7.4 Applicazioni: finalmente i sondaggi

Torniamo al problema guida proposto nel paragrafo 1.1.

> 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. Precisamente, quanto vale
>
> $$\mathbf{P}(\mathbf{3900} \leq \mathbf{F} \leq \mathbf{4100})\ ?$$

Ora mediante il TLC possiamo risolverlo in modo computazionalmente meno articolato:

1. modellizziamo la situazione mediante la distribuzione binomiale, come già fatto all'inizio del percorso; essa ha parametri $n = 10.000$ e $p = 0,4$
2. per il TLC, il numero di individui F si approssima mediante la v.a. normale X di parametri
$$\mu = np = 4.000 \quad \text{e} \quad \sigma^2 = np(1-p) = 2400$$
3. calcoliamo quindi la probabilità richiesta

$$\mathbf{P}(\mathbf{3900} \leq \mathbf{F} \leq \mathbf{4100}) \underset{\substack{\uparrow \\ \textit{per il TLC}}}{\simeq} P(3900 \leq X \leq 4100) \underset{\substack{\uparrow \\ \textit{standardizzando } X}}{=} P\left(\frac{3900-4000}{\sqrt{2400}} \leq Z \leq \frac{4100-4000}{\sqrt{2400}}\right)$$

$$\simeq P(-2,04 \leq Z \leq 2,04) \underset{\substack{\uparrow \\ \textit{ricorrendo alle tavole}}}{\simeq} 0,9586$$

Pertanto[30] la probabilità che la percentuale di cittadini favorevoli ad A sia compresa tra il 39% e il 41% dell'intera popolazione è circa $\boxed{0,9586}$.

[30] Nell'ultimo passaggio, conviene calcolare $P(-2,04 \leq Z \leq 2,04)$ sfruttando la simmetria della densità normale e dell'intervallo di variabilità di Z. Pertanto tale probabilità è uguale a $2P(0 \leq Z \leq 2,04) = 2(P(Z \leq 2,04) - P(Z \leq 0))$. Dalle tavole si deduce che $P(Z \leq 2,04) \simeq 0,9793$; invece per $P(Z \leq 0)$ **non serve** ricorrere alle tavole.

Osservazione: i due approcci a confronto

- Il TLC permette di risolvere il problema iniziale senza dover svolgere molti conti. Invece il ricorso al modello binomiale, come si accorse De Moivre, comporta parecchie valutazioni di probabilità ($P(F = 3900) + P(F = 3901) + ... + P(F = 4100)$).

- Proviamo comunque a ricorrere alla sola distribuzione binomiale e usiamo un calcolatore[31] per portare a termine il calcolo impostato già nel paragrafo 1.1. Risulta

$$P(3900 \leq F \leq 4100) \simeq 0,95877$$

  I due numeri (quello trovato con la normale e quello trovato con la binomiale) sono "vicini" ma comunque diversi. Qual è quello "giusto"?
  In realtà ... nessuno dei due. Infatti i due valori di probabilità sono stati ottenuti all'interno di due modelli diversi, binomiale e normale, che sono due *schematizzazioni* della situazione reale, tra le varie *possibili*. E tali modelli sono stati costruiti sulla base di nostre precise *scelte* sulla situazione reale: non sono la situazione reale. Concludiamo così che nessuno dei due valori è "più giusto" dell'altro e che entrambi rappresentano due possibili soluzioni del problema in esame.

[31] Con Excel si può usare la funzione $DISTRIB.BINOM.N(num\_successi; prove; probabilitá\_s; cumulativo)$ con cumulativo=VERO.