

DISTRIBUZIONE NORMALE

La **distribuzione normale** o di **Gauss** è la distribuzione di **variabile continua** di gran lunga più usata perché è in grado di descrivere la maggior parte dei fenomeni che accadono in natura. Il termine **normale**, deriva dal fatto che variabili continue come il peso, l'altezza, la larghezza delle foglie di una pianta, ecc assumono **valori normali**, ossia valori che oscillano sempre in un certo range.

Ti ho spiegato questo concetto in un video che trovi in fondo alla pagina ([clicca qui](#)). Se invece sei interessato alle applicazioni, puoi consultare gli **esercizi sulla distribuzione normale** accuratamente selezionati e visionabili al seguente link:

VAI AGLI ESERCIZI

Questa distribuzione, è particolarmente importante non solo perché molti fenomeni osservati nella vita reale sono descritti dalla distribuzione normale, ma anche perché può essere utilizzata come *distribuzione limite* in quanto è una distribuzione alla quale tendono altre distribuzioni. Se vuoi approfondire leggi:

- [Approssimazione distribuzione binomiale con normale](#)

- [Approssimazione distribuzione di Poisson con normale](#)

Inoltre, all'aumentare della numerosità del campione, si dimostra che i dati osservati seguono un andamento normale. Quest'ultima caratteristica viene messa in evidenza nel famoso **Teorema del Limite Centrale** di cui te ne ho parlato in [questo articolo](#).

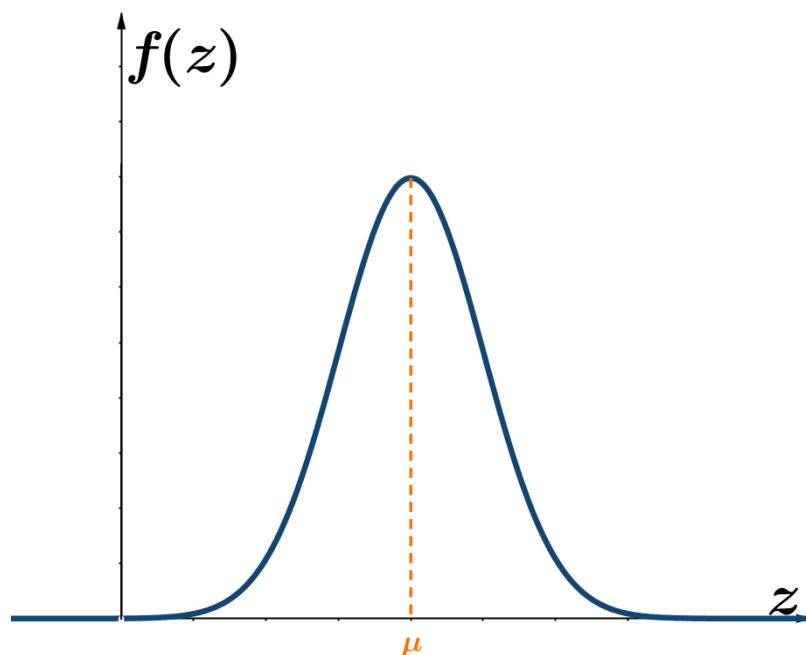
In alternativa ho registrato un video che trovi a fine pagina in cui ti spiego il **significato concreto della distribuzione normale e il motivo per cui è così famosa!**

Definizione e grafico della distribuzione normale

Una variabile aleatoria continua X con media μ e scarto quadratico medio σ ha distribuzione normale (e si indica con $X \sim N(\mu, \sigma)$) se la sua funzione di densità è:

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

dove μ e σ sono i parametri della distribuzione normale che indicano rispettivamente il valore medio e la deviazione standard (o scarto quadratico medio). Ti parlerò di questi in dettaglio più avanti nell'articolo. Il generico **grafico della distribuzione normale** (detto anche **funzione di densità della normale**) è rappresentato qui sotto.



I valori che si trovano vicino il centro della curva (quindi attorno al valore medio) sono quelli più comuni nella popolazione mentre i valori

che stanno all'estremità della curva (quindi vicino le code) sono quelli più rari.

Caratteristiche della distribuzione normale

Di seguito ti elenco le caratteristiche fondamentali della curva normale di Gauss facendo riferimento alla figura qui sotto.

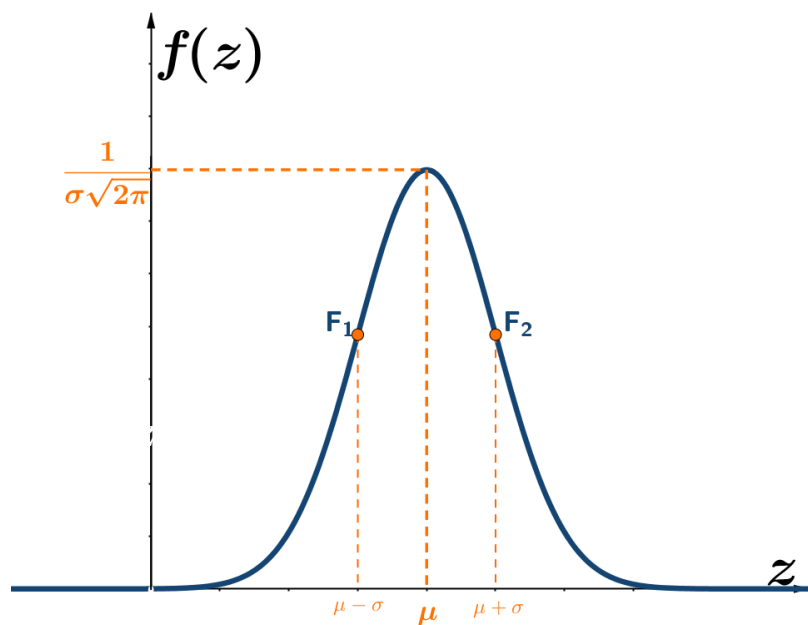
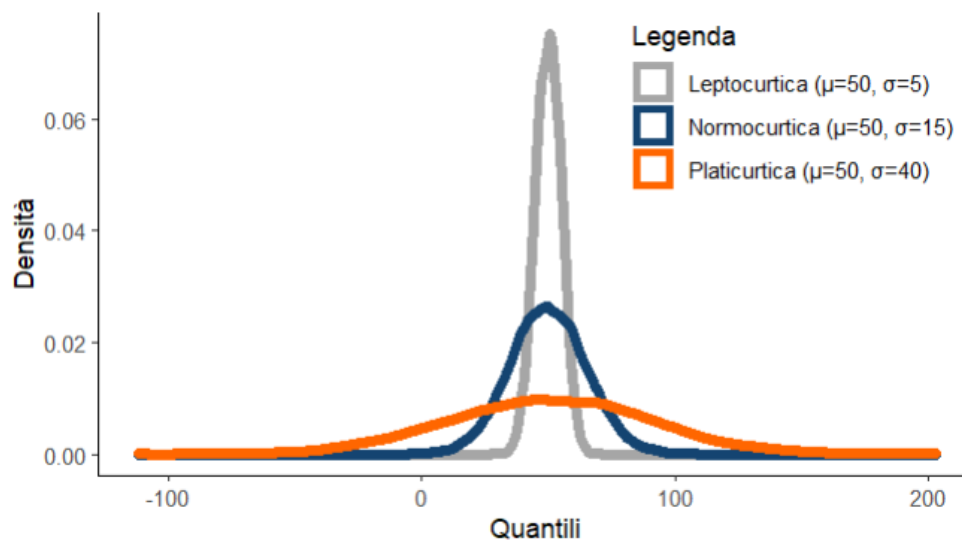


FIGURA 2

Forma

La distribuzione normale, come già visto, ha sempre la forma a campana ma questa può variare in base alla deviazione standard. Infatti, esistono 3 tipi di curve normali a seconda del valore della deviazione standard:

- curva **leptocurtica**, cioè più appuntita della curva normale (con **bassa deviazione standard**)
- curva **normocurtica** o **mesocurtica**, cioè come una curva normale (con **regolare deviazione standard**)
- curva **platicurtica**, cioè più piatta della curva normale (con **alta deviazione standard**)



Il plot soprastante è stato realizzato completamente in R Studio. Se vuoi imparare a **graficare la distribuzione normale in R** contattami cliccando [qui](#).

Curtosi

Per comprendere quale delle tre forme appena descritte assume la distribuzione normale, esiste un coefficiente di forma chiamato **curtosi**. Te ne ho parlato approfonditamente in [questa lezione](#).

Simmetria

La distribuzione normale è simmetrica rispetto al valor medio μ (nella figura 2 vedi che la retta $x = \mu$ spacca la curva esattamente a metà)

Valore massimo e moda

Il **valore massimo della distribuzione normale** si ottiene in corrispondenza del valore medio μ nel quale, la funzione di densità

$f(x)$ assume valore $\frac{1}{\sigma\sqrt{2\pi}}$ come puoi vedere nel grafico sopra. Il

valore massimo della funzione di densità della distribuzione normale rappresenta la **moda** che quindi coincide con il valore medio μ .

Asintoti e monotonia

Dalla figura 2 si può vedere che tanto più x si allontana da μ quanto più $f(x)$ decresce e tende asintoticamente a zero.

Punti di flesso

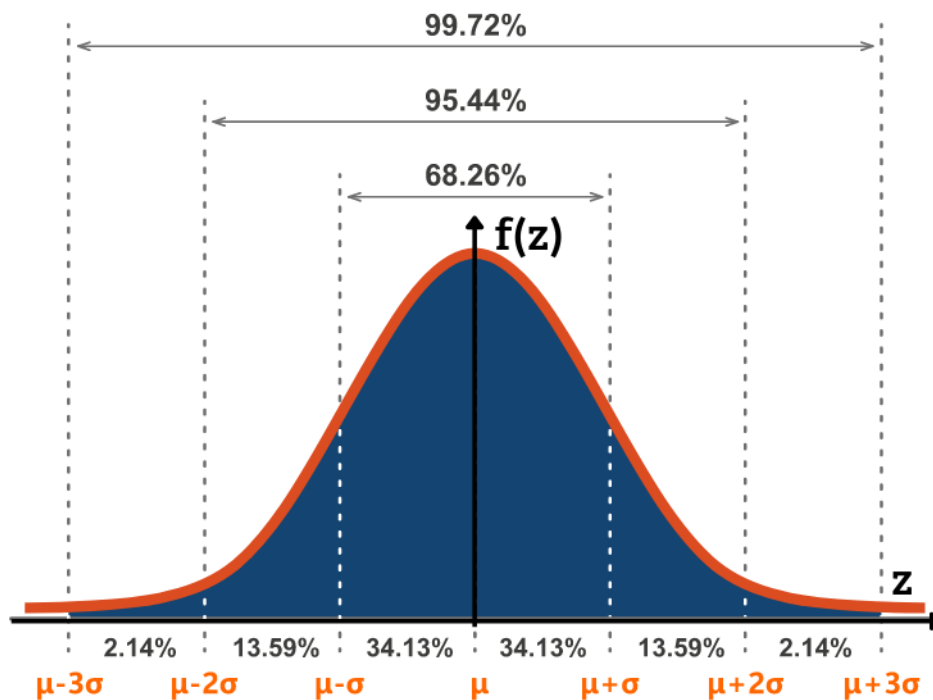
Se stai studiando statistica matematica ti è utile sapere che i punti $\mu - \sigma$ e $\mu + \sigma$ sono punti di flesso della curva normale (osserva la figura 2).

Deviazione standard

Esiste una legge empirica chiamata **disuguaglianza di Cebicev** tramite la quale è possibile dedurre il valore di probabilità in intervalli specifici dettati dal numero di deviazioni standard delle quali ci si discosta dal valore medio. In altre parole, se conosciamo la deviazione standard oltre che la media dei nostri valori normali, dal **teorema di Cebicev** segue che:

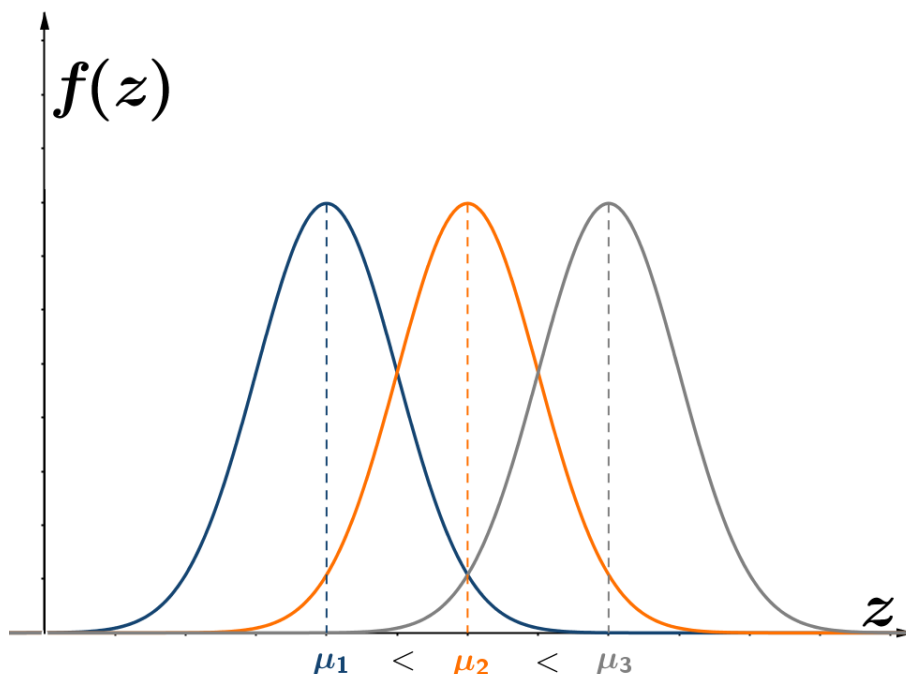
- Il **68, 26** delle osservazioni rientra in $(\mu - \sigma, \mu + \sigma)$ (detto anche "più o meno una deviazione standard").
- Il **95, 44%** delle osservazioni rientra in $(\mu - 2\sigma, \mu + 2\sigma)$ (detto anche "più o meno due deviazioni standard")
- Il **99, 72%** delle osservazioni rientra in $(\mu - 3\sigma, \mu + 3\sigma)$ (detto anche "più o meno tre deviazioni standard")

La figura qui sotto rappresenta quanto appena detto.

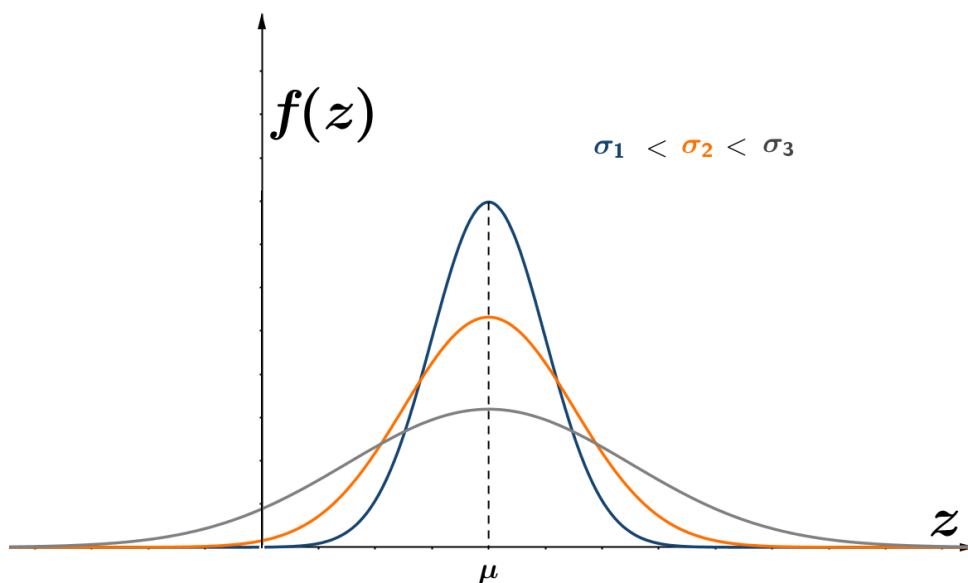


I parametri della distribuzione normale

μ e σ , come detto sono i parametri che definiscono la distribuzione normale. Il valore medio o **valore atteso** μ denota il **centro della distribuzione normale**, ossia quel valore a cui è corrisponde il valore massimo del grafico. Inoltre, questo caratterizza la posizione della curva sull'asse delle ascisse: al variare di μ la curva si sposta lungo l'asse x , ma resta invariata nella sua forma (vedi grafico qui sotto).



La **deviazione standard** σ , invece, caratterizza la forma della curva, in quanto è una misura della *dispersione* dei valori attorno al valore medio: al variare di σ la curva cambia forma: infatti al crescere di σ la curva si appiattisce e si allarga, mentre al diminuire di σ la curva si restringe e si alza.



Funzione di ripartizione cumulata normale

Per calcolare la probabilità sottesa dalla curva in un determinato intervallo di valori z , devi calcolare la **funzione di ripartizione cumulata della distribuzione normale** che è data da:

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Tuttavia, il calcolo della funzione cumulata non può essere eseguito mediante la formula suddetta perché questo richiederebbe svolgere un integrale non di facile risoluzione. Per tale motivo, si fa uso di una tavola (figura qui sotto) in cui è possibile leggere i valori di probabilità in corrispondenza dell'area sottesa identificata dai valori z.

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
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	0.99995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	0.99997									
5.0	0.9999997									
6.0	0.999999999									

Imparare a leggere la tavola della distribuzione normale è facile.

Consulta [questo articolo](#) per saperne di più oppure vai direttamente agli esercizi svolti cliccando sul bottone blu qui sotto.

VAI AGLI ESERCIZI

Se hai appena iniziato a studiare la distribuzione normale, ti consiglio di completare questa lettura prima di andare a consultare gli esercizi.

Formula di standardizzazione

Sicuramente hai già sentito parlare di **distribuzione normale standard** o normale standardizzata ma se così non fosse non preoccuparti perché sto per spiegarti cos'è la standardizzazione e come standardizzare un valore.

Standardizzare una variabile X con media μ e deviazione standard σ significa applicare la seguente formula:

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

dove X può essere una variabile normale continua che assume infiniti valori oppure può anche essere un singolo valore.

Il significato della formula si può così spiegare:

1. sottrarre la media μ consente di centrare la variabile X nell'origine, ossia la nuova variabile Z ottenuta ha $\mu = 0$;
2. dividere per la deviazione standard σ , permette di normalizzare la variabile in modo che Z abbia $\sigma = 1$

Il risultato è una variabile Z avente distribuzione normale standard (media 0 e varianza 1) i cui valori da essa assunti sono numeri puri, adimensionali, senza unità di misura. Per indicare una variabile normale standardizzata Z si usa la seguente notazione: $Z \sim N(0, 1)$.

Esempio

Supponiamo che il valore $x = 4.1$ è assunto da una variabile casuale X avente media $\mu = 3$ e varianza $\sigma^2 = 4$. Calcolare il punteggio z corrispondente.

Svolgimento

Applico la formula di standardizzazione suddetta facendo attenzione a fare la radice della varianza al denominatore per ottenere la deviazione standard:

$$z = \frac{x - \mu}{\sigma} = \frac{4.1 - 3}{\sqrt{4}} = 0.55$$

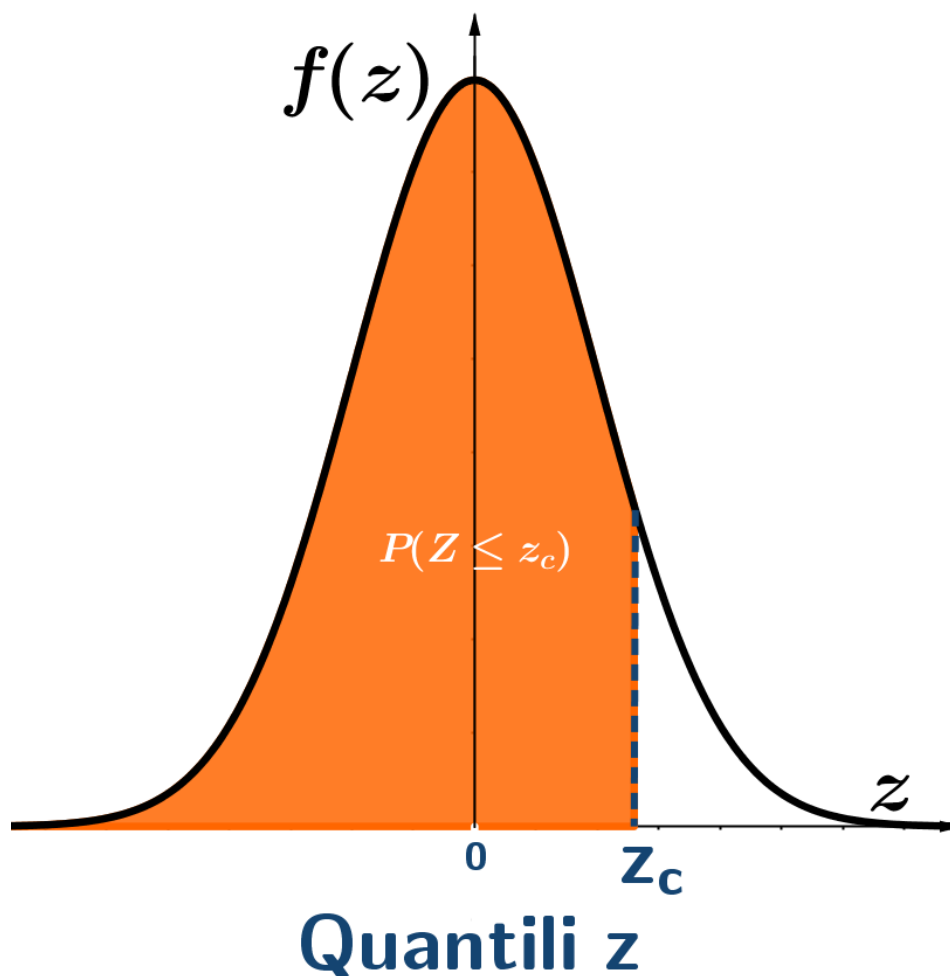
Qui sotto trovi l'espressione della funzione di densità normale standard

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Qui sotto invece la funzione cumulata della distribuzione normale standard:

$$\begin{aligned} F(z_c) &= P(Z \leq z_c) = \\ &= \int_{-\infty}^{z_c} f(z) dz = \\ &= \int_{-\infty}^{z_c} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \end{aligned}$$

Molto spesso la funzione cumulativa della normale viene indicata con il simbolo $\Phi(z_c)$ invece che $F(z_c)$. Come mostra la figura sottostante questa sta a indicare il valore dell'area sottesa dalla curva normale a sinistra del valore critico o quantile z_c .



Perché standardizzare?

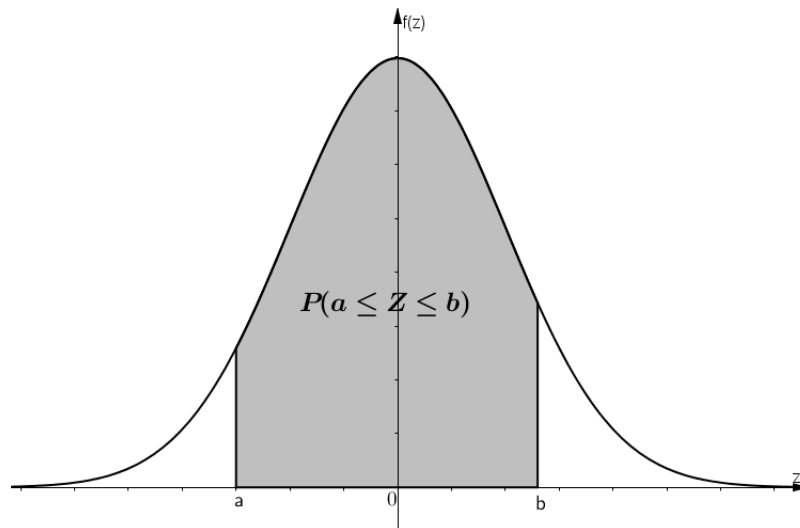
Ma a cosa serve standardizzare? I principali motivi per cui si standardizza sono:

- Confrontare elementi di diverse variabili e diverse unità di misura.
- Ai fini del calcolo delle probabilità permette di utilizzare semplicemente una tavola statistica anziché dover calcolare integrali difficili.
- Semplifica la vita perché, grazie alla tavola, non abbiamo bisogno di effettuare calcoli specifici per ogni valore di media e varianza: infatti, standardizzando i valori della distribuzione si ottiene sempre media nulla e varianza pari a 1.

Calcolo dell'area sottesa dalla curva normale

Per determinare la probabilità che la variabile causale normale assuma un valore compreso in un determinato intervallo (a, b) è

necessario calcolare l'area sotto la curva compresa tra a e b (vedi figura in basso).



Questo significherebbe calcolare il seguente integrale:

$$P(a \leq X \leq b) = \int_a^b f(x) dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

che riscritto in termini di funzione di ripartizione diventa:

$$P(a \leq X \leq b) = F(b) - F(a)$$

Come già detto il valore di tale probabilità si può ricavare facilmente consultando la tavola normale. Nella sezione esercizi (clicca sul bottone blu presente qui sotto) ti ho fornito **tutte le tipologie di problemi sul calcolo della probabilità di una variabile normale** che ti possono capitare in un esame di statistica.

[VAI AGLI ESERCIZI](#)

Guarda il video che ho registrato in cui ti faccio capire con un esempio pratico **cos'è la distribuzione normale e perché descrive molti fenomeni in natura**.