

e la probabilità che si verifichi un evento compreso tra c e b ($a < c < b$) sarà data da:

$$p(c < X < b) = \int_c^b f(x) dx$$

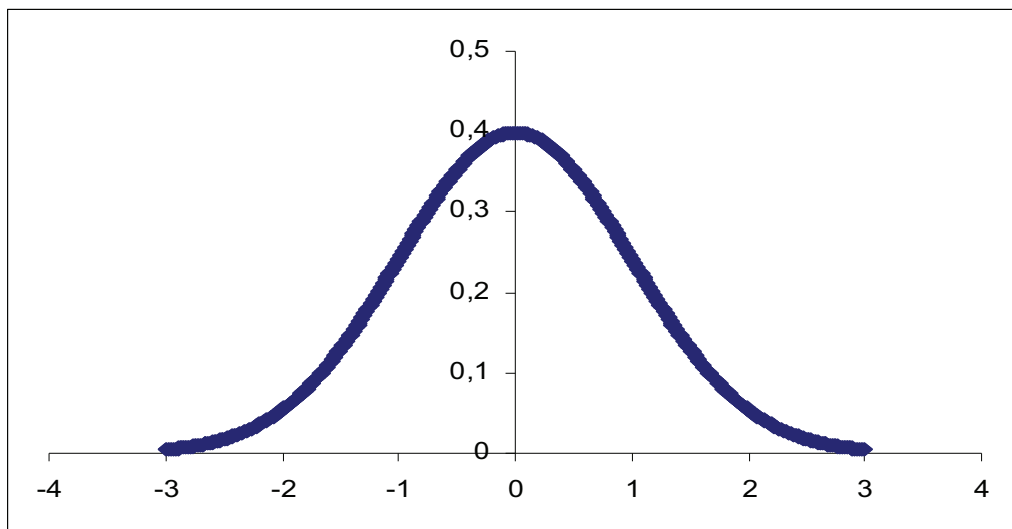
LA VC NORMALE O GAUSSIANA

Una vc si dice normale o gaussiana (da Gauss che la propose come modello descrittivo degli errori di misura) se la sua fd è la seguente:

$$f(X) = N(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2}$$

dove μ_x e σ_x^2 rispettivamente rappresentano il valor medio e la varianza di X ; $-\infty \leq X \leq \infty$ è una vc continua; $\pi = 3,1415$ ed $e = 2,7183$ (base dei logaritmi neperiani) sono note costanti matematiche.

La sua rappresentazione grafica è la seguente:



ed ovviamente la probabilità dell'evento certo sarà data da

$$p(X) = \int_a^b f(x) dx = 1$$

Oltre ai due valori caratteristici appena esaminati se ne possono definire altri; tra essi una certa importanza ha la media quadratica:

$$E(X^2) = \int_a^b x^2 f(x) dx$$

È facile dimostrare che:

$$V(X) = E(X^2) - [E(X)]^2$$

Per la dimostrazione basta svolgere il quadrato dell'altra formulazione di $V(X)$, semplificare ed ottenere la seconda formulazione che è di maggiore praticità a fini computazionali.

Lo studio analitico della funzione evidenza:

1) la curva è simmetrica rispetto all'ordinata del punto di massimo;

2) quest'ultimo si trova in corrispondenza del valore $x = \mu$; segue che la mediana (MED , valore che divide una distribuzione di frequenze in due parti esattamente uguali) e la moda (MOD , valore cui corrisponde il massimo valore di una distribuzione di frequenze) coincidono, nella normale, con la media aritmetica;

3) la curva è definita tra meno infinito e più infinito;

4) La curva presenta due punti di flesso (cambiamento di concavità) in corrispondenza con i valori $x = \mu \pm \sigma$

L'assetto grafico della curva è determinato dai parametri μ e σ ; il primo determina il posizionamento della curva sull'asse delle ascisse; per questo si definisce μ come un **parametro di posizione**.

Il secondo, essendo una misura di variabilità con riferimento alla media, mostra quanto siano più o meno dispersi i valori della distribuzione intorno al valore medio. Allora, bassi valori di σ indicano valori della distribuzione (probabilità) *poco dispersi* o anche, come si dice, *molto concentrati*, intorno a μ , al contrario alti valori di σ indicano valori della distribuzione *molto dispersi* rispetto alla media. Pertanto il parametro σ è detto **parametro di forma** della distribuzione.

Se una vc ha una distribuzione normale la probabilità che x assuma un certo valore in un certo intervallo, poniamo a - b , si ottiene da:

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

che in termini grafici altro non è se non la superficie delimitata a sinistra dall'ordinata nel punto a , a destra dall'ordinata del punto b , inferiormente dall'asse delle ascisse e superiormente dalla curva normale tra a e b . Ovviamente, la probabilità dell'evento certo, cioè

$$p(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x) dx =$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2} dx = 1$$

da cui si ha anche che:

$$p(\mu \leq X \leq \infty) =$$

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

Esempio, se una vc normale ha media pari a 3,6 e varianza pari a 81, la probabilità che x sia compreso tra -4,2 e 7,5 si ha risolvendo l'integrale

$$p(-4.2 \leq X \leq 7.5) =$$

$$= \int_{-4.2}^{7.5} f(x) dx = \int_{-4.2}^{7.5} \frac{1}{9\sqrt{2\pi}} e^{\frac{-(x-3.6)^2}{4(81)}} dx$$

Per fortuna esiste la possibilità di operare in modo estremamente più semplice, ma a tale fine occorre definire una particolare vc normale, detta vc normale standardizzata, la cui caratteristica è quella di avere media pari a zero e varianza unitaria, cioè:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} = N(Z; 0, 1)$$

Si può dimostrare che data una normale $N(X; \mu, \sigma^2)$ si può sempre passare ad una $N(Z; 0, 1)$ semplicemente trasformando le x in z con la relazione

$$Z = \frac{(x - \mu)}{\sigma}$$

Siccome per la normale standardizzata esistono tavole che contengono la determinazione degli integrali coinvolti con il calcolo di

$$p(z_1 \leq Z \leq z_2) = \int_{z_1}^{z_2} f(Z) dz$$

allora basta passare da X a Z , risolvere il nostro problema su Z ed averlo risolto per X senza dover calcolare alcun integrale.

Tutto questo sarà molto più chiaro con alcuni esempi numerici; prima vediamo più da vicino come sono costruite le tavole per la normale standardizzata.

In primo luogo: la tabulazione avviene solo per la parte positiva della distribuzione, dal momento che essendo la media della standardizzata uguale a zero basta avere

$\text{prob}(0 \leq Z \leq a)$ per avere $\text{prob}(-a \leq Z \leq 0)$

Poi, le tavole forniscono l'area sotto la normale standardizzata secondo il seguente schema:

L'immissione rappresenta l'area sottostante la distribuzione standardizzata dalla media aritmetica a Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.49865	.49869	.49874	.49878	.49882	.49886	.49889	.49893	.49897	.49900
3.1	.49903	.49906	.49910	.49913	.49916	.49918	.49921	.49924	.49926	.49929
3.2	.49931	.49934	.49936	.49938	.49940	.49942	.49944	.49946	.49948	.49950
3.3	.49952	.49953	.49955	.49957	.49958	.49960	.49961	.49962	.49964	.49965
3.4	.49966	.49968	.49969	.49970	.49971	.49972	.49973	.49974	.49975	.49976
3.5	.49977	.49978	.49978	.49979	.49980	.49981	.49981	.49982	.49983	.49983
3.6	.49984	.49985	.49985	.49986	.49986	.49987	.49987	.49988	.49988	.49989
3.7	.49989	.49990	.49990	.49990	.49991	.49991	.49992	.49992	.49992	.49992
3.8	.49993	.49993	.49993	.49994	.49994	.49994	.49994	.49995	.49995	.49995
3.9	.49995	.49995	.49996	.49996	.49996	.49996	.49996	.49997	.49997	.49997

Allora se $z = 0.22$ la superficie al di sotto della standardizzata (tra 0 e z) è pari a 0.0871, cioè è circa il 9% dell'intera distribuzione, se invece è pari a 0.30 la superficie è 0.1179, cioè circa il 12% della distribuzione, e così via.

Le tavole della normale standardizzata sono riportate in appendice ad ogni testo di statistica.

Vediamo allora un po' di esempi numerici e la soluzione di alcuni problemi.

Esempi:

Si calcoli usando la tavola della normale standardizzata la probabilità che:

$$-1.96 < Z < 1.96$$

Data la simmetria su 0 della distribuzione, basta moltiplicare per 2 il valore che si trova sulla tavola in corrispondenza di 0.96, cioè 0.4750.

Questo valore indica la probabilità tra 0 e 1.96, quindi $0.4750 \times 2 = 0.95$ dice che la probabilità richiesta, in termini percentuali, è il 95%.

Si calcoli ora la probabilità che in una normale con media pari a 10 e varianza pari a 4, X assuma un valore compreso tra 8 e 12.

Per usare la standardizzata si devono determinare su quest'ultima distribuzione quei valori che corrispondono a 8 e a 12;

essi sono:

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{8 - 10}{2} = -1$$

$$z_2 = \frac{X_2 - \mu}{\sigma} = \frac{12 - 10}{2} = +1$$

Allora dalle tavole della normale standardizzata:

$$\text{prob}(0 \leq Z \leq z_2) = 0.3413$$

$$\text{prob}(z_1 \leq Z \leq z_2) = 2 \cdot 0.3413 = 0.6826$$

e pertanto la probabilità richiesta per X è approssimativamente del 68%.

Sia X una vc normale con media 16000 e scarto quadratico medio pari a 2000.

Calcolare la probabilità che X sia compreso tra 15000 e 18000.

$$z_1 = \frac{15000 - 16000}{2000} = -0.5$$

$$z_2 = \frac{18000 - 16000}{2000} = 1$$

Allora:

$$\text{prob}(-0.5 \leq Z \leq 1) = \text{prob}(0 \leq Z \leq 0.5) = 0.1915 + \\ + \text{prob}(0 \leq Z \leq 1) = 0.3413$$

$$\text{prob}(-0.5 \leq Z \leq 1) = 0.1915 + 0.3413 = 0.5328$$

allora la probabilità richiesta per X è di circa il 53%.