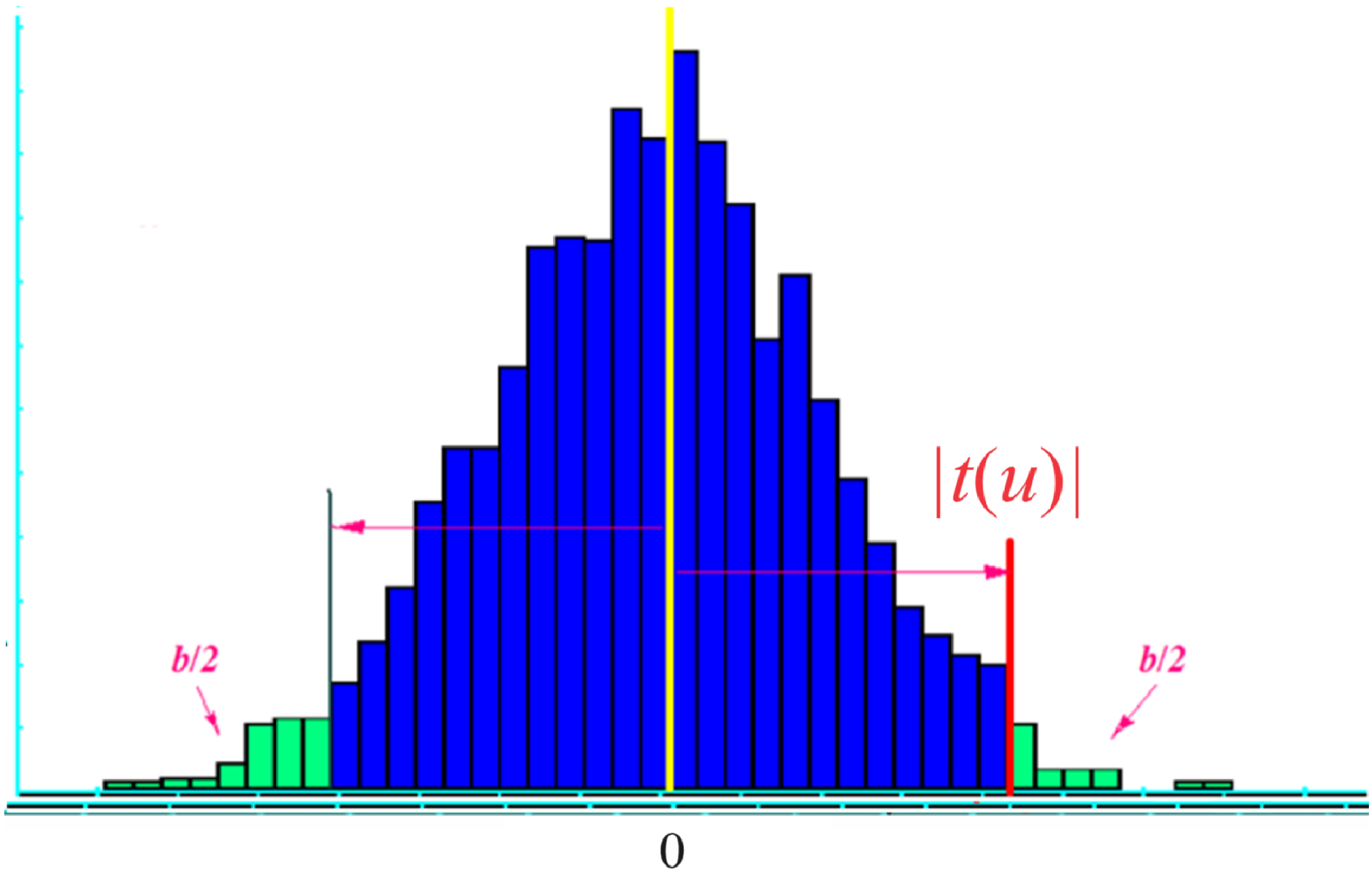
**Step 2.** (two sides test)

*Approach A (p-value method):* Calculate the probability

$$\mathbf{b = P\{|}\boldsymbol{T}(n\text{-}1)\mathbf{| \geq |}t(u)\mathbf{|\}}\ ,$$

where $\boldsymbol{T}(n\text{-}1)$ is a variable which has Student distribution with (***n*-1**) degrees of freedom.

***b*** = Probability (***p-value***) of wrong decision of excluding estimation value ***u*** when this value should be a "good" value of estimation



$|t(u)|$
b/2
b/2
0

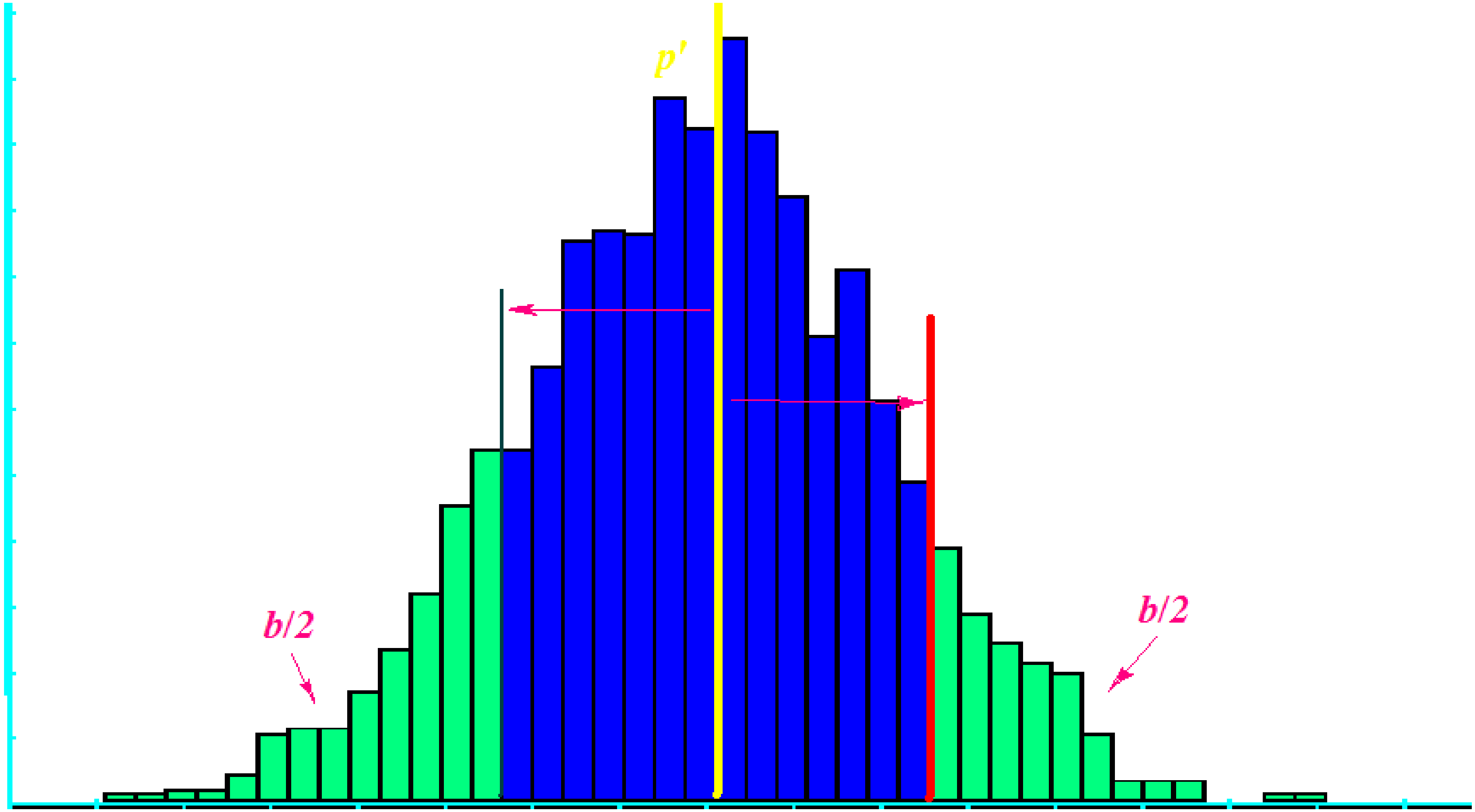
***Step 3***. (two sides test)

Compare **b** to the significance level **α**

(**α = 10%, 5%, 1%, …**)

- If **b ≤ α** → ***reject the hypothesis*** $H_0$, conclude that μ differs from ***u***, because possibility of getting mistake in decision is "very small"

* If **b > α** → ***accept the hypothesis*** $H_0$, confirm **μ = *u***, because possibility of having mistake by rejecting the hypothesis is too high.


p'
b/2
b/2

***Two tails (sides) test:***

(Null) Hypothesis $\quad H_0: \mu = u$

Alternative Hypothesis $\quad H_a: \mu$ ***differs from u***

# *Critical value Approach*

**Step 2.** (two sides test)

*Approach B (critical value method):* With the given significance level **α** (**α = 10%, 5%, 1%, …**), calculate the critical value $t_{1-\alpha/2}$, the number point such that **1-α/2** is the probability of the Student distributed variable $T(n-1)$ taking values smaller or equal to $t_{1-\alpha/2}$:

$$1-\alpha/2 = P\{T(n-1) \leq t_{1-\alpha/2}\}$$

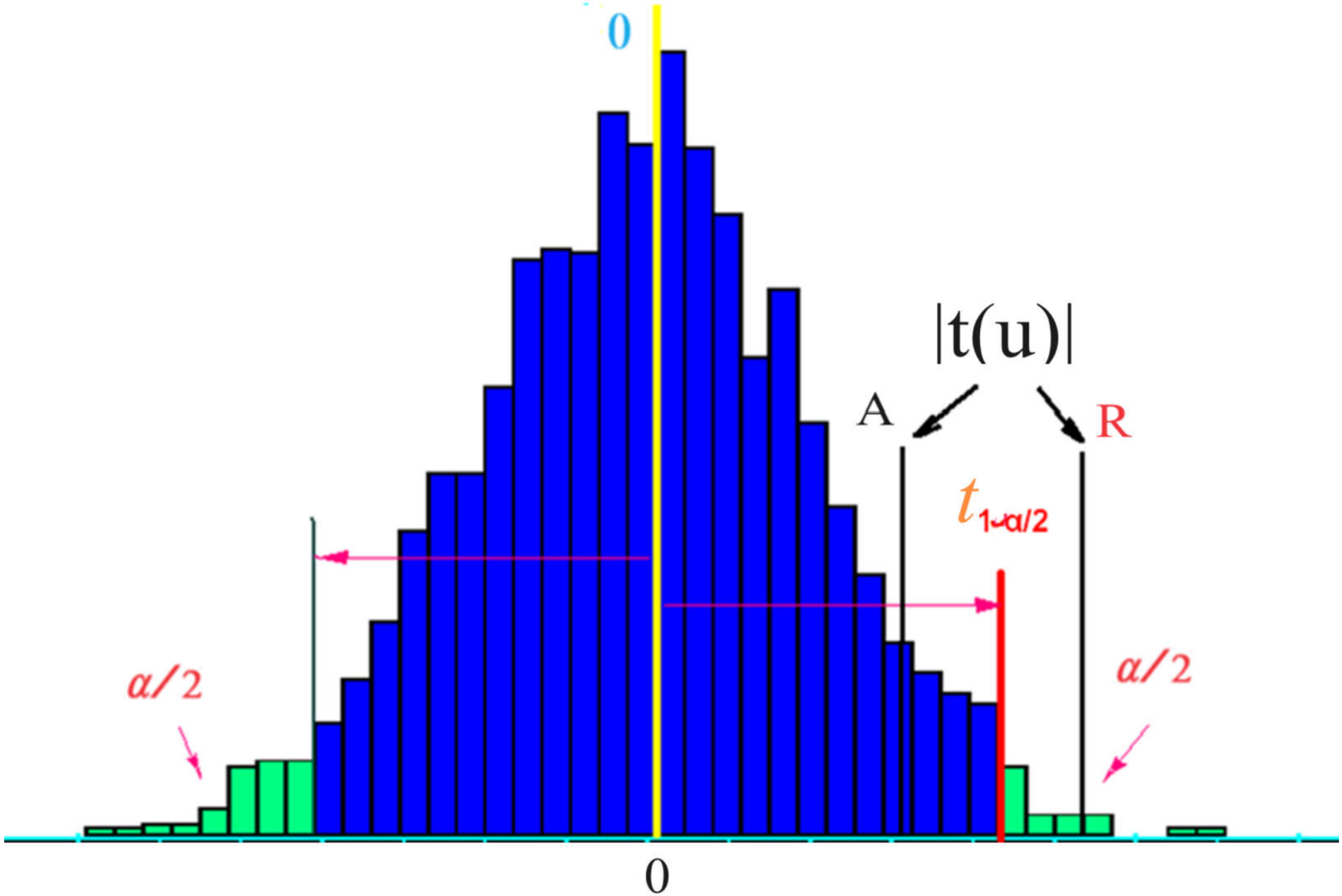
***Step 3***. (two sides test)

Compare **absolute value** of the **t-test statistic** ($|t(u)|$) too the *critical value* $t_{1-\alpha/2}$

(**$\alpha$ = 10%, 5%, 1%, …**)

- If $|t(u)| \geq t_{1-\alpha/2}$ → ***reject the hypothesis*** $H_0$, conclude that $\mu$ differs from $u$.

* If $|t(u)| < t_{1-\alpha/2}$ → ***accept the hypothesis*** $H_0$, confirm $\mu = u$.



0
|t(u)|
A
R
$t_{1-\alpha/2}$
$\alpha/2$
$\alpha/2$
0

***Two tails (sides) test:***

(Null) Hypothesis **$H_0$: $\mu$ = *u***

Alternative Hypothesis **$H_a$: $\mu$ *differs from u***

# *Confidence Interval Approach*

# **Approach C**. Using confidence interval

With ***confidence level*** of (1-$\alpha$), we can use confidence interval for hypothesis testing:

$$\left[\text{Mean}(X) - \underbrace{T(n\text{-}1, 1\text{-}\alpha/2) * SD(X)/\sqrt{n}}_{\text{margin of error}} \; ; \; \text{Mean}(X) + T(n\text{-}1, 1\text{-}\alpha/2) * \underbrace{SD(X)/\sqrt{n}}_{\text{standard error}}\right]$$


# Decide

- ***Reject the Hypothesis*** **H** if the confidence interval does not contain the point ***u***
- ***Accept the Hypothesis*** **H** if the confidence interval contains the point ***u***

# Using Excel to Compute Student P-value and Critical Value

- Excel has two functions for computing P-value and critical respect to Student distribution:

  - TDIST is used to compute the P-values.

  - TMINV is used to compute the Critical values.