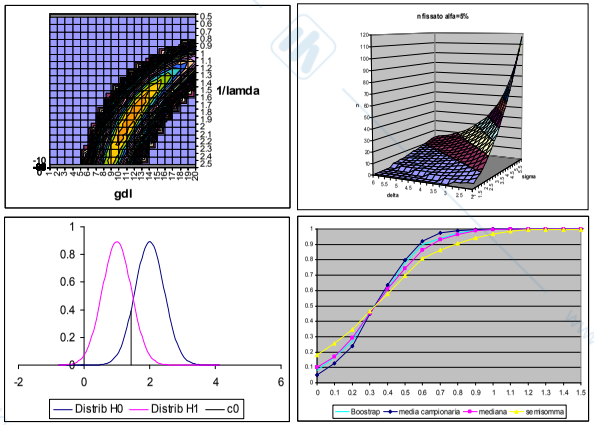
UNIVERSITÀ CATTOLICA DEL SACRO CUORE
LARGO A.GEMELLI, 1 - 20123 MILANO

Facoltà di Scienze Bancarie Finanziarie e Assicurative

Statistica I

prof. Diego Zappa

Appunti ad uso degli studenti



Milano – Febbraio 2021

Da *Guida dello studente*

DENOMINAZIONE DEL CORSO	STATISTICA I
CREDITI FORMATIVI (ORE DI AULA)	10 CFU (70 ORE)
COLLOCAZIONE NEL PIANO DI STUDI	PRIMO ANNO (LT – TUTTI I CORSI)
PERIODO DI SVOLGIMENTO	SECONDO SEMESTRE
DOCENTE	PROF. DIEGO ZAPPA

OBIETTIVO DEL CORSO E RISULTATI DI APPRENDIMENTO ATTESI

Trattasi di un corso a contenuto prevalentemente metodologico, in cui vengono presentati i principali argomenti relativi alle macro-aree (a) statistica descrittiva, (b) calcolo delle probabilità e campionamento.
I principali macro-argomenti saranno: indici sintetici per variabili unidimensionali, introduzione allo studio di fenomeni bivariati, schemi di campionamento e costruzione di variabili casuali campionarie. Il corso prevede lo svolgimento di lezioni a contenuto sia metodologico sia applicato, senza soluzione di continuità.
Al termine del corso i partecipanti sapranno come si effettua la sintesi informativa di una variabile statistica (o di più variabili statistiche), comprendere i primi rudimenti relativi allo studio di fenomeni aleatori.

PROGRAMMA DEL CORSO

OBIETTIVI FORMATIVI CHE LO STUDENTE POTREBBE ACQUISIRE NEL CORSO

I parte: statistica descrittiva

Dopo avere descritto il contenuto delle scienze statistiche, verrà introdotta la tassonomia dei caratteri statistici, di cui si proporranno le principali rappresentazioni tabellari e grafiche.
Seguirà la presentazione dei principali indici per descrivere variabili statistiche unidimensionali e bidimensionali. In particolare verranno presentati i seguenti argomenti:

- Indici di posizione.
- Indici di dispersione e indici di variabilità globale.
- Indici di forma.

Seguirà l'introduzione all'analisi statistica di fenomeni bidimensionali, di cui si commenteranno i concetti di dipendenza e di indipendenza in presenza di:

- simmetria informativa (indici di connessione e concetto di indipendenza stocastica);
- asimmetria informativa (introduzione ai modelli di regressione lineari nei parametri, impiego del metodo dei minimi quadrati e misure di adattamento ai dati).

Dopo il completamento della trattazione dei precedenti argomenti la studentessa/lo studente è in grado di effettuare le principali analisi descrittive di dataset e di produrre semplici report interpretativi del fenomeno oggetto di analisi.

II parte: calcolo delle probabilità e campionamento

Si inizierà con una introduzione al calcolo delle probabilità. In particolare verranno presentati i seguenti argomenti:

- Definizione assiomatica, frequentista e classica di probabilità.
- Lo schema dell'urna.
- Definizione di variabile casuale.
- Principali variabili casuali: uniforme (discreta e continua), di Bernoulli, binomiale, normale (o gaussiana).
- La legge dei grandi numeri e teorema del limite centrale (solo enunciati).

Si procederà quindi con l'introduzione di elementi utili per affrontare argomenti di inferenza statistica, tra cui:

– principali tecniche di campionamento;
– definizione di verosimiglianza;
– distribuzione di alcune variabili casuali campionarie (in particolare della media campionaria).

Dopo il completamento della trattazione dei precedenti argomenti, la studentessa/lo studente è in grado di affrontare le prime nozioni relative alle procedure decisionali in condizioni di incertezza.

BIBLIOGRAFIA

Dispense a cura del docente
G. CICCHITELLI, *Statistica – Principi e metodi,* Pearson, 2018.

DIDATTICA DEL CORSO

Si alterneranno lezioni metodologiche a lezioni con prevalente contenuto esemplificativo.

METODO E CRITERI DI VALUTAZIONE

Prova scritta suddivisa in due parti. Ai fini della valutazione finale, le due parti hanno uguale peso (50%).

Le prove comprendono esercizi sia applicati che teorici. È prevista una sezione obbligatoria che, se non superata, rende la prova non sufficiente. Il superamento di questa sezione, senza aver svolto alcun esercizio della sezione istituzionale, non rende la prova sufficiente.

LUOGO e ORARIO RICEVIMENTO

Orario e luogo di ricevimento
Il Prof. Diego Zappa riceve gli studenti come da avviso pubblicato sul sito www.unicatt.it presso il Dipartimento di Scienze statistiche (edificio di via Lanzone, 18, III piano).
(controllare pagina web http://docenti.unicatt.it/ita/Diego_Zappa)

Premessa

Il materiale contenuto in queste pagine è stato composto con l'intento di aiutare gli studenti nel seguire le lezioni, evitando, per quanto possibile, di impegnarli nel "ricopiare formule".
Gli argomenti meno complessi sono presenti in forma estremamente sintetica, mentre concetti/argomenti più complessi sono corredati anche da brevi commenti.

Si suggerisce di utilizzare tale materiale affiancando la **lettura sistematica** del testo di riferimento e **svolgendo/seguendo** gli esercizi proposti in aula.

Avvertenza: il materiale potrà subire aggiornamenti durante il corso delle lezioni

Statistica

Metodo della ricerca empirica

Terminologia/notazione

Popolazione: il rilevabile (di dimensione ***N***)
Campione: il rilevato (di dimensione ***n***)
Unità statistica: $h = 1,...,\boldsymbol{N}$ $(h = 1,...,\boldsymbol{n})$

Dell'unità statistica, *h-esima*, posso misurare uno o più caratteri (indicati esempio con *X*).

Caratteri :

- qualitativi
  - sconnessi (nominali)
  - rettilinei (ordinali)
- quantitativi
  - discreti
  - continui

–o–o–o–o–o–o–o–o–

Un carattere (*X*) può assumere la modalità x_i, per $i=1,..,k$

Tramite il c.d. *spoglio* dei dati si ricava n_i :

- *frequenza assoluta* con cui la modalità x_i si è manifestata.
- *frequenza assoluta* con cui si sono osservate modalità all'*interno* della *classe* $x_{i-1} -| x_i$ (intervallo di valori)
  - –| classe chiusa a dx e aperta a sx
  - |– classe chiusa a sx e aperta a dx
  - |–| classe chiusa a sx e a dx

–o–o–o–o–o–o–o–o–

La successione $\{x_i, n_i\}$ per $i=1,..,k$ si definisce "*serie*"
(Se *X* è raggruppata in classi si parla di *seriazione*)

Con riferimento ad una popolazione composta da *N* individui classificati in base a *k* modalità, si ha

$$N = \sum_{i=1}^{k} n_i$$

e si definisce

$$f_i = \frac{n_i}{N}$$

la *frequenza relativa* della modalità x_i. Ovviamente deve valere

$$\sum_{i=1}^{k} \frac{n_i}{N} = \sum_{i=1}^{k} f_i = 1$$

—o—o—o—o—o—o—o—o—

Esempio (tema 21/06/2012)

La seguente tabella riporta informazioni su D = dividendo per azione (in percentuale) e A= tipologia del titolo, riferito a 10 titoli azionari (n.b. dati simulati).

D	2	0	1	1	4	2	1	0	2	4
A	E	F	S	I	E	S	I	F	F	E

Si effettui lo spoglio dei dati, costruendo la corrispondente tabella a doppia entrata.

R. Effettuando lo spoglio di A si ottiene

dove - E,F,I,S sono le k=4 modalità
- 3,3,2,2 sono frequenze assolute
- 0.3, 0.3, 0.2, 0.2 sono frequenze relative

i	A	n_i	f_i
1	E	3	0.3
2	F	3	0.3
3	I	2	0.2
4	S	2	0.2
		N=10	**F=1**

—o—o—o—o—o—o—o—o—

Con riferimento ad un carattere *ordinabile* (e ordinato in senso crescente) si definisce

Funzione di ripartizione il luogo dei punti (x_j , F_j), dove

$$F_j = \sum_{i=1}^{j} \frac{n_i}{N} = \sum_{x_i \le x_j} f_i \qquad per\ j = 1, \dots, k$$

con le seguenti proprietà

1- F_j è definita su tutto $\mathbb{R}$

2- per $x < x_{\min}$, $F_j = 0$; per $x > x_{\max}$, $F_j = 1$

3- per $x_j < x_{j+1}$, $F_j \le F_{j+1}$

4- per $x_j \le x_{j*} < x_{j+1}$, $F_j = F_{j*}$

La funzione di retro-ripartizione è

$$\bar{F}_j \equiv G_j = \sum_{i=j}^{k} \frac{n_i}{N} = \sum_{x_i \ge x_j} f_i \qquad per\ j = 1, \dots, k$$

—o—o—o—o—o—o—o—o—

Esempio di funzione di ripartizione e di retroripartizione

(vedi prova 9/4/2010)

X	n_i	f_i	F_j	G_j
648	3	0.3	0.3	1
649	1	0.1	0.4	0.7
650	3	0.3	0.7	0.6
651	2	0.2	0.9	0.3
652	1	0.1	1	0.1
	N=10	**1**		

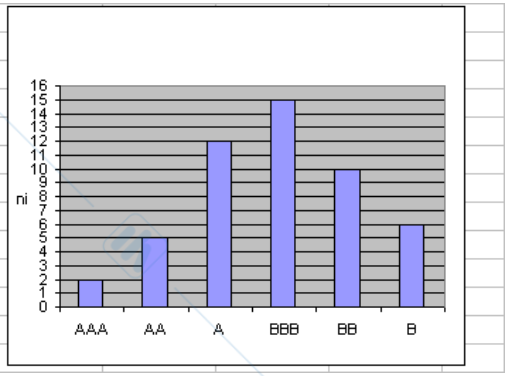
—o—o—o—o—o—o—o—o—

Rappresentazioni Grafiche

Caratteri qualitativi

- Barre
  - in ordinata: frequenze assolute o relative

Rating	ni
AAA	2
AA	5
A	12
BBB	15
BB	10
B	6

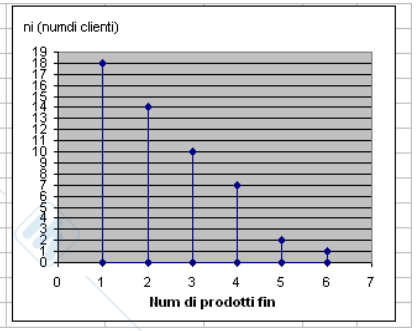


- Torte
- ...

Caratteri quantitativi

- Discreti
  - Istogrammi
    - in ordinata: frequenze assolute o relative

Num di prodotti fin	ni (num di clienti)
1	18
2	14
3	10
4	7
5	2
6	1

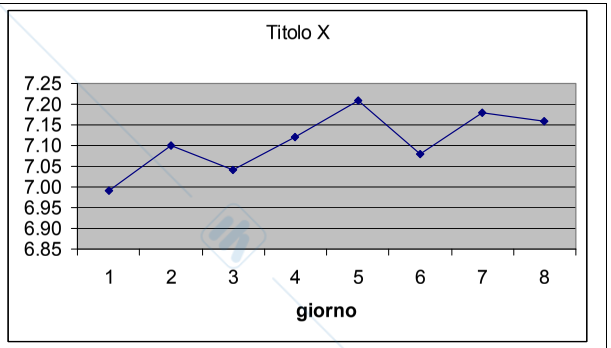


  - Torte
  - ...

- Seriazioni (Serie storica)
  - in ascissa: tempo, *t*.
  - in ordinata: valore della variabile al tempo *t*.

Dati

Giorno	**Prezzo chiusura Titolo X**
1	6.99
2	7.10
3	7.04
4	7.12
5	7.21
6	7.08
7	7.18
8	7.16



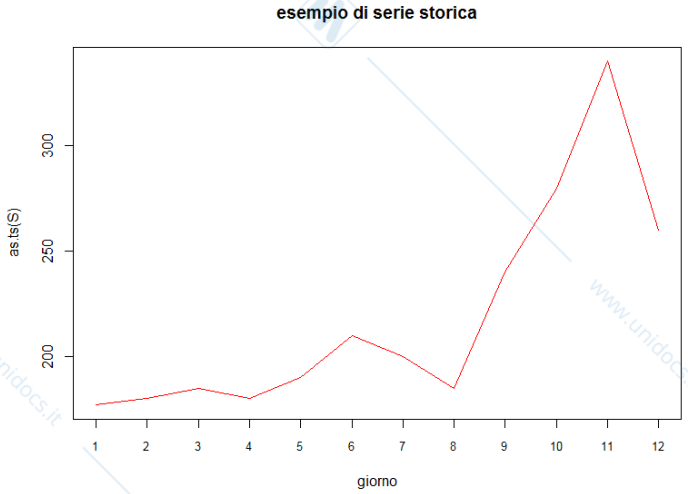
—o—o—o—o—o—o—o—o—o—

Esempio serie storica (15/07/2011)

La seguente tabella riporta le serie di dati raccolti nell'arco di 12 giorni su: *S*=Spread medio giornaliero di rendimento BTP-Bund, *D*=presenza/assenza in giornata di decisioni sul rating del debito pubblico di qualche paese europeo.

Giorno	1	2	3	4	5	6	7	8	9	10	11	12
S	177	180	185	180	190	210	200	185	240	280	340	260

Si raccolgano i dati della variabile *S* nelle classi 170–|200, 200–|250, 250–|350 e si chiami la nuova serie *S**. Rappresentare, con opportuni grafici, le serie *S* e *S**.

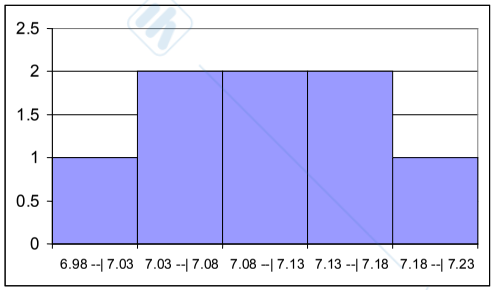


—o—o—o—o—o—o—o—o—o—

- Continui
  - Istogrammi (a barre contigue)
    - in ordinata: densità di frequenza (frequenze ridotte) $d_i = n_i / w_i$
      dove w_i=ampiezza della classe (ovvero $x_i - x_{i-1}$)

1.1. *Raggruppamento in classi di ampiezza costante 0.05*

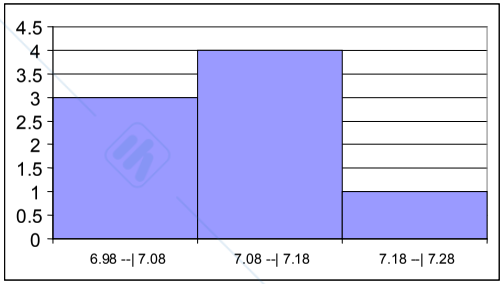
Classi: x_{i-1} —| x_i	n_i
6.98 —| 7.03	1
7.03 —| 7.08	2
7.08 —| 7.13	2
7.13 —| 7.18	2
7.18 —| 7.23	1



(n.b. si usa anche la notazione x_{i-1} —\| $x_i \equiv (x_{i-1}, x_i]$

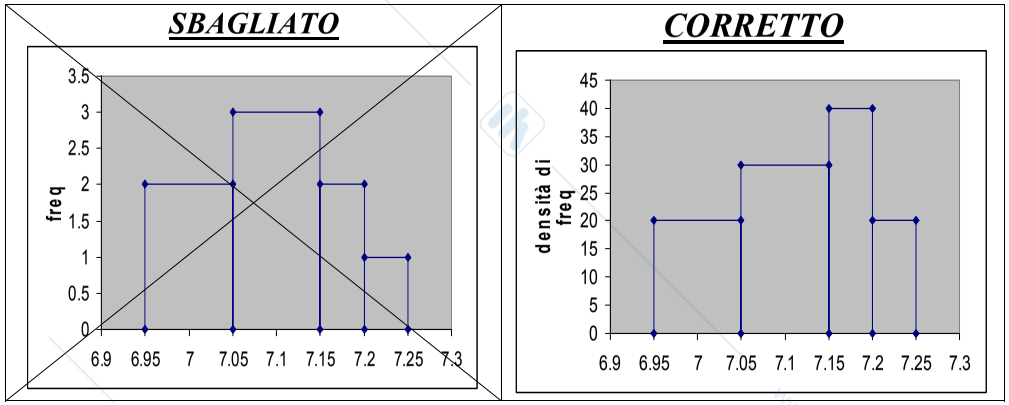
1.2. *Raggruppamento in classi di ampiezza costante 0.10*

Classi: x_{i-1} —| x_i	n_i
6.98 —| 7.08	3
7.08 —| 7.18	4
7.18 —| 7.28	1



1.3. *Raggruppamento in classi di ampiezza non costante*

Classi: x_{i-1} —| x_i	n_i	W_i	$d_i = \frac{n_i}{w_i}$
6.95 —| 7.05	2	0.1	20
7.05 —| 7.15	3	0.1	30
7.15 —| 7.20	2	0.05	40
7.20 —| 7.25	1	0.05	20



-o-o-o-o-o-o-o-o-o-

Esempio di istogramma con classi di ampiezza non costante (prova 15/07/2011)

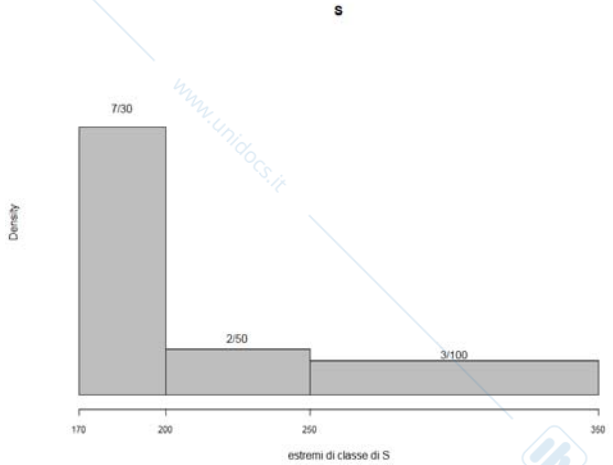
Giorno	1	2	3	4	5	6	7	8	9	10	11	12
S	177	180	185	180	190	210	200	185	240	280	340	260

Si raccolgano i dati della variabile *S* nelle classi 170–|200, 200–|250, 250–|350 e si chiami la nuova serie *S**. Rappresentare, con opportuni grafici, *S**.

R. Una riclassificazione possibile è

S	Classi_S1	Classi_S2
177	170-|200	(170,200]
180	170-|200	(170,200]
185	170-|200	(170,200]
180	170-|200	(170,200]
190	170-|200	(170,200]
210	200-|250	(200,250]
200	170-|200	(170,200]
185	170-|200	(170,200]
240	200-|250	(200,250]
280	250-|350	(250,350]
340	200-|250	(250,350]
260	250-|350	(250,350]

Classi_S1	n_i	w_i	d_i
170-|200	7	30	7/30
200-|250	2	50	2/50
250-|350	3	100	3/100

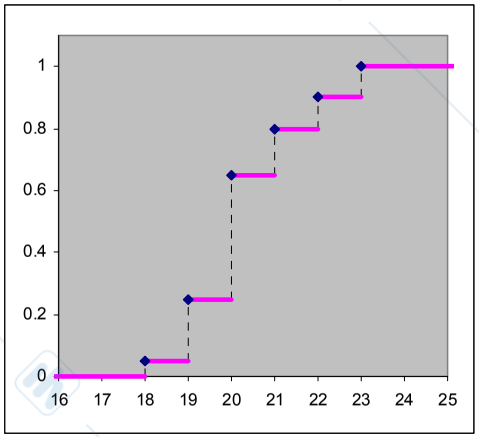


-o-o-o-o-o-o-o-o-o-

Rappresentazioni Grafiche: Funzione di ripartizione

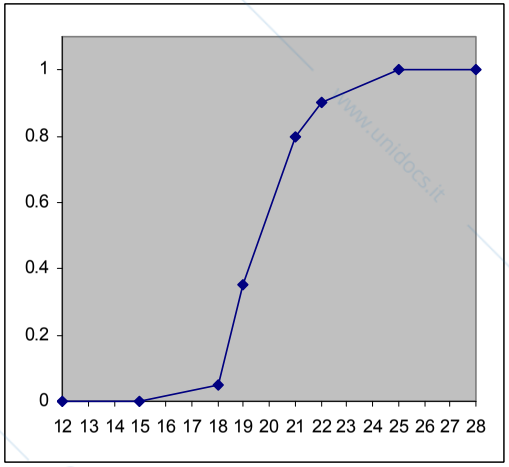
Caratteri numerabili

Età in anni compiuti	Freq rel	Fi
18	0.05	0.05
19	0.20	0.25
20	0.40	0.65
21	0.15	0.80
22	0.10	0.90
23	0.10	1.00



Caratteri non numerabili

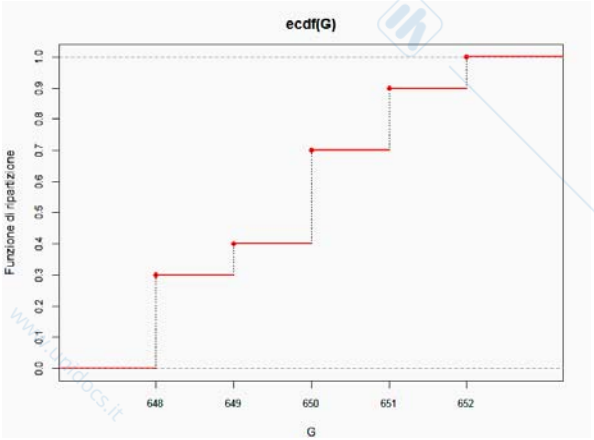
Età (IN CLASSI)	Freq rel	Fi
15-|18	0.05	0.05
18-|19	0.3	0.35
19-|21	0.45	0.8
21-|22	0.1	0.9
22-|25	0.1	1



—o—o—o—o—o—o—o—o—

Esempio di funzione di ripartizione e di retroripartizione
(vedi prova 9/4/2010)

X	n_i	f_i	F_j	G_j
648	3	0.3	0.3	1
649	1	0.1	0.4	0.7
650	3	0.3	0.7	0.6
651	2	0.2	0.9	0.3
652	1	0.1	1	0.1
	N=10	**1**		



—o—o—o—o—o—o—o—o—

Indici di posizione (Medie) $\alpha(X)$

Sia

$$\alpha(\boldsymbol{X}) : \boldsymbol{X} \rightarrow \mathbb{R}^1$$

$\alpha(\boldsymbol{X})$ si chiama indice di posizione se valgono le seguenti proprietà **:**

1. **Internalità :**

$$x_{(1)} = \min(\boldsymbol{X}) \leq \alpha(\boldsymbol{X}) \leq \max(\boldsymbol{X}) = x_{(n)}$$

2. **Moltiplicativa (proporzionalità)**

$$\alpha(c \cdot \boldsymbol{X}) = c \cdot \alpha(\boldsymbol{X})$$

3. **Monotonicità:** date due v.s. X e Y tali che

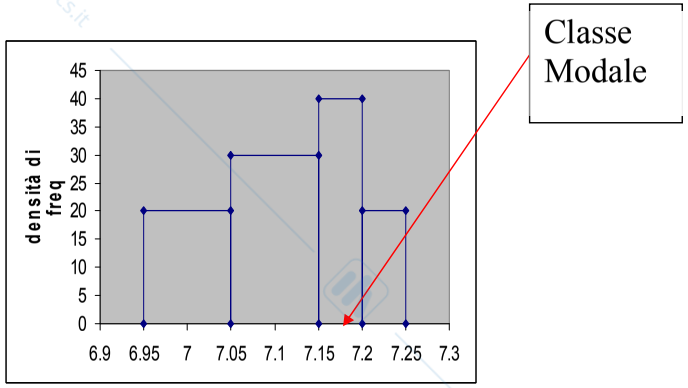
$$\boldsymbol{X} \preccurlyeq \boldsymbol{Y}$$

allora deve valere

$$\alpha(\boldsymbol{X}) \leq \alpha(\boldsymbol{Y})$$

Moda : Mo(X)

"modalità a cui è associata la maggiore frequenza o la classe a cui è associata la maggiore densità di frequenza"



Percentile di ordine p : x_p (o $x_{(p)}$) con 0<p<1

"modalità o classe che occupa la posizione *p*-esima di una distribuzione ordinata di valori"

A)

$$x_p: \left(\sum_{x \le x_p} f_i \ge p\right) \wedge \left(\sum_{x \ge x_p} f_i \ge (1-p)\right)$$

B) sia $F(X)$ la *funzione* di ripartizione. Allora

$$x_p : F(x_p) = p$$

- Se F è una funzione invertibile allora, $F^{-1}(p) = x_p$
- Se non esiste inversa, la soluzione potrebbe non essere unica

Es. p=0.25 ($x_{0.25}$=I quartile)

p=0.5 ($x_{0.5}$=Mediana / II quartile)

p=0.75 ($x_{0.75}$=III quartile)

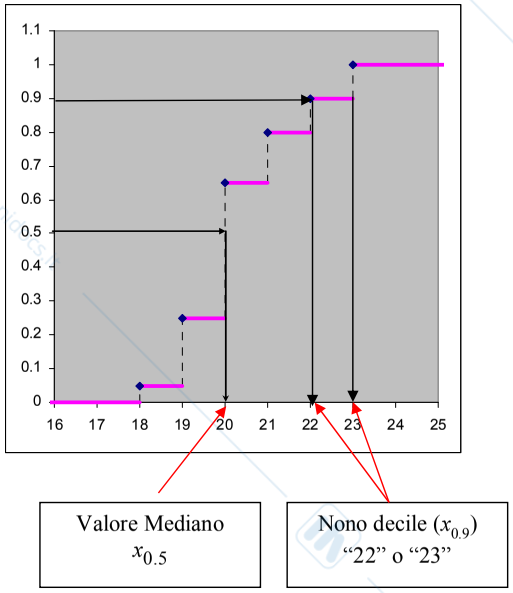
p=0.1 · c ($x_{0.1 \cdot c}$=decile c-esimo) con c=1,2,...,9

Caratteri numerabili

Con l'impostazione A

	Età in anni compiuti	Freq rel	Fi	Retro-ripartizione
	18	0.05	0.05	1
I quartile	*19*	0.2	0.25	0.95
I quartile / Mediana	***20***	0.4	0.65	0.75
III quartile	*21*	0.15	0.8	0.35
nono decile	22	0.1	0.9	0.2
nono decile	23	0.1	1	0.1

Con l'impostazione B

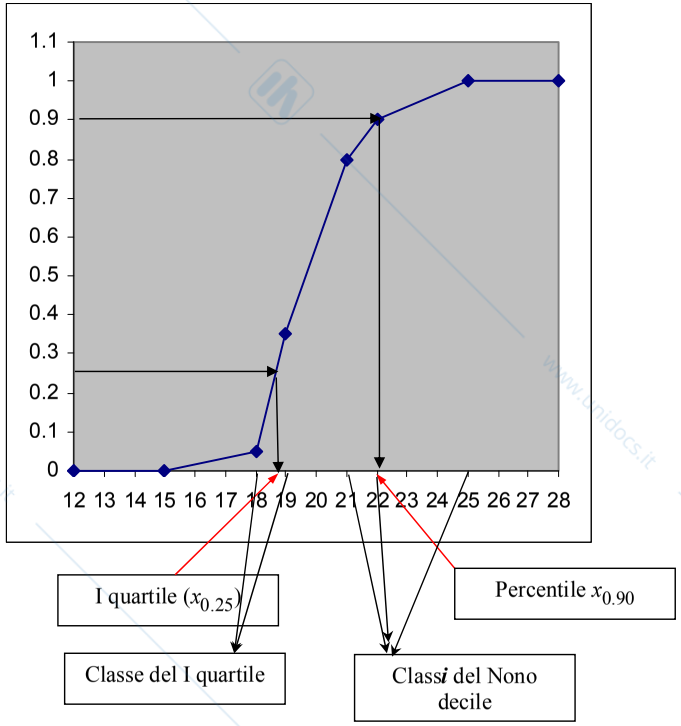


Caratteri non numerabili

Con l'impostazione A

	Età (IN CLASSI)	Freq rel	Fi	Retro-ripartizione
	15-|18	0.05	0.05	1
I quartile	*18-|19*	0.3	0.35	0.95
Mediana / III quartile	*19-|21*	0.45	0.8	0.65
nono decile	21-|22	0.1	0.9	0.2
nono decile	22-|25	0.1	1	0.1

Con l'impostazione B



–o–o–o–o–o–o–o–o–o–

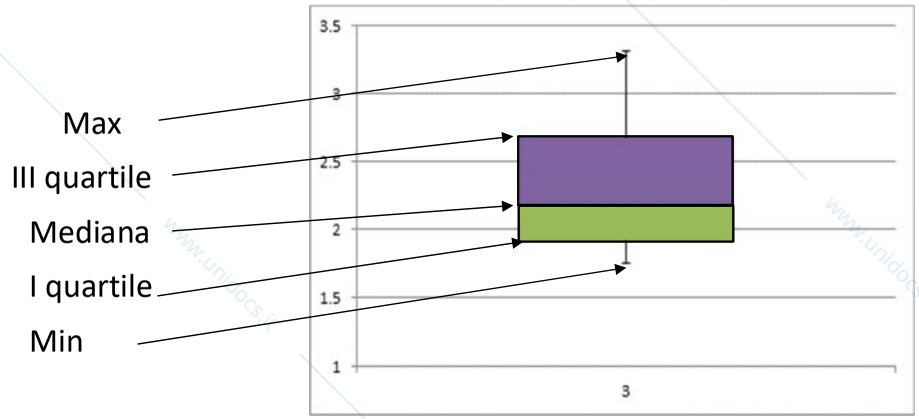
(v. file FdR rip.xls)

Il Box-Plot (o Box and Wiskers Plot)

Sia :

$x_{(1)}$= min
$x_{0.25}$=I quartile
$x_{0.5}$=Mediana
$x_{0.75}$=III quartile
$x_{(n)}$= max

Data una serie di valori, il seguente grafico prende il nome di Box-Plot



Posto $x_{0.75} - x_{0.25}$=RI := *Range Interquatilico*, se

$x_{(1)} > x_{0.25} - \text{RI} \cdot 1.5$ **allora** Min = $x_{(1)}$ **altrimenti** Min = $x_{0.25} - \text{RI} \cdot 1.5$
$x_{(n)} < x_{0.75} + \text{RI} \cdot 1.5$ **allora** Max = $x_{(n)}$ **altrimenti** Max = $x_{0.75} + \text{RI} \cdot 1.5$

Tutti i valori

$$\{x \quad : \quad x < x_{0.25} - \text{RI} \cdot 1.5 \quad \cup \quad x > x_{0.75} + \text{RI} \cdot 1.5 \}$$

vengono chiamati "outlier".

Esempio

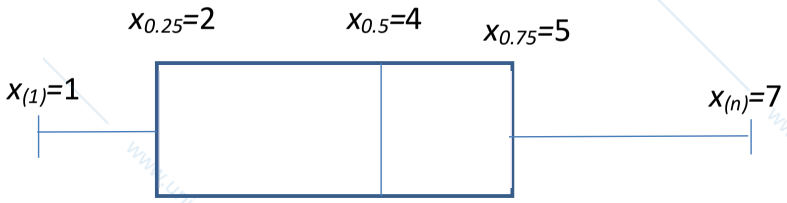
Sia
X={1,2,2,3,4,5,5,5,7} ovvero

x_i	n_i	F_j	G_j
1	1	1/9	1
2	2	3/9	8/9
3	1	4/9	6/9
4	1	5/9	5/9
5	3	8/9	4/9
7	1	1	1/9

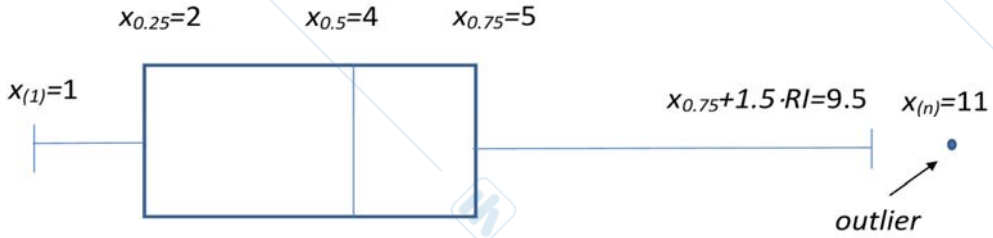
Da cui

$x_{(1)}$= 1
$x_{0.25}$=2
$x_{0.5}$=4
$x_{0.75}$=5
$x_{(n)}$= 7
RI= 5-2=3
RI·1.5= 4.5

Box-Plot

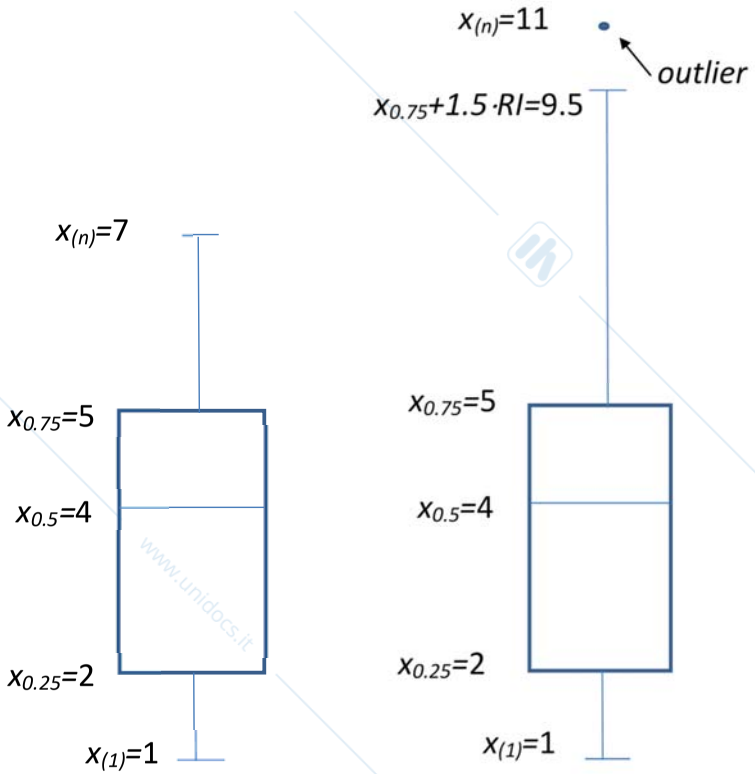


Se avessimo avuto $x_{(n)}$= 11 allora



(v. file Esempi)

(rappresentazione “verticale”)



-o-o-o-o-o-o-o-o-o-

Esempio di boxplot (21/06/2012)

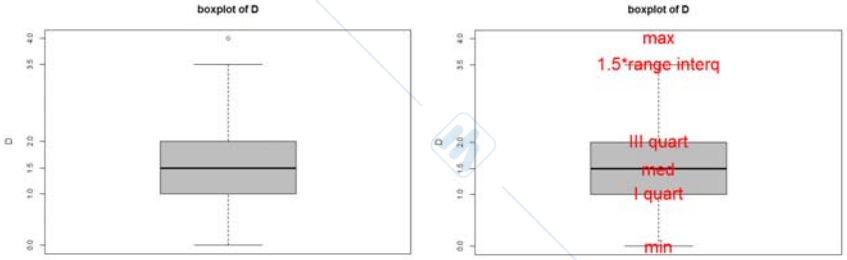
D	2	0	1	1	3	2	1	0	2	4
A	E	F	S	I	E	S	I	F	F	E

1.1) Si effettui lo spoglio dei dati, costruendo la corrispondente tabella a doppia entrata. Si rappresenti con un grafico opportuno la distribuzione condizionata *A|D=2* e *si rappresenti la marginale D con un box-plot*.

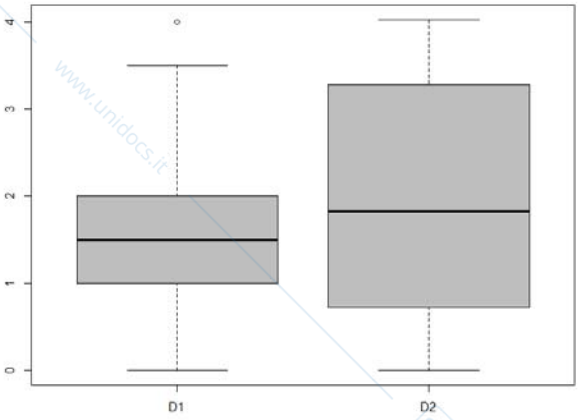
```
> quantile(D)
  0%  25%  50%  75% 100%
 0.0  1.0  1.5  2.0  4.0

> boxplot(D, range=1.5, border=1, col="grey", ylab="D", yaxt="n")
> axis(2, c(quantile(D), 3.5), cex.axis=.8)

> text(x=1,c(quantile(D), 3.5), c("min","I quart","med","III
                    quart","max","1.5*range interq"),col="red",
cex=2)
```



(Un confronto tra due distribuzioni)



-o-o-o-o-o-o-o-o-o-

Medie analitiche

Si consideri X o una sua trasformazione $g(X)$

Per variabili X non negative o trasformazioni $g(X)$ su supporto $\mathbb{R}^+$, si definisce media potenziata di ordine r di X (o di $g(X)$)

$$M_r(X) = \sqrt[r]{\frac{1}{N}\sum_{i=1}^{k} x_i^r n_i} = \sqrt[r]{\sum_{i=1}^{k} x_i^r f_i} \equiv \mu_r$$

$$M_r\big(g(X)\big) = \sqrt[r]{\frac{1}{N}\sum_{i=1}^{k} g(x_i)^r\, n_i} = \sqrt[r]{\sum_{i=1}^{k} g(x_i)^r\, f_i} \equiv \mu_r[g(X)]$$

Proprietà di $M_r(X)$

- $M_r(X)$ è un indice di posizione con $r \in \mathbb{R}$
- Per r=0 il limite

$$\lim_{r\to 0^+} M_r(X) = \lim_{r\to 0^-} M_r(X) = M_0(X) = \sqrt[N]{\prod_{i=1}^{k} x_i^{n_i}}$$

  Prende il nome di media geometrica. In particolare vale

$$\log M_0(X) = \frac{1}{N}\sum_{i=1}^{k} n_i \log x_i$$

- $\lim_{r\to-\infty} M_r(X) = x_{(1)} \quad \lim_{r\to+\infty} M_r(X) = x_{(n)}$
- Per $r_1 \le r_2$ vale $M_{r_1}(X) \le M_{r_2}(X)$

N.B.

1- $[M_r(X)]^r = \frac{1}{N}\sum_{i=1}^{k} x_i^r n_i$ prende il nome di momento dall'origine di ordine *r* che è definito anche per variabili *X* non positive o sue trasformazioni *g*(*X*).

2- $M_r(X)$ per *r*=1 si chiama media (aritmetica) di *X*

$$M(X) = \frac{1}{N}\sum_{i=1}^{k} x_i n_i \equiv \bar{x} \equiv \mu$$

$$M\big(g(X)\big) = \frac{1}{N}\sum_{i=1}^{k} g(x_i) n_i \equiv \overline{g(X)} \equiv \mu[g(X)]$$

3- se *X* è in classi, per la classe x_{i-1}-|x_i si considera, per convenzione, il solo valore centrale $\frac{x_i+x_{i-1}}{2} := {}_c x_i$

-0-0-0-0-0-0-

Esempio su medie (13/02/2015)

La tabella seguente riporta quotazioni dei bond decennali russi (B) e del prezzo del petrolio (P) , rilevati in cinque giorni di contrattazioni. (n.b. dati fittizi):

giorno	*Lun*	*Mart*	*Merc*	*Giov*	*Ven*
B	142	140	139	140	142
P	50	49	49	51	53

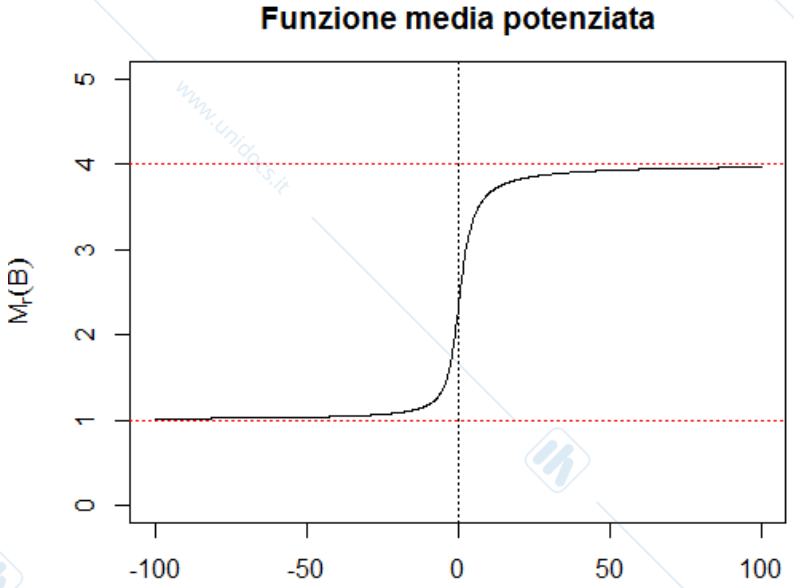
1.2a) Calcolare, se possibile, media, deviazione standard, asimmetria e momento secondo dall'origine della marginale B

M(B) = 140.6

$M(B^2)$ = 19769.8

Funzione media potenziata per B* = B-min(B)+1 ovvero

B*	4	2	1	2	4



-0-0-0-0-0-0-

Proprietà di $M(X)$

A) M(X) è un <u>operatore lineare</u> (ovvero la media di una combinazione lineare è la combinazione lineare delle medie).

Es. Se Y=a+bX allora M(Y) = a+b·M(X)

$$M(Y) = \sum_{i=1}^{k}(a + b \cdot x_i)f_i =$$

$$= \sum_{i=1}^{k}(a)f_i + \sum_{i=1}^{k}(b \cdot x_i)f_i = a + b \cdot M(X)$$

Se Y=a + bX +cZ allora M(Y) = a+b·M(X)+c·M(Z)

B) Qual è la soluzione a θ: $\sum_{i=1}^{k}(g(x_i) - \theta)f_i = 0$ ovvero θ: $M(g(X) - \theta)$=0 ?

La soluzione è per $\theta = \overline{g(X)}$. Quindi se $g(X) = X$ <u>allora</u> M(X – μ) = 0 !!!

C) Qual è la soluzione a

$$\min_{\theta} \sum_{i=1}^{k}(g(x_i) - \theta)^2 f_i = \min_{\theta} M[(g(X) - \theta)^2]$$

si ha per $\theta = \overline{g(X)}$

Il gradiente è $-2\sum_{i=1}^{k}(g(x_i) - \theta)f_i = 0$ da cui $\theta = \overline{g(X)}$
L'hessiano è $+2\sum_{i=1}^{k} f_i > 0$.
Per cui $\theta = \overline{g(X)}$ è soluzione di minimo.
Quindi se $g(X) = X$ <u>allora</u> $\min_{\theta} M[(X - \theta)^2]$ si ha per θ =μ.

<u>Esempio su proprietà di minimo della media aritmetica</u>
Si consideri la tabella seguente

X	f_i
x_1	f_1
x_2	f_2

con $f_1 < f_2$ e $x_1 < x_2$. Dimostrare che

$$M\left[\left(X - Me(X)\right)^2\right] \geq M\left[\left(X - M(X)\right)^2\right]$$

Si osservi che dovendo essere $f_2 = 1 - f_1 > 0.5$ si ha $Me(X) = x_2$.
Da cui

$$M\left[\left(X - Me(X)\right)^2\right] = \sum_{i=1}^{2}(x_i - x_2)^2 f_i = (x_1 - x_2)^2 f_1 =$$
$$= (x_1^2 f_1 + x_2^2 f_1 - 2x_1 x_2 f_1)$$

$$M\left[\left(X - M(X)\right)^2\right] = \sum_{i=1}^{2}\left(x_i - (x_1 f_1 + x_2 f_2)\right)^2 f_i =$$
$$= \left(x_1^2 f_1 + x_2^2(1 - f_1)\right) - \left(x_1 f_1 + x_2(1 - f_1)\right)^2$$
$$= \left(x_1^2 f_1 + x_2^2(1 - f_1)\right) - \left(x_1^2 f_1^2 + x_2^2(1 - f_1)^2 + 2x_1 x_2 f_1(1 - f_1)\right)$$

ovvero si tratta di dimostrare che

$$(x_1^2 f_1 + x_2^2 f_1 - 2x_1 x_2 f_1)$$
$$\geq \left(x_1^2 f_1 + x_2^2(1 - f_1)\right)$$
$$- \left(x_1^2 f_1^2 + x_2^2(1 - f_1)^2 + 2x_1 x_2 f_1(1 - f_1)\right)$$

$$(x_2^2 f_1 - 2x_1 x_2 f_1)$$
$$\geq \left(x_2^2(1 - f_1)\right)$$
$$- \left(x_1^2 f_1^2 + x_2^2 + x_2^2 f_1^2 - 2x_2^2 f_1 + 2x_1 x_2 f_1(1 - f_1)\right)$$

$$(x_2^2 f_1) \geq \left(x_2^2(-f_1)\right) - \left(x_1^2 f_1^2 + x_2^2 f_1^2 - 2x_2^2 f_1 + 2x_1 x_2 f_1(-f_1)\right)$$

$$2x_2^2 f_1 \geq -(x_1^2 f_1^2 + x_2^2 f_1^2 - 2x_2^2 f_1 - 2x_1 x_2 f_1^2)$$

$$0 \geq -(x_1^2 f_1^2 + x_2^2 f_1^2 - 2x_1 x_2 f_1^2) = -\left(x_1 f_1 - x_2 f_1\right)^2 \quad VERO\ !$$

Indici di dispersione

Sia $g(X,\theta)= |X-\theta|$ la funzione *scarto* dall'indice di posizione θ. Si ponga

$$\mu_r[g(X,\theta)] = \sqrt[r]{\sum_{i=1}^{k} g(x_i,\theta)^r f_i} \equiv R_r(\theta) \quad \text{per } r > 0$$

ovvero

$$R_r(\theta) = \sqrt[r]{\frac{1}{N}\sum_{i=1}^{k} |x_i-\theta|^r n_i} = \sqrt[r]{\sum_{i=1}^{k} |x_i-\theta|^r f_i} \quad \text{per } r > 0$$

è la media potenziata di $g(X,\theta)$ di ordine $r > 0$ ed è detto *indice di dispersione dal centro* θ di ordine r.

Proprietà di $R_r(\theta)$

- $R_r(\theta) \geq 0 \quad \forall r$
- $\lim_{r\to 0} R_r(\theta) = \min |x_i - \theta|$

  $\lim_{r\to +\infty} R_r(\theta) = \max|x_i - \theta|$
- Per $r_1 \leq r_2$ vale $R_{r_1}(\theta) \leq R_{r_2}(\theta)$

N.B.

- Se $r = 2$ e $\theta=\mu$ si ha

  $R_2(\mu) \equiv \sigma \Rightarrow$ ***scarto quadratico medio***

  $[R_2(\mu)]^2 \equiv \sigma^2 \Rightarrow$ ***varianza***

$$\sigma^2 = M[(X-\mu)^2] = \sum_{i=1}^{k}(x_i-\mu)^2 f_i$$

$$= \sum_{i=1}^{k} x_i^2 f_i + \sum_{i=1}^{k}(\mu)^2 f_i - 2\sum_{i=1}^{k}(x_i \cdot \mu) f_i$$

$$= M[X^2] + (\mu)^2 - 2\mu \sum_{i=1}^{k} x_i f_i$$

$$= M[X^2] + \mu^2 - 2\mu^2 = M[X^2] - \mu^2$$

$$:= Var(X)$$

- $[R_r(\mu_X)]^r = M[|X-\mu|^r]$ prende il nome di momento centrale di ordine r

Proprietà di $Var(X)$

- $Var(X) \geq 0$

- **NON** è un operatore lineare

Se $\boldsymbol{Y = a + bX}$ allora

$$Var(Y) = M[(Y-\mu_Y)^2] = M\left[\left(a+bX-(a+b\mu_X)\right)^2\right]$$
$$= M[(bX-b\mu_X)^2] = M[b^2(X-\mu_X)^2]$$
$$= b^2 M[(X-\mu_X)^2] = b^2 Var(X)$$

Se $\boldsymbol{Y = a + bX + cZ}$ allora

$$Var(Y) = M[(Y-\mu_Y)^2]$$

$$= M\left[\left(a+bX+cZ-(a+b\mu_X+c\mu_Z)\right)^2\right]$$

$$= M\left[\left(bX+cZ-(b\mu_X+c\mu_Z)\right)^2\right]$$

$$= M[(bX-b\mu_X)^2] + M[(cZ-c\mu_Z)^2] + 2M[(bX-b\mu_X)(cZ-c\mu_Z)]$$

$$= b^2 Var(X) + c^2 Var(Z) + 2bc\, M[(X-\mu_X)(Z-\mu_Z)]$$

$$= b^2 Var(X) + c^2 Var(Z) + 2bc\, Cov(X,Z)$$

Esempio su proprietà dello scarto quadratico medio di una combinazione lineare (*per una piena dimostrazione si vedano gli argomenti sulla regressione*)

Si dimostri che

$$\sqrt{Var(bX+cY)} \le \sqrt{Var(bX)} + \sqrt{Var(cY)}$$
$$Var(bX+cY) \le Var(bX) + Var(cY) + 2bc\sqrt{Var(X)}\sqrt{Var(Y)}$$
$$Cov(X,Y) \le \sqrt{Var(X)}\sqrt{Var(Y)}$$

Poiché

$$\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \rho$$

dove ρ (vedi argomenti su regressione) è il coeff di correlazione lineare che è sempre ≤ 1, si ha che la diseguaglianza è vera!

N.B.
- la varianza di una combinazione lineare è maggiore della combinazione lineare delle varianze
- lo scarto quadratico medio di una combinazione lineare è minore della combinazione lineare degli scarti quadratici medi

Il coefficiente di variazione

Con riferimento ad una variabile definita su $\mathbb{R}^+$, il rapporto

cv = $\frac{\sigma}{\mu}$ => prende il nome di coefficiente di variazione

È utile per:

- confrontare fenomeni aventi unità misura differente (es. la volatilità di un portafoglio in euro vs la volatilità di un paniere in dollari)
- verificare se al variare della media (condizionata) rimane costante il rapporto con lo scarto quadratico medio (condizionato).

In particolare si definisce:
"omoschedastico": un fenomeno che al variare della media condizionata non cambia in varianza (tipico di molti fenomeni sperimentali)

"eteroschedastico": un fenomeno che al variare della media condizionata cambia anche in varianza (tipico dei fenomeni finanziari e di molti fenomeni assicurativi)

Esempio: tasso medio e dispersione della sinistrosità al variare della classe Bonus Malus di un portafoglio clienti di tre compagnie di assicurazione (X,Y,W).

Classe BM	$\mu(X)$	$\sigma(X)$		$\mu(Y)$	$\sigma(Y)$		$\mu(W)$	$\sigma(W)$		cv(X)	cv(Y)	cv(W)
1-4	0.01	0.02		0.01	0.05		0.01	0.4		2	5	4
5-8	0.25	0.5		0.25	0.5		0.25	0.4		2	2	0.625
8-11	0.9	1.8		0.9	0.6		0.9	0.4		2	0.667	0.44
11-15	1.1	2.2		1.1	0.8		1.1	0.4		2	1.38	0.36

La Tabella "a doppia entrata"

(*introduzione allo studio di fenomeni bivariati*)

X \ Y	y_1	y_2	...	y_j	...	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	...	...	...	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	...	...	...	n_{2k}	$n_{2.}$
...	...	...	...	...	...	...	...
x_i	...	...	...	n_{ij}	...	...	$n_{i.}$
...	...	...	...	...	...	...	...
x_h	n_{h1}	n_{h2}	...	...	...	n_{hk}	$n_{h.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.k}$	$N=\Sigma n_{ij}$

Le variabili condizionate sono

$X|Y=y_j$ (ovvero avendo fissato j) con pesi n_{ij} per j=1,2,...k
$Y|X=x_i$ (ovvero avendo fissato i) con pesi n_{ij} per i=1,2,...h

Le marginali sono X e Y con pesi $n_{i.}$ e $n_{.j}$ rispettivamente.

Attenzione quando le distribuzioni condizionate vengono espresse sulla base di frequenze relative, $f_{ij}=n_{ij}/N$ e $f_{i.}=n_{i.}/N$ e/o $f_{.j}=n_{.j}/N$.

Infatti da

X \ Y	y_1	y_2	...	y_j	...	y_k	$f_{i.}$
x_1	f_{11}	f_{12}	...	...	...	f_{1k}	$f_{1.}$
x_2	f_{21}	f_{22}	...	...	...	f_{2k}	$f_{2.}$
...	...	...	...	...	...	...	...
x_i	...	...	...	f_{ij}	...	...	$f_{i.}$
...	...	...	...	...	...	...	...
x_h	f_{h1}	f_{h2}	...	...	...	f_{hk}	$f_{h.}$
$f_{.j}$	$f_{.1}$	$f_{.2}$	...	$f_{.j}$	...	$f_{.k}$	1

Si deduce che

$X|Y=y_j$ (ovvero avendo fissato la colonna j) ha pesi $\frac{f_{ij}}{f_{.j}}$ per j=1,2,...k

$Y|X=x_i$ (ovvero avendo fissato la riga i) ha pesi $\frac{f_{ij}}{f_{i.}}$ per i=1,2,...h

Infatti solo in questo caso $\sum_{i=1}^{h} \frac{f_{ij}}{f_{.j}} = 1$ e $\sum_{j=1}^{k} \frac{f_{ij}}{f_{i.}} = 1$

–o–o–o–o–o–o–o–o–o–

Esempio guida

Frequenze assolute

X \ Y	M	F	$n_{i.}$
19	5	5	10
20	8	10	18
21	5	10	15
22	2	5	7
$n_{.j}$	20	30	50

Frequenze relative

X \ Y	M	F	$f_{i.}$
19	0,1	0,1	0,2
20	0,16	0,2	0,36
21	0,1	0,2	0,3
22	0,04	0,1	0,14
$f_{.j}$	0,4	0,6	1

Frequenze condizionate (per riga) $\frac{f_{ij}}{f_{.j}}$

X|Y	M	F	
19	0,1 /0,4 = 0,25	0,1 /0,6 = 0,165	
20	0,16/0,4 = 0,4	0,2 /0,6 = 0,335	
21	0,1 /0,4 = 0,25	0,2 /0,6 = 0,335	
22	0,04/0,4 = 0,1	0,1 /0,6 = 0,165	
	0,4 /0,4 = 1	0,6 /0,6 = 1	

-o-o-o-o-o-o-o-o-o-

Esempio (tema 21/06/2012)

La seguente tabella riporta informazioni su *D* = dividendo per azione (in percentuale) e *A*= tipologia del titolo, riferito a 10 titoli azionari (n.b. dati simulati).

D	2	0	1	1	4	2	1	0	2	4
A	E	F	S	I	E	S	I	F	F	E

Si effettui lo spoglio dei dati, costruendo la corrispondente tabella a doppia entrata.

Effettuando lo spoglio congiunto di *A* ***e*** *D* si ottiene la tabella a lato dove
- n_{ij} si chiamano frequenze (assolute) congiunte
- $n_{i.}$ si chiamano frequenze (assolute) marginali di riga
- $n_{.j}$ si chiamano frequenze (assolute) marginali di colonna
- la tabella si chiama tabella a doppia entrata e descrive il fenomeno congiunto (*A,D*)

			j=	*1*	*2*	*3*	*4*	
				D				
i		n_{ij}		**0**	**1**	**2**	**4**	$n_{i.}$
1	**A**	E		0	0	1	2	3
2		F		2	0	1	0	3
3		I		0	2	0	0	2
4		S		0	1	1	0	2
		$n_{.j}$		**2**	**3**	**3**	**2**	**N= 10**

La variabile marginale D è descritta da

	D	
	0 1 2 4	
$n_{.j}$	2 3 3 2	N= 10

La variabile marginale A è descritta da

		$n_{i.}$
A	E	3
	F	3
	I	2
	S	2
		N= 10

Un esempio di variabile condizionata D|A="F"

		D				
	n_{ij}	0	1	2	4	$n_{i.}$
A	F	2	0	1	0	3

Un esempio di variabile condizionata A|D=2

		D
	n_{ij}	2
A	E	1
	F	1
	I	0
	S	1
	$n_{.j}$	3

Esempio (tema 9/4/2010)

Sono stati raccolti nella tabella seguente i prezzi per litro di benzina (B), esposti da 10 distributori dislocati sul territorio nazionale (Z: N=nord, C=centro, S=sud).

B	1333	1345	1337	1348	1329	1347	1331	1336	1334	1345
Z	S	C	C	S	C	N	C	N	S	N

Utilizzando per la variabile B le classi 1320-|1330 , 1330-|1335 , 1335-|1350 , raggruppare B e Z in una tabella a doppia entrata.

Usando le frequenze assolute

		Z			
	n_{ij}	C	N	S	$n_{i.}$
B (Classi)	1320-|1330	1	0	0	1
	1330-|1335	1	0	2	3
	1335-|1350	2	3	1	6
	$n_{.j}$	4	3	3	N=10

Le stesse interpretazioni fatte in precedenza possono essere svolte usando le frequenze relative

		Z			
	f_{ij}	C	N	S	$f_{i.}$
B (Classi)	1320-|1330	0.1	0	0	0.1
	1330-|1335	0.1	0	0.2	0.3
	1335-|1350	0.2	0.3	0.1	0.6
	$f_{.j}$	0.4	0.3	0.3	F=1

Importante dettaglio emerge quando si considerano le distribuzioni condizionate relative

es. Se si considera la distribuzione B|Z=”C”, si avrebbe

		Z
	f_{ij}	C
B (Classi)	1320-|1330	0.1
	1330-|1335	0.1
	1335-|1350	0.2
	$f_{.j}$	**0.4**

Ovvero in modo equivalente

B (Classi)	f_{i1}
1320-|1330	0.1
1330-|1335	0.1
1335-|1350	0.2
	$f_{.1}$ =0.4

Non si tratta di una coerente distribuzione (la somma delle freq relative non è pari ad 1). Bisogna riproporzionare le frequenze congiunte rispetto alla frequenza marginale (0.4) della modalità condizionante, da cui ottenere

B (Classi)	f_{i1} / $f_{.1}$
1320-|1330	0.1/0.4=0.25
1330-|1335	0.1/0.4=0.25
1335-|1350	0.2/0.4=0.50
	F=1

-o-o-o-o-o-o-o-o-o-

<u>Esempio scatterplot di tabella a doppia entrata (13/04/2012)</u>

La tabella seguente riporta le distribuzioni <u>condizionate</u> relative al differenziale BTP-Bund (*S*: B=”< 300”, M=”300–|400”, A=”> 400”) dei contratti di apertura di conto deposito (*D*) stipulati da una banca (n.b. dati fittizi):

		D		
		0 –| 20	20 –| 60	60 –| 80
S	B	0.5	0.5	0
	M	0.2	0.6	0.2
	A	0	0	1

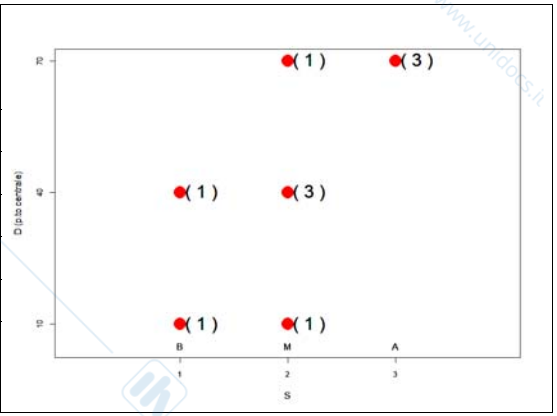
Ricostruire le frequenze assolute congiunte, sapendo che $n_{1.}$=2, $n_{2.}$=5, $n_{3.}$=3.

		D			
		0 –| 20	20 –| 60	60 –| 80	$n_{i.}$
S	B	1	1	0	2
	M	1	3	1	5
	A	0	0	3	3

Fare un opportuno grafico del fenomeno congiunto (*S*,*D*) .

N.b. verranno usati i punti centrali delle classi. La tabella può essere riscritta come segue

S	D	D (punto centrale)	n_{ij}
B	0 –| 20	10	*1*
M	0 –| 20	10	*1*
B	20 –| 60	40	*1*
M	20 –| 60	40	*3*
M	60 –| 80	70	*1*
A	60 –| 80	70	*3*

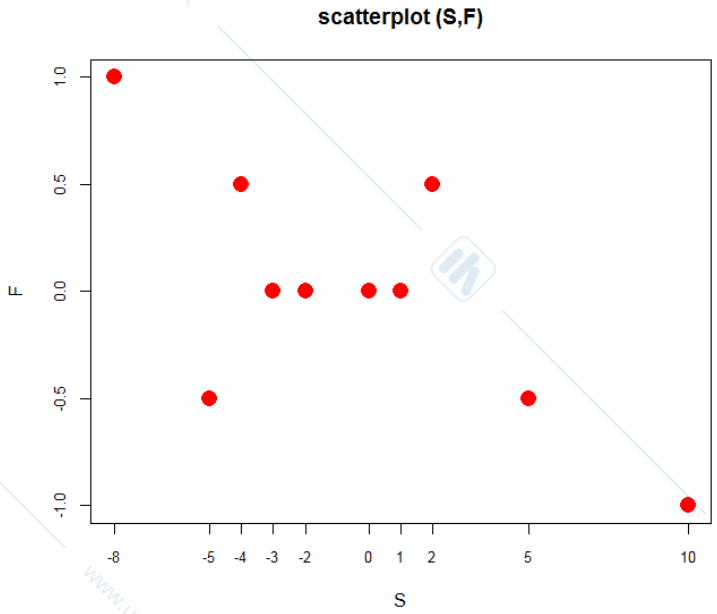


Esempio su scatterplot (17/01/2014)

La tabella seguente riporta le variazioni dello spread Btp-Bund (*S*) decennali e dell'indice FTSEMIB (*F*) registrati a fine giornata in 10 giorni di contrattazione :

S	-8	2	0	5	10	-5	-3	-2	1	-4
F	+1	+0.5	0	-0.5	-1	-0.5	0	0	0	+.5

Si rappresenti con un grafico opportuno il fenomeno congiunto (*S, F*).



Alcuni indici sintetici per distribuzioni condizionate

Medie condizionate : $M(Y|X=x_i)$ e $M(X|Y=y_j)$

$$M(Y|X=x_i) \equiv \mu_{Y|X=x_i} \equiv \mu_Y(x_i) = \frac{1}{n_{i.}}\sum_{j=1}^{k} y_j n_{ij} = \sum_{j=1}^{k} y_j \frac{f_{ij}}{f_{i.}}$$

per i=1,2,...,h

$$M(X|Y=y_j) \equiv \mu_{X|Y=y_j} \equiv \mu_X(y_j) = \frac{1}{n_{.j}}\sum_{i=1}^{h} x_i n_{ij} = \sum_{i=1}^{h} x_i \frac{f_{ij}}{f_{.j}}$$

per j=1,2,...,k

Varianze condizionate: $Var(Y|X=x_i)$ e $Var(X|Y=y_j)$

$$Var(Y|X=x_i) \equiv \sigma^2_{Y|X=x_i} \equiv \sigma^2_{Y(x_i)} = \frac{1}{n_{i.}}\sum_{j=1}^{k}\left(y_j - \mu_Y(x_i)\right)^2 n_{ij}$$

per i=1,2,...,h

$$Var(X|Y=y_j) \equiv \sigma^2_{X|Y=y_j} \equiv \sigma^2_{X(y_i)} = \frac{1}{n_{.j}}\sum_{i=1}^{h}\left(x_i - \mu_X(y_j)\right)^2 n_{ij}$$

per j=1,2,...,k

(si hanno scritture analoghe a quelle delle medie condizinate nel caso di impiego delle frequenze relative)

Con riferimento ad un fenomeno bivariato, si possono enunciare i seguenti teoremi.

Teo. <u>Media a due stadi</u>

Si consideri la v.s. (quantitativa) Y e una variabile X (covariata di Y). Allora vale che

$$M_X[M_Y(Y|X)] = \mu_Y$$

Dim.:

$$M_X[M_Y(Y|X)] = M_X\left[\frac{1}{n_{i.}}\sum_{j=1}^{k} y_j n_{ij}\,|X = x_i\right] =$$

$$= \frac{1}{N}\sum_{i=1}^{h}\left[\frac{1}{n_{i.}}\sum_{j=1}^{k} y_j n_{ij}\right] n_{i.} =$$

$$= \frac{1}{N}\sum_{i=1}^{h}\sum_{j=1}^{k} y_j n_{ij} = \frac{1}{N}\sum_{j=1}^{k} y_j \sum_{i=1}^{h} n_{ij} = \frac{1}{N}\sum_{j=1}^{k} y_j n_{.j} = \mu_Y$$

■

Es.

Si riprenda la tabella

		D			
		0 –| 20	20 –| 60	60 –| 80	$n_{i.}$
S	B	1	1	0	2
	M	1	3	1	5
	A	0	0	3	3

Si utilizzino i punti centrali delle classi di D e la si rinomini in

		Y			
		10	40	70	$n_{i.}$
X	B	1	1	0	2
	M	1	3	1	5
	A	0	0	3	3

Si calcoli $M_X[M_Y(Y|X)] = \mu_Y$

Essendo

	$M_Y(Y\|X)$	$n_{i.}$
$M_Y(Y\|X=B)$	25	2
$M_Y(Y\|X=M)$	40	5
$M_Y(Y\|X=A)$	70	3
		10

Si ha $M_X[M_Y(Y|X)] = \mu_Y = 46$

Es.

X \ Y	M	F	$n_{i.}$
19	5	5	10
20	8	10	18
21	5	10	15
22	2	5	7
$n_{.j}$	20	30	50

	$M_X(X\|Y)$	$n_{.j}$
$M_X(X\|Y=y_1)$	20,2	20
$M_X(X\|Y=y_2)$	20,5	30
		50

$M_Y[M_X(X|Y)] = \mu_X = 20{,}38$

-------- o - o --------

Teo. <u>Scomposizione della varianza</u>
Si consideri la v.s. (quantitativa) Y e una variabile X (covariata di Y). Allora vale che

$$M_X[Var_Y(Y|X)] + Var_X[M_Y(Y|X)] = \sigma_Y^2$$

Dim.:

$$\begin{aligned}
\mathrm{Var}(Y) &= \frac{1}{N}\sum_{j=1}^{k}(y_j - \mu_Y)^2 n_{.j} \\
&= \frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{h}(y_j - \mu_Y(x_i) + \mu_Y(x_i) - \mu_Y)^2 n_{ij} \\
&= \frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{h}\left[(y_j - \mu_Y(x_i)) + (\mu_Y(x_i) - \mu_Y)\right]^2 n_{ij} \\
&= \frac{1}{N}\left(\sum_{j=1}^{k}\sum_{i=1}^{h}(y_j - \mu_Y(x_i))^2 n_{ij}\right. \\
&\qquad + \sum_{j=1}^{k}\sum_{i=1}^{h}(\mu_Y(x_i) - \mu_Y)^2 n_{ij} \\
&\qquad \left. + 2\sum_{j=1}^{k}\sum_{i=1}^{h}(y_j - \mu_Y(x_i))(\mu_Y(x_i) - \mu_Y) n_{ij}\right) \\
&= \frac{1}{N}\sum_{i=1}^{h}\left(\frac{1}{n_{i.}}\sum_{j=1}^{k}(y_j - \mu_Y(x_i))^2 n_{ij}\right) n_{i.}\eta^2 = \frac{Var[M(Y|X)]}{Var(Y)}
\end{aligned}$$

$$\begin{aligned}
&+ \frac{1}{N}\sum_{i=1}^{h}(\mu_Y(x_i) - \mu_Y)^2 n_{i.} \\
&\qquad + \frac{2}{N}\sum_{i=1}^{h}(\mu_Y(x_i) - \mu_Y)\sum_{j=1}^{k}(y_j - \mu_Y(x_i)) n_{ij}
\end{aligned}$$

$$= M_X[Var_Y(Y|X)] + Var_X[M_Y(Y|X)] + 0 \quad \blacksquare$$

Es.
Si riprenda la tabella

		Y			
		10	40	70	$n_{i.}$
X	B	1	1	0	2
	M	1	3	1	5
	A	0	0	3	3

Si calcoli $M_X[Var_Y(Y|X)] + Var_X[M_Y(Y|X)] = \sigma_Y^2$

	$M_Y(Y\|X)$	$n_{i.}$
$M_Y(Y\|X{=}B)$	25	2
$M_Y(Y\|X{=}M)$	40	5
$M_Y(Y\|X{=}A)$	70	3
		10

	$Var_Y(Y\|X)$	$n_{i.}$
$Var_Y(Y\|X{=}B)$	225	2
$Var_Y(Y\|X{=}M)$	360	5
$Var_Y(Y\|X{=}A)$	0	3
		10

Si ha $M_X[M_Y(Y|X)] = \mu_Y = 46$

$Var_X[M_Y(Y|X)] = 279$ $\qquad$ $M_X[Var_Y(Y|X)] = 225$

$Var_X[M_Y(Y|X)] + M_X[Var_Y(Y|X)]$=279+225= $Var(Y)$ = 504

Es.

X \ Y	M	F	$n_{i.}$
19	5	5	10
20	8	10	18
21	5	10	15
22	2	5	7
$n_{.j}$	20	30	50

	$M_X(X\|Y)$	$n_{.j}$
$M_X(X\|Y=y_1)$	20,2	20
$M_X(X\|Y=y_2)$	20,5	30
		50

	$Var_X(X\|Y)$	$n_{.j}$
$Var_X(X\|Y=y_1)$	0,86	20
$Var_X(X\|Y=y_2)$	0,91667	30
		50

$Var_Y[M_X(X|Y)]$ = 0,0216 $M_Y[Var_X(X|Y)]$ = 0,894

$Var_Y[M_X(X|Y)] + M_Y[Var_X(X|Y)]$ =0,0216+0,894=Var(X)=0,9156

-------- o - o --------

Vedi file xls.

Una generalizzazione: *i miscugli (detti anche misture)*

Si definisce miscuglio una v.s. X costituita da h v.s. X_1, X_2 ...X_h avente peso π_1, π_2 ,..., π_h con $\sum_i \pi_i = 1$

Media di miscugli (es. portafoglio titoli)

Sia $\boldsymbol{X} = \{X_1, X_2,...,X_h\}$

Sia la Matrice dei dati

T	X_1	X_2	...	X_j	...	X_h
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1h}
2	x_{21}	x_{22}	...	...	...	x_{2h}
...	...	...	...	...	...	...
i	...	...	...	x_{ij}	...	...
...	...	...	...	...	...	...
T	x_{T1}	x_{T2}		x_{Tj}	...	x_{Th}

Sia la Matrice dei pesi

T	X_1	X_2	...	X_j	...	X_h
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1h}
2	n_{21}	n_{22}	...	...	...	n_{2h}
...	...	...	...	...	...	...
i	...	...	...	n_{ij}	...	...
...	...	...	...	...	...	...
T	n_{T1}	n_{T2}		n_{Tj}	...	n_{Th}
	$\mathbf{n_1}$	$\mathbf{n_2}$		$\mathbf{n_j}$	**...**	$\mathbf{n_h}$

Allora

$M(X_j) \equiv \bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{T} x_{ij} n_{ij}$ per j=1,2,...,h

Posto $N = \sum_{j=1}^{h} n_j$ si ha inoltre

$$M(\boldsymbol{X}) \equiv \bar{\boldsymbol{x}}_{\boldsymbol{P}} = \frac{1}{N}\sum_{j=1}^{h}\sum_{i=1}^{T} x_{ij}n_{ij} = \frac{1}{N}\sum_{j=1}^{h}\left(\frac{1}{n_j}\sum_{i=1}^{T} x_{ij}n_{ij}\right)n_j$$
$$= \frac{1}{N}\sum_{j=1}^{h} \bar{x}_j n_j$$

Esempio

Matrice dei dati

Tempo	X1	X2	X3
1	3	7	5
2	3	9	6
3	4	8	4
4	3	7	6
5	6	7	4
6	6	10	4

Matrice dei pesi

Tempo	X1	X2	X3
1	1	1	3
2	1	1	2
3	2	1	1
4	2	1	1
5	1	1	1
6	3	1	2
Nj	**10**	**6**	**10**

Media	4.4	8	4.9

Media portafoglio	5.423077

Varianza di miscuglio

Si ricordi che dato $\boldsymbol{X} = \{X_1, X_2,...,X_h\}$

$M(X_j) \equiv \bar{x}_j$ per j=1,2,...,h

$\mathrm{Var}(X_j) \equiv \sigma_j^2 = \frac{1}{n_j}\sum_{i=1}^{T}(x_{ij} - \bar{x}_j)^2 n_{ij}$ per j=1,2,...,h

$M(\boldsymbol{X}) \equiv \bar{\boldsymbol{x}}_P = \frac{1}{N}\sum_{j=1}^{h}\bar{x}_j n_j$

$$\mathrm{Var}(\boldsymbol{X}) = \sigma_P^2 = M[(\boldsymbol{X} - \bar{\boldsymbol{x}}_P)^2] = \frac{1}{N}\sum_{j=1}^{h}\sum_{i=1}^{T}(x_{ij} - \bar{\boldsymbol{x}}_P)^2 n_{ij}$$
$$= \frac{1}{N}\sum_{j=1}^{h}\sum_{i=1}^{T}(x_{ij} - \bar{x}_j + \bar{x}_j - \bar{\boldsymbol{x}}_P)^2 n_{ij}$$
$$= \frac{1}{N}\sum_{j=1}^{h}\sum_{i=1}^{T}\left((x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{\boldsymbol{x}}_P)\right)^2 n_{ij}$$
$$= \frac{1}{N}\sum_{j=1}^{h}\left(\frac{1}{n_j}\sum_{i=1}^{T}(x_{ij} - \bar{x}_j)^2 n_{ij}\right)n_j + \frac{1}{N}\sum_{j=1}^{h}(\bar{x}_j - \bar{\boldsymbol{x}}_P)^2 n_j$$
$$+ \frac{2}{N}\sum_{j=1}^{h}(\bar{x}_j - \bar{\boldsymbol{x}}_P)\sum_{i=1}^{T}(x_{ij} - \bar{x}_j)n_{ij}$$
$$= M[\sigma_j^2] + Var[\bar{x}_j]$$

="Var within" + "Var Between" := $\sigma_P^{2\,*} + \bar{\sigma}_P^2$

(Segue) esempio su portafoglio titoli

Titoli	Media dei singoli Titoli	VAR dei singoli Titoli	nj
X1	4.4	1.84	10
X2	8	1.33334	6
X3	4.9	0.69	10

VAR Within	1.280769
VAR Between	2.040237
VAR Portafoglio	3.321006
VAR Within / Var Portafoglio	0.386

L'ultimo rapporto (in senso lato) è una misura di quanto differente è il rischio degli asset nel portafoglio. Se alto significa che in portafoglio ci sono titoli con volatilità differenti.
Analogamente il rapporto VAR Between / Var Portafoglio (0.614) dice quanto differente è la redditività dei singoli asset.

La funzione di regressione

Si consideri la tabella

		Y			
		0 –| 20	20 –| 60	60 –| 80	$n_{i.}$
X	1	1	0	0	1
	2	0	3	3	6
	3	0	2	1	3
	$n_{.j}$	1	5	4	10

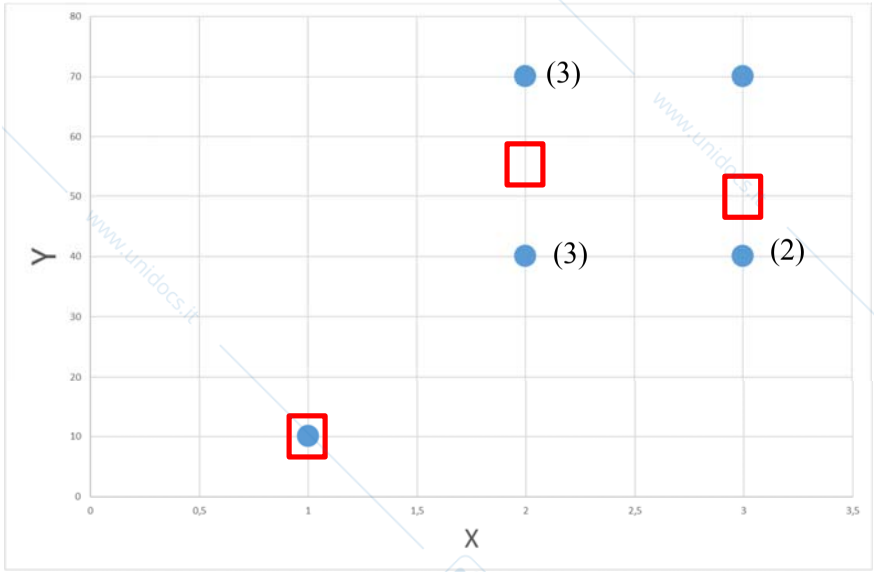
Si è già visto come disegnare il grafico congiunto e il calcolo delle medie condizionate Y|X ovvero

	$M_Y(Y\|X)$	$n_{i.}$
M[Y| X=1]	10	1
M[Y| X=2]	55	6
M[Y| X=3]	50	3
		10

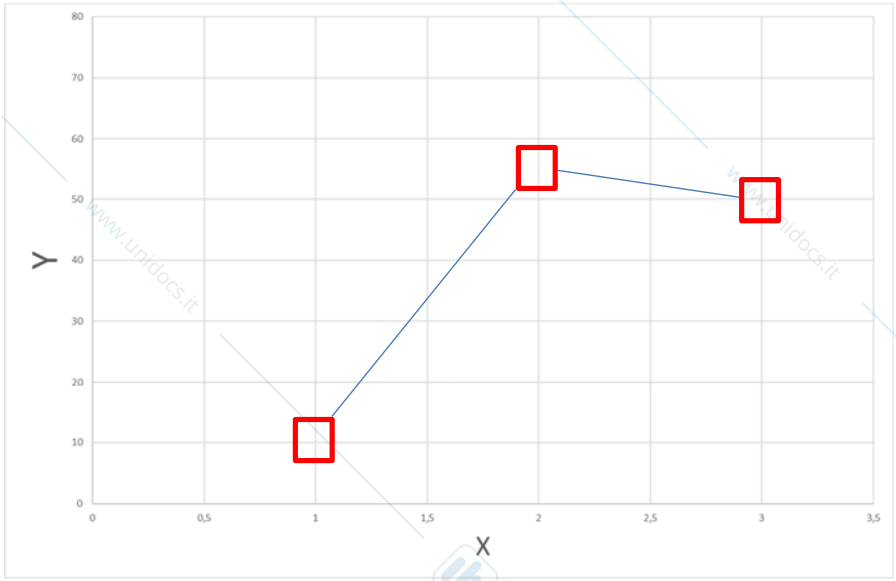
Da cui $Var[M_Y(Y|X)] = 174$, $Var(Y) = 369$.

Da cui dedurre per differenza (in base al teorema sulla scomposizione della varianza) $M[Var_Y(Y|X)] = 195$.

Di seguito viene riportato il grafico (X,Y) a cui sono state aggiunte (in rosso) le coordinate delle medie condizionate



Supponiamo di rimuovere le coordinate della tabella, e di unire con una spezzata le coordinate delle medie condizionate. Si ottiene



che prende il nome di *funzione di regressione* ovvero è il luogo dei punti passanti *PER* le medie condizionate. Una misura della capacità delle medie condizionate di rappresentare il fenomeno bivariato è dato dal *rapporto di correlazione*

$$\eta^2 = \frac{Var[M(Y|X)]}{Var(Y)}$$

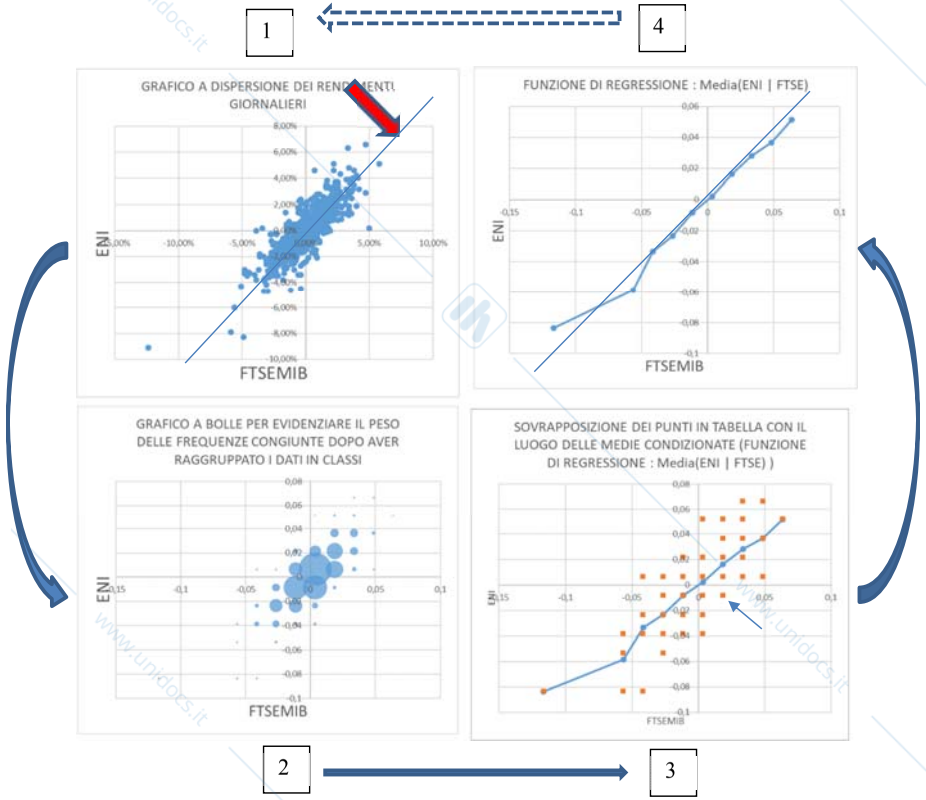
che per il nostro esempio è pari a $\eta^2 = \frac{174}{369} = 0{,}4715$.
Significa che le sole medie condizionate hanno la capacità di *spiegare* la dipendenza *in media* di Y da X per il 47,15% della variabilità complessiva.

Domanda: è possibile associare alla funzione di regressione un polinomio?

-o-o-o-o-o-o-o-o-o-

Esempio tratto da : *P_0-portfolio 2016-2017.xlsx*

1- dati singoli
2- dati in tabella
3- dati in tabella con funzione di regressione
4- funzione di regressione



Q.
- cosa suggerisce la funzione di regressione?
- cosa avremmo potuto fare partendo direttamente dai (molti) dati in figura 1?
- nel caso in cui partissimo direttamente dai dati singoli, come posso valutare la capacità di descrivere il fenomeno bivariato?

Modelli interpolanti : la Regressione

Si ipotizzi di disporre delle variabili $\{Y,X\}$. Si ipotizzi di essere interessati allo studio della dipendenza di Y da X . Si ipotizzi che esista una funzione (ignota nella forma) del tipo

$$Y = f(X;\boldsymbol{\theta})$$

Assumendo $f(.)$ derivabile fino ad almeno al p-esimo ordine, usando lo sviluppo in serie di Taylor, si supponga di sviluppare in serie $f(.)$ e di essere interessati allo studio della funzione polinomiale (approssimante quella vera)

$$Y = a_0 + a_1 X + \cdots + a_p X^p$$

Si ottiene un modello *lineare nei parametri*.
Alternative possono essere modelli aventi struttura non lineare (nei parametri). In questo corso (ed in genere nelle applicazioni) si è interessati a modelli al più linearizzabili nei parametri.

I modelli comunemente usati per descrivere la dipendenza di Y da X possono, ad esempio, essere

$Y=a_0+a_1 X+...+a_p X^p$
$Y=a_0+a_1 X$
$Y=a_1 X$
$Y=a_0+a_1 g(X)$
$Exp(Y)=(a_0+a_1 g(X))$
$ln(Y)=(a_0+a_1 g(X))$
...

(Temi: **TUTTI**)

(Una possibile generalizzazione)
Se si ipotizza di disporre di m+1 variabili $\{Y,X_1,\dots,X_m\}$ e di essere interessati allo studio della dipendenza di Y da $X_1,\dots,X_m$ e che esista una funzione

$$Y=f(X_1,\dots,X_m\,;\,\alpha_0,\alpha_1,\dots,\alpha_m)$$

di cui non si conosce la struttura, assumendo $f(.)$ derivabile fino ad almeno il primo ordine, usando ad esempio lo sviluppo in serie di Taylor, oggetto dello studio potrà essere

$$Y = a_0 + a_1X_1 + \cdots + a_pX_m$$

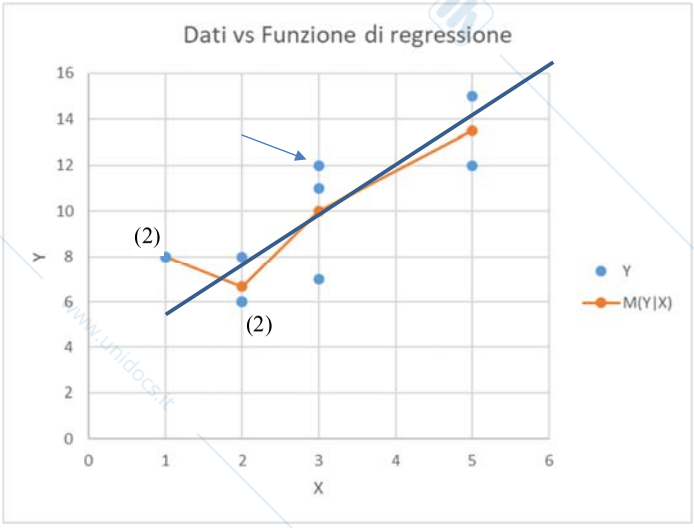
Ad esempio se si considerano le sole variabili X_1, X_2 il modello di riferimento potrà ad esempio essere

$$Y = a_0 + a_1X_1 + a_2X_2$$

Il metodo dei minimi quadrati

Esempio guida

	Y						
X	6	7	8	11	12	15	$n_{i.}$
1	0	0	2	0	0	0	2
2	2	0	1	0	0	0	3
3	0	1	0	1	1	0	3
5	0	0	0	0	1	1	2
$n_{.j}$	2	1	3	1	2	1	10



Q. qual è l'ordine massimo del polinomio che possiamo usare nell'esempio guida?

Si consideri il generico polinomio : $Y = a_0 + a_1 X + \cdots + a_p X^p$

(n.b. di ordine $p \leq$ (#medie condizionate−1) !)

Problema

Come calcolare $\{a_0, a_1, \ldots, a_p\}$ sapendo che il modello potrà al più passare ***tra*** i punti e non *per* i punti?

Strategia

Cercare $\{a_0, a_1, \ldots, a_p\}$ in modo da perdere la minore quantità possibile di informazioni, avendo scelto una opportuna funzione per misurare la perdita di informazioni.

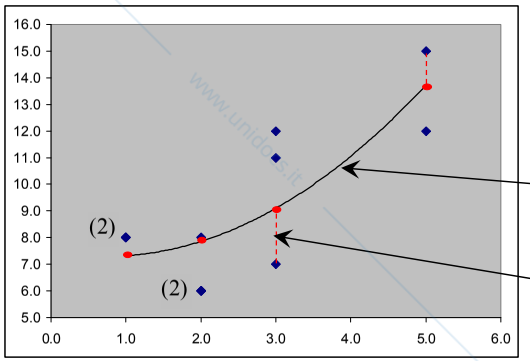
Strumento

Metodo dei minimi quadrati

Si supponga, in generale, di avere dati raggruppati in una Tabella "a doppia entrata". Ci si ponga l'obiettivo

$$\min_{a_0, a_1, \ldots, a_p} \sum_i \sum_j \left(y_j - \left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p\right)\right)^2 n_{ij}$$

L'argomento del problema di minimo è noto anche come *devianza residua* (*devres)*.
Quindi si può anche affermare che il problema si pone l'obiettivo di minimizzare la devianza residua.



Si dovrà quindi risolvere

$$\begin{cases} \dfrac{\partial devres}{\partial a_0} = (-2) \sum_i \sum_j \left(y_j - \left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p\right)\right) n_{ij} = 0 \\ \ldots\ldots\ldots \\ \dfrac{\partial devres}{\partial a_p} = (-2) \sum_i \sum_j x_i^p \left(y_j - \left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p\right)\right) n_{ij} = 0 \end{cases}$$

che è un sistema di equazioni *lineare* nelle incognite $a_0, a_1, \ldots, a_p$.
Si dimostra che l'Hessiano (matrice delle derivate parziali seconde) è positivo nella soluzione del sistema e quindi la soluzione del sistema di equazioni *È* un minimo.

La soluzione restituirà valori da associare ad $a_0, a_1, \ldots, a_p$ che verranno indicati con $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p$

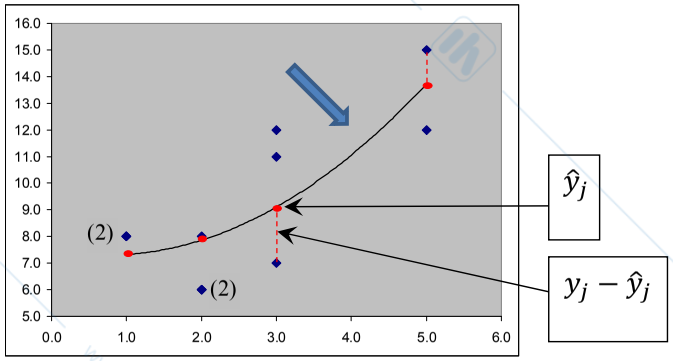
Disponendo di $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p$, si definiscono

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X + \cdots + \hat{a}_p X^p \qquad \textbf{modello interpolante}$$

$$\hat{y}_j = \hat{a}_0 + \hat{a}_1 x_i + \cdots + \hat{a}_p x_i^p \qquad \textbf{valore interpolato} \text{ di } y_j \text{ in } x_i$$

$$Y - (\hat{a}_0 + \hat{a}_1 X + \cdots + \hat{a}_p X^p) = Y - \hat{Y} \qquad \textbf{variabile residuo}$$

$$y_j - (\hat{a}_0 + \hat{a}_1 x_i + \cdots + \hat{a}_p x_i^p) = y_j - \hat{y}_j \qquad \textbf{residuo} \text{ in } (x_i, y_j)$$



Si noti che **posto** $Y = \boldsymbol{a_0} + a_1 X + \cdots + a_p X^p$ da

$$\frac{\partial devres}{\partial a_0} = -2 \sum_i \sum_j \left(y_j - \left(a_0 + a_1 x_i + \cdots + a_p x_i^p \right) \right) n_{ij} = 0$$

e sostituendo al posto di $a_0, a_1, \ldots, a_p$ i valori $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p$, soluzione del sistema di equazioni ottenuto col metodo dei minimi quadrati, si ricava che

$$\frac{1}{N} \sum_i \sum_j \left(y_j - \left(\hat{a}_0 + \hat{a}_1 x_i + \cdots + \hat{a}_p x_i^p \right) \right) n_{ij} = 0$$

quindi $M\left(Y - \left(\hat{a}_0 + \hat{a}_1 X + \cdots + \hat{a}_p X^p\right)\right) = 0$

ovvero $M\left(Y - \hat{Y}\right) = 0$.

Da cui si deduce che la **somma dei residui è nulla** e che

$$M(Y) = M\left(\hat{Y}\right)$$

Se invece $Y = f(X; a_0, a_1, \ldots, a_p)$ (es. $Y = \exp(a_0 + a_1 X + \ldots + a_p X^p)$)

$$\frac{\partial devres}{\partial a_0} = -2 \sum_i \sum_j \frac{\partial f(x_i; a_0, a_1, \ldots, a_p)}{\partial a_0} \left(y_j - f(x_i; a_0, a_1, \ldots, a_p)\right) n_{ij} = 0$$

che non garantisce che $M(Y) = M(\hat{Y})$.

—o—o—o—o—o—o—o—o—o—

N.B. ***TUTTI*** i calcoli possono essere fatti "riscrivendo" la tabella a doppia entrata in modo equivalente come segue

X\Y	y_1	y_2	…	Y_j	…	y_{k-1}	y_k
x_1	n_{11}	n_{12}	…	n_{1j}	…	$n_{1,k-1}$	n_{1k}
x_2	n_{21}	n_{22}	…	n_{2j}	…	$n_{2,k-1}$	n_{2k}
…	…	…	…	…	…	…	…
X_i	…	…	…	n_{ij}	…	…	…
…	…	…	…	…	…	…	…
x_{h-1}	…	…	…	…	…	$n_{h-1,k-1}$	$n_{h-1,k}$
x_h	…	…	…	…	…	$n_{h,k-1}$	n_{hk}

X	Y	n_{ij}	$n_{ij} \equiv n_i$
x_1	y_1	n_{11}	n_1
x_1	y_2	n_{12}	n_2
…	…	…	…
x_i	y_j	n_{ij}	n_i
…	…	…	…
x_{h-1}	y_{k-1}	$n_{h-1,k-1}$	n_{hxk-1}
x_h	y_k	n_{hk}	n_{hxk}

Si noti che, le frequenze congiunte, n_{ij}, possono quindi essere scritte come se fossero delle frequenze associate alle righe corrispondenti in tabella. In tal caso, indicizzando ciascuna riga solo con i (per i=1,…,hxk) e ponendo $(x_i,y_j) \equiv (x_i,y_i)$ e $n_{ij} \equiv n_i$ si ha che il problema dei minimi quadrati può essere in modo equivalente ritrascritto in forma più semplice e senza l'uso della doppia sommatoria come segue

$$\min_{a_0,a_1,\dots,a_p} \sum_i \left(y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p\right)\right)^2 n_i$$

$$\begin{cases} \dfrac{\partial devres}{\partial a_0} = (-2)\sum_i \left(y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p\right)\right) n_i = 0 \\ \dots \\ \dfrac{\partial devres}{\partial a_p} = (-2)\sum_i x_i^p \left(y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_p x_i^p\right)\right) n_i = 0 \end{cases}$$

in cui si nota l'evidente notazione più semplice.

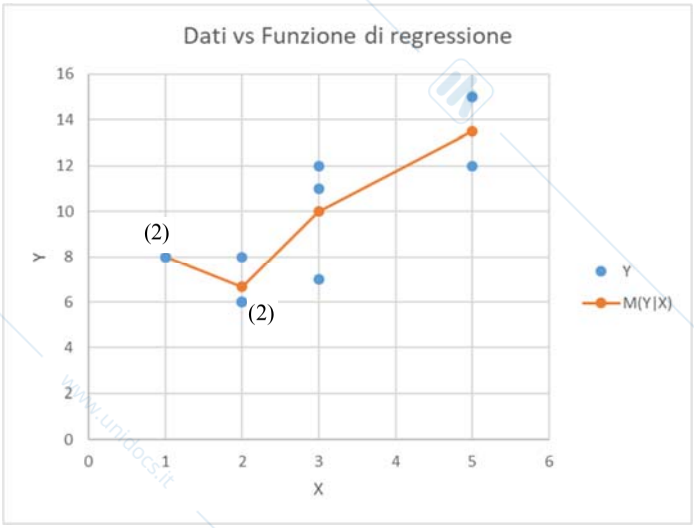
N.B. (terminologia) si parla di

- regressione pesata (ponderata) se <u>almeno un</u> n_i è > 1
- regressione semplice se <u>tutti gli</u> n_i sono = 1

In generale, per motivi di semplicità, si userà il termine "regressione" senza specifico distinguo tra i due casi a meno che il contesto non lo renda necessario.

Esempio (<u>*rilevante*</u>)(a).Hp: $Y = a_0 + a_1 X$

	Y						
X	6	7	8	11	12	15	$n_{i.}$
1	0	0	2	0	0	0	2
2	2	0	1	0	0	0	3
3	0	1	0	1	1	0	3
5	0	0	0	0	1	1	2
$n_{.j}$	2	1	3	1	2	1	10



Si risolva

$$\min_{a_0,a_1} \sum_i \sum_j \left(y_j - (a_0 + a_1 x_i)\right)^2 n_{ij}$$

Se si usa la rappresentazione della tabella come prima suggerito ovvero

X	Y	n_i
1	8	2
2	6	2
2	8	1
3	7	1
3	11	1
3	12	1
5	12	1
5	15	1

si ha quindi

$$\min_{a_0,a_1} \sum_i \sum_j \left(y_j - (a_0 + a_1 x_i)\right)^2 n_{ij}$$
$$\equiv \min_{a_0,a_1} \sum_i \left(y_i - (a_0 + a_1 x_i)\right)^2 n_i$$

da cui le derivate prime

$$\begin{cases} \dfrac{\partial devres}{\partial a_0} = (-2)\sum_i \left(y_i - (a_0 + a_1 x_i)\right) n_i = 0 \\ \dfrac{\partial devres}{\partial a_1} = (-2)\sum_i x_i \left(y_i - (a_0 + a_1 x_i)\right) n_i = 0 \end{cases}$$
$$\begin{cases} \sum_i \frac{1}{N}\left(y_i - (a_0 + a_1 x_i)\right) n_i = 0 \\ \sum_i \frac{1}{N} x_i \left(y_i - (a_0 + a_1 x_i)\right) n_i = 0 \end{cases}$$

$$\begin{cases} \sum_i \frac{1}{N}(y_i) n_i - \sum_i \frac{1}{N}(a_0) n_i - \sum_i \frac{1}{N}(a_1 x_i) n_i = 0 \\ \sum_i \frac{1}{N}(x_i y_i) n_i - \sum_i \frac{1}{N}(x_i a_0) n_i - \sum_i \frac{1}{N}(a_1 x_i^2) n_i = 0 \end{cases}$$

$$\begin{cases} M(Y) - a_0 - a_1 M(X) = 0 \\ M(XY) - a_0 M(X) - a_1 M(X^2) = 0 \end{cases}$$

$$\begin{cases} a_1 = \dfrac{\boldsymbol{M(XY) - M(X)M(Y)}}{M(X^2) - [M(X)]^2} = \dfrac{\boldsymbol{Cov(X,Y)}}{Var(X)} \\ a_0 = M(Y) - a_1 M(X) \end{cases}$$

N.B. L'Hessiano (matrice delle derivate seconde - elementi divisi per $2N$) è

$$H = \begin{bmatrix} 1 & M(X) \\ M(X) & M(X^2) \end{bmatrix}$$

il cui determinante, $|H| = M(X^2) - [M(X)]^2$, è quindi $Var(X) > 0$. Quindi la soluzione trovata è un minimo.

N.B. Si rammenti che quando si è parlato di varianza di una combinazione lineare di variabili, era emersa una statistica nota come covarianza. Tale statistica compare qui al numeratore del coefficiente angolare. In generale si definisce covarianza

$$\begin{aligned} \boldsymbol{Cov(X,Y)} &= \boldsymbol{M}[(X - \mu_X)(Y - \mu_Y)] \\ &= \boldsymbol{M}[XY - \mu_X Y - X\mu_Y + \mu_X \mu_Y] \\ &= \boldsymbol{M}[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\ &= \boldsymbol{M}[XY] - \mu_X \mu_Y \end{aligned}$$

che coincide con l'espressione al numeratore di a_1.

In particolare $Cov(X,Y)$ può essere positiva o negativa mentre il denominatore di a_1 è sempre positivo.

Il segno della *Cov* quindi definisce l'inclinazione della retta di regressione ed esprime l'intensità (non normalizzata) della dipendenza lineare tra X *e* Y.

Si osservi inoltre che, usando la "nuova" tabella, i calcoli delle statistiche necessarie per i parametri possono essere eseguiti "in colonna" come da esempio seguente.

X	Y	n_i	XY	X^2	Y^2	$X \cdot n_i$	$Y \cdot n_i$	$XY \cdot n_i$	$X^2 \cdot n_i$	$Y^2 \cdot n_i$
1	8	2	8	1	64	2	16	16	2	128
2	6	2	12	4	36	4	12	24	8	72
2	8	1	16	4	64	2	8	16	4	64
3	7	1	21	9	49	3	7	21	9	49
3	11	1	33	9	121	3	11	33	9	121
3	12	1	36	9	144	3	12	36	9	144
5	12	1	60	25	144	5	12	60	25	144
5	15	1	75	25	225	5	15	75	25	225
		$\sum n_i$ =10				$\sum x_i n_i$ =27	$\sum y_i n_i$ =93	$\sum x_i y_i n_i$ =281	$\sum x_i^2 n_i$ = 91	$\sum y_i^2 n_i$ = 947

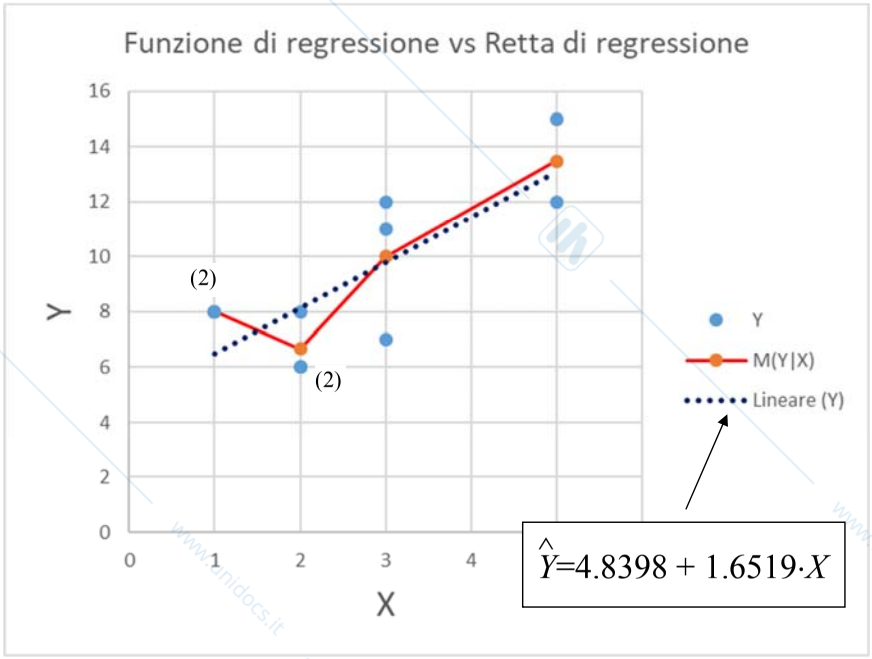
da cui agevolmente ottenere i momenti necessari per i calcoli

Per l'esempio

$$\begin{cases} \hat{a}_1 = \dfrac{M(XY) - M(X)M(Y)}{M(X^2) - [M(X)]^2} = \dfrac{Cov(X,Y)}{Var(X)} = \dfrac{\frac{281}{10} - \frac{27}{10} \cdot \frac{93}{10}}{\frac{91}{10} - \left(\frac{27}{10}\right)^2} = \dfrac{2.99}{1.81} = 1.65193 \\ \hat{a}_0 = M(Y) - \hat{a}_1 M(X) = \dfrac{93}{10} - 1.65193 \cdot \dfrac{27}{10} = 4.83978 \end{cases}$$

In definitiva si ottiene il modello interpolante

$$\hat{Y} = 4.8398 + 1.6519 \cdot X$$



Valori interpolati

$x_i(=X)$	Y	n_i	$\hat{y}_i\ (=\hat{Y})$ $=4.83978+$ $1.65193\cdot x_i$
1.0	8	2	6.4917
2.0	6	2	8.1436
2.0	8	1	8.1436
3.0	7	1	9.7956
3.0	11	1	9.7956
3.0	12	1	9.7956
5.0	12	1	13.0994
5.0	15	1	13.0994

da cui

$x_i(=X)$	$\hat{y}_i$	n_i
1.0	6.4917	2
2.0	8.1436	3
3.0	9.7956	3
5.0	13.0994	2

N.B. le frequenze associate ai valori interpolati sono le frequenze della marginale *X*

Residui

$x_i(=X)$		$y_i\ (=Y)$	n_i
1.0		8.0	2
2.0		6.0	2
2.0		8.0	1
3.0		7.0	1
3.0		11.0	1
3.0		12.0	1
5.0		12.0	1
5.0		15.0	1

da cui

$y_i - \hat{y}_i$ $(=residui)$	n_i
1.5083	2
-2.1436	2
-0.1436	1
-2.7956	1
1.2044	1
2.2044	1
-1.0994	1
1.9006	1

N.B. le frequenze associate ai residui sono le frequenze (congiunte) delle singole coordinate

Essendo il modello dotato di termine noto si ha che

$M(Y) = M(\hat{Y}) = 9.3$
$M(Y - \hat{Y}) = 0$ ovvero $\sum_i (y_i - \hat{y}_i) n_i = 0$

La devianza residua ovvero il valore numerico della funzione di perdita calcolata usando la soluzione per i parametri ottenuti coi minimi quadrati è pari a $\sum_i (y_i - \hat{y}_i)^2 n_i = 32.7069$.

-o-o-o-o-o-o-o-o-o-o-o-

<u>Teorema : scomposizione della varianza di un modello di regressione</u>

Si consideri un modello del tipo $Y = a_0 + a_1 X + \cdots + a_p X^p$ (ovvero un modello con presente il termine noto a_0).
Allora vale la scomposizione della varianza di Y come segue:

$$Var(Y) = Var(\hat{Y}) + Var(Y - \hat{Y})$$

= Varianza **<u>spiegata</u>** dal modello +
+Varianza **<u>residua</u>** (non spiegata)

$$= \bar{\sigma}^2_{a_0 + a_1 X + \cdots + a_p X^p} + \sigma^{*2}_{a_0 + a_1 X + \cdots + a_p X^p}$$

<u>DIMOSTRAZIONE</u>

$\mathrm{Var}(Y) = M[(Y - \mu_Y)^2] =$

$$= M\left[\left(Y - \hat{Y} + \hat{Y} - \mu_Y\right)^2\right] =$$

$$= M\left[\left(\hat{Y} - \mu_Y\right)^2\right] + M\left[\left(Y - \hat{Y}\right)^2\right] + 2 \cdot M\left[\left(\hat{Y} - \mu_Y\right)\left(Y - \hat{Y}\right)\right] =$$

= Varianza dei valori interpolati + Varianza dei residui + 0

$= \mathrm{Var}(\hat{Y}) + \mathrm{Var}[Y - \hat{Y}]$

N.B. => $M[(\hat{Y}-\mu_Y)(Y-\hat{Y})] =$
$= M[\hat{Y}(Y-\hat{Y})] - \mu_Y M(Y-\hat{Y}) = 0$

Infatti:

1) da $\frac{\partial devres}{\partial a_0} = 0$ si ha che $M[Y-\hat{Y}] = 0$.

2) Stante la struttura lineare nei parametri, moltiplicando per $\hat{a}_i$ l'equazione del gradiente corrispondente a $\frac{\partial devres}{\partial a_i}$, ripetendo l'operazione per tutte le equazioni, sostituendo al posto di $a_0, a_1, \dots, a_p$ i valori $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_p$,

$$\begin{cases} \sum_i \sum_j \hat{a}_0 \left(y_j - \left(\hat{a}_0 + \hat{a}_1 x_i + \cdots + \hat{a}_p {x_i}^p\right)\right) n_{ij} = 0 \\ \dots \\ \sum_i \sum_j \hat{a}_p {x_i}^p \left(y_j - \left(\hat{a}_0 + \hat{a}_1 x_i + \cdots + \hat{a}_p {x_i}^p\right)\right) n_{ij} = 0 \end{cases}$$

e quindi sommando tutte le equazioni, si ottiene

$$\sum (\hat{a}_0 + \hat{a}_1 x_i + \cdots + \hat{a}_p {x_i}^p)[y_i - (\hat{a}_0 + \hat{a}_1 x_i + \cdots + \hat{a}_p {x_i}^p)] n_i = 0$$

ovvero,

$$M[\hat{Y}(Y-\hat{Y})] = 0$$

-o-o-o-o-o-o-o-o-o-

N.B. SE $Y=f(X;a_1,\dots,a_p)$, ovvero senza a_0, allora potrebbe NON valere la scomposizione della varianza in quanto potrebbe essere $M[Y-\hat{Y}] \neq 0$

-o-o-o-o-o-o-o-o-o-o-

N.B. anche se il modello non è lineare e anche se $M(Y-\hat{Y}) \neq 0$, ovvero anche se $M(Y) \neq M(\hat{Y})$, la varianza residua σ^{*2} si calcolerà sempre come $M[(Y-\hat{Y})^2]$ sebbene non si tratti propriamente di una varianza.

-o-o-o-o-o-o-o-o-o-o-

Per l'esempio (a) da

$\hat{y}_i\ (=\hat{Y})$	n_i
6.4917	2
8.1436	3
9.7956	3
13.0994	2

e da

$y_i - \hat{y}_i$ (=*residui*)	n_i
1.5083	2
-2.1436	2
-0.1436	1
-2.7956	1
1.2044	1
2.2044	1
-1.0994	1
1.9006	1

o da

$\hat{y}_i(=\hat{Y})$	$y_i - \hat{y}_i$ (=*residui*)	n_i
6.4917	1.5083	2
8.1436	-2.1436	2
8.1436	-0.1436	1
9.7956	-2.7956	1
9.7956	1.2044	1
9.7956	2.2044	1
13.0994	-1.0994	1
13.0994	1.9006	1

si ha quindi

$$Var[Y] = Var[\hat{Y}] + Var[Y-\hat{Y}] = \bar{\sigma}^2_{a_0+a_1X} + \sigma^{*2}_{a_0+a_1X}$$

$$8.21 = 4.9393 + 3.270718$$

-o-o-o-o-o-o-o-o-o-o

Misure di adattamento

SE vale la scomposizione della varianza, ovvero

$$Var[Y] = Var(\hat{Y}) + Var[Y - \hat{Y}]$$

allora la *bontà di adattamento* di un *modello di regressione* si esprime con l'*indice di adattamento* (noto anche come *indice di determinazione*)

adattamento ai dati nullo

$\Uparrow$

$$0 \le R^2 = \frac{Var(\hat{Y})}{Var(Y)} = 1 - \frac{Var(Y-\hat{Y})}{Var(Y)} \le 1$$

$\Downarrow$

perfetto adattamento ai dati

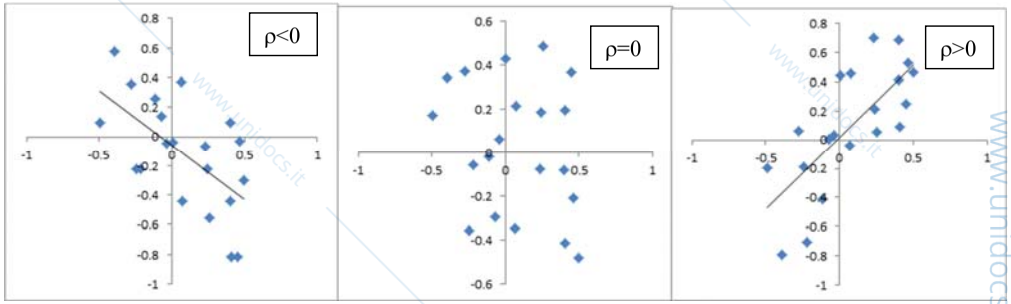
In particolare se : $Y = a_0 + a_1 X$ si ottiene

$$R^2 = \rho^2 = \frac{Var(\hat{Y})}{Var(Y)} = \frac{Var(\hat{a}_0 + \hat{a}_1 X)}{Var(Y)} = \frac{\hat{a}_1^2 Var(X)}{Var(Y)}$$
$$= \frac{Cov(X,Y)^2}{Var(X) \cdot Var(Y)}$$

N.B.

ρ^2 è l'<u>indice di determinazione lineare</u>

$-1 \le \rho \le 1$ è il <u>coefficiente di correlazione lineare</u>



-o-o-o-o-o-o-o-o-o-o-o

Riprendendo l'esempio (a) si ha

$$\rho^2 = \frac{Var(\hat{Y})}{Var(Y)} = \frac{4.9393}{8.21} = \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)\cdot\text{Var}(Y)} = \frac{[2.99]^2}{1.81\cdot 8.21} = 0.6016$$

N.B. Si deduce che

$$\bar{\sigma}^2_{a_0+a_1X} = Var(Y)\ [\ \rho^2] = 4.9393$$

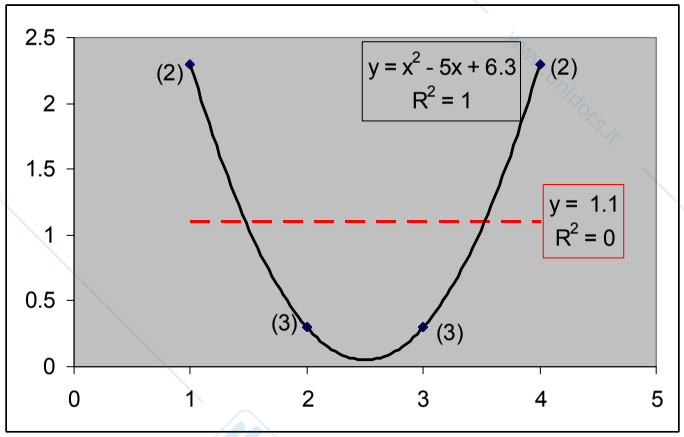
e

$$\sigma^{*2}_{a_0+a_1X} = Var(Y)\ [1-\rho^2] = 3.270718$$

—o—o—o—o—o—o—o—o—

Esempio (b) [casi estremi]

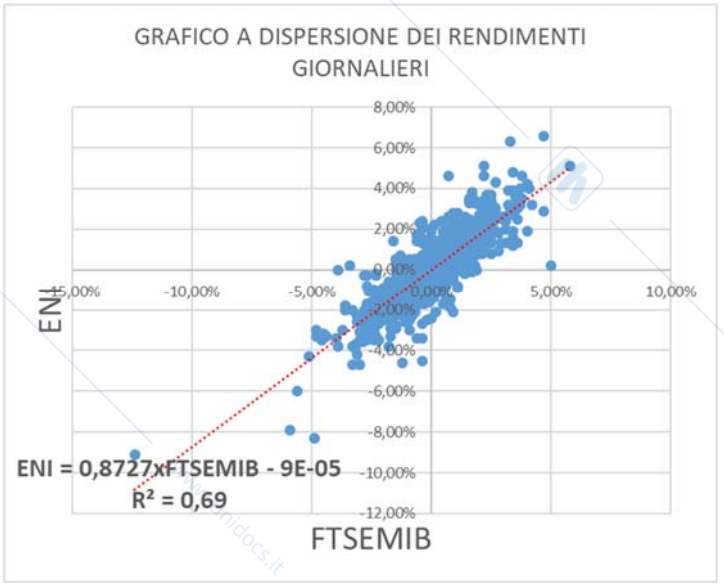
X	Y	n_{ij}
1.0	2.3	2
2.0	0.3	3
3.0	0.3	3
4.0	2.3	2



-o-o-o-o-o-o-o-o-o-o-o

Esempio tratto da : *P_0-portfolio 2016-2017.xlsx*

Usando i dati già commentati per la costruzione della funzione di regressione, il modello di regressione polinomiale di ordine 1 sarebbe stato :



Confronto tra Modelli

È frequente dover effettuare confronto tra modelli al fine di misurare quale tra quelli proposti si adatta meglio ai dati

Alcuni possibili casi

a) I) $Y = a_{0,I} + l_1(X; a_{1,I}, \dots, a_{p,I})$
II) $Y = a_{0,II} + l_2(X; a_{1,II}, \dots, a_{p,II})$

$\Rightarrow$ Confronti possibili :

$R_I^2 \gtreqless R_{II}^2$ oppure $\bar{\sigma}_I^2 \gtreqless \bar{\sigma}_{II}^2$ oppure $\sigma_I^{*2} \gtreqless \sigma_{II}^{*2}$

Es. I) $Y = a_{0,I} + a_{1,I}X$ II) $Y = a_{0,II} + a_{1,II}\ln(X)$

−o−o−o−o−o−o−o−o−

b) I) $Y = a_{0,I} + l_1(X; a_{1,I}, \dots, a_{p,I})$ II) $Y = l_2(X; a_{1,II}, \dots, a_{p,II})$

$\Rightarrow$ possibile confronto: $\sigma_I^{*2} \gtreqless \sigma_{II}^{*2}$

Es. I) $Y = a_{0,I} + a_{1,I}\exp(X)$ II) $Y = a_{1,II}X^2$

−o−o−o−o−o−o−o−o−

c) I) $Y = a_{0,I} + l_1(X; a_{1,I}, \dots, a_{p,I})$
II) $q(Y) = a_{0,II} + l_2(X; a_{1,II}, \dots, a_{p,II})$

Dopo aver calcolato i parametri e se $\exists\ q^{-1}(\cdot)$ allora, calcolato σ_{II}^{*2}, utilizzando $\hat{Y} = q^{-1}[\hat{a}_{0,II} + l_2(X; \hat{a}_{1,II}, \dots, \hat{a}_{p,II})]$

$\Rightarrow$ possibile confronto : $\sigma_I^{*2} \gtreqless \sigma_{II}^{*2}$

Es. I) $Y = a_{0,I} + a_{1,I}\exp(X)$ II) $\ln(Y) = a_{0,II} + a_{1,II}X$

Per il modello II si dovrà calcolare la devianza residua utilizzando $\hat{Y} = exp[\hat{a}_{0,II} + \hat{a}_{1,II}X]$

−o−o−o−o−o−o−o−o−

d) I) $Y = a_{0,I} + l_1(X; a_{1,I}, \dots, a_{p,I})$
II) $Y = h\left(a_{0,II} + l_2(X; a_{1,II}, \dots, a_{p,II})\right)$

Se il modello (II) è linearizzabile allora, si linearizzi il modello (II), si calcolino i parametri; quindi, utilizzando $\hat{Y} = h[\hat{a}_{0,II} + l_2(X; \hat{a}_{1,II}, \dots, \hat{a}_{p,II})]$, si calcoli σ_{II}^{*2}

$\Rightarrow$ possibile confronto : $\sigma_I^{*2} \gtreqless \sigma_{II}^{*2}$

Es. I) $Y = a_{0,I} + a_{1,I}\exp(X)$ II) $Y = a_{0,II}\exp(a_{1,II}X)$

Per il modello II si dovrà linearizzare in

$$\ln(Y) = \ln(a_{0,II}) + a_{1,II}X$$

Quindi calcolati i parametri calcolare la devianza residua utilizzando

$\hat{Y} = \hat{a}_{0,II}\exp(\hat{a}_{1,II}X)$
dove
$\hat{a}_{0,II} = \exp(\hat{a}'_{0,II})$ e $a'_{0,II} = \ln(a_{0,II})$

−o−o−o−o−o−o−o−o−

e) I) $Y = l_1(X; a_{1,I}, \dots, a_{p,I})$

II) $q(Y) = h\left(l_2(X; a_{1,II}, \dots, a_{p,II})\right)$

Se il modello (II) è linearizzabile e se $\exists\ q^{-1}(\cdot)$ allora, si linearizzi il modello (II), si calcolino i parametri, quindi, utilizzando

$$\hat{Y} = q^{-1}\{h[l_2(X; \hat{a}_{1,II}, \dots, \hat{a}_{p,II})]\}$$

si calcoli $\sigma^{*\,2}_{II}$

$\Rightarrow$ possibile confronto : $\sigma^{*\,2}_{I} \gtreqless \sigma^{*\,2}_{II}$

Es. vedi infra

−o−o−o−o−o−o−o−o−

N.B. Si deduce che, in generale, se i due modelli non hanno la variabile dipendente espressa nella stessa forma funzionale o se uno dei due non possiede termine noto, allora il confronto potrà essere effettuato SOLO sulla base della varianza residua, opportunamente calcolata, di entrambi i modelli.

-o-o-o-o-o-o-o-o-o-o-

Dall'**esempio(a)**

Si consideri come (II) modello: $\sqrt{Y} = \ln(b_0 X^{b_1})$

Si deduce che

$$\sqrt{Y} = \ln(b_0) + b_1 \ln(X)$$

Si introducano le trasformazioni

$Y^* = \sqrt{Y}$ $\quad b_0^* = ln(\mathrm{b}_0)$ $\quad X^* = \ln(X)$

si ha

$Y^* = b_0^* + \mathrm{b}_1 X^*$

Utilizzando

X*= ln(X)	Y* = (Y)$^{0.5}$	n_{ij}
0	2.8284	2
0.6931	2.4495	2
0.6931	2.8284	1
1.0986	2.6458	1
1.0986	3.3166	1
1.0986	3.4641	1
1.6094	3.4641	1
1.6094	3.873	1

Si ottiene

$$\hat{\mathrm{b}}_1 = \frac{\mathrm{Cov}(X^*, Y^*)}{\mathrm{Var}(X^*)} = 0.564034$$

$$\hat{b}_0^* = \mathrm{M}(Y^*) - \hat{\mathrm{b}}_1 \mathrm{M}(X^*) = 2.530043$$

Il confronto (caso e) dell'elenco dei confronti) col modello $Y \cong \mathrm{a}_0 + \mathrm{a}_1 X$ **deve** essere fatto calcolando la varianza residua su un modello che abbia come variabile dipendente Y (e quindi non $(Y)^{0.5}$). Dal modello $(Y)^{0.5} = ln\{\mathrm{b}_0 X^{\mathrm{b}_1}\}$ si può scrivere

$$Y = [ln\{\mathrm{b}_0 X^{\mathrm{b}_1}\}]^2$$

Essendo $\hat{b}_0^* = \ln(\hat{b}_0)$ si ha che, sotto opportune condizioni (...),

$\hat{\mathrm{b}}_0 = exp(\hat{b}_0^*) = \exp(2.530043) = 12.55405$

I residui si possono quindi calcolare come $Y - [ln\{\hat{\mathrm{b}}_0 X^{\hat{\mathrm{b}}_1}\}]^2$ ovvero

X	Y	$\hat{Y} = [ln\{12.55405X^{0.564034}\}]^2$	$Y-\hat{Y}$	n_{ij}
1	8.0	6.40112	1.598881	2
2	6.0	8.53225	-2.53225	2
2	8.0	8.53225	-0.53225	1
3	7.0	9.92059	-2.92059	1
3	11.0	9.92059	1.079406	1
3	12.0	9.92059	2.079406	1
5	12.0	11.8186	0.181392	1
5	15.0	11.8186	3.181392	1

da cui si nota che

$M(\hat{Y})=9.1798 \neq M(Y)=8.88778$

Quindi $M(Y-\hat{Y})= 0.1202 \neq 0$

$\sigma^{*2}_{II} = \frac{1}{N}\Sigma(y_i-\hat{y}_i)^2 n_{ij} = 4.2393$

N.B. per definizione di devianza residua, il calcolo della (impropriamente detta) varianza residua avviene SENZA sottrarre il quadrato della media dei residui ovvero $[0.1202]^2$

Essendo $\sigma^{*2}_{I} = 3.270718 < \sigma^{*2}_{II} = 4.2393$

si deduce che il modello (I) si adatta meglio ai dati.

Connessione

"Esiste connessione quando NON esiste *indipendenza stocastica*"

quindi

"Se 2 o più variabili sono *indipendenti in senso stocastico* allora NON esiste connessione"

Def: X e Y sono *indipendenti in senso stocastico* se

$$n_{ij} = \hat{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{N} \quad \forall i,j$$

ovvero (equivalentemente) se

$$f_{ij} = \hat{f}_{ij} = f_{i.} \cdot f_{.j} \quad \forall i,j$$

Da $f_{ij} = \hat{f}_{ij} = f_{i.} \cdot f_{.j}$ si deduce che X e Y sono indipendenti in senso stocastico se

$$\frac{f_{ij}}{f_{i.}} = f_{.j} \quad \text{e} \quad \frac{f_{ij}}{f_{.j}} = f_{i.} \quad \forall i,j$$

ovvero se tutte le distribuzioni di frequenza condizionate *somigliano* alla distribuzione (relativa) della marginale

Esempio: 2 titoli di borsa classificati in base alle variazioni di segno registrate in 36 giorni di contrattazione

Frequenze assolute

X \ Y	–	+	$n_{i.}$
–	3	9	12
+	6	18	24
$n_{.j}$	9	27	36

Distribuzioni relative condizionate X|Y e Y|X

X | Y	–	+	$f_{i.}$		Y | X	–	+	$f_{.j}$
–	0.33	0.33	0.33		–	0.25	0.25	0.25
+	0.67	0.67	0.67		+	0.75	0.75	0.75

$\Rightarrow \exists$ indipendenza stocastica

-o-o-o-o-o-o-o-o-o-o-o-

Domanda: "Se NON esiste indipendenza stocastica, come misurare l'intensità della dipendenza funzionale?"

Indice di Pearson

$$\chi^2 = \sum_{j=1}^{k}\sum_{i=1}^{h}\frac{\left(n_{ij}-\hat{n}_{ij}\right)^2}{\hat{n}_{ij}} = \sum_{j=1}^{k}\sum_{i=1}^{h}\frac{\left(n_{ij}^2+\hat{n}_{ij}^2-2n_{ij}\hat{n}_{ij}\right)}{\hat{n}_{ij}}$$

$$= N\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{\left(n_{ij}^2+\left(\frac{n_{i.}\cdot n_{.j}}{N}\right)^2-2\left(\frac{n_{i.}\cdot n_{.j}}{N}\right)n_{ij}\right)}{n_{i.}\cdot n_{.j}} =$$

$$= N\sum_{j=1}^{k}\sum_{i=1}^{h}\left(\frac{n_{ij}^2}{n_{i.}n_{.j}}+\frac{n_{i.}\cdot n_{.j}}{N^2}-2\frac{n_{ij}}{N}\right) =$$

$$= N\left(\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{n_{ij}^2}{n_{i.}n_{.j}}+\frac{N\cdot N}{N^2}-2\frac{N}{N}\right) = N\left(\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{n_{ij}^2}{n_{i.}\cdot n_{.j}}-1\right) =$$

$$= N\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{\left(f_{ij}-\hat{f}_{ij}\right)^2}{\hat{f}_{ij}} = N\left(\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{f_{ij}^2}{f_{i.}\cdot f_{.j}}-1\right)$$

Massimo valore raggiungibile con l'*Indice di Pearson*

Considerando che

$$\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{n_{ij}^2}{n_{i.}\cdot n_{.j}} = \sum_{j=1}^{k}\sum_{i=1}^{h}\frac{n_{ij}\cdot n_{ij}}{n_{i.}\cdot n_{.j}} \le \sum_{j=1}^{k}\sum_{i=1}^{h}\frac{n_{ij}\cdot n_{i.}}{n_{i.}\cdot n_{.j}} = \sum_{j=1}^{k}\sum_{i=1}^{h}\frac{n_{ij}}{n_{.j}}$$

$$= \sum_{j=1}^{k}\frac{1}{n_{.j}}\sum_{i=1}^{h}n_{ij} = \sum_{j=1}^{k}1 = k$$

si ha

$$\chi^2 = N\left(\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{n_{ij}^2}{n_{i.}\cdot n_{.j}}-1\right) \le N(k-1)$$

Analogamente considerando che vale anche

$$\frac{n_{ij}^2}{n_{i.}\cdot n_{.j}} \le \frac{n_{ij}\cdot n_{.j}}{n_{i.}\cdot n_{.j}}$$

si ha

$$\chi^2 \le N(h-1)$$

In sintesi

$\max \chi^2 = N[\min\{(h-1),(k-1)\}]$

Indice di connessione χ^2 normalizzato

Indipendenza stocastica	$\Leftarrow 0 \leq \frac{\chi^2}{\max \chi^2} \leq 1 \Rightarrow$	Dipendenza funzionale

Esempio: Assumendo N=200

	Y		
X	D	E	$f_{i.}$
A	0,2	0,3	0,5
B	0,1	0	0,1
C	0	0,4	0,4
$f_{.j}$	0,3	0,7	1

$$\chi^2 = N\left(\sum_{j=1}^{k}\sum_{i=1}^{h}\frac{f_{ij}^2}{f_{i.}\cdot f_{.j}} - 1\right) =$$

$$= 200\left(\left[\frac{0,2^2}{0,5 \text{ x } 0,3} + \frac{0,3^2}{0,5 \text{ x } 0,7} + \frac{0,1^2}{0,1 \text{ x } 0,3} + \frac{0,4^2}{0,4 \text{ x } 0,7}\right] - 1\right) =$$

$$= 200(1,428671 - 1) = 85,72$$

$\max \chi^2 = 200[(2-1)] = 200$

$$\frac{\chi^2}{\max \chi^2} = \frac{85,72}{200} = 42,86\%$$

Esempio: confronto del segno della variazione percentuale di IntesaSanPaolo vs Generali al termine di 937 giornate di contrattazione

		IntesaSanPaolo		
		Negativo	Positivo	
Generali	Negativo	370	112	482
	Positivo	113	342	455
		483	454	937

Tabella con $\frac{n_{12}^2}{n_{1.}\cdot n_{.2}} = \frac{112^2}{454*482}$

		IntesaSanPaolo	
		Negativo	Positivo
Generali	Negativo	0,588043263	0,05732356
	Positivo	0,058102974	0,566219683

$\chi^2 = 937(1,2697 - 1) = 252,7089$

$\max \chi^2 = 937[(2-1)] = 937$

$$\frac{\chi^2}{\max \chi^2} = 26,97\%$$

-o-o-o-o-o-o-o-o-o-

Relazione tra indipendenza stocastica, indipendenza in media, incorrelazione

Q. Se esiste indipendenza stocastica, esiste dipendenza in media? NO

Si ricordi che si ha indipendenza in media se $\eta^2 = \frac{Var[M(Y|X)]}{Var(Y)} = 0$ ovvero se $Var[M(Y|X)]$=0.

Ma $Var[M(Y|X)] = \frac{1}{N}\sum_{i=1}^{h}\left(\mu_Y(x_i) - \mu_Y\right)^2 n_{i.}$

Quindi devo dimostrare che se esiste indipendenza stocastica $\mu_Y(x_i) = \mu_Y \; \forall i.$

Però $\mu_Y(x_i) = \frac{1}{n_{i.}}\sum_{j=1}^{k} y_j n_{ij} = \sum_{j=1}^{k} y_j \frac{f_{ij}}{f_{i.}}$ che nel caso di indip. stoc. è pari a $\mu_Y(x_i) = \sum_{j=1}^{k} y_j \frac{f_{i.}f_{.j}}{f_{i.}} = \sum_{j=1}^{k} y_j f_{.j} = \mu_Y$.

Q. Se esiste indipendenza stocastica, esiste correlazione (ovvero covarianza diversa da zero)? NO

Richiamando la formula della covarianza è noto che $M(XY) - M(X)M(Y) = Cov(X,Y)$. Quindi per dimostrare che $Cov(X,Y) = 0$ basta dimostrare che $M(XY) = M(X)M(Y)$ quando c'è indipendenza stocastica.
Se esiste indipendenza stocastica

$$M(XY) = \frac{1}{N}\sum_i\sum_j x_i y_j n_{ij} = \frac{1}{N}\sum_i\sum_j x_i y_j \frac{n_{i.}n_{.j}}{N}$$
$$= \frac{1}{N}\sum_i x_i n_{i.} \frac{1}{N}\sum_j y_j n_{.j} = M(X)M(Y)$$

e quindi $Cov(X,Y) = 0$ e quindi $\rho = 0$.

Q. Se esiste indipendenza in media, esiste correlazione (ovvero covarianza diversa da zero)? NO

Come detto sopra se esiste indipendenza in media $\mu_Y(x_i) = \mu_Y \; \forall i.$
Richiamando la formula della covarianza è noto che $M(XY) - M(X)M(Y) = Cov(X,Y)$. Quindi per dimostrare che $Cov(X,Y) = 0$ basta dimostrare che $M(XY) = M(X)M(Y)$ quando c'è indipendenza in media. Sapendo che

$$M(XY) = \frac{1}{N}\sum_i\sum_j x_i y_j n_{ij} = \frac{1}{N}\sum_i x_i n_{i.}\sum_j y_j \frac{n_{ij}}{n_{i.}}$$
$$= \frac{1}{N}\sum_i x_i n_{i.}\left(\mu_Y(x_i)\right)$$

se esiste indipendenza in media si avrà

$$M(XY) = \frac{1}{N}\sum_i x_i n_{i.}\left(\mu_Y\right) = \mu_Y\frac{1}{N}\sum_i x_i n_{i.} = \mu_Y\mu_X$$

da cui si deduce che in caso di indipendenza in media

$$M(XY) = M(X)M(Y)$$

e quindi $Cov(X,Y) = 0$ e quindi $\rho = 0$.

Quindi
"Se esiste indipendenza stocastica esiste indipendenza in media"
"Se esiste indipendenza stocastica, esiste incorrelazione"
"Se esiste indipendenza in media esiste incorrelazione"

Ovvero

indipendenza stocastica =>
indipendenza in media =>
incorrelazione

N.B.
Il viceversa NON è vero, ovvero

"Se esiste incorrelazione allora può esistere dipendenza in media"
"Se esiste indipendenza in media può esistere connessione !!!"