Università degli Studi di Padova Facoltà di Medicina e Chirurgia

Corso di Laurea triennale Tecniche della Prevenzione

PERCORSO SRTAORDINARIO 2007/08

Insegnamento di **STATISTICA MEDICA**

Docente:Dott.ssa Egle Perissinotto

Modulo III

Contenuti:

- Definizione ed elementi di calcolo delle probabilità
- Le variabili casuali
- La distribuzione normale e la normale standardizzata

Fenomeni caratterizzati da incertezza (misurata con la probabilità)

Esempi relativi alla vita quotidiana:

- La probabilità di vincere alla lotteria (evento la cui probabilità è esattamente prevedibile)
- La probabilità di essere investito attraversando la strada (evento la cui probabilità non è esattamente prevedibile)
- La probabilità che piova (probabilità basata sulla credibilità che noi stessi diamo all’evento)
- Teorema centrale della Statistica

Definizione di:

EVENTO: ogni possibile risultato di un esperimento od osservazione

PROBABILITA': rapporto tra il numero di volte in cui l'evento si manifesta ed il numero di volte in cui si è ripetuta la rilevazione o osservazione.

$$P(E) = \frac{\text{Numero di eventi favorevoli}}{\text{Numero totale di eventi (prove)}}$$

Pertanto : $0 \leq P(E) \leq 1$

Definiamo **EVENTI INCOMPATIBILI** gli eventi per i quali il verificarsi di uno di essi esclude il verificarsi degli altri.

Es. Il gruppo sanguigno con modalità:

0 ; A ; B ; AB.

Nella popolazione X, siano:

$P(0) = 0,46$

$P(A) = 0,43$

$P(B) = 0,08$

$P(AB) = 0,03$

Si definisce **CLASSE COMPLETA** di eventi incompatibili l'insieme di tutti gli eventi incompatibili relativi ad un certo fenomeno.

La probabilità di una classe completa di eventi incompatibili (cioè la probabilità del verificarsi di uno qualsiasi di essi) corrisponde all' **EVENTO CERTO** e ha probabilità pari a 1.

Nell'esempio del gruppo sanguigno:

$$P(0 \cup A \cup B \cup AB) = P(0) + P(A) + P(B) + P(AB) =$$

$$= 0{,}46 + 0{,}43 + 0{,}08 + 0{,}03 = 1$$

LA DISTRIBUZIONE DELLE
FREQUENZE RELATIVE EMPIRICHE
INFORMA SU

↓

LA DISTRIBUZIONE DELLA
PROBABILITÀ TEORICA
CHE INFORMA SU

↓

LE **PROBABILITÀ TEORICHE**
DI VERIFICARSI DI DIFFERENTI VALORI

Esempio: Trattamento del diabete

	Modalità (eventi)	Frequenza relativa
1	Nessuno	f_1
2	Solo dieta	f_2
3	Ipoglicemizzanti orali	f_3
4	Insulina	f_4
…	………..	….
k	Altro	f_k
	Totale	1

Nell'esempio della variabile "Trattamento del diabete":

Le modalità della variabile:

- sono incompatibili
- formano una classe completa di eventi

Essendo n_i ($i=1,...,k$) le frequenze assolute ed n il loro totale, si può scrivere:

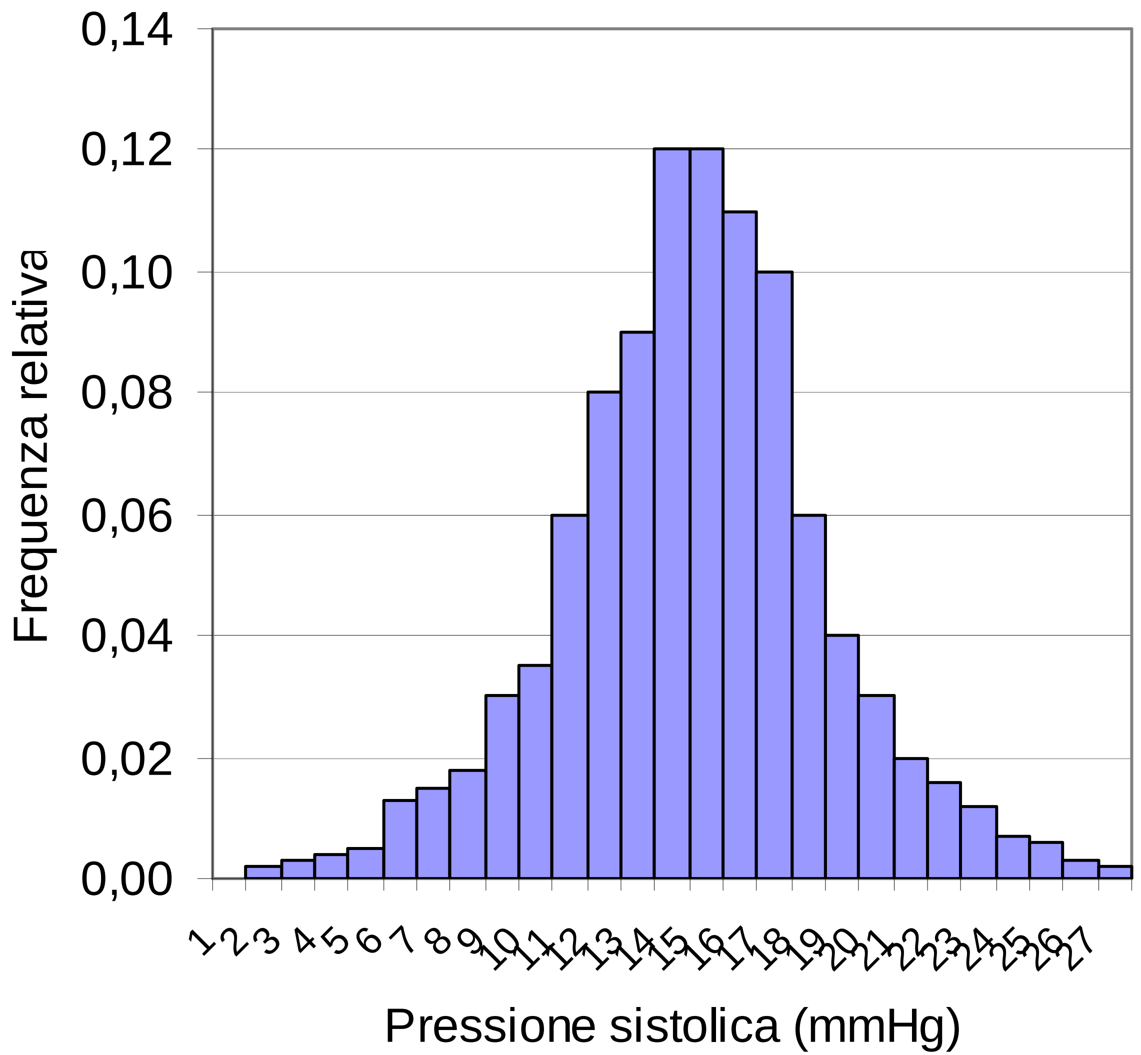
$$\frac{n_i}{n} = f_i = p_i$$

avendosi:

$$\sum_{i=1}^{k} p_i = 1$$

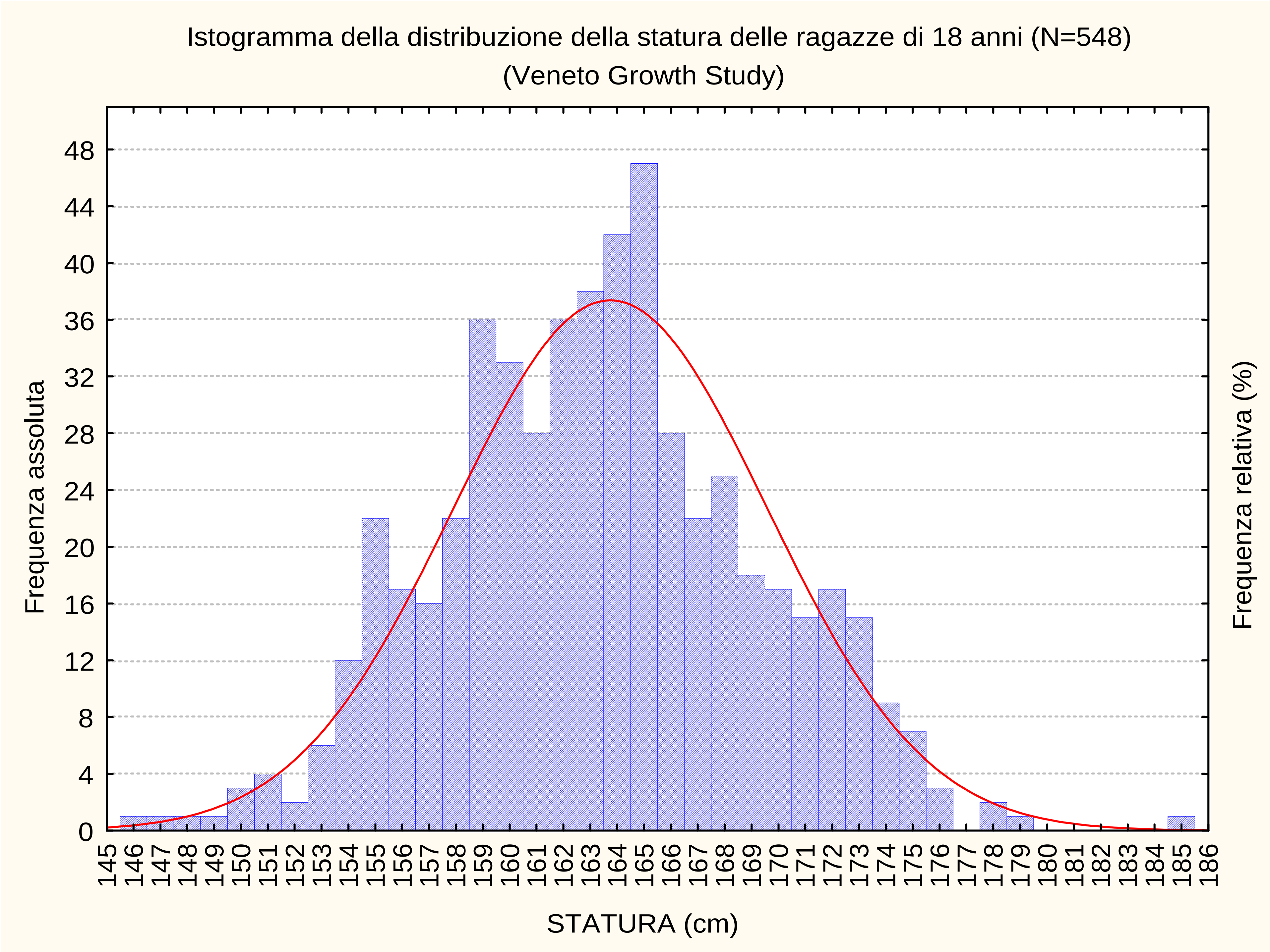
Le variabili casuali
(v.c.)

- Sono modelli teorici utili a descrivere i fenomeni Sono sempre specificate da due entità:
  a) l'insieme dei valori assunti dalla variabile
  b) le frequenze associate a ciascun valore (o densità di frequenza, o funzioni di densità di frequenza)
- Gli eventi specificati dai valori assunti dalle variabili sono sempre incompatibili, e formano una classe completa (*spazio degli eventi*)
- La somma delle frequenze (o l'integrale della funzione di densità di frequenza esteso a tutto il campo di esistenza della variabile) vale uno.

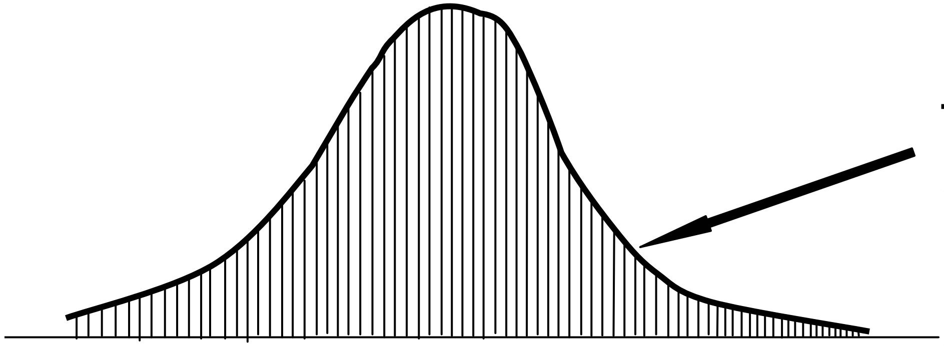


LA VARIABILE CASUALE CONTINUA

È una <u>variabile casuale</u> che può assumere una infinità continua di valori


Istogramma della distribuzione della statura delle ragazze di 18 anni (N=548)
(Veneto Growth Study)
Frequenza assoluta
Frequenza relativa (%)
0
4
8
12
16
20
24
28
32
36
40
44
48
145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186
STATURA (cm)

La si può immaginare come una successione di istogrammi costituiti da barre di cui si riduce progressivamente l’ampiezza dell’intervallo di valori fino a ridurli a dimensione infinitesima

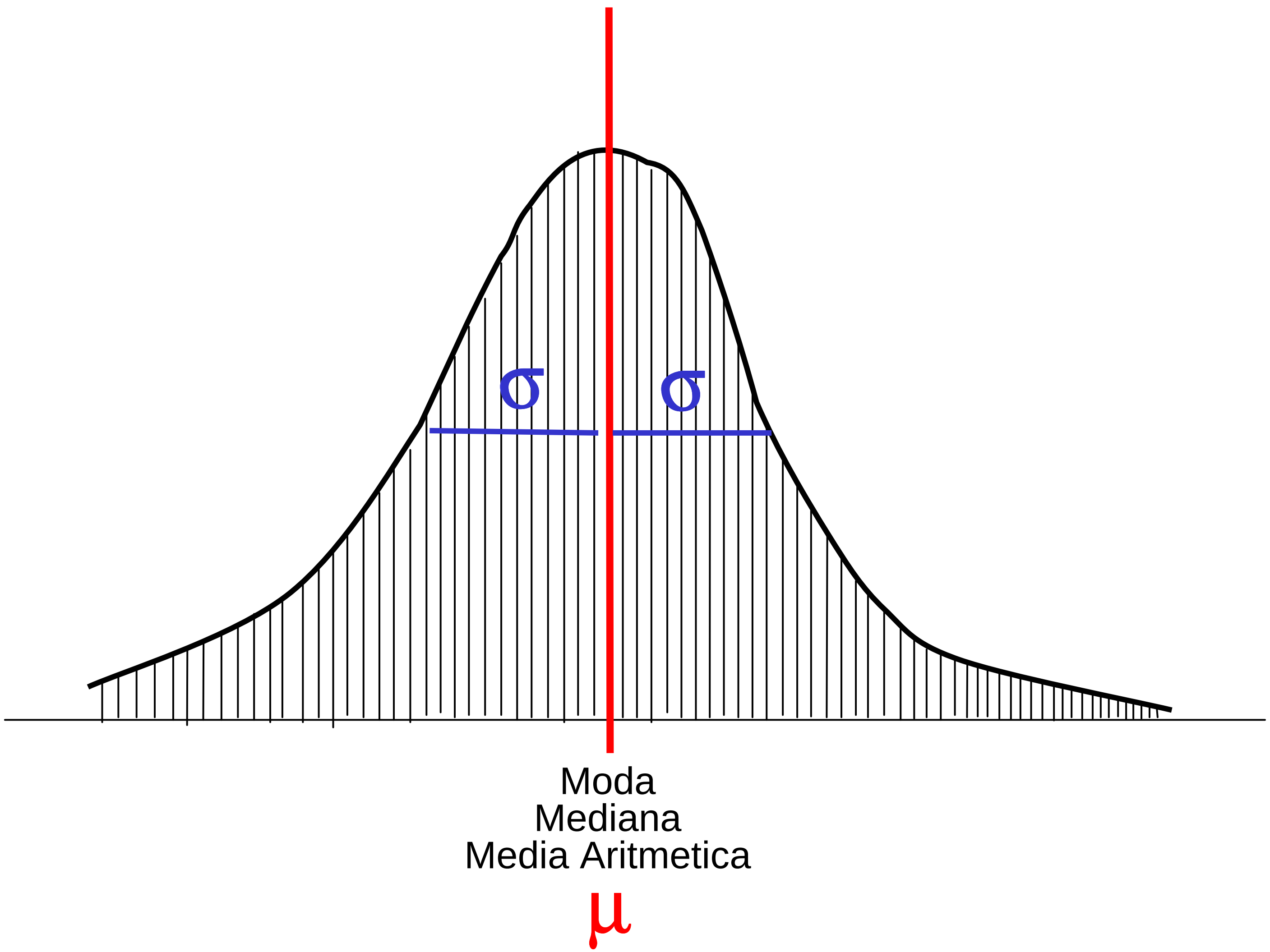


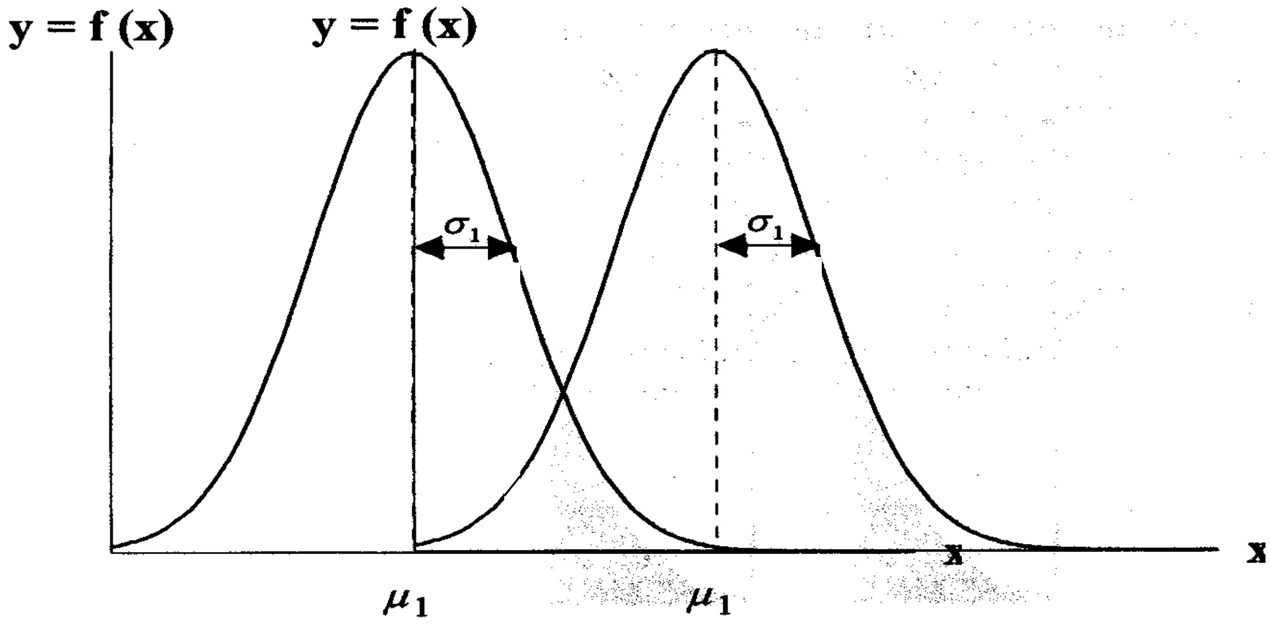
Tale rappresentazione grafica prende il nome di **FUNZIONE Di DENSITÀ** di probabilità.

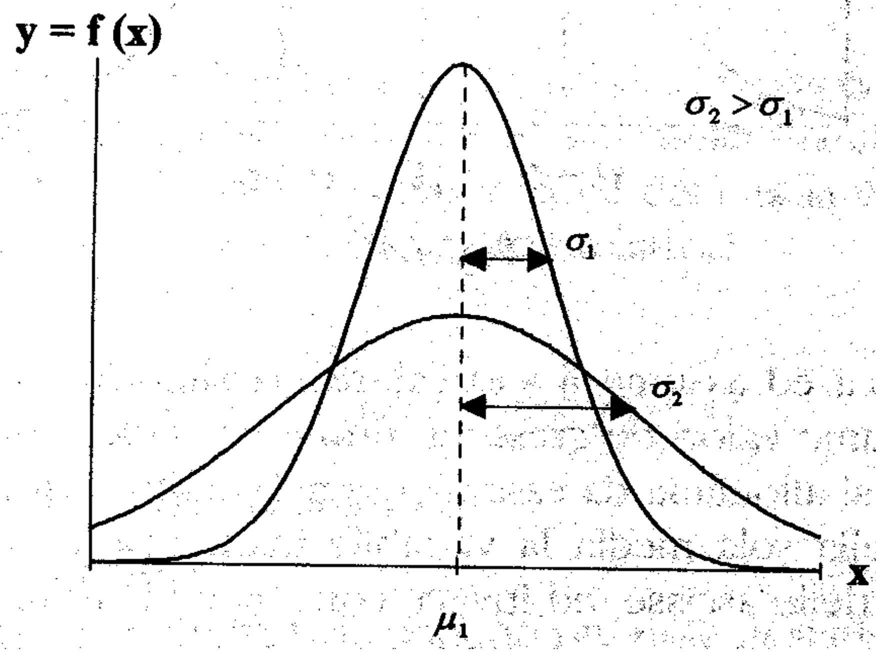
Il modello **NORMALE** o **GAUSSIANA** è un modello di distribuzione teorica di probabilità seguito da molte variabili naturali.

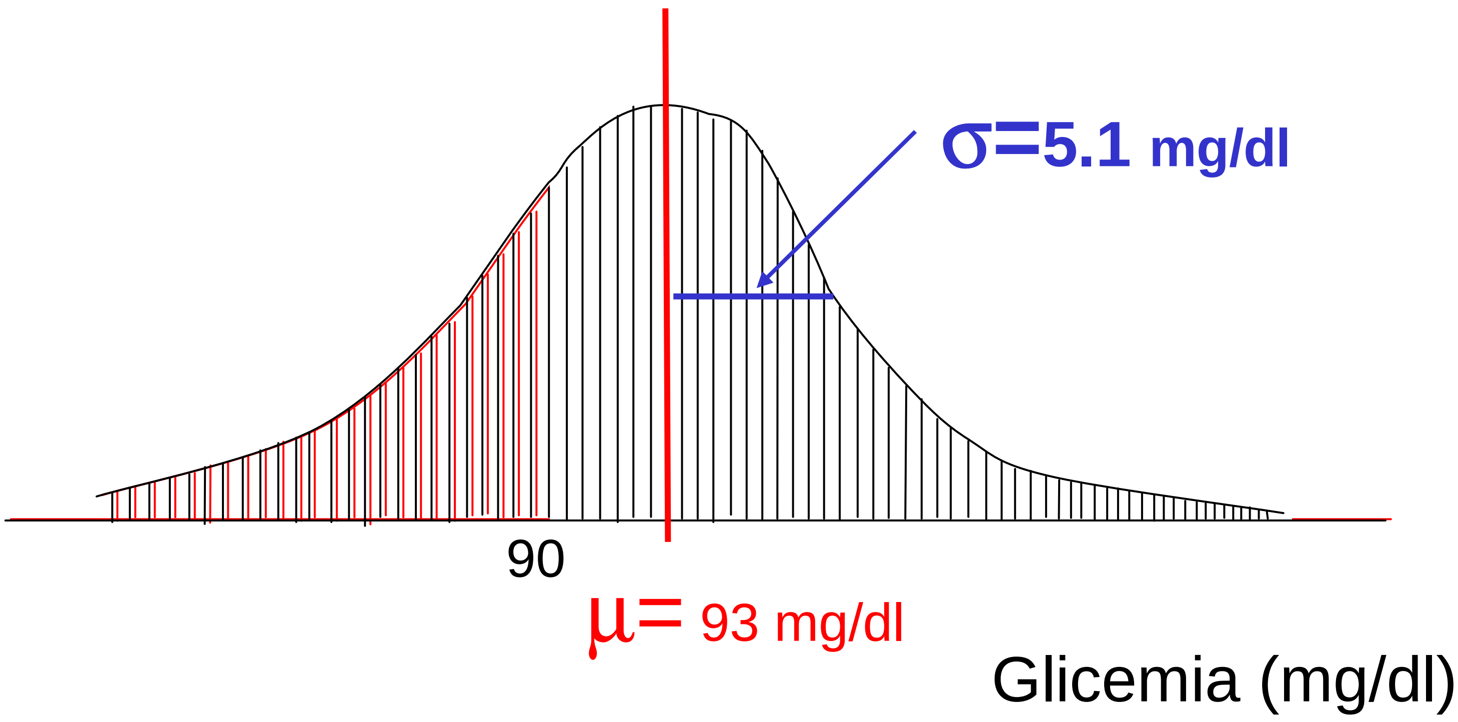
- Tutte le gaussiane hanno la stessa forma e sono definite da $-\infty$ a $+\infty$.
- Sono simmetriche rispetto l'asse $\mathbf{x = \mu}$
- Presentano due flessi (cambio di concavità) nei

  punti $\mu - \sigma$ **e** $\mu + \sigma$
- Media, mediana e moda coincidono.
- Tutte le distribuzioni Normali sono individuate da due parametri: μ e σ e sono genericamente indicate con: **N** $(\mu; \sigma)$


σ
σ
Moda
Mediana
Media Aritmetica
μ


y = f (x)
y = f (x)
σ_1
σ_1
x
x
μ_1
μ_1


y = f (x)
$\sigma_2 > \sigma_1$
σ_1
σ_2
x
μ_1



$$\Pr\left\{ \text{glicemia} <= 90 \right\} = ?$$

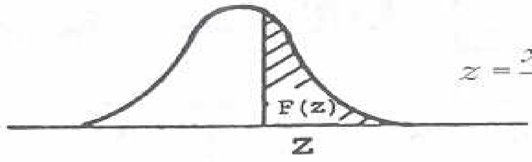
Distribuzione NORMALE STANDARDIZZATA

- Quando $\mu = 0$ e $\sigma = 1$ la distribuzione normale è chiamata NORMALE STANDARDIZZATA, e viene indicata N(0,1)
- Questa variabile casuale viene indicata con la Z

La *funzione di ripartizione* di Z è riportata nelle tavole della normale

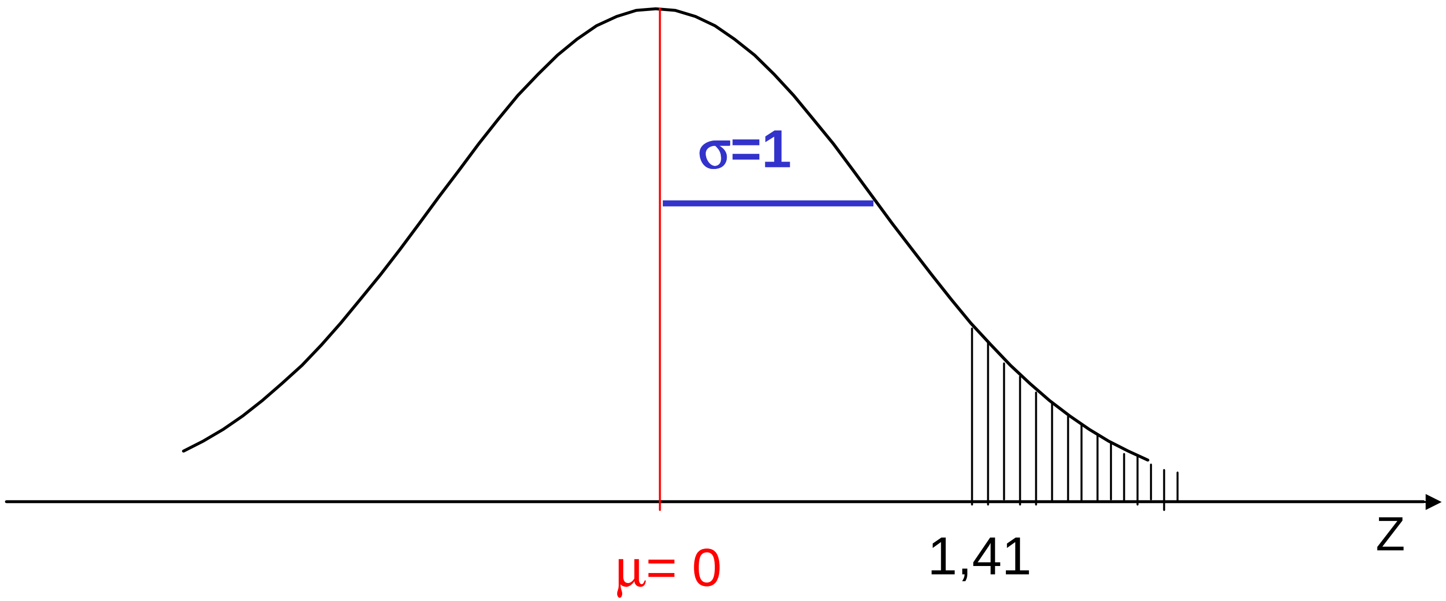
TAVOLE DELLA DISTRIBUZIONE NORMALE STANDARDIZZATA *F(z)*

Tavola 1. - *Aree a destra del valore di z (Probabilità di avere un valore di z superiore ad un certo livello)*



$$z = \frac{x - \mu}{\sigma}$$

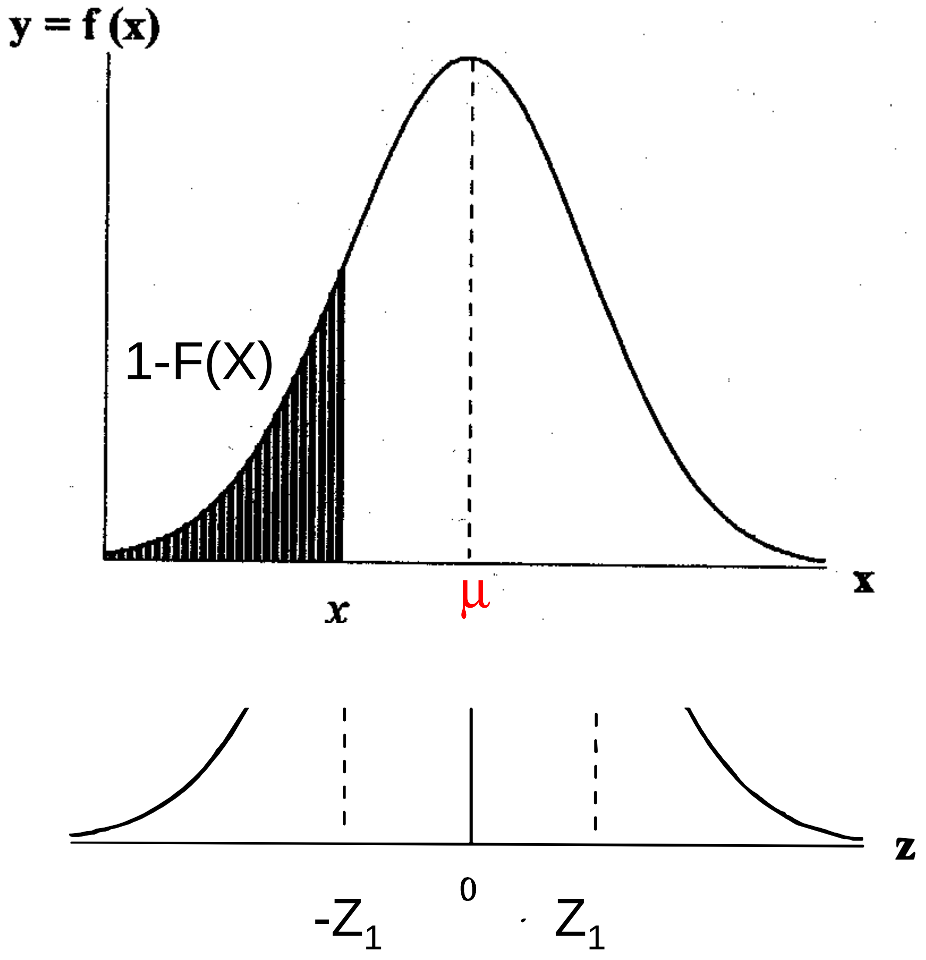
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4771	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1824	.1867
0.9	.1841	.1814	.1788	.1762	.1836	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0978	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.00999	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00766	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00582	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00279	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00168	.00164	.00159	.00154	.00148	.00144	.00139

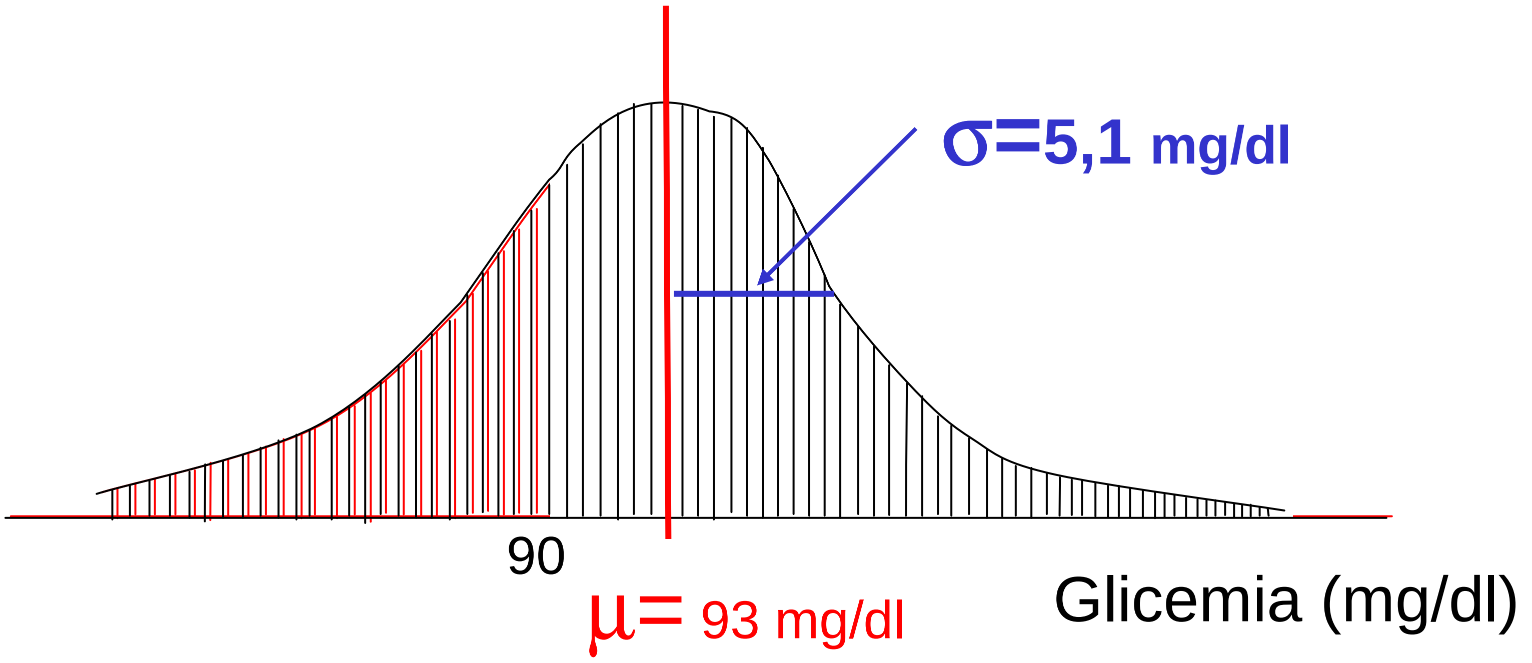


$$\Pr\left\{ Z \geq 1{,}41 \right\} = ? \quad = 0{,}0793 \cong 8\%$$

Per il calcolo delle aree sottese alla curva relativamente a **qualsiasi** variabile casuale normale è possibile ricorrere alle tavole della Z, calcolando lo ***scarto standardizzato*** :

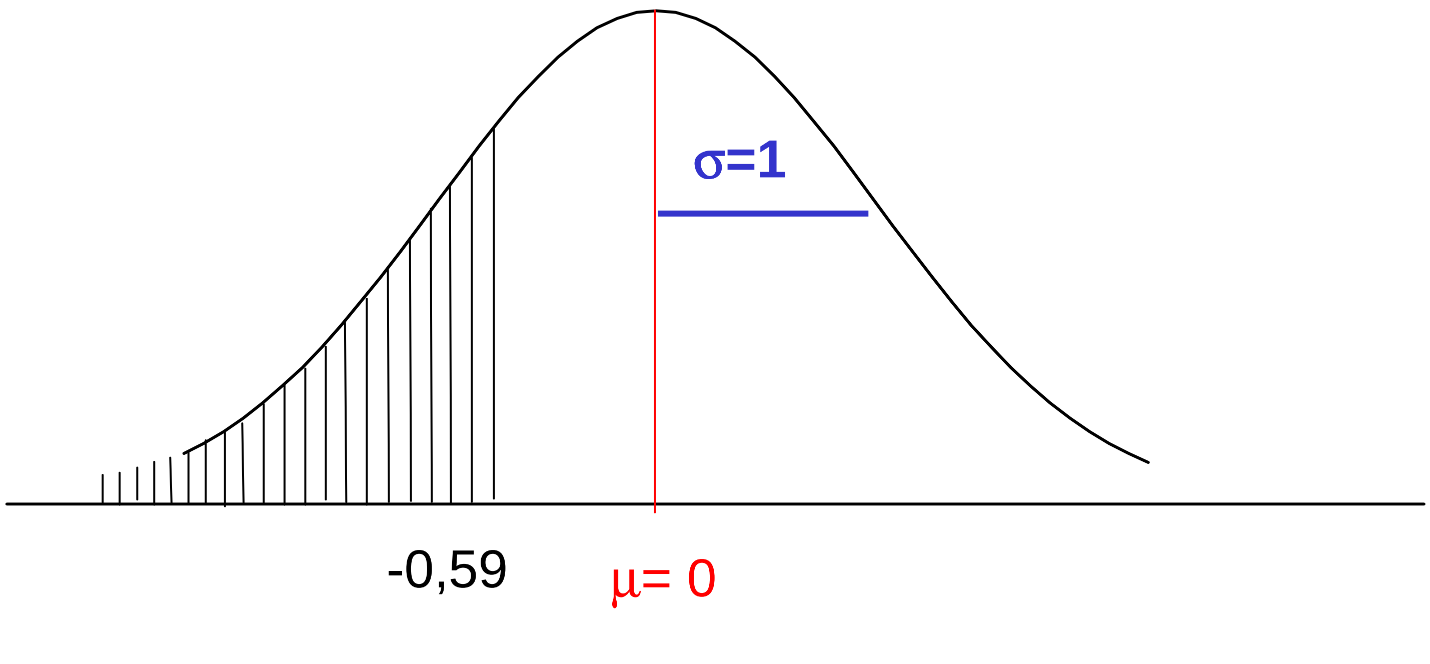
$$Z = \frac{X - \mu}{\sigma}$$





$$Pr\left\{\text{glicemia} <= 90\right\} = ? = Pr\left\{Z <= -0,59\right\} = 0,28$$

$$z = \frac{x - \mu}{\sigma} = \frac{90 - 93}{5,1} = -0,59$$


σ=1
-0,59
μ= 0

ESERCIZIO

Sapendo che la durata del periodo di rieducazione alla coordinazione in soggetti anziani dopo ictus è distribuita in modo normale con media (μ) paria 64 giorni e deviazione standard (σ) paria 12 giorni, qual è la probabilità che un soggetto casualmente estratto abbia necessità di un periodo di rieducazione maggiore di 90 giorni?