

Tema 5 - Muestreo estadístico

Juan Gabriel Gomila, Arnau Mir y Ricardo Alberich

Conceptos básicos de muestreo

- ▶ En todo estudio estadístico distinguiremos entre **población**, (conjunto de sujetos con una o varias características que podemos medir y deseamos estudiar), y **muestra**, (subconjunto de una población.)
- ▶ Dos tipos de análisis estadístico:
 - ▶ **Exploratorio o descriptivo:** estadística descriptiva.
 - ▶ **Inferencial o confirmatorio:** estadística inferencial.

Conceptos básicos de muestreo

Pasos en un estudio inferencial:

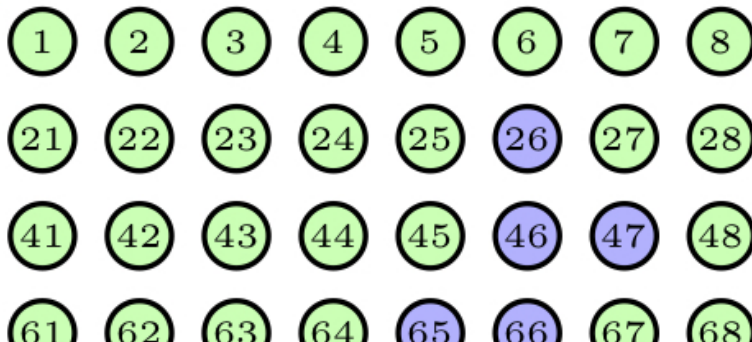
- ▶ Establecer la característica que se desea estimar o la hipótesis que se desea contrastar.
- ▶ Determinar la información (los datos) que se necesita para hacerlo.
- ▶ Diseñar un experimento que permita recoger estos datos; este paso incluye:
 - ▶ Decidir qué tipo de muestra se va a tomar y su tamaño.
 - ▶ Elegir las técnicas adecuadas para realizar las inferencias deseadas a partir de la muestra que se tomará.

Conceptos básicos de muestreo

- ▶ Tomar una muestra y medir los datos deseados sobre los individuos que la forman.
- ▶ Aplicar las técnicas de inferencia elegidas con el *software* adecuado.
- ▶ Obtener conclusiones.
- ▶ Si las conclusiones son fiables y suficientes, redactar un informe; en caso contrario, volver a empezar.

Tipos de muestreo: Muestreo aleatorio con y sin reposición

- ▶ Un **muestreo aleatorio** consiste en seleccionar una muestra de la población de manera que todas las muestras del mismo tamaño sean **equiprobables**.
- ▶ Dos tipos básicos de muestreo:
 - ▶ con reposición:



Tipos de muestreo: Muestreo aleatorio con reposición

En R

```
sample(1:100, 15, replace=TRUE)
```

```
## [1] 96 36 2 89 66 29 83 79 64 32 21 77 32 84 5
```

Veamos un ejemplo más elaborado. Consideremos la tabla de datos *iris* que contiene 150 flores de 3 especies diferentes: **setosa**, **versicolor** y **virginica**. La tabla de datos contiene 5 variables: la longitud y amplitud del pétalo, la longitud y la amplitud del sépalo y la especie de la flor:

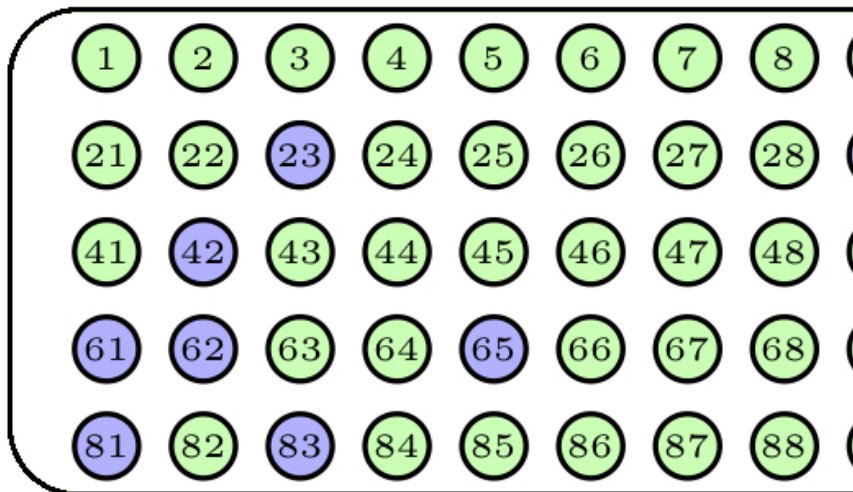
Tipos de muestreo: Muestreo aleatorio con reposición

```
head(iris)
```

##		Sepal.Length	Sepal.Width	Petal.Length
## 1		5.0999999999999996	3.5000000000000000	1.3999999999999999
## 2		4.9000000000000004	3.0000000000000000	1.3999999999999999
## 3		4.7000000000000002	3.2000000000000002	1.3000000000000000
## 4		4.5999999999999996	3.1000000000000001	1.5000000000000000
## 5		5.0000000000000000	3.6000000000000001	1.3999999999999999
## 6		5.4000000000000004	3.8999999999999999	1.7000000000000000
##		Petal.Width	Species	
## 1		0.20000000000000001	setosa	
## 2		0.20000000000000001	setosa	
## 3		0.20000000000000001	setosa	
## 4		0.20000000000000001	setosa	
## 5		0.20000000000000001	setosa	
## 6		0.40000000000000002	setosa	

Tipos de muestreo: Muestreo aleatorio sin reposición

- ▶ sin reposición:



Tipos de muestreo: Muestreo aleatorio sin reposición

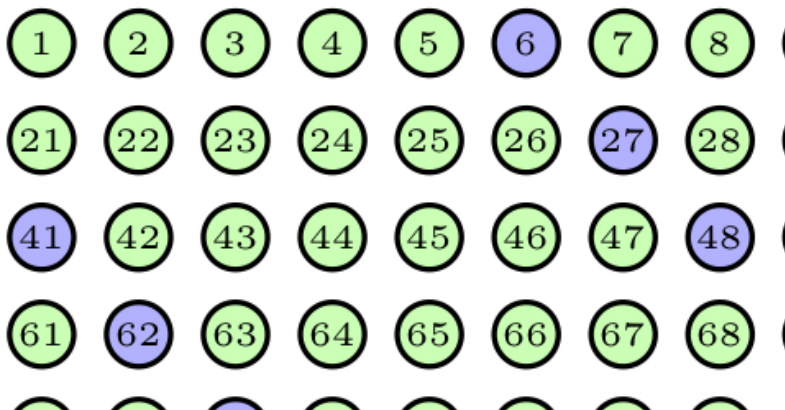
En R

```
sample(1:100, 15, replace=FALSE)
```

```
## [1] 23 48 21 25 47 17 74 15 24 7 22 20 71 9 3
```

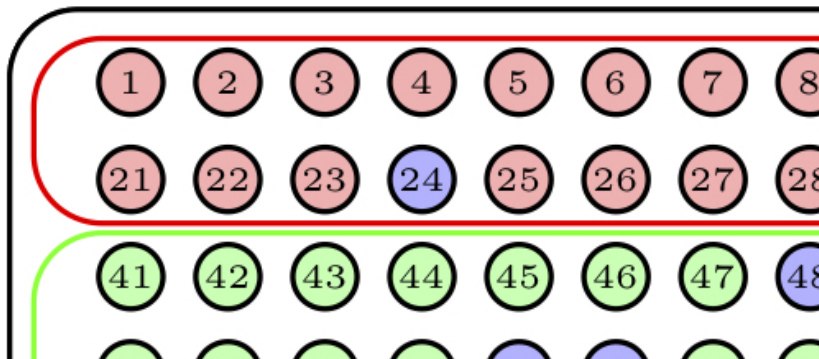
Muestreo sistemático

- Suponemos que disponemos de una lista ordenada de todos los individuos de la población. Muestreo sistemático: tomarlos a intervalos constantes escogiendo al azar el primer individuo que elegimos:



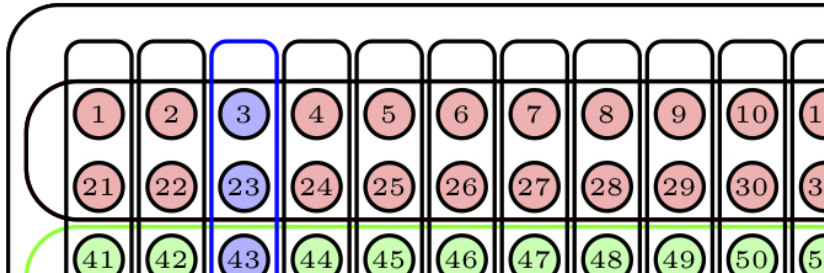
Muestreo aleatorio estratificado

- ▶ Este tipo de muestreo se utiliza cuando la población está clasificada en **estratos** que son de interés para la propiedad estudiada.
- ▶ Se toma una muestra aleatoria de cada estrato y se unen en una muestra global. A este proceso se le llama **muestreo aleatorio estratificado**:



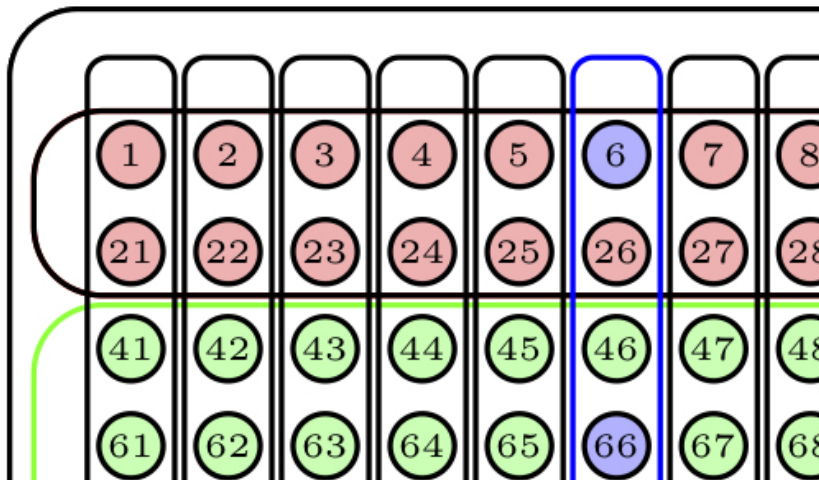
Muestreo por conglomerados

- ▶ El proceso de obtener y estudiar una muestra aleatoria en algunos casos es caro o difícil, incluso aunque dispongamos de la lista completa de la población.
- ▶ Una alternativa posible sería, en vez de extraer una muestra aleatoria de todos los individuos de la población, escoger primero al azar unos subconjuntos en los que la población está dividida, a las que llamamos en este contexto **conglomerados** (*clusters*):



Muestreo polietápico

- Una vez seleccionada la muestra aleatoria de conglomerados, tomar de alguna manera una muestra aleatoria de cada uno de ellos:



Guía rápida

- ▶ `sample(x, n, replace=...)` genera una muestra aleatoria de tamaño `n` del vector `x`, con reposición si igualamos `replace` a `TRUE` y sin reposición si lo igualamos a `FALSE` (su valor por defecto). Si `x` es un número natural `x`, representa el vector `1,2,...,x`.
- ▶ `set.seed` permite fijar la semilla de aleatoriedad.
- ▶ `replicate(n, expresión)` evalúa `n` veces la expresión, y organiza los resultados como las columnas de una matriz (o un vector, si el resultado de cada expresión es unidimensional).