Unidad 3

• Muestreo en Poblaciones Finitas

3.1 Tipos de muestreo aleatorio simple.
3.2 Estimación total, la media, la proporción poblacionales y el total de clase.
3.3 Varianza teórica de los estimadores de la media, el total y la proporción poblacionales.
3.4 Propiedades de la cuasivarianza en poblaciones finitas.
3.5 Estimación de la varianza de los estimadores de la media, el total y la proporción poblacionales.
3.6 Aplicaciones del muestreo de poblaciones finitas.
3.7 Intervalos de confianza para los estadísticos poblacionales: media, total, proporción y total de clase.
3.8 Determinación del tamaño óptimo muestral para la estimación de parámetros poblacionales.

11.1. TIPOS DE MUESTREO ALEATORIO SIMPLE

Este capítulo trata de la extracción de muestras en poblaciones finitas, lo cual representa una novedad con respecto a los análisis previos. Al igual que en dichos capítulos, consideramos únicamente *muestreo aleatorio simple,* en el que cada elemento individual de la población tiene la misma probabilidad de ser escogido en cada extracción muestral[1]. Cuando hemos tratado los problemas de estimación y contrastación de hipótesis, lo hemos hecho como si la población de la que hemos extraído la muestra de trabajo era de tamaño infinito. Este supuesto ignora consideraciones muy importantes cuando la población no es infinita y, especialmente, si es de tamaño relativamente reducido. En tal caso, el investigador debe preocuparse acerca de que su muestra sea *suficientemente representativa* de la población, del modo que vamos a aprender en este capítulo.

Además, las propiedades de los estadísticos y, concretamente, su esperanza y varianza, dependen de que el tamaño de la población de la que se extraen sea o no infinita. En este sentido, el lector debe ser consciente de que todo el tratamiento que hemos efectuado en los capítulos previos es válido solamente bajo el supuesto de poblaciones infinitas. *Cuando la población es finita, hay que utilizar las expresiones de la esperanza matemática y varianza que para los principales estadísticos muestrales vamos a obtener en este capítulo.*

Dicho esto, adelantemos que puede decirse que, si la población es finita, pero su tamaño es suficientemente grande en relación con los tamaños muestrales que se están utilizando, entonces la teoría de poblaciones infinitas es válida prácticamente a todos los efectos. De igual modo, vamos a concluir de nuestro análisis que el tipo de muestreo desempeña un papel fundamental en estas consideraciones, pues mientras que un muestreo sin reemplazamiento preserva el carácter finito de una población, el muestreo con reemplazamiento hace que una población finita adopte las características de una población infinita. Es fácil intuir la razón: si los elementos no se reemplazan en una población finita, la probabilidad que cada elemento tiene de ser extraído *depende* de los resultados de las extracciones previas, lo que no ocurre en poblaciones infinitas. Por el contrario, con reemplazamiento, las probabilidades de extracción *no dependen* de los resultados de las extracciones previas, al igual que ocurre en las poblaciones infinitas.

Volvemos a tomar conceptos ya introducidos en las Secciones 4.5 y 7.1, por lo que es conveniente su lectura antes de comenzar este capítulo.

Utilizamos en este capítulo un parámetro nuevo, N, el tamaño de la población. El cociente n/N es clave. Cuando es inferior a un 5 %, estamos en la situación mencionada en que, aun siendo finita, la población es, a todos los efectos, infinita, al menos en relación con la muestra. Por otra parte, recordemos que en el Capítulo 4 ya obtuvimos el número de muestras distintas de tamaño n que pueden extraerse de una población finita de tamaño N:

[1] Hay que entender esta condición con mucho cuidado: si un elemento está repetido, es la probabilidad de cada unidad la que es igual a la de las demás. En una urna con 2 bolas blancas y 3 rojas, la probabilidad de extraer una bola blanca no es igual a la de extraer una bola roja, pero no es de eso de lo que estamos hablando aquí. Estamos diciendo que, en un muestreo aleatorio simple, cada bola blanca tiene la misma probabilidad de ser extraída que la otra, e igual a la probabilidad de cada bola roja, cada una igual a 1/5 en este caso.

Cuadro 11.1. NÚMERO DE MUESTRAS POSIBLES

	Las muestras varían con el orden de sus elementos	*Las muestras no varían con el orden de sus elementos*
Sin reemplazamiento	Variaciones sin repetición de N elementos, tomados n a n: $\binom{N}{n} n!$	Combinaciones sin repetición de N elementos, tomados n a n: $\binom{N}{n}$
Con reemplazamiento	Variaciones con repetición de N elementos, tomados n a n: N^n	Combinaciones con repetición de N elementos, tomados n a n: $\binom{N-1+n}{n} = \binom{N-1+n}{N-1} = \frac{(N-1+n)!}{(N-1)!n!}$

11.1.1. Muestreo sin reemplazamiento

Definamos la variable *indicador* del elemento i-ésimo de la población, $1 \leqslant i \leqslant N$, denotada por e_i, a aquella variable que toma el valor 1 si dicho elemento pasa a formar parte de la muestra, y toma el valor 0 en caso contrario. Denotemos por π_i la probabilidad de que e_i sea igual a 1, siendo $1-\pi_i$ la del suceso contrario, y por π_{ij} la probabilidad de que los elementos i y j formen, simultáneamente, parte de la muestra. Por ser e_i una variable binomial, tenemos:

$$E(e_i) = 1\, P(e_i = 1) + 0\, P(e_i = 0) = \pi_i$$

$$Var(e_i) = E(e_i^2) - [E(e_i)]^2 = \pi_i - \pi_i^2 = \pi_i(1 - \pi_i)$$

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

donde hemos utilizado el hecho de que el producto $e_i e_j$ es igual a 1 si los elementos poblacionales i-ésimo y j-ésimo están en la muestra, y es igual a cero si alguno de ellos no forma parte de la misma.

La probabilidad π_i de que el elemento i-ésimo de la población pertenezca a una muestra de tamaño n, es decir, que $e_i = 1$, cuando las muestras no difieren con el orden de sus elementos es:

$$\pi_i^{SR} = \frac{n.^o\, muestras\, con\, dicho\, elemento}{n.^o\, total\, de\, muestras} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

puesto que el número de muestras de tamaño n que pueden extraerse de una población de tamaño N, que contengan al elemento i-ésimo, es igual al número de muestras sin reemplazamiento que pueden formarse de tamaño $n - 1$, de una población de tamaño $N - 1$.

Por similar razón, la probabilidad de que un par de elementos poblacionales distintos esté simultáneamente incluidos en la muestra, es:

$$\pi_{ij}^{SR} = \frac{n.^o \, muestras \, con \, ambos \, elementos}{n.^o \, total \, de \, muestras} = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n}{N}\frac{n-1}{N-1}$$

por lo que, en el muestreo sin reemplazamiento, tenemos:

$$E(e_i) = \frac{n}{N} \quad ; \quad E(e_i^2) = \frac{n}{N};$$

$$Var(e_i) = \frac{n}{N}\left(1 - \frac{n}{N}\right)$$

$$Cov(e_i, e_j) = -\frac{(N-n)n}{(N-1)N^2}$$

11.1.2. Muestreo con reemplazamiento

Supongamos que hay k tipos diferentes de elementos en la población, y sea P_i la probabilidad de elegir el elemento i-ésimo de la población, cada vez que tomamos un elemento de la misma para formar parte de la muestra, cumpliéndose, por tanto: $P_1 + P_2 + \ldots + P_k = 1$. Tal elemento poblacional puede aparecer 0, 1, 2, .., n veces en una muestra de tamaño n. El número de veces que aparece el elemento i-ésimo en la muestra es, por tanto, una variable aleatoria e_i con distribución binomial, $B(n, P_i)$, de modo que:

$$E(e_i) = n\,P_i$$
$$Var(e_i) = n\,P_i\,(1 - P_i)$$

Si todos los elementos son distintos, entonces $k = N$ y $P_i = 1/N$.

Si existe un número diferente M_i de elementos de cada tipo, $\sum_{i=1}^{k} M_i = N$, entonces la probabilidad de extraer un elemento de tipo i-ésimo ya no sería $1/N$, sino M_i/N. En cuanto a la muestra global, el vector aleatorio $(e_1, e_2, \ldots, e_k)$ sigue una *distribución multinomial*, de acuerdo con la cual, la probabilidad de que al seleccionar una muestra de tamaño n, haya r_1 elementos de tipo 1, r_2 elementos de tipo 2, etc, es:

$$P(e_1 = r_1, e_2 = r_2, \ldots, e_k = r_k) = \frac{n!}{r_1! r_2! \ldots r_k!} P_1^{r_1} P_2^{r_2} \ldots P_k^{r_k}$$

$$\text{con } \sum_{1}^{k} P_i = 1 \text{ y } \sum_{1}^{k} r_i = n$$

donde cada probabilidad será: $P_i = M_i/N$, siendo M_i el número de elementos *del tipo i-ésimo* que existen en la población. Reconocemos esta distribución como la *distribución multinomial* [Sección 8.6], en la que sabemos que la covarianza entre cada dos de sus variables componentes es:

$$Cov(e_i, e_j) = -nP_iP_j \quad \text{para } i \neq j$$

También sabemos que la distribución de probabilidad de cada componentes es $B(n, P_i)$, lo que explica la esperanza y varianza de e_i que antes obtuvimos.

Así, considerando el caso particular en que *las probabilidades de selección de cada elemento poblacional sean iguales*: $P_i = 1/N$, $1 \leqslant i \leqslant N$, se tiene:

$$E(e_i) = \frac{n}{N}\ ;\ Var(e_i) = \frac{n}{N}\left(1 - \frac{1}{N}\right);$$

$$E(e_i^2) = \frac{n(n-1)}{N^2} + \frac{n}{N} = \frac{Nn - n + n^2}{N^2}$$

y, por tanto:

$$Cov(e_i, e_j) = -\frac{n}{N^2};$$

$$E(e_i e_j) = \frac{n^2 - n}{N^2}$$

11.2. ESTIMACIÓN DEL TOTAL, LA MEDIA, LA PROPORCIÓN POBLACIONALES Y EL TOTAL DE CLASE

Estamos ahora en condiciones de considerar la estimación de algunos estadísticos poblacionales de interés, como son: el *total* poblacional de una variable, es decir, la suma de los valores de la variable a lo largo de todos los elementos que componen la población, la *media* poblacional de la variable, es decir, el cociente entre el total y el tamaño poblacional y, por último, para una determinada característica de interés, la *proporción* poblacional, es decir, el porcentaje de elementos de la población que satisfacen dicha característica. Por último, el *total de clase* es el número de elementos que en la población satisfacen la característica en estudio. Si analizamos la renta en una pequeña población formada por 500 hogares censados, el *total* es la renta agregada de todos los hogares, mientras que la *media* es la renta promedio en todos los hogares. Nótese que ésta es la *media poblacional.* Un ejemplo de proporción poblacional sería el porcentaje, de entre las 500 familias, que ingresan más de 200.000 ptas./mes, mientras que el *total de clase* sería el número de familias, entre las 500, que ingresan más de dicha cantidad. Nótese que ésta es la única ocasión en este texto en que hablamos de la *media poblacional,* y no de su esperanza matemática, precisamente por el carácter finito de la población.

Las variables mayúsculas denotarán estadísticos poblacionales, y con minúsculas denotaremos sus análogos muestrales. Si X_i denota la variable en estudio, denotamos el total y la media por:

$$\textit{Total poblacional:}\quad X = \sum_{i=1}^{N} X_i$$

$$\textit{Media poblacional:}\quad \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{X}{N}$$

donde debe notarse que existe siempre una relación exacta entre *media* y *total* poblacionales: La media es igual al total dividido por el tamaño de la población. La media poblacional es denotada por $\overline{X}$, distinguiéndola de la media muestral, que seguirá siendo denotada por $\bar{x}$. Otro ejemplo del total poblacional es la masa salarial percibida por el conjunto de los trabajadores de un país (la población) mientras que la media poblacional sería el salario medio de cada uno de ellos.

Dada una determinada característica, definimos la variable A_i, que toma el valor 1 si el elemento i-ésimo de la población satisface dicha característica, tomando el valor numérico 0 en caso contrario. Tenemos así los estadísticos:

$$\textit{Proporción poblacional:}\quad P = \frac{1}{N}\sum_{i=1}^{N} A_i$$

$$\textit{Total de clase:}\quad A = \sum_{i=1}^{N} A_i = NP$$

volviéndonos a encontrar con una relación exacta entre ambos. Denotamos por p la proporción muestral. Otro ejemplo de proporción es la tasa de paro, que es el porcentaje de la población activa (la población estadística, en este caso), que satisface la condición de estar empleado. El total de clase es entonces, el número de ocupados.

Consideremos genéricamente un estadístico poblacional θ, *que pueda representarse como función de* Y_i, una variable que toma un determinado valor numérico para cada elemento poblacional:

$$\theta = \sum_{1}^{N} Y_i$$

Los estadísticos poblacionales mencionados, total, media, proporción y total de clase, pueden representarse de este modo para distintas elecciones de la variable Y_i. En el caso del *total*, $Y_i = X_i$, para la *media poblacional:* $Y_i = X_i/N$, para la *proporción:* $Y_i = A_i/N$, y para el *total de clase:* $Y_i = A_i$.

11.2.1. Muestreo sin reemplazamiento

Consideremos primero el caso del muestreo sin reemplazamiento, es decir, aquél en que las unidades extraídas para formar parte de la muestra no son devueltas a la población [Sección 4.5.2].

Supongamos que estimamos θ mediante una combinación lineal de las observaciones de Y_i obtenidas en la muestra:

$$\dot{\theta} = \sum_{1}^{n} a_i Y_i = \sum_{1}^{N} (a_i e_i) Y_i$$

Nótese el cambio de límites del sumatorio en la igualdad de la derecha: podemos convertir una suma desde 1 hasta n en una suma hasta N sin más que multiplicar por la variable e_i, puesto que ésta es igual a 0 en los elementos poblacionales que no forman parte de la muestra.

Este *estimador lineal ponderado* tiene esperanza matemática:

$$E(\hat{\theta}) = \sum_1^N a_i E(e_i) Y_i = \sum_1^n a_i Y_i \frac{n}{N}$$

(ver Sección 11.1.1.) por lo que será insesgado si tomamos: $a_i = N/n$, en cuyo caso, el estimador queda definido por:

$$\hat{\theta} = \frac{N}{n} \sum_1^n Y_i$$

expresión en la que, según quién sea Y_i, estaremos estimando el total, la media o la proporción poblacionales. Substituyendo estos a_i, tenemos las expresiones para los estimadores insesgados:

$$TOTAL: \; Y_i = X_i \;\Rightarrow\; \hat{X} = \frac{N}{n} \sum_1^n X_i = N\bar{x}$$

$$MEDIA: \; Y_i = \frac{X_i}{N} \;\Rightarrow\; \hat{\bar{X}} = \frac{N}{n} \sum_1^n \frac{X_i}{N} = \sum_1^n \frac{X_i}{n} = \bar{x}$$

$$PROPORCION: \; Y_i = \frac{A_i}{N} \;\Rightarrow\; \hat{P} = \frac{N}{n} \sum_1^n \frac{A_i}{N} = \frac{1}{n} \sum_1^n A_i = p$$

$$TOTAL\ DE\ CLASE: \; Y_i = A_i \;\Rightarrow\; \hat{A} = \frac{N}{n} \sum_1^n A_i = Np = N\hat{P}$$

En un caso más general, en que hubiera elementos repetidos en la población, la probabilidad de extraer cada uno de ellos no sería igual a $1/N$, y las expresiones similares serían más complicadas.

11.2.2. Muestreo con reemplazamiento

Consideremos ahora el caso en que las unidades seleccionadas para formar parte de la muestra son reintegradas a la población. Si definimos el estimador $\hat{\theta}$ de un parámetro θ como hicimos en el caso del muestreo sin reemplazamiento, tendremos:

$$\hat{\theta} = \sum_1^n a_i Y_i = \sum_1^N a_i Y_i e_i$$

$$E(\hat{\theta}) = \sum_1^N a_i Y_i E(e_i) = \sum_1^N a_i Y_i n P_i$$

donde hemos utilizado resultados de la Sección 11.1.2, y, *en el caso de selección con igual probabilidad* para todos los elementos en la población, en que se tiene: $P_i = 1/N$, tenemos:

$$E(e_i) = n\,P_i = n\frac{1}{N} = \frac{n}{N}$$

por lo que escogiendo: $a_i = \dfrac{N}{n}$ obtenemos nuevamente estimadores integrados.

Por tanto, las expresiones analíticas de los *estimadores insesgados del total, media y proporción poblacionales son las mismas que las que obtuvimos para el muestreo sin reemplazamiento,* lo que no ocurriría bajo supuestos diferentes del que hemos hecho acerca de la igual probabilidad de extracción de cada elemento poblacional.

11.3. VARIANZAS TEÓRICAS DE LOS ESTIMADORES DE LA MEDIA, EL TOTAL Y LA PROPORCIÓN POBLACIONALES

Una vez que sabemos cómo estimar de manera insesgada el total, la media, la proporción y el total de clase, obtenemos en esta sección las expresiones analíticas para sus varianzas. Como en todo problema de estimación, la varianza del estimador es la medida de precisión en la estimación, por lo que indica el grado de confianza que asociamos al valor numérico del estimador, como aproximación al verdadero valor, desconocido, del parámetro poblacional que se estima. Conocer la varianza nos permitirá, además, construir intervalos de confianza, al igual que con cualquier parámetro poblacional desconocido, y contrastar hipótesis sencillas, si conocemos la distribución de probabilidad poblacional. En realidad, en este momento vamos a deducir las expresiones analíticas de la varianza de cada estimador, que dependen, a su vez, de parámetros desconocidos. Por tanto, para poder utilizarlas en la construcción de intervalos de confianza, habremos de estimar, a su vez, los valores numéricos que toman las expresiones que a continuación deducimos:

$$Var(\dot{X}) = E[\dot{X} - E(\dot{X})]^2 = E[N\bar{x} - X]^2 = E[N\bar{x} - N\bar{X}]^2 = N^2\,E[\bar{x} - \bar{X}]^2 =$$

$$= N^2\,E\left[\frac{\sum_1^n X_i}{n} - \frac{n\bar{X}}{n}\right]^2 = \frac{N^2}{n^2}\,E\left[\sum_1^n (X_i - \bar{X})\right]^2 = \frac{N^2}{n^2}\,E\left[\sum_1^N (X_i - \bar{X})\,e_i\right]^2 =$$

$$= \frac{N^2}{n^2}\left[\sum_1^N (X_i - \bar{X})^2\,E(e_i^2) + \sum_{i \neq j} (X_i - \bar{X})(X_j - \bar{X})\,E(e_i e_j)\right] =$$

$$= \frac{N^2}{n^2}\,[N\sigma^2\,E(e_i^2) - N\sigma^2\,E(e_i e_j)] = \frac{N^3}{n^2}\,\sigma^2\,[E(e_i^2) - E(e_i e_j)]$$

donde σ^2 denota la varianza poblacional, es decir $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$, y hemos utilizado que:

$$\sum_{1}^{N}(X_i - \overline{X}) = 0 \Rightarrow \left[\sum_{1}^{N}(X_i - \overline{X})\right]^2 = 0 = \sum_{1}^{N}(X_i - \overline{X})^2 + \sum_{i \neq j}(X_i - \overline{X})(X_j - \overline{X})$$

por lo que:

$$\sum_{1}^{N}(X_i - \overline{X})^2 = -\sum_{i \neq j}(X_i - \overline{X})(X_j - \overline{X})$$

pero el miembro izquierdo es igual a: $N\sigma^2$, de modo que el miembro de la derecha es igual a: $-N\sigma^2$, lo que hemos incorporado en la penúltima igualdad.

Aunque esta expresión está directamente relacionada con la varianza del Total, será utilizada para calcular las varianzas de todos los demás estimadores, como veremos pronto.

11.3.1. Muestreo sin reemplazamiento

Utilizando las expresiones apropiadas para $E(e_i^2)$ y $E(e_i e_j)$ en este tipo de muestreo, se tiene la *varianza del estimador del total poblacional*:

$$Var_{SR}(X) = \frac{N^3}{n^2}\sigma^2\left[\frac{n}{N} - \frac{n}{N}\frac{n-1}{N-1}\right] = N^2\frac{N-n}{N-1}\frac{\sigma^2}{n}$$

donde hemos utilizado los resultados de la Sección 11.1.1. Como *la varianza del estimador de la media poblacional*, que es $\overline{x}$, es:

$$Var(\overline{X}) = Var_{SR}(\overline{x}) = Var_{SR}\left(\frac{1}{N}X\right) = \frac{1}{N^2}Var_{SR}(X) = \frac{N-n}{N-1}\frac{\sigma^2}{n}$$

Ya vimos que el estimador de la media poblacional es la media muestral. Ahora tenemos también la varianza de dicho estimador, que no es sino una corrección, debido a que la población es finita, sobre la expresión habitual en poblaciones infinitas para la varianza de la media muestral: $Var(\overline{x}) = \sigma^2/n$. Cuando el tamaño de la población, en relación con el de la muestra, N/n, aumenta, entonces el factor de corrección que hemos introducido tiende a uno, por lo que la expresión de la varianza de la media muestral coincide con la varianza de dicho estadístico en el caso de poblaciones infinitas.

Recordemos, por otra parte, que el estimador de la proporción poblacional es la media aritmética de la variable dicotómica $A_i = \{0,1\}$, según que el elemento i-ésimo $1 \leqslant i \leqslant n$ de la muestra satisfaga o no la característica en estudio. Por ser una media aritmética, podemos usar la expresión anterior para obtener *la varianza del estimador de la proporción*:

$$Var_{SR}(\hat{P}) = \frac{N-n}{N-1}\frac{\sigma_A^2}{n}$$

donde σ_A^2 es la varianza de la variable dicotómica, es decir:

$$\sigma_A^2 = \frac{1}{N}\sum_1^N (A_i - P)^2 = \frac{1}{N}\sum_1^N A_i^2 - P^2 = \frac{1}{N}\sum_1^N A_i - P^2 = P - P^2 = P(1-P)$$

ya que, puesto que A_i sólo toma valores 0 y 1, entonces: $A_i^2 = A_i$.

Finalmente:

$$Var_{SR}(\hat{P}) = \frac{N-n}{N-1}\,\frac{P(1-P)}{n}$$

Ya habíamos visto que la proporción muestral es un estimador insesgado de la proporción poblacional. Ahora tenemos la varianza de dicho estimador como una corrección, debido al tamaño finito de la población, sobre la expresión habitual para la varianza de la proporción. Al igual que ocurría con la media muestral, cuando el tamaño poblacional tiende a infinito, en relación con el de la muestra, la expresión de la varianza de la proporción muestral en poblaciones finitas coincide con la varianza de dicho estimador en poblaciones infinitas.

El estimador del *total de clase*, es decir, del número de elementos que en la población satisfacen la característica en estudio, se obtiene sin más que multiplicar el estimador de la proporción de dichos elementos, por el tamaño poblacional N. En consecuencia, *la varianza del estimador del total de clase* es:

$$Var_{SR}(\hat{A}) = N^2\, Var_{SR}(\hat{P})$$

11.3.2. Muestreo con reemplazamiento

Utilizando las expresiones para $E(e_i^2)$ y $E(e_ie_j)$ en este tipo de muestreo (Sección 11.1.2), se tiene la varianza del estimador del total poblacional:

$$Var_{CR}(\hat{X}) = \frac{N^3}{n^2}\sigma^2\left[\frac{Nn - n + n^2}{N^2} - \frac{n^2 - n}{N^2}\right] = \frac{N^2}{n}\sigma^2$$

Asimismo, recordando que el estimador de la media poblacional es la media muestral, tenemos:

$$Var_{CR}(\hat{\bar{X}}) = Var_{CR}(\bar{x}) = Var_{CR}\left(\frac{1}{N}\hat{X}\right) = \frac{1}{N^2}\, Var_{CR}(\hat{X}) = \frac{\sigma^2}{n}$$

Por otra parte, la proporción muestral es un estimador insesgado de la proporción poblacional, y es una media muestral, la de la variable A, por lo que podemos utilizar el resultado que acabamos de obtener:

$$Var_{CR}(\hat{P}) = \frac{\sigma_A^2}{n} = \frac{P(1-P)}{n}$$

mientras que estimamos la varianza del total de clase mediante:

$$Var_{CR}(\hat{A}) = N^2 \frac{P(1-P)}{n}$$

Como puede apreciarse, a diferencia del muestreo sin reemplazamiento, las expresiones de las varianzas para el muestreo con reemplazamiento son idénticas a las que se obtienen en el caso de poblaciones infinitas. Ello se debe a que, al reemplazar los elementos muestrales que se van extrayendo se tiene, de hecho, una población infinita.

11.4. PROPIEDADES DE LA CUASIVARIANZA EN POBLACIONES FINITAS

Comencemos obteniendo estimadores insesgados de la varianza y cuasivarianza poblacionales. En este caso, las expresiones varían, si bien ligeramente, dependiendo de que el muestreo sea con o sin reemplazamiento. La cuasivarianza poblacional es:

$$S^2 = \frac{1}{N-1} \sum_1^N (X_i - \bar{X})^2 = \frac{N}{N-1} \sigma^2$$

y si consideramos la cuasivarianza muestral:

$$s^2 = \frac{1}{n-1} \sum_1^n (X_i - \bar{x})^2$$

como estimador, podemos calcular su esperanza matemática, que es:

$$E(s^2) = \frac{n}{n-1} E\left[\frac{1}{n} \sum_1^N (X_i - \bar{x})^2 e_i\right] = \frac{n}{n-1} E\left[\frac{1}{n} \sum_1^N X_i^2 e_i + \frac{1}{n} \sum_1^N \bar{x}^2 e_i - 2\frac{1}{n} \bar{x} \sum_1^N X_i e_i\right] =$$

$$= \frac{n}{n-1} E\left[\frac{1}{n} \sum_1^N X_i^2 e_i - \bar{x}^2\right] = \frac{n}{n-1} \left[\frac{1}{n} \sum_1^N X_i^2 E(e_i) - E(\bar{x}^2)\right] =$$

$$= \frac{n}{n-1} \left[\frac{1}{n} \sum_1^N X_i^2 \frac{n}{N} - E(\bar{x}^2)\right] = \frac{n}{n-1} \left[\frac{1}{N} \sum_1^N (X_i^2 - \bar{X}^2) - Var(\bar{x})\right] = \frac{n}{n-1} [\sigma^2 - Var(\bar{x})]$$

donde hemos utilizado el supuesto de que: $E(e_i) = n/N$ tanto si el muestreo es con o sin reemplazamiento, es decir, estamos suponiendo igual posibilidad de selección de cada elemento poblacional.

Observemos ahora que todas las expresiones anteriores para las varianzas teóricas de los estimadores insesgados que hemos considerado, pueden escribirse en función de las cuasivarianzas poblacional y muestral.

11.4.1. Muestreo sin reemplazamiento

En el caso del *muestreo sin reemplazamiento,* podemos representar *la varianza del total poblacional* que obtuvimos en la Sección 11.3.1 como:

$$Var_{SR}(\hat{X}) = N^2 \frac{N-n}{N-1} \frac{\sigma^2}{n} = N^2 \frac{N-n}{N-1} \frac{N-1}{N} \frac{s^2}{n}$$

y expresiones similares pueden obtenerse para los estimadores de la *media, proporción y total de clase poblacionales.* Ahora bien, en este tipo de muestreo, tenemos:

$$E_{SR}(\hat{s}^2) = \frac{n}{n-1}[\sigma^2 - Var_{SR}(\bar{x})] = \frac{n}{n-1}\left[\sigma^2 - \frac{N-n}{N-1}\frac{\sigma^2}{n}\right] =$$

$$= \frac{\sigma^2}{n-1}\left[\frac{Nn - n - N + n}{N-1}\right] = \frac{N}{N-1}\sigma^2 = s^2$$

por lo que *en el muestreo sin reemplazamiento, la cuasivarianza muestral es un estimulador insesgado de la cuasivarianza poblacional* y, por tanto, *la varianza muestral es un estimador insesgado de la varianza poblacional.*

11.4.2. Muestreo con reemplazamiento

En el caso del *muestreo con reemplazamiento* [Sección 11.3.2], tenemos:

$$E_{CR}(\hat{s}^2) = \frac{n}{n-1}[\sigma^2 - Var_{CR}(\bar{x})] = \frac{n}{n-1}\left(\sigma^2 - \frac{\sigma^2}{n}\right) = \sigma^2 = \frac{N-1}{N}s^2$$

por lo que *en el muestreo con reemplazamiento, la cuasivarianza muestral es un estimador insesgado de la varianza poblacional,* pero *un estimador sesgado de la cuasivarianza poblacional.* Nuevamente, este tipo de muestreo se comporta como si pudiese convertir una población finita, en una población infinita.

11.5. ESTIMACIÓN DE LA VARIANZA DE LOS ESTIMADORES DE LA MEDIA, EL TOTAL Y LA PROPORCIÓN POBLACIONALES

Una vez vistas las expresiones analíticas para las varianzas teóricas de las varianzas de los estimadores considerados, nos queda por resolver el problema de evaluar numéricamente dichas varianzas, una vez que disponemos de una muestra particular, puesto que dependen de parámetros desconocidos.

11.5.1. Varianza del estimador de la media poblacional

En el caso de la media muestral, estimador insesgado de la media poblacional, ya vimos que en el *muestreo sin reemplazamiento*:

$$Var_{SR}(\bar{x}) = \frac{N-n}{N-1}\,\frac{\sigma^2}{n}$$

Un *estimador insesgado de la varianza de la media muestral* puede obtenerse, en el *muestreo sin reemplazamiento*, mediante:

$$\widehat{Var_{SR}(\bar{x})} = \frac{N-n}{N}\,\frac{\acute{s}^2}{n}$$

ya que como: $E(\acute{s}^2) = s^2 = \frac{N}{N-1}\sigma^2$, se tiene:

$$E(\widehat{Var_{SR}(\bar{x})}) = \frac{N-n}{N-1}\,\frac{\sigma^2}{n} = Var_{SR}(\bar{x})$$

y tenemos un estimador insesgado de la varianza de la media muestral.

En el caso del muestreo *con reemplazamiento* utilizamos:

$$\widehat{Var_{CR}(\bar{x})} = \frac{\acute{s}^2}{n}$$

como estimador insesgado de la varianza de la media muestral, σ^2/n.

11.5.2. Varianza del estimador del total poblacional

El estimador insesgado del total poblacional es el total muestral, es decir, N veces la media muestral, por lo que el *estimador insesgado de su varianza* se obtiene en cada tipo de muestreo multiplicando por N^2 al estimador insesgado de la varianza de la media muestral.

$$\widehat{Var_{SR}(\hat{X})} = N(N-n)\frac{\acute{s}^2}{n}$$

$$\widehat{Var_{CR}(\hat{X})} = N^2\frac{\acute{s}^2}{n}$$

11.5.3. Varianza del estimador de la proporción poblacional

El estimador insesgado de la proporción poblacional es la proporción muestral, pero el estimador insesgado de *la varianza de ésta última* es algo más complejo de obtener. Observemos, en primer lugar, que la cuasivarianza muestral de la variable indicatriz A_i puede escribirse:

$$s_A^2 = \frac{1}{n-1}\sum_{i=1}^{n}(A_i - p)^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} A_i^2 + np^2 - 2p\sum_{i=1}^{n} A_i\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} A_i + np^2 - 2pnp\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} A_i - np^2\right] = \frac{1}{n-1}(np - np^2) = \frac{1}{n-1}np(1-p)$$

Ya hemos probado anteriormente que, en el *muestreo sin reemplazamiento* (Sección 11.4.1), la cuasivarianza muestral es un estimador insesgado de la cuasivarianza poblacional, de modo que se tiene: $E(s_A^2) = S_A^2$, por lo que:

$$E_{SR}(s_A^2) = E\left[\frac{n}{n-1}p(1-p)\right] = S_A^2 = \frac{N}{N-1}P(1-P)$$

lo que sugiere que, como antes vimos, para estimar la varianza del estimador de la proporción, que es (Sección 11.3.1):

$$Var_{SR}(\hat{P}) = \frac{N-n}{N}\frac{N}{N-1}\frac{P(1-P)}{n}$$

utilicemos:

$$\widehat{Var_{SR}(\hat{P})} = \frac{N-n}{N}\frac{p(1-p)}{n-1}$$

que implica:

$$E[\widehat{Var_{SR}(\hat{P})}] = E[\widehat{Var_{SR}(p)}] = \frac{N-n}{N}\frac{1}{n}E\left[\frac{n}{n-1}p(1-p)\right] =$$

$$= \frac{N-n}{N}\frac{1}{n}\frac{N}{N-1}P(1-P) = Var_{SR}(p) = Var_{SR}(\hat{P})$$

por lo que habremos estimado insesgadamente la varianza de la proporción muestral, que es el estimador insesgado de la proporción poblacional.

El mismo tipo de consideraciones hace que en el *muestreo con reemplazamiento* utilicemos como estimador de la varianza del estimador de la proporción poblacional, que es la proporción muestral, la expresión:

$$\widehat{Var_{CR}(\hat{P})} = \frac{P(1-P)}{n-1}$$

que es, de nuevo, insesgado:

$$E[\widehat{Var_{CR}(\hat{P})}] = E\left(\frac{\hat{P}(1-\hat{P})}{n-1}\right) = \frac{1}{n-1}E_{CR}\left(\frac{n-1}{n}\hat{s}_A^2\right) = \frac{1}{n}E_{CR}(\hat{s}_A^2) = \frac{\sigma^2}{n} = Var_{CR}(p)$$

En ambos casos se obtiene el estimador de la varianza del estimador del total de clase sin más que multiplicar por N^2 la varianza del estimador de la proporción, ya que estimamos el Total de clase mediante: $\hat{A} = N\hat{P}$. Por ejemplo, en el muestreo *sin reemplazamiento.*

$$\widehat{Var_{SR}(\hat{A})} = N^2\,\widehat{Var_{SR}(\hat{P})} = N(N-n)\frac{\hat{P}(1-\hat{P})}{n-1}$$

$$\widehat{Var_{CR}(\hat{A})} = N^2\,\widehat{Var_{CR}(\hat{P})} = \frac{N^2\hat{P}(1-\hat{P})}{n-1}$$

Estas expresiones pueden escribirse asimismo en función del total de clase muestral. Denotando por a el *total de clase muestral*, tenemos:

$$\hat{P} = p = \frac{1}{n}\sum_1^n A_i = \frac{a}{n}; \quad 1-\hat{P} = 1-p = \frac{n-a}{n};$$

de modo que:

$$\widehat{Var_{SR}(\hat{P})} = \widehat{Var_{SR}(p)} = \frac{N-n}{N}\,\frac{1}{n-1}\,\frac{a(n-a)}{n^2}$$

en el caso del *muestreo sin reemplazamiento*, y:

$$\widehat{Var_{CR}(\hat{P})} = \frac{a(n-a)}{n^2(n-1)}$$

en el caso del *muestreo con reemplazamiento.*

11.6. APLICACIONES DEL MUESTREO EN POBLACIONES FINITAS

En la sección anterior hemos derivado estimadores insesgados para el total, la media y la proporción poblacionales, y hemos calculado sus varianzas teóricas; finalmente, hemos deducido expresiones factibles para la estimación de tales varianzas, utilizando la información disponible en la muestra. Una de las conclusiones básicas es que la varianza teórica de los estimadores insesgados depende de que el muestreo se haya realizado con o sin reemplazamiento. También hemos visto que el reemplazamiento proporciona expresiones para las varianzas de los estimadores que son idénticas a las que tienen en poblaciones infinitas.

La primera implicación de dicho estudio es que, *si se conoce la distribución poblacional*, entonces podemos establecer afirmaciones probabilísticas acerca de los valores posibles de los estadísticos muestrales, previamente a la extracción de una muestra y al cálculo de los mismos. Esto es lo que analizamos en esta sección.

En otras ocasiones, *desconoceremos las características de la distribución poblacional*, y extraeremos una muestra para, a partir de los estadísticos muestrales, estimar los parámetros poblacionales. Como todo problema de estimación, este ejercicio puede efectuarse de modo puntual, o mediante intervalos de confianza. Este último es el análisis que efectuamos en la sección siguiente.

A continuación repasamos algunas cuestiones sencillas que utilizan los conceptos básicos de las secciones anteriores.

Ejemplo 11.1. El gasto medio diario de una familia en una determinada población es, en miles de ptas., $\mu = 6{,}5$, con una desviación típica de: $\sigma = 6{,}2$. ¿Cuál es la probabilidad de que al seleccionar una muestra aleatoria simple de 100 familias obtengamos en ellas un gasto medio mayor o igual que 5,8?

En este caso, damos por conocidos los estadísticos poblacionales, y se nos pide la probabilidad de un suceso aleatorio referente a la distribución de la media muestral, a priori, es decir, antes de obtener la información muestral. La población es prácticamente infinita, por lo que la media muestral tiene como esperanza matemática la media poblacional, y como desviación típica, la desviación típica poblacional, dividida por la raíz cuadrada del tamaño muestral, que es 10, es decir 0,62. Suponemos que el conjunto total de las calificaciones sigue un comportamiento Normal. Ello implica que la media muestral es una variable aleatoria con distribución de probabilidad asimismo Normal [2]. La probabilidad solicitada es, por tanto, de 0,871.

La probabilidad de que el gasto medio de una de estas familias exceda de 5,8 es de 0,545, muy inferior a la anterior. La probabilidad con la media es mayor porque la media muestral es mucho más precisa que las observaciones aisladas, y puesto que 5,8 es una cifra de gasto relativamente baja, podemos estar muy seguros de que casi cualquier muestra de 100 familias satisfará dicha condición, lo que no ocurriría necesariamente con familias individuales. Al considerar la media muestral, las familias que tienen un gasto bajo tienden a compensarse con las que tienen un gasto alto. Este tipo de compensaciones no se produce al considerar una sola familia, por lo que no podemos garantizar tanto como con la media muestral que vayamos a obtener un valor igual o superior a 5,8.

En cada caso analizamos cuestiones diferentes: al observar una sola familia, nos preocupamos acerca de la distribución de las cifras de gasto individuales, es decir, de la distribución poblacional; en el caso de la media muestral, nos interesa la distribución de la media en una muestra de 100 familias.

La probabilidad de que el gasto medio de una muestra de 100 familias esté comprendida entre 5,5 y 7,0 es de: 0,7366. Como interpretación, puede decirse que si se toman 50 muestras, cada una conteniendo 100 familias, el número esperado de ellas que tendrá media igual o inferior a 7,0, es de 39. El número de muestras que tendrán un gasto medio comprendido entre 5,5 y 7,0 es de 37.

---

[2] En cualquier caso, el Teorema Central del Límite [Teorema 8.26] justifica que tratemos la media muestral como aproximadamente Normal.

Ejemplo 11.2. La distribución de los ingresos mensuales de una población de 6.400 familias tiene una media: $\mu = 365.000$ pts., con desviación típica de 72.000 pts. Si se extrae una muestra de tamaño 400 sin reemplazamiento, la distribución de la media muestral tiene esperanza matemática igual a 365.000 pts, y desviación típica:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{72000}{20}\sqrt{\frac{6400-400}{6400-1}} = 3600\sqrt{\frac{6000}{6399}} = 3600(0.968) = 3485$$

donde hemos utilizado el hecho de que el muestreo es sin reemplazamiento, lo que ocurre en encuestas de este tipo: tanto al encuestar precios para la construcción del IPC, como al entrevistar a familias para generar la Encuesta de Población Activa, o si se efectúa una encuesta de rentas familiares, no tendría ningún sentido preguntar dos veces a la misma familia. Si el muestreo hubiese sido con reemplazamiento, la desviación típica de la media muestral hubiese sido:

$$\sigma_{\bar{x}} = \frac{72000}{\sqrt{400}} = 3600$$

un 3,3 % superior al que hemos obtenido. Cuando la muestra es mayor, en relación con el tamaño de la población, la diferencia entre las desviaciones típicas de la media bajo ambos tipos de muestreo se hace menos evidente. Cuando el tamaño de la muestra es aproximadamente un 5 % del tamaño de la población no tiene mucho interés efectuar la corrección de población finita.

En un primer análisis, los ejercicios anteriores no parecen tener mucho interés práctico. Generalmente desconocemos los parámetros poblacionales, por lo que no podemos efectuar inferencias acerca de una muestra. Además, ¿cuál sería el interés práctico de interesarnos por las características de una muestra cuando conocemos las características de la población total? Dicho interés existe, sin embargo, en los denominados problemas de *control de calidad*. En ellos, suponemos que las características poblacionales son conocidas, si no con certeza, al menos sí con una cierta garantía, basada en la experiencia de funcionamiento de un determinado proceso y estamos muy interesados en que la calidad de dicho proceso no se deteriore, lo que ocurrirá si varían las características de su distribución de probabilidad. El siguiente ejemplo presenta una situación de control de calidad:

Ejemplo 11.3. Una planta empaquetadora de detergentes envasa cada paquete con un promedio de $\mu = 2$ Kg, con desviación típica de: $\sigma = 250$ gr. Para controlar la calidad del empaquetado, se toman cada día 64 envases y, si el peso medio es inferior a un cierto valor crítico, se reajusta el proceso de envasado, lo cual es un procedimiento costoso. ¿Cuál debe ser el valor crítico utilizado si no quiere detener el proceso en más de uno de cada 10 días?

Esto equivale a diseñar un proceso de control del envasado de modo que la probabilidad de detener dicho proceso cada día, sea de 0,10. Suponemos que las mediciones de peso de cada envase siguen un comportamiento Normal, con lo que la media muestral también lo será, y las tablas de la Normal nos dicen que 0,10 es la probabilidad por debajo de $-1{,}28$. La esperanza matemática de la media muestral es, en este proceso de control de calidad: $\mu = 2000$ gr., con desviación típica:

$$\sigma_{\bar{x}} = \frac{250}{\sqrt{64}} = \frac{250}{8} = 31{,}25$$

Por tanto, el valor crítico debe escogerse:

$$0{,}10 = P\left[\frac{\lambda - 2000}{31{,}25} \leqslant -1{,}28\right] \Rightarrow \lambda = (31{,}25)(-1{,}28) + (2000) = 1960$$

y debemos decidir detener el proceso de envasado si el peso medio de los 64 envases medidos es inferior a este valor numérico.

Al hacer este ejercicio, hemos supuesto implícitamente que el número de envases rellenados cada día es prácticamente ilimitado, en relación con el tamaño muestral de 64. Supongamos ahora por un momento que al día se rellenasen 400 envases. En tal caso, la desviación típica de la media de la muestra (que es, obviamente, sin reemplazamiento) sería:

$$\sigma_{\bar{x}} = \frac{250}{\sqrt{64}}\sqrt{\frac{400-64}{400-1}} = \frac{250}{8}\sqrt{\frac{336}{399}} = 31{,}25(0{,}918) = 28{,}68$$

El valor crítico debería ser:

$$0{,}10 = P\left[\frac{\lambda - 2000}{28{,}68} \leqslant -1{,}28\right] \Rightarrow \lambda = (28{,}68)(-1{,}28) + (2000) = 1963{,}29$$

En este ejemplo, la empresa conoce, basada en una larga experiencia previa, las características de su proceso de envasado, y está muy interesada en detectar posibles desviaciones del mismo, es decir, *pérdidas de calidad*. Para ello se lleva a cabo diariamente el muestreo descrito, y es interesante conocer de antemano las características que revestirá dicho proceso de muestreo.

El tratamiento de *la proporción* es similar al de la media muestral que ya hemos analizado. Recuérdese que la proporción no es sino la media muestral de una variable indicatriz, que hemos denotado por A_i, que indica si el elemento i-ésimo de la población satisface o no una determinada característica en estudio. Por tanto, al igual que cualquier otra media muestral, el teorema central del límite justifica que podamos suponer que la distribución de la proporción puede aproximarse por una Normal. La esperanza matemática de la proporción muestral, p, es la proporción poblacional, P. La varianza de la proporción muestral es igual a $P(1-P)/n$, siendo n el tamaño de la muestra, si ésta es con reemplazamiento, o si su tamaño es inferior a un 5 % del tamaño de la población. En otro caso, su varianza debe ajustarse por el mismo factor de corrección que utilizamos en el caso de la varianza de la media muestral [Sección 11.3]. El siguiente ejemplo trata este caso:

Ejemplo 11.4. Se sabe que el 60 % de los alumnos de la Facultad de CCEE de la Universidad Complutense de Madrid tiene un promedio de calificaciones inferior a 6. Si tomamos una muestra aleatoria de tamaño 100, ¿cuál es la probabilidad de que, de esa muestra, la mitad de los alumnos o más tenga calificación inferior a 6?

La esperanza matemática de la proporción de alumnos que satisfacen dicha condición en la muestra de 100 alumnos es 0,60. La desviación típica es:

$$\sigma_p = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{(0.60)(0.40)}{100}} = \sqrt{\frac{0.24}{100}} = 0.049$$

y sabemos que la probabilidad solicitada es:

$$P\left[\frac{p-P}{\sigma_p} > \frac{0.50-0.60}{0.049}\right] = P[N(0.1) > -2.04] = 0,9793$$

es decir, que estamos considerando un suceso de muy probable ocurrencia.

Algunos autores recomiendan incorporar una corrección por continuidad, que consiste en sumar a la proporción solicitada, 0,50, el término: $1/(2n)$, que es 0,005 en nuestro caso. Ello conduce a una probabilidad de 0.9738 en la situación anterior.

El problema puede resolverse igualmente, partiendo de la aproximación Normal a la distribución del *número de alumnos* que tienen una calificación superior a 6. Cada alumno responde a una distribución binomial, que toma el valor 1 si el alumno satisface dicha condición, y 0 en caso contrario. El número esperado de alumnos en la muestra con calificación superior a 6 es, de acuerdo con los parámetros poblacionales: $\mu = (100)\ (0,60) = 60$, y su desviación típica es:

$$\sigma_{n_A} = \sqrt{nP(1-P)} = \sqrt{100(0,60)(0,40)} = 4,90$$

por lo que:

$$P\left[\frac{n_A - E(n_A)}{\sigma_{n_A}} > \frac{50-60}{4.90}\right] = P[N(0,1) > -\ 2,04] = 0,9793$$

como antes.

Al resolver este problema, nuevamente hemos supuesto de manera implícita que el tamaño poblacional, el número total de estudiantes, es prácticamente ilimitado, o al menos que el tamaño muestral de 100 no excede del 5 % del tamaño de la población. Supongamos ahora que el conjunto total de alumnos fuese de 800. Entonces, la desviación típica de la proporción muestral sería:

$$\sigma_p = \sqrt{\frac{P(1-P)}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{(0,60)(0,40)}{100}}\sqrt{\frac{800-100}{800-1}} =$$

$$\sqrt{\frac{0,24}{100}}\sqrt{\frac{700}{899}} = (0,049)(0,882) = 0,043$$

con una probablidad:

$$P\left[\frac{p-P}{\sigma_p} > \frac{0,50-0,60}{0,043}\right] = P[N(0,1) > -2,33] = 0,9901$$

11.7. INTERVALOS DE CONFIANZA PARA LOS ESTADÍSTICOS POBLACIONALES: MEDIA, TOTAL, PROPORCIÓN Y TOTAL DE CLASE

En las secciones previas obtuvimos estimadores insesgados de cuatro parámetros poblacionales de gran relevancia: la media y total poblacionales, así como la proporción y total de clase. En ocasiones, es importante disponer de intervalos de confianza de los mismos, pues pueden utilizarse para contrastar hipótesis acerca de sus valores teóricos, cuando éstos son desconocidos.

Las varianzas de los estimadores insesgados y, en particular, de la media y la proporción muestrales, dependen de parámetros poblacionales. Por tanto, cuando éstos son desconocidos, también lo son dichas varianzas, lo que impide la construcción de intervalos de confianza. Pero hemos mostrado en la Sección 11.5 cómo estimar de modo insesgado las varianzas de estos estimadores, por lo que estamos en condiciones de generar intervalos de confianza para los parámetros poblacionales (media y total poblacionales, proporción poblacional y total de clase) y, con ello, contrastar hipótesis acerca de sus valores desconocidos.

Sin embargo, en la construcción de un intervalo es crucial conocer la familia de distribuciones de probabilidad que sigue el estimador. El estimador insesgado de la media poblacional es la media muestral, y el del total es la media, multiplicada por el tamaño poblacional, N. En consecuencia, si la población de la que se muestrea es aproximadamente Normal[3], entonces las tablas de probabilidad de la distribución Normal nos permitirán generar intervalos de confianza. Sólo tendremos que tener cuidado de utilizar la expresión adecuada para la varianza del estimador, según el tipo de muestreo que se haya realizado. Cuando no lo es, el Teorema Central del Límite justifica que podemos aproximar la distribución de la media y la proporción muestrales por distribuciones Normales si el tamaño muestral es suficientemente grande.

El proceso de elaborar intervalos de confianza es totalmente idéntico al que describimos en el Capítulo 10, lo que mostramos en el siguiente ejemplo para la media:

[3] Estrictamente hablando, no puede serlo, por ser finita.

Ejemplo 11.5. Con objeto de estimar el salario medio de todos los profesores de la Universidad Complutense, se encuesta a 900 profesores, tomados al azar de una relación completa de los mismos. El salario medio mensual de los encuestados es de 250.000 pts. con una desviación típica muestral de 810.000 pts. La desviación típica de la media muestral es:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{810.000}{\sqrt{900}} = \frac{810.000}{30} = 27.000 \text{ pts.}$$

por lo que, con un 95 % de confianza, la esperanza matemática se halla en el intervalo: [250.000 − (1,96)(27.000); 250.000 + (1,96)(27.000)] = (197.080 ptas.;302.920 ptas.). La estimación puntual del salario medio poblacional, es decir, del salario medio de *todos* los profesores de la universidad, es la media muestral, 250.000 ptas.

En realidad, para poder utilizar con todo rigor los valores críticos que la distribución Normal asocia a cada nivel de probabilidad deberíamos conocer la desviación típica poblacional, lo que no ocurre en nuestro caso. Cuando no se conoce sino una estimación suya, como aquí ocurre, entonces los niveles críticos deberían provenir de una distribución t de Student. Sin embargo, para tamaños muestrales superiores a 30, los valores críticos que se obtienen de las tablas de probabilidad de las distribuciones Normal y de Student son prácticamente indistinguibles.

Si quisiésemos una mayor confianza en la estimación, por ejemplo, del 99 %, entonces comenzaríamos determinando, a partir de las tablas de la distribución de probabilidad Normal, que el intervalo (−2,58;2,58) comprende dicha cantidad de probabilidad. Por tanto, tomando a la izquierda y a la derecha de la media muestral 2,58 veces su desviación típica, que es en este ejemplo de 27.000 ptas., obteniendo un intervalo: (180.340 ptas.;319.660 ptas.), que es más amplio que el intervalo del 95 % de confianza, como consecuencia de que proporciona una mayor confianza de contener al salario medio poblacional, que es desconocido.

El intervalo de confianza del 90 % se construye tomando 1,645 veces la desviación típica. En nuestro ejemplo, el intervalo de confianza del 90 % para el salario medio es de: (205.585 ptas.;294.415 ptas.). La estimación puntual del salario medio poblacional continúa siendo en ambos casos de 250.000 ptas., la media muestral, pues tal estimación no depende del nivel de confianza.

Si quisiésemos obtener un estimador puntual de la nómina mensual que la Universidad Complutense destina a sus docentes (el *Total poblacional*), bastaría con aplicar la media muestral a cada uno de sus profesores. Lo mismo ocurre si quisiésemos obtener un intervalo de confianza para la nómina total de profesores. Así, supongamos que la Universidad tiene contratados a 2.500 docentes. Su nómina estimada sería de: (250.000 ptas.)(2.500 profesores) = 625 millones de ptas. El intervalo de confianza del 99 % para la nómina mensual, que acabamos de estimar en 625 millones de ptas., es (450.850.000 ptas.;799.150.000 ptas.) y el intervalo de confianza del 95 % es de (492.700.000 ptas.;757.300.000 ptas.).

La estimación puntual y por intervalos de la proporción poblacional se resuelve de modo totalmente similar al de la media poblacional. Ello se debe a que, si la población es Normal, la proporción poblacional p también obedece a una distribución Normal, con esperanza matemática igual a la proporción poblacional P, y varianza igual a: $\sigma_p^2 = P(1 - P)/n$, siendo n el tamaño de la muestra en que se calculó p. Si la población fuese finita, entonces debemos introducir en la varianza la corrección de poblaciones finitas, llegando a:

$$\sigma_p = \sqrt{\frac{P(1-P)}{n}}\ \sqrt{\frac{N-n}{N-1}}$$

que es irrelevante cuando el tamaño muestral n es inferior al 5 % del tamaño poblacional N.

La segunda dificultad estriba en que, desconociendo la verdadera proporción poblacional P, no podemos calcular la varianza de su estimador, la proporción muestral. Por tanto, en las expresiones anteriores sustituimos el valor desconocido de P por su estimación, p. Esta aproximación debería conducirnos a una distribución t de Student en la que calcular los valores críticos correspondientes a los distintos niveles de confianza pero, como dijimos antes, la distinción entre dicha distribución y la Normal es mínima para tamaños muestrales superiores a 30. El siguiente ejemplo analiza este caso:

Ejemplo 11.6. En una muestra aleatoria simple de 400 alumnos y alumnas de primer curso de la licenciatura de Economía de la Universidad Complutense aficionados al fútbol, un 42 % señaló al Atlético de Madrid como su equipo de fútbol preferido.

Si, con esta información quisiéramos estimar el porcentaje de alumnos de primer año que son hinchas de dicho equipo, lo haríamos mediante la proporción muestral, 0,42. Si queremos disponer de un intervalo de confianza del 90 %, comenzaríamos estimando la varianza de la proporción muestral, mediante:

$$\sigma_P = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0{,}42)(0{,}58)}{400}} = 0{,}0247$$

de modo que diríamos que, con una confianza del 90 %, la proporción poblacional.para todos los alumnos de primer curso, que es desconocida, está comprendida entre $0{,}42 + (1{,}645)(0{,}0247)$ y $0{,}42 + (1{,}645)(0{,}0247)$, es decir, entre 0,38 y 0,46.

Si quisiésemos estimar el *número* de alumnos que tienen a dicho equipo de fútbol como favorito, es decir, el total de clase, lo haríamos multiplicando el tamaño de la población por la proporción estimada. Supongamos que queremos estimar el número total. de entre los 12.000 alumnos de la Universidad Complutense que satisfacen dicha condición, y que creemos (lo que puede ser cuestionable) que los alumnos de primer curso de Economía constituyen una sub-población representativa respecto a los gustos futbolísticos. En tal caso, tendríamos una estimación del total de clase de: $A = (12.000)(0{,}42) = 5.040$ alumnos que tienen al Atlético de Madrid como su equipo de fútbol favorito.

Para obtener un intervalo de confianza, basta multiplicar el tamaño de la población, 12.000 alumnos, por el intervalo de igual nivel de confianza para la proporción. Así, para obtener un intervalo de confianza del 90 % para el total de clase, multiplicamos 12.000 por los extremos inferior y superior antes obtenidos, 0,38 y 0,46, y obtenemos que, el intervalo de confianza del 90 % acerca del número de alumnos de la Universidad Complutense que comparten dicha característica sería de entre 4.560 y 5.520 alumnos. La estimación puntual que antes obtuvimos ocupa la posición central en este intervalo.

Supongamos ahora que no considerásemos al grupo de alumnos de primer curso de Economía como representativo de los gustos deportivos de los alumnos de toda la universidad, por lo que pretendemos extraer inferencia acerca tan sólo de dicho grupo, que vamos a suponer constituido por un total de 1.500 alumnos. En tal caso, el tamaño muestral es un porcentaje apreciable del tamaño poblacional: 400/1500 = 26,7 %, por lo que no podemos ignorar la corrección que en la varianza debemos introducir cuando el tamaño de la población no es suficientemente superior al de la muestra (lo que denominamos *tamaño finito*). En este caso, estimaríamos la desviación típica de la proporción muestral por:

$$\sigma_P = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{(0{,}42)(0{,}58)}{400}}\sqrt{\frac{1500-400}{1500-1}} = (0{,}0247)(0{,}857) = 0{,}021$$

por lo que el intervalo de confianza del 90 % estaría comprendido entre 0,385 y 0,455. La estimación del total de clase sería: $A = (0{,}42)(1500) = 630$ alumnos favorables al Atlético de Madrid, mientras que el intervalo de confianza del 90 % para dicho total de clase sería de: [(0,385)(1.500); (0,455)(1.500)] = [577 alumnos;683 alumnos].

11.8. DETERMINACIÓN DEL TAMAÑO ÓPTIMO MUESTRAL PARA LA ESTIMACIÓN DE PARÁMETROS POBLACIONALES

Hasta ahora no nos hemos cuestionado acerca del tamaño que debe tener una muestra y, sin embargo, es éste un aspecto de gran importancia. Generalmente, aumentar el tamaño muestral viene asociado con un incremento del coste de muestreo, si bien este aspecto no lo estamos considerando. Por otra parte, si, con motivo de reducir costes, extraemos una muestra muy pequeña, entonces nuestras estimaciones, basadas en estadísticos muestrales, no tendrán suficiente fiabilidad. Por tanto, se trata de conseguir un equilibrio entre ambas alternativas, y decidir acerca del tamaño muestral que nos permita, con un mínimo coste, garantizar en nuestra estimación un nivel de confianza fijado de antemano, así como que estemos por debajo de un máximo error admisible, por encima del cual la estimación carecería de interés.

Dado un nivel de confianza a, denominamos *error de estimación*, denotado por E_M a la máxima diferencia que queremos permitir, con nivel de confianza $100(1-a)$ %, entre el parámetro desconocido y el estadístico utilizado como estimador. Por ejemplo, para la media, tenemos:

$$-E_M \leqslant \bar{x} - \bar{X} \leqslant E_M \Leftrightarrow -\frac{E_M}{\sigma_{\bar{x}}} \leqslant \frac{\bar{x}-\bar{X}}{\sigma_{\bar{x}}} \leqslant \frac{E_M}{\sigma_{\bar{x}}} \Leftrightarrow -\frac{E_M}{\sigma}\sqrt{n} \leqslant \frac{\bar{x}-\bar{X}}{\sigma}\sqrt{n} \leqslant \frac{E_M}{\sigma}\sqrt{n}$$

Ahora bien, $\bar{X}$ es un parámetro desconocido, y es la esperanza matemática de $\bar{x}$. La variable en el centro está tipificada, por lo que si la población es Normal, es una variable $N(0,1)$. Fijado un nivel de confianza $1-a$, tomamos de las tablas de dicha distribución la abscisa λ_a que deja a su derecha una probabilidad de $a/2$, de modo que tenemos:

$$E_M = \lambda_a \frac{\sigma_x}{\sqrt{n}}$$

donde hemos representado E_M como función, tanto de la desviación típica poblacional σ_x, como del tamaño muestral. Supongamos ahora que fijamos un máximo error admisible E_M, así como un determinado nivel de confianza a. En tal caso, despejando en la expresión anterior, tenemos:

$$n = \lambda_a^2 \frac{a_x^2}{E_M^2}$$

en la que podemos apreciar que:

a) cuanto menor sea el máximo error admisible, es decir, cuanto más exigentes seamos en términos del error que toleramos en la estimación, mayor será el tamaño muestral que debemos recoger,

b) cuanto mayor sea el nivel de confianza deseado, mayor será el nivel crítico de las tablas, λ_a, y mayor será el tamaño muestral necesario para garantizar dicho nivel de confianza,

c) cuanto mayor sea la varianza de la distribución de la que extraemos la muestra, menor será la precisión informativa de los elementos que de ella extraemos y, en consecuencia, será preciso un mayor tamaño muestral para poder garantizar los niveles de precisión requeridos de antemano.

Ejemplo 11.7. Con objeto de estimar el precio medio del kg. de pollo en un determinado mes, un encuestador visita un cierto número de puntos de venta. Supongamos que queremos cometer un error máximo en dicha estimación de 5 ptas., y que la desviación típica del precio del pollo en encuestas de meses previos ha permanecido siempre en torno a 12 ptas. Ello implica que, de mantenerse dicha desviación típica del precio, la media muestral tiene una desviación típica igual a $12/n$, siendo n el tamaño muestral que se escoja. Al 90 % de confianza, el número de puntos de venta que deben visitarse es de:

$$n = \lambda_a^2 \frac{\sigma_x^2}{E_M^2} = 1{,}645^2 \frac{144}{25} = 15{,}58$$

es decir, 16 puntos de venta.

Para alcanzar niveles de confianza del 95 % y 99 %, deberían visitarse:

$$n = \lambda_a^2 \frac{\sigma_x^2}{E_M^2} = 1{,}96^2 \frac{144}{25} = 22{,}13$$

$$n = \lambda_a^2 \frac{\sigma_x^2}{E_M^2} = 2{,}58^2 \frac{144}{25} = 38{,}34$$

23 y 39 puntos de venta, respectivamente. Vemos que el número de visitas aumenta con el nivel de confianza que se pretende lograr.

Si el máximo error admisible fuese de 3 ptas., el número de puntos de venta necesarios al 90 % de confianza sería:

$$n = \lambda_a^2 \frac{\sigma_x^2}{E_M^2} = 1{,}65^2 \frac{144}{9} = 43{,}56$$

44, apreciablemente superior al que era preciso visitar con un máximo error, más tolerable, de 5 ptas./kg. Los puntos de venta precisos con este máximo error de 3 ptas./kg. pasan a ser de 62 y 107 para niveles de confianza del 95 % y 99 %.

Supongamos ahora que, trabajando con un máximo error de 3 ptas./kg., y al 90 % de confianza, se visitan 44 puntos de venta, obteniendo un precio medio de 325 ptas./kg., con desviación típica muestral de 14,2 pts./kg.. Ello hace que el error de estimación sea de:

$$E_M = \lambda_a \frac{\sigma_x}{\sqrt{n}} = 1{,}645 \frac{14{,}2}{\sqrt{44}} = 3{,}52 \text{ ptas./kg}$$

que excede del nivel máximo que nos habíamos fijado de antemano. Ello se debe a que la desviación típica muestral ha resultado superior a la poblacional, lo que arroja dudas acerca de que el valor histórico de $\sigma_x = 12$ ptas./kg. continúe siendo válido. Cuando esto ocurre, generalmente se sustituye el valor poblacional histórico de 12 por el muestral de 14,2, y determinamos nuevamente el tamaño muestral preciso, obteniendo:

$$n = \lambda_a^2 \frac{\sigma_x^2}{E_M^2} = 1{,}645^2 \frac{14{,}2^2}{3^2} = 61$$

61 puntos de venta. Es decir, para lograr un error de estimación de 3 ptas., inferior al de 3.52 ptas. que obtuvimos, deberíamos haber visitado un mayor número de puntos de venta.

En el ejemplo anterior nos hemos abstraído del posible tamaño finito de la población. Estrictamente hablando, los cálculos anteriores sólo son válidos si dicha población fuese infinita, es decir, si los tamaños muestrales que han surgido en el ejemplo anterior no excediesen del 5 % del número total de puntos de venta, lo cual parece un supuesto razonable. Consideremos ahora el caso de una población de tamaño finito. El intervalo de confianza para la media poblacional es, en tal caso:

$$\bar{x} \pm \lambda_a \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \bar{x} \pm E_M$$

donde podemos ver que la expresión del error de estimación se ve afectada por el factor de corrección de poblaciones finitas:

$$E_M = \lambda_a \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

y, despejando, tenemos la expresión para el tamaño muestral que garantiza un determinado nivel de confianza y un máximo error de estimación prefijado:

$$n = \frac{\lambda_a^2 \sigma_x^2 N}{(N-1) E_M^2 + \lambda_a^2 \sigma_x^2}$$

Ejemplo 11.8. Volviendo a la estimación del precio del kg. de pollo, con un máximo error de 5 ptas./kg., y una desviación típica de 12 ptas./kg., supongamos que el número total de puntos de venta es de 100. Tendríamos en tal caso:

$$n = \frac{\lambda_a^2 \sigma_x^2 N}{(N-1) E_M^2 + \lambda_a^2 \sigma_x^2} = \frac{1{,}645^2 (144)(100)}{(99)(25) + 1{,}645^2 (144)} = \frac{38.967}{2.865} = 13{,}6$$

es decir, 14 puntos de venta. El lector puede comprobar sin gran dificultad que, al igual que ocurre con poblaciones infinitas, el tamaño muestral, es decir, el número de puntos de venta precisos, crece con el nivel de confianza que se pretende alcanzar, así como con la varianza poblacional, y también cuando exigimos más de la estimación, permitiéndonos un menor error máximo admisible.

El caso de la proporción se analiza de modo totalmente similar, como vamos a ver:

Ejemplo 11.9. Supongamos que, con el objeto de diseñar correctamente las estaciones de servicio, una empresa distribuidora de gasolina quiere estimar el porcentaje de sus clientes que utiliza gasolina sin plomo. El máximo error de estimación, para la proporción es:

$$E_M = \lambda_a \sqrt{\frac{P(1-P)}{n}}$$

Es posible que la empresa haya realizado encuestas similares en ocasiones previas, y que tenga una cierta idea acerca de la varianza de la proporción muestral, para lo cual debe tener una cierta idea previa acerca del valor desconocido de la proporción que quiere estimar, P. Además, si lo conociera no tendría sentido estimarlo salvo si cree que ha podido haber algún cambio. Incluso entonces, la estrategia que vamos a sugerir es adecuada. Cuando P es desconocida, se sustituye en la expresión anterior por 0,50. Esto se debe a que para dicho valor, el radicando es máximo y, con él, el error de estimación, lo que hace que sesguemos al análisis hacia extraer muestras excesivamente grandes, lo cual, a su vez, garantizará aún más el cumplimiento de los criterios que exijamos acerca del nivel de confianza y del error de estimación que realmente cometamos.

Despejando en la expresión del error de estimación, y sustituyendo P por 0,50, tenemos:

$$n = \frac{\lambda_a^2 P(1-P)}{E_M^2} = \frac{(1{,}645)^2 (0{,}25)}{(0{,}05)^2} = 270{,}60$$

donde hemos utilizado un nivel de confianza del 90 % y un máximo error de 5 %. Deberíamos encuestar entonces a 273 clientes. A niveles de confianza del 95 % y 99 %, dichos números se elevarían a 385 y 666 clientes.

Si la empresa tuviera el convencimiento de que la proporción de clientes que demandan gasolina sin plomo es del 30 %, o si quisiese contrastar que tal porcentaje, estimado en ocasiones previas, no ha variado significativamente, efectuaría un seguimiento de:

$$n = \frac{\lambda_a^2 P(1-P)}{E_M^2} = \frac{(1{,}645)^2 (0{,}30)(0{,}70)}{(0{,}05)^2} = 227{,}3$$

clientes.

Supongamos ahora que la población con la que trabajamos sea finita. En tal caso, la varianza de la proporción muestral se ve afectada por el factor de corrección ya conocido, por lo que el error de estimación es, en este caso:

$$E_M = \lambda_a \sqrt{\frac{P(1-P)}{n}} \sqrt{\frac{N-n}{N-1}}$$

y despejando, llegamos a la expresión para el tamaño muestral:

$$n = \frac{\lambda_a^2 P(1-P) N}{(N-1) E_M^2 + \lambda_a^2 P(1-P)}$$

Veamos otra situación de control de calidad:

Ejemplo 11.10. Consideremos una empresa que fabrica 2.400 artículos homogéneos al día, un porcentaje de los cuales resulta defectuoso. De muestreos realizados anteriormente, se conoce que el porcentaje de elementos que resultan defectuosos es del 7,5 %. Queremos hallar el tamaño de la muestra que debe tomarse diariamente con objeto de estimar el porcentaje de elementos defectuosos, si se pretende tener un error máximo en la estimación de 1 %, y se trabaja al 90 % de confianza:

$$n = \frac{\lambda_a^2 P(1-P) N}{(N-1) E_M^2 + \lambda_a^2 P(1-P)} = \frac{(1{,}645)^2 (0{,}075)(0{,}925)(2.400)}{(2.399)(0{,}01)^2 + (1{,}645)^2 (0{,}075)(0{,}925)} = \frac{450{,}6}{0{,}4276} = 1053{,}8$$

artículos, un número muy elevado en relación con la producción diaria, que podría resultar inviable. Ello se debe al *reducido error máximo* que pretendemos lograr en la estimación.

El lector puede comprobar que, si nos permitiésemos hasta un 3 % de error en la estimación, entonces el tamaño muestral debería ser de 193, sensiblemente inferior al anterior. En tal caso, con una muestra de 264 elementos, todavía inferior a la primera, podríamos alcanzar un nivel de confianza del 95 %, y con 423 elementos lograríamos un nivel de confianza del 99 %.

EJERCICIOS

11.1. Para controlar la calidad de una fábrica de chips para ordenador se toma aleatoriamente una muestra al final del día de 64 piezas de la producción total, que es de 5.000 unidades y se contrasta si son defectuosas. Si el porcentaje de piezas defectuosas es λ o superior, se corrige el proceso técnico de fabricación, lo cual es costoso. Calcule el valor de λ de modo que la probabilidad de ajustar un proceso que genera un 10 % de piezas defectuosas sea del 0,95.

11.2. Repita el ejercicio anterior bajo el supuesto de que la producción diaria fuese tan sólo de 300 unidades.

11.3. De los 2 millones de familias del área metropolitana de Madrid, un 26 % está suscrito al periódico *La Gaceta*. Encuentre la probabilidad de que, al encuestar a 10.000 familias, la proporción de suscriptores sea superior al 30 %. ¿Cambiaría su respuesta si en vez de Madrid se tratase de una población de la provincia con 1.500 habitantes?

11.4. Utilizando la expresión para el tamaño muestral preciso para alcanzar un cierto nivel de confianza y no exceder de un máximo error admisible, demuestre que dicho tamaño crece con el nivel de confianza que se pretende alcanzar, así como con la varianza poblacional, y también cuando exigimos más de la estimación, permitiéndonos un menor error máximo admisible.

11.5. El servicio de Sanidad de una determinada provincia está efectuando una encuesta para determinar si los niños que viven en un ambiente rural han recibido adecuadamente un programa de vacunación recomendado. Si se desea que la estimación de la proporción de niños que están vacunados no se separe en más de 0,05 % de la verdadera proporción (desconocida) con una probabilidad de 0,98, ¿cuál es el número mínimo de niños que deben encuestarse? ¿Cuál sería dicho número si pudiésemos suponer que la verdadera proporción de niños vacunados es del 65 %?

11.6. Supongamos que la tasa de paro viene siendo del 12 %, pero queremos actualizar nuestra estimación con vistas a diseñar el escenario de política económica para el próximo año. Para ello, queremos tener una confianza del 99 % en que la nueva estimación no se aparte más de un 0.001 de la verdadera tasa de paro. ¿A cuántas personas activas deberemos encuestar para estimar la tasa de paro? ¿Cómo cambiaría su respuesta si no tuviésemos ninguna confianza en el posible valor de la tasa de paro? ¿Y si el estudio se efectuase para estimar la tasa de paro de una pequeña población de 2.500 personas?

11.7. En una Facultad con 3.000 alumnos se está tratando de estimar el porcentaje de alumnos que estarían a favor de un nuevo método de evaluación del profesorado, y se quiere que con una confianza del 95 % la estimación no se separe más de 0,03 de su verdadero valor. ¿A cuántos alumnos habría que encuestar? ¿Cuál sería la respuesta si el número total de alumnos fuese sólo de 800?

11.8. Sea X la cantidad de nicotina, en gramos, en cada cigarro de una nueva marca de tabaco. Encuentre el número de cigarrillos que habría que analizar para que el intervalo de confianza del 99 % sea de 0,10 alrededor de la media muestral, si podemos suponer que la desviación típica es de $\sigma = 0{,}3$.

11.9. El precio medio de un artículo homogéneo en los 30 puntos de venta que se han visitado, ha sido de 16,8, con una cuasivarianza de 35. Determine el tamaño que debe tener una nueva muestra para que el intervalo de confianza aproximado del 95 % para el precio μ esté comprendido en un entorno de 0,50 alrededor de la media muestral resultante. ¿Cambiaría su respuesta si hubiese un total de 500 puntos de venta?

11.10. Un fabricante de detergentes envasa su producto en paquetes de 3 kg. La media y desviación típica de los envases viene siendo de 3,04 kg y 0,01, respectivamente. Si el peso medio puede reducirse en 0,01 kg, se ahorrarían unas 700.000 ptas. al año, para lo que se han hecho ciertos ajustes en el proceso de envasado.

a) ¿Cuántos paquetes habría que examinar para que el máximo error de estimación sea de 0.001 con confianza del 90 %?

b) Una muestra de 1.200 paquetes produjo una media de 3,024 kg con cuasidesviación típica de $s = 0{,}011$. Calcule un intervalo de confianza del 90 % para μ.
c) Estime el ahorro por año que se ha conseguido con estos ajustes.
d) Estime la proporción de paquetes que pesan ahora menos de 3 kg.

11.11. En una encuesta pre-electoral, sea p el porcentaje de votantes a favor de un cierto candidato. ¿Cuál debe ser el tamaño muestral para que el máximo error de estimación sea de:

a) ¿0,03 con confianza del 95 %?
b) ¿0,02 con confianza del 95 %?
c) ¿0,03 con confianza del 90 %?

Compare sus respuestas a las tres preguntas. Resuelva el ejercicio si se tratase de una población de 2.000 habitantes.