
Muestreo

Selección de actividades resueltas

PID_00279896

Àngel J. Gil Estallo

Jose Fco. Martínez Bosca

Arnau Mir Torres

Lluís M. Pla Aragonés

Àngel A. Juan Pérez (editor)

Àngel J. Gil Estallo**Jose Fco. Martínez Boscá****Arnau Mir Torres****Lluís M. Pla Aragonés****Àngel A. Juan Pérez (editor)**

La revisión de este recurso de aprendizaje UOC ha sido coordinada por la profesora: Laura Calvet Liñán

Segunda edición: febrero 2021

© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)

Av. Tibidabo, 39-43, 08035 Barcelona

Autoría: Àngel J. Gil Estallo, Jose Fco. Martínez Boscá, Arnau Mir Torres, Lluís M. Pla Aragonés

Producción: FUOC

Todos los derechos reservados

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita del titular de los derechos.

Índex

Introducción	5
Mapa conceptual	6
Actividades.....	7

Introducción

Imaginemos que hemos diseñado un nuevo programa de contabilidad y queremos comprobar su grado de aceptación entre las empresas de Cataluña. ¿Cómo podemos hacer dicha comprobación? Una manera simple de responder a esta pregunta es simplemente escoger aleatoriamente un conjunto de empresas y hacer la encuesta correspondiente. Ahora bien, si la elección del conjunto anterior no se hace de forma “representativa”, el resultado de la encuesta puede no ser fiable.

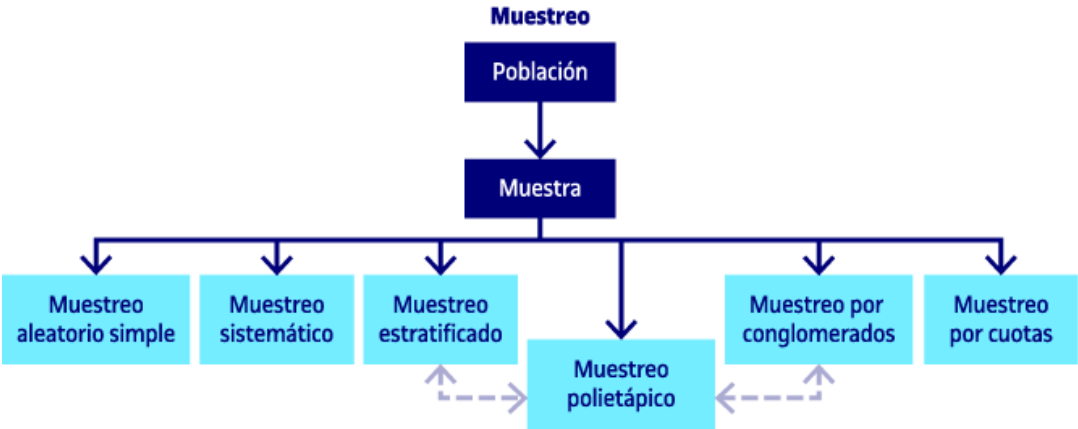
Tenemos que hacer, por tanto, un estudio de cómo escoger una muestra representativa de todas las empresas de Cataluña de cara a obtener un resultado fiable de la encuesta sobre el programa de contabilidad.

Las técnicas de muestreo nos dicen cómo hacer un estudio como el anterior para que la muestra que escojamos sea lo menos sesgada posible. Cuando decimos sesgada, queremos decir que la muestra tenga más empresas de un determinado tipo que de otro. Por ejemplo, si en nuestra muestra aparecen un 60 % de empresas que se dedican al sector de las comunicaciones y un 20 % de empresas que se dedican al sector de asesoramiento jurídico, nuestra muestra será sesgada, ya que tendrá el triple de empresas que se dedican al sector de las comunicaciones respecto de las empresas que se dedican al asesoramiento jurídico.

Para poder hacer la lista de empresas de nuestra muestra, en primer lugar necesitamos estudiar con qué información contamos. O sea, ¿tenemos un listado de todas las empresas de Cataluña?, ¿dicho listado contiene información sobre el tipo de empresa?, etc. Dependiendo de la información que tengamos, podemos realizar un tipo de muestreo u otro.

A continuación presentamos unos problemas resueltos sobre muestreo que nos pueden ayudar a resolver cuestiones como la que planteamos anteriormente.

Mapa conceptual



Actividades

Actividad 1: nivel de satisfacción de un servidor de correo

Técnicas de muestreo. Muestreo probabilístico

Queremos hacer un estudio de la satisfacción que tienen los alumnos de la UOC con su servidor de correo. Para realizar este estudio, se decide pasar una encuesta a quinientos alumnos. Decid qué tipo de muestreo haríais en cada uno de los casos siguientes explicando cómo haríais el muestreo y detallando cómo determinaríais la muestra:

- Si disponemos de un listado de todos los alumnos de la UOC.
- Si queremos considerar las comarcas de Cataluña, Islas Baleares, Comunidad Valenciana y otras como zonas geográficas. En este caso, también disponemos de un listado de los alumnos en el que está indicada la comunidad autónoma a la que pertenece cada alumno.
- Si tenemos la tabla siguiente en la que se indica el porcentaje de alumnos por franjas de edad (entre 20 y 40 años, entre 40 y 60 años y > 60 años) y por tipo de servidor.

Servidor correo/edad	20-40 años	40-60 años	> 60 años
Gmail	20 %	5 %	1 %
Hotmail	20 %	5 %	2 %
Yahoo	20 %	3 %	1 %
Otros	20 %	2 %	1 %

Solución

- En este caso tenemos que hacer un muestreo aleatorio simple. Vamos a suponer que tenemos quince mil alumnos. Para simular en R una muestra de quinientos alumnos podemos usar directamente las instrucciones:

```
test1<-sample(1:15000,500,replace=TRUE)
```

con lo que obtenemos quinientos valores (cada vez que se ejecute se obtiene una muestra diferente), de los que mostramos los diez primeros:

```
test1[1:10]
## [1] 8821 8053 10924 14387 3738 7106 11776 2376 11187 495
```

- En este caso debemos realizar un muestreo por conglomerados usando proporcionalmente los alumnos pertenecientes a cada comunidad autónoma. Por ejemplo, imaginemos que el 60 % de los alumnos de la UOC son de Cataluña, el 15 % de la Comunidad Valenciana, el 15 % de Baleares y el 10 % restante pertenece a otras comunidades autónomas. Entonces nuestra muestra tendrá $0.6 \cdot 500 = 300$ alumnos de Cataluña, $0.15 \cdot 500 = 75$ alum-

nos de la Comunidad Valenciana, 75 alumnos de Baleares y 50 alumnos de otras comunidades. Después elegimos aleatoriamente unas comarcas de Cataluña, Valencia, Baleares y otras comunidades autónomas. Supongamos, para fijar ideas, que elegimos diez comarcas de Cataluña, diez comarcas de la Comunidad Valenciana, diez comarcas de Baleares y diez comarcas de otras comunidades autónomas. Seguidamente, hemos de elegir treinta estudiantes de cada comarca de Cataluña, siete estudiantes de cinco comarcas de la Comunidad Valenciana y ocho estudiantes de las otras cinco comarcas, lo mismo para las comarcas de Baleares y cinco estudiantes para las comarcas de otras comunidades autónomas; todas las elecciones se hacen de forma aleatoria. Así, habremos formado una muestra de quinientos estudiantes.

c) En este caso debemos realizar un muestreo por cuotas. Calculamos cuántos estudiantes hemos de tener en nuestra muestra por franjas de edad y por servidor:

Servidor correo/edad	20-40 años	40-60 años	>60 años
Gmail	$20\% \cdot 500 = 100$	$5\% \cdot 500 = 25$	$1\% \cdot 500 = 5$
Hotmail	$20\% \cdot 500 = 100$	$5\% \cdot 500 = 25$	$2\% \cdot 500 = 10$
Yahoo	$20\% \cdot 500 = 100$	$3\% \cdot 500 = 15$	$1\% \cdot 500 = 5$
Otros	$20\% \cdot 500 = 100$	$2\% \cdot 500 = 10$	$1\% \cdot 500 = 5$

Seguidamente, debemos encargar a nuestros entrevistadores que entrevisten a cien alumnos entre 20-40 años que usen Gmail, a cien que usen Hotmail, a cien que usen Yahoo y a cien que usen otros servidores de correo. También tienen que entrevistar a veinticinco alumnos entre 40-60 años que usen Gmail, a veinticinco que usen Hotmail, a quince que usen Yahoo y a diez que usen otros servidores de correo. Por último, tienen que entrevistar a cinco alumnos mayores de 60 años que usen Gmail, a diez que usen Hotmail, a cinco que usen Yahoo y a cinco que usen otros servidores de correo.

Actividad 2: estudio de la velocidad de banda ancha

Técnicas de muestreo. Muestreo probabilístico

Se quiere realizar un estudio de la velocidad de banda ancha en la ciudad de Palma de Mallorca. Queremos encuestar a trescientos clientes de banda ancha. Decid qué tipo de muestreo realizaríais en cada uno de los casos siguientes y explicad cómo hacerlo:

- Si disponemos de un listado de todos los clientes de banda ancha de todos los ISP (*Internet Service Provider*) de Internet.
- Si tenemos un mapa de la ciudad separada en cincuenta zonas geográficas donde tenemos una lista de usuarios por zona. En cada zona hay más o menos el mismo número de clientes.
- Si tenemos la tabla siguiente en la que se indica el porcentaje de clientes por franjas de edad (< 20 años, entre 20 y 40 años, entre 40 y 60 años y > 60 años) y por ISP (Telefónica, Ono, Orange, Jazztel, etc.).

ISP/edad	<20 años	20-40 años	40-60 años	> 60 años
Telefónica	20 %	30 %	10 %	10 %
Ono	5 %	15 %	3 %	2 %
Otros (Orange...)	1 %	2 %	1 %	1 %

Solución

a) En este caso tenemos que hacer un muestreo aleatorio simple. Vamos a suponer que tenemos ciento cincuenta mil clientes. En este caso usaremos la instrucción siguiente (para ahorrar espacio se muestran solo los cinco primeros valores):

```
sample(1:150000, 300, replace=TRUE) [1:5]
## [1] 140390 110720 61810 8989 128287
```

b) En este caso debemos realizar un muestreo por conglomerados donde estos son las zonas. Por ejemplo, supongamos que elegimos aleatoriamente diez zonas entre las cincuenta. Para elegir estas diez zonas, podemos usar nuevamente R de la misma forma en que lo hemos usado en el apartado anterior. Con la instrucción:

```
test2<-sample(50, 10, replace=TRUE)
test2[1:10]
## [1] 42 33 10 26 44 30 22 46 50 47
```

Luego, dentro de cada zona elegida tenemos de elegir aleatoriamente treinta clientes.

c) En este caso debemos realizar un muestreo por cuotas. Calculemos cuántos clientes hemos de tener en nuestra muestra por edad y por ISP:

ISP/edad	< 20 años	20-40 años	40-60 años	> 60 años
Telefónica	20 % · 300 = 60	30 % · 300 = 90	10 % · 300 = 30	10 % · 300 = 30
Ono	5 % · 300 = 15	15 % · 300 = 45	3 % · 300 = 9	2 % · 300 = 6
Otros (Orange...)	1 % · 300 = 3	2 % · 300 = 6	1 % · 300 = 3	1 % · 300 = 3

Por último, debemos encargar a nuestros entrevistadores que entrevisten a sesenta clientes de menos de 20 años que tengan Telefónica, a quince que tengan Ono y a tres que tengan otros ISP. También tienen que entrevistar a noventa clientes entre 20-40 años que tengan Telefónica, a cuarenta y cinco que tengan Ono y a seis que tengan otros ISP. También tienen que entrevistar a treinta clientes entre 40-60 años que tengan Telefónica, a nueve que tengan Ono y a tres que tengan otros ISP. Por último, tienen que entrevistar a treinta clientes mayores de 60 años que tengan Telefónica, seis clientes que tengan Ono y tres clientes que tengan otros ISP.

Actividad 3: estudio del uso de Internet

Técnicas de muestreo. Muestreo probabilístico

Se quiere realizar un estudio sobre el uso de Internet en un barrio de la ciudad de Barcelona. Queremos extraer una muestra de ciento cincuenta hogares. Decid qué tipo de muestreo haríais y cómo lo haríais en cada uno de los siguientes casos:

- Si tenemos un listado de todos los hogares de este barrio.
- Si tenemos un mapa del barrio de la ciudad separada por quince zonas geográficas.
- Si tenemos la información siguiente del barrio:

Tipo de hogar	< 25 años	Entre 25 y 40 años	Entre 40 y 65 años	> 65 años
Familia con hijos	5	50	40	35
Familia sin hijos	15	26	19	7
Estudiantes	30	10	0	0
Otros	12	15	18	23

Solución

- En este caso tenemos que hacer un muestreo aleatorio simple. Vamos a suponer que tenemos mil hogares. Simulemos en R una muestra de ciento cincuenta hogares. Para ello usamos las siguientes instrucciones (que solo muestran los diez primeros resultados, para ahorrar espacio):

```
test3<-sample(1:1000,150,replace=TRUE)
test3[1:10]
## [1] 88 60 32 31 17 70 95 32 80 49
```

- En este caso debemos realizar un muestreo por conglomerados donde estos son las zonas. Por ejemplo, supongamos que elegimos aleatoriamente cinco zonas entre las quince. Para elegir estas cinco zonas, podemos usar nuevamente R de la misma forma en que lo hemos usado en el apartado anterior, con lo que se obtiene:

```
test4<-sample(1:15,5,replace=TRUE)
test4
## [1] 5 11 2 15 1
```

Luego, dentro de cada zona elegida hemos de elegir aleatoriamente treinta hogares.

- En este caso debemos realizar un muestreo por cuotas. En total tenemos trescientos cinco hogares usando la tabla anterior. Calculamos cuántos clientes hemos de tener en nuestra muestra, por edad media y por tipo de hogar:

Tipo de hogar	< 25 años	Entre 25 y 40 años	Entre 40 y 65 años	> 65 años
Familia con hijos	$5 \cdot \frac{150}{305} \approx 2$	$50 \cdot \frac{150}{305} \approx 25$	$40 \cdot \frac{150}{305} \approx 20$	$35 \cdot \frac{150}{305} \approx 17$
Familia sin hijos	$15 \cdot \frac{150}{305} \approx 7$	$26 \cdot \frac{150}{305} \approx 13$	$19 \cdot \frac{150}{305} \approx 9$	$7 \cdot \frac{150}{305} \approx 4$
Estudiantes	$30 \cdot \frac{150}{305} \approx 15$	$10 \cdot \frac{150}{305} \approx 5$	0	0
Otros	$12 \cdot \frac{150}{305} \approx 6$	$15 \cdot \frac{150}{305} \approx 7$	$18 \cdot \frac{150}{305} \approx 9$	$23 \cdot \frac{150}{305} \approx 11$

Por último, hemos de encargar a nuestros entrevistadores que entrevisten a dos hogares con edad media de menos de veinticinco años en los que haya hijos, a siete hogares sin hijos, a quince hogares donde viven estudiantes y a seis hogares con otras características. También tienen que entrevistar a veinticinco hogares con edad media entre 20-40 años en los que haya hijos, a trece hogares sin hijos, a cinco hogares donde viven estudiantes y a siete hogares con otras características. También tienen que entrevistar a veinte hogares con edad media entre 40-65 años en los que haya hijos, a nueve hogares sin hijos y a nueve hogares con otras características. Por último, tienen que entrevistar a diecisiete hogares con edad media mayor de 65 años con hijos, a cuatro hogares sin hijos y a once hogares con otras características.