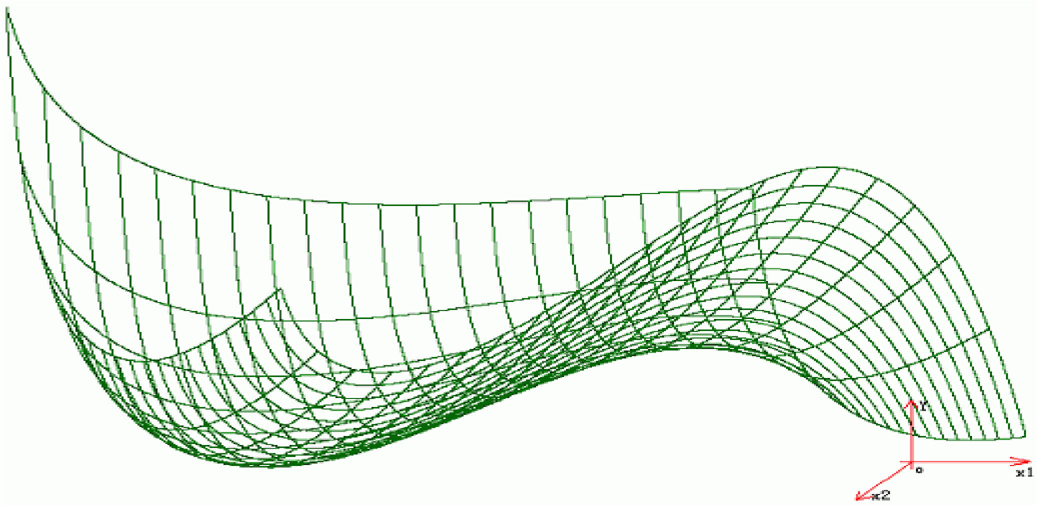
Τ.Ε.Ι. ΠΑΤΡΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ ΣΕ ΟΙΚΟΝΟΜΙΚΑ ΔΕΔΟΜΕΝΑ



Ιορδάνου Γαρυφαλλιά
Ευφραίμης Ανδρέας

Επίβλεψη:Μπουμπούλη Αθανασία

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Πάτρα,Δεκέμβριος 2010

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ

ΚΕΦΑΛΑΙΟ 1

ΥΠΟΔΕΙΓΜΑ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

1.1 ΕΙΣΑΓΩΓΗ.....5

1.2 ΤΟ ΜΟΝΤΕΛΟ ΤΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.....7

1.3 Η ΚΑΤΑΝΟΜΗ ΤΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.....9

1.4 ΕΚΤΙΜΙΣΗ ΚΑΙ ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.....11

1.5 ΜΕΤΑΒΛΗΤΟΤΙΤΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ.....13

1.6 ΙΔΙΟΤΙΤΕΣ ΤΩΝ ΕΚΤΙΜΗΤΩΝ ΚΑΙ Ο ΕΚΤΙΜΗΤΗΣ ΔΙΑΚΥΜΑΝΣΕΩΣ σ^2 15

1.7 ΤΥΠΙΚΑ ΣΦΑΛΜΑΤΑ,ΟΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ ΚΑΙ ΤΑ ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΕΝΟΣ ΔΕΙΓΜΑΤΟΣ.....17

1.8 ΜΕΘΟΔΟΣ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ.....21

1.9 ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΠΑΛΙΝΔΡΟΜΗΣΗ.....23

1.10 ΕΚΤΙΜΙΣΗ ΠΟΛΥΜΕΤΑΒΛΗΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ.....25

1.11 ΑΚΡΑΙΕΣ ΤΙΜΕΣ ΚΑΙ ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΠΑΛΙΝΔΡΟΜΗΣΕΩΝ....27

ΚΕΦΑΛΑΙΟ 2

ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

2.1 ΕΙΣΑΓΩΓΗ.....29

2.1.1 Η ΕΝΝΟΙΑ ΤΗΣ ΓΡΑΜΜΙΚΟΤΗΤΑΣ.....30

2.2 Η ΜΕΘΟΔΟΣ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΟΝΩΝ ΣΤΗΝ ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ.....31

2.2.1 ΠΡΟΒΛΗΜΑ ΤΩΝ ΜΗ ΓΡΑΜΜΙΚΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ....32

2.2.2 Η ΕΠΙΛΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ.33

2.2.3 ΔΙΑΦΟΡΕΣ ΜΕΤΑΞΥ ΓΡΑΜΜΙΚΩΝ ΚΑΙ ΜΗ ΓΡΑΜΜΙΚΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ 37
2.3 ΜΗ ΓΡΑΜΜΙΚΕΣ ΜΟΡΦΕΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ........39
2.3.1 ΠΟΛΥΩΝΥΜΙΚΗ ΜΟΡΦΗ........39
2.3.2 ΑΝΤΙΣΤΡΟΦΗ ΜΟΡΦΗ........40
2.3.3. ΕΚΘΕΤΙΚΗ ΜΟΡΦΗ........43
2.3.3.1. ΥΠΟΔΕΙΓΜΑ ΣΤΑΘΕΡΗΣ ΜΕΓΕΝΘΥΣΗΣ........44
2.3.4ΛΟΓΙΣΤΙΚΗ ΚΑΜΠΥΛΗ........46
2.3.5 ΣΥΝΑΡΤΗΣΗ ΣΤΑΘΕΡΩΝ ΕΛΑΣΤΙΚΟΤΗΤΩΝ........47
2.4ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΑΝΑΜΕΣΑ ΣΤΗΝ ΓΡΑΜΜΙΚΗ ΚΑΙ ΣΤΗ ΛΟΓΑΡΙΘΜΙΚΑ ΓΡΑΜΜΙΚΗ ΜΟΡΦΗ........48
2.4.1 BOX-COX........49
2.4.2 BERA-McALEER........51
2.4.3 McKINNON-WHITE-DAVIDSON........53

<u>ΚΕΦΑΛΑΙΟ 3</u>

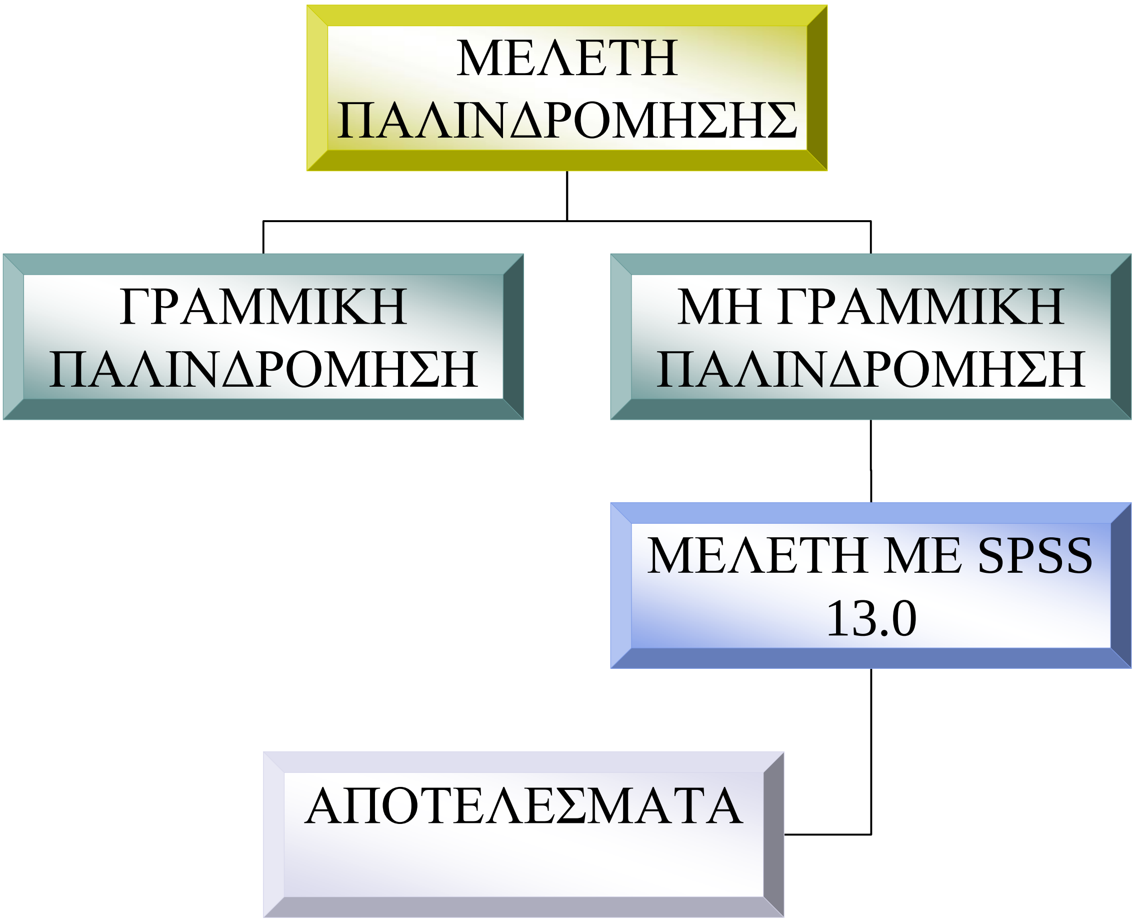
<u>ΕΦΑΡΜΟΓΕΣ ΜΕ SPSS 13.0</u>

3.1 ΜΟΝΤΕΛΟ ΚΟΣΤΟΥΣ 55
3.2 ΔΑΠΑΝΗ ΓΙΑ ΤΡΟΦΗΜΑ ΚΑΙ ΑΚΑΘΑΡΙΣΤΟ ΕΘΝΙΚΟ ΕΙΣΟΔΗΜΑ......69
3.3 ΑΠΟΤΑΜΙΕΥΣΗ ΚΑΙ ΔΙΑΘΕΣΙΜΟ ΕΙΣΟΔΗΜΑ........75

<u>ΒΙΒΛΙΟΓΡΑΦΙΑ</u>........80

ΠΕΡΙΛΗΨΗ

Μέσα από την μελέτη και την ανάπτυξη του θέματος της μη γραμμικής παλινδρόμησης , στόχος μας είναι να παρουσιάσουμε και να αναλύσουμε οικονομικά μοντέλα τα οποία δεν υπακούν στους κανόνες της γραμμικότητας .Μέσα από διαδικασίες και τεχνικές που χρησημοποιούνται φαίνεται θεωρητικά και πρακτικά η τεχνική εξομάλυνσης των μη γραμμικών μορφών παλινδρομήσεων και μετατροπής τους σε γραμμικές έτσι ωστε να είναι δυνατή η επεξεργασία και η μελέτη τους .



ΚΕΦΑΛΑΙΟ 1

ΥΠΟΔΕΙΓΜΑ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

1.1 ΕΙΣΑΓΩΓΗ

Στη στατιστική η γραμμική παλινδρόμηση αναφέρετε σε οποιαδήποτε προσέγγιση ώστε να μοντελοποιηθεί η σχέση μεταξύ ενός ή περισσοτέρων μεταβλητών δεδομένου y και μία ή περισσότερες δεδομένου x ,τέτοιο ώστε το μοντέλο να εξαρτάται γραμμικά από τις άγνωστες παραμέτρους για να εκτιμηθούν από τα δεδομένα. Αυτού του είδους το μοντέλο ονομάζετε γραμμικό μοντέλο. Συνήθως , η γραμμική παλινδρόμηση αναφέρετε σ' ένα μοντέλο κατά το οποίο ο μέσος του y δεδομένου την αξία του x είναι μία συνάρτηση του x. Σε κάποιες περιπτώσεις , η γραμμική παλινδρόμηση μπορεί να αναφέρετε σ' ένα μοντέλο στο οποίο η διάμεσος ή κάποια άλλη μεταβλητή της κανονικής κατανομής του y δεδομένου x εκφράζετε σαν γραμμική λειτουργία του x .

Η γραμμική παλινδρόμηση ήταν ο πρώτος τύπος παλινδρόμησης που μελετήθηκε ενεργά και χρησιμοποιήθηκε ιδιαίτερα σε πρακτικές εφαρμογές. Βασικός παράγοντας ήταν ότι τα μοντέλα που εξαρτώνται γραμμικά από τις άγνωστες παραμέτρους τους είναι ευκολότερο να προσαρμοστούν από τα μη γραμμικά και τα στατιστικά αποτελέσματα που αφορούν τους εκτιμητές είναι ευκολότερο να προσδιοριστούν.

Η γραμμική παλινδρόμηση έχει πολλές πρακτικές εφαρμογές από τις οποίες οι περισσότερες ανήκουν σε δύο ευρύ κατηγορίες :

i. Αν ο στόχος είναι η πρόβλεψη η γραμμική παλινδρόμηση μπορεί να χρησιμοποιηθεί ώστε να ενσωματωθεί σε ένα μοντέλο με σκοπό να παρατηρηθεί η αντίδραση του y σε σχέση με την αξία του x.

ii. Δεδομένου μία μεταβλητή y και ένα πλήθος από μεταβλητές x_1 έως x_p που σχετίζονται με το y, τότε η γραμμική παλινδρόμηση μπορεί να χρησιμοποιηθεί για να προσδιορίσει πόσο ισχυρή είναι η σχέση μεταξύ y και x_i , ποια x_i δεν έχουν καθόλου σχέση με το y και να ταυτοποιήσουν ποια x_i σε ένα πολυωνυμικό υπόδειγμα έχουν ουσιαστικές πληροφορίες για το y , διότι αν κάποια από αυτό είναι γνωστά τα υπόλοιπα δεν είναι ιδιαίτερα χρήσιμα.

Τα μοντέλα της γραμμικής παλινδρόμησης συχνά χρησιμοποιούν τη μέθοδο ελαχίστων τετραγώνων κυρίως γιατί είναι σχετικά εύχρηστη και χρήσιμη , ακόμα και στην επέκταση της γραμμικής παλινδρόμησης σε μη γραμμική λόγω της ιδιαιτερότητας κάποιων μοντέλων , όπως θα δούμε παρακάτω.

1.2 ΤΟ ΜΟΝΤΕΛΟ ΤΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΙΣΗΣ

Το γραμμικό μοντέλο της παλινδρόμησης μπορεί να χωριστεί σε δύο κατηγορίες . Η μία κατηγορία είναι η απλή γραμμική παλινδρόμηση που αναφέρεται σε σχέσεις που περιλαμβάνουν μία μόνο ερμηνευτική μεταβλητή.

Στο συγκεκριμένο μοντέλο έχουμε τ ζεύγη παρατηρήσεων (y_t,x_t) για ένα ορισμένο χρονικό διάστημα κι αν υποθέσουμε πάντα ότι η σχέση μεταξύ των παρατηρήσεων μας είναι γραμμική τότε η μαθηματική μορφή των συναρτησιακών σχέσεων είναι :

$$y_t = b_o + b_1 x_t$$

Αυτή η σχέση είναι προσδιοριστική και σημαίνει οτι όλες οι μεταβλητές θα έχουν την ίδια συμπεριφορά. Αυτό στην πραγματικότητα δεν ισχύει και αυτή η σχέση δεν μπορεί να ικανοποιήσει όλα τα ζεύγη ειδικά αν μιλάμε για οικονομικά μεγέθη. Οι διαφορές μπορούν να ληφθούν υπόψη με την προσθήκη μίας τυχαίας μεταβλητής $\mathbf{u_t}$ οπότε η προσδιοριστική σχέση γίνετε στοχαστική της μορφής :

$$y_t = b_0 + b_1 x_t + u_t$$

Οι βασικοί λόγοι που την ύπαρξη αυτού του όρου είναι ότι υπάρχουν κι άλλοι αστάθμητοι παράγοντες που επηρρεάζουν το υπόδειγμα. Για παράδειγμα , η ανθρώπινη συμπεριφορά η οποία είναι απρόβλεπτη , με αποτέλεσμα να επηρρεάζει σε σημαντικό βαθμό όλα τα υποδείγματα οικονομικής φύσεως. Επιπλέον επειδή τα σφάλματα μετρήσεως των μεταβλήτων είναι αναπόφευκτα ακόμα και η θεωρητική σχέση ανάμεσα στις μεταβλητές να είναι ακριβής , πάλι θα υπάρχει απόκλιση στις μετρήσεις .

Στην παραπάνω σχέση , λόγω της στοχαστικής της φύσης συνεπάγεται πως για κάθε τιμή της μεταβλητής x δεν υπάρχει μόνο μία τιμή για την y αλλά ολόκληρη κατανομή τιμών που εξαρτάται από το διαταρακτικό όρο. Για την εκτίμηση αυτής της σχέσεως έχουμε T ζεύγη παρατηρήσεων (x , y) αλλά δεν έχουμε για τον $\mathbf{u_t}$ οπότε θα κάνουμε ορισμένες υποθέσεις για την συμπεριφορά του.

$$y_t = b_0 + b_1 x_t + u_t$$
$$u_t \sim (0, s^2)$$

Οι υποθέσεις μας σημαίνουν τα εξής : η μεταβλητή $\mathbf{u_t}$ είναι τυχαία μεταβλητή που παίρνει τιμές θετικές και αρνητικές άλλα κατά μέσο όρο η τιμή της είναι 0. Η διακύμανση της είναι σταθερή για όλες τις τιμές της x , δηλαδή η διασπορά των τιμών της $\mathbf{u_t}$ από τον μέσο δεν αλλάζει όταν μεταβάλετε η τιμή της x_t , αλλά παραμένει η ίδια. Όταν συμβαίνει αυτό ο διαταρακτικός μας όρος ονομάζεται ομοσκεδαστικός . **($Eu^2_t = \sigma^2$)**

Ακόμη σημαντική είναι η υπόθεση ότι οι διαταρακτικοί όροι δε συσχετίζονται μεταξύ τους , επομένως η συνδιακύμανση του διαταρακτικού όρου της παρατηρήσεως t , με το διαταρακτικό όρο οποιασδήποτε άλλης παρατηρήσεως s είναι μηδέν. **($Eu_tu_s=0$ για $t \neq s$)**

Τέλος ,υποθέτουμε πως η μεταβλητή x δεν είναι στοχαστική και πως οι τιμές της παραμένουν σταθερές και δεν είναι όλες ίσες μεταξύ τους . Οι τιμές φυσικά που παίρνει ο διαταρακτικός όρος μεταβάλλονται , όπως επίσης και οι τιμές της εξαρτημένης μεταβλητής.

1.3 Η ΚΑΤΑΝΟΜΗ ΤΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΙΣΗΣ

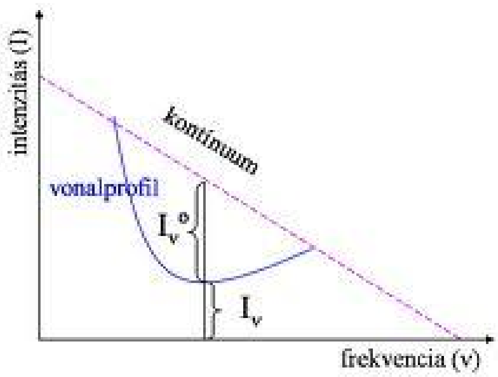
Η μεταβλητή y είναι συνάρτηση της τυχαίας μεταβλητής u και επομένως η y είναι τυχαία μεταβλητή. Επιπλέον η κατανομή της y είναι κατανομή υπό συνθήκη , δεδομένης της τιμής της x. Άρα ο μέσος $E(Y_t)$ και η διακύμανση $V(Y_t)$ που είναι πραγματικά υπό συνθήκη μέσος και πραγματικά υπό συνθήκη διακύμανση μπορούν να δειχθούν από τις ακόλουθες σχέσεις :

$$E\left(Y_t\right) = b_0 + b_1 X_t \quad \text{και} \quad V\left(Y_t\right) = s^{2}$$

Οι παραπάνω σχέσεις μπορούν να αποδεδειχθούν μαθηματικά με βάση τον ορισμό της προσδοκώμενης τιμής και τον ορισμό της διακύμανσης . Επομένως η μεταβλητή Y_t , σύμφωνα με τις υποθέσεις του υποδείγματος μας που προαναφέρθηκαν , ακολουθεί μια κατανομή με βάση τις παραπάνω σχέσεις .

Η σχέση της $y_{(t)}$ ανάμεσα στους μέσους της μη εξαρτημένης μεταβλητής και στις αντίστοιχες τιμές του x είναι η γραμμή παλινδρομήσεως που ζητάμε ιδιαίτερα όταν μιλάμε για πληθυσμό . Ουσιαστικά απεικονίζει τη σχέση ανάμεσα στους μέσους της εξαρτημένης μεταβλητής y και στις αντίστοιχες τιμές της ανεξάρτητης μεταβλητής x . Από αυτή τη σχέση βγαίνει η μορφή της γραμμής παλινδρομήσεως στο δείγμα και από την ανάγκη μας να την υπολογίσουμε ώστε να ξέρουμε από ποια σημεία διέρχεται .

Έτσι μπορούμε να εκτιμήσουμε τους συντελεστές β_0 και β_1 και έχουμε τη μορφή : $\hat{y}_t = \hat{b}_0 + \hat{b}_1 x_t$ και η διαφορά των υπολογισμένων τιμών από τις πραγματικές είναι της μορφής $u_t = y_t - \hat{y}_t$ που είναι το κατάλοιπο και η εκτίμηση του διαταρακτικού όρου u_t .



Σχήμα 1: Κατανομή υπολοίπων στον πληθυσμό

1.4 ΕΚΤΙΜΗΣΗ ΚΑΙ ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΙΣΗΣ

Συνήθως , για την εκτίμηση των παραμέτρων της γραμμικής παλινδρόμησης χρησιμοποιούμε τη μέθοδο ελαχίστων τετραγώνων (least squares method) . Αυτό συμβαίνει λόγω της σχετικής απλότητας της και λόγω ότι οι εκτιμητές έχουν πολλές επιθυμητές ιδιότητες . Επειδή το εύρος των εκτιμητών μπορεί να είναι άπειρο ουσιαστικά χρησιμοποιούμε τις τιμές εκεί που ελαχιστοποιείται η συνάρτηση :

$$\min \sum_{1}^{N} \hat{e}_i^2 = \min \sum_{i=1}^{N} (y_i - \hat{b}_0 - \hat{b}_1 x_i)^2$$

Αν παραγωγίσουμε ως προς $\hat{b}_0$ και ως προς $\hat{b}_1$ έχουμε το παρακάτω σύστημα: $\frac{\partial \sum e_i^2}{\partial \hat{b}_0} = 0$ και $\frac{\partial \sum \hat{e}_i^2}{\partial \hat{b}_1} = 0$.

Η λύση του συστήματος μας δίνει και τις εξισώσεις προσδιορισμού των εκτιμητών μας , οι οποίες είναι : $\hat{b}_0 = \overline{y} - \hat{b}_1 \overline{x}$ και $\hat{b}_1 = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}$.

Για να κατανοήσουμε τη λειτουργία της γραμμής παλινδρόμησης , την οποία εκτιμάμε με τη μέθοδο ελαχίστων τετραγώνων , πρέπει να λάβουμε υπόψη κάποιες βασικές ιδιότητες της :

i. Η γραμμή παλινδρομήσεως του δείγματος διέρχεται από το σημείο $(\overline{x}, \overline{y})$, δηλαδή από το σημείο που ορίζεται από το μέσο των x και y . Αυτό μπορεί να διαπιστωθεί και από τη σχέση $\hat{y}_i = \hat{b}_0 + \hat{b}_1\overline{x}$ όπου $\hat{y}$ ισούται με $\overline{y}$.

ii. Το άθροισμα των τιμών της y από το δείγμα είναι ίσο με το άθροισμα των τιμών που υπολογίζουμε από την παλινδρόμηση , δηλαδή $\sum \hat{y}_i = \sum y_i$. Αυτή η σχέση δηλώνει ότι ο εκτιμητής μας είναι αμερόληπτος διότι :

$$\sum \hat{y}_i = \sum (\hat{b}_0 + \hat{b}_1 x_i) = n\hat{b}_0 + \hat{b}_1 \sum x_i$$
$$\sum y_i = \sum (b_0 + b_1 x_i) = n b_0 + b_1 \sum x_i$$

Οπότε $E\left(\hat{b}_0\right) = b_0$ και $E\left(\hat{b}_1\right) = b_1$

iii. Το άθροισμα των καταλοίπων της γραμμής παλινδρομήσεως είναι 0. Συνεπώς $\sum e_i = 0$ επειδή $\sum \hat{y}_i - \sum y_i = 0$

iv. Επίσης το άθροισμα των γινομένων των τιμών της x και των καταλοίπων είναι 0 , δηλαδή $\sum x_i \hat{e}_i = 0$.

v. Το άθροισμα των γινομένων των καταλοίπων και των τιμών y που υπολογίζουμε από την παλινδρόμηση του δείγματος είναι 0 , δηλαδή $\sum \hat{y}_i \hat{e}_i = 0$.

Σημαντικό είναι να υπολογίσουμε την ελαστικότητα της συνάρτησης στο αντιπροσωπευτικό σημείο που είναι το $(\overline{x}, \overline{y})$, γιατί δεν έχει νόημα να υπολογίζουμε πολλά σημεία ελαστικότητας .

$$Eyx = \frac{\Delta y}{\Delta x}\frac{x_i}{y_i} \qquad \text{και} \qquad \frac{\Delta y}{\Delta x} = \frac{\partial x_i}{\partial y_i} = \frac{\partial\left(\hat{b}_0 + \hat{b}_1 x_i\right)}{\partial x_i} = \hat{b}_1$$

Οπότε $$Eyx = \frac{\partial y_i}{\partial x_i}\frac{\overline{x_i}}{\overline{y_i}} = \hat{b}_1 \frac{\overline{x_i}}{\overline{y_i}} = a$$

Για μία ποσοστιαία μεταβολή x θα έχουμε μία αύξηση της y κατά α% .

1.5 ΜΕΤΑΒΛΗΤΟΤΙΤΑ ΠΑΛΙΝΔΡΟΜΙΣΗΣ ΚΑΙ ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ

Το υπόδειγμα μας αναφέραμε πως χωρίζεται σε δύο μέρη : το συστηματικό μέρος $b_0 + b_1 x_t$ και το διαταρακτικό ή τυχαίο όρο (u_t) . Μέρος της μεταβλητότητας που παρατηρείται στις τιμές της y οφείλεται στις μεταβολές της μεταβλητής x . Σκοπός είναι να μπορέσουμε να προσδιορίσουμε πόση είναι η μεταβλητότητα στις τιμές της y που ερμηνεύεται από την παλινδρόμηση . Μπορούμε να χωρίσουμε τη συνολική μεταβλητότητα της y σε δύο μέρη: στη μεταβλητότητα που ερμηνεύεται από την παλινδρόμηση και στη μεταβλητότητα που οφείλετε στους τυχαίους παράγοντες .

Είναι μαθηματικά αποδεδειγμένο και θα αναφερόμαστε στη σχέση που δείχνει τη συνολική μεταβλητότητα της y και ορίζεται ως το άθροισμα των τετραγώνων των αποκλίσεων των παρατηρήσεων της y από το μέσο τους , δηλαδή $\sum (y_i - \bar{y})^2$ ως συνολικό άθροισμα τετραγώνων **(SST)** , στον όρο $\sum (\hat{y}_i - \bar{y})^2$ που εκφράζει τη μεταβλητότητα που ερμηνεύεται από την παλινδρόμηση , ως άθροισμα τετραγώνων παλινδρομήσεως **(RSS)** και στον όρο $\sum e_i^{\,2}$, που εκφράζει την μεταβλητότητα που ερμηνεύεται από τυχαίου παράγοντες , ως άθροισμα των τετραγώνων των καταλοίπων **(ESS)** . Η σχέση μεταξύ τους φαίνεται από την παρακάτω εξίσωση :

TSS = RSS + ESS

Αναφερόμαστε στον συντελεστή προσδιορισμού του δείγματος εκφράζοντας την αναλογία της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από τη παλινδρόμηση δηλαδή την μεταβλητότητα μιας εξαρτημένης μεταβλητής σε σχέση με μία ανεξάρτητη μεταβλητή και το συμβολίζουμε με $\mathbf{R^2}$. Το R^2 δεν μπορεί να πάρει τιμές αρνητικές ή μεγαλύτερες από τη μονάδα $0 \leq R^2 \leq 1$ και σχετίζεται με τους όρους της μεταβλητότητας της y με τις σχέσεις :

$$R^2 = \frac{RSS}{TSS} \quad \text{και} \quad R^2 = 1 - \frac{ESS}{TSS}$$

Παρατηρήθηκε ότι η εισροή επεξηγηματικών μεταβλητών (x’s) είχε ως αποτέλεσμα να αυξάνεται η τιμή του R^2 . Γι’ αυτό και χρησιμοποιήθηκε ο διορθωμένος συντελεστής προσδιορισμού $\bar{R}^2$ με μορφή :

$$\bar{R}^2 = \frac{RSS/(n-k-1)}{TSS/(n-1)} \quad \text{και ισχύει} \quad \bar{R}^2 < R^2$$

Γνωρίζουμε ότι ως μέτρο του βαθμού συσχετίσεως των δύο μεταβλητών χρησιμοποιείται ο συντελεστής συσχετίσεως **r** . Ουσιαστικά είναι ο συντελεστής συσχετίσεως του δείγματος και ορίζεται ως :

$$r(x,y)=\frac{\sum\left(x_i-\overline{x}\right)}{\sqrt{\sum\left(y_i-\overline{y}\right)^2\sum\left(x_i-\overline{x}\right)^2}}$$

Από την παραπάνω σχέση είναι φανερό πως ο συντελεστής συσχετίσεως του δείγματος είναι η τετραγωνική ρίζα του συντελεστή προσδιορισμού . Αν και υπάρχει τόσο στενή σχέση μεταξύ τους ,υπάρχει σημαντική διαφορά προς την ερμηνεία τους . Ο συντελεστής συσχετίσεως του δείγματος είναι ένας εκτιμητής του συντελεστής συσχετίσεως του πληθυσμού (ρ) το οποίο είναι μία άγνωστη παράμετρος της συνδυασμένης κατανομής δύο τυχαίων μεταβλητών ενώ ο συντελεστής προσδιορισμού αναφέρεται στην αναλογία της μεταβλητότητας της Y που ερμηνεύει η μεταβλητή X , για την οποία υποθέτουμε ότι είναι τυχαία μεταβλητή . Επιπλέον , ο συντελεστής συσχετίσεως του δείγματος είναι μέτρο μόνο της γραμμικής συσχετίσεως ή εξαρτήσεως δύο μεταβλητών , με αποτέλεσμα να έχει περιορισμένη χρήση.

1.6 ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΕΚΤΙΜΗΤΩΝ ΚΑΙ Ο ΕΚΤΙΜΗΤΗΣ ΔΙΑΚΥΜΑΝΣΕΩΣ σ^2

Οι συντελεστές που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων έχουν ορισμένες επιθυμητές ιδιότητες . Σύμφωνα με το θεώρημα των Gauss-Markov , για το κλασσικό γραμμικό υπόδειγμα οι εκτιμητές που προκύπτουν από τη μέθοδο ελαχίστων τετραγώνων είναι άριστοι γραμμικοί αμερόληπτοι εκτιμητές . Ουσιαστικά αυτό σημαίνει ότι :

i. Είναι αμερόληπτοι εκτιμητές $E\left(\hat{b}_0\right) = \hat{b}_0$ και $E\left(\hat{b}_i\right) = \hat{b}_i$

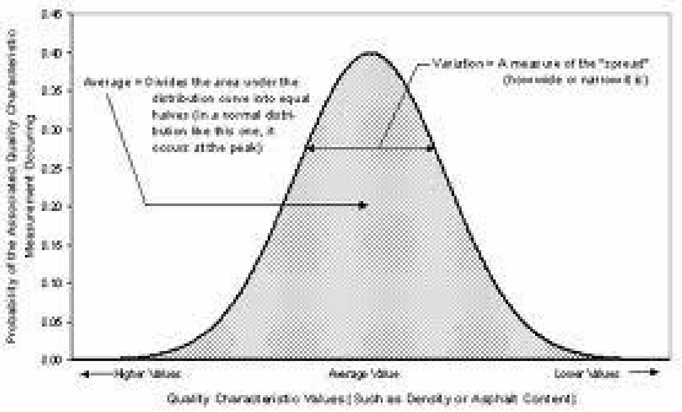
ii. Είναι γραμμικές συναρτήσεις των παρατηρήσεων των εξαρτημένων μεταβλητών y_i

iii. Έχουν την μικρότερη δυνατή διακύμανση μεταξύ των υπολοίπων αμερόληπτων εκτιμητών .

iv. Τέλος , είναι συνεπείς $\lim_{x \to \infty} \hat{b}_1 = \hat{b}_1$

Πρέπει να επισημάνουμε ότι ο καλύτερος αμερόληπτος εκτιμητής είναι αυτός με τη μικρότερη διακύμανση .

Οι διακυμάνσεις των συντελεστών $\hat{b}_0$ και $\hat{b}_1$ είναι συναρτήσεις της άγνωστης διακυμάνσεως (σ^2) του διαταρακτικού όρου (u) . Για να εκτιμηθούν οι διακυμάνσεις των συντελεστών θ πρέπει να έχουμε μία εκτίμηση για την (σ^2) , η οποία θα βασίζεται στα κατάλοιπα της γραμμής παλινδρομήσεως . Ένας αμερόληπτος εκτιμητής της s^2 δίνεται από τη σχέση : $s^2 = \dfrac{\sum y_i^2 - \hat{b}_1 \sum x_i y_i}{k-2}$ Η θετική τετραγωνική ρίζα της διακυμάνσεως s^2 ονομάζεται τυπικό σφάλμα εκτιμήσεως της y . Η σημασία του είναι ανάλογη προς τη σημασία του τυπικού σφάλματος ενός εκτιμητή .



Σχήμα 2

1.7 ΤΥΠΙΚΑ ΣΦΑΛΜΑΤΑ,ΟΙ ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ ΚΑΙ ΤΑ ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΕΝΟΣ ΔΕΙΓΜΑΤΟΣ.

Για να είμαστε σε θέση να γνωρίζουμε ότι εκτιμήθηκε σωστά ένα υπόδειγμα πρέπει να κάνουμε έναν έλεγχο υποδείγματος . Όπως γνωρίζουμε το τυπικό σφάλμα χρησιμοποιείται ως μέτρο ακρίβειας του εκτιμητή . Το τυπικό σφάλμα βάση της μεθόδου ελαχίστων τετραγώνων και της διασποράς των τιμών των εκτιμητών της y είναι της μορφής :

$$e.s.e = \left(\hat{b}_0\right) = \hat{s}\sqrt{\frac{1}{\sum\left(x_i - \overline{x}\right)^2}}$$

και

$$e.s.e\left(\hat{b}_1\right) = \hat{s}\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum\left(x_i - \overline{x}\right)^2}} \qquad \text{με} \qquad s^2 = \sqrt{\frac{RSS}{n-2}}$$

Όσο πιο μικρό το τυπικό σφάλμα τόσο πιο καλή η εκτίμηση της άγνωστης παραμέτρου και αντίστροφα .

Τα τυπικά σφάλματα είναι ο κύριος λόγος για να κατασκευάσουμε τα διαστήματα εμπιστοσύνης . Βάση εκείνων μπορούμε να υπολογίσουμε σε ποια όρια κινούνται οι άγνωστοι συντελεστές και κατ' επέκταση οι εκτιμητές . Όταν έχουμε στατιστικές που ακολουθούν την τυποποιημένη κανονική κατανομή αντικαθιστάμε με τα τυπικά σφάλματα των $\hat{b}_0$ και $\hat{b}_1$. Επειδή η διακύμανση σ^2 είναι άγνωστη , την αντικαθιστάμε με τον αμερόληπτο εκτιμητή της s^2 οπότε οι διακυμάνσεις των εκτιμητών μεταβάλλονται και τα διαστήματα εμπιστοσύνης είναι της εξής μορφής :

$$\Pr\left[-t_{a/2,n-k-1} \leq t \leq t_{a/2,n-k-1}\right] = 1-a$$

και σύμφωνα με τα παραπάνω γίνεται

$$\Pr\left[-t_{a/2,n-k-1} \leq \frac{\hat{b}_0 - \hat{b}_1}{e.s.e\left(\hat{b}_0\right)} \leq t_{a/2,n-k-1}\right] = 1-a$$

Για το $\hat{b}_0$ παίρνει τη μορφή :

$$\Pr[\hat{b}_0 - e.s.e\left(\hat{b}_0\right)t_{a/2,n-k-1} \leq b_0 \leq \hat{b}_0 + e.s.e\left(\hat{b}_0\right)t_{a/2,n-k-1}] = 1-a$$

Ομοίως για το $\hat{b}_1$:

$$\Pr\left[\hat{b}_1 - e.s.e\left(\hat{b}_1\right)t_{a/2,n-k-1} \leq b_1 \leq \hat{b}_1 + e.s.e\left(\hat{b}_1\right)t_{a/2,n-k-1}\right] = 1-a$$

Τις περισσότερες φορές όμως μας ενδιαφέρει να ελέγξουμε στατιστικά συγκεκριμένες υποθέσεις για τους συντελεστές β_0 και β_1 . Όταν αναλύουμε οικονομικά δεδομένο το ενδιαφέρον μας κινείται κυρίως γύρω από την ύπαρξη κάποιας σχέσης μεταξύ της εξαρτημένης και της ερμηνευτικής μεταβλητής. Αν δεν υπάρχει σχέση ανάμεσα στις μεταβλητές y και x , τότε η γραμμή παλινδρομήσεως περνάει από την αρχή των αξόνων , δηλαδή $\beta_0 = 0$ και φυσικά $R^2 = 0$. Η μηδενική υπόθεση που θέλουμε να ελέγξουμε , είναι :

$$H_0 : b_0 = 0$$

$$H_1 : b_0 \neq 0$$

Αυτή η υπόθεση ονομάζεται έλεγχος σημαντικότητας , γιατί ελέγχουμε αν ο συντελεστής β_0 είναι σημαντικά διαφορετικός από το μηδέν . Για τον έλεγχο αυτής της υποθέσεως χρησιμοποιούμε ως στατιστική ελέγχου την κατανομή t και εφόσον η μηδενική υπόθεση είναι $\beta_0 = 0$ τότε :

$$\frac{\hat{b}_0 - b_0}{e.s.e\left(\hat{b}_0\right)} \sim t_{a/2,n-k-1} \quad \text{και} \quad e.s.e\left(\hat{b}_0\right) = \hat{s}\sqrt{\frac{1}{\sum\left(x_i - \overline{x}\right)}}$$

Αυτός είναι και ο λόγος που χρησιμοποιούμε την t κατανομή , διότι ο λόγος μίας κανονικής και μίας x^2 κατανομής μας οδηγεί στην t .

Για α επίπεδο σημαντικότητας , η μηδενική υπόθεση απορρίπτεται αν :

$$|t| = \frac{\hat{b}_1}{s_{\hat{b}1}} \geq t_{T-2,a/2}$$

Η απόρριψη της μηδενικής υποθέσεως σημαίνει ότι υπάρχει σχέση ανάμεσα στην y και στην x . Άρα η μεταβλητή x είναι σημαντική στην ερμηνεία της συμπεριφοράς της y .

Ο έλεγχος υποθέσεως της $\beta_0 = 0$ μπορεί να γίνει επίσης και με την κατανομή F και τον πίνακα ANOVA . Μέχρι στιγμής έχουμε ορίσει τα TSS , RSS , ESS καθώς και την σχέση που τα συνδέει TSS = ESS + RSS . Ο πίνακας ANOVA μας δίνει πληροφοριακά στοιχεία για τα αθροίσματα τετραγώνων καθώς και πρόσθετη πληροφόρηση για τον έλεγχο στατιστικής σημαντικότητας του μοντέλου . Η στατιστική F είναι της μορφής :

$$F = \frac{\sum \hat{y}^2}{\sum \hat{u}^2 / (T-2)}$$

Η υπόθεση $\beta_0 = 0$ απορρίπτεται όταν $F \geq F_{a,v_1,v_2}$. Όπου F_a είναι η τιμή της F από τους πίνακες με 1(βαθμοί ελευθερίας αριθμητή) και T-2 (βαθμοί ελευθερίας παρονομαστή) βαθμούς ελευθερίας σε α επίπεδο σημαντικότητας . Να παρατηρηθεί ότι οι βαθμοί ελευθερίας είναι παρατηρήσεις που μεταβάλλονται ελεύθερα . Ακόμη , ότι η στατιστική F ουσιαστικά είναι το τετράγωνο της t .

Κατ' επέκταση τα παραπάνω μπορούν να συνοψισθούν στον παρακάτω πίνακα γνωστός ως πίνακας ANOVA , είναι μία ανάλυση της διακυμάνσεως για την απλή παλινδρόμηση .

ΠΙΝΑΚΑΣ ANOVA

Πηγή Μεταβλητότητας	Άθροισμα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσο Άθροισμα Τετραγώνων	F
Παλινδρόμηση	RSS	1	RSS/1	$\frac{RSS/1}{ESS/N-2}$
Υπόλοιπο	ESS	N-2	ESS/N-2	
Σύνολο	TSS	N-1	TSS/N-1	

Η στατιστική F χρησιμοποιείται για να ελέγξει τη σημαντικότητα του μοντέλου και απαντά στον εξής έλεγχο :

H_0 : όλοι οι εκτιμητές εκτός από τον σταθερό όρο να είναι 0

H_1 : διαφορετικά

Η κατανομή F δεν έχει δίπλευρο , άρα όταν $F \geq F_{a,v_1,v_2}$ απορρίπτεται η υπόθεση H_0 και ο εκτιμητής μας είναι στατιστικά σημαντικός .

1.8 ΜΕΘΟΔΟΣ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

Αν x είναι μία τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας f(x) και $(x_1,\ldots,x_n)$ ένα τυχαίο δείγμα , $f(x_1,\ldots, x_n)=f(x_1)\ldots f(x_n)=\prod_i f(x_i)$ ονομάζεται πιθανότητα ή πιθανοφάνεια του δείγματος και παριστάνεται με το σύμβολο L . Στο παράδειγμα μας , η συνάρτηση πιθανότητας f(x) της μεταβλητής x δεν εξαρτάται από καμιά παράμετρο , όπως ο μέσος , η διακύμανση κτλ. Για να τονίσουμε ότι η κατανομή μιας μεταβλητής είναι η συνάρτηση μίας ή περισσότερων παραμέτρων $\theta_1,\theta_2,\ldots,\theta_\kappa$ μπορούμε να την παραστήσουμε με $f(x| \theta_1,\theta_2,\ldots,\theta_\kappa)$. Εφόσον η x είναι μία τυχαία μεταβλητή με συνάρτηση πυκνότητας $f(x| \theta_1,\theta_2,\ldots,\theta_\kappa)$ και $(x_1,\ldots,x_n)$ είναι ένα τυχαίο δείγμα , η συνάρτηση : $L(x_1,\ldots,x_n| \theta_1,\theta_2,\ldots,\theta_\kappa)= f(x_1| \theta_1,\ldots,\theta_\kappa)\ldots f(x_n| \theta_1,\ldots,\theta_\kappa)= \prod_i f(x_i | q_1,\ldots,q_k)$ ονομάζεται συνάρτηση πιθανοφάνειας του δείγματος . Θα πρέπει να τονίσουμε ότι η συγκεκριμένη συνάρτηση δεν είναι συνάρτηση κατανομής πιθανοτήτων . Είναι συνάρτηση των παραμέτρων για δεδομένες τιμές των τυχαίων μεταβλητών .

Στο κανονικό γραμμικό υπόδειγμα οι διαταρακτικοί όροι είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κανονική κατανομή με μέσο το μηδέν και σταθερή διακύμανση $σ^2$. Οπότε , η συνάρτηση πυκνότητας του διαταρακτικού όρου είναι :

$$f\left(u_t\right)=\frac{1}{s\sqrt{2p}}e^{-\frac{1}{2}\left(\frac{u_t-0}{s}\right)^2} .$$

Κατά συνέπεια, η συνάρτηση πιθανοφάνειας του δείγματος είναι :

$$L=\frac{1}{\left(2ps^{2}\right)^{\frac{T}{2}}}e^{-\frac{1}{2}\frac{\sum u_t^2}{s^2}}$$

Όπου $u_t = y_t - β_0 - β_1 x_t$

Για να εκτιμήσουμε τους συντελεστές $β_0$ και $β_1$ θα χρησιμοποιήσουμε τη μέθοδο μέγιστης πιθανοφάνειας . Αντί να μεγιστοποιήσουμε τη σχέση , μεγιστοποιούμε το λογάριθμο της , βρίσκουμε μερικές παραγώγους ως προς $β_0$, $β_1$ και $σ^2$ και τις εξισώνουμε με το 0 . Οπότε έχουμε :

$$\sum y_t = Tb_0^* + b_1^*\sum x_t \quad \text{και} \quad \sum y_t x_t = b_0^*\sum x_t + b_1^*\sum x_t^2$$

Όπου b_0^* και b_1^* παριστάνουν τους μεγίστης πιθανοφάνειας εκτιμητές . Οι παραπάνω εξισώσεις είναι ίδιες με τις κανονικές εξισώσεις που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων . Η μεγιστοποίηση της σχέσης ως προς $β_0$ και $β_1$ σημαίνει ελαχιστοποίηση του τελευταίου όρου , αφού είναι αρνητικός . Δηλαδή , η ελαχιστοποίηση του αθροίσματος των τετραγώνων των αποκλίσεων . Επειδή συμπεραίνουμε ότι οι μεγίστης πιθανοφάνειας εκτιμητές είναι ίδιοι με τους εκτιμητές που προκύπτουν από τη μέθοδο ελαχίστων τετραγώνων ο εκτιμητής της διακυμάνσεως είναι : $s^{*2}=\frac{\sum \hat{u}^2}{T}$. Ο εκτιμητής s^{*2} όμως δεν είναι αμερόληπτος εκτιμητής της διακυμάνσεως , είναι όμως συνεπής και ασυμπτωτικά αποτελεσματικός . το ίδιο ισχύει και για τους εκτιμητές b_0^* και b_1^* .

1.9 ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Το υπόδειγμα της απλής γραμμικής αναφέρεται σε σχέσεις που περιλαμβάνουν μία μόνο ερμηνευτική μεταβλητή . Η συμπεριφορά όμως των περισσοτέρων οικονομικών μεταβλητών είναι η συνάρτηση όχι μίας αλλά πολλών μεταβλητών . Έστω μία συνάρτηση y με κ ερμηνευτικές μεταβλητές και ένα δείγμα με Τ παρατηρήσεις , τότε το υπόδειγμα μας είναι :

$$y_t = b_0 + b_1 x_{i_1} + b_2 x_{i_2} + \ldots\ldots + b_k x_{i_k} + e_i$$

Ουσιαστικά αυτή η σχέση αποτελεί το υπόδειγμα της γραμμικής πολυμεταβλητής παλινδρομήσεως που είναι κατ' επέκταση το μοντέλο της απλής παλινδρομήσεως για περισσότερες από μία μεταβλητές . Για το συγκεκριμένο υπόδειγμα πρέπει να λάβουμε υπόψη κάποιες βασικές υποθέσεις που πρέπει να ισχύουν για όλες τις παρατηρήσεις .

i. Τα υπόλοιπα ε_i είναι τυχαία μεταβλητή
ii. Ακολουθούν κανονική κατανομή $e_i \sim N(o, s^2)$, $Var(e_i) = s^2$ και $m = 0$.

iii. Δεν υπάρχουν ακριβείς γραμμικές σχέσεις ανάμεσα στις ανεξάρτητες μεταβλητές .
iv. Οι ερμηνευτικές μεταβλητές δεν είναι στοχαστικές . Οι τιμές τους παραμένουν σταθερές και δεν είναι όλες ίσες μεταξύ τους .
v. Το μέγεθος των παρατηρήσεων είναι μεγαλύτερο από τον αριθμό των ανεξάρτητων μεταβλητών (N > K) .
vi. Κάθε τιμή i της εξαρτημένης μεταβλητής είναι γραμμικός συνδυασμός των ανεξάρτητων μεταβλητών και του στοχαστικού όρου .

Ο υπολογισμός των εκτιμητών γίνεται με την μέθοδο ελαχίστων τετραγώνων . Δηλαδή με την ελαχιστοποίηση των τετραγώνων των υπολοίπων .

$$Min\sum_{i=1}^{T}\hat{e}_i^2 = Min\sum_{i=1}^{T}\left(\hat{y}_i - \hat{b}_0 - \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i} - .... - \hat{b}_k x_{ki}\right)$$

Επειδή η λύση του συστήματος είναι πολύπλοκη χρησιμοποιούμε πίνακες και καταλήγουμε στο εξής αποτέλεσμα :

$$\hat{b} = (x'x)^{-1}(x'y)$$

Για το γραμμικό υπόδειγμα είδαμε πως στην περίπτωση μιας ανεξάρτητης μεταβλητής , οι εκτιμητές που προκύπτουν με τη μέθοδο ελαχίστων τετραγώνων είναι άριστοι , γραμμικοί και αμερόληπτοι . Αυτό ισχύει και για το πολυμεταβλητό γραμμικό υπόδειγμα με Κ ανεξάρτητες μεταβλητές και οι διακυμάνσεις και συνδιακυμάνσεις δίνονται από τη σχέση

$V\left(\hat{b}\right) = s^{2}\left(x'x\right)^{-1}$, οπότε έχουμε την γενική μορφή του θεωρήματος των Gauss-Markov .

1.10 ΕΚΤΙΜΗΣΗ ΠΟΛΥΜΕΤΑΒΛΗΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

Στην περίπτωση του πολλαπλού γραμμικού υποδείγματος για το στατιστικό έλεγχο , όπως και στην απλή παλινδρόμηση , θα υποθέσουμε ότι ο διαταρακτικός όρος είναι μία τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή . Γνωρίζουμε ότι $e_i \sim N(0, s^{2})$. Με τη σειρά τους και με βάση το πολλαπλό γραμμικό υπόδειγμα και οι εκτιμητές $\hat{b}_0, \hat{b}_1, \ldots.., \hat{b}_k$ ακολουθούν την κανονική κατανομή . Στην περίπτωση των πινάκων η στατιστική συνάρτηση διαμορφώνεται ως εξής :

$$\frac{\hat{b}_i - b_i}{s\sqrt{cij}} \sim N(x'x)^{-1}$$

Όπου σ αντικαθιστάται από μία εκτίμηση $\hat{s} = \sqrt{\dfrac{\sum e_i}{n-k-1}}$ και cij είναι τι διαγώνιο στοιχείο του . Άρα ο έλεγχος είναι $t_{\hat{b}_i} = \dfrac{\hat{b}_i - b_i}{e.s.e\left(\hat{b}_i\right)} \sim t_{a/2, n-k-1}$.

Στην περίπτωση που θέλουμε να συγκρίνουμε ένα υπόδειγμα της μορφής $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots\ldots + b_k x_{kj} + e_i$ (μη περιορισμένο υπόδειγμα Ι) με το εξής $y_i = b_0 + b_1 x_{1i} + \ldots\ldots + b_{k-2} x_{k-2} + e_i$ (

περιορισμένο υπόδειγμα ΙΙ) πρέπει να εκτελέσουμε τον παρακάτω έλεγχο :

$H_0 : \hat{b}_2 = \hat{b}_1 = 0$ (το ΙΙ είναι κατάλληλο)

$H_1 =$ διαφορετικά (το Ι είναι κατάλληλο)

Πρέπει να επισημάνουμε ότι το διάστημα εμπιστοσύνης στο πολυμεταβλητό υπόδειγμα είναι της μορφής :

$\hat{b}_i - e.s.e\left(\hat{b}_i\right) t_{a/2, n-k-1} \leq b_i \leq \hat{b}_i + e.s.e\left(\hat{b}_i\right) t_{a/2, n-k-1}$ για κάθε εκτιμητή .

Στο πολλαπλό υπόδειγμα όπως είπαμε παραπάνω , όλοι οι υπολογισμοί για τη χρήση όλων των υπολοίπων παραμέτρων γίνονται με μορφή πινάκων .Για παράδειγμα στην ανάλυση διακύμανσης θα έχουμε ότι :

$$RSS = \hat{b}\,'x'y - n\overline{y}^2$$
$$ESS = e_i\,'e_i = y'y - \hat{b}\,'x'y$$
$$TSS = \sum y_i^2 = y'y - n\overline{y}^2$$

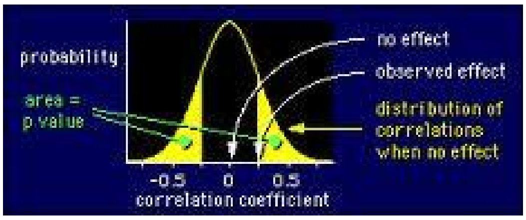
Άρα και στον πίνακα ANOVA , που εκφράζει την παλινδρόμηση του υποδείγματος μέσω της στατιστικής F , θα υπάρχει μία διαφορά στους βαθμούς ελευθερίας που θα καθορίζεται από τον αριθμό των παρατηρήσεων του υποδείγματος μας .

Η στατιστική F χρησιμοποιείται για τον έλεγχο καταλληλότητας του υποδείγματος , όποτε σε ένα υπόδειγμα της μορφής $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + ... + b_k x_{kj} + e_i$, ο έλεγχος είναι

$$H_0 : \hat{b}_2 = \hat{b}_1 = ... = \hat{b}_k = 0$$

H_1: διαφορετικά

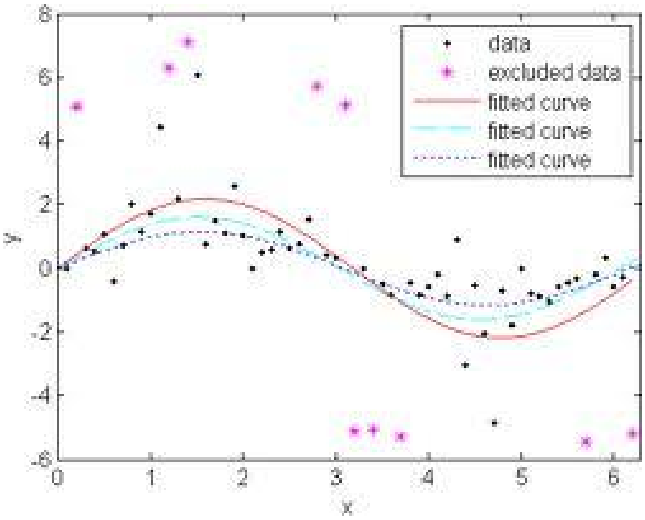
Η $F = \frac{(R_u^2 - R_r^2)/r}{\left(1-R_u^2\right)/N-k} \square F_{r,N-k,a}$ συγκρίνεται με το αποτέλεσμα από τον πίνακα ANOVA για να πραγματοποιήσουμε τον έλεγχο. Όπου r είναι ο αριθμός των περιορισμών , N το μέγεθος του δείγματος και k ο αριθμός των ανεξάρτητων μεταβλητών του μη περιορισμένου υποδείγματος .



Σχήμα 3: Διάγραμμα ελέγχου F στατιστικής

1.11 ΑΚΡΑΙΕΣ ΤΙΜΕΣ ΚΑΙ ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΠΑΛΙΝΔΡΟΜΗΣΕΩΝ

Μερικές φορές οι τιμές ορισμένων παρατηρήσεων ενός δεδομένου δείγματος διαφέρουν ή απέχουν τόσο πολύ από τις τιμές των υπολοίπων παρατηρήσεων ώστε επηρεάζουν σημαντικά τις εκτιμήσεις των συντελεστών του υποδείγματος . Στην περίπτωση του απλού υποδείγματος , οι παρατηρήσεις αυτές , που μπορούν να χαρακτηριστούν ως ακραίες , μπορούν εύκολα να εντοπιστούν από το διάγραμμα διασποράς . Επειδή ο εντοπισμός τους όμως δεν είναι εφικτός με αυτόν τον τρόπο για $k \geq 2$, απαιτείται εξέταση των καταλοίπων . Τα διαγράμματα διασποράς δίνουν διαφορετική εικόνα από αυτή που θα μας δώσει η εκτιμημένη γραμμή παλινδρόμησης . Η ύπαρξη ακραίων τιμών πρέπει να προβληματίζει , αλλά η λύση δεν μπορεί να είναι πάντοτε η παράλειψη τους από το δείγμα . Απαιτείται περαιτέρω εξέταση των συνθηκών που χαρακτηρίζουν το οικονομικό φαινόμενο .



Σχήμα 3

Σε προηγούμενη ενότητα μιλήσαμε για τον διορθωμένο συντελεστή $\bar{R}^2$ και για την καταλληλότητα του ως συντελεστή στην σύγκριση της ερμηνευτικής ικανότητας δύο ή περισσοτέρων παλινδρομήσεων που διαφέρουν ως προς των αριθμό των παραμέτρων ή και το μέγεθος του δείγματος . Εκτός όμως από τον διορθωμένο συντελεστή προσδιορισμού υπάρχουν και άλλα κριτήρια .

Τα κυριότερα από αυτά είναι το κριτήριο πληροφοριών Akaike ή AIC και το Μπεϋεσιανό κριτήριο Schwarz ή SBC . Τα κριτήρια αυτά δίνονται από τις ακόλουθες σχέσεις :

$$AIC = \log \frac{\sum \hat{e}_i^2}{N} + \frac{2k}{N} \quad \text{και} \quad SBC = \log \frac{\sum \hat{e}_i^2}{N} + \frac{k \log N}{N}$$

Με βάση τα παραπάνω κριτήρια επιλέγεται ως καταλληλότερο το υπόδειγμα με την μικρότερη τιμή .

ΚΕΦΑΛΑΙΟ 2

ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

2.1 ΕΙΣΑΓΩΓΗ

Στο προηγούμενο κεφάλαιο εξετάσαμε συναρτησιακές σχέσεις ανάμεσα σε δύο η περισσότερες μεταβλητές οι οποίες είναι γραμμικές . Η μορφή όμως της συναρτησιακής σχέσης ανάμεσα σε δύο ή περισσότερες οικονομικές μεταβλητές μπορεί να μην είναι γραμμική . Μια τέτοια κατάσταση μπορεί να εκρέει είτε από την οικονομική θεωρία είτε από τα εμπειρικά δεδομένα. Π.χ. οι συναρτήσεις κόστους δεν είναι γραμμικές όπως αυτό προκύπτει από την οικονομική θεωρία.Το ζήτημα λοιπόν είναι αν οι μέθοδοι εκτιμήσεως και η ανάλυση του γραμμικού μοντέλου μπορούν να εφαρμοσθούν σε σχέσεις που δεν ειναι γραμμικές .Εξ' αιτίας της όχι τόσο περιοριστικής ανάλυσης της γραμμικής παλινδρόμησης , πολλές σχέσεις που δεν είναι γραμμικές μπορούν εύκολα να μετασχηματιστούν σε γραμμικές , οπότε οι μέθοδοι εκτιμήσεως του γραμμικού υποδείγματος μπορούν να εφαρμοστούν στις μετασχηματισμένες αλλά γραμμικές πια σχέσεις .

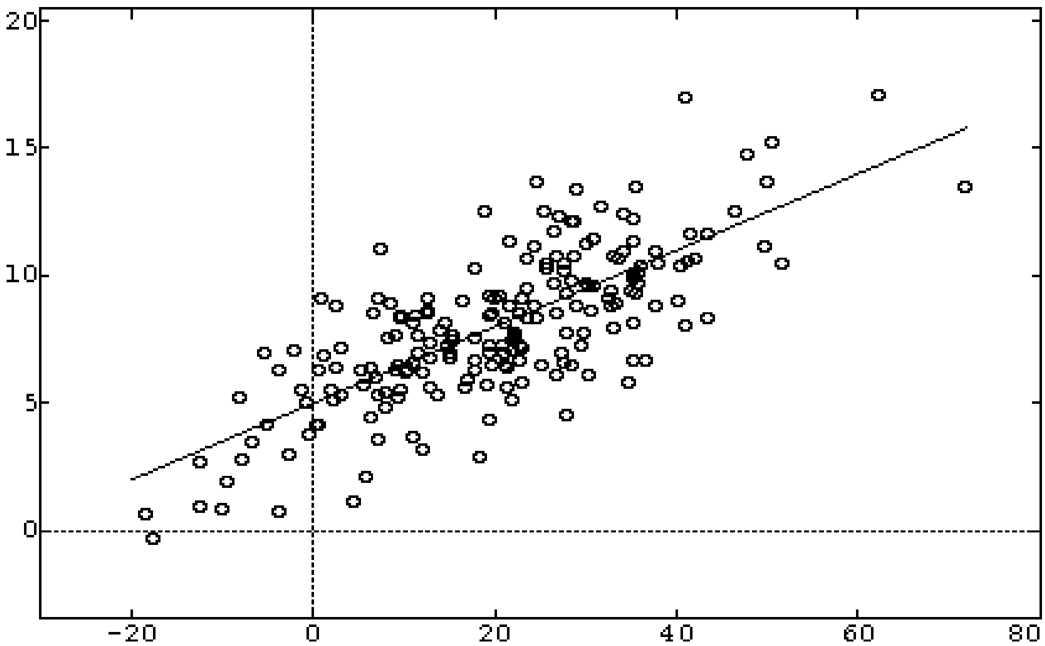
Όταν οδηγούμαστε σε ένα μοντέλο μη γραμμικής μορφής , προτιμούμε να προσαρμόσουμε ενα τέτοιο μη γραμμικό μοντέλο, οπότε αυτό είναι δυνατόν, παρά να προσαρμόσουμε ενα εναλακτικό ,λιγότερο ρεαλιστικό γραμμικό μοντέλο.Κάθε μοντέλο που **δεν** είναι της μορφής :

$$\mathbf{Y = \alpha + \beta\, X + \varepsilon}$$

θα καλείται μη γραμμικό μοντέλο με την έννοια οτι <u>δεν είναι γραμμικό ως πρός τις παραμέτρους.</u>

2.1.1 Η ΕΝΝΟΙΑ ΤΗΣ ΓΡΑΜΜΙΚΟΤΗΤΑΣ

Η συνάρτηση **Y = f(X)** είναι γραμμική αν μπορεί να παρασταθεί με ευθεία γραμμή , που σημαίνει ότι η κλίση της ,δηλαδή η <u>πρώτη παράγωγος είναι σταθερή</u> . Πιο αναλυτικά η συνάρτηση **Y= f(X)** ειναι γραμμική ως προς το Χ αν και μονο αν η πρώτη παράγωγος δεν εξαρτάται απο το **X** που συνεπάγεται η <u>δεύτερη παράγωγος να είναι ίση με το μηδέν</u> .
Για την εκτίμηση όμως μεγαλύτερη σημασία έχει αν η συνάρτηση είναι <u>γραμμική ως προς τις παραμέτρους</u> .Για να συμβαίνει αυτό θα πρέπει οι παράμετροι να μην είναι συναρτήσεις γινομένων ή κλασμάτων ούτε να εμφανίζονται ως δυνάμεις .
Όπως είπαμε και παραπάνω πολλές σχέσεις που δεν είναι γραμμικές μπορούν να μετασχηματιστούν εύκολα σε γραμμικές ,στις οποίες μπορεί να εφαρμοστεί η μέθοδος ανάλυσης της γραμμικής παλινδρόμησης .



Σχήμα 4:Γραμμική παλινδρόμηση με θετική συσχέτιση ανάμεσα στις δύο μεταβλητές .

2.2 Η ΜΕΘΟΔΟΣ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΟΝΩΝ ΣΤΗΝ ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (LEAST SQUARES ESTIMATION IN NONLINEAR REGRESSION)

Όπως γνωρίζουμε με την μέθοδο των ελαχίστων τετραγώνων ,επιλέγουμε εκείνη την γραμμή για την οποία το άθροισμα των τετραγώνων των αποκλίσεων,των παρατηρήσεων της Υ απο την γραμμή παλινδρομήσεως του δείγματος είναι ελάχιστο .Στις περιπτώσεις τώρα των μη γραμμικών ελαχίστων τετραγώνων ο τυπικός συμβολισμός είναι διαφορετικός από τις περιπτώσεις των γραμμικών . Στον παρακάτω πίνακα φαίνονται οι διαφορές αυτές .

	Γραμμικα	Μη Γραμμικα
Αποκριση	Y	Y
Υποδεικτες των παρατηρησεων	$q_1, q_2, \ldots, q_p$	$u = 1, 2, \ldots, n$
Ανεξαρτητες μεταβλητες	$X_1, \mathbf{K}, X_n$	$x_1, \mathbf{K}, x_n$
Παράμετροι	$b_1, b_2, \ldots, b_p$	$q_1, q_2, \ldots, q_p$

2.2.1 ΠΡΟΒΛΗΜΑ ΤΩΝ ΜΗ ΓΡΑΜΜΙΚΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Ο αντικειμενικός σκοπός περιλαμβάνει την προσαρμογή των παραμέτρων του μοντέλου , έτσι ώστε να ταιριάζουν καλύτερα στο σύνολο των δεδομένων. Ένα απλό σύνολο δεδομένων αποτελείται από n σημεία (ζεύγη δεδομένων) (x_i, y_i) , I = 1, ..., n, όπου x_i είναι μια ανεξάρτητη μεταβλητή και y_i είναι μια εξαρτημένη μεταβλητή. Η συνάρτηση του μοντέλου έχει τη μορφή $f(x, \boldsymbol{\beta})$. Οι τιμές των παραμέτρων για τις οποίες τα δεδομένα ταιριάζουν καλύτερα στο μοντέλο ,είναι και αυτές που ψάχνουμε . Η μέθοδος των ελαχίστων τετραγώνων βρίσκει την βέλτιστη επιλογή όταν το άθροισμα (S) των τετραγώνων των καταλοίπων είναι ελάχιστο .

$$\boxed{S = \sum_{i=1}^{n} {r_i}^2}$$

Ως κατάλοιπα ορίζουμε την διαφορά μεταξύ της τιμής της εξαρτημένης μεταβλητής και της προβλεπόμενης τιμής από το εκτιμώμενο μοντέλο .

$$\boxed{r_i = y_i - f(x_i, \hat{\boldsymbol{\beta}}),}$$

Ένα παράδειγμα ενός μοντέλου είναι αυτό της ευθείας γραμμής . Με σημείο τομής β_0 και κλίση β_1 . Η συνάρτηση του μοντέλου δίνεται από:

$$\boxed{f(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x.}$$

<u>Ένα σημείο δεδομένων μπορεί να αποτελείται από μία ή περισσότερες ανεξάρτητες μεταβλητές</u> .

2.2.2 Η ΕΠΙΛΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Τα προβλήματα των ελαχίστων τετραγώνων εμπίπτουν σε δύο κατηγορίες, γραμμικές και μη γραμμικές.Το πρόβλημα των γραμμικών ελαχίστων τετραγώνων έχει επιλυθεί , αλλά το πρόβλημα των μη γραμμικών ελαχίστων τετραγώνων δεν έχει λυθεί μόνιμα ,έτσι συχνά επιλύεται μέσω της επαναληπτικής ευγενοποίησης .Σε κάθε επανάληψη το σύστημα προσεγγίζει ένα γραμμικό, συνεπώς , ο υπολογισμός είναι παρόμοιος και στις δύο περιπτώσεις.

Το ελάχιστο άθροισμα των τετραγώνων το βρίσκουμε θέτοντας την πρώτη παράγωγο(κλίση) στο μηδέν. Δεδομένου ότι το μοντέλο περιέχει παραμέτρους m , υπάρχουν m εξισώσεις κλίσης.

$$\boxed{\frac{\partial S}{\partial \beta_j} = 2\sum_i r_i \frac{\partial r_i}{\partial \beta_j} = 0,\ j = 1, \ldots, m}$$

και δεδομένου $r_i = y_i - f(x_i, \boldsymbol{\beta})$ οι εξισώσεις κλίσης γίνονται

$$\boxed{-2\sum_i \frac{\partial f(x_i, \boldsymbol{\beta})}{\partial \beta_j} r_i = 0,\ j = 1, \ldots, m}$$

Οι εξισώσεις κλίσης εφαρμόζονται σε όλα τα προβλήματα ελαχίστων τετραγώνων. Κάθε συγκεκριμένο πρόβλημα απαιτεί συγκεκριμένες εκφράσεις για το μοντέλο και την μερική του παράγωγο.

Γραμμικά ελαχίστα τετράγωνα

Ένα μοντέλο παλινδρόμησης είναι γραμμικό όταν αποτελείται από ένα γραμμικό συνδυασμό παραμέτρων, δηλαδή

$$\boxed{f(x_i, \beta) = \sum_{j=1}^{m} \beta_j \phi_j(x_i)}$$

όπου οι συντελεστές, φ ι, είναι συναρτήσεις του x i.

Μη-γραμμικά ελάχιστα τετράγωνα

Δεν υπάρχει δεδομένη λύση σε ενα μη-γραμμικό πρόβλημα ελαχίστων τετραγώνων. Αντ 'αυτού, χρησιμοποιούνται αριθμητικοί αλγόριθμοι για να βρεθεί την τιμή των παραμέτρων που ελαχιστοποιούν το στόχο. Οι περισσότεροι αλγόριθμοι περιλαμβάνουν αρχικές τιμές για τις παραμέτρους. Στη συνέχεια, οι παράμετροι εξευγενίζονται επαναληπτικά, δηλαδή, οι τιμές λαμβάνονται από τη διαδοχική προσέγγιση.

$$\boxed{{\beta_j}^{k+1} = {\beta_j}^{k} + \Delta\beta_j}$$

(K) είναι ένας αριθμός επανάληψεων και το διάνυσμα των αυξήσεων, $\Delta\beta_j$ είναι γνωστό ως (shift vector). Σε καποιους ευρέως χρησημοποιούμενους αλγόριθμους , σε κάθε επανάληψη του μοντέλου μπορεί να γραμμικοποιεί κατά προσέγγιση μια πρώτη επέκταση Taylor για περίπου $\boldsymbol{\beta}^k$

$$\boxed{\begin{aligned} f(x_i, \boldsymbol{\beta}) &= f^k(x_i, \boldsymbol{\beta}) + \sum_j \frac{\partial f(x_i, \boldsymbol{\beta})}{\partial \beta_j} \left(\beta_j - {\beta_j}^k\right) \\ &= f^k(x_i, \boldsymbol{\beta}) + \sum_j J_{ij} \Delta\beta_j . \end{aligned}}$$

Η Jacobean, J, είναι συνάρτηση σταθερών ορων , η ανεξάρτητη μεταβλητή και οι παραμέτροι, αλλάζουν από τη μια επανάληψη στην επόμενη. Τα υπολείμματα βρίσκονται απο τον τύπο:

$$\boxed{r_i = y_i - f^k(x_i, \boldsymbol{\beta}) - \sum_{j=1}^{m} J_{ij} \Delta\beta_j = \Delta y_i - \sum_{j=1}^{m} J_{ij} \Delta\beta_j} .$$

Για να ελαχιστοποιηθεί το άθροισμα των τετραγώνων των καταλοίπων r $_i$, η εξίσωση της κλίσης τίθεται στο μηδέν και επιλύεται ως προς $\Delta\beta_j$

$$-2\sum_{i=1}^{n} J_{ij}\left(\Delta y_i - \sum_{j=1}^{m} J_{ij}\Delta\beta_j\right) = 0$$

όπου με ανακατατάξεις , γεννά m ταυτόχρονες γραμμικές εξισώσεις, τις κανονικές εξισώσεις(normal equations).

$$\sum_{i=1}^{n}\sum_{k=1}^{m} J_{ij}J_{ik}\Delta\beta_k = \sum_{i=1}^{n} J_{ij}\Delta y_i \qquad (j = 1, \ldots, m)$$

2.2.3 ΔΙΑΦΟΡΕΣ ΜΕΤΑΞΥ ΓΡΑΜΜΙΚΩΝ ΚΑΙ ΜΗ ΓΡΑΜΜΙΚΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Το συναρτησιακό μοντέλο, f, σε LLSQ (γραμμική μεθοδος ελαχίστων τετραγώνων) είναι ένας γραμμικός συνδυασμός των παραμέτρων της μορφής :

$$f = X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots$$

i. Το μοντέλο μπορεί να αντιπροσωπεύει μια ευθεία γραμμή, μια παραβολή ή οποιοδήποτε άλλη πολυωνυμική συναρτηση . Σε NLLSQ (μη-γραμμική μέθοδος ελαχίστων τετραγώνων) τα στοιχεία εμφανίζονται ως συναρτήσεις, όπως η β $^{2,\ \beta}$ e ^{x} και ούτω καθεξής. Εάν η παράγωγος $\partial f / \partial \beta_j$ είναι είτε σταθερή είτε εξαρτάται μόνο από τις τιμές των ανεξάρτητων μεταβλητών, το μοντέλο είναι γραμμικό στις παραμέτρους. Διαφορετικά, το μοντέλο είναι μη γραμμικό.

ii. Πολλοί αλγόριθμοι για να επιλύσουν NLLSQ απαιτούν αρχικές τιμές για τις παραμέτρους, η LLSQ δεν απαιτεί.

iii. Πολλοί αλγόριθμοι για να επιλύσουν NLLSQ απαιτούν το Jacobian να έχει υπολογιστεί. Οι αναλυτικές εκφράσεις για τη μερική παραγώγο μπορεί να είναι περίπλοκες . Αν οι αναλυτικές εκφράσεις είναι αδύνατο να επιλύσουν τη μερική παραγώγο πρέπει αυτή να υπολογιστεί με αριθμητική προσέγγιση.

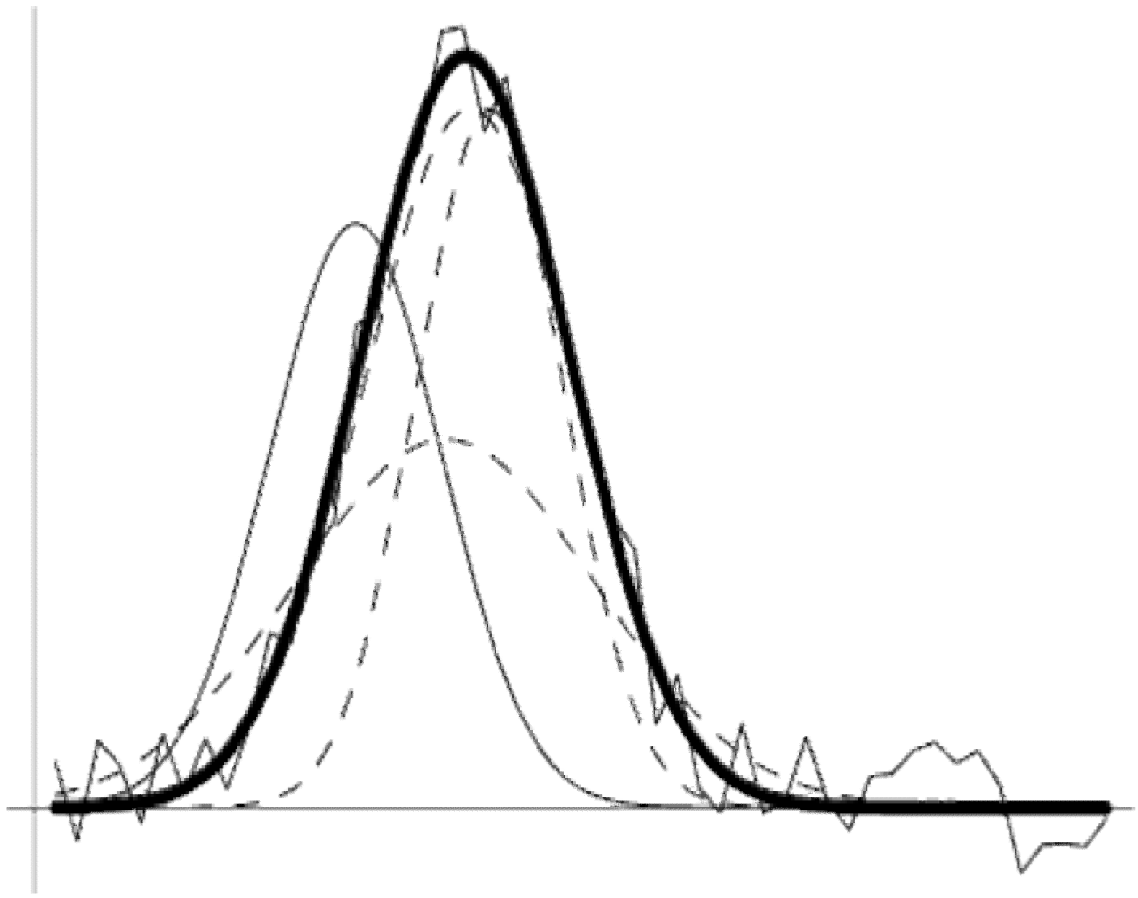
iv. Σε NLLSQ η αδυναμία του αλγορίθμου να βρεί ένα ελάχιστο είναι συνηθισμένο φαινόμενο ενώ στην LLSQ δεν είναι.

v. Η μέθοδος NLLSQ είναι συνήθως μια επαναληπτική διαδικασία. Η επαναληπτική διαδικασία πρέπει να διακοπεί όταν ένα κριτήριο σύγκλισης είναι ικανοποιητικό. Οι λύσεις της LLSQ υπολογίζονται με άμεσες μεθόδους, αν και τα προβλήματα με μεγάλους αριθμούς παραμέτρων επιλύονται συνήθως με επαναληπτικές μέθοδους, όπως η Gauss-Seidel.

vi. Σε LLSQ η λύση είναι μοναδική, αλλά σε NLLSQ μπορεί να υπάρχουν πολλαπλά ελάχιστα στο άθροισμα των τετραγώνων.

vii. Υπό την προϋπόθεση ότι τα κατάλοιπα είναι ασυσχέτιστα με τις μεταβλητές πρόβλεψης, η LLSQ αποδίδει αμερόληπτες εκτιμήσεις, αλλά ακόμη και υπό αυτόν τον όρο οι εκτιμήσεις της NLLSQ δεν ειναι αμερόληπτες .

Οι διαφορές αυτές πρέπει λαμβάνονται υπ'οψην κάθε φορά που επιδιώκεται η λύση σε ενα NLLSQ πρόβλημα .



Σχημα 2 : NON LINEAR LEAST SQUARES

2.3 ΜΗ ΓΡΑΜΜΙΚΕΣ ΜΟΡΦΕΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

2.3.1 ΠΟΛΥΩΝΥΜΙΚΗ ΜΟΡΦΗ

Έστω το υπόδειγμα:

$$\boxed{Y = b_0 + b_1 X + b_2 X^2 + u} \qquad (1)$$

Που μπορεί να χρησιμοποιηθεί για την περιγραφή της καμπύλης του οριακού κόστους ,όπου Υ= οριακό κόστος και Χ= προιόν. .Το υπόδειγμα (1) μετατρέπεται εύκολα σε γραμμικό αν θέσουμε $X = X_1^*$ και $X^2 = X_2^*$

Άρα η (1) γίνεται : $\boxed{Y = b_0 + b_1 X_1^* + b_2 X_2^* + u}$ (2)

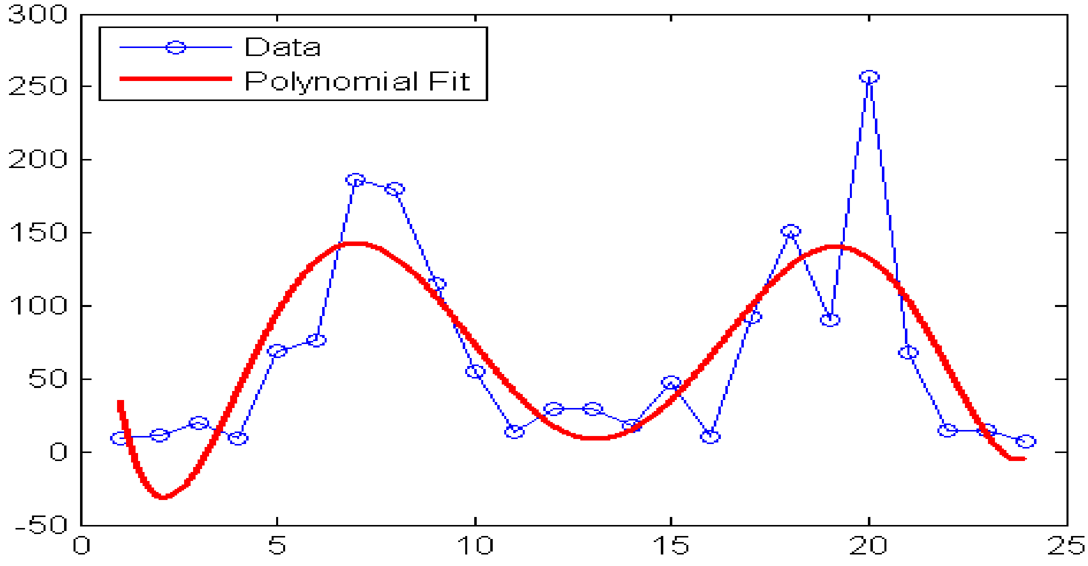
Και το πιο γενικό υπόδειγμα :

$\boxed{Y = b_0 + b_1 X + b_2 X^2 + ... + b_k X^k + u}$ (3)μετατρέπεται σε γραμμικό αν θέσουμε όπου $X^s = X_s^*$ για s=1,2,.....,k, οπότε έχουμε

$$\boxed{Y = b_0 + b_1 X_1^* + b_2 X_2^* + ... + b_k X_k^* + u} \quad (4)$$

Επομένως για να εκτιμήσουμε τους συντελεστές του (3) εφαρμόζουμε τις μεθόδους εκτίμησης της γραμμικής παλινδρόμησης στο (4) αφού έχουν κοινούς συντελεστές .

Το υπόδειγμα (4) μπορούμε να το χρησημοποιήσουμε για τον έλεγχο της γραμμικότητας της συνάρτησης .

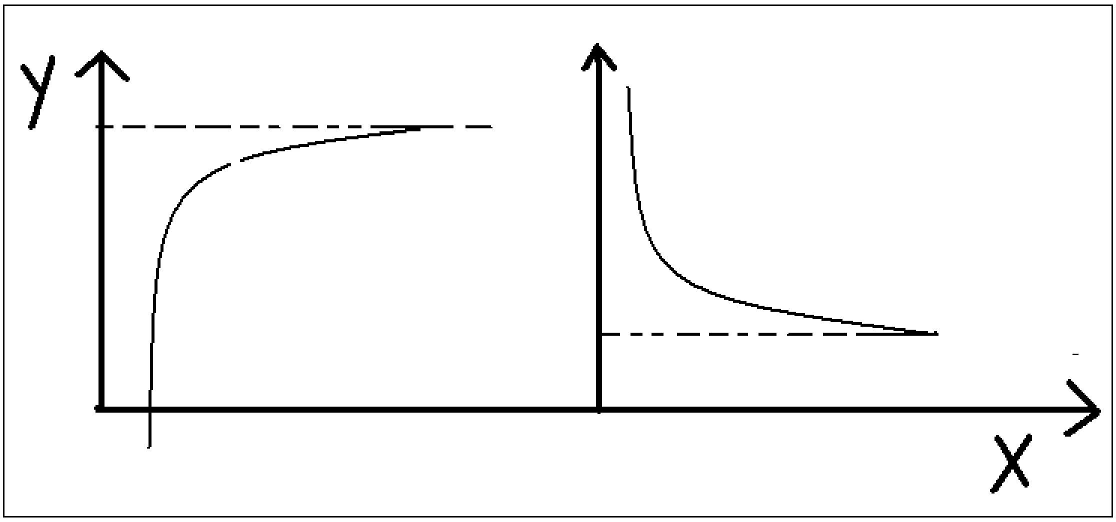


2.3.2 ΑΝΤΙΣΤΡΟΦΗ ΜΟΡΦΗ

Έστω το υπόδειγμα:

$$\boxed{Y = b_0 + b_1 \frac{1}{\mathrm{X}_t} + u_t} \qquad (1)$$

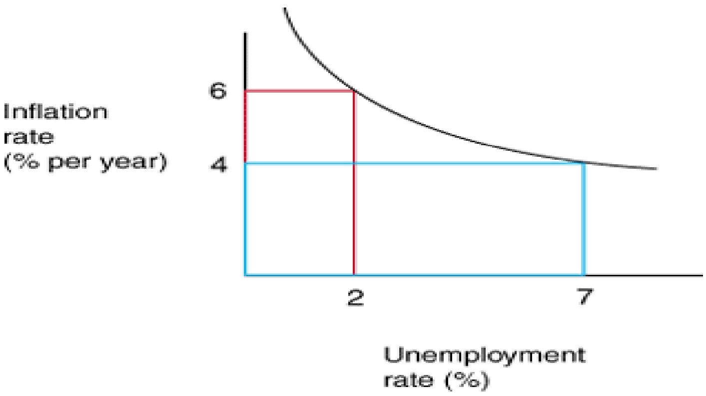
Το υπόδειγμα (1) είναι χρήσιμο για την ανάλυση οικονομικών φαινομένων ,όπου η μέση τιμή της μεταβλητής Y τείνει προς ένα ασυμπτωτικό επίπεδο όταν η τιμή της μεταβλητής X αυξάνεται .O συντελεστής b_0 είναι το ανώτατο όριο της μέσης τιμής της Y αν b_1 **p** 0 και το κατώτατο όριο αν b_1 **f** 0



(1) (2)

Καμπυλη Phillips(1) :

Η καμπύλη αυτή περιγράφει την αντίστροφη σχέση μεταξύ του ρυθμού μεταβολής των μισθών και του επιπέδου της ανεργίας .

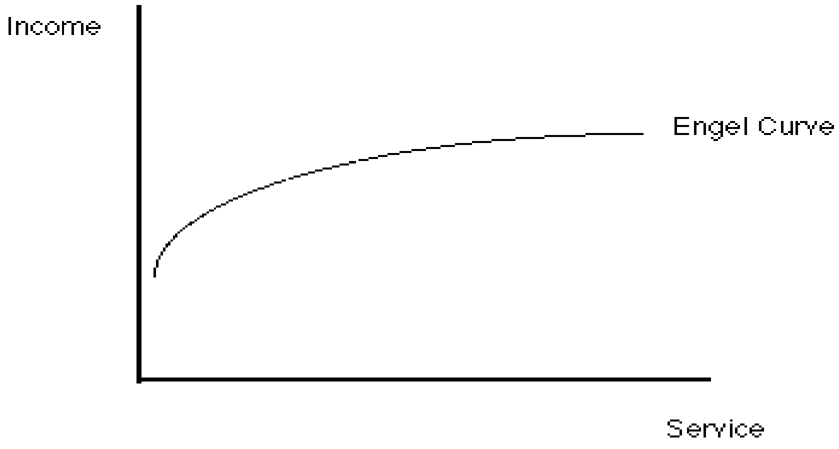


Καμπυλη Engel(2) :

Η καμπύλη αυτή περιγράφει την σχέση ανάμεσα στις δαπάνες για συγκεκριμένα αγαθά ή κατηγορίες αγαθών και στο εισόδημα.

Figure 4.31

Engel Curve for Service



Το υπόδειγμα $\boxed{Y = b_0 + b_1 \frac{1}{X_t} + u_t}$ μετατρέπεται εύκολα σε γραμμικό,αν θέσουμε όπου 1/X = X* ,άρα έχουμε:

$$\boxed{Y = b_0 + b_1 X_t^* + u_t}$$

Όπως και στην πολυωνυμική μορφή για να εκτιμήσουμε τους συντελεστές b_0 και b_1 εφαρμόζουμε τις μεθόδους εκτιμήσεως της γραμμικής παλινδρόμησης .

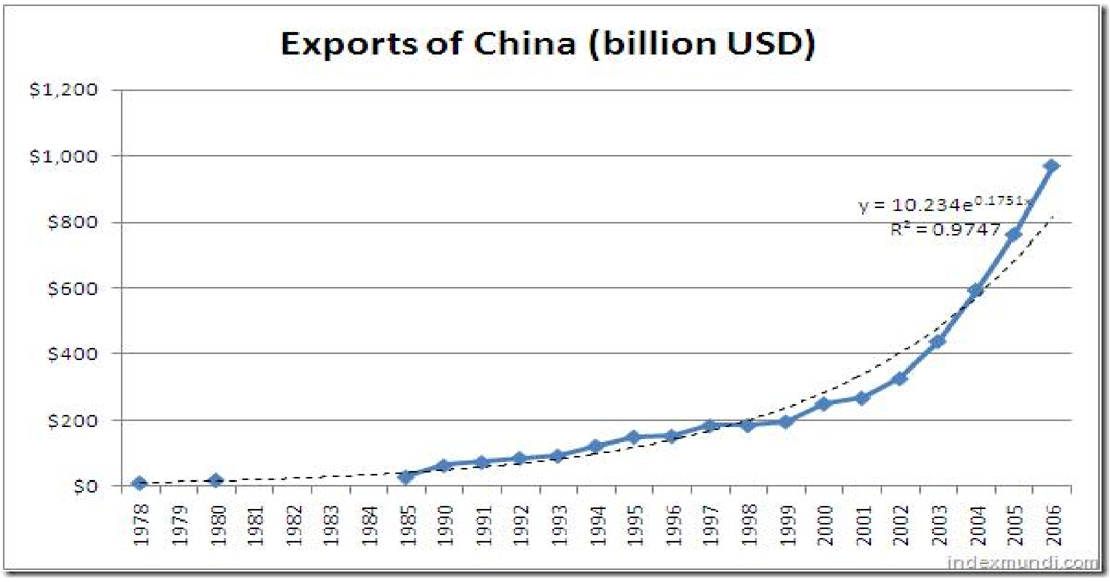
2.3.3. ΕΚΘΕΤΙΚΗ ΜΟΡΦΗ

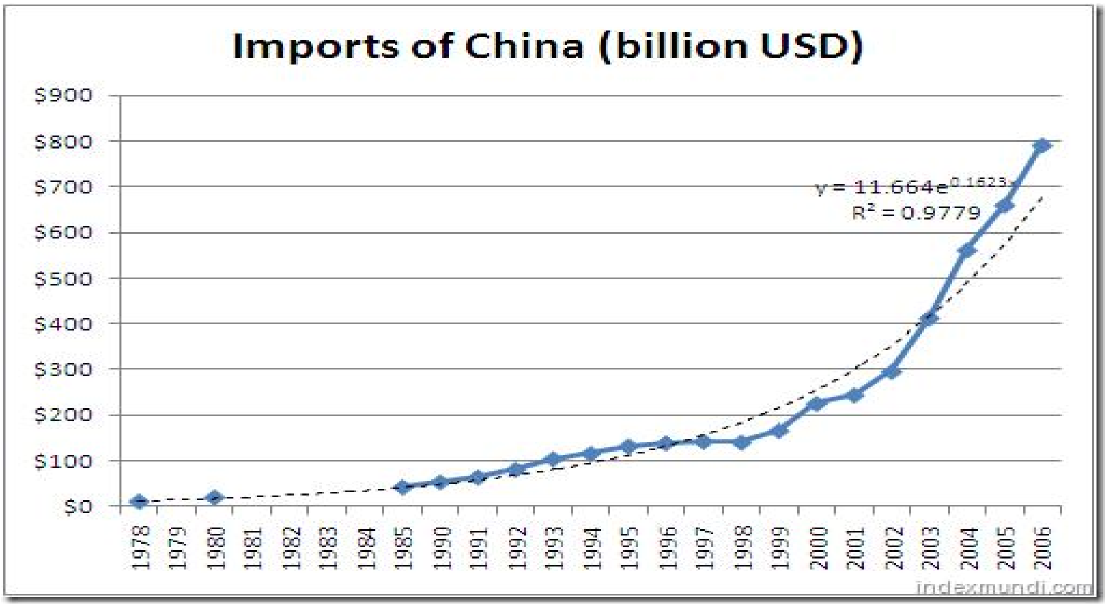
Έστω η συνάρτηση:

$$\boxed{\mathrm{Y} = e^{b_0 + b_1 \mathrm{X} + u}} \quad (1)$$

Η συνάρτηση (1) που είναι γνωστή ως εκθετική συνάρτηση (exponential function) , μετατρέπεται εύκολα σε γραμμική αν χρησημοποιήσουμε τον λογαριθμικό μετασχηματισμό ώς προς βάση το e:

$$\boxed{\log Y = b_0 + b_1 \mathrm{X} + u} \quad (2)$$





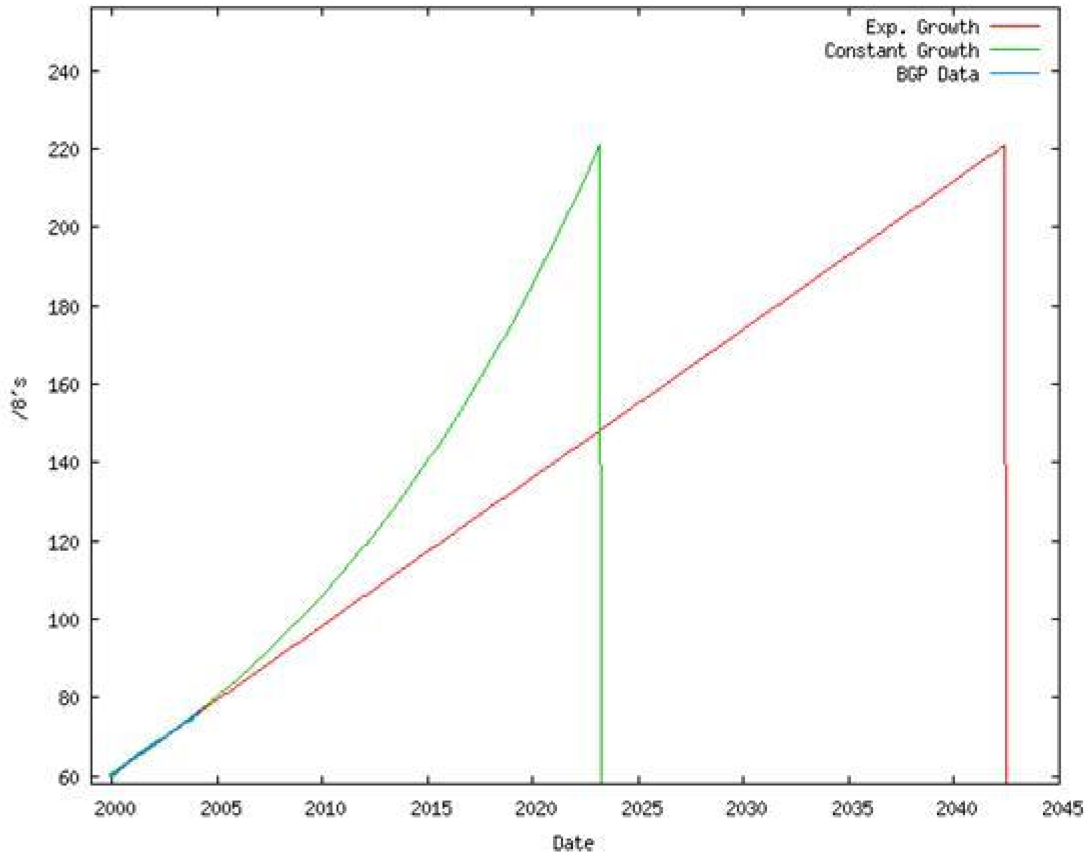
2.3.3.1. ΥΠΟΔΕΙΓΜΑ ΣΤΑΘΕΡΗΣ ΜΕΓΕΝΘΥΣΗΣ

Μια άλλη διατύπωση της εκθετικής συνάρτησης είναι η εξής :

$$\boxed{Y = b_0 e^{b_1 \mathrm{X} + u}} \quad (3)$$ που γίνεται

$$\boxed{\log Y = \log b_0 + b_1 \mathrm{X} + u} \quad (4)$$

Η παραπάνω συνάρτηση για Χ= t (χρόνος) ,έχει χρησιμοποιηθεί για την μελέτη αυτών των μεταβλητών που περιγράφουν η υποδηλώνουν την πρόοδο της επιστήμης , όπως π.χ. ο αριθμός των επιστημόνων (dependent).Σ'αυτή την περίπτωση , b_0 παριστάνει την μεταβλητή Υ (αριθμό επιστημονων) στην αρχή της περιόδου .Βασικό χαρακτηριστικό της προηγούμενης συνάρτησης είναι οτι συνεπάγεται σταθερό ρυθμό μεταβολής της μεταβλητής Υ ,για αυτό και το υπόδειγμα (4) ειναι γνωστό ως υπόδειγμα σταθερής μεγένθυσης (constant growth model).

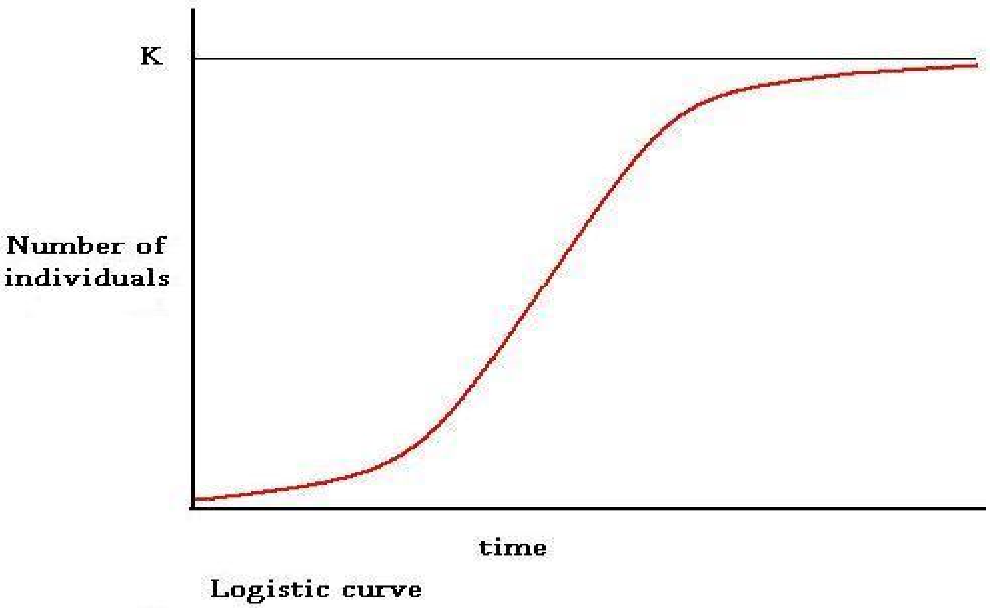


2.3.4ΛΟΓΙΣΤΙΚΗ ΚΑΜΠΥΛΗ

Έστω η συνάρτηση :

$$\boxed{Y_t = \frac{g}{1+e^{b_0+b_1 t}}}$$

Όπου γ>0 και b_1 <0 ,είναι γνωστή ως λογιστική καμπύλη (logistic curve). Η παραπάνω συνάρτηση δεν είναι γραμμική ούτε ως προς τις μεταβλητές ούτε ως προς τις παραμέτρους και επωμένος δε μπορεί να εκτιμηθεί κατα τα γνωστά.



Η παράμετρος Κ (σχήμα)ή γ είναι το ανώτατο όριο που μπορεί να φτάσει η μεταβλητή Υ .Όσο μεγαλύτερη είναι η τιμή του συντελεστή b_1 τόσο ταχύτερα η Υ προσεγγίζει το ανώτατο όριο γ .Η λογιστική συνάρτηση μπορεί να χρησημοποιηθεί για την περιγραφή φαινομένων ,όπως η ανάπτυξη (growth) και η εξάπλωση ενός νέου προιόντος και γενικά μιας βιομηχανίας .Αν η παράμετρος γ ήταν γνωστή ,τότε το πρόβλημα της εκτίμησης γίνεται πολύ απλό γιατί: $\boxed{Y^* = b_0 + b_1 t}$ όπου $\boxed{Y^* = \log(\frac{g}{\mathrm{Y}} - 1)}$

Στην συνάρτηση αυτή μπορούμε να χρησημοποιήσουμε την μέθοδο των ελαχίστων τετραγώνων για να εκτιμήσουμε τους συντελεστές b_0 και b_1.

2.3.5 ΣΥΝΑΡΤΗΣΗ ΣΤΑΘΕΡΩΝ ΕΛΑΣΤΙΚΟΤΗΤΩΝ

Έστω η συνάρτηση :

$$\boxed{Y = b_0 \mathrm{X}_1^{b_1} \mathrm{X}_2^{b_2} \ldots \mathrm{X}_k^{b_k}} (1)$$

Όπου οι μεταβλητές $X_1, X_2 \mathbf{K}, X_k$ εμφανίζονται πολλαπλασιαστικά .Η παραπάνω συνάρτηση είναι γνωστή ως συνάρτηση σταθερών ελαστικοτήτων γιατί οι ελαστικότητες της Υ ,ως προς κάθε ανεξάρτητη μεταβλητή X_j $b_0^* = \log b_0$ είναι σταθερές και ίσες με τον αντίστοιχο συντελεστή b_j.

Δηλαδή :

$$\boxed{\frac{\partial \mathrm{Y}}{\partial \mathrm{X}_j} \cdot \frac{X_j}{Y} = b_j} \quad \text{για} \quad \boxed{j = 1, 2, \ldots.., k}$$

Η γνωστή συνάρτηση παραγωγής Cobb-Douglas αποτελεί ειδική περίπτωση της παραπάνω συναρτήσεως για Κ=2 .Δηλαδή,

$\boxed{Y = b_0 X_1^{b_1} X_2^{b_2}}$ όπου Y=προιόν X1=εργασία ,X2=κεφάλαιο

Η συνάρτηση (1) που δεν είναι γραμμική ούτε ως προς τις μεταβλητές ούτε ως προς τις παραμέτρους ,μετατρέπεται στην ακόλουθη λογαριθμικά γραμμική συνάρτηση :

$$\boxed{Y^* = b_0^* + b_1 X_1^* + b_2 X_2^* + ... + b_k X_k^*} (2)$$

Οπου $\boxed{Y^* = \log Y}$,

$\boxed{X_j^* = \log X_j}$

$\boxed{b_0^* = \log b_0}$

2.4 ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΑΝΑΜΕΣΑ ΣΤΗΝ ΓΡΑΜΜΙΚΗ ΚΑΙ ΣΤΗ ΛΟΓΑΡΙΘΜΙΚΑ ΓΡΑΜΜΙΚΗ ΜΟΡΦΗ

Ο καθορισμός της μορφής της συναρτησιακής σχέσης που συνδέει την εξαρτημένη ή ερμηνεύομενη μεταβλητή Y και τις ανεξάρτητες η ερμηνευτικές μεταβλητές X ,είναι μέρος της εξειδικεύσεως του υποδείγματος.Η επιλογή ανάμεσα σε μία γραμμική η λογαριθμικά γραμμική μορφή ήταν πάντα ένα πρόβλημα στην πράξη.Κατ'αρχάς ,η σύγκριση ενός γραμμικού και ενός λογαριθμικά γραμμικού υποδείγματος με βάση τον συντελεστή προσδιορισμού δεν ειναι έγκυρη ,γιατι :

$\boxed{\Sigma(\mathrm{Y} - \bar{Y})^2 \neq \Sigma(\log Y - \log \bar{Y})^2}$.Απαιτείται επομένως ,η υιοθέτηση κριτιρίων που να παρακάμπτουν το πρόβλημα αυτό. Τα βασικότερα κριτήρια είναι :BOX-COX,BERA-McALEER και MACKINNON-WHITE-DAVIDSON .

2.4.1BOX-COX

Έστω ο μετασχηματισμός :

$\boxed{Y(l) = \frac{\mathrm{Y}^{l} - 1}{l}}$ για $l \neq 0$ ....= logY για λ = 0.

Έστω το υπόδειγμα :

$$\boxed{Y_t(l) = b_0 + b_1 \mathrm{X}_t + u_t}$$

Το οποίο για λ=1 είναι γραμμικό και για λ=0 ειναι λογαριθμικά γραμμικό,σύμφωνα με την παραπάνω υπόθεση.Το λ μπορεί να εκτιμηθεί με την μέθοδο της μέγιστης πιθανοφάνειας και στην συνέχεια να ελεγχθεί η τιμή του ,δηλαδή να ελεγχθούν οι υποθέσεις λ=0 και λ=1 .Αν λ=0,ως εξαρτημένη μεταβλητή χρησημοποιείται η logY ,ενώ αν λ=1 ως εξαρτημένη μεταβλητή χρησημοποιείται η Y.Αν πάλι και οι δύο υποθέσεις γίνουν αποδεκτές ή και οι δύο απορριφθούν,μπορεί να χρησημοποιηθεί η εκτιμημένη τιμή του λ ,οπότε η εξαρτημένη μεταβλητή θα είναι Y(λ).

Η μέθοδος που προτείνουν οι BOX-COX είναι η εξης :

1ο Βήμα

Διαιρούμε κάθε παρατήρηση Y_t με τον γεωμετρικό μέσο των παρατηρήσεων της Υ.

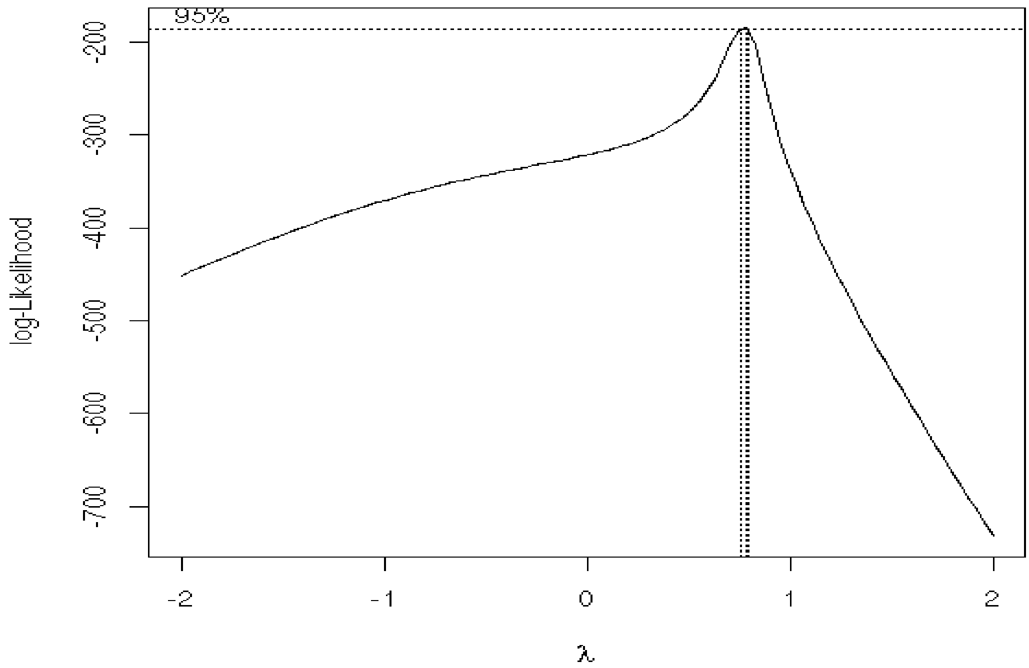
2ο Βήμα

Δίνοντας διάφορες τιμές στο λ εκτιμάμε τις αντίστοιχες παλινδρομήσεις ανάμεσα στο Υ(λ) και στο Χ και υπολογίζουμε το άθροισμα των τετραγώνων των καταλοίπων.

3ο Βήμα

Επιλέγουμε αυτή την τιμή του λ για την οποία το άθροισμα των τετραγώνων των καταλοίπων είναι ελάχιστο.

*επίσης μπορούμε να διαλέξουμε ανάμεσα στο γραμμικό και το λογαριθμικά γραμμικό υπόδειγμα αυτό με το μικρότερο άθροισμα των τετραγώνων των καταλοίπων.



2.4.2BERA-McALEER

Έστω,η μηδέν υπόθεση :

$\boxed{H_0 : \log Y_t = b_0 + b_1 X_t + u_t}$ όπου $u_t \sim (0, s_u^2)$

Και η εναλλακτική :

$\boxed{H_1 : Y_t = b_0 + b_1 X_t + e_t}$ όπου $e_t \sim (0, s_e^2)$

Ο έλεγχος των παραπάνω υποθέσεων μπορεί να γίνει ώς εξής :

1^{o} Βήμα

Εκτιμάμε τα δύο υποδείγματα και βρίσκουμε τις υπολογισμένες τιμές ,δηλαδή $\widehat{LogY_t}$ και $\hat{Y}_t$ αντίστοιχα .Η υπολογισμένη τιμή του Y_t απο την λογαριθμικά γραμμική μορφή ειναι αντιλογάριθμος $\widehat{LogY_t}$,ενώ η υπολογισμένη τιμή του $\log Y_t$,απο την γραμμική μορφή είναι $\log\hat{Y}_t$.

2^{o} Βήμα

Εκτιμάμε τις εξής δύο παλινδρομήσεις :

Αντιλ. $\widehat{LogY_t} = b_0 + b_1 X_1 + n_t$

$\log \hat{Y}_t = b_0 + b_1 X_1 + e_t$

Και έστω $\hat{n}_t$ και $\hat{e}_t$ τα αντίστοιχα κατάλοιπα .

3^{o} Βήμα

Με τα κατάλοιπα $\hat{n}_t$ και $\hat{e}_t$ εκτιμάμε τις εξής δύο βοηθητικές παλινδρομήσεις :

$$\boxed{\log Y_t = b_0 + b_1 X_t + q_0 \hat{n}_t + n^*_t}$$

$$\boxed{Y_t = b_0 + b_1 X_t + q_1 \hat{e}_t + e_t^*}$$

Ο έλεγχος των δύο παραπάνω υποθέσεων βασίζεται στους συντελεστές q_0 και q_1,στις βοηθητικές παλινδρομήσεις και γίνεται με την στατιστική t.Έτσι αν γίνει δεκτή η υπόθεση $q_0 = 0$ επιλέγεται η λογαριθμικά γραμμική μορφή ,ενω αν γίνει δεκτή η υπόθεση $q_1 = 0$,επιλέγεται η γραμμική μορφή.Το πρόβλημα βέβαια το έχουμε όταν και οι δύο υποθέσεις γίνονται δεκτές η απορρίπτονται.

2.4.3McKINNON-WHITE-DAVIDSON

Είναι ουσιαστικά παραλλαγή του προηγούμενου κριτηρίου και αναφέρεται εν συντομία MWD.Η διαδικασία έχει ως εξής :

1^{o} Βήμα

Εναι ακριβώς το ίδιο με το κριτήριο Bera-McAleer .

2^{o} Βήμα

Εκτιμάμε τις εξής δύο βοηθητικές παλινδρομίσεις :

$$\boxed{\log Y_t = b_0 + b_1 X_1 + q_0(\hat{Y}_t - anti\,l.\log\hat{}\,Y_t) + e_t^*}$$

$$\boxed{Y_t = b_0 + b_1 X_t + q_1(\log \hat{g}\, Y_t - \log \hat{Y}_t) + e_t^*}$$

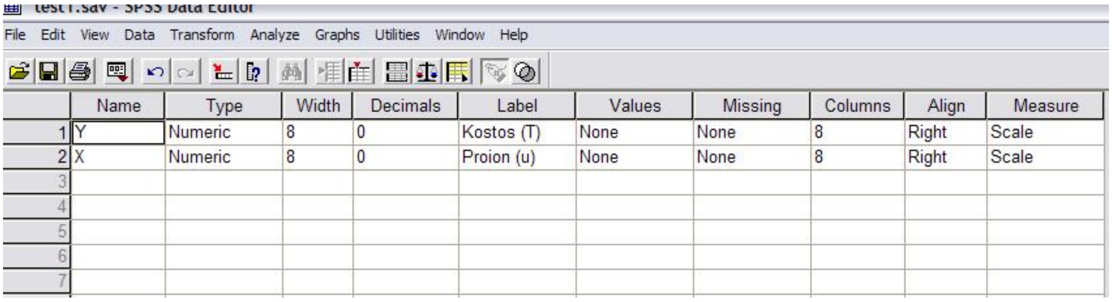
Ο έλεγχος των δύο παραπάνω υποθέσεων βασίζεται στους συντελεστές q_0 και q_1, και γίνεται με την στατιστική t. Έτσι αν γίνει δεκτή η υπόθεση $q_0 = 0$ επιλέγεται η λογαριθμικά γραμμική μορφή ,ενώ αν γίνει δεκτή η υπόθεση $q_1 = 0$,επιλέγεται η γραμμική μορφή.

ΚΕΦΑΛΑΙΟ 3

ΕΦΑΡΜΟΓΕΣ ΜΕ SPSS 13.0

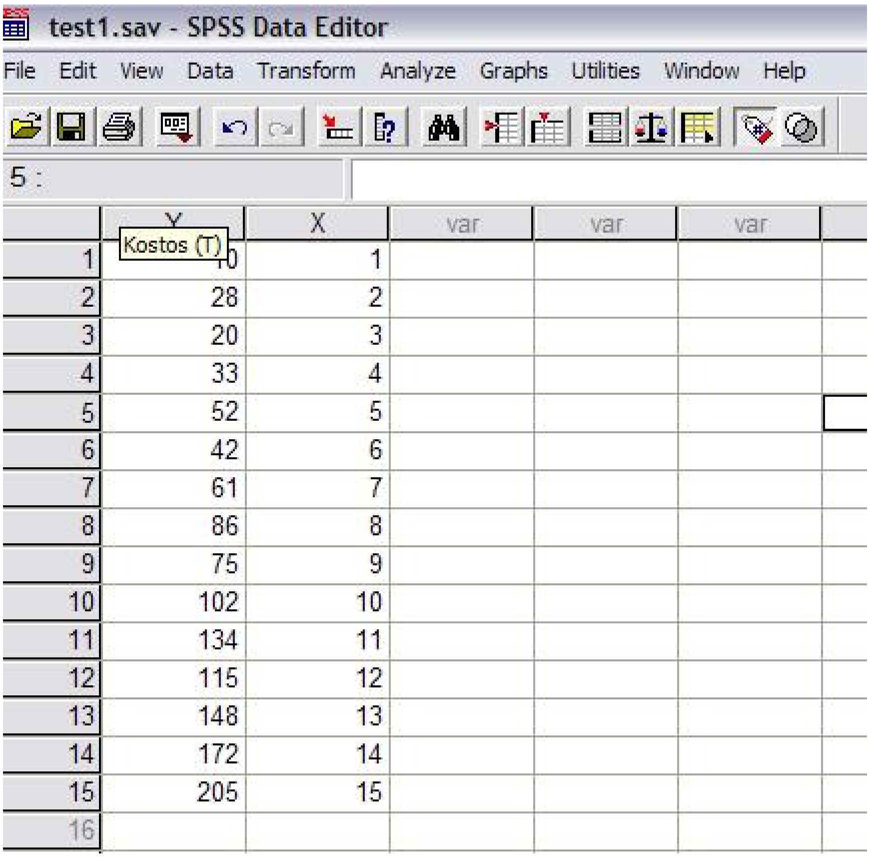
3.1 ΜΟΝΤΕΛΟ ΚΟΣΤΟΥΣ

Περνάμε τις μεταβλητές του μοντέλου μας στο variable view του spss έτσι ώστε να αρχίσουμε με την εισαγωγή των δεδομένων.

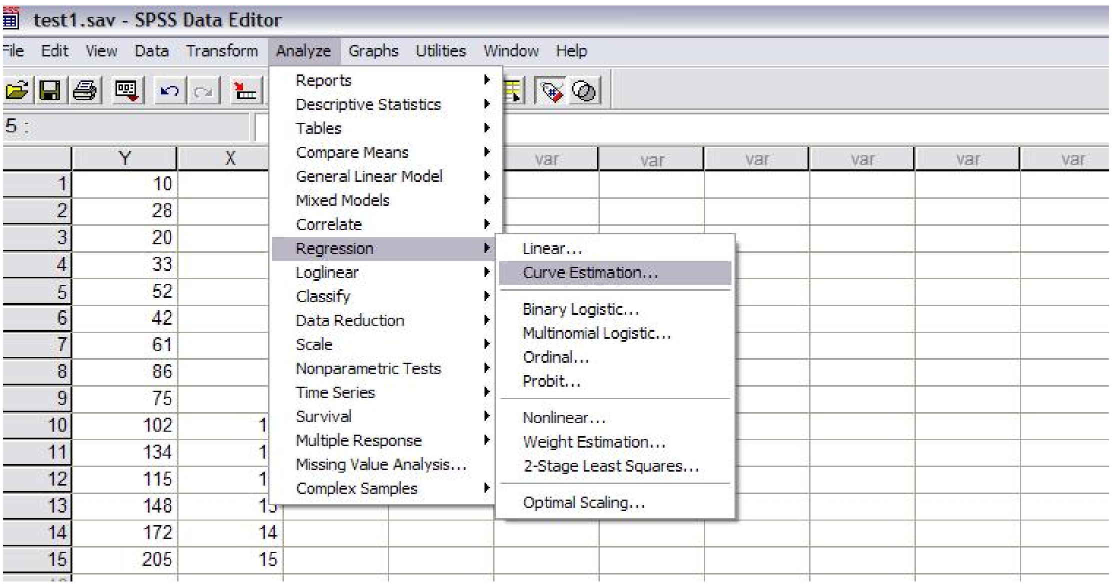


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Y	Numeric	8	0	Kostos (T)	None	None	8	Right	Scale
2	X	Numeric	8	0	Proion (u)	None	None	8	Right	Scale
3										
4										
5										
6										
7										

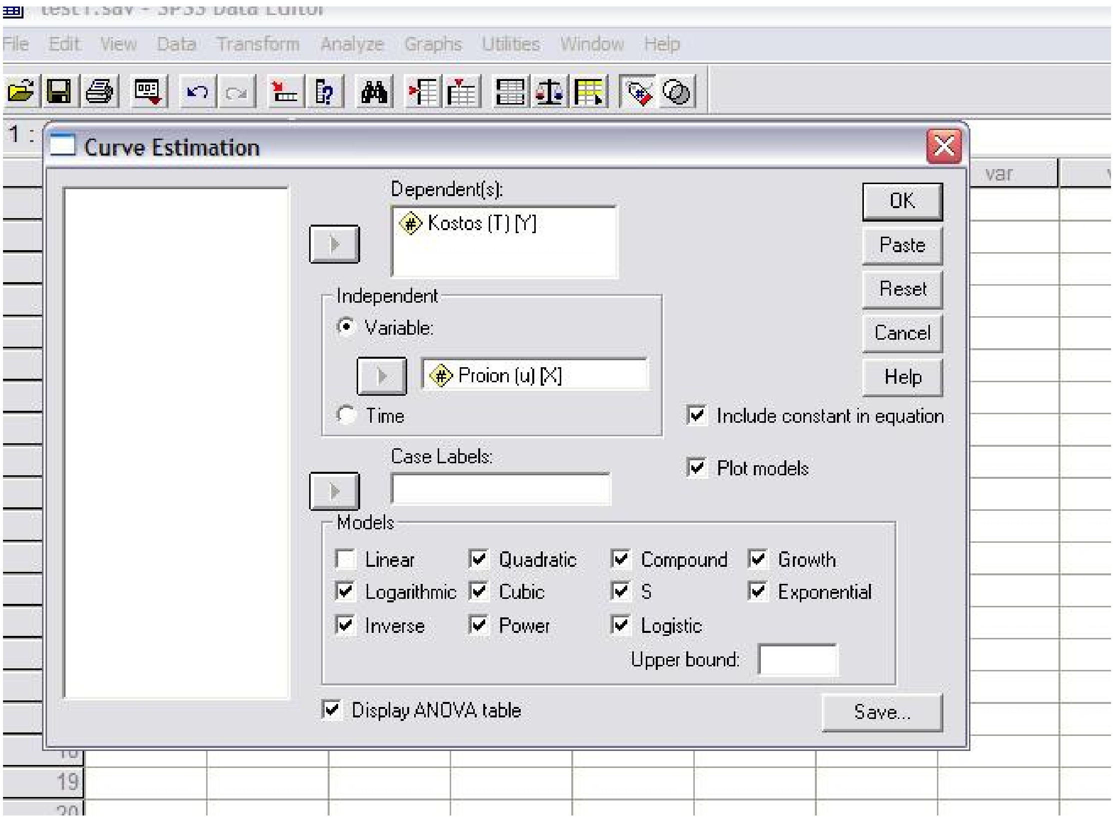
Εισάγουμε τα δεδομένα μας στις στήλες Y=Κόστος και X=Μονάδες Προιόντος .



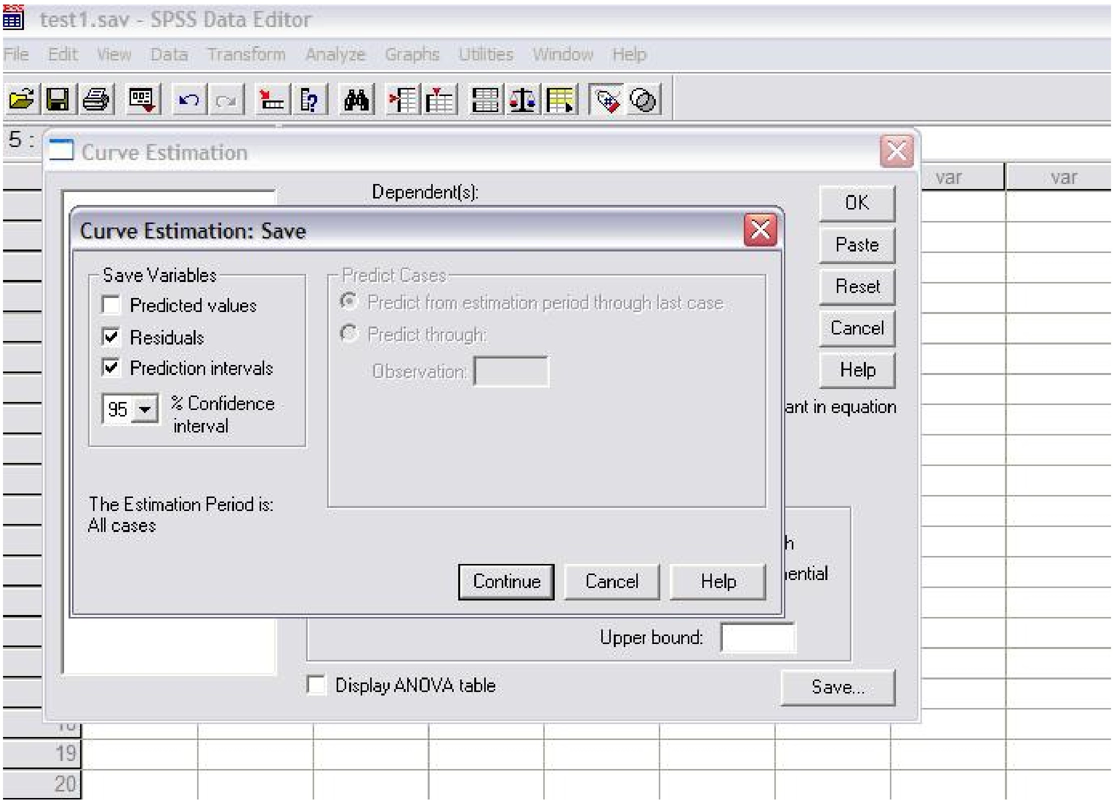
Επιλέγουμε απο το μενού Analyze την διαδρομή Regression>Curve Estimation.



Τοποθετούμε στο πλαίσιο της εξαρτημένης μεταβλητής το Y=κόστος και στο πλαίσιο της ανεξάρτητης το X=προιόν.Επιλέγουμε όλες τις πιθανές παλινδρομήσεις που μπορεί να ταιριάζουν στο δείγμα μας και επιλέγουμε το πλαίσιο ANOVA.



Αφού επιλέξουμε και στο μενού save η εκτίμηση μας να περιέχει τα υπόλοιπα και διάστημα εμπιστοσίνης παταμε continue και μετα ok.



Το spss μας δίνει τα αποτελέσματα για όλες τις παλινδρομήσεις που επιλέξαμε .Το στοιχείο που μας ενδιαφέρει πιο πολύ είναι το R Square,το οποίο εκφράζει το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από το μοντέλο.Όσο μεγαλύτερο είναι τόσο καλύτερα εκφράζει η κάθε μορφή παλινδρόμησης το δείγμα.Επίσης ο έλεγχος για την συνολική σημαντικότητα του μοντέλου δίνεται από την στήλη F του πίνακα ANOVA.Ακολουθούν τα αποτελέσματα για όλες τις πιθανές μορφές :

Logarithmic

Model Summary

R	**R Square**	Adjusted R Square	Std. Error of the Estimate
.854	**.729**	.708	31.934

The independent variable is Proion (u).

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	35684.279	1	35684.279	34.991	.000
Residual	13257.455	13	1019.804		
Total	48941.733	14			

The independent variable is Proion (u).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
ln(Proion (u))	64.563	10.914	.854	5.915	.000
(Constant)	-34.550	21.911		-1.577	.139

Inverse

Model Summary

R	**R Square**	Adjusted R Square	Std. Error of the Estimate
.641	**.411**	.365	47.109

The independent variable is Proion (u).

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	20090.892	1	20090.892	9.053	.010
Residual	28850.842	13	2219.296		
Total	48941.733	14			

The independent variable is Proion (u).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 / Proion (u)	-154.068	51.206	-.641	-3.009	.010
(Constant)	119.616	16.621		7.197	.000

Quadratic

Model Summary

R	**R Square**	Adjusted R Square	Std. Error of the Estimate
.985	**.971**	.966	10.902

The independent variable is Proion (u).

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	47515.609	2	23757.805	199.908	.000
Residual	1426.124	12	118.844		
Total	48941.733	14			

The independent variable is Proion (u).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Proion (u)	2.149	2.793	.163	.770	.456
Proion (u) ** 2	.664	.170	.827	3.913	.002
(Constant)	13.440	9.710		1.384	.192

Cubic

Model Summary

R	**R Square**	Adjusted R Square	Std. Error of the Estimate
.986	**.973**	.966	10.952

The independent variable is Proion (u).

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	47622.390	3	15874.130	132.350	.000
Residual	1319.344	11	119.940		
Total	48941.733	14			

The independent variable is Proion (u).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Proion (u)	**8.997**	**7.781**	.681	1.156	**.272**
Proion (u) ** 2	**-.372**	**1.111**	-.463	-.335	**.744**
Proion (u) ** 3	**.043**	**.046**	.797	.944	**.366**
(Constant)	**2.870**	**14.854**		.193	**.850**

Compound

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.961	**.924**	.918	.247

The independent variable is Proion (u).

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	9.604	1	9.604	157.230	.000
Residual	.794	13	.061		
Total	10.398	14			

The independent variable is Proion (u).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Proion (u)	1.203	.018	2.614	67.705	.000
(Constant)	14.595	1.960		7.447	.000

The dependent variable is ln(Kostos (T)).

Growth

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.961	**.924**	.918	.247

The independent variable is Proion (u).

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	9.604	1	9.604	157.230	.000
Residual	.794	13	.061		
Total	10.398	14			

The independent variable is Proion (u).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Proion (u)	.185	.015	.961	12.539	.000
(Constant)	2.681	.134		19.962	.000

The dependent variable is ln(Kostos (T)).

Exponential

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.961	**.924**	.918	.247

The independent variable is Proion (u).

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.

Regression	9.604	1	9.604	157.230	.000
Residual	.794	13	.061		
Total	10.398	14			

The independent variable is Proion (u).

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Proion (u)	.185	.015	.961	12.539	.000
(Constant)	14.595	1.960		7.447	.000

The dependent variable is ln(Kostos (T)).

Logistic

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.961	.924	.918	.247

The independent variable is Proion (u).

ANOVA

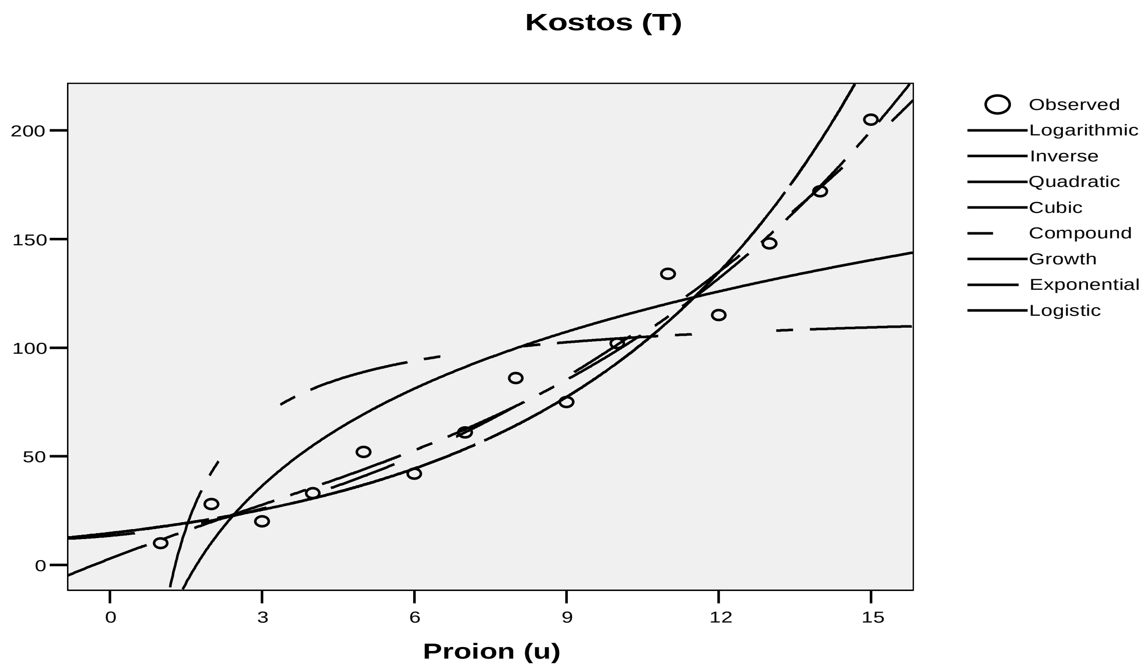
	Sum of Squares	df	Mean Square	F	Sig.
Regression	9.604	1	9.604	157.230	.000
Residual	.794	13	.061		
Total	10.398	14			

The independent variable is Proion (u).

Coefficients

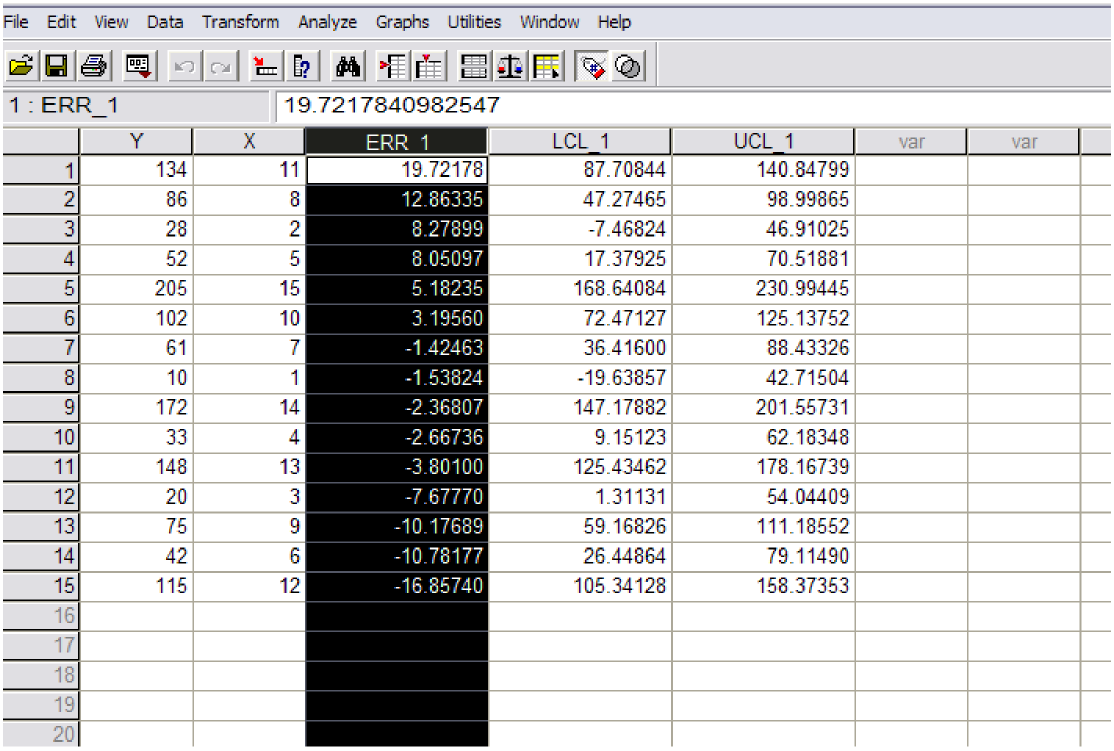
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Proion (u)	.831	.012	.382	67.705	.000
(Constant)	.069	.009		7.447	.000

The dependent variable is ln(1 / Kostos (T)).



Βλέπουμε ότι η μορφή που ερμηνεύει καλύτερα την εξαρτημένη μεταβλητή είναι η πολυωνυμία $3^{ου}$ βαθμού (CUBIC).Το R Square της Cubic είναι 0.973,μεγαλύτερο από όλες τις άλλες μορφές .Για να είναι όμως το μοντέλο μας σωστό θα <u>πρέπει όπως γνωρίζουμε από την θεωρία τα υπόλοιπα να ακολουθούν κανονική κατανομή.</u>

Επιλέγουμε την στήλη ERR_1 και ακολουθούμε την διαδρομή GRAPHS>HISTOGRAM(display normal curve)>OK.

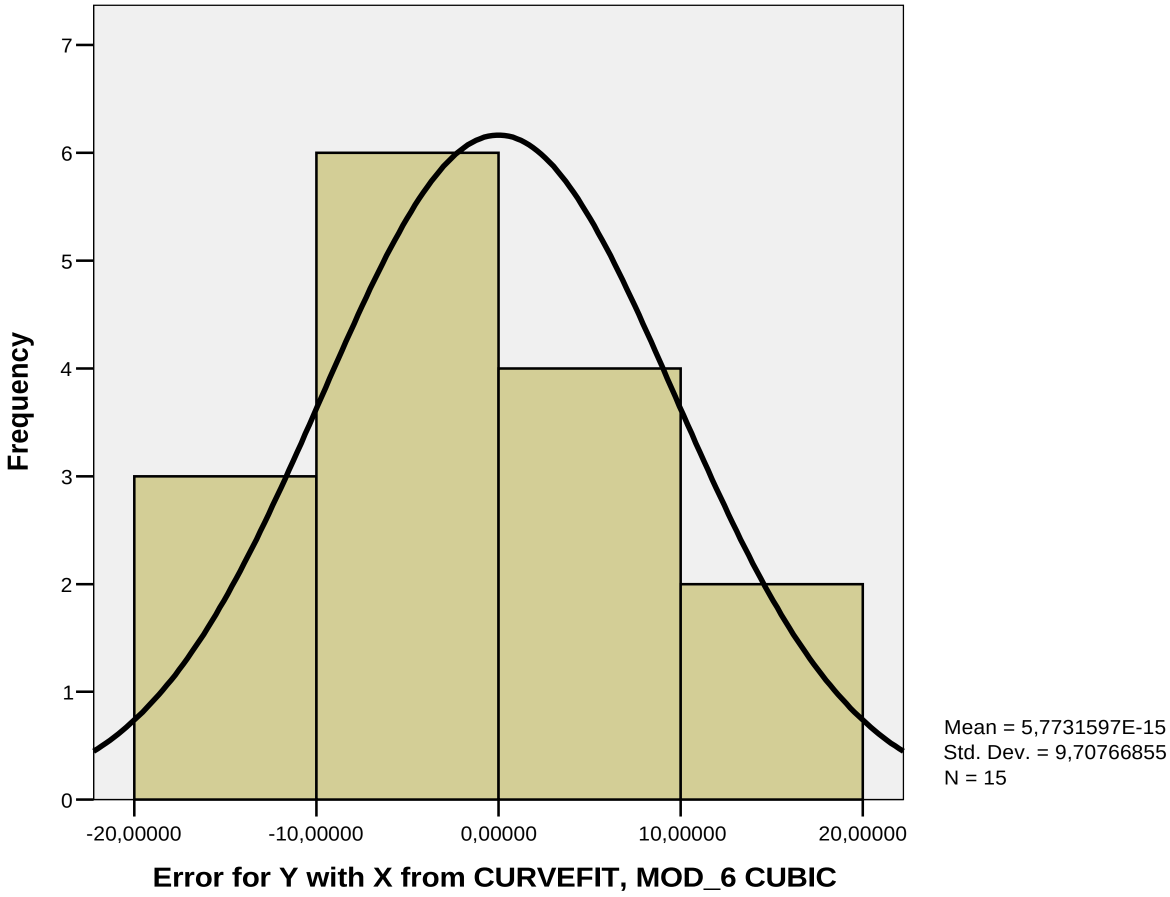


File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : ERR_1 19.7217840982547

	Y	X	ERR_1	LCL_1	UCL_1	var	var
1	134	11	19.72178	87.70844	140.84799		
2	86	8	12.86335	47.27465	98.99865		
3	28	2	8.27899	-7.46824	46.91025		
4	52	5	8.05097	17.37925	70.51881		
5	205	15	5.18235	168.64084	230.99445		
6	102	10	3.19560	72.47127	125.13752		
7	61	7	-1.42463	36.41600	88.43326		
8	10	1	-1.53824	-19.63857	42.71504		
9	172	14	-2.36807	147.17882	201.55731		
10	33	4	-2.66736	9.15123	62.18348		
11	148	13	-3.80100	125.43462	178.16739		
12	20	3	-7.67770	1.31131	54.04409		
13	75	9	-10.17689	59.16826	111.18552		
14	42	6	-10.78177	26.44864	79.11490		
15	115	12	-16.85740	105.34128	158.37353		
16							
17							
18							
19							
20							

Μέσω αυτής της διαδρομής εμφανίζεται το ιστόγραμμα συχνοτήτων των υπολοίπων που από ότι βλέπουμε ακολουθούν κανονική κατανομή.Άρα η μορφή που διαλέξαμε είναι ικανοποιητική.



Στην συνέχεια τρέχουμε μόνο την cubic μορφή με την ίδια διαδικασία έτσι ώστε το spss να μας δώσει σχηματικά την μοναδική και βέλτιστη μορφή.

Γνωρίζουμε από την θεωρία ότι ο τύπος της cubic είναι ο :

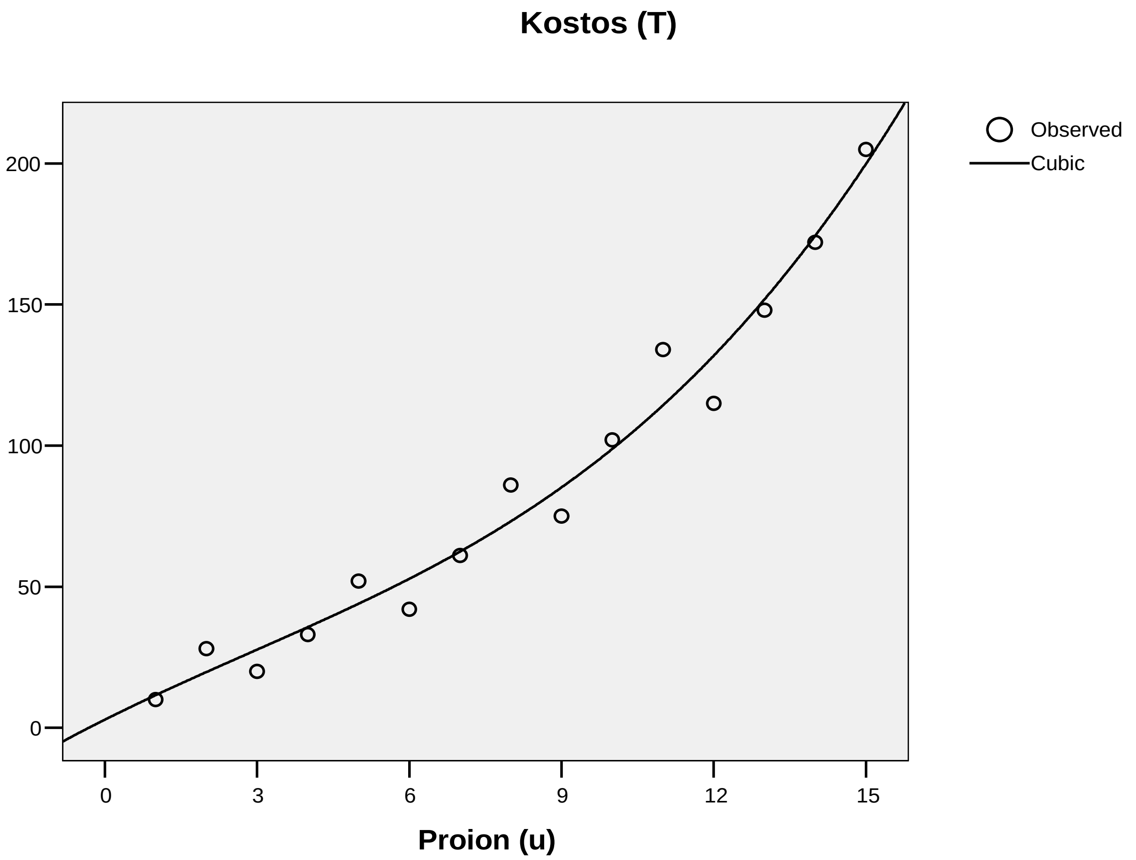
$$\boxed{\mathrm{Y} = b_0 + b_1\mathrm{X} + b_2\mathrm{X}^2 + \ldots + b_k\mathrm{X}^k + u}$$

Άρα η ευθεία μας στο δείγμα αυτό δίνεται από τον πίνακα Coefficients

Και είναι η :

$$Y = 2.870 + 8.997X - 0.372X^2 + 0.043X^3$$

$$(0,850)(0,272)(0,744)(0,366)$$



3.2 ΔΑΠΑΝΗ ΓΙΑ ΤΡΟΦΗΜΑ ΚΑΙ ΑΚΑΘΑΡΙΣΤΟ ΕΘΝΙΚΟ ΕΙΣΟΔΗΜΑ

Περνάμε τις μεταβλητές του μοντέλου μας στο variable view του spss έτσι ώστε να αρχίσουμε με την εισαγωγή των δεδομένων.

test2.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	year	Numeric	8	0		None	None	8	Right	Scale
2	Y	Numeric	8	1	dapani gia trofi	None	None	8	Right	Scale
3	X	Numeric	8	1	a.ethniko eiso	None	None	8	Right	Scale
4										
5										
6										
7										
8										
9										
10										
11										

Εισάγουμε τα δεδομένα μας(χιλ.ευρώ).

test2.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 :

	year	Y	X	var	var	var
1	1958	5.5	16.4			
2	1959	5.6	16.8			
3	1960	5.6	17.5			
4	1961	6.0	19.3			
5	1962	6.1	19.5			
6	1963	6.4	21.4			
7	1964	6.7	23.1			
8	1965	7.1	25.1			
9	1966	7.5	26.6			
10	1967	7.8	27.6			
11	1968	8.0	29.4			
12	1969	8.3	32.2			
13	1970	8.4	34.6			
14	1971	8.5	37.1			
15	1972	8.8	40.1			
16	1973	9.3	43.1			
17	1974	9.3	41.1			
18	1975	9.5	43.1			
19	1976	9.6	45.3			
20	1977	9.5	46.5			
21	1978	9.9	48.8			
22	1979	9.9	50.2			
23						
24						
25						
26						

Όπως και στο μοντέλο 3.1 ,και αφού ακολουθήσουμε την ίδια ακριβώς διαδικασία ,συγκρίνουμε όλες τις πιθανές παλινδρομήσεις .Στο μοντέλο αυτό η μορφή που μας ικανοποιεί είναι η αντίστροφη(inverse).Βλέπουμε από τους παρακάτω πίνακες ότι το μοντέλο μας αυτό ερμηνεύει το 97,5% της πληροφορίας του Y(R square).

Inverse

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.988	.975	.974	.247

The independent variable is a.ethniko eisodima.

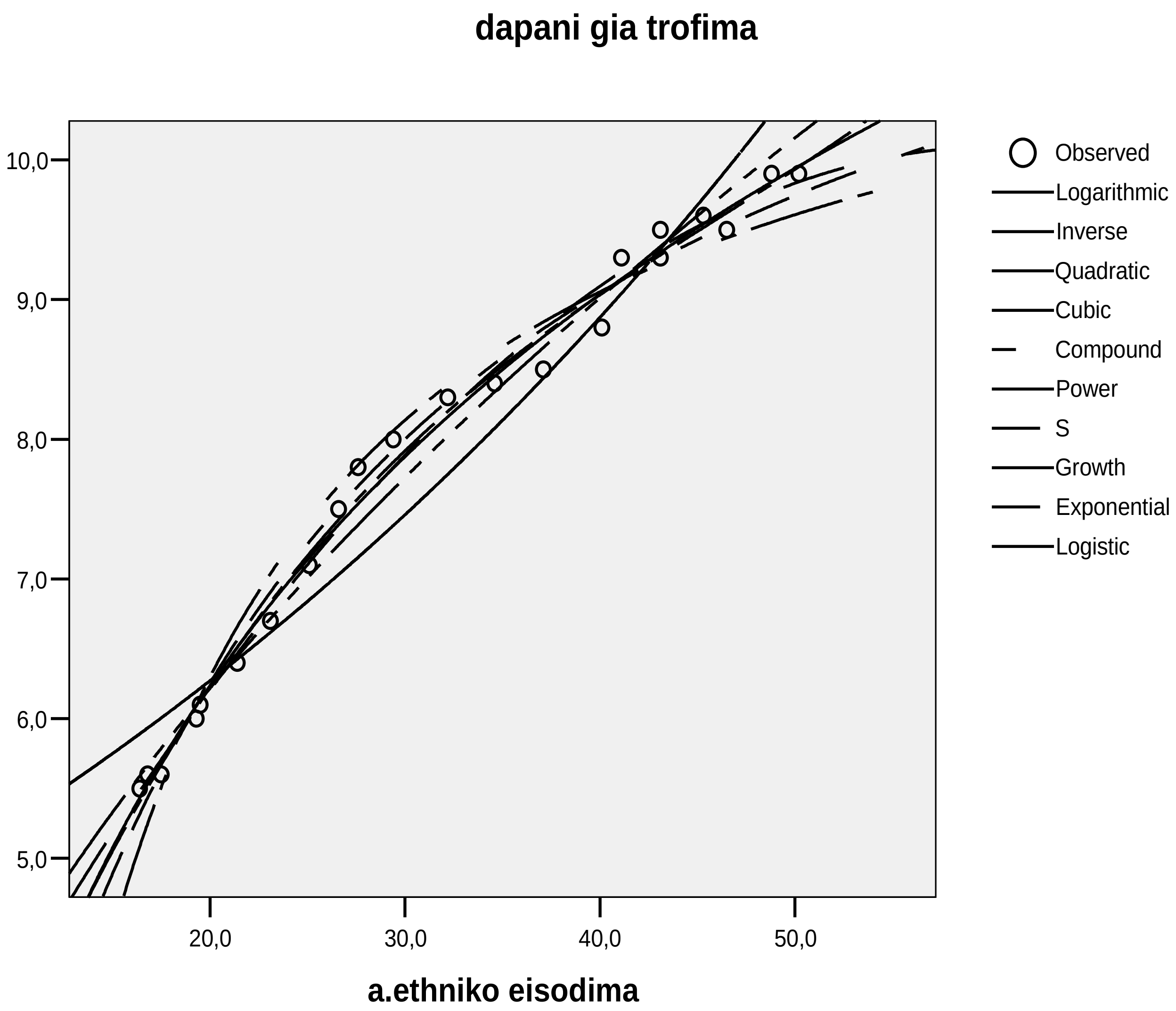
ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	47.980	1	47.980	787.492	.000
Residual	1.219	20	.061		
Total	49.199	21			

The independent variable is a.ethniko eisodima.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 / a.ethniko eisodima	-110.339	3.932	-.988	-28.062	.000
(Constant)	11.814	.150		78.844	.000



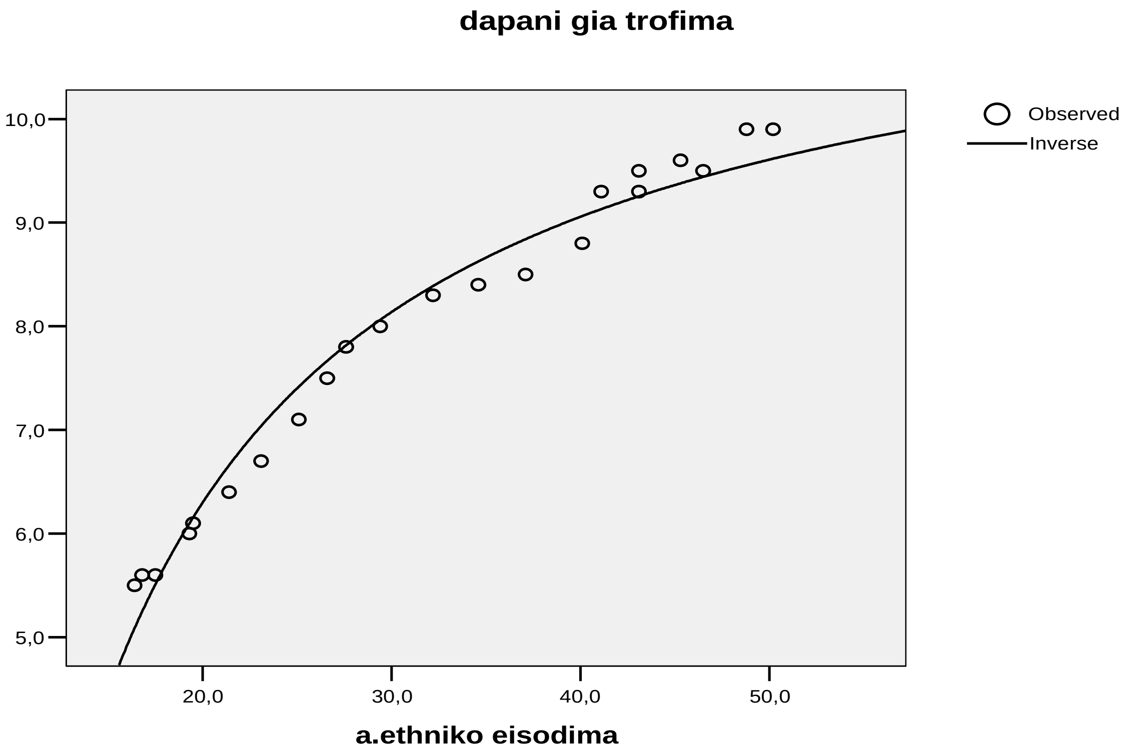
Στο παράπανω σχήμα βλέπουμε τις πιθανές ευθείες παλινδρόμησης .

Γνώριζουμε από την θεωρία ότι ο τύπος της αντίστοφης (INVERSE) είναι ο :

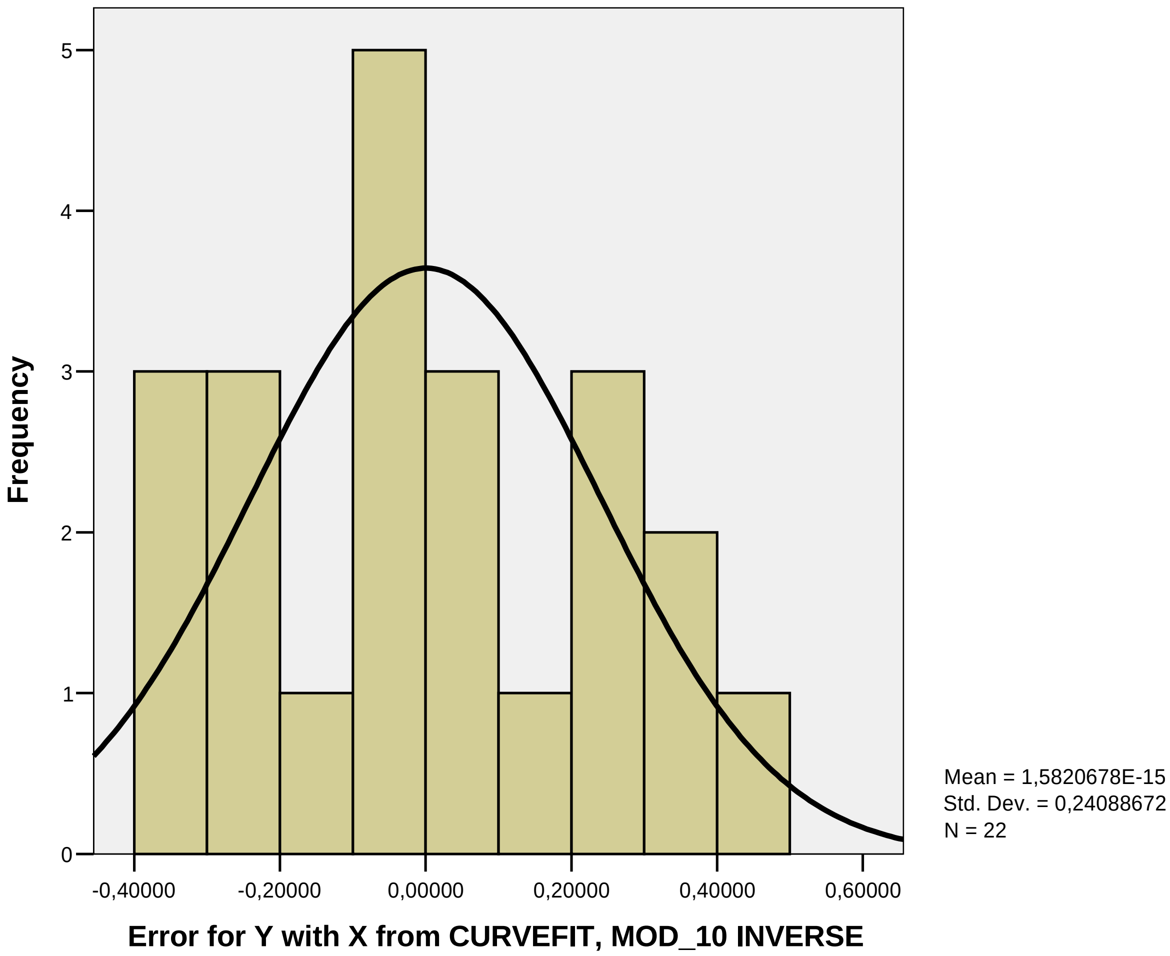
$$\boxed{Y = b_0 + b_1 \frac{1}{\mathrm{X}_t} + u_t}$$

Άρα η ευθεία μας στο δείγμα αυτό δινεταί απο τον πίνακα Coefficients. Και είναι η :

$$Y = 11,814 - 110,339 \frac{1}{X} + u$$

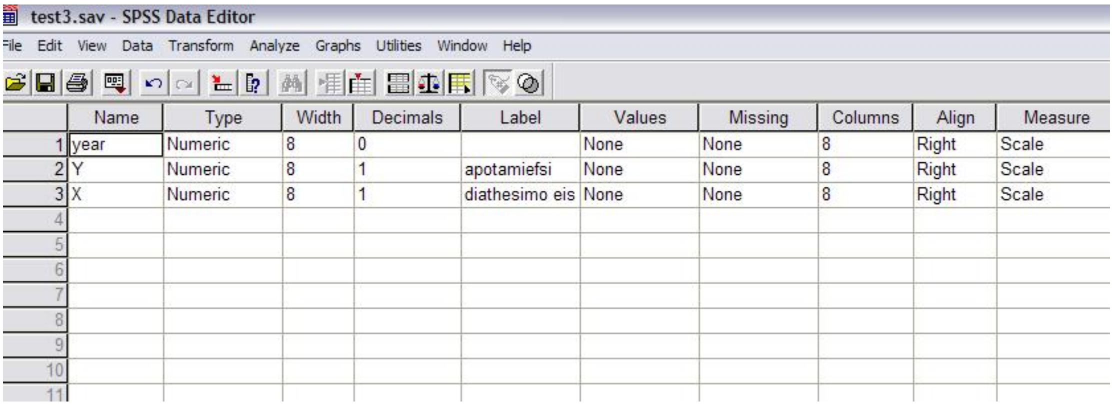


Για να είναι όμως το μοντέλο μας σωστό θα <u>πρέπει όπως γνωρίζουμε από την θεωρία τα υπόλοιπα να ακολουθούν κανονική κατανομή.</u>Έτσι ακλουθώντας πάλι την διαδικασία πού περιγράφεται στο 3.1 ,εμφανίζεται το ιστόγραμμα συχνοτήτων των υπολοίπων που από ότι βλέπουμε ακολουθούν κανονική κατανομή.Άρα η μορφή που διαλέξαμε είναι ικανοποιητική.


Frequency
5
4
3
2
1
0
-0,40000
-0,20000
0,00000
0,20000
0,40000
0,60000
Error for Y with X from CURVEFIT, MOD_10 INVERSE
Mean = 1,5820678E-15
Std. Dev. = 0,24088672
N = 22

3.3 ΑΠΟΤΑΜΙΕΥΣΗ ΚΑΙ ΔΙΑΘΕΣΙΜΟ ΕΙΣΟΔΗΜΑ

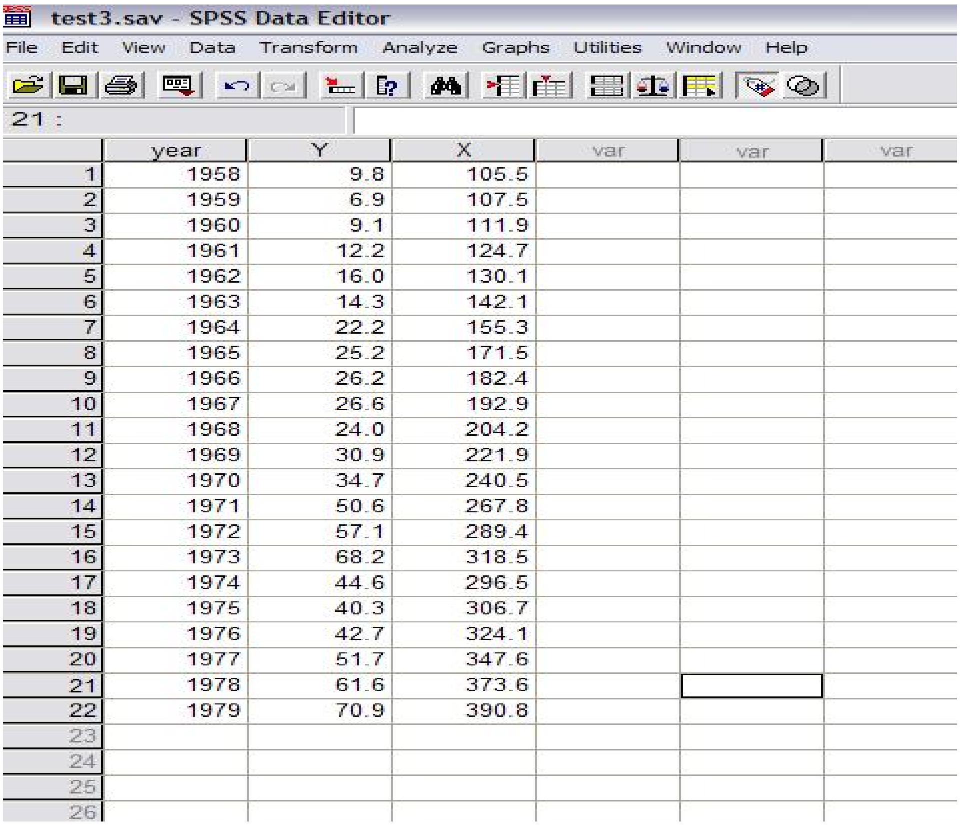
Περνάμε τις μεταβλητές του μοντέλου μας στο variable view του spss έτσι ώστε να αρχίσουμε με την εισαγωγή των δεδομένων.


test3.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	year	Numeric	8	0		None	None	8	Right	Scale
2	Y	Numeric	8	1	apotamiefsi	None	None	8	Right	Scale
3	X	Numeric	8	1	diathesimo eis	None	None	8	Right	Scale
4										
5										
6										
7										
8										
9										
10										
11										

Εισάγουμε τα δεδομένα μας(δισεκ.δρχ.).


test3.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

21 :

	year	Y	X	var	var	var
1	1958	9.8	105.5			
2	1959	6.9	107.5			
3	1960	9.1	111.9			
4	1961	12.2	124.7			
5	1962	16.0	130.1			
6	1963	14.3	142.1			
7	1964	22.2	155.3			
8	1965	25.2	171.5			
9	1966	26.2	182.4			
10	1967	26.6	192.9			
11	1968	24.0	204.2			
12	1969	30.9	221.9			
13	1970	34.7	240.5			
14	1971	50.6	267.8			
15	1972	57.1	289.4			
16	1973	68.2	318.5			
17	1974	44.6	296.5			
18	1975	40.3	306.7			
19	1976	42.7	324.1			
20	1977	51.7	347.6			
21	1978	61.6	373.6			
22	1979	70.9	390.8			
23						
24						
25						
26						

Όπως και στο μοντέλο 3.1και 3.2 ,και αφού ακολουθήσουμε την ίδια ακριβώς διαδικασία ,συγκρίνουμε όλες τις πιθανές παλινδρομήσεις .Στο μοντέλο αυτό η μορφή που μας ικανοποιεί είναι η αντίστροφη-S(inverse-S).Βλέπουμε από τους παρακάτω πίνακες ότι το μοντέλο μας αυτό ερμηνεύει το 95,0% της πληροφορίας του Y(R square).

S

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.975	.950	.947	.158

The independent variable is diathesimo eisodima.

ANOVA

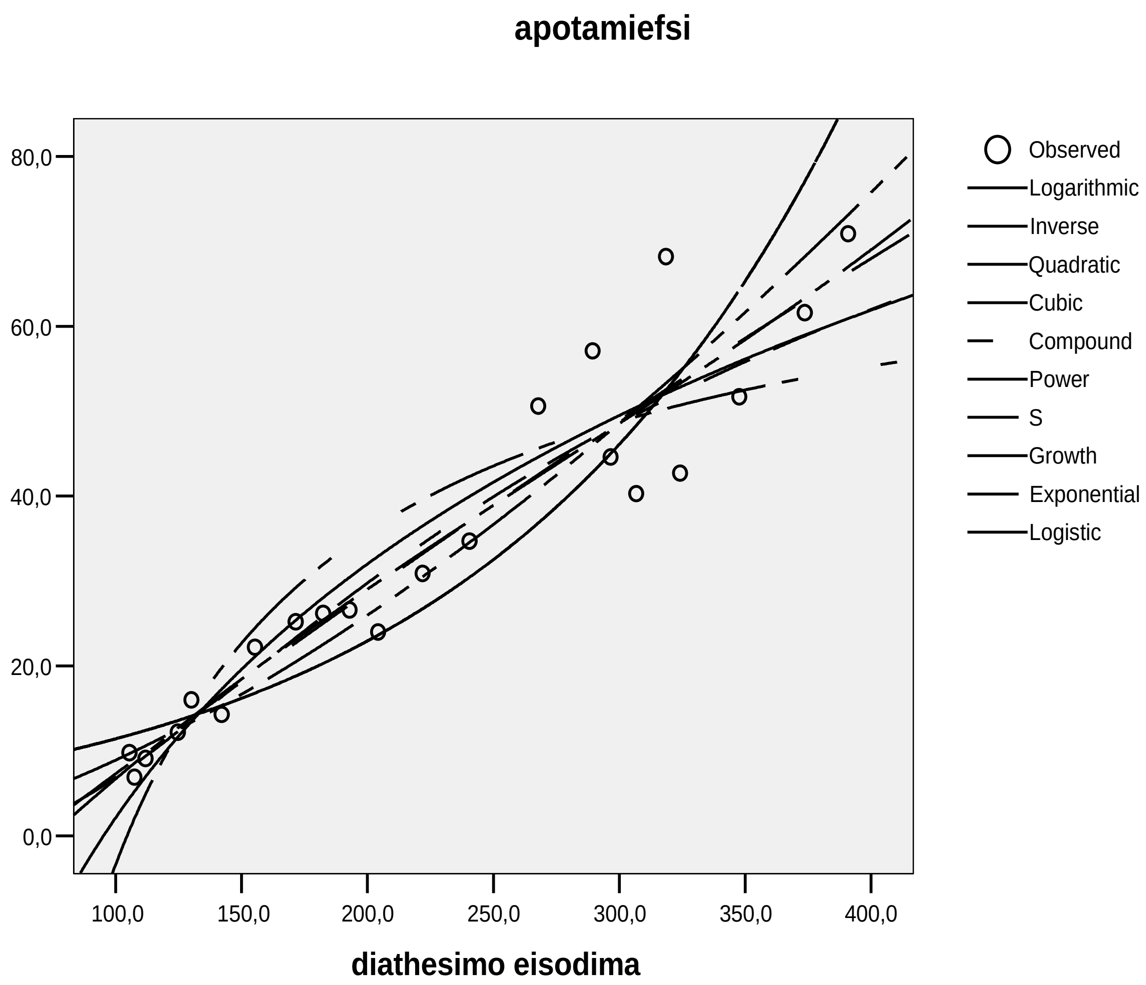
	Sum of Squares	df	Mean Square	F	Sig.
Regression	9.473	1	9.473	379.792	.000
Residual	.499	20	.025		
Total	9.971	21			

The independent variable is diathesimo eisodima.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
1 / diathesimo eisodima	-293.450	15.058	-.975	-19.488	.000
(Constant)	4.861	.086		56.727	.000

The dependent variable is ln(apotamiefsi).



Στο παράπανω σχήμα βλέπουμε τις πιθανές ευθείες παλινδρόμησης .

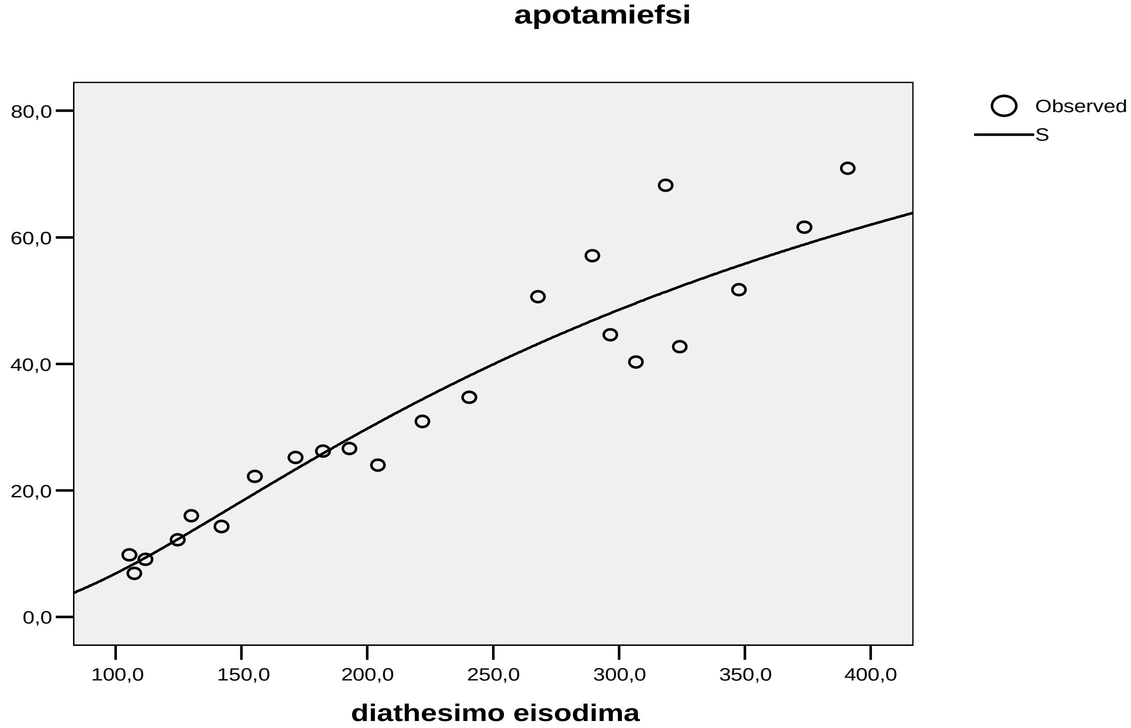
Γνώριζουμε από την θεωρία ότι ο τύπος της αντίστοφης -S(INVERSE-S) είναι ο :

$$\boxed{Y = b_0 + b_1 \frac{1}{\mathrm{X}_t} + u_t}$$

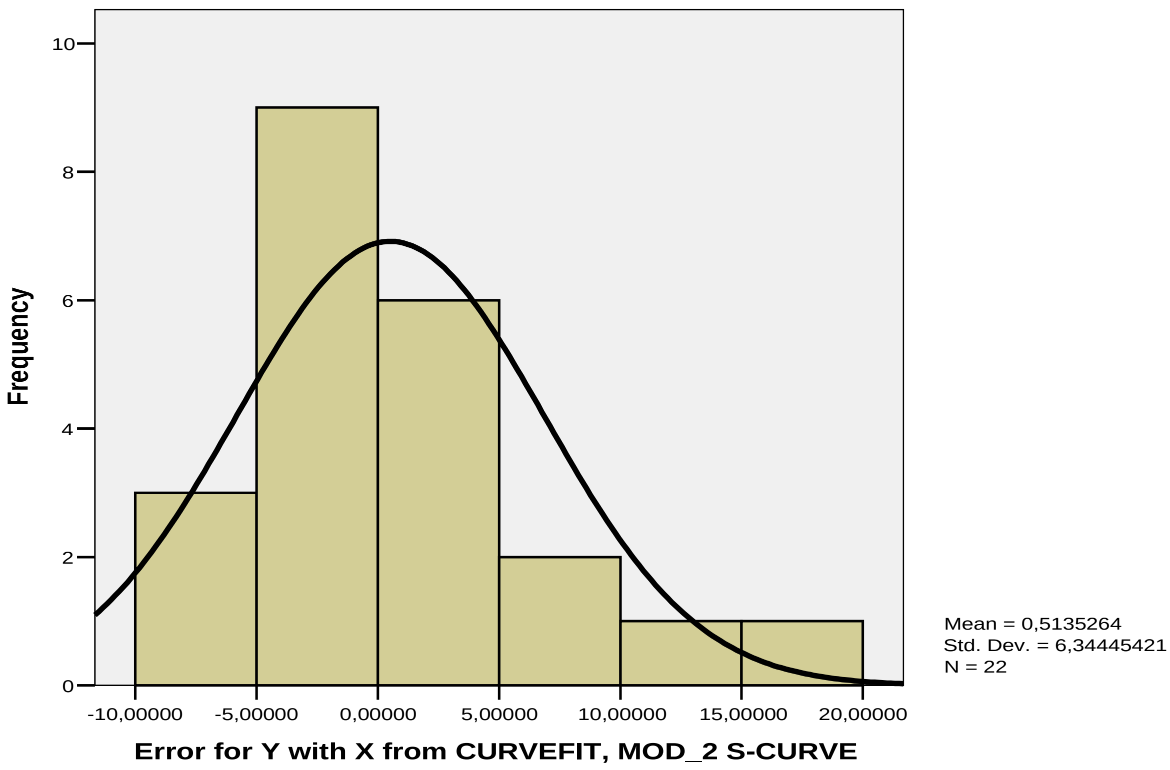
Άρα η ευθεία μας στο δείγμα αυτό δινεταί απο τον πίνακα Coefficients.

Και είναι η :

$$Y = 4,861 - 293,45 \frac{1}{X} + u$$



Για να είναι όμως το μοντέλο μας σωστό θα πρέπει όπως γνωρίζουμε από την θεωρία τα υπόλοιπα να ακολουθούν κανονική κατανομή.Έτσι ακλουθώντας πάλι την διαδικασία πού περιγράφεται στο 3.1 κ 3.2 ,εμφανίζεται το ιστόγραμμα συχνοτήτων των υπολοίπων που από ότι βλέπουμε ακολουθούν κανονική κατανομή.Άρα η μορφή που διαλέξαμε είναι ικανοποιητική.



ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Norman P., Smith H. (1997), *Εφαρμοσμένη ανάλυση παλινδρόμησης*, Αθήνα:Παπαζήση

2. Γ.Κ.Χρηστου(2006) , *Εισαγωγή στην οικονομετρία*, Αθήνα:Γκουτενμπέργκ

3. William H. Greene (2003), *Econometric Analysis*,Λονδίνο:Πεντις Χολ

4. Arthur E. Albert, Leland A. Grader Jr. (1967), *Stochastic approximation and non linear regression*,Νέα Υόρκη:Άγνωστος

5.Κιόχος Α.Π.(1993), *Σταστιστική* ,Αθήνα:Interbooks

6.Λαζαρίδης Α.(2000),*Οικονομετρία*,Θεσσαλονίκη: Ζυγός

7.ΔημητρίουΓ.(1991),*ΜαθήματαΑνάλυσηςΠαλινδρόμησης*,Πειραιάς: Σταμούλης

8.Γκαμαλέτσος Θ.(1972),Εφαρμοσμέμη Οικονομετρία,Αθήνα:Παπαζήση