

## **ΕΡΓΑΣΙΑ ΣΤΙΣ ΠΡΟΒΛΕΨΕΙΣ**

Στο συμπιεσμένο αρχείο περιλαμβάνονται όλα τα αρχεία Excel με τα δεδομένα που χρειάζεστε καθώς και η εκφώνηση των ασκήσεων (αρχείο Word).

Κάθε εργασία μπορεί να εκπονηθεί από 3 άτομα. Η εργασία είναι προαιρετική. Παραδίδετε ηλεκτρονικά ένα αρχείο Excel που θα αναπτύξετε και ένα αρχείο Word με συνοπτική έκθεση ευρημάτων. Ο βαθμός της εργασίας μετράει **προσθετικά 1 βαθμό στον τελικό βαθμό** (ο οποίος εφαρμόζεται εφόσον έχετε περάσει το μάθημα με τις εξετάσεις).

**ΠΡΟΣΟΧΗ:** Η εργασία που θα παραδώσετε θα πρέπει να είναι υψηλού επιπέδου, επαγγελματικής ποιότητας. Πρόχειρες εργασίες δεν θα ληφθούν διόλου υπόψη. Θα παραδώσετε τόσο το Excel όσο και κείμενο (word ή pdf). Οι εργασίες θα ελεγχθούν επίσης και ως προς το πρωτότυπό τους. Εργασίες που έχουν προκύψει από αντιγραφή θα μηδενισθούν και δεν θα γίνουν άλλες εργασίες δεκτές από τους συντελεστές των εργασιών αυτών.

**ΗΜΕΡΟΜΗΝΙΑ ΠΑΡΑΔΟΣΗΣ: Παρασκευή 14/11, 18:00 ΑΥΣΤΗΡΑ**

**ΜΟΝΟΝ ΗΛΕΚΤΡΟΝΙΚΗ ΠΑΡΑΔΟΣΗ ΜΕΣΩ ΑΝΑΡΤΗΣΗΣ ΣΤΟ E-CLASS**

### **ΑΣΚΗΣΗ 1**

Στο αρχείο «Ποσοστό του πληθυσμού που χρησιμοποιεί το Ίντερνετ.xls» περιλαμβάνονται στοιχεία από το Our World in Data σχετικά με τη χρήση Ίντερνετ ετησίως σε ποσοστό του πληθυσμού της Ελλάδας, καθώς και άλλων 4 χωρών, από το 1990 έως και το 2023 (ή έως το 2022 όπου δεν υπάρχουν διαθέσιμα στοιχεία). Σε όλα τα παρακάτω ερωτήματα θεωρήστε τα δεδομένα των ετών 1990 έως και 2015 ως δεδομένα εκμάθησης (training data), ενώ τα δεδομένα των ετών 2016 έως 2023 (ή 2022) ως δεδομένα δοκιμών (testing data).

Θα παραδώσετε τις απαντήσεις σας σε αρχείο Excel, όπου θα φαίνονται με ευκρινή τρόπο τυχόν γραφήματα, δείκτες, κλπ.

Πραγματοποιείτε τα ακόλουθα:

- (1) Για **τη χρονοσειρά που αφορά την Ελλάδα (γραμμή 7)**, προβλέψτε τη χρήση Ίντερνετ για κάθε έτος από το 2016 και μετά, με χρήση των εξής μεθόδων:
  - Κινούμενος μέσος τριών και πέντε σημείων
  - Απλή εκθετική εξομάλυνση με  $\alpha=0.3$
  - Εκθετική εξομάλυνση με τάση με  $\alpha=0.2$  και  $\beta=0.8$
  - Γραμμική παλινδρόμηση (όπου το έτος 1990 θεωρείται έτος 1)
- (2) Για κάθε μία από τις μεθόδους, υπολογίστε το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα για τα έτη 2016-2023.

**Για τις τρεις πρώτες μεθόδους, εφόσον έχετε υπολογίσει την πρόβλεψη για κάθε έτος, μπορείτε να χρησιμοποιείτε τα δεδομένα αυτού του έτους για την εξαγωγή της πρόβλεψης του επόμενου έτους.**

- (3) Συγκρίνετε τις μεθόδους μεταξύ τους και υποδείξτε την καλύτερη. Στη συνέχεια, για την απλή εκθετική εξομάλυνση καθώς και για την εκθετική εξομάλυνση με τάση, υπολογίστε την τιμή του  $\alpha$  και αντίστοιχα τις τιμές των  $\alpha$  και  $\beta$  που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα και επαναξιολογήστε τις μεθόδους.
- (4) Για **τις χρονοσειρές που αφορούν την Αμερική (γραμμή 6) και την Κένυα (γραμμή 9)**, προβλέψτε τη χρήση Ίντερνετ για το έτος 2022, χρησιμοποιώντας τη μέθοδο της γραμμικής παλινδρόμησης με βάση τα δεδομένα των ετών 1990 έως και 2007. Ποιο είναι το απόλυτο και τετραγωνικό σφάλμα σε κάθε περίπτωση;
- (5) Επαναλάβετε το ερώτημα (4), αυτή τη φορά χρησιμοποιώντας τα δεδομένα των ετών 2004 έως και 2021 και συγκρίνετε τις δύο περιπτώσεις για την κάθε χώρα.

## **ΑΣΚΗΣΗ 2**

Στο αρχείο «Κατά κεφαλήν ΑΕΠ.xlsx» περιλαμβάνονται στοιχεία από το Our World in Data για το κατά κεφαλήν Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ) ετησίως της Ελλάδας, καθώς και άλλων 4 χωρών, εκφρασμένο σε σταθερά διεθνή δολάρια (αγοραστική δύναμη του 2021) για τα έτη 1990 έως και 2022. Σε όλα τα παρακάτω ερωτήματα, θα θεωρήσετε τα δεδομένα των ετών 1990 έως και 2015 ως δεδομένα εκμάθησης (training data), ενώ τα δεδομένα των ετών 2016 έως και 2022 ως δεδομένα δοκιμών (testing data).

Θα παραδώσετε τις απαντήσεις σας σε αρχείο Excel, όπου θα φαίνονται με ευκρινή τρόπο τυχόν γραφήματα, δείκτες, κλπ.

Πραγματοποιείτε τα ακόλουθα:

- (1) Για **τη χρονοσειρά που αφορά την Ελλάδα (γραμμή 7)**, προβλέψτε το κατά κεφαλήν ΑΕΠ για κάθε έτος από το 2016 και μετά, με χρήση των εξής μεθόδων:
  - Κινούμενος μέσος τεσσάρων σημείων
  - Απλή εκθετική εξομάλυνση με  $\alpha=0.3$
  - Εκθετική εξομάλυνση με τάση με  $\alpha=0.2$  και  $\beta=0.8$
  - Γραμμική παλινδρόμηση (όπου το έτος 1990 θεωρείται έτος 1)
- (2) Για κάθε μία από τις μεθόδους, υπολογίστε το μέσο απόλυτο και το μέσο τετραγωνικό σφάλμα για τα έτη 2016-2022.

**Για τις τρεις πρώτες μεθόδους, εφόσον έχετε υπολογίσει την πρόβλεψη για κάθε έτος, μπορείτε να χρησιμοποιείτε τα δεδομένα αυτού του έτους για την εξαγωγή της πρόβλεψης του επόμενου έτους.**

- (3) Συγκρίνετε τις μεθόδους μεταξύ τους και υποδείξτε την καλύτερη. Στη συνέχεια, για την απλή εκθετική εξομάλυνση καθώς και για την εκθετική εξομάλυνση με τάση, υπολογίστε την τιμή του  $\alpha$  και αντίστοιχα τις τιμές των  $\alpha$  και  $\beta$  που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα και επαναξιολογήστε τις μεθόδους.
- (4) Για **τις χρονοσειρές που αφορούν την Αμερική (γραμμή 6) και την Κένυα (γραμμή 9)**, προβλέψτε το κατά κεφαλήν ΑΕΠ για το έτος 2022, χρησιμοποιώντας τη μέθοδο της γραμμικής παλινδρόμησης με βάση τα δεδομένα των ετών 1990 έως και 2015. Ποιο είναι το απόλυτο και τετραγωνικό σφάλμα σε κάθε περίπτωση;
- (5) Με βάση τους υπολογισμούς σας για το ερώτημα (4), ποια είναι η πρόβλεψή σας για το κατά κεφαλήν ΑΕΠ για το έτος 2024; Αναζητήστε το αντίστοιχο αρχείο από το website Our World in Data -θα πρέπει να κάνετε αναζήτηση στο Internet- και δείτε πώς διαφοροποιείται η πρόβλεψή σας από τα πραγματικά στοιχεία.
- (6) Συνδυάστε τα δεδομένα σας με το αρχείο «Ποσοστό του πληθυσμού που χρησιμοποιεί το Ίντερνετ.xlsx» για τα έτη 1990-2022, ώστε να υπολογίστε τον συντελεστή συσχέτισης ( $r$ ) και τον συντελεστή προσδιορισμού ( $r^2$ ) μεταξύ των δύο χρονοσειρών: κατά κεφαλήν ΑΕΠ και χρήση Ίντερνετ (% του πληθυσμού) για την Ελλάδα, την Αμερική και την Κένυα, αντίστοιχα. Υπάρχει συσχέτιση μεταξύ των δύο χρονοσειρών για την κάθε χώρα; Ποιο είναι το πόρισμά σας;
- (7) Αφαιρέστε από τα δεδομένα σας το έτος 2020, το οποίο αντιστοιχεί στην περίοδο της πανδημίας COVID-19, και επαναλάβετε όλες τις αναλύσεις που πραγματοποιήσατε στα ερωτήματα (1), (2), (3) και (6). Πώς επηρεάζει η αφαίρεση του έτους 2020 την ακρίβεια των προβλέψεων και τη συμπεριφορά των μοντέλων σας; Μεταβάλλεται η συσχέτιση μεταξύ του κατά κεφαλήν ΑΕΠ και της χρήσης Ίντερνετ (% του πληθυσμού) για την κάθε χώρα;

## **ΑΣΚΗΣΗ 3**

Στο αρχείο «Αφίξεις σε τουριστικά καταλύματα.xlsx» περιλαμβάνονται μηνιαία στοιχεία από τη EUROSTAT σχετικά με τον αριθμό των αφίξεων σε τουριστικά καταλύματα στην Ελλάδα για την περίοδο: Ιανουάριος 2011 - Αύγουστος 2025. Σε όλα τα παρακάτω ερωτήματα, θα θεωρήσετε τα δεδομένα των ετών 2011 έως και 2022 ως δεδομένα εκμάθησης (training data), ενώ τα δεδομένα των ετών 2023 έως και 2025 (Αύγουστος) ως δεδομένα δοκιμών (testing data).

Θα παραδώσετε τις απαντήσεις σας σε αρχείο Excel, όπου θα φαίνονται με ευκρινή τρόπο τυχόν γραφήματα, δείκτες, κλπ.

Πραγματοποιείτε τα ακόλουθα:

- (1) Δημιουργείτε τη χρονοσειρά των μηνιαίων αφίξεων σε τουριστικά καταλύματα και απεικονίστε τη γραφικά. Σχολιάστε συνοπτικά τη συμπεριφορά της, αναφέροντας την ύπαρξη και τη σταθερότητα τυχόν εποχικότητας, τη γενική τάση (αύξηση, μείωση ή σταθερότητα) και τη μεταβλητότητα των δεδομένων διαχρονικά.
- (2) Ξεκινώντας από το έτος 2011 και αξιοποιώντας σταδιακά τα δεδομένα των επόμενων ετών, εξάγετε προβλέψεις για τις μηνιαίες αφίξεις σε τουριστικά καταλύματα στην Ελλάδα για κάθε έτος από το 2023 και μετά, με χρήση των εξής μεθόδων:
  - Κινούμενος μέσος δώδεκα σημείων
  - Εκθετική εξομάλυνση με τάση με  $\alpha=0.3$  και  $\beta=0.2$
  - Εκθετική εξομάλυνση με πολ/κή εποχικότητα με  $\alpha=0.3$  και  $\gamma=0.2$ , θεωρώντας ότι η εποχικότητα είναι μηνιαία με ετήσιο κύκλο ( $L=12$ )
- (3) Για κάθε μία από τις μεθόδους, υπολογίστε το σφάλμα ανά μήνα, καθώς και το μέσο απόλυτο και μέσο τετραγωνικό σφάλμα ανά έτος για την περίοδο 2023-2025. Σχολιάστε ποια μέθοδος αποδίδει καλύτερα με βάση τα σφάλματα.
- (4) Υπολογίστε το ετήσιο άθροισμα αφίξεων σε τουριστικά καταλύματα στην Ελλάδα για την περίοδο 2011-2022. Στη συνέχεια, συνδυάστε τα αποτελέσματα αυτά με το αρχείο «Κατά κεφαλήν ΑΕΠ.xlsx», από το οποίο θα χρησιμοποιήσετε τα δεδομένα της ίδιας περιόδου (2011-2022), ώστε να υπολογίσετε τον συντελεστή συσχέτισης ( $r$ ) και τον συντελεστή προσδιορισμού ( $r^2$ ) μεταξύ των δύο χρονοσειρών: ετήσιο άθροισμα αφίξεων σε τουριστικά καταλύματα και κατά κεφαλήν ΑΕΠ για την Ελλάδα. Υπάρχει συσχέτιση μεταξύ τουριστικής κίνησης και οικονομικής ευημερίας;
- (5) Αφαιρέστε από τη χρονοσειρά ολόκληρα τα έτη 2020-2021 (COVID-19) και επαναλάβετε όλες τις αναλύσεις που πραγματοποιήσατε στα ερωτήματα (2) και (3) μόνο για τη μέθοδο της εκθετικής εξομάλυνσης με εποχικότητα (ίδιες τιμές για  $\alpha$  και  $\gamma$ ), καθώς και τους υπολογισμούς για το ερώτημα (4). Στη συνέχεια, συγκρίνετε τα αποτελέσματα με εκείνα της πλήρους χρονοσειράς και σχολιάστε πώς η παρουσία του outlier (COVID-19) επηρεάζει την ακρίβεια των προβλέψεων και πώς μεταβάλλεται η συσχέτιση μεταξύ των ετησίων αφίξεων και του κατά κεφαλήν ΑΕΠ;