

ΠΑΛΑΙΑ ΘΕΜΑΤΑ

(I) [50 μονάδες] Αναλύουμε δεδομένα από τις Η.Π.Α. σχετικά με την καθημερινή κατανάλωση τσιγάρων με την μέθοδο της παλινδρόμησης. Η εξαρτημένη μεταβλητή, *cigs* είναι ο αριθμός των τσιγάρων που καπνίζονται ανά ημέρα. Οι ανεξάρτητες μεταβλητές έχουν ως εξής:

το *income* (εισόδημα) είναι το ετήσιο εισόδημα,

η *cigpric* είναι η τιμή ενός πακέτου τσιγάρων (σε σεντ),

η *educ* (μόρφωση) είναι έτη εκπαίδευσης,

η *age* (ηλικία) μετριέται σε έτη,

και η *restaurn* (εστιατόρια) είναι μία ψευδομεταβλητή με:

1 = εάν το πρόσωπο κατοικεί σε πολιτεία με περιορισμούς για το κάπνισμα στα εστιατόρια και

0 = εάν το πρόσωπο κατοικεί σε πολιτεία χωρίς περιορισμούς για το κάπνισμα στα εστιατόρια.

Εκτιμούμε το παρακάτω οικονομετρικό υπόδειγμα με την μέθοδο των συνήθη ελαχίστων τετραγώνων (OLS). Στις παρενθέσεις βρίσκονται τα τυπικά σφάλματα, ακριβώς κάτω από τις εκτιμήσεις των παραμέτρων.

$$(1) \quad \begin{aligned} \text{cigs} = & -3.64 + .880 \log(\text{income}) - .751 \log(\text{cigpric}) \\ & (24.08) \quad (.728) \quad (5.773) \\ & -.501 \text{educ} + .771 \text{age} - .0090 \text{age}^2 - 2.83 \text{restaurn} \\ & (.167) \quad (.160) \quad (.0017) \quad (1.11) \end{aligned}$$

$$n = 807, R^2 = .0526,$$

Αν u είναι τα κατάλοιπα από την προηγούμενη παλινδρόμηση, εκτελούμε την ακόλουθη παλινδρόμηση:

$$(2) \quad \begin{aligned} \log(u^2) = & -1.92 + .29 \log(\text{income}) + .195 \log(\text{cigpric}) \\ & (2.56) \quad (.774) \quad (.615) \\ & -.080 \text{educ} + .204 \text{age} - .00239 \text{age}^2 - 0.627 \text{restaurn} \\ & (.018) \quad (.017) \quad (.00019) \quad (.118) \end{aligned}$$

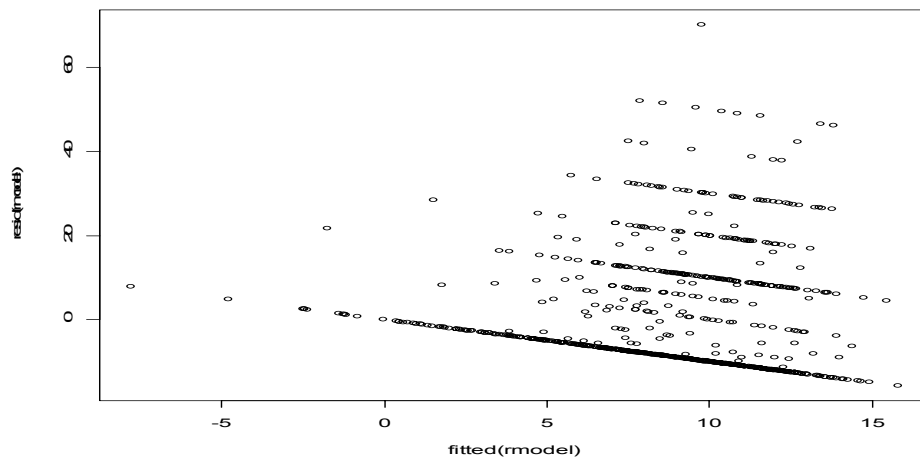
$$n = 807, R^2 = .0324,$$

Έπειτα, εφαρμόζουμε την διαδικασία εφικτού γενικευμένου εκτιμητή των ελαχίστων τετραγώνων για την ετεροσκεδαστικότητα (με τα βιβλία του κ. Χρήστου: εκτίμηση

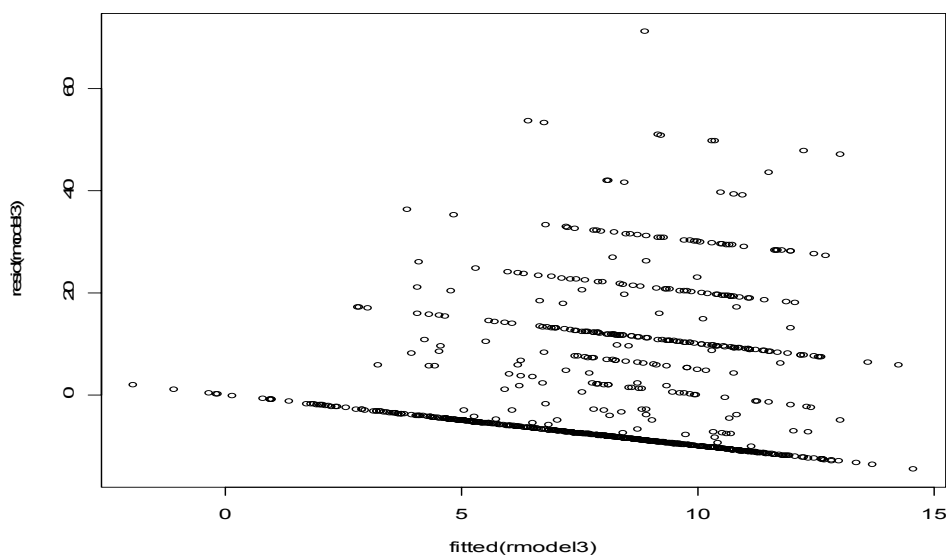
του υποδείγματος όταν εκτιμάμε την συνάρτηση ετεροσκεδαστικότητας) και παίρνουμε τα εξής νέα αποτελέσματα.

$$\begin{aligned}
 (3) \quad & \text{cigs} = 5.64 + 1.30 \log(\text{income}) - 2.94 \log(\text{cigpric}) \\
 & \quad (17.80) \quad (.44) \quad (4.46) \\
 & - .463 \text{educ} + .482 \text{age} - .0056 \text{age}^2 - 3.46 \text{restaurn} \\
 & \quad (.120) \quad (.097) \quad (.0009) \quad (.80) \\
 & n = 807, R^2 = .1134.
 \end{aligned}$$

Σχήμα Ι. Κατάλοιπα (y) έναντι Προσαρμοσμένων Τιμών (x) για το Μοντέλο (1)



Σχήμα ΙΙ. Κατάλοιπα (y) έναντι Προσαρμοσμένων Τιμών (x) για το Μοντέλο (2)



(1) [5 μονάδες] Από τα σχήματα I και II, υπάρχει ετεροσκεδαστικότητα πριν και μετά την άρση της ετεροσκεδαστικότητας, αντίστοιχα; Αιτιολογήστε το συμπέρασμα σας.

(2) [5 μονάδες] Συγκρίνοντας τα μοντέλα (1) και (3), για ποια ή για ποιες μεταβλητές αλλάζει η σημαντικότητα;

(3) [10 μονάδες] Χρησιμοποιώντας τα αποτελέσματα από την Παλινδρόμηση (2), γράψτε τις σταθμίσεις τις οποίες χρησιμοποιούμε στην Παλινδρόμηση (3).

(4) [5 μονάδες] Πόσο λιγότερο ή περισσότερο καπνίζουν οι απόφοιτοι Πανεπιστημίου από τους απόφοιτους λυκείου;

(5) [5 μονάδες] Από ποια ηλικία και μετά ελαττώνεται το κάπνισμα, βάση της Παλινδρόμησης (1). Αιτιολογήστε αριθμητικά την απάντησή σας.

(6) [10 μονάδες] Οι περιορισμοί για το κάπνισμα στα εστιατόρια επηρεάζουν σημαντικά στην ελάττωση του καπνίσματος; Εκτελεστεί τον κατάλληλο έλεγχο.

(7) [10 μονάδες] Εκτελώντας την παλινδρόμηση $\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \text{σφάλμα μας}$ δίνει $R^2=0.0324$. Εκτελέστε τον έλεγχο του White για ετεροσκεδαστικότητα. Χρησιμοποιήστε τον κατάλληλο πίνακα, βρείτε το πιο χρήσιμο σημείο και γράψτε το συμπέρασμα σας βάση αυτού.

(II) [60 μονάδες] Στις αρχές της δεκαετίας του '80, οι αμερικάνικες εταιρίες παραγωγής χλωριούχου βαρίου θεώρησαν ότι η Κίνα παρείχε στις Η.Π.Α. χλωριούχο βάριο σε απίστευτα χαμηλή τιμή η οποία ήταν κάτω της τιμής του κόστους ή κάτω της τιμής πώλησης στην Κίνα. Η διαδικασία αυτή στα διεθνή οικονομικά καλείται *dumping*. Επίσης *dumping* καλείται κάθε είδους «αρπαχτής» στην αγορά. Οι αμερικάνικες εταιρίες παραγωγής χλωριούχου βαρίου παραπονέθηκαν στην αμερικάνικη Επιτροπή Διεθνούς Εμπορίου τον Οκτώβριο του 1983 με μία καταγγελία κατά των Κινέζικων εταιριών. Η επιτροπή αυτή αποφάσισε υπέρ των αμερικάνικων εταιριών τον Οκτώβριο του 1984. Αναλύουμε μηνιαία δεδομένα από τον Φεβρουάριο του 1978 μέχρι τον Δεκέμβριο του 1988, παλινδρομώντας τις εξής μεταβλητές.

chnimp = ο όγκος των εισαγωγών του χλωριδίου βαρίου από την Κίνα

chempi = ένας δείκτης χημικής παραγωγής – μεταβλητή έλεγχου για την ζήτηση χλωριδίου βαρίου στην αγορά

gas = ο όγκος της παραγωγής της βενζίνης – δεύτερη μεταβλητή έλεγχου για την ζήτηση στην αγορά

rtwex = ένας δείκτης συναλλαγματικής ισοτιμίας, ο οποίος μετρά τη δύναμη του δολαρίου έναντι διάφορων άλλων νομισμάτων

befile6 = ψευδομεταβλητή, 1= για τους έξι μήνες πριν την καταγγελία και 0= για τους υπόλοιπους μήνες

affile6 = ψευδομεταβλητή, 1= για τους έξι μήνες μετά την καταγγελία και 0= για τους υπόλοιπους μήνες

afdec6 = ψευδομεταβλητή, 1= για τους έξι μήνες μετά την απόφαση της επιτροπής και 0= για τους υπόλοιπους μήνες

Η εξίσωση της παλινδρόμησης είναι η εξής:

$$\log(\text{chnimp})_t = \beta_0 + \beta_1 \log(\text{chempi})_t + \beta_2 \log(\text{gas})_t + \beta_3 \log(\text{rfwex})_t + \beta_4 \text{befile6}_t + \beta_5 \text{affile6}_t + \beta_6 \text{afdec6}_t + u_t$$

Πίνακας. Εξαρτημένη Μεταβλητή: $\log(\text{chnimp})$ [Εισαγωγές από την Κίνα]

Συντελεστής	Μέθοδος Ελαχίστων τετραγώνων	Εκτίμηση κατά Cochrane-Orcutt
$\log(\text{chempi})$	3.12 (0.48)	2.95 (0.65)
$\log(\text{gas})$.196 (.907)	1.05 (0.99)
$\log(\text{rtwex})$.983 (.400)	1.14 (0.51)
befile6	.060 (.261)	- .016 (.321)
affile6	- .032 (.264)	- .033 (.323)
afdec6	- .565 (.286)	- .577 (.343)
Τεταγμένη της αρχής	- 17.80 (20.05)	- 37.31 (23.22)
$\hat{\rho}$	—	- .293 (.084)
Παρατηρήσεις R -τετράγωνο	131 .305	130 .193

Στον παραπάνω πίνακα δίνονται οι εκτιμητές ελαχίστων τετραγώνων (OLS) και οι εκτιμητές με την μέθοδο Cochrane-Orcutt. Στις παρενθέσεις βρίσκονται τα τυπικά σφάλματα (ακριβώς κάτω από τις εκτιμήσεις των παραμέτρων).

Το χλωριούχο βάριο χρησιμοποιείτε ως υλικό σε καθαριστικά και για την παραγωγή βενζίνης. Απαντήστε στα παρακάτω ερωτήματα.

(1) [5 μονάδες] Συμπληρώστε τα κενά στην παρακάτω εξίσωση για την εκτίμηση κατά Cochrane-Orcutt.

$$\log(\text{chnimp}) = \frac{\quad}{(\quad)} + \frac{\quad}{(\quad)} \log(\text{chempi}) + \frac{\quad}{(\quad)} \log(\text{gas})$$

$$+ \frac{\quad}{(\quad)} \log(\text{rfwex}) + \frac{\quad}{(\quad)} \text{befile6} + \frac{\quad}{(\quad)} \text{affile6} + \frac{\quad}{(\quad)} \text{afdec6}$$

$$n = \quad, \quad R^2 = \quad$$

(2) [10 μονάδες] Είναι οι εισαγωγές από την Κίνα στις Η.Π.Α. όντως υψηλότερες λίγους μήνες πριν από την καταγγελία; Εκτελέστε τον κατάλληλο έλεγχο.

- (3) [5 μονάδες] Παρουσιάζεται μεταβολή στις εισαγωγές από την Κίνα στις Η.Π.Α. μετά την καταγγελία;
- (4) [5 μονάδες] Πόση είναι η μείωση των εισαγωγών μετά την ευνοϊκή απόφαση της επιτροπής για τις αμερικάνικες εταιρίες;
- (5) [5 μονάδες] Μία αύξηση στην συνολική παραγωγή χημικών προϊόντων αυξάνει την ζήτηση του χλωριούχου βαρίου; Ερμηνεύστε αυτήν την σχέση;
- (6) [5 μονάδες] Προτού να δείτε τα αποτελέσματα της μελέτης, θα αναμένατε την ύπαρξη σχέσης μεταξύ των μεταβλητών *chrimp* και *gas*. Εξηγήστε γιατί. Συμφωνεί το συμπέρασμα σας με τα αποτελέσματα της μελέτης;
- (7) [5 μονάδες] Μία υποτίμηση του δολαρίου έναντι των άλλων νομισμάτων, θα μείωνε την εισαγωγή χλωριούχου βαρίου από την Κίνα στις Η.Π.Α.; Εξηγήστε γιατί;
- (8) [10 μονάδες] Εξηγήστε αναλυτικά τα βήματα της μεθόδου Cochrane-Orcutt.
- (9) [5 μονάδες] Για ποιο λόγο εφαρμόζεται ή μέθοδο Cochrane-Orcutt; Εξηγήστε ποιο είναι πραγματικά το πρόβλημα αν δεν την εφαρμόσουμε;
- (10) [5 μονάδες] Εξηγήστε πως προκύπτει η τιμή της $\hat{\rho}$ -0.293, δώστε τύπο η την μεθοδολογία από την οποία προκύπτει;

(1) [10 μονάδες] Θεωρήστε γραμμική παλινδρόμηση με διαστρωματικά δεδομένα. Τι εννοούμε με τον όρο ανθεκτικά τυπικά σφάλματα, εξηγήστε; Είναι οι στατιστικές t και F ανθεκτικές;

(2) [10 μονάδες] Στην γραμμική παλινδρόμηση, τι επιπτώσεις υπάρχουν στην εκτίμηση και στα τυπικά σφάλματα όταν υπάρχει σφάλμα στις μετρήσεις στην εξαρτημένη και σε μία από τις ερμηνευτικές μεταβλητές;

3) [10 μονάδες] Εξηγήστε πως ελέγχουμε για AR(1) αυτοσυσχέτιση με τον έλεγχο Durbin-Watson; Αναφέρεται επίσης και τα εξής βήματα H_0 , H_1 , τεστ, κατανομή του τεστ, περιοχή απόρριψης, και συμπέρασμα, αφού πρώτα γράψετε το μοντέλο (υπόδειγμα).

4) [50 μονάδες] Έστω ότι η $math10$ συμβολίζει το ποσοστό των μαθητών στα γυμνάσια του Μίσιγκαν που πέρασαν ένα τυπικό τεστ μαθηματικών. Θέλουμε να εκτιμήσουμε την επίδραση της μεταβλητής «δαπάνες ανά μαθητή» ($expend$) επί της απόδοσης στα μαθηματικά. Ένα απλό μοντέλο είναι το εξής:

$$math10 = \beta_0 + \beta_1 \log(expend) + \beta_2 \log(enroll) + \beta_3 poverty + u$$

όπου η $poverty$ είναι το ποσοστό των μαθητών που ζουν σε κατάσταση φτώχειας.

(α) [5 μονάδες] Η μεταβλητή $lnchprg$ είναι το ποσοστό των μαθητών που δικαιούνται να συμμετέχουν στο κρατικό επιχορηγούμενο σίτισης. Γιατί αυτή είναι μια λογική αντιπροσωπευτική μεταβλητή για την $poverty$;

(β) [5 μονάδες] Ο πίνακας που ακολουθεί παρουσιάζει εκτιμήσεις των ελαχίστων τετραγώνων με και χωρίς την $lnchprg$ ως ερμηνευτική μεταβλητή. Σε παρένθεση δίνονται τα τυπικά σφάλματα. Εξηγήστε γιατί η επίδραση των δαπανών ανά μαθητή επί της $math10$ είναι μικρότερη στην στήλη (2) από ότι στην στήλη (1).

(γ) [10 μονάδες] Συνεχίζοντας την ερώτηση (β), η επίδραση στην στήλη (2) συνεχίζει να είναι στατιστικώς μεγαλύτερη από το μηδέν; Εκτελέστε το κατάλληλο στατιστικό τεστ. Αναφέρεται επίσης και τα εξής βήματα H_0 , H_1 , τεστ, κατανομή του τεστ, περιοχή απόρριψης, και συμπέρασμα.

Ανεξάρτητες μεταβλητές	(1)	(2)
$\log(expend)$	11.13 (3.30)	7.75 (3.04)
$\log(enroll)$	0.022 (0.615)	-1.26 (0.58)
$lnchprg$	—————	-0.324 (0.036)
Τεταγμένη της αρχής	-69.24 (26.72)	-23.14 (24.99)
Παρατηρήσεις	428	428
R -τετράγωνο	0.0297	0.1893

(δ) [10 μονάδες] Είναι τα ποσοστά των μαθητών που πέρασαν το τεστ μικρότερα στα μεγαλύτερα σχολεία, όταν οι λοιποί παράγοντες παραμένουν σταθεροί; Εξηγήστε την άποψή σας.

(ε) [10 μονάδες] Ερμηνεύστε τον συντελεστή της $lnchprg$ στη στήλη (2).

(στ) [10 μονάδες] Τι συμπέρασμα βγάζετε για την σημαντική αύξηση στο R -τετράγωνο από την στήλη (1) στην στήλη (2);

(5) [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] $\text{Cov}\{Y, X_2\} =$

1) [50 μονάδες] Θεωρήστε το πρόβλημα από την προηγούμενη άσκηση. Από την παλινδρόμηση όλων των ετών παίρνουμε

$$(4.1) \quad \text{Καταθέσεις} = y = 62.42 + 0.0376 \times (\text{Εισόδημα} = x)$$

Δοθέντος ότι από μία νέα παλινδρόμηση παίρνουμε τους εκτιμητές του δεύτερου μέλους:

$$(4.2) \quad \log[(y - 62.42 - 0.0376x)^2] = 4.59 + 0.000400x$$

α) [15 μονάδες] Γράψτε αναλυτικά τους συντελεστές στάθμισης με τους οποίους στην αρχική παλινδρόμηση (4.1) πετυχαίνουμε ομοσκεδαστικότητα.

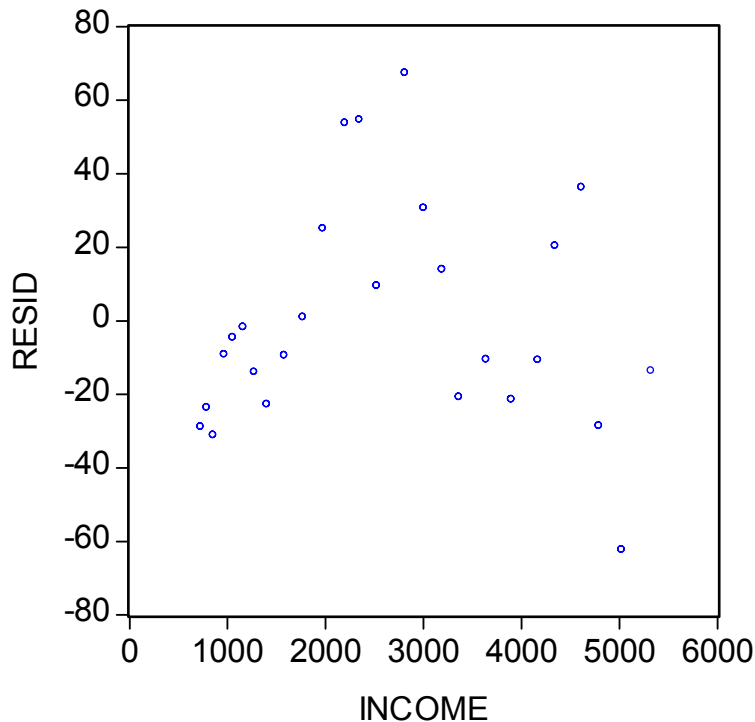
β) [10 μονάδες] Στην δεύτερη παλινδρόμηση, το $R^2=0.085$ και ο εκτιμώμενος συντελεστής με εκτιμώμενη τιμή 0.0004, έχει π -τιμή=0.15 (p-value). Υπάρχουν ενδείξεις ότι υπάρχει ετεροσκεδαστικότητα και γιατί;

γ) [10 μονάδες] Στην παλινδρόμηση (4.1) η στατιστική του Durbin-Watson είναι 0.86. Υπάρχουν ενδείξεις για αυτοσυσχέτιση και γιατί;

δ) [15 μονάδες] Βασισμένοι το παρακάτω διάγραμμα διασποράς (των καταλοίπων με την x μεταβλητή) αναφέρεται

i) [10 μονάδες] αν υπάρχει ετεροσκεδαστικότητα και αυτοσυσχέτιση. Εξηγήστε λεπτομερώς. Αν αναγνωρίζεται ετεροσκεδαστικότητα αναφέρεται την μορφή της. .
Αν αναγνωρίζεται αυτοσυσχέτιση, είναι θετική η αρνητική;

ii) [5 μονάδες] Αν οι απαντήσεις σας στην δ-i) έρχονται σε αντίφαση με τις απαντήσεις των β) και γ), παρακαλώ εξηγήστε.



2) [40 μονάδες] Χρησιμοποιούμε τα δεδομένα για τα έτη 1980 και 1990 που περιλαμβάνουν τις τιμές ενοικίων και άλλες μεταβλητές για πόλεις που «φιλοξενούν» κολέγια. Στόχος μας είναι να διαπιστώσουμε αν μια ισχυρότερη παρουσία σπουδαστών επηρεάζει τις τιμές των ενοικίων. Το μοντέλο είναι:

$$\log(\text{rent}_{it}) = \beta_0 + \delta_0 y90_t + \beta_1 \log(\text{pop}_{it}) + \beta_2 \log(\text{avginc}_{it}) + \beta_3 \text{pctstu}_{it} + u_{it}$$

όπου $y90$ είναι ψευδομεταβλητή, η pop είναι ο πληθυσμός πόλης, η avginc είναι το μέσο εισόδημα, και η pctstu είναι ο πληθυσμός σπουδαστών ως ποσοστό του πληθυσμού της πόλης (στη διάρκεια του σχολικού έτους).

Οι εκτιμητές ελαχίστων τετραγώνων δίνουν, με (κλασικά τυπικά σφάλματα), και [ανθεκτικά τυπικά σφάλματα]:

$$\hat{\log}(\text{rent}_{it}) = -.6 + .3y90_t + .5 \log(\text{pop}_{it}) + .55 \log(\text{avginc}_{it}) + .006 \text{pctstu}_{it} + u_{it}$$

(.5)	(.04)	(.02)	(.05)	(.002)
[.3]	[.1]	[.15]	[.23]	[.004]

$$n=130, R^2=0.9$$

α) [5 μονάδες] Τι συμπέρασμα βγάζεται από τον συντελεστή δ_0 (ερμηνεία), είναι στατιστικά σημαντικός, χρησιμοποιήστε το κλασικό τυπικό σφάλμα.

β) [5 μονάδες] Τι συμπέρασμα βγάζεται από τον συντελεστή δ_0 (ερμηνεία), είναι στατιστικά σημαντικός, χρησιμοποιήστε το κλασικό τυπικό σφάλμα.

γ) [5 μονάδες] Γιατί η μεταβλητή pctstu δεν είναι σε \log μορφή όπως οι υπόλοιπες.

δ) [5 μονάδες] Στην β) αν χρησιμοποιήσετε το ανθεκτικό τυπικό σφάλμα για ετεροσκεδαστικότητα (στην αγκύλη) αλλάζει το συμπέρασμα σας.

ε) [5 μονάδες] Υποπτεύεστε από κάποιο στοιχείο ότι υπάρχει ετεροσκεδαστικότητα; Εξηγήστε. Επηρεάζει η ετεροσκεδαστικότητα την εκτίμηση των συντελεστών γενικώς;

στ) [5 μονάδες] Πως ονομάζονται τα δεδομένα αν στις δύο χρονικές στιγμές εξετάζουμε τις ίδιες πόλεις και πως αν εξετάζουμε διαφορετικές εκτός από ελάχιστες. Πως ονομάζονται τα δεδομένα αν στις δύο χρονικές στιγμές εξετάζουμε τις ίδιες πόλεις αλλά μόνο για ελάχιστες δεν έχουμε δεδομένα και για τις δύο χρονικές στιγμές.

ζ) [5 μονάδες] Για ποιο λόγο αναλύουμε διάφορες σε δεδομένα όπως τα παραπάνω; Στην στ), σε ποιες από τις τρεις περιπτώσεις μπορούμε να πάρουμε διαφορές; Σε ποια από αυτές παρουσιάζεται κάποιο πρόβλημα και γιατί; Εξηγήστε.

η) [5 μονάδες] Αναλύοντας τις διαφορές το $R^2=0.3$ το οποίο διαφέρει πολύ από το προηγούμενο $R^2=0.9$. Ποιο δεν θα εμπιστευόσασταν και γιατί;

3) [22 μονάδες] Έστω $\{x_t: t=-1,0,1,\dots\}$ είναι μια αλληλουχία ανεξάρτητων, ταυτόσημα κατανομημένων τυχαίων μεταβλητών με μέσο ίσο με το 0 και διακύμανση ίση με την μονάδα. Θεωρήστε ότι:

$$x_t = e_t - 0.7e_{t-1} - 0.3e_{t-2}, \quad t=1,2,\dots$$

Βρείτε

α) [3 μονάδες] $E(x_t)=$

β) [5 μονάδες] $\text{Var}(x_t)=$

γ) [7 μονάδες] $\text{Cov}(x_t, x_{t+1})=$

δ) [7 μονάδες] $\text{Cov}(x_t, x_{t+2})=$

4) [15 μονάδες]

α) [10 μονάδες] Γράψτε τον παρακάτω τύπο για $g=4$

$$\hat{v} = \sum_{t=1}^n \hat{a}_t^2 + 2 \sum_{h=1}^g [1-h/(g+1)] \left(\sum_{t=h+1}^n \hat{a}_t \hat{a}_{t-h} \right),$$

β) [5 μονάδες] Εξηγήστε που χρησιμεύει αυτός ο τύπος και πότε χρησιμοποιείται για $g=4$.

5) [50 μονάδες] Εκτιμήστε για το παρακάτω μοντέλο:

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad \text{where} \quad u_t = \rho u_{t-1} + \varepsilon_t \quad \text{with} \quad |\rho| < 1$$

και δεδομένα.

t	Y_t	X_t
1	2	-2
2	4	4
3	3	-2

α) [15 μονάδες] Εκτιμήστε τα β_0 και β_1 με την μέθοδο των ελαχίστων τετραγώνων.

β) [10 μονάδες] Υπολογίστε τα κατάλοιπα

γ) [25 μονάδες] Να γίνει ο έλεγχος αυτοσυσχέτισης πρώτης τάξης με το κριτήριο Durbin-Watson.

6) [50 μονάδες] Θεωρήστε το παρακάτω μοντέλο παλινδρόμησης με σφάλμα μετρήσεις στις δύο ερμηνευτικές μεταβλητές. Επίσης τα δυο σφάλματα μετρήσεων συσχετίζονται μεταξύ τους και με τα σφάλματα της παλινδρόμησης .

$$\begin{cases} y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + e \\ x_1 = x_1^* + v \\ x_2 = x_2^* + u \end{cases}$$

$$E\{x_1\} = E\{x_2\} = E\{e\} = E\{v\} = E\{u\} = 0,$$

$$\text{Var}\{x_1\} = \sigma_{11} = 1, \text{Var}\{x_2\} = \sigma_{22} = 1, \text{Var}\{e\} = \sigma_{ee} = 1, \text{Var}\{v\} = \sigma_{vv} = 1, \text{Var}\{u\} = \sigma_{uu} = 1,$$

$$\text{Cov}\{x_1, x_2\} = \sigma_{12} = 0.9,$$

$$\text{Cov}\{e, v\} = \sigma_{ev} = 0.3, \text{Cov}\{e, u\} = \sigma_{eu} = 0.4, \text{Cov}\{u, v\} = \sigma_{uv} = 0.5,$$

$$\text{Cov}\{x_1, e\} = 0, \text{Cov}\{x_2, e\} = 0,$$

$$\text{Cov}\{x_1, v\} = 0, \text{Cov}\{x_2, v\} = 0,$$

$$\text{Cov}\{x_1, u\} = 0, \text{Cov}\{x_2, u\} = 0$$

α) [15 μονάδες] Δώστε μία οικονομική εφαρμογή στην οποία ενδεχομένως να ισχύει το παραπάνω μοντέλο, χωρίς να λάβετε υπόψη τις δοθέντες τιμές για τις διακυμάνσεις και συνδιακυμάνσεις. Εξηγήστε ακριβώς τις ισχύουσες υποθέσεις στην εφαρμογή.

β) [5 μονάδες] Βρείτε την $\text{Cov}\{x_1, x_2\}$ και σχολιάστε την τιμή της. Η τιμή αυτή θα βοηθήσει την παλινδρόμηση ή θα παρουσιάσει κάποιο πρόβλημα. Εξηγήστε.

γ) [5 μονάδες] Είναι τα σφάλματα μέτρησης μεγάλα σε σχέση με τις άλλες ποσότητες. Εξηγήστε.

δ) [15 μονάδες] Υπολογίστε την $\text{Var}\{y\} = \sigma_{yy}$, λαμβάνοντας υπόψη και την συσχέτιση των μεταβλητών και ότι $\beta_0 = 1, \beta_1 = 2, \beta_2 = -1$.

ε) [15 μονάδες] Υπολογίστε την $\text{Cov}\{y, x_1\}$ λαμβάνοντας υπόψη και την συσχέτιση των μεταβλητών και ότι $\beta_0 = 1, \beta_1 = 2, \beta_2 = -1$.

Υπόδειξη: Στις δυο τελευταίες ερωτήσεις χρησιμοποιήστε τις παρακάτω ιδιότητες.

Θεώρημα

Αν $U = a_1 U_1 + a_2 U_2 + \dots + a_n U_n$ και $V = b_1 V_1 + b_2 V_2 + \dots + b_m V_m$ είναι γραμμικοί συνδυασμοί τυχαίων μεταβλητών U_i και V_j με σταθερούς συντελεστές a_i και b_j .

Ισχύουν τα εξής:

α) $E(U) = a_1 E(U_1) + a_2 E(U_2) + \dots + a_n E(U_n)$

β) $\text{Var}(U) = \sum_{i=1}^n a_i^2 \text{Var}(U_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(U_i, U_j)$, όταν οι μεταβλητές U_k είναι ανεξάρτητες ο δεύτερος όρος είναι 0 αφού όλες οι συνδιακυμάνσεις είναι 0.

γ) $\text{Cov}(U, V) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(U_i, V_j)$

.....