Muestreo Aleatorio Simple

13/Febrero/2021

El muestreo aleatorio simple (M.A.S.) o muestreo aleatorio irrestricto es la forma más sencilla de muestreo probabilístico y proporciona la base teórica para formas de muestreo más complejas.

Existen dos formas de llevar a cabo el M.A.S:
- Con reemplazo: Se puede seleccionar varias veces la misma unidad de muestreo o elemento.
- Sin reemplazo: Solo se puede seleccionar una vez la misma unidad de muestreo o elemento.

¿Con o sin reemplazo?

En una M.A.S. sin reemplazo, las probabilidades de selección varían en función del tiempo o acciones realizadas (las variables son dependientes). En un M.A.S. con reemplazo las probabilidades de selección se mantienen constantes durante todo el tiempo (las variables son independientes).

Estimadores y propiedades

Sean $y_1, y_2, \ldots, y_n$ una muestra aleatoria con distribución desconocida utilizamos como estimador de μ, $\hat{\mu} = \bar{y}$, entonces $\hat{\mu}$ tiene las siguientes propiedades:

1. $E(\hat{\mu}) = \mu$

2. $$Var(\hat{\mu}) = \begin{cases} \dfrac{\sigma^2}{n}, & \text{si el M.A.S. es con reemplazo} \\[2ex] \dfrac{\sigma^2}{n}\left[\dfrac{N-n}{N-1}\right], & \text{si el M.A.S. es sin remplazo} \end{cases}$$

Demostración

$$E[\hat{\mu}] = E\left[\frac{\sum y_i}{n}\right] = \frac{1}{n}\sum E[y_i] = \frac{1}{\cancel{n}}\cancel{n}\mu = \mu$$

$$Var(\hat{\mu}) = Var\left(\frac{\sum y_i}{n}\right)$$

$$= \frac{1}{n^2} Var\left(\sum y_i\right)$$

$$= \frac{1}{n^2}\left[\sum_i Var(y_i) + 2\sum_i\sum_j cov(y_i, y_j)\right]$$

Muestreo Aleatorio Simple

13/Febrero/2021

Si es M.A.S. con reemplazo:

$$Var(\hat{\mu}) = \frac{1}{n^2}\left[\sum Var(y_i)\right]$$
$$= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Si es M.A.S. sin reemplazo:

$$Var(\hat{\mu}) = \frac{1}{n^2}\left[\sum Var(y_i) + 2\sum_i\sum_j cov(y_i, y_j)\right]$$
$$= \frac{1}{n^2}\left[n\sigma^2 + 2\sum_i\sum_j \frac{-\sigma^2}{N-1}\right]$$
$$= \frac{1}{n^2}\left[n\sigma^2 - \frac{\cancel{2}n(n-1)\sigma^2}{\cancel{2}(N-1)}\right]$$
$$= \frac{1}{n^2} \cdot n\sigma^2 \left[1 - \frac{n-1}{N-1}\right]$$
$$= \frac{\sigma^2}{n}\left[\frac{N-1-n+1}{N-1}\right] = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$$

Donde $\left[\frac{N-n}{N-1}\right]$, es el factor de corrección para población finitas. Si $N \to \infty$ es término vale 1

Definamos la cota para el error de estimación de la siguiente manera:

$$B = 2\sqrt{Var(\hat{\mu})} = 2\,DE(\hat{\mu})$$

Los límites de error en la estimación son:
- Límite inferior: $\hat{\mu} - B$
- Límite superior: $\hat{\mu} + B$

Finalmente, definimos el margen de error como:

$$M.E. = \frac{B}{\hat{\mu}} * 100\%$$

Nota
Si el margen de error se encuentra encima del 5.00% se recomienda seleccionar un nuevo tamaño de muestra.

Si desconocemos σ^2, proponemos un estimador insesgado:

$$\hat{Var}(\hat{\mu}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

Muestreo Aleatorio Simple

20|Febrero|2021

Demostración

$$E\left[\frac{s^2}{n}\left(\frac{N-n}{N-1}\right)\right] = \frac{1}{n}\cdot\frac{N-n}{N-1}E\left[\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}\right]$$

$$= \frac{1}{n}\cdot\frac{N-n}{N-1}\cdot\frac{1}{n-1}E\left[\sum_{i=1}^{n}(y_i-\bar{y})^2\right]$$

$$= \frac{1}{n}\cdot\frac{N-n}{N-1}\cdot\frac{1}{n-1}E\left[\sum_{i=1}^{n}(y_i-\mu)+(\mu-\bar{y})^2\right]$$

$$= \frac{1}{n}\cdot\frac{N-n}{N-1}\cdot\frac{1}{n-1}E\left[\sum_{i=1}^{n}(y_i-\mu)^2+2(y_i-\mu)(\mu-\bar{y})+(\mu-\bar{y})^2\right]$$

$$= \frac{1}{n(n-1)}\cdot\frac{N-n}{N-1}E\left[\sum_{i=1}^{n}(y_i-\mu)^2+2(\mu-\bar{y})\sum_{i=1}^{n}(y_i-\mu)+\sum_{i=1}^{n}(\mu-\bar{y})^2\right]$$

$$= \frac{1}{n(n-1)}\cdot\frac{N-n}{N-1}\left[\sum_{i=1}^{n}E(y_i-\mu)^2+E\left(2(\mu-\bar{y})\sum_{i=1}^{n}(y_i-\mu)\right)+E\left(n(\mu-\bar{y})^2\right)\right]$$

$$= \frac{1}{n(n-1)}\cdot\frac{N-n}{N-1}\left[n\sigma^2+2(\mu-\bar{y})E\left[\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}\mu\right]+n\,E[(\bar{y}-\mu)^2]\right]$$

$$= \frac{1}{(n-1)n}\cdot\frac{N-n}{N-1}\left[n\sigma^2+E\left(2n(\mu-\bar{y})(\bar{y}-\mu)+n(\bar{y}-\mu)^2\right]\right.$$

$$= \frac{1}{n(n-1)}\cdot\frac{N-n}{N-1}\left[n\sigma^2+E\left(-2n(\bar{y}-\mu)^2+n(\bar{y}-\mu)^2\right)\right]$$

$$= \frac{1}{n(n-1)}\cdot\frac{N-n}{N-1}\left[n\sigma^2-n\,E[(\bar{y}-\mu)^2]\right]$$

$$= \frac{1}{n(n-1)}\cdot\frac{N-n}{N-1}\left[n\sigma^2-\cancel{n}\frac{\sigma^2}{\cancel{n}}\left(\frac{N-n}{N-1}\right)\right]$$

$$= \frac{\sigma^2}{n(n-1)}\left(\frac{N-n}{N-1}\right)\left(n-\left(\frac{N-n}{N-1}\right)\right)$$

$$= \frac{\sigma^2}{n(n-1)}\left(\frac{Nn-\cancel{n}-N+\cancel{n}}{N-1}\right)\left(\frac{N-n}{N-1}\right)$$

$$= \frac{\sigma^2}{n(\cancel{n-1})}\left(\frac{N-n}{N-1}\right)\left(\frac{N(\cancel{n-1})}{N-1}\right)$$

$$E\left[\frac{s^2}{n}\left(\frac{N-1}{N-n}\right)\right] = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\left(\frac{N}{N-1}\right)$$

$\therefore$ no es insesgado

Por lo que:

$$E\left[\frac{s^2}{n}\left(\frac{N-n}{\cancel{N-1}}\right)\left(\frac{\cancel{N-1}}{N}\right)\right] = E\left[\frac{s^2}{n}\left(\frac{N-n}{N}\right)\right] = \hat{Var}(\bar{y})$$

$\therefore$ es insesgado

$\boxed{\therefore \frac{s^2}{n}\left(\frac{N-n}{N}\right) \text{ es un estimador de } \sigma^2}$

En este caso la cota para el error de estimación se puede obtener como:

$$\hat{B} = 2\sqrt{\hat{Var}(\hat{\mu})} = 2DE(\hat{\mu})$$

Los límites de error en la estimación se pueden estimar como:

- Límite inferior: $\hat{\mu} - \hat{B}$
- Límite superior: $\hat{\mu} + \hat{\beta}$

Se puede estimar el margen de error con la siguiente función:

$$M.E = \frac{\hat{B}}{\hat{\mu}} * 100\%$$

Muestreo Aleatorio Simple

20|Febrero|2021

Ejemplos

1. Un hospital de la localidad recibió una auditoría contable para determinar su situación financiera, para lo cual debería analizar las cuentas abiertas de los 1021 pacientes, sin embargo, por cuestiones de tiempo no podrían analizar las todas. Se decidió seleccionar una muestra aleatoria simple de 200 cuentas y se obtuvo que de esas 200 cuentas el hospital debera recibir un de \$1,844,400 pesos, con una varianza de 445.21 pesos. Estimar la deuda promedio por paciente.

$N = 1021$
$n = 200$
$\sum y_i = 1844400$
$s^2 = 445.21$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1844400}{200} = 9222$$, es deuda promedio

La varianza de la deuda promedio:

$$Var(\bar{y}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) = \frac{445.21}{200}\left(\frac{821}{1021}\right) = 1.789$$

La cota de error de estimación:

$$\hat{B} = 2\sqrt{Var(\bar{y})} = 2\sqrt{1.789} = 2.6758$$

Con límites en el error de estimación definidos en el intervalo de confianza:

$$[\hat{\mu} - \hat{B}; \hat{\mu} + \hat{B}] = [9222 - 2.6758, 9222 + 2.6758] = [9219.32, 9224.67]$$

Y un margen de error o error de muestreo es:

$$ME = \frac{\hat{B}}{\hat{\mu}} * 100\% = \frac{2.6758}{9222} * 100\% = 0.029\%$$

∴ La deuda promedio por paciente es de \$9222, esta deuda se encuentra en los rangos de 9219 y 9225 pesos

2. En ese mismo hospital se tienen 484 empleados y la administración desea saber las horas de trabajo efectivo semanales de todos los empleados. Para ello, se decide tomar una muestra aleatoria simple de 9 trabajadores y para cada uno de ellos se registra las horas de trabajo efectivo por semana.

Muestreo Aleatorio Simple

20|Febrero|2021

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
34	32	52	43	40	41	45	43	39

$N = 484$
$n = 9$

$$\sum y_i = 34+32+52+43+40+41+45+43+39 = 369$$

$$\bar{y} = \frac{\sum x_i}{n} = \frac{369}{9} = 41,$$ horas promedio trabajadas semanalmente por empleado

$$\hat{t} = N\bar{y} = 484(41) = 19844$$ horas trabajadas

Con una varianza estimada:

$$\hat{Var}(\hat{t}) = \frac{s^2 N}{n}(N-n) = \frac{35(484)}{9}(475) = 849055.556$$

y una cota en el error de estimación de:

$$\hat{B} = 2\sqrt{\hat{Var}(\hat{t})} = 2(945.545) = 1891.090$$

Lo que nos da un intervalo de confianza del 95%

$$IC = (\hat{t} - \hat{B}, \hat{t} + \hat{B}) = (17952.909, 21735.090)$$

Un margen de error o error de muestreo de:

$$ME = \frac{\hat{B}}{\hat{t}} \times 100\% = \frac{1891.090}{19844} \times 100 = 9.5297\%$$

$\therefore$ El total de horas trabajadas durante la semana es de 19844 horas, sin embargo el rango está entre 17952.909 y 21735.090 con 95% de confianza. Se recomienda una muestra más grande.

Estimación de un total poblacional a partir de un Muestreo Aleatorio Simple

Sean $y_1, y_2, \ldots, y_n$ variables aleatorias dependientes que siguen una distribución con media μ y varianza σ^2. Estimaremos $\hat{t} = \sum_{i=1}^{n} y_i = n\mu$, para ello utilizaremos:

$$\hat{t} = N\hat{\mu} = N\bar{y}$$

El cual tiene las siguientes propiedades:

1. $E(\hat{t}) = N\mu$

2. $Var(\hat{t}) = \begin{cases} \frac{N^2\sigma^2}{n}\left(\frac{N-n}{N-1}\right), & \text{si } \sigma^2 \text{ es conocida} \\ \frac{s^2 N}{n}(N-n), & \text{si } \sigma^2 \text{ es desconocida} \end{cases}$

Demostración

$$E(\hat{t}) = E[N\hat{\mu}] = NE[\hat{\mu}] = N\mu$$

$$Var(\hat{t}) = Var(N\hat{\mu}) = N^2 Var(\hat{\mu}) = \frac{N^2\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

Con una cota para el error de estimación obtenida por:

$$B = 2\sqrt{Var(\hat{t})} = 2DE(\hat{t})$$

Los límites de error en la estimación se pueden calcular como:
- Límite inferior: $\hat{t} - B$
- Límite superior: $\hat{t} + B$

El margen de error puede estimarse como:

$$M.E = \frac{B}{\hat{t}} * 100\%$$

Estimación de una proporción Poblacional a partir de un M.A.S.

Sean $y_1, y_2, \ldots, y_n$ variables aleatorias con distribución binomial con parámetros (n, p), media igual a np y varianza dada por $np(1-p)$. Para estimar p, utilizaremos:

$$\hat{p} = \frac{x}{n}, \text{ donde } x = \text{número de éxitos en la muestra}$$

también se puede estimar:

$$\hat{p} = \sum_{i=1}^{n} y_i, \text{ donde } y_i \begin{cases} 1, & \text{si es cierto} \\ 0, & \text{si no} \end{cases}$$

Posee las siguientes propiedades:

1. $E(\hat{p}) = p$

2. $Var(\hat{p}) = \begin{cases} \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right), & \text{si } \sigma^2 \text{ es conocida} \\ \dfrac{s^2}{n-1}\left(\dfrac{N-n}{N}\right), & \text{si } \sigma^2 \text{ es desconocida} \end{cases}$

Demostración

$$E[\hat{p}] = E[X/n] = \frac{1}{\cancel{n}}\cancel{n}p = p$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(x) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$, sea $\sigma^2 = p(1-p)$ y

agregando el factor para correciones finitas

$$Var(\hat{p}) = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$$

Con una cota para el error de estimación dada por:

$$B = 2\sqrt{Var(\hat{p})} = 2DE(\hat{p})$$

Los límites de error en la estimación (Intervalo de Confianza al 95%), se pueden calcular como
- Límite inferior: $\hat{p} - B$
- Límite superior: $\hat{p} + B$

El margen de error puede estimarse como:

$$M.E: \frac{B}{\hat{p}} * 100\%$$

La estimación de σ^2 es:

$$\sigma^2 = \hat{p}\hat{q} = \hat{p}(1-\hat{p})$$

Muestreo Aleatorio Simple

27/Febrero/2021

Selección del tamaño de muestra

Para seleccionar un tamaño de muestra adecuado se emplean las dos fórmulas siguientes:

$$n = \frac{N\sigma^2}{(N-1)D+\sigma^2} \quad \text{ó} \quad n = \frac{N\sigma^2}{(N-1)D+\hat{p}\hat{q}}$$

donde D se estima de la siguiente manera:

- Si se desea estimar una media o una proporción:

$$D = \frac{B^2}{4}$$

- Si se desea estimar un total:

$$D = \frac{B^2}{4N^2}$$

Si desconocemos σ^2 se puede estimar de diferentes maneras:

- Si conocemos la proporción:

$$\sigma^2 = \hat{p}\hat{q} = \hat{p}(1-\hat{p})$$

- Si conocemos el rango de variabilidad:

$$\sigma^2 = \frac{R^2}{16}$$

- Si no conocemos algunos de los datos anteriores:

$$\sigma^2 = s^2$$

Ejemplo

Se desea estimar la cantidad promedio de dinero de las cuentas por cobrar de un hospital. De estudios previos se sabe que la varianza de las deudas es de un rango de variación de 1000. Se sabe que hay 1021 cuentas abiertas y se desea un error máximo de $30 pesos en la estimación. Si se va a utilizar un muestreo aleatorio simple, ¿cuántas cuentas abiertas hay que verificar?

$N = 1021$
$B = 30$
$R = 1000$

$$D = \frac{(30)^2}{4} = \frac{900}{4} = 225$$

Muestreo Aleatorio Simple

27/Febrero/2021

Se desconoce σ^2 pero se conoce R:

$$\sigma^2 = \frac{R^2}{16} = \frac{(1000)^2}{16} = \frac{1000000}{16} = 62500$$

Estimando n:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} = \frac{(1021)(62500)}{(1021-1)(225)+62500}$$

$$= \frac{63812500}{(1020)(225)+62500} = \frac{63812500}{229500+6250}$$

$$= \frac{63812500}{292000} = 218.5359589 \approx 219$$

$\therefore$ Se deben verificar 219 cuentas para hacer un muestreo apropiado