



La regressione bivariata

Regressione

Scopo dell'analisi della regressione

La regressione esamina la relazione lineare tra una o più variabili esplicative (o var. indipendenti o "predittori") e una variabile criterio (o var. dipendente)

Si ipotizza che la var. indipendente "determini", "influenzi" o "predica" la var. dipendente

Regressione bivariata (o semplice)

Una variabile indipendente e una variabile dipendente

Regressione multipla

Due o più variabili indipendenti e una variabile dipendente

Regressione

Punto di partenza: Matrice delle varianze e covarianze

	mediaVoti	QI	Mot_Intr	Cosc
mediaVoti	,667			
QI	,246	,624		
Mot_Intr	,169	,298	,961	
Cosc	,276	,218	,189	,863

sotto la diagonale: le
covarianze tra le variabili (es.
tra «mediaVoti» e «Cosc»)

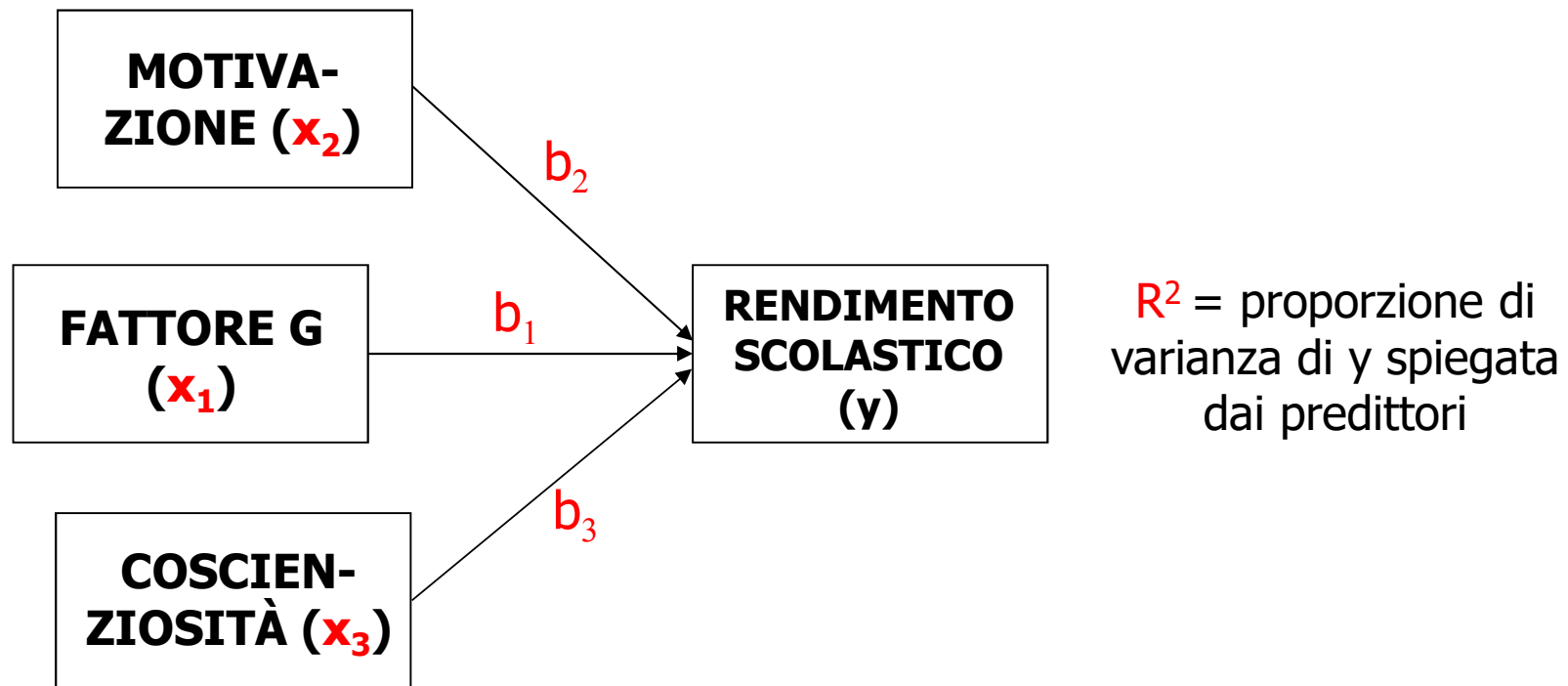
$$\frac{\sum (x - \bar{x})(y - \bar{y})}{n - 1}$$

sulla diagonale
principale: le varianze
delle variabili

$$\frac{\sum (y - \bar{y})^2}{n - 1}$$

Regressione

Punto di arrivo: Stima dei parametri



Regressione

Regressione bivariata (o semplice)

Dal punto di vista matematico:

L'equazione che lega y a x (equazione di regressione) è:

$$y' = a + bx$$

Variabili:

y' → valore della var. dipendente previsto in base all'equazione di regressione

x → variabile indipendente

Parametri

a → valore atteso per la var. dipendente quando $x = \text{zero}$

b → impatto che x ha su y (di quante unità cambia y quando x aumenta di una unità)

Regressione

Regressione bivariata (o semplice)

In genere, tuttavia, il valore previsto della var. dipendente (y') non coincide con il valore osservato (y)

Dobbiamo introdurre un termine di errore o residuo (e), che rende conto dell'errore che si commette nel predire y da x

$$y = \boxed{a + bx} + e$$

↑
 y'

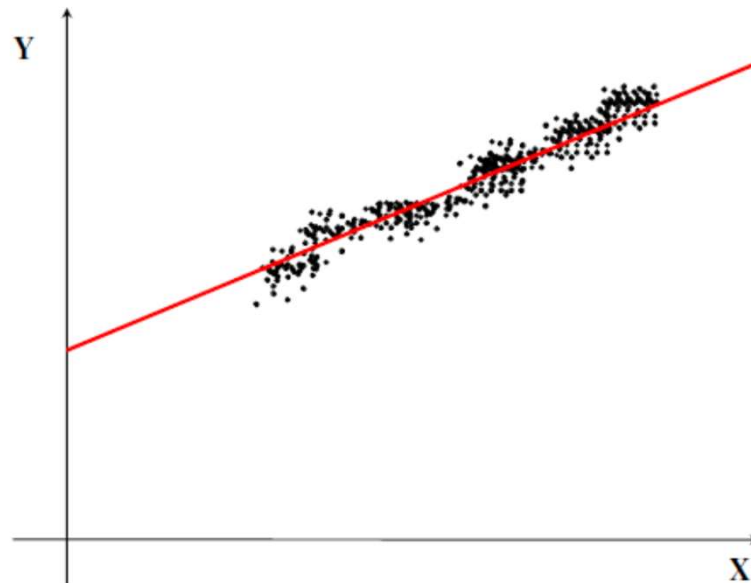
$$e = y' - y$$

Regressione

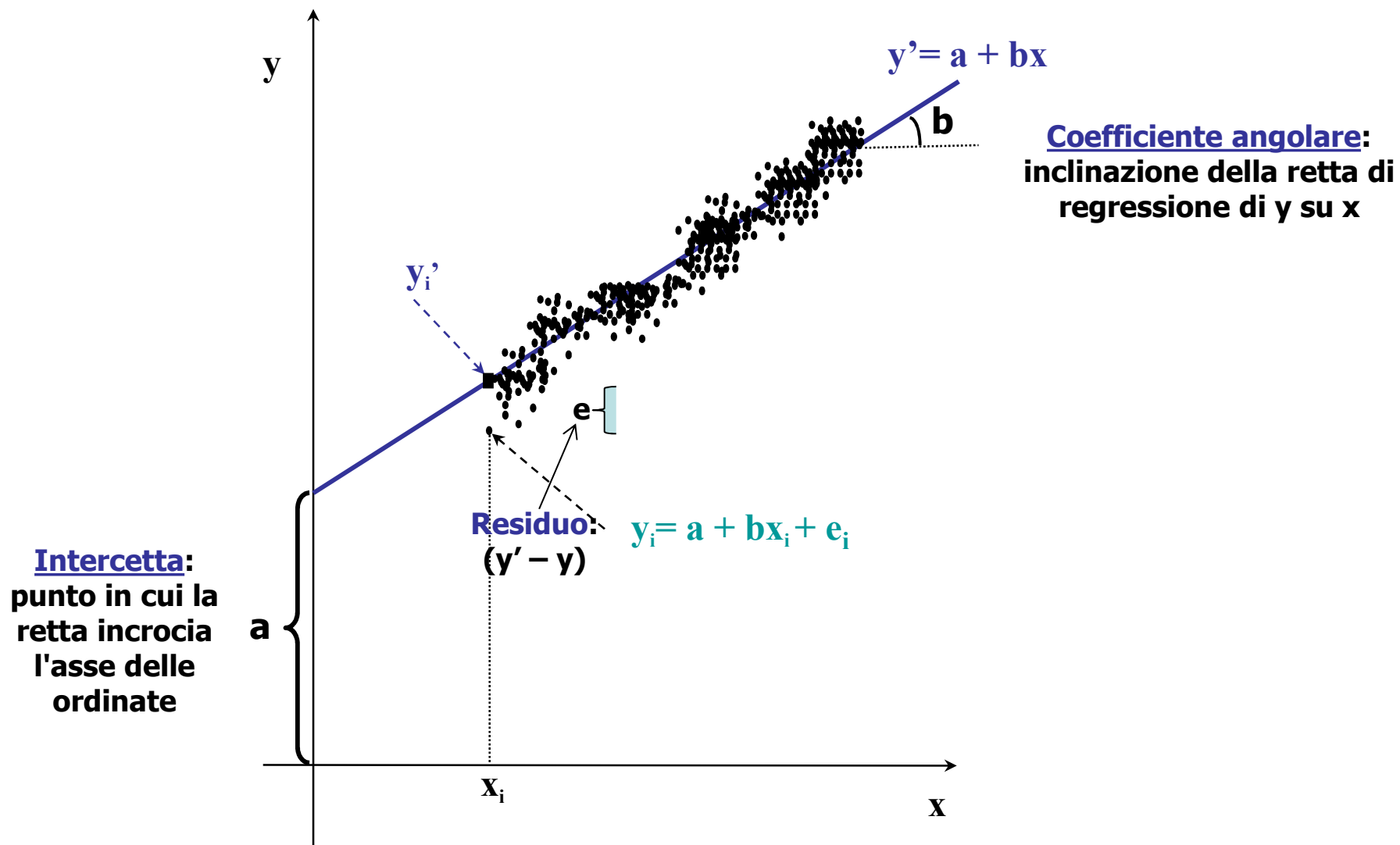
Regressione bivariata (o semplice)

Dal punto di vista grafico:

Si tratta di individuare la retta che consente di predire al meglio i punteggi nella var. dipendente a partire da quelli nella var. indipendente (è la retta che “interpola” meglio la nuvola di punti della distribuzione congiunta delle due variabili)



Regressione



Regressione

Formule dei minimi quadrati per il calcolo di a e b:

**La retta di regressione deve essere identificata:
occorre calcolare a e b**

**Ciò viene fatto con il metodo dei minimi quadrati
("least squares")**

Equazione dei minimi quadrati:

$$\Sigma(y_i - y_i')^2 = \Sigma(y_i - (a + bx_i))^2 = \min$$

**Identifica la retta che riduce al minimo l'errore
che viene commesso nello stimare y da x**

Regressione

Formule dei minimi quadrati per il calcolo di a e b:

$$a = \bar{y} - b\bar{x}$$

il coefficiente "a" rappresenta il valore previsto di y quando x è uguale a 0

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = \frac{\text{cov}(X, Y)}{\text{var}(X)}$$

Il coefficiente "b" rappresenta il cambiamento atteso in y associato all'aumento di una unità in x

La regressione bivariata

Esercizio 1

Participant	X Social Support	Y Life Satisfaction
1	4	6
2	6	5
3	8	7
4	10	9
5	12	9

$$\bar{X} = 8.0$$

$$\bar{Y} = 7.2$$

$$S_X = 3.16$$

$$S_Y = 1.79$$

$$\text{cov}(X, Y) = 5.00 \quad r_{xy} = .884$$

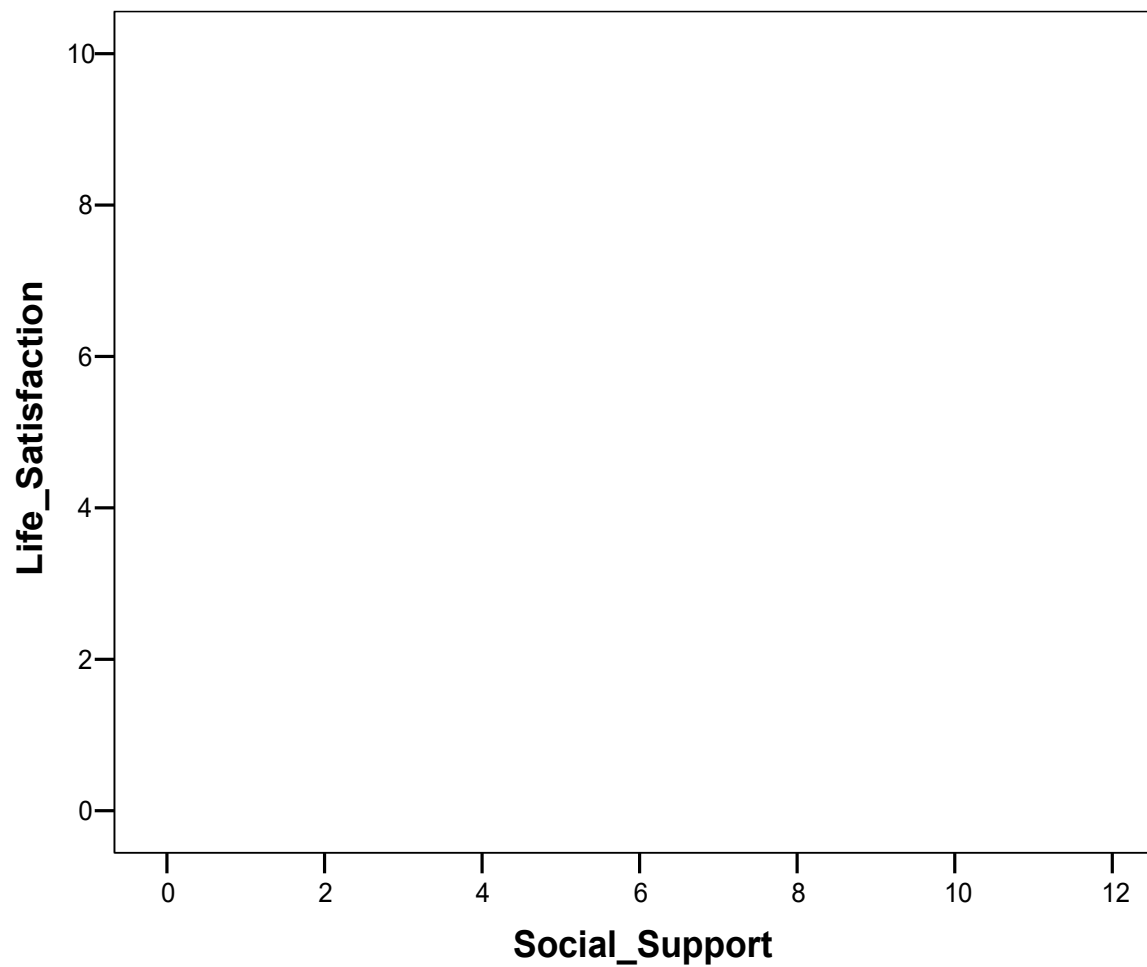
Calcola il coefficiente b: _____

Calcola il coefficiente a: _____

Se un soggetto ha un punteggio pari a 9 nel supporto sociale, il suo punteggio nella soddisfazione di vita previsto in base all'equazione di regressione (Y') è pari a: _____

La regressione bivariata

Esercizio 2



traccia la retta di regressione ...

Stime standardizzate

Il coefficiente di regressione esprime la relazione tra Y e X nell'unità di misura delle 2 variabili

In alcuni casi è preferibile esprimere questa relazione in una scala di misura immediatamente comprensibile. A tal fine si deve standardizzare il coefficiente di regressione

Il coefficiente di regressione standardizzato viene indicato con il termine "peso beta" o più semplicemente β

Stime standardizzate

Il coefficiente standardizzato si ottiene moltiplicando il coefficiente "grezzo" (non standardizzato) per il rapporto delle deviazioni standard della var. indipendente e della var. dipendente:

$$\beta = b (s_x/s_y)$$

Come si interpreta? Indica «di quante deviazioni standard» y aumenta all'aumentare di una deviazione standard in x

Nella regressione bivariata è uguale al coefficiente di correlazione, ovvero: $\beta = r_{yx}$

Regressione

La regressione bivariata in SPSS

The screenshot displays the SPSS software interface. The 'Analyze' menu is open, and the 'Regression' option is selected, leading to the 'Lineare...' (Linear...) dialog box. The background shows a data table with the following columns: 'anno', 'classe', 'genere', 'sto', 'fil', 'mat', 'fis', 'sci', 'edf', 'comp', 'inf', 'art', and 'rel'. The 'rel' column contains categorical values: Buono, Esonerato, Distinto, and Ottimo.

	anno	classe	genere	sto	fil	mat	fis	sci	edf	comp	inf	art	rel		
1	1998	4I		6	6	7	7	6	8	10	7	8	Buono		
2	1998	4F		6	7	6	6	6	7	9	6	7	Esonerato		
3	1998	4G		7	7	6	7	7	10	10	.	8	Distinto		
4	1999	4F		6	7	6	6	7	7	9	7	7	Distinto		
5	1998	4I					7	6	7	10	6	7	Esonerato		
6	1999	4L					6	7	8	10	6	8	Esonerato		
7	1998	4H					8	8	9	10	8	7	Esonerato		
8	1998	4I					8	8	8	9	6	7	Buono		
9	1998	4G					7	6	8	10	.	7	Buono		
10	1998	4D					6	6	8	9	6	6	Esonerato		
11	1998	4B					6	7	9	9	.	8	Esonerato		
12	1998	4A					8	7	9	10	.	7	Esonerato		
13	1998	4C					6	6	9	9	7	8	Esonerato		
14	1998	4C					6	6	9	9	7	7	Esonerato		
15	1998	4M					7	6	10	9	6	8	Ottimo		
16	1997	4C					6	5	9	7	7	6	Esonerato		
17	1997	4C					5	6	7	7	6	6	Buono		
18	1997	4I					7	7	9	8	6	8	Esonerato		
19	1998	4C					7	6	9	9	8	7	Distinto		
20	1998	4H	1	33	6	6	6	6	9	8	6	6	Buono		
21	1998	4H	1	43	6	4	6	6	6	9	9	6	7	Esonerato	
22	1998	4I	1	7	6	7	6	6	7	9	9	6	9	Esonerato	
23	1998	4L	1	12	6	7	7	6	7	8	10	6	9	Buono	
24	1997	4C	1	27	8	6	7	7	6	7	9	10	8	7	Distinto
25	1998	4D	1	8	5	7	7	6	6	7	9	9	6	7	Ottimo
26	1998	4B	2	31	9	7	9	9	9	9	10	.	9	Buono	
27	1997	4D	2	41	6	4	7	7	6	5	7	9	6	7	Esonerato

Si può calcolare dalla finestra di dialogo «Lineare»
(selezionabile dal menu «Analizza», procedura «Regressione»)

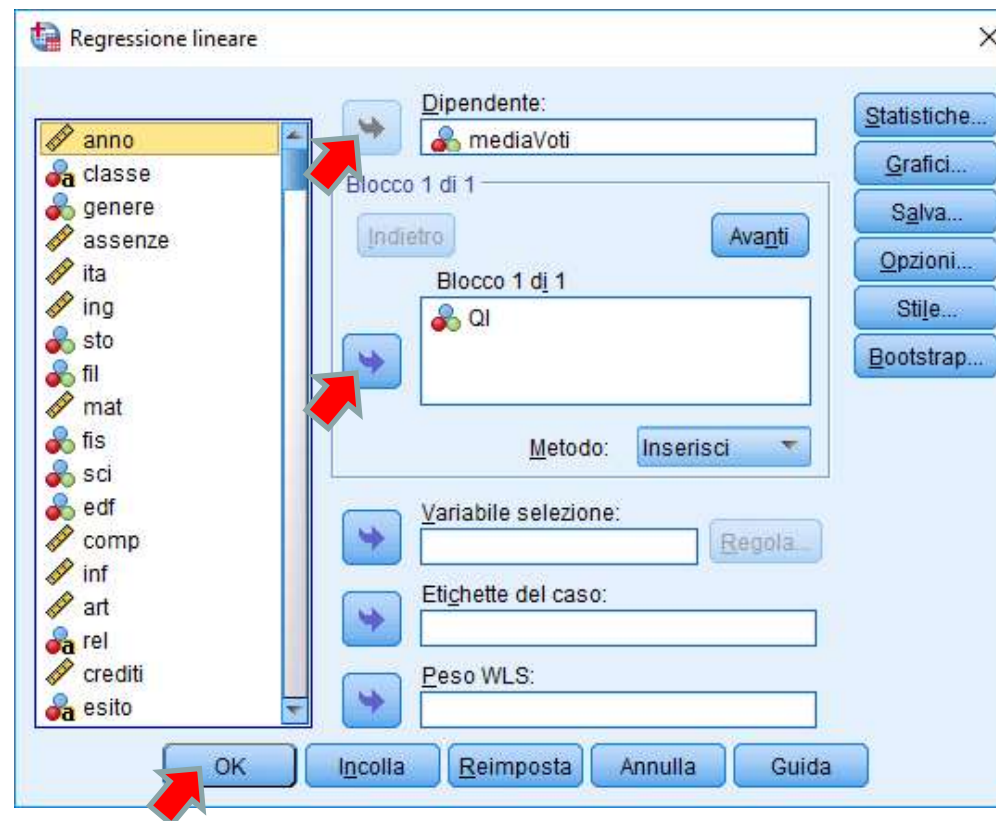
Regressione

La regressione bivariata in SPSS

1. Selezionare la variabile dipendente (es. mediaVoti) e inserirla nel pannello «Dipendente»

2. Selezionare la variabile indipendente (es. QI) e inserirla nel pannello «Blocco 1 di 1»

3. Cliccare su «OK»



Variabili immesse/rimosse^a

Modello	Variabili immesse	Variabili rimosse	Metodo
1	QI ^b	.	Inserisci

a. Variabile dipendente: mediaVoti

b. Sono state immesse tutte le variabili richieste.

Regressione

Stima dei parametri in SPSS

intercetta

a

Coefficienti^a

Modello	Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
	B	Errore std.	Beta		
1	(Costante)	5,791		19,224	,000
	Q1	,445	,403	5,792	,000

a. Variabile dipendente: mediaVoti

Var. indipendente (X)

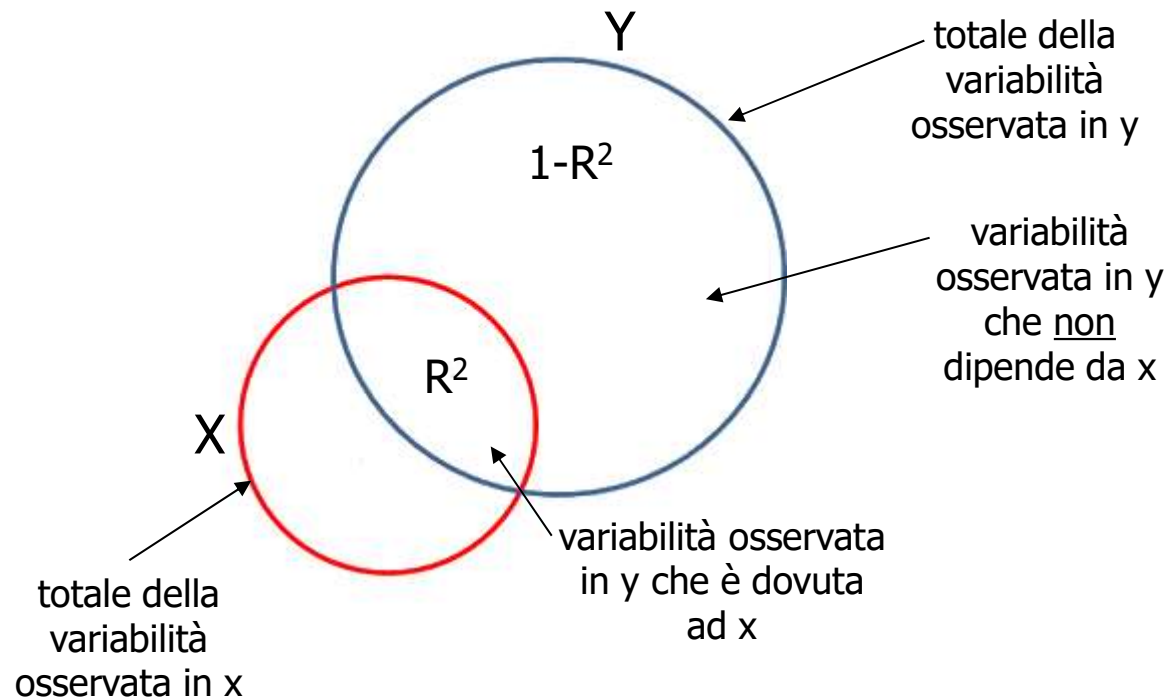
stima di b

stima di β

Regressione

Adeguatezza della equazione di regressione

Il coefficiente di determinazione (R^2) è un indice della proporzione di varianza di y che viene spiegata dalla regressione



Adeguatezza della equazione di regressione

Da $\sqrt{(1-R^2)}$ è possibile ricavare l'errore standard della stima: questo coefficiente rappresenta la dispersione dei punti intorno alla retta

$$S_e = \sqrt{(1-r^2)}S_y = \sqrt{\frac{\sum (y - y')^2}{n-2}}$$

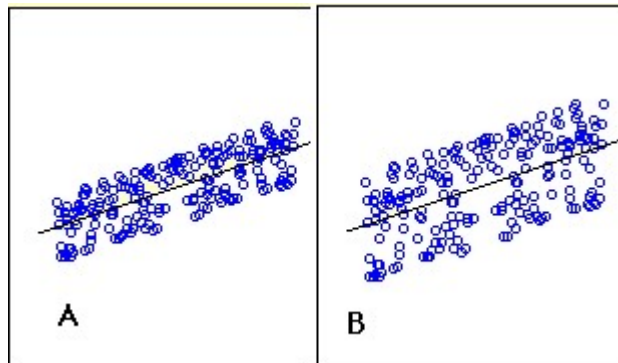
È una misura dell'accuratezza della stima (è inversamente proporzionale all'accuratezza: rappresenta una sorta di «errore medio» che si commette nel predire y)

Regressione

Adeguatezza della equazione di regressione

Se $R^2 = 1$, allora $S_e = 0$: l'errore che si commette nel predire y da x è uguale a zero (tutti i valori di y cadono sulla retta di regressione y')

Al diminuire di r , aumenta l'errore (aumenta la dispersione dei punti attorno alla retta)



L'errore standard della stima è maggiore in B che in A

Regressione

Adeguatezza della equazione di regressione in SPSS

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,403 ^a	,162	,158	,753

a. Predittori: (costante), QI

Coefficiente di determinazione

Errore standard della stima

Statistiche descrittive

	N	Media	Deviazione std.
QI	211	3,77	,790
mediaVoti	178	7,50	,817
Numero di casi validi (listwise)	175		

La regressione bivariata

Esercizio 3

Subject	Social Support X	Life Satisfaction Y	Y'	Residual Y - Y'
1	4	6		
2	6	5		
3	8	7		
4	10	9		
5	12	9		

$$\bar{X} = 8.0 \quad \bar{Y} = 7.2$$

$$S_X = 3.16 \quad S_Y = 1.79$$

$$\text{cov}(X, Y) = 5.00 \quad r_{xy} = .884$$

**Calcola l'errore standard
della stima ...**

Regressione

Verifica delle ipotesi (test di significatività)

Significatività statistica dei coefficienti b:

Ipotesi nulla e alternativa sono:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

**dove β è il coefficiente di regressione
nella popolazione**

Regressione

Verifica delle ipotesi (test di significatività)

Significatività statistica dei coefficienti b:

La significatività di b viene esaminata utilizzando il test della t di Student (con $n - 2$ gradi di libertà):

$$t = \frac{b - \beta}{S_b}$$

dove S_b è l'errore standard di b

Ricorda: la formula della maggior parte dei test statistici è:

Quale **STATISTICA**? Quella osservata sul campione: in questo caso b

statistica - parametro
errore standard

Quale **PARAMETRO**? Quello ipotizzato nella popolazione in base ad H_0 : in questo caso $\beta = 0$

La deviazione standard della distribuzione campionaria (in questo caso di b)

Regressione

Verifica delle ipotesi (test di significatività)

Significatività statistica dei coefficienti b:

$$t = \frac{b - \beta}{s_b} \quad \text{si semplifica in} \quad t = \frac{b}{s_b}$$

dove

$$s_b = \frac{s_e}{\sqrt{\sum (X_i - \bar{X})^2}}$$

Il test t è pari al rapporto tra il coefficiente b e il suo errore standard:

**valori di t superiori a | 2 | indicano
che il coefficiente è significativo: rifiuto H₀**

Regressione

Verifica delle ipotesi (test di significatività)

Significatività statistica dei coefficienti b in SPSS

Coefficienti^a

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore std.	Beta		
1	(Costante)	5,791	,301		19,224	,000
	QI	,445	,077	,403	5,792	,000

a. Variabile dipendente: mediaVoti

b Errore standard della stima t di student significatività

$$t = \frac{b}{s_b} = \frac{.445}{.077} = 5.792$$

Regressione

Valori critici della distribuzione t di Student

Con $n = 175$, i gradi di libertà sono pari a:
 $175 - 2 = 173$

5.792 > 1.98
 quindi rifiuto H_0

gdl	Ipotesi monodirezionale					
	.10	.05	.025	.01	.005	.0005
	Ipotesi bidirezionale					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

La regressione bivariata

Esercizio 4

La regressione bivariata è stata utilizzata per esaminare il grado in cui la motivazione intrinseca allo studio influenza il voto al termine dell'anno scolastico. La motivazione intrinseca è stata misurata con una scala i cui punteggi variano da un minimo di 1 ad un massimo di 5.

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,214 ^a	,046	,041	,804

a. Predittori: (costante), Mot_Intr

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore std.	Beta		
1	(Costante)	6,934	,205		33,838	,000
	Mot_Intr	,183	,063	,214	2,896	,004

a. Variabile dipendente: mediaVoti

1. Quale è la variabile dipendente?
2. Quale è la var. indipendente?
3. La relazione tra motivazione intrinseca e voto è: positiva negativa assente
4. Il coefficiente di regressione b è significativo? Cosa si può concludere?
5. Se la motivazione aumenta di una unità, di quanto aumenta il voto?
6. La % di varianza del voto che viene spiegata dalla motivazione è pari a:
7. Uno studente ottiene un punteggio nella motivazione pari a 4. Il punteggio previsto nella variabile dipendente (Y') è pari a:

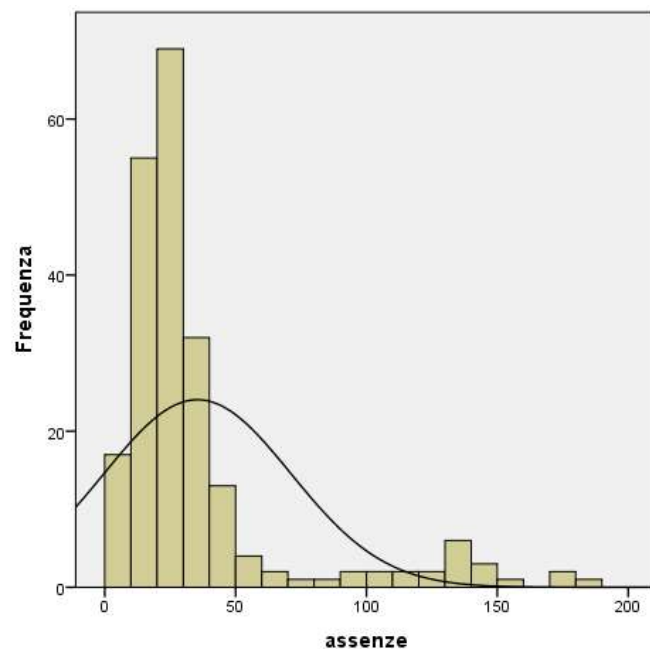
La regressione bivariata

Esercizio 5

Esaminiamo ora la relazione tra apertura mentale degli studenti (openness) e le ore di assenza effettuate durante l'anno scolastico (assenze). Sono state calcolate anche alcune statistiche descrittive.

Statistiche descrittive

	N Statistica	Intervallo Statistica	Minimo Statistica	Massimo Statistica	Media		Deviazione std. Statistica	Asimmetria		Curtosi	
					Statistica	Errore std.		Statistica	Errore std.	Statistica	Errore std.
assenze	215	187	2	189	35,41	2,433	35,681	2,457	,166	5,537	,330
Openness	212	4	1	5	3,89	,063	,915	-,683	,167	,128	,333
Numero di casi validi (listwise)	212										



Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,401 ^a	,161	,157	31,522

a. Predittori: (costante), Openness

Coefficienti^a

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore std.	Beta		
1	(Costante)	93,297	9,484		9,838	,000
	Openness	-15,059	2,373	-,401	-6,347	,000

a. Variabile dipendente: assenze

La regressione bivariata

Esercizio 5

1. Cosa si può dire a proposito della forma della distribuzione della variabile "assenze"?
2. Il campo di variazione delle ore di assenza è pari a: _____
3. C'è una relazione tra apertura mentale e ore di assenza? Se sì, di che tipo?
4. Se la variabile apertura mentale aumenta di una unità, come cambiano le ore di assenza in media?
5. Se l'apertura mentale aumenta di 5 unità, come cambiano le assenze in media?
6. Se il punteggio nell'apertura mentale è uguale a zero, il valore atteso nelle ore di assenza è pari a: _____
7. Il coefficiente di regressione non standardizzato è pari a: _____. È significativo? _____
8. La percentuale di varianza delle ore di assenza che viene spiegata dall'apertura mentale è pari a: _____
9. Se un soggetto ha un'apertura mentale pari a 3.5, le ore di assenza durante l'anno scolastico previste in base all'equazione di regressione (Y') sono pari a: _____