

4

Inferencia estadística

Tema 4.1 · Inferencia Estadística

Inferencia estadística: estimación puntual e intervalar. Distribución de la media muestral. Cómo se saca el valor z de cualquier nivel de confianza con la tabla de la normal. Intervalos de confianza para la media (σ conocida) y para la proporción. Tamaño de la muestra en función del margen de error y el nivel de confianza. Diez ejercicios resueltos al estilo PAU.

Temas del bloque 4

4.1 Inferencia Estadística (este apunte)

Muestras, distribución de la media muestral, el valor z de cualquier nivel de confianza, intervalos para la media y la proporción, tamaño de la muestra.

Apuntes de preparación para la PAU · 2.º de Bachillerato

Ejemplar de muestra gratuito, igual para todos. Prohibida su venta o redistribución sin permiso escrito del titular.

© 2026 PAUmode. Todos los derechos reservados. Queda prohibida la reproducción total o parcial, la distribución, la comunicación pública y la transformación de esta obra sin autorización escrita del titular. El incumplimiento podrá dar lugar a las acciones civiles y penales previstas en la legislación sobre propiedad intelectual.

4.1 Inferencia Estadística

1.1 Matrices · 1.2 Sistemas y PL · 1.3 Cramer/Inecuaciones · 2.1 Integrales · 2.2 Derivadas · 2.3 Funciones · 3.1 Probabilidad · 3.2 Distribuciones · **4.1 Inferencia.**

4.1.1 · Muestras y estadísticos

La **inferencia estadística** permite extraer conclusiones sobre una **población** a partir de una **muestra aleatoria**. La población tiene parámetros desconocidos (media μ , proporción p); la muestra permite calcular **estadísticos** (media muestral $\bar{x}$, proporción muestral $\hat{p}$) que estiman esos parámetros. Dar directamente el estadístico como valor del parámetro es la **estimación puntual**; darlo dentro de un rango es la **estimación por intervalo**, y es la que interesa, porque la puntual no dice nada del error que se comete. En la PAU de CCSS el objetivo es construir un **intervalo de confianza** (IC): un rango de valores que contiene el parámetro con una probabilidad determinada.

Vocabulario esencial:

Población: conjunto completo de individuos (tamaño N , parámetros μ , p).

Muestra: subconjunto extraído al azar (tamaño n , estadísticos $\bar{x}$, $\hat{p}$).

Media muestral: $\bar{x} = (x_1 + x_2 + \dots + x_n) / n$

Proporción muestral: $\hat{p} = \text{número de casos favorables} / n$

Distribución en el muestreo (Teorema Central del Límite):

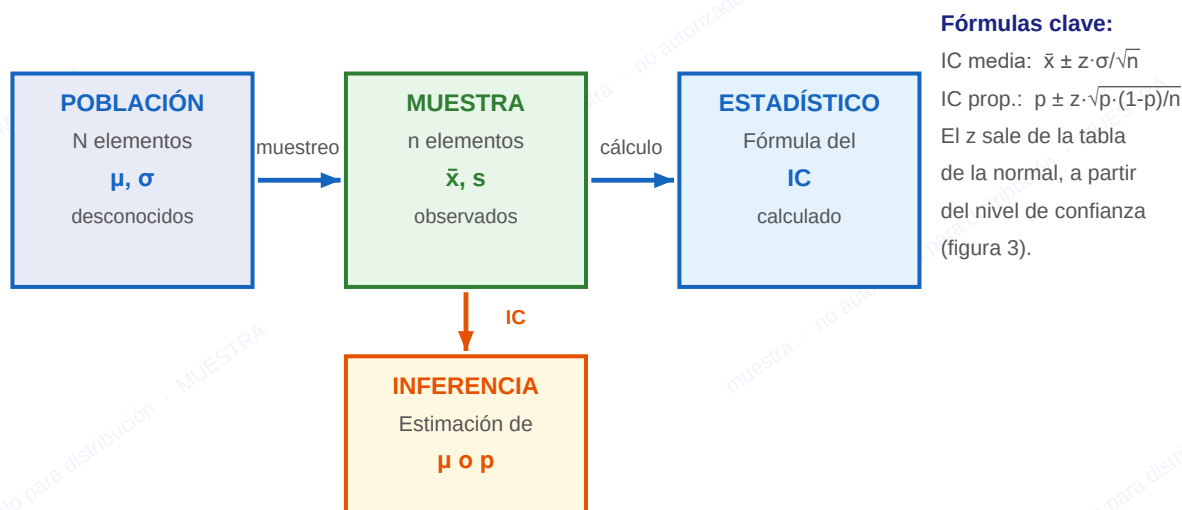
Si n es grande ($n \geq 30$), $\bar{x}$ sigue aproximadamente $N(\mu, \sigma/\sqrt{n})$.

El error típico de la media es: $s_x = \sigma / \sqrt{n}$

El error típico de la proporción: $s_p = \sqrt{p \cdot (1-p)/n}$

Nivel de confianza. El nivel de confianza $(1 - \alpha) \cdot 100\%$ indica la frecuencia con la que el IC contendría el parámetro verdadero si se repitiera el muestreo muchas veces. Un IC al 95 % no significa que haya un 95 % de probabilidad de que μ esté en ese intervalo concreto: μ es fijo; lo que varía es el intervalo de muestra en muestra.

Esquema de muestreo e inferencia estadística



La inferencia estima parámetros poblacionales a partir de la muestra.

Figura 1 - Esquema de muestreo e inferencia. A partir de la población (parámetros desconocidos μ y p) se extrae una muestra de tamaño n . Con los datos muestrales se calcula el estadístico (media o proporción) y se construye el intervalo de confianza para estimar el parámetro poblacional.

4.1.2 · La distribución de la media muestral

Si de una población se extraen muchas muestras del mismo tamaño n , cada una da una media distinta. Esas medias también forman una distribución, y es la pieza sobre la que se apoya todo lo demás. Cuando la población es normal $N(\mu, \sigma)$, o cuando la muestra es grande ($n \geq 30$), la media muestral sigue una normal centrada en la misma μ pero **mucho menos dispersa**: su desviación típica es σ dividido entre la raíz de n .

Distribución de la media muestral:

$\bar{x}$ sigue $N(\mu, \sigma / \sqrt{n})$

$\sigma / \sqrt{n}$ se llama error típico de la media.

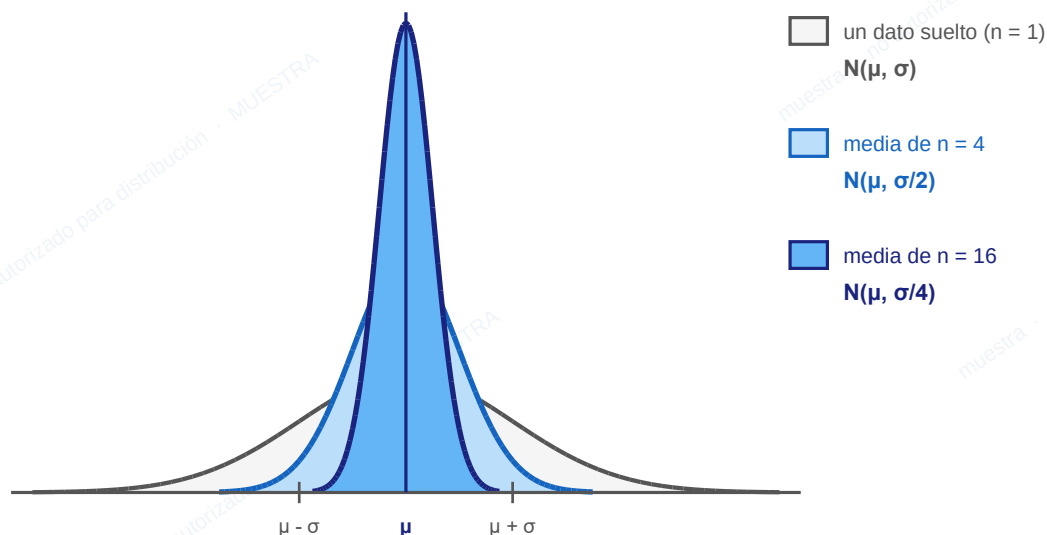
Para calcular una probabilidad sobre la media de una muestra se tipifica igual que siempre, pero dividiendo entre el error típico y no entre σ :

$$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

La proporción muestral funciona igual:

$\hat{p}$ sigue $N(p, \sqrt{p \cdot (1-p)/n})$

Al crecer el tamaño de la muestra, la campana se estrecha



Error típico de la media: $\sigma/\sqrt{n}$. $n = 4 \rightarrow \sigma/2$; $n = 16 \rightarrow \sigma/4$; $n = 100 \rightarrow \sigma/10$.

Las tres encierran la misma área: al estrecharse, se hacen más altas. Por eso n grande da un error pequeño.

Figura 2 - Un dato suelto frente a las medias de muestras de $n = 4$ y de $n = 16$. Las tres campanas están centradas en la misma μ y encierran la misma área, pero la de las medias es cada vez más estrecha y más alta: promediar cuatro datos reduce la dispersión a la mitad y promediar dieciséis la reduce a la cuarta parte. Ese estrechamiento es lo que permite estimar μ con un margen de error pequeño.

Error típico. Tipificar con σ cuando la pregunta es sobre una media. Si el enunciado pregunta por **un** individuo se divide entre σ ; si pregunta por la **media de una muestra de n** se divide entre $\sigma/\sqrt{n}$, que es un número más pequeño. Olvidar la raíz de n da probabilidades muy distintas de las correctas.

4.1.3 · Intervalo de confianza para la media

Cuando se conoce la desviación típica poblacional σ y la muestra es suficientemente grande ($n \geq 30$), la media muestral $\bar{x}$ sigue una distribución normal y el intervalo de confianza al nivel $(1 - \alpha) \cdot 100\%$ es:

IC para la media μ (σ conocida, n grande):

$$IC = (\bar{x} - z \cdot \sigma/\sqrt{n}; \bar{x} + z \cdot \sigma/\sqrt{n})$$

Forma abreviada: $\bar{x} \pm e$ donde $e = z \cdot \sigma/\sqrt{n}$

e = semiamplitud (margen de error o radio del intervalo)

Amplitud total del IC = $2 \cdot e$

El único dato que cambia de un nivel de confianza a otro es z , y ese z no hay que aprenderse: se saca de la tabla de la normal que se entrega con el examen. El procedimiento es siempre el mismo, valga el nivel lo que valga:

Cómo se obtiene z de cualquier nivel de confianza:

- 1. Del nivel de confianza al riesgo: $\alpha = 1 - \text{nivel}$.**
- 2. Se reparte el riesgo entre las dos colas: $\alpha/2$ en cada una.**
- 3. Se calcula el área acumulada: $1 - \alpha/2$.**
- 4. Se busca ese número DENTRO de la tabla, no en el borde.**
- 5. z es la suma de la fila y la columna donde aparece.**

Ejemplo con el 97 % (nivel que pide el modelo oficial de la prueba):

$$\alpha = 1 - 0,97 = 0,03$$

$$\alpha/2 = 0,015 \text{ y } 1 - \alpha/2 = 0,9850$$

$$0,9850 \text{ está en la fila } 2,1 \text{ y la columna } 0,07 \Rightarrow z = 2,17$$

**Si el número exacto no está en la tabla se toma el valor más cercano
(los criterios oficiales admiten esto y también interpolar).**

Error típico. Confundir σ con s (desviación muestral). En estos ejercicios se trabaja con σ (poblacional); si el enunciado da s , se usa como aproximación de σ cuando $n \geq 30$. No uses la fórmula t de Student: en este temario siempre se usa z .

De un nivel de confianza cualquiera a su valor z

Ejemplo: nivel de confianza del 97 %

- 1) $\alpha = 1 - 0,97 = 0,03$
- 2) $1 - \alpha/2 = 1 - 0,015 = 0,9850$
- 3) se busca 0,9850 DENTRO de la tabla
- 4) **fila 2,1 y columna 0,07 $\Rightarrow z = 2,17$**

z	0,05	0,06	0,07	0,08
2,0	0,9798	0,9803	0,9808	0,9812
2,1	0,9842	0,9846	0,9850	0,9854
2,2	0,9878	0,9881	0,9884	0,9887

La fila da el 2,1 y la columna el 0,07: el cruce es 0,9850

Trozo de la tabla que se entrega en el examen

Algunos niveles, leídos así en la tabla:

90 % $\rightarrow z = 1,645$	92 % $\rightarrow z = 1,75$	93 % $\rightarrow z = 1,81$	94 % $\rightarrow z = 1,88$	95 % $\rightarrow z = 1,96$
96 % $\rightarrow z = 2,05$	97 % $\rightarrow z = 2,17$	98 % $\rightarrow z = 2,33$	99 % $\rightarrow z = 2,576$	

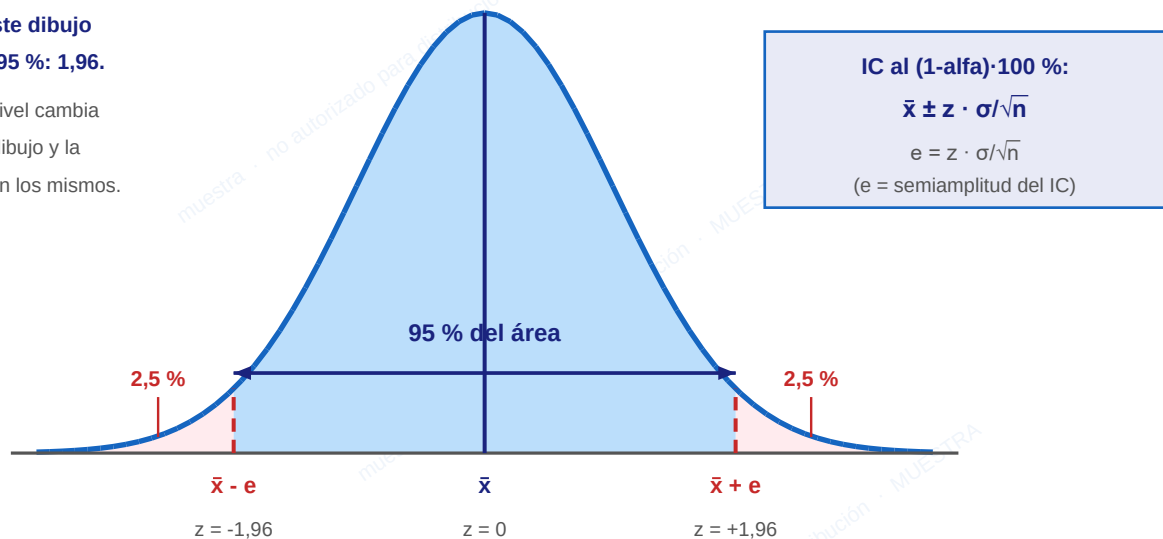
1,645 es el punto medio entre 1,64 y 1,65. 2,576 es el valor exacto (2,5758) redondeado; la media de 2,57 y 2,58 sería 2,575.

Figura 3 - De un nivel de confianza cualquiera a su valor z. A la izquierda, el camino paso a paso con el 97 % como ejemplo; a la derecha, el trozo de la tabla oficial donde se lee 0,9850. Abajo, nueve niveles de confianza con el z que les corresponde.

Intervalo de confianza para la media — nivel del 95 %

El z de este dibujo es el del 95 %: 1,96.

Con otro nivel cambia solo z; el dibujo y la fórmula son los mismos.



2,5 % (cola roja) + 95 % (zona azul) + 2,5 % (cola roja) = 100 %. Las dos colas juntas son el riesgo $\alpha = 5$ %.

Figura 4 - Intervalo de confianza al 95 % sobre la campana normal. El área azul sombreada (95 %) corresponde al IC. Las dos colas (2,5 % cada una) representan el riesgo $\alpha = 5$ % de que el IC no contenga la media real. Los límites del IC se obtienen con $\bar{x} \pm 1,96 \cdot \sigma/\sqrt{n}$.

Truco. Solo hay tres valores de z que conviene llevar aprendidos, porque la tabla no los da exactos: **1,645** (90 %), **1,96** (95 %) y **2,576** (99 %). Cualquier otro nivel se lee en la tabla en veinte segundos, y el examen puede pedir cualquiera: el 97 %, el 98 %, el 96 %, el 94 %, el 92 %... El modelo oficial de la prueba pide un intervalo al 97 % y un tamaño de muestra al 92 %, así que dar por hecho que solo caen tres niveles deja el ejercicio en blanco.

4.1.4 · Intervalo de confianza para la proporción

Cuando se estudia una característica cualitativa (porcentaje de votos, tasa de aprobados, fracción de piezas defectuosas), el parámetro de interés es la **proporción poblacional** p . Su estimador es la proporción muestral $\hat{p} = k/n$, donde k es el número de individuos con la característica en la muestra. Con n grande ($n \cdot \hat{p} \geq 5$ y $n \cdot (1 - \hat{p}) \geq 5$), se aplica la aproximación normal:

IC para la proporción p (n grande):

$$\text{IC} = (\hat{p} - z \cdot \sqrt{\hat{p} \cdot (1 - \hat{p}) / n} ; \\ \hat{p} + z \cdot \sqrt{\hat{p} \cdot (1 - \hat{p}) / n})$$

$$\text{Semiamplitud: } e = z \cdot \sqrt{\hat{p} \cdot (1 - \hat{p}) / n}$$

Procedimiento PAU:

1. Calcula $\hat{p} = k / n$.
2. Saca z del nivel de confianza con la tabla (apartado 4.1.3).
3. Calcula $e = z \cdot \sqrt{\hat{p} \cdot (1 - \hat{p}) / n}$.
4. $\text{IC} = (\hat{p} - e ; \hat{p} + e)$.
5. Comprueba que los dos extremos estén entre 0 y 1.

Interpretación del IC para la proporción. El IC (0,58 ; 0,72) para la proporción de votos significa que, con el nivel de confianza elegido, estimamos que entre el 58 % y el 72 % de la población votaría a ese partido. La anchura del IC depende de n : a mayor muestra, IC más estrecho.

Error típico. No confundir $\hat{p}$ (estadístico muestral observado) con p (parámetro poblacional desconocido). En el IC para la proporción usamos $\hat{p}$ en la fórmula porque p es lo que queremos estimar.

4.1.5 · Tamaño mínimo de la muestra

Antes de realizar una encuesta o estudio, puede interesar calcular el **tamaño mínimo de muestra** n que garantiza un margen de error máximo e con un nivel de confianza dado. Despejando n de la fórmula de la semiamplitud:

Tamaño mínimo de muestra para estimar la MEDIA:

$$n \geq (z \cdot \sigma / e)^2$$

n = número de individuos a encuestar (redondear SIEMPRE hacia arriba).

e = margen de error deseado.

σ = desviación típica poblacional conocida o estimada.

Tamaño mínimo de muestra para estimar la PROPORCIÓN:

$$n \geq z^2 \cdot p \cdot (1-p) / e^2$$

Si p es desconocida, usar $p = 0,5$ (caso más desfavorable):

$$n \geq z^2 \cdot 0,25 / e^2$$

Esto garantiza que el IC tenga semiamplitud como máximo e , sea cual sea el verdadero valor de p .

Truco. El tamaño de la muestra **siempre se redondea hacia arriba**, aunque la parte decimal sea mínima. Si sale $n \geq 76,5625$ la respuesta es $n = 77$, y si sale $n \geq 1067,1$ la respuesta es $n = 1068$. El motivo es que la fórmula dice « n como poco tanto»: con 76 individuos el margen de error se pasaría del que pide el enunciado, y el apartado se da por fallado. Es un punto fácil de perder, y se pierde por redondear como se redondea siempre.

Comparativa: IC para la media frente a IC para la proporción

Característica	IC para la Media	IC para la Proporción
Parámetro	μ (media poblacional)	p (proporción poblacional)
Estimador	$\bar{x}$ (media muestral)	$\hat{p}$ (proporción muestral)
Fórmula IC	$\bar{x} \pm z \cdot \sigma/\sqrt{n}$	$\hat{p} \pm z \cdot \sqrt{\hat{p}(1-\hat{p})/n}$
Semiamplitud e	$e = z \cdot \sigma/\sqrt{n}$	$e = z \cdot \sqrt{\hat{p}(1-\hat{p})/n}$
Tamaño de muestra	$n \geq (z \cdot \sigma / e)^2$	$n \geq z^2 \cdot p(1-p) / e^2$
Caso peor (n)	Requiere σ conocida	$p = 0,5$ maximiza $p(1-p) = 0,25$

En las dos columnas, el z es el mismo: el que corresponda al nivel de confianza del enunciado.

Figura 5 - Comparativa entre el IC para la media y el IC para la proporción. Ambos siguen la estructura "estimador $\pm z$ -error típico". La diferencia está en el error típico: $\sigma/\sqrt{n}$ para la media, $\sqrt{\hat{p}(1-\hat{p})/n}$ para la proporción.

PAU - Atención. En la PAU de CCSS un problema de inferencia puede pedir: a) calcular el IC dado $\bar{x}$ o $\hat{p}$, n , σ o s , y el nivel de confianza; b) interpretar el IC en el contexto del problema; c) calcular n mínimo para un error e dado. Pueden venir los tres en un mismo ejercicio, y el nivel de confianza no tiene por qué ser del 90 %, del 95 % o del 99 %: el modelo oficial de la prueba pide un intervalo al 97 % y un tamaño de muestra al 92 %. Antes de nada, saca el z del nivel que ponga el enunciado.

Ejercicios - Sección 4.1

Diez ejercicios al estilo PAU: los cuatro primeros sobre IC para la media, los tres siguientes sobre IC para la proporción, y los tres últimos sobre tamaño de la muestra y problemas completos. Tres de ellos usan niveles de confianza que no son los tres de siempre (97 %, 96 % y 92 %), porque son los que se olvidan y el examen también los pide.

1. Se toma una muestra de 100 estudiantes y se obtiene una nota media de 75 puntos. La desviación típica poblacional es $\sigma = 12$. Calcula el intervalo de confianza al 95 % para la nota media de todos los estudiantes.

IC para la media: el enunciado da σ (poblacional) y la muestra es grande ($n = 100$), que son las dos condiciones para usar z .

Datos: $\bar{x} = 75$, $\sigma = 12$, $n = 100$. Se usa el IC para la media porque lo que se estima es una nota media, no un porcentaje.

Primero el z, que sale del nivel de confianza y de la tabla:

$$\alpha = 1 - 0,95 = 0,05, \text{ y } 1 - \alpha/2 = 1 - 0,025 = 0,9750$$

0,9750 está en la tabla en la fila 1,9 y la columna 0,06 $\Rightarrow z = 1,96$

Ahora el margen de error e, que es lo que se suma y se resta a la media:

$$e = z \cdot \sigma / \sqrt{n} = 1,96 \cdot 12 / \sqrt{100} = 1,96 \cdot 12 / 10$$

$$e = 1,96 \cdot 1,2 = 2,352$$

El intervalo es la media muestral más y menos ese margen:

$$IC = (\bar{x} - e ; \bar{x} + e) = (75 - 2,352 ; 75 + 2,352)$$

$$IC = (72,648 ; 77,352)$$

Interpretación: con una confianza del 95 %, la nota media de todos los estudiantes está entre 72,65 y 77,35 puntos.

Resultado: IC al 95 % para la media: (72,648 ; 77,352)

2. Una muestra de 64 personas da una estatura media de 180 cm. La desviación típica de la población es $\sigma = 15$ cm. Calcula el IC al 99 % para la estatura media poblacional.

Mismo esquema que el anterior, pero al 99 %: sirve para ver que subir la confianza ensancha el intervalo.

Datos: $\bar{x} = 180$, $\sigma = 15$, $n = 64$.

El z del 99 % es uno de los tres que conviene llevar aprendidos, porque la tabla no lo da exacto:

$$\alpha = 1 - 0,99 = 0,01, \text{ y } 1 - \alpha/2 = 0,9950$$

En la tabla, 2,57 da 0,9949 y 2,58 da 0,9951: 0,9950 cae en medio

El valor exacto es 2,5758, que redondeado da $z = 2,576$

Margen de error. La raíz de 64 es exacta, así que sale un número redondo:

$$\sqrt{64} = 8$$

$$e = 2,576 \cdot 15 / 8 = 38,64 / 8 = 4,83$$

$$IC = (180 - 4,83 ; 180 + 4,83) = (175,17 ; 184,83)$$

Con los mismos datos, al 95 % el margen sería $1,96 \cdot 15 / 8 = 3,675$, más pequeño. Subir la confianza sube z , sube el error y ensancha el intervalo: se gana seguridad y se pierde precisión.

Resultado: IC al 99 % para la media: (175,17 ; 184,83)

3. El tiempo medio de espera en una fila, medido en 36 observaciones, es $\bar{x} = 5,2$ minutos. La desviación típica es $\sigma = 0,8$ minutos. Calcula el IC al 97 % para la media poblacional.

Nivel de confianza que no es ninguno de los tres de siempre: hay que ir a la tabla a buscar el z . Es el nivel que pide el modelo oficial de la prueba.

Datos: $\bar{x} = 5,2$, $\sigma = 0,8$, $n = 36$.

El 97 % no viene aprendido de casa, así que se saca de la tabla paso a paso:

$$\alpha = 1 - 0,97 = 0,03, \text{ luego } \alpha/2 = 0,015$$

$$1 - \alpha/2 = 1 - 0,015 = 0,9850$$

Se busca 0,9850 DENTRO de la tabla: aparece en la fila 2,1 y la columna 0,07

$$z = 2,1 + 0,07 = 2,17$$

Con el z ya se hace lo mismo que en cualquier otro nivel:

$$\sqrt{36} = 6$$

$$e = 2,17 \cdot 0,8 / 6 = 1,736 / 6 \text{ aprox. } 0,2893$$

$$\text{IC} = (5,2 - 0,2893 ; 5,2 + 0,2893)$$

$$\text{IC} = (4,911 ; 5,489)$$

El z del 97 % (2,17) queda entre el del 95 % (1,96) y el del 99 % (2,576), que es justo donde tiene que estar. Ese es el modo rápido de comprobar que no se ha leído mal la tabla.

Resultado: IC al 97 % para la media: (4,911 ; 5,489) minutos

4. Una empresa mide el consumo eléctrico mensual de 144 hogares y obtiene $\bar{x} = 250$ kWh con $\sigma = 30$ kWh. a) Calcula el IC al 95 %. b) ¿Cuál es la amplitud del intervalo? c) ¿Cuánto mediría el intervalo si la confianza bajase al 90 %?

La amplitud es $2 \cdot e$, no e . Y el apartado c) es una pregunta de razonar: cómo afecta el nivel de confianza a la anchura del intervalo.

Datos: $\bar{x} = 250$, $\sigma = 30$, $n = 144$. La raíz de 144 es 12.

a) Al 95 % el z es 1,96 ($1 - \alpha/2 = 0,9750$ en la tabla):

$$e = 1,96 \cdot 30 / 12 = 58,8 / 12 = 4,9$$

$$IC = (250 - 4,9 ; 250 + 4,9) = (245,1 ; 254,9) \text{ kWh}$$

b) La amplitud es el intervalo entero, de un extremo al otro, y el margen e es solo la mitad:

$$\text{Amplitud} = 2 \cdot e = 2 \cdot 4,9 = 9,8 \text{ kWh}$$

c) Al 90 % el z baja a 1,645, y todo lo demás se queda igual:

$$e = 1,645 \cdot 30 / 12 = 49,35 / 12 = 4,1125$$

$$\text{Amplitud} = 2 \cdot 4,1125 = 8,225 \text{ kWh}$$

$$IC = (245,8875 ; 254,1125) \text{ kWh}$$

Bajar la confianza del 95 % al 90 % estrecha el intervalo de 9,8 a 8,225 kWh. Con la misma muestra no se puede tener más confianza y más precisión a la vez: para ganar las dos cosas hay que aumentar n.

Resultado: a) IC = (245,1 ; 254,9) kWh b) amplitud = 9,8 kWh c) al 90 %, amplitud = 8,225 kWh

5. En una muestra de 400 personas, 240 declaran usar transporte público a diario. Calcula el IC al 95 % para la proporción de usuarios de transporte público.

Aquí lo que se estima es un porcentaje, no una media: cambia el error típico, que pasa a ser $\sqrt{\hat{p} \cdot (1-\hat{p})/n}$.

Lo primero es la proporción muestral, que es el estimador:

$$\hat{p} = k / n = 240 / 400 = 0,60$$

Se comprueba que la muestra da para usar la normal:

$$n \cdot \hat{p} = 400 \cdot 0,60 = 240 \geq 5 \text{ y } n \cdot (1-\hat{p}) = 400 \cdot 0,40 = 160 \geq 5$$

El z del 95 % es 1,96. Ahora el error típico de la proporción:

$$\sqrt{\hat{p} \cdot (1-\hat{p})/n} = \sqrt{0,60 \cdot 0,40 / 400} = \sqrt{0,0006} \text{ aprox. } 0,02449$$

$$e = 1,96 \cdot 0,02449 \text{ aprox. } 0,0480$$

$$IC = (0,60 - 0,048 ; 0,60 + 0,048) = (0,552 ; 0,648)$$

Los dos extremos están entre 0 y 1, como tiene que ser en una proporción. Interpretación: con un 95 % de confianza, entre el 55,2 % y el 64,8 % de la población usa transporte público a diario.

Resultado: IC al 95 % para la proporción: (0,552 ; 0,648)

6. Una encuesta a 200 alumnos revela que 70 suspendieron una asignatura. Calcula el IC al 99 % para la proporción de alumnos que suspenden.

IC para la proporción al 99 %. Con una muestra pequeña y una confianza alta sale un intervalo muy ancho, y eso también hay que saber decirlo.

Proporción muestral:

$$\hat{p} = 70 / 200 = 0,35$$

Condición para poder usar la normal:

$$n \cdot \hat{p} = 70 \geq 5 \text{ y } n \cdot (1 - \hat{p}) = 130 \geq 5$$

El z del 99 % es 2,576. Error típico y margen de error:

$$\sqrt{0,35 \cdot 0,65 / 200} = \sqrt{0,0011375} \text{ aprox. } 0,03373$$

$$e = 2,576 \cdot 0,03373 \text{ aprox. } 0,0869$$

$$IC = (0,35 - 0,0869 ; 0,35 + 0,0869)$$

$$IC = (0,263 ; 0,437)$$

Con el 99 % de confianza, la proporción real de suspensos está entre el 26,3 % y el 43,7 %. Es un intervalo de 17 puntos de ancho: con 200 alumnos y tanta confianza, la estimación dice poco. Para estrecharlo hay que encuestar a más gente.

Resultado: IC al 99 % para la proporción: (0,263 ; 0,437)

7. En una muestra de 500 empresas, 360 tienen página web propia. Calcula el IC al 96 % para la proporción de empresas con página web.

Otro nivel de confianza que no es de los tres de siempre, esta vez en una proporción. El z se busca igual: el nivel solo cambia ese número.

Proporción muestral:

$$\hat{p} = 360 / 500 = 0,72$$

El z del 96 %, con los mismos pasos de siempre:

$$\alpha = 1 - 0,96 = 0,04, \text{ y } 1 - \alpha/2 = 1 - 0,02 = 0,9800$$

En la tabla no está 0,9800 exacto: lo más cercano es 0,9798, en la fila 2,0 y la columna 0,05 $\Rightarrow z = 2,05$

Error típico de la proporción y margen de error:

$$\sqrt{0,72 \cdot 0,28 / 500} = \sqrt{0,0004032} \text{ aprox. } 0,02008$$

$$e = 2,05 \cdot 0,02008 \text{ aprox. } 0,0412$$

$$IC = (0,72 - 0,0412 ; 0,72 + 0,0412)$$

$$IC = (0,679 ; 0,761)$$

Con el 96 % de confianza, entre el 67,9 % y el 76,1 % de las empresas tiene página web propia. Coger el valor más cercano de la tabla es lo que admiten los criterios de corrección; interpolar también vale y da casi lo mismo.

Resultado: IC al 96 % para la proporción: (0,679 ; 0,761)

8. Se quiere estimar la media de una variable con $\sigma = 10$, con un margen de error máximo de $e = 2$ y un nivel de confianza del 92 %. ¿Cuál es el tamaño mínimo de muestra necesario?

Tamaño de muestra al 92 %, que es el nivel que pide el modelo oficial de la prueba en este apartado. Y el redondeo va SIEMPRE hacia arriba.

Datos: $\sigma = 10$, $e = 2$. La incógnita es n .

El z del 92 % se busca en la tabla como cualquier otro:

$$\alpha = 1 - 0,92 = 0,08, \text{ y } 1 - \alpha/2 = 1 - 0,04 = 0,9600$$

Lo más cercano en la tabla es 0,9599, en la fila 1,7 y la columna 0,05 $\Rightarrow z = 1,75$

La fórmula sale de despejar n en $e = z \cdot \sigma / \sqrt{n}$:

$$n \geq (z \cdot \sigma / e)^2 = (1,75 \cdot 10 / 2)^2$$

$$n \geq (8,75)^2 = 76,5625$$

El resultado dice « n como poco 76,5625», así que se redondea hacia arriba: con 76 individuos el margen de error se pasaría de 2.

$$n = 77$$

Aunque hubiera salido 76,01 la respuesta seguiría siendo 77. En el tamaño de muestra nunca se redondea al entero más próximo: siempre hacia arriba.

Resultado: Tamaño mínimo de muestra: $n = 77$

9. Se quiere estimar la proporción de votantes de un partido con un margen de error máximo del 3 % y un nivel de confianza del 95 %. ¿Cuántas personas hay que encuestar como mínimo? No se conoce la proporción de antemano.

Cuando no se sabe nada de p hay que ponerse en el caso peor, $p = 0,5$, que es el que hace más grande el producto $p \cdot (1-p)$ y por tanto más grande la muestra.

Datos: $e = 0,03$ (el 3 % escrito como proporción), $z = 1,96$ (nivel del 95 %).

Como p es desconocida se toma $p = 0,5$, porque $p \cdot (1-p)$ es máximo ahí: $0,5 \cdot 0,5 = 0,25$. Cualquier otro valor de p daría una muestra más pequeña, así que con esta el error se cumple seguro.

$$n \geq z^2 \cdot 0,25 / e^2$$

$$n \geq (1,96)^2 \cdot 0,25 / (0,03)^2$$

$$n \geq 3,8416 \cdot 0,25 / 0,0009$$

$$n \geq 0,9604 / 0,0009 \text{ aprox. } 1067,1$$

Redondeo hacia arriba, como siempre en tamaños de muestra:

$$n = 1068$$

Con 1068 encuestados el margen de error no pasa del 3 % con un 95 % de confianza, sea cual sea la proporción real. Es más o menos el tamaño de las encuestas electorales que se publican, y no es casualidad.

Resultado: Tamaño mínimo de muestra: $n = 1068$

10. En una muestra de 625 alumnos de bachillerato, 400 aprueban la PAU. a) Calcula el IC al 95 % para la proporción de aprobados. b) ¿Cuántos alumnos habría que encuestar para reducir el margen de error al 2 % con el mismo nivel de confianza? c) La edad media de la muestra es $\bar{x} = 22$ años con $\sigma = 4$. Calcula el IC al 90 % para la edad media.

Problema de tres apartados como los de examen: IC para la proporción, tamaño de muestra y IC para la media, todo con los mismos datos.

Datos generales: $n = 625$, $k = 400$ aprobados.

a) Proporción muestral y su intervalo al 95 % ($z = 1,96$):

$$\hat{p} = 400 / 625 = 0,64$$

$$\sqrt{0,64 \cdot 0,36 / 625} = \sqrt{0,00036864} = 0,0192$$

$$e = 1,96 \cdot 0,0192 = 0,037632$$

$$IC = (0,64 - 0,0376 ; 0,64 + 0,0376) = (0,602 ; 0,678)$$

b) Ahora se fija el error y la incógnita es n . Aquí sí hay una estimación de p (la del apartado a), así que se usa 0,64 y no el caso peor:

$$n \geq z^2 \cdot \hat{p} \cdot (1-\hat{p}) / e^2$$

$$n \geq (1,96)^2 \cdot 0,64 \cdot 0,36 / (0,02)^2$$

$$n \geq 3,8416 \cdot 0,2304 / 0,0004 = 0,8851 / 0,0004 \text{ aprox. } 2212,8$$

Redondeando hacia arriba: $n = 2213$ alumnos.

c) La edad es una media, no una proporción, así que se cambia de fórmula. Al 90 % el z es 1,645:

$$\sqrt{625} = 25$$

$$e = 1,645 \cdot 4 / 25 = 6,58 / 25 = 0,2632$$

$$IC = (22 - 0,2632 ; 22 + 0,2632) = (21,737 ; 22,263) \text{ años}$$

Entre el 60,2 % y el 67,8 % de los alumnos aprueban. Para bajar el margen del 3,8 % al 2 % hacen falta 2213 alumnos, tres veces y media la muestra actual: la precisión se paga cara, porque n crece con el cuadrado. El intervalo de la edad, en cambio, es muy estrecho porque σ es pequeña frente a n .

Resultado: a) $IC = (0,602 ; 0,678)$ b) $n = 2213$ c) IC de la edad = $(21,737 ; 22,263)$ años

Fin del Bloque 4 — Inferencia estadística, y fin de la asignatura: con este tema se cierran los nueve temas de Matemáticas CCSS.