

Computación estadística!

▼ Cuaderno 1

07 de Marzo de 2022

Subir un archivo desde excel

1. En el menu izquierdo abrir files
2. Conectar el Drive
3. Buscar el archivo dentro de MyDrive
4. Copiar el Path

Crear un notebook que haga las siguientes tareas:

1. Cargar un archivo de excel con una ruta permanente en drive (apoyarse en el modulo pandas)
2. Cargar una matriz 4×4 en el notebook
3. Copiar una imagen del internet y llevarla a Notebook
4. Crear una tabla en ($LATEX$)
5. Copiar una ecuación y llevarla a ($LATEX$)

Crear una cuenta en [GitHub](#)

1. Entrar al link
2. Crear Cuenta
3. Crear un nuevo repositorio público
4. Subir los Archivos .ipynb

▼ Solución

▼ 1. Cargar un archivo con ruta permanente

```
Link_archivo='/content/drive/MyDrive/My_Comp_Est_2022-1/california_housing_train.csv'
```

Double-click (or enter) to edit

```
import pandas as pd
Data=pd.read_csv(Link_archivo)
```

```
print(Data)
```

| | longitude | latitude | housing_median_age | total_rooms | total_bedrooms | \ |
|-------|-----------|----------|--------------------|-------------|----------------|---|
| 0 | -114.31 | 34.19 | 15.0 | 5612.0 | 1283.0 | |
| 1 | -114.47 | 34.40 | 19.0 | 7650.0 | 1901.0 | |
| 2 | -114.56 | 33.69 | 17.0 | 720.0 | 174.0 | |
| 3 | -114.57 | 33.64 | 14.0 | 1501.0 | 337.0 | |
| 4 | -114.57 | 33.57 | 20.0 | 1454.0 | 326.0 | |
| ... | ... | ... | ... | ... | ... | |
| 16995 | -124.26 | 40.58 | 52.0 | 2217.0 | 394.0 | |
| 16996 | -124.27 | 40.69 | 36.0 | 2349.0 | 528.0 | |
| 16997 | -124.30 | 41.84 | 17.0 | 2677.0 | 531.0 | |
| 16998 | -124.30 | 41.80 | 19.0 | 2672.0 | 552.0 | |
| 16999 | -124.35 | 40.54 | 52.0 | 1820.0 | 300.0 | |

| | population | households | median_income | median_house_value |
|---|------------|------------|---------------|--------------------|
| 0 | 1015.0 | 472.0 | 1.4936 | 66900.0 |
| 1 | 1129.0 | 463.0 | 1.8200 | 80100.0 |
| 2 | 333.0 | 117.0 | 1.6509 | 85700.0 |
| 3 | 515.0 | 226.0 | 3.1917 | 73400.0 |
| 4 | 624.0 | 262.0 | 1.9250 | 65500.0 |

```
...      ...      ...      ...      ...
16995      907.0      369.0      2.3571      111400.0
16996      1194.0      465.0      2.5179      79000.0
16997      1244.0      456.0      3.0313      103600.0
16998      1298.0      478.0      1.9797      85800.0
16999      806.0      270.0      3.0147      94600.0
```

```
[17000 rows x 9 columns]
```

▼ 2. Cargar una matriz de 4×4

Para este ejemplo opté por hacer una matriz de covarianza, las cuales generalmente tienen muchas aplicaciones estadísticas (se usa principalmente en el cálculo matricial de los coeficientes de la regresión lineal mediante Mínimos Cuadrados Ordinarios, entre otros usos).

[Link para construir ecuaciones en *LaTeX*](#)

[Link 2](#)

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & Cov(X_1, X_3) & Cov(X_1, X_4) \\ Cov(X_2, x1) & Var(X_2) & Cov(X_2, X_3) & Cov(X_2, X_4) \\ Cov(X_3, x1) & Cov(X_3, X_2) & Var(X_3) & Cov(X_3, X_4) \\ Cov(X_4, x1) & Cov(X_4, X_2) & Cov(X_4, X_3) & Var(X_4) \end{pmatrix}$$

De manera más general:

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22}^2 & \dots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nm}^2 \end{pmatrix}$$

3. Subir Una imagen desde la web

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} ; -\infty < z < \infty$$

$$F(z) = p(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du$$

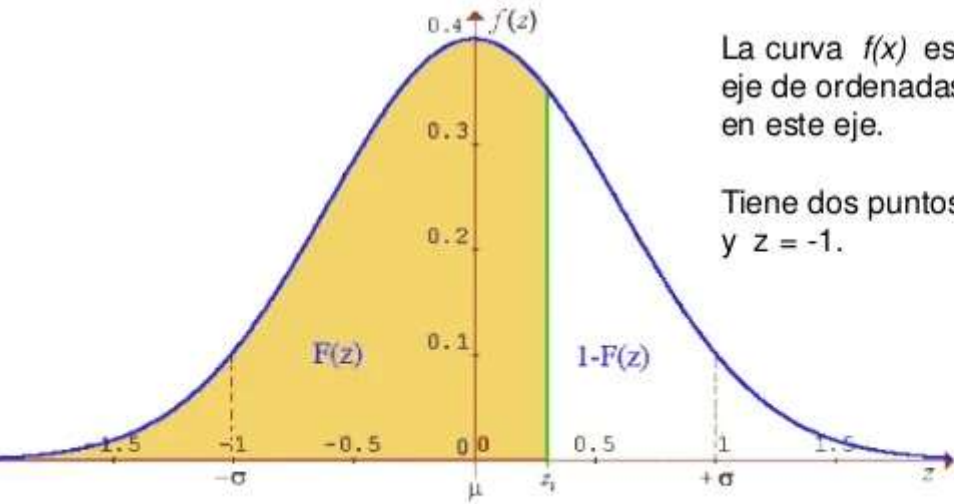
Característica de la distribución normal tipificada (reducida o estándar):

No depende de ningún parámetro.

Su media es 0, su varianza es 1 y su desviación típica es 1.

La curva $f(x)$ es simétrica respecto al eje de ordenadas y tiene un máximo en este eje.

Tiene dos puntos de inflexión en $z = 1$ y $z = -1$.



16

▼ 4. Cargar una tabla en $L^A T^E X$, ejemplos de plantillas online:

\$\$

| | | |
|-----|-----|-----|
| foo | bar | baz |
| 1 | 2 | 3 |

| <i>column1</i> | <i>column2</i> | <i>column3</i> | <i>column4</i> |
|----------------|----------------|----------------|----------------|
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |

| <i>column1</i> | <i>column2</i> | <i>column3</i> | <i>column4</i> |
|----------------|----------------|----------------|----------------|
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |

| <i>column1</i> | <i>column2</i> | <i>column3</i> | <i>column4</i> |
|----------------|----------------|----------------|----------------|
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |

| <i>column1</i> | <i>column2</i> | <i>column3</i> | <i>column4</i> |
|----------------|----------------|----------------|----------------|
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |
| <i>foo</i> | <i>bar</i> | <i>foo</i> | <i>bar</i> |

\$\$

Se construyo una tabla con algunas medidas empleadas en Estadística:

| <i>Nombre</i> | <i>Fórmula</i> |
|----------------------------|---|
| <i>Desviación estándar</i> | $\sigma = \sqrt{\frac{1}{n} \times \sum_{i=1}^N (X_i - \bar{X})^2}$ |
| <i>Varianza</i> | $\sigma^2 = \frac{1}{n} \times \sum_{i=1}^N (X_i - \bar{X})^2$ |
| <i>Desviación media</i> | $D_m = \frac{1}{n} \times \sum_{i=1}^N X_i - \bar{X} $ |
| <i>Mediana</i> | $X \times \frac{n+1}{2}$ si N impar $\frac{1}{2} \times (X_{\frac{N}{2}} + X(\frac{N}{2} + 1))$ si N par |
| <i>Moda</i> | <i>Datos más repitidos</i> |
| <i>Media</i> | $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ |

4. Subir una ecuación en *L^AT_EX*

$\widehat{Y_i} = a + b \cdot X_i$

Con :

$$a = \frac{\sum X_i^2 \cdot \sum Y_i - \sum X_i \cdot \sum X_i \cdot Y_i}{n \cdot \sum X_i^2 - (\sum X_i)^2}$$

Pendiente :

$$b = \frac{n \cdot \sum X_i \cdot Y_i - \sum X_i \cdot \sum Y_i}{n \cdot \sum X_i^2 - (\sum X_i)^2}$$