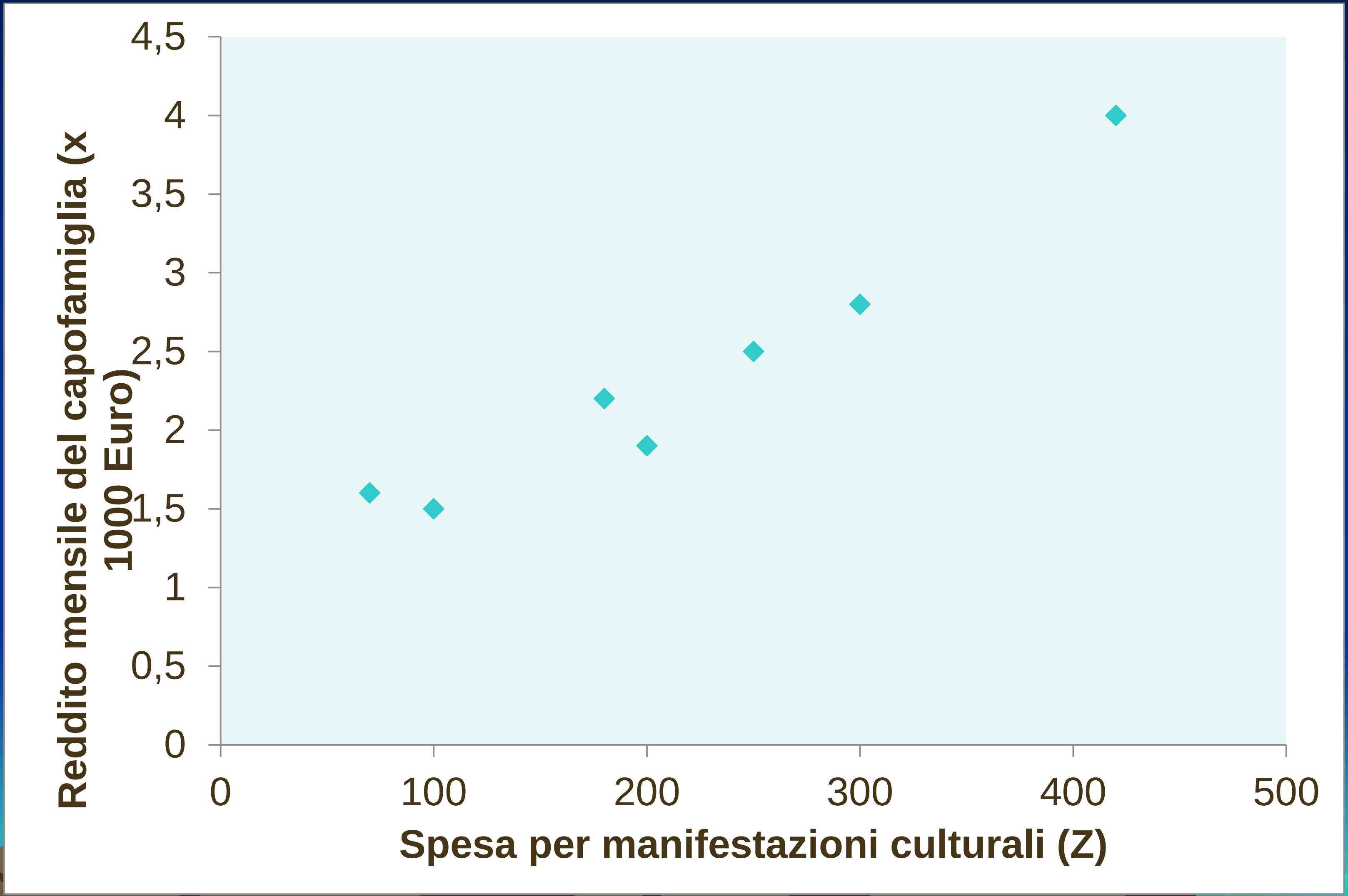
Soluzione esercizi seconda settimana

Es. 7 famiglie

	Spesa per manifestazioni culturali (Z)	Reddito mensile del capofamiglia (x 1000 Euro) (Y)
A	200	1,9
B	420	4,0
C	250	2,5
D	70	1,6
E	180	2,2
F	300	2,8
G	100	1,5

- Costruire il diagramma di dispersione
- Confrontare la variabilità di Z con quella di Y (utilizzando l'analisi robusta e non robusta)

Diagramma di dispersione



Confronto della variabilità

- Analisi non robusta: utilizzare CV e non le varianze

$$CV = \frac{\sigma}{M} \cdot 100$$

- VAR(Z) = 12306,122
- VAR(Y) = 0,637
- M(Z) = 217,143   M(Y)=2,357
- CV(Z) = $(12306{,}122)^{½}$ / 217,143=0,511
- CV(Y) = $(0{,}637)^{½}$ / 2,357=0,339

Confronto della variabilità

- Analisi robusta: utilizzare MAD relativi

- Me(Z) = 200 Me(Y)=2,2
- MAD(Z) = 100 MAD(Y)=0,6
- MAD’(Z) = 0,5 MAD’(Y) = 0,27

- La spesa per manifestazioni culturali è più variabile del reddito mensile

ESERCIZIO RIASSUNTIVO

Nella seguente tabella è riportata la distribuzione dei dipendenti di una grande azienda in base alla retribuzione lorda mensile:

Retribuzioni	Numero di dipendenti
1000 – 1200	30
1200 – 1500	130
1500 – 2000	150
2000 – 2500	50
2500 – 3500	30
3500 - 5000	20

I) Si calcoli la media e la mediana delle retribuzioni e le si commentino.

II) Si calcoli lo scostamento quadratico medio e il MAD delle retribuzioni e si commenti il significato dei risultati ottenuti.

III) Si dica, motivando la risposta, quale trasformazione subirebbero la media, la mediana, lo scostamento quadratico medio e il MAD delle retribuzioni, calcolati ai punti precedenti, se:

a)tutte le retribuzioni fossero aumentate di 50 euro,

b)tutte le retribuzioni fossero incrementate del 7%

c)tutte le retribuzioni fossero incrementate del 2% e, dopo questo aumento, aumentate di 100 euro .

IV) Si rappresenti graficamente la suddetta distribuzione e si dica quali informazioni si possono ricavare e si calcoli la moda.

V) Si calcoli l’indice di asimmetria di Fisher e lo si commenti

Retribuzioni	n_i	f_i	F(x)	$\lvert x_i - Me \rvert$
1000 – 1200 (1100)	30	0,073	0,073	550
1200 – 1500 (1350)	130	0,317	0,390	300
1500 – 2000 (1750)	150	0,366	0,756	100
2000 – 2500 (2250)	50	0,122	0,878	600
2500 – 3500 (3000)	30	0,073	0,951	1350
3500 – 5000 (4250)	20	0,049	1	2600

M = (1100·30 +…+ 4250·20) / 410 = 1850

$$Me = 1500 + \frac{500}{0,366}(0,5 - 0,39) = 1650$$

Scostamento quadratico medio:

$$\sigma = \sqrt{\frac{(1100-1850)^2 30 + \ldots + (4250-1850)^2 20}{410}} = 722,04$$

Retribuzioni	f_i	$\lvert x_i - Me\rvert$
1000 – 1200 (1100)	0,073	550
1200 – 1500 (1350)	0,317	300
1500 – 2000 (1750)	0,366	100
2000 – 2500 (2250)	0,122	600
2500 – 3500 (3000)	0,073	1350
3500 – 5000 (4250)	0,049	2600

$Me = 1650$

$\lvert x_i - Me\rvert$ ordinati	f_i	F(x)
100	0,366	0,366
300	0,317	0,683
550	0,073	0,756
600	0,122	0,878
1350	0,073	0,951
2600	0,049	1

MAD = 300

$y_i = a + b \cdot x_i$ $M(X)= 1850$ $Me(X)=1650$ $\sigma(X)=722{,}04$ $MAD(X)=300$

$M(Y)= a + b \cdot M(X)$ $Me(Y) = a + b \cdot Me(X)$

$\sigma(Y) = b \cdot \sigma(X)$ $MAD(Y) = b \cdot MAD(X)$

a) tutte le retribuzioni fossero aumentate di 50 euro,

$M = 1850 +50 = 1900$ $Me = 1650 + 50 = 1700$

$\sigma = 722{,}04$ $MAD = 300$

b) tutte le retribuzioni fossero incrementate del 7%

$M = 1850 \cdot 1{,}07 = 1979{,}5$ $Me = 1650 \cdot 1{,}07 = 1765{,}5$

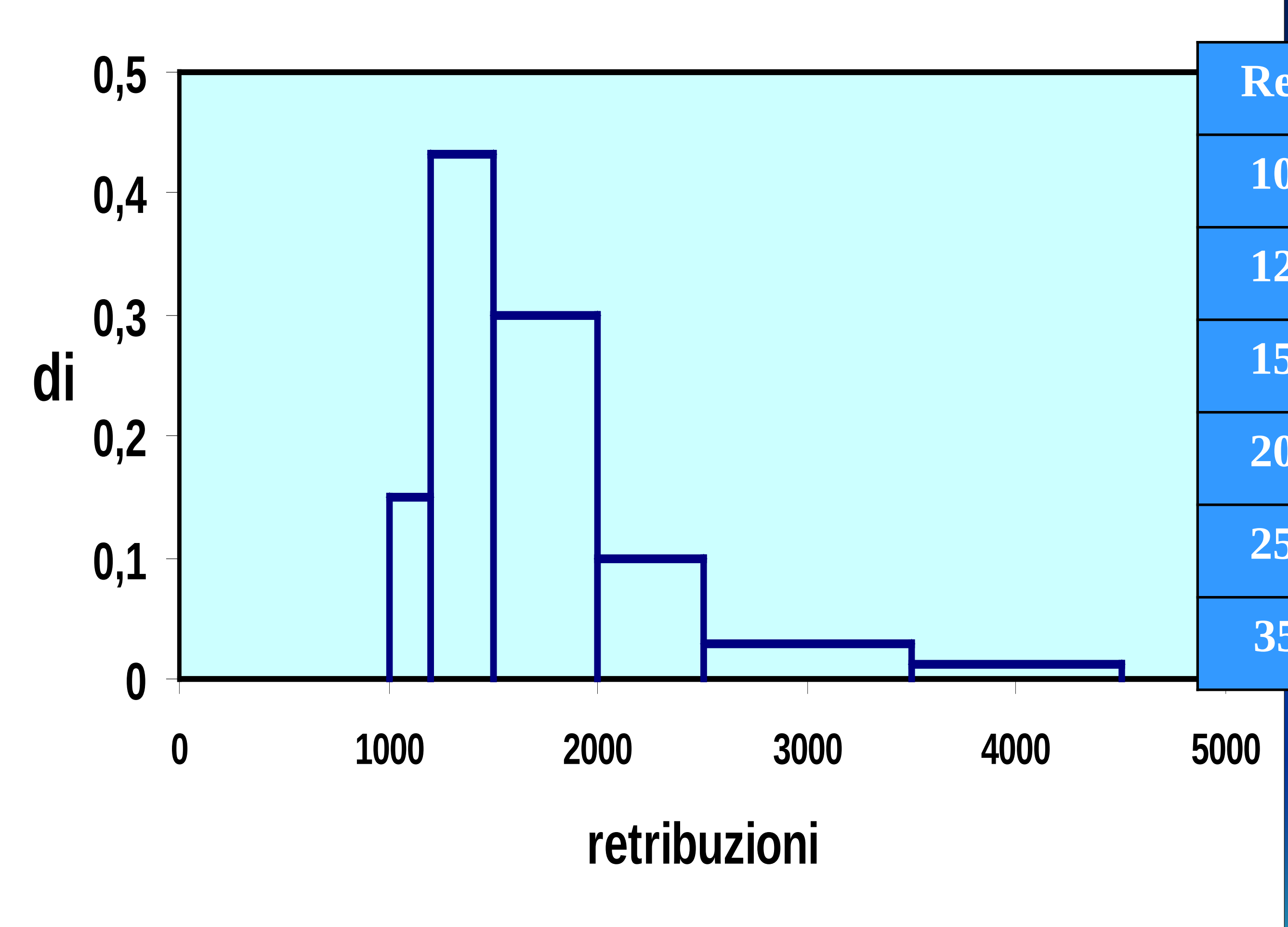
$\sigma = 722{,}04 \cdot 1{,}07 = 772{,}58$ $MAD = 300 \cdot 1{,}07 = 321$

c) tutte le retribuzioni fossero incrementate del 2% e, dopo questo aumento, aumentate di 100 euro.

$M = 1850 \cdot 1{,}02 + 100 = 1987$ $Me = 1650 \cdot 1{,}02 + 100 = 1783$

$\sigma = 722{,}04 \cdot 1{,}02 = 736{,}48$ $MAD = 300 \cdot 1{,}02 = 306$

Analisi della forma di distribuzione



Retribuzioni	d_i
1000 – 1200	0,15
1200 – 1500	0,433
1500 – 2000	0,30
2000 – 2500	0,10
2500 – 3500	0,03
3500 - 5000	0,013

Mo = 1350 €   Forma di distribuzione: asimmetria positiva

Soluzione: indice di asimmetria di Fisher

Retribuzioni	f_i	$(x_i - M)^3 f_i$
1000 – 1200 (1100)	0,073	-30.868.902,44
1200 – 1500 (1350)	0,317	-39.634.146,34
1500 – 2000 (1750)	0,366	-365.853,6585
2000 – 2500 (2250)	0,122	7.804.878,049
2500 – 3500 (3000)	0,073	111.283.536,6
3500 – 5000 (4250)	0,049	674.341.463,4
		722.560.975,6

M=1850

$$\gamma(\mathbf{X}) = \frac{\sum_{i=1}^{r}(x_i - M(\mathbf{X}))^3 f_i}{[\sigma(\mathbf{X})]^3} = \frac{722.560.975,6}{722,04^3} = 1,9195$$

➔ asimmetria positiva

Esercizio: 4 ipermercati. Sono riportati il fronte espositivo (in metri) relativo ai detergenti per la casa (X) ed il numero di referenze (Y)

Ipermercato	X	Y
A	48	178
B	70	222
C	80	192
D	70	159

Si determini il fronte espositivo Z per referenza di ciascun supermercato

Spiegando il significato dei simboli utilizzati, si scrivano le espressioni della media aritmetica e dello scostamento quadratico medio del fronte espositivo per referenza e si effettuino i relativi calcoli

Calcolo del fronte espositivo per referenza (Z)

Ipermercato	X	Y	Z=X/Y
A	48	178	0,27
B	70	222	0,32
C	80	192	0,42
D	70	159	0,44

Scrivere l'espressione della media aritmetica

Media aritmetica del fronte espositivo per referenza (z)

$$z_i = \frac{x_i}{y_i}$$

$$M(z) = \frac{\sum_{i=1}^{r} z_i w_i}{\sum_{i=1}^{r} w_i}$$

Come variabile peso si utilizza il denominatore del rapporto (y numero di referenze)

$$M(z) = \frac{\sum_{i=1}^{r} (x_i / y_i) y_i}{\sum_{i=1}^{r} y_i} = \frac{\sum_{i=1}^{r} x_i}{\sum_{i=1}^{r} y_i}$$

Espressione della media del fronte espositivo per referenza

Ipermercato	X	Y	Z=X/Y
A	48	178	0,27
B	70	222	0,32
C	80	192	0,42
D	70	159	0,44

$$M(Z) = \frac{\sum_{i=1}^{4} z_i y_i}{\sum_{i=1}^{4} y_i} = \frac{\sum_{i=1}^{4} x_i}{\sum_{i=1}^{4} y_i}$$

Calcolo della media del fronte espositivo per referenza

Ipermercato	X	Y	Z=X/Y
A	48	178	0,27
B	70	222	0,32
C	80	192	0,42
D	70	159	0,44

$$M(Z) = \frac{\sum_{i=1}^{4} z_i y_i}{\sum_{i=1}^{4} y_i} = \frac{\sum_{i=1}^{4} x_i}{\sum_{i=1}^{4} y_i} = \frac{48+70+80+70}{178+222+192+159} = 0.357$$

Espressione dello scostamento quadratico medio del fronte espositivo per referenza

Ipermercato	X	Y	Z=X/Y
A	48	178	0,27
B	70	222	0,32
C	80	192	0,42
D	70	159	0,44

$$\sigma(Z) = \sqrt{\frac{\sum_{i=1}^{4}[z_i - M(Z)]^2 y_i}{\sum_{i=1}^{4} y_i}}$$

Calcolo dello scostamento quadratico medio del fronte espositivo per referenza

Ipermercato	X	Y	Z=X/Y
A	48	178	0,27
B	70	222	0,32
C	80	192	0,42
	70	159	0,44

$$\sigma(Z) = \sqrt{\frac{\sum_{i=1}^{4}[z_i - M(Z)]^2 y_i}{\sum_{i=1}^{4} y_i}} =$$

$$\sigma(Z) = \sqrt{\frac{(0{,}27 - 0{,}357)^2 178 + \cdots + (0{,}44 - 0{,}357)^2 159}{751}} = 0.069$$

Esercizio: Rendimenti, in quintali per ettaro, d’una certa varietà di frumento

Rendimenti, in quintali per ettaro, d’una certa varietà di frumento, per 100 appezzamenti di terreno

Rendimenti	n. appezzamenti
54-58	16
58-62	23
62-66	25
66-70	21
70-74	15
	100

- Quartili?
- Significato del terzo quartile
- Box-plot
- Si illustrino le informazioni traibili dalla suddetta rappresentazione

Soluzione: calcolo quartili

Rendimenti	n. appezzamenti	f_i	F_i
54-58	16	0,16	0,16
58-62	23	0,23	0,39
62-66	25	0,25	0,64
66-70	21	0,21	0,85
70-74	15	0,15	1
	100	1	

$$x_{0,25} = 58 + \frac{4}{0,23}(0,25 - 0,16) = 59,56$$

$$x_{0,50} = 62 + \frac{4}{0,25}(0,5 - 0,39) = 63,76$$

$$x_{0,75} = 66 + \frac{4}{0,21}(0,75 - 0,64) = 68,09$$

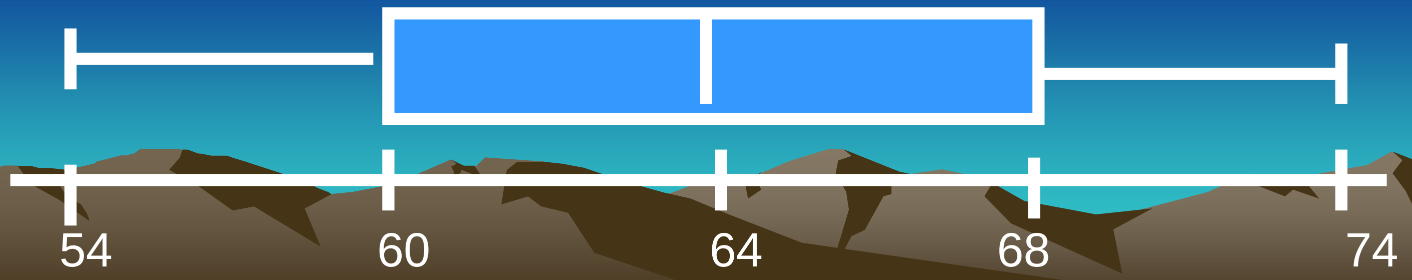
Soluzione: costruzione del boxplot

- Dopo aver trovato i quartili occorre calcolare i punti di troncamento

Boxplot

- PT_{inf} = 59,56- 1,5(68,09-59,56)=46,765
- Dato che 46,765 < x_{min} → $PT_{inf} = x_{min}$= 54

- PT_{sup} = 68,09 + 1,5(68,09-59,56)=80,885
- Dato che 80,885 > x_{max} → $PT_{sup} = x_{max}$=74



Distribuzione quasi simmetrica, nessun outlier

ESERCIZIO: Var. % rispetto all’anno precedente dei prezzi di due beni

Anno	Bene A	Bene B
1997	–	–
1998	+2,9%	–1,0%
1999	+3,5%	+3,6%
2000	+1,8%	+4,5%
2001	–0,6%	+2,5%

Calcolare:

- N.I. dei prezzi con base 1997=100
- N.I. composti dei prezzi con base 1997=100
- peso A = 20%; peso B = 80%
- Variazione complessiva dal 1997 al 2001 e tasso medio annuo dei N.I. composti

Ricostruzione NI a base mobile = **100 + var. %**

Anno	Var % Bene A	Var % Bene B	N.I base mobile Bene A	N.I base mobile Bene B
1997	–	–	–	–
1998	**+2,9%**	**–1,0%**	**102,9**	**99,0**
1999	**+3,5%**	**+3,6%**	**103,5**	**103,6**
2000	**+1,8%**	**+4,5%**	**101,8**	**104,5**
2001	**–0,6%**	**+2,5%**	**99,4**	**102,5**

Ricostruzione NI a base fissa = **relazione tra N.I. a base fissa e base mobile**

Anno	N.I base mobile Bene A	NI base mobile N.I Bene B	NI base fissa Bene A	NI base fissa Bene B
1997	–	–	100	100
1998	102,9	99,0	102,9	99,0
1999	103,5	103,6	106,5	102,6
2000	101,8	104,5	108,4	107,2
2001	99,4	102,5	107,8	109,9

Interpretazione

N.I. composti (base 1997) $\Rightarrow$ media ponderata N.I. semplici a base fissa:

Anno	NI base fissa Bene A	NI base fissa Bene B	NI composti
1997	100	100	100
1998	102,9	99,0	102,9*0,2+99,0*0,8=99,8
1999	106,5	102,6	106,5*0,2+102,6*0,8=103,4
2000	108,4	107,2	107,4
2001	107,8	109,9	109,5

Interpretazione

$$\sqrt[4]{1,095}-1=0,023$$

Variazione complessiva N.I. composti (base 1997)

Anno	NI composti
1997	**100**
1998	**99,8**
1999	**103,4**
2000	**107,4**
2001	**109,5**

Variazione complessiva
109,5/100 = 1,095 ⇒ + 9,5%

Tasso medio annuo di variazione:

$$\sqrt[4]{1,095}-1=0,023$$

⇒ + 2,3%

(In alternativa: calcolo in funzione dei N.I. a base mobile)

CORRELAZIONE FRA DUE S.S.

- Esempio: X = numero di extracomunitari iscritti al collocamento, Y = numero di discount
- Calcolare e r_{XY} tra le variabili originarie, i NI a base fissa, le variazioni percentuali a base fissa, i NI a base mobile, le variazioni percentuali a base mobile

Correlazione *spuria* $\Rightarrow$ relazione tra i livelli

Anni	X	Y
1993	72.644	600
1994	85.993	1.300
1995	96.287	1.930
1996	136.942	2.328
1997	140.100	2.523

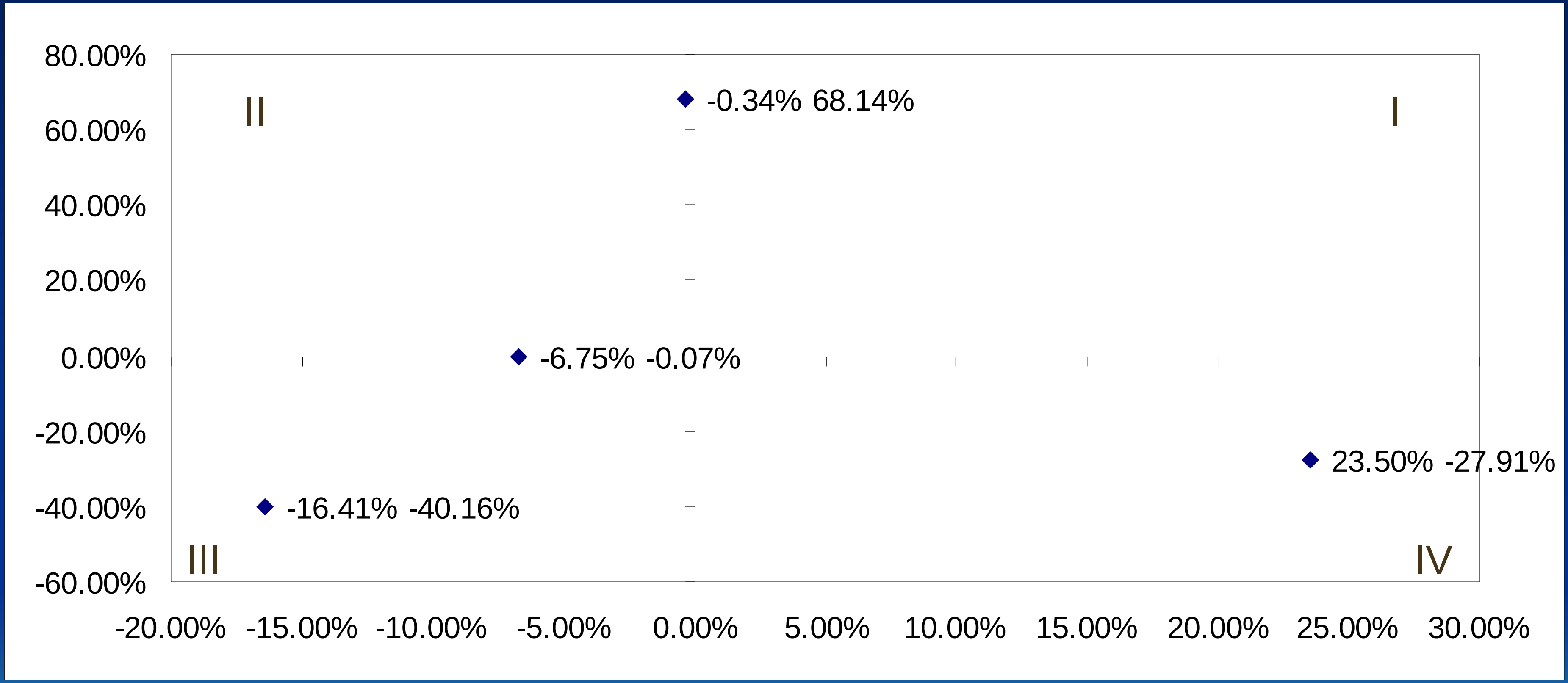
$$r_{xy} = \frac{COV(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{17.977.023,36}{(27.300,88 \cdot 705,42)} = 0,933$$

NI base mobile X (numero di extracomunitari) e Y (numero di discount

Anni	n. i. base mobile	n. i. base mobile	Var % X	Var % Y	Scost media X	Scost media Y
1993	-	-				
1994	118,38	216,67	18,38	116,67	-0,34	68,14
1995	111,97	148,46	11,97	48,46	-6,75	-0,07
1996	142,22	120,62	42,22	20,62	23,50	-27,91
1997	102,31	108,38	2,31	8,38	-16,41	-40,16
Media	118,72	148,53	18,72	48,53	0,00	0,00
Var	0,0217	0,1758	0,0217	0,1758	Cov(Nix,NIy)=-0,000496	

r_{xy}(tra n. i. a base mobile) =-0,000496/(0,0217*0,1758)^{½} = -0,008

Scatter sugli scostamenti NI base mobile o var. percentuali



Osservazioni finali

- **Non esiste relazione lineare tra le variazioni annue di X e Y**
- **Si ottiene r_{xy} = -0,008 anche effettuando il calcolo sulle *variazioni %* rispetto all'anno precedente (proprietà di invarianza per trasformazioni lineari crescenti)**

X = PREZZI (in euro)
Y = QUANTITA' VENDUTE (in n. di pezzi)

- Calcolare rxy sui dati originali, sui NI a base fissa e sulle variazioni percentuali. Commentare i risultati

Anni	X	Y	v(X) %	v(Y)%
1997	1,50	200	-	-
1998	1,68	208	+12	+4
1999	1,78	229	+6	+10,1
2000	1,96	243	+10,1	+6,1
2001	2,25	245	+14,8	+0,8
2002	2,43	265	+8,0	+8,2
2003	2,60	288	+7,0	+8,7

Coefficiente di correlazione

- Calcolato sui livelli
- $r_{xy} = 0{,}97$

- Calcolato sulle variazioni percentuali
- $r_{v(x)v(y)} = -0{,}998$

Lucidi regressione

ESEMPIO (7 supermercati) r_{xy}=0,96

	N. dipendenti (X)	Fatturato in milioni di € (Y)
A	10	1,9
B	18	3,1
C	20	3,2
D	8	1,5
E	30	6,2
F	12	2,8
G	14	2,3
Medie	16	3

Calcolo di a e b

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
A	10	1,9	100	3,61	19
B	18	3,1	324	9,61	55,8
C	20	3,2	400	10,24	64
D	8	1,5	…	…	…
E	30	6,2	…	…	…
F	12	2,8	…	…	…
G	14	2,3	…	…	…
Tot.	112	21	2128	77,28	402,6

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$a = \frac{21 \cdot 2.128 - 112 \cdot 402,6}{7 \cdot 2.128 - 112^2} = -\frac{403,2}{2.352} = -0,17$$

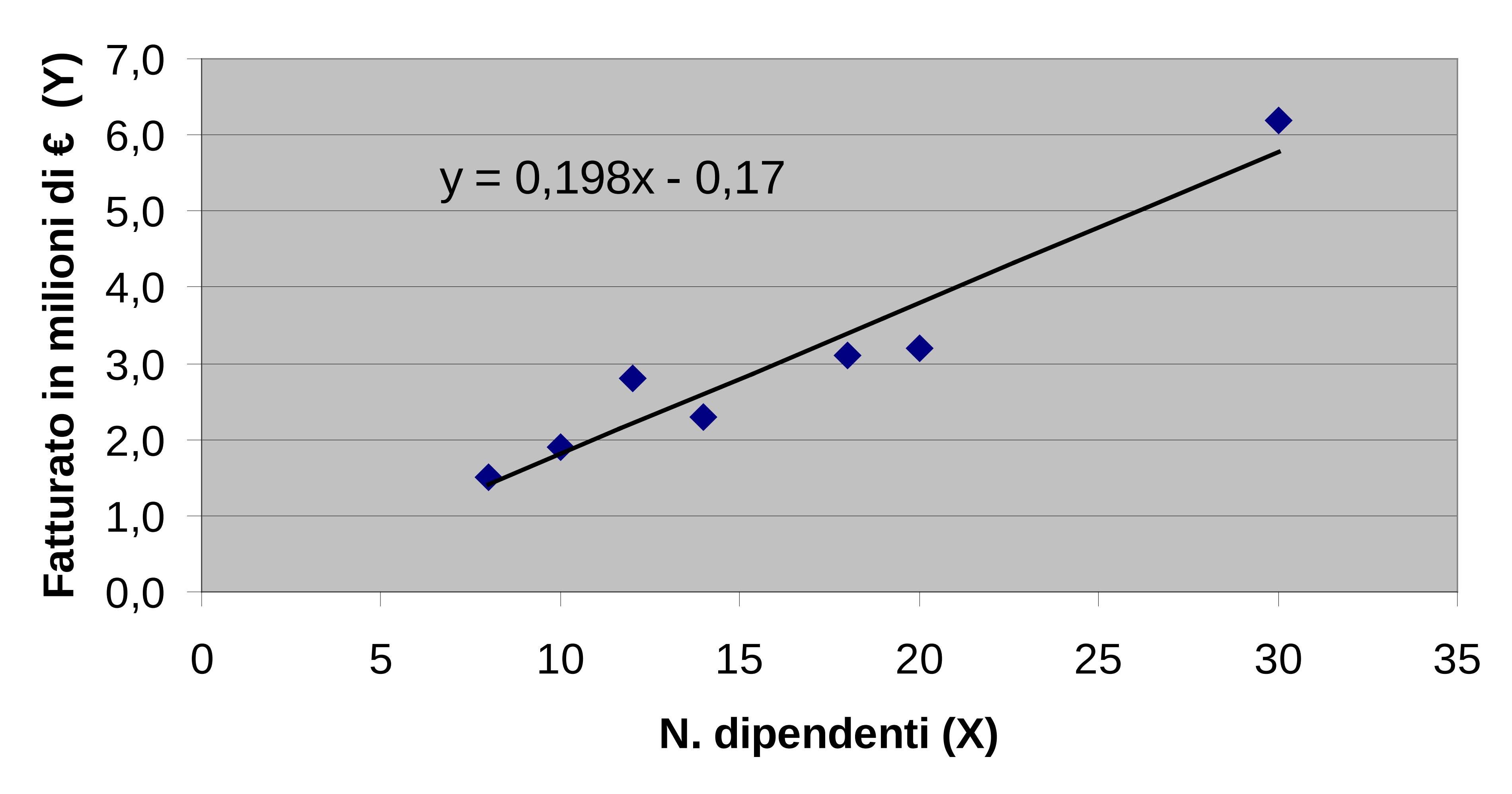
Calcolo di a e b

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
A	10	1,9	100	3,61	19
B	18	3,1	324	9,61	55,8
C	20	3,2	400	10,24	64
D	8	1,5	…	…	…
E	30	6,2	…	…	…
F	12	2,8	…	…	…
G	14	2,3	…	…	…
Tot.	112	21	2128	77,28	402,6

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{7 \cdot 402{,}6 - 112 \cdot 21}{7 \cdot 2.128 - 112^2} = \frac{466{,}2}{2.352} = 0{,}198$$

Scatter con retta di regressione



Interpretazione dei parametri ESEMPIO (7 supermercati)

- $a = -0{,}17 \rightarrow$ fatturato teorico quando N. di dipendenti $= 0$
- $b = 0{,}198 \rightarrow$ incremento medio nel fatturato quando il numero di dipendenti aumenta di 1 unità

Calcolo dei valori teorici e dei residui

y_i=-0,17+0,198x_i

	x_i	y_i	Valori teorici	Residui	$x_i \times residuo_i$
A	**10**	**1,9**	-0,17+0,198*10=1,81	0,09	0,89
B	**18**	**3,1**	-0,17+0,198*18=3,40	-0,30	-5,34
C	**20**	**3,2**	-0,17+0,198*20= 3,79	-0,59	-11,86
D	**8**	**1,5**	1,41	0,09	0,69
E	**30**	**6,2**	5,78	0,43	12,75
F	**12**	**2,8**	2,21	0,59	7,11
G	**14**	**2,3**	2,60	-0,30	-4,25
Tot.	**112**	**21**	**21**	**0**	**0**

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

$$\sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} x_i e_i = 0$$

Modi alternativi di esprimere b (p. 229)

$$b = \frac{COV(X,Y)}{VAR(X)} = r_{xy}\frac{\sigma_y}{\sigma_x}$$

- Dimostrazione

$$b = \frac{COV(\mathbf{X},\mathbf{Y})}{\sqrt{VAR(\mathbf{X})VAR(\mathbf{Y})}}\frac{\sigma(\mathbf{Y})}{\sigma(\mathbf{X})}$$

ESEMPIO (7 supermercati):

$r_{xy} = 0{,}961$ $\sigma_x = 6{,}928$ $\sigma_y = 1{,}428$

$$b = \frac{COV(X,Y)}{VAR(X)} = r_{xy}\frac{\sigma_y}{\sigma_x}$$ $$b = 0{,}961 \cdot \frac{1{,}428}{6{,}928} = 0{,}198$$

$M_y = 3 \quad M_x = 16$ $$a = M_y - bM_x$$

$$a = 3 - 0{,}198 \cdot 16 = -0{,}17$$

BONTA' DI ADATTAMENTO

- Occorre analizzare i residui $e_i = (y_i - \hat{y}_i)$

DEVIANZA RESIDUA

$$DEV(E) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

- **L'adattamento è buono quando *DEV*(*E*) è "piccola"**
- **Problemi:**
- ***DEV*(*E*) cresce all'aumentare del *numero di osservazioni* (*n*)**
- ***DEV*(*E*) dipende dall'*unità di misura* e dall'*ordine di grandezza* di *Y***

In qualsiasi modello di regressione con o senza intercetta è valida la relazione che segue

$$\sum_{i=1}^{n} y_i^{\,2} = \sum_{i=1}^{n} \hat{y}_i^{\,2} + \sum_{i=1}^{n} e_i^{\,2}$$

- **Questa relazione sfrutta la terza proprietà delle stime dei minimi quadrati (vincolo della derivata parziale rispetto a b posta uguale a 0)**

$$\sum_{i=1}^{n} x_i (y_i - \hat{y}_i) = 0$$

Esempio supermercati (continua)

$y_i=-0,17+0,198x_i$

	x_i	y_i	Valori teorici	Residui	$X_i \times residuo_i$	y_i^2	(Valori teorici)^2	residui^2
A	**10**	**1,9**	1,81	0,09	0,89	3.61	3.279	0.008
B	**18**	**3,1**	3,40	-0,30	-5,34	9.61	11.536	0.088
C	**20**	**3,2**	3,79	-0,59	-11,86	10.24	14.386	0.351
D	**8**	**1,5**	1,41	0,09	0,69	2.25	2.000	0.007
E	**30**	**6,2**	5,78	0,43	12,75	38.44	33.351	0.181
F	**12**	**2,8**	2,21	0,59	7,11	7.84	4.871	0.351
G	**14**	**2,3**	2,60	-0,30	-4,25	5.29	6.779	0.092
Tot.	**112**	**21**	**21**	**0**	**0**	77.28	76.201	1.079

$$\sum_{i=1}^{n} y_i^{\,2} = \sum_{i=1}^{n} \hat{y}_i^{\,2} + \sum_{i=1}^{n} e_i^{\,2}$$

77.28=76.201+1.079

BONTA' DI ADATTAMENTO

- Retta di regressione: $\hat{y}_i = a + bx_i$

DEVIANZA TOTALE

$$DEV(Y) = \sum_{i=1}^{n} (y_i - M_y)^2$$

DEVIANZA DI REGRESSIONE

$$DEV(\hat{Y}) = \sum_{i=1}^{n} (\hat{y}_i - M_y)^2$$

DEVIANZA RESIDUA

$$DEV(E) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

Esempio 7 supermercati (continua)

$\hat{y}_1 = -0{,}17 + 0{,}198 * 10$ **Calcolo di R^2 (δ)**

	x_i	y_i	$\hat{y}_i$	e_i^2	$(\hat{y}_i - M_y)^2$
A	10	1,9	1,81	0.008	1,416
B	18	3,1	3,394	0.088	0,155
C	20	3,2	3,79	0.351	0,624
D	8	1,5	1,414	0.007	…
E	30	6,2	5,77	0.181	…
F	12	2,8	2,206	0.351	…
G	14	2,3	2,602	0.092	…
Tot.	112	21	21	1,079	13,201

- *DEV(Y)* = $7 \cdot (1{,}428)^2 = 14{,}28$

$M_y = 3$

$Dev_{TOT} = Dev_{REGR} + Dev_{RES}$

$14{,}28 = 13{,}201 + 1{,}079$

$$\delta = \frac{13{,}201}{14{,}28} = 1 - \frac{1{,}079}{14{,}28} = 0{,}924$$