SPSS

FREE WEBINAR

Linear Regression for Accurate Interpretations: Analysis of Reading Speed

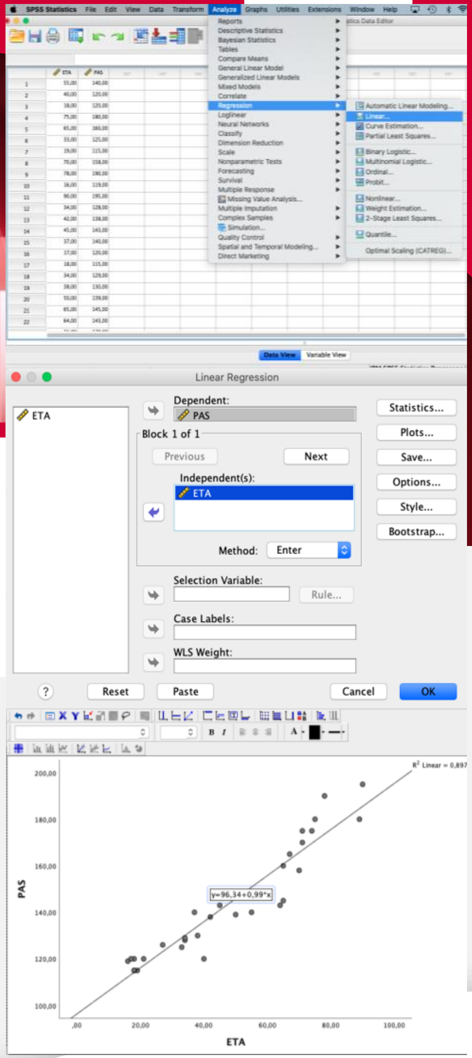
Tuesday, March 26th

2:30 PM

Statistics for Data Analysis

IBM.
Gold Partner

WEBINAR AGENDA



About Us	Statistics for Data Analysis	Linear Regression for Accurate Interpretations: Statistics for Data Analysis
25+ years of history… in brief	Services included in the solution (LaunchBox, Add-On, APP, webinar, Support, …)	• **Introduction** to Linear Regression • **Regression line**: the coefficients a, b, Beta • Regression model **evaluation**: R-squared • Statistical **inference**: t Student, F Fisher • Practical **example** • Q&A

Scarica la nostra APP sugli store

GET IT ON Google Play

Download on the App Store

Linear Regression for Accurate Interpretations: Statistics for Data Analysis

Introduction

- Most statistical techniques define a statistical model some **relationships** between variables of interest.

- Linear Regression (simple or multiple) is a technique designed to study and quantify the effects of one or more **independent variables** (explanatory or predictor variables) on a **dependent variable** (variable of interest).

- A simple statistical model is an **efficient** and **compact representation** of the data collected to describe an empirical phenomenon.

The statistical model and the representation we make, works for:

- Efficient and compact **description**
- Future **prediction**
- **Inference** on the population

That is: understanding the phenomenon

Statistics for Data Analysis

SPSS

Regression Assumptions:

- The dependent variable (Y) is normally distributed (Gaussian curve)
- The independent variables (X) are correlated with the dependent variable (Y)
- Homoscedasticity of errors
- The residuals have a Gaussian distribution

Statistics for Data Analysis

SPSS

Linear Regression for Accurate Interpretations:

Statistics for Data Analysis

Regression line: the coefficients a, b, Beta

Regression Line

Statistics for Data Analysis

SPSS

The purpose of the regression line is to represent the linear relationship between the independent variable and the dependent variable.

The line can be described by two coefficients: the constant term and the angular coefficient.

$$y = f(x\beta) + \varepsilon$$

Systematic component   Random Component

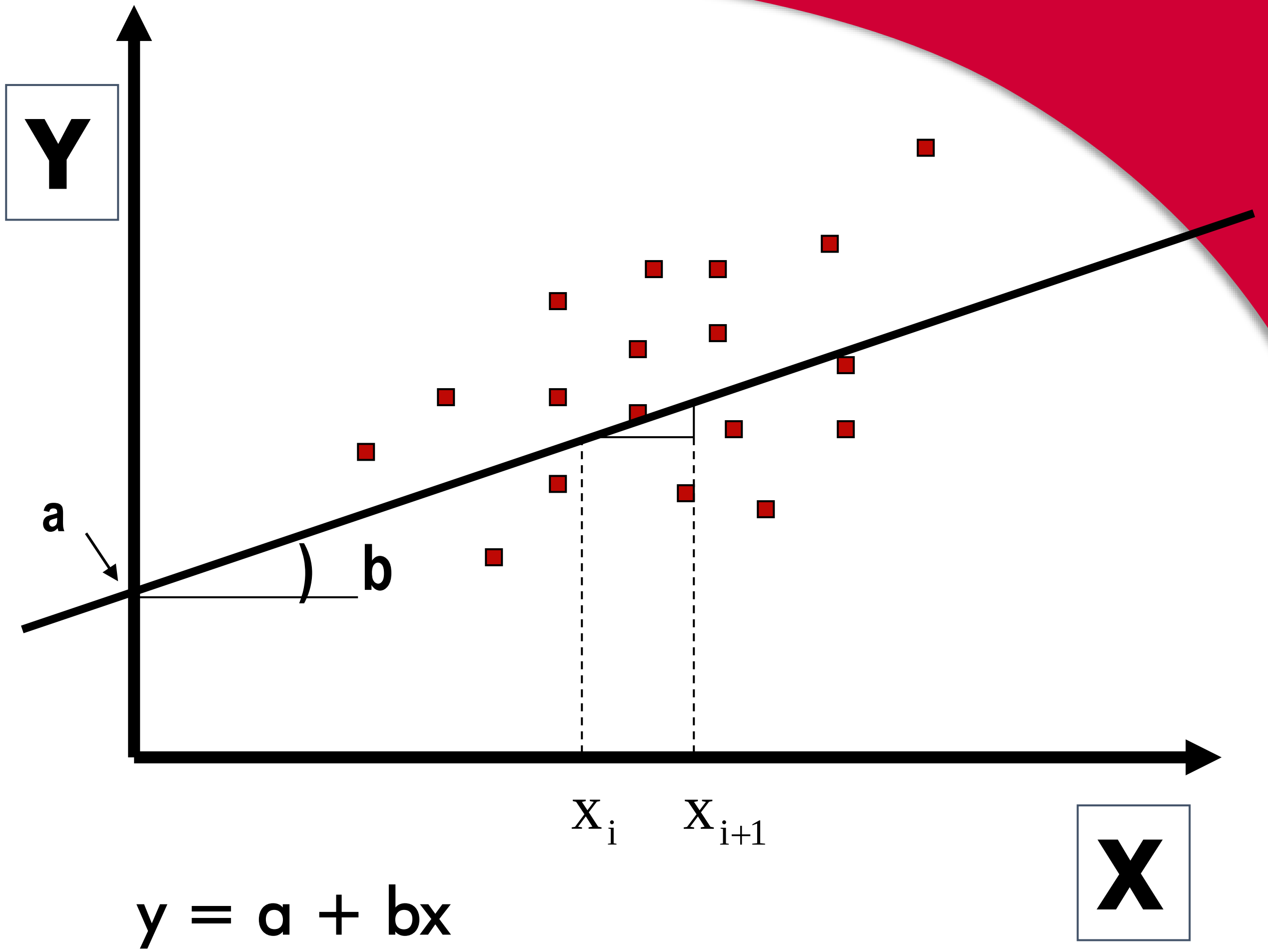
The aim of a statistical model is to represent as much information as possible through the systematic component; this is achieved by estimating the parameters contained in:

$$f(x\beta)$$

Regression Coefficient

Constant Coefficient: a
the intercept of the line indicates the expected (mean) value of the dependent variable when the independent variable is equal to 0,

Regression Coefficient: b
is the angular coefficient of the straight line: it indicates the expected change in the dependent variable when the independent variable changes by one unit.



$y = a + bx$

a = intercept value

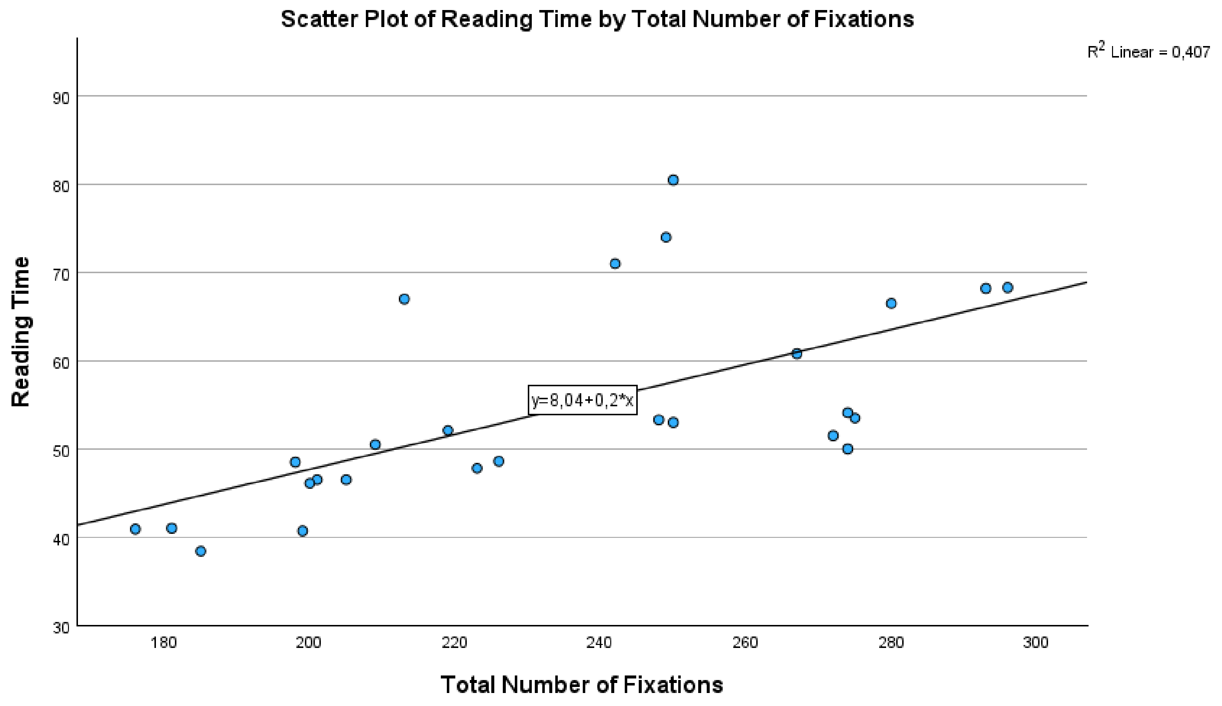
b = angular coefficient

Regression Coefficient

Statistics for Data Analysis

METHOD OF LEAST SQUARES

The model parameters are estimated by means of least squares. The slope of the regression line, determined by the coefficient b, is intended to minimise the distance between the values estimated by the model and the actual values.



Linear Regression for Accurate Interpretations:

Statistics for Data Analysis

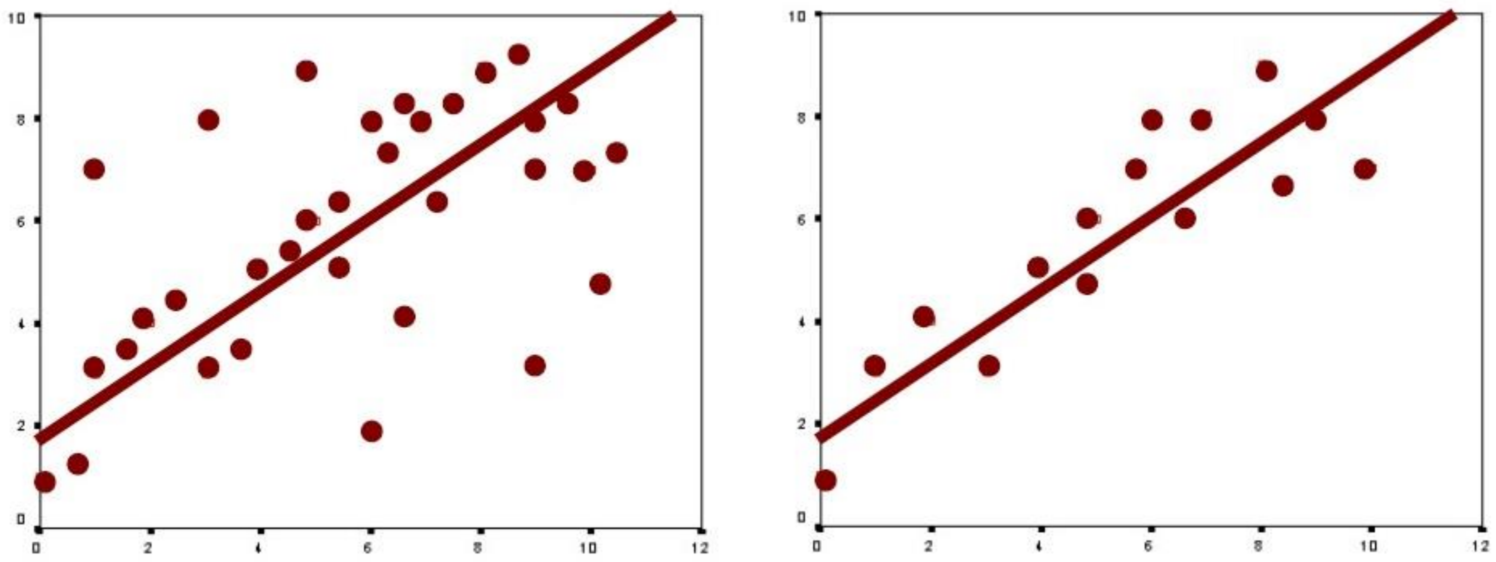
Regression model evaluation

Regression Model Evaluation

Goodness of fit:
Not all regression lines have the same predictive power, i.e. the same ability to fit the observed data.
The model will fit the data better the more it reduces the prediction error compared to not using such an estimate.

Low

High



Regression R-Square

Without **regression**, the only reliable prediction of a participant's reading time is the average of the entire group; thus the **error** associated with the prediction model is equivalent to the total **variance** of our data.
The variable variance to be predicted represents both the error we make in using the mean as a predictor and all the information we could account for using a better predictor than the mean.

Expecet value

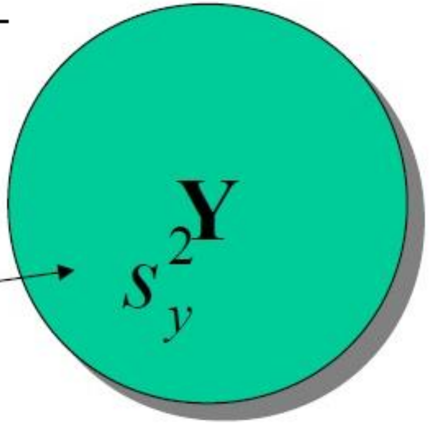
Mean

$$\hat{y}_i = M_y$$

Variance

$$s_y^2 = \frac{\sum (y_i - M)^2}{n-1}$$

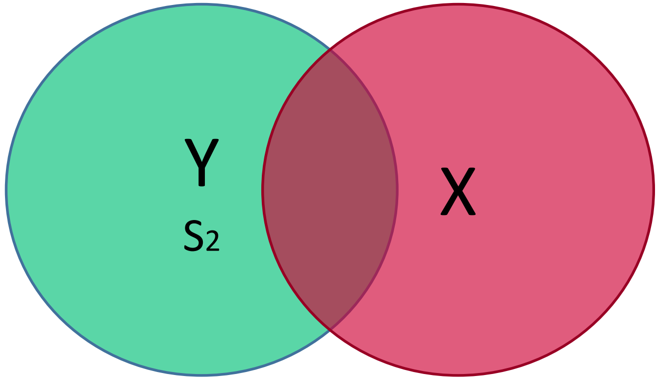
we are unable to explain the deviations from the mean (the variance).



Regression R-Square

By introducing one or more (independent) variables into our model with regression, we can explain part of the variance of the dependent variable. If the regression predicts the expected value for the dependent variable (e.g. reading time) better than the average, then we will have an R-squared > 0.

The R-squared value can vary between 0 and 1 and tells us how much the error in our forecasts is reduced by to the fact that we use regression; that is, how much percentage of variance we are able to explain with our model.



Statistics for Data Analysis

SPSS

Linear Regression for Accurate Interpretations:

Statistics for Data Analysis

Statistica Inference

Statistical Inference of Regression

Regression analysis can make a difference

01

How can we infer and determine that the **coefficients** we have found are actually effective in predicting the change in the independent variable?

02

The **null hypothesis** (H0) is that the coefficients we have found have the value of zero and therefore their contribution in predicting the dependent variable is zero.

03

The significance test (**p-value**) tells us how likely it is that the null hypothesis is true, conventionally we reject **H0** is if **p** < **0.05**. This is equivalent to saying that the probability that we have randomly found our parameter is less than 5%.

Linear Regression for Accurate Interpretations: Statistics for Data Analysis

Example

Example Introduction

While reading a passage, the eyes move over the paper/screen making many quick fixations. The time it takes us to read a given passage will depend both on the total number of fixations we make and on their duration.

In the dataset there are the variables: Subject (participant), TOTNumberF (total number of fixations), TOTDurationF (average fixation duration), TOTReading (total reading time).

The objective of the analysis is to assess the contribution of the independent variables (TOTNumF, TOTDurationF) on the dependent variable (TOTReadin).

OUR TEAM

@SPS

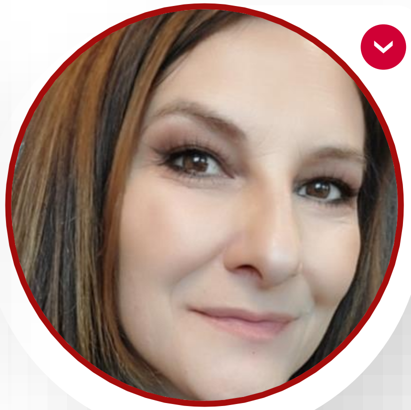
THANK YOU for joining us

from

Cristina Carini

Sales Consultant @SPS
FAO & WFP Account Manager

c.carini@spss.it
366 753 5607



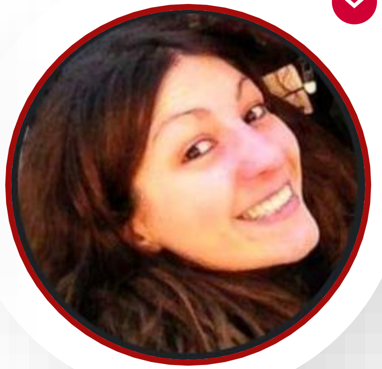
Laura Zerbini

Sales Consultant @SPS
For Educational & Research
University, AOU,
Research Institutions and IRCCS

l.zerbini@spss.it
335 1360538

Elisa Falchieri

Sales & Marketing Manager @SPS



Serena Demingo

Digital Marketing @SPS

Ramon Pedrini

Trainer @SPS