

Εφαρμογή Πολλαπλής Γραμμικής Παλινδρόμησης

Απόκριση: $\ln(\text{FEV})$

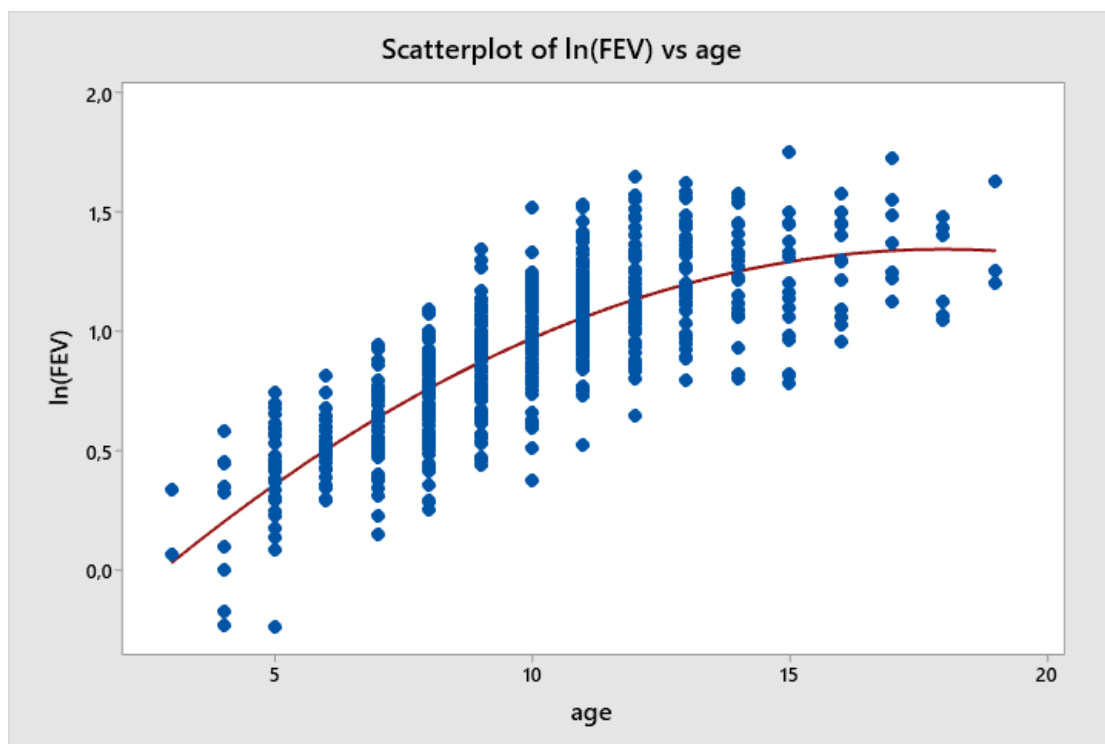
Συμμεταβλητές: ηλικία, ύψος, κάπνισμα (0: μη καπνιστής, 1: καπνιστής)

A) Συμμεταβλητές: ηλικία, κάπνισμα (0: μη καπνιστής, 1: καπνιστής)

Συμμεταβλητή: ηλικία μόνο

1^ο Μοντέλο: Πολυωνομικό Μοντέλο 2^{ου} βαθμού:

$$E(\ln \text{FEV}) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ηλικία}^2$$



Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	-0,5457	0,0713	(-0,6857; -0,4056)	-7,65	0,000	
age	0,2103	0,0139	(0,1831; 0,2375)	15,17	0,000	27,38
age^2	-0,005845	0,000645	(-0,007113; -0,004578)	-9,06	0,000	27,38

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Seq MS	F-Value	P-Value
Regression	2	46,4897	64,10%	46,4897	23,2449	581,21	0,000
age	1	43,2101	59,58%	9,2047	43,2101	1080,41	0,000
age^2	1	3,2797	4,52%	3,2797	3,2797	82,00	0,000
Error	651	26,0362	35,90%	26,0362	0,0400		
Lack-of-Fit	14	0,6877	0,95%	0,6877	0,0491	1,23	0,245
Pure Error	637	25,3485	34,95%	25,3485	0,0398		
Total	653	72,5259	100,00%				

Tests use the sequential sums of squares

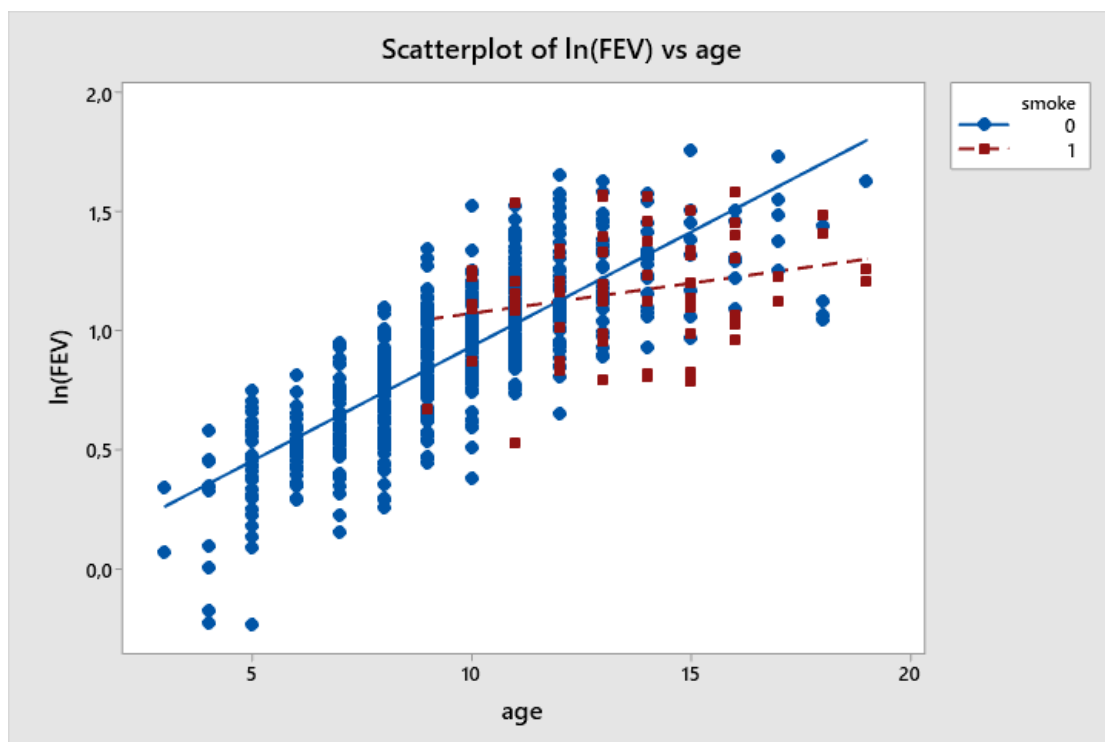
Ελέγγω $H_0: \beta_2 = 0$ έναντι δίπλευρης εναλλακτικής σε ε.σ. $\alpha=0,05$.

$\hat{\beta}_2 = -0,0058$, $s.e(\hat{\beta}_2) = 0,000645$, $T = -9,06$, $p - value = 0,000$

Απορρίπτω τη μηδενική.

2^ο Μοντέλο: Συμμεταβλητές: ηλικία, κάπνισμα

$$E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{κάπνισμα} + \beta_3 \times (\text{κάπνισμα} \times \text{ηλικία})$$



Regression Equation

$$\ln(\text{FEV}) = -0,0266 + 0,09596 \text{ age} + 0,846 \text{ smoke} - 0,0707 \text{ age*smoke}$$

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	-0,0266	0,0306	(-0,0867; 0,0335)	-0,87	0,385	
age	0,09596	0,00308	(0,08991; 0,10202)	31,12	0,000	1,29
smoke	0,846	0,153	(0,544; 1,147)	5,51	0,000	32,77
age*smoke	-0,0707	0,0114	(-0,0930; -0,0484)	-6,22	0,000	34,07

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC
0,204930	62,36%	62,19%	27,6910	61,82%	-211,24	-188,92

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Seq MS	F-Value	P-Value
Regression	3	45,2283	62,36%	45,228	15,0761	358,99	0,000
age	1	43,2101	59,58%	40,671	43,2101	1028,90	0,000
smoke	1	0,3960	0,55%	1,277	0,3960	9,43	0,002
age*smoke	1	1,6222	2,24%	1,622	1,6222	38,63	0,000
Error	650	27,2976	37,64%	27,298	0,0420		
Lack-of-Fit	24	2,8293	3,90%	2,829	0,1179	3,02	0,000
Pure Error	626	24,4683	33,74%	24,468	0,0391		
Total	653	72,5259	100,00%				

Tests use the sequential sums of squares

Προσαρμοσμένες ευθείες:

Μη-καπνιστές:

$$\ln \widehat{FEV} = -0,0266 + 0,09596 \times \text{ηλικία}$$

Καπνιστές :

$$\ln \widehat{FEV} = (-0,0266 + 0,846) + (0,09596 - 0,0707) \times \text{ηλικία}$$

$H_0: \beta_3 = 0$, δηλαδή η επίδραση της ηλικίας είναι ίδια για καπνιστές και μη.

P-value=0,000, άρα απορρίπτω την μηδενική και συμπεραίνω....

3^ο Μοντέλο: Ένα 'πλήρες' πολυωνυμικό μοντέλο 2^ο βαθμού με ηλικία και κάπνισμα

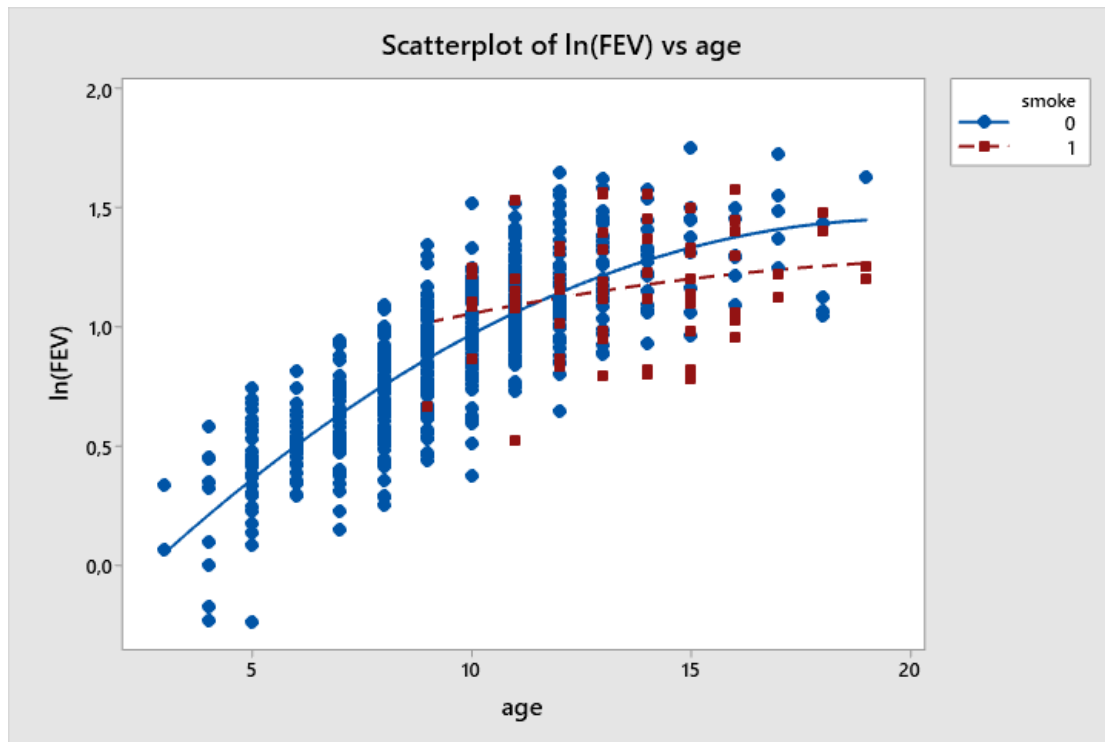
$$E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ηλικία}^2 + \beta_3 \times \text{κάπνισμα} \\ + \beta_4 \times (\text{κάπνισμα} \times \text{ηλικία}) + \beta_5 \times (\text{κάπνισμα} \times \text{ηλικία}^2)$$

Καμπύλη για μη καπνιστές:

$$E(\ln FEV) = \beta_0 + \beta_1 \times \etaλικία + \beta_2 \times \etaλικία^2$$

Καμπύλη για καπνιστές:

$$E(\ln FEV) = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) \times \etaλικία + (\beta_2 + \beta_5) \times \etaλικία^2$$



Regression Equation

$$\ln(\text{FEV}) = -0,4885 + 0,1948 \text{ age} - 0,004882 \text{ age}^2 + 1,044 \text{ smoke} - 0,130 \text{ smoke*age} + 0,00347 \text{ smoke*age}^2$$

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	-0,4885	0,0758	(-0,6373; -0,3397)	-6,45	0,000	
age	0,1948	0,0152	(0,1649; 0,2247)	12,80	0,000	33,45
age^2	-0,004882	0,000737	(-0,006329; -0,003434)	-6,62	0,000	36,19
smoke	1,044	0,729	(-0,388; 2,476)	1,43	0,153	789,37
smoke*age	-0,130	0,107	(-0,341; 0,080)	-1,22	0,223	3211,19
smoke*age^2	0,00347	0,00387	(-0,00412; 0,01107)	0,90	0,370	890,66

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC
0,198613	64,76%	64,48%	26,1287	63,97%	-250,13	-218,92

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Seq MS	F-Value	P-Value
Regression	5	46,9642	64,76%	46,9642	9,3928	238,11	0,000
age	1	43,2101	59,58%	6,4632	43,2101	1095,39	0,000
age^2	1	3,2797	4,52%	1,7305	3,2797	83,14	0,000
smoke	1	0,1233	0,17%	0,0808	0,1233	3,13	0,078
smoke*age	1	0,3194	0,44%	0,0586	0,3194	8,10	0,005
smoke*age^2	1	0,0318	0,04%	0,0318	0,0318	0,81	0,370
Error	648	25,5617	35,24%	25,5617	0,0394		
Lack-of-Fit	22	1,0934	1,51%	1,0934	0,0497	1,27	0,182
Pure Error	626	24,4683	33,74%	24,4683	0,0391		
Total	653	72,5259	100,00%				

Tests use the sequential sums of squares

Έστω πως θέλω να εξετάσω αν χρειάζεται η πληροφορία αν κάποιος καπνίζει στο να προβλέψω την απόκριση. Δηλαδή, έχουν οι καπνιστές διαφορετική καμπύλη από τους μη καπνιστές?

$$H_0: \beta_3 = \beta_4 = \beta_5 = 0 \text{ vs } H_1: \text{όχι } H_0, \text{ ή αλλιώς}$$

$$H_0: \beta_3 = 0 \text{ και } \beta_4 = 0 \text{ και } \beta_5 = 0 \text{ vs } H_1: \text{όχι } H_0$$

Δηλαδή συγκρίνω το 3^ο Μοντέλο (Full Model) με το 1^ο Μοντέλο (Reduced Model).

Η στατιστική συνάρτηση ελέγχου είναι

$F = \frac{(SSR_{FULL} - SSR_{RED}) / df^*}{MSE_{FULL}}$, όπου df^* είναι απλά ο αριθμός των υποθέσεων που ελέγχεται. Επίσης SSR είναι άθροισμα τετραγώνων λόγω παλινδρόμησης. Κάτω από την μηδενική υπόθεση η σ.σ.ε. έχει την $F_{df^*, df(ERROR) FULL}$

Απορρίπτω αν F είναι μεγάλο, δηλαδή μεγαλύτερο από το άνω α ποσοστιαίο σημείο της $F_{df^*, df(ERROR) FULL}$ κατανομής. Εδώ έχουμε

$F = \frac{(46,9642 - 46,4897) / 3}{25,5617 / 648} = \frac{0,4745}{0,0394} = 12,04 > F_{3, 648; 0,05} \cong 2,62$, άρα απορρίπτω την μηδενική υπόθεση, που σημαίνει πως σε ε.σ. 5% μπορώ να ισχυριστώ πως οι καμπύλες είναι διαφορετικές.

B) Μοντέλα με ηλικία, ύψος και κάπνισμα

1^ο Μοντέλο (ύψος, ηλικία)

$$\text{Έχουμε } E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος}$$

2^ο Μοντέλο (ύψος, ηλικία, κάπνισμα)

$$\text{Έχουμε } E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος} + \beta_3 \times \text{κάπνισμα}$$

Δηλαδή:

$$\text{Για μη-καπνιστές: } E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος}$$

$$\text{Για καπνιστές: } E(\ln FEV) = (\beta_0 + \beta_3) + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος}$$

3^ο Μοντέλο (ύψος, ηλικία, κάπνισμα, αλληλεπίδραση καπνίσματος με ηλικία)

$$\text{Έχουμε } E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος} + \beta_3 \times \text{κάπνισμα} \\ + \beta_4 \times (\text{κάπνισμα} \times \text{ηλικία})$$

$$\text{Για μη-καπνιστές: } E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος}$$

$$\text{Για καπνιστές: } E(\ln FEV) = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) \times \text{ηλικία} + \beta_2 \times \text{ύψος}$$

Η επίδραση της ηλικίας στο $\ln FEV$ για τους **μη καπνιστές** είναι β_1 (Για σταθερό ύψος, μεταβολή της ηλικίας κατά ένα έτος συνεπάγεται β_1 μονάδες διαφορά κατά μέσο όρο σε $\ln FEV$).

Η επίδραση της ηλικίας στο $\ln FEV$ για τους **καπνιστές** είναι $\beta_1 + \beta_4$ (Για σταθερό ύψος, μεταβολή της ηλικίας κατά ένα έτος συνεπάγεται $\beta_1 + \beta_4$ μονάδες διαφορά κατά μέσο όρο σε $\ln FEV$).

4^ο Μοντέλο (ύψος, ηλικία, κάπνισμα, αλληλεπίδραση καπνίσματος με ηλικία, αλληλεπίδραση καπνίσματος με ύψος)

$$\text{Έχουμε } E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος} + \beta_3 \times \text{κάπνισμα} \\ + \beta_4 \times (\text{κάπνισμα} \times \text{ηλικία}) + \beta_5 \times (\text{κάπνισμα} \times \text{ύψος})$$

$$\text{Για μη-καπνιστές: } E(\ln FEV) = \beta_0 + \beta_1 \times \text{ηλικία} + \beta_2 \times \text{ύψος}$$

$$\text{Για καπνιστές: } E(\ln FEV) = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) \times \text{ηλικία} + (\beta_2 + \beta_5) \times \text{ύψος}$$

Η επίδραση της ηλικίας στο $\ln FEV$ για τους **μη καπνιστές** είναι β_1 (Για σταθερό ύψος, μεταβολή της ηλικίας κατά ένα έτος συνεπάγεται β_1 μονάδες διαφορά κατά μέσο όρο σε $\ln FEV$).

Η επίδραση της ηλικίας στο $\ln FEV$ για τους **καπνιστές** είναι $\beta_1 + \beta_4$ (Για σταθερό ύψος, μεταβολή της ηλικίας κατά ένα έτος συνεπάγεται $\beta_1 + \beta_4$ μονάδες διαφορά κατά μέσο όρο σε $\ln FEV$).

Η επίδραση του ύψους στο $\ln FEV$ για τους **μη καπνιστές** είναι β_2 (Για σταθερή ηλικία, μεταβολή του ύψους κατά μία ίντσα συνεπάγεται β_2 μονάδες διαφορά κατά μέσο όρο σε $\ln FEV$).

Η επίδραση του ύψους στο $\ln FEV$ για τους **καπνιστές** είναι $\beta_2 + \beta_5$ (Για σταθερή ηλικία, μεταβολή του ύψους κατά μία ίντσα συνεπάγεται $\beta_2 + \beta_5$ μονάδες διαφορά κατά μέσο όρο σε $\ln FEV$).