

#1. Ένα άρθρο σε μία οικονομική εφημερίδα αναφέρει τα αποτελέσματα μίας ανάλυσης παλινδρόμησης ανάμεσα στην απόδοση των μετοχών (Y) και την αναλογία της ονομαστικής αξίας προς τη χρηματιστηριακή αξία (X) των μετοχών αυτών για μία συγκεκριμένη χρονική περίοδο. Η εξίσωση παλινδρόμησης η οποία υπολογίστηκε με τη μέθοδο των ελαχίστων τετραγώνων είναι: $\hat{Y} = 1,21 + 3,1X$. Δίνεται επίσης η εκτίμηση της τυπικής απόκλισης του γωνιακού συντελεστή παλινδρόμησης που είναι ίση με 2,89. Το μέγεθος του δείγματος που χρησιμοποιήθηκε είναι $n = 18$.

(α) Υπάρχουν ενδείξεις, σε ε.σ.σ. $\alpha = 5\%$, ότι υπάρχει μία γραμμική σχέση μεταξύ της απόδοσης των μετοχών και της αναλογίας της ονομαστικής προς τη χρηματιστηριακή αξία αυτών;

(β) Στην παραπάνω ανάλυση παλινδρόμησης αναφέρθηκε επίσης ότι ο συντελεστής προσδιορισμού είναι ίσος με 0,067. Σας εκπλήσσει το αποτέλεσμα αυτό; Ναι ή όχι και γιατί;

Λύση. (α) Ελέγχουμε σε ε.σ.σ. $\alpha = 5\%$, τις ακόλουθες υποθέσεις: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$. Χρησιμοποιούμε την ελεγχοσυνάρτηση: $T_{n-2} = \frac{b_1}{s_{b_1}} = \frac{3,1}{2,89} = 1,072$. Κάτω από την μηδενική υπόθεση: $T_{n-2} \sim t_{n-2}$. Σε ε.σ.σ. $\alpha = 5\%$, απορρίπτουμε την μηδενική υπόθεση αν $|T_{n-2}| > t_{n-2; \frac{\alpha}{2}}$. Από τους πίνακες τιμών των ποσοστιαίων σημείων της κατανομής t με $n-2 = 16$ β.ε. βρίσκουμε ότι $t_{n-2; \frac{\alpha}{2}} = t_{16; 0,025} = 2,12$. Άρα η μηδενική υπόθεση δεν απορρίπτεται, συνεπώς δεν φαίνεται να υπάρχουν, από τα δεδομένα μας, ισχυρές ενδείξεις γραμμικής σχέσης ανάμεσα στις μεταβλητές X και Y .

(β) Δεν μας εκπλήσσει αυτό το αποτέλεσμα διότι από το Ερώτημα (α) δεν διαπιστώθηκε μέσω του ελέγχου, στατιστικά σημαντική γραμμική σχέση ανάμεσα στις μεταβλητές X και Y . Η μεταβλητή X δεν προέκυψε να είναι στατιστικά σημαντική για το μοντέλο και για την ερμηνεία των μεταβολών των τιμών της μεταβλητής Y . Συνεπώς, πολύ λογικά, η εξίσωση παλινδρόμησης δεν «εξηγεί» (δεν «ερμηνεύει») μεγάλο ποσοστό της συνολικής μεταβλητότητας της μεταβλητής Y και για αυτόν το λόγο περιμένουμε μία χαμηλή τιμή του συντελεστή προσδιορισμού R^2 . Πράγματι, η τιμή του συντελεστή προσδιορισμού είναι μόλις 6,7%.

#2. Ένας διευθυντής πωλήσεων ενδιαφέρεται να μελετήσει τη σχέση μεταξύ των πωλήσεων της εταιρείας του (σε χιλ. Ευρώ) και των εξόδων για διαφήμιση (σε χιλ. Ευρώ). Για το σκοπό αυτό συνέλεξε δεδομένα 24 μηνών ($n = 24$) (όσον αφορά τις πωλήσεις και έξοδα για διαφήμιση) και χρησιμοποιώντας απλό γραμμικό υπόδειγμα κατέληξε στα ακόλουθα αποτελέσματα:

Πίνακας Συντελεστών του Υποδείγματος

	Συντελεστής	Τυπικό Σφάλμα Εκτίμησης	t - Στατιστικό
Σταθερά	105.5	26.239	4.021
Διαφήμιση	2.015	0.7598	2.652

(α) Συμπληρώστε τον πίνακα συντελεστών του υποδείγματος.

Λύση. (α) Έχουμε $T = \frac{b_0}{s_{b_0}} \Rightarrow b_0 = Ts_{b_0} = 105.5$. Επίσης, $T = \frac{b_1}{s_{b_1}} \Rightarrow s_{b_1} = \frac{b_1}{T} = 0.7598$.

Παρακάτω δίνεται ο Πίνακας Ανάλυσης Διακύμανσης (ANOVA table).

Πίνακας Ανάλυσης Διακύμανσης ANOVA				
	Άθροισμα	Β.ε.	Μέσο Άθροισμα	F
	Τετραγώνων		Τετραγώνων	
Παλινδρόμηση	11809,406	1	11809,406	7,034
Κατάλοιπα	36935,8	22	1678,90	
Σύνολο	48745,206	23		

(β) Να συμπληρωθεί ο πίνακας.

Είναι: $\frac{11809,406}{MSE} = 7,034 \Rightarrow MSE = \frac{11809,406}{7,034} = 1678,90$. Επίσης, $\frac{SSE}{22} = 1678,9 \Rightarrow SSE = 36935,8$

(γ) Να γράψετε το υπόδειγμα της απλής παλινδρόμησης που έχει εκτιμηθεί από τον διευθυντή πωλήσεων και να ερμηνεύσετε τους συντελεστές του.

Η εκτιμώμενη ευθεία απλής γραμμικής παλινδρόμησης είναι: $\hat{Y} = 105,5 + 2,015X$

Ερμηνεία των συντελεστών: Για μηδενικές δαπάνες σε διαφήμιση ($X = 0$) οι πωλήσεις (σε χιλιάδες ευρώ) αναμένεται κατά μέσο όρο να ανέλθουν σε 105,5 χιλιάδες ευρώ. Για κάθε μία μονάδα αύξησης των εξόδων σε διαφήμιση (για κάθε 1000 ευρώ αύξηση) οι πωλήσεις αναμένεται να αυξηθούν κατά περίπου 2,015 χιλιάδες ευρώ.

(δ) Να διεξάγετε τους ακόλουθους ελέγχους υποθέσεων που αφορούν την παράμετρο β_1 . Σε ποια συμπεράσματα καταλήγετε; Να χρησιμοποιήσετε επίπεδο στατιστικής σημαντικότητας $\alpha = 0.05$.

(i) Ελέγχουμε σε ε.σ.σ. $\alpha = 0.05$, τις εξής υποθέσεις: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$. Χρησιμοποιούμε τη στατιστική συνάρτηση ελέγχου $T_{n-2} = \frac{b_1}{s_{b_1}} = \frac{2,015}{0,758} = 2,6520$. Απορρίπτουμε την $H_0 : \beta_1 = 0$ σε ε.σ.σ. $\alpha = 0.05$, αν

$|T_{n-2}| > t_{n-2;0.025}$. Από πίνακες της κατανομής t_{22} βρίσκουμε ότι: $t_{22;0.025} = 2,074$. Η μηδενική υπόθεση απορρίπτεται σε ε.σ.σ. $\alpha = 0.05$. Συνεπώς, τα δεδομένα παρέχουν ενδείξεις ότι οι διαφημιστικές δαπάνες φαίνεται να επιδρούν στατιστικά σημαντικά στις πωλήσεις της εταιρείας.

(ii) Ελέγχουμε σε ε.σ.σ. $\alpha = 0.05$, τις εξής υποθέσεις: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$. Χρησιμοποιούμε τη στατιστική συνάρτηση ελέγχου $T_{n-2} = \frac{b_1}{s_{b_1}}$, για την οποία βρήκαμε ότι: $T_{n-2} = 2,6520$. Απορρίπτουμε την $H_0 : \beta_1 = 0$ σε

ε.σ.σ. $\alpha = 0.05$, αν $T_{n-2} > t_{n-2;0.05}$. Από πίνακες της κατανομής t_{22} βρίσκουμε ότι: $t_{22;0.05} = 1,717$. Η μηδενική υπόθεση απορρίπτεται σε ε.σ.σ. $\alpha = 0.05$. Συνεπώς, τα δεδομένα παρέχουν ισχυρές ενδείξεις ότι οι διαφημιστικές δαπάνες φαίνεται να ασκούν στατιστικά σημαντική θετική επίδραση στις πωλήσεις της εταιρείας.

(ε) Να υπολογιστεί ο συντελεστής προσδιορισμού του υποδείγματος αυτού και να ερμηνευτεί η τιμή του.

Είναι: $R^2 = \frac{SSR}{SST} = \frac{11809,406}{48745,206} = 0,24226$. Δηλαδή μόνο το 24,22% της συνολικής μεταβλητότητας των πωλήσεων της εταιρείας ερμηνεύεται (επεξηγείται) μέσω των δαπανών της εταιρείας για διαφήμιση.

(ζ) Να υπολογιστεί το τυπικό σφάλμα εκτίμησης της εξίσωση παλινδρόμησης και να ερμηνευτεί η τιμή του.

Είναι: $s_e = s = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE} = \sqrt{1678,90} = 40,97$. Είναι η εκτίμηση της τυπικής απόκλιση των (εκτιμώμενων από το δείγμα) σφαλμάτων γύρω από την (εκτιμώμενη) ευθεία παλινδρόμησης. Τα σφάλματα απέχουν περίπου $\pm 40,97 \cong 41$ μονάδες από την ευθεία παλινδρόμησης (σχετικά μεγάλο (σίγουρα όχι αμελητέο) τυπικό σφάλμα εκτίμησης).

(η) Να διεξάγετε στατιστικό έλεγχο σημαντικότητας της εξίσωσης παλινδρόμησης σε επίπεδο στατιστικής σημαντικότητας $\alpha = 0.05$. Σε ποιο συμπέρασμα καταλήγετε;

Ελέγχουμε σε ε.σ.σ. $\alpha = 0.05$, τις ακόλουθες υποθέσεις:

$H_0 : \beta_0 = \beta_1 = 0$ vs $H_1 : \beta_i \neq 0$, για ένα τουλάχιστον $i = 0, 1$.

Εναλλακτικά, H_0 : η εξίσωση παλινδρόμησης δεν εξηγεί καθόλου τις μεταβολές της Y .

έναντι της H_1 : η εξίσωση παλινδρόμησης εξηγεί ένα μέρος των μεταβολών της Y .

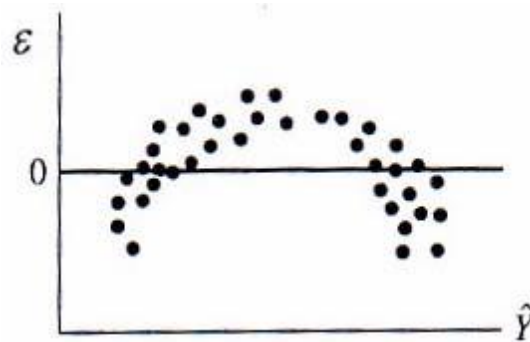
Χρησιμοποιούμε τη σ.σ. ελέγχου $F_{1,22}$ η οποία κάτω από την μηδενική υπόθεση ακολουθεί την κατανομή F με 1 β.ε. (για τον αριθμητή) και 22 β.ε. (για τον παρονομαστή), αντίστοιχα. Από τους πίνακες της κατανομής F για αυτούς τους βαθμούς ελευθερίας βρίσκουμε ότι: $F_{1,22;0.05} = 4,3$. Απορρίπτουμε τη μηδενική υπόθεση σε ε.σ.σ. $\alpha = 0.05$, αν $F_{1,22} = 7,034 > F_{1,22;0.05} = 4,3$. Συνεπώς, η μηδενική υπόθεση απορρίπτεται σε ε.σ.σ. $\alpha = 0.05$ και επομένως τα δεδομένα παρέχουν ενδείξεις ότι η εξίσωση παλινδρόμησης εξηγεί ένα μέρος των μεταβολών της μεταβλητής Y .

#3. Α. Σε ένα πρόβλημα απλής γραμμικής παλινδρόμησης η ευθεία γραμμικής παλινδρόμησης έχει εκτιμηθεί από τη σχέση $\hat{Y} = 0,85 - 1,2X$ και ο συντελεστής προσδιορισμού υπολογίστηκε ίσος με 0,81. Να υπολογίσετε τον συντελεστή συσχέτισης των μεταβλητών X και Y .

Λύση. Α. Στο μοντέλο απλής γραμμικής παλινδρόμησης ισχύει ότι: $r_{X,Y}^2 = R^2 \Rightarrow r_{X,Y}^2 = 0,81 \Rightarrow r_{X,Y} = -0,9$. Ο συντελεστής γραμμικής συσχέτισης διατηρεί το πρόσημο του εκτιμώμενου συντελεστή παλινδρόμησης b_1 . Εδώ έχουμε $b_1 = -0,12$, συνεπώς $r_{X,Y} = -0,9$. Η τιμή του συντελεστή γραμμικής συσχέτισης μεταξύ των μεταβλητών X και Y παρέχει ενδείξεις έντονης αρνητικής γραμμικής συσχέτισης των δύο μεταβλητών.

Β. Σε ένα πρόβλημα απλής γραμμικής παλινδρόμησης έχουν εκτιμηθεί οι παράμετροι του υποδείγματος και έχουν υπολογιστεί τα κατάλοιπα. Να κατασκευάσετε διαγράμματα καταλοίπων στα οποία να αποτυπώνονται ενδείξεις για παραβίαση των βασικών υποθέσεων του απλού γραμμικού υποδείγματος.

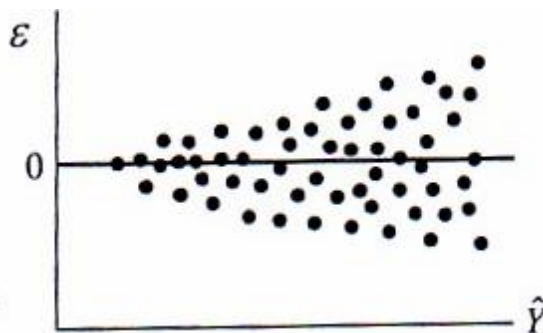
(i) Έλλειψη γραμμικότητας.



Σχήμα 1

Στο Σχήμα 1, παρατηρούμε ότι τα κατάλοιπα δεν συγκεντρώνονται τυχαία γύρω από την ευθεία $\varepsilon = 0$ (όπως πρέπει να συμβαίνει για να τηρείται η υπόθεση της γραμμικότητας), αλλά ακολουθούν ένα κυκλικό πρότυπο (αρνητικές-θετικές-αρνητικές τιμές). Αυτή η «κυκλική» συμπεριφορά φανερώνει ενδείξεις για παλινδρόμηση π.χ. δευτέρου βαθμού ως προς τη μεταβλητή X .

(ii) Έλλειψη ομοσκεδαστικότητας (ενδείξεις για ετεροσκεδαστικότητα (μη σταθερή διακύμανση) των καταλοίπων).

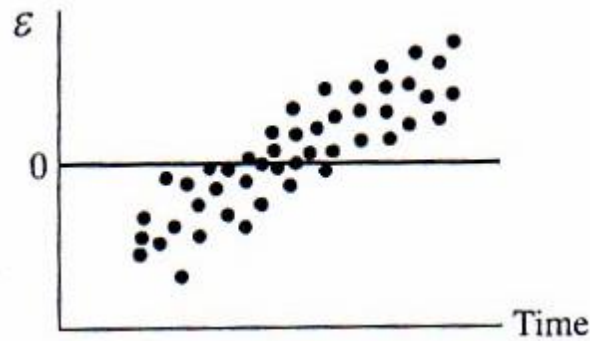


Σχήμα 2

Σε διάφορες οικονομικές και εμπορικές εφαρμογές, τα κατάλοιπα σε σχέση με τη μεταβλητή \hat{Y} δίνουν διαγράμματα, όπως του Σχήματος 2, μορφής τραπεζίου (δηλαδή αυξανόμενων των τιμών της \hat{Y} αυξάνει η διασπορά των καταλοίπων).

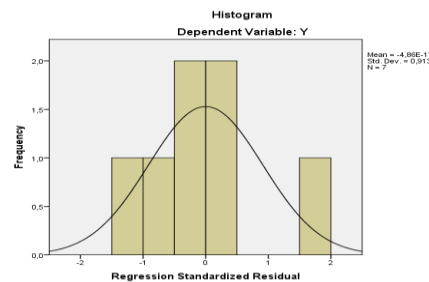
(iii) Έλλειψη ανεξαρτησίας καταλοίπων (ή ενδείξεις συσχέτισης των καταλοίπων). Εξαρτημένα Y εμφανίζονται συνήθως όταν παίρνουμε παρατηρήσεις από την ίδια πειραματική μονάδα σε διαφορετικές χρονικές στιγμές (π.χ. μετράμε την πίεση ή το βάρος του ίδιου ατόμου ανά εβδομάδα). Επίσης, σε περιπτώσεις όπου χρησιμοποιούνται μηχανές (όργανα μέτρησης, κ.τλ) που αλλάζει η απόδοσή τους με τη χρήση ή ο

χειριστής βελτιώνεται (ή χειροτερεύει) με την πάροδο του χρόνου. Σε τέτοιες περιπτώσεις προκύπτουν διαγράμματα καταλοίπων σαν αυτό του Σχήματος 3 που υποδηλώνει ενδείξεις εξάρτησης-συσχέτισης μεταξύ των καταλοίπων.



Σχήμα 3

(δ) Έλλειψη κανονικότητας των καταλοίπων. Κατασκευάζοντας τα ιστογράμματα των καταλοίπων ή και των τυποποιημένων τιμών τους (διαιρούμε τις τιμές των καταλοίπων με την τυπική τους απόκλιση s_e) μπορούμε να λάβουμε εμπειρικές ενδείξεις για να διερευνήσουμε αν τα κατάλοιπα (ή ισοδύναμα η εξαρτημένη μεταβλητή) στο μοντέλο μας κατανέμονται κανονικά ή όχι.



Σχήμα 4

Το Σχήμα 4, παρέχει (μάλλον) ικανοποιητικές ενδείξεις ότι η μεταβλητή Y μπορεί να θεωρηθεί ότι κατανέμεται σύμφωνα με την κανονική κατανομή (αν και φυσικά υπάρχουν σαφέστερες αποκλίσεις από την κανονικότητα προς την δεξιά ουρά της κατανομής των καταλοίπων).

#4. Α. Ποιο πρόβλημα αντιμετωπίζεται με τη χρησιμοποίηση του διορθωμένου συντελεστή πολλαπλού προσδιορισμού στα υποδείγματα της πολλαπλής παλινδρόμησης.

Η προσθήκη πολλών ανεξαρτήτων μεταβλητών μπορεί να οδηγήσει σε "τεχνητή" αύξηση της τιμής του R^2 που δεν θα έχει καμία αξία όταν μάλιστα ο αριθμός των ανεξάρτητων μεταβλητών (k) είναι υψηλός σε σχέση με το μέγεθος τους δείγματος (n). Το πρόβλημα αντιμετωπίζεται με τον διορθωμένο R^2_{adj} , που λαμβάνει υπόψη του, τους βαθμούς ελευθερίας. Με τη χρήση του διορθωμένου συντελεστή πολλαπλού προσδιορισμού απαντάμε στο ερώτημα: Είναι η αύξηση του R^2 τόσο σημαντική ώστε να αξίζει την απώλεια ενός β.ε;

Β. Ποια είναι η ερμηνεία του συντελεστή μερικού προσδιορισμού της μεταβλητής X_2 στο υπόδειγμα $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$;

Ο συντελεστής μερικού προσδιορισμού της X_2 δίνει το ποσοστό της ανερμήνευτης διασποράς της Y που δεν εξηγεί η επίδραση της X_1 και που θα ερμηνεύσει (επεξηγήσει) η νέα μεταβλητή X_2 . Δηλαδή, τι ποσοστό της

ανεξήγητης (από τις επιδράσεις της X_1) διασποράς της ανεξάρτητης μεταβλητής Y μπορεί να ερμηνεύσει η νέα ανεξάρτητη μεταβλητή X_2 που υπάρχει στο μοντέλο.

#5. Μία κατασκευαστική εταιρεία ενδιαφέρεται να μελετήσει τη σχέση μεταξύ του αριθμού των οικοδομικών αδειών (Y) και των επιτοκίων (X). Για το λόγο αυτό συγκέντρωσε στοιχεία που αφορούν τις μεταβλητές αυτές για τα έτη 1990-1997 τα οποία παρουσιάζονται στον παρακάτω πίνακα:

Έτος	1990	1991	1992	1993	1994	1995	1996	1997
Y (σε χιλιάδες)	2.2	3	2.8	1.9	1.8	1.5	1.3	2.1
X (σε %)	6.5	6	6.5	7.5	8.5	9.5	9	7.5

Με βάση τα παραπάνω δεδομένα να απαντήσετε στα ακόλουθα ερωτήματα:

(α) Να εκτιμηθεί η ευθεία παλινδρόμησης του αριθμού των οικοδομικών αδειών ως προς την τιμή των επιτοκίων. Να ερμηνεύσετε τις τιμές των συντελεστών γραμμικής παλινδρόμησης.

$$\text{Υπολογίζουμε: } \bar{x} = 7,625, \bar{y} = 2,075, \sum_{i=1}^8 x_i y_i = 121,75, \sum_{i=1}^8 x_i^2 = 476,5.$$

$$\text{Βρίσκουμε την εκτιμώμενη ευθεία παλινδρόμησης: } \hat{Y} = b_0 + b_1 X = 5,3 - 0,42X.$$

Για μηδενικό επιτόκιο ο μέσος αριθμός των οικοδομικών αδειών αναμένεται να είναι περίπου ίσος με 5,3.

Για κάθε αύξηση του επιτοκίου κατά μία ποσοστιαία μονάδα αναμένεται μείωση του μέσου αριθμού των οικοδομικών αδειών κατά 0,42% (περίπου κατά μισή ποσοστιαία μονάδα), δηλαδή μείωση του αριθμού των αδειών κατά 0,5%.

(β) (Homework! Κάντε τις πράξεις!!). Αν υποθέσουμε ότι το επιτόκιο το επόμενο έτος (1998) θα είναι 7% να προβλεφθεί ο μέσος αριθμός οικοδομικών αδειών (σε χιλιάδες) το επόμενο έτος. Να κατασκευαστεί το 95% δ.ε. για το μέσο αριθμό οικοδομικών αδειών σε αυτήν την περίπτωση. Για ποια τιμή της μεταβλητής X το διάστημα θα έχει το μικρότερο εύρος;

Για $x_0 = 0.07 = 7\%$ βρίσκουμε την προβλεπόμενη τιμή $\hat{Y}_0 = b_0 + b_1 X_0 = 5,3 - 0,42X_0$. Το διάστημα πρόβλεψης

$$\text{είναι: } \hat{Y}_0 \pm t_{n-2; 0.025} s_{\hat{Y}_0}, \text{ όπου } s_{\hat{Y}_0} = s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

$$\text{Όμως, } s_e = \sqrt{\frac{SSE}{n-2}}. \text{ SSE} = \text{SST} - \text{SSR}, \text{ SST} = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2, n = 8. \text{ SSR} = b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2.$$

Εδώ $\alpha = 0.05$. Το εύρος d του διαστήματος εμπιστοσύνης είναι: $d = 2t_{n-2;0.025} s_{\hat{Y}_0}$ και επιτυγχάνει την ελάχιστη τιμή του όταν η ποσότητα εντός της τετραγωνικής ρίζας στην έκφραση $s_{\hat{Y}_0} = s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$ γίνεται ελάχιστη. Αυτό συμβαίνει όταν $x_0 = \bar{x} = 7,625$.

Καλό καλοκαίρι!!