

Απλή γραμμική παλινδρόμηση

Simple linear regression

Γούργουλης Βασίλειος, Ph.D.
Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

1

- νέφος σημείων
- πρόβλεψη
- πραγματικές – προβλεπόμενες τιμές
- ευθεία (γραμμή) παλινδρόμησης
- συντελεστές παλινδρόμησης
- συντελεστής προσδιορισμού R^2

2

Ανιχνεύει την σχέση
μεταξύ **δύο** ποσοτικών μεταβλητών

- Τι **σχέση** υπάρχει μεταξύ των δύο μεταβλητών
- **Εκτίμηση** (πρόβλεψη)

Από τις τιμές μιας **ανεξάρτητης** (independent) μεταβλητής (**X**)
να προβλεφθούν
οι τιμές μιας **εξαρτημένης** (dependent) μεταβλητής (**Y**)

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

3

Μοντέλο απλής παλινδρόμησης

άτομο	Y	X
A_1	Y_1	X_1
A_2	Y_2	X_2
A_3	Y_3	X_3
...	...	...
A_N	Y_N	X_N

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

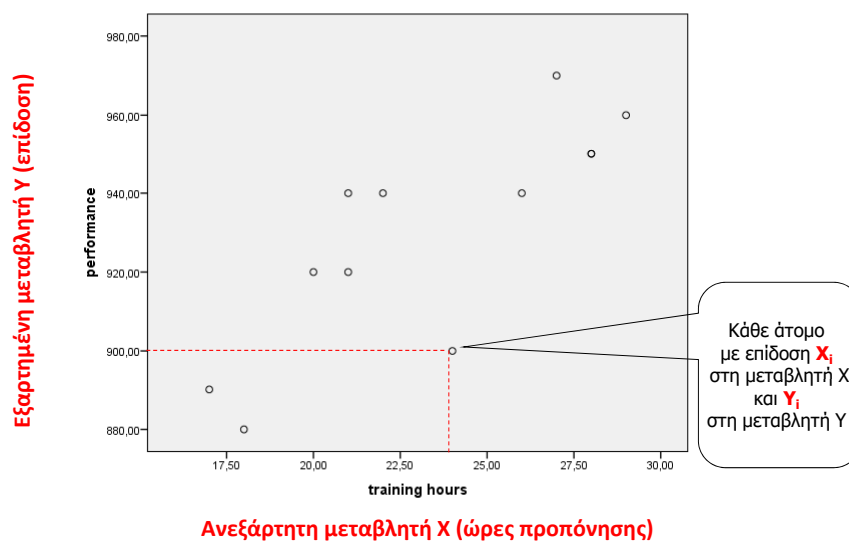
4

α/α	επίδοση στη σκοποβολή (Y_i)	εβδομαδιαίες ώρες προπόνησης (X_i)
1	920	20
2	940	21
3	880	18
4	890	17
5	900	24
6	940	26
7	950	28
8	960	29
9	940	22
10	920	21
11	950	28
12	970	27

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

5

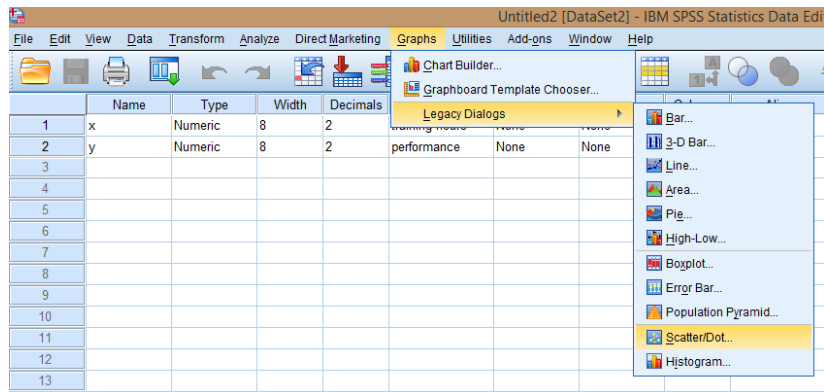
Γραφική αναπαράσταση δεδομένων (νέφος σημείων)



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

6

Γραφική αναπαράσταση δεδομένων

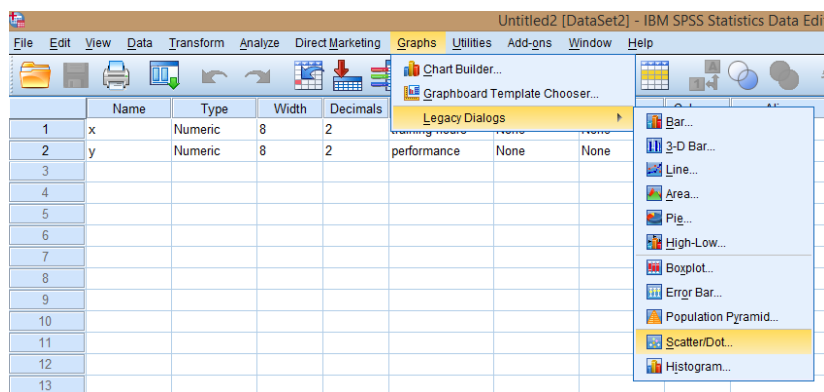


Graphs
Legacy Dialogs
Scatter/Dot..

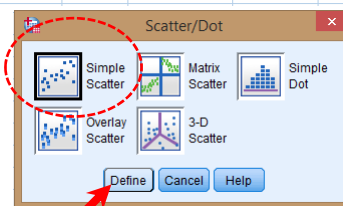
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

7

Γραφική αναπαράσταση δεδομένων



Graphs
Legacy Dialogs
Scatter/Dot..

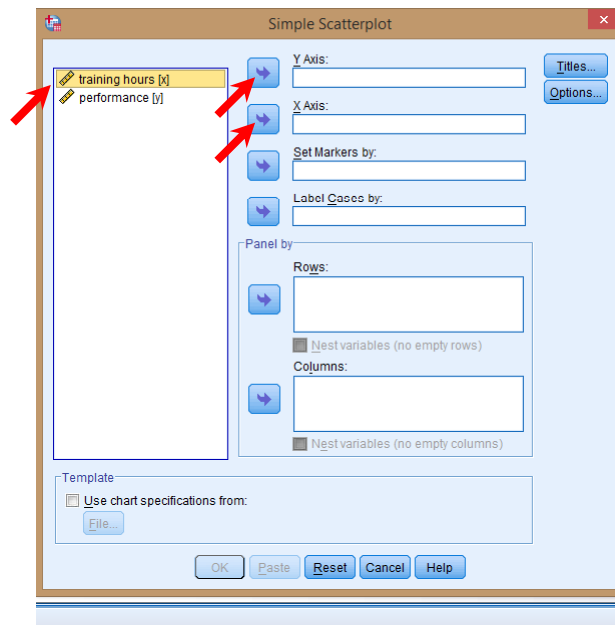


Simple Scatter

Define

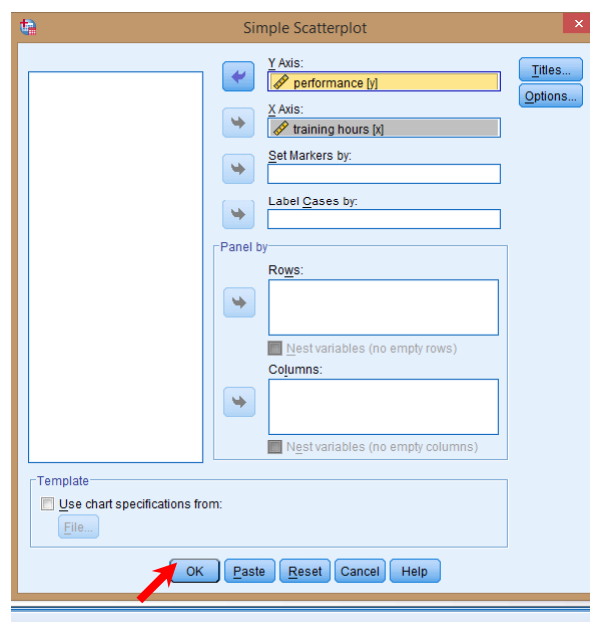
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

8



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

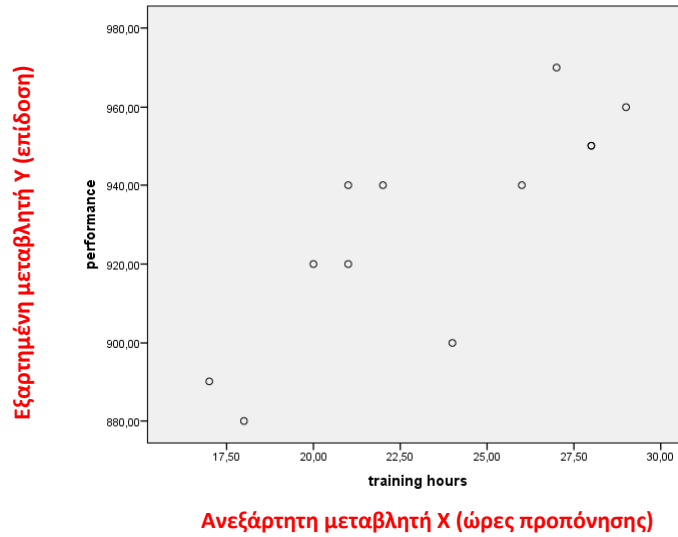
9



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

10

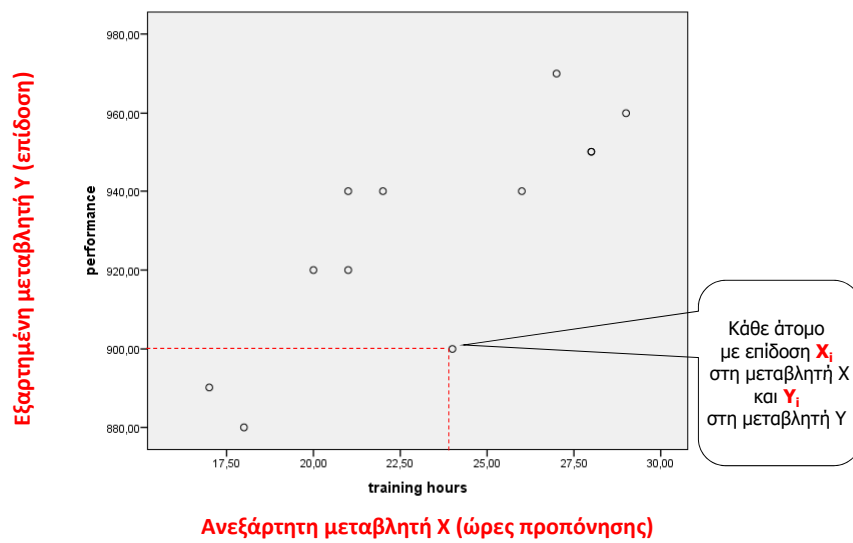
Αναπαράσταση του κάθε ατόμου ως σημείο
με συντεταγμένες (X_i, Y_i) τις τιμές του κάθε ατόμου στις μεταβλητές X και Y



11

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

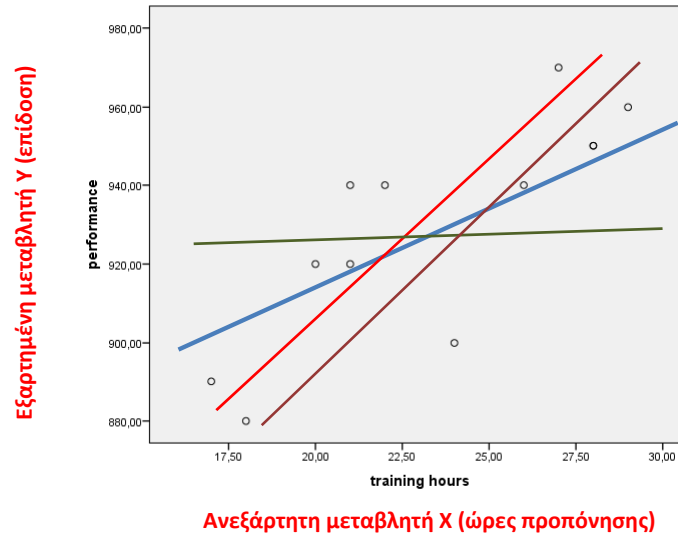
Αναπαράσταση του κάθε ατόμου ως σημείο
με συντεταγμένες (X_i, Y_i) τις τιμές του κάθε ατόμου στις μεταβλητές X και Y



12

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

Από αυτό το «νέφος» σημείων μπορούν να διέλθουν **άπειρες** γραμμές παλινδρόμησης



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

13

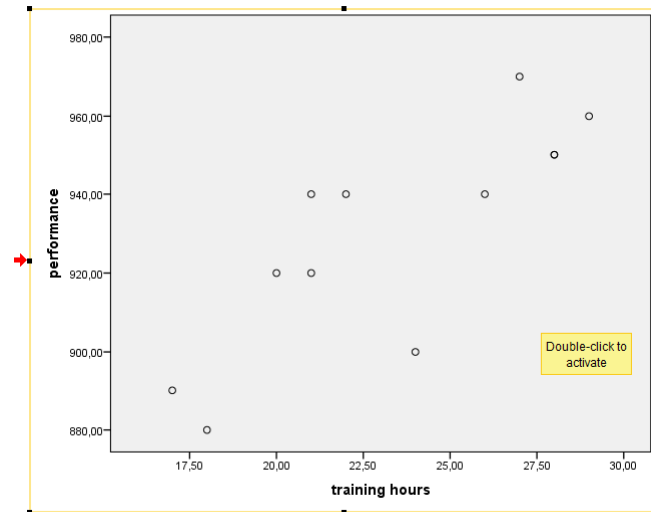
Ποια ευθεία (γραμμή παλινδρόμησης)
εκφράζει καλύτερα (με μεγαλύτερη ακρίβεια) τα δεδομένα;

... **μία και μόνο μία** γραμμή παλινδρόμησης
μπορεί να εκφράσει με τον καλύτερο τρόπο
αυτό το «νέφος» σημείων.

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

14

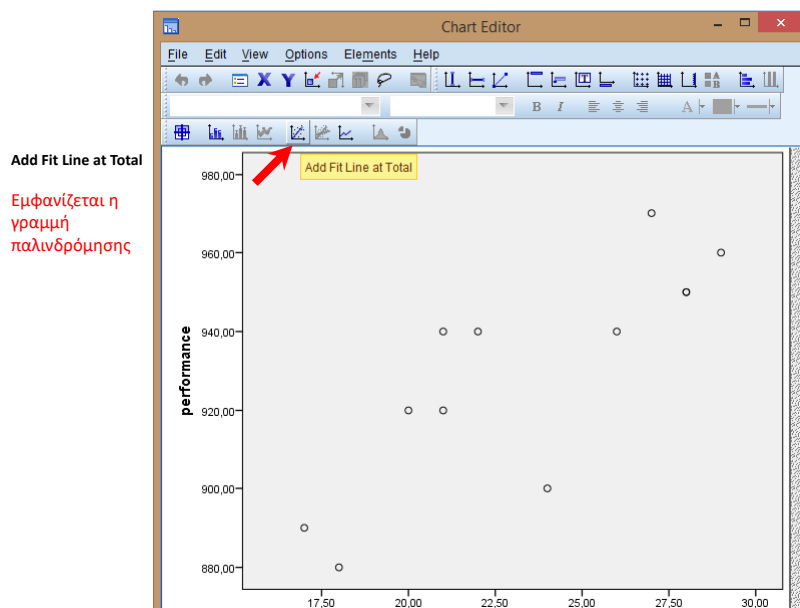
... κάνοντας **διπλό** (αριστερό) **κλικ** πάνω στο γράφημα ...



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

15

... κάνοντας **διπλό** (αριστερό) **κλικ** πάνω στο γράφημα ... ενεργοποιείται το πλαίσιο Chart Editor ...



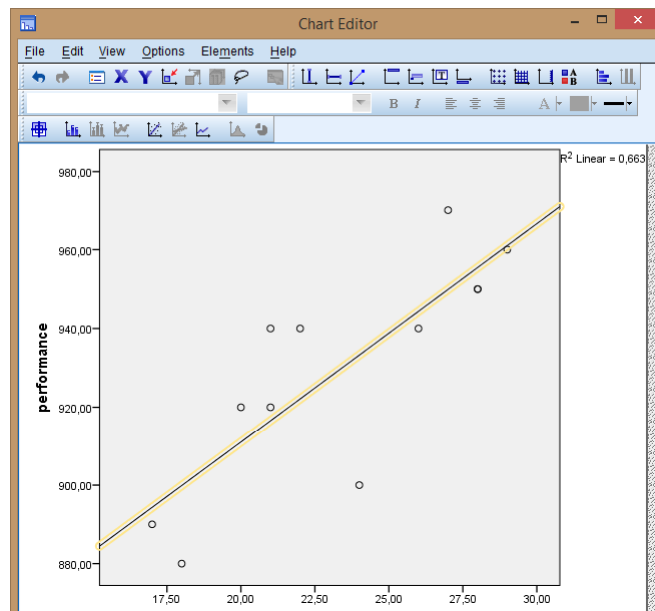
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

16

... κάνοντας **διπλό** (αριστερό) **κλικ** πάνω στο γράφημα ... ενεργοποιείται το πλαίσιο Chart Editor ...

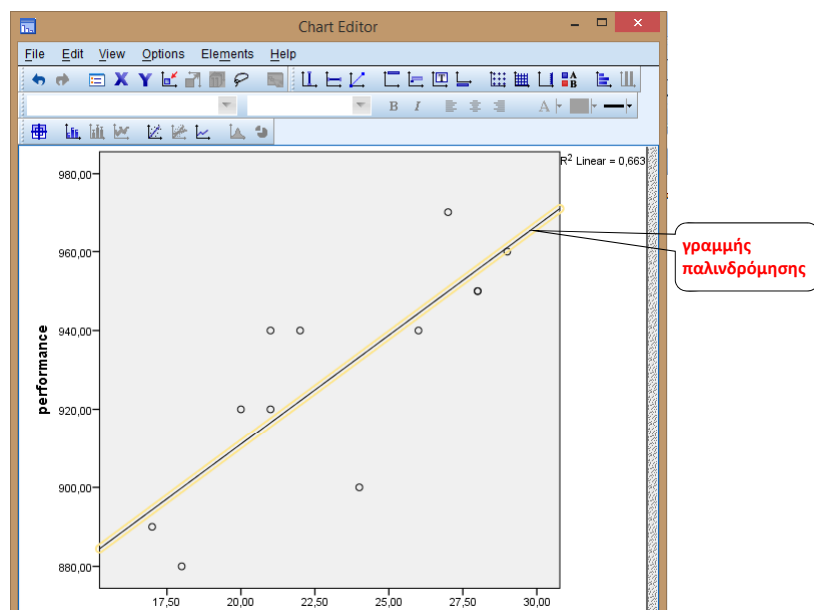
Add Fit Line at Total

Εμφανίζεται η
γραμμή
παλινδρόμησης



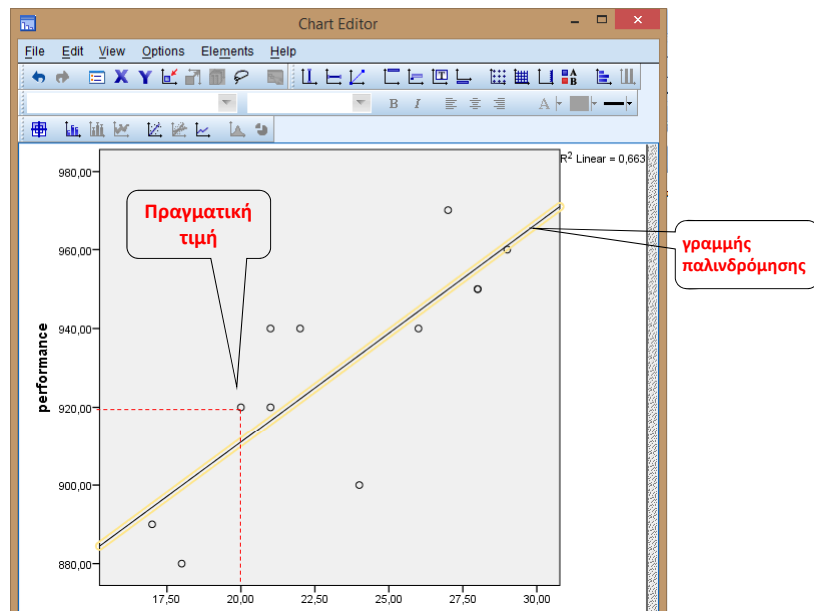
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

17



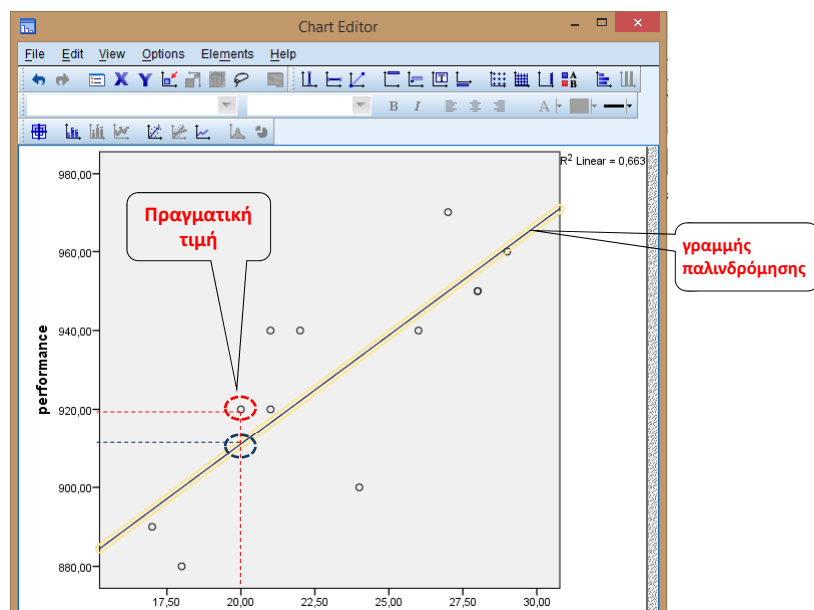
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

18



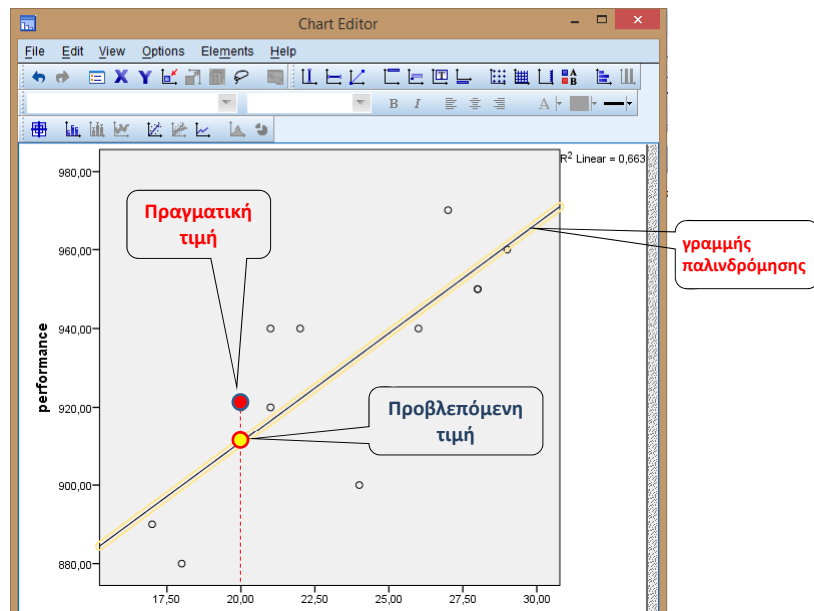
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

19



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

20



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

21

**Πώς καθορίζεται η διεύθυνση
αυτής της γραμμής παλινδρόμησης;**

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

22

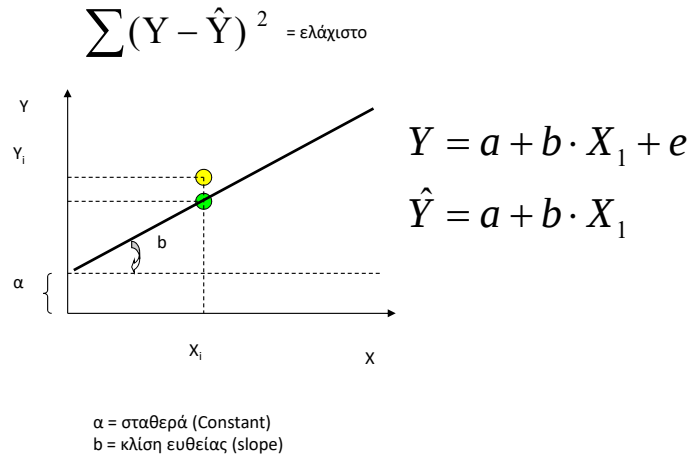
Η γραμμή παλινδρόμησης

που μπορεί να εκφράσει με τον καλύτερο τρόπο (με την μεγαλύτερη ακρίβεια)

αυτό το «νέφος» σημείων,

είναι η γραμμή βάσει της οποίας

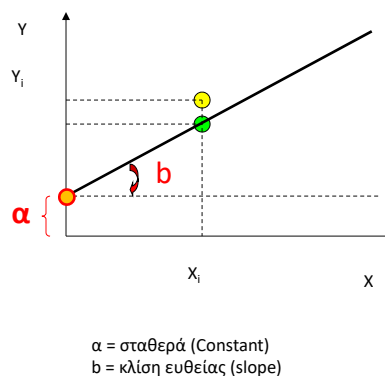
το **άθροισμα των τετραγώνων της διαφοράς ΟΛΩΝ των πραγματικών από τις προβλεπόμενες τιμές** είναι το **ελάχιστο** δυνατό.



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

23

Πώς καθορίζονται
οι **συντελεστές παλινδρόμησης (α, b)**
που ορίζουν τη **διεύθυνση** της γραμμής παλινδρόμησης;



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

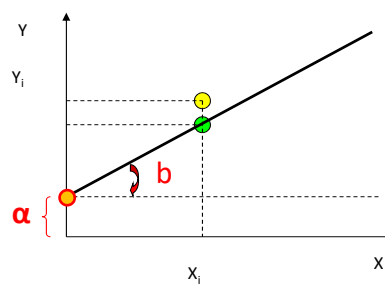
24

Καθορισμός **συντελεστών παλινδρόμησης** (unstandardized coefficients)

$$b_{yx} = \frac{N \sum_{i=1}^N (X_i \cdot Y_i) - \sum_{i=1}^N X_i \cdot \sum_{i=1}^N Y_i}{N \sum_{i=1}^N X_i^2 - (\sum_{i=1}^N X_i)^2}$$

$$a = \frac{\sum_{i=1}^N Y_i - b \cdot \sum_{i=1}^N X_i}{N} = \bar{Y} - b \cdot \bar{X}$$

α = σταθερά (Constant)
 b = κλίση ευθείας (slope)



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

25

α/α	επίδοση στη σκοποβολή (Y_i)	εβδομαδιαίες ώρες προπόνησης (X_i)
1	920	20
2	940	21
3	880	18
4	890	17
5	900	24
6	940	26
7	950	28
8	960	29
9	940	22
10	920	21
11	950	28
12	970	27

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

26

a/a	επίδοση στη σκοποβολή (Y _i)	εβδομαδιαίες ώρες προπόνησης (X _i)	(X _i) ²	X _i Y _i
1	920	20		
2	940	21		
3	880	18		
4	890	17		
5	900	24		
6	940	26		
7	950	28		
8	960	29		
9	940	22		
10	920	21		
11	950	28		
12	970	27		
	$\sum_{i=1}^N Y_i =$	$\sum_{i=1}^N X_i =$	$\sum_{i=1}^N X_i^2 =$	$\sum_{i=1}^N X_i Y_i =$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

27

a/a	επίδοση στη σκοποβολή (Y _i)	εβδομαδιαίες ώρες προπόνησης (X _i)	(X _i) ²	X _i Y _i
1	920	20	400	18400
2	940	21	441	19740
3	880	18	324	15840
4	890	17	289	15130
5	900	24	576	21600
6	940	26	676	24440
7	950	28	784	26600
8	960	29	841	27840
9	940	22	484	20680
10	920	21	441	19320
11	950	28	784	26600
12	970	27	729	26190
	$\sum_{i=1}^N Y_i = 11160$	$\sum_{i=1}^N X_i = 281$	$\sum_{i=1}^N X_i^2 = 6769$	$\sum_{i=1}^N X_i Y_i = 262380$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

28

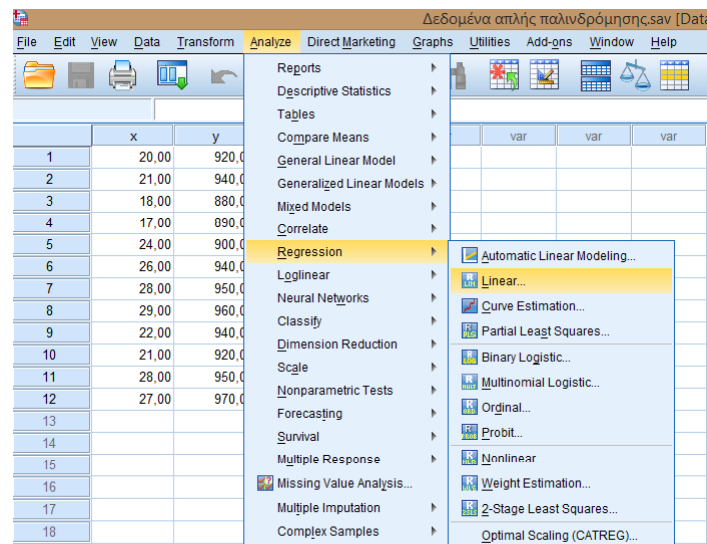
$$b_{yx} = \frac{N \sum_{i=1}^N (X_i \cdot Y_i) - \sum_{i=1}^N X_i \cdot \sum_{i=1}^N Y_i}{N \sum_{i=1}^N X_i^2 - (\sum_{i=1}^N X_i)^2}$$

$$b_{yx} = \frac{12 \cdot 262380 - 281 \cdot 11160}{12 \cdot 6769 - (281)^2} = \frac{3148560 - 3135960}{81228 - 78961} = \frac{12600}{2267} = 5.56$$

$$\alpha_{yx} = \frac{\sum_{i=1}^N Y_i - b_{yx} \cdot \sum_{i=1}^N X_i}{N} = \frac{11160 - 5,56 \cdot 281}{12} = 799,8$$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

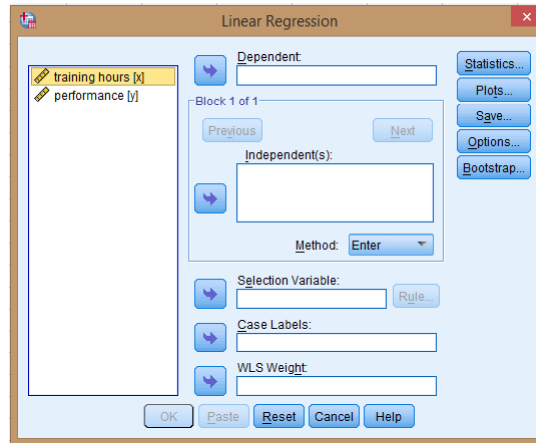
29



Analyze
Regression
Linear ...

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

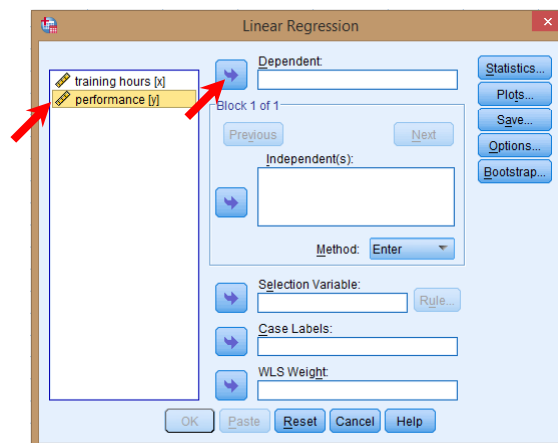
30



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

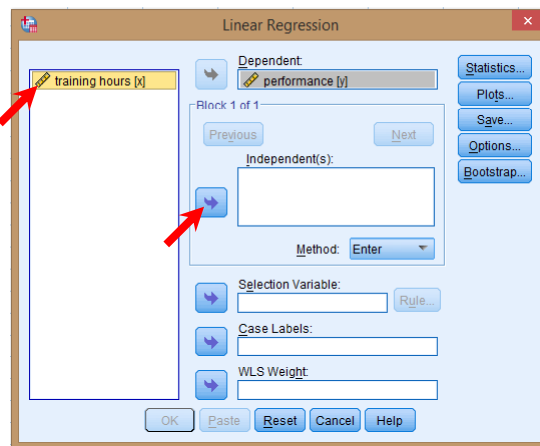
31

... καθορισμός εξαρτημένης (**Dependent**) μεταβλητής ...



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

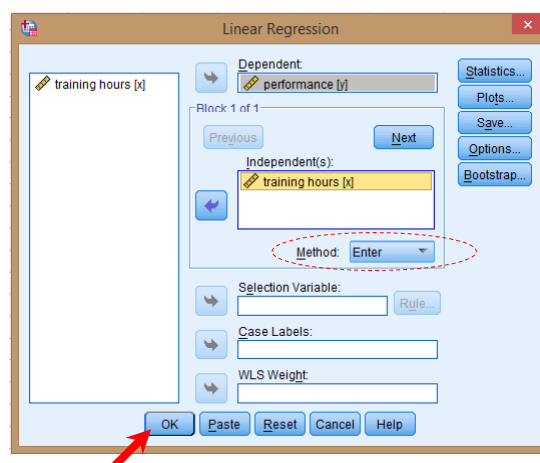
32



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

33

... καθορισμός ανεξάρτητης (**Independent**) μεταβλητής ...



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

34

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	training hours	.	Enter

a. All requested variables entered.

b. Dependent Variable: performance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	17,21654

a. Predictors: (Constant), training hours

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5835,906	1	5835,906	19,689	,001 ^a
	Residual	2964,094	10	296,409		
	Total	8800,000	11			

a. Predictors: (Constant), training hours

b. Dependent Variable: performance

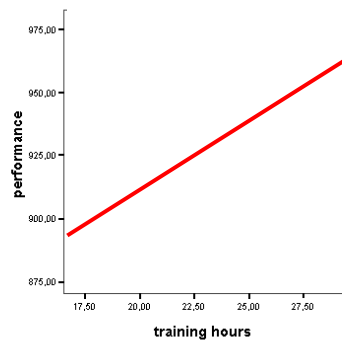
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	799,850	29,750		26,886	,000
	training hours	5,558	1,253	,814	4,437	,001

a. Dependent Variable: performance

35

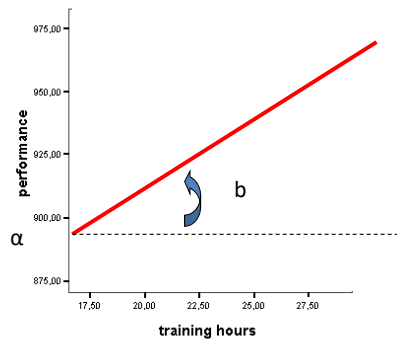
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



Πώς καθορίζεται η διεύθυνση της γραμμής παλινδρόμησης;

36

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



Από τους συντελεστές παλινδρόμησης (α , b)

$$Y = a + b \cdot X_1 + e$$

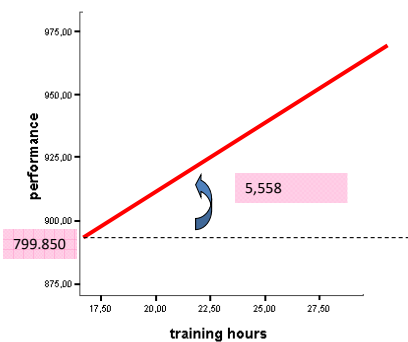
$$\hat{Y} = a + b \cdot X_1$$

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

37

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



$$Y = a + b \cdot X_1 + e$$

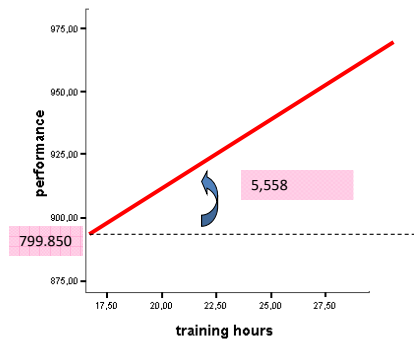
$$\hat{Y} = a + b \cdot X_1$$

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

38

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



Συντελεστές παλινδρόμησης (α , b)

$$Y = a + b \cdot X_1 + e$$

$$\hat{Y} = a + b \cdot X_1$$

$$b = \frac{N \cdot \sum_{i=1}^N X_i Y_i - \sum_{i=1}^N X_i \cdot \sum_{i=1}^N Y_i}{N \cdot \sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i \right)^2} = \frac{\sum_{i=1}^N X_i Y_i - N \cdot \bar{X} \cdot \bar{Y}}{\sum_{i=1}^N X_i^2 - N \cdot \bar{X}^2}$$

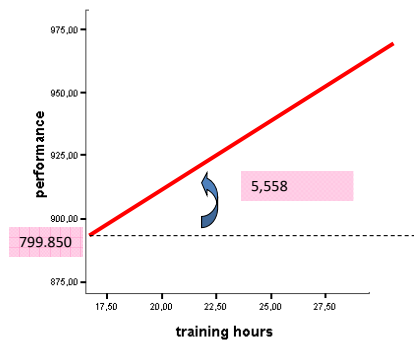
$$a = \frac{\sum_{i=1}^N Y_i - b \cdot \sum_{i=1}^N X_i}{N} = \bar{Y} - b \cdot \bar{X}$$

		Coefficients ^a		t	Sig.
		Unstandardized Coefficients	Standardized Coefficients		
Model		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

39



Συντελεστές παλινδρόμησης (α , b)

$$Y = a + b \cdot X_1 + e$$

$$\hat{Y} = a + b \cdot X_1$$

$$b = \frac{N \cdot \sum_{i=1}^N X_i Y_i - \sum_{i=1}^N X_i \cdot \sum_{i=1}^N Y_i}{N \cdot \sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i \right)^2}$$

$$b_{yx} = \frac{12 \cdot 262380 - 281 \cdot 11160}{12 \cdot 6769 - (281)^2} = \frac{3148560 - 3135960}{81228 - 78961} = \frac{12600}{2267} = 5.56$$

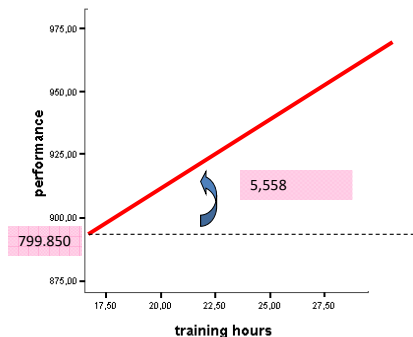
$$\alpha_{yx} = \frac{\sum_{i=1}^N Y_i - b_{yx} \cdot \sum_{i=1}^N X_i}{N} = \frac{11160 - 5,56 \cdot 281}{12} = 799,8$$

		Coefficients ^a		t	Sig.
		Unstandardized Coefficients	Standardized Coefficients		
Model		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

40



Standardized Coefficients Beta

Τυπικός συντελεστής προσδιορισμού

Εκφράζει
την **κλίση** της γραμμής παλινδρόμησης
όταν και οι δύο μεταβλητές
(εξαρτημένη Y και ανεξάρτητη X)
εκφράζονται σε z - τιμές

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

41

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

Standardized Coefficients Beta

Τυπικός συντελεστής προσδιορισμού

$$Beta = B_1 \cdot \frac{S_X}{S_Y} =$$

συντελεστής παλινδρόμησης ανεξάρτητης μεταβλητής $\cdot$ $\frac{\text{τυπική απόκλιση ανεξάρτητης μεταβλητής } X}{\text{τυπική απόκλιση εξαρτημένης μεταβλητής } Y}$

Descriptive Statistics			
	Mean	Std. Deviation	N
performance	930,0000	28,28427	12
training hours	23,4167	4,14418	12

$$Beta = B_1 \cdot \frac{S_X}{S_Y} = 5,558 \cdot \frac{4,14418}{28,28427} = 0,814$$

Εκφράζει
την **κλίση** της γραμμής παλινδρόμησης
όταν και οι δύο μεταβλητές
(εξαρτημένη Y και ανεξάρτητη X)
εκφράζονται σε z - τιμές

Ας το δούμε αυτό στην πράξη μετασχηματίζοντας τις αρχικές τιμές της εξαρτημένης και της ανεξάρτητης μεταβλητής σε z-τυπικές τιμές και ας εκτελέσουμε την απλή παλινδρόμηση για αυτές τις μετασχηματισμένες τιμές

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

42

Επαλήθευση
(τυπικού συντελεστή προσδιορισμού Beta)

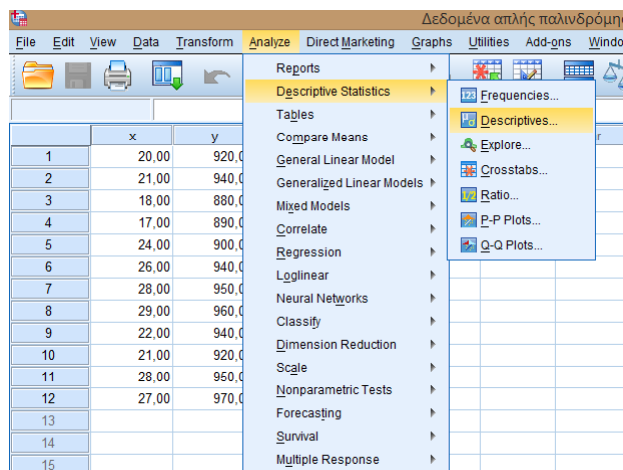
... **μετασχηματίζοντας**
τις αρχικές τιμές της εξαρτημένης και της ανεξάρτητης μεταβλητής
σε **z-τυπικές τιμές**
και εκτελώντας την απλή παλινδρόμηση
για αυτές τις μετασχηματισμένες τιμές

Μετατροπή αρχικών τιμών σε z – τυπικές τιμές
$$z = \frac{X_i - \bar{X}}{S}$$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

43

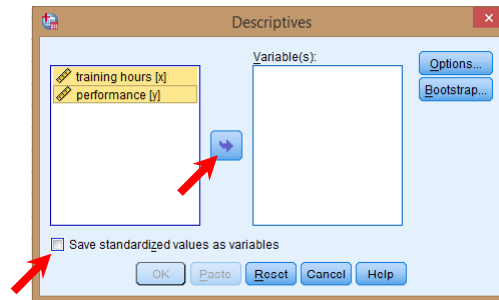
Μετατροπή αρχικών τιμών σε z – τυπικές τιμές
$$z = \frac{X_i - \bar{X}}{S}$$



Analyze
Descriptive Statistics
Descriptives...

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

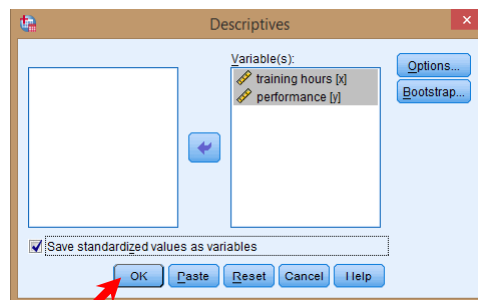
44



Saved standardized values as variables

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

45



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

46

z-τυπικές τιμές

*Δεδομένα απλής παλινδ

	x	y	Zx	Zy	var
1	20,00	920,00	-,82445	-,35355	
2	21,00	940,00	-,58315	,35355	
3	18,00	880,00	-1,30705	-1,76777	
4	17,00	890,00	-1,54836	-1,41421	
5	24,00	900,00	,14076	-1,06066	
6	26,00	940,00	,62336	,35355	
7	28,00	950,00	1,10597	,70711	
8	29,00	960,00	1,34727	1,06066	
9	22,00	940,00	-,34184	,35355	
10	21,00	920,00	-,58315	-,35355	
11	28,00	950,00	1,10597	,70711	
12	27,00	970,00	,86467	1,41421	
13					

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

47

... εκτελώντας την απλή παλινδρόμηση για τις **z-τυπικές τιμές** ...

*Δεδομένα απλής παλινδρόμησης.sav [DataSet1]

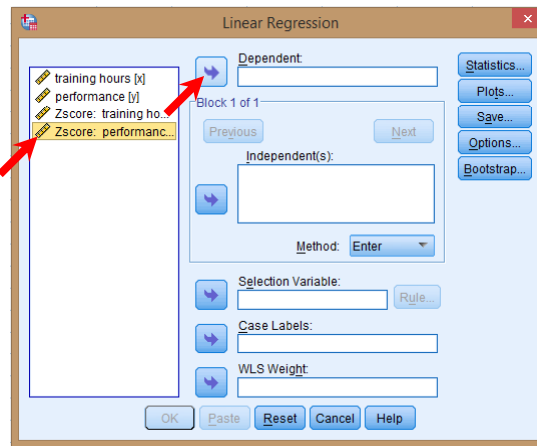
	x	y	var	var	var
1	20,00	920,00	-,35355		
2	21,00	940,00	,35355		
3	18,00	880,00	-1,76777		
4	17,00	890,00	-1,41421		
5	24,00	900,00			
6	26,00	940,00			
7	28,00	950,00			
8	29,00	960,00			
9	22,00	940,00			
10	21,00	920,00			
11	28,00	950,00			
12	27,00	970,00			
13					
14					
15					
16					

File Edit View Data Transform **Analyze** Direct Marketing Graphs Utilities Add-ons Window Help

- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression**
 - Automatic Linear Modeling...
 - Linear...**
 - Curve Estimation...
 - Partial Least Squares...
 - Binary Logistic...
 - Multinomial Logistic...
 - Ordinal...
 - Probit...
 - Nonlinear...
 - Weight Estimation...
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...

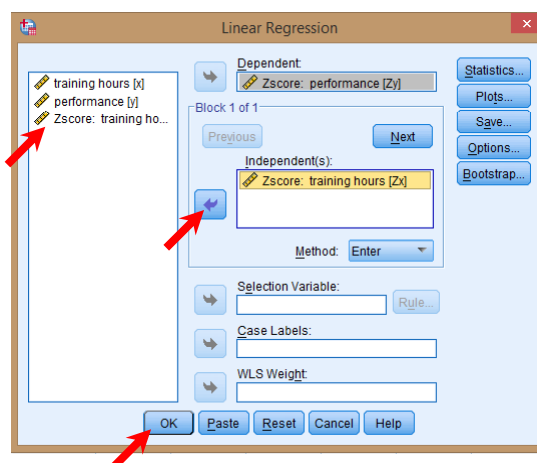
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

48



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

49



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

50

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Zscore: performance	,0000000	1,00000000	12
Zscore: training hours	,0000000	1,00000000	12

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Zscore: training hours	.	Enter

a. All requested variables entered.

b. Dependent Variable: Zscore: performance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	,60869671

a. Predictors: (Constant), Zscore: training hours

Correlations

		Zscore: performance	Zscore: training hours
Pearson Correlation	Zscore: performance	1,000	,814
	Zscore: training hours	,814	1,000
Sig. (1-tailed)	Zscore: performance	.	,001
	Zscore: training hours	,001	.
N	Zscore: performance	12	12
	Zscore: training hours	12	12

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	7,295	1	7,295	19,689	,001 ^a
Residual	3,705	10	,371		
Total	11,000	11			

a. Predictors: (Constant), Zscore: training hours

b. Dependent Variable: Zscore: performance

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,000	,176		,000	1,000
	Zscore: training hours	,814	,184	,814	4,437	,001

a. Dependent Variable: Zscore: performance

51

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

Αρχικές τιμές - unstandardized

Descriptive Statistics

	Mean	Std. Deviation	N
performance	930,0000	28,28427	12
training hours	23,4167	4,14418	12

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	training hours	.	Enter

a. All requested variables entered.

b. Dependent Variable: performance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	17,21654

a. Predictors: (Constant), training hours

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	5835,906	1	5835,906	19,689	,001 ^a
Residual	2964,094	10	296,409		
Total	8800,000	11			

a. Predictors: (Constant), training hours

b. Dependent Variable: performance

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	799,850	29,750		26,886	,000
	training hours	5,558	1,253	,814	4,437	,001

a. Dependent Variable: performance

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

Z - τυπικές τιμές - standardized

$$z = \frac{X_i - \bar{X}}{S}$$

Descriptive Statistics

	Mean	Std. Deviation	N
Zscore: performance	,0000000	1,00000000	12
Zscore: training hours	,0000000	1,00000000	12

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Zscore: training hours	.	Enter

a. All requested variables entered.

b. Dependent Variable: Zscore: performance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	,60869671

a. Predictors: (Constant), Zscore: training hours

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	7,295	1	7,295	19,689	,001 ^a
Residual	3,705	10	,371		
Total	11,000	11			

a. Predictors: (Constant), Zscore: training hours

b. Dependent Variable: Zscore: performance

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,000	,176		,000	1,000
	Zscore: training hours	,814	,184	,814	4,437	,001

a. Dependent Variable: Zscore: performance

52

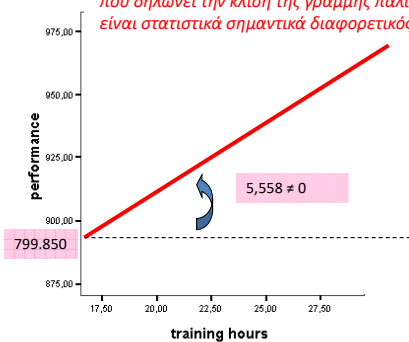
Υπάρχει πραγματικά σχέση
μεταξύ της εξαρτημένης (Y) και της ανεξάρτητης (X)
μεταβλητής;

53

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

Υπάρχει πραγματικά σχέση μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής;

Ο συντελεστής B (Unstandardized Coefficient),
που δηλώνει την κλίση της γραμμής παλινδρόμησης,
είναι στατιστικά σημαντικά διαφορετικός από το 0;



Για να υπάρχει πραγματικά σχέση
μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής
θα πρέπει ο **συντελεστής B** (Unstandardized Coefficient)
που δηλώνει την κλίση της γραμμής παλινδρόμησης
να είναι στατιστικά σημαντικά **διαφορετικός από το 0**.

Coefficients ^a					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	799,850	29,750		26,886	,000
training hours	5,558	1,253	,814	4,437	,001

a. Dependent Variable: performance

54

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

Υπάρχει πραγματικά σχέση μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής;

Ο συντελεστής B (Unstandardized Coefficient), που δηλώνει την κλίση της γραμμής παλινδρόμησης, είναι στατιστικά σημαντικά διαφορετικός από το 0;



Για να υπάρχει πραγματικά σχέση μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής θα πρέπει ο **συντελεστής B** (Unstandardized Coefficient) που δηλώνει την κλίση της γραμμής παλινδρόμησης να είναι στατιστικά σημαντικά **διαφορετικός από το 0**.

Αν ο συντελεστής B (Unstandardized Coefficient) που δηλώνει την κλίση της γραμμής παλινδρόμησης **ΔΕΝ** είναι στατιστικά σημαντικά **διαφορετικός από το 0** τότε **ΔΕΝ** υπάρχει πραγματικά σχέση μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής.

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

55

Υπάρχει πραγματικά σχέση μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής;

Ο συντελεστής B (Unstandardized Coefficient), που δηλώνει την κλίση της γραμμής παλινδρόμησης, είναι στατιστικά σημαντικά διαφορετικός από το 0;



Για να υπάρχει πραγματικά σχέση μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής θα πρέπει ο **συντελεστής B** (Unstandardized Coefficient) που δηλώνει την κλίση της γραμμής παλινδρόμησης να είναι στατιστικά σημαντικά **διαφορετικός από το 0**.

Αν ο συντελεστής B (Unstandardized Coefficient) που δηλώνει την κλίση της γραμμής παλινδρόμησης **ΔΕΝ** είναι στατιστικά σημαντικά διαφορετικός από το 0 τότε **ΔΕΝ** υπάρχει πραγματικά σχέση μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής.

Αυτός η έλεγχος επιτυγχάνεται μέσω της **t – τιμής** από τον πίνακα Coefficients.

$$t = \frac{b_i}{S_{b_i}}$$

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

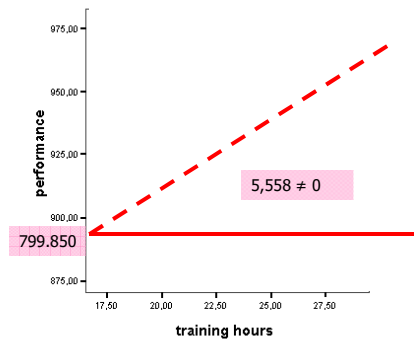
a. Dependent Variable: performance

b_i = ο συντελεστής παλινδρόμησης της συγκεκριμένης μεταβλητής x_i

S_{b_i} = το τυπικό σφάλμα του συγκεκριμένου συντελεστή παλινδρόμησης

56

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



Έλεγχος του
«αν ο συντελεστής B (Unstandardized Coefficient)
που δηλώνει την κλίση της γραμμής παλινδρόμησης
είναι στατιστικά σημαντικά διαφορετικός από το 0»
σότε
υπάρχει πραγματικά σχέση
μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής.

Αυτός η έλεγχος επιτυγχάνεται μέσω της **t – τιμής**
από τον πίνακα Coefficients.

$$t = \frac{b_i}{S_{b_i}} \quad b_i = \text{ο συντελεστής παλινδρόμησης της συγκεκριμένης μεταβλητής } X_i$$

S_{b_i} = το τυπικό σφάλμα του συγκεκριμένου συντελεστή παλινδρόμησης

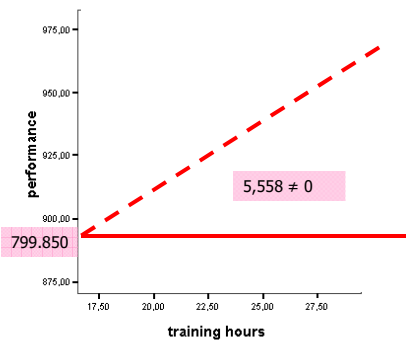
$$t_{\text{traininghours}} = \frac{b_{\text{traininghours}}}{S_{\text{traininghours}}} = \frac{5,558}{1,253} = 4,437$$

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

57



Έλεγχος του
«αν ο συντελεστής B (Unstandardized Coefficient)
που δηλώνει την κλίση της γραμμής παλινδρόμησης
είναι στατιστικά σημαντικά διαφορετικός από το 0»
σότε
υπάρχει πραγματικά σχέση
μεταξύ της εξαρτημένης Y και της ανεξάρτητης X μεταβλητής.

Αυτός η έλεγχος επιτυγχάνεται μέσω της **t – τιμής**
από τον πίνακα Coefficients.

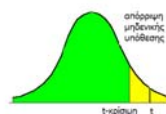
$$t = \frac{b_i}{S_{b_i}} \quad b_i = \text{ο συντελεστής παλινδρόμησης της συγκεκριμένης μεταβλητής } X_i$$

S_{b_i} = το τυπικό σφάλμα του συγκεκριμένου συντελεστή παλινδρόμησης

$$t_{\text{traininghours}} = \frac{b_{\text{traininghours}}}{S_{\text{traininghours}}} = \frac{5,558}{1,253} = 4,437$$

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	799,850	29,750		,000
	training hours	5,558	1,253	,814	,001

a. Dependent Variable: performance



Εφόσον
το επίπεδο σημαντικότητας που
αντιστοιχεί
στην **t – τιμή**
είναι μικρότερο από 0.05
ο συντελεστής B
είναι στατιστικά σημαντικά
διαφορετικός από το 0,
άρα οι δύο μεταβλητές έχουν
σχέση μεταξύ τους.

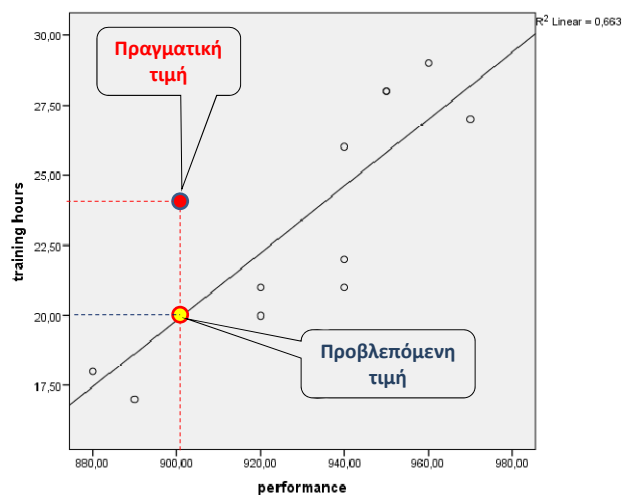
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

**Κατά πόσο το συγκεκριμένο γραμμικό μοντέλο
είναι κατάλληλο για να εκφράσει τα αρχικά δεδομένα;**

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

59

**Κατά πόσο το συγκεκριμένο γραμμικό μοντέλο
είναι κατάλληλο για να εκφράσει τα αρχικά δεδομένα;**



Ποια είναι η σχέση μεταξύ πραγματικών και προβλεπόμενων τιμών;

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

60

Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.814 ^a	.663	.629	17,21654

a. Predictors: (Constant), training hours

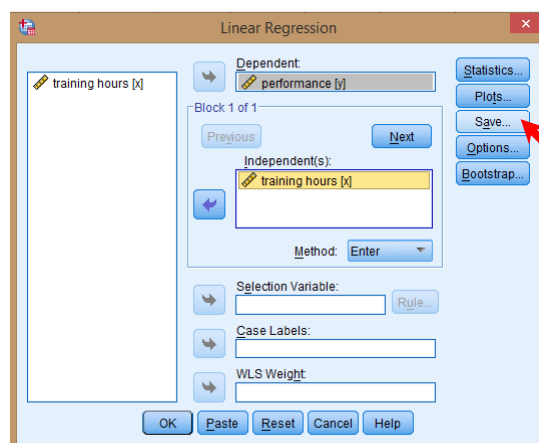
R - είναι η τετραγωνική ρίζα του R-Squared
και είναι ο **συντελεστής συσχέτισης**
μεταξύ των **παρατηρούμενων** και **προβλεπόμενων** τιμών της εξαρτημένης μεταβλητής.

Ποιες είναι οι **προβλεπόμενες** τιμές;

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

61

Ποιες είναι οι **προβλεπόμενες** τιμές;



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

62

Ποιες είναι οι προβλεπόμενες τιμές;

Predicted Values

Unstandardized

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

63

Ποιες είναι οι προβλεπόμενες τιμές;

	x	y	PRE_1	var
1	20,00	920,00	911,01015	
2	21,00	940,00	916,56815	
3	18,00	880,00	899,89413	
4	17,00	890,00	894,33613	
5	24,00	900,00	933,24217	
6	26,00	940,00	944,35818	
7	28,00	950,00	955,47419	
8	29,00	960,00	961,03220	
9	22,00	940,00	922,12616	
10	21,00	920,00	916,56815	
11	28,00	950,00	955,47419	
12	27,00	970,00	949,91619	
13				

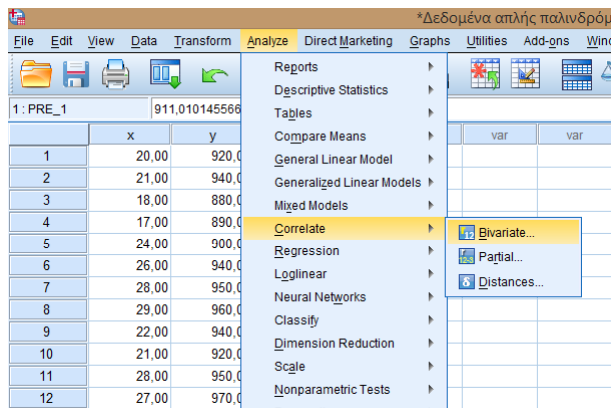
Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

64

Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;

Για να καθοριστεί η **σχέση** μεταξύ πραγματικών και προβλεπόμενων τιμών υπολογίζεται ο **συντελεστής συσχέτισης** μεταξύ τους

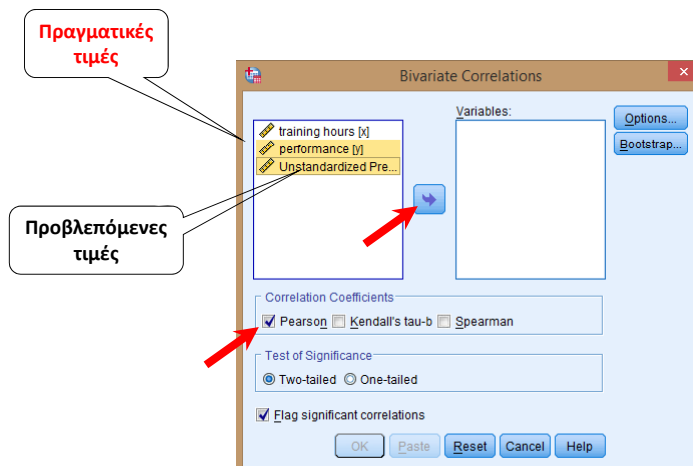


Analyze
Correlate
Bivariate...

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

65

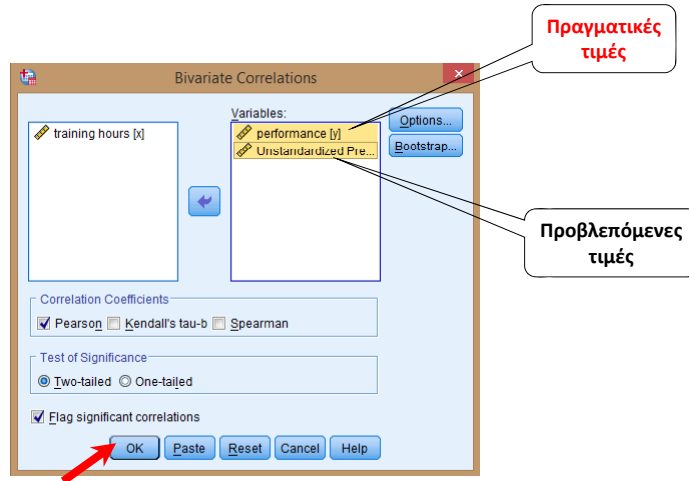
Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

66

Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;



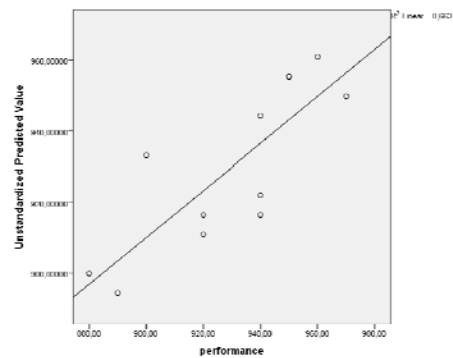
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

67

Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;

Correlations			
		performance	Unstandardized Predicted Value
performance	Pearson Correlation	1	,814**
	Sig. (2-tailed)	.	,001
	N	12	12
Unstandardized Predicted Value	Pearson Correlation	,814**	1
	Sig. (2-tailed)	,001	.
	N	12	12

**. Correlation is significant at the 0.01 level (2-tailed).



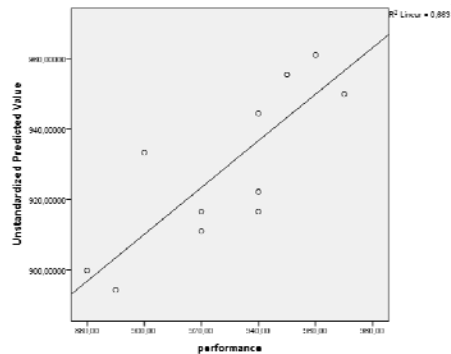
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

68

Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;

Correlations			
		performance	Unstandardized Predicted Value
performance	Pearson Correlation	1	,814**
	Sig. (2-tailed)	.	,001
	N	12	12
Unstandardized Predicted Value	Pearson Correlation	,814**	1
	Sig. (2-tailed)	,001	.
	N	12	12

**. Correlation is significant at the 0.01 level (2-tailed).



ο **συντελεστής συσχέτισης** μεταξύ των **παρατηρούμενων** και **προβλεπόμενων** τιμών της εξαρτημένης μεταβλητής είναι ίσος με το **R**.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	17,21654

a. Predictors: (Constant), training hours

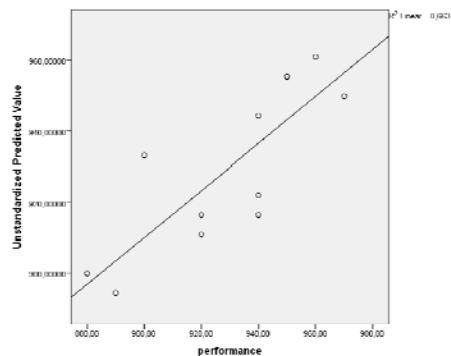
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

69

Ποια είναι η **σχέση** μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;

Correlations			
		performance	Unstandardized Predicted Value
performance	Pearson Correlation	1	,814**
	Sig. (2-tailed)	.	,001
	N	12	12
Unstandardized Predicted Value	Pearson Correlation	,814**	1
	Sig. (2-tailed)	,001	.
	N	12	12

**. Correlation is significant at the 0.01 level (2-tailed).



ο **συντελεστής συσχέτισης** μεταξύ των **παρατηρούμενων** και **προβλεπόμενων** τιμών της εξαρτημένης μεταβλητής είναι ίσος με το **R**.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	17,21654

a. Predictors: (Constant), training hours

Τι ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, ερμηνεύεται από την ανεξάρτητη μεταβλητή X;

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

70

Τι ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, ερμηνεύεται από την ανεξάρτητη μεταβλητή X;

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.814 ^a	.663	.629	17,21654

a. Predictors: (Constant), training hours

Συντελεστής προσδιορισμού R^2 (coefficient of determination) **εκφράζει το ποσοστό της διακύμανσης** (της μεταβλητότητας) **της εξαρτημένης μεταβλητής Y, που ερμηνεύεται** (που μπορεί να προβλεφθεί) **από την ανεξάρτητη μεταβλητή X**

$$R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} \quad \frac{\text{προβλεπόμενη μεταβολή του Y}}{\text{συνολική μεταβολή του Y}}$$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

71

Τι ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, ερμηνεύεται από την ανεξάρτητη μεταβλητή X;

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.814 ^a	.663	.629	17,21654

a. Predictors: (Constant), training hours

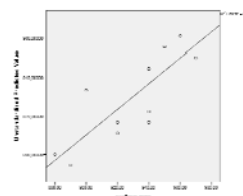
Συντελεστής προσδιορισμού R^2 (coefficient of determination) **εκφράζει το ποσοστό της διακύμανσης** (της μεταβλητότητας) **της εξαρτημένης μεταβλητής Y, που ερμηνεύεται** (που μπορεί να προβλεφθεί) **από την ανεξάρτητη μεταβλητή X**

$$R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} \quad \frac{\text{προβλεπόμενη μεταβολή του Y}}{\text{συνολική μεταβολή του Y}}$$

Όταν $R^2 = 1$
 όλες οι παρατηρούμενες τιμές βρίσκονται πάνω στην γραμμή παλινδρόμησης (τέλεια γραμμική σχέση μεταξύ των δύο μεταβλητών)
Όταν $R^2 = 0$
 δεν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών

$R^2 = 0,663$

Το 66.3% της διακύμανσης της εξαρτημένης μεταβλητής Y, ερμηνεύεται (μπορεί να προβλεφθεί) από την ανεξάρτητη μεταβλητή X



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

72

Τι ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, ερμηνεύεται από την ανεξάρτητη μεταβλητή X;

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.814 ^a	.663	.629	17,21654

a. Predictors: (Constant), training hours

Ο προσαρμοσμένος συντελεστής προσδιορισμού **Adjusted R²** εκφράζει το ποσοστό της διακύμανσης (της μεταβλητότητας) της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X **λαμβάνοντας υπόψη του το μέγεθος του δείγματος**

Τι ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, ερμηνεύεται από την ανεξάρτητη μεταβλητή X;

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.814 ^a	.663	.629	17,21654

a. Predictors: (Constant), training hours

Ο προσαρμοσμένος συντελεστής προσδιορισμού **Adjusted R²** εκφράζει το ποσοστό της διακύμανσης (της μεταβλητότητας) της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X **λαμβάνοντας υπόψη του το μέγεθος του δείγματος**

Adjusted R² = 0,629

Το 62.9% της διακύμανσης της εξαρτημένης μεταβλητής Y, ερμηνεύεται από την ανεξάρτητη μεταβλητή X

To Adjusted R-square υπολογίζεται βάσει της εξίσωσης $1 - ((1 - Rsq)(N - 1) / (N - k - 1))$, Όπου N = μέγεθος δείγματος και k=αριθμός ανεξάρτητων μεταβλητών

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2) \cdot (N - 1)}{(N - k - 1)} = 1 - \frac{(1 - 0.663) \cdot (12 - 1)}{(12 - 1 - 1)} = 1 - \frac{0.337 \cdot 11}{10} = 0,629$$

Όταν το μέγεθος του δείγματος N είναι μεγάλο τότε το Adjusted R-square είναι κοντά με το R - square

**Κατά πόσο το συγκεκριμένο γραμμικό μοντέλο
είναι κατάλληλο για να εκφράσει τα αρχικά δεδομένα;**

**Το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y,
που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι διαφορετικό από το μηδέν (0);**

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5835,906	1	5835,906	19,689	,001 ^a
	Residual	2964,094	10	296,409		
	Total	8800,000	11			

a. Predictors: (Constant), training hours

b. Dependent Variable: performance

Η διακύμανση της εξαρτημένης μεταβλητής Y, **που ερμηνεύεται από την ανεξάρτητη μεταβλητή X** είναι το **Regression Sum of Squares** (= 5835,906).

Η **συνολική** διακύμανση της εξαρτημένης μεταβλητής Y είναι το **Total Sum of Squares** (= 8800,000).

Το υπόλοιπο μέρος της διακύμανσης της εξαρτημένης μεταβλητής, που **δεν ερμηνεύεται** (δεν μπορεί να προβλεφθεί) **από την ανεξάρτητη μεταβλητή** είναι το **Residual Sum of Squares** (=2964,094) και είναι η διαφορά μεταξύ Total Sum of Squares και Regression Sum of Squares

75

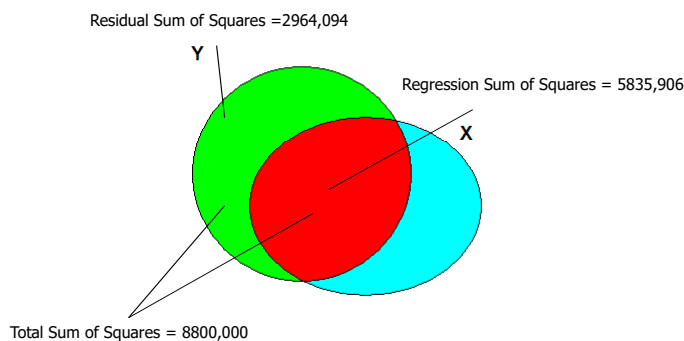
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

**Το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y,
που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι διαφορετικό από το μηδέν (0);**

Η διακύμανση της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι το **Regression Sum of Squares** (= 5835,906).

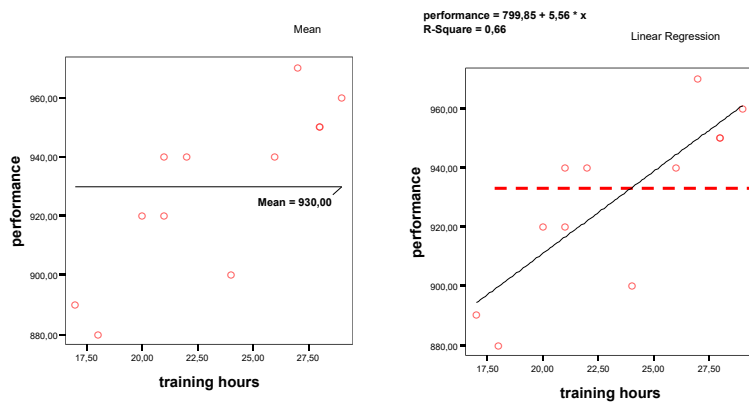
Η συνολική διακύμανση της εξαρτημένης μεταβλητής Y είναι το **Total Sum of Squares** (= 8800,000).

Το υπόλοιπο μέρος της διακύμανσης της εξαρτημένης μεταβλητής, που δεν ερμηνεύεται (δεν μπορεί να προβλεφθεί) από την ανεξάρτητη μεταβλητή είναι το **Residual Sum of Squares** (=2964,094) και είναι η διαφορά μεταξύ Total Sum of Squares και Regression Sum of Squares



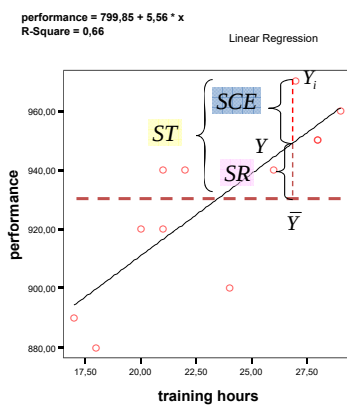
76

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

77



Έλεγχος γραμμικότητας του μοντέλου

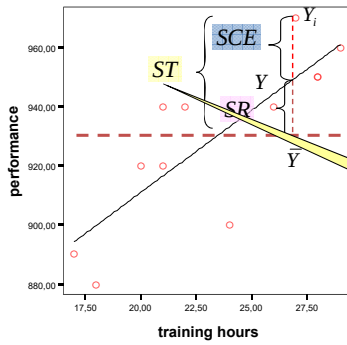
$$\sum (Y_i - \bar{Y})^2 = ST = \text{συνολική διακύμανση της } Y$$

$$ST = SR + SCE$$

78
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

performance = 799,85 + 5,56 * x
R-Square = 0,66

Linear Regression



Έλεγχος γραμμικότητας του μοντέλου

$$\sum (Y_i - \bar{Y})^2 = ST = \text{συνολική διακύμανση της } Y$$

$$ST = SR + SCE$$

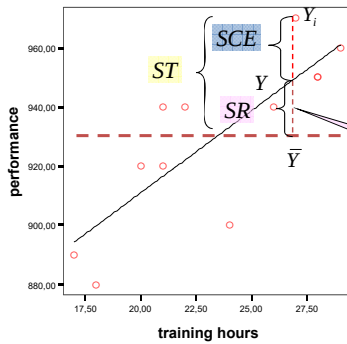
συνολική
διακύμανση
της Y

79

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

performance = 799,85 + 5,56 * x
R-Square = 0,66

Linear Regression



Έλεγχος γραμμικότητας του μοντέλου

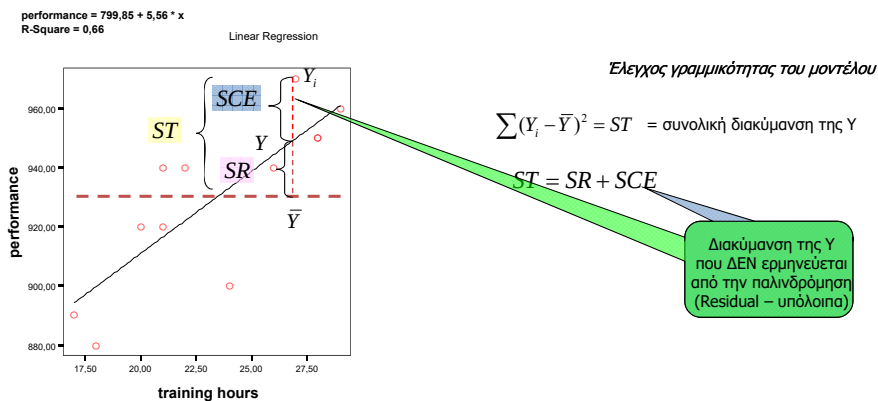
$$\sum (Y_i - \bar{Y})^2 = ST = \text{συνολική διακύμανση της } Y$$

$$ST = SR + SCE$$

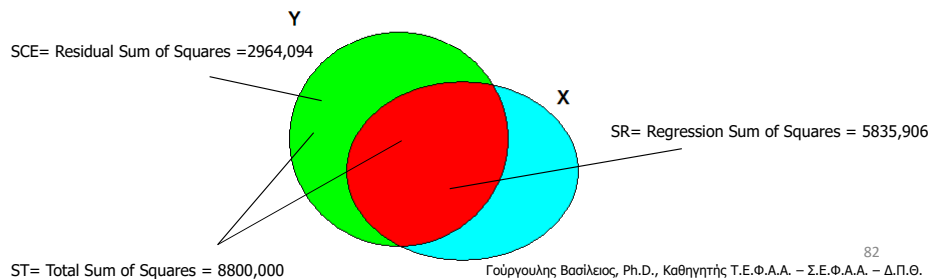
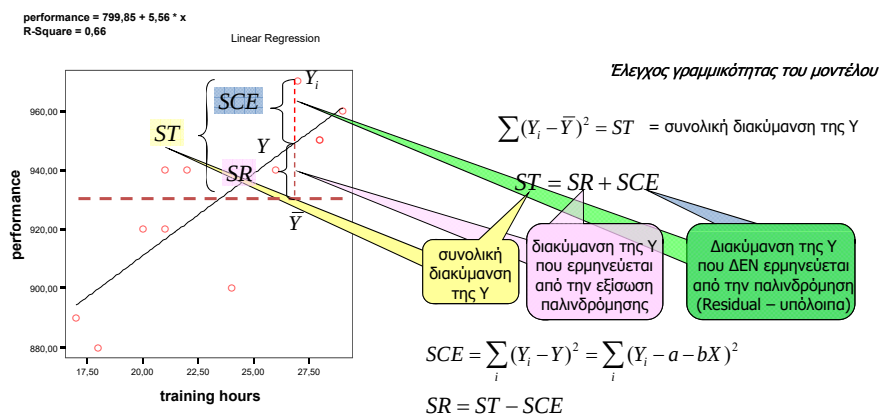
διακύμανση της Y
που ερμηνεύεται
από την εξίσωση
παλινδρόμησης

80

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



81
Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.



Κατά πόσο το συγκεκριμένο γραμμικό μοντέλο είναι κατάλληλο για να εκφράσει τα αρχικά δεδομένα;

Το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι διαφορετικό από το μηδέν (0);

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5835,906	1	5835,906	19,689	,001 ^a
	Residual	2964,094	10	296,409		
	Total	8800,000	11			

a. Predictors: (Constant), training hours

b. Dependent Variable: performance

Το πηλίκο μεταξύ της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X (Regression Sum of Squares = 5835,906), προς τη συνολική διακύμανση της εξαρτημένης μεταβλητής Y (Total Sum of Squares = 8800,000), δηλαδή το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X, ισούται με τον συντελεστή προσδιορισμού R-square = $(5835,906/8800,000) = 0,663$.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	17,21654

a. Predictors: (Constant), training hours

83

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

Κατά πόσο το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι στατιστικά σημαντικά διαφορετικό από το μηδέν (0);

Για να είναι το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X, δηλαδή ο συντελεστής προσδιορισμού R-square = $(5835,906/8800,000) = 0,663$, στατιστικά σημαντικά διαφορετικός από το μηδέν (0), θα πρέπει το ποσό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X (Regression Sum of Squares = 5835,906), να είναι κατά πολύ μεγαλύτερο από το ποσό της διακύμανσης της εξαρτημένης μεταβλητής Y, που δεν ερμηνεύεται (δεν μπορεί να προβλεφθεί) από την ανεξάρτητη μεταβλητή X (Residual Sum of Squares = 2964,094)

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5835,906	1	5835,906	19,689	,001 ^a
	Residual	2964,094	10	296,409		
	Total	8800,000	11			

a. Predictors: (Constant), training hours

b. Dependent Variable: performance

Διαιρώντας τα αθροίσματα τετραγώνων (Sum of Squares) με τους αντίστοιχους βαθμούς ελευθερίας (df) προκύπτουν τα μέσα τετράγωνα (Mean Squares).

Διαιρώντας το Mean Square Regression (=5835,906) με το Mean Square Residual (=296,409) προκύπτει η $F = 19,689$

84

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

Κατά πόσο το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι στατιστικά σημαντικά διαφορετικό από το μηδέν (0);

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5835,906	1	5835,906	19,689	,001 ^a
	Residual	2964,094	10	296,409		
	Total	8800,000	11			

a. Predictors: (Constant), training hours

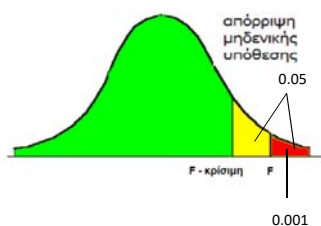
b. Dependent Variable: performance

Όταν η F – τιμή είναι μεγαλύτερη από την F – κρίσιμη (βρίσκεται δεξιά της) σημαίνει ότι το Mean Square Regression ΕΙΝΑΙ πολύ μεγαλύτερο από το Mean Square Residual,

οπότε το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X (Mean Square Regression) είναι στατιστικά σημαντικά κατά πολύ μεγαλύτερο από το ποσό της διακύμανσης της εξαρτημένης μεταβλητής Y, που δεν ερμηνεύεται από την ανεξάρτητη μεταβλητή X (Mean Square Residual).

Συνεπώς το R-square είναι στατιστικά σημαντικά διαφορετικό από το μηδέν (0).

Το επίπεδο σημαντικότητας που αντιστοιχεί στην F – τιμή (κόκκινη περιοχή) είναι μικρότερο από το επίπεδο σημαντικότητας 0.05 (κίτρινη + κόκκινη περιοχή).



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

85

Κατά πόσο το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι στατιστικά σημαντικά διαφορετικό από το μηδέν (0);

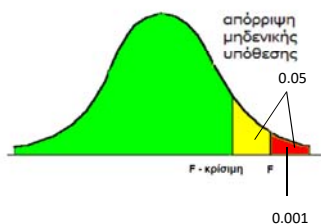
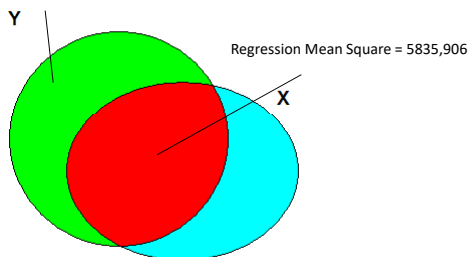
ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5835,906	1	5835,906	19,689	,001 ^a
	Residual	2964,094	10	296,409		
	Total	8800,000	11			

a. Predictors: (Constant), training hours

b. Dependent Variable: performance

Residual Mean Square =296,409



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

86

Σύνοψη

Ποια ευθεία (γραμμή παλινδρόμησης)
εκφράζει καλύτερα (με μεγαλύτερη ακρίβεια) τα δεδομένα;



Η γραμμή βάσει της οποίας
το άθροισμα των τετραγώνων της διαφοράς
ΟΛΩΝ των πραγματικών από τις προβλεπόμενες τιμές
είναι το ελάχιστο δυνατό.

$$\sum (Y - \hat{Y})^2 = \text{ελάχιστο}$$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

87

Σύνοψη

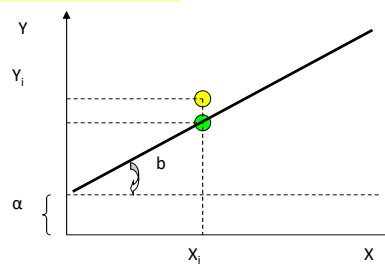
Πώς καθορίζεται η διεύθυνση αυτής της γραμμής παλινδρόμησης;



Καθορίζοντας τους συντελεστές παλινδρόμησης (α, b)

$$Y = a + b \cdot X_1 + e$$

$$\hat{Y} = a + b \cdot X_1$$



Πώς καθορίζονται οι συντελεστές παλινδρόμησης;



Από τον πίνακα Coefficients (Unstandardized Coefficients B)

Model	Coefficients ^a				
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	799,850	29,750		26,886	,000
training hours	5,558	1,253	,814	4,437	,001

a. Dependent Variable: performance

$$\hat{Y} = a + b \cdot X_1 \Rightarrow$$

Performance = 799.850 + 5.558 • training hours

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

88

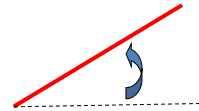
Σύνοψη

Πότε υπάρχει πραγματικά σχέση

μεταξύ της εξαρτημένης (Y) και της ανεξάρτητης (X) μεταβλητής;



Όταν ο **συντελεστής B** (Unstandardized Coefficient) που δηλώνει την κλίση της γραμμής παλινδρόμησης είναι στατιστικά σημαντικά **διαφορετικός από το 0**.



Πότε ο συντελεστής B (Unstandardized Coefficient) που δηλώνει την κλίση της γραμμής παλινδρόμησης είναι στατιστικά σημαντικά **διαφορετικός από το 0**.



Όταν το επίπεδο σημαντικότητας που αντιστοιχεί στην **t – τιμή** είναι μικρότερο από 0.05.



Coefficients ^a					
Model	Unstandardized Coefficients		Standardized Coefficients		Sig.
	B	Std. Error	Beta	t	
1	799,850	29,750		26,886	,000
	5,558	1,253	,814	4,437	,001
a. Dependent Variable: performance					

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

89

Σύνοψη

Πως καθορίζεται και ποια είναι η σχέση μεταξύ **πραγματικών** και **προβλεπόμενων** τιμών;



Βάσει του **συντελεστή συσχέτισης R** μεταξύ των **παρατηρούμενων** και **προβλεπόμενων** τιμών της εξαρτημένης μεταβλητής.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	17,21654
a. Predictors: (Constant), training hours				

Πως καθορίζεται το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y, που **ερμηνεύεται** από την ανεξάρτητη μεταβλητή X;



Βάσει του **συντελεστή προσδιορισμού R²** (coefficient of determination)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,814 ^a	,663	,629	17,21654
a. Predictors: (Constant), training hours				

$$R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} = 66.3\%$$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

90

Σύνοψη

Πως καθορίζεται αν το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής Y , που ερμηνεύεται από την ανεξάρτητη μεταβλητή X είναι διαφορετικό από το μηδέν (0);



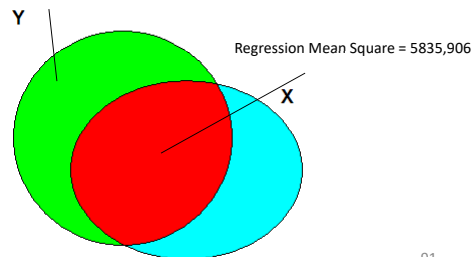
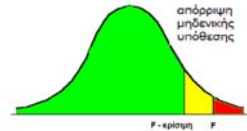
Όταν το επίπεδο σημαντικότητας της F – τιμής είναι μικρότερο από 0.05 και η F -τιμή είναι μεγαλύτερη από την F – κρίσιμη (βρίσκεται δεξιά της) σημαίνει ότι το Mean Square Regression ΕΙΝΑΙ πολύ μεγαλύτερο από το Mean Square Residual.

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5835,906	1	5835,906	19,689	,001 ^a
	Residual	2964,094	10	296,409		
	Total	8800,000	11			

a. Predictors: (Constant), training hours

b. Dependent Variable: performance

Residual Mean Square = 296,409



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

91

Συγγραφή αποτελεσμάτων ...

Από την εφαρμογή της απλής γραμμικής παλινδρόμησης, για την πρόβλεψη των τιμών της ποσοτικής εξαρτημένης μεταβλητής (επίδοση – performance) βάσει των τιμών στην ανεξάρτητη μεταβλητή (ώρες προπόνησης – training hours), διαπιστώθηκε ότι

το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που ερμηνεύεται από την ανεξάρτητη μεταβλητή ανέρχεται σε 66,3% ($R^2 = 0.663$) και είναι στατιστικά σημαντικό ($F_{1,10} = 19.689$; $p < 0.05$).

Η εξίσωση πρόβλεψης είναι:

$$\text{Performance} = 799.850 + 5.558 \times \text{training hours}$$

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

92

Ερωτήσεις;



Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

93

Youtube video_ Linear Regression

Linear Regression (part 1): https://www.youtube.com/watch?v=YcNGKam_wwQ

Linear Regression (part 2): <https://www.youtube.com/watch?v=U4QL8QCbil8>

Linear Regression (part 3): <https://www.youtube.com/watch?v=YdCLztrl73s>

Linear Regression (part 4): <https://www.youtube.com/watch?v=ynijS2McieQ>

Linear Regression (part 5): <https://www.youtube.com/watch?v=Q6CkRWXvCZw>

Linear Regression - SPSS (part 1): <https://www.youtube.com/watch?v=0AGLdgUtIjg>

Linear Regression - SPSS (part 2): <https://www.youtube.com/watch?v=VEQPX6d-EQw>

Linear Regression - SPSS (part 3): https://www.youtube.com/watch?v=e6yLXpsxK_4

Linear Regression - SPSS (part 4): https://www.youtube.com/watch?v=LPoy_dfuTFU

Γούργουλης Βασίλειος, Ph.D., Καθηγητής Τ.Ε.Φ.Α.Α. – Σ.Ε.Φ.Α.Α. – Δ.Π.Θ.

94