



Μεθοδολογικό σημείωμα εφαρμογής εμπιστευτικότητας στα στοιχεία της Απογραφής Πληθυσμού – Κατοικιών 2021

Η Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ) στο πλαίσιο διασφάλισης του στατιστικού απορρήτου, σύμφωνα με την εθνική και ευρωπαϊκή νομοθεσία ([Νομικό πλαίσιο](#)) αλλά και την δέσμευσή της για τη διατήρηση της εμπιστοσύνης των απογραφόμενων της Απογραφής Πληθυσμού – Κατοικιών 2021, εφάρμοσε όλα τα κατάλληλα μέτρα προστασίας της εμπιστευτικότητας των στοιχείων, σύμφωνα με τις [συστάσεις](#) της Ευρωπαϊκής Στατιστικής Υπηρεσίας (Eurostat).

Συγκεκριμένα, στα πινακοποιημένα στοιχεία της Απογραφής Πληθυσμού – Κατοικιών 2021 εφαρμόστηκε η μέθοδος προσθήκης «θορύβου» σε κελιά (Cell Key Method (CKM)), ώστε αφενός να διασφαλίζεται η εμπιστευτικότητα των στοιχείων και αφετέρου να διατηρούνται τα χαρακτηριστικά τους και η χρησιμότητά τους, δεδομένου ότι η επίδραση της μεθόδου στα στοιχεία δεν είναι σημαντική.

Η ανωτέρω μέθοδος εμπιστευτικότητας διασφαλίζει το στατιστικό απόρρητο με την πρόσθεση ή αφαίρεση ενός μικρού αριθμού στις τιμές κελιών των πινάκων που προσδιορίζονται κατά την εφαρμογή της μεθόδου, με αποτέλεσμα ένα ποσοστό των δημοσιευμένων τιμών των μεταβλητών να διαφέρει ελαφρώς από τις πραγματικές τιμές. Σε ένα τυπικό σύνολο πινακοποιημένων στοιχείων, λιγότερο από το 20% των τιμών των κελιών έχει «διαταραχθεί» ελαφρώς, με τις πολύ μικρές τιμές να έχουν μεγαλύτερη πιθανότητα διαταραχής. Επισημαίνεται ότι τυχόν αρχικές μηδενικές τιμές των κελιών των πινάκων παραμένουν μηδενικές και μετά την εφαρμογή της μεθόδου.

Πιο συγκεκριμένα, κατά την εφαρμογή της μεθόδου, κάθε εγγραφή των πρωτογενών δεδομένων (microdata) λαμβάνει ένα τυχαίο κλειδί, γνωστό ως record key, που είναι ένας αριθμός με προκαθορισμένο εύρος. Αυτό το κλειδί είναι μοναδικό για κάθε στατιστική μονάδα (άτομο, νοικοκυριό, κατοικία) και δεν μεταβάλλεται ποτέ. Κατά την κατάρτιση πινάκων συχνότητας, το κάθε κελί που αντιπροσωπεύει έναν αριθμό στατιστικών μονάδων αντιστοιχίζεται με έναν μοναδικό αριθμό (cell key), του οποίου η τιμή υπολογίζεται ως το άθροισμα των record keys των στατιστικών μονάδων που ανήκουν σε αυτό το κελί. Η τιμή του cell key, σε συνδυασμό με έναν πίνακα πιθανοτήτων (p-table), προσδιορίζει τα κελιά στα οποία θα προστεθεί θόρυβος, καθώς και το μέγεθος του θορύβου. Εάν τα ίδια κελιά εμφανίζονται σε διαφορετικούς πίνακες, το ίδιο επίπεδο θορύβου θα εφαρμοστεί σταθερά, διατηρώντας τη συνέπεια

των στοιχείων. Ο θόρυβος είναι αρκετά μικρός ώστε να μην αλλοιώσει σημαντικά τα στοιχεία και αρκετά μεγάλος ώστε να προστατεύει από τους κινδύνους αποκάλυψης.

Σημειώνεται ότι η εφαρμογή της μεθόδου μπορεί να διαταράξει την προσθετικότητα (additivity) των τιμών στους πίνακες. Δηλαδή, το σύνολο μιας γραμμής ή στήλης που εμφανίζεται στον πίνακα δεν ισούται πάντα με το συνολικό άθροισμα των τιμών αυτής της γραμμής ή στήλης. Αυτό συμβαίνει επειδή η μέθοδος προσθέτει ή αφαιρεί θόρυβο στις τιμές κελιών ανεξάρτητα από τις άλλες τιμές. Για παράδειγμα, στον ακόλουθο πίνακα δημογραφικών στοιχείων αν αθροίσουμε τον αριθμό των ατόμων των 3 ομάδων ηλικιών, το άθροισμα διαφέρει ελαφρώς από το σύνολο που εμφανίζεται στην τελευταία γραμμή του πίνακα, επειδή σε κάθε ομάδα ηλικιών έχει εφαρμοστεί διαφορετική ποσότητα θορύβου.

Μόνιμος πληθυσμός κατά ομάδες ηλικιών	
	Σύνολο
Κάτω των 14 ετών	520
15-64 ετών	10082
65 ετών και άνω	688
Σύνολο	11292

Η μη προσθετικότητα των στοιχείων λόγω εφαρμογής της μεθόδου εμπιστευτικότητας κρίνεται αμελητέα, καθώς με την εφαρμογή της εξασφαλίζεται ότι δεν είναι δυνατή η ταυτοποίηση των στατιστικών μονάδων με μικρές συχνότητες και, παράλληλα, διατηρείται η συνολική χρησιμότητα και ακρίβεια των στοιχείων για ανάλυση και λήψη αποφάσεων.

Ως εκ τούτου, με την εφαρμογή της μεθόδου αυτής, η ΕΛΣΤΑΤ είναι σε θέση να διαθέσει στους χρήστες περισσότερους πίνακες στοιχείων της Απογραφής Πληθυσμού-Κατοικιών 2021 από ό,τι στο παρελθόν, και με αρκετά μεγαλύτερη ανάλυση, τηρώντας τη δέσμευσή της αφενός για την ικανοποίηση των αναγκών των χρηστών και αφετέρου για τη διασφάλιση του στατιστικού απορρήτου.

Επισημαίνεται ότι στα ήδη ανακοινωθέντα αποτελέσματα της Απογραφής Πληθυσμού-Κατοικιών έτους 2021, όπως έχουν δημοσιευθεί σε:

- ΦΕΚ (Μόνιμος και Νόμιμος Πληθυσμός)
- Δημοσιεύματα με infographics
- Πίνακες σε excel

τα οποία έχουν αναρτηθεί στην ιστοσελίδα της ΕΛΣΤΑΤ, στον σύνδεσμο <https://www.statistics.gr/el/2021-census-pop-hous>, θα υπάρχουν μικρές αποκλίσεις λόγω της εφαρμογής των κανόνων στατιστικής εμπιστευτικότητας.