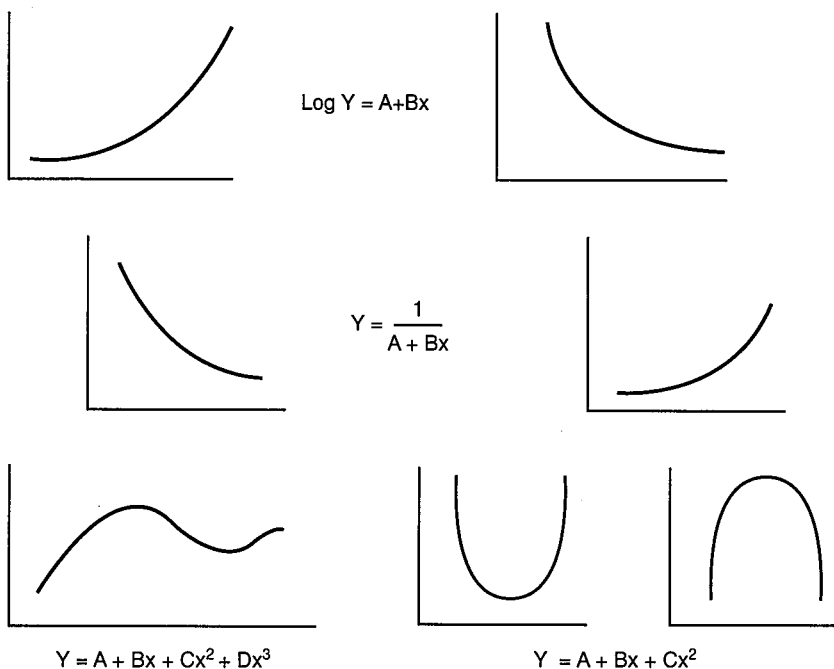


Μη γραμμική παλινδρόμηση

Χρησιμοποιώντας την πολλαπλή γραμμική παλινδρόμηση «εκβιάζουμε», κατά μια έννοια, τα δεδομένα μας να πάρουν σχήμα γραμμικό. Υπάρχουν όμως περιπτώσεις κατά τις οποίες η γραμμικότητα δεν είναι προφανής, διότι η γενική σχέση, δηλαδή αυτή μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης, δεν μπορεί να περιγραφεί γραμμικά (ευθεία γραμμή, επίπεδο, γραμμικό υπερπλάνο). Σ' αυτές τις περιπτώσεις αναπτύσσεται μια μη γραμμική σχέση μεταξύ των μεταβλητών, η οποία μπορεί να περιγραφεί από ένα μη γραμμικό σχήμα. Η γραμμή αυτή (περίπτωση Y και X) θα είναι μια καμπύλη «καλής προσαρμογής» στη γενική σχέση των μεταβλητών. Ουσιαστικά δηλαδή, πέραν του διαφορετικού σχήματος της περιγραφής των σχέσεων, η μη γραμμική παλινδρόμηση δεν διαφέρει από την γραμμική.

Γίνεται φανερό ότι θα πρέπει να επιλέξουμε ποια καμπύλη θα προσαρμοζεται καλύτερα στα δεδομένα μας, διότι δεν είναι όλες του ίδιου σχήματος. Η επιλογή της κατάλληλης συνάρτησης που θα περιγράφει την καμπύλη δεν είναι εύκολη υπόθεση και πρέπει να λαμβάνουμε πάντοτε υπόψη τη φύση των μεταβλητών. Οποσδήποτε μια διαγραμματική αναπαράσταση της σχέσης θα μας βοηθήσει περισσότερο στην επιλογή της κατάλληλης συνάρτησης.

Εν συνεχεία, παραθέτονται μερικά μη γραμμικά σχήματα με τις αντίστοιχες συναρτήσεις που τα περιγράφουν:



Πολυώνυμο δευτέρου βαθμού

Όταν ο εκθέτης του αγνώστου X μιας συνάρτησης ισούται με την μονάδα, τότε λέμε ότι η συνάρτηση είναι πρώτου βαθμού. Όταν ο άγνωστος έχει ως δύναμη τον αριθμό δύο, τότε λέμε ότι η συνάρτηση είναι δευτέρου βαθμού και εκφράζεται με τον τύπο $Y = A + BX + CX^2$. Αυτή η δευτέρου βαθμού καμπύλη, έχει μόνο μία αλλαγή διεύθυνσης, όπως βλέπουμε στο αντίστοιχο σχήμα, ανεξαρτήτως των τιμών των συντελεστών A , B , C ή των τιμών του X . Επίσης παρουσιάζεται συμμετρική σε σχέση με το υψηλότερο ή το χαμηλότερο σημείο της. Τα πιο πολλά όμως δεδομένα δεν μπορούν να έχουν μια τέτοια συμμετρία. Πολλές φορές, σαν ενδιάμεση λύση, χρησιμοποιούμε μόνο το ένα κομμάτι της συμμετρικής καμπύλης.

Η καμπύλη που περιγράφει την $Y = A + BX + CX^2 + DX^3$ έχει μόνο μία αλλαγή του σχήματος, με δύο μέγιστα και δύο ελάχιστα. Όσο δε, αυξάνεται η δύναμη τόσο περισσότερα κρίσιμα (ακραία) σημεία εμφανίζονται. Ένα πολυώνυμο k -τάξεως περιλαμβάνει $k-1$ κορυφές και κοιλάτες. Εξ αιτίας λοιπόν, της δυσκολίας που προκύπτει λόγω της ιδιαιτερότητας του σχήματος των εξισώσεων μεγάλου βαθμού, λίγα είναι τα δεδομένα που υπερβαίνουν την τρίτου βαθμού εξίσωση παλινδρόμησης.

Υπενθυμίζουμε ότι στην εξίσωση $Y = A + BX + CX^2$ το A είναι το σημείο τομής με τον άξονα Y , το B η κλίση της καμπύλης και το C δείχνει την διεύθυνση καθώς και την καμπυλότητα της γραμμής.

Έστω ότι έχει εκτιμηθεί ότι η εξίσωση παλινδρόμησης των δεδομένων ενός πληθυσμού εκφράζεται από το πολυώνυμο $Y_c = a + bX + cX^2$. Οι συντελεστές - εκτιμητές a , b και c βρίσκονται με τον ίδιο τρόπο, όπως και στην απλή γραμμική παλινδρόμηση. Παίρνουμε, δηλαδή, διαδοχικά την μερική παράγωγο ως προς a , b και c και την εξισώνουμε με μηδέν. Οι ρίζες του συστήματος που προκύπτει θα είναι οι τιμές των a , b και c .

Οι εξισώσεις του συστήματος που προκύπτουν από την διαδοχική παραγωγή του πολυωνύμου $Y_c = a + bX + cX^2$ έχουν την μορφή:

$$\sum Y = na + b \sum X + c \sum X^2$$

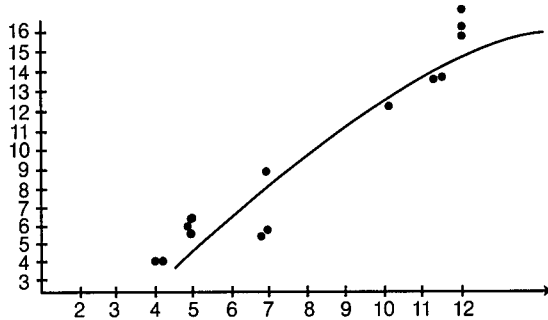
$$\sum XY = a \sum X + b \sum X^2 + c \sum X^3$$

$$\text{και } \sum X^2 Y = a \sum X^2 + b \sum X^3 + c \sum X^4$$

Παράδειγμα 2.1. Έστω ο πίνακας των δεδομένων των μεταβλητών X και Y . Δίπλα από αυτές της μεταβλητές σχηματίζουμε τις παραστάσεις του συστήματος.

X	Y	X^2	X^3	X^4	XY	X^2Y	Y^2
5	6	25	125	625	30	150	36
4	4	16	64	256	20	80	25
7	6	49	343	2401	42	294	36
11	13	121	1331	14641	143	1573	169
6	6	36	216	1296	36	216	36
7	6	49	343	2401	42	294	36
7	8	49	343	2401	56	392	64
4	5	16	64	256	20	80	25
11	14	121	1331	14641	154	1694	196
10	12	100	1000	10000	120	1200	144
12	15	144	1728	20736	180	2160	225
9	11	81	729	6561	99	891	121
8	9	64	512	4096	72	576	81
7	8	49	343	2401	56	392	64
5	6	25	125	625	30	150	36
12	16	144	1728	20736	192	2304	256
125	146	1089	10325	104073	1292	12446	1550

Η σχηματιζόμενη σχέση που δημιουργείται μεταξύ των X και Y αναπαρίσται στο διάγραμμα:



Αντικαθιστώντας στις εξισώσεις του συστήματος τις αντίστοιχες τιμές του πίνακα, έχουμε:

$$146 = a \cdot 16 + b \cdot 125 + c \cdot 1089 \quad \textcircled{1}$$

$$1292 = a \cdot 125 + b \cdot 1089 + c \cdot 10325 \quad \textcircled{2}$$

$$12446 = a \cdot 1089 + b \cdot 10325 + c \cdot 104073 \quad \textcircled{3}$$

Λύνουμε το σύστημα και έχουμε:

$$\text{Από την } \textcircled{1} \text{ και } \textcircled{2} \quad a = \frac{146 - 125b - 1089c}{16} \quad \text{και} \quad b = \frac{151,375 - 1817,2c}{112,44}.$$

Εκ των ανωτέρω η πρώτη παράσταση γίνεται $a = 58,1875c - 1,425$.

Τέλος από την $\textcircled{3}$ έχουμε $c = 0,1006$. Αντικαθιστώντας παίρνουμε $a = 4,43$ και $b = -0,2757$

Επομένως η εξίσωση παλινδρόμησης γίνεται:

$$Y_c = 4,43 - 0,2757X + 0,1006X^2$$

Οι εκτιμώμενες τιμές του Y , δηλαδή οι Y_c , δίνονται από τον πίνακα:

X	Y_c
3	4,5083
4	4,9368
5	5,57
6	6,3974
7	7,4295
8	8,6628

X	Y_c
9	10,0973
10	11,733
11	13,57
12	15,608
13	17,8473
14	20,2976

Τυπική απόκλιση της παλινδρόμησης

Η απόκλιση του πληθυσμού από την καμπύλη παλινδρόμησης εκτιμάται ως ε-

ξής: $S_{Y';X,X^2} = \sqrt{\frac{\sum (Y - Y_c)^2}{n - m}}$, όπου n ο αριθμός των ατόμων και m ο βαθμός του πολυωνύμου της παλινδρόμησης συν ένα.

Η ανωτέρω μαθηματική έκφραση γίνεται:

$$S_{Y';X,X^2} = \sqrt{\frac{\sum (Y^2 - 2YY_c + Y_c^2)}{n - m}}. \text{ Όμως } E(Y Y_c) = Y_c E(Y) = Y_c^2, \text{ διότι } E(Y) = Y_c.$$

Επομένως έχουμε:

$$\begin{aligned} S_{Y';X,X^2} &= \sqrt{\frac{\sum (Y^2 - Y Y_c)}{n - m}} = \sqrt{\frac{\sum (Y^2 - Y(a + bX + cX^2))}{n - m}} = \\ &= \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY - c \sum X^2 Y}{n - m}}. \end{aligned}$$

Η τυπική απόκλιση παλινδρόμησης του προηγούμενου παραδείγματος 2.1 γίνεται:

$$\begin{aligned} S_{Y';X,X^2} &= \sqrt{\frac{1550 - 4,43 \times 146 + 0,2757 \times 1292 - 0,1006 \times 12446}{16 - 3}} = \\ &= \sqrt{0,5659} \Rightarrow S_{Y';X,X^2} = 0,75 \end{aligned}$$

Συντελεστής προσδιορισμού

Όπως στην γραμμική συσχέτιση, έτσι και στην μη γραμμική, ο συντελεστής προσδιορισμού είναι ο λόγος της διακύμανσης που εξηγείται (από την σχέση που βρέθηκε) προς την ολική. Επομένως εκφράζεται από την παράσταση:

$$r_{Y';X,X^2}^2 = 1 - \left[\frac{\sum (Y - Y_c)^2}{\sum (Y - \bar{Y})^2} \right] \text{ ή χρησιμοποιούμε τον εκτιμητή του } \tilde{r}_{Y';X,X^2}^2, \text{ ο οποίος εκφράζεται: } \tilde{r}_{Y';X,X^2}^2 = 1 - \left[\frac{\sum (Y - Y_c)^2 / (n - m)}{\sum (Y - \bar{Y})^2 / (n - 1)} \right].$$

Το ολικό άθροισμα τετραγώνων μπορεί να εκφραστεί και ως ακολούθως:

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

Το εξηγήσιμο άθροισμα τετραγώνων εκφράζεται:

$$\sum (\bar{Y} - Y_c)^2 = a \sum Y + b \sum XY + c \sum X^2 Y - \bar{Y} \sum Y$$

Το δε, μη εξηγήσιμο άθροισμα τετραγώνων εκφράζεται:

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

Λαμβάνοντας υπόψη τα δεδομένα του προηγούμενου παραδείγματος έχουμε, με $\bar{Y} = 9,125$.

$$\text{Ολικό άθροισμα τετραγώνων: } \sum (Y - \bar{Y})^2 = 1550 - 9,125 \times 146 = 217,75$$

$$\begin{aligned} \text{Εξηγήσιμο άθροισμα τετραγώνων: } \sum (\bar{Y} - Y_c)^2 &= 4,43 \times 146 - 0,2757 \times \\ &\times 1292 + 0,1006 \times 12446 - 9,125 \times 146 = 210,3932 \end{aligned}$$

$$\begin{aligned} \text{Μη εξηγήσιμο άθροισμα τετραγώνων: } \sum (Y - Y_c)^2 &= 217,75 - 210,3932 = \\ &= 7,3568 \end{aligned}$$

Επομένως ο συντελεστής προσδιορισμού υπολογίζεται σύμφωνα με τους παραπάνω υπολογισμούς: $\tilde{r}_{Y;X}^2 = 1 - \left[\frac{\sum (Y - Y_c)^2 / (n - m)}{\sum (Y - \bar{Y})^2 / (n - 1)} \right] =$

$$= 1 - \frac{7,3568 / 13}{217,75 / 15} = 1 - \frac{0,57}{14,52} = 0,96$$

Από τα ανωτέρω συμπεραίνουμε ότι το 96% της μεταβλητότητας του Y οφείλεται στη μεταβλητότητα του X . Επομένως ο συντελεστής συσχέτισης, για μη γραμμικά μοντέλα, ο οποίος ισούται με την τετραγωνική ρίζα του συντελεστή προσδιορισμού, είναι:

$$r_{Y;X} = \sqrt{0,96} = 0,98.$$

Επιλογή βαθμού του πολυνύμου

Όταν χρησιμοποιούμε μια μη γραμμική έκφραση ως εκτίμηση μιας σχέσης, δεν γνωρίζουμε, εάν μια άλλη μεγαλύτερου βαθμού, περιγράφει καλύτερα τη

σχέση. Για να το διαπιστώσουμε θα πρέπει να ελέγξουμε αν ο συντελεστής προσδιορισμού της παλινδρόμησης που προσδιορίζεται από το πολυώνυμο $υ$ -ψηλότερου βαθμού, είναι μεγαλύτερος από τον αντίστοιχο συντελεστή του πολυωνύμου με τον χαμηλότερο βαθμό. Ήτοι αν $r_{Y';X,X^2}^2 > r_{Y';X}^2$.

Ο συντελεστής προσδιορισμού $r_{Y';X}^2$, περιγράφει την αναλογία της εξηγήσιμης μεταβλητότητας, δηλαδή αυτής που οφείλεται στην ανεξάρτητη μεταβλητή, όπως εκφράζεται από την γραμμική σχέση, προς την ολική. Ο συντελεστής προσδιορισμού $r_{Y';X,X^2}^2$ περιγράφει το ίδιο πράγμα αλλά όπως εκφράζεται από την μη γραμμική σχέση (δευτέρου βαθμού).

Αν οι δύο συντελεστές δεν διαφέρουν στατιστικώς σημαντικά, τότε η μη γραμμική σχέση (δευτέρου βαθμού) δεν προσφέρεται περισσότερο για την μέτρηση της σχέσης μεταξύ των μεταβλητών, από την γραμμική σχέση.

Μερικός συντελεστής προσδιορισμού

Ο μερικός συντελεστής προσδιορισμού δευτέρου βαθμού $r_{Y';X^2,X}^2$, δείχνει το ποσοστό της ανεξήγητης μεταβλητότητας, όταν χρησιμοποιούμε τη γραμμική σχέση $(1 - r_{Y';X}^2)$, που εξηγείται τώρα από την χρήση της μη γραμμικής σχέσης $(r_{Y';X,X^2}^2 - r_{Y';X}^2)$. Εκφράζεται με την σχέση ορισμού του, που είναι:

$$r_{Y';X^2,X}^2 = \frac{r_{Y';X,X^2}^2 - r_{Y';X}^2}{1 - r_{Y';X}^2}.$$

Η μεταβλητή αυτή ακολουθεί την t -κατανομή με τιμή $t = \sqrt{\frac{r_{Y';X^2,X}^2 (n-3)}{1 - r_{Y';X^2,X}^2}}$.

Ακολουθώντας πραγματοποιούμε τον έλεγχο κανονικά, όπως π.χ. στον συντελεστή συσχέτισης.

Ο ίδιος προβληματισμός είναι πιθανόν να αναπτυχθεί και για πολυώνυμα μεγαλύτερου βαθμού. Π.χ. είναι προτιμότερο να χρησιμοποιήσουμε ένα πολυώνυμο τρίτου βαθμού, απ' ένα πολυώνυμο δευτέρου βαθμού; Εκφράζεται δε, αυτή η απορία με την σχέση $r_{Y';X,X^2,X^3}^2 > r_{Y';X,X^2}^2$

Προσδιορίζουμε επίσης τον μερικό συντελεστή προσδιορισμού:

$$r_{Y';X^3,X,X^2}^2 = \frac{r_{Y';X,X^2,X^3}^2 - r_{Y';X,X^2}^2}{1 - r_{Y';X,X^2}^2}.$$

Στο προηγούμενο παράδειγμα 2.1 το $r_{Y';XX^2}^2 = 0,96$ αντιστοιχεί σε δευτέρου βαθμού σχέση. Αν τώρα θεωρήσουμε την σχέση, ως γραμμική και υπολογίσουμε τον αντίστοιχο συντελεστή προσδιορισμού, έχουμε $r_{Y';X}^2 = 0,931$.

Για να επιλέξουμε ποια από τις δύο παλινδρομήσεις ταιριάζει καλύτερα στα δεδομένα μας, υπολογίζουμε τον μερικό συντελεστή προσδιορισμού.

$$r_{Y';X^2,X}^2 = \frac{r_{Y';XX^2}^2 - r_{Y';X}^2}{1 - r_{Y';X}^2} = \frac{0,96 - 0,931}{1 - 0,931} = 0,42$$

Άρα το 42%, της μεταβλητότητας του Y , που δεν είχε ερμηνευτεί, τώρα με την προσθήκη του X^2 , ερμηνεύεται.

$$\text{Σχηματίζουμε το } t = \sqrt{\frac{r_{Y';X^2,X}^2 (n-3)}{1 - r_{Y';X^2,X}^2}} = \sqrt{\frac{0,42 (16-3)}{1 - 0,42}} = \sqrt{\frac{5,46}{0,58}} = 3,068.$$

Με 13 βαθμούς ελευθερίας, στο επίπεδο σημαντικότητας $\alpha = 0,05$, η κρίσιμη τιμή είναι $t_0 = 1,771$. Εφ' όσον $1,771 < 3,068$ η δευτέρου βαθμού καμπύλη εξηγεί, κατά ένα ποσοστό σημαντικά μεγαλύτερο απ' αυτό της γραμμικής καμπύλης, την ολική μεταβλητότητα του Y .

Παλινδρόμηση μετά από μετασχηματισμό των μεταβλητών

Όπως είδαμε, ανάλογες του βαθμού του πολυνύμου, είναι και οι δυσκολίες που προκύπτουν στην εκτίμηση των συντελεστών των μη γραμμικών εξισώσεων. Για τον λόγο αυτό, πολλές φορές, χρησιμοποιούμε κάποιον μετασχηματισμό των δεδομένων μας, κυρίως για να μπορέσουμε να τα μετατρέψουμε σε μια γραμμική μορφή. Αυτός ο μετασχηματισμός μπορεί να πραγματοποιηθεί ή μέσω των λογαρίθμων ή μέσω των αντίστροφων συναρτήσεων ή μέσω εκθετικών μορφών συναρτήσεων ή τέλος μέσω άλλων τινών μετασχηματισμών.

Ένα πρόβλημα που προκύπτει είναι ότι οι μετασχηματισμοί αυτοί επιφέρουν τροποποιήσεις στις κλίμακες των αξόνων του καρτεσιανού χώρου. Π.χ. ένας μετασχηματισμός του Y σε $\log Y$ συμπιέζει τις μικρές τιμές και διαστέλλει τις μεγάλες. Το ίδιο συμβαίνει στον άξονα των X , όταν λογαριθμίσουμε το X . Οι μετασχηματισμοί αυτοί επιφέρουν καμπυλώσεις στις ευθείες που η κλίση τους εξαρτάται από το b . Στις πιο πολλές περιπτώσεις όμως, ένα πολυώνυμο μεγάλου βαθμού λογαριθμείται και στα δύο μέρη του. Αυτό βοηθάει στο να έχουμε μια ασυμπτωτική μορφή της λογαριθμικής γραμμής, όταν πλησιάζει προς το μηδέν διότι οι λογάριθμοι δεν δέχονται αρνητικές τιμές.

Ο κάθε μετασχηματισμός παρέχει διαφορετικά πλεονεκτήματα. Π.χ. ο μετασχηματισμός $Y = \frac{1}{a+bX}$ δημιουργεί μια καμπύλη που είναι ασυμπτωτική προς τον άξονα X , ο μετασχηματισμός $Y = a + b\left(\frac{1}{X}\right)$ μια ασυμπτωτική ως προς τον άξονα τον Y και ο $\frac{1}{Y} = a + b\left(\frac{1}{X}\right)$ μια ασυμπτωτική παράλληλη και προς τους δυο άξονες.

Παρ' όλα τα ανωτέρω όμως, ο βασικός μας σκοπός με την χρήση των μετασχηματισμών παραμένει η μετατροπή των καμπυλών σε γραμμικής μορφής παράσταση. Υπάρχουν περιπτώσεις που ο μετασχηματισμός δημιουργεί καλύτερες «μαθηματικές» συνθήκες για την χρήση και επίλυση των προβλημάτων. Π.χ. η εύρεση του ελαχίστου του αθροίσματος τετραγώνων ενός πολυνύμου εκθετικής (exponentational) μορφής δεν επιλύεται. Ο μετασχηματισμός όμως, της εκθετικής μορφής σε λογαριθμικής, μετατρέπει το εκθετικό πολυνύμο σε γραμμικό, που η επίλυσή του είναι γνωστή.

Ας δούμε στην απλή παλινδρόμηση πως χρησιμοποιείται ένας λογαριθμικός μετασχηματισμός. Εάν τα δεδομένα αναπαριστώνται σ' ένα διάγραμμα, που και οι δύο άξονες είναι λογαριθμικοί και το νέφος των σημείων παρουσιάζει γραμμικότητα, τότε ξέρουμε ότι χρειαζόμαστε τον λογαριθμικό μετασχηματισμό και στα δυο μέρη. Θα έχει δε τη μορφή $\log Y = \log a + b \log X$.

Για να βρούμε τα $\log a$ και b χρησιμοποιούμε το κάτωθι σύστημα:

$$\sum \log Y = n \times \log a + b \times \sum \log X \quad \text{και}$$

$$\sum (\log X \times \log Y) = \log a \times \sum \log X + b \times \sum (\log X)^2.$$

Η συνολική μεταβλητότητα ορίζεται ως ακολούθως:

$$\sum (\log y)^2 = \sum (\log Y)^2 - (\overline{\log Y}) \sum \log Y, \quad \text{όπου} \quad (\overline{\log Y}) = \left[\frac{\sum \log Y}{n} \right], \quad \text{ενώ το}$$

ερμηνεύσιμο άθροισμα τετραγώνων είναι:

$$\sum (\log y)_c^2 = \log a \times \sum \log Y + b \sum (\log X \times \log Y) - (\overline{\log Y}) \times \sum \log Y.$$

Η ανεξήγητη μεταβλητότητα υπολογίζεται από την διαφορά των δύο ανωτέρων.

Στο παράδειγμα 2.1 έχουμε:

X	Y	logX	logY	logXlogY	$(\log X)^2$	$(\log Y)^2$
5	6	0,70	0,78	0,546	0,49	0,6084
4	5	0,60	0,70	0,42	0,36	0,49
7	6	0,85	0,78	0,663	0,7225	0,6084
11	13	1,04	1,11	1,1544	1,0816	1,2321
6	6	0,78	0,78	0,6084	0,6084	0,6084
7	6	0,85	0,78	0,663	0,7225	0,6084
7	8	0,85	0,90	0,765	0,7225	0,81
4	5	0,60	0,70	0,42	0,36	0,49
11	14	1,04	1,15	1,196	1,0816	1,3225
10	12	1,00	1,08	1,08	1	1,1664
12	15	1,08	1,18	1,2744	1,1664	1,3924
9	11	0,95	1,04	0,988	0,9525	1,0816
8	9	0,90	0,95	0,855	0,81	0,9025
7	8	0,85	0,90	0,765	0,7225	0,81
5	6	0,70	0,85	0,595	0,49	0,7225
12	16	1,08	1,20	1,296	1,1664	1,44
125	146	13,87	14,88	13,2892	12,4569	14,2936

Σχηματίζουμε το σύστημα των εξισώσεων που έχει τις τιμές:

$$\begin{cases} 14,88 = 16 \times \log a + 13,87 \times b \\ 13,2892 = 13,87 \times \log a + 12,4569 \times b \end{cases} \Rightarrow \begin{cases} \log a = 0,93 - 0,87 \times b \\ 0,3901 = 0,39 \times b \end{cases}$$

$$\begin{cases} \text{το οποίο έχει λύσεις} \\ b \approx 1 \end{cases} \left. \begin{array}{l} \log a = 0,06 \\ b \approx 1 \end{array} \right\}$$

Η εξίσωση λοιπόν θα πάρει τη μορφή:

$$(\log Y)_c = \log a + b \times \log X \Rightarrow (\log Y)_c = 0,06 + 1 \times \log X$$

Το συνολικό άθροισμα τετραγώνων θα είναι:

$$\sum (\log Y)^2 - (\overline{\log Y}) \sum \log Y = 14,2936 - 0,93 \times 14,88 = 0,4552, \text{ διότι}$$

$$(\overline{\log Y}) = \left[\frac{(\sum \log Y)}{n} \right] = \frac{14,88}{16} = 0,93$$

Το εξηγήσιμο άθροισμα τετραγώνων είναι:

$$\log a \times \sum \log Y + b \sum (\log X \times \log Y) - (\overline{\log Y}) \times \sum \log Y =$$

$$= 0,06 \times 14,88 + 1 \times 13,2892 - 0,93 \times 14,88 = 0,3436$$

Ο συντελεστής προσδιορισμού υπολογίζεται, ως ο λόγος του εξηγήσιμου αθροίσματος τετραγώνων προς το συνολικό.

$$r_{\log X, \log Y} = \frac{0,3436}{0,4552} = 0,755$$

Πρέπει να έχουμε όμως υπ' όψη μας ότι η εξίσωση εκτιμάει τον $\log Y$ και όχι τον Y . Επομένως θα πρέπει να απολογαριθμίσουμε την εξίσωση, για συγκεκριμένες τιμές του X . Δηλαδή για συγκεκριμένη τιμή του $X=7$ έχουμε:

$$(\log Y)_c = \log a + b \times \log X \Rightarrow (\log Y)_c = 0,06 + 1 \times 0,85 \Rightarrow (\log Y)_c = 0,91, \text{ διότι γνωρίζουμε ότι } \log X = 0,85.$$

Υπολογίζουμε το antilog του 0,91, για να βρούμε την προβλεπόμενη τιμή του Y από την τιμή $X=7$.

Από τους πίνακες των λογαρίθμων βρίσκουμε ότι: $\text{antilog}(0,91) = 8,13$.