

ΠΑΛΙΑ ΘΕΜΑΤΑ

(I) [100 μονάδες] Αναλύουμε 506 κοινότητες στην περιοχή της Βοστώνης για τις οποίες έχουμε τις ακόλουθες μεταβλητές:

1. LOG(PRICE) = log(διάμεσος από τις τιμές των κατοικιών των κοινοτήτων)
2. LOG(NOX) = log(ποσότητα διοξειδίου του αζώτου στην ατμόσφαιρα, σωματίδια ανά εκατομμύριο)
3. DIST = απόσταση σε πέντε εμπορικά κέντρα
4. ROOMS = μέσος αριθμός δωματίων σε κάθε σπίτι
5. STRATIO = μέση αναλογία φοιτητών ανά δάσκαλο στα σχολεία
6. CRIME = κατά κεφαλήν εγκλήματα
7. RADIAL = δείκτης πρόσβασης σε περιφερειακούς αυτοκινητόδρομους
8. PROPTAX = φόροι ιδιοκτησίας ανά \$1000
9. LOWSTAT = ποσοστό της χαμηλής τάξης (με χαμηλό εισόδημα)

Οι έξοδοι (output) των αποτελεσμάτων δύο παλινδρομήσεων και ενός F-test (Wald test) για τον έλεγχο περιορισμών αποκλεισμού:

Μοντέλο υπό Περιορισμούς (υπό δέσμευση - restricted)

Εξαρτημένη μεταβλητή: log(price)
 Παρατηρήσεις: 506

	Μεταβλητή	Εκτιμητές	T.Σφάλματα	t-τιμή	Pr(> t)	

β_0	Τετραγμένη	11.998755	0.258513	46.414	<2e-16	***
β_1	log(nox)	-0.635366	0.094981	-6.689	6.05e-11	***
β_2	log(dist)	-0.284785	0.032601	-8.735	<2e-16	***
β_3	rooms	0.100043	0.016297	6.139	1.70e-09	***
β_4	stratio	-0.037857	0.004828	-7.841	2.75e-14	***
β_5	crime	-0.012694	0.001348	-9.414	<2e-16	***
β_6	radial	0.015229	0.002603	5.850	8.94e-09	***
β_7	proptax	-0.007533	0.001348	-5.587	3.81e-08	***
β_8	lowstat	-0.028776	0.001875	-15.347	<2e-16	***

Άθροισμα τετραγώνων των καταλοίπων: 18.75 με 497 βαθμούς ελευθερίας (β.ε.)

R-τετράγωνο: 0.7783, Προσαρμοσμένο R-τετράγωνο: 0.7748

F-στατιστική: 218.1 με 8 και 497 β.ε., π-τιμή: < 2.2e-16

e-p = 10^{-p}, *** π-τιμή < 0.001, ** 0.001 < π-τιμή < 0.01, * 0.01 < π-τιμή < 0.05, • 0.05 < π-τιμή < 0.1

.....

Μοντέλο χωρίς Περιορισμούς (χωρίς δέσμευση - unrestricted)

Εξαρτημένη μεταβλητή: log(price)
 Παρατηρήσεις: 506

	Μεταβλητή	Εκτιμητές	Τ.Σφάλματα	t-τιμή	Pr(> t)	
γ ₀	Τεταγμένη	11.179117	0.219947	50.826	<2e-16	***
γ ₁	log(nox)	-0.365884	0.112879	-3.241	0.00127	**
γ ₂	log(dist)	-2.947737	0.395422	-7.455	4.09e-13	***
γ ₃	[log(dist)] ²	0.347868	0.208758	1.666	0.09628	.
γ ₄	rooms	1.386446	0.241305	5.746	1.61e-08	***
γ ₅	rooms ²	1.170978	0.202292	5.789	1.27e-08	***
γ ₆	stratio	-0.029146	0.004523	-6.444	2.77e-10	***
γ ₇	crime	-3.670875	0.311180	-11.797	<2e-16	***
γ ₈	crime ²	1.580078	0.272352	5.802	1.18e-08	***
γ ₉	radial	0.059914	0.014923	4.015	6.88e-05	***
γ ₁₀	proptax	-0.006975	0.001260	-5.537	5.02e-08	***
γ ₁₁	lowstat	-4.638712	0.287716	-16.123	<2e-16	***
γ ₁₂	lowstat ²	0.968187	0.217370	4.454	1.04e-05	***
γ ₁₃	log(nox)*radial	-0.021317	0.008040	-2.651	0.00827	**

Άθροισμα τετραγώνων των καταλοίπων: 15.01 με 492 βαθμούς ελευθερίας (β.ε.)

R-τετράγωνο: 0.8224, Προσαρμοσμένο R-τετράγωνο: 0.8177

F-στατιστική: 175.3 με 13 και 492 β.ε., π-τιμή: < 2.2e-16

e-p = 10^{-p}, *** π-τιμή < 0.001, ** 0.001 < π-τιμή < 0.01, * 0.01 < π-τιμή < 0.05, • 0.05 < π-τιμή < 0.1

F-test (Wald test) για τον έλεγχο περιορισμών αποκλεισμού:

H₀: γ₃ = γ₅ = γ₈ = γ₁₂ = γ₁₃=0

H₁: η H₀ δεν ισχύει

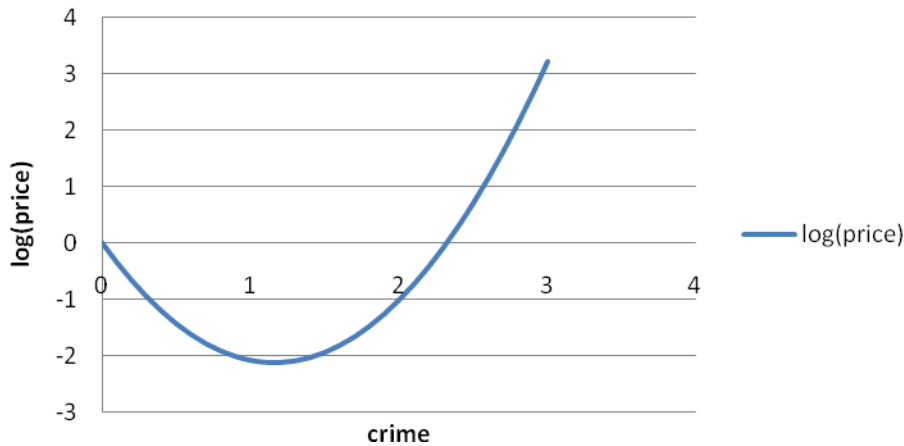
F-test = **24.445** με 5 και 492 βαθμούς ελευθερίας

π-τιμή < 0.01

- (1) [5 μονάδες] Γράψτε την εξίσωση της παλινδρόμησης για το μοντέλο με περιορισμούς (όπως διατυπώνεται στα παραδείγματα των σημειώσεων και του βιβλίου).
- (2) [5 μονάδες] Σχολιάστε κατά πόσο το μοντέλο χωρίς περιορισμούς είναι ικανοποιητικό. Αναφέρεται τα κριτήρια τα οποία χρησιμοποιείται.
- (3) [5 μονάδες] Στο μοντέλο με περιορισμούς, αιτιολογήστε αν οι μεταβλητές έχουν τα κατάλληλα πρόσημα; Αν όχι, προσπαθήστε να δώσετε κάποια λογική εξήγηση.
- (4) [5 μονάδες] Ποιές μεταβλητές είναι στατιστικά σημαντικές σε επίπεδο σημαντικότητας 0.05 και 0.1;
- (5) [5 μονάδες] Αναφέρεται αν από το μοντέλο απουσιάζουν κάποιες σημαντικές μεταβλητές και ποιες είναι αυτές. Γενικά τι πρόβλημα υπάρχει όταν δεν περιλαμβάνουμε στο μοντέλο μας κάποιες βασικές μεταβλητές.
- (6) [5 μονάδες] Πως προκύπτει η t-τιμή 1.666 που αντιστοιχεί στην μεταβλητή $[\log(dist)]^2$, στο μοντέλο χωρίς περιορισμούς.
- (7) [10 μονάδες] Βρείτε την πλησιέστερη τιμή, χρησιμοποιώντας τον κατάλληλο πίνακα, της t-τιμής 1.666 που αντιστοιχεί στην μεταβλητή $[\log(dist)]^2$, στο μοντέλο χωρίς περιορισμούς.
- (8) [5 μονάδες] Στο μοντέλο με περιορισμούς, πόσο είναι η ελαστικότητα της τιμής των κατοικιών προς την ποσότητα διοξειδίου του αζώτου στην ατμόσφαιρα και ποια είναι η ερμηνεία της;
- (9) [10 μονάδες] Στο μοντέλο με περιορισμούς, ελέγχουμε την υπόθεση: $H_0: \beta_1 = -0.5$ και $H_1: \beta_1 \neq -0.5$, εξηγήστε τι στην ουσία ελέγχουμε και εκτελέστε τον έλεγχο. Τι συμπεραίνεται από τον έλεγχο;
- (10) [10 μονάδες] Στο μοντέλο με περιορισμούς, βρείτε την ακριβή ποσοστιαία μεταβολή της *price* όταν μεταβάλλεται η μεταβλητή *rooms* κατά δύο μονάδες και οι άλλες παραμένουν σταθερές, χρησιμοποιώντας τον τύπο: $\% \Delta price = [\exp(\beta_3 \Delta rooms) - 1]$. Διαφέρει πολύ από την ποσοστιαία μεταβολή που προκύπτει άμεσα από το μοντέλο;
- (11) [10 μονάδες] Στο μοντέλο με περιορισμούς, εξηγήστε πως μεταβάλλεται η μεταβλητή *price* όταν μεταβάλλεται η μεταβλητή *crime* από 2.2 σε 2.6 και οι άλλες μεταβλητές παραμένουν σταθερές, χρησιμοποιώντας έναν μόνο τύπο από τους ακόλουθους τύπους:
 i) $\Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * crime] \Delta crime$
 ii) $\% \Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * crime] \% \Delta crime$
 iii) $\% \Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * crime] \Delta crime$
 iv) $\Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * crime] \% \Delta crime$
- (12) [10 μονάδες] Στο μοντέλο χωρίς περιορισμούς, εξηγήστε πως μεταβάλλεται η μεταβλητή *price* όταν μεταβάλλεται η μεταβλητή *dist* από 3.5 κατά 6% και οι άλλες μεταβλητές παραμένουν σταθερές, χρησιμοποιώντας έναν μόνο τύπο από τους ακόλουθους τύπους:
 i) $\Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * \log(dist)] \Delta dist$
 ii) $\% \Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * \log(dist)] \% \Delta dist$
 iii) $\% \Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * \log(dist)] \Delta dist$
 iv) $\Delta price \approx [\text{συντελεστής πρωτοβάθμιου όρου} + 2 * (\text{συντελεστής δευτεροβάθμιου όρου}) * \log(dist)] \% \Delta dist$

(13) [10 μονάδες] Βρείτε το ελάχιστο της συνάρτησης $\log(\text{price}) = \gamma_7 \text{ crime} + \gamma_8 \text{ crime}^2$ όπως φαίνεται στο παρακάτω σχήμα. Αν δεν έχετε τον τύπο βρείτε που μηδενίζεται η πρώτη παράγωγο. Τι πληροφορία μας δίνει αυτό το σημείο; Συμφωνείται με αυτή την σχέση μεταξύ $\log(\text{price})$ και crime ; Είναι η αναμενόμενη; Αν δεν είναι η αναμενόμενη σε κάποια διάστημα της crime , δώστε μία πιθανή λογική εξήγηση;

Σχέση $\log(\text{price})$ με crime



(14) [10 μονάδες] Στο μοντέλο χωρίς περιορισμούς, εξηγήστε πως μεταβάλλεται η μεταβλητή price όταν μεταβάλλεται η μεταβλητή nox κατά 3%, η μεταβλητή radial παίρνει την τιμή 5, και οι άλλες παραμένουν σταθερές, χρησιμοποιώντας τον τύπο: $\% \Delta \text{price} = [\gamma_1 + \gamma_{13} \text{radial}] \% \Delta \text{nox}$. Πως προκύπτει αυτός ο τύπος;

(15) [10 μονάδες] Πως υπολογίζεται η τιμή της F στατιστικής 24.445 για τον έλεγχο περιορισμών αποκλεισμού που δίνεται στην έξοδο του υπολογιστή (output). Γράψτε τον κατάλληλο τύπο και αντικαταστήστε τα σωστά νούμερα.

(16) [5 μονάδες] Στην παραπάνω F στατιστική για τον έλεγχο περιορισμών αποκλεισμού γιατί οι βαθμοί ελευθερίας είναι 5 και 492.

(17) [10 μονάδες] Σε συνέχεια της προηγούμενης ερώτησης, εξηγήστε πως προκύπτει ότι η p -τιμή είναι < 0.01 . Από ποιον πίνακα και ποιο είναι το κοντινότερο σημείο του πίνακα στο οποίο βασιζόμαστε;

(18) [5 μονάδες] Τι συμπέρασμα βγάζετε από την παραπάνω F στατιστική σχετικά με την γραμμικότητα του μοντέλου;

(1) [20 μονάδες] Περιγράψτε την στατιστική του πολλαπλασιαστή του Lagrange (με τα βιβλία του κ. Χρήστου – Στατιστική X^2). Τι ελέγχουμε με αυτήν την στατιστική; Δώστε τα πέντε βασικά βήματα ενός ελέγχου υποθέσεων (H_0 , H_1 , τεστ, κατανομή του τεστ, περιοχή απόρριψης, και συμπέρασμα).

(2) [10 μονάδες] Γράψτε μόνο τις γενικές ιδιότητες της διακύμανσης, και της διακύμανσης, για γραμμικούς συνδυασμούς, για διακριτές μεταβλητές.

(3) [20 μονάδες] Θεωρήστε την γραμμική παλινδρόμηση, $Y = b_0 + b_1X_1 + b_2X_2$, με τις οικονομικές μεταβλητές Y , X_1 , και X_2 .

Αν $\text{Var}\{X_1\}=1$, $\text{Var}\{X_2\}=3.7$, και $\text{Cov}\{X_1, X_2\}=1.75$, γράψτε τις ακόλουθες ποσότητες σε σχέση με τους συντελεστές (εξηγήστε αναλυτικά, ποιες ιδιότητες και πως τις χρησιμοποιείται):

(α) [10 μονάδες] $\text{Var}\{Y\} =$

(β) [10 μονάδες] $\text{Cov}\{Y, X_2\} =$

(4) [80 μονάδες] Θεωρούμε το υπόδειγμα (μοντέλο) του ερωτήματος (2) και εφαρμόζουμε την μεθοδολογία του ερωτήματος (1) για τον έλεγχο σημαντικότητας της μεταβλητής X_2 στο υπόδειγμα. Παρόλο που αυτός έλεγχος μπορεί να πραγματοποιηθεί με απλούστερα τεστ, π.χ. t-τεστ, καλείστε να εφαρμόσετε την στατιστική του πολλαπλασιαστή του Lagrange.

Οι τιμές των μεταβλητών δίνονται στον επόμενο πίνακα.

Πινάκας 1.

Y	-1	0	0	1
X ₁	-1	-1	1	1
X ₂	-2	-1	1	3

Τα αποτελέσματα των δύο παλινδρομήσεων της στατιστικής του πολλαπλασιαστή του Lagrange καταγράφονται στους παρακάτω δύο πίνακες στην επόμενη σελίδα.

(α) [5 μονάδες] Ποιες μεταβλητές παλινδρομούμε στην δεύτερη παλινδρόμηση;

(β) [5 μονάδες] Ερμηνεύστε την κλίση στην πρώτη παλινδρόμηση;

Πινάκας 2. Παλινδρόμηση 1

R^2 (1)=
Προσαρμοσμένο
(Διορθωμένο) R^2 (2)=

ANOVA

	df (β.ε.)	SS (Α.Τ.)	MS (Μ.Τ.)	F-τεστ	π-τιμή
Παλινδρόμηση	1	1	----	(5)=	----
Κατάλοιπα	2	(4)=	----		
Σύνολο	3	(3)=			

	Συντελεστές	Τυπικό Σφάλμα	t-τεστ	π-τιμή	Κάτω φράγμα Δ.Ε. 95%	Άνω φράγμα Δ.Ε. 95%
Τεταγμένη της αρχής	(6)=	0,35	0,00	1,00	-1,52	1,52
x ₁	0,50	0,35	(7)=	(8)=	-1,02	2,02

Κατάλοιπα

Παρατήρηση	Προβλέψεις	Κατάλοιπα
1	(9)=	(10)=
2	-0,5	0,5
3	0,5	-0,5

Πινάκας 3. Παλινδρόμηση 2

R ²	0,901
Προσαρμοσμένο (Διορθωμένο) R ²	0,70

ANOVA

	<i>df (β.ε.)</i>	<i>SS (A.T.)</i>	<i>MS (M.T.)</i>	<i>F-τεστ</i>	<i>π-τιμή</i>
Παλινδρόμηση	2	3,6	1,8	4,5	0,32
Κατάλοιπα	1	0,4	0,4		
Σύνολο	3	4			

	<i>Συντελεστές</i>	<i>Τυπικό Σφάλμα</i>	<i>t-τεστ</i>	<i>π-τιμή</i>	<i>Κάτω φράγμα Δ.Ε. 95%</i>	<i>Άνω φράγμα Δ.Ε. 95%</i>
Τεταγμένη της αρχής	-0,3	0,33	-0,90	0,53	-4,51	3,91
x1	-2,1	0,77	-2,73	0,22	-11,86	7,66
x2	1,2	0,40	3,00	0,20	-3,88	6,28

(γ) [50 μονάδες] Συμπληρώστε τα 10 ζητούμενα νούμερα στον Πίνακα 2, αριθμούμενα ως **(1)**-**(10)**. Για την π-τιμή, (8), δώστε ένα κάτω φράγμα. Γράψτε παρακάτω τους τύπους που χρησιμοποιήσατε:

(1)

(2)

(3)

(4)

(5)

(6)

(7)

(8)

(9)

(10)

(δ) [20 μονάδες] Ολοκληρώστε την στατιστική του πολλαπλασιαστή του Lagrange (με τα βιβλία του κ. Χρήστου – Στατιστική X^2). Δώστε τα πέντε βασικά βήματα ενός ελέγχου υποθέσεων (H_0 , H_1 , τεστ, κατανομή του τεστ, περιοχή απόρριψης, και συμπέρασμα)

*****8

1) [105 μονάδες] Ο πίνακας που ακολουθεί δημιουργήθηκε χρησιμοποιώντας πραγματικά δεδομένα:

Εξαρτημένη μεταβλητή: $\log(\text{μισθός})$			
Ανεξάρτητες Μεταβλητές	Μοντέλο (1)	Μοντέλο (2)	Μοντέλο (3)
$\log(\text{πωλήσεις})$	0,224 (0,027)	0,158 (0,040)	0,188 (0,040)
$\log(\text{αξία})$	_____	0,112 (0,050)	0,100 (0,049)
κέρδη	_____	-0,0023 (0,0022)	0,0022 (0,0021)
προϋπηρεσία_1	_____	_____	0,0171 (0,0055)
προϋπηρεσία_2	_____	_____	-0,0092 (0,0033)
τεταγμένη της αρχής	4,94 (0,20)	4,62 (0,25)	4,57 (0,25)
R-τετράγωνο	0,281	0,304	0,353
Αριθμός Παρατηρήσεων = $n = 177$			

Οι μεταβλητές ορίζονται ως εξής:

μισθός = ο μισθός του γενικού διευθυντή της εταιρίας

πωλήσεις = οι πωλήσεις της εταιρίας

αξία = η αξία της επιχείρησης στην αγορά

κέρδη = είναι τα κέρδη ως ποσοστό των πωλήσεων

προϋπηρεσία_1 = είναι τα έτη στη θέση του γενικού διευθυντή στην εταιρεία

προϋπηρεσία_2 = είναι το σύνολο των ετών της παραμονής του διευθυντή στην εταιρία

Σημειώστε ότι σε παρένθεση αναφέρονται τα τυπικά σφάλματα των εκτιμητών:

(α) [10 μονάδες] Σχολιάστε την επίδραση της μεταβλητής αξία στο μισθό του γενικού διευθυντή. Ερμηνεύστε τον συντελεστή της. Εξηγήστε την απάντησή σας στατιστικά, με το κατάλληλο τεστ.

(β) [10 μονάδες] Σχολιάστε την επίδραση της μεταβλητής κέρδη στο μισθό του γενικού διευθυντή. Ερμηνεύστε τον συντελεστή της. Εξηγήστε την απάντησή σας στατιστικά, με το κατάλληλο τεστ.

(γ) [20 μονάδες] Σχολιάστε την επίδραση των μεταβλητών *προϋπηρεσία_1* και *προϋπηρεσία_2* στο μισθό του γενικού διευθυντή. Ερμηνεύστε τους συντελεστές αυτών. Εξετάστε ταυτοχρόνως την στατιστική σημαντικότητα των προαναφερθέντων συντελεστών, με το κατάλληλο τεστ.

(δ) [10 μονάδες] Τι συμπεραίνετε για το γεγονός ότι η πιο πολυετής παραμονή στην ίδια εταιρία, όταν διατηρήσουμε σταθερούς τους άλλους παράγοντες, συνδέεται με ένα χαμηλότερο μισθό;

(ε) [15 μονάδες] Χρησιμοποιώντας την Στατιστική *F* συγκρίνεται τα μοντέλα (1) και (2). Αναφέρεται όλα τα σχετικά: μηδενική και εναλλακτική υπόθεση, κριτήριο, και συμπέρασμα.

(στ) [10 μονάδες] Από τα παραπάνω αποτελέσματα, πιο από τα μοντέλα (1), (2) και (3) θα θεωρούσατε ως το καλύτερο; Συμπεραίνετε το ίδιο από τις τιμές των R-τετραγώνων; Παρακαλώ αιτιολογήστε το.

(ζ) [10 μονάδες] Από το μοντέλο (3) θα απομακρύνεται κάποια από τις μεταβλητές. Γενικά η παραμονή μιας ασήμαντης μεταβλητής, δημιουργεί προβλήματα στις τιμές των άλλων μεταβλητών;

(η) [10 μονάδες] Από το μοντέλο (3) πιστεύεται ότι απουσιάζει μια σημαντική μεταβλητή; Αν ναι, αναφέρεται την μεταβλητή που θεωρείτε σημαντική. Γενικά η απουσία μιας σημαντικής μεταβλητής, δημιουργεί προβλήματα στις τιμές των άλλων μεταβλητών;

(θ) [10 μονάδες] Περιληπτικά αναφέρεται ποια είναι τα βασικά συμπεράσματα σας από την παραπάνω μελέτη.

2) [10 μονάδες] Πότε ένας εκτιμητής καλείται αμερόληπτος και πότε συνεπής. Μπορεί ένας εκτιμητής να είναι και αμερόληπτος και συνεπής;

3) [10 μονάδες] Σε τι διαφέρουν τα Ενοποιημένα ή Σφαιρικά (Πάνελ) δεδομένα από τις Χρονοσειρές;

4) [15 μονάδες] Στον παρακάτω τύπο σχολιάστε πως μεταβάλλεται η διακύμανση του εκτιμητή της κλίσης ως προς τη μεταβολή σχετικών ποσοτήτων, αιτιολογώντας τις μεταβολές αυτές.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}, \text{ όπου}$$

$SST_j = \sum (x_{ij} - \bar{x}_j)^2$ και R_j^2 είναι το R^2 από την παλινδρόμηση της x_j επί όλων των άλλων x 's

5) [10 μονάδες] Αποδείξτε και εξηγήστε αναλυτικά ότι: $\sum_{i=1}^n (x_i - \bar{x})x_i = \sum_{i=1}^n (x_i - \bar{x})^2$

6) [20 μονάδες] Αποδείξτε και εξηγήστε αναλυτικά ότι:

$$\text{Var} \left\{ \sum_{i=1}^n (c_i - \bar{c})^2 + \sum_{i=1}^n (c_i - \bar{c})(z_i - \mu_z) \right\} = \sigma_z^2 \sum_{i=1}^n (c_i - \bar{c})^2$$

όπου τα c_i είναι σταθερές και z_i είναι ανεξάρτητες τυχαίες μεταβλητές με αναμενόμενη τιμή μ_z και διακύμανση σ_z^2 .

(7) [60 μονάδες] Εκτελώντας γραμμική παλινδρόμηση στις οικονομικές μεταβλητές, Y και X, οι τιμές των οποίων δίνονται στον επόμενο πίνακα, υπολογίστε τις ζητούμενες ποσότητες. $Y = b_0 + b_1 X$.

Y	-3	-1	0	4
X	-3	-2	1	3

(α) [15 μονάδες] Υπολογίστε τον συντελεστή του X, b_1 .

(β) [5 μονάδες] Προβλέψτε την τιμή του Y για $X=-1.5$.

(γ) [10 μονάδες] Εκτιμήστε τα κατάλοιπα, δηλαδή u .

(δ) [10 μονάδες] Εκτιμήστε την διακύμανση των καταλοίπων, $\hat{\sigma}^2 = S^2$.

(ε) [10 μονάδες] Να ελεγχθεί η υπόθεση $H_0: \beta_1 = 0$ έναντι της $H_e: \beta_1 \neq 0$, σε επίπεδο σημαντικότητας $\alpha=0.05$, και αναφέρετε τα συμπεράσματα σας.

(στ) [10 μονάδες] Υπολογίστε τον συντελεστή R^2 και ερμηνεύστε το μέγεθος του.

8) [20 μονάδες] Αποδείξτε και εξηγήστε αναλυτικά ότι:

α) [10 μονάδες] $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i$

β) [10 μονάδες] $\sum_{i=1}^n (x_i - \bar{x})y_i = \sum_{i=1}^n x_i(y_i - \bar{y})$

9) [20 μονάδες] Αν $n=3$ όπου τα a_i είναι σταθερές με τιμές $a_1=-1$, $a_2=0$, και $a_3=1$, και αν x_1, x_2 , και x_3 είναι ανεξάρτητες τυχαίες μεταβλητές με αναμενόμενη τιμή $\mu_x=10.3$ και διακύμανση $\sigma_x^2 = 0.5$, δείξτε και εξηγήστε αναλυτικά ότι:

$$\text{Var} \left\{ \sum_{i=1}^3 (a_i - \bar{a})^2 + \sum_{i=1}^3 (a_i - \bar{a})(x_i - \mu_x) \right\} = 1$$

10) [40 μονάδες] Θεωρήστε το υπόδειγμα

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

και τον εκτιμητή $\tilde{\beta}_1 = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1) y_i}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$, ο οποίος δεν είναι ο εκτιμητής ελαχίστων τετραγώνων.

α) [15 μονάδες] Δείξτε ότι

$$\sum_{i=1}^n (x_{i1} - \bar{x}_1) y_i = \beta_1 \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1) x_{i2} + \sum_{i=1}^n (x_{i1} - \bar{x}_1) u_i$$

β) [15 μονάδες] Χρησιμοποιώντας την σχέση στην α) δείξτε ότι

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1) x_{i2}}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$$

γ) [10 μονάδες] Εξηγήστε ποτέ ο εκτιμητής $\tilde{\beta}_1$ είναι αμερόληπτος και συνεπής. Δώστε τον τύπο της μεροληψίας του $\tilde{\beta}_1$.

11) [30 μονάδες] Απαντήστε τις ακόλουθες ερωτήσεις.

α) [10 μονάδες] Πως θα εξετάζατε γραφικά και αριθμητικά, με απλό τρόπο, στην απλή παλινδρόμηση αν το μοντέλο είναι γραμμικό;

β) [10 μονάδες] Αν υπάρχει μη γραμμική σχέση στην απλή παλινδρόμηση, αναφέρεται δύο τρόπους κατάλληλους για μη γραμμική παλινδρόμηση. Πως θα εξετάζατε ότι αυτοί οι δύο τρόποι θα βελτίωναν το υπόδειγμα σας.

γ) [10 μονάδες] Η πολυσυγγραμμικότητα είναι πλεονέκτημα η μειονέκτημα και γιατί; Μπορεί να υπάρχει πολυσυγγραμμικότητα στην απλή παλινδρόμηση; Εξηγήστε γιατί; Πως θα εξετάζατε, με απλό τρόπο, την ύπαρξη πολυσυγγραμμικότητας;

12) [35 μονάδες] Ελέγξτε αν υπάρχουν διαφορές στις συναρτήσεις παλινδρόμησης ανάμεσα στις χρονικές περιόδους 1970-1981 και 1982-1995, εκτελώντας το κατάλληλο στατιστικό τεστ.

Πίνακας. Καταθέσεις και προσωπικά εισοδήματα (δισ. \$) ΗΠΑ, 1970-1995

A/A	Έτος	Καταθέσεις	Εισόδημα	Έτος	Καταθέσεις	Εισόδημα
1	1970	61	727.1	1982	205.5	2347.3
2	1971	68.6	790.2	1983	167	2522.4
3	1972	63.6	855.3	1984	235.7	2810
4	1973	89.6	965	1985	206.2	3002
5	1974	97.6	1054.2	1986	196.5	3187.6
6	1975	104.4	1159.2	1987	168.4	3363.1
7	1976	96.4	1273	1988	189.1	3640.8
8	1977	92.5	1401.4	1989	187.8	3894.5
9	1978	112.6	1580.1	1990	208.7	4166.8
10	1979	130.1	1769.5	1991	246.4	4343.7
11	1980	161.8	1973.3	1992	272.6	4613.7
12	1981	199.1	2200.2	1993	214.4	4790.2
13				1994	189.4	5021.7
14				1995	249.3	5320.8

Παλινδρομώντας τις καταθέσεις (y) στα εισοδήματα (x) στις παρακάτω χρονικές περιόδους παίρνουμε:

A) για 1970-1981, $R^2=0.9021$, $SST=?$ (υπολογίστε το)

B) για 1982-1995, $R^2=0.2971$, $SST=14234.2$

Γ) για 1970-1995, $R^2=0.7672$, $SST=99863.8$

Υπόδειξη: Μπορείτε να χρησιμοποιήσετε τον τύπο του R^2

13) [10 μονάδες] Υπολογίστε το προσαρμοσμένο R-τετράγωνο σε ένα μοντέλο παλινδρόμησης με $R^2=0.7$, με 30 παρατηρήσεις, και όταν το μοντέλο περιέχει

α) μία μεταβλητή

β) πέντε μεταβλητές

γ) δέκα μεταβλητές

Υπάρχει περίπτωση το R^2 να είναι πολύ μικρότερο από το προσαρμοσμένο R-τετράγωνο. Εξηγήστε περιληπτικά. Αν ναι, δώστε ένα αριθμητικό παράδειγμα.

14) [10 μονάδες] Προβλέψτε την τιμή του y για $x=1.5$ στο παρακάτω εκτιμώμενο μοντέλο

$$1 \hat{\sigma} g(y) = 0.5 + 1.2x$$

στο οποίο τα κατάλοιπα ακολουθούν κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση 0.8.

15) [40 μονάδες] Θεωρούμε μετρήσεις ύψη (σε ίντσες) 28 φοιτητριών και 80 φοιτητών (ύψος). Ένα υπόδειγμα παλινδρόμησης του ύψους σε μία τεταγμένη της αρχής και σε μία

ψευδομεταβλητή (φοιτήτρια), η οποία παίρνει την τιμή 1 για φοιτήτριες και την τιμή 0 για φοιτητές, δίνει τα ακόλουθα αποτελέσματα:

$$\widehat{ύψος} = 72.0 - 4.83 \times \text{φοιτήτρια}, \quad R^2 = 0.45,$$

(0.3) (0.56)

Στις παρενθέσεις δίνονται τα τυπικά σφάλματα

(α) [5 μονάδες] Ερμηνεύστε την τεταγμένη της αρχής.

(β) [10 μονάδες] Ερμηνεύστε την κλίση.

(γ) [5 μονάδες] Ερμηνεύστε το R^2 .

(δ) [10 μονάδες] Ελέγξτε τη υπόθεση ότι οι φοιτήτριες, κατά μέσο όρο, είναι πιο κοντές από τους φοιτητές, με επίπεδο σημαντικότητας, 1%.

(ε) [10 μονάδες] Θεωρείτε πιθανό ότι ο όρος των σφαλμάτων είναι ομοσκεδαστικό; Εξηγήστε γιατί.