

Supuestos prácticos. Bloque 1.

MUESTREO:

Una empresa nacional del sector textil ha decidido realizar una encuesta entre sus trabajadores para conocer el nivel de satisfacción en su puesto de trabajo, con el objetivo de introducir cambios para mejorar la motivación de los empleados.

El número de empleados es 368678, pero quiere centrarse en conocer la opinión de los empleados que trabajan en las tiendas, que según el departamento de recursos humanos son 211245. El responsable de llevar a cabo la investigación ha decidido, en un primer acercamiento a la realidad objeto de estudio, realizar 964 encuestas.

Identifica en esta información, la población, el marco muestral, la unidad muestral y la muestra.

A partir de la información que contiene el enunciado, podemos identificar los conceptos que se nos piden:

Unidad muestral: Trabajadores totales de la empresa (*cada uno de los elementos que componen la población y que, a priori, se encuentran en disposición de ser seleccionados*).

Marco muestral: trabajadores de la empresa que, además, desarrollan su actividad en las tiendas (*parte de la población de la que vamos a seleccionar la muestra, según una serie de requisitos señalados en la investigación*).

Población (N): 211245 trabajadores (*conjunto de elementos de los que se desea obtener información dada su relación con el ámbito objeto de estudio*).

Muestra (n): 964 trabajadores (*conjunto de unidades seleccionadas para su análisis de entre el total de la población o marco muestral. Permite obtener información útil relativa a la población considerada*).

En un segundo paso, el responsable de la investigación considera aceptable asumir un error de 2,5% con un nivel de confianza del 95%, y le gustaría conocer el tamaño de la muestra para reconsiderar si es necesario aumentarlo o disminuirlo.

A pesar de que $N > 100.000$, al conocer el tamaño de la población, utilizamos la fórmula del cálculo del tamaño de la muestra para poblaciones finitas.

$$N = 211245 \quad / \quad p = q = 0,5 \quad / \quad N.C. = 95\% \rightarrow z = 1,96 \quad / \quad e = 0,0025$$

$$n = \frac{z^2 N p}{e^2 (N-1) + z^2 p q} = \frac{(1,96)^2 \cdot 2111245 \cdot 0,5 \cdot 0,5}{(0,025)^2 \cdot (2111245-1) + (1,96)^2 \cdot 0,5 \cdot 0,5} = \frac{3,8416 \cdot 2111245 \cdot 0,25}{0,000625 \cdot 211244 + 3,8416 \cdot 0,25} = \frac{202879,698}{132,0275 + 0,9604} = \frac{202879,698}{132,9879} = 1525,55 \approx 1526$$

Por lo tanto, el tamaño de la muestra adecuado para un error muestral del 2,5% con un nivel de confianza del 95% debe ser de 1526 encuestas, superior a lo que se había supuesto en un primer paso de la investigación.

¿Con qué error se está trabajando en el caso de utilizar el tamaño de la muestra propuesta en el inicio de la investigación? Considerado que el N.C. es el mismo que en el apartado anterior.

$$N = 211245 \quad / \quad n = 964 \quad / \quad p = q = 0,5 \quad / \quad \text{N.C.} = 95\% \rightarrow z = 1,96$$

Como en el apartado anterior, a pesar de que $N > 100000$, utilizamos la fórmula para una población finita:

$$e = z \cdot \sqrt{\frac{(N-n) \cdot p \cdot q}{n \cdot (N-1)}} = 1,96 \cdot \sqrt{\frac{(211245-964) \cdot 0,5 \cdot 0,5}{964 \cdot (211245-1)}} = 1,96 \cdot \sqrt{\frac{52570,25}{203639216}} = 1,96 \cdot \sqrt{0,00025815} = 1,96 \cdot 0,016 = 0,0314 \rightarrow e = 3,14\%$$

En el caso de continuar la investigación con el tamaño de la muestra indicado al comienzo del ejercicio, el error muestral será del 3,14% para un nivel de confianza del 95%

El responsable de la investigación se plantea llevar a cabo un muestreo estratificado por sexo. Considerando los dos tamaños de las muestras que tenemos, calcula el tamaño de los estratos mediante:

a) Afijación simple

$$\text{Si } n = 964 \quad / \quad i = 2$$

En la afijación simple, todos los estratos tendrán el mismo tamaño (n_i):

$$n_i = \frac{n}{i} = \frac{964}{2} = 482$$

Por lo tanto, se realizará la encuesta a 482 hombres y 482 mujeres.

$$\text{Si } n = 1526 \quad / \quad i = 2$$

$$n_i = \frac{n}{i} = \frac{1526}{2} = 763$$

En este caso, la encuesta se realizará a 763 hombres y 763 mujeres.

- b) Afijación proporcional, teniendo en cuenta que la distribución por sexos es de 117489 mujeres y 93456 hombres.

$$\underline{Si\ n = 964} \ / \ N = 211245 \ / \ N_H = 93756 \ / \ N_M = 117489$$

En la afijación proporcional, el número de elementos de la muestra que se incluyen en cada estrato será proporcional a la representación de cada uno en la población total:

$$n_H = \frac{n}{N} \cdot N_H = \frac{964}{211245} \cdot 93756 = 427,85 \approx 428$$

$$n_M = \frac{n}{N} \cdot N_M = \frac{964}{211245} \cdot 117489 = 536,15 \approx 536$$

En el caso de utilizar este tipo de afijación, se realizará la encuesta a 428 hombres y 536 mujeres.

$$\underline{Si\ n = 1526} \ / \ N = 211245 \ / \ N_H = 93756 \ / \ N_M = 117489$$

$$n_H = \frac{n}{N} \cdot N_H = \frac{1526}{211245} \cdot 93756 = 677,27 \approx 677$$

$$n_M = \frac{n}{N} \cdot N_M = \frac{1526}{211245} \cdot 117489 = 848,72 \approx 849$$

En el caso de utilizar este tipo de afijación, se realizará la encuesta a 677 hombres y 849 mujeres.

Para finalizar la investigación, la empresa está interesada en conocer la opinión de los talleres que se encargan de parte de la producción de la empresa. Para ello va a realizar 150 encuestas telefónicas, diferenciando entre talleres de patronaje y corte de tejido, talleres de costura y talleres de etiquetado y finalización de la prenda. Como en el caso anterior, el responsable decide utilizar un muestreo estratificado. Calcula el tamaño de los estratos, mediante:

- a) Afijación simple

$$n = 150 \ / \ i = 3$$

$$n_i = \frac{n}{i} = \frac{150}{3} = 50$$

Se realizarán encuestas telefónicas a 50 talleres de cada uno de los tipos que se han indicado.

- b) Afijación proporcional, teniendo en cuenta que la distribución por el tipo de talleres es de 594 de costura, 278 de patronaje y corte de tejido y 75 de etiquetado y finalización de la prenda.

$$n = 150 \ / \ N = 947 \ / \ N_C = 594 \ / \ N_P = 278 \ / \ N_E = 75$$

$$n_C = \frac{n}{N} \cdot N_C = \frac{150}{947} \cdot 594 = 94,08 \approx 94$$

$$n_P = \frac{n}{N} \cdot N_P = \frac{150}{947} \cdot 278 = 44,03 \approx 44$$

$$n_E = \frac{n}{N} \cdot N_E = \frac{150}{947} \cdot 75 = 11,88 \approx 12$$

Por lo tanto, en el caso de utilizar este tipo de afijación, se realizará la encuesta telefónica a 94 talleres de costura, 44 talleres de patronaje y corte y 12 talleres de etiquetado y finalización de las prendas.

COVARIANZA, CORRELACIÓN Y RECTA DE REGRESIÓN:

Un jefe de ventas considera que existe una relación entre el número de visitas (y) que un comercial realiza a nuevas empresas y el número de pedidos (x) que finalmente formalizan estas empresas. Durante 5 semanas obtuvo los siguientes resultados:

Pedido formalizados (x_i)	5	7	2	1	9
Visitas realizadas (y_i)	15	18	10	8	20

- Calcula el coeficiente de correlación lineal
- Si esta semana el comercial ha formalizado 6 pedidos de empresas nuevas, ¿cuántas visitas podemos suponer que ha realizado?

Lo primero que tenemos que hacer, antes de contestar a las preguntas que se plantean en el enunciado, es construir una tabla con los elementos que necesitamos para resolver el ejercicio:

	f_i	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
	1	5	15	25	225	75
	1	7	18	49	324	126
	1	2	10	4	100	20
	1	1	8	1	64	8
	1	9	20	81	400	180
Σ	5	24	71	160	1113	409

- Como en el enunciado no nos han dicho nada, suponemos que la frecuencia absoluta (f_i) = 1
- N es el número total de datos $\rightarrow N = \Sigma f_i$
- Medias aritméticas

$$\bar{x} = \frac{\sum xi}{N} = \frac{24}{5} = 4,8$$

$$\bar{y} = \frac{\sum yi}{N} = \frac{71}{5} = 14,2$$

4. Varianzas y desviaciones típicas

$$\sigma^2_x = \frac{\sum (xi)^2}{N} - (\bar{x})^2 = \frac{(24)^2}{5} - (4,8)^2 = 8,96$$

$$\sigma_x = \sqrt{\sigma^2_x} = \sqrt{8,96} = 2,993$$

$$\sigma^2_y = \frac{\sum (yi)^2}{N} - (\bar{y})^2 = \frac{(71)^2}{5} - (14,2)^2 = 20,96$$

$$\sigma_y = \sqrt{\sigma^2_y} = \sqrt{20,96} = 4,578$$

5. Covarianza

$$\sigma_{xy} = \frac{\sum xi \cdot yi}{N} - (\bar{x} \cdot \bar{y}) = \frac{409}{5} - (4,8 \cdot 14,2) = 13,64$$

La covarianza es positiva = *correlación directa* → al aumentar el número de visitas, aumentará el número de pedidos realizados por las nuevas empresas.

Una vez que tenemos hechos todos los cálculos anteriores, podemos comenzar a resolver las preguntas propuestas en el ejercicio.

a) Lo primero que nos piden es el coeficiente de correlación lineal de Pearson

El coeficiente de correlación de Pearson nos permite establecer la relación que existe entre dos variables. Es un valor que oscila entre -1 y +1

- *Un valor menor que 0 indica que existe una correlación negativa, es decir, que las dos variables están asociadas en sentido inverso. Cuánto más se acerca a -1, mayor es la fuerza de esa relación invertida. Cuando es exactamente -1, eso significa que tienen una correlación negativa perfecta.*
- *Un valor mayor que 0 indica que existe una correlación positiva. En este caso las variables estarían asociadas en sentido directo. Cuanto más cerca de +1, más alta es su asociación. Un valor exacto de +1 indicaría una relación lineal positiva perfecta.*
- *Finalmente, una correlación de 0, o próxima a 0, indica que no hay relación lineal entre las dos variables.*

Teniendo en cuenta los datos que hemos estado calculando hasta ahora:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{13,64}{2,993 \cdot 4,578} = 0,995$$

El valor de r está muy cercano a 1, por lo que existe una correlación positiva casi perfecta. Además, la estimación realizada estará muy cerca del valor real.

- b) Para responder a esta pregunta, debemos hallar la recta de regresión de y (dato que no conocemos y nos piden calcular) sobre x (dato conocido = número de pedidos realizados por nuevas empresas).

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma^2_x} \cdot (x - \bar{x})$$

$$y = 14,2 + \frac{13,64}{8,96} \cdot (x - 4,8) \rightarrow y = 14,2 + 1,52(x - 4,8) \rightarrow y = 1,52x + 6,9$$

Si tenemos en cuenta que en el enunciado nos dicen que $x = 6$, calculamos el número de visitas que ha realizados, sustituyendo en valor de x en la expresión anterior:

$$y = (1,52 \cdot 6) + 6,9 = 16$$

Por lo tanto, podemos suponer que el comercial ha realizado 16 visitas a empresas nuevas.