

Θέμα εργασίας

Οδηγίες

Η εργασία θα αξιολογηθεί με άριστα τη **μία μονάδα**. Θα πρέπει να τηρηθούν οι παρακάτω οδηγίες αλλιώς θα υπάρχει απώλεια βαθμού για κάθε παράλειψη τήρησης οδηγίας.

• Η εργασία θα πρέπει να είναι αυθεντική. Μπορεί να υπάρχει συνεργασία μεταξύ των ομάδων πάνω στα ερωτήματα της εργασίας αλλά η αντιγραφή απαγορεύεται.

• Η εργασία θα πρέπει να είναι γραμμένη στον υπολογιστή σε πρόγραμμα επεξεργασίας κειμένου, όπως Word. Αν είναι γραμμένη σε άλλο πρόγραμμα επεξεργασίας κειμένου, θα πρέπει να μετατραπεί σε pdf.

• Στην αρχή (πρώτη σελίδα) θα πρέπει να αναφέρονται τα ονόματα και τα ΑΕΜ των φοιτητών/τριών της ομάδας.

• Οι απαντήσεις και λύσεις των ζητημάτων θα πρέπει να δίνονται 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 τα ΑΕΜ των φοιτητών/τριών της ομάδας και ο κάθε πίνακας από το SPSS θα πρέπει να έχει στην επικεφαλίδα τα ΑΕΜ των φοιτητών/τριών της ομάδας. 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 υποβληθεί ηλεκτρονικά μέσω της ιστοσελίδας του μαθήματος στο elearning το αργότερο ως τις **10/7/2022**. Θα πρέπει να υποβληθούν ηλεκτρονικά τρία αρχεία, ως εξής:

1. Το πρώτο αρχείο θα περιέχει την αναφορά της εργασίας σε μορφή κειμένου. Το όνομα του αρχείου θα πρέπει να είναι ΑΕΜ<ΑΕΜ>ΑΕΜ<ΑΕΜ>Report.docx όπου <ΑΕΜ> είναι ο Αριθμός Ειδικού Μητρώου (χωρίς τις ανισότητες) του/της φοιτητή/τριας της ομάδας (αν είναι ομάδα ενός ατόμου θα πρέπει το όνομα του αρχείου να είναι ΑΕΜ<ΑΕΜ>Report.docx) και docx δηλώνει ότι είναι γραμμένο σε Word (άλλες επιτρεπτές μορφές είναι doc και pdf). Για παράδειγμα, αν η ομάδα δύο ατόμων έχει ΑΕΜ 9820 και 10000, το αρχείο αναφοράς θα πρέπει να έχει όνομα ΑΕΜ9820ΑΕΜ10000Report.docx, ΑΕΜ9820ΑΕΜ10000Report.doc ή ΑΕΜ9820ΑΕΜ10000Report.pdf.

2. Το δεύτερο αρχείο θα είναι το αρχείο δεδομένων SPSS (κατάληξη .sav) που θα δημιουργήσετε για τη Μελέτη Α. Σε αντιστοιχία με το όνομα του αρχείου αναφοράς το όνομα του αρχείου αυτού θα πρέπει να είναι ΑΕΜ<ΑΕΜ>ΑΕΜ<ΑΕΜ>StudyA.sav για ομάδα δύο ατόμων και ΑΕΜ<ΑΕΜ>StudyA.sav για ομάδα ενός ατόμου.

3. Το τρίτο αρχείο θα είναι το αρχείο δεδομένων SPSS (κατάληξη .sav) που θα δημιουργήσετε για τη Μελέτη Β. Σε αντιστοιχία με το όνομα του αρχείου για τη μελέτη Α το όνομα του αρχείου αυτού θα πρέπει να είναι ΑΕΜ<ΑΕΜ>ΑΕΜ<ΑΕΜ>StudyB.sav για ομάδα δύο ατόμων και ΑΕΜ<ΑΕΜ>StudyB.sav για ομάδα ενός ατόμου.

Παρακαλώ να εξετάσετε αν τα ονόματα των τριών αρχείων είναι σωστά πριν τα υποβάλλετε. Αρχείο με όνομα που δεν είναι στην προβλεπόμενη μορφή μπορεί να αγνοηθεί και η εργασία να μην αξιολογηθεί! Τα αρχεία δε θα πρέπει να έχουν συμπίεστεί σε ένα αρχείο για την υποβολή.

Περιγραφή προβλήματος και δεδομένα

Η εργασία αφορά την πανδημία του κορονοϊού που χτύπησε πολλές χώρες του πλανήτη και όλες τις χώρες της Ευρώπης. Τα δεδομένα που θα χρησιμοποιήσετε δίνονται στο αρχείο ECDC-7Days-Testing.xlsx, που μεταφορτώθηκε παλιότερα από την ιστοσελίδα <https://www.stelios67pi.eu>, όπου υπάρχουν και πολλά άλλα στοιχεία που ενδεχομένως να θέλετε να αντλήσετε. Το αρχείο έχει εβδομαδιαία δεδομένα για αριθμό κρουσμάτων και τεστ και κυρίως για το δείκτη θετικότητας (στήλη positivity_rate) για χώρες της Ευρώπης όπως τα καταγράφει το European Center for Disease Control (ECDC).

Στα ζητήματα της εργασίας που γίνεται αναφορά σε μια χώρα της Ευρώπης από τις 25 Ευρωπαϊκές χώρες που δίνονται στο αρχείο EuropeanCountries.xlsx, θα ορίσετε αυτήν τη χώρα σύμφωνα με τον αριθμό της ομάδας σας. Η χώρα που αντιστοιχεί στην ομάδα είναι αυτή με αύξοντα αριθμό το υπόλοιπο της διαίρεσης του ΑΕΜ με το 25 αυξημένο κατά ένα (αν η ομάδα αποτελείται από δύο μέλη, του ενός από τα δύο μέλη). Για παράδειγμα, για ΑΕΜ=10076 ο αύξων αριθμός χώρας είναι 2 και αντιστοιχεί στο Βέλγιο. Ας ονομάσουμε τη χώρα που επιλέξατε ως χώρα Α. Η χώρα Β θα είναι η αλφαβητικά γειτονική με τη χώρα Α που σας αναλογεί στη λίστα των 25 χωρών (πριν ή μετά, ελεύθερη επιλογή). Υπάρχει η δυνατότητα να χρησιμοποιηθεί μια άλλη γειτονική Ευρωπαϊκή χώρα αν κριθεί πως δεν υπάρχουν ικανοποιητικά δεδομένα για τα ζητήματα της εργασίας από την Ευρωπαϊκή χώρα που αρχικά ορίστηκε και αυτό αφορά τη χώρα Α και τη χώρα Β. Θα πρέπει στην αρχή της αναφοράς σας να δηλώσετε την αλλαγή και να την αιτιολογήσετε.

Η ομάδα μπορεί επίσης να αντλήσει στοιχεία για τα ζητήματα της εργασίας από άλλες πηγές αν κρίνει πως είναι πιο εύκολα επεξεργάσιμα ή πιο ακριβή και καλύτερα επικαιροποιημένα.

Θα γίνουν δύο μελέτες με συγκεκριμένα ερωτήματα που δίνονται παρακάτω.

Μελέτη Α

Στην Μελέτη Α, θέλουμε να διερευνήσουμε αν ο εβδομαδιαίος δείκτης θετικότητας σε μια χώρα Α διαφέρει από αυτόν της Ελλάδας καθώς και από μια άλλη χώρα Β. Πρώτα θα βρείτε, ξεκινώντας από το τέλος του 2021 και

πηγαίνοντας προς τα πίσω, τη βδομάδα στην οποία ο δείκτης θετικότητας της χώρας Α είναι μέγιστος. Έστω ότι αυτή είναι η βδομάδα WA. Θα θεωρήσετε τη χρονική περίοδο 12 βδομάδων πριν την κορύφωση του δείκτη θετικότητας της χώρας Α, δηλαδή τις βδομάδες από WA - 12 ως και WA - 1. Θα κάνετε το ίδιο για τη χώρα Β και την Ελλάδα, δηλαδή θα βρείτε την εβδομάδα κορύφωσης του δείκτη θετικότητας και θα θεωρήσετε τις 12 βδομάδες πριν από αυτήν. Με αυτόν τον τρόπο θα ορίσετε τα τρία δείγματα για τις χώρες Α, Β και Ελλάδα μεγέθους 12. Θα πρέπει να μεταφέρετε σε αρχείο δεδομένων SPSS τα δεδομένα για τον εβδομαδιαίο δείκτη θετικότητας για τις τρεις χώρες (σύνολο 36 τιμές). Στην αναφορά σας θα αναφέρετε τις χρονικές περιόδους που επιλέξατε για την κάθε χώρα.

Τα ζητήματα στη Μελέτη Α είναι:

1. Σχολιάστε την *κατανομή* του εβδομαδιαίου δείκτη θετικότητας την περίοδο πριν την κορύφωση στις τρεις χώρες (Α, Β και Ελλάδα) με βάση τα αντίστοιχα δείγματα των 12 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 τρία δείγματα.
 - Ιστόγραμμα για το κάθε δείγμα.
 - Ένα σχήμα που να περιέχει τα τρία θηκογράμματα, ένα για κάθε δείγμα.

Να γίνουν αναλυτικά σχόλια για τον πίνακα και τα σχήματα συγκρίνοντας τα αποτελέσματα για τα τρία δείγματα. Είναι κανονική η κατανομή του εβδομαδιαίου δείκτη θετικότητας σε κάθε μια από τις τρεις χώρες? Υπάρχουν διαφορές στην κατανομή του εβδομαδιαίου δείκτη θετικότητας στις δύο χώρες Α και Β καθώς και μεταξύ της χώρας Α και της Ελλάδας με βάση τα τρία δείγματα;

2. Για κάθε ένα από τα τρία δείγματα Α, Β και Ελλάδα των 12 παρατηρήσεων, ελέγξτε με κατάλληλο 95% διάστημα εμπιστοσύνης αν η μέση τιμή του εβδομαδιαίου δείκτη θετικότητας μπορεί να είναι 5%. Στην απάντησή σας συμπεριλάβετε κατάλληλο πίνακα αποτελεσμάτων στο SPSS και για τα τρία δείγματα και σχολιάστε τα αποτελέσματα. Συγκρίνετε επίσης την ακρίβεια της εκτίμησης της μέσης τιμής του εβδομαδιαίου δείκτη θετικότητας στις τρεις χώρες.

Με βάση τα δείγματα για τις χώρες Α και Β, ελέγξτε αν η μέση τιμή του του εβδομαδιαίου δείκτη θετικότητας μπορεί να είναι ίδια στις δύο χώρες χρησιμοποιώντας κατάλληλο 95% διάστημα εμπιστοσύνης (θεωρείστε κανονική κατανομή και ίσες διασπορές). Στην απάντησή σας συμπεριλάβετε κατάλληλο πίνακα αποτελεσμάτων στο SPSS και σχολιάστε τα αποτελέσματα. Επαναλάβετε το ίδιο, για τις χώρες Α και Ελλάδα.

Μελέτη Β

Στην Μελέτη Β, θέλουμε να διερευνήσουμε αν μπορούμε σε μια χώρα να προβλέψουμε τους θανάτους από κορονοϊό από τον εβδομαδιαίο δείκτη θετικότητας κάποιας προηγούμενης εβδομάδας. Θα απαντήσετε αυτό το ερώτημα για τη χώρα Α και Β. Θα πρέπει να βρείτε για ποια υστέρηση εβδομάδας του εβδομαδιαίου δείκτη θετικότητας με αναφορά στην εβδομάδα που αναφέρονται οι θάνατοι είναι καλύτερο το μοντέλο απλής γραμμικής παλινδρόμησης του εβδομαδιαίου αριθμού θανάτων (ανά εκατομμύριο κατοίκων) ως προς τον εβδομαδιαίο δείκτη θετικότητας κάποιας εβδομάδας πριν. Τον εβδομαδιαίο αριθμό θανάτων (ανά εκατομμύριο κατοίκων) για τη χώρα Α μπορείτε να το διαβάσετε στο σχετικό γράφημα στην ιστοσελίδα <https://www.stelios67pi.eu/testing.html>. Θα δοκιμάσετε τα μοντέλα για υστέρηση από ένα ως και τέσσερις βδομάδες πριν. Θα πρέπει να επιλέξετε μια περίοδο 4 μηνών (16 εβδομάδων, ελεύθερη επιλογή) και θα πρέπει να μεταφέρετε σε αρχείο δεδομένων SPSS τα δεδομένα για τον εβδομαδιαίο αριθμό θανάτων (μια στήλη) και τον εβδομαδιαίο δείκτη θετικότητας μιας ως και 4 βδομάδες πριν (καλύτερα σε 4 διαφορετικές στήλες) για τη χώρα Α και τη χώρα Β.

Τα ζητήματα στη Μελέτη Β είναι:

3. Χρησιμοποιώντας το SPSS και το δείγμα για τη χώρα Α, κάνετε κατάλληλο διάγραμμα διασποράς και υπολογίστε τον αντίστοιχο συντελεστή συσχέτισης για τον εβδομαδιαίο δείκτη θετικότητας με υστέρηση (υστέρηση 1, 2, 3, 4) για τη χώρα Α και τον εβδομαδιαίο αριθμό θανάτων (ανά εκατομμύριο κατοίκων) για τη χώρα Α. Θα το επαναλάβετε αυτό τέσσερις φορές, μια για κάθε υστέρηση (αν θέλετε, υπάρχει δυνατότητα στο SPSS να συμπεριλάβετε σε έναν πίνακα τα διαγράμματα διασποράς και επίσης σε έναν πίνακα τα αποτελέσματα για τον συντελεστή συσχέτισης). Σχολιάστε αν φαίνεται να εξαρτάται ο εβδομαδιαίος αριθμός θανάτων από τον εβδομαδιαίο δείκτη θετικότητας με υστέρηση και για ποια υστέρηση φαίνεται η εξάρτηση να είναι πιο ισχυρή. Θα επαναλάβετε το ίδιο για τη χώρα Β. Συγκρίνετε τα αποτελέσματα σας για τις δύο χώρες Α και Β.
4. Για το δείγμα της χώρας Α και χρησιμοποιώντας το SPSS, εκτιμήστε το μοντέλο απλής γραμμικής παλινδρόμησης με τη μέθοδο ελαχίστων τετραγώνων για την εξάρτηση του εβδομαδιαίου αριθμού θανάτων από τον εβδομαδιαίο δείκτη θετικότητας κάποιας υστέρησης. Αυτό θα το κάνετε για τη κάθε μια από τις τέσσερις υστερήσεις. Σχολιάστε το κάθε ένα από τα τέσσερα 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 κάθε μοντέλου (σταθερός όρος, κλίση και τυπική απόκλιση σφαλμάτων παλινδρόμησης). Φαίνεται κάποιο από τα τέσσερα μοντέλα παλινδρόμησης να είναι πιο κατάλληλο για να

προβλέπει τον εβδομαδιαίο αριθμό θανάτων από τον εβδομαδιαίο δείκτη θετικότητας κάποιας υστέρησης; Αιτιολογείστε την απάντησή σας με αναφορά στις παραμέτρους του κάθε μοντέλου. Θα επαναλάβετε το ίδιο για τη χώρα Β. Συγκρίνετε τα αποτελέσματα σας για τις δύο χώρες Α και Β.

5. Αναφέρετε τυχόν παρατηρήσεις σας για την εγκυρότητα των αποτελεσμάτων στα ζητήματα των δύο μελετών.