

1ο ΕΠΑΝΑΛΗΠΤΙΚΟ ΘΕΜΑ:

Δίνεται ο παρακάτω πίνακας με τα ετήσια στοιχεία μιας επιχείρησης:

έτη	Ψ: πωλήσεις (σε χιλ. ευρώ)	X1: δαπάνες διαφήμισης (σε χιλ. ευρώ)	X2: 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

- Ερώτημα 1: Να βρεθεί η τιμή του συντελεστή γραμμικής συσχέτισης ($\rho=;$) που συνδέει τις πωλήσεις (Ψ) με τις δαπάνες για διαφήμιση (X1):

Λύση:

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

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

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

(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

Ο συντελεστής συσχέτισης θα είναι: $\rho = 1.352,143 / (1.372,612) = +0,9850$

- **Ερώτημα 2:** Να υπολογίσετε τους συντελεστές a και b του απλού γραμμικού μοντέλου της μορφής: $\text{πωλήσεις}(t) = a + b * \text{δαπάνες_διαφήμισης}(t) + \varepsilon(t)$

Λύση:

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

Y_i	X_i	$(X_i * Y_i)$	$(X_i)^2$
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)

$$b = [7 * (8.140) - (559*85)] / [(7 * 1.281) - (85*85)] \\ = (56.980 - 47.515) / (8.967 - 7.225) = 9.465 / 1.742 = 5,4334$$

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

- **Ερώτημα 3:** 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-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 $\alpha=5\%$).

- **Ερώτημα 4:** Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για τον συντελεστή β του απλού γραμμικού μοντέλου της μορφής:
 $\text{πωλήσεις}(t) = \alpha + \beta * \text{δαπάνες_διαφήμισης}(t) + \varepsilon(t)$

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

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

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

- 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

Ερώτημα 5: 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%)**

Ερώτημα 6: 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).

Ερώτημα 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 $k=2 \rightarrow$ **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 $\rightarrow$ **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.