Capítulo 5. Muestreo. Tipos de Muestreo

5.1 Resumen Teórico de Muestreo. Tipos de Muestreo

5.1.1 Introducción

Cuando hay que hacer un estudio estadístico sobre una población, lo más habitual es que no se pueda acceder a todos los individuos que la componen; es necesario, entonces, elegir una muestra, realizar el estudio sobre ella y después intentar extrapolar los datos a toda la población en general. Gran parte de la calidad del estudio estriba en que la muestra sea cuidadosamente elegida, en el sentido de que sea lo más representativa posible, por lo tanto es muy importante el proceso que sigamos para la extracción de la muestra; a dicho proceso es a lo que se denomina muestreo.

Existen varios tipos de muestreo, pero básicamente se pueden agrupar en dos categorías: muestreo aleatorio (los individuos son elegidos al azar) y no aleatorio (los componentes de la muestra son elegidos según otros criterios, no por azar). Dentro del primero también se pueden distinguir cuatro subtipos: muestreo aleatorio simple, estratificado, sistemático y por conglomerados o áreas. Éste último y el muestreo no aleatorio no los consideraremos en este resumen.

En general, llamaremos N al tamaño de la población (número de individuos que la componen, en el caso de que sea finita) y n al de la muestra.

5.1.2 Muestreo aleatorio simple

Es aquel en el que los n individuos de la muestra son elegidos al azar. Dependiendo de la naturaleza de la población, puede ser útil numerar sus N elementos antes de elegir la muestra.

En este tipo de muestreo se presuponen dos condiciones: que cualquier elemento de la población tenga la misma probabilidad de ser seleccionado para la muestra, y que cualquier subconjunto de n unidades de la población tenga la misma probabilidad de ser elegido como parte de la muestra.

Es importante destacar un aspecto de este tipo de muestreo. En los problemas de las Pruebas de Acceso a la Universidad que siguen a continuación, siempre se considerará que la elección de la muestra se realiza *con reemplazamiento*, es decir, que un mismo elemento puede aparecer varias veces en la misma muestra. El motivo es que, desde el punto de vista teórico de la inferencia estadística, una población finita en la que los individuos son elegidos con reemplazamiento puede ser considerada como infinita, condición necesaria para asegurar alguno de los resultados que se exponen más adelante sobre la distribución de las medias muestrales.

5.1.3 Muestreo aleatorio estratificado

Es el que se utiliza cuando en la población se pueden distinguir varios colectivos (*estratos*) cuya presencia queremos reflejar en la muestra. Llamaremos $N_1, N_2, N_3,...$ al tamaño de los estratos (con $N_1 + N_2 + N_3 + ... = N$), y $n_1, n_2, n_3,...$ al número de individuos de los respectivos estratos que hay en la muestra (con $n_1 + n_2 + n_3 + ... = n$).

Según el criterio que elijamos para reflejar los estratos en la muestra, tenemos dos subtipos en este muestreo: con afijación igual (también llamada constante o simple) y con afijación proporcional.

- Muestreo aleatorio estratificado con afijación igual.

  En este caso no se toma en cuenta el número de individuos que componen cada estrato, sino que todos tienen la misma presencia en la muestra. Por ejemplo, si hay 5 estratos, de cada uno se elegirían $n/5$ individuos para la muestra, independientemente del peso que cada uno de ellos tuviera en la población. Es decir, $n_1 = n_2 = ... = n/5$.

- Muestreo aleatorio estratificado con afijación proporcional.

  Aquí sí que se toma en cuenta el tamaño de cada estrato. Lo que se pretende es que la muestra mantenga, en su composición, la misma proporción de individuos que cada estrato tenga en la población.

  Así planteado, los valores de n_1, n_2, etc. se obtendrían a partir de una sencilla proporción:

$$\frac{N_1}{N} = \frac{n_1}{n} \rightarrow n_1 = \frac{N_1 \cdot n}{N} \text{, y análogamente } n_2 = \frac{N_2 \cdot n}{N} \text{, etc.}$$

5.1.4 Muestreo aleatorio sistemático

Para realizar este tipo de muestreo es necesario ordenar a los individuos de la población, asignándoles de este modo un número ordinal a cada uno. El proceso sería el siguiente: dividimos N entre n, nos da como resultado h (llamado *coeficiente de elevación*), y después elegimos, al azar, uno de los h primeros individuos de la población, por ejemplo el que ocupa el lugar k, y a partir de ahí la muestra se iría obteniendo escogiendo individuos de h en h, es decir: $k, k + h, k + 2h, k + 3h, ..., k + (n - 1) \cdot h$.

5.1.5 Distribución de las Medias y las Proporciones Muestrales

Una vez obtenida la muestra de la población, y realizado el estudio sobre ella, llega la fase en que hay que obtener conclusiones sobre toda la población. Nosotros vamos a centrarnos en el caso de que queramos estimar la media de la población (por ejemplo, la estatura media de las niñas de 11 años), o la proporción de individuos de esa población que tienen una determinada característica (por ejemplo, la proporción de adolescentes que han viajado fuera de España). Para ello se utilizan los resultados teóricos que se exponen a continuación.

Distribución de las medias muestrales

Vamos a considerar ahora todas las muestras posibles de tamaño n que se puedan extraer de una población, y la variable aleatoria formada por sus correspondientes medias muestrales. Si llamamos μ y σ a la media y la desviación típica de la población (respectivamente), y siendo $\overline{X}$ la variable aleatoria formada por las medias muestrales, entonces podemos asegurar que:

- La media de $\overline{X}$ es μ.
- La desviación típica de $\overline{X}$ es $\dfrac{\sigma}{\sqrt{n}}$. *(Este resultado sólo es válido para poblaciones infinitas o para poblaciones finitas en las que el muestreo se ha hecho con reemplazamiento)*
- Si $X \approx N(\mu,\sigma)$, entonces $\overline{X} \approx N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.
- *(Teorema Central del Límite)* Si X no sigue una ley normal, pero $n \geq 30$, entonces se puede considerar que $\overline{X} \approx N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

El último punto nos asegura en la práctica la distribución normal de las medias muestrales en el caso de que el tamaño de las muestras sea grande ($n \geq 30$).

Distribución de las proporciones muestrales

Lo que vamos a estudiar ahora de todas las muestras posibles de tamaño n no es su media, sino la proporción de sus individuos que tienen una determinada característica. Llamaremos p al valor de esa proporción en toda la población, y $\hat{P}$ a la variable aleatoria constituida por las proporciones muestrales. Entonces también se puede demostrar que:

- La media de $\hat{P}$ es p.
- La desviación típica de $\hat{P}$ es $\sqrt{\dfrac{p \cdot q}{n}}$, donde $q = 1 - p$.
- Si $n \geq 30$, entonces se puede considerar que $\hat{P} \approx N\left(p, \sqrt{\dfrac{p \cdot q}{n}}\right)$.