

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

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ

ΜΑΘΗΜΑ 10

Συμπερασματολογία για 2 ποσοτικές μεταβλητές (ΑΠΛΗ ΠΑΛΙΝΔΡΟΜΗΣΗ)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-1

10. Δύο ποσοτικές μεταβλητές ΠΕΡΙΕΧΟΜΕΝΑ

- Εισαγωγή
 - ✓ Το μοντέλο της γραμμικής παλινδρόμησης
 - ✓ Δείκτης γραμμική συσχέτισης του Pearson
- Απλή Γραμμική Παλινδρόμηση
 - ✓ Το μοντέλο
 - ✓ Έλεγχοι υποθέσεων & Ερμηνεία παραμέτρων
 - ✓ Εφαρμογή στο SPSS (Παράδειγμα 10-1)
 - ✓ Έλεγχος Προϋποθέσεων
 - ✓ Διαγνωστικά διαγράμματα καταλοίπων
 - ✓ Μετασχηματισμοί της μεταβλητής απόκρισης
- Σύγκριση με το paired t-test

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-2

10.1. Εισαγωγή

10.1.1. Το μοντέλο της γραμμικής παλινδρόμησης

Έστω ότι έχουμε 2 ποσοτικές μεταβλητές

- X: επεξηγηματική ή ανεξάρτητη
- Y: απόκρισης ή εξαρτημένη

Είναι λογικό ότι αν πιστεύουμε ότι η X επηρεάζει (επιδρά) με κάποιο τρόπο την Y τότε θα υπάρχει μια συνάρτηση $h(x)$ τέτοια ώστε:

$$y = h(x)$$

και επειδή μιλάμε για τυχαία φαινόμενα προσθέτουμε και ένα τυχαίο όρο (σφάλμα)

$$y = h(x) + \varepsilon$$

$\varepsilon \sim \text{Distribution}(\theta)$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-3

10.1. Εισαγωγή

10.1.1. Το μοντέλο της γραμμικής παλινδρόμησης

Έστω ότι έχουμε 2 ποσοτικές μεταβλητές

- X: επεξηγηματική ή ανεξάρτητη
- Y: απόκρισης ή εξαρτημένη

Η πιο απλή μορφή αυτής της λογικής είναι να θέσουμε τη συνάρτηση ίση με τη γραμμική συνάρτηση δηλ.

$$h(x) = \beta_0 + \beta_1 x$$

και να χρησιμοποιήσουμε την «αγαπημένη» μας κανονική κατανομή ως τυχαίο σφάλμα

$$y = \beta_0 + \beta_1 x + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-4

10.1. Εισαγωγή
10.1.1. Το μοντέλο της γραμμικής παλινδρόμησης

Πιο σωστή προσέγγιση (GLM)

- X: επεξηγηματική ή ανεξάρτητη (γνωστή/σταθερή)
- Y: απόκριση ή εξαρτημένη (τυχαία)
- $Y \sim \text{Distribution}(\theta)$
- $g(\theta) = h(x)$
 - ✓ h(x): σταθερή (μη τυχαία) συνιστώσα
 - ✓ g(θ): συνδετική συνάρτηση (link function) μεταξύ τυχαίας (στοχαστικής) και σταθερής (ντετερμινιστικής) συνιστώσας του μοντέλου
 - ✓ Συνήθως h(x) ⇔ γραμμική συνάρτηση ως προς X ⇔ γραμμικός συνδυασμός/προσδιορισμός (linear predictor)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-3

10.1. Εισαγωγή
10.1.1. Το μοντέλο της γραμμικής παλινδρόμησης

Πιο σωστή προσέγγιση (GLM type)

- X: επεξηγηματική ή ανεξάρτητη (γνωστή/σταθερή)
- Y: απόκρισης ή εξαρτημένη (τυχαία)
- $Y \sim \text{Normal}(\mu, \sigma^2)$ $[\theta = (\mu, \sigma^2)^T]$
- $\mu = \beta_0 + \beta_1 x$ $[g(\theta) = \mu]$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-6

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

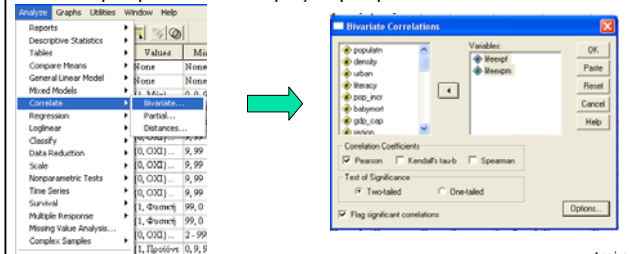
- Ο δείκτης γραμμικής συσχέτισης του Pearson $\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}$
 - Παίρνει τιμές από -1 έως 1
 - ✓ 1 = πλήρη θετική γραμμική σχέση
 - ✓ -1 = πλήρη αρνητική γραμμική σχέση
 - ✓ 0 = οι δύο μεταβλητές είναι ασυσχέτιστες / ανεξάρτητες
- Είναι ελεύθερος μονάδων $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{S_{xy}}{S_x S_y}$
- Ποσοτικοποιεί το βαθμό της γραμμικής εξάρτησης μεταξύ των δύο μεταβλητών
- Δεν ξεχωρίζει ποια είναι X και ποια Y

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-7

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

Παράδειγμα 10-1 [world95]

- Να εξετασθεί η γραμμική σχέση μεταξύ
 - Αναμενόμενου χρόνου ζωής ανδρών και γυναικών Lifexpf – lifexhpr
 - Πληθυσμού και πυκνότητας πληθυσμού.



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-8

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

Παράδειγμα 10-1 [world95]

- Na εξετασθεί η γραμμική σχέση μεταξύ
 - Αναμενόμενου χρόνου ζωής ανδρών και γυναικών Lifexpf – lifexpm

| | | lifexpf Average female life expectancy | lifexpm Average male life expectancy |
|---|---|---|---|
| lifexpf Average female life expectancy | Pearson Correlation Sig. (2-tailed) N | 1 109 | .982* .000 109 |
| lifexpm Average male life expectancy | Pearson Correlation Sig. (2-tailed) N | .982* .000 109 | 1 109 |

** .Correlation is significant at the 0.01 level (2-tailed).

Ο πίνακας είναι συμμετρικός
Η διαγώνιος είναι 1 εφόσον είναι οι συσχετίσεις της κάθε μεταβλητής με τον εαυτό της

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-9

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

Παράδειγμα 10-1 [world95]

- Na εξετασθεί η γραμμική σχέση μεταξύ
 - Αναμενόμενου χρόνου ζωής ανδρών και γυναικών Lifexpf – lifexpm

| | | lifexpf Average female life expectancy | lifexpm Average male life expectancy |
|---|---|---|---|
| lifexpf Average female life expectancy | Pearson Correlation Sig. (2-tailed) N | 1 109 | .982* .000 109 |
| lifexpm Average male life expectancy | Pearson Correlation Sig. (2-tailed) N | .982* .000 109 | 1 109 |

** .Correlation is significant at the 0.01 level (2-tailed)

Μέγεθος δείγματος
Av p-value<0.01 τότε **
Av 0.01<p-value<0.05 τότε *

Ο δείκτης γρ.συσχέτισης του Pearson
p-value για τον έλεγχο H₀: ρ = 0 έναντι της H₁: ρ ≠ 0

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-10

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

Παράδειγμα 10-1 [world95]

- Na εξετασθεί η γραμμική σχέση μεταξύ
 - Πληθυσμού και πυκνότητας πληθυσμού

| | | populattn Population in thousands | density Number of people / sq. kilometer |
|---|---|---|---|
| populattn Population in thousands | Pearson Correlation Sig. (2-tailed) N | 1 109 | -.018 .850 109 |
| density Number of people / sq. kilometer | Pearson Correlation Sig. (2-tailed) N | -.018 109 | 1 109 |

Είναι λογικό αυτό;

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-11

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

Παράδειγμα 10-1 [world95]

- Na εξετασθεί η γραμμική σχέση μεταξύ
 - Πληθυσμού και πυκνότητας πληθυσμού

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-12

10.1. Εισαγωγή

10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

Επιπλέον σχόλια

- Ο δείκτης προϋποθέτει Χ,Υ να είναι τυχαίες μεταβλητές
- Σαν δείκτης μπορεί να χρησιμοποιηθεί ανεξάρτητα κανονικότητας
- Ο έλεγχος προϋποθέτει κανονικότητα ή μεγάλο δείγμα
- Εναλλακτικά μπορούμε να χρησιμοποιήσουμε τους μη παραμετρικούς δείκτες συσχέτισης
- Αν η σχέση είναι ισχυρή αλλά μη γραμμική ο δείκτης θα μας δείξει πόσο καλά προσεγγίζεται από τη γραμμική σχέση
- Σύμφωνα με τους Chatfield & Collins (1980, σελ. 40-41)
 - Ο έλεγχος είναι σχετικά συντηρητικός δηλ. μικρά r θα μας δώσουν εξάρτηση (κάποιοι είδους) ειδικά για μεγάλα δείγματα
 - Εμπειρικός κανόνας: ισχυρή γρ. σχέση για $r > 0.70$
 - Ο ρ δεν εκτιμάται αξιόπιστα για μικρά δείγματα ($n < 12$)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-13

10.1. Εισαγωγή

10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ (ΣΥΜΜΑΖΕΜΑ)

- Μόνο δείκτες συσχέτισης ή p-value
- 1 δεκαδικό για τα r
- Μονάδες στη διαγώνιο
- Ομαδοποίηση παρόμοιων μεταβλητών
- χρησιμοποιήστε σύμβολα (αστεράκια) ή χρώματα να τονίσετε σημαντικές σχέσεις
- Αν παρόλα αυτά δε βγαίνει νόημα
 - σβήστε τα νούμερα (κρατήστε μόνο σύμβολα)
 - Χρησιμοποιήστε path diagrams

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-15

10.1. Εισαγωγή

10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ

| | | log_den | log_pop | pop_dens | den_dens | urban | life_exp | life_exp_f | life_exp_m | life_exp_t | pop_inc |
|------------|---|---------|---------|----------|----------|-------|----------|------------|------------|------------|---------|
| log_den | 1 | | | | | | | | | | |
| log_pop | | 1 | | | | | | | | | |
| pop_dens | | | 1 | | | | | | | | |
| den_dens | | | | 1 | | | | | | | |
| urban | | | | | 1 | | | | | | |
| life_exp | | | | | | 1 | | | | | |
| life_exp_f | | | | | | | 1 | | | | |
| life_exp_m | | | | | | | | 1 | | | |
| life_exp_t | | | | | | | | | 1 | | |
| pop_inc | | | | | | | | | | 1 | |

*** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Διαφάνεια 10-14

10.1. Εισαγωγή

10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ

| | | log_den | log_pop | pop_dens | den_dens | urban | life_exp | life_exp_f | life_exp_m | life_exp_t | pop_inc |
|------------|---|---------|---------|----------|----------|-------|----------|------------|------------|------------|---------|
| log_den | 1 | | | | | | | | | | |
| log_pop | | 1 | | | | | | | | | |
| pop_dens | | | 1 | | | | | | | | |
| den_dens | | | | 1 | | | | | | | |
| urban | | | | | 1 | | | | | | |
| life_exp | | | | | | 1 | | | | | |
| life_exp_f | | | | | | | 1 | | | | |
| life_exp_m | | | | | | | | 1 | | | |
| life_exp_t | | | | | | | | | 1 | | |
| pop_inc | | | | | | | | | | 1 | |

*** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-16

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ

Correlations

| Statistics | log_den | log_pop | populati | density | urban | lifeexpf | lifeexpm | literacy | pop_inc |
|---------------------|-----------------------------|-----------------------------|-------------------------|---------------------------------|-----------------------------|--------------------------------|------------------------------|---------------------|----------------------------------|
| | Log (base 10) of Population | Log (base 10) of Population | Population in thousands | Number of people / sq kilometer | People living in cities (%) | Average female life expectancy | Average male life expectancy | People who read (%) | Population increase (% per year) |
| Pearson Correlation | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| log_den | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| log_pop | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| populati | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| density | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| urban | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 |
| lifeexpf | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 |
| lifeexpm | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 |
| literacy | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 |
| pop_inc | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-17

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ

Correlations

| Statistics | log_den | log_pop | populati | density | urban | lifeexpf | lifeexpm | literacy | pop_inc |
|---------------------|-----------------------------|-----------------------------|-------------------------|---------------------------------|-----------------------------|--------------------------------|------------------------------|---------------------|----------------------------------|
| | Log (base 10) of Population | Log (base 10) of Population | Population in thousands | Number of people / sq kilometer | People living in cities (%) | Average female life expectancy | Average male life expectancy | People who read (%) | Population increase (% per year) |
| Pearson Correlation | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| log_den | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| log_pop | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| populati | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| density | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| urban | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 |
| lifeexpf | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 |
| lifeexpm | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 |
| literacy | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 |
| pop_inc | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-18

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ

Correlations

| Statistics | log_den | log_pop | populati | density | urban | lifeexpf | lifeexpm | literacy | pop_inc |
|---------------------|-----------------------------|-----------------------------|-------------------------|---------------------------------|-----------------------------|--------------------------------|------------------------------|---------------------|----------------------------------|
| | Log (base 10) of Population | Log (base 10) of Population | Population in thousands | Number of people / sq kilometer | People living in cities (%) | Average female life expectancy | Average male life expectancy | People who read (%) | Population increase (% per year) |
| Pearson Correlation | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| log_den | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| log_pop | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| populati | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| density | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| urban | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 | 0.6 |
| lifeexpf | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 | 0.6 |
| lifeexpm | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 | 0.6 |
| literacy | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 | 0.6 |
| pop_inc | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 1 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-19

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ/ ΡΑΘΗ ΔΙΑΓΡΑΜ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-20

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ/ ΑΛΛΕΣ ΕΠΙΛΟΓΕΣ ΣΤΟ SPSS

Μη παραμετρικοί δείκτες του Kendall και του Spearman

Έκλεχοι διπλής ή μονής ουράς

Σημείωση σημαντικών συσχετίσεων

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-21

10.1. Εισαγωγή
10.1.2. Δείκτης γραμμικής συσχέτισης του Pearson

ΠΙΝΑΚΕΣ ΣΥΣΧΕΤΙΣΗΣ/ ΑΛΛΕΣ ΕΠΙΛΟΓΕΣ ΣΤΟ SPSS

Μέσοι & τυπικές αποκλίσεις

Συνδιακυμάνσεις

Χειρισμός αγνοούμενων τιμών:
*Αφαίρεση ανά ζεύγος μεταβλητών
*Αφαίρεση όλης της παρατήρησης από όλες τις συσχετίσεις

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-22

10.2. Απλή γραμμική παλινδρόμηση
10.2.1. Το μοντέλο

Το μοντέλο:

- $Y = \beta_0 + \beta_1 x + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$
- ή ισοδύναμα
- $Y \sim N(\mu, \sigma^2)$, $E(Y) = \mu = \beta_0 + \beta_1 x$

Μοντέλο και δεδομένα:

- Y_i, X_i ζεύγη τιμών για $i = 1, 2, \dots, n$
 - $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$
 - $Y_i \sim N(\mu_i, \sigma^2)$, $\mu_i = \beta_0 + \beta_1 x_i$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-23

10.2. Απλή γραμμική παλινδρόμηση
10.2.1. Το μοντέλο

ΕΚΤΙΜΗΣΕΙΣ+ΟΡΟΛΟΓΙΑ

- $\hat{\beta}_0, \hat{\beta}_1$: Δειγματικές Εκτιμήσεις των β_0 και β_1
- \hat{y}_i : Αναμενόμενη/προβλεπόμενη τιμή (σύμφωνα με το μοντέλο) της μεταβλητής απόκρισης για το i άτομο
- e_i : Κατάλοιπο/σφάλμα της παλινδρόμησης (εκτίμηση του ε_i)
- $\hat{\sigma}^2$: Εκτίμηση της διακύμανσης των σφαλμάτων

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-24

10.2. Απλή γραμμική παλινδρόμηση

10.2.1. Το μοντέλο

ΕΚΤΙΜΗΣΕΙΣ+ΟΡΟΛΟΓΙΑ

- $\hat{\sigma}^2$: Εκτίμηση της διακύμανσης των σφαλμάτων

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

- R^2 : Συντελεστής προσδιορισμού (coefficient of determination, Ryan, 1997, σελ. 29)

$$R^2 = 1 - \frac{(n-2)\hat{\sigma}^2}{(n-1)s_Y^2}$$

- ✓ δείκτης καλής προσαρμογής
- ✓ Τιμές από 0 έως 1
- ✓ Ποσοστό διακύμανσης που εξηγείται από το μοντέλο (για μεγάλο n)
- ✓ Στην απλή παλινδρόμηση ίσο με r^2
- R_{adj}^2 : δείκτης καλής προσαρμογής
- ✓ Τιμές από 0 έως 1
- ✓ Ποσοστό διακύμανσης που εξηγείται από το μοντέλο
- ✓ Πιο χρήσιμος στην πολλαπλή παλινδρόμηση

$$R_{adj}^2 = 1 - \frac{\hat{\sigma}^2}{s_Y^2}$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-25

10.2. Απλή γραμμική παλινδρόμηση

10.2.1. Το μοντέλο

ΕΚΤΙΜΗΣΕΙΣ+ΟΡΟΛΟΓΙΑ

- $\hat{\beta}_0, \hat{\beta}_1$: Δειγματικές Εκτιμήσεις των β_0 και β_1

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$= \frac{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} = \frac{s_y}{s_x} r$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-26

10.2. Απλή γραμμική παλινδρόμηση

10.2.1. Το μοντέλο

ΠΡΟΫΠΟΘΕΣΕΙΣ:

- Ανεξαρτησία σφαλμάτων (και Y)
- Κανονικότητα σφαλμάτων (και Y)
- Ομοσκεδαστικότητα σφαλμάτων (και Y)
- Γραμμικότητα μεταξύ X και Y
- Δουλεύουμε με κατάλοιπα e_i

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-27

10.2. Απλή γραμμική παλινδρόμηση

10.2.2. Έλεγχοι υποθέσεων & Ερμηνεία

Μας ενδιαφέρει η σχέση X & Y:

- $H_0: \beta_1=0$ έναντι της εναλλακτικής $H_1: \beta_1 \neq 0$
- ΔΕΥΤΕΡΕΥΩΝ ΕΛΕΓΧΟΣ: $H_0: \beta_0=0$ έναντι της εναλλακτικής $H_1: \beta_0 \neq 0$
- Κατά πόσο μπορούμε να προβλέψουμε την Y με τη χρήση της X; [ΟΤΑΝ ΜΑΣ ΕΝΔΙΑΦΕΡΕΙ ΠΡΟΒΛΕΨΗ]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-28

10.2. Απλή γραμμική παλινδρόμηση

10.2.2. Έλεγχοι υποθέσεων & Ερμηνεία

Μας ενδιαφέρει η σχέση X & Y :

- $H_0: \beta_1=0$ έναντι της εναλλακτικής $H_1: \beta_1 \neq 0$
 - ✓ Ισοδύναμο με τον έλεγχο για συσχέτιση μεταξύ X και Y
 - ✓ Δίνει την κλίση της ευθείας
 - ✓ Μας ενδιαφέρει για την ερμηνεία των αιτιολογικών σχέσεων μεταξύ φαινομένων – μεταβλητών
 - ✓ **ΕΡΜΗΝΕΙΑ:** Εξετάζει πόσο αναμένουμε να αυξηθεί η Y με μία μονάδα αύξησης της X
 - ✓ Η τιμή του β_1 επηρεάζεται από την κλίμακα (μονάδες μέτρησης) των X & Y . Το ρ (και r) και ο αντίστοιχος έλεγχος δεν επηρεάζονται.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-29

10.2. Απλή γραμμική παλινδρόμηση

10.2.2. Έλεγχοι υποθέσεων & Ερμηνεία

Μας ενδιαφέρει η σχέση X & Y :

- ΔΕΥΤΕΡΕΥΩΝ ΕΛΕΓΧΟΣ: $H_0: \beta_0=0$ έναντι της εναλλακτικής $H_1: \beta_0 \neq 0$
 - ✓ Μας δίνει το σημείο που η ευθεία τέμνει τον κάθετο άξονα YY' δηλαδή την τιμή του Y όταν $X=0$
 - ✓ **ΕΡΜΗΝΕΙΑ:** Η αναμενόμενη τιμή του Y όταν $X=0$
 - ✓ Πολλές φορές η τιμή αυτή δεν έχει ερμηνεία (διότι η τιμή $X=0$ δεν παρατηρείται ποτέ στην πράξη)
 - ✓ Άλλες φορές θέτουμε $\beta_0=0$ εκ-των-προτέρων και ανεξαρτήτως ελέγχου λόγω κοινής λογικής
 - ✓ Πολλές φορές «βολεύει» για λόγους ερμηνείας αντί της X να χρησιμοποιήσουμε την $X^* = X - \bar{X}$. Τότε
 - ✓ Το β_1 δεν αλλάζει
 - ✓ Το β_0 είναι ίσο με την αναμενόμενη τιμή του Y όταν X είναι ίσο με το δειγματικό μέσο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-30

10.2. Απλή γραμμική παλινδρόμηση

10.2.2. Έλεγχοι υποθέσεων & Ερμηνεία

Μας ενδιαφέρει η σχέση X & Y :

- **ΟΤΑΝ ΜΑΣ ΕΝΔΙΑΦΕΡΕΙ ΠΡΟΒΛΕΨΗ:** Κατά πόσο μπορούμε να προβλέψουμε την Y με τη χρήση της X ;
 - Μπορούμε να προβλέψουμε την αναμενόμενη τιμή του Y για κάθε τιμή του X
 - Το σ^2 & το R^2 μας δίνουν την ακρίβεια της πρόβλεψης.
 - ✓ $R^2 > 0.7 \Leftrightarrow$ καλές προβλέψεις,
 - ✓ $R^2 > 0.9 \Leftrightarrow$ πολύ καλές προβλέψεις
 - ΠΡΟΣΟΧΗ: οι προβλέψεις είναι σωστές – αποδεκτές και αξιόπιστες μόνο για τις τιμές του X που έχουμε παρατηρήσει
 - ✓ Δεν μπορούμε να προβλέψουμε κάτι που δεν το έχουμε μελετήσει
 - ✓ Πολλές φορές κάνουμε πρόβλεψη για τιμές εκτός του παρατηρούμενου εύρους τιμών της X (extrapolation)
 - ✓ Αυτές οι προβλέψεις χρησιμοποιούνται σαν οδηγός μόνο
 - ✓ Υποθέτουμε ότι η ίδια σχέση υπάρχει και στις άλλες τιμές του X (αυτές που δεν έχουμε παρατηρήσει).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-31

10.2. Απλή γραμμική παλινδρόμηση

10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1 [05_dataset5.dat]

- Ο υπεύθυνος των logistics μιας εταιρείας, ενδιαφέρεται να εκτιμήσει το χρόνο παράδοσης (άρα και το αντίστοιχο κόστος) κάθε φορτίου ανάλογα με την απόσταση.
- Για το λόγο αυτό πήρε ένα τυχαίο δείγμα 10 φορτωτικών και κατέγραψε την απόσταση σε μίλια και τις ημέρες παράδοσης.
- Να κατασκευαστεί ένα μοντέλο που θα βοηθήσει τον υπεύθυνο της εταιρείας

| Φορτωτική | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------------------------|-----|-----|------|-----|-----|-----|------|-----|-----|------|
| Απόσταση σε μίλια | 825 | 215 | 1070 | 550 | 480 | 920 | 1350 | 325 | 670 | 1215 |
| Χρόνος παράδοσης σε ημέρες | 3.5 | 1.0 | 4.0 | 2.0 | 1.0 | 3.0 | 4.5 | 1.5 | 3.0 | 5.0 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-32

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1

- Μονάδα μελέτης: φορτωτική
- Μέγεθος δείγματος: $n=2$ φορτωτικές
- Χαρακτηριστικά $p=3$
 - ✓ Αριθμός φορτωτικής
 - ✓ Απόσταση
 - ✓ Χρόνος παράδοσης
- Ποια είναι X & ποια Y;

| code | distance | delivery | Var |
|------|----------|----------|-----|
| 2 | 215 | 1.0 | |
| 3 | 1070 | 4.0 | |
| 4 | 550 | 2.0 | |
| 5 | 480 | 1.0 | |
| 6 | 920 | 3.0 | |
| 7 | 1350 | 4.5 | |
| 8 | 325 | 1.5 | |
| 9 | 670 | 3.0 | |
| 10 | 1215 | 5.0 | |
| 11 | | | |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-33

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Διαγραμματική απεικόνιση
SCATTERPLOT

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-35

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1
ΑΝΑΛΥΣΗ – ΒΗΜΑΤΑ

- Ανάλυση ανά μία μεταβλητή
- Διαγραμματική απεικόνιση (Scatter-plot)
- Δείκτες συσχέτισης
- Μοντέλο Παλινδρόμησης
- Έλεγχος Προϋποθέσεων (Ανάλυση καταλοίπων)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-34

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Διαγραμματική απεικόνιση
SCATTERPLOT + ΓΡΑΜΜΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-36

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Συσχέτιση

- Ρίχνουμε μια ματιά σε κανονικότητα (ελέγχους – QQplot – Ιστογράμματα)

ΕΛΕΓΧΟΙ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

| Tests of Normality | | | | | | |
|--------------------|---------------------------------|----|-------|--------------|----|------|
| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
| | Statistic | df | Sig. | Statistic | df | Sig. |
| distance | .112 | 10 | .200* | .970 | 10 | .892 |
| delivery | .142 | 10 | .200* | .937 | 10 | .520 |

^a. This is a lower bound of the true significance.
^a. Lilliefors Significance Correction

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-37

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Συσχέτιση

QQPLOTS

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-38

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Συσχέτιση

ΔΕΙΚΤΗΣ ΓΡΑΜΜΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ ΤΟΥ PEARSON

Correlations

| | | distance | delivery |
|----------|---------------------|----------|----------|
| distance | Pearson Correlation | 1 | .949** |
| | Sig. (2-tailed) | | .000 |
| | N | 10 | 10 |
| delivery | Pearson Correlation | .949** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 10 | 10 |

** . Correlation is significant at the 0.01 level

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-39

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Συσχέτιση

ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΔΕΙΚΤΕΣ

Correlations

| | | distance | delivery |
|-----------------|-------------------------|-------------------------------|----------|
| Kendall's tau_b | distance | Correlation Coefficient 1.000 | .841** |
| | | Sig. (2-tailed) | .001 |
| | | N | 10 |
| delivery | Correlation Coefficient | .841** | 1.000 |
| | Sig. (2-tailed) | .001 | |
| | N | 10 | 10 |
| Spearman's rho | distance | Correlation Coefficient 1.000 | .945** |
| | | Sig. (2-tailed) | .000 |
| | | N | 10 |
| delivery | Correlation Coefficient | .945** | 1.000 |
| | Sig. (2-tailed) | .000 | |
| | N | 10 | 10 |

** . Correlation is significant at the 0.01 level (2-tailed).

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-40

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση
ΠΕΡΙΛΗΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΟΝΤΕΛΟΥ

Εξαρτημένη μεταβλητή

ανεξάρτητη μεταβλητή

Διαφάνεια 10-41

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση
ΠΕΡΙΛΗΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΟΝΤΕΛΟΥ

Multiple Correlation Coefficient
(Συντελεστής πολλαπλής συσχέτισης)

- ✓ Παιρνει τιμές από 0 έως 1
- ✓ Στην απλή παλινδρόμηση $R=|r|$

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .949 ^a | .900 | .888 | .4800 |

a. Predictors: (Constant), distance

Διαφάνεια 10-42

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση
ΠΕΡΙΛΗΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΟΝΤΕΛΟΥ

$R^2 =$ % διακύμανσης που εξηγείται από το μοντέλο

- ✓ Χρησιμοποιείται ως μέτρο καλής προσαρμογής ή πρόβλεψης
- ✓ αυξάνει με κάθε μεταβλητή που προσθέτουμε
- ✓ Δεν ΠΡΕΠΕΙ να χρησιμοποιείται ως κριτήριο επιλογής μοντέλου (ΓΕΝΙΚΑ)
- ✓ Μπορούμε να συγκρίνουμε μοντέλα με ίδιο αριθμό μεταβλητών (άρα και μοντέλα απλής γραμμικής παλινδρόμησης)
- ✓ Στην απλή παλινδρόμηση $R^2=r^2$

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .949 ^a | .900 | .888 | .4800 |

a. Predictors: (Constant), distance

Διαφάνεια 10-43

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση
ΠΕΡΙΛΗΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΟΝΤΕΛΟΥ

$R_{adj}^2 =$ % διακύμανσης που εξηγείται από το μοντέλο διορθωμένο για τον αριθμό των μεταβλητών

- ✓ Λαμβάνει υπόψη του τις μεταβλητές
- ✓ Χρησιμοποιείται ως μέτρο καλής προσαρμογής ή πρόβλεψης
- ✓ ΔΕΝ αυξάνει με κάθε μεταβλητή που προσθέτουμε
- ✓ ΜΠΟΡΕΙ να χρησιμοποιηθεί ως κριτήριο επιλογής μοντέλου (ΓΕΝΙΚΑ)
- ✓ Στην απλής γραμμικής παλινδρόμησης δε διαφέρει πολύ από το R^2 .

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .949 ^a | .900 | .888 | .4800 |

a. Predictors: (Constant), distance

Διαφάνεια 10-44

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση ΠΕΡΙΛΗΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΟΝΤΕΛΟΥ

Τυπικό απόκλιση των σφαλμάτων

- ✓ Εκτίμηση του σ
- ✓ Συνδέεται άμεσα με την ακρίβεια των προβλέψεων του μοντέλου

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .949 ^a | .900 | .888 | .4800 |

a. Predictors: (Constant), distance

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-45

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση ΠΕΡΙΛΗΠΤΙΚΟΣ ΠΙΝΑΚΑΣ ΜΟΝΤΕΛΟΥ

Υποσημείωση που δείχνει τις μεταβλητές που χρησιμοποιούνται ως ανεξάρτητες:

- ✓ Εδώ έχουμε τον σταθερό όρο και την «απόσταση»

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .949 ^a | .900 | .888 | .4800 |

a. Predictors: (Constant), distance

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-46

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση ΠΙΝΑΚΑΣ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ ΜΟΝΤΕΛΟΥ

- ✓ Στην απλή παλινδρόμηση ελέγχει την υπόθεση: $H_0: \beta_1=0$ έναντι της εναλλακτικής $H_1: \beta_1 \neq 0$
- ✓ ΠΡΟΣΟΧΗ: στην πολλαπλή (που ακολουθεί, ενοτ. 8) η υπόθεση είναι διαφορετική
- ✓ ΕΛΕΓΧΕΙ ΚΑΤΑ ΠΟΣΟ ΤΟ ΤΡΕΧΩΝ ΜΟΝΤΕΛΟ ΔΙΑΦΕΡΕΙ ΑΠΟ ΤΟ ΣΤΑΘΕΡΟ (δηλαδή το μοντέλο $y=\beta_0+\epsilon$)

ANOVA^a

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1 | Regression | 16.682 | 1 | 16.682 | 72.396 | .000 ^b |
| | Residual | 1.843 | 8 | .230 | | |
| | Total | 18.525 | 9 | | | |

a. Predictors: (Constant), distance
b. Dependent Variable: delivery

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-47

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση ΠΙΝΑΚΑΣ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

ΕΚΤΙΜΗΣΕΙΣ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

- ✓ ΗΜΕΡΕΣ ΠΑΡΑΔΟΣΗΣ = $0.118 + 0.00359 \text{ ΜΙΛΙΑ} + \epsilon$, $\epsilon \sim \text{NORMAL}(0, 0.48^2)$

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .118 | .355 | | .333 | .748 |
| | distance | .00359 | .000421 | .949 | 8.509 | .000 |

a. Dependent Variable: delivery

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-48

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΠΙΝΑΚΑΣ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

ΤΥΠΙΚΑ ΣΦΑΛΜΑΤΑ ΕΚΤΙΜΗΣΕΩΝ ΠΑΡΑΜΕΤΡΩΝ

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{Var(\hat{\beta}_0)} = 0.355, \hat{\sigma}_{\hat{\beta}_1} = \sqrt{Var(\hat{\beta}_1)} = 0.000421$$

| Coefficients ^a | | | | | | |
|---------------------------|------------|-----------------------------|------------|---------------------------|-------|------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .118 | .355 | | .333 | .748 |
| | distance | .00359 | .000421 | .949 | 8.509 | .000 |

^a. Dependent Variable: delivery

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-49

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΠΙΝΑΚΑΣ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

Ελεγχος συντηρήσεως t για τον έλεγχο αν κάθε παράμετρος είναι μηδέν ή όχι $t = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$

$$t_{\hat{\beta}_0} = \frac{\hat{\beta}_0}{\hat{\sigma}_{\hat{\beta}_0}} = \frac{0.118}{0.355} = 0.333, t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{0.00359}{0.000421} = 8.527$$

| Coefficients ^a | | | | | | |
|---------------------------|------------|-----------------------------|------------|---------------------------|-------|------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .118 | .355 | | .333 | .748 |
| | distance | .00359 | .000421 | .949 | 8.509 | .000 |

^a. Dependent Variable: delivery

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-50

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΠΙΝΑΚΑΣ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

P-values για τους έλεγχους στατιστικής σημαντικότητας κάθε παραμέτρου

| Coefficients ^a | | | | | | |
|---------------------------|------------|-----------------------------|------------|---------------------------|-------|------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .118 | .355 | | .333 | .748 |
| | distance | .00359 | .000421 | .949 | 8.509 | .000 |

^a. Dependent Variable: delivery

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-51

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΠΙΝΑΚΑΣ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

Τυποποιημένοι συντελεστές

- ✓ Οι συντελεστές αν χρησιμοποιήσουμε τις τυποποιημένες μεταβλητές Z_x και Z_y
- ✓ Ο τυποποιημένος συντελεστής του β_0 είναι πάντα 0

Ερμηνεία: πόσες τυπικές αποκλίσεις αναμένουμε να αυξηθεί η Y όταν X αυξάνεται κατά μία τυπική απόκλιση;

| Coefficients ^a | | | | | | |
|---------------------------|------------|-----------------------------|------------|---------------------------|-------|------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .118 | .355 | | .333 | .748 |
| | distance | .00359 | .000421 | .949 | 8.509 | .000 |

^a. Dependent Variable: delivery

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-52

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΠΙΝΑΚΑΣ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

Τυποποιημένοι συντελεστές
✓ Στην γραμμική παλινδρόμηση ο τυποποιημένος συντελεστής του β_0 είναι ίσος με τον συντελεστή του Pearson

| | delivery | distance |
|----------|-----------------|----------|
| delivery | | .949** |
| distance | .949** | |
| | Sig. (2-tailed) | .000 |
| | N | 10 |

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .118 | .355 | | .333 | .748 |
| 1 | distance | .00359 | .000421 | .949 | 8.509 | .000 |

^a. Dependent Variable: delivery

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διοφάνεια 10-53

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΕΡΜΗΝΕΙΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΠΑΡΑΜΕΤΡΟΣ $\beta_1=0.00359$

- ✓ **ΥΠΑΡΧΕΙ ΣΧΕΣΗ; ΝΑΙ**
 $P=0.000 < 0.05$ δηλαδή απορρίπτουμε τη $H_0 \Rightarrow$ Συνεπώς η απόσταση επηρεάζει τον χρόνο παράδοσης
- ✓ **ΤΙ ΣΧΕΣΗ; ΘΕΤΙΚΗ**
 $\beta_1 > 0$ συνεπώς θετική σχέση \Rightarrow όσο αυξάνει η απόσταση τόσο μεγαλώνει ο χρόνος παράδοσης
- ✓ **ΠΟΣΟ ΕΠΗΡΕΑΖΕΙ Η ΑΠΟΣΤΑΣΗ ΤΗΝ ΠΑΡΑΔΟΣΗ;**
 - Με κάθε επιπλέον μίλι ο αναμενόμενος χρόνος παράδοσης αυξάνει κατά 0.00359 μέρες (περίπου 5 λεπτά)
 - Με κάθε επιπλέον 100 μίλια ο αναμενόμενος χρόνος παράδοσης αυξάνει κατά 0.359 μέρες (περίπου 8.6 ώρες)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διοφάνεια 10-54

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΕΡΜΗΝΕΙΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΠΑΡΑΜΕΤΡΟΣ β_1

- Έστω $X_1=X$ και $X_2=X+1$ τότε
- $\mu_1 = \beta_0 + \beta_1 X_1 = \beta_0 + \beta_1 X$
- $\mu_2 = \beta_0 + \beta_1 X_2 = \beta_0 + \beta_1 (X+1)$
- $\Delta\mu = \mu_2 - \mu_1 = \beta_0 + \beta_1 (X+1) - \beta_0 - \beta_1 X = \beta_1$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διοφάνεια 10-55

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΕΡΜΗΝΕΙΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΠΑΡΑΜΕΤΡΟΣ $\beta_0=0.118$

- ✓ **ΠΡΕΠΕΙ ΝΑ ΥΠΑΡΧΕΙ ΣΤΗΝ ΕΞΙΣΩΣΗ; ΟΧΙ**
 $P=0.748 > 0.05$ δηλαδή δεν απορρίπτουμε τη $H_0 \Rightarrow$ Συνεπώς η σταθερά μπορεί να θεωρηθεί 0 και να αφαιρεθεί από το μοντέλο
- ✓ **ΕΡΜΗΝΕΙΑ;**
 - Όταν η απόσταση είναι μηδενική τότε ο χρόνος παράδοσης είναι 0.118 μέρες (2.8 ώρες)
 - Χρόνος παράδοσης όταν το φορτίο είναι πολύ κοντά
 - ΠΡΟΣΟΧΗ είναι εκτός των ορίων της X
- ✓ **ΝΑ ΤΟ ΑΦΑΙΡΕΣΟΥΜΕ; ΜΑΛΛΟΝ ΝΑΙ**
 Εδώ η λογική και ο έλεγχος μας λέει ναι

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διοφάνεια 10-56

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΕΡΜΗΝΕΙΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΠΡΟΒΛΕΠΤΙΚΗ ΙΚΑΝΟΤΗΤΑ/ ΠΡΟΣΑΡΜΟΓΗ ΜΟΝΤΕΛΟΥ

✓ $R=r=0.95$ & $R^2=0.89$;

- Υψηλή σχέση
- Υψηλή προσαρμογή και καλή προβλεπτική ικανότητα
- 89% της διακύμανσης εξηγείται από το μοντέλο

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-57

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΕΡΜΗΝΕΙΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΤΥΠΟΠΟΙΗΜΕΝΟΣ ΣΥΝΤΕΛΕΣΤΗΣ $\beta_1=0.949$

✓ Αν η απόσταση αυξηθεί κατά μία τυπική απόκλιση (380 μίλια) τότε ο χρόνος παράδοσης αναμένουμε να αυξηθεί κατά 0.95 τυπικές αποκλίσεις (δηλαδή κατά $0.949 \cdot 1.435 = 1.36$ μέρες).

Descriptive Statistics

| | N | Minimum | Maximum | Mean | Std. Deviation |
|--------------------|----|---------|---------|--------|----------------|
| distance | 10 | 215 | 1350 | 762.00 | 379.746 |
| delivery | 10 | 1.0 | 5.0 | 2.850 | 1.4347 |
| Valid N (listwise) | 10 | | | | |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-58

10.2. Απλή γραμμική παλινδρόμηση
10.2.3. Εφαρμογή στο SPSS

Παράδειγμα 10-1: Παλινδρόμηση

ΕΡΜΗΝΕΙΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΤΥΠΟΠΟΙΗΜΕΝΟΣ ΣΥΝΤΕΛΕΣΤΗΣ β_1

$$\hat{\beta}_0^{(st)} = \bar{Z}_x - \hat{\beta}_1^{(st)} \bar{Z}_y = 0 \quad \hat{\beta}_1^{(st)} = \frac{s_{Z_y}}{s_{Z_x}} r_{Z_x Z_y} = r_{Z_x Z_y}$$

$$r_{Z_x Z_y} = \frac{\sum_{i=1}^n (Z_{x,i} - \bar{Z}_x)(Z_{y,i} - \bar{Z}_y)}{\sqrt{\sum_{i=1}^n (Z_{x,i} - \bar{Z}_x)^2 \sum_{i=1}^n (Z_{y,i} - \bar{Z}_y)^2}} = \frac{\sum_{i=1}^n Z_{x,i} Z_{y,i}}{\sqrt{\sum_{i=1}^n Z_{x,i}^2 \sum_{i=1}^n Z_{y,i}^2}} =$$

$$= \frac{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_x} \frac{Y_i - \bar{Y}}{s_y} \right)}{\sqrt{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_x} \right)^2 \sum_{i=1}^n \left(\frac{Y_i - \bar{Y}}{s_y} \right)^2}} = r_{XY}$$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-59

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΠΡΟΫΠΟΘΕΣΕΙΣ

- ✓ Κανονικότητα σφαλμάτων (και Y)
- ✓ Ανεξαρτησία σφαλμάτων (και Y)
- ✓ Ομοσκεδαστικότητα σφαλμάτων (και Y)
- ✓ Γραμμικότητα μεταξύ X και Y

✓ Δουλεύουμε με κατάλοιπα e_i

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-60

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΚΑΝΟΝΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

Διαγραμματική απεικόνιση κατανομής των καταλοίπων

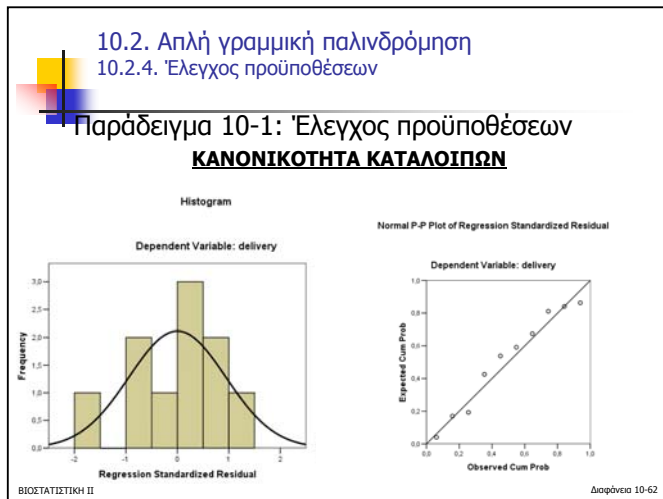
ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διοφάνεια 10-61

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΚΑΝΟΝΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

Αποθήκευση τυποποιημένων καταλοίπων

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διοφάνεια 10-63



10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΚΑΝΟΝΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

Έλεγχος κανονικότητας τυποποιημένων καταλοίπων (EXPLORE)

| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|-----------------------------|---------------------------------|----|-------------------|--------------|----|------|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| ZRE:1 Standardized Residual | .140 | 10 | .200 ^a | .934 | 10 | .487 |

^a. This is a lower bound of the true significance.
^a. Lilliefors Significance Correction

OK με την προϋπόθεση κανονικότητας

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διοφάνεια 10-64

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

Η ανεξαρτησία των καταλοίπων δεν είναι εύκολα ελεγχόμενη

- Time sequence plot
- Έλεγχος τυχαιότητας (RUNS TEST)
- Έλεγχος για αυτοσυσχετίσεις [κυρίως για χρονολογικά δεδομένα]
 - ✓ δείκτης Durbin – Watson (συσχέτιση διαδοχικών σφαλμάτων – αυτοσυσχέτιση 1ης τάξης)
 - ✓ Box-Ljung έλεγχος αυτοσυσχετίσεων και ACF Plots
 - ✓ AR μοντέλα

Για λεπτομέρειες βλ. Ryan 1997 σελ. 46-47

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-65

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

- TIME SEQUENCE PLOT

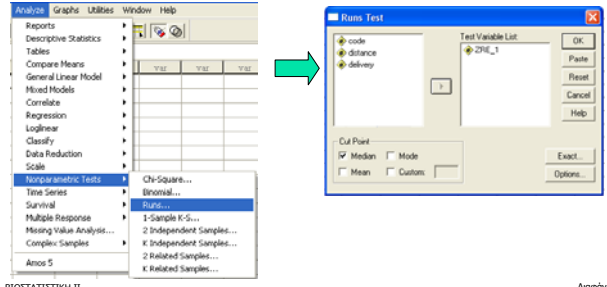


ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-66

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

- RUNS TEST



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-67

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

- RUNS TEST

| Runs Test | |
|-------------------------|-----------------------------------|
| | ZRE_1 Standardized Residual |
| Test Value ^a | .16337 |
| Cases < Test Value | 5 |
| Cases >= Test Value | 5 |
| Total Cases | 10 |
| Number of Runs | 3 |
| Z | -1.677 |
| Asymp. Sig. (2-tailed) | .094 |
| Exact Sig. (2-tailed) | .079 |
| Point Probability | .032 |

a. Median

Δεν απορρίπτεται η υπόθεση της τυχαιότητας για α=0.05

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-68

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

- Η ανεξαρτησία των καταλοίπων δεν είναι εύκολα ελεγχόμενη
- Μπορούμε να ελέγξουμε αυτοσυσχέτιση με τον δείκτη Durbin – Watson [κυρίως για χρονολογικά δεδομένα]
 - ✓ $0 < D < 4$
 - ✓ $0 < D < 2$ θετική αυτοσυσχέτιση
 - ✓ $2 < D < 4$ αρνητική αυτοσυσχέτιση
 - ✓ $D = 2$ ⇔ δεν υπάρχει αυτοσυσχέτιση
 - ✓ Έλεγχος από πίνακες:
 - Μόνο για $15 \leq n \leq 100$
 - $D < d_L$ απορρίπτουμε την υπόθεση μηδενικής αυτοσυσχέτισης
 - $D > d_U$ δεν απορρίπτουμε την υπόθεση μηδενικής αυτοσυσχέτισης
 - Διαφορετικά δεν μπορούμε να αποφασίσουμε
 - Χοντρικά μπορούμε να απορρίψουμε τη μηδενική αυτοσυσχέτιση για $1.5 < D < 2.5$ [d_L για $n=100, \rho=1$]

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-69

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

Model Summary^b

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------------------|----------|-------------------|----------------------------|---------------|
| 1 | .949 ^a | .900 | .888 | .4800 | .753 |

a. Predictors: (Constant), distance
b. Dependent Variable: delivery

↑
Έλεγχος Durbin-Watson
Επικίνδυνα χαμηλή τιμή

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-71

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

↑
Έλεγχος Durbin-Watson

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-70

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

- ACF PLOTS και έλεγχοι BOX-LJUNG

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνειο 10-72

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

- ACF PLOTS και έλεγχοι BOX-LJUNG

| Lag | Autocorrelation | Std. Error ^a | Value | df | Sig. ^b |
|-----|-----------------|-------------------------|--------|----|-------------------|
| 1 | .500 | .274 | 3.327 | 1 | .068 |
| 2 | -.107 | .258 | 3.408 | 2 | .174 |
| 3 | -.397 | .242 | 6.199 | 3 | .102 |
| 4 | -.577 | .224 | 12.864 | 4 | .012 |
| 5 | -.381 | .204 | 16.345 | 5 | .006 |
| 6 | -.106 | .183 | 16.682 | 6 | .011 |
| 7 | .092 | .158 | 17.020 | 7 | .017 |
| 8 | .142 | .129 | 18.231 | 8 | .020 |

a. The underlying process assumed is independence (white noise).
b. Based on the asymptotic chi-square approximation.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-73

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΑΝΕΞΑΡΤΗΣΙΑ ΚΑΤΑΛΟΙΠΩΝ

- ACF PLOTS και έλεγχοι BOX-LJUNG

| Lag | Partial Autocorrelation | Std. Error |
|-----|-------------------------|------------|
| 1 | .500 | .316 |
| 2 | -.190 | .316 |
| 3 | -.505 | .316 |
| 4 | -.258 | .316 |
| 5 | .111 | .316 |
| 6 | -.060 | .316 |
| 7 | -.287 | .316 |
| 8 | -.192 | .316 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-74

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA X
- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA \hat{Y}
- ΔΙΑΓΡΑΜΜΑ RESIDUALS² ANA \hat{Y}
- ΕΛΕΓΧΟΙ ΙΣΟΤΗΤΑΣ ΔΙΑΚΥΜΑΝΣΕΩΝ ΣΕ ΟΜΑΔΟΠΟΙΗΣΕΙΣ ΤΗΣ X (ΣΕ 4 ή περισσότερες)
- Βλ. Gunst & Mason (1980, σελ. 237)
- Βλ. Draper & Smith (1998, 3rd edition, σελ. 56-59, 62-67)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-75

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA X

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-76

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA \hat{Y}

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-77

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA \hat{Y}

Scatterplot

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-78

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA \hat{Y}

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-79

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA Y ;

ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΙΙ Διαφάνεια 10-80

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

■ ΔΙΑΓΡΑΜΜΑ RESIDUALS² ANA \hat{Y}

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-81

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ

■ ΕΛΕΓΧΟΣ ΙΣΟΤΗΤΑΣ ΔΙΑΚΥΜΑΝΣΕΩΝ ANA ΤΕΤΑΡΤΗΜΟΡΙΑ Χ

Test of Homogeneity of Variances

| Standardized Residual | | | |
|-----------------------|-----|-----|------|
| Levene Statistic | df1 | df2 | Sig. |
| .880 | 3 | 6 | .503 |

Δε φαίνεται διαφορά. Προσοχή όμως έχουμε λίγες παρατηρήσεις

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-82

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΓΡΑΜΜΙΚΟΤΗΤΑ Χ & Υ

■ ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA \hat{Y}
[ΠΡΕΠΕΙ ΝΑ ΕΙΝΑΙ ΣΕ ΜΙΑ ΠΑΡΑΛΛΗΛΗ ΖΩΝΗ]
[-2 ΕΩΣ 2 ΓΙΑ ST.RESIDUALS]

■ Scatterplot X & Y

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-83

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

■ Παράδειγμα 10-1: Έλεγχος προϋποθέσεων
ΓΡΑΜΜΙΚΟΤΗΤΑ Χ & Υ

■ ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA \hat{Y}
Scatterplot

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-84

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y

- ΑΝΤΙΠΑΡΑΔΕΙΓΜΑ
 - $X \sim \text{NORMAL}(0,1)$ [ΔΕΙΓΜΑ 100 ΤΙΜΩΝ]
 - $Y = 5 + x + 10x^2$
- ΜΟΝΤΕΛΟ ΓΡ. ΠΑΛΙΝΔΡΟΜΗΣΗΣ:
 - $Y = 15 + 6.67X + \varepsilon$
 - $\varepsilon \sim \text{Normal}(0, 18.34^2)$
 - $R^2 = 0.12$

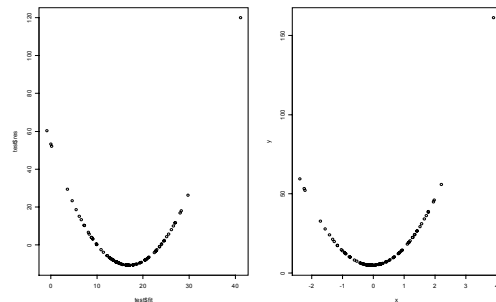
ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-85

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-86

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y

- ΑΝΤΙΠΑΡΑΔΕΙΓΜΑ #2
 - $X \sim \text{SEQ}(0,5,0.1)$ [ΔΕΙΓΜΑ 100 ΤΙΜΩΝ]
 - $Y = 5x^2$
- ΜΟΝΤΕΛΟ ΓΡ. ΠΑΛΙΝΔΡΟΜΗΣΗΣ:
 - $Y = -20 + 25X + \varepsilon$
 - $\varepsilon \sim \text{Normal}(0, 9.88^2)$
 - $R^2 = 0.93$

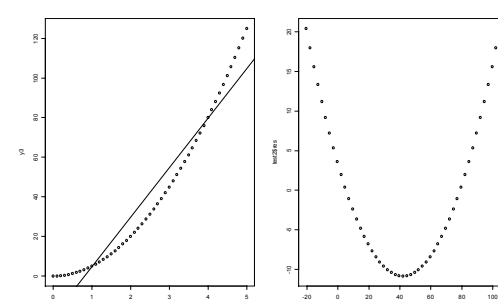
ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-87

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-88

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y

- ΑΝΤΙΠΑΡΑΔΕΙΓΜΑ #3
 - $X \sim \text{SEQ}(0.1, 10, 0.1)$ [ΔΕΙΓΜΑ 100 ΤΙΜΩΝ]
 - $Y = -3 + 2 \ln(x)$
- ΜΟΝΤΕΛΟ ΓΡ. ΠΑΛΙΝΔΡΟΜΗΣΗΣ:
 - $Y = -2.77 + 0.57x + \varepsilon$
 - $\varepsilon \sim \text{Normal}(0, 0.83^2)$
 - $R^2 = 0.80$

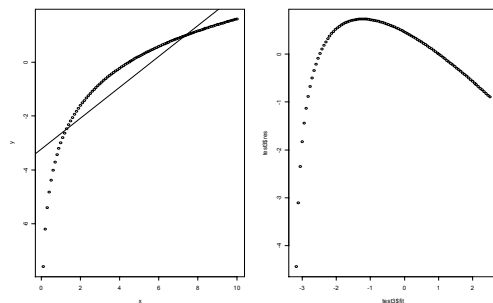
ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-89

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-90

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y

- ΑΝΤΙΠΑΡΑΔΕΙΓΜΑ #4
 - $X \sim \text{SEQ}(0.1, 10, 0.1)$ [ΔΕΙΓΜΑ 100 ΤΙΜΩΝ]
 - $Y = \exp(-3) x^2$
- ΜΟΝΤΕΛΟ ΓΡ. ΠΑΛΙΝΔΡΟΜΗΣΗΣ:
 - $Y = -0.855 + 0.5x + \varepsilon$
 - $\varepsilon \sim \text{Normal}(0, 0.375^2)$
 - $R^2 = 0.939$

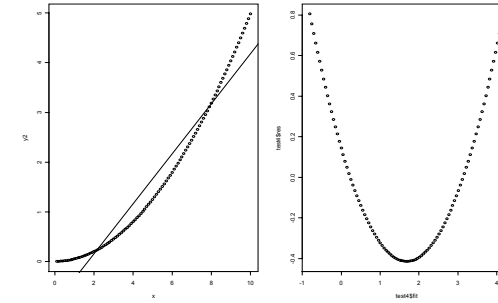
ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-91

10.2. Απλή γραμμική παλινδρόμηση
10.2.4. Έλεγχος προϋποθέσεων

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΓΡΑΜΜΙΚΟΤΗΤΑ X & Y



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-92

10.2. Απλή γραμμική παλινδρόμηση

10.2.5. Ακραίες τιμές

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΚΡΑΙΕΣ ΤΙΜΕΣ

- ΔΙΑΓΡΑΜΜΑ ST. RESIDUALS ANA \hat{Y}
[ΝΑ ΕΙΝΑΙ ΜΕΤΑΞΥ -2 ΚΑΙ 2]
- ΑΠΛΟ SCATTERPLOT Χ+Υ
- HISTOGRAM ΚΑΤΑΛΟΙΠΩΝ (ST.)
- DELETED RESIDUALS
- ΠΙΟ ΠΡΟΧΩΡΗΜΕΝΕΣ ΤΕΧΝΙΚΕΣ
- INFLUENTIAL OBSERVATIONS: ΠΑΡΑΤΗΡΗΣΕΙΣ ΠΟΥ ΕΠΗΡΕΑΖΟΥΝ – ΑΛΛΟΙΩΝΟΥΝ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-93

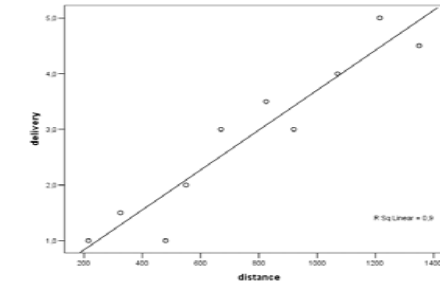
10.2. Απλή γραμμική παλινδρόμηση

10.2.5. Ακραίες τιμές

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΚΡΑΙΕΣ ΤΙΜΕΣ

- ΑΠΛΟ SCATTERPLOT Χ+Υ



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-95

10.2. Απλή γραμμική παλινδρόμηση

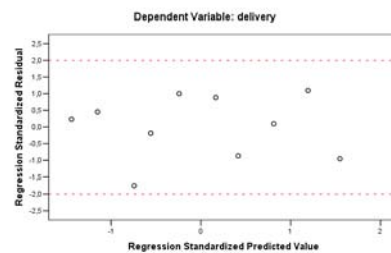
10.2.5. Ακραίες τιμές

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΚΡΑΙΕΣ ΤΙΜΕΣ

- ΔΙΑΓΡΑΜΜΑ RESIDUALS ANA \hat{Y}

Scatterplot



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-94

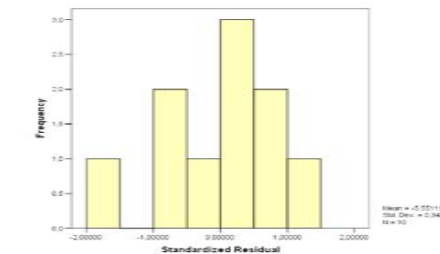
10.2. Απλή γραμμική παλινδρόμηση

10.2.5. Ακραίες τιμές

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΚΡΑΙΕΣ ΤΙΜΕΣ

- ΑΠΛΟ ΙΣΤΟΓΡΑΜΜΑ ΚΑΤΑΛΟΙΠΩΝ



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-96

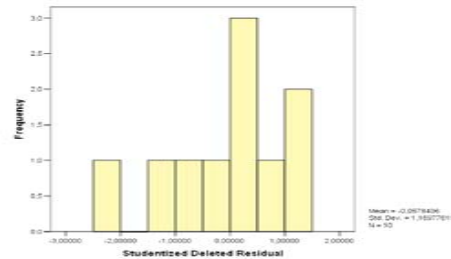
10.2. Απλή γραμμική παλινδρόμηση

10.2.5. Ακραίες τιμές

Παράδειγμα 10-1: Έλεγχος προϋποθέσεων

ΑΚΡΑΙΕΣ ΤΙΜΕΣ

- ΑΠΛΟ ΙΣΤΟΓΡΑΜΜΑ ΚΑΤΑΛΟΙΠΩΝ (studentized deleted)



ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-97

10.2. Απλή γραμμική παλινδρόμηση

10.2.6. Μετασχηματισμοί της μεταβλητής απόκρισης

- Με μετασχηματισμούς της Y μπορούν να λυθούν προβλήματα

- ✓ Κανονικότητας
- ✓ Ομοσκεδαστικότητας
- ✓ Μη γραμμικότητας

Βλ. Draper & Smith (1998, 3rd edition, σελ. 56-59, 277-319)

- Με μετασχηματισμούς της X μπορούν να λυθούν προβλήματα

- ✓ Μη γραμμικότητας

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-98

10.2. Απλή γραμμική παλινδρόμηση

10.2.6. Μετασχηματισμοί της μεταβλητής απόκρισης

- Συνηθισμένοι μετασχηματισμοί του Y

- ✓ $\ln(y) = \beta_0 + \beta_1 X + \varepsilon$
 - $Y = \exp(\beta_0 + \beta_1 X + \varepsilon) = B_0(B_1)^X E$, $E \sim \text{LOGNORMAL}$
 - Προσοχή είναι διαφορετικό από $Y = \exp(\beta_0 + \beta_1 X) + \varepsilon$, $\varepsilon \sim \text{Normal}$
- ✓ $Y^{-1} = \beta_0 + \beta_1 X + \varepsilon$

- ✓ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ BOX-COX

- $Z_\lambda = (Y^\lambda - 1)/\lambda$, $\lambda \neq 0$, $Z_0 = \ln(x)$
- Για θετικά Y μόνο
- Μπορούμε να επιλέξουμε λ με ML ή με R^2 .
- Δύσκολη η ερμηνεία

Βλ. Draper & Smith (1998, 3rd edition, σελ. 56-59, 277-319)

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-99

10.2. Απλή γραμμική παλινδρόμηση

10.2.6. Μετασχηματισμοί της μεταβλητής απόκρισης

- Συνηθισμένοι μετασχηματισμοί του X & Y

- ✓ $y = \beta_0 + \beta_1 \ln(X) + \varepsilon$
- ✓ $\ln(y) = \beta_0 + \beta_1 \ln(X) + \varepsilon$
 - $Y = \exp(\beta_0 + \beta_1 \ln X + \varepsilon) = B_0(X)^{\beta_1} E$, $E \sim \text{LOGNORMAL}$
- ✓ $Y = \beta_0 + \beta_1 X^{-1} + \varepsilon$
- ✓ $Y^{-1} = \beta_0 + \beta_1 X^{-1} + \varepsilon$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II

Διαφάνεια 10-100

10.2. Απλή γραμμική παλινδρόμηση

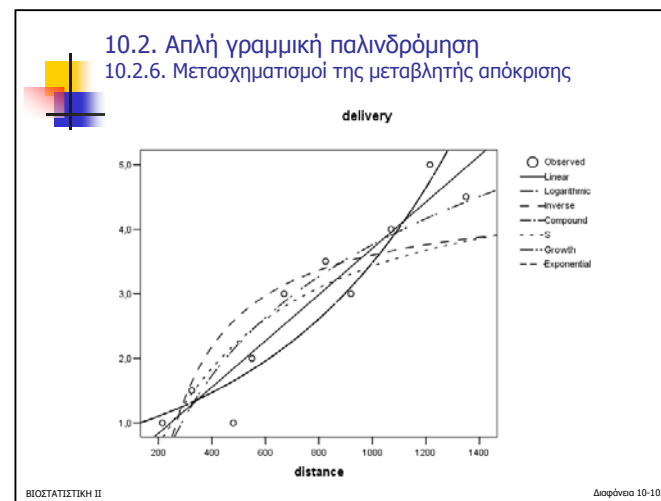
10.2.6. Μετασχηματισμοί της μεταβλητής απόκρισης

$Y = b_0 + b_1 \ln(X)$
 $Y = b_0 + b_1/X$

$\ln(Y) = \ln(b_0) + \ln(b_1) \cdot x$
 $\ln(Y) = b_0 + b_1/x$

$\ln(Y) = b_0 + b_1 x$
 $\ln(Y) = \ln(b_0) + b_1 x$

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-101



10.2. Απλή γραμμική παλινδρόμηση

10.2.6. Μετασχηματισμοί της μεταβλητής απόκρισης

Model Summary and Parameter Estimates

Dependent Variable: delivery

| Equation | Model Summary | | | | | Parameter Estimates | |
|-------------|---------------|--------|-----|-----|------|---------------------|----------|
| | R Square | F | df1 | df2 | Sig. | Constant | b1 |
| Linear | .900 | 72.396 | 1 | 8 | .000 | .118 | .004 |
| Logarithmic | .839 | 41.612 | 1 | 8 | .000 | -11.528 | 2.213 |
| Inverse | .662 | 15.639 | 1 | 8 | .004 | 4.550 | -948.814 |
| Compound | .836 | 40.781 | 1 | 8 | .000 | .826 | 1.001 |
| S | .722 | 20.730 | 1 | 8 | .002 | 1.645 | -412.806 |
| Growth | .836 | 40.781 | 1 | 8 | .000 | -.192 | .001 |
| Exponential | .836 | 40.781 | 1 | 8 | .000 | .826 | .001 |

The independent variable is distance.

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-102

10.2. Απλή γραμμική παλινδρόμηση

10.3. Σχέση παλινδρόμησης με το paired t-test

Και στις 2 προσεγγίσεις έχουμε 2 ποσοτικές μεταβλητές όσον αφορά το software

- Τι κάνει το ένα και τι το άλλο;

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-104

10.2. Απλή γραμμική παλινδρόμηση
10.3. Σχέση παλινδρόμησης με το paired t-test

ΠΑΡΑΔΕΙΓΜΑ: WORLD95
ΣΥΓΚΡΙΣΗ ΘΗΝΣΙΜΟΤΗΤΑΣ ΑΝΔΡΩΝ ΚΑΙ ΓΥΝΑΙΚΩΝ

Paired Samples Correlations

| Pair | | N | Correlation | Sig. |
|------|---|-----|-------------|------|
| 1 | Average female life expectancy & Average male life expectancy | 109 | .982 | .000 |

Ο δείκτης συσχέτισης 2 διαδοχικών χρονικά μετρήσεων ονομάζεται και intraclass correlation coefficient

Paired Samples Test

| Pair | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|------|---|--------------------|----------------|-----------------|---|-------|--------|-----|-----------------|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| 1 | Average female life expectancy - Average male life expectancy | 5.239 | 2.269 | .217 | 4.808 | 5.669 | 24.109 | 108 | .000 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-105

10.2. Απλή γραμμική παλινδρόμηση
10.3. Σχέση παλινδρόμησης με το paired t-test

ΠΑΡΑΔΕΙΓΜΑ: WORLD95

ANOVA^a

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|-----|-------------|----------|-------------------|
| 1 | Regression | 11648.224 | 1 | 11648.224 | 2952.585 | .000 ^a |
| | Residual | 422.125 | 107 | 3.945 | | |
| | Total | 12070.349 | 108 | | | |

a. Predictors: (Constant), Average male life expectancy
b. Dependent Variable: Average female life expectancy

Paired Samples Test

| Pair | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|------|---|--------------------|----------------|-----------------|---|-------|--------|-----|-----------------|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| 1 | Average female life expectancy - Average male life expectancy | 5.239 | 2.269 | .217 | 4.808 | 5.669 | 24.109 | 108 | .000 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-106

10.2. Απλή γραμμική παλινδρόμηση
10.3. Σχέση παλινδρόμησης με το paired t-test

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|-----|-------------|----------|-------------------|
| 1 | Regression | 11648.224 | 1 | 11648.224 | 2952.585 | .000 ^a |
| | Residual | 422.125 | 107 | 3.945 | | |
| | Total | 12070.349 | 108 | | | |

a. Predictors: (Constant), Average male life expectancy
b. Dependent Variable: Average female life expectancy

Paired Samples Test

| Pair | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|------|---|--------------------|----------------|-----------------|---|-------|--------|-----|-----------------|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| 1 | Average female life expectancy - Average male life expectancy | 5.239 | 2.269 | .217 | 4.808 | 5.669 | 24.109 | 108 | .000 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-107

10.2. Απλή γραμμική παλινδρόμηση
10.3. Σχέση παλινδρόμησης με το paired t-test

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------------------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -2.550 | 1.351 | | -1.887 | .062 |
| | Average male life expectancy | 1.120 | .021 | .982 | 54.338 | .000 |

a. Dependent Variable: Average female life expectancy

Paired Samples Test

| Pair | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|------|---|--------------------|----------------|-----------------|---|-------|--------|-----|-----------------|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| 1 | Average female life expectancy - Average male life expectancy | 5.239 | 2.269 | .217 | 4.808 | 5.669 | 24.109 | 108 | .000 |

ΒΙΟΣΤΑΤΙΣΤΙΚΗ II Διαφάνεια 10-108