

# Διαχείριση Μεγάλης Ποσότητας Πληροφοριών -ΔΜΠΠ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

ΤΜΗΜΑ ΑΡΧΕΙΟΝΟΜΙΑΣ, ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ  
ΜΟΥΣΕΙΟΛΟΓΙΑΣ

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

*Ανάλυση ιαλινδρόμησης (Regression Analysis)*



ΚΕΡΚΥΡΑ

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



## Πίνακας περιεχομένων

|       |                                                    |    |
|-------|----------------------------------------------------|----|
| 1.    | Άσκηση α – Ανάλυση απλής παλινδρόμησης .....       | 4  |
| 1.1   | Δεδομένα .....                                     | 4  |
| 1.2   | Εισαγωγή δεδομένων.....                            | 5  |
| 1.3   | Επεξεργασία δεδομένων.....                         | 6  |
| 1.3.1 | Προϋποθέσεις .....                                 | 6  |
| 1.3.2 | Διατύπωση υπόθεσης.....                            | 6  |
| 1.3.3 | Υπολογισμός του συντελεστή συσχέτισης Pearson..... | 6  |
| 1.3.4 | Ανάλυση απλής παλινδρόμησης.....                   | 7  |
| 1.4   | Ερμηνεία των αποτελεσμάτων .....                   | 8  |
| 2.    | Άσκηση β – Ανάλυση πολλαπλής παλινδρόμησης .....   | 10 |
| 2.1   | Δεδομένα .....                                     | 10 |
| 2.2   | Εισαγωγή δεδομένων.....                            | 10 |
| 2.3   | Επεξεργασία δεδομένων.....                         | 11 |
| 2.3.1 | Προϋποθέσεις .....                                 | 11 |
| 2.3.2 | Διατύπωση υπόθεσης.....                            | 11 |
| 2.3.3 | Υπολογισμός του συντελεστή συσχέτισης Pearson..... | 11 |
| 2.3.4 | Ανάλυση πολλαπλής παλινδρόμησης.....               | 12 |
| 2.4   | Ερμηνεία αποτελεσμάτων .....                       | 13 |
| 3.    | Ασκήσεις κατανόησης.....                           | 15 |
| 3.1   | Δεδομένα .....                                     | 15 |
| 4.    | Ορολογία .....                                     | 16 |
| 5.    | Βιβλιογραφία .....                                 | 17 |
| 6.    | Σημειώματα .....                                   | 18 |

## 1. Άσκηση α – Ανάλυση απλής παλινδρόμησης

### 1.1 Δεδομένα

Τα Αρχεία Νομού Κέρκυρας είναι περιφερειακή υπηρεσία των Γενικών Αρχείων του Κράτους. Βασικός σκοπός των Γ.Α.Κ. είναι η εποπτεία, διάσωση, συγκέντρωση, ταξινόμηση και ευρετηρίαση του αρχειακού υλικού του νομού, καθώς και η διάθεση και η με ποικίλους τρόπους αξιοποίηση όλου αυτού του αρχειακού πλούτου (<http://gak.ker.sch.gr/>). Στεγάζεται δε στους χώρους του Παλαιού Φρουρίου.



Υποθέστε πως σας ζητάνε να μελετήσετε την ικανοποίηση στον χώρο εργασίας\* ενός αρχείου και να εξετάσετε τη σχέση της με την επαγγελματική πρόοδο. Για τον σκοπό αυτό δίνετε σε μία ομάδα εργαζομένων ένα ερωτηματολόγιο το οποίο περιέχει μία κλίμακα μέτρησης της ικανοποίησης στον χώρο εργασίας. Στόχος σας είναι εξετάσετε εάν μπορείτε να προβλέψετε την επαγγελματική πρόοδο, βασισμένοι στην βαθμολογία της ικανοποίησης του ερωτηματολογίου.

Οι βαθμολογίες που συγκεντρώσατε από τα ερωτηματολόγια συγκεντρώνονται στον παρακάτω πίνακα:

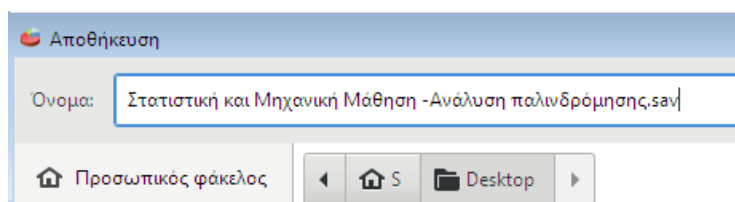
| Εργαζόμενος | Ικανοποίηση στον χώρο εργασίας | Επαγγελματική πρόοδο |
|-------------|--------------------------------|----------------------|
| 1           | 74                             | 56                   |
| 2           | 57                             | 58                   |
| 3           | 45                             | 43                   |
| 4           | 81                             | 75                   |
| 5           | 56                             | 61                   |
| 6           | 62                             | 60                   |
| 7           | 33                             | 42                   |
| 8           | 83                             | 93                   |
| 9           | 72                             | 59                   |
| 10          | 53                             | 64                   |

\* Κατά τον Abraham Maslow, Αμερικανό ψυχολόγο, η ικανοποίηση της ανάγκης για εργασία ανήκει στη δεύτερη βαθμίδα ιεράρχησης των αναγκών η οποίες συνθέτουν την ανθρώπινη προσωπικότητα (<https://www.simplypsychology.org/maslow.html#gsc.tab=0>).

## 1.2 Εισαγωγή δεδομένων



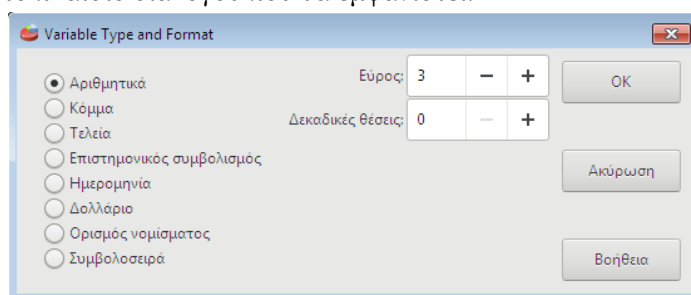
Στο στατιστικό πακέτο του εργαστηρίου σας να δημιουργηθεί το αρχείο “Ανάλυση παλινδρόμησης\_α.sav”. Αποθηκεύστε το αρχείο.



Να συμπληρωθεί η καρτέλα Απεικ.Μεταβλητών (Variable View) ως εξής:

| Όνομα                 | Τύπος      | Εύρος | Δεκαδικά | Ετικέτα       | Ταμπέλες Τιμών | Ελλειπούσες Τιμές | Στήλες | Ευθυγρ/ση | Μέτρηση | Ρόλος    |
|-----------------------|------------|-------|----------|---------------|----------------|-------------------|--------|-----------|---------|----------|
| Ικανοποίηση_Εργασία   | Αριθμητικά | 2     | 0        | Ικανοποίηση   | Κανένα         | Κανένα            | 8      | Δεξιά     | Κλίμακα | Εισαγωγή |
| Επαγγελματική_Πρόοδος | Αριθμητικά | 2     | 0        | Επαγγελματική | Κανένα         | Κανένα            | 8      | Δεξιά     | Κλίμακα | Εισαγωγή |

- Στη στήλη Όνομα να δώσετε τα παρακάτω ονόματα μεταβλητών: Ικανοποίηση\_Εργασία, Επαγγελματική\_Πρόοδος. Προσοχή! Χωρίς κενά.
- Κάντε διπλό κλικ στο κελί στη στήλη Type (Τύπος). Να δώσετε τον σωστό τύπο δεδομένων στο πλαίσιο διαλόγου που θα εμφανιστεί:



- Ομοίως στις στήλες Εύρος (Width) και Ευθυγρ/ση (Align)
- Στη στήλη Μέτρηση (Measure) να δώσετε τη σωστή κλίμακα μέτρησης για κάθε μεταβλητή (Scale/Κλίμακα και για τις δύο).

- Στην καρτέλα Απεικ.Δεδομένων (Data View), πληκτρολογήστε τα δεδομένα της άσκησης. Θα πρέπει να φαίνεται ως εξής:

| Ικανοποίηση_Εργασία | Επαγγελματική_Πρόοδος |
|---------------------|-----------------------|
| 74                  | 56                    |
| 57                  | 58                    |
| 45                  | 43                    |
| 81                  | 75                    |
| 56                  | 61                    |
| 62                  | 60                    |
| 33                  | 42                    |
| 83                  | 93                    |
| 72                  | 59                    |
| 53                  | 64                    |



- Αποθηκεύστε το αρχείο.

### 1.3 Επεξεργασία δεδομένων

#### 1.3.1 Προϋποθέσεις

Η ανάλυση παλινδρόμησης προϋποθέτει δύο συνθήκες:

- οι μεταβλητές να συσχετίζονται σε ικανοποιητικό βαθμό και
- η συσχέτιση να είναι ευθύγραμμη.

#### 1.3.2 Διατύπωση υπόθεσης

Πριν προβούμε σε εκτίμηση (πρόβλεψη), πρέπει να εξετάσουμε εάν οι μεταβλητές συσχετίζονται. Εάν δεν συσχετίζονται, δεν έχει νόημα καμμία πρόβλεψη. Οι υποθέσεις είναι:

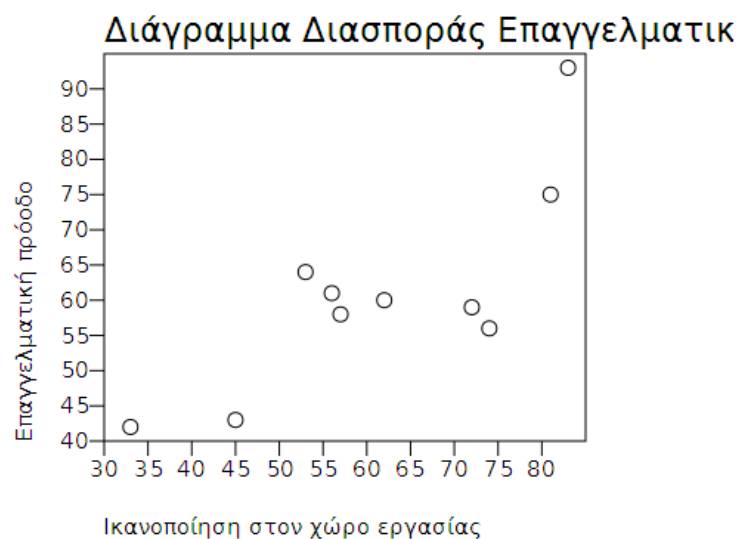
- Μηδενική υπόθεση  $H_0$ : ΔΕΝ υπάρχει γραμμική σχέση ανάμεσα στη μεταβλητή κριτήριο (εξαρτημένη - Επαγγελματική πρόοδος) και στην προβλεπτική μεταβλητή (ανεξάρτητη - Ικανοποίηση στον χώρο Εργασίας)
- Εναλλακτική υπόθεση  $H_1$ : ΥΠΑΡΧΕΙ γραμμική σχέση ανάμεσα στη μεταβλητή κριτήριο (εξαρτημένη - Επαγγελματική πρόοδος) και στην προβλεπτική μεταβλητή (ανεξάρτητη - Ικανοποίηση στον χώρο Εργασίας)

Υπόθεση διπλής κατεύθυνσης

#### 1.3.3 Υπολογισμός του συντελεστή συσχέτισης Pearson

Εξετάζουμε εάν ισχύουν οι προϋποθέσεις

- ✎ Γραφήματα (Graphs) → Διάγραμμα Διασποράς (Scater Plot) → Y Άξονας (Y Axis): Εξαρτημένη μεταβλητή → X Άξονας (X Axis): Ανεξάρτητη μεταβλητή → OK
- ✎ Ανάλυση (Analyze) → Διμεταβλητή Συχέτιση (Correlate Bivariate) → Εξαρτημένη και Ανεξάρτητη μεταβλητή (Variables) ) → Τεστ Στ. Σημαντικότητας 2πλής Κατ/νσης → Επιλογή Διακρ. Στ. Σημαντικότητας συσχετίσεων → OK



Μπορούμε να θεωρήσουμε τη συσχέτιση ευθύγραμμη.

### Συσχετίσεις

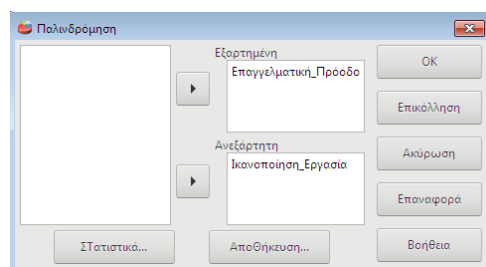
|                                |                      | Ικανοποίηση στον χώρο εργασίας | Επαγγελματική πρόοδος |
|--------------------------------|----------------------|--------------------------------|-----------------------|
| Ικανοποίηση στον χώρο εργασίας | Pearson Συσχέτιση    | 1,000                          | ,801 <sup>a</sup>     |
|                                | Στ.Σημ. (2-κατ/νσης) |                                | ,005                  |
|                                | N                    | 10                             | 10                    |
| Επαγγελματική πρόοδος          | Pearson Συσχέτιση    | ,801 <sup>a</sup>              | 1,000                 |
|                                | Στ.Σημ. (2-κατ/νσης) | ,005                           |                       |
|                                | N                    | 10                             | 10                    |

a. Σημαντικό σε επίπεδο .05

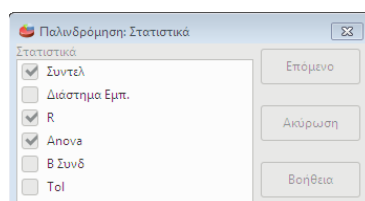
Υπάρχει υψηλή θετική συσχέτιση των δύο μεταβλητών:  $r(8) = .801$ ,  $p = .005$ .

### 1.3.4 Ανάλυση απλής παλινδρόμησης

➤ Ανάλυση (Analyze) → Παλινδρόμηση (Regression) → Γραμμική (Linear) →



➤ Στατιστικά → Επόμενο



➤ OK

Θα εμφανιστούν οι πίνακες:

#### Σύνοψη Μοντέλου (Επαγγελματική πρόοδος)

| R   | R Τετράγωνο | Προσ. R Τετράγωνο | Τυπικό Σφάλμα Εκτίμησης |
|-----|-------------|-------------------|-------------------------|
| ,80 | ,64         | ,60               | 9,36                    |

#### ANOVA (Επαγγελματική πρόοδος)

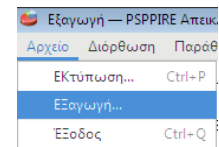
|              | Άθρ. Τετραγώνων | ΒΕ | ΜΟ Τετραγώνων | F     | Στ.Σημ. |
|--------------|-----------------|----|---------------|-------|---------|
| Παλινδρόμηση | 1252,62         | 1  | 1252,62       | 14,31 | ,005    |
| Σφάλμα       | 700,28          | 8  | 87,53         |       |         |
| Σύνολο       | 1952,90         | 9  |               |       |         |

#### Συντελεστές (Επαγγελματική πρόοδος)

|                                | Αστάθμιτοι Συντελεστές |             | Σταθμισμένοι Συντελεστές |  | t    | Στ.Σημ. |
|--------------------------------|------------------------|-------------|--------------------------|--|------|---------|
|                                | B                      | Τυπ. Σφάλμα | β                        |  |      |         |
| (Σταθερά)                      | 15,80                  | 12,33       | ,00                      |  | 1,28 | ,232    |
| Ικανοποίηση στον χώρο εργασίας | ,74                    | ,19         | ,80                      |  | 3,78 | ,005    |



Αποθηκεύστε το αρχείο εξόδου στον φάκελο εργασίας με το όνομα “Στατιστική & Μηχανική Μάθηση -Ανάλυση παλινδρόμησης\_α.pdf”.



#### 1.4 Ερμηνεία των αποτελεσμάτων

- Ο πρώτος πίνακας ονομάζεται Σύνοψη Μοντέλου. Ο συντελεστής  $R^2$  είναι 0,60 και ελέγχει τη συνολική προσαρμογή του μοντέλου. Αυτό σημαίνει ότι το 60% της συνολικής διακύμανσης της επαγγελματικής προόδου οφείλεται στην ικανοποίηση εντός του χώρου εργασίας. Το τυπικό σφάλμα εκτίμησης είναι μία ένδειξη για το πόσο επιτυχημένο είναι το μοντέλο πρόβλεψής μας. Όσο πιο μικρό είναι, τόσο περισσότερο επιτυχημένη είναι η πρόβλεψή μας.
- Ο δεύτερος πίνακας δίνει μία ανάλυση της διακύμανσης (ANOVA), η οποία εξετάζει τη μηδενική υπόθεση. Το αποτέλεσμα είναι Στ. Σημ. (Sign.)=.005 το οποίο είναι μικρότερο από  $\alpha=0,05$ . Άρα το αποτέλεσμα είναι στατιστικά σημαντικό, οπότε απορρίπτουμε τη μηδενική υπόθεση, η οποία ισχυρίζεται ότι δεν υπάρχει γραμμική σχέση των δύο μεταβλητών.

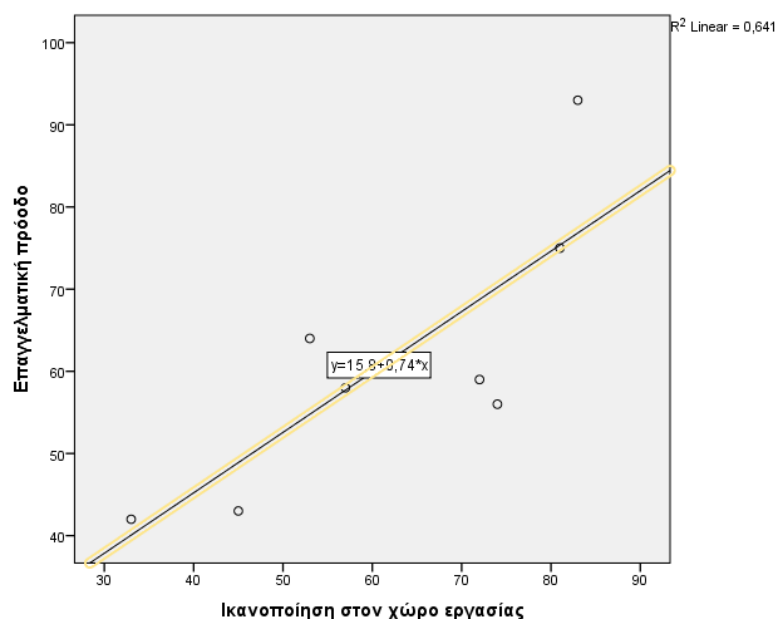
Το ANOVA λειτουργεί και ως δικλείδα ασφαλείας. Εάν βρίσκαμε Στ. Σημ.  $> 0,05$ , τότε θα έπρεπε να δεχτούμε τη μηδενική, οπότε εκεί θα σταματάγαμε τις εργασίες, αφού δεν θα μπορούσε να γίνει κάποια πρόβλεψη.

- Από τον τρίτο πίνακα Συντελεστές, παίρνουμε πληροφορίες για τη συσχέτιση, δηλαδή για το προβλεπτικό μας μοντέλο. Στην πραγματικότητα αναζητούμε την εξίσωση της ευθείας  $y=a+bx$  η οποία περνάει πιο κοντά από τα σημεία του διαγράμματος διασποράς. αντικαθιστώντας από τον πίνακα, έχουμε:

$$y=15,8+0,74*x \quad \text{ή}$$

$$\text{Επαγγελματική πρόοδος} = 15,8 + 0,74 * \text{Ικανοποίηση στην εργασία}$$

Μάλιστα, εάν θα θέλαμε να σχεδιάσουμε τη γραμμική σχέση, δηλαδή τη γραμμή παλινδρόμησης, αυτή θα ήταν:



Το  $a$  (Σταθερά) είναι η τιμή της Επαγγελματικής προόδου όταν η Ικανοποίηση στον χώρο εργασίας είναι μηδέν. Θετική τιμή σημαίνει ότι η πρόοδος είναι καλή. Το  $b$ , ο συντελεστής της σχέσης, μας πληροφορεί για το μέγεθος της πρόβλεψης καθώς και για το είδος της σχέσης (θετική ή αρνητική). Έτσι, υπάρχει θετική συσχέτιση μεταξύ των δύο μεταβλητών, και κάθε φορά που η ικανοποίηση στην εργασία αυξάνεται κατά μία μονάδα, η επαγγελματική πρόοδος θα αυξάνεται κατά 0,74 μονάδες. Επίσης, ο δείκτης  $\beta$  έχει Στ. Σημ. (Sig.)=.005 το οποίο είναι στατιστικά σημαντικό ( $\alpha=0.05$ ), που σημαίνει ότι η μεταβλητή η (μεταβλητή) Ικανοποίηση στην Εργασία δύναται όντως να αποτελέσει προβλεπτικό παράγοντα για την (μεταβλητή) Επαγγελματική Πρόοδο.

Ας επιχειρήσουμε μια πρόβλεψη. Εάν στην κλίμακα ικανοποίησης στον χώρο εργασίας κάποιος απαντήσει 100, σημαίνει ότι η επαγγελματική του πρόοδος θα είναι  $y=15,8+0,74*100=15,8+74=89,8$ . Η τιμή αυτή είναι η καλύτερη δυνατή πρόβλεψη που είμαστε σε θέση να επιτύχουμε με βάση τη μελέτη μας και δεν σημαίνει απαραίτητα ότι κάποιος με ικανοποίηση εργασίας 100 θα επιτύχει επαγγελματική πρόοδο 89,8. Και αυτό επειδή υπάρχει το τυπικό σφάλμα εκτίμησης.

Για να κατανοήσουμε το τυπικό σφάλμα εκτίμησης, ας επιχειρήσουμε μία πρόβλεψη για  $x=45$ . Σε αυτήν την περίπτωση θα πάρουμε  $y=15,8+0,74*45=48,3$ . Όπως βλέπουμε όμως από τα δεδομένα, το τρίτο άτομο έχει δώσει επαγγελματική πρόοδο 43. Δηλαδή, η πρόβλεψη που κάναμε περιέχει ένα σφάλμα  $43-48,3=-5,3$  βαθμούς. Το σφάλμα αυτό είναι εντός των ορίων της τιμής του τυπικού σφάλματος εκτίμησης ((Std. Error of the Estimate=9,36). Το αρνητικό πρόσημο δηλώνει ότι η πραγματική τιμή είναι μικρότερη από την προβλεπόμενη.

- Το αποτέλεσμα διατυπώνεται ως εξής (APA Style 7<sup>th</sup> edition, σ. 180):
  - ✓ «Προκειμένου να διερευνήσουμε εάν υπάρχει συσχέτιση μεταξύ της ικανοποίησης των εργαζομένων στον χώρο εργασίας των αρχείων και της επαγγελματικής προόδου τους διεξήχθη ανάλυση απλής παλινδρόμησης. Το διάγραμμα διασποράς έδειξε ότι υπάρχει ισχυρή γραμμική συσχέτιση  $r(8)=.801$ ,  $p=.005$ . Το αποτέλεσμα της απλής γραμμικής παλινδρόμησης έδειξε ότι υπάρχει στατιστικώς σημαντική σχέση μεταξύ ικανοποίησης και προόδου,  $p=.005$ . Ο συντελεστής  $R^2$  είναι 0,60, που σημαίνει ότι το 60% της συνολικής διακύμανσης της επαγγελματικής προόδου οφείλεται στην ικανοποίηση εντός του χώρου εργασίας.»

## 2. Άσκηση β – Ανάλυση πολλαπλής παλινδρόμησης

### 2.1 Δεδομένα

Στην προηγούμενη έρευνα, υποθέστε ότι ερευνούσατε έναν επιπλέον παράγοντα, την εκπαίδευση. Πιο συγκεκριμένα, εάν γνωρίζετε την ικανοποίηση των εργαζομένων ενός αρχείου στον χώρο τους και τον βαθμό του πτυχίου τους, θα μπορούσατε να προβλέψετε την επαγγελματική τους πρόοδο;

Για τις ανάγκες της έρευνας, συμπληρώθηκαν ερωτηματολόγια και προέκυψαν τα παρακάτω δεδομένα:

| Εργαζόμενος | Ικανοποίηση<br>στον χώρο<br>εργασίας | Εκπαίδευση<br>(βαθμός στο<br>πτυχίο σε<br>κλίμακα 100) | Επαγγελματική<br>πρόοδο |
|-------------|--------------------------------------|--------------------------------------------------------|-------------------------|
| 1           | 74                                   | 78                                                     | 56                      |
| 2           | 57                                   | 60                                                     | 58                      |
| 3           | 45                                   | 48                                                     | 43                      |
| 4           | 81                                   | 83                                                     | 75                      |
| 5           | 56                                   | 57                                                     | 61                      |
| 6           | 62                                   | 67                                                     | 60                      |
| 7           | 33                                   | 32                                                     | 42                      |
| 8           | 83                                   | 83                                                     | 93                      |
| 9           | 72                                   | 67                                                     | 59                      |
| 10          | 53                                   | 56                                                     | 64                      |

### 2.2 Εισαγωγή δεδομένων



Στον φάκελο Στατιστική να αντιγράψετε και να επικολλήσετε το αρχείο "Ανάλυση παλινδρόμησης\_α.sav". Στη συνέχεια να μετονομάσετε το αρχείο σε "Ανάλυση παλινδρόμησης\_β.sav". Να προσθέσετε τη μεταβλητή "Εκπαίδευση" και να εισάγετε τα δεδομένα.

## 2.3 Επεξεργασία δεδομένων

Επειδή η πρόβλεψη χρησιμοποιεί περισσότερες από μία μεταβλητές, κατάλληλη είναι η ανάλυση πολλαπλής παλινδρόμησης (Multiple regression).

### 2.3.1 Προϋποθέσεις

Η πολλαπλή ανάλυση παλινδρόμησης προϋποθέτει τις συνθήκες της απλής ανάλυσης παλινδρόμησης:

- οι μεταβλητές να συσχετίζονται σε ικανοποιητικό βαθμό,
- η συσχέτιση να είναι ευθύγραμμη και
- να μην υπάρχει συγγραμμικότητα.

### 2.3.2 Διατύπωση υπόθεσης

Πριν προβούμε σε εκτίμηση (πρόβλεψη), πρέπει να εξετάσουμε εάν η προβλεπτική μεταβλητή σχετίζεται με τις υπόλοιπες. Εάν δεν συσχετίζονται, δεν έχει νόημα καμμία πρόβλεψη.

- Μηδενική υπόθεση  $H_0$ : ΔΕΝ υπάρχει γραμμική σχέση ανάμεσα στη μεταβλητή κριτήριο (εξαρτημένη - Επαγγελματική πρόοδος) και στις προβλεπτικές μεταβλητές (ανεξάρτητες – Ικανοποίηση στον χώρο Εργασίας & Εκπαίδευση)
- Εναλλακτική υπόθεση  $H_1$ : ΥΠΑΡΧΕΙ γραμμική σχέση ανάμεσα στη μεταβλητή κριτήριο (εξαρτημένη - Επαγγελματική πρόοδος) και στις προβλεπτικές μεταβλητές (ανεξάρτητες – Ικανοποίηση στον χώρο Εργασίας & Εκπαίδευση)

Υπόθεση διπλής κατεύθυνσης

### 2.3.3 Υπολογισμός του συντελεστή συσχέτισης Pearson

Εξετάζουμε εάν ισχύουν οι δύο πρώτες προϋποθέσεις.

- ✎ Εξετάζουμε τον δείκτη συσχέτισης Pearson των μεταβλητών Ικανοποίηση Εργασίας και Επαγγελματική πρόοδο (βλ. 1.3.3).
- ✎ Επίσης τον δείκτη συσχέτισης Pearson της Εκπαίδευση και την Επαγγελματική πρόοδο. Εφαρμόζοντας περιγραφικά στατιστικά έχουμε:

| Ανάλυση                | Γραφήματα | Βοηθήματα | Παράθυρα       | Βοήθεια |
|------------------------|-----------|-----------|----------------|---------|
| Περιγραφικά Στατιστικά |           |           | Συχνότητες...  |         |
| Σύγκριση Μέσων         |           |           | Περιγραφικά... |         |

Περιγραφικά Στατιστικά

|                                 | N  | Μέσος Όρος | ΤΑ    | Ελάχιστο | Μέγιστο |
|---------------------------------|----|------------|-------|----------|---------|
| Ικανοποίηση στον χώρο εργασίας  | 10 | 61,60      | 16,04 | 33       | 83      |
| Εκπαίδευση (βαθμός πτυχίου 100) | 10 | 63,10      | 16,11 | 32       | 83      |
| Επαγγελματική πρόοδο            | 10 | 61,10      | 14,73 | 42       | 93      |
| Έγκυρο N (ανά γραμμή)           | 11 |            |       |          |         |
| Ελλειπούσες N (ανά υπόθεση)     | 1  |            |       |          |         |

Εκ πρώτης όψεως, μπορούμε να διακρίνουμε ίδιους ή παρόμοιους δείκτες.

Ζητώντας διάγραμμα διασποράς και συντελεστή συσχέτισης, έχουμε:



**Συσχετίσεις**

|                                 |                      | Εκπαίδευση (βαθμός πτυχίου 100) | Επαγγελματική πρόοδος |
|---------------------------------|----------------------|---------------------------------|-----------------------|
| Εκπαίδευση (βαθμός πτυχίου 100) | Pearson Συσχέτιση    | 1,000                           | ,786 <sup>a</sup>     |
|                                 | Στ.Σημ. (2-κατ/νσης) |                                 | ,007                  |
|                                 | N                    | 10                              | 10                    |
| Επαγγελματική πρόοδος           | Pearson Συσχέτιση    | ,786 <sup>a</sup>               | 1,000                 |
|                                 | Στ.Σημ. (2-κατ/νσης) | ,007                            |                       |
|                                 | N                    | 10                              | 10                    |

a. Σημαντικό σε επίπεδο .05

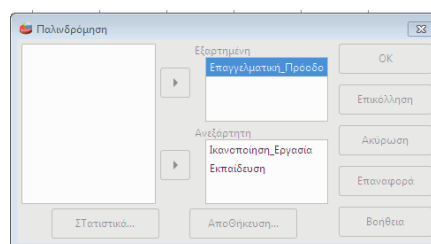
Όπως φαίνεται από το διάγραμμα διασποράς η συσχέτιση φαίνεται να είναι ευθύγραμμη. Από τον πίνακα Συσχετίσεις, υπάρχει υψηλή θετική συσχέτιση των μεταβλητών:  $r(8)=.786$ ,  $p=.007$ .

Επίσης, για τον δείκτη συσχέτισης Pearson της μεταβλητής Ικανοποίηση Εργασίας και της Επαγγελματική πρόοδος έχουμε:  $r(8)=.801$ ,  $p=.005$  (βλ. 1.3.3).

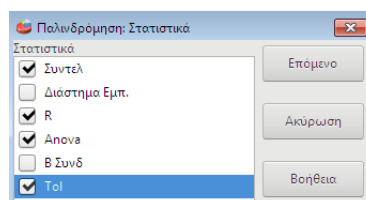
Υπάρχει, λοιπόν, συσχέτιση μεταξύ των προβλεπτικών μεταβλητών και της μεταβλητής κριτηρίου.

### 2.3.4 Ανάλυση πολλαπλής παλινδρόμησης

➤ Ανάλυση (Analyze) → Παλινδρόμηση (Regression) → Γραμμική (Linear) →



➤ Στατιστικά



➤ → Επόμενο → OK

Με αποτελέσματα:

**Σύνοψη Μοντέλου (Επαγγελματική πρόοδος)**

| R   | R Τετράγωνο | Προσ. R Τετράγωνο | Τυπικό Σφάλμα Εκτίμησης |
|-----|-------------|-------------------|-------------------------|
| ,80 | ,64         | ,54               | 10,00                   |

**ANOVA (Επαγγελματική πρόοδος)**

|              | Άθρ. Τετραγώνων | ΒΕ | ΜΟ Τετραγώνου | F    | Στ.Σημ. |
|--------------|-----------------|----|---------------|------|---------|
| Παλινδρόμηση | 1252,73         | 2  | 626,37        | 6,26 | ,028    |
| Σφάλμα       | 700,17          | 7  | 100,02        |      |         |
| Σύνολο       | 1952,90         | 9  |               |      |         |

**Συντελεστές (Επαγγελματική πρόοδος)**

|                                 | Αστάθμιτοι Συντελεστές |             | Σταθμισμένοι Συντελεστές |      | t    | Στ.Σημ. | Στατιστικά Συγγραμμικότητας |       |
|---------------------------------|------------------------|-------------|--------------------------|------|------|---------|-----------------------------|-------|
|                                 | B                      | Τυπ. Σφάλμα | β                        |      |      |         | Tolerance                   | VIF   |
| (Σταθερά)                       | 15,89                  | 13,43       |                          | ,00  | 1,18 | ,271    |                             |       |
| Ικανοποίηση στον χώρο εργασίας  | ,77                    | 1,15        |                          | ,84  | ,67  | ,523    | ,03                         | 30,68 |
| Εκπαίδευση (βαθμός πτυχίου 100) | -,04                   | 1,15        |                          | -,04 | -,03 | ,974    | ,03                         | 30,68 |



Αποθηκεύστε το αρχείο εξόδου στον φάκελο εργασίας Στατιστική με το όνομα “Ανάλυση παλινδρόμησης\_β.pdf”.

## 2.4 Ερμηνεία αποτελεσμάτων

- Ελέγχοντας τη συνολική προσαρμογή του μοντέλου, ο συντελεστής  $R^2$  είναι 0,64. Αυτό σημαίνει ότι το 64% της συνολικής διακύμανσης της επαγγελματικής πρόοδου οφείλεται στην ικανοποίηση εντός του χώρου εργασίας.
- Ο δεύτερος πίνακας δίνει μία ανάλυση της διακύμανσης (ANOVA), η οποία εξετάζει τη μηδενική υπόθεση. Το αποτέλεσμα είναι Στ. Σημ. (Sign.)=.028 το οποίο είναι μικρότερο από  $\alpha=0,05$ . Άρα το αποτέλεσμα είναι στατιστικά σημαντικό, οπότε απορρίπτουμε τη μηδενική υπόθεση, η οποία ισχυρίζεται ότι δεν υπάρχει γραμμική σχέση των μεταβλητών.

$$F(2, 7)=6,26, p=.028$$

- ✓ Εναλλακτικά, από τον πίνακα κρίσιμων τιμών, βρίσκουμε ότι η τιμή  $F=6,26$  για 2 και 7 βαθμούς ελευθερίας και για επίπεδο στατιστικής σημαντικότητας  $\alpha=0,05$  είναι μεγαλύτερη από την αντίστοιχη κρίσιμη τιμή του πίνακα (4,74). Άρα απορρίπτουμε τη μηδενική και δεχόμαστε την εναλλακτική.

Συμπεραίνουμε ότι, σε σχέση με τη μη εφαρμογή του μοντέλου, η εφαρμογή του βελτίωσε σημαντικά την ικανότητά μας να προβλέψουμε την επαγγελματική πρόοδο στον χώρο των αρχείων (μεταβλητή κριτήριο).

- Από τον τρίτο πίνακα Συντελεστές, παίρνουμε πληροφορίες για τη συσχέτιση, δηλαδή για το προβλεπτικό μας μοντέλο. Στην πραγματικότητα αναζητούμε την εξίσωση της ευθείας  $y=\alpha+\beta_1x_1+\beta_2x_2$  η οποία περνάει πιο κοντά από τα σημεία του διαγράμματος διασποράς. αντικαθιστώντας από τον πίνακα, έχουμε:

$$y = 15,89 + 0,77 \cdot x_1 + (-0,04) \cdot x_2 \quad \text{ή}$$

$$\text{Επαγγελματική πρόοδος} = 15,89 + 0,77 \cdot \text{Ικανοποίηση στην εργασία} + (-0,04) \cdot \text{Εκπαίδευση βαθμός πτυχίου}$$

Για κάθε μεταβολή στην ικανοποίηση στον χώρο εργασίας κατά μία μονάδα, η προβλεπόμενη επαγγελματική πρόοδος θα αυξηθεί κατά 0,77 μονάδες, με την προϋπόθεση ότι η εκπαίδευση θα παραμείνει σταθερή (ίδιος βαθμός στο πτυχίο).

- Ποιος ο βαθμός σπουδαιότητας της κάθε προβλεπτικής μεταβλητής; Η απάντηση προκύπτει από τους συντελεστές παλινδρόμησης (δηλαδή, τις αντίστοιχες κλίσεις των ευθειών  $\beta_1$  και  $\beta_2$  για κάθε προβλεπτική μεταβλητή). Στο παράδειγμά μας, σημαντικότερο ρόλο παίζει η ικανοποίηση στον χώρο εργασίας των αρχείων, με συντελεστή 0,77, ενώ ο βαθμός του πτυχίου φαίνεται να έχει ελαφρώς αρνητική επιρροή (-0,04).

Η σύγκριση των συντελεστών παλινδρόμησης δεν αποτελεί ασφαλής ένδειξη για την σπουδαιότητα της κάθε προβλεπτικής μεταβλητής, επειδή επηρεάζεται από τη διακύμανση της κάθε μεταβλητής. Για να είναι ασφαλής, πρέπει να μετατρέψουμε τους συντελεστές παλινδρόμησης σε *τυποποιημένους (σταθμισμένους) συντελεστές παλινδρόμησης b (beta)*. Σε αυτήν την περίπτωση, οι διακυμάνσεις των τυποποιημένων συντελεστών είναι ίσες (ίδιος μέσος όρος (0) και τυπική απόκλιση (1)). Από την αντίστοιχη στήλη, βλέπουμε ότι  $\beta_1=0,84$  και  $\beta_2=-0,04$ . Δηλαδή, η ικανοποίηση στον χώρο εργασίας παίζει τον σημαντικότερο ρόλο στην επαγγελματική επίδοση.

- Το φαινόμενο της συγγραμμικότητας (collinearity) ή πολυσκεδαστικότητας (multicollinearity) είναι εκείνο το οποίο υπάρχει υψηλή συσχέτιση των προβλεπτικών μεταβλητών, με συνέπεια η ανάλυση πολλαπλής παλινδρόμησης (δηλαδή η εξίσωση παλινδρόμησης) να μη δίνει αξιόπιστα αποτελέσματα. Πιο συγκεκριμένα, εάν οι προβλεπτικές μεταβλητές εμφανίζουν υψηλή συσχέτιση, ενδέχεται δύο διαφορετικά δείγματα του ίδιου πληθυσμού να δώσουν τελείως διαφορετικά προβλεπτικά μοντέλα (εξισώσεις παλινδρόμησης).

Στην τελευταία στήλη του πίνακα συντελεστές, θα πρέπει ο δείκτης ανοχής (Tolerance) να είναι μεγαλύτερος από 0,1 και ο παράγοντας πληθωριστικής διασποράς (Variance Inflation Factor –VIF) να είναι μικρότερος από 10. Επειδή δεν συναληθεύουν και οι δύο συνθήκες, δεν μπορούμε να εμπιστευτούμε το προβλεπτικό μοντέλο:  $y = 15,89 + 0,77 \cdot x_1 + (-0,04) \cdot x_2$ .

Ένας πρώτος οπτικός τρόπος εκτίμησης εμφάνισης της συγγραμμικότητας είναι εάν κάποιος δείκτης έχει πολύ υψηλή συσχέτιση (μεγαλύτερη 0,8).



Δοκιμάστε να εισάγετε τους εξής βαθμούς πτυχίου: 88, 89, 67, 78, 81, 67, 99, 98, 67, 56. Μπορείτε να λάβετε προβλεπτικό μοντέλο;

### 3. Ασκήσεις κατανόησης

#### 3.1 Δεδομένα

Στον παρακάτω πίνακα φαίνονται τα δεδομένα που προέκυψαν από μία έρευνα για τον καρκίνο του πνεύμονα σε 21 χώρες της Ευρωπαϊκής Ένωσης. Στην στήλη 2 βρίσκεται ο αριθμός των τσιγάρων ανά άτομο την ημέρα και στη στήλη τρία ο αριθμός των θανάτων από καρκίνο του πνεύμονα ανά 10.000 κατοίκους.

| Χώρες | Αριθμός<br>τσιγάρων | ΚΙΠΝ |
|-------|---------------------|------|
| 1     | 11                  | 26   |
| 2     | 9                   | 21   |
| 3     | 9                   | 24   |
| 4     | 9                   | 21   |
| 5     | 8                   | 19   |
| 6     | 8                   | 13   |
| 7     | 8                   | 19   |
| 8     | 6                   | 11   |
| 9     | 6                   | 23   |
| 10    | 5                   | 15   |
| 11    | 5                   | 13   |
| 12    | 5                   | 4    |
| 13    | 5                   | 18   |
| 14    | 5                   | 12   |
| 15    | 5                   | 3    |
| 16    | 4                   | 11   |
| 17    | 4                   | 15   |
| 18    | 4                   | 6    |
| 19    | 3                   | 13   |
| 20    | 3                   | 4    |
| 21    | 3                   | 14   |

Ζητείται να κάνετε πρόβλεψη για τον αριθμό των ατόμων που θα πεθάνουν από καρκίνο του πνεύμονα, όταν ο αριθμός των τσιγάρων που καπνίζουν σε μια χώρα κατά μέσο όρο θα είναι 6.

#### 4. Ορολογία



Δημιουργήστε αρχείο κειμένου (π.χ. Word) με το όνομα “Ανάλυση παλινδρόμησης\_Ορολογία”

Δημιουργήστε τον πίνακα και αντιστοιχίστε:

Regression analysis, Criterion variable, Equation of the straight line, Multiple regression, Regression coefficients, Predictive variable, Logistic regression, Linear regression, Error of estimate

|                           |  |
|---------------------------|--|
| Ανάλυση παλινδρόμησης     |  |
| Προβλεπτική μεταβλητή     |  |
| Μεταβλητή κριτήριο        |  |
| Πολλαπλή παλινδρόμηση     |  |
| Γραμμική παλινδρόμηση     |  |
| Εξίσωση ευθείας γραμμής   |  |
| Συντελεστής παλινδρόμησης |  |
| Σφάλμα εκτίμησης          |  |
| Λογιστική παλινδρόμηση    |  |



Αποθηκεύστε το αρχείο.

## 5. Βιβλιογραφία

- Ρούσσος, Π. & Τσαούσης, Γ. (2020). *Στατιστική εφαρμοσμένη στις κοινωνικές επιστήμες με τη χρήση του SPSS και του R*. Αθήνα: Gutenberg. <http://www.socialstats.gr/index.html>
- Field, A. (2016). *Η διερεύνηση της στατιστικής με τη χρήση του SPSS*. Αθήνα: Προπομπός.
- jamovi (n.d.). *jamovi Stats.Open.Now*. <https://www.jamovi.org/>
- Zaiontz, S. (n.d.). *Real statistics using Excel*. <https://www.real-statistics.com/>
- Smith, T. J., & McKenna, C. M. (2013). A comparison of logistic regression pseudo R<sup>2</sup> indices. *Multiple Linear Regression Viewpoints*, 39(2), 17-26.  
[http://www.glmj.org/archives/articles/Smith\\_v39n2.pdf](http://www.glmj.org/archives/articles/Smith_v39n2.pdf)

## 6. Σημειώματα

### Σημείωμα ιστορικού εκδόσεων έργου

Το παρόν έργο αποτελεί την έκδοση 2.0

Έχουν προηγηθεί οι κάτωθι εκδόσεις:

- Σ. Τζαναβάρης. «Εργαλεία Στατιστικής στην Επιστήμη της Πληροφορίας». Έκδοση: 1.0. Κέρκυρα, 2018.

### Σημείωμα αναφοράς

Copyright Σ. Τζαναβάρης. «Στατιστική». Έκδοση: 2.0. Κέρκυρα, 2021.

### Σημείωμα χρήσης έργων τρίτων

#### Διατήρηση σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

### Σημείωμα αδειοδότησης

Σημείωμα Αδειοδότησης Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως Μη Εμπορική ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.