Enunciado

¿Cuánta gente deberá encuestarse en Argentina si se desea conocer dentro de un 1 % la intención de voto a un candidato con un nivel de confianza de 95 %, sabiendo que aproximadamente (a) el 45 % (b) el 5 % del electorado votará por él?

Resolución

Realizar la pregunta ¿Votará usted a este candidato? es un experimento de Bernullí, ya que los resultados posibles son *si* o *no* (1 o 0). El resultado de una encuesta entonces, es decir, el número k de encuestados que votarán por un candidato es una variable aleatoria con distribución binomial. La probabilidad de éxito p de dicha variable aleatoria puede estimarse como $\hat{p}=k/n$, donde n es el número de personas encuestadas. El problema nos pregunta entonces por n bajo la condición de que la intención de voto p se conozca dentro del 1 %. Esto último refiere a que el error pretendido para p es 0.01. Pero además, se desea conocer p con un nivel de confianza del 95 %. Esto es pedir para p un intervalo tal que, por construcción, el 95 % de las veces incluya al valor real de p, es decir, a la verdadera intención de voto.

Para resolver el problema apelaremos al Teorema Central del Límite, y diremos que la variable aleatoria k puede pensarse como la suma de muchas variables aleatorias (cada uno de los 1 y 0 que resultan de los n experimentos de Bernulli realizados). Vale decir entonces que la distribución de k será bien descripta por una distribución Gaussiana y que los parámetros μ y σ de dicha distribución los obtendremos de usar que, en rigor, la distribución de k es binomial. Así es que $\mu = np$ y $\sigma = \sqrt{np(1-p)}$. Y además, gracias a esta aproximación, podemos afirmar que un intervalo de ancho 2σ bastará para cubrir el 95 % de nivel de confianza pedido.

Con esto podemos escribir que:

$$
\begin{aligned}
P(\hat{k} - 2\sigma_{\hat{k}} < &k < \hat{k} + 2\sigma_{k}) = 0{,}95 \\
P(\hat{k}/n - 2\sigma_{\hat{k}}/n < &k/n < \hat{k}/n + 2\sigma_{\hat{k}}/n) = 0{,}95 \\
P(\hat{p} - 2\sigma_{\hat{k}}/n < &p < \hat{p} + 2\sigma_{\hat{k}}/n) = 0{,}95 \\
P(\hat{p} - 2\sqrt{np(1-p)}/n < &p < \hat{p} + 2\sqrt{np(1-p)}/n) = 0{,}95 \\
P(\hat{p} - 2\sqrt{p(1-p)/n} < &p < \hat{p} + 2\sqrt{p(1-p)/n}) = 0{,}95
\end{aligned}
$$

Y ahora pedimos que el error en p sea 0.01, esto es: $2\sigma_p = 0{,}01$

$$
\begin{aligned}
2\sqrt{p(1-p)/n} &= 0{,}01 \\
4p(1-p)/n &= (0{,}01)^2 \\
n &= \frac{4p(1-p)}{(0{,}01)^2} \\
n &= 40000p(1-p)
\end{aligned}
$$

Finalmente, si $p = 0{,}45$ se obtiene $n = 9900$ y si $p = 0{,}05$ se obtiene $n = 1900$.