ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

7. Παλινδρόμηση

Γενικά

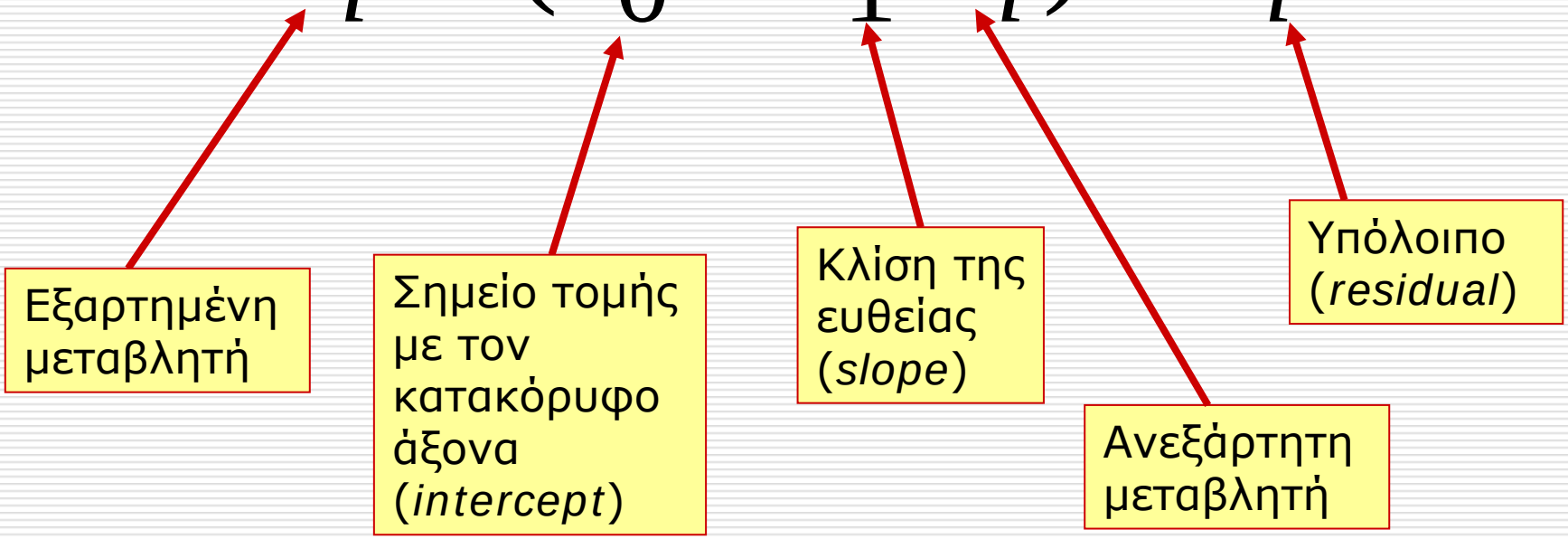
- Επέκταση της έννοιας της συσχέτισης:
  - Πώς μπορούμε να προβλέπουμε τη μια μεταβλητή από την άλλη;
- **Απλή παλινδρόμηση (*simple regression*):** Κατασκευή μοντέλου πρόβλεψης της μιας μεταβλητής από την άλλη
- **Πολλαπλή παλινδρόμηση (*multiple regression*):** Κατασκευή μοντέλου πρόβλεψης της μιας μεταβλητής από πολλές άλλες

Εισαγωγή στην παλινδρόμηση

- Κατασκευή μοντέλου πρόβλεψης **εξαρτημένης μεταβλητής** (*dependent variable*) από ανεξάρτητες μεταβλητές (*independent variables*)
- Απλούστερο μοντέλο: Προσαρμογή ευθείας (γραμμικό μοντέλο) στα δεδομένα
- Υπολογισμός της ευθείας με μέθοδο **ελαχίστων τετραγώνων** (*least squares*)

Το απλό γραμμικό μοντέλο

$$Y_i = (b_0 + b_1 X_i) + \varepsilon_i$$



- b_0, b_1: **Συντελεστές παλινδρόμησης** (*regression coefficients*)

Η μέθοδος των ελαχίστων τετραγώνων

- Επιλογή της ευθείας (εύρεση των συντελεστών b0, b1) ώστε να ελαχιστοποιείται η ποσότητα

$$\sum_i \varepsilon_i^2 = \sum_i \left(Y_i - b_0 - b_1 X_i\right)^2$$

- Υπολογίζεται με μαθηματικό τρόπο (ακρότατα συνάρτησης)

Αξιολόγηση της προσαρμογής (goodness of fit)

- Πόσο καλά προσαρμόζεται η ευθεία στα δεδομένα;

$$SS_T = \sum_i (Y_i - \bar{Y})^2$$

Συνολικό άθροισμα τετραγώνων (Total Sum of Squares)

$$SS_R = \sum_i (Y_i - b_0 - b_1 X_i)^2$$

Άθροισμα τετραγώνων υπολοίπων (Residual Sum of Squares)

$$SS_M = SS_T - SS_R$$

Άθροισμα τετραγώνων μοντέλου (Model Sum of Squares)

Ερμηνεία των αθροισμάτων τετραγώνων

- SS_T: Η απόκλιση των δεδομένων από το «χειρότερο μοντέλο» (μέση τιμή)
- SS_R: Η απόκλιση των δεδομένων από το «καλύτερο μοντέλο» (ευθεία)
- SS_M: Η διαφορά ανάμεσα στο «χειρότερο» και στο «καλύτερο μοντέλο»
  - Μεγάλο SS_M: σημαντική η συνεισφορά του μοντέλου στην πρόβλεψη της Y
  - Μικρό SS_M: το μοντέλο ελάχιστα βελτιώνει την «χειρότερη πρόβλεψη» της μέσης τιμής

Μέτρο αξιολόγησης του μοντέλου: R^2

- Η ποιότητα της προσαρμογής του μοντέλου μπορεί να μετρηθεί ως ποσοστό «βελτίωσης της πρόβλεψης» που οφείλεται στο μοντέλο

$$R^2 = \frac{SS_M}{SS_T} = \frac{SS_T - SS_R}{SS_T} = 1 - \frac{SS_R}{SS_T}$$

- Ερμηνεία: το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο
- Συμπίπτει με το τετράγωνο του συντελεστή Pearson

Μέτρο αξιολόγησης του μοντέλου: F-test

$$MS_M = \frac{SS_M}{\text{degrees\_of\_freedom}} = \frac{SS_M}{\#\text{variables}}$$

$$MS_R = \frac{SS_R}{\text{degrees\_of\_freedom}} = \frac{SS_R}{n - \#regr.coefficients}$$

Μέσα αθροίσματα τετραγώνων (Mean Sum of Squares)

$$F = \frac{MS_M}{MS_R}$$

Ερμηνεία: Για ένα καλό μοντέλο το MS_M θα είναι μεγάλο και το MS_R μικρό άρα «συνολικά» το F θα είναι μεγάλο (sig. < 0.05)

Σημαντικότητα των συντελεστών

- Ερμηνεία του b_1: η αλλαγή που επέρχεται στην εξαρτημένη μεταβλητή αν η ανεξάρτητη αλλάξει κατά μια μονάδα
- Σε κακό μοντέλο: $b_1 \approx 0$
- Για να ελέγξουμε αν η τιμή του b_1 είναι σημαντικά διαφορετική του 0 χρησιμοποιούμε t-test (Sig.<0.05)

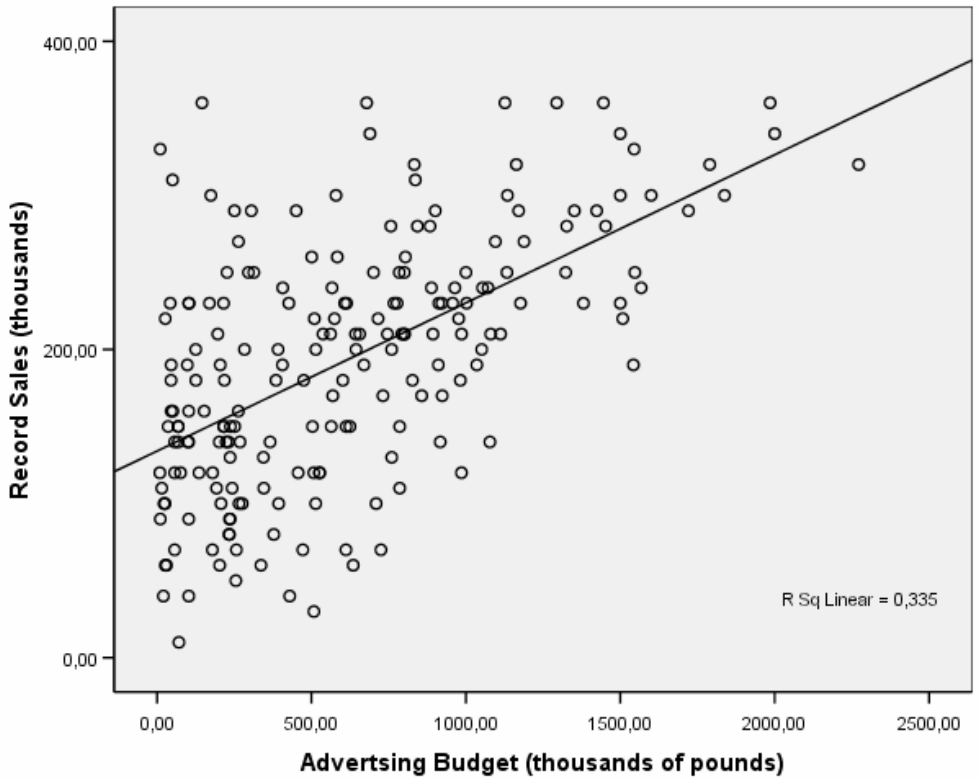
Απλή παλινδρόμηση με το SPSS

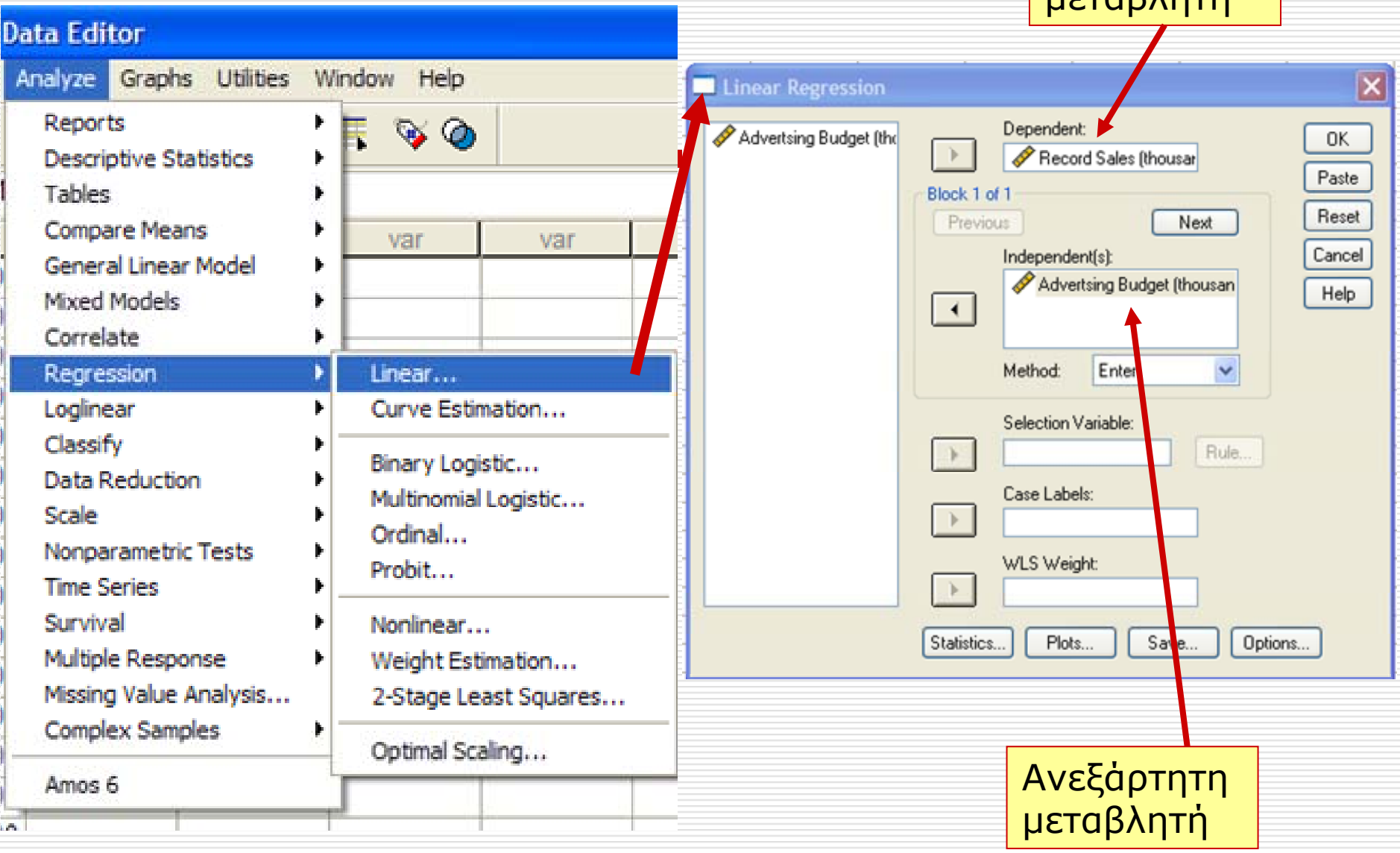
file: Record1.sav

- Ανεξάρτητη μεταβλητή: adverts (ποσό διαφήμισης δίσκου)
- Εξαρτημένη μεταβλητή: sales (αριθμός πωλήσεων δίσκου)

Statistics

		Advertsing Budget (thousands of pounds)	Record Sales (thousands)
N	Valid	200	200
	Missing	0	0
Mean		614,4123	193,2000
Median		531,9160	200,0000
Std. Deviation		485,65521	80,69896
Minimum		9,10	10,00
Maximum		2271,86	360,00




Εξαρτημένη μεταβλητή
Data Editor
Analyze
Graphs
Utilities
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Amos 6
Linear...
Curve Estimation...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling...
var
var
Linear Regression
Advertsing Budget (tho
Dependent:
Record Sales (thousar
Block 1 of 1
Previous
Next
Independent(s):
Advertsing Budget (thousan
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
OK
Paste
Reset
Cancel
Help
Statistics...
Plots...
Save...
Options...
Ανεξάρτητη μεταβλητή

Αποτελέσματα παλινδρόμησης – Συνολική προσαρμογή του μοντέλου

Model Summary

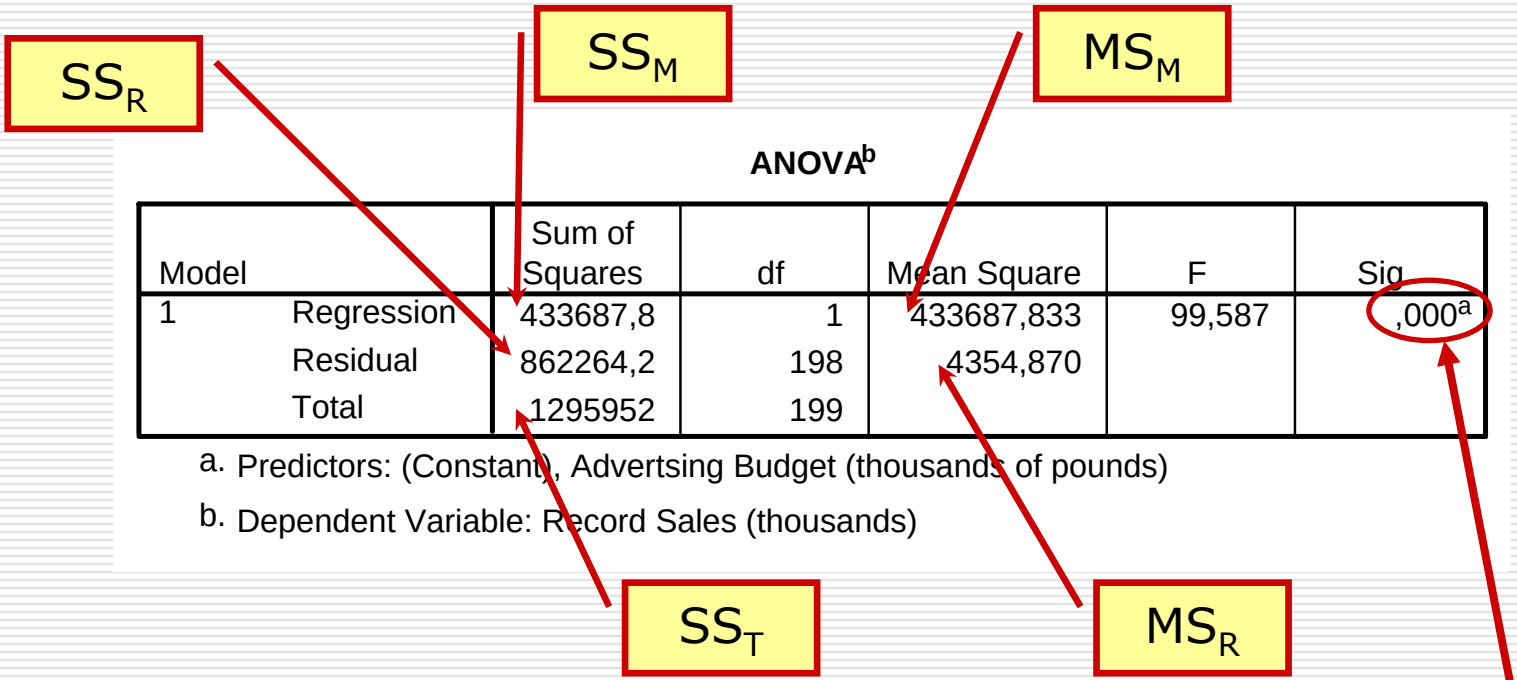
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,578[a]	,335	,331	65,99144

a. Predictors: (Constant), Advertsing Budget (thousands of pounds)

Συντελεστής συσχέτισης Pearson

Το μοντέλο εξηγεί το 33.5% της μεταβλητότητας των πωλήσεων

Αποτελέσματα παλινδρόμησης – Συνολική προσαρμογή του μοντέλου



ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig
1	Regression	433687,8	1	433687,833	99,587	,000[a]
	Residual	862264,2	198	4354,870		
	Total	1295952	199			

a. Predictors: (Constant), Advertsing Budget (thousands of pounds)

b. Dependent Variable: Record Sales (thousands)

- Συμπέρασμα: Από το F-test (sig<0,001) συμπεραίνουμε ότι το μοντέλο συνεισφέρει σημαντικά στην πρόβλεψη του αριθμού των πωλήσεων

Αποτελέσματα παλινδρόμησης – παράμετροι του μοντέλου

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	134,140	7,537		17,799	,000
	Advertsing Budget (thousands of pounds)	,096	,010	,578	9,979	,000

a. Dependent Variable: Record Sales (thousands)

Record Sales=134,140+0,096*Advertising Budget

sig<0,001 άρα οι δύο παράμετροι είναι σημαντικά διαφορετικές από το 0 και επομένως η συνεισφορά του μοντέλου στην πρόβλεψη των πωλήσεων είναι σημαντική

Ερμηνεία των παραμέτρων του μοντέλου

- b0=134,140: Σε περίπτωση που δεν ξοδευτούν καθόλου χρήματα για διαφήμιση (X=0) το μοντέλο προβλέπει ότι θα πουληθούν 134140 δίσκοι
- b1=0,096: Αν το ποσό της διαφήμισης αυξηθεί κατά 1 μονάδα (£1000) το μοντέλο προβλέπει 96 (=0,096*1000) επιπλέον πωλήσεις δίσκων (**!!**)
- Παράδειγμα πρόβλεψης: Πόσοι δίσκοι θα πωληθούν αν ξοδέψουμε £100000; Με αντικατάσταση (X=100) παίρνουμε Y=143,75 (δηλ. περίπου 144000 δίσκους)

Πολλαπλή παλινδρόμηση (*multiple regression*)

- Επέκταση του γραμμικού μοντέλου με περισσότερες ανεξάρτητες μεταβλητές

$$Y_i = (b_0 + b_1 X_i + \cdots + b_k X_k) + \varepsilon_i$$

- Η εξίσωση της ευθείας επεκτείνεται στο επίπεδο (2 ανεξ. μεταβλητές) και στο υπερεπίπεδο (>2 ανεξ. μετ/τές)

Αξιολόγηση του μοντέλου

- SS_T, SS_R, SS_M υπολογίζονται με πιο πολύπλοκο τρόπο αλλά η ερμηνεία τους είναι η ίδια
- Υπολογίζεται συντελεστής πολλαπλής συσχέτισης (multiple R) που δείχνει πόσο ισχυρή είναι η συσχέτιση εξαρτημένης με όλες τις ανεξάρτητες
- Το R^2 ερμηνεύεται με ακριβώς τον ίδιο τρόπο ως ποσοστό μεταβλητότητας που εξηγείται από το μοντέλο

Μέθοδοι παλινδρόμησης

- Βασικό πρόβλημα: Πώς επιλέγουμε τις ανεξάρτητες μεταβλητές που θα χρησιμοποιηθούν για το μοντέλο;
- Οι ανεξάρτητες μεταβλητές συνήθως είναι συσχετισμένες μεταξύ τους
- Υπάρχουν μεθοδολογίες επιλογής των καταλληλότερων μεταβλητών

Επιλογή μεταβλητών

- **Αναγκαστική εισαγωγή** (*Forced Entry – Enter*): όλες οι μεταβλητές ταυτόχρονα
- **Εισαγωγή και εξαγωγή με βήματα** (*Stepwise*): η σειρά καθορίζεται από μαθηματικά κριτήρια
- **Προς τα εμπρός εισαγωγή** (*Forward*)
- **Προς τα πίσω εξαγωγή** (*Backward*)
- Γενικές οδηγίες:
  - Η επιλογή με βήματα δίνει διαφορετικά μοντέλα, δεν αφήνει τον ερευνητή να επιλέξει.
  - Προτιμότερο να στηριζόμαστε σε θεωρητικά βιβλιογραφικά αποτελέσματα.

Ακρίβεια του μοντέλου

- Βασικά ερωτήματα:
  - Το μοντέλο προσαρμόζεται καλά στα δεδομένα ή επηρεάζεται από λίγες περιπτώσεις;
  - Μπορεί το μοντέλο να γενικευτεί και σε άλλα δείγματα;

Διαγνωστικά προσαρμογής του μοντέλου (*diagnostics*)

- **Παράτυπα σημεία** (*outliers*): Δεδομένα (cases) που διαφέρουν σημαντικά από τα υπόλοιπα
- Μπορούν να επηρεάσουν σημαντικά τις τιμές των συντελεστών της παλινδρόμησης
- Μπορούν να εντοπιστούν από τα μεγάλα **υπόλοιπα** (*residuals*) που δίνουν

Υπόλοιπα

- Γενικά: Τα υπόλοιπα υπολογίζονται ως διαφορές ανάμεσα στις παρατηρήσεις και τις εκτιμήσεις της παλινδρόμησης
  - μικρά υπόλοιπα $\Rightarrow$ καλή προσαρμογή
  - μεγάλα υπόλοιπα $\Rightarrow$ κακή προσαρμογή
  - σημεία με ιδιαίτερα μεγάλα υπόλοιπα $\Rightarrow$ παράτυπα σημεία

Μετασχηματισμοί υπολοίπων

- Για μελέτη – σύγκριση υπολοίπων τα τυποποιούμε (*standardized residuals*) διαιρώντας με την τυπική τους απόκλιση
  - Τυποποιημένα υπόλοιπα με απόλυτη τιμή > 3 προβληματίζουν
  - Αν πάνω από 1% των τυπ. υπολοίπων είναι > 2.5 έχουμε ένδειξη κακής προσαρμογής
  - Αν πάνω από 5% των τυπ. υπολοίπων είναι >2 έχουμε ένδειξη κακής προσαρμογής
- *Studentized residuals*: Τα υπόλοιπα διαιρεμένα με εκτιμητή της τυπ. απόκλισης που μεταβάλλεται από σημείο σε σημείο. Θεωρούνται ακριβέστερα

Δεδομένα με σημαντική επιρροή (*Influential cases*)

- Άλλος τρόπος ελέγχου παράτυπων σημείων:
  - Υπάρχουν σημεία που έχουν αδικαιολόγητα μεγάλη επίδραση στο μοντέλο;
- *Adjusted Predicted Value*: υπολογίζεται για κάθε case αφαιρώντας την από το δείγμα και εκτιμώντας την με το μοντέλο που προκύπτει
- *DFFit*: Διαφορά ανάμεσα στην Adj. Pred. value και στην αρχική Pred. value → *Standardized DFFit*
- *Deleted Residual*: Διαφορά ανάμεσα στην Adj. Pred. value και στην παρατηρούμενη τιμή → *Studentized deleted residual*

Δεδομένα με σημαντική επιρροή (*Influential cases*)

- *Cook's distance*: Μέτρο συνολικής επίδρασης ενός σημείου στο μοντέλο. Δεδομένα με τιμή >1 προβληματίζουν
- Άλλα μέτρα:
  - leverage values
  - Mahalanobis distances
  - DFBeta & Standardized DFBeta
  - Covariance Ratio (CVR)

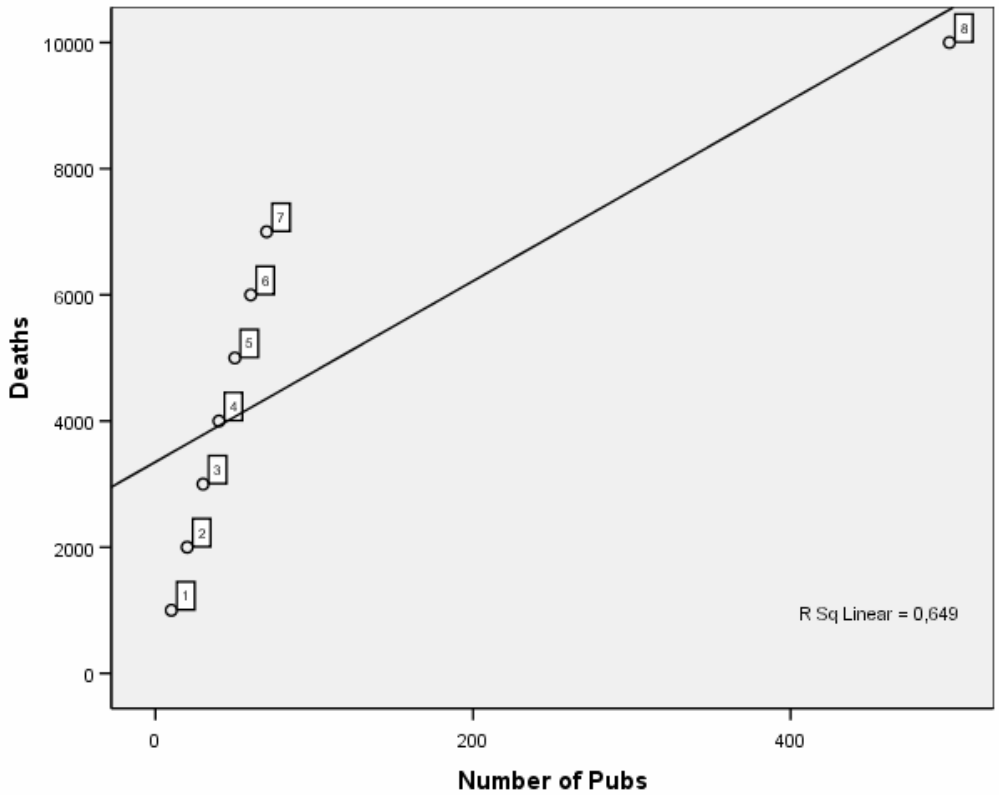
Παράδειγμα
(File: pubs.sav)

- Ανεξάρτητη μεταβλητή: Number of pubs
- Εξαρτημένη μεταβλητή: Number of deaths (σε μια χρονική περίοδο)

Case Summaries[a]

		Number of Pubs	Deaths
1		10	1000
2		20	2000
3		30	3000
4		40	4000
5		50	5000
6		60	6000
7		70	7000
8		500	10000
Total	N	8	8

a. Limited to first 100 cases.



Διαφορά residual – influence statistics

Case Summaries[a]

	Standardized Residual	Cook's Distance	Centered Leverage Value	Standardized DFFIT	DFBETA Intercept	DFBETA pubs	Standardized DFBETA Intercept	Standardized DFBETA pubs
1	-1,33839	,21328	,04074	-,74402	-509,6518	1,39249	-,74317	,36886
2	-,87895	,08530	,03196	-,40964	-321,1277	,80153	-,40766	,18484
3	-,41950	,01814	,02424	-,17697	-147,1066	,33016	-,17494	,07132
4	,03995	,00015	,01759	,01606	13,45081	-,02658	,01572	-,00564
5	,49940	,02294	,01200	,20042	161,44976	-,27267	,19337	-,05933
6	,95885	,08092	,00748	,40473	297,67748	-,41116	,38333	-,09618
7	1,41830	,17107	,00402	,68084	422,81664	-,44422	,62996	-,12023
8	-,27966	227,14286	,86196	-460379232,7	3351,955	-85,66108	92676016,02	-430238878,2
Total N	8	8	8	8	8	8	8	8

a. Limited to first 100 cases.

Πολύ μικρό υπόλοιπο

Πολύ μεγάλη επίδραση στο μοντέλο

Γενίκευση του μοντέλου – βασικές υποθέσεις (1/2)

- Τύπος μεταβλητών: Οι ανεξάρτητες είναι ποσοτικές ή δίτιμες και η εξαρτημένη συνεχής
- Καμιά μεταβλητή δεν έχει διασπορά 0
- Οι ανεξάρτητες δεν πρέπει να έχουν μεγάλες συσχετίσεις μεταξύ τους (*multicollinearity*)
- Τα υπόλοιπα πρέπει να έχουν σταθερή διασπορά (*homoscedasticity*). Προβληματική η διαφορετική διασπορά (*heteroscedasticity*).

Γενίκευση του μοντέλου – βασικές υποθέσεις (2/2)

- Ανεξάρτητα σφάλματα (*independent errors*). Υποθέτουμε ότι δεν υπάρχει αυτοσυσχέτιση (*autocorrelation*)
- Σφάλματα κανονικά κατανεμημένα (υποθέτουμε ότι τα υπόλοιπα ακολουθούν κανονική κατανομή με μέση τιμή 0)
- Ανεξαρτησία των τιμών της εξαρτημένης μεταβλητής
- Η πραγματική σχέση είναι γραμμική

Ακρίβεια του μοντέλου για άλλα δείγματα – *Cross Validation*

- *Adjusted R^2*: «Διόρθωση» του R^2
  - ερμηνεύεται ως το ποσοστό της μεταβλητότητας της Y που θα ερμηνευόταν από το μοντέλο του πληθυσμού
- Τυχαία διαμέριση των δεδομένων σε *training set* και *test set*.
  - Το μοντέλο δημιουργείται από το training και προβλέπει τα σημεία του test – ακολουθεί αξιολόγηση

Έλεγχοι για παραβίαση των υποθέσεων

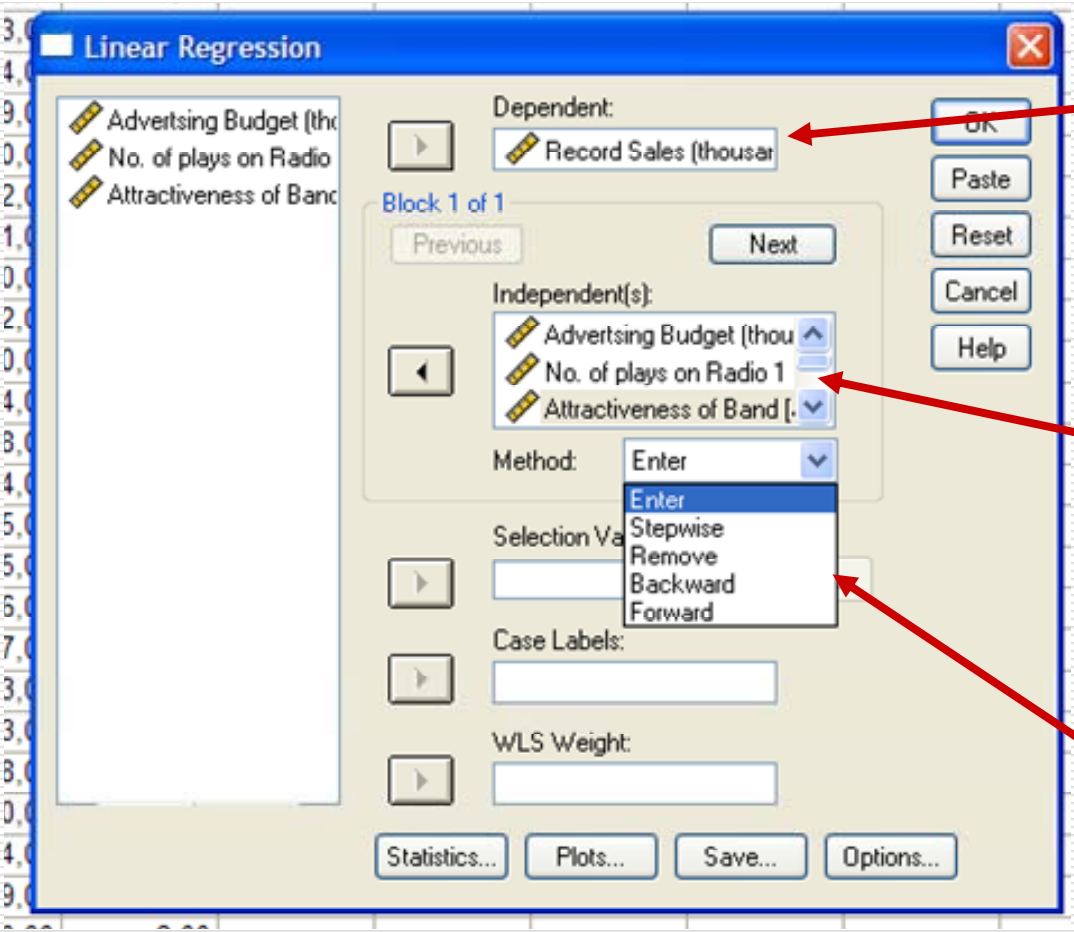
- Multicollinearity:
  - Εξέταση πίνακα συσχετίσεων ανεξάρτητων μεταβλητών
  - Variance inflation factor (VIF)
  - Tolerance
- Heteroscedasticity - Normality:
  - Γραφικές παραστάσεις υπολοίπων
- Autocorrelation:
  - Durbin-Watson test

Πολλαπλή παλινδρόμηση με το SPSS

file: Record2.sav

- Ανεξάρτητές μεταβλητές:
  - adverts (ποσό διαφήμισης δίσκου)
  - airplay (αριθμός ραδιοφωνικών μεταδόσεων του δίσκου από συγκεκριμένο σταθμό)
  - attract (ελκυστικότητα του καλλιτέχνη / συγκροτήματος 0-10 από προηγούμενη έρευνα)
- Εξαρτημένη μεταβλητή: sales (αριθμός πωλήσεων δίσκου)

Analyze->Regression->Linear



Εξαρτημένη μεταβλητή

Ανεξάρτητες μεταβλητές

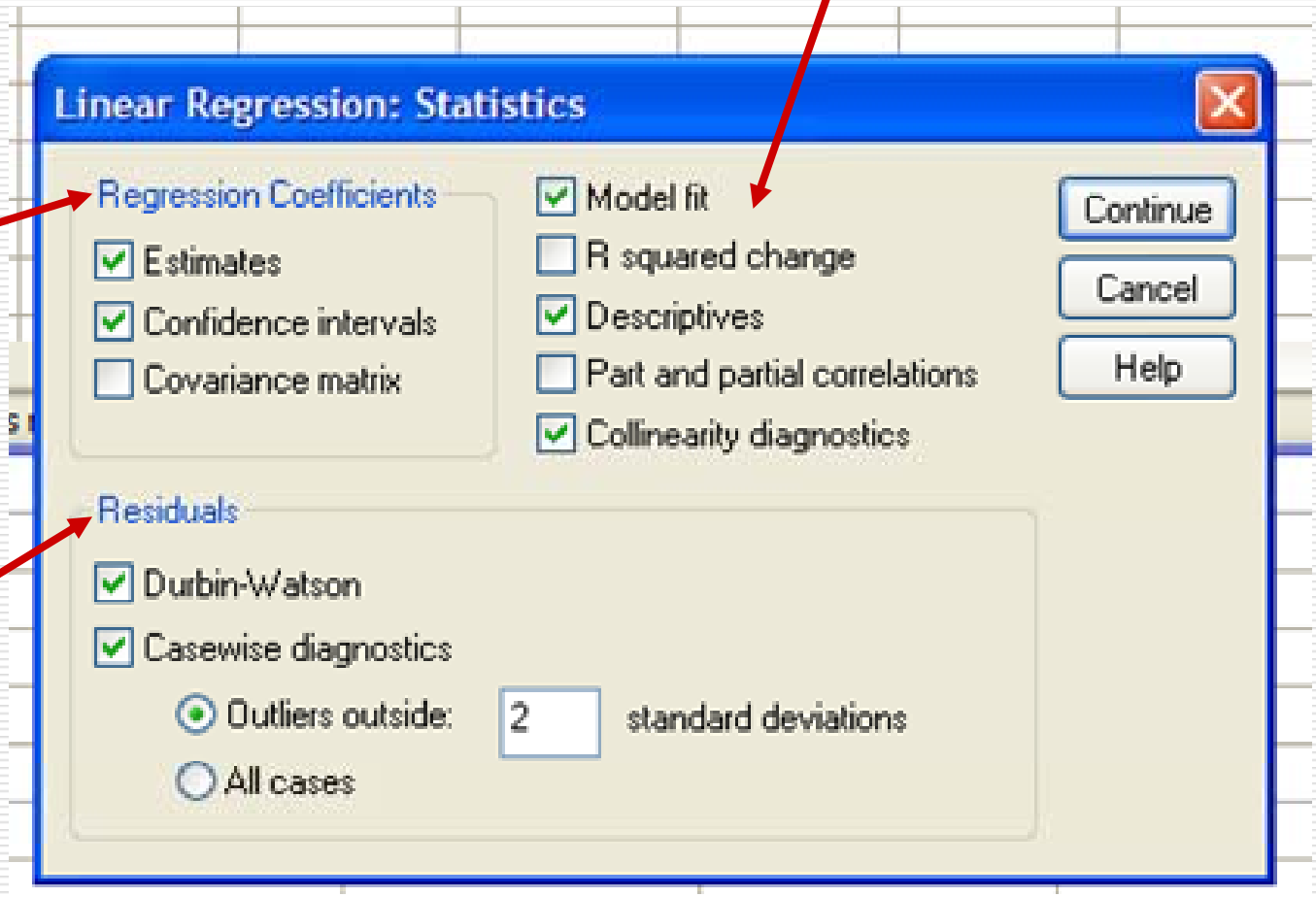
Επιλογή μεθόδου εισαγωγής μεταβλητών

Statistics

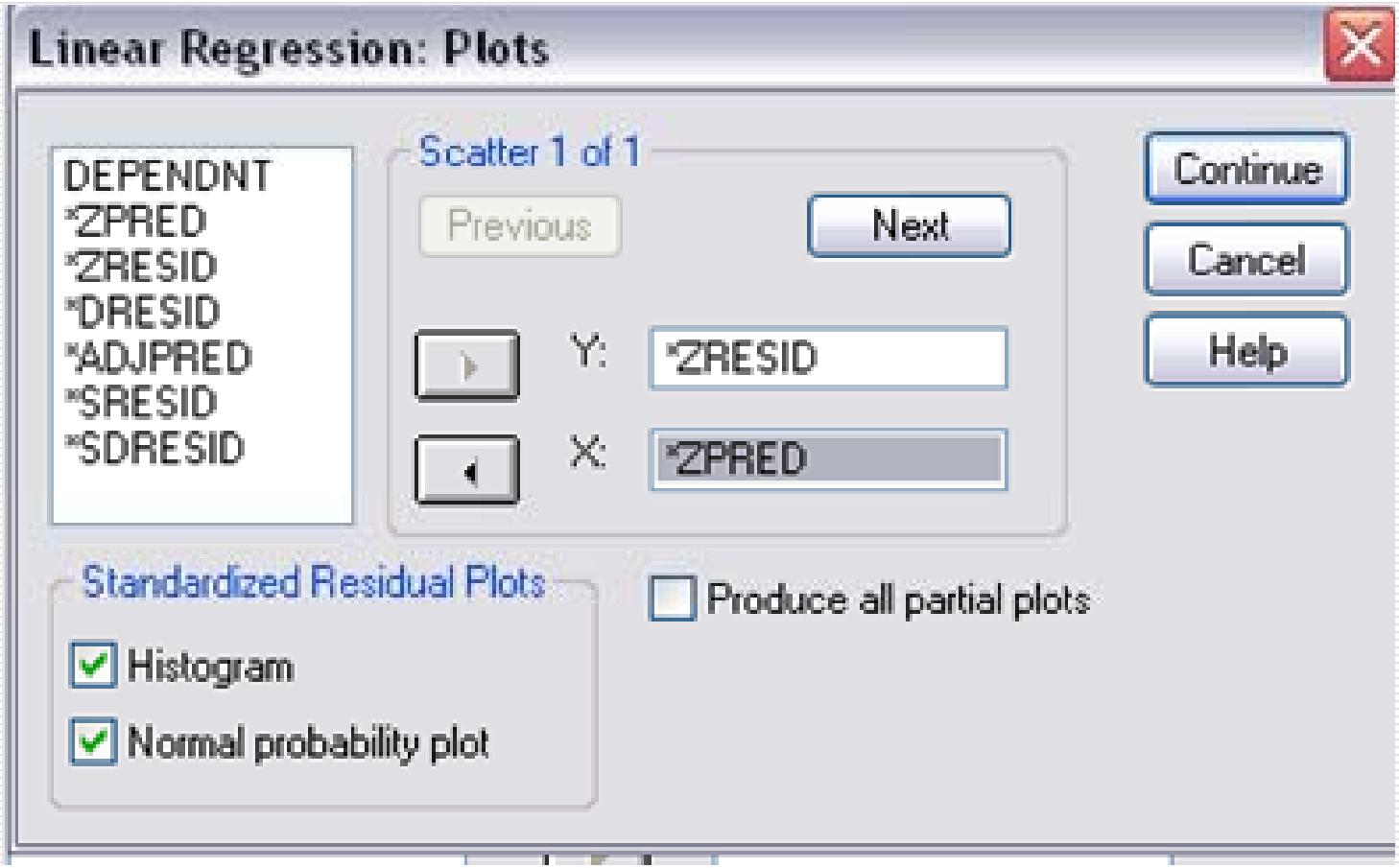
Προσαρμογή του μοντέλου

Στατιστικά μέτρα για τους συντελεστές της παλινδρόμησης

Στατιστικά μέτρα υπολοίπων

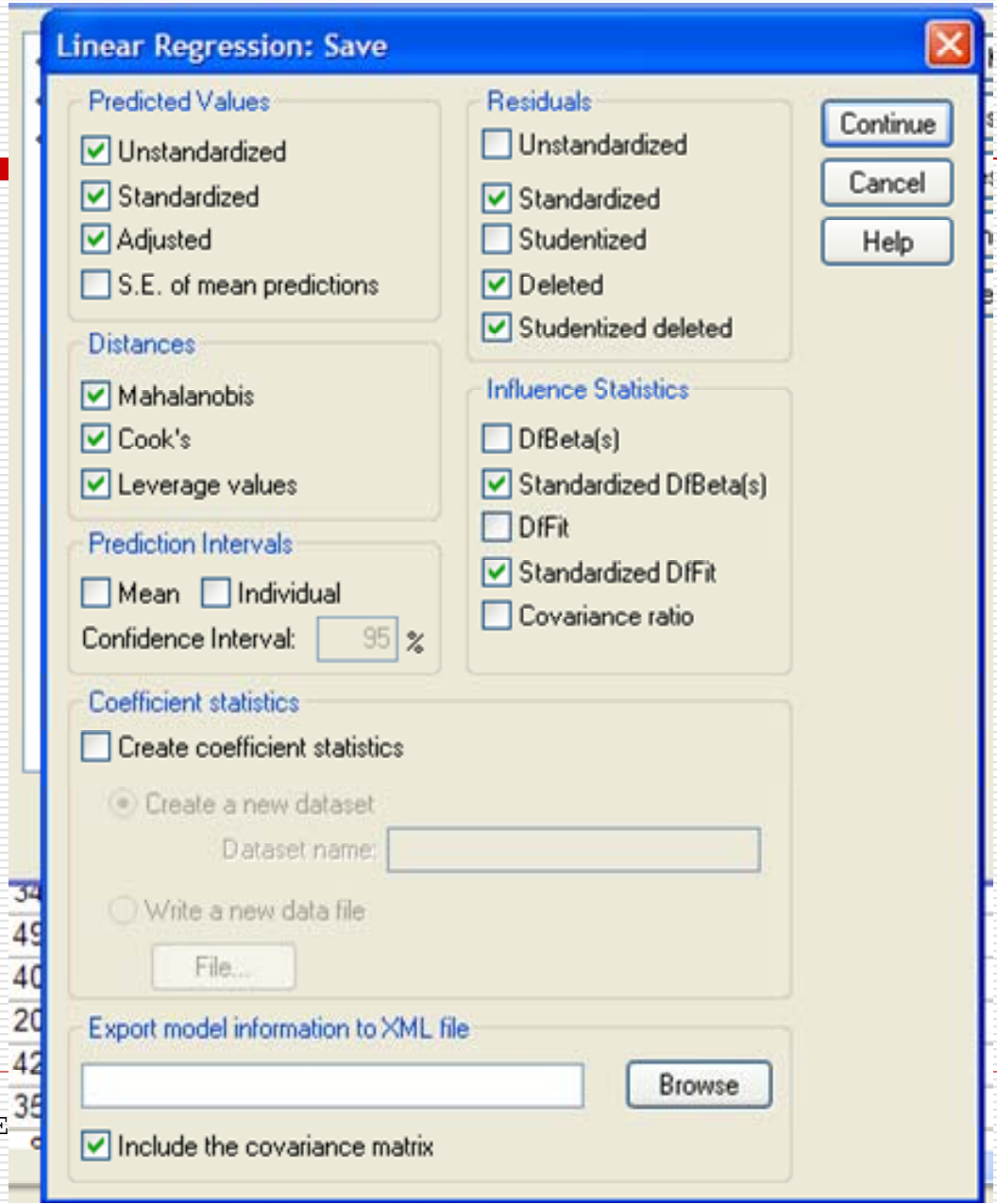


Plots-Γραφικές παραστάσεις για ανάλυση των υπολοίπων

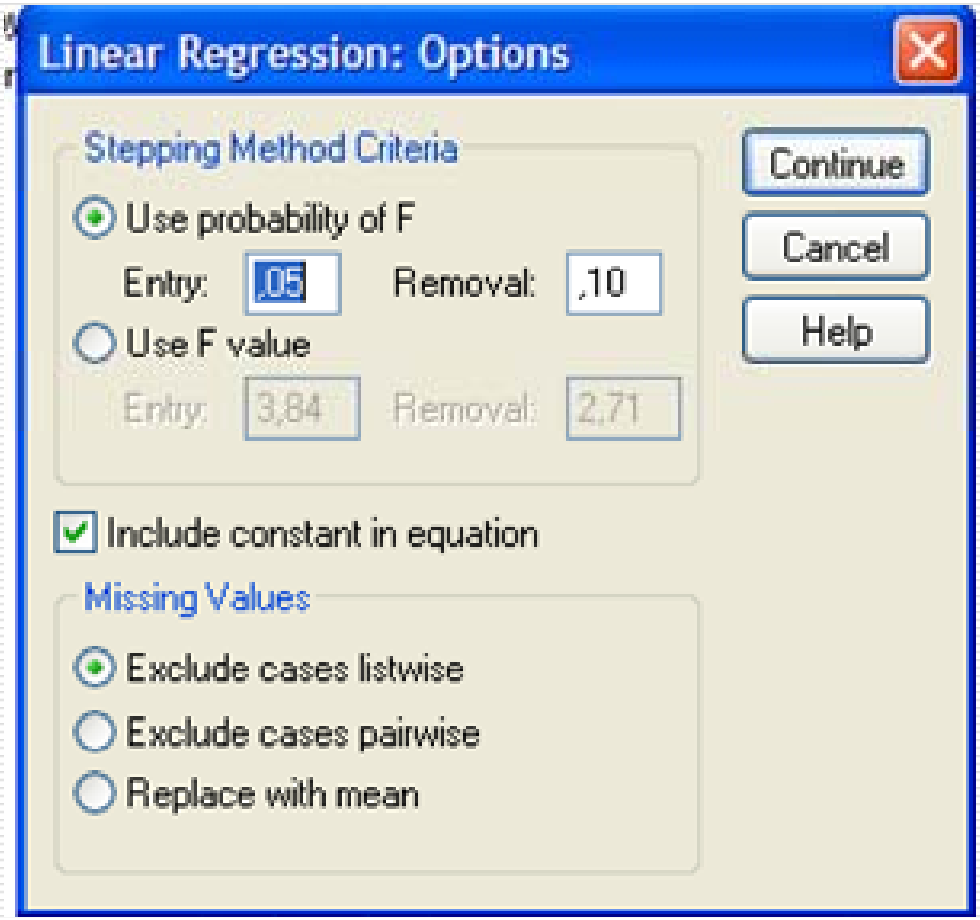


Save regression diagnostics

Υπόλοιπα, Προβλέψεις και Στατιστικά μέτρα υπολοίπων αποθηκεύονται ως νέες μεταβλητές



Κριτήρια για αλγόριθμους με βήματα, διαχείριση χαμένων τιμών και υπολογισμός σταθεράς



Αποτελέσματα παλινδρόμησης

Descriptive Statistics

	Mean	Std. Deviation	N
Record Sales (thousands)	193,2000	80,69896	200
Advertsing Budget (thousands of pounds)	614,4123	485,65521	200
No. of plays on Radio 1 per week	27,5000	12,26958	200
Attractiveness of Band	6,7700	1,39529	200

- Περίληψη όλων των μεταβλητών

Συσχετίσεις ανάμεσα στις μεταβλητές (δεν φαίνεται multicollinearity)

Correlations

		Record Sales (thousands)	Advertsing Budget (thousands of pounds)	No. of plays on Radio 1 per week	Attractiveness of Band
Pearson Correlation	Record Sales (thousands)	1,000	,578	,599	,326
	Advertsing Budget (thousands of pounds)	,578	1,000	,102	,081
	No. of plays on Radio 1 per week	,599	,102	1,000	,182
	Attractiveness of Band	,326	,081	,182	1,000
Sig. (1-tailed)	Record Sales (thousands)	.	,000	,000	,000
	Advertsing Budget (thousands of pounds)	,000	.	,076	,128
	No. of plays on Radio 1 per week	,000	,076	.	,005
	Attractiveness of Band	,000	,128	,005	.
N	Record Sales (thousands)	200	200	200	200
	Advertsing Budget (thousands of pounds)	200	200	200	200
	No. of plays on Radio 1 per week	200	200	200	200
	Attractiveness of Band	200	200	200	200

Προσαρμογή του μοντέλου

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,815[a]	,665	,660	47,08734	1,950

a. Predictors: (Constant), Attractiveness of Band, Advertsing Budget (thousands of pounds), No. of plays on Radio 1 per week

b. Dependent Variable: Record Sales (thousands)

- Το μοντέλο εξηγεί 66.5% της συνολικής μεταβλητότητας
- Το adjusted R2 δεν είναι πολύ μικρότερο και δείχνει ότι το μοντέλο μπορεί να γενικευτεί στον πληθυσμό
- Το D-W είναι κοντά στο 2 οπότε σύμφωνα με εμπειρικό κανόνα τα σφάλματα είναι ανεξάρτητα (**ανησυχούμε για τιμές <1 ή >3 !!**)

Σημαντικότητα του μοντέλου

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	861377,4	3	287125,806	129,498	,000[a]
	Residual	434574,6	196	2217,217		
	Total	1295952	199			

a. Predictors: (Constant), Attractiveness of Band, Advertsing Budget (thousands of pounds), No. of plays on Radio 1 per week

b. Dependent Variable: Record Sales (thousands)

- Το F-test δίνει sig.<0.001 οπότε το μοντέλο είναι πολύ σημαντικό στην εξήγηση της μεταβλητότητας

Παράμετροι του μοντέλου

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-26,613	17,350		-1,534	,127	-60,830	7,604		
	Advertsing Budget (thousands of pounds)	,085	,007	,511	12,261	,000	,071	,099	,986	1,015
	No. of plays on Radio 1 per week	3,367	,278	,512	12,123	,000	2,820	3,915	,959	1,043
	Attractiveness of Band	11,086	2,438	,192	4,548	,000	6,279	15,894	,963	1,038

a. Dependent Variable: Record Sales (thousands)

sales=-26.61 + 0.085*adverts
+3.367*airplay + 11.086*attract

Ερμηνεία του μοντέλου (1/2)

- Όλοι οι συντελεστές των μεταβλητών είναι θετικοί οπότε όσο αυξάνουν οι τιμές των μεταβλητών αυξάνονται οι πωλήσεις
- Το μέγεθος του κάθε συντελεστή δείχνει πόσο αυξάνονται οι πωλήσεις όταν αυξηθεί κατά 1 μονάδα η αντίστοιχη μεταβλητή κρατώντας τις υπόλοιπες σταθερές

Ερμηνεία του μοντέλου (2/2)

- Τα t-tests για κάθε συντελεστή των μεταβλητών δίνουν sig.<0.001 και επομένως όλες οι μεταβλητές είναι σημαντικές
- Το μέγεθος του t μας δείχνει ότι η διαφήμιση και η ραδιοφωνική μετάδοση είναι εξίσου σημαντικές ενώ η ελκυστικότητα λιγότερο σημαντική (το ίδιο προκύπτει και από τους standardized coefficients)

Multicollinearity

- Πρόβλημα όταν:
  - max(VIF)>10
  - mean(VIF)>>1
  - Tolerance<0.1 ή 0.2
- Εδώ δεν υπάρχει πρόβλημα!
- Υπάρχουν και άλλα στατιστικά μέτρα

Casewise diagnostics

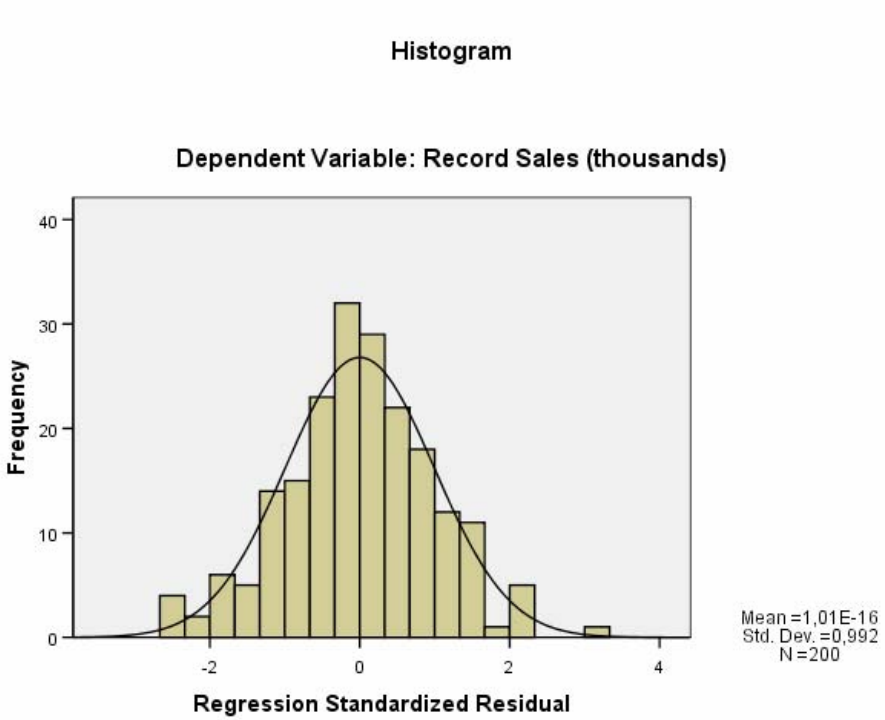
Casewise Diagnostics[a]

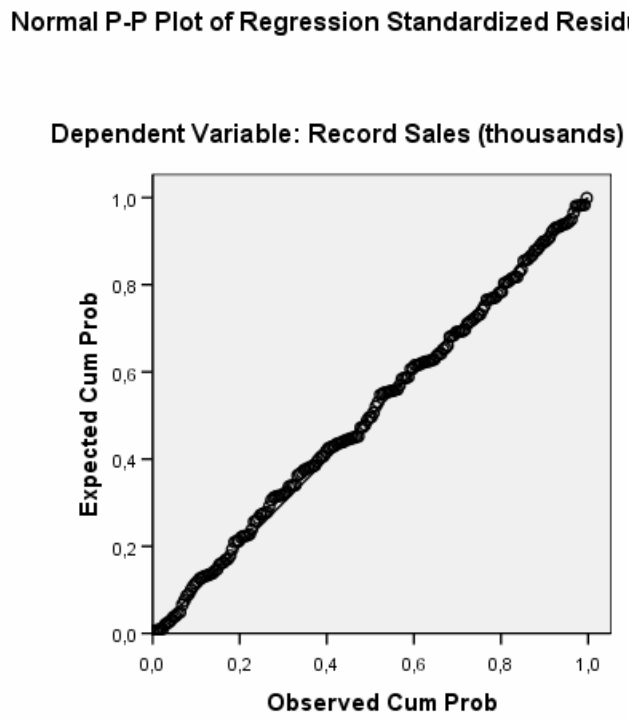
Case Number	Std. Residual	Record Sales (thousands)	Predicted Value	Residual
1	2,125	330,00	229,9203	100,07975
2	-2,314	120,00	228,9490	-108,949
10	2,114	300,00	200,4662	99,53375
47	-2,442	40,00	154,9698	-114,970
52	2,069	190,00	92,5973	97,40266
55	-2,424	190,00	304,1231	-114,123
61	2,098	300,00	201,1897	98,81030
68	-2,345	70,00	180,4156	-110,416
100	2,066	250,00	152,7133	97,28666
164	-2,577	120,00	241,3240	-121,324
169	3,061	360,00	215,8675	144,13246
200	-2,064	110,00	207,2061	-97,20606

a. Dependent Variable: Record Sales (thousands)

- Λογικός ο αριθμός των μεγάλων υπολοίπων (12/200)

Γραφική ανάλυση υπολοίπων





Γραφική ανάλυση υπολοίπων