STATISTICA A – K (63 ore)

Marco Riani

mriani@unipr.it

http://www.riani.it

Esercizio

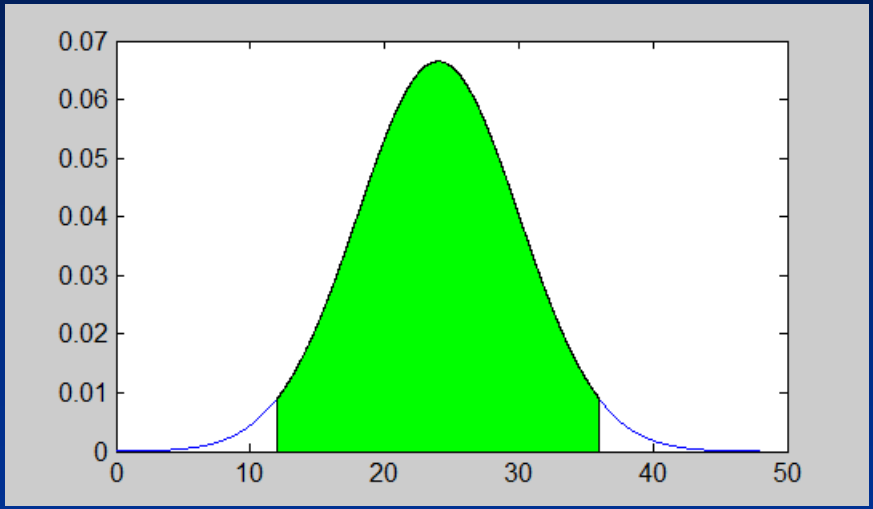
- La durata di un macchinario si distribuisce secondo una distribuzione normale di media 2 anni e scarto quadratico medio 0,5 anni. Si determini:

1. prob che il macchinario duri più di 28 mesi.
2. l'intervallo di ampiezza 2 anni al quale corrisponde la massima prob di contenere la durata effettiva del macchinario. Calcolare tale probabilità.
3. Se il costo di acquisto del macchinario è di 1000 euro e il costo del suo funzionamento è stimato in 150 euro all'anno, si calcolino la media e la varianza del costo complessivo del macchinario.

Soluzione

- T= v.a. che descrive la durata del macchinario T~N(24 mesi 6^2 mesi)
- Pr(T>28)?
- Pr(T>28)=1-Pr(T<28)=1-F(4/6)=0,25249

Intervallo di ampiezza 2 anni al quale corrisponde la massima prob di contenere la durata effettiva del macchinario.



T= v.a. che descrive la durata del macchinario
T~N(24 mesi 6^2 mesi)

Dalla forma campanulare e simmetrica attorno a μ della densità di una generica N(μ, σ^2), si ottiene che l'intervallo di ampiezza 2 anni che contiene la massima probabilità per una N(24,6^2) è l'intervallo di ampiezza 2 anni attorno alla media (E(T) = 24), ossia [12 mesi,36 mesi].

Pr(12 mesi ≤T≤36 mesi)

- Dato che T~N(24 mesi 6^2 mesi)

- Pr(12 <T<36)
- =Pr(-2<(T-E(T))/σ(T)<2)=0,9545

Media e varianza del costo complessivo del macchinario

- C= v.a. che descrive il costo complessivo
- CA=costo acquisto = 1000 €
- CM = costo manutenzione annuo =150 €
- T = v.a. che descrive la durata in mesi
- $C=CA+(CM/12) \times T$
- $C=1000+(25/2)\ T$ con $T \sim N(24m\ \ 6^2m)$
- $E(C) = ?$ $E(C)= 1000+ (25/2)\ E(T)=1300$
- $VAR(C)=?$ $VAR(C)= (25/2)^2 VAR(T)= 5625$

Se avessi espresso tutto in anni

- C= v.a. che descrive il costo complessivo
- CA=costo acquisto = 1000 €
- CM = costo manutenzione annuo =150 €
- T_A = v.a. che descrive la durata in anni
- $C=CA+CM \times T_A$
- C=
- E(C) = 1000+ 150 $E(T_A)$=1300
- VAR(C)= 150^2 $VAR(T_A)$= 5625

Esercizio

- Il tempo impiegato da un meccanico in un negozio di biciclette per assemblare un certo tipo di bicicletta può essere considerato una v.c. normale con media 32 minuti e deviazione standard 3,5 minuti. Si calcoli la probabilità che il tempo medio per assemblare 10 biciclette
  - Non superi 33 minuti
  - Sia compreso tra 28,5 e 31,5 minuti

Soluzione

- X=v.c. tempo impiegato
- $X \sim N(32, 3{,}5^2)$ n=10

$$\bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$$

$$\bar{X} \sim N(32, \frac{3{,}5^2}{10})$$

- Pr media campionaria <33?

$$\Pr(\bar{X} < 33) = \Pr(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < \frac{33 - 32}{3{,}5 / \sqrt{10}})$$

$$\Pr(\bar{X} < 33) = \Pr(Z < 0{,}9035) = 0{,}8169$$

Il valore 0.8169 è stato ottenuto dalla funzione di Excel =DISTRIB.NORM.ST(0,9035). Utilizzando le tavole F(0,90)=0,81594

$$\overline{X} \sim N(32, \frac{3{,}5^2}{10})$$

- Calcolo di

$$\Pr(28{,}5 < \overline{X} < 31{,}5) = ?$$

$$\Pr(\overline{X} < 31{,}5) = \Pr(Z < -0{,}45175) = 0{,}32572$$

$$\Pr(\overline{X} < 28{,}5) = \Pr(Z < -3{,}16228) = 0{,}00078$$

$$\Pr(28{,}5 < \overline{X} < 31{,}5) = 0{,}32494$$

I valori 0,32572 e 0,00078 sono stati ottenuti con le funzioni di Excel =DISTRIB.NORM.ST(-0,45175) e =DISTRIB.NORM.ST(-3,16228).

Utilizzando le tavole si ottiene F(-0,45)-F(-3,16)= 0.32636- 0.00079=0.32557

Esercizio

- Sia $f(x)=1/2 \qquad -1<x<1$
- Si calcoli
- $E(X) \qquad E(X+2)$
- $E(X^2) \qquad \sigma^2 \qquad E(X/4+7)$

Soluzione

- Sia f(x)=1/2 $-1<x<1$
- E(X)=0
- E(X+2)=?
- E(X+2)=2
- $E(X^2) = 1/3$ $\sigma^2 = 1/3$
- E(X/4+7)=?
- E(X/4+7)=7

Esercizio

- Una lotteria mette in palio uno scooter del valore di 3000 Euro. Vengono venduti 10000 biglietti al prezzo di 1€. Se si acquista un biglietto qual è il guadagno atteso? Qual è il guadagno atteso se si comperano 100 biglietti. Si confronti la varianza del guadagno nei due casi

Esempio

Distribuzione della v.c.
X = “guadagno”

x_i	p_i
- 1	9999/10000
2999	1/10000
	1

E(X) = -7/10 € guadagno atteso se si acquista un biglietto

E(100*X) =-70 € guadagno atteso se si acquistano 100 biglietti

Esempio

Distribuzione della v.c.
X = “guadagno”

x_i	p_i
- 1	9999/10000
2999	1/10000
	1

VAR(X) = 899.91 €² se si acquista un biglietto
VAR(100*X) =10000VAR(X)= 899910 €² se si acquistano 100 biglietti

Esercizio: il gioco dell'intruso (odd man game)

- 3 persone giocano all'«odd man game». Ciascuno lancia una moneta. Chi ottiene una faccia diversa da quella degli altri due è l'intruso («odd man») e perde.
- Qual è la probabilità che via sia un intruso in un determinato turno di gioco assumendo che le monete non siano truccate?
- Qual è la probabilità che siano necessari un numero di turni pari di gioco per determinare il perdente («l'odd man»)?

Soluzione

- Spazio degli eventi
- Ω={TTT, TTC, TCT, CTT, CCT, CTC, TCC, CCC}

- Casi favorevoli che determinano la conclusione del gioco al primo turno
- {TTC, TCT, CTT, CCT, CTC, TCC}
- P(vi sia un intruso) = 3/4

Probabilità che siano necessari un numero di turni pari di gioco per determinare il perdente («l'odd man»)

- Pr conclusione turno 2 = (1/4)(¾)
- Pr conclusione turno 4 = $(1/4)^3(¾)$
- Pr conclusione turno 6 = $(1/4)^5(¾)$
- …..
- Pr conclusione turno pari=

$$\frac{3}{4}\sum_{j=0}^{\infty}\left(\frac{1}{4}\right)^{2j+1} = \frac{3}{16}\sum_{j=0}^{\infty}\left(\frac{1}{4}\right)^{2j} = \frac{3}{16}\sum_{j=0}^{\infty}\left(\frac{1}{16}\right)^{j} = \frac{3}{16}\frac{1}{1-\frac{1}{16}} = \frac{1}{5}$$

Esercizio: il gioco dell'intruso (odd man game)

- Si risponda ai quesiti dell'esercizio precedente assumendo stavolta che il numero dei giocatori sia uguale a 4 (in questo caso «l'odd man» è quello che ottiene una faccia diversa da quella degli altri 3).

Soluzione

- Spazio degli eventi
- Ω={16 possibili casi}

- Casi favorevoli che determinano la conclusione del gioco al primo turno

{CTTT, TCTT, TTCT, TTTC

TCCC, CTCC, CCTC, CCCT}

- P(vi sia un intruso) = 1/2

Probabilità che siano necessari un numero di turni pari di gioco per determinare il perdente («l'odd man»)

- Pr conclusione turno 2 = (1/2)(1/2)
- Pr conclusione turno 4 = $(1/2)^3(1/2)$
- Pr conclusione turno 6 = $(1/2)^5(1/2)$
- …..
- Pr conclusione turno pari=

$$\frac{1}{2}\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^{2j+1} = \frac{1}{4}\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^{2j} = \frac{1}{4}\sum_{j=0}^{\infty}\left(\frac{1}{4}\right)^{j} = \frac{1}{4}\frac{1}{1-\frac{1}{4}} = \frac{1}{3}$$

Esercizio: il gioco dell'intruso (odd man game)

- Si risponda ai quesiti dell'esercizio precedente assumendo stavolta che il numero dei giocatori sia uguale a n (in questo caso «l'odd man» è quello che ottiene una faccia diversa da quella degli altri n-1).
- Does this seem like a feasible game as n gets large?

Soluzione

- Spazio degli eventi Ω?
- Ω={2^n possibili casi}

- Casi favorevoli che determinano la conclusione del gioco al primo turno

{CTT…T, TCT…T, …, TT…TC

TCC…C, CTC…C, …., CC…CT}

P(vi sia un intruso) = $2n/ 2^n = n/ 2^{n-1}$

Probabilità che siano necessari un numero di turni pari di gioco per determinare il perdente («l'odd man»)

- Pr conclusione turno 2 = $(1-n/2^{n-1})(n/2^{n-1})$
- Pr conclusione turno 4 = $(1-n/2^{n-1})^3(n/2^{n-1})$
- Pr conclusione turno 6 = $(1-n/2^{n-1})^5(n/2^{n-1})$
- .....
- Pr conclusione turno pari=

$$\frac{n}{2^{n-1}} \sum_{j=0}^{\infty} \left(1 - \frac{n}{2^n - 1}\right)^{2j+1}$$

Probabilità che siano necessari un numero di turni pari di gioco per determinare il perdente («l'odd man»)

- Pr conclusione turno pari=

$$\frac{n}{2^{n-1}}\sum_{j=0}^{\infty}\left(1-\frac{n}{2^{n}-1}\right)^{2j+1}=\frac{n}{2^{n-1}}\left(1-\frac{n}{2^{n}-1}\right)\sum_{j=0}^{\infty}\left(1-\frac{n}{2^{n}-1}\right)^{2j}$$

$$=\frac{n}{2^{n-1}}\left(1-\frac{n}{2^{n-1}}\right)\sum_{j=0}^{\infty}\left[\left(1-\frac{n}{2^{n-1}}\right)^{2}\right]^{j}=\frac{n}{2^{n-1}}\left(1-\frac{n}{2^{n-1}}\right)\frac{1}{1-\left(1-\frac{n}{2^{n-1}}\right)^{2}}$$

Does this seem like a feasible game as n gets large?

- P(vi sia un intruso) = $2n/ 2^n = n/ 2^{n-1}$

- $n/ 2^{n-1} \rightarrow 0$ se $n \rightarrow \infty$

$$= \lim_{n\to\infty} \frac{n}{2^{n-1}} \left(1 - \frac{n}{2^{n-1}}\right) \frac{1}{1-\left(1 - \frac{n}{2^{n-1}}\right)^2} = 0$$

Esercizio

- Un fornitore di pneumatici sostiene che la durata media di un certo tipo di pneumatici per camion è di 45000 Km. Un'impresa sottopone a test l'affermazione del produttore osservando un campione di 56 pneumatici utilizzati dai propri veicoli.
- Qual è la conclusione a cui giunge l'impresa se trova una durata media di 43740 con un s_{cor}=2749 km (si ponga α=0,01)
- Si calcoli il p-value

Soluzione

H_0: μ = 45000 Km

H_1: μ < 45000 $\Rightarrow$ la durata effettiva dei pneumatici è inferiore a quella dichiarata

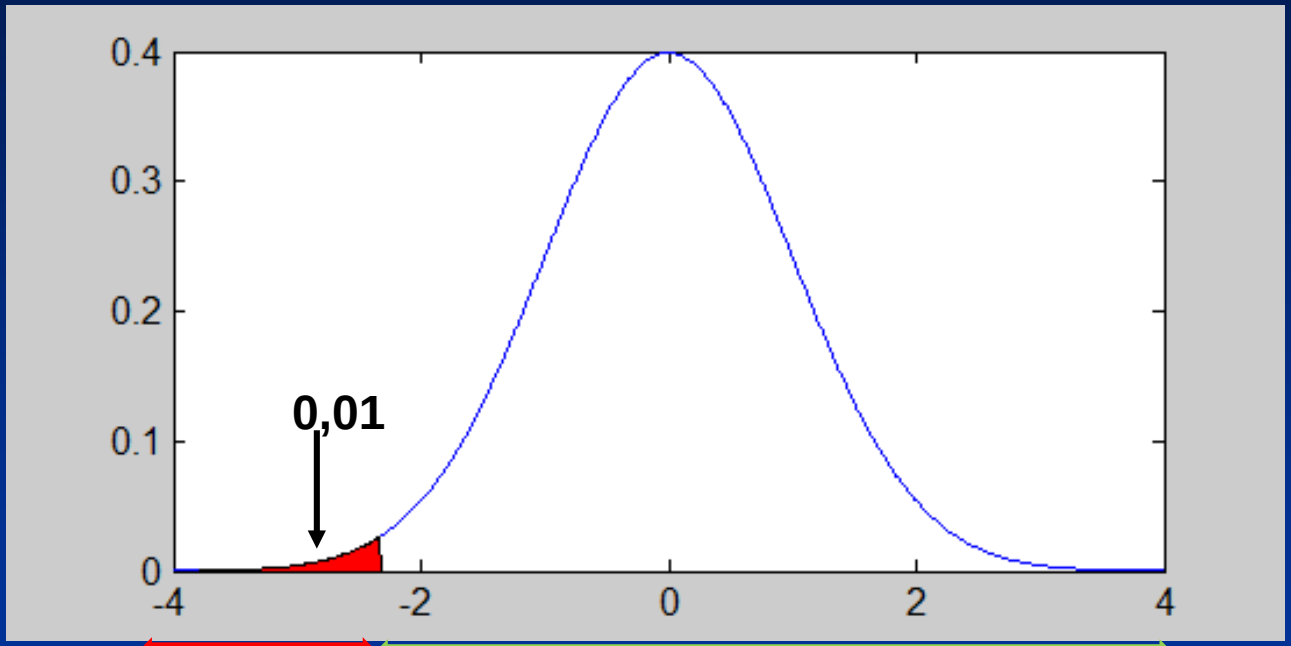
$\bar{x} = 43740$ s_{cor}=2749 n=56

Teorema centrale del limite

$$Z(\bar{X}) = \frac{\bar{X} - \mu_0}{s_{cor} / \sqrt{n}} \sim \mathrm{N}(0,1)$$

$$Z(\bar{x}) = \frac{43740 - 45000}{2749 / \sqrt{56}} = -3{,}43$$

H_1: $\mu < 45000$ α=0,01 F(-2,33)=0,01



Il valore osservato del test (-3,43) cade nella zona di rifiuto

p-value = F(-3,43) = 0,0003

Esercizio

- Di seguito sono riportati i dati di durata (in migliaia di Km) di un convertitore catalitico in un campione di 15 osservazioni.
- 115,4 85,2 89,1 118,3 88,4 109,3 104,3 69,3 105,5 106,8 103,1 101,6 102,9 89,6 109,3
- Si verifichi l'ipotesi che la durata media sia pari a 100 contro l'alternativa che essa sia minore. Si assuma un livello di significatività $\alpha=0,05$. Si calcoli il p-value del test.

Soluzione

H_0: μ = 100

H_1: μ < 100 $\Rightarrow$ la durata media effettiva è inferiore a 100000 Km

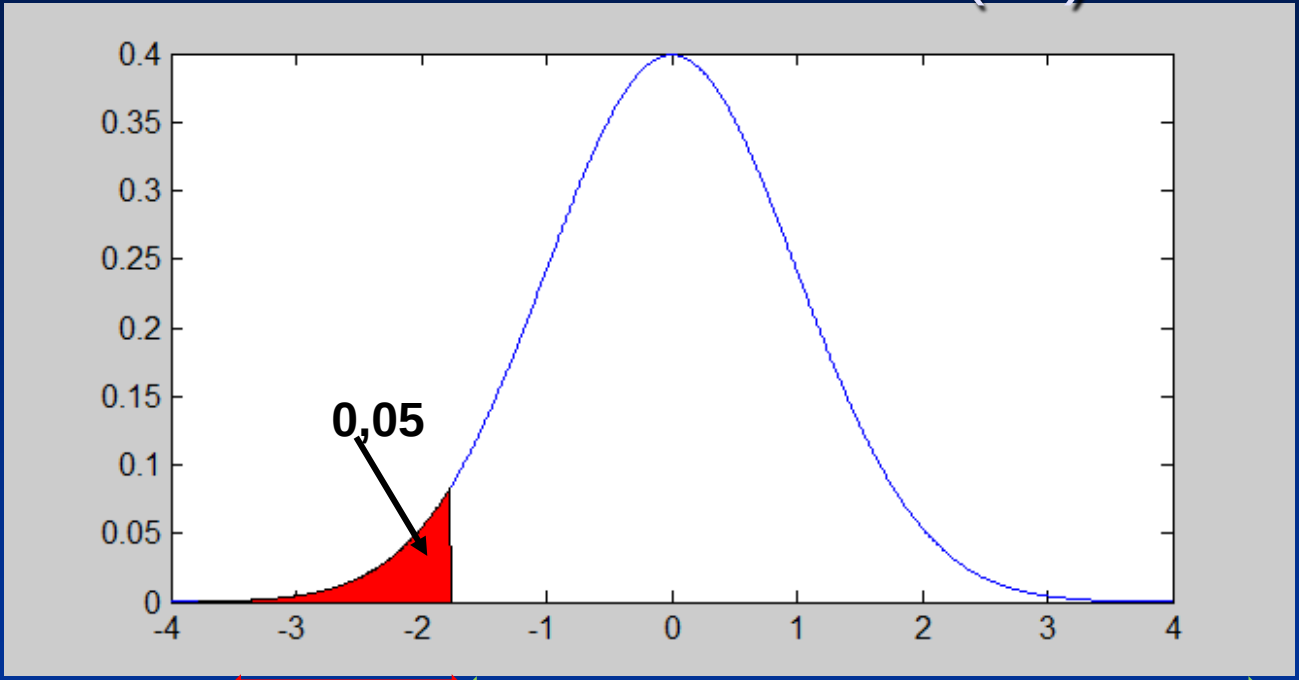
$\bar{x} = 99{,}87$ s_{cor}=13,05 n=15

Ip. di distribuzione normale dell'universo

$$Z(\bar{X}) = \frac{\bar{X} - \mu_0}{s_{cor} / \sqrt{n}} \sim t(14)$$

$$Z(\bar{x}) = \frac{99{,}87 - 100}{13{,}05 / \sqrt{15}} = -0{,}0376$$

H_1: μ <100 α=0,05 $F_{t(14)}$(-1,761)=0,05



Il valore osservato del test (-0,0376) cade nella zona di accettazione

Esercizio

L’Istituto Superiore di Sanità ha stimato che le spese a carico del Sistema Sanitario Nazionale per la riabilitazione di un paziente che ha avuto un ictus è di 42372 euro. L’amministrazione di una ASL, per verificare se i costi nella ASL sono in linea con la media nazionale, ha raccolto le informazioni sul costo della riabilitazione di 64 pazienti. Il costo medio è risultato pari a 44143 euro con uno scarto quadratico medio (campionario) corretto di 9156 euro.

- (a) Calcolare l'intervallo di confidenza al livello del 99% per la vera media dei costi nell’ASL considerata.
- (b) Dopo aver impostato l'ipotesi nulla e l'ipotesi alternativa, si testi se la differenza tra il costo medio nazionale e il costo medio stimato nell’ASL è significativa al livello di significatività dell'1%. Commentare i risultati ottenuti.
- Come sarebbero cambiate le conclusioni se il livello di significatività fosse stato del 10%?

Soluzione (a)

Dati a priori

$\mu_0 = 42372$

Dati campionari

$\overline{x} = 44143 \quad n = 64$

$s_{corr} = 9156$

(a) Calcolare l'intervallo di confidenza al livello del 99% per la vera media dei costi nell'ASL considerata.

$$P\left\{\overline{X} - z(\alpha)\frac{s_{cor}}{\sqrt{n}} \leq \mu \leq \overline{X} + z(\alpha)\frac{s_{cor}}{\sqrt{n}})\right\} = 1 - \alpha$$

z(α)=2,58 n>30 (Teorema centrale del limite)

$$Pr(41190,19 < \mu < 47095,81) = 0,99$$

Si noti che il valore a priori 42372 è compreso nell'intervallo di confidenza

Soluzione (b)

H_0: $\mu = \mu_0$ =42372

H_1: $\mu >$ 42372 $\Rightarrow$ l'ASL presenta un costo medio significativamente superiore a quello della media nazionale α=0,01 α=0,1

$\overline{x} = 44143$ $s_{corr} = 9156$ $n = 64$

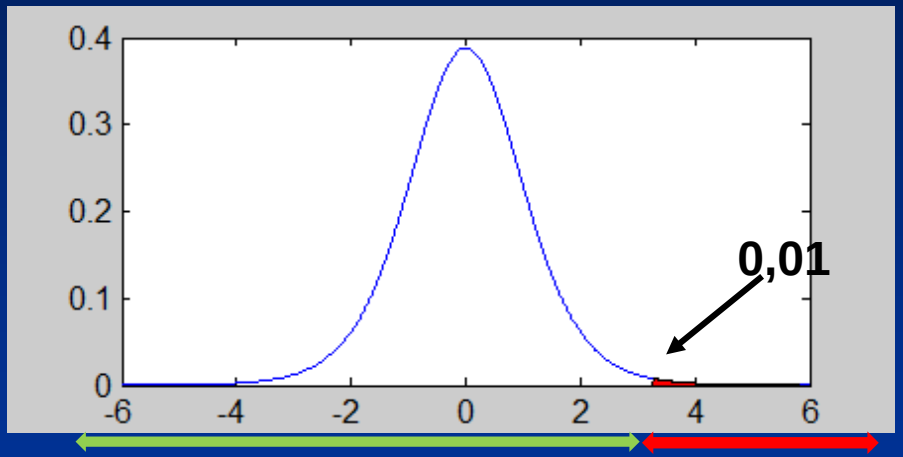
Teorema centrale del limite

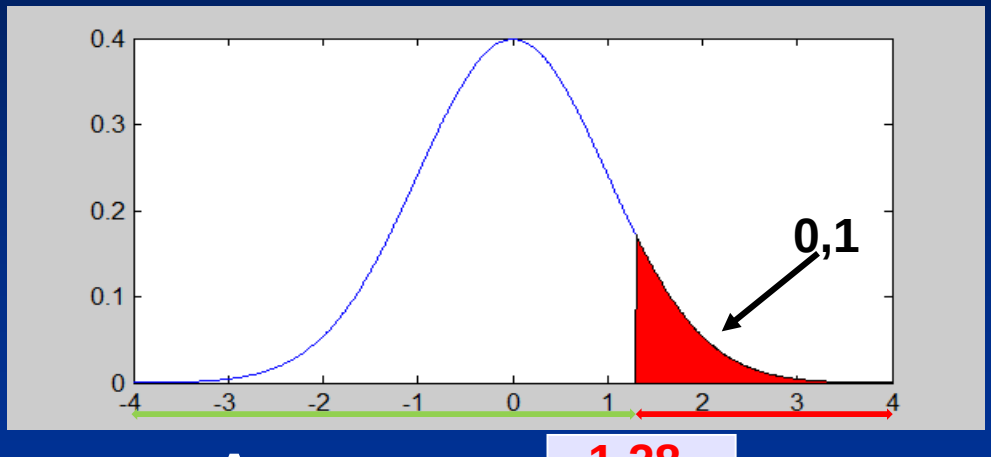
$$Z(\overline{X}) = \frac{\overline{X} - \mu_0}{s_{cor} / \sqrt{n}} \sim \mathrm{N}(0,1)$$

$$Z(\overline{x}) = \frac{44143 - 42372}{9156 / \sqrt{64}} = 1{,}547$$

H_1: $\mu > 42372$

α=0,01 F (2,33)=0,99 **α=0,1** F (1,28)=0,90



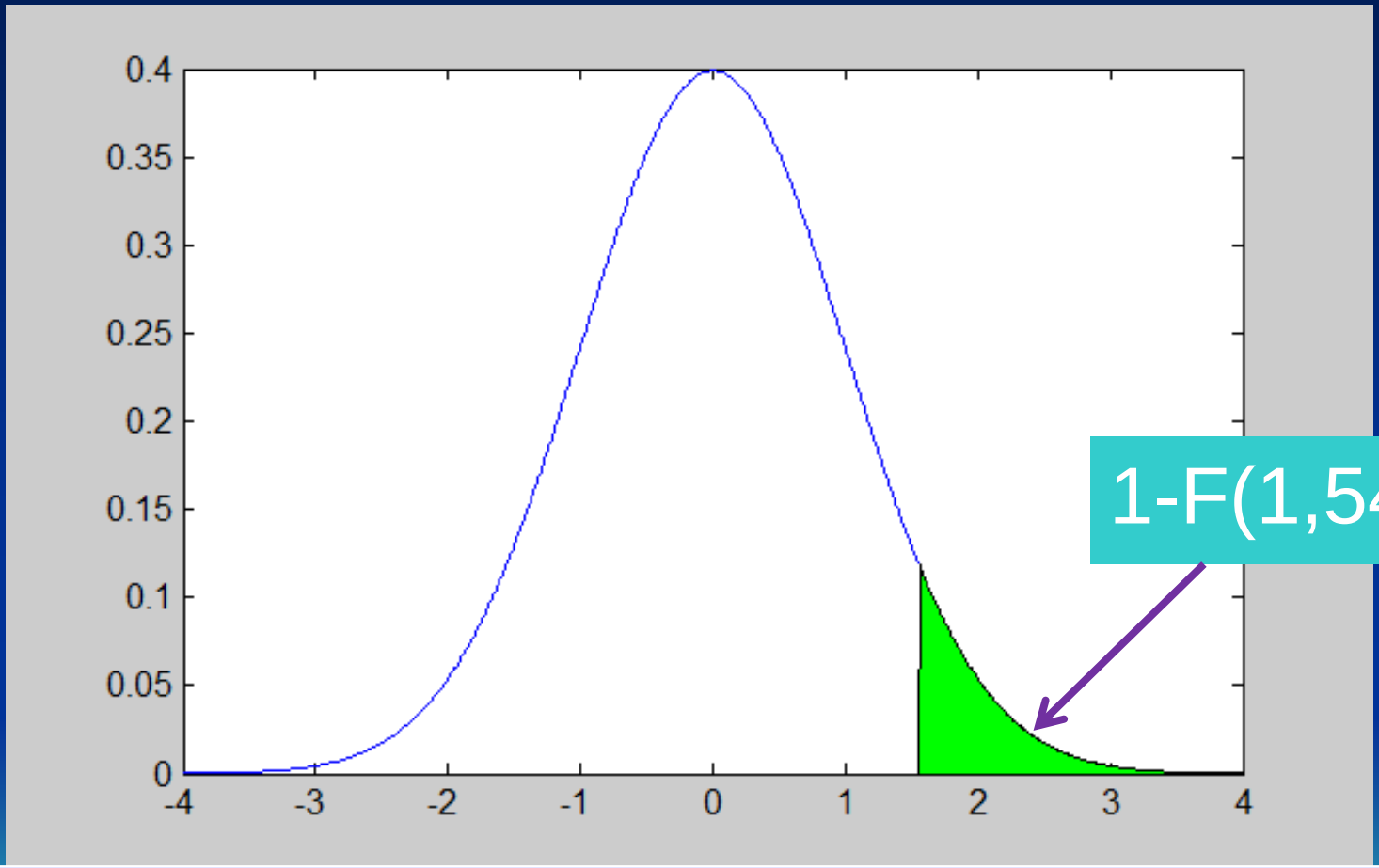


$$Z(\bar{x}) = 1{,}547$$

Con **α=0,01 i**l valore osservato del test cade nella zona di accettazione

Con **α=0,1 i**l valore osservato del test cade nella zona di rifiuto

Calcolo del p-value H_1: μ > 42372



0 **1,547**

1-F(1,547)= 0,061

Esercizio

Si assuma che la pressione sistolica media di un adulto sano sia 120 (mm Hg) e lo scarto quadratico medio 5,6.
Assumendo che la pressione abbia una distribuzione normale calcolare la probabilità che:

- selezionando un individuo sano scelto a caso questi abbia una pressione sistolica superiore a 125;
- scegliendo a caso 4 individui, la media della loro pressione sistolica sia superiore a 125;
- scegliendo a caso 25 individui, la media della loro pressione sistolica sia superiore a 125;
- selezionando 6 individui sani quattro di essi abbiano una pressione inferiore a 125.

Soluzione

- X= v.c. che rappresenta la pressione sistolica media di un individuo sano ~N(120 $5,6^2$)
- Probabilità che selezionando un individuo sano scelto a caso questi abbia una pressione sistolica superiore a 125

$$Pr(X>125)=1-Pr(X<125)=1-Pr(Z<(125-120)/5,6)=$$

$$1-F(0,89)=1-0,81=0,19$$

Soluzione

- X= v.c. che rappresenta la pressione sistolica media di un individuo sano ~N(120 $5{,}6^2$)
- Probabilità che scegliendo a caso 4 individui, la media della loro pressione sistolica sia superiore a 125

$$\overline{X}_n \sim N(\mu\ \sigma^2/n) \qquad \overline{X}_4 \sim N(120\ 5{,}6^2/4)$$

$$Pr(\overline{x}_4 > 125) = 1 - Pr(\overline{x}_4 < 125)$$

$$1 - Pr(\overline{x}_4 < 125) = 1 - Pr(Z(\overline{x}_4) < \tfrac{125-120}{5{,}6/2}) = 0{,}037$$

Soluzione

- X= v.c. che rappresenta la pressione sistolica media di un individuo sano ~N(120 $5{,}6^2$)
- Probabilità che scegliendo a caso 25 individui, la media della loro pressione sistolica sia superiore a 125

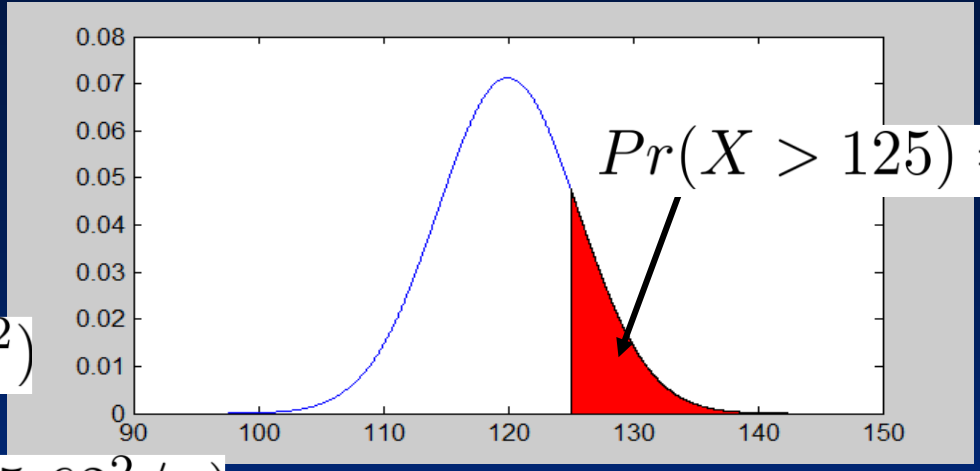
$$\overline{X}_n \sim N(\mu\ \sigma^2/n) \qquad \overline{X}_{25} \sim N(120\ 5{,}6^2/25)$$

$$Pr(\overline{x}_{25} > 125) = 1 - Pr(\overline{x}_{25} < 125)$$

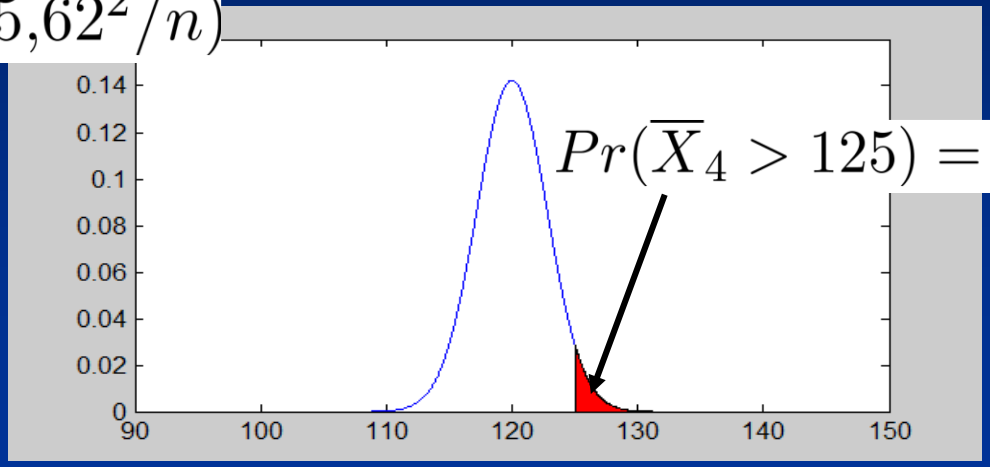
$$1 - Pr(Z(\overline{x}_{25}) < \tfrac{125-120}{5{,}6/5}) = 1 - Pr(Z(\overline{x}_{25}) < 4,464) \approx 0$$

Rappresentazione grafica delle diverse probabilità

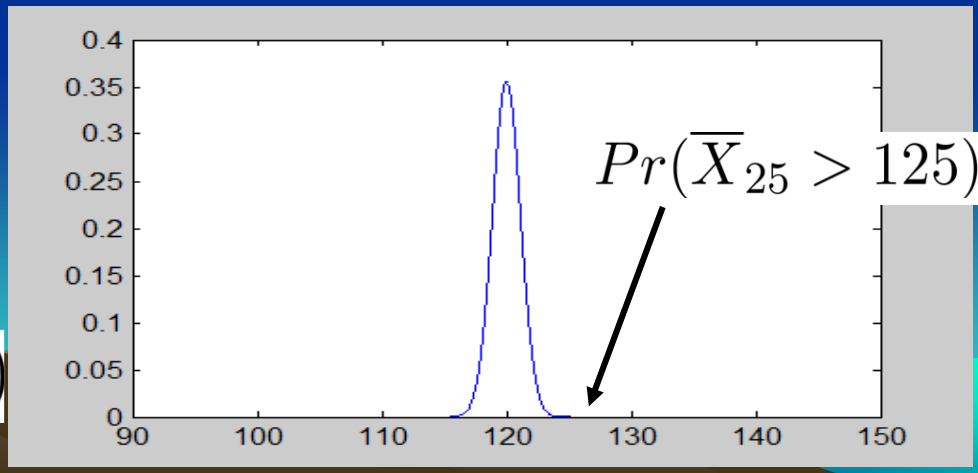
$X \sim N(\mu\ \sigma^2) = N(120\ 5{,}62^2)$



$\overline{X}_n \sim N(\mu\ \sigma^2/n) = N(120\ 5{,}62^2/n)$



$\overline{X}_4 \sim N(120\ \ 5{,}62^2/4)$



$\overline{X}_{25} \sim N(120\ \ 5{,}62^2/25)$

Soluzione

- Probabilità che selezionando 6 individui sani quattro di essi abbiano un pressione inferiore a 125.

Universo Bernoulliano Y con
$\pi = Pr(X<125) = 0,81$

S=v.c. che rappresenta il numero di successi su n prove da un Universo Bernoulliano Y con prob. di successo= π

S =v.a. binomiale:
$S \sim B(n, \pi)$

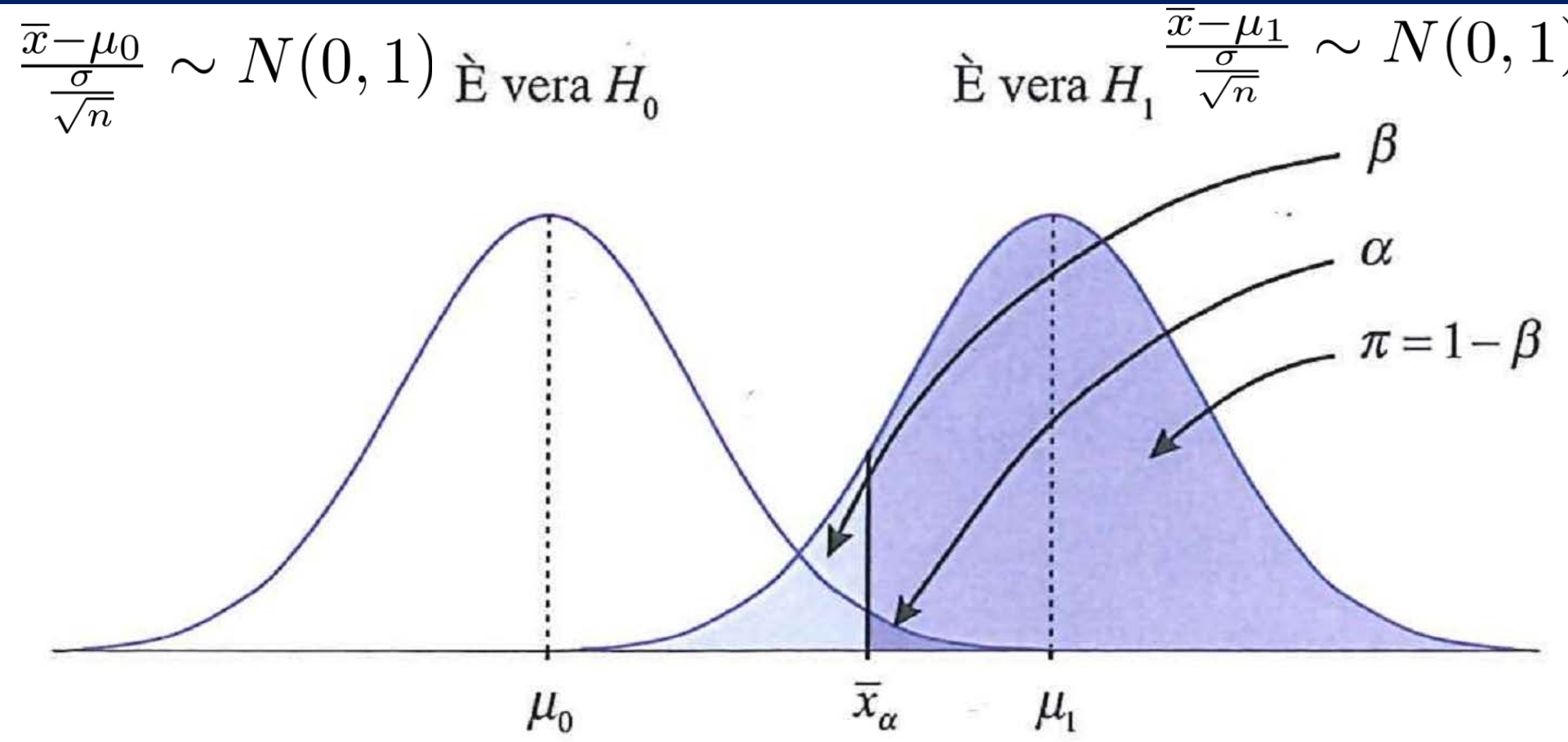
$$P(S = s) = \binom{n}{s} \pi^s (1 - \pi)^{n-s}$$

$$P(S = 4) = \binom{6}{4} 0,81^4 (1 - 0,81)^{6-4} = 0,233$$

Esercizio

- Si consideri la verifica di ipotesi sulla media di una popolazione normale. Si definisce la potenza di un test la probabilità di rifiutare un'ipotesi nulla falsa (ossia la probabilità di non commettere un errore di seconda specie)
- Si considerino le seguenti ipotesi nulla e alternativa
- H_0: $\mu=\mu_0$
- H_1: $\mu=\mu_1$ (con $\mu_1 > \mu_0$)

Errore di prima specie (α) errore seconda specie (β) e potenza del test (1-β)

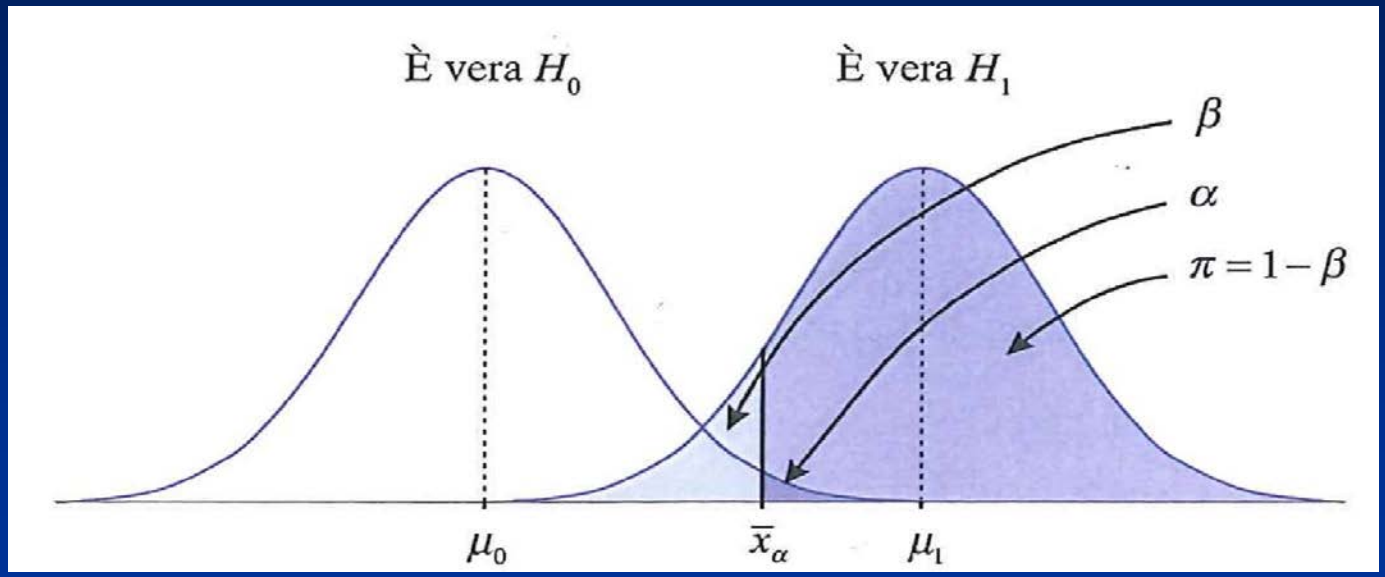


$\bar{x}_\alpha$ = valore soglia che separa la zona di accettazione dalla zona di rifiuto

Quesiti

- Si dimostri che la potenza del test (1-β) è
  - Funzione crescente della dimensione campionaria (*n*)
  - Funzione crescente della differenza tra μ_1 e μ_0
  - Funzione decrescente di σ (standard deviation dell'universo)
  - Funzione crescente di α (probabilità di commettere errore di prima specie)

Obiettivo: trovare l'espressione analitica che definisce la potenza del test (1-β)



1-β = prob. che la v.c. media campionaria, quando l'ipotesi alternativa è vera, assuma valori più grandi di $\bar{x}_\alpha$

$\bar{x}_\alpha$ = valore soglia che separa la zona di accettazione dalla zona di rifiuto

Obiettivo preliminare Trovare l'espressione analitica di $\bar{x}_\alpha$

$$Pr\left(\overline{x} > \overline{x}_\alpha \middle| \mu = \mu_0\right) = \alpha$$

$$Pr\left(\overline{x} < \overline{x}_\alpha \middle| \mu = \mu_0\right) = 1 - \alpha$$

$$Pr\left(\frac{\overline{x}-\mu_0}{\frac{s_{cor}}{\sqrt{n}}} < \frac{\overline{x}_\alpha-\mu_0}{\frac{s_{cor}}{\sqrt{n}}} \middle| \mu = \mu_0\right) = 1 - \alpha$$

$$F_{\mu_0}\left(\frac{\overline{x}_\alpha-\mu_0}{\frac{s_{cor}}{\sqrt{n}}}\right) = 1 - \alpha$$

$$\frac{\overline{x}_\alpha-\mu_0}{\frac{s_{cor}}{\sqrt{n}}} = z_{1-\alpha}$$

$$F_{\mu_0}(z_{1-\alpha}) = 1 - \alpha$$

$$\overline{x}_\alpha = \mu_0 + z_{1-\alpha}\frac{s_{cor}}{\sqrt{n}}$$

Obiettivo finale trovare la prob che la v.c. media campionaria, quando l'ipotesi alternativa è vera, assuma valori più grandi di $\bar{x}_\alpha$

$$\overline{x}_\alpha = \mu_0 + z_{1-\alpha}\frac{s_{cor}}{\sqrt{n}}$$

$$Pr(\overline{x} > \overline{x}_\alpha | \mu = \mu_1) = 1 - Pr(\overline{x} < \overline{x}_\alpha | \mu = \mu_1)$$

$$Pr(\overline{x} < \overline{x}_\alpha | \mu = \mu_1) = Pr\left(\frac{\overline{x}-\mu_1}{s_{cor}/\sqrt{n}} < \frac{\overline{x}_\alpha-\mu_1}{s_{cor}/\sqrt{n}} | \mu = \mu_1\right)$$

$$Pr(\overline{x} < \overline{x}_\alpha | \mu = \mu_1) = F_{\mu_1}\left(\frac{\overline{x}_\alpha-\mu_1}{\frac{s_{cor}}{\sqrt{n}}}\right)$$

$$Pr(\overline{x} < \overline{x}_\alpha | \mu = \mu_1) = F_{\mu_1}\left(-\frac{(\mu_1-\mu_0)\sqrt{n}}{s_{cor}} + z_{1-\alpha}\right)$$

$$Pr(\overline{x} > \overline{x}_\alpha | \mu = \mu_1) = 1 - F_{\mu_1}\left(-\frac{(\mu_1-\mu_0)\sqrt{n}}{s_{cor}} + z_{1-\alpha}\right)$$

Obiettivo finale trovare la prob che la v.c. media campionaria, quando l'ipotesi alternativa è vera, assuma valori più grandi di $\bar{x}_{\alpha}$

$$Pr(\overline{x} > \overline{x}_{\alpha} | \mu = \mu_1) = 1 - F_{\mu_1}\left(-\frac{(\mu_1 - \mu_0)\sqrt{n}}{s_{cor}} + z_{1-\alpha}\right)$$

- Tanto maggiore è la differenza tra μ_1 e $\mu 0$ tanto più F_{μ_1} diminuisce.
- Se $\mu_1 \rightarrow +\infty$ $F_{\mu_1}(-\infty)=0$ e la potenza $\rightarrow +1$
- All'aumentare di n F diminuisce
- Se $n \rightarrow +\infty$ $F_{\mu_1}(-\infty)=0$ e la potenza $\rightarrow +1$
- All'aumentare di s_{cor} la potenza diminuisce (se $s_{cor} \rightarrow +\infty$ la potenza $\rightarrow F_{\mu_1}(z_{1-\alpha})$)

Obiettivo finale trovare la prob che la v.c. media campionaria, quando l'ipotesi alternativa è vera, assuma valori più grandi di $\bar{x}_{\alpha}$

$$Pr(\overline{x} > \overline{x}_{\alpha} | \mu = \mu_1) = 1 - F_{\mu_1}\left(-\frac{(\mu_1 - \mu_0)\sqrt{n}}{s_{cor}} + z_{1-\alpha}\right)$$

- Se $\mu_1 \rightarrow \mu_0$?
- la potenza $\rightarrow 1 - F_{\mu_0}(z_{1-\alpha}) = 1-(1-\alpha) = \alpha$

Esercizio

- Per una generica voce di inventario di una determinata impresa, sia X la differenza tra il valore inventariato ed il valore certificato. Da un campione di 120 voci un certificatore contabile ha ottenuto $\bar{x}=25,3$ $s^2_{cor}=13240$
- Si sottoponga a test l'ipotesi che l'inventario non sia gonfiato specificando opportunamente l'ipotesi alternativa (si ponga $\alpha=0,01$)
- Si calcoli il p-value
- Si calcoli la prob. di rifiutare l'ipotesi nulla nel caso in cui la vera media di X fosse pari a 30

Soluzione

H_0: $\mu = 0$

H_1: $\mu > 0 \Rightarrow$ l'inventario è gonfiato

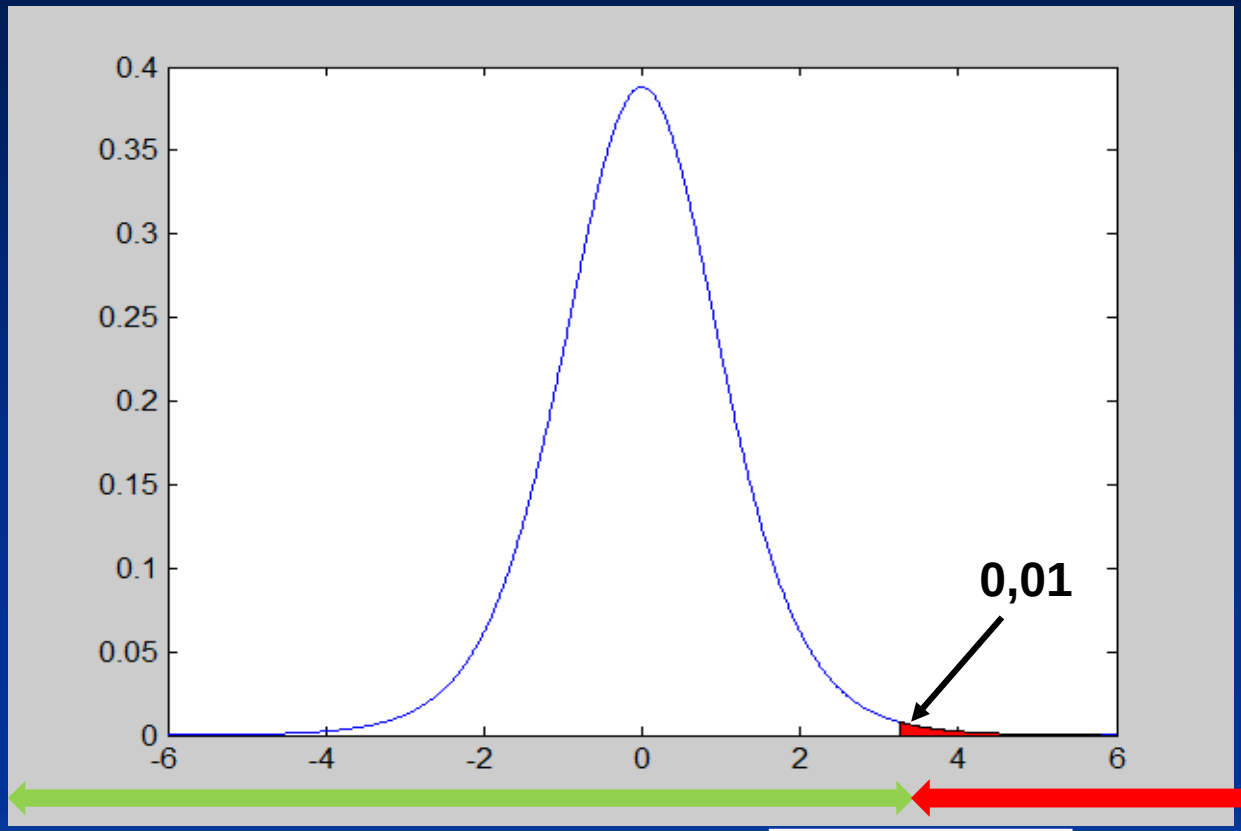
$\bar{x} = 25{,}3$ s_{cor}=115,065 n=120

Teorema centrale del limite

$$Z(\bar{X}) = \frac{\bar{X} - \mu_0}{s_{cor} / \sqrt{n}} \sim \mathrm{N}(0,1)$$

$$Z(\bar{x}) = \frac{25{,}3 - 0}{115{,}065 / \sqrt{120}} = 2{,}4086$$

$H_1: \mu > 0$ α=0,01 F(2,33)=0,99



p-value =
1-F(2,41)=0,008

Zona di accettazione **2,33** Zona di rifiuto

$$Z(\bar{x}) = \frac{25{,}3 - 0}{115{,}065/\sqrt{120}} = 2{,}4086$$

$t_{obs} = Z(\bar{x})$ = 2,41 cade nella zona di rifiuto

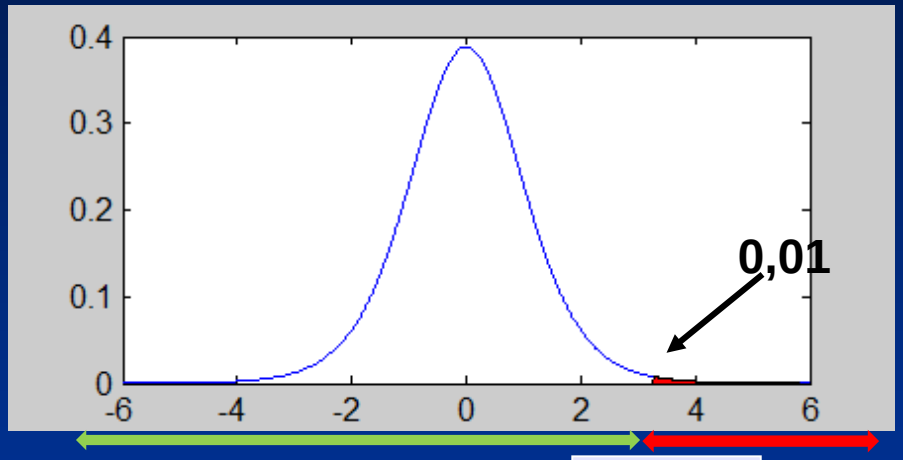
Soluzione (continua)

- Si calcoli la prob. di rifiutare l'ipotesi nulla nel caso in cui la vera media di X fosse pari a 30
- Pr che il valore del test cada nella zona di rifiuto quando μ=30

Qual è il valore soglia che $\bar{x}_\alpha$ separa la zona di accettazione da quella di rifiuto in termini di valori originari?

Qual è il valore soglia che x_α separa la zona di accettazione da quella di rifiuto in termini di valori originari?

s_{cor}=115,065 n=120



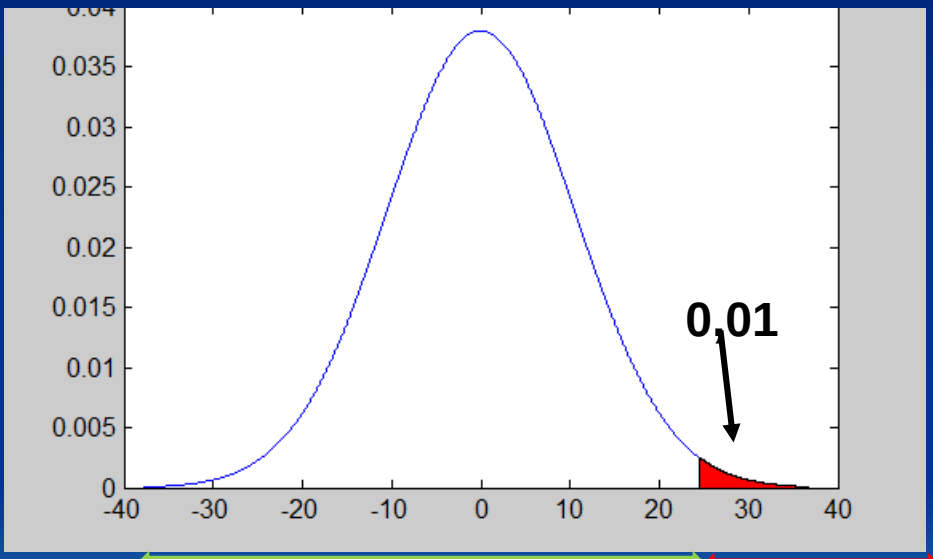
$$\frac{\bar{x}_\alpha - 0}{115{,}065/\sqrt{120}} = 2{,}33$$

Il valore soglia $\bar{x}_\alpha$ è 24,474

Prob. di rifiutare l'ipotesi nulla quando μ=30 ➔ prob. di trovare un valore più grande di 24,474 quando μ=30

Distribuzione media campionaria quando è vera μ=0

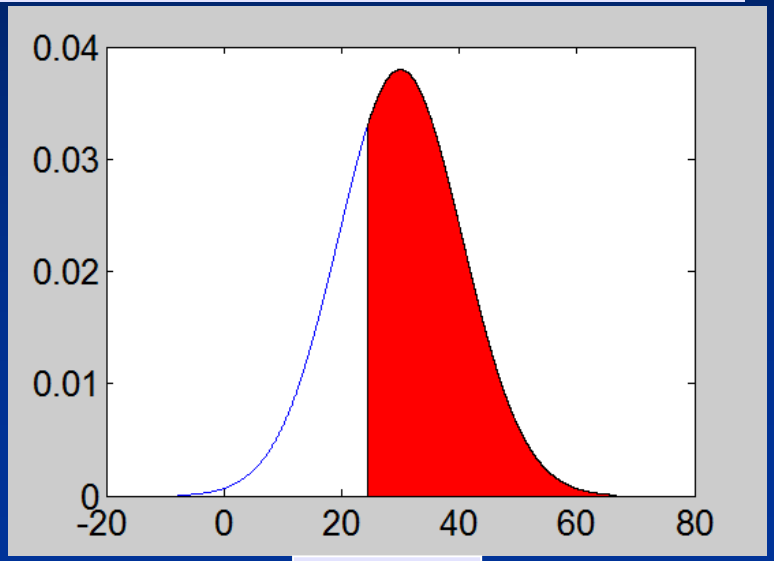
$$Z(\overline{X}) = \frac{\overline{X} - 0}{s_{cor} / \sqrt{n}} \sim N(0,1)$$



24,474

Distribuzione media campionaria quando è vera μ=30

$$Z(\overline{X}) = \frac{\overline{X} - 30}{s_{cor} / \sqrt{n}} \sim N(0,1)$$



24,474

Area rossa = prob. di rifiutare l'ipotesi nulla quando μ=30 (potenza del test = 1-β))

$$1 - F\left(\frac{24{,}474 - 30}{115{,}065 / \sqrt{120}}\right) \approx 0{,}70$$

Soluzione alternativa per il calcolo della potenza del test

- Applichiamo direttamente la formula che avevamo trovato in maniera analitica

$$Pr(\overline{x} > \overline{x}_\alpha | \mu = \mu_1) = 1 - F_{\mu_1}\left(-\frac{(\mu_1 - \mu_0)\sqrt{n}}{s_{cor}} + z_{1-\alpha}\right)$$

μ_1=30 μ_0=0 n=120 $Z_{1-\alpha}$=2,33 s_{cor}=115,065

$$Pr(\overline{x} > \overline{x}_\alpha | \mu = 30) = 1 - F_{\mu_1}\left(-\frac{30\sqrt{120}}{115{,}065} + 2{,}33\right) = 0{,}7$$

TEST SULLA FREQUENZA RELATIVA (grandi campioni)

$H_0: \pi = \pi_0$ (π_0 = valore prefissato)

Consideriamo come statistica-test la frequenza relativa campionaria P che, sotto H_0, gode delle seguenti proprietà:

$E(P) = \pi_0$, $VAR(P) = \pi_0(1 - \pi_0)/n$,

$$Z(P) = \frac{P - \pi_0}{\sigma(P)} \sim N(0,1)$$

↑

Teorema centrale del limite (n grande)

Quindi la frequenza relativa campionaria standardizzata P secondo H_0 è distribuita secondo N(0,1).

Rifiutiamo H_0 quando osserviamo frequenze relative campionarie lontane da $\pi_0 \rightarrow$ frequenze relative campionarie standardizzate lontane da $0 \rightarrow$ sulle code della distribuzione $\rightarrow$ legate a probabilità basse.

I passi sono analoghi al test sulla media:

- **scelta H_1**
- **scelta α**
- **def. zone di rifiuto e accettazione**
- **calcolo $z(p)$ e decisione**
- **in alternativa, calcolo *P-value***

Esempio: ricordo della pubblicità

H_0: $\pi = 0{,}3$ **(soglia di "efficacia")**

H_1: $\pi < 0{,}3$ **(pubblicità inefficace)**

Campione di 600 telespettatori (150 la ricordano):

$p = 0{,}25$ **Fisso $\alpha = 0{,}01$**

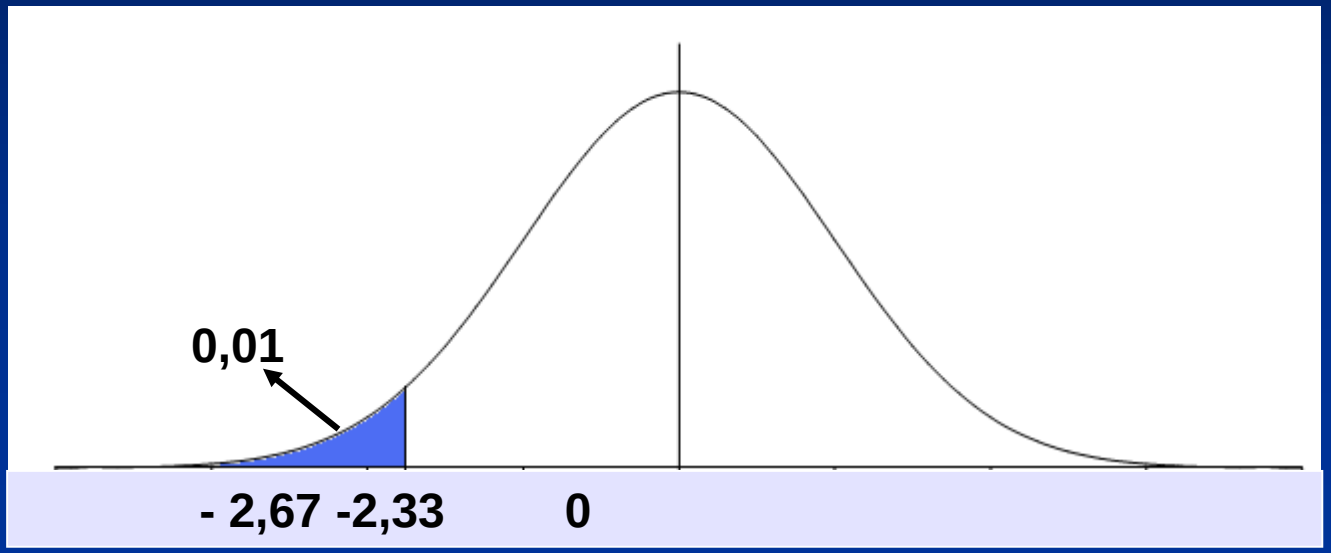
$$\sigma(P) = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{0{,}3 \cdot 0{,}7}{600}} = 0{,}0187$$

$$Z(P) = \frac{P - \pi_0}{\sigma(P)} \sim N(0,1)$$

$$z(p) = \frac{0{,}25 - 0{,}3}{0{,}0187} = -2{,}67$$

Approccio diretto H_1: $\pi < 0{,}3$

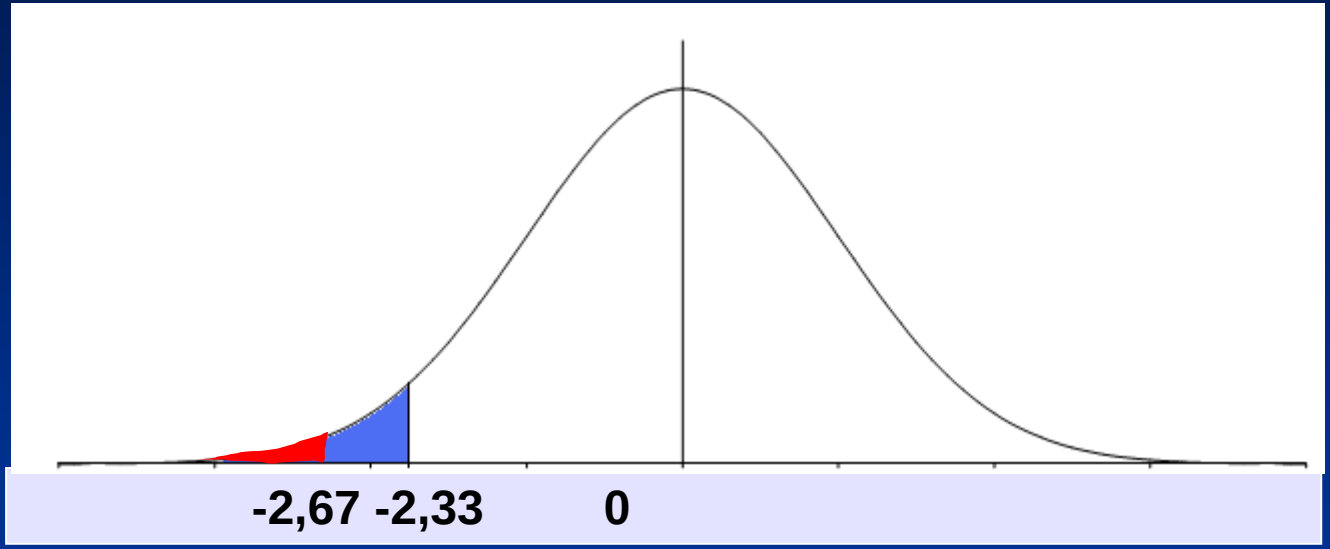
si fissa $\alpha = 0{,}01 \Rightarrow F(-2{,}33) = 0{,}01$



-2,67 è un valore estremo $\Rightarrow$ cade infatti nella zona di rifiuto $\Rightarrow$ rifiutiamo H_0 e concludiamo che la pubblicità è poco efficace

Approccio inverso: p-value

H_1: $\pi < 0{,}3$



p-value = P(Z(X)<-2,67)=*F* [-2,67] = 0,00379 $\Rightarrow$ forte evidenza contro H_0

$\Rightarrow$pubblicità poco efficace

VERIFICA DI IPOTESI SU DUE UNIVERSI (GRANDI CAMPIONI)

ESEMPIO 1

reddito procapite annuo delle famiglie in cui il capofamiglia è laureato e

reddito procapite annuo delle famiglie in cui il capofamiglia è diplomato.

n1=120 famiglie (capofamiglia laureato)

n2=150 famiglie (capofamiglia diplomato).

$\bar{x}_1$**=32.000 €**

$\bar{x}_2$**= 30.000 €**

Esiste una differenza nel reddito procapite annuo a favore delle famiglie con capofamiglia laureato?

ESEMPIO 2

Studio sull'incidenza dell'emicrania sulle persone dedite ad attività sportiva,

n1=150 ragazzi

n2= 200 ragazze.

Si rileva che il 20% dei ragazzi e il 22,5% delle ragazze soffre di emicrania abituale.

Esistono differenze nell'incidenza dell'emicrania nei maschi e nelle femmine?

Ipotesi nulla: i due campioni provengono da universi aventi il medesimo valore del parametro di interesse :

$$H_0: \theta_1=\theta_2$$

(θ_1 e θ_2 parametri nel primo e secondo universo. Es: reddito annuo medio o percentuale di soggetti con emicrania)

N. B.

Si chiede che campioni siano indipendenti cioè che provengono da universi che rappresentano gruppi distinti e ben “separati” di unità statistiche.

Ipotesi sulle medie

H_0: $\mu_1=\mu_2$ μ_1 e μ_2 valori prefissati

Consideriamo come statistica test la "Differenza tra due medie campionarie":

$$DM = \overline{X}_1 - \overline{X}_2$$

Che, sotto H_0, gode delle proprietà:

$$E(DM) = E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2 = 0$$

$$VAR(DM) = VAR(\overline{X}_1 - \overline{X}_2) = VAR(\overline{X}_1) + VAR(\overline{X}_2)$$

$$VAR(DM) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Se si ipotizza

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

$$VAR(DM) = \sigma^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]$$

Si applica il Teorema centrale del limite:

$$Z(DM) = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{VAR(DM)}} \sim N(0,1)$$

per n_1 e n_2 entrambi grandi

Se le varianze dei due universi sono ignote

$$s(DM) = \sqrt{\frac{s^2_{cor1}}{n_1} + \frac{s^2_{cor2}}{n_2}}$$

$$Z(DM) = \frac{\overline{X}_1 - \overline{X}_2}{s(DM)} \sim N(0,1)$$

I passi sono analoghi al test sulla media:

- **scelta H_1**
- **scelta α**
- **def. zone di rifiuto e accettazione**
- **Calcolo $dm = \bar{x}_1 - \bar{x}_2$, $scor_1$, $scor_2$, s(DM),**

quindi $$z(dm) = \frac{\bar{x}_1 - \bar{x}_2}{s(DM)}$$

- **in alternativa, calcolo *P-value***

Es 1: reddito procapite famiglie

- H_0: $\mu_1=\mu_2$ $\alpha = 0{,}05$
- H_1: $\mu_1>\mu_2$ $\Rightarrow$ **(esiste una differenza a favore delle famiglie con capofamiglia laureato)**

capofamiglia laureato

$n_1 = 120$

$\bar{x}_1 = 32.000$

$s_{cor1}=5.500$

capofamiglia diplomato

$n_2 = 150$

$\bar{x}_2 = 30.000$

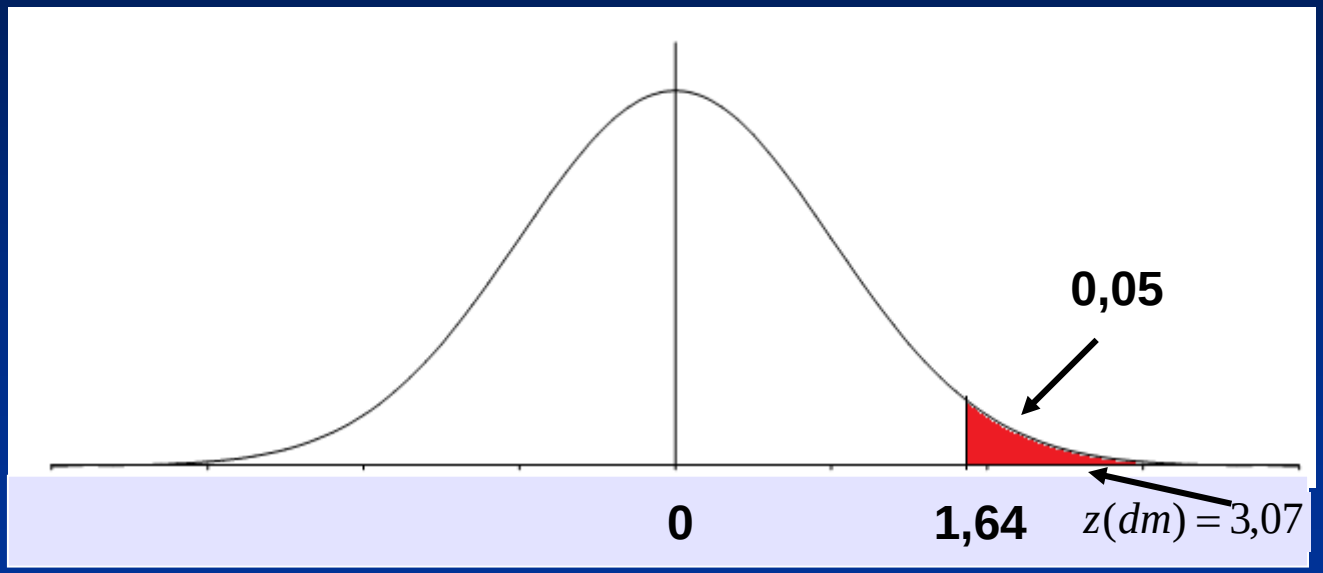
$s_{cor2}=5.100$

$$s(DM) = \sqrt{\frac{s_{cor1}^2}{n_1} + \frac{s_{cor2}^2}{n_2}} = \sqrt{\frac{5.500^2}{120} + \frac{5.100^2}{150}} = 652{,}29$$

$$z(dm) = \frac{32.000 - 30.000}{652{,}29} = 3{,}07$$

Approccio diretto

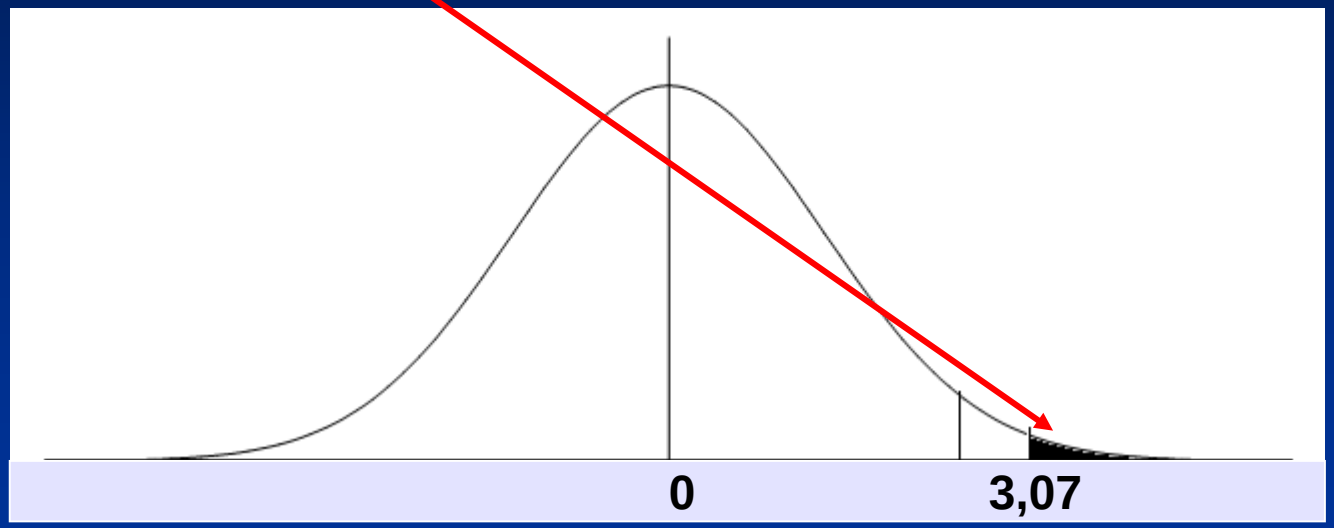
- **se α = 0,05 $\Rightarrow$ F(1,64) = 0,95**



3.07 è un valore estremo $\Rightarrow$ **cade infatti nella regione di rifiuto** $\Rightarrow$ rifiuto H_0 e concludo che **esiste una differenza nel reddito procapite annuo a favore delle famiglie con capofamiglia laureato**

Approccio inverso: P-value

- ***P-value*** **= *P*{*Z*(DM) ≥ + 3,07} = [1 – *F*(+3,07)] = 1 – 0,99893 = 0,00107**



- **⇒ forte evidenza contro H_0.**
- **La differenza tra μ_1 e μ_2 *è significativa:* il reddito procapite annuo delle famiglie con capofamiglia laureato è significativamente maggiore di quello delle famiglie con capofamiglia diplomato.**

IPOTESI SULLE FREQUENZE RELATIVE

H_0: $\pi_1=\pi_2$, π_1 e π_2 valori prefissati (Es. frequenza di soggetti con emicrania).

Consideriamo come statistica test la "Differenza tra due frequenze relative campionarie" (DP):

$$DP = P_1 - P_2$$

Che, sotto H_0, gode delle proprieta':

1) $$E(DP) = E(P_1) - E(P_2) = \pi_1 - \pi_2 = 0$$

2) $$VAR(DP) = VAR(P_1) + VAR(P_2) = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$$

$$VAR(DP) = \pi(1-\pi)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

3) Si applica il Teorema centrale del limite

$$Z(DP) = \frac{P_1 - P_2}{\sqrt{VAR(DP)}} \sim N(0,1)$$

- per n_1 e n_2 entrambi grandi

Poiché π è ignoto, viene stimato dai dati campionari:

$$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} \quad \text{Stima di } \pi$$

$$\mathrm{s(DP)} = \sqrt{p(1-p)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

$$\mathrm{Z(DP)} = \frac{P_1 - P_2}{s(DP)} \sim N(0,1)$$

I passi sono analoghi agli altri test

- **scelta H_1**
- **scelta α**
- **def. zone di rifiuto e accettazione**
- **Calcolo dp = p_1- p_2, *s(DP), quindi***

$$z(dp) = \frac{p_1 - p_2}{s(DP)}$$

In alternativa: calcolo *P-value*

Es. incidenza emicrania

$$Z(DP) = \frac{P_1 - P_2}{s(DP)} \sim N(0,1)$$

$$s(DP) = \sqrt{p(1-p)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

- H_0: $\pi_1 = \pi_2$
- H_1: $\pi_1 \neq \pi_2$ $\Rightarrow$ **(esiste una differenza nell'incidenza dell'emicrania nei maschi e nelle femmine)**

	maschi	femmine
•	$n_1 = 150$	$n_2 = 200$
•	$p_1 = 0,20$	$p_2 = 0,225$

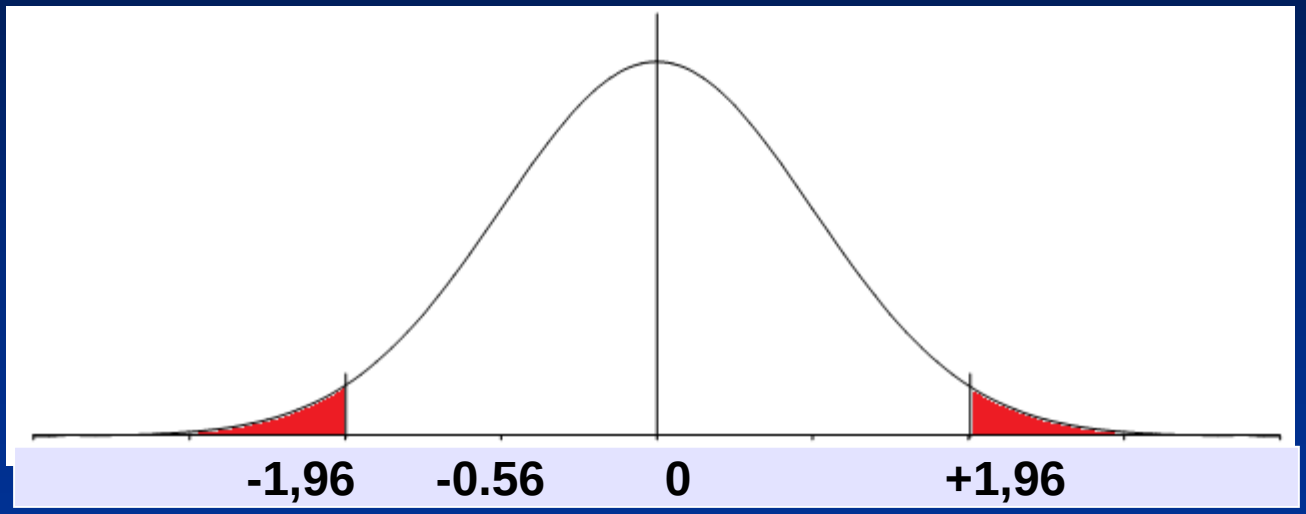
$$p = \frac{0,20 \cdot 150 + 0,225 \cdot 200}{150 + 200} = 0,2143$$

$$s(DP) = \sqrt{0,2143 \cdot (1 - 0,2143)\left[\frac{1}{150} + \frac{1}{200}\right]} = 0,0443$$

$$z(dp) = \frac{0,20 - 0,225}{0,0443} = -0,56$$

Approccio diretto

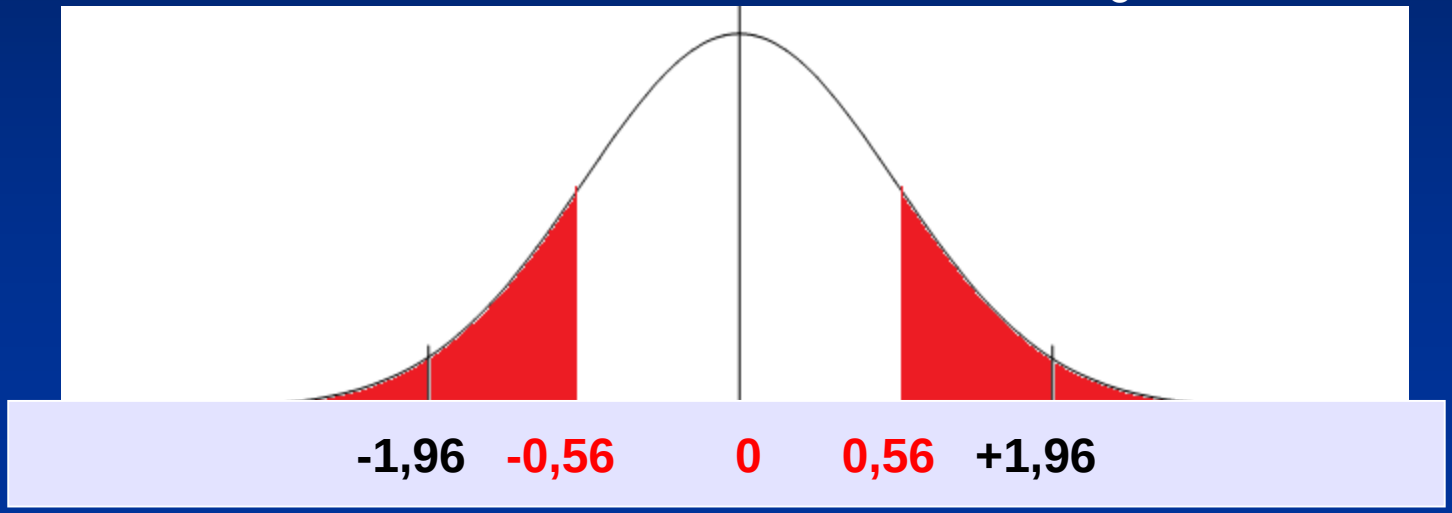
- **si fissa α = 0,05 $\Rightarrow$ z(0,05) = 1,96**



- -0.56 NON è un valore estremo $\Rightarrow$ **cade infatti nella regione di accettazione$\Rightarrow$ NON c'è evidenza per rifiutare** H_0 e dire che **esiste una differenza nell'incidenza dell'emicrania tra maschi e femmine**

Approccio inverso: P-value

- ***P-value*** **= P{Z(DP)≤- 0,56} + *P*{*Z*(DP) ≥ + 0,56} = 2*F*(-0,56) = 2· 0,28774 = 0,57548**
- **⇒ nessuna evidenza contro H_0.**



- **La differenza tra π_1 e π_2 non *è significativa:* l'incidenza dell'emicrania nei ragazzi dediti ad attività sportiva è uguale a quella delle ragazze.**

Esercizio

In una prova sul carico di rottura di due tipi di corda si dispone di 2 campioni di ampiezza 26 e 35 rispettivamente. Nel primo campione la media è 185,3Kg, nel secondo campione la media è 175,2Kg. Per esperienza passata si sa che le deviazioni standard delle popolazioni generatrici sono rispettivamente pari a 14,8 e 10,6. Si costruisca un intervallo di confidenza per la differenza tra le medie ad un livello di confidenza di 0,95.

Soluzione

Ip. Distribuzione normale dei carichi di rottura delle due corde

Primo tipo di corda

$$\bar{x}_1 \sim N(\mu_1,\ \sigma_1^2/\mathrm{n}_1)$$

$$\bar{x}_1 = 185{,}3 \quad n_1 = 26 \quad \sigma_1 = 14{,}8$$

Secondo tipo di corda

$$\bar{x}_2 \sim N(\mu_2,\ \sigma_2^2/\mathrm{n}_2)$$

$$\bar{x}_2 = 175{,}2 \quad n_1 = 35 \quad \sigma_2 = 10{,}6$$

OBIETTIVO: intervallo di confidenza per la differenza tra le medie ad un livello di confidenza di 0,95.

Occorre trovare la distribuzione di DM

$$DM = \bar{x}_1 - \bar{x}_2$$

Soluzione

Ip. Distribuzione normale dei carichi di rottura delle due corde

Primo tipo di corda

$$\bar{x}_1 \sim N(\mu_1,\ \sigma_1^2/\mathrm{n}_1)$$

$$\bar{x}_1 = 185{,}3\ \ n_1 = 26\ \ \sigma_1 = 14{,}8$$

Secondo tipo di corda

$$\bar{x}_2 \sim N(\mu_2,\ \sigma_2^2/\mathrm{n}_2)$$

$$\bar{x}_2 = 175{,}2\ \ n_2 = 35\ \ \sigma_2 = 10{,}6$$

$$DM = \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2,\ VAR(\bar{x}_1 - \bar{x}_2)\right)$$

$$Z(DM) = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{VAR(\bar{x}_1 - \bar{x}_2)}} \sim N(0,1)$$

$$\sqrt{VAR(DM)} = \sqrt{VAR(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{14{,}8^2}{26} + \frac{10{,}6^2}{35}} = 3{,}41$$

Stima per intervallo di confidenza della differenza tra le due medie

$$\Pr\left(-1{,}96 < \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{VAR(\overline{X}_1 - \overline{X}_2)}} < 1{,}96\right) = 0{,}95$$

$$Pr((\overline{x}_1 - \overline{x}_2) - 1,96\sqrt{VAR(\overline{x}_1 - \overline{x}_2)} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + 1,96\sqrt{VAR(\overline{x}_1 - \overline{x}_2)} = 0,95$$

$$Pr\left((185,3 - 175,2) - 1,96 \times 3,41 < \mu_1 - \mu_2 < (185,3 - 175,2) + 1,96 \times 3,41\right) = 0,95$$

$$Pr\left(3,42 < \mu_1 - \mu_2 < 16,78\right) = 0,95$$

Osservazione: il fatto che entrambi gli estremi dell'intervallo di confidenza siano positivi suggerisce che la media della prima popolazione (μ_1) è verosimilmente superiore a quella della seconda (μ_2)

Esercizio

- Utilizzando i dati dell'esercizio precedente, si verifichi l'ipotesi che le medie delle due popolazioni generatrici siano uguali al livello di significatività del 0,01.
- Si calcoli il p-value del risultato del test e si commentino i risultati ottenuti

$\bar{x}_1 = 185{,}3 \;\; n_1 = 26 \;\; \sigma_1 = 14{,}8$

$\bar{x}_2 = 175{,}2 \;\; n_2 = 35 \;\; \sigma_2 = 10{,}6$

Soluzione

- H_0: $\mu_1 = \mu_2$
- H_1: $\mu_1 \neq \mu_2$ $\Rightarrow$ **(esiste una differenza tra il carico di rottura dei due tipi di corda)**

Se è vera l'ipotesi nulla H_0: $\mu_1 = \mu_2$

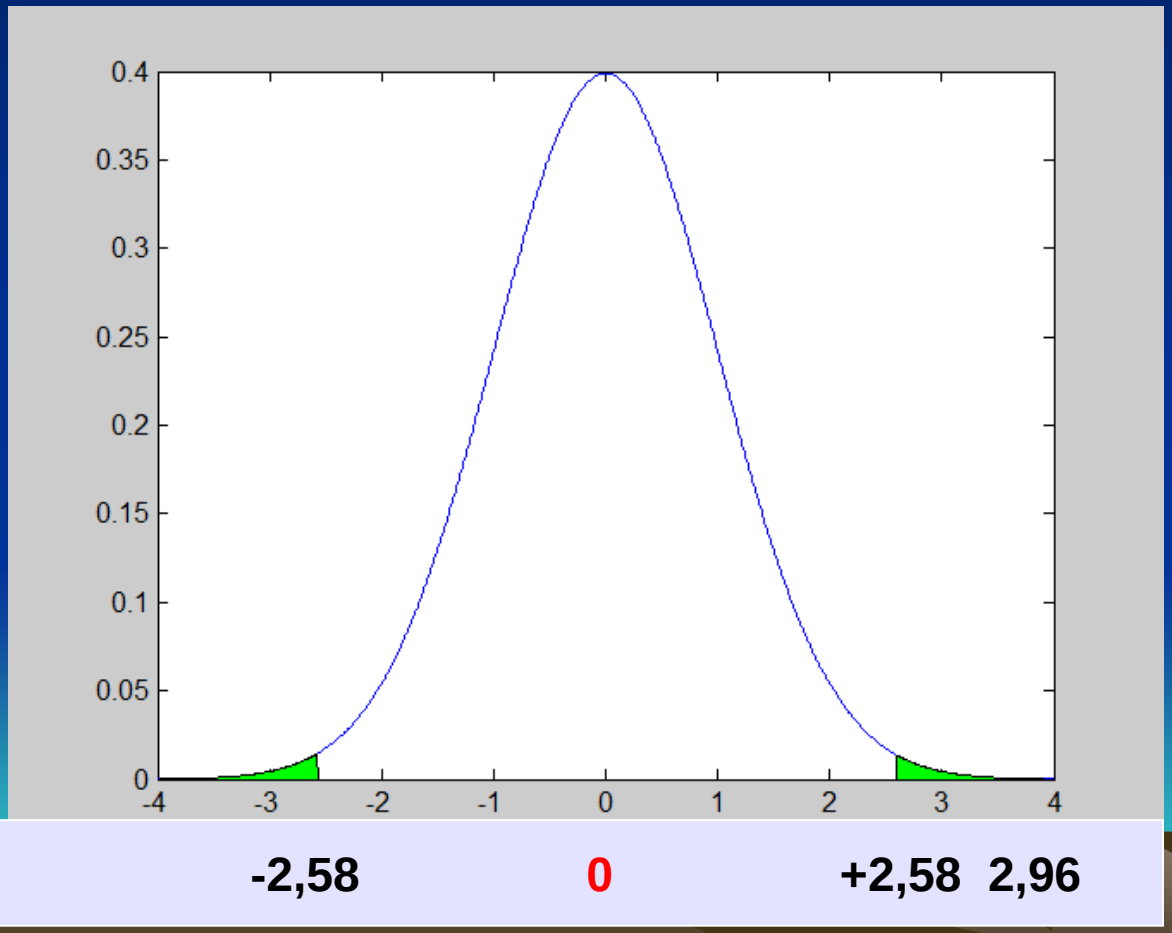
$$Z(DM) = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{VAR(\bar{x}_1 - \bar{x}_2)}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{VAR(\bar{x}_1 - \bar{x}_2)}} \sim N(0,1)$$

$$\sqrt{VAR(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{14{,}8^2}{26} + \frac{10{,}6^2}{35}} = 3{,}41$$

$$Z(DM) = \frac{10{,}1}{3{,}41} = 2,96$$

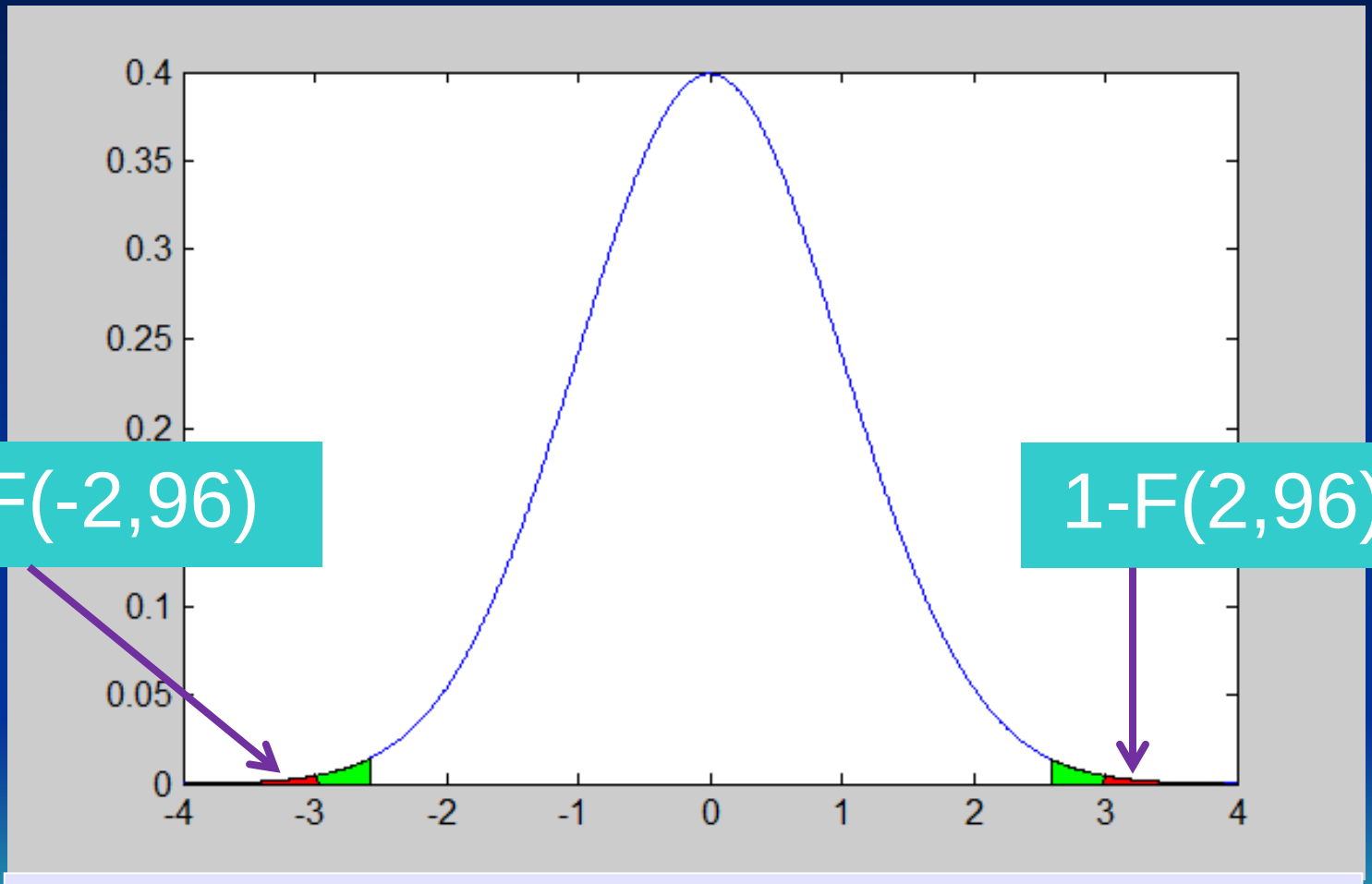
Soluzione

- H_0: $\mu_1=\mu_2$
- H_1: $\mu_1\neq\mu_2$ $\Rightarrow$ **(esiste una differenza tra il carico di rottura dei due tipi di corda)** $\alpha=0{,}01$.



2,96 cade nella zona di rifiuto. C'è una prob. inferiore all'1% di osservare il risultare campionario 2,96 quando l'ipotesi nulla è vera.

Calcolo del p-value



F(-2,96)+1-F(2,96)= 2(1-F(2,96))=0,003

Esercizio

- Un ricercatore desidera stimare la media di una popolazione che presenta una deviazione standard σ con un campione di numerosità *h* in modo tale che sia uguale a 0,90 la probabilità che la media del campione non differisca dalla media della popolazione per più dell'8% della deviazione standard. Si determini *h*.

Soluzione

Se l'intervallo di confidenza è al 90%

$$P\left\{\overline{X} - 1{,}645\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + 1{,}645\frac{\sigma}{\sqrt{n}}\right\} = 0{,}90$$

$$P\left\{|\overline{X} - \mu| \le 1{,}645\frac{\sigma}{\sqrt{n}}\right\} = 0{,}90$$

Se vogliamo che l'errore di stima della media non superi 0,08 σ

$$1{,}645\frac{\sigma}{\sqrt{n}} \le 0{,}08\sigma$$

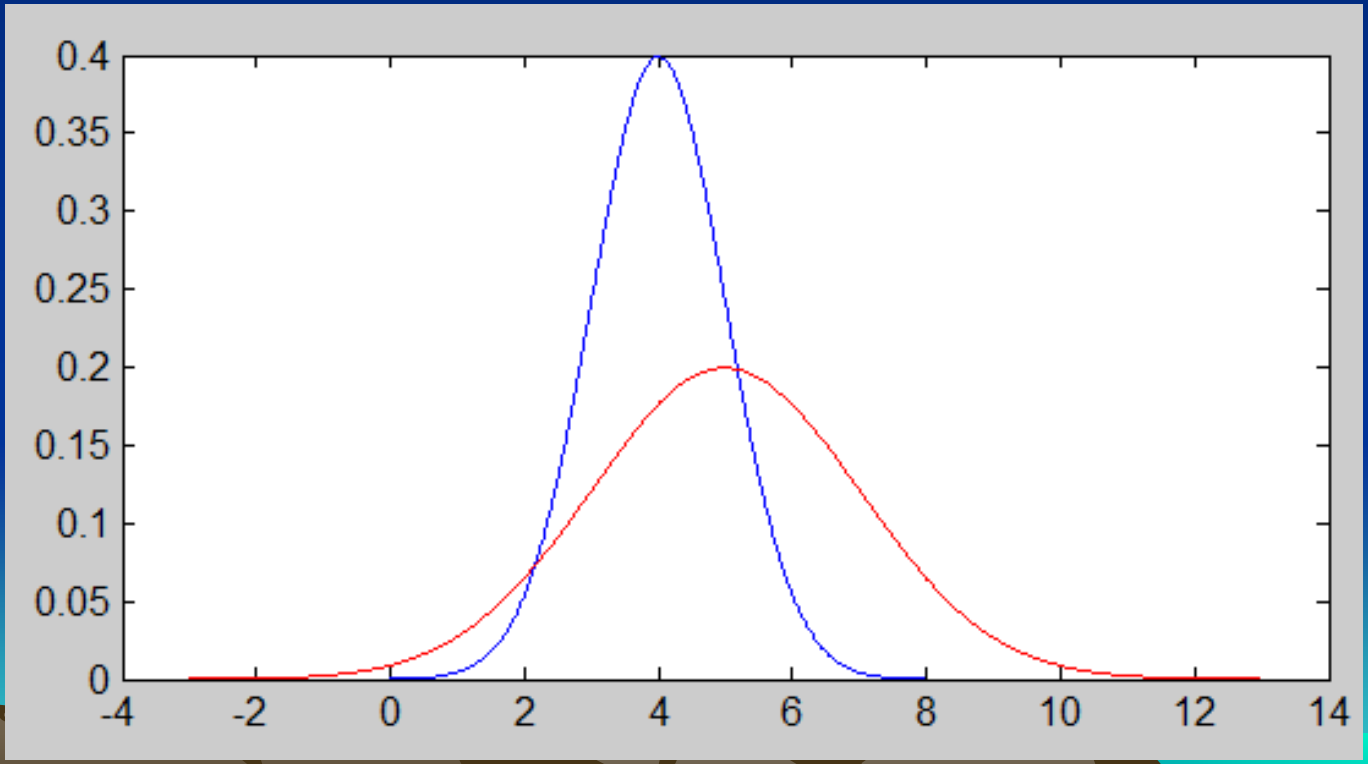
$$n \ge \left(\frac{1{,}645}{0{,}08}\right)^2 \approx 423$$

Esercizio

- Siano X_1 e X_2 due v.c. indipendenti con distribuzione N(4,1) e N(5,4) rispettivamente.
- Si rappresenti graficamente la densità delle due distribuzioni
- Si calcoli $P(X_1<X_2)$

Soluzione

- Linea Blu ➔ X_1 ~N(4,1)
- Linea rossa ➔ X_2 ~ N(5,4)



Distribuzione di $Z=X_1-X_2$

$X_1 \sim N(4,1)$ $X_2 \sim N(5,4)$

- Qualsiasi combinazione lineare di v.c. normali presenta distribuzione normale
- $E(Z) = E(X_1-X_2)=E(X_1)-E(X_2)=4-5=-1$
- $VAR(Z) = VAR(X_1-X_2)=VAR(X_1)+VAR(X_2)$ $=1+4=5$
- $Z \sim N(-1, 5)$
- $P(Z<0)= P(X_1<X_2)= P(X_1-X_2<0) = F(1/5^{0.5})=0.6726$

Esercizio

- Sia X_1 X_2 X_3 un campione casuale estratto dalla distribuzione normale N(2,9). Si calcoli
- $P(X_1+4X_2-4X_3>8)$
- $P(2X_1+4X_2-4X_3>8)$

Distribuzione di $Z= X_1+4X_2-4X_3$

$X_1\ X_2\ X_3 \sim N(2,9)$

- Qualsiasi combinazione lineare di v.c. normali presenta distribuzione normale
- $E(Z) = E(X_1+4X_2-4X_3)=E(X_1)+4E(X_2)-4E(X_3) =2+8-8=2$
- $VAR(Z) = VAR(X_1+4X_2-4X_3) =VAR(X_1)+16VAR(X_2)+16VAR(X_3) =33*9=297$
- $Z\sim N(2, 297)$ $P(Z>8)$?
- $P(Z>8)= 1-F(6/297^{0.5})=0.3639$

Distribuzione di $Z= 2X_1+4X_2-4X_3$

$X_1\ X_2\ X_3 \sim N(2,9)$

- Qualsiasi combinazione lineare di v.c. normali presenta distribuzione normale
- $E(Z) = E(2X_1+4X_2-4X_3)=E(X_1)+4E(X_2)-4E(X_3)$ $=4+8-8=4$
- $VAR(Z) = VAR(2X_1+4X_2-4X_3)$ $=4VAR(X_1)+16VAR(X_2)+16VAR(X_3)$ $=36*9=324$
- $Z \sim N(4, 324)$
- $P(Z>8)= 1-F(4/324^{0.5})=0.4121$

Esercizi da svolgere per LUN 24 marzo

Esercizio

Un partito politico ha commissionato un'indagine sull'orientamento della popolazione al prossimo referendum. Al partito interessa sapere se la percentuale dei votanti è la stessa nelle regioni chiave A e B. Nella regione A su 500 intervistati 300 hanno dichiarato che voteranno sì, nella regione B, su 600 intervistati, 340 hanno dichiarato che voteranno sì.

Si definisca l'ipotesi nulla e l'ipotesi alternativa e si dica quali conclusioni si ottengono assumendo $\alpha=0,05$. Si calcoli il p-value del test.

Esercizio

- Una moneta viene lanciata 80 volte, ottenendo 45 volte l'esito «testa».
- Al livello di significatività del 5% vi è sufficiente evidenza per ritenere che la moneta sia truccata?

Esercizio

- Nel processo di controllo del peso delle confezioni di un determinato prodotto l’azienda esamina un campione di 800 confezioni e trova che 15 di esse hanno un peso fuori norma.
- Si determini l’intervallo di confidenza al 97% della proporzione di pezzi fuori norma.
- Si testi, al livello di significatività dell'1%, l'ipotesi che la proporzione di pezzi fuori norma sia pari a 1,25%.
- Se la proporzione di pezzi fuori norma nell'universo fosse uguale a 1,5%, effettuando cinque estrazioni
  - si calcoli la probabilità di trovare esattamente due pezzi fuori norma;
  - si scriva l'espressione che consente di calcolare la probabilità di ottenere un numero di pezzi fuori norma compreso tra due e quattro (estremi compresi).

Esercizio

Un tipo di componente viene fornito in confezioni da 400 pezzi. Ne testiamo un campione di 16 per stimare la frazione di difettosi: vogliamo fare un test al livello di significatività α del 5% che ci permetta di rifiutare l'intera partita se vi è evidenza statistica che i pezzi difettosi (nella confezione) sono più del 15%

Quesiti

- Qual `e il parametro incognito su cui basare il test? Come vanno scelte ipotesi nulla e alternativa? Se nel campione si trovano 3 difettosi, cosa si decide? Quanti difettosi si possono accettare al massimo nel campione senza rifiutare la fornitura?
- Se una confezione ha il 25% di difettosi, con che probabilità questo test la rifiuta?

Esercizio

- Si consideri un dado a 20 facce tutte uguali
- Qual è il valore atteso?
- Quante volte è necessario lanciarlo affinché la probabilità di ottenere almeno un 20 sia maggiore o uguale a 0.5?
- Lanciandolo 20 volte, qual è il numero medio di 20 ottenuti?
- Pr di ottenere almeno una volta la faccia 20 in 20 lanci?

Esercizio

- Nel gioco del lotto un numero ha una probabilità p di uscire ad ogni estrazione.
- Si scriva la densità della v.c. (X) che descrive il tempo di attesa dell'uscita del numero all'estrazione k-esima (v. casuale geometrica), k=1, 2, 3, ….
- Si dimostri che la somma delle probabilità è 1
- Si calcoli il valore atteso e la varianza di X
- Si calcoli l'espressione che definisce P(X>k)

Esercizio

- Dimostrare che nel gioco del lotto la probabilità che siano necessari i+j tentativi prima di ottenere il primo successo, dato che ci sono già stati i insuccessi consecutivi, è uguale alla probabilità non condizionata che almeno j tentativi siano necessari prima del primo successo.
- *Morale: il fatto di avere già osservato i insuccessi consecutivi non cambia la distribuzione del numero di tentativi necessari per ottenere il primo successo*

Esercizio

- Sia X una v.c. definita nell'intervallo [0 +∞)

$$f_X(x) = cxe^{-\frac{x^2}{2}}$$

- Calcolare il valore di c affinché $f_X(x)$ sia effettivamente una densità
- Rappresentarla graficamente la funzione di densità
- Calcolare la funzione di ripartizione e rappresentarla graficamente
- Calcolare P(X>x)

Esercizio

Un'azienda produce occhiali utilizzando tre diversi macchinari. Il primo macchinario produce mediamente un paio di occhiali difettosi ogni 100, il secondo ogni 200, il terzo ogni 300. Gli occhiali vengono imballati in scatole identiche, contenenti 100 paia.

Ogni scatola contiene occhiali scelti a caso tra quelli prodotti da una sola delle tre macchine.

Si supponga che il primo macchinario abbia una produzione doppia rispetto agli altri due, cioè una scatola scelta a caso ha probabilità 1/2 di essere prodotta dal primo macchinario, 1/4 dal secondo e 1/4 dal terzo.

Quesiti

- Un ottico riceve una scatola con 100 paia di occhiali.

1. Qual è la probabilità che trovi almeno un paio di occhiali difettoso?

2. Se l'ottico trova esattamente due paia difettose, qual è la probabilità che gli occhiali siano stati prodotti dal primo macchinario?

Esercizio

- Si lancia ripetutamente una coppia di dadi non truccati e si sommano i risultati.
- 1. Si calcoli la prob di ottenere un 7 come somma.
- 2. Si calcoli la prob che occorrano meno di 6 lanci per ottenere almeno un 7.
- 3. Si calcoli la prob che occorrano più di 6 lanci per ottenere almeno un 7

Esercizio

- Per 10 paesi dell'Unione Europea si è osservato il prezzo in euro di un litro di benzina (X) e il numero di veicoli pro capite circolanti (Y). Si conoscono i seguenti risultati relativi alle due variabili.

$$\sum_{i=1}^{10} x_i = 8.79 \qquad \sum_{i=1}^{10} y_i = 8.63$$
$$\sum_{i=1}^{10} x_i^2/10 = 0.77385 \qquad \sum_{i=1}^{10} y_i^2/10 = 0.7695$$

- Inoltre è noto che la devianza residua è pari a 0.01157.

Richieste

- Calcolare i parametri *a* e *b* della retta di regressione assumendo Y come variabile dipendente (in mancanza di altre informazioni fare opportune ipotesi, giustificandole, sul segno del coefficiente angolare).
- Commentare la bontà di adattamento del modello.
- Calcolare l'intervallo di confidenza di β al livello di confidenza del 95% e commentare i risultati ottenuti.