

ΕΡΩΤΗΣΕΙΣ ΕΠΑΝΑΛΗΨΗΣ:

Ερώτηση 1η: Επιδιώκουμε να χτίσουμε ένα απλό γραμμικό μοντέλο με τη μέθοδο των ελαχίστων τετραγώνων (OLS) που να εξετάζει την πιθανή σχέση ανάμεσα στη συμμετοχή του πληθυσμού στα τραπεζικά επαγγέλματα (ψ =συμμ_τραπ_επαγγ , %) και την παρουσία του πληθυσμού των προτεσταντών (X =ποσ_προτεστ , %) σε ένα πλήθος (i) περιοχών της χώρας. Ποια από τις παρακάτω σχέσεις αντιπροσωπεύει την πραγματική σχέση για του πληθυσμού;

α) $\text{συμμ_τραπ_επαγγ}(i) = \alpha + \beta * \text{ποσ_προτεστ}(i)$

β) $\text{συμμ_τραπ_επαγγ}(i) = \alpha + \beta * \text{ποσ_προτεστ}(i) + \varepsilon(i)$

γ) $\text{ποσ_προτεστ}(i) = \alpha + \beta * \text{συμμ_τραπ_επαγγ}(i)$

δ) $\text{ποσ_προτεστ}(i) = \alpha + \beta * \text{συμμ_τραπ_επαγγ}(i) + \varepsilon(i)$

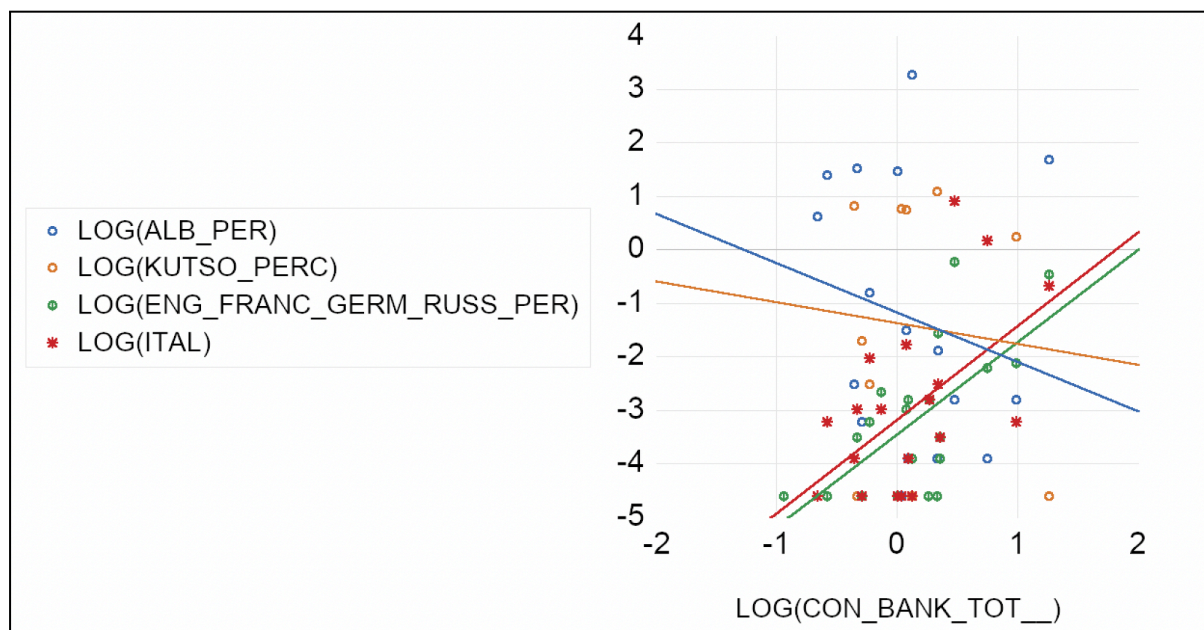
(συνέχεια...) **Ερώτηση 2η:** Ποια από τις παρακάτω προτάσεις είναι η σωστή;

α) Οι πραγματικοί συντελεστές α και β του μοντέλου είναι αδύνατον να υπολογιστούν

β) Οι εκτιμητές των πραγματικών συντελεστών α και β μπορούν να υπολογιστούν

γ) Η σχέση ανάμεσα στους δύο παράγοντες που εξετάζονται στο μοντέλο μας είναι κατ' ανάγκη αιτιακή

Ερώτηση 3η: Δίνεται το ακόλουθο σχήμα (βλ. **διάγραμμα διασποράς λογαριθμικών τιμών**) που εξετάζει τη γραμμική σχέση ανάμεσα στη συγκέντρωση του εργατικού πληθυσμού στα τραπεζικά επαγγέλματα (**con_bank_tot**) και την παρουσία της αλβανικής (**alb_per**), κουτσοβλαχικής (**kutso_perc**), αγγλικής-γαλλικής-γερμανικής-ρωσικής (**eng_franc_germ_russ_perc**), και ιταλικής (**ital**) γλώσσας, αντίστοιχα:



Ποιο είναι το συμπέρασμα που θα βγάζατε κοιτώντας το παραπάνω γράφημα;

Αρχικά, φαίνεται πως τόσο η παρουσία της αγγλικής-γαλλικής-γερμανικής-ρωσικής γλώσσας όσο και η αντίστοιχη της ιταλικής σχετίζονται **θετικά** με την συμμετοχή του εργατικού δυναμικού στα τραπεζικά επαγγέλματα (δείτε τις θετικές κλίσεις για την πράσινη και την κόκκινη γραμμή). Αντίθετα, φαίνεται πως τόσο η παρουσία της αλβανικής γλώσσας όσο και της κουτσοβλαχικής, αντίστοιχα, σχετίζονταν **αρνητικά** με τη συμμετοχή του εργατικού δυναμικού στα τραπεζικά επαγγέλματα (δείτε τις αρνητικές κλίσεις για την μπλε και την πορτοκαλί γραμμή).

Απ' ότι φαίνεται, οι **πράσινες κουκίδες** (βλ. αγγλική-γαλλική-γερμανική-ρωσική γλώσσα) “ματσάρουν” (δηλ. προσαρμόζονται) καλύτερα πάνω στην πράσινη γραμμή. Αντίθετα, οι **μπλε κουκίδες** (βλ. αλβανική γλώσσα) αποκλίνουν, σε αρκετές περιπτώσεις, σημαντικά από την μπλε γραμμή. Αυτό σημαίνει πως η σχέση ανάμεσα στην αλβανική γλώσσα και τη συμμετοχή στα τραπεζικά επαγγέλματα μάλλον δε μπορεί να περιγραφεί ικανοποιητικά μέσα από μια γραμμική σχέση. Λέμε “μάλλον” αφού δεν έχουμε υπόψη μας τα αποτελέσματα του ελέγχου σημαντικότητας για την ύπαρξη του συντελεστή γραμμικής συσχέτισης ...

ΠΡΟΣΟΧΗ: το διάγραμμα διασποράς μου δίνει απλά ενδείξεις ... Αν θέλουμε να δούμε ακριβώς τι συμβαίνει (με κάποιο επίπεδο βεβαιότητας) τότε απαιτείται η ανάλυση συσχέτισης. Δείτε στον παρακάτω πίνακα τα σχετικά αποτελέσματα:

Covariance Analysis: Ordinary
Date: 04/27/25 Time: 14:32
Sample: 1 26
Included observations: 26

Correlation t-Statistic Probability	CON_BANK...	ALB_PER <UTSO_PE...	NG_FRAN...	ITAL
CON_BANK_TOT__	1.000000 ----- -----			
ALB_PER	0.054835 0.269038 0.7902	1.000000 ----- -----		
KUTSO_PERC	0.057025 0.279821 0.7820	-0.162094 -0.804737 0.4289	1.000000 ----- -----	
ENG_FRANC_GE...	0.612376 3.794773 0.0009	0.001266 0.006203 0.9951	-0.159372 -0.790868 0.4368	1.000000 ----- -----
ITAL	0.355461 1.863071 0.0747	-0.076712 -0.376920 0.7095	-0.136725 -0.676162 0.5054	0.807309 6.701905 0.0000
				1.000000 ----- -----

Επιβεβαιώνονται τελικά οι ενδείξεις μας από το διάγραμμα διασποράς;;;

Η πιο έντονα **θετική σχέση** προκύπτει ανάμεσα στην αγγλική-γαλλική-γερμανική-ρωσική γλώσσα και τη συμμετοχή του εργατικού δυναμικού στα τραπεζικά επαγγέλματα (**$\rho=+0,61$**). Μάλιστα, η σχέση αυτή είναι και στατιστικά σημαντική (απορρίπτουμε την $H_0: \rho=0$ ακόμη και στο επίπεδο σημαντικότητας 99%).

Τι συμβαίνει με τις άλλες γλώσσες;;;

Ο συντελεστής γραμμικής συσχέτισης ανάμεσα στην παρουσία της ιταλικής γλώσσας και τη συμμετοχή στα τραπεζικά επαγγέλματα έχει **θετικό πρόσημο**, αλλά χαμηλότερη ένταση σε σύγκριση με την αντίστοιχη σχέση για την αγγλική-γαλλική-γερμανική-ρωσική γλώσσα (**$\rho=+0,35$**). Ο συντελεστής ρ είναι στατιστικά σημαντικός μονάχα στο επίπεδο σημαντικότητας 90% ($\alpha=10\%$). Δεν μπορούμε να απορρίψουμε την $H_0: \rho=0$ όταν διεξάγουμε τον έλεγχο στο $\alpha=5\%$ και $\alpha=1\%$, αντίστοιχα.

Τέλος, σε ότι αφορά την παρουσία της αλβανικής και κουτσοβλαχικής γλώσσας, αντίστοιχα, **δε φαίνεται να υπάρχει κάποια γραμμική σχέση** με τη συμμετοχή στα τραπεζικά επαγγέλματα. Οι συντελεστές γραμμικής συσχέτισης είναι πολύ κοντά στο **$\rho=0$** . Και στις δύο αυτές περιπτώσεις δεν μπορούμε να απορρίψουμε την $H_0: \rho=0 \rightarrow$ έχουμε τιμές του p-value πάρα πολύ μεγάλες ($p\text{-value}>0,05$)

Ερώτηση 4η: Δίνεται ο πίνακας με τους συντελεστές γραμμικής συσχέτισης (ρ) για την παρουσία των προτεσταντών (*protestant*), των ορθόδοξων (*orthodox*), των τραπεζοϋπαλλήλων (*con_bank_emp*), των τραπεζιτών (*con_bank_perc*), του τραπεζικού τομέα (*con_bank_tot*), και των μεσιτών (*con_brok_perc*) στους 26 νομούς της Ελλάδας το έτος 1907 (τις τιμές των συντελεστών συνοδεύουν οι στατιστικές τιμές *t-stat* και τα *p-values* των ελέγχων για την ύπαρξη της στατιστικής σημαντικότητας του συντελεστή γραμμικής συσχέτισης):

Correlation t-Statistic Probability	PROTESTANT	ORTHODOX
PROTESTANT	1.000000 ----- -----	
ORTHODOX	-0.651800 -4.210433 0.0003	1.000000 ----- -----
CON_BANK_EMP_...	0.456833 2.515891 0.0190	-0.197296 -0.985929 0.3340
CON_BANK_PERC	0.042430 0.208050 0.8369	-0.156382 -0.775657 0.4455
CON_BANK_TOT__	0.462872 2.558142 0.0173	-0.352671 -1.846363 0.0772
CON_BROK_PERC	0.539804 3.141500 0.0044	-0.416048 -2.241414 0.0345

Πως θα χαρακτηρίζατε τη σχέση ανάμεσα στους τραπεζοϋπαλλήλους (*con_bank_emp*) και τους προτεστάντες (*protestant*), και ορθόδοξους (*orthodox*), αντίστοιχα;

Ανάμεσα στους τραπεζοϋπαλλήλους και τους προτεστάντες παρατηρείται μια **θετική** (σχετικά έντονη) γραμμική σχέση ($\rho=+0,45$). Μάλιστα, η σχέση αυτή είναι και **στατιστικά σημαντική** αφού η τιμή του *p-value* που συνοδεύει τον συντελεστή συσχέτισης είναι πάρα πολύ μικρή ($p\text{-value}=0,01<0,05$)

$H_0: \rho=0$

$H_1: \rho>0 \text{ ή } \rho<0$

Μπορούμε να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας του συντελεστή γραμμικής συσχέτισης στο επίπεδο σημαντικότητας 95% ($\alpha=5\%$), όχι όμως και στο επίπεδο σημαντικότητας 99% ($\alpha=1\%$).

Ανάμεσα στους τραπεζοϋπαλλήλους και τους ορθόδοξους παρατηρείται μια **αρνητική** (αν και όχι αρκετά έντονη) γραμμική σχέση ($\rho=-0,19$). Ωστόσο, η σχέση αυτή ΔΕΝ βρέθηκε πως ήταν στατιστικά σημαντική αφού η τιμή του *p-value* που συνοδεύει τον συντελεστή συσχέτισης είναι πάρα πολύ μεγάλη ($p\text{-value}=0,3340>0,05$)

$H_0: \rho=0$

$H_1: \rho>0 \text{ ή } \rho<0$

Σ' αυτήν την περίπτωση, ΔΕΝ μπορούμε να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας του συντελεστή γραμμικής συσχέτισης (ακόμη και στο επίπεδο σημαντικότητας 90%, δηλαδή για $\alpha=10\%$, η μηδενική υπόθεση του $\rho=0$ ΔΕΝ μπορεί να απορριφθεί).

Πως θα χαρακτηρίζατε τη σχέση ανάμεσα στους μεσίτες (*con_brok_emp*) και τους προτεστάντες (*protestant*), και ορθόδοξους (*orthodox*), αντίστοιχα;

Ανάμεσα στους μεσίτες και τους προτεστάντες παρατηρείται μια **θετική** (σχετικά έντονη) γραμμική σχέση ($\rho=+0,53$). Μάλιστα, η σχέση αυτή είναι και **στατιστικά σημαντική** αφού η τιμή του p -value που συνοδεύει τον συντελεστή συσχέτισης είναι πάρα πολύ μικρή ($p\text{-value}=0,00<0,05$)

$H_0: \rho=0$

$H_1: \rho>0$ ή $\rho<0$

Μπορούμε να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας του συντελεστή γραμμικής συσχέτισης ακόμη και στο επίπεδο σημαντικότητας 99% ($\alpha=1\%$) αφού $p\text{-value}<0,01$.

Ανάμεσα στους μεσίτες και τους ορθόδοξους παρατηρείται μια **αρνητική** (σχετικά έντονη) γραμμική σχέση ($\rho=-0,41$). Μάλιστα, η σχέση αυτή βρέθηκε πως ήταν στατιστικά σημαντική αφού η τιμή του p -value που συνοδεύει τον συντελεστή συσχέτισης είναι πάρα πολύ μικρή ($p\text{-value}=0,03<0,05$)

$H_0: \rho=0$

$H_1: \rho>0$ ή $\rho<0$

Ωστόσο, ΔΕΝ μπορούμε να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας του συντελεστή γραμμικής συσχέτισης για το επίπεδο σημαντικότητας 99% ($\alpha=1\%$) αφού $p\text{-value}=0,03>0,01$.

Ερώτηση 5η: Δίνεται ο παρακάτω πίνακας με τα ποσοστά των προτεσταντών και τα ποσοστά συμμετοχής του εργατικού δυναμικού στο επάγγελμα του μεσίτη (n=7).

Περιοχές	Ποσοστά προτεσταντών (%)	Μεσίτες (%)
Κέρκυρα	0,60	0,31
Αττική	0,29	0,51
Μέγαρα	0,24	0,04
Σύρος	0,15	0,49
Κέα	0,14	0,06
Πάτρα	0,13	0,71
Ζάκυνθος	0,09	0,23

Να υπολογιστεί ο συντελεστής γραμμικής συσχέτισης:

Λύση:

(Xi-Xmean)	(Yi-Ymean)	(Xi-Xmean)*(Yi-Ymean)	(Xi-Xmean)^2	(Yi-Ymean)^2
0,60-0,23= 0,37	0,31-0,34= -0,03	-0,0111	0,14	0,0009
0,29-0,23= 0,06	0,51-0,34= 0,17	0,0102	0,0036	0,0289
0,24-0,23= 0,01	0,04-0,34= -0,3	-0,003	0,0001	0,09
0,15-0,23= -0,08	0,49-0,34= 0,15	-0,012	0,0064	0,0225
0,14-0,23= -0,09	0,06-0,34= -0,28	0,0252	0,0081	0,0784
0,13-0,23= -0,1	0,71-0,34= 0,37	-0,037	0,01	0,1369
0,09-0,23= -0,14	0,23-0,34= -0,11	0,0154	0,0196	0,0121
		-0,0123	0,1847	0,3697

$$Y_{\text{mean}} = \frac{1}{n} * \sum_{i=1}^n (Y_i) = 1/7 * (0,31 + 0,51 + \dots + 0,71 + 0,23) = 0,34$$

$$X_{\text{mean}} = \frac{1}{n} * \sum_{i=1}^n (X_i) = 1/7 * (0,60 + 0,29 + \dots + 0,13 + 0,09) = 0,23$$

Ο συντελεστής συσχέτισης θα είναι: **$\rho = -0,0123 / (0,2631) = -0,05$**

(χρησιμοποιούμε τον σχετικό τύπο που βρίσκεται στις διαφάνειες της 1ης συνάντησης)

Ερώτηση 6η: Δίνεται ο πίνακας με τα αποτελέσματα της μεθόδου των ελαχίστων τετραγώνων (OLS) για τον προσδιορισμού του απλού γραμμικού μοντέλου της μορφής:

$$\text{con_trade_per}(i) = \alpha + \beta * \text{protestant}(i) + \varepsilon(i)$$

όπου:

con_trade_per = συμμετοχή του εργατικού δυναμικού στο εμπόριο (%)

protestant = παρουσία προτεσταντών (%)

i=1,2,3, ... , 26 (νομοί)

Dependent Variable: CON_TRADE_PER				
Method: Least Squares				
Date: 04/27/25 Time: 16:17				
Sample: 1 26				
Included observations: 26				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.759843	0.559662	15.65202	0.0000
PROTESTANT	14.72547	4.382579	3.360000	0.0026
R-squared	0.319913	Mean dependent var		9.473462
Adjusted R-squared	0.291576	S.D. dependent var		3.136893
S.E. of regression	2.640258	Akaike info criterion		4.853433
Sum squared resid	167.3030	Schwarz criterion		4.950210
Log likelihood	-61.09463	Hannan-Quinn criter.		4.881302
F-statistic	11.28960	Durbin-Watson stat		1.619620
Prob(F-statistic)	0.002601			

Να δώσετε την ερμηνεία των συντελεστών της παλινδρόμησης:

Αν αυξηθεί κατά 1 μονάδα (δηλ κατά 1%) η παρουσία των προτεσταντών τότε θα αυξηθεί και η συμμετοχή του εργατικού πληθυσμού στο εμπόριο κατά περίπου **14,7** μονάδες (δηλ κατά 14,7%).

Η εκτίμηση για τον σταθερό όρο της σχέσης (c) παίρνει την τιμή **8,75** → αυτό σημαίνει πως ακόμη και αν η παρουσία των προτεσταντών είναι μηδενική (0%) θα υπάρχει ένα ελάχιστο επίπεδο παρουσίας εργαζομένων στο εμπόριο ίση με 8,75 μονάδες (δηλ 8,75%). Στην περίπτωση αυτή έχει νόημα η ερμηνεία του εκτιμητή του σταθερού όρου.

Να δώσετε την ερμηνεία του συντελεστή προσδιορισμού:

Η τιμή του συντελεστή προσδιορισμού (R-squared) είναι ίση με **0,31**. Αυτό σημαίνει πως περίπου το 31% της συνολικής μεταβλητότητας των τιμών της εξαρτημένης μεταβλητής μας (συμμετοχή στο εμπόριο) μπορεί να εξηγηθεί από την

μεταβλητότητα (βλ. κίνηση) των τιμών της ανεξάρτητης μεταβλητής μας (παρουσία προτεσταντών).

Να διατυπώσετε την μηδενική και την εναλλακτική υπόθεση του ελέγχου για την στατιστική σημαντικότητα των συντελεστών α και β της πραγματικής σχέσης που συνδέει την εξαρτημένη με την ανεξάρτητη μεταβλητή:

$H_0: \alpha = 0$ (μη στατιστικά σημαντικός συντελεστής)

$H_1: \alpha > 0$ ή $\alpha < 0$ (στατιστικά σημαντικός συντελεστής)

$H_0: \beta = 0$ (μη στατιστικά σημαντικός συντελεστής)

$H_1: \beta > 0$ ή $\beta < 0$ (στατιστικά σημαντικός συντελεστής)

Με βάση τα αποτελέσματα του πίνακα που δίνεται παραπάνω να αξιολογήσετε την στατιστική σημαντικότητα των συντελεστών α και β στο επίπεδο βεβαιότητας 95% και 99%:

Τόσο για τον συντελεστή α όσο και για τον β οι στατιστικές τιμές του ελέγχου της στατιστικής σημαντικότητας **t-stat** είναι πολύ μεγάλες. Ωστόσο, για να απαντήσουμε με σαφήνεια απαιτείται η σύγκριση των στατιστικών τιμών με τις αντίστοιχες κριτικές τιμές από τους πίνακες της κατανομής της t . Για να αποφύγουμε αυτή τη σύγκριση χρησιμοποιούμε τον κανόνα για τις τιμές του p-value:

Αν $p\text{-value} < 0,05$ (5%) → τότε μπορούμε να απορρίψουμε την μηδενική υπόθεση του ελέγχου της στατιστικής σημαντικότητας (με βεβαιότητα 95%). Αυτό σημαίνει πως ο συντελεστής είναι στατιστικά σημαντικός.

Αν $p\text{-value} > 0,05$ (5%) → τότε ΔΕΝ μπορούμε να απορρίψουμε την μηδενική υπόθεση του ελέγχου της στατιστικής σημαντικότητας (με βεβαιότητα 95%). Αυτό σημαίνει πως ο συντελεστής ΔΕΝ είναι στατιστικά σημαντικός.

Με βάση τα στοιχεία του παραπάνω πίνακα οι τιμές των p-value είναι πάρα πολύ μικρές τόσο για τον έλεγχο της σημαντικότητας του συντελεστή α όσο και για τον αντίστοιχο του β . Αυτό σημαίνει πως οι συντελεστές α και β , αντίστοιχα, είναι στατιστικά σημαντικοί. Επειδή οι τιμές των $p\text{-value} < 0,01$ (1%) είμαστε σε θέση να απορρίψουμε την μηδενική υπόθεση της μη σημαντικότητας των συντελεστών α και β ακόμη και με βεβαιότητα 99% ($\alpha=1\%$).

Ερώτηση 7η (sos): Δίνεται ο παρακάτω πίνακας με τις τιμές για τα ποσοστά του αγράμματου πληθυσμού και τη συμμετοχή του εργατικού δυναμικού στη βιομηχανία.

Περιοχές	Αγράμματοι (Ψ, %)	Συμμετοχή στη Βιομηχανία (Χ, %)
Μήλος	37	46
Βόλος	43	39
Κέα	34	36
Παρνασσός	33	21
Ύδρα	42	14
Ευρυτανία	25	10
Βόνιτσα	26	9

7.1 Να υπολογίσετε με τη μέθοδο των ελαχίστων τετραγώνων τους εκτιμητές των συντελεστών α και β της σχέσης: $\text{αγράμματοι}(i) = \alpha + \beta * \text{βιομηχανία}(i) + \epsilon_i$

Λύση:

Yi	Xi	(Xi * Yi)	(Xi)^2
37	46	1.702	2.116
43	39	1.677	1.521
34	36	1.224	1.296
33	21	693	441
42	14	588	196
25	10	250	100
26	9	234	81
240 (sum)	175 (sum)	6.368 (sum)	5.751 (sum)

$$b = [7 * (6.368) - (175 * 240)] / [(7 * 5.751) - (175 * 175)]$$

$$= (44.576 - 42.000) / (40.257 - 30.625) = 2.576 / 9.632 = 0,2674$$

$$a = (240/7) - (175/7) * 0,0751 = 34,2857 - 25 * 0,2674 = 34,2857 - 6,6860 = 27,59$$

(χρησιμοποιούμε τους τύπους που δώσαμε στο υλικό της 4ης συνάντησης)

7.2 Να δώσετε την ερμηνεία των συντελεστών της παλινδρόμησης που υπολογίσατε στο προηγούμενο ερώτημα:

Αν η συμμετοχή του εργατικού δυναμικού στη βιομηχανία αυξηθεί κατά 1 μονάδα (δηλ κατά 1%) τότε οι αγράμματοι θα αυξηθούν κατά περίπου 0,2674 μονάδες (δηλ κατά 0,267%).
Ακόμη και αν η συμμετοχή στην βιομηχανία είναι μηδέν θα υπάρχει ένα επίπεδο αγράμματων ίσο με 27,59 μονάδες (δηλ 27,59%).

7.3 Να υπολογίσετε τα σφάλματα για τις δεδομένες τιμές της ανεξάρτητης μεταβλητής:

Λύση:

Yi	Xi	Y (fitted)	residuals
37	46	=27,59+0,2674*(46)=39,89	=(37-39,89)=-2,89
43	39	=27,59+0,2674*(39)=38,01	=(43-38,01)=4,99
34	36	=27,59+0,2674*(36)=37,21	=(34-37,21)=-3,21
33	21	=27,59+0,2674*(21)=33,20	=(33-33,20)=-0,2
42	14	=27,59+0,2674*(14)=31,33	=(42-31,33)=10,67
25	10	=27,59+0,2674*(10)=30,26	=(25-30,26)=-5,26
26	9	=27,59+0,2674*(9)=29,99	=(26-29,99)=-3,99

7.4 Να υπολογίσετε τον συντελεστή προσδιορισμού και στη συνέχεια να δώσετε την ερμηνεία του:

$$R\text{-squared} = SSR / SST = \sum (Y_i \text{ fitted} - Y_{\text{mean}})^2 / \sum (Y_i - Y_{\text{mean}})^2 = 0,33 \text{ (33\%)}$$

(χρησιμοποιούμε τους τύπους που δώσαμε στο υλικό της 4ης συνάντησης)

Yi	Xi	(Y_fitted - Y_mean)^2	(Y_fitted - Y_mean)^2
----	----	-----------------------	-----------------------

37	46	$=(39,89-34,28)^2=31,47$	$=(37-34,28)^2=7,39$
43	39	$=(38,01-34,28)^2=14,59$	$=(43-34,28)^2=76,03$
34	36	$=(37,21-34,28)^2=8,58$	$=(34-34,28)^2=0,078$
33	21	$=(33,20-34,28)^2=1,16$	$=(33-34,28)^2=1,63$
42	14	$=(31,33-34,28)^2=8,70$	$=(42-34,28)^2=59,59$
25	10	$=(30,26-34,28)^2=15,36$	$=(25-34,28)^2=86,11$
26	9	$=(29,99-34,28)^2=18,40$	$=(26-34,28)^2=68,55$
Y_mean=34,28		SSR=98,26	SST=296,37

R-squared=0,33 (33%) → περίπου το 33% της συνολικής μεταβλητότητας των τιμών του αγράμματος πληθυσμού (δηλ της εξαρτημένης μεταβλητής μας) μπορεί να εξηγηθεί από τη μεταβλητότητα των τιμών της ανεξάρτητης μεταβλητής μας (δηλ της συμμετοχής του εργατικού δυναμικού στην βιομηχανία). Το υπόλοιπο κομμάτι (βλ. $1-0,33 = 0,67$ ή 67%) εξηγείται από άλλους παράγοντες που δεν συμπεριλαμβάνονται στο μοντέλο μας.

7.5 Να υπολογίσετε το τυπικό σφάλμα της παλινδρόμησης:

(χρησιμοποιούμε τον τύπο που δώσαμε στο υλικό της 11ης συνάντησης)

Ισχύει: $SST = SSR + SSE \rightarrow SSE = SST - SSR = 198,11$

τυπικό σφάλμα παλινδρόμησης = τετραγωνική ρίζα ($SSE / n-k-1$)
= τετραγωνική ρίζα ($198,11 / 7-2$) = 6,29

7.6 Να διατυπώσετε τον έλεγχο της στατιστικής σημαντικότητας για τον συντελεστή β, να υπολογίσετε την στατιστική τιμή ελέγχου t-stat, και να δώσετε την ερμηνεία του αποτελέσματος:

(χρησιμοποιούμε τους τύπους που δώσαμε στο υλικό της 7ης και 11ης συνάντησης)

1ο βήμα: Διατύπωση Υποθέσεων του Ελέγχου

Ho (μηδενική υπόθεση): $\beta = 0 \rightarrow$ ο συντελεστής δεν είναι στατιστικά σημαντικός

H1 (εναλλακτική υπόθεση): $\beta > 0$ ή $\beta < 0 \rightarrow$ ο συντελεστής είναι στατιστικά σημαντικός

2ο βήμα: Υπολογισμός της Στατιστικής Τιμής του Ελέγχου (t-stat)

t-stat = $b / s.e(b) = 0,2674 / 0,17 = 1,577$

όπου

b: ο εκτιμητής του συντελεστή β (υπολογίστηκε στο πρώτο ερώτημα της άσκησης)

s.e(b): τυπικό σφάλμα του εκτιμητή b (δείτε το υλικό της 11ης συνάντησης)

$s.e(b) = 6,29 * 0,027 = 0,17$

Yi	Xi	(X _i - X _{mean}) ²
37	46	=(46-25) ² =441
43	39	=(39-25) ² =196
34	36	=(36-25) ² =121
33	21	=(21-25) ² =16
42	14	=(14-25) ² =121
25	10	=(10-25) ² =225
26	9	=(9-25) ² =256
	X_{mean}=25	1.376 (sum)

3ο βήμα: Σύγκριση της Στατιστικής Τιμής Ελέγχου (t-stat) με τις Κριτικές Τιμές (t-crit)

t-stat = 1,577

Η κριτική τιμή δίνεται από τους πίνακες της κατανομής t → **t_{crit} (n-2) = 2,57**

Κανόνας: αν η απόλυτη τιμή της στατιστικής τιμής ελέγχου είναι μεγαλύτερη από την κριτική τιμή τότε μπορούμε να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας.

Στην περίπτωσή μας: **t-stat = 1,57 < 2,57 → ΔΕΝ** μπορούμε να απορρίψουμε την H₀ (ο συντελεστής β **ΔΕΝ** φαίνεται να είναι στατιστικά σημαντικός).

ΔΕΝ μπορούμε να απορρίψουμε την H₀ ακόμη και για α=10%.

7.7 Να κατασκευάσετε ένα διάστημα εμπιστοσύνης για τον συντελεστή β για α=5% και α=1%, αντίστοιχα, και να δώσετε την ερμηνεία του αποτελέσματος:

(χρησιμοποιούμε τους τύπους που δώσαμε στο υλικό της 10ης συνάντησης)

$$P[b - t_{n-2} * s.e(b) < \beta < b + t_{n-2} * s.e(b)] = 95\%$$

$$P[0,2674 - 2,57 * 0,17 < \beta < 0,2674 + 2,57 * 0,17] = 95\%$$

$$P(-0,1695 < \beta < 0,7043) = 95\%$$

$$P[b - t_{n-2} * s.e(b) < \beta < b + t_{n-2} * s.e(b)] = 99\%$$

$$P[0,2674 - 4,03 * 0,17 < \beta < 0,2674 + 4,03 * 0,17] = 99\%$$

$$P(-0,4177 < \beta < 0,9525) = 99\%$$

Ο πραγματικός συντελεστής β βρίσκεται μέσα στο διάστημα $[-0,169, +0,704]$ με βεβαιότητα 95% ($\alpha=5\%$).

Ο πραγματικός συντελεστής β βρίσκεται μέσα στο διάστημα $[-0,417, +0,952]$ με βεβαιότητα 99% ($\alpha=1\%$).

Ερώτηση 8η (sos): Δίνονται οι παρακάτω πίνακες με τα αποτελέσματα των παλινδρομήσεων με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων (OLS):

$$\text{bank_cler_cont_perc}_i = \alpha + \beta * \text{prot_perc}_i + \epsilon_i, k=1 \quad (1)$$

$$\text{bank_cler_cont_perc}_i = \alpha + \beta * \text{prot_perc}_i + \text{cit_franc}_i + \epsilon_i, k=2 \quad (2)$$

όπου

bank_cler_cont_perc_i: τραπεζοϋπάλληλοι (ως % επί του εργατικού δυναμικού)

prot_perc_i: προτεστάντες (ως % επί του συνολικού πληθυσμού)

cit_franc_i: άτομα με γαλλική ιθαγένεια (ως % επί του συνολικού πληθυσμού)

$i=1,2,3,4, \dots, 69$: ελληνικές επαρχίες (*provinces*)

Πως θα αξιολογούσατε στατιστικά τη χρήση του παράγοντα “προτεστάντες” (*prot_perc*) ως πιθανό προσδιοριστικό παράγοντα της παρουσίας των τραπεζοϋπαλλήλων στις ελληνικές επαρχίες (*bank_cler_cont_perc*);

Dependent Variable: BANK_CLER_CONT_PERC				
Method: Least Squares				
Date: 04/28/25 Time: 14:31				
Sample: 1 69				
Included observations: 69				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.051074	0.011638	4.388713	0.0000
PROT_PERC	0.444085	0.126529	3.509740	0.0008
R-squared	0.155302	Mean dependent var		0.064783
Adjusted R-squared	0.142694	S.D. dependent var		0.098348
S.E. of regression	0.091061	Akaike info criterion		-1.926008
Sum squared resid	0.555576	Schwarz criterion		-1.861251
Log likelihood	68.44727	Hannan-Quinn criter.		-1.900317
F-statistic	12.31828	Durbin-Watson stat		2.131428
Prob(F-statistic)	0.000807			

Απάντηση: Ο παράγοντας “προτεστάντες” (*prot_perc*) φαίνεται πως είναι **στατιστικά σημαντικός** για την ερμηνεία των τιμών της εξαρτημένης μεταβλητής μας (βλ. *τραπεζοϋπάλληλοι*). Η τιμή του p-value που συνοδεύει την στατιστικά τιμή ελέγχου (t-stat) για την στατιστική σημαντικότητα του συντελεστή β είναι πάρα πολύ μικρή ($p\text{-value} < 0,01$). Αυτό σημαίνει πως είμαστε σε θέση, ακόμη και με βεβαιότητα 99% ($\alpha=1\%$), να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας του συντελεστή β. Σε ότι αφορά όμως την **ερμηνευτική ικανότητα** του μοντέλου μας, αυτή θα μπορούσε να χαρακτηριστεί ως σχετικά χαμηλή. Πιο συγκεκριμένα, η τιμή του συντελεστή προσδιορισμού (R-squared) είναι ίση με **0,1553** → Αυτό σημαίνει πως μόλις το **15,54%** της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής μας (Ψ) μπορεί να εξηγηθεί από την μεταβλητότητα των τιμών της ανεξάρτητης μεταβλητής μας (X). Κατά συνέπεια, ένα μεγάλο κομμάτι της συνολικής μεταβλητότητας της Ψ, περίπου το $100\% - 15,54\% = 84,46\%$, παραμένει ανερμήνευτο. Το κομμάτι αυτό θα εξηγείται από άλλους παράγοντες που δε λαμβάνει υπόψη το απλό γραμμικό μοντέλο μας ($k=1$).

Πως θα αξιολογούσατε στατιστικά την προσθήκη ενός ακόμη παράγοντα στο μοντέλο μας, και πιο συγκεκριμένα του παράγοντα “γαλλική ιθαγένεια” (*cit_franc*);

Dependent Variable: BANK_CLER_CONT_PERC

Method: Least Squares

Date: 04/28/25 Time: 14:35

Sample: 1 69

Included observations: 69

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.044628	0.011687	3.818767	0.0003
PROT_PERC	0.308500	0.137523	2.243272	0.0282
CIT_FRANC	0.452804	0.205232	2.206306	0.0308
R-squared	0.213323	Mean dependent var		0.064783
Adjusted R-squared	0.189484	S.D. dependent var		0.098348
S.E. of regression	0.088542	Akaike info criterion		-1.968183
Sum squared resid	0.517415	Schwarz criterion		-1.871048
Log likelihood	70.90233	Hannan-Quinn criter.		-1.929647
F-statistic	8.948589	Durbin-Watson stat		2.126057
Prob(F-statistic)	0.000364			

Απάντηση: Ο νέος παράγοντας (βλ. “γαλλική ιθαγένεια”) που προσθέτουμε στο μοντέλο μας φαίνεται πως ήταν **στατιστικά σημαντικός** για την επεξήγηση των τιμών της παρουσίας των τραπεζοϋπαλλήλων (Ψ) στις 69 επαρχίες της χώρας. Πιο συγκεκριμένα, η τιμή του p-value για τον έλεγχο της στατιστικής σημαντικότητας του συντελεστή β2 είναι πολύ μικρή (**p-value=0,03<0,05**). Αυτό σημαίνει πως είμαστε σε θέση να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας του συντελεστή β2 με επίπεδο βεβαιότητας το 95% (α=5%). Ωστόσο, στο επίπεδο βεβαιότητας 99% (α=1%) αποτυγχάνουμε να απορρίψουμε την H₀ (**p-value=0,03>0,01**). Ακόμη, κοιτώντας την τιμή του συντελεστή προσδιορισμού (*R-squared*), αλλά και την αντίστοιχη του διορθωμένου συντελεστή προσδιορισμού (*Adjusted R-squared*), θα λέγαμε πως η **ερμηνευτική ικανότητα** του μοντέλου μας βελτιώθηκε σημαντικά. Με την προσθήκη της νέας ανεξάρτητης μεταβλητής είμαστε σε θέση να εξηγούμε ένα μεγαλύτερο κομμάτι της συνολικής μεταβλητότητας των τιμών της εξαρτημένης μεταβλητής μας (Ψ). Με τη βοήθεια του υπολογισμού του **μερικού συντελεστή προσδιορισμού** (βλ.υλικό 13ης συνάντησης) μπορούμε να δούμε τη δυναμική αυτής της προσθήκης: **(R-squared_νέο - R-squared_παλιό) / (1-R_squared_παλιό) = (0,21-0,15)/(1-0,15) = 7,05%**. Ερμηνεία: η προσθήκη του νέου παράγοντα (βλ. *γαλλική ιθαγένεια*) οδήγησε στην αύξηση της επεξήγησης του αρχικά ανερμήνευτου κομματιού της συνολικής μεταβλητότητας της Ψ κατά περίπου 7%.

Να διατυπώσετε την μηδενική και την εναλλακτική υπόθεση του ελέγχου της από κοινού στατιστικής σημαντικότητας των συντελεστών β στο πολλαπλό μοντέλο μας (k=2). Να αξιολογήσετε στατιστικά το αποτέλεσμα του ελέγχου.

Απάντηση: Ο έλεγχος της **από κοινού σημαντικότητας** των συντελεστών β της παλινδρόμησης με δύο ανεξάρτητες μεταβλητές (k=2) έχει ως εξής:

1ο βήμα: Διατύπωση της μηδενικής και της εναλλακτικής υπόθεσης

$H_0: \beta_1 = \beta_2 = 0$ (όλα τα β είναι από κοινού μηδέν)

H_1 : τουλάχιστον ένα από τα β_1 και β_2 είναι στατιστικά σημαντικό

2ο βήμα: Υπολογισμός της στατιστικής ελέγχου (F-stat)

Η διαδικασία αυτή, γνωστή ως ANOVA (βλ. υλικό 7ης συνάντησης), αποδίδει την τιμή της F-stat. Στην δική μας περίπτωση, η τιμή της F-stat δίνεται έτοιμη στα αποτελέσματα του πίνακα για το μοντέλο με $k=2 \rightarrow \mathbf{F-stat = 8,94}$

3ο βήμα: Σύγκριση της στατιστικής τιμής ελέγχου με τις κριτικές τιμές από τους πίνακες της κατανομής F. Στην δική μας περίπτωση, χρησιμοποιούμε την τιμή του p-value από τους πίνακες που συνοδεύει την F-stat $\rightarrow \mathbf{p-value = 0,000 < 0,01}$ (1%). Πολύ μικρή τιμή του p-value συνεπάγεται την απόρριψη της μηδενικής υπόθεσης του ελέγχου. Συμπέρασμα; τουλάχιστον ένα από τα β_1 και β_2 είναι **στατιστικά σημαντικό**. Άρα, το πολυμεταβλητό μοντέλο μας ($k=2$) έχει κάποια στατιστικά σημαντική ερμηνευτική ικανότητα.

Ερώτηση 9η (sos sos): Δίνεται ο παρακάτω πίνακας με τα ετήσια στοιχεία μιας επιχείρησης (το λύσαμε αναλυτικά μέσα στην τάξη):

έτη	Ψ : πωλήσεις (σε χιλ. ευρώ)	X_1 : δαπάνες διαφήμισης (σε χιλ. ευρώ)	X_2 : ικανοποίηση εργαζομένων (0-10)
-----	-------------------------------------	---	---

2000	30	5	4
2001	62	9	6
2002	85	12	9
2003	72	10	8
2004	68	9	7
2005	97	15	9
2006	145	25	10

Να βρεθεί η τιμή του συντελεστή γραμμικής συσχέτισης που συνδέει τις πωλήσεις με τις δαπάνες για διαφήμιση:

Λύση:

(Xi-Xmean)	(Yi-Ymean)	(Xi-Xmean)*(Yi-Ymean)	(Xi-Xmean)^2	(Yi-Ymean)^2
5-12,14=-7,14	30-79,85=-49,85	355,929	50,9796	2.485,0225
9-12,14=-3,14	62-79,85=-17,85	56,049	9,8596	318,6225
12-12,14=-0,14	85-79,85=5,15	-0,721	0,0196	26,5225
10-12,14=-2,14	72-79,85=-7,85	16,799	4,5796	61,6225
9-12,14=-3,14	68-79,85=-11,85	37,209	9,8596	140,4225
15-12,14=2,86	97-79,85=17,15	49,049	8,1796	294,1225
25-12,14=12,86	145-79,85=65,15	837,829	165,3796	4.244,5225
		1.352,143	248,8572	7.570,8575

$$Y_{\text{mean}} = \frac{1}{n} * \sum_{i=1}^n (Y_i) = 1/7 * (30 + 62 + \dots + 97 + 145) = 79,85$$

$$X_{\text{mean}} = \frac{1}{n} * \sum_{i=1}^n (X_i) = 1/7 * (5 + 9 + \dots + 15 + 25) = 12,14$$

Ο συντελεστής συσχέτισης θα είναι: **$\rho = 1.352,143 / (1.372,612) = +0,9850$**
(χρησιμοποιούμε τον σχετικό τύπο που βρίσκεται στις διαφάνειες της 1ης συνάντησης)

Να υπολογίσετε τους συντελεστές α και β του απλού γραμμικού μοντέλου της μορφής: $\text{πωλήσεις}(t) = \alpha + \beta * \text{δαπάνες_διαφήμισης}(t) + \varepsilon(t)$

Λύση:

Yi	Xi	(Xi * Yi)	(Xi)^2
----	----	-----------	--------

30	5	150	25
62	9	558	81
85	12	1.020	144
72	10	720	100
68	9	612	81
97	15	1.455	225
145	25	3.625	625
559 (sum)	85 (sum)	8.140 (sum)	1.281 (sum)

$$b = [7 * (8.140) - (559*85)] / [(7 * 1.281) - (85*85)]$$

$$= (56.980 - 47.515) / (8.967 - 7.225) = 9.465 / 1.742 = 5,4334$$

$$a = (559/7) - (85/7)*5,4334 = 79,8571 - 12,1428*5,4334 = 79,8571 - 65,977 = 13,8801$$

(χρησιμοποιούμε τους τύπους που δώσαμε στο υλικό της 4ης συνάντησης)

Να αξιολογήσετε στατιστικά την σημαντικότητα του συντελεστή β του απλού γραμμικού μοντέλου της μορφής: $\text{πωλήσεις}(t) = \alpha + \beta * \text{δαπάνες_διαφήμισης}(t) + \varepsilon(t)$

(χρησιμοποιούμε τους τύπους που δώσαμε στο υλικό της 7ης και 11ης συνάντησης)

1ο βήμα: Διατύπωση Υποθέσεων του Ελέγχου

Ho (μηδενική υπόθεση): $\beta = 0 \rightarrow$ ο συντελεστής δεν είναι στατιστικά σημαντικός

H1 (εναλλακτική υπόθεση): $\beta > 0$ ή $\beta < 0 \rightarrow$ ο συντελεστής είναι στατιστικά σημαντικός

2ο βήμα: Υπολογισμός της Στατιστικής Τιμής του Ελέγχου (t-stat)

$$t\text{-stat} = b / s.e(b) = 5,4334 / s.e(b) = ; ; ;$$

όπου

b: ο εκτιμητής του συντελεστή β (υπολογίστηκε στο προηγούμενο ερώτημα)

s.e(b): τυπικό σφάλμα του εκτιμητή b (δείτε το υλικό της 11ης συνάντησης) = ; ; ;

Σφάλμα παλινδρόμησης: τετραγωνική ρίζα ($SSE / n-k-1$) = ; ; ; (6,364)

Yi	Xi	Y (fitted)	residuals	residuals^2
30	5	=13,88+5,433*(5)=41,045	=(30-41,04)=-11,04	121,8816
62	9	=13,88+5,433*(9)=62,777	=(62-62,77)=-0,77	0,5929

85	12	=13,88+5,433*(12)=79,076	=(85-79,07)=5,93	35,1649
72	10	=13,88+5,433*(10)=68,21	=(72-68,21)=3,79	14,3641
68	9	=13,88+5,433*(9)=62,777	=(68-62,77)=5,23	27,3529
97	15	=13,88+5,433*(15)=95,375	=(97-95,37)=1,63	2,6569
145	25	=13,88+5,433*(25)=145,70	=(145-145,7)=-0,7	0,49
				SSE=202,503

Yi	Xi	(X _i -X _{mean}) ²
30	5	=(5-12,24) ² =50,979
62	9	=(9-12,24) ² =10,497
85	12	=(12-12,24) ² =0,0576
72	10	=(10-12,24) ² =5,0176
68	9	=(9-12,24) ² =10,497
97	15	=(15-12,24) ² =7,617
145	25	=(25-12,24) ² =162,81
	X_{mean}=25	247,4828 (sum)

$$s.e(b) = 6,364 * 0,0636 = 0,4047$$

$$t\text{-stat} = b / s.e(b) = 5,4334 / 0,4047 = 13,42$$

3ο βήμα: Σύγκριση της Στατιστικής Τιμής Ελέγχου (t-stat) με τις Κριτικές Τιμές (t-crit)

$$t\text{-stat} = 13,42$$

Η κριτική τιμή δίνεται από τους πίνακες της κατανομής $t \rightarrow t_{crit}(n-2) = 2,57$

Κανόνας: αν η απόλυτη τιμή της στατιστικής τιμής ελέγχου είναι μεγαλύτερη από την κριτική τιμή τότε μπορούμε να απορρίψουμε την μηδενική υπόθεση της μη στατιστικής σημαντικότητας.

Στην περίπτωσή μας: $t\text{-stat} = 13,42 > 2,57 \rightarrow$ Άρα, μπορούμε να απορρίψουμε την H_0 (ο συντελεστής β είναι στατιστικά σημαντικός στο $\alpha=5\%$).

Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για τον συντελεστή β του απλού γραμμικού μοντέλου της μορφής:

$$πωλήσεις(t) = \alpha + \beta * \text{δαπάνες_διαφήμισης}(t) + \varepsilon(t)$$

$$P[b - t_{n-2} * s.e(b) < \beta < b + t_{n-2} * s.e(b)] = 95\%$$

$$P[5,4334 - 2,57 * 0,4047 < \beta < 5,4334 + 2,57 * 0,4047] = 95\%$$

$$P(4,3933 < \beta < 6,4734) = 95\%$$

Προσθέτουμε στο μοντέλο μια ακόμη ανεξάρτητη μεταβλητή, και πιο συγκεκριμένα τον παράγοντα “ικανοποίηση των εργαζομένων”. Τρέχουμε με τη μέθοδο των ελαχίστων τετραγώνων το νέο μοντέλο ($k=2$) της μορφής:

$$\text{πωλήσεις}(t) = \alpha + \beta_1 * \text{δαπάνες_διαφήμισης}(t) + \beta_2 * \text{ικανοποίηση_εργαζομένων}(t) + \varepsilon(t)$$

Δίνονται οι παρακάτω πίνακες:

ANOVA table						
Source	DF	Sum of Square	Mean Square	F Statistic	P-value	
Regression (between $\hat{y}_i$ and $\bar{y}$)	2	7533.772066	3766.886033	406.296689	0.0000239943	
Residual (between y_i and $\hat{y}_i$)	4	37.085077	9.271269			
Total (between y_i and $\bar{y}$)	6	7570.857143	1261.809524			

Coefficient Table Iteration 1 (adjusted R-squared = 0.993)							
	Coeff	SE	t-stat	lower $t_{0.025}(4)$	upper $t_{0.975}(4)$	Stand Coeff	p-value
b	-7.079684	5.347453	-1.323936	-21.926592	7.767225	0	0.256114
Δαπάνες_διαφ_χιλ_ευρώ	4.141316	0.346434	11.954131	3.179461	5.103171	0.750829	0.000280598
ικανοποίηση_εργαζομένων	4.840489	1.077726	4.491393	1.848243	7.832735	0.282101	0.0108942

Να αξιολογήσετε στατιστικά τη σημαντικότητα του παράγοντα “ικανοποίηση των εργαζομένων”:

Ο παράγοντας “ικανοποίηση των εργαζομένων” είναι στατιστικά σημαντικός αφού η τιμή του p-value του ελέγχου της στατιστικής σημαντικότητας είναι ίση με **0,01089 < 0,05 (5%)**. Αυτό σημαίνει πως είμαστε σε θέση να απορρίψουμε τη μηδενική υπόθεση → Άρα, ο συντελεστής β_2 είναι **στατιστικά σημαντικός** για $\alpha=5\%$. Ωστόσο, ΔΕΝ είμαστε σε θέση να απορρίψουμε την H_0 στο επίπεδο βεβαιότητας $\alpha=1\%$ → **0,01089 > 0,01 (1%)**

Με τη βοήθεια των στοιχείων του πίνακα ANOVA (δίνεται παραπάνω) να υπολογίσετε την τιμή του συντελεστή προσδιορισμού, και να δώσετε την ερμηνεία του αποτελέσματος:

$$R\text{-squared} = SSR / SST = 7533,77 / 7570,85 = \mathbf{0,9951 \text{ ή } 99,51\%}$$

Ερμηνεία: περίπου το **99,5%** του συνόλου της μεταβλητότητας των τιμών των πωλήσεων της επιχείρησης (Ψ) μπορεί να εξηγηθεί από την μεταβλητότητα των τιμών των πωλήσεων (X_1) και του βαθμού ικανοποίησης των εργαζομένων (X_2). Να διατυπώσετε την μηδενική και την εναλλακτική υπόθεση του ελέγχου της από κοινού στατιστικής σημαντικότητας των συντελεστών β στο πολλαπλό μοντέλο μας ($k=2$). Να αξιολογήσετε στατιστικά το αποτέλεσμα του ελέγχου.

Απάντηση: Ο έλεγχος της **από κοινού σημαντικότητας** των συντελεστών β της παλινδρόμησης με δύο ανεξάρτητες μεταβλητές ($k=2$) έχει ως εξής:

1ο βήμα: Διατύπωση της μηδενικής και της εναλλακτικής υπόθεσης

$H_0: \beta_1 = \beta_2 = 0$ (όλα τα β είναι από κοινού μηδέν)

H_1 : τουλάχιστον ένα από τα β_1 και β_2 είναι στατιστικά σημαντικό

2ο βήμα: Υπολογισμός της στατιστικής ελέγχου (F-stat)

Η διαδικασία αυτή, γνωστή ως ANOVA (βλ. υλικό 7ης συνάντησης), αποδίδει την τιμή της F-stat. Στην δική μας περίπτωση, η τιμή της F-stat δίνεται έτοιμη στα αποτελέσματα του πίνακα για το μοντέλο με $k=2 \rightarrow \mathbf{F-stat = 406,29}$

3ο βήμα: Σύγκριση της στατιστικής τιμής ελέγχου με τις κριτικές τιμές από τους πίνακες της κατανομής F. Στην δική μας περίπτωση, χρησιμοποιούμε την τιμή του p-value από τους πίνακες που συνοδεύει την F-stat $\rightarrow \mathbf{p-value = 0,0000 < 0,01}$ (1%). Πολύ μικρή τιμή του p-value συνεπάγεται την απόρριψη της μηδενικής υπόθεσης του ελέγχου. Συμπέρασμα; τουλάχιστον ένα από τα β_1 και β_2 είναι **στατιστικά σημαντικό**. Άρα, το πολυμεταβλητό μοντέλο μας ($k=2$) έχει κάποια στατιστικά σημαντική ερμηνευτική ικανότητα.