

Επαναληπτικό Θέμα

Δίνεται ο παρακάτω πίνακας με τα ετήσια στοιχεία 8 οικονομιών ($n=8$) για το έτος 2015 ($t=2015$):

παρατηρήσεις (i)	Ρυθμός ανάπτυξης (Ψ_i , %)	Συνεισφορά της βιομηχανίας στο ΑΕΠ (X_i , %)
1	2,2	23,7
2	3,4	24
3	0,6	24,4
4	1,6	24,2
5	5	33,5
6	3,4	19,8
7	3,7	26,4
8	0,9	20,3

- 1) Να υπολογίσετε την τιμή του συντελεστή γραμμικής συσχέτισης που συνδέει τους ρυθμούς ανάπτυξης (Ψ) με την συνεισφορά της βιομηχανίας στο ΑΕΠ (X).
Έπειτα, να χαρακτηρίσετε τη σχέση.

i	Ψ_i	X_i	$(\Psi_i - \bar{\Psi})$	$(X_i - \bar{X})$	$(\Psi_i - \bar{\Psi}) * (X_i - \bar{X})$	$(\Psi_i - \bar{\Psi})^2$	$(X_i - \bar{X})^2$
1	2,2	23,7	(2,2-2,6)=-0,4	(23,7-24,5)=-0,8	-0,4*(-0,8)=0,32	0,4*0,4=0,16	0,8*0,8=0,64
2	3,4	24	(3,4-2,6)=0,8	(24-24,5)=-0,5	0,8*(-0,5)=-0,4	0,8*0,8=0,64	0,5*0,5=0,25
3	0,6	24,4	(0,6-2,6)=-2	(24,4-24,5)=-0,1	-2*(-0,1)=0,2	2*2=4	0,1*0,1=0,01
4	1,6	24,2	(1,6-2,6)=-1	(24,2-24,5)=-0,3	-1*(-0,3)=0,3	1*1=1	0,3*0,3=0,09
5	5	33,5	(5-2,6)=2,4	(33,5-24,5)=9	2,4*9=21,6	2,4*2,4=5,76	9*9=81
6	3,4	19,8	(3,4-2,6)=0,8	(19,8-24,5)=-4,7	0,8*(-4,7)=-3,76	0,8*0,8=0,64	4,7*4,7=22,09
7	3,7	26,4	(3,7-2,6)=1,1	(26,4-24,5)=1,9	1,1*1,9=2,09	1,1*1,1=3,7	1,9*1,9=3,61
8	0,9	20,3	(0,9-2,6)=-1,7	(20,3-24,5)=-4,2	-1,7*(-4,2)=7,14	1,7*1,7=0,9	4,2*4,2=17,64
					sum=27,49	sum=16,30	sum=125,33

$$\bar{X} = \frac{(23,7+24+24,4+24,2+33,5+19,8+26,4+20,3)}{8} = 24,5$$

$$\bar{\Psi} = \frac{(2,2+3,4+0,6+1,6+5+3,4+3,7+0,9)}{8} = 2,6$$

$$r = \frac{27,49}{\sqrt{(16,30)*(125,33)}} = + 0,61$$

Συμπέρασμα: ο συντελεστής συσχέτισης λαμβάνει την τιμή **$\rho=+0,61$** . Υπάρχει μια σχετικά έντονη θετική σχέση ανάμεσα στους δύο παράγοντες που εξετάζονται.

$$r = \frac{\sum_1^n (X_i - \bar{X}) * (\Psi_i - \bar{\Psi})}{\sqrt{\sum_1^n (X_i - \bar{X})^2 * \sum_1^n (\Psi_i - \bar{\Psi})^2}}$$

2) Με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων να εκτιμήσετε το μοντέλο:

$$\Psi_i = \alpha + \beta * X_i + \varepsilon_i$$

όπου ψ: ρυθμοί ανάπτυξης

χ: συνεισφορά βιομηχανίας στο ΑΕΠ

i	Ψi	Xi	(Xi*Ψi)	X_i^2
1	2,2	23,7	=52,14	=561,69
2	3,4	24	=81,60	=576
3	0,6	24,4	=14,64	=595,36
4	1,6	24,2	=38,72	=585,64
5	5	33,5	=167,50	=1.122,25
6	3,4	19,8	=67,32	=392,04
7	3,7	26,4	=97,68	=696,96
8	0,9	20,3	=18,27	=412,09
	sum=20,8	sum=196,3	sum=537,87	sum=4.942,09

$$\hat{\beta} = \frac{n \sum_{i=1}^n (X_i \Psi_i) - \sum_{i=1}^n X_i \sum_{i=1}^n \Psi_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} = \frac{8 * (537,87) - (20,8) * (196,3)}{8 * (4.942,09) - (196,3)^2} = \frac{4.302,96 - 4.083}{39.536,72 - 38.533,7} = \frac{219,92}{1.003,03} = + 0,22$$

$$\hat{a} = \bar{\Psi} - \bar{X} * \hat{\beta} = 2,6 - 0,22 * 24,5 = - 2,79$$

3) Να δώσετε την ερμηνεία του συντελεστή b της παλινδρόμησης:

Μια αύξηση της συνεισφοράς της βιομηχανίας στο ΑΕΠ κατά 1 μονάδα (δηλ κατά 1%) θα οδηγήσει στην αύξηση του ρυθμού οικονομικής ανάπτυξης κατά περίπου **0,22 μονάδες** (δηλ 0,22%).

4) Να υπολογίσετε τις εκτιμώμενες τιμές της εξαρτημένης μεταβλητής σας για κάθε δεδομένη τιμή της ανεξάρτητης μεταβλητής:

i	Ψ_i	X_i	$\widehat{\Psi}_i$
1	2,2	23,7	$= -2,79 + 0,22 \cdot (23,7) = 2,42$
2	3,4	24	$= -2,79 + 0,22 \cdot (24) = 2,49$
3	0,6	24,4	$= -2,79 + 0,22 \cdot (24,4) = 2,58$
4	1,6	24,2	$= -2,79 + 0,22 \cdot (24,2) = 2,53$
5	5	33,5	$= -2,79 + 0,22 \cdot (33,5) = 4,58$
6	3,4	19,8	$= -2,79 + 0,22 \cdot (19,8) = 1,57$
7	3,7	26,4	$= -2,79 + 0,22 \cdot (26,4) = 3,02$
8	0,9	20,3	$= -2,79 + 0,22 \cdot (20,3) = 1,68$

5) Να υπολογίσετε τα σφάλματα της παλινδρόμησης:

i	Ψ_i	X_i	$\widehat{\Psi}_i$	$(\Psi_i - \widehat{\Psi}_i)$	$(\Psi_i - \widehat{\Psi}_i)^2$
1	2,2	23,7	$=-2,79+0,22*(23,7)=2,42$	$(2,2 - 2,42)=-0,22$	0,05
2	3,4	24	$=-2,79+0,22*(24)=2,49$	$(3,4 - 2,49)=0,91$	0,83
3	0,6	24,4	$=-2,79+0,22*(24,4)=2,58$	$(0,6 - 2,58)=-1,98$	3,91
4	1,6	24,2	$=-2,79+0,22*(24,2)=2,53$	$(1,6 - 2,53)=-0,93$	0,87
5	5	33,5	$=-2,79+0,22*(33,5)=4,58$	$(5 - 4,58)=0,42$	0,18
6	3,4	19,8	$=-2,79+0,22*(19,8)=1,57$	$(3,4 - 1,57)=1,83$	3,36
7	3,7	26,4	$=-2,79+0,22*(26,4)=3,02$	$(3,7 - 3,02)=0,68$	0,47
8	0,9	20,3	$=-2,79+0,22*(20,3)=1,68$	$(0,9 - 1,68)=-0,78$	0,60
				sum=-0,07	sum=10,27

6) Να υπολογίσετε τον συντελεστή προσδιορισμού και να δώσετε την ερμηνεία του.

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_1^n (\Psi_i - \widehat{\Psi}_i)^2}{\sum_1^n (\Psi_i - \bar{\Psi})^2} = 1 - \frac{10,27}{16,30} = 1 - 0,63 = 0,37$$

Ερμηνεία: Περίπου το 37% της συνολικής μεταβλητότητας των τιμών της εξαρτημένης μεταβλητής μπορεί να εξηγηθεί από τη μεταβλητότητα των τιμών της ανεξάρτητης μεταβλητής μας. Το υπόλοιπο 63% ερμηνεύεται από άλλους παράγοντες που δεν συμπεριλαμβάνονται στο μοντέλο μας.

Εναλλακτικά: $r^2 = R^2 = (0,61)^2 = 0,37$ (ή 37%)

7) Να εξεταστεί η στατιστική σημαντικότητα του συντελεστή b της παλινδρόμησης.

1ο βήμα: διατύπωση υποθέσεων

$H_0: \beta = 0 \rightarrow$ ο συντελεστής β είναι στατιστικά σημαντικός

$H_1: \beta \neq 0 \rightarrow$ ο συντελεστής β δεν είναι στατιστικά σημαντικός

2ο βήμα: υπολογισμός της στατιστικής τιμής ελέγχου (t-stat)

$$t - stat = \frac{\hat{\beta} - 0}{s.e(\hat{\beta})}$$

3ο βήμα: σύγκριση της t-stat με τις κριτικές τιμές από τους πίνακες

Για να ολοκληρωθεί η διαδικασία πρέπει να υπολογιστεί η τιμή του τυπικού σφάλματος του εκτιμητή της παλινδρόμησης:

$$s.e(\hat{\beta}) = \hat{\sigma} * \frac{1}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

όπου

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{10,27}{8-1-1}} = 1,31 \text{ (τυπικό σφάλμα παλινδρόμησης)}$$

$$\text{Άρα: } s.e(\hat{\beta}) = \hat{\sigma} * \frac{1}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} = 1,31 * \frac{1}{\sqrt{125,33}} = 0,12$$

Η στατιστική τιμή ελέγχου θα είναι:

$$t - stat = \frac{\hat{\beta} - 0}{s.e(\hat{\beta})} = \frac{0,22}{0,1169} = 1,88$$

για $\alpha=5\%$: $t\text{-crit}=2,44 \rightarrow t\text{-stat} < t\text{-crit} \rightarrow$ **Do not reject the null hypothesis (ο συντελεστής β ΔΕΝ είναι στατιστικά σημαντικός)**

8) Να κατασκευάσετε ένα 95% και ένα 99% διάστημα εμπιστοσύνης για τον συντελεστή β.

$$P(\hat{\beta} - t_{\alpha/2, n-k-1} * s.e(\hat{\beta}) < \beta < \hat{\beta} + t_{\alpha/2, n-k-1} * s.e(\hat{\beta})) = 95\%$$

$$P(0,22 - 2,44 * 0,1169 < \beta < 0,22 + 2,44 * 0,1169) = 95\%$$

$$P(-0,06 < \beta < 0,50) = 95\%$$

$$P(\hat{\beta} - t_{\alpha/2, n-k-1} * s.e(\hat{\beta}) < \beta < \hat{\beta} + t_{\alpha/2, n-k-1} * s.e(\hat{\beta})) = 99\%$$

$$P(0,22 - 3,70 * 0,1169 < \beta < 0,22 + 3,70 * 0,1169) = 99\%$$

$$P(-0,21 < \beta < 0,65) = 99\%$$

**Δίνεται το αποτέλεσμα μιας απλής γραμμικής παλινδρόμησης
(Ψ: έκταση σιδηροδρόμου, Χ: αστικός πληθυσμός)**

Dependent Variable: LOG(RAIL_LINES__TOTAL_ROUTE_KM_)

Method: Least Squares

Date: 05/14/25 Time: 19:08

Sample: 1 49

Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.979340	1.195865	-1.655154	0.1046
LOG(URBAN_POPULATION)	0.641322	0.073709	8.700733	0.0000
R-squared	0.616961	Mean dependent var	8.374907	
Adjusted R-squared	0.608811	S.D. dependent var	1.319081	
S.E. of regression	0.825021	Akaike info criterion	2.493143	
Sum squared resid	31.99097	Schwarz criterion	2.570360	
Log likelihood	-59.08201	Hannan-Quinn criter.	2.522439	
F-statistic	75.70276	Durbin-Watson stat	2.073934	
Prob(F-statistic)	0.000000			

Έπειτα, προσθέτουμε στο μοντέλο μας δύο νέες ανεξάρτητες μεταβλητές (βλ. δείκτης οικονομικών ανισοτήτων του Gini και επίπεδα ατμοσφαιρικής ρύπανσης).

Dependent Variable: LOG(RAIL_LINES__TOTAL_ROUTE_KM_)

Method: Least Squares

Date: 05/14/25 Time: 19:11

Sample: 1 49

Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.814663	2.689311	1.790296	0.0801
LOG(URBAN_POPULATION)	0.719023	0.067327	10.67953	0.0000
LOG(GINI_INDEX)	-1.717268	0.782309	-2.195126	0.0334
LOG(PM2_5_AIR_POLLUTION__MEAN_A...	-0.692491	0.216561	-3.197678	0.0025
R-squared	0.719861	Mean dependent var	8.374907	
Adjusted R-squared	0.701185	S.D. dependent var	1.319081	
S.E. of regression	0.721062	Akaike info criterion	2.261925	
Sum squared resid	23.39688	Schwarz criterion	2.416359	
Log likelihood	-51.41716	Hannan-Quinn criter.	2.320517	
F-statistic	38.54478	Durbin-Watson stat	2.221986	
Prob(F-statistic)	0.000000			

9) Να υπολογίσετε τον μερικό συντελεστή προσδιορισμού και να δώσετε την ερμηνεία του.

$$R^2_{partial} = \frac{(R^2_{\text{νέο}} - R^2_{\text{παλιό}})}{1 - R^2_{\text{παλιό}}} = \frac{(0,7198 - 0,6169)}{1 - 0,6169} = \frac{0,1029}{0,3831} = 0,26 \text{ ή } 26\%$$

Ερμηνεία: οι δύο νέοι παράγοντες που προσθέτουμε στο μοντέλο μας ερμηνεύουν το 26% του ανερμηνευτού τμήματος του αρχικού μας μοντέλου.

10) Να διατυπώσετε τον έλεγχο από κοινού σημαντικότητας των συντελεστών β στο πολλαπλό μοντέλο. Να δώσετε την ερμηνεία του αποτελέσματος.

1ο βήμα: διατύπωση υποθέσεων

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ (από κοινού μηδέν)

H_1 : τουλάχιστον ένα β είναι στατιστικά σημαντικό

2ο βήμα: υπολογισμός της στατιστικής τιμής F-stat

F-stat=38,54

3ο βήμα: εξαγωγή συμπεράσματος

Η τιμή p-value που συνοδεύει την F-stat είναι πάρα πολύ μικρή ($<0,05$). Άρα, είμαστε σε θέση να απορρίψουμε την μηδενική υπόθεση της από κοινού ασημαντότητας των συντελεστών β . Τουλάχιστον ένα από τα β είναι στατιστικά σημαντικό.