Esempio: Verifica di ipotesi sulla media campionaria

[**Da Luccio e Caudek, *Statistica per Psicologi*, Laterza, 2001**]

Esempio 7.1. Supponiamo di trovarci in Africa, e di stare in un paese in cui esistono due etnie: l'etnia A, costituita da individui molto alti; i dati relativi alla leva ci dicono per i giovani che si tratta di una universo composto di persone di statura abbastanza alta: $\mu = 170$ cm, con $\sigma = 60$; vi è poi l'etnìa B, pigmoide, la cui statura ha $\mu = 140$ cm, con ancora $\sigma = 60$. Ora, si presenta a noi un gruppo di 25 giovani, che sostiene di appartenere all'etnìa A. Noi ne misuriamo la statura, e vediamo che la media di questo gruppo è di 150 cm.

Eseguire un test statistico appropriato per sottoporre a verifica:

$$H_0 : \mu = 170cm, \qquad \text{(Etnia A)}$$

contro l'ipotesi sostantiva unilaterale sinistra

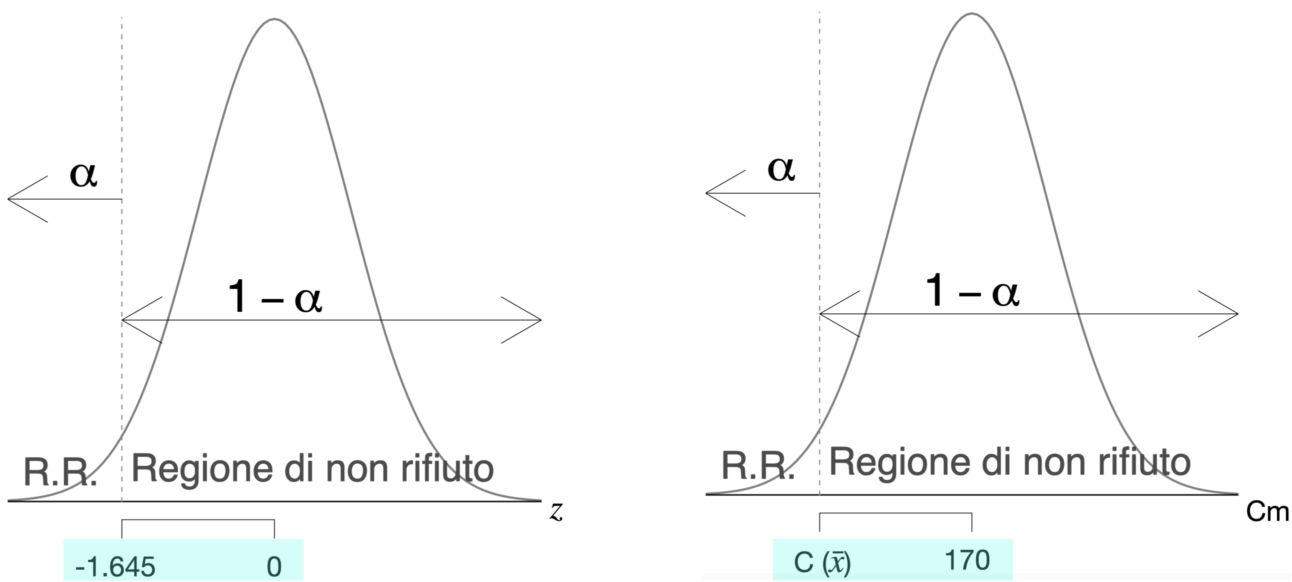
$$H_a : \mu < 170cm\,. \qquad \text{(Etnia B?)}$$

* Impostiamo il livello di significatività ad $\alpha = 0.05$

* Riportiamo i dati campionari disponibili:
  - n = 25;
  - $\bar{x} = 150$cm;
  - deviazione standard della popolazione delle altezze **nota**, con valore $\sigma = 60$

* Tracciamo la **distribuzione campionaria della nostra statistica** media campionaria, identificando chiaramente la regione di rifiuto e di non rifiuto **sotto** H_0, in base alle ipotesi nulla e sperimentale:

(1) **Ragionando in punti z** (2) **Ragionando in punti "grezzi" (Cm)**

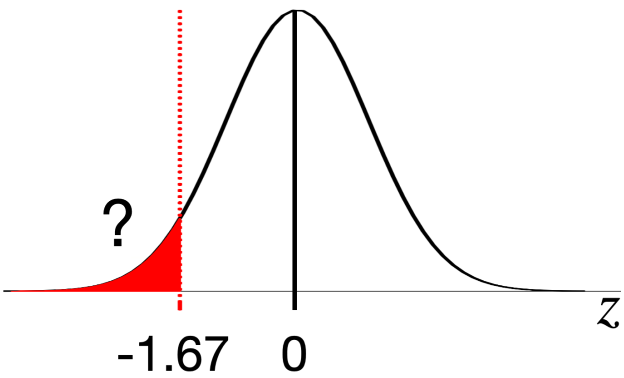


Esempio: Verifica di ipotesi sulla media campionaria

(1) Ragionando in punti z

Calcoliamo il punto z corrispondente ad una media campionaria pari a $\bar{x} = 150$cm, sotto l'ipotesi (*nulla*) che si tratti effettivamente di 25 ragazzi appartenenti all'etnia "A":

$$z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{150 - 170}{60/\sqrt{25}} = -1.67$$



Calcoliamo l'integrale della funzione di densità normale dal valore $-\infty$ a -1.67 ricorrendo alla tavola A:

p-valore = 0,0475 < 0.05

infatti, $z_{\bar{x}} = -1.67 < z_\alpha = 1.645$ cade nella regione di rifiuto (R.R.) dell'ipotesi nulla.

Decisione:

RIGETTIAMO
$H_0 : \mu = 170cm$,
A favore di
$H_a : \mu < 170cm$
con probabilità di Errore del Primo tipo pari al 5%.

Il risultato è statisticamente significativo rispetto all'ipotesi nulla: "*…i ragazzi sono troppo bassi (in media) per appartenere ragionevolmente al gruppo etnico A* $(z_{\bar{x}} = -1.67; \quad p < 0.05)$".

(2) Ragionando in punti "grezzi" (Cm)

Troviamo il valore critico (in Cm) per $\alpha = 0.05$ e confrontiamolo con il nostro valore medio campionario:

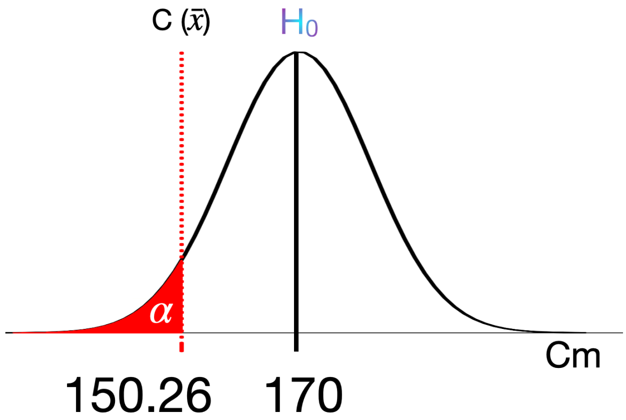
Dal punto z:

$$z_\alpha = -1.645$$

Al punteggio "grezzo" di altezza, sotto H_0:

$$C(\bar{x}) = -1.645\sigma/\sqrt{n} + \mu_0$$

$$= -1.645 \times 60/\sqrt{25} + 170 = 150.26$$



La probabilità di trovare una media campionaria di 150 cm o inferiore (coda sinistra) assumendo vera l'ipotesi H_0 è inferiore al livello di significatività prestabilito, di $\alpha = 0.05$.

infatti, $\bar{x} = 150 < C_{\bar{x}} = 150.26$ cade nella regione di rifiuto (R.R.) dell'ipotesi nulla.

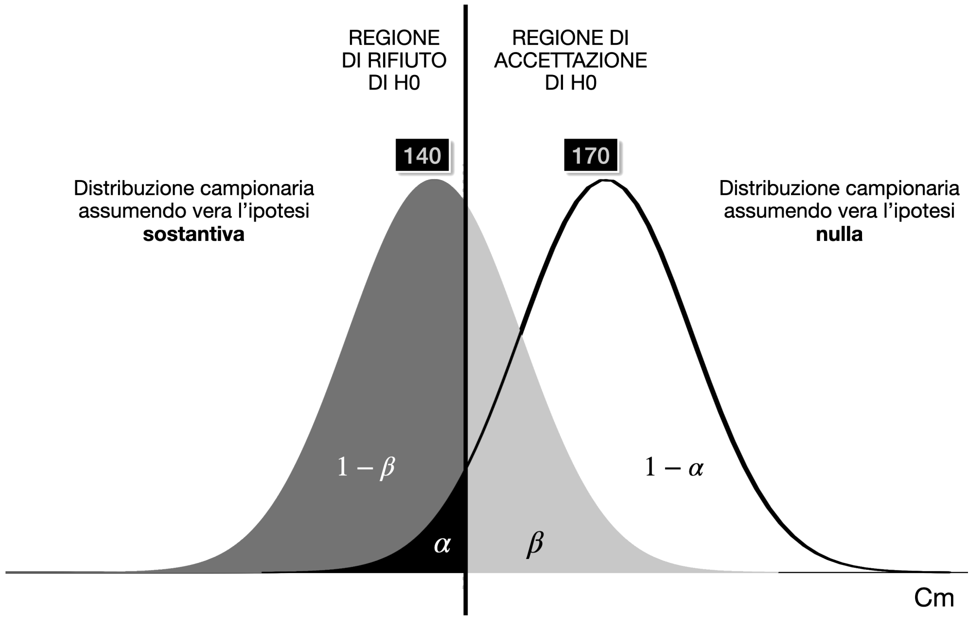
Cosa possiamo concludere?

"*…i 25 ragazzi sono troppo bassi (in media) per appartenere ragionevolmente al gruppo etnico A, infatti* ***l'evento osservato è sufficientemente improbabile****, ammesso che l'ipotesi nulla sia vera* $(\alpha = 0.05)$."

Esercizio: Potenza del test sulla media campionaria

Ipotizzando un valore di $H_a : \mu = 140cm$, con $\sigma = 60$ per l'ipotesi alternativa, qual'è la potenza del test $(1 - \beta)$ precedentemente condotto?

Tracciamo *anche* la distribuzione sotto H_a (140cm, Etnia B):



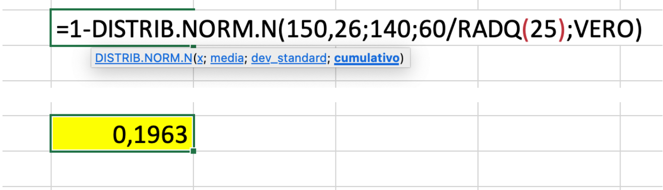
C ($\bar{x}$)=150.26

Avendo impostato precedentemente il test statistico con $\alpha = 0.05$ e punteggio Critico (C) in altezza media (Cm):

$$C(\bar{x}) = -z_\alpha \sigma / \sqrt{n} + \mu_0$$

$$= -1.645 \times 60/\sqrt{25} + 170 = 150.26$$

Per trovare l'area β dovremmo calcolare l'integrale verso dx, sotto H_a:



Oppure standardizzare il criterio del test (C=150.26) sotto H_a e consultare la tabella delle aree della normale standard (Tabl. A):

$$\frac{150.26 - \mu_a}{\sigma/\sqrt{n}} = \frac{150.26 - 140}{60/5} = 0.855$$

Esercizio: Potenza del test sulla media campionaria

a cui corrisponde un'area sulla coda destra di

$$P(z \geq 0.855) = 0.1963$$

La **potenza del test** statistico precedentemente condotto, assumendo $H_a : \mu = 140$cm e con $\sigma = 60$ è uguale a:

$$(1 - \beta) = 0.8037$$

e rappresenta la probabilità stimata di non commettere un errore del secondo tipo.

- Si tratta ovviamente di una stima, basata sull'assunzione $H_a : \mu = 140$ (…il dato disponibile per l'Etnìa B).