TEST DI IPOTESI (DISTR. NORMALE)

__Definizione__ : Un' ipotesi formulata in termini di parametri di una popolazione (ad esempio media o varianza) è detta IPOTESI STATISTICA.

Il procedimento che consente di rifiutare o accettare un' ipotesi statistica utilizzando i dati di un campione, viene chiamato TEST DI IPOTESI.

Nella costruzione del test si avanzano due ipotesi :

- IPOTESI NULLA H_0 : ipotesi vera fino a prova contraria
- IPOTESI ALTERNATIVA H_1 : complementare di H_0

L'obiettivo del test è verificare se i dati smentiscono l'ipotesi nulla H_0.

RIGETTO : " i dati del campione non confermano l'ipotesi "

ACCETTO : " i dati del campione non consentono di rigettare l'ipotesi "

Come si può decidere quale è H_0 e quale H_1?

a) nell'ipotesi alternativa ci si mette ciò che si spera o che ci si aspetta di poter concludere come risultato del test;

b) l'ipotesi nulla H_0 è posta con lo scopo di essere screditata, quindi ciò che si oppone alla conclusione che il ricercatore cerca di raggiungere rappresenta l'ipotesi nulla;

c) nell'ipotesi nulla deve sempre comparire un segno di uguaglianza ($=$, $\leq$ o $\geq$);

d) le due ipotesi sono complementari.

<u>Osservazione</u>: Le possibili conclusioni per un test di ipotesi sono:

1) se l'ipotesi nulla H_0 è rifiutata, si conclude che l'ipotesi alternativa H_1 è probabilmente vera;

2) se l'ipotesi nulla non è rifiutata, si conclude che i dati non forniscono una sufficiente evidenza per sostenere l'ipotesi alternativa.

Definizione: La REGIONE DI RIFIUTO corrisponde all'insieme dei valori di una statistica test che conducono al rifiuto dell'ipotesi nulla. L'insieme dei valori che portano all'accettazione dell'ipotesi nulla si chiama REGIONE DI ACCETTAZIONE.
I VALORI CRITICI sono i valori della statistica test che separano le regioni di rifiuto e di accettazione.

TIPI DI ERRORE E LIVELLO DI SIGNIFICATIVITA'

Nell'accettare o rifiutare una determinata ipotesi nulla H_0, si può agire correttamente (e cioè accettare un'ipotesi vera o rifiutare un'ipotesi falsa), oppure si possono commettere due tipi di errore:

I) ERRORE DEL I TIPO: rifiutare H_0 quando è vera;

II) ERRORE DEL II TIPO: accettare H_0 quando è falsa.

- $P(\text{ERRORE I TIPO}) = \alpha$ LIVELLO DI SIGNIFICATIVITA': α

- $P(\text{ERRORE II TIPO}) = \beta$ RISCHIO DEL CONSUMATORE: β

Seguiamo il test di ipotesi su un campione normale avente media μ e varianza σ^2, avente come stimatore di μ : $\hat{\mu} = \bar{X}_n$

$$\text{Supponiamo} : \begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \neq \mu_0 \end{cases}$$

Sembra ragionevole accettare H_0 quando $\bar{X}_n$ non è troppo lontano da μ_0.

- $P(\text{ERRORE I SPECIE}) = \alpha$

$$\parallel$$

$$P_{\mu_0}\left(|\bar{X}_n - \mu_0| > c \right) = \alpha$$

dove scriviamo P_{μ_0} per intendere che la probabilità precedente viene calcolata con l'assunzione che $\mu = \mu_0$. Infatti la definizione di errore di I specie prevede che esso si verifichi quando i dati ci portano a rifiutare H_0 mentre in realtà esso è vera.

Quando $\mu = \mu_0$

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

$$P_{\mu_0}(|\bar{X}_n - \mu_0| > c) = \alpha$$

$$P_{\mu_0}\left(\left| \frac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}} \right| > \frac{c}{\sigma / \sqrt{n}} \right) = \alpha$$

$$P_{\mu_0}\left(|Z| > \frac{\sqrt{n}}{\sigma} c \right) = \alpha$$

$$\Updownarrow$$

$$2 P_{\mu_0}\left(Z > \frac{\sqrt{n}}{\sigma} c \right) = \alpha$$

$$P_{\mu_0}\left(Z > \frac{\sqrt{n}}{\sigma} c \right) = \frac{\alpha}{2}$$

Siccome per definizione di $z_{\alpha/2}$ vale

$$P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$$

si deduce che

$$\frac{\sqrt{n}}{\sigma} c = z_{\alpha/2}$$

Il test di ipotesi si traduce nel calcolare

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

• $P(\text{ERRORE II SPECIE}) = \beta$

$$P_\mu\left(\left|\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right) = \beta$$

Scriviamo P_μ per intendere che la probabilità dipende da μ, la quale è la media vera ma sconosciuta.

$$P_\mu\left(\left|\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right) =$$

$$= P_\mu\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) =$$

$$= P_\mu\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu + \mu - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) =$$

$$= P_\mu\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) =$$

$$= P_\mu\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2} < Z < \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) =$$

$$= \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) - \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) =$$

$$= \beta(\mu)$$

$\beta(\mu)$: curva OC (curva caratteristica)

$F(\mu) = 1 - \beta(\mu)$ POTENZA DEL TEST

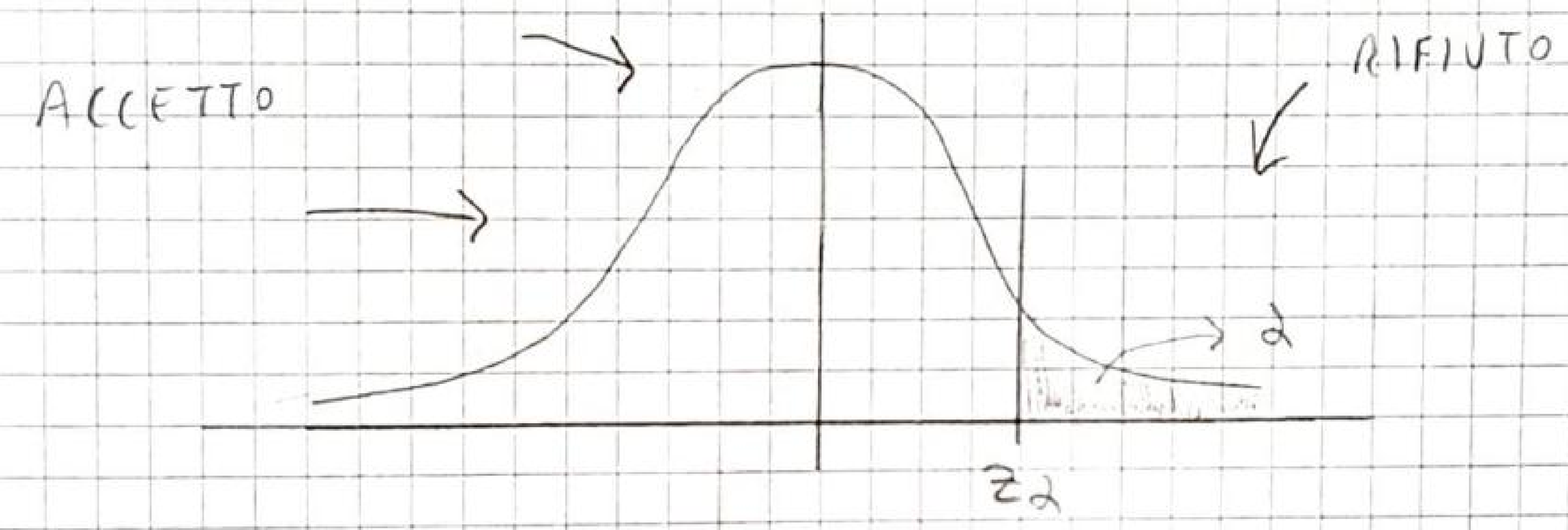
TEST UNILATERALI (A UNA CODA)

1° CASO

$H_0 : \mu \leq \mu_0$

$H_1 : \mu > \mu_0$

REGIONE DI RIFIUTO :	$z > z_\alpha$
REGIONE DI ACCETTAZIONE :	$z < z_\alpha$

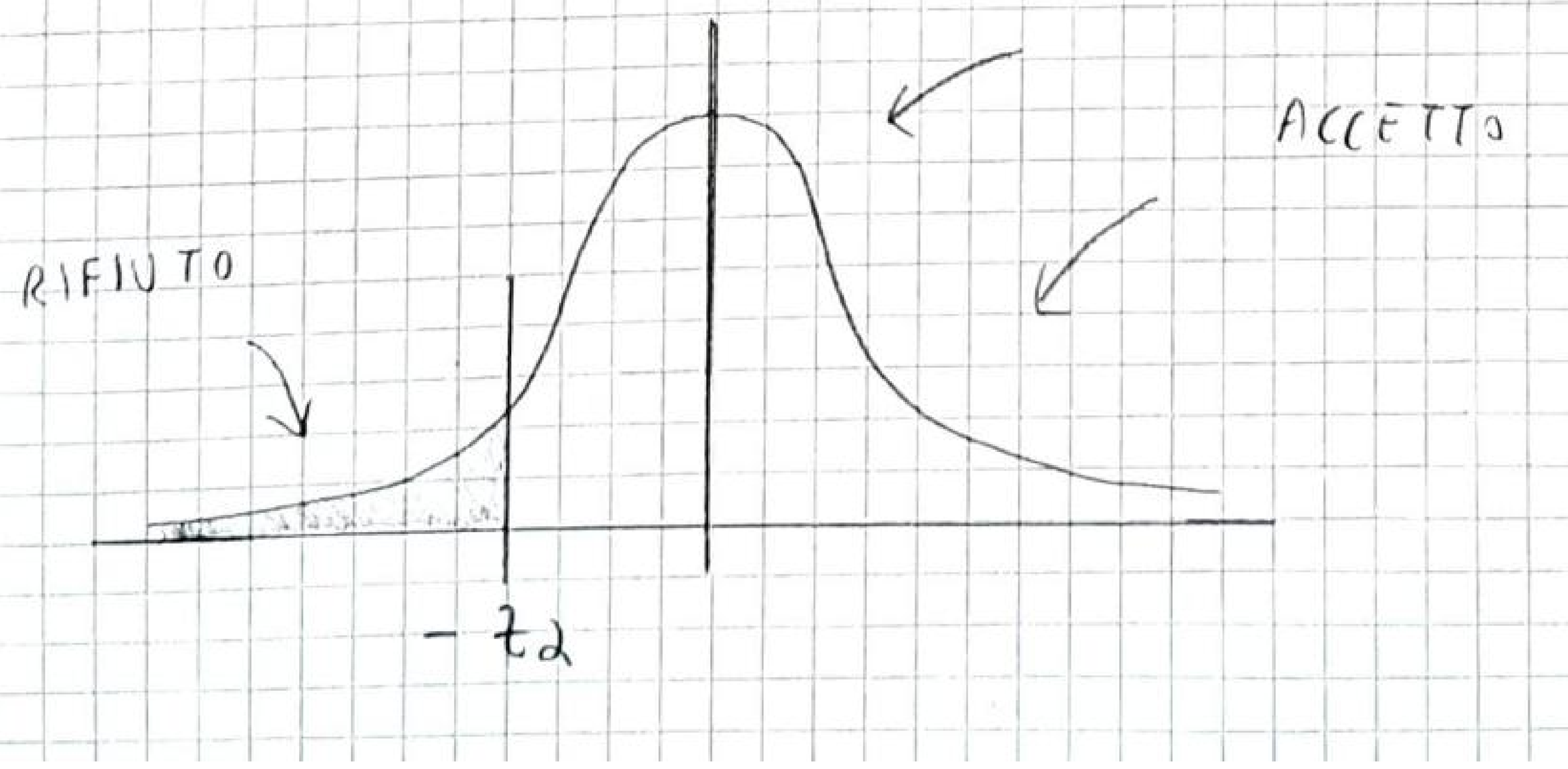


2° CASO

$H_0 : \mu \geq \mu_0$

$H_1 : \mu < \mu_0$

REGIONE DI RIFIUTO :	$z < -z_\alpha$
REGIONE DI ACCETTAZIONE :	$z > -z_\alpha$

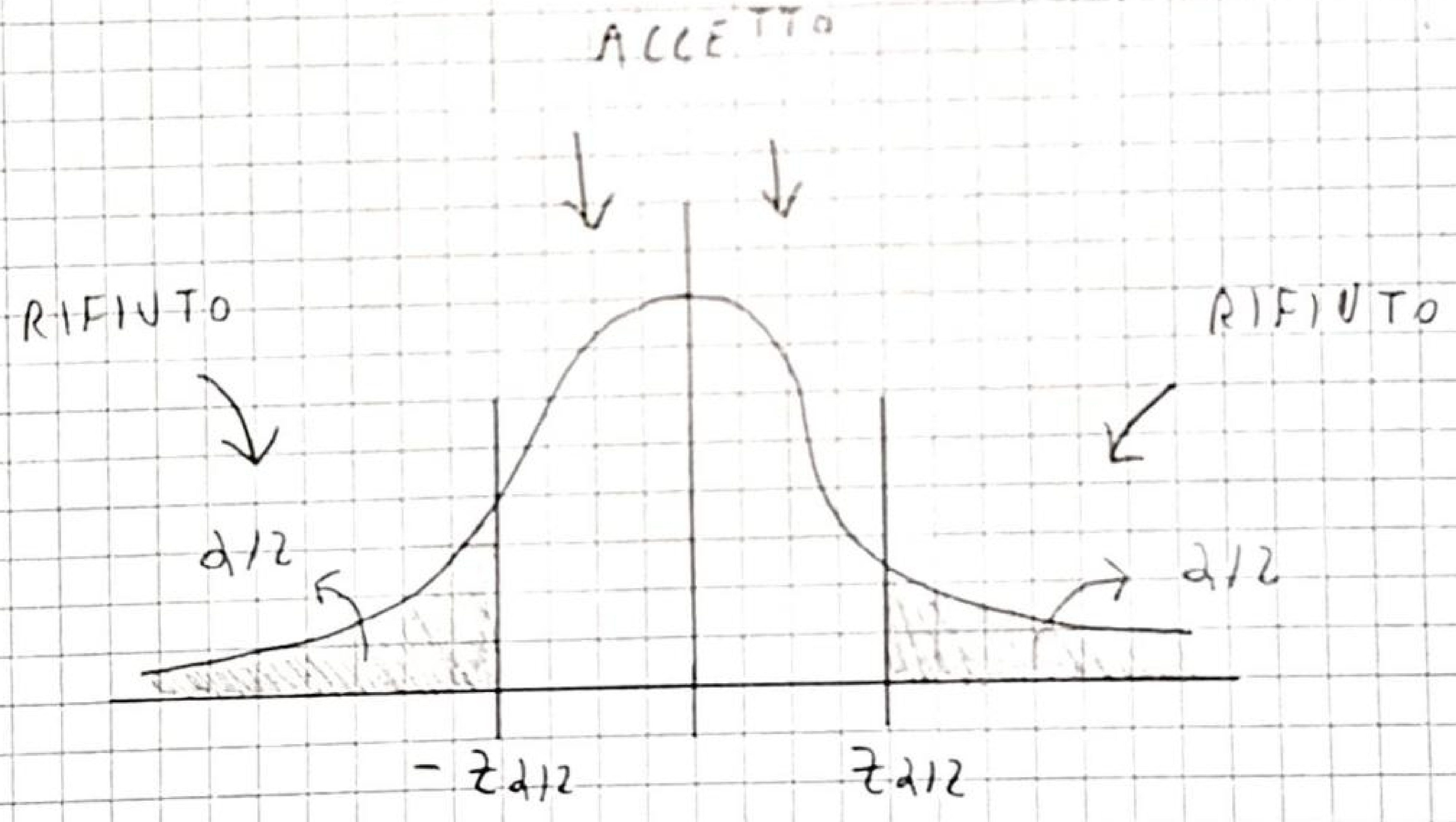


TEST BILATERALI (A DUE CODE)

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

REGIONE DI RIFIUTO : $z < -z_{\alpha/2} \vee z > z_{\alpha/2}$
REGIONE DI ACCETTAZIONE : $-z_{\alpha/2} < z < z_{\alpha/2}$



$$\frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

Come cambia la regione critica al variare di α?

All'aumentare di α la regione critica aumenta, e quindi diventa più "difficile" accettare H_0.

Ovviamente esiste un α critico che fa da spartiacque tra accettazione e rifiuto.

<u>Definizione</u>: Si dice P-VALUE (P-DEI-DATI) il più piccolo valore di α per cui i dati campionari consentono di rifiutare l'ipotesi nulla.

Un p-value quasi uguale a zero significa che siamo praticamente certi di non sbagliare rifiutando l'ipotesi nulla; un p-value dell'ordine dei soliti livelli di significatività (10%, 5% e 1%) indica che la decisione se rifiutare o no l'ipotesi nulla è critica; un p-value maggiore indica invece che, a qualsiasi livello ragionevole di significatività, sbagliamo rifiutando l'ipotesi nulla (in questo caso si può dire che il test ci porta ad accettare l'ipotesi).

Se $z_0 = \dfrac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}}$ inteso come

valore numerico (e non come variabile aleatoria)

(TAVOLE NORMALE STANDARD)

$$
\text{P - VALUE} \begin{cases} P(Z > z_0) & \text{TEST UNILATORALE CON : } H_0 : \mu \leq \mu_0 \\ P(Z < z_0) & \text{TEST UNILATERALE con : } H_0 : \mu \geq \mu_0 \\ 2P(Z > |z_0|) & \text{TEST BILATERALE con : } H_0 : \mu = \mu_0 \end{cases}
$$

Finora abbiamo considerato i test con σ^2 NOTA.

- <u>Se σ^2 NON E' NOTA</u>

$\sigma^2 \longrightarrow S^2$

$$
T = \frac{\bar{X}_n - \mu_0}{S / \sqrt{n}} \sim t_{n-1} \quad \left(\text{TEST DI STUDENT} \right)
$$

Per $n \geq 30 \longrightarrow$ si comporta come se fosse un campione NORMALE