

eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

## Material de estudio

# *OCW 2019: Curso práctico para el análisis e inferencia estadística con Mathematica*

---

## Tema 1. Variables y series estadísticas

### **Equipo docente del curso**

*Arrospide Zabala, Eneko  
Martín Yagüe, Luis  
Unzueta Inchaurre, Aitziber  
Soto Merino, Juan Carlos  
Durana Apaolaza, Gaizka  
Bikandi Irazabal, Iñaki*

Departamento de Matemática Aplicada  
Escuela de Ingeniería de Bilbao, Edificio II-I

**OCW**  
Open CourseWare



## TEMA 1. VARIABLES Y SERIES ESTADÍSTICAS

### Definiciones básicas

---

#### Dato

Un dato (ó carácter ó variable) es una cualidad o propiedad inherente de los individuos de una población. Las variables se pueden clasificar en cuantitativas (ó categóricas) y cualitativas (ó numéricas).

---

#### Estadística descriptiva

La Estadística descriptiva es la parte de la Estadística relacionada con la recolección, clasificación y descripción de los datos obtenidos en una muestra de una población. Su propósito es facilitar el uso de esos datos, generalmente, con el apoyo de tablas, medidas numéricas o gráficas.

---

#### Variable estadística

Los datos de los individuos de una población que se pretende estudiar constituyen variables estadísticas.

La descripción de una variable estadística supone:

- clasificarla
  - obtener las tablas con las frecuencias absolutas y relativas de sus valores
  - realizar representaciones gráficas (diagramas de barras, histogramas, etc.)
  - calcular los valores numéricos (estadísticos) que la resumen
- 

#### Serie estadística

Es una colección de los diferentes valores observados de una variable estadística.

En *Mathematica* la serie se considera como una lista que se asigna a una variable. Se puede operar con los diferentes valores de la variable estadística accediendo a los elementos de la lista.

La definición de la lista puede realizarse:

- tecleando los valores entre llaves separados por comas
  - importando la información de un fichero estructurado (*txt*, *xlsx*, ...)
- 

### Introducción de una serie por teclado

La definición de la lista que contiene la serie estadística puede realizarse tecleando los valores entre llaves separados por comas.

**Ejemplo.** Definición de una serie como una lista asignada a una variable.

```
altura = {1.87, 1.69, 1.71, 1.75, 1.77, 1.81}
```

```
{1.87, 1.69, 1.71, 1.75, 1.77, 1.81}
```

`altura[[2]]` (\* acceso a un elemento de la serie \*)

1.69

`Length[altura]` (\* número de elementos (observaciones) de la serie \*)

6

## Importación de datos desde un fichero externo

La definición de la lista puede realizarse, también, importando la información de un fichero estructurado (*txt*, *xlsx*, ...).

**Ejemplo.** Importación de datos de un fichero *Excel* guardado en un directorio del usuario. Supóngase que contiene observaciones de longitud y diámetro, en milímetros, de once tornillos fabricados por dos máquinas, denotadas como *A* y *B*.

`SetDirectory["C:\\Users\\USUARIO\\Desktop"];` (\* establecer directorio de trabajo \*)

`Directory[]` (\* comprobar directorio actual de trabajo \*)

C:\Users\USUARIO\Desktop

- importación de los datos de la hoja *Excel*

`dat = Import["OCW2019Importar.xlsx"]`

```
{{Máquina, Longitud (mm), Diámetro (mm)}, {A, 17., 5.},
 {A, 17., 5.}, {B, 16., 3.}, {B, 15., 4.}, {B, 16., 5.}, {A, 16., 6.},
 {B, 17., 5.}, {B, 18., 7.}, {A, 19., 4.}, {A, 19., 5.}, {A, 15., 6.}}
```

- creación de la lista con los datos de la hoja *Excel*

`dat2 = Flatten[dat, 1]`

```
{{Máquina, Longitud (mm), Diámetro (mm)}, {A, 17., 5.},
 {A, 17., 5.}, {B, 16., 3.}, {B, 15., 4.}, {B, 16., 5.}, {A, 16., 6.},
 {B, 17., 5.}, {B, 18., 7.}, {A, 19., 4.}, {A, 19., 5.}, {A, 15., 6.}}
```

- representación tabular de los datos de la lista obtenida

`Grid[dat2, Dividers -> All]`

Máquina	Longitud (mm)	Diámetro (mm)
A	17.	5.
A	17.	5.
B	16.	3.
B	15.	4.
B	16.	5.
A	16.	6.
B	17.	5.
B	18.	7.
A	19.	4.
A	19.	5.
A	15.	6.

- eliminación de la cabecera (primer elemento de la lista `dat2`)

`datos = Drop[dat2, 1]`

```
{{A, 17., 5.}, {A, 17., 5.}, {B, 16., 3.}, {B, 15., 4.}, {B, 16., 5.},
 {A, 16., 6.}, {B, 17., 5.}, {B, 18., 7.}, {A, 19., 4.}, {A, 19., 5.}, {A, 15., 6.}}
```

- definición de las variables estadísticas

**maq = datos[[All, 1]] (\* variable aleatoria cualitativa \*)**

{A, A, B, B, B, A, B, B, A, A, A}

**long = IntegerPart[datos[[All, 2]]] (\* variable aleatoria cuantitativa continua \*)**

{17, 17, 16, 15, 16, 16, 17, 18, 19, 19, 15}

**diam = IntegerPart[datos[[All, 3]]] (\* variable aleatoria cuantitativa continua \*)**

{5, 5, 3, 4, 5, 6, 5, 7, 4, 5, 6}

- obtención de las distintas observaciones de un tornillo (individuo)

**datos[[2]] (\* segundo tornillo: fabricado por A, longitud=17mm, diámetro=5mm \*)**

{A, 17., 5.}

- selección de información de la estructura de datos
  - diámetro de los tornillos fabricados por la máquina A

**dA = IntegerPart[Cases[datos, {"A", \_, x\_] => x]]**

{5, 5, 6, 4, 5, 6}

- longitud de los tornillos fabricados por la máquina B

**lB = IntegerPart[Cases[datos, {"B", x\_, \_} => x]]**

{16, 15, 16, 17, 18}

- diámetro de los tornillos fabricados por la máquina A cuya longitud es 19mm

**lA19 = IntegerPart[Cases[datos, {"A", 19, x\_] => x]]**

{4, 5}

**Ejemplo.** Importación de un fichero *Excel* de un ejemplo de conjunto de datos del programa.

- importación del fichero completo

**city = Import["ExampleData/cities.xlsx"]**

{{{City, Country, Population}, {Tokyo, Japan,  $8.3 \times 10^6$ }, {Chicago, United States,  $2.8 \times 10^6$ }, {London, United Kingdom,  $7.4 \times 10^6$ }, {Berlin, Germany,  $3.4 \times 10^6$ }}, {{}}

- creación de la lista con los datos de la hoja *Excel*

**city2 = Flatten[city, 1]**

{{City, Country, Population}, {Tokyo, Japan,  $8.3 \times 10^6$ }, {Chicago, United States,  $2.8 \times 10^6$ }, {London, United Kingdom,  $7.4 \times 10^6$ }, {Berlin, Germany,  $3.4 \times 10^6$ }, {}}

- representación tabular de los datos de la lista

**Grid[city2, Dividers -> All]**

City	Country	Population
Tokyo	Japan	$8.3 \times 10^6$
Chicago	United States	$2.8 \times 10^6$
London	United Kingdom	$7.4 \times 10^6$
Berlin	Germany	$3.4 \times 10^6$

- eliminación de la cabecera

```
ciudad1 = Drop[Flatten[city, 1], 1]
```

```
{{Tokyo, Japan, 8.3×106}, {Chicago, United States, 2.8×106},  
{London, United Kingdom, 7.4×106}, {Berlin, Germany, 3.4×106}, {}}
```

- eliminación del último elemento vacío

```
ciudad2 = Drop[ciudad1, -1]
```

```
{{Tokyo, Japan, 8.3×106}, {Chicago, United States, 2.8×106},  
{London, United Kingdom, 7.4×106}, {Berlin, Germany, 3.4×106}}
```

- definición de las variables estadísticas

```
ciudad = ciudad2[[All, 1]] (* variable aleatoria cualitativa *)
```

```
{Tokyo, Chicago, London, Berlin}
```

```
país = ciudad2[[All, 2]] (* variable aleatoria cualitativa *)
```

```
{Japan, United States, United Kingdom, Germany}
```

```
población = ciudad2[[All, 3]] (* variable aleatoria cuantitativa discreta *)
```

```
{8.3×106, 2.8×106, 7.4×106, 3.4×106}
```

- importación selectiva
  - observaciones de Tokyo y London (filas 2 y 4)

```
TokyoLondon = Import["ExampleData/cities.xlsx", {"Data", 1, {2, 4}}]
```

```
{{Tokyo, Japan, 8.3×106}, {London, United Kingdom, 7.4×106}}
```

- población de las ciudades (filas 2, 3, 4 y 5 de la columna 3) sin cabecera

```
pobl = Import["ExampleData/cities.xlsx", {"Data", 1, {2, 3, 4, 5}, 3}]
```

```
{8.3×106, 2.8×106, 7.4×106, 3.4×106}
```