**Lezione 6: Forma di distribuzione**
Corso di Statistica
Facoltà di Economia
Università della Basilicata

Prof. Massimo Aria

- aria@unina.it

## Standardizzazione di una variabile

**Standardizzare** una variabile statistica significa trasformare la distribuzione originaria in una distribuzione espressa in una unità di misura standard che non risenta dell'effetto scala di misura e dell'effetto media.

# Variabile standardizzata

Una **variabile standardizzata** Z si ottiene sottraendo a tutti i valori la media e rapportando gli scarti così ottenuti allo SQM.

Applicazioni:

- Quando si vogliono confrontare distribuzioni di variabili misurate con diverse scale di misura
- Quando si vogliono confrontare distribuzioni di variabili misurate nella stessa scala di misura ma che hanno diversa intensità media.

Variabile standardizzata

$$X = x_1, x_2, x_3, \ldots, x_N \Rightarrow Z = Z(X)$$

$$Z(X) = \frac{X - \mu}{\sigma}$$

| $X$ | $Z = Z(X)$ |
|---|---|
| $x_1$ | $\frac{x_1 - \mu}{\sigma} = z_1$ |
| $x_2$ | $\frac{x_2 - \mu}{\sigma} = z_2$ |
| $x_3$ | $\frac{x_3 - \mu}{\sigma} = z_3$ |

Processo di standardizzazione

Proprietà

$$\mu_Z = 0$$

$$\boxed{\mu_Z = \frac{1}{N}\sum z_i = \frac{1}{N}\frac{1}{\sigma_X}\sum (x_i - \mu_X) = \frac{0}{N\sigma_X} = 0}$$

$$\sigma_X^2 = 1$$

$$\boxed{\sigma_Z^2 = \frac{1}{N}\sum z_i^2 = \frac{1}{N}\sum_i \left(\frac{x_i - \mu_X}{\sigma_X}\right)^2 = \frac{1}{N}\frac{1}{\sigma_X^2}\sum (x_i - \mu_X)^2 = \frac{1}{\sigma_X^2}\sigma_X^2 = 1}$$

Proprietà della variabile standardizzata

# Forma di una distribuzione di frequenza

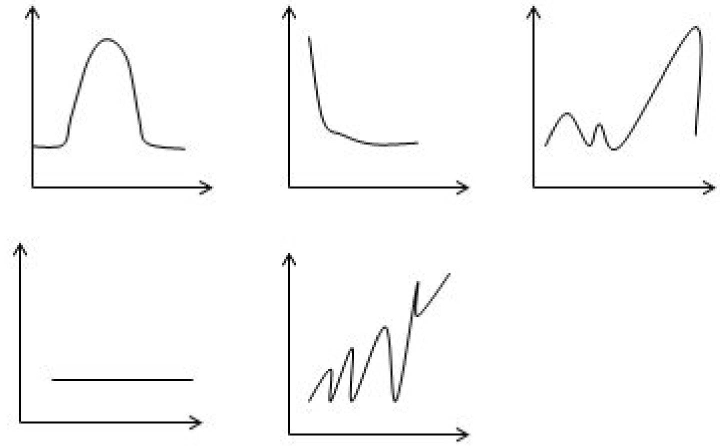
La posizione e la variabilità di una distribuzione di frequenza non esauriscono le informazioni contenute nei dati.

Due variabili statistiche possono avere la stessa posizione e la stessa variabilità ma differire per il peso dei valori che si trovano sulle code, cioè quelli che assumono misure molto distanti dalla media.

Nella statistica descrittiva si definiscono alcune misure concernenti la forma di una distribuzione e che vanno sotto il nome di **asimmetria** e **curtosi**.

## Esempi di forma

Alcuni esempi
della forma di una distribuzione



## Forma normale di una distribuzione

Molti fenomeni assumono un forma cosiddetta normale.

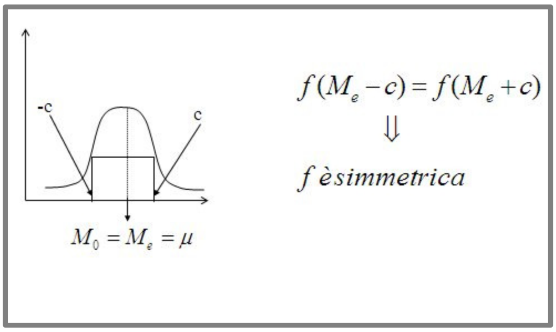
**Normale** è una distribuzione che:
- presenta con una forma campanulare e simmetrica rispetto alla posizione centrale
- Gli indici di posizione assumono uguale valore (Media=Mediana=Moda)

Si dice **simmetrica** invece una forma che, rispetto alla posizione centrale assume uguale struttura delle frequenze sia nella parte destra che nella sinistra:

f(Me-c)=f(Me+c)

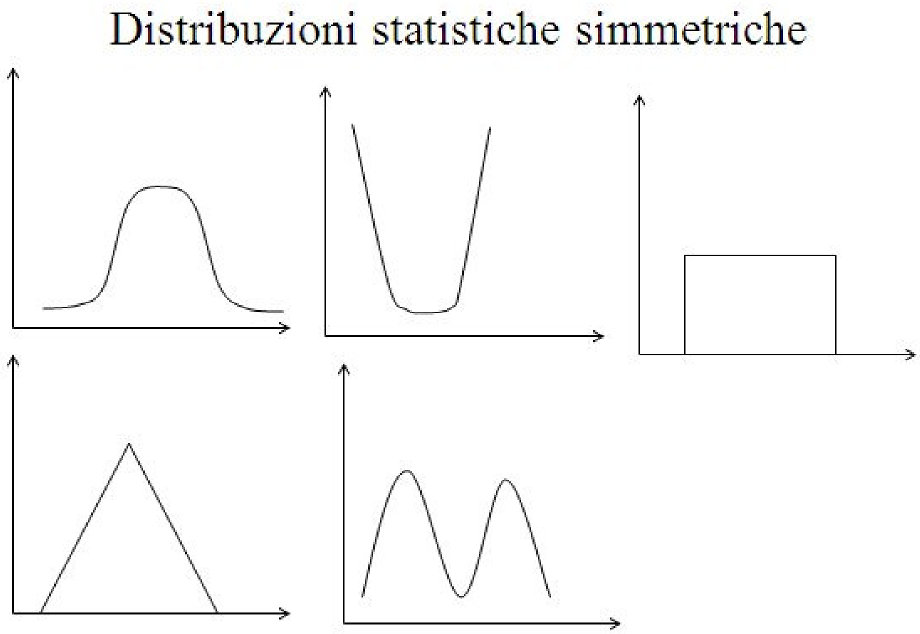
(la frequenza è la stessa sia per la modalità “Mediana meno una costante” sia per quella “Mediana più una costante”, qualunque sia la costante)

NB. Una normale è una particolare distribuzione simmetrica.
Una distribuzione simmetrica potrebbe non essere normale.




Distribuzione Normale

## Esempi di forme simmetriche




## Asimmetria di una distribuzione

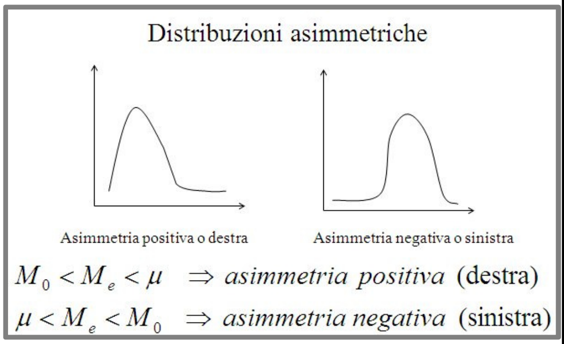
Si dice **asimmetrica** una distribuzione la cui forma non si presenta speculare rispetto alla posizione centrale.

Si parla di:

• **Asimmetria positiva** quando la forma è caratterizzata da una coda allungata verso destra

• **Asimmetria negativa** quando la forma è caratterizzata da una coda allungata verso sinistra

Un metodo empirico per individuare la presenza di asimmetria è quello di confrontare gli indici di posizione della distribuzione considerata.




Distribuzioni asimmetriche

## Indice normalizzato di asimmetria

Un semplice **indice di asimmetria** si ottiene mediante *la differenza tra media e mediana rapportata allo SQM* che si dimostra essere il massimo di questo scarto.

Esso è un indice normalizzato in quanto, essendo rapportato al proprio massimo, varia in valore assoluto tra 0 e 1 a prescindere dall'unità di misura della variabile originaria.

Indice normalizzato di asimmetria

$$A = \frac{\mu - M_e}{\sigma}$$

$$poichè \left| \mu - M_e \right| \leq \sigma \Rightarrow -1 \leq A \leq 1$$

$$A < 0 \Rightarrow asimmetria\ negativa$$

$$A = 0 \Rightarrow simmetria$$

$$A > 0 \Rightarrow asimmetria\ positiva$$

Indice normalizzato di asimmetria

## Indice di asimmetria di Fisher

Un ulteriore indice di asimmetria, proposto da Fisher, è definito come *la media aritmetica delle terze potenze della variabile standardizzata Z*.

Questo indice è positivo, negativo o nullo rispettivamente per una distribuzione asimmetrica positiva, negativo o simmetrica.

L'**indice di Fisher** non è normalizzato, per cui assume valori in tutto l'asse dei numeri reali.

Indice di Fisher

$$\gamma_1 = \frac{1}{N} \sum \left( \frac{x_l - \mu}{\sigma} \right)^3$$

$$Può\ assumere\ valori\ tra\ +\infty\ e\ -\infty$$

$$\gamma < 0 \Rightarrow asimmetria\ negativa$$

$$\gamma = 0 \Rightarrow simmetria$$

$$\gamma > 0 \Rightarrow asimmetria\ positiva$$

Indice di asimmetria di Fisher

## Curtosi di una distribuzioni

Un altro aspetto della forma di una distribuzione di frequenza è la **curtosi**.

Essa riguarda lo studio del maggiore o minore appuntimento, e conseguentemente, il maggiore o minor peso delle code rispetto alla parte centrale della forma.

## Indice di curtosi di Pearson

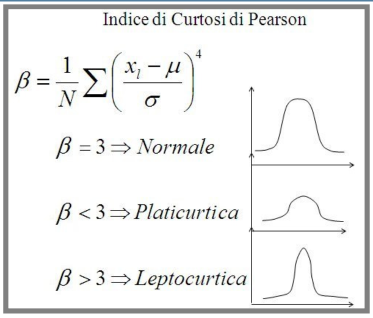
L'**indice di curtosi di Pearson** misura la curtosi come media aritmetica delle quarte potenze della variabile standardizzata Z.

Questo indice assume valore pari a 3 nel caso in cui la distribuzione assuma una forma normale.

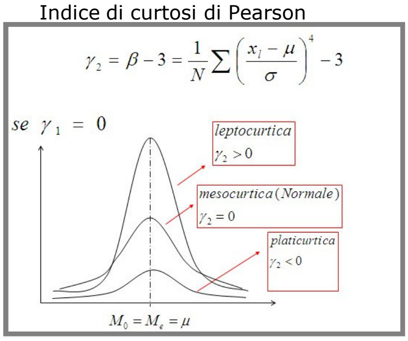
Quando la distribuzione ha una forma maggiormente appuntita rispetto alla normale si parla di **forma leptocurtica** e l'indice sarà > 3.

Quando la distribuzione ha una forma meno appuntita rispetto alla normale si parla di **forma platicurtica** e l'indice sarà < 3.

Sottraendo la costante 3 all'indice di Pearson si ottiene una versione centrata rispetto alla distribuzione normale:
- indice=0 → forma normale
- indice>0 → forma leptocurtica
- indice<0 → forma platicurtica




Indice di curtosi di Pearson




Indice di curtosi centrato rispetto alla normale

## Esempio applicativo degli indici di forma

| $x_i$ | $n_i$ | $z_i = \frac{(x_i - \mu)}{\sigma}$ | $(z_i)^3 \cdot n_i$ | $(z_i)^4 \cdot n_i$ |
|---|---|---|---|---|
| 10 | 2 | -2,11 | -18,90 | 39,97 |
| 12 | 3 | -1,17 | -4,81 | 5,63 |
| 13 | 4 | -0,70 | -1,36 | 0,95 |
| 15 | 7 | 0,25 | 0,10 | 0,03 |
| 16 | 5 | 0,72 | 1,85 | 1,32 |
| 17 | 3 | 1,19 | 5,05 | 6,00 |
| 18 | 1 | 1,66 | 4,58 | 7,62 |
| | **25** | | **-13,50** | **61,52** |

$Media\ \mu = \frac{362}{25} = 14{,}48 \qquad Mediana\ Me = 15$

$Devianza\ SS = 112{,}24 \qquad Varianza\ \sigma^2 = 4{,}49 \qquad SQM\ \sigma = \sqrt{4{,}49} = 2{,}1$

$A = \frac{\mu - Me}{\sigma} = -0{,}25 \qquad \gamma_1 = \frac{1}{N}\sum\left(\frac{(x_i-\mu)}{\sigma}\right)^3 \cdot n_i = \frac{-13{,}50}{25} = -0{,}54$

$\beta = \frac{1}{N}\sum\left(\frac{(x_i-\mu)}{\sigma}\right)^4 \cdot n_i = \frac{61{,}52}{25} = 2{,}46 \qquad \gamma_2 = \beta - 3 = -0{,}54$

## Nella prossima lezione

Nella prossima lezione si affronteranno i seguenti argomenti:

- distribuzioni doppie di frequenza
- profili riga e di colonna
- distribuzioni condizionate