ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ ΕΙΔΙΚΑ ΘΕΜΑΤΑ

Κεφάλαιο 16

Απλή Γραμμική Παλινδρόμηση και Συσχέτιση

Επιμέλεια παρουσιάσεων: Δρ. Αλέκα Καλαπόδη

Ανάλυση Παλινδρόμησης …

Ο στόχος μας είναι να ***αναλύσουμε τη σχέση*** μεταξύ συνεχών μεταβλητών. Η ***ανάλυση παλινδρόμησης*** είναι το πρώτο εργαλείο που θα μελετήσουμε.

Η ανάλυση παλινδρόμησης χρησιμοποιείται για την πρόβλεψη της τιμής μιας μεταβλητής (***εξαρτημένη μεταβλητή***) με βάση την τιμή άλλων μεταβλητών (***ανεξάρτητες μεταβλητές***).

Εξαρτημένη μεταβλητή: συμβολίζεται με **Y**

Ανεξάρτητες μεταβλητές: συμβολίζονται με $X_1, X_2, \ldots, X_k$

Ανάλυση Παλινδρόμησης …

Εάν μας ενδιαφέρει να καθορίσουμε ***μόνο*** το εάν ***υπάρχει*** σχέση, χρησιμοποιούμε ***ανάλυση συσχέτισης***, μια τεχνική που έχουμε ήδη δει.

Στο κεφάλαιο αυτό θα εξετάσουμε τη σχέση μεταξύ ***δύο μεταβλητών***, με ***απλή γραμμική παλινδρόμηση***.

Οι μαθηματικές εξισώσεις που περιγράφουν αυτές τις σχέσεις καλούνται και ***μοντέλα***, και ταξινομούνται σε δύο κατηγορίες: ντετερμινιστικά ή πιθανοθεωρητικά.

Μοντέλα …

Ντετερμινιστικό Μοντέλο: μια εξίσωση ή σύνολο εξισώσεων που μας επιτρέπει να ***καθορίσουμε πλήρως*** την τιμή της εξαρτημένης μεταβλητής από τις τιμές των ανεξάρτητων μεταβλητών.

Σε αντίθεση με …

Πιθανοθεωρητικό Μοντέλο: μια μέθοδος για να υπολογίσουμε την ***τυχαιότητα*** που υπάρχει στις πραγματικές διαδικασίες.

Π.χ. έχουν πωληθεί όλα τα σπίτια ίσου εμβαδού στην ίδια ακριβώς τιμή;

Ένα Μοντέλο …

Για να δημιουργήσουμε ένα πιθανοθεωρητικό μοντέλο, ξεκινάμε με ένα ντετερμινιστικό μοντέλο το οποίο ***προσεγγίζει τη σχέση*** που θέλουμε να προσδιορίσουμε και προσθέτουμε έναν ***τυχαίο όρο*** που μετράει το ***σφάλμα*** της ντετερμινιστικής συνιστώσας;.

Ντετερμινιστικό Μοντέλο:

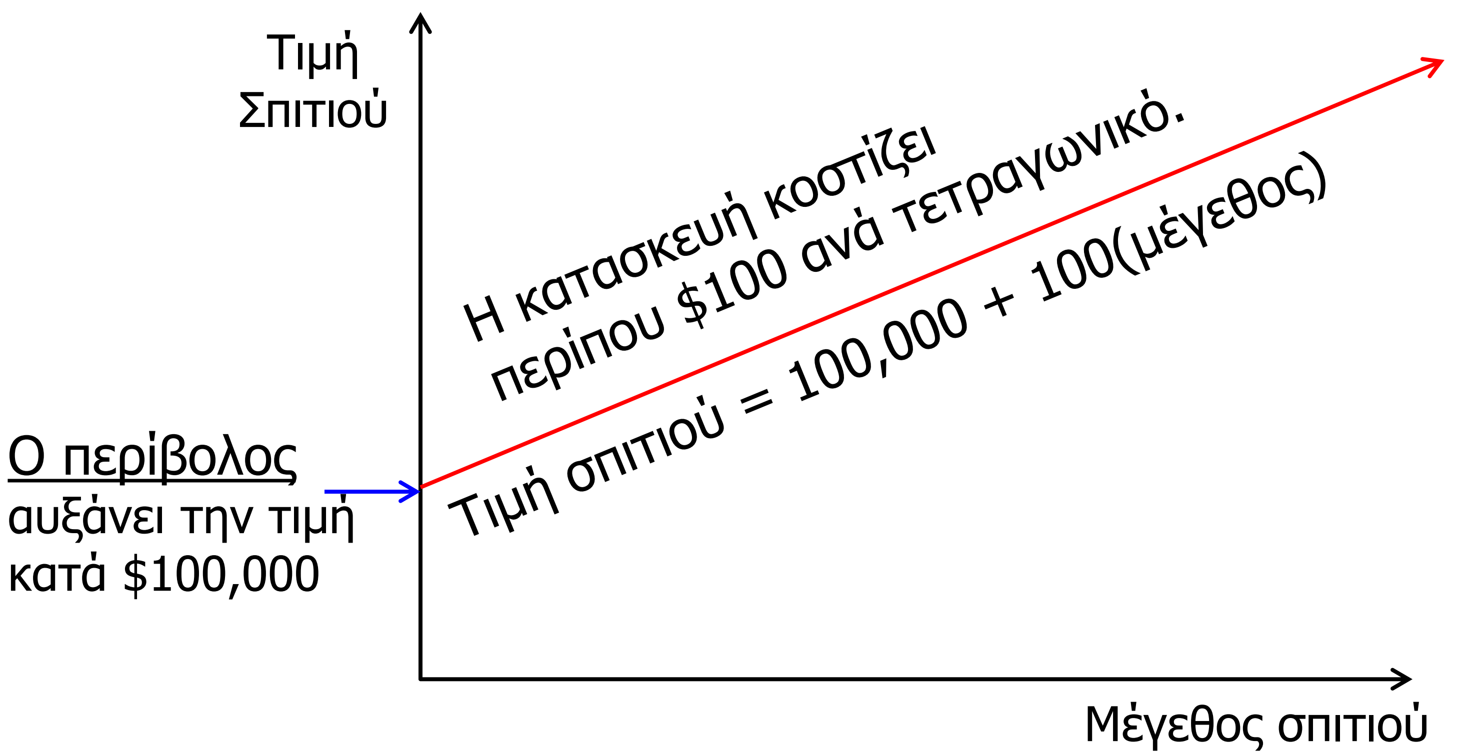
Το κόστος κατασκευής ενός νέου σπιτιού είναι περίπου \$100 ανά τετραγωνικό πόδι (ft^2) ενώ ο περίβολος αυξάνει την αξία περίπου κατά \$100,000. Επομένως, η εκτιμώμενη τιμή πώλησης (**y**) είναι:

$$\mathbf{y = \$100{,}000 + (100\$/ft^2)(x)}$$

(όπου **x** τα τετραγωνικά πόδια του σπιτιού)

Ένα Μοντέλο …

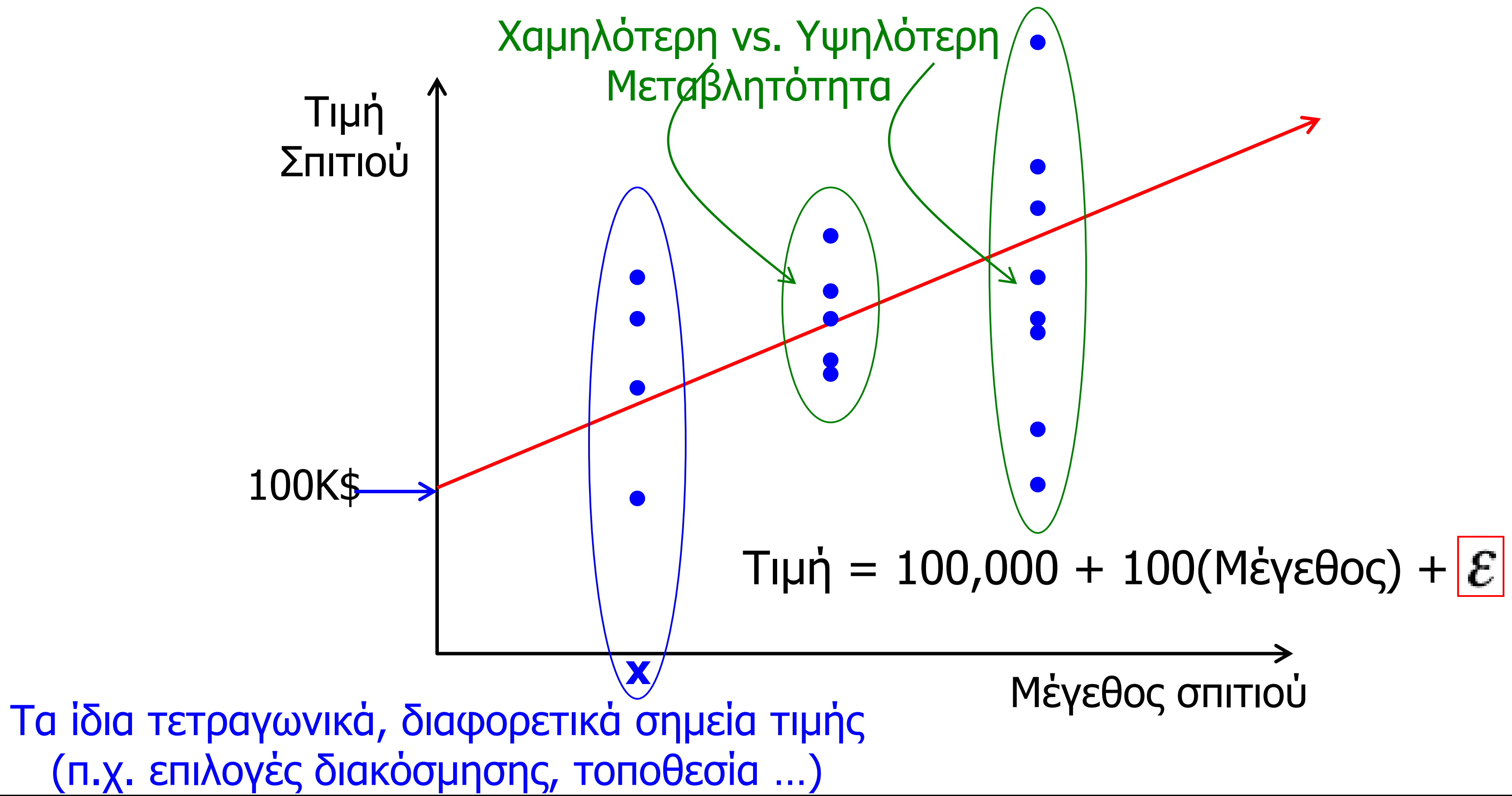
Ένα μοντέλο της σχέσης μεταξύ του μεγέθους του σπιτιού (ανεξάρτητη μεταβλητή) και της τιμής του (εξαρτημένη μεταβλητή) θα ήταν:



Στο μοντέλο αυτό, η τιμή είναι πλήρως **καθορισμένη** από το μέγεθος.

Ένα Μοντέλο …

Στην πραγματικότητα όμως, η τιμή του σπιτιού διαφοροποιείται ακόμα και μεταξύ σπιτιών ίδιου μεγέθους:



Τυχαίος Όρος ...

Αναπαριστούμε την τιμή ενός σπιτιού ως συνάρτηση του μεγέθους του στο Πιθανοθεωρητικό Μοντέλο:

$$y = 100{,}000 + 100x + \varepsilon$$

όπου ε είναι ο ***τυχαίος όρος*** (***ή μεταβλητή σφάλματος***).

Είναι η διαφορά μεταξύ της ***πραγματικής*** τιμής πώλησης και της ***εκτιμώμενης*** με βάση το μέγεθος του σπιτιού. Η τιμή του διαφοροποιείται μεταξύ των πωλήσεων, ακόμα κι αν τα τετραγωνικά (δηλ. **x**) παραμένουν ίδια.

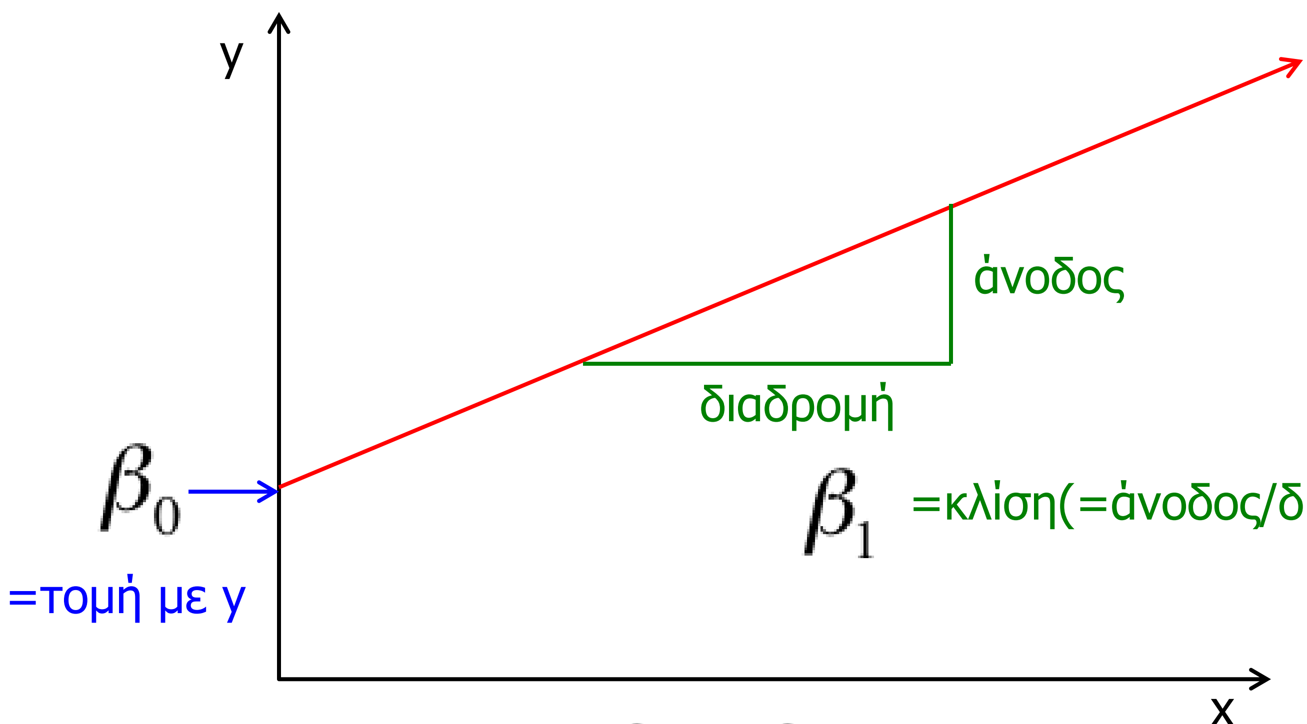
Απλή Γραμμική Παλινδρόμηση …

Ένα μοντέλο ευθείας γραμμής με μια ανεξάρτητη μεταβλητή καλείται ***γραμμικό μοντέλο πρώτης τάξης*** ή ***σμοντέλο απλής γραμμικής παλινδρόμησης***. Δίνεται από την:

$$\overbrace{y}^{\text{εξαρτημένη μεταβλητή}} = \underbrace{\beta_0}_{\text{Τομή με y}} + \underbrace{\beta_1}_{\text{Κλίση ευθείας}} \overbrace{x}^{\text{ανεξάρτητη μεταβλητή}} + \underbrace{\varepsilon}_{\text{Μεταβλητή σφάλματος}}$$

Απλή Γραμμική Παλινδρόμηση ...

Σημειώστε ότι και το β_0 και το β_1 είναι ***παράμετροι του πληθυσμού*** οι οποίες είναι συνήθως άγνωστες και επομένως ***εκτιμώνται*** από τα δεδομένα.



$$y = \beta_0 + \beta_1 x + \varepsilon$$

Εκτίμηση Συντελεστών ...

Θα εκτιμήσουμε το β_0 χρησιμοποιώντας το b_0 και το β_1 χρησιμοποιώντας το b_1, την τομή με y και την κλίση (αντίστοιχα) της ***ευθείας ελαχίστων τετραγώνων*** ή ***ευθείας παλινδρόμησης*** που δίνεται από την:

$$\boxed{\hat{y} = b_0 + b_1 x}$$

(Θυμίζουμε ότι είναι μια εφαρμογή της μεθόδου ελαχίστων τετραγώνων και δημιουργεί μια ευθεία η οποία ***ελαχιστοποιεί*** το άθροισμα των τετραγώνων των αποκλίσεων μεταξύ των σημείων και της ευθείας)

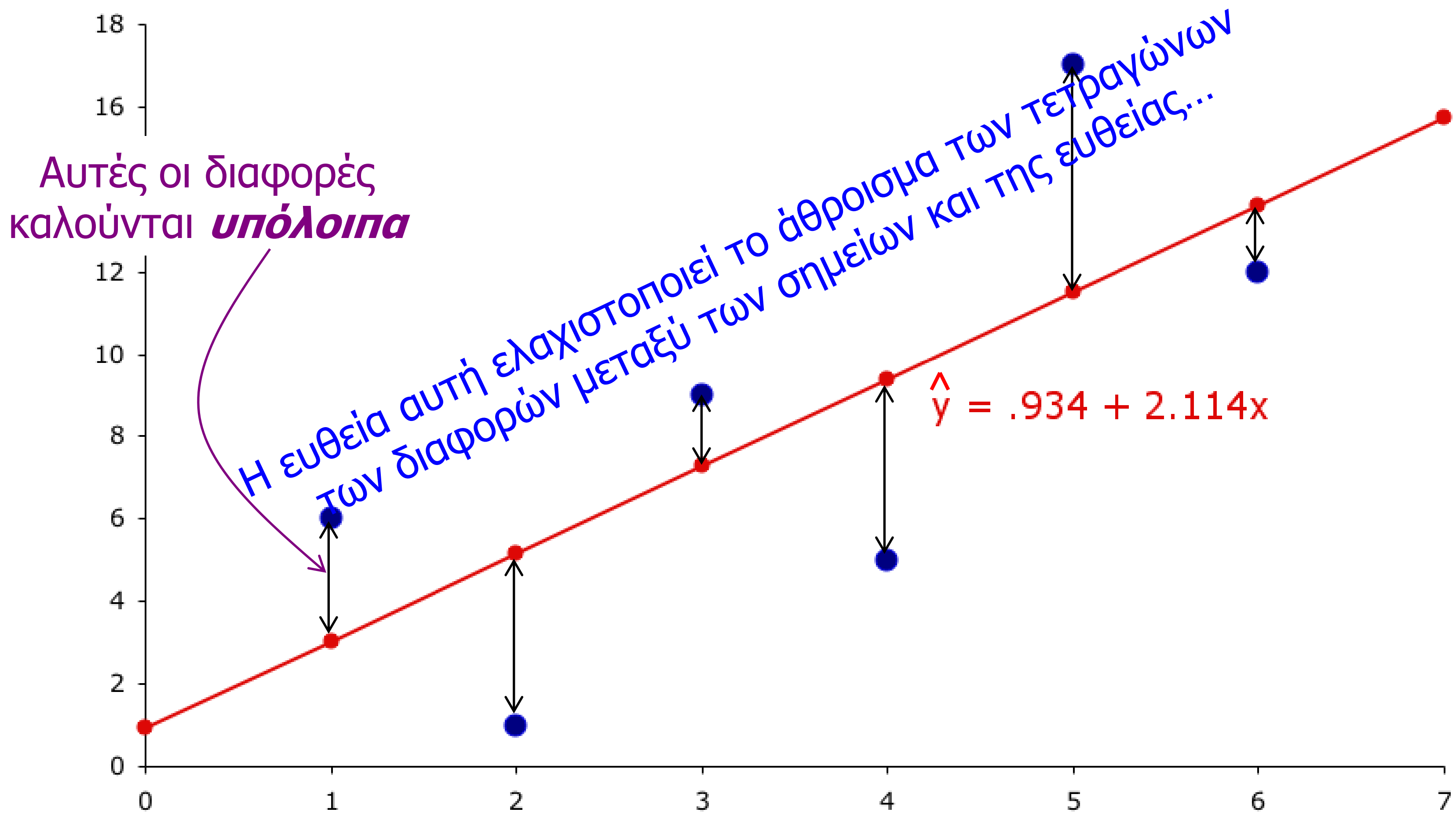
Παράδειγμα 16.1

Το πριμ απόδοσης (σε χιλιάδες δολάρια) έξι υπαλλήλων με διαφορετικά χρόνια προϋπηρεσίας δίνεται στον παρακάτω πίνακα. Θέλουμε να καθορίσουμε τη γραμμική σχέση μεταξύ του πριμ και των ετών προϋπηρεσίας.

Έτη προϋπηρεσίας x	1	2	3	4	5	6
Ετήσιο πριμ y	6	1	9	5	17	12

Ευθεία Ελαχίστων Τετραγώνων...

Παράδειγμα 16.1



Παράδειγμα 16.2…

Οι έμποροι αυτοκινήτων στη Βόρεια Αμερική χρησιμοποιούν το «Κόκκινο Βιβλίο» για να καθορίσουν την τιμή των μεταχειρισμένων αυτοκινήτων που παίρνουν με ανταλλαγή όταν οι πελάτες τους αγοράζουν καινούργιο.

Το βιβλίο εκδίδεται κάθε μήνα και καταγράφει τις τιμές για όλα τα βασικά μοντέλα αυτοκινήτων.

Έχει εναλλακτικά τιμές για κάθε μοντέλο ανάλογα με την κατάστασή του και τον εξοπλισμό του.

Οι τιμές καθορίζονται με βάση τη μέση τιμή σε πρόσφατες δημοπρασίες.

Παράδειγμα 16.2…

Ωστόσο, το Κόκκινο Βιβλίο δεν δίνει την τιμή με βάση την ένδειξη των χιλιομέτρων, παρά το γεγονός ότι αυτό αποτελεί σημαντικό παράγοντα για τους αγοραστές μεταχειρισμένων αυτοκινήτων.

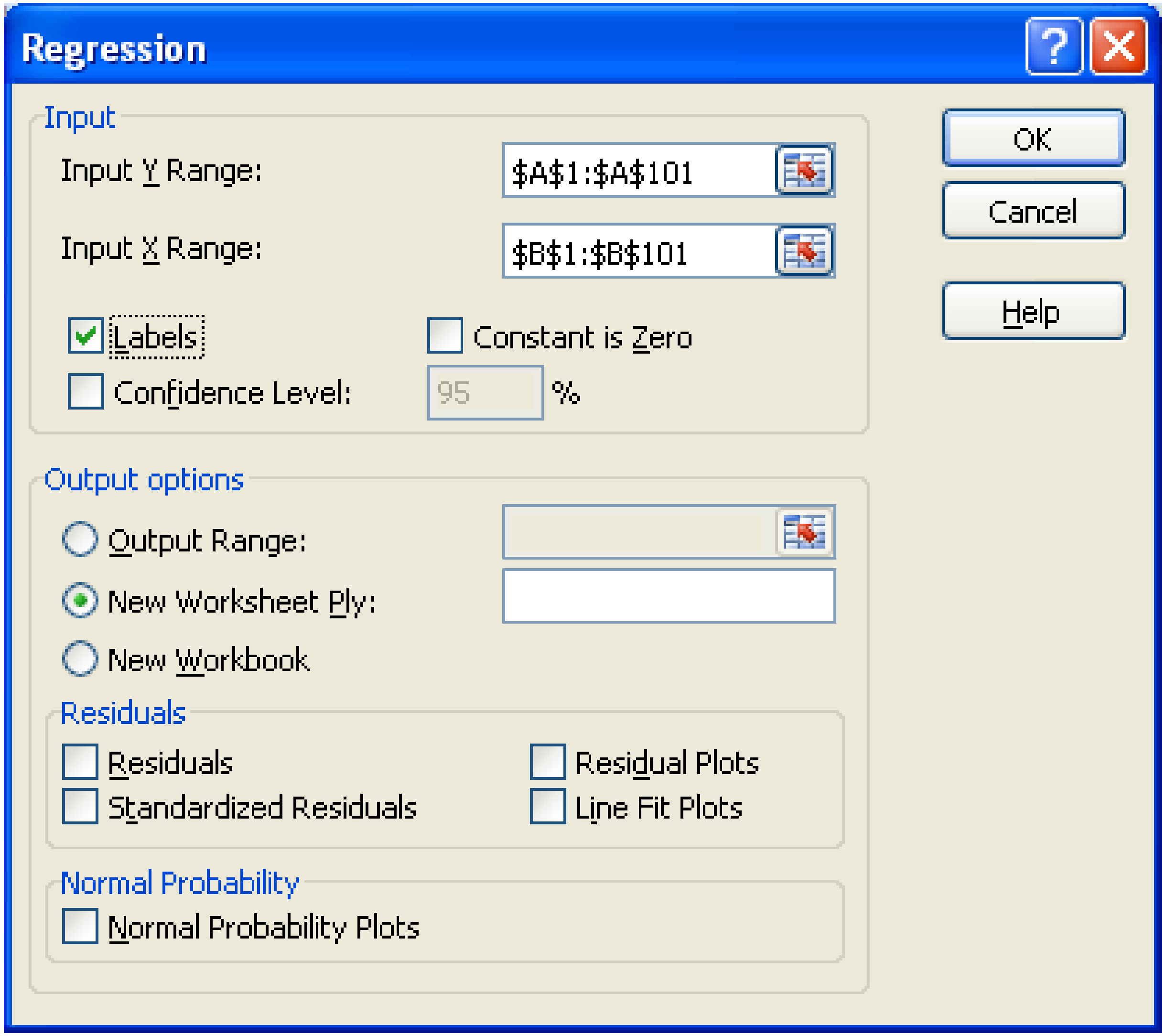
Για να εξετάσει αυτό το θέμα, ένας πωλητής επέλεξε τυχαία 100 αυτοκίνητα Toyota Camrys τριών ετών, τα οποία πωλήθηκαν σε δημοπρασίες τον τελευταίο μήνα.

Ο πωλητής κατέγραψε την τιμή πώλησης (σε χιλιάδες δολάρια) και τον αριθμό των μιλίων (σε χιλιάδες) του κοντέρ. (Xm16-02).

Ο πωλητής θέλει να υπολογίσει την ευθεία παλινδρόμησης.

Παράδειγμα 16.2…

Χρησιμοποιεί το Excel ….



Παράδειγμα 16.2…

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.8052				
5	R Square	0.6483				
6	Adjusted R Square	0.6447				
7	Standard Error	0.3265				
8	Observations	100				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	1	[illegible]	19.26	180.64	5.75E-24
13	Residual	98	[illegible].45	0.11		
14	Total	99	29.70			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	17.25	0.182	94.73	3.57E-98	
18	Odometer	-0.0669	0.0050	-13.44	5.75E-24	

Από όλα όσα έχει υπολογίσει μας ενδιαφέρουν μόνο αυτά

$$\hat{y} = b_0 + b_1 x = 17.250 - 0.0669\,x$$

Παράδειγμα 6.2…

ΕΡΜΗΝΕΙΑ

Όπως ήταν αναμενόμενο …

Ο συντελεστής κλίσης, b_1, είναι –0.0669, δηλαδή, κάθε επιπλέον μίλι στο κοντέρ μειώνει την τιμή κατά 0.0669 δολάρια ή 6.69 σεντς.

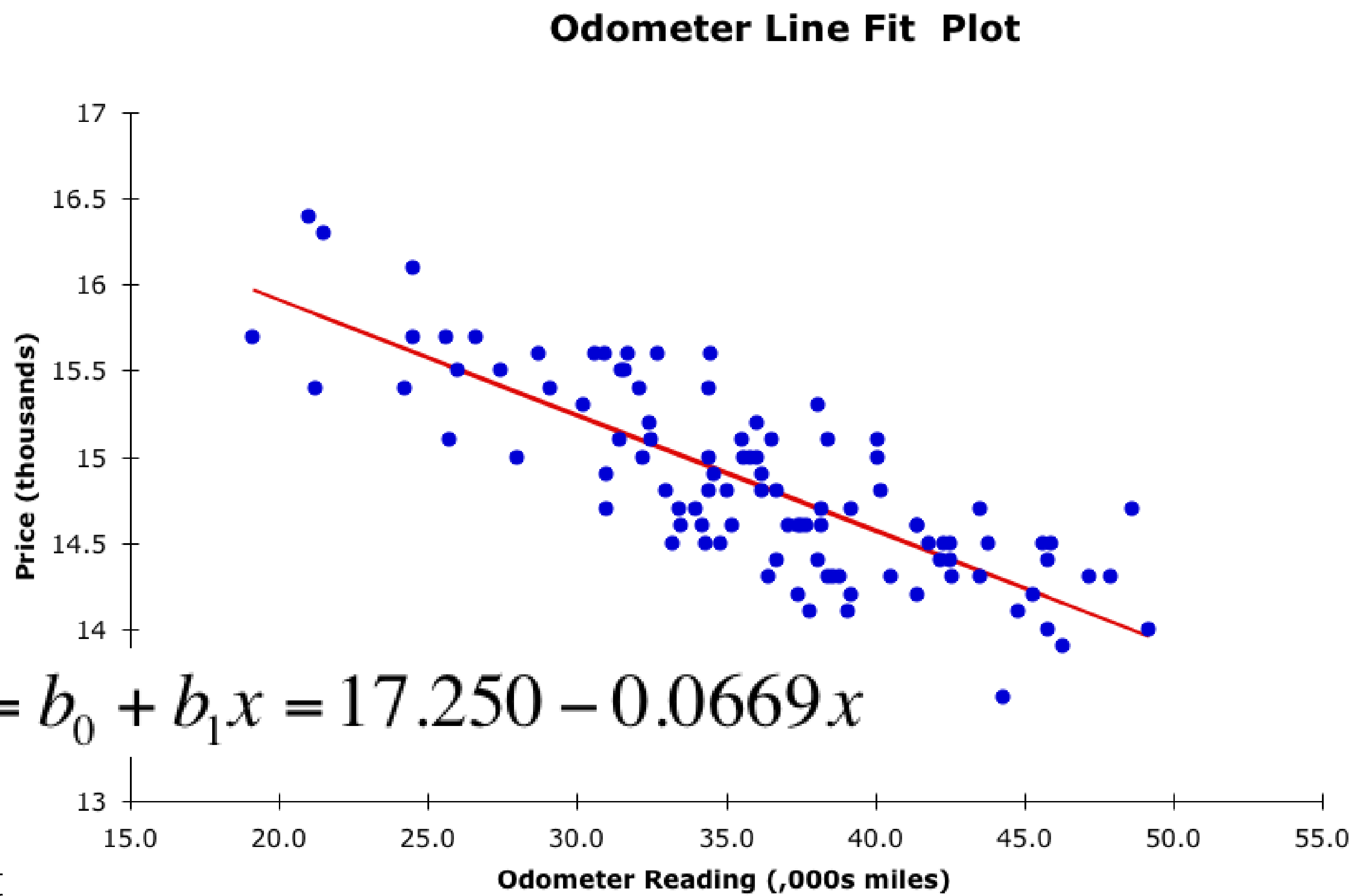
Το ίχνος στον y, b_0, είναι 17,250. Μια ερμηνεία είναι ότι όταν x = 0 (δεν έχει κινηθεί) η τιμή πώλησης είναι \$17,250. Ωστόσο, δεν έχουμε δεδομένα για αυτοκίνητα με λιγότερα από 19,100 μίλια, επομένως αυτή η εκτίμηση δεν είναι σωστή.

$$\hat{y} = b_0 + b_1 x = 17.250 - 0.0669x$$

Παράδειγμα 6.2…

ΕΡΜΗΝΕΙΑ

Κατασκευάζουμε το διάγραμμα των δεδομένων και την ευθεία παλινδρόμησης



$$\hat{y} = b_0 + b_1 x = 17.250 - 0.0669x$$

Απαιτούμενες Συνθήκες …

Για να ισχύουν τα προηγούμενα, πρέπει να πληρούνται τέσσερεις συνθήκες :

- Η κατανομή πιθανοτήτων του ε να είναι κανονική.
- Ο μέσος της κατανομής είναι 0, δηλαδή $E(\varepsilon) = 0$.
- Η τυπική απόκλιση του ε, σ_ε, είναι σταθερή για κάθε τιμή του x.
- Η τιμή του ε που αντιστοιχεί σε κάθε τιμή του y είναι α-νεξάρτητη του ε για κάθε άλλη τιμή του y.

Αξιολόγηση του μοντέλου …

Η μέθοδος ελαχίστων τετραγώνων δημιουργεί πάντοτε μια ευθεία, ακόμα κι αν δεν υπάρχει σχέση μεταξύ των μεταβλητών, ή κι αν η σχέση είναι μη γραμμική.

Επομένως, πέρα από τον καθορισμό των συντελεστών της ευθείας ελαχίστων τετραγώνων, πρέπει να την αξιολογήσουμε για να δούμε πόσο καλά “ταιριάζει” στα δεδομένα. Θα δούμε στη συνέχεια αυτές τις μεθόδους αξιολόγησης. Βασίζονται στο άθροισμα των τετραγώνων των σφαλμάτων (SSE).

Άθροισμα Τετραγώνων Σφάλματος (SSE)…

Το άθροισμα τετραγώνων σφάλματος υπολογίζεται ως:

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Και χρησιμοποιείται για τον υπολογισμό του ***τυπικού σφάλματος εκτίμησης***:

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}}$$

Αν s_ε είναι μηδέν, τότε όλα τα σημεία είναι πάνω στην ευθεία παλινδρόμησης.

Τυπικό Σφάλμα Εκτίμησης ...

	A	B	C	D		
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.8052				
5	R Square	0.6483				
6	Adjusted R Square	0.6447				
7	Standard Error (s_ε)	0.3265				
8	Observations	100				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	1	19.26	19.26	180.64	0.00
13	Residual	98	10.45	0.11		
14	Total	99	29.70			
15						
16		***Coefficients***	*Standard Error*	*t Stat*	*P-value*	
17	**Intercept**	**17.25**	0.182	94.73	0.0000	
18	**Odometer**	**-0.0669**	0.0050	-13.44	0.0000	
19						

Αν s_ε είναι μικρό, η προσαρμογή είναι εξαιρετική και το γραμμικό μοντέλο μπορεί να χρησιμοποιηθεί για πρόβλεψη. Αν s_ε είναι μεγάλο, το μοντέλο μας δεν είναι καλό…

Αλλά πότε είναι ***μικρό*** και πότε είναι ***μεγάλο*** ;

Τυπικό Σφάλμα Εκτίμησης …

Κρίνουμε την τιμή του s_ε συγκρίνοντάς το με το μέσο της εξαρτημένης μεταβλητής $\bar{y}$.

Στο παράδειγμά μας,

$s_\varepsilon = .3265$ και

$\bar{y} = 14.841$

άρα (μιλώντας σχετικά) φαίνεται να είναι "μικρό", άρα το μοντέλο γραμμικής παλινδρόμησης της τιμής πώλησης των αυτοκινήτων ως συνάρτηση της ένδειξης του κοντέρ είναι "καλό".

Έλεγχος της κλίσης …

Εάν δεν υπάρχει γραμμική σχέση μεταξύ δύο μεταβλητών, θα περιμέναμε η ευθεία παλινδρόμησης να είναι ***οριζόντια***, δηλαδή, να έχουμε ***μηδενική κλίση***.

Θέλουμε να δούμε εάν υπάρχει γραμμική σχέση, δηλαδή να δούμε εάν η κλίση (β_1) είναι διαφορετική από το μηδέν. Η υπόθεσή μας γίνεται:

$$H_1: \beta_1 \neq 0$$

Άρα η μηδενική υπόθεση είναι:

$$H_0: \beta_1 = 0$$

Έλεγχος της κλίσης …

Ο στατιστικός έλεγχος για την υπόθεσή μας:

$$\boxed{t = \frac{b_1 - \beta_1}{s_{b_1}}}$$

όπου s_{b_1} είναι η τυπική απόκλιση του b_1, ορισμένη ως:

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$$

Εάν το σφάλμα (ε) ακολουθεί κανονική κατανομή, ο έλεγχος ακολουθεί την **t**-κατανομή με n–2 βαθμούς ελευθερίας. Η περιοχή απόρριψης εξαρτάται από το εάν έχουμε μονόπλευρο ή αμφίπλευρο έλεγχο (συνήθως έχουμε αμφίπλευρο).

Παράδειγμα 6.4…

Να ελέγξετε εάν υπάρχει γραμμική σχέση μεταξύ της τιμής και της ένδειξης του κοντέρ. (5% επίπεδο σημαντικότητας)

Έχουμε:

$$H_1: \beta_1 \neq 0$$

$$H_0: \beta_1 = 0$$

(εάν η μηδενική υπόθεση ισχύει, δεν υπάρχει γραμμική σχέση)

Η περιοχή απόρριψης είναι:

$$t < -t_{\alpha/2,\nu} = -t_{.025,98} \approx -1.984 \quad or \quad t > t_{\alpha/2,\nu} = t_{.025,98} \approx 1.984$$

Παράδειγμα 16.4...

ΥΠΟΛΟΓΙΣΜΟΣ

Μπορούμε να υπολογίσουμε το **t** με το χέρι ή με το Excel ...

$$t < -t_{\alpha/2,\nu} = -t_{.025,98} \approx -1.984 \quad or \quad t > t_{\alpha/2,\nu} = t_{.025,98} \approx 1.984$$

	Coefficients	Standard Error	t Stat	P-value
Intercept	17.25	0.182	94.73	0.0000
Odometer	-0.0669	0.0050	-13.44	0.0000

p-τιμή

Συγκρίνουμε

Βλέπουμε ότι το **t** για το "κοντέρ" (δηλ. την κλίση b_1) είναι –13.44 που είναι μεγαλύτερο από το $\mathbf{t}_{\text{Critical}} = -1.984$. Παρατηρούμε ότι η p-τιμή is 0.000.

Υπάρχουν πολλά στοιχεία που οδηγούν στο ότι υπάρχει γραμμική σχέση μεταξύ της τιμής και της ένδειξης του κοντέρ.

Έλεγχος της κλίσης …

Εάν θέλουμε να ελέγξουμε για ***θετική*** ή ***αρνητική*** γραμμική σχέση κάνουμε μονόπλευρους ελέγχους, δηλαδή οι υποθέσεις μας είναι:

H_1: $\beta_1 < 0$ (έλεγχος για αρνητική κλίση)

ή

H_1: $\beta_1 > 0$ (έλεγχος για θετική κλίση)

Φυσικά, η μηδενική υπόθεση παραμένει: H_0: $\beta_1 = 0$.

Συντελεστής Προσδιορισμού …

Μέχρι τώρα οι έλεγχοι δείχνουν εάν ***υπάρχει*** μια γραμμική σχέση. Είναι χρήσιμο να μετρήσουμε και το ***πόσο ισχυρή είναι αυτή η σχέση***. Αυτό γίνεται υπολογίζοντας τον ***συντελεστή προσδιορισμού*** R^2.

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} \quad \text{ή} \quad R^2 = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2}$$

Ο συντελεστής προσδιορισμού είναι το τετράγωνο του συντελεστή συσχέτισης (**r**), συνεπώς $R^2 = (r)^2$

Συντελεστής Προσδιορισμού …

Όπως είδαμε στην ανάλυση διασποράς, μπορούμε να χωρίσουμε την μεταβλητότητα του y σε δύο μέρη:

Μεταβλητότητα του y = SSE + SSR

SSE – Sum of Squares Error – μέτρο της μεταβλητότητας του y που παραμένει ανεξήγητη (Άθροισμα Τετραγώνων Σφάλματος)

SSR – Sum of Squares Regression – μέτρο της μεταβλητότητας του y που εξηγείται από τη μεταβλητότητα της ***ανεξάρτητης μεταβλητής*** x. (Άθροισμα Τετραγώνων Παλινδρόμησης)

Συντελεστής Προσδιορισμού **ΥΠΟΛΟΓΙΣΜΟΣ**

Το υπολογίζουμε με το χέρι ή με το Excel…

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{100-1}\left[53{,}155.9 - \frac{(3{,}601.1)(1{,}484.1)}{100}\right] = -2.909$$

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = \frac{(-2.909)^2}{(43.509)(.3000)} = .6483$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{100-1}\left[133{,}986.59 - \frac{(3{,}601.1)^2}{100}\right] = 43.509$$

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] = \frac{1}{100-1}\left[22{,}055.23 - \frac{(1{,}484.1)^2}{100}\right] = .3000$$

	A	B
1	SUMMARY OUTPUT	
2		
3	*Regression Statistics*	
4	Multiple R	0.8052
5	R Square	0.6483
6	Adjusted R Square	0.6447
7	Standard Error	0.3265
8	Observations	100

Συντελεστής Προσδιορισμού

ΕΡΜΗΝΕΙΑ

R^2 έχει τιμή .6483. Άρα το 64.83% της μεταβλητότητας των τιμών πώλησης (y) ερμηνεύεται από τη μεταβλητότητα των ενδείξεων του κοντέρ (x). Το υπόλοιπο 35.17% είναι ***ανεξήγητο***, δηλαδή οφείλεται σε σφάλμα.

Αντίθετα από τους ελέγχους, ο ***συντελεστής προσδιορισμού*** **δεν** έχει ***κρίσιμη τιμή*** η οποία να μας επιτρέπει να βγάλουμε συμπεράσματα.

Γενικά, όσο μεγαλύτερη είναι η τιμή του R^2, τόσο ***καλύτερα*** το μοντέλο προσαρμόζεται στα δεδομένα.

$R^2 = 1$: Απόλυτη ταύτιση της ευθείας και των δεδομένων.

$R^2 = 0$: Δεν υπάρχει γραμμική σχέση μεταξύ x και y.

Συντελεστής Συσχέτισης

Μπορούμε να χρησιμοποιήσουμε τον ***συντελεστή συσχέτισης*** για να ελέγξουμε την ύπαρξη γραμμικής σχέσης μεταξύ δύο μεταβλητών.

Θυμίζουμε:

Το εύρος του συντελεστή συσχέτισης είναι

μεταξύ –1 and +1.

• Αν r = –1 (αρνητική συσχέτιση) ή r = +1 (θετική συσχέ-τιση) κάθε σημείο είναι πάνω στην ευθεία παλινδρόμησης.

• Αν r = 0 δεν υπάρχει γραμμικό μοτίβο

Συντελεστής Συσχέτισης

Ο συντελεστής συσχέτισης του ***πληθυσμού*** συμβολίζεται με ρ (rho)

Εκτιμάμε την τιμή του από τα δεδομένα με τον ***συντελεστή συσχέτισης του δείγματος***:

$$\boxed{r = \frac{s_{xy}}{s_x s_y}}$$

Ό στατιστικός έλεγχος για $\rho = 0$ είναι:

$$t = r\sqrt{\frac{n-2}{1-r^2}}, \quad \nu = n-2$$

δηλαδή **t**-κατανομή με n–2 βαθμούς ελευθερίας.

Παράδειγμα 16.6…

Μπορούμε να διεξάγουμε **t**-έλεγχο του ***συντελεστή συσχέτισης*** για να καθορίσουμε με διαφορετικό τρόπο εάν η τιμή πώλησης και η ένδειξη του κοντέρ είναι ***γραμμικά εξαρτημένες***.

Η υπόθεσή μας είναι:

H_1: $\rho \neq 0$

(δηλ. υπάρχει γραμμική σχέση) και η μηδενική υπόθεση είναι:

H_0: $\rho = 0$

(δηλ. δεν υπάρχει γραμμική σχέση)

Παράδειγμα 16.6…

ΥΠΟΛΟΓΙΣΜΟΣ

Έχουμε ήδη αποδείξει ότι:

$$s_{xy} = -2.909 \qquad \begin{aligned} s_x^2 &= 43.509 \;\Rightarrow\; s_x = \sqrt{43.509} = 6.596 \\ s_y^2 &= .3000 \;\Rightarrow\; s_y = \sqrt{.3000} = .5477 \end{aligned}$$

Επομένως ο συντελεστής συσχέτισης είναι:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{-2.909}{(6.596)(.5477)} = -.8052$$

και η τιμή για τον έλεγχο γίνεται:

$$t = r\sqrt{\frac{n-2}{1-r^2}} = -.8052\sqrt{\frac{100-2}{1-(-.8052)^2}} = -13.44$$

Παράδειγμα 16.6...

ΥΠΟΛΟΓΙΣΜΟΣ

Μπορούμε να χρησιμοποιήσουμε και το Excel παίρνοντας το output:

	A	B
1	**Correlation**	
2		
3	*Price and Odometer*	
4	Pearson Coefficient of Correlation	-0.8052
5	t Stat	-13.44
6	df	98
7	P(T<=t) one tail	0
8	t Critical one tail	1.6606
9	P(T<=t) two tail	0
10	t Critical two tail	1.9845

p-τιμή

συγκρίνουμε

Μπορούμε επίσης να κάνουμε μονόπλευρο έλεγχο για θετική ή αρνητική γραμμική σχέση

Πάλι, απορρίπτουμε τη μηδενική υπόθεση (ότι δεν υπάρχει γραμμική συσχέτιση) και αποδεχόμαστε την εναλλακτική (ότι οι μεταβλητές μας συνδέονται με γραμμικό τρόπο).

Χρήση της εξίσωσης Παλινδρόμησης …

Θα μπορούσαμε να χρησιμοποιήσουμε την εξίσωση παλινδρόμησης:

$$\hat{y} = 17.250 - .0669x$$

για να προβλέψουμε την τιμή πώλησης ενός αυτοκινήτου με 40 (,000) μίλια:

$$\hat{y} = 17.250 - .0669x = 17.250 - .0669(40) = 14,574$$

Καλούμε αυτήν την τιμή (\$14,574) ***σημειακή πρόβλεψη***. Μάλλον όμως η πραγματική τιμή πώλησης θα είναι διαφορετική …