

4. Statistiche elementari non parametriche

Le statistiche elementari non parametriche comprendono in genere, oltre alla *mediana*, il *valore minimo* osservato, il *valore massimo* osservato, il *range* (cioè la differenza tra valore massimo e valore minimo) e spesso i *quartili*. In base al principio elementare per cui un segmento di retta può essere suddiviso in quattro parti mediante tre tacche equidistanti tra di loro e dagli estremi del segmento, i quartili sono tre. Il primo quartile è il valore che lascia alla sua sinistra il 25% dei dati (e che lascia alla sua destra il 75% dei dati). Il secondo quartile è il valore che lascia alla sua sinistra il 50% dei dati (e che lascia alla sua destra il 50% dei dati): il secondo quartile è quindi la mediana. Il terzo quartile è il valore che lascia alla sua sinistra il 75% dei dati (e che lascia alla sua destra il 25% dei dati).

Anche nel caso delle statistiche non parametriche, fatto uguale 100 il numero di dati di una distribuzione, il percentile è il valore che lascia alla sua sinistra la percentuale di dati desiderata (e corrispondentemente alla destra il complemento a 100). Così ad esempio il 2,5-esimo percentile è il valore che lascia alla sua sinistra il 2,5% dei dati osservati (e corrispondentemente lascia alla sua destra il 97,5% dei dati osservati), il 50-esimo percentile (che nel caso delle statistiche non parametriche è la mediana) è il valore che lascia alla sua sinistra il 50% dei dati osservati (e corrispondentemente lascia alla sua destra il 50% dei dati osservati). Si noti che il primo quartile corrisponde al 25-esimo percentile non parametrico, il secondo quartile corrisponde al 50-esimo percentile non parametrico (cioè alla mediana), il terzo quartile corrisponde al 75-esimo percentile non parametrico.

4.1. Calcolo del percentile corrispondente a un dato

Per calcolare il percentile corrispondente a un dato, percentile che indica la posizione del dato all'interno della distribuzione osservata:

→ calcolare il rango del dato. Il rango è il numero di posizione dei singoli dati nella lista dei dati ordinati in ordine numerico crescente. Quando due dati sono uguali, a ciascuno viene assegnato come numero di posizione la media dei numeri di posizione dei dati che risultano uguali. Così, per esempio, se il quinto e il sesto dato, in una lista ordinata, sono tutti e due uguali, e pari a 223, il rango del quinto e del sesto dato sarà uguale, e pari a 5,5. Il procedimento viene esteso anche al caso in cui più di due dati siano uguali;

→ essendo n il numero dei dati, il percentile P (p -esimo) non parametrico, che indica la posizione del dato all'interno della distribuzione osservata, viene calcolato come

$$P = 100 \cdot \text{rango} / (n + 1)$$

4.2. Calcolo del dato corrispondente a un percentile

In un insieme di n dati, il P -esimo percentile non parametrico corrisponde al dato il cui numero C (numero progressivo, non necessariamente intero, nella serie dei dati ordinati in ordine numerico crescente) è

$$C = P \cdot (n + 1) / 100$$

Qualora C non sia un intero (accade quasi sempre) viene effettuata una interpolazione lineare fra il valore corrispondente all'intero che immediatamente precede C e il valore corrispondente all'intero che immediatamente segue C .