

Quantili e percentili

Oltre alla mediana, che divide a metà un insieme di dati ordinati, si possono definire altri indici di posizione, detti **quantili** e **percentili**, che dividono l'insieme di dati ordinati in un dato numero di parti uguali. Questi **indici di posizione non centrale** sono usati soprattutto per ampi insiemi di dati. I **quartili** sono un caso particolare dei quantili, e si ottengono dividendo l'insieme di dati ordinati in quattro parti uguali.

Definizione 1

Il **primo quartile** Q_1 è un valore tale che il 25 % dei dati ordinati è minore o uguale a Q_1 . Il primo quartile Q_1 è detto anche **25-esimo percentile** e indicato con $P_{0.25}$.

Il **terzo quartile** Q_3 è un valore tale che il 75 % dei dati ordinati è minore o uguale a Q_3 ed è detto anche **75-esimo percentile** e indicato con $P_{0.75}$.

Il secondo quartile Q_2 (50-esimo percentile) coincide con la mediana.

Per calcolare i quartili si segue una regola simile a quella usata per il calcolo della mediana.

Regola per il calcolo dei quartili

- 1 – Si ordinano gli n dati assegnati in ordine crescente;
- 2 – si calcola il prodotto $k = np$, dove $p = 0.25$ per il primo quartile e $p = 0.75$ per il terzo quartile;
- 3 – se k è un intero, il quartile si ottiene facendo la media del k -esimo e del $(k+1)$ -esimo valore dei dati ordinati;
- 4 – se k non è intero, si arrotonda k per eccesso al primo intero successivo e si sceglie come quartile il corrispondente valore dei dati ordinati.

La regola può essere generalizzata in modo semplice per trovare un qualsiasi altro percentile.

Ad esempio per trovare il 95-esimo percentile, ossia quel valore tale che il 95% dei dati ordinati è minore o uguale ad esso, si usa la stessa regola, con $p = 0.95$.

Molti software calcolano i percentili con una regola un po' più complessa, basata sull'interpolazione lineare fra dati adiacenti, perciò i valori trovati possono differire leggermente da quelli ricavati con la regola più semplice qui indicata.

Esempio 1

Calcolare il primo e il terzo quartile dell'insieme di dati

32.2 32.0 30.4 31.0 31.2 31.3 30.3 29.6 30.5 30.7

Dati ordinati

29.6 30.3 30.4 30.5 30.7 31.0 31.2 31.3 32.0 32.2

Primo quartile

$$n = 10 \quad p = 0.25 \quad k = np = 2.5$$

k non è intero, perciò si arrotonda per eccesso $k = 3$: il primo quartile è il terzo dei dati ordinati

$$Q_1 = 30.4.$$

Terzo quartile

$$n = 10 \quad p = 0.75 \quad k = np = 7.5$$

k non è intero, perciò si arrotonda per eccesso $k = 8$: il terzo quartile è l'ottavo dei dati ordinati

$$Q_3 = 31.3.$$

Secondo quartile (mediana)

$$n = 10 \quad p = 0.5 \quad k = np = 5$$

k è intero, perciò si fa la media tra il quinto e il sesto dato e si ottiene

$$Q_2 = \frac{30.7 + 31.0}{2} = 30.85$$

(Questo valore coincide con quello che si trova con la regola della mediana).

Esempio 2

Calcolare il primo e il terzo quartile e il 95-esimo percentile per i dati della seguente tabella, che sono il risultato di 80 determinazioni, in una data unità di misura, dell'emissione giornaliera di un gas inquinante da un impianto industriale:

15.8	26.4	17.3	11.2	23.9	24.8	18.7	13.9	9.0	13.2
22.7	9.8	6.2	14.7	17.5	26.1	12.8	28.6	17.6	23.7
26.8	22.7	18.0	20.5	11.0	20.9	15.5	19.4	16.7	10.7
19.1	15.2	22.9	26.6	20.4	21.4	19.2	21.6	16.9	19.0
18.5	23.0	24.6	20.1	16.2	18.0	7.7	13.5	23.5	14.5
14.4	29.6	19.4	17.0	20.8	24.3	22.5	24.6	18.4	18.1
8.3	21.9	12.3	22.3	13.3	11.8	19.3	20.0	25.7	31.8
25.9	10.5	15.9	27.5	18.1	17.9	9.4	24.1	20.1	28.5

Dati ordinati

6.2	7.7	8.3	9.0	9.4	9.8	10.5	10.7	11.0	11.2
11.8	12.3	12.8	13.2	13.3	13.5	13.9	14.4	14.5	14.7
15.2	15.5	15.8	15.9	16.2	16.7	16.9	17.0	17.3	17.5
17.6	17.9	18.0	18.0	18.1	18.1	18.4	18.5	18.7	19.0
19.1	19.2	19.3	19.4	19.4	20.0	20.1	20.1	20.4	20.5
20.8	20.9	21.4	21.6	21.9	22.3	22.5	22.7	22.7	22.9
23.0	23.5	23.7	23.9	24.1	24.3	24.6	24.6	24.8	25.7
25.9	26.1	26.4	26.6	26.8	27.5	28.5	28.6	29.6	31.8

Primo quartile

$$n = 80 \quad p = 0.25 \quad k = np = 20$$

k è intero, perciò si fa la media tra il 20-esimo e il 21-esimo dato e si ottiene

$$Q_1 = \frac{14.7 + 15.2}{2} = 14.95$$

Terzo quartile

$$n = 80 \quad p = 0.75 \quad k = np = 60$$

k è intero, perciò si fa la media tra il 60-esimo e il 61-esimo dato e si ottiene

$$Q_3 = \frac{22.9 + 23.0}{2} = 22.95$$

95-esimo percentile

$$n = 80 \quad p = 0.95 \quad k = np = 76$$

k è intero, perciò si fa la media tra il 76-esimo e il 77-esimo dato e si ottiene

$$P_{0.95} = \frac{27.5 + 28.5}{2} = 28.0$$

Il 95-esimo percentile fornisce un'importante informazione: soltanto il 5% dei dati sono maggiori di 28.0, ossia, con riferimento al tipo di dati descritti nell'esempio 2, soltanto nel 5% dei giorni l'emissione di gas inquinanti supera la soglia di 28.0 unità.