



04 July 2023

Webinar SPS

Cluster Analysis with Statistics for Data Analysis

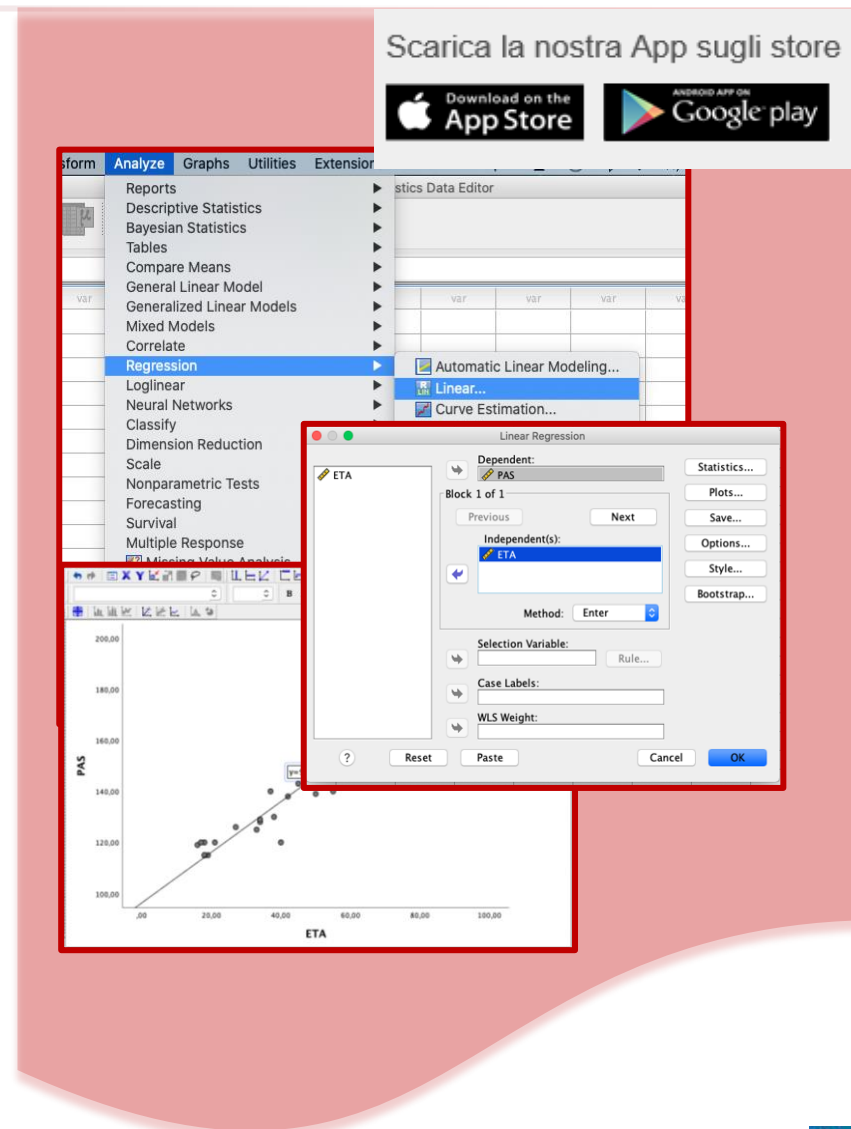
Statistics for Data Analysis Introduction

Introduction

Theory

Example with Statistics for Data Analysis

Acknowledgement



Statistics for Data Analysis Introduction

Introduction

Theory

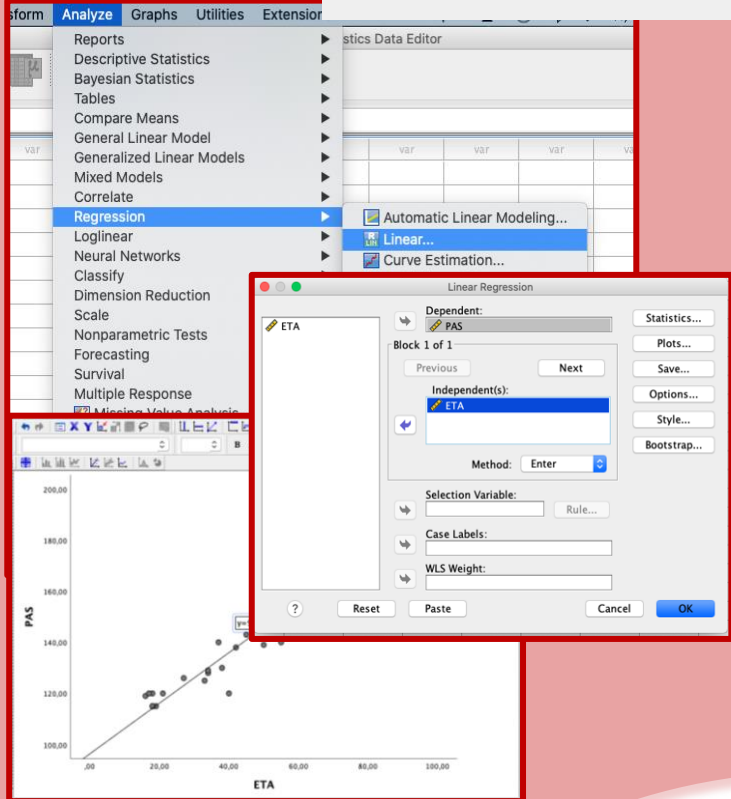
Example with Statistics for Data Analysis

Acknowledgement

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, and Multiple Response. The 'Regression' option is highlighted, and the 'Linear...' option is selected. The 'Linear Regression' dialog box is open, showing 'Dependent:' as 'PAS' and 'Independent(s):' as 'ETA'. The 'Method:' is set to 'Enter'. The 'Selection Variable:', 'Case Labels:', and 'WLS Weight:' fields are empty. The 'Statistics...' button is highlighted. In the background, a scatter plot of 'PAS' (Y-axis) versus 'ETA' (X-axis) is visible, showing a positive linear correlation with a regression line.

Desktop icon



Statistics for Data Analysis (LaunchBox v1.2.1.0)

> Dashboard

Add-On

Strumenti Aggiuntivi

Manuali d'uso

Avvia Statistics for Data Analysis

V. 28.0.1.1

Avvia

Webinar

22 giu

How to Process Questionnaire Data with Statistics for Data Analysis

mercoledì, 22 giugno, 2022

10:30 - 11:30

[Iscriviti](#)

5 lug

La Segmentazione con Statistics for Data Analysis

martedì, 05 luglio, 2022

10:30 - 11:30

[Iscriviti](#)

News

Per t

Scopri l'a

[Mostra alt](#)

E' an

Scopri co

[Mostra altro](#)

Rilasciata la patch per la Vulnerabilità Log4j

Clicca su "Mostra altro" per installare le fix pack

[Mostra altro](#)

Lingua

Contattaci

Prossimi Training

SURV Online

Dedicato all'Analisi di Sopravvivenza

18-19-20-21-22 luglio 2022

LOG-COX Online

Dedicato alla Regressione Logistica e alla Cox Regression

27-28-30 settembre 2022

INTRO Online

Corso Introduttivo a Statistics for Data Analysis

3-4-5-6-7 ottobre 2022

PSCORE Online

Dedicato al Propensity Score matching

18-19-21 Ottobre 2022

TEST Online

Dedicato ai Test Statistici

3-4-7-9-10 novembre 2022

REG Online

Dedicato ai Modelli di Regressione

21-22-23-24-25 novembre

MULTI Online

Dedicato all'Analisi Multivariata

12-13-14-15-16 dicembre

Premium



PScore

Premium



Youden



Scarica la nostra App sugli store



Search with:

Statistics for Data Analysis

Statistics for Data Analysis Introduction

Introduction

Theory

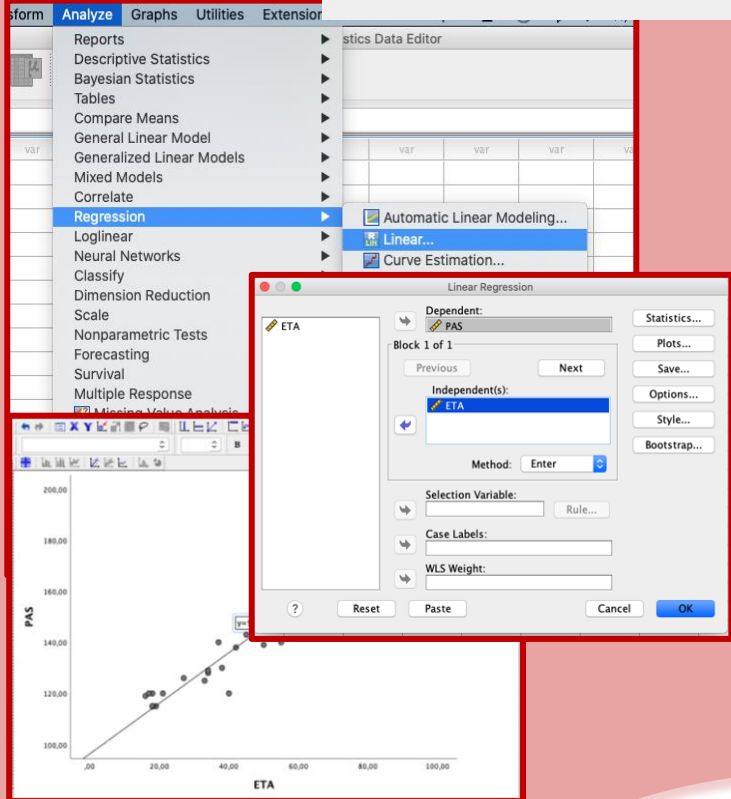
Example with Statistics for Data Analysis

Acknowledgement

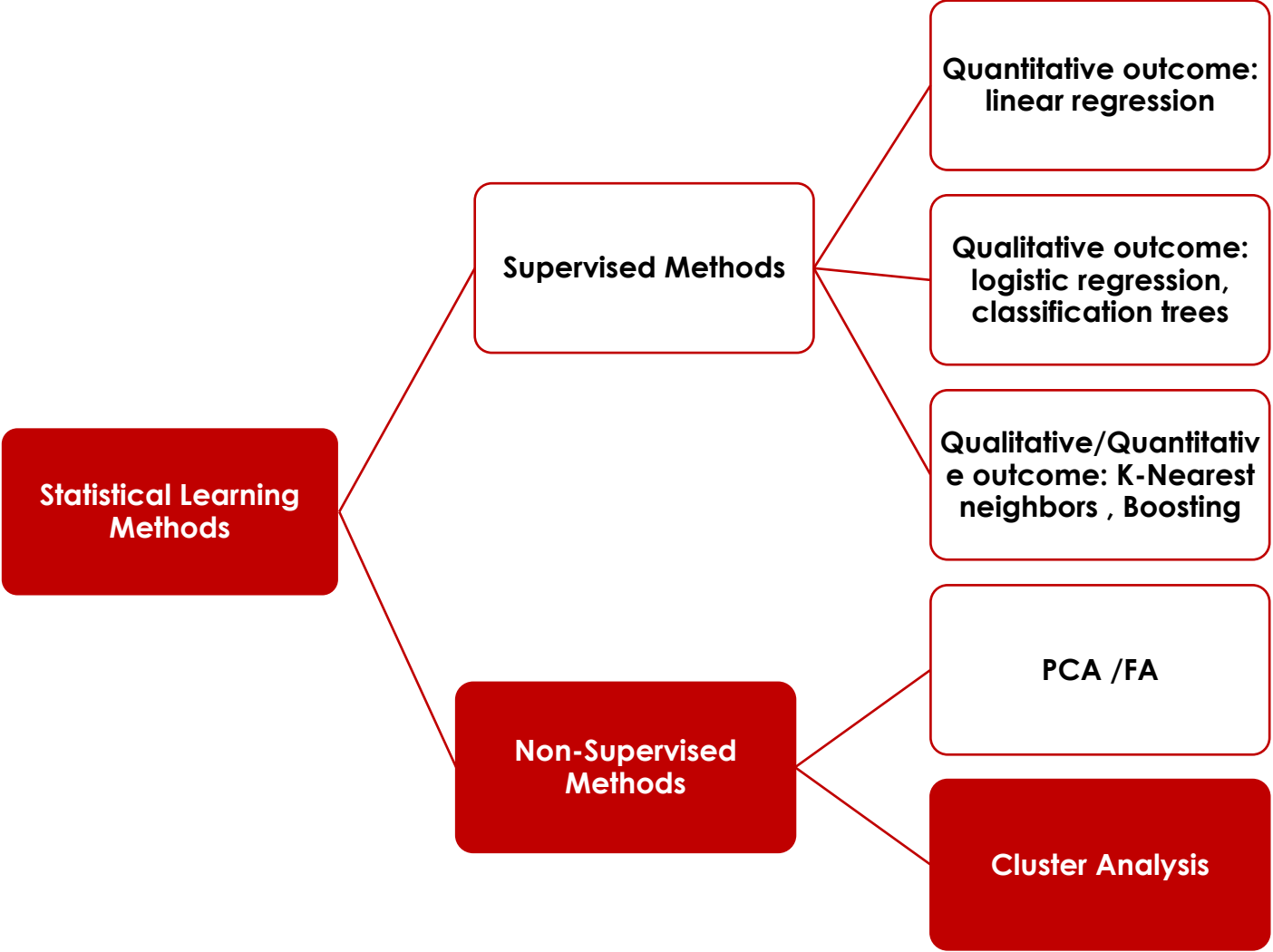
Scarica la nostra App sugli store

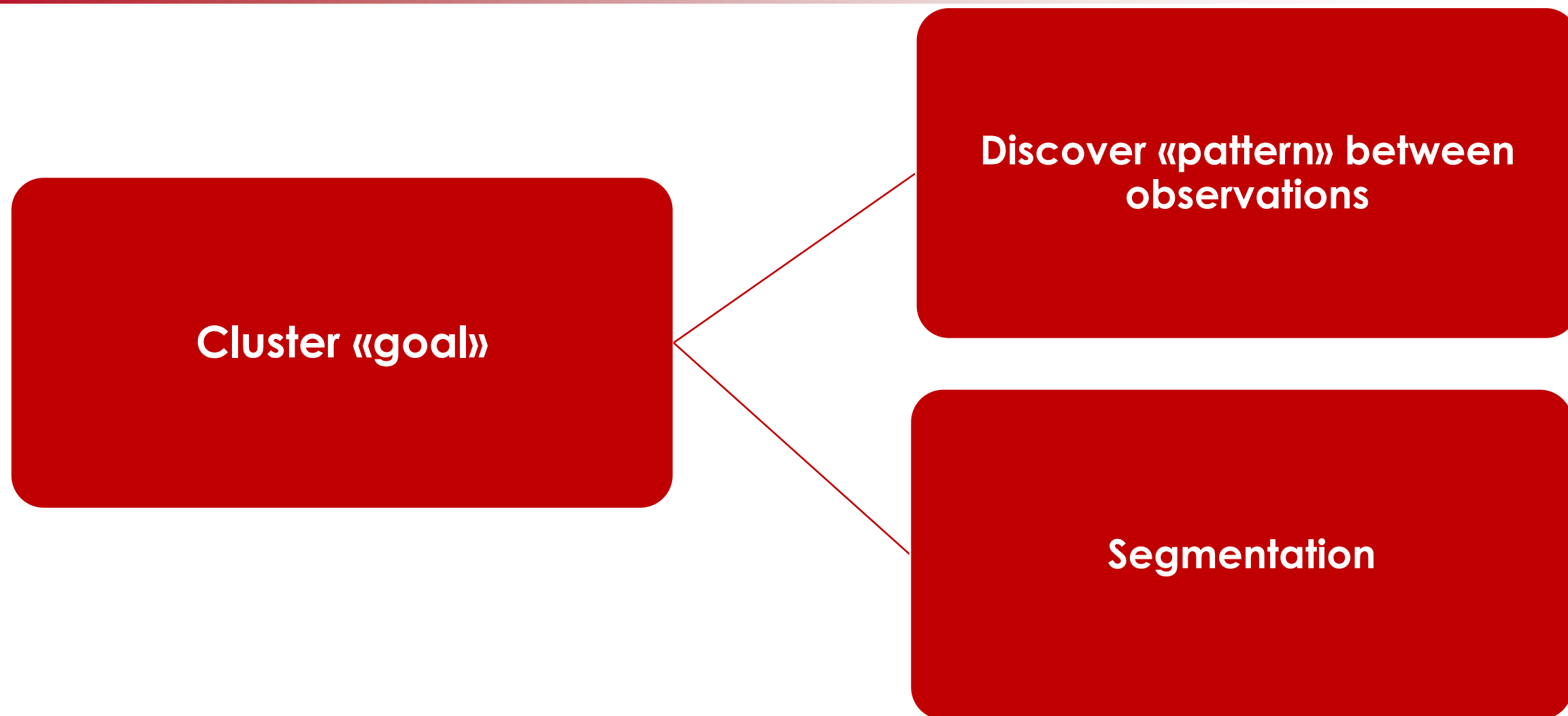
Download on the App Store

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, and Multiple Response. The 'Regression' option is highlighted, and a submenu is open showing 'Automatic Linear Modeling...', 'Linear...', and 'Curve Estimation...'. The 'Linear Regression' dialog box is open, showing 'Dependent:' as 'PAS' and 'Independent(s):' as 'ETA'. The 'Method:' is set to 'Enter'. The 'Selection Variable:', 'Case Labels:', and 'WLS Weight:' fields are empty. The 'Statistics...', 'Plots...', 'Save...', 'Options...', 'Style...', and 'Bootstrap...' buttons are visible on the right. A scatter plot of PAS vs ETA is shown in the background, with a regression line fitted to the data points.







When can we apply CA?

**Heterogeneous Observations
(statistical units)**

CA aims to group observations in order to obtain homogeneous groups. It has many applications, for example:

- **Scientific field:** with large dataset, for classify animals species, trees, minerals, geographic areas...
- **Marketing field:** market segmentation.
- **Sport field:** classify athletes' roles, based on a set of features.
- **Clinic field:** patients' segmentation, based on the medical record.

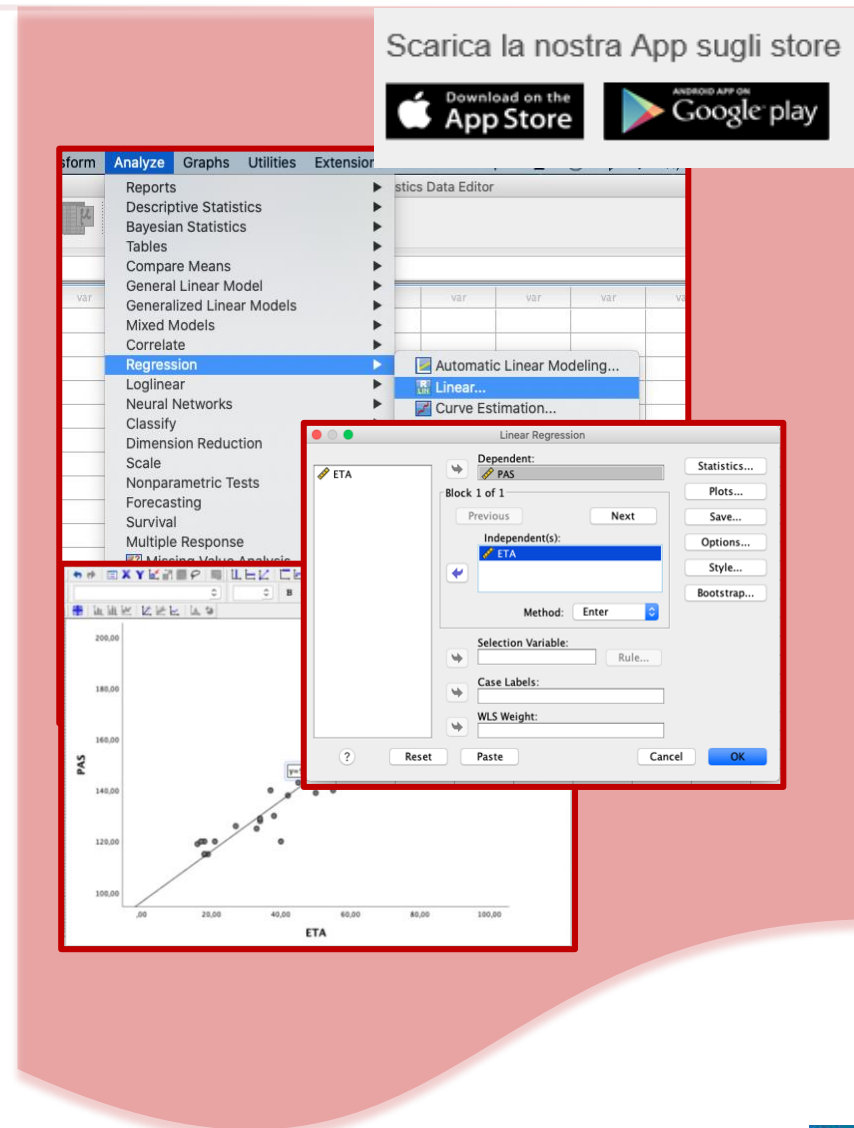
Statistics for Data Analysis Introduction

Introduction

Theory

Example with Statistics for Data Analysis

Acknowledgement



Statistician objective:

The main goal of Cluster Analysis is to identify groups or clusters of similar objects within a dataset. This technique aims to split the data such that observations within each cluster are more similar to each other (max homogeneity inside each cluster, max heterogeneity between clusters).

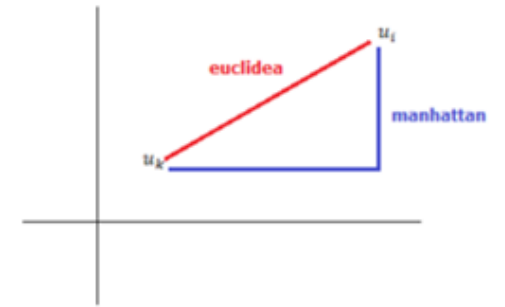
Step:

1. Dataset preparation.
2. Measuring «distance» between statistical units (quantitative variables) or «similarity» (qualitative variables).
3. Clustering algorithm.
4. Results analysis.

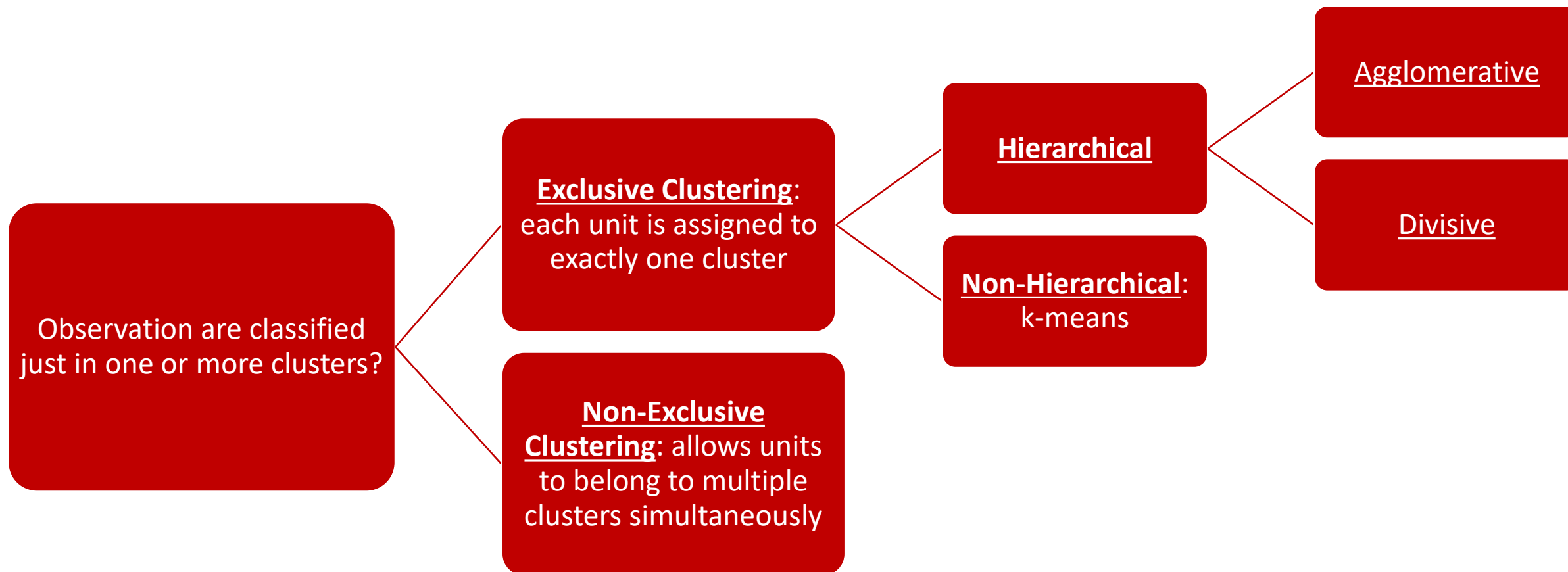
«Distance» computation between 2 statistical units:

-*Euclidean* distance $\rightarrow d(x_i, x_j) = \sqrt{\sum_{r=1}^p (x_{ir} - x_{jr})^2}$

-*Manhattan* distance $\rightarrow d(x_i, x_j) = \sum_{r=1}^p |x_{ir} - x_{jr}|$

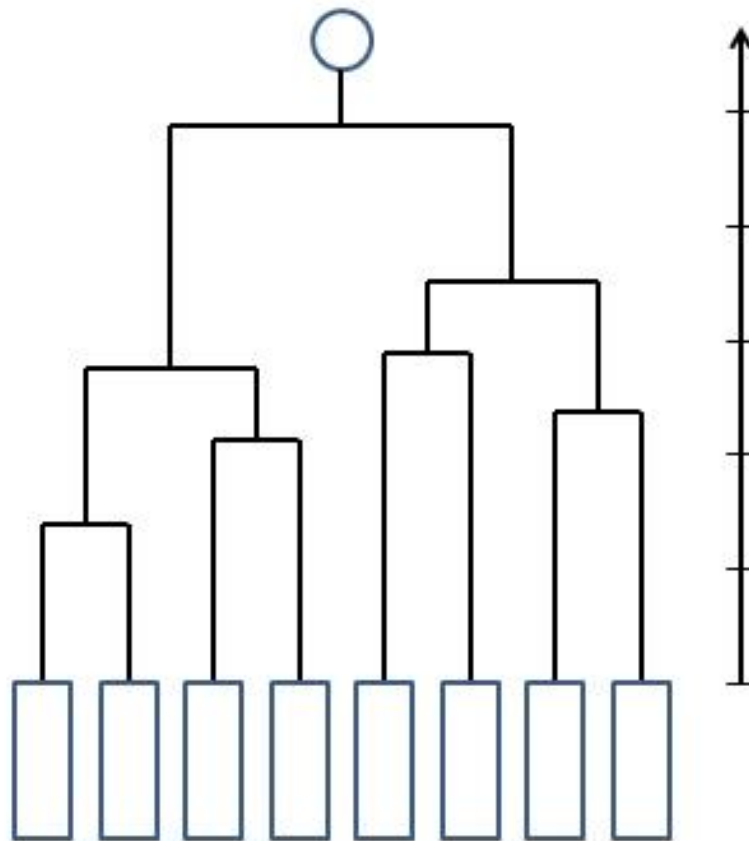


Note that distance is symmetric and positive (≥ 0)!

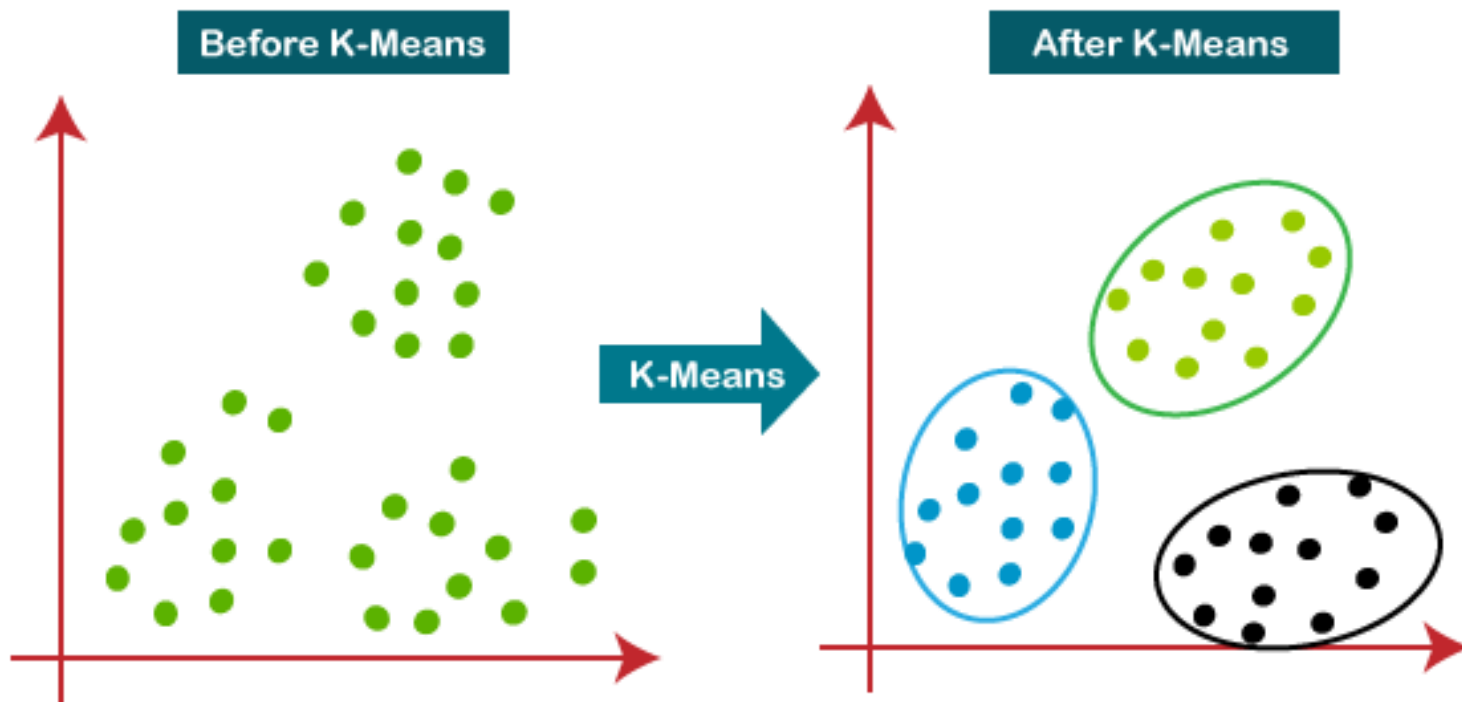


Agglomerative →

← Divisive



Distance →



	Hierarchical Clustering	Non-Hierarchical Clustering (K-Means)
	Very informative	Fast computation
	Complete (it evaluates all possible aggregations)	Flexible
	Large sample size: computational problems	-Initial cluster size choice -Not applicable with qualitative variables
Common problem →	Final cluster size choice and interpretation	

Statistics for Data Analysis Introduction

Introduction

Theory

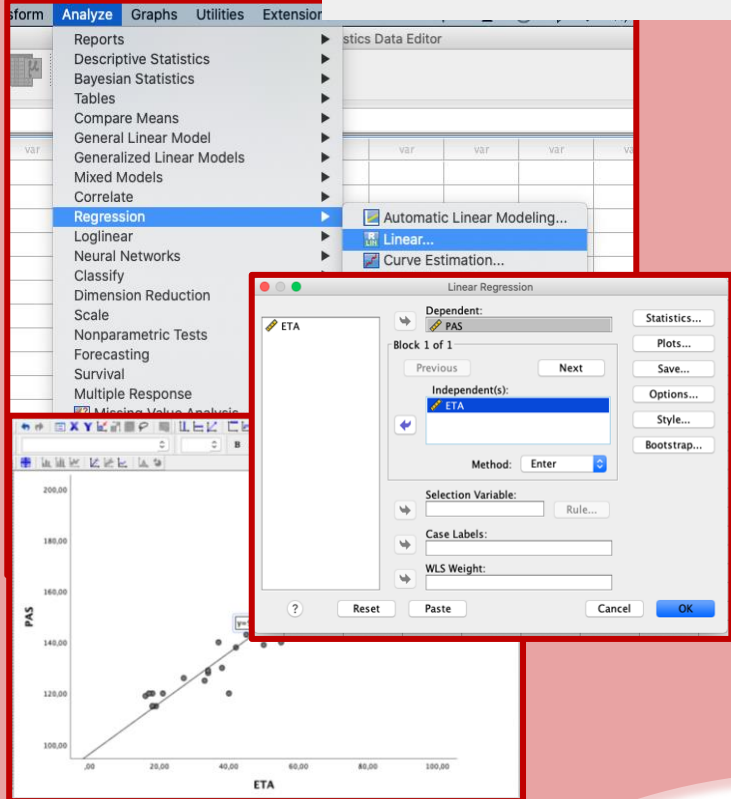
Example with Statistics for Data Analysis

Acknowledgement

Scarica la nostra App sugli store

Download on the App Store

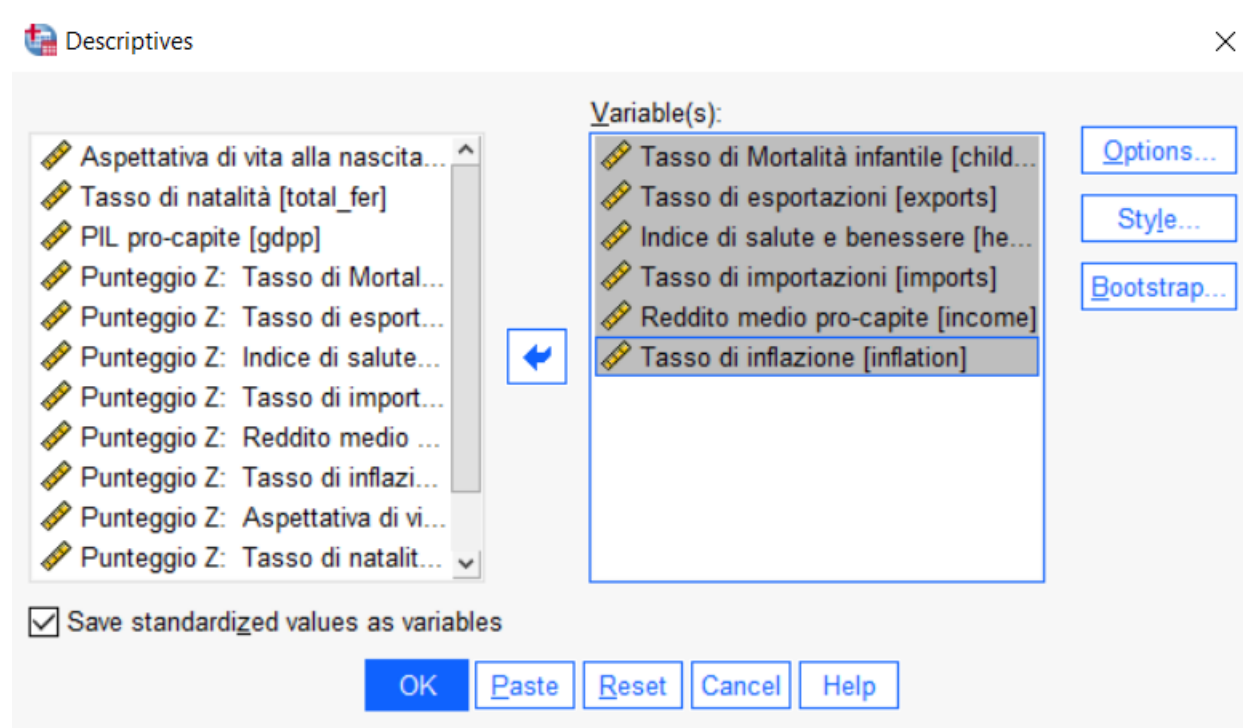
ANDROID APP ON Google play



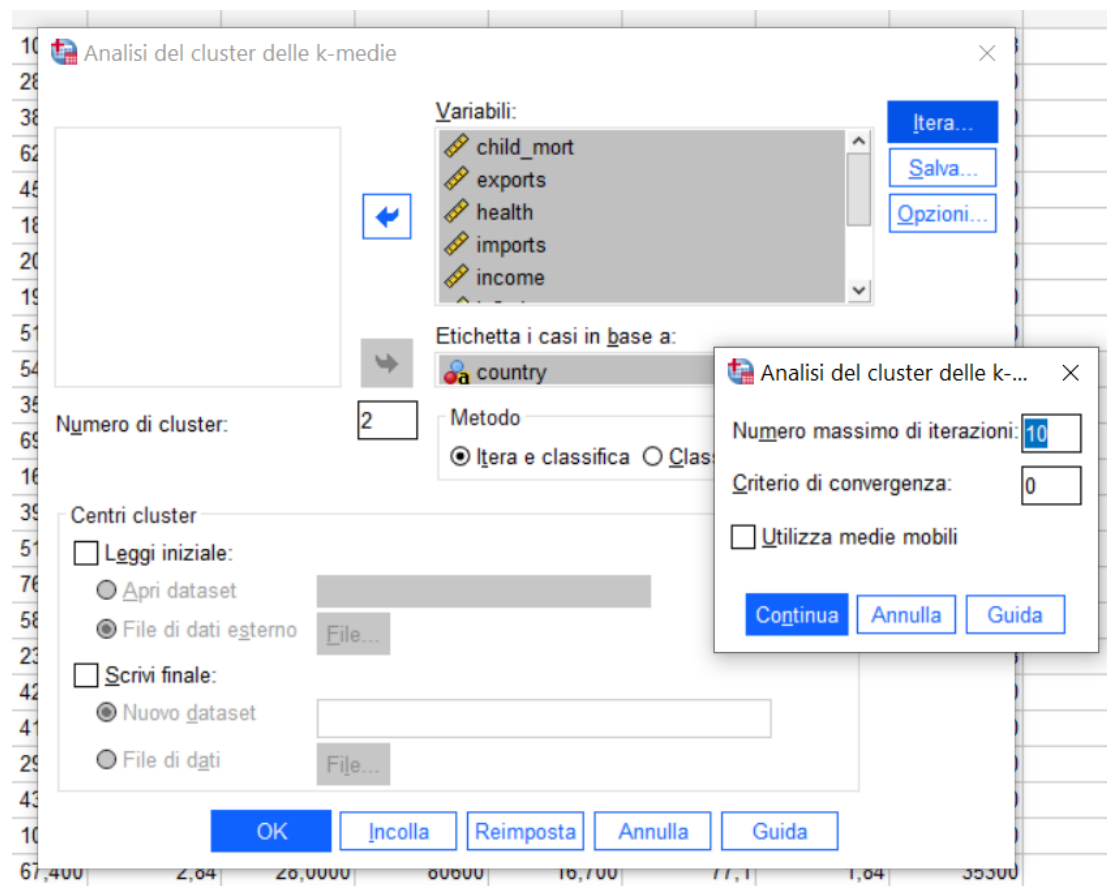
The screenshot displays the SPSS software interface. The 'Analyze' menu is open, showing options like Reports, Descriptive Statistics, Bayesian Statistics, Tables, Compare Means, General Linear Model, Generalized Linear Models, Mixed Models, Correlate, Regression, Loglinear, Neural Networks, Classify, Dimension Reduction, Scale, Nonparametric Tests, Forecasting, Survival, and Multiple Response. The 'Regression' option is selected, and the 'Linear...' option is highlighted. The 'Linear Regression' dialog box is open, showing 'PAS' as the dependent variable and 'ETA' as the independent variable. The 'Method' is set to 'Enter'. The 'Statistics...' button is highlighted. In the background, a scatter plot of PAS vs ETA is visible, showing a positive linear correlation.

Example with Statistics for Data Analysis

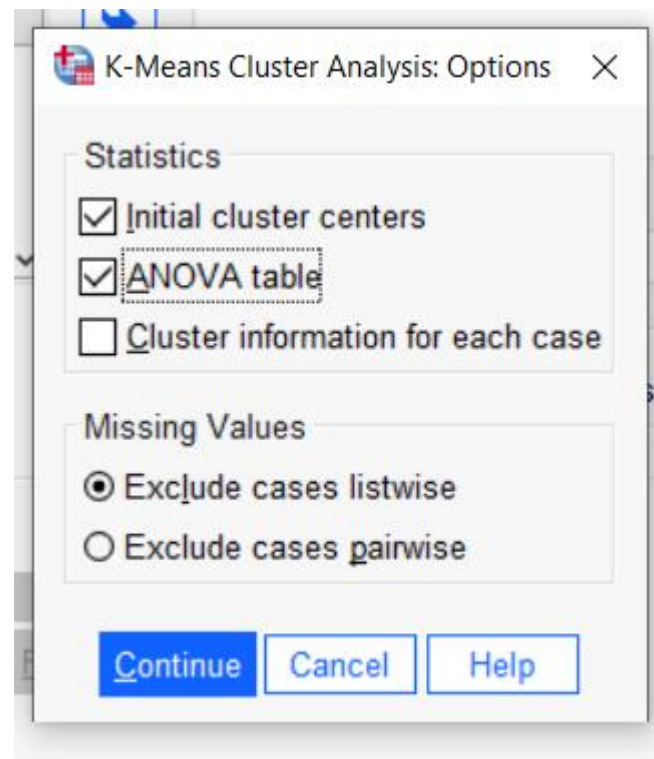
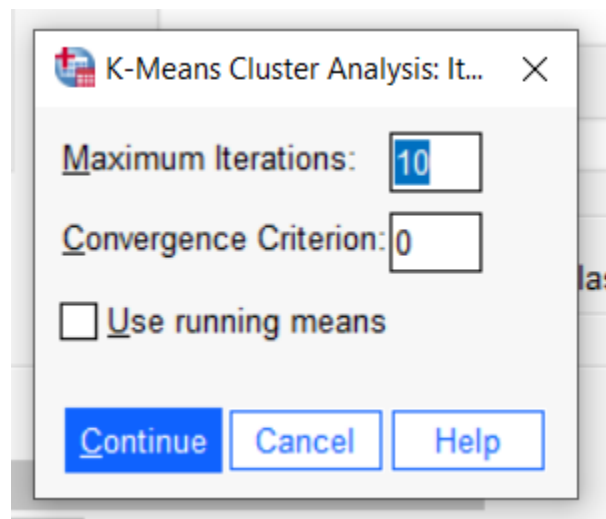
1) Analyze → Descriptive Statistics → Descriptives... → check «Save standardized values as variables»



2) Analyze → Classify → K-Means cluster....



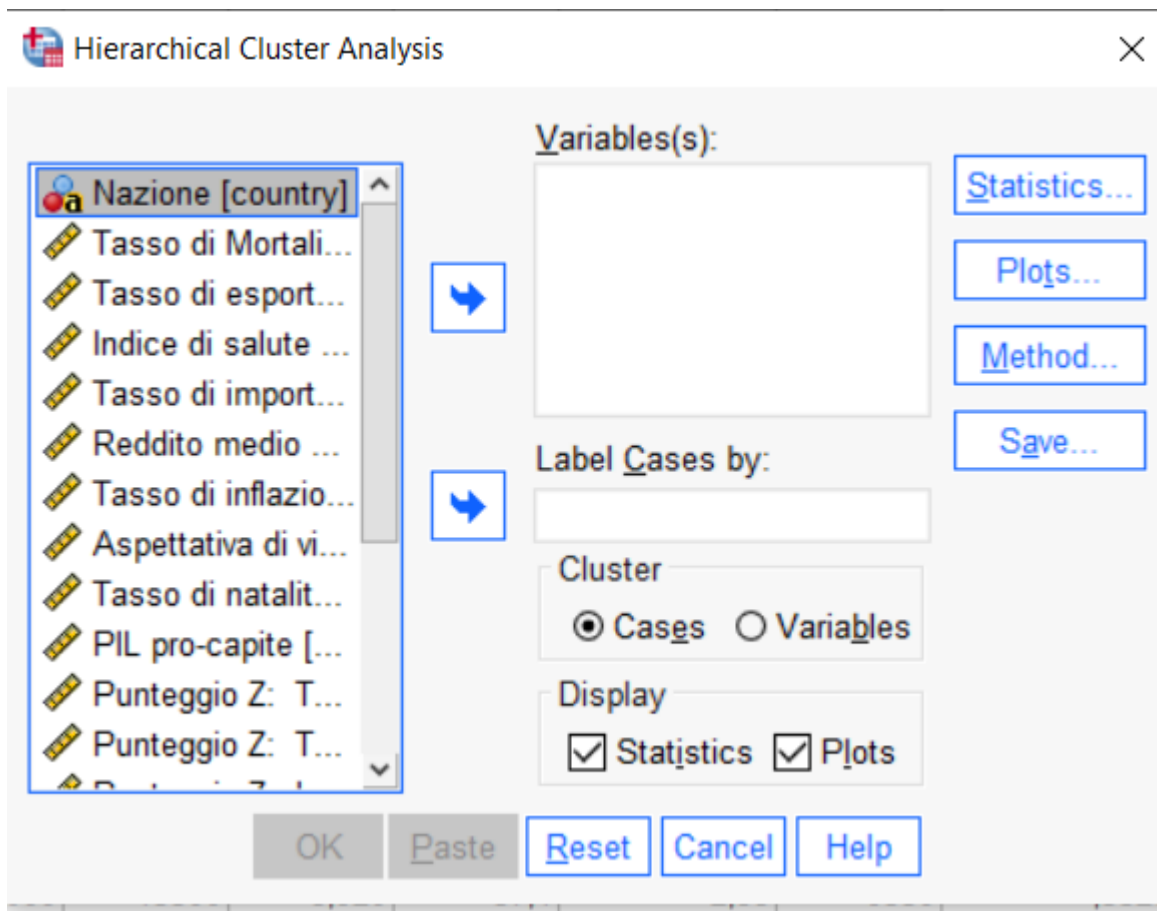
2.1) Menù: «Iterate» ed «Options»



3.1) Output: final centroids for each cluster

Centri finali del cluster		
	Cluster	
	1	2
child_mort	46,0	5,6
exports	37,587	55,969
health	6,42	8,48
imports	46,9923	46,4594
income	9603	48963
inflation	8,713	3,852
life_expec	68,3	80,1
total_fer	3,20	1,89
gdpp	5349	45091

4) Analyze → Classify → Hierarchical Cluster



- ✓ OK with quantitative/qualitative variables
- ✓ Exploratory analysis
- ✓ Units segmentation
- ✓ Large number of observations: non-hierarchical clustering
- ✓ Repeat the analysis with different #cluster sizes
- ✓ First step for further analysis (e.g. regression,...)

@Sales

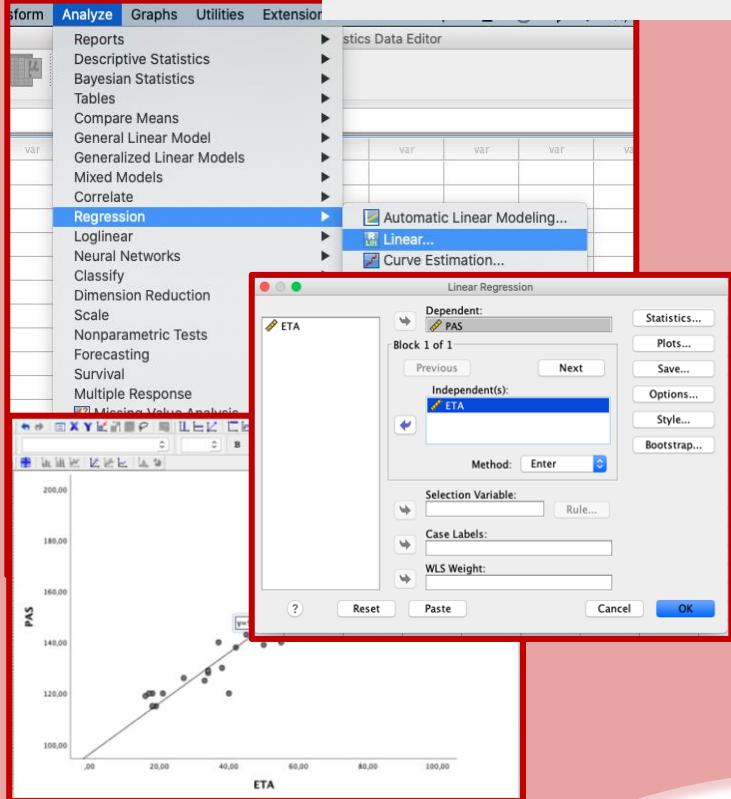
ACKNOWLEDGEMENT

Sales & Training Teams @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 the 'Linear...' option is highlighted. The 'Linear Regression' dialog box is open, showing 'Dependent:' as 'PAS' and 'Independent(s):' as 'ETA'. The 'Method:' is set to 'Enter'. The 'Selection Variable:' is empty. The 'Case Labels:' and 'WLS Weight:' fields are also empty. The 'Statistics...' button is visible. In the background, a scatter plot of 'PAS' (Y-axis) versus 'ETA' (X-axis) is shown, with a linear regression line fitted to the data points.