11 LAS MUESTRAS ESTADÍSTICAS

Página 266

1. Una ganadería tiene 3 000 vacas. Se quiere extraer una muestra de 120. Explica cómo se obtiene la muestra:

a) Mediante muestreo aleatorio simple.

b) Mediante muestreo aleatorio sistemático.

a) — Se numeran las vacas del 1 al 3 000.

— Se sortean 120 números de entre los 3 000.

— La muestra estará formada por las 120 vacas a las que correspondan los números obtenidos.

b) Coeficiente de elevación: $h = \frac{3000}{120} = 25$

— Se sortea un número del 1 al 25. Supongamos que sale el 9.

— Las vacas seleccionadas para la muestra serían las que correspondieran a los números 9, 34, 59, 84, 109, …, 2 984.

Página 267

2. Una ganadería tiene 2 000 vacas. Son de distintas razas: 853 de A, 512 de B, 321 de C, 204 de D y 110 de E.

Queremos extraer una muestra de 120:

a) ¿Cuántas hay que elegir de cada raza para que el muestreo sea estratificado con reparto proporcional?

b) ¿Cómo ha de ser la elección dentro de cada estrato?

a) Llamamos n_1 al número de vacas que debemos elegir de la raza A, n_2 al de raza B, n_3 al de C, n_4 al de D y n_5 al de E.

Ha de cumplirse que:

$$\frac{120}{2000} = \frac{n_1}{853} = \frac{n_2}{512} = \frac{n_3}{321} = \frac{n_4}{204} = \frac{n_5}{110}$$

Así, obtenemos:

$$n_1 = 51{,}18 \quad n_2 = 30{,}72 \quad n_3 = 19{,}26 \quad n_4 = 12{,}24 \quad n_5 = 6{,}6$$

La parte entera de estos números suma:

51 + 30 + 19 + 12 + 6 = 118. Faltan 2 para llegar a 120.

Por tanto, debemos elegir:

51 vacas de raza A, 31 vacas de B, 19 de C, 12 de D y 7 de E.

b) Dentro de cada estrato, la elección ha de ser aleatoria.

Página 268

1. Obtén aleatoriamente cuatro números enteros comprendidos entre 1 y 95.

Por ejemplo:

Ran# [0.226] × 95 + 1 = [22.47] $\rightarrow$ 22

Ran# [0.048] × 95 + 1 = [5.56] $\rightarrow$ 5

Ran# [0.277] × 95 + 1 = [27.315] $\rightarrow$ 27

Ran# [0.842] × 95 + 1 = [80.99] $\rightarrow$ 80

O mejor:

95 × ×

Ran# = [21.47] $\rightarrow$ 22

Ran# = [4.56] $\rightarrow$ 5

Hemos obtenido los números 22, 5, 27 y 80.

2. Obtén cinco números enteros elegidos aleatoriamente entre 1 y 800.

Por ejemplo:

Ran# [0.104] × 800 + 1 = [84.2] $\rightarrow$ 84

Ran# [0.098] × 800 + 1 = [79.4] $\rightarrow$ 79

Ran# [0.835] × 800 + 1 = [669] $\rightarrow$ 669

Ran# [0.449] × 800 + 1 = [360.2] $\rightarrow$ 360

Ran# [0.622] × 800 + 1 = [498.6] $\rightarrow$ 498

O mejor:

800 × ×

Ran# = [83.2] $\rightarrow$ 84

Ran# = [78.4] $\rightarrow$ 79

Hemos obtenido los números 84, 79, 669, 360 y 498.

Página 269

3. De una población de *N* = 856 elementos, deseamos extraer una muestra de tamaño *n* = 10. Mediante el uso de números aleatorios, designa cuáles son los 10 individuos que componen la muestra.

Para multiplicar por 856 los números que aparezcan en pantalla, introducimos:

856 ⊠ ⊠ (factor constante)

Ahora recurrimos a los números aleatorios. Por ejemplo, podemos obtener:

Ran#	0.836	=	714.76	→ 715
Ran#	0.419	=	358.664	→ 359
Ran#	0.554	=	474.224	→ 475
Ran#	0.567	=	485.352	→ 486
Ran#	0.530	=	453.68	→ 454
Ran#	0.057	=	48.792	→ 49
Ran#	0.993	=	850.008	→ 851
Ran#	0.396	=	338.976	→ 339
Ran#	0.013	=	11.128	→ 12
Ran#	0.636	=	544.416	→ 545

Los individuos elegidos para la muestra serían los correspondientes a los números 715, 359, 475, 486, 454, 49, 851, 339, 12 y 545.

4. De una población de 543 individuos, queremos extraer una muestra de tamaño 40 mediante números aleatorios. Obtén los cinco primeros elementos de dicha muestra.

Para multipicar por 543 los números que aparezcan en pantalla, introducimos:

543 ⊠ ⊠ (factor constante)

Ahora recurrimos a los números aleatorios. Por ejemplo, podemos obtener:

Ran#	0.237	=	128.691	→ 129
Ran#	0.071	=	38.553	→ 39
Ran#	0.614	=	333.402	→ 334
Ran#	0.497	=	269.871	→ 270
Ran#	0.475	=	257.925	→ 258

Los cinco primeros elementos de la muestra serían los correspondientes a los números 129, 39, 334, 270 y 258.

Página 272

EJERCICIOS Y PROBLEMAS PROPUESTOS

PARA RESOLVER

1 **En cada uno de los siguientes casos, di si el colectivo es población o es muestra:**

a) En una floristería, añaden al riego de todas las macetas unas gotas de un cierto producto para probar su eficacia.

b) En un gran invernadero, se seleccionan 200 plantas que serán regadas con unas gotas de un producto para analizar su eficacia.

a) Es **población**, porque riega con el producto todas las macetas.

b) Es **muestra**, porque no riega con el producto todas las macetas, sino una parte de ellas.

2 **Un fabricante de elásticos quiere estudiar su resistencia a la rotura. Para ello, los estira hasta que se rompen y anota el grado de estiramiento que alcanzan sin romperse.**

¿Puede realizar dicho estiramiento sobre la población o es imprescindible realizarlo sobre una muestra? ¿Por qué?

Es imprescindible hacerlo sobre una muestra, porque interesa romper la menor cantidad de elásticos posible.

3 **Solo uno de los siguientes procedimientos nos permite obtener una muestra representativa. Di cuál es y, en los otros, estudia el sentido del sesgo y su importancia:**

a) Para estudiar las frecuencias relativas de las letras, se toman al azar 20 libros de la biblioteca de un centro escolar y se cuenta las veces que aparece cada letra en la página 20 de los libros seleccionados.

b) Para conocer la opinión de sus clientes sobre el servicio ofrecido por unos grandes almacenes de cierta ciudad, se selecciona al azar, entre los que poseen tarjeta de compra, a 100 personas entre las que han gastado menos de 1 000 € el último año, otras 100 entre las que han gastado entre 1 000 € y 5 000 €, y 100 más entre las que han gastado más de 5 000 €.

c) Para calcular el número medio de personas que están adscritas a cada cartilla en un Centro de Salud de la Seguridad Social, los médicos toman nota de todas las cartillas de las personas que acuden a las consultas durante un mes.

a) Es una muestra representativa.

b) No es representativa, porque hay mucha más gente en un intervalo (por ejemplo, entre 1 000 € y 5 000 €) que en otro (más de 5 000 €), y hemos tomado el mismo número de representantes. Además, hay otra mucha gente sin tarjeta que no se ha tomado en cuenta.

c) No es representativa, ya que lo que más se va a ver son las cartillas que corresponden a familias numerosas. Está claro que cuanta más gente tenga esa cartilla más fácil es que ese mes se tome nota de ella.

4 **De un colectivo de 500 personas, elige una muestra de 20 mediante:**

a) Un muestreo aleatorio sistemático.

b) Un muestreo aleatorio simple.

Utiliza la tecla [Ran#] **de la calculadora.**

Para los dos casos, numeramos a las personas del 1 al 500.

a) $h = \dfrac{500}{20} = 25$

Origen: 25 [×] [Ran#] [+] 1 [=] [14.075] (por ejemplo)

Deberemos elegir las personas cuyos números sean:

14, 39, 64, 89, 114, 139, 164, 189, 214, 239, 264, 289, 314, 339, 364, 389, 414, 439, 464, 489.

b) Con la tecla [Ran#] de la calculadora, hacemos: 500 [×] [×] [Ran#] [=] [Ran#] [=] [Ran#] [=]… hasta obtener 20 resultados distintos

5 **En un conjunto de 1 000 conductores hay:**

— 50 taxistas.

— 75 camioneros.

— 25 conductores de autobús.

El resto son conductores de vehículos corrientes y se reparten así:

— 250 con más de 20 años de experiencia.

— 425 con una experiencia de entre 5 y 20 años.

— 175 con una experiencia de 0 a 5 años.

Para confeccionar una muestra de 40 individuos mediante muestreo aleatorio estratificado proporcional, ¿cuántos hay que seleccionar de cada uno de los seis estratos?

Llamamos n_1 al número de taxistas que tendríamos que seleccionar, n_2 al número de camioneros, n_3 al número de conductores de autobuses, n_4 al número de conductores con más de 20 años de experiencia, n_5 al de conductores con una experiencia de entre 5 y 20 años y n_6 al de conductores con una experiencia de 0 a 5 años. Entonces:

$$\frac{n_1}{50} = \frac{n_2}{75} = \frac{n_3}{25} = \frac{n_4}{250} = \frac{n_5}{425} = \frac{n_6}{175} = \frac{400}{1\,000}$$

Así, debemos elegir:

n_1 = 2 taxistas

n_2 = 3 camioneros

n_3 = 1 conductor de autobús

n_4 = 10 conductores con más de 20 años de experiencia

n_5 = 17 con experiencia entre 5 y 20 años

n_6 = 7 con experiencia entre 0 y 5 años

6 **En cierta provincia hay cuatro comarcas, C_1, C_2, C_3 y C_4, con un total de 1 500 000 personas censadas. De ellas, 300 000 residen en C_1, 450 000 en C_2 y 550 000 en C_3.**

Se quiere realizar un estudio sobre las costumbres alimenticias en esa provincia basado en una muestra de 3 000 personas.

a) ¿Qué tipo de muestreo deberíamos realizar si queremos que en la muestra resultante haya representación de todas las comarcas?

b) ¿Qué número de personas habría que seleccionar en cada comarca, atendiendo a razones de proporcionalidad?

c) ¿Cómo seleccionarías las personas en cada comarca?

Justifica las respuestas.

a) Deberíamos realizar un muestreo aleatorio estratificado.

b) El número de personas que residen en C_4 es:

$$1\,500\,000 - (300\,000 + 450\,000 + 550\,000) = 200\,000$$

Llamamos n_1, n_2, n_3 y n_4 al número de personas que tendríamos que seleccionar en cada comarca (C_1, C_2, C_3 y C_4, respectivamente). Entonces:

$$\frac{n_1}{300\,000} = \frac{n_2}{450\,000} = \frac{n_3}{450\,000} = \frac{n_4}{200\,000} = \frac{3\,000}{1\,500\,000}$$

Por tanto, debemos elegir:

n_1 = 600 personas de C_1

n_2 = 900 personas de C_2

n_3 = 1 100 personas de C_3

n_4 = 400 personas de C_4

c) Dentro de cada comarca, podríamos seleccionarlos mediante un muestreo aleatorio simple, o mediante un muestreo sistemático.

7 **En un centro de enseñanza con 981 alumnos y alumnas, se va a hacer un sondeo sobre tendencias políticas.**

Se va a escoger una muestra de 84 estudiantes. En el centro hay 5 cursos (1.°, 2.°, 3.°, 4.° y 5.°) con un número de alumnos y alumnas en cada uno de ellos de 345, 234, 190, 140 y 72.

¿Cuántos alumnos deberemos escoger de cada curso si deseamos que el muestreo sea estratificado con reparto proporcional?

$$\frac{84}{981} = \frac{a}{345} = \frac{b}{234} = \frac{c}{190} = \frac{d}{140} = \frac{e}{72}$$

Así: $a = 30$, $b = 20$, $c = 16$, $d = 12$, $e = 6$

Página 273

8 **Queremos seleccionar una muestra de 50 alumnos de 2.° de Bachillerato. En cada uno de los siguientes casos debes decidir si el muestreo debe ser aleatorio simple o estratificado por sexos (chicos-chicas) para estudiar las variables indicadas:**

a) Estatura.

b) Tiempo que emplean los alumnos en ir de su casa al instituto.

c) Agudeza visual (porcentaje de alumnado con gafas).

d) Incidencia de caries dental.

e) Práctica de fútbol.

f) Lectura de algún periódico.

g) Número de hermanos.

a) En la estatura de chicos y chicas de esa edad suele haber diferencias significativas. El muestreo debe ser estratificado en este caso.

b) Simple.

c) Simple.

d) Simple.

e) Estratificado. Hay una gran diferencia entre el porcentaje de chicos y chicas que juegan al fútbol.

f) Simple.

g) Simple.

9 **Una biblioteca pública está organizada en cinco secciones con los números de libros que se indican en esta tabla:**

SECCIÓN 1	SECCIÓN 2	SECCIÓN 3	SECCIÓN 4	SECCIÓN 5
500	860	1 200	700	740

Se quiere seleccionar una muestra del 5% de los libros mediante muestreo estratificado aleatorio, considerando como estratos las secciones.

Determina el número de libros que habría que seleccionar en cada sección si:

a) Consideramos afijación igual.

b) Consideramos afijación proporcional.

Tenemos un total de 4 000 libros.

a) El 5% de 4 000 son 200 libros. Como tenemos cinco secciones, debemos elegir 200 : 5 = 40 libros de cada sección.

b) Como queremos una muestra del 5%, elegimos aleatoriamente un 5% de libros de cada sección. Así, debemos escoger:

— 25 libros de la sección 1.

— 43 libros de la sección 2.

— 60 libros de la sección 3.

— 35 libros de la sección 4.

— 37 libros de la sección 5.

Elegimos un total de 200 libros.

PARA PROFUNDIZAR

10 **Si cuentas el número de personas y el número de perros que viven en tu portal y todos los compañeros y compañeras hacéis lo mismo, obtendréis una muestra con la que podréis estimar el número de perros que hay en vuestra población.**

a) ¿Cómo es de fiable esta estimación?

b) ¿Es aleatoria la muestra que has utilizado?

c) ¿Se te ocurre un procedimiento mejor para seleccionar la muestra?

a) Es poco fiable.

b) La muestra no es aleatoria porque no la hemos elegido al azar entre los habitantes de la ciudad que se quiere estudiar.

Si en ese portal hay muchas viviendas, pueden representar, en el mejor de los casos, a las familias de ese barrio (céntrico o periférico, con ciertas características socioeconómicas, culturales…), pero no a los demás barrios de la población.

c) Utilizar una muestra de viviendas elegidas al azar entre las de esa población.

11 **Para hacer un estudio sobre los hábitos ecológicos de las familias de una ciudad, se han seleccionado, por sorteo, las direcciones, calle y número que serán visitadas. Si en un portal vive más de una familia, se sorteará entre ellas la que será seleccionada. ¿Obtendremos con este procedimiento una muestra aleatoria?**

Las familias que viven en viviendas unifamiliares tienen mayor probabilidad de ser elegidas.

12 **La validez de la información que nos proporciona una encuesta depende, en gran medida, de la cuidadosa elaboración del cuestionario. ¿Qué defectos adviertes en las siguientes preguntas?:**

a) ¿Cuántos libros leíste el año pasado?

b) ¿Cuánto tiempo dedicas al deporte?

Poco ☐ **Mediano** ☐ **Mucho** ☐ **Muchísimo** ☐

c) ¿Qué opinión tienes del alcalde?

Muy buena ☐ **Buena** ☐ **Indiferente** ☐

d) ¿Qué opinas sobre el cambio climático?

a) Salvo que se vayan apuntando los libros leídos, que casi nadie hace, la respuesta que se dé es aproximada.

b) Las opciones que se dan de respuesta son muy subjetivas. Dos personas que dediquen el mismo tiempo, pueden dar respuestas distintas.

c) Es una pregunta que, dependiendo de la época en que se haga, de la ideología del encuestado, etc., puede variar mucho.

d) Las respuestas serán tan distintas que no se pueden tabular ni estudiar posteriormente.

Página 273

AUTOEVALUACIÓN

1. **Un guionista de cine tiene dudas sobre cómo resolver el final de su próxima película de intriga. Decide preguntar sobre la viabilidad de dos posibles finales a futuros espectadores. Di dos razones, al menos, por las que se justifique que el sondeo debe hacerlo sobre una muestra (y no consultar a toda la población).**

- La población (los futuros espectadores de la película), además de ser muy numerosa, aún no está bien definida.
- Los individuos participantes en la muestra “se estropean”: al conocer de antemano el posible final de la película, dejarán de disfrutar plenamente la emoción de la intriga en la película finalizada.

2. **Selecciona mediante muestreo aleatorio sistemático una muestra de 14 individuos de un total de 584. Utiliza para ello la tecla Ran# de tu calculadora.**

El coeficiente de elevación es $h = \dfrac{584}{14} = 41{,}71$.

Tomamos $h = 42$.

Sorteamos quién será el primero, del 1 al 42:

Ran# $\boxed{0.667}$ × 42 = $\boxed{28.014}$

El primer elemento que obtenemos en este caso será el individuo 29.

Por tanto, elegiremos a estos individuos:

29, 71, 113, 155, 197, 239, 281, 323, 365, 407, 449, 491, 533 y 575.

3. **En un centro de enseñanza con 1 324 alumnos y alumnas, se va a hacer un sondeo sobre afición a la lectura. Se va a escoger una muestra de 80 estudiantes. En el centro hay 6 cursos: 1.°, 2.°, 3.°, 4.°, 5.° y 6.°, con 411, 338, 175, 153, 130 y 117 alumnos, respectivamente.**

a) ¿Cuántos hay que escoger de cada curso si se desea que el muestreo sea estratificado con reparto proporcional?

b) Dentro de cada estrato, ¿cómo se seleccionan los individuos que forman parte de la muestra?

a) $\dfrac{80}{1\,324} = \dfrac{a}{411} = \dfrac{b}{338} = \dfrac{c}{175} = \dfrac{d}{153} = \dfrac{e}{130} = \dfrac{f}{117}$

80 ÷ 1 324 = $\boxed{0.060422296}$ × × 411 = $\boxed{24.83}$ ← a

338 = $\boxed{20.42}$ ← b

175 = $\boxed{10.57}$ ← c

153 = $\boxed{9.24}$ ← d

130 = $\boxed{7.85}$ ← e

117 = $\boxed{7.06}$ ← f

A la suma de las partes enteras (24 + 20 + 10 + 9 + 7 + 7 = 77) le faltan 3 unidades para llegar a 80. Añadimos una unidad a las tres que tienen mayor parte decimal, a, c y e. Por tanto:

$a = 25$, $b = 20$, $c = 11$, $d = 9$, $e = 8$, $f = 7$

b) En cada estrato, los correspondientes elementos de la muestra se eligen aleatoriamente.