

## Θέμα εργασίας

### Οδηγίες

Η εργασία θα αξιολογηθεί με άριστα τη **μία μονάδα**. Θα πρέπει:

- **Να είναι ατομική. Η συνεργασία πάνω στα ερωτήματα είναι επιθυμητή αλλά η αντιγραφή απαγορεύεται.**
- Στην αρχή (πρώτη σελίδα) να αναφέρεται το όνομα και το ΑΕΜ του φοιτητή / της φοιτήτριας.
- Να περιέχει τις απαντήσεις και λύσεις με σχόλια, όπως ζητούνται και με τη σειρά που ζητούνται.
- Τα γραφήματα και οι πίνακες αποτελεσμάτων από το SPSS θα πρέπει να παρουσιάζονται με τη σειρά που ζητούνται και σε κατάλληλα σημεία μέσα στο κείμενο της εργασίας.
- **Το κάθε γράφημα από το SPSS θα πρέπει να έχει στον τίτλο το ΑΕΜ του φοιτητή / της φοιτήτριας και ο κάθε πίνακας από το SPSS θα πρέπει να έχει στην επικεφαλίδα το ΑΕΜ του φοιτητή / της φοιτήτριας. Θα συμπεριλάβετε ΜΟΝΟ σχετικούς πίνακες και σχήματα στο SPSS που υποστηρίζουν τις απαντήσεις σας, διαφορετικά θα μετρήσει αρνητικά στο βαθμό σας.**
- Η εργασία θα πρέπει να είναι γραμμένη στον υπολογιστή σε πρόγραμμα επεξεργασίας κειμένου, όπως Word. Αν είναι γραμμένη σε άλλο πρόγραμμα επεξεργασίας κειμένου, να σταλεί σε pdf.

Η εργασία θα πρέπει να παραδοθεί ηλεκτρονικά μέσω της ιστοσελίδας του μαθήματος στο *teaching* το αργότερο ως τις **11/7/2021**. Θα πρέπει να παραδοθεί ηλεκτρονικά ΜΟΝΟ ένα αρχείο που περιέχει ολόκληρη την εργασία. Το όνομα του αρχείου θα πρέπει να είναι *Stat<ΑΕΜ>.docx* όπου *<ΑΕΜ>* είναι ο Αριθμός Ειδικού Μητρώου του/της φοιτητή/τριας που παραδίδει την εργασία (χωρίς τις ανισότητες) και *.docx* δηλώνει ότι είναι γραμμένο σε Word (άλλες επιτρεπτές μορφές είναι *.doc* και *.pdf*), π.χ. αν το ΑΕΜ είναι 9820 το αρχείο Word θα πρέπει να έχει όνομα *Stat9820.docx*, *Stat9820.doc* ή *Stat9820.pdf*. **Παρακαλώ να εξετάσετε αν το όνομα του αρχείου είναι σωστό πριν το υποβάλλετε. Αρχείο με όνομα που δεν είναι στην προβλεπόμενη μορφή μπορεί να αγνοηθεί και η εργασία να μην αξιολογηθεί!**

### Περιγραφή προβλήματος και δεδομένα

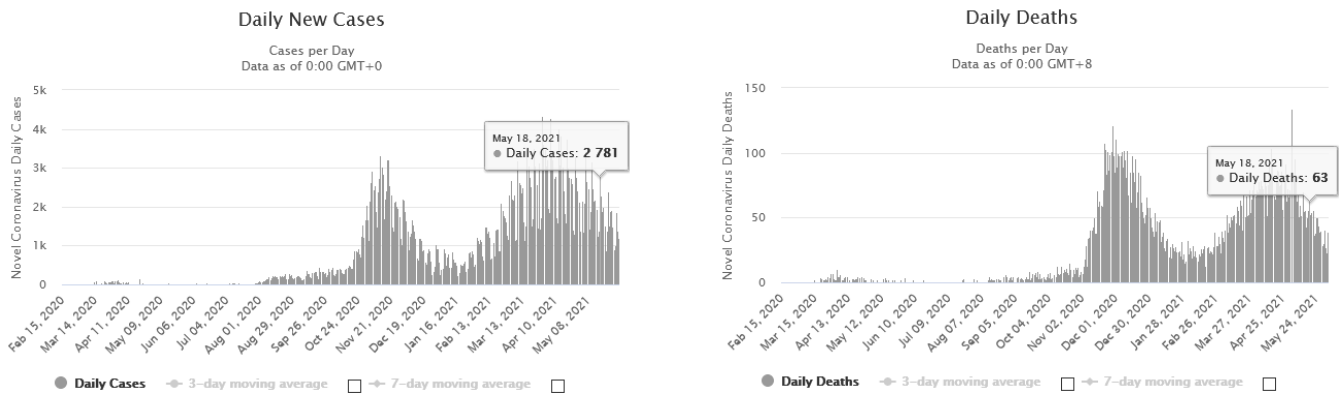
Όλοι/ες γνωρίζουμε πως η πανδημία του κορονοϊού χτύπησε πολλές χώρες του πλανήτη και όλες τις χώρες της Ευρώπης. Στην εργασία θα χρησιμοποιήσετε στοιχεία κρουσμάτων και θανάτων κορονοϊού (covid-19) σε διάφορες χώρες της Ευρώπης, όπως έχουν καταγραφεί στον ιστοχώρο <https://www.worldometers.info/coronavirus/>, που θα χρησιμοποιήσετε για να εξάγετε δεδομένα (δίνονται και εναλλακτικές λύσεις παρακάτω). Δίνεται επίσης το αρχείο σε μορφή excel *CoronavirusCountries.xlsx*, που περιέχει σε διάταξη τις 37 χώρες της Ευρώπης με πληθυσμό πάνω από 1 εκ κατοίκους που χτυπήθηκαν από τον κορονοϊό.

Θα χρησιμοποιήσετε δύο από τις 37 χώρες για την στατιστική ανάλυση σας σύμφωνα με τη διάταξη που δίνονται στο αρχείο και με βάση τα 4 τελευταία ψηφία του ΑΕΜ σας ως εξής. Η πρώτη χώρα, συμβολίζεται στο εξής ως Α, έχει αύξοντα αριθμό στο αρχείο *CoronavirusCountries.xlsx* που είναι το υπόλοιπο της διαίρεσης του διψήφιου αριθμού των δύο τελευταίων ψηφίων του ΑΕΜ σας με το 37. Η δεύτερη χώρα, συμβολίζεται στο εξής ως Β, έχει αντίστοιχα αύξοντα αριθμό στο αρχείο *CoronavirusCountries.xlsx* που είναι το υπόλοιπο της διαίρεσης του δεύτερου διψήφιου αριθμού από το ΑΕΜ σας με το 37, όπου αυτός ο διψήφιος αριθμός αποτελείται από το τέταρτο από το τέλος ψηφίο του ΑΕΜ (πρώτο ψηφίο του διψήφιου αριθμού) και το τρίτο από το τέλος ψηφίο του ΑΕΜ (δεύτερο ψηφίο του διψήφιου αριθμού). Για παράδειγμα το ΑΕΜ 9820 έχει ως Α τη Τουρκία με αύξοντα αριθμό 20, που είναι το υπόλοιπο της διαίρεσης του 20 με το 37 και ως Β τη Νορβηγία με αύξοντα αριθμό 24, που είναι το υπόλοιπο της διαίρεσης του 98 με το 37. Αν και οι δύο αύξοντες αριθμοί είναι ίδιοι, τότε ο δεύτερος αύξων αριθμός θα αυξηθεί κατά 1 (και αν είναι 37 το επόμενο είναι 1). Για παράδειγμα αν το ΑΕΜ είναι 9824 τότε ο πρώτος αύξων αριθμός είναι 24 και ο δεύτερος είναι 25, δηλαδή το υπόλοιπο της διαίρεσης του 98 με το 37 είναι 24 και +1 είναι 25.

Στο αρχείο *CoronavirusCountries.xlsx* στο όνομα της κάθε χώρας υπάρχει υπερ-σύνδεσμος που σας μεταφέρει στην ιστοσελίδα του ιστοχώρου <https://www.worldometers.info/coronavirus/> για αυτήν τη χώρα (μπορείτε φυσικά να μεταβείτε μέσα από την κεντρική σελίδα του ιστοχώρου). Εκεί θα εστιάσετε στα γραφήματα για τα ημερήσια νέα κρούσματα (Daily New Cases) και για τους ημερήσιους νέους θανάτους (Daily New Deaths).

Για την λήψη των δεδομένων που θα αναλύσετε, εναλλακτικά μπορείτε να συλλέξετε τα στοιχεία για τον αριθμό ημερήσιων νέων κρουσμάτων και τον αριθμό ημερήσιων νέων θανάτων από άλλες βάσεις δεδομένων. Αναφέρονται η βάση δεδομένων του European Centre for Disease Prevention and Control, <https://www.ecdc.europa.eu/en/publications-data/data-daily-new-cases-covid-19-eueea-country> (έχει στοιχεία από 1/3/2021), η βάση δεδομένων COVID19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, <https://github.com/CSSEGISandData/COVID-19>, η βάση δεδομένων Our World in Data, <https://github.com/owid/covid-19-data/tree/master/public/data>.

Στα παρακάτω σχήματα δίνεται η εικόνα δύο γραφημάτων (στις 4/6/2021) για την Ελλάδα από τον ιστοχώρο <https://www.worldometers.info/coronavirus/>.



Σε κάθε ένα από τα δύο παραπάνω σχήματα φαίνεται σε ένθετο κάποια πληροφορία και κυρίως η τιμή του αντίστοιχου δείκτη για κάποια μέρα που δηλώνεται με το ύψος της αντίστοιχης μπάρας (για τις 18 Μαΐου στα δύο σχήματα). Μπορείτε εύκολα να δείτε την κάθε τιμή τοποθετώντας το δείκτη του υπολογιστή (cursor) στην κορυφή της αντίστοιχης μπάρας. Με αυτόν τον τρόπο μπορείτε να συλλέξετε τα στοιχεία που θα χρειαστείτε στα ζητήματα της εργασίας που δίνονται παρακάτω. Δυστυχώς ο ιστοχώρος αυτός δεν παρέχει τη δυνατότητα λήψης των δεδομένων σε αρχείο. Εναλλακτικά θα πρέπει να πάρετε αυτήν την πληροφορία με λήψη αρχείων στις εναλλακτικές βάσεις δεδομένων που δίνονται παραπάνω (σημειώνετε ότι μπορεί να υπάρχουν διαφορές στα στοιχεία της κάθε πηγής και στο πλαίσιο της εργασίας όλες οι πηγές που αναφέρονται θεωρούνται έγκυρες).

Συγκεκριμένα, θα χρειαστεί πρώτα να ορίσετε την κορύφωση της “καμπύλης” για τα ημερήσια νέα κρούσματα και για τους ημερήσιους νέους θανάτους για κάποιο κύμα που θα επιλέξετε. Για παράδειγμα για την Ελλάδα αυτό μπορεί να είναι το πρώτο κύμα, Μάρτιος – Μάιος 2020 ή το δεύτερο κύμα, Οκτώβριος 2020 – Ιανουάριος 2021, ή και το τρίτο κύμα που ξεκινά τον Φεβρουάριο 2021 αλλά δε φαίνεται να έχει ολοκληρωθεί (διαβάστε παρακάτω για να κρίνετε αν μπορείτε να χρησιμοποιήσετε ένα κύμα που δεν έχει ολοκληρωθεί). Η ημέρα της κορύφωσης θα οριστεί αιτιολογημένα από εσάς. Ένας απλός τρόπος είναι να ορίσετε την κορύφωση από τη μέγιστη τιμή στο χρονικό διάστημα του κύματος. Εδώ θέλει προσοχή, καθώς κάποια μεγάλη τιμή μπορεί να μην αντιστοιχεί σε κορύφωση, αλλά να είναι μια απόμακρη τιμή που μπορεί να προέκυψε από τον τρόπο καταγραφής ή κάποιο μεμονωμένο γεγονός. Για παράδειγμα, στο τρίτο κύμα για την Ελλάδα (σχήμα στα δεξιά) στις 3/5/2021 καταγράφηκαν 134 θάνατοι που είναι πολύ περισσότεροι από αυτούς που καταγράφηκαν τις άλλες κοντινές μέρες καθώς και από την κορύφωση που φαίνεται να είναι πιο πριν, περί τα μέσα Απριλίου. Άρα ως κορύφωση του δεύτερου κύματος στην Ελλάδα θα ήταν πιο σωστό να επιλέξετε την 15/4/2021 με 104 ημερήσιους θανάτους ή κάποια άλλη ημέρα που θα επιλέξετε με κάποιο άλλο κριτήριο (που θα πρέπει να αναφέρετε). Η αντίστοιχη κορύφωση για τα ημερήσια νέα κρούσματα θα μπορούσε να είναι 30/3/2021 με 4332 ημερήσια κρούσματα.

Αφού έχετε ορίσει τη μέγιστη τιμή για τα ημερήσια νέα κρούσματα και τους ημερήσιους νέους θανάτους για το κύμα που επιλέξατε, θα ορίσετε τα δείγματα σας για την ανάλυση στα ζητήματα του μέρους Α και του μέρους Β.

Για το μέρος Α, θα ορίσετε τις  $n$  τιμές για τα ημερήσια νέα κρούσματα από τη μέρα μετά την κορύφωση ως και  $n$  μέρες μετά την κορύφωση. Θα ορίσετε εσείς την τιμή  $n$  που θα πρέπει να 20 ή μεγαλύτερη. Για την περίπτωση της Ελλάδας και για τις ενδεικτικές επιλογές κορύφωσης που αναφέρθηκαν παραπάνω, αυτό σημαίνει ότι το σύνολο των  $n=20$  τιμών είναι από τις 31/3/2021 ως και τις 19/4/2021. Αντίστοιχα θα ορίσετε τις  $n=20$  συνεχόμενες τιμές για τους ημερήσιους νέους θανάτους από τη μέρα μετά την κορύφωση ως και 20 μέρες μετά την κορύφωση. Για την περίπτωση της Ελλάδας και σύμφωνα με την παραπάνω επιλογή κορύφωσης αυτό σημαίνει ότι το σύνολο των 20 τιμών είναι από τις 16/4/2021 ως και τις 5/5/2021. Στο συγκεκριμένο παράδειγμα τυχαίνει το χρονικό παράθυρο των 20 τιμών του συνόλου των ημερήσιων νέων κρουσμάτων να είναι αρκετά πριν από το χρονικό παράθυρο των 20 τιμών του συνόλου των ημερήσιων νέων θανάτων (υστέρηση 16 ημερών), αλλά για άλλες χώρες μπορεί τα δύο αυτά χρονικά παράθυρα να διαφέρουν πολύ λιγότερο.

Αν σε ένα από τα δύο σύνολα υπάρχουν τιμές που κρίνετε ότι είναι λανθασμένες ως προς το φαινόμενο που μελετάτε, π.χ. αρνητική τιμή μετά από διόρθωση του καταλόγου ή υπερβολικά μεγάλη τιμή λόγω ενημέρωσης / διόρθωσης του καταλόγου ή μηδενική τιμή λόγω μη-καταγραφής την ημέρα αυτή, τότε μπορείτε να αφαιρέσετε ή να διορθώσετε αυτήν την τιμή καθώς και άλλες τιμές στο δείγμα σας. Αν αφαιρέσετε κάποια τιμή από το πρώτο δείγμα, π.χ. των ημερήσιων κρουσμάτων, θα αφαιρέσετε και την αντίστοιχη τιμή από το δεύτερο δείγμα, π.χ. των ημερήσιων θανάτων, και η τιμή που αφαιρείται στο δεύτερο μπορεί να είναι για την ίδια ημερομηνία (αν υπάρχει) ή η αντίστοιχη τιμή στη διάταξη που έχουν στο δείγμα (π.χ. δέκατη τιμή στο δεύτερο σύνολο αν η λανθασμένη που εντοπίστηκε στο πρώτο σύνολο είναι η δέκατη). Σε κάθε περίπτωση τα δύο δείγματα, για κρούσματα και θανάτους, θα πρέπει να έχουν το ίδιο μέγεθος και να υπάρχει αντιστοιχία των ημερών στα δύο δείγματα. Για παράδειγμα για την Ελλάδα επισημάνθηκε παραπάνω πως στις 3/5/2021 καταγράφηκαν 134 θάνατοι, πολύ περισσότεροι από τις άλλες ημέρες εκείνης της περιόδου. Επιπλέον παρατηρούμε πως η ημέρα στις 3/5/2021 ήταν η επόμενη της Κυριακής του Πάσχα, όπου δεν καταγράφηκαν στοιχεία για αυτό και στις 2/5/2021 στον κατάλογο η τιμή των νέων κρουσμάτων και θανάτων είναι 0. Μια πιθανή αντιμετώπιση είναι να μοιράσετε την τιμή των νέων θανάτων της 3/5/2021 στις δύο ημέρες 2/5/2021 και 3/5/2021, δηλαδή 67 στις 2/5/2021 και 67 στις 3/5/2021 και τότε θα πρέπει

το ίδιο να κάνετε για τα νέα κρούσματα (αλλά οι ημερομηνίες αυτές δεν περιέχονται στο δείγμα των νέων κρουσμάτων όπως ορίστηκαν παραπάνω στο παράδειγμα της Ελλάδας). Μια άλλη πιθανή αντιμετώπιση είναι να αφαιρέσετε τις ημέρες 2/5/2021 και 3/5/2021 από το δείγμα των ημερήσιων θανάτων, που είναι η 17η και 18η τιμή του δείγματος. Αντίστοιχα θα πρέπει να αφαιρέσετε τη 17η και 18η τιμή στο δείγμα των ημερήσιων κρουσμάτων και να συνεχίσετε με τα δείγματα των 18 τιμών ή να επεκτείνετε τις αντίστοιχες περιόδους κατά δύο ημέρες ώστε να έχετε 20 τιμές σε κάθε δείγμα.

Στη συνέχεια θα διαιρέσετε την κάθε τιμή ημερήσιων νέων θανάτων με την αντίστοιχη τιμή ημερήσιων νέων κρουσμάτων (η πρώτη τιμή του δεύτερου συνόλου (θανάτων) ως προς την πρώτη τιμή του πρώτου συνόλου (κρουσμάτων) και όμοια για τις δεύτερες τιμές, τρίτες τιμές κτλ). Θα μετατρέψετε τις τιμές που προκύπτουν σε ποσοστά, πολλαπλασιάζοντας με εκατό. Αν κάποια τιμή ημερήσιων νέων θανάτων ή ημερήσιων νέων κρουσμάτων είναι 0, η τιμή του ποσοστού θα είναι επίσης 0. Αυτές οι τιμές εκφράζουν το ημερήσιο ποσοστό θνητότητας στη χώρα την περίοδο μετά την κορύφωση ως προς τα ημερήσια νέα κρούσματα, δηλαδή προσεγγιστικά μας λέει για τα νέα κρούσματα που έχουμε σε μια μέρα στη χώρα τι ποσοστό από αυτά θα οδηγηθεί σε θάνατο. Το σύνολο αυτών των τιμών θα αποτελέσει το δείγμα για κάθε μια από τις δύο χώρες Α και Β για τα ζητήματα του μέρους Α.

## **Ζητήματα της εργασίας**

Πριν να περάσετε και απαντήσετε τα ζητήματα για τη μελέτη Α και Β θα πρέπει να παρουσιάσετε συνοπτικά πως δημιουργήσατε τα δείγματα που χρησιμοποιούνται στις δύο μελέτες, π.χ. με κάποιο γράφημα, περιγραφή της επιλογής της ημέρας κορύφωσης για τα ημερήσια νέα κρούσματα και ημερήσιους νέους θανάτους, καθώς και προβλήματα που συναντήσατε στην επιλογή των δεδομένων.

### **Μελέτη Α**

Για τη μελέτη Α, θα θεωρήσετε το δείγμα για τη χώρα Α των  $n$  τιμών για το ημερήσιο ποσοστό θνητότητας ως προς τα ημερήσια νέα κρούσματα στη χώρα την περίοδο μετά την κορύφωση, που θα ονομάζεται δείγμα Α και το αντίστοιχο δείγμα για τη χώρα Β που θα ονομάζεται δείγμα Β. Στο εξής η τυχαία μεταβλητή που μετρήθηκε στα δείγματα Α και Β θα ονομάζεται απλά *ημερήσιο ποσοστό θνητότητας*.

Θα μελετήσετε την κατανομή του ημερήσιου ποσοστού θνητότητας στη χώρα την περίοδο μετά την κορύφωση, και κυρίως τη μέση τιμή του, για τις δύο χώρες που σας αντιστοιχούν.

1. Σχολιάστε την *κατανομή* του ημερήσιου ποσοστού θνητότητας την περίοδο μετά την κορύφωση στις δύο χώρες Α και Β με βάση τα αντίστοιχα δείγματα των  $n$  μετρήσεων. Θα πρέπει να συμπεριλάβετε τα παρακάτω, χρησιμοποιώντας το SPSS:

- Έναν πίνακα με συνοπτικά μέτρα κεντρικής τάσης (μέση τιμή και διάμεσο), μεταβλητότητας (διασπορά, τυπική απόκλιση, εύρος δεδομένων, πρώτο και τρίτο τεταρτομόριο). Ο πίνακας θα πρέπει να περιέχει τα μέτρα και για τα δύο δείγματα.
- Ιστόγραμμα για το κάθε δείγμα.
- Ένα σχήμα που να περιέχει τα δύο θηκογράμματα, ένα για κάθε δείγμα.

Να γίνουν αναλυτικά σχόλια για τον πίνακα και τα σχήματα συγκρίνοντας τα αποτελέσματα για τα δύο δείγματα. Είναι κανονική η κατανομή του ημερήσιου ποσοστού θνητότητας την περίοδο μετά την κορύφωση σε κάθε μια από τις δύο χώρες? Υπάρχουν διαφορές στην κατανομή του ημερήσιου ποσοστού θνητότητας την περίοδο μετά την κορύφωση στις δύο χώρες Α και Β με βάση τα δύο δείγματα;

2. Για κάθε ένα από τα δείγματα Α και Β των  $n$  παρατηρήσεων, ελέγξτε με κατάλληλο 95% διάστημα εμπιστοσύνης αν η μέση τιμή του ημερήσιου ποσοστού θνητότητας μπορεί να είναι 5%. Στην απάντησή σας συμπεριλάβετε κατάλληλο πίνακα αποτελεσμάτων στο SPSS και για τα δύο δείγματα και σχολιάστε τα αποτελέσματα. Συγκρίνετε επίσης την ακρίβεια της εκτίμησης της μέσης τιμής του ημερήσιου ποσοστού θνητότητας στις δύο χώρες. Σημειώστε τυχόν επιφύλαξη που έχετε για την εγκυρότητα των αποτελεσμάτων.
3. Με βάση τα δείγματα Α και Β, ελέγξτε αν η μέση τιμή του ημερήσιου ποσοστού θνητότητας μπορεί να είναι ίδια στις δύο χώρες χρησιμοποιώντας κατάλληλο 95% διάστημα εμπιστοσύνης. Στην απάντησή σας συμπεριλάβετε κατάλληλο πίνακα αποτελεσμάτων στο SPSS και σχολιάστε τα αποτελέσματα. Σημειώστε τυχόν επιφύλαξη που έχετε για την εγκυρότητα των αποτελεσμάτων.

### **Μελέτη Β:**

Θέλουμε να εξετάσουμε αν μετά την κορύφωση του κύματος, φαίνεται ο αριθμός ημερήσιων νέων κρουσμάτων να μειώνεται γραμμικά, δηλαδή ο αριθμός των ημερήσιων νέων κρουσμάτων να εξαρτάται γραμμικά από το χρόνο. Για αυτό θα θεωρήσετε τα δείγματα των  $n$  ζευγαρωτών παρατηρήσεων για τη χώρα Α, όπου το πρώτο στοιχείο κάθε ζευγαριού είναι ο αύξων αριθμός της ημέρας και το δεύτερο η αντίστοιχη τιμή ημερήσιων νέων κρουσμάτων στη χώρα. Το ίδιο θα κάνετε και για τον αριθμό ημερήσιων νέων θανάτων. Η ανάλυση θα γίνει σύμφωνα με τα παρακάτω ερωτήματα.

4. Χρησιμοποιώντας το SPSS, κάνετε κατάλληλο διάγραμμα διασποράς και υπολογίστε τον αντίστοιχο συντελεστή συσχέτισης για τον αύξοντα αριθμό της ημέρας και τον αριθμό ημερήσιων νέων κρουσμάτων χρησιμοποιώντας το δείγμα για τη χώρα Α. Σχολιάστε αν εξαρτάται ο αριθμός ημερήσιων νέων κρουσμάτων από την ημέρα. Κάνετε το ίδιο για τον αριθμό ημερήσιων νέων θανάτων.
5. Για το δείγμα της χώρας Α που περιέχει τον αύξοντα αριθμό της ημέρας και τον αριθμό ημερήσιων νέων κρουσμάτων και χρησιμοποιώντας το SPSS, εκτιμήστε το μοντέλο γραμμικής παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων. Κάνετε το ίδιο για τον αριθμό ημερήσιων νέων θανάτων. Σχολιάστε το κάθε ένα από τα δύο μοντέλα που εκτιμήσατε με έμφαση στην καταλληλότητα του μοντέλου για προβλέψεις. Στην απάντησή σας θα συμπεριλάβετε τους πίνακες του SPSS που δίνουν την εκτίμηση των παραμέτρων του μοντέλου (σταθερός όρος, κλίση και τυπική απόκλιση σφαλμάτων παλινδρόμησης). Φαίνεται τα δύο μοντέλα παλινδρόμησης που προσεγγίζουν τη γραμμική μείωση των ημερήσιων νέων κρουσμάτων και τη γραμμική μείωση των ημερήσιων νέων θανάτων να συμφωνούν; Αιτιολογείστε την απάντησή σας με αναφορά στις παραμέτρους του κάθε μοντέλου.

Θέλουμε να προβλέψουμε τον αριθμό ημερήσιων νέων θανάτων από τον αριθμό νέων κρουσμάτων κάποιες μέρες πριν. Στο δείγμα που έχετε για τη χώρα Α έχετε επιλέξει τις δύο περιόδους για ημερήσια κρούσματα και ημερήσιους θανάτους με κάποια υστέρηση. Για το παράδειγμα της Ελλάδας που χρησιμοποιήθηκε παραπάνω, είχε επιλέξει η περίοδος από 31/3/2021 ως 19/4/2021 για τα ημερήσια κρούσματα και η περίοδος από 16/4/2021 ως 5/5/2021 για τους ημερήσιους θανάτους, όπου η διαφορά (υστέρηση) είναι 16 ημέρες. Για αυτήν την υστέρηση  $\tau$  (για την Ελλάδα  $\tau=16$ ) θέλουμε να διερευνήσουμε αν μπορούμε να προβλέψουμε τους ημερήσιους νέους θανάτους μια μέρα  $t$  από τα ημερήσια νέα κρούσματα την ημέρα  $t-\tau$ , και να κάνουμε το ίδιο για δύο ακόμα τιμές υστέρησης. Τις άλλες δύο υστερήσεις μπορείτε να τις επιλέξετε ελεύθερα σε σχέση με την υστέρηση  $\tau$  για τη χώρα Α (δηλαδή τη διαφορά της περιόδου των ημερήσιων νέων κρουσμάτων και των ημερήσιων νέων θανάτων). Για κάθε μια από τις δύο νέες τιμές υστέρησης θα πρέπει να διαμορφώσετε αντίστοιχα το δείγμα ζευγαρωτών παρατηρήσεων ημερήσιων κρουσμάτων και θανάτων, κρατώντας τις τιμές των ημερήσιων θανάτων σταθερές και μεταθέτοντας χρονικά τις τιμές των αντίστοιχων ημερήσιων κρουσμάτων σύμφωνα με την επιλογή της υστέρησης. Η ανάλυση θα γίνει σύμφωνα με τα παρακάτω ερωτήματα.

6. Χρησιμοποιώντας το SPSS, κάνετε κατάλληλο διάγραμμα διασποράς και υπολογίστε τον αντίστοιχο συντελεστή συσχέτισης για τον αριθμό ημερήσιων νέων θανάτων και των ημερήσιων νέων κρουσμάτων  $\tau$  ημέρες πριν χρησιμοποιώντας το δείγμα της χώρας Α. Σχολιάστε αν εξαρτάται ο αριθμός ημερήσιων νέων θανάτων από τον αριθμό ημερήσιων νέων κρουσμάτων  $\tau$  ημέρες πριν. Κάνετε το ίδιο για τις άλλες δύο υστερήσεις που επιλέξατε.
7. Χρησιμοποιώντας το SPSS, εκτιμήστε το μοντέλο γραμμικής παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων για την εξάρτηση του αριθμού ημερήσιων νέων θανάτων από τον αριθμό των ημερήσιων νέων κρουσμάτων  $\tau$  ημέρες πριν χρησιμοποιώντας το δείγμα Α. Κάνετε το ίδιο για τις άλλες δύο υστερήσεις που επιλέξατε. Σχολιάστε το κάθε ένα από τα τρία μοντέλα που εκτιμήσατε με έμφαση στην καταλληλότητα του μοντέλου για προβλέψεις. Στην απάντησή σας θα συμπεριλάβετε τους πίνακες του SPSS που δίνουν την εκτίμηση των παραμέτρων του μοντέλου (σταθερός όρος, κλίση και τυπική απόκλιση σφαλμάτων παλινδρόμησης). Φαίνεται κάποια από τα τρία μοντέλα παλινδρόμησης (για κάποια από τα τις τρεις τιμές υστέρησης  $\tau$  που επιλέξατε) να είναι πιο κατάλληλο για την πρόβλεψη του αριθμού ημερήσιων νέων θανάτων από τον αριθμό των ημερήσιων νέων κρουσμάτων  $\tau$  ημέρες πριν; Αιτιολογείστε την απάντησή σας με αναφορά στις παραμέτρους και τα στατιστικά στοιχεία του κάθε μοντέλου.