Απλή Γραμμική Παλινδρόμηση Ι

1. Εισαγωγή

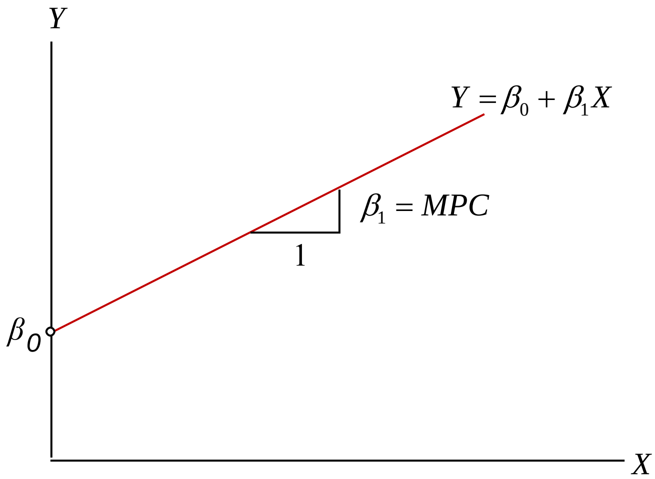
Έστω ότι θέλουμε να ερευνήσουμε εμπειρικά τη σχέση που υπάρχει ανάμεσα στις δαπάνες κατανάλωσης (Y) και στο διαθέσιμο εισόδημα (X), των οικογενειών.

Σύμφωνα με την Κεϋνσιανή θεωρία, η κατανάλωση αυξάνεται καθώς αυξάνεται το εισόδημα, όχι όμως τόσο πολύ όσο το εισόδημα.
Η απλούστερη μαθηματική σχέση που μπορεί να περιγράψει τον παραπάνω νόμο, είναι ένα γραμμικό μοντέλο της μορφής

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

$$\beta_0 > 0$$

$$0 < \beta_1 < 1$$



Ο συντελεστής β_0 (τετμημένη της ευθείας), μας δείχνει την κατανάλωση, όταν το διαθέσιμο εισόδημα είναι μηδέν. Σύμφωνα με την οικονομική θεωρία, ακόμη και σε μηδενικό επίπεδο εισοδήματος, υπάρχει κατανάλωση και, επομένως, πρέπει $\beta_0 > 0$.

Ο συντελεστής β_1 είναι η οριακή κατανάλωση (**M**arginal **P**ropensity of **C**onsume) , δηλαδή $\beta_1 = \frac{dy}{dx}$, και μας δείχνει την μεταβολή της κατανάλωσης που αντιστοιχεί σε μοναδιαία μεταβολή του εισοδήματος. Σύμφωνα με τον Keynes, η κατανάλωση αυξάνεται όταν αυξάνεται το εισόδημα – επομένως πρέπει $\beta_1 > 0$ – όχι όμως τόσο πολύ όσο το εισόδημα, επομένως πρέπει $\beta_1 < 1$.

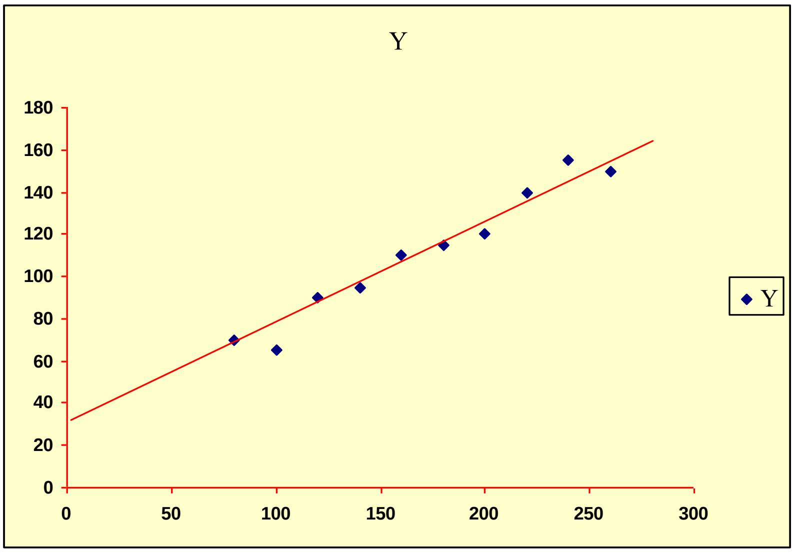
Έστω, τώρα, ότι θέλουμε να μελετήσουμε τη συνάρτηση της κατανάλωσης στα Ελληνικά νοικοκυριά. Στόχος μας είναι ο προσδιορισμός των συντελεστών $\beta_{0,}\ \beta_1$, βάσει ενός δείγματος παρατηρήσεων $(Y_t,\ X_t)$ $t=1,2,3,\ldots,T$, που αφορούν τη μηνιαία κατανάλωση Y και το μηνιαίο εισόδημα X (1 χρηματική μονάδα = 10€) και το οποίο λαμβάνεται σύμφωνα με τους κανόνες της δειγματοληψίας από το σύνολο (πληθυσμό) των Ελληνικών νοικοκυριών. Έστω, λοιπόν, ότι διαθέτουμε το παρακάτω δείγμα:

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Εξαιτίας της γραμμικότητας του μοντέλου, περιμένουμε ότι όλα τα ζεύγη των παρατηρήσεων $(Y_t,\ X_t)$ $t=1,2,3,\ldots,T$ θα βρίσκονται πάνω σε μια ευθεία, δηλαδή για κάθε $t=1,2,\ \ldots,T$ να ισχύει η σχέση

$$Y_t=\beta_0+\beta_1 X_t \qquad (1)$$

Όμως, από το διάγραμμα διασποράς βλέπουμε ότι κάτι τέτοιο δεν ισχύει.



Γιατί συμβαίνει αυτό;

Η σχέση $Y_t = \beta_0 + \beta_1 X_t$ είναι μια σχέση **προσδιοριστική** (deterministic). Αυτό σημαίνει ότι όλες οι οικογένειες με το ίδιο διαθέσιμο εισόδημα, έχουν την ίδια κατανάλωση. Στην πραγματικότητα όμως αυτό δεν ισχύει. Και αυτό επειδή, εκτός από το διαθέσιμο εισόδημα, υπάρχουν και άλλοι παράγοντες που επηρεάζουν το εισόδημα. Για παράδειγμα, το πλήθος των μελών της οικογένειας, η ηλικία τους, κ.α.

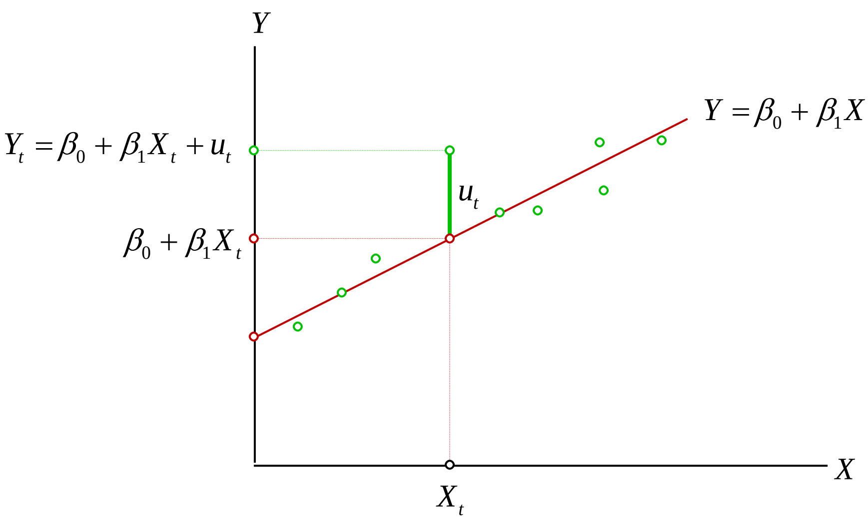
Παρατηρούμε, ωστόσο, ότι τα σημεία συσσωρεύονται γύρω από μια «νοητή» ευθεία, επομένως η υπόθεση της γραμμικής σχέσης μεταξύ του εισοδήματος και της κατανάλωσης δεν είναι μια υπόθεση που πρέπει να απορριφθεί. Πρέπει όμως να βρούμε έναν τρόπο, ώστε να συμπεριλάβουμε στο μοντέλο μας τις αποκλίσεις των σημείων από την ευθεία.

Αυτό γίνεται με την προσθήκη μιας τυχαίας μεταβλητής u_t, οπότε η προσδιοριστική σχέση (1) μετατρέπεται στη **στοχαστική σχέση**

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad (2)$$

Η σχέση αυτή μας λέει ότι η κατανάλωση σχετίζεται γραμμικά με το εισόδημα, όμως η σχέση μεταξύ αυτών των δύο δεν είναι ακριβής αλλά διαταράσσεται από κάποιους παράγοντες, τους οποίους δεν μπορούμε ή δεν θέλουμε να συμπεριλάβουμε, στο μοντέλο. Όλοι αυτοί οι παράγοντες αναπαριστάνονται από την τυχαία μεταβλητή u_t, η οποία ονομάζεται **διαταρακτικός όρος** (disturbance term), ακριβώς επειδή διαταράσσει την προσδιοριστική σχέση $Y_t = \beta_0 + \beta_1 X_t$.

Η στοχαστική σχέση $Y_t = \beta_0 + \beta_1 X_t + u_t$, μας λέει ότι για ένα δεδομένο επίπεδο εισοδήματος X_t, περιμένουμε η κατανάλωση να είναι μια τιμή Y_t, η οποία βρίσκεται σε μια κοντινή, αλλά τυχαία απόσταση u_t, από την ευθεία $Y = \beta_0 + \beta_1 X$. Μια τιμή, η οποία μπορεί να βρίσκεται οπουδήποτε γύρω, ή και ακόμα και ακριβώς επάνω, στην ευθεία $Y = \beta_0 + \beta_1 X$.



Η εξίσωση $Y_t = \beta_0 + \beta_1 X_t + u_t$ ονομάζεται **μοντέλο γραμμικής παλινδρόμησης**, και αυτό που θα μας απασχολήσει είναι το πώς, από τα διαθέσιμα δεδομένα – το δείγμα – μπορούμε να παράγουμε τις εκτιμήσεις (δηλαδή, τις αριθμητικές τιμές) των παραμέτρων του μοντέλου.

Η σημασία του διαταρακτικού όρου

Ο διαταρακτικός όρος u_t, είναι ένα υποκατάστατο ή μία προσέγγιση όλων εκείνων των μεταβλητών που παραλείπονται από το μοντέλο, οι οποίοι όμως επηρεάζουν την Y. Προκύπτει, λοιπόν, το ερώτημα: Γιατί δεν συμπεριλαμβάνουμε όλες αυτές τις μεταβλητές αναλυτικά στο μοντέλο, χρησιμοποιώντας, για παράδειγμα, ένα μοντέλο πολλαπλής παλινδρόμησης; Υπάρχουν αρκετοί λόγοι που δεν το κάνουμε.

1. Η θεωρία που προσδιορίζει την συμπεριφορά της Y μπορεί να είναι, και τις περισσότερες φορές πράγματι είναι, ελλιπής. Μπορεί, για παράδειγμα, να γνωρίζουμε με βεβαιότητα ότι το εισόδημα επηρεάζει την κατανάλωση, αλλά μπορεί να αγνοούμε ή να μήν είμαστε βέβαιοι για κάποιες άλλες μεταβλητές που «προσβάλουν» την κατανάλωση. Έτσι, η μεταβλητή u_t, μπορεί να χρησιμοποιηθεί ως υποκατάστατο όλων αυτών των μεταβλητών που εξαιρούνται ή παραλείπονται από το μοντέλο.
2. Ακόμη και αν γνωρίζουμε ποιές είναι κάποιες από τις υπόλοιπες μεταβλητές που επηρεάζουν την συμπεριφορά της Y, μπορεί να μην έχουμε ποσοτικές πληροφορίες γι΄ αυτές. Στις εμπειρικές έρευνες, το φαινόμενο να μήν διαθέτουμε τα ιδανικά δεδομένα είναι πολύ συχνό. Για παράδειγμα, σε μια έρευνα για την κατανάλωση, θα θέλαμε, κατ΄ αρχήν, να συμπεριλάβουμε σαν μεταβλητή στο μοντέλο μας τον πλούτο, της οικογένειας. Όμως, γενικά, δεν διαθέτουμε πληροφορίες για την μεταβλητή αυτή. Έτσι, όσο και αν θεωρητικά είναι σημαντική, αναγκαζόμαστε να την εξαιρέσουμε.
3. Ας υποθέσουμε, ότι στο μοντέλο κατανάλωσης – εισοδήματος, θέλουμε εκτός από το εισόδημα να συμπεριλάβουμε το πλήθος των μελών της οικογένειας, το θρήσκευμα, τη μόρφωση, την γεωγραφική περιοχή στην οποία ζούνε, παράγοντες για τους οποίους μπορούμε να θεωρήσουμε ότι επηρεάζουν την κατανάλωση. Όμως, αρκετές φορές, η κοινή επίδραση όλων αυτών των μεταβλητών μπορεί να είναι τόσο μικρή ώστε να μην χρειάζεται να τις συμπεριλάβουμε αναλυτικά στο μοντέλο. Αυτή η κοινή επίδραση μπορεί κατόπιν να θεωρηθεί ως μία τυχαία μεταβλητή u_t.
4. Ακόμα και αν καταφέρουμε να συμπεριλάβουμε στο μοντέλο μας όλες τις μεταβλητές, υπάρχουν ακόμη κάποια περιθώρια για την συμπεριφορά της Y να παρουσιάζει κάποια τυχαιότητα την οποία δεν μπορούμε να ερμηνεύσουμε όσο και αν προσπαθήσουμε. Έτσι , θεωρούμε ότι είναι ο διαταρκτικός όρος που αντικατοπτρίζει αυτή την κρυμμένη «τυχαιότητα».
5. Παρόλο που στο μοντέλο της κλασικής παλινδρόμησης υποθέτουμε ότι οι μετρήσεις των μεταβλητών X και Y είναι ακριβείς, στις πράξη τα δεδομένα μπορεί να υπόκεινται σε σφάλματα μετρήσεων. Στην περίπτωση αυτή, ο διαταρακτικός όρος, αναπαριστά τα σφάλματα των μετρήσεων.
6. Τέλος, πάντοτε επιθυμούμε, το μοντέλο που κατασκευάζουμε να είναι όσο το δυνατόν ποιό απλό. Αν μπορούμε να ερμηνεύσουμε την συμπεριφορά της Y, βάσει δύο – τριών μεταβλητών, και αν η θεωρία μας δεν είναι αρκετά ισχυρή ώστε να υποστηρίξει την χρήση και άλλων μεταβλητών, τότε δεν υπάρχει λόγος να τις συμπεριλάβουμε. Ο διαταρακτικός όρος, τώρα, είναι ένα υποκατάστατο όλων αυτών των μεταβλητών που παραλείπουμε από το μοντέλο.

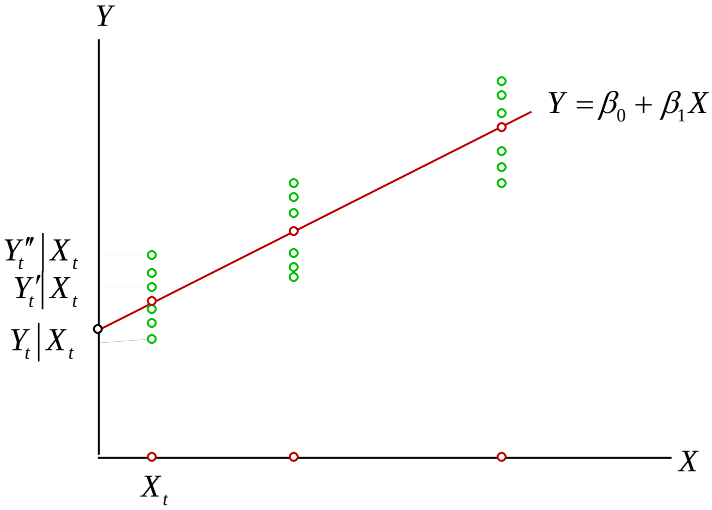
2. Κλασική Γραμμική Παλινδρόμηση

Οι βασικές υποθέσεις

Η σχέση

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad (2)$$

Αποτελεί την οικονομετρική μορφή της γραμμικής σχέσης που υποθέτουμε ότι συνδέει τις μεταβλητές Y και X. Η στοχαστική φύση της σχέσης (2) συνεπάγεται ότι για κάθε τιμή της X δεν υπάρχει μία μόνο τιμή για την Y, αλλά μια ολόκληρη κατανομή τιμών, η οποία εξαρτάται από τον διαταρακτικό όρο.



Μια ολοκληρωμένη εξειδίκευση της παραπάνω γραμμικής σχέσης ανάμεσα στην Y και την X περιγράφεται από τις ακόλουθες υποθέσεις:

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad (3)$$

$$u_t \sim \left(0, \sigma^2\right) \qquad (4)$$

α) u_t είναι μία τυχαία μεταβλητή

β) $Eu_t = 0$

γ) $\operatorname{var}\left(u_t^2\right) = Eu_t^2 = \sigma^2$

$$Cov\left(u_t, u_s\right) = Eu_t u_s = 0, \qquad t \neq s \qquad (5)$$

Η μεταβλητή X δεν είναι στοχαστική. Οι τιμές της παραμένουν σταθερές και δεν είναι όλες ίσες μεταξύ τους. (6)

Οι υποθέσεις (3) – (6) αποτελούν το **υπόδειγμα της κλασικής γραμμικής παλινδρόμησης** (classical linear regression model), στο οποίο μπορούμε να εφαρμόσουμε τις κλασικές μεθόδους για να εκτιμήσουμε τις άγνωστες παραμέτρους β_0, β_1, σ^2.

Η κατανομή της *Y* και η γραμμή παλινδρόμησης

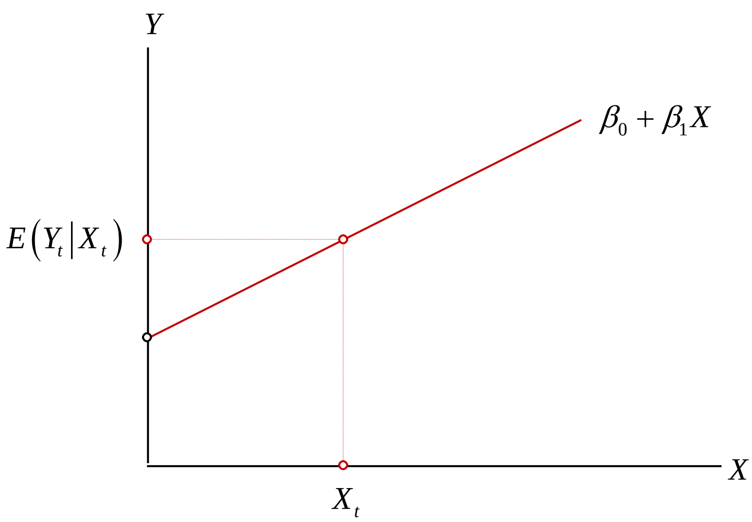
Αποδεικνύεται ότι

$$E\left(Y_t \middle| X_t\right)=\beta_0+\beta_1 X_t \qquad (7)$$

$$Var\left(Y_t \middle| X_t\right)=\sigma^2 \qquad (8)$$

Επομένως, η μεταβλητή Y_t, όπως ορίζεται στο υπόδειγμα που καθορίζουν οι σχέσεις (3) – (6) , ακολουθεί μία κατανομή με μέση τιμή $E\left(Y_t\right)=\beta_0+\beta_1 X_t$ και διακύμανση $Var\left(Y_t\right)=\sigma^2$, ίση με τη διακύμανση του διαταρακτικού όρου.

Το παραπάνω συμπέρασμα, μας λέει ότι η ευθεία $Y=\beta_0+\beta_1 X$ δεν είναι τίποτα άλλο, παρά μια ευθεία πάνω στην οποία βρίσκονται οι δεσμευμένες μέσες τιμές $E\left(Y_t \middle| X_t\right)$ της μεταβλητής Y .



Η σχέση $E\left(Y_t \middle| X_t\right)=\beta_0+\beta_1 X_t$, **ονομάζεται γραμμή παλινδρόμησης στον πληθυσμό** (population regression line) και είναι η σχέση που υπάρχει ανάμεσα στους υπό συνθήκη μέσους της εξαρτημένης μεταβλητής Y , και στις αντίστοιχες τιμές της ανεξάρτητης μεταβλητής X .

Η γραμμή της παλινδρόμησης όμως στον πληθυσμό μας είναι άγνωστη, εφόσον δεν γνωρίζουμε τις τιμές των παραμέτρων $\beta_0,\ \beta_1$. Στόχος μας είναι, και θα δούμε με ποιόν τρόπο φτάνουμε σε αυτόν, από ένα δείγμα παρατηρήσεων $\left(Y_t,\ X_t\right)$ $t=1,2,3,\ldots,T$ να εκτιμήσουμε αυτές τις παραμέτρους.

Έστω, λοιπόν, ότι $\hat{\beta}_0,\ \hat{\beta}_1$ είναι οι εκτιμήσεις των $\beta_0,\ \beta_1$. Τότε

$$\hat{Y}_t=\hat{\beta}_0+\hat{\beta}_1 X_t \qquad (9)$$

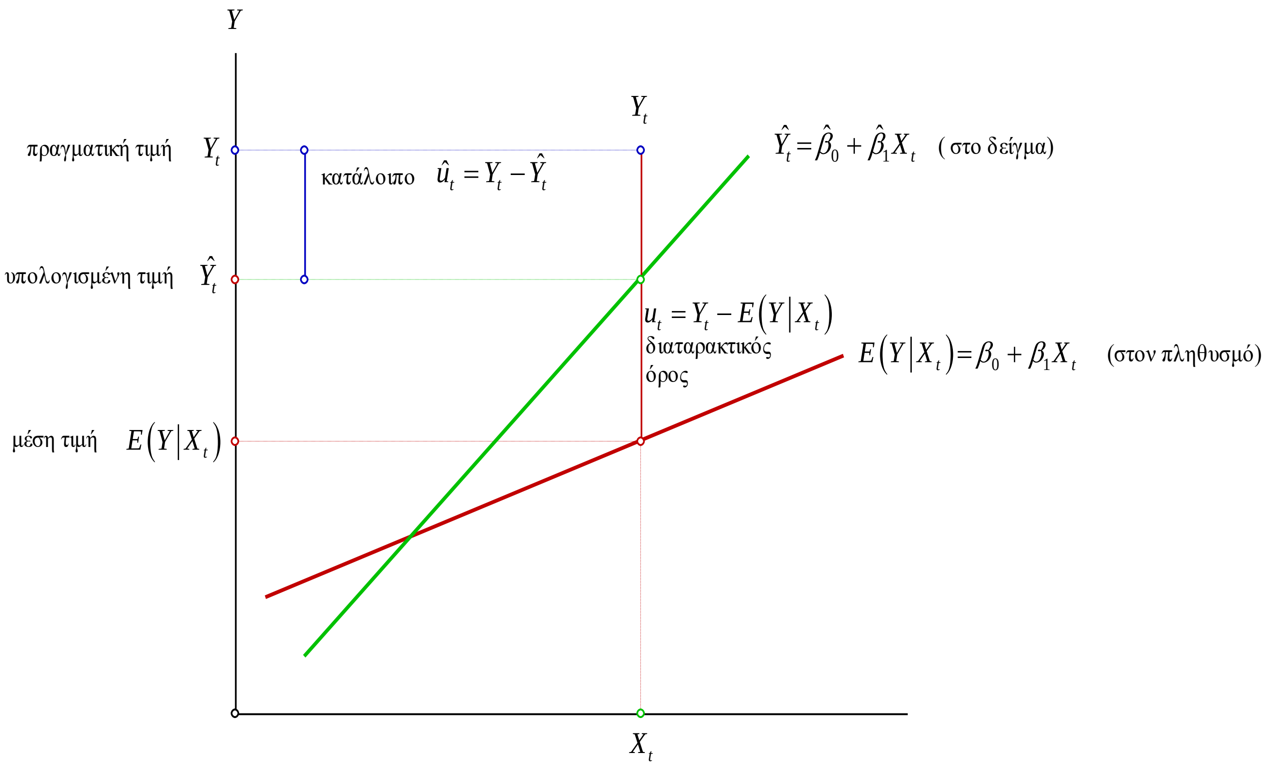
είναι η **γραμμή παλινδρόμησης στο δείγμα** (sample regression line) και η $\hat{Y}_t$ είναι η τιμή της Y που υπολογίζουμε από την γραμμή παλινδρόμησης του δείγματος.

Είναι φανερό ότι οι υπολογισμένες τιμές $\hat{Y}_t$ δεν θα είναι όλες ίσες με τις πραγματικές τιμές Y_t, δηλαδή τις τιμές του δείγματος.

Η διαφορά μεταξύ των πραγματικών τιμών Y_t και των υπολογισμένων τιμών $\hat{Y}_t$, δηλαδή η

$$\hat{u}_t = Y_t - \hat{Y}_t$$

ονομάζεται **κατάλοιπο** (residual) και μπορεί να θεωρηθεί ως **εκτίμηση του διαταρακτικού όρου** u_t.

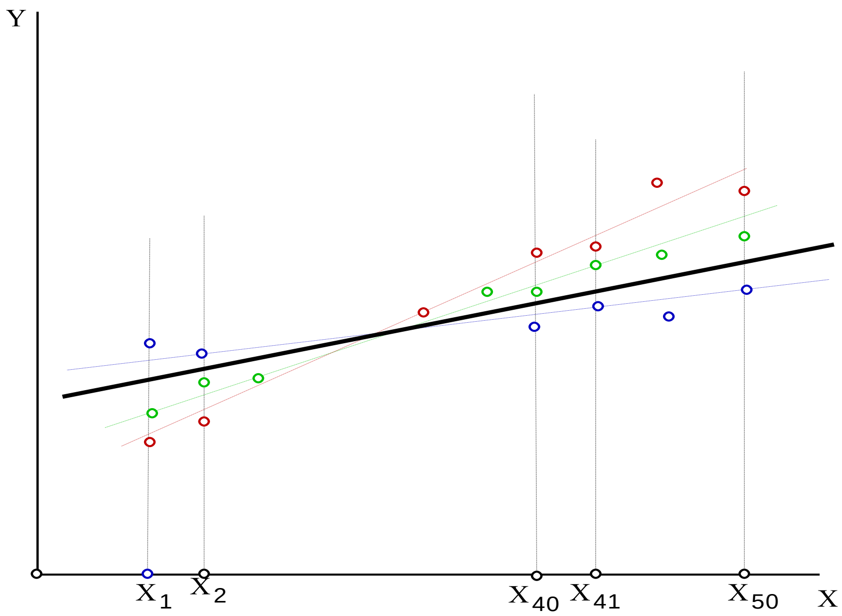


$(Y_t,\ X_t)$ $\quad t = 1,2,3,\ldots,T$	Δείγμα από *T* ζεύγη παρατηρήσεων
Y_t	Πραγματικές τιμές – Οι παρατηρήσεις του δείγματος
$E(Y_t\|X_t) = \beta_0 + \beta_1 X_t$	Γραμμή παλινδρόμησης στον πληθυσμό (άγνωστη)
$\beta_0,\ \beta_1$	Συντελεστές της παλινδρόμησης στον πληθυσμό (άγνωστοι)
$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$	Γραμμή παλινδρόμησης στο δείγμα (εκτίμηση της γραμμής παλινδρόμησης στον πληθυσμό)
$\hat{\beta}_0,\ \hat{\beta}_1$	εκτιμήσεις των $\beta_0,\ \beta_1$.
$\hat{Y}_t$	Υπολογισμένες τιμές (εκτιμήσεις των $E(Y_t\|X_t)$
$u_t = Y_t - E(Y_t\|X_t)$	Διαταρακτικοί όροι (άγνωστοι)
$\hat{u}_t = Y_t - \hat{Y}_t$	Κατάλοιπα (εκτιμήσεις των διαταρακτικών όρων)
X	Ανεξάρτητη μεταβλητή ή παλινδρομητής (regressor)

Y εξαρτημένη μεταβλητή ή παλινδρομούμενη μεταβλητή (regressand)

Συνοψίζοντας, μπορούμε να πούμε ότι, ο κύριος στόχος της ανάλυσης της παλινδρόμησης είναι να **εκτιμήσουμε** την ευθεία $E(Y_t|X_t)=\beta_0+\beta_1 X_t$ της παλινδρόμησης στον πληθυσμό, βάσει ενός μοναδικού δείγματος παρατηρήσεων, το οποίο χρησιμοποιούμε για να υπολογίσουμε την $\hat{Y}_t=\hat{\beta}_0+\hat{\beta}_1 X_t$ που είναι η γραμμή παλινδρόμησης στο δείγμα.

Αυτό σημαίνει, ότι διαφορετικά δείγματα δίνουν διαφορετικές γραμμές παλινδρόμησης στο δείγμα, και ότι εμείς από το μοναδικό δείγμα που διαθέτουμε κάθε φορά, θα υπολογίσουμε μία από αυτές.



Η ευθεία παλινδρόμησης στον πληθυσμό $E(Y_t|X_t)=\beta_0+\beta_1 X_t$

Ευθεία παλινδρόμησης που υπολογίζεται από το 1^{ο} δείγμα

Ευθεία παλινδρόμησης που υπολογίζεται από το 2^{ο} δείγμα

⋮

Ευθεία παλινδρόμησης που υπολογίζεται από το ν – στο δείγμα

- Η ευθεία παλινδρόμησης στον πληθυσμό δεν είναι τίποτα άλλο παρά η πραγματική αλλά άγνωστη σε εμάς ευθεία, πάνω στην οποία βρίσκονται οι δεσμευμένες μέσες τιμές $\mathrm{E}(Y|X_t)$ της μεταβλητής Y.
- Για κάθε δείγμα που συλλέγουμε, έχουμε και μία διαφορετική ευθεία παλινδρόμησης στο δείγμα, που είναι μία εκτίμηση της ευθείας της παλινδρόμησης στον πληθυσμό.

- Το ερώτημα που τίθεται είναι αν, δεδομένου ότι διαθέτουμε ένα μοναδικό δείγμα και ότι η ευθεία της παλινδρόμησης στο δείγμα δεν είναι τίποτα άλλο παρά μια προσέγγιση της παλινδρόμησης στον πληθυσμό, μπορούμε να έχουμε μια μέθοδο ώστε αυτή η προσέγγιση να είναι η καλύτερη.

3. Η αρχή των Ελαχίστων Τετραγώνων

Η μέθοδος των Ελαχίστων Τετραγώνων οφείλεται στον Carl Friedrich Gauss. Κάτω από τις υποθέσεις βασικές υποθέσεις της Κλασσικής Γραμμικής Παλινδρόμησης (2η Διάλεξη), η μέθοδος των Ελαχίστων Τετραγώνων έχει κάποιες πολύ ελκυστικές στατιστικές ιδιότητες που την καθιστούν μία από τις ποιό ισχυρές και δημοφιλείς μεθόδους στην ανάλυση της παλίνδρομης.

Έστω το μοντέλο δύο μεταβλητών της παλινδρόμησης στον πληθυσμό

$$Y_t = \beta_0 + \beta_1 X_t + u_t \tag{10}$$

Αυτό εκτιμάται από το μοντέλο της παλινδρόμησης στο δείγμα

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + \hat{u}_t \tag{11}$$

$$= \hat{Y}_t + \hat{u}_t \tag{12}$$

όπου $\hat{Y}_t$ είναι η εκτιμώμενη (υπό συνθήκη μέση τιμή) της Y_t.

Το πρόβλημά μας είναι η εκτίμηση της (2).

Ξαναγράφουμε την (3) ως εξής:

$$\begin{aligned} \hat{u}_t &= Y_t - \hat{Y}_t \\ &= Y_t - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_t \end{aligned} \tag{13}$$

Η σχέση αυτή μας δείχνει ότι τα κατάλοιπα $\hat{u}_t$, είναι απλώς οι διαφορές ανάμεσα στις πραγματικές και στις εκτιμώμενες τιμές της Y.

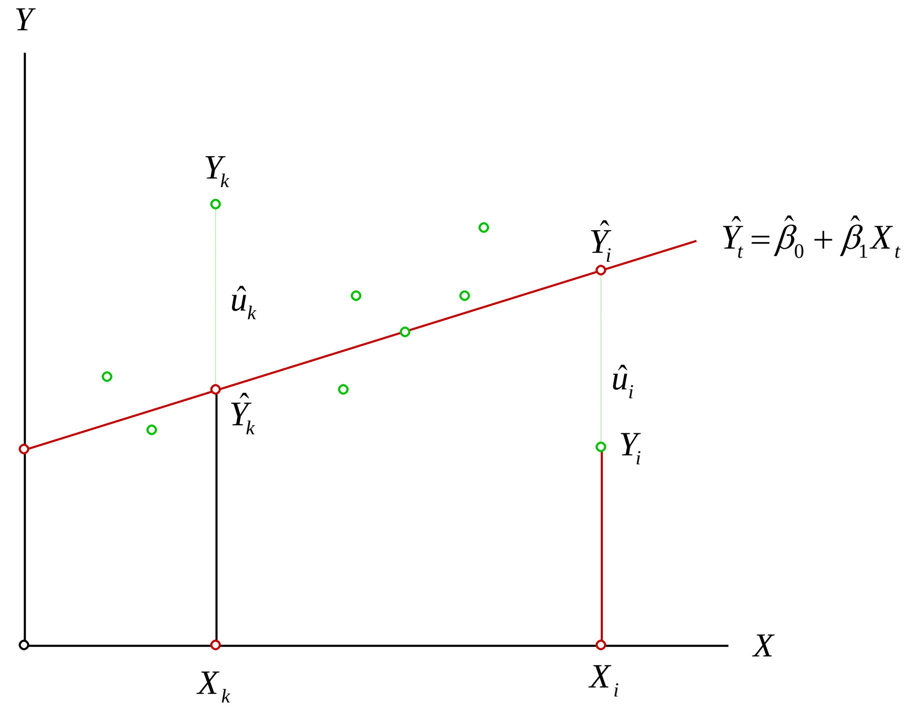
Έστω τώρα ότι διαθέτουμε ένα δείγμα $(Y_t,\ X_t)$ $\quad t = 1, 2, 3, \ldots, T$ παρατηρήσεων των Y και X, και θέλουμε να προσδιορίσουμε το μοντέλο παλινδρόμησης στο δείγμα, έτσι ώστε αυτό να βρίσκεται όσο το δυνατόν ποιό κοντά στις πραγματικές τιμές της Y. Το κριτήριο που χρησιμοποιούμε είναι το εξής:

Η παλινδρόμηση στο δείγμα επιλέγεται με τέτοιον τρόπο, ώστε τι άθροισμα των τετραγώνων των καταλοίπων να είναι όσο το δυνατόν μικρότερο.

Ζητάμε δηλαδή οι συντελεστές $\hat{\beta}_0,\ \hat{\beta}_1$, να είναι τέτοιοι ώστε το άθροισμα

$$\begin{aligned} \sum \hat{u}_t^2 &= \sum \left(Y_t - \hat{Y}_t \right)^2 \\ &= \sum \left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t \right)^2 \end{aligned} \tag{14}$$

να είναι ελάχιστο.



Η αρχή των Ελαχίστων Τετραγώνων

4. Οι Κανονικές Εξισώσεις και οι Εκτιμητές των Συντελεστών

Από την σχέση (5), φαίνεται ότι το άθροισμα των τετραγώνων των καταλοίπων είναι μία συνάρτηση δύο μεταβλητών, δηλαδή των εκτιμητών $\hat{\beta}_0, \hat{\beta}_1$. Δηλαδή

$$\sum \hat{u}_t = f\left(\hat{\beta}_0, \hat{\beta}_1\right)$$

Το πρόβλημα ελαχιστοποίησης του παραπάνω αθροίσματος είναι ένα κλασικό πρόβλημα του Διαφορικού Λογισμού συναρτήσεων πολλών μεταβλητών. Η διαδικασία ελαχιστοποίησης παράγει το παρακάτω σύστημα εξισώσεων για τους εκτιμητές $\hat{\beta}_0, \hat{\beta}_1$.

$$\begin{aligned} \sum Y_t &= T \cdot \hat{\beta}_0 + \hat{\beta}_1 \cdot \sum X_t \\ \sum Y_t \cdot X_t &= \hat{\beta}_0 \cdot \sum X_t + \hat{\beta}_1 \cdot \sum X_t^2 \end{aligned} \qquad (15)$$

όπου T είναι το μέγεθος του δείγματος, και όλα τα αθροίσματα $\sum$ υπολογίζονται από $i=1$, μέχρι $i=T$. Δηλαδή, $\sum \; = \sum_{i=1}^{T}$.

Το παραπάνω σύστημα (6), είναι γνωστό ως σύστημα των κανονικών εξισώσεων και η επίλυσή του παράγει τους εκτιμητές $\hat{\beta}_0, \hat{\beta}_1$.

Η επίλυση του παραπάνω συστήματος μας δίνει τις παρακάτω λύσεις για τους εκτιμητές $\hat{\beta}_0, \hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\sum\left(X_t - \bar{X}\right)\cdot\left(Y_t - \bar{Y}\right)}{\sum\left(X_t - \bar{X}\right)^2} = \frac{\sum x_t \cdot y_t}{\sum x_t^2} \quad (16)$$

και

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \cdot \bar{X} \quad (17)$$

όπου,

$\bar{Y} = \frac{1}{T}\sum Y_t$, ο μέσος των τιμών της μεταβλητής Y , στο δείγμα

$\bar{X} = \frac{1}{T}\sum X_t$, ο μέσος των τιμών της μεταβλητής X , στο δείγμα.

$x_t = X_t - \bar{X}$

$y_t = Y_t - \bar{Y}$

$x_t^2 = \left(X_t - \bar{X}\right)^2$

Παράδειγμα 1

Έστω ότι θέλουμε να μελετήσουμε τη συνάρτηση κατανάλωσης στα Ελληνικά νοικοκυριά και διαθέτουμε το παρακάτω δείγμα.

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

όπου (Y) οι δαπάνες κατανάλωσης και (X) το διαθέσιμο εισόδημα , των οικογενειών.

(α) Η εκτίμηση των συντελεστών της ευθείας της παλινδρόμησης στο δείγμα

Για να διευκολυνθούμε στους υπολογισμούς μας κατασκευάζουμε τον παρακάτω πίνακα.

	Y_t	X_t	$x_t = X_t - \bar{X}$	$y_t = Y_t - \bar{Y}$	$x_t^2 = \left(X_t - \bar{X}\right)^2$	$x_t \cdot y_t$
	65	100	-70	-46	4900	3220
	70	80	-90	-41	8100	3690
	90	120	-50	-21	2500	1050
	95	140	-30	-16	900	480
	110	160	-10	-1	100	10
	115	180	10	4	100	40
	120	200	30	9	900	270
	140	220	50	29	2500	1450
	150	260	90	39	8100	3510
	155	240	70	44	4900	3080
ΣΥΝΟΛΑ	**1110**	**1700**	0	0	**33000**	**16800**

$\boxed{\sum Y_t}$ $\boxed{\sum X_t}$ $\boxed{\sum x_t^2}$ $\boxed{\sum x_t \cdot y_t}$

T = **10**

$\bar{Y} = \frac{1}{T}\sum Y_t$ = **111**

$\bar{X} = \frac{1}{T}\sum X_t$ = **170**

Επομένως

$$\hat{\beta}_1 = \frac{\sum x_t \cdot y_t}{\sum x_t^2} = \frac{16800}{33000} = 0,5091$$

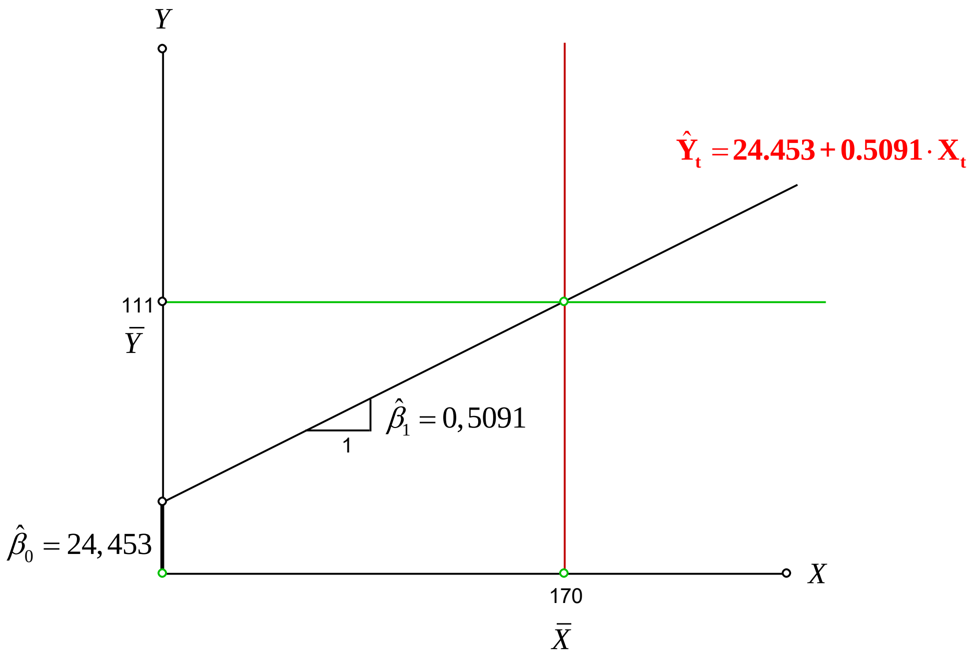
και

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \cdot \bar{X} = 111 - 0,5091 \times 170 = 24,453$$

Έτσι, η εκτιμώμενη συνάρτηση κατανάλωσης που βασίζεται στο παραπάνω δείγμα είναι η :

$$\mathbf{\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t = 24.453 + 0.5091 \cdot X_t}$$

Οι παραπάνω υπολογισμοί του πίνακα μπορούν να γίνουν εύκολα με Excel.



Η ευθεία παλινδρόμησης για τα δεδομένα του Παραδείγματος 1

(β) Η ερμηνεία των αποτελεσμάτων

Η (εβδομαδιαία) συνάρτηση κατανάλωσης είναι:

$$\hat{Y}_t = 24.453 + 0.5091 \cdot X_t$$

Συμφωνεί με την οικονομική θεωρία, καθώς $\hat{\beta}_0 > 0,\ 0 < \hat{\beta}_1 < 1$.

Κάθε σημείο της ευθείας της παλινδρόμησης, μας δίνει μια εκτίμηση της αναμενόμενης (ή μέσης) τιμής της Y , που αντιστοιχεί στην επιλεγμένη τιμή X .

Έτσι, για επίπεδο διαθεσίμου εβδομαδιαίου εισοδήματος 100€, η εκτιμώμενη μέση κατανάλωση θα είναι $\hat{Y}_t = 24.453 + 0.5091 \cdot 100 = 55.363$ €.

Ο συντελεστής παλινδρόμησης $\hat{\beta}_0 = 24.543$ (τεταγμένη της ευθείας), μας δείχνει το μέσο επίπεδο εβδομαδιαίας κατανάλωσης, όταν το διαθέσιμο εισόδημα είναι μηδενικό. Είναι δυνατόν, δηλαδή, μια οικογένεια η οποία δεν διαθέτει εισόδημα να (π.χ. εξαιτίας ανεργίας) να παρουσιάζει ένα ελάχιστο επίπεδο κατανάλωσης, δαπανώντας χρήματα που προέρχονται από δανεισμό ή αποταμιεύσεις.

Ο συντελεστής παλινδρόμησης $\hat{\beta}_1 = 0.5091$ (κλίση της ευθείας), παριστάνει την μεταβολή στην προσδοκώμενη τιμή της εξαρτημένης μεταβλητής όταν η ερμηνευτική μεταβλητή μεταβάλλεται κατά μία μονάδα. Με άλλα λόγια, είναι η **οριακή ροπή της κατανάλωσης**.

Επειδή $\hat{\beta}_1 = 0.5091$ γνωρίζουμε ότι, για επίπεδα διαθέσιμου εβδομαδιαίου εισοδήματος μεταξύ 80 και 260 ευρώ, όταν το εβδομαδιαίο διαθέσιμο εισόδημα αυξηθεί κατά 1€ , η εβδομαδιαία κατανάλωση θα αυξηθεί, κατά μέσο όρο, κατά 0,5091 λεπτά. Ή, με άλλα λόγια 50,91% της αύξησης του εβδομαδιαίου διαθέσιμου εισοδήματος θα απορροφηθεί από τις δαπάνες κατανάλωσης.

Παράδειγμα 2

Στην μικροοικονομία είναι γνωστό ότι η κατανάλωση ενός αγαθού εξαρτάται γενικά από την τιμή του, την τιμή άλλων αγαθών που είναι ανταγωνιστικά προς αυτό ή υποκατάστατά του, και από το εισόδημα του καταναλωτή. Κανονικά, για να συμπεριλάβουμε όλους αυτούς τους παράγοντες σε μιά συνάρτηση κατανάλωσης θα πρέπει να χρησιμοποιήσουμε ένα μοντέλο πολλαπλής παλινδρόμησης. Προς το παρόν, θα αρκεστούμε να υποθέσουμε ότι με τους υπόλοιπους παράγοντες να παραμένουν σταθεροί, η ποσότητα κατανάλωσης ενός προϊόντος εξαρτάται μόνον από την τιμή του. (μέθοδος ceteris paribus).

Τα παρακάτω αποτελέσματα αφορούν τη συνάρτηση κατανάλωσης του καφέ (Y) σε φλιτζάνια ανά άτομο και ανά ημέρα, και της τιμής (X) ανά κιλό. Τιμές και για τις δύο μεταβλητές σε €. Χρησιμοποιήθηκε ένα δείγμα 11 παρατηρήσεων και η μέθοδος των ελαχίστων τετραγώνων, από όπου πήραμε:

$$\hat{Y}_t = 2.6911 - 0.4795 X_t$$

Η ερμηνεία για την παραπάνω εκτιμώμενη συνάρτηση ζήτησης του καφέ έχει ως εξής:

Εφόσον $\hat{\beta}_1 = -0.4795$, η αύξηση στην τιμή του καφέ κατά 1€, αναμένεται να επιφέρει μείωση στη μέση κατανάλωση του καφέ περίπου κατά μισό φλιτζάνι.

Εφόσον $\hat{\beta}_0 = 2.6911$, γνωρίζουμε ότι αν η τιμή του καφέ μηδενιστεί, η μέση κατανάλωση του καφέ ανά άτομο αναμένεται να είναι 2.6911 φλιτζάνια την ημέρα. Παρόλο που δεν είναι πάντοτε εύκολο να δώσουμε μια φυσική ερμηνεία στη τεταγμένη της παλινδρόμησης, μπορούμε να σκεφτούμε ότι ακόμα και στην περίπτωση που η τιμή του καφέ μηδενιζόταν οι άνθρωποι δεν θα κατανάλωναν απεριόριστες ποσότητες εξαιτίας των βλαβερών επιπτώσεων της καφεΐνης στον οργανισμό.