



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información

# #dataviz



## **7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)**

### **7.1. Muestreo: tipos y ejemplos**



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

#### CENSO

Imagen completa de una población, requiere de gran cantidad de recursos humanos y físicos. Este estudio puede llegar a perder interés de la investigación al ser tan dispendiosa.



#### MUESTRA

Selección de un subconjunto de elementos que hacen parte de un grupo mas grande. Puede llegar a ser mas precisa que un censo y obtener conclusiones aproximadas a la realidad.





## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

#### **\_probabilístico**

##### **\_aleatorio simple**

##### **\_aleatorio sistemático**

##### **\_aleatorio estratificado**

##### **\_aleatorio por conglomerados**

#### **\_no probabilístico**

##### **\_por cuotas**

##### **\_intencional o de conveniencia**

##### **\_bola de nieve**

##### **\_discrecional**



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

# \_probabilístico

- Los métodos de muestreo probabilísticos son aquellos que se basan en el principio de **equiprobabilidad**. Es decir, aquellos en los que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra y, consiguientemente, todas las posibles muestras de tamaño  $n$  tienen la misma probabilidad de ser seleccionadas.
- Sólo estos métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables.



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

- A veces, para estudios exploratorios, el muestreo probabilístico resulta excesivamente costoso y se acude a métodos no probabilísticos, aun siendo conscientes de que no sirven para realizar generalizaciones (estimaciones inferenciales sobre la población), pues no se tiene certeza de que la muestra extraída sea representativa, ya que no todos los sujetos de la población tienen la misma probabilidad de ser elegidos. En general **se seleccionan a los sujetos siguiendo determinados criterios procurando, en la medida de lo posible, que la muestra sea representativa.**

## **\_no probabilístico**

\_por cuotas

\_intencional o de conveniencia

\_bola de nieve

\_discrecional



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

**\_probabilístico**

**\_aleatorio simple**

\_aleatorio sistemático

\_aleatorio estratificado

\_aleatorio por conglomerados

**\_no probabilístico**

\_por cuotas

\_intencional o de conveniencia

\_bola de nieve

\_discrecional



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

# **\_probabilístico**

\_aleatorio simple

**\_aleatorio sistemático**

\_aleatorio estratificado

\_aleatorio por conglomerados

**\_no probabilístico**

\_por cuotas

\_intencional o de conveniencia

\_bola de nieve

\_discrecional



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

# \_probabilístico

\_aleatorio simple

\_aleatorio sistemático

# \_aleatorio estratificado

\_aleatorio por conglomerados

no probabilíst

En una fábrica que consta de 600 trabajadores queremos tomar una muestra de 20. Sabemos que hay 200 trabajadores en la sección A, 150 en la B, 150 en la C y 100 en la D.

$$\frac{20}{600} = \frac{x_1}{200}$$

$$x_1 = 6.6 \approx 7 \text{ trabajadores de A}$$

$$\frac{20}{600} = \frac{x_2}{150}$$

$$x_2 = 5 \quad 5 \text{ trabajadores de B}$$

$$\frac{20}{600} = \frac{x_3}{150}$$

$$x_3 = 5 \quad 5 \text{ trabajadores de C}$$

$$\frac{20}{600} = \frac{x_4}{100}$$

$$x_4 = 3.3 \approx 3 \text{ trabajadores de D}$$

\_bola de

\_discrecional



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

# \_probabilístico

\_aleatorio simple

\_aleatorio sistemático

\_aleatorio estratificado

# \_aleatorio por conglomerados

La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

- **Afijación Simple:** A cada estrato le corresponde igual número de elementos muestrales.
- **Afijación Proporcional:** La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.
- **Afijación Óptima:** Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica. Tiene poca aplicación ya que no se suele conocer la desviación.



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

# **\_probabilístico**

\_aleatorio simple

\_aleatorio sistemático

\_aleatorio estratificado

# **\_aleatorio por conglomerados**

\_no probabilístico

\_por cuotas

[http://www.cis.es/cis/opencm/ES/1\\_encuestas/estudios/ver.jsp?estudio=14212](http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=14212)

\_bola de nieve

\_discrecional



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

- En este tipo de muestreo se fijan unas "cuotas" que consisten en un número de individuos que reúnen unas determinadas condiciones, por ejemplo: 20 individuos de 25 a 40 años, de sexo femenino y residentes en Gijón.
- Una vez determinada la cuota se eligen los primeros que se encuentren que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión.

**\_no probabilístico**

**\_por cuotas**

**\_intencional o de conveniencia**

**\_bola de nieve**

**\_discrecional**



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

- Este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras "representativas" mediante la inclusión en la muestra de grupos supuestamente típicos. Es muy frecuente su utilización en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto.
- También puede ser que el investigador seleccione directa e intencionadamente los individuos de la población. El caso más frecuente de este procedimiento es el utilizar como muestra los individuos a los que se tiene fácil acceso (los profesores de universidad emplean con mucha frecuencia a sus propios alumnos).

**\_no probabilístico**

**\_por cuotas**

**\_intencional o de conveniencia**

**\_bola de nieve**

**\_discrecional**



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

**\_probabilístico**

Se localiza a algunos individuos, los cuales conducen a otros, y estos a otros, y así hasta conseguir una muestra suficiente. Este tipo se emplea muy frecuentemente cuando se hacen estudios con poblaciones "marginales", delincuentes, sectas, determinados tipos de enfermos, etc.

**\_no probabilístico**

**\_por cuotas**

**\_intencional o de conveniencia**

**\_bola de nieve**

**\_discrecional**



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

**\_probabilístico**

**\_aleatorio simple**

**\_aleatorio sistemático**

**\_aleatorio estratificado**

**\_aleatorio por conglomerados**

**\_no probabilístico**

**\_por cuotas**

**\_intencional o de conveniencia**

**\_bola de nieve**

**\_discrecional**



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.1. Muestreo: tipos y ejemplos

#### **\_probabilístico**

##### **\_aleatorio simple**

##### **\_aleatorio sistemático**

##### **\_aleatorio estratificado**

##### **\_aleatorio por conglomerados**

#### **\_no probabilístico**

##### **\_por cuotas**

##### **\_intencional o de conveniencia**

##### **\_bola de nieve**

##### **\_discrecional**



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.2. Cálculo del tamaño muestral: el nivel de confianza y el margen de error



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.2. Cálculo del tamaño muestral: el nivel de confianza y el margen de error

$$\frac{k^2 N p q}{e^2 (N - 1) + k^2 p q}$$

- N: es el tamaño de la población o universo (número total de posibles encuestados)
- k: es una constante que depende del nivel de confianza que asignemos. El **nivel de confianza** indica la probabilidad de que los resultados de nuestra investigación sean ciertos: un 95,5 % de confianza es lo mismo que decir que nos podemos equivocar con una probabilidad del 4,5%. Los valores de k se obtienen de la tabla de la [distribución normal](#) estándar N(0,1)
- e: es el **error muestral** deseado, en tanto por uno. El error muestral es la diferencia que puede haber entre el resultado que obtenemos preguntando a una muestra de la población y el que obtendríamos si preguntáramos al total de ella
- p: proporción de individuos que poseen en la población la característica de estudio. Este dato es generalmente desconocido y se suele suponer que  $p=q=0.5$  que es la opción más segura.
- q: proporción de individuos que no poseen esa característica, es decir, es  $1-p$



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.2. Cálculo del tamaño muestral: el nivel de confianza y el margen de error

<http://www.raosoft.com/samplesize.html>



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.2. Cálculo del tamaño muestral: el nivel de confianza y el margen de error

- [METROSCOPIA](#)
- [Barómetro del CIS](#)
- INE-EPA (METODOLOGÍA, [MARGEN DE ERROR](#))
- EU Labour Force Survey (METODOLOGÍA, [MARGEN DE ERROR](#))
- [EL MARGEN DE ERROR EN LA FÍSICA](#)



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

### 7.2. Cálculo del tamaño muestral: el nivel de confianza y el margen de error

1. Cálculo del número de encuestas que hay que hacer para un estudio sobre el tejido empresarial en la comarca de \_\_\_\_\_



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata

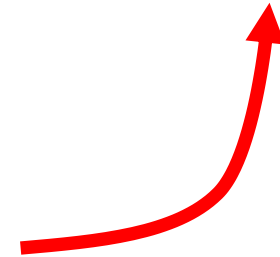


## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="button" value="Enviar"/>
----------------------	----------------------	----------------------	---------------------------------------

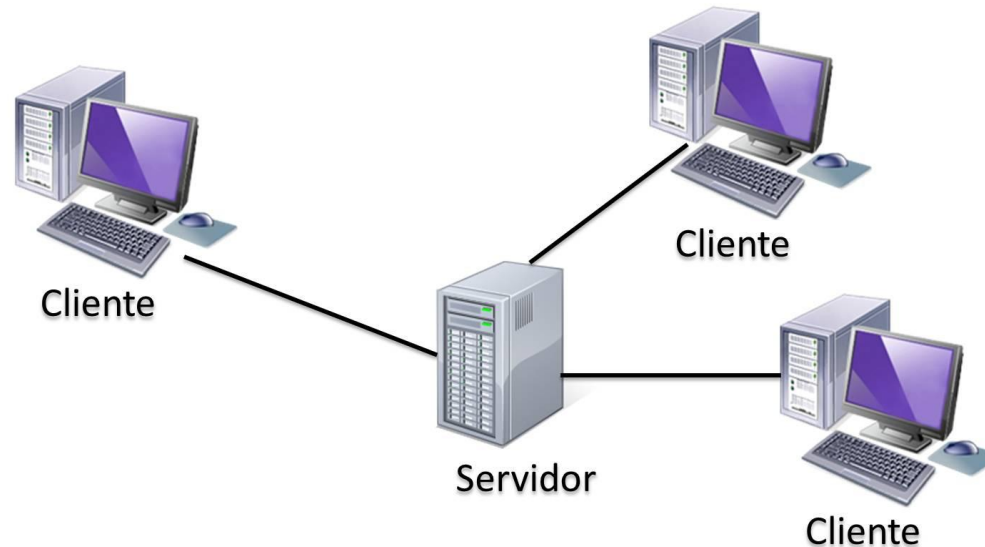
**lenguaje SQL**





## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata





## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

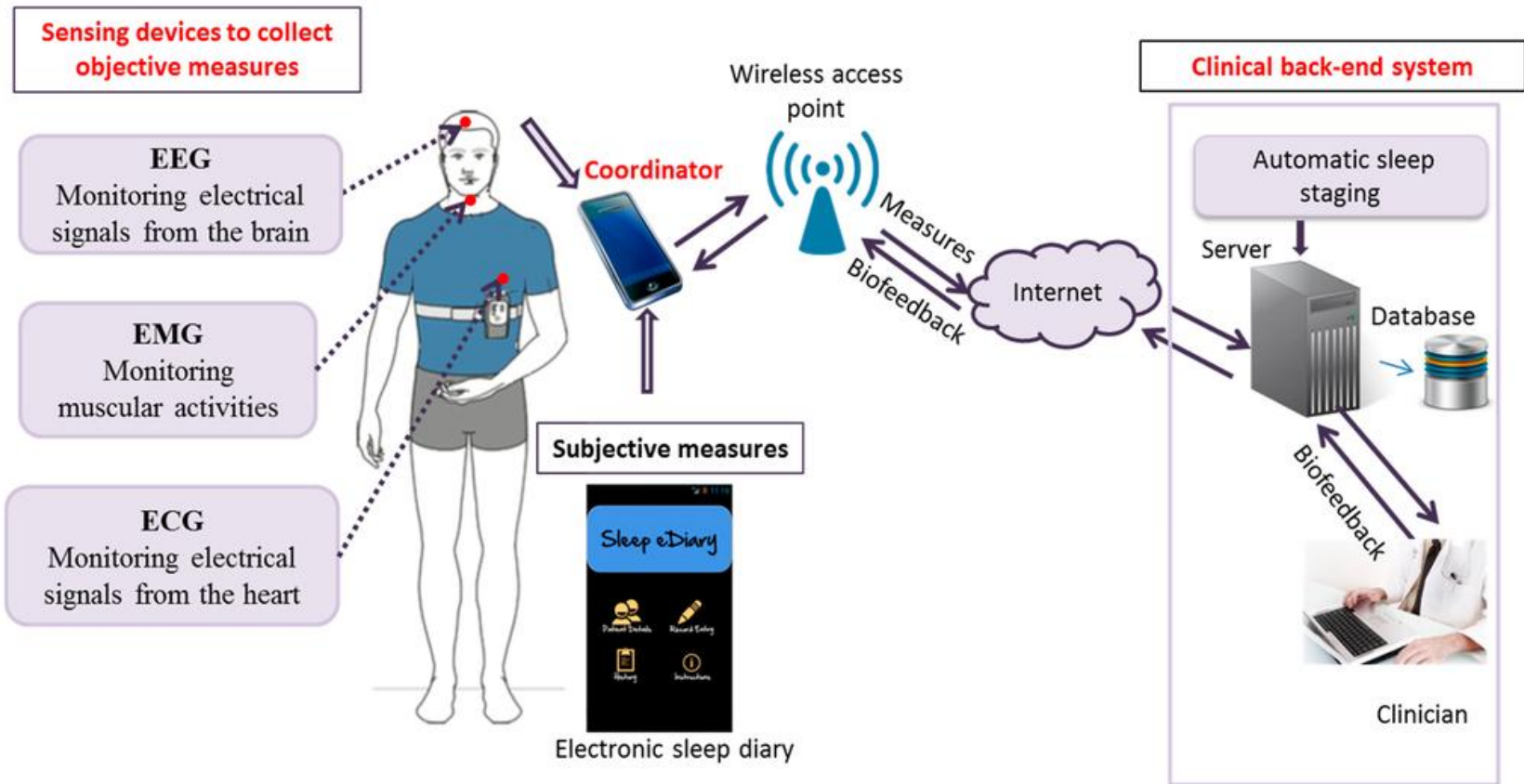
7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata





## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata





## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata

postgresql

\_base de datos

\_módulo de geolocalización de los datos: postGIS

\_módulo de servidor de páginas web: APACHE+PHP

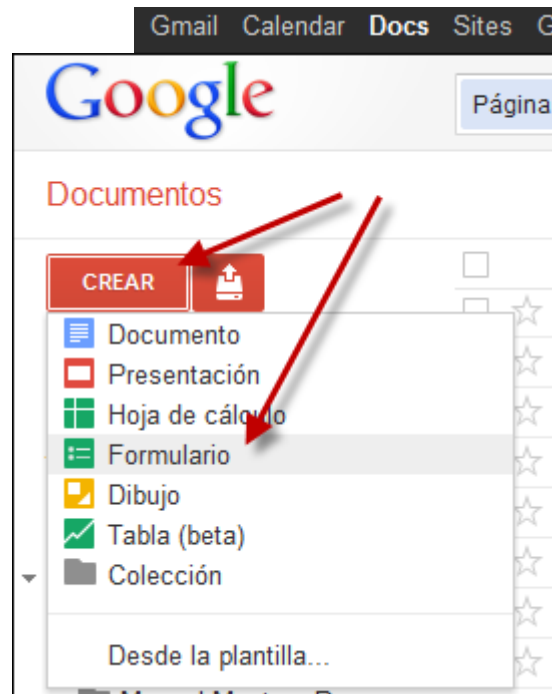
\_conectores: ODBC, JDBC

QGIS (Sistema de Información Geográfica - GIS)



## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata





## 7. Muestreo estadístico. Bases de datos, tratamiento y análisis visual de la información (#dataviz)

7.3. Procesamiento: almacenamiento de bases de datos, sistemas de consulta, lectura y escritura remota contra bases de datos (sistemas de encuestación online), introducción al bigdata

# #bigdata



## 8. Tratamiento y análisis de la información



## 8. Tratamiento y análisis de la información

### 8.1. ACCESS Y EXCEL COMO HERRAMIENTA DE ANÁLISIS DE DATOS

POSTGRESQL

MYSQL

ORACLE

HADOOP

MS-SQL SERVER

CLUSTRIX

...

# lenguaje SQL





## 8. Tratamiento y análisis de la información

### 8.2. CROSSTABLES EN EXCEL: TABLAS DINÁMICAS

1. Pirámide poblacional
2. Añadir sexo “hombre/mujer” en access
3. Añadir nombres de los municipios
4. crosstable estado civil/sexo
5. crosstable estado civil/sexo/grupo de edad
6. crosstable nivel de estudios completado/sexo
7. crosstable costa-interior/nivel de estudios
8. crosstable salarios/educacion & salarios/educación/sexo
9. crosstable necesidades específicas
10. ...



## 8. Tratamiento y análisis de la información

### 8.3. REPRESENTACIÓN GRÁFICA DE DATOS EN EXCEL: GRÁFICAS BÁSICAS



## 8. Tratamiento y análisis de la información

### 8.4. REPRESENTACIÓN GRÁFICA DE DATOS EN EXCEL: GR. AVANZADAS

#### Regresiones

1. En función de la edad, el salario es mayor?
2. Cuanto mayor nivel educativo, el salario es mayor?
3. Dispersión de los municipios:  
tasa de ocupación/nivel educativo promedio



## 8. Tratamiento y análisis de la información

### 8.5. REPRESENTACIÓN GEOGRÁFICA DE DATOS ESTADÍSTICOS EN QGIS

1. Municipios según el salario
2. Municipios según el porcentaje de mujeres
3. Municipios según la tasa de ocupación
4. Carencias según municipio