Chapter 1

Concepto de población, marco y muestra. Muestreo probabilístico. Distribución de un estimador en el muestreo. Error cuadrático medio y sus componentes. Intervalos de confianza: Estimadores insesgados y sesgados. Métodos de selección. Probabilidad de la unidad de pertenecer a la muestra y propiedades. Comparación con el muestreo no probabilístico: Muestreo por cuotas.

1.1. Concepto de población, marco y muestra.

Cuando queremos estudiar una serie de características en una población podemos abordar el estudio de dos formas:

- Observando dichas características en todos los componentes de esa población. En ese caso se dice que estamos realizando un censo. En muchas ocasiones realizar censos no es posible, bien por motivos de coste, porque para observar un elemento de la población haya que destruirlo, porque la población sea infinita o por otras causas.
- Observando un conjunto de elementos de la población y estrapolando la información qe obtengamos al resto de la población. En este caso se dice que estamos realizando una encuesta. Al conjunto de técnicas matemáticas para seleccionar ese subconjunto de la población de forma que sea representativo de la misma se le llama muestreo. A las técnicas para deducir información acerca de la población a través de un subconjunto de la misma se le llama inferencia estadística.

Población objetivo: Colección de elementos sobre los que se desea investigar una determinada característica.

Población investigada: Subconjunto de la población objetivo que realmente es objeto de estudio. La parte no estudiada puede ser por no poder acceder a ella, negativas, etc.

Unidad elemental de muestreo: Cada elemento de la población investigada.

Unidad de muestreo compuesta: Conjunto de varias unidades elementales que se investigan conjuntamente.

Marco: Listado de todas las unidades de muestreo, que se utilizará para seleccionar la muestra. Lo ideal es que el marco coincida con la población objetivo, pero puede no ser así debido a errores, falta de actualizaciones, omisiones y otras causas. La diferencia entre el marco y la población objetivo debe ser lo suficientemente pequeï¿œa como para que las inferencias que se hagan sobre la población sean válidas.

Muestra: Colección o conjunto de unidades de muestreo seleccionada del marco correspondiente a la población que se quiere investigar.

Tamaï¿œo muestral: Número de elementos que componen la muestra.

Se llama **marco en sentido amplio** a aquel que además de incluir el listado de las unidades de la población a investigar incluye información adicional sobre las mismas. Esta información puede utilizarse para mejorar el diseï¿œo de los procesos de muestreo. Como ejemplos de información adicional podemos citar variables auxiliares que tengan alguna correlación con la variable en estudio, o el conocimiento de estimaciones de características de la población provenientes de una encuesta anterior.

Si nuestro marco consta de unidades de muestreo compuestas, y tenemos listados parciales de las unidades simples que componen cada unidad compuesta ten, se dice que dispondremos de marcos múltiples.

Si el muestreo se hace de forma que se conoce la probabilidad de obtener cada una de las muestras posibles, se dice que el muestreo es probabilístico. Si la población objetivo es finita, se habla de muestreo en poblaciones finitas.

1.2. Muestreo probabilístico.

Un método de muestreo es un mecanismo mediante el que seleccionaremos la muestra a partir de la población a investigar.

El método de muestreo que hayamos definido se llama muestreo probabilístico si cumple las siguientes condiciones:

- Podemos definir el conjunto de muestras distintas posibles que generará el método de muestreo aplicado a una población específica. Esto quiere decir que podemos especificar las unidades de la población que pertenecen a cada muestra de las posibles.
- Cada muestra posible tiene asignada una probabilidad de selección.

Es fácil comprobar que en este supuesto, el procedimiento de muestreo aplicado a una población es un fenómeno aleatorio probabilizable. En general sólo se considerarán métodos de muestreo en los que no haya ninguna muestra con probabilidad nula, es decir, métodos de muestreo no restringidos.

En la práctica, rara vez se define un procedimiento mediante todas las muestras posibles y sus probabilidades de obtención, ya que esto sería muy laborioso a poco que la población investigada fuese grande. En su lugar, y de forma equivalente, se define el procedimiento de muestreo asignando a cada unidad de muestreo la probabilidad de que esté incluida en la muestra.

Formalmente, el procedimiento de muestreo será probabilístico si, siendo S el conjunto de todas las posibles muestras a obtener mediante nuestro procedimiento de muestreo $\{s_1, s_2, \ldots, s_n\}$y siendo $\mathcal{F}$ la familia de todos los subconjuntos posibles de S, forman un espacio probabilizable, y además las función que asigna a cada muestra posible su probabilidad de obtención es una medida de probabilidad sobre este espacio.

1.3. Distribución de un estimador en el muestreo.

Una vez hemos definido un procedimiento de muestro probabilístico, debemos construir un estimador que nos permita inferir a partir de la muestra seleccionada las características poblacionales a investigar (total, media,

proporción, etc). Estos estimadores serán por tanto variables aleatorias cuya distribución de probabilidad será función de las probabilidades de las muestras.

Más formalmente, sea la característica X de los elementos de la población U, que toma el valor X_i para cada unidad U_i de la población. Consideramos ahora una función θ de los valores X_i, a la que llamaremos parámetro poblacional. Seleccionamos una muestra aleatoria s de nuestra población, y a partir de los valores de la característica en los elementos de la muestra queremos obtener una estimación de θ para el total de la población, mediante una función $\hat{\theta}$ basada en los valores que toma la característica en los elementos de la muestra.

A la función $\hat{\theta}$ que asocia cada muestra a un valor numérico se le llama estimador del parámetro poblacional θ, y alos valores de dicha función para cada muestra se les llama estimaciones. Así, podemos formalizar el estimador como una aplicación medible del espacio de todas las muestras posibles a $\mathbb{R}$,$\hat{\theta} : S \to \mathbb{R}$ y por tanto se puede definir $\hat{\theta}$ como una variable aleatoria sobre la recta real.A la función de probabilidad inducida por la variable aleatoria $\hat{\theta}$ se la llama distribución de probabilidad en el muestreo del estimador $\hat{\theta}$. Por tanto, sea $T = \left\{t \in \mathbb{R}/\exists\left(X_1,\ldots,X_n\right) \in S\left(X\right), \hat{\theta}\left(X_1,\ldots,X_n\right) = t\right\}$. El subconjunto T de $\mathbb{R}$ constituye el conjunto de valores del estimador. Se define la ley de probabilidad del estimador como $P\left(\hat{\theta}\left(X_1,\ldots,X_n\right) = t\right) = \sum_{\left\{S_i/\hat{\theta}(S_i(X))=t\right\}} P\left(S_i\right)$, es decir, la probabilidad de cada valor del estimador es igual a la suma de las probabilidades de obtener cada muestra que origine ese mismo valor del estimador. Por tanto, la distribución de probabilidad del estimador en el muestreo será el conjunto de pares formados por todos los valores posibles del estimador y las probabilidades de que el estimador tome esos valores.

Frecuentemente se expresa un procedimiento de estimación como el conjunto formado por el marco, el procedimiento de muestreo y el estimador utilizado. Al conjunto formado por las muestras posibles, sus probabilidades y el estimador se le llama diseï¿œo muestral.

A partir de la definición de muestreo probabilístico y de distribución de un estimador en el muestreo, podemos utilizar las herramientas de la inferencia estadística para obtener propiedades de los estimadores.

1.4. Error cuadrático medio y sus componentes.

Dado que nuestro estimador es una variable aleatoria, podemos definir para el mismo todas las propiedades asignadas a las variables aleatorias.

Definimos la esperanza matemática o media del estimador $\hat{\theta}$ del parámetro poblacional θ como $E\left(\hat{\theta}\right) = \sum_s \hat{\theta}\left(s_i\right) P\left(s_i\right) = \sum_{\mathbb{R}} tP\left(\hat{\theta} = t\right)$.

Definimos la varianza del estimador $\hat{\theta}$ del parámetro poblacional θ como $V\left(\hat{\theta}\right) = \sigma^2_{\hat{\theta}} = E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2\right] = \sum_{\mathbb{R}} \left(t - E\left(\hat{\theta}\right)\right)^2 P\left(\hat{\theta} = t\right) = E\left(\hat{\theta}^2\right) - \left(E\left(\hat{\theta}\right)\right)^2$. Es una medida de la concentración de los valores del estimador en torno a su valor medio.

Definimos el error de muestreo o desviación típica del estimador como la raíz cuadrada de su varianza.

Definimos el error relativo de muestreo, o coeficiente de variacióndel estimador, como su desviación típica dividido por su esperanza.

Definimos el sesgo del estimador $\hat{\theta}$ del parámetro poblacional θ como la diferencia entre el valor esperado del estimador y el valor real del parámetro: $B\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$. Se dice que el estimador es insesgado si su sesgo es cero, y sesgado si su sesgo es distinto de cero. El estimador es consistente cuando su sesgo tiende a cero al aumentar el tamaï¿œo de la muestra.

Definimos acuracidad o error cuadrático medio del estimador $\hat{\theta}$ del parámetro poblacional θ como $ECM\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - \theta\right)^2\right]$.

La precisión de un estimador se analiza en función de su error de muestreo, su error cuadrático medio y su sesgo. El error cuadrático medio se puede descomponer de la siguiente forma:

$$ECM\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right) + E\left(\hat{\theta}\right) - \theta\right)^2\right] = E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right) + B\left(\hat{\theta}\right)\right)^2\right] =$$
$$= E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2\right] + \left(B\left(\hat{\theta}\right)\right)^2 + 2B\left(\hat{\theta}\right) E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right) = Var\left(\hat{\theta}\right) + \left(B\left(\hat{\theta}\right)\right)^2 = \sigma^2_{\hat{\theta}} + \left(B\left(\hat{\theta}\right)\right)^2$$

Así, el error cuadrático medio se puede descomponer como la suma de la varianza del estimador más el cuadrado de su sesgo, o, expresado de otra forma, la suma de los cuadrados del error de muestreo y del sesgo. En general, es conveniente que el error cuadrático medio sea lo más pequeï¿œo posible. Es deseable obtener estimadores insesgados, siempre que eso no implique un aumento de su varianza. Asímismo, es deseable obtener estimadores de varianza pequeï¿œa, siempre que eso no implique la aparición de un sesgo. En la práctica, se puede admitir el uso de estimadores sesgados siempre que el cociente $\left|\frac{B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}}\right|$ sea igual o menor que $0,1$. Para poder comparar distintos estimadores, y elegir el más conveniente para estimar nuestro parámetro, debemos distinguir entre estimadores sesgados e insesgados.

1.4.1. Comparación de estimadores insesgados.

Si un estimador es insesgado, su error cuadrático medio coincide con su varianza. Por tanto, para comparar varios estimadores insesgados del mismo parámetro, hay que considerar sus errores de muestreo, siendo mejor el estimador cuyo error de muestreo sea menor. Además, el error relativo sólo varía en función del error de muestreo, por tanto podemos hacer depender la decisión sólo del error de muestreo.

1.4.2. Comparación de estimadores sesgados.

Si un estimador es sesgado, la magnitud para analizar su precisión es su error cuadrático medio. Por lo tanto, para comparar varios estimadores sesgados en cuanto a precisión se utiliza el ECM, y será mejor el que menor ECM presente.

Pero en la práctica el cálculo del error cuadrático medio puede ser problemático, e incluso su valor puede variar dependiendo del valor del parámetro que se investiga. Es por esto que se utiliza el siguiente razonamiento:

Sabemos que $ECM\left(\hat{\theta}\right) = \sigma^2_{\hat{\theta}} + \left(B\left(\hat{\theta}\right)\right)^2$, y por tanto, podemos formar un triángulo rectángulo en el que $\sigma_{\hat{\theta}}$ y $B\left(\hat{\theta}\right)$ sea la longitud de los catetos, y $\sqrt{ECM\left(\hat{\theta}\right)}$ la de la hipotenusa. Por tanto, la tangente del ángulo que forma la hipotenusa con el cateto de longitud $\sigma_{\hat{\theta}}$, α, será: $\tan\alpha = \frac{B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}}$, que es la contribución del sesgo y la desviación típica al error cuadrático medio. Por tanto, cuanto menor sea este cociente, menor será la contribución del sesgo al ECM.

Por esta razón, se calcula el cociente $\left|\frac{B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}}\right|$ para cada estimador a comparar, siendo más preciso el estimador que presenta una relación del sesgo al error de muestreo más pequeï¿œa. También se puede utitlizar el coeficiente de variación, siendo el mejor estimador el que menor coeficiente de variación presente.

Si los estimadores a comparar tienen un sesgo despreciable, es decir, $\left|\frac{B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}}\right| < \frac{1}{10}$ se consideran insesgados.

1.5. Estimación por intervalos de confianza.

Si realizamos una afirmación acerca de un parámetro poblacional a partir de la información contenida en una muestra, lo podemos hacer basándonos e el valor puntual de un estadístico basado en la misma, o bien mediante un intervalo de confianza. Un intervalo de confianza al nivel de confianza α es un intervalo real para el que hay una probabilidad $1-\alpha$ de que contenga al valor real del parámetro. Al valor $1-\alpha$ se le suele llamar coeficiente de confianza. Veremos cómo utilizar estimadores para hallar intervalos de confianza.

1.5.1. Intervalos de confianza si el estimador es insesgado.

Se trata de construir un intervalo de confianza para el parámetro θ mediante un estimador insesgado del mismo, es decir, $E\left(\hat{\theta}\right) = \theta$.

1.5.1.1. El estimador tiene una distribución normal.

En ese caso, si $\hat{\theta} \sim N\left(\theta, \sigma^2_{\hat{\theta}}\right)$, entonces $E\left(\hat{\theta}\right) = \theta$, y $\dfrac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \sim N(0,1)$ y se puede calcular λ_α tal que $P\left[-\lambda_\alpha \leq \dfrac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] = 1-\alpha$

$$
\begin{aligned}
P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] - P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq -\lambda_\alpha\right] &= 1-\alpha \\
P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] - \left\{1 - P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \geq -\lambda_\alpha\right]\right\} &= 1-\alpha \\
P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] - \left\{1 - P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right]\right\} &= 1-\alpha \\
2P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] - 1 &= 1-\alpha \\
P\left[\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] &= 1-\frac{\alpha}{2}
\end{aligned}
$$

por tanto, λ_α es el valor que hace que la función de distribución de la normal estándar valga $1-\dfrac{\alpha}{2}$, y el intervalo de confianza será:

$$\hat{\theta} - \lambda_\alpha \sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + \lambda_\alpha \sigma_{\hat{\theta}}$$

Lo normal es no conocer el valor de $\sigma_{\hat{\theta}}$, sino de una estimación del mismo con datos muestrales conocidos. En estos casos no podemos asegurar con exactitud que el intervalo de confianza es como hemos visto. Para estos casos, podemos usar la distribución t de Student con $n-1$ grados de libertad para calcular el intervalo de confianza. En este caso, tenemos que $\dfrac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \sim t_{n-1}$, y de forma similar, el intervalo de confianza será $\hat{\theta} - \lambda_\alpha \hat{\sigma}_{\hat{\theta}} \leq \theta \leq \hat{\theta} + \lambda_\alpha \hat{\sigma}_{\hat{\theta}}$ solo que en este caso λ_α es el valor que hace que la función de distribución t de Student con $n-1$ grados de libertad valga $1-\dfrac{\alpha}{2}$.

1.5.1.2. El estimador no tiene una distribución normal.

En este caso utilizamos la desigualdad de Tchebichev, que dice:

$$P\left\{\left|\hat{\theta} - E\left(\hat{\theta}\right)\right| < k\right\} \geq 1 - \frac{\sigma^2_{\hat{\theta}}}{k^2} \quad \forall k > 0$$

Como $\hat{\theta}$ es insesgado para θ, $E\left(\hat{\theta}\right) = \theta$ y $P\left\{\left|\hat{\theta}-\theta\right| < k\right\} \geq 1 - \dfrac{\sigma^2_{\hat{\theta}}}{k^2}$. Para un nivel de significación α tomamos $k = \dfrac{\sigma_{\hat{\theta}}}{\sqrt{\alpha}}$, y entonces $P\left\{\left|\hat{\theta}-\theta\right| < \dfrac{\sigma_{\hat{\theta}}}{\sqrt{\alpha}}\right\} \geq 1-\alpha$, y el intervalo será $\hat{\theta} - \dfrac{\sigma_{\hat{\theta}}}{\sqrt{\alpha}} \leq \theta \leq \hat{\theta} + \dfrac{\sigma_{\hat{\theta}}}{\sqrt{\alpha}}$. En general, este intervalo es más ancho que si la distribución es normal, por lo que la propiedad de normalidad es bastante deseable.

1.5.2. Estimadores sesgados.

Si el estimador $\hat{\theta}$ es sesgado para θ, $B\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta \neq 0$. Por el teorema central del límite, si la muestra es suficientemente grande, se cumple $\dfrac{\hat{\theta} - E\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}} \sim N(0,1)$, y por tanto,

$$P\left[-\lambda_\alpha \leq \frac{\hat{\theta} - E\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] = 1-\alpha \Rightarrow P\left[-\lambda_\alpha \leq \frac{\hat{\theta} - \theta - B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] = 1-\alpha \Rightarrow$$
$$\Rightarrow P\left[-\lambda_\alpha \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} - \frac{B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}} \leq \lambda_\alpha\right] = 1-\alpha$$

y en el cálculo del intervalo aparece en término$\dfrac{B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}}$. Como ya sabemos, si $\left|\dfrac{B\left(\hat{\theta}\right)}{\sigma_{\hat{\theta}}}\right| < \dfrac{1}{10}$ la influencia del sesgo es despreciable con lo que el intervalo de confianza es el mismo que para estimadores insesgados. Si esto no se cumple, el sesgo influirá en el intervalo de confianza, y se tiene que el intervalo es $\hat{\theta} - \lambda_\alpha \hat{\sigma}_{\hat{\theta}} - B\left(\hat{\theta}\right) \leq \theta \leq \hat{\theta} + \lambda_\alpha \hat{\sigma}_{\hat{\theta}} - B\left(\hat{\theta}\right)$. Ahora el intervalo no está centrado en $\hat{\theta}$, está desplazado en una cantidad $B\left(\hat{\theta}\right)$ respecto al intervalo sin sesgo. Si queremos centrar el intervalo en torno a $\hat{\theta}$ tenemos que tomar el peor de los casos, y alargar el extremo del intervalo más pórximo a$\hat{\theta}$ hast conseguir un intervalo simétrico. Por tanto, el intervalo será más largo que para un estimador sin sesgo.

1.6. Métodos de selección. Probabilidad de la unidad de pertenecer a la muestra y propiedades.

Veremos ahora las distintas formas en las que se pueden seleccionar las unidades que pertenecen a la muestra. Inicialmente se pueden clasificar en dos grandes clases: métodos de muestreo con reposición y métodos de muestreo sin reposición.

Un procedimiento aleatorio de muestreo es sin reposición si todas las muestras que tienen algún elemento repetido son imposibles. Es decir, las unidades seleccionadas no se devuelven a la población para seleccionar la siguiente unidad de la muestra, de ahí el nombre. Como norma general, no se tiene en cuenta el modo de colocar los elementos en la muestra, es decir, muestras con los mismos elementos en distinto orden son consideradas iguales. Por tanto, el espacio muestral contiene $C_{N,n} = \binom{N}{n}$ muestras de tamaï¿œo n distintas posibles en una población de tamaï¿œo N. Si las muestras con los mismos elementos colocados en distinto orden son distintasm el espacio muestral contiene $V_{N.n} = \binom{N}{n} n!$ muestras.

Un procedimiento aleatorio de muestreo es con reposición cuando las muestras que tienen algún elemento repetido son posibles. Es decir, las unidades seleccionadas se devuelven a la población para seleccionar la siguiente unidad de la muestra. Si se tiene en cuenta el orden de colocación en la muestra, el espacio muestral tendrá $VR_{N,n} = N^n$ muestras posibles. Si no se tiene en cuenta el orden, el espacio muestral tiene $CR_{N,n} = \binom{N-n-1}{n}$ muestras posibles.

Adicionalmente, se pueden clasificar los métodos de selección en otros dos grandes grupos:

Selección con probabilidades iguales: Todos los elementos de la población tienen la misma probabilidad de pertenecer a la muestra.

Selección con probabilidades desiguales: Los elementos de la población no tienen la misma probabilidad de pertenecer a la muestra.

1.6.1. Selección sin reposición.

Consideremos una población de tamaï¿œo N, con unidades $\{u_1, u_2, \ldots, u_N\}$. Seleccionamos sin reposición una muestra $(\tilde{\boldsymbol{x}})$ de tamaï¿œo n. Por tanto, cada unidad puede aparecer como máximo una vez en la muestra. Para cada unidad de la población, u_i, definimos la variable aleatoria e_i como sigue:

$$e_i = \begin{cases} 1 & si\, u_i \in (\tilde{\boldsymbol{x}}) \\ 0 & si\, u_i \notin (\tilde{\boldsymbol{x}}) \end{cases}$$

y con la distribución de probabilidad $P(e_i = 1) = \pi_i$, $P(e_i = 0) = 1 - \pi_i$. Así hemos definido una variable aleatoria en función de la probabilidad de que la unidad i-ésima correspondiente pertenezca a la muestra. Las propiedades de esta variabla aleatoria (que presenta una distribución de Bernouilli) son:

$$E(e_i) = 1\pi_i + 0(1 - \pi_i) = \pi_i \qquad E\left(e_i^2\right) = 1^2\pi_i + 0^2(1 - \pi_i) = \pi_i$$
$$Var(e_i) = E\left(e_i^2\right) - [E(e_i)]^2 = \pi_i - \pi_i^2 = \pi_i(1 - \pi_i)$$

Ahora consideramos la variable aleatoria producto, $e_i e_j$ con $i \neq j$, que evidentemente se define:

$$e_i e_j = \begin{cases} 1 & si\ (u_i, u_j) \in (\tilde{\boldsymbol{x}}) \\ 0 & si\ (u_i, u_j) \notin (\tilde{\boldsymbol{x}}) \end{cases}$$

y con la distribución de probabilidad $P(e_i e_j = 1) = \pi_{ij}$, $P(e_i e_j = 0) = 1 - \pi_{ij}$. Entonces,

$$E(e_i e_j) = 1\pi_{ij} + 0(1 - \pi_{ij}) = \pi_{ij}$$
$$Cov(e_i, e_j) = E(e_i e_j) - E(e_i)E(e_j) = \pi_{ij} - \pi_i\pi_j$$

Veamos algunas propiedades de estas probabilidades:

1. $\sum_{i=1}^N \pi_i = n$ ya que $\sum_{i=1}^N \pi_i = \sum_{i=1}^N E(e_i) = E\left(\sum_{i=1}^N e_i\right) = E(n) = n$.
2. $\sum_{i=1, i\neq j}^N \pi_{ij} = (n-1)\pi_j$, ya que $\sum_{i=1, i\neq j}^N \pi_{ij} = \sum_{i=1, i\neq j}^N E(e_i e_j) = E\left(\sum_{i=1, i\neq j}^N (e_i e_j)\right) = E\left(e_j\left(\sum_{i=1}^N e_i - e_j\right)\right) = nE(e_j) - E\left(e_j^2\right) = (n-1)\pi_j$.
3. $\sum_{i=1, i\neq j}^N (\pi_{ij} - \pi_i\pi_j) = -\pi_j(1 - \pi_j)$, ya que $\sum_{i=1, i\neq j}^N (\pi_{ij} - \pi_i\pi_j) = \sum_{i=1, i\neq j}^N \pi_{ij} - \sum_{i=1, i\neq j}^N \pi_i\pi_j = (n-1)\pi_j - \pi_j\left(\sum_{i=1}^N \pi_i - \pi_j\right) = (n-1)\pi_j - n\pi_j - \pi_j^2 = -\pi_j(1 - \pi_j)$.

1.6.2. Selección con reposición.

Consideremos una población de tamaï¿œo N, con unidades $\{u_1, u_2, \ldots, u_N\}$. Seleccionamos con reposición una muestra $(\tilde{\boldsymbol{x}})$ de tamaï¿œo n. Con este método de selección cada unidad puede aparecer en la muestra de cero a n veces. Definimos para cada unidad u_i la variable aleatoria de apoyo e_i como el número de veces que la unidad i-ésima, u_i pertenece a la muestra de tamaï¿œo n. Estas variables aleatorias se distribuyen según una binomial de tamaï¿œo n, y probabilidad P_i, siendo P_i la probabilidad que tiene la unidad i-ésima de entrar en la muestra en cada extracción. Se tiene que cumpir que $\sum_{i=1}^N P_i = 1$. Por tanto,

$$E(e_i) = nP_i$$
$$Var(e_i) = nP_i(1 - P_i)$$

Sean $t_1, t_2, \ldots, t_N$ el número de veces que aparece cada unidad en la muestra. Por tanto, $\sum_{i=1}^N t_i = n$. La distribución de probabilidad para la muestra será en este caso una multinomial:

$$P(\tilde{\boldsymbol{x}}) = \frac{n!}{t_1! t_2! \cdots t_N!} P_1^{t_1} P_2^{t_2} \cdots P_N^{t_n}$$

Su función generatriz de momentos es:

$$g_{(e_1,\ldots,e_N)}(\theta_1,\ldots,\theta_N) = E\left(e^{\boldsymbol{\theta}'\mathbf{e}}\right) = E\left(e^{\theta_1 e_1+\cdots+\theta_N e_N}\right) = \sum_{t_1+t_2+\cdots+t_N=n} e^{\theta_1 t_1+\cdots+\theta_N t_N}\frac{n!}{t_1!t_2!\cdots t_N!}P_1^{t_1}P_2^{t_2}\cdots P_N^{t_n} =$$
$$= \sum_{t_1+t_2+\cdots+t_N=n}\left(e^{\theta_1}P_1\right)^{t_1}\cdots\left(e^{\theta_N}P_N\right)^{t_N}\frac{n!}{t_1!t_2!\cdots t_N!} = \left(\sum_{i=1}^{N}P_i e^{\theta_i}\right)^n$$

y por tanto: $E(e_i e_j) = \dfrac{\partial^2 g(0,\ldots,0)}{\partial\theta_i\partial\theta_j} = n(n-1)P_iP_j$. $Cov(e_i,e_j) = E(e_ie_j) - E(e_i)E(e_j) = -nP_iP_j$, y así hemos definido el vector esperanza matemática y la matriz de covarianzas para nuestra variable multinomial.

1.7. Comparación con el muestreo no probabilístico: Muestreo por cuotas.

El muestreo no probabilístico es aquel en el que la selección de la muestra no está sometida a criterios probabilísticos, y por tanto no se conoce la probabilidad de cada unidad de estar presente en la muestra. Por tanto, no conoceremos la distribución de probabilidad de los estimadores ni podremos calcular los errores lo que nos impedirá evaluar objetivamente los resultados y su calidad. Algunos tipod e muestreo no probabilístico son:

Muestreo opinático: Se eligen unidades que se creen que son especialmente representativas de la población que se quiere investigar. Con base en la información recabada en esas unidades se hacen estimaciones sobre las carácterísticas de la población.

Muestreo aplicando criterio: La elección de los componentes de la muestra se deja a criterio del entrevistador.

Muestreo por cuotas: Los entrevistadores tienen libertad de elegir la muestra, siempre que esta se omponga de un determinado número de individuos según una o varias características, por ejemplo, determinado número de mujers y de hombres. El diseï¿œo de la encuesta sigue los principios del muestreo probabilístico hasta llegar al momento de selección de la muestra. En esta etapa se da a cada encuestador un número de encuestas a un deternimado grupo de unidades. El margen de maniobra del entrevistador puede introducir sesgos en el proceso de selección, y el desconocimiento de las probabilidades de selección impide evitar errores debido a ponderaciones incorrectas en el proceso de estimación, ni podemos calcular los errores debidos al muestreo.

Chapter 2

Muestreo con probabilidades iguales. Estimadores lineales. Varianzas de los estimadores y sus estimaciones. Comparación entre el muestreo con y sin reposición. Consideraciones sobre el tamaï¿œo de la muestra.

2.1. Muestreo con probabilidades iguales.

Cuando la probabilidad que tiene cualquier unidad de la población de ser elegida para la muestra es la misma para todas la unidades decimos que estamos ante un método de muestreo con probabilidades iguales. Habrá que distinguir entre muestreo sin reposición y con reposición, y entre los casos en que el orden de colocación de los elementos intervenga o no sea así.

2.1.1. Muestreo sin reposición.

2.1.1.1. No interviene el orden.

En este caso cada unidad de muestreo puede aparecer como máximo una vez en la muestra. Por tanto, en la primera extracción cada unidad tendrá una probabilidad $P(u_i) = \frac{1}{N}$, en la segunda la probabilidad de elegir una de las restantes será $P(u_j) = \frac{1}{N-1}$, y así sucesivamente.

En este caso, el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las combinaciones sin repetición de N elementos tomados de n en n, es decir $C_{N,n} = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}$, y la probabilidad de una muestra cualquiera será $P(u_1, u_2, \ldots, u_n) = \frac{1}{\binom{N}{n}}$. Calculando a partir de las probabilidades de elegir una unidad, tenemos:

$$P(u_1, u_2, \ldots, u_n) = n!P(\{u_1, u_2, \ldots, u_n\}) = n!P(u_1)\,P(u_2/u_1)\,P(u_3/u_1u_2)\cdots P(u_n/u_1u_2\ldots u_{n-1}) =$$
$$= n!\frac{1}{N}\frac{1}{N-1}\frac{1}{N-2}\cdots\frac{1}{N-(n-1)} = \frac{1}{\frac{N!}{n!\,(N-n)!}}$$

Calculemos ahora la probabilidad que tiene una unidad u_i de pertenecer a la muestra, ya que es necesario para desarrollar toda la teoría del muestreo. La probabilidad de una unidad de pertenecer a la muestra será la suma de la probabilidad que tiene de salir en la priemer extracción, más la probabilidad de salir en la segunda extracción no habiendo salido en la primera, más la probabilidad de salir en la tercera extracción no habiendo salido en la primera ni en la segunda... y sí hasta la extracción n-ésima. Así podemos escribir:

$$P(u_i = u_1) = \frac{1}{N}$$

$$P(u_i = u_2 \cap u_i \neq u_1) = P(u_i = u_2 / u_i \neq u_1)\, P(u_i \neq u_1) = \frac{1}{N-1}\frac{N-1}{N} = \frac{1}{N}$$

$$P(u_i = u_3 \cap u_i \neq u_2 \cap u_i \neq u_1) = P(u_i = u_3 / u \neq u_2 \cap u_i \neq u_1)\, P(u_i \neq u_2 / u_i \neq u_1)\, P(u_i \neq u_1) = \frac{1}{N-2}\frac{N-2}{N-1}\frac{N-1}{N} = \frac{1}{N}$$

$$\vdots$$

$$P(u_i = u_n \cap u_i \neq u_{n-1} \cap \cdots \cap u_i \neq u_1) = P(u_i = u_n / u_i \neq u_{n-1} \cap \cdots \cap u_i \neq u_1) \cdots P(u_i \neq u_2 / u_i \neq u_1)\, P(u_i \neq u_1) =$$

$$= \frac{1}{N-(n-1)}\frac{N-(n-1)}{N-(n-2)} \cdots \frac{N-2}{N-1}\frac{N-1}{N} = \frac{1}{N}$$

y por tanto, $\pi_i = \frac{n}{N}$.

También será necesario calcular la probabilidad de dos unidades distintas de pertenecer simultáneamente a la muestra. El número de muestras posibles será $\binom{N}{n}$, y el número de muestras en las que estén presentes las dos unidades que nos interesan será $\binom{N-2}{n-2}$. APlicando la visión frecuentista del cálculo de probabilidades, la probabilidad será el número de casos favorables entre el número de casos posibles:

$$\pi_{ij} = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{\frac{(N-2)!}{(n-2)!\,(N-n)!}}{\frac{N!}{n!\,(N-n)!}} = \frac{(N-2)!n!}{N!\,(n-2)!} = \frac{n(n-1)}{N(N-1)}$$

Si definimos la variable aleatoria e_i como el número de veces que la unidad u_i puede aparecer en la muestra, esta variable puede tomar valores entre 0 y 1 para cada unidad, y sigue una distribución de Bernouilli de parámetro $p = \frac{n}{N}$. Por tanto, $E[e_i] = \frac{n}{N}$, $Var(e_i) = \frac{n(N-n)}{N^2}$, $Cov(e_i, e_j) = -\frac{n(N-n)}{N^2(N-1)}$.

2.1.1.2. Interviene el orden.

En este caso, el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las variaciones sin repetición de N elementos tomados de n en n, es decir $V_{N,n} = \binom{N}{n} n! = \frac{N!}{(N-n)!}$, y la probabilidad de una muestra cualquiera será $P(u_1, u_2, \ldots, u_n) = \frac{1}{\binom{N}{n} n!}$. Las probabilidades de una unidad de pertenecer a la muestra son las mismas.

2.1.2. Muestreo con reposición.

2.1.2.1. No interviene el orden.

En este caso, el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las combinaciones con repetición de N elementos tomados de n en n, es decir $CR_{N,n} = \binom{N+n-1}{n} = \frac{(N+n-1)!}{n!\,(N-1)!}$, y la probabilidad de todas las muestras no será la misma, ya que este método de selección no produce muestras equiprobables.

2.1.2.2. Interviene el orden.

En este caso, el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las variaciones con repetición de N elementos tomados de n en n, es decir $VR_{N,n} = N^n$, y la probabilidad de una muestra cualquiera será $P(u_1, u_2, \ldots, u_n) = \frac{1}{N^n}$.

Si definimos la variable aleatoria e_i como el número de veces que la unidad u_i puede aparecer en la muestra, esta variable puede tomar valores entre 0 y n para cada unidad, y sigue una distribución binomial de parámetros n y $p = \frac{1}{N}$. Por tanto, $E[e_i] = \frac{n}{N}$, $Var(e_i) = \frac{n(N-1)}{N^2}$, $Cov(e_i, e_j) = -nP_iP_j = -\frac{n}{N^2}$.

2.2. Estimadores lineales insesgados.

Supongamos que tenemos una población de tamaï¿œo N, para la que hemos definido una característica X_i que toma el valor X_i en cada unidad u_i. Supongamos que tenemos un parámetro poblacional, función de los N valores de las X_i, que es el que queremos estimar. En general este parámetro se puede expresar como una suma de elementos Y_i, que son función de los valores que la característica X_i presenta, $\theta = \sum_{i=1}^{N} Y_i$ donde $Y_i = Y(X_i)$. Por ejemplo:

Total poblacional: $X = \theta(X_1, \ldots, X_N) = \sum_{i=1}^{N} X_i$, donde $Y_i = X_i$.

Media poblacional: $X = \theta(X_1, \ldots, X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i$, donde $Y_i = \frac{X_i}{N}$.

Total de clase: $A = \theta(A_1, \ldots, A_N) = \sum_{i=1}^{N} A_i$, donde $Y_i = A_i$. Donde A_i se usa para características cualitativas dicotómicas: $A_i = 1$ si u_1 presenta la característica, y $A_i = 0$ si u_1 no presenta la característica.

Proporción de clase: $A = \theta(A_1, \ldots, A_N) = \frac{1}{N}\sum_{i=1}^{N} A_i$, donde $Y_i = \frac{A_i}{N}$.

En general las mejores propiedades suelen presentarlas los estimadores lineales, de la forma $\hat{\theta} = \sum_{i=1}^{n} w_i Y_i$.

- Todas las mediciones de la variable de estudio que aparecen en la muestra intervienen en el estimador.
- La importancia de la aportación al estimador de cada unidad muestral puede controlarse mediante su coeficiente.
- Cuando $w_i = 1$ todas las unidades muestrales intervienen con la misma importancia en el estimador.
- Cuando las unidades de la muestra son compuestas, el valor de w_i puede regular la importancia de cada unidad compuesta asociándola con su tamaï¿œo o con el número de unidades elementales que contiene.
- Los coeficientes pueden depender del tamaï¿œo de las unidades muestrales, de su orden en la muestra o de las probabilidades que tienen de pertenecer a la muestra.
- Las funciones lineales son las más sencillas de manejar matemáticamente.

2.2.1. Muestreo sin reposición.

Queremos que el estimador sea insesgado, es decir, que su esperanza sea igual al parámetro que queremos estimar. Veamos cual ha de ser el valor de los coeficientes para que esto ocurra.

$$E\left(\hat{\theta}\right) = E\left(\sum_{i=1}^{n} w_i Y_i\right) = E\left(\sum_{i=1}^{N} w_i Y_i e_i\right) = \sum_{i=1}^{N} w_i Y_i E\left(e_i\right)$$

Ya que $e_i = 1$ si u_i pertenece a la muestra y$e_i = 0$ si u_i no pertenece a la muestra. Como $\theta = \sum_{i=1}^{N} Y_i$, entonces se tiene que cumplir que $w_i E\left(e_i\right) = 1$ y por tanto, $w_i = \dfrac{1}{E\left(e_i\right)}$. Para muestreo sin reposición y probabilidades iguales $E\left(e_i\right) = \dfrac{n}{N}$ y por tanto$w_i = \dfrac{N}{n}$. Con esto podemos construir los estimadores:

Total poblacional: Como $Y_i = X_i$, $\hat{X} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} X_i = \dfrac{N}{n} \sum_{i=1}^{n} X_i = N\bar{x}$.

Media poblacional: Como $Y_i = \dfrac{X_i}{N}$, $\hat{\bar{X}} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} \dfrac{X_i}{N} = \dfrac{1}{n} \sum_{i=1}^{n} X_i = \bar{x}$.

Total de clase: Como $Y_i = A_i$, $\hat{A} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} A_i = \dfrac{N}{n} \sum_{i=1}^{n} A_i$.

Proporción de clase: Como $Y_i = \dfrac{A_i}{N}$, $\hat{P} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} \dfrac{A_i}{N} = \dfrac{1}{n} \sum_{i=1}^{n} A_i$.

2.2.2. Muestreo con reposición.

Queremos que el estimador sea insesgado, es decir, que su esperanza sea igual al parámetro que queremos estimar. Veamos cual ha de ser el valor de los coeficientes para que esto ocurra.

$$E\left(\hat{\theta}\right) = E\left(\sum_{i=1}^{n} w_i Y_i\right) = E\left(\sum_{i=1}^{N} w_i Y_i e_i\right) = \sum_{i=1}^{N} w_i Y_i E\left(e_i\right)$$

Ya que e_i es igual al número de veces que u_i aparece en la muestra, por tanto $Y_i e_i$ es lo mismo que sumar Y_i tantas veces como la unidadu_i aparece en la muestra. Como $\theta = \sum_{i=1}^{N} Y_i$, entonces se tiene que cumplir que $w_i E\left(e_i\right) = 1$ y por tanto, $w_i = \dfrac{1}{E\left(e_i\right)}$. Para muestreo con reposición y probabilidades iguales $E\left(e_i\right) = \dfrac{n}{N}$ y por tanto$w_i = \dfrac{N}{n}$. Con esto podemos construir los estimadores:

Total poblacional: Como $Y_i = X_i$, $\hat{X} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} X_i = \dfrac{N}{n} \sum_{i=1}^{n} X_i = N\bar{x}$.

Media poblacional: Como $Y_i = \dfrac{X_i}{N}$, $\hat{\bar{X}} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} \dfrac{X_i}{N} = \dfrac{1}{n} \sum_{i=1}^{n} X_i = \bar{x}$.

Total de clase: Como $Y_i = A_i$, $\hat{A} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} A_i = \dfrac{N}{n} \sum_{i=1}^{n} A_i$.

Proporción de clase: Como $Y_i = \dfrac{A_i}{N}$, $\hat{P} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{N}{n} \dfrac{A_i}{N} = \dfrac{1}{n} \sum_{i=1}^{n} A_i$.

Podemos observar que coinciden con los estimadores de estoa parámetros si el muestreo se realiza sin reposición.

2.3. Varianzas de los estimadores y sus estimaciones.

Es importante conocer la expresión de la varianza de estos estimadores, así como una forma de estimar la misma, para poder evaluar la calidad de los mismos.

2.3.1. Varianza del estimador en muestreo sin reposición.

$$Var\left(\hat{\theta}\right) = Var\left(\frac{N}{n}\sum_{i=1}^{n} Y_i\right) = Var\left(\frac{N}{n}\sum_{i=1}^{N} Y_i e_i\right) = \frac{N^2}{n^2}\left[\sum_{i=1}^{N} Y_i^2 V\left(e_i\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} Y_i Y_j Cov\left(e_i, e_j\right)\right] =$$
$$= \frac{N^2}{n^2}\left[\frac{n\left(N-n\right)}{N^2}\sum_{i=1}^{N} Y_i^2 - \frac{n\left(N-n\right)}{N^2\left(N-1\right)}\sum_{i=1}^{N}\sum_{j\neq i}^{N} Y_i Y_j\right] = \frac{N-n}{n}\left[\sum_{i=1}^{N} Y_i^2 - \frac{1}{\left(N-1\right)}\sum_{i=1}^{N}\sum_{j\neq i}^{N} Y_i Y_j\right] =$$
$$= \frac{N-n}{n}\left[\sum_{i=1}^{N} Y_i^2 - \frac{1}{\left(N-1\right)}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} Y_i Y_j - \sum_{i=1}^{N} Y_i^2\right)\right] = \frac{N-n}{n\left(N-1\right)}\left[\left(N-1\right)\sum_{i=1}^{N} Y_i^2 + \left(\sum_{i=1}^{N} Y_i^2 - Y^2\right)\right] =$$
$$= \frac{N-n}{n\left(N-1\right)}\left[N\sum_{i=1}^{N} Y_i^2 - Y^2\right] = \frac{N^2\left(N-n\right)}{n\left(N-1\right)}\left[\sum_{i=1}^{N}\frac{Y_i^2}{N} - \bar{Y}^2\right] = \frac{N^2\left(N-n\right)}{n\left(N-1\right)}\sigma_Y^2 = \frac{N^2\left(N-n\right)}{N}\frac{S_Y^2}{n}$$

Y se obtienen las siguientes varianzas:

Total poblacional: Como $Y_i = X_i$, $Var\left(\hat{X}\right) = N^2\left(1-f\right)\frac{S_X^2}{n}$, con $f = \frac{n}{N}$ y $S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$, cuasivarianza poblacional.

Media poblacional: Como $Y_i = \frac{X_i}{N}$, $Var\left(\hat{\bar{X}}\right) = \left(1-f\right)\frac{S_X^2}{n}$, con $f = \frac{n}{N}$ y $S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$, cuasivarianza poblacional.

Total de clase: Como $Y_i = A_i$, $Var\left(\hat{A}\right) = \frac{N^3}{N-1}\left(1-f\right)\frac{P\left(1-P\right)}{n}$.

Proporción de clase: Como $Y_i = \frac{A_i}{N}$, $Var\left(\hat{P}\right) = \frac{N}{N-1}\left(1-f\right)\frac{P\left(1-P\right)}{n}$.

2.3.2. Estimación de las varianzas.

Los estimadores insesgados para estas varainzas son los siguientes:

Total poblacional: Como $Y_i = X_i$, $\hat{Var}\left(\hat{X}\right) = N^2\left(1-f\right)\frac{s_X^2}{n}$, con $f = \frac{n}{N}$ y $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2$, cuasivarianza muestral.

Media poblacional: Como $Y_i = \frac{X_i}{N}$, $\hat{Var}\left(\hat{\bar{X}}\right) = \left(1-f\right)\frac{s_X^2}{n}$, con $f = \frac{n}{N}$ y $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2$, cuasivarianza muestral.

Total de clase: Como $Y_i = A_i$, $\hat{Var}\left(\hat{A}\right) = \frac{N^3}{N-1}\left(1-f\right)\frac{p\left(1-p\right)}{n}$, con $f = \frac{n}{N}$ y $p = \frac{1}{n}\sum_{i=1}^{n} A_i$, proporción muestral.

Proporción de clase: Como $Y_i = \frac{A_i}{N}$, $\hat{Var}\left(\hat{P}\right) = \frac{N}{N-1}\left(1-f\right)\frac{p\left(1-p\right)}{n}$, con $f = \frac{n}{N}$ y $p = \frac{1}{n}\sum_{i=1}^{n} A_i$, proporción muestral.

2.3.3. Varianza del estimador en muestreo con reposición.

$$Var\left(\hat{\theta}\right) = Var\left(\frac{N}{n}\sum_{i=1}^{n} Y_i\right) = Var\left(\frac{N}{n}\sum_{i=1}^{N} Y_i e_i\right) = \frac{N^2}{n^2}\left[\sum_{i=1}^{N} Y_i^2 V\left(e_i\right) + 2\sum_{i=1}^{N}\sum_{j>i}^{N} Y_i Y_j Cov\left(e_i, e_j\right)\right] =$$
$$= \frac{N^2}{n^2}\left[\frac{n\left(N-1\right)}{N^2}\sum_{i=1}^{N} Y_i^2 - 2\frac{n}{N^2}\sum_{i=1}^{N}\sum_{j>i}^{N} Y_i Y_j\right] = \frac{1}{n}\left[\left(N-1\right)\sum_{i=1}^{N} Y_i^2 - 2\sum_{i=1}^{N}\sum_{j>i}^{N} Y_i Y_j\right] \cdots$$
$$\left(\sum_{i=1}^{N} Y_i\right)^2 = \sum_{i=1}^{N} Y_i^2 + 2\sum_{i=1}^{N}\sum_{j>i}^{N} Y_i Y_j$$
$$-2\sum_{i=1}^{N}\sum_{j>i}^{N} Y_i Y_j = \sum_{i=1}^{N} Y_i^2 - \left(\sum_{i=1}^{N} Y_i\right)^2 = \sum_{i=1}^{N} Y_i^2 - \theta^2$$
$$\frac{1}{n}\left[\left(N-1\right)\sum_{i=1}^{N} Y_i^2 - 2\sum_{i=1}^{N}\sum_{j>i}^{N} Y_i Y_j\right] = \frac{1}{n}\left[\left(N-1\right)\sum_{i=1}^{N} Y_i^2 + \sum_{i=1}^{N} Y_i^2 - \theta^2\right] = \frac{1}{n}\left[N\sum_{i=1}^{N} Y_i^2 - \theta^2\right]$$
$$Var\left(\hat{\theta}\right) = \frac{1}{n}\left[N\sum_{i=1}^{N} Y_i^2 - \theta^2\right]$$

Y se obtienen las siguientes varianzas:

Total poblacional: Como $Y_i = X_i$, $Var\left(\hat{X}\right) = \frac{1}{n}\left[N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2\right] = \frac{N^2}{n}\left[\bar{X^2} - \bar{X}^2\right] = \frac{N^2}{n}\sigma_X^2$, con $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$, varianza poblacional.

Media poblacional: Como $Y_i = \frac{X_i}{N}$, $Var\left(\hat{\bar{X}}\right) = \frac{1}{n}\left[N\sum_{i=1}^{N}\frac{X_i^2}{N^2} - \left(\sum_{i=1}^{N}\frac{X_i}{N}\right)^2\right] = \frac{1}{n}\left[\bar{X^2} - \bar{X}^2\right] = \frac{1}{n}\sigma_X^2$, con $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$, varianza poblacional.

Total de clase: Como $Y_i = A_i$, $Var\left(\hat{A}\right) = \frac{1}{n}\left[N\sum_{i=1}^{N} A_i^2 - \left(\sum_{i=1}^{N} A_i\right)^2\right] = \frac{N^2}{n}\left[P - P^2\right] = \frac{N^2}{n}P(1-P)$.

Proporción de clase: Como $Y_i = \frac{A_i}{N}$, $Var\left(\hat{P}\right) = \frac{1}{n}\left[N\sum_{i=1}^{N}\frac{A_i^2}{N^2} - \left(\sum_{i=1}^{N}\frac{A_i}{N}\right)^2\right] = \frac{1}{n}\left[P - P^2\right] = \frac{1}{n}P(1-P)$.

2.3.4. Estimación de las varianzas.

Los estimadores insesgados para estas varainzas son los siguientes:

Total poblacional: Como $Y_i = X_i$, $\hat{Var}\left(\hat{X}\right) = N^2\frac{s_X^2}{n}$, con $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2$, cuasivarianza muestral.

Media poblacional: Como $Y_i = \frac{X_i}{N}$, $\hat{Var}\left(\hat{\bar{X}}\right) = \frac{s_X^2}{n}$, con $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2$, cuasivarianza muestral.

Total de clase: Como $Y_i = A_i$, $\hat{Var}\left(\hat{A}\right) = N^2\frac{p\left(1-p\right)}{n}$, con $p = \frac{1}{n}\sum_{i=1}^{n} A_i$, proporción muestral.

Proporción de clase: Como $Y_i = \frac{A_i}{N}$, $\hat{Var}\left(\hat{P}\right) = \frac{p\left(1-p\right)}{n}$, con $p = \frac{1}{n}\sum_{i=1}^{n} A_i$, proporción muestral.

2.4. Comparación entre el muestreo con y sin reposición.

Es interesante comparar la eficiencia de los estimadores si se realiza el muestreosin reposición respecto a si se realiza con reposición. La forma de comprobar cual es más eficiente es comparar sus varianzas. El estimador que a igual tamaï¿œo de la muestra tenga menor varianza será el más eficiente.

Antes de empezar, es conveniente tener en cuenta que $S_X^2 = \frac{N}{N-1}\sigma_X^2$. Veamos la varianza del estimador del total:

$$\left.\begin{array}{l} Var\left(\hat{X}_{SR}\right) = N^2\left(1-f\right)\frac{S_X^2}{n} = N^2\left(\frac{N-n}{N}\right)\frac{N}{N-1}\frac{\sigma_X^2}{n} = N^2\frac{N-n}{N-1}\frac{\sigma_X^2}{n} \\ Var\left(\hat{X}_{CR}\right) = N^2\frac{\sigma_X^2}{n} \end{array}\right\} \Rightarrow \frac{Var\left(\hat{X}_{SR}\right)}{Var\left(\hat{X}_{CR}\right)} = \frac{N-n}{N-1} < 1 \Rightarrow Var\left(\hat{X}_{SR}\right) < V$$

$$\left.\begin{array}{l} Var\left(\hat{\bar{X}}_{SR}\right) = \left(1-f\right)\frac{S_X^2}{n} = \left(\frac{N-n}{N}\right)\frac{N}{N-1}\frac{\sigma_X^2}{n} = \frac{N-n}{N-1}\frac{\sigma_X^2}{n} \\ Var\left(\hat{\bar{X}}_{CR}\right) = \frac{\sigma_X^2}{n} \end{array}\right\} \Rightarrow \frac{Var\left(\hat{\bar{X}}_{SR}\right)}{Var\left(\hat{\bar{X}}_{CR}\right)} = \frac{N-n}{N-1} < 1 \Rightarrow Var\left(\hat{\bar{X}}_{SR}\right) < Var\left(\hat{\bar{X}}_{CR}\right.$$

$$\left.\begin{array}{l} Var\left(\hat{A}_{SR}\right) = \frac{N^3}{N-1}\left(1-f\right)\frac{P\left(1-P\right)}{n} = \frac{N^2}{N-1}\left(N-n\right)\frac{P\left(1-P\right)}{n} \\ Var\left(\hat{A}_{CR}\right) = \frac{N^2}{n}P(1-P) \end{array}\right\} \Rightarrow \frac{Var\left(\hat{A}_{SR}\right)}{Var\left(\hat{A}_{CR}\right)} = \frac{N-n}{N-1} < 1 \Rightarrow Var\left(\hat{A}_{SR}\right) < Var\left(A\right.$$

$$\left.\begin{array}{l} Var\left(\hat{P}_{SR}\right) = \frac{N}{N-1}\left(1-f\right)\frac{P\left(1-P\right)}{n} = \frac{1}{N-1}\left(N-n\right)\frac{P\left(1-P\right)}{n} \\ Var\left(\hat{P}_{CR}\right) = \frac{1}{n}P(1-P) \end{array}\right\} \Rightarrow \frac{Var\left(\hat{P}_{SR}\right)}{Var\left(\hat{P}_{CR}\right)} = \frac{N-n}{N-1} < 1 \Rightarrow Var\left(\hat{P}_{SR}\right) < Var\left(\hat{P}\right.$$

Por tanto, el muestreo sin reposición es más eficiente que con reposición. El grado de eficiencia que se consigue depende de la relación entre N y n. Cuanto menor sea n respecto al tamaï¿œo de la población menor será la ganancia en precisión del muestreo sin reemplazamiento. Si $N-1 \approx N$, entonces $V_{SR} = \left(1-f\right)V_{CR}$, es decir, cuanto menor sea la fracción de muestreo mayor será la ganancia en precisión. A $\left(1-f\right)$ se le llama factor de corrección para poblaciones finitas.

2.5. Consideraciones sobre el tamaï¿œo de la muestra.

La importancia del tamaï¿œo muestral, radica en la precisión del estimador y la representatividad de muestra. es decir, cuanto mayor sea el tamaï¿œo muestral más preciso será el estimador.

Sin embargo, este razonamiento nos llevaría a proponer muestras de tamaï¿œo N, es decir, que contengan a toda la población. Normalmente esto no es posible por problemas de coste, disponibilidad... Es por esto que es necesario definir un criterio respecto a lo que se espera del estimador, y a partir de ahí tomar la muestra más pequeï¿œa que nos permita cumplir con ese criterio. Ese criterio normalmente se expresa en función del error de muestreo deseado.

Normalmente el tamaï¿œo de la muestra dependerá de parámetros poblacionales desconocidos que habrá que estimar.

2.5.1. Tamaï¿œo de la muestra para un error de muestreo dado.

Sea $e = \sigma_{\hat{\theta}}$ el error de muestreo máximo que estamos dispuestos a admitir. Veamos cual es el tamaï¿œo de la muestra a seleccionar para cometer ese error:

2.5.1.1. Muestreo sin reemplazamiento.

Total poblacional: $e^2 = N^2\left(1-\frac{n}{N}\right)\frac{S_X^2}{n} = \frac{N^2S_X^2}{n} - \frac{N^2S_X^2}{N}$ y por tanto, $n = \frac{N^2S_X^2}{e^2+NS_X^2}$.

Media poblacional: $e^2 = \left(1-\frac{n}{N}\right)\frac{S_X^2}{n} = \frac{S_X^2}{n} - \frac{S_X^2}{N}$ y por tanto, $n = \frac{NS_X^2}{Ne^2+S_X^2}$.

Total de clase: $e^2 = \frac{N^3}{N-1}\left(1-\frac{n}{N}\right)\frac{P\left(1-P\right)}{n} = N^2\left(\frac{N}{N-1}\frac{P\left(1-P\right)}{n} - \frac{N}{N-1}\frac{P\left(1-P\right)}{N}\right)$ y por tanto, $n = \frac{\frac{N}{N-1}P\left(P-1\right)}{\frac{e^2}{N^2}+\frac{N}{N-1}\frac{P\left(1-P\right)}{N}} = \frac{N^3P\left(P-1\right)}{e^2\left(N-1\right)+N^2P\left(P-1\right)}$

Proporción de clase: $e^2 = \frac{N}{N-1}\left(1-\frac{n}{N}\right)\frac{P(1-P)}{n} = \frac{N}{N-1}\frac{P(1-P)}{n} - \frac{N}{N-1}\frac{P(1-P)}{N}$, $n = \frac{NP(1-P)}{e^2(N-1)+P(1-P)}$.

2.5.1.2. Muestreo con reemplazamiento.

Total poblacional: $e^2 = N^2\frac{\sigma_X^2}{n}$ y por tanto, $n = \frac{N^2\sigma_X^2}{e^2}$.

Media poblacional: $e^2 = \frac{\sigma_X^2}{n}$ y por tanto, $n = \frac{\sigma_X^2}{e^2}$.

Total de clase: $e^2 = \frac{N^2}{n}P(1-P)$ y por tanto, $n = \frac{N^2P(P-1)}{e^2}$

Proporción de clase: $e^2 = \frac{1}{n}P(1-P)$, $n = \frac{P(1-P)}{e^2}$.

2.5.2. Tamaï¿œo de la muestra para un error de muestreo y un coeficiente de confianza dados.

Muchas veces no queremos fijar tanto la varianza de nuestro estimador cmo establecer la máxima aplitud del intervalo de confianza que a un nivel de confianza dado que generaremos con el mismo, es decir queremos fijar un e_α y un α tales que:

$$P\left(-e_\alpha \leq \hat{\theta} - \theta \leq e_\alpha\right) = 1-\alpha$$

Por tanto:

$$P\left(\frac{-e_\alpha}{\sigma\left(\hat{\theta}\right)} \leq \frac{\hat{\theta}-\theta}{\sigma\left(\hat{\theta}\right)} \leq \frac{e_\alpha}{\sigma\left(\hat{\theta}\right)}\right) = 1-\alpha$$

Y suponiendo una distribución normal, $\lambda_\alpha = \frac{e_\alpha}{\sigma\left(\hat{\theta}\right)}$, tal que $P(x \leq \lambda_\alpha) = 1-\frac{\alpha}{2}$, de las tablas de la distribución normal estándar. Por tanto, necesitaremos que $\sigma\left(\hat{\theta}\right) = \frac{e_\alpha}{\lambda_\alpha}$, y con este valor podemos aplicar las fórmulas del caso anterior.

Capítulo 3

Muestreo con probabilidades desiguales. Estimadores lineales. Varianza de los estimadores y sus estimaciones. Probabilidades óptimas de selección. Métodos de selección con reposición y sin reposición y probabilidades proporcionales al tamaï¿œo.

3.1. Muestreo con probabilidades desiguales.

Cuando la probabilidad que tiene cualquier unidad de la población de ser elegida para la muestra no es la misma para todas la unidades decimos que estamos ante un método de muestreo con probabilidades desiguales. Habrá que distinguir entre muestreo sin reposición y con reposición, y en los casos en que el orden de colocación de los elementos intervenga o no sea así.

3.1.1. Muestreo sin reposición.

Como norma general, no se tiene en cuenta el orden de colocación de los elementos en la muestra, es decir, muestras con los mismos elementos extraídos en distinto orden son iguales. En este caso, el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las combinaciones sin repetición de N elementos tomados de n en n, es decir $C_{N,n} = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}$. Si consideramos que dos muestras con los mismos elementos ordenados de distinta forma son distintas el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las variaciones sin repetición de N elementos tomados de n en n, es decir $V_{N,n} = \binom{N}{n} n! = \frac{N!}{(N-n)!}$.

Consideremos una población de tamaï¿œo N, de unidades $\{u_1, u_2, \ldots, u_N\}$. Seleccionamos sin reposición una muestra $(\widetilde{\boldsymbol{x}})$ de tamaï¿œo n. Si definimos la variable aleatoria e_i como el número de veces que la unidad u_i aparece en la muestra, esta variable puede tomar valores entre 0 y 1 para cada unidad, y sigue una distribución de Bernouilli

de parámetro $p = \pi_i$, es decir:

$$e_i = \begin{cases} 1 & u_i \in (\widetilde{\boldsymbol{x}}) \\ 0 & u_i \notin (\widetilde{\boldsymbol{x}}) \end{cases} P(e_i = 1) = \pi_i;\ P(e_i = 0) = 1 - \pi_i$$

Por tanto,$E[e_i] = \pi_i$, $E\left[e_i^2\right] = \pi_i$, $Var(e_i) = \pi_i(1-\pi_i)$. Si para cada $i, j = 1, 2, \ldots, N$ con $i \neq j$ consideramos la variable aleatoria producto $e_i e_j$que estará definida:

$$e_i e_j = \begin{cases} 1 & (u_i, u_j) \in (\widetilde{\boldsymbol{x}}) \\ 0 & (u_i, u_j) \notin (\widetilde{\boldsymbol{x}}) \end{cases} P(e_i e_j = 1) = \pi_{ij};\ P(e_i e_j = 0) = 1 - \pi_{ij}$$

Por tanto, $E[e_i e_j] = \pi_{ij}$; $Cov(e_i, e_j) = E(e_i e_j) - E(e_i)E(e_j) = \pi_{ij} - \pi_i \pi_j$.
Propiedades de las probabilidades:

1. $\sum_{i=1}^N \pi_i = n$, ya que $\sum_{i=1}^N \pi_i = \sum_{i=1}^N E(e_i) = E\left(\sum_{i=1}^N e_i\right) = E(n) = n$.
2. $\sum_{i=1, i\neq j}^N \pi_i = n - \pi_j$, ya que $\sum_{i=1}^N \pi_i = \sum_{i=1, i\neq j}^N \pi_i + \pi_j = n$.
3. $\sum_{i=1, i\neq j}^N \pi_{ij} = (n-1)\pi_j$, ya que $\sum_{i=1, i\neq j}^N \pi_{ij} = \sum_{i=1, i\neq j}^N E(e_i e_j) = E\left(\sum_{i=1, i\neq j}^N e_i e_j\right) = E\left(e_j \sum_{i=1, i\neq j}^N e_i\right) = E\left(e_j \sum_{i=1}^N e_i - e_j^2\right) = n\pi_j - \pi_j$.
4. $\sum_{i=1, i\neq j}^N (\pi_{ij} - \pi_i \pi_j) = -\pi_j(1-\pi_j)$, ya que $\sum_{i=1, i\neq j}^N (\pi_{ij} - \pi_i\pi_j) = \sum_{i=1, i\neq j}^N \pi_{ij} - \sum_{i=1, i\neq j}^N \pi_i\pi_j = (n-1)\pi_j - \pi_j \sum_{i=1, i\neq j}^N \pi_i = = (n-1)\pi_j - \pi_j(n - \pi_j) = n\pi_j - \pi_j - n\pi_j + \pi_j^2 = -\pi_j(1-\pi_j)$.

3.1.2. Muestreo con reposición.

Como norma general, no se tiene en cuenta el orden de colocación de los elementos en la muestra, es decir, muestras con los mismos elementos extraídos en distinto orden son iguales. En este caso, el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las combinaciones con repetición de N elementos tomados de n en n, es decir $CR_{N,n} = \binom{N-n+1}{n} = \dfrac{(N+n-1)!}{n!\,(N-1)!}$. Si consideramos que dos muestras con los mismos elementos ordenados de distinta forma son distintas el número de muestras de tamaï¿œo n de un espacio muestral con N unidades será el de las variaciones con repetición de N elementos tomados de n en n, es decir $VR_{N,n} = N^n$.

Consideremos una población de tamaï¿œo N, de unidades $\{u_1, u_2, \ldots, u_N\}$. Seleccionamos con reposición una muestra $(\widetilde{\boldsymbol{x}})$ de tamaï¿œo n. Si definimos la variable aleatoria e_i como el número de veces que la unidad u_i aparece en la muestra, esta variable puede tomar valores entre 0 y n para cada unidad, y sigue una distribución de binomial de parámetros n y P_i, siendo P_i la probabilidad de selección de la unidad i-ésima en cada extracción (probabilidad unitaria de selección). Por tanto, $E[e_i] = nP_i$, $E\left[e_i^2\right] = n^2P_i^2 + nP_i(1-P_i)$, $Var(e_i) = nP_i(1-P_i)$. La probabilidad de una muestra cualquiera seguirá una distribución multinomial, ya que cada unidad u_i puede seleccionarse t_i veces, con $\sum_{i=1}^N t_i = n$, por tanto:

$$P(\widetilde{\boldsymbol{x}}) = P(e_1 = t_1, e_1 = t_1, \ldots, e_N = t_N) = \frac{n!}{t_1!t_1!\cdots t_N!} P_1^{t_1} P_2^{t_2} \cdots P_N^{t_N}$$

Calculamos su función generatriz de momentos:

$$g_{e_1, e_2, \ldots, e_N}(\theta_1, \theta_2, \ldots, \theta_N) = E\left(e^{\theta_1 e_1 + \theta_2 e_2 + \cdots + \theta_N e_N}\right) = \left[P_1 e^{\theta_1} + P_2 e^{\theta_2} + \cdots + P_N e^{\theta_N}\right]^n$$

A partir de esta función calculamos $E(e_i e_j) = \dfrac{\partial^2 g(0, \ldots, 0)}{\partial\theta_i \partial\theta_j} = n(n-1)P_iP_j$. $Cov(e_i, e_j) = E(e_i e_j) - E(e_i)E(e_j) = -nP_iP_j$, y así hemos definido el vector esperanza matemática y la matriz de covarianzas para nuestra variable multinomial.

3.2. Estimadores lineales insesgados.

Supongamos que tenemos una población de tamaï¿œo N, para la que hemos definido una característica X_i que toma el valor X_i en cada unidad u_i. Supongamos que tenemos un parámetro poblacional, función de los N valores de las X_i, que es el que queremos estimar. En general este parámetro se puede expresar como una suma de elementos Y_i, que son función de los valores que la característica X_i presenta, $\theta = \sum_{i=1}^{N} Y_i$ donde $Y_i = Y(X_i)$. Por ejemplo:

Total poblacional: $X = \theta(X_1, \ldots, X_N) = \sum_{i=1}^{N} X_i$, donde $Y_i = X_i$.

Media poblacional: $X = \theta(X_1, \ldots, X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i$, donde $Y_i = \frac{X_i}{N}$.

Total de clase: $A = \theta(A_1, \ldots, A_N) = \sum_{i=1}^{N} A_i$, donde $Y_i = A_i$. Donde A_i se usa para características cualitativas dicotómicas: $A_i = 1$ si u_1 presenta la característica, y $A_i = 0$ si u_1 no presenta la característica.

Proporción de clase: $A = \theta(A_1, \ldots, A_N) = \frac{1}{N}\sum_{i=1}^{N} A_i$, donde $Y_i = \frac{A_i}{N}$.

En general las mejores propiedades suelen presentarlas los estimadores lineales de la forma $\hat{\theta} = \sum_{i=1}^{n} w_i Y_i$.

- Todas las mediciones de la variable de estudio que aparecen en la muestra intervienen en el estimador.
- La importancia de la aportación al estimador de cada unidad muestral puede controlarse mediante su coeficiente.
- Cuando $w_i = 1$ todas las unidades muestrales intervienen con la misma importancia en el estimador.
- Cuando las unidades de la muestra son compuestas, el valor de w_i puede regular la importancia de cada unidad compuesta asociándola con su tamaï¿œo o con el número de unidades elementales que contiene.
- Los coeficientes pueden depender del tamaï¿œo de las unidades muestrales, de su orden en la muestra o de las probabilidades que tienen de pertenecer a la muestra.
- Las funciones lineales son las más sencillas de manejar matemáticamente.

3.2.1. Muestreo sin reposición.

Queremos que el estimador sea insesgado, es decir, que su esperanza sea igual al parámetro que queremos estimar. Veamos cual ha de ser el valor de los coeficientes para que esto ocurra.

$$E\left(\hat{\theta}\right) = E\left(\sum_{i=1}^{n} w_i Y_i\right) = E\left(\sum_{i=1}^{N} w_i Y_i e_i\right) = \sum_{i=1}^{N} w_i Y_i E(e_i)$$

Ya que $e_i = 1$ si u_i pertenece a la muestra y$e_i = 0$ si u_i no pertenece a la muestra. Como $\theta = \sum_{i=1}^{N} Y_i$, entonces se tiene que cumplir que $w_i E(e_i) = 1$ y por tanto, $w_i = \frac{1}{E(e_i)}$. Para muestreo sin reposición $E(e_i) = \pi_i$ y por tanto$w_i = \frac{1}{\pi_i}$. Con esto podemos construir los estimadores, obteniendo el estimador de Horvitz y Thompson para los distintos parámetros poblacionales:

Total poblacional: Como $Y_i = X_i$, $\hat{X} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \frac{X_i}{\pi_i}$, $\hat{X}_{HT} = \sum_{i=1}^{n} \frac{X_i}{\pi_i}$.

Media poblacional: Como $Y_i = \frac{X_i}{N}$, $\hat{\bar{X}} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \frac{1}{\pi_i}\frac{X_i}{N} = \frac{1}{N}\sum_{i=1}^{n} \frac{X_i}{\pi_i}$, $\hat{\bar{X}}_{HT} = \frac{1}{N}\sum_{i=1}^{n} \frac{X_i}{\pi_i}$.

Total de clase: Como $Y_i = A_i$, $\hat{A} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \frac{1}{\pi_i} A_i$, $\hat{A}_{HT} = \sum_{i=1}^{n} \frac{A_i}{\pi_i}$.

Proporción de clase: Como $Y_i = \frac{A_i}{N}$, $\hat{P} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \frac{1}{\pi_i}\frac{A_i}{N}$, $\hat{P}_{HT} = \frac{1}{N}\sum_{i=1}^{n} \frac{A_i}{\pi_i}$.

3.2.2. Muestreo con reposición.

Queremos que el estimador sea insesgado, es decir, que su esperanza sea igual al parámetro que queremos estimar. Veamos cual ha de ser el valor de los coeficientes para que esto ocurra.

$$E\left(\hat{\theta}\right) = E\left(\sum_{i=1}^{n} w_i Y_i\right) = E\left(\sum_{i=1}^{N} w_i Y_i e_i\right) = \sum_{i=1}^{N} w_i Y_i E\left(e_i\right)$$

Ya que e_i es igual al número de veces que u_i aparece en la muestra, por tanto $Y_i e_i$ es lo mismo que sumar Y_i tantas veces como la unidadu_i aparece en la muestra. Como $\theta = \sum_{i=1}^{N} Y_i$, entonces se tiene que cumplir que $w_i E\left(e_i\right) = 1$ y por tanto, $w_i = \dfrac{1}{E\left(e_i\right)}$. Para muestreo con reposición $E\left(e_i\right) = nP_i$ y por tanto$w_i = \dfrac{1}{nP_i}$. Con esto podemos construir los estimadores, obteniendo el estimador de Hansen y Hurwitz para los distintos parámetros poblacionales:

Total poblacional: Como $Y_i = X_i$, $\hat{X} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{X_i}{nP_i}$, $\hat{X}_{HH} = \sum_{i=1}^{n} \dfrac{X_i}{nP_i}$.

Media poblacional: Como $Y_i = \dfrac{X_i}{N}$, $\hat{\bar{X}} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{1}{nP_i}\dfrac{X_i}{N}$, $\hat{\bar{X}}_{HH} = \dfrac{1}{N}\sum_{i=1}^{n} \dfrac{X_i}{nP_i}$.

Total de clase: Como $Y_i = A_i$, $\hat{A} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{A_i}{nP_i}$, $\hat{A}_{HH} = \sum_{i=1}^{n} \dfrac{A_i}{nP_i}$.

Proporción de clase: Como $Y_i = \dfrac{A_i}{N}$, $\hat{P} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} \dfrac{1}{N}\dfrac{A_i}{nP_i}$, $\hat{P}_{HH} = \dfrac{1}{N}\sum_{i=1}^{n} \dfrac{A_i}{nP_i}$.

3.3. Varianzas de los estimadores y sus estimaciones.

Es importante conocer la expresión de la varianza de estos estimadores, así como una forma de estimar la misma, para poder evaluar la calidad de los mismos.

3.3.1. Varianza del estimador en muestreo sin reposición.

$$Var\left(\hat{\theta}_{HT}\right) = Var\left(\sum_{i=1}^{n} \frac{Y_i}{\pi_i}\right) = Var\left(\sum_{i=1}^{N} \frac{Y_i}{\pi_i} e_i\right) = \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2} V\left(e_i\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j} Cov\left(e_i, e_j\right) =$$
$$= \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2}\pi_i\left(1-\pi_i\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\left(\pi_{ij} - \pi_i\pi_j\right) = \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i}\left(1-\pi_i\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\left(\pi_{ij} - \pi_i\pi_j\right)$$

3.3.2. Estimación de las varianzas.

Para poder obtener una estimación de la varianza sólo podemos utilizar los datos muestrales, mientras que el valor de la varianza que tenemos se calcula a partir de la población entera. Necesitamos por tanto un estimador de la varianza. Utilizaremos la raíz cuadrada de este estimador como error de muestreo. Un estimador insesgado de la varianza viene dado por la expresión:

$$\hat{Var}\left(\hat{\theta}_{HT}\right) = \sum_{i=1}^{n} \frac{Y_i^2}{\pi_i^2}\left(1-\pi_i\right) + \sum_{i=1}^{n}\sum_{j\neq i}^{n} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\frac{\left(\pi_{ij} - \pi_i\pi_j\right)}{\pi_{ij}}$$

Veamos si es insesgado:

$$E\left[\hat{Var}\left(\hat{\theta}_{HT}\right)\right] = E\left[\sum_{i=1}^{n} \frac{Y_i^2}{\pi_i^2}\left(1-\pi_i\right)\right] + E\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\frac{\left(\pi_{ij} - \pi_i\pi_j\right)}{\pi_{ij}}\right]$$
$$= E\left[\sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2}\left(1-\pi_i\right)e_i\right] + E\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\frac{\left(\pi_{ij} - \pi_i\pi_j\right)}{\pi_{ij}} e_i e_j\right] = \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2}\left(1-\pi_i\right)E\left(e_i\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\frac{\left(\pi_{ij} - \pi_i\pi_j\right)}{\pi_{ij}}$$
$$= \sum_{i=1}^{N} \frac{Y_i^2}{\pi_i^2}\left(1-\pi_i\right)\pi_i + \sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\frac{\left(\pi_{ij} - \pi_i\pi_j\right)}{\pi_{ij}}\pi_{ij} = Var\left(\hat{\theta}_{HT}\right)$$

3.3.2.1. Estimador de la varianza de Yates y Grundy.

Otro estimador insesgado de la varianza viene dado por la expresión:

$$\hat{Var}\left(\hat{\theta}_{HT}\right)=\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}$$

Para comprobarlo realizamos las siguientes transformaciones en la expresión de la varianza, y teniendo en cuenta que $\sum_{i=1,i\neq j}^{N}(\pi_{ij}-\pi_i\pi_j)=-\pi_j(1-\pi_j)$:

$$Var\left(\hat{\theta}_{HT}\right)=\sum_{i=1}^{N}\frac{Y_i^2}{\pi_i^2}\pi_i(1-\pi_i)+2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}(\pi_{ij}-\pi_i\pi_j)=\sum_{i=1}^{N}\frac{Y_i^2}{\pi_i^2}\sum_{j=1,j\neq i}^{N}(\pi_i\pi_j-\pi_{ij})+2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}(\pi_{ij}-\pi_i\pi_j)=$$

$$=\sum_{i=1}^{N}\sum_{j=1,j>i}^{N}\left(\frac{Y_i^2}{\pi_i^2}+\frac{Y_j^2}{\pi_j^2}\right)(\pi_i\pi_j-\pi_{ij})-2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}(\pi_i\pi_j-\pi_{ij})=\sum_{i=1}^{N}\sum_{j=1,j>i}^{N}\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2(\pi_i\pi_j-\pi_{ij})$$

Y además:

$$E\left[\hat{Var}\left(\hat{\theta}_{HT}\right)\right]=E\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\right]=E\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}e_ie_j\right]=\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2\underline{(\pi}$$

$$=\sum_{i=1}^{N}\sum_{j=1,j>i}^{N}\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2(\pi_i\pi_j-\pi_{ij})$$

3.3.3. Varianza del estimador en muestreo con reposición.

$$Var\left(\hat{\theta}_{HH}\right)=Var\left(\sum_{i=1}^{n}\frac{Y_i}{nP_i}\right)=Var\left(\sum_{i=1}^{N}\frac{Y_i}{nP_i}e_i\right)=\sum_{i=1}^{N}\frac{Y_i^2}{n^2P_i^2}V(e_i)+\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{Y_i}{nP_i}\frac{Y_j}{nP_j}Cov(e_i,e_j)=$$

$$=\sum_{i=1}^{N}\frac{Y_i^2}{n^2P_i^2}nP_i(1-P_i)-\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{Y_i}{nP_i}\frac{Y_j}{nP_j}nP_iP_j=\frac{1}{n}\sum_{i=1}^{N}\frac{Y_i^2}{P_i}-\frac{1}{n}\sum_{i=1}^{N}Y_i^2-\frac{1}{n}\sum_{i=1}^{N}\sum_{j\neq i}^{N}Y_iY_j$$

$$\left(\sum_{i=1}^{N}Y_i\right)^2=\sum_{i=1}^{N}Y_i^2+\sum_{i=1}^{N}\sum_{j\neq i}^{N}Y_iY_j$$

$$-\sum_{i=1}^{N}\sum_{j\neq i}^{N}Y_iY_j=\sum_{i=1}^{N}Y_i^2-\left(\sum_{i=1}^{N}Y_i\right)^2=\sum_{i=1}^{N}Y_i^2-\theta^2$$

$$\frac{1}{n}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i}-\sum_{i=1}^{N}Y_i^2-\sum_{i=1}^{N}\sum_{j\neq i}^{N}Y_iY_j\right]=\frac{1}{n}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i}-\sum_{i=1}^{N}Y_i^2+\sum_{i=1}^{N}Y_i^2-\theta^2\right]=\frac{1}{n}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i}-\theta^2\right]$$

$$Var\left(\hat{\theta}_{HH}\right)=\frac{1}{n}\left[\sum_{i=1}^{N}\left(\frac{Y_i}{P_i}\right)^2P_i-\theta^2\right]$$

Veamos también que:

$$\sum_{i=1}^{N}\left(\frac{Y_i}{P_i}-\theta\right)^2P_i=\sum_{i=1}^{N}\frac{Y_i^2}{P_i^2}P_i-2\theta\underbrace{\sum_{i=1}^{N}Y_i}_{\theta}+\theta^2\underbrace{\sum_{i=1}^{N}P_i}_{1}=\sum_{i=1}^{N}\left(\frac{Y_i}{P_i}\right)^2P_i-\theta^2$$

$$Var\left(\hat{\theta}_{HH}\right)=\frac{1}{n}\sum_{i=1}^{N}\left(\frac{Y_i}{P_i}-\theta\right)^2P_i$$

3.3.4. Estimación de las varianzas.

Para poder obtener una estimación de la varianza sólo podemos utilizar los datos muestrales, mientras que el valor de la varianza que tenemos se calcula a partir de la población entera. Necesitamos por tanto un estimador de la varianza. Utilizaremos la raíz cuadrada de este estimador como error de muestreo. Un estimador insesgado de la varianza viene dado por la expresión:

$$\hat{Var}\left(\hat{\theta}_{HH}\right) = \frac{1}{n\left(n-1\right)}\left[\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}\right)^2 - n\hat{\theta}_{HH}^2\right]$$

Veamos si es insesgado:

$$E\left[\hat{Var}\left(\hat{\theta}_{HH}\right)\right] = \frac{1}{n\left(n-1\right)}E\left[\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}\right)^2 - n\hat{\theta}_{HH}^2\right] = \frac{1}{n\left(n-1\right)}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i^2}E\left(e_i\right) - nE\left(\hat{\theta}_{HH}^2\right)\right] =$$

$$= \frac{1}{n\left(n-1\right)}\left[n\sum_{i=1}^{N}\frac{Y_i^2}{P_i} - n\left(V\left(\hat{\theta}_{HH}\right) + \left[E\left(\hat{\theta}_{HH}\right)\right]^2\right)\right] = \frac{1}{n-1}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i} - V\left(\hat{\theta}_{HH}\right) - \theta^2\right] = \frac{1}{n-1}\left[nV\left(\hat{\theta}_{HH}\right) - V\left(\hat{\theta}_{HH}\right)\right]$$

Veamos también que:

$$\sum_{i=1}^{n}\left(\frac{Y_i}{P_i} - \hat{\theta}_{HH}\right)^2 = \sum_{i=1}^{n}\frac{Y_i^2}{P_i^2} - 2\hat{\theta}_{HH}\underbrace{\sum_{i=1}^{n}\frac{Y_i}{P_i}}_{n\hat{\theta}_{HH}} + n\hat{\theta}_{HH}^2 = \sum_{i=1}^{n}\left(\frac{Y_i}{P_i}\right)^2 - n\hat{\theta}_{HH}^2$$

$$\hat{Var}\left(\hat{\theta}_{HH}\right) = \frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\left(\frac{Y_i}{P_i} - \hat{\theta}_{HH}\right)^2$$

3.4. Probabilidades óptimas de selección.

3.5. Métodos de selección con reposición y sin reposición y probabilidades proporcionales al tamaï¿œo.

En muchas ocasiones, especialmente con unidades de muestreo compuestas, es conveniente asignar a las unidades probabilidades de selección que sean proporcionales al tamaï¿œo de la unidad en cuestión. Vamos a ver distintos esquemas de selección para estos casos.

3.5.1. Estimadores sin reposición.

3.5.1.1. Modelo polinomial o esquema de urna generalizado.

Sea M_i el entero positivo asociado a la unidad compuesta u_i que representa su tamaï¿œo. Sea $M = \sum_{i=1}^{N} M_i$, es decir, el tamaï¿œo total de las unidades de la población. Se selecciona la primera unidad de la muestra con una probabilidad $p_i = \frac{M_i}{M}$. dado que el muestreo es sin reposición, la siguiente unidad, j se selecciona con una probabilidad $p_j = \frac{M_j}{M - M_i}$, y así sucesivamente hasta seleccionar la muestra completa.

Este modelo equivale a tener una urna con M_i bolas por cada unidad, y seleccionar en cada extracción una bola al azar incorporando a la muestra la unidad a la que está asociada, y una vez seleccionada retirar de la urna las M_i bolas asociadas a la unidad. Por eso a este esquema se le llama esquema de urna generalizado. Como $\sum_{i=1}^{N} P_i = \sum_{i=1}^{N} \frac{M_i}{M} = \frac{\sum_{i=1}^{N} M_i}{M} = 1$, el modelo está bien definido.

Veamos las probabilidades de que una unidad pertenezca a la muestra:

$$P\left(u_i = u_1\right) = \frac{M_i}{M}$$

$$P\left(u_i = u_2 \cap u_i \neq u_1\right) = P\left(u_i = u_2 / u_i \neq u_1\right) P\left(u_i \neq u_1\right) = \sum_{j \neq i} \frac{M_i}{M - M_j} \frac{M_j}{M} = \frac{M_i}{M} \sum_{j \neq i} \frac{M_j}{M - M_j} = \frac{M_i}{M} \sum_{j \neq i} \frac{M_j / M}{M/M - M_j/M} = P_i$$

$$P\left(u_i = u_3 \cap u_i \neq u_2 \cap u_i \neq u_1\right) = P\left(u_i = u_3 / u_i \neq u_2 \cap u_i \neq u_1\right) P\left(u_i \neq u_2 / u_i \neq u_1\right) P\left(u_i \neq u_1\right) =$$

$$= \sum_{j \neq i} \sum_{k \neq i, k \neq j} P\left(u_i = u_3 / u_j = u_2 \cap u_k = u_1\right) P\left(u_j = u_2 / u_k = u_1\right) P\left(u_k = u_1\right) = \sum_{j \neq i} \sum_{k \neq i, k \neq j} \frac{M_i}{M - M_j - M_k} \frac{M_j}{M - M_k} \frac{M_k}{M} =$$

$$= \frac{M_i}{M} \sum_{j \neq i} \sum_{k \neq i, k \neq j} \frac{M_i}{M - M_j - M_k} \frac{M_j}{M - M_k} \frac{M_k}{M} =$$

$$\vdots$$

$$P\left(u_i = u_n \cap u_i \neq u_{n-1} \cap \cdots \cap u_i \neq u_1\right) = P\left(u_i = u_n / u_i \neq u_{n-1} \cap \cdots \cap u_i \neq u_1\right) \cdots P\left(u_i \neq u_2 / u_i \neq u_1\right) P\left(u_i \neq u_1\right) =$$

$$= \frac{1}{N - (n-1)} \frac{N - (n-1)}{N - (n-2)} \cdots \frac{N-2}{N-1} \frac{N-1}{N} = \frac{1}{N}$$

$$\pi_i = P\left(u_i \in (\tilde{x})\right) = P\left(u_i = u_1\right) + P\left(u_i = u_2 \cap u_i \neq u_1\right) = P_i \left(1 + \sum_{j \neq i} \frac{P_j}{1 + P_j}\right)$$

$$\pi_{ij} = P\left(u_j = u_2 \cap u_i = u_1\right) + P\left(u_i = u_2 \cap u_j = u_1\right) = \frac{M_i}{M - M_j} \frac{M_j}{M} + \frac{M_j}{M - M_i} \frac{M_i}{M} = P_i P_j \left(\frac{1}{1 - P_i} + \frac{1}{1 - P_j}\right)$$

3.5.1.2. Modelo de Ikeda.

Ikeda propuso el método de selección siguiente: la primera unidad se elige con probabilidades proporcionales al tamaï¿œo, y el resto sin reposición y con probabilidades iguales. Sea $P_i = \frac{M_i}{M}$ la probabilidad asignada a la unidad i, calculemos π_i y π_{ij}. La probabilidad de que la unidad i esté presente es igual a la probabilidad de elegirla la primera más la probabilidad de no elegirla la primera y sí en una de las siguientes elecciones.

$$\pi_i = P_i + (1 - P_i) \frac{n-1}{N-1} = \frac{N-n}{N-1} P_i + \frac{n-1}{N-1}$$

De modo análogo calculamos π_{ij} como la probabilidad de elegir la unidad i la primera y la j en una de las siguientes elecciones, más la probabilidad de elegir la unidad j la primera y la i en una de las siguientes elecciones, más la probabilidad de no elegir ninguna de las dos la primera y elegirlas en una de las siguientes elecciones.

$$\pi_{ij} = P_i \frac{n-1}{N-1} + P_j \frac{n-1}{N-1} + \left(1 - (P_i + P_j) \frac{n-1}{N-1} \frac{n-2}{N-2}\right) = \frac{n-1}{N-1} \left[\frac{N-n}{N-2} (P_i + P_j) + \frac{n-2}{N-2}\right]$$

La probabilidad de una muestra en particular, s, es igual a la probabilidad de obtener la unidad i en la primera selección por la probabilidad de las $n-1$ restantes, sumado para las n unidades que componen la muestras:

$$P(s) = \sum_{i=1}^{n} P_i \frac{1}{\binom{N-1}{n-1}} = \frac{1}{\binom{N-1}{n-1}} \sum_{i=1}^{n} \frac{M_i}{M} = \frac{1}{M} \frac{1}{\binom{N-1}{n-1}} \sum_{i=1}^{n} M_i = k \sum_{i=1}^{n} M_i$$

es decir, la probabilidad de una muestra es proporcional al tamaï¿œo de sus unidades sumado.

3.5.2. Con reposición.

3.5.2.1. Modelo polinomial.

Sea M_i el entero positivo asociado a la unidad compuesta u_i que representa su tamaï¿œo. Sea $M = \sum_{i=1}^{N} M_i$, es decir, el tamaï¿œo total de las unidades de la población. Se seleccionan las unidades de la muestra con una probabilidad $P_i = \dfrac{M_i}{M}$. Este muestreo es con probabilidades proporcionales al tamaï¿œo, $P_i = kM_i$. Un método práctico para seleccionar muestras con esta configuración es definir el intervalo $[1, M]$ y dividirlo en N subintervalos I_i, cada uno de una longitud M_i. Se elige de forma aleatoria un número δ en el intervalo definido, y se incorpora a la muestra la unidad u_i tal que $\delta \in I_i$.

3.5.2.2. Método de Lahiri.

3.5.3. Esquema mixto de selección de Sánchez-Crespo y Gabeiras.

Se considera un esquema de urna en el que cada unidad u_i está representada por M_i bolas. Se selecciona una bola al azar, se incorpora la unidad correspondiente a la muestra y no se reemplaza la bola seleccionada. Así, este modelo tiene probabilidades gradualmente variables, y se podrá extraer la inidad i tantas veces como el mínimo entre su tamaï¿œo y el tamaï¿œo de la muestra.

Es un método mixto de selección: por un lado es sin reposición pues cuando se retira un representante de la unidad i no se repone, pero tiene características del muestreo con reposición, pues una unidad puede estar presente más de una vez en la muestra.

Definimos la variable aleatoria e_i como el número de veces que la unidad i está presente en la muestra. Se distribuye según una Hipergeométrica de parámetros M, n y P_i, siendo P_i la probabilidad de elegir la unidad i-esima para la muestra.

$$E(e_i) = nP_i \;\; Var(e_i) = \frac{M-n}{M-1} nP_i(1-P_i)$$

La probabilidad de una muestra de tamaï¿œo n tendrá una distribución hipergeométrica generalizada. Sea t_i el número de veces que aparece la unidad i-esima en la muestra con $\sum_{i=1}^{N} t_i = n$, y:

$$P(\tilde{x}) = p(e_1 = t_1, e_2 = t_2, \ldots, e_N = t_N) = \frac{\binom{M_1}{t_1}\binom{M_2}{t_2}\cdots\binom{M_N}{t_N}}{\binom{\sum_{i=1}^{N} M_i}{\sum_{i=1}^{N} t_i}} = \frac{\binom{MP_1}{t_1}\binom{MP_2}{t_2}\cdots\binom{MP_N}{t_N}}{\binom{M}{n}}$$

$$E(e_i e_j) = MP_iP_j \frac{n(n-1)}{(M-1)} \;\; Cov(e_i, e_j) = \frac{M-n}{M-1} nP_iP_j$$

3.5.3.1. Estimador lineal insesgado de Sánchez-Crespo y Gabeiras.

Estimamos la característica pobacional $\theta = \sum_{i=1}^{N} Y_i$. El estimador lineal será $\hat{\theta} = \sum_{i=1}^{n} \omega_i Y_i$, y para que sea insesgado debe cumplir que $E\left(\hat{\theta}\right) = \theta$. Por tanto,

$$E\left(\hat{\theta}\right) = E\left(\sum_{i=1}^{n} \omega_i Y_i\right) = E\left(\sum_{i=1}^{N} \omega_i Y_i e_i\right) = \sum_{i=1}^{N} \omega_i Y_i E(e_i) = \sum_{i=1}^{N} \omega_i Y_i nP_i = \sum_{i=1}^{N} Y_i$$

Y por tanto, para ue sea insesgado, $\omega_i = \dfrac{1}{nP_i}$ y el estimador insesgado de Sánchez-Crespo y Gabeiras será $\hat{\theta}_{SCG} = \sum_{i=1}^{n} \dfrac{Y_i}{nP_i}$ que coincide con la expresión del estimador de Hansen y Hurwitz.

Su varianza será