Metodi statistici per lo studio dei fenomeni biologici

Alla fine di questa lezione dovreste essere in grado di:

- spiegare il concetto di distribuzione di probabilità
- descrivere la distribuzione di probabilità binomiale
- utilizzare la distribuzione di probabilità binomiale
- spiegare l'approssimazione gaussiana alla binomiale

Caso e probabilità (1)

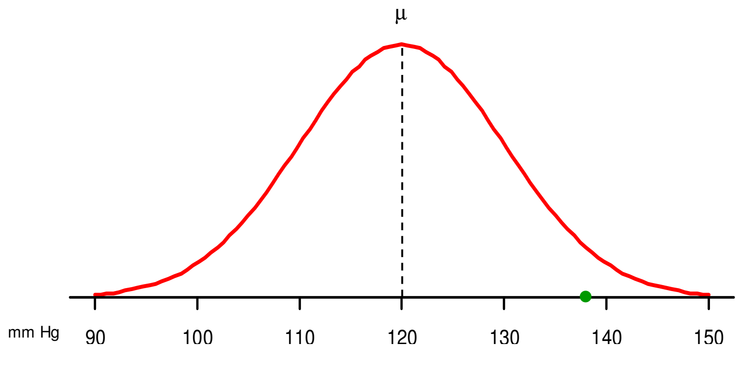
- Tutti i fenomeni che sono influenzati da molte cause, ciascuna di entità modesta, imprevedibili e spesso ignote
- Quando si parla di caso si intende parlare di eventi probabilisticamente indipendenti.
- Il risultato 'casuale' è il frutto della concomitanza accidentale di due fenomeni, la cui probabilità può essere calcolata a partire dalle probabilità individuali.
- Se la singola osservazione è imprevedibile, l'azione del caso su insiemi di osservazioni può essere anticipata con regole precise. Come?

Caso e probabilità (2)

- Possiamo identificare un modello (probabilistico) del fenomeno, che permetta di predire la frequenza con cui, se è solo il caso ad agire, la variabile in studio assumerà tutti i suoi possibili valori

→ **Distribuzione di probabilità**

- Questo insieme di valori, calcolati in condizioni di indipendenza, viene utilizzato come 'riferimento, per valutare (probabilisticamente) se due eventi sono indipendenti



Come interpretare un risultato?

Interpretiamo il singolo risultato 'confrontandolo' con la distribuzione di probabilità della variabile, da cui la misura osservata verosimilmente deriva.

E' importante definire che relazione probabilistica esiste fra campione e popolazione

Distribuzioni di probabilità

- ✓ **Discreta**: Specifica tutti i possibili risultati della variabile casuale e la loro probabilità
- ✓ Esempi: **Binomiale, Poisson**
- ✓ **Continua**: Associa una probabilità a un possibile range di risultati
- ✓ Esempi: **Gaussiana, *t* di Student, chi-quadrato**

Esempio

Secondo la ditta produttrice, un nuovo farmaco antidolorifico ha una probabilità di 0,70 di far passare il mal di testa

Cioè su 100 persone trattate ci si aspetta che ne guariscano 70 anche se non sappiamo quali

Abbiamo trattati 7 malati con il nuovo farmaco, ma abbiamo osservato un miglioramento in 3 casi. Come interpretiamo questo risultato? Il farmaco non è efficace come si dice?

Qual è il contesto di valutazione?

- Ciascuna prova può avere come risultato uno dei due possibili eventi (successo - insuccesso)

- La probabilità di un successo, **p**, rimane costante in ogni prova (0,70). La probabilità dell'insuccesso è **1-p** (0,30)

- Le prove sono indipendenti, ovvero il risultato di una prova non è influenzato dal risultato di qualsiasi prova (conoscere l'effetto di in farmaco su un paziente non fornisce informazioni sull'effetto dello stesso farmaco sul paziente successivo)

Qual è il contesto di valutazione?

- Vogliamo calcolare la probabilità di osservare 3 guarito su 7 trattati

- Una sequenza possibile è G-NG-NG-G-G-NG-NG

- La cui probabilità è
  - eventi indipendenti
  - p = 0,70
  - 1-p = 0,30

$p*(1-p)*(1-p)*p*p*(1-p)*(1-p) =$

$p^3*(1-p)^4 =$

$0.7^3*0.3^4 = 0.00278$

MA

Qual è il contesto di valutazione?

- Ci sono 35 sequenze possibili con 3 guariti su 7

G-G-G-NG-NG-NG-NG
G-G-NG-G-NG-NG-NG
G-G-NG-NG-G-NG-NG
……………………………………
G-NG-NG-G-G-NG-NG
……………………………………
NG-NG-NG-NG-G-G-G

- Ognuna ha probabilità $p^3*(1-p)^4$

- Quindi la probabilità totale di 3 guariti su 7 è $35 * p^3 * (1-p)^4$

Come possiamo calcolare quante sono le possibili sequenze (combinazioni) di un generico numero x di successi in n prove ?

$$ {}_nC_x = \binom{n}{x} = \frac{n!}{x!(n-x)!} $$

Coefficiente binomiale

n ! = prodotto di tutti i numeri interi da n fino ad 1 [0! = 1]

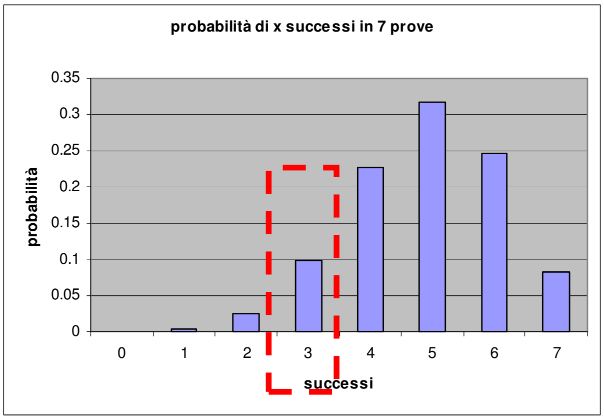
$$ {}_7C_3 = \frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{(3\cdot2\cdot1)\cdot(4\cdot3\cdot2\cdot1)} = 35 $$

Distribuzione binomiale

Ci dice la probabilità di osservare *x* volte un dato risultato in *n* tentativi

P (X = x) =

$$\binom{n}{x} p^x (1-p)^{n-x}$$



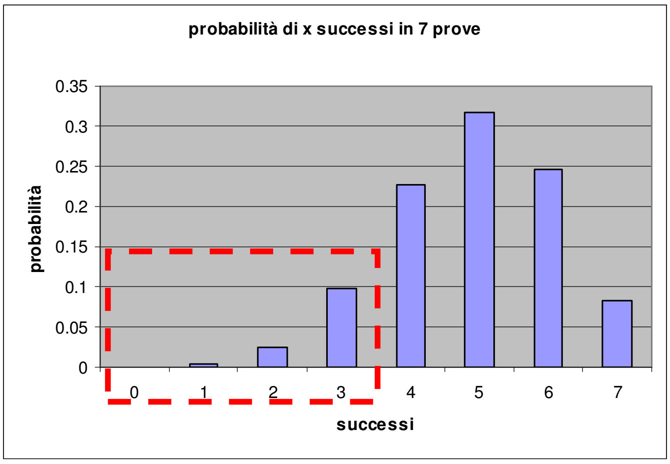
Soluzione

P(x = 3 successi in 7 prove) = P(x = 3 | 7 prove)

$$\frac{7!}{(3!)\cdot(4!)}(0.7)^3(0.3)^4 = 35\cdot 0.343\cdot 0.0081 =$$

$$= 0.0972$$

Qual è la probabilità di osservare non più di 3 guariti su 7 trattati?



Soluzione

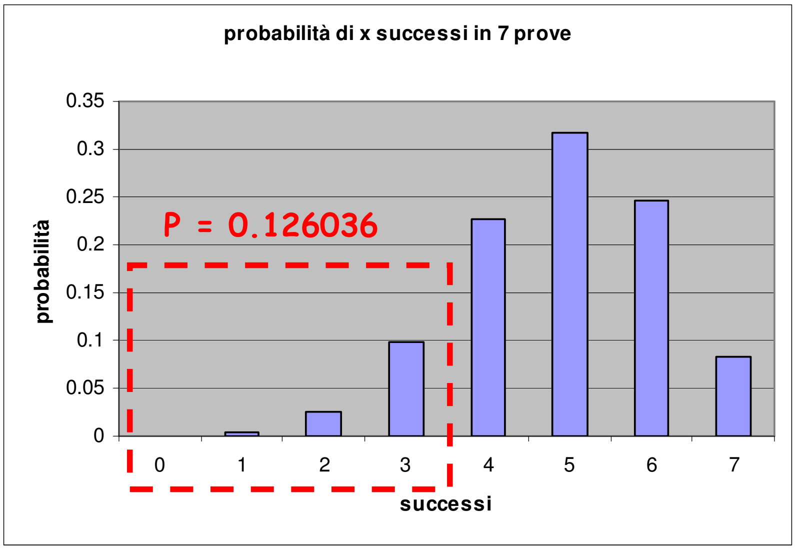
Qual è la probabilità associata ad uno qualunque dei possibili risultati della X che si possono osservare?

P(x ≤ 3 successi in 7 prove) = P(x = ≤ 3 | 7 prove) =

= P(0) + P(1) + P(2) + P(3) =

$$\frac{7!}{(0!)\cdot(7!)}(0.7)^0(0.3)^7 \quad + \dots + \quad \frac{7!}{(3!)\cdot(4!)}(0.7)^3(0.3)^4 =$$

= 0.0002187+0.00357+0.02500+0.09724 = 0.126036



Probabilità

Circa la metà di tutti i neonati è di sesso maschile. Nell'ospedale A nascono in media 60 bambini al mese, mentre nell'ospedale B nascono in media circa 20 bambini al mese.

In un mese qualsiasi, in quale dei due ospedali è più probabile osservare più dell'80% di nati di sesso femminile.

a. Ospedale A (60 parti al mese)
b. Ospedale B (20 parti al mese)
c. La probabilità è la stessa nei due ospedali

Ospedale A (60 parti)

$P(F) > 0.8 = P(49_F) + P(50_F) + P(51_F) + \ldots\ldots\ldots\ldots + P(60_F) =$

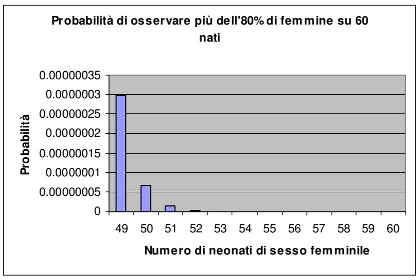
$$\frac{60 \cdot 59 \cdot \ldots \cdot 1}{(49 \cdot 48 \cdot \ldots \cdot 1) \cdot (11 \cdot 10 \cdot \ldots \cdot 1)} (0.5)^{49} (0.5)^{11} + \ldots..$$

Ospedale B (20 parti)

$P(F) > 0.8 = P(17_F) + P(18_F) + P(19_F) + P(20_F) =$

$$\frac{20 \cdot 19 \cdot \ldots \cdot 1}{(17 \cdot 16 \cdot \ldots \cdot 1) \cdot (3 \cdot 2 \cdot 1)} (0.5)^{17} (0.5)^{3} + \ldots..$$

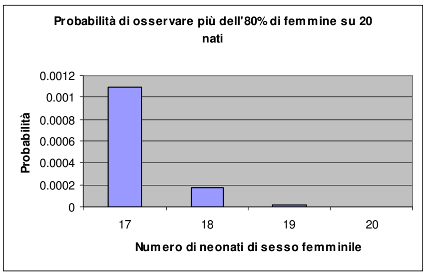
Ospedale A (60 parti)



P (F>80% su 60) =

3.78064E-07

Ospedale B (20 parti)



P (F>80% su 20) =

0.001288414

n = 5; pT = 0,45; pC = 0,45		
successi	A	B
0	0,05	0,05
1	0,21	0,21
2	0,34	0,34
3	0,28	0,28
4	0,11	0,11
5	0,02	0,02

Per confrontare due gruppi misuriamo la differenza nel numero di eventi osservati nei due gruppi.

Per interpretare la differenza dobbiamo conoscere come si comporterebbero le differenze se fosse solo il caso ad agire.

Cosa succederebbe se i gruppi fossero simili?

n = 5; pT = 0,45; pC = 0,45									
successi	A	B	A			B			
				0	1	2	3	4	5
0	0,05	0,05	0	0,0025	0,0105	0,0170	0,0140	0,0055	0,0010
1	0,21	0,21	1	0,0105	0,0441	0,0714	0,0588	0,0231	0,0042
2	0,34	0,34	2	0,0170	0,0714	0,1156	0,0952	0,0374	0,0068
3	0,28	0,28	3	0,0140	0,0588	0,0952	0,0784	0,0308	0,0056
4	0,11	0,11	4	0,0055	0,0231	0,0374	0,0308	0,0121	0,0022
5	0,02	0,02	5	0,0010	0,0042	0,0068	0,0056	0,0022	0,0004

Per confrontare due gruppi misuriamo la differenza nel numero di eventi osservati nei due gruppi.

Per interpretare la differenza dobbiamo conoscere come si comporterebbero le differenze se fosse solo il caso ad agire.

Cosa succederebbe se i gruppi fossero simili?

n = 5; pT = 0,45; pC = 0,45

successi	A	B	A	B 0	B 1	B 2	B 3	B 4	B 5
0	0,05	0,05	0	0,0025	0,0105	0,0170	0,0140	0,0055	0,0010
1	0,21	0,21	1	0,0105	0,0441	0,0714	0,0588	0,0231	0,0042
2	0,34	0,34	2	0,0170	0,0714	0,1156	0,0952	0,0374	0,0068
3	0,28	0,28	3	0,0140	0,0588	0,0952	0,0784	0,0308	0,0056
4	0,11	0,11	4	0,0055	0,0231	0,0374	0,0308	0,0121	0,0022
5	0,02	0,02	5	0,0010	0,0042	0,0068	0,0056	0,0022	0,0004

B-A	
5	0,0010
4	0,0097
3	0,0439
2	0,1188
1	0,2101
0	0,2531
-1	0,2101
-2	0,1188
-3	0,0439
-4	0,0097
-5	0,0010

Per confrontare due gruppi misuriamo la differenza nel numero di eventi osservati nei due gruppi.

Per interpretare la differenza dobbiamo conoscere come si comporterebbero le differenze se fosse solo il caso ad agire.

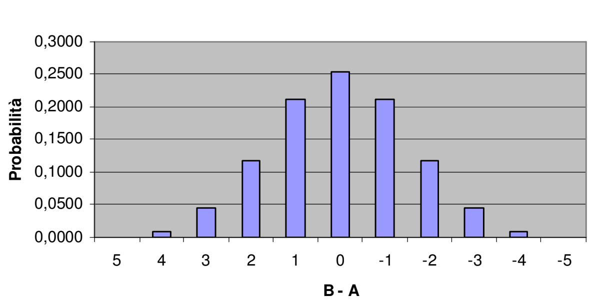
Cosa succederebbe se i gruppi fossero simili?

n = 5; pT = 0,45; pC = 0,45

successi	A	B	A	B 0	B 1	B 2	B 3	B 4	B 5
0	0,05	0,05	0	0,0025	0,0105	0,0170	0,0140	0,0055	0,0010
1	0,21	0,21	1	0,0105	0,0441	0,0714	0,0588	0,0231	0,0042
2	0,34	0,34	2	0,0170	0,0714	0,1156	0,0952	0,0374	0,0068
3	0,28	0,28	3	0,0140	0,0588	0,0952	0,0784	0,0308	0,0056
4	0,11	0,11	4	0,0055	0,0231	0,0374	0,0308	0,0121	0,0022
5	0,02	0,02	5	0,0010	0,0042	0,0068	0,0056	0,0022	0,0004

B-A	
5	0,0010
4	0,0097
3	0,0439
2	0,1188
1	0,2101
0	0,2531
-1	0,2101
-2	0,1188
-3	0,0439
-4	0,0097
-5	0,0010

Differenze B-A nel numero di successi

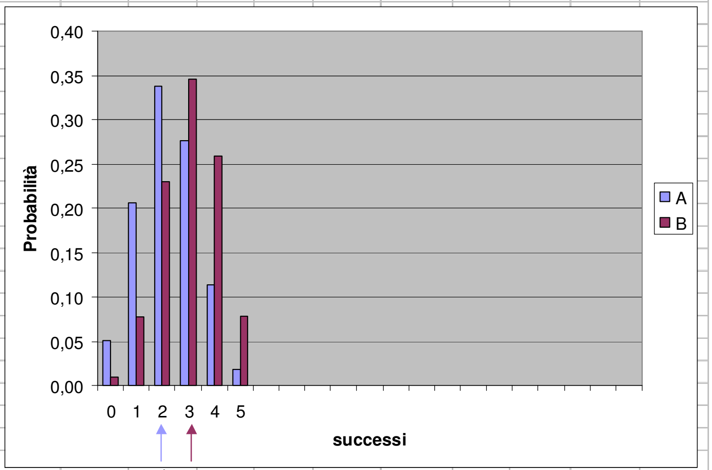


Distribuzione binomiale

n	p	q	n	p	q
5	0,45	0,55	5	0,6	0,4

successi	A	B
0	0,05	0,01
1	0,21	0,08
2	0,34	0,23
3	0,28	0,35
4	0,11	0,26
5	0,02	0,08

E se invece i gruppi fossero diversi?

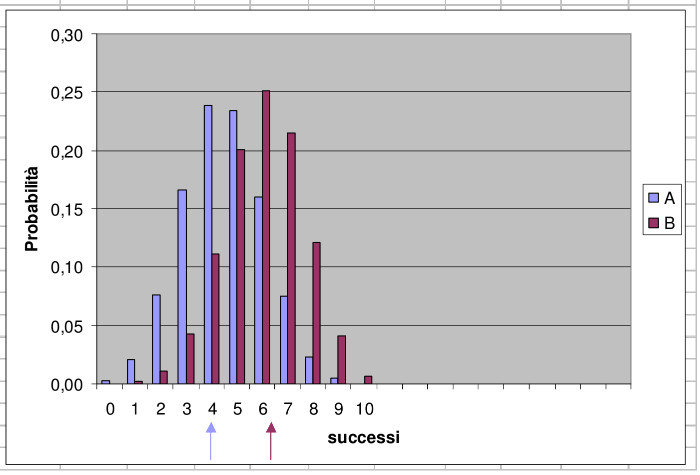


Media = np

Distribuzione binomiale

A			B		
n	p	q	n	p	q
10	0,45	0,55	10	0,6	0,4

successi	A	B
0	0,00	0,00
1	0,02	0,00
2	0,08	0,01
3	0,17	0,04
4	0,24	0,11
5	0,23	0,20
6	0,16	0,25
7	0,07	0,21
8	0,02	0,12
9	0,00	0,04
10	0,00	0,01

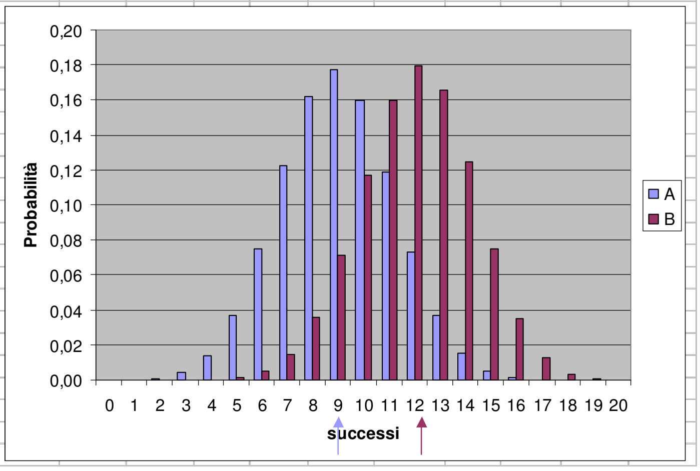


Media = np

Distribuzione binomiale

	A			B	
n	p	q	n	p	q
20	0,45	0,55	20	0,6	0,4

successi	A	B
0	0,00	0,00
1	0,00	0,00
2	0,00	0,00
3	0,00	0,00
4	0,01	0,00
5	0,04	0,00
6	0,07	0,00
7	0,12	0,01
8	0,16	0,04
9	0,18	0,07
10	0,16	0,12
11	0,12	0,16
12	0,07	0,18
13	0,04	0,17
14	0,01	0,12
15	0,00	0,07
16	0,00	0,03
17	0,00	0,01
18	0,00	0,00
19	0,00	0,00
20	0,00	0,00



Media = np

Come usiamo la distribuzione binomiale?

- Per confrontare un risultato osservato con la distribuzione attesa

- Per confrontare la proporzione di eventi di un certo tipo in differenti gruppi di soggetti

- La media (o valore atteso) è uguale a np

  La deviazione standard è uguale a $\sqrt{npq}$

- Con numerosità sufficientemente ampie si possono semplificare i calcoli utilizzando la distribuzione gaussiana al posto della binomiale

GALLO_Metodi statistici per lo studio dei fenomeni biologici/31