

Analisi Dati con Excel

Giovanni Della Lunga

giovanni.dellalunga@gmail.com

La prima regola di ogni tecnologia è che l'automazione applicata ad un'operazione efficiente ne aumenterà l'efficienza. La seconda è che l'automazione applicata ad un'operazione inefficiente ne aumenterà l'inefficienza.

Bill Gates

Funzioni Statistiche

Statistica Descrittiva con Excel



| Funzioni Statistiche

- » Excel fornisce una vasta gamma di funzioni Statistiche, che vanno dai calcoli di base, come la media, la mediana e la moda, a quelli più complessi come i test di distribuzione e di probabilità statistica.
- » Fra queste vi sono una serie di funzioni per il calcolo dei momenti che vale la pena descrivere in maggior dettaglio.
- » In pratica vengono comunemente utilizzati i primi quattro momenti:
 - media
 - varianza
 - skewness (o asimmetria)
 - curtosi

| Funzioni Statistiche: Media, Moda e Mediana

- » La funzione **MEDIA()** viene utilizzata per calcolare la media aritmetica di un insieme di valori.
- » L'argomento di questa funzione deve indicare il blocco di celle che contiene i valori di cui calcolare il valor medio.
- » Notate che le celle contenenti il valore 0 (zero) saranno incluse nel calcolo della media mentre le celle vuote (senza alcun dato) restano escluse da questo calcolo.
- » Esempio: Se hai i dati nelle celle da A1 a A10, puoi calcolare la media con:

=MEDIA(A1:A10)

Funzioni Statistiche: Media, Moda e Mediana

- » La **MODA()**, invece, è **il valore che appare con maggiore frequenza** in un insieme di dati.
- » Può essere particolarmente utile in insiemi di dati dove ci sono valori che si ripetono spesso, fornendo un'idea di quale valore è più comune.
- » In alcuni insiemi di dati, può esserci più di una moda (o nessuna).

Nota Importante: Se i dati non contengono duplicati, la funzione MODA restituirà un errore #N/A. Per evitare errori, potete utilizzare la funzione SE.ERRORE per gestire i casi in cui non ci sia una moda:

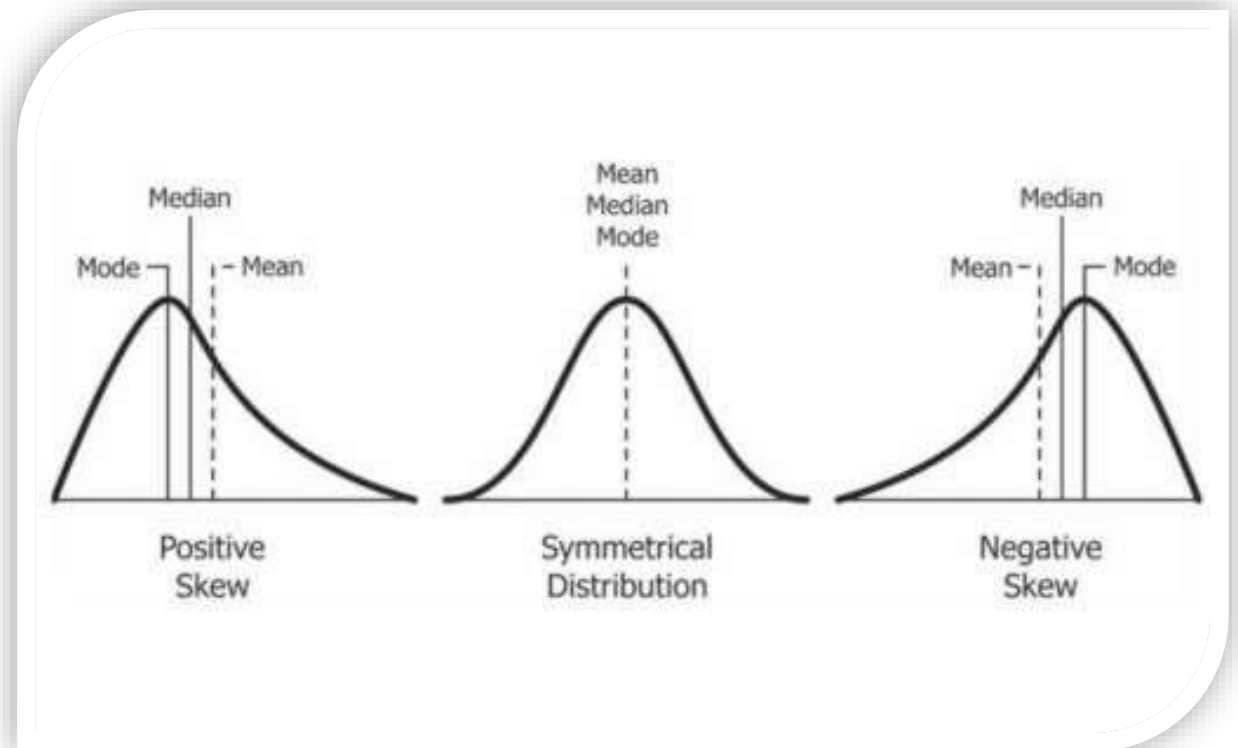
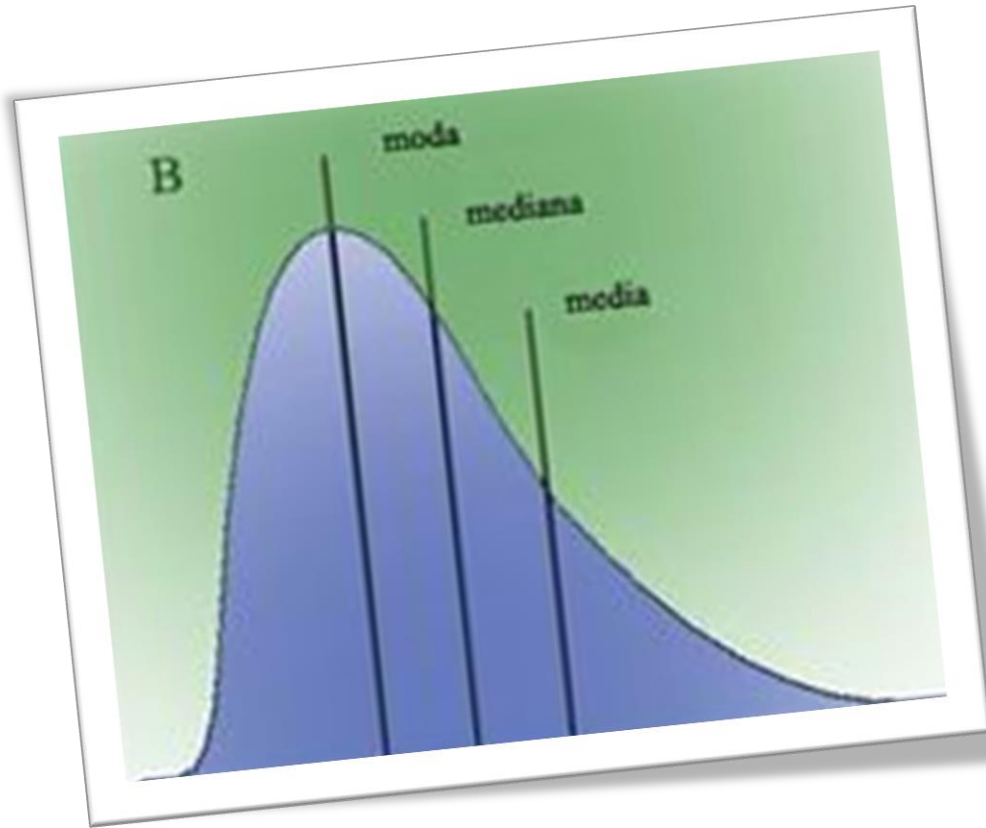
`=SE.ERRORE(MODA(A1:A10), "Nessuna moda")`

| Funzioni Statistiche: Media, Moda e Mediana

- » La **MEDIANA()** è il valore che si trova esattamente a metà dell'insieme di dati quando questi sono ordinati in ordine crescente o decrescente.
- » Se il numero di valori è dispari, la mediana è il valore centrale; se è pari, è la media dei due valori centrali.
- » Esempio: Per trovare la mediana dei dati nelle celle da A1 a A10:

=MEDIANA(A1:A10)

Funzioni Statistiche: Media, Moda e Mediana



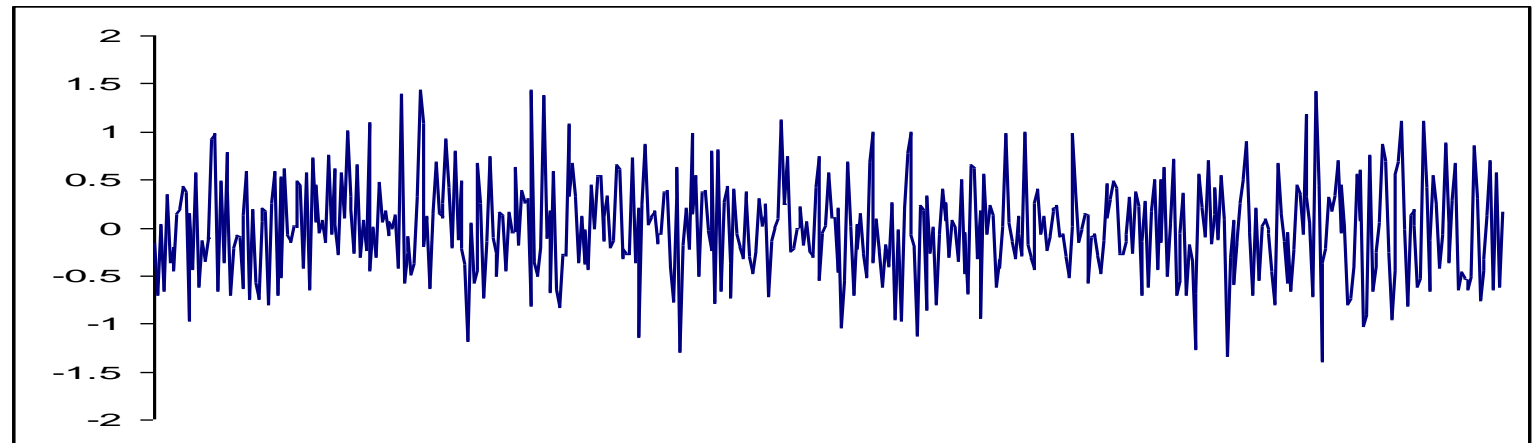
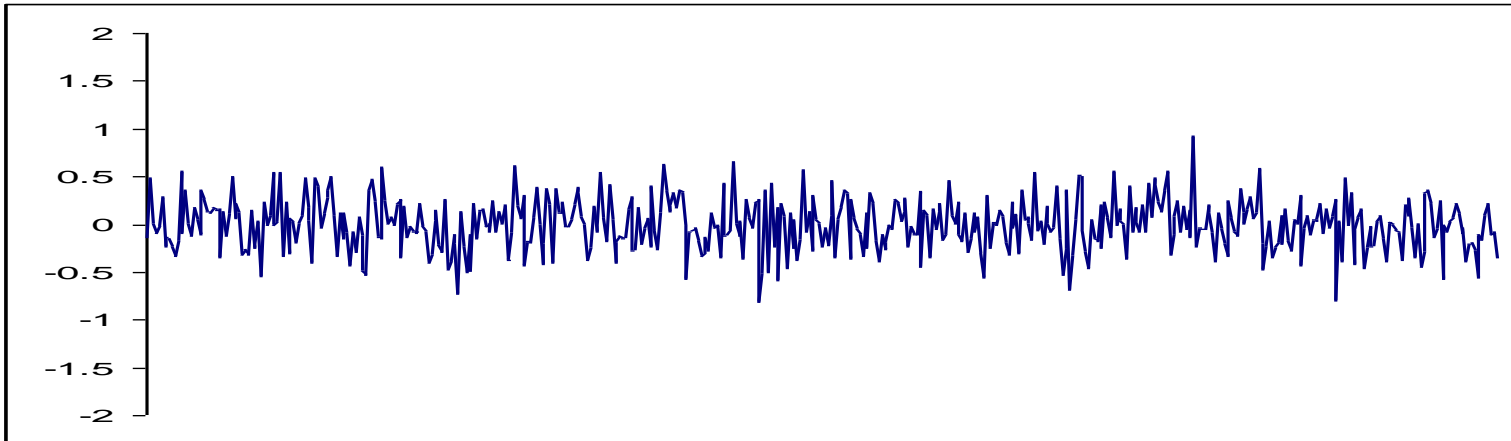
Funzioni Statistiche: Varianza

- » La varianza può essere calcolata tramite la funzione VAR(). Anche questa funzione prende come argomento il blocco di celle contenente i dati di cui si desidera stimare la varianza.
- » Lo stimatore utilizzato da Excel per il calcolo è il seguente

$$s = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

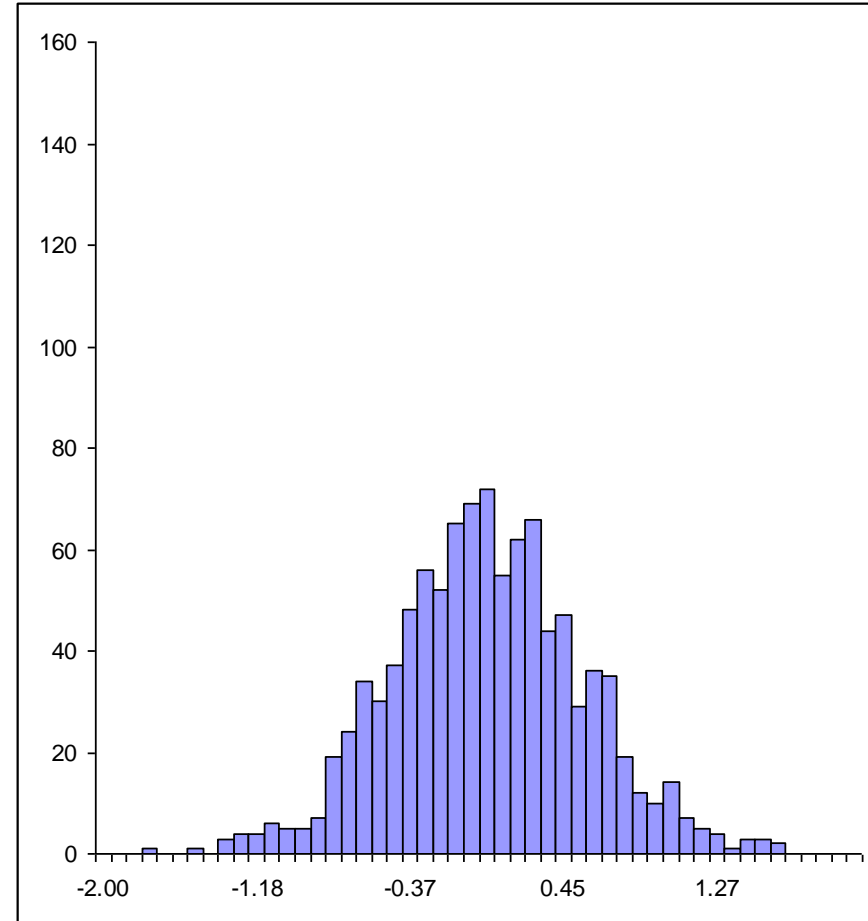
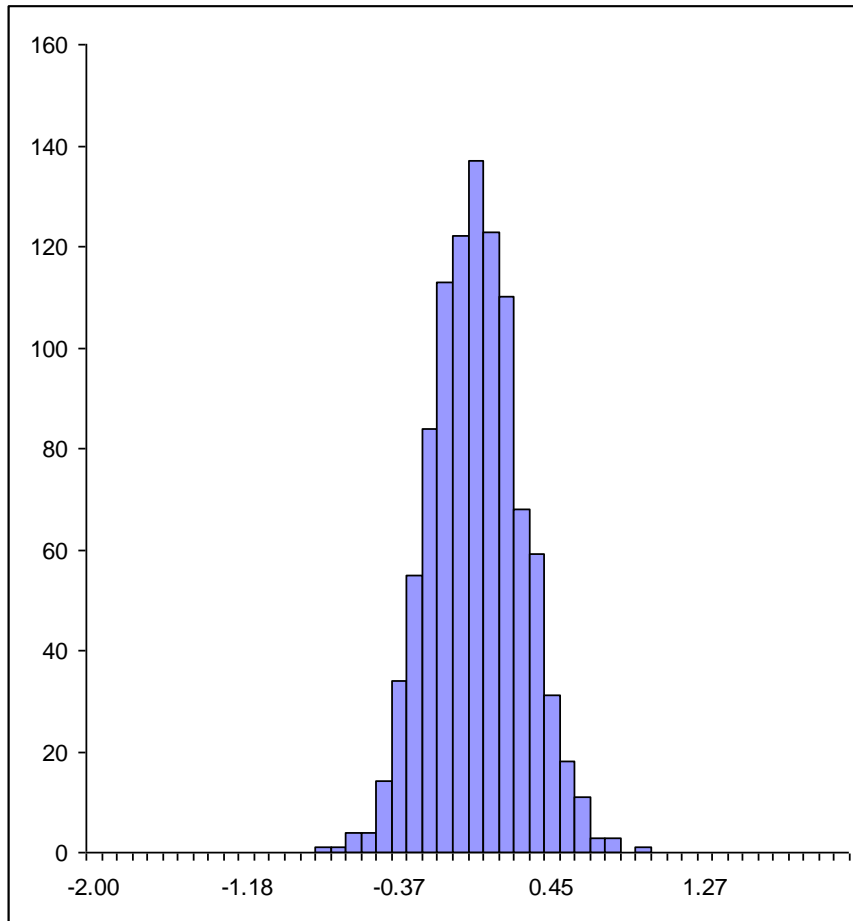
Il significato della deviazione standard

» Due serie storiche di cui la seconda ha standard deviation doppia dell'altra...



Il significato della deviazione standard

» ... e le rispettive distribuzioni di probabilità!

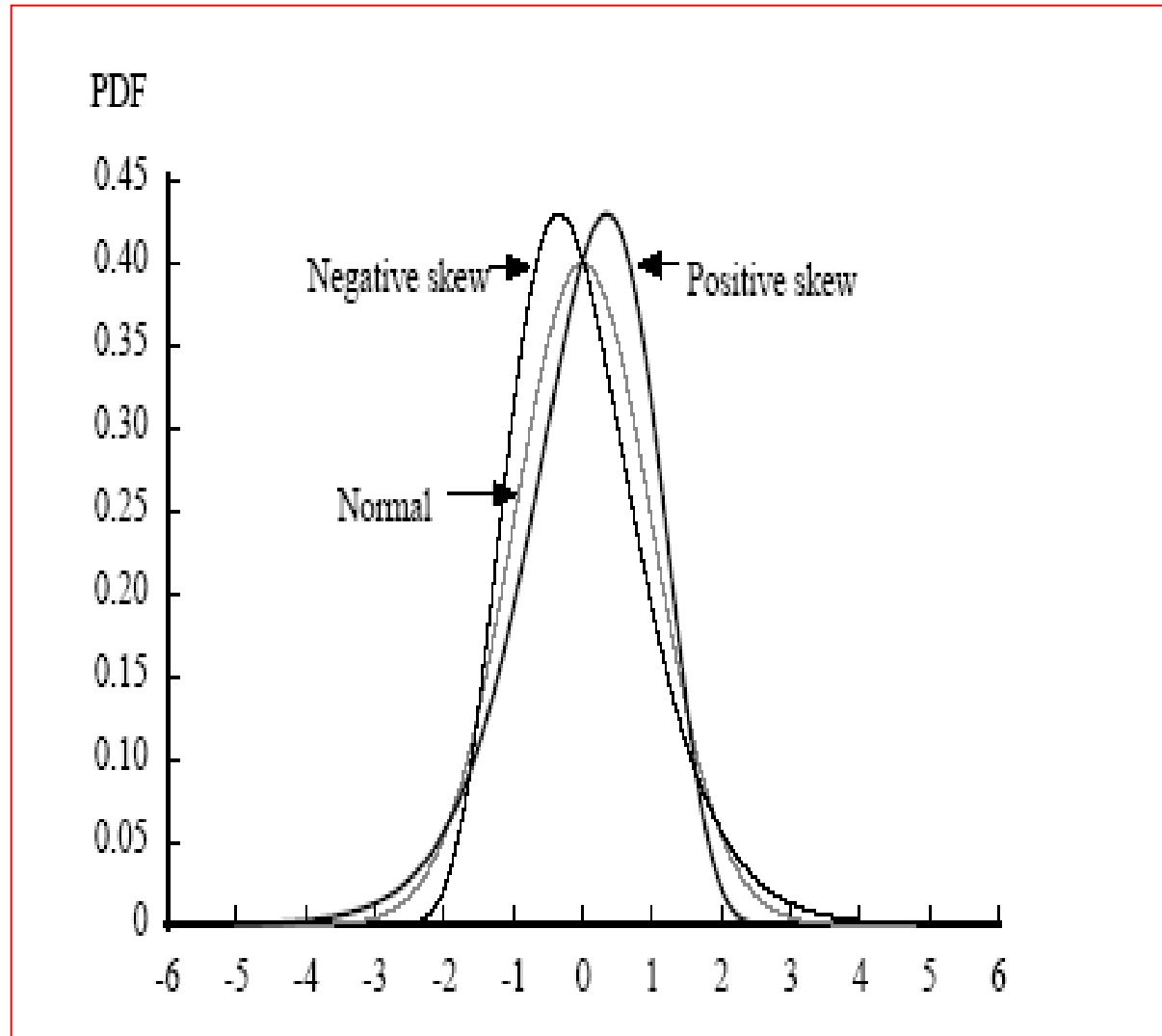


| Funzioni Statistiche: Skewness o Asimmetria

- » Il momento centrale di ordine 3 ci dà informazioni sul **grado di asimmetria** di una distribuzione attorno alla sua media ed è comunemente indicato col termine **skewness**.
- » L'asimmetria positiva indica una distribuzione con una coda asimmetrica che si estende verso i valori più positivi.
- » L'asimmetria negativa indica una distribuzione con una coda asimmetrica che si estende verso i valori più negativi.
- » Excel mette a disposizione la funzione **ASIMMETRIA()** che utilizza il seguente stimatore

$$a = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Funzioni Statistiche: Skewness o Asimmetria

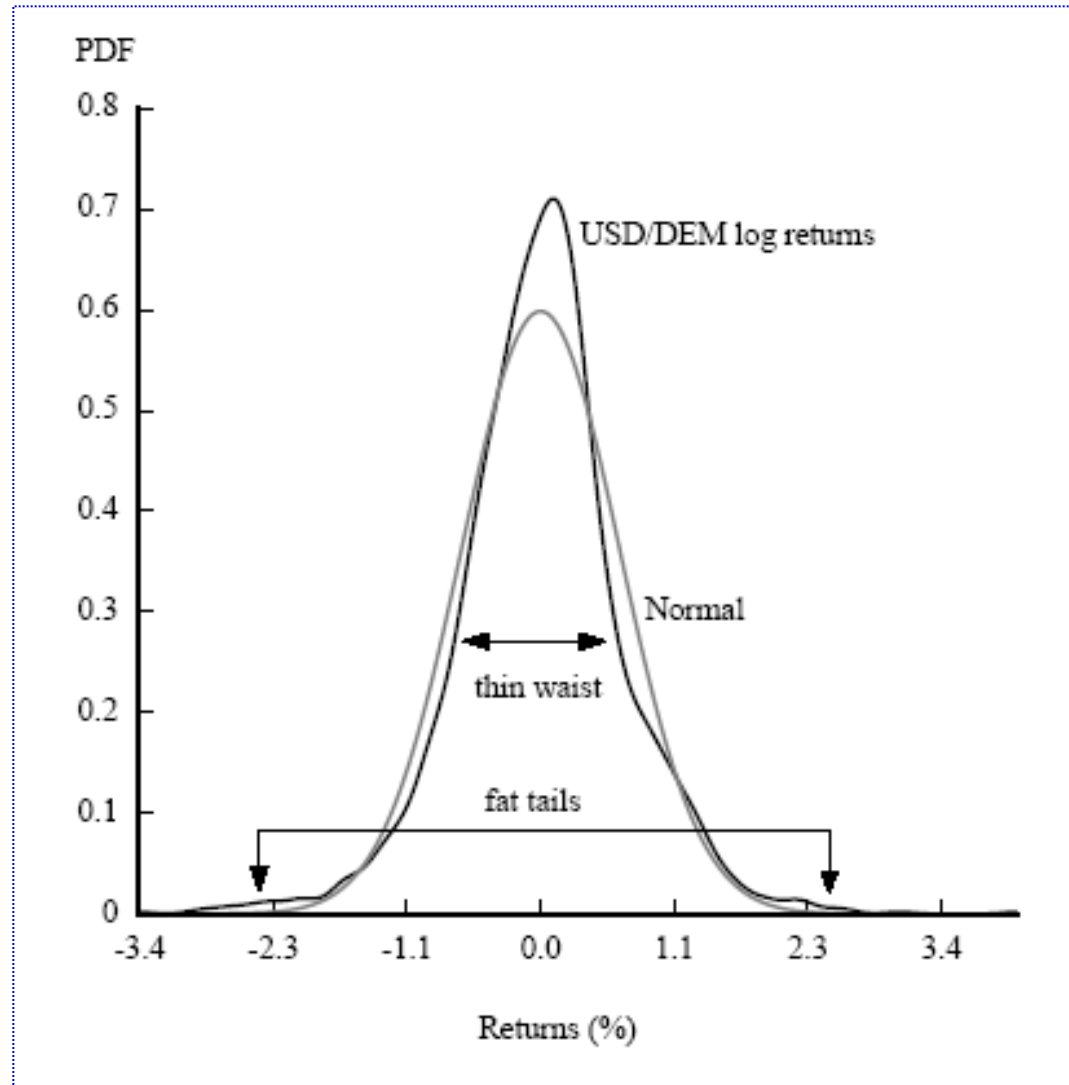


Funzioni Statistiche: Curtosi

- » Infine **CURTOSI()** restituisce la curtosi ed è caratteristica delle cosiddette “code grasse”.
- » Lo stimatore utilizzato da Excel riporta in realtà la cosiddetta “curtosi in eccesso” ovvero la differenza fra il valore stimato e 3.
- » Questo è dovuto al fatto che la distribuzione normale o gaussiana ha curtosi pari a 3 e questo indicatore viene spesso utilizzato come indice per comprendere quando la distribuzione di un insieme di dati si allontani dalla normalità.

$$\left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

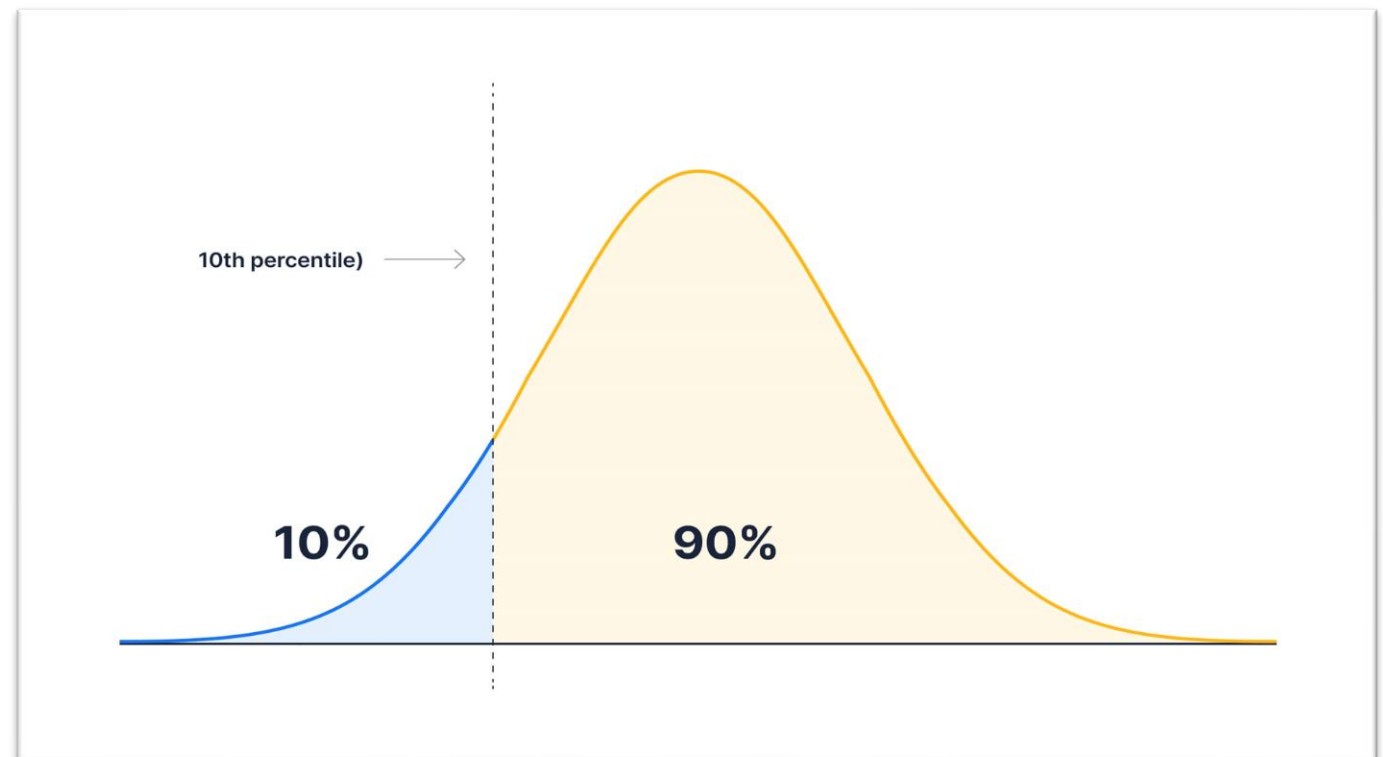
Funzioni Statistiche: Curtosi



» Nell'immagine un esempio di distribuzione empirica dei rendimenti di un titolo in cui si evidenzia il fenomeno della "leptocurtosi" (code grasse)

Funzioni Statistiche: Percentile

- » Un percentile è una misura statistica che indica il valore al di sotto del quale cade una certa percentuale di osservazioni in un dataset.
- » Ad esempio, il 25° percentile (anche noto come primo quartile) è il valore sotto il quale si trova il 25% dei dati. Il 50° percentile è la mediana, e il 75° percentile (terzo quartile) è il valore sotto il quale si trova il 75% dei dati.
- » I percentili sono utili per comprendere la distribuzione dei dati e per identificare i valori estremi e i punti di suddivisione in un dataset.



| Funzioni Statistiche: Percentile

- » È possibile utilizzare questa funzione per stabilire una soglia di accettazione.
- » Ad esempio, in un esame, si può decidere di esaminare i candidati con un punteggio superiore al 90° percentile.
- » In particolare il 50-esimo percentile è una funzione che abbiamo già trovato ...
- » Esempio: Per trovare il 50-esimo percentile dei dati nelle celle da A1 a A100:

=PERCENTILE(A1:A100; 0.5)

Funzioni Statistiche: Riepilogo

Medie, Frequenza e Rango	
<u>MEDIA</u>	Restituisce la media di un elenco di numeri specificati
<u>MEDIA.VALORI</u>	Restituisce la media di un elenco di numeri forniti, conteggiando i valori testuali e il valore logico FALSO come 0 e il valore logico VERO come 1
<u>MEDIA.SE</u>	Calcola la media delle celle in un intervallo specificato, che soddisfino un dato criterio (Nuova in Excel 2007)
<u>MEDIA.PIÙ.SE</u>	Calcola la media delle celle in un intervallo specificato, che soddisfino più criteri (Nuova in Excel 2007)
<u>MEDIANA</u>	Restituisce la mediana (il valore medio) di un elenco di numeri fornito
<u>MODA</u>	Restituisce la moda (il valore che ricorre più di frequente) di un elenco di numeri forniti (Sostituita dalla funzione MODA.SNGL in Excel 2010)
MODA.SNGL	Restituisce la moda (il valore che ricorre più di frequente) di un elenco di numeri forniti (Nuova in Excel 2010 - sostituisce la funzione MODA)
MODA.MULT	Restituisce una matrice verticale dei valori che si verificano più frequentemente in una matrice o in un intervallo di dati (Nuova in Excel 2010)

Funzioni Statistiche: Riepilogo

Medie, Frequenza e Rango	
MEDIA.GEOMETRICA	Restituisce la media geometrica di un insieme di numeri specificati
MEDIA.ARMONICA	Restituisce la media armonica di un insieme di numeri specificati
MEDIA.TRONCATA	Restituisce la media della parte interna di un set di dati
<u>FREQUENZA</u>	Calcola la frequenza di occorrenza dei valori di un intervallo e restituisce una matrice verticale di numeri
<u>RANGO</u>	Restituisce il rango statistico di un dato valore, all'interno di una serie di valori forniti (Sostituita dalla funzione RANGO.UG in Excel 2010)
RANGO.UG	Restituisce il rango di un numero in un elenco di numeri, ovvero la sua grandezza relativa agli altri valori nell'elenco; se più valori hanno lo stesso rango, viene restituito il rango medio (Nuova in Excel 2010 - sostituisce la funzione RANGO)
RANGO.MEDIA	Restituisce il rango statistico di un dato valore, all'interno di una serie di valori forniti (se più di un valore ha lo stesso rango, viene restituito il rango medio) (Nuova in Excel 2010)

Funzioni Statistiche: Riepilogo

Medie, Frequenza e Rango	
CURTOSI	Restituisce la curtosi di un insieme di dati
PERCENTILE	Restituisce il k-esimo percentile di valori in un intervallo specificato, dove K è nel range 0 - 1 (inclusivo) (Sostituito dalla funzione INC.PERCENTILE in Excel 2010)
INC.PERCENTILE	Restituisce il k-esimo percentile di valori in un intervallo specificato, dove K è nel range 0 - 1 (inclusivo) (Nuova in Excel 2010 - sostituisce la funzione PERCENTILE)
ESC.PERCENTILE	Restituisce il k-esimo percentile di valori in un intervallo specificato, dove K è nel range 0 - 1 (esclusivo) (Nuova in Excel 2010)
PERCENT.RANGO	Restituisce il rango di un valore in un insieme di dati, in percentuale (0-1 inclusivo) (Sostituita dalla funzione INC.PERCENT.RANGO in Excel 2010)
INC.PERCENT.RANGO	Restituisce il rango di un valore in un insieme di dati, in percentuale (0-1 inclusivo) (Nuova in Excel 2010 - sostituisce la funzione PERCENT.RANGO)
ESC.PERCENT.RANGO	Restituisce il rango di un valore in un insieme di dati, in percentuale (0-1 esclusivo) (Nuova in Excel 2010)
<u>QUARTILE</u>	Restituisce il quartile di un set di dati, in base ai valori del percentile da 0-1, estremi inclusi (Sostituita dalla funzione INC.QUARTILE in Excel 2010)

Funzioni Statistiche: Riepilogo

Deviazione e Varianza	
MEDIA.DEV	Restituisce la media delle deviazioni assolute dei dati dalla loro media
<u>DEV.ST</u>	Restituisce la deviazione standard di un insieme di valori forniti (che rappresentano un campione di una popolazione) (Sostituita dalla funzione DEV.ST.C in Excel 2010)
<u>DEV.ST.C</u>	Restituisce la deviazione standard di un insieme di valori forniti (che rappresentano un campione di una popolazione) (Nuova in Excel 2010 - sostituisce la funzione DEV.ST)
DEV.ST.VALORI	Restituisce la deviazione standard di un insieme di valori forniti (che rappresentano un campione di una popolazione), conteggiando i valori testuali e il valore logico FALSO come 0 e il valore logico VERO come 1
<u>DEV.ST.POP</u>	Restituisce la deviazione standard di un insieme di valori forniti (che rappresentano un'intera popolazione) (Sostituita dalla funzione DEV.ST.P in Excel 2010)
<u>DEV.ST.P</u>	Restituisce la deviazione standard di un insieme di valori forniti (che rappresentano un'intera popolazione) (Nuova in Excel 2010 - sostituisce la funzione DEV.ST.POP)
DEV.ST.POP.VALORI	Restituisce la deviazione standard di un insieme di valori forniti (che rappresentano un'intera popolazione), conteggiando i valori testuali e il valore logico FALSO come 0 e il valore logico VERO come 1
VAR	Restituisce la varianza di un insieme di valori forniti (che rappresentano un campione di una popolazione) (Sostituita dalla funzione VAR.C in Excel 2010)

Altre Funzioni

Trovare il valore più grande e il più piccolo	
<u>MAX</u>	Restituisce il valore più grande da un elenco di numeri specificato
MAX.VALORI	Restituisce il valore più grande da un elenco di valori specificati, conteggiando i valori testuali e il valore logico FALSO come 0 e il valore logico VERO come 1
<u>MAX.PIÙ.SE</u>	Restituisce il valore massimo tra le celle specificate da un dato set di condizioni (Nuova in Excel 2016)
<u>MIN</u>	Restituisce il valore più piccolo da un elenco di numeri specificati
MIN.VALORI	Restituisce il valore più piccolo da un elenco di valori specificati, conteggiando i valori testuali e il valore logico FALSO come 0 e il valore logico VERO come 1
<u>MIN.PIÙ.SE</u>	Restituisce il valore minimo tra le celle specificate secondo un determinato insieme di condizioni (Nuova in Excel 2016)
<u>GRANDE</u>	Restituisce il k-esimo valore più grande di un set di dati
<u>PICCOLO</u>	Restituisce il k-esimo valore più piccolo di un set di dati

| Altre Funzioni

Contare le celle	
<u>CONTA.NUMERI</u>	Restituisce il numero di valori numerici in un insieme di celle o valori forniti
<u>CONTA.VALORI</u>	Restituisce il numero delle celle non-vuote in un insieme di celle o valori specificati
<u>CONTA.VUOTE</u>	Restituisce il numero di celle vuote in un intervallo specificato
<u>CONTA.SE</u>	Restituisce il numero di celle (di un intervallo specificato), che soddisfino un dato criterio
<u>CONTA.PIÙ.SE</u>	Restituisce il numero di celle (di un intervallo specificato), che soddisfino una serie di criteri forniti (Nuova in Excel 2007)