

# Intervalli di Confidenza

Corso di Statistica di base

Giancarlo Ferrari

## Introduzione

- Nella lezione precedente sono state esaminate le proprietà delle distribuzioni di medie campionarie
- In tutti i casi abbiamo assunto che conoscevamo la media  $\mu$  della popolazione e la sua deviazione standard  $\sigma$  e questo ha consentito di ideare un esperimento ideale nel corso del quale venivano prelevati tutti i possibili campioni di dimensione  $n$
- L'importante risultato al quale si era giunti è che era stato possibile individuare i valori (limite minimo e massimo) che racchiudono una ben definita proporzione dei valori della intera distribuzione delle medie campionarie utilizzando le proprietà della deviato standardizzata

## Introduzione

- Ad esempio il 95% dei valori campionari era compreso entro  $\pm 1,96$  errori standard
- Se avessimo voluto trovare l'intervallo entro il quale erano contenuti il 99% dei valori avremmo dovuto moltiplicare l'errore standard per 2,58 (il valore della devziata standardizzata che delimita un'area pari a 0,005 a destra di +2,58 e a sinistra di -2,58)

## Introduzione

- Spesso i parametri della popolazione sono sconosciuti ed infatti l'obiettivo che abbiamo è proprio quello di ottenere una stima di tali parametri e di solito vengono utilizzati due metodi di stima
- Il primo è chiamato metodo della *stima puntuale* mentre il secondo è chiamato *stima intervallare*

## Stima puntuale della media campionaria

- Il metodo della stima puntuale è relativamente semplice e si utilizzano i metodi già descritti nelle prime lezioni
- Si abbiano una serie di osservazioni  $x_1 + x_2 + x_3 + \dots + x_n$
- La media aritmetica semplice è  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
- La stima puntuale può essere migliorata e resa più informativa attraverso la stima intervallare

## Stima intervallare della media campionaria

- Questa tecnica fornisce un intervallo di possibili valori entro i quali si ritiene sia compreso il parametro in esame – in questo caso la media  $\mu$  della popolazione (sconosciuta) – con un certo grado di confidenza
- Questo range di valori è chiamato *Intervallo di Confidenza*

## Stima intervallare della media campionaria

- Vediamo come poter sfruttare le proprietà della distribuzione delle medie campionarie per ottenere una stima intervallare della media
- Abbiamo già visto che data una variabile casuale  $\bar{x}$  (intesa come media campionaria) stimata su un campione pari a  $n$  la devziata standardizzata è distribuita normalmente e  $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

## Stima intervallare della media campionaria

- Per una distribuzione normale standardizzata sappiamo inoltre che il 95% dei valori di  $Z$  è compreso tra  $-1,96$  e  $+1,96$
- Il 99% dei valori di  $Z$  è compreso tra  $-2,58$  e  $+2,58$



Stima intervallare della media campionaria

- In altri termini la probabilità che  $Z$  assuma un valore compreso tra -1,96 e +1,96 può essere espressa nel seguente modo:

$$\Pr(-1,96 \leq Z \leq +1,96) = 0,95$$

che può essere espressa in maniera equivalente come:

$$\Pr\left(-1,96 \leq \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \leq +1,96\right) = 0,95$$

## Stima intervallare della media campionaria

- La disuguaglianza può essere manipolata all'interno della parentesi senza modificare la probabilità.
- Moltiplicando i tre termini per l'errore standard  $\sigma/\sqrt{n}$  si ottiene:

$$\Pr\left(-1,96 \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq +1,96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

Stima intervallare della media campionaria

- Sottraendo da ciascun termine il valore  $\bar{x}$  si ha:

$$\Pr\left(-1,96 \cdot \frac{\sigma}{\sqrt{n}} - \bar{x} \leq -\mu \leq +1,96 \cdot \frac{\sigma}{\sqrt{n}} - \bar{x}\right) = 0,95$$

- Moltiplicando per -1 si inverte la direzione della disuguaglianza

$$\Pr\left(1,96 \cdot \frac{\sigma}{\sqrt{n}} + \bar{x} \geq \mu \geq -1,96 \cdot \frac{\sigma}{\sqrt{n}} + \bar{x}\right) = 0,95$$

Stima intervallare della media campionaria

- La disuguaglianza può essere riorganizzata come:

$$\Pr\left(\bar{x} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

Pertanto il valore della media vera (sconosciuta) della popolazione  $\mu$  sarà compresa entro i limiti dell'intervallo di confidenza al 95% dato da:

$$\bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}}$$

## Stima intervallare della media campionaria

- Il significato del risultato ottenuto è il seguente: **SE SELEZIONASSIMO 100 CAMPIONI CASUALI DALLA POPOLAZIONE ED UTILIZZIAMO QUESTI CAMPIONI PER CALCOLARE 100 DIVERSI INTERVALLI DI CONFIDENZA 95 DI QUESTI INTERVALLI DI CONFIDENZA CONTERRANNO LA MEDIA  $\mu$**
- Si ricordi che  $\bar{x}$  è la variabile casuale mentre  $\mu$  è un parametro fisso costante e pertanto l'intervallo

$$\Pr\left(\bar{x} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

rappresenta la probabilità (95%) a priori che l'intervallo  $\bar{x} \pm 1,96 \cdot SE$  contenga  $\mu$

## Stima intervallare della media campionaria

- Naturalmente se avessimo voluto ottenere un intervallo di confidenza pari al 99% di probabilità di includere il valore sconosciuto  $\mu$  avremmo dovuto moltiplicare l'Errore Standard per 2,58

$$\Pr\left(\bar{x} - 2,58 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2,58 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0,99$$

- 99 di 100 intervalli di confidenza ottenuti da 100 campioni casuali indipendenti conterranno la media vera  $\mu$

## Stima intervallare della media campionaria

- Ora è possibile vedere bene l'effetto che ha la dimensione campionaria nella stima intervallare di una media
- Si supponga di aver estratto un campione di 100 individui da una popolazione molto grande (supponiamo che la variabile sia il peso) e di averne stimato sia il valore medio che la deviazione standard campionaria:

$$\bar{x} = 65 \text{ e deviazione standard } s = 8$$

## Stima intervallare della media campionaria

- L'intervallo di confidenza al 95% sarebbe pari a

$$LS = 65 + \left( 1,96 \cdot \frac{8}{\sqrt{100}} \right) = 66,57$$

$$LI = 65 - \left( 1,96 \cdot \frac{8}{\sqrt{100}} \right) = 63,43$$

- Il significato di questo risultato è che vi è il 95% di probabilità che l'intervallo contenga la media vera della popolazione  $\mu$  (sconosciuta)



## Stima intervallare della media campionaria

- Se gli stessi risultati fossero stati ottenuti con un campione di 200 individui l'intervallo di confidenza al 95% sarebbe pari a

$$LS = 65 + \left( 1,96 \cdot \frac{8}{\sqrt{200}} \right) = 66,11$$

$$LI = 65 - \left( 1,96 \cdot \frac{8}{\sqrt{200}} \right) = 63,89$$

L'intervallo è naturalmente di dimensioni inferiori e pertanto la stima intervallare è più precisa.

Stima intervallare della media campionaria

La distribuzione t di Student

## Stima intervallare della media campionaria – la distribuzione t di Student

- Nel calcolare gli intervalli di confidenza per la media campionaria facendo uso della stima della media e della deviazione standard ottenuta dal campione abbiamo dato per scontato che l'intervallo di confidenza potesse essere calcolato utilizzando i valori della deviazione standardizzata  $Z$
- Ciò è stato possibile perché abbiamo utilizzato una numerosità campionaria elevata e comunque tale che l'assunzione di normalità della distribuzione della media campionaria non veniva violata

## Stima intervallare della media campionaria – la distribuzione t di Student

- In realtà se la stima della media e della deviazione standard viene effettuata su campioni di piccole dimensioni (un campione di 15 soggetti può essere considerato un campione di piccole dimensioni) le deviazioni dalla normalità possono essere rilevanti
- In questi casi invece che usare la distribuzione della Deviata Standardizzata Z si usa la distribuzione t di Student

## Stima intervallare della media campionaria – la distribuzione t di Student

- La t di Student si distribuisce in maniera analoga alla Z e la sua formula è praticamente equivalente, infatti per un campione casuale di dimensione  $n$  selezionato da una popolazione normale si ha:

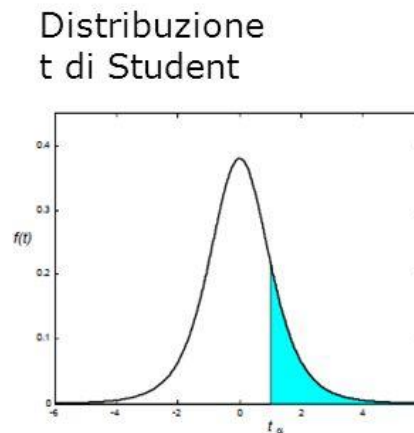
$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

che è nota come distribuzione t di Student con  $n-1$  gradi di libertà (i gradi di libertà si indicano con la lettera greca  $\nu$  – nu)

Nella distribuzione t di Student il valore della deviazione standard  $s$  che compare al denominatore è la stima campionaria di  $\sigma$

# Stima intervallare della media campionaria – la distribuzione t di Student

- Come la distribuzione normale la distribuzione t di Student è simmetrica e l'area totale sotto la curva vale 1
- Al crescere dei gradi libertà i valori di t si avvicinano quelli della devziata standardizzata Z



Ad esempio:  
gdl=9  $\alpha=0.05$ :  
 $t_{0.05}=1.833$

Se gdl=9, trovare il valore di  $t_\alpha$  tale che la somma dell'area a destra di  $t_\alpha$  e dell'area a sinistra di  $-t_\alpha$  vale  $\alpha = 0.05$

Area totale delle due code =  $\alpha = 0.05$   
area a destra di  $t_\alpha$  (una coda) =  $\alpha/2 = 0.025$ .

$t_\alpha = t_{0.025} = 2.262$

v	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	v
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
$\infty$	1.282	1.645	1.960	2.326	2.576	$\infty$

# Stima intervallare della media campionaria – la distribuzione t di Student

- Si osservi che i gradi di libertà non sono altro che la numerosità campionaria e che al crescere di essa t e Z tendono allo stesso valore
- La distribuzione t tiene conto del fatto che con numerosità campionarie basse la distribuzione delle medie campionarie può oscillare entro intervalli più ampi e pertanto i valori della deviato standardizzata sono più elevati

df	Area nella Coda di Destra sotto la Curva di Distribuzione t					
	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
31	1.309	1.696	2.040	2.453	2.744	3.375
32	1.309	1.694	2.037	2.449	2.738	3.365
33	1.308	1.692	2.035	2.445	2.733	3.356
34	1.307	1.691	2.032	2.441	2.728	3.348
35	1.306	1.690	2.030	2.438	2.724	3.340
36	1.306	1.688	2.028	2.434	2.719	3.333
37	1.305	1.687	2.026	2.431	2.715	3.326
38	1.304	1.686	2.024	2.429	2.712	3.319
39	1.304	1.685	2.023	2.426	2.708	3.313
40	1.303	1.684	2.021	2.423	2.704	3.307
∞	1.282	1.645	1.960	2.326	2.576	3.090

# Stima intervallare della media campionaria – la distribuzione t di Student

- La tavola a destra è una tabella ridotta delle aree calcolate per la distribuzione t
- A differenza delle tavole Z le aree sono indicate nella prima riga
- Nella prima colonna indicata come  $df$  (degrees of freedom) vengono indicati i gradi di libertà (la numerosità campionaria)

$df$	Area nella Coda di Destra sotto la Curva di Distribuzione t					
	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
31	1.309	1.696	2.040	2.453	2.744	3.375
32	1.309	1.694	2.037	2.449	2.738	3.365
33	1.308	1.692	2.035	2.445	2.733	3.356
34	1.307	1.691	2.032	2.441	2.728	3.348
35	1.306	1.690	2.030	2.438	2.724	3.340
36	1.306	1.688	2.028	2.434	2.719	3.333
37	1.305	1.687	2.026	2.431	2.715	3.326
38	1.304	1.686	2.024	2.429	2.712	3.319
39	1.304	1.685	2.023	2.426	2.708	3.313
40	1.303	1.684	2.021	2.423	2.704	3.307
$\infty$	1.282	1.645	1.960	2.326	2.576	3.090



## Stima intervallare della media campionaria – la distribuzione t di Student

- Ad esempio se avessimo calcolato l'intervallo di confidenza al 95% utilizzando un campione di 20 individui (supponiamo che la variabile sia il peso) e, come in precedenza, di averne stimato sia il valore medio che la deviazione standard campionaria:

$$\bar{x} = 65 \text{ e deviazione standard } s = 8$$

- L'intervallo di confidenza al 95% utilizzando la distribuzione t di Student ( $t_{\alpha/2=0,025; \nu=19}$ ) sarebbe pari a:

$$UL = 65 + \left( 2,093 \cdot \frac{8}{\sqrt{20}} \right) = 68,74$$

$$LL = 65 - \left( 2,093 \cdot \frac{8}{\sqrt{20}} \right) = 61,25$$

Stima intervallare di una proporzione

## Stima intervallare di una proporzione

- Nel calcolare l'intervallo di confidenza per una proporzione campionaria si può adottare lo stesso criterio utilizzato per una variabile di tipo continuo sfruttando il fatto che una variabile casuale binomiale per valori di  $n$  sufficientemente grandi segue una distribuzione normale
- Supponiamo che si voglia studiare la distribuzione della media campionaria  $p$  stimata su  $n$  prove indipendenti pari a 100
- Si supponga inoltre che la caratteristica che si vuole studiare è 'occhi azzurri' in una popolazione dove la proporzione 'occhi azzurri' è uguale a  $\pi$
- Il numero  $s$  di individui con la caratteristica 'occhi azzurri' dovrebbe seguire una distribuzione binomiale

## Stima intervallare di una proporzione

- La disuguaglianza utilizzata in precedenza per  $\mu$  la si può utilizzare per  $\pi$  (parametro fisso costante) dove l'errore standard della media campionaria è:

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

## Stima intervallare di una proporzione

- In analogia con la disuguaglianza utilizzata in precedenza per  $\mu$  l'intervallo

$$\Pr\left(p - 1,96 \cdot \sqrt{\frac{\pi(1 - \pi)}{n}} \leq \pi \leq p + 1,96 \cdot \sqrt{\frac{\pi(1 - \pi)}{n}}\right) = 0,95$$

rappresenta la probabilità (95%) a priori che l'intervallo contenga  $\pi$

## Stima intervallare di una proporzione

- Il significato del risultato ottenuto è analogo a quanto ottenuto in precedenza: **SE SELEZIONASSIMO 100 CAMPIONI CASUALI DALLA POPOLAZIONE ED UTILIZZIAMO QUESTI CAMPIONI PER CALCOLARE 100 DIVERSI INTERVALLI DI CONFIDENZA 95 DI QUESTI INTERVALLI DI CONFIDENZA CONTERRANNO LA MEDIA  $\pi$**

## Stima intervallare della proporzione campionaria

- Anche in questo caso è possibile vedere bene l'effetto che ha la dimensione campionaria nella stima intervallare di una proporzione
- Si supponga che nel campione di 100 individui estratti da una popolazione molto grande la proporzione «occhi azzurri» sia  $p = 0,06$  (o 6%)

$$\bar{p} = \frac{6}{100} = 0,06 \text{ e } ES = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,06(1-0,06)}{100}} = 0,024$$

Stima intervallare della proporzione campionaria

- Pertanto l'intervallo di confidenza al 95% sarà:

$$LS = \frac{6}{100} = 0,06 + (1,96 \cdot 0,024) = 0,1065 \text{ (o } 10,65\%)$$

$$LI = \frac{6}{100} = 0,06 - (1,96 \cdot 0,024) = 0,0134 \text{ (o } 1,34\%)$$



## Stima intervallare di una proporzione

- Il significato del risultato ottenuto è analogo a quanto ottenuto in precedenza: **SE SELEZIONASSIMO 100 CAMPIONI CASUALI DALLA POPOLAZIONE ED UTILIZZIAMO QUESTI CAMPIONI PER CALCOLARE 100 DIVERSI INTERVALLI DI CONFIDENZA 95 DI QUESTI INTERVALLI DI CONFIDENZA CONTERRANNO LA MEDIA  $\pi$**

Stima intervallare di una proporzione con intervallo di confidenza esatto

- In analogia con quanto si è visto per le medie campionarie stimate su campioni di piccole dimensioni anche per le proporzioni campionarie vi possono essere deviazioni dalla normalità
- In particolare l'approssimazione della binomiale alla normale è buona a condizione che  $(n \cdot \pi)$  e  $n \cdot (1 - \pi)$  sono entrambi  $> 5$

## Stima intervallare di una proporzione con intervallo di confidenza esatto

- Si abbia come esempio la stessa stima di  $p = 0,06$  ottenuta però su un campione pari a 50
- Si osservi che  $np=3$  e che  $n(1-p)=47$
- Se utilizzassimo la distribuzione normale per l'intervallo di confidenza al 95% il risultato sarebbe:

$$LS = 0,06 + (1,96 \cdot \sqrt{(0,06 \cdot 0,94)/50}) = 0,1258 \text{ (12,58\%)}$$

$$LI = 0,06 - (1,96 \cdot \sqrt{(0,06 \cdot 0,94)/50}) = -0,0058 \text{ (-0,58\%)}$$

Naturalmente non è possibile ottenere un valore negativo poiché una distribuzione di probabilità può assumere solo valori che vanno da 0 a 1

## Stima intervallare di una proporzione con intervallo di confidenza esatto

- In realtà poiché uno dei prodotti  $n\pi$  e  $n(1 - \pi)$  è inferiore a 5 non è appropriato utilizzare le proprietà della distribuzione normale (infatti  $n \cdot p = 3$ )
- In questo caso è necessario utilizzare gli intervalli di confidenza esatti (metodo di Clopper Pearson) che per un campione pari a 50 e  $p=0,06$  sono pari a:

$$UL = 0,1448$$

$$LL = 0,0203$$

Stima intervallare di una proporzione con intervallo di confidenza esatto

- Si osservi che il Limite Superiore è maggiore di quello stimato con l'approssimazione Normale ed inoltre che il Limite Inferiore non è inferiore a 0

$$UL = 0,1448$$

$$LL = 0,0203$$

- Si osservi inoltre che l'intervallo è asimmetrico (come era naturale aspettarsi)

Stima intervallare di una proporzione con intervallo di confidenza esatto

- Si osservi che il Limite Superiore è maggiore di quello stimato con l'approssimazione Normale ed inoltre che il Limite Inferiore non è inferiore a 0

$$UL = 0,1448$$

$$LL = 0,0203$$

- Si osservi inoltre che l'intervallo è asimmetrico (come era naturale aspettarsi)

Stima intervallare di una proporzione con intervallo di confidenza esatto

- Metodo di Clopper Pearson per intervalli di confidenza esatti

$$\frac{1}{1 + \frac{n - x + 1}{x} F_{1(n-x+1), 2x, \alpha/2}} \leq p \leq \frac{\frac{x + 1}{n - x} F_{2(x+1), 2(n-x), \alpha/2}}{1 + \frac{x + 1}{n - x} F_{2(x+1), 2(n-x), \alpha/2}}$$

Stima intervallare di una proporzione con intervallo di confidenza esatto

- I limiti di confidenza esatti al 95% per  $p$  sono i due valori  $p_{inf}$  e  $p_{sup}$  per i quali il valore osservato ( $s = 3$  per  $n=50$  dell'esempio precedente) diventa significativo ad un test unilaterale al livello dello 0,025

$p_{inf}$  risponde alla domanda: qual è il valore di  $p < 0,06$  (inferiore al valore medio) tale che la probabilità di 3 o più è 0,025 su 50 prove indipendenti?

$p_{sup}$  risponde alla domanda: qual è il valore di  $p > 0,06$  (superiore al valore medio) tale che la probabilità di 3 o meno è 0,025 su 50 prove indipendenti?



## Stima intervallare di una proporzione

- I limiti dell'intervallo di confidenza esatti sono tabulati (assieme ad altre tavole statistiche) in apposite tavole scientifiche

