1 Introducción

1.1 Estimación de p en población infinita

Para estimar una proporción (un $p_{exito} = p$), se toma una muestra de tamaño n, y si X : "cantidad de éxitos en la muestra", entonces la estimación y su correspondiente desvío estándar es

$$\widehat{p} = \frac{X}{n} \quad \text{con} \quad s_{\widehat{p}} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

1.2 Estimación de p en población finita

Ahora se tiene un **lote** de N elementos entre los cuales hay e exitos. Obviamente $p = \frac{e}{N}$. Si se toma una muestra(sin reemplazo) de tamaño n, y si X : "cantidad de éxitos en la muestra", entonces la estimación y su correspondiente desvío estándar es

$$\widehat{p} = \frac{X}{n} \quad \text{con} \quad s_{\widehat{p}} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\sqrt{1-\frac{n}{N}}$$

Remark 1 *El factor $\sqrt{1-\frac{n}{N}}$ se lo llama factor de corrección por población finita (**cpf**). Este factor hace disminuir aún más el desvío del estimador al aumentar el tamaño de la muestra. Notar que cuando se muestrea toda la población ($n = N$), resulta **cpf** =0, y entonces $s_{\widehat{p}} = 0$, como debe ser, ya que en este caso la estimación coincide con el verdadero p.*

2 Ejemplo de uso

Supongase que se tiene una máquina que produce artículos en serie, siendo la aparición de artículos buenos, sucesos independientes con probabilidad p. Se trataría entonces de una población infinita.

Supongamos que se han fabricado 100 artículos, y que se encuentran en depósito.

Yo de esos 100, tomo una muestra al azar de tamaño 70 y los inspecciono, encontrando 56 buenos.

2.1 Caso A

Me piden una estimación del p de la máquina.

Yo lo estimo con la muestra, y para el desvío del estimador **NO** uso la corrección por población finita. Ya que, aunque los 70 los saqué de los 100, **mi interés reside en el parámetro p de la máquina**.

$$\widehat{p} = \frac{X}{n} = \frac{56}{70} = 0.80 \quad \text{con } s_{\widehat{p}} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \sqrt{\frac{0.8(1-0.8)}{70}} = 0.048$$

2.2 Caso B

Un cliente quiere comprarme todos los artículos buenos que tenga, y yo decido venderle, los buenos que están en el lote de los 100.

Necesito estimar la proporción de buenos que tendré **en el lote**. Para ello uso la muestra de 70, pero para calcular el desvío, **SI** uso la corrección por población finita, ya que ahora **mi interés reside en el parámetro** p **del lote**.

$$\widehat{p} = \frac{X}{n} = \frac{56}{70} = 0.80 \quad \text{con} \quad s_{\widehat{p}} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\sqrt{1-\frac{n}{N}} = \sqrt{\frac{0.8(1-0.8)}{70}}\sqrt{1-\frac{70}{100}} = 0.026$$

(En este caso la cantidad total de buenos estimada será $100\widehat{p} = 80$, con desvío $100 s_{\widehat{p}} = 2.6$).

3 Otro ejemplo

Aquí se estudió la presencia de una enfermedad estacional en el mes de Julio/05 en dos localidades. Se relevaron **totalmente** las dos localidades, o sea se trata de un **censo**, obteniendo:

	N habitantes	N enfermos	tasa(por 1000hs)
Loc. 1	100	1	10
Loc. 2	1.000.000	2000	2

3.1 Caso C

Si el interés primario de un funcionario es estimar la cantidad de dosis de vacuna a entregar a esas localidades, es suficiente esta información. Aquí, al tratarse de toda la población de cada localidad, es valido considerar las **cpf**=0, y los desvíos estándar serían nulos. La información presentada es válida para Julio/05.

3.2 Caso D

Si soy un médico epidemiólogo, y quiero evaluar la presencia de brotes infecciosos en esas localidades en otros años, **mi interés es el brote infeccioso.** Por más que la muestra sea la totalidad, yo debo pensar mi problema como en el caso A, y calcular desvíos de tasas, sin usar corrección. Para el epidemiólogo el brote infeccioso equivale a la máquina, y el piensa como "muestras" a los 100, y 1.000.000 habitantes de las localidades(aunque sean la totalidad). Por eso aquí **sí** corresponde calcular desvíos, sin considerar corrección por población finita. Se tendrá entonces, considerando las "tasa(por 1000hs)" = $1000\widehat{p}$:

- Para Loc. 1

$$\widehat{t} = 1000\frac{1}{100} = 10 \quad \text{con} \quad s_{\widehat{t}} = 1000\sqrt{\frac{0.01(1-0.01)}{100}} = 9.95$$

- Para Loc. 2

$$\widehat{t} = 1000\frac{2000}{1000000} = 2 \quad \text{con} \quad s_{\widehat{t}} = 1000\sqrt{\frac{0.002(1-0.002)}{1000000}} = 0.047$$

O sea en el próximo año la tasa para la localidad 1 será de alrededor de 10, pudiendo variar bastante ya que $s_{\widehat{t}} = 9.95$. En cambio para la localidad 2 se espera una tasa muy cercana a 2.