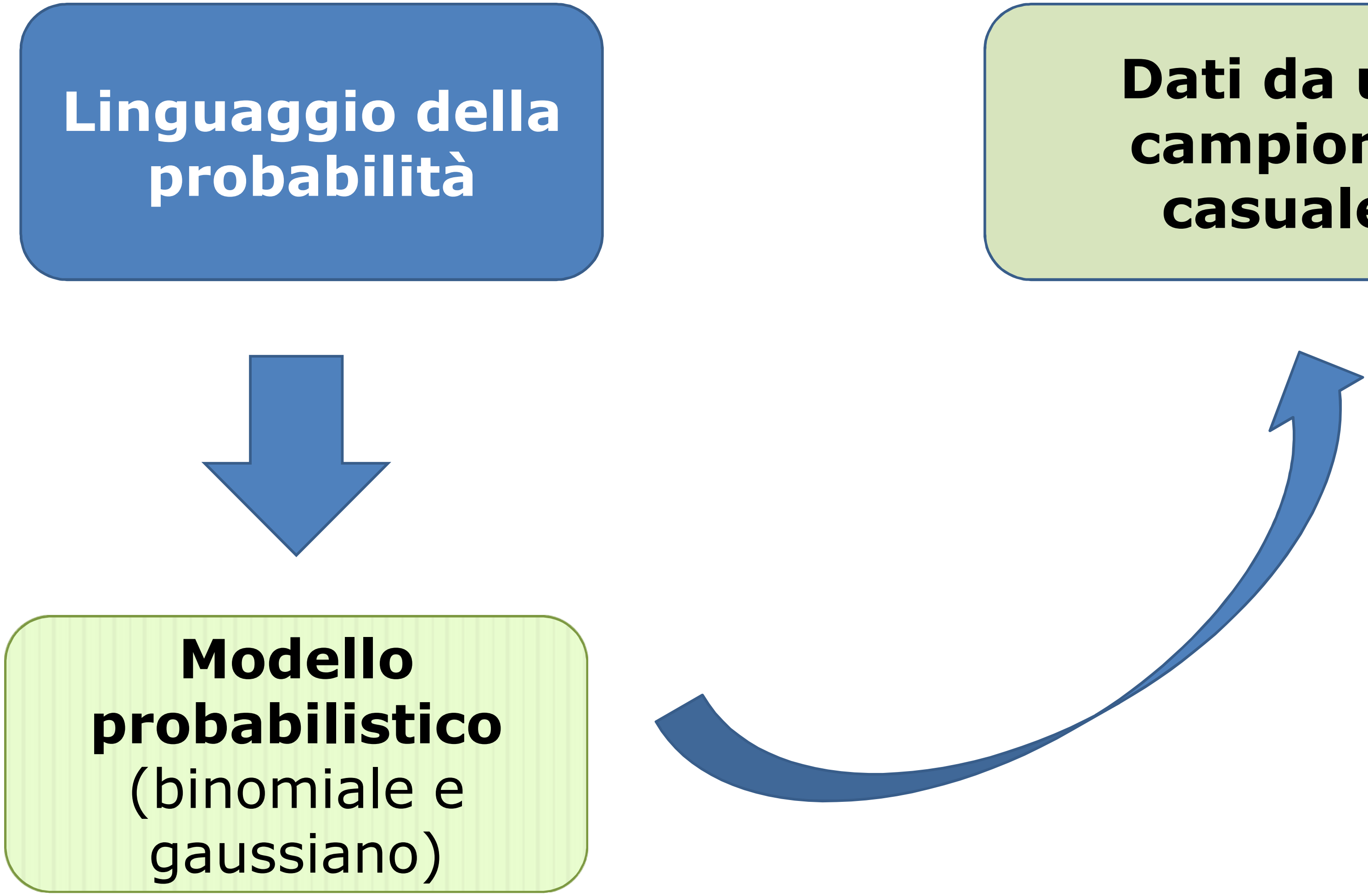
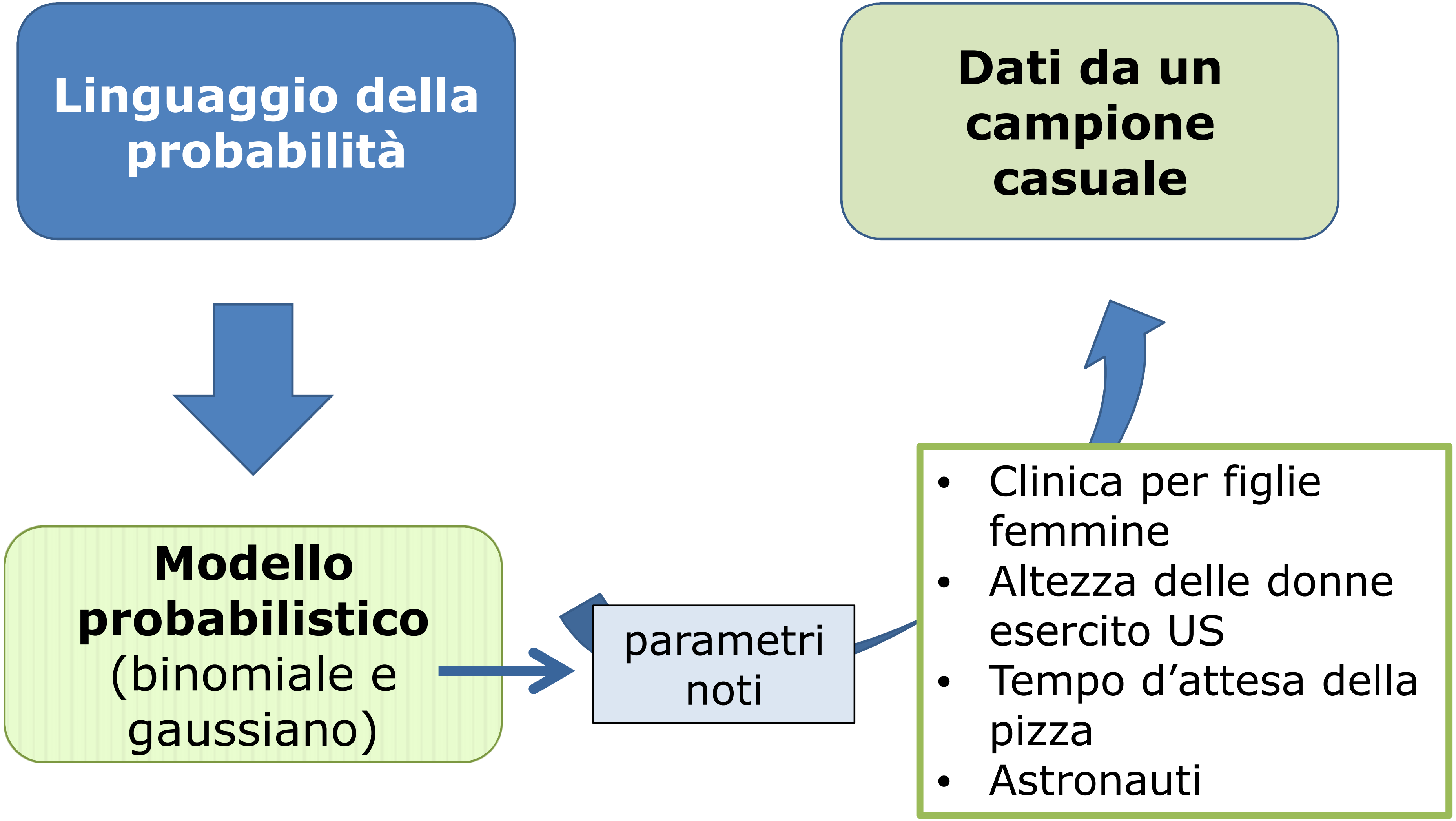
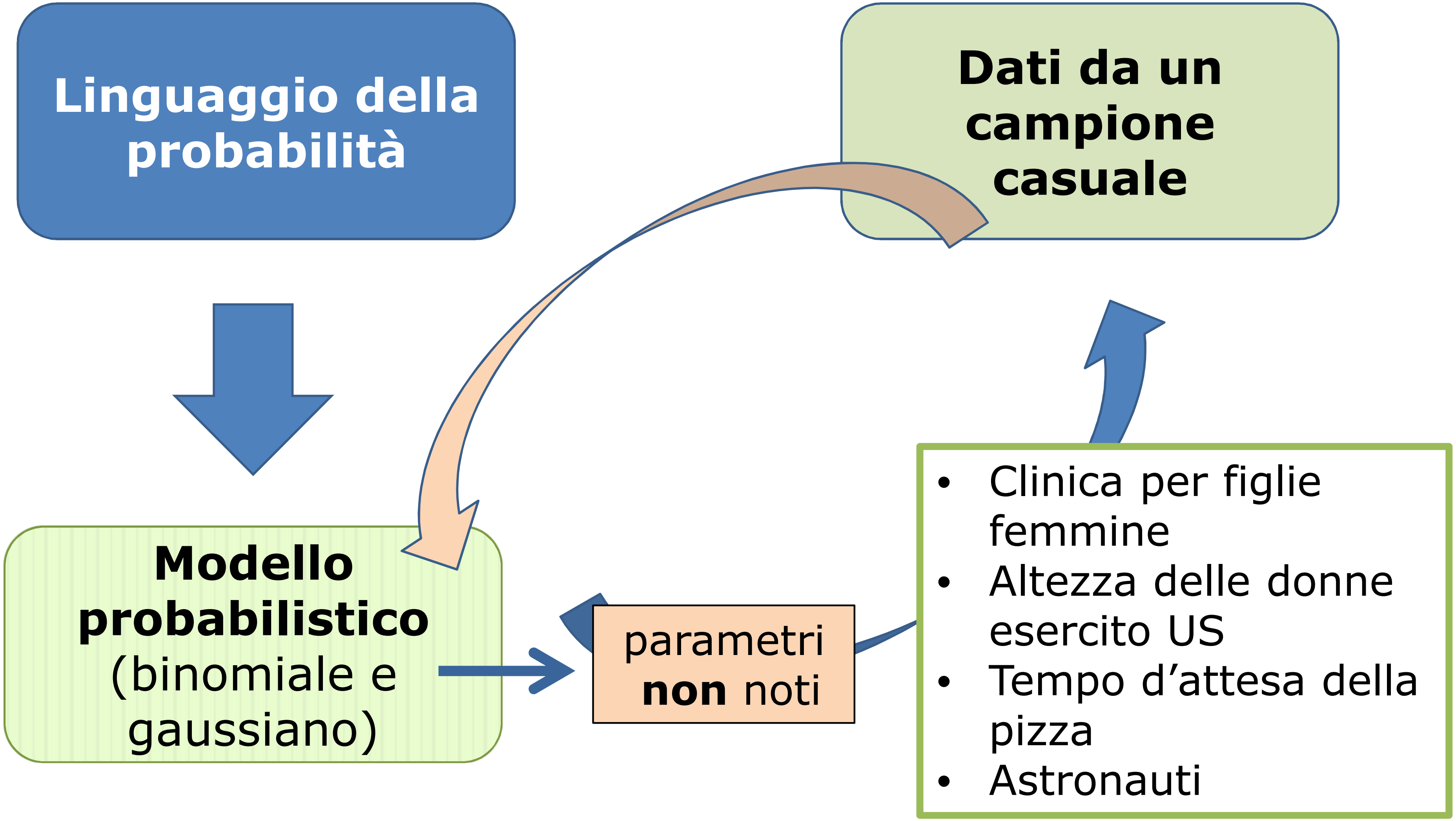
STATISTICA

I modelli probabilistici e l'inferenza


Linguaggio della probabilità
Dati da un campione casuale
Modello probabilistico
(binomiale e gaussiano)


Linguaggio della probabilità
Dati da un campione casuale
Modello probabilistico (binomiale e gaussiano)
parametri noti
• Clinica per figlie femmine
• Altezza delle donne esercito US
• Tempo d'attesa della pizza
• Astronauti


Linguaggio della probabilità
Dati da un campione casuale
Modello probabilistico (binomiale e gaussiano)
parametri **non** noti
Clinica per figlie femmine
Altezza delle donne esercito US
Tempo d’attesa della pizza
Astronauti

W la gaussiana!

$$X \sim N(\mu, \sigma^2) \Longrightarrow \qquad X' = aX + b \sim \boldsymbol{N}(a\mu + b, a^2\sigma^2)$$

$$X_1 \sim N(\mu_1, {\sigma_1}^2)\ e\ \ X_2 \sim N(\mu_2, {\sigma_2}^2)\ \boldsymbol{indipendenti} \ \Longrightarrow$$

$$X_1 + X_2 \sim \boldsymbol{N}(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2)$$

W la gaussiana!

$$X \sim N(\mu, \sigma^2) \Longrightarrow \qquad X' = aX + b \sim \boldsymbol{N}(a\mu + b, a^2\sigma^2)$$

$$X_1 \sim N(\mu_1, {\sigma_1}^2)\ e\ \ X_2 \sim N(\mu_2, {\sigma_2}^2)\ \boldsymbol{indipendenti} \ \Longrightarrow$$

$$X_1 + X_2 \sim \boldsymbol{N}(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2)$$

$$X_1 \sim N(\mu, \sigma^2)\ e\ \ X_2 \sim N(\mu, \sigma^2)\ \boldsymbol{indipendenti}\ (\boldsymbol{e\ identiche}) \Longrightarrow$$

$$X_1 + X_2 \sim \boldsymbol{N}(2\mu, 2\sigma^2)$$

$$\boxed{\frac{X_1 + X_2}{2} \sim \boldsymbol{N}\left(\mu, \frac{\sigma^2}{2}\right)}$$

W la gaussiana!

$$X \sim N(\mu, \sigma^2) \Longrightarrow \qquad X' = aX + b \sim \boldsymbol{N}(a\mu + b, a^2\sigma^2)$$

$$X_1 \sim N(\mu_1, {\sigma_1}^2) \ e \ \ X_2 \sim N(\mu_2, {\sigma_2}^2) \ \boldsymbol{indipendenti} \ \Longrightarrow$$

$$X_1 + X_2 \sim \boldsymbol{N}(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2)$$

$$X_1, \ldots, X_n \ \ \boldsymbol{indipendenti\ e\ identiche}, X_i \sim N(\mu, \sigma^2) \Longrightarrow$$

$$X_1 + X_2 + \cdots + X_n \sim \boldsymbol{N}(n\mu, n\sigma^2)$$

$$\boxed{\bar{\boldsymbol{X}}_{\boldsymbol{n}} = \frac{\boldsymbol{X_1} + \boldsymbol{X_2} + \cdots + \boldsymbol{X_n}}{\boldsymbol{n}} \sim \boldsymbol{N}\left(\boldsymbol{\mu}, \frac{\boldsymbol{\sigma^2}}{\boldsymbol{n}}\right)}$$

W la gaussiana!

$$X \sim N(\mu, \sigma^2) \Longrightarrow \qquad X' = aX + b \sim \boldsymbol{N}(a\mu + b, a^2\sigma^2)$$

$$X_1 \sim N(\mu_1, {\sigma_1}^2) \ e \ \ X_2 \sim N(\mu_2, \ldots \Longrightarrow$$

$$X_1 + X \ldots)$$

campione casuale

$\boldsymbol{X_1, \ldots, X_n}$ ***indipendenti e identiche*** $X_i \sim N(\mu, \sigma^2) \Longrightarrow$

$$X_1 + X_2 + \cdots + X_n \sim \boldsymbol{N}(n\mu, n\sigma^2)$$

$$\boldsymbol{\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$

un inciso importante

X_1 e X_2 $\boldsymbol{indipendenti}$ se $\{P(A \cap B) = P(A)P(B)\}$

$\forall\, a < b, c < d$ si ha che

$P(a < X_1 \leq b$ e $c < X_2 \leq d) = P(a < X_1 \leq b) \times P(c < X_2 \leq d)$

Esempio all’arsenico, II

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\mathrm{g}/\mathrm{L}$.

Sapendo che la legge impone una soglia massima per l’arsenico di 10 $\mu\mathrm{g}/\mathrm{L}$, si può affermare che l'acqua potabile di Milano non è a norma di legge?

Esempio all’arsenico, II

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\text{g/L}$.

Sapendo che la legge impone una soglia massima per l’arsenico di 10 $\mu\text{g/L}$, si può affermare che l'acqua potabile di Milano non è a norma di legge?

$$\bar{x} = \frac{15 + 12.5 + 7.0 + 13.0 + 7.5}{5} = 11\ \mu\text{g/L} > 10\ \mu\text{g/L}$$

è un risultato attendibile per **tutta** l’acqua di Milano?

Esempio all'arsenico, II

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\mathrm{g/L}$.

Sapendo che la legge impone una soglia massima per l'arsenico di 10 $\mu\mathrm{g/L}$, si può affermare che l'acqua potabile di Milano non è a norma di legge?

Sia X la variabile che descrive la quantità di arsenico in un campione di acqua milanese. Supp. $\boldsymbol{X \sim N(\mu, \sigma^2)}$ con $\boldsymbol{\mu}$ **incognita** e $\sigma = 1.6\ \mu\mathrm{g/L}$ (semplificazione momentanea)

Esempio all'arsenico, II

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\mathrm{g/L}$.

Sapendo che la legge impone una soglia massima per l'arsenico di 10 $\mu\mathrm{g/L}$, si può affermare che l'acqua potabile di Milano non è a norma di legge?

Sia X la variabile che descrive la quantità di arsenico in un campione di acqua milanese. Supp. $\boldsymbol{X \sim N(\mu, \sigma^2)}$ con $\boldsymbol{\mu}$ **incognita** e $\sigma = 1.6\ \mu\mathrm{g/L}$ (semplificazione momentanea)

1. Usiamo $\bar{x} = \mathbf{11}$ **come stima di** $\boldsymbol{\mu}$ ⇨ $X \sim N(11, 1.6^2)$

Esempio all'arsenico, II

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu g/L$.

Sapendo che la legge impone una soglia massima per l'arsenico di 10 $\mu g/L$, si può affermare che l'acqua potabile di Milano non è a norma di legge?

Sia X la variabile che descrive la quantità di arsenico in un campione di acqua milanese. Supp. $\boldsymbol{X \sim N(\mu, \sigma^2)}$ con $\boldsymbol{\mu}$ **incognita** e $\sigma = 1.6$ $\mu g/L$ (semplificazione momentanea)

1. Usiamo $\bar{x} = \mathbf{11}$ **come stima di** μ $\Rightarrow$ $X \sim N(11, 1.6^2)$
2. $X_1, X_2, \cdots, X_5$ variabili ***i.i.d,*** $\sim X$ $\Rightarrow$ $\boldsymbol{\bar{X}_5 \sim N(11, 1.6^2/5)}$

Esempio all'arsenico, II

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\mathrm{g/L}$.

Sapendo che la legge impone una soglia massima per l'arsenico di 10 $\mu\mathrm{g/L}$, si può affermare che l'acqua potabile di Milano non è a norma di legge?

Sia X la variabile che descrive la quantità di arsenico in un campione di acqua milanese. Supp. $\boldsymbol{X \sim N(\mu, \sigma^2)}$ con $\boldsymbol{\mu}$ **incognita** e $\sigma = 1.6\ \mu\mathrm{g/L}$ (semplificazione momentanea)

1. Usiamo $\bar{x} = 11$ **come stima di** μ ⇒ $X \sim N(11, 1.6^2)$
2. $X_1, X_2, \cdots, X_5$ variabili ***i.i.d.*** $\sim X \Rightarrow \bar{X}_5 \sim \boldsymbol{N(11, 1.6^2/5)}$
3. $P(\bar{X}_5 > 10) = P\left(\frac{\bar{X}_5 - 11}{1.6/\sqrt{5}} > \frac{10-11}{1.6/\sqrt{5}}\right) = P(Z > -1.40) = 1 - \Phi(-1.40) =$
$= \mathbf{0.92}$

Esempio all'arsenico, II

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la **quantità presente nell'acqua di arsenico**, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 μg/L.

Sapendo che la legge impone una soglia massima per l'arsenico di 10 μg/L, si può affermare che l'acqua potabile di Milano non è a norma di legge?

Sia X la variabile che descrive la quantità di arsenico in un campione di acqua milanese. Supp. $\boldsymbol{X \sim N(\mu, \sigma^2)}$ con $\boldsymbol{\mu}$ **incognita** e $\sigma = 1.6$ μg/L (semplificazione momentanea)

1. Usiamo $\bar{x} = 11$ **come**
2. $X_1, X_2, \cdots, X_5$ variabili ***i.i***
3. $P(\bar{X}_5 > 10) = P\left(\frac{\bar{X}_5 - 11}{1.6/\sqrt{5}} > \frac{10-11}{1.6/\sqrt{5}}\right) = P(Z > -1.40) = 1 - \Phi(-1.40) =$
$= \mathbf{0.92}$

Se la stima $\bar{x}$ di $\boldsymbol{\mu}$ è corretta, il 92% dei campioni indica un contenuto medio di As fuori legge

La media campionaria

Siano $X_1, X_2, \cdots, X_n$ variabili casuali ***i.i.d***

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

media campionaria: v.c. che *predice* il valore della media aritmetica dei dati nel campione

La media campionaria

Siano $X_1, X_2, \cdots, X_n$ variabili casuali ***i.i.d***

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

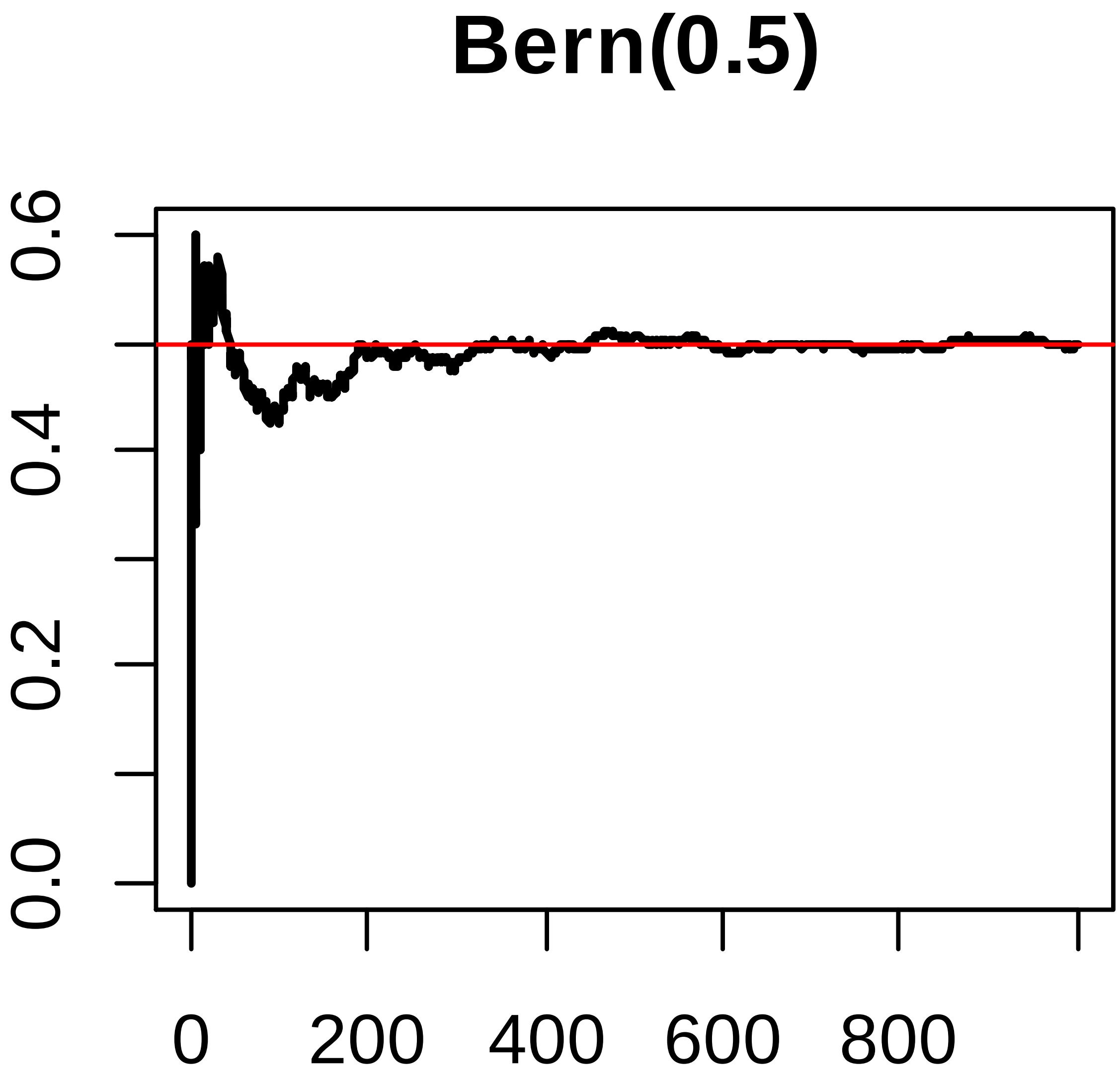
media campionaria: v.c. che *predice* il valore della media aritmetica dei dati nel campione
E' uno *stimatore* della media $E(X_i)$

$X_1, \dots, X_n, \dots$ successione di v.a. **i.i.d** con media $E(X_i) = \mu$ finita.

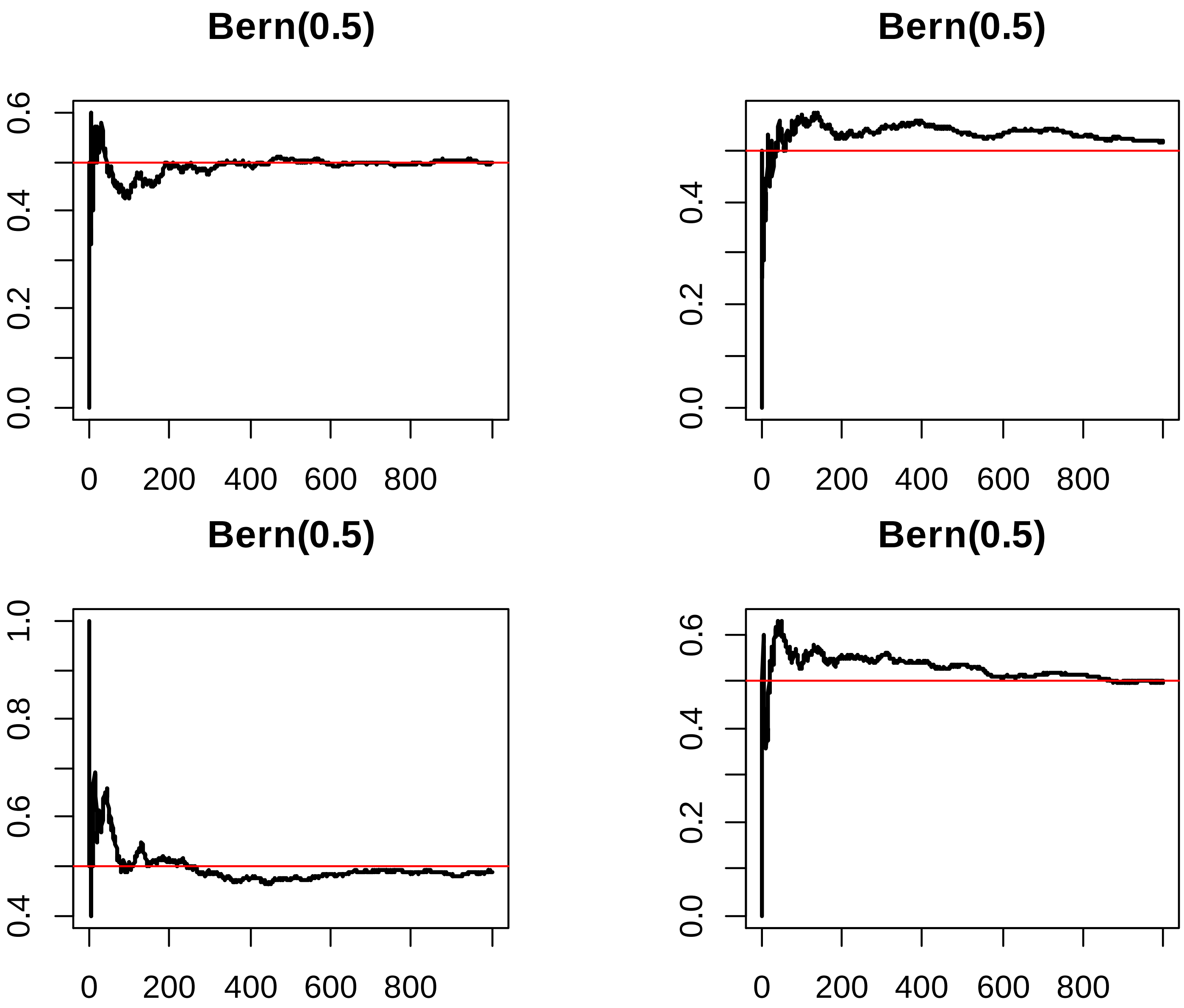
$$\bar{x}_n = \frac{x_1 + x_2 + \cdots + x_n}{n} \xrightarrow[n \to \infty]{} \mu \quad \textit{quasi certamente}$$

Legge dei grandi numeri

Legge (forte) dei grandi numeri



Legge (forte) dei grandi numeri



La media campionaria

Siano $X_1, X_2, \cdots, X_n$ variabili casuali ***i.i.d***

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

media campionaria

Se $X_i \sim N(\mu, \sigma^2)$ $\Rightarrow$ $\bar{X}_n \sim \boldsymbol{N}\left(\boldsymbol{\mu}, \frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}}\right)$ $\bar{X}_n$ stimatore di $\boldsymbol{\mu}$

Se $X_i \sim b(p)$ $\Rightarrow$ $\bar{X}_n \sim \frac{1}{n} Bin(n, p)$ $\bar{X}_n$ stimatore di $\boldsymbol{p}$

La media campionaria

Siano $X_1, X_2, \cdots, X_n$ variabili casuali ***i.i.d***

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

media campionaria

Se $X_i \sim N(\mu, \sigma^2)$ $\Rightarrow$ $\bar{X}_n \sim \boldsymbol{N}\left(\boldsymbol{\mu}, \frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}}\right)$

Se $X_i \sim b(p)$ $\Rightarrow$ $\bar{X}_n \sim \frac{1}{n} Bin(n, p)$

$$\boxed{\textbf{Se } \boldsymbol{np \geq 5 \;\&\; np(1-p) \geq 5} \Rightarrow \bar{X}_n \sim \boldsymbol{N}\left(\boldsymbol{p}, \frac{\boldsymbol{p(1-p)}}{\boldsymbol{n}}\right)}$$

La media campionaria

Siano $X_1, X_2, \cdots, X_n$ variabili casuali ***i.i.d***

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

media campionaria

Se $X_i \sim N(\mu, \sigma^2)$ $\Rightarrow$ $\bar{X}_n \sim \boldsymbol{N}\left(\boldsymbol{\mu}, \frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}}\right)$

Se $X_i \sim b(p)$ $\Rightarrow$ $\bar{X}_n \sim \frac{1}{n} Bin(n, p)$

Teo. Centrale del Limite

Se $X_i \sim ???$ & $n > 30$ $\Rightarrow$ $\bar{X}_n \sim \boldsymbol{N}\left(\boldsymbol{E}(\boldsymbol{X}_1), \frac{\boldsymbol{Var}(\boldsymbol{X}_1)}{\boldsymbol{n}}\right)$

Binomiale e Gaussiana

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. Di **140** coppie su cui è stata messa in atto, **90** hanno avuto una femmina.

X variabile casuale che dice il sesso del neonato: 1=F, 0=M

$$P(X=1)=p, \qquad P(X=0)=q=1-p, \qquad con\ qualche\ tecnica$$

X è una variabile bernoulliana di *parametro* $p \in (0,1)$

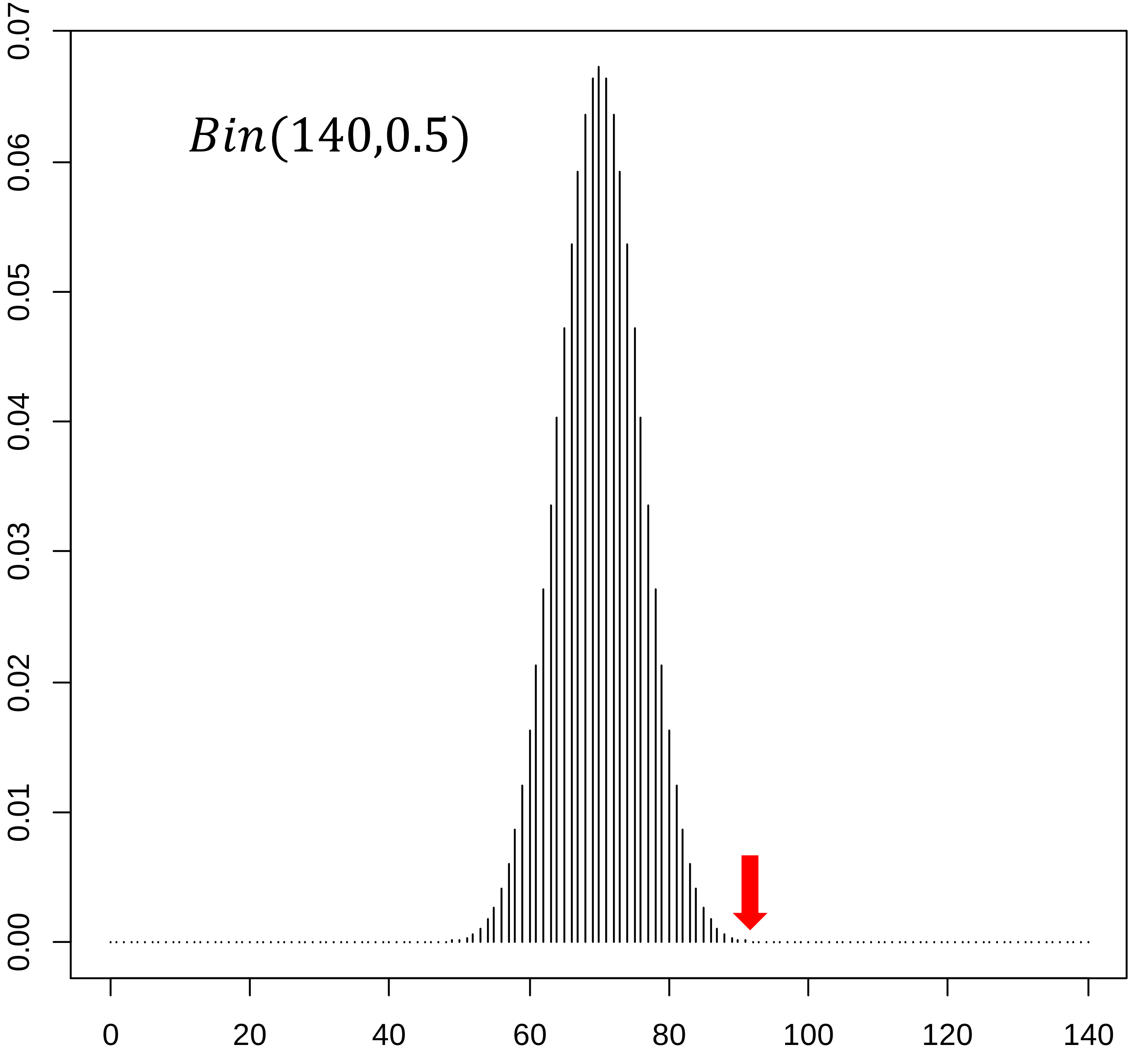
$X_1 + X_2 + \ldots + X_{140}$ conta il numero di F nei 140 parti

$S_n = X_1 + X_2 + \ldots + X_n$ è una variabile **Binomiale**: $Bin(n,p)$

Se

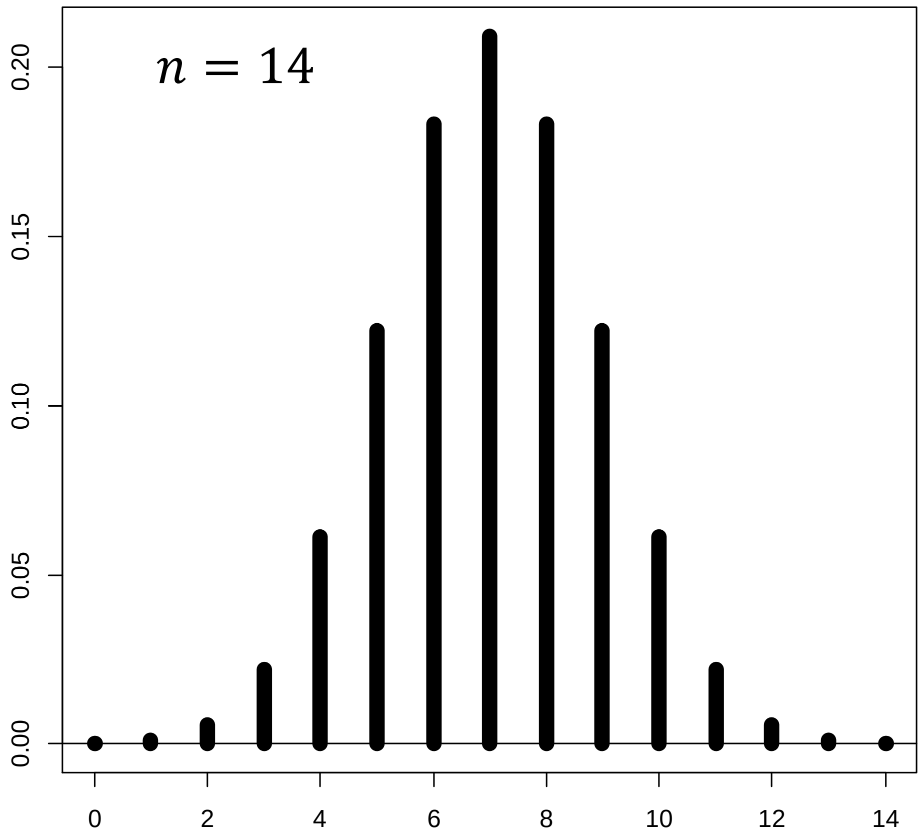
$$p = 0.5, n = 140 \Rightarrow P(S_{140} = 90) = \frac{140!}{90!(140-90)!} 0.5^{90}(1-0.5)^{140-90} = 0.000213747$$

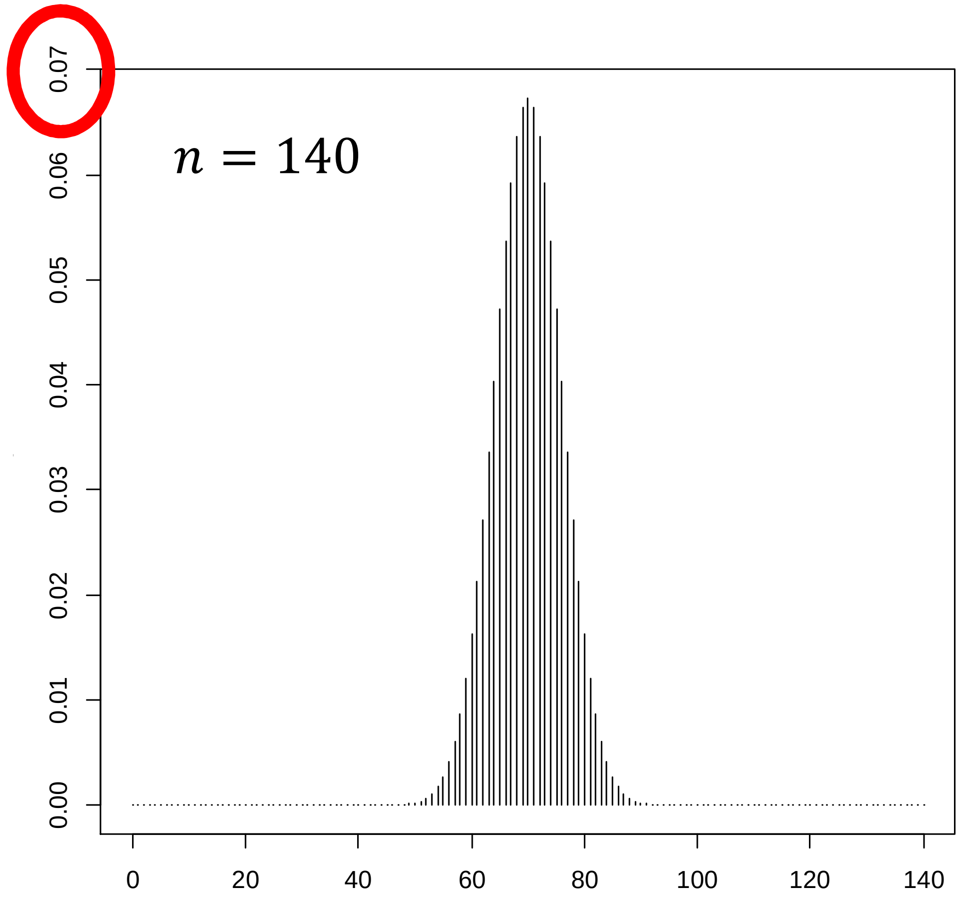
Binomiale e Gaussiana



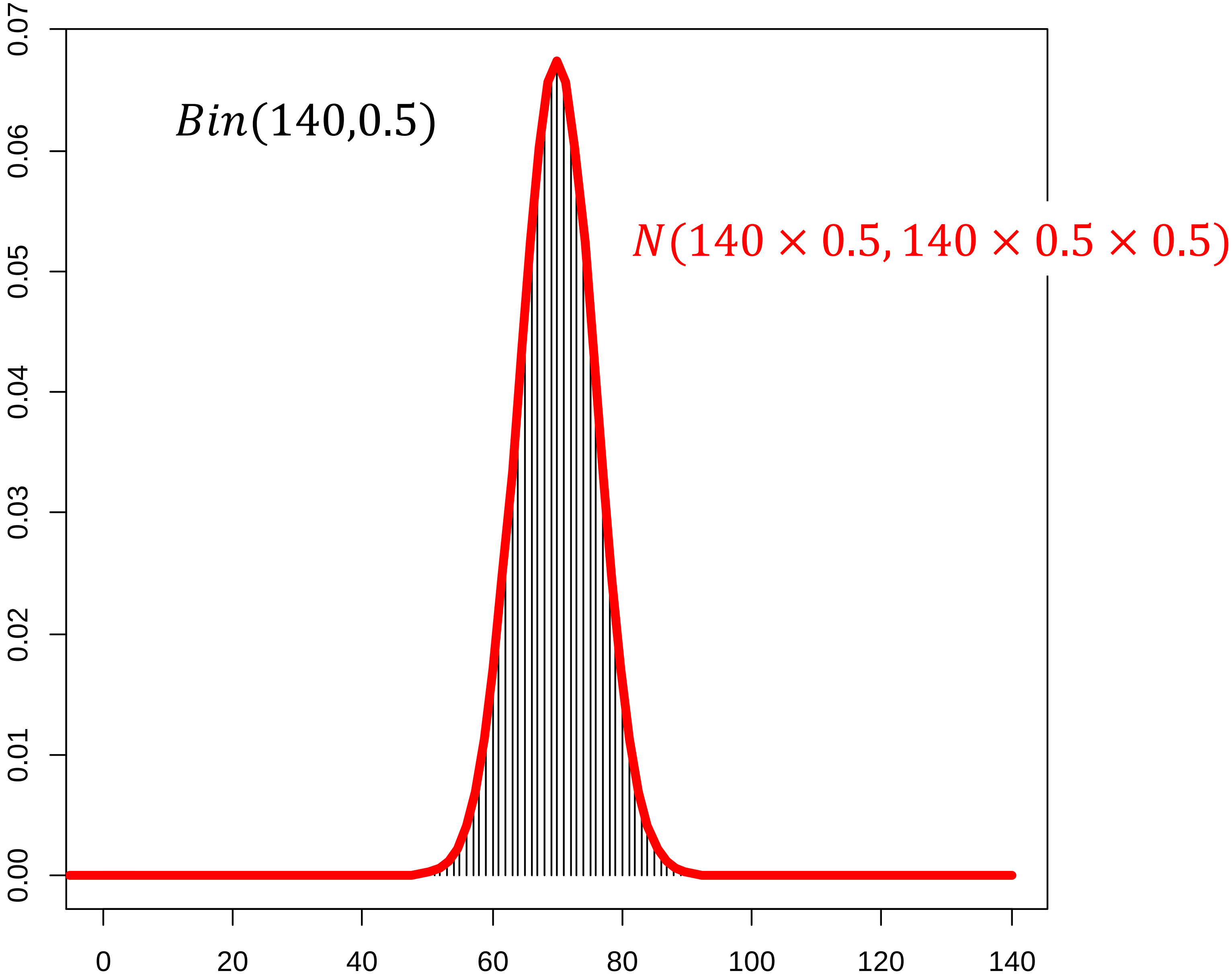
un promemoria

Quando n è grande, sono moltissimi i valori k tali che $P(X = k) = 0$, perché c'è il vincolo della somma=1! Per questo è meglio guardare $X = k\ o\ più$ piuttosto che solo $X = k$.





Binomiale e Gaussiana



Binomiale e Gaussiana

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. Di **140** coppie su cui è stata messa in atto, **90** hanno avuto una femmina.

Se $p = 0.5, n = 140 \Rightarrow \boldsymbol{P(S_{140} \geq 90)} = P\left(\frac{S_{140}}{140} \geq \frac{90}{140}\right) = P(\bar{X}_{140} \geq 0.64)$

$$\bar{X}_n \sim \boldsymbol{N}\left(\mathbf{0.5}, \frac{\mathbf{0.5(1-0.5)}}{\mathbf{140}}\right) \Rightarrow \approx P\left(Z \geq \frac{0.64 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{140}}}\right) = P(Z \geq 3.31) =$$

$$= 1 - P(Z < 3.31) = 1 - 0.9995 = 0.0005$$

La tecnica sembra funzionare: in condizioni "naturali" avere 90 o più F in 140 parti ha una probabilità bassissima.

Binomiale e Gaussiana

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. Di **140** coppie su cui è stata messa in atto, **90** hanno avuto una femmina.

Se $p = 0.5, n = 140 \Rightarrow P(S_{140} \geq 90) = P\left(\frac{S_{140}}{140} \geq \frac{90}{140}\right) = P(\bar{X}_{140} \geq 0.64)$

$$\bar{X}_n \sim \boldsymbol{N}\left(\mathbf{0.5}, \frac{\mathbf{0.5(1-0.5)}}{\mathbf{140}}\right) \Rightarrow \approx P\left(Z \geq \frac{0.64 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{140}}}\right) = P(Z \geq 3.31) =$$

$$= 1 - P(Z < 3.31) = 1 - 0.9995 = 0.0005$$

La tecnica sembra funzionare: in condizioni “naturali” avere 90 o più F in 140 parti ha una probabilità bassissima.

$$\hat{p}_n = \frac{90}{140} = 0.643$$

Media campionaria come stimatore

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la quantità presente nell'acqua di arsenico, ottenendo i seguenti valori campionari: 15, 12.5, 7.0, 13.0, 7.5 $\mu g/L$.

Abbiamo usato $\bar{x} = 11\ (media\ aritmetica)$ **come stima di** $E(X_1)$

$\bar{X}_n$ ***stimatore*** di $E(X_1)$, $\bar{x}_n$ ***stima*** di $E(X_1)$

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. Di 140 coppie su cui è stata messa in atto, 90 hanno avuto una femmina.

Usiamo $\bar{x} = \frac{90}{140} = 0.64 = 64\%$ (proporzione osservata) **come stima della proporzione** p di F che si ottiene con l'uso della tecnica del IVF.

N.B. $p = E(X_1)$

Media campionaria co
stimatore

Perchè ci serve uno stimatore, e non ci basta la stima?

Durante le verifiche sulla qualità d
misurata la quantità presente nell'a
seguenti valori campionari: 15, 12

Abbiamo usato $\boldsymbol{\bar{x} = 11}$ **(*media aritmeti*** **come** **a di** $\boldsymbol{E(X_1)}$

$\boxed{\bar{X}_n \textbf{ \textit{stimatore} } \text{di } \boldsymbol{E(X_1)} \quad , \quad \bar{x}_n \textbf{ \textit{stima} } \text{di } \boldsymbol{E(X_1)}}$

Il Genetics and IVF Institute di Fairfax in Virginia ha sviluppato una tecnica in grado, sostengono, di aumentare la probabilità per una coppia di avere una figlia femmina. Di 140 coppie su cui è stata messa in atto, 90 hanno avuto una femmina.

Usiamo $\bar{x} = \hat{p}_n = \frac{90}{140} = 0.64 = 64\%$ (proporzione osservata) **come stima della proporzione** $\boldsymbol{p}$ di F che si ottiene con l'uso della tecnica del IVF.

N.B. $\boldsymbol{p = E(X_1)}$

Media campionaria come stimatore

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la quantità presente nell'acqua di arsenico, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\text{g/L}$.

Abbiamo usato $\bar{x} = 11$ $(media\ aritmetica)$ **come stima di** $E(X_1)$

se fossimo andati a raccogliere i campioni in altri punti dell'acquedotto, avremmo avuto altri valori, e una stima diversa

Media campionaria come stimatore

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la quantità presente nell'acqua di arsenico, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\mathrm{g/L}$.

Abbiamo usato $\bar{x} = 11\ (media\ aritmetica)$ **come stima di** $E(X_1)$

se fossimo andati a raccogliere i campioni in altri punti dell'acquedotto, avremmo avuto altri valori, e una stima diversa

$$\bar{X}_5 \sim N\left(8.3, 1.6^2/5\right)$$

$$P(\bar{X}_5 > 10) = P\left(\frac{\bar{X}_5 - 8.3}{1.6/\sqrt{5}} > \frac{10 - 8.3}{1.6/\sqrt{5}}\right) = P(Z > 2.37) = 1 - \Phi(2.37) = \mathbf{0.009!}$$

Media campionaria come stimatore

Durante le verifiche sulla qualità dell'acqua potabile di Milano si è misurata la quantità presente nell'acqua di arsenico, ottenendo i seguenti **valori campionari**: 15, 12.5, 7.0, 13.0, 7.5 $\mu\mathrm{g/L}$.

Abbiamo usato $\bar{x} = 11$ $(media\ aritmetica)$ **come stima di** $E(X_1)$

se fossimo andati a raccogliere i campioni in altri punti dell'acquedotto, avremmo avuto altri valori, e una stima diversa

Per dire qualcosa attorno all'incertezza sulla stima ci serve la variabile casuale *stimatore*!

$\bar{X}_5 \sim \boldsymbol{N}(\mathbf{8.3}, \mathbf{1.6^2/5})$

$P(\bar{X}_5 > 10) = P\left(\frac{\bar{X}_5 - 8.3}{1.6/\sqrt{5}} > \frac{10 - 8.3}{1.6/\sqrt{5}}\right) = P(Z > 2.37) = 1 - \Phi(2.37) = \mathbf{0.009!}$

Inferenza sulla **media di una popolazione**

17/10/2016, su un quotidiano on-line:

Sì: 33.8%
No: 37.0%
Indeciso: 29.2%

«Ed è per questo che, …, "la sfida" del 4 dicembre "si giocherà sul filo dei decimali". Anche perché tutti i sondaggi che vengono pubblicati da giornali e tv in queste settimane hanno come sempre un margine d'errore del 3 per cento.»

Inferenza sulla **media di una popolazione**

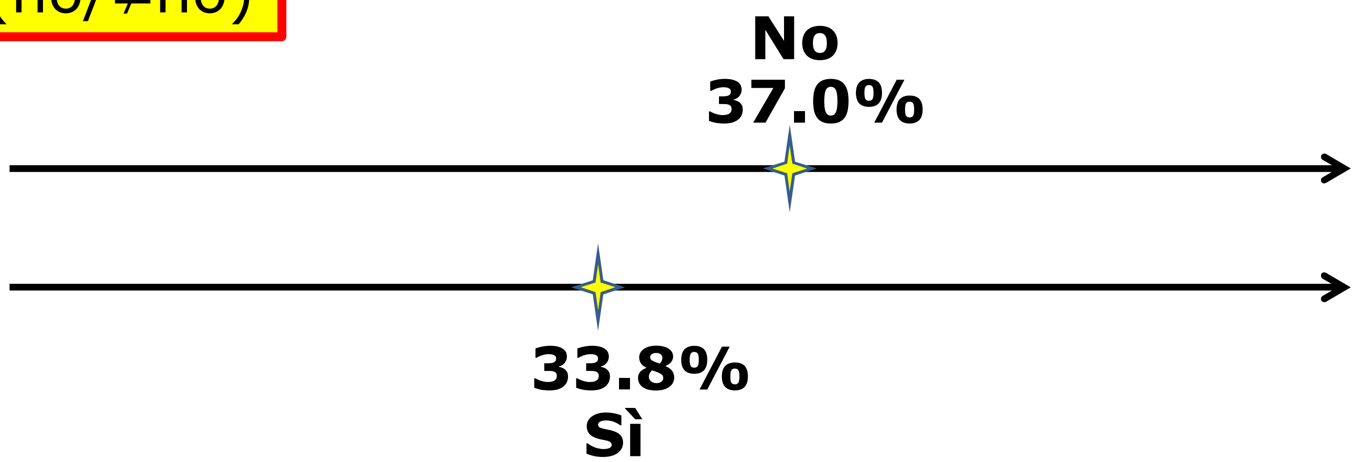
17/10/2016, su un quotidiano on-line:

Sì: $\hat{p}_n$=33.8%
No: $\hat{p}_n$=37.0%
Indeciso: 29.2%

$\boxed{X_i \sim bern(p_S)}$ (sì/≠sì)

$\boxed{Y_i \sim bern(p_N)}$ (no/≠no)

*«Ed è per questo che, …, "la sfida" del 4 dicembre "si giocherà **sul filo dei decimali**". Anche perché tutti i sondaggi che vengono pubblicati da giornali e tv in queste settimane hanno come sempre un **margine d'errore** del **3 per cento**.»*



Inferenza sulla **media di una popolazione**

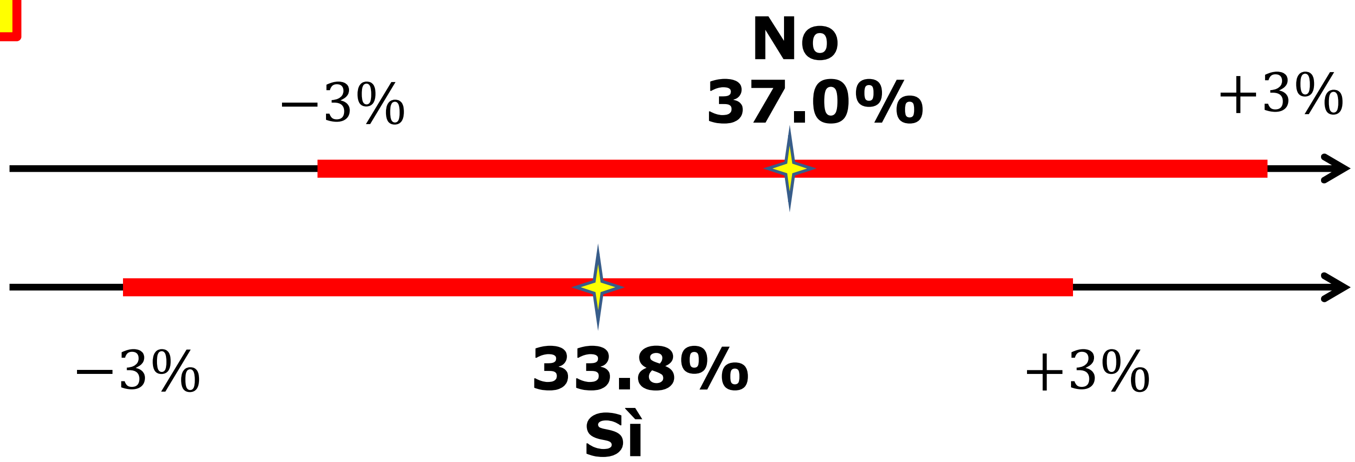
17/10/2016, su un quotidiano on-line:

Sì: $\hat{p}_n = 33.8\%$
No: $\hat{p}_n = 37.0\%$
Indeciso: 29.2%

*«Ed è per questo che, …, "la sfida" del 4 dicembre "si giocherà **sul filo dei decimali**". Anche perché tutti i sondaggi che vengono pubblicati da giornali e tv in queste settimane hanno come sempre un **margine d'errore** del **3 per cento**.»*

$\boxed{X_i \sim bern(p_S)}$

$\boxed{Y_i \sim bern(p_N)}$



Inferenza sulla **media di una popolazione**

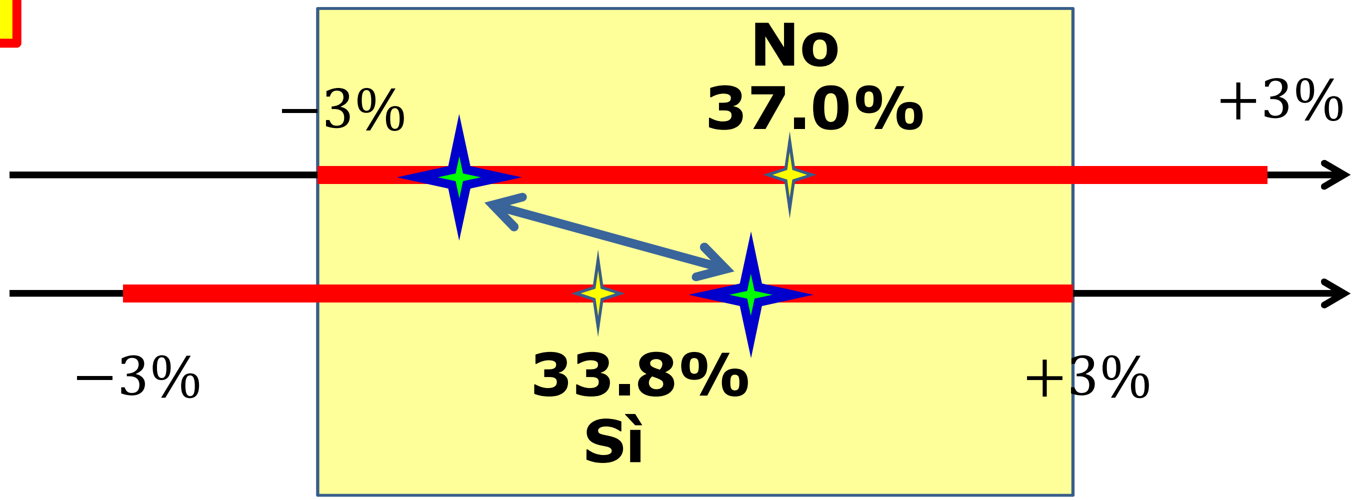
17/10/2016, su un quotidiano on-line:

Sì: $\hat{p}_n$=33.8%
No: $\hat{p}_n$=37.0%
Indeciso: 29.2%

$X_i \sim bern(p_S)$

$Y_i \sim bern(p_N)$

*«Ed è per questo che, …, "la sfida" del 4 dicembre "si giocherà **sul filo dei decimali**". Anche perché tutti i sondaggi che vengono pubblicati da giornali e tv in queste settimane hanno come sempre un **margine d'errore** del **3 per cento**.»*



ARPA LOMBARDIA
Agenzia Regionale per la Protezione dell'Ambiente

Portale ARPA Lombardia | Contatti e PEC | Sito accessibile | Intranet

RegioneLombardia

QUALITA' DELL'ARIA

LE MAPPE | STAZIONI FISSE | STAZIONI MOBILI | LE EMISSIONI

Chi siamo | Rete di rilevamento | Inquinanti | Inventario | Modellistica | Attivita' e progetti | Approfondimenti

Stazione di rilevamento

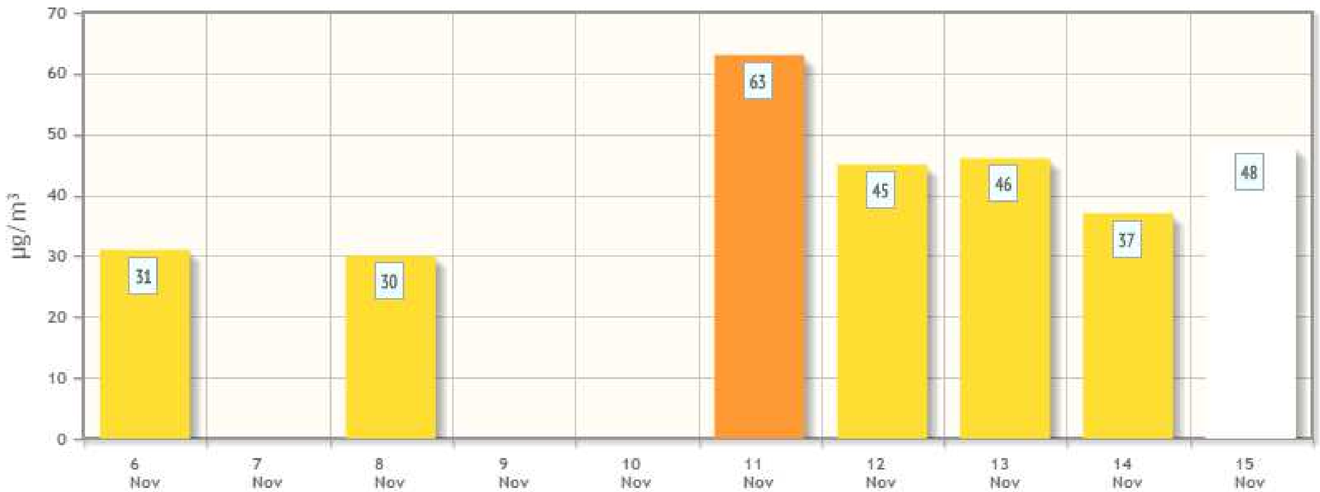
Milano Pascal Città Studi

Stazione	Milano Pascal Città Studi
Località	Milano
Indirizzo	via Ponzio 34/6 – Pascal Città Studi
Inquinanti	Benzene NO2 O3 PM10 PM2.5 SO2

Gli inquinanti monitorati **15 novembre 2016**

I dati riportati sono da considerarsi incerti fino alla loro validazione da parte del competente Centro Regionale per il Monitoraggio della Qualità dell'Aria.
I valori limite di PM2.5 e benzene fanno riferimento ad un periodo medio di valutazione annuale.

PM10	48 μg/m³ media giornaliera	Valore limite 50 μg/m³



Dati rilevati degli ultimi 10 gg.

ARPA LOMBARDIA
Agenzia Regionale per la Protezione dell'Ambiente

Portale ARPA Lombardia | Contatti e PEC | Sito accessibile | Intranet

RegioneLombardia

QUALITA' DELL'ARIA

LE MAPPE | STAZIONI FISSE | STAZIONI MOBILI | LE EMISSIONI

Chi siamo | Rete di rilevament

X_i valore giornaliero di PM10

Stazione di rilevamento

Milano Pascal Città Studi

$$\bar{x}_5 = \frac{63 + 45 + 46 + 37 + 48}{5} = 47.8$$

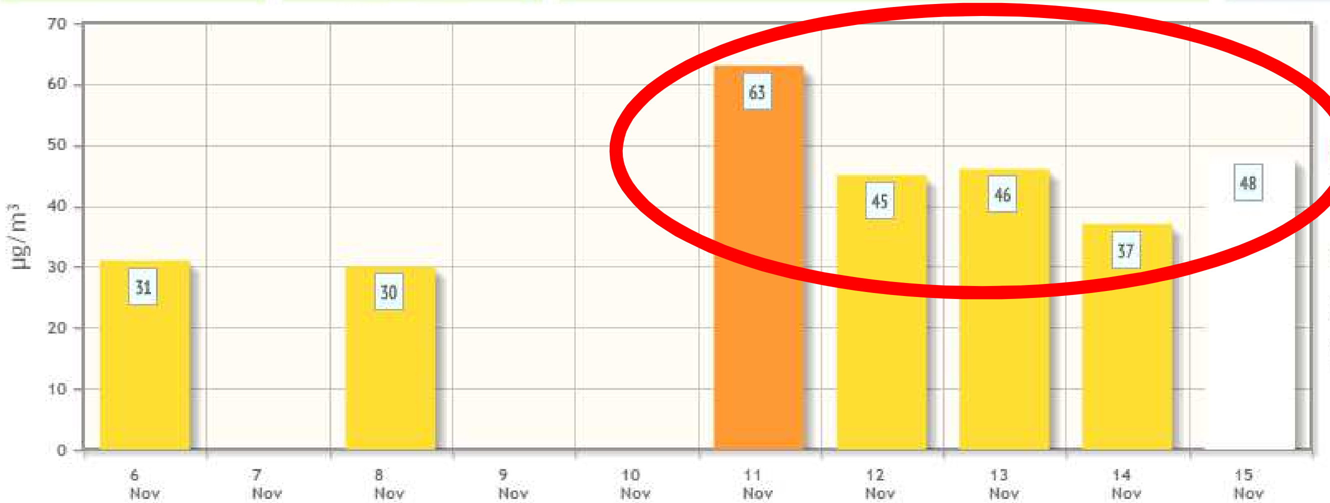
Indirizzo via Ponzio 34/6 – Pascal Città Studi

e il margine di errore?

Gli inquinanti monitorati **15 novembre 2016**

I dati riportati sono da considerarsi incerti fino alla loro validazione da parte del competente Centro Regionale per il Monitoraggio della Qualità dell'Aria.
I valori limite di PM2.5 e benzene fanno riferimento ad un periodo medio di valutazione annuale.

PM10 | 48 µg/m³ media giornaliera | Valore limite 50 µg/m³



Dati rilevati degli ultimi 10 gg.

Inferenza sulla **media di una popolazione**

$\bar{X}_n$ stimatore di μ

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

$X_1, X_2, \ldots, X_n$ i.i.d possibili osservazioni

$X_i \sim N(\mu,\ \sigma^2)$ modello, con $\boldsymbol{\sigma^2}$ **nota**

Intervallo di confidenza di livello $\boldsymbol{1-\alpha}$

Contiene il *vero valore* della media della popolazione, μ, con probabilità $\boldsymbol{1-\alpha}$

$$P\left(\bar{X}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha$$

Inferenza sulla **media di una popolazione**

$$\left(\bar{X}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\ ,\ \bar{X}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right)$$

E' il modello che prevede tutti i possibili intervalli di confidenza di livello $1-\alpha$

100 campioni
100 intervalli di confidenza
di liv. $1-\alpha = 0.95$

95 contengono il vero valore di μ

possiamo avere fiducia che **il "nostro"** intervallo sia uno di questi.

Inferenza sulla **media di una popolazione**

$\bar{X}_n$ stimatore di μ

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

$X_1, X_2, \dots, X_n$ i.i.d possibili osservazioni

$X_i \sim N(\mu, \sigma^2)$ modello, con $\boldsymbol{\sigma^2}$ **nota**

Tavola 1a: Valori critici della Variabile Casuale Normale Standardizzata. $P(Z \geq z_\alpha) = \alpha$.

α	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001	
z_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905	3.7190	
α	0.00009	0.00008	0.00007	0.00006	0.00005	0.00004	0.00003	0.00002	0.00001
z_α	3.7455	3.7750	3.8082	3.8461	3.8906	3.9444	4.0128	4.1075	4.2649

$$P\left(\bar{X}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha$$

Inferenza sulla **media di una popolazione**

$\bar{X}_n$ stimatore di μ

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

$X_1, X_2, \dots, X_n$ i.i.d possibili osservazioni

$X_i \sim N(\mu, \sigma^2)$ modello, con σ^2 **nota**

Tavola 1a: Valori critici della Variabile Casuale Normale Standardizzata. $P(Z \geq z_\alpha) = \alpha$.

α	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001	
z_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905	3.7190	
α	0.00009	0.00008	0.00007	0.00006	0.00005	0.00004	0.0000[illegible]	0.00002	0.00001
z_α	3.7455	3.7750	3.8082	3.8461	3.8906	3.9444	4.0128	4.1075	4.2649

$\alpha = 0.05$
$1 - \alpha = 0.95$

$$P\left(\bar{X}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X}_n + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha$$

$\alpha = 0.001$
$1 - \alpha = 0.999$

$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$z_{\frac{\alpha}{2}} = z_{0.0005} = 3.2905$

Inferenza sulla **media di una popolazione**

$\bar{X}_n$ stimatore di μ

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

$X_1, X_2, \ldots, X_n$ i.i.d possibili osservazioni

$X_i \sim N(\mu,\ \sigma^2)$ modello, con $\boldsymbol{\sigma^2}$ **nota**

Intervallo di confidenza di livello $\boldsymbol{1-\alpha}$

Contiene il *vero valore* della media della popolazione, μ, con probabilità $\boldsymbol{1-\alpha}$

$$P\left(\bar{X}_n - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \le \mu \le \bar{X}_n + \boxed{z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}}\right) = 1-\alpha$$

margine d'errore

Inferenza sulla **media di una popolazione**

$(\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}) : \{\mathbf{34, 19, 6, 12, 24, 12, 20, 27, 33, 29}\}$ campione casuale

$\bar{X}_n$ stimatore di μ

$X_1, X_2, \dots, X_n$ i.i.d possibili osservazioni

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

$X_i \sim N(\mu, \sigma^2)$ modello, con $\boldsymbol{\sigma^2}$ **nota**

$$P\left(\bar{\boldsymbol{X}}_{\boldsymbol{n}} - \boldsymbol{z}_{\frac{\boldsymbol{\alpha}}{\boldsymbol{2}}}\sqrt{\frac{\boldsymbol{\sigma^2}}{\boldsymbol{n}}} \le \mu \le \bar{\boldsymbol{X}}_{\boldsymbol{n}} + \boldsymbol{z}_{\frac{\boldsymbol{\alpha}}{\boldsymbol{2}}}\sqrt{\frac{\boldsymbol{\sigma^2}}{\boldsymbol{n}}}\right) = 1 - \alpha$$

$n = 10$

$\sigma^2 = 100\ (\mu g/m^3)^2$

$\alpha = 0.05,\ z_{\frac{0.05}{2}} = 1.96$

$\bar{\boldsymbol{x}} = \mathbf{21.6}\ \boldsymbol{\mu g/m^3}$

$$\left(\mathbf{21.6 - 1.96}\sqrt{\frac{\mathbf{100}}{\mathbf{10}}}, \mathbf{21.6 + 1.96}\sqrt{\frac{\mathbf{100}}{\mathbf{10}}}\right)$$

$(\mathbf{15.40, 27.80})\ \boldsymbol{\mu g/m^3}$

Inferenza sulla **media di una popolazione**

$(\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}) : \{\mathbf{34}, \mathbf{19}, \mathbf{6}, \mathbf{12}, \mathbf{24}, \mathbf{12}, \mathbf{20}, \mathbf{27}, \mathbf{33}, \mathbf{29}\}$ campione casuale

$\bar{X}_n$ stimatore di μ

$X_1, X_2, \dots, X_n$ i.i.d possibili osservazioni

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

$X_i \sim N(\mu, \sigma^2)$ modello, con $\boldsymbol{\sigma^2}$ **nota**

$$P\left(\bar{\boldsymbol{X}}_{\boldsymbol{n}} - \boldsymbol{z}_{\frac{\boldsymbol{\alpha}}{\mathbf{2}}}\sqrt{\frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}}} \le \mu \le \bar{\boldsymbol{X}}_{\boldsymbol{n}} + \boldsymbol{z}_{\frac{\boldsymbol{\alpha}}{\mathbf{2}}}\sqrt{\frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}}}\right) = 1 - \alpha$$

$n = 10$
$\sigma^2 = 100\ (\mu g/m^3)^2$
$\alpha = 0.05,\ z_{\frac{0.05}{2}} = 1.96$
$\bar{x} = 21.6, \mu g/m^3$

$$\left(\mathbf{21.6 - 1.96}\sqrt{\frac{\mathbf{100}}{\mathbf{10}}}, \mathbf{21.6} + \boxed{\mathbf{1.96}\sqrt{\frac{\mathbf{100}}{\mathbf{10}}}}\right)$$

$6.2\ \mu g/m^3$ è il margine d’errore