ΣΗΜΕΙΩΣΕΙΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ

ΕΙΣΑΓΩΓΗ.

Ο ορισμός της οικονομετρίας περιλαμβάνει την εμπειρική εκτίμηση των οικονομικών σχέσεων. Χρησιμοποιώντας δηλαδή την οικονομική θεωρία την στατιστική θεωρία αλλά και τα απαραίτητα σετ δεδομένων από ένα δείγμα ενός πληθυσμού ελέγχει αλλά και μέτρα τις ορισμένες σχέσεις ανάμεσα σε διάφορες οικονομικές μεταβλητές.

Οι σκοποί της οικονομετρίας είναι βασικά τρεις: Πρώτον ο έλεγχος της οικονομικής θεωρίας ως προς την εμπειρική της τεκμηρίωση παράλληλα με την διερεύνηση των οποίων δυνατοτήτων για αναδιατύπωση κάποιων σχέσεων η οποία και καλείται διαρθρωτική ανάλυση (structural analysis), δεύτερον η διατύπωση εναλλακτικών προτάσεων οικονομικής πολιτικής (policy evaluation)και τρίτον η διενέργεια προβλέψεων όσον αφορά την εξέλιξη των διάφορων τιμών ορισμένων οικονομικών μεγεθών σε περιοχές του δείγματος εκτίμησης (forecasting).

Αναφερόμενοι στην οικονομετρική ανάλυση θα πρέπει να τονίσουμε και τα στάδια αυτής. Το πρώτο στάδιο αφορά την εξειδίκευση του υποδειγματος δηλαδή στον καθορισμό των μεταβλητών που θα το απαρτίζουν , στην καταγραφή αυτών σε εξωγενείς και ενδογενείς καθώς και στην μαθηματική διατύπωση του υποδείγματος. Το δεύτερο στάδιο αναφέρεται στην κατάλληλη επιλογή των οικονομετρικών τεχνικών για την εκτίμηση των συντελεστών των μεταβλητών μας και ονομάζεται εκτίμηση του υποδείγματος. Τέλος το τρίτο στάδιο αφορά των έλεγχο του υποδείγματος με την παράλληλη εφαρμογή οικονομικών, στατιστικών αλλά και οικονομετρικών κριτηρίων για το έλεγχο των αποτελεσμάτων της εκτιμήσεως.

Κύριος σκοπός αυτών των σημειώσεων είναι μια πρώτη γνωριμία του φοιτητή με τις προκαταρκτικές έννοιες της οικονομετρίας, το θεωρητικό υπόβαθρο αυτής καθώς και η ενασχόληση και επίλυση κάποιων προβλημάτων. Τα προβλήματα αυτά θα επιλύονται με δυο τρόπους πρώτα υπολογιστικά και δεύτερον χρησιμοποιώντας το λογισμικό στατιστικό πρόγραμμα SPSS.

Ανεξάρτητα από τους οποιονδήποτε ελέγχους αν αξιολογήσουμε ένα οικονομετρικό υπόδειγμα αυτό αξιολογείται ως προς την ικανότητα του από την επίδοση του σε προβλέψεις των ενδογενών μεταβλητών που αυτό χρησιμοποιεί. Ασφαλώς ένα υπόδειγμα το οποίο και θεωρείται «καλό» σύμφωνα με τα κριτήρια των ελέγχων δεν είναι απαραίτητο και να δίνει ακριβείς και ασφαλείς προβλέψεις.

Σχήμα 1-Διαδικασία Οικονομετρικής Αναλύσεως.

9.5 Παρουσίαση του απλού διμετάβλητου γραμμικού υποδείγματος.

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Ας υποθέσουμε ότι κάποιος επιχειρηματίας αποφασίζει να αγοράσει ένα κατάστημα εμπορίας ποτών ,εμβαδού S το οποίο βρίσκεται σε μια συνοικία ενός πληθυσμού P. Ο επιχειρηματίας αυτός γνωρίζει ότι τα κέρδη του μαγαζιού είναι γραμμική εξάρτηση της επιφάνειας αλλά και του πληθυσμού που βρίσκεται. Επίσης γνωρίζει το κέρδος 13 γειτονικών καταστημάτων όμοιας επιχειρηματικότητας. Σε τι κέρδος μπορεί να ελπίζει αν το κατάστημα που ενδιαφέρεται να αγοράσει είναι 25 τ.μ; Ο επιχειρηματίας έχει τα εξής στοιχεία στην διάθεση του.

ΠΛΗΘΥΣΜΟΣ (Δεκ.κατ.)	ΕΠΙΦΑΝΕΙΑ(τ.μ)	ΚΕΡΔΟΣ (Χιλ.ευρω)
70	21	198
35	26	209
55	14	197
25	10	156
28	12	85
43	20	187
15	15	43
33	28	211
23	19	120
4	11	62
42	10	176
20	14	117
56	36	273

Αναφερόμενοι στο παραπάνω πρόβλημα θα προσπαθήσουμε να κάνουμε μια απλή εισαγωγή στην έννοια της οικονομετρίας δίνοντας μια απλή αλλά και περιεκτική διάσταση για το τι είναι και σε τι μπορεί να μας βοηθήσει.

Η παραπάνω οικονομική σχέση μπορεί να παρασταθεί "στοχαστικά" ως εξής και να εκτιμηθεί με βάση ένα οικονομετρικό μοντέλο:

$$Y = F(X, \varepsilon) = a + \beta X_\iota + \varepsilon_\iota \qquad (1)$$

Όπου Υ είναι η εξαρτημένη **(Dependent)** μεταβλητή, ενώ X_ι είναι η ανεξάρτητη **(Independent)** μεταβλητή, ε_ι είναι η στοχαστική μεταβλητή και α,β είναι σταθεροί συντελεστές. Η σχέση απεικονίζει δυο βασικά χαρακτηριστικά :

- Την εξαρτημένη μεταβλητή Y η οποία μεταβάλλεται συστηματικά σαν αποτέλεσμα της μεταβολής της ανεξάρτητης μεταβλητής X_{ι}. Αυτή η συστηματική σχέση περιγράφεται από την ευθεία γραμμή της σχέσης (1).
- Τις διασκορπισμένες παρατηρήσεις γύρω από την ευθεία γραμμή οι οποίες οφείλονται στο γεγονός ότι:

Εκτός από την μεταβλητή X υπάρχουν και άλλον παράγοντες που επηρεάζουν την μεταβλητή Y και υπάρχει συμφυής μεταβλητικότητα της Y, καθώς και ότι τα υποδείγματα παλινδρόμησης ενσωματώνουν τα δύο χαρακτηριστικά της στοχαστικής σχέσης όταν υποθέσουμε ότι:

Για κάθε επίπεδο της ανεξάρτητης μεταβλητής X υπάρχει μία κατανομή πιθανότητας (Probability Distribution) της εξαρτημένης μεταβλητής Y, και οι μέσοι αυτών των κατανομών πιθανοτήτων μεταβάλλονται συστηματικά για κάθε μεταβολή της X.

Παρέχοντας όπως θα δούμε παρακάτω μια εκτίμηση για το μοντέλο (1) θα μπορέσουμε να απαντήσουμε στο ερώτημα μας.

9.5 **Οι υποθέσεις του απλού γραμμικού υποδείγματος.**

Σε γενική μορφή μπορούμε να ξαναδιατυπώσουμε το προς εκτίμηση υπόδειγμα ως

$Y=f(X, \varepsilon_i)= \alpha + \beta X, + \varepsilon_i \quad \iota =1,2,...,\eta$

Όπου:

Y: ονομάζεται εξαρτημένη ή ερμηνευόμενη μεταβλητή.

X: είναι η ανεξάρτητη ή ερμηνευτική μεταβλητή και

i: είναι μία αντιπροσωπευτική από τις η παρατηρήσεις του δείγματος.

Η παραπάνω σχέση ονομάζεται εξίσωση γραμμικής παλινδρόμησης και οι υποθέσεις που συνιστούν το κλασσικό γραμμικό υπόδειγμα παλινδρόμησης είναι οι εξής:

1. Η συναρτησιακή μορφή του υποδείγματος είναι γραμμική: $Y = \alpha+ \beta X +\varepsilon$, δηλαδή ότι η μαθηματική μορφή η οποία και συνδέει την ανεξάρτητη με την εξαρτημένη μεταβλητή είναι γραμμικής μορφής.

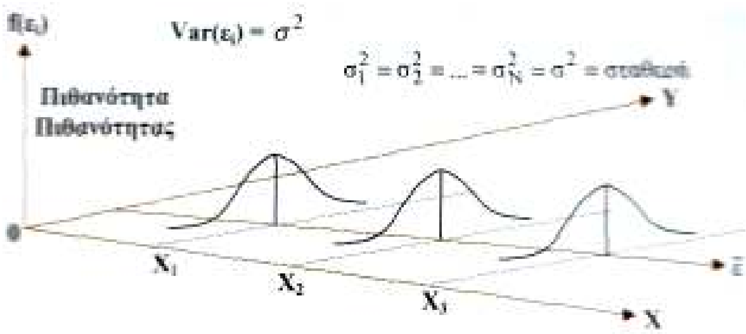
Η γραμμικότητα αυτή αναφέρεται στους συντελεστές παλινδρόμησης και όχι στις μεταβλητές του υποδείγματος.

2. Ο μέσος του όρου σφάλματος είναι μηδέν: Ε $(\varepsilon_i /X_i)=0$, δηλαδή η μεταβλητή ε_ι είναι μια τυχαία μεταβλητή η οποία και μπορεί να παίρνει τόσο αρνητικές αλλά και θετικές τιμές αλλά η μέση της τιμή (μαθηματική ελπίδα), υπό τον περιορισμό ότι η τιμή των ανεξάρτητων μεταβλητών είναι δεδομένες, είναι μηδέν.

Η σημασία της υπόθεσης αυτής συνίσταται στο γεγονός ότι οι μη εμφανείς παράγοντες οι οποίοι και ¨υπολογίζονται¨ στον διαταρακτικό όρο δεν επηρεάζουν συστηματικά την μέση τιμή της εξαρτημένης μεταβλητής.

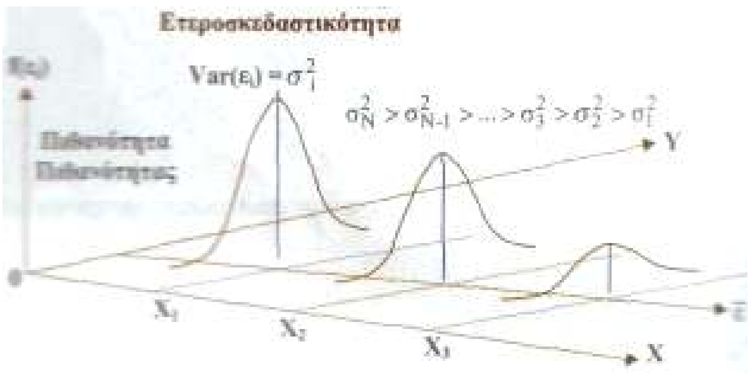
3. Η διακύμανση όλων των όρων σφάλματος είναι η ίδια σταθερά Var $(\varepsilon_i /X_i)=\sigma^2$.

Η υπόθεση αυτή μας λέει ότι η διασπορά των τιμών της τυχαίας μεταβλητής γύρω από τον μέσο της δεν αλλάζει όταν μεταβάλλεται η τιμή της ανεξάρτητης μεταβλητής X_ι. Όταν η διακύμανση παραμένει σταθερή ο διαταρακτικός όρος χαρακτηρίζεται ομοσκεδαστικός ενώ όταν η διακύμανση δεν είναι σταθερή ετεροσκεδαστικός.



Ομοσκεδαστικότητα.

(Πηγή:Οικονομετρία Α.Ανδρικοπουλου)

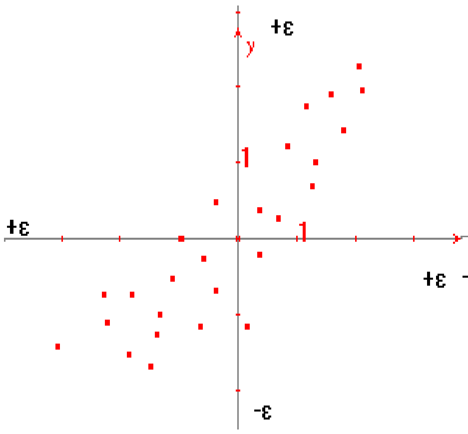


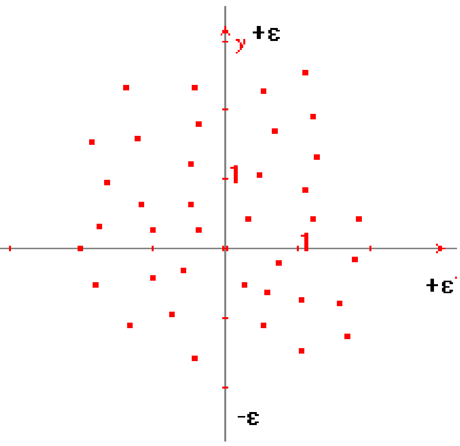
Ετεροσκεδαστικότητα (Πηγή:Οικονομετρία Α.Ανδρικοπουλου)

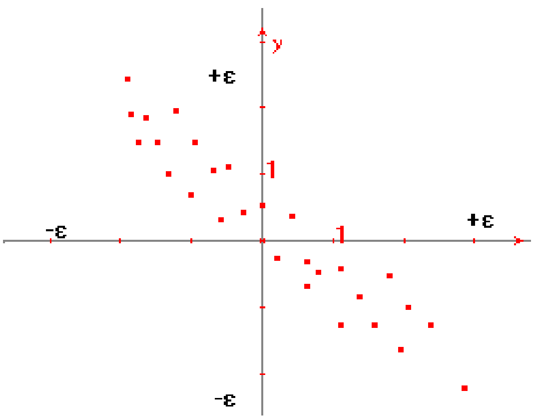
4. Η συνδιακύμανση μεταξύ των όρων σφάλματος είναι μηδέν: Cov (ε_i,ε_j)=0.

$$Cov(\varepsilon_i,\varepsilon_j) = E[\varepsilon_i - E(\varepsilon_i)][\varepsilon_j - E(\varepsilon_j)] = E(\varepsilon_i,\varepsilon_j) = 0$$

Δηλαδή η σχέση αυτή μας λέει ότι οι διαταρακτικοί όροι χαρακτηρίζονται από την απουσία της αυτοσυσχέτισης καθώς και ότι για κάθε X_ι οι αποκλίσεις των κάθε τιμών Y από τις μέσες τιμές δεν μας δίνουν υποδείγματα των κάτωθι μορφών.







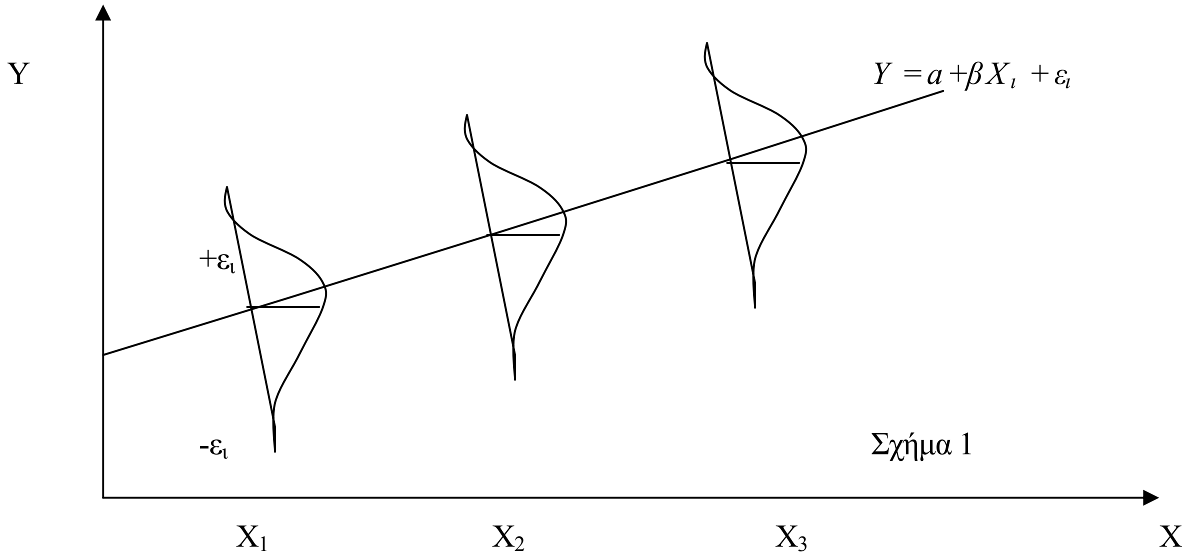
5. Η συνδιακύμανση των όρων σφάλματος και των παρατηρήσεων της ανεξάρτητης μεταβλητής είναι πάντα μηδέν: Cov (ε_i,X_ι)=0, για κάθε ί,j =1,2,.,.,η $+\varepsilon_\iota$

Η υπόθεση αυτή μας τονίζει η ανεξάρτητη μεταβλητή X δεν είναι στοχαστική και πως οι τιμές παραμένουν σταθερές σε μια επαναληπτική διαδικασία.

6. Οι όροι σφάλματος, ανεξάρτητοι μεταξύ τους, ακολουθούν την κανονική κατανομή: $\varepsilon_i \sim N(0,\sigma^2)$ με μέσο μ=0 και διασπορά σ^2.

Η τελευταία υπόθεση τίθεται για μικρά δείγματα, όπου μικρά δείγματα στην οικονομετρία θεωρούνται αυτά με αριθμό παρατηρήσεων κάτω από 30 ή κάτω από 20. Η υπόθεση αυτή δεν είναι αναγκαία για μεγάλα δείγματα, αφού βάσει του κεντρικού οριακού θεωρήματος $\varepsilon_i \to N(0,\sigma^2)$.Η υπόθεση 6 μας διευκολύνει στην στατιστική επαγωγή και την κατασκευή ελέγχων υποθέσεων σχετικά με την συμπεριφορά των εκτιμητριών.

Οι υποθέσεις αυτές μπορούν να συνοψισθούν στο ΣΧΗΜΑ [1].



Σχήμα 1

Η υπόθεση της γραμμικότητας σημαίνει ότι, όσον αφορά το προσδιορισμό, τα σημεία των ζευγών των παρατηρήσεων (X_i, Y_i) θα βρίσκονται γύρω ευθεία γραμμή, Y= α+ β X, στον διαγραμματικό χώρο (X, Y). Με άλλα λόγια οι μέσοι $E(Y_\iota / X = X_\iota)$ των κατανομών $f(Y_\iota/X)$ θα βρίσκονται πάνω στην ίδια ευθεία. Η υπόθεση αυτή δεν είναι τόσο προσδιοριστική όσο φαίνεται, γιατί αναφέρεται στον τρόπο που οι προς εκτίμηση παράμετροι εισέρχονται στην εξίσωση, και όχι αναγκαστικά στην γραμμικότητα της σχέσης μετά: και Y.

Οι υποθέσεις αυτές βασικές για την μέθοδο της ελαχιστοποίησης των τετράγωνων (OLS) περιλαμβάνονται στο γνωστό θεώρημα των Gauss-Markov.

9.5 Η ΜΕΘΟΔΟΣ ΤΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

3.1 Η Ελαχιστοποίηση των Τετραγώνων των Καταλοίπων

Σκοπός της εκτίμησης της συναρτησιακής μας σχέσης $Y_\iota = \alpha + \beta X_\iota + \varepsilon_i = E(Y_\iota / X_\iota) + \varepsilon_\iota$

είναι η εκτίμηση των αγνώστων παραμέτρων, άρα και η εκτίμηση του προσδιοριστικού μέρους της

εξίσωσης που συμβολίζεται σαν $\hat{Y}_\iota = \hat{\alpha} + \hat{\beta} X_\iota$ όπου $\hat{\alpha}, \hat{\beta}$ είναι οι εκτιμήτριες των παραμέτρων α και β

αντίστοιχα.

Το μη ερμηνευμένο μέρος $\hat{\varepsilon}_\iota = Y_i - \hat{Y}_\iota$, ονομάζεται κατάλοιπο και αποτελεί μια εκτίμηση του αγνώστου όρου σφάλματος ε_ι (βλ .ΣΧΗΜΑΑ [1]).

Για να βρούμε όμως τιμές για τις άγνωστες παραμέτρους χρειαζόμαστε κάποιο κριτήριο προσαρμογής το οποίο και να μας δίνει εκτιμήτριες οι οποίες έχουν κάποιες συγκεκριμένες ιδιότητες, όσον αφορά την σχέση τους με τις αντίστοιχες παραμέτρους. Σα κριτήριο πολύ συχνά χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων[1], η όποια συνίσταται στην ελαχιστοποίηση των τετραγώνων των καταλοίπων:

$$\min \sum_{i=1}^{n} \hat{\varepsilon}_\iota^2 = \sum_{\iota=1}^{\nu} (Y_\iota - \hat{\alpha} - \hat{\beta} X_\iota)^2$$

Οι συνθήκες πρώτης τάξης ως προς τα $\hat{\alpha}, \hat{\beta}$ δίνουν:

$$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_\iota^2}{\partial \hat{\alpha}} = \sum_{\iota=1}^{\nu} 2 (Y_\iota - \hat{\alpha} - \hat{\beta} X_\iota)(-1) \text{ δηλαδή } \sum \hat{\varepsilon}_\iota = 0$$

$$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_\iota^2}{\partial \hat{\beta}} = \sum_{\iota=1}^{\nu} 2 (Y_\iota - \hat{\alpha} - \hat{\beta} X_\iota)(-X_\iota) \text{ δηλαδή } \sum X_\iota \hat{\varepsilon}_\iota = 0$$

[1] Υπάρχουν και άλλα κριτήρια όπως η μέθοδος της μέγιστης πιθανοφάνειας.

Αυτές υπό την κάτωθι μορφή συνιστούν τις ονομαζόμενες κανονικές εξισώσεις ελαχίστων τετραγώνων:

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} X_i \qquad \sum_{i=1}^{n} X_i Y_i = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}\sum_{i=1}^{n} X_i^2$$

Η λύση αυτών μας δίνει τις εκτιμήτριες ελάχιστων τετράγωνων ως εξής:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X} \qquad \hat{\beta} = \frac{\sum_{\iota=1}^{\nu}(X_\iota - \bar{X})Y_\iota}{\sum_{\iota=1}^{\nu}(X_\iota - \bar{X})^2} = \frac{\sum_{\iota=1}^{\nu}(X_\iota - \bar{X})(Y_\iota - \bar{Y})}{\sum_{\iota=1}^{\nu}(X_\iota - \bar{X})^2} = \frac{s_{xy}}{s_x^2}$$

όπου: $s_{x,y} = \dfrac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{n-1}$ και $s_x^2 = \dfrac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$

Έχοντας εκτιμήσει τις παραμέτρους α και β έχει προσδιοριστεί και η ευθεία παλινδρομήσεως

$\hat{Y}_\iota = \hat{\alpha} + \hat{\beta} X_\iota$ που ονομάζεται ευθεία ελαχίστων τετραγώνων.

Θα πρέπει σε αυτό το σημείο να αναφέρουμε ότι οι εκτιμητές αυτοί ικανοποιούν τα εξής:

1. Είναι αμερόληπτοι εκτιμητές (Α.Ε) του πληθυσμού δηλ. $E(\hat{a}) = a$ και $E(\hat{\beta}) = \beta$.
2. Έχουν την μικρότερη δυνατή διακύμανση μεταξύ όλων των αμερόληπτων εκτιμητών.
3. Είναι συνεπείς εκτιμητές δηλ. όταν το μέγεθος του δείγματος είναι άπειρο $N \to \infty$ τότε

$\hat{\alpha} \to \alpha, \hat{\beta} \to \beta$

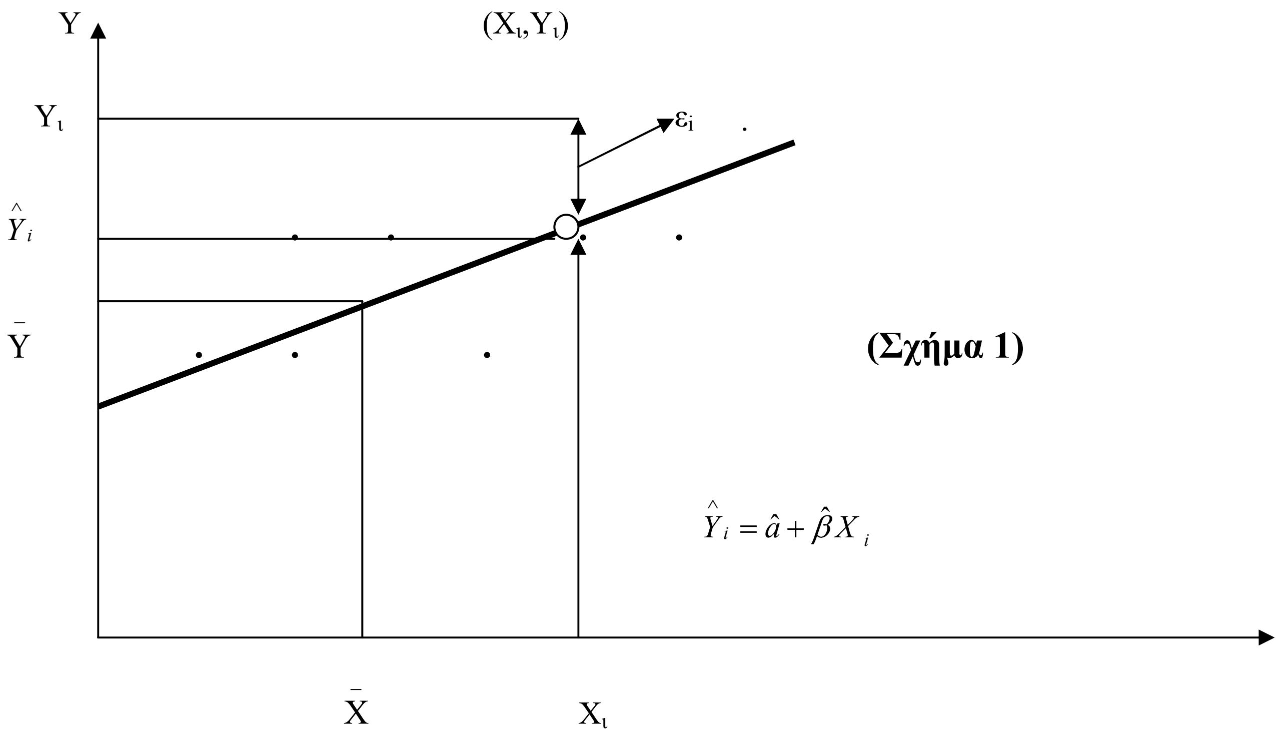
Σημείωση

Η παρατηρούμενη τιμή Y είναι η πραγματική τιμή της μεταβλητής, ενώ η $\hat{Y}$ είναι η τιμή που περιμένουμε για το δοθέν X_ι, είναι δηλαδή η εκτιμηθείς τιμή. Όσο η διαφορά της εκτιμηθείσας τιμής από την πραγματική είναι μικρή τόσο καλύτερο είναι το μοντέλο.

Ερμηνεία της παραμέτρου β

Το β είναι η κλίση της ευθείας και ονομάζεται συντελεστής παλινδρομήσεως ή γωνιακός συντελεστής. Προσδιορίζει τη αναμενόμενη μεταβολή που επέρχεται στην εξαρτημένη μεταβλητή όταν η ανεξάρτητη μεταβληθεί κατά μια μονάδα. Όταν ο συντελεστής παλινδρομήσεως είναι θετικός αριθμός, τότε η εξάρτηση είναι θετική, ενώ όταν είναι αρνητικός η εξάρτηση είναι αρνητική.

Ας δούμε και γραφικά τι παριστάνει η εκτίμηση του μοντέλου μας.



(Σχήμα 1)

4. ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΤΥΠΙΚΕΣ ΑΠΟΚΛΙΣΕΙΣ.

Τυπικό σφάλμα εκτίμησης (standard error of the estimate)

Γνωρίζουμε ότι οι εκτιμητές με την μέθοδο της ελαχιστοποίησης των τετράγωνων είναι αμερόληπτοι εκτιμητές όπως περιγράψαμε και στην προηγούμενη ενότητα.

Η διακύμανση των εκτιμητών αυτών δίνεται από τους εξής τύπους:

$$Var(\hat{\beta}) = (se(\hat{\beta}))^2 = \frac{\sigma^2}{\sum_{\iota=1}^{\nu}(X_i - \overline{X})^2}, \quad Var(\hat{a}) = (se(\hat{a}))^2 = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{\iota=1}^{\nu}(X_i - \overline{X})^2}\right) \qquad \textbf{(1)}$$

Η τετραγωνική ρίζα της διακύμανσης είναι γνωστή και ως τυπικό σφάλμα (s.e). Από τις υποθέσεις των Gauss-Markov γνωρίζουμε ότι $\varepsilon_i \sim N(0,\sigma^2)$. Αν λοιπόν στηριχτούμε στην υπόθεση αυτή τότε,

$$\hat{\beta} \sim N(\beta,(se(\hat{\beta}))^2, \hat{a} \sim N(a,(se(\hat{a}))^2 \qquad \textbf{(2)}$$

Μπορούμε να κανονικοποιήσουμε τώρα την παράσταση μας αφαιρώντας τον μέσο και διαιρώντας με το τυπικό σφάλμα. Άρα,

$$\frac{\hat{\beta}-\beta}{(se(\hat{\beta}))^2} \sim \mathrm{N}(0,1) \quad \text{και} \quad \frac{\hat{\alpha}-\alpha}{(se(\hat{\alpha}))^2} \sim \mathrm{N}(0,1) \qquad \textbf{(3)}$$

Θα θέλαμε λοιπόν να χρησιμοποιήσουμε τις σχέσεις τόσο για την δημιουργία διαστημάτων εμπιστοσύνης αλλά και για την δημιουργία ελέγχου υποθέσεων, αλλά αυτό δεν είναι εφικτό καθώς τα $(se(\hat{\beta}))$, $(se(\hat{a}))$ δεν μπορούν να υπολογιστούν.

Τα πραγματικά τυπικά σφάλματα δεν μπορούν να εκτιμηθούν καθώς εξετάζοντας την σχέση (1) η διασπορά σ^2 εκφράζει την διασπορά των υπολοίπων. Όμως εμείς δεν παρατηρούμε τα ε_ι αλλά τα $\hat{\varepsilon}_\iota$.

Από την σχέση $\mathrm{Y} = \hat{\alpha} + \hat{\beta}\mathrm{X}_\iota + \hat{\varepsilon}_i$ και λύνοντας ως προς τα $\hat{\varepsilon}_\iota$,

$\hat{\varepsilon}_i = \mathrm{Y} - \hat{\alpha} - \hat{\beta}\mathrm{X}_\iota$ όπου $\hat{\sigma}^2 = \frac{\sum \varepsilon_\iota^2}{n-2} = \frac{RSS}{n-2} \Leftrightarrow \sigma^2 = \sqrt{\frac{RSS}{n-2}}$ είναι εκτιμητής του σ^2.

Γνωρίζουμε ότι η μαθηματική ελπίδα των υπολοίπων $E(\varepsilon_\iota)=0$. Για να πάρουμε την διακύμανση θα πρέπει να διαιρέσω με το n-2. Γιατί 2 παράμετροι είναι αναγκαίοι να εκτιμηθούν για να βρω τα υπόλοιπα. Έχουμε λοιπόν ότι $\sigma^2 = \sqrt{\frac{RSS}{n-2}}$ όπου είναι εκτιμητής του σ^2 άρα $\hat{\sigma}$ είναι εκτιμητής του σ.

Χρησιμοποιωντας λοιπόν $\hat{\sigma}$ πάρουμε να εκτιμήσουμε το αληθινό τυπικό σφάλμα , έχοντας τα εκτιμώμενα τυπικά σφάλματα

$$ese(\hat{\beta}) = \hat{\sigma}\sqrt{\frac{1}{\sum_{\iota=1}^{\nu}(X_i-\overline{X})^2}} \quad ese(\hat{a}) = \hat{\sigma}\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{\sum_{\iota=1}^{\nu}(X_i-\overline{X})^2})}$$

Κανονικοποιώντας τώρα [1]θα έχουμε ότι,

$$\frac{\hat{\beta}-\beta}{ese(\hat{\beta})} \sim t(n-2) \text{ και } \frac{\hat{a}-a}{ese(\hat{a})} \sim t(n-2)$$

Σημείωση

Σε αυτό το σημείο τονίζουμε ότι τα εκτιμώμενα τυπικά σφάλματα τα οποία και ακολουθούν μια t-student κατανομή (εφόσον το δείγμα μας αποτελείται από λιτότερες από τριάντα παρατηρήσεις) με n-2 βαθμούς ελευθέριας έναντι των τυπικών σφαλμάτων τα οποία ακολουθούν μια κανονική κατανομή. Στην περίπτωση όπου το δειγμα μας αποτελείται από τριάντα και παραπάνω παρατηρήσεις θα χρησιμοποιήσουμε το κριτήριο της Ζ μεταβολής.

Στην πραγματικότητα όμως και στις δυο περιπτώσεις δεν ξεχωρίζουμε τα εκτιμώμενα τυπικά σφάλματα από τα τυπικά σφάλματα χρησιμοποιώντας το σύμβολο s.e.

Διαστήματα εμπιστοσύνης

Το Δ.Ε διάστημα εμπιστοσύνης για το συντελεστή β_i ορίζεται ως εξής υπό την προϋπόθεση ότι η μεταβλητή $\frac{\hat{\beta}_\iota-\beta_\iota}{\hat{\sigma}_\iota}$ μας ακολουθεί την t-κατανομή με N-K βαθμούς ελευθέριας:

$$P[-t_{(a/2,N-k)} \le t \le t_{(a/2,N-k)}] = 1-a \Leftrightarrow$$

[1] Η κανονικοποιημένη ποσότητα ακολουθεί t κατανομή με n-2 βαθμούς ελευθερίας.

$$P[-t_{(a/2,N-k)} \leq \frac{\hat{\beta}_\iota - \beta_\iota}{\hat{\sigma}_\iota} \leq t_{(a/2,N-k)}] = 1 - a \Leftrightarrow$$

$$P[\beta_\iota - t_{(a/2,N-k)} \hat{\sigma}_\iota \leq \hat{\beta}_\iota \leq \beta_\iota + t_{(a/2,N-k)} \hat{\sigma}_\iota] = 1 - a$$

Το 100(1-α)% ΔΕ για την εκτιμηθείς τιμή της άγνωστης παρατηρήσεως Υ για δοθέν X_ι δίνεται:

$$\hat{y} \pm t_{(n-2),a/2} \sqrt{s^2 (1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_i (x_i - \bar{x})^2})}$$

Όπου α = επίπεδο σημαντικότητας.

<u>Σημείωση</u>

Ανάλογα με τον έλεγχο αξιοπιστίας των συντελεστών του μοντέλου μας χρησιμοποιούμε την t-κατανομή ή την κανονική κατανομή αναλόγως με τον πληθυσμό του δείγματος μας σε σχέση με τον αριθμό 30.

5. ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ ΣΤΟ ΑΠΛΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ[2].

Α. ΔΙΠΛΕΥΡΟΣ ΕΛΕΓΧΟΣ

Στον δίπλευρο έλεγχο σκοπός μας είναι η διερεύνηση της παρακάτω υποθέσεως: Εάν υπάρχει στατιστικά σημαντική σχέση ανάμεσα στην εξαρτημένη και στην ανεξάρτητη μεταβλητή. Με αλλά λόγια εάν η κλίση της συνάρτησης παλινδρόμησης μας θα είναι θετική ή αρνητική με βάση την συναρτησιακή σχέση ανάμεσα στα Χ,Υ. Ο έλεγχος διαμορφώνεται ως εξής:

H_0: $\beta=0$ (Μηδενική Υπόθεση;)

H_1: $\beta \succ 0$ (Εναλλακτική Υπόθεση;)

Β. ΔΕΞΙΟΣ ΜΟΝΟΠΛΕΥΡΟΣ ΕΛΕΓΧΟΣ

Ας υποθέσουμε ότι με βάση την οικονομική θεωρία αναμένουμε ότι ο συντελεστής βι είναι θετικός και θέλουμε να διαπιστώσουμε εμπειρικά αν ο β; είναι, πρώτο, θετικός και, δεύτερο, στατιστικά σημαντικός. Σε αυτή την περίπτωση, οι H_0 και Η, υποθέσεις εκφράζονται ως εξής:

H_0: $\beta=0$ (Μηδενική Υπόθεση;)

H_1: $\beta \succ 0$ (Εναλλακτική Υπόθεση;)

Γ. ΑΡΙΣΤΕΡΟΣ ΜΟΝΟΠΛΕΥΡΟΣ ΕΛΕΓΧΟΣ

Ας υποθέσουμε ότι με βάση την οικονομική θεωρία αναμένουμε ότι ο συντελεστής βι είναι αρνητικός και θέλουμε να διαπιστώσουμε εμπειρικά αν ο β; είναι, πρώτο, είναι αρνητικός και δεύτερο, στατιστικά σημαντικός. Σε αυτή την περίπτωση, οι H_0 και Η, υποθέσεις εκφράζονται ως εξής:

H_0: $\beta=0$ (Μηδενική Υπόθεση;)

H_1: $\beta \prec 0$ (Εναλλακτική Υπόθεση;)

---

[2] Η θεωρία των ελέγχων έχει αναπτυχθεί σε προηγούμενο κεφάλαιο.

6. ΠΑΡΑΓΟΝΤΕΣ ΠΟΥ ΕΠΗΡΕΑΖΟΥΝ ΑΚΡΙΒΕΙΑ ΣΤΟ ΑΠΛΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ.

Η ακρίβεια σε ένα γραμμικό μοντέλο μετριέται από τα τυπικά σφάλματα (standard error-s.e). Για παράδειγμα όσο μικρότερο είναι το $se(\hat{\beta}_2)$ τόσο πιο ακριβής είναι και η εκτίμηση της κλίσης.

Ο τύπος για το standard error του συντελεστή κλίσης δίνεται ως εξής:

$$se(\hat{\beta}_2) = \hat{\sigma}\sqrt{\frac{1}{\sum_{\iota}(\mathrm{X}_{\iota} - \bar{\mathrm{X}})}} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{\iota}(\mathrm{X}_{\iota} - \bar{\mathrm{X}})}} \quad \text{όμως}$$

$$Var(X) = \frac{\sum_i (X_i - \bar{X})}{n-1} \Leftrightarrow \sum_i (X_i - \bar{X}) = (n-1)Var(X) \text{ και ισχύει ότι } \hat{\sigma}^2 \approx \mathrm{var}(\varepsilon)$$

Άρα ο τύπος γίνεται: $se(\hat{\beta}_2) = \sqrt{\frac{\mathrm{var}(\varepsilon)}{(n-1)Var(X)}}$

Σε αυτό το σημείο πρέπει να αναφέρουμε ότι η ακρίβεια εξαρτάται από τρεις κύριους παράγοντες.

a. Από το μέγεθος του δείγματος. Αφού το n εμφανίζεται και στον παρανομαστή του κλάσματος όσο μεγαλύτερο είναι το μέγεθος τόσο και μικρότερη η τιμή του standard error άρα τόσο πιο ακριβής η εκτίμηση αυτού.

b. Επίσης όσο μεγαλύτερη είναι η απόκλιση του Χ τόσο και μικρότερη η τιμή του standard error άρα τόσο πιο ακριβής η εκτίμηση αυτού.

c. Var (ε). Όσο μεγαλύτερη είναι η απόκλιση των ε_i στην κατακόρυφη διεύθυνση τόσο το λιγότερο ακριβής είναι η εκτίμηση του standard error of the slope parameter.

7. ΔΙΑΦΟΡΕΣ ΣΥΝΑΡΤΗΣΙΑΚΕΣ ΜΟΡΦΕΣ.

Παλινδρόμηση καλείται η μέθοδος εκτίμησης μιας μεταβλητής, της εξαρτημένης από άλλες ανεξάρτητες μεταβλητές. Έτσι αν η Y μεταβλητή πρόκειται να εκτιμηθεί από τις X_i, i = 1,…,p με βάση μια εξίσωση, η εξίσωση αυτή καλείται εξίσωση παλινδρόμησης.

Είδη εξισώσεων παλινδρόμησης

Απλό γραμμικό μοντέλο

$Y = \alpha + \beta x$

Γενικό γραμμικό μοντέλο

$Y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$

Πολυωνυμικό μοντέλο

$Y = \beta_o + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p$

Εκθετικό μοντέλο

$Y = \exp(\beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)$

Το εκθετικό μοντέλο αυτό με έναν λογαριθμικό μετασχηματισμό μετατρέπεται σε γραμμικό. Αναφερόμενοι στο εκθετικό μοντέλο θα παρουσιάσουμε το εξής παράδειγμα. Ας θεωρήσουμε την εξής συνάρτηση παραγωγής Cobb-Douglas όπου Q είναι το παραγόμενο προϊών , L είναι το εργατικό δυναμικό, K είναι το κεφαλαίο και A, b, a σταθερές.

$Q = AK^a L^b \quad (1)$

Η εξίσωση αυτή με την υπάρχουσα μορφή θα λέγαμε ότι είναι αρκετά δύσκολο να εκτιμηθεί όποτε προσπαθούμε να την μετασχηματίσουμε σε κάποια άλλη μορφή. Η ιδία της όμως η μορφή (εκθετική) μας επιτρέπει να την λογαριθμοποιήσουμε οπότε:

$Q = AK^a L^b \Leftrightarrow \ln Q = \ln A + a \ln K + b \ln L \quad (2)$

Παρατηρούμε ότι η εξίσωση (2) είναι σε γραμμική μορφή, άρα ποιο εύκολο να εκτιμηθεί.

Κατάλληλοι μετασχηματισμοί μπορούν να εφαρμοστούν και σε άλλα μοντέλα μετατρέποντας τα σε γραμμικά.

8. ΑΝΑΛΥΣΗ ΔΙΑΣΠΟΡΑΣ.

Είναι γνωστό ότι ,

$$\varepsilon_t = \mathrm{Y}_t - \hat{\mathrm{Y}}_t = (\mathrm{Y}_t - \overline{\mathrm{Y}}) - (\hat{\mathrm{Y}}_t - \overline{\mathrm{Y}}) \Leftrightarrow$$

$$(\mathrm{Y}_t - \overline{\mathrm{Y}}) = (\mathrm{Y}_t - \hat{\mathrm{Y}}_t) + (\hat{\mathrm{Y}}_t - \overline{\mathrm{Y}})$$

$$(\mathrm{Y}_t - \overline{\mathrm{Y}})^2 = (\mathrm{Y}_t - \hat{\mathrm{Y}}_t)^2 + (\hat{\mathrm{Y}}_t - \overline{\mathrm{Y}})^2 + 2(\mathrm{Y}_t - \hat{\mathrm{Y}}_t)(\hat{\mathrm{Y}}_t - \overline{\mathrm{Y}}) \Leftrightarrow$$

διότι $2(\mathrm{Y}_t - \hat{\mathrm{Y}}_t)(\hat{\mathrm{Y}}_t - \overline{\mathrm{Y}}) = 0$ άρα

$$\sum (\mathrm{Y}_t - \overline{\mathrm{Y}})^2 = \sum (\mathrm{Y}_t - \hat{\mathrm{Y}}_t)^2 + \sum (\hat{\mathrm{Y}}_t - \overline{\mathrm{Y}})^2$$

<u>Αθροίσματα τετραγώνων-(Total Sum of Squares)</u>

Το άθροισμα $SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \text{ΑΤΠ}$ ονομάζεται συνολικό άθροισμα τετραγώνων και εκφράζει την ολική μεταβολή (total variation).Είναι ένα μέτρο της συνολικής μεταβολής για την εξαρτημένη μεταβλητή. Το άθροισμα αυτό αναλύεται σε δυο άλλες συνιστώσες, εκ των οποίων η μια οφείλεται στην παλινδρόμηση και η άλλη εκφράζει το υπόλοιπο μεταβολής (residual variation) και είναι η μεταβολή που οφείλεται σε άλλους παράγοντες εκτός από το x. *Παρατηρήστε ότι εάν διαιρέσουμε τον SST με το n-1 έχουμε την Var(Y).*

<u>Άθροισμα Τετραγώνων Παλινδρόμησης-(Explained Sum Of Squares).</u>

Το άθροισμα που οφείλεται στην παλινδρόμηση (SS due to regression) και εξηγείται από αυτήν δίνεται από τον εξής τύπο:

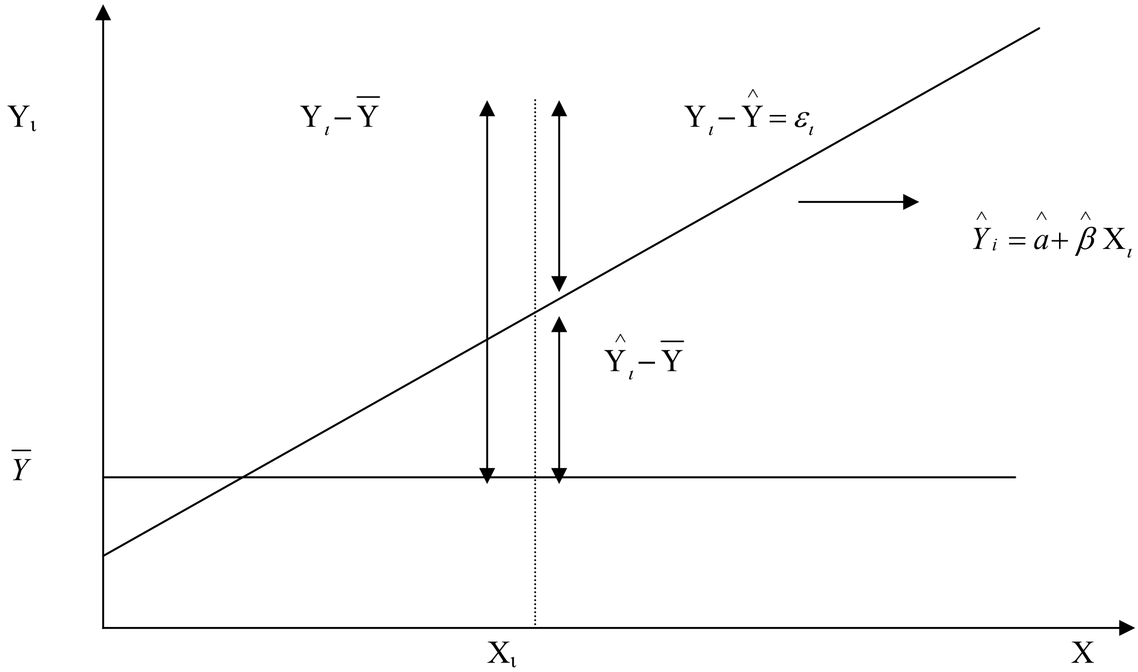
$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \text{ΣΤΑ}$$

<u>Άθροισμα Τετραγώνων των Καταλοίπων-(Residual Sum of Squares).</u>

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{\varepsilon}^2_t = \text{ΑΤΚ}$$

Επομένως ισχύει: $TSS = ESS + RSS$

Στο παρακάτω σχήμα θα προσπαθήσουμε να αναλύσουμε την μεταβλητικότητα της Y και τα συστατικά αυτής,



Η μεταβλητικοτητα της εξαρτημένης μεταβλητής παρουσιάζετε στον κάτωθι πίνακα.

Πηγή Μεταβλητικοτητας	**Άθροισμα Τετράγωνων**	**Βαθμοί Ελευθέριας**	
Παλινδρόμηση	ΑΤΠ	Κ-1	ΑΤΠ/κ-1
X			
Κατάλοιπα	ΑΤΚ	Ν-κ	ΑΤΚ/Ν-κ
ε	ΣΤΑ	Ν-1	ΣΤΑ/Ν-1

R^2 συντελεστής προσδιορισμού.

Η αναλογία της μεταβλητικότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από την παλινδρόμηση ονομάζεται συντελεστής προσδιορισμού (coefficient of determination) και δίνεται από τον τύπο: $R^2 = \frac{ESS}{TSS}$, όπου $0 \le R^2 \le 1$

Ο συντελεστής προσδιορισμού R^2 μετράει το ποσοστό της μεταβλητικότητας της μεταβλητής Y η οποία και ερμηνεύεται από την παλινδρόμηση του δείγματος.

Ο συντελεστής προσδιορισμού R^2 συμπίπτει με το r^2, όπου r ο δειγματικός συντελεστής συσχέτισης στο απλό γραμμικό μοντέλο. Επίσης όταν το R^2 είναι κοντά στην μονάδα τότε μιλάμε ότι το μοντέλο μας είναι καλό στην ερμηνεία της απόκλισης του Y ,υπάρχει τέλεια προσαρμογή της ευθείας παλινδρομήσεως (Fit Model).

Πρόβλημα χρησιμοποιώντας το R^2.

Πολλοί ερευνητές χρησιμοποιούν το συντελεστή προσαρμογής ως ένα ιδανικό κριτήριο για την επιλογή ενός οικονομετρικού μοντέλου. Είναι όμως αυτό σωστό; Το πρόβλημα με τον συντελεστή προσδιορισμού είναι ότι αυξάνει με την προσθήκη μιας προσθετής ερμηνευτικής μεταβλητής στο μοντέλο μας. Για παράδειγμα εάν στο παράδειγμα της κατανάλωσης φαγητού με το οικογενειακό εισόδημα εμείς προσθέσουμε την μεταβλητή της ύπαρξης τηλεφωνικής γραμμής θα παρατηρήσουμε τι η τιμή του R^2 θα αυξηθεί. Άρα καταλαβαίνουμε ότι η χρήση του R^2 για την επιλογή ενός οικονομετρικού μοντέλου θα είναι ένα κίνητρο για την προσθήκη και άλλων επεξηγηματικών μεταβλητών. Η λύση σε αυτό το πρόβλημα είναι η χρήση του adjusted R^2.
Πρώτα ας θεωρήσουμε ότι:

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

Adjusted R^2 είναι μια κανονικοποιημένη μορφή της τελικής έκφρασης:

$$\overline{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$$, όπου κ ο αριθμός των παραμέτρων του μοντέλου μας[2].

F – έλεγχος

Ο έλεγχος αυτός εξετάζει τη γραμμική παλινδρόμηση μεταξύ των X και Y. Δηλαδή, εδώ έχουμε τον εξής έλεγχο υποθέσεων:

H_o: όλοι οι συντελεστές παλινδρόμησης εκτός από το α είναι μηδέν.

<u>έναντι</u>

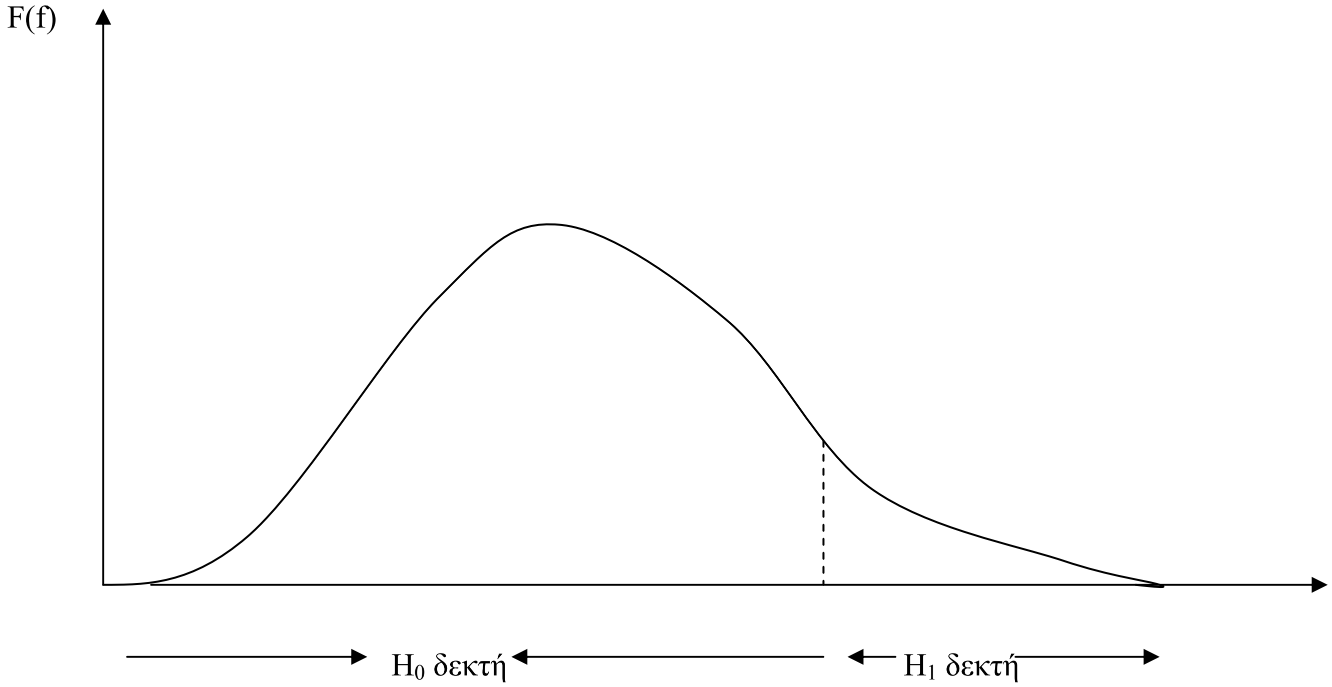
H_1: όλοι οι συντελεστές είναι διαφορετικοί του μηδενός.

<u>*Μεθοδολογία:*</u>

Για τον έλεγχο της H_o υπολογίζεται ο παρακάτω λόγος: $F = \frac{MSR}{MSE} = \frac{SSR/_{n-2}}{SSE/_{1}} = \frac{R^2/k-1}{(1-R^2)/N-k}$

Το F ακολουθεί την F-κατανομή με (1 , n-2) β.ε.

Ο έλεγχος αυτός απεικονίζετε ως εξής:



[2] Εδώ κ=2 οι παράμετροι μας ο σταθερός όρος και η κλίση.

Συντελεστής Συσχέτισης.

Ο συντελεστής συσχέτισης (Correlation Coefficient) είναι ένας δείκτης που μετράει τον βαθμό της γραμμής συσχέτισης ανάμεσα στις μεταβλητές X και Y και ορίζεται σαν η τετραγωνική ρίζα του συντελεστή προσδιορισμού, δηλαδή, $r = \pm\sqrt{R^2} = \dfrac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2}\sqrt{\sum (Y-\bar{Y})^2}}$. Αν και υπάρχει τόσο στενή σχέση ανάμεσα στον συντελεστή προσδιορισμού, r^2 , και τον συντελεστή συσχέτισης, R, υπάρχει όμως σημαντική διαφορά ως προς την ερμηνεία τους και, φυσικά, τις ιδιότητες του. Συγκεκριμένα, ο συντελεστής προσδιορισμού είναι πάντοτε θετικός και η τιμή του βρίσκεται ανάμεσα στη μονάδα και το μηδέν. Αντίθετα ο συντελεστής συσχέτισης κυμαίνεται μεταξύ της αρνητικής και θετικής μονάδας. Δηλαδή :

- Όταν r=1 τότε υπάρχει πλήρης θετική συσχέτιση ανάμεσα στην Y και X, όταν r=-1 υπάρχει πλήρης αρνητική συσχέτιση , και όταν r=0,οι μεταβλητές Y και X δεν συσχετίζονται.

- Ο συντελεστής συσχέτισης είναι συμμετρικός , $r_{xy} = r_{yx}$, δεν εξαρτάται από τον γραμμικό μετασχηματισμό των μεταβλητών του υποδείγματος, και έχει μόνο εφαρμογή όταν η σχέση ανάμεσα στην Y και X είναι γραμμική.

- Ο συντελεστής συσχέτισης του δείγματος είναι ένας εκτιμητή: του συντελεστή συσχέτισης στον πληθυσμό,

ΑΣΚΗΣΗ-ΕΦΑΡΜΟΓΗ

Ο ρόλος όπως είδαμε της οικονομετρίας είναι να προσδιορίσει ποσοτικά, οικονομικά αλλά και άλλα μοντέλα. Αρχικά θα προσπαθήσουμε με ένα απλό παράδειγμα να κάνουμε μια εισαγωγή στην εκτίμηση ενός οικονομετρικού υποδείγματος.

Παράδειγμα.

Σύμφωνα με τον νόμο του Άγγλου οικονομολόγου Engel (1848) όσο το εισόδημα ενός νοικοκυριού αυξάνει τόσο η κατανάλωση σε φαγητό αυξάνει κατ' αναλογία ή η εισοδηματική ελαστικότητα για την ζήτηση για φαγητό είναι μεγαλύτερη του μηδενός αλλά μικρότερη της μονάδας.

Το οικονομικό μοντέλο που περιγράφει αυτήν την σχέση δίνεται ως εξής:

Κατανάλωση σε Φαγητό $=\beta_1+\beta_2$ **εισόδημα* , όπου $\beta_1>0$ και $0<\beta_2<1$.

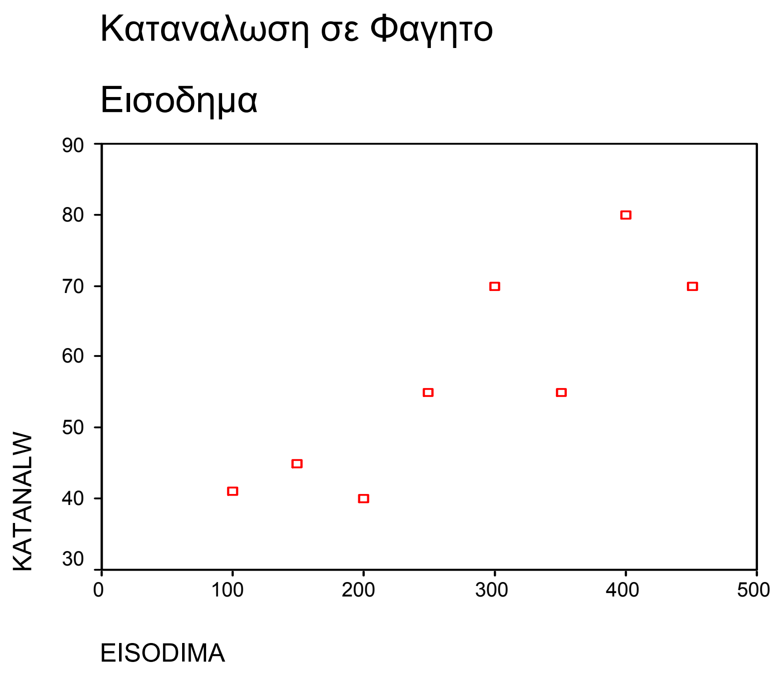
Οι οικονομέτρες θα προσπαθήσουν να εκτιμήσουν τα β_1,β_2. Επίσης θα προσπαθήσουν να προσδιορίσουν το κατά πόσον εμπιστοσύνη μπορούμε να έχουμε σε αυτές τις εκτιμήσεις τους (χρησιμοποιώντας τυπικές αποκλίσεις). Έπειτα θα χρησιμοποιήσουν τεστ για την ισχύ του νόμου του Engel. Για να γίνουν όλα αυτά θα χρησιμοποιήσουν το κάτωθι οικονομετρικό μοντέλο.

Κατανάλωση σε Φαγητό $=\beta_1 + \beta_2$ **εισόδημα* $+\varepsilon_{i.}$

Το Απλό Γραμμικό Μοντέλο.

Πολλές φορές μας ενδιαφέρει η στατιστική ανεξαρτησία κάποιων μεταβλητών είτε αυτές είναι ανεξάρτητες ή και εξαρτημένες. Στην περίπτωση όπου υπάρχει μόλις μια επεξηγηματική μεταβλητή το διμετάβλητο οικονομετρικό μοντέλο (Gujarati) είναι το κατάλληλο για την εκτίμηση. Ας θεωρήσουμε το κάτωθι σετ των δεδομένων για το άνωθι παράδειγμα.

Εισόδημα (X):	100	150	200	250	300	350	400	450
Κατανάλωση σε Φαγητό (Y):	41	45	40	55	70	55	80	70.



Παραπάνω δίνεται ένα διάγραμμα διασποράς των δεδομένων. Η επεξηγηματική μεταβλητή X απεικονίζεται στον οριζόντιο άξονα χχ' ενώ η εξαρτημένη μεταβλητή στον κάθετο. Κάθε νοικοκυριό απεικονίζεται από ένα σημείο στο διάγραμμα μας. Βέβαια είναι φανερή η θετική και θα λέγαμε γραμμική σχέση μεταξύ των δυο αυτών μεταβλητών το ερώτημα που υφίσταται εδώ είναι με ποιον τρόπο θα καταφέρουμε να εκτιμήσουμε αυτήν την θετική σχέση μεταξύ των δυο αυτών μεταβλητών.

Το οικονομετρικό μοντέλο που θα χρησιμοποιήσουμε είναι το εξής:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X + \hat{\varepsilon}_\iota \qquad \iota = 1,2,\ldots,\eta \text{ όπου} \qquad (1)$$

Y_ι: το εισόδημα του νοικοκυριού i.

β_1: σταθερός όρος.

β_2: κλίση.

n: το μέγεθος του δείγματος.

X_ι: η κατανάλωση φαγητού του νοικοκυριού i.

ε_i: ο διαταρακτικός όρος του νοικοκυριού i.

Κύριο μέλημα μας είναι η εκτίμηση των β_1, β_2 που ονομάζονται και μεταβλητές του υποδείγματος μας. Η μέθοδος η οποία και θα χρησιμοποιηθεί είναι η μέθοδος (OLS, ordinaryY least square) ελαχίστων τετραγώνων η οποία και θα αναλυθεί εκτενέστερα αργότερα.

Σύμφωνα με αυτήν οι εκτιμήσεις των παραμέτρων β_1, β_2 συμβολίζονται ως $\hat{\beta}_1, \hat{\beta}_2$ και δίνονται από τους εξής τύπους:

$$\hat{\beta}_2 = \frac{\sum_{\iota=1}^{\nu} (X_\iota - \bar{X}) Y_\iota}{\sum_{\iota=1}^{\nu} (X_\iota - \bar{X})^2} \tag{2}$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \bar{X} \tag{3}$$

Για τον υπολογισμό αυτών σχεδιάζουμε το κάτωθι πίνακα.

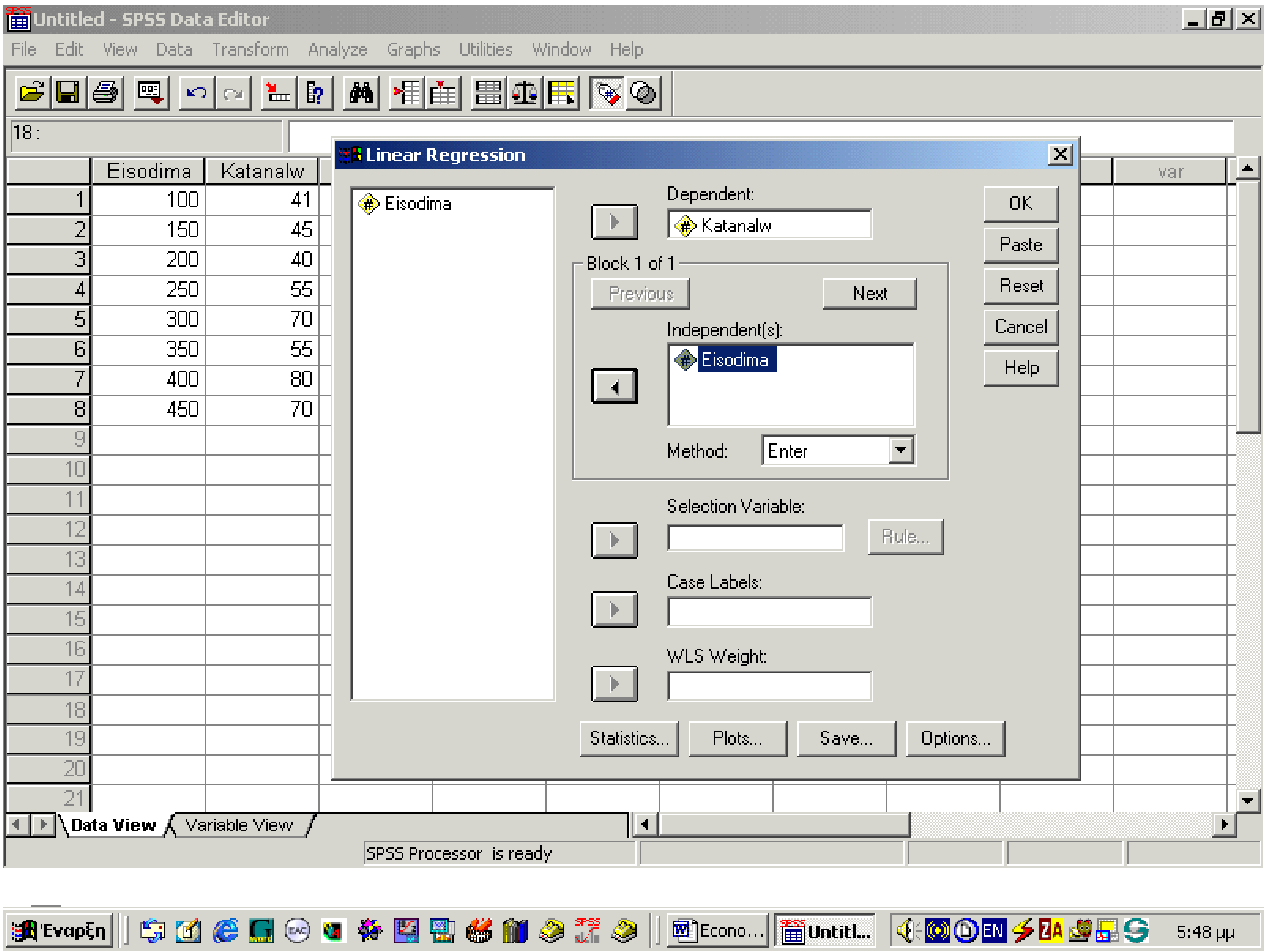
X	Y	$X - \bar{X}$	$(X - \bar{X})^2$	$(X - \bar{X})Y$
100	41	-175	30625	-7175
150	45	-125	15625	-5625
200	40	-75	5625	-3000
250	55	-25	625	-1375
300	70	25	625	-1750
350	55	75	5625	4125
400	80	125	15625	10000
450	70	175	30625	12250
$\sum_{\iota=1}^{8} X = 2200$	$\sum_{1}^{8} Y = 456$		$\sum_{\iota=1}^{8} (X - \bar{X})^2$	$\sum_{\iota=1}^{8} (X - \bar{X})Y$
$\bar{X} = 275$	$\bar{Y} = 275$			

Χρησιμοποιώντας λοιπόν έναν τέτοιο πίνακα μπορούμε να υπολογίσουμε με ένα απλό κομπιουτεράκι τα $\hat{\beta}_1, \hat{\beta}_2$. Πως όμως μπορούμε να χρησιμοποιήσουμε το SPSS για τον υπολογισμό των $\hat{\beta}_1, \hat{\beta}_2$;

Πριν προχωρήσουμε στον υπολογισμό των συντελεστών ας δούμε πως μπορούμε να απεικονίσουμε το διάγραμμα διασποράς με την βοήθεια του SPSS.

- Επιλέξτε Graphs από τη γραμμή μενού στην κορυφή του παραθύρου, για να εμφανίσετε ένα πτυσσόμενο μενού (Εικόνα). Έπειτα επιλέξτε scatter και εν συνεχεία simple και, Define, για να ορίσετε το διάγραμμα διασποράς το οποίο θέλετε.
- Στο menu που εμφανίζεται και στον άξονα των ΄ψ εισάγετε την μεταβλητή katanalwsh ενώ στον ΄χχ εισάγετε την μεταβλητή eisodhma.
- Επιλέγοντας το κουμπί titles μπορείτε να δώσετε τίτλο στο διάγραμμα διασποράς, ενώ με το κουμπί OK έχουμε την γραφική απεικόνιση του.
- Ας περάσουμε τώρα στον υπολογισμό των συντελεστών.

Εικονα 1



- Στο παράθυρο του Data Editor καταχωρίστε τα δεδομένα συγκεκριμένα τα X,Y ή φορτώστε το ανάλογο αρχείο εάν τα δεδομένα σας είναι σε μορφή π.χ excel.
- Επιλέξτε Analyze από τη γραμμή μενού στην κορυφή του παραθύρου, για να εμφανίσετε ένα πτυσσόμενο μενού (Εικόνα 1).
- Από αυτό το πτυσσόμενο μενού, επιλέξτε Regression (Παλινδρόμηση) για να ανοίξετε ένα μικρότερο μενού.
- Από το νέο μενού επιλέξτε Linear (Γραμμική) για να ανοίξετε το πλαίσιο διαλόγου που αναφέρετε ως Linear Regression.
- Επιλέξτε τη μεταβλητή "katanalw''και πατήστε στο κουμπί δίπλα στο πλαίσιο κειμένου Dependent : (Εξαρτημένη) για να την προσθέσετε εκεί.
- Επιλέξτε τη μεταβλητή 'Eisodima" και πατήστε στο κουμπί δίπλα στο πλαίσιο κειμένου Independent (ανεξάρτητη) για να την προσθέσετε εκεί.
- Πατήστε στο κουμπί Statistics και προσέξτε να είναι επιλεγμένη η επιλογή των estimates, εάν όχι την επιλέγετε εσείς. Με αυτόν τον τρόπο υπολογίζεται τους εκτιμητές των β_1, β_2.Επιλεγοντας Continue επιστρέφουμε στο αρχικό menu όπου πατώντας το κουμπί O έχουμε καταφέρει να υπολογίσουμε τις εκτιμήσεις παίρνοντας τον ακόλουθο πίνακα.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	28,321	7,711		3,673	,010	9,453	47,189
	EISODIMA	,104	,026	,854	4,029	,007	,041	,168

a. Dependent Variable: KATANALW

(Πίνακας 1)

Από τον πίνακα 1 έχουμε ότι $\hat{\beta}_1 = 28.321, \hat{\beta}_2 = 0.104$

Άρα το οικονομετρικό μας μοντέλο έχει την εξής μορφή:

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X + \hat{\varepsilon}_i \Leftrightarrow \hat{Y} = 28{,}321 + 0{,}104X$$

Σύμφωνα με αυτό το μοντέλο και παρατηρώντας το Σχήμα 1 μπορούμε να αναρωτηθούμε εάν το σημείο στην γραμμή παλινδρόμησης ανταποκρίνεται στην εκάστοτε παρατήρηση ι. Βέβαια δεν μπορούμε να το ισχυριστούμε αυτό αλλά μπορούμε να πούμε ότι η γραμμή παλινδρόμησης εκφράζει την πρόβλεψη για την κατανάλωση φαγητού δεδομένου του εισοδήματος που μια οικογένεια μπορεί να έχει. Η πραγματική κατανάλωση φαγητού για κάθε οικογένεια εκφράζεται ως Y_i, ενώ η διαφορά (απόσταση) μεταξύ της γραμμής παλινδρόμησης (fitted value) και της εκάστοτε παρατήρησης εκφράζει το υπόλοιπο (κατάλοιπο) ε_i. Εάν το υπόλοιπο είναι θετικό μπορούμε να ισχυριστούμε ότι η πραγματική κατανάλωση φαγητού για κάθε νοικοκυριό είναι μεγαλύτερη από ότι το μοντέλο προβλέπει και αντιστρόφως.

Θα πρέπει όμως σε αυτό το σημείο να αναφέρουμε το κατά πόσο 'ταιριαστό' είναι το μοντέλο το οποίο και έχουμε επιλέξει για την εκτίμηση των μεταβλητών μας . Αναφερόμενοι στην θεωρία μας και πάλι θα πρέπει να τονίσουμε την ύπαρξη των συντελεστών προσδιορισμού και το κατά πόσον είναι πιο κοντά αριθμητικά στην μονάδα.

Model Summary (b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,854 (a)	,730	,685	8,387	,730	16,233	1	6	,007	

a Predictors: (Constant), EISODIMA
b Dependent Variable: KATANALW

Με βάση το SPSS θα έχουμε ότι R=0.854, R^2=0.730 και Adjusted R^2=0.685. Σχολιάζοντας λοιπόν τον συντελεστή προσδιορισμού να αναφέρουμε ότι το 85.4% της συνολικής μεταβλητικότητας ερμηνεύεται από το εισόδημα των καταναλωτών ενώ το υπόλοιπο 14.6 % από υπόλοιπους παράγοντες.

Εκτός όμως από αυτό μας ενδιαφέρει και αν ισχύουν και κάποιες υποθέσεις που αφορούν το μοντέλο μας.

Για παράδειγμα ερωτάμε εάν το οικογενειακό εισόδημα επηρεάζει την δαπάνη σε φαγητό. Για να εξετάσουμε αυτήν την ερώτηση θα χρησιμοποιήσουμε την κατασκευή ενός ελέγχου υπόθεσης ως εξής:

<u>Έλεγχος υπόθεσης 2 ουρών-Δίπλευρος (Two tailed test)</u>

H_0: $\beta_2=0$ (Μηδενική Υπόθεση; το εισόδημα έχει καμία επίδραση στην δαπάνη για φαγητό.)

H_1: $\beta_2 \neq 0$ (Εναλλακτική Υπόθεση; Το εισόδημα δεν έχει καμία επίδραση)

Είναι ένα t-test από το οποίο και βρίσκουμε ότι :

$t = \dfrac{\beta - 0}{se(\hat{\beta})} = \dfrac{0.104}{0.026} = 4.029$ Τα αποτελέσματα περιέχονται στον πίνακα 1.

Γνωρίζοντας πως ελέγχουμε μια τέτοια υπόθεση με ένα επίπεδο σημαντικότητας α=0,05 θα έχουμε ότι θα απορρίπτουμε την αρχική μας υπόθεση H_0 εάν $t>t_{6,0.025}$ ή $t<-t_{6,0.025}$ και αποδεχόμαστε την H_0 εάν $t<t_{6,0.025}$. Συγκεκριμένα έχουμε ότι 4.029<2.45 το οποίο και ισχύει άρα απορρίπτουμε την H_0 συνεπώς το εισόδημα δεν έχει καμία επίδραση στην δαπάνη φαγητού.

Πόσο σημαντικό είναι το αποτέλεσμα που έχουμε βρει; Αυτό εξαρτάται από την τιμή p-value η οποία είναι η πιθανότητα του να έχουμε μια τιμή τόσο ακραία όσο αυτή που αποκτούμαι.

P-value=P(t(6)>4,029)=0.05

- Εάν p-value<0,10 υπάρχει ήπια ένδειξη έναντι της H_0.
- Εάν p-value<0,05 υπάρχει ένδειξη έναντι της H_0.
- Εάν p-value<0,010 υπάρχει ισχυρή ένδειξη έναντι της H_0.

Στην περίπτωση αυτή έχουμε ότι υπάρχει ισχυρή ένδειξη αφού p-value<0,010.

Στην περίπτωση που εμείς τώρα θα θέλουμε να εξετάσουμε εάν το φαγητό είναι κανονικό αγαθό δηλαδή εάν έχει θετική επίδραση στο εισόδημα θα πρέπει να σχηματίσουμε τον εξής έλεγχο όσον αφορά το β (Μονόπλευρος)

H_0: β_2=0 (Μηδενική Υπόθεση; το εισόδημα δεν έχει καμία επίδραση στην δαπάνη φαγητού

H_1: $\beta_2>0$ (Εναλλακτική Υπόθεση; Το εισόδημα έχει κάποια επίδραση-Κανονικό αγαθό)

Παρατηρούμε λοιπόν ότι εδώ έχουμε μονόπλευρο έλεγχο αρά οι κανόνες για την αποδοχή ή απόρριψη της H_0 αλλάζουν. Συγκεκριμένα απορρίπτουμε την H_0 εάν $t<t_{6,0.05}$ και δεχόμαστε την αρχική υπόθεση εάν $t>t_{6,0.05}$.Εχουμε λοιπόν ότι 4.029>1.94 άρα απορρίπτουμε την H_0.

Ομοίως μπορούμε να απαντούμε στους εξής έλεγχους υποθέσεων για το εάν το φαγητό είναι αναγκαίο αγαθό,

H_0: β_1=0 (Μηδενική Υπόθεση; Το φαγητό δεν είναι αναγκαίο ούτε πολυτέλεια.)

H_1: $\beta_1>0$ (Εναλλακτική Υπόθεση; Το φαγητό είναι αναγκαίο)

Και εάν το φαγητό είναι πολυτέλεια.

H_0: β_1=0 (Μηδενική Υπόθεση; Το φαγητό είναι πολυτέλεια.)

H_1: $\beta_1<0$ (Εναλλακτική Υπόθεση; Το φαγητό δεν είναι πολυτέλεια)

Διαστήματα Εμπιστοσύνης(Δ.Ε).

Ας περάσουμε τώρα στον υπολογισμό των διαστημάτων εμπιστοσύνης για τους συντελεστές του γραμμικού μας μοντέλου. Να σημειώσουμε ότι το SPSS μας προσφέρει τα Δ.Ε στον πίνακα 1 ως Confidence Intervals for B δίνοντας το άνω και κάτω όριο του Δ.Ε

Για τον υπολογισμό των Δ.Ε θα χρησιμοποιήσουμε τον εξής τύπο:

$$P[\beta_\iota - t_{(a/2,N-k)}\hat{\sigma}_\iota \le \hat{\beta}_\iota \le \beta_\iota + t_{(a/2,N-k)}\hat{\sigma}_\iota] = 1-a \Leftrightarrow$$
$$P[\beta_1 - t_{(0.05,8)}\hat{\sigma}_\iota \le \hat{\beta}_1 \le \beta_1 + t_{(0.05,8)}\hat{\sigma}_\iota] = 0.95 \Leftrightarrow$$
$$P[28.321 - 2.45 * 7.711 \le \hat{\beta}_1 \le 28.321 + 28.321 + 2.45 * 7.711] = 0.95$$

Άρα το Δ.Ε για το β_1 είναι το εξής :

$$28.321 - 2.45 * 7.711 \le \hat{\beta}_1 \le 28.321 + 28.321 + 2.45 * 7.711 \Leftrightarrow$$
$$9.453 \le \hat{\beta}_1 \le 47.189.$$

Ομοίως το Δ.Ε για τον συντελεστή β_2 θα δίνεται ως εξής:

$$P[\beta_2 - t_{(a/2,N-k)}\hat{\sigma}_1 \le \hat{\beta}_2 \le \beta_2 + t_{(a/2,N-k)}\hat{\sigma}_1] = 1-a \Leftrightarrow$$
$$P[\beta_2 - t_{(0.05,8)}\hat{\sigma}_1 \le \hat{\beta}_2 \le \beta_2 + t_{(0.05,8)}\hat{\sigma}_1] = 0.95 \Leftrightarrow$$
$$P[0.104 - 2.45 * 0.026 \le \hat{\beta}_2 \le 0.104 + 28.321 + 2.45 * 0.026] = 0.95$$

Άρα το Δ.Ε για το β_2 είναι το εξής :

$$0.104 - 2.45 * 0.026 \le \hat{\beta}_2 \le 0.104 + 28.321 + 2.45 * 0.026 \Leftrightarrow$$
$$0.041 \le \hat{\beta}_2 \le 0.168$$

Προβλεφθείσες Τιμές-Υπόλοιπα.

Στην περίπτωση που θέλουμε να υπολογίσουμε τις προβλέψεις του μοντέλου μας με βάση τα δεδομένα που έχουμε θα ενεργήσουμε ως εξής:

- Επιλέγουμε Analyze->Regression->Linear.Πατήστε στο κουμπί Save, και επιλέγουμε την επιλογή Unstandardized στο υπομενού Predicted Values.Με αυτόν τον τρόπο υπολογίζονται οι προβλεφθείσες τιμές, με βάση το μοντέλο μας, για την κατανάλωση.
- Εάν τώρα επιθυμούμε να υπολογίσουμε τα υπόλοιπα του μοντέλου μας θα πρέπει να ενεργήσουμε ως εξής. Επιλέγουμε *Analyze->Regression->Linear*.Πατήστε στο κουμπί Save, και επιλέγουμε την επιλογή Unstandardized στο ειπωμένου Residuals.

Παρατηρούμε ότι τόσο οι προβλέφθηκες τιμές όσο και τα υπόλοιπα του μοντέλου μας δίνονται στον Data Editor ως εξής:

pre_1	res_1
38.75000	2.25000
43.96429	1.03571
49.17857	-9.17857
54.39286	.60714
59.60714	10.39286
64.82143	-9.82143
70.03571	9.96429
75.25000	-5.25000

Στην περίπτωση τώρα που επιθυμούμε να υπολογίσουμε την κατανάλωση για φαγητό την οποία και θα έχει μια οικογένεια δεδομένου ότι το εισόδημα της θα είναι 320 εύρο θα πρέπει απλά να θέσουμε στο μοντέλο μας την τιμή X=320.

$$\hat{Y}=\hat{\beta}_1+\hat{\beta}_2X+\hat{\varepsilon}_i \Leftrightarrow \hat{Y}=28{,}321+0{,}104X \Leftrightarrow$$

$$\hat{Y}=28{,}321+0{,}104*320=61.601$$

Άρα θα έχουμε ότι η κατανάλωση μας για φαγητό δεδομένου ότι το εισόδημα της οικογενείας θα είναι 320 ευρω θα εινα 61.601 εύρο.

9. ΟΙ ΚΛΑΣΣΙΚΕΣ ΥΠΟΘΕΣΕΙΣ ΤΩΝ GAUSS-MARKOV.

Στην περίπτωση του απλού γραμμικού μοντέλου η μέθοδος των ελαχίστων τετραγώνων παράγει επιθυμητούς εκτιμητές των μεταβλητών καθώς στηρίζεται σε

Προϋπόθεση 1.Correct Specification

Η προϋπόθεση αυτή στηρίζεται στο γεγονός ότι το μοντέλο (1) στην περίπτωση μας αποτελεί το πραγματικό μοντέλο. Καταλαβαίνουμε λοιπόν ότι αυτή η προϋπόθεση είναι και δύσκολο να ικανοποιηθεί. Κοινές παραβιάσεις αυτής της προϋπόθεσης αποτελούν:

I. Λάθος συναρτησιακή μορφή του υποδείγματος. π.χ. ασφαλώς θα ήταν τεράστιο λάθος να εκτιμήσουμε την Καμπύλη του Philips μέσω μιας γραμμικής σχέσης παρά μιας αντίστροφης.

II. Παράλειψη μιας σχετικής μεταβλητής. π.χ. επίσης θα θεωρείτο ως λάθος η παράλειψη της μεταβλητής των interest rate από μια aggregate consumption function.

III. Συνυπολογισμός μια μεταβλητής μη σχετικής με την οικονομική θεωρία.

Προϋπόθεση 2. $E(\varepsilon_i) = 0, \forall$ i

Προϋπόθεση 3. Δεν υπάρχει γραμμική σχέση μεταξύ των κ επεξηγηματικών μεταβλητών

Προϋπόθεση 4. $Var(\varepsilon_i) = \sigma^2, \ \forall$ i

Προϋπόθεση 5. $Cov(\varepsilon_i, \varepsilon_j) = 0, \ \forall \text{ i} \neq j$

Προϋπόθεση 6. Οι επεξηγηματικές μεταβλητές είναι μη στοχαστικές.

9.2 ΤΡΙΜΕΤΑΒΛΗΤΟ ΓΡΑΜΜΙΚΟ ΥΠΟΔΕΙΓΜΑ

Το υπόδειγμα που αναπτύχθηκε στην προηγούμενη ενότητα αναφέρεται σε σχέσεις οι οποίες και περιλαμβάνουν μια μόνο ανεξάρτητη μεταβλητή. Τα υποδείγματα τα οποία και περιλαμβάνουν περισσότερες από μια ανεξάρτητες μεταβλητές καλούνται πολυμεταβλητα υποδείγματα παλινδρόμησης. Σε αυτήν την παράγραφο θα αναλύσουμε την περίπτωση όπου η οικονομική σχέση περιλαμβάνει δυο ανεξάρτητες μεταβλητές. Ασφαλώς το κεφάλαιο αυτό τελειώνει με την μέθοδο των μητρών.

Έστω ότι η στοχαστική σχέση η οποία και συνδέει την εξαρτημένη μεταβλητή με τις δυο ανεξάρτητες μεταβλητές είναι γραμμική και ακολουθεί την σχέση $Y=f(x_1, x_2, \varepsilon) = \beta_0 + \beta_0 x_1 + \beta_0 x_1 + \varepsilon$. Σε αυτό σημείο θα πρέπει να τονίσουμε ότι οι υποθέσεις των Gauss-Markov ισχύουν ως έχει αλλά θα πρέπει να προσθέσουμε δυο επιπλέον. Πρώτα, το γεγονός ότι δεν υπάρχει ακριβής γραμμική σχέση ανάμεσα στις ανεξάρτητες μεταβλητές του μοντέλου μας, αποφεύγοντας έτσι το πρόβλημα της πλήρους πολυσυγραμμικοτητας (perfect multicollinearity). Δεύτερον ότι οι βαθμοί ελευθέριας θα πρέπει να είναι θετικοί. Αυτό σημαίνει ότι ο αριθμός των παρατηρήσεων του δείγματος θα πρέπει να είναι μεγαλύτερος από τον αριθμό των συντελεστών του οικονομετρικού μας υποδείγματος.

9.2 Η Ελαχιστοποίηση των Τετραγώνων των Καταλοίπων

Σκοπός μας όπως και στο απλό γραμμικό μοντέλο η εκτίμηση της συναρτησιακής μας σχέσης $Y=f(x_1, x_2, \varepsilon) = \beta_0 + \beta_0 x_1 + \beta_0 x_1 + \varepsilon$ η οποία και εκτιμείται προσαρμόζοντας ένα οικονομετρικό

υπόδειγμα που έχει την εξής μορφή $Y=\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\varepsilon}_\iota$.Το μη ερμηνευμένο μέρος $\hat{\varepsilon}_\iota = Y_i - \hat{Y}_\iota$, ονομάζεται κατάλοιπο και αποτελεί μια εκτίμηση του αγνώστου όρου σφάλματος ε_ι .Για να βρούμε όμως τιμές για τις άγνωστες παραμέτρους χρειαζόμαστε κάποιο κριτήριο που στην περίπτωση μας είναι η μέθοδος των ελαχίστων τετραγώνων. Σύμφωνα με αυτήν:

$$\min \sum_{i=1}^{n} \hat{\varepsilon}_\iota^2 = \sum_{t=1}^{\nu} (Y_\iota - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2)^2$$

Οι συνθήκες πρώτης τάξης ως προς τα $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ δίνουν αμερόληπτους και αποτελεσματικούς εκτιμητες οι οποίοι και έχουν τις εξής μορφές:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \overline{x_1} - \hat{\beta}_2 \overline{x_2} \qquad \textbf{(1)}$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{\nu} (Y_i X_1) \sum_{t=1}^{\nu} (X_2)^2 - \sum_{t=1}^{\nu} (Y_i X_2) \sum_{t=1}^{\nu} (X_1 X_2)}{\sum_{t=1}^{\nu} (X^2{}_1) \sum_{t=1}^{\nu} (X^2{}_2) - \sum_{t=1}^{\nu} (X_1 X_2)} \qquad \textbf{(2)}$$

$$\hat{\beta}_2 = \frac{\sum_{\iota=1}^{\nu}(Y_i X_2)\sum_{\iota=1}^{\nu}(X_1)^2 - \sum_{\iota=1}^{\nu}(Y_i X_2)\sum_{\iota=1}^{\nu}(X_1 X_2)}{\sum_{\iota=1}^{\nu}(X^2{}_1)\sum_{\iota=1}^{\nu}(X^2{}_2) - \sum_{\iota=1}^{\nu}(X_1 X_2)} \qquad \textbf{(3)}$$

Καθώς το υπόδειγμα μας έχει υπολογιστεί με την μέθοδο των Ελαχίστων τετραγώνων, οι ιδιότητες που ισχύουν είναι ανάλογες με το απλό καθώς και με το πολλαπλό γραμμικό υπόδειγμα. Δηλαδή,

- Η γραμμική παλινδρόμηση διέρχεται από τα σημεία $\overline{Y}, \overline{X_1}, \overline{X_2}$.
- Τα υπόλοιπα δεν συσχετίζονται με τις ανεξάρτητες μεταβλητές.
- Ισχύει ότι $\overline{Y} = \overline{\hat{Y}}$
- Τέλος οι συντελεστές της παλινδρόμησης είναι αμερόληπτοι και αποτελεσματικοί.

9.3 Συντελεστής Προσδιορισμού- Διορθωμένος Συντελεστής Προσδιορισμού

Ο συντελεστής πολλαπλού προσδιορισμού παρέχει το ποστοστο της μεταβλητικοτητας της εξαρτημένης μεταβλητής Y, το οποίο και εξηγείται από την παλινδρόμηση. Ο συντελεστής αυτός υπολογίζεται από την εξής σχέση,

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{N}(\hat{Y} - \overline{Y})^2}{\sum_{i=1}^{N}(Y - \overline{Y})^2} = \frac{\hat{\beta}_1 \sum_{i=1}^{N}(X_1 - \overline{X_1})(Y - \hat{Y}) + \beta_2 \sum_{i=1}^{N}(X_2 - \overline{X_2})(Y - \hat{Y})}{\sum_{i=1}^{N}(Y - \hat{Y})^2}.$$

Είναι γνωστό όμως ότι ο συντελεστής αυτό επηρεάζεται τόσο από το μέγεθος του δείγματος αλλά και από τον αριθμό των ανεξάρτητων μεταβλητων. Ο διορθωμένος συντελεστής προσδιορισμού δίνεται απ τον εξής τύπο:

$$\overline{R}^2 = \frac{\sum_{i=1}^{N}(Y - \hat{Y})^2 \big/ N-k}{\sum_{i=1}^{N} Y^2 \big/ N-1}$$, όπου N ο αριθμός των παρατηρήσεων και κ, ο αριθμός των ανεξάρτητων μεταβλητών του υποδείγματος.

9.4 Μερικοί Συντελεστές Προσδιορισμού.

Οι μερικοί συντελεστές προσδιορισμού μετρούν το καθαρό ποσοστό της μεταβλητικότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από την ανεξάρτητη μεταβλητή, απαλλαγμένη από την επίδραση άλλων μεταβλητών στο υπόδειγμα παλινδρόμησης. Συγκεκριμένα, με βάση το τριμεταβλητο γραμμικό υπόδειγμα μπορούμε υπολογίσουμε τρεις συντελεστές προσδιορισμού:

Τον μερικό συντελεστή προσδιορισμού ανάμεσα στην Y και Xι , υπό τον περιορισμό ότι X_2 παραμένει σταθερή .

$r^2_{y,x_1} = \dfrac{\hat{\beta}_1^{\,2} \sum_{i=1}^{N} (X_1 - \overline{X_1})}{\sum_{i=1}^{N} (Y - \overline{Y})^2}$, τον μερικό συντελεστή προσδιορισμού ανάμεσα στην Y και X_2 , υπό τον περιορισμό ότι X_1 παραμένει σταθερή $r^2_{y,x_2} = \dfrac{\hat{\beta}_2^{\,2} \sum_{i=1}^{N} (X_2 - \overline{X_2})}{\sum_{i=1}^{N} (Y - \overline{Y})^2}$, και τέλος τον μερικό συντελεστή ανάμεσα στις X_1, X_2 δεδομένου ότι η Y παραμένει σταθερή $r^2_{x_1,x_2} = \dfrac{\sum_{i=1}^{N} [(X_1 - \overline{X_1})(X_2 - \overline{X_2})]^2}{\sum_{i=1}^{N} (X_1 - \overline{X_1})^2 \sum_{i=1}^{N} (X_2 - \overline{X_2})^2}$

9.5 *Διακυμάνσεις και τυπικά σφάλματα των εκτιμητών ελαχίστων τετραγώνων.*

9.6 *Έλεγχοι υποθέσεων στο τριμεταβλητο γραμμικό υπόδειγμα*

A. ΔΙΠΛΕΥΡΟΣ ΕΛΕΓΧΟΣ

Στον δίπλευρο έλεγχο σκοπός μας είναι η διερεύνηση της παρακάτω υποθέσεως: Εάν υπάρχει στατιστικά σημαντική σχέση ανάμεσα στην εξαρτημένη και στην ανεξάρτητη μεταβλητή. Με αλλά λόγια εάν η κλίση της συνάρτησης παλινδρόμησης μας θα είναι θετική ή αρνητική με βάση την συναρτησιακή σχέση ανάμεσα στα X,Y. O έλεγχος διαμορφώνεται ως εξής:

H_0: β=0 (Μηδενική Υπόθεση;)

H_1: $\beta \succ 0$ (Εναλλακτική Υπόθεση;)

B. ΔΕΞΙΟΣ ΜΟΝΟΠΛΕΥΡΟΣ ΕΛΕΓΧΟΣ

Ας υποθέσουμε ότι με βάση την οικονομική θεωρία αναμένουμε ότι ο συντελεστής βι είναι θετικός και θέλουμε να διαπιστώσουμε εμπειρικά αν ο β; είναι, πρώτο, θετικός και, δεύτερο, στατιστικά σημαντικός. Σε αυτή την περίπτωση, οι H_0 και H, υποθέσεις εκφράζονται ως εξής:

H_0: β=0 (Μηδενική Υπόθεση;)

H_1: $\beta \succ 0$ (Εναλλακτική Υπόθεση;)

Γ. ΑΡΙΣΤΕΡΟΣ ΜΟΝΟΠΛΕΥΡΟΣ ΕΛΕΓΧΟΣ

Ας υποθέσουμε ότι με βάση την οικονομική θεωρία αναμένουμε ότι ο συντελεστής βι είναι αρνητικός και θέλουμε να διαπιστώσουμε εμπειρικά αν ο β; είναι, πρώτο, είναι αρνητικός και δεύτερο, στατιστικά σημαντικός. Σε αυτή την περίπτωση, οι H_0 και Η, υποθέσεις εκφράζονται ως εξής:

H_0: $\beta=0$ (Μηδενική Υπόθεση;)

H_1: $\beta \prec 0$ (Εναλλακτική Υπόθεση;)

Δ. ΕΛΕΓΧΟΣ ΙΣΟΤΗΤΑΣ ΣΥΝΤΕΛΕΣΤΩΝ

Ο έλεγχος ισότητας συντελεστών περιλαμβάνει απάντηση στο εξής ερώτημα:

H_0: $\beta_1 = \beta_2$ (Μηδενική Υπόθεση;)

H_1: $\beta_1 \neq \beta_2$ (Εναλλακτική Υπόθεση;)

Υπό την προϋπόθεση ότι οι κλασσικές υποθέσεις ικανοποιούνται η παραπάνω σχέση ακολουθεί κανονική κατανομή με N-3 βαθμούς ελευθερίας.

Άρα $$t = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - (\beta_1 - \beta_2)}{\sqrt{\operatorname{var}(\hat{\beta}_1 - \hat{\beta}_2)}} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\sigma^2_{\hat{\beta}_1 - \hat{\beta}_2}}}, \text{οπου}$$

$$\sigma^2_{\hat{\beta}_1 - \hat{\beta}_2} = \sqrt{\operatorname{var}(\hat{\beta}_1) + \operatorname{var}(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)}$$

Εάν $t > t_{a/2, N-3}$ η αρχική υπόθεση γίνεται δεκτή και αντίστροφα.

9.7 Αξιολόγηση του συνόλου του υποδείγματος.

Η αξιολόγηση του συνόλου του υποδείγματος μπορεί να ελεγχθεί με την F-στατιστική απαντώντας στον εξής έλεγχο υποθέσεων:

H_o: όλοι οι συντελεστές παλινδρόμησης εκτός από το β_0 είναι μηδέν.

<u>έναντι</u>

H_1: μερικοί από τους συντελεστές είναι διαφορετικοί του μηδενός. Ο υπολογισμός της F-στατιστικής γίνεται με τον εξής τύπο $F = \dfrac{\dfrac{\sum_{i=1}^{n}(\hat{Y} - \overline{Y})^2}{k-1}}{\dfrac{\sum_{i=1}^{n}(Y - \hat{Y})^2}{N-k}}$. Όπως γνωρίζουμε και από τους υπόλοιπους έλεγχους η αρχική μα υπόθεση θα γίνεται αποδεκτή εάν $F < F_{(a,k-1,N-k)}$ ενώ θα απορρίπτεται εάν $F > F_{(a,k-1,N-k)}$

<u>ΑΣΚΗΣΕΙΣ</u>

1. Μια εταιρεία έρευνας προστασίας του καταναλωτή ερευνά την ζήτηση (Q_D) και την αντίστοιχη τιμή (P) που για μια χρηματοοικονομική υπηρεσία η οποία προσφέρεται από μια τοπική τράπεζα. Τα δεδομένα είναι τα εξής:

Q_D	33	87	128	131	189	228	256	308	299	327	360	336	303	342	434	391	336	250
P	69.6	65.8	55.3	51.2	56.8	51.2	43.8	41.3	53.7	52	36.2	33.2	38.3	35.2	30.5	36.5	25.7	36,1

a. Ποια είναι η εξίσωση παλινδρόμησης; Ερμηνεύσατε τα αποτελέσματα.

b. Ποια η τιμή του συντελεστή συσχέτισης; Σχολιάστε την γραφική παράσταση των μεταβλητών.

c. Ποιος ο εκτιμώμενος ρυθμός μεταβολής της ζήτησης για αυτήν την χρηματοοικονομική υπηρεσία σε σχέση με την τιμή;

d. Ποια η εκτίμηση του τυπικού σφάλματος του υποδείγματα;.

e. Πλάι η τιμή του εκτιμώμενου τυπικού σφάλματος; Ορίστε διάστημα εμπιστοσύνης 90% για την κλίση.

2. Από 12 κτήματα που αποτελούν ένα τυχαίο δείγμα του συνόλου των καλλιεργημένων με ορισμένο γεωργικό προϊόν εδαφών σε μια περιοχή συγκεντρώθηκαν στοιχεία παραγωγής (Y σε τόνους) και εκτάσεων (X σε στρέμματα) για μια καλλιεργητική εποχή. Με βάση τα στοιχεία αυτά υπολογίστηκαν τα ακόλουθα αθροίσματα:

$\Sigma X = 480 \quad \Sigma Y = 92 \quad \Sigma X^2 = 21050 \quad \Sigma Y^2 = 850 \quad \Sigma XY = 4370$

α) Να εκτιμηθεί η εξίσωση παλινδρόμησης της Y πάνω στην X.

β) Να εξετασθεί σε επίπεδο σημαντικότητας 0,05 αν στο σύνολο των κτημάτων της περιοχής οι τιμές των συντελεστών β_0 και β_1 της εξίσωσης παλινδρόμησης είναι ίσες με το μηδέν.

3. Δίνονται τα ακόλουθα στοιχεία κατανάλωσης ρυζιού (Y σε τόνους) και της τιμής ανά τόνο ρυζιού (X σε χιλιάδες δρχ.) για μια χρονική περίοδο 10 ετών, όπως φαίνεται στον παρακάτω πίνακα:

Έτη	1	2	3	4	5	6	7	8	9	10
X	12	16	18	20	26	28	32	37	42	48
Y	84	86	80	70	72	66	62	65	60	63

α) Να εκτιμηθεί η γραμμική εξίσωση κατανάλωσης ρυζιού.

β) Να εκτιμηθεί η ποσότητα ρυζιού που κατά μέσο όρο αναμένεται να καταναλωθεί σε περίπτωση διαμόρφωσης της τιμής του ρυζιού σε 52 χιλ. δρχ. και να καθοριστεί το διάστημα εμπιστοσύνης σε επίπεδο σημαντικότητας α = 0,05 για την πληθυσμιακή κλίση β,.

γ) Ελέγξατε σε επίπεδο 5% αν οι συντελεστές της εξίσωσης αυτής είναι στατιστικά σημαντικοί.

δ) Ποια η ελαστικότητα τιμής ζήτησης;

4. Ο παρακάτω πίνακας δείχνει τα έτη υπηρεσίας χ και το ύψος των μηνιαίων τακτικών αποδοχών γ σε χιλιάδες δρχ. 8 εργατών ενός τμήματος παραγωγής σ' ένα εργοστάσιο:

Έτη (χ)	1	3	4	5	5	6	6	8
Αποδοχές(Υ)	150	160	165	165	170	170	175	180

Ζητείται να βρεθεί ο συντελεστής γραμμικής συσχετίσεως ρ και να σχολιασθεί το αποτέλεσμα.

5. Ο παρακάτω πίνακας δείχνει τα ετήσια εισοδήματα χ σε δεκάδες χιλιάδες δρχ. και τα αντίστοιχα ετήσια έξοδα τηλεφωνικών λογαριασμών γ σε δεκάδες χιλιάδες δρχ. μιας ομάδας 8 νοικοκυριών:

χ:	300	350	400	450	500	550	600
γ:	9	8	10	13	13	16	15

Ζητείται:

α) Να βρεθεί ο συντελεστής γραμμικής συσχετίσεως ρ.

β) Να υπολογισθούν οι παράμετροι α και β της εξισώσεως παλινδρομήσεως.

6. Ο παρακάτω πίνακας αναφέρεται σε 10 ομοειδείς εμπορικές επιχειρήσεις και δείχνει τις ετήσιες πωλήσεις τους γ και τις αντίστοιχες ετήσιες διαφημιστικές τους δαπάνες χ. Τα δεδομένα εκφράζονται σε εκατ. δρχ.:

χ:	20	22	25	30	32	35	40	50	60	70
γ:	250	270	260	350	330	300	370	480	450	500

Ζητείται:

α) Να βρεθεί ο συντελεστής γραμμικής συσχετίσεως ρ και να σχολιασθεί το αποτέλεσμα.

β) Να προσδιορισθεί η εξίσωση γραμμικής παλινδρομήσεως.

7. Ο παρακάτω πίνακας δείχνει τα ετήσια εισοδήματα χ σε εκατομμύρια δρχ. και τις αντίστοιχες δαπάνες διατροφής γ σε εκατομμύρια δρχ. μιας ομάδας 7 νοικοκυριών:

χ:	3	3,3	3,5	3,8	4	4,2	4,3
γ:	1,3	1,3	1,4	1,5	1,6	1,7	1,7

Ζητείται:

α) Να βρεθεί ο συντελεστής γραμμικής συσχετίσεως ρ.

β) Να βρεθεί η εξίσωση παλινδρομήσεως γ = α + β χ.

8. Για να προσδιοριστεί η επίδραση ενός νέου φαρμάκου στην αρτηριακή πίεση, χορηγήθηκε το φάρμακο σ'ένα άρρωστο για 8 μέρες και σε διαφορετικές δόσεις. Τα αποτελέσματα φαίνονται στον πίνακα:

Ημερήσιος αριθμός χάπι	4	3	2	1	4	5	3	2
Αρτηριακή πίεση, Υ	16	14	13	Π	15	10	12	

α) Να εκτιμηθεί η ευθεία Υ=α+βχ+ε.

β) Να γίνει γραφική απεικόνιση και να σχεδιαστεί η ευθεία Υ=α+βχ. Να χρησιμοποιηθεί αυτή η ευθεία για να εκτιμηθεί η πίεση για 2,5 χάπια την ήμερα.

γ) Να υπολογιστεί η διασπορά των σφαλμάτων κι ο συντελεστής συσχέτισης..

9. Ο συντελεστής συσχέτισης Γ σε ένα τυχαίο δείγμα, μεγέθους ν = 18, δυο μεταβλητών Χ και Υ, είναι ίσος με - 0,89. Να ελέγξετε, σε επίπεδο σημαντικότητας α = 0,05 αν υπάρχει γραμμική συσχέτιση μεταξύ των Χ και Υ.

10. Στα παρακάτω εργοστάσια εργαζόταν 1990 το εξής προσωπικό; M_1= μηχανικοί, M_2= τεχνολόγοι μηχανικοί-εργοδηγοί, Ε= εργάτες, E_1= χαλυβουργεία, E_2= τσιμεντοβιομηχανία, E_3= εργοστάσια χαρτιού, E_4.= εργοστάσιο ξύλου, E_5-= ναυπηγεία,

E_6= αλευροβιομηχανία

	E_1	E_2	E_3	E_4	E_5	E_6
M_1	7	10	4	4	5	5
M_2	10	14	9	11	20	16
E	530	460	420	480	500	820

α) Να γίνει γραμμική παλινδρόμηση 1) του αριθμού των εργατών στον αριθμό των τεχνολόγων μηχανικών-εργοδηγών, 2) του αριθμού μηχανικών στον αριθμό των εργατών.

β) Μια βιομηχανία ζυμαρικών έχει 295 εργάτες. Πόσους μηχανικούς και πόσους τεχνολόγους μηχανικούς-εργοδηγούς πρέπει να έχει;

11. Στον ακόλουθο πίνακα δίνονται για 10 πυρκαγιές που συνέβησαν σε μια πόλη. Ο χρόνος (X) σε min, που χρειάστηκε η πυροσβεστική μέχρι να φτάσει στον τόπι πυρκαγιάς και το κόστος (Y) σε εκατ. δρχ. της ζημιάς που προκάλεσε κάθε πυρκαγιά:

Πυρκαγιά	Χρόνος (min)	Κόστος κατάσβεσης (εκατ.ευρο)
1	23	31
2	30	45
3	15	20
4	24	36
5	27	38
6	19	25
7	12	25
8	21	28
9	18	24
10	20	30

Ζητούνται:

α) Η εξίσωση γραμμικής παλινδρόμησης της Y πάνω στη X.

β) Να ελέγξετε τη μηδενική υπόθεση Ho : β = O με εναλλακτική την H_1 : β > 0. θεωρούμε α = 0,05.

γ) Βρείτε ένα 90 % διάστημα εμπιστοσύνης για το β και να ερμηνεύσετε το αποτέλεσμα.

12. Τα ποσά μιας χημικής ένωσης γ που διαλύονται σε 100 gr νερού σε διαφορές θερμοκρασίες χ, φαίνονται στον πίνακα

χ° C	Y, gr		
0	8	6	7
15	12	10	11
30	25	20	23
45	31	32	33
60	42	40	43
75	47	51	46

α) Να βρεθεί η ευθεία παλινδρόμησης,

β) Να απεικονιστούν οι τιμές χ και γ και να γίνει το γράφημα της ευθείας παλινδρόμησης

γ) Να εκτιμηθεί το ποσό της χημικής ένωσης, που θα διαλυθεί σε 100 gr νερού στους 50^0 C.

13. Ο ρυθμός των ελαττωματικών προϊόντων ανά 1000 κομμάτια, που παράγει μια μηχανή, σε μια περίοδο 6 ήμερων είναι η εξής:

Ημέρα, χ	1	2	3	4	5	6
Ρυθμός ελαττωματικών,	8	11	16	19	25	29

α)Να απεικονιστούν οι τιμές χ και γ γραφικά,

β) Να βρεθεί η :γραμμή παλινδρόμησης (ευθεία ελαχίστων τετραγώνων) του ρυθμού ελαττωματικών ως προς την ήμερα παραγωγής,

γ) Να βρεθεί το τυπικό σφάλμα της εκτίμησης.

14. Γίνεται μια μελέτη για αν βρεθεί ο αριθμός των αστοχιών μιας αυτόματης παραγωγικής διαδικασίας, όταν η μηχανή εργάζεται 4,6,8,10 και 12 εβδομάδες χωρίς ημερήσια προληπτική συντήρηση. Τα αποτελέσματα ήταν τα εξής:

Αριθμός εβδομάδων χωρίς συντήρηση (Χ)	4	6	8	10	12
Αριθμός αστοχιών (Υ)	8	6	8	14	16

α)Να απεικονιστούν οι τιμές Χ και Υ γραφικά,

β) Να βρεθεί η γραμμή παλινδρόμησης (ευθεία ελαχίστων τετραγώνων) του ρυθμού ελαττωματικών ως προς την ήμερα παραγωγής,

γ) Να βρεθεί το τυπικό σφάλμα της εκτίμησης. Και να εκτιμηθεί η διασπορά..

15. Στον παρακάτω πίνακα φαίνεται η μελέτη που γίνεται για να προσδιοριστεί η σχέση μεταξύ των μηνιαίων διαφημιστικών εσόδων και των πωλήσεων ενός supermarket.

Διαφημίσεις (χιλ. ευρω)	40	20	25	20	30	50	40	20	50	40	25	50
Πωλήσεις (χιλ. ευρω)	390	380	395	360	440	420	460	400	540	500	410	500

α) Να παρασταθούν τα δεδομένα σε ορθογώνιο σύστημα αξόνων χ και γ.

β) Να βρεθεί η γραμμή παλινδρόμησης και να γίνει το γράφημα της.

γ) Να εκτιμηθεί η διασπορά 3 .

δ) Να εκτιμηθούν οι πωλήσεις για έξοδα διαφημίσεων 80 και 100 χιλ. δρχ.

ε) Να βρεθεί ένα 90% δ,ε. για το μέσο επίπεδο πωλήσεων, όταν γίνονται έξοδα διαφημίσεων 50 χιλ. δρχ.

στ) Να βρεθούν 90% Δ.Ε για τα α,β της παλινδρόμησης.

16. Στον παρακάτω πίνακα φαίνεται ο αριθμός αυτοκινητιστικών ατυχημάτων , σε σχέση με το αριθμό χ των αυτοκίνητων που καταμετρούνται στα διόδια,

α) Να παρασταθούν τα δεδομένα σε ορθογώνιο σύστημα αξόνων χ και ψ.

β) Να προσαρμοστούν τα δεδομένα σε παραβολή ελαχίστων τετραγώνων.

Αριθμός αυτοκίνητων (X)	200	300	400	500	600	700
Αριθμός ατυχημάτων (Y)	1	4	8	7	3	1

17. Για τα δεδομένα της άσκησης 15 να βρεθούν;

α) ένα 90% δ.ε. για το μέσο ρυθμό ελαττωματικών προϊόντων για μια περίοδο 4 ήμερων,

β) ένα 90% δ.ε. για την αναμενόμενη τιμή του ρυθμού ελαττωματικών γ,

Ε(γ), για κάθε αριθμό ήμερων, δηλ. όταν χ=1,2, 3,4,5,6 γ) το γράφημα της ζώνης εμπιστοσύνης,

δ) να γίνει έλεγχος της υπόθεσης για την Ε(Υ) ότι ο ρυθμός ελλατ5ωματικων είναι τουλάχιστον 16 για 4 ημέρες παραγωγής..

18. Για τα δεδομένα της άσκησης 11.8 θέλουμε να βρούμε:

α) ένα 90% δ.ε. για το μέσο αριθμό αστοχιών, όταν μεσολαβούν 8 εβδομάδες χωρίς συντήρηση,

β) ένα 90% δ.ε. για την αναμενόμενη τιμή της γ, Ε(γ), για κάθε αριθμό εβδομάδων χωρίς συντήρηση, δηλ. όταν χ=4,6,8,10,12,

γ) το γράφημα της ζώνης εμπιστοσύνης,

δ) να γίνει έλεγχος της υπόθεσης, για την Ε(γ), ότι ο αριθμός αστοχιών ξεπερνά τις 14 αν περάσουν 10 εβδομάδες χωρίς συντήρηση .

19. Στον παρακάτω πίνακα φαίνονται οι γεννήσεις (ανά 1000 κατοίκους) σε σχέση με ημερολογιακό χρόνο,

α) Να παρασταθούν γραφικά τα δεδομένα του πίνακα,

β) Να προσαρμοστούν τα δεδομένα σε ορθογώνιο σύστημα αξόνων χ και γ.

γ) Να προσαρμοστούν τα δεδομένα σε παραβολή ελαχίστων τετραγώνων,

δ) Είναι η εξίσωση της παραβολής αξιόπιστη για μακροχρόνια πρόβλεψη;

Χρόνος	1950	1955	1960	1965	1970	1975	1980	1985
Γεννήσεις ανά 1000 κατοίκους	19,3	17,9	17,1	13	14,4	16,5	18,5	18,8

20. Τα τελευταία 8 χρόνια ο αριθμός Χ των υποψήφιων που μπήκαν στα ΑΕΙ/ΑΤΕΙ και ο αριθμός των φοιτητών που έχουν καθυστερήσει πάνω από ένα εξάμηνο να πάρουν πτυχίο δίνονται στον κάτωθι πίνακα.

Έτος	1982	1983	1984	1985	1986	1987	1988	1989
Αριθμός φοιτητών Χ	80,000	82,000	83,500	86,000	85,000	84,000	84,500	83,000
Αριθμός φοιτητών με καθυστέρηση Υ	2500	2600	3000	3600	3000	2750	2800	3000

α) Να βρεθεί η μέση τιμή και η διασπορά του συνόλου των υποψήφιων που μπαίνουν στα ΑΕΙ/ΑΤΕΙ.

β) Να βρεθεί η μέση τιμή και η διασπορά των φοιτητών με καθυστέρηση πάνω από ένα εξάμηνο να πάρουν πτυχίο.

γ) Να γίνει γραμμική παλινδρόμηση της Υ στη Χ.

δ) Ο αριθμός των υποψήφιων που μπαίνουν στα ΑΕΙ/ΤΕΙ το 1990 είναι 75000. Ποιος θα είναι κατά μέσο ορό ο αριθμός αυτών που καθυστερούν τουλάχιστον ένα εξάμηνο, να πάρουν πτυχίο;

ε) Να βρεθεί ένα 95% δ. ε. για τον αριθμό αυτών που καθυστερούν στη λήψη πτυχίου για το 1990. Το ίδιο για το β.

στ) Να ελεγχθεί σε επίπεδο σημαντικότητας 5% αν το ποσοστό αυτών που καθυστερούν στη λήψη πτυχίου ήταν μεγαλύτερο το 1984 από ότι το 1989.

21. Η παραγωγή ενός εργοστάσιου σε χιλ. κομμάτια από το 1979 μέχρι και το 1989 ήταν:

X	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Y	,42	4,7	5,1	5,3	5,6	5,9	6,4	7,1	7,6	7,7	8

α) Να γίνει γραμμική παλινδρόμηση της Y στην X.

β) Να προβλεφθεί η μέση παραγωγή του εργοστασίου για τα χρονιά 1990, 1991 και γ) το 1991 η παραγωγή ήταν 7,9 χιλ. κομμάτια. Το περιμέναμε;

22. Για τα δεδομένα της άσκησης Π.8 να βρεθούν:

α) ένα 90% δ.ε. για τον αριθμό ατυχημάτων, όταν περνούν τα διόδια 600 αυτοκίνητα,

β) ένα 90% δ.ε. για την αναμενόμενη τιμή του γ,Ε(γ), για κάθε αριθμό αυτοκίνητων που περνούν τα διόδια, δηλ. όταν χ=200,300,400,500,600,700,

γ) το γράφημα της ζώνη εμπιστοσύνης,

δ) να γίνει έλεγχος της υπόθεσης, για την Ε(γ),ότι ο αριθμός ατυχημάτων ξεπερνά τα 8 αν περάσουν τα διόδια τουλάχιστον 400 αυτοκίνητα, και

ε) ένα 90% διάστημα πρόβλεψης για τον αριθμό ατυχημάτων, όταν περάσουν τα διόδια 500 αυτοκίνητα.

23. Αν N_0 αριθμός των βακτηριδίων σε μια χημική ένωση σε μια χρονική στιγμή σε μια χρονική στιγμή και N_1 ο αριθμός τους , :t χρονικές μονάδες αργότερα, τότε συνήθως δεχόμαστε το μοντέλο

$N_1=N_0\,K\,e^{\beta t}$ όπου n είναι ο μέγιστος αριθμός βακτηριδίων /mgr που μπορεί να υπάρξει . Αν για δυο διαφορετικές θερμοκρασιακές καταστάσεις δεχτούμε

Α) η= 100 βακτ/mgr, Β) η= 135 βακτ/mgr, τότε: t

α) Να εκτιμηθεί η ευθεία παλινδρόμησης και για τις δυο περιπτώσεις από τη σχέση $Y_t= α+βt+ε$,όπου: $Y_t=\ln N_1/N_0$, $α=\ln K$

β) Να εκτιμηθεί ο αριθμός βακτηριδίων /mgr όταν $t=40^0$ C.= 80°C

γ) Να βρεθεί 95% δ.ε. για τον αριθμό βακτηριδίων όταν έχουμε 200 gr της χημικής ένωσης σε θερμοκρασίες 50 και 70^0 C.

23. Μία στατιστική εταιρία έκανε μια έρευνα για να βρει την σχέση ανάμεσα στο εισόδημα 8 νοικοκυριών και τη κατανάλωση πετρελαίου. Βρέθηκαν τα παρακάτω αποτελέσματα:

Νοικοκυριό	Εισόδημα σε ευρώ	Πετρέλαιο σε ευρώ
1	300	82
2	450	90
3	600	80
4	750	110
5	900	140
6	1050	110
7	1200	160
8	1350	140

Α να βρεθεί το scatter-plot

Β να βρεθούν τα Β0 και Β1 και να ερμηνευθούν

Γ να βρεθούν τα υπόλοιπα.

24. Έστω ότι έχουμε μια ομάδα αναλυτών που εργάζεται για μια εταιρία ηλεκτρικής ενέργειας η οποία προκειμένου να εισέλθει στην αγορά κάνει έρευνα για την σχέση που έχει η κατανάλωση σε ενέργεια σε έναν σπίτι-νοικοκυριό και το εισόδημα του νοικοκυριού:

Νοικοκυριά	Εισόδημα σε ευρώ	Κατανάλωση σε ευρώ
1	200	82
2	300	90
3	400	80
4	500	110
5	600	140
6	700	110
7	800	160
8	900	140

A. Να δείξετε το scatter-plot.

B. Να βρεθούν τα Β0 και Β1 και να δοθεί η ερμηνεία τους.

C. Να υπολογιστούν τα υπόλοιπα.

25. Η Ε.Σ.Υ.Ε ανακοίνωσε για το έτος 2002 μια μελέτη για τα τροχαία ατυχήματα (χ) σε σχέση με τους οδηγούς που έχουν καταναλώσει αλκοόλ (ψ)

Χι :1, 3 ,4 ,6 ,8 ,9 ,11, 14

Υι :1, 2, 4, 4, 5, 7, 8, 9

Ζητείται

α)Διάγραμμα διασποράς

β)Να βρεθεί το κατάλληλο γραμμικό μοντέλο και να γίνει προσδιορισμός B_0, B_1.

γ) Υπολογισμός residuals.

26. Όταν ο εθνικός προϋπολογισμός αυξάνει κάθε χρόνο ένας καθοριστικός παράγοντας γι'αυτό είναι ότι οι δαπάνες για την άμυνα της χώρας αυξάνουν. Ο επόμενος πίνακας δείχνει τις κατά κεφαλήν δαπάνες μιας χώρας αμυντικούς σκοπούς σε σχέση με το εθνικό της εισόδημα από το 1976 έως το 1982.

Έτος	Αμυντικές Δαπάνες (εκκ.δολαρια)	Εθνικό Εισόδημα (δις.Δολαρια)
1976	423	2,827
1977	465	2,959
1978	499	3,115
1979	553	3,142
1980	631	3,187
1981	739	3,244
1982	846	3,166

Ζητούνται :

α)Να γίνει το διάγραμμα διασποράς

β)Να βρεθεί το κατάλληλο γραμμικό μοντέλο και να γίνει προσδιορισμός B_0 και B_1

γ) να υπολογιστούν τα υπόλοιπα (residuals).

δ) Ποια η κατά κεφαλή αμυντική δαπάνη όταν το εθνικό εισόδημα είναι 4.521

ΒΙΒΛΙΟΓΡΑΦΙΑ

a. Εισαγωγή στην οικονομετρία. Τόμος Α. Γ.Κ.Χρήστου. Εκδόσεις Gutenberg.

ΣΕΜΙΝΑΡΙΟ 1.ΕΝΝΟΙΕΣ ΚΑΙ ΜΕΝΟΥ ΤΟΥ ΣΤΑΤΙΣΤΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ SPSS

ΣΕΜΙΝΑΡΙΟ 2-3. ΤΟ ΑΠΛΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ- ΘΕΩΡΙΑ

Δίνεται ο ακόλουθος πίνακας.

Εισόδημα (X):	100	150	200	250	300	350	400	450
Κατανάλωση σε Φαγητό (Υ):	41	45	40	55	70	55	80	70.

1. Σε διάγραμμα διασπορές να σχεδιάσετε την μεταβλητή Y έναντι στην μεταβλητή X. Μπορείτε να περιγράψετε μια σχέση ανάμεσα στις δυο αυτές μεταβλητές;
2. Να υπολογίσετε τους εκτιμητές του οικονομετρικού σας μοντέλου. Ποιες οι ερμηνείες που δίνετε για τις τιμές των εκτιμητών αυτών;
3. Να υπολογίσετε τα υπόλοιπα του μοντέλου σας και να βρείτε το μεγαλύτερο θετικό. Πως το ερμηνεύετε;
4. Είναι το φαγητό κανονικό, κατώτερο ή ανώτερο αγαθό; Είναι αναγκαιότητα ή είδος πολυτελείας; Να σχηματίσετε τους απαραίτητους έλεγχους υποθέσεων ώστε να απαντήσετε στα ανάλογα ερωτήματα.
5. Να υπολογίσετε ένα Δ.Ε 95% για την οριακή επίδραση του εισοδήματος στην δαπάνη για φαγητό. Να δώσετε μια ερμηνεία για το Δ.Ε που υπλογισατε. Πως το διάστημα εμπιστοσύνης που υπολογίσατε συνδέεται με την απάντηση σας στο ερώτημα 4;
6. Να υπολογίσετε την εισοδηματική ελαστικότητα για φαγητό για ένα νοικοκυριό με εισόδημα 225 ευρό.
7. Ο νόμος του Engel (Engel, 1848) προβλέπει ότι, όταν το εισόδημα αυξάνει η δαπάνη για φαγητό αυξάνει ομοίως αλλά λιγότερο αναλογικά με την αύξηση του εισοδήματος. Τα αποτελέσματα σας επιβεβαιώνουν τον νόμο αυτό;

ΣΕΜΙΝΑΡΙΟ 4. ΑΠΛΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΕΝΑΛΛΑΚΤΙΚΕΣ ΜΟΡΦΕΣ-Η ΣΗΜΑΣΙΑ ΤΗΣ ΓΡΑΜΜΙΚΟΤΗΤΑΣ

X	**1**	**1**	**2**	**2**	**3**	**3**	**4**	**4**	**5**	**5**
Y	**195**	**205**	**290**	**310**	**370**	**400**	**435**	**475**	**485**	**535**

ΣΕΜΙΝΑΡΙΟ 5. ΑΠΛΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΕΝΑΛΛΑΚΤΙΚΕΣ ΜΟΡΦΕΣ

ΣΕΜΙΝΑΡΙΟ 6-7 . ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΨΕΥΔΟΜΕΤΑΒΛΗΤΕΣ

***ΣΕΜΙΝΑΡΙΟ 8.* ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΠΟΛΛΥΣΥΓΓΡΑΜΜΙΚΟΤΗΤΑΣ**

***ΣΕΜΙΝΑΡΙΟ 8.* ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ**

ΣΕΜΙΝΑΡΙΟ 9-10. ΕΠΑΝΑΛΗΠΤΙΚΟ ΜΑΘΗΜΑ

1.