Distribuzione Normale

La distribuzione normale, comunemente nota come distribuzione gaussiana, rappresenta una delle distribuzioni di probabilità più fondamentali e ampiamente utilizzate nel campo della statistica. Questa distribuzione è caratterizzata da una forma ad arco, simile a una campana, che è perfettamente simmetrica rispetto al suo centro, corrispondente alla media. La distribuzione normale viene definita matematicamente da due parametri principali: la media (μ), che determina la posizione centrale della distribuzione, e la deviazione standard (σ), che misura la dispersione dei dati attorno alla media.

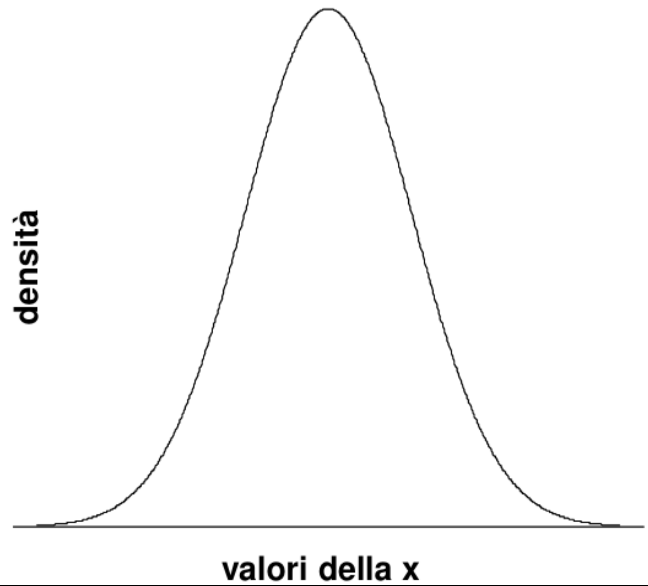
Una delle caratteristiche più significative della distribuzione normale è che la maggior parte dei dati tende a concentrarsi attorno alla media. In effetti, circa il 68,27% dei valori di una variabile che segue una distribuzione normale si trova entro una deviazione standard dalla media, mentre il 95,45% dei valori rientra entro due deviazioni standard e il 99.7% entro tre deviazioni standard. Questa proprietà, è particolarmente utile per comprendere la variabilità dei dati e fare previsioni.

La ragione della vasta applicabilità della distribuzione normale in statistica deriva dal fatto che molte variabili reali, come l'altezza, il peso, e i punteggi ottenuti in test standardizzati, tendono ad avvicinarsi a questa forma distribuzionale. Di conseguenza, l'impiego di tecniche statistiche basate sulla distribuzione normale permette di analizzare e interpretare i dati in modo efficiente e accurato. La capacità di approssimare una varietà di fenomeni naturali e sociali rende la distribuzione normale uno strumento fondamentale per statistici, ricercatori e data scientist.

L'origine della distribuzione normale risale al 1733, quando il matematico Abraham de Moivre la utilizzò come un metodo per valutare approssimativamente la funzione di probabilità binomiale. Tuttavia, la distribuzione ha guadagnato una rilevanza significativa con i lavori di Carl Friedrich Gauss nel 1809, che la applicò nel contesto della teoria degli errori. Gauss dimostrò che gli errori nelle misurazioni seguono una distribuzione normale, consolidando ulteriormente la sua importanza in statistica.

In termini di definizione matematica, si considera una variabile casuale continua X che può assumere qualsiasi valore all'interno dell'intervallo reale, da meno infinito a più infinito ($-\infty$; $+\infty$). Questa definizione permette di analizzare variabili che possono teoricamente mostrare un ampio range di valori, offrendo così un quadro chiaro dell'andamento dei dati nel contesto della distribuzione normale. Essenzialmente, la distribuzione normale funge da fondamento per molte aree della statistica e della scienza dei dati, continuando a influenzare il modo in cui comprendiamo e interpretiamo le informazioni quantitative.

X ha una distribuzione di probabilità normale se la sua densità di probabilità risulta essere:



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

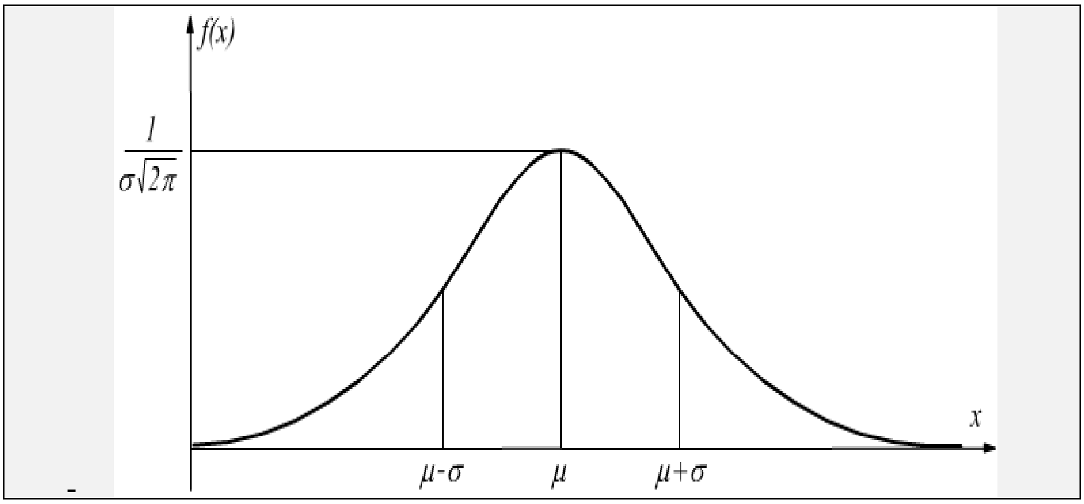
$\pi = 3{,}14$

$e = 2{,}72$

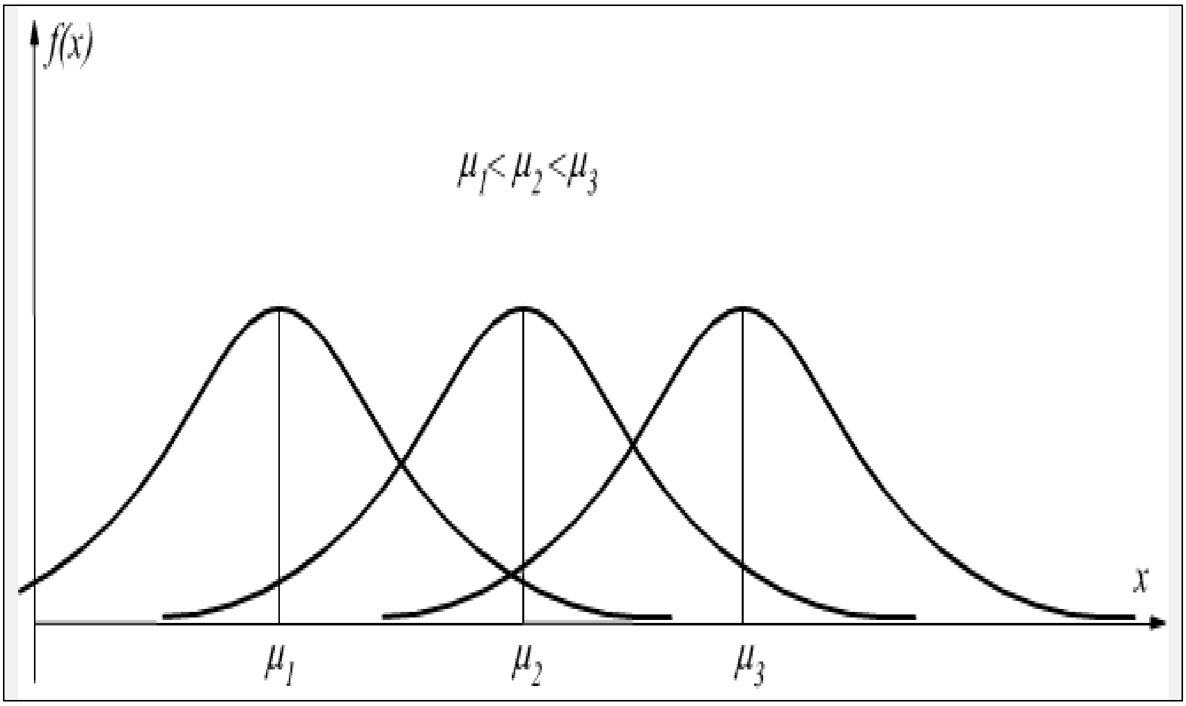
con σ = deviazione standard (scarto quadratico medio) e μ = media. Se X ha una distribuzione normale di probabilità descritta dalla formula suddetta, X viene detta variabile normale.

Le caratteristiche sulle quali si fonda f(x) nella distribuzione normale sono:

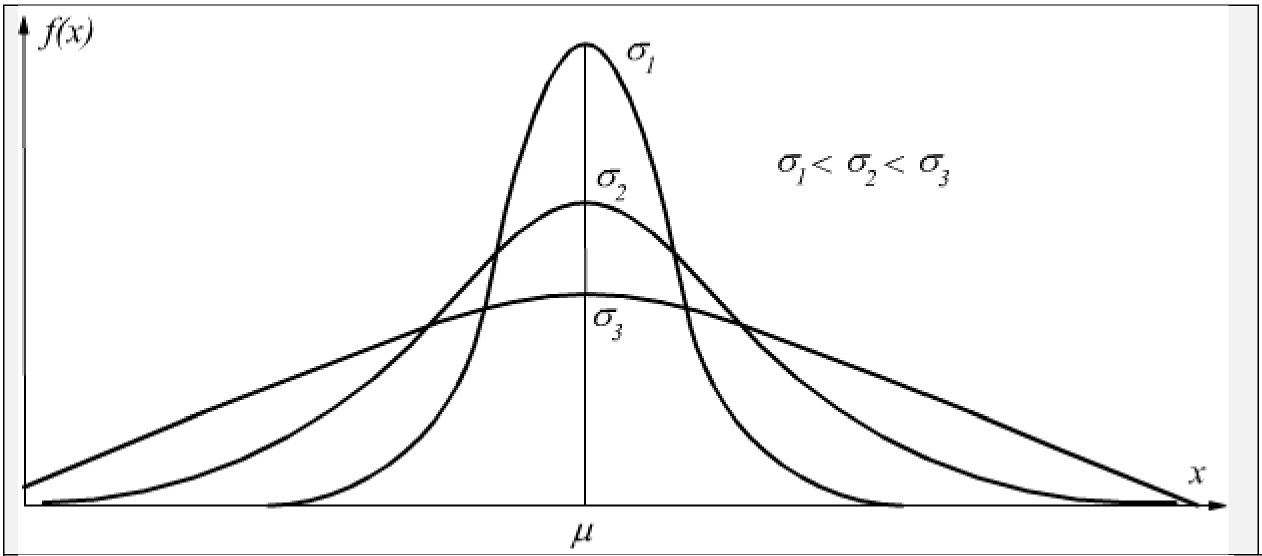
- La forma a campana della curva;
- ha media pari a μ e varianza pari a σ2
- è simmetrica rispetto alla media
- ha media, moda e mediana coincidenti
- ha il massimo di x= μ quando $\frac{1}{\sigma\sqrt{2\pi}}$



Al variare di μ la curva, la distribuzione si sposta lungo l'asse x, ma resta invariata la sua forma.



Il parametro σ descrive come è configurata la curva, in quanto indica quanto i valori si distribuiscono intorno al punto più alto della curva stessa. In altre parole, σ misura la variabilità dei dati rispetto al loro valore centrale, influenzando così l'ampiezza e la forma della curva stessa.

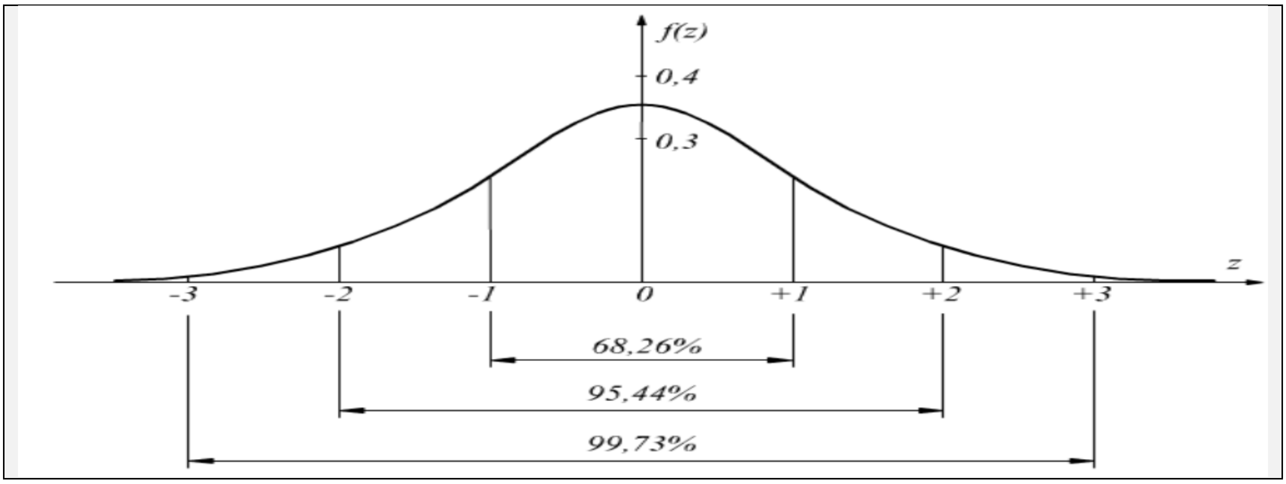


Standardizzazione: Trasformazione lineare dei dati effettuata ricorrendo ai due parametri fondamentali di una distribuzione: la media e lo scarto quadratico medio, secondo la formula riportata di seguito:

$$Z = \frac{x_i - \bar{x}}{\sigma}$$

Media aritmetica = 0

Scarto quadratico medio = 1



- z è la variabile normale standardizzata.
- Il 68,27% della distribuzione dei valori è compreso fra -1 e +1.
- Il 95,45% della distribuzione dei valori è compreso fra -2 e +2.
- Il 99,7% della distribuzione dei valori è compreso fra -3 e +3.

Confrontare distribuzioni aventi differenti unità di misura;

Esempio

Altezze ➔ Media: 170 cm ; S.Q.M. = 10 cm

Pesi: Media ➔ 70 kg ; S.Q.M. = 5 kg

	Altezza	**Peso**
A	**190**	**85**
B	**165**	**75**
C	**185**	**80**

Individuo A:

$$Z_A = \frac{x_i - \bar{x}}{\sigma} = \frac{(190-170)}{10} = \frac{20}{10} = 2$$

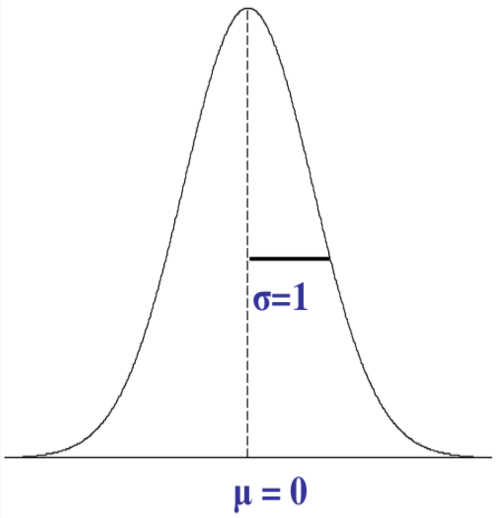
$$Z_P = \frac{x_i - \bar{x}}{\sigma} = \frac{(85-70)}{5} = \frac{15}{5} = 3$$

Individuo B:

$$Z_A = \frac{x_i - \bar{x}}{\sigma} = \frac{(165-170)}{10} = \frac{-5}{10} = -\frac{1}{2}$$

$$Z_P = \frac{x_i - \bar{x}}{\sigma} = \frac{(75-70)}{5} = \frac{5}{5} = 1$$

La curva normale standardizzata



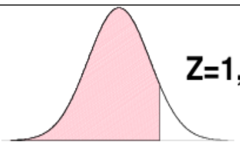
$\mu - \sigma \le x \le \mu + \sigma \longrightarrow$ **68,27%**

$\mu - 2\sigma \le x \le \mu + 2\sigma \longrightarrow$ **95,45%**

$\mu - 3\sigma \le x \le \mu + 3\sigma \longrightarrow$ **99,73%**

$\mu - 1{,}96\sigma \le x \le \mu + 1{,}96\sigma \longrightarrow$ **95%**

Area sottostante la curva della distribuzione normale standardizzata



Z=1,23

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

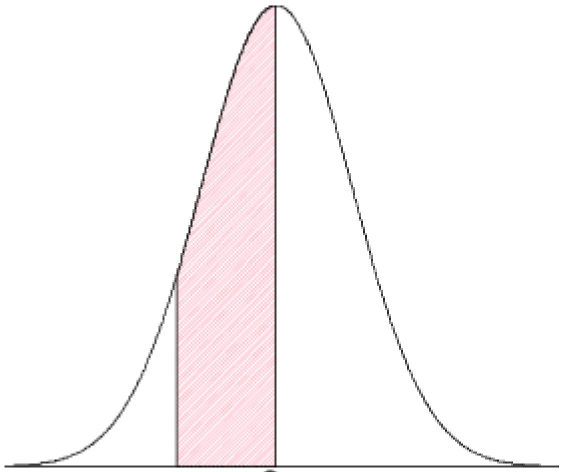
Esempio

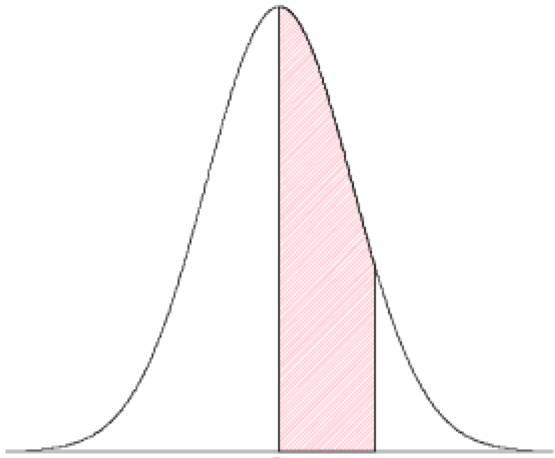
Una compagnia di trasporto ha calcolato che i propri camion coprono una distanza media annua di 50.000 km., con uno s.q.m. di 12.000 km. Si suppone che la distribuzione segua un andamento approssimativamente normale. Qual è la probabilità che un camion copra una distanza: A) compresa tra 34.000 e 50.000 km? B) compresa tra 34.000 e 38.000 km? C) inferiore a 30.000 o superiore a 60.000 km?

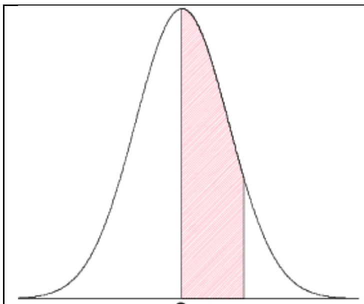
A) Compresa tra 34 e 50

$$z_1 = \frac{34-50}{12} = \frac{-16}{12} = -1{,}33$$

$$z_2 = \frac{50-50}{12} = \frac{0}{12} = 0$$







Z	0,00	0,01	0,02	0,03	0,04
0,0	0,5000	0,5040	0,5080	0,5120	0,5160
0,1	0,5398	0,5438	0,5478	0,5517	0,5557
0,2	0,5793	0,5832	0,5871	0,5910	0,5948
0,3	0,6179	0,6217	0,6255	0,6293	0,6331
0,4	0,6554	0,6591	0,6628	0,6664	0,6700
0,5	0,6915	0,6950	0,6985	0,7019	0,7054
0,6	0,7257	0,7291	0,7324	0,7357	0,7389
0,7	0,7580	0,7611	0,7642	0,7673	0,7704
0,8	0,7881	0,7910	0,7939	0,7967	0,7995
0,9	0,8159	0,8186	0,8212	0,8238	0,8264
1,0	0,8413	0,8438	0,8461	0,8485	0,8508
1,1	0,8643	0,8665	0,8686	0,8708	0,8729
1,2	0,8849	0,8869	0,8888	0,8907	0,8925
1,3	0,9032	0,9049	0,9066	0,9082	0,9099
1,4	0,9192	0,9207	0,9222	0,9236	0,9251
1,5	0,9332	0,9345	0,9357	0,9370	0,9382

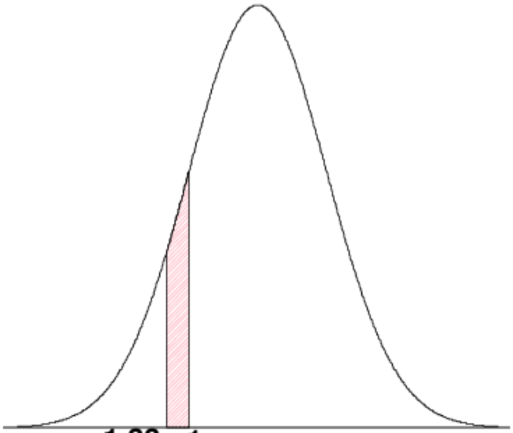
$P(-1{,}33 \leq z \leq 0) =$

$= 0{,}9082 - 0{,}5 = 0{,}4082$

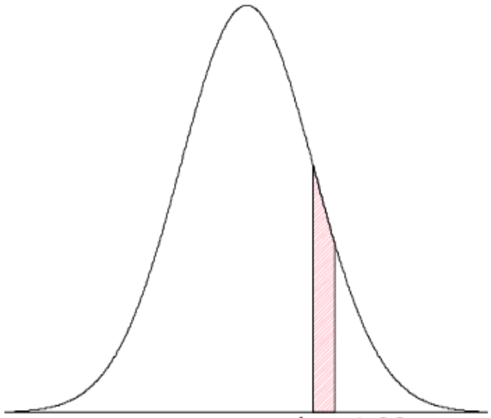
$\boxed{40{,}82\%}$

B) Compresa tra 34 e 38

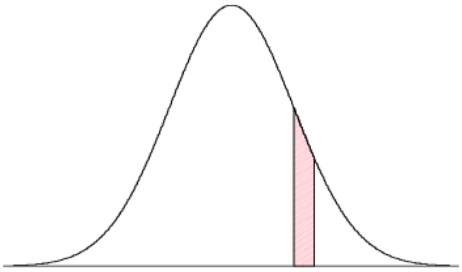
$z_1 = \dfrac{34-50}{12} = \dfrac{-16}{12} = -1{,}33$	$z_2 = \dfrac{38-50}{12} = \dfrac{-12}{12} = -1$



- 1,33 -1



+1 +1,33



+1 +1,33

Z	0,00	0,01	0,02	0,03	0,04
0,0	0,5000	0,5040	0,5080	0,5120	0,5160
0,1	0,5398	0,5438	0,5478	0,5517	0,5557
0,2	0,5793	0,5832	0,5871	0,5910	0,5948
0,3	0,6179	0,6217	0,6255	0,6293	0,6331
0,4	0,6554	0,6591	0,6628	0,6664	0,6700
0,5	0,6915	0,6950	0,6985	0,7019	0,7054
0,6	0,7257	0,7291	0,7324	0,7357	0,7389
0,7	0,7580	0,7611	0,7642	0,7673	0,7704
0,8	0,7881	0,7910	0,7939	0,7967	0,7995
0,9	0,8159	0,8186	0,8212	0,8238	0,8264
1,0	0,8413	0,8438	0,8461	0,8485	0,8508
1,1	0,8643	0,8665	0,8686	0,8708	0,8729
1,2	0,8849	0,8869	0,8888	0,8907	0,8925
1,3	0,9032	0,9049	0,9066	0,9082	0,9099
1,4	0,9192	0,9207	0,9222	0,9236	0,9251
1,5	0,9332	0,9345	0,9357	0,9370	0,9382

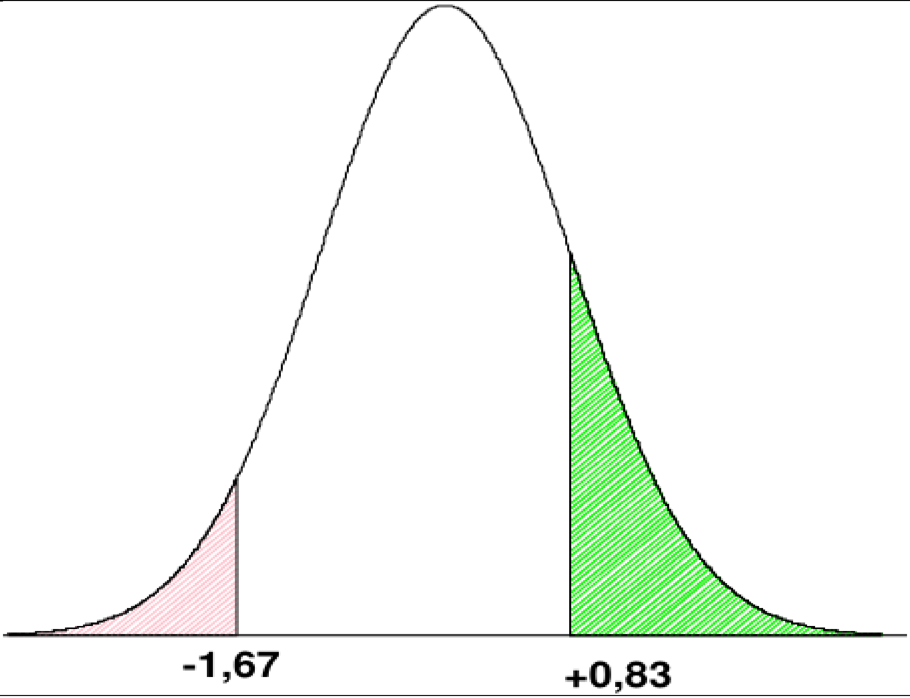
$P(-1{,}33 \leq z \leq -1) =$

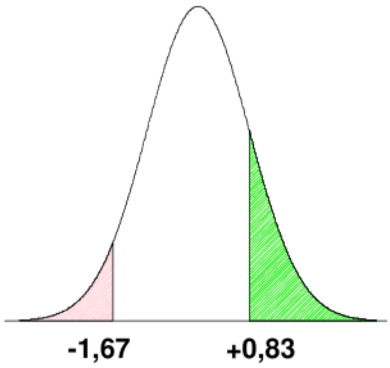
$= 0{,}9082 - 0{,}8413 = 0{,}0669$

6,69%

C) Inferiore a 30 o superiore a 60

$z_1 = \frac{30-50}{12} = \frac{-20}{12} = -1,67$	$z_2 = \frac{60-50}{12} = \frac{10}{12} = 0,83$





Z	0,02	0,03	0,04	0,05	0,06	0,07
0,0	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279
0,1	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675
0,2	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064
0,3	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443
0,4	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808
0,5	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157
0,6	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486
0,7	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794
0,8	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078
0,9	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340
1,0	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577
1,1	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790
1,2	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980
1,3	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147
1,4	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292
1,5	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418
1,6	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525
1,7	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616
1,8	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693

$P(z \leq -1,67) + P(z \geq 0,83) =$

$= (1 - 0,7967) + (1 - 0,9525) =$

$= 0,2033 + 0,0475 = 0,2508$

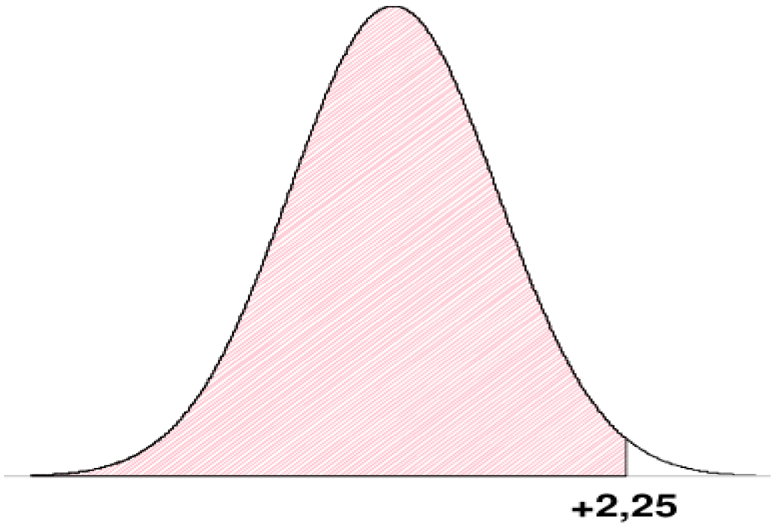
$\boxed{25,08\%}$

Esempio 2

Il risultato ottenuto ad un test da un gruppo di studenti si distribuisce come la curva normale con media pari a 73 e s.q.m. pari a 8. A) Qual è la probabilità che uno studente abbia riportato un voto non superiore a 91? B) Qual è la probabilità che uno studente abbia riportato un voto compreso tra 65 e 89? C) e compreso tra 81 e 89?

A) Non superiore a 91

$$z_1 = \frac{91-73}{8} = \frac{18}{8} = 2{,}25$$



$P(z \leq +2{,}25) = 0{,}9878$

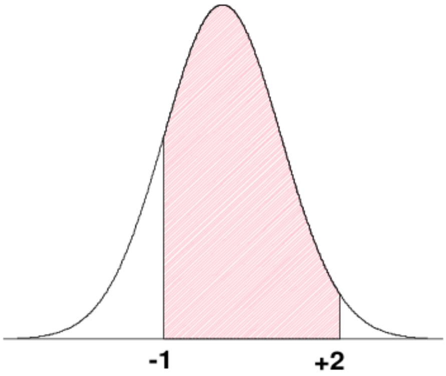
98,78%

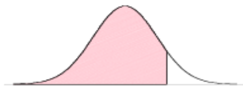
Z	0,00	0,01	0,02	0,03	0,04	0,05
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	$\boxed{0{,}9878}$
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946

B) Compreso tra 65 e 89

$$z_1 = \frac{65-73}{8} = \frac{-8}{8} = -1$$

$$z_2 = \frac{89-73}{8} = \frac{16}{8} = 2$$





z	0,00	0,01	0,02	0,03
0,6	0,7257	0,7291	0,7324	0,7357
0,7	0,7580	0,7611	0,7642	0,7673
0,8	0,7881	0,7910	0,7939	0,7967
0,9	0,8159	0,8186	0,8212	0,8238
1,0	0,8413	0,8438	0,8461	0,8485
1,1	0,8643	0,8665	0,8686	0,8708
1,2	0,8849	0,8869	0,8888	0,8907
1,3	0,9032	0,9049	0,9066	0,9082
1,4	0,9192	0,9207	0,9222	0,9236
1,5	0,9332	0,9345	0,9357	0,9370
1,6	0,9452	0,9463	0,9474	0,9484
1,7	0,9554	0,9564	0,9573	0,9582
1,8	0,9641	0,9649	0,9656	0,9664
1,9	0,9713	0,9719	0,9726	0,9732
2,0	0,9772	0,9778	0,9783	0,9788
2,1	0,9821	0,9826	0,9830	0,9834
2,2	0,9861	0,9864	0,9868	0,9871
2,3	0,9893	0,9896	0,9898	0,9901
2,4	0,9918	0,9920	0,9922	0,9925

$P(-1 \leq z \leq +2) =$

$= (0{,}9772 - 0{,}5) + (0{,}8413 - 0{,}5) =$

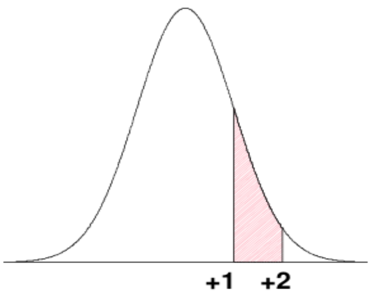
$= 0{,}4772 + 0{,}3413 = 0{,}8185$

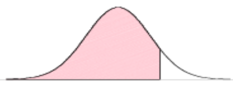
81,85%

B) Compreso tra 81 e 89

$$z_1 = \frac{81-73}{8} = \frac{8}{8} = +1$$

$$z_2 = \frac{89-73}{8} = \frac{16}{8} = +2$$





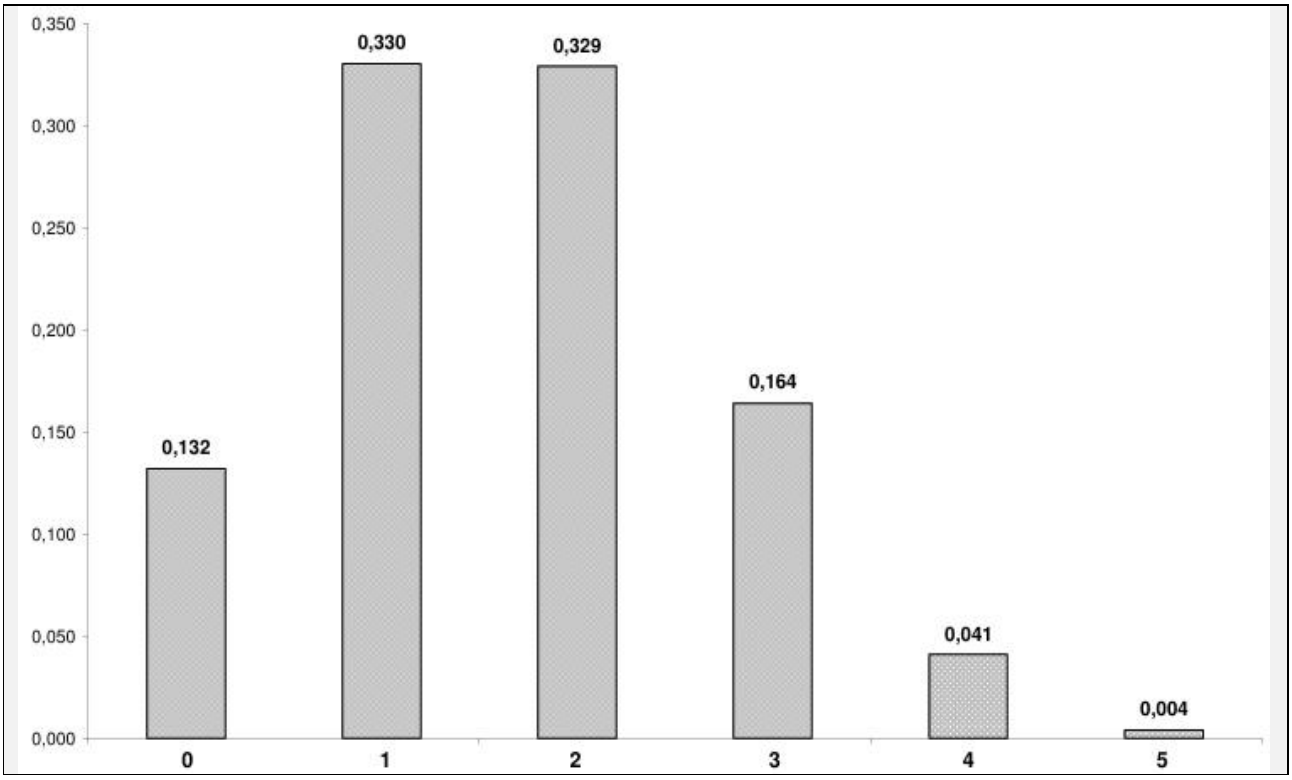
z	0,00	0,01	0,02	0,03
0,6	0,7257	0,7291	0,7324	0,7357
0,7	0,7580	0,7611	0,7642	0,7673
0,8	0,7881	0,7910	0,7939	0,7967
0,9	0,8159	0,8186	0,8212	0,8238
1,0	0,8413	0,8438	0,8461	0,8485
1,1	0,8643	0,8665	0,8686	0,8708
1,2	0,8849	0,8869	0,8888	0,8907
1,3	0,9032	0,9049	0,9066	0,9082
1,4	0,9192	0,9207	0,9222	0,9236
1,5	0,9332	0,9345	0,9357	0,9370
1,6	0,9452	0,9463	0,9474	0,9484
1,7	0,9554	0,9564	0,9573	0,9582
1,8	0,9641	0,9649	0,9656	0,9664
1,9	0,9713	0,9719	0,9726	0,9732
2,0	0,9772	0,9778	0,9783	0,9788
2,1	0,9821	0,9826	0,9830	0,9834
2,2	0,9861	0,9864	0,9868	0,9871
2,3	0,9893	0,9896	0,9898	0,9901
2,4	0,9918	0,9920	0,9922	0,9925

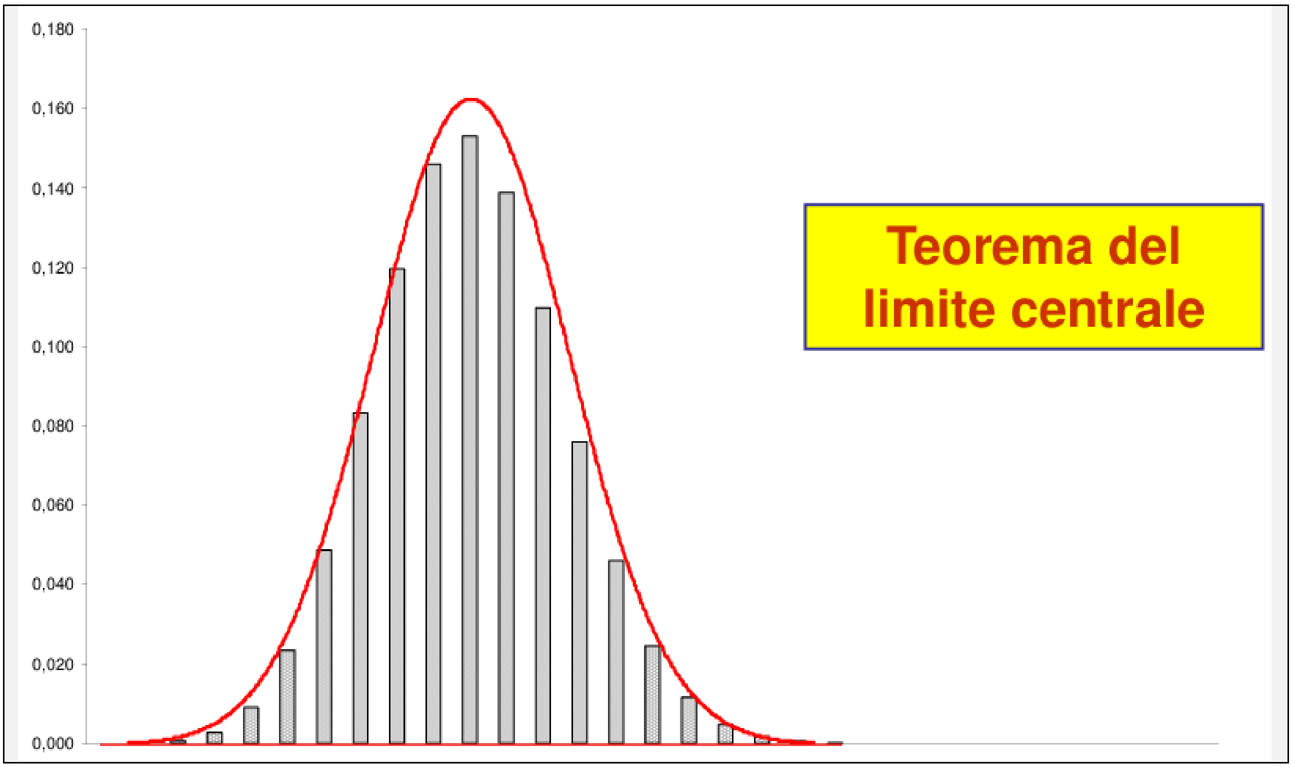
$P(+1 \leq z \leq +2) =$

$= (0{,}9772 - 0{,}8413) = 0{,}1359$

13,59%

Binomiale e Normale





Teorema del limite centrale:

- Il Teorema del Limite Centrale è uno dei risultati fondamentali della teoria della probabilità e della statistica. Esso afferma che, sotto certe condizioni, la somma (o la media) di un gran numero di variabili casuali indipendenti e identicamente distribuite tende a seguire una distribuzione normale (gaussiana), indipendentemente dalla forma della distribuzione originale delle variabili stesse.
- Questo teorema ha numerose applicazioni pratiche, in quanto giustifica l'uso della distribuzione normale in molte situazioni della vita reale, come nel campionamento e nelle inferenze statistiche, anche quando le popolazioni di partenza non seguono una distribuzione normale. Inoltre, è alla base di molte tecniche statistiche, come il calcolo degli intervalli di confidenza e i test di ipotesi.