

# Muestreo Estadístico: Muestreo Aleatorio Simple con y sin Reemplazamiento \*

García Prado, Sergio  
sergio@garciparedes.me

8 de octubre de 2017

## 1. Demostraciones

A continuación se demuestran distintas propiedades relacionadas con los estadísticos referidos al muestreo aleatorio simple con y sin reemplazamiento.

### 1.1. Probabilidad de inclusión de primer y segundo nivel en muestreo aleatorio simple sin reemplazamiento (m.a.s)

La probabilidad de inclusión en el muestreo aleatorio simple sin reemplazamiento viene condicionada por dos factores diferentes: (a) el número de posibles muestras que incluyan el elemento  $k$ -ésimo determinado por  $\binom{N-1}{n-1}$  y el de posibles muestras que incluyan los elementos  $k$  y  $l$  tales que  $k \neq l$  determinado por  $\binom{N-2}{n-2}$  así como, (b) la probabilidad de elegir una determinada muestra de tamaño  $n$  de entre todas las posibles, definida como  $\frac{1}{\binom{N}{n}}$ .

Nótese que dichos valores provienen del área de teoría combinatoria, donde se cumple la siguiente propiedad: “Sean  $n$  elementos distintos y  $k \leq n$ . A las distintas agrupaciones no ordenadas de  $k$  elementos elegidos entre  $n$  distintos se las denomina combinaciones de  $n$  elementos tomados en grupos de  $k$ , y el número de dichas combinaciones es  $\binom{n}{k}$ ” [LMPMSB14]

Por tanto, la probabilidad de inclusión queda determinada por el sumatorio de las probabilidades de todas aquellas muestras que incluyen al elemento  $k$  para el caso del primer nivel y de los elementos  $\{k, l\}$ , con  $k \neq l$  para el segundo nivel:

$$\pi_k = \sum_{k \in s} p(s) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} \frac{n}{N} \quad (1)$$

$$\pi_{kl} = \sum_{\{k, l\} \in s} p(s) = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)} \quad (2)$$

### 1.2. Probabilidad de inclusión de primer nivel en muestreo aleatorio con reemplazamiento

Para el caso del muestreo aleatorio simple con reemplazamiento, la probabilidad de inclusión de primer nivel se define de manera diferente respecto de la del caso sin reemplazamiento. Esto se debe a que una observación  $k$  puede aparecer más de una vez en una determinada muestra  $s$ .

---

\*URL: <https://github.com/garciparedes/statistical-sampling-simple-random-sample>

En este caso, para probar la probabilidad de inclusión del elemento  $k$ -ésimo en la muestra nos apoyaremos en la distribución *Binomial*, que modeliza la probabilidad de selección con reemplazamiento de un elemento de entre  $N$  posibles con probabilidad  $p$ . En este caso escogeremos  $p = \frac{1}{N}$ , que es la probabilidad de selección de una observación en la muestra.

Por tanto, definiremos la variable  $X \sim B(n, \frac{1}{N})$ . Para probar la probabilidad de que el elemento sea seleccionado al menos una vez, necesitamos conocer  $P(X \geq 1) = P(X = 1) + P(X = 2) + \dots + P(X = n)$ . Sin embargo, esta probabilidad puede ser calculada de manera sencilla como  $P(X \geq 1) = 1 - P(X = 0)$ . Puesto que  $P(X = 0) = \binom{n}{0} (\frac{1}{N})^0 (1 - \frac{1}{N})^{n-0} = (1 - \frac{1}{N})^n$ . Dado que la probabilidad de inclusión  $\pi_k = P(X \geq 1)$  por definición, entonces:

$$\pi_k = 1 - \left(1 - \frac{1}{N}\right)^n \quad (3)$$

### 1.3. Varianza del $\pi$ -estimador para proporciones en muestreo aleatorio simple sin reemplazamiento (m.a.s)

Para la demostración de  $\widehat{Var}(\widehat{P}_\pi)$  nos apoyaremos en distintas propiedades tanto de la varianza, que utilizaremos para reducir dicha definición a la varianza del total poblacional, definido tal y como se indica en la ecuación (4), que supondremos como cierto.

$$Var(\widehat{\tau}_\pi) = \frac{N^2(1-f)}{n} \sigma^2 \quad (4)$$

El primer paso será por tanto obtener una estimación de dicho valor, para lo cual se utilizará la cuasi-varianza muestral tal y como se indica en la ecuación (??):

$$\widehat{Var}(\widehat{\tau}_\pi) = \frac{N^2(1-f)}{n} S^{*2} = \frac{N^2(1-f)}{n-1} S^2 \quad (5)$$

A partir de dicha ecuación, lo siguiente es añadir la división entre  $N$  al estimador del total poblacional, para obtener un promedio, lo cual produce el cuadrado de dicho valor en la varianza, por propiedades de la misma. Dicha transformación se muestra en la ecuación (6)

$$\widehat{Var}\left(\frac{\widehat{\tau}_\pi}{N}\right) = \widehat{Var}(\widehat{\mu}) = \frac{N^2(1-f)}{nN^2} S^{*2} = \frac{(1-f)}{n-1} S^2 \quad (6)$$

Es trivial probar que  $\mu = P$  cuando la variable estudiada es dicotómica, por tanto, se puede seguir dicho paso sin más demostraciones. A partir de dicha presuposición de dicotomía, entonces se sabe que  $S^2 = \widehat{P}_\pi(1 - \widehat{P}_\pi)$  puesto que en dicho caso la variable sigue una distribución de *Bernoulli*. Entonces queda probada la ecuación (7).

$$\widehat{Var}(\widehat{P}_\pi) = \frac{1-f}{n-1} \widehat{P}_\pi(1 - \widehat{P}_\pi) \quad (7)$$

### 1.4. Intervalo de confianza del $\pi$ -estimador para el total poblacional en muestreo aleatorio simple sin reemplazamiento (m.a.s)

En este caso se pretende probar el intervalo de confianza de nivel  $1 - \alpha$  para la distribución de probabilidades referida a la variable aleatoria  $\tau$ , es decir, a la suma del total poblacional para la variable  $Y$ . Es decir, se pretende probar la igualdad descrita en la ecuación (8). Dicha demostración se llevará a cabo siguiendo la notación de valores de probabilidad.

$$\left[ \hat{\tau}_\pi \pm z_{1-\alpha/2} \sqrt{\widehat{Var}[\hat{\tau}_\pi]} \right] \equiv \Pr \left( \tau \in \left[ \hat{\tau}_\pi \pm z_{1-\alpha/2} \sqrt{Var[\hat{\tau}_\pi]} \right] \right) = 1 - \alpha \quad (8)$$

Un paso previo para la demostración del intervalo de confianza es la demostración de la ecuación (9). Por el *Teorema Central del Límite* se puede asumir que la distribución  $\tau$  sigue una distribución normal con media  $\hat{\tau}$  y varianza  $Var[\hat{\tau}_\pi]$ . Por tanto, este puede ser tipificado a una distribución normal estándar. Esta propiedad se indica en la ecuación (9)

$$\frac{\hat{\tau}_\pi - \tau}{\sqrt{Var[\hat{\tau}_\pi]}} \sim N(0, 1) \quad (9)$$

Por tanto, para demostrar que el intervalo de confianza sigue un nivel  $1 - \alpha$  basta con aplicar distintas propiedades algebraicas sobre este, hasta llegar a la distribución normal estándar tal y como se indicó en la ecuación (9), que debido a su simetría permite sustituir el intervalo multiplicando por 2 el valor de  $\alpha$ . Por último, debido a las propiedades de la distribución normal estándar, queda demostrado que en hasta el percentil  $z_{1-\alpha}$  se localiza la probabilidad  $1 - \alpha$ .

$$\Pr \left( \tau \in \left[ \hat{\tau}_\pi \pm z_{1-\alpha/2} \sqrt{Var[\hat{\tau}_\pi]} \right] \right) = \quad (10)$$

$$= \Pr \left( \tau - \hat{\tau}_\pi \in \left[ \pm z_{1-\alpha/2} \sqrt{Var[\hat{\tau}_\pi]} \right] \right) \quad (11)$$

$$= \Pr \left( \frac{\tau - \hat{\tau}_\pi}{\sqrt{Var[\hat{\tau}_\pi]}} \in \left[ \pm z_{1-\alpha/2} \right] \right) \quad (12)$$

$$= \Pr \left( \frac{\hat{\tau}_\pi - \tau}{\sqrt{Var[\hat{\tau}_\pi]}} \notin \left[ \pm z_{1-\alpha/2} \right] \right) \quad (13)$$

$$= \Pr \left( N(0, 1) \notin \left[ \pm z_{1-\alpha/2} \right] \right) \quad (14)$$

$$= \Pr \left( N(0, 1) \geq z_{1-\alpha} \right) \quad (15)$$

$$= 1 - \alpha \quad (16)$$

## Referencias

- [LPMMSB14] Esperanza Larrinaga Miner, María Felisa Pérez Martínez, and Araceli Suárez Barrio. Grado en ingeniería informática, uva: Matemática discreta. 2014.
- [SSW03] Carl-Erik Särndal, Bengt Swensson, and Jan Wretman. *Model assisted survey sampling*. Springer Science & Business Media, 2003.
- [TG18] Jesús Alberto Tapia García. Muestreo Estadístico 1, 2017/18.