Two simple Definitions of Percentiles

There is no single universally accepted definition of a percentile. Different methods can yield different results, particularly in small datasets

What is a percentile?

A percentile represents the percentage of scores in a dataset that fall below a given value. It helps us understand how an individual score compares to the rest of the dataset

for eg

- If your shyness score is 35 out of 50, the number doesn't tell you much

However, if you know that 65% of people scored lower than you, are in the 65th percentile - meaning you are shyer than 65% of the population

- Definition 1 of percentiles
  The 65th percentile is the lowest score that is greater than 65% of the score

- Definition 2 of percentiles
  The 65th percentile is the smallest score that is greater than or equal to 65% of the scores

While these definitions are simple, they can lead to significantly different results, especially with small datasets

The challenge of Rounding
For example, if there are 50 total scores;
65% of 50 = 32.5
But since we can't have half a score, how do we determine the 65th percentile

- Third Definition of Percentile (Interpolation Method)
  When referring to a percentile, we use this third definition, which is based on interpolation. This method ensures smoother rounding and better accuracy compared to the first two definitions.

Step by step method to compute percentiles

(1) Calculate the Rank (R)

$$R = \frac{P}{100} \times (N+1)$$

where P is desired percentile, N is the total number of values in the dataset.

(2) Break R into Integer (IR) and Fraction (FR) Parts

IR = Integer Portion of R

FR = Decimal Portion of R

(3) Find the Two Closest values

Locate the value at Rank IR

Locate the value at Rank IR + 1

(4) Interpolate (If needed)

If FR = 0 the percentile is the value at Rank IR

P^{th} percentile = Value at IR + (FR × (value at IR+1 - Value at IR))

Examples

Eg 1 25th percentile for 8 Number

Number	Rank
3	1
5	2
7	3
8	4
9	5
11	6
13	7
15	8

Step1 Compute R i.e. $R = \frac{P}{100} \times (N+1)$

$= (25/100) \times (8+1) = 2.25$

IR = 2 FR = 0.25

Step2 : Find values

value at Rank 2 = 5

Value at Rank 3 = 7

Step 3 Apply interpolation

$\{5 + (0.25) \times (7-5)\} = 5.5$

25^{th} percentile = 5.5

Special case: Exact Rank Match

For even-numbered datasets, the 50^{th} percentile (Median) might land exactly on a data point

Eg2: 50^{th} percentile for 5 number

Number	Rank
2	1
3	2
5	3
9	4
11	5

Rank = $P/100\ (N+1)$

$= (50/100) \times (6)$

$= 3$

Since IR = 3, FR = 0 . the 50^{th} Percentile = 5

Date / /
Page No.

Pearson Correlation Coefficient (r) - Explained Simply

What is pearson Correlation?

⇒ Pearson correlation (denoted as r for samples and ρ for populations) measures the strength and direction of a linear relationship between two variables.

- If $r = 1$, there is a perfect positive linear relationship (As one variable increases, the other increases proport
- If $r = -1$, there is a perfect negative linear relationship. (As one variable increases, the other decreases proportion
- If $r = 0$, no linear relationship exists between the two variables.

Real World Examples

- Perfect Positive Correlation ($r = 1$)
  The relationship between Celsius and Farenheit temperatures (Exact, direct increase)
- Perfect Negative Correlation ($r = -1$)
  The amount of fuel in a cars tank vs distance driven (as one increases, the other decreases)
- No Correlation
  Shoe size and intellegence (no relationship)

Date / /
Page No.

Key properties of Pearson's Correlation (r) - Simplified

(1) Range of Pearson's r
- r always falls between -1 and 1
  - r = 1 → perfect positive linear relationship
  - r = -1 → perfect negative linear relationship
  - r = 0 → no linear relationship

(2) Symmetry Property
- The correlation of X with Y is the same as the correlation of Y with X.
  - eg The correlation between weight and Height is the same as between Height and weight

(3) Effect of Linear Transformations
Pearson's r doesnot change if you:
- Multiply a variable by a constant (eg converting inches to feet)
- Add or subtract a constant (eg adding 5 points to all exam scores)

Takeaway: Pearson's r only captures linear relationships and is independent of unit changes or shifts in scale!