LEZIONI DI STATISTICA MEDICA

Distribuzioni di probabilità:

Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

VARIABILE CASUALE

Si definisce variabile casuale (o aleatoria) una variabile che assume differenti valori con definite probabilità

ESEMPI

- il valore ottenuto nel lancio di due dadi

X	2	3	4	5	6	7	8	9	10	11	12
P(X)	**1/36**	**2/36**	**3/36**	**4/36**	**5/36**	**6/37**	**5/36**	**4/36**	**3/36**	**2/36**	**1/36**

- Numero di aborti su n gravidanze
- Numero di incidenti stradali in Veneto
- Numero di bronchiti in un anno
- Livello di colesterolo in un soggetto diabetico
- Pressione sistolica di un giovane

VARIABILE CASUALE DISCRETA

1. Assume un numero definito di valori

2. *Funzione* o *Distribuzione di probabilità*: $f(X) = P(X = x_i)$

dove f(X) è tale che:

1. $f(x) \geq 0$* per ogni *x

2. $\Sigma_x f(x) = 1$

3. Funzione di distribuzione F (X) $F(X) = P(X \leq x_i) = \sum_{t \leq x} f(t)$

Esempi:

- Numero di teste su 100 lanci
- Numero di figli maschi su 4 nascite
- Numero di decessi in un mese

DISTRIBUZIONI DI PROBABILITA' TEORICHE

1. VARIABILI DISCRETE

ESPERIMENTO BERNOULLIANO

Solo 2 possibili esiti in ogni prova:

Successo (S) → $Pr(S)=\pi$

Insuccesso (I) → $Pr(I)=1\text{-}Pr(S)=\mathbf{1\text{-}}\pi$

Prove indipendenti

Esempi:

1. Stato di vita
2. Lancio di una moneta

VARIABILE CASUALE BERNOULLIANA

Variabile che assume solo due valori con definite probabilità

X	f(X)
0	1-π
1	π

Esempio: Consideriamo la variabile casuale bernoulliana X= "Avere gruppo sanguigno B"; sappiamo che la probabilità di appartenere a tale gruppo sanguigno vale 0.08. La sua distribuzione di probabilità sarà:

X= 1 avere gruppo sanguigno B   Pr(X=1)=0.08= π

X= 0 <u>non</u> avere gruppo sanguigno B   Pr(X=0)=1-0.08=0.92=1- π

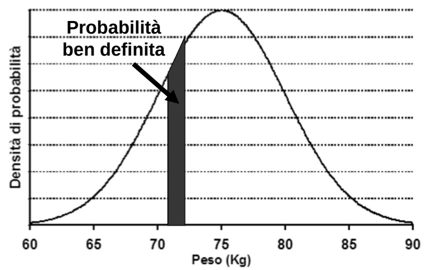
Distribuzione di probabilità di *X*:

X	f(X)
1	0.08
0	0.92

2.VARIABILI CONTINUE

VARIABILE CASUALE CONTINUA:

Variabile che può assumere infiniti valori all'interno di un intervallo continuo.



- 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 di un intorno, per quanto piccolo, di x_i
  $P(x_i-\delta\leq X\leq x_i+\delta)\geq 0$

VARIABILE CASUALE CONTINUA

1. **Può assumere <u>infiniti</u> valori all'interno di un <u>intervallo continuo</u>**
2. **La funzione di densità di probabilità (f.d.p) è una funzione continua**

Proprietà:

1. $f(x) \geq 0$ per ogni x
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$
3. $P(a \leq x \leq b) = \int_{a}^{b} f(x)dx$

3. **Funzione di distribuzione F (X)**

$$F(X) = \int_{-\infty}^{x} f(t)dt$$

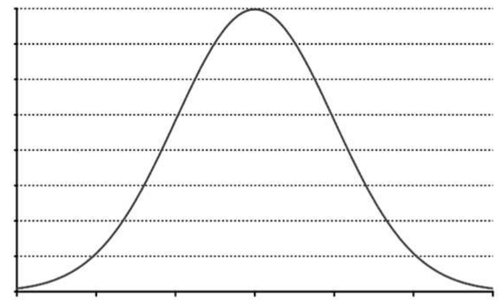
Esempi:

- Pressione sistolica
- Livello ematico nel sangue
- Altezza
- Peso

SESM

DISTRIBUZIONE NORMALE

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



SESM

Infatti:

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

- 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, per n sufficientemente grande (n>30) e p non troppo piccolo (p>0.05).

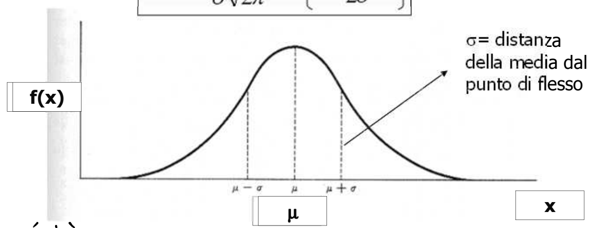
- Trasformazioni matematiche (log, √,…) possono "normalizzare" una variabile che naturalmente non lo sarebbe

SESM

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

se la sua p.d.f è data da:

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

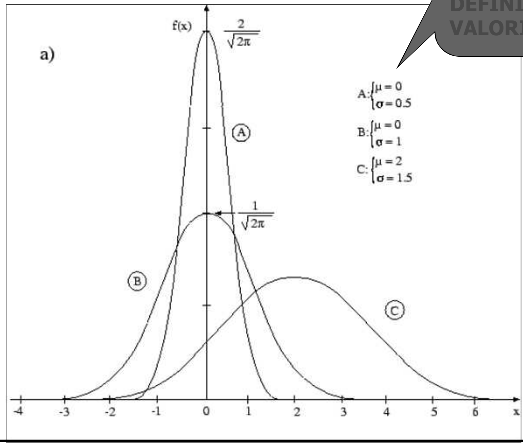


Proprietà:

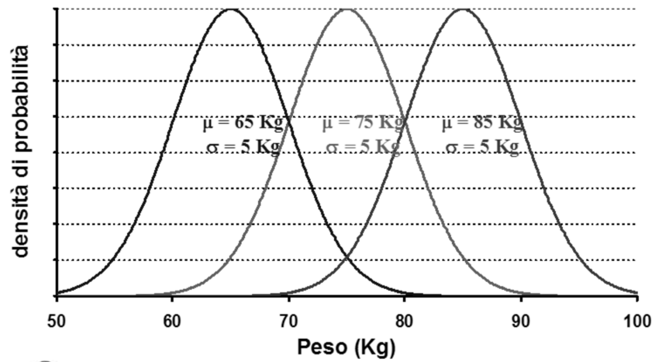
- Distribuzione simmetrica → f(-x)=f(x)
- media=moda=mediana
- $P(\mu - \sigma \leq x \leq \mu + \sigma) = 0.68$
- $P(\mu - 1.96\sigma \leq x \leq \mu + 1.96\sigma) = 0.05$

SESM

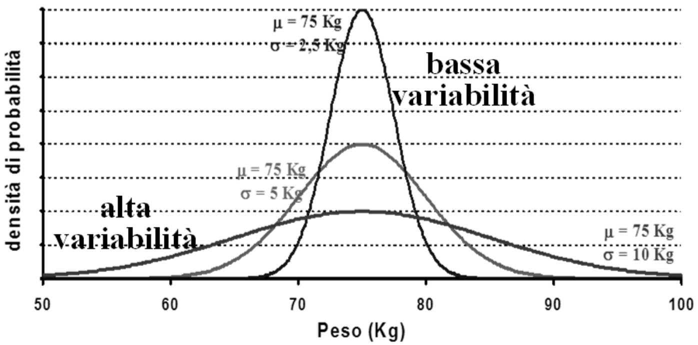
LA FORMA SPECIFICA DELLA DISTRIBUZIONE E' DEFINITA DAI VALORI DI μ e σ



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***:

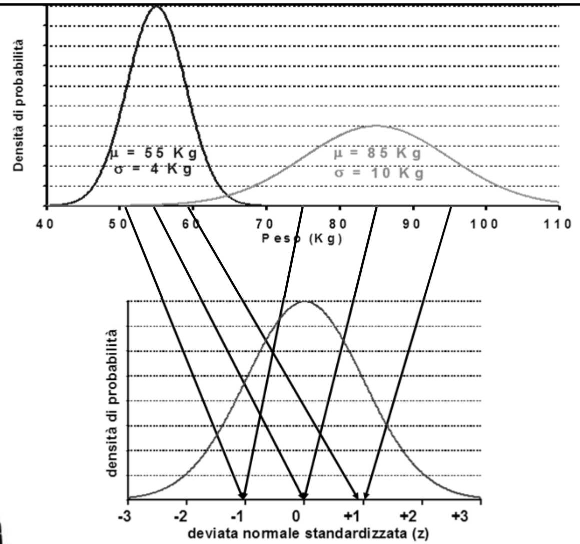
Definizione:: Sia X una variabile casuale distribuita normalmente,

$X \sim N(\mu, \sigma)$, allora la nuova variabile **Z:**

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

avrà una distribuzione normale con $\mu=0$ e $\sigma=1$. **[Z~N(0, 1)]**

Z è detta <u>DEVIATA NORMALE STANDARDIZZATA</u>



Esistono delle tavole (tavole della z) che danno la probabilità che Z sia maggiore di un valore qualsiasi. P(Z≥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	0,4960	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≥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	[illegible]	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	[illegible]	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	[illegible]	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	[illegible]	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	[illegible]	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	[illegible]	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	[illegible]	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	[illegible]	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	[illegible]	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	[illegible]	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	[illegible]	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	[illegible]	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	[illegible]	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	[illegible]	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	[illegible]	0,0694	0,0681
1,5	0,0668	0,0655	0,06430	0,0630	0,0618	0,0606	0,0594	[illegible]	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	[illegible]	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	[illegible]	0,0375	0,0367
1,8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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≥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	[illegible]	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	[illegible]	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	[illegible]	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	[illegible]	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	[illegible]	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	[illegible]	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	[illegible]	0,2546	0,2514	0,2483	0,2451
0,7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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:

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

1. $Z>1.30$
2. $Z<-0.85$
3. $0.50<Z<1.05$

B. Utlizzando le tavole di Z, calcolare quel valore che ha una probabilità del 35% di essere superato.

Esercizio 2: Si supponga che nella popolazione maschile adulta italiana la variabile "peso in kg" sia: ***X~N(75, 8):***

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

1. Un soggetto preso a caso abbia un peso ≤63 kg
2. Un soggetto abbia un peso compreso tra 69 e 92

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

Soluzione:

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/ml** ed una **deviazione standard= 9 mg/ml.**
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: $\mathbf{z = (90-105)/9 = -1.67}$
   e quello relativo a 125: $\mathbf{z = (125-105)/9 = 2.22}$

$\mathbf{P(-1.67 \le Z \le 2.22) = 1-[P(Z \le -1.67)+P(Z>2.22)] = 1-[P(Z>1.67)+P(Z>2.22)] =}$

$\mathbf{= 1-(0.049+0.013) = 0.938 = 93.8\%}$

2. Cerco il valore di Z tale per cui:

$P(Z \leq c)=0.1 \Rightarrow C=-1.28 \Rightarrow -1.28=(x-105)/9 \Rightarrow X=93.5$

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

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

Quindi l'intervallo ricercato sarà:

$\mu \mp 1.96\,\sigma = 105 \mp 1.96\ 9 \Rightarrow$ 87.4-122.6