

# Percentiles



**Macarena Villamea**

Tech Lead Data Science

---

Master en Data Science  
2022-2023

# INDEX

- 1** Introduction to percentile
- 2** Different ways to calculate percentiles
- 3** Calculate percentiles in Python
- 4** Quartiles

# Percentiles

In statistics, a **k-th percentile** (percentile score or centile) is a score below which a given percentage k of scores in its frequency distribution falls (exclusive definition) or a score at or below which a given percentage falls (inclusive definition).

El **percentil 25** es el valor bajo el cual se encuentran el 25% de las observaciones.



## PERCENTILES

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

## PERCENTILES

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

**n=10**



## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51  
**20%**

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

10 datos

**n=10**

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**  
**20%**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

**n=10**

**q=20%**

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**  
**20%**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=$$

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**  
**20%**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1$$

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**  
**20%**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

**n=10**

**q=20%**

$$(n-1)q+1=(10-1)0.2+1=2.8$$

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**  
**20%**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

**n=10**

**q=20%**

$$(n-1)q+1=(10-1)0.2+1=2.8$$

Parte

decimal = 0

20-percentil es el dato  
que está en esta  
posición

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**  
**20%**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

**n=10**

**q=20%**

$$(n-1)q+1=(10-1)0.2+1=2.8$$

Parte decimal  
distinta de 0

¡Hay que hacer un poco  
más de cuentas!

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**  
**20%**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**10 datos**

**n=10**

**q=20%**

$$(n-1)q+1=(10-1)0.2+1=2.8$$

Entre **2** y **3**



## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**  
i j

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} \begin{array}{l} \text{fraction} \\ (n-1)q+1=(10-1)0.2+1=2.8 \end{array}$$

## PERCENTILES

**LINEAR INTERPOLATION** → obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**i**   **j**

20-percentil=

=**i**+(**j-i**)**fraction**

=**23**+(**33-23**)**0.8**

=**31**

**n=10**

**q=20%**

$$(n-1)q+1=(10-1)0.2+1=2.8$$

**fraction**

## PERCENTILES

### MIDPOINT INTERPOLATION

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

## PERCENTILES

### MIDPOINT INTERPOLATION

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8$$

## PERCENTILES

### MIDPOINT INTERPOLATION

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

**i**   **j**

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8$$

Entre **2** y **3**

## PERCENTILES

### MIDPOINT INTERPOLATION

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

i j

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8$$

$$\begin{aligned} 20\text{-percentil} &= \\ &= \frac{(j+i)}{2} \\ &= \frac{(33+23)}{2} \\ &= 28 \end{aligned}$$

## PERCENTILES

### NEAREST INTERPOLATION

Calculemos el **20-percentil** de **67, 84, 61, 78, 12, 23, 45, 74, 33, 51**

**12, 23, 33, 45, 51, 61, 67, 74, 78, 84**

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8$$

## PERCENTILES

### NEAREST INTERPOLATION

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8 \rightarrow \text{redondea a 3}$$



## PERCENTILES

### NEAREST INTERPOLATION

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8 \rightarrow \text{redondea a 3}$$

## PERCENTILES

### NEAREST INTERPOLATION

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8 \rightarrow \text{redondea a 3}$$

## PERCENTILES

### HIGHEST INTERPOLATION

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8 \rightarrow \text{entre 2 y 3}$$

## PERCENTILES

### LOWEST INTERPOLATION

Calculemos el **20-percentil** de 67, 84, 61, 78, 12, 23, 45, 74, 33, 51

12, 23, 33, 45, 51, 61, 67, 74, 78, 84

$$\left. \begin{array}{l} n=10 \\ q=20\% \end{array} \right\} (n-1)q+1=(10-1)0.2+1=2.8 \rightarrow \text{entre 2 y 3}$$



# Percentiles

**`DataFrame.quantile(q=0.5, axis=0, numeric_only=True, interpolation='linear')`**

- q: % del percentil
- axis: 0 para filas, 1 para columnas
- numeric\_only: si es True no se computan datetimes y timedeltas
- interpolation: tipo de interpolación.
  - linear: lineal
  - lower: al menor
  - higher: al mayor
  - nearest: al más cercano
  - midpoint: punto medio



# Percentiles

```
import pandas as pd
```

```
data = pd.read_csv('data_percentiles.csv')  
data
```

	a
0	12
1	23
2	33
3	45
4	51
5	61
6	67
7	74
8	78
9	84

```
# linear  
data.quantile(0.2, interpolation='linear')
```

```
a    31.0  
Name: 0.2, dtype: float64
```

```
# lower  
data.quantile(0.2, interpolation='lower')
```

```
a    23  
Name: 0.2, dtype: int64
```

```
# higher  
data.quantile(0.2, interpolation='higher')
```

```
a    33  
Name: 0.2, dtype: int64
```

```
# nearest  
df.quantile(0.2, interpolation='nearest')
```

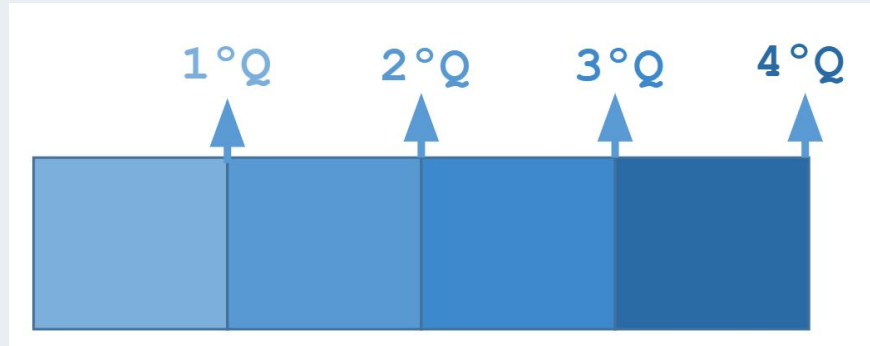
```
a    33  
Name: 0.2, dtype: int64
```

```
# midpoint  
df.quantile(0.2, interpolation='midpoint')
```

```
a    28.0  
Name: 0.2, dtype: float64
```

## QUARTILES

- **Primer cuartil:** 25-percentil
- **Segundo cuartil:** 50-percentil
  - Coincide con la **mediana**
- **Tercer cuartil:** 75-percentil
- **Cuarto cuartil:** 100-percentil
  - Es el **máximo** valor



## SUMMARY

### Today we've learnt:

- Percentiles: definition
- Calculate percentiles by
  - linear interpolation
  - midpoint interpolation
  - nearest interpolation
  - highest interpolation
  - lowest interpolation
- Calculate percentiles in Python
- Quartiles: definition





