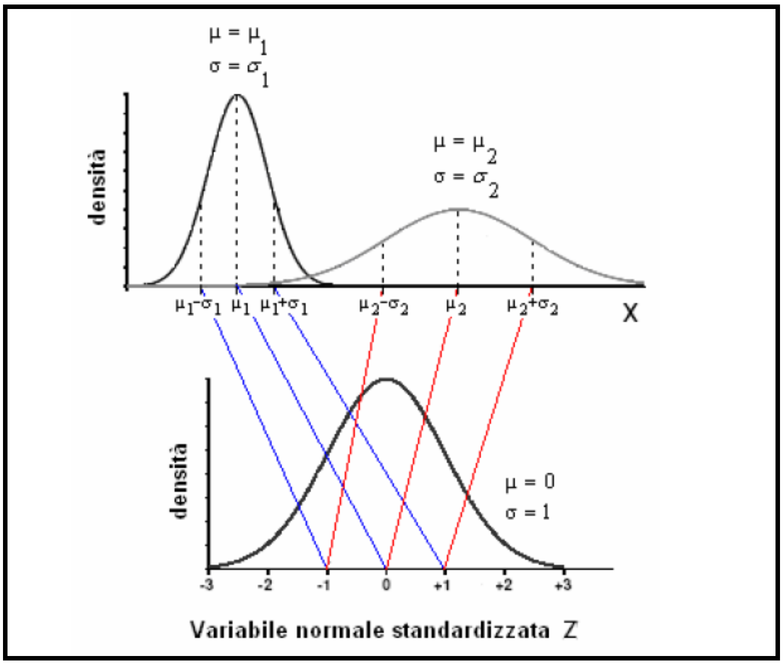
FACOLTA' DI INGEGNERIA

CORSO DI LAUREA IN INGEGNERIA CIVILE

CORSO DI IDROLOGIA

PROF. PASQUALE VERSACE



SCHEDA DIDATTICA N°7

LA DISTRIBUZIONE NORMALE

A.A. 2012-13

La distribuzione NORMALE

Uno dei più importanti esempi di distribuzione di probabilità continua è dato dalla *distribuzione Normale (curva normale o distribuzione Gaussiana)*; è una delle più usate in statistica sia perché molti fenomeni si distribuiscono "normalmente", sia perché altre funzioni di probabilità (ad es. la binomiale) possono essere approssimate da essa.
Per i matematici è di grande utilità il ragionare su distribuzioni basate su un numero di casi infinitamente "grandi". Infatti, il *Teorema Limite Centrale* stabilisce che la distribuzione della somma o della media di un campione casuale di *n* valori estratti da una popolazione ha una distribuzione normale per un numero elevato *n*. La distribuzione normale è detta anche curva degli errori accidentali o di Gauss.

La *distribuzione Normale* è definita dall'equazione:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right] \qquad (1)$$

con *variabile casuale*

$x = \in \left]-\infty ; +\infty\right[$

e *parametri*

μ = media $\in \left]-\infty ; +\infty\right[$

σ^2 = varianza $\in \left]0 ; +\infty\right[$;

Per indicare una v.c. X distribuita in maniera Normale con parametri μ e σ^2 si usa in genere la notazione $X \sim N(\mu, \sigma^2)$.

Sulla $f_X(x)$ Normale si possono fare alcune considerazioni:

1) $f_X(x) \geq 0 \; \forall x$;
2) $\int_{-\infty}^{+\infty} f_X(x)dx = 1$;
3) $f_X(x)$ è tanto più grande quanto più è piccolo l'esponente, ovvero quanto più x è vicino a μ, e raggiunge il suo massimo per $x = \mu$ (moda, media e mediana coincidono);
4) $f_X(x)$ possiede due flessi, cioè due punti in cui cambia concavità , in $\mu-\sigma$ e $\mu+\sigma$;
5) $f_X(x)$ è simmetrica intorno ad $x = \mu$, cioè per ogni coppia di ascisse $x_1=\mu-c$ e $x_2=\mu+c$ aventi la stessa distanza c dall'asse di simmetria, $f_X(x_1) = f_X(x_2)$;

Poiché π ed e sono costanti per conoscere l'esatte forma della distribuzione è necessario conoscere la media μ e lo scarto quadratico medio σ. Si avranno, pertanto, diverse curve normali quanti sono le possibili combinazioni di medie e scarti quadratici medi. Per esempio, due curve con medie uguali e scarto diverso differiranno per quanto riguarda la forma che sarà più o meno appuntita: se minore è lo scarto quadratico medio, la curva è più appuntita. Come illustrato delle figure 1 e 2 e nell'esempio successivo,

- μ determina la posizione della curva sull'asse delle ascisse. È il ***parametro di posizione***, al variare del suo valore, la curva non cambia nella forma ma subisce una traslazione rispetto all'asse orizzontale;
- σ^2 determina la maggiore o minore concentrazione della curva intorno a μ. È il ***parametro di scala***: al suo variare cambia la forma della curva di distribuzione. In particolare, per bassi valori di σ, l'area sotto la curva è concentrata intorno alla media, mentre per alti valori la curva è "schiacciata" rispetto all'asse orizzontale.

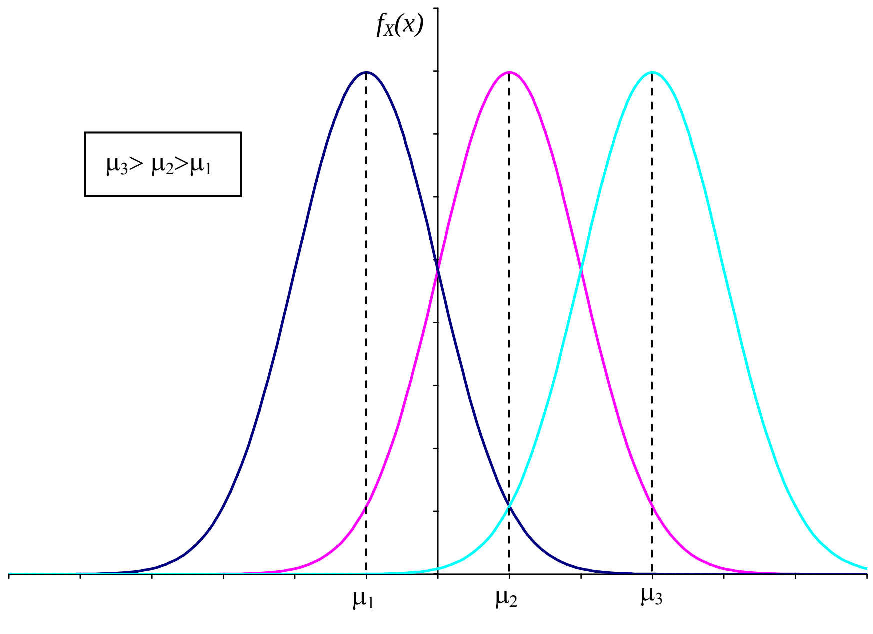


Figura 1 - Distribuzione di una variabile casuale Normale con varianza fissa e differenti medie

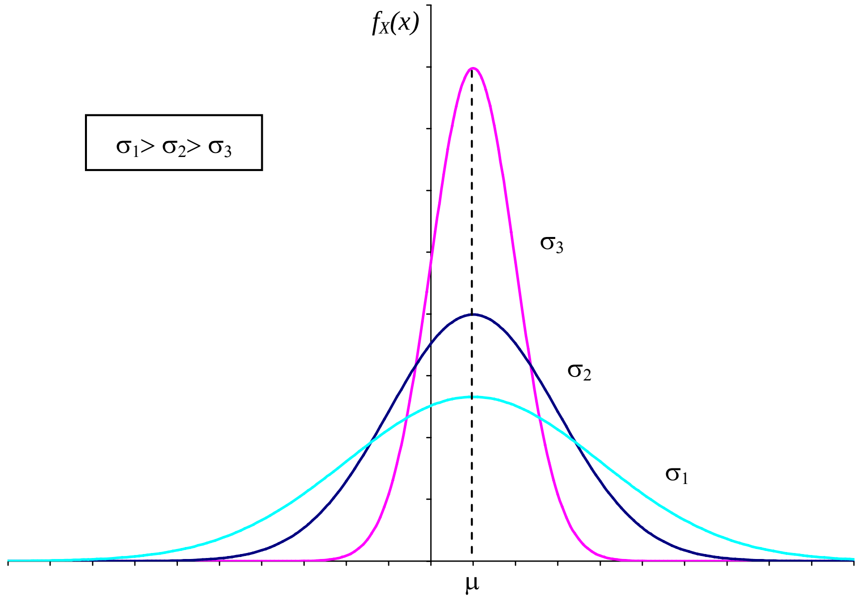


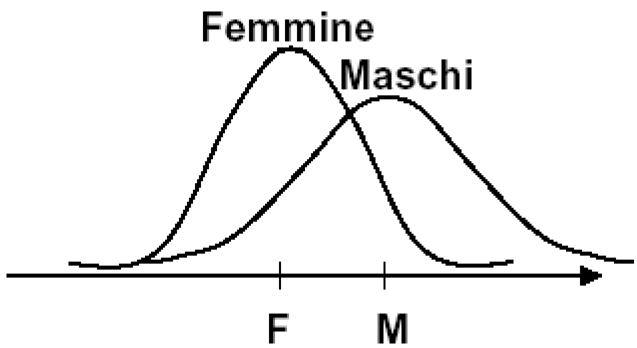
Figura 2- Distribuzione di una variabile casuale Normale con media fissa e differenti varianze

---

ESEMPIO: Altezza di uomini e donne

Nella figura che segue la moda della curva che descrive la distribuzione dell'altezza delle donne è più alta di quella degli uomini: significa che le donne sono più alte degli uomini?

NO!

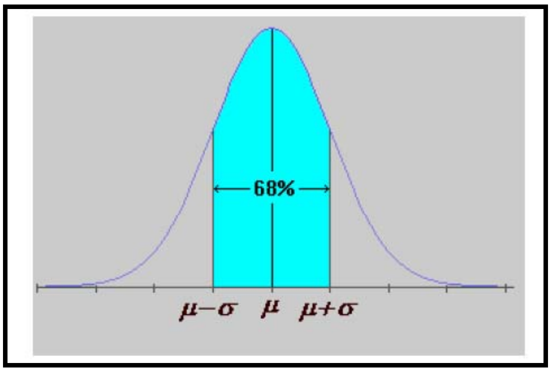


Gli uomini sono in media più alti delle donne perché la loro curva è più a destra, cioè verso valori di *X* (altezza) maggiori. Poiché le altezze dei maschi hanno una variabilità maggiore, la curva dei maschi è più bassa e larga (entrambe le aree sottese valgono 1).

Se un fenomeno si distribuisce secondo una distribuzione Normale si ha che:

- circa il 68% di tutti i valori cade nell'intervallo di + e – 1 deviazione standard dalla media

$$P[\mu-\sigma \leq x \leq \mu+\sigma] = 0.6826$$



analogamente:

- il 95% dei valori cade nell'intervallo di + e – 1.96 deviazioni standard dalla media

$$P[\mu-1.96\sigma \leq x \leq \mu+1.96\sigma] = 0.95$$

- e per 3 sigma …

$$P[\mu-3\sigma \leq x \leq \mu+3\sigma] = 0.9974.$$

Essendo infinite le curve normali al variare della media e dello scarto quadratico medio è sorta la necessità di costruire una curva a cui fare riferimento e di cui sono state calcolate le aree.

Un caso particolare della v.c. Normale, utile nelle applicazioni, è la v.c. *Normale Standardizzata*, convenzionalmente indicata con Z (oppure U), così definita:

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

e per la quale si ha **E[Z] = 0** e **Var[Z] = 1**.

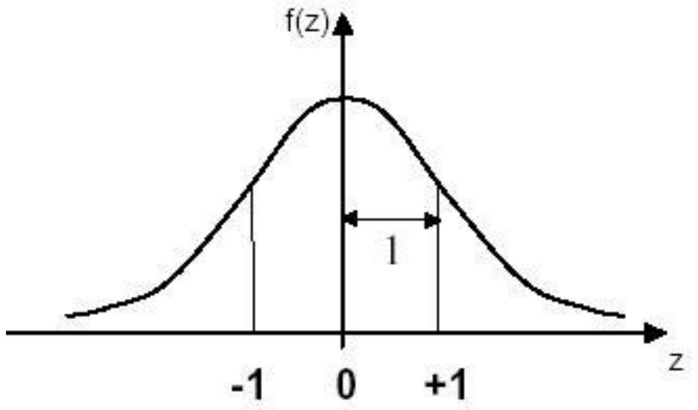
La distribuzione normale standardizzata, infatti, è una particolare normale di valor medio nullo e varianza unitaria. E' possibile dimostrare che:

$$f_Z(z)dz = f_X(x)dx$$

e verificare quindi:

$$E[Z] = \int_{-\infty}^{+\infty} z \cdot f_Z(z)dz = \int_{-\infty}^{+\infty} \frac{x-\mu}{\sigma} \cdot f_X(x)dx = \frac{1}{\sigma}\left[\int_{-\infty}^{+\infty} x \cdot f_X(x)dx - \mu \int_{-\infty}^{+\infty} f_X(x)dx\right] = \frac{1}{\sigma}[\mu - \mu \cdot 1] = 0$$

$$Var[Z] = \int_{-\infty}^{+\infty} z^2 \cdot f_Z(z)dz = \int_{-\infty}^{+\infty} \left(\frac{x-\mu}{\sigma}\right)^2 f_X(x)dx = \frac{1}{\sigma^2}\int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$



Data una v.c. $X \sim N\ (\mu, \sigma^2)$ è possibile passare ad una v.c. Normale Standardizzata $Z \sim N\ (0,1)$ (e viceversa) attraverso la relazione (2).

Per la variabile standardizzata la *p.d.f.* e la *c.d.f.* diventano rispettivamente:

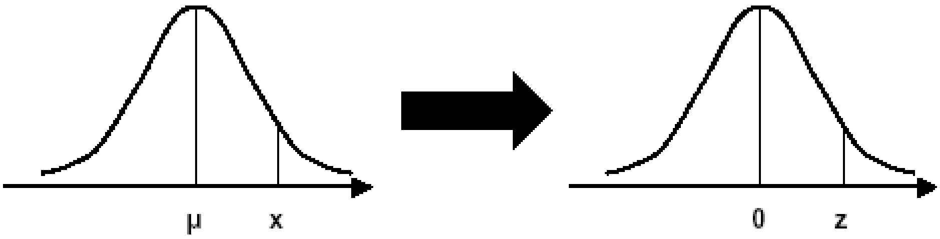
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right]$$

$$F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{v^2}{2}\right) dv$$

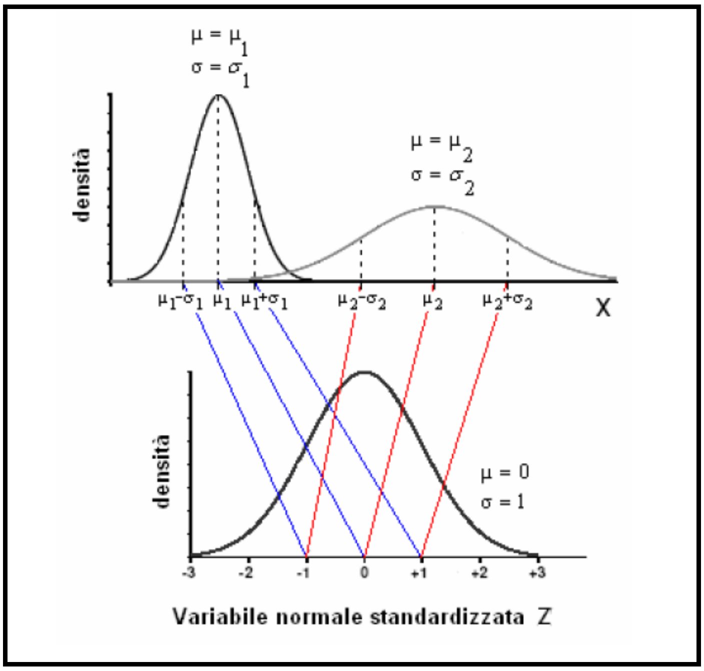
Per la variabile casuale *Z*, non essendo la distribuzione dipendente da alcun parametro, sono stati tabulati i valori della *p.d.f.* e della *c.d.f.* (vedi Tabella 1 allegata).

Inoltre è possibile dimostrare che:

$$F_Z(z) = F_X(x) \qquad (3)$$



Graficamente:

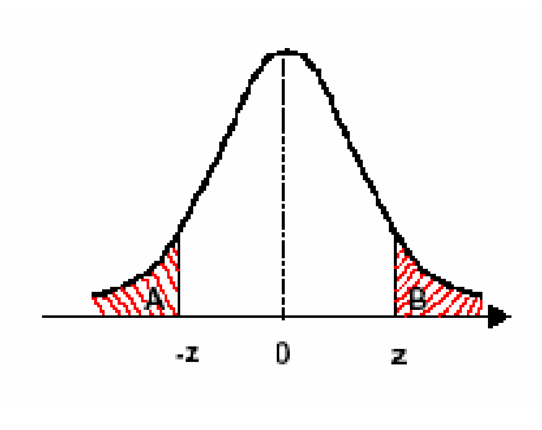


Graficamente la trasformazione della variabile originaria in variabile standardizzata si ottiene passando dal sistema cartesiano ortogonale al sistema con asse Z con la curva al valore massimo nell'origine e flessi a -1, +1. Fatta la trasformazione, si possono usare, quindi, le apposite tavole per il calcolo delle le porzioni di aree sottese.
In tabella 1 è riportata la tavola delle aree sotto la curva normale standardizzata comprese tra l'ascissa 0 e qualsiasi valore positivo fino a 3,99. Servendosi di questa tavola è possibile trovare la aree (e quindi probabilità) comprese tra due ascisse qualsiasi, ricordando la simmetria della curva intorno alla media che coincide con il valore z=0.

Come si procede per calcolare le probabilità nel caso di una v.c. Normale con l'ausilio della v.c. standardizzata?

- Si definiscono la v.c. X, i valori di μ e σ e l'evento di interesse
- Si calcola il valore standardizzato z
- Si disegna la curva normale individuando sul grafico l'area di interesse
- Si usano tavole, simmetria e probabilità dell'evento complementare (1- …) per calcolare il valore della probabilità (area) che si desidera.

Se ho un valore di z < 0 ? Si osserva che: $F_Z(-z)= 1- F_Z(z)$



AREA A = AREA B

AREA B =1 – $F_Z(z)$

N.B.-*Consultando la tavola è importante ricordare che i valori z sono sull'asse delle ascisse e non confonderli con le aree. I valori di z possono essere negativi, le aree non lo sono mai!*

ESEMPI: ***Variabile standardizzata***

1) **Si calcoli, utilizzando la tabella 1, la probabilità $P[Z \leq 3]$ ovvero $F_Z(3)$.**

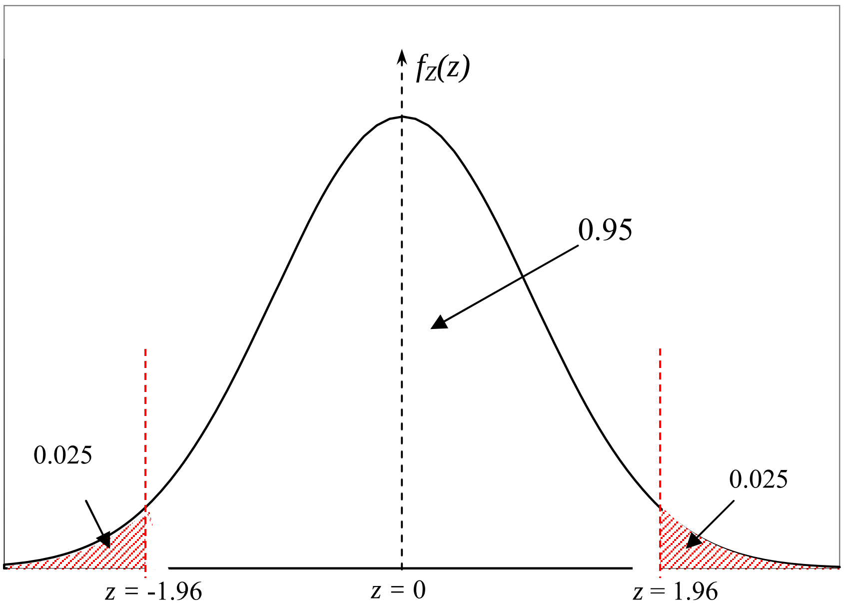
   Nella tabella 1 procedere verso il basso nella colonna segnata *z* fino a raggiungere il valore 3. Quindi procedere verso destra fino alla colonna segnata 0. Il valore 0.4987 che si individua è l'area sottesa a partire dall'ascissa 0; per ottenere l'area richiesta, tenendo conto della simmetria della curva, bisogna aggiungere a tale valore 0.5.

   $F_Z(3)=0.4987 + 0.5 = 0.9987$ rappresenta la probabilità che *z* sia minore od uguale a 3.

2) **Si calcoli la probabilità $P[-1.96 \leq Z \leq 1.96]$.**

   E' necessario calcolare $[F_Z(1.96) - F_Z(-1.96)]$ ovvero l'area compresa tra $z = -1.96$ e $z = 1.96$.

   Per trovare l'area compresa tra 0 e 1.96 procedere verso il basso nella colonna segnata *z* fino a raggiungere il valore 1.9. Quindi procedere a destra fino alla colonna segnata 6. Il valore 0.475 per simmetria rappresenta anche l'area compresa tra –1.96 e 0. Pertanto la probabilità richiesta risulta $P[-1.96 \leq Z \leq 1.96] = 0.475 + 0.475 = 0.95$.



3) **Trovare la probabilità $P[Z \leq -0.6]$ ovvero $F_Z(-0.6)$.**

   L'area richiesta si trova a sinistra del valore $z = 0$; sulla tabella è, invece, possibile trovare il valore dell'area compresa tra 0 e 0.6. che risulta pari a 0.2258.

   Considerando la simmetria della curva si ha

$$F_Z(-z)= 1- F_Z(z)$$

   da cui

$$F_Z(-0.6)= 1- F_Z(0.6) = 1 - (0.5 + 0.2258) = 0.2742$$

4) **Trovare la probabilità $P[Z \geq -1.28]$.**

L'area richiesta = (area compresa tra z = -1.28 e z = 0) + (area a destra di z = 0). Sulla tabella troviamo il valore del primo termine individuando il valore per 1.28, mentre il secondo termine è pari a 0.5.

Quindi si ha:

$$P[Z \geq -1.28] = 0.3997 + 0.5 = 0.8997$$

5) **Trovare la probabilità $P[Z \geq 2.05] = 1-F_Z(2.05)$.**

L'area richiesta è l'area a destra di z = 2.05. Sulla tabella troviamo il valore dell'area compresa tra 0 e 2.05; per ottenere $F_Z(2.05)$ aggiungiamo l'area a sinistra di z=0 cioè 0.5.

$$F_Z(2.05) = 0.4798 + 0.5 = 0.9798$$

Quindi si ha:

$$P[Z \geq 2.05] = 1-F_Z(2.05) = 1 - 0.9798 = 0.0202.$$

Analogamente partendo dal valore della probabilità è possibile, attraverso la tabella, individuarne il corrispondente frattile.

6) **Trovare il valore della variabile standardizzata *Z* cui corrisponde una probabilità di non superamento pari a 0.95.**

E' necessario, quindi, calcolare il frattile $z_{0.95}$ a cui corrisponde $F_Z(z_{0.95})$=0.95.

Poiché nella tabella sono riportate le aree a destra di z = 0, bisogna individuare il valore 0.45 = 0.95- 0.5. Tale valore risulta compreso tra quelli riportati tra z = 1.64 e z = 1.65.

Effettuando un'interpolazione lineare tra tali valori si ha:

$$\frac{1.65-1.64}{0.4505-0.4495} = \frac{1.65-z}{0.4505-0.45}$$

$$z = 1.65-(0.4505-0.45)\cdot\frac{1.65-1.64}{0.4505-0.4495} = 1.645$$

ESEMPIO ***distribuzione Normale***

1) Si assuma che la v.c. altezza delle donne di età maggiore ai 18 anni, in Calabria, è distribuita secondo la legge Normale.

La media e lo scarto quadratico medio dell'altezza delle donne sono rispettivamente 158 cm e 15 cm.

a) **Si calcoli la probabilità che l'altezza di una donna sia compresa tra x_1= 140 cm e x_2 = 170 cm.**

Per rispondere al quesito è necessario utilizzare la tabella e pertanto considerare i valori standardizzati.

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{140 - 158}{15} = -1.2$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{170 - 158}{15} = 0.8$$

Dalla tabella si ha:

$F_Z(-1.2) = 1- F_Z(1.2) = 1- (0.5 + 0.3849) = 1- 0.8849 = 0.1151$

$F_Z(0.8) = 0.5 + 0.2881 = 0.7881$

Per la relazione (3) risulta:

$F_Z(-1.2) = F_X(140)$

$F_Z(0.8) = F_X(170)$

La probabilità richiesta è quindi:

$P[\ 170 \leq X \leq 140] = F_X(170) - F_X(140) = F_Z(0.8) - F_Z(-1.2) = 0.7881- 0.1151= 0.673.$

Su un campione di 1000 donne ad esempio il 67.3 %, ovvero 673, hanno un altezza compresa tra 140 e 170 cm.

b) si calcoli la probabilità che l'altezza di una donna sia maggiore di 185 cm.

Considerando la variabile standardizzata

$$z = \frac{x - \mu}{\sigma} = \frac{185 - 158}{15} = 1.8$$

dalla tabella si ha :

$P[Z \leq 1.8] = F_Z(1.8) = 0.5 + 0.4641= 0.9641$

Quindi:

$P[Z \geq 1.8] =1- F_Z(1.8) = 0.0359$

$P[X \geq 185] = 1 - F_X(185) = 0.0359$

c) l'altezza cui corrisponde una probabilità di non superamento del 30%.

$F_X(x_{0.3}) = F_Z(z_{0.3}) = 0.3.$

La variabile standardizzata $z_{0.3}$ in questo caso sarà negativa e, quindi, è necessario considerare la relazione

$$F_Z(-z)= 1- F_Z(z)$$

Si ha:

$F_Z(-z)= 1- F_Z(z) = 1- 0.3 = 0.7$

Poichè nella tabella sono riportati i valori per $z > 0$ è necessario cercare l'area pari a 0.7- 0.5 = 0.2.

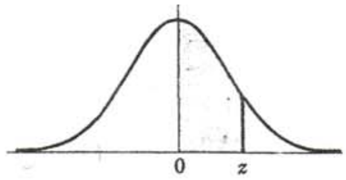
Tale valore si ha in corrispondenza di 0.525 (è stata effetuata un'interpolazione lineare) per cui $z_{0.3}$ = - 0.525.

Per tornare alla v.c. X (altezza in cm) si utilizza la relazione (2).

$$x_{0.3} = \sigma \cdot z_{0.3} + \mu = 15 \cdot (-0.525) + 158 = 150.125 \text{ cm}$$

Tabella 1- Aree sotto la curva Normale Standardizzata (SCHAUM)

AREE
sotto la
CURVA NORMALE
STANDARDIZZATA
da 0 a z



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000