

Γραμμικά Μοντέλα

Έλεγχοι Δομικών Αλλαγών

Διδάσκουσα: Λουκία Μελιγκοτσίδου

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Τμήμα Μαθηματικών

Έλεγχος δομικών αλλαγών με τη χρήση ψευδομεταβλητών

[Έλεγχος δομικής αλλαγής]: Έλεγχος αν υπάρχει αλλαγή στους συντελεστές της παλινδρόμησης.

- Ψευδομεταβλητή (dummy variable): δυαδική μεταβλητή που παίρνει τιμές 1 και 0 ανάλογα με την ύπαρξη ή όχι κάποιου χαρακτηριστικού.
- Για χρονοσειρές έχουμε δομικές αλλαγές που αντιπροσωπεύουν κάποια κρίση, καταστροφή, πόλεμο, αλλαγή δημοσιονομικής πολιτικής.
- Για διαστρωματικά στοιχεία έχουμε δομικές αλλαγές που αντιστοιχούν σε ποιοτικά χαρακτηριστικά που επηρεάζουν τους συντελεστές του γραμμικού μοντέλου (π.χ νοικοκυριό σε αστική ή αγροτική περιοχή).

[Ψευδομεταβλητές χρονικής επίδρασης]: Έστω το απλό γραμμικό υπόδειγμα κατανάλωσης-εισοδήματος της οικονομίας για τα έτη 1920 ως 2020

$$C_t = \beta_1 + \beta_2 Y_t + \epsilon_t, \quad t = 1920, \dots, 2020.$$

Ενδιαφερόμαστε να εξετάσουμε την υπόθεση ότι η ύπαρξη πολέμων επηρεάζει σημαντικά την καταναλωτική συμπεριφορά. Έστω d_t δίτιμη μεταβλητή (ψευδομεταβλητή) με $d_t = 1$, αν υπάρχει πόλεμος και $d_t = 0$, αν δεν υπάρχει. Στο παράδειγμά μας, δηλαδή, $d_t = 1$, $t = 1940, \dots, 1945$ (δεύτερος παγκόσμιος πόλεμος).

Θεωρούμε τα παρακάτω γραμμικά μοντέλα χρονοσειρών.

- M1: $C_t = \beta_1 + \delta_1 d_t + \beta_2 Y_t + \epsilon_t$

Το υπόδειγμα M1 θεωρεί ότι ο πόλεμος προκαλεί μια παράλληλη μετατόπιση του επιπέδου της κατανάλωσης κατά σταθερά δ_1 , δηλαδή αλλαγή της σταθεράς από β_1 σε $\beta_1 + \delta_1$ στην περίοδο πολέμου. Από την οικονομική θεωρία περιμένουμε ότι η μετατόπιση θα είναι προς τα κάτω ($\delta_1 < 0$).

- M2: $C_t = \beta_1 + \beta_2 Y_t + \delta_2 d_t Y_t + \epsilon_t$

Το υπόδειγμα M2 θεωρεί ότι ο πόλεμος αλλάζει την οριακή ροπή του εισοδήματος για κατανάλωση, δηλαδή αλλαγή του συντελεστή κλίσης από β_2 σε $\beta_2 + \delta_2$ στην περίοδο πολέμου. Από την οικονομική θεωρία περιμένουμε ότι η μετατόπιση θα είναι προς τα κάτω ($\delta_2 < 0$).

- M3: $C_t = \beta_1 + \delta_1 d_t + \beta_2 Y_t + \delta_2 d_t Y_t + \epsilon_t$

Το υπόδειγμα M2 θεωρεί ότι ο πόλεμος προκαλεί και παράλληλη μετατόπιση του επιπέδου κατανάλωσης και αλλαγή στην οριακή ροπή του εισοδήματος για κατανάλωση, δηλαδή και αλλαγή της σταθεράς από β_1 σε $\beta_1 + \delta_1$ και αλλαγή του συντελεστή κλίσης από β_2 σε $\beta_2 + \delta_2$ στην περίοδο πολέμου. Από την οικονομική θεωρία περιμένουμε ότι η διπλή μετατόπιση θα είναι προς τα κάτω.

Αν ισχύουν οι κλασικές υποθέσεις, τα υποδείγματα M1-M3 μπορούν να εκτιμηθούν με μέθοδο ελαχίστων τετραγώνων. Οι έλεγχοι δομικών αλλαγών μπορούν να γίνουν ελέγχοντας τη στατιστική σημαντικότητα των αντίστοιχων συντελεστών.

[Γενίκευση στην πολλαπλή παλινδρόμηση]: Το γενικό γραμμικό μοντέλο γράφεται, σε μορφή πινάκων

$$Y = X\beta + \epsilon$$

Αν $D = \text{diag}(d_1, d_2, \dots, d_T)$ διαγώνιος πίνακας με τις τιμές της ψευδομεταβλητής, τότε για να έχουμε επίδραση της ψευδομεταβλητής στη σταθερά και σε όλους τους συντελεστές κλίσης της πολλαπλής παλινδρόμησης πρέπει να φτιάξουμε έναν νέο πίνακα σχεδιασμού X^* ως εξής $X^* = [X \quad DX]$. Επομένως το μοντέλο με δομική αλλαγή σε όλες τις παραμέτρους της παλινδρόμησης γράφεται ως

$$Y = X^*\beta^* + \epsilon,$$

$\beta^* = (\beta', \delta)'$. Έλεγχοι στατιστικής σημαντικότητας στους συντελεστές $\delta_1, \delta_2, \dots$ επιβεβαιώνουν ή απορρίπτουν την υπόθεση της δομικής αλλαγής στη σταθερά ή τον αντίστοιχο συντελεστή κλίσης.

[Ψευδομεταβλητές επιδράσεων παραγόντων]: Έστω το απλό γραμμικό υπόδειγμα κατανάλωσης-εισοδήματος για διαστρωματικά στοιχεία (νοικοκυριά)

$$C_i = \beta_1 + \beta_2 Y_i + \epsilon_i, \quad i = 1, \dots, N.$$

Για να μελετηθεί η επίδραση διαφόρων παραγόντων όπως το επίπεδο εκπαίδευσης ή η περιοχή διαμονής εισάγονται στο υπόδειγμα ποιοτικές μεταβλητές (ψευδομεταβλητές επιδράσεων παραγόντων).

Έστω d_{1i} δίτιμη μεταβλητή (ψευδομεταβλητή) με $d_{1i} = 1$, αν ένα μέλος του νοικοκυριού έχει πτυχίο πανεπιστημίου και $d_{1i} = 0$, διαφορετικά. Επίσης, d_{2i} ψευδομεταβλητή με $d_{2i} = 1$, αν ένα μέλος του νοικοκυριού έχει επιχειρηματική δραστηριότητα και $d_{2i} = 0$, διαφορετικά, και d_{3i} ψευδομεταβλητή με $d_{3i} = 1$, αν το νοικοκυριό διαμένει σε αστική περιοχή και $d_{3i} = 0$, διαφορετικά.

Στο μοντέλο

$$C_i = \beta_1 + \delta_1 d_{1i} + \delta_2 d_{2i} + \delta_3 d_{3i} + \beta_2 Y_i + \epsilon_i, \quad i = 1, \dots, N,$$

η παράμετρος δ_1 εκφράζει τη μεταβολή (αύξηση ή μείωση) στην πάγια κατανάλωση (ανεξαρτήτως εισοδήματος) αν

υπάρχει πανεπιστημιακή μόρφωση, η παράμετρος δ_2 εκφράζει τη μεταβολή στην πάγια κατανάλωση αν υπάρχει επιχειρηματική δραστηριότητα και η παράμετρος δ_3 εκφράζει τη μεταβολή στην πάγια κατανάλωση αν το νοικοκυριό διαμένει σε αστική περιοχή συγκριτικά με τη μη αστική.

Αν ισχύουν οι κλασικές υποθέσεις τα διάφορα μοντέλα μπορούν να εκτιμηθούν με τη μέθοδο ελαχίστων τετραγώνων και να πραγματοποιηθούν έλεγχοι στατιστικής σημαντικότητας επί μέρους συντελεστών (t-tests) καθώς και έλεγχοι γραμμικών περιορισμών (F-tests) κατά τα γνωστά.

Στο παράδειγμά μας, έλεγχος της από κοινού υπόθεσης $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$ αντιστοιχεί σε έλεγχο της υπόθεσης ότι κανένας από τους υπό μελέτη παράγοντες δεν επηρεάζει την καταναλωτική δαπάνη.

[Αλληλεπίδραση παραγόντων]: Ορίζουμε ψευδομεταβλητή αλληλεπίδρασης $d_{23i} = d_{2i}d_{3i}$, η οποία εκφράζει την από κοινού επίδραση των παραγόντων "περιοχή διαμονής" και "μορφωτικό επίπεδο" επιπλέον των ατομικών τους επιδράσεων. Στο μοντέλο

$$C_i = \beta_1 + \delta_1 d_{1i} + \delta_2 d_{2i} + \delta_3 d_{3i} + \delta_{23} d_{23i} + \beta_2 Y_i + \epsilon_i, \quad i = 1, \dots, N,$$

απόρριψη της $H_0 : \delta_{23} = 0$ αντιστοιχεί στο συμπέρασμα ότι η ταυτόχρονη παρουσία των χαρακτηριστικών "διαμονή σε αστική περιοχή" και "κατοχή πτυχίου πανεπιστημίου" οδηγεί σε μεταβολή της πάγιας κατανάλωσης επιπλέον των κυρίων επιδράσεων των αντίστοιχων παραγόντων.

[Ψευδομεταβλητές εποχικών επιδράσεων]: Μέσω ψευδομεταβλητών μπορεί να απομονωθεί και ξεχωριστή εκτίμηση των επιδράσεων της εποχικότητας.

Έστω η συνάρτηση κατανάλωσης παγωτών εφαρμοσμένη σε τριμηνιαία δεδομένα

$$C_t = \beta_1 + \beta_2 Y_t + \epsilon_t, \quad t = 1, \dots, T,$$

όπου C_t η ζήτηση για παγωτά και Y_t το εισόδημα.

Για να ληφθεί υπόψη η εποχικότητα εισάγουμε 3 ψευδομεταβλητές, d_{1t}, d_{2t}, d_{3t} , με $d_{1t} = 1$, αν είναι φθινόπωρο και $d_{1t} = 0$, διαφορετικά, $d_{2t} = 1$, αν είναι άνοιξη και $d_{2t} = 0$, διαφορετικά και $d_{3t} = 1$, αν είναι καλοκαίρι και $d_{3t} = 0$, διαφορετικά. Στο υπόδειγμα

$$C_t = \beta_1 + \delta_1 d_{1t} + \delta_2 d_{2t} + \delta_3 d_{3t} + \beta_2 Y_t + \epsilon_t, \quad t = 1, \dots, T,$$

μοντελοποιούνται οι επιδράσεις της εποχικότητας στην πάγια κατανάλωση παγωτών (ανεξαρτήτως εισοδήματος), η οποία εκφράζεται από τη σταθερά του γραμμικού μοντέλου.

Στο παραπάνω υπόδειγμα το β_1 εκφράζει την πάγια κατανάλωση παγωτών (ανεξαρτήτως εισοδήματος) το χειμώνα (εποχή που έχει λεφθεί ως "βάση"), το δ_1 εκφράζει τη μεταβολή (αναμένεται να είναι αύξηση) στην κατανάλωση παγωτών το φθινόπωρο σε σχέση με το χειμώνα (δηλαδή $\beta_1 + \delta_1$ η πάγια κατανάλωση παγωτών το φθινόπωρο), το δ_2 εκφράζει την αύξηση στην κατανάλωση παγωτών την άνοιξη σε σχέση με το χειμώνα και το δ_3 εκφράζει την αύξηση στην κατανάλωση παγωτών το καλοκαίρι σε σχέση με το χειμώνα.

Η εκτίμηση του υποδείγματος με τις ψευδομεταβλητές εποχικότητας μπορεί να δείξει τις ξεχωριστές επιδράσεις που έχουν στην κατανάλωση παγωτών οι διαφορετικές εποχές του έτος. Με αντίστοιχο τρόπο μπορούν να μελετηθούν οι επιδράσεις της εποχικότητας στην κλίση του μοντέλου.

Παραδείγματα

Ας θεωρήσουμε το υπόδειγμα εισοδήματος-κατανάλωσης με μία ψευδομεταβλητή που επιδρά τόσο στη σταθερά όσο και στην κλίση του μοντέλου.

$$C_i = \beta_1 + \delta_1 d_i + \beta_2 Y_i + \delta_2 d_i Y_i + \epsilon_i, \quad i = 1, \dots, N,$$

Το μοντέλο αυτό υποθέτει ότι ο υπό μελέτη παράγοντας (ο οποίος μπορεί να είναι ποιοτική μεταβλητή, ψευδομεταβλητή χρονικής επίδρασης ή εποχικότητας) επιρεάζει τόσο την πάγια κατανάλωση (σταθερά μοντέλου) όσο και τη ροπή προς κατανάλωση (κλίση παλινδρόμησης).

Αν πράγματι τα δεδομένα μας επιβεβαιώνουν (με κατάλληλα t-tests) ότι και οι 4 συντελεστές του παραπάνω γραμμικού μοντέλου είναι στατιστικά σημαντικοί στο γράφημα της κατανάλωσης C ως προς το εισόδημα Y θα βλέπαμε τα δεδομένα να συγκεντρώνονται ξεκάθαρα γύρω από δύο διαφορετικές ευθείες, όπως φαίνεται στο σχήμα 1. Αυτές οι δύο ευθείες έχουν αρκετά διαφορετικές θετικές κλίσεις και τέμνουν την κατακόρυφη ευθεία $x = 0$ σε διαφορετικά σημεία.

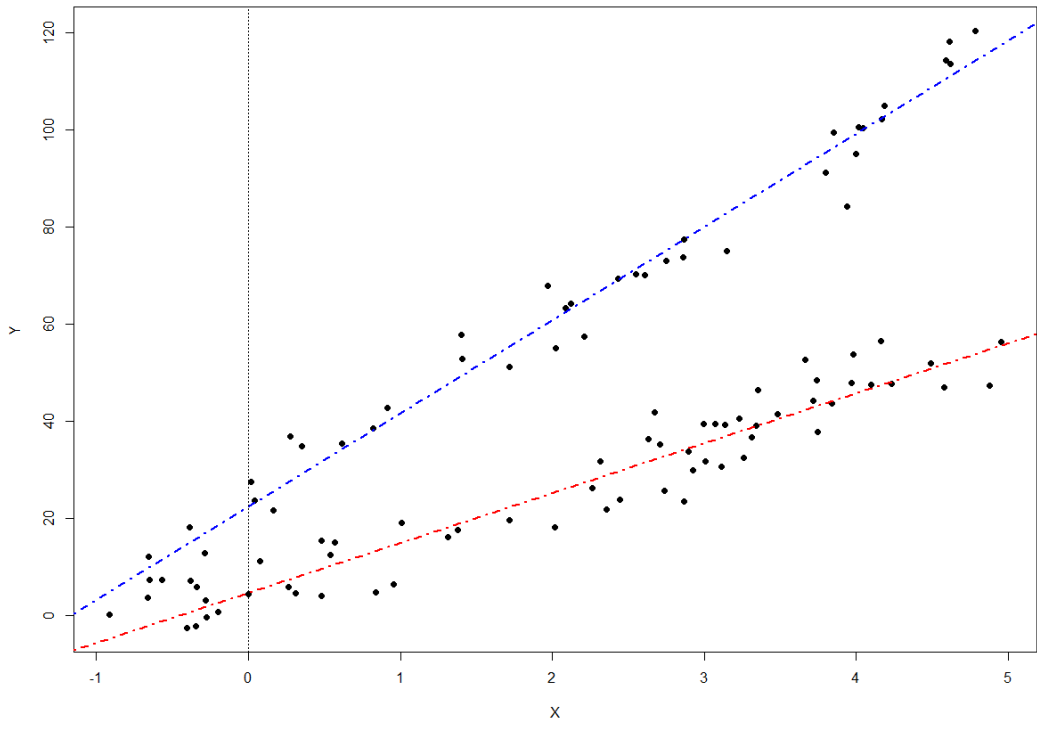


Figure 1: Δεδομένα με επίδραση της ψευδομεταβλητής στη σταθερά και στην κλίση

Αν η παράμετρος δ_2 δεν ήταν στατιστικά σημαντική, τότε θα βλέπαμε τις δύο ευθείες να είναι σχεδόν παράλληλες στο γράφημα, όπως φαίνεται στο σχήμα 2.

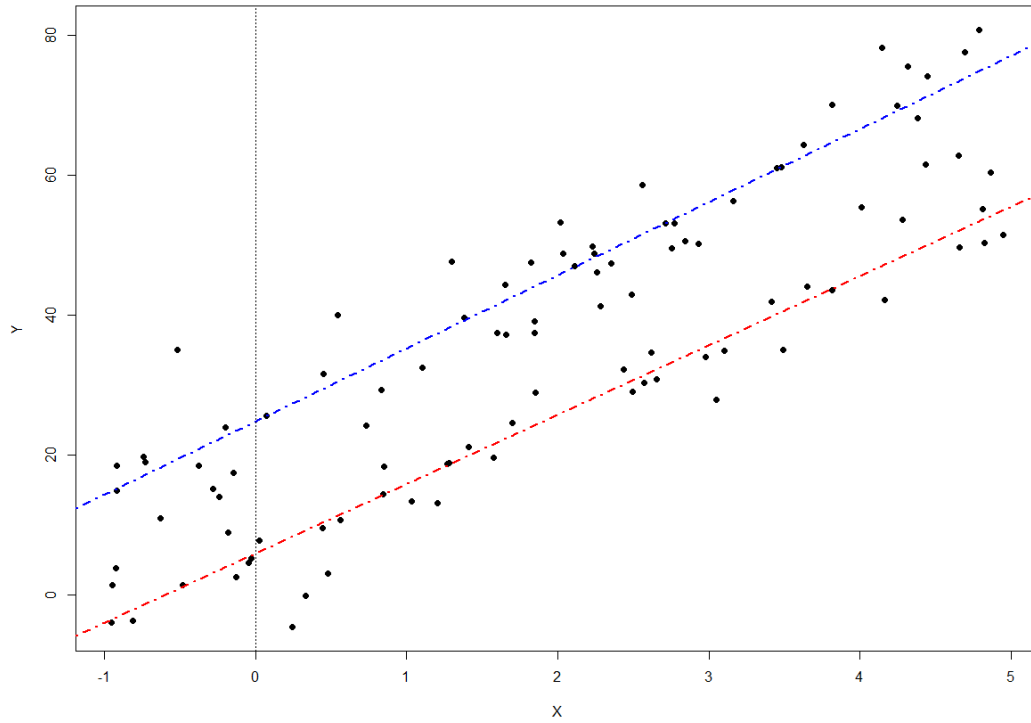


Figure 2: Παράλληλες Ευθείες Παλινδρόμησης

Αν η παράμετρος δ_1 δεν ήταν στατιστικά σημαντική τότε θα βλέπαμε τις δύο ευθείες να τέμνονται σχεδόν στο ίδιο σημείο στον άξονα των x , όπως φαίνεται στο σχήμα 3.

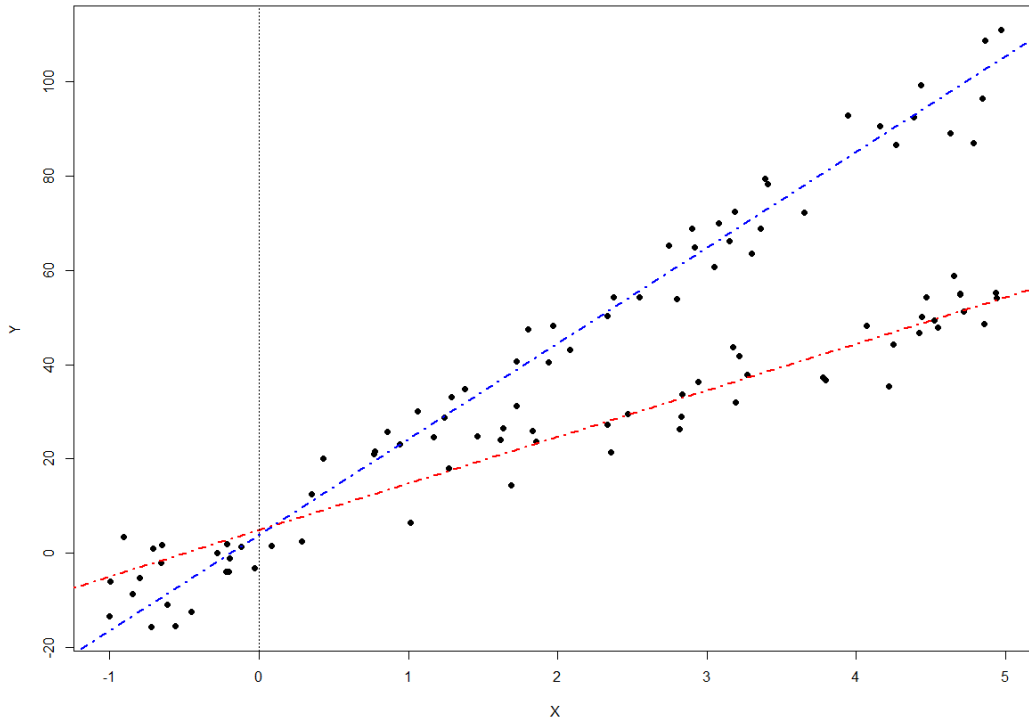


Figure 3: Τέμνουσες Ευθείες Παλινδρόμησης

Στην περίπτωση που η ποιοτική μεταβλητή έχει τρία επίπεδα (π.χ αστική, ημιαστική, αγροτική περιοχή νοικοκυριού) τότε πρέπει να φτιάξουμε δύο αντίστοιχες ψευδομεταβλητές και να μελετήσουμε χωριστά τις επιδράσεις των τριών επιπέδων στη σταθερά και την κλίση της ευθείας παλινδρόμησης.

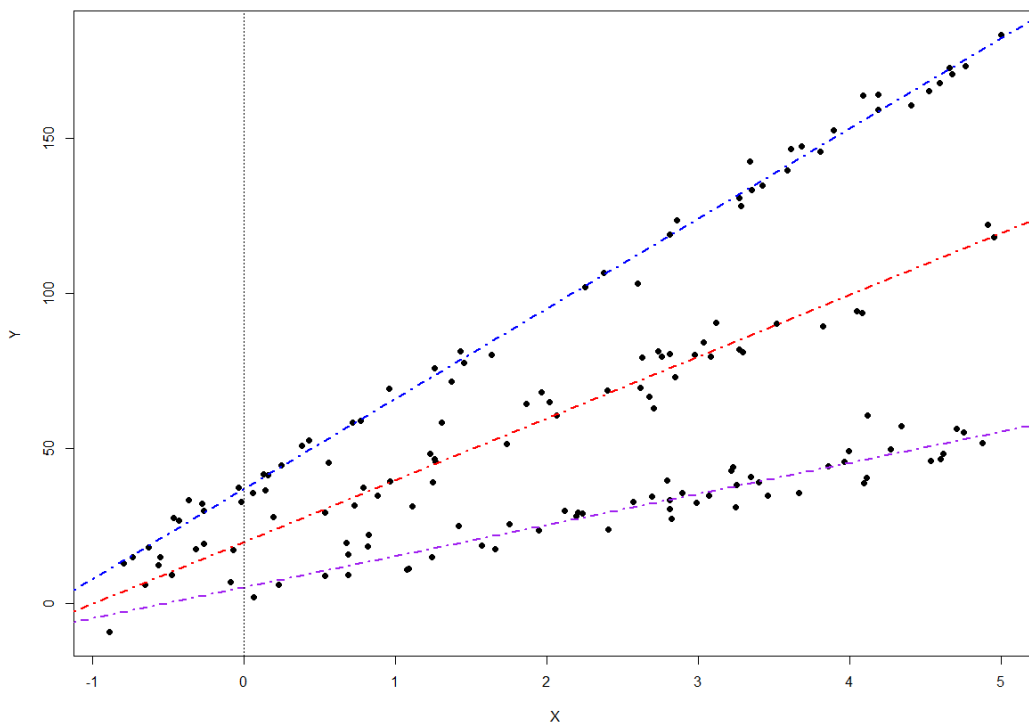


Figure 4: Ποιοτική Επεξηγηματική Μεταβλητή με 3 Επίπεδα

Στο σχήμα 4, βλέπουμε τις παρατηρήσεις να συγκεντρώνονται γύρω από τρεις ευδιάκριτες ευθείες, όλες με διαφορετικές θετικές κλίσεις και σταθερές.