

ΙΙΜΣ, ΕΠΑΓΓΕΛΜΑΤΙΚΗ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗ ΥΓΕΙΑ, ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΟΙΚΟΝΟΜΙΚΗ ΑΠΟΤΙΜΗΣΗ
ΑΚ. ΕΤΟΣ 2006-2007, 3ο εξάμηνο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

ΜΑΘΗΜΑ 12

Συμπερασματολογία για την επίδραση πολλών μεταβλητών σε μια ποσοτική (Πολλαπλή Παλινδρόμηση) [μέρος 2ο]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-95

11. Επίδραση πολλών μεταβλητών σε μια ποσοτική ΠΕΡΙΕΧΟΜΕΝΑ Β' ΜΕΡΟΥΣ

- Δείκτες μερικής συσχέτισης (Partial correlations) (+ Συνέχεια παραδείγματος 11-1)
- Πολυωνυμική Παλινδρόμηση
- Χρήση κατηγορικών συμμεταβλητών σε παλινδρομικά μοντέλα
- Σχέση παλινδρόμησης και ANOVA/t-tests
- Αναλ. Συνδιακύμανσης (ANCOVA)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-96

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations)

- Έστω $X_1, X_2, X_3, \dots, X_p$
- Περιγράφουν τη γραμμική σχέση 2 μεταβλητών (π.χ. X_1 & X_2) αφού ελέγξουμε (after controlling for) για την επίδραση των υπόλοιπων μεταβλητών (δηλ. X_3, \dots, X_p).
 - ✓ [αφού ελέγξουμε (after controlling for) \Leftrightarrow εξουδετέρωση των επιδράσεων που οφείλονται σε άλλες μεταβλητές]
- ΥΠΟΛΟΓΙΣΜΟΣ
 - ✓ Υπολογίζουμε κατάλοιπα
 - ✓ e_1 της παλινδρόμησης $X_1 = \beta_0^{(1)} + \beta_3^{(1)} X_3^{(1)} + \dots + \beta_p^{(1)} X_p^{(1)}$
 - ✓ e_2 της παλινδρόμησης $X_2 = \beta_0^{(2)} + \beta_3^{(2)} X_3^{(2)} + \dots + \beta_p^{(2)} X_p^{(2)}$
 - ✓ $r_{12}^{\text{partial}} = \text{Cor}(e_1, e_2)$
- Έχει ίδιες ιδιότητες με το δείκτη του Pearson.
- Προϋποθέσεις: Κανονικότητα, γραμμικότητα κλπ. (βλ. παλινδρόμηση)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-97

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

ΔΙΑΔΙΚΑΣΙΑ ΣΤΟ SPSS

1. Επιλέγουμε το menu:
Analyze>Correlate>Partial...
2. Επιλέγουμε 2 ή περισσότερες ποσοτικές μεταβλητές για τις οποίες επιθυμούμε να υπολογίσουμε τους δείκτες μερικής συσχέτισης
3. Επιλέγουμε 2 ή περισσότερες μεταβλητές για των οποίων τις επιδράσεις στις μεταβλητές του βήματος 2 θέλουμε να ελέγξουμε (δηλ. να εξουδετερώσουμε).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-98

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

Μεταβλητές για τις οποίες θα υπολογίσουμε τους δείκτες μερικής συσχέτισης

Μεταβλητές των οποίων την επιδράσεις ελέγχουμε (εξουδετερώνουμε)

Διαφάνεια 12-99

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

ΑΠΛΟΙ ΔΕΙΚΤΕΣ ΓΡΑΜΜΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ ΤΟΥ PEARSON

Διαφάνεια 12-101

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

Correlations

Control Variables		price Price in US Dollars	living Size of Living Area in Sq. ft.
lotsize Lot Size in Sq. ft.	price Price in US Dollars	1.000	.778
		Correlation	.000
		Significance (2-tailed)	.27
	living Size of Living Area in Sq. ft.	.778	1.000
		Correlation	.000
		Significance (2-tailed)	.27

Διαφάνεια 12-100

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

Correlations

Control Variables		price Price in US Dollars	living Size of Living Area in Sq. ft.	lotsize Lot Size in Sq. ft.
-none- ^a	price Price in US Dollars	1.000	.963	.907
		Correlation	.000	.000
		Significance (2-tailed)	0	.28
	living Size of Living Area in Sq. ft.	.963	1.000	.961
		Correlation	.000	.000
		Significance (2-tailed)	0	.28
	lotsize Lot Size in Sq. ft.	.907	.961	1.000
		Correlation	.000	.000
		Significance (2-tailed)	.28	0
lotsize Lot Size in Sq. ft.	price Price in US Dollars	1.000	.778	
		Correlation	.000	
		Significance (2-tailed)	0	.27
	living Size of Living Area in Sq. ft.	.778	1.000	
		Correlation	.000	
		Significance (2-tailed)	.27	0

^a Cells contain zero-order (Pearson) correlations.

Διαφάνεια 12-102

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

■ ΣΤΗΝ ΠΑΛΙΝΔΡΟΜΗΣΗ

Διαφάνεια 12-103

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	Correlations		
		B	Std. Error	Beta				Zero-order	Partial	Part
1	(Constant)	22149.972	9552.441			2.319	.028			
	living Size of Living Area in Sq. ft	77.024	11.967	1.180	6.436	.000	.963	.778	.327	
	lotsize Lot Size in Sq. ft	-2.345	1.900	-.226	-1.234	.228	.907	-.231	-.063	

a. Dependent Variable: price Price in US Dollars

ΔΕΙΚΤΕΣ ΜΕΡΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ ΤΙΜΗΣ + ΜΕΓΕΘΟΥΣ ΟΙΚΗΜΑΤΟΣ

Control Variables		price Price in US Dollars	living Size of Living Area in Sq. ft	lotsize Lot Size in Sq. ft
Correlation	-none ^a	1.000	.963	.907
	price Price in US Dollars	1.000	.963	.961
	living Size of Living Area in Sq. ft	.963	1.000	.961
	lotsize Lot Size in Sq. ft	.907	.961	1.000
	price Price in US Dollars	1.000	.778	
	living Size of Living Area in Sq. ft	.778	1.000	

Διαφάνεια 12-105

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	Correlations		
		B	Std. Error	Beta				Zero-order	Partial	Part
1	(Constant)	22149.972	9552.441			2.319	.028			
	living Size of Living Area in Sq. ft	77.024	11.967	1.180	6.436	.000	.963	.778	.327	
	lotsize Lot Size in Sq. ft	-2.345	1.900	-.226	-1.234	.228	.907	-.231	-.063	

a. Dependent Variable: price Price in US Dollars

ΔΕΙΚΤΕΣ ΓΡΑΜΜΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ ΤΟΥ PEARSON

Control Variables		price Price in US Dollars	living Size of Living Area in Sq. ft	lotsize Lot Size in Sq. ft
Correlation	-none ^a	1.000	.963	.907
	price Price in US Dollars	1.000	.963	.961
	living Size of Living Area in Sq. ft	.963	1.000	.961
	lotsize Lot Size in Sq. ft	.907	.961	1.000
	price Price in US Dollars	1.000	.778	
	living Size of Living Area in Sq. ft	.778	1.000	

Διαφάνεια 12-104

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	Correlations		
		B	Std. Error	Beta				Zero-order	Partial	Part
1	(Constant)	22149.972	9552.441			2.319	.028			
	living Size of Living Area in Sq. ft	77.024	11.967	1.180	6.436	.000	.963	.778	.327	
	lotsize Lot Size in Sq. ft	-2.345	1.900	-.226	-1.234	.228	.907	-.231	-.063	

a. Dependent Variable: price Price in US Dollars

ΔΕΙΚΤΕΣ ΜΕΡΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ ΤΙΜΗΣ + ΜΕΓΕΘΟΥΣ ΟΙΚΟΠΕΔΟΥ

Control Variables		price Price in US Dollars	living Size of Living Area in Sq. ft	lotsize Lot Size in Sq. ft
Correlation	-none ^a	1.000	.963	.907
	price Price in US Dollars	1.000	.963	.961
	living Size of Living Area in Sq. ft	.963	1.000	.961
	lotsize Lot Size in Sq. ft	.907	.961	1.000
	price Price in US Dollars	1.000	.778	
	living Size of Living Area in Sq. ft	.778	1.000	

Διαφάνεια 12-106

11.6. Δείκτες μερικής συσχέτισης (Partial Correlations) Παράδειγμα 11-1 (συνέχεια)

ΜΗΜΑΤΙΚΟΙ ΔΕΙΚΤΕΣ ΣΥΣΧΕΤΙΣΗΣ

- ΤΙΜΗΣ + ΜΕΓΕΘΟΥΣ ΟΙΚΗΜΑΤΟΣ
- ΤΙΜΗΣ + ΜΕΓΕΘΟΥΣ ΟΙΚΟΠΕΔΟΥ

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	Correlations			
		B	Std. Error	Beta				Zero-order	Partial	Part	
1	(Constant)	22149.972	9552.441			2.319	.028				
	living: Size of Living Area in Sq. ft.	77.024	11.967			6.436	.000	.963	.778		.327
	lotsize: Lot Size in Sq. ft.	-2.345	1.900			-1.234	.228	.907	-.231		-.063

a. Dependent Variable: price: Price in US Dollars

Part Correlation: Συσχέτιση μεταξύ εξαρτημένης μεταβλητής Y (στο παρ ⇔ ΤΙΜΗ) και της κάθε ανεξάρτητης X όταν έχουμε ελέγξει την τελευταία για τις επιδράσεις των υπόλοιπων ανεξάρτητων μεταβλητών.

$r_{YX_i}^{PART} = \text{Cor}(Y, e_i)$ με e_i τα κατάλοιπα της παλινδρόμησης

$$X_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

➤ Επίσης ονομάζεται και δείκτης Ημι-μερικής Συσχέτισης (semipartial correlation).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-107

11.7. Πολυωνυμική Παλινδρόμηση

- Πολλές φορές η παλινδρόμηση μπορεί να είναι πολυωνυμικής μορφής.
- Είναι παρόμοιο με πολλαπλή παλινδρόμηση με ανεξάρτητες μεταβλητές τις δυνάμεις της μεταβλητής X
- Δηλ.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_p X^p + \varepsilon$$
- Γιατί είναι σημαντική; Διότι μπορούμε (βάζοντας αρκετούς όρους) να προσεγγίσουμε ικανοποιητικά οποιαδήποτε τύπου σχέση (δηλ. Συνάρτηση βλ. Taylor expansion στον *απειροστικό λογισμό*)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-108

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-1 (SPSS: Curve estimation)

• $Y = b_0 + b_1 X + b_2 X^2$

• $Y = b_0 + b_1 X + b_2 X^2 + b_3 X^3$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-109

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-1 (SPSS: Curve estimation)

Model Summary and Parameter Estimates

Dependent Variable: price: Price in US Dollars

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.926	352.672	1	28	.000	13597.842	62.829		
Quadratic	.927	171.703	2	27	.000	26514.620	49.154	.003	
Cubic	.939	134.496	3	26	.000	233853.4	-283.007	.173	-2.8E-005

The independent variable is living: Size of Living Area in Sq. ft.

Price in US Dollars

Size of Living Area in Sq. ft.

○ Observed
— Linear
- - Quadratic
- - - Cubic

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-110

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-1 (SPSS: Curve estimation)

ΤΥΠΩΝΕΙ ΠΕΡΙΣΣΟΤΕΡΕΣ ΛΕΠΤΟΜΕΡΕΙΕΣ ΓΙΑ ΚΑΘΕ ΜΟΝΤΕΛΟ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-111

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-1 (SPSS: Curve estimation)

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
living Size of Living Area in Sq. ft	-283.007	146.491	-4.336	-1.932	.064
living Size of Living Area in Sq. ft ** 2	.173	.074	10.778	2.341	.027
living Size of Living Area in Sq. ft ** 3	-2.8E-005	.000	-5.552	-2.304	.029
(Constant)	233853.4	93421.807		2.503	.019

ΜΗ ΣΤΑΤΙΣΤΙΚΑ ΣΗΜΑΝΤΙΚΟΣ ΟΡΟΣ

ΣΤΑΤΙΣΤΙΚΑ ΣΗΜΑΝΤΙΚΟΙ ΟΡΟΙ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-112

11.7. Πολυωνυμική Παλινδρόμηση ΕΡΜΗΝΕΙΑ

ΠΡΟΣΟΧΗ

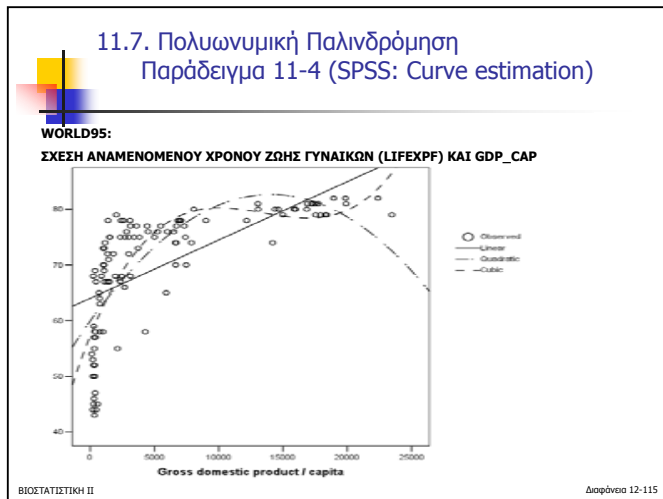
- ΕΔΩ ΤΟ R² ΕΙΝΑΙ ΗΔΗ ΥΨΗΛΟ ΜΕ ΤΟ ΓΡΑΜΜΙΚΟ ΜΟΝΟ ΟΡΟ
- ΠΡΟΣΘΕΣΗ ΠΟΛΥΩΝΥΜΙΚΩΝ ΟΡΩΝ ΔΕ ΜΑΣ ΒΕΛΤΙΩΝΕΙ ΠΟΛΥ ΤΟ ΜΟΝΤΕΛΟ ΑΛΛΑ ΜΑΣ ΔΥΣΚΟΛΕΥΕΙ ΠΟΛΥ ΤΗΝ ΕΡΜΗΝΕΙΑ
- ΣΥΝΕΠΩΣ ΕΔΩ ΕΙΝΑΙ ΚΑΛΥΤΕΡΑ ΝΑ ΜΕΙΝΟΥΜΕ ΣΤΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ ΑΚΟΜΑ ΚΑΙ ΚΑΠΟΙΟ ΑΛΛΟ ΠΟΛΥΩΝΥΜΙΚΟ ΜΟΝΤΕΛΟ ΕΙΝΑΙ ΛΙΓΟ ΚΑΛΥΤΕΡΟ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-113

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-4

WORLD95:
ΣΧΕΣΗ ΑΝΑΜΕΝΟΜΕΝΟΥ ΧΡΟΝΟΥ ΖΩΗΣ ΓΥΝΑΙΚΩΝ (LIFEXPF) ΚΑΙ GDP_CAP

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-114



11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-4 (SPSS: Curve estimation)

WORLD95:
ΣΧΕΣΗ ΑΝΑΜΕΝΟΜΕΝΟΥ ΧΡΟΝΟΥ ΖΩΗΣ ΓΥΝΑΙΚΩΝ (LIFEXPF) ΚΑΙ GDP_CAP

Model Summary and Parameter Estimates

Dependent Variable: lifexpf Average female life expectancy

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.412	75.115	1	107	.000	64.016	.001		
Quadratic	.544	63.352	2	106	.000	59.951	.003	-1.1E-007	
Cubic	.604	53.315	3	105	.000	57.270	.006	-4.5E-007	1.12E-011

The independent variable is gdp_cap Gross domestic product / capita.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-116

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-4 (SPSS: Stepwise procedures)

WORLD95:
ΣΧΕΣΗ ΑΝΑΜΕΝΟΜΕΝΟΥ ΧΡΟΝΟΥ ΖΩΗΣ ΓΥΝΑΙΚΩΝ (LIFEXPF) ΚΑΙ GDP_CAP ΥΠΟΛΟΓΙΖΟΥΜΕ ΩΣ ΚΑΙ ΤΗ 10η ΔΥΝΑΜΗ ΜΕ ΤΟ TRANSFORM>COMPUTE

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-117

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-4 (SPSS: Stepwise procedures)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-118

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-4 (SPSS: Stepwise procedures)

Variables Entered/Removed ^a			
Model	Variables Entered	Variables Removed	Method
1	gdp_cap Gross domestic product		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-
2	G2		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-
3	G3		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-
4	G4		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-

^a Dependent Variable: lifeexpf Average female life expectancy

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.642 ^a	.412	.407	8.141
2	.738 ^b	.544	.536	7.202
3	.777 ^c	.604	.592	6.750
4	.820 ^d	.673	.661	6.159

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-119

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-4 (SPSS: Stepwise procedures)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-121

11.7. Πολυωνυμική Παλινδρόμηση Παράδειγμα 11-4 (SPSS: Stepwise procedures)

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
4	(Constant)	53.446	1.439		37.138	.000
	gdp_cap Gross domestic product / capita	.011	.001	.6860	8.320	.000
	G2	-1.8E-006	.000	-.21556	-6.062	.000
	G3	1.09E-010	.000	.25969	5.198	.000
	G4	-2.2E-015	.000	-.10657	-4.703	.000

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-120

11.7. Πολυωνυμική Παλινδρόμηση ΕΡΜΗΝΕΙΑ ΤΕΤΡΑΓΩΝΙΚΩΝ ΜΟΝΤΕΛΩΝ

$$Y = b_0 + b_1x + b_2x^2$$

- ΜΕΤΑΒΟΛΗ ΜΙΑΣ ΜΟΝΑΔΑΣ ΣΤΗΝ Χ ΣΥΝΕΠΙΓΕΤΑΙ ⇔ $\mu_{x+1} - \mu_x = b_1 + b_2(2x+1)$ [ΕΞΑΡΤΑΤΑΙ ΑΠΟ ΤΟ Χ]
- Αν $b_2 > 0$ ⇔ ελάχιστο για $x = -b_1 / (2b_2)$
- Αν $b_2 < 0$ ⇔ μέγιστο για $x = -b_1 / (2b_2)$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-122

11.8. Χρήση κατηγορικών συµµεταβλητών σε παλινδροµικά µοντέλα

ΤΟ ΜΟΝΤΕΛΟ ΤΗΣ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ

- Έστω ότι έχουμε $k=1,2,\dots,K$ ομάδες και σε κάθε ομάδα $j=1,2, \dots, n_k$ παρατηρήσεις
 $Y_{kj} = \mu_k + \varepsilon, \varepsilon \sim N(0, \sigma^2) \Leftrightarrow$
 $Y_{kj} = \mu + a_k + \varepsilon, \varepsilon \sim N(0, \sigma^2)$
- a_k ονομάζεται επίδραση του k επιπέδου της κατηγορικής µεταβλητής
- Για να έχουμε σωστή παραµετροποίηση πρέπει να χρησιµοποιήσουµε 1 περιορισµό
- Ερµηνεία µας γίνεται ανάλογα την παραµετροποίηση

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διοφάνεια 12-123

11.8. Χρήση κατηγορικών συµµεταβλητών σε παλινδροµικά µοντέλα

- Για να χρησιµοποιήσουµε κατηγορικές µεταβλητές χρειαζόµαστε να φτιάξουµε ψευδοµεταβλητές (Dummy variables)
- Για k επίπεδα χρειαζόµαστε $k-1$ ψευδοµεταβλητές
- 2 συνηθισµένοι τύποι
 - ✓ Μεταβλητές 0-1 (Γωνιακοί περιορισµοί – Corner Constraints) $a_1 = 0$
 [π.χ. $a_1=0$] : Γίνεται χρήση ενός βασικού επιπέδου αναφοράς (baseline reference category)
 - ✓ Ψευδοµεταβλητές Μηδενικού Αθροίσµατος (sum-to-zero constraints): Κάθε παράµετρος µετράει αποκλίσεις από το «µέσο» των επιδράσεων

$$\sum_{k=1}^K a_k = 0 \Leftrightarrow a_1 = -\sum_{k=2}^K a_k$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διοφάνεια 12-124

11.8. Χρήση κατηγορικών συµµεταβλητών σε παλινδροµικά µοντέλα

- Γωνιακοί περιορισµοί – Corner Constraints

Επίπεδο	D_2	D_3	...	D_{k-1}	D_k
1	0	0	...	0	0
2	1	0	...	0	0
...
$k-1$	0	0	...	1	0
k	0	0	...	0	1

- Αν $X=1$ (επίπεδο αναφοράς) τότε όλες οι dummies θα έχουν τιμή µηδέν (0)
- Αν $X=i>1$ τότε η i dummy θα έχει τιμή ένα (1) ενώ όλες οι άλλες είναι µηδέν (0).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διοφάνεια 12-125

11.8. Χρήση κατηγορικών συµµεταβλητών σε παλινδροµικά µοντέλα

- Περιορισµοί Μηδενικού Αθροίσµατος – Sum-to-zero Constraints

Επίπεδο	D_2	D_3	...	D_{k-1}	D_k
1	-1	-1	...	-1	-1
2	1	0	...	0	0
...
$k-1$	0	0	...	1	0
k	0	0	...	0	1

- Αν $X=1$ (επίπεδο αναφοράς) τότε όλες οι dummies θα έχουν τιμή µειόν ένα (-1)
- Αν $X=i>1$ τότε η i dummy θα έχει τιμή ένα (1) ενώ όλες οι άλλες είναι µηδέν (0).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διοφάνεια 12-126

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests

Γωνιακοί περιορισμοί – Corner Constraints

- $Y = \beta_0 + \beta_2 D_2 + \epsilon$
- Αν $X=1$ (επίπεδο 1 – αναφοράς)
 - ◊ $D_2 = 0 \Leftrightarrow Y = \beta_0 + \epsilon$
 - ◊ η μέση (αναμενόμενη) τιμή του 1ου επιπέδου είναι ίση με τη σταθερά β_0
- Αν $X=2$ (επίπεδο 2)
 - ◊ $D_2 = 1 \Leftrightarrow Y = \beta_0 + \beta_2 + \epsilon$
 - ◊ η μέση (αναμενόμενη) τιμή του 2ου επιπέδου είναι ίση με $\beta_0 + \beta_2$
- Συνεπώς β_2 είναι η διαφορά των μέσων τιμών στα δύο επίπεδα (αυτό δεν κάνει το t-test;)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-127

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests

ΠΑΡΑΔΕΙΓΜΑ 4-6 [05_dataset3]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
score	Equal variances assumed	.621	.440	-1.528	20	.142	-4.73333	3.09734	-11.19428	1.72761
	Equal variances not assumed			-1.509	19.846	.143	-4.73333	3.01752	-11.03090	1.56424

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	83.167	2.088		39.826	.000
	d2	-4.733	3.097	.323	-1.528	.142

a. Dependent Variable: score

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-128

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests

Περιορισμοί μηδενικού αθροίσματος– STZ Constraints

- $Y = \beta_0 + \beta_2 D_2 + \epsilon$
- Αν $X=1 \Leftrightarrow D_2 = -1 \Leftrightarrow Y = \beta_0 - \beta_2 + \epsilon$
 - ◊ η μέση (αναμενόμενη) τιμή του 1ου επιπέδου είναι ίση με τη σταθερά $\beta_0 - \beta_2$
- Αν $X=2$ (επίπεδο 2)
 - ◊ $D_2 = 1 \Leftrightarrow Y = \beta_0 + \beta_2 + \epsilon$
 - ◊ η μέση (αναμενόμενη) τιμή του 2ου επιπέδου είναι ίση με $\beta_0 + \beta_2$
- Συνεπώς β_2 είναι η απόκλιση κάθε επιπέδου από το συνολικό μέσο
- Συνεπώς β_0 είναι εκτίμηση για το συνολικό μέσο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-129

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests

ΠΑΡΑΔΕΙΓΜΑ 4-6 [05_dataset1.dat]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
score	Equal variances assumed	.621	.440	-1.528	20	.142	-4.73333	3.09734	-11.19428	1.72761
	Equal variances not assumed			-1.509	19.846	.143	-4.73333	3.01752	-11.03090	1.56424

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	83.83	1.549		55.230	.000
	d2stz	-2.367	1.549	.323	-1.528	.142

a. Dependent Variable: score

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-130

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests

- Οι 2 παλινδρομήσεις είναι ισοδύναμες
- Οι έλεγχοι για το β_2 είναι ισοδύναμοι με τα independent samples t-tests.
- Το ίδιο συμβαίνει και με την ANOVA

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-131

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests
Παράδειγμα [05_dataset03]■ **Γωνιακοί περιορισμοί – Corner Constraints**

- $k=3$ επίπεδα
- $Y = \beta_0 + \beta_2 D_2 + \beta_3 D_3 + \varepsilon$
- Αν $X=1$ (επίπεδο 1 – αναφοράς)
 - ⇔ $D_2 = 0, D_3 = 0$ ⇔ $Y = \beta_0 + \varepsilon$
 - ⇔ η μέση (αναμενόμενη) τιμή του 1ου επιπέδου είναι ίση με τη σταθερά β_0
- Αν $X=2$ (επίπεδο 2)
 - ⇔ $D_2 = 1, D_3 = 0$ ⇔ $Y = \beta_0 + \beta_2 + \varepsilon$
 - ⇔ η μέση (αναμενόμενη) τιμή του 2ου επιπέδου είναι ίση με $\beta_0 + \beta_2$
- Αν $X=3$ (επίπεδο 2)
 - ⇔ $D_2 = 0, D_3 = 1$ ⇔ $Y = \beta_0 + \beta_3 + \varepsilon$
 - ⇔ η μέση (αναμενόμενη) τιμή του 2ου επιπέδου είναι ίση με $\beta_0 + \beta_3$
- Συνεπώς β_2 είναι η διαφορά του 2ου επιπέδου από το επ. Αναφοράς
- Συνεπώς β_3 είναι η διαφορά του 3ου επιπέδου από το επ. αναφοράς

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-132

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests
Παράδειγμα [05_dataset03]■ **Γωνιακοί περιορισμοί – Corner Constraints**

- ΓΕΝΙΚΟΙ ΚΑΝΟΝΕΣ
- Για $1 < i < k$
- Αν $X=i$ (επίπεδο i)
 - ⇔ $D_i = 1, D_k = 0$ για $k \neq i$ ⇔ $Y = \beta_0 + \beta_i + \varepsilon$
 - ⇔ η μέση (αναμενόμενη) τιμή του $<i>$ επιπέδου είναι ίση με $\beta_0 + \beta_i$
- Συνεπώς β_0 είναι ο μέσος του επιπέδου αναφοράς
- Συνεπώς β_i είναι η διαφορά του $<i>$ επιπέδου από το επίπεδο αναφοράς

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-133

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests
Παράδειγμα [05_dataset03]■ **Γωνιακοί περιορισμοί – Corner Constraints**

ANOVA					
grade					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	250.000	2	125.000	3.348	.070
Within Groups	448.000	12	37.333		
Total	698.000	14			

 $H_0: \mu_1 = \mu_2 = \mu_3$

ANOVA ^a ΠΑΛΙΝΔΡΟΜΗΣΗ						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	250.000	2	125.000	3.348	.070 ^b
	Residual	448.000	12	37.333		
	Total	698.000	14			

 $H_0: \beta_2 = \beta_3 = 0$

a. Predictors: (Constant), d3, d2

b. Dependent Variable: grade

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-134

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests Παράδειγμα [05_dataset03]

Γωνιακοί περιορισμοί – Corner Constraints

Dependent Variable: grade

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	75.000	2.733	27.447	.000	69.046	80.954
[method=1]	5.000	3.864	1.294	.220	-3.420	13.420
[method=2]	10.000	3.864	2.588	.024	1.580	18.420
[method=3]	0 ^a					

Coefficients^a ΠΑΛΙΝΔΡΟΜΗΣΗ

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.
		B	Std. Error	Beta			
1	(Constant)	80.000	2.733			29.277	.000
	d2	5.000	3.864	.346		1.294	.220
	d3	-5.000	3.864	-.346		-1.294	.220

a. Dependent Variable: grade

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-135

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests Παράδειγμα [05_dataset03]

Περιορισμοί μηδενικού αθροίσματος – Sum-to-zero Constraints

- κ=3 επίπεδα
- $Y = \beta_0 + \beta_2 D_2 + \beta_3 D_3 + \varepsilon$
- Αν $X=1$ (επίπεδο 1 – αναφοράς)
 - $\Leftrightarrow D_2 = -1, D_3 = -1 \Leftrightarrow Y = \beta_0 - \beta_2 - \beta_3 + \varepsilon$
 - \Leftrightarrow η μέση (αναμενόμενη) τιμή του 1ου επιπέδου είναι ίση με τη σταθερά $\beta_0 - \beta_2 - \beta_3$
- Αν $X=2$ (επίπεδο 2)
 - $\Leftrightarrow D_2 = 1, D_3 = 0 \Leftrightarrow Y = \beta_0 + \beta_2 + \varepsilon$
 - \Leftrightarrow η μέση (αναμενόμενη) τιμή του 2ου επιπέδου είναι ίση με $\beta_0 + \beta_2$
- Αν $X=3$ (επίπεδο 3)
 - $\Leftrightarrow D_2 = 0, D_3 = 1 \Leftrightarrow Y = \beta_0 + \beta_3 + \varepsilon$
 - \Leftrightarrow η μέση (αναμενόμενη) τιμή του 3ου επιπέδου είναι ίση με $\beta_0 + \beta_3$
- Συνεπώς β_2 είναι η διαφορά του 2ου επιπέδου από το συνολικό μέσο (β_0)
- Συνεπώς β_3 είναι η διαφορά του 3ου επιπέδου από το συνολικό μέσο (β_0)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-136

11.9. Σχέση παλινδρόμησης και ANOVA/t-tests Παράδειγμα [05_dataset03]

Περιορισμοί μηδενικού αθροίσματος – Sum-to-zero Constraints

- ΓΕΝΙΚΟΙ ΚΑΝΟΝΕΣ
- Για $1 < i < k$
- Αν $X=i$ (επίπεδο i)
 - $\Leftrightarrow D_i = 1, D_k = 0$ για $k \neq i \Leftrightarrow Y = \beta_0 + \beta_i + \varepsilon$
 - \Leftrightarrow η μέση (αναμενόμενη) τιμή του $<i>$ επιπέδου είναι ίση με $\beta_0 + \beta_i$
- Συνεπώς β_0 είναι ο συνολικός μέσος (ο μέσος των μέσων)
- Συνεπώς β_i είναι η διαφορά του $<i>$ επιπέδου από το συνολικό μέσο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-137

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

- Γυρίζουμε στο παράδειγμα 8-1
- Έχουμε και την περιοχή όπου (πιθανώς) να επηρεάζει την τιμή των σπιτιών
- Έχουμε 3 περιοχές άρα θα χρησιμοποιήσουμε 2 dummies με περιοχή αναφοράς την 1η (area2, area3)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-138

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

Συνεπώς το μοντέλο μας θα γίνει

$$\text{Price}_i = \beta_0 + \beta_1 \text{Living}_i + \beta_2 \text{area2}_i + \beta_3 \text{area3}_i + \varepsilon$$

(χρησιμοποιούμε μόνο την έκταση της οικίας για να είναι πιο εύκολο το παράδειγμα)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-139

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

ΔΙΑΔΙΚΑΣΙΑ

- Φτιάχνουμε τις dummies (area2, area3)
- Προσαρμόζουμε το παλινδρομικό μοντέλο με Y την τιμή (price) και X το μέγεθος του οικήματος (living) και τις dummies των περιοχών (area2, area3)
- Επιλέγουμε το μοντέλο μας
- Ερμηνεύουμε τις παραμέτρους

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-140

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.972 ^a	.945	.939	8194.834

a. Predictors: (Constant), area3, area2, living
Size of Living Area in Sq. ft

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	41121.1	11750.6		3.50	.002
	living Size of Living Area in Sq. ft	43.6	8.2	.67	5.32	.000
	area2	4015.0	5081.7	.06	.79	.437
	area3	24283.1	9950.4	.35	2.44	.022

a. Dependent Variable: price
Price in US Dollars

$$\text{Price}_i = 41121 + 43.6 \text{Living}_i + 40015 \text{area2}_i + 24283 \text{area3}_i + \varepsilon$$

Διαφάνεια 12-141

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

$$\text{Price}_i = 41121 + 43.6 \text{Living}_i + 4015 \text{area2}_i + 24283 \text{area3}_i + \varepsilon$$

ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ

- Subdivision=A ⇔ area2=area3=0 ⇔

$$\text{Price}_i = \beta_0 + \beta_1 \text{Living}_i \quad [\text{εδώ } \text{Price}_i = 41121 + 43.6 \text{Living}_i] \Leftrightarrow$$

Συνεπώς

- ✓ β_0 : σταθερά του παλινδρομικού μοντέλου για την περιοχή A
- ✓ β_0 : Αναμενόμενη τιμή της περιοχής A όταν το οικήμα έχει μηδενική έκταση (???) [εδώ 41121\$]
- ✓ β_1 : κλίση του παλινδρομικού μοντέλου για την περιοχή A
- ✓ β_1 : Αναμενόμενη αύξηση τιμής στην περιοχή A ανά τετραγωνικό έκταση κτιρίου [εδώ 43.6\$]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-142

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

$$\text{Price}_i = 41121 + 43.6 \text{ Living}_i + 4015 \text{ area2}_i + 24283 \text{ area3}_i + \epsilon$$

ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ

- Subdivision=B ⇔ area2=1, area3=0 ⇔

$$\text{Price}_i = (\beta_0 + \beta_2) + \beta_1 \text{Living}_i \quad [\text{εδώ } \text{Price}_i = 45136 + 43.6 \text{ Living}_i] \Leftrightarrow$$

Συνεπώς

- ✓ $(\beta_0 + \beta_2)$: σταθερά του παλινδρομικού μοντέλου για την περιοχή B
- ✓ $(\beta_0 + \beta_2)$: Αναμενόμενη τιμή της περιοχής B όταν το οίκημα έχει μηδενική έκταση (???) [εδώ 45136\$]
- ✓ β_1 : κλίση του παλινδρομικού μοντέλου για την περιοχή B (ίδια με A)
- ✓ β_1 : Αναμενόμενη αύξηση τιμής στην περιοχή B ανά τετραγωνικό έκταση κτιρίου [εδώ 43.6\$]
- ✓ β_2 : Διαφορά τιμής μεταξύ 2 σπιτιών περιοχών A και B ίδιου μεγέθους (στη B είναι 4015\$ πιο ακριβό)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-143

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

$$\text{Price}_i = 41121 + 43.6 \text{ Living}_i + 4015 \text{ area2}_i + 24283 \text{ area3}_i + \epsilon$$

ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ

- Subdivision=C ⇔ area2=0, area3=1 ⇔

$$\text{Price}_i = (\beta_0 + \beta_3) + \beta_1 \text{Living}_i \quad [\text{εδώ } \text{Price}_i = 65040 + 43.6 \text{ Living}_i] \Leftrightarrow$$

Συνεπώς

- ✓ $(\beta_0 + \beta_3)$: σταθερά του παλινδρομικού μοντέλου για την περιοχή C
- ✓ $(\beta_0 + \beta_3)$: Αναμενόμενη τιμή της περιοχής C όταν το οίκημα έχει μηδενική έκταση (???) [εδώ 65040\$]
- ✓ β_1 : κλίση του παλινδρομικού μοντέλου για την περιοχή C (ίδια με A,B)
- ✓ β_1 : Αναμενόμενη αύξηση τιμής στην περιοχή C ανά τετραγωνικό έκταση κτιρίου [εδώ 43.6\$]
- ✓ β_3 : Διαφορά τιμής μεταξύ 2 σπιτιών περιοχών A και C ίδιου μεγέθους (στη C είναι 24283\$ πιο ακριβό)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-144

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

$$\text{Price}_i = 41121 + 43.6 \text{ Living}_i + 4015 \text{ area2}_i + 24283 \text{ area3}_i + \epsilon$$

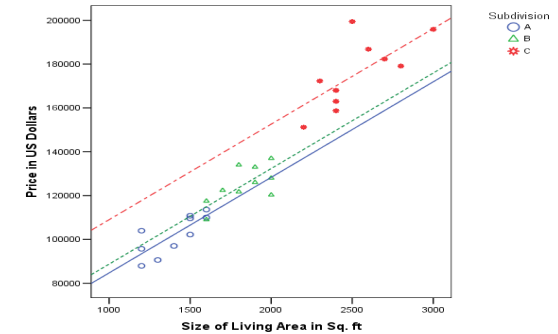
ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ (συνοπτική)

- ✓ β_0 : το πάγιο κόστος μίας κατοικίας στην περιοχή A είναι 41121\$
- ✓ β_1 : το κόστος κάθε τετραγωνικού είναι 43.6 για όλες τις περιοχές
- ✓ β_2 : ένα σπίτι στην περιοχή B έχει 4015\$ επιπλέον πάγιο κόστος σε σχέση με ένα σπίτι της περιοχής A
- ✓ β_3 : ένα σπίτι στην περιοχή C έχει 24283\$ επιπλέον πάγιο κόστος σε σχέση με ένα σπίτι της περιοχής A
- ✓ [ΠΡΟΣΟΧΗ]: Το β_1 είναι κοινό [ΠΑΡΑΛΛΗΛΕΣ ΓΡΑΜΜΕΣ]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-145

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA) Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 12-146

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)

Μοντέλο παράλληλων γραμμών (Παράδειγμα 11-1)

$$\text{Price}_i = 41121 + 43.6 \text{ Living}_i + 4015 \text{ area2}_i + 24283 \text{ area3}_i + \epsilon$$

ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

- ✓ $\beta_2 \Leftrightarrow p=0.437 > 0.05 \Leftrightarrow$ δεν υπάρχει σημαντική διαφοροποίηση του πάγιου κόστους μεταξύ των περιοχών A & B \Leftrightarrow μπορούμε να αφαιρέσουμε την dummy area2 και να έχουμε ένα μοντέλο για τις 2 περιοχές
- ✓ $\beta_3 \Leftrightarrow p=0.022 < 0.05 \Leftrightarrow$ υπάρχει σημαντική διαφοροποίηση του πάγιου κόστους μεταξύ των περιοχών A & C
- ✓ β_0 & $\beta_1 \Leftrightarrow$ στατιστικά σημαντικά
- ✓ ΠΡΟΣΟΧΗ: συγκρίσεις (+έλεγχοι υποθέσεων) γίνονται μόνο σε σχέση με το επίπεδο αναφοράς. Αν θέλαμε να συγκρίνουμε το πάγιο κόστος των περιοχών B+C θα έπρεπε να αλλάξουμε το επίπεδο αναφοράς.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-147

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)

Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

Είναι πιο ρεαλιστικό μοντέλο να υποθέσουμε διαφορετική τιμή ανά τετραγωνικό

\Leftrightarrow διαφορετικές κλίσεις

\Leftrightarrow διαφορετικά παλινδρομικά μοντέλα

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-148

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)

Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

Είναι πιο ρεαλιστικό μοντέλο να υποθέσουμε διαφορετική τιμή ανά τετραγωνικό

\Leftrightarrow διαφορετικές κλίσεις

\Leftrightarrow διαφορετικά παλινδρομικά μοντέλα

ή ένα ενιαίο (συνδιακύμανσης)

$$\text{Price}_i = \beta_0 + \beta_1 \text{ Living}_i + \beta_2 \text{ area2}_i + \beta_3 \text{ area3}_i + \beta_4 \text{ Living}_i \times \text{area2}_i + \beta_5 \text{ Living}_i \times \text{area3}_i + \epsilon$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-149

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)

Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

Οι πολλαπλασιαστικοί όροι $\text{Living}_i \times \text{area2}_i$ & $\text{Living}_i \times \text{area3}_i$

- ✓ είναι οι παράμετροι που μετρούν τις αλληλεπιδράσεις της ποσοτικής living και της κατηγορικής subdivision στην μεταβλητή απόκρισης Price
- ✓ εισάγουν διαφορετικές κλίσεις στο μοντέλο
- ✓ μας βοηθούν να ελέγξουμε την ισότητα των κλίσεων (και για αυτό δεν κάνουμε 3 διαφορετικά μοντέλα)

ΣΤΗΝ ΠΡΑΞΗ:

- ✓ όταν έχουμε πολλές επεξηγηματικές μεταβλητές (ποσοτικές + κατηγορικές) δεν είναι δυνατόν λάβουμε όλες τις αλληλεπιδράσεις
- ✓ Παίρνουμε μόνο αυτές που έχουν λογική ερμηνεία ή πρέπει να συμπεριληφθούν σύμφωνα με κάποια επιστημονική θεωρία ή σενάριο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-150

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

ΔΙΑΔΙΚΑΣΙΑ

- Φτιάχνουμε τις dummies (area2, area3)
- Φτιάχνουμε τις αλληλεπιδράσεις livar2, livar3
- Προσαρμόζουμε το παλινδρομικό μοντέλο με Y την τιμή (price) και X το μέγεθος του οικήματος (living), τις dummies των περιοχών (area2, area3) και τις αλληλεπιδράσεις livar2, livar3
- Επιλέγουμε το μοντέλο μας
- Ερμηνεύουμε τις παραμέτρους

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-151

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΑΡΑΔΕΙΓΜΑΤΟΣ

$$Price_i = 41410 + 43.4 Living_i + 17625 area2_i + 17913 area3_i - 7.4 Living_i \times area2_i + 3.4 Living_i \times area3_i + \epsilon$$

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	41410.0	24395.0		1.697	.103
	living Size of Living Area in Sq. ft	43.4	17.3	.664	2.504	.019
	area2	17624.8	41148.6	.254	.428	.672
	area3	15913.2	38132.3	-.230	-.417	.680
	livar2	-7.4	25.0	-.196	-.295	.770
	livar3	3.4	20.8	.125	.163	.872

^a Dependent Variable: price Price in US Dollars

2-152

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

Price_i = 41410 + 43.4 Living_i + 17625 area2_i + 17913 area3_i - 7.4 Living_i × area2_i + 3.4 Living_i × area3_i + ε

ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ

- Subdivision=A ⇔ area2=area3=0 ⇔
Price_i = β₀ + β₁ Living_i [*εδώ Price_i = 41410 + 43.4 Living_i*] ⇔
Συνεπώς
 - ✓ β₀ : σταθερά του παλινδρομικού μοντέλου για την περιοχή A
 - ✓ β₀ : Αναμενόμενη τιμή της περιοχής A όταν το οίκημα έχει μηδενική έκταση (???) [*εδώ 41410\$*]
 - ✓ β₁ : κλίση του παλινδρομικού μοντέλου για την περιοχή A
 - ✓ β₁ : Αναμενόμενη αύξηση τιμής στην περιοχή A ανά τετραγωνικό έκτασης κτιρίου [*εδώ 43.4\$*]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-153

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

Price_i = 41410 + 43.4 Living_i + 17625 area2_i + 17913 area3_i - 7.4 Living_i × area2_i + 3.4 Living_i × area3_i + ε

ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ

- Subdivision=B ⇔ area2=1, area3=0 ⇔
Price_i = (β₀+β₂) + (β₁+β₄)Living_i [*Price_i = 59035 + 36 Living_i*] ⇔
 - ✓ β₀+β₂ : σταθερά του παλινδρομικού μοντέλου για την περιοχή B
 - ✓ β₀+β₂ : Αναμενόμενη τιμή της περιοχής B όταν το οίκημα έχει μηδενική έκταση (???) [*εδώ 59035\$*]
 - ✓ β₂ : Διαφορά των αναμενόμενων τιμών των περιοχών A & B όταν το οίκημα έχει μηδενική έκταση (???) [*εδώ στην περιοχή B έχουμε πάγιο κόστος μεγαλύτερο κατά 17625\$ από ότι στην περιοχή A*].
 - ✓ β₁ + β₄ : κλίση του παλινδρομικού μοντέλου για την περιοχή B
 - ✓ β₁ + β₄ : Αναμενόμενη αύξηση τιμής στην περιοχή B ανά τετραγωνικό έκτασης κτιρίου [*εδώ 36\$*]
 - ✓ β₄ : Διαφορά τιμής ανά τετραγωνικό για την περιοχή B σε σχέση με την περιοχή A [*εδώ στη περιοχή B το τετραγωνικό κοστίζει 7.4 λιγότερα \$ από ότι στην A*]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-154

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

$$\text{Price}_i = 41410 + 43.4 \text{ Living}_i + 17625 \text{ area2}_i + 17913 \text{ area3}_i - 7.4 \text{ Living}_i \times \text{area2}_i + 3.4 \text{ Living}_i \times \text{area3}_i + \varepsilon$$

ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ

- Subdivision=C \Leftrightarrow area2=0, area3=1 \Leftrightarrow
 $\text{Price}_i = (\beta_0 + \beta_3) + (\beta_1 + \beta_5) \text{Living}_i$ [$\text{Price}_i = 59323 + 46.8 \text{ Living}_i$] \Leftrightarrow
 - $\beta_0 + \beta_3$: σταθερά του παλινδρομικού μοντέλου για την περιοχή C
 - $\beta_0 + \beta_3$: Αναμενόμενη τιμή της περιοχής C όταν το οίκημα έχει μηδενική έκταση (???) [εδώ 59035\$]
 - β_3 : Διαφορά των αναμενόμενων τιμών των περιοχών A & C όταν το οίκημα έχει μηδενική έκταση (???) [εδώ στην περιοχή C έχουμε πάγιο κόστος μεγαλύτερο κατά 17913\$ από ότι στην περιοχή A].
 - $\beta_1 + \beta_5$: κλίση του παλινδρομικού μοντέλου για την περιοχή C
 - $\beta_1 + \beta_5$: Αναμενόμενη αύξηση τιμής στην περιοχή C ανά τετραγωνικό έκτασης κτιρίου [εδώ 46.8\$]
 - β_5 : Διαφορά τιμής ανά τετραγωνικό για την περιοχή C σε σχέση με την περιοχή A [εδώ στη περιοχή C το τετραγωνικό κοστίζει 3.4 περισσότερα \$ από ότι στην A]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-155

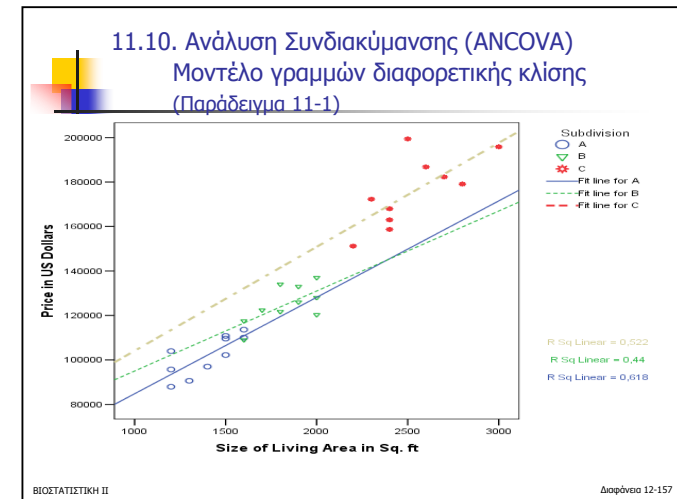
11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

$$\text{Price}_i = 41410 + 43.4 \text{ Living}_i + 17625 \text{ area2}_i + 17913 \text{ area3}_i - 7.4 \text{ Living}_i \times \text{area2}_i + 3.4 \text{ Living}_i \times \text{area3}_i + \varepsilon$$

ΕΡΜΗΝΕΙΑ ΠΑΡΑΜΕΤΡΩΝ (συνοπτική)

- β_0 : το πάγιο κόστος μίας κατοικίας στην περιοχή A είναι 41121\$
- β_1 : το κόστος κάθε τετραγωνικού είναι 43.6 για όλες τις περιοχές
- β_2 : το επιπλέον πάγιο κόστος για την περιοχή B είναι 17625\$ μεγαλύτερο σε σχέση με την περιοχή A
- β_3 : το επιπλέον πάγιο κόστος για την περιοχή C είναι 17913\$ μεγαλύτερο σε σχέση με την περιοχή A
- β_4 : ένα σπίτι στην περιοχή B έχει 7.4\$ μικρότερο κόστος ανά τετραγωνικό σε σχέση με ένα σπίτι της περιοχής A
- β_5 : ένα σπίτι στην περιοχή C έχει 3.4\$ μεγαλύτερο κόστος ανά τετραγωνικό σε σχέση με ένα σπίτι της περιοχής A

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-156



11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης
(Παράδειγμα 11-1)

$$\text{Price}_i = 41410 + 43.4 \text{ Living}_i + 17625 \text{ area2}_i + 17913 \text{ area3}_i - 7.4 \text{ Living}_i \times \text{area2}_i + 3.4 \text{ Living}_i \times \text{area3}_i + \varepsilon$$

ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

- β_4 & β_5 \Leftrightarrow δεν υπάρχει διαφορά στο κόστος ανά τετραγωνικό για τις 3 περιοχές
- β_2 & β_3 \Leftrightarrow δεν υπάρχει διαφορά στο πάγιο κόστος για τις 3 περιοχές
- β_1 \Leftrightarrow το πάγιο κόστος της περιοχής A μπορεί να θεωρηθεί μηδέν
- β_2 \Leftrightarrow η τιμή αλλάζει ανάλογα με την έκταση του οικήματος.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 12-158

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
 Μοντέλο γραμμών διαφορετικής κλίσης και κοινής σταθεράς (Παράδειγμα 11-1)

Είναι πιο ρεαλιστικό μοντέλο να υποθέσουμε διαφορετική τιμή ανά τετραγωνικό και ίδιο πάγιο κόστος (αν αυτό έχει νόημα)

⇔ διαφορετικά παλινδρομικά μοντέλα με ίδιες σταθερές

$$\Leftrightarrow \text{Price}_i = \beta_0 + \beta_1 \text{Living}_i + \beta_4 \text{Living}_i \times \text{area2}_i + \beta_4 \text{Living}_i \times \text{area3}_i + \varepsilon$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-159

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
 Μοντέλο γραμμών διαφορετικής κλίσης και κοινής σταθεράς (Παράδειγμα 11-1)

$$\text{Price}_i = 50618 + 36.9 \text{Living}_i + 3.7 \text{Living}_i \times \text{area2}_i + 12.5 \text{Living}_i \times \text{area3}_i + \varepsilon$$

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	50617.6	15762.9		3.211	.004
	living Size of Living Area in Sq. ft	36.9	11.3	.565	3.271	.003
	livar2	3.7	3.5	.098	1.063	.297
	livar3	12.5	5.4	.460	2.331	.028

a. Dependent Variable: price Price in US Dollars

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-160

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
 Μοντέλο γραμμών διαφορετικής κλίσης και κοινής σταθεράς (Παράδειγμα 11-1)

$$\text{Price}_i = 50618 + 36.9 \text{Living}_i + 3.7 \text{Living}_i \times \text{area2}_i + 12.5 \text{Living}_i \times \text{area3}_i + \varepsilon$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-161

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
 Μοντέλο γραμμών διαφορετικής κλίσης και μηδενικής σταθεράς (Παράδειγμα 11-1)

Είναι πιο ρεαλιστικό μοντέλο να υποθέσουμε διαφορετική τιμή ανά τετραγωνικό και μηδενικό πάγιο κόστος

⇔ διαφορετικά παλινδρομικά μοντέλα με ίδιες σταθερές

$$\Leftrightarrow \text{Price}_i = \beta_1 \text{Living}_i + \beta_4 \text{Living}_i \times \text{area2}_i + \beta_4 \text{Living}_i \times \text{area3}_i + \varepsilon$$

[SPSS: Analyze>Regression>Linear
 OPTIONS
 Include constant in equation]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-162

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Μοντέλο γραμμών διαφορετικής κλίσης και μηδενικής σταθεράς (Παράδειγμα 11-1)

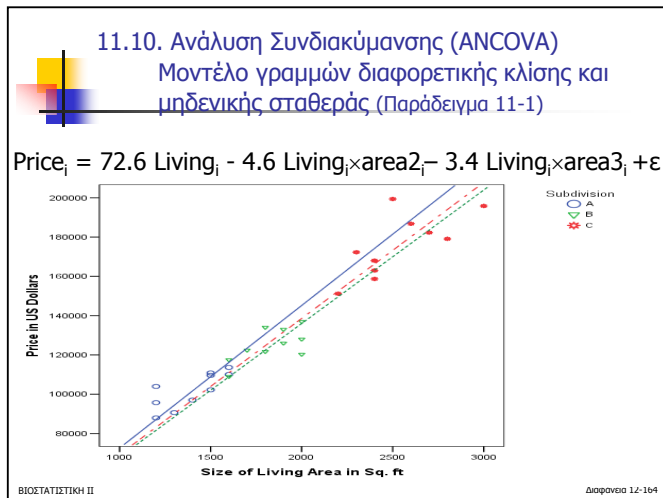
$$\text{Price}_i = 72.6 \text{ Living}_i - 4.6 \text{ Living}_i \times \text{area2}_i - 3.4 \text{ Living}_i \times \text{area3}_i + \varepsilon$$

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	living Size of Living Area in Sq. ft	72.6	2.1	1.043	34.021	.000
	livar2	-4.6	2.7	-.035	-1.696	.101
	livar3	-3.4	2.4	-.036	-1.374	.181

a. Dependent Variable: price Price in US Dollars
b. Linear Regression through the Origin

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-163



11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Κλιμακωτές Διαδικασίες Επιλογής Μεταβλητών (Παράδειγμα 11-1)

Εδώ η σταθερά φαίνεται να βελτιώνει την προσαρμογή του μοντέλου οπότε μάλλον πρέπει να την αφήσουμε στο μοντέλο

Κάνοντας backward (βάζοντας και τη σταθερά μέσα ως μεταβλητή) καταλήγουμε στο ακόλουθο μοντέλο

$$\text{Price}_i = 35892 + 48 \text{ Living}_i + 17625 \text{ area2}_i + 18176 \text{ area3}_i + \varepsilon$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-165

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)
Κλιμακωτές Διαδικασίες Επιλογής Μεταβλητών (Παράδειγμα 11-1)

Variables Entered/Removed^{b,c}

Model	Variables Entered	Variables Removed	Method
1	livar3, livar2, constant, area3, area2, living Size of Living Area in Sq. ft		Enter
2		livar3	Backward (criterion: Probability of F-to-remove >= .100).
3		livar2	Backward (criterion: Probability of F-to-remove >= .100).
4		area2	Backward (criterion: Probability of F-to-remove >= .100).

Model Summary^{f,g}

Model	R	R Square ^a	Adjusted R Square	Std. Error of the Estimate
1	.998 ^b	.997	.996	8484.693
2	.998 ^c	.997	.996	8317.891
3	.998 ^d	.997	.996	8194.834
4	.998 ^e	.997	.997	8137.613

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 12-166

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)

Κλιμακωτές Διαδικασίες Επιλογής Μεταβλητών
(Παράδειγμα 11-1)

Κάνοντας backward (αφαιρώντας τις area2 & area 3 ⇔ προκαθορίζουμε ίσες σταθερές) καταλήγουμε στο ακόλουθο μοντέλο

$$\text{Price}_i = 38181 + 47 \text{ Living}_i + 7.6 \text{ Living}_i \times \text{area3}_i + \varepsilon$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-167

11.10. Ανάλυση Συνδιακύμανσης (ANCOVA)

Κλιμακωτές Διαδικασίες Επιλογής Μεταβλητών
(Παράδειγμα 11-1)

Variables Entered/Removed^{b,c}

Model	Variables Entered	Variables Removed	Method
1	livar3, livar2, constant, living Size of Living Area in Sq. ft	.	Enter
2		livar2	Backward (criterion: Probability of F-to-remove >= .100).

Model Summary^{d,e}

Model	R	R Square ^a	Adjusted R Square	Std. Error of the Estimate
1	.998 ^b	.997	.996	8195.753
2	.998 ^c	.997	.996	8215.610

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

Διαφάνεια 12-168