



DISTRIBUCIÓN EN EL MUESTREO DE UNA PROPORCIÓN

En una población, se desea conocer la proporción p de individuos, siendo la característica X una variable aleatoria tipo Bernoulli (éxito o fracaso). Si la muestra aleatoria es de tamaño n , el estadístico $\hat{p}$ es un estimador de la proporción p .

Distribución: Si el tamaño de la muestra n , es suficientemente grande ($n \geq 30$) y p no toma valores muy próximos a 0 o a 1, la distribución P se puede aproximar por la de una distribución normal:

$$P \in N\left(\mu = p, \sigma = \sqrt{\frac{p \cdot q}{n}}\right)$$

Hipótesis de máxima indeterminación

Si no conocemos p , suponemos que $p = \frac{1}{2}$

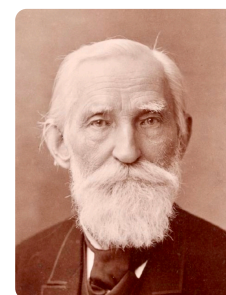
INTERVALO DE CONFIANZA PARA LA PROPORCIÓN (con desviación típica conocida).

Confianza, error y tamaño de la muestra

$$\left(\hat{p} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}, \hat{p} + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}\right)$$

Por tanto, p está en dentro de este intervalo con un nivel de confianza $1 - \alpha$

$$P\left(\hat{p} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}\right) = 1 - \alpha$$



Pafnuty Chebyshev
 Пафну́тий Чебышёв

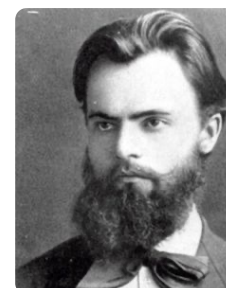
Error máximo admisible

El error máximo admisible en la estimación de la proporción, utilizando el $I.C.$ es su radio

$$E = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

De aquí podemos despejar el tamaño mínimo (n) y el nivel de confianza ($1 - \alpha$).

$$\text{Tamaño mínimo de la muestra (entero): } n \geq \left(\frac{Z_{\alpha/2}}{E}\right)^2 \cdot (\hat{p} \cdot \hat{q})$$



Aleksandr Lyapunov
 Алекса́ндр Ляпуно́в

Ejemplos:

- En las últimas elecciones se ha tomado una muestra de 450 personas a la salida de los colegios electorales y 125 de ellas afirmaron haber votado al partido A. Halla el intervalo de confianza para el porcentaje de votantes con un nivel de confianza del 90 % ?

$$125 \text{ personas de } 450 \rightarrow \hat{p} = \frac{125}{450} = 0,28$$

$$1 - \alpha = 0,90 \rightarrow \alpha = 0,10 \rightarrow z_{\alpha/2} = z_{0,05} = 1,645$$

$$\left(0,28 - 1,645 \cdot \sqrt{\frac{0,28 \cdot 0,72}{450}}; 0,28 + 1,645 \cdot \sqrt{\frac{0,28 \cdot 0,72}{450}} \right) = (0,25; 0,31)$$

Con la muestra elegida, la probabilidad de que la proporción de votantes esté en el intervalo de confianza (0,25; 0,31) es 0,90.

- De una muestra de 500 personas, 325 tienen teléfono móvil. Determina un intervalo de confianza para estimar la proporción de usuarios de teléfono móvil de la población con un nivel de confianza:

a) Del 95 %

b) Del 99 %

$$325 \text{ personas de } 500 \rightarrow \hat{p} = \frac{325}{500} = 0,65$$

$$1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow z_{\alpha/2} = z_{0,025} = 1,96$$

$$\left(0,65 - 1,96 \cdot \sqrt{\frac{0,65 \cdot 0,35}{500}}; 0,65 + 1,96 \cdot \sqrt{\frac{0,65 \cdot 0,35}{500}} \right) = (0,61; 0,69)$$

Con la muestra elegida, la probabilidad de que la proporción de las personas que tiene teléfono móvil esté en el intervalo de confianza (0,61; 0,69) es 0,95.



Jarl Lindeberg

$$\hat{p} = 0,65$$

$$1 - \alpha = 0,99 \rightarrow \alpha = 0,01 \rightarrow z_{\alpha/2} = z_{0,005} = 2,58$$

$$\left(0,65 - 2,58 \cdot \sqrt{\frac{0,65 \cdot 0,35}{500}}; 0,65 + 2,58 \cdot \sqrt{\frac{0,65 \cdot 0,35}{500}} \right) = (0,59; 0,71)$$

Con la muestra elegida, la probabilidad de que la proporción de las personas que tiene teléfono móvil esté en el intervalo de confianza (0,59; 0,71) es 0,99.

- De una muestra de 300 bombillas, 24 han resultado ser defectuosas. Al determinar el intervalo de confianza para la proporción de bombillas defectuosas de la población con un nivel de significación de 0,01, ¿cuál es el error máximo admisible?

$$24 \text{ bombillas defectuosas de } 300 \rightarrow \hat{p} = \frac{24}{300} = 0,08$$

$$E = z_{\alpha/2} \cdot \sqrt{\hat{p} \cdot \hat{q}} = 2,58 \cdot \sqrt{0,08 \cdot 0,92} = 0,04$$

Este resultado indica que, al considerar el 8% como la proporción de bombillas defectuosas, el mayor error que podemos cometer en la estimación es 0,04.



Harald Cramér



4. De una muestra de 1.000 habitantes, 650 leen cierto periódico. ¿Cuál debe ser el tamaño mínimo de una muestra, si se pretende que el error máximo admisible cometido, al estimar la proporción poblacional, sea inferior al 5 %, para un nivel de confianza del 90 % ?

$$650 \text{ habitantes leen cierto periódico de } 1.000 \rightarrow \hat{p} = \frac{650}{1.000} = 0,65$$

$$1 - \alpha = 0,90 \rightarrow \alpha = 0,1 \rightarrow z_{\alpha/2} = z_{0,05} = 1,645$$

$$n = \frac{1,645^2 \cdot 0,65 \cdot 0,35}{0,05^2} = 246,25$$

Para que se cumplan las condiciones, y como n tiene que ser un número entero, debe ser mayor o igual que 247.

5. En una pequeña ciudad se quiere estudiar la incidencia del aumento del paro. Para ello se selecciona una muestra formada por 100 personas, de las que 14 están en situación de desempleo.

a) Determina el intervalo de confianza al 99 % para la proporción poblacional de parados en esta ciudad.

b) Halla el intervalo de confianza al 95 % para la misma proporción, y compáralo con el obtenido en el apartado anterior.

$$\left((23 - 34) - 1,645 \cdot \frac{3^2}{45} + \frac{3^2}{45}, (23 - 34) + 1,645 \cdot \frac{3^2}{45} + \frac{3^2}{45} \right) = (-12,04; -9,96)$$

Calculamos la proporción de parados, $\hat{p}$, $\hat{q}$ de no parados, $\hat{q}$.

$$\hat{p} = \frac{14}{100} = 0,14 \quad \hat{q} = 1 - \hat{p} = 1 - 0,14 = 0,86$$

a) $1 - \alpha = 0,99 \rightarrow \alpha = 0,01 \rightarrow z_{\alpha/2} = z_{0,005} = 2,58$

$$\left(0,14 - 2,58 \cdot \sqrt{\frac{0,14 \cdot 0,86}{100}}, 0,14 + 2,58 \cdot \sqrt{\frac{0,14 \cdot 0,86}{100}} \right) = (0,09; 0,23)$$

b) $1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow z_{\alpha/2} = z_{0,025} = 1,96$

$$\left(0,14 - 1,96 \cdot \sqrt{\frac{0,14 \cdot 0,86}{100}}, 0,14 + 1,96 \cdot \sqrt{\frac{0,14 \cdot 0,86}{100}} \right) = (0,07; 0,21)$$

(26 - 23) - 2,58 · $\frac{4^2}{15}$; (26 - 23) + 2,58 · $\frac{4^2}{15}$ = (-0,52; 6,52)

Al disminuir el nivel de confianza, disminuye la amplitud del intervalo.

6. En un control de calidad se analizó una muestra aleatoria de 750 tornillos, resultando defectuosos 80 de ellos. Hallar un intervalo de confianza para la proporción de tornillos defectuosos en el conjunto de la producción, con el:

a) 95 % de confianza.

b) 99 % de confianza

Calculamos la proporción de tornillos defectuosos, $\hat{p}$, y de tornillos no defectuosos, $\hat{q}$.

$$\hat{p} = \frac{80}{750} = 0,11 \quad \hat{q} = 1 - \hat{p} = 1 - 0,11 = 0,89$$

a) $1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow z_{\alpha/2} = z_{0,025} = 1,96$

$$\left(0,11 - 1,96 \cdot \sqrt{\frac{0,11 \cdot 0,89}{750}}, 0,11 + 1,96 \cdot \sqrt{\frac{0,11 \cdot 0,89}{750}} \right) = (0,09; 0,13)$$

b) $1 - \alpha = 0,99 \rightarrow \alpha = 0,01 \rightarrow z_{\alpha/2} = z_{0,005} = 2,58$

$$\left(0,11 - 2,58 \cdot \sqrt{\frac{0,11 \cdot 0,89}{750}}, 0,11 + 2,58 \cdot \sqrt{\frac{0,11 \cdot 0,89}{750}} \right) = (0,08; 0,14)$$



EJERCICIOS ABAU DISTRIBUCIÓN, IC, TAMAÑO Y NIVEL CONFIANZA PARA LA PROPORCIÓN p

7. En un estanque se desea estimar el porcentaje de peces dorados. Para eso, se toma una muestra aleatoria de 700 peces y se encuentra que exactamente 70 de ellos son dorados.
- Halla, con un nivel de confianza del 99 %, un intervalo para estimar la proporción de peces dorados en el estanque.
 - En el intervalo anterior, ¿cuánto vale el error de estimación?
 - Considerando dicha muestra, ¿qué le ocurriría al error de estimación si aumentase el nivel de confianza? Justifica la respuesta

(Septiembre 2018 Opc B) Solución:

p = proporción peixes dourados

$$n = 700; \hat{p} = \frac{70}{700} = 0,1$$

$$a) \text{ IC para } p: (\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$$

$$1 - \alpha = 0,99 \Rightarrow \alpha = 0,01 \Rightarrow \alpha/2 = 0,005 \Rightarrow 1 - \alpha/2 = 0,995$$

$$z_{\alpha/2} = 2,575 \begin{cases} 2,57 \\ 2,58 \end{cases}$$

$$L_1 = 0,1 - 2,575 \sqrt{\frac{0,1 \times 0,9}{700}} = 0,1 - 2,575 \times 0,011339 = 0,0708$$

$$L_2 = 0,1 + 2,575 \times 0,011339 = 0,1292$$

O intervalo de confianza para a proporción e $IC(0,0708, 0,1292)$
 7,07% 12,92%

A un nivel de confianza do 99% a proporción de peixes dourados estará entre 7,07% e 12,92%

$$b) e = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2,575 \times 0,011339 = 0,0292 \rightarrow e = 2,92\%$$

$$c) \text{ n.c.} = 1 - \alpha = 0,99 \Rightarrow z_{\alpha/2} > 2,575 \Rightarrow e = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ aumenta.}$$

(Ou ben calculando de novo o valor do erro)



8. Como resultado de una encuesta en la que se utilizó el supuesto de máxima indeterminación ($p = 1 - p = 1/2$) se afirma que, con un 97,56 % de confianza, el porcentaje de individuos de una población que considera el alcohol y/o las drogas como causa principal de los accidentes de tráfico, está entre e

Exemplos de resposta / Solucións

- a) Calcula el número de individuos de esa población a los que se les realizó la encuesta.
 b) De los que se les realizó la encuesta, ¿cuántos contestaron que la causa principal de los accidentes es el alcohol y/o las drogas?

(Junio 2017 Opc B) Solución:

(a) **1'25 puntos.** Calcula o número de individuos desa poboación aos que se lles realizou a enquisa.

Sexa

" p : proporción de individuos da poboación, que considera o alcol e/ou as drogas como causa principal dos accidentes de tráfico". **Parámetro poboacional descoñecido**

- Como resultado da enquisa na que se utilizou o suposto de máxima indeterminación, temos que o máximo erro que cometeron na estimación do intervalo ven dado por:

$$E = z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} = z_{\alpha/2} \sqrt{\frac{1}{4n}}$$

- Como consecuencia do intervalo que obtiveron (baixo ese principio), o radio do intervalo foi $\frac{0'625 - 0'575}{2} = 0'025$ é dicir, que cometeron un erro na estimación dun 2,5% **0'25 puntos.**
 – Calculamos, para unha confianza do 97'56%, $z_{\alpha/2} = z_{0'0122} = 2'25$ **0'25 puntos.**

- Resolver $z_{\alpha/2} \sqrt{\frac{1}{4n}} = 2'25 \sqrt{\frac{1}{4n}} = 0'025 \Rightarrow n = 2025$ **0'50 puntos.**

- Responder no contexto do exercicio:

"A enquisa realizóuselles a 2025 individuos desa poboación" **0'25 puntos.**

- (b) **0'75 puntos.** Dos que se lles realizou a enquisa, ¿cantos contestaron que a causa principal dos accidentes é o alcol e/ou as drogas?

p : proporción de individuos da poboación, que considera o alcol e / ou as drogas como causa principal dos accidentes de tráfico

$\downarrow n = 2025$

X : número de individuos, en mostras de 2025 individuos, que considera o alcol e / ou as drogas como causa principal dos accidentes de tráfico

valor particular para a mostra dada $\rightarrow x$

$\hat{P}$: proporción de individuos, en mostras de 2025 individuos, que considera o alcol e / ou as drogas

como causa principal dos accidentes de tráfico

valor particular para a mostra dada $\rightarrow \hat{p} = \frac{x}{2025}$

- Calcular o valor da proporción muestral $\hat{p} = \frac{0'575 + 0'625}{2} = 0'6$ **0'25 puntos.**
 – Calcular o 60% de 2025 = 1215, ou así $\hat{p} = \frac{x}{n} = \frac{x}{2025} = 0'6 \Rightarrow x = 1215$ **0'25 puntos.**

"Contestaron que o alcol e/ou as drogas son a causa principal dos accidentes de tráfico 1215 individuos, da mostra de 2025 enquisados" **0'25 puntos.**



S	$\bar{S}$
0'32	0'68
0'54	0'46
0'83	0'17
	1



9. Se realiza un estudio acerca de la proporción de universitarios gallegos que practican algún deporte. Elegida una muestra aleatoria de 200 universitarios gallegos se comprobó que 146 de ellos practican algún deporte. A partir de la muestra dada, calcula un intervalo del 95 % de confianza para la proporción.

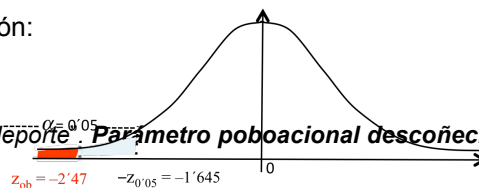
Exemplos de resposta / Solucións

(Septiembre 2016 Opc A, Ej 4, sólo el apartado b.) Solución:

para $\alpha = 0'05$, $(-\infty, -1'645)$

Sexan

" p : proporción de universitarios galegos que practican algún deporte" **Parámetro poboacional descoñecido**
(é α que nos mandan contrastar) $H_0: p = 0'8$



Estatístico proporción mostral $\equiv \hat{P}$: proporción de universitarios galegos que practican algún deporte,

en mostras de 200 universitarios galegos $\rightarrow \hat{p} = \frac{146}{200} = 0'73$
valor particular do estatístico $\hat{P}$, para a mostra dada

- (b) 0'75 puntos. A partir da mostra dada, calcula un intervalo do 95% de confianza para a proporción de universitarios galegos que practican algún deporte. Interpreta o intervalo obtido.

$H_1: p < 0'8$

Calcular numericamente os extremos do intervalo

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0'73 - 1'96 \cdot \sqrt{\frac{0'73 \cdot 0'27}{200}} = 0'73 - 0'061 = 0'669 \approx 0'67 \quad \text{0'25 puntos}$$

$$\hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0'73 + 1'96 \cdot \sqrt{\frac{0'73 \cdot 0'27}{200}} = 0'73 + 0'061 = 0'791 \approx 0'79 \quad \text{0'25 puntos}$$

$$z_{ob} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = -2'47$$

Interpretación

"Estimamos que a porcentaxe de universitarios galegos que practican algún deporte, está entre un 67% e un 79%, aproximadamente, cun 95% de confianza" 0'25 puntos.