

ΤΕΛΙΚΗ ΓΡΑΠΤΗ ΕΞΕΤΑΣΗ (Μάθημα: “ΕΦΑΡΜΟΓΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΣΕ ΕΠΙΧΕΙΡΗΣΙΑΚΑ ΠΡΟΒΛΗΜΑΤΑ”, Α-Λ, ΛΟ-ΧΡΗ)

Όνομα:

Επίθετο:

Α.Μ:

Έτος Σπουδών:

Αν δικαιούστε προφορική εξέταση συμπληρώστε “ΝΑΙ”

Έχετε παραδώσει την προαιρετική εργασία (κυκλώστε); ΝΑΙ ΟΧΙ

Σχόλιο-Παρατήρηση

.....

.....

Ομάδα Α

ΘΕΜΑ 1ο (7 μονάδες)

Στον παρακάτω πίνακα δίνονται 6 παρατήσεις (**n=6**) για τις τιμές δύο μεταβλητών που αποτυπώνουν τις δυναμικές της διαδικασίας του εκσυγχρονισμού των εθνικών οικονομιών κατά το έτος 2000 (t=2000):

Χώρα	Αστικός Πληθυσμός (%)	Έκταση σιδηροδρομικών γραμμών (σε χιλ. χιλιόμετρα)
Βουλγαρία	69	4
Ισπανία	76	14
Η.Β	79	17
Κροατία	53	3
Ινδία	28	63
Ιταλία	67	16

- 1) Να υπολογίσετε την τιμή του συντελεστή γραμμικής συσχέτισης του Pearson και να δώσετε την ερμηνεία του αποτελέσματος. Θα χαρακτηρίζατε το πρόσημο της τιμής του συντελεστή ως αναμενόμενο; (χρησιμοποιήστε τη γνωστή φόρμουλα εκτελώντας τις απαιτούμενες πράξεις με το κομπιουτεράκι σας σε στήλες πίνακα) (1 μονάδα)
- 2) Με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων (OLS) να υπολογίσετε τους συντελεστές $\hat{\alpha}$ και $\hat{\beta}$ του απλού γραμμικού μοντέλου της μορφής $\Psi(i) = \alpha + \beta * X(i) + \varepsilon(i)$
όπου Ψ : έκταση σιδηροδρομικών γραμμών (σε χιλ. χιλιόμετρα)
 X : αστικός πληθυσμός (%)
 $i=1,2,3,4,5,6$ (αριθμός παρατηρήσεων)
(Οι βασικοί υπολογισμοί σας θα πρέπει να φαίνονται σε στήλες πίνακα). (1 μονάδα)
- 3) Να δώσετε την ερμηνεία του αποτελέσματος του συντελεστή $\hat{\beta}$ που υπολογίσατε στο προηγούμενο ερώτημα. (0,5 μονάδες)
- 4) Να υπολογίσετε τις προβλεπόμενες τιμές $\hat{\Psi}_i$ για κάθε δεδομένη τιμή του X_i . (0,5 μονάδες)
- 5) Να κατασκευάσετε μια νέα στήλη του πίνακα που να περιέχει τα κατάλοιπα ($\hat{u}_i$). Με μοναδικό κριτήριο τις τιμές των καταλοίπων “πόσο καλά” θα λέγατε πως σε θέση το εκτιμώμενο μοντέλο σας να ερμηνεύει την πραγματικότητα; (0,5 μονάδες)

- 6) Να υπολογίσετε το τυπικό σφάλμα της παλινδρόμησης ($\hat{\sigma}$). (0,5 μονάδες)
- 7) Να υπολογίσετε το τυπικό σφάλμα του εκτιμητή $\hat{\beta}$, δηλαδή το $s.e(\hat{\beta})$. Έπειτα, να κατασκευάσετε ένα 95% και ένα 99% διάστημα εμπιστοσύνης για τον προσδιορισμό του πραγματικού συντελεστή β . Να δώσετε την ερμηνεία των αποτελεσμάτων σας. (Δίνονται οι κριτικές τιμές από τον πίνακα της κατανομής t : για $\alpha=5\% \rightarrow t_{(a/2, 4)} = 2,77$ και για $\alpha=1\% \rightarrow t_{(a/2, 4)} = 4,60$) (1 μονάδα)
- 8) Να υπολογίσετε την τιμή του συντελεστή προσδιορισμού (R-squared) του μοντέλου σας και να δώσετε την ερμηνεία του αποτελέσματος. Χρησιμοποιήστε την φόρμουλα που περιλαμβάνει τα αθροίσματα των τετραγώνων (SS). (1 μονάδα)
- 9) Να διεξάγετε τον έλεγχο t για την στατιστική σημαντικότητα του συντελεστή β στο $\alpha=5\%$ και $\alpha=10\%$, αντίστοιχα. Τι συμπέρασμα βγάξετε; Δίνεται ότι για $\alpha=10\% \rightarrow t_{(a/2, 4)} = 2,13$ (1 μονάδα)

ΘΕΜΑ 2ο (3 μονάδες)

Δίνεται στα αριστερά ο πίνακας με τα αποτελέσματα ενός απλού γραμμικού μοντέλου ($k=1$) της μορφής $\log(pol_i) = \alpha + \beta * \log(ineq_i) + \varepsilon_i$ (όπου **pol**: τα επίπεδα της ατμοσφαιρικής ρύπανσης και **ineq**: οι τιμές του δείκτη των οικονομικών ανισοτήτων του Gini). Έπειτα, προσθέτουμε στο αρχικό μας υπόδειγμα 2 νέες ανεξάρτητες μεταβλητές: Τον παράγοντα “**TR**” (σιδηροδρομικές εκτάσεις) και τον παράγοντα “**URB**” (αστικός πληθυσμός). Τα αποτελέσματα της πολλαπλής παλινδρόμησης βρίσκονται στον πίνακα της δεξιάς πλευράς.

Dependent Variable: LOG(POL)
Method: Least Squares
Date: 06/05/25 Time: 14:17
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.656370	1.082275	-0.606472	0.5486
LOG(INEQ)	1.020911	0.300023	3.402778	0.0019
R-squared	0.271940	Mean dependent var	3.019267	
Adjusted R-squared	0.248454	S.D. dependent var	0.445279	
S.E. of regression	0.386020	Akaike info criterion	0.992838	
Sum squared resid	4.619362	Schwarz criterion	1.083536	
Log likelihood	-14.38183	Hannan-Quinn criter.	1.023355	
F-statistic	11.57890	Durbin-Watson stat	2.117059	
Prob(F-statistic)	0.001857			

Dependent Variable: LOG(POL)
Method: Least Squares
Date: 06/05/25 Time: 14:30
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.121776	1.041944	1.076618	0.2905
LOG(INEQ)	0.741041	0.270579	2.738728	0.0104
LOG(TR)	-0.016892	0.044402	-0.380445	0.7064
URB	-0.011628	0.003162	-3.677167	0.0010
R-squared	0.522379	Mean dependent var	3.019267	
Adjusted R-squared	0.472970	S.D. dependent var	0.445279	
S.E. of regression	0.323259	Akaike info criterion	0.692484	
Sum squared resid	3.030387	Schwarz criterion	0.873879	
Log likelihood	-7.425988	Hannan-Quinn criter.	0.753518	
F-statistic	10.57253	Durbin-Watson stat	1.944372	
Prob(F-statistic)	0.000073			

- 1) Να δώσετε την ερμηνεία του συντελεστή $\hat{\beta}$ στο απλό γραμμικό μοντέλο. (0,5 μονάδες)
- 2) Είναι ο παράγοντας “οικονομικές ανισότητες” (**ineq**) στατιστικά σημαντικός για τον προσδιορισμό των επιπέδων της ατμοσφαιρικής ρύπανσης; Να δικαιολογήσετε την απάντησή της με τη βοήθεια της τιμής του ελέγχου t ή με τη βοήθεια της τιμής του p -value. (0,5 μονάδες)
- 3) Με μοναδικό κριτήριο την τιμή του συντελεστή προσδιορισμού πως θα χαρακτηρίζατε την ερμηνευτική ικανότητα του απλού γραμμικού μοντέλου σας; (0,5 μονάδες)
- 4) Βελτιώθηκε η ερμηνευτική ικανότητα του μοντέλου μας με την προσθήκη των 2 νέων μεταβλητών; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 5) Είναι και οι δύο παράγοντες που προστέθηκαν στο μοντέλο μας στατιστικά σημαντικοί; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 6) Να διατυπώσετε την μηδενική και την εναλλακτική υπόθεση του ελέγχου από κοινού σημαντικότητας των συντελεστών β της παλινδρόμησης (F). Ποιο συμπέρασμα θα βγάξετε κοιτώντας την στατιστική τιμή του ελέγχου F που παρατίθεται στον πίνακα του πολυμεταβλητού μοντέλου; Αιτιολογήστε την απάντησή σας. (0,5 μονάδες)

ΤΕΛΙΚΗ ΓΡΑΠΤΗ ΕΞΕΤΑΣΗ (Μάθημα: “ΕΦΑΡΜΟΓΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΣΕ ΕΠΙΧΕΙΡΗΣΙΑΚΑ ΠΡΟΒΛΗΜΑΤΑ”, Α-Λ, ΛΟ-ΧΡΗ)

Όνομα:

Επίθετο:

Α.Μ:

Έτος Σπουδών:

Αν δικαιούστε προφορική εξέταση συμπληρώστε “ΝΑΙ”

Έχετε παραδώσει την προαιρετική εργασία (κυκλώστε); ΝΑΙ ΟΧΙ

Σχόλιο-Παρατήρηση

.....

.....

Ομάδα Β

ΘΕΜΑ 1ο (7 μονάδες)

Στον παρακάτω πίνακα δίνονται 6 παρατήσεις (**n=6**) για τις τιμές δύο μεταβλητών που αποτυπώνουν την κοινωνική ευαλωτότητα στο εθνικό επίπεδο κατά το έτος 2010 (t=2010):

Χώρα	Κορίτσια στην παραγωγή (%)	Οικονομικές ανισότητες (Gini index)
1	20	48
2	19	36
3	9	29
4	3	30
5	8	38
6	2	35

- 1) Να υπολογίσετε την τιμή του συντελεστή γραμμικής συσχέτισης του Pearson και να δώσετε την ερμηνεία του αποτελέσματος. Θα χαρακτηρίζετε το πρόσημο της τιμής του συντελεστή ως αναμενόμενο; (χρησιμοποιήστε τη γνωστή φόρμουλα εκτελώντας τις απαιτούμενες πράξεις με το κομπιουτεράκι σας σε στήλες πίνακα) (1 μονάδα)

- 2) Με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων (OLS) να υπολογίσετε τους συντελεστές $\hat{\alpha}$ και $\hat{\beta}$ του απλού γραμμικού μοντέλου της μορφής $\Psi(i) = \alpha + \beta * X(i) + \varepsilon(i)$

όπου Ψ : κορίτσια στην παραγωγή (%)

X : οικονομικές ανισότητες (δείκτης του Gini)

$i=1,2,3,4,5,6$ (αριθμός παρατηρήσεων)

(Οι βασικοί υπολογισμοί σας θα πρέπει να φαίνονται σε στήλες πίνακα). (1 μονάδα)

- 3) Να δώσετε την ερμηνεία του αποτελέσματος για τον συντελεστή $\hat{\beta}$ που υπολογίσατε στο προηγούμενο ερώτημα. (0,5 μονάδες)

- 4) Να υπολογίσετε τις προβλεπόμενες τιμές $\hat{\Psi}_i$ για κάθε δεδομένη τιμή του X_i . (0,5 μονάδες)

- 5) Να κατασκευάσετε μια νέα στήλη του πίνακα που να περιέχει τα κατάλοιπα ($\hat{u}_i$). Με μοναδικό κριτήριο τις τιμές των καταλοίπων “πόσο καλά” θα λέγατε πως σε θέση το εκτιμώμενο μοντέλο σας να ερμηνεύει την πραγματικότητα; (0,5 μονάδες)

6) Να υπολογίσετε το τυπικό σφάλμα της παλινδρόμησης ($\hat{\sigma}$). (0,5 μονάδες)

7) Να υπολογίσετε το τυπικό σφάλμα του εκτιμητή $\hat{\beta}$, δηλαδή το $s.e(\hat{\beta})$. Έπειτα, να κατασκευάσετε ένα 90% και ένα 99% διάστημα εμπιστοσύνης για τον προσδιορισμό του πραγματικού συντελεστή β . Να δώσετε την ερμηνεία των αποτελεσμάτων σας. (Δίνονται οι κριτικές τιμές από τον πίνακα της κατανομής t: για $\alpha=10\% \rightarrow t_{(a/2, 4)} = 2,13$ και για $\alpha=1\% \rightarrow t_{(a/2, 4)} = 4,60$) (1 μονάδα)

8) Να υπολογίσετε την τιμή του συντελεστή προσδιορισμού (R-squared) του μοντέλου σας και να δώσετε την ερμηνεία του αποτελέσματος. Χρησιμοποιήστε την φόρμουλα που περιλαμβάνει τα αθροίσματα των τετραγώνων (SS). (1 μονάδα)

9) Να διεξάγετε τον έλεγχο t για την στατιστική σημαντικότητα του συντελεστή β στο $\alpha=1\%$ και $\alpha=10\%$, αντίστοιχα. Τι συμπέρασμα βγάξετε; (1 μονάδα)

ΘΕΜΑ 2ο (3 μονάδες)

Δίνεται στα αριστερά ο πίνακας με τα αποτελέσματα ενός απλού γραμμικού μοντέλου ($k=1$) της μορφής $\log(\text{ineq}_i) = \alpha + \beta * \log(\text{pol}_i) + \varepsilon_i$ (όπου **pol**: τα επίπεδα της ατμοσφαιρικής ρύπανσης και **ineq**: οι τιμές του δείκτη των οικονομικών ανισοτήτων του Gini). Έπειτα, προσθέτουμε στο αρχικό μας υπόδειγμα 2 νέες ανεξάρτητες μεταβλητές: Τον παράγοντα “**TR**” (σιδηροδρομικές εκτάσεις) και τον παράγοντα “**URB**” (αστικός πληθυσμός). Τα αποτελέσματα της πολλαπλής παλινδρόμησης βρίσκονται στον πίνακα της δεξιάς πλευράς.

Dependent Variable: LOG(INEQ)
Method: Least Squares
Date: 06/05/25 Time: 16:04
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.796108	0.238828	11.70763	0.0000
LOG(POL)	0.266370	0.078280	3.402778	0.0019
R-squared	0.271940	Mean dependent var	3.600349	
Adjusted R-squared	0.248454	S.D. dependent var	0.227447	
S.E. of regression	0.197178	Akaike info criterion	-0.350728	
Sum squared resid	1.205254	Schwarz criterion	-0.260030	
Log likelihood	7.787007	Hannan-Quinn criter.	-0.320211	
F-statistic	11.57890	Durbin-Watson stat	1.932226	
Prob(F-statistic)	0.001857			

Dependent Variable: LOG(INEQ)
Method: Least Squares
Date: 06/05/25 Time: 16:05
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.691351	0.415601	6.475804	0.0000
LOG(POL)	0.277303	0.101253	2.738728	0.0104
URB	0.001705	0.002321	0.734484	0.4686
LOG(TR)	-0.032390	0.026557	-1.219650	0.2324
R-squared	0.314986	Mean dependent var	3.600349	
Adjusted R-squared	0.244122	S.D. dependent var	0.227447	
S.E. of regression	0.197745	Akaike info criterion	-0.290460	
Sum squared resid	1.133994	Schwarz criterion	-0.109065	
Log likelihood	8.792590	Hannan-Quinn criter.	-0.229426	
F-statistic	4.444966	Durbin-Watson stat	1.958825	
Prob(F-statistic)	0.010933			

- 1) Να δώσετε την ερμηνεία του συντελεστή $\hat{\beta}$ στο απλό γραμμικό μοντέλο. (0,5 μονάδες)
- 2) Είναι ο παράγοντας “ατμοσφαιρική ρύπανση” (**pol**) στατιστικά σημαντικός για τον προσδιορισμό των επιπέδων της οικονομικής ανισότητας; Να δικαιολογήσετε την απάντησή της με τη βοήθεια της τιμής του ελέγχου t ή με τη βοήθεια της τιμής του p-value. (0,5 μονάδες)
- 3) Με μοναδικό κριτήριο την τιμή του συντελεστή προσδιορισμού πως θα χαρακτηρίζατε την ερμηνευτική ικανότητα του απλού γραμμικού μοντέλου σας; (0,5 μονάδες)
- 4) Βελτιώθηκε η ερμηνευτική ικανότητα του μοντέλου μας με την προσθήκη των 2 νέων μεταβλητών; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 5) Είναι και οι δύο παράγοντες που προστέθηκαν στο μοντέλο μας στατιστικά σημαντικοί; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 6) Να διατυπώσετε την μηδενική και την εναλλακτική υπόθεση του ελέγχου από κοινού σημαντικότητας των συντελεστών β της παλινδρόμησης (F). Ποιο συμπέρασμα θα βγάzaτε κοιτώντας την στατιστική τιμή του ελέγχου F που παρατίθεται στον πίνακα του πολυμεταβλητού μοντέλου; Αιτιολογήστε την απάντησή σας. (0,5 μονάδες)

ΤΕΛΙΚΗ ΓΡΑΠΤΗ ΕΞΕΤΑΣΗ (Μάθημα: “ΕΦΑΡΜΟΓΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΣΕ ΕΠΙΧΕΙΡΗΣΙΑΚΑ ΠΡΟΒΛΗΜΑΤΑ”, Α-Λ, ΛΟ-ΧΡΗ)

Όνομα:

Επίθετο:

Α.Μ:

Έτος Σπουδών:

Αν δικαιούστε προφορική εξέταση συμπληρώστε “ΝΑΙ”

Έχετε παραδώσει την προαιρετική εργασία (κυκλώστε); ΝΑΙ ΟΧΙ

Σχόλιο-Παρατήρηση

.....

.....

Ομάδα Γ

ΘΕΜΑ 1ο (7 μονάδες)

Στον παρακάτω πίνακα δίνονται 6 παρατήσεις ($n=6$) για τις τιμές δύο μεταβλητών που αποτυπώνουν τις δυναμικές της διαδικασίας του εκσυγχρονισμού των εθνικών οικονομιών κατά το έτος 2000 ($t=2000$):

Χώρα	Έκταση σιδηροδρομικών γραμμών (σε χιλ. χιλιόμετρα)	Αστικός Πληθυσμός (%)
Ισπανία	14	76
Βουλγαρία	4	69
Η.Β	17	79
Ινδία	63	28
Κροατία	3	53
Ιταλία	16	67

- 1) Να υπολογίσετε την τιμή του συντελεστή γραμμικής συσχέτισης του Pearson και να δώσετε την ερμηνεία του αποτελέσματος. Θα χαρακτηρίζατε το πρόσημο της τιμής του συντελεστή ως αναμενόμενο; (χρησιμοποιήστε τη γνωστή φόρμουλα εκτελώντας τις απαιτούμενες πράξεις με το κομπιουτεράκι σας σε στήλες πίνακα) (1 μονάδα)

- 2) Με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων (OLS) να υπολογίσετε τους συντελεστές $\hat{\alpha}$ και $\hat{\beta}$ του απλού γραμμικού μοντέλου της μορφής $\Psi(i) = \alpha + \beta \cdot X(i) + \varepsilon(i)$

όπου Ψ : αστικός πληθυσμός (%)

X : έκταση σιδηροδρομικών γραμμών (σε χιλ. χιλιόμετρα)

$i=1,2,3,4,5,6$ (αριθμός παρατηρήσεων)

(Οι βασικοί υπολογισμοί σας θα πρέπει να φαίνονται σε στήλες πίνακα). (1 μονάδα)

- 3) Να δώσετε την ερμηνεία του αποτελέσματος του συντελεστή $\hat{\beta}$ που υπολογίσατε στο προηγούμενο ερώτημα. (0,5 μονάδες)

- 4) Να υπολογίσετε τις προβλεπόμενες τιμές $\hat{\Psi}_i$ για κάθε δεδομένη τιμή του X_i . (0,5 μονάδες)

- 5) Να κατασκευάσετε μια νέα στήλη του πίνακα που να περιέχει τα κατάλοιπα ($\hat{u}_i$). Με μοναδικό κριτήριο τις τιμές των καταλοίπων “πόσο καλά” θα λέγατε πως σε θέση το εκτιμώμενο μοντέλο σας να ερμηνεύει την πραγματικότητα; (0,5 μονάδες)

- 6) Να υπολογίσετε το τυπικό σφάλμα της παλινδρόμησης ($\hat{\sigma}$). (0,5 μονάδες)
- 7) Να υπολογίσετε το τυπικό σφάλμα του εκτιμητή $\hat{\beta}$, δηλαδή το $s.e(\hat{\beta})$. Έπειτα, να κατασκευάσετε ένα 95% και ένα 90% διάστημα εμπιστοσύνης για τον προσδιορισμό του πραγματικού συντελεστή β . Να δώσετε την ερμηνεία των αποτελεσμάτων σας. (Δίνονται οι κριτικές τιμές από τον πίνακα της κατανομής t : για $\alpha=5\% \rightarrow t_{(\alpha/2, 4)} = 2,77$ και για $\alpha=10\% \rightarrow t_{(\alpha/2, 4)} = 2,13$) (1 μονάδα)
- 8) Να υπολογίσετε την τιμή του συντελεστή προσδιορισμού (R-squared) του μοντέλου σας και να δώσετε την ερμηνεία του αποτελέσματος. Χρησιμοποιήστε την φόρμουλα που περιλαμβάνει τα αθροίσματα των τετραγώνων (SS). (1 μονάδα)
- 9) Να διεξάγετε τον έλεγχο t για την στατιστική σημαντικότητα του συντελεστή β στο $\alpha=5\%$ και $\alpha=10\%$, αντίστοιχα. Τι συμπέρασμα βγάζετε; (1 μονάδα)

ΘΕΜΑ 2ο (3 μονάδες)

Δίνεται στα αριστερά ο πίνακας με τα αποτελέσματα ενός απλού γραμμικού μοντέλου ($k=1$) της μορφής $\log(tr_i) = \alpha + \beta * ineq_i + \varepsilon_i$ (όπου tr : έκταση σιδηροδρομικών γραμμών σε χιλ. χιλιόμετρα και $ineq$: δείκτης οικονομικών ανισοτήτων του Gini). Έπειτα, προσθέτουμε στο αρχικό μας υπόδειγμα 1 νέα ανεξάρτητη μεταβλητή: Τον παράγοντα “ urb ” (αστικός πληθυσμός). Τα αποτελέσματα της πολλαπλής παλινδρόμησης βρίσκονται στον πίνακα της δεξιάς πλευράς.

Dependent Variable: LOG(TR)
Method: Least Squares
Date: 06/05/25 Time: 16:45
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.374810	1.018188	3.314526	0.0023
INEQ	-0.058376	0.026396	-2.211569	0.0345
R-squared	0.136275	Mean dependent var	1.181980	
Adjusted R-squared	0.108413	S.D. dependent var	1.408245	
S.E. of regression	1.329720	Akaike info criterion	3.466506	
Sum squared resid	54.81281	Schwarz criterion	3.557203	
Log likelihood	-55.19734	Hannan-Quinn criter.	3.497023	
F-statistic	4.891037	Durbin-Watson stat	0.359787	
Prob(F-statistic)	0.034500			

Dependent Variable: LOG(TR)
Method: Least Squares
Date: 06/05/25 Time: 16:44
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.972078	1.426896	1.382075	0.1772
URB	0.017079	0.012351	1.382849	0.1769
INEQ	-0.050381	0.026651	-1.890411	0.0684
R-squared	0.188031	Mean dependent var	1.181980	
Adjusted R-squared	0.133900	S.D. dependent var	1.408245	
S.E. of regression	1.310576	Akaike info criterion	3.465318	
Sum squared resid	51.52828	Schwarz criterion	3.601365	
Log likelihood	-54.17775	Hannan-Quinn criter.	3.511094	
F-statistic	3.473622	Durbin-Watson stat	0.446823	
Prob(F-statistic)	0.043963			

- 1) Να δώσετε την ερμηνεία του συντελεστή $\hat{\beta}$ στο απλό γραμμικό μοντέλο. (0,5 μονάδες)
- 2) Είναι ο παράγοντας “οικονομικές ανισότητες” ($ineq$) στατιστικά σημαντικός για τον προσδιορισμό του επιπέδου της έκτασης των σιδηροδρομικών γραμμών; Να δικαιολογήσετε την απάντησή της με τη βοήθεια της τιμής του ελέγχου t ή με τη βοήθεια της τιμής του p -value. (0,5 μονάδες)
- 3) Με μοναδικό κριτήριο την τιμή του συντελεστή προσδιορισμού πως θα χαρακτηρίζατε την ερμηνευτική ικανότητα του απλού γραμμικού μοντέλου σας; (0,5 μονάδες)
- 4) Βελτιώθηκε η ερμηνευτική ικανότητα του μοντέλου μας με την προσθήκη της νέας μεταβλητής; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 5) Είναι ο παράγοντας που προστέθηκε στο μοντέλο μας στατιστικά σημαντικός; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 6) Να διατυπώσετε την μηδενική και την εναλλακτική υπόθεση του ελέγχου από κοινού σημαντικότητας των συντελεστών β της παλινδρόμησης (F). Ποιο συμπέρασμα θα βγάzaτε κοιτώντας την στατιστική τιμή του ελέγχου F που παρατίθεται στον πίνακα του πολυμεταβλητού μοντέλου; Αιτιολογήστε την απάντησή σας. (0,5 μονάδες)

ΤΕΛΙΚΗ ΓΡΑΠΤΗ ΕΞΕΤΑΣΗ (Μάθημα: “ΕΦΑΡΜΟΓΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΣΕ ΕΠΙΧΕΙΡΗΣΙΑΚΑ ΠΡΟΒΛΗΜΑΤΑ”, Α-Λ, ΛΟ-ΧΡΗ)

Όνομα:

Επίθετο:

Α.Μ:

Έτος Σπουδών:

Αν δικαιούστε προφορική εξέταση συμπληρώστε “ΝΑΙ”

Έχετε παραδώσει την προαιρετική εργασία (κυκλώστε); ΝΑΙ ΟΧΙ

Σχόλιο-Παρατήρηση

.....

.....

Ομάδα Δ

ΘΕΜΑ 1ο (7 μονάδες)

Στον παρακάτω πίνακα δίνονται 6 παρατήσεις (**n=6**) για τις τιμές δύο μεταβλητών που αποτυπώνουν την κοινωνική ευαλωτότητα στο εθνικό επίπεδο κατά το έτος 2010 ($t=2010$):

Χώρα	Κορίτσια στην παραγωγή (%)	Οικονομικές ανισότητες (Gini index)
1	2	35
2	19	36
3	20	48
4	8	38
5	3	30
6	9	29

- 1) Να υπολογίσετε την τιμή του συντελεστή γραμμικής συσχέτισης του Pearson και να δώσετε την ερμηνεία του αποτελέσματος. Θα χαρακτηρίζετε το πρόσημο της τιμής του συντελεστή ως αναμενόμενο; (χρησιμοποιήστε τη γνωστή φόρμουλα εκτελώντας τις απαιτούμενες πράξεις με το κομπιουτεράκι σας σε στήλες πίνακα) (1 μονάδα)

- 2) Με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων (OLS) να υπολογίσετε τους συντελεστές $\hat{\alpha}$ και $\hat{\beta}$ του απλού γραμμικού μοντέλου της μορφής $\Psi(i) = \alpha + \beta * X(i) + \varepsilon(i)$

όπου Ψ : οικονομικές ανισότητες (δείκτης του Gini)

X : κορίτσια στην παραγωγή (%)

$i=1,2,3,4,5,6$ (αριθμός παρατηρήσεων)

(Οι βασικοί υπολογισμοί σας θα πρέπει να φαίνονται σε στήλες πίνακα). (1 μονάδα)

- 3) Να δώσετε την ερμηνεία του αποτελέσματος για τον συντελεστή $\hat{\beta}$ που υπολογίσατε στο προηγούμενο ερώτημα. (0,5 μονάδες)

- 4) Να υπολογίσετε τις προβλεπόμενες τιμές $\hat{\Psi}_i$ για κάθε δεδομένη τιμή του X_i . (0,5 μονάδες)

- 5) Να κατασκευάσετε μια νέα στήλη του πίνακα που να περιέχει τα κατάλοιπα ($\hat{u}_i$). Με μοναδικό κριτήριο τις τιμές των καταλοίπων “πόσο καλά” θα λέγατε πως σε θέση το εκτιμώμενο μοντέλο σας να ερμηνεύει την πραγματικότητα; (0,5 μονάδες)

- 6) Να υπολογίσετε το τυπικό σφάλμα της παλινδρόμησης ($\hat{\sigma}$). (0,5 μονάδες)
- 7) Να υπολογίσετε το τυπικό σφάλμα του εκτιμητή $\hat{\beta}$, δηλαδή το $s.e(\hat{\beta})$. Έπειτα, να κατασκευάσετε ένα 95% και ένα 99% διάστημα εμπιστοσύνης για τον προσδιορισμό του πραγματικού συντελεστή β . Να δώσετε την ερμηνεία των αποτελεσμάτων σας. (Δίνονται οι κριτικές τιμές από τον πίνακα της κατανομής t : για $\alpha=5\% \rightarrow t_{(a/2, 4)} = 2,77$ και για $\alpha=1\% \rightarrow t_{(a/2, 4)} = 4,60$) (1 μονάδα)
- 8) Να υπολογίσετε την τιμή του συντελεστή προσδιορισμού (R-squared) του μοντέλου σας και να δώσετε την ερμηνεία του αποτελέσματος. Χρησιμοποιήστε την φόρμουλα που περιλαμβάνει τα αθροίσματα των τετραγώνων (SS). (1 μονάδα)
- 9) Να διεξάγετε τον έλεγχο t για την στατιστική σημαντικότητα του συντελεστή β στο $\alpha=5\%$ και $\alpha=10\%$, αντίστοιχα. Τι συμπέρασμα βγάξετε; Δίνεται ότι για $\alpha=10\% \rightarrow t_{(a/2, 4)} = 2,13$ (1 μονάδα)

ΘΕΜΑ 2ο (3 μονάδες)

Δίνεται στα αριστερά ο πίνακας με τα αποτελέσματα ενός απλού γραμμικού μοντέλου ($k=1$) της μορφής $urb_i = \alpha + \beta * pol_i + \varepsilon_i$ (όπου **pol**: τα επίπεδα της ατμοσφαιρικής ρύπανσης και **urb**: τα ποσοστά του αστικού πληθυσμού). Έπειτα, προσθέτουμε στο αρχικό μας υπόδειγμα 1 νέα ανεξάρτητη μεταβλητή: Τον παράγοντα “**ineq**” (δείκτης οικονομικών ανισοτήτων του Gini). Τα αποτελέσματα της πολλαπλής παλινδρόμησης βρίσκονται στον πίνακα της δεξιάς πλευράς.

Dependent Variable: URB
Method: Least Squares
Date: 06/05/25 Time: 17:10
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	90.65846	6.441087	14.07502	0.0000
POL	-1.158393	0.260330	-4.449706	0.0001
R-squared	0.389762	Mean dependent var	64.54545	
Adjusted R-squared	0.370077	S.D. dependent var	19.21603	
S.E. of regression	15.25131	Akaike info criterion	8.345900	
Sum squared resid	7210.679	Schwarz criterion	8.436597	
Log likelihood	-135.7073	Hannan-Quinn criter.	8.376417	
F-statistic	19.79988	Durbin-Watson stat	2.129832	
Prob(F-statistic)	0.000103			

Dependent Variable: URB
Method: Least Squares
Date: 06/05/25 Time: 17:12
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	84.80873	11.81971	7.175196	0.0000
POL	-1.240695	0.297467	-4.170873	0.0002
INEQ	0.205119	0.345937	0.592938	0.5577
R-squared	0.396831	Mean dependent var	64.54545	
Adjusted R-squared	0.356620	S.D. dependent var	19.21603	
S.E. of regression	15.41336	Akaike info criterion	8.394855	
Sum squared resid	7127.154	Schwarz criterion	8.530901	
Log likelihood	-135.5151	Hannan-Quinn criter.	8.440630	
F-statistic	9.868653	Durbin-Watson stat	2.070463	
Prob(F-statistic)	0.000509			

- 1) Να δώσετε την ερμηνεία του συντελεστή $\hat{\beta}$ στο απλό γραμμικό μοντέλο. (0,5 μονάδες)
- 2) Είναι ο παράγοντας “ατμοσφαιρική ρύπανση” (**pol**) στατιστικά σημαντικός για τον προσδιορισμό των ποσοστών του αστικού πληθυσμού; Να δικαιολογήσετε την απάντησή της με τη βοήθεια της τιμής του ελέγχου t ή με τη βοήθεια της τιμής του p -value. (0,5 μονάδες)
- 3) Με μοναδικό κριτήριο την τιμή του συντελεστή προσδιορισμού πως θα χαρακτηρίζατε την ερμηνευτική ικανότητα του απλού γραμμικού μοντέλου σας; (0,5 μονάδες)
- 4) Βελτιώθηκε η ερμηνευτική ικανότητα του μοντέλου μας με την προσθήκη της νέας μεταβλητής; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 5) Είναι ο νέος παράγοντας που προστέθηκε στο μοντέλο μας στατιστικά σημαντικός; Αιτιολογήστε την απάντησή σας με τη βοήθεια στοιχείων από τον νέο πίνακα. (0,5 μονάδες)
- 6) Να διατυπώσετε την μηδενική και την εναλλακτική υπόθεση του ελέγχου από κοινού σημαντικότητας των συντελεστών β της παλινδρόμησης (F). Ποιο συμπέρασμα θα βγάzaτε κοιτώντας την στατιστική τιμή του ελέγχου F που παρατίθεται στον πίνακα του πολυμεταβλητού μοντέλου; Αιτιολογήστε την απάντησή σας. (0,5 μονάδες)