STATISTICA A – K (63 ore)

Marco Riani

mriani@unipr.it

http://www.riani.it

Esercizio

- In un'urna vi sono N/2 palline bianche e N/2 palline nere. Si supponga di estrarre un campione con ripetizione di dimensione n. Si consideri la v.c. X numero di palline bianche in n estrazioni. Come deve essere scelta la numerosità n per fare in modo che var(X) risulti pari a 5?

Soluzione

- X =numero di palline bianche estratte
- π =1/2
- $X \sim B(n, \pi)$
- $VAR(X) = n \pi (1- \pi) = n \times 0.5 \times 0.5$
- Obiettivo $VAR(X)= 5$
- Vincolo $n \times 0.5 \times 0.5 =5$
- $n = 5/(1/4)= 20$

Esercizio

- Un candidato si presenta ad un concorso in cui viene sottoposto ad una prova a quiz costituita da 8 affermazioni alle quali bisogna rispondere SI o NO. Si suppone che per superare la prova si debba rispondere correttamente a più di 6 domande. Quale è la probabilità che rispondendo a caso il candidato superi la prova?

Soluzione

- **n =8 π=0,5   S~B(8; 0,5)**
- **Pr (S≥7) =**

$$P(S>6)=\binom{8}{7}\frac{1}{2}^{7}\frac{1}{2}^{1}+\binom{8}{8}\frac{1}{2}^{8}=0.0352$$

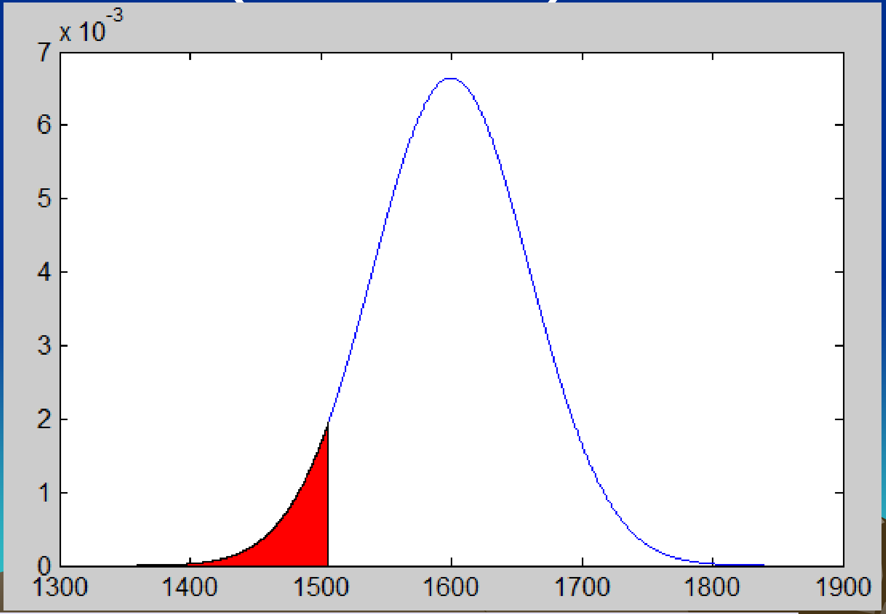
Esercizio

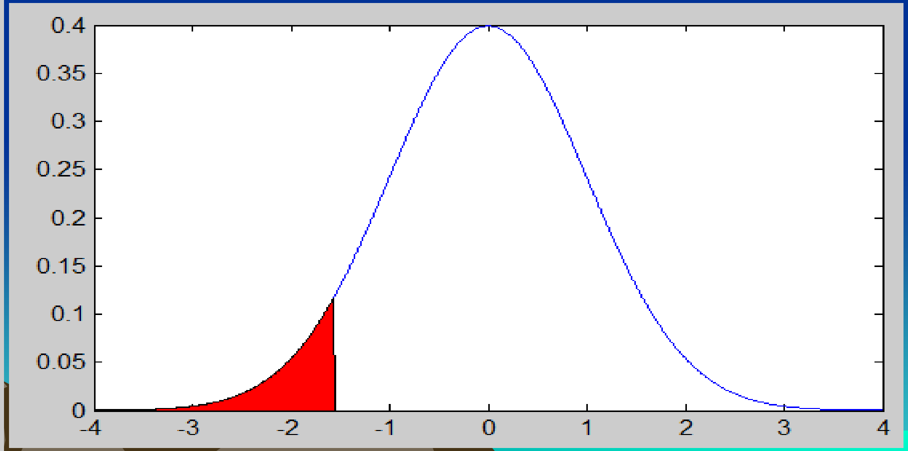
- Una ditta produttrice di una macchina sa che la durata della stessa si distribuisce come una normale con $\mu = 1600$ e $\sigma^2 = 3600$.
- L'azienda risarcisce 500 euro all'acquirente se la durata della macchina è <1506 gg
- Calcolare le seguenti probabilità
- in seguito alla vendita della macchina la ditta debba risarcire 500 €
- su 5 macchine vendute la ditta debba risarcire al massimo 500 euro
- su 5 macchine vendute la ditta debba risarcire almeno 500 euro.
- su 100 macchine vendute la ditta debba risarcire più di 5000 euro (è sufficiente scrivere la formula di calcolo).
- su 120 macchine vendute la ditta debba risarcire meno di 5000 euro (è sufficiente scrivere la formula di calcolo).

Soluzione

- Durata macchina $X \sim N(1600 \text{ gg}, 60^2)$.
- L'azienda risarcisce 500 euro all'acquirente se la durata della macchina è <1506gg
- Calcolare le seguenti probabilità
- Pr la ditta debba risarcire 500 €=Pr X<1506 gg
- $Pr(X<1506)=$ $Pr(Z<(1506-1600)/60)=$

$Pr(Z<-1,57) \approx 0,0582$





x=1506 μ=1600

z=(1506-1600)/60

Su 5 macchine vendute la ditta debba risarcire al massimo 500 euro

- La ditta deve risarcire 500 € se X < 1506
- Pr risarcimento = 0,0582
- Numero di prove n=5
- Pr di risarcire al massimo 500 euro = Pr che tutte le 5 macchine durino più di 1506 gg + Pr che una sola duri meno di 1506 gg
- Y= numero di macchine con durata inferiore a 1506 gg

Y~B(5   0,0582)

Pr di risarcire al massimo 500 Euro = Pr (Y=0) +Pr(Y=1)

$$Pr(Y=0)+Pr(Y=1)=\binom{5}{0}0,0582^0(1-0,0582)^5+\binom{5}{1}0,0582^1(1-0,0582)^4$$

$$=(1-0,0582)^5+5\times 0,0582^1(1-0,0582)^4=0,9699$$

Su 5 macchine vendute la ditta debba risarcire debba risarcire almeno 500 €

- La ditta deve risarcire 500 € se $X < 1506$
- Pr risarcimento = 0,0582
- Numero di prove n=5
- Y= numero di macchine con durata inferiore a 1506 gg
    - Y~B(5, 0,0582)

Pr di risarcire almeno 500 euro =

Pr (Y=1)+ Pr (Y=2)+…+ Pr (Y=5)

$$\sum_{i=1}^{5} Pr(Y=i) = \sum_{i=1}^{5} \binom{5}{i} 0,0582^{i}(1-0,0582)^{5-i}$$

$$\sum_{i=1}^{5} Pr(Y=i) = 1 - Pr(Y=0) = 1-(1-0,0582)^{5} = 0,2590$$

Su 100 macchine vendute la ditta debba risarcire più di 5000 euro

- La ditta deve risarcire 500 € se X < 1506
- Pr risarcimento = 0,0582
- Numero di prove n=100
- Y= numero di macchine con durata inferiore a 1506 gg

Y~B(100 0,0582)

L'az. risarcisce più di 5000 euro quando i guasti sono >10

$$\sum_{i=11}^{100} Pr(Y=i) = \sum_{i=11}^{100} \binom{100}{i} 0,0582^{i}(1-0,0582)^{100-i}$$

Utilizzando Excel o la calcolatrice si ottiene

$$\sum_{i=11}^{100} Pr(Y=i) = 0,03116187612$$

Su 120 macchine vendute la ditta debba risarcire meno di 5000 euro

- La ditta deve risarcire 500 € se X < 1506
- Pr risarcimento = 0,0582
- Numero di prove n=120
- Y= numero di macchine con durata inferiore a 1506 gg

Y~B(120   0,0582)

L'az. risarcisce meno di 5000 euro quando i guasti sono <10

$$\sum_{i=0}^{9} Pr(Y=i) = \sum_{i=0}^{9} \binom{120}{i} 0,0582^{i}(1-0,0582)^{120-i}$$

Utilizzando Excel o la calcolatrice si ottiene

$$\sum_{i=0}^{9} Pr(Y=i) = 0,83830072767$$

Esercizio

Il ministero degli interni ha comunicato che il 95% dei tifosi aspetta almeno 3 minuti per entrare dentro lo stadio, mentre il 2% aspetta più di 14 minuti.
Supponendo che il tempo di attesa Y sia distribuito normalmente $Y \sim N(\mu, \sigma^2)$

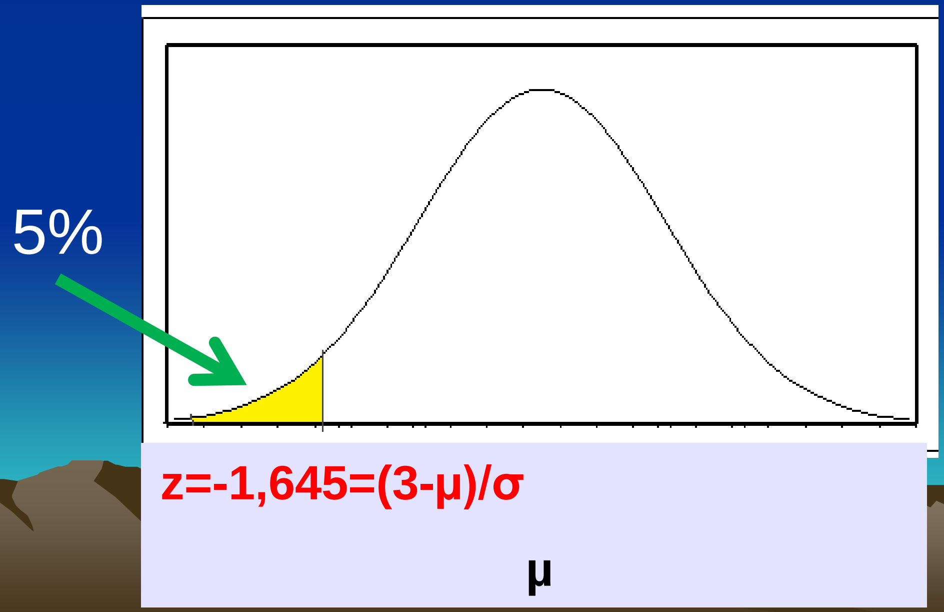
Calcolare

- Il tempo medio di attesa e la relativa varianza
- Pr(Y < 3 min)
- Pr(3 min < Y < 7 min)
- Pr(7 min < Y < 10 min)
- Pr(Y > 10 min)

Tempo di attesa $Y \sim N(\mu, \sigma^2)$.

- Obiettivo: trovare μ e σ^2
- il 95% dei tifosi aspetta almeno 3 minuti→
- $Pr(Y>3)=0,95$
- $Pr(Z>(3-\mu)/\sigma)=0,95$ oppure $Pr(Z<(3-\mu)/\sigma)=0,05$

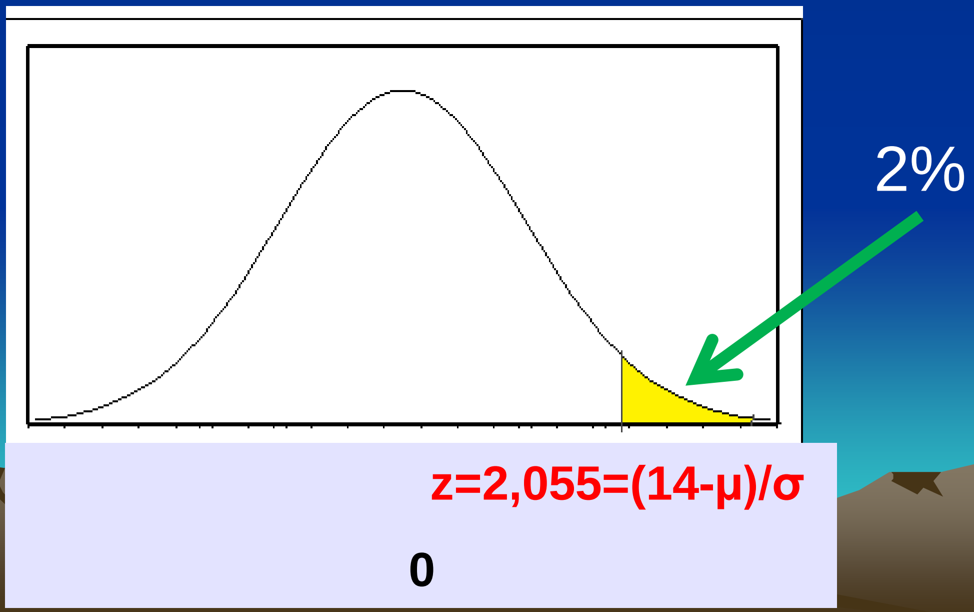
$$(3-\mu)/\sigma=-1,645$$



Tempo di attesa Y ~ N(μ, σ^2).

- Obiettivo: trovare μ e σ^2
- il 2% aspetta più di 14 minuti ➔
- Pr(Y>14)=0,02
- Pr(Z>(14-μ)/σ)=0,02 oppure Pr(Z<(14-μ)/σ)=0,98

(14-μ)/σ=2,055



Risolvendo il sistema

$$\begin{cases} (3-\mu)/\sigma=-1,645 \\ (14-\mu)/\sigma=2,055 \end{cases}$$

Si ottiene

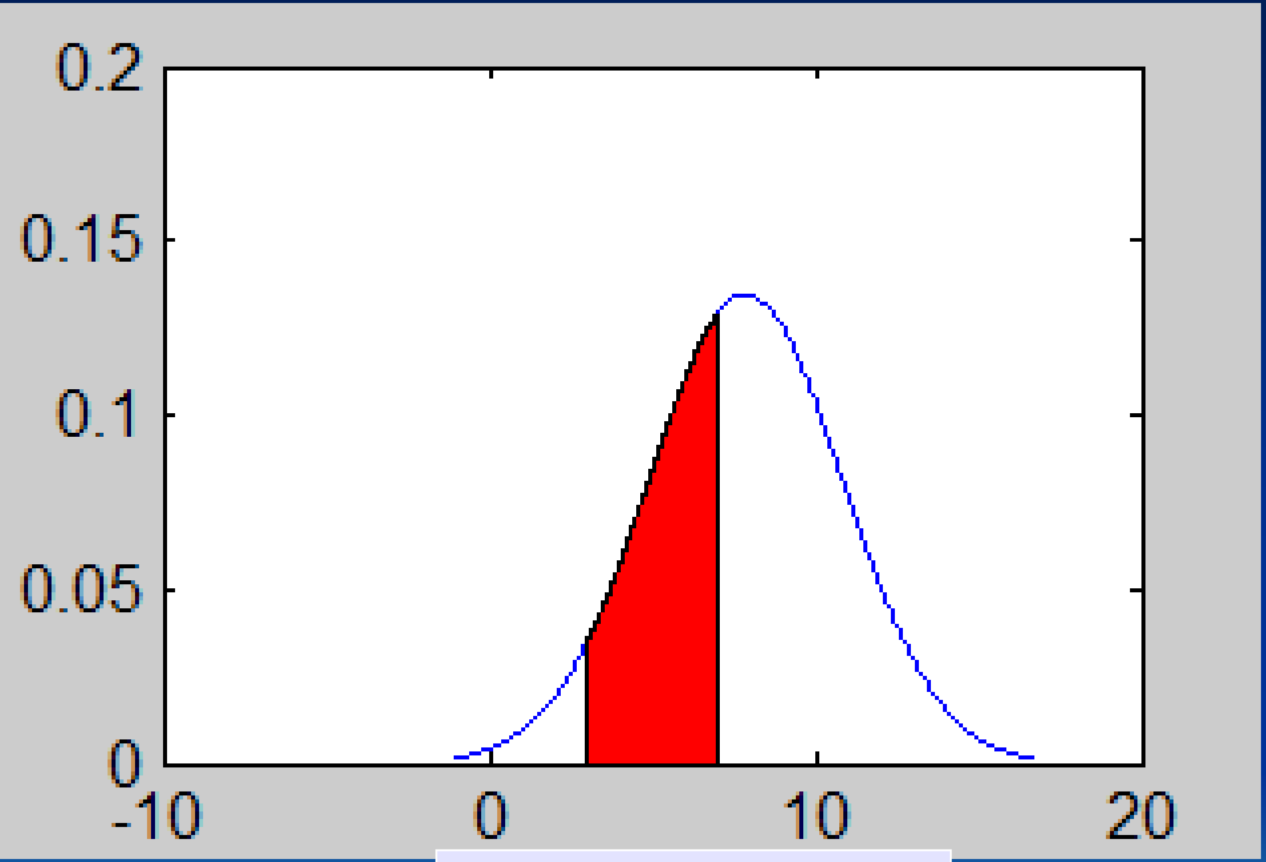
$\mu \approx 7,87$ $\sigma \approx 2,96$

- Controllo della soluzione trovata
- $F((3-7,87)/2,96) \approx F(-1,645) \approx 0,05$
- $F((14-7,87)/2,96) \approx F(2,07) \approx 0,98$

Pr(Y < 3 min)

- Informazione iniziale
- il 95% dei tifosi aspetta almeno 3 minuti→
- Ossia Pr(Y>3)=0,95
- Quindi
- Pr(Y < 3)=?
- Pr(Y < 3)=0,05

Pr(3min < Y < 7min)   Y~N(7,87 2,96²)



3 7

F(7)-F(3)=

F((7-7,87)/2,96)-F((3-7,87)/2,96)=0,336

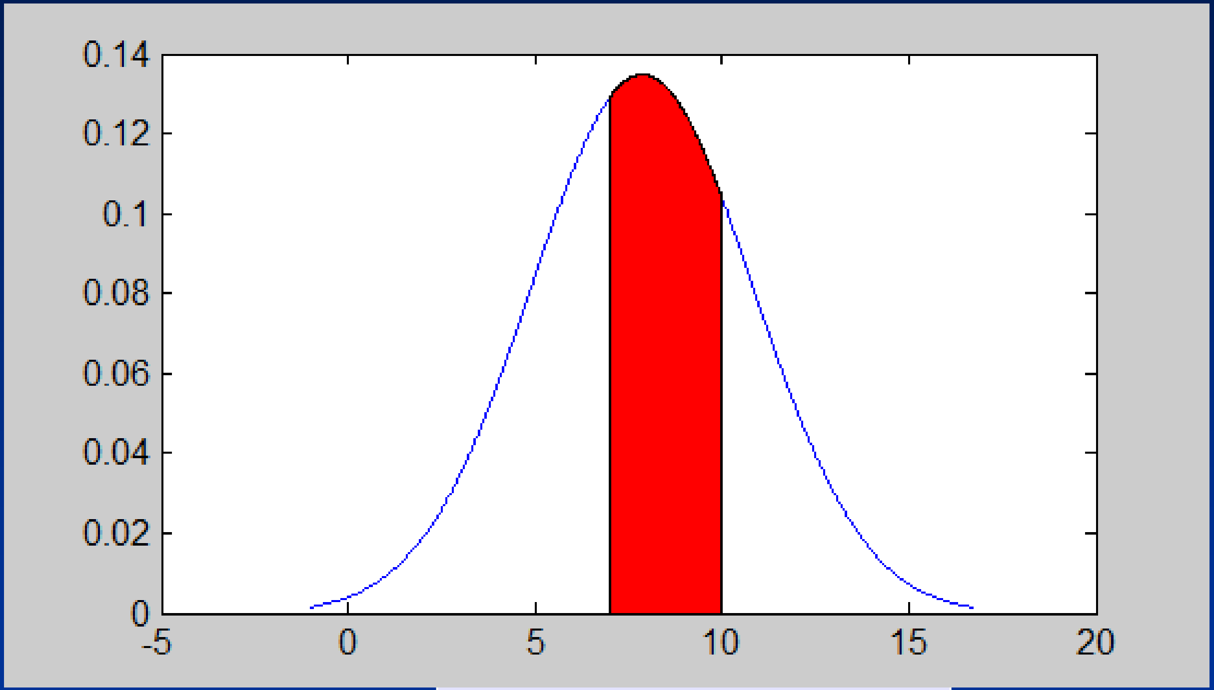
Osservazione: come calcolare F(b)-F(a) nella calcolatrice

Es. Pr(3 < Y < 7)

$Y \sim N(7{,}87\ 2{,}96^2)$

- Dopo essere andati in modo statistico
- SHIFT STAT 5 1:P( ) ➔ Funzione di ripartizione
- P((7-7.87)/2.96)-P((3-7.87)/2.96) = 0.334453

Pr(7min<Y<10min) Y~N(7,87 2,96²)

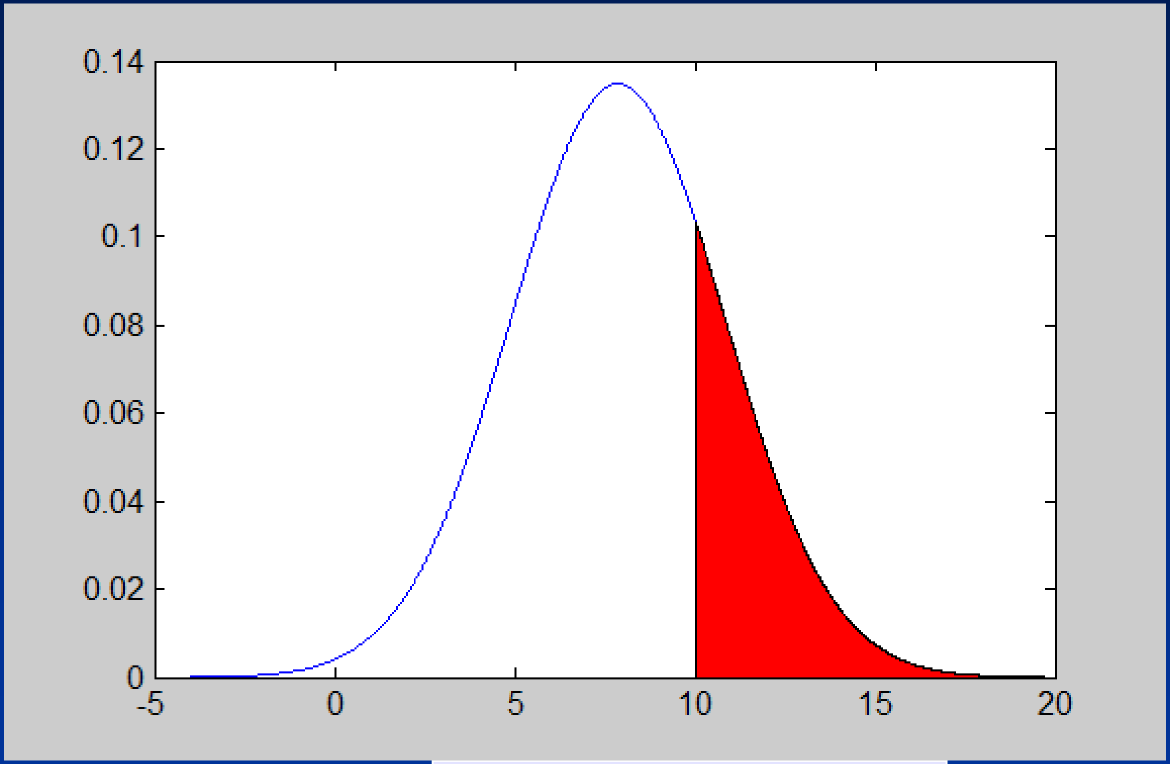


7 10

F(10)-F(7)=

F((10-7,87)/2,96)-F((7-7,87)/2,96)=0,378

Pr(Y >10 min)   Y~N(7,87 2,96²)



10

Pr(Y>10)=1- F(10)

1- F((10-7,87)/2,96) = 0,236

Esercizio

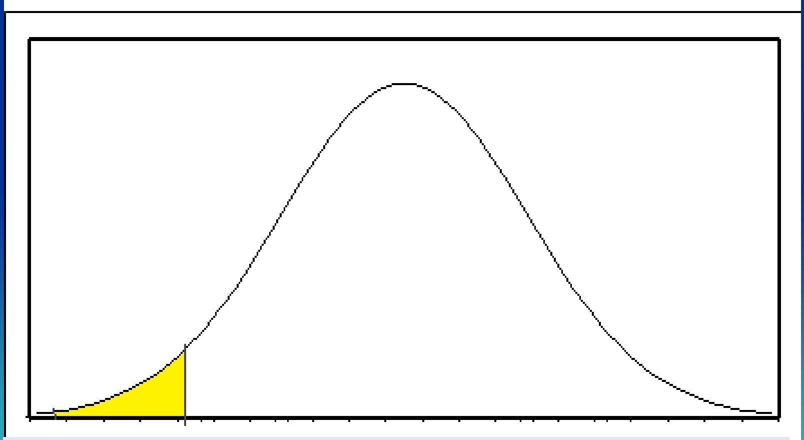
- Un’azienda produttrice di vino offre confezioni contenenti nominalmente 55 cl
- Si osserva che
  il 10% delle confezioni contiene meno di 53,5 cl di vino
  e che il 15% più di 56 cl.
- Assumendo che la distribuzione del contenuto sia normale calcolare:
- il contenuto medio delle bottiglie di vino e la sua varianza
- la prob. che selezionando a caso una bottiglia essa abbia un contenuto minore di quello nominale
- la prob. che selezionando a caso 3 bottiglie al massimo due abbiano un contenuto inferiore a quello nominale

Soluzione

- X= contenuto di una bottiglia di vino ~ N(μ, σ^2).
- Obiettivo: trovare il contenuto medio delle bottiglie di vino e la sua varianza
- Informazioni note il 10% delle confezioni contiene meno di 53,5 cl di vino, il 15% più di 56 cl.

$Pr(X<53,5)=0,10$

$Pr(Z<(53,5-\mu)/\sigma)=0,10$

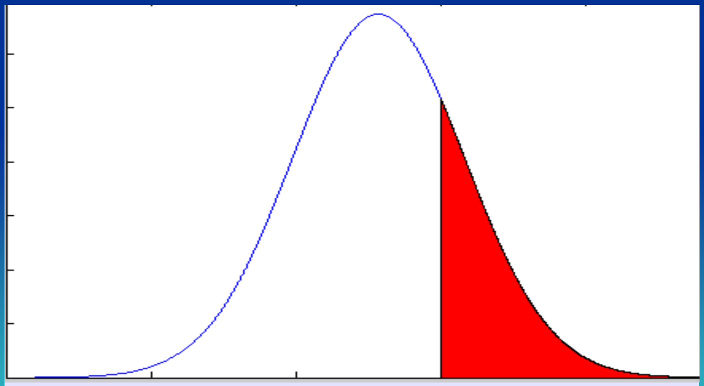


$z=-1,28=(53,5-\mu)/\sigma$

0

$Pr(X>56)=0,15$

$Pr(Z<(56-\mu)/\sigma)=0,85$



$z=1,04=(56-\mu)/\sigma$

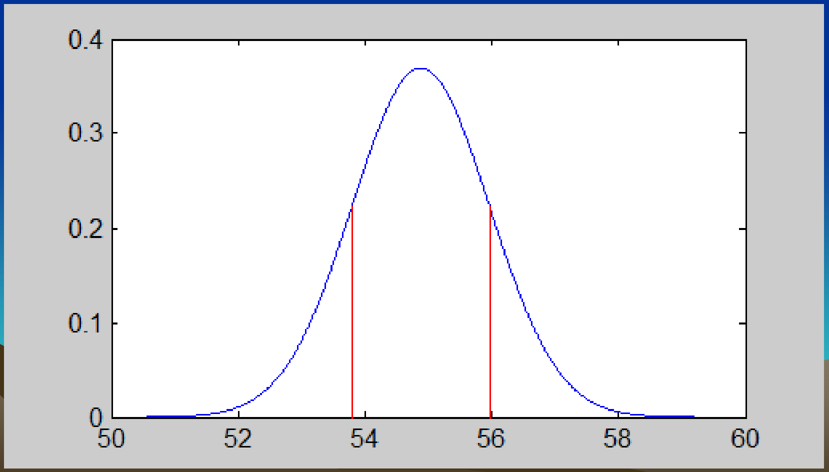
0

Soluzioni

- Obiettivo: trovare il contenuto medio delle bottiglie di vino e la sua varianza
- Risolvendo il sistema

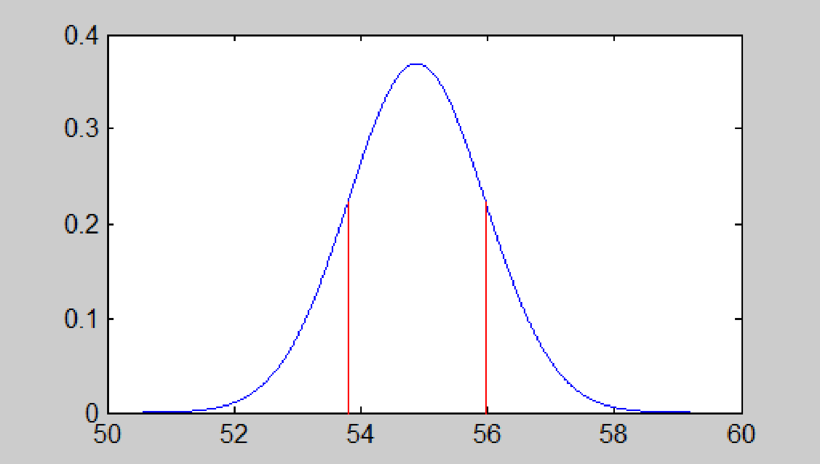
$$\begin{cases} (53{,}5-\mu)/\sigma) = -1.28 \\ (56-\mu)/\sigma) = 1.04 \end{cases}$$

- μ=54,88 cl σ=1,08 cl   var=$1{,}08^2$ cl^2



Soluzioni

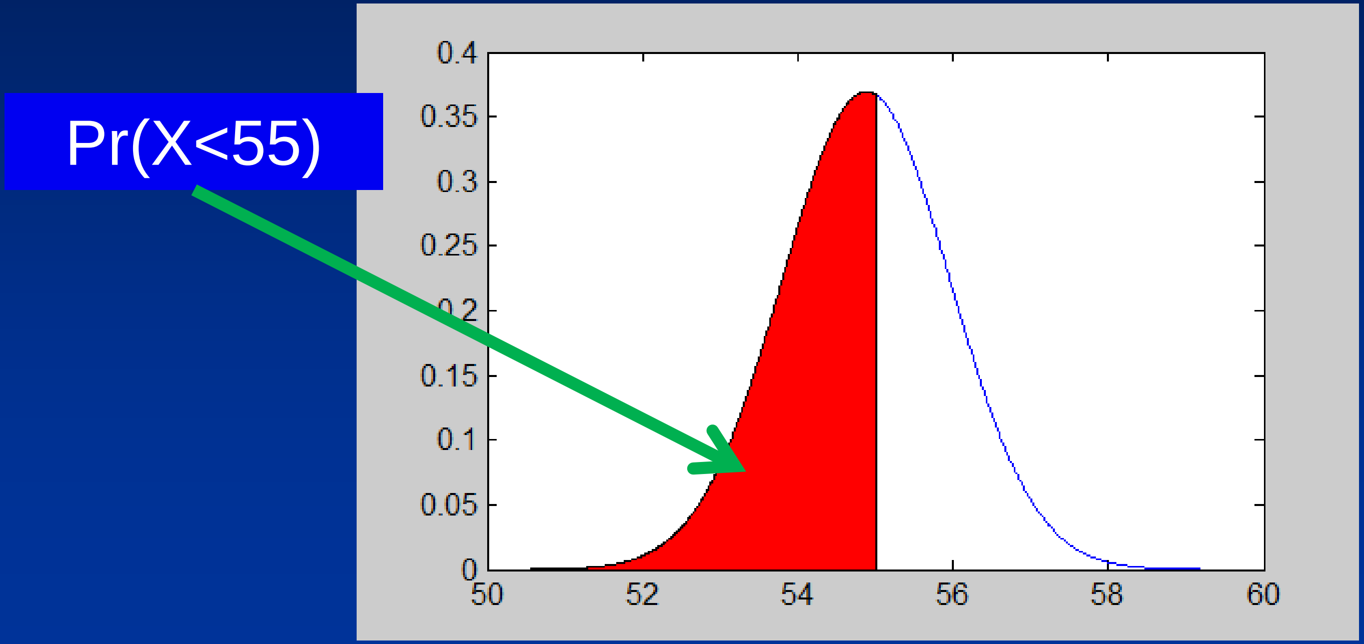
- Obiettivo: la prob. che selezionando a caso una bottiglia essa abbia un contenuto minore di quello nominale (55 cl)
  - μ=54,88 cl   σ=1,08 cl



Obiettivo: prob. in una N(54,88 $1{,}08^2$) di ottenere valori inferiori a 55 cl (Pr (X<55))

Soluzioni

Obiettivo: prob. in una X~N(54,88 $1{,}08^2$) di ottenere valori inferiori a 55 cl



Pr(X<55) = Pr (Z< (55-54,88)/1,08)=

= Pr (Z<0,11) = 0,5438

Soluzioni

- Prob. che selezionando a caso 3 bottiglie al massimo due abbiano un contenuto inferiore a quello nominale

π=Pr(successo)= Pr(contenuto inferiore a quello nominale) =0,5438

Prob. richiesta= Prob. di ottenere al massimo due successi su 3 prove

X ~B (3 0,5438)

$$\sum_{i=0}^{2} Pr(X=i) = \sum_{i=0}^{2} \binom{3}{i} 0,5438^{i}(1-0,5438)^{3-i}$$

$$\sum_{i=0}^{2} Pr(X=i) = 1 - Pr(X=3) = 1 - \binom{3}{3} 0,5438^{3}(1-0,5438)^{0} = 1 - 0,161 = 0,839$$

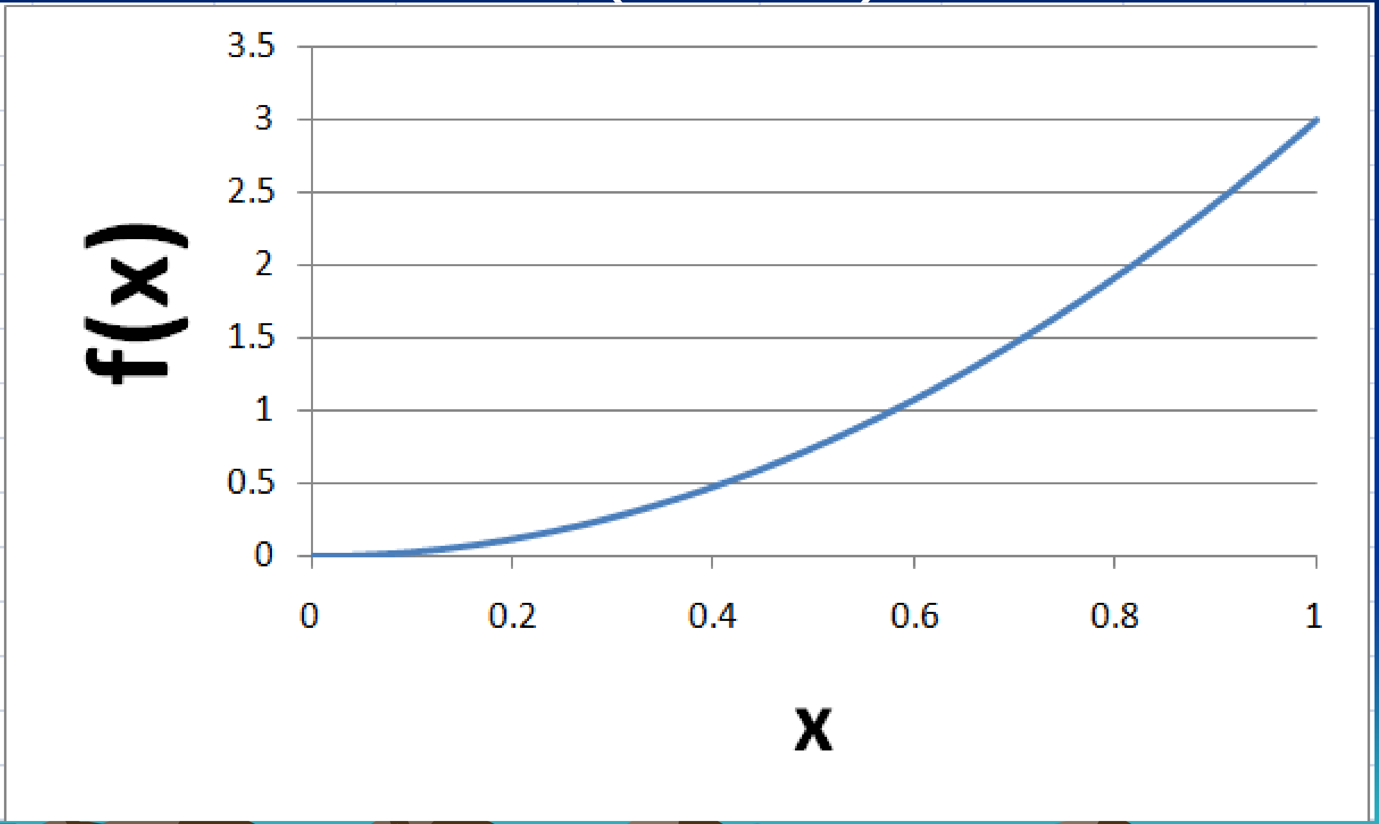
La probabilità richiesta non è piccola!

Esercizio

- Sia X_1 X_2 …, X_{80} un campione casuale proveniente da una popolazione distribuita secondo il modello $f(x)=3x^2$ $(0<x<1)$. Si determini la probabilità che la media campionaria sia minore di 0,8.

Distribuzione del fenomeno nell'universo

Rappresentazione grafica di $f(x)=3x^2$ $(0<x<1)$



$$\int_0^1 f(x)dx = \int_0^1 3x^2 dx = 1$$

Soluzione

- Dato che

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

- Per calcolare la probabilità che la media campionaria sia minore di 0,8 è necessario trovare la media (μ) e la varianza (σ^2) dell'universo

Soluzione

- Passo 1. Calcolare la media e la varianza dell'universo X che presenta densità $f(x)=3x^2$ (0<x<1)

$$E(X)=\int xf(x)dx=\int_0^1 x3x^2dx=\int_0^1 3x^3dx=\frac{3}{4}$$

$$E(X^2)=\int x^2f(x)dx=\int_0^1 x^23x^2dx=\int_0^1 3x^4dx=\frac{3}{5}$$

$$Var(X)=E(X^2)-[E(X)]^2=\frac{3}{5}-\left[\frac{3}{4}\right]^2=\frac{3}{80}$$

- X presenta distribuzione (non normale) con $E(X)=\mu=3/4$ e $VAR(X)=\sigma^2=3/80$

- X presenta distribuzione non normale con $E(X)=\mu=3/4$ e $VAR(X)=\sigma^2=3/80$

- La media campionaria di un campione di 80 osservazioni estratte da X presenta la seguente distribuzione approssimata (per il teorema centrale del limite)

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

$$\overline{X} \approx N\left(\frac{3}{4}; \frac{3}{80}\frac{1}{80}\right)$$

$$Pr(\overline{X} < 0,8) = Pr\left(Z = \frac{(\overline{X} - \mu)}{\sigma/\sqrt{n}} < \frac{0,8 - 3/4}{\sqrt{\frac{3/80}{80}}}\right) = F(2.3094) = 0,98956$$

Esercizio

- Dato un universo con media 6.12, varianza 46 e indice di asimmetria di Fisher pari a 3, calcolare
  - il valore atteso la varianza e l'indice di asimmetria di Fisher della v.a. secondo elemento del campione.
  - il valore atteso e la varianza dello stimatore $T=(X_1+2X_2)/3$.

Soluzione

- U~ (6.12 46)
- Indice di asimmetria di Fisher (γ)= 3

- Dato che $X_1 \dots X_n$ are random variables IID (independent and identically distributed) with the same distribution of X
- $E(X_2)$=? $VAR(X_2)$=? $\gamma(X_2)$?

- $E(X_2)$=6.12 $VAR(X_2)$=46
- Indice di asimmetria di Fisher di X_2 = 3

Soluzione

- U~ (6.12 46)
- $T=(X_1+2X_2)/3$ E(T)?

$E(T)=E((X_1+2X_2)/3)$

$=(1/3)\,[E(X_1)+2\,E(X_2)]=$

$(1/3)(6.12+2\;6.12)= 6.12 = \mu$

Soluzione

- U~ (6.12 46)
- $T=(X_1+2X_2)/3$.   VAR(T)?

- $VAR(T)= VAR((X_1+2X_2)/3)$

$= (1/9)\,[VAR(X_1)+4\ VAR(X_2)]$

$= (5/9)\ 46$

Esercizio

- Si definisce errore quadratico medio (MSE=mean square error) di uno stimatore T di un parametro θ la quantità
- $E(T-\theta)^2$.
  - Dimostrare che se lo stimatore T è corretto il suo MSE coincide con la sua varianza
  - Dimostrare che se lo stimatore T è distorto il suo MSE può essere scritto come: $MSE(T)=VAR(T) + Bias^2$

Soluzione: dimostrare che se lo stimatore T è corretto il suo MSE, $E(T-\theta)^2$, coincide con la sua varianza

- Se T è uno stimatore non distorto di θ allora $E(T)=\theta$ quindi
- $MSE=E(T-\theta)^2$.
- $MSE=E(T-\theta)^2=E(T-E(T))^2=VAR(T)$

Soluzione: Dimostrare che se lo stimatore T è distorto il suo MSE può essere scritto come: MSE(T)=VAR(T) + Bias^2

- $MSE(T)=E(T - \theta)^2 = E(T- E(T) + E(T) - \theta)^2$
- Svolgendo il quadrato si ottiene:
- $MSE(T)=E(T - E(T))^2 + (E(T) - \theta)^2$

  $2*(E(T) - \theta)*E(T - E(T))$
- Il doppio prodotto è zero quindi

$MSE(T)= E(T - E(T))^2 + (E(T) - \theta)^2 = VAR(T)+Bias^2$

Esercizio

Il direttore di un centro commerciale vuole modificare l'orario di apertura del centro. In un campione casuale di 300 clienti, 246 si sono dichiarati favorevoli al nuovo orario proposto.

- Si determini l'intervallo di confidenza della frequenza relativa dell'universo
- con probabilità 0,95
- con probabilità 0,995

e si commentino in termini comparati i suddetti intervalli

$$P\left\{p - z(\alpha/2)s(p) \le \pi \le p + z(\alpha/2)s(p)\right\} = 1 - \alpha$$

$$s(p) = \sqrt{\frac{p(1-p)}{n}}$$

- n = 300 p = 246/300 = 0,82
- Z(P) ~ N(0,1) TCL

$$P\{p - z(\alpha/2)s(p) \le \pi \le p + z(\alpha/2)s(p)\} = 1-\alpha$$

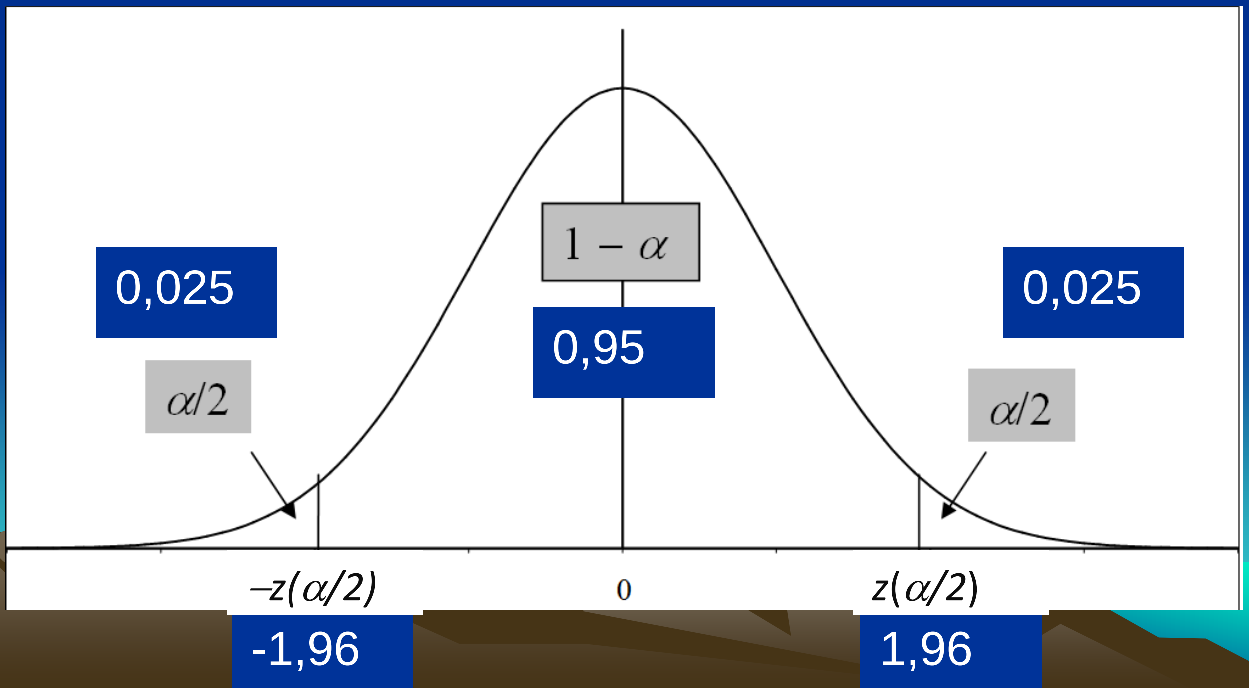
$$s(p) = \sqrt{\frac{0{,}82 \times 0{,}18}{300}} = 0{,}022$$

$$s(p) = \sqrt{\frac{p(1-p)}{n}}$$

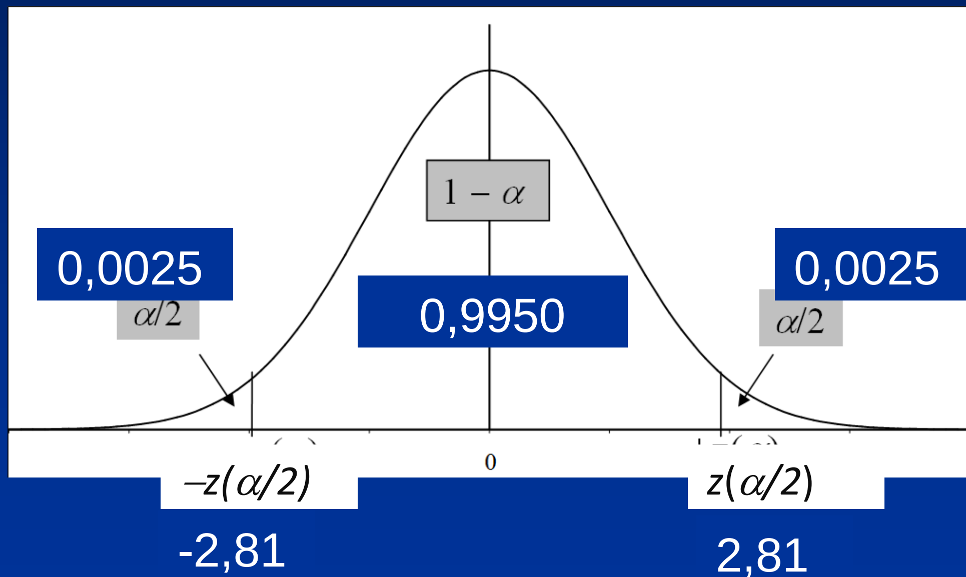
- Intervallo di confidenza al 95%

z(α/2) = ± 1,96 ➔ 0,82 ±1,96 ·0,022

Pr{0,777 ≤ π ≤ 0,863} = 0,95



- n = 300 p = 246/300 = 0,82
- Z(P) ~ N(0,1) teorema centrale del limite
- Intervallo di confidenza al 99,5% (0,995)



F(2,81)=0,9975

$$P\{p - z(\alpha/2)s(p) \le \pi \le p + z(\alpha/2)s(p)\} = 1-\alpha$$

Pr{0,758 ≤π≤ 0,882} = 0,995

Confronto tra i due intervalli di confidenza

$$\Pr\{0{,}777 \leq \pi \leq 0{,}863\} = 0{,}95$$

$$\Pr\{0{,}758 \leq \pi \leq 0{,}882\} = 0{,}995$$

Esercizio

- La deviazione standard della statura degli studenti iscritti ad una università è 5,8 cm. Quanti studenti si devono estrarre a sorte dalla popolazione se si vuole con probabilità del 90% che l'errore di stima della media non superi i 2 cm.

Soluzione: informazioni note X~(μ 5,8²)

- Se l'intervallo di confidenza è al 90% si ottiene

$$P\left\{\bar{X} - 1{,}65\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1{,}65\frac{\sigma}{\sqrt{n}}\right\} = 0{,}90$$

$$P\left\{|\bar{X} - \mu| \le 1{,}65\frac{\sigma}{\sqrt{n}}\right\} = 0{,}90$$

Se vogliamo che l'errore di stima della media non superi i 2 cm $|\bar{X} - \mu| \le 2$

$$1{,}65\frac{\sigma}{\sqrt{n}} \le 2$$

$$n \ge \frac{(1{,}65 * 5{,}8)^2}{4} \approx 23$$

Esercizio: stima della percorrenza media delle vetture diesel di un certo modello al primo guasto

- n=400 $\overline{X}$ =34.000 Km; s_{cor}=9000 Km
- Calcolare l'intervallo di confidenza di μ al 95% e al 99%

$$P\left\{ \overline{X} - z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \right\} = 1-\alpha$$

Esercizio: stima della percorrenza media delle vetture diesel di un certo modello al primo guasto

- n=400 $\overline{X}$ =34.000 Km; s_{cor}=9000 Km
- Livello di confidenza (1-α)=0,95 z(0,025)=1,96

- P{33118 < μ< 34882}=0,95
- Livello di confidenza (1-α)=0,99 z(0,005)=2,58
- P{32839 < μ< 35161}=0,99

Esercizio

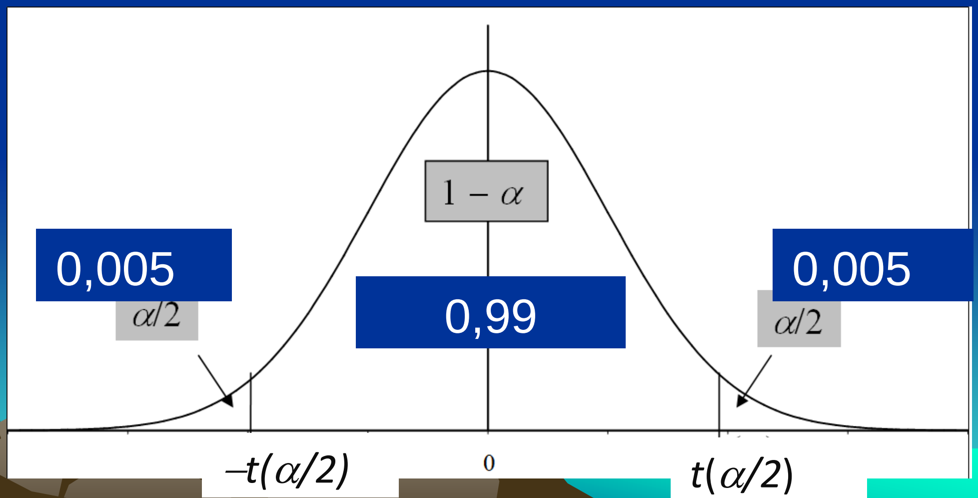
- I dati che seguono si riferiscono alla durata (in migliaia di Km) di una cinghia da automobile in un campione di 15 osservazioni
- 115,4 85,2 89,1 118,3 88,4 109,3 104,3 69,3 105,5 106,8 103,1 101,6 102,9 89,6 109,3
- Facendo le opportune ipotesi, si costruisca un intervallo di confidenza per la media al 99%

Soluzione

- n=15 $\bar{x}$ =99,87 mila Km;
  s^2_{cor}=170,24 Pr( )=0,99
- Ip. Distribuzione normale nell'universo

$$P\left\{ \bar{X} - t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \leq \mu \leq \bar{X} + t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \right\} = 1 - \alpha$$

t(α) deve essere cercato in una v.a. T di student con14 gradi di libertà



0,005

0,99

0,005

-2,977

2,977

Soluzione

- n=15 $\overline{x}$ =99,87 mila Km; s^2_{cor}=170,24
- Ip. Distribuzione normale nell'universo

$$P\left\{\overline{X} - t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + t(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\right\} = 1-\alpha$$

$$P\left\{99{,}87 - 2{,}977\frac{13{,}05}{\sqrt{15}} \le \mu \le 99{,}87 + 2{,}977\frac{13{,}05}{\sqrt{15}}\right\} = 0{,}99$$

$$P\{89{,}84 \le \mu \le 109{,}9\} = 0{,}99$$

Esercizio

- Di seguito sono riportati i Km percorsi in un giorno da un campione di taxi operante in una grande città
- 173 195 115 122 154 149 120 148 152 68 132 91 120 148 103 101
- Sulla base di questo campione assumendo che la popolazione generatrice sia normale è stato determinato il seguente intervallo di confidenza (116,55 144,7). Si calcoli il livello di confidenza su cui è stato calcolato

Soluzione

Media campionaria=130,6875 n=16 s_{cor}=32,21122

$$P\left\{\overline{X} - t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + t(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\right\} = P\{116{,}55 \le \mu \le 144{,}7\} = 1 - \alpha$$

- Equazione da risolvere $\overline{x} + t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} = 144{,}7$

$$t(\alpha/2) = (144{,}7 - \overline{x})\sqrt{n}\,/\,s_{cor} = 1{,}74$$

Dalla tavola t(α/2)=1,74 con g=15 corrisponde ad α di poco superiore a 0,1 ossia ad un 1- α di poco inferiore a 0,9 (Utilizzando la funzione di Excel distrib.t(1,74;15;2) si ottiene α =0,102329)

Variante al precedente esercizio

- Se i dati di base fossero stati i seguenti:
- 172 195 115 122 154 149 120 148 152 68 132 91 120 148 103 101
- Quale sarebbe stato il livello di confidenza dell’intervallo (116,55 144,7)?
- Media campionaria=130,625
- S_{cor}=32,1245 t(α/2)=1,75 ➔ α≈0,10
- ➔ 1- α ≈ 0,9

Esercizio

Nella seguente distribuzione di frequenze è riportato il numero di dipendenti di 50 aziende tessili operanti in una determinata provincia.

Numero di dipendenti	Frequenze assolute
5	12
8	11
12	11
14	8
15	7
545	1

Si calcoli l'intervallo di confidenza al 99% della media dell'universo del numero di dipendenti commentando i risultati ottenuti (con o senza il valore anomalo)

Stima di μ in distribuzioni di frequenze

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i n_i}{n} = \sum_{i=1}^{r} x_i f_i$$

$$n = \sum_{i=1}^{r} n_i \qquad \text{e} \qquad f_i = n_i/n.$$

Stima corretta di σ in presenza di distribuzioni di frequenze

$$s_{cor} = \sqrt{\frac{\sum_{i=1}^{r} \left(x_i - \overline{x}\right)^2 \cdot n_i}{\sum_{i=1}^{r} n_i - 1}}$$

Soluzione (con il valore anomalo)

- M=20,84 s_{cor}^2=5735,525

$$P\left\{\overline{X} - z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + z(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\right\} = 1-\alpha$$

- $s_{cor}/n^{0.5}$ = 10,710   z(α/2)=2,58

Estremo inferiore = -6,793 → 0

Estremo superiore = 48,47

Pr(0 < μ < 48,47) =0,99

Soluzione (senza il valore anomalo)

- $M=10{,}143 \quad s_{cor}^2= 14{,}375$
- $s_{cor}/49^{0.5} = 0{,}542 \qquad z(\alpha)=2{,}58$

$$P\left\{ \overline{X} - z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \right\} = 1-\alpha$$

Estremo inferiore = 8,745

Estremo superiore = 11,540

$Pr(8{,}745 < \mu < 11{,}540) = 0{,}99$

Esercizio

Un’azienda produce rotoli di stoffa della lunghezza di 70m. Tali rotoli possono presentare difetti di diversa natura. L’azienda è interessata a stimare il numero medio di difetti presenti nei rotoli prodotti. In un campione casuale di 85 rotoli si è trovata la seguente distribuzione

n. difetti	0	1	2	3	4	5	6
Frequenza	16	26	22	13	5	2	1

Si determini l’intervallo di confidenza al 99% per la media dei difetti presenti nei rotoli di stoffa

Soluzione

- Media campionaria=1,7059 n=85
- S^2= 1,760554 S_{cor}=1,3347

$$P\left\{\overline{X}-2{,}58\frac{s_{cor}}{\sqrt{n}}\leq\mu\leq\overline{X}+2{,}58\frac{s_{cor}}{\sqrt{n}}\right\}=0{,}99$$

$$P\{1{,}33\leq\mu\leq 2{,}08\}=0{,}99$$

Esercizio

Con riferimento all’esercizio precedente, si consideri che un rotolo risulta vendibile se presenta un massimo di 3 difetti. Sulla base dello stesso campione di cui all’esercizio precedente, si costruisca un intervallo di confidenza al 95% per la proporzione di rotoli considerati vendibili

Soluzione

n. difetti	0	1	2	3	4	5	6
Frequenza	16	26	22	13	5	2	1

- Proporzione di successi nel campione= (16+26+22+13)/85=0,9059=p

$$P\{p-1{,}96s(p)\leq\pi\leq p+1{,}96s(p)\}=0{,}95$$

$$s(p)=\sqrt{\frac{p(1-p)}{n}}$$

$$P\left\{0{,}91-1{,}96\sqrt{\frac{0{,}91(1-0{,}91)}{85}}\leq\pi\leq 0{,}91+1{,}96\sqrt{\frac{0{,}91(1-0{,}91)}{85}}\right\}=0{,}95$$

$$P\{0{,}84\leq\pi\leq 0{,}97\}=0{,}95$$

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.
- 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 e si calcoli l'espressione che consente di calcolare la probabilità di ottenere un numero di pezzi fuori norma compreso tra due e quattro (estremi compresi).
  - rappresentare graficamente la densità

Soluzione: intervallo di confidenza

- p= 15/800= 0,01875 z(0,015)= ? s(p)=?
- z(0,015)= 2,17 n=800

s(p)=[0,01875(1-0,01875)]^{0,5}/(800^{0,5})=0,0047956

$$P\left\{p - 2{,}17s(p) \leq \pi \leq p + 2{,}17s(p)\right\} = 0{,}97$$

- Estremo inferiore = 0,01875-2,17*0,0047956=0,008
- Estremo superiore = 0,01875+2,17*0,0047956=0,029

Pr(0,008<π<0,029)=0,97

Parte 2

- 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 e si calcoli l'espressione che consente di calcolare la probabilità di ottenere un numero di pezzi fuori norma compreso tra due e quattro (estremi compresi).
  - rappresentare graficamente la densità

Soluzione: parte 2

- π =0,015
- n=5
- X=numero di pezzi fuori norma~(5, 0,015)

- Pr(X=2)=0,00215

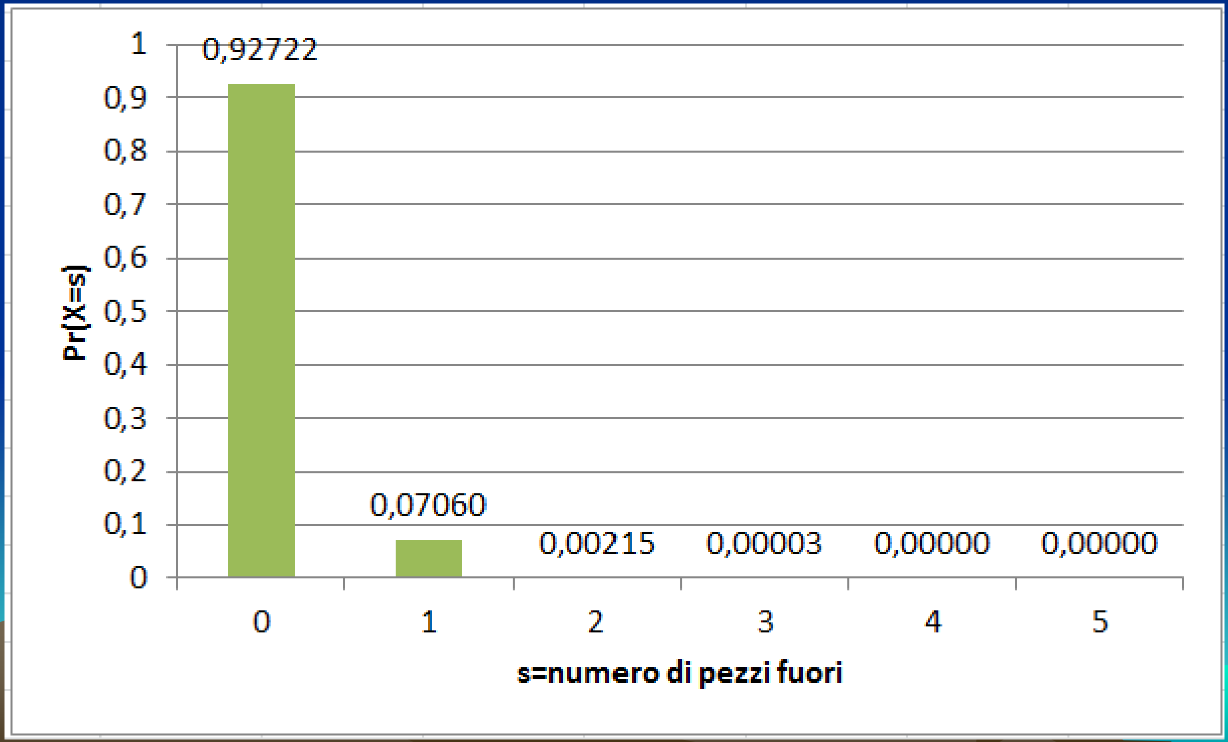
$$Pr(X=2)=\binom{5}{2}0,015^2(1-0,015)^{5-2}=0,00215$$

Probabilità di ottenere un numero di pezzi fuori norma compreso tra due e quattro (estremi compresi).

$$\sum_{i=2}^{4} Pr(X=i)=\sum_{i=2}^{4}\binom{5}{i}0,015^i(1-0,015)^{5-i}=0,002183$$

Soluzione: rappresentazione grafica densità

- π =0,015
- n=5
- X=numero di pezzi fuori norma~(5, 0,015)



Esercizio

- Data una scheda telefonica da 5 euro di cui non si sa se sia mai stata usata e nel caso sia stata usata non si conosce l'ammontare ancora disponibile, è ragionevole ipotizzare per tale ammontare X la seguente funzione di densità $f(x)=1/5$ per $[0 \leq x \leq 5]$

- Verificare che $f(x)=1/5$ per $[0 \leq x \leq 5]$ è una densità e rappresentarla graficamente
- Calcolare il credito residuo atteso (E(X))
- Calcolare la varianza del credito residuo (VAR(X))
- Devo fare una telefonata da 2 € calcolare la prob che la scheda sia sufficiente per fare la telefonata
- Ho 60 schede tutte con un ammontare che si distribuisce come descritto sopra. Qual è la prob che l'ammontare complessivo sia superiore a 170 €

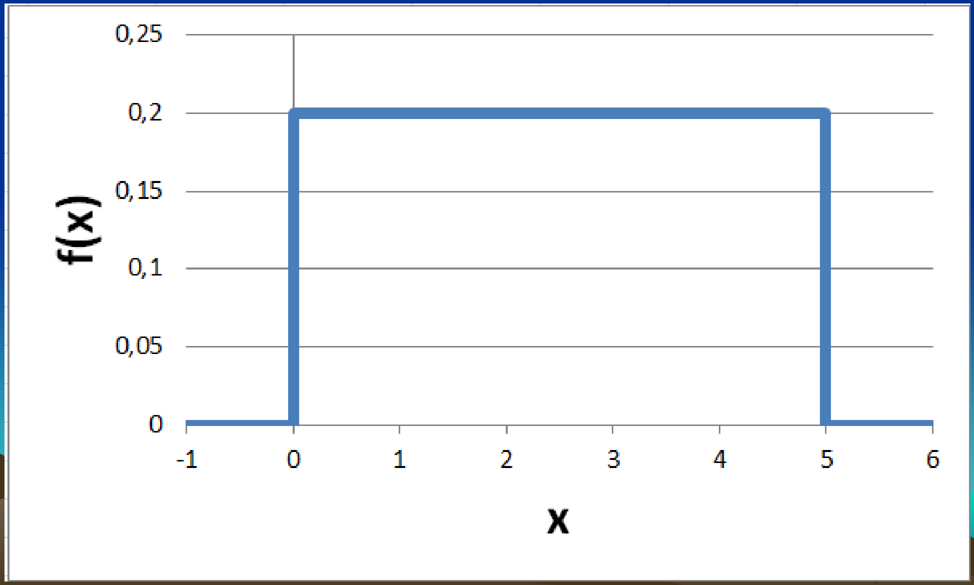
Soluzione

- Verificare che f(x)=1/5 per [0 ≤x≤5] sia una densità

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^5 \frac{1}{5}dx = 1$$

$$f(x) \geq 0$$

f(x)=1/5=0,2 0<x<5



Calcolo di E(X) e VAR(X)

$$E(X) = \int_0^5 x f(x) dx = \int_0^5 x \frac{1}{5} dx = 2{,}5$$

$$\begin{aligned} VAR(X) &= E(X^2) - [E(X)]^2 \\ &= \int_0^5 x^2 f(x) dx - \left[\int_0^5 x f(x) dx\right]^2 \\ &= 5^2/12 \end{aligned}$$

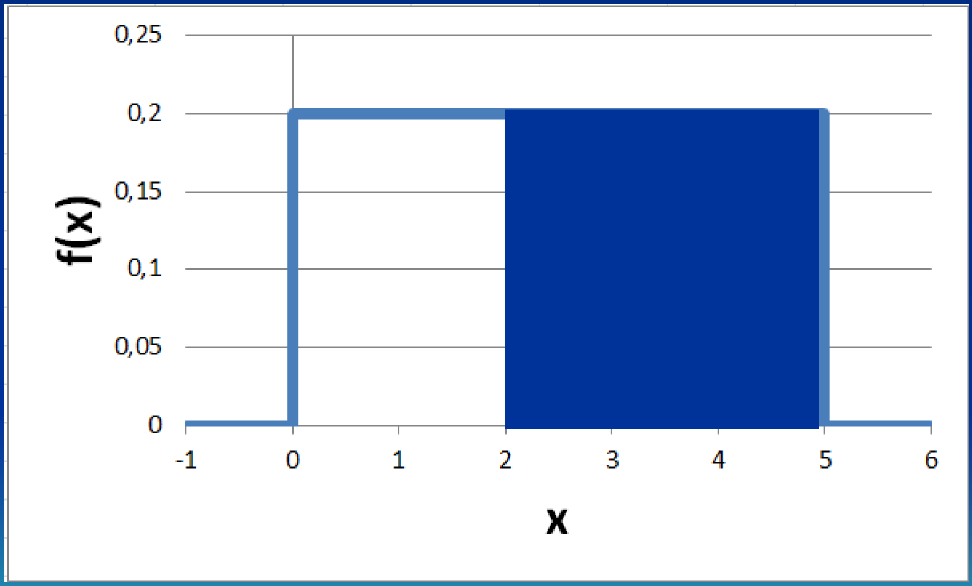
$f(x)=1/5=0,2 \quad 0<x<5$

Soluzione (continua)

- Devo fare una telefonata da 2 € calcolare la prob che la scheda sia sufficiente per fare la telefonata
- Pr(credito residuo>2)

Calcolo di Pr(X>2)

$$Pr(X > 2) = \int_2^5 f(x)dx = \int_2^5 \frac{1}{5}dx = 3/5$$



La prob richiesta è l'area di un rettangolo con base 3 e altezza 0,2

Soluzione (continua)

- Ho 60 schede tutte con un ammontare che si distribuisce come descritto sopra. Qual è la prob che l’ammontare complessivo sia superiore a 170 €
- Pr(credito residuo totale>170)

$Pr(X_1+X_2+. . .+X_{60})>170$

- Sia X_i la variabile casuale che descrive la disponibilità residua (in euro) dell'i-esima scheda telefonica, i = 1, . . . , 60
- $T = X_1+X_2+. . .+X_{60}$
- la variabile casuale che descrive l'ammontare complessivo delle 60 schede
- Obiettivo: calcolare Pr(T>170)
- Qual è la distribuzione di T?

$f(x)=1/5=0,2 \quad 0<x<5 \quad E(X)=2,5 \quad VAR(X)=25/12$

$Pr(T)>170$

- Dato che $X_1+X_2+...+X_{60}$ sono v.c. iid
- $T= X_1+X_2+...+X_{60} \approx N(E(T)\ VAR(T))$
- $E(T)? \quad VAR(T)?$

- Calcolo E(T) e VAR(T)

$$E(T)= E(X_1)+E(X_2)+...+E(X_{60})=60\times2,5=150$$

$$VAR(T)= VAR(X_1)+VAR(X_2)+...+VAR(X_{60})=$$
$$=(25/12)\times60=125$$

$$T \approx N(150\ 125)$$

Pr(T)>170

- T ≈ N(150 125)

- Pr(T)>170=1-Pr(T)<170

$$1-F((170-150)/125^{0,5})$$
$$=1-F(1{,}788854)=0{,}03682$$

Esercizio

- Si consideri una popolazione distribuita secondo il seguente modello

X	Pi
2	0.3
5	0.6
7	0.1

- Si elenchino tutti i campioni di ampiezza 3 che si possono estrarre con ripetizione da tale popolazione assegnando a ciascun campione la relativa probabilità
- Si determini la distribuzione campionaria della media e la si rappresenti graficamente
- Si calcoli il valore atteso e la varianza della media campionaria
- Si determini la distribuzione campionaria della mediana ed il suo valore atteso

Soluzione: spazio dei campioni ($27=3^3$) e relative probabilità

Tabella 1: Spazio dei campioni e relative probabilità

Campioni	Prob	Campioni	Prob	Campioni	Prob
2,2,2	0,027	5,2,2	0,054	7,2,2	0,009
2,2,5	0,054	5,2,5	0,108	7,2,5	0,018
2,2,7	0,009	5,2,7	0,018	7,2,7	0,003
2,5,2	0,054	5,5,2	0,108	7,5,2	0,018
2,5,5	0,108	5,5,5	0,216	7,5,5	0,036
2,5,7	0,018	5,5,7	0,036	7,5,7	0,006
2,7,2	0,009	5,7,2	0,018	7,7,2	0,003
2,7,5	0,018	5,7,5	0,036	7,7,5	0,006
2,7,7	0,003	5,7,7	0,006	7,7,7	0,001
Tot. prob.	0,3		0,6		0,1

X	Pi
2	0.3
5	0.6
7	0.1

Universo X

$$E(X) = \mu = 2 \times 0.3 + 5 \times 0.6 + 7 \times 0.1 = 4.3$$

Distribuzione della media campionaria

Tabella 1: Spazio dei campioni e relative probabilità

Campioni	Prob	Campioni	Prob	Campioni	Prob
2,2,2	0,027	5,2,2	0,054	7,2,2	0,009
2,2,5	0,054	5,2,5	0,108	7,2,5	0,018
2,2,7	0,009	5,2,7	0,018	7,2,7	0,003
2,5,2	0,054	5,5,2	0,108	7,5,2	0,018
2,5,5	0,108	5,5,5	0,216	7,5,5	0,036
2,5,7	0,018	5,5,7	0,036	7,5,7	0,006
2,7,2	0,009	5,7,2	0,018	7,7,2	0,003
2,7,5	0,018	5,7,5	0,036	7,7,5	0,006
2,7,7	0,003	5,7,7	0,006	7,7,7	0,001
Tot. prob.	0,3		0,6		0,1

$\bar{X}$	p_i
2	0.027
3	0.162
3.67	0.027
4	0.324
4.67	0.108
5	0.216
5.33	0.009
5.67	0.108
6.33	0.018
7	0.001
	1

X	Pi
2	0.3
5	0.6
7	0.1

Universo X

$$E(X) = \mu = 2 \times 0.3 + 5 \times 0.6 + 7 \times 0.1 = 4.3$$

$$var(X) = \sigma^2 = (2-4.3)^2 0.3 + (5-4.3)^2 0.6 + (7-4.3)^2 0.1 = 2.61$$

$\overline{X}$	p_i
2	0.027
3	0.162
3.67	0.027
4	0.324
4.67	0.108
5	0.216
5.33	0.009
5.67	0.108
6.33	0.018
7	0.001
	1

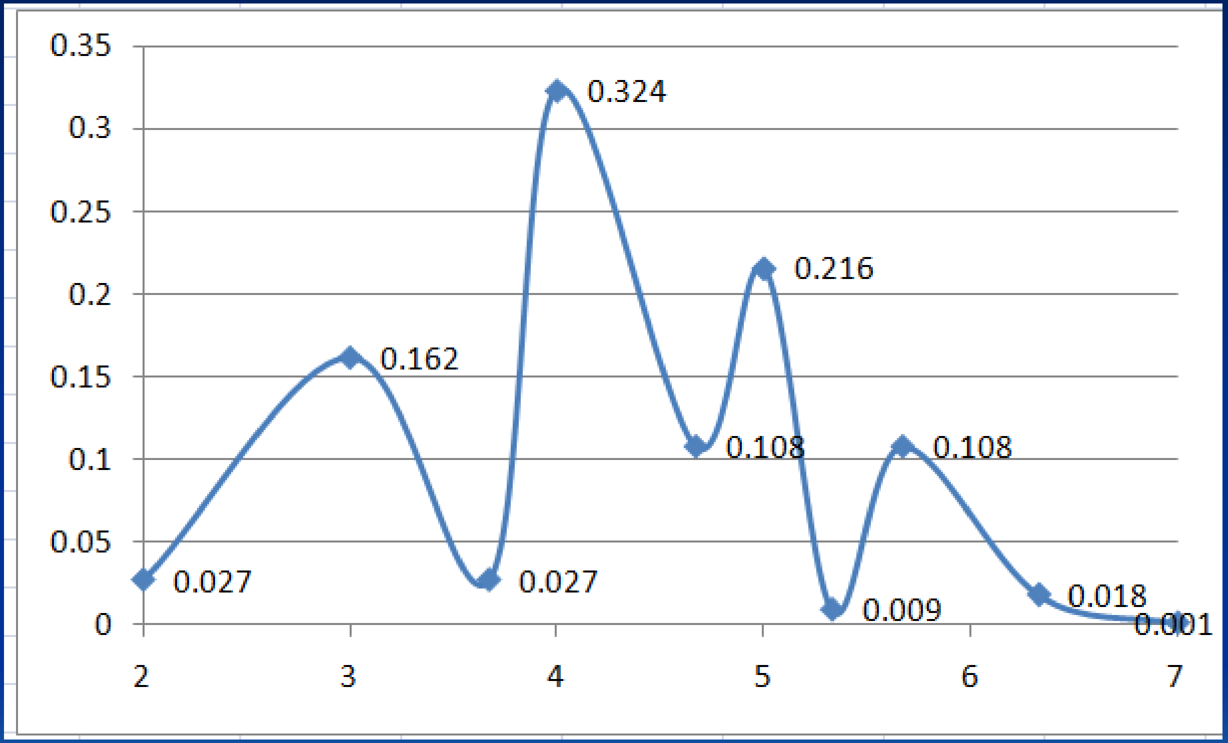
Distribuzione della media campionaria

$$E(\overline{X}) = \sum_{i=1}^{10} \overline{x}_i p_i = 2 \times 0.027 + \cdots + 7 \times 0.001 = 4.3 = \mu$$

$$var(\overline{X}) = (2-4.3)^2 0.027 + \cdots + (7-4.3)^2 0.001 = \frac{2.61}{3} = \frac{\sigma^2}{n}$$

Rappresentazione grafica della distribuzione della media campionaria

$\bar{X}$	p_i
2	0.027
3	0.162
3.67	0.027
4	0.324
4.67	0.108
5	0.216
5.33	0.009
5.67	0.108
6.33	0.018
7	0.001
	1



- Quando n è elevato la distribuzione della media campionaria è normale. Quando n è piccolo la distribuzione dipende da quella dell'universo

X	Pi
2	0.3
5	0.6
7	0.1

Universo X

$$E(X) = \mu = 2 \times 0.3 + 5 \times 0.6 + 7 \times 0.1 = 4.3$$

Distribuzione della mediana campionaria

Tabella 1: Spazio dei campioni e relative probabilità

Campioni	Prob	Campioni	Prob	Campioni	Prob
2,2,2	0,027	5,2,2	0,054	7,2,2	0,009
2,2,5	0,054	5,2,5	0,108	7,2,5	0,018
2,2,7	0,009	5,2,7	0,018	7,2,7	0,003
2,5,2	0,054	5,5,2	0,108	7,5,2	0,018
2,5,5	0,108	5,5,5	0,216	7,5,5	0,036
2,5,7	0,018	5,5,7	0,036	7,5,7	0,006
2,7,2	0,009	5,7,2	0,018	7,7,2	0,003
2,7,5	0,018	5,7,5	0,036	7,7,5	0,006
2,7,7	0,003	5,7,7	0,006	7,7,7	0,001
Tot. prob.	0,3		0,6		0,1

Me	p_i
2	0.216
5	0.756
7	0.028
	1

X	P_i
2	0.3
5	0.6
7	0.1

Universo X
E(X)=4.3

Distribuzione della mediana campionaria

Me	p_i
2	0.216
5	0.756
7	0.028
	1

E(Me)=4.408. In questo caso lo stimatore mediana campionaria è distorto Bias=0.108

Esercizio

- Sia $X_1, \dots, X_n$ un campione casuale estratto da un universo X con la seguente distribuzione di Cauchy (T di student con un solo grado di libertà)

$$f(x;\theta,d) = \frac{1}{\pi d}\frac{1}{1+\left(\frac{x-\theta}{d}\right)^2} \qquad -\infty < x < \infty \qquad d > 0$$

Richieste

- Verificare che $f(x; \theta, d)$ è una densità
- Rappresentare graficamente $f(x; \theta, d)$
- Calcolare la funzione di ripartizione $F(x)$
- Calcolare la mediana di X
- Calcolare $E(X)$
- Illustrare se in presenza di un campione casuale estratto da questa densità è possibile applicare il teorema centrale del limite

Richieste

- Verificare che 6,314 (ossia il numero all'incrocio della prima riga e della prima colonna della tabella di p. 150 del testo di inferenza) è il quantile che lascia alla sua sinistra una probabilità pari a 0,95
- Trovare il quantile 0,995 (ossia il valore che lascia alla sua destra una probabilità pari a 0,005). Verificare che tale numero risulta uguale a 63,656 (v. tabella di p. 150 del libro di inferenza)

Soluzione

- Verifica che è una densità

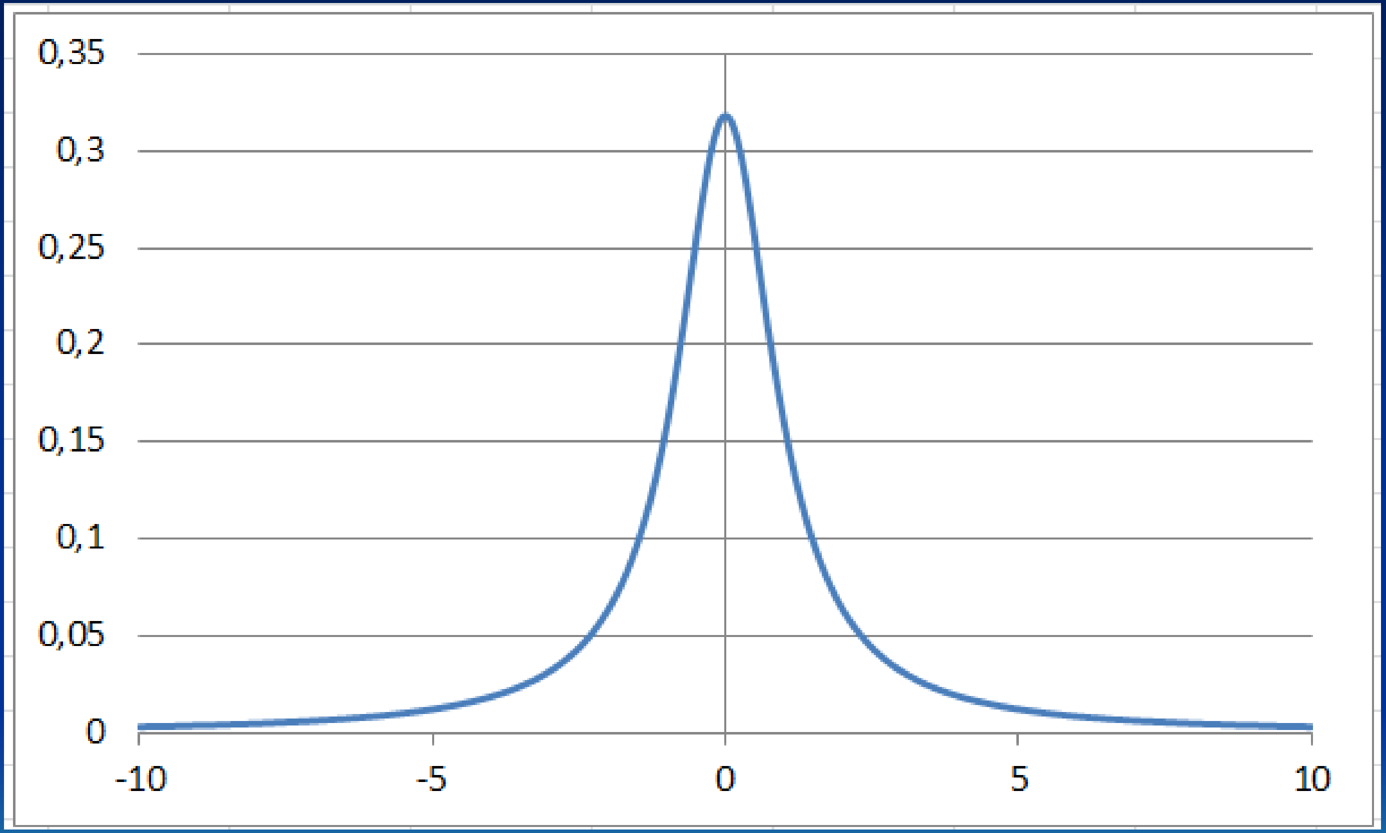
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{\infty} f(x;\theta,d)dx = \frac{1}{\pi d}\int_{-\infty}^{\infty} \frac{1}{1+\left(\frac{x-\theta}{d}\right)^2}dx$$

$$\int_{-\infty}^{\infty} f(x;\theta,d)dx = \frac{1}{\pi} arctg\frac{x-\theta}{d}\Big|_{-\infty}^{+\infty} = 1$$

- Per chi desidera ripassare le proprietà dell'arcotangente http://it.wikipedia.org/wiki/Arcotangente

Rappresentazione grafica θ=0, d=1



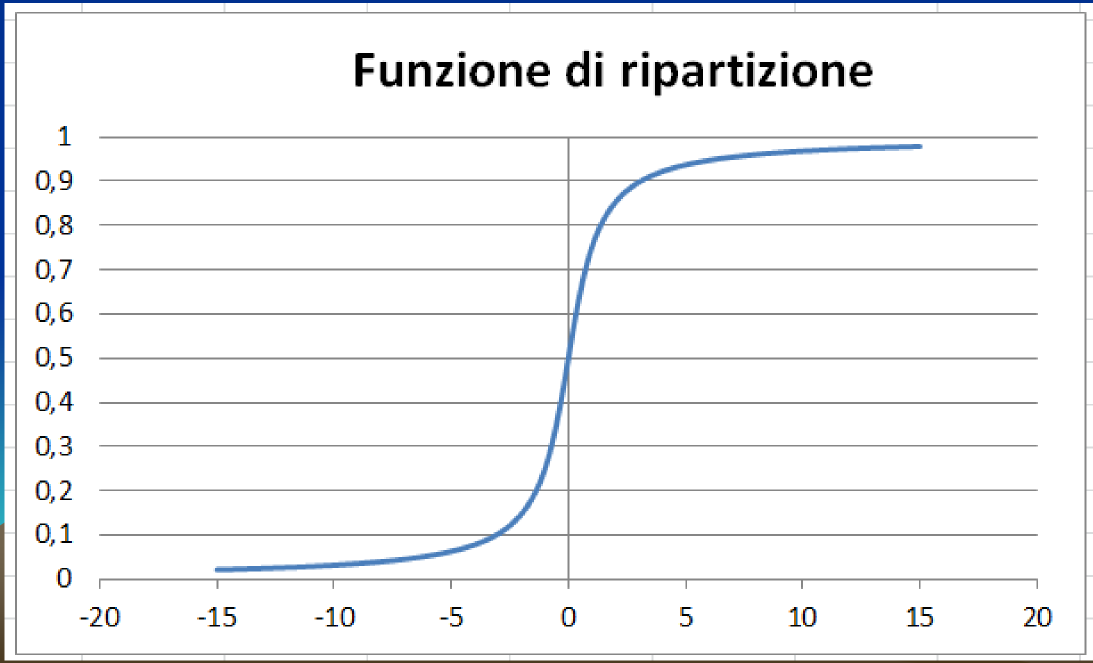
$$f(x;\theta,d)=\frac{1}{\pi d}\frac{1}{1+\left(\frac{x-\theta}{d}\right)^2} \qquad -\infty < x < \infty \qquad d>0$$

Calcolo della funzione di ripartizione

$$F(x) = \frac{1}{\pi d}\int_{-\infty}^{x} \frac{1}{1+\left(\frac{t-\theta}{d}\right)^2}dt$$

$$F(x) = \frac{1}{\pi} arctg \frac{t-\theta}{d}\Big|_{-\infty}^{x}$$

$$F(x) = \frac{1}{2} + \frac{1}{\pi} arctg \frac{x-\theta}{d}$$



θ=0, d=1

Calcolo della mediana

- La mediana (Me) per un v.c. continua con funzione di densità f(x) e ripartizione F(x) è definita come la soluzione della seguente equazione

$$F(Me) = \int_{-\infty}^{Me} f(x)dx = 0{,}5$$

$$Me = F^{-1}(0,5)$$

Mediana = quantile che lascia alla sua destra ed alla sua sinistra una probabilità pari a 0,5

Calcolo della mediana

$$f(x) = \frac{1}{\pi d} \int_{-\infty}^{Me} \frac{1}{1+\left(\frac{x-\theta}{d}\right)^2} = 0{,}5$$

$$\frac{1}{\pi} arctg \frac{x-\theta}{d} \Big|_{-\infty}^{Me} = 0{,}5$$

$$\frac{1}{\pi} \left(arctg \frac{Me-\theta}{d} + \frac{\pi}{2} \right) = 0{,}5$$

$$arctg \frac{Me-\theta}{d} = 0$$

$$Me = \theta$$

Calcolo del valore atteso

- Per semplificare i calcoli possiamo considerare la variabile di Cauchy in forma «standardizzata» z= (x-θ)/d

$$f(x;\theta,d) = \frac{1}{\pi d}\frac{1}{1+\left(\frac{x-\theta}{d}\right)^2} \qquad -\infty < x < \infty \qquad d > 0$$

$$f(z;0,1) = \frac{1}{\pi}\frac{1}{1+z^2} \qquad -\infty < z < \infty$$

$$E(Z) = \int_{-\infty}^{+\infty} z f(z) dz = 2\frac{1}{\pi}\int_0^{+\infty} \frac{z}{1+z^2} dz$$

$$E(Z) = \frac{1}{\pi}\log(1+z^2)\Big|_0^{+\infty}$$

$$E(Z) = +\infty$$

- Domanda: illustrare se in presenza di un campione casuale estratto da questa densità è possibile applicare il teorema centrale del limite
- Risposta: non è possibile applicare il teorema centrale del limite in quanto $E(X)=\infty$. Di conseguenza lo scostamento standardizzato della media campionaria non si distribuisce come una v.c. normale standardizzata

Verificare che 6,314 (ossia il numero all'incrocio della prima riga e della prima colonna della tabella di p. 150 del testo di inferenza) è il quantile che lascia alla sua sinistra una probabilità pari a 0,95

Occorre verificare che

$$\int_{-\infty}^{6{,}314} f(z)dz = 0{,}95$$

$$F(6{,}314) = 0{,}95$$

$$F(z) = \frac{1}{2} + \frac{1}{\pi} arctg z$$

$$F(6{,}314) = \frac{1}{2} + \frac{1}{\pi} arctg(6{,}314)$$

$$F(6{,}314) = \frac{1}{2} + \frac{1}{\pi} 1{,}4137 = 0{,}95$$

Trovare il quantile 0,995 (ossia il valore che lascia alla sua destra una probabilità pari a 0,005). Verificare che tale numero risulta uguale a 63,656 (v. tabella di p. 150 del libro di inferenza)

Occorre trovare $x_{0,995 \text{ tale}}$ per cui $F(x_{0,995})=0,995$

$$F(x_{0,995}) = \frac{1}{2} + \frac{1}{\pi} arctg(x_{0,995}) = 0,995$$

$$arctg(x_{0,995}) = 0,495 \times \pi$$

$$x_{0,995} = tg(0,495 \times \pi) = 63,65674$$

Osservazione: in Excel =INV.T(0,01;1) =63,65674