C.I. di Metodologia clinica

Modulo 5.

I metodi per la sintesi e la comunicazione delle informazioni sulla salute

Quali errori influenzano le stime?

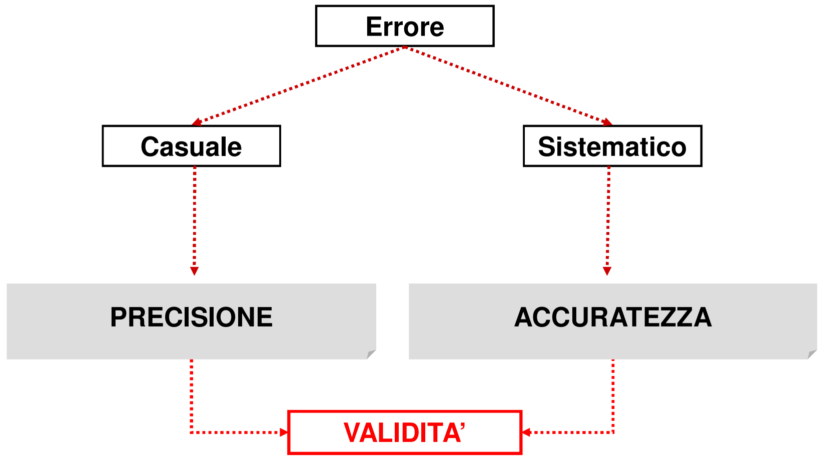
L'errore casuale

I metodi per la produzione delle informazioni sulla salute

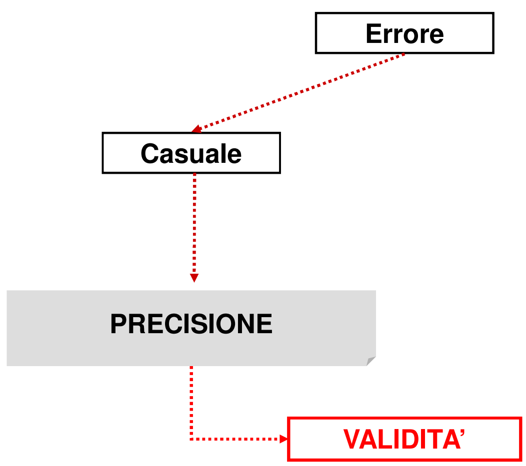
Alla fine di questa lezione dovreste essere in grado di:

- descrivere il significato di errore casuale
- spiegare il significato di distribuzione di campionamento
- calcolare la probabilità dell'intersezione di eventi indipendenti
- calcolare la probabilità dell'unione di più eventi
- descrivere l'effetto della numerosità sulla variabilità delle stime

Gli errori nella ricerca clinica



Gli errori nella ricerca clinica

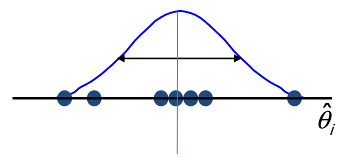


Il modello statistico

Effetto studiato +

Errore casuale =

Effetto osservato

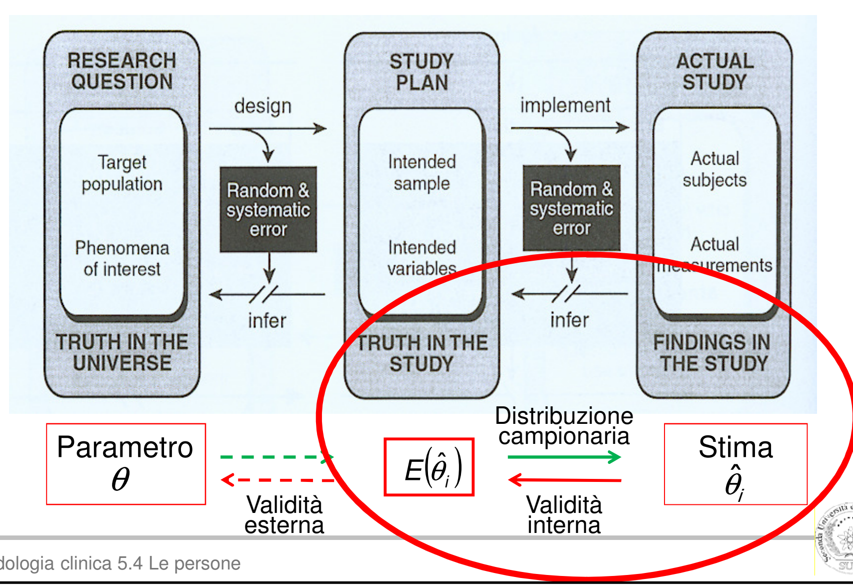


$E(\hat{\theta}_i) = \theta$

N.B. In assenza di errore sistematico!

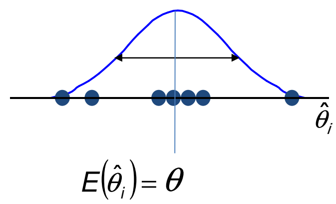
Il risultato individuale (stima) è imprevedibile, ma le potenziali conseguenze del caso (errore casuale) su insiemi di osservazioni possono essere anticipate con regole precise (distribuzioni di campionamento)

La struttura della ricerca clinica

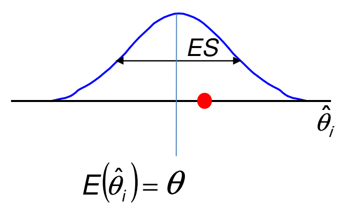


L'errore casuale della stima

E' la variazione intorno al parametro delle stime ottenute in tutti i possibili campioni con uguale numerosità, dovuta a cause indeterminate e non controllabili.



Il risultato dell'unico campione osservato è 'confrontato' con la distribuzione di probabilità (di campionamento) di tutte le stime che si potrebbero teoricamente osservare con campioni di uguale numerosità.



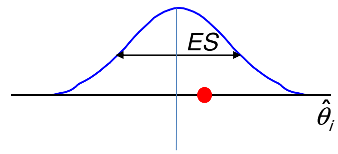
- $E(\hat{\theta}_i)$ è la media di tutte le (teoriche) stime campionarie
- L'errore standard (ES) è la deviazione standard di tutte le (teoriche) stime campionarie

Alcuni esempi

$(\hat{\theta}_i)$	$E(\hat{\theta}_i)$	$ES(\hat{\theta}_i)$	
$\hat{\mu}$	μ	$\sqrt{\dfrac{\hat{\sigma}^2}{n}}$	
$\hat{\mu}_A - \hat{\mu}_B = \hat{\delta}$	$\mu_A - \mu_B = \delta$	$\sqrt{\hat{\sigma}_p^2 \left(\dfrac{1}{n_A} + \dfrac{1}{n_B} \right)}$	Campioni indipendenti
$\hat{\mu}_A - \hat{\mu}_B = \hat{\delta}$	$\mu_A - \mu_B = \delta$	$\sqrt{\dfrac{\hat{\sigma}_{diff}^2}{n}}$	Campioni appaiati
$\hat{\pi}$	π	$\sqrt{\dfrac{\hat{\pi}(1-\hat{\pi})}{n}}$	
$\hat{\pi}_A - \hat{\pi}_B$	$\pi_A - \pi_B$	$\sqrt{\hat{\pi}(1-\hat{\pi}) \left(\dfrac{1}{n_A} + \dfrac{1}{n_B} \right)}$	

L'errore casuale della stima

Il risultato dell'unico campione osservato è 'confrontato' con la distribuzione di probabilità (di campionamento) di tutte le stime che si potrebbero teoricamente osservare con campioni di uguale numerosità.



$E(\hat{\theta}_i) = \theta$

- $E(\hat{\theta}_i)$ è la media di tutte le (teoriche) stime campionarie
- L'errore standard (ES) è la deviazione standard di tutte le (teoriche) stime campionarie
- In un intervallo di ± 2 *ES* intorno a $E(\hat{\theta}_i)$ è compreso circa il 95% di tutte le stime

Che relazione probabilistica esiste fra campione e popolazione? (Quali sono le stime che si potrebbero teoricamente osservare in un campione e con quali probabilità?)

Distribuzioni di campionamento

Discrete

Vengono specificati tutti i possibili valori (stime) e le loro probabilità

- Binomiale, Poisson, …

La somma di tutte le probabilità è uguale a 1

Continue

Viene associata una probabilità a un possibile intervallo di valori

- Gaussiana, *t* di Student, chi-quadrato, …

L'area totale sotto la curva è uguale a 1

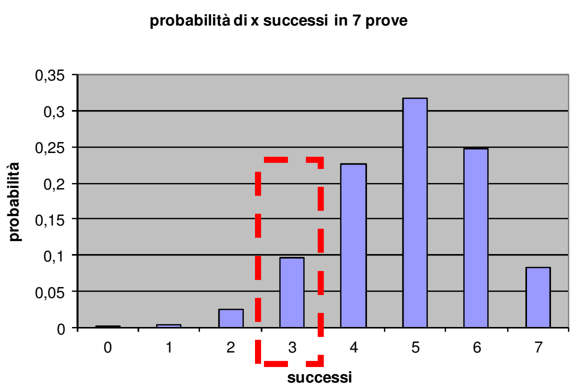
Esempio 1

Secondo la ditta produttrice, un nuovo farmaco antidolorifico ha una probabilità di 0,70 di far passare il mal di testa [$E(\hat{\theta}_i) = \theta$ = 0.70]

Abbiamo trattati 7 malati con il nuovo farmaco, ma abbiamo osservato un miglioramento in 3 casi (43%). [$\hat{\theta}_i$]
Come interpretiamo questo risultato?

Qual è la probabilità di osservare 3 guariti su 7 trattati?
(Più in generale, qual è la probabilità di osservare *x* eventi su *n* tentativi?)

Distribuzione binomiale



Probabilità di avere 3 guariti su 7 trattati = **0,0972**

Probabilità di osservare *x* volte un dato risultato in *n* tentativi

P (X = x) = $\dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

Come ci arriviamo (1)?

- Una sequenza possibile è G-NG-NG-G-G-NG-NG
- La cui probabilità è $0.7*0.3*0.3*0.7*0.7*0.3*0.3 =$
  $0.7^3*0.3^4 = 0.00278$

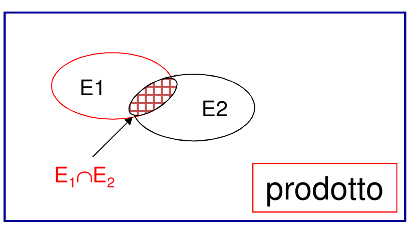
SE

- la probabilità di guarire è uguale ‘a priori’ per tutti i pazienti
- la probabilità di guarire non cambia in funzione dei risultati precedenti (cioè conoscere l’effetto di in farmaco su un paziente non fornisce informazioni sull’effetto dello stesso farmaco sul paziente successivo) → prove indipendenti

Intersezione di eventi indipendenti

prodotto

Due eventi sono indipendenti se



$P(E_1 \cap E_2) = P(E_1 \text{ e } E_2) =$

$= P(E_1) * P(E_2)$

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

Non possiamo sapere ‘a priori’ se due eventi sono indipendenti, se non per pianificazione

All’inverso, possiamo utilizzare la stima in condizioni di indipendenza come ‘riferimento’: se due eventi fossero indipendenti, allora la probabilità sarebbe …

Come ci arriviamo (2)?

- Ci sono 35 combinazioni possibili con 3 guariti su 7

$$\frac{n!}{x!(n-x)!}$$

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

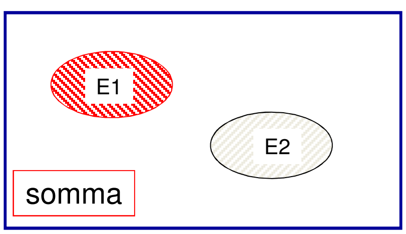
- Quindi la probabilità totale di 3 guariti su 7 è $35 * p^3 * (1-p)^4$

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

Unione di eventi

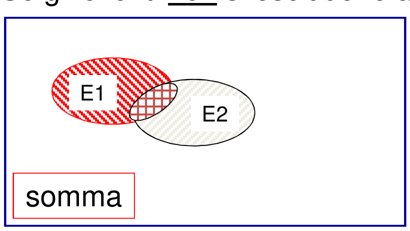
somma

Se gli eventi si escludono a vicenda



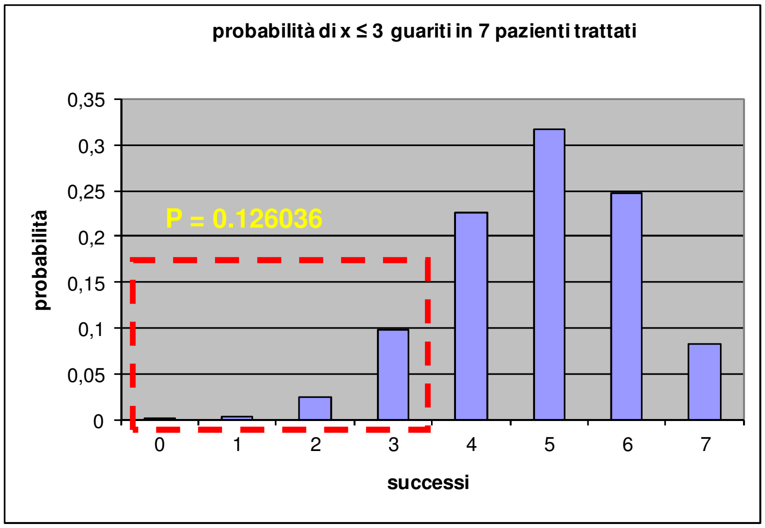
$P(E_1 \cup E_2) = P(E_1 \text{ o } E_2) =$
$= P(E_1) + P(E_2)$

Se gli eventi <u>non</u> si escludono a vicenda



$P(E_1 \cup E_2) = P(E_1 \text{ o } E_2) =$
$= P(E_1) + P(E_2) - P(E_1 \cap E_2)$

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



La somma di tutte le probabilità è 1

Esempio 2

Confrontiamo due trattamenti, A e B, che in realtà hanno la stessa efficacia. Che risultati possiamo osservare?

$$E\left(\hat{\delta}_i\right) = \delta = 0$$

- La probabilità di guarire con il trattamento A è 0,45.
- La probabilità di guarire con il trattamento B è 0,45.
- Sono trattate 5 persone per gruppo.

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

Probabilità di osservare 0, 1 … 5 guarigioni in 5 soggetti trattati per ognuno dei due trattamenti A e B

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

Probabilità associate a tutte le combinazioni possibili del numero di successi in A e B (prodotto).

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

A	B 0	1	2	3	4	5
0	0,0025	0,0105	0,0170	0,0140	0,0055	0,0010
1	0,0105	0,0441	0,0714	0,0588	0,0231	0,0042
2	0,0170	0,0714	0,1156	0,0952	0,0374	0,0068
3	0,0140	0,0588	0,0952	0,0784	0,0308	0,0056
4	0,0055	0,0231	0,0374	0,0308	0,0121	0,0022
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

Probabilità associate a tutte le differenze possibili nel numero di successi fra B e A (somma).

Le differenze possibili vanno da +5 (quando si osservano 5 guarigioni con B e nessuna con A) a -5 (quando si osservano 5 guarigioni con A e nessuna con B).

La probabilità di osservare una differenza pari a 0 (il valore reale) è uguale a 0,25.

La somma delle probabilità di tutte le differenze è 1

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

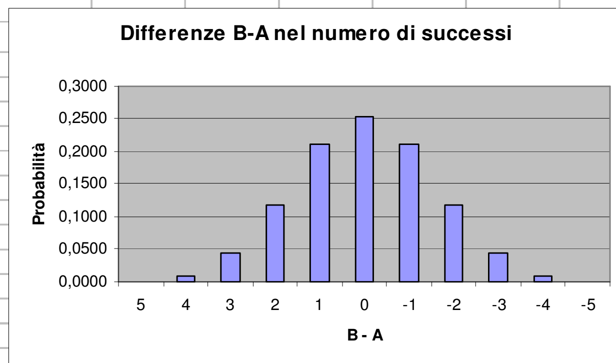
A	B 0	1	2	3	4	5
0	0,0025	0,0105	0,0170	0,0140	0,0055	0,0010
1	0,0105	0,0441	0,0714	0,0588	0,0231	0,0042
2	0,0170	0,0714	0,1156	0,0952	0,0374	0,0068
3	0,0140	0,0588	0,0952	0,0784	0,0308	0,0056
4	0,0055	0,0231	0,0374	0,0308	0,0121	0,0022
5	0,0010	0,0042	0,0068	0,0056	0,0022	0,0004

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

- Se non c'è un reale effetto del trattamento, è comunque possibile che per caso si osservino differenze tra i gruppi, con probabilità diverse

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

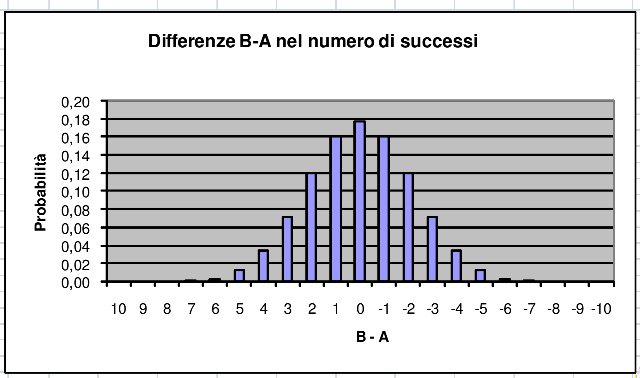
B-A	
5	0,0010
4	0,0097
3	0,0439
2	0,1188
1	0,2101
0	0,2531
-1	0,2101
-2	0,1188
-3	0,0439
-4	0,0097
-5	0,0010



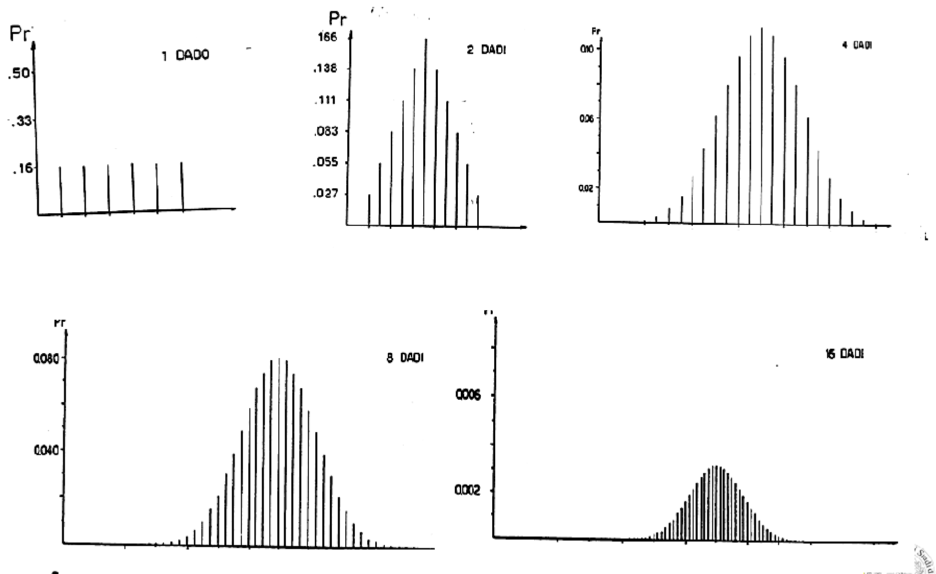
n = 10; pT = 0,45; pC = 0,45

successi	A	B	A	B 0	1	2	3	4	5	6	7	8	9	10
0	0,00	0,00	0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1	0,02	0,02	1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	0,08	0,08	2	0,00	0,00	0,01	0,01	0,02	0,02	0,01	0,01	0,00	0,00	0,00
3	0,17	0,17	3	0,00	0,00	0,01	0,03	0,04	0,04	0,03	0,01	0,00	0,00	0,00
4	0,24	0,24	4	0,00	0,00	0,02	0,04	0,06	0,06	0,04	0,02	0,00	0,00	0,00
5	0,23	0,23	5	0,00	0,00	0,02	0,04	0,06	0,05	0,04	0,02	0,00	0,00	0,00
6	0,16	0,16	6	0,00	0,00	0,01	0,03	0,04	0,04	0,03	0,01	0,00	0,00	0,00
7	0,07	0,07	7	0,00	0,00	0,01	0,01	0,02	0,02	0,01	0,00	0,00	0,00	0,00
8	0,02	0,02	8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
9	0,00	0,00	9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
10	0,00	0,00	10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

B-A	
10	0,00
9	0,00
8	0,00
7	0,00
6	0,00
5	0,0122
4	0,0341
3	0,0718
2	0,1194
1	0,1606
0	**0,1771**
-1	0,1606
-2	0,1194
-3	0,0718
-4	0,0341
-5	0,0122
-6	0,00
-7	0,00
-8	0,00
-9	0,00
-10	0,00



Verso la gaussiana



Probabilità e numerosità

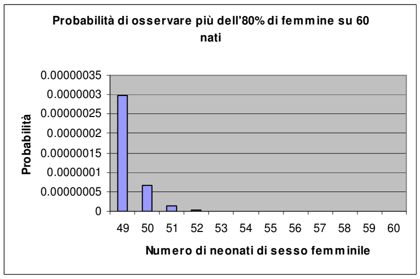
Circa la metà di tutti i neonati è di sesso maschile. Nell'ospedale A nascono in media 60 bambini al mese, mentre nell'ospedale B nascono in media circa 20 bambini al mese.

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

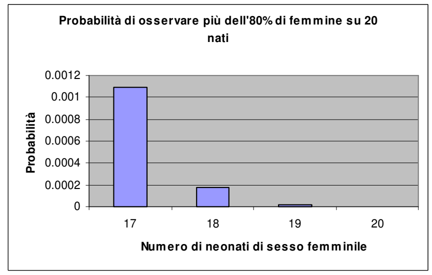
Probabilità e numerosità

Ospedale A (60 parti)



P (F>80% su 60) = 3,78064E-07

Ospedale B (20 parti)



P (F>80% su 20) = 0,001288414

E' più probabile osservare valori estremi (es. differenze più grandi) con basse numerosità

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)

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

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