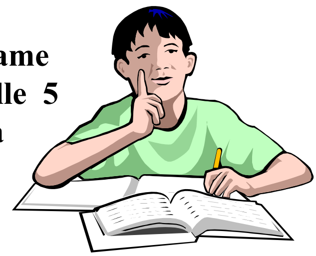
Variabili aleatorie

Distribuzione binomiale

Si supponga che uno studente affronti un esame composto da 3 domande chiuse. Una sola delle 5 alternative di risposta proposta per ciascuna domanda è vera



DOMANDA n. 1	
Il momento omogeneo centrale di ordine 1 di una variabile è:	
Risposte ammesse	
A	1
B	0
C	La media
D	La deviazione standard
E	Nessuna delle precedenti

Supponiamo anche che lo studente non abbia studiato e che risponda al test a caso.

Possiamo prevedere che risultato otterrà?

Variabili aleatorie

Distribuzione binomiale

Possiamo ricondurre il problema precedente al concetto di v.a.

•Abbiamo l'insieme omega degli eventi semplici (il risultato delle tre domande)

•Abbiamo il supporto della variabile aleatoria (0,1,2,3 i punteggi che lo studente può ottenere a seconda delle risposte *azzeccate*)

•Abbiamo un valore di probabilità associati a ciascun valore del supporto della v.a.

d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3
no	no	no	sì	no	no	no	sì	no	no	no	sì	sì	sì	no	sì	no	sì	no	sì	sì	sì	sì	sì

0 **1** **2** **3**

P(0) **P(1)** **P(2)** **P(3)**

Statistica inferenziale

Variabile aleatoria unidimensionale discreta

Esempio lancio di un dado

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4 \omega_5, \omega_6\}$$

X= 1 2 3 4 5 6

P(1) P(2) P(3) P(4) P(5) P(6)

Ω

Ci sono sei eventi possibili (le sei facce del dado)

X

A ciascun evento (faccia) associamo un codice numerico (supporto della v.a.)

P(x)

A ciascun codice viene assegnata una probabilità che in tutti i casi vale 1/6

Variabili aleatorie

Distribuzione binomiale

La probabilità di rispondere correttamente è uguale a 1/5

DOMANDA n. 1

Il momento omogeneo centrale di ordine 1 di una variabile è:

Risposte ammesse

A	1
B	0
C	La media
D	La deviazione standard
E	Nessuna delle precedenti

La probabilità di rispondere non correttamente è uguale a 4/5 (1-1/5)

Variabili aleatorie

Distribuzione binomiale

d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3
no	no	no	sì	no	no	no	sì	no	no	no	sì	sì	sì	no	sì	no	sì	no	sì	sì	sì	sì	sì

0 **1** **2** **3**

La probabilità della sequenza no - no - no è uguale al prodotto delle probabilità degli eventi semplici (ipotesi di indipendenza).

$$P_{(no,no,no)} = \frac{1}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} = \frac{1}{125} = 0{,}008$$

La probabilità che lo studente prenda zero è uguale a 0,008

Variabili aleatorie

Distribuzione binomiale

sequenza 1			sequenza 2			sequenza 3			sequenza 4			sequenza 5			sequenza 6			sequenza 7			sequenza 8		
d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3	d1	d2	d3
no	no	no	sì	no	no	no	sì	no	no	no	sì	sì	sì	no	sì	no	sì	no	sì	sì	sì	sì	sì

0 **1** **2** **3**

La probabilità di ciascuna sequenza è uguale al prodotto delle probabilità degli eventi semplici (successo/insuccesso) che lo compongono (ipotesi di indipendenza).

sequenza	p	voto	voto
sequenza 1	p(1)	0	p(1)
sequenza 2	p(2)	1	p(2) + p(3) + p(4)
sequenza 3	p(3)		
sequenza 4	p(4)		
sequenza 5	p(5)	2	p(5) + p(6) + p(7)
sequenza 6	p(6)		
sequenza 7	p(7)		
sequenza 8	p(8)	3	p(8)

Al supporto della v.a. (0,1,2,3) è possibile associare un valore di probabilità

Il risultato ottenuto è un esempio di v.a.u. discreta in particolare un esempio di avariabile aleatoria binomiale

Variabili aleatorie

Distribuzione binomiale

Quando gli eventi possono assumere soltanto due valori (testa o croce) la distribuzione teorica di probabilità assume una forma precisa chiamata funzione binomiale di probabilità

p= probabilità che si presenti un certo evento (successo)

q= probabilità che non si presenti un certo evento (insuccesso)

$$q = 1 - p$$

Distribuzioni di probabilità

Distribuzione binomiale

*La probabilità che l'evento si presenti esattamente **k** volte in **n** prove è data da*

$$P_k = \binom{N}{k} p^k q^{n-k}$$

Combinazioni di classe k di N elementi

Distribuzioni di probabilità

Distribuzione binomiale

Esempio: probabilità di ottenere due volte testa, lanciando una moneta tre volte

1° lancio	2° lancio	3° lancio
C	*C*	*C*
C	*C*	*T*
C	*T*	*C*
C	*T*	*T*
T	*C*	*C*
T	*C*	*T*
T	*T*	*C*
T	*T*	*T*

23

13

12

Combinazioni di classe 2 di 3 elementi

Distribuzioni di probabilità

Distribuzione binomiale

Esempio: probabilità di ottenere due volte testa, lanciando una moneta tre volte

q	*p*	*p*
p	*q*	*p*
p	*p*	*q*

q*p*p $=p^2*q$

p*q*p $=p^2*q$

p*p*q $=p^2*q$

p=0,5;

q=0,5

$$P_2=\binom{3}{2}p^2q^1 \quad \boxed{=\frac{n!}{k!(n-k)!}*p^k*q^{n-q}} \quad =\frac{6}{2*(1)}*0{,}5^2*0{,}5 \quad =$$

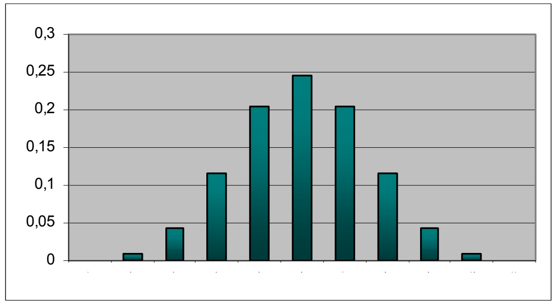
=6/2 * 0,125 = 3* 0,125= 0,375

Distribuzioni di probabilità

Distribuzione binomiale

Esempio: probabilità di ottenere n volte testa, lanciando una moneta 10 volte

n=10	p=0,5
k	p
0	0,001
1	0,01
2	0,044
3	0,117
4	0,205
5	0,246
6	0,205
7	0,117
8	0,044
9	0,01
10	0,001



Distribuzioni di probabilità

Distribuzione binomiale

Valori attesi

$$\mu = E(X) = \sum_{i=1}^{N} x_i * p(x_i) = \boxed{Np}$$

$$\sigma^2 = VAR(X) = E\left(X-\mu\right)^2 = \boxed{Np(1-p)}$$

Distribuzioni di probabilità

Distribuzione binomiale

Riassumendo:

Funzione di probabilità	$P_k = \binom{N}{k} p^k q^{n-k}$
Argomenti	f(X,p,N)
Media	Np
Varianza	Np(1-p)

Distribuzioni di probabilità

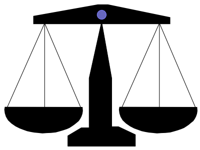
Distribuzione di Poisson

La distribuzione di Poisson approssima la distribuzione binomiale quando N è grande e p è prossimo a 0 (evento raro)

Funzione di probabilità	$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$
Argomenti	f(X, λ)
Media	λ
Varianza	λ

Distribuzioni di probabilità

Distribuzione Normale



Si immagini di dover misurare (es: "pesare") un oggetto. Ripetendo più volte la misurazione otterremo risultati lievemente diversi determinati dal peso dell'oggetto (T) e da un errore dovuto ad una perturbazione accidentale (e)

$$x_n = T + e_n$$

La probabilità di commettere un errore piccolo è maggiore di quella di ottenere un errore grande

C'è la stessa probabilità di ottenere un errore + o un errore -

Distribuzioni di probabilità

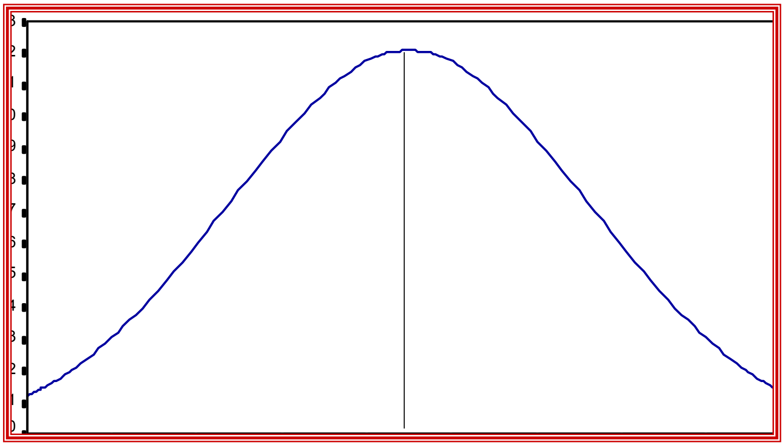
Distribuzione Normale

Graficamente la curva normale ha l'aspetto caratteristico di una curva a campana, è perfettamente simmetrica, con due punti di flesso $\mu + \sigma \; ; \; \mu - \sigma$.

La moda coincide con la mediana e la media della distribuzione

Distribuzioni di probabilità

Distribuzione Normale



Distribuzioni di probabilità

Distribuzione Normale

Funzione di densità $$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Argomenti $f(X,\ \mu,\sigma^2)$

Media μ

Varianza σ^2

Distribuzioni di probabilità

Distribuzione Normale

Intervalli tipici

$$P(\mu - \sigma \leq x \leq \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 0.6826$$

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = \int_{\mu-2\sigma}^{\mu+2\sigma} f(x)dx = 0.9545$$

$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx = 0.9974$$

Distribuzioni di probabilità

Distribuzione Normale Standardizzata

Una curva normale con media=0 e varianza =1 si dice curva standardizzata

$$z = \frac{x - \mu}{\sigma}$$

$x - \mu$ → Centratura

σ → Uniformazione

Funzione di densità

$$f(z) = \frac{1}{\sqrt{2\Pi}} e^{-z^2/2}$$

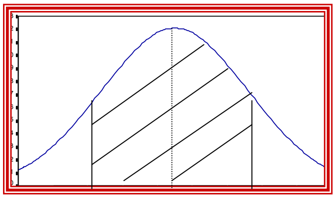
Argomenti	f(Z)
Media	0
Varianza	1

Distribuzioni di probabilità

Distribuzione Normale standardizzata

Intervalli tipici

Probabilità di avere un punteggio z compreso tra –1 e + 1 =
l'area sottostante le curva e delimitata dai punti di ascissa -1 +1

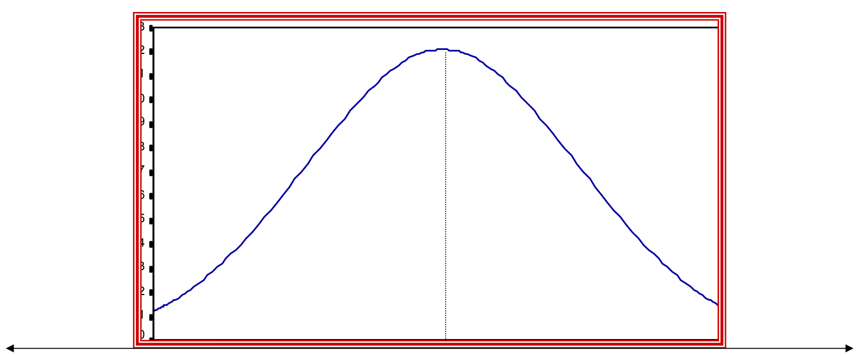


-1 0 +1

$$\int_{-1}^{+1} f(z)dz = 0{,}6826$$

Distribuzioni di probabilità

Distribuzione Normale Standardizzata



Mediante la tavola di probabilità è possibile ricavare l'area sottostante ad ogni porzione della curva compresa tra la media ed una certa ascissa |z|

Distribuzioni di probabilità

Uso Distribuzione Normale Standardizzata

ES:

A quale area della distribuzione normale standardizzata corrisponde un punteggio z compreso tra - 1.51 e + 1.51 ?

Consultare la tavola

Riga 1.5

Colonna .1

Valore corrispondente .4345

Area compresa tra 0 e + 1.51 =.4345 =43,45%

Area compresa tra –1.51 e +1.51 =.4345 *2 =0.8690 = 86,9%

Distribuzioni di probabilità

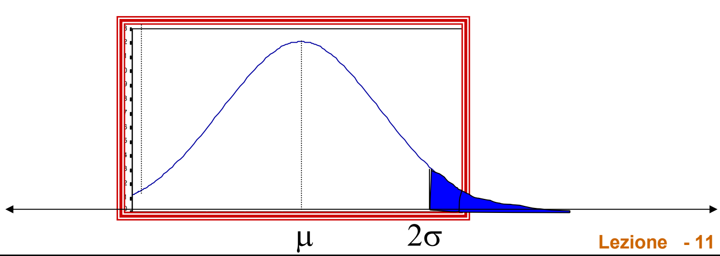
Uso Distribuzione Normale Standardizzata

ES:

I valori del test relativo ad una abilità cognitiva si distribuiscono normalmente con media ($\mu = 43$) e deviazione standard ($\sigma = 7$).
Il test viene somministrato a 150 soggetti; quanti studenti otterranno, probabilisticamente, un punteggio superiore a 57?

Riformulazione del quesito:

Probabilità di ottenere un risultato > = $\mu + 2\sigma \longrightarrow$ distribuzione *z*



Distribuzioni di probabilità

Uso Distribuzione Normale Standardizzata

ES:

Probabilità di ottenere un risultato >= μ+ 2σ $\longrightarrow$ distribuzione *z*

Consultare le tavole

Riga 2.0 colonna 0 (2volte la deviazione standard)

Valore osservato =.4772 $\longrightarrow$ il 47.72% ottiene un valore compreso tra la media e 57.

Il 50% ottiene un valore inferiore alla media

50% + 47.72% = 97.72% ottiene un valore minore o uguale a 57

100 % - 97.72= 2.28 % ottiene un valore maggiore di 57

150 *0,0228 = 3,42 soggetti ottengono un punteggio superiore a 57

Distribuzioni di probabilità

Altre distribuzioni "standardizzate"

Punti T = 10 z + 50

Media=50 Dev.std. =10

Il punteggio si muove tra 0 a 100

non utilizza i valori decimali

Scala più utilizzata nei reattivi psicologici

z	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
T	0	10	20	30	40	50	60	70	80	90	100

Distribuzioni di probabilità

Altre distribuzioni "standardizzate"

	test	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	4	3.57	4	4.05
2	2	6	5.95	10	11.90
3	3	4	7.14	14	16.67
	4	7	8.33	21	25.00
4	5	7	8.33	28	33.33
5	6	7	8.33	35	41.67
	7	7	8.33	42	50.00
6	8	7	8.33	49	58.33
	9	8	9.52	57	67.85
	10	5	7.14	62	73.81
	11	5	5.95	67	79.76
7	12	6	7.14	73	86.90
8	13	4	5.95	77	91.67
9	14	4	4.76	81	96.43
	15	3	3.57	84	100.00

Distribuzioni di probabilità

Altre distribuzioni "standardizzate"

Stanine

Range 1-9 Media= 5 Deviazione standard = 2 (1.96)

The SAS System
The FREQ Procedure

stanine	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3	3.57	3	3.57
2	5	5.95	8	9.52
3	13	15.48	21	25.00
4	7	8.33	28	33.33
5	14	16.67	42	50.00
6	25	29.76	67	79.76
7	5	5.95	72	85.71
8	5	5.95	77	91.67
9	7	8.33	84	100.00

Distribuzioni di probabilità

Distribuzione Chi quadro

Si chiama Chi quadro la sommatoria dei quadrati di N variabili normali standardizzate distribuite in modo indipendente

$$\chi^2 = \sum_{i=1}^{N} z_i^2$$

Argomenti *ν (=gradi di libertà)*

range $0 \le \chi^2 \le +\infty$

Distribuzioni di probabilità

Distribuzione Chi quadro

Gradi di libertà

Quando le variabili non sono indipendenti occorre stabilire le condizioni che le vincolano

	Maschi	Femmi.	totale
Giovani	50	10	60
Anziani	20	20	40
totale	70	30	100

Distribuzioni di probabilità

Distribuzione F

Rapporto tra due v.a. chi quadro indipendenti, ciascuna divisa per il rispettivo numero di gradi di libertà

$$F = \frac{\chi_a^2}{\chi_b^2} \frac{\nu_b}{\nu_a}$$

La distribuzione F è alla base dell'analisi della varianza tecnica fondamentale in Psicologia

Distribuzioni campionarie

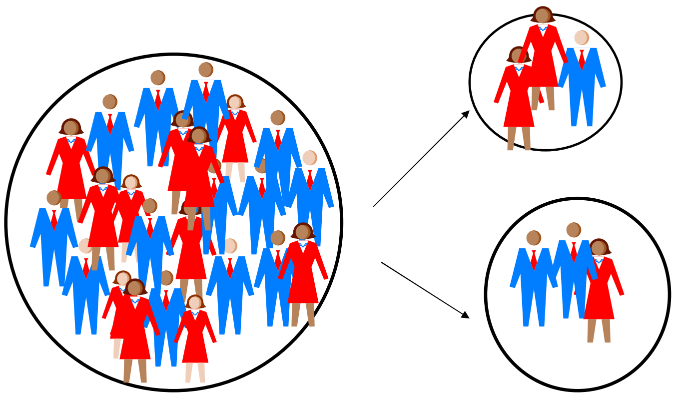
Campione bernoulliano =campione nel quale dopo ogni estrazione l'unità selezionata viene reintrodotta

L'estrazione di ogni unità è indipendente dalle precedenti estrazioni

Si immagini di estrarre da una certa popolazione (finita o infinita) un campione bernoulliano ($x_1,x_2,x_3,.. x_n$)

Distribuzioni campionarie

Possiamo ora pensare all'estrazione di tutti i possibili campioni estraibili dalla stessa popolazione



Distribuzioni campionarie

Possiamo ora pensare all'estrazione di tutti i possibili campioni estraibili dalla stessa popolazione

Campione 1	x_1 ,	x_2 ,	x_3	, … ,	x_n
Campione 2	x_1 ,	x_2 ,	x_3	, … ,	x_n
……………	….	,…	,…..	, ….	, ….
Campione K	x_1 ,	x_2 ,	x_3	, … ,	x_n
Variabile aleatoria	X_1 ,	X_2 ,	X_3	, … ,	X_n

Distribuzioni campionarie

Variabile aleatoria $X_1, \ X_2, \ X_3, \dots, X_n$

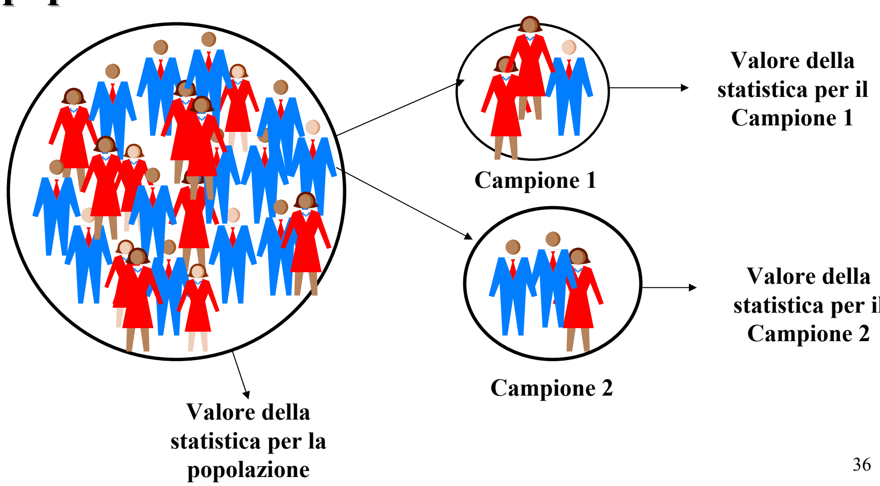
$$P(X_1, \ X_2, \ X_3, \dots, X_n) = \prod p(x_i) \text{ con } i = 1,2,3, \dots, n$$

Ogni campione sarà caratterizzato da un certo valore per una generica statistica

L'insieme di tutti i valori che la statistica assume per i diversi campioni e della funzione di probabilità ad essi associata definisce la distribuzione campionaria della statistica

Distribuzioni campionarie

Possiamo ora pensare all'estrazione di tutti i possibili campioni estraibili dalla stessa popolazione



Distribuzioni campionarie

Tutti i campioni	p dei campioni	valori statistica	probab.
Campione 1 (a,b,c) →	p(c1) →	f(x)1	p(c1)
Campione 2 (a,b,d) →	p(c2) →	f(x)2	p(c2)
Campione 3 (d,b,c) →	p(c3) →	f(x)3	p(c3)
Campione 4 (a,a,c) →	p(c4) →	f(x)4	p(c4)
Campione 5 (b,b,c) →	p(c5) →	f(x)5	p(c5)
Campione 6 (c,b,c) →	p(c6) →	f(x)6	p(c6)
Camp.N (x_1,x_2,x_3) →	p(cN) →	f(x)N	p(cN)

Distribuzioni campionarie

v.a. somma campionaria

Si definisca a partire da una popolazione avente media= μ e varianza= σ^2 la variabile aleatoria somma campionaria

$$U = X_1 + X_2 + X_3 + \ldots + X_n$$

Il valore atteso di U = E(U)=n μ

La varianza è pari a VAR(U)=n σ^2

Teorema del limite centrale:

la variabile aleatoria $Z_U = \dfrac{U - E(U)}{\sigma\sqrt{n}}$ **all'aumentare di n si distribuisce come un v.a. Normale standardizzata qualunque sia la distribuzione delle X**

Distribuzioni campionarie

v.a. media campionaria

Si prenda una popolazione formata da 6 individui

id	Kg
1	65
2	72
3	78
4	57

$\mu = 68$

$\sigma^2 = 82$

Estraiamo da questa popolazione alcuni dei possibili campioni estraibili

Distribuzioni campionarie

v.a. media campionaria

id	Kg
1	65
2	72
3	78

$\overline{X} = 71{,}7$

id	Kg
1	65
2	72
4	57

$\overline{X} = 64{,}7$

id	Kg
1	65
3	78
4	57

$\overline{X} = 66{,}7$

id	Kg
2	72
3	78
4	57

$\overline{X} = 69{,}0$

Valori medie campionarie	n	*f*	probabilità
64,7	1	0,25	0,25
66,7	1	0,25	0,25
69,0	1	0,25	0,25
71,0	1	0,25	0,25

Distribuzioni campionarie

v.a. media campionaria

Valori medie campionarie	**n**	***f***	probabilità
64,7	n°camp	n°/N	0,…
64,8	n°camp	n°/N	0,…
64,9	n°camp	n°/N	0,…
65.0	n°camp	n°/N	0,…
65.1	n°camp	n°/N	0,…
65.2	n°camp	n°/N	0,…
…	n°camp	n°/N	0,…
μ	n°camp	n°/N	0,…
…	n°camp	n°/N	0,…
….	n°camp	n°/N	0,…
71,0	n°camp	n°/N	0,…

Distribuzioni campionarie

v.a. media campionaria

Valori medie campionarie	**n**	***f***	probabilità
64,7	n°camp	n°/N	0,…
64,8	n°camp	n°/N	0,…
64,9	n°camp	n°/N	0,…
65.0	n°camp	n°/N	0,…
65.1	n°camp	n°/N	0,…
65.2	n°camp	n°/N	0,…
…	n°camp	n°/N	0,…
μ	n°camp	n°/N	0,…
…	n°camp	n°/N	0,…
….	n°camp	n°/N	0,…
71,0	n°camp	n°/N	0,…

Questa è una variabile aleatoria.

I diversi valori della statistica campionaria rappresentano gli esiti possibili dell'esperimento "*estrazione di un campione da una popolazione*". Ad ogni Evento è possibile associare una probabilità (la probabilità di estrarre campioni aventi quel valore per la statistica).

Distribuzioni campionarie

v.a. media campionaria

Siano $\mathbf{X_1 + X_2 + X_3 + \ldots + X_n}$ **delle v.a. associate all'estrazione di un campione di ampiezza n estratto da una popolazione con media** μ **e varianza** σ^2**.**

La media campionaria è a sua volta una v.a. definita da

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n} U$$

Distribuzioni campionarie

v.a. media campionaria

VALORI CARATTERISTICI

$$\mu_{\bar{x}} = E(\bar{X}) = \frac{1}{n} n\mu = \mu$$

$$VAR(\bar{X}) = \frac{1}{n} \sigma^2$$

Fattore di normalizzazione per la varianza se il campione non è bernoulliano $\left(\frac{N-n}{N-1} \right)$

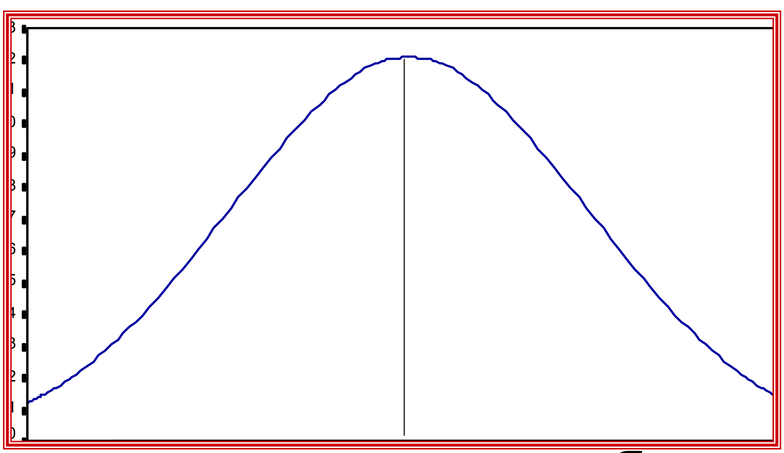
Distribuzioni campionarie

v.a. media campionaria

Se si estraggono ripetutamente campioni casuali di dimensione n da un qualsiasi universo qualunque sia la sua forma con media μ e varianza σ^2, all'aumentare della dimensione del campione la distribuzione della media campionaria tenderà ad avvicinarsi alla normale ed avrà come media μ e come varianza σ^2/n

Distribuzioni di probabilità

V.a. Distribuzione media campionaria (per N>30)



$$\mu \longrightarrow \frac{\sigma}{\sqrt{N}}$$

v.a. varianza campionaria

A differenza di quanto visto per la media campionaria non si verifica l'eguaglianza perfetta tra valore atteso della varianza campionaria e valore della varianza dell'universo

$$E\left(s^2\right) = \frac{n-1}{n}\sigma^2 \Leftrightarrow \hat{s}^2 = \frac{n}{n-1}s^2$$

v.a. media campionaria

variabile	**valore atteso**	**varianza**
somma campionaria	$n\mu$	$n\sigma^2$
media campionaria	μ	$\frac{1}{n}\sigma^2$
varianza campionaria	$\frac{n-1}{n}\sigma^2$	funzione del momento centrale di ordine 4
diferenza tra due medie	$\mu_1 - \mu_2$	$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$