

Επαγωγική Στατιστική

Γραμμική παλινδρόμηση-Linear Regression

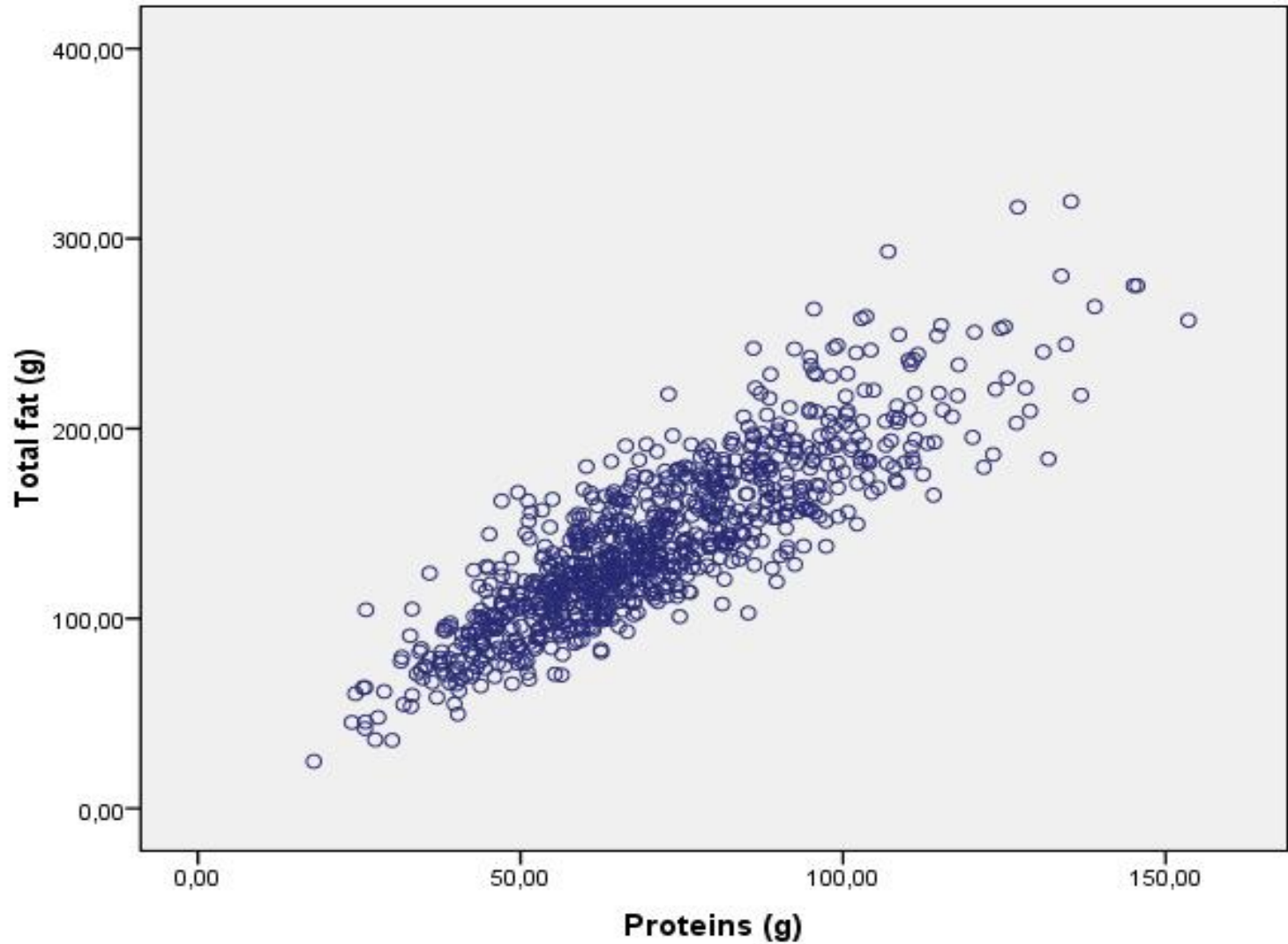
Χαράλαμπος Γναρδέλλης

Πρόγραμμα Μεταπτυχιακών Σπουδών
Βιώσιμη Αλιεία, Υδατοκαλλιέργεια

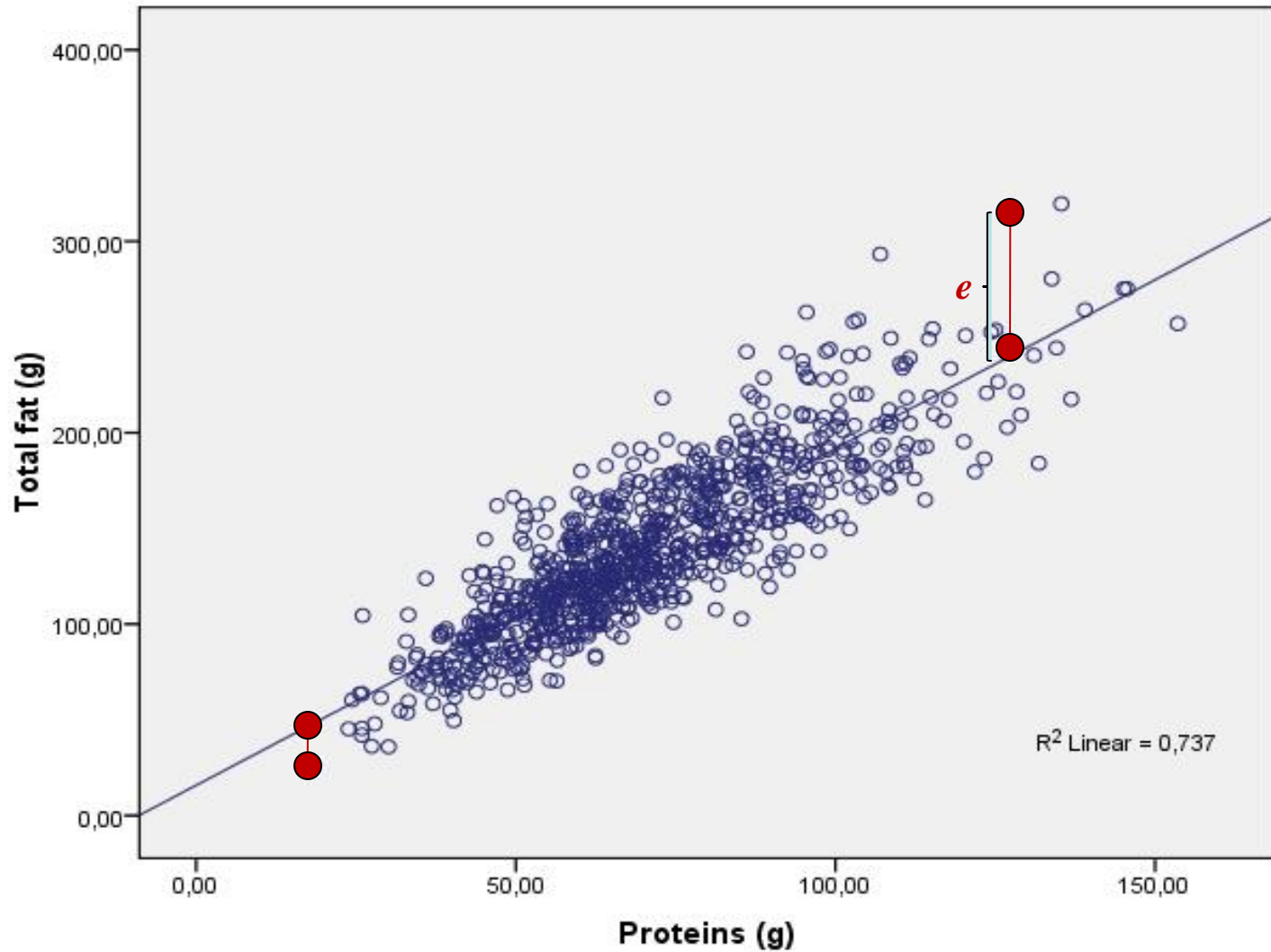
Γραμμική παλινδρόμηση

- Οι συντελεστές συσχέτισης προσδιορίζουν το βαθμό (ή αλλιώς την ένταση) της σχέσης που υπάρχει μεταξύ δύο ποσοτικών μεταβλητών υπό την προϋπόθεση ότι η σχέση αυτή είναι γραμμική.
- Αν η γραμμική σχέση των δύο μεταβλητών οριστεί με όρους εξάρτησης της μίας από την άλλη, δηλαδή, αν η μεταβολή των τιμών της μιας μεταβλητής θεωρηθεί ότι προκύπτει με γραμμικό τρόπο από τη μεταβολή των τιμών της άλλης, τότε η ανάλυση της σχέσης των δύο μεταβλητών πραγματοποιείται με τη βοήθεια ενός υποδείγματος απλής γραμμικής παλινδρόμησης.

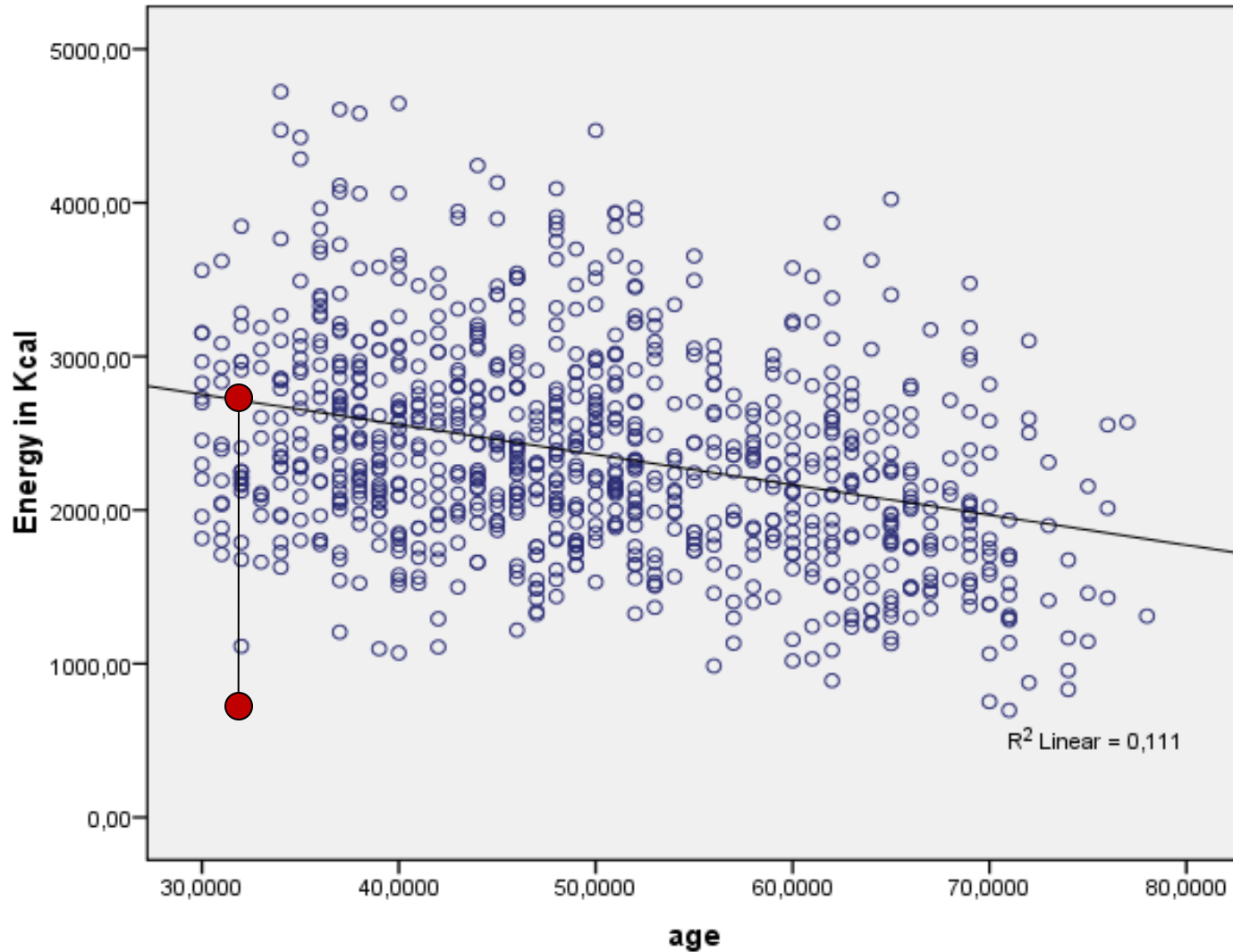
Διάγραμμα διασποράς δύο γραμμικά συσχετιζόμενων ποσοτικών μεταβλητών



Ευθεία παλινδρόμησης δύο γραμμικά συσχετιζόμενων μεταβλητών



Ευθεία παλινδρόμησης δύο γραμμικά συσχετιζόμενων μεταβλητών



- Η απλή γραμμική παλινδρόμηση ορίζει τη σχέση δύο ποσοτικών μεταβλητών X και Y , μέσω ενός υποδείγματος της μορφής

$$y = b_0 + b_1x + e,$$

όπου x και y είναι οι τιμές των δύο μεταβλητών για μια οποιαδήποτε παρατήρηση των δειγματικών δεδομένων.

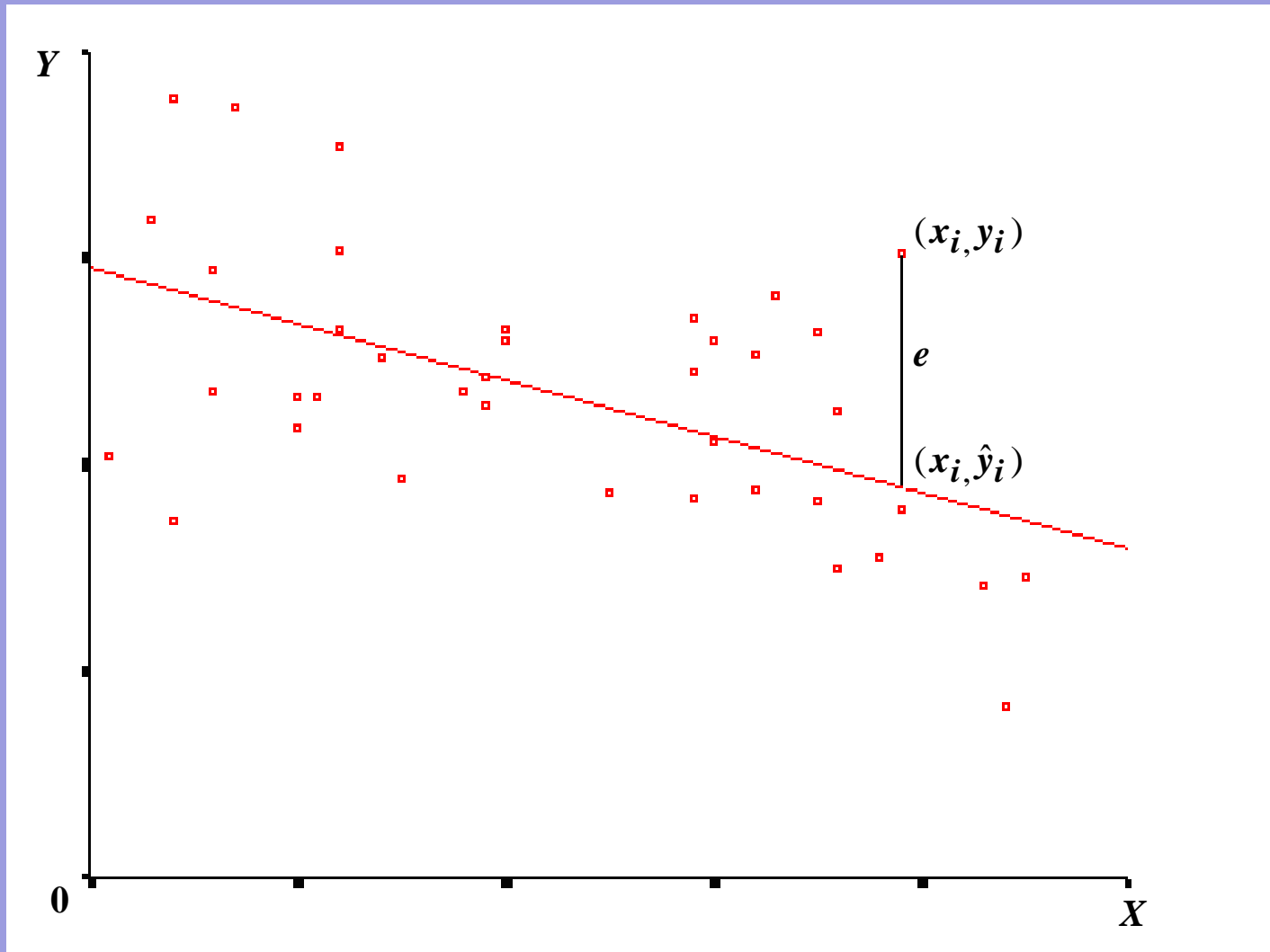
- Το πρώτο μέρος του υποδείγματος συνοψίζει τη γραμμική σχέση των δύο μεταβλητών, ενώ η ποσότητα e , η οποία ονομάζεται *υπόλοιπο* ή *σφάλμα*, αφορά την απόκλιση από τη γραμμικότητα, δηλαδή την απόκλιση των τιμών της Y από την ευθεία

$$b_0 + b_1x$$

- Η ευθεία $b_0 + b_1x$ ονομάζεται ευθεία της παλινδρόμησης

- Μέσω του υποδείγματος της απλής γραμμικής παλινδρόμησης οι τιμές της Y εκτιμώνται (ή προβλέπονται) από τις τιμές της X . Για το λόγο αυτό η Y ονομάζεται *εξαρτημένη μεταβλητή του υποδείγματος (dependent variable ή response variable)* ενώ η μεταβλητή X ονομάζεται *ανεξάρτητη (independent ή predictor variable)*.
- Το υπόδειγμα της απλής γραμμικής παλινδρόμησης είναι η απλούστερη περίπτωση ενός γενικότερου υποδείγματος που χρησιμοποιείται στη στατιστική συμπερασματολογία, σύμφωνα με το οποίο *η μεταβολή των τιμών μιας ποσοτικής μεταβλητής ερμηνεύεται γραμμικά από τη μεταβολή των τιμών ενός συνόλου k άλλων ποσοτικών μεταβλητών*.

Προσδιορισμός της ευθείας της παλινδρόμησης για τις μεταβλητές X και Y



Αν δηλαδή η σχέση των μεταβλητών X και Y είναι γραμμική τότε αυτή μπορεί να συνοψιστεί με τη βοήθεια μιας ευθείας γραμμής η οποία προσεγγίζει με βέλτιστο τρόπο τα σημεία του διαγράμματος. Κάθε σημείο επομένως του διαγράμματος με συντεταγμένες (x_i, y_i) , μπορεί να θεωρηθεί ότι βρίσκεται επάνω σε μια ευθεία γραμμή, η οποία ονομάζεται **ευθεία της παλινδρόμησης**, και η οποία προσεγγίζει με τον καλύτερο δυνατό τρόπο το σύνολο των σημείων. Αυτή η παραδοχή απλοποιεί τη μορφή του νέφους των σημείων, υποκαθιστώντας το με τα αντίστοιχα σημεία (x_i, \hat{y}_i) που προκύπτουν από τις κατακόρυφες προβολές των δειγματικών τιμών της Y στην ευθεία της παλινδρόμησης.

- Οι κατακόρυφες προβολές των δειγματικών τιμών της Y , ορίζουν τα υπόλοιπα $e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_i)$.
- Αν όλα τα υπόλοιπα είναι ίσα με 0, θα έχουμε την πλήρη προσαρμογή της ευθείας επί των σημείων του διαγράμματος. Η πλήρης προσαρμογή της ευθείας είναι απίθανο να προκύψει (εκτός και αν οι δύο μεταβλητές είναι απολύτως γραμμικά εξαρτημένες), μπορούν όμως να ελαχιστοποιηθούν τα υπόλοιπα. Η ελαχιστοποίηση των υπολοίπων ισοδυναμεί με την ελαχιστοποίηση της ποσότητας

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2,$$

η οποία ονομάζεται *άθροισμα τετραγώνων των υπολοίπων* (*residual sum of squares*) ή *άθροισμα τετραγώνων των σφαλμάτων* (*error sum of squares*).

- Η ευθεία δηλαδή της παλινδρόμησης κατασκευάζεται με την ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων, γι' αυτό το λόγο ονομάζεται και *ευθεία των ελαχίστων τετραγώνων*.
- Από την ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων προκύπτουν οι τιμές των συντελεστών του υποδείγματος, b_0 και b_1 .

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Ο συντελεστής b_1 (κλίση) της ευθείας της παλινδρόμησης, ορίζεται ως η εκτιμώμενη μεταβολή της εξαρτημένης μεταβλητής (της Y δηλαδή) για κάθε μία μονάδα αύξησης της ανεξάρτητης (της X). Ο ορισμός αυτός προκύπτει από την παρακάτω σχέση

$$y_{x+1} - y_x = [b_0 + b_1(x + 1)] - (b_0 + b_1x) = b_1$$

Αν το πρόσημο του συντελεστή b_1 είναι θετικό, τότε για κάθε μονάδα αύξησης της X η Y αυξάνεται κατά b_1 , ενώ αν το πρόσημο του συντελεστή είναι αρνητικό για κάθε μονάδα αύξησης της X η Y ελαττώνεται κατά b_1 .

Επιπλέον, ο σταθερός όρος b_0 του υποδείγματος ορίζεται ως η εκτιμώμενη τιμή της εξαρτημένης μεταβλητής Y για την τιμή 0 της μεταβλητής X ,

$$b_0 = b_0 + b_1 \cdot 0$$

- Το υπόδειγμα της γραμμικής παλινδρόμησης μπορεί αρχικά να χρησιμοποιηθεί σε περιγραφικό επίπεδο, συνοψίζοντας τη σχέση της εξαρτημένης μεταβλητής με την ανεξάρτητη υπό τη μορφή της γραμμικής εξίσωσης

$$y = b_0 + b_1x + e,$$

- ενώ στη συνέχεια μπορεί να χρησιμοποιηθεί και επαγωγικά, ελέγχοντας αν οι σχέσεις που προσδιορίστηκαν περιγραφικά (μέσω της δειγματικής εξίσωσης της παλινδρόμησης) ισχύουν και σε πληθυσμιακό επίπεδο.

Δηλαδή, ο προσδιορισμός των συντελεστών του υποδείγματος της παλινδρόμησης για τα δειγματικά δεδομένα

$$y = b_0 + b_1x + e,$$

αποτελεί το πρώτο βήμα για την εκτίμηση των συντελεστών του πληθυσμιακού υποδείγματος της παλινδρόμησης, το οποίο ορίζεται από την εξίσωση

$$y = \beta_0 + \beta_1x + \varepsilon.$$

Θεωρώντας δηλαδή ότι η σχέση των μεταβλητών X και Y είναι γραμμική και ορίζοντάς την για τα δειγματικά δεδομένα μέσω του δειγματικού υποδείγματος, εναπομένει να εκτιμηθούν ή να ελεγχθούν επαγωγικά οι συντελεστές του αντίστοιχου πληθυσμιακού υποδείγματος.

- Ο κυριότερος επαγωγικός έλεγχος που μπορεί να γίνει στις τιμές των συντελεστών του υποδείγματος αφορά τον συντελεστή b_1 . Βασιζόμενοι στην κλίση b_1 της ευθείας των ελαχίστων τετραγώνων, μπορούμε να ελέγξουμε την κλίση β_1 της πληθυσμιακής ευθείας της παλινδρόμησης ως προς την τιμή 0 .
- Αν η κλίση της πληθυσμιακής ευθείας της παλινδρόμησης είναι ίση με το 0 , τότε

$$y = \beta_0 + 0 \cdot x = \beta_0.$$

Επομένως, η τιμή της Y ανεξαρτήτως της τιμής της X στην οποία αντιστοιχεί, ισούται με το συντελεστή β_0 της παλινδρόμησης. Δηλαδή δεν υπάρχει γραμμική σχέση μεταξύ της X και της Y .

Ο έλεγχος δηλαδή γίνεται με μηδενική υπόθεση

$$H_0 : \beta_1 = 0$$

και εναλλακτική

$$H_A : \beta_1 \neq 0,$$

μέσω της ποσότητας

$$t = \frac{b_1}{se(b_1)},$$

η οποία, εφόσον ισχύει η μηδενική υπόθεση, ακολουθεί την κατανομή t με $n-2$ βαθμούς ελευθερίας.

Αντίστοιχος έλεγχος γίνεται και για το συντελεστή b_0 .

Με μηδενική υπόθεση την

$$H_0 : \beta_0 = 0,$$

και εναλλακτική την

$$H_A : \beta_0 \neq 0,$$

η ποσότητα

$$t = \frac{b_0}{se(b_0)},$$

ακολουθεί την κατανομή t με $n-2$ βαθμούς ελευθερίας

Τέλος, μπορούμε να υπολογίσουμε το 95% διάστημα εμπιστοσύνης για τον πληθυσμιακό συντελεστή β_1 της παλινδρόμησης χρησιμοποιώντας τη γνωστή μορφή των διαστημάτων εμπιστοσύνης

$$(b_1 - 1,96se(b_1) , b_1 + 1,96se(b_1)).$$

Αντίστοιχα διαστήματα εμπιστοσύνης μπορούν να κατασκευαστούν και για το συντελεστή β_0

$$(b_0 - 1,96se(b_0) , b_0 + 1,96se(b_0)).$$

Αξιολόγηση του υποδείγματος της παλινδρόμησης Ο συντελεστής προσδιορισμού

Μετά τον προσδιορισμό της ευθείας των ελαχίστων τετραγώνων, δια μέσου της εξίσωσης

$$\hat{y} = b_0 + b_1x$$

εναπομένει η αξιολόγηση της προσαρμογής της ευθείας αυτής επί των δειγματικών τιμών. Ένας τρόπος για να αξιολογήσουμε την προσαρμογή της ευθείας των ελαχίστων τετραγώνων είναι να υπολογίσουμε το συντελεστή προσδιορισμού (*coefficient of determination*).

Ο συντελεστής προσδιορισμού της δειγματικής ευθείας της παλινδρόμησης, συμβολιζόμενος με R^2 , ορίζεται ως το τετράγωνο του δειγματικού συντελεστή συσχέτισης, δηλαδή

$$R^2=r^2.$$

Επειδή ο δειγματικός συντελεστής συσχέτισης παίρνει τιμές στο διάστημα $[-1, 1]$, ο συντελεστής προσδιορισμού παίρνει τιμές στο διάστημα $[0,1]$. Όταν $R^2=1$, όλα τα σημεία που αναπαριστούν τις δειγματικές τιμές των X και Y βρίσκονται τοποθετημένα επί της ευθείας των ελαχίστων τετραγώνων. Όταν $R^2=0$, δεν υπάρχει γραμμική σχέση μεταξύ των δειγματικών τιμών των X και Y .

Επειδή γενικότερα ισχύει

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

όπου

$$\sum_{i=1}^n (y_i - \bar{y})^2$$

Συνολικό άθροισμα τετραγώνων

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Άθροισμα τετραγώνων επεξηγούμενο από την παλινδρόμηση

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Άθροισμα τετραγώνων των σφαλμάτων

Αποδεικνύεται ότι ο συντελεστής προσδιορισμού, μπορεί να ερμηνευθεί ως το ποσοστό της μεταβλητότητας των τιμών της Y που επεξηγείται από το υπόδειγμα της γραμμικής παλινδρόμησης.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Άθροισμα τετραγώνων επεξηγούμενο από την παλινδρόμηση

$$\sum_{i=1}^n (y_i - \bar{y})^2$$

Συνολικό άθροισμα τετραγώνων

Ο δειγματικός συντελεστής προσδιορισμού αξιολογείται και αυτός επαγωγικά, σε σχέση με τον πληθυσμιακό συντελεστή ρ^2 .

Ο έλεγχος γίνεται με μηδενική υπόθεση

$$H_0 : \rho^2 = 0,$$

και εναλλακτική

$$H_A : \rho^2 \neq 0,$$

μέσω της ποσότητας

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - 2)}.$$

Η οποία όταν ισχύει η μηδενική υπόθεση ακολουθεί την κατανομή F με 1 και $n-2$ βαθμούς ελευθερίας

$$\text{Ενέργεια} = b_0 + b_1(\text{ηλικία})$$

R^2 , Συντελεστής
προσδιορισμού

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,300 ^a	,090	,089	744,56403

a. Predictors: (Constant), Ηλικία

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7,751E7	1	7,751E7	139,821	,000 ^a
	Residual	7,833E8	1413	554375,595		
	Total	8,608E8	1414			

a. Predictors: (Constant), Ηλικία

b. Dependent Variable: Ενεργειακή Πρόσληψη σε Kcal

Επαγωγική αξιολόγηση του R^2

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		
	B	Std. Error	Beta			Lower Bound	Upper Bound	
1	(Constant)	3499,216	90,219			3322,239	3676,193	
	Ηλικία	-19,687	1,665	-,300	-11,825	,000	-22,954	-16,421

a. Dependent Variable: Ενεργειακή Πρόσληψη σε Kcal

Για κάθε έτος ηλικίας ελαττώνεται η πρόσληψη κατά 20 περίπου (19,7) kcal

Επαγωγική αξιολόγηση του συντελεστή b_1

Διάστημα εμπιστοσύνης του συντελεστή b_1

Υπόδειγμα πολλαπλής γραμμικής παλινδρόμησης

Η ερμηνευτική αξία του υποδείγματος της απλής γραμμικής παλινδρόμησης, μπορεί να διευρυνθεί περαιτέρω, αν σε αυτό συμπεριληφθούν περισσότερες από μία ανεξάρτητες μεταβλητές, οι οποίες επιδρούν στις τιμές της εξαρτημένης μεταβλητής. Η χρήση και η ερμηνεία ενός υποδείγματος γραμμικής παλινδρόμησης με πολλές ανεξάρτητες μεταβλητές –γνωστό ως υπόδειγμα πολλαπλής γραμμικής παλινδρόμησης (*multiple linear regression model*)– είναι φυσική προέκταση του υποδείγματος της απλής γραμμικής παλινδρόμησης.

Πολλαπλή γραμμική παλινδρόμηση

- Στο υπόδειγμα της πολλαπλής γραμμικής παλινδρόμησης υποθέτουμε ότι η σχέση μιας ποσοτικής μεταβλητής Y και μιας σειράς k ανεξάρτητων μεταβλητών X_1, X_2, \dots, X_k είναι γραμμική. Η εκτίμηση των τιμών της Y για τα δειγματικά δεδομένα γίνεται από την εξίσωση

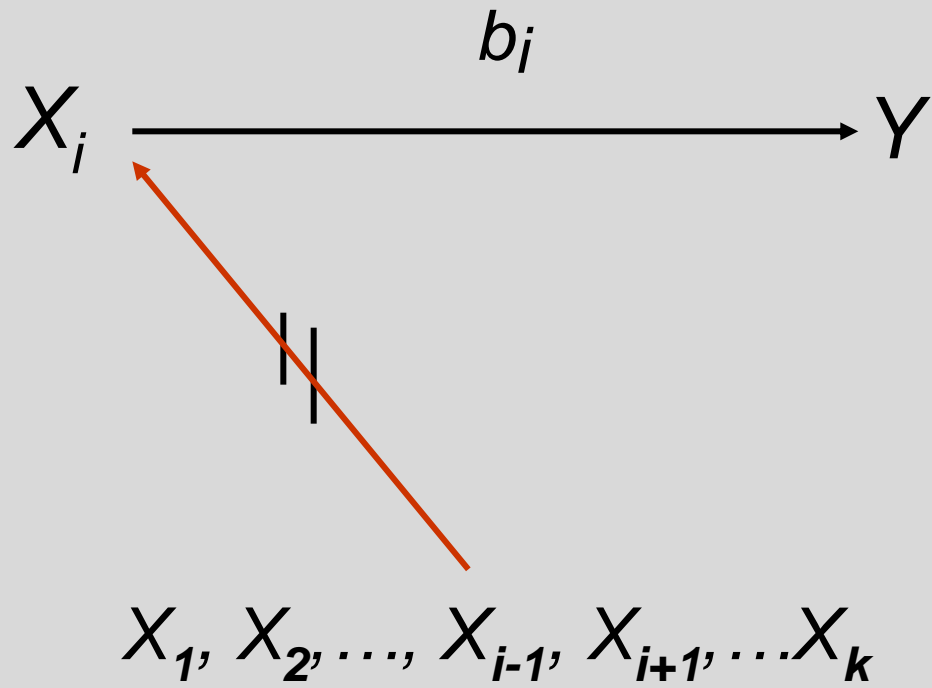
$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + e$$

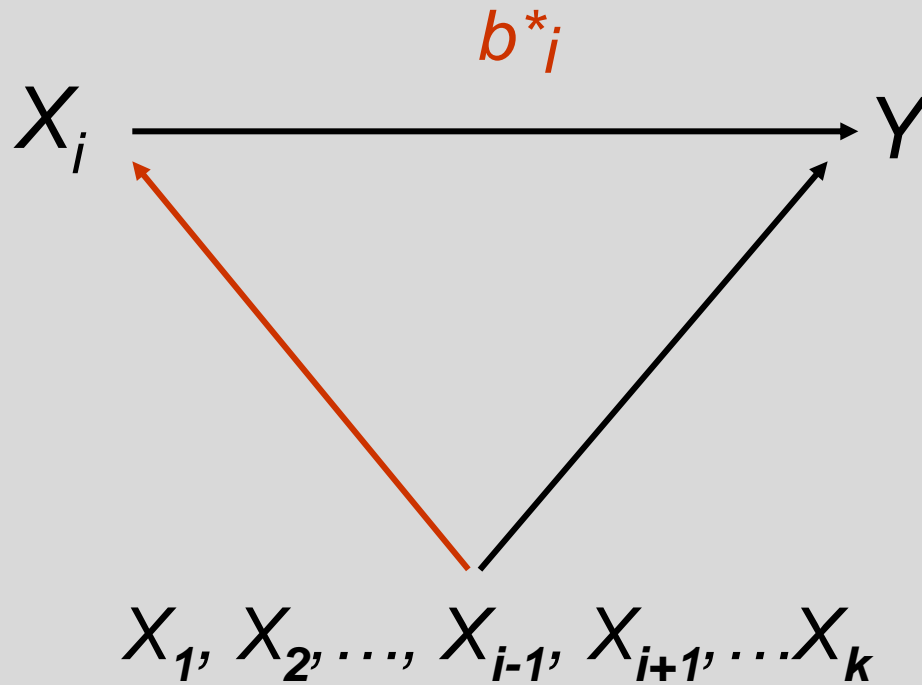
- Το πρώτο μέρος του υποδείγματος συνοψίζει τη γραμμική σχέση της Y με τις μεταβλητές X_1, X_2, \dots, X_k ενώ το σφάλμα e αφορά την απόκλιση από τη γραμμικότητα.

Ερμηνεία των συντελεστών του υποδείγματος της γραμμικής παλινδρόμησης

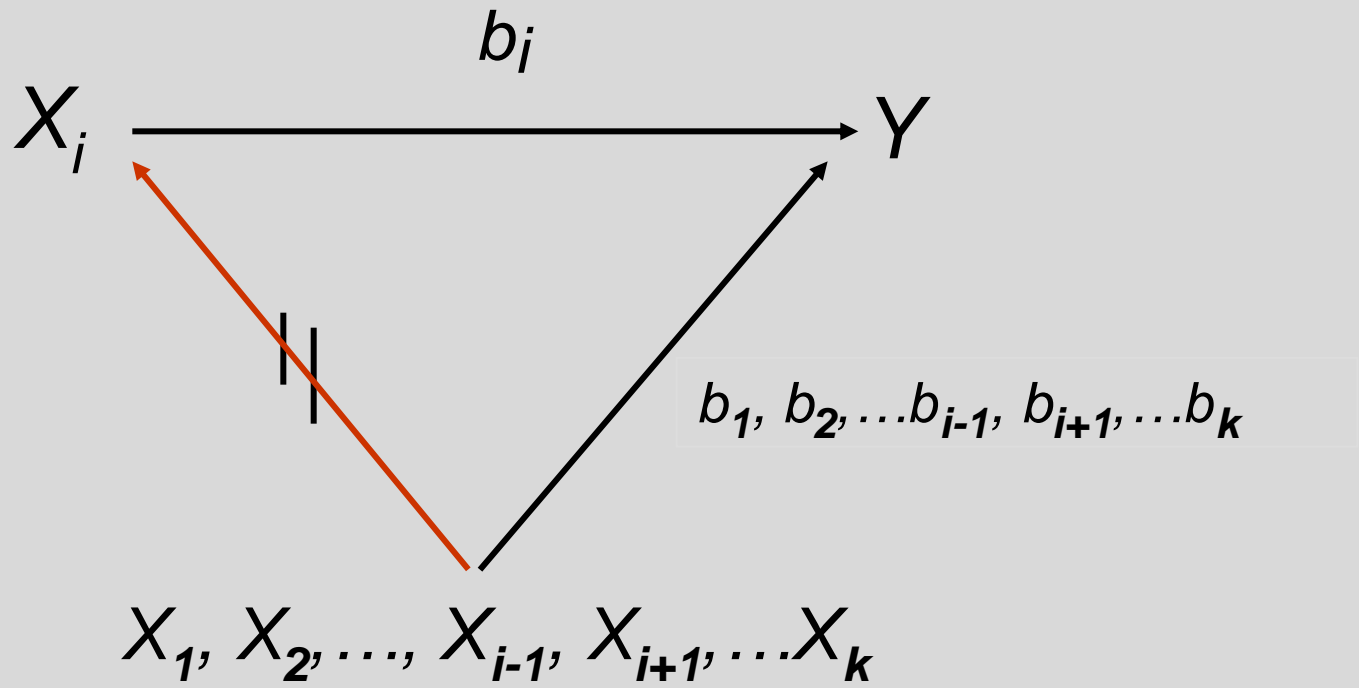
$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + e$$

- Ο σταθερός όρος b_0 εκτιμά την τιμή της εξαρτημένης μεταβλητής Y όταν όλες οι ανεξάρτητες μεταβλητές παίρνουν την τιμή 0.
- Ένας οποιοσδήποτε συντελεστής $b_i, i = 1, 2, \dots, k$ εκτιμά τη μεταβολή της εξαρτημένης μεταβλητής Y για μια μονάδα αύξησης της ανεξάρτητης μεταβλητής X_i όταν οι τιμές των άλλων ανεξάρτητων μεταβλητών παραμένουν σταθερές ή (ισοδύναμα), όταν οι γραμμικές επιδράσεις των υπόλοιπων μεταβλητών στη X_i έχουν απομακρυνθεί από το υπόδειγμα.





Συγχυτικοί παράγοντες (confounders)



- Στην περίπτωση κατά την οποία μια ανεξάρτητη μεταβλητή του υποδείγματος είναι *δίτιμη* (*binary*), εκφρασμένη αριθμητικά με τις τιμές **0** και **1** (π.χ. **0**=γυναίκες, **1**=άνδρες), η ερμηνεία του συντελεστή της παραμένει ανάλογη με αυτή των ποσοτικών μεταβλητών.

Δηλαδή, η ποσότητα b_i

που αφορά τη δίτιμη μεταβλητή X_i εξακολουθεί

να ορίζει τη μεταβολή της Y για μια μονάδα αύξησης της X_i

(δηλαδή πόσο διαφέρει η τιμή της Y στις δύο κατηγορίες της

X_i) όταν οι τιμές των άλλων ανεξάρτητων μεταβλητών παραμένουν σταθερές.

- **Αν η τιμή του b_i είναι θετική, $b_i > 0$,**
τότε ο συντελεστής b_i ορίζει πόσο μεγαλύτερη είναι η τιμή της Y στην κατηγορία με κωδικό **1** έναντι της κατηγορίας με κωδικό **0** (π.χ. πόσο μεγαλύτερη είναι η τιμή της Y στους άνδρες απ' ότι στις γυναίκες).
- **Αν η τιμή του b_i είναι αρνητική, $b_i < 0$,**
τότε ο συντελεστής b_i ορίζει πόσο μικρότερη είναι η τιμή της Y στην κατηγορία με κωδικό **1** έναντι της κατηγορίας με κωδικό **0** (π.χ. πόσο μικρότερη είναι η τιμή της Y στους άνδρες απ' ότι στις γυναίκες).

- Όταν χρησιμοποιείται ως ανεξάρτητη μεταβλητή ενός υποδείγματος μια κατηγορική μεταβλητή με περισσότερες των δύο κατηγοριών, θα πρέπει να κωδικοποιηθεί με τρόπο αντίστοιχο των δίτιμων μεταβλητών.
- Αν η μεταβλητή περιλαμβάνει k κατηγορίες, πρέπει να δημιουργηθούν $k-1$ νέες ψευδομεταβλητές (*dummy variables*) με τιμές 0 και 1.
- Για το σκοπό αυτό, μία από τις κατηγορίες της μεταβλητής επιλέγεται ως κατηγορία αναφοράς, ενώ δημιουργούνται $k-1$ ψευδομεταβλητές που αντιπροσωπεύουν τις υπόλοιπες κατηγορίες. Η τιμή κάθε μιας από τις οριζόμενες ψευδομεταβλητές είναι 1, αν μία παρατήρηση ανήκει στην αντίστοιχη κατηγορία, και 0, αν δεν ανήκει.

- Αν για παράδειγμα χρησιμοποιηθεί μια μεταβλητή που ορίζει το επίπεδο εκπαίδευσης, με κατηγορίες

αναλφάβητος/η,

απόφοιτος/η στοιχιώδους εκπαίδευσης,

απόφοιτος/η μέσης εκπαίδευσης

απόφοιτος/η τριτοβάθμιας εκπαίδευσης,

πρέπει αρχικά να οριστεί μία κατηγορίας αναφοράς, π.χ. η κατηγορία *αναλφάβητος/η*.

Στη συνέχεια να δημιουργηθούν τρεις ψευδομεταβλητές που αντιπροσωπεύουν τους απόφοιτους *στοιχιώδους εκπαίδευσης* (X_1), *μέσης εκπαίδευσης* (X_2) και *τριτοβάθμιας εκπαίδευσης* (X_3). Κάθε μία από αυτές τις ψευδομεταβλητές παίρνει την τιμή 1, αν κάποιος είναι του ιδίου επιπέδου εκπαίδευσης με αυτό που ορίζει η ψευδομεταβλητή, και 0, αν δεν είναι.

Έτσι για έναν απόφοιτο στοιχειώδους εκπαίδευσης οι δημιουργούμενες ψευδομεταβλητές θα πάρουν τις τιμές

$$X_1=1, X_2=0, X_3=0$$

για έναν απόφοιτο μέσης εκπαίδευσης, τις τιμές

$$X_1=0, X_2=1, X_3=0$$

και για έναν απόφοιτο τριτοβάθμιας εκπαίδευσης, τις τιμές

$$X_1=0, X_2=0, X_3=1.$$

Εφόσον οι $k-1$ ψευδομεταβλητές εισαχθούν σε ένα υπόδειγμα παλινδρόμησης, ο συντελεστής που θα προκύψει για κάθε μία εκφράζει τη μέση διαφορά της εξαρτημένης μεταβλητής μεταξύ της κατηγορίας που αντιπροσωπεύει η ψευδομεταβλητή και της κατηγορίας αναφοράς.

- Η κατηγορία αναφοράς δεν μπορεί να μπει στο υπόδειγμα, διότι σε αυτή την περίπτωση οι ψευδομεταβλητές θα ήταν μεταξύ τους γραμμικά εξαρτημένες, γεγονός που θα δημιουργούσε πρόβλημα συγγραμμικότητας στο υπόδειγμα.
- Μπορεί όμως με εναλλαγή της κατηγορίας αναφοράς να συγκριθούν μεταξύ τους όλες οι κατηγορίες της μεταβλητής.

Χρήση του υποδείγματος της γραμμικής παλινδρόμησης

- Το υπόδειγμα της γραμμικής παλινδρόμησης

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

μπορεί αρχικά να αξιολογηθεί ως προς την προσαρμογή του στα δειγματικά δεδομένα (μέσω του συντελεστή πολλαπλού προσδιορισμού R^2 , ο οποίος εκφράζει το ποσοστό της μεταβλητότητας της Y που ερμηνεύεται από το υπόδειγμα),

- στη συνέχεια όμως μπορεί να χρησιμοποιηθεί και επαγωγικά, ελέγχοντας αν οι σχέσεις που προσδιορίστηκαν περιγραφικά (μέσω της δειγματικής εξίσωσης της παλινδρόμησης) ισχύουν και σε πληθυσμιακό επίπεδο.

Επαγωγική αξιολόγηση του υποδείγματος της γραμμικής παλινδρόμησης

- Η επαγωγική αξιολόγηση του υποδείγματος της γραμμικής παλινδρόμησης αφορά:
 - ✓ τον έλεγχο της γραμμικής σχέσης μεταξύ της εξαρτημένης μεταβλητής και των ανεξαρτήτων (μέσω της αξιολόγησης του συντελεστή πολλαπλού προσδιορισμού R^2)
 - ✓ τον έλεγχο της επίδρασης κάθε ανεξάρτητης μεταβλητής στην εξαρτημένη (μέσω της αξιολόγησης των συντελεστών b_i)
 - ✓ την κατασκευή διαστημάτων εμπιστοσύνης για τους αντίστοιχους πληθυσμιακούς συντελεστές β_i της παλινδρόμησης.

Εκτίμηση του χρόνου βραδινού ύπνου αναφορικά με το φύλο, την ηλικία, τις ώρες ημερήσιας εργασίας και την περιοχή διαμονής

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,481	,231	,229	1,763

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	962,400	4	240,600	77,387	,000
Residual	4399,324	1415	3,109		
Total	5361,724	1419			

Αξιολόγηση του υποδείγματος ως προς την προσαρμογή του στα δειγματικά δεδομένα

Επαγωγική αξιολόγηση του υποδείγματος ως προς την προσαρμογή του στα πληθυσμιακά δεδομένα

$$\Omega\rho\rho\rho\rho\rho\rho = 9,150 + 0,682(\rho\rho\rho\rho) + 0,004(\eta\rho\rho\rho\rho) - 0,224(\epsilon\rho\rho\rho\rho) - 0,111(\rho\rho\rho\rho\rho)$$

Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig	95% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	9,150	,284		32,217	,000	8,593	9,707
Φύλο (άνδρες vs γυναίκες)	,682	,105	,173	6,507	,000	,476	,887
Ηλικία σε έτη	,004	,005	,022	,784	,433	-,005	,013
Ώρες ημερήσιας εργασίας	-,224	,016	-,434	-14,362	,000	-,255	-,194
Περιοχή διαμονής (Αθήνα vs λοιπή χώρα)	-,111	,099	-,028	-1,131	,258	-,305	,082

Συντελεστές
του
υποδείγματος

Επαγωγική
αξιολόγηση των
συντελεστών

Διαστήματα
εμπιστοσύνης των
συντελεστών