

LA DISTRIBUZIONE NORMALE (CURVA DI GAUSS)

DISTRIBUZIONE NORMALE o CURVA DI GAUSS

1. E' la più importante distribuzione statistica continua e trova numerose applicazioni nello studio dei fenomeni biologici.
2. Fu proposta da Gauss (1809) nell'ambito della teoria degli errori, ed è stata attribuita anche a Laplace (1812), che ne definì le proprietà principali in anticipo rispetto alla trattazione più completa fatta da Gauss.
3. Il nome “normale” deriva dalla convinzione che molti fenomeni fisico-biologici si distribuiscono con frequenze più elevate nei valori centrali e con frequenze progressivamente minori verso gli estremi della variabile.
4. E' detta anche **CURVA DEGLI ERRORI ACCIDENTALI** in quanto, soprattutto nelle discipline fisiche, la distribuzione degli errori commessi nel misurare ripetutamente una stessa grandezza, è molto bene approssimata da questa curva.

Esempio 1 - Distribuzione di 211 pazienti affetti da cirrosi biliare primitiva rispetto ai valori della concentrazione sierica di albumina

Valori albumina (g/l)	Frequenze		
	Assolute	Percentuali	Percentuali Cumulate
20-22	2	0,95	0,95
22-24	7	3,32	4,27
24-26	8	3,79	8,06
26-28	10	4,74	12,80
28-30	20	9,48	22,27
30-32	19	9,00	31,28
32-34	28	13,27	44,55
34-36	40	18,96	63,51
36-38	28	13,27	76,78
38-40	22	10,43	87,20
40-42	11	5,21	92,42
42-44	10	4,74	97,16
44-46	3	1,42	98,58
46-48	2	0,95	99,53
48-50	1	0,47	100,00
Totale	211	100,00	

1 D.S.

M=Me=Mo

1 D.S.

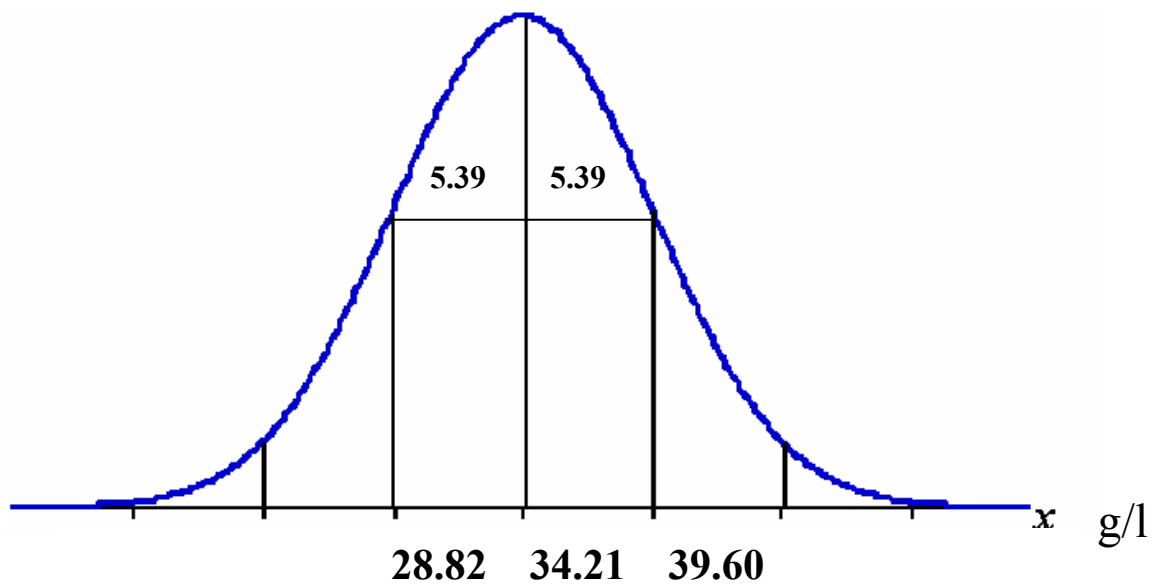
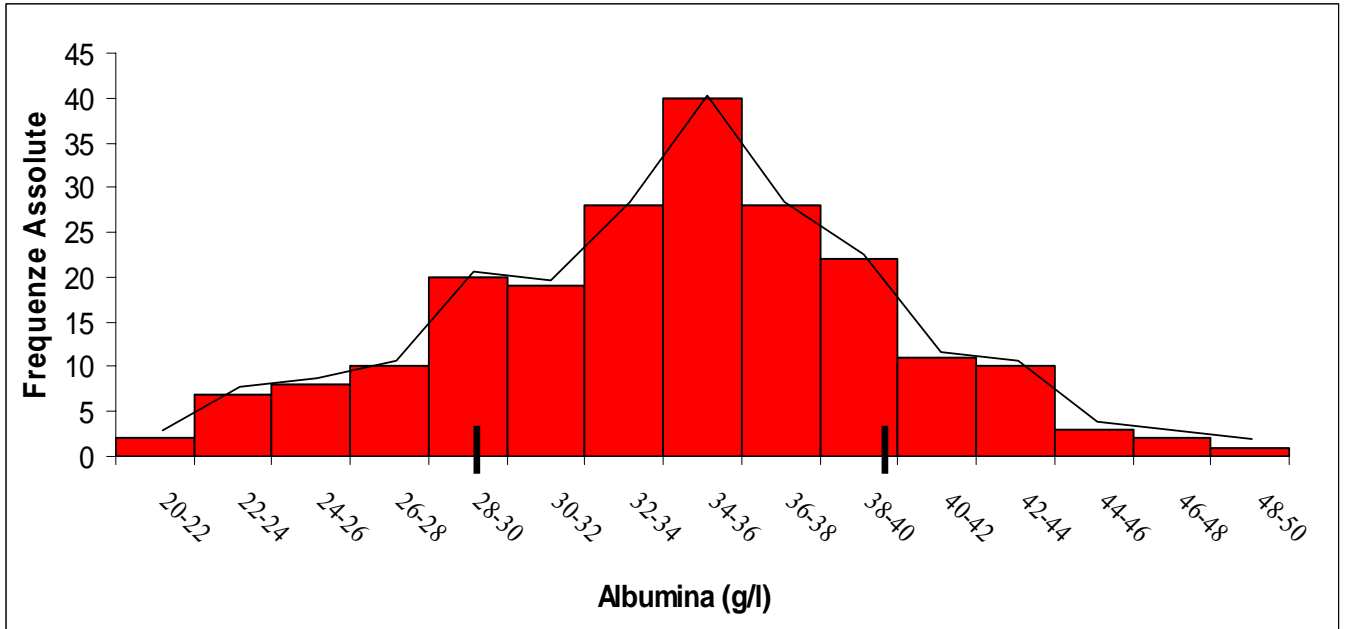
MEDIA ARITMETICA (μ)= 34.21 g/l

DEVIAZIONE STANDARD (σ) = 5.39 g/l

I: 34.21 – 5.39 = 28.82 g/l

L: 34.21 + 5.39 = 39.60 g/l

Figura 1 - Istogramma di frequenza della concentrazione sierica di albumina in 211 pazienti con cirrosi biliare primitiva

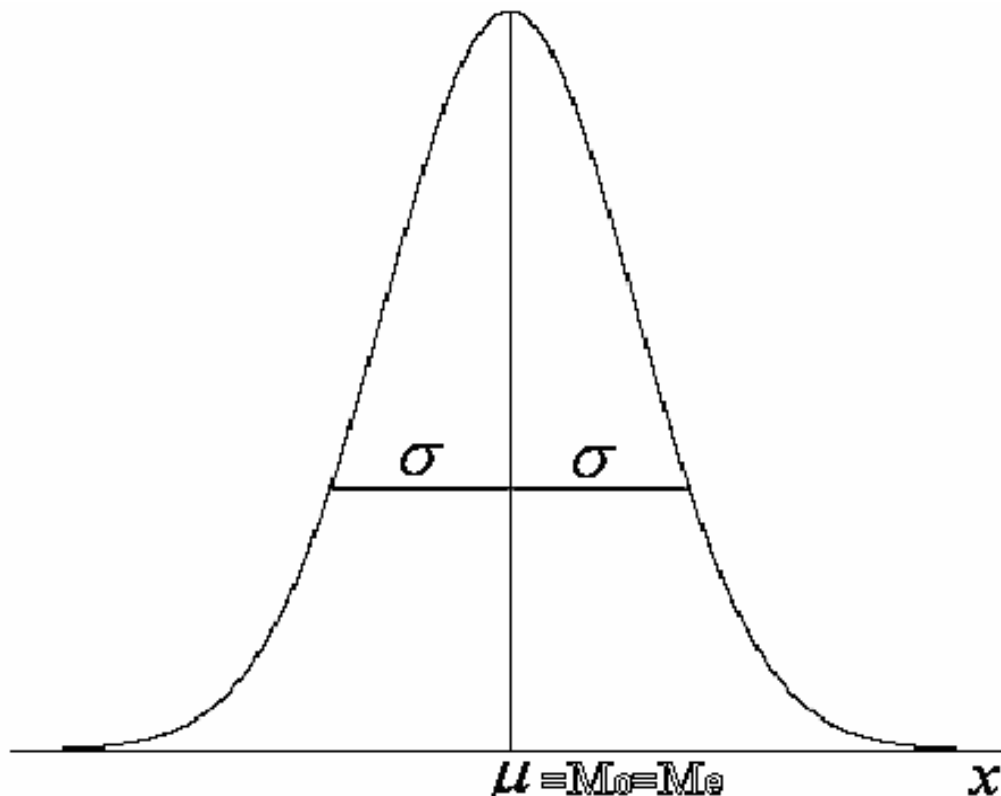


Quando le distribuzioni di frequenza prima risultano crescenti, raggiungono un massimo e poi cominciano a decrescere fino ad arrivare allo zero, si può parlare di variabili che tendono a distribuirsi “normalmente”, ossia che seguono un andamento secondo la curva di Gauss la cui espressione analitica è del tipo:

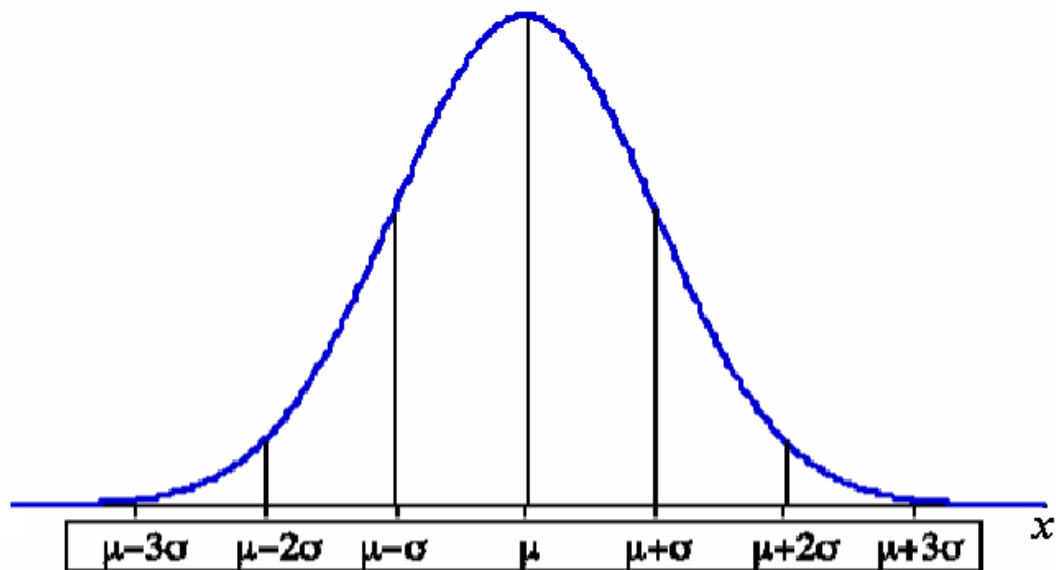
$$y_i = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\sum (x_i - \mu)^2 f_i}{2\sigma^2}\right); \quad \mu \sim \bar{x}$$

$\sigma \sim \text{D.S.}$

→ la Y dipende solamente da due parametri: μ e σ .

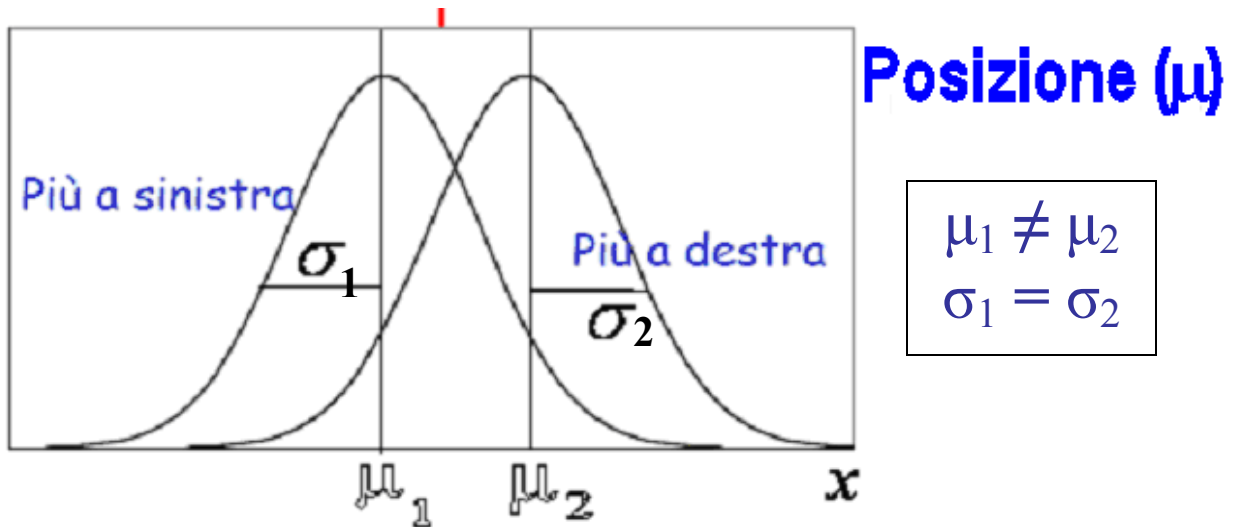


CARATTERISTICHE DELLA DISTRIBUZIONE NORMALE



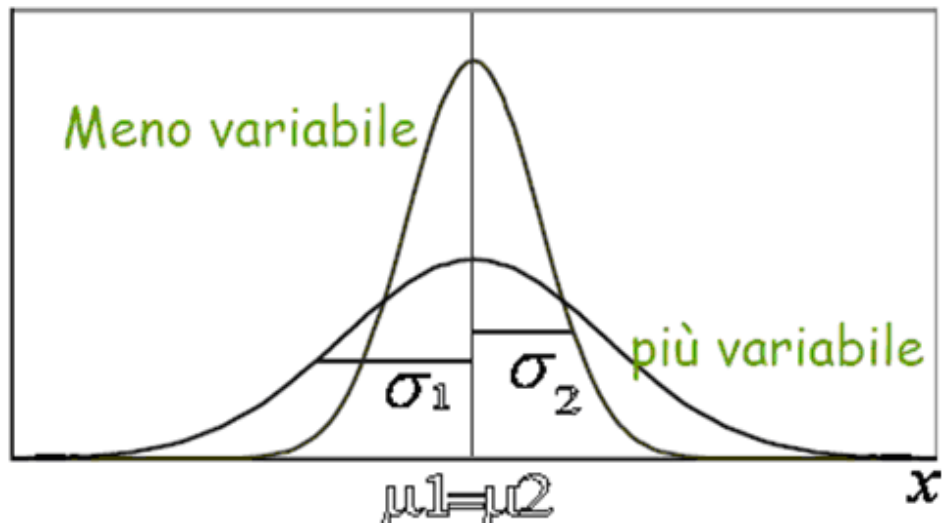
1. è **simmetrica** rispetto al valore medio (μ)
2. il valore di $x = \mu$ (media aritmetica) coincide anche con la moda e la mediana della distribuzione
3. è **asintotica** all'asse delle x da entrambi i lati
4. è **crescente** per $x < \mu$ e **decescente** per $x > \mu$
5. possiede due punti di flesso per $x = \mu \pm \sigma$
6. l'**area sotto la curva** è = 1 (essendo = 1 la probabilità del verificarsi di un valore di x compreso tra $-\infty$ e $+\infty$)

I 2 parametri di una Distribuzione Normale ne caratterizzano la POSIZIONE e la FORMA



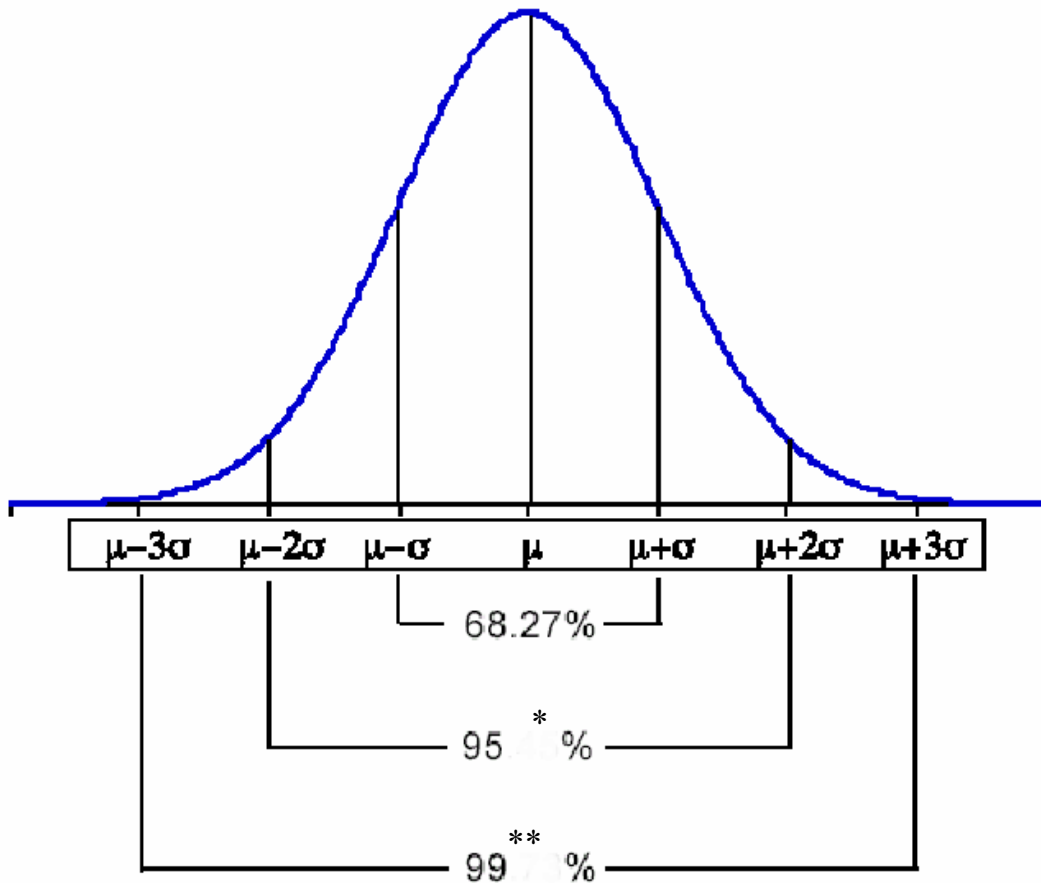
Forma (σ)

$$\begin{matrix} \sigma_1 \neq \sigma_2 \\ \mu_1 = \mu_2 \end{matrix}$$



AREE (PROBABILITA') PER INTERVALLI NOTI

Area sotto la curva negli intervalli $\mu \pm k \sigma$, per $k = 1, 2, 3$



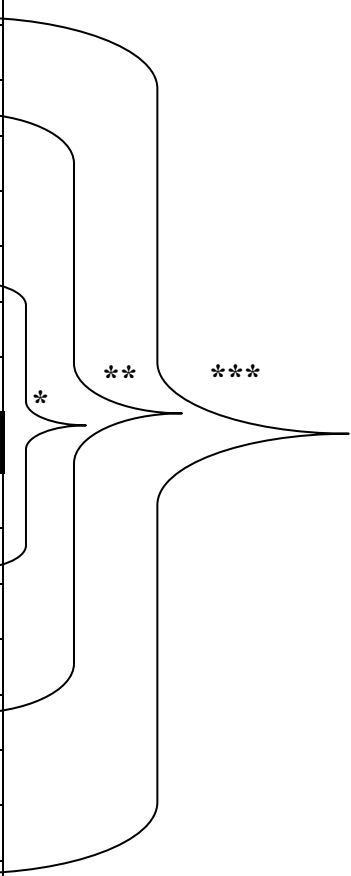
N.B. In teoria risultano:

* $\mu \pm 1.96\sigma$

** $\mu \pm 2.58\sigma$

Per la distribuzione dell'esempio 1 si ha:

Valori albumina (g/l)	Frequenze (Fr)		
	Assolute	Percentuali	Fr% Cumulate
20-22•	2	0,95	0,95
22-24•	7	3,32	4,27
24-26	8	3,79	8,06
26-28	10	4,74	12,80
28-30•	20	9,48	22,27
30-32	19	9,00	31,28
32-34	28	13,27	44,55
34-36	40	18,96	63,51
36-38	28	13,27	76,78
38-40•	22	10,43	87,20
40-42	11	5,21	92,42
42-44	10	4,74	97,16
44-46•	3	1,42	98,58
46-48	2	0,95	99,53
48-50•	1	0,47	100,00
Totale	211	100,00	



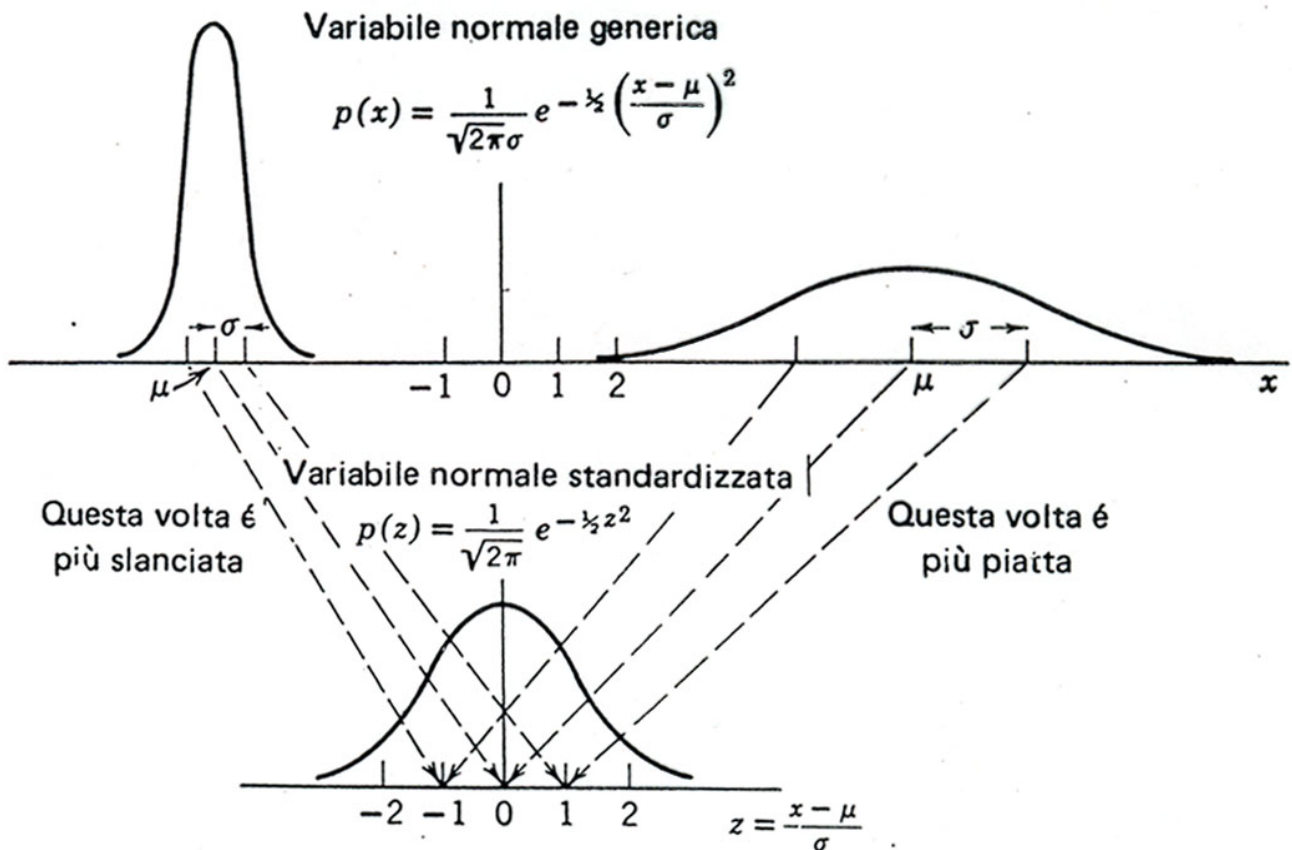
* $Fr(\mu \pm \sigma) = (34.21 \pm 5.39) = Fr\%(39.6) - Fr\%(28.8) = 64.93\%$ del totale

** $Fr(\mu \pm 2\sigma) = (34.21 \pm 10.78) = Fr\%(44.9) - Fr\%(23.4) = 94.31\%$

*** $Fr(\mu \pm 3\sigma) = (34.21 \pm 16.17) = Fr\%(50.4) - Fr\%(18.0) = 100\%$

DISTRIBUZIONE NORMALE STANDARDIZZATA

Una distribuzione Normale che ha media = 0 e DS = 1 è chiamata distribuzione normale standardizzata.



La distribuzione gaussiana dipende dai parametri μ e σ e questi a loro volta dai valori e dall'unità di misura della variabile x_i in esame.

Volendo una distribuzione normale standardizzata, ossia che non dipenda dall'unità di misura della variabile, si può trasformare quest'ultima mediante la relazione:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

La variabile z ha media $\mu=0$ e $\sigma=1$ e la funzione della curva normale standardizzata è:

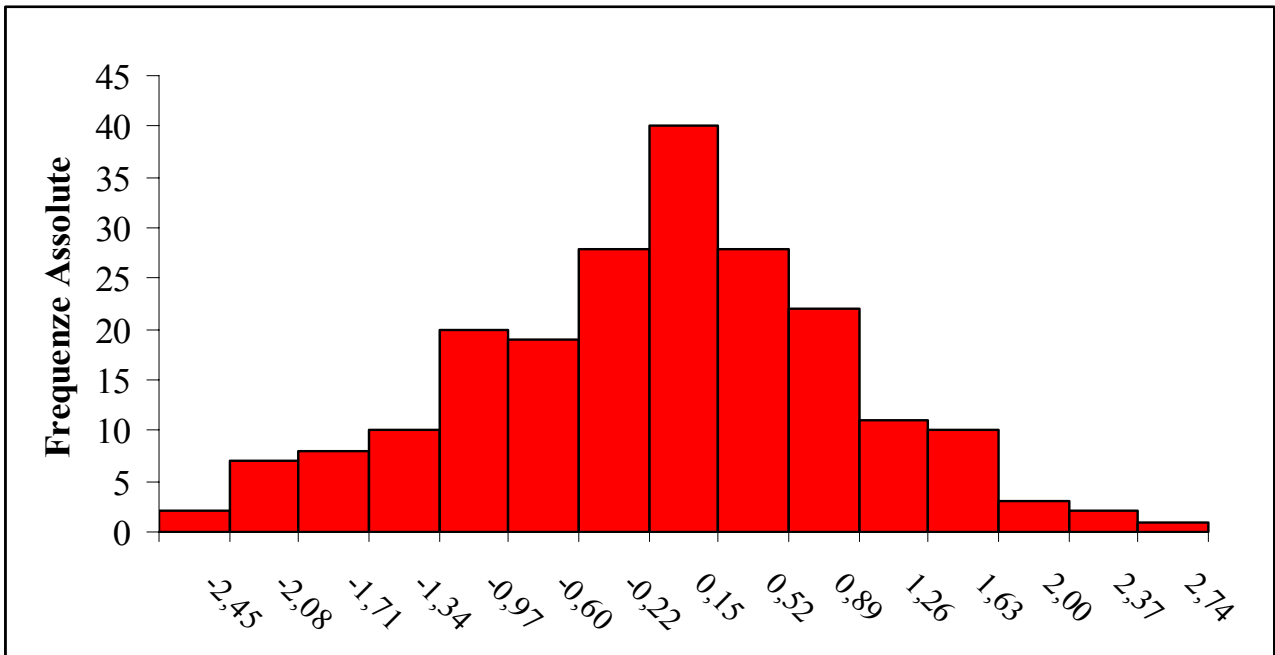
$$z = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Per la distribuzione di frequenze dell'esempio 1, standardizzando i valori centrali di ogni classe di albumina, si ha la seguente tabella:

Valori albumina		Frequenze	$Z_i f_i$	$Z_i^2 f_i$
Grezzi (X)	Standardizzati (Z)			
21	-2.45	2	-4.90	12.01
23	-2.08	7	-14.60	30.28
25	-1.71	8	-13.70	23.39
27	-1.34	10	-13.40	17.96
29	-0.97	20	-19.40	18.82
31	-0.60	19	-11.40	6.84
33	-0.22	28	-6.20	1.36
35	0.15	40	6.00	0.90
37	0.52	28	14.50	7.57
39	0.89	22	19.50	17.42
41	1.26	11	13.86	17.46
43	1.63	10	16.30	26.57
45	2.00	3	6.00	12.00
47	2.37	2	4.70	11.23
49	2.74	1	2.74	7.51
Totale	2.19	211	0	211.32

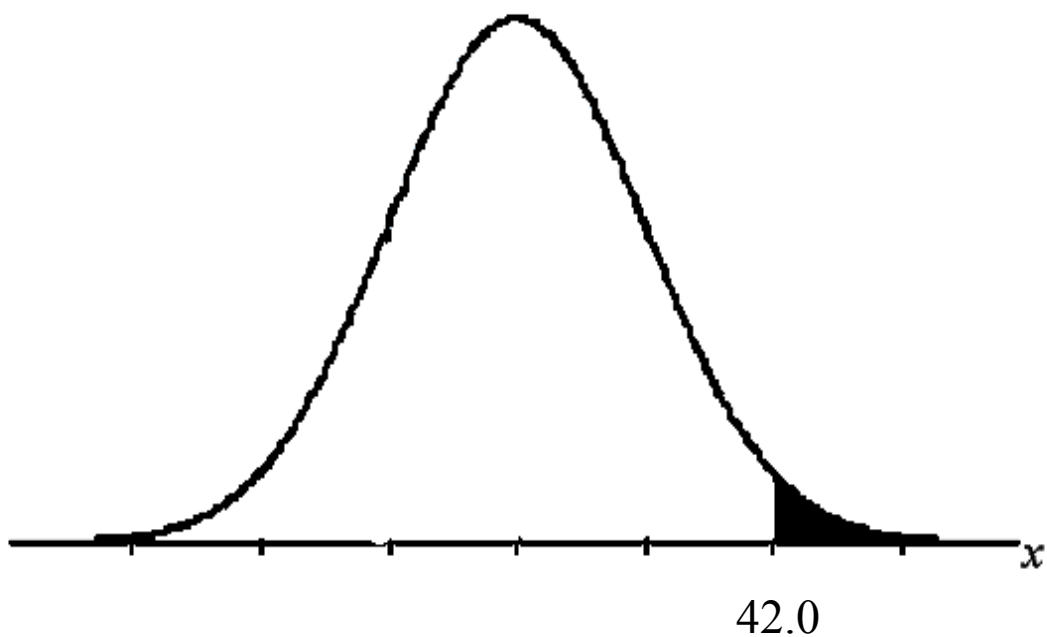
MEDIA ARITMETICA = 0

DEVIAZIONE STANDARD $\sqrt{211.32/211} = 1$



Ora ci chiediamo: qual è la probabilità per un paziente affetto da cirrosi biliare primitiva di avere un valore di albumina ≥ 42.0 g/l?

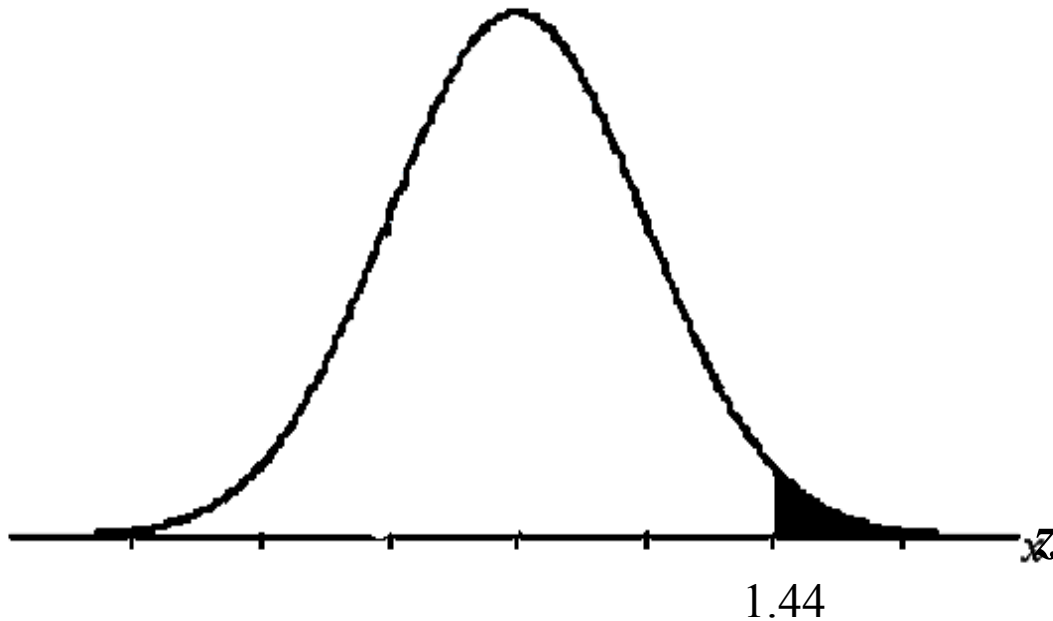
Cioè, qual è la $P(X \geq 42.0)$



Trasformando il valore di 42.0 g/l in valori di Z

$$Z = (X - \mu) / \sigma = (42.00 - 34.21) / 5.39 = 1.44$$

Dalle Tavole della Distribuzione Normale Standardizzata si ha:



$$P(Z \geq 1.44) = 1 - 0.9251 = 0.0749 = 7.49\%$$

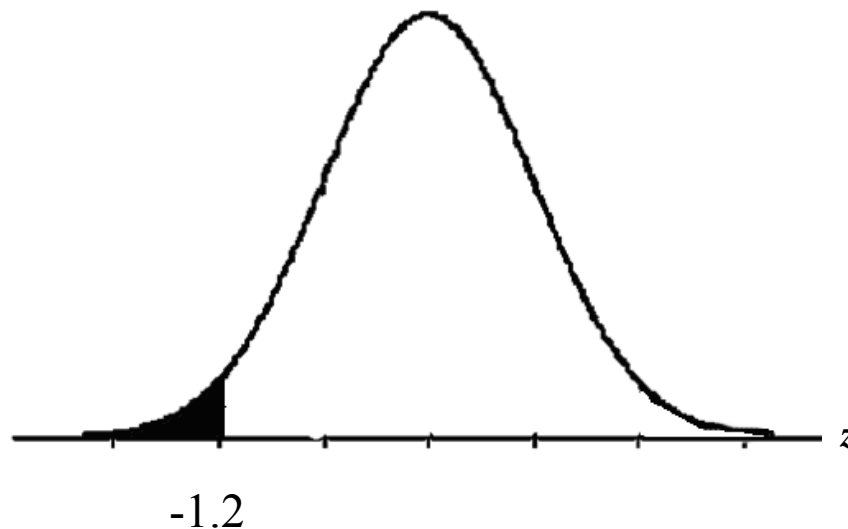
Un paziente con albumina > 42 g/l ha il 7.5% di probabilità di essere affetto da cirrosi biliare primitiva, e il 92.5% dei pazienti cirrotici presenta un valore < 42 g/l

Esempio 1. Dalla letteratura scientifica risulta che in una popolazione (P) apparentemente sana il valore medio dell'HDL-colesterolo è di 57 mg/100ml e DS 10 mg/100ml.

Sapendo che i valori dell'HDL (high density lipoprotein) si distribuiscono normalmente, si vuole stimare la probabilità che un soggetto appartenente alla P abbia valori di HDL <45 mg/100ml.

$$Z = (45-57)/10 = -1.2$$

$$P(Z < -1.2) = 1 - 0.8849 = 0.1151 = 11.51\%$$



La percentuale di soggetti sani con valori di HDL <45mg/100ml nella popolazione è del 11.51%, mentre la Probabilità che un soggetto appartenente alla popolazione abbia valori di HDL >45 mg/100ml è del 88.49%.

Si vuole stimare la probabilità che un soggetto, estratto a caso da una popolazione apparentemente sana, abbia valori di HDL compresi tra 45 mg/100 ml e 75 mg/100ml. Mediante la standardizzazione si ha:

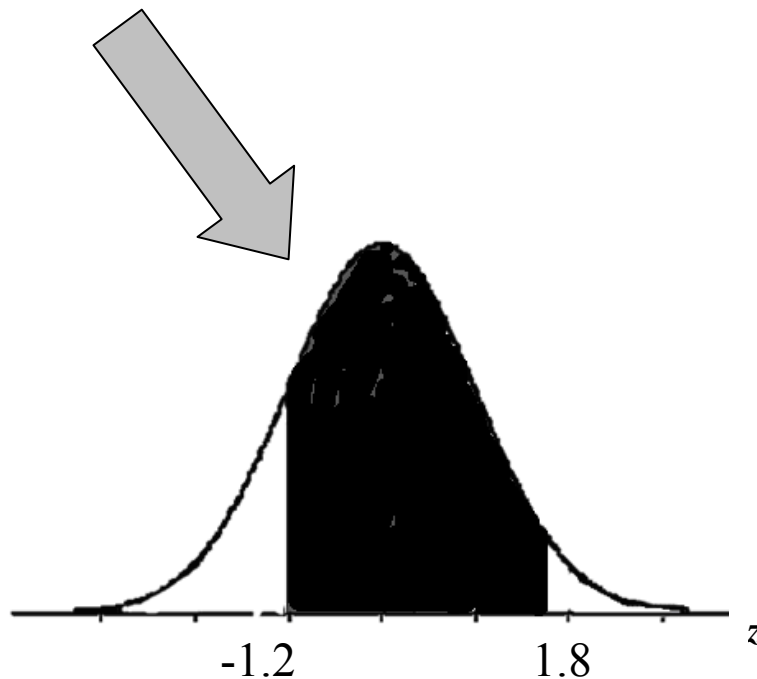
$$Z_1 = (45-57)/10 = -1.2$$

$$Z_2 = (75-57)/10 = 1.8$$

Dalla Tavola della Distribuzione Normale Standardizzata si ha:

$$P(-1.2 < Z < 1.8) = P(Z < 1.8) - [1 - P(Z < 1.2)] = 0.9641 - [1 - 0.8849] = 0.9641 -$$

$$0.1151 = 0.849 = 84.9\%$$



Si vuole stimare la probabilità che un soggetto, estratto a caso da una popolazione apparentemente sana, abbia valori di HDL compresi tra 40 mg/100 ml e 80 mg/100ml (Range di Normalità).

Mediante la standardizzazione si ha:

$$Z_1 = (40-57)/10 = -1.7$$

$$Z_2 = (80-57)/10 = 2.3$$

Dalla Tavola della Distribuzione Normale Standardizzata si ha:

$$P(-1.7 < Z < 2.3) = P(Z < 2.3) - [1 - P(Z < 1.7)] = 0.9893 - [1 - 0.9554] = 0.9893 - 0.0446 = 0.9447 = 94.5\%$$

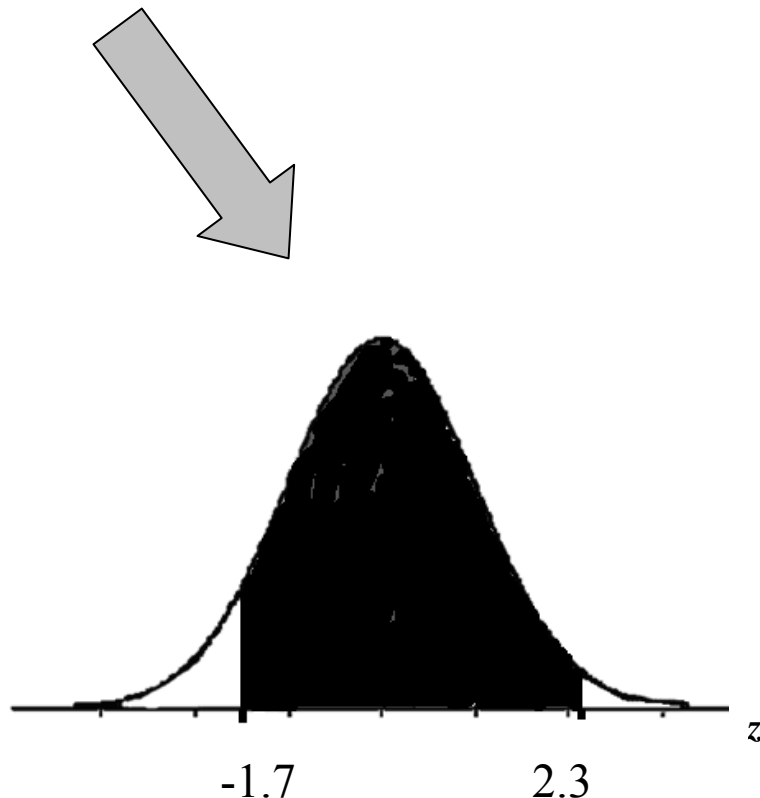
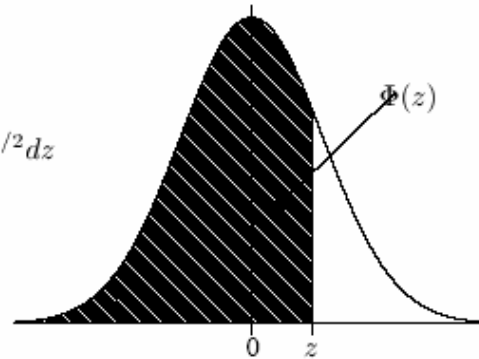


Tavola 1: Funzione di ripartizione della Variabile Casuale Normale Standardizzata

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$



<i>z</i>	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