

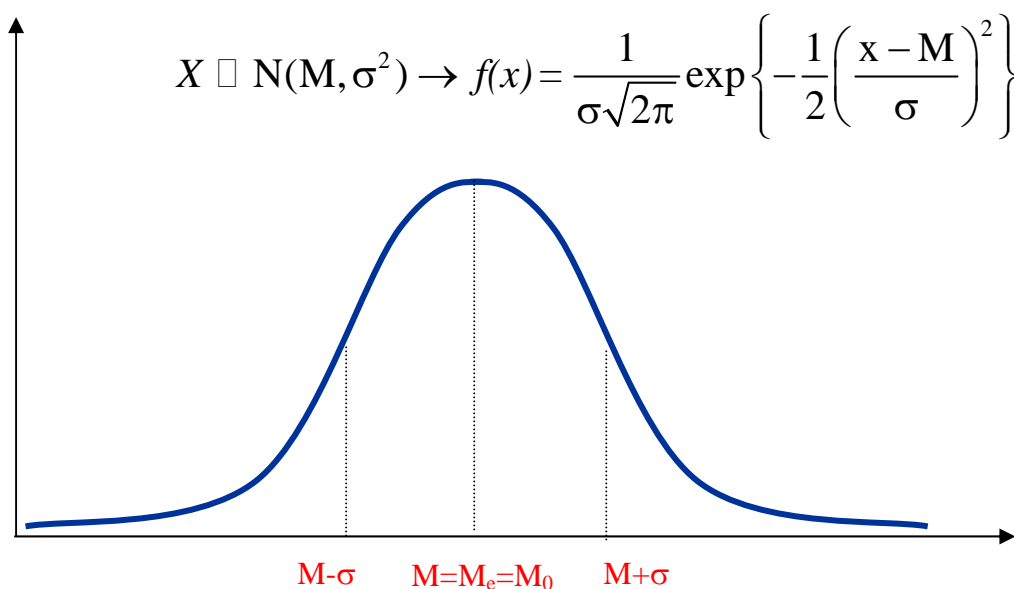
# DISTRIBUZIONE NORMALE

E' una distribuzione teorica che può essere assunta come modello di numerose distribuzioni di frequenza di dati osservati

- è detta **curva degli errori accidentali** cioè di quegli errori che si commettono misurando più volte, in via indipendente e con il medesimo grado di accuratezza, una certa grandezza
- è la distribuzione più usata in statistica
- è un modello continuo

## Proprietà

1. è simmetrica rispetto alla media
2. ha forma campanulare
3. dipende da due parametri, la media  $M$  e la varianza  $\sigma^2$
4. è massima in corrispondenza della media
5. ha due punti di flesso in corrispondenza di  $M-\sigma$  e  $M+\sigma$
6. è asintotica rispetto all'asse delle ascisse

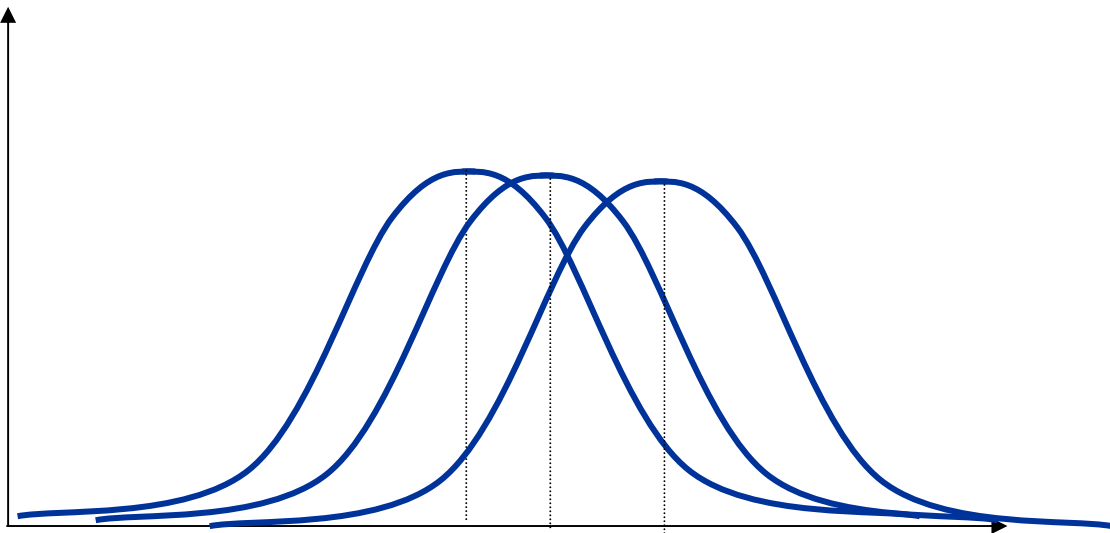


➤ Nella distribuzione normale è possibile determinare la frequenza relativa associata ad un qualsiasi intervallo espresso in termini della media e dello scostamento quadratico medio. In particolare:

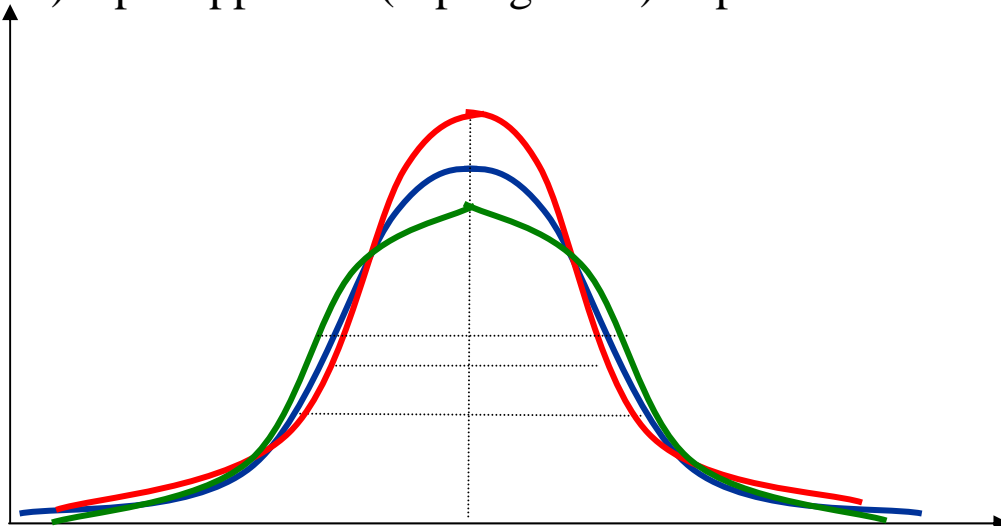
- nell'intervallo  $(M-\sigma, M+\sigma)$  è compreso il 68,27% dei casi
- nell'intervallo  $(M-2\sigma, M+2\sigma)$  è compreso il 95,45% dei casi
- nell'intervallo  $(M-3\sigma, M+3\sigma)$  è compreso il 99,73% dei casi

➤ La curva normale varia al variare di  $M$  e  $\sigma$ :

- al variare di  $M$  subisce traslazioni orizzontali



- al variare di  $\sigma$  la curva assume forma più appuntita ( $\sigma$  più piccolo) o più appiattita ( $\sigma$  più grande) rispetto alla Media



E' necessario quindi ricondurre le infinite curve normali caratterizzate da differenti valori di  $M$  e di  $\sigma$  ad un'unica curva:

## CURVA NORMALE STANDARDIZZATA

$$Z = \frac{x - M}{\sigma} \rightarrow \text{variabile standardizzata}$$

$$E(Z)=0 \quad \text{Var}(Z)=1$$

$$Y \square Z(0,1) \rightarrow f(y) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (z)^2 \right\}$$

$$E(Y)=0 \quad \text{Var}(Y)=1$$

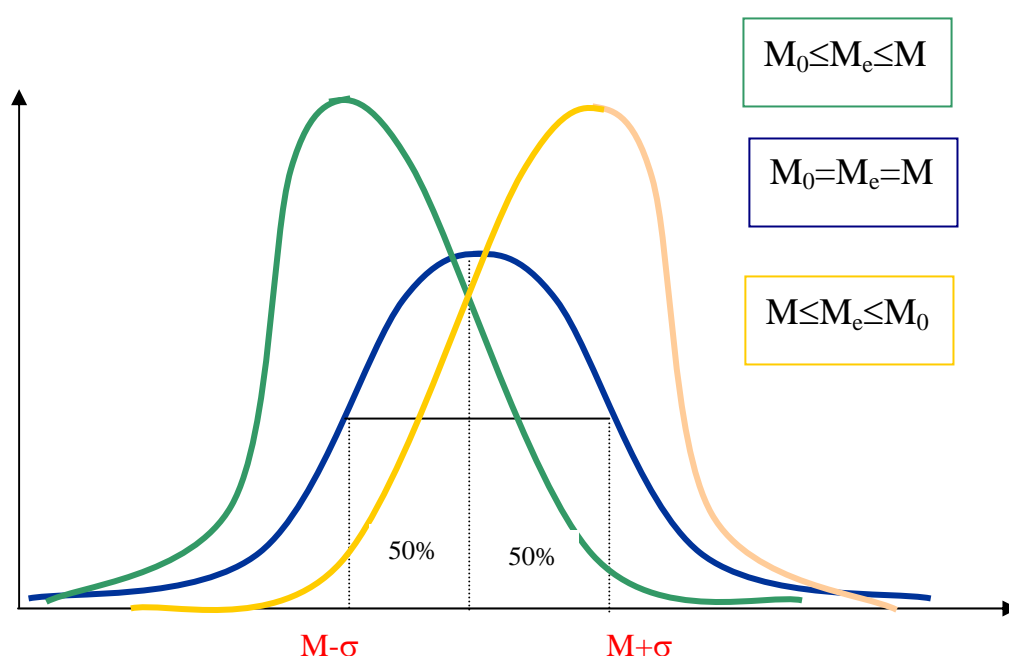
- nell'intervallo  $(M-1, M+1)$  è compreso il 68,27% dei casi
- nell'intervallo  $(M-2, M+2)$  è compreso il 95,45% dei casi
- nell'intervallo  $(M-3, M+3)$  è compreso il 99,73% dei casi

**nb.** l'area al di sotto della curva normale (standardizzata) è pari ad 1

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{x-M}{\sigma} \right)^2 \right\} = 1$$

## ASIMMETRIA E CURTOSI

- una distribuzione si dice simmetrica rispetto ad un polo di simmetria (ad esempio la media aritmetica  $M$ ) se considerata una costante “ $c$ ” i valori  $(M-c)$  e  $(M+c)$  assumono la stessa frequenza
- una curva, rappresentativa di una distribuzione continua, è simmetrica dunque se ad ascisse equidistanti dalla media corrisponde lo stesso valore in ordinata  $f(M-c)=f(M+c)$



- **indici di asimmetria**

$$a_q = \frac{Q_3 - 2M_e + Q_1}{Q_3 - Q_1}$$

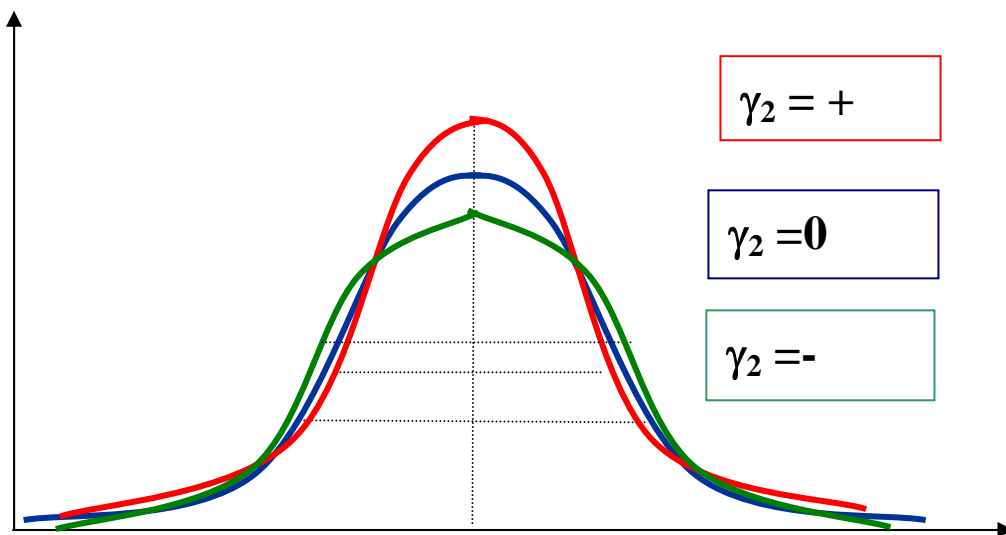
$$\gamma_1 = \frac{\sum_{i=1}^s (x_i - M)^3 n_i}{s\sigma^3}$$

1.  $a_q$  si annulla anche se c'è asimmetria
2.  $\gamma_1 = 0$  simmetria,  $\gamma_1 = +$  asimmetria positiva,  $\gamma_1 = -$  asimmetria negativa

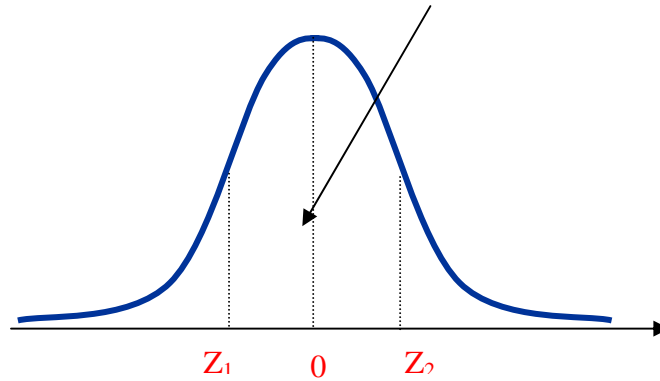
- un indice di curtosi misura l'allontanamento dalla normalità o disnormalità

$$\gamma_2 = \frac{\sum_{i=1}^s (x_i - M)^4 n_i}{s\sigma^4} - 3$$

$\gamma_2 = 0$  normale,  $\gamma_2 = +$  ipernormalità,  $\gamma_2 = -$  iponormalità



## APPLICAZIONI

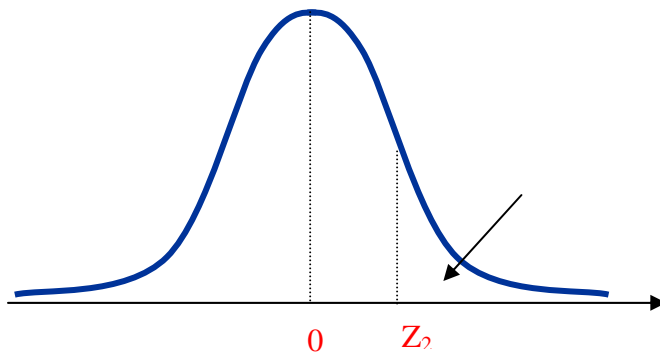


$$\begin{aligned} P(x_1 \leq X \leq x_2) &= P\left(\frac{x_1 - M}{\sigma} \leq Z \leq \frac{x_2 - M}{\sigma}\right) = \\ &= P(z_1 \leq Z \leq 0) + P(0 \leq Z \leq z_2) \end{aligned}$$

**es.** il diametro di un lotto di sfere è normale con  $M=8000$  e  $\sigma=50$ ; si determini la frazione di sfere comprese tra 7930 e 8080

Dati:  $x_1=7930$                        $x_2=8080$   
 $M=8000$                                  $\sigma=50$

$$\begin{aligned} P(7930 \leq X \leq 8080) &= P\left(\frac{7930 - 8000}{50} \leq Z \leq \frac{8080 - 8000}{50}\right) = \\ &= P(-1,4 \leq Z \leq 1,6) = P(-1,4 \leq Z \leq 0) + P(0 \leq Z \leq 1,6) = \\ &= 0,41924 + 0,44520 = 0,8644 = 86,4\% \end{aligned}$$



$$P(X \geq x_2) = P\left(Z \geq \frac{x_2 - M}{\sigma}\right) =$$

$$= P(Z \geq z_2) = 0,5 - P(0 \leq Z \leq z_2)$$

**es.** il lotto di circuiti difettosi è normale se la  $M=9$  e  $\sigma=2,954$ ; si determini la probabilità di avere più di 15 circuiti difettosi

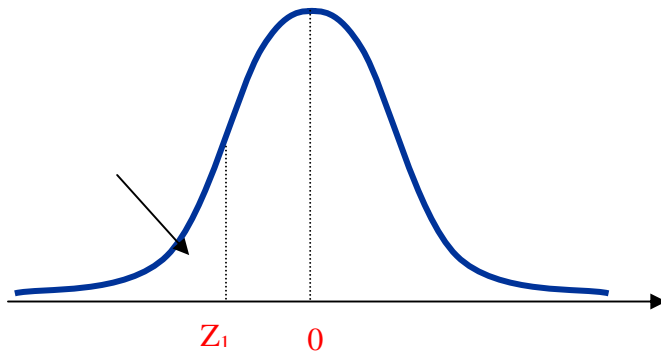
Dati:  $x_2=15$

$$M=9$$

$$\sigma=2,954$$

$$P(X \geq 15) = P\left(Z \geq \frac{15-9}{2,954}\right) =$$

$$= P(Z \geq 2,03) = 0,5 - 0,47882 = 0,02 = 2\%$$



$$\begin{aligned}
 P(X \leq x_1) &= P\left(Z \leq \frac{x_1 - M}{\sigma}\right) = \\
 &= P(Z \leq z_1) = 0,5 - P(z_1 \leq Z \leq 0)
 \end{aligned}$$

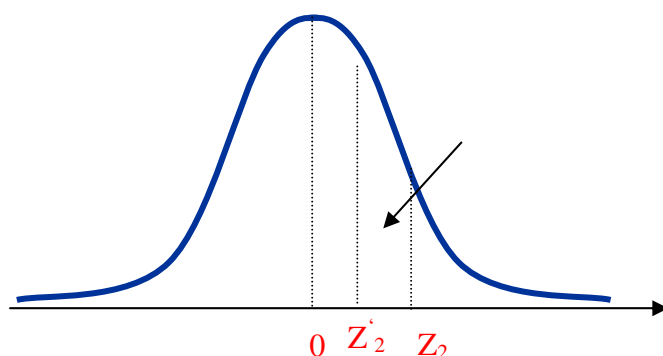
**es.** supponiamo che la statura dei ventenni italiani sottoposti alla visita militare è normale con  $M=172$  e  $\sigma=6$ ; calcolare la percentuale di ventenni di statura inferiore a 160 cm

Dati:  $x_1=160$

$M=172$

$\sigma=6$

$$\begin{aligned}
 P(X \leq 160) &= P\left(Z \leq \frac{160 - 172}{6}\right) = \\
 &= P(Z \leq -2) = 0,5 - P(-2 \leq Z \leq 0) = \\
 &= 0,5 - 0,47725 = 0,02275 = 2,35\%
 \end{aligned}$$

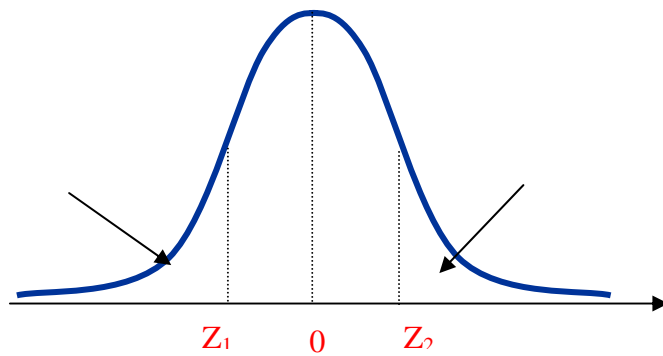


$$\begin{aligned}
 P(x_2' \leq X \leq x_2) &= P\left(\frac{x_2' - M}{\sigma} \leq Z \leq \frac{x_2 - M}{\sigma}\right) = \\
 &= P(z_2' \leq Z \leq z_2) = P(0 \leq Z \leq z_2) - P(0 \leq Z \leq z_2')
 \end{aligned}$$

**es.** la distribuzione degli investimenti di un gruppo di aziende si può ritenere normale con  $M=80$  milioni e  $\sigma=5$  milioni; calcolare il numero di aziende con investimenti compresi tra 83 e 92 milioni

Dati:  $x_2'=83$        $x_2=92$   
 $M=80$                $\sigma=5$

$$\begin{aligned}
 P(x_2' \leq X \leq x_2) &= P\left(\frac{83-80}{5} \leq Z \leq \frac{92-80}{5}\right) = \\
 &= P(0,6 \leq Z \leq 2,4) = P(0 \leq Z \leq 2,4) - P(0 \leq Z \leq 0,6) \\
 &= 0,492 - 0,225 = 0,267 = 26\%
 \end{aligned}$$



$$P(x_2 \leq X \leq x_1) = P\left(\frac{x_2 - M}{\sigma} \leq Z \leq \frac{x_1 - M}{\sigma}\right) =$$

$$= P(z_2 \leq Z \leq z_1) = P(Z \geq z_2) + P(Z \leq z_1)$$

**es.** la lunghezza media di un campione di foglie di lauro è  $M=151$  e  $\sigma=15$  mm; calcolare quante foglie misurano meno di 128 mm e più di 185 mm

Dati:  $x_1=128$        $x_2=185$   
 $M=151$                $\sigma=15$

$$P(185 \leq X \leq 128) = P\left(\frac{185 - 151}{15} \leq Z \leq \frac{128 - 151}{15}\right) =$$

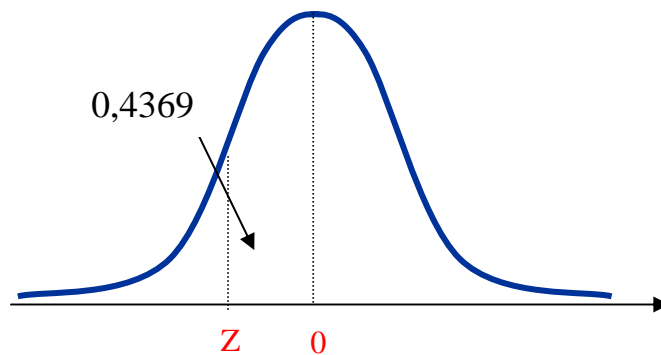
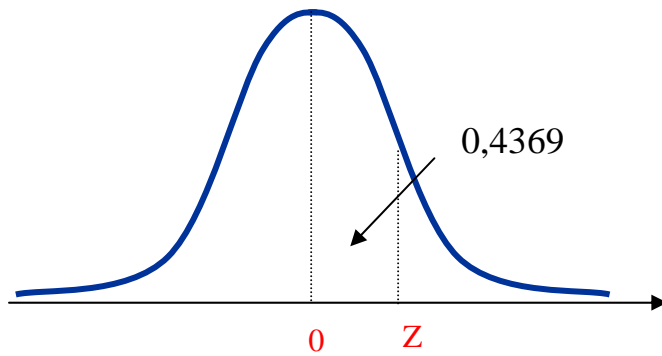
$$P(Z \geq 2,3) + P(Z \leq -1,5) = 0,01 + 0,07 = 0,08 = 8\%$$

**Problema inverso:** data la probabilità trovare il valore di  $z$

es. se la probabilità al di sotto della curva normale compresa tra 0 e  $z$  è pari a 0,4369, trovare il valore di  $z$

$$\text{dati } P(0 \leq Z \leq z) = 0,4369$$

soluzione:  $z = \pm 1,53$



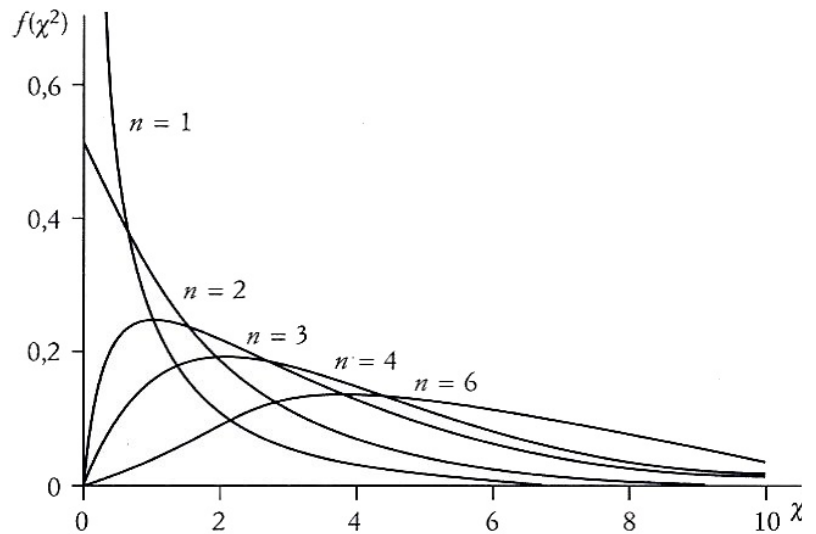
# PARTICOLARI TRASFORMAZIONI DELLA V.C.

## NORMALE

### V.C. $\chi_n^2$ Chi-quadrato

Somma dei quadrati di n v.c.  
normali standardizzate  
indipendenti

$$\sum_i \left( \frac{X_i - \mu_i}{\sigma_i} \right)^2 \square \chi_n^2$$



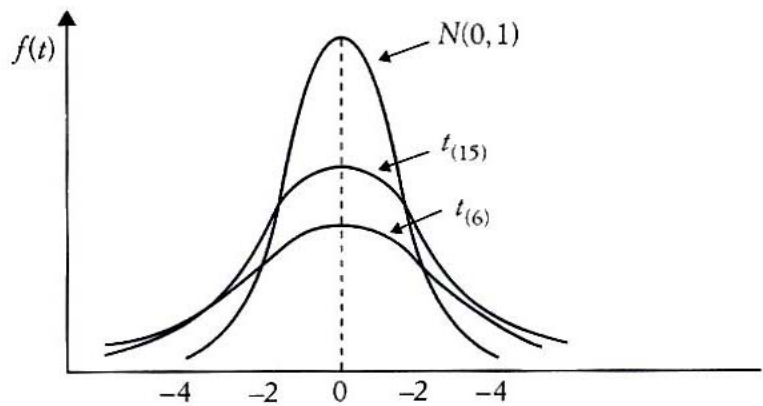
### Applicazioni:

- inferenza sulla varianza di una popolazione normale,
- verifica di un'ipotesi funzionale, test di adattamento
- verifica dell'ipotesi di indipendenza stocastica.

### V.C. $t_n$ T di Student

Rapporto tra una v.c. normale standardizzata e la radice di un v.c. chi-quadrato, con n gradi di libertà, indipendenti tra loro

$$\frac{Z}{\sqrt{\chi^2/n}} \square t_n$$



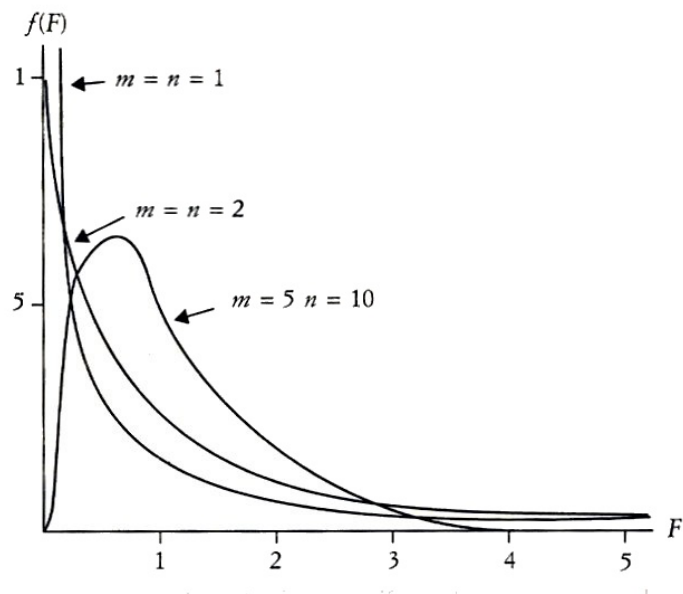
### Applicazioni:

- inferenza sulla media di una popolazione normale,
- inferenza sulla differenza tra le medie di due popolazioni normali,
- inferenza sui parametri del modello di regressione.

## V.C. $F(m,n)$ F di Snedecor e Fisher

Rapporto di due v.c. chi-  
quadrato indipendenti rapportate  
ai rispettivi gradi di libertà

$$\frac{X/m}{Y/n} \square F(m,n)$$



### Applicazioni:

- inferenza sul rapporto tra le varianze di due popolazioni normali,
- inferenza nel modello di regressione.

## PROPRIETÀ DELLA NORMALE

- Data una v.c. Normale X, ogni *trasformazione lineare* di X sarà anch'essa normale.
- La v.c. Normale è stabile, cioè ogni *combinazione lineare* di variabili casuali normali dà luogo ad una variabile casuale che si distribuisce ancora normalmente.

## Teorema del Limite Centrale

Siano  $X_1, X_2, \dots, X_n$   $n$  variabili casuali indipendenti e  
identicamente distribuite con media  $\mu$  e

varianza  $0 < \sigma^2 < +\infty$

la variabile casuale  $Z_n = \frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n\sigma^2}}$

per  $n \rightarrow \infty$  tende a una variabile casuale normale standardizzata

$$Z_n \rightarrow Z \sim N(0,1)$$

## Applicazione del teorema del limite centrale

Approssimazione della variabile casuale binomiale alla variabile casuale  
normale standardizzata

### **Teorema di De Moivre-Laplace**

Considerando che la variabile casuale binomiale  $X = Y_1 + Y_2 + \dots + Y_n$   
è una somma di  $n$  variabili casuali  $Y$  bernoulliane indipendenti di  
media  $p$  e varianza  $pq$

allora per  $n \rightarrow \infty$  la variabile casuale  $Z = \frac{X - np}{\sqrt{npq}} \sim Z(0,1)$

## Esempio di approssimazione alla normale

E' noto che nel 25% delle cavie alle quali è stata somministrata una certa dose di farmaco sottoposto a sperimentazione, avviene una mutazione. A 150 cavie viene somministrata tale dose, qual è la probabilità che i mutanti siano più di 40?

$$X \sim B(150, 0,25)$$

per  $n \rightarrow \infty$  la variabile casuale  $X \cong N(np, (np(1-p))^{1/2})$

$$P(X > 40,5) = P\left(Z > \frac{40,5 - np}{\sqrt{npq}}\right) =$$

$$P\left(Z > \frac{40,5 - 150 \cdot 0,25}{\sqrt{150 \cdot 0,25 \cdot 0,75}}\right) = P(Z > 0,56) = 0,5 - 0,2123 = 0,2877$$