APPROFONDIMENTI – SCHEDA 2

1. Diverse definizioni dei quantili

La funzione di distribuzione cumulata per dati sperimentali è una funzione a scalini, non è quindi invertibile. Questo comporta che non esiste un modo univoco per "invertirla".
Il software statistico SAS presenta 5 diverse definizioni.
Scriviamo un numero come la somma fra la sua parte intera e la sua parte decimale

$y = u + d$ con u parte intera di y e d parte decimale

ad esempio 8.4 = 8 + 0.4.
Indichiamo con $x_{(k)}$ il k-esimo dato sperimentale ordinato.
Sia n il numero di unità sperimentali e α un numero fra 0 e 1 che rappresenta il quantile.

Def. 1 di SAS:
Sia $n\alpha = u + d$
$q_\alpha = x_{(u)} + d\,(x_{(u+1)} - x_{(u)}) = (1 - d)\,x_{(u)} + d\,x_{(u+1)}$
cioè una media pesata fra $x_{(u)}$ e $x_{(u+1)}$ con peso la parte decimale
Se u = 0 si sceglie $x_{(1)}$

Def. 2 di SAS:
$q_\alpha =$ l'osservazione più vicina a $n\alpha$

Def. 3 di SAS:
$q_\alpha = x_{(u)}$ se d = 0 pseudoinversa della funzione di distribuzione cumulata
$q_\alpha = x_{(u+1)}$ se d > 0

Def. 4 di SAS:
Sia $(n+1)\alpha = u + d$
$x_{(u)} + d\,(x_{(u+1)} - x_{(u)}) = (1 - d)\,x_{(u)} + d\,x_{(u+1)}$
cioè una media pesata fra $x_{(u)}$ e $x_{(u+1)}$ con peso la parte decimale
Se u = n+1 si sceglie $x_{(n)}$

Def. 5 di SAS:
$q_\alpha = (x_{(u)} + d\,x_{(u+1)})/2$ se d = 0 pseudoinversa della funzione di distribuzione cumulata mediata
$q_\alpha = x_{(u+1)}$ se d > 0

Il SAS ha come default la definizione 5.

La definizione 3 è quella più corretta dal punto di vista matematico, è facile da calcolare a mano e ha la proprietà che il valore di ciascun quantile è un valore assunto dai dati.

Minitab usa la definizione 4.

Osserviamo che in nessun caso viene calcolata l'interpolazione lineare fra il valore della cumulata in $F(x_{(u)})$ e $F(x_{(u+1)})$ che comporta:

$$q_\alpha = x_{(u)} + \frac{\alpha - F(x_{(u)})}{F(x_{(u+1)}) - F(x_{(u)})}\left(x_{(u+1)} - x_{(u)}\right)$$

Le interpolazioni fatte sa SAS riguardano solo le ascisse.