

## ΠΑΛΙΝΔΡΟΜΗΣΗ (ΜΙΑ ΕΙΣΑΓΩΓΗ)

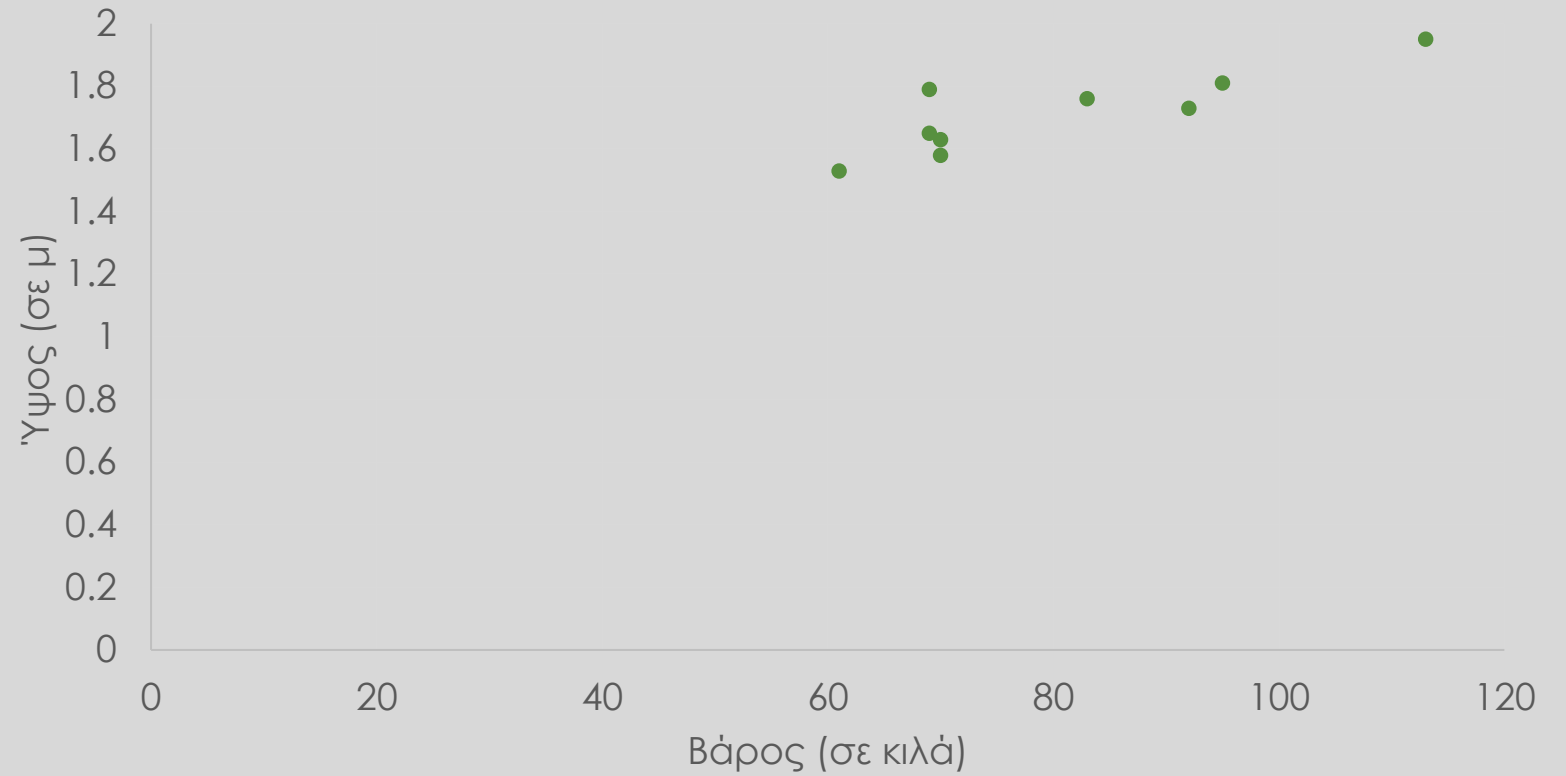
Ανεξάρτητη/εξαρτημένη μεταβλητή,  
ντετερμινιστική/στατιστική σχέση, αριθμός παρατηρήσεων,  
μηδενική/εναλλακτική υπόθεση, έλεγχος κανονικότητας

# Ορολογία και σημαντικές έννοιες

- **Ποσοτικές** λέγονται οι **μεταβλητές** οι οποίες παίρνουν μόνο αριθμητικές τιμές και μπορούν να ταξινομηθούν σε διακριτές (π.χ. αριθμός παιδιών ανά οικογένεια) ή συνεχείς (π.χ. βάρος).
- Το  $\alpha$  (**άλφα**) είναι το **επίπεδο σημαντικότητας** που χρησιμοποιείται για τον υπολογισμό του **επιπέδου εμπιστοσύνης**. Συνήθως  $\alpha = 0,05$  (επιπέδου εμπιστοσύνης =  $0,95$  (95%))
- Ο **αριθμός παρατηρήσεων** συμβολίζεται με  $n$ .
- **Οπτικοποίηση δεδομένων (διαγράμματα διασποράς)**
- **Έλεγχος κανονικότητας**
- Η **απλή γραμμική παλινδρόμηση** είναι μια προσέγγιση για τη μοντελοποίηση της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής  $Y$  και μίας ανεξάρτητης μεταβλητής  $X$ .
- Η σχέση αυτή λέγεται **στατιστικά σημαντική** εάν η πιθανότητα ισχύς της εναλλακτικής υπόθεσης είναι  $> 95\%$ . Τότε η πιθανότητα να ισχύει η  $H_0$  είναι μικρότερη του επιπέδου σημαντικότητας  $\alpha$  ( $< 0,05$ ) και την απορρίπτουμε.

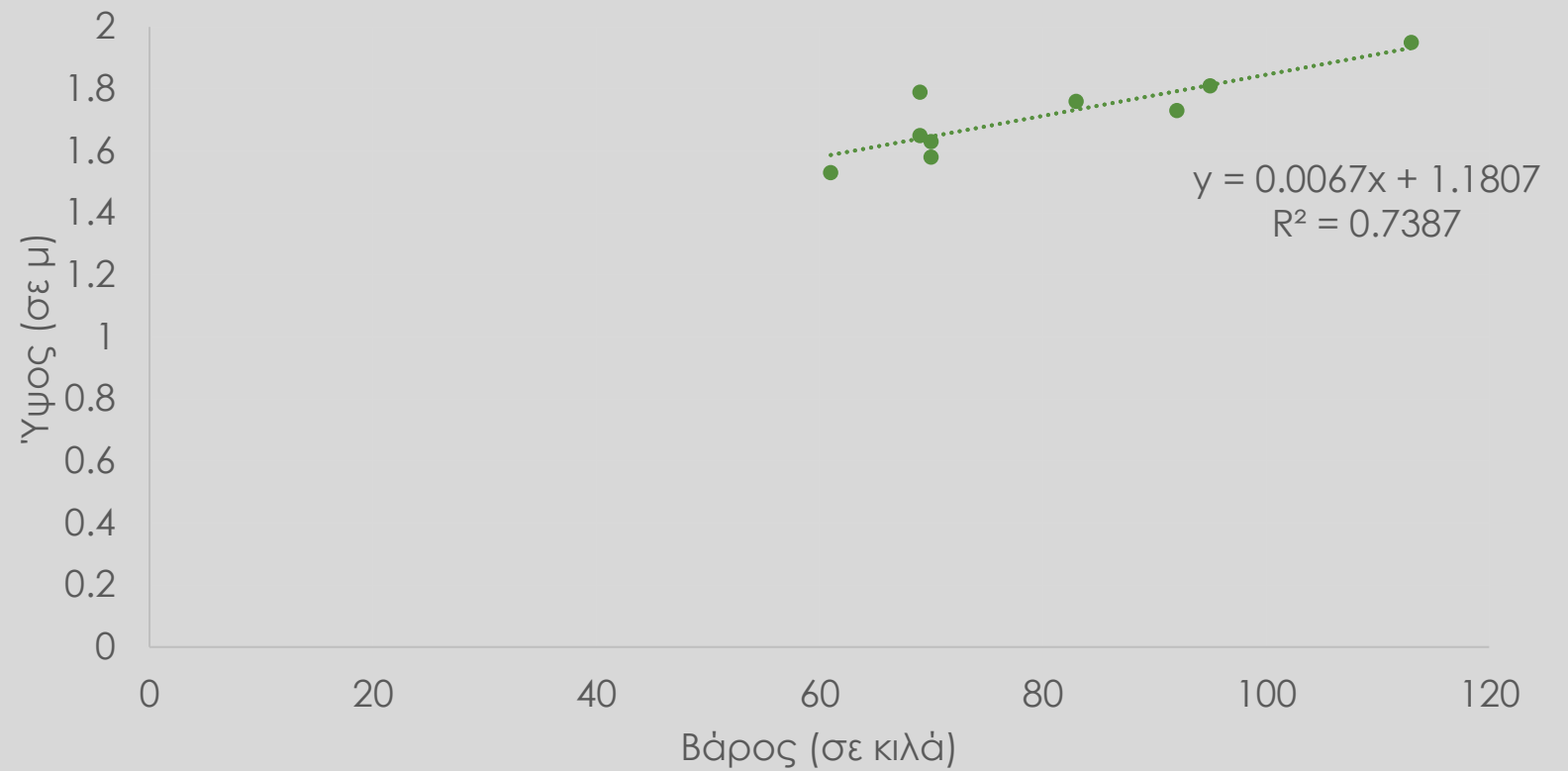
# Οπτικοποίηση δεδομένων

x	y
Βάρος σε (κιλά)	Ύψος (σε μ)
95	1.81
83	1.76
61	1.53
92	1.73
69	1.65
70	1.58
113	1.95
69	1.79
70	1.63



# Οπτικοποίηση δεδομένων

x	y
Βάρος σε (κιλά)	Ύψος (σε μ)
95	1.81
83	1.76
61	1.53
92	1.73
69	1.65
70	1.58
113	1.95
69	1.79
70	1.63



# Απλή γραμμική παλινδρόμηση

$$y_i = \alpha + \beta x_i$$

$y_i$  – εξαρτημένη μεταβλητή, “dependent” variable

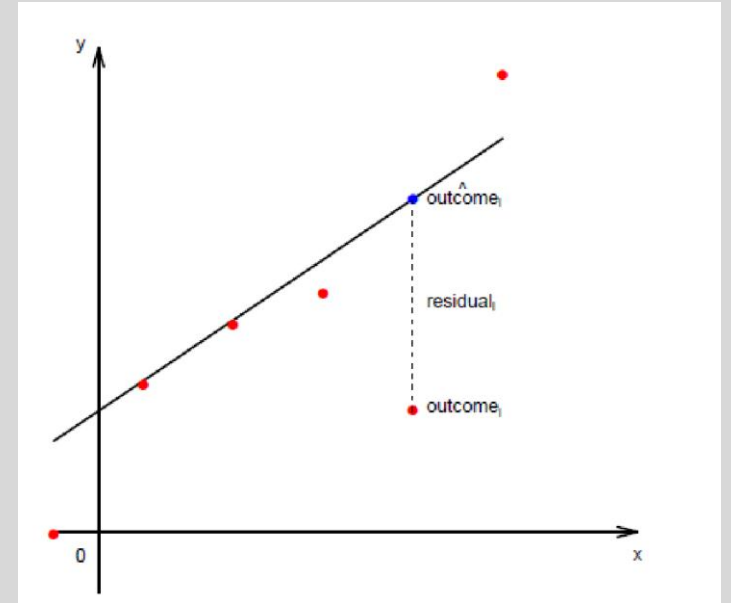
$\alpha$  – σημείο τομής με τον άξονα των  $y$ , intercept

$\beta$  – συντελεστής παλινδρόμησης (κλίση), regression coefficient (slope)

$x_i$  – ανεξάρτητη μεταβλητή, predictor, regressor

$\varepsilon_i$  – σφάλματα, υπόλοιπα (κατάλοιπα), errors, residuals.

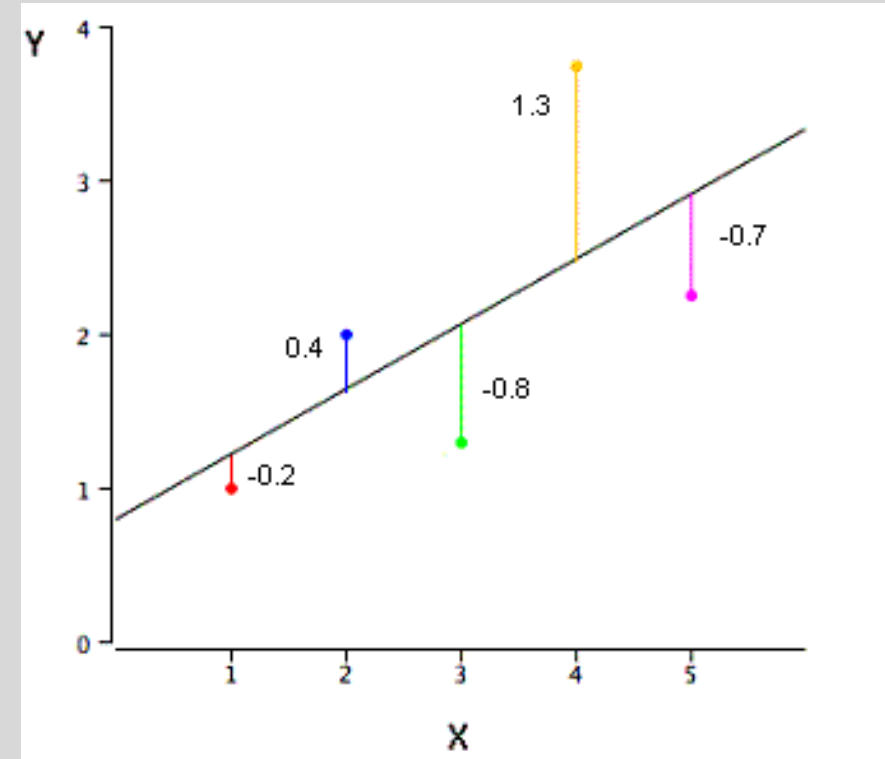
$$y_i = \alpha + \beta x_i + \varepsilon_i$$



# Υπολογισμός των παραμέτρων

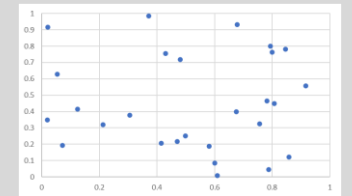
- Γίνεται με τη μέθοδο των ελαχίστων τετραγώνων και σκοπός είναι η ελαχιστοποίηση του αθροίσματος των τετραγώνων των υπολοίπων:

$$SS_{res} = \sum_i (y_i - (\alpha + \beta x_i))^2$$



# Σημαντικότητα των παραμέτρων

- Προφανώς, οι παράμετροι της παλινδρόμησης θα ποικίλουν όταν πάρουμε διαφορετικά δείγματα.
- Είναι συνεπώς σημαντικό να υπολογίσουμε τη σημαντικότητα των παραμέτρων του μοντέλου.
- Θα ελέγξουμε τη μηδενική υπόθεση να δούμε εάν  $\mathbf{b} = \mathbf{0}$  (δηλ. οριζόντια γραμμή)
- Αυτό γίνεται με ένα t-test
- Εάν  $p < 0,05$  απορρίπτεται η μηδενική Υπόθεση
- Το  $R^2$  (μεταξύ 0 και 1) είναι ο συντελεστής προσδιορισμού και δείχνει πόση διασπορά εξηγείται από το μοντέλο.



$$t = \frac{\hat{\beta} - \beta_0}{S_{\hat{\beta}}} = \frac{\hat{\beta}}{S_{\hat{\beta}}}$$

# Παράδειγμα (ανάλυση με excel/calc)

- Ερώτημα:
- Ισχύει η μηδενική υπόθεση  $H_0$  ( $b=0$ );
  - Εάν ναι ( $p>0.05$ ), τότε δεν υπάρχει στατιστικά σημαντική σχέση μεταξύ ύψους και βάρους
- Ισχύει η εναλλακτική υπόθεση  $H_1$  ( $b\neq 0$ );
  - Εάν ναι ( $p<0.05$ ), τότε υπάρχει στατιστικά σημαντική σχέση μεταξύ ύψους και βάρους
    - Που είναι σημείο τομής με τον άξονα των  $y$  ( $a$ , intercept);
    - Τιμή του συντελεστή παλινδρόμησης ( $b$ , κλίση), regression coefficient (slope);
    - Τιμή του συντελεστή προσδιορισμού ( $R^2$ ); Πόση διασπορά εξηγείται από το μοντέλο;

x	y
Βάρος σε (κιλά)	Ύψος (σε μ)
95	1.81
83	1.76
61	1.53
92	1.73
69	1.65
70	1.58
113	1.95
69	1.79
70	1.63



Save Off

File Home Insert Page Layout Formulas Data Review

Cut Copy Paste Format Painter

Clipboard

Calibri 11

B I U

Q22

	A	B	C	D	E	F
1						
2						
3		Υψος (σε μ)	Βάρος σε (κιλά)			
4		1.81	95			
5		1.76	83			
6		1.53	61			
7		1.73	92			
8		1.65	69			
9		1.58	70			
10		1.95	113			
11		1.79	69			
12		1.63	70			
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25					120	
26					100	
27					80	
28						
29						

Sheet4 Sheet5 Sheet6 Sheet7 Sheet1

Home

New

Open

Info

Save

Save As

Print

Share

Export

Publish

Close

Account

Feedback

Options

Good

New

Recent

Clipboard: Paste, Cut, Copy, Format Painter

Font: Calibri (Body) 10, Bold, Italic, Underline

Chart 3

	A	B	C	D
7		1.73	92	
8		1.65	69	
9		1.58	70	
10		1.95	113	
11		1.79	69	
12		1.63	70	
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				

### Excel Options

- General
- Formulas
- Data
- Proofing
- Save
- Language
- Ease of Access
- Advanced
- Customize Ribbon
- Add-ins**

### View and manage Microsoft Office Add-ins.

Name	Location
<b>Active Application Add-ins</b>	
Analysis ToolPak	C:\...Office16\Library\Analysis\ANALYS32.XLL
<b>Inactive Application Add-ins</b>	
Analysis ToolPak - VBA	C:\...ice16\Library\Analysis\ATPVBAEN.XLAM
Date (XML)	C:\...s\Microsoft Shared\Smart Tag\MOFL.DLL
Euro Currency Tools	C:\...root\Office16\Library\EUROTOOL.XLAM
Inquire	C:\...soft Office\Office16\DCF\NativeShim.dll
Microsoft Actions Pane 3	
Microsoft Data Streamer for Excel	C:\...softDataStreamerforExcel.vsto\vstolocal
Microsoft Power Map for Excel	C:\...p Excel Add-in\EXCELPLUGINSHELL.DLL
Microsoft Power Pivot for Excel	C:\...l Add-in\PowerPivotExcelClientAddIn.dll
Microsoft Power View for Excel	C:\...l Add-in\AdHocReportingExcelClient.dll
Solver Add-in	C:\...Office16\Library\SOLVER\SOLVER.XLAM
<b>Document Related Add-ins</b>	
No Document Related Add-ins	

Add-in: Analysis ToolPak  
Publisher: Microsoft Corporation  
Compatibility: No compatibility information available  
Location: C:\Program Files\Microsoft Office\root\Office16\Library\Analysis\ANALYS32.XLL  
Description: Provides data analysis tools for statistical and engineering analysis

Manage: Excel Add-ins **Go...**

### Add-ins

Available Add-ins available:

- Analysis ToolPak
- Euro Currency Tools
- Solver Add-in

**OK**

Cancel

Browse...

Automation...

---

### Analysis ToolPak

Provides data analysis tools for statistical and engineering analysis

AutoSave  Off stats  Thomas Tscheulin TT

File Home Insert Page Layout Formulas **Data** Review View Help

Get Data: From Text/CSV, From Web, From Table/Range, Recent Sources, Existing Connections

Queries & Connections: Refresh, Properties, Edit Links

Data Types: Stocks, Geography

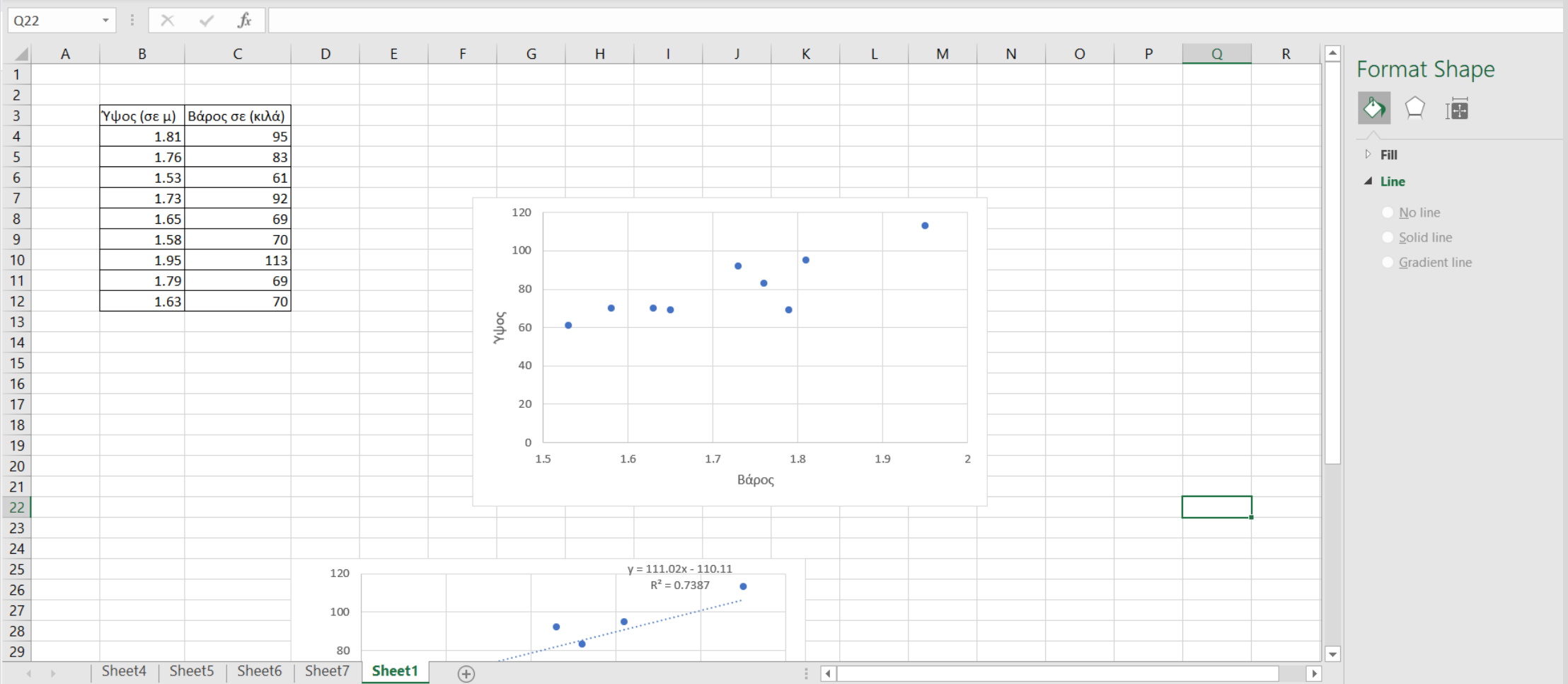
Sort & Filter: Sort, Filter, Clear, Reapply, Advanced

Data Tools: Text to Columns, Data Tools

Forecast: What-If Analysis, Forecast Sheet

Outline: Group, Ungroup, Subtotal

Analysis: **Data Analysis**



## Data Analysis



### Analysis Tools

- Histogram
- Moving Average
- Random Number Generation
- Rank and Percentile
- Regression**
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two Sample for Means

OK

Cancel

Help

Regression



Input

Input Y Range:

Input X Range:

Labels  Constant is Zero

Confidence Level:  %

OK

Cancel

Help

Output options

Output Range:

New Worksheet Ply:

New Workbook

Residuals

Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

Normal Probability

Normal Probability Plots

File Home Insert Page Layout Formulas **Data** Review View Help Chart Design Format

Get Data: From Text/CSV, From Web, From Table/Range, Recent Sources, Existing Connections

Queries & Connections: Refresh, Properties, Edit Links

Data Types: Stocks, Geography

Sort & Filter: Sort, Filter, Clear, Reapply, Advanced

Data Tools: Text to Columns, What-If Analysis

Chart 5

x	y
Βάρος σε (κιλά)	Υψος σε (μ)
95	1.81
83	1.76
61	1.53
92	1.73
69	1.65
70	1.58
113	1.95
69	1.79
70	1.63

Regression

Input

Input Y Range:  **8**

Input X Range:  **8**

Labels  Constant is Zero

Confidence Level:  %

Output options

Output Range:

New Worksheet Ply:

New Workbook

Residuals

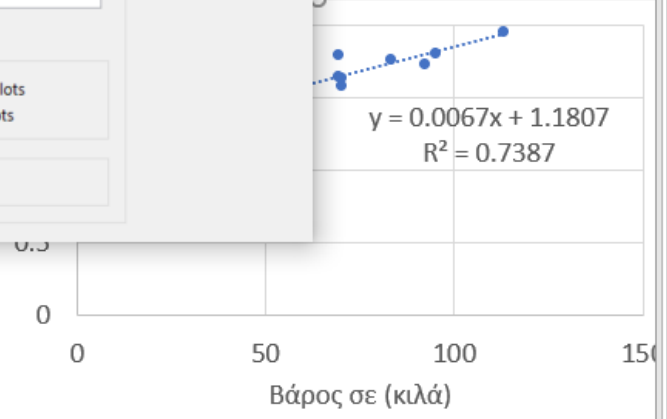
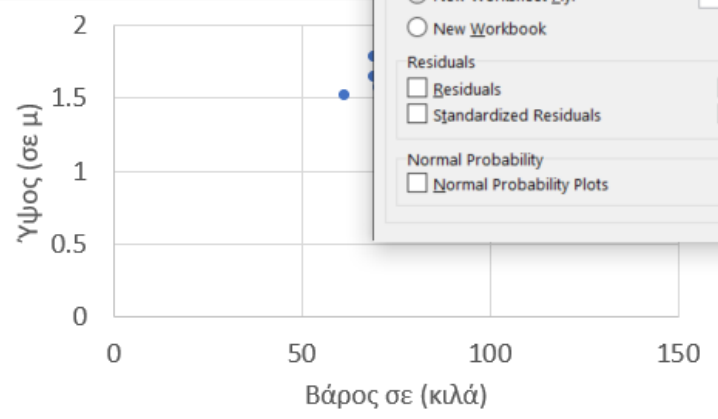
Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

Normal Probability

Normal Probability Plots

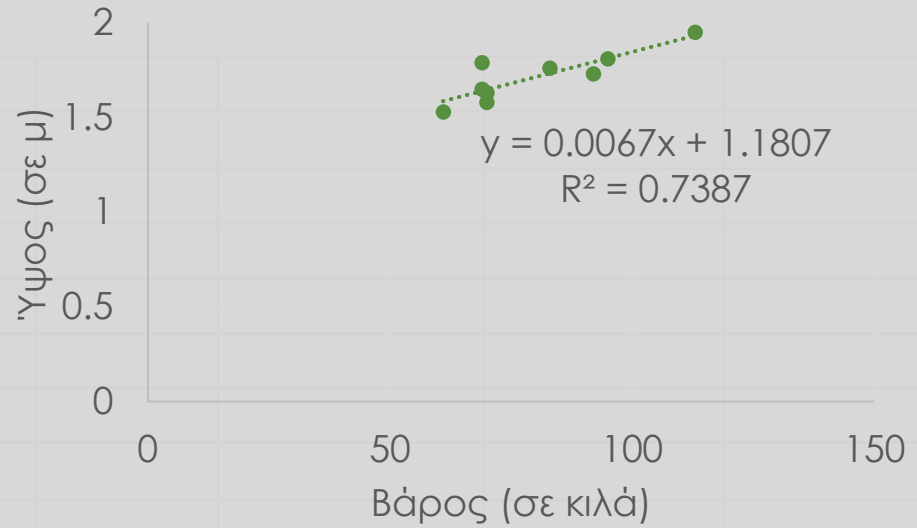
OK Cancel Help



SUMMARY OUTPUT

*Regression Statistics*

Multiple R	0.859475618
R Square	0.738698339
Adjusted R Square	0.70136953
Standard Error	0.071256912
Observations	9



ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	0.10047939	0.100479	19.78896	0.002977097
Residual	7	0.035542833	0.005078		
Total	8	0.136022222			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	1.18066386	0.12232001	9.652254	2.75E-05	0.891422999	1.46990472	0.891423	1.469904721
Βάρος σε (κιλά)	0.006653775	0.001495742	4.448478	0.002977	0.003116908	0.01019064	0.00311691	0.010190642

# Παράδειγμα (ανάλυση με excel/calc)

- Ερώτημα:
- Ισχύει η μηδενική υπόθεση  $H_0$  ( $b=0$ );
  - ΌΧΙ,  $p < 0.05$
- Ισχύει η εναλλακτική υπόθεση  $H_1$  ( $b \neq 0$ );
  - ΝΑΙ,  $p < 0.05$ 
    - Που είναι σημείο τομής με τον άξονα των  $y$  ( $a$ , intercept);
      - $a=1.181$
    - Τιμή του συντελεστή παλινδρόμησης ( $b$ , κλίση), regression coefficient (slope);
      - $b=0.007$
    - Τιμή του συντελεστή προσδιορισμού ( $R^2$ ); Πόση διασπορά εξηγείται από το μοντέλο;
      - $R^2=0.739$
      - Το μοντέλο εξηγεί 73.9% της διασποράς



# Άσκηση (ανάλυση με excel/calc)

- Ένα μαγαζί θέλει να εκτιμήσει την πιθανή σχέση μεταξύ βροχόπτωσης και πώλησης ομπρελών.
- Ποια βήματα πρέπει να ακολουθήσουν;
- Παρατηρήσεις:

Μηνιαία βροχόπτωση (σε mm)	πώληση ομπρελών
82	15
92.5	25
83.2	17
131.9	41
23.2	5
56	13
78.6	16
5.7	9
43.8	12

a	
b	
R <sup>2</sup>	
ρ	

# Άσκηση (ανάλυση με excel/calc)

- Ένα μαγαζί θέλει να εκτιμήσει την πιθανή σχέση μεταξύ βροχόπτωσης και πώλησης ομπρελών.
- Ποια βήματα πρέπει να ακολουθήσουν;
- Παρατηρήσεις:

Μηνιαία βροχόπτωση (σε mm)	πώληση ομπρελών
82	15
92.5	25
83.2	17
131.9	41
23.2	5
56	13
78.6	16
5.7	9
43.8	12

a	0.744
b	0.245
R <sup>2</sup>	0.794
p	0.001

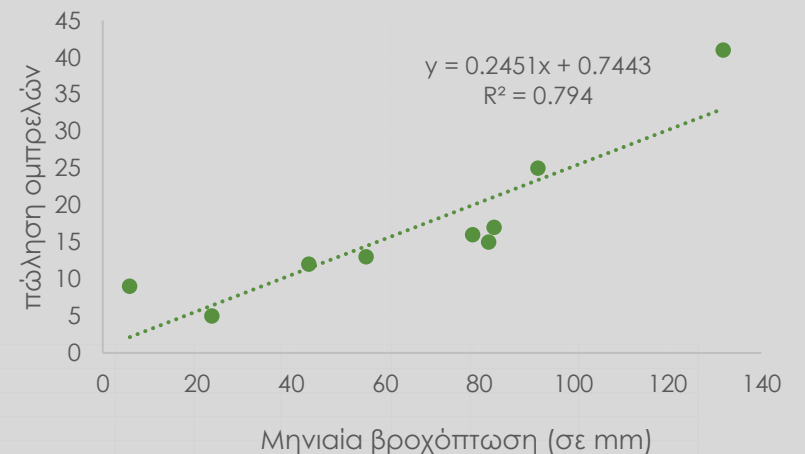
## SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8910536
R Square	0.7939765
Adjusted R Square	0.7645446
Standard Error	5.1295363
Observations	9

## ANOVA

	df	SS	MS	F	Significance F
Regression	1	709.8150019	709.815	26.976709	0.001261426
Residual	7	184.1849981	26.31214		
Total	8	894			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.7442996	3.566367442	0.2087	0.8406257	-7.68881932	9.177419	-7.68881932	9.177418566
Μηνιαία βροχόπτωση (σε mm)	0.2451019	0.047190235	5.193911	0.0012614	0.133514692	0.356689	0.133514692	0.356689038



# Άσκηση (ανάλυση με excel/calc)

- Ερώτημα:
- Ισχύει η μηδενική υπόθεση  $H_0$  ( $b=0$ );
  - ΌΧΙ,  $p < 0.05$
- Ισχύει η εναλλακτική υπόθεση  $H_1$  ( $b \neq 0$ );
  - ΝΑΙ,  $p < 0.05$ , άρα η πώληση ομπρελών εξαρτάται στατιστικά σημαντικά από την βροχόπτωση
    - Που είναι σημείο τομής με τον άξονα των  $y$  ( $a$ , intercept);
      - $a=0.744$
    - Τιμή του συντελεστή παλινδρόμησης ( $b$ , κλίση), regression coefficient (slope);
      - $b=0.245$
    - Τιμή του συντελεστή προσδιορισμού ( $R^2$ ); Πόση διασπορά εξηγείται από το μοντέλο;
      - $R^2=0.794$
      - Το μοντέλο εξηγεί 79.4% της διασποράς