Tarea 2 Muestreo Estadistico.

Jhonatan Smith Garcia

15/4/2021

Se plantea solucion caso de estudio diapositiva 4, semana 5:

MAE CASO DE ESTUDIO: Un Ingeniero esta interesado en estudiar la productividad de las maquinas en una multinacional, dicha productividad esta representada por la produccion mensual (en toneladas) que dichas maquinas generan del producto. La empresa esta dividida en TRES plantas de produccion. Las caracterısticas de cada planta, asi como los resultados del estudio se muestran en la siguiente tabla. Adicionalmente, se registraron las maquinas que necesitan mantenimiento.

TABLA DE DATOS:

	Planta 1	Planta 2	Planta 3
Nro. Máq. por Planta	500	1000	4000
Nro. Máq. en la muestra	50	80	160
Nro. Máq. para mantenimiento	4	6	5
Producción Promedio (ton)	40	60	100
Desv. Est. De la producción	12	17	10

Ahora, las preguntas planteadas se encuentran en las diapositivas.

Solucion

1) En primera instancia se identifica la poblacion objetivo, la pregunta problema y demas temas relacionados para realizar un correcto analisis estadistico. Para este caso puntual, la pregunta problema o fenomeno de interes es la productividad de las maquinas en una empresa multinacional. Dicha productividad se calcula basado en la produccion mensual, teniendo en cuenta**tres** plantas de produccion. Cada planta tiene sus caracteristicas de produccion, cada una de ellas son indispensables para la medida de la produccion final. Es decir; la produccion final depende de la produccion de cada una de estas tres plantas. De esto ultimo se entiende que, la pregunta problema se responde a travez de un analisis de estratos (Cada una de las plantas). Luego, realiza un MAS para cada una de las plantas, seleccionando aleatoriamente una muestra de cada estrato (cada planta de produccion) de manera independiente y aqui, claramente, la seleccion de una planta, no tiene relacion ni incidencia sobre otras (muestras independientes). Bajo estas condiciones se asegura que el ingeniero realizo un Muestre Aleatorio Estratificado.

2) Antes que nada, se debe de tener en cuenta cada uno de los siguientes datos de la tabla:

- $H = 3$ El numero de estratos (Las tres plantas de produccion)
- N_h, con $h = 1, 2, 3$ El numero de maquinas en la h-ésima planta de produccion.
- n_h con $h = 1, 2, 3$ Numero de maquinas seleccionadas del h-ésimo estrato

- $N = N_1 + N_2 + N_3 = 5500$ El numero de maquinas en las respectivas plantas (Planta 1, tiene N_1 maquinas, asi con las otras dos). (N será tomado como el tamaño poblacional total)
- Por notacion $\bar{y}_h$, con $h = 1, 2, 3$ Produccion promedio en el h-esimo estrato (planta) Ademas, este es un estimador insesgado para la media poblacional. Este será el parametro a estimar.
- Por notacion S_h^2, con $h = 1, 2, 3$ La varianza muestral en el h-esimo estrato.
- $\alpha = 0.05$ -Como en cada caso, α es fijo y ademas, cada una de las muestras de los estratos es de un n mayor o igual a 30, entonces se tomará a $B \approx 2 * \widehat{Var(\bar{y_h})}$ para cada h. Esto debido a la aproximacion normal Z

Antes que nada, para hallar un IC al 95% se necesita realizar el calculo de la media muestral para cada estrado(dado en la tabla de resultados), con su respectiva varianza para cada planta. De esta manera:

Para la planra de produccion #1: $\widehat{Var(\bar{y_1})} = \frac{(N_1-n_1)*S_1^2}{N_1*n_1}$ asi, reemplazando los valores de la tabla obtenemos que:

$\widehat{Var(\bar{y_1})} = \frac{(500-50)*12^2}{500*50} \approx 2.592$

Con esto, se sabe que $B_1 = 2*\sqrt{2.592} \approx 3.22$ Teniendo en cuenta que B_1 es el limite del error estimado asociado al estrato $h = 1$ para la estimacion de la media muestral en el mismo h entonces un Ic de *aproximadamente* el 95% es $\bar{y_1} \pm B_1$= 40 $\pm$ 3.22

Por tanto el Ic solicitado para la media de la produccion mensual de la planta #1 es (36.78,43.22) implicando que, con una confianza de aproximadamente el 95% el valor real de la produccion real de toneladas promedio al mes esta entre el intervalo anterior.

Para la planra de produccion #2: Haciendo el mismo analisis hecho para la planta 1 se realizará para la planta #2.

$\widehat{Var(\bar{y_2})} = \frac{(N_2-n_2)*S_2^2}{N_2*n_2}$ asi, reemplazando los valores de la tabla obtenemos que:

$\widehat{Var(\bar{y_2})} = \frac{(1000-80)*17^2}{1000*80} \approx 3.3235$

Con esto, se sabe que $B_2 = 2 * \sqrt{3.3235} \approx 3.65$ Teniendo en cuenta que B_2 es el limite del error estimado asociado al estrato $h = 2$ para la estimacion de la media muestral en el mismo h entonces un Ic de *aproximadamente* el 95% es $\bar{y_2} \pm B_2$= 60 $\pm$ 3.65

Por tanto el Ic solicitado para la media de la produccion mensual de la planta #1 es (56.35,63.65) implicando que, con una confianza de aproximadamente el 95% el valor real de la produccion real de toneladas promedio al mes esta entre el intervalo anterior.

Para la planta de produccion #3: Haciendo el mismo analisis hecho para la planta 1 se realizará para la planta #2.

$\widehat{Var(\bar{y_3})} = \frac{(N_3-n_3)*S_3^2}{N_3*n_3}$ asi, reemplazando los valores de la tabla obtenemos que:

$\widehat{Var(\bar{y_3})} = \frac{(4000-160)*10^2}{4000*160} = 0.6$

Con esto, se sabe que $B_3 = 2*\sqrt{0.6} \approx 1.55$ Teniendo en cuenta que B_3 es el limite del error estimado asociado al estrato $h = 3$ para la estimacion de la media muestral en el mismo h entonces un Ic de *aproximadamente* el 95% es $\bar{y_3} \pm B_3$= 100 $\pm$ 1.55

Por tanto el Ic solicitado para la media de la produccion mensual de la planta #1 es (98.45,101.55) implicando que, con una confianza de aproximadamente el 95% el valor real de la produccion real de toneladas promedio al mes esta entre el intervalo anterior.

Entonces para estimar la produccion media mensual de la multinacional se describe de la siguiente manera:

$\bar{y}_{est} = \frac{1}{N} \sum_{h=1}^{H} N_h * \bar{y}_h$

Y claramente, todos los parametros de esta expresion son conocidos.Asi, reemplazando, tendriamos que:

$\bar{y}_{est} = \frac{1}{N}\sum_{h=1}^{H} N_h * \bar{y}_h = \frac{1}{5500}[(500)(40) + (60)(1000) + (100)(4000)] \approx 87.273$

Por tanto, en promedio; la multinacional produce aproximadamente 87 toneladas mensuales.

Conociendo este dato, la varianza del estimador esta dada por:

$\widehat{Var(\bar{y}_{est})} = \frac{1}{N^2}\sum_{h=1}^{3} N_h^2 * \widehat{Var(\bar{y}_h)} = \frac{1}{5500^2}[(500^2)(2.592) + (1000^2)(3.3235) + (4000^2)(0.6)] \approx 0.45$

Sea $B = 2 * \sqrt{0.45} \approx 1.35$ de esta forma, un IC de *aproximadamente* 95% para la produccion promedio mensual de la multinacional será (85.923,88.623)

Finalmente se concluye que, el valor real de la produccion emsnual de la multinacional es un valor dentro de este intervalo, con una confianza aproximada del 95%.

Estimacion total producido: Para la estimacion del total producido en la multinacional se usará los datos ya calculados. En particular, la media ya estimada.

$\widehat{\tau}_{est} = \sum_{h=1}^{H} N_h * \bar{y}_h = (500 * 40) + (1000 * 60) + (4000 * 100) = 480000$

La estimacion del total producido por la multinacional mensualmente es de unas 480000 toneladas.

Ahora, para calcular un IC al 90% de confianza, se tiene que: $\widehat{Var(\widehat{\tau}_{est})} = \sum_{h=1}^{H} N_h^2 * \widehat{Var(\bar{y}_h)}$ =(500$^{2)(2.592)+(1000}$2)(3.3235)+(4000^2)(0.6)=13571500

Ahora, para calcular el limite del error B se tiene que:

$B = Z_{0.95} * \sqrt{\widehat{Var(\widehat{\tau}_{est})}} = 1.644854 * 3683.952 = 6059.563$

De esta manera un IC con una confianza al 90% para la produccion del total de la empresa estara dado por $B \pm \widehat{\tau}$ =(473940.4,486059.6). Lo que implica que el valor real de la totalidad mensual producida se encuentra en el IC dado.

Estimar proporcion de maquinas en mantenimiento Para calcular dicha proporcion, reemplazamos en la respectiva formula de la proporcion para muestreo estratificado.

$$\widehat{p}_{est} = \frac{1}{N}\sum_{h=1}^{H} N_h * p_h$$

Pero de esta formula, se debe tener en cuenta que:

-$\widehat{p}_{est}$ es la proporcion de maquinas en mantenimiento(dato a estimar) -Dada por $p_h = \frac{a_h}{n_h}$ cada una de las proporciones de las muestras. -$\widehat{Var(p_h)} = \frac{p_h(1-p_h)}{n_h-1}\frac{N_h}{N_h}$ la varianza estimada para la proporcion de maquinas en mantenimiento.

Al realizar todos los calculos se tiene que: $\widehat{Var(p_1)} = \frac{4/50(46/50)}{50-1}\frac{500-50}{500} \approx 0.0014$

$\widehat{Var(p_2)} = \frac{6/80(74/80)}{80-1}\frac{1000-80}{1000} \approx 0.0008$

$\widehat{Var(p_3)} = \frac{5/160(155/160)}{160-1}\frac{4000-160}{4000} \approx 0.0002$

Finalmente,

$\widehat{p}_{est} = \frac{1}{5500}[(500 * 0.0014) + (1000 * 0.008) + (4000 * 0.002)] \approx 0.0436$

Este dato se traduce como, la proporcion de maquinas en mantenimiento de la multinacional es de aproximadamente el 0.0435 (4.36%)

$$\widehat{Var(p_{est})} = \frac{1}{5500^2}[(500^2 * 0.0014) + (1000^2 * 0.0008) + (4000^2 * 0.002)] \approx 0.00013$$

Para calcular el IC solicitado esta dado por $\widehat{p}_{est} \pm Z_{0.975} * \widehat{Var(p_{est})} = 0.0436 \pm 1.96 * \sqrt{0.0013}$=(0.0213,0.0659) Esto implica que el valor real de la proporcion de la smaquinas en mantenimiento de la empresa esta en el intervalo dado con una confianza del 95%.

Tamaño de muestra: Esta es la infromacion suministrada para calcular el tamaño de la muestra.

L.E.E:$B = Z_{0.975} * \widehat{Var(\bar{y_{est}})} = 3$ o bien $D = \frac{B^2}{Z^2_{0.975}} \approx 2.343$, $\sigma_1^2 = 12^2$,$\sigma_2^2 = 17^2$,$\sigma_3^2 = 10^2$, D=2.343, N=5500,$N_1 = 500$, $N_2 = 1000$ y $N_3 = 5500$

Y para calcular n se usa la expresion:

$n = \frac{\sum_{h=1}^{3} \frac{N_h^2 \sigma^2}{wh}}{N^2 D + \sum_{h=1}^{3} N_h \sigma_h^2}$

De esta expresion anterior los unicos datos desconocidos son los w_h para ello, se calculan todos siguiendo todos la asignacion oprtima de Neyman ya que se desconocen costos del muestreo y se asumen todos ellos son iguales.

Por tanto $w_h = \frac{N_h \sigma_h}{\sum_h^H N_h \sigma_h}$ para h=1 seria:

$w_1 = \frac{N_1 \sigma_1}{\sum_{h=1}^{3} N_h \sigma_h} = \frac{500*12}{(500*12)+(1000*17)+(4000*10)}$

y reemplazando los valores en esta expresion se tiene que $w_1 = 2/21$,$w_2 = 17/63$ y $w_3 = 40/63$.

Luego:

$N^2 * D = 5500^2 * 3 * 3/(1.96)^2 = 70868908.7$

$\sum_{h=1}^{3} \frac{N_h^2 \sigma_h^2}{w_h} = \frac{500^2 * 12^2}{2/21} + \frac{1000^2 * 17^2}{17/63} + \frac{4000^2 * 10^2}{40/63} \approx 3607870370.37037$ $\sum_{h=1}^{3} N_h \sigma_h^2 = N_1 \sigma_1^2 + N_2 \sigma_2^2 + N_3 \sigma_3^2 =$ 761000

Finalmente, reemplazando todos los valores en la ecuacion anterior; el valor final de $n \approx 50.3682109$, por tanto el tamaño de la muestra a tomar es n=51.

Ejercicio 19 taller 2

Para este ejercicio se tiene que hay un total de 1000 personas. Si se toma a N=1000 como el numerto total de habitantes y ademas sea Y_i= la variable aleatoria definida como el numero de personas que responde "si" a la encuesta en la muestra i-esima i=1,2 (ya que hay dos grupos muestreados) Entonces, dadas las muestras del ejercicio, se tienen las siguientes proporciones:

- $p_1 = \frac{Y_1}{n_1} = \frac{60}{80} = 0.75$ Esta es la proporcion de n_1 de personas censadas que votaron "si"
- $q_1 = 1 - p_1 = 0.25$ Este es el numero de personas censadas que que NO votaron "si"
- $p_2 = \frac{Y_2}{n_2} = \frac{8}{20} = 0.4$ Proporcion habitantes NO censados que votaron "Si"
- $q_2 = 1 - p_2 = 0.6$ Esta es la proporcion de habitantes NO censados que no votaron "Si"

Al igual que en el ejercicio anterior, usamos la misma formula para calcular $\widehat{p}_{est}$. de estta manera:

$\widehat{p}_{est} = \frac{1}{1000}[(0.75 * 800) + (0.4 * 200)] = 0.68$ De esto se entiende que, aproximadamente el 68% de los habitantes votaron "Si", es decir; a favor.

Ahora, para calcular el IC pedido, tambien se hará uso de la formula anterior de la varianza para la proporcion. Asi, para cada p, se tiene:

-$\widehat{Var(p_h)} = \frac{p_h(q_h)}{n_h - 1} \frac{N_h}{N_h}$

-$\widehat{Var(p_1)} = \frac{0.75*0.25}{70} * \frac{800-80}{800} = 0.00213607$ -$\widehat{Var(p_2)} = \frac{0.4*0.6}{20} * \frac{200-20}{200} = 0.0108$

Ahora:

-$\widehat{Var(p_{est})} = \frac{1}{1000^2}[(0.00213607)(800^2) + (200^2)(0.0108)] \approx 0.0015$

Ahora, Sea $B = 2 * \sqrt{(\widehat{Var(p_{est})})} = 0.07622335$ de esta manera, un IC de aproximadamente 95% para la proporcion de votantes del municipio que votaron a favor es

$\widehat{p}_{est} \pm B = 0.68 \pm 0.07622335 = (0.6037767, 0.7562234)$ Y esto, finalmente se interpreta que, con una confianza del 95% el dato real de la proporcion de votantes a favor en el municipio se encuentra en el anterior IC dado.

NOTA PARA EL DOCENTE= Una duda profe. El error B se puede aproximar gracias a una Z cuando n es mayor o igual a 30. Sin embargo, en este caso n es "grande" pero los n_h son, 80 y 20 respectivamente. Ambos deben cumplir la condicion de ser mayor a 30 para usar la aproximacion del error por una normal? O no se fija en n_h si no en n, el tamaño de la muestra una vez sumado los estratos. La duda surge debido al tamaño del segundo estrato en este ejercicio y que apenas vengo a caer en cuenta de esta difeencia. Gracias de antemano!!