

Πολυμεταβλητό: έχω πολλές ανεξάρτητες μεταβλητές

(1)

Πολυμεταβλητό γραμμικό υπόδειγμα

Πολυμεταβλητό γραμμικό υπόδειγμα

- Με το πολυμεταβλητό γραμμικό υπόδειγμα παλινδρόμησης εκτιμούμε μια γραμμική (ως προς τις παραμέτρους) σχέση ανάμεσα σε δύο ή περισσότερες μεταβλητές X_k , $k=1,2,\dots$ («ανεξάρτητες μεταβλητές») και μια ποσοτική μεταβλητή Y («εξαρτημένη μεταβλητή»).
- Οι ανεξάρτητες μεταβλητές X_i συνήθως είναι υπό τον έλεγχο του ερευνητή.
- Στόχος είναι να μελετήσουμε πώς οι (συνήθως ελεγχόμενες) μεταβολές των τιμών των X_i επιδρούν γραμμικά στις τιμές που παίρνει η Y , δηλ. πώς «η Y εξαρτάται γραμμικά από τις X_i ».
- Με ένα εκτιμημένο γραμμικό μοντέλο για την σχέση των X_i και Y μπορούμε να:
 - εξετάσουμε στατιστικά αν υπάρχει γραμμική σχέση ανάμεσα στις X_i και Y
 - διεξάγουμε προβλέψεις των τιμών της Y δεδομένων των τιμών των X_i

3

εξαρτημένη μεταβλητή
η X Y = απόδοση χωραφιάς

ανεξάρτητες X_1 = ποσότητα λιπάσματος
 X_2 = 217 αμm κόνιο
 X_3 = μέση θερμοκρασία
 X_4 = κ + 1 ...

Το Κλασικό Υπόδειγμα

- Πολυμεταβλητό Γραμμικό Υπόδειγμα στον Πληθυσμό:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

- β_0 : συντελεστής τομής (σταθερός όρος)
- β_j , $j=1,\dots,k$: συντελεστές κλίσεως
- ϵ : διαταρακτικός όρος / σφάλμα

$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ είναι το «συστηματικό» (ή προσδιοριστικό) μέρος του μοντέλου

Το β_i καθορίζει την επίδραση της ανεξάρτητης μεταβλητής X_i στην εξαρτημένη μεταβλητή Y .

Είναι επέκταση του απλού γραμμικού υποδείγματος.

4

είναι μία γενίκευση του
απλού γραμμικού
υποδείγματος

απόδοση $Y = 25 + (30)X$ ποσότητα λιπάσματος

αν αυξήσω το X κατά μία μονάδα, τότε το Y αυξάνεται κατά 30 μονάδες

Βασικές Υποθέσεις

- Οι βασικές υποθέσεις του υποδείγματος:

1. Η αληθής σχέση στον πληθυσμό είναι γραμμική:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + u_i$$

2. Ο διαταρακτικός όρος έχει μέση τιμή 0 και σταθερή διακύμανση (ομοσκεδαστικότητα) για όλες τις παρατηρήσεις:

$$u_i \sim (0, \sigma_u^2), \quad \forall i = 1, 2, \dots, n$$

3. Οι διαταρακτικοί όροι διαφορετικών παρατηρήσεων είναι ανεξάρτητοι και συνεπώς ασυσχέτιστοι (απουσία αυτοσυσχέτισης):

$$\sigma_{u_i u_j} = 0, \quad \forall i \neq j \quad \text{ή} \quad E(u_i u_j) = 0, \quad \forall i \neq j$$

γραμμικό μοντέλο

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

εξ. γραμμικό μοντέλο

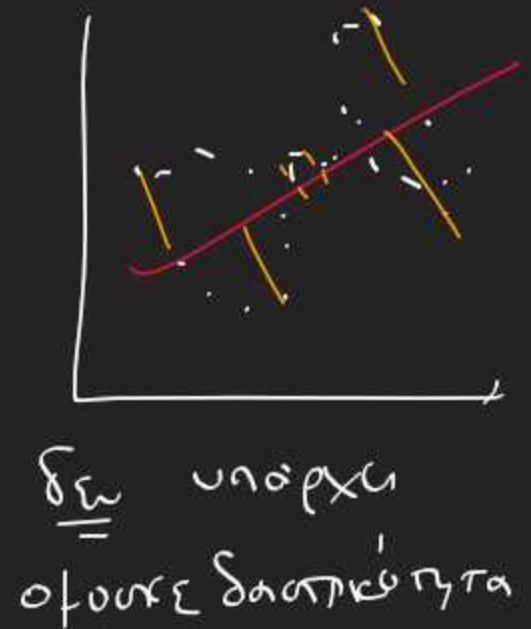
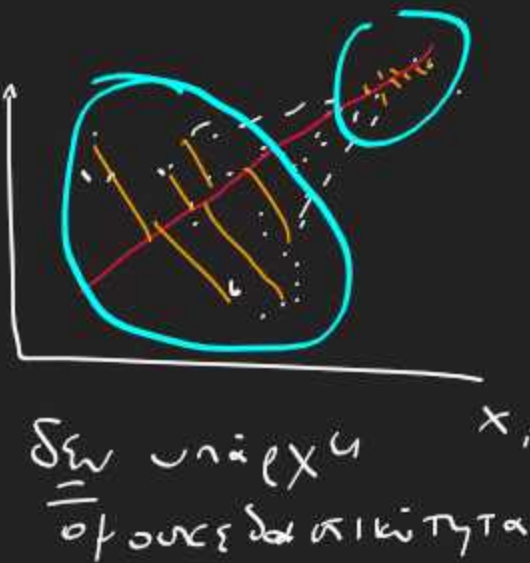
$$Y = \beta_0 + \beta_1 X_1$$

$$Y = \beta_0 + \beta_1 \sqrt{X_1}$$

$$Y = \beta_0 + \frac{1}{X_1}$$



Εχόλιν για υπόθεση 1:



Εχόλιν για υπόθεση 3:

ή Δε χρησιμοποιώ γραμμικό μοντέλο για το χρόνο.

Βασικές Υποθέσεις

- Οι βασικές υποθέσεις του υποδείγματος (συνέχεια):

- Οι ερμηνευτικές μεταβλητές X είναι μη-στοχαστικές και έχουν πεπερασμένη μεταβλητότητα. $\sigma^2 < \infty$

502.

- Επιπλέον, δεν υπάρχει ακριβής γραμμική σχέση ανάμεσα σε δύο ή περισσότερες ερμηνευτικές μεταβλητές (απουσία απόλυτης πολυσυγγραμμικότητας).

- Οι υποθέσεις αυτές είναι απαραίτητες για την ισχύ του θεωρήματος **Gauss-Markov**: οι εκτιμητές ελαχίστων τετραγώνων των συντελεστών του πολυμεταβλητού γραμμικού υποδείγματος είναι **άριστοι γραμμικοί αμερόληπτοι εκτιμητές (BLUE)**.

7

3

Λευ διαλέγω ανεξάρτητες
εξαρτητές X_1, X_2, X_3, \dots
που να σχετίζονται μεταξύ τους

πχ Y - κώδικας χωραφισ

X_1 - ποσότητα λιπάσματος

X_2 - έτος θερμοκρασία

X_3 - έτος θερμοκρασία

Συνήθως στη στατική ανάλυση
πρώτα βρίσκω πινάκα
σχετίσεων και μετά
βρίσκω το ταίρι.

πχ η ένταξη συσχετίζεται (ρεαγισό)

$$\begin{matrix} X_1 & X_2 & X_3 & X_4 \\ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{matrix} & \begin{bmatrix} 1 & 0.2 & -0.6 & 0.05 \\ 0.2 & 1 & 0.3 & -0.1 \\ -0.6 & 0.3 & 1 & -0.8 \\ 0.05 & -0.1 & -0.8 & 1 \end{bmatrix} \end{matrix}$$

Για το μοντέλο θα διαλέγω μία από δύο μεταβλητές X_3, X_4 κρατώντας X_1

Βασικές Υποθέσεις

- Υπόθεση κανονικότητας του διαταρακτικού όρου:

$$\epsilon_i \sim N(0, \sigma^2), \quad \forall i = 1, 2, \dots, n$$

- γίνεται για να είναι δυνατός ο έλεγχος υποθέσεων και η κατασκευή διαστημάτων εμπιστοσύνης για τους συντελεστές με τις συνήθεις μεθόδους.

- Για να μπορεί να εκτιμηθεί το υπόδειγμα, θα πρέπει:

$$k < n$$

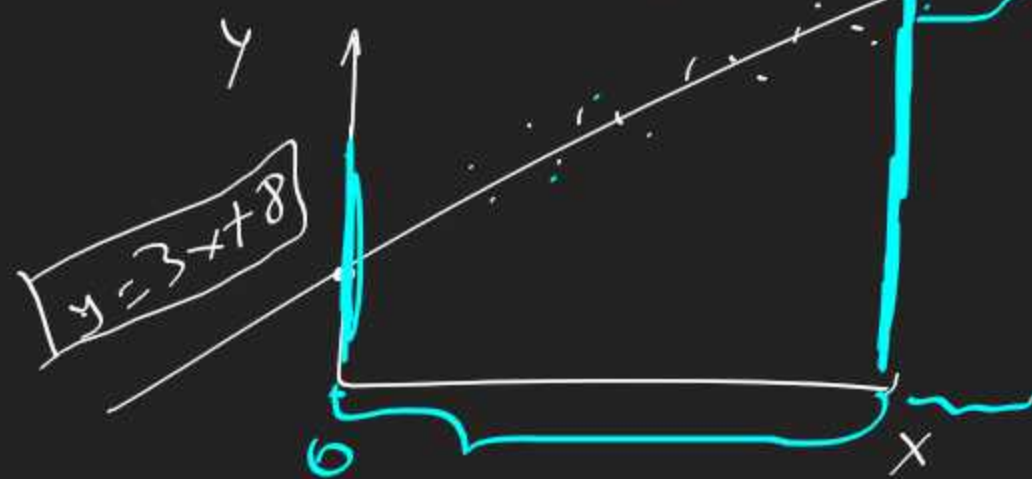
- Αναμενόμενη τιμή και διακύμανση των Y_i :

Από τις βασικές υποθέσεις:

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$$

$$\sigma_{Y_i}^2 = \sigma^2$$

10



πχ Y - κώδικας χωραφισ

X_1 - ποσότητα λιπάσματος

(όσο περισσότερο λιπάσμα
τόσο το καλύτερο)

Ερμηνεία των Συντελεστών Κλίσεως

- Ερμηνεία συντελεστών κλίσεως: *κερνή παραγωγή*

$$\beta_j = \frac{\partial E(Y)}{\partial X_j}, \quad j = 2, \dots, k$$

- η μεταβολή της αναμενόμενης τιμής της Y όταν η X_j μεταβάλλεται κατά μία μονάδα, για σταθερές τιμές των υπόλοιπων ερμηνευτικών μεταβλητών.

- ονομάζονται και μερικοί συντελεστές παλινδρόμησης

$$\text{πχ } Y = 300 + 20X_1 + 30X_2 - 15X_3 + \epsilon_i$$

β_1 : Αν αυξήσω την τιμή του X_1 κατά μία μονάδα
και οι υπόλοιπες μεταβλητές (X_2, X_3) μείνουν σταθερές,
τότε η μεταβλητή Y θα αυξηθεί κατά $\beta_1 (= 20)$
μονάδες

nx Αλγόριθμο Γραμμική Πρόσδυξη

(4)

$$\boxed{Y = 200 + 10 \cdot X} + \varepsilon_i$$

Y αμοδογ

X ποσότητα

↓ κόστος

Πρόβλεψη Αν $x = 20$ kg. Ποση αμοδογ θα έχω?

$$Y = 200 + 10 \cdot 20$$

$$= 200 + 200$$

$$= \underline{\underline{400 \text{ kg}}}$$

ΕΚΤΙΜΗΣΗ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ

Μορφή πίνακων

• Εκτιμώμενο υπόδειγμα:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik}$$

Κατάλοιπα: $e_i = \hat{u}_i = Y_i - \hat{Y}_i$

• Εκτιμητές των συντελεστών

$$X' = X^T$$

$$\hat{\beta} = (X'X)^{-1} X'Y$$

όπου

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$$

$$X = \begin{pmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \dots & X_{kn} \end{pmatrix}_{(n \times (k+1))}$$

$$Y = (Y_1, Y_2, \dots, Y_n)$$

Αλγ. Βρεϊντγου: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

↑
X' X

$$\beta_1 = \frac{\sum x_{1i} \cdot y_i \cdot \sum x_{2i}^2 - \sum x_{2i} y_i \cdot \sum x_{1i} x_{2i}}{\sum x_{1i}^2 \cdot \sum x_{2i}^2 - (\sum x_{1i} x_{2i})^2}$$

$$\beta_2 = \frac{\sum x_{2i} \cdot y_i \cdot \sum x_{1i}^2 - \sum x_{1i} y_i \cdot \sum x_{1i} x_{2i}}{\sum x_{1i}^2 \cdot \sum x_{2i}^2 - (\sum x_{1i} x_{2i})^2}$$

Y	X ₁	X ₂