Esempio 8.3 p. 136 Luccio & Caudek.
(Commentato)

Supponiamo di volere sottoporre a test l'ipotesi nulla

$$H_0 : \mu = 100, \text{ con } \sigma = 25,$$

contro l'ipotesi sostantiva (alternativa, sperimentale…) unidirezionale destra

$$H_1 : \mu > 100, \text{ con } \sigma = 25$$

per la quale ipotizziamo in via del tutto eccezionale, sulla base di studi precedenti, il valore:

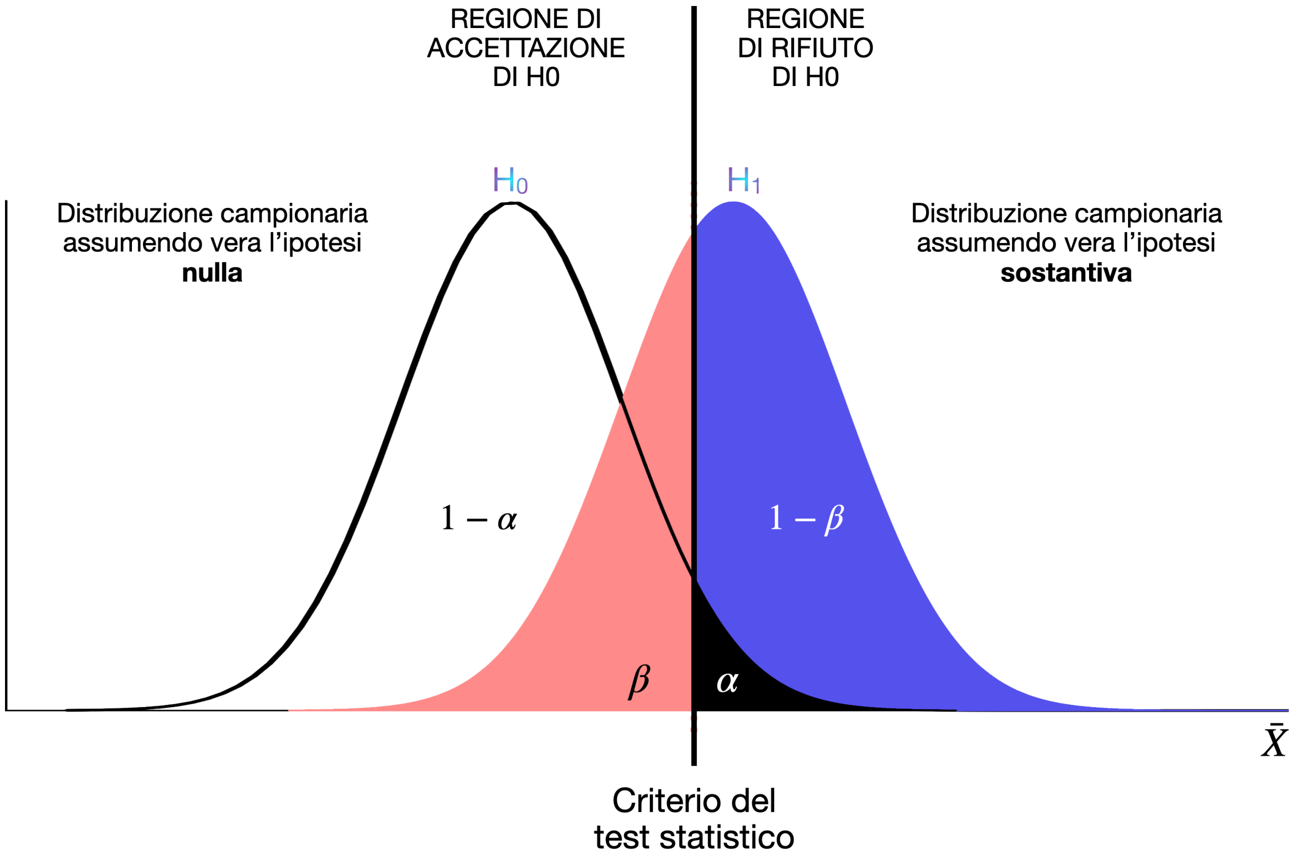
$$H_1 : \mu = 105, \text{ con } \sigma = 25,$$

con $\alpha = \beta = 0.05$.

Si trovi la grandezza del campione tale per cui i livelli di errore del primo e del secondo tipo siano quelli prefissati.

Osservazioni:

1) Fissando l'errore di primo tipo α, che è relativo all'area di rifiuto sull'area totale della distribuzione campionaria di $\bar{x}$ corrispondente all'ipotesi nulla, contemporaneamente abbiamo stabilito (sulla coda sinistra) della distribuzione dell'ipotesi alternativa anche il livello di β, ovvero l'errore di secondo tipo:



Ponendo un asse di decisione in corrispondenza di un valore tale per cui tutti i punteggi medi che cadono al di là di questo verranno ritenuti

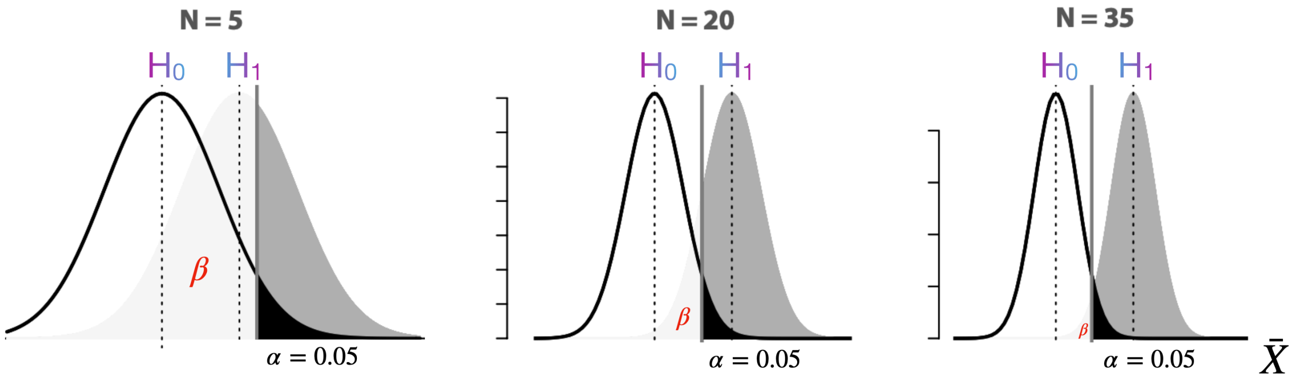
Esempio 8.3 p. 136 Luccio & Caudek.
(Commentato)

appartenenti all'area di rifiuto dell'ipotesi nulla (coda destra), e quindi attribuiti alla distribuzione dell'ipotesi sostantiva, contemporaneamente si pone un'area di rifiuto anche per la distribuzione relativa a H1 (area β, coda sinistra sotto H1).

2) Il livello dell'area β è eccessivo e la probabilità di commettere un errore di secondo tipo appare troppo elevata.

3) L'area β sebbene determinata dalla scelta del criterio del test, basato sull'area α dell'errore di primo tipo, tende progressivamente a ridursi all'aumentare di n in virtù della formula dell'errore standard della distribuzione campionaria di $\bar{x}$ (*consistenza della media campionaria, legge dei grandi numeri*):

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{(Eq. 7.6, Luccio)}$$



Se standardizziamo il valore di media campionaria che corrisponde al criterio del test per l'ipotesi nulla con $\alpha = 0.05$, otteniamo

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = z_\alpha = 1.645.$$

Allo stesso modo, per quanto riguarda l'ipotesi alternativa, avremo che un'area (sinistra) $\beta = 0.05$ sarà determinata da un valore z negativo

$$\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} = -z_\beta = -1.645.$$

- Stabiliti i valori α e β; stabiliti i punti z corrispondenti sotto H0 e H1,
- posto che ci siano noti i rispettivi valori attesi e che la deviazione standard sia σ;
- possiamo allora impostare il seguente sistema di equazioni, e risolverlo per il numero n di soggetti:

Esempio 8.3 p. 136 Luccio & Caudek.
(Commentato)

$$\begin{cases} \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} & = z_\alpha & = 1.645 \\ \dfrac{\bar{x}-\mu_1}{\sigma/\sqrt{n}} & = -z_\beta & = -1.645 \end{cases}$$

8.10 Luccio, p. 136

Dato che

$$\bar{x} = -z_\beta \sigma/\sqrt{n} + \mu_1,$$

possiamo riscrivere il sistema come

$$\begin{cases} \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} & = 1.645 \\ \bar{x} & = -1.645\sigma/\sqrt{n} + \mu_1 \end{cases}$$

sostituendo nella prima equazione

$$\frac{-1.645\sigma/\sqrt{n} + \mu_1 - \mu_0}{\sigma/\sqrt{n}} = 1.645$$

ricaviamo facilmente il valore per radice di n:

$$(\mu_1 - \mu_0) = 1.645\sigma/\sqrt{n} + 1.645\sigma/\sqrt{n}$$

$$(\mu_1 - \mu_0) = (1.645 + 1.645)\,\sigma/\sqrt{n}$$

$$\frac{1}{(\mu_1 - \mu_0)} = \frac{\sqrt{n}}{(1.645 + 1.645)\,\sigma}$$

$$\frac{(1.645 + 1.645)\,\sigma}{(\mu_1 - \mu_0)} = \sqrt{n}$$

con soluzione generale per il numero n di soggetti

$$\frac{\left(|z_\alpha| + |z_\beta|\right)^2 \sigma^2}{(\mu_1 - \mu_0)^2} = n$$

$$\frac{(1.645 + 1.645)^2\, 25^2}{(105 - 100)^2} = 270.6025$$

Saranno necessari n = 271 soggetti per assicurare che in questo test i livelli di errore del primo e del secondo tipo siano quelli prefissati (5%). Si tratta ovviamente di una stima, basata sull'assunzione che $H_1 : \mu = 105$.

Esempio 7.1 p. 115 Luccio & Caudek.
(`Commentato`)

1) Eseguire un test statistico appropriato per sottoporre a verifica:

$H_0 : \mu \geq 170cm,$ Etnia A

contro l'ipotesi sostantiva unidirezionale sinistra

$H_1 : \mu < 170cm\,.$

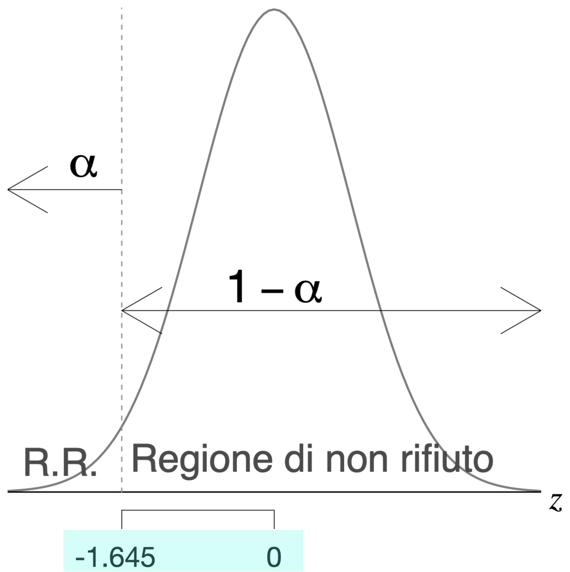
* Impostiamo il livello di significatività ad $\alpha = 0.05$

* Riportiamo i dati campionari disponibili:

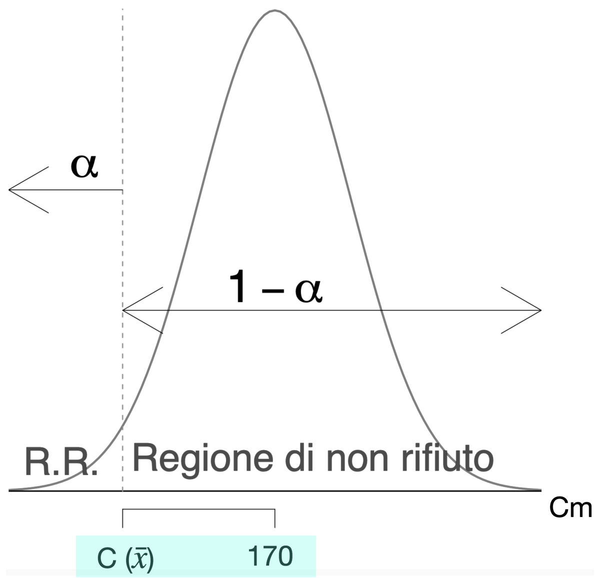
n = 25;
$\bar{x} = 150$cm;
deviazione standard della popolazione delle altezze **nota**,
con valore $\sigma = 60$

* Tracciamo la **distribuzione campionaria del nostro indice statistico** (media campionaria), identificando chiaramente la regione di rifiuto e di non rifiuto **sotto H0**, in base alle ipotesi nulla e sperimentale:
*

(1) **Ragionando in punti z**



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

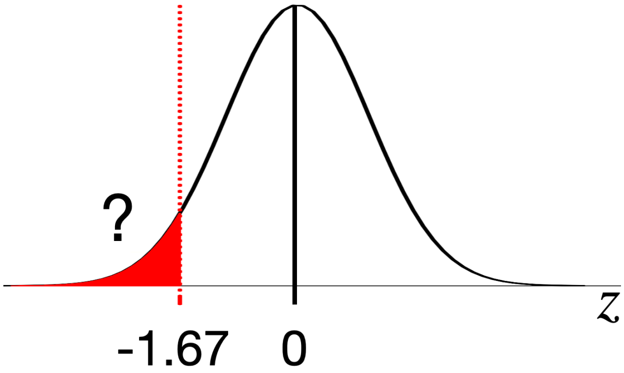


Esempio 7.1 p. 115 Luccio & Caudek.
(Commentato)

(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":

$$z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{150 - 170}{60/\sqrt{5}} = \text{-1.67}$$



Calcoliamo l'integrale della funzione di densità normale dal valore $-\infty$ a -1.67 ricorrendo alla tavola 1 Luccio:

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
$H_0 : \mu \geq 170cm,$
A favore di
$H_1 : \mu < 170cm$
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:

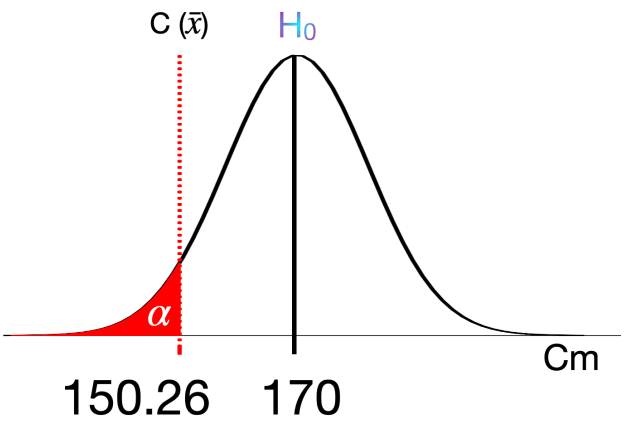
Dal punto z:

$$z_{\alpha} = -1.645$$

Al punteggio "grezzo" di altezza, sotto H0:

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

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



La probabilità di trovare una media campionaria di 150 cm o inferiore (coda sinistra) assumendo vera l'ipotesi H0 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* $(\alpha = 0.05)$."

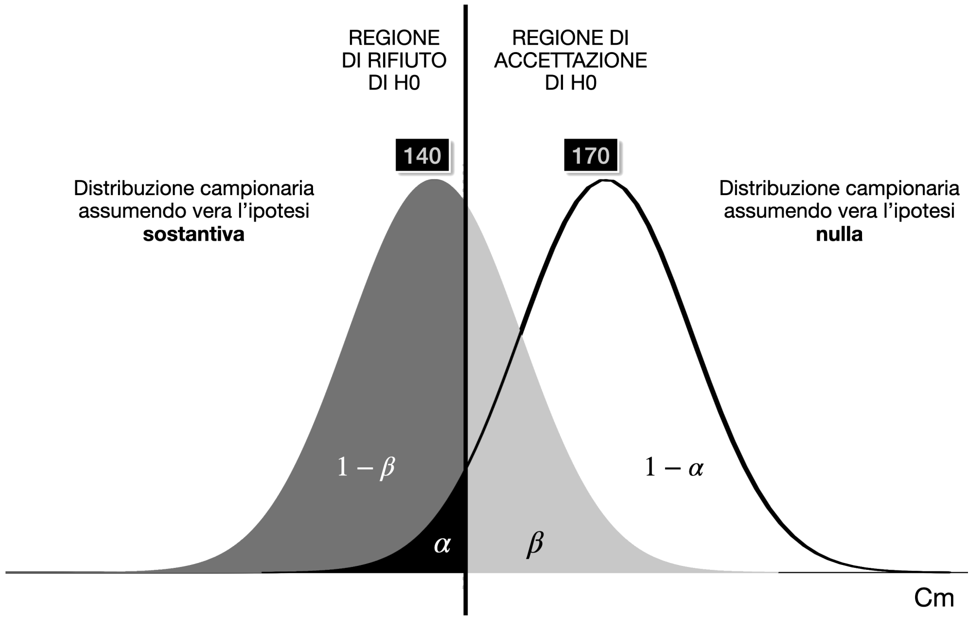
Esercizio: Potenza del test
(Sulla base di **7.1** p. 115 Luccio & Caudek)

(1) Ipotizzando un valore di $H_1 : \mu = 140cm,$ con $\sigma = 60$ per l'ipotesi alternativa, qual'è la potenza del test $(1 - \beta)$ precedentemente condotto?

(2) Quale valore di grandezza campionaria n sarebbe necessario per eguagliare $\alpha = \beta = 0.05$

---

(1) Tracciamo *anche* la distribuzione sotto H1 (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 H1:

```
=1-DISTRIB.NORM.N(150,26;140;60/RADQ(25);VERO)
DISTRIB.NORM.N(x; media; dev_standard; cumulativo)

0,1963
```

Oppure standardizzare il criterio del test (C=150.26) sotto H1 e consultare la tabella delle aree della normale standard:

Esercizio: Potenza del test
(Sulla base di **7.1** p. 115 Luccio & Caudek)

$$\frac{150.26 - \mu_1}{\sigma/\sqrt{n}} = \frac{150.26 - 140}{60/5} = 0.855$$

a cui corrisponde un'area sulla coda destra di

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

La **potenza del test** statistico precedentemente condotto, assumendo $H_1 : \mu = 140cm$ e con $\sigma = 60$ è uguale a:

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

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

(2) Circa il secondo punto, per eguagliare $\alpha = \beta = 0.05$

Impostiamo il sistema di equazioni:

$$\begin{cases} \dfrac{\bar{x} - 170}{60/\sqrt{n}} = & -z_\alpha = & -1.645 \\ \dfrac{\bar{x} - 140}{60/\sqrt{n}} = & z_\beta = & 1.645 \end{cases}$$

con soluzione nota:

$$n = \frac{\left(|z_\alpha| + |z_\beta|\right)^2 \sigma^2}{\left(\mu_1 - \mu_0\right)^2} = \frac{\left(1.645 + 1.645\right)^2 60^2}{(140 - 170)^2} = 43.2964.$$

- Saranno necessari n = 43 soggetti per assicurare che in questo test i livelli di errore del primo e del secondo tipo siano quelli prefissati (5%).
- Si tratta ovviamente di una stima, basata sull'assunzione $H_1 : \mu = 140$ (…il dato disponibile per l'Etnia B)