

# 2. Muestreo

## 2.1. Introducción

Como se ha visto en el capítulo anterior, todas nuestras conclusiones dependen de los datos que tenemos. Por ello, uno de los aspectos básicos de todo problema estadístico es el modo de seleccionar la muestra, problema también conocido como método de muestreo.

Supongamos que queremos realizar un estudio para una determinada población. Por ejemplo, consideremos la situación en la que queremos saber las notas de ingreso en la Facultad de Farmacia que han obtenido los alumnos de primer curso. En este caso, lo más sencillo es considerar a todos los alumnos de primer curso y ver la nota de ingreso de cada uno de ellos. En este caso estamos observando el valor del experimento aleatorio en todos los individuos de la población; esto se conoce como un *censo* o un *muestreo exhaustivo*. Consideremos sin embargo la situación en la que queremos saber el porcentaje de españoles que han sufrido un esguince. En este caso, la población en estudio es muy grande para poder ser evaluada completamente. Por tanto, estamos obligados a trabajar con una parte de la población, la *muestra*.

En el caso de un censo, las conclusiones son absolutamente fiables. Sin embargo, en el segundo caso todas las conclusiones que vamos a obtener se basan en que la muestra sea representativa de la población, es decir, que los individuos seleccionados proporcionan una información que no tiene ningún sesgo respecto a la correspondiente información en el conjunto de la población. Si esto es así, entonces las conclusiones que se extraigan para la muestra serán aplicables a toda la población en estudio.

Sin embargo, no siempre podemos asegurar que la muestra sea re-

representativa. Supongamos por ejemplo que estamos estudiando la altura media de los españoles y seleccionamos una muestra de tamaño 100 eligiendo los individuos al azar en la población; podría ocurrir que estos 100 individuos fuesen jugadores de baloncesto, de forma que su altura media (muestral) sea muy superior a la media real de toda la población.

Hay varios factores que influyen en la probabilidad de que una muestra no sea representativa. El primer factor que influye es el número de individuos que se seleccionan. Supongamos por ejemplo que estamos estudiando si una moneda está trucada. Si lanzamos la moneda 10 veces y obtenemos 10 caras, podemos empezar a sospechar que la moneda está trucada, pero no estaremos completamente seguros. Sin embargo, si lanzamos la moneda 100 veces y obtenemos 100 caras, nuestra seguridad de que la moneda está trucada es mucho mayor. El número de veces que se repite el experimento, es decir, el número de individuos seleccionados para formar la muestra, se llama **tamaño de muestra**, y se denota por  $n$ .

El otro aspecto que influye en la representatividad de la muestra es la forma de seleccionar los individuos de la misma. En primer lugar, cualquier forma de seleccionar los individuos tiene que tener un cierto grado de incertidumbre (es decir, *aleatoriedad*), en el sentido de que no podemos saber *a priori* los individuos que van a ser seleccionados. Además, es conveniente que cualquier individuo de la población pueda ser seleccionado (o al revés, que un individuo concreto no tenga necesariamente que estar en la muestra), y en muchas ocasiones, que la probabilidad de que los distintos individuos de la población sean seleccionados sea la misma. Por ejemplo, si vamos a seleccionar a 10 individuos de la clase, no parece lógico que descartemos directamente a los últimos 20 alumnos por orden alfabético.

La obtención de muestras que no son representativas debido a la forma de seleccionar los individuos de la muestra ha ocurrido varias veces a lo largo de la Historia, y ha dado lugar a conclusiones erróneas. Un caso muy conocido es el siguiente: en la cuarta reelección del presidente Roosevelt se había realizado una encuesta telefónica preguntando la intención de voto. Sin embargo, esta encuesta dio lugar a unas predicciones muy diferentes de los resultados reales; en este caso concreto, esto fue debido a una mala selección de los individuos de la muestra. En efecto, en estos años (1944), el teléfono era un artículo de lujo y al

restringir la selección a los hogares con teléfono se obtuvo una muestra representativa de la población con un alto poder adquisitivo, cuya intención de voto no coincide en general con el sentir de todo el país.

Para evitar situaciones de este estilo existen distintos tipos de muestreo. En este tema veremos las características de los más comunes.

Un último aspecto a tener en cuenta: nunca podremos estar completamente seguros de que la muestra es representativa; y por ello, como veremos en los últimos capítulos de este manual, todas las conclusiones que se obtengan irán ligadas a un valor de probabilidad que nos indica la fiabilidad de dichas conclusiones. Sin embargo, si seleccionamos la muestra con criterio sí podremos reducir la probabilidad de que no lo sea.

## 2.2. Muestreo aleatorio simple

Este tipo de muestreo es el más sencillo. Se usa cuando consideramos que todos los individuos son indistinguibles respecto a la característica en estudio. En otras palabras, cuando no podemos dividir la población en grupos de individuos de forma que la respuesta de los individuos de un grupo es de esperar que sea diferente de la respuesta de los individuos de otro grupo.

Consideremos por ejemplo que estamos estudiando la nota de entrada en el Grado de Farmacia para los alumnos de primer curso. Entonces, si organizamos a los alumnos por orden alfabético, no hay ninguna razón lógica para que un alumno que se apellide Álvarez tenga una nota superior (o inferior) a un alumno que se apellida Martínez.

El muestreo aleatorio simple se basa en escoger los individuos de la población al azar. Este tipo de muestreo es el más sencillo y es el que se usa en los desarrollos matemáticos que veremos a lo largo del curso.

El muestreo aleatorio puede ser *con reemplazamiento* o *sin reemplazamiento*.

- En el muestreo con reemplazamiento, en cada realización del experimento se elige al azar entre todos los individuos de la población, incluso los que ya han sido seleccionados anteriormente. En el ejemplo anterior, si tenemos 500 alumnos y queremos una mues-

tra de tamaño 50, consiste en sortear 50 números entre 1 y 500, teniendo en cuenta que puede salir el mismo número varias veces.

- Por el contrario, en el muestreo sin reemplazamiento el individuo seleccionado se excluye para los experimentos sucesivos, de forma que el sorteo se realiza entre los individuos de la población que no han sido seleccionados. En el ejemplo anterior, sería sortear un número entre 1 y 500 para seleccionar al primer individuo de la muestra. Si se obtiene el 325, entonces para el segundo individuo de la muestra, se sortea un número entre 1 y 500 excluyendo el 325. Y así sucesivamente.

En principio parece más razonable el muestreo sin reemplazamiento, pues evita que un mismo individuo sea seleccionado varias veces. La aparición repetida de un mismo individuo básicamente se traduce en una reducción del tamaño de muestra, ya que nos proporciona una información que ya conocemos, y además otorga a ese individuo una importancia que no tiene en la población. La aparición repetida de un individuo en la muestra es especialmente probable en poblaciones cuyo tamaño sea pequeño en comparación con el tamaño de muestra. Por ejemplo, si tenemos que seleccionar 5 individuos de entre un conjunto de 6, es muy posible que si realizamos un muestreo con reemplazamiento alguno de los individuos sea seleccionado más de una vez.

Sin embargo, el muestreo sin reemplazamiento hace que las probabilidades de cada resultado varíen en cada realización, puesto que las condiciones en que se realiza el sorteo van cambiando; supongamos por ejemplo que tenemos 25 bolas, 10 blancas y 15 negras y queremos hallar la probabilidad de obtener una bola blanca en la cuarta bola elegida. En el muestreo con reemplazamiento, esta probabilidad coincide con la probabilidad de obtener bola blanca en la primera bola seleccionada, que es  $10/25$ , ya que siempre estamos en las mismas condiciones al realizar el sorteo. Sin embargo, en el muestreo sin reemplazamiento, esta probabilidad depende de las bolas que hayan sido seleccionadas anteriormente. Por ejemplo, si todas fueron blancas, ahora la probabilidad es  $7/22$ , pero si todas fueron negras, sería  $10/22$ .

¿Cómo proceder entonces? En la mayor parte de las ocasiones el tamaño de la población es muy grande en comparación con el tamaño de

muestra, con lo que es poco probable que haya repeticiones de individuos seleccionados si se realiza un muestreo con reemplazamiento; así, se realiza un muestreo sin reemplazamiento pero se funciona con los desarrollos matemáticos del muestreo con reemplazamiento, que son mucho más sencillos; esto es una aproximación, pero muy cercana a la realidad para poblaciones muy grandes, ya que las muestras que se excluyen con un muestreo sin reemplazamiento son muy pocas en comparación con el número de muestras que hay en un muestreo con reemplazamiento.

En principio el muestreo aleatorio parece una forma de actuar bastante razonable para obtener muestras representativas. Sin embargo, en poblaciones humanas hay mucha diversidad y muchas veces hay varios grupos que tienen un comportamiento similar en función de la característica; al usar el muestreo aleatorio simple tenemos el riesgo de que no se obtenga ningún individuo de alguno de los grupos y entonces la muestra perdería representatividad. Por ejemplo, las intenciones de voto en la ciudad y en el campo varían y con el muestreo aleatorio simple podemos llegar a dar más peso del que le corresponde a alguno de estos dos grupos.

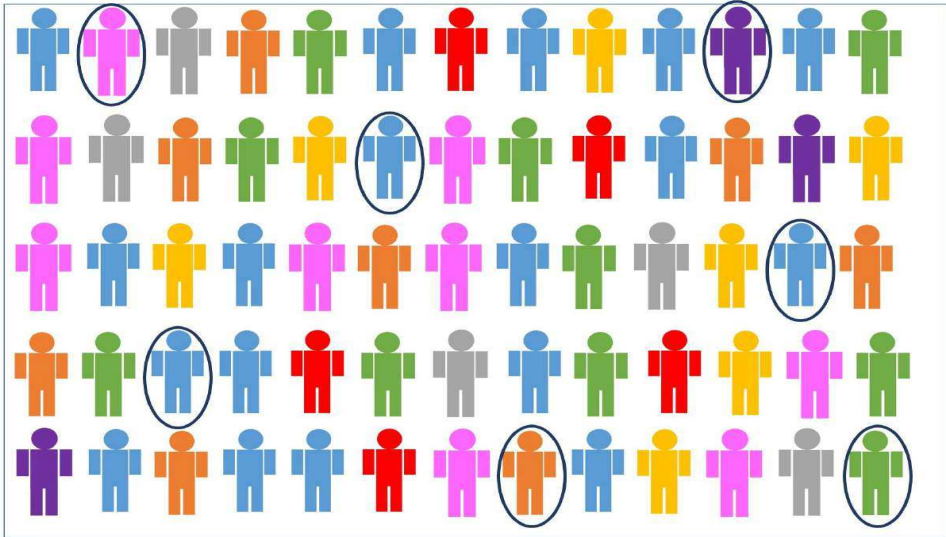
Un esquema explicativo del muestreo aleatorio se puede ver en la figura 2.1.

## 2.3. Muestreo estratificado

Como se ha visto anteriormente, el muestreo aleatorio simple tiene el problema de que hay un riesgo de que la muestra se extraiga por azar de una parte concreta de la población, o que una parte de la población esté más (o menos) representada en la muestra de lo que debería, con lo que la muestra no sería representativa de toda la población; esta situación es muy común en poblaciones de seres vivos.

### Ejemplo 12.

*Supongamos que estamos interesados en lo que gastan en ocio los empleados de una empresa. En este caso, es de esperar que el gasto en ocio dependa de los ingresos, por lo que tiene sentido dividir a los empleados en tres grupos en función de los ingresos: obreros, mandos y directivos. Nótese que el grupo de directivos tiene muy pocos individuos en comparación con el número total de empleados de la empresa. Por*



**Figura 2.1.** Esquema del muestreo aleatorio. En este caso, se trata de un muestreo aleatorio sin reposición.

*ello, si hacemos un muestreo aleatorio, no sería extraño que no apareciera ningún directivo en la muestra, con lo que los resultados no serían representativos de toda la población. De esta manera, parece lógico que dividamos la población en estos tres grupos y determinemos en primer lugar cuántos directivos, cuántos mandos y cuántos obreros se van a seleccionar para la muestra.*

Cuando estamos en situaciones como la anterior, lo más razonable es realizar un muestreo estratificado. Este muestreo consta de los siguientes pasos:

- En primer lugar, se dividen los individuos se divide la población en **estratos**. Un estrato es un grupo de individuos que se espera sea muy homogéneo respecto a la característica en estudio y que su comportamiento sea diferente del de los otros estratos para la característica en estudio; en el ejemplo anterior, tenemos tres estratos que se corresponden con los directivos, mandos y obreros. Nótese que *se espera* que los individuos de un estrato sean muy homogéneos, pero no es necesariamente así. En el ejemplo anterior,

esperamos que todos los directivos gasten más dinero en ocio que los obreros, pero tal vez encontremos un directivo cuyo gasto en ocio sea muy reducido. También esperamos que el comportamiento en ocio sea diferente si estudiamos cada grupo por separado, pero tal vez no sea así.

- Una vez dividida la población en estratos  $E_1, \dots, E_r$ , se fija el número  $n_i$  de individuos del estrato  $E_i$  que van a ser seleccionados, y a continuación se realiza la selección de estos individuos, generalmente mediante un muestreo aleatorio. De esta forma, si tenemos  $k$  estratos, el número total de individuos seleccionados es

$$n := n_1 + \dots + n_k.$$

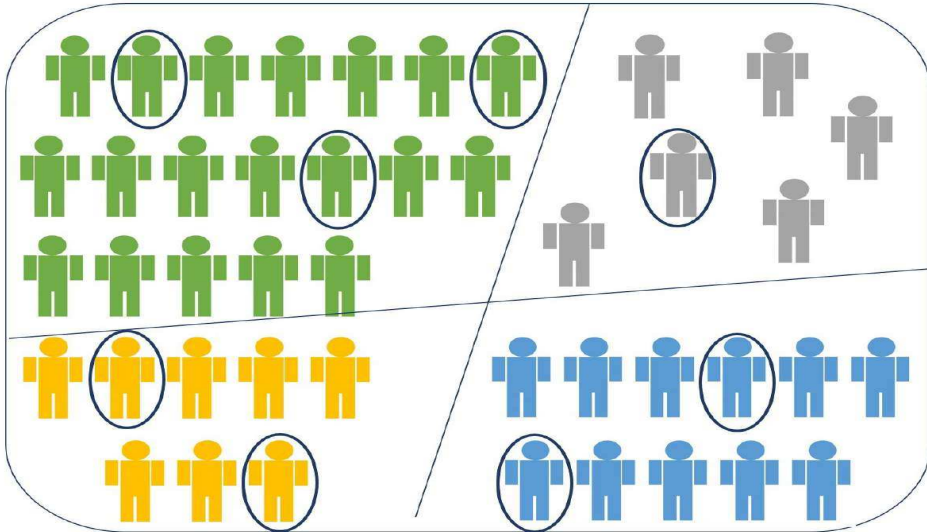
Por ejemplo, podemos decidir seleccionar  $n_1 = 2$  directivos,  $n_2 = 8$  mandos y  $n_3 = 40$  obreros para obtener una muestra de tamaño 50.

En este muestreo tenemos el problema de repartir el valor  $n$  entre los distintos estratos, decidiendo de antemano el número de individuos  $n_i$  que se van a seleccionar del estrato  $E_i$ . Parece razonable (aunque dependiendo del estudio no es necesariamente así) que los valores  $n_i$  sean proporcionales a los tamaños de los estratos en la población, lo que se conoce como *afijación proporcional*. Es decir, que si  $E_i$  representa al 10% de la población, entonces se deberían extraer de  $E_i$  el 10% de los individuos de la muestra.

Un esquema explicativo del muestreo sistemático se puede ver en la figura 2.2.

## 2.4. Muestreo por conglomerados

Consideremos el siguiente ejemplo para explicar el muestreo por conglomerados. Supongamos que estamos estudiando el tiempo de hospitalización de un paciente. Entonces, podríamos considerar un muestreo aleatorio y sortear entre todos los pacientes aquellos que van a ser incluidos en la muestra. Sin embargo, al actuar de este modo es probable que se seleccionen por ejemplo dos pacientes de Sevilla, cinco de Madrid,



**Figura 2.2.** Esquema del muestreo estratificado. La población se ha dividido en cuatro estratos, y luego se ha hecho un muestreo aleatorio en cada uno de ellos. Nótese que no se extrae el mismo número de individuos en cada estrato.

etc. Esto hace que la selección de la muestra sea un proceso costoso en tiempo y dinero. Por otra parte, no tenemos ninguna razón para suponer que los pacientes de un hospital tengan un tiempo de hospitalización más largo que el de otro hospital. Entonces, parece lógico seleccionar un hospital al azar y considerar todos los pacientes de ese hospital como la muestra.

Cuando estamos en situaciones como la anterior, lo más razonable es realizar un muestreo por conglomerados. Este muestreo consta de los siguientes pasos:

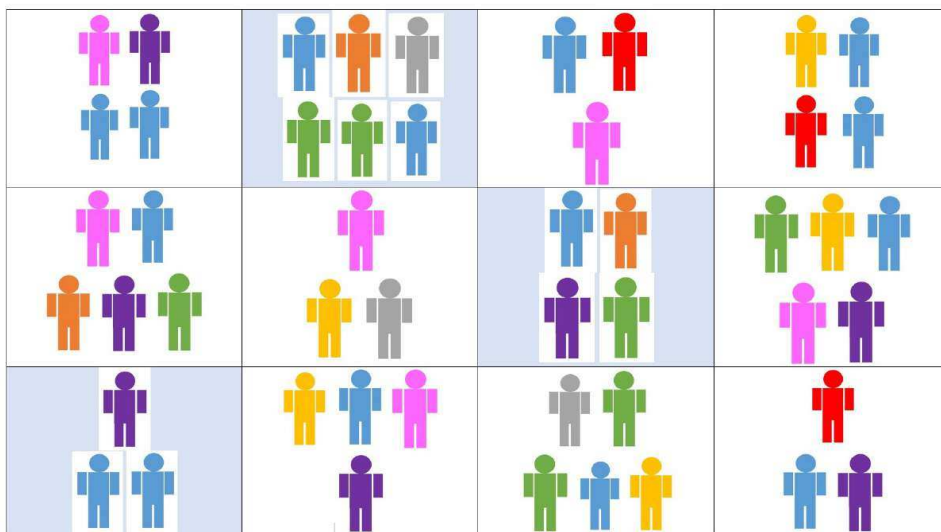
- Al contrario que en el muestreo estratificado, el muestreo por conglomerados consiste en dividir la población en subpoblaciones llamadas **conglomerados**. Mientras que en los estratos eran subgrupos muy homogéneos respecto a la característica en estudio, los conglomerados son subgrupos que son tan variables como el conjunto de la población; en otras palabras, son como poblaciones en miniatura que de alguna forma están separadas. En el ejemplo anterior, los conglomerados son los hospitales. Así, se divide la

población en  $k$  conglomerados  $C_1, \dots, C_k$ .

- Una vez dividida la población en conglomerados, se seleccionan varios de ellos al azar (por ejemplo  $C_1, C_4, C_7$ ) y se realiza un muestreo aleatorio (en ocasiones un censo) sobre los individuos que están dentro de ese conglomerado. Finalmente se juntan todos los individuos seleccionados para formar la muestra total. Este muestreo es muy popular porque permite abaratar el coste de realizar la muestra.

Nótese que, al contrario que en el muestreo estratificado, este muestreo no busca que la muestra sea más representativa que con el muestreo aleatorio, sino que busca una muestra que sea tan representativa como la del muestreo aleatorio pero obtenida de una forma más económica o rápida.

Un esquema explicativo del muestreo sistemático se puede ver en la figura 2.3.



**Figura 2.3.** Esquema del muestreo por conglomerados. La población se ha dividido en 12 conglomerados, de los que se han seleccionado tres.

## 2.5. Muestreo polietápico

El muestreo polietápico es un caso generalización del muestreo por conglomerados, en el que la unidad final de muestreo no son los conglomerados sino subdivisiones de estos. Este tipo de muestreo será interesante cuando los conglomerados contengan un elevado número de individuos y resulte aconsejable hacer una selección dentro de este conglomerado. De esta forma, en el muestreo polietápico se comienza dividiendo la población en varios conglomerados y seleccionando al azar varios de ellos. Ahora bien, estos conglomerados pueden ser todavía muy grandes, lo que aconseja dividir los conglomerados elegidos en conglomerados más pequeños y seleccionar varios de ellos. Este proceso continua hasta que obtenemos conglomerados de tamaño reducido que sean manejables, de forma que entonces ya se pueden seleccionar los individuos para la muestra final.

### Ejemplo 13.

*Supongamos que estamos estudiando el nivel de matemáticas de los niños de 12 años en la Unión Europea. En este caso, podemos considerar que no hay diferencias entre los distintos países y así podemos considerar que cada país es un conglomerado. Supongamos que en esta primera etapa se seleccionan al azar Italia y Francia. En este caso, realizar un muestreo aleatorio en estos dos países sería muy costoso.*

*Entonces, podemos considerar que no hay diferencia entre las distintas regiones de cada uno de estos países, con lo que asumimos que vamos a obtener los mismos resultados en todas las regiones. Así, podemos dividir cada país en sus regiones y seleccionar varias de ellas al azar. Por ejemplo, podemos seleccionar (por sorteo) Bretaña y Alsacia como regiones a considerar en Francia, y Lacio, Sicilia y Cerdeña como regiones a seleccionar en Italia. Esta sería la segunda etapa.*

*Ahora podemos considerar que estas regiones son todavía conglomerados muy grandes y pasar a considerar las provincias de cada región como conglomerados. Esta sería la tercera etapa.*

*Finalmente, se hace un muestreo aleatorio dentro de cada provincia. Esta sería la etapa final.*

*Como se ve, realizar este tipo de muestreo reduce en gran medida los costes, y permite obtener una información tan fiable (teóricamente) como la que se obtendría con un muestreo aleatorio simple.*

El ejemplo anterior nos sirve también para ver los riesgos de un muestreo por conglomerados: estamos asumiendo que todos los países son homogéneos, lo que no es exacto; luego, que todas las regiones dentro de un país son homogéneas, lo que tampoco es exacto; y así sucesivamente. Así, es de esperar que nuestra muestra final no sea tan representativa como la que se obtendría con un muestreo aleatorio. Sin embargo, esperamos que sea casi tan representativa, con la ventaja de que se ha obtenido mucho más fácilmente.

## 2.6. Muestreo sistemático

Este muestreo es muy útil cuando hay una ordenación natural de los individuos de la población (por ejemplo porque estamos evaluando una variable a lo largo del tiempo) y esta ordenación podría influir en el resultado del experimento.

Por ejemplo, supongamos que estamos estudiando la temperatura en una ciudad y que nuestra población son los minutos a lo largo del día; entonces, tenemos que seleccionar los minutos en los que se va a medir la temperatura; por otra parte, la temperatura varía a lo largo del día según el momento en que se mida.

En situaciones como la anterior, si hacemos un muestreo aleatorio nos arriesgamos a que no aparezcan medidas de la temperatura dentro de una franja de tiempo o muchas mediciones dentro de una franja muy reducida. Y esto haría que nuestra muestra no fuese representativa. Para evitar este inconveniente se recurre al muestreo sistemático. Este muestreo consiste en dividir el tamaño total de la población entre el tamaño de muestra; por ejemplo, si en el caso anterior queremos obtener una muestra de tamaño 24, entonces dividimos a toda la población ordenada en 24 grupos, que en nuestro caso serán las horas del día. Nótese que todos los grupos tienen el mismo número de individuos, en nuestro caso 60, que se corresponden con los 60 minutos de cada hora.

A continuación se escoge un individuo al azar en el primer grupo, que en nuestro caso es seleccionar al azar un minuto de la primera hora. Supongamos que ha sido seleccionado el minuto 7. Es decir, tenemos el individuo correspondiente a las 00:07 horas.

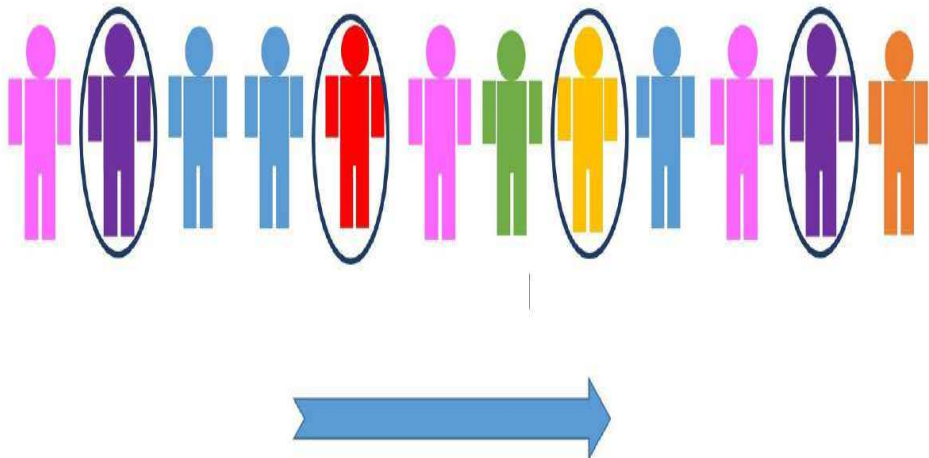
Finalmente, los individuos seleccionados para la muestra son los in-

dividuos en posición 7 en todos los grupos. Entonces la muestra final sería:

00 : 07, 01 : 07, 02 : 07, ..., 23 : 07.

La razón de escoger siempre el mismo individuo en cada grupo es la siguiente: Debe tenerse en cuenta aquí que si se escogiese un individuo al azar en cada franja, nos arriesgamos a que aparezcan los individuos 59 (para el primer grupo), 1 (para el segundo grupo), 57(para el tercer grupo), ... lo que nos conduciría a que hay tramos en los que no tenemos información. En este caso tenemos dos medidas muy similares y luego no hay información durante casi dos horas. Y precisamente el muestreo sistemático intenta evitar esta situación y repartir los individuos seleccionados para la muestra de forma uniforme según el orden.

Un esquema explicativo del muestreo sistemático se puede ver en la figura 2.4.



**Figura 2.4.** Esquema del muestreo sistemático. La flecha represente indica la ordenación de los individuos. Se tiene una muestra de tamaño 4 en una población de tamaño 12, luego hay cuatro grupos de tamaño 3, en cada uno de los cuales se ha escogido el segundo individuo.

## 2.7. Muestreos específicos para ciencias de la salud

Además de los tipos de muestreo vistos anteriormente, existen otras formas de realizar el experimento más orientadas a problemas médicos. En estos muestreos se estudia si poseer una determinada característica es un factor de riesgo para desarrollar una enfermedad. Por ejemplo, una situación típica es ver si fumar es un factor de riesgo para desarrollar cáncer de pulmón.

### 2.7.1. Muestreo por cohortes

Veamos la situación más sencilla de este tipo de muestreo. La idea es hacer una comparación de la frecuencia de enfermedad (o de un determinado desenlace) entre dos poblaciones, una de las cuales está expuesta a un determinado factor de exposición o *factor de riesgo* al que no está expuesta la otra. Es decir, se selecciona una muestra de individuos *sanos* y estos individuos se dividen en dos submuestras, según tengan o no el factor de riesgo. A continuación, se observa para cada una de las submuestras la frecuencia con que la enfermedad aparece al cabo de un tiempo. La finalidad de este tipo de muestreo es observar si el factor de riesgo hace más probable (por frecuente) la aparición de la enfermedad.

Los pasos para realizar un estudio de cohortes son los siguientes:

- Seleccionar de entre los individuos sanos de la población dos muestras, una de entre los individuos que tienen el factor de riesgo y otra entre los individuos que no tienen el factor de riesgo.
- Medir las variables de resultado, es decir, la presencia o ausencia de enfermedad en los grupos al cabo de un tiempo.

A partir de este esquema básico, el muestreo por cohortes puede ampliarse para obtener mucha más información. Así, es posible seleccionar no solo dos muestras (ausencia o presencia del factor de riesgo), sino que permite aplicar una gradación en la presencia del factor de riesgo. Así, podemos considerar varias muestras (cohortes) para estudiar con una mayor profundidad la influencia del factor de riesgo. En la situación con

la que se inició esta sección, en la que queríamos ver si fumar influía en el desarrollo posterior del cáncer de pulmón, podríamos considerar varias cohortes, como «no fuma», «fuma poco», «fuma moderadamente», «fuma mucho», y comparar los desarrollos posteriores de la enfermedad.

Otra ventaja de este muestreo es que permite hacer un seguimiento a lo largo del tiempo, lo que proporciona mucha más información. En el caso anterior, es posible que tras 10 años, la proporción de enfermos sea la misma en las dos poblaciones, pero un muestreo por cohortes permite ver si la enfermedad se desarrolla antes en la muestra de sujetos con factor de riesgo, o si la enfermedad se desarrolla de forma más virulenta para los individuos con factor de riesgo.

Además, este tipo de muestreo tiene la ventaja de que permite estudiar factores de exposición raros, que de otra forma no aparecerían en el muestreo por ser muy poco frecuentes.

Sin embargo, este muestreo tiene una serie de desventajas. En primer lugar, es un muestreo muy costoso, en el sentido de que obliga a hacer un seguimiento durante largos períodos de tiempo a los individuos de la muestra. Así, esto obliga a hacer revisiones periódicas de todos los individuos. Por otra parte, tiene el problema de que es posible que a lo largo del tiempo se pierdan individuos de las muestras, por ejemplo por fallecimiento por otras causas ajenas a la enfermedad, o porque el individuo abandone el estudio y no acuda a las revisiones periódicas.

## 2.7.2. Muestreo caso-control

En este tipo de muestreo tenemos la misma situación que en el muestreo de cohortes y el objetivo es el mismo. En el muestreo caso-control se seleccionan dos muestras, una de las cuales es de individuos con la enfermedad (casos) y la otra es una muestra de individuos sanos (control). Una vez seleccionados los individuos en cada grupo, se investiga si estuvieron expuestos o no al factor de riesgo; a continuación se compara la proporción de individuos expuesto al factor de riesgo en la muestra de enfermos con la proporción de individuos expuestos al factor de riesgo en la muestra de individuos que no han desarrollado la enfermedad.

Los pasos para realizar un muestreo caso-control son:

- Seleccionar una muestra de población con la enfermedad. A los individuos de esta muestra se les llama *casos*.

- Seleccionar una muestra de la población que esté libre de la enfermedad, que será el grupo *control*.
- Medir la proporción de individuos expuestos al factor de riesgo en las dos muestras.

El muestreo caso control tiene la ventaja de que es un tipo de muestreo muy rápido de efectuar, y proporciona resultados inmediatos. Además permite estudiar con unos mismos datos muchos factores de riesgo simultáneamente, pues basta estudiar si cada uno de ellos está presente o no en los individuos de la muestra de casos y de la muestra de control.

Sin embargo, tiene el inconveniente de que la información que proporciona una información bastante pobre, pues solo permite comparar las proporciones de individuos con riesgo en cada una de las muestras. Además, si el factor de riesgo es muy poco frecuente, puede ocurrir que no aparezcan individuos con ese factor de riesgo en ninguna de las dos muestras.

Finalmente, el muestreo caso-control tiene el inconveniente de que no permite concluir causalidad. Para explicar este proceso, consideremos la siguiente situación: «¿Las personas altas tienen tendencia a pesar mucho?» En este caso, tenemos como factor de riesgo «ser alto» y como característica «pesar mucho». Por lo tanto, escogemos dos muestras de individuos, unos que pesan mucho y otros que no, y observamos si los individuos de cada muestra son altos o no. Supongamos que las conclusiones son que efectivamente ser alto influye en pesar mucho. Ahora el muestreo caso-control permite intercambiar los papeles de factor de riesgo y característica. Es decir, podemos considerar los mismos individuos y considerar como característica de estudio ser alto y como factor de riesgo pesar mucho. Y repitiendo el estudio, podríamos sacar la conclusión de que pesar mucho influye en ser alto. En definitiva, no sabemos qué factor provoca qué característica. Esto no ocurre con el muestreo por cohortes, en el que se parte de individuos sin la enfermedad, lo que determina claramente que es el factor de riesgo el que influye en su desarrollo.