# Εκπαιδευτικό σεμινάριο

**Εισαγωγή στην επιστήμη δεδομένων με τη χρήση της R**

## 1. Σκοπός

To πρόγραμμα σκοπεύει να προσφέρει βασικές γνώσεις και δεξιότητες σε εργαλεία και τεχνικές της επιστήμης δεδομένων (Data Science) όπως Εισαγωγή δεδομένων σε πρόγραμμα στατιστικής επεξεργασίας, Επεξεργασία δεδομένων και στατιστική ανάλυση δεδομένων μέσω στατιστικών ομαδοποιήσεων, γραφικές παραστάσεις, και επικοινωνία αποτελεσμάτων, μέσα από τη γλώσσα στατιστικής επιστήμης R.

Με το τέλος του προγράμματος οι συμμετέχοντες θα είναι σε θέση να:

1. Δημιουργούν ομαδοποιημένα στατιστικά στοιχεία για τα δεδομένα (grouped summaries)
2. Εισάγουν δεδομένα μορφής .csv, .xls, .sav στην R
3. Εισάγουν δεδομένα από διαδικτυακές πηγές, στο πρόγραμμα της R
4. Καθαρίζουν δεδομένα (data cleaning)
5. Επιλέγουν το κατάλληλο είδος γραφικής απεικόνισης για τις μεταβλητές που θα αναπαραστήσουν
6. Δημιουργούν γραφικές παραστάσεις απεικόνισης ομαδοποιημένων στοιχείων
7. Δημιουργούν περιληπτικούς πίνακες για απεικόνιση βασικών στοιχείων των δεδομένων
8. Δημιουργούν αυτοματοποιημένα reports σε μορφή PDF, HTML και MS WORD

## 2. Ενότητες εκπαιδευτικού προγράμματος

### **Ενότητα Α: Γνωριμία με το περιβάλλον της R και το RStudio. Βασικές εντολές της R (60 λεπτά).**

Περιδιάβαση στο περιβάλλον της R, το περιβάλλον RStudio, και βασικές εντολές για αρχική εξοικείωση με τη γλώσσα προγραμματισμού R.

### **Ενότητα Β: Εισαγωγή Δεδομένων Excel, CSV, TXT στην R (60 λεπτά)**

Τα δεδομένα μπορεί να είναι σε διάφορες μορφές. Η ενότητα αυτή επικεντρώνεται στο να μάθουν οι συμμετέχοντες τα βασικά εργαλεία όπως το πακέτο *readr*, και *readxl*, για να μπορούν εύκολα, γρήγορα και αποτελεσματικά να εισάγουν δεδομένα στην R.

### **Eνότητα Γ: Επεξεργασία δεδομένων (120 λεπτά)**

Η ενότητα αυτή έχει ως κύριο στόχο να παρέχει στους συμμετέχοντες τις γνώσεις και δεξιότητες για μετασχηματισμό των δεδομένων (data wrangling) και στατιστική ανάλυση με το πακέτο *dplyr* και *tidyr* στην R.

### **Ενότητα Δ: Γραφική αναπαράσταση δεδομένων (Data visualisation) (90 λεπτά)**

Η γραφική παράσταση θα πρέπει κυρίως να αντιπροσωπεύει τα δεδομένα. Πρέπει να δίνει στον ίδιο τον αναλυτή χρήσιμες πληροφορίες για να μπορέσει είτε να συνεχίσει την διερεύνηση των δεδομένων, ή να καταλήξει σε ένα μήνυμα που θα μεταδώσει.

### **Ενότητα Ε: Αυτοματοποιημένα reports σε μορφή PDF, HTML και MS WORD (90 λεπτά)**

Στην ενότητα αυτή οι συμμετέχοντες θα εκπαιδευτούν πώς να χρησιμοποιήσουν το πακέτο Rmarkdown για την εξαγωγή αναφορών (reports) σε διάφορες μορφές (.pdf, .doc., .html)

## 3. Διάρκεια σεμιναρίου:

Η χρονική διάρκεια του σεμιναρίου είναι συνολικά 8 ώρες [7 ωρες εκπαίδευση + 1 ώρα διάλειμμα ανάμεσα στις ενότητες]

# Ενδεικτικό κόστος σεμιναρίου

1,400€ + VAT

Μέγιστος αριθμός συμμετεχόντων 8 άτομα

# Τόπος σεμιναρίου

Το σεμινάριο θα διεξαχθεί στο χώρο του πελάτη.

# Γιατί με την R

-Αναπαραγωγίσιμη Ανάλυση Δεδομένων (reproducible analysis)

Δυνατότητα αναπαραγωγής της ανάλυσης. Κωδικοποιήση της ανάλυσης – Πίνακες, γραφήματα, ακόμη και κείμενο και, στη συνέχεια, εκτέλεση της ανάλυσης ξανά με ένα κλικ, με νέα δεδομένα, π.χ. κάθε μέρα, εβδομάδα, μήνα κ.λπ.

-Μαθησιακοί πόροι (Learning resources)

Η εκμάθηση και η επέκταση των γνώσεών στην R, δεν ήταν ποτέ πιο εύκολη. Υπάρχει ένας τεράστιος αριθμός δωρεάν πόρων εκμάθησης στο διαδίκτυο – βιβλία [https://www.rstudio.com/resources/books/], διαδικτυακά σεμινάρια και βίντεο [https://www.rstudio.com/resources/webinars/] Cheatsheets [ https://www.rstudio.com/resources/cheatsheets/]. Επιπλέον, υπάρχουν ειδικά φόρουμ Q & A για την R [https://community.rstudio.com/], μια ειδική κοινότητα για το R στο twitter [#Rstats https://twitter.com/search?q=%23RStats], και αποκλειστικές ομάδες χρηστών R σε 4141 πόλεις σε όλο τον κόσμο [https://benubah.github.io/r-community-explorer/rugs.html]

-Αυτοματοποιημένες αναφορές

Το R markdown είναι ένα βοηθητικό πρόγραμμα στο R όπου μπορείτε να μετατρέψετε τις αναλύσεις σας σε έγγραφα, αναφορές, παρουσιάσεις και πίνακες εργαλείων υψηλής ποιότητας για το κοινό σας. Μορφές εξόδου, όπως HTML, PDF, MS Word, διαφάνειες HTML5, βιβλία, πίνακες εργαλείων, λαμπερές εφαρμογές ή ακόμα και ιστότοποι.

-Δωρεάν

Η R είναι δωρεάν

**Εκπαιδευτής**

Λεύκιος Παίκούσης

Data scientist

<https://cy.linkedin.com/in/lefkios>

Πιστοποιημένος εκπαιδευτής επαγγελματικής κατάρτισης ΑΝΑΑΔ