

ΑΣΚΗΣΕΙΣ

1. Θεωρήστε το παράδειγμα που αναφέρεται στη συσχέτιση του βαθμού ικανοποίησης των εργαζομένων σε ένα εργαστήριο σε σχέση με τις οκτώ μεταβλητές που ορίστηκαν εκεί. (X_1 =ηλικία, X_2 =φύλο, X_3 =εβδομαδιαίος μισθός, X_4 =χρόνια προϋπηρεσίας στο εργαστήριο, X_5 =χρόνια εμπειρίας σε σχετική δουλειά, X_6 =επίπεδο εκπαίδευσης, X_7 =ποσοστό συνολικού εισοδήματος και X_8 =διοικητική θέση).

(α) Παρατηρώντας την εκτύπωση του προγράμματος του υπολογιστή για το μερικό μοντέλο που περιλαμβάνει τις μεταβλητές X_2 , X_3 , X_6 και X_8 να γράψετε την εξίσωση ελαχίστων τετραγώνων για το μοντέλο αυτό.

(β) Να ερμηνεύσετε τις τιμές των εκτιμητριών των συντελεστών των μεταβλητών X_2 , X_3 , X_6 και X_8 σε σχέση με το βαθμό ικανοποίησης Y στην εργασία.

(γ) Χρησιμοποιώντας την εκτύπωση του υπολογιστή για το πλήρες μοντέλο (με όλες τις μεταβλητές X_1 , X_2 , ..., X_8) που δίνεται στο παράδειγμα, να γράψετε την εξίσωση ελαχίστων τετραγώνων για το μοντέλο αυτό.

(δ) Εξηγήστε γιατί οι συντελεστές των μεταβλητών X_2 , X_3 , X_6 και X_8 είναι διαφορετικοί από αυτούς που βρέθηκαν στο ερώτημα (β).

(ε) Να κατασκευασθεί ένα 95% διάστημα εμπιστοσύνης για την εκτιμήτρια β_6 της μεταβλητής X_6 στο μερικό μοντέλο με τις τέσσερις μεταβλητές της ερώτησης (β) με χρήση της εκτύπωσης του προγράμματος, όπως αυτή δίνεται στο παράδειγμα.

(στ) Να ερμηνευθεί το διάστημα αυτό.

(ζ) Η ερμηνεία που δόθηκε στο προηγούμενο ερώτημα συμφωνεί με τις απόψεις που έχετε για το επίπεδο εκπαίδευσης;

(η) Να κατασκευασθεί ένα 95% διάστημα εμπιστοσύνης για το συντελεστή της μεταβλητής X_6 στο πλήρες μοντέλο με βάση την εκτύπωση του προγράμματος που δίνεται στο σχετικό παράδειγμα. Εξηγήστε γιατί αυτό το διάστημα εμπιστοσύνης είναι διαφορετικό από εκείνο που βρήκατε στο προηγούμενο ερώτημα.

2. Ένας εκτιμητής ακίνητης περιουσίας χρησιμοποιεί τις μεθόδους της πολλαπλής παλινδρόμησης προκειμένου να καθορίσει μια κατ' αρχήν τιμή για την αξία μιας μονοκατοικίας. Προκειμένου να καθορίσει μια κατ' αρχήν τιμή, θεωρεί τρεις ανεξάρτητες μεταβλητές που πιστεύει ότι επηρεάζουν την τιμή αυτή. Οι μεταβλητές αυτές είναι:

X_1 = Ηλικία του σπιτιού,

X_2 = Εμβαδόν (σε m^2),

X_3 = Τρόπος κατασκευής ($X_3=0$, αν η κατασκευή είναι από πέτρα και $X_3=1$ αν η κατασκευή είναι με τούβλο).

Οι τιμές με τις οποίες πουλήθηκαν πρόσφατα 45 μονοκατοικίες στην περιοχή αυτή χρησιμοποιήθηκαν για προσαρμογή του μοντέλου. Ένα μέρος της εκτύπωσης του προγράμματος για το μοντέλο αυτό και της ανάλυσης με τη μέθοδο των ελαχίστων τετραγώνων δίνεται στον πίνακα που ακολουθεί.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	(a)	2.882 E04	(e)	(g)	(h)
ERROR	(b)	(d)	(f)		
TOTAL	(c)	4.063 E04			
R SQUARE (i)					
PARAMETER	ESTIMATE	T FOR H_0 PARAMETER=0	PR> T	STD ERROR OF ESTIMATE	
INTERCEPT	2.088 E03				
X_1	-9.357 E02	-2.66	(l)	(o)	
X_2	4.581 E01	(k)	(m)	1.656 E01	
X_3	(j)	1.88	(n)	3.190 E03	

Να συμπληρώσετε τα κενά στις θέσεις από (a) έως (o) στον πίνακα αυτό.

3. Ένας έφορος με ιδιαίτερη κλίση στη Στατιστική ενδιαφέρεται να βρει μια γρήγορη μέθοδο με την οποία να προσεγγίζεται ο φόρος εισοδήματος που αντιστοιχεί σε κάθε φορολογούμενο με την χρήση μερικών μεταβλητών για τις οποίες είναι εύκολο να έχει διαθέσιμα στοιχεία. Για

το σκοπό αυτό επέλεξε τυχαία 100 φορολογικές δηλώσεις του προηγούμενου έτους, τις οποίες και εξέτασε. Για κάθε μια από αυτές έκανε καταγραφή των τιμών των τριών μεταβλητών τις οποίες όρισε και του αντίστοιχου φόρου εισοδήματος που ο φορολογούμενος πλήρωσε. Η εκτύπωση του υπολογιστή για την προσαρμογή της ευθείας πολλαπλής παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων δίνεται στον πίνακα που ακολουθεί.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	(a)	(d)	(e)	(g)	(h)
ERROR	(b)	1.012 E06	(f)		
TOTAL	(c)	3.882 E06			
R SQUARE (i)					
PARAMETER	ESTIMATE	T FOR H ₀ PARAMETER=0	PR> T	STD ERROR OF ESTIMATE	
INTERCEPT	-8.416 E-01				
X ₁	3.608 E-03	2.03	(m)	(o)	
X ₂	(j)	(l)	.0500	7.3751 E-03	
X ₃	(k)	1.79	(n)	4.8831 E-02	

Να συμπληρωθούν τα κενά από το (a) έως το (o).

4. Χρησιμοποιώντας την εκτιμηθείσα εξίσωση παλινδρόμησης με την μέθοδο του αποκλεισμού μεταβλητών για το παράδειγμα που αναφέρεται στην ικανοποίηση από τις συνθήκες εργασίας, στο οποίο καταλήξαμε με τις ανεξάρτητες μεταβλητές X₃ και X₆, να προβλέψετε το βαθμό ικανοποίησης από την εργασία για ένα εργαζόμενο άνδρα ηλικίας 45 ετών που δουλεύει σε μη διευθυντική θέση, έχει εβδομαδιαίο μισθό 40.000 δρχ., δουλεύει στην εταιρεία αυτή επί 17 χρόνια χωρίς άλλη επαγγελματική πείρα με 13 χρόνια εκπαίδευσης και με μοναδικό εισόδημα τον μισθό από την εργασία αυτή.

5. (α) Για το ίδιο παράδειγμα, κρίνοντας από τον πίνακα που αναφέρεται στην εκτύπωση του υπολογιστή από το 70 βήμα του αποκλεισμού μεταβλητών, να εκφέρετε την άποψή σας για το κατά πόσο υψηλότερες αμοιβές τείνουν να σχετίζονται με μεγαλύτερη ικανοποίηση από την εργασία.

(β) Υπάρχουν ενδείξεις ότι περισσότερη εκπαίδευση τείνει να σχετίζεται με μεγαλύτερη ικανοποίηση στη δουλειά;

(γ) Αν ένας εργαζόμενος έχει δύο χρόνια εκπαίδευσης περισσότερο από ένα άλλο εργαζόμενο, πόσα περισσότερα χρήματα θα πρέπει να παίρνει ο πρώτος σε εβδομαδιαίο μισθό ώστε να έχει την ίδια ικανοποίησης από την δουλειά του όπως ο άλλος εργαζόμενος;

6. (α) Να χρησιμοποιηθεί η εκτιμηθείσα εξίσωση παλινδρόμησης για το πρόβλημα αυτό που περιλαμβάνει τις μεταβλητές X_3 και X_6 για να προβλεφθεί ο μέσος βαθμός ικανοποίησης στη δουλειά όλων των εργαζομένων με εβδομαδιαίο μισθό 36.000 δρχ. και 17 χρόνια εκπαίδευσης.

(β) Ποιος είναι ο αριθμός των βαθμών ελευθερίας που αναφέρονται στην κατανομή t που χρησιμοποιείται για την κατασκευή του διαστήματος πρόβλεψης;

(γ) Για ένα 95% διάστημα πρόβλεψης για το Y που έχει ως άκρα τις τιμές 4.19 έως 11.98, να βρεθεί το τυπικό σφάλμα του $\hat{Y}$.

(δ) Για ένα 95% διάστημα εμπιστοσύνης για το $\mu_{Y|x}$ με άκρα 7.18 και 8.99, να βρεθεί το τυπικό σφάλμα του $\hat{\mu}_{Y|x}$.

7. Ένας μεσίτης προκειμένου να μελετήσει τις τιμές πώλησης διαμερισμάτων για κάποια συγκεκριμένη χρονική περίοδο, συγκέντρωσε τα στοιχεία που δίνονται στον πίνακα που ακολουθεί.

Στοιχεία για Πωλήσεις Διαμερισμάτων

Τιμή πώλησης Y (σε εκ.δρχ.)	X ₁	X ₂	D ₁	D ₂	D ₃	X ₃	X ₄
4.4	0	88	0	0	0	1	6
5.0	1	100	0	1	0	2	7
5.2	0	97	0	1	1	1	14
3.2	16	90	1	0	0	6	1
3.2	4	83	0	1	0	1	3
5.0	8	110	1	0	0	3	2
5.1	0	97	0	0	1	1	4
3.8	11	94	0	0	1	2	1
4.3	3	89	0	0	1	1	7
5.3	5	109	1	0	0	1	3
3.0	8	73	0	0	1	4	1
5.3	8	120	0	1	0	7	2
4.9	7	108	0	0	1	3	2
5.9	1	110	1	0	1	1	8
3.5	3	76	0	0	0	1	5
5.2	1	100	1	0	0	2	4
5.5	2	112	0	1	0	1	6
5.0	7	112	0	1	0	3	3
6.9	7	141	1	0	1	2	4
4.3	0	83	0	0	1	3	8
4.4	0	88	0	0	0	1	7
5.7	19	146	0	1	1	1	1
4.5	12	103	1	0	1	14	1
3.7	3	77	0	0	1	8	2
4.8	1	91	1	0	0	1	4

Οι μεταβλητές αυτές είναι οι εξής:

X₁=παλαιότητα του διαμερίσματος (σε έτη)

X₂=επιφάνεια (σε m²)

D₁=1, αν το διαμέρισμα είναι 2 δωματίων και 0, διαφορετικά

D₂=1, αν το διαμέρισμα είναι 3 δωματίων και 0, διαφορετικά

D₃=1, αν το διαμέρισμα έχει χώρο parking και 0, διαφορετικά

X₃=όροφος του διαμερίσματος

X_4 =απόσταση από το κέντρο της πόλης (σε χιλιόμετρα)

Άλλες μεταβλητές που περιλαμβάνονται στο μοντέλο είναι οι $X_1^2, X_2^2, X_3^2, X_4^2$ και οι όροι αλληλεπίδρασης $X_1X_2, X_1X_3, X_1X_4, X_2X_3, X_2X_4, X_3X_4, D_2X_2, D_3X_2, D_1D_3, D_2D_3$, όπως επίσης και οι όροι αλληλεπίδρασης $D_1D_3X_2$ και $D_2D_3X_2$. Παρατηρούμε δηλαδή ότι υπάρχουν 24 μεταβλητές και επομένως υπάρχουν 25 παράμετροι που πρέπει να εκτιμηθούν από 25 μόνο παρατηρήσεις.

(α) Εξηγήστε τους λόγους για τους οποίους η μέθοδος αποκλεισμού μεταβλητών δεν είναι δυνατόν να εφαρμοσθεί στο πρόβλημα αυτό.

(β) Εξηγήστε τους λόγους για τους οποίους κάθε μια από τις μεταβλητές αλληλεπίδρασης έχουν περιληφθεί στο μοντέλο.

(γ) Η μέθοδος της βηματικής παλινδρόμησης για το συγκεκριμένο μοντέλο δίνει την εκτύπωση που ακολουθεί για το πρώτο βήμα της βηματικής παλινδρόμησης.

Πρώτο Βήμα Βηματικής Παλινδρόμησης

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	1	1.473 E09	(b)	55.87	.0001
ERROR	23	6.064 E08	(c)		
TOTAL	24	(a)			
R SQUARE .7084					
PARAMETER	ESTIMATE	T FOR H ₀ PARAMETER=0	PR> T	STD ERROR OF ESTIMATE	
INTERCEPT	3737.205-01				
X_2	43.137	7.47	.0001	5.771	

(α) Να συμπληρώσετε τους αριθμούς που αντιστοιχούν στις θέσεις (a), (b) και (c).

(β) Ποια από τις 24 μεταβλητές έχει τον υψηλότερο συντελεστή συσχέτισης με το Y και ποια είναι η τιμή του συντελεστή αυτού;

(γ) Ποια ερμηνεία δίνεται στον συντελεστή 43.137 που δίνεται στη συγκεκριμένη εκτύπωση;

(δ) Το δεύτερο βήμα της βηματικής παλινδρόμησης του μοντέλου δίνει ως εκτύπωση τον πίνακα που ακολουθεί.

Δεύτερο Βήμα Βηματικής Παλινδρόμησης

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	2	1.988 E09	9.941 E08	239.97	.0001
ERROR	22	9.114 E08	4.143 E06		
TOTAL	24	2.079 E09			

		R SQUARE (a)		
PARAMETER	ESTIMATE	T FOR H ₀ PARAMETER=0	PR> T	STD ERROR OF ESTIMATE
INTERCEPT	-10834.152			
X ₂	62.730E-03	(b)	.0001	2.884
X ₁ *X ₂	-0,915	(c)	.0001	.082

Να αντικατασταθούν τα (a), (b) και (c) με τους κατάλληλους αριθμούς.

(ε) Ποια μεταβλητή έχει προστεθεί στο μοντέλο;

(στ) Λαμβάνοντας υπόψη τις μεταβλητές που υπάρχουν τώρα στο μοντέλο και τους συντελεστές τους, όπως αυτοί δίνονται στην εκτύπωση, ερμηνεύσατε το αποτέλεσμα που είχε η προσθήκη της νέας μεταβλητής στο μοντέλο.

(ζ) Το τρίτο βήμα στη βηματική παλινδρόμηση του μοντέλου δίνει την εκτύπωση που ακολουθεί. Να συμπληρωθούν τα στοιχεία (a) έως (e).

(η) Ποια ερμηνεία δίνετε στην νέα μεταβλητή που προστέθηκε στο μοντέλο και στο συντελεστή της;

(θ) Γιατί δεν έχει απομακρυνθεί καμιά μεταβλητή από το μοντέλο στο βήμα αυτό;

Τρίτο Βήμα Βηματικής Παλινδρόμησης

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	(a)	2.015 E09	6.716 E08	218.43	.0001
ERROR	(b)	(d)	3.075 E06		
TOTAL	(c)	(e)			
R SQUARE .9689					
PARAMETER	ESTIMATE	T FOR H ₀ PARAMETER=0	PR> T	STD ERROR OF ESTIMATE	
INTERCEPT	-12437.967				
X ₂	65.059E-03	24.94	.0001	2.608	
D ₂	-2455.459	-2.94	.0078	835.304	
X ₁ *X ₂	-0.922	-13.03	.0001	7.075 E-022	

(ι) Το τέταρτο βήμα της βηματικής παλινδρόμησης στο μοντέλο αυτό δίνει την ακόλουθη εκτύπωση. Να συμπληρωθούν τα κενά από (a) έως (r).

Τέταρτο Βήμα Βηματικής Παλινδρόμησης

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	(a)	2.046 E09	(g)	(i)	(j)
ERROR	(b)	(d)	(h)		
TOTAL	(c)	(e)			
R SQUARE (f)					
PARAMETER	ESTIMATE	T FOR H ₀ PARAMETER=0	PR> T	STD ERROR OF ESTIMATE	
INTERCEPT	-11618.450				
X ₂	64.685E-03	(k)	(o)	1.913	
D ₂	-3644.513	(l)	(p)	669.601	
D ₂ *D ₃ *X ₂	4.327	(m)	(q)	.989	
X ₁ *X ₂	-1.020	(n)	(r)	5.643 E-02	

(κ) Το πέμπτο και τελευταίο βήμα της βηματικής παλινδρόμησης στο μοντέλο αυτό δίνει την εκτύπωση που ακολουθεί. Να γράψετε την εξίσωση που αντιστοιχεί στο μοντέλο αυτό.

Πέμπτο Βήμα Βηματικής Παλινδρόμησης

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR>F
MODEL	5	2.059 E09	4.118 E08	386.34	.0001
ERROR	19	2.025 E07	1.066 E06		
TOTAL	24	2.079 E09			

		R SQUARE .9903		
PARAMETER	ESTIMATE	T FOR H ₀ PARAMETER=0	PR> T	STD ERROR OF ESTIMATE
INTERCEPT	-13507.748			
X ₃	286.798	3.46	.0026	82.931
X ₂	66.223E-03	41.38	.0001	1.600
D ₂	-3765.273	-6.98	.0001	539.317
D ₂ *D ₃ *X ₂	5.446	6.34	.0001	0.858
X ₁ *X ₂	-1.118	-20.89	.0001	5.350 E-02

(λ) Δώστε την ερμηνεία κάθε μιας από τις μεταβλητές του μοντέλου εξηγώντας ταυτόχρονα πώς αυτή επηρεάζει την τιμή πώλησης του διαμερίσματος.

(μ) Στην διαδικασία βηματικής παλινδρόμησης που ακολουθήσαμε σταματήσαμε μετά την εισαγωγή της μεταβλητής X₃ στο μοντέλο. Αυτό έγινε γιατί καμιά άλλη μεταβλητή δεν έδινε στατιστικά σημαντικό αποτέλεσμα σε επίπεδο σημαντικότητας $\alpha_1=0.05$, όπου α_1 είναι το επίπεδο σημαντικότητας, για να συμπεριληφθεί μια μεταβλητή στο μοντέλο. Αν κάποιος δεν γνωρίζει το μέγεθος του α_1 , ποιες είναι οι ενδείξεις που θα έπαιρνε για το μέγεθος του α_1 από τις διάφορες εκτυπώσεις που προηγήθηκαν;