

# Εξόρυξη δεδομένων με χρήση τεχνικών μηχανικής μάθησης

Τμ. Μηχανικών Πληροφορικής και Υπολογιστών



# Ανάλυση Παλινδρόμησης

# Ανάλυση Παλινδρόμησης – Παράδειγμα Υπολογισμού Ευθείας

Έστω ότι μια εταιρεία θέλει να μελετήσει την συσχέτιση ανάμεσα στον αριθμό των ωρών που απασχολήθηκαν οι εργαζόμενοι της σε ένα project και τις αμοιβές που έλαβαν.

Ο αριθμός των ωρών αποτελεί την ανεξάρτητη μεταβλητή

$X$ =Αριθμός ωρών

Οι αμοιβές αποτελούν την εξαρτημένη μεταβλητή

$Y$ =Αμοιβές

# Ανάλυση Παλινδρόμησης – Παράδειγμα Υπολογισμού Ευθείας

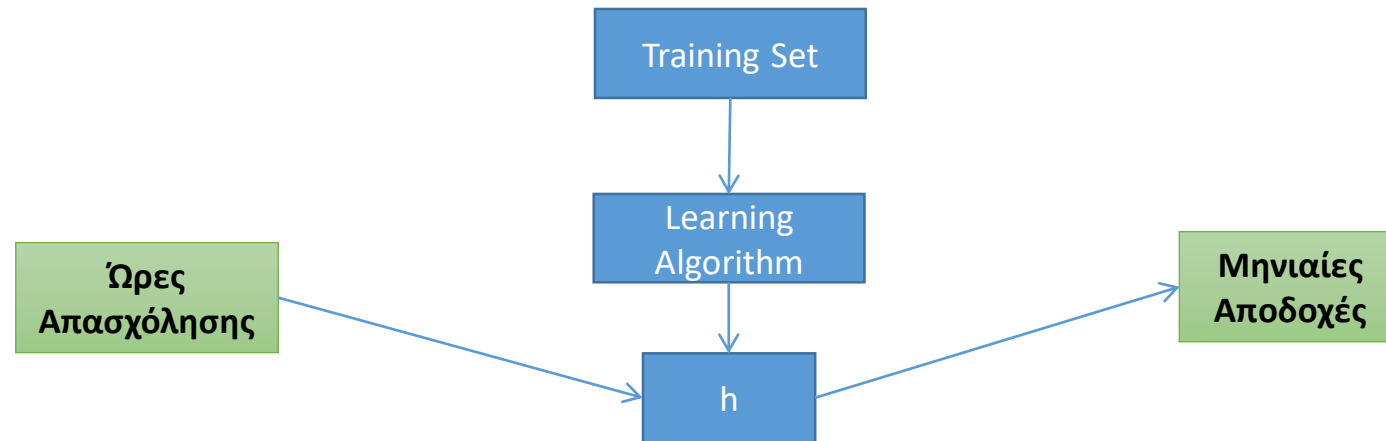
Προκειμένου να υπολογίσουμε την ευθεία της εξίσωσης έχουμε συγκεντρώσει 10 διαφορετικές περιπτώσεις εργαζομένων.

Οπότε στον διπλανό πίνακα μπορούμε να παρατηρήσουμε τις ώρες απασχόλησης μαζί με τις αντίστοιχες αμοιβές τους.

A/A	Ώρες	Μισθός
1	15	641
2	8	422
3	22	780
4	4	341
5	22	800
6	11	500
7	17	780
8	9	550
9	7	500
10	21	760

# Ανάλυση Παλινδρόμησης

Προκειμένου να υπολογίσουμε την εξίσωση θα χρησιμοποιήσουμε τα 10 δείγματα ως training set βάσει των οποίων θα δημιουργήσουμε την αντίστοιχη εξίσωση των ελαχίστων τετραγώνων (Learning Algorithm). Οπότε στην συνέχεια για συγκεκριμένες τιμές εισόδου θα προβλέπουμε τις αντίστοιχες τιμές εξόδου.



# Ανάλυση Παλινδρόμησης-Μέθοδος ελαχίστων τετραγώνων

Ας ξανά θυμηθούμε τον τρόπο υπολογισμού των  $\beta_0, \beta_1$

$$\hat{\beta}_1 = \frac{v \sum_{i=1}^v x_i y_i - (\sum_{i=1}^v x_i)(\sum_{i=1}^v y_i)}{v \sum_{i=1}^v x_i^2 - (\sum_{i=1}^v x_i)^2}, \hat{\beta}_0 = \frac{1}{v} \sum_{i=1}^v y_i - \hat{\beta}_1 \cdot \frac{1}{v} \sum_{i=1}^v x_i$$

# Ανάλυση Παλινδρόμησης

$X_i$	$Y_i$	$X_i^2$	$X_i Y_i$
15	641	225	9615
8	422	64	3376
22	780	484	17160
4	341	16	1364
22	800	484	17600
11	500	121	5500
17	780	289	13260
9	550	81	4950
7	500	49	3500
21	760	441	15960
136	6074	2254	92285
$\sum X_i$	$\sum Y_i$	$\sum X_i^2$	$\sum X_i Y_i$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2},$$

$$\hat{\beta}_1 = \frac{10 * 92285 - 136 * 6074}{10 * 2254 - 136^2},$$

$$\hat{\beta}_1 = \frac{96786}{4044},$$

$$\hat{\beta}_1 = 23,933$$

# Ανάλυση Παλινδρόμησης

$X_i$	$Y_i$	$X_i^2$	$X_i Y_i$
15	641	225	9615
8	422	64	3376
22	780	484	17160
4	341	16	1364
22	800	484	17600
11	500	121	5500
17	780	289	13260
9	550	81	4950
7	500	49	3500
21	760	441	15960
136	6074	2254	92285
$\sum X_i$	$\sum Y_i$	$\sum X_i^2$	$\sum X_i Y_i$

$$\hat{\beta}_0 = \frac{1}{v} \sum_{i=1}^v y_i - \hat{\beta}_1 \cdot \frac{1}{v} \sum_{i=1}^v x_i$$

$$\hat{\beta}_0 = \frac{1}{10} 6074 - 23,933 * \frac{1}{10} * 136$$

$$\hat{\beta}_0 = 281.9112$$

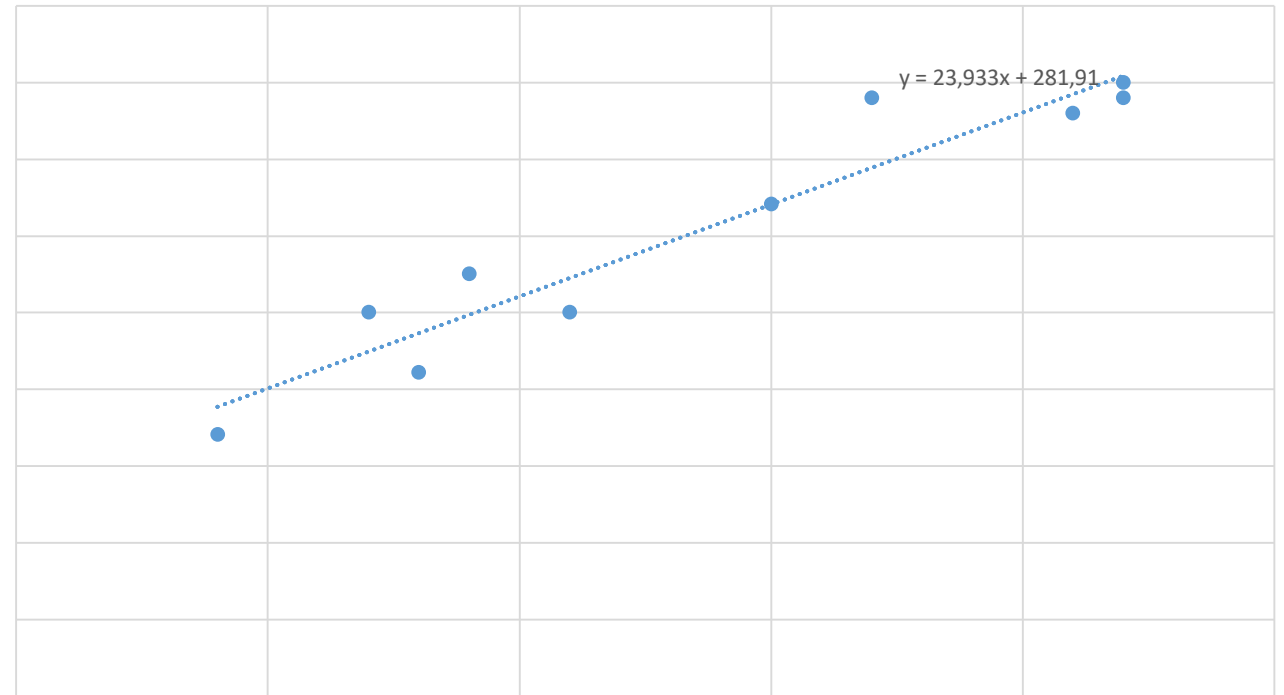


# Ανάλυση Παλινδρόμησης

$X_i$	$Y_i$	$X_i^2$	$X_i Y_i$
15	641	225	9615
8	422	64	3376
22	780	484	17160
4	341	16	1364
22	800	484	17600
11	500	121	5500
17	780	289	13260
9	550	81	4950
7	500	49	3500
21	760	441	15960
136	6074	2254	92285

$$\sum X_i \quad \sum Y_i \quad \sum X_i^2 \quad \sum X_i Y_i$$

$$y = \hat{\beta}_0 + \hat{\beta}_1 X = 281,91 + 23,933X$$



# Ανάλυση Παλινδρόμησης

$X_i$	$Y_i$	$X_i^2$	$X_i Y_i$	$\hat{y}_i$	$\hat{y}_i - \bar{y}_i$	$y_i - \bar{y}_i$	$(\hat{y}_i - \bar{y}_i)^2$	$(y_i - \bar{y}_i)^2$
15	641	225	9615	640.91	33.87	33.96	1146.838	1153.282
8	422	64	3376	473.37	-134	-185.04	17866.6	34239.8
22	780	484	17160	808.44	201.4	172.96	40560.35	29915.16
4	341	16	1364	377.64	-229	-266.04	52623.44	70777.28
22	800	484	17600	808.44	201.4	192.96	40560.35	37233.56
11	500	121	5500	545.17	-61.9	-107.04	3827.526	11457.56
17	780	289	13260	688.77	81.73	172.96	6679.956	29915.16
9	550	81	4950	497.31	-110	-57.04	12041.33	3253.562
7	500	49	3500	449.44	-158	-107.04	24837.44	11457.56
21	760	441	15960	784.5	177.5	152.96	31493.12	23396.76
136	6074	2254	92285		3.588	3.6	231637	252799.7

$$R^2 = SSR/SST$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

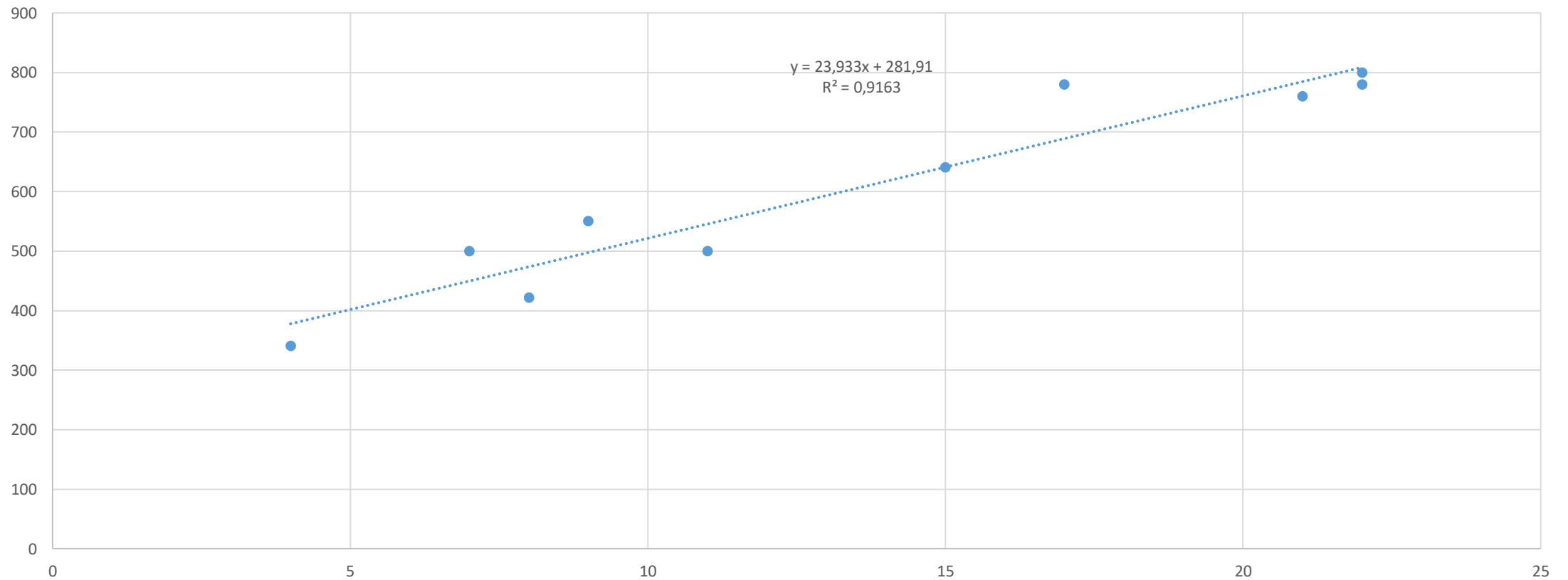
$$R^2 = SSR/SST$$

$$R^2 = 252799.7 / 231637$$

$$R^2 = 0,9163$$

# Ανάλυση Παλινδρόμησης

Τίτλος γραφήματος



## Παράδειγμα 2

# Ανάλυση Παλινδρόμησης

X	Y	X <sup>2</sup>	XY
1	2	1	2
2	4	4	8
3	5	9	15
4	4	16	16
5	5	25	25
15	20	55	66

$$\sum X_i$$

$$\sum Y_i$$

$$\sum X_i^2$$

$$\sum X_i Y_i$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2},$$

$$\hat{\beta}_1 = \frac{5 * 66 - 15 * 20}{5 * 55 - 15^2},$$

$$\hat{\beta}_1 = \frac{30}{50},$$

$$\hat{\beta}_1 = 0,6$$

# Ανάλυση Παλινδρόμησης

X	Y	X <sup>2</sup>	XY
1	2	1	2
2	4	4	8
3	5	9	15
4	4	16	16
5	5	25	25
15	20	55	66

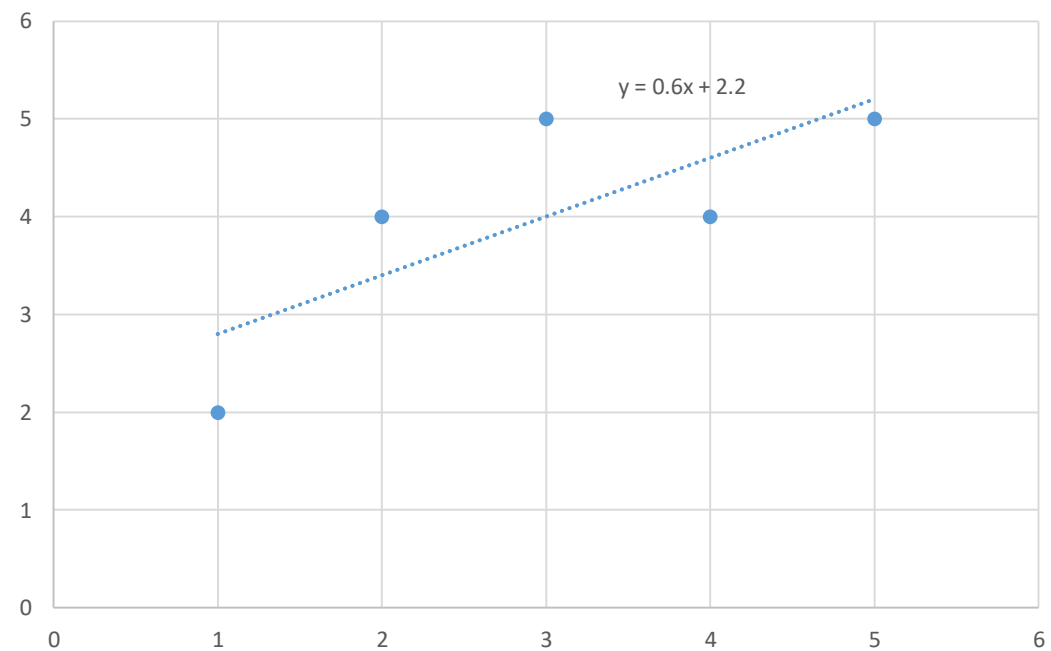
$$\hat{\beta}_0 = \frac{1}{v} \sum_{i=1}^v y_i - \hat{\beta}_1 \cdot \frac{1}{v} \sum_{i=1}^v x_i$$

$$\hat{\beta}_0 = \frac{1}{5} 20 - 0,6 * \frac{1}{5} * 15$$

$$\hat{\beta}_0 = 2.2$$

# Ανάλυση Παλινδρόμησης

X	Y	X <sup>2</sup>	XY
1	2	1	2
2	4	4	8
3	5	9	15
4	4	16	16
5	5	25	25
15	20	55	66



# Ανάλυση Παλινδρόμησης

X	Y	X <sup>2</sup>	XY	$y_i - \bar{y}_i$	$(y_i - \bar{y}_i)^2$	$\hat{y}_i$	$\hat{y}_i - \bar{y}_i$	$(\hat{y}_i - \bar{y}_i)^2$
1	2	1	2	-2	4	2.8	-1.2	1.44
2	4	4	8	0	0	3.4	-0.6	0.36
3	5	9	15	1	1	4	0	0
4	4	16	16	0	0	4.6	0.6	0.36
5	5	25	25	1	1	5.2	1.2	1.44
15	20	55	66	0	6	20	0	3.6

$$R^2 = \text{SSR}/\text{SST}$$

$$\text{SSR} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2$$

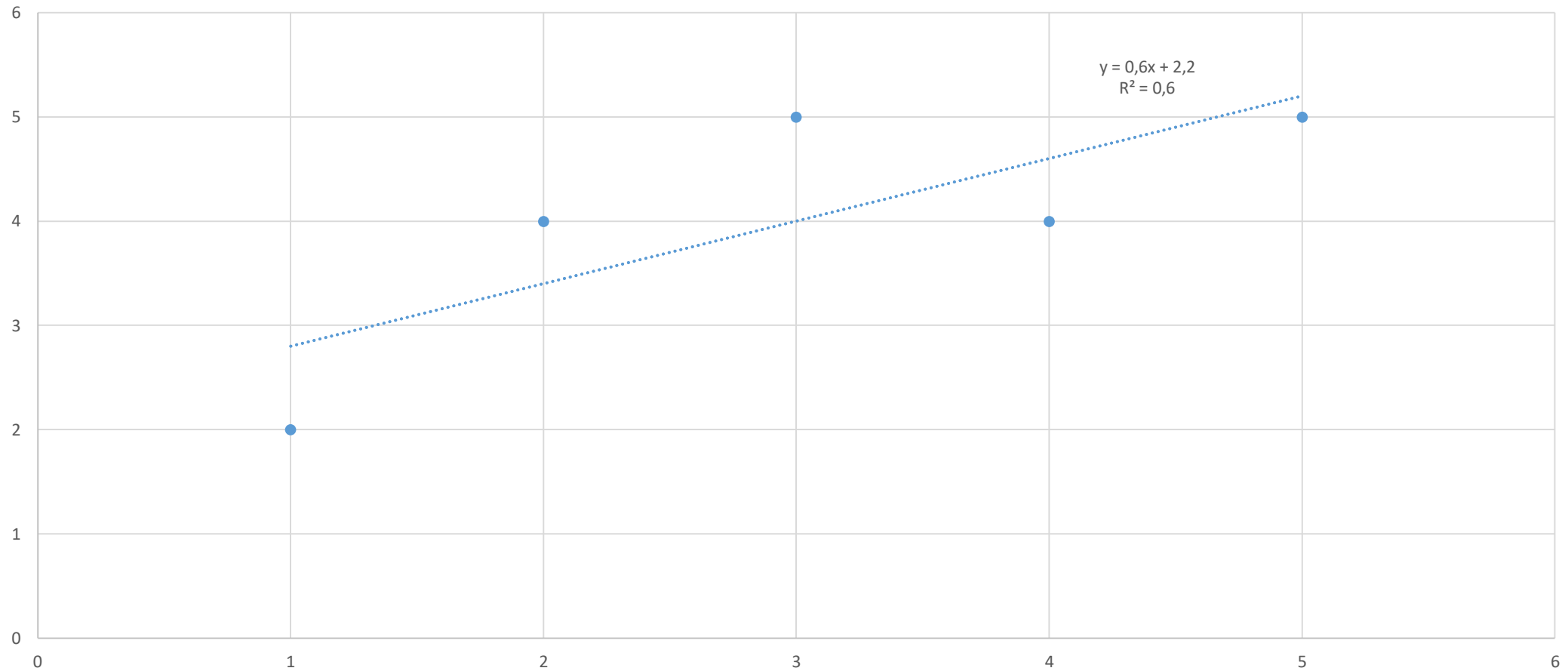
$$\text{SST} = \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

$$R^2 = 3.6/6$$

$$R^2 = 0,6$$



# Ανάλυση Παλινδρόμησης



ΤΕΛΟΣ

