Statistica Inferenziale

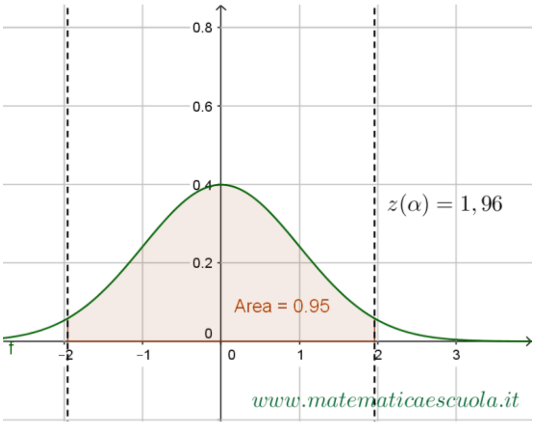
[1] **Problema**

Una Ditta produce dischetti il cui diametro dovrebbe essere di cm 22. Viene estratto un campione di 150 dischetti per il quale viene accertato un diametro medio di cm 21,8 con uno scarto quadratico medio di cm 0,5. Adottando un livello di significatività del 5% si vuole verificare se la produzione è sotto controllo.

Risoluzione

Il problema in esame rientra in quelli con il campione da considerare numeroso e per la risoluzione faremo riferimento al modello della curva gaussiana.

- Il valore medio del diametro dei dischetti prodotti dall'azienda deve essere M=22 cm. Questa è l'ipotesi nulla H_0 che riteniamo sia vera. L'ipotesi alternativa bilaterale è H1:M≠22 cm. Il livello di significatività del test è α=5%, quindi il livello di fiducia è 1-α=0,95=95%.
- L'area della regione esterna della gaussiana è α=0,05, quindi, per la simmetria della curva rispetto all'asse delle ordinate, deduciamo che ciascuna delle code esterne ha area 0,05/2=0,025.
- Ricerca dei punti -z(α), z(α), che delimitano l'intervallo di confidenza.



  Il punto z(α)>0 è tale che

  $$P\left(z \le z(\alpha)\right) = 1 - 0,025 = 0,975$$

- Dal confronto della tabella della gaussiana standardizzata deduciamo che z(α)=1,96. Infatti in corrispondenza di 1,96 figura il numero 0,9750 . Osserviamo che nella figura riportata (a scopo didattico) si evince che l'area del sottografico della gaussiana relativo all'intervallo di confidenza [-1,96;1,96] è 0,95.
- **Occorre ora determinare il valore standardizzato corrispondente al valore medio del campione** estratto $\bar{x} = 21,8cm$ per stabilire se rientra nell'intervallo di confidenza]-1,96;1,96[; se si verifica questo caso allora si accetterà l'ipotesi nulla H_0, altrimenti si rifiuterà. Precisiamo che l'area di rischio sull'accettazione della veridicità dell'ipotesi nulla è quella esterna all'intervallo di confidenza (per il quale si ha fiducia che al 95% l'ipotesi fatta su H_0 sia vera) e se il valore standardizzato del valore medio $\bar{x} = 21,8cm$ del campione cade nell'area di rischio è consigliabile non accettare come vera l'ipotesi nulla. E' bene precisare altresì che in questa eventualità si suole affermare che "si commetterebbe un errore del primo tipo (o di prima specie)" e ciò perché si scarterebbe un'ipotesi ritenuta vera che invece noi scarteremmo ritenendola falsa; in realtà la stessa potrebbe essere vera, sebbene con probabilità del 5%.

  Il valore da calcolare è

[1] Il testo di questo problema è riportato come n.47 a Pag. 281 del volume Nuovi elementi di matematica - Calcolo delle probabilità e statistica inferenziale, A.A. M. Trovato, R. Manfredi- Ghisetti e Corvi Editori

$$z\left(\bar{x}\right)=\frac{\bar{x}-M}{\frac{\hat{\sigma}}{\sqrt{150}}}$$

dove M=22 cm è il valore medio dell'universo dei dischetti prodotti dalla Ditta e $\hat{\sigma}$ è lo **scarto quadratico medio corretto del campione** (relativo alle n=150 unità estratte) .

*** ***

Note teoriche di riferimento

1) Il valore standardizzato di un valore x, cioè il calcolo del punto z(x) corrispondente sull'asse delle ascisse per la **curva gaussiana normale** (valore medio m=0 e scarto q.m.=σ=1), si ottiene dividendo lo scarto di quel valore x dal valore medio M dell'universo dei dati per il rapporto tra lo scarto quadratico medio corretto del campione e la radice quadrata della numerosità del campione. In simboli

$$z\left(x\right)=\frac{x-M}{\frac{\hat{\sigma}}{\sqrt{n}}}.$$

2) Ricordiamo che $\hat{\sigma}$ è definito come la radice quadrata della somma dei quadrati degli n scarti dei valori x_i del campione dal loro valore medio $\bar{x}$ divisa per il numero dei dati diminuito di una unità. Dunque risulta

$$\hat{\sigma}=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n-1}}$$

3) Sussiste un'importante relazione tra lo scarto quadratico medio corretto $\hat{\sigma}$ e lo scarto quadratico medio σ del campione da utilizzare quando sono noti lo s.q.m. del campione e la sua numerosità n. La presentiamo:

$$\hat{\sigma}=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n-1}}=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n}\cdot\frac{n}{n-1}}=\sqrt{\frac{n}{n-1}}\cdot\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n}}=\sqrt{\frac{n}{n-1}}\cdot\sigma$$

*** ***

- Calcoliamo ora il valore standardizzato corrispondente al valore medio del campione.

Valore dello scarto quadratico medio corretto del campione

$$\hat{\sigma}=\sqrt{\frac{150}{149}}\cdot 0{,}5cm==\simeq 0{,}502cm$$

Calcolo del valore standardizzato relativo al valore medio del campione

$$z\left(\overline{x}\right)=\frac{\overline{x}-M}{\frac{\hat{\sigma}}{\sqrt{150}}}=\frac{\left(21,8-22,0\right)cm}{\frac{0,502cm}{\sqrt{150}}}\simeq -4,87946$$

Il valore standardizzato ottenuto è abbondantemente al di fuori dell'intervallo di confidenza

]-1,96;1,96[, quindi **a livello di significatività del 5% l'ipotesi H_0 va rifiutata** e quindi si conclude che la qualità della produzione dei dischetti non sia affatto sotto controllo.