QUANTILI

Date n osservazioni relative alla variabile quantitativa X si definisce **quantile di ordine** p (con $0 < p < 1$), e si indica con x_p, quel particolare valore della variabile per cui la proporzione di osservazioni inferiori o uguali a x_p è almeno p (e quindi la proporzione di osservazioni maggiore o uguale a x_p è almeno 1-p).

Se, per esempio, per una distribuzione di redditi annui è noto che il quantile di ordine 0.25 è pari a 5 mila euro, questo vuol dire che almeno un quarto delle unità considerate ha un reddito annuo inferiore o uguale a 5 mila euro; se in una distribuzione di stature $x_{0,8}$ vale 178 centimetri, significa che almeno l'80% delle unità ha una statura inferiore o uguale a 178 centimetri.

La determinazione del quantile di ordine p viene effettuata tenendo presente la posizione occupata dagli n valori della variabile, che quindi devono essere ordinati in modo non decrescente.

Nel caso in cui i dati siano costituiti da una sequenza di n valori sarà quindi necessario innanzitutto elencare i suoi elementi in modo da ottenere una sequenza ordinata in modo non decrescente, il cui generico elemento verrà in seguito indicato mediante il simbolo $x_{(i)}$, con $x_{(i-1)} \leq x_{(i)} \leq x_{(i+1)}$ per $i = 2, 3, \ldots, n-1$.

Data una sequenza di n osservazioni relative a una variabile quantitativa X, il quantile x_p di ordine p è l'osservazione che nella sequenza ordinata occupa il posto corrispondente alla parte intera superiore di np, indicato con $\lceil np \rceil$

Il simbolo $\lceil np \rceil$ rappresenta quindi la parte intera superiore di np per cui

$$np \leq \lceil np \rceil < np+1$$

Esempio 1

Considerata la seguente sequenza di osservazioni rilevate su 5 piantine di cui si è misurata l'altezza ottenendo i seguenti valori in centimetri

23.4	18.2	21.0	22.7	19.1

si determini il quantile di ordine $p = 0.5$ della variabile.

In questo caso $n = 5$, per cui

$$\lceil np \rceil = \lceil 5 \times 0.5 \rceil = \lceil 2.5 \rceil = 3.$$

Il quantile di ordine 0.5, $x_{0.5}$, occupa quindi il terzo posto nella sequenza ordinata

18.2	19.1	21.0	22.7	23.4

per cui risulta

$$x_{0.5} = x_{(3)} = 21.$$

Esempio 2
Si supponga ora che alle unità dell'esempio precedente si sia aggiunta una nuova piantina, la cui altezza in centimetri è pari a 21.5. Si vuole determinare il quantile di ordine p = 0.5 delle 6 unità statistiche.
In questo caso la sequenza ordinata è la seguente

18.2 19.1 21.0 21.5 22.7 23.4

per cui, dato che n = 6,

$$\lceil np \rceil = \lceil 6 \times 0.5 \rceil = \lceil 3 \rceil = 3.$$

Il quantile di ordine 0.5 occupa quindi il terzo posto nella sequenza ordinata per cui risulta ancora
$x_{0.5} = x_{(3)} = 21.0$.

Considerata una generica sequenza, tutti i suoi termini sono evidentemente quantili di un certo ordine della variabile X, ma fra questi alcuni sembrano più indicativi di altri perché fanno riferimento a quei valori di p che corrispondono a quelli di uso più comune e sono considerati, quindi, come altrettanti valori caratteristici della X.

Sono frequentemente utilizzati i quantili $x_{0.25}$, $x_{0.5}$ e $x_{0.75}$ che, per la loro importanza nella descrizione delle caratteristiche della variabile X, hanno un nome particolare e vengono detti **quartili**.
Il quartile più importante è il secondo, $x_{0.5}$, comunemente chiamato **mediana**, perché indica quel particolare valore della variabile tale che almeno la metà delle unità presenta un valore minore o uguale a $x_{0.5}$ e almeno la metà delle unità presenta un valore maggiore o uguale a $x_{0.5}$.

Per esempio l'ISTAT fa espressamente riferimento al "reddito mediano" in numerosi studi volti a confrontare la distribuzione del reddito in Italia rispetto a quella di altri Paesi europei, oppure per analizzare condizioni di vita e livello di povertà delle famiglie italiane nei diversi anni.

Altri quantili di uso frequente sono i nove **decili** $x_{0.1}$, $x_{0.2}$, ..., $x_{0.9}$ e i novantanove **centili** $x_{0.01}$, $x_{0.02}$, ..., $x_{0.99}$. In questi casi la mediana corrisponde al quinto decile o al cinquantesimo centile.

Esempio 3
Considerata la seguente sequenza di n=10 valori di una variabile quantitativa continua X

3.20 3.25 2.80 2.96 3.00 3.18 3.12 1.87 1.99 2.02

si determinino i suoi tre quartili.

La sequenza ordinata è

1.87 1.99 2.02 2.80 2.96 3.00 3.12 3.18 3.20 3.25

per cui il primo quartile occupa il posto $\lceil np \rceil = \lceil 10 \times 0.25 \rceil = \lceil 2.5 \rceil = 3$, la mediana il posto $\lceil np \rceil = \lceil 10 \times 0.5 \rceil = \lceil 5 \rceil = 5$ e il terzo quartile il posto $\lceil np \rceil = \lceil 10 \times 0.75 \rceil = \lceil 7.5 \rceil = 8$.

Risulta quindi

$x_{0.25} = 2.02$
$x_{0.5} = 2.96$
$x_{0.75} = 3.18$

Esempio 4
Considerata la seguente sequenza di n=14 valori di una variabile quantitativa discreta X

3 1 5 -2 -3 -5 0 10 -9 20 12 6 -21 8

si determinino i suoi tre quartili.

La sequenza ordinata è

-21 -9 -5 -3 -2 0 1 3 5 6 8 10 12 20

per cui il primo quartile occupa il posto $\lceil np \rceil = \lceil 14 \times 0.25 \rceil = \lceil 3.5 \rceil = 4$, la mediana il posto $\lceil np \rceil = \lceil 14 \times 0.5 \rceil = \lceil 7 \rceil = 7$ e il terzo quartile il posto $\lceil np \rceil = \lceil 14 \times 0.75 \rceil = \lceil 10.5 \rceil = 11$. Risulta quindi

$x_{0.25} = -3$
$x_{0.5} = 1$
$x_{0.75} = 8$

Se nella sequenza degli n valori alcuni di questi si presentano più di una volta, l'addensamento delle frequenze in corrispondenza di alcune determinazioni fa sì che uno stesso valore della variabile possa corrispondere a più quantili di ordine diverso.
Se, per esempio, si considera la seguente sequenza ordinata dei voti in statistica ottenuti da 20 studenti

18 18 18 20 20 22 22 23 23 23 23 24 24 24 25 26 27 27 28 30

i tre quartili occupano rispettivamente i posti $\lceil 20 \times 0.25 \rceil = 5$, $\lceil 20 \times 0.5 \rceil = 10$ e $\lceil 20 \times 0.75 \rceil = 15$ e sono quindi dati da $x_{0.25} = 20$, $x_{0.5} = 23$ e $x_{0.75} = 25$. Il valore 20 corrisponde però anche al secondo decile, così come il valore 23 è sia mediana, sia quarto e sesto decile.
Alle medesime conclusioni si giunge ovviamente se si fa riferimento, invece che alla sequenza, alla distribuzione di frequenza corrispondente, per cui il calcolo dei quantili si effettua sempre nello stesso modo. Nel caso di una distribuzione di frequenza conviene però utilizzare i valori delle frequenze assolute cumulate.

In una distribuzione di frequenza il quantile di ordine p della variabile quantitativa X corrisponde alla determinazione c_j a cui è associata la prima frequenza assoluta cumulata N_j maggiore o uguale a $\lceil np \rceil$.

Data per esempio la distribuzione dei voti considerati in precedenza

Distribuzione dei voti

x	Frequenza assoluta	Frequenza assoluta cumulata
18	3	3
20	2	5
22	2	7
23	4	11
24	3	14
25	1	15
26	1	16
27	2	18
28	1	19
30	1	20
	20	

dalle frequenze assolute cumulate contenute nell'ultima colonna della tabella precedente si ottengono gli stessi quartili ottenuti sulla sequenza dei voti.

Esempio 5

Data la seguente distribuzione di frequenza relativa a una variabile quantitativa discreta X

Distribuzione di una variabile X

x	Frequenza assoluta
-2	4
-1	6
0	10
1	10
2	12
3	8
	50

si determini il valore del secondo, del quinto e del settimo decile.

Le unità esaminate sono 50, per cui i posti occupati dai tre decili richiesti sono rispettivamente il decimo, il venticinquesimo e il trentacinquesimo. Dalle frequenze assolute cumulate riportate nella tabella seguente

Distribuzione di una variabile X

x	Frequenza assoluta	Frequenza assoluta cumulata
-2	4	4
-1	6	10
0	10	20
1	10	30
2	12	42
3	8	50
	50	

risulta $x_{0.2} = -1$, $x_{0.5} = 1$ e $x_{0.7} = 2$.