Statistics for Data Analysis

Powered by

SPSS

Gold Business Partner

IBM

Competency
Data Science & Business Analytics

Webinar
Linear Regression

17th March 2023

Webinar | Agenda

Statistics for Data Analysis

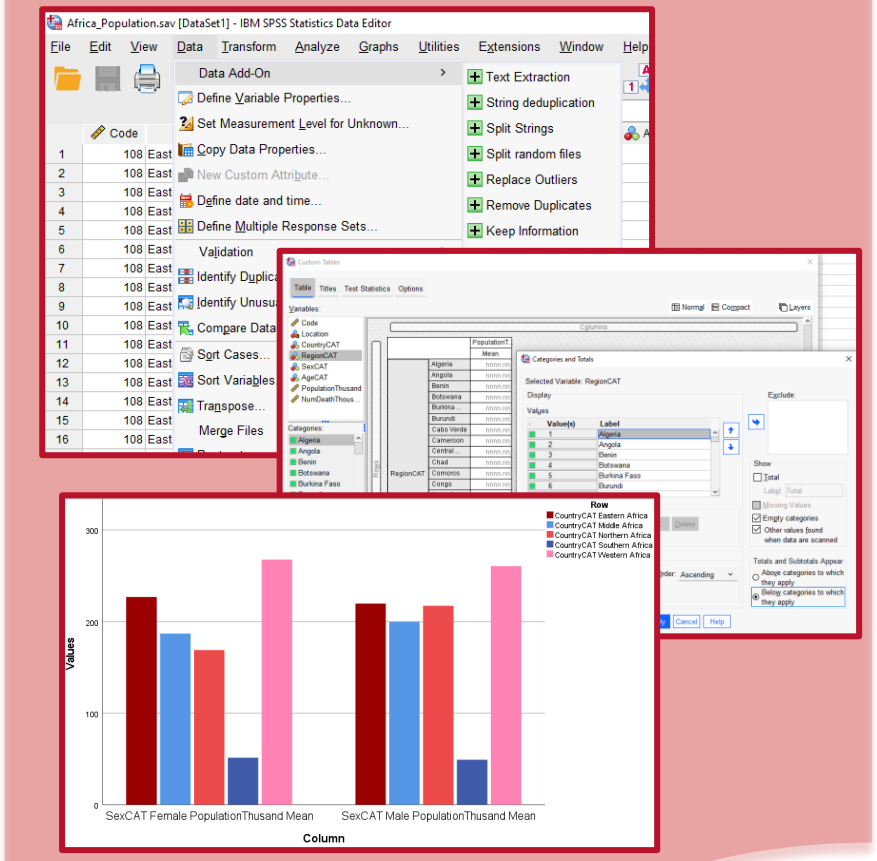
Services included in the Solution (LaunchBox, Add-On, APP, webinar, Support, …)

Linear Regression

- Introduction
- Relations between variables
- Estimation of the Linear Regression Line
- Model Interpretation
- Q&A

Thanks & Regards

Thanks and contacts of the Training & Sales Staff



https://en.statisticsfordataanalysis.com/

Webinar | Agenda

Statistics for Data Analysis

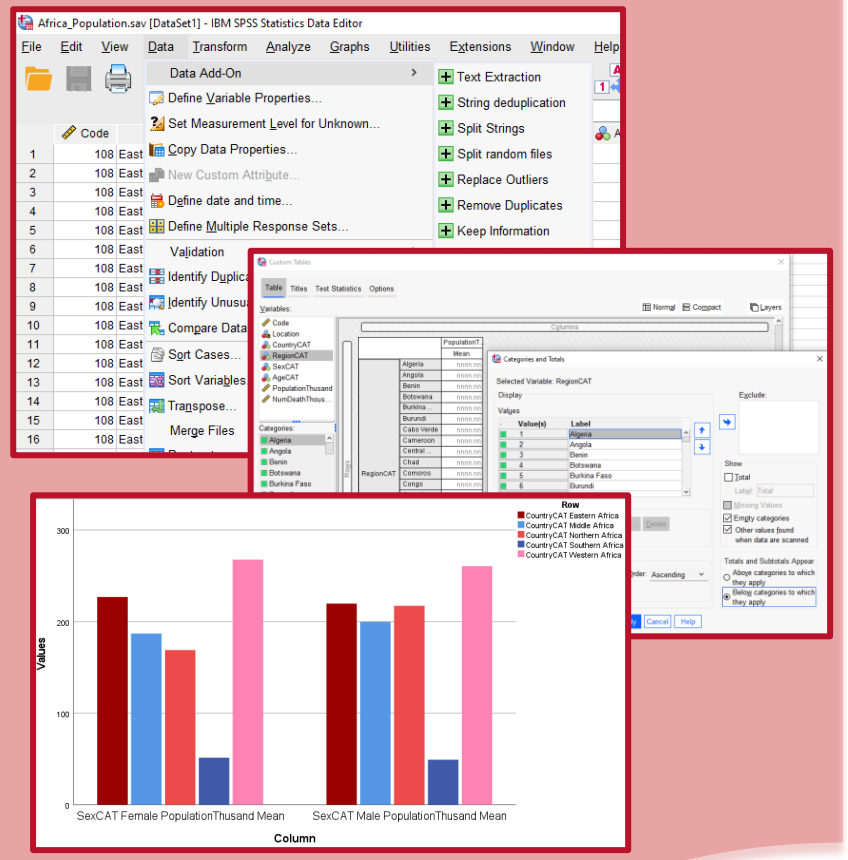
Services included in the Solution (LaunchBox, Add-On, APP, webinar, Support, …)

Linear Regression

- Introduction
- Relations between variables
- Estimation of the Linear Regression Line
- Model Interpretation
- Q&A

Thanks & Regards

Thanks and contacts of the Training & Sales Staff



https://en.statisticsfordataanalysis.com/

SPS | Statistics for Data Analysis

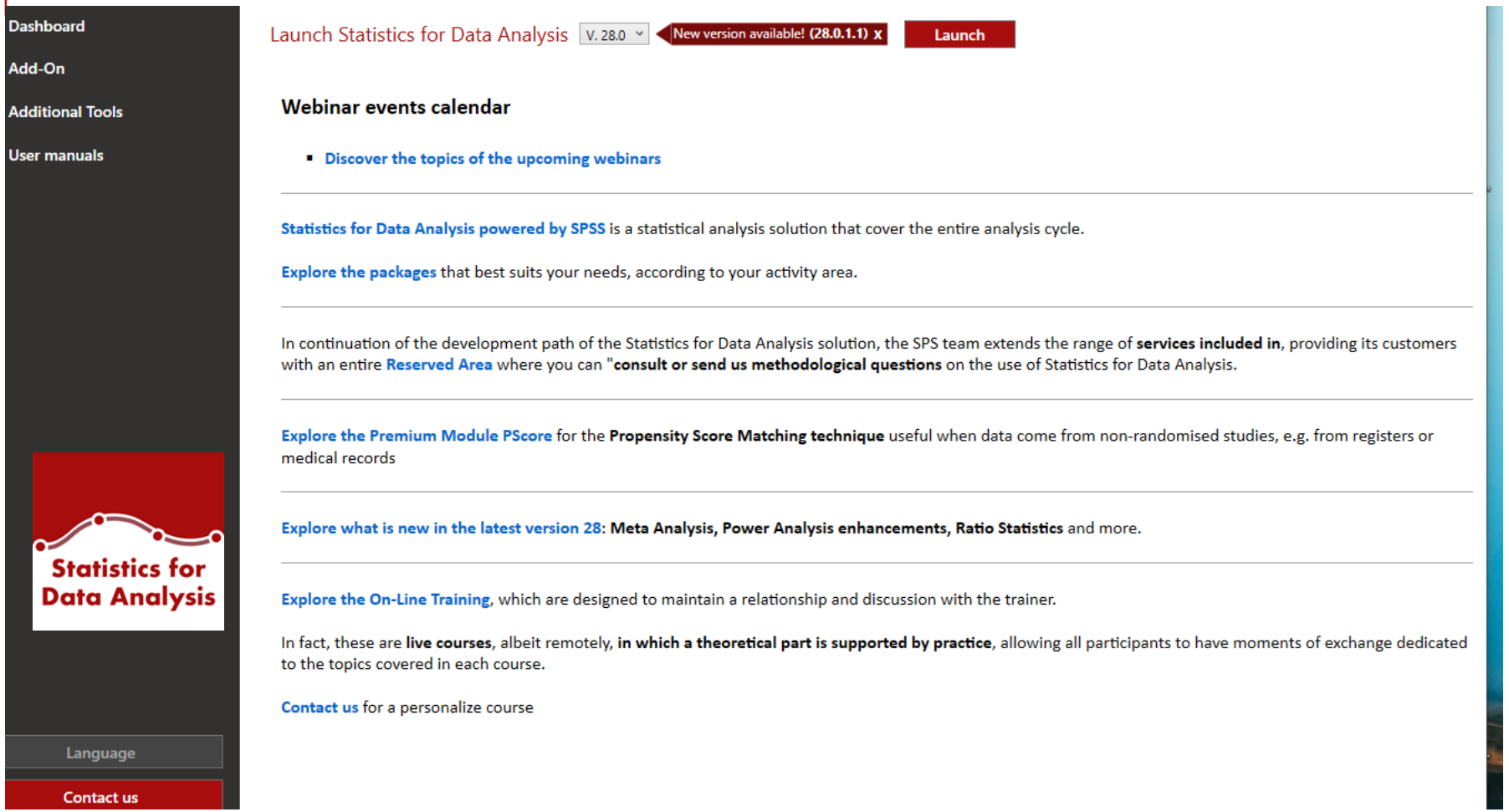
Statistics for Data Analysis is a solution designed and created by a team of professionals of SPS specializing in **Data Analysis**. Includes:

- **Operational tools** to share our data analysis knowledge:
  - LaunchBox
  - Add-On
  - Premium Modules
  - APP
- **Value-Addes Services**
  - Technical support (assistenza@spss.it)
  - OnLine Training
  - On the job Training
  - Webinar
  - Tips & Tricks

- Software licenses powered by **SPSS Statistics**

Operational Tool | LaunchBox

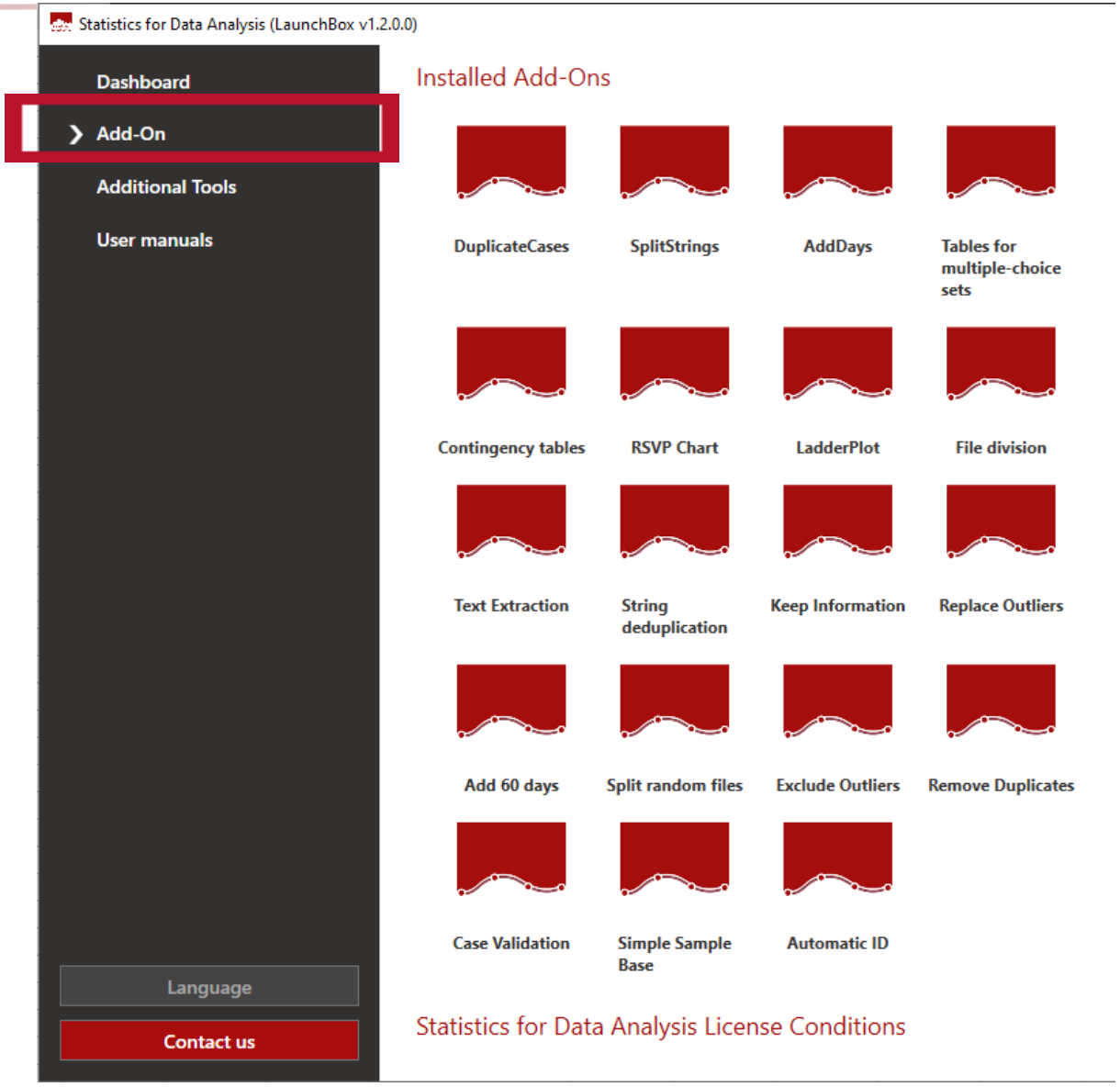
Statistics world-wide dashboard with information on how to install various releases, stand-alone solutions and more



LaunchBox | Add-Ons

Self-installing add-ons:

additional features developed by SPS that strengthen the Statistics for Data Analysis solution, already available natively in the dedicated menus



Linear Regression Introduction

- A quantitative, dependent variable, (Y), whose variability is (can be) explained by another quantitative variable, to which it is related by a linear function.
- A quantitative, explanatory variable (X), which will 'possibly' have an explanatory/predictive effect on Y
- A function that properly relates X vs Y. In the case of Linear Regression, the function will be the straight line

$$Y = a + bx$$

LaunchBox | Customer Area

To enjoy the video content available and

Using the services of the customer area is very easy:

- **Request** access by clicking on the button Login
- **Register by filling in the required fields**

Within 48 working hours our team will provide you with access, after checking that your email is associated with an active contract.

Not a customer but would like to find out how to access the contents of this area?

Contact us

We will show you how!

Webinar | Agenda

Statistics for Data Analysis

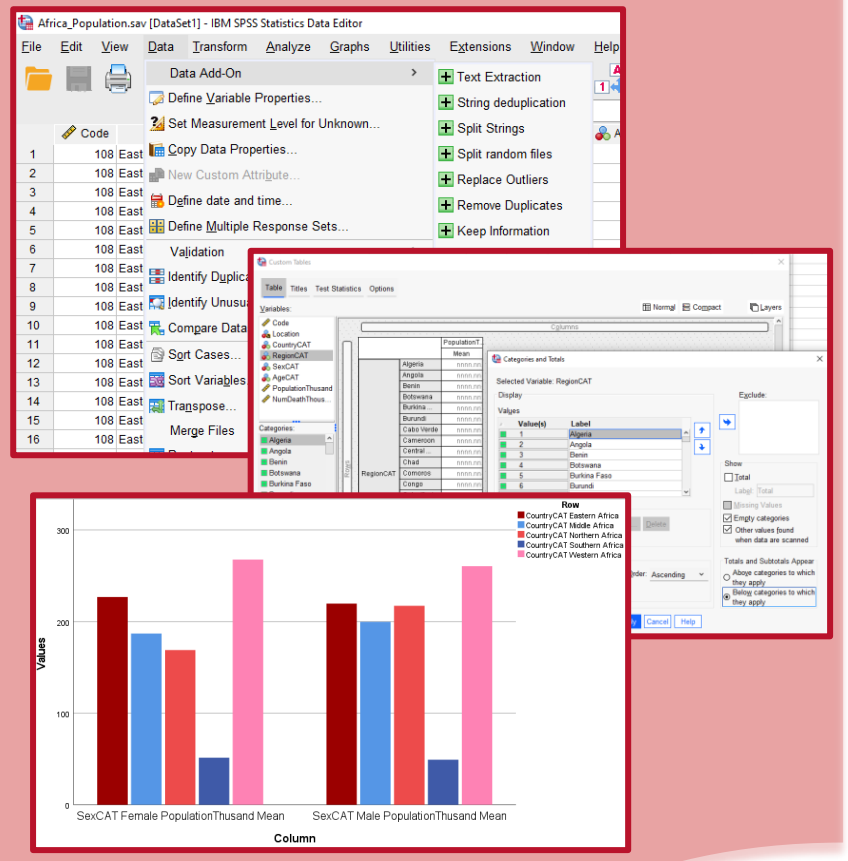
Services included in the Solution (LaunchBox, Add-On, APP, webinar, Support, …)

Linear Regression

- Introduction
- Relations between variables
- Estimation of the Linear Regression Line
- Model Interpretation
- Q&A

Thanks & Regards

Thanks and contacts of the Training & Sales Staff



https://en.statisticsfordataanalysis.com/

Relations between variables

The Simple Linear Regression model translates the influence of one variable on another into functional terms

- ***Descriptive-explanatory purposes:***
  Is there a relationship?
  What is the most appropriate analytical form?

- ***Predictive purposes:***
  Once the model is built, you can predict the mean value that Y assumes given a specific value of X

Simple Linear Regression: Example

Is blood pressure affected by age?

Quantitative dependent variable, normally distributed

Quantitative explanatory variable

Simple Linear Regression Analysis

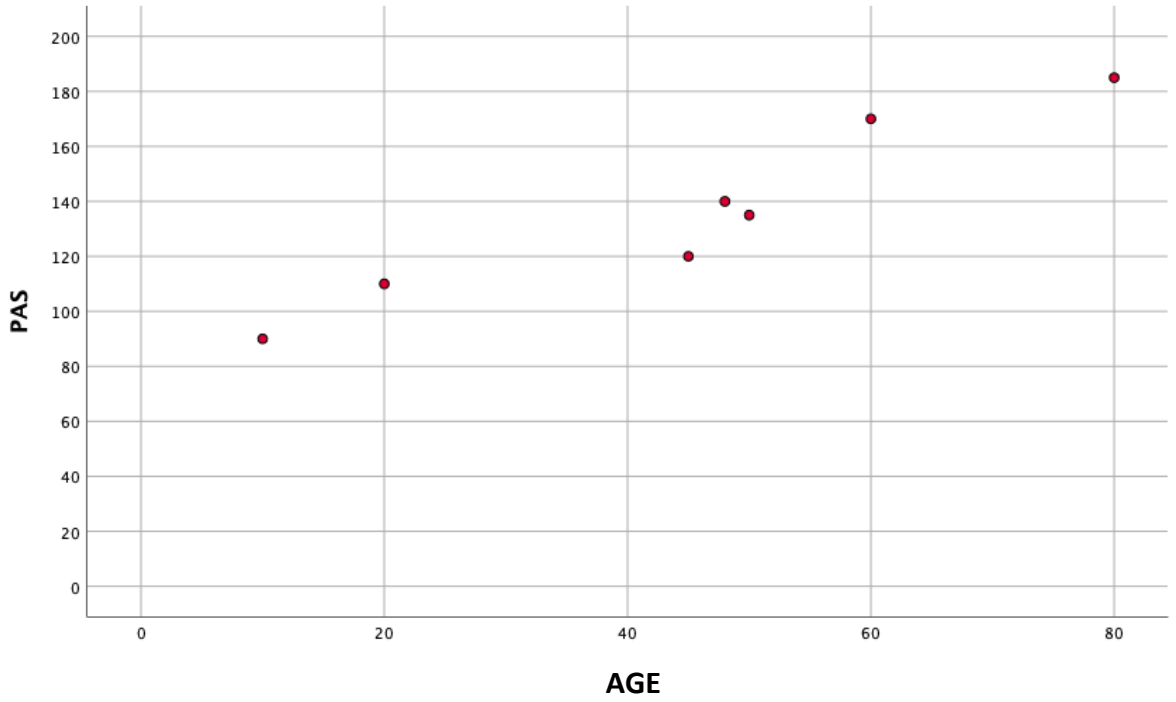
Simple Linear Regression: Example

(2 quantitative variables detected on each unit)

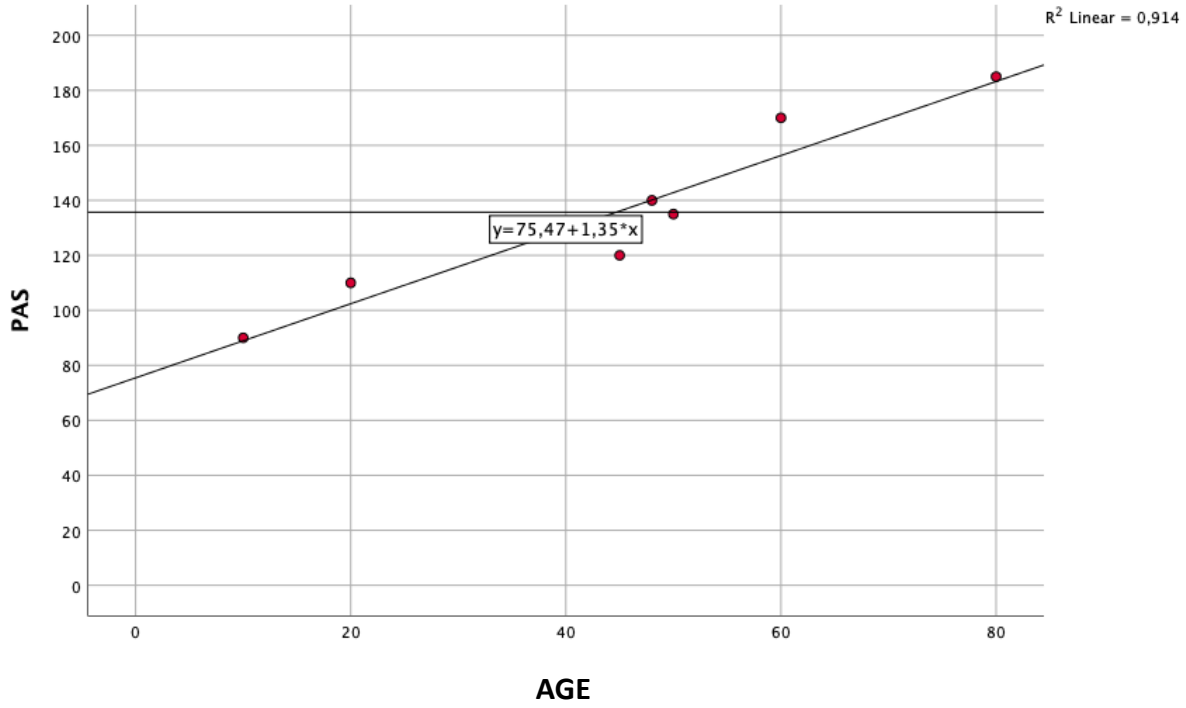
SBP	AGE
120	45
135	50
130	48
170	60
…	…

Simple Linear Regression: Example

Scatterplot



Regression Line X vs Y

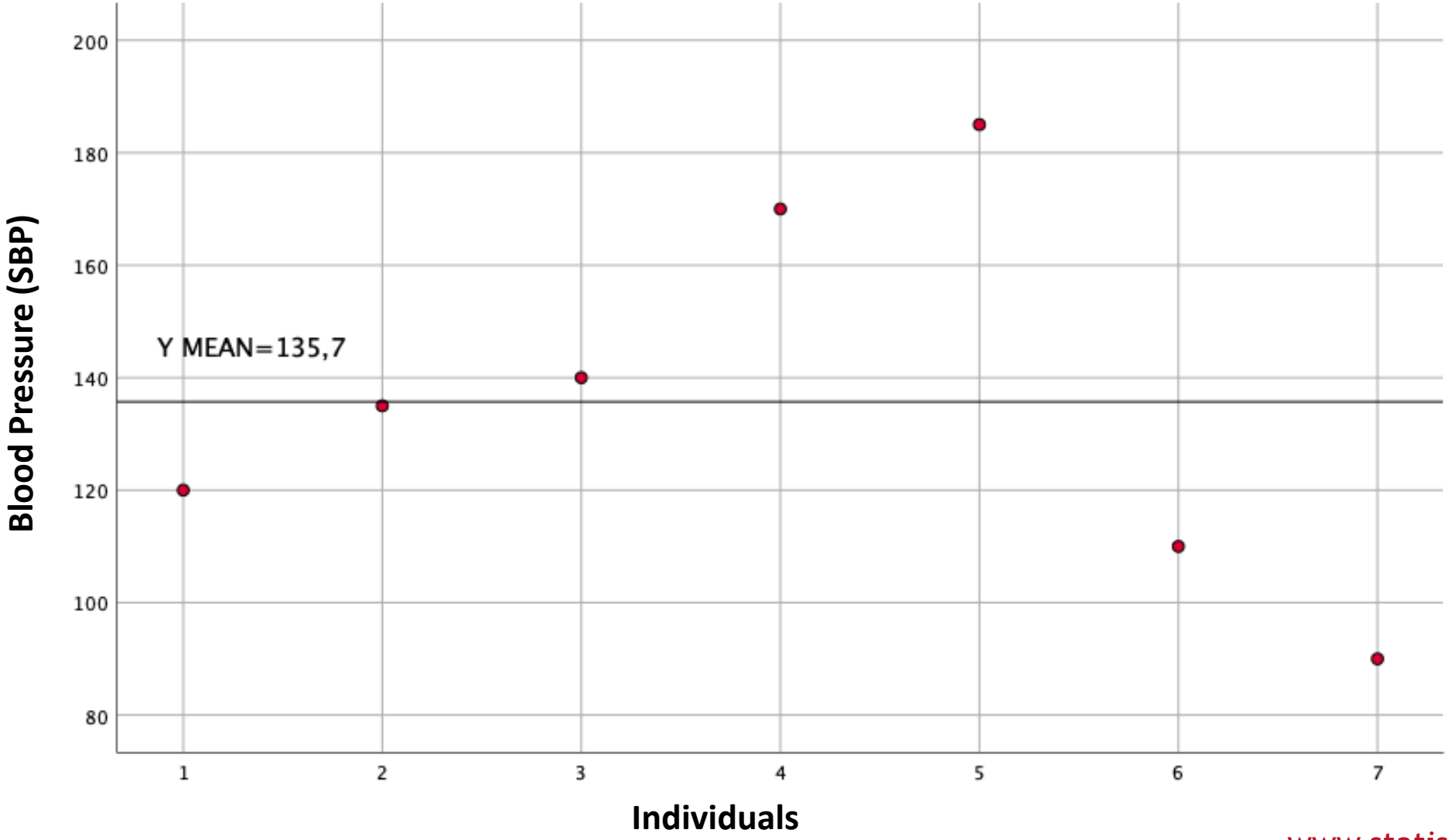


Estimate the Linear Regression Line

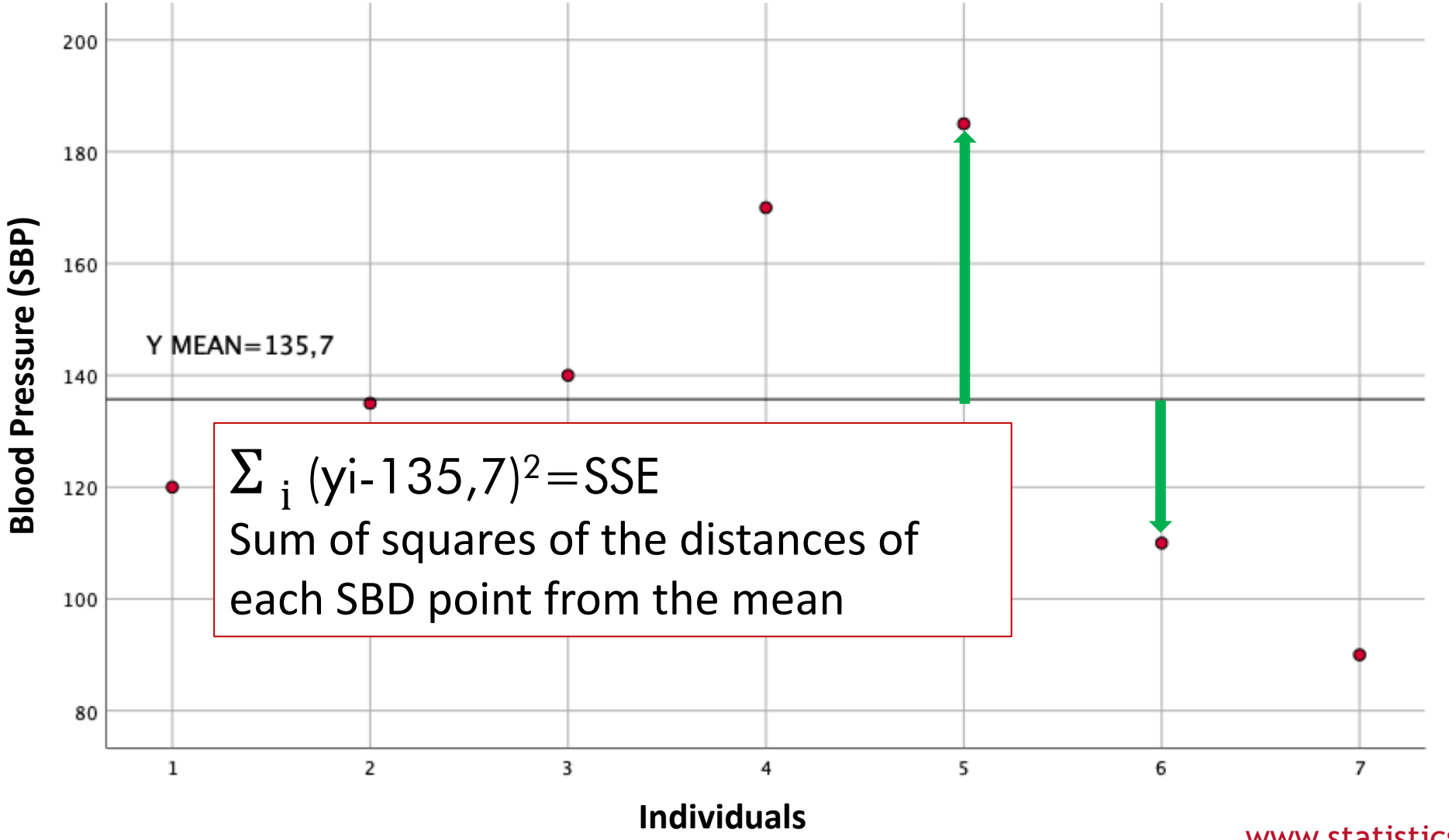
Let's proceed by step

WITH ONLY THE "BLOOD PRESSURE" VARIABLE WHICH WOULD BE THE USEFUL INDEXES TO ASSUME THE VALUE OF THE NEXT DETECTION???

Estimate the Linear Regression Line



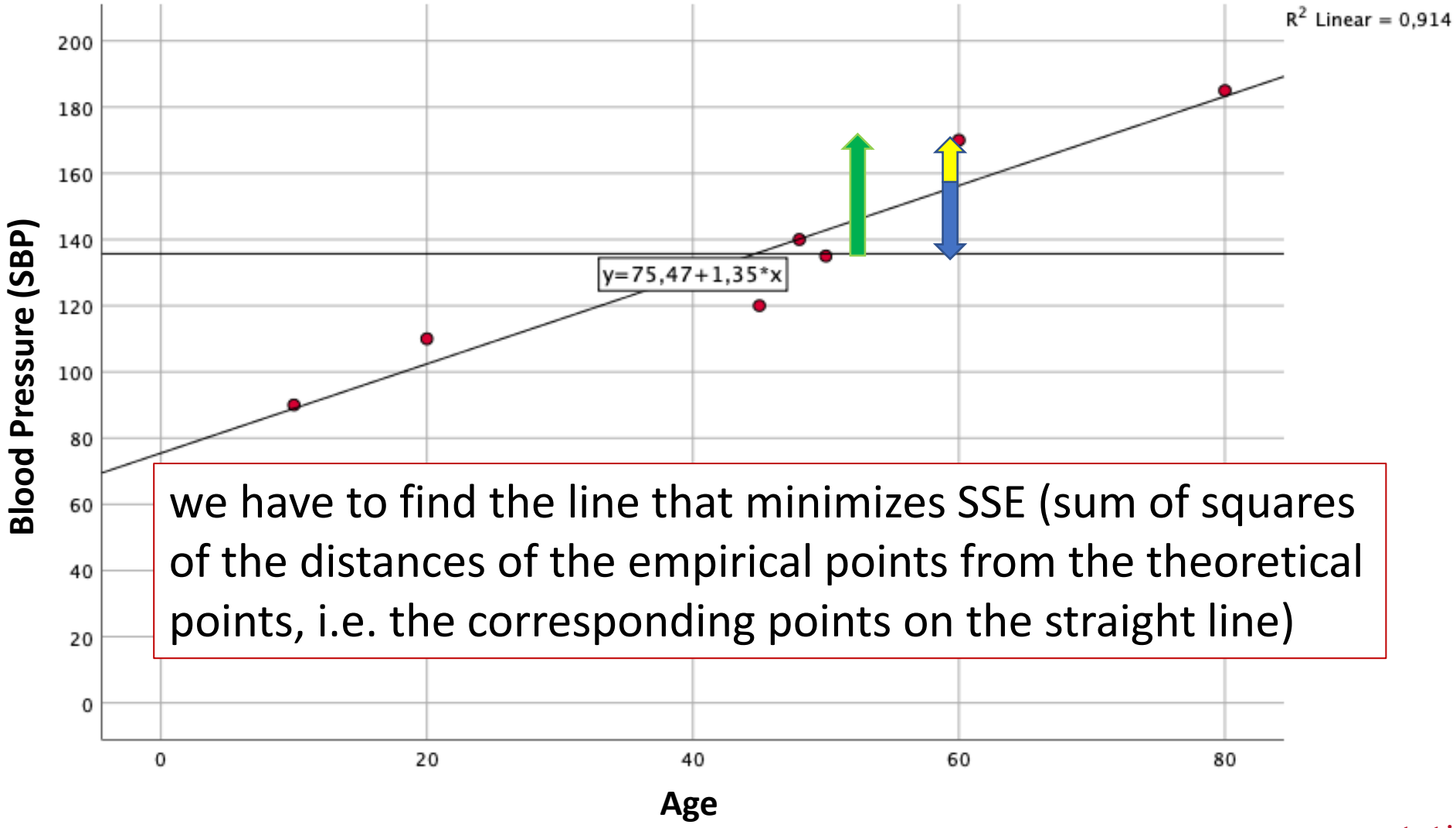
Estimate the Linear Regression Line



$\sum_i (yi-135,7)^2 = SSE$

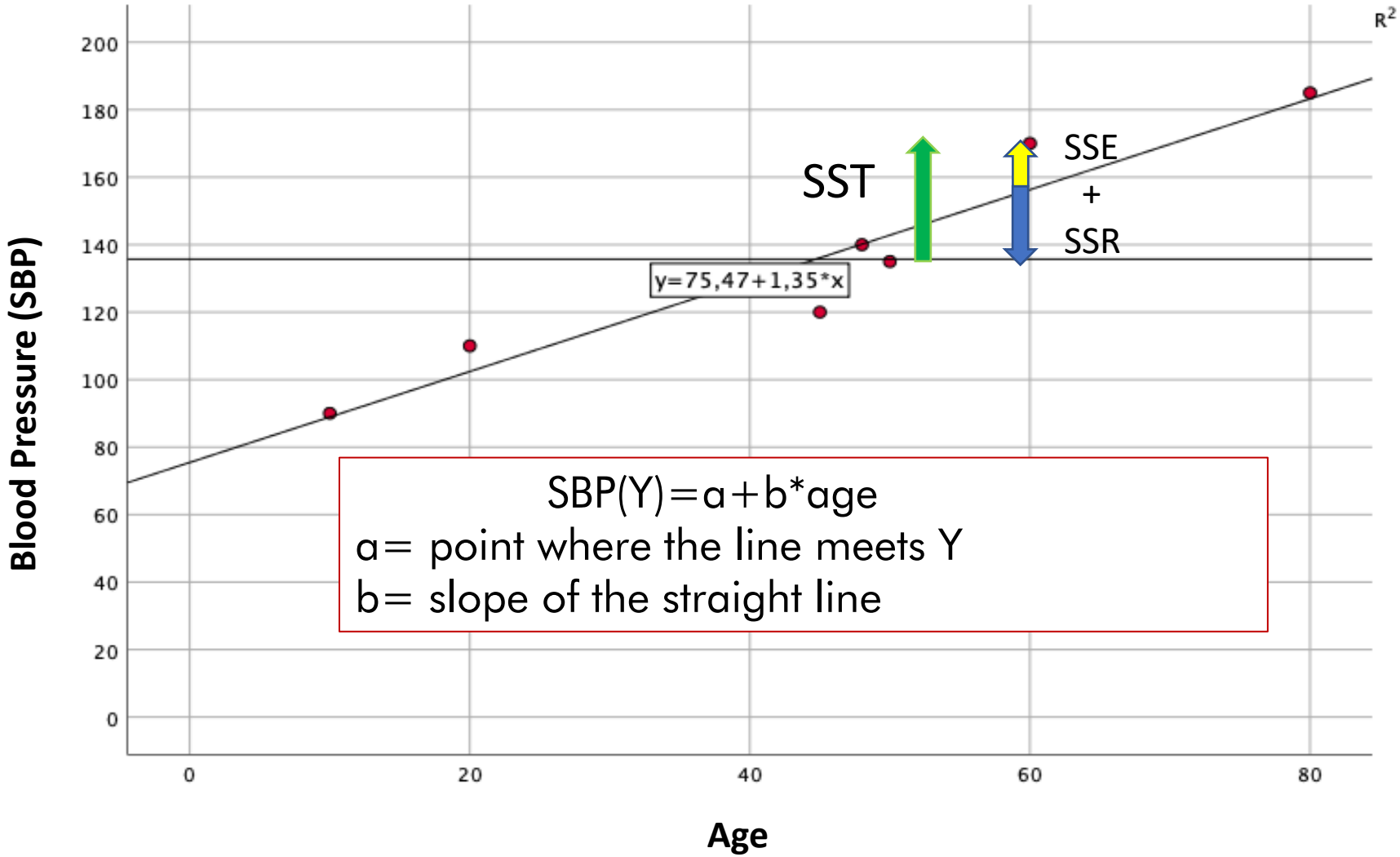
Sum of squares of the distances of each SBD point from the mean

Estimate the Linear Regression Line

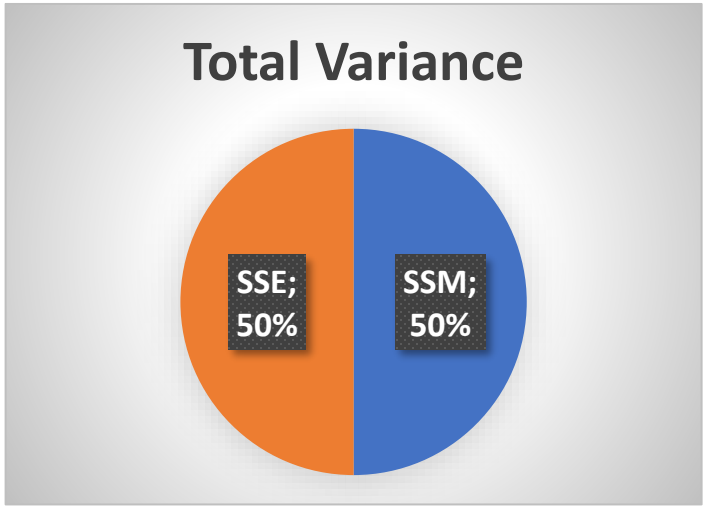


we have to find the line that minimizes SSE (sum of squares of the distances of the empirical points from the theoretical points, i.e. the corresponding points on the straight line)

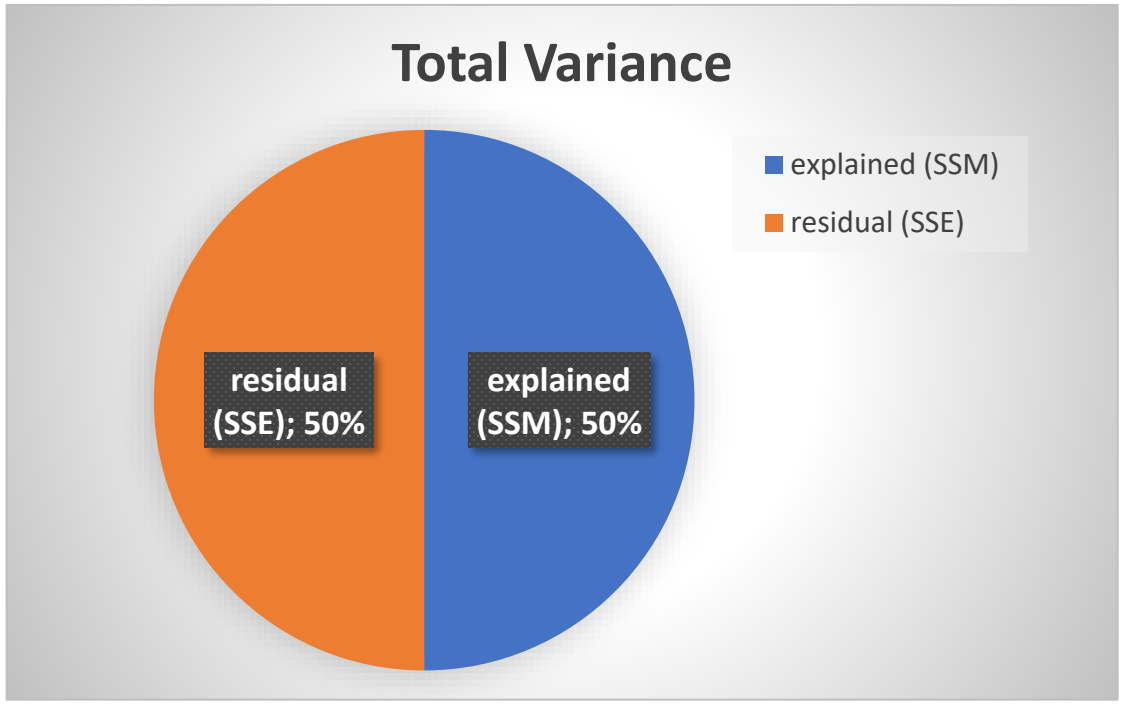
Estimate the Linear Regression Line



$$s^2 = \frac{\sum(x_{ij} - \bar{\bar{x}})^2}{n-1} = \frac{SST}{DFT} = MST$$



Estimate the Linear Regression Line



$$s^2 = \frac{\sum\left(x_{ij} - \bar{\bar{x}}\right)^2}{n-1} = \frac{SST}{DFT} = MST$$

Total deviance (SST) for 32 data points is 300 units.

An x-variable is then regressed against the data, accounting for 150 units of variance (SSM).

What is the R^2?

What is the F ratio?

$$R^2 = SSR/SST = 150/300 = 0.5$$

$$F_{15}^{1,30} = \frac{SSR/1}{SST/(n-2)} = \frac{150/1}{300/30} =$$

The Angular Coefficient b

1) It measures how much Y increases as X changes by unit;

2) It has the same unit of measure as Y;

3) The + sign means that there is an agreement between the 2 variables (as X increases, Y also increases); the – sign means disagreement (as X increases, Y decreases);

4) Its significance is tested using a t-test, with the null hypothesis $H_0 = 0$

5) The confidence interval shows within which values, with probability equal to the value considered for the calculation of CI, you expect to find the data population

Statistics for Data Analysis

The Regression Line in Statistics for Data Analysis

SPSS

Prickly pears

prickly pears.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

1 : edible_perc 64,930000000000 Visible: 24 of 24 Variables

	edible_perc	Density	L	a	b	Tint	Caliber	length	weight	edible_weight	peel_thickness	date	AreaDOPEtna	Agr
1	64,93000000...	1,03520	35,5033333333...	21,1783333333...	10,86000000...	1,950122774708	5,025000000...	8,925000000...	120,6000000...	78,30000000...	0,308333333333	02-Oct-2004	Etna	Belpasso
2	63,30229046...	1,04560	42,2833333333...	14,4666666666...	23,31500000...	0,620487525913	5,383333333...	8,700000000...	132,4333333...	83,83333333...	0,333333333333	02-Oct-2004	Etna	Belpasso
3	62,61450381...	1,04800	35,8866666666...	18,5000000000...	6,893333333...	2,683752417795	5,800000000...	10,82500000...	174,6666666...	109,3666666...	0,316666666667	11-Oct-2004	Etna	Belpasso
4	62,10895361...	1,03600	46,1700000000...	12,1000000000...	28,28333333...	0,427813789039	5,250000000...	9,100000000...	123,6000000...	76,76666666...	0,316666666667	11-Oct-2004	Etna	Belpasso
5	59,78449105...	1,01000	54,7175000000...	-3,7325000000...	31,22250000...	-0,1195451997...	5,350000000...	8,612500000...	129,9250000...	77,67500000...	0,375000000000	21-Aug-2004	Etna	Biancavilla
6	57,43000000...	1,08000	37,1512500000...	19,3050000000...	9,590000000...	2,013034410845	5,630000000...	8,012500000...	144,1250000...	82,77500000...	0,481250000000	21-Aug-2004	Etna	Biancavilla
7	63,89000000...	1,07000	51,8925000000...	9,465000000000	27,16000000...	0,348490427099	5,087500000...	7,831250000...	116,8750000...	74,67500000...	0,333750000000	21-Aug-2004	Etna	Biancavilla
8	62,14504596...	1,19920	61,2625000000...	-2,4850000000...	33,51750000...	-0,0741403744...	5,250000000...	8,193750000...	122,3750000...	76,05000000...	0,422500000000	27-Aug-2004	Etna	Biancavilla
9	52,24219880...	1,05600	37,8137500000...	14,7887500000...	8,062500000...	1,834263565891	5,162500000...	9,200000000...	120,9750000...	63,20000000...	0,457142857143	27-Aug-2004	Etna	Biancavilla
10	54,03093721...	1,04560	44,3812500000...	11,4212500000...	24,97125000...	0,457375982380	4,975000000...	8,200000000...	109,9000000...	59,38000000...	0,500000000000	27-Aug-2004	Etna	Biancavilla
11	59,36796614...	1,04000	54,4837500000...	9,172500000000	31,70875000...	0,289273465526	5,262500000...	8,025000000...	138,8375000...	82,42500000...	0,437500000000	04-Sep-2004	Etna	Biancavilla
12	57,16000000...	1,05920	40,1012500000...	18,2575000000...	7,283750000...	2,506607173503	5,776250000...	8,148750000...	152,2250000...	87,05000000...	0,570000000000	04-Sep-2004	Etna	Biancavilla
13	62,92000000...	1,06720	60,8862500000...	-1,9912500000...	29,20625000...	-0,0681789000...	5,856250000...	8,787500000...	158,6500000...	99,82500000...	0,412500000000	04-Sep-2004	Etna	Biancavilla
14	64,78904351...	1,06320	38,5162500000...	23,9937500000...	6,336250000...	3,786742947327	5,187500000...	7,662500000...	113,1750000...	73,32500000...	0,343750000000	20-Sep-2004	Etna	Biancavilla
15	66,86000000...	1,06000	61,7762500000...	0,756250000000	26,26125000...	0,028797182160	6,050000000...	8,606250000...	155,6250000...	104,0500000...	0,318750000000	20-Sep-2004	Etna	Biancavilla
16	63,57000000...	1,06000	52,6150000000...	14,3600000000...	23,78000000...	0,603868797309	5,400000000...	8,087500000...	131,3250000...	83,48750000...	0,387500000000	20-Sep-2004	Etna	Biancavilla
17	59,73000000...	1,04000	38,1350000000...	18,6666666666...	22,77166666...	0,819732123253	5,466666666...	9,341666666...	141,0333333...	84,23333333...	0,475000000000	28-Sep-2004	Etna	Biancavilla
18	57,71000000...	1,02640	29,2783333333...	17,9833333333...	8,388333333...	2,143850586132	5,166666666...	9,266666666...	126,1500000...	72,80000000...	0,416666666667	28-Sep-2004	Etna	Biancavilla
19	64,21000000...	1,05840	133,366666666...	8,030000000000	19,38833333...	0,414166595031	5,366666666...	8,283333333...	133,3666666...	85,63333333...	0,383333333333	28-Sep-2004	Etna	Biancavilla
20	73,21000000...	1,04000	55,3083333333...	6,665000000000	36,47833333...	0,182711198428	5,616666666...	9,116666666...	143,3333333...	104,9333333...	0,266666666667	11-Oct-2004	Etna	Biancavilla
21	66,31764854...	1,03000	40,1566666666...	14,2333333333...	13,52833333...	1,052112849575	5,350000000...	9,150000000...	132,1166666...	87,61666666...	0,366666666667	11-Oct-2004	Etna	Biancavilla

Data View Variable View

Statistics for Data Analysis

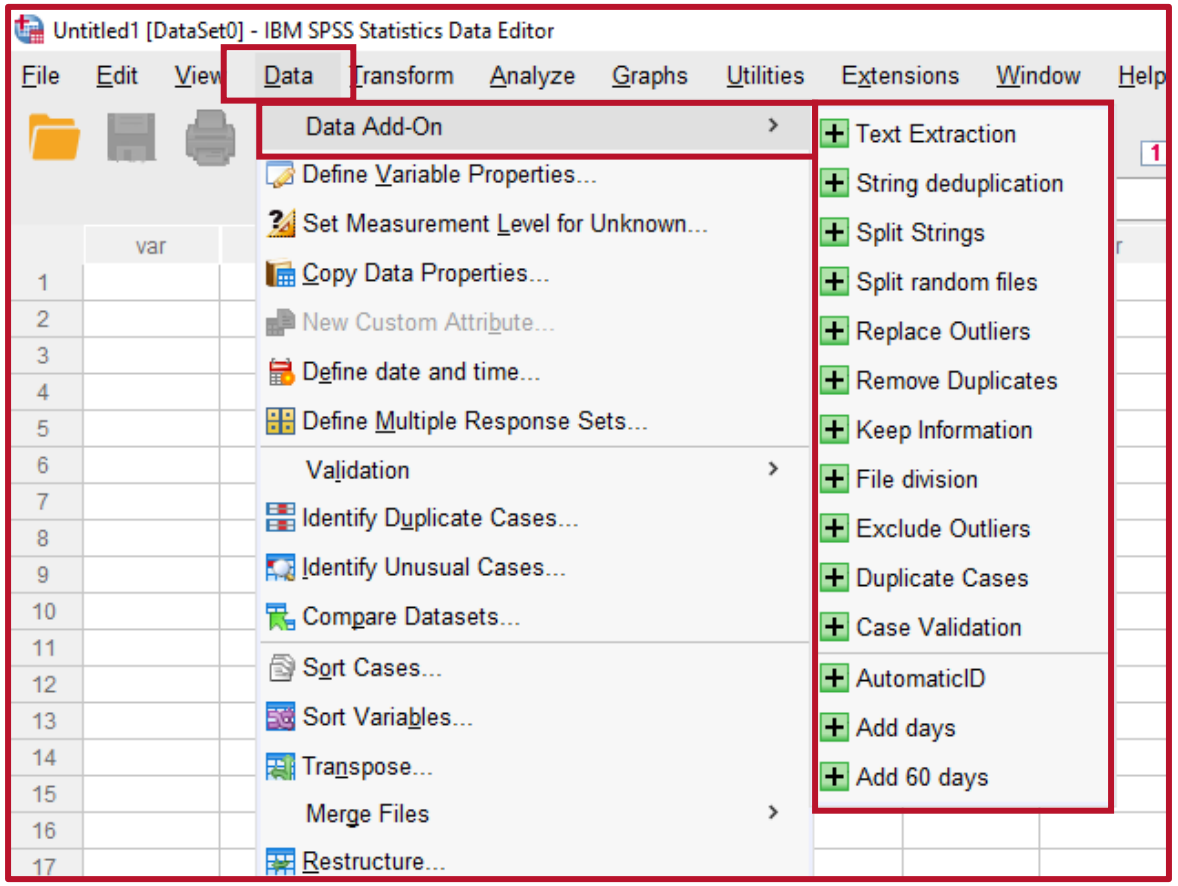
Before running the Regression Model…

SPSS

Before starting the Analysis, you need to prepare the Dataset.

In the menu Data, you can find several **Add-Ons** referring to the Data Preparation phase.

Add-Ons are **Integrating** procedures created and suggested by the **collaboration** with our customers, to **facilitate** and **automate** your analysis activities.

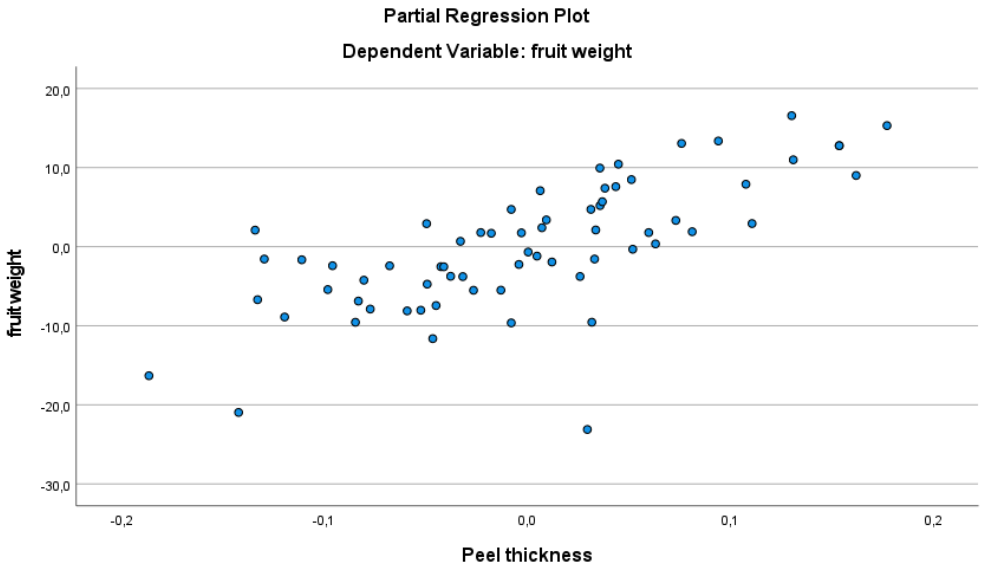


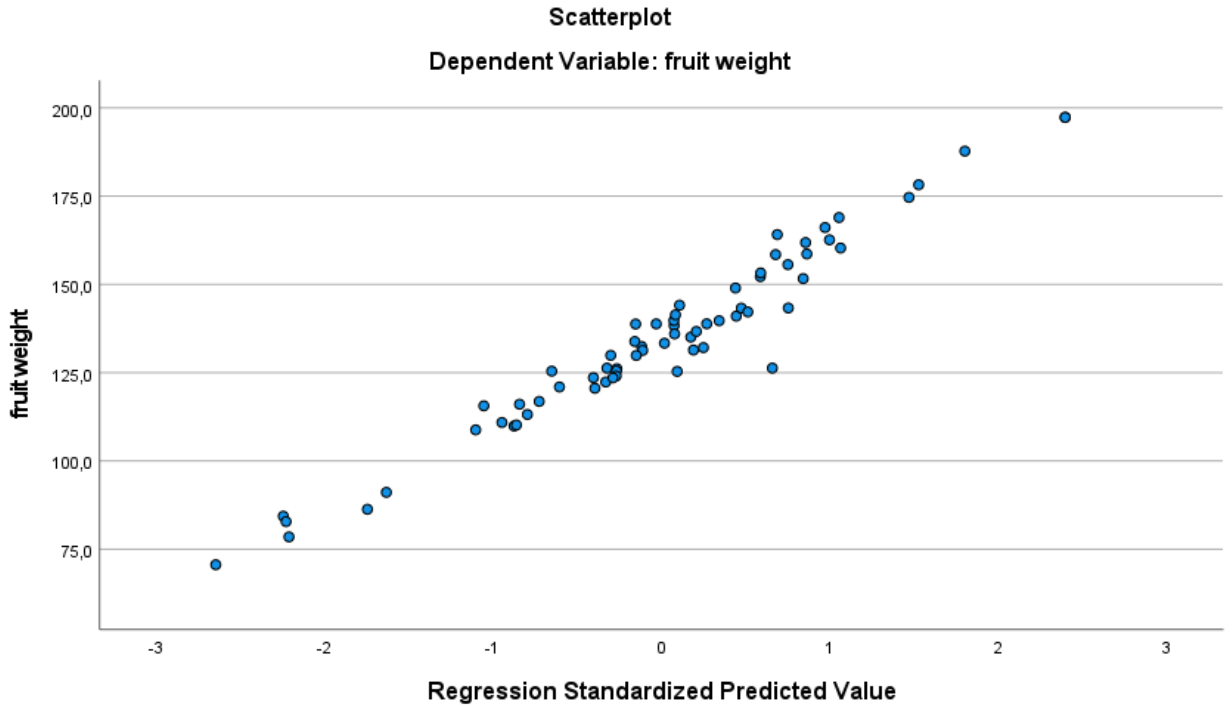
Linear Regression Output

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13,479	5,439		2,478	0,016
	Edible weight	1,463	0,064	0,943	22,762	<0.001
2	(Constant)	-3,530	5,252		-0,672	0,504
	Edible weight	1,386	0,054	0,893	25,846	<0.001
	Peel thickness	63,407	10,680	0,205	5,937	<0.001
3	(Constant)	-38,029	7,387		-5,148	<0.001
	Edible weight	1,168	0,058	0,753	20,170	<0.001
	Peel thickness	72,515	8,867	0,235	8,178	<0.001
	length	5,491	0,957	0,208	5,737	<0.001

a. Dependent Variable: fruit weight





Webinar | Agenda

Statistics for Data Analysis

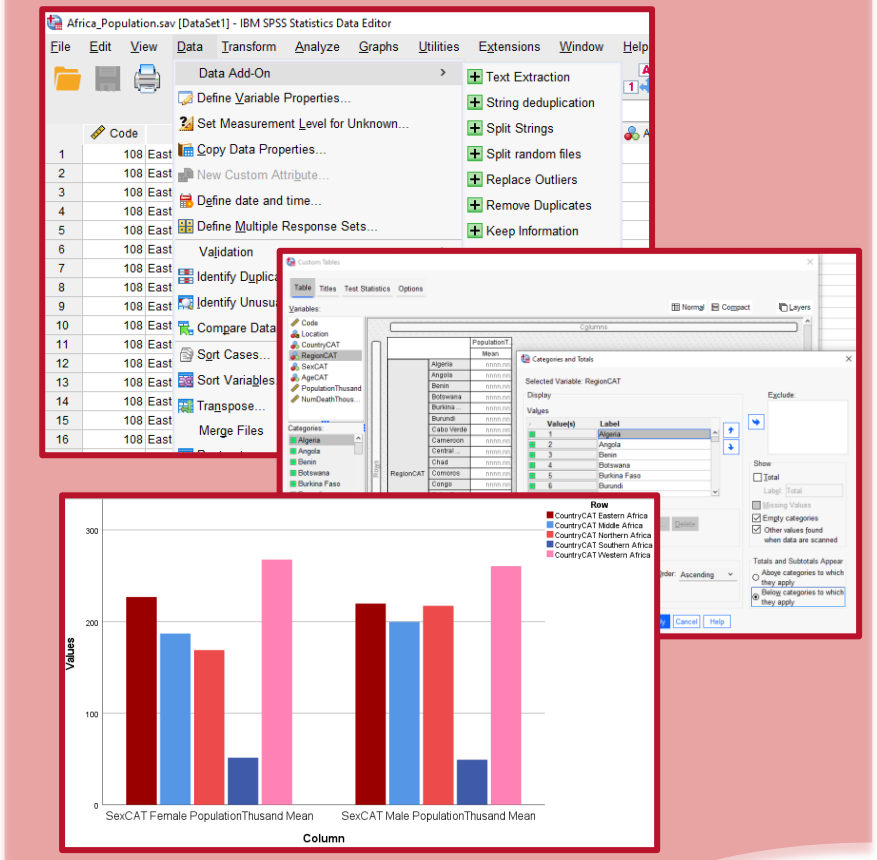
Services included in the Solution (LaunchBox, Add-On, APP, webinar, Support, …)

Linear Regression

- Introduction
- Relations between variables
- Estimation of the Linear Regression Line
- Model Interpretation
- Q&A

Thanks & Regards

Thanks and contacts of the Training & Sales Staff



https://en.statisticsfordataanalysis.com/

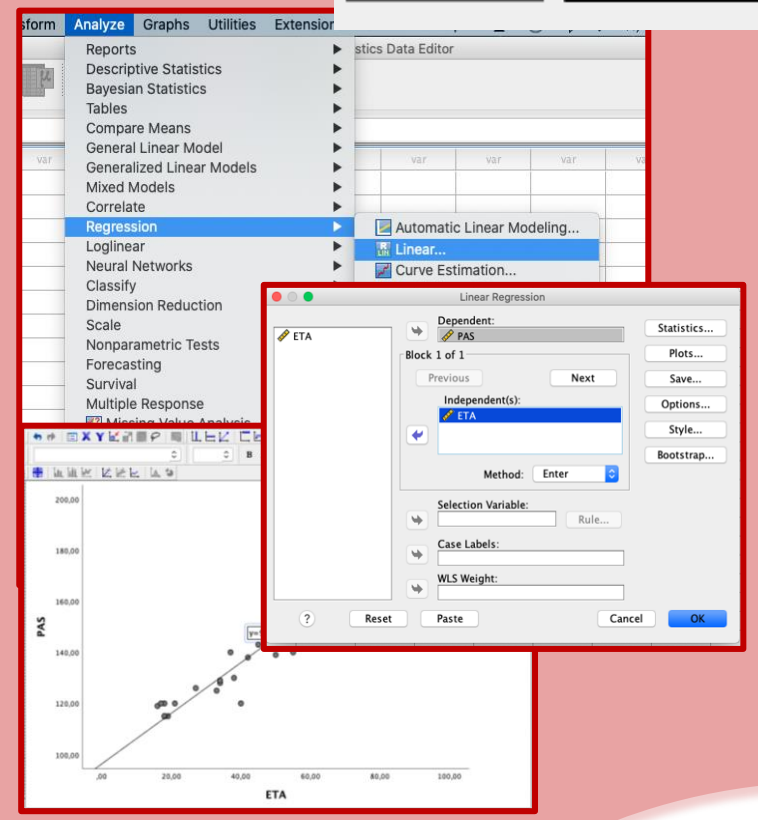
Statistics for Data Analysis

QUESTIONS & ANSWERS

SPSS

Webinar | Linear Regression

SPSS



@Sales

THANKS

Sales & Training Teams @SPS