

# Muestreo

## Selección de actividades resueltas

© Jose Fco. Martínez Boscá, Arnau Mir Torres, Lluís  
M. Pla Aragonés &  
Ángel A. Juan (Editor)

© FUOC 2009

## Introducción

Imaginemos que hemos diseñado un nuevo programa de contabilidad y queremos comprobar el grado de aceptación entre las empresas de Cataluña. ¿Cómo podemos hacer dicha comprobación? Una manera simple de responder a dicha 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 ser no 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.

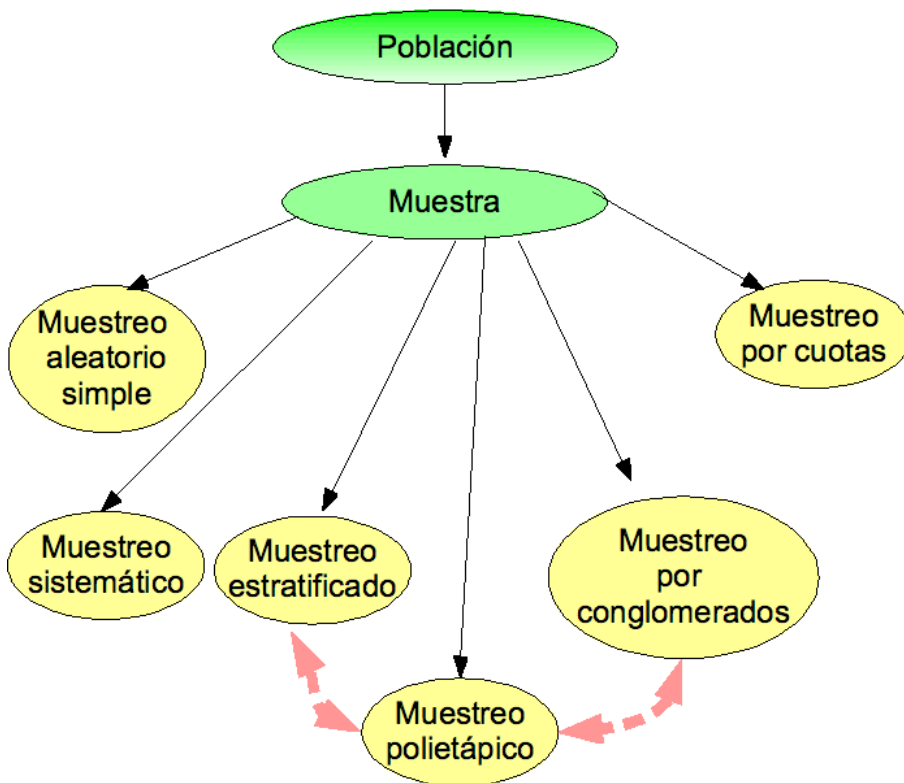
Las técnicas de muestreo nos dicen cómo hacer un estudio como el anterior para que la muestra que escojamos sea lo menos sesgada posible. Cuando decimos sesgada, queremos decir que la muestra tenga más empresas de un determinado tipo que de otro determinado tipo. Por ejemplo, si en nuestra muestra aparecen 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

# MUESTREO



## Actividades

### Muestreo

#### Actividad 1: Nivel de satisfacción de un servidor de correo.

Técnicas de muestreo. Muestreo probabilístico.

Queremos hacer un estudio de satisfacción que tienen los alumnos de la UOC con su servidor de correo. Para realizar este estudio, se decide pasar una encuesta a 500 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, País Valenciano y otros como zonas geográficas. En este caso, también disponemos de un listado de los alumnos donde hay indicada la comunidad autónoma de la que pertenece cada alumno.
- si tenemos la tabla siguiente donde 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



b) En este caso hemos de 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% del País Valenciano, el 15% de Baleares y el 10% restante perteneciente a otras comunidades autónomas. Entonces nuestra muestra tendrá  $0,6 \cdot 500 = 300$  alumnos de Cataluña,  $0,15 \cdot 500 = 75$  alumnos del País Valenciano, 75 alumnos de Baleares y 50 alumnos de otras comunidades. Después elegimos aleatoriamente unas comarcas de Cataluña, País Valenciano, Baleares y de otras comunidades autónomas. Supongamos, para fijar ideas, que elegimos 10 comarcas de Cataluña, 10 comarcas del País Valenciano, 10 comarcas de Baleares y 10 comarcas de otras comunidades autónomas. Seguidamente, hemos de elegir 30 estudiantes para cada comarca de Cataluña, 7 estudiantes de 5 comarcas del País Valenciano y 8 estudiantes de las otras 5 comarcas, lo mismo para las comarcas de Baleares y 5 estudiantes para las comarcas de otras comunidades autónomas todas las elecciones de forma aleatoria. De esta forma, habremos formado una muestra de 500 estudiantes.

c) En este caso hemos de realizar un muestreo por cuotas. Calculemos 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, hemos de encargar a nuestros entrevistadores que entrevisten a 100 alumnos entre 20-40 años que usen GMAIL, a 100 que usen HOTMAIL, a 100 que usen YAHOO y a 100 que usen otros servidores de correo. También tienen que entrevistar a 25 alumnos entre 40-60 años que usen GMAIL, a 25 que usen HOTMAIL, a 15 que usen YAHOO y a 10 que usen otros servidores de correo. Por último, tienen que entrevistar a 5 alumnos mayores de 60 años que usen GMAIL, a 10 que usen HOTMAIL, a 5 que usen YAHOO y a 5 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 300 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 50 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 donde 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 150000 clientes. Usaremos la instrucción `sample(1:150000,300,replace=TRUE)`

b) En este caso hemos de realizar un muestreo por conglomerados donde éstos son las zonas. Por ejemplo, supongamos que elegimos aleatoriamente 10 zonas entre las 50. Para elegir estas 10 zonas, podemos usar nuevamente R de la misma forma en que lo hemos usado en el apartado anterior. Con la instrucción

```
sample(1:50,10,replace=TRUE)
```

con lo que obtenemos por ejemplo las siguientes zonas:

```
[1] 22 38 31 21 7 31 23 19 30 22
```

Luego, dentro de cada zona elegida hemos de elegir aleatoriamente 30 clientes.

c) En este caso hemos de 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, hemos de encargar a nuestros entrevistadores que entrevisten a 60 clientes de menos de 20 años que tengan Telefónica, a 15 que tengan Ono y a 3 que tengan otros ISPs. También tienen que entrevistar a 90 clientes entre 20-40 años que tengan Telefónica, a 45 que tengan Ono y a 6 que tengan otros ISPs. También tienen que entrevistar a 30 clientes entre 40-60 años que tengan Telefónica, a 9 que tengan Ono y a 3 que tengan otros ISPs. Por último, tienen que entrevistar a 30 clientes mayores de 60 años que tengan Telefónica, 6 clientes que tengan Ono y 3 clientes que tengan otros ISPs.

**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 150 hogares. Decid qué tipo de muestreo haríais y cómo lo haríais en cada uno de los siguientes casos:

- a) Si tenemos un listado de todos los hogares de este barrio donde se indica el nivel adquisitivo de cada hogar: bajo, medio y alto.
- b) Si tenemos un mapa del barrio de la ciudad separada por 15 zonas geográficas.
- c) Si tenemos la información siguiente del barrio:

Tipo de hogar	Edad media de cada 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**

a) En este caso tenemos que hacer un muestreo aleatorio simple. Vamos a suponer que tenemos 1000 hogares. Simulemos en R una muestra de 150 hogares. Para ello usamos las siguientes instrucciones:

```
muestra<-sample(1:100,150,replace=TRUE)
> muestra
 [1] 92 54 47 37 79 72 20 45 4 32 31 93 84 51 50 5 3 28 60 73
63 97 61 78 24 28 19 5 47 21 83 45 83 64 30
 [36] 68 70 52 4 48 70 97 5 78 13 4 67 1 77 12 31 87 91 17 60
99 5 27 37 66 57 37 14 78 99 36 99 25 91 51
 [71] 93 46 1 36 92 61 1 10 84 79 26 37 28 32 59 79 34 19 24 39
3 59 78 26 43 62 26 24 85 20 73 19 55 8 73
 [106] 19 87 82 16 15 94 39 18 24 85 100 57 23 89 54 69 21 77 88 4
2 45 47 23 81 6 65 40 15 78 95 31 38 18 66
 [141] 77 65 49 76 22 11 20 69 98 3
> muestra[1:14]
 [1] 92 54 47 37 79 72 20 45 4 32 31 93 84 51
```

Ésta última nos muestra los 14 primeros hogares escogidos en la muestra.

b) En este caso hemos de realizar un muestreo por conglomerados donde éstos son las zonas. Por ejemplo, supongamos que elegimos aleatoriamente 5 zonas entre las 15. Para elegir estas 5 zonas, podemos usar nuevamente R de la misma forma en que lo hemos usado en el apartado anterior, con lo que se obtiene

```
> sample(1:15,5,replace=TRUE)
 [1] 5 13 15 10 11
```

Luego, dentro de cada zona elegida hemos de elegir aleatoriamente 30 hogares.

c) En este caso hemos de realizar un muestreo por cuotas. En total tenemos 305 hogares usando la tabla anterior. Calculemos cuántos clientes hemos de tener en nuestra muestra edad media y por tipo de hogar:

Tipo de hogar	Edad media de cada 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 2 hogares con edad media de menos de 25 años que haya hijos, a 7 hogares sin hijos, a 15



hogares donde viven estudiantes y a 6 hogares con otras características. También tienen que entrevistar a 25 hogares con edad media entre 20-40 años que haya hijos, a 13 hogares sin hijos, a 5 hogares donde viven estudiantes y a 7 hogares con otras características. También tienen que entrevistar a 20 hogares con edad media entre 40-65 años que haya hijos, a 9 hogares sin hijos y a 9 hogares con otras características. Por último, tienen que entrevistar a 17 hogares con edad media mayor de 65 años con hijos, a 4 hogares sin hijos y a 11 hogares con otras características.

### **Direcciones de interés**

[http://en.wikipedia.org/wiki/Sampling\\_\(statistics\)](http://en.wikipedia.org/wiki/Sampling_(statistics))  
Definición y tipos de muestreo en estadística.

[http://en.wikipedia.org/wiki/Simple\\_random\\_sample](http://en.wikipedia.org/wiki/Simple_random_sample)  
Muestreo aleatorio simple.

[http://en.wikipedia.org/wiki/Stratified\\_sampling](http://en.wikipedia.org/wiki/Stratified_sampling)  
Muestreo estratificado.

[http://en.wikipedia.org/wiki/Systematic\\_sampling](http://en.wikipedia.org/wiki/Systematic_sampling)  
Muestreo sistemático.