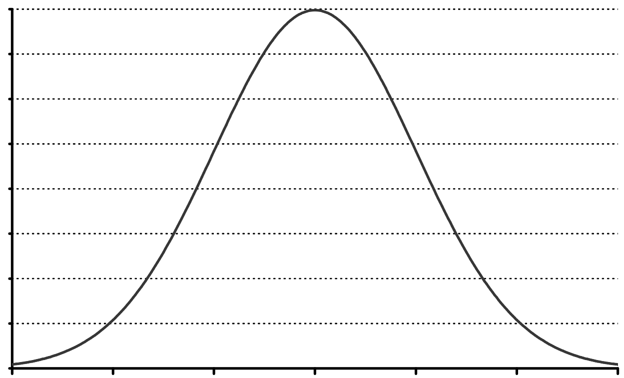
LEZIONI DI STATISTICA MEDICA

Distribuzione Normale

Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

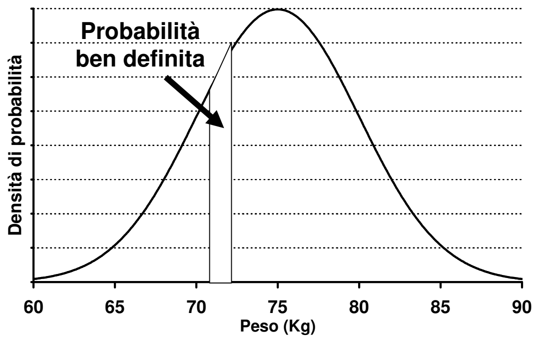
DISTRIBUZIONE NORMALE

La distribuzione normale (distribuzione gaussiana, distribuzione degli errori accidentali) occupa un ruolo centrale nell'ambito della statistica medica.



VARIABILE CASUALE CONTINUA:

Qual è la probabilità che il peso di un individuo di questa popolazione sia esattamente 73Kg 133g? Praticamente zero.



- La probabilità di un singolo valore x_i per una variabile continua è pari a 0
  $P(X=x_i)=0$

- E' positiva e definita la probabilità per un intorno, per quanto piccolo, di x_i
  $P(x_i-\delta \le X \le x_i+\delta) \ge 0$

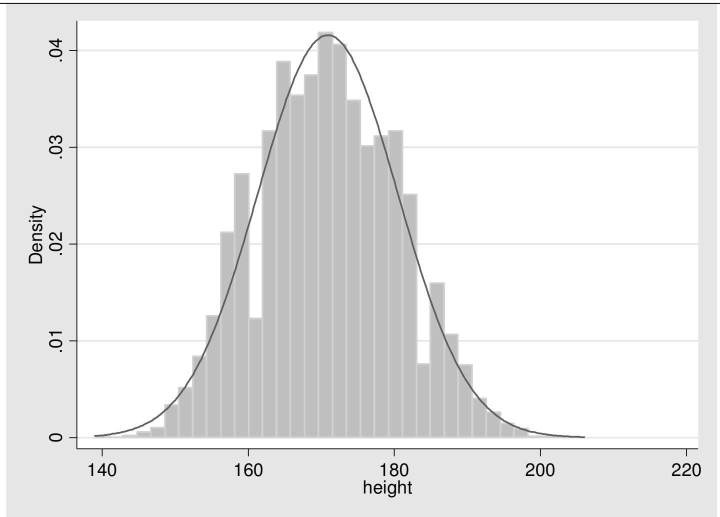
- La maggior parte delle variabili biologiche (peso, statura, …) dipendono dalla somma di svariati fattori genetici e ambientali. ⇨ La maggior parte delle variabili biologiche segue la distribuzione normale.

- La maggior parte delle altre distribuzioni teoriche di probabilità (binomiale, di poisson, t di Student) tendono alla distribuzione normale

- Trasformazioni matematiche (log, √,…) possono "normalizzare" una variabile che naturalmente non lo sarebbe

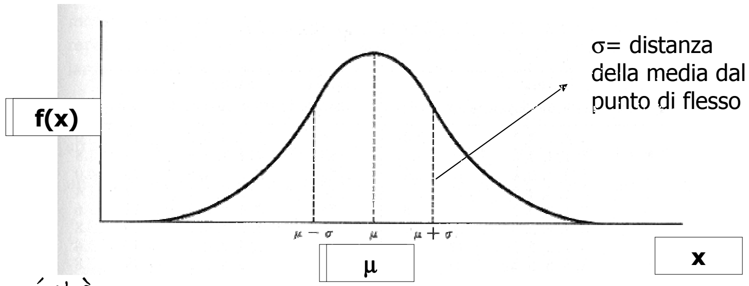
- La variabile $Y=X_1+X_2+....X_n$ segue una distribuzione normale per n sufficientemente grande e X indipendenti **[teorema del limite centrale]**

Distribuzione delle altezze di circa 15000 soggetti di età 20-44 (dati ECRHS)



La distribuzione empirica (diagramma a barre) può essere approssimata con una curva teorica (la distribuzione normale).

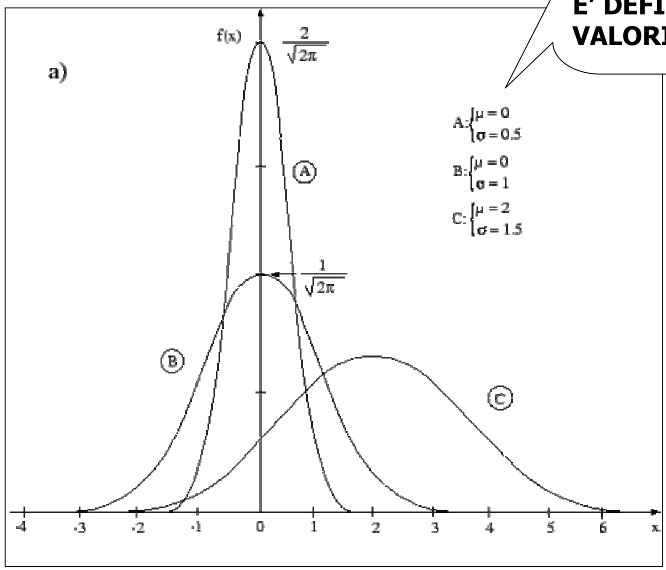
Definizione: Una variabile casuale X ha una distribuzione normale, $\boldsymbol{X \sim N(\mu, \sigma)}$



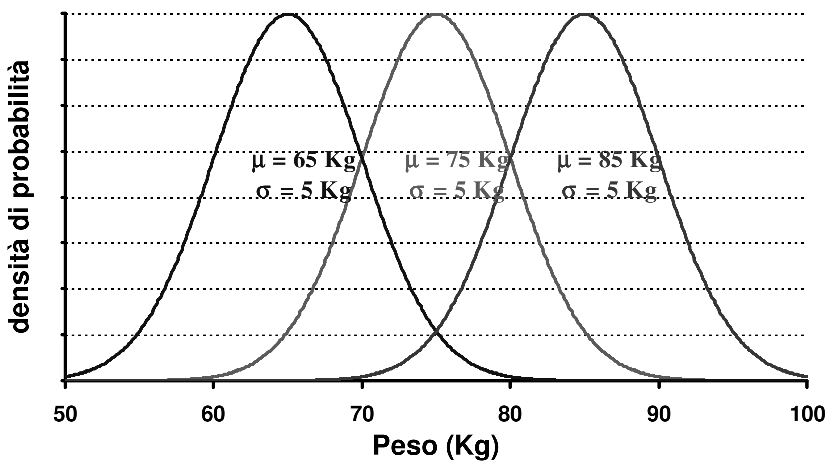
Proprietà:

- Distribuzione simmetrica ⟹ f(μ-x)=f(μ+ x)
- media=moda=mediana
- $P(\mu-\sigma \leq x \leq \mu+\sigma)=0.68$
- $P(\mu-1.96\sigma \leq x \leq \mu+1.96\sigma)=0.95$

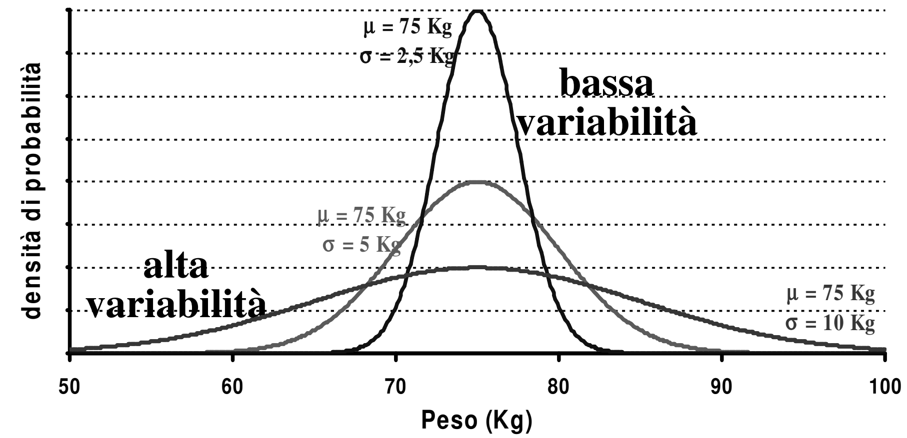
LA FORMA SPECIFICA DELLA DISTRIBUZIONE E' DEFINITA DAI VALORI DI μ e σ



Queste 3 distribuzioni differiscono per la media (misura di posizione)



Queste 3 distribuzioni differiscono per la deviazione standard (misura di dispersione)



Esiste un numero infinito di distribuzioni normali diverse fra loro.

E' possibile ricondurre tutte queste diverse distribuzioni ad un'unica **distribuzione standard**?

Sì, attraverso la ***trasformazione normale***:

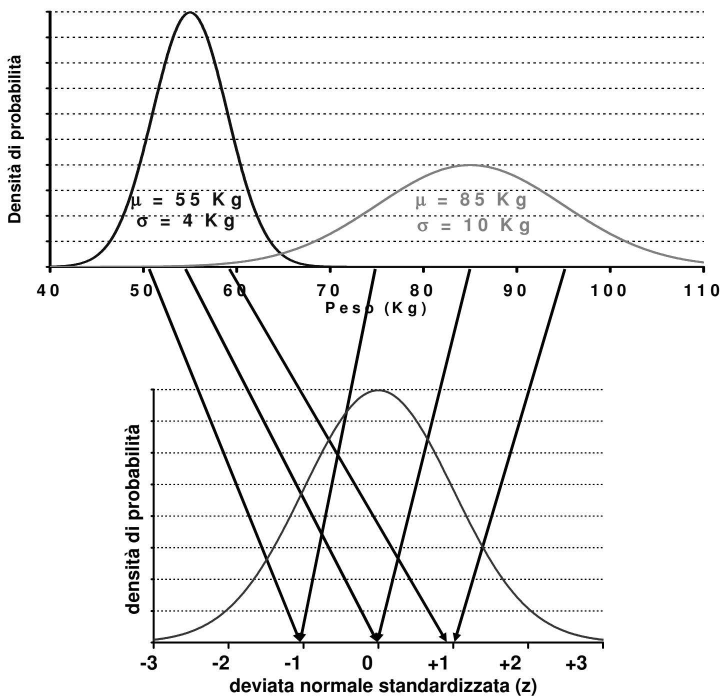
Definizione:: Sia *X* una variabile casuale distribuita normalmente,

X~N(μ, σ), allora la nuova variabile **Z:**

$$\boxed{z = (x-\mu)/\sigma}$$

avrà una distribuzione normale con *μ=0 e σ=1 .* **[Z~N(0, 1)]**

Z è detta DEVIATA NORMALE STANDARDIZZATA



Esistono delle tavole (*tavole della z*) che danno la probabilità che Z sia maggiore di un valore qualsiasi. $\boxed{P(Z \geq z)}$ *(Test a una coda)*

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	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,06430	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Qual è la probabilità che Z sia maggiore o uguale a 1,87?

$P(Z \geq z)$

0,0307 = 3,07%

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	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,06430	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Qual è la probabilità che Z sia maggiore o uguale a 0,75?

$P(Z \geq z)$

0,2266 = 22,66%

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	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,06430	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Esercizio 1: Sia X="peso in kg" e ***X~N(75, 8):***

A. Utilizzando le tavole di Z, calcolare la probabilità che:

1. il peso sia ≤ 63 kg
2. il peso sia compreso tra 69 e 92

B. Qual è il valore del peso tale per cui l'80% ha valori inferiori?

Soluzione:

A. 1. Calcoliamo il valore di Z relativo 63: **$z = (63-75)/8 = -1.5$**

$P(Z \leq -1.5) = P(Z \geq 1.5) = 0.0668 = 6.7\%$

Perché la distribuzione è simmetrica

2. Calcoliamo il valore di Z relativo 69: **$z = (69-75)/8 = -0.75$**

e quello relativo a 92: **$z = (92-75)/8 = 2.125$**

$P(-0.75 \leq Z \leq 2.125) = 1-[P(Z \leq -0.75)+P(Z>2.125)] =$

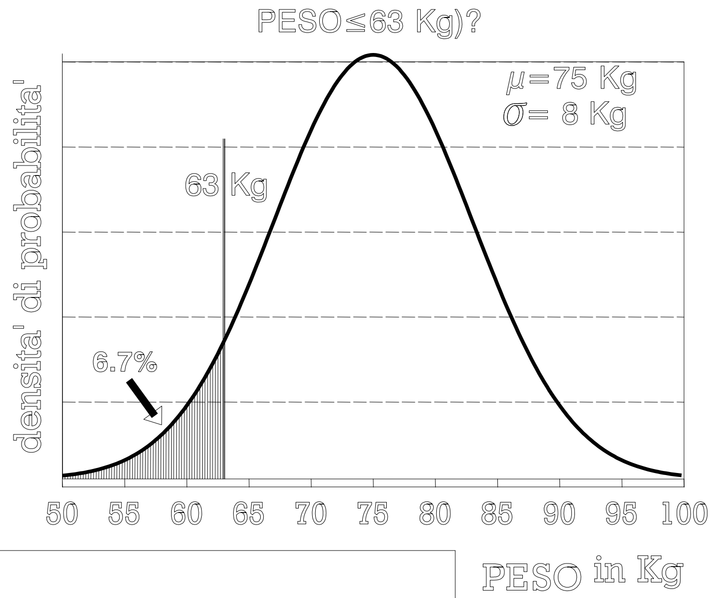
$= 1-[P(Z>0.75)+P(Z>2.125)] =$

$= 1-(0.2266+0.0170) = 0.76118 = 76.1\%$

B. Calcoliamo il valore di X tale per cui **$P(Z<z)=0.80$** $\longrightarrow$ **$P(Z \geq z)=0.20$**

perciò: **$z = 0.84$**

Quindi il valore di X sarà pari a : **X= 0.84*8+75=81.72 kg**



Esercizio 1:

Si assuma che tra i non diabetici, il livello ematico di glucosio a digiuno sia distribuito in maniera approssimativamente normale con **media=105 mg/100ml** ed una **deviazione standard= 9 mg/100ml.**
Calcolare:

1. Quale % di non diabetici ha livelli compresi tra 90 e 125 mg/ml
2. Qual è il valore di glicemia tale per cui il 90% dei soggetti ha valori superiori
3. Quali livelli di glicemia comprendono il 95% dei non diabetici

Soluzione:

X= livello ematico di glucosio

1. Calcoliamo il valore di Z relativo 90: **z = (90-105)/9 =-1.67**
   e quello relativo a 125: **z = (125-105)/9 =2.22**

P(-1.67≤ Z≤ 2.22)=1-[P(Z≤ -1.67)+P(Z > 2.22)]=1-[P(Z >1.67)+P(Z>2.22)]=
=1-(0.049+0.013)=0.938=93.8%

2. Cerco il valore di Z tale per cui:

P(Z<=c)=0.1 ⟹ **c=-1.28** ⟹ **-1.28=(x-105)/9** ⟹ **X=93.48 mg/ml**
P(Z>c)=0.9

3. Livelli di glicemia che comprendono il 95% dei non diabetici

$\mathbf{Pr(x_1 \leq X \leq x_2)=0.95}$

$\mathbf{Pr(z_1 \leq Z \leq z_2)=0.95}$ ⟹ $\mathbf{z_1 = -1.96 = (x_1 - \mu)/\sigma}$

$\mathbf{Pr(-1.96 \leq Z \leq +1.96)=0.95}$ $\qquad$ $\mathbf{z_2 = +1.96 = (x_2 - \mu)/\sigma}$

$\mathbf{Pr(\mu -1.96\, \sigma \leq x \leq \mu -1.96\, \sigma)=0.95}$

Quindi l’intervallo ricercato sarà:

$\mu \pm \mathbf{1.96\, \sigma = 105 \pm 1.96 \cdot 9}$ ⟹ **(87.36-122.64)**