

Corso di Statistica Facoltà di Economia

a.a. 2010-2011

francesco mola

Lezione n° 5

sommario

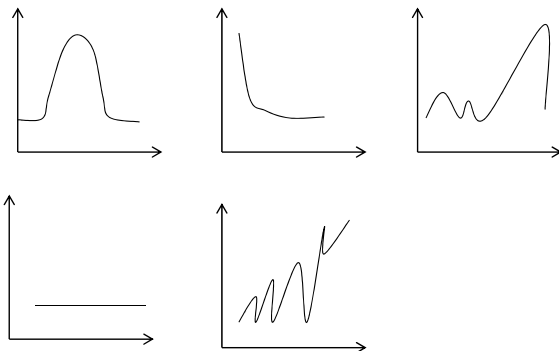
- Indici di forma
- Variabile standardizzata

lez5_2010-2011

statistica-francesco mola

2

Forma di una distribuzione

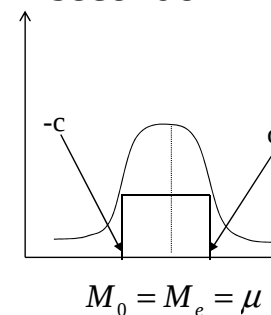


lez5_2010-2011

statistica-francesco mola

3

Molti fenomeni della realtà si distinguono
secondo il modello della curva Normale



$$f(M_e - c) = f(M_e + c)$$

$\Downarrow$

f è simmetrica

$M_0 < M_e < \mu \Rightarrow$ asimmetria positiva (destra)

$\mu < M_e < M_0 \Rightarrow$ asimmetria negativa (sinistra)

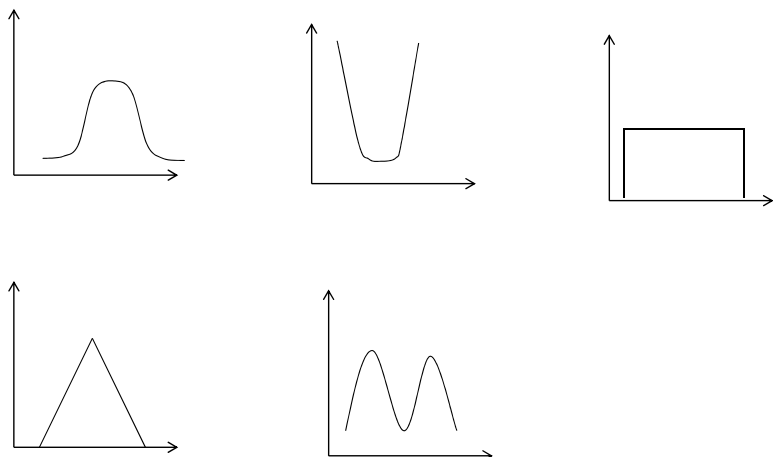
lez5_2010-2011

statistica-francesco mola

4

Forma di una distribuzione

Distribuzioni statistiche simmetriche



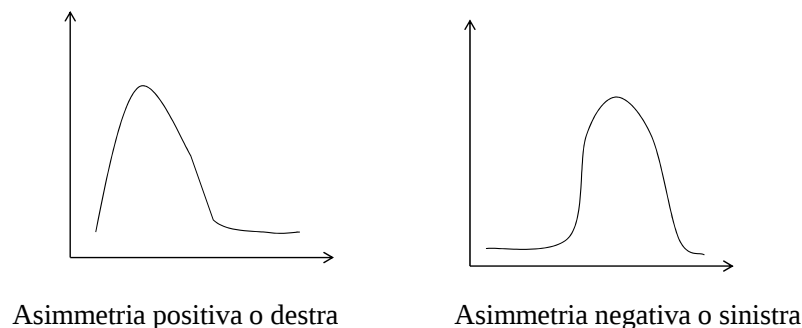
lez5_2010-2011

statistica-francesco mola

5

Forma di una distribuzione

Distribuzioni asimmetriche



Asimmetria positiva o destra

Asimmetria negativa o sinistra

lez5_2010-2011

statistica-francesco mola

6

Indici di asimmetria

$$A_1 = \mu - M_e$$

$$A_2 = \frac{\mu - M_e}{\sigma} \text{ poichè si dimostra che } |\mu - M_e| \leq \sigma \Rightarrow$$

$$\Rightarrow -1 \leq A_2 \leq 1$$

$$A_2 < 0 \Rightarrow \text{asimmetria negativa}$$

$$A_2 = 0 \Rightarrow \text{simmetria}$$

$$A_2 > 0 \Rightarrow \text{asimmetria positiva}$$

lez5_2010-2011

statistica-francesco mola

7

Indice di Fisher

$$\gamma_1 = \frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^3$$

Può assumere valori tra $+\infty$ e $-\infty$

$$\gamma_1 < 0 \quad \text{Negativa}$$

$$\gamma_1 = 0 \quad \text{Nulla}$$

$$\gamma_1 > 0 \quad \text{Positiva}$$

lez5_2010-2011

statistica-francesco mola

8

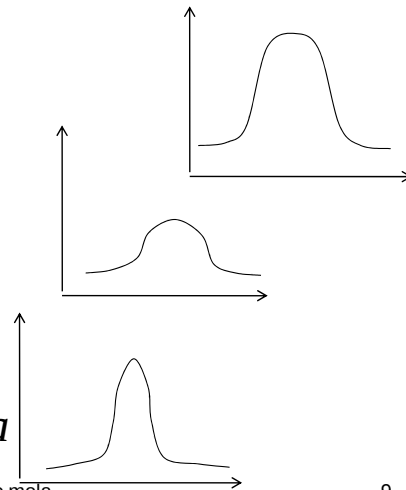
Indice di Curtosi

$$P_2 = \frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^4$$

$P_2 = 3 \Rightarrow \text{Normale}$

$P_2 < 3 \Rightarrow \text{Platicurtica}$

$P_2 > 3 \Rightarrow \text{Leptocurtica}$



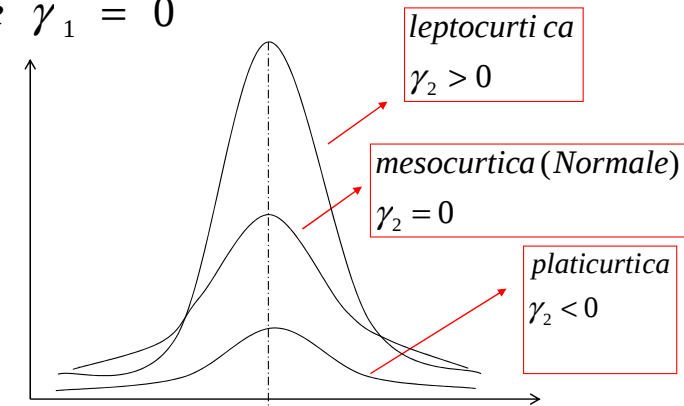
lez5_2010-2011

statistica-francesco mola

9

$$\gamma_2 = P_2 - 3 = \frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3$$

se $\gamma_1 = 0$



lez5_2010-2011

$M_0 = M_e = \mu$

statistica-francesco mola

10

Calcoliamo gli indici di Asimmetria e Curtosi per la tabella relativa al tabacco

MARCA	condensato	peso	monossido	prezzo
Alpine	14.1	0.86	13.6	medio
Benson&Hedges	16	1.06	16.6	alto
BullDurham	29.8	2.03	23.5	alto
Camellights	8	0.67	10.2	medio
Carlton	4.1	0.4	5.4	basso
Chesterfield	15	1.04	15	alto
GoldenLights	8.8	0.76	9	medio
Kent	12.4	0.95	12.3	medio
Kool	16.6	1.12	16.3	alto
L&M	14.9	1.02	15.4	medio
LarkLights	13.7	1.01	13	medio
Marlboro	15.1	0.9	14.4	alto
Merit	7.8	0.57	10	medio
MultiFilter	11.4	0.78	10.2	medio
NewportLights	9	0.74	9.5	medio
Now	1	0.13	1.5	basso
OldGold	17	1.26	18.5	alto
PallMallLight	12.8	1.08	12.6	medio
Raleigh	15.8	0.96	17.5	alto
SalemUltra	4.5	0.42	4.9	basso
Tareyton	14.5	1.01	15.9	medio
True	7.3	0.61	8.5	medio
ViceroyRichLight	8.6	0.69	10.6	medio
VirginiaSlims	15.2	1.02	13.9	alto
WinstonLights	12	0.82	14.9	medio

lez5_2010-2011

statistica-francesco mola

11

	Condensato	Peso	Monossido
Asimmetria A_1	-0,584	-0,023	-0,472
A_2	-0,105	-0,068	-0,101
Curtosi γ_1	0,71	0,909	-0,181
γ_2	2,162	3,132	0,312

lez5_2010-2011

statistica-francesco mola

12

Variabile standardizzata

$$X = x_1, x_2, x_3, \dots, x_n \Rightarrow Z = Z(X)$$

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

$$\mu = 0$$

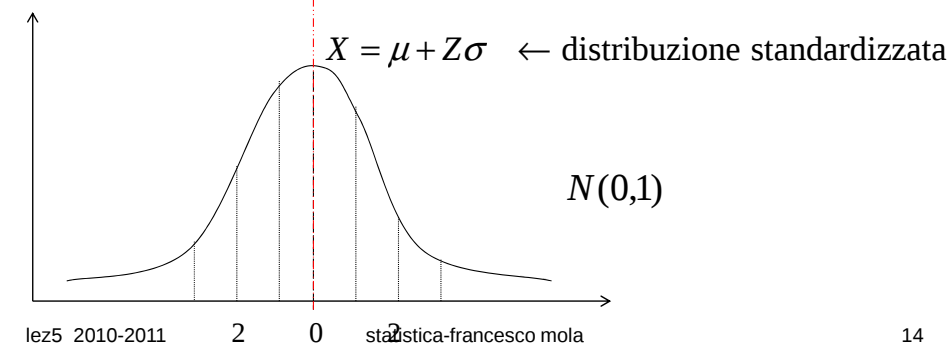
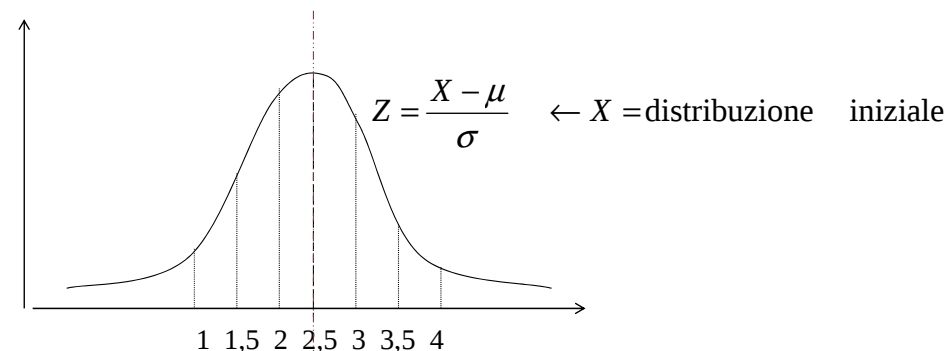
X	$Z = Z(X)$
x_1	$\frac{x_1 - \mu}{\sigma} = z_1$
x_2	$\frac{x_2 - \mu}{\sigma} = z_2$
x_3	$\frac{x_3 - \mu}{\sigma} = z_3$

$$\sigma^2 = 1$$

lez5_2010-2011

statistica-francesco mola

13



lez5_2010-2011

statistica-francesco mola

14

...tabacco

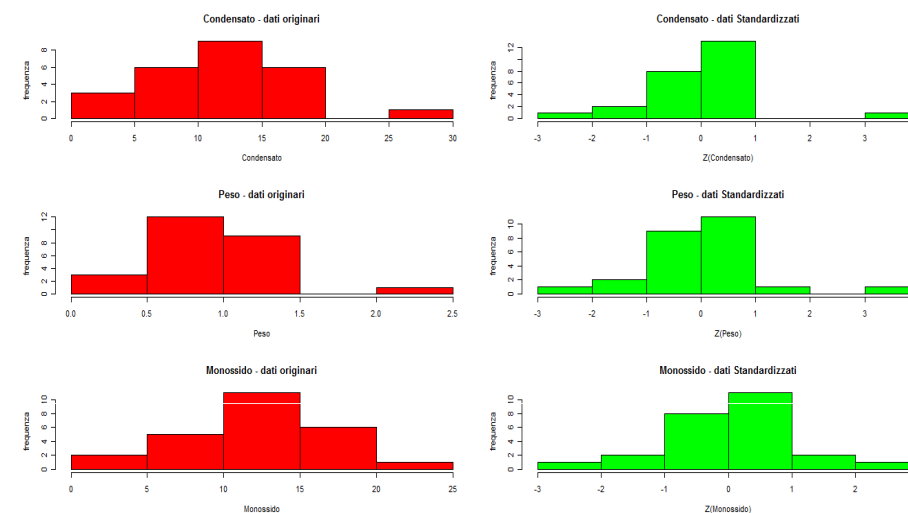
MARCA	valori originari (X)			valori standardizzati (Z)		
	condensato	peso	monossido	condensato	peso	monossido
Alpine	14,1	0,86	13,6	0,339	-0,047	0,231
Benson&Hedges	16	1,06	16,6	0,682	0,529	0,877
BullDurham	29,8	2,03	23,5	3,168	3,325	2,363
CamelLights	8	0,67	10,2	-0,759	-0,595	-0,501
Carlton	4,1	0,4	5,4	-1,462	-1,373	-1,535
Chesterfield	15	1,04	15	0,502	0,472	0,532
GoldenLights	8,8	0,76	9	-0,615	-0,336	-0,760
Kent	12,4	0,95	12,3	0,033	0,212	-0,049
Kool	16,6	1,12	16,3	0,790	0,702	0,812
L&M	14,9	1,02	15,4	0,483	0,414	0,618
LarkLights	13,7	1,01	13	0,267	0,385	0,102
Marlboro	15,1	0,9	14,4	0,520	0,068	0,403
Merit	7,8	0,57	10	-0,795	-0,883	-0,544
MultiFilter	11,4	0,78	10,2	-0,147	-0,278	-0,501
NewportLights	9	0,74	9,5	-0,579	-0,393	-0,652
Now	1	0,13	1,5	-2,020	-2,152	-2,375
OldGold	17	1,26	18,5	0,862	1,106	1,286
PallMallLight	12,8	1,08	12,6	0,105	0,587	0,016
Raleigh	15,8	0,96	17,5	0,646	0,241	1,071
SalemUltra	4,5	0,42	4,9	-1,390	-1,316	-1,643
Tareyton	14,5	1,01	15,9	0,411	0,385	0,726
True	7,3	0,61	8,5	-0,886	-0,768	-0,867
ViceroyRichLight	8,6	0,69	10,6	-0,651	-0,537	-0,415
VirginiaSlims	15,2	1,02	13,9	0,538	0,414	0,295
WinstonLights	12	0,82	14,9	-0,039	-0,163	0,511

lez5_2010-2011

statistica-francesco mola

15

Proviamo a rappresentare graficamente le distribuzioni originarie e i dati standardizzati



Cosa è cambiato?



lez5_2010-2011

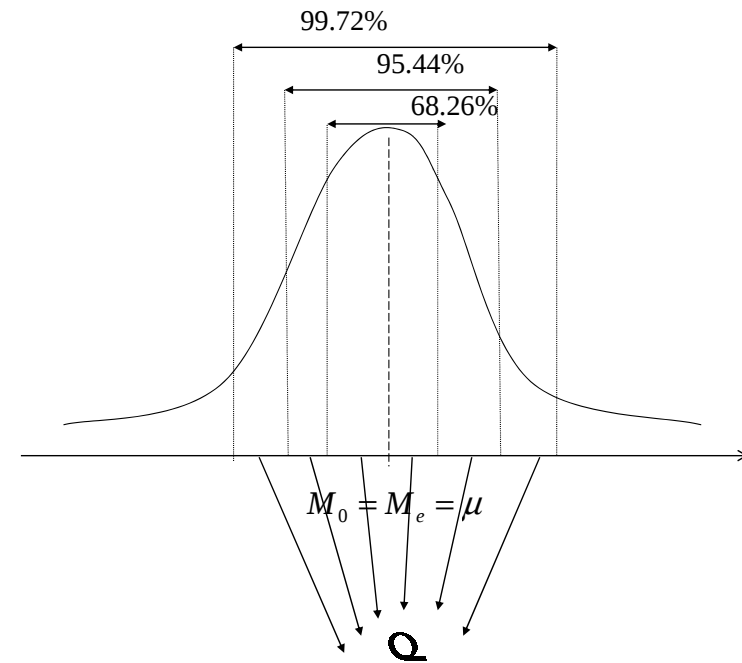
statistica-francesco mola

16

...la simmetria con riferimento alla Curva Normale

Equazione della funzione Normale o
di Gauss

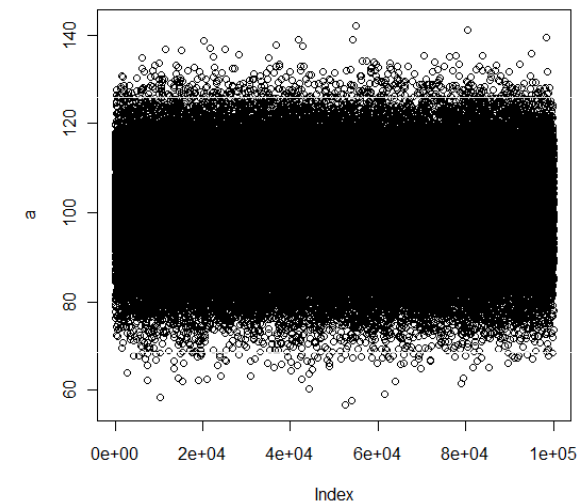
$$n(x; \mu; \sigma^2) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

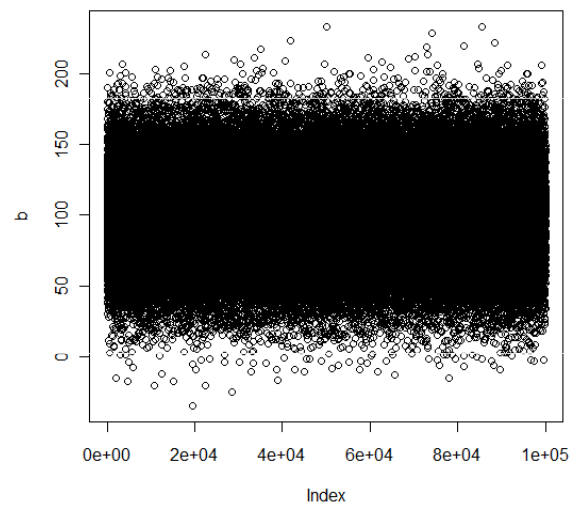


Quando si standardizzano dati provenienti da una distribuzione normale

Consideriamo due variabili a e b. Si tratta di 100000 osservazioni provenienti da due distribuzioni normali. La variabili hanno entrambe media 100 e la variabilità di b è molto superiore a quella della variabile a.

Vediamo cosa accade con la standardizzazione

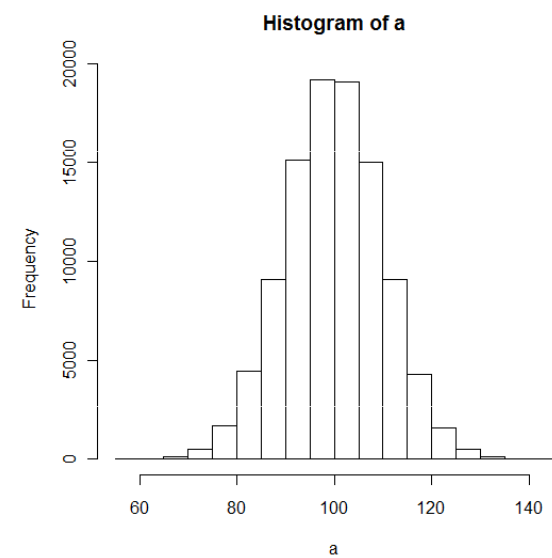




lez5_2010-2011

statistica-francesco mola

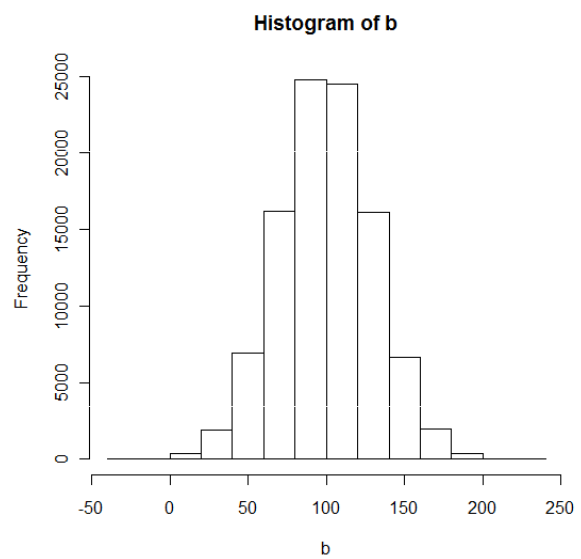
21



lez5_2010-2011

statistica-francesco mola

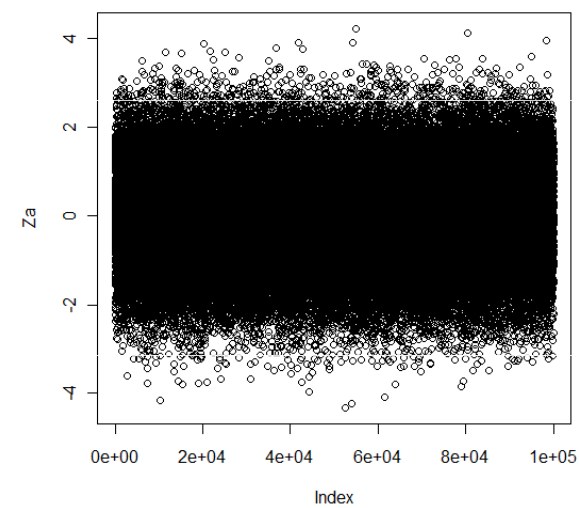
22



lez5_2010-2011

statistica-francesco mola

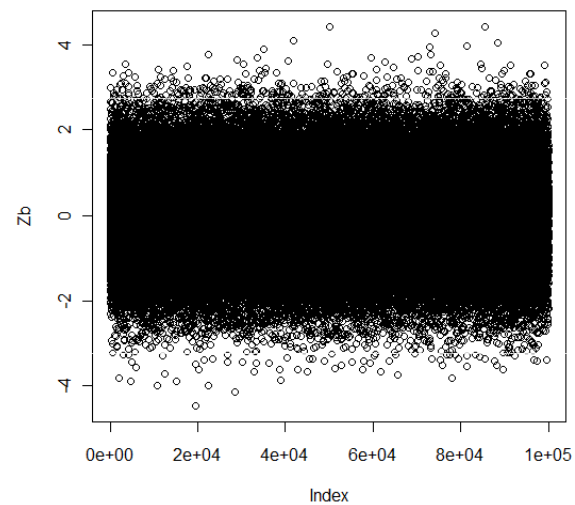
23



lez5_2010-2011

statistica-francesco mola

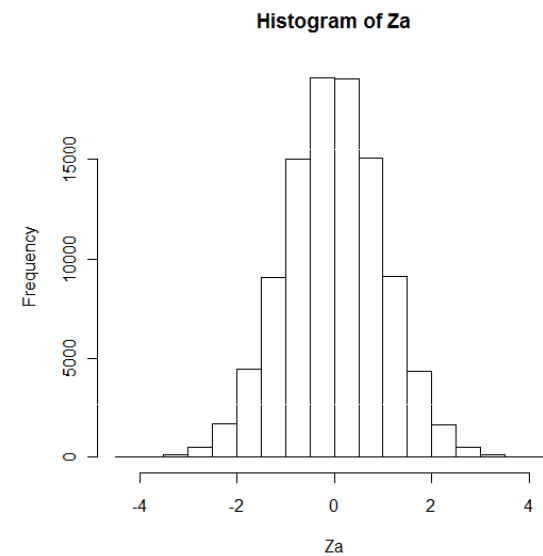
24



lez5_2010-2011

statistica-francesco mola

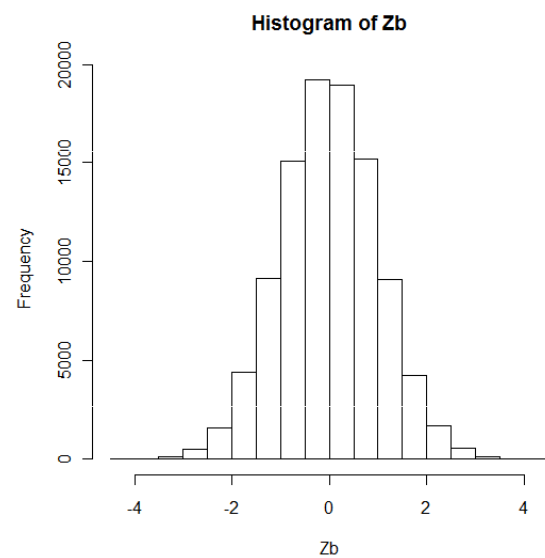
25



lez5_2010-2011

statistica-francesco mola

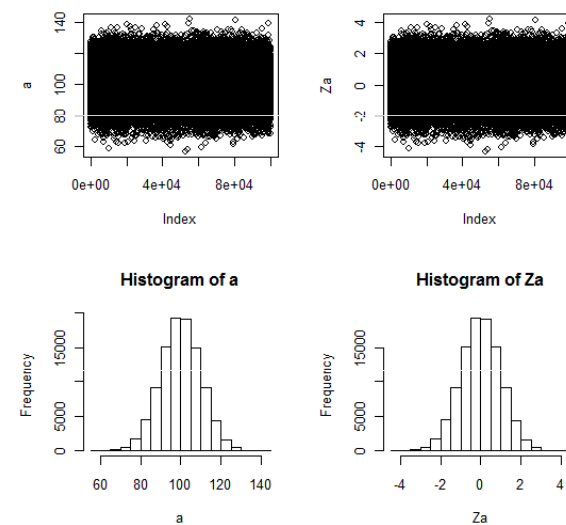
26



lez5_2010-2011

statistica-francesco mola

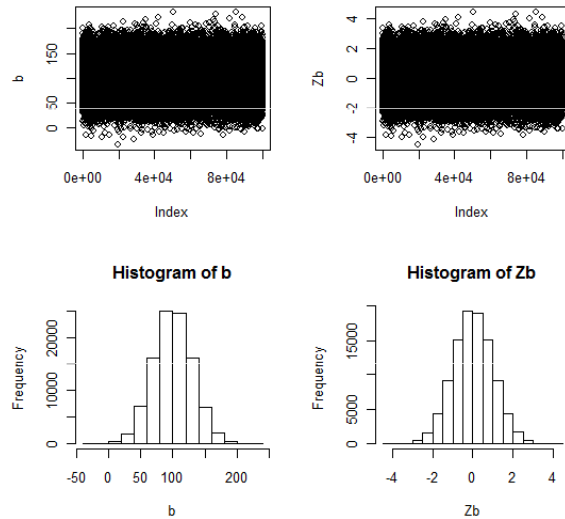
27



lez5_2010-2011

statistica-francesco mola

28



Una curva Normale è completamente rappresentata da μ e σ

Gode di simmetria rispetto a μ $\gamma_1 = \frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^3 = 0$

Ha forma campanulare

Le ascisse dei punti di flesso sono in $\mu \pm \sigma$

Gli estremi della curva sono asintotici

La conoscenza di un modello teorico corrispondente ad una distribuzione empirica consente:

- Regolarizzare un istogramma
- Ricostruire dati mancanti
- Confrontare istogrammi basati su classi di diversa ampiezza
- Descrivere sinteticamente i dati con i parametri del modello

E' possibile ricostruire le frequenze teoriche $\hat{n}_i$ relative ad Una certa classe X_i, X_{i+1} semplicemente integrando la funzione Normale su detto intervallo

$$\int_{x_i}^{x_{i+1}} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx$$

Applicazioni

- Calcolo della frazione delle osservazioni che cadono tra due valori z e $z + \Delta z$
- Calcolo della frazione che supera o è inferiore ad un certo z
- Determinazione delle frequenze assolute moltiplicando per n

Esempio

$$X = (x_0 - x_1, n_1), (x_1 - x_2, n_2) \cdots (x_{k-1} - x_k, n_k)$$

$$\mu = 160$$

$$\sigma = 10$$

$$n = 200$$

$$\hat{n}_i \Rightarrow 170 - 180$$

$$z_1 = \frac{170 - 160}{10} = 1$$

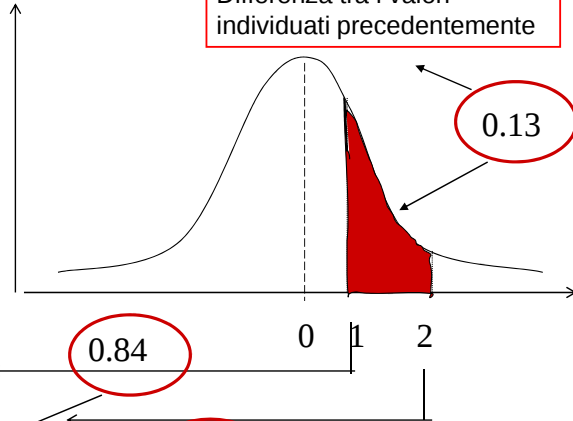
$$z_2 = \frac{180 - 160}{10} = 2$$

E' l'area fino al punto 1.

Dedotto dalle tavole

lez5_2010-2011

Non nota



0.84

0.97

0.13

E' l'area tra il punto 1 ed il punto 2.

Differenza tra i valori individuati precedentemente

E' l'area fino al punto 2.

Dedotto dalle tavole

33

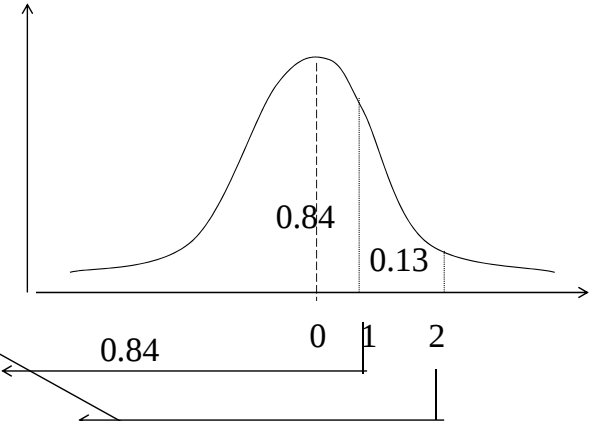
Esempio

$$0.97 - 0.84 = 0.13$$

$$0.13 * 200 = 26$$

E' la cardinalità!

lez5_2010-2011



0.84

0.13

0.84

0.97

Stima della frequenza della classe 170-180

statistica-francesco mola

34