1.4 I sondaggi prima delle elezioni

Una popolazione costituita da **10.000** individui è chiamata a votare tra due candidati, diciamo A e B. Mediante un sondaggio effettuato su un campione della popolazione, si è stimato che la probabilità che l'**individuo** sia favorevole ad A è del **40**%.
Sulla base di ciò, si vuole stimare il numero **F** di individui che voteranno A, supponendo che tutti si presentino alle votazioni. Precisamente, quanto vale $\mathbf{P}(\mathbf{3900} \leq \mathbf{F} \leq \mathbf{4100})$?

Osservazione
In sostanza, a partire dalla percentuale di individui del campione che sono a favore di A, si stima la probabilità che il generico individuo della popolazione sia a favore di A.

Risoluzione

Proviamo a modellizzare la situazione proposta mediante lo schema delle prove ripetute. Affinché ciò abbia senso, assumiamo che le prove siano *indipendenti*[5], ossia che il voto della persona non sia influenzato dal voto delle altre, e che le prove avvengano tutte nelle *medesime condizioni.*

Precisamente interpretiamo la questione nel modo seguente:

- una sequenza di $\boldsymbol{n = 10.000}$ prove, ciascuna delle quali corrisponde ad una persona;

- ciascuna prova ha due soli esiti possibili: *la persona vota A* oppure *la persona vota B*; dall'esito del sondaggio stimiamo che la probabilità che l'individuo sia favorevole ad A è $\boldsymbol{p = 0{,}4}$.

Secondo il nostro modello la probabilità che il numero di individui F sia, ad esempio, 3900 è[6]

$$\mathrm{P}(F = 3900) = \binom{10000}{3900} \cdot 0{,}4^{3900} \cdot 0{,}6^{6100}$$

e la probabilità richiesta si può esprimere nella forma

$$\mathrm{P}(\mathbf{3900} \leq \mathbf{F} \leq \mathbf{4100}) = \mathrm{P}(F = \mathbf{3900}) + \mathrm{P}(F = \mathbf{3901}) + ... + \mathrm{P}(F = \mathbf{4100})$$

Trattandosi di parecchie valutazioni di probabilità, utilizziamo un calcolatore[7]. Alla fine otteniamo che

$$P(3900 \leq F \leq 4100) \simeq \boxed{0{,}95978}$$

[5] Non possiamo dire che ciò avvenga realmente. Si tratta solamente di una nostra ipotesi che assumiamo, almeno in prima approssimazione, allo scopo di costruire un modello ragionevolmente semplice.

[6] Ricordiamo che la probabilità di ottenere k successi è data da

$$\mathrm{P}(k \text{ successi}) = \binom{n}{k} p^k (1-p)^{n-k}$$

[7] Con Excel si può usare la funzione $DISTRIB.BINOM.N(num\_successi; prove; probabilit\acute{a}\_s; cumulativo)$ con *cumulativo*=FALSO. Essa restituisce la probabilità di avere un numero di successi esattamente pari a *num_successi*. E tale calcolo va effettuato per ciascun valore da 3900 a 4100 come si vede in figura.

n	10000	p	0,4
k	P (k successi)		
3900	0,00101		P totale
3901	0,00106		0,95978
3902	0,00110		
3903	0,00115		
3904	0,00119		
3905	0,00124		
3906	0,00129		
3907	0,00134		
3908	0,00140		
3909	0,00145		
3910	0,00151		
3911	0,00156		
3912	0,00162		
3913	0,00168		
3914	0,00174		
3915	0,00181		
3916	0,00187		
3917	0,00194		
3918	0,00201		
3919	0,00208		
3920	0,00215		
4090	0,00151		
4091	0,00145		
4092	0,00140		
4093	0,00134		
4094	0,00129		
4095	0,00124		
4096	0,00120		
4097	0,00115		
4098	0,00110		
4099	0,00106		
4100	0,00102		

Osservazione
Se nella funzione di Excel $DISTRIB.BINOM.N(num\_successi; prove; probabilitá\_s; cumulativo)$ poniamo *cumulativo*=VERO, essa restituisce la probabilità di avere un numero di successi **al più** pari a *num_successi*.
Come useresti questa funzione al nostro scopo?