

# Appendix

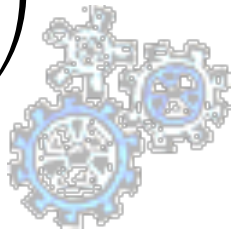


# Media Aritmetica

- Per effettuare la correzione di errori accidentali
  - permette di sostituire i valori di ogni elemento senza cambiare il totale
    - Sostituzione di valori NULL
- Monotona crescente

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\frac{1}{n+k} \left( \sum_{i=1}^n x_i + k\bar{x} \right) = \bar{x}$$



# Media Geometrica

$$x_g = \sqrt[n]{\prod_{i=1}^n x_i}$$

<i>Prodotto</i>	<i>Variazioni Prezzi</i>	
	1996	1997
A	100	200
B	100	50
<i>Media</i>	100	125

- Per bilanciare proporzioni
- dati moltiplicativi
- La media aritmetica dei logaritmi è il logaritmo della media geometrica
- Monotona crescente

$$x_g = 100$$

$$\log x_g = \frac{1}{n} \sum_{i=1}^n \log x_i$$



# Media Armonica

- Monotona decrescente
- Per misure su dimensioni fisiche
- E.g., serie temporali

$$x_a = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$



# Mediana

- Il valore centrale in un insieme ordinato di dati
- Robusta
  - poco influenzata dalla presenza di dati anomali

*1 7 12 18 23 34 54*

$$\bar{x} = 21.3$$

$$M = 23$$



# Mediana e Quartili

- Divide un insieme di dati a metà`
  - statistica robusta (non influenzata da valori con rilevanti differenze)
  - ulteriori punti di divisione
- interquartili
  - mediane degli intervalli dei dati superiore e inferiore
  - Un quarto dei dati osservati è sopra/sotto il quartile
- percentili
  - di grado p: il p% dei dati osservati è sopra/sotto il percentile
  - mediana: 50-esimo percentile
  - primo quartile: 25-esimo percentile
  - secondo quartile: 75-esimo percentile
- max, min
  - range = max-min



# Percentili

- Rappresentati con  $x_p$
- Utilizziamo le lettere per esprimerli

<i>Etichetta</i>	<i>P</i>
M	$\frac{1}{2}=0.5$
F	$\frac{1}{4}=0.25$
E	$\frac{1}{8}=0.125$
D	$\frac{1}{16}=0.0625$
C	$\frac{1}{32}=0.03125$
B	$\frac{1}{64}$
A	$\frac{1}{128}$
Z	$\frac{1}{256}$
Y	$\frac{1}{512}$
X	$\frac{1}{1024}$



# Moda

- Misura della frequenza dei dati

*a a b b c c a d b c a e c b a a*

moda = a ( $f = 6$ )

- Significativo per dati categorici
- Non risente di picchi
- Molto instabile





# Range, Deviazione media

- Intervallo di variazione

$$r = \max - \min$$

- Scarti interquantili

$$r_p = x_{100-p} - x_p$$

- Scarto medio assoluto

$$S_n = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

- Scarto medio assoluto dalla mediana

– In generale,  $S_{.5} \leq S_n$

$$S_M = \frac{1}{n} \sum_{i=1}^n |x_i - M|$$



# Varianza, deviazione standard

- misure di mutua variabilità tra i dati di una serie
- Devianza empirica
- Varianza
- Coefficiente di variazione
  - misura relativa

$$dev = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$V = \frac{s}{\bar{x}}$$



# Simmetria

- Si ha simmetria quando media, moda e mediana coincidono
  - condizione necessaria, non sufficiente
  - Asimmetria sinistra: moda, mediana, media
  - Asimmetria destra: media, mediana, moda



# Simmetria (Cont.)

- Indici di asimmetria

- medie interquartili

$$\bar{x}_p = (x_{1-p} + x_p) / 2$$

- Momenti centrali

$$m_k = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^k$$

- indice di Fisher

- $\gamma$  nullo per distribuzioni simmetriche

- $\gamma > 0$ : sbilanciamenti a destra

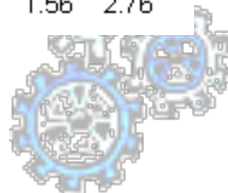
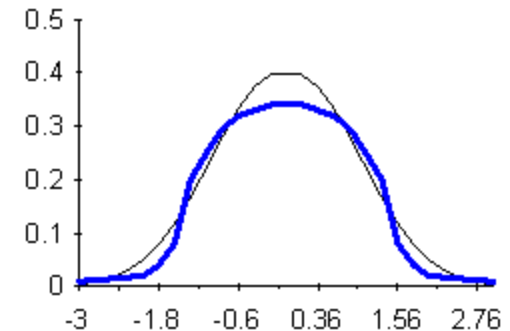
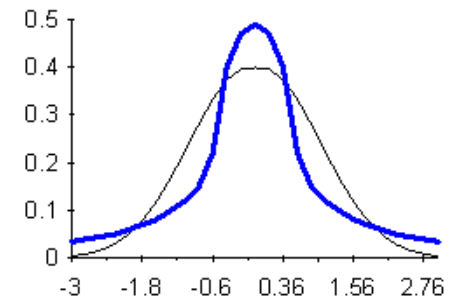
- $\gamma < 0$ : sbilanciamento a sinistra

$$\gamma = \frac{m_3}{\hat{S}^3}$$



# Curtosi

- Grado di appiattimento della curva di distribuzione rispetto alla curva normale
  - mesocurtica: forma uguale alla distribuzione normale;
  - leptocurtica: una frequenza minore delle classi intermedie, frequenza maggiore delle classi estreme e dei valori centrali;
  - platicurtica: una frequenza minore delle classi centrali e di quelle estreme, con una frequenza maggiore di quelle intermedie
    - numero più ridotto di valori centrali.

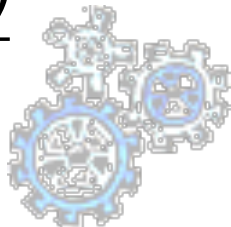


# Curtosi (cont.)

- Indice di Pearson
  - $\beta=3$ : distribuzione mesocurtica
  - $\beta > 3$ : distribuzione leptocurtica
  - $\beta < 3$ : distribuzione platicurtica
- Coefficiente di curtosi
  - Una distribuzione leptocurtica ha  $K \sim 1/2$
  - platicurtosi:  $k \sim 0$

$$\beta = \frac{m_4}{\hat{s}^4}$$

$$K = \frac{1/2 (x_{.75} - x_{.25})}{(x_{.90} - x_{.10})}$$



# Coefficienti di Correlazione

- Covarianza 
$$Cov(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- Coefficiente di Pearson

$$r_{xy} = \frac{Cov(x, y)}{S_x S_y}$$

