

The SPSS logo is displayed in white text on a red square background in the top-left corner.

Gold  
Business  
Partner

The IBM logo is located in the top-right corner, consisting of its characteristic eight horizontal stripes.

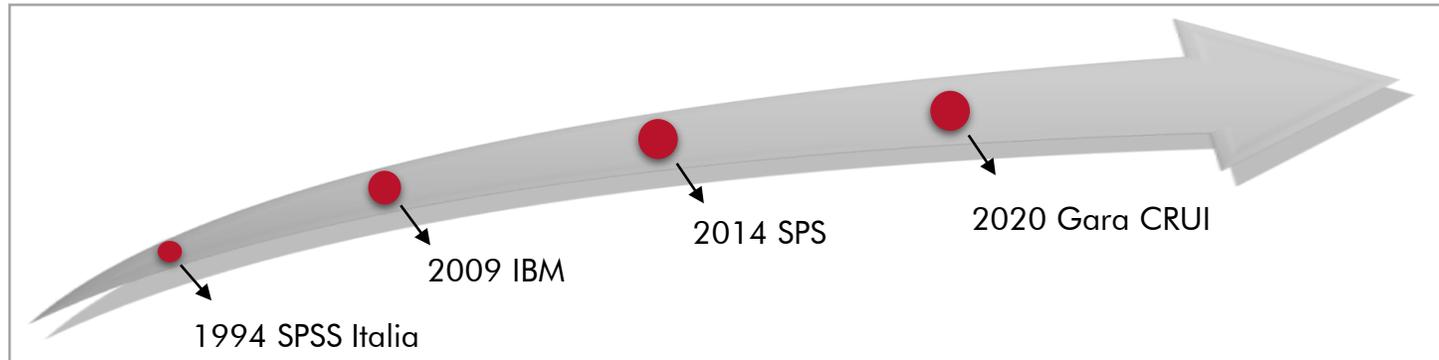
Competency  
Data Science &  
Business Analytics



# Webinar

Come selezionare le variabili in un modello di regressione lineare multipla

## Roadmap



## Posizionamento

- +25 anni di esperienza sui prodotti SPSS
- IBM BP **Software Support Provider** con competenza IBM Expert in Data Science and Business Analytics
- Gold Business Partner IBM



## Presenza sul mercato

- Scuole e Università
- Enti di ricerca (CNR, ISS, CREA ecc.)
- Sanità (IRCCS, AO, AOU ecc.)
- Pubblica Amministrazione
- Associazioni, Fondazioni
- Aziende (GdO, Finance, Retail, ecc.)

Il contratto stipulato tra CRUI e SPS S.r.l. in data 21 ottobre 2019 si riferisce alla fornitura licenze Campus IBM SPSS Statistics del "Catalogo licenze CRUI Campus PA" e di licenze aggiuntive IBM SPSS Amos e IBM SPSS Statistics per usi amministrativi.

Tutte le informazioni sul contratto sono disponibili al link

<https://www.spss.it/crui-universita-italiane>

Il referente commerciale per l'esecuzione contratto in SPS Srl è:

Laura Zerbini (+39 335 1360538, [laura.zerbini@spss.it](mailto:laura.zerbini@spss.it))

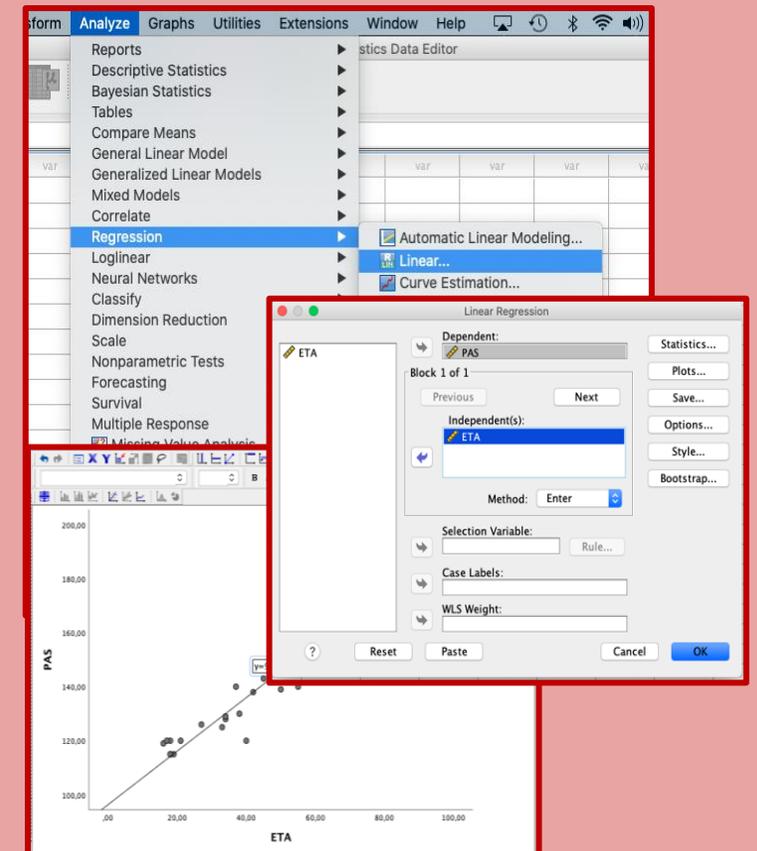


**Laura Zerbini**

Consulente in Italia di soluzioni  
SPSS per l'analisi statistica de...



Arianna Azzellino docente @SPS



Contesto introduttivo

Regressione lineare

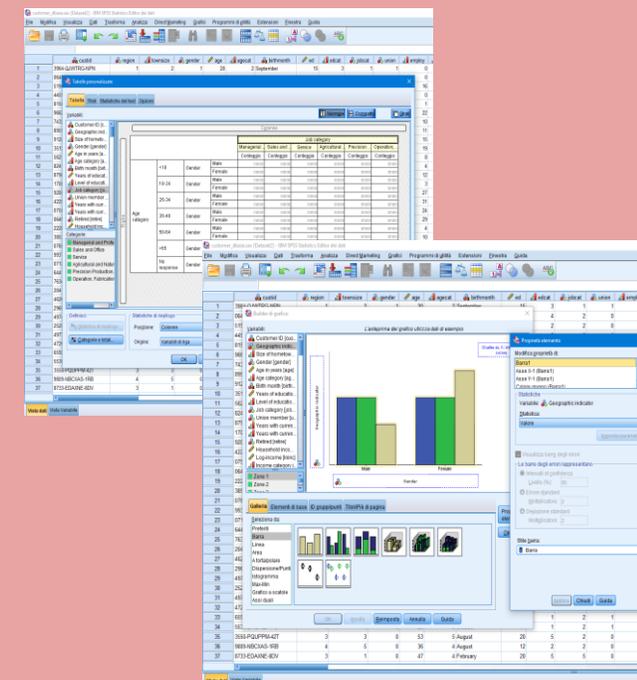
Assunti su cui poggia la regressione lineare

Come selezionare le variabili?

Esempio con IBM SPSS Statistics

Ringraziamenti

Scarica la nostra App sugli store



Contesto introduttivo

Regressione lineare

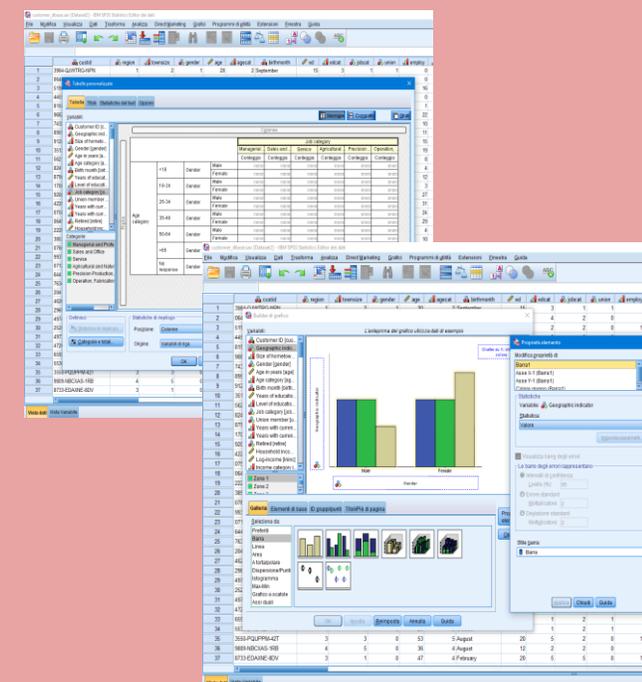
Assunti su cui poggia la regressione lineare

Come selezionare le variabili?

Esempio con IBM SPSS Statistics

Ringraziamenti

Scarica la nostra App sugli store



Contesto introduttivo

Regressione lineare

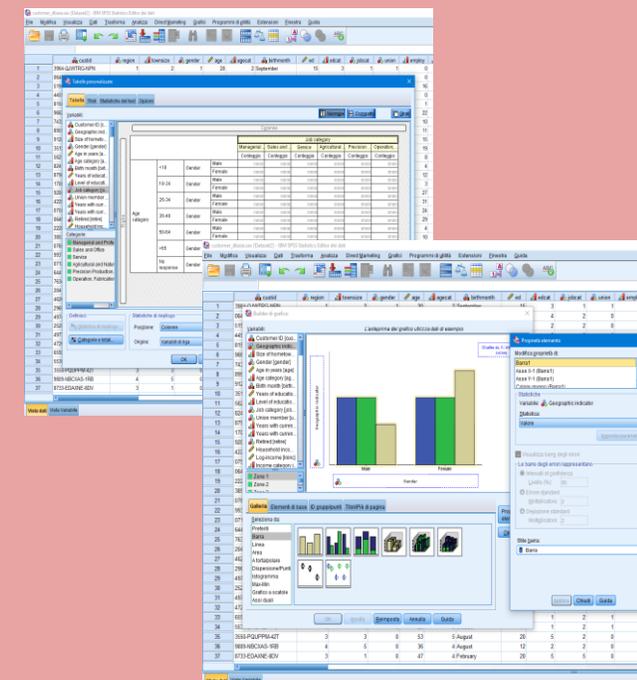
Assunti su cui poggia la regressione lineare

Come selezionare le variabili?

Esempio con IBM SPSS Statistics

Ringraziamenti

Scarica la nostra App sugli store

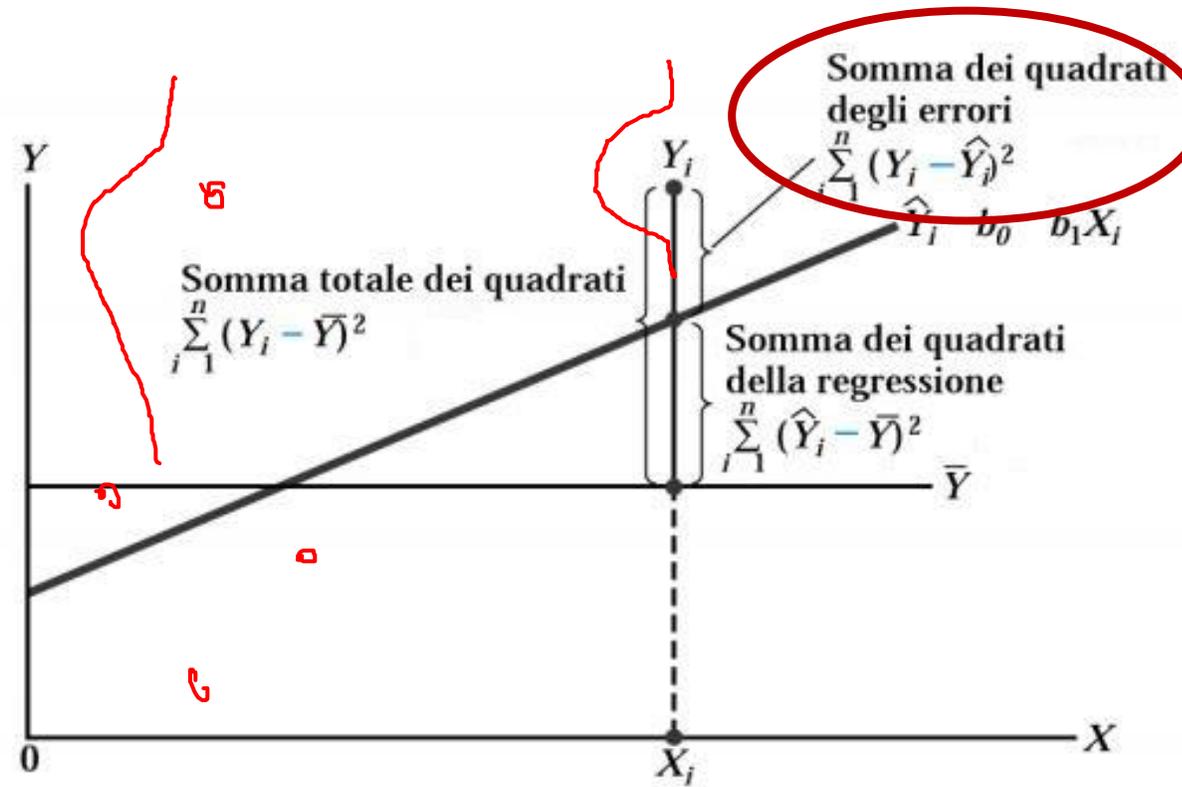


Forma generale del modello di regressione lineare multipla:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

- $\beta_0$ : rappresenta l'intercetta del modello (valore del modello nel caso in cui tutte le variabili esplicative  $X$  valgano 0).
- $\beta_j$  ( $j = 1, 2, \dots, k$ ): è chiamato coefficiente di regressione, e misura di quanto varia  $Y$  (var. target) al variare unitario di  $X_j$ , al netto delle altre  $k-1$  variabili esplicative (ovvero tenendo il loro valore costante).
- $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ : componente deterministica del modello.
- $\epsilon$ : errore/componente casuale del modello (si suppone essere costante ed intorno allo 0).

I parametri del modello vengono stimati attraverso i minimi quadrati (OLS), che hanno l'obiettivo di minimizzare i residui al quadrato:



$$F = \frac{(TSS-RSS)/k}{RSS/(n-k-1)} = \frac{MSS/k}{RSS/(n-k-1)}$$

Il suo test di ipotesi è così strutturato:

$$\begin{cases} H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1: \text{almeno un } \beta_j \text{ diverso da } 0 \end{cases}$$

- Accetto  $H_0$  se  $p\text{-value}(\text{sign.}) > 0.05 \rightarrow$  modello da rigettare.
- Rifiuto  $H_0$  se  $p\text{-value}(\text{sign.}) \leq 0.05 \rightarrow$  modello da prendere in considerazione.

*Sign. test F*

		ANOVA <sup>a</sup>				
Modello		Somma dei quadrati	gl	Media quadratica	F	Sign.
1	Regressione	,286	4	,072	52,697	,000 <sup>b</sup>
	Residuo	,160	118	,001		
	Totale	,447	122			

a. Variabile dipendente: area della zona forestale bruciata dall'incendio (in ettari - ha)

b. Predittori: (costante), RH)umidità relativa esterna in %, ISI)indicatore quantitativo di velocità di diffusione del fuoco nei primi istanti (in punti), DMC)indicatore quantitativo di umidità negli strati organici superficiali della foresta (in punti), temperatura esterna in °C

**Bontà del modello** (quanta variabilità di Y il modello ci spiega, attraverso l'indice di determinazione lineare):

$$R^2 = \left(1 - \frac{RSS}{TSS}\right) = \frac{MSS}{TSS}$$

Indice compreso tra 0 ed 1, in particolare:

- $R^2=0$ : il modello non spiega niente circa la variabilità di Y.
- $R^2=1$ : il modello spiega l'intera variabilità di Y (modello ideale!)

- $Y$  gaussiana
- Le variabili esplicative  $X$  sono correlate con  $Y$
- Omoschedasticità degli errori (costanti)
- Residui hanno andamento gaussiano
- Assenza di autocorrelazione dei residui
- Assenza di multicollinearità tra i regressori

Contesto introduttivo

Regressione lineare

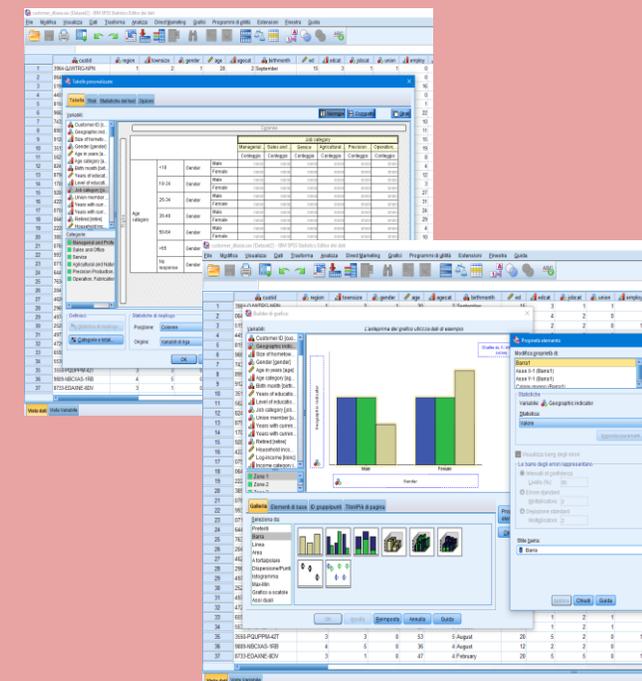
Assunti su cui poggia la regressione lineare

Come selezionare le variabili?

Esempio con IBM SPSS Statistics

Ringraziamenti

Scarica la nostra App sugli store



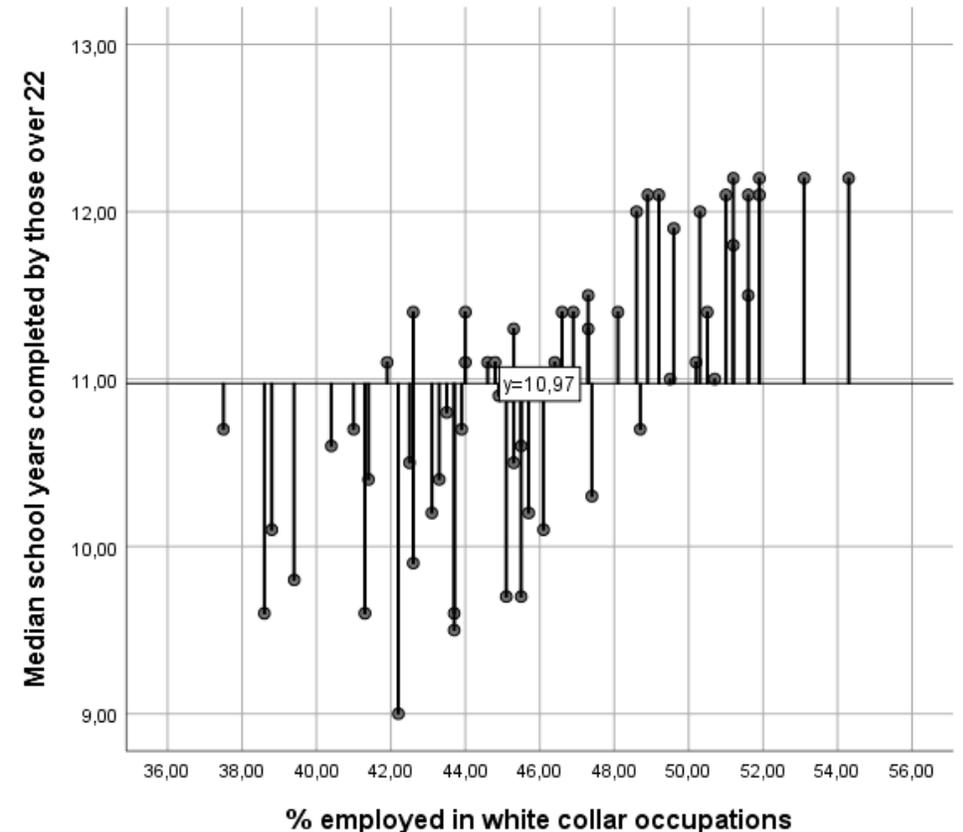
Interpretabilità  
modello

Selezione variabili  
significative

F ANOVA-Regressione:

Varianza spiegata dalla variabile-x / df

Varianza residua / df

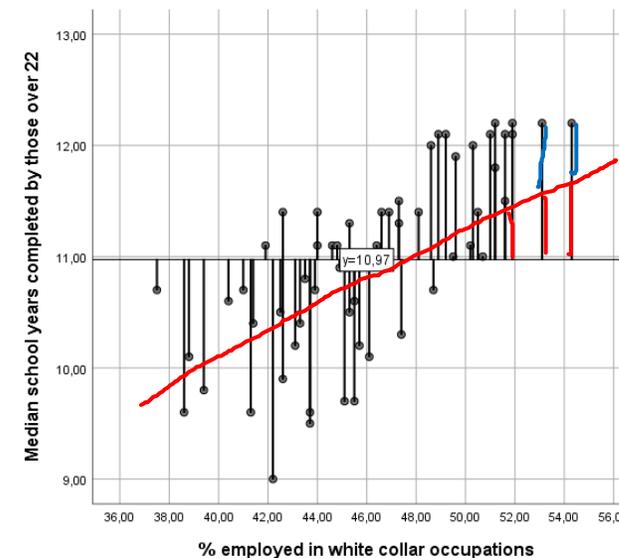
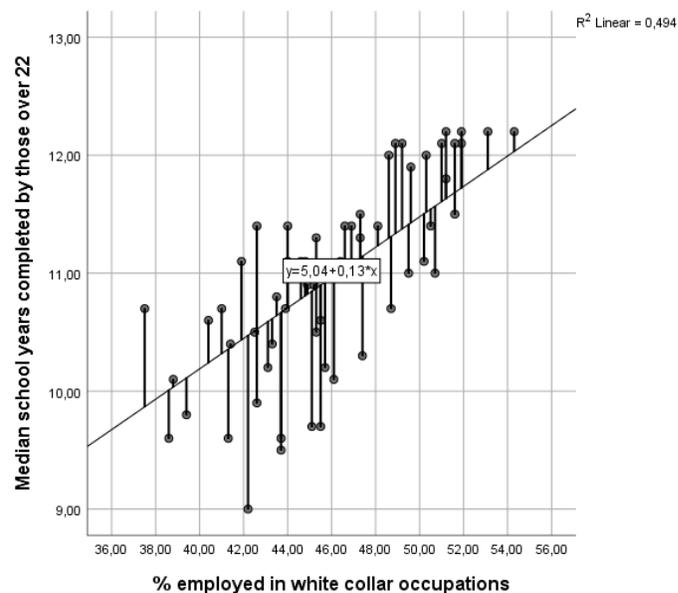


F ANOVA-Regressione:

Varianza spiegata dalla variabile-x / df

Varianza residua / df

**Varianza  
Spiegata**



**Varianza  
Non spiegata**

Stepwise: Confronto in termini di F ad ogni step

$$\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}$$

- 1) Stepwise Forward (in Avanti)
- 2) Stepwise Backward (all'Indietro)
- 3) Stepwise Stepwise (misto)

Contesto introduttivo

Regressione lineare

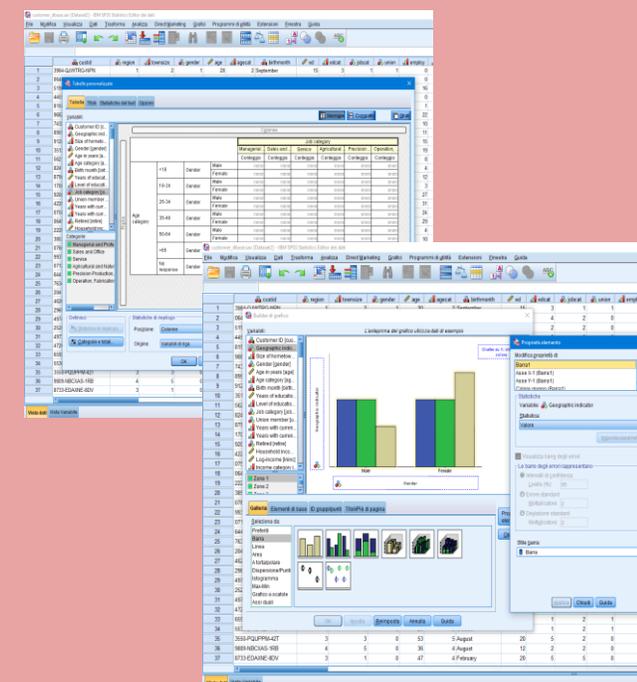
Assunti su cui poggia la regressione lineare

Come selezionare le variabili?

Esempio con IBM SPSS Statistics

Ringraziamenti

Scarica la nostra App sugli store



Pollution.sav [DataSet6] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

1 : mort 921,87 Visible: 16 of 16 Variables

	jult	ovr65	popn	educ	hous	dens	nonw	wwdrk	poor	hc	nox	sdiox	humid	mort	var
1	71,00	8,10	3,34	11,40	81,50	3243,00	8,80	42,60	11,70	21,00	15,00	59,00	59,00	921,87	
2	72,00	11,10	3,14	11,00	78,80	4281,00	3,50	50,70	14,40	8,00	10,00	39,00	57,00	997,88	
3	74,00	10,40	3,21	9,80	81,60	4260,00	0,80	39,40	12,40	6,00	6,00	33,00	54,00	962,35	
4	79,00	6,50	3,41	11,10	77,50	3125,00	27,10	50,20	20,60	18,00	8,00	24,00	56,00	982,29	
5	77,00	7,60	3,44	9,60	84,60	6441,00	24,40	43,70	14,30	43,00	38,00	206,00	55,00	1071,29	
6	80,00	7,70	3,45	10,20	66,80	3325,00	38,50	43,10	25,50	30,00	32,00	72,00	54,00	1030,38	
7	74,00	10,90	3,23	12,10	83,90	4679,00	3,50	49,20	11,30	21,00	32,00	62,00	56,00	934,70	
8	73,00	9,30	3,29	10,60	86,00	2140,00	5,30	40,40	10,50	6,00	4,00	4,00	56,00	899,53	
9	70,00	9,00	3,31	10,50	83,20	6582,00	8,10	42,50	12,60	18,00	12,00	37,00	61,00	1001,90	
10	72,00	9,50	3,36	10,70	79,30	4213,00	6,70	41,00	13,20	12,00	7,00	20,00	59,00	912,35	
11	79,00	7,70	3,39	9,60	69,20	2302,00	22,20	41,30	24,20	18,00	8,00	27,00	56,00	1017,61	
12	76,00	8,60	3,20	10,90	83,40	6122,00	16,30	44,90	10,70	88,00	63,00	278,00	58,00	1024,89	
13	77,00	9,20	3,21	10,20	77,00	4101,00	13,00	45,70	15,10	26,00	26,00	146,00	57,00	970,47	
14	71,00	8,80	3,29	11,10	86,30	3042,00	14,70	44,60	11,40	31,00	21,00	64,00	60,00	985,95	
15	75,00	8,00	3,26	11,90	78,40	4259,00	13,10	49,60	13,90	23,00	9,00	15,00	58,00	958,84	
16	85,00	7,10	3,22	11,80	79,90	1441,00	14,80	51,20	16,10	1,00	1,00	1,00	54,00	860,10	
17	75,00	7,50	3,35	11,40	81,90	4029,00	12,40	44,00	12,00	6,00	4,00	16,00	58,00	936,23	
18	73,00	8,20	3,15	12,20	84,20	4824,00	4,70	53,10	12,70	17,00	8,00	28,00	38,00	871,77	
19	74,00	7,20	3,44	10,80	87,00	4834,00	15,80	43,50	13,60	52,00	35,00	124,00	59,00	959,22	
20	72,00	6,50	3,53	10,80	79,50	3694,00	13,10	33,80	12,40	11,00	4,00	11,00	61,00	941,18	
21	85,00	7,30	3,22	11,40	80,70	1844,00	11,50	48,10	18,50	1,00	1,00	1,00	53,00	891,71	

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

## Ringraziamenti e Conclusioni

LAURA ZERBINI @SPS

Scarica la nostra App sugli store

Download on the App Store

ANDROID APP ON Google play

The screenshot displays the SPSS software interface. The 'Analyze' menu is open, showing options like Reports, Descriptive Statistics, Bayesian Statistics, Tables, Compare Means, General Linear Model, Generalized Linear Models, Mixed Models, Correlate, Regression, Loglinear, Neural Networks, Classify, Dimension Reduction, Scale, Nonparametric Tests, Forecasting, Survival, Multiple Response, and Missing Value Analysis. The 'Regression' option is selected, and a submenu is visible with 'Automatic Linear Modeling...', 'Linear...', and 'Curve Estimation...'. The 'Linear Regression' dialog box is open, showing 'Dependent:' as 'PAS', 'Independent(s):' as 'ETA', and 'Method:' as 'Enter'. The 'Statistics...' button is highlighted. In the background, a scatter plot shows a positive linear correlation between 'ETA' (x-axis, 0.00 to 100.00) and 'PAS' (y-axis, 100.00 to 200.00).