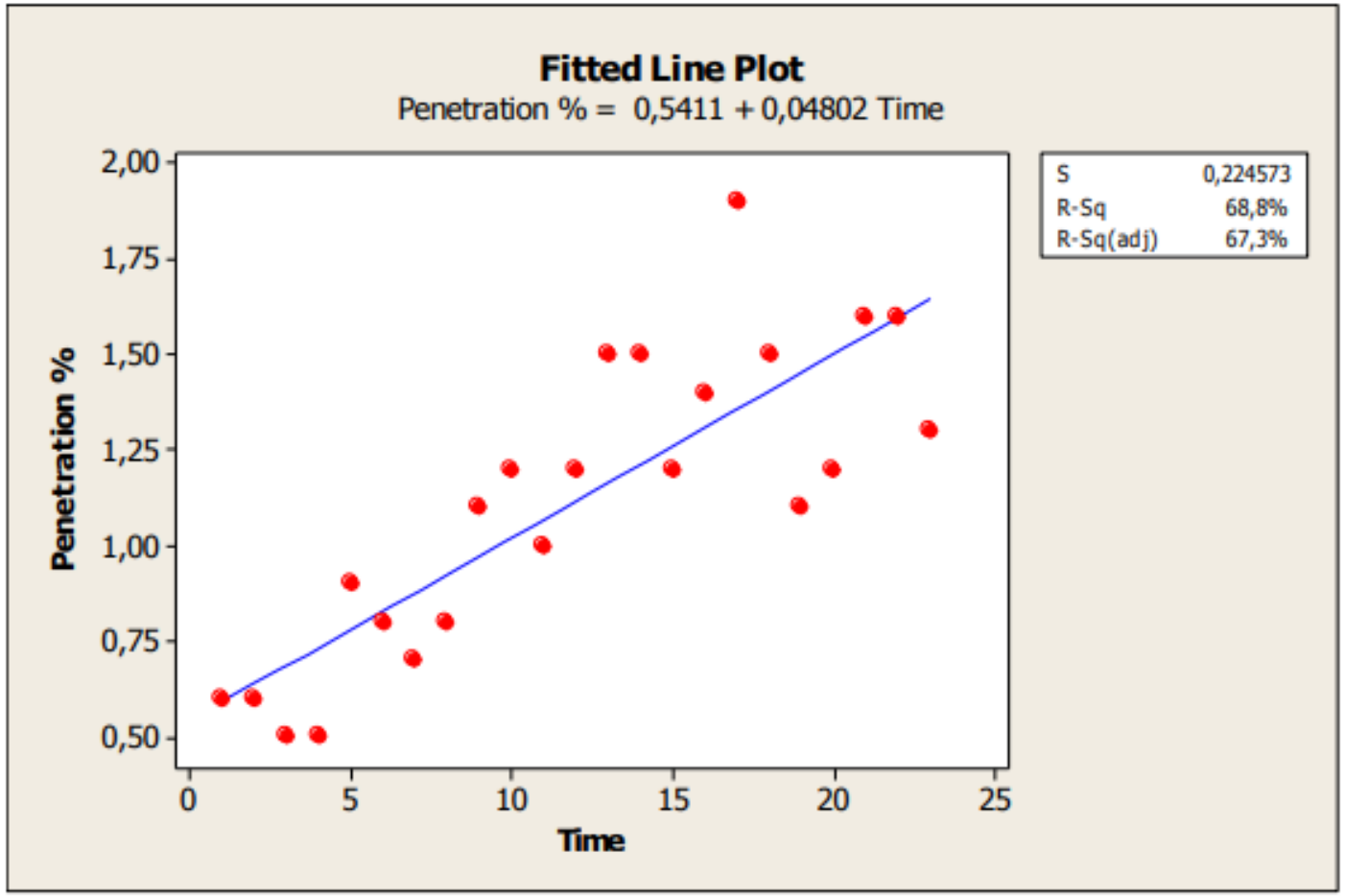
National and Kapodistrian
UNIVERSITY OF ATHENS

Τεχνικές Ανάλυσης και Πρόβλεψης Τηλεπικοινωνιακών Αγορών

Απλή Γραμμική Παλινδρόμηση

Απλή Γραμμική Παλινδρόμηση



Ευθεία Παλινδρόμησης

- Η ευθεία που προσαρμόζεται καλύτερα σε μία συλλογή X–Y σημείων είναι η γραμμή αυτή, που ελαχιστοποιεί το άθροισμα των τετραγωνικών κάθετων αποστάσεων της γραμμής από τα προς μελέτη σημεία.
- Αυτή η γραμμή είναι γνωστή ως γραμμή ελαχίστων τετραγώνων ή προσαρμοσμένη γραμμή παλινδρόμησης (fitted regression line) και η εξίσωσή της ονομάζεται προσαρμοσμένη εξίσωση παλινδρόμησης (fitted regression equation)

Ευθεία Παλινδρόμησης

- Η προσαρμοσμένη ευθεία γραμμή είναι της μορφής:

$$\hat{Y} = b_0 + b_1 X$$

- Ο πρώτος όρος, είναι η τομή της ευθείας με τον άξονα Υ (Y-intercept) και ο δεύτερος όρος, είναι η κλίση (slope).
- Η κλίση αναπαριστά, το πόσο αλλάζει το Υ, όταν το Χ αυξάνει κατά μία μονάδα.
- Το άμεσο αντικείμενό μας είναι να καθοριστούν τιμές για τα b_0 και b_1.

Ευθεία Παλινδρόμησης

- Η μέθοδος των ελαχίστων τετραγώνων επιλέγει τις τιμές για τα b_0 και b_1, οι οποίες ελαχιστοποιούν το άθροισμα των τετραγωνικών σφαλμάτων (αποστάσεις).

$$SSE = \sum (Y - \hat{Y})^2 = \sum (Y - b_0 - b_1 X)^2$$

$$b_0 = \frac{\sum Y \sum X^2 - \sum X \sum XY}{n \sum X^2 - (\sum X)^2}$$

$$b_1 = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

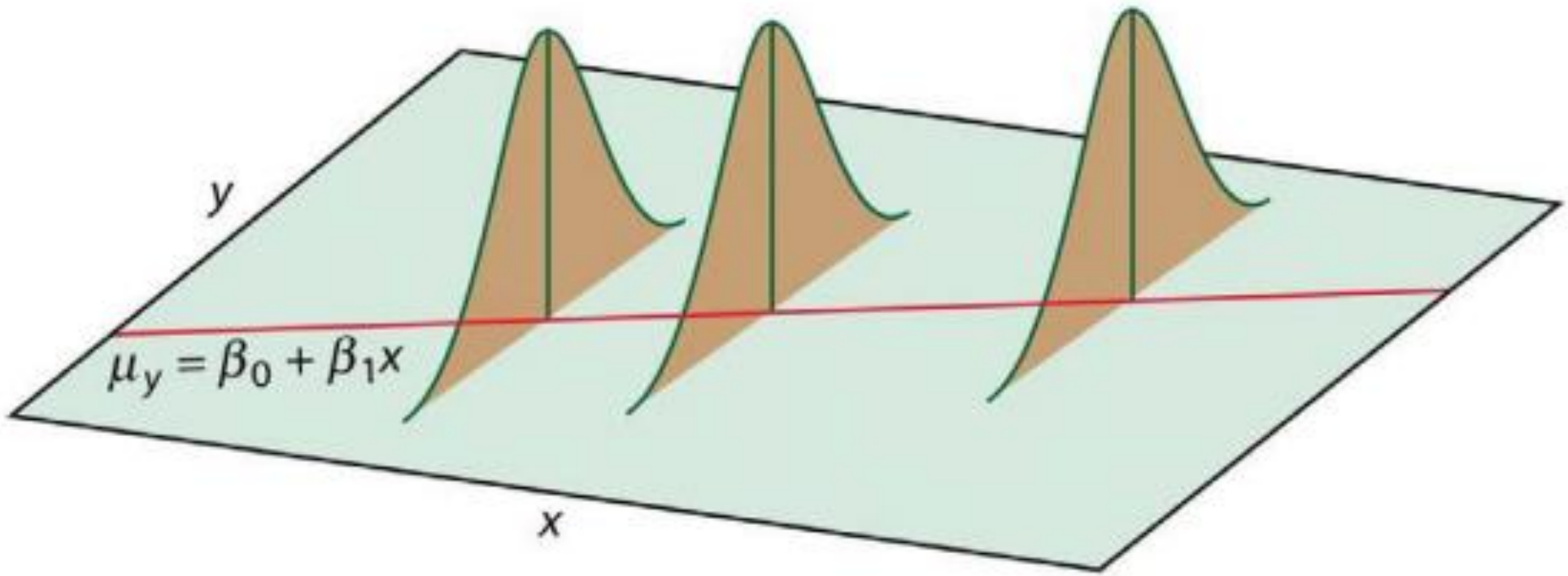
Στατιστικό Μοντέλο για Παλινδρόμηση Ευθείας Γραμμής

- Η εξαρτημένη μεταβλητή Y σχετίζεται με την ανεξάρτητη μεταβλητή X, με τη σχέση:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Το $\beta_0 + \beta_1 X$ είναι η μέση τιμή για ένα δεδομένο X.
- Οι αποκλίσεις (ε) υποθέτουμε, ότι είναι ανεξάρτητες, τυχαίες και κανονικά κατανεμημένες, με μέσο 0 ($E(\varepsilon)=0$) και τυπική απόκλιση σ ($E(\varepsilon^2)=\sigma^2$), δηλαδή $\varepsilon=N(0, \sigma^2)$. Οι άγνωστες σταθερές είναι οι β_0, β_1 και σ.

Στατιστικό Μοντέλο για Παλινδρόμηση Ευθείας Γραμμής



Τυπικό Σφάλμα Εκτίμησης

- Το τυπικό σφάλμα εκτίμησης (standard error of the estimate) υπολογίζει την κατανομή των σημείων δεδομένων γύρω από την προσαρμοσμένη γραμμή στην κατεύθυνση Y.

- Αποτελεί ένα μέτρο για τη διασπορά (dispersion) ανάλογο της τυπικής απόκλισης του δείγματος.

Τυπικό Σφάλμα Εκτίμησης

› Το τυπικό σφάλμα εκτίμησης υπολογίζει την ποσότητα, κατά την οποία οι πραγματικές τιμές του Υ διαφέρουν από τις εκτιμημένες τιμές.

$$s_{y \cdot x} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum XY}{n-2}}$$

Τυπικό Σφάλμα Εκτίμησης

- Για σχετικά μεγάλα δείγματα, περιμένουμε περίπου το 67% των διαφορών $Y-Y^{\wedge}$ να είναι σε ένα διάστημα $\pm s_{y \cdot x}$ από το 0 και περίπου το 95% αυτών των διαφορών να είναι σε ένα διάστημα $2(\pm s_{y \cdot x})$ από το 0.
- Μία ανάλυση παλινδρόμησης με ένα μικρό τυπικό σφάλμα εκτίμησης σημαίνει, ότι όλα τα σημεία δεδομένων βρίσκονται πολύ κοντά στην προσαρμοσμένη γραμμή παλινδρόμησης. Αν το τυπικό σφάλμα εκτίμησης είναι μεγάλο, τα σημεία δεδομένων είναι ευρέως διασκορπισμένα γύρω από την προσαρμοσμένη γραμμή.

Πρόβλεψη του Υ

- Η προσαρμοσμένη γραμμή παλινδρόμησης μπορεί να χρησιμοποιηθεί για να εκτιμήσει την τιμή του Υ για μία δεδομένη τιμή του Χ.

- Για να βρούμε ένα σημείο πρόβλεψης (point forecast), ή για να προβλέψουμε το Υ για μία δεδομένη τιμή του Χ, απλά υπολογίζουμε την εκτιμώμενη συνάρτηση παλινδρόμησης στο Χ.

Πρόβλεψη του Υ

- Η (προσαρμοσμένη) γραμμή παλινδρόμησης του δείγματος είναι μία εκτίμηση της γραμμής παλινδρόμησης του πληθυσμού (population regression line), που βασίζεται στα σημεία δεδομένων.
- Άλλα τυχαία ζεύγη τιμών θα έδειχναν διαφορετική προσαρμοσμένη γραμμή παλινδρόμησης
- Για ν δείγματα ν διαφορετικές γραμμές παλινδρόμησης, όπως και στην περίπτωση, στην οποία πολλά διαφορετικά δείγματα από τον ίδιο πληθυσμό έχουν διαφορετικούς μέσους δείγματος.

Πρόβλεψη του Υ

- Υπάρχουν δύο πηγές αβεβαιότητας, που σχετίζονται με ένα σημείο πρόβλεψης, το οποίο παράγεται από μία προσαρμοσμένη εξίσωση παλινδρόμησης:
  - Αβεβαιότητα λόγω της διασποράς των σημείων δεδομένων γύρω από τη γραμμή παλινδρόμησης του δείγματος.
  - Αβεβαιότητα λόγω της διασποράς της γραμμής παλινδρόμησης του δείγματος γύρω από τη γραμμή παλινδρόμησης του πληθυσμού.
- Μπορεί να κατασκευαστεί ένα διάστημα πρόβλεψης (interval forecast) του Υ, λαμβάνοντας υπόψη τις παραπάνω δύο αβεβαιότητες.

Τυπικό Σφάλμα Πρόβλεψης

› Το τυπικό σφάλμα της πρόβλεψης (standard error of the forecast) s_f υπολογίζει τη μεταβλητότητα της προβλεπόμενης τιμής Υ, από την γραμμική σχέση, γύρω από την πραγματική τιμή του Υ για μία τιμή Χ.

$$s_f = \sqrt{s_{y \cdot x}^2 + s_{y \cdot x}^2 \left(\frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X - \bar{X})^2} \right)} \quad \text{ή}$$

$$s_f = s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X - \bar{X})^2}}$$

Τυπικό Σφάλμα Πρόβλεψης

- Ο πρώτος όρος κάτω από την ρίζα, $s_{y \cdot x}^2$, υπολογίζει τη διασπορά των σημείων δεδομένων γύρω από τη γραμμή παλινδρόμησης του δείγματος (πρώτη πηγή αβεβαιότητας).

- Ο δεύτερος όρος κάτω από τη ρίζα υπολογίζει τη διασπορά της γραμμής παλινδρόμησης του δείγματος γύρω από της γραμμή παλινδρόμησης του πληθυσμού (δεύτερη πηγή αβεβαιότητας).

Τυπικό Σφάλμα Πρόβλεψης

- Το τυπικό σφάλμα της πρόβλεψης εξαρτάται από το X, την τιμή του X για την οποία επιθυμούμε μία πρόβλεψη του Y.
- Το s_f έχει ελάχιστη τιμή όταν $X=\bar{X}$, δεδομένου ότι ο αριθμητής στον τρίτο όρο κάτω από τη ρίζα της Εξίσωσης (κάτω εξίσωση) θα είναι $(\bar{X}-\bar{X})^2=0$.
- Αν κρατήσουμε όλα τα άλλα σταθερά, όσο πιο μακριά είναι το X από το $\bar{X}$, τόσο πιο μεγάλο είναι το τυπικό σφάλμα της πρόβλεψης

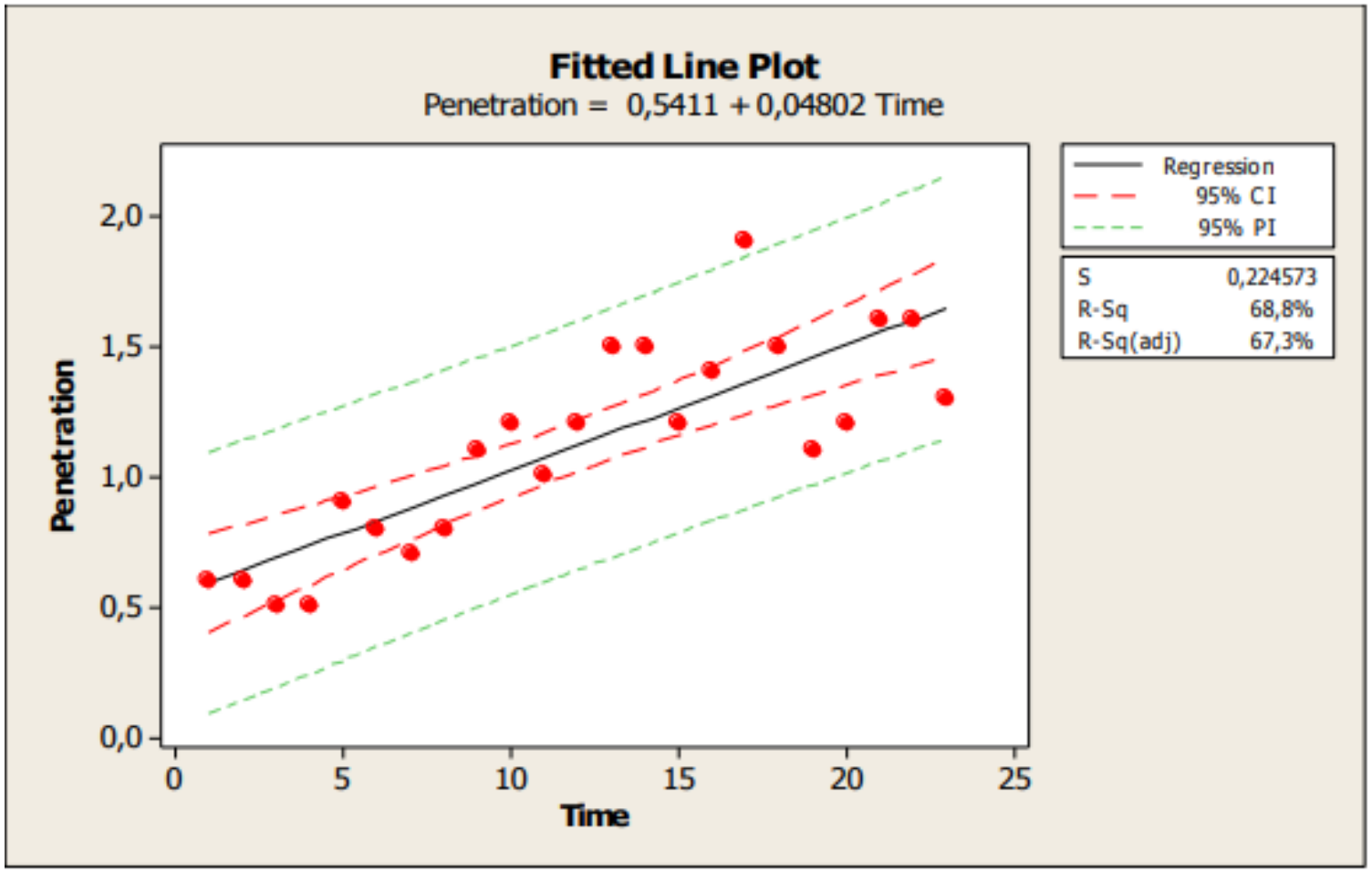
Διάστημα Πρόβλεψης

- Αν το στατιστικό μοντέλο της απλής γραμμικής παλινδρόμησης είναι κατάλληλο, ένα διάστημα πρόβλεψης για το Y δίνεται από τη σχέση:

$$\hat{Y} \pm ts_f$$

- όπου το t είναι ένα εκατοστιαίο σημείο της t κατανομής με df=n-2.
- Αν το μέγεθος του δείγματος είναι μεγάλο (n-2≥30), το t εκατοστιαίο σημείο μπορεί να αντικατασταθεί με το αντίστοιχο εκατοστιαίο σημείο Z της τυπικής κανονικής κατανομής.

Διάστημα Πρόβλεψης



Ανάλυση της Διασποράς

› Από τις παρατηρούμενες τιμές έχουμε:

$$Y = \hat{Y} + (Y - \hat{Y})$$

ή

$$Y = b_0 + b_1 X + (Y - b_0 - b_1 X)$$

› όπου Υ είναι η παρατηρούμενη τιμή, ο όρος εκτός παρένθεσης ερμηνεύεται από τη γραμμική σχέση και ο όρος εντός παρένθεσης είναι το κατάλοιπο (residual), ή η απόκλιση από τη γραμμική σχέση.

Ανάλυση της Διασποράς

$$\sum (Y-\bar{Y})^2=\sum (\hat{Y}-\bar{Y})^2+\sum (Y-\hat{Y})^2$$

ή SST = SSR + SSE ή

Ολική μεταβλητότητα του Y = Εξηγήσιμη μεταβλητότητα από την παλινδρόμηση (γραμμική σχέση) + τα κατάλοιπα.

όπου:

$$SST=\sum (Y-\bar{Y})^2$$

$$SSR=\sum (\hat{Y}-\bar{Y})^2$$

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

SS = Sum of squares και T, R, E σημαίνουν total, regression, error.

Ανάλυση της Διασποράς

› Διαιρώντας με n-1 και τροποποιώντας τη σχέση των αθροισμάτων των τετραγώνων των διαφορών παίρνουμε:

$$\frac{\sum (Y-\bar{Y})^2}{n-1}=\frac{\sum (\hat{Y}-\bar{Y})^2}{n-1}+\frac{(n-2)\sum (Y-\hat{Y})^2}{(n-1)(n-2)} \quad \text{ή}$$

$$s_y^2=\frac{SSR}{n-1}+\frac{(n-2)}{(n-1)}s_{y\cdot x}^2$$

Ανάλυση της Διασποράς

› Τα αθροίσματα τετραγώνων, που σχετίζονται με την ανάλυση της μεταβλητότητας του Y και τους αντίστοιχους βαθμούς ελευθερίας, φαίνονται στον Πίνακα, ο οποίος είναι γνωστός ως ANOVA πίνακας, ή πίνακας ανάλυσης της διασποράς (ANalysis Of VAriance table).

Source	Sum of Squares	Df	Mean Square
Regression	SSR	1	MSR=SSR/1
Error	SSE	n – 2	MSE=SSE/(n-2)
Total	SST	n – 1	

Το μέσο τετραγωνικό σφάλμα είναι:

$$MSE = \frac{SSE}{n-2} = \frac{\sum (Y - \hat{Y})^2}{n-2} = s^2_{y \cdot x}$$

Ανάλυση της Διασποράς

› Έστω ότι ισχύει:

$$SST = \sum (Y - \bar{Y})^2 = 3,3932$$

$$SSE = \sum (Y - \hat{Y})^2 = 1,059$$

$$SSR = \sum (\hat{Y} - \bar{Y})^2 = SST - SSE = 3,3932 - 1,059 = 2,3342$$

› Η ανάλυση της μεταβλητότητας είναι:

SST	=	**SSR**	+	**SSE**
3,3932	=	**2,3342**	+	**1,059**
Συνολική Μεταβολή		**Εξηγήσιμη Μεταβολή**		**Μη εξηγήσιμη Μεταβολή**

Ανάλυση της Διασποράς

- Από τη μεταβλητότητα που παραμένει μετά την πρόβλεψη ένα ποσοστό αυτής:

$$\frac{SSR}{SST}=\frac{2,3342}{3,3932}=0,69$$

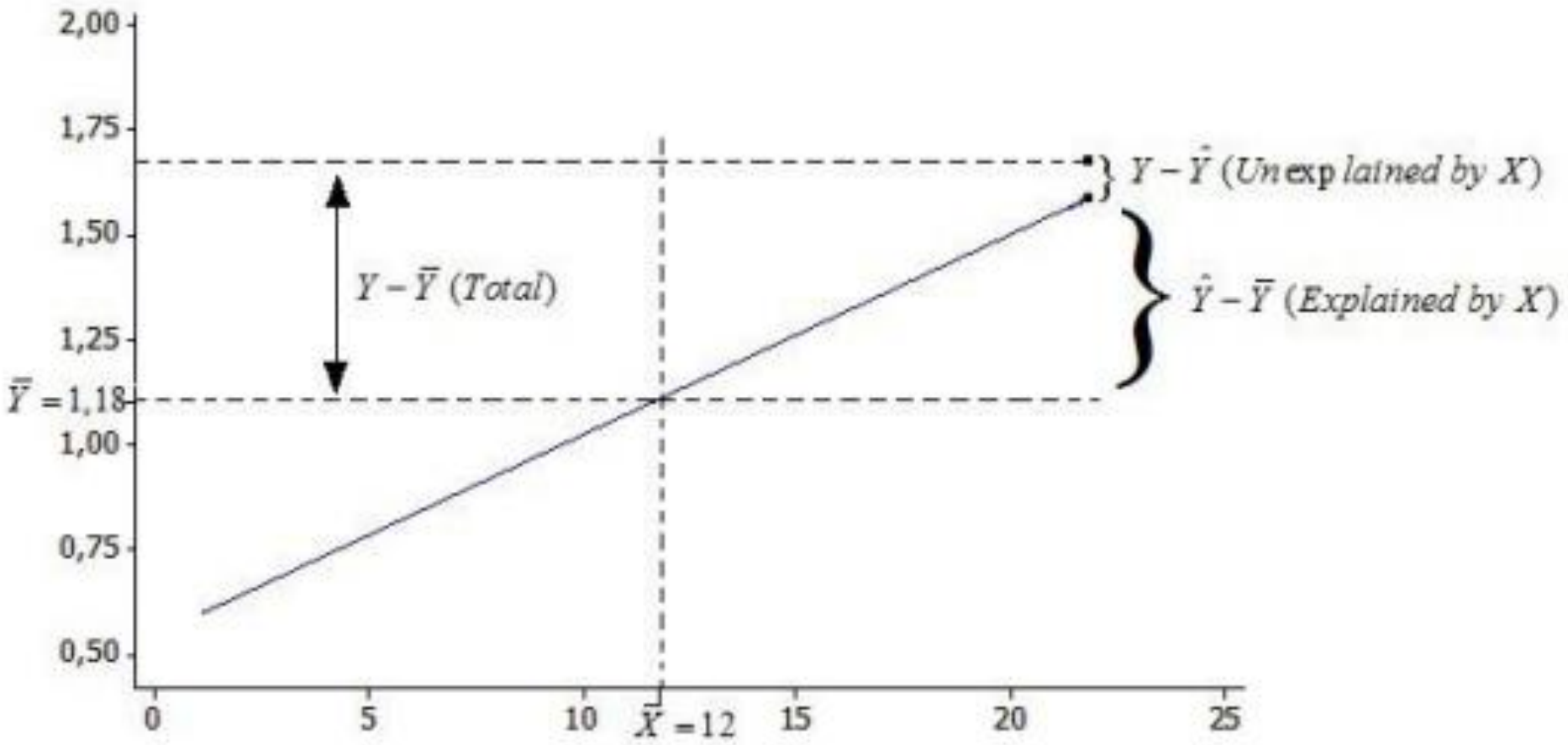
- έχει ερμηνευθεί από τη σχέση του Y με το X. Ένα ποσοστό, 1-0,69=0,31, της μεταβλητότητας του Y γύρω από το $\overline{Y}$ παραμένει μη ερμηνεύσιμο.
- Από αυτή την οπτική γωνία, η γνώση της σχετικής μεταβλητής X οδηγεί σε καλύτερες προβλέψεις του Y, από αυτές που μπορούν να γίνουν από το $\overline{Y}$ μία ποσότητα που δεν εξαρτάται από το X.

Ανάλυση της Διασποράς

› Ο αντίστοιχος πίνακας ANOVA είναι:

Πηγή	Άθροισμα Τετραγώνων	Df	Τετράγωνο Μέσου
Παλινδρόμηση	2,33	1	2,33
Σφάλμα	1,06	21	0,05
Σύνολο	3,39	22	

Συντελεστής Προσδιορισμού



Συντελεστής Προσδιορισμού

- Το SST υπολογίζει την συνολική διασπορά γύρω από το $\overline{Y}$
- Το μέρος του συνόλου που ερμηνεύεται από την κίνηση στο X είναι το SSR.
- Η υπόλοιπη ή μη ερμηνευμένη διασπορά είναι το SSE.

Συντελεστής Προσδιορισμού

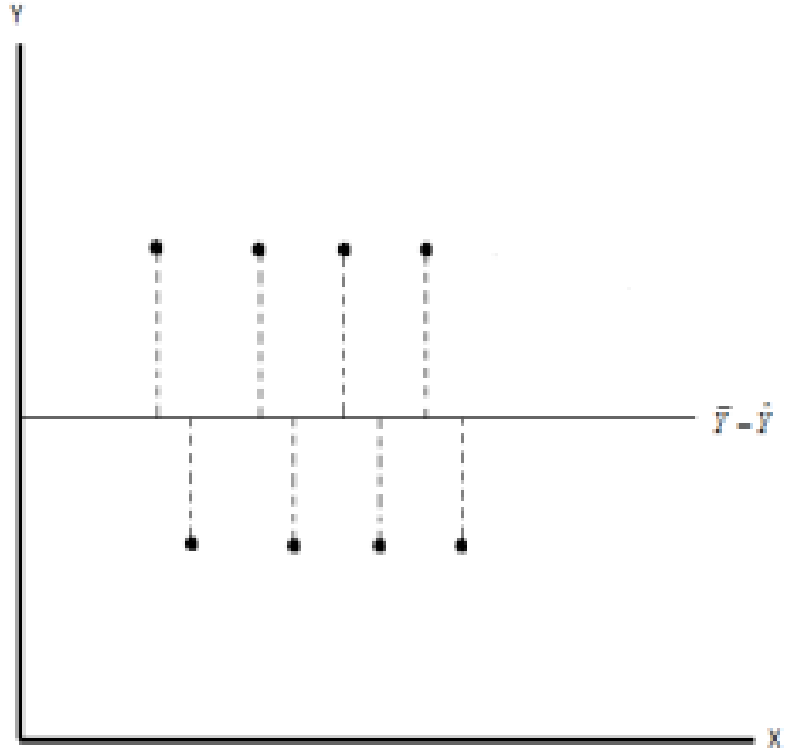
› Ο λόγος της ερμηνευμένης προς τη συνολική μεταβολή καλείται συντελεστής προσδιορισμού του δείγματος και συμβολίζεται με r^2.

$$r^2 = \frac{Explained\ Variation}{Total\ Variation} = \frac{SSR}{SST} = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2}$$

$$r^2 = 1 - \frac{Unexplained\ Variation}{Total\ Variation} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2}$$

› Ο συντελεστής προσδιορισμού υπολογίζει το ποσοστό της μεταβλητότητας στο Υ, το οποίο μπορεί να ερμηνευθεί μέσω της γνώσης της μεταβλητότητας (διαφορές) στην ανεξάρτητη μεταβλητή Χ.

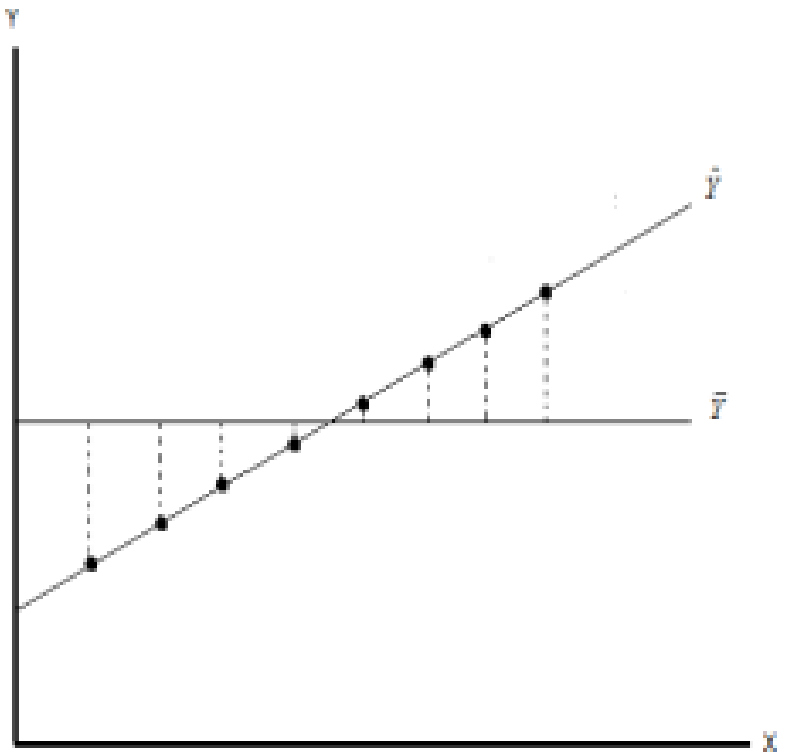
Συντελεστής Προσδιορισμού



$r^2 = 1 - \sum(Y - \hat{Y})^2 / \sum(Y - \bar{Y})^2$

$= 1 - 1 = 0$

(a) No Linear Correlation



$r^2 = 1 - \sum(Y - \hat{Y})^2 / \sum(Y - \bar{Y})^2$

$= 1 - 0 = 1$

(b) Perfect Linear Correlation

Προσαρμοσμένος Συντελεστής Προσδιορισμού

› Ο προσαρμοσμένος συντελεστής προσδιορισμού, προσαρμοσμένος με τους βαθμούς ελευθερίας:

$$\bar{r}^2 = r^2(adj) = 1 - \frac{SSE/(n-2)}{SST/(n-1)} = 1 - \frac{\sum (Y-\hat{Y})^2/(n-2)}{\sum (Y-\bar{Y})^2/(n-1)} = 1 - \frac{s^2_{y\cdot x}}{s^2_y}$$

Έλεγχος Υποθέσεων

- Το στατιστικό μοντέλο για την απλή γραμμική παλινδρόμηση προϋποθέτει, ότι η γραμμική σχέση μεταξύ του Υ και του Χ ισχύει για όλες τις επιλογές των Χ-Υ ζευγών του πληθυσμού. Αυτό γίνεται, γιατί υπάρχει – αν υπάρχει – μία αληθής σχέση μεταξύ του Χ και του Υ της μορφής $\mu_y=\beta_0+\beta_1 X$.
- Δεδομένης της ένδειξης του δείγματος, μπορούμε εμείς να συμπεράνουμε, ότι η αληθινή αυτή σχέση ισχύει για όλα τα Χ και Υ;

Έλεγχος Υποθέσεων

- Λαμβάνοντας υπόψη την παρακάτω υπόθεση ως μηδενική υπόθεση:

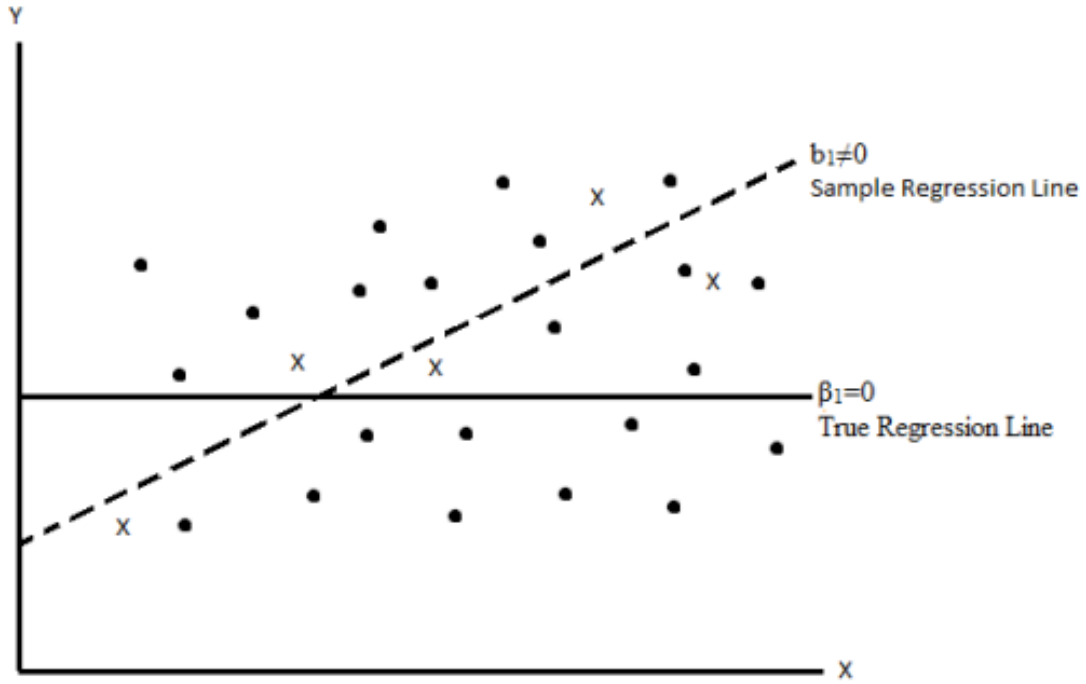
$$H_0: \beta_1 = 0,$$

όπου το β_1 είναι η κλίση της γραμμής παλινδρόμησης του πληθυσμού

- Αν αυτή η υπόθεση είναι αληθής, δεν υπάρχει σχέση μεταξύ του Y και του X στον πληθυσμό.
- Αν το H_0 δεν απορριφθεί σημαίνει, ότι παρά το γεγονός, ότι το δείγμα έχει παράξει μία προσαρμοσμένη γραμμή με μία μη μηδενική τιμή για το b_1, εμείς πρέπει να συμπεράνουμε, ότι δεν υπάρχουν επαρκείς ενδείξεις, που να υποδεικνύουν, ότι το Y σχετίζεται με το X. Αυτό συμβαίνει γιατί δεν μπορούμε να αποκλείσουμε την πιθανότητα, η γραμμή παλινδρόμησης του πληθυσμού να είναι οριζόντια.

Έλεγχος Υποθέσεων

Μία οριζόντια γραμμή παλινδρόμησης ($\beta_1=0$) είναι ισοδύναμη με την δήλωση $H_0: r=0$, όπου r είναι ο συντελεστής συσχέτισης του πληθυσμού.



Έλεγχος Υποθέσεων

› Αν η μηδενική υπόθεση H0:$\beta_1=0$ είναι αληθής, τότε η στατιστική t με τιμή

$$t = \frac{(b_1 - \beta_1)}{s_{b_1}} = \frac{b_1}{s_{b_1}}$$

ακολουθεί την t κατανομή με n-2 βαθμούς ελευθερίας.

› Το s_{b1} είναι η τυπική απόκλιση του b_1 και δίδεται από τη σχέση:

$$s_{b_1} = \frac{s_{y \cdot x}}{\sqrt{\sum (X - \bar{X})^2}}$$

Έλεγχος Υποθέσεων

- Ένας εναλλακτικός έλεγχος για την μηδενική υπόθεση H_0: $\beta_1=0$ είναι επίσης διαθέσιμος από τον πίνακα ANOVA.
- Αν οι υποθέσεις του στατιστικού μοντέλου για τη γραμμική παλινδρόμηση είναι κατάλληλες και αν η μηδενική υπόθεση H_0:$\beta_1=0$ είναι αληθής, ο λόγος:

$$F = \frac{MSR}{MSE}$$

- ακολουθεί την F κατανομή με df = 1, n-2.

Έλεγχος Υποθέσεων

- Όταν η H_0 είναι αληθής, το MSE είναι εκτιμητής του σ^2 και της διασποράς του σφάλματος (ε) στο στατιστικό μοντέλο για παλινδρόμηση ευθείας γραμμής.

- Από την άλλη μεριά, αν η εναλλακτική υπόθεση $H_1: \beta_1 \neq 0$ είναι αληθής, ο αριθμητής του F-λόγου τείνει να είναι μεγαλύτερος από τον παρονομαστή. Στην περίπτωση αυτή οι μεγάλοι F-λόγοι είναι συνεπείς με την εναλλακτική υπόθεση.
- Σε επίπεδο σημαντικότητας α, η περιοχή απόρριψης είναι η $F > F_{\alpha}$.

Ανάλυση Καταλοίπων

- Τα συμπεράσματα μπορεί να είναι παραπλανητικά, αν οι υποθέσεις, που έγιναν στη διαμόρφωση του μοντέλου, είναι συνολικά ασυμβίβαστες με τα δεδομένα.

- Είναι απαραίτητο να ελέγξουμε τα δεδομένα προσεκτικά για ενδείξεις κάποιας παραβίασης των υποθέσεων.

Ανάλυση Καταλοίπων

- Οι υποθέσεις για το μοντέλο παλινδρόμησης ευθείας γραμμής είναι:

1. Η υποκείμενη σχέση είναι γραμμική
2. Τα σφάλματα είναι ανεξάρτητα
3. Τα σφάλματα έχουν σταθερή διασπορά
4. Τα σφάλματα είναι κανονικά κατανεμημένα

Ανάλυση Καταλοίπων

- Η πληροφόρηση στην διασπορά, που δεν μπορεί να ερμηνευθεί από την προσαρμοσμένη συνάρτηση παλινδρόμησης, περιέχεται στα κατάλοιπα
- Για να ελέγξουμε τις μετρικές ενός δοκιμαστικού μοντέλου, μπορούμε να εξετάσουμε τα κατάλοιπα, χαράζοντάς τα με διάφορους τρόπους:

1. Γράφημα ιστογράμματος καταλοίπων
2. Γράφημα καταλοίπων σαν συνάρτηση των προσαρμοσμένων τιμών
3. Γράφημα καταλοίπων σαν συνάρτηση της ανεξάρτητης μεταβλητής
4. Γράφημα καταλοίπων στο χρόνο αν τα δεδομένα είναι χρονολογικά

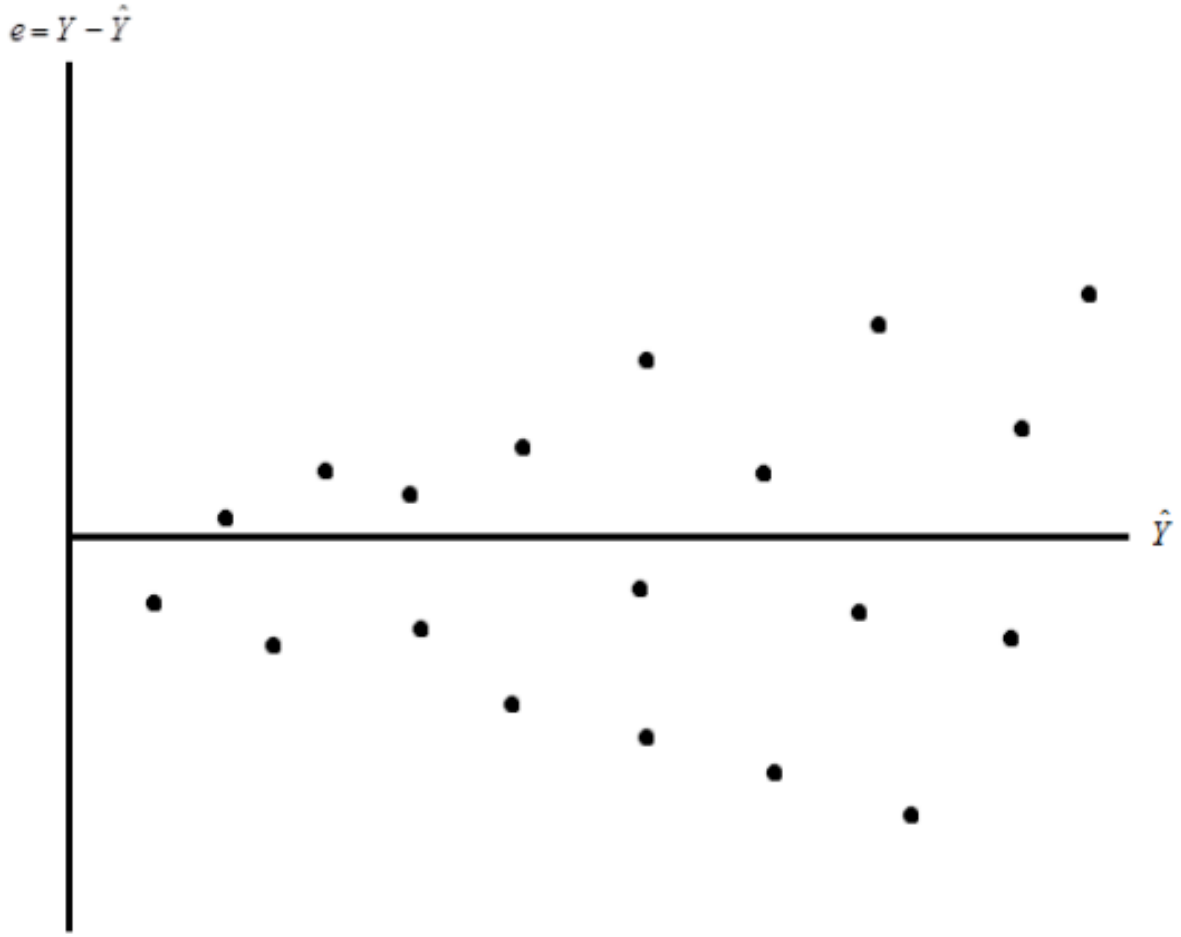
Ανάλυση Καταλοίπων

- Ένα ιστόγραμμα των καταλοίπων παρέχει έναν έλεγχο της κανονικότητας της υπόθεσης.
- Τυπικά, μέτριες αποκλίσεις από μία καμπύλη σε σχήμα καμπάνας δεν αποδυναμώνουν τις υποθέσεις από τους ελέγχους, ή τα διαστήματα πρόβλεψης, που βασίζονται στην t κατανομή, ειδικά αν το σετ δεδομένων είναι μεγάλο.
- Η παραβίαση της υπόθεσης της κανονικότητας μόνο, συνήθως δεν είναι τόσο σοβαρή, όσο η παραβίαση κάποιας από τις υπόλοιπες υποθέσεις.

Ανάλυση Καταλοίπων

- Αν μία γραφική παράσταση των καταλοίπων με τις προσαρμοσμένες τιμές υποδεικνύει, ότι η γενική φύση της σχέσης μεταξύ του Υ και του Χ διαμορφώνει μία καμπύλη αντί για μία ευθεία γραμμή, ένας κατάλληλος μετασχηματισμός των δεδομένων μπορεί να μετατρέψει μία μη γραμμική σχέση σε μία άλλη, που είναι σχεδόν γραμμική.
- Ένας μετασχηματισμός μπορεί ακόμα να σταθεροποιήσει τη διασπορά.

Ανάλυση Καταλοίπων



Ανάλυση Καταλοίπων

- Η υπόθεση της ανεξαρτησίας είναι η πιο κρίσιμη. Η έλλειψη ανεξαρτησίας μπορεί δραστικά να διαστρεβλώσει τα συμπεράσματα, που βγήκαν από τους t ελέγχους.

- Η υπόθεση ανεξαρτησίας είναι ιδιαίτερα επικίνδυνη για τις χρονοσειρές δεδομένων, τα οποία προκύπτουν συχνά σε επιχειρηματικά και οικονομικά προβλήματα πρόβλεψης.

Ανάλυση Καταλοίπων

- Για κατάλοιπα σειρών δεδομένων – το οποίο σημαίνει, για κατάλοιπα, που παράγονται χρησιμοποιώντας μεθόδους παλινδρόμησης σε χρονο-διατεταγμένα δεδομένα – η ανεξαρτησία μπορεί να ελεγχθεί με ένα διάγραμμα των καταλοίπων με το χρόνο.

Ανάλυση Καταλοίπων

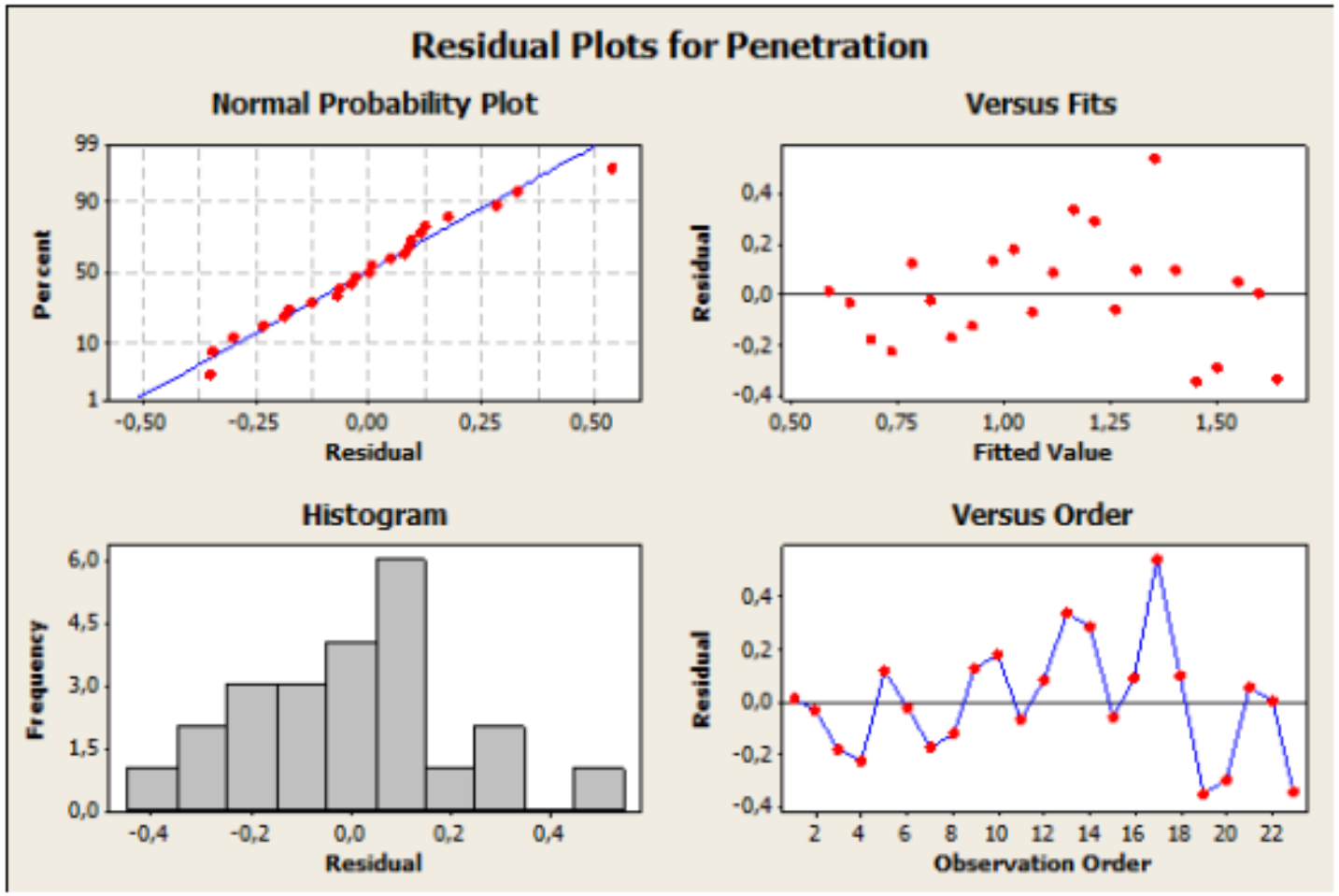
› Δεν θα έπρεπε να υπάρχει συστηματικό σχέδιο, όπως μία σειρά από υψηλές τιμές ακολουθούμενη από μία σειρά με χαμηλές τιμές. Επιπλέον, οι αυτοσυσχετίσεις δείγματος των καταλοίπων (sample autocorrelations of the residuals):

$$r_k(e) = \frac{\sum_{t=k+1}^{n} e_t e_{t-k}}{\sum_{t=1}^{n} e_t^2}, \quad k = 1, 2, \ldots, K$$

όπου n είναι ο αριθμός των καταλοίπων και K είναι χαρακτηριστικά n/4, θα έπρεπε να είναι όλες μικρές.

› Η ανεξαρτησία γίνεται εμφανής, αν ο κάθε ένας από τους συντελεστές αυτοσυσχέτισης των καταλοίπων είναι μέσα στο διάστημα $0 \pm 2/\sqrt{n}$ για όλα τα k.

Ανάλυση Καταλοίπων



Εφαρμογή

Τρίμηνο	Διείσδυση
1	0,6
2	0,6
3	0,5
4	0,5
5	0,9
6	0,8
7	0,7
8	0,8
9	1,1
10	1,2
11	1
12	1,2

Τρίμηνο	Διείσδυση
13	1,5
14	1,5
15	1,2
16	1,4
17	1,9
18	1,5
19	1,1
20	1,2
21	1,6
22	1,6
23	1,3

Μετασχηματισμοί Μεταβλητών

Όταν ένα διάγραμμα διασποράς υποδεικνύει ότι υπάρχει μία μη γραμμική σχέση μεταξύ του Y και του X, υπάρχουν δύο βασικές προσεγγίσεις:

- Η πρώτη είναι να προσαρμόσουμε τα δεδομένα με μία καμπύλη – μη γραμμική – συνάρτηση παλινδρόμησης και να χρησιμοποιήσουμε τη προσαρμοσμένη σχέση για σκοπούς πρόβλεψης.
- Η δεύτερη προσέγγιση περιλαμβάνει το μετασχηματισμό της μεταβλητής X σε άλλη μορφή, έτσι ώστε η προκύπτουσα σχέση με το Y να είναι γραμμική.

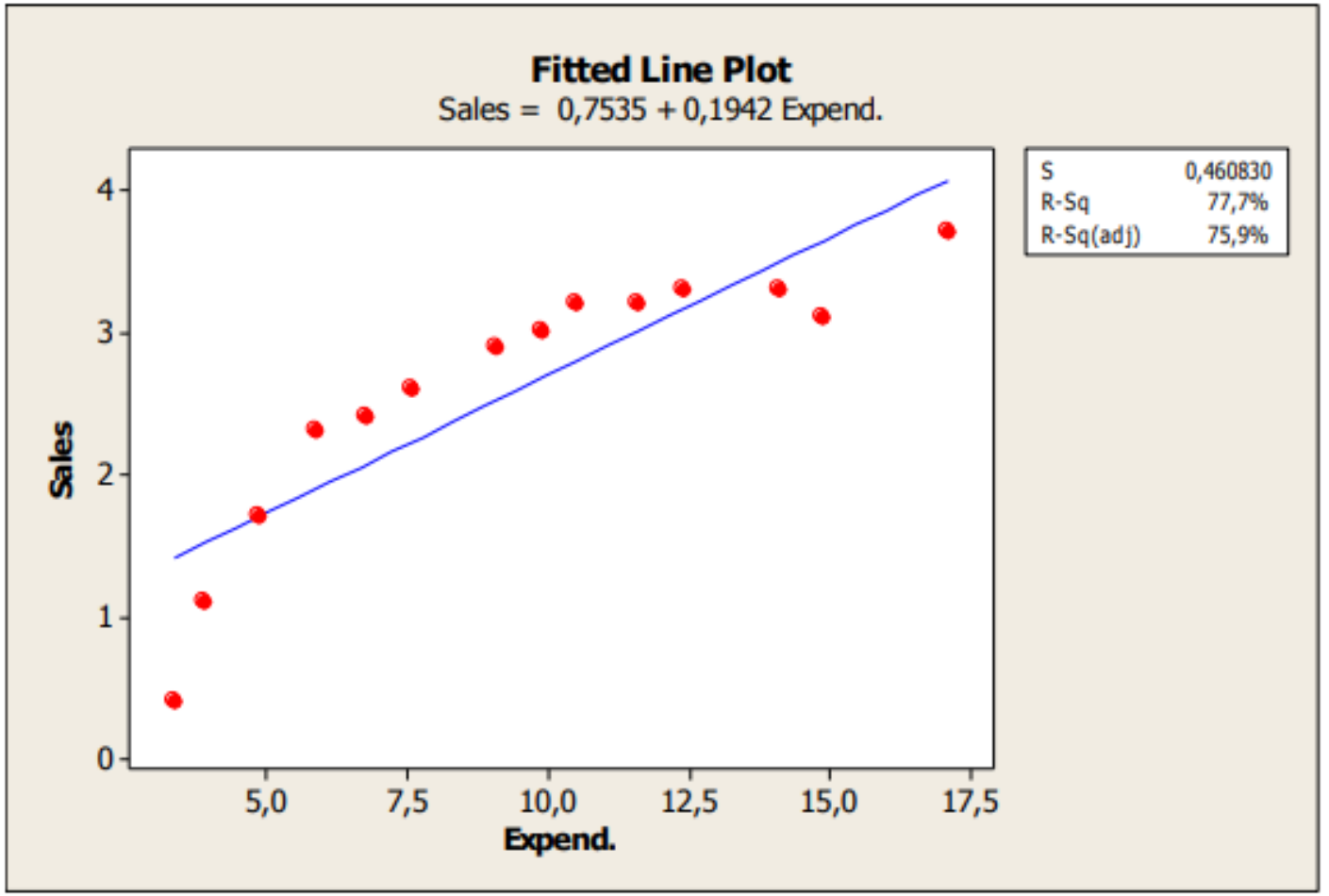
Μετασχηματισμοί Μεταβλητών

- Τέσσερις από τους πιο κοινούς μετασχηματισμούς (συναρτήσεις), που χρησιμοποιούνται για να παραχθούν νέες μεταβλητές πρόβλεψης, είναι οι: αντίστροφη, λογαριθμική, τετραγωνική ρίζα και τετράγωνο:

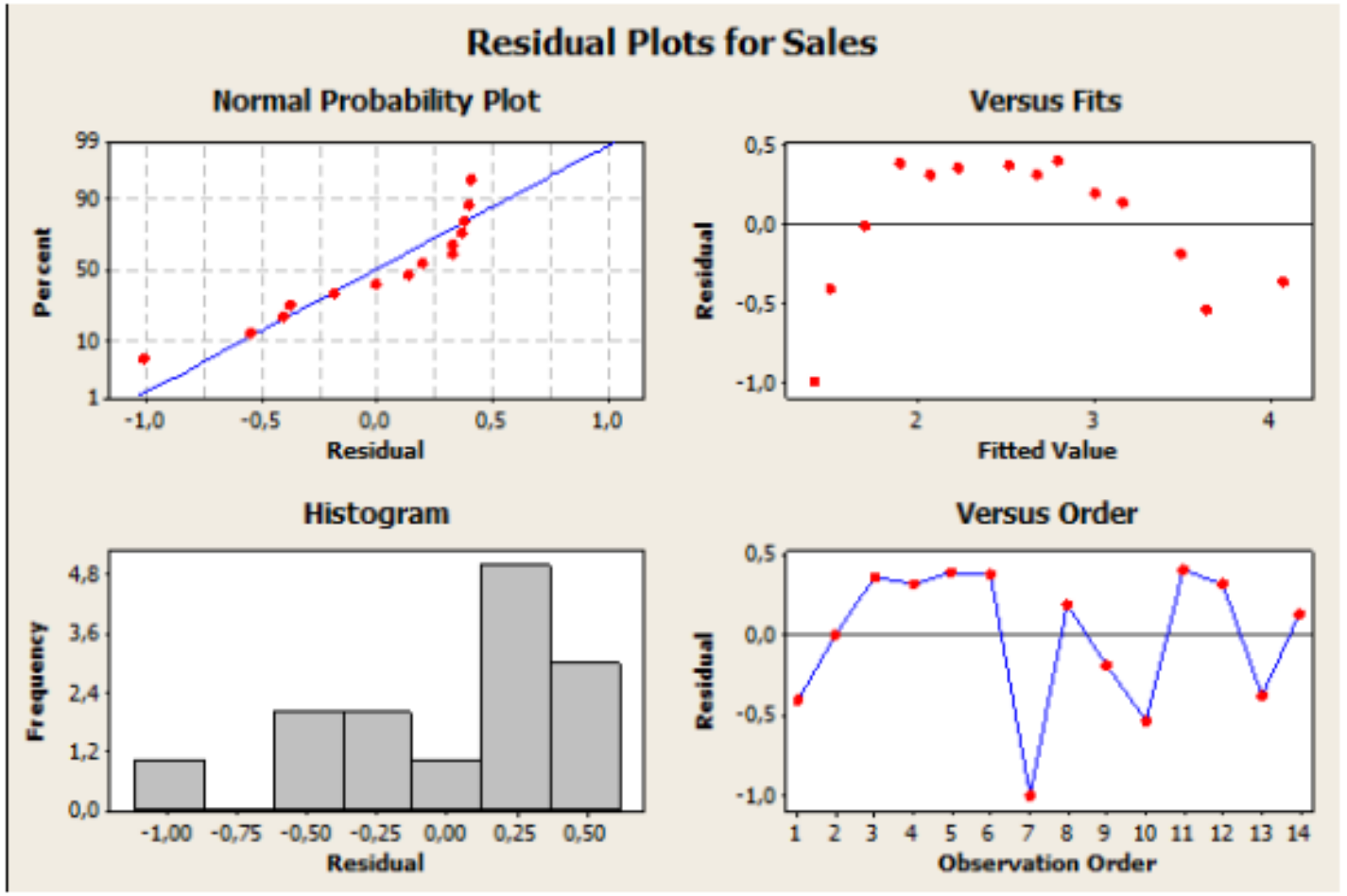
$$\frac{1}{X},\ \log X,\ \sqrt{X},\ X^2$$

- Όταν για όλες αυτές τις μεταβλητές δημιουργηθούν γραφικές παραστάσεις με το Υ, η ελπίδα είναι ότι η μη γραμμική σχέση μεταξύ του Υ και του Χ θα γίνει μία γραμμική σχέση μεταξύ του Υ και ενός εκ των μετασχηματισμών του Χ.

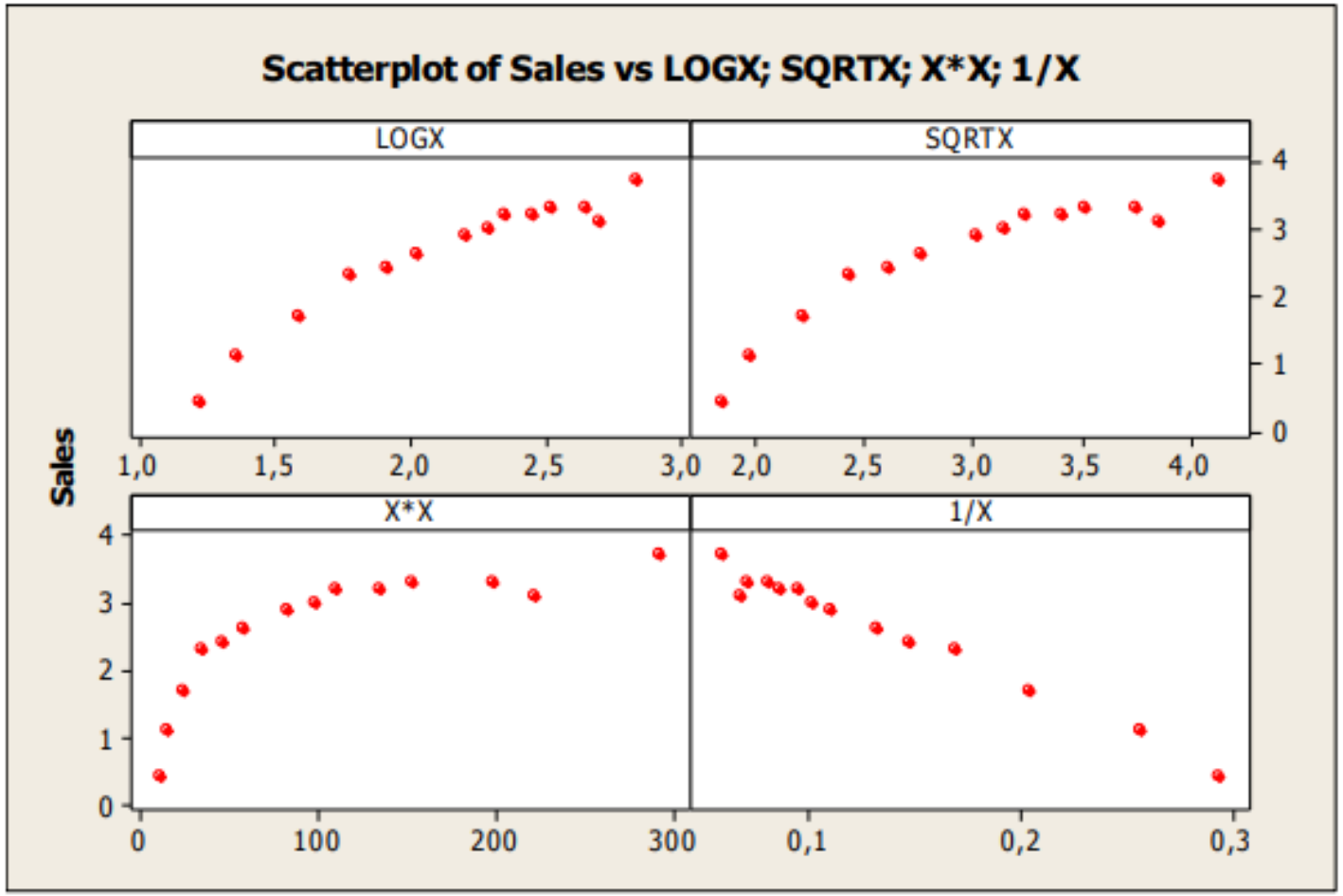
Μετασχηματισμοί Μεταβλητών



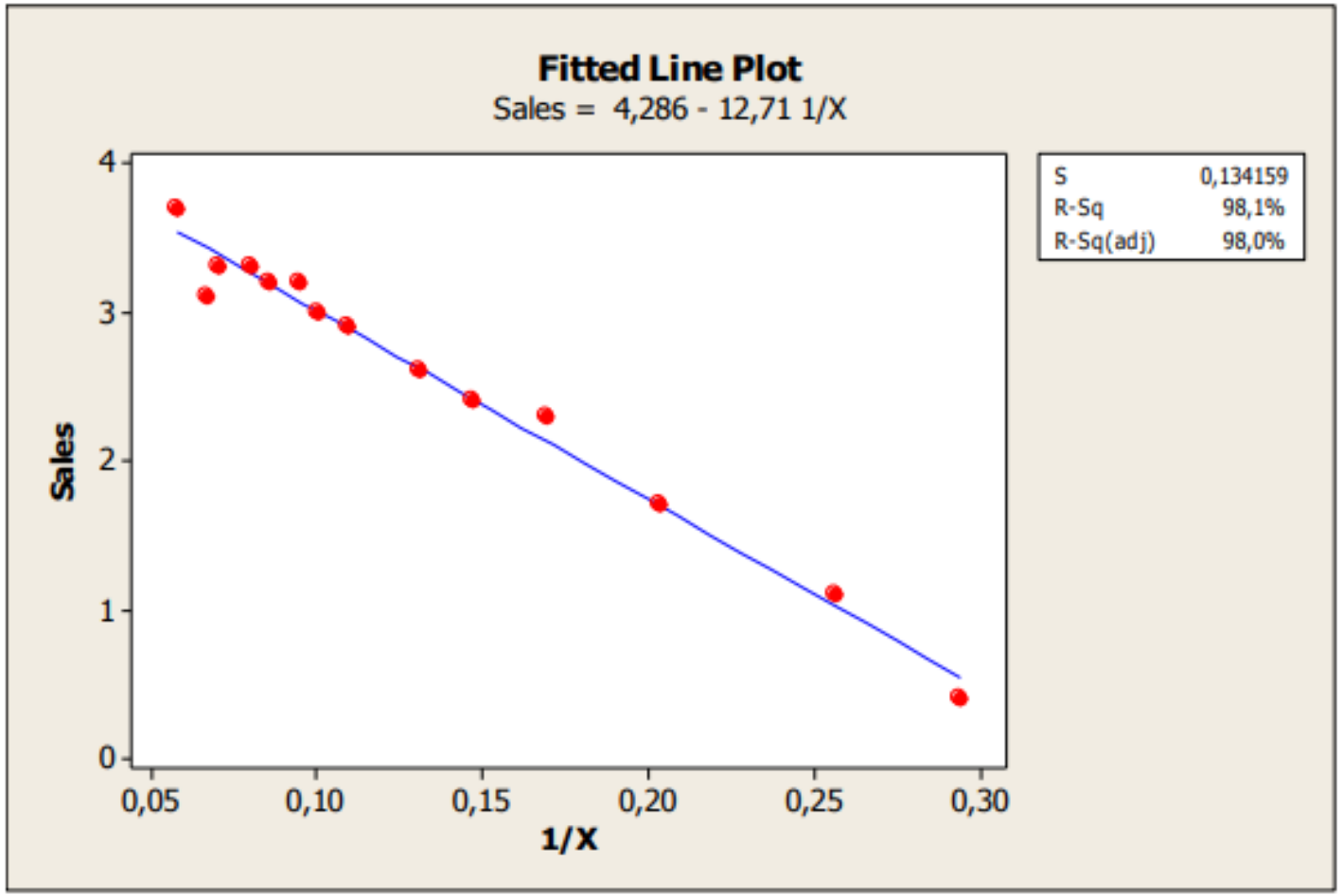
Μετασχηματισμοί Μεταβλητών



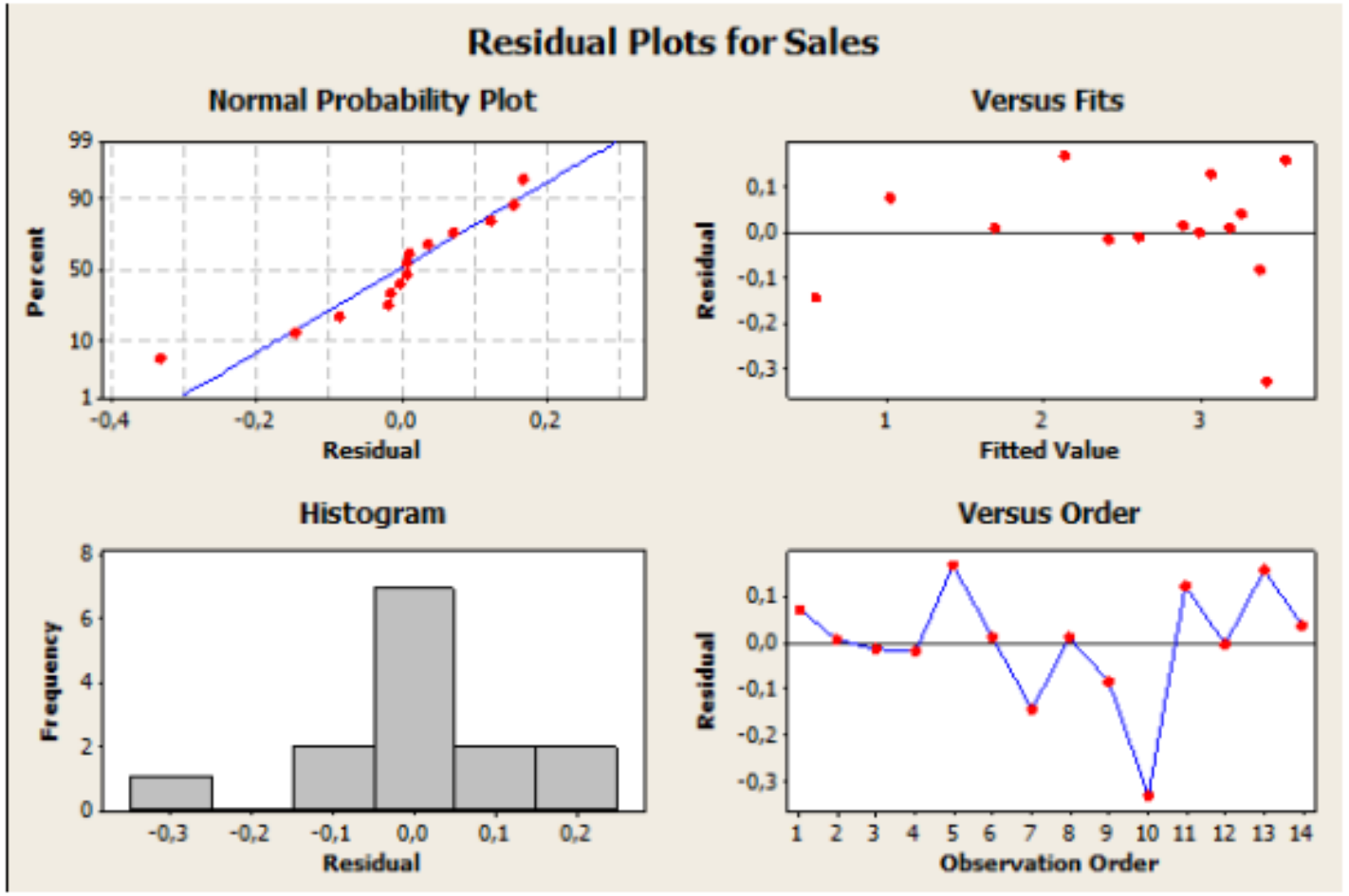
Μετασχηματισμοί Μεταβλητών



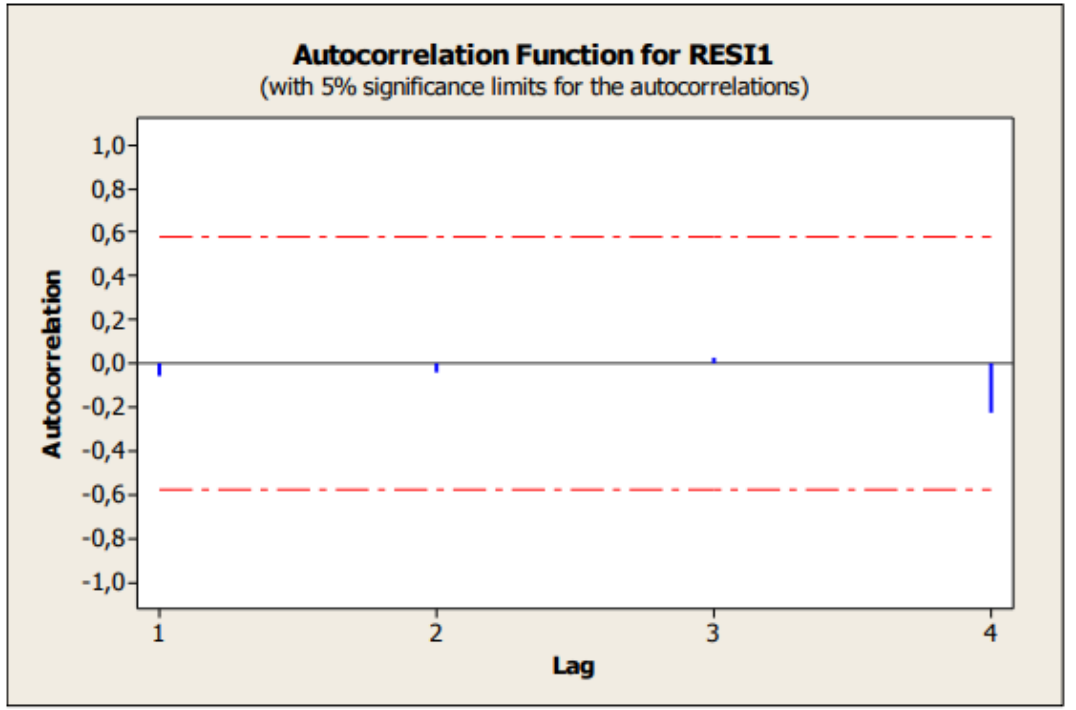
Μετασχηματισμοί Μεταβλητών



Μετασχηματισμοί Μεταβλητών



Μετασχηματισμοί Μεταβλητών



Lag	ACF	T	LBQ
1	-0,055110	-0,21	0,05
2	-0,043994	-0,16	0,09
3	0,027625	0,10	0,10
4	-0,223745	-0,83	1,23

Καμπύλες Ανάπτυξης

- Οι καμπύλες ανάπτυξης είναι καμπυλόγραμμες σχέσεις μεταξύ της μεταβλητής του ενδιαφέροντος και του χρόνου.
- Δείχνουν τον ετήσιο ρυθμό ανάπτυξης, που πρέπει να διατηρείται, έτσι ώστε να επιτυγχάνονται τα προσχεδιασμένα μελλοντικά επίπεδα τιμών. Αυτός ο ετήσιος ρυθμός ανάπτυξης μπορεί να είναι, ή να μην είναι λογικός.

Καμπύλες Ανάπτυξης

- Αν μία μεταβλητή, που υπολογίζεται με την πάροδο του χρόνου, αυξάνει στο ίδιο ποσοστό σε κάθε χρονική περίοδο, λέγεται ότι παρουσιάζει εκθετική ανάπτυξη (exponential growth).
- Αν η μεταβλητή αυξάνει κατά την ίδια ποσότητα κάθε χρονική περίοδο, λέγεται ότι παρουσιάζει γραμμική ανάπτυξη (linear growth).
- Μερικές φορές ένας απλός μετασχηματισμός θα μετατρέψει μία μεταβλητή, που έχει εκθετική ανάπτυξη, σε μία μεταβλητή, που έχει γραμμική ανάπτυξη.

Ερωτήσεις???