

# Jugando a adivinar: los misterios de la estimación de parámetros

Universidad de Málaga (España)

Texto de clase

Autor: Antonio Matas-Terrón

Fecha de la versión: 30 de septiembre de 2024

## Empezamos con un reto

En el corazón de una ciudad bulliciosa como Villanueva del Trabuco está el centro educativo donde trabajamos, conocido como **“Future Learners’ Hub”**. La reputación del centro se había extendido por todas partes, con estudiantes que vienen desde países lejanos.

Una buena mañana, te cita el director del centro en su despacho. Se trata de un caballero de expresión seria, pero amable, “tengo una tarea para ti. Como sabes, la junta se ha interesado por el progreso y las preferencias de nuestros estudiantes. Quieren optimizar los programas del centro para adaptarlos mejor a nuestros usuarios.”

El director te entrega un expediente grueso. “Esto”, dijo, “contiene los datos que tenemos hasta ahora de nuestro alumnado, pero no está todo, falta más de la mitad y es necesario que sepamos que pasa con el total”.

Examinas brevemente los archivos y te das cuenta de que hay muchísimos grupos distintos, tanto por tipo de estudios, como por edades, procedencia, etc., y que apenas hay una pequeña porción de toda la información que se pide. “Ésto es un trabajo bastante grande”, murmuras. “Comprender qué pasa en una muestra es una cosa, pero hacer una estimación a partir de ahí, sobre todas la gente, es otra cosa”.

“Sí”, te dice el director, “pero confío en ti”. Es como un rompecabezas, y sólo tenemos algunas de las piezas. El desafío es importante, pero necesitamos visualizar el panorama completo con lo que tenemos para poder determinar qué actuaciones tomar en el futuro inmediato en el centro.

Miras hacia arriba y barruntas “No se trata sólo de números y datos. Se trata de comprender qué ocurre aquí. ¡Menudo marrón me ha caído!”.

Al salir del despacho del director, sabes que la tarea que tienes por delante va a ser una de las experiencias más desafiantes de su carrera... Sobre todo porque no prestas atención a las clases de Métodos y ahora no sabes cómo se hace.

Moraleja: ahora estás aquí, y tienes la oportunidad de prestar atención, y hacer el esfuerzo por comprender y recordad esta parte tan básica del ejercicio profesional a la que vamos a llamar “Estimación de Parámetros”.

Antes de comenzar, recordemos dos conceptos básicos:

- **Parámetro:** es un valor numérico que caracteriza o resume una característica de una población completa, que suele ser desconocido.
- **Estadístico:** es un valor numérico que caracteriza o resume una característica de una muestra.

## Introducción a la estimación de parámetros

Tenemos claro que casi siempre trabajamos con muestras procedentes de una población, y sin embargo, también es cierto que casi siempre nos interesa saber cómo es esa población y no solamente quedarnos con los resultados de la muestra. Por ejemplo, si tenemos los resultados del nivel medio de tensión arterial de una parte los trabajadores de nuestra empresa, es posible tratar de estimar (suponer) cuál será el resultado promedio para el total de trabajadores. Para ello realizaremos la estimación de parámetros.

NOTA: el objetivo de la “estimación de parámetros” es tratar de conjeturar cuál puede ser el parámetro de una población partiendo de un estadístico muestral.

A partir de las distribuciones muestrales, sabemos que un estadístico se puede distribuir con una determinada función de probabilidad puesto que dicho estadístico puede adoptar distintos valores (se trata de una variable aleatoria). Por el contrario, un parámetro siempre es una constante. Por tanto, la estimación de parámetros asume el reto de tratar de averiguar el valor de esa constante (el parámetro), partiendo de una variable aleatoria (el estadístico).

Cuando el estadístico de la distribución muestral se utiliza para estimar el parámetro, pasa a llamarse “estimador”. Esto es así porque la gente experta en estadística les gusta poner nombres a las cosas, y de paso, hacernos un lío.

En cualquier caso, los estimadores tienen que cumplir una serie de requisitos, si es posible, que vamos a tratar de explicar de forma amena:

- **Ausencia de sesgo.** Imagina que un estimador es como un jugador de dardos. Cuando este jugador es un verdadero profesional, lanza sus dardos siempre al centro del blanco. Si nuestro estimador es “insesgado”, significa que, en promedio, ¡da justo en el blanco! Esto se traduce en:  $E(\hat{E}) = \theta$ . ¡Pleno al centro!
- **Consistencia.** Ahora, piensa en un estimador como un atleta en entrenamiento. Al principio, podría no ser el mejor, pero con cada práctica (o en nuestro caso, con muestras más grandes) va mejorando.
- **Eficiencia.** Imaginemos a dos estimadores,  $\hat{E}_1$  y  $\hat{E}_2$ , en una carrera de eficiencia. ¿Cómo decidimos quién gana? El que tiene menos varianza se lleva el trofeo. Otro ejemplo: si tuviéramos que elegir entre dos GPS para

guiarnos, elegiríamos el que nos da direcciones más precisas y consistentes: el de menor varianza.

- **Suficiencia.** Finalmente, piensa en un detective que tiene todas las pistas para resolver un caso. ¡No necesita más información porque ya tiene todo! En nuestro mundo, un estimador es “suficiente” cuando ha recopilado y usado toda la información que la muestra le proporcionó. Es como tener todas las piezas del rompecabezas y no dejar ninguna fuera.

Hasta aquí la *intro* de la estimación. Vamos a pasar a ver cómo se hace. Para ello, solamente una última anotación. La estimación de parámetros implica dos procedimientos (también se puede entender que es un solo procedimiento con dos fases):

- La estimación puntual.
- La estimación por intervalos.

### Estimación puntual

Consiste en decir que el parámetro es igual al estimador ¡y quedarse tan pancho!

Imagina que te encuentras en un centro escolar y decides descubrir el cociente intelectual medio de todos esos alumnos que corren por los pasillos. Para eso, pasas un test de inteligencia a una muestra y descubres que el CI promedio es de, digamos, 105. ¿Qué haces ahora? Podemos asumir que el CI del colegio es el mismo que el de la muestra. Pero aquí es donde las cosas se ponen interesantes. Al hacer una estimación puntual, hay bastantes posibilidades de equivocarse.

Matemáticamente hablando, este error es simplemente la diferencia entre nuestro estimador y el verdadero valor del parámetro. Imagínalo como una especie de baile entre números:  $E = |\hat{E} - \theta|$  donde  $E$  es el error,  $\hat{E}$  es el estimador, y  $\theta$  es el parámetro. Y surge una pregunta ¿Cuánto es el error  $E$ ? ¡Pues tu sabrás!

### Estimación por intervalos

Esta es una segunda forma que además, necesita realizar una estimación puntual previamente. Con la estimación por intervalos, en lugar de decir “¡es este!”, se dice “bueno, podría estar entre estos” sin jugártelo todo a una sola carta (en este caso, a sólo un valor). Así se tiene una especie de una póliza de seguros de que tu parámetro está ahí, entre los valores que dices. El inconveniente es que no se puede decir cuál es el valor en concreto. Puedes estar cerca o lejos, pero nunca completamente seguro. En cualquier caso, es más probable acertar con una estimación por intervalos que con una puntual. Porque con la estimación puntual, aunque se puede ser *super* preciso, también es muy fácil equivocarse. En cambio, el intervalo da más margen de acierto, aunque con menos precisión. En este caso, el error que podemos tener no se expresa en términos absolutos (como en la estimación puntual) sino en términos de máximo admitido:  $E \geq |\hat{E} - \theta|$ .

Para terminar, un apunte: cuando hablamos de estos intervalos, nos apoyamos

en distribuciones muestrales. ¡Y aquí viene la parte emocionante! El rango donde se espera encontrar el parámetro se llama “intervalo de confianza”. Es como decir: “confío en que esté aquí”. Además:

- El nivel de confianza (N.C.) se representa por  $(1 - \alpha)$  y suele expresarse en porcentajes.
- Pero, siempre hay un riesgo de equivocarse, y eso se llama nivel de riesgo. Se representa como  $\alpha$  y NO se expresa nunca en porcentajes.

Si te das cuenta, el nivel de riesgo se representa con un misterioso  $\alpha$  y al nivel de confianza como  $(1 - \alpha)$ . Es un juego de opuestos, donde confianza y riesgo bailan juntos.

Por último, vamos a ver el proceso para la construcción de un intervalo de confianza que estima un parámetro. Este proceso siempre es el mismo:

- 1.- Identificamos qué distribución muestral corresponde con el estadístico/estimador en cuestión.
- 2.- Establecemos el nivel de riesgo que vamos a asumir. Habitualmente, se suele situar en 0.05, o en 0.01, aunque siempre es posible utilizar cualquier otro  $\alpha$ .
- 3.- Se hace una estimación puntual, que se quedará en el centro del intervalo.
- 4.- Calculamos el error típico de estimación. Este error y la fórmula concreta para obtenerlo, dependerá de cómo se calcula la desviación típica de la distribución muestral (que a su vez depende del tipo de estadístico).
- 5.- Se construye el intervalo sumando y restando el error típico a la estimación puntual.

Dicho todo esto, y para aclararlo bien, hay que afrontar el siguiente reto: conocer cómo se calculan los intervalos de confianza para los parámetros más solicitados: media y proporción.

## Intervalo de confianza para la media

Como sabemos, la distribución muestral de la media (o promedio) tiene la forma de una campana de Gauss (distribución normal). Pues bien, conociendo esto, la construcción del intervalo de confianza para la estimación de un parámetro media (que indicaremos como  $\mu$  porque hay que recordar que los parámetros se escriben en letras griegas) a partir de la media de una muestra (escrito como  $\bar{X}$  porque los estadísticos se indican con letras latinas) sigue el siguiente procedimiento:

- 1.- Definir el nivel de riesgo (seguridad). Si no se nos indica, establecemos el nivel de riesgo  $\alpha$  que nos interese. Recordemos que el nivel de riesgo se expresa como proporción, y es el complementario

de nivel de confianza (N.C.) que se expresa como porcentaje. Como se ha dicho anteriormente, generalmente se elige un  $\alpha$  de 0.05 (N.C.=95%), o bien de 0.01 (N.C.=99%). Establecer el nivel de riesgo es fundamental porque define la zona de la curva normal entre los que se supone que estará el parámetro: los puntos  $-z_{\frac{\alpha}{2}}$  y  $z_{\frac{\alpha}{2}}$  de la curva normal estandarizada.

2.- Establecer los límites del intervalo. Se aplica la fórmula del cálculo (que incluye la estimación puntual, más y menos el error de estimación).

Fórmula para  $n$  mayor a 30 casos:

$$IC = \bar{X} \pm z_{\alpha/2} \frac{S_x}{\sqrt{n-1}}$$

Fórmula para  $n$  igual o menor a 30 casos:

$$IC = \bar{X} \pm t_{\alpha/(2,n-1)} \frac{S_n}{\sqrt{n-1}}$$

3.- Interpretamos los resultados.

Veamos el siguiente ejemplo, 300 antiguos alumnos de una facultad de ingeniería tardaron 5 años (cinco años, 0 meses, 0 días, etc.) en obtener su titulación. A partir de ahí, se quiere estimar el número de años que de promedio, tardan los estudiantes de ingeniería en graduarse, con una seguridad de que dicho promedio sea correcto del 95%. Además, se sabe que la desviación típica de la muestra es de 2 puntos.

Procedemos a estimar los años promedio para graduarse. En primer lugar, anotamos los datos que tenemos:

$$n = 300$$

$$\bar{X} = 5$$

$$S_x = 2$$

$$\alpha = 0.05$$

1.- Definir el nivel de riesgo. En esta caso, nos dicen que el nivel de confianza debe ser el 95%, por lo que  $\alpha$  será 0.05. Recordemos que el nivel de confianza es  $(1 - \alpha)\%$ .

2.- Construir los límites del intervalo. Puesto que  $n$  es mayor a 30, se utiliza la expresión basada en puntuaciones  $z$ :

$$IC = \bar{X} \pm z_{\alpha/2} \frac{S_x}{\sqrt{n-1}}$$

$$IC_{\bar{X}} = 5 \pm 1.96 \sqrt{\frac{4}{300-1}} \rightarrow (4.774, 5.226)$$

3.-Interpretación. Por tanto, el promedio de años para la graduación del alumnado de ingeniería, está entre los 4.774 años y los 5.226 años, y esto lo decimos con una seguridad del 95%.

En el caso de tener una muestra pequeña, de por ejemplo sólo 20 alumnos, el procedimiento será el siguiente:

Anotamos los datos que tenemos:

$$n = 20$$

$$\bar{X} = 5$$

$$S_x = 2$$

$$\alpha = 0.05$$

1.- Nivel de riesgo:  $\alpha = 0.05$  (N.C. = 95%).

2.- Construir los límites del intervalo. Puesto que  $n$  es menor a 30, se utiliza la expresión basada en puntuaciones  $t$ :

$$IC = \bar{X} \pm t_{n-1, \alpha} \sqrt{\frac{S_x^2}{n-1}}$$

$$IC_{\bar{X}} = 5 \pm 2.093 \sqrt{\frac{4}{19}} \rightarrow (4.04 - 5.96)$$

3.-Interpretación. El promedio de graduación del alumnado de ingeniería está entre los 4.04 años y los 5.96 años, con un grado de seguridad de que esto es así del 95%.

**NOTA:** ¿Te has dado cuenta de qué ocurre con la amplitud del intervalo cuando disminuimos el tamaño muestral?

## Intervalo de confianza para las proporciones

En el caso de proporciones, el reto es estimar el parámetro  $\pi$  (proporción de un suceso en una población) a partir del estadístico  $p$  (proporción de un suceso en una muestra de tamaño  $n$ ). En este caso, la distribución muestral no es una normal sino una binomial  $B(n, p)$ . No obstante, ya sabemos que un tamaño de muestra  $n$  grande permite que usemos la distribución normal en lugar de la binomial.

El procedimiento de estimación sigue las mismas tres fases de la estimación de medias, aunque hay que tener en cuenta las siguientes advertencias:

- Si el tamaño de la muestra es muy pequeño (y en esto los matemáticos no se ponen de acuerdo, pero se suele tomar como referente  $n < 30$ ) tendríamos que calcular las probabilidades de cada suceso.
- Si el tamaño de la población  $N$  es infinita o finita muy grande. En este último caso, se incluye un añadido a la fórmula.

Veamos el procedimiento:

- 1.- Definir el nivel de riesgo  $\alpha$ .
- 2.- Establecer los límites del intervalo. Puesto que vamos a trabajar con muestras grandes, de  $n > 30$ , podemos usar como referente a la curva normal. No obstante, diferenciaremos entre poblaciones finitas o infinitas:

Fórmula para  $N$  finitas:

$$IC = p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

Fórmula para  $N$  infinitas:

$$IC = p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

- 3.- Interpretamos los resultados.

Realicemos el siguiente ejemplo:

Se sabe que en muestra aleatoria de 300 centros escolares, el 80% del alumnado promociona en primera convocatoria de 1º de ESO. Asumiendo que el número de centros escolares se considera infinita, la pregunta sería, ¿cuál es la proporción de promoción a nivel nacional?

Anotamos los datos que tenemos:

$$n = 300$$

$$p = 0.80$$

$$\alpha = 0.05 \quad N : \text{desconocido/infinito}$$

- 1.- Nivel de riesgo:  $\alpha = 0.05$  (N.C. = 95%).

- 2.- Construir los límites del intervalo:

$$IC = p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}};$$
$$IC = 0.8 \pm 1.96 \sqrt{\frac{0.8 \cdot 0.2}{300}} \rightarrow (0.75 - 0.84)$$

3.-Interpretación. La proporción de alumnado promocionado en la primera convocatoria en 1º de la ESO a nivel nacional está entre el 75% y el 84% con un grado de seguridad del 95%.

Vamos ahora a un segundo ejemplo donde vamos a estimar un intervalo de confianza para la proporción de alumnos que presentan alguna disfunción leve en un colegio, basándonos en una muestra pequeña de estudiantes de 20 estudiantes que seleccionamos admitiendo el reemplazo (un estudiante puede ser elegido más de una vez) para poder aplicar la distribución binomial.

Supongamos que en un colegio hay 200 alumnos en total, y queremos estimar la proporción de estudiantes que tienen alguna disfunción leve. Para ello, seleccionamos una muestra aleatoria de 20 alumnos, con reemplazo. De esos 20, observamos que 4 presentan algún tipo de disfunción leve. Nuestro objetivo es construir un intervalo de confianza al 80% para la proporción de alumnos con disfunción leve en el colegio.

### Solución paso a paso

1. Proporción de la muestra:

Dado que 4 de los 20 alumnos tienen disfunción leve, la proporción observada en la muestra es:

$$\hat{p} = \frac{4}{20} = 0.2$$

2. Distribución binomial: Al haber seleccionado con reemplazo, los resultados pueden modelarse como un experimento binomial, donde cada selección tiene dos posibles resultados: el alumno tiene o no disfunción leve. La probabilidad de éxito (que un alumno tenga disfunción leve) es la proporción de la muestra, es decir,  $p = 0.2$ . El número de éxitos sigue una distribución binomial con parámetros:

- $n = 20$  (número de alumnos seleccionados),
- $p = 0.2$  (proporción observada).

La fórmula para calcular la probabilidad binomial es:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

Donde  $\binom{n}{k}$  es el coeficiente binomial.

A continuación te muestro cómo realizar el cálculo de la probabilidad binomial para el caso de 2 y 3 alumnos con disfunción leve, utilizando la fórmula de la distribución binomial.

Cálculo para  $k = 2$  (2 alumnos con disfunción leve)

1. Coeficiente binomial para  $k = 2$  y  $n = 20$ :

$$\binom{20}{2} = \frac{20!}{2!(20-2)!} = \frac{20 \times 19}{2 \times 1} = 190$$

2. Probabilidad de éxito para  $k = 2$ :

$$p^k = (0.2)^2 = 0.04$$



3. Probabilidad de fracaso para  $n - k = 18$ :

$$(1 - p)^{n-k} = (0.8)^{18} \approx 0.0148$$

4. Cálculo de la probabilidad para  $k = 2$ :

$$P(X = 2) = \binom{20}{2} \cdot (0.2)^2 \cdot (0.8)^{18}$$

Sustituyendo los valores:

$$P(X = 2) = 190 \cdot 0.04 \cdot 0.0148 \approx 0.1369$$

Cálculo para  $k = 3$  (3 alumnos con disfunción leve)

1. Coeficiente binomial para  $k = 3$  y  $n = 20$ :

$$\binom{20}{3} = \frac{20!}{3!(20-3)!} = \frac{20 \times 19 \times 18}{3 \times 2 \times 1} = 1140$$

2. Probabilidad de éxito para  $k = 3$ :

$$p^k = (0.2)^3 = 0.008$$

3. Probabilidad de fracaso para  $n - k = 17$ :

$$(1 - p)^{n-k} = (0.8)^{17} \approx 0.0185$$

4. Cálculo de la probabilidad para  $k = 3$ :

$$P(X = 3) = \binom{20}{3} \cdot (0.2)^3 \cdot (0.8)^{17}$$

Sustituyendo los valores:

$$P(X = 3) = 1140 \cdot 0.008 \cdot 0.0185 \approx 0.2054$$

La siguiente tabla muestra las probabilidades para diferentes números de alumnos con disfunción leve en la muestra de 20 alumnos, incluyendo sólo hasta 9 casos, puesto que a partir de entonces la probabilidad es despreciable y la probabilidad acumulada es prácticamente del 100%.

Número de alumnos con disfunción leve	Proporción en la muestra	Probabilidad	Probabilidad acumulada
0	0.00	0.0115	0.0115
1	0.05	0.0576	0.0691
2	0.10	0.1369	0.2060
3	0.15	0.2054	0.4114
4	0.20	0.2182	0.6296
5	0.25	0.1746	0.8042
6	0.30	0.1118	0.9160
7	0.35	0.0558	0.9718
8	0.40	0.0217	0.9935
9	0.45	0.0066	1.0000

### 3. Intervalo de Confianza:

Para construir un intervalo de confianza al 80%, necesitamos encontrar un rango de proporciones que acumule aproximadamente un 80% de probabilidad. Observamos que la probabilidad acumulada de que entre 2 y 5 alumnos presenten disfunción leve es:

$$P(2 \leq X \leq 5) = 0.1369 + 0.2054 + 0.2182 + 0.1746 = 0.7351$$

Este rango cubre aproximadamente el 73.51% de la distribución, lo que está relativamente cerca del intervalo deseado. Podemos ajustar ligeramente el rango incluyendo hasta 6 alumnos:

$$P(2 \leq X \leq 6) = 0.7351 + 0.1118 = 0.8469$$

Este intervalo nos da un 84.69% de confianza, lo que es razonablemente cercano a un intervalo de confianza al 80%.

4.- Interpretación: basándonos en la muestra de 20 alumnos y la proporción observada de alumnos con disfunción leve, podemos decir que, con un nivel de confianza del 80%, la proporción de alumnos con disfunción leve en la población total del colegio probablemente se encuentre entre el 10% y el 30%.

**Para ampliar:** En estos vídeos se explica todo lo anterior. Se puede repasar el contenido con ellos:

- <https://youtu.be/c6e-PlmXpyg?si=g7rkli-aph-bfiIh&t=404> (a partir del minuto 6:44)
- <https://youtu.be/2wugQGg1GNY?si=SSX78ofO8X4BDnLB>

## Ejercicios resueltos

1.- Se tiene una muestra aleatoria de 30 personas a las que se ha pasado una prueba de lectura. El promedio fue de 50 puntos, con una desviación típica de 5 puntos. ¿Cuál será el promedio en la prueba de lectura, en la población origen de la muestra, con un grado de seguridad del 95%?

Datos:

- $n = 30$  (tamaño de la muestra) que consideraremos muestra pequeña.
- $\bar{X} = 50$  (media de la muestra)
- $s = 5$  (desviación estándar de la muestra)
- $\alpha = 0.05$  (dado que queremos un 95% de confianza)

$$IC = \bar{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n-1}}$$

Dado que  $n = 30$ , usaremos la prueba  $t$  con los grados de libertad  $df = n - 1 = 29$ . Buscando en la tabla  $t$  de Student para  $\alpha/2 = 0.025$  y  $df = 29$ , encontramos un valor de  $t \approx 2.045$ .

Sustituyendo en la fórmula:

$$IC = 50 \pm 2.045 \cdot \frac{5}{\sqrt{30-1}}$$

$$IC = 50 \pm 2.045 \cdot 0.912$$

$$IC = 50 \pm 1.867$$

Esto nos da un intervalo de:

$$(48.133, 51.867)$$

Por lo tanto, con un 95% de confianza, el promedio de la prueba de lectura en la población origen de la muestra estará entre 48.133 y 51.867 puntos.

2.- Se está realizando un estudio sobre la satisfacción del alumnado con sus docentes. Para ello, se han recogido datos en una muestra aleatoria de 100 alumnos. En el instrumento que se les pasa, se obtiene una media de 100 con una desviación típica de 15 puntos. ¿Cuál será el nivel de satisfacción del alumnado si establecemos un nivel de riesgo (de equivocarse) de 0.01?

Dado que:

- $n = 100$  (tamaño de la muestra)
- $\bar{X} = 100$  (media de la muestra)

- $s = 15$  (desviación estándar de la muestra)
- $\alpha = 0.01$  (nivel de riesgo de equivocarse)

Dado que estamos trabajando con una muestra grande ( $n > 30$ ), podemos usar la distribución  $z$ . Buscando en la tabla  $z$  para  $\alpha/2 = 0.005$  (debido al riesgo del 1% o 0.01 dividido entre 2), encontramos un valor de  $z \approx 2.576$ .

$$IC = \bar{X} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n-1}}$$

Sustituyendo en la fórmula:

$$IC = 100 \pm 2.576 \cdot \frac{15}{\sqrt{100-1}}$$

$$IC = 100 \pm 2.576 \cdot 1.5$$

$$IC = 100 \pm 3.864$$

Esto nos da un intervalo de:

$$(96.136, 103.864)$$

Por lo tanto, con un 99% de confianza (nivel de riesgo del 0.01), el nivel de satisfacción del alumnado con sus docentes en la población origen de la muestra estará entre 96.136 y 103.864 puntos.

3.- En un estudio con una muestra aleatoria de 50 personas, se obtuvo un grado de estrés medio de 80 puntos, con una desviación típica de 10. ¿Qué nivel de estrés medio tendrá la población origen? Se quiere obtener los resultados con un 95% de seguridad.

Dado que:

- $n = 50$  (tamaño de la muestra)
- $\bar{X} = 80$  (media de la muestra)
- $s = 10$  (desviación estándar de la muestra)
- $\alpha = 0.05$  (nivel de riesgo de equivocarse)

Usaremos la fórmula del intervalo de confianza para la media cuando se conoce la desviación estándar de la muestra. Buscamos en la tabla  $z$  para  $\alpha/2 = 0.025$  (debido al riesgo del 5% o 0.05 dividido entre 2), encontramos un valor de  $z \approx 1.96$ .

$$IC = \bar{X} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n-1}}$$

Sustituyendo en la fórmula:

$$IC = 80 \pm 1.96 \cdot \frac{10}{\sqrt{50-1}}$$

$$IC = 80 \pm 1.96 \cdot 1.4142$$

$$IC = 80 \pm 2.7718$$

Esto nos da un intervalo de:

$$(77.2282, 82.7718)$$

Por lo tanto, con un 95% de confianza, el nivel de estrés medio en la población origen de la muestra estará entre 77.2282 y 82.7718 puntos.

4.- En una muestra aleatoria de 47 jóvenes parejas, 9 de ellas tienen problemas en su relación. ¿Cuál será la proporción de jóvenes parejas con problemas en la población de donde proceden las de la muestra, a un nivel de riesgo de error del 0.05?

Datos del problema:

- Tamaño de la muestra ( $n$ ): 47
- Número de éxitos ( $x$ ): 9 (jóvenes parejas con problemas)
- Proporción muestral ( $\hat{p}$ ):  $\hat{p} = \frac{9}{47} \approx 0.191$
- Nivel de confianza: 95% (corresponde a  $\alpha = 0.05$ , y  $Z_{\alpha/2}$  es aproximadamente 1.96).

El intervalo de confianza se calcula usando la siguiente fórmula:

$$IC = \hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Donde: -  $\hat{p}$  es la proporción muestral. -  $Z_{\alpha/2}$  es el valor crítico de la distribución normal estándar para un nivel de confianza del 95% (en este caso, 1.96). -  $n$  es el tamaño de la muestra.

En este caso, hablaremos de margen de error para ser conscientes de que la terminología no tiene que ser rígida:

$$ME = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Sustituyendo los valores:

$$ME = 1.96 \cdot \sqrt{\frac{0.191(1 - 0.191)}{47}} = 1.96 \cdot \sqrt{\frac{0.191 \cdot 0.809}{47}} = 1.96 \cdot \sqrt{\frac{0.1542}{47}} = 1.96 \cdot \sqrt{0.0032}$$

$$ME = 1.96 \cdot 0.0573 \approx 0.1123$$

Cálculo del intervalo de confianza:

Ahora que tenemos el margen de error ((  $ME = 0.1123$  )), podemos calcular el intervalo de confianza sumando y restando el margen de error a la proporción muestral ( $\hat{p}$ ):

$$IC = \hat{p} \pm ME = 0.191 \pm 0.1123$$

Esto nos da el siguiente intervalo:

$$IC = [0.191 - 0.1123, 0.191 + 0.1123] = [0.0787, 0.3033]$$

Interpretación:

Con un 95% de confianza, podemos decir que la proporción de jóvenes parejas con problemas en la población está entre 0.0787 (7.87%) y 0.3033 (30.33%).

5.- En las últimas huelgas convocadas por las asociaciones de estudiantes no se han registrado exactamente el número personas que secundaron la huelga. Solamente se sabe que en una muestra aleatoria de 200 estudiantes, 120 afirma haber estado de huelga. Si la población de estudiantes total es de 32500, y asumiendo un nivel de confianza del 95% ¿Cuál puede ser el porcentaje de estudiantes que realmente secundó la huelga?

Dado que:

- $n = 200$  (tamaño de la muestra)
- $p = \frac{120}{200} = 0.6$  (proporción de la muestra que afirma haber estado de huelga)
- $q = 1 - p = 0.4$
- $\alpha = 0.05$  (nivel de confianza del 95%)

Utilizando la aproximación normal (distribución  $z$ ) el valor de  $z \approx 1.96$  para  $\alpha/2 = 0.025$ .

$$IC = p \pm z_{\alpha/2} \cdot \sqrt{\frac{p \cdot q}{n}}$$

Sustituyendo en la fórmula:

$$IC = 0.6 \pm 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{200}}$$

$$IC = 0.6 \pm 1.96 \cdot 0.0346$$

$$IC = 0.6 \pm 0.0679$$

Esto nos da un intervalo de:

$$(0.5321, 0.6679)$$

Por lo tanto, con un 95% de confianza, entre el 53.21% y el 66.79% de todos los estudiantes secundó la huelga.

## Estudios de caso

1.- Se tienen los datos reales de un grupo de alumnos sobre las técnicas de estudio utilizadas para estudiar un documento de 6 páginas. En los datos se incluye lo siguiente:

- **Sujeto:** código del alumno.
- **Tiempo:** número de minutos que tardaron en estudiar el artículo.
- **Técnicas\_total:** número de técnicas que han usado.
- **Puntutacion\_total:** calificación que obtuvieron en la prueba.

Los datos se pueden descargar desde aquí:

<https://acortar.link/aXQpdm>

A partir de estos datos, realizar lo siguiente:

- Estimar los parámetros poblacionales e interpretarlos.
- Explicar y justificar los procedimientos elegidos.

2.- En este caso debes suponer que trabajas como en el departamento de orientación de un Centro de Profesores (CEP). De acuerdo con nuestros registros, contamos con información detallada de 135 docentes, distribuida en las siguientes categorías:

- **Edad:** Es una variable cuantitativa con una media de 56 años y una desviación estándar de 20 años.
- **Predisposición para la Actualización Docente:** Esta es una variable dicotómica, donde el 86% de los docentes muestra una disposición favorable hacia la realización de cursos de actualización.

- **Especialización Formativa:** En este aspecto, el 35% de los docentes posee formación en ciencias, mientras que el restante 65% se ha especializado en humanidades.
- **Localización del Centro Escolar:** En términos de ubicación, el 75% de los docentes pertenece a centros educativos de localización rural, y el 25% restante a centros urbanos.
- **Cociente Intelectual:** Esta variable cuantitativa refleja una media de 95 con una desviación estándar de 9 puntos.
- **Antecedentes Disciplinarios:** En relación con la apertura de expedientes por faltas leves, es relevante destacar que el 98% de los docentes no presenta antecedentes de este tipo.

Los datos los tienes en el siguiente enlace:

<https://acortar.link/0x1zun>

A partir de estos datos, realizar lo siguiente:

- Estimar los parámetros poblacionales e interpretarlos.
- Explicar y justificar los procedimientos elegidos.