

## **Έρευνα Ποιοτικών Χαρακτηριστικών Ημεδαπών Τουριστών 2018**

**ΕΙΔΟΣ:** Δειγματοληπτική έρευνα

### **Πλαίσιο Δειγματοληψίας**

Η Έρευνα Ποιοτικών Χαρακτηριστικών Ημεδαπών Τουριστών είναι μια ετήσια δειγματοληπτική έρευνα νοικοκυριών. Το δείγμα των νοικοκυριών για την έρευνα του έτους 2018 προκύπτει από το δείγμα της Έρευνας Εργατικού Δυναμικού η οποία είναι μία δισταδιακή στρωματοποιημένη δειγματοληπτική έρευνα με πρωτογενή μονάδα δειγματοληψίας την επιφάνεια (ένα ή περισσότερα συνεχόμενα οικοδομικά τετράγωνα η ολόκληρος οικισμός) και τελική μονάδα το νοικοκυριό. Η Έρευνα Εργατικού Δυναμικού είναι μια συνεχής έρευνα που παράγει τριμηνιαίες εκτιμήσεις με δείγμα που ανανεώνεται κατά το ένα έκτο ανά τρίμηνο και καλύπτει τον πληθυσμό των ιδιωτικών νοικοκυριών. Ο δειγματοληπτικός σχεδιασμός της Έρευνας Εργατικού Δυναμικού βασίστηκε στα στοιχεία της Απογραφής Πληθυσμού Κατοικιών 2001. Ο αριθμός των νοικοκυριών, καθώς και των ατόμων ανά νοικοκυριό προέκυψε κατά Περιφέρεια από τα στοιχεία της Απογραφής Πληθυσμού-Κατοικιών 2011.

Το δείγμα της Έρευνας Εργατικού Δυναμικού, αποτέλεσε το πλαίσιο για την επιλογή των ερευνώμενων δειγματοληπτικών νοικοκυριών της Έρευνας Ποιοτικών Χαρακτηριστικών Ημεδαπών Τουριστών 2018. Η χρήση του δείγματος της Έρευνας Εργατικού Δυναμικού για τη συλλογή των στατιστικών στοιχείων της Έρευνας Ποιοτικών Χαρακτηριστικών Ημεδαπών Τουριστών διευκολύνει στο σχεδιασμό του δείγματος και στη μείωση του κόστους κατάρτισης καταλόγων.

### **Δειγματοληπτική μονάδα**

Μονάδα έρευνας είναι το ιδιωτικό νοικοκυριό και όλα τα μέλη του.

### **Σχέδιο δειγματοληψίας**

Το δείγμα της Έρευνας Ποιοτικών Χαρακτηριστικών Ημεδαπών Τουριστών 2018, αποτελεί υποδείγμα του δείγματος νοικοκυριών της Έρευνας Εργατικού Δυναμικού των τριμήνων Q4\_2017-Q3\_2018, και επιλέχθηκε από τα νοικοκυριά της έρευνας τα οποία δεν ερευνούνταν πλέον.

Η Έρευνα Εργατικού Δυναμικού είναι μία δισταδιακή δειγματοληπτική έρευνα με πρωτογενή μονάδα την επιφάνεια (ένα ή περισσότερα οικοδομικά τετράγωνα), και δευτερογενή δειγματοληπτική μονάδα το νοικοκυριό.

Η συγκεκριμένη έρευνα είναι τριμηνιαία και το δείγμα της ανανεώνεται κατά 1/6 κάθε τρίμηνο. Ο πληθυσμός αναφοράς της συμπίπτει με εκείνον της Έρευνας Ποιοτικών Χαρακτηριστικών Ημεδαπών Τουριστών, ο σχεδιασμός βασίστηκε στα δεδομένα της Απογραφής Πληθυσμού Κατοικιών 2001 και πραγματοποιήθηκε προσαρμογή της βάσει των στοιχείων της Απογραφής Πληθυσμού Κατοικιών 2011.

Για την Έρευνα Ποιοτικών Χαρακτηριστικών Ημεδαπών Τουριστών, ως πρωτογενείς μονάδες θεωρούνται οι επιφάνειες (ένα ή περισσότερα οικοδομικά τετράγωνα) και δευτερογενείς μονάδες τα νοικοκυριά, εκ των οποίων ερευνούνται το σύνολο των ατόμων τα οποία διαμένουν σε αυτά.

### **Κριτήρια στρωμάτωσης του πληθυσμού**

Ο σχεδιασμός της έρευνας περιλαμβάνει δύο επίπεδα στρωματοποίησης α) το πρώτο στάδιο είναι η γεωγραφική στρωμάτωση, η οποία βασίζεται στον καταμερισμό της Χώρας σε 13 Περιφέρειες, ενώ η Περιφέρεια Πρωτεύουσας και το Πολεοδομικό συγκρότημα Θεσσαλονίκης αποτελούν ιδιαίτερα γεωγραφικά στρώματα, β) Το δεύτερο στάδιο στρωμάτωσης περιλαμβάνει την ομαδοποίηση των δημοτικών και τοπικών κοινοτήτων σε κάθε Περιφέρεια και ως εκ τούτου και των επιφανειών που ανήκουν σε αυτές, κατά βαθμό αστικότητας σε 8 κατηγορίες ως ακολούθως

Αστικές περιοχές 50.000 κατοίκων και άνω
Αστικές περιοχές 30.000 – 49.999 κατοίκων
Αστικές περιοχές 10.000 – 29.999 κατοίκων
Αστικές περιοχές 5.000 – 9.999 κατοίκων
Αστικές περιοχές 2.000 – 4.999 κατοίκων
Αγροτικές περιοχές 1.000 – 1.999 κατοίκων
Αγροτικές περιοχές 500 – 999 κατοίκων
Αγροτικές περιοχές 1 – 499 κατοίκων

Ο συνολικός αριθμός των στρωμάτων στις 13 Περιφέρειες ανέρχεται σε 93, ενώ οι Περιφέρεια Πρωτεύουσας και το Πολεοδομικό Συγκρότημα Θεσσαλονίκης χωρίστηκαν σε 42 και 12 υποστρώματα αντίστοιχα, με βάση τα οικοδομικά τετράγωνα από τα οποία αποτελούνται οι Δήμοι. Ο συνολικός αριθμός των στρωμάτων ανήλθε σε 147.

### **Μέγεθος και κριτήρια κατανομής του δείγματος**

Το μέγεθος του ετήσιου δείγματος ανήλθε σε 8.705 νοικοκυριά που ανήκουν σε 1.118 πρωτογενείς μονάδες δειγματοληψίας.

### **1<sup>ο</sup> Στάδιο Δειγματοληψίας**

Στο 1<sup>ο</sup> στάδιο δειγματοληψίας, σε κάθε (τελικό) στρώμα  $h$ , επιλέχθηκαν  $a_h$  πρωτογενείς μονάδες με πιθανότητα ανάλογη του μεγέθους τους. Ο αριθμός  $a_h$  των επιλεγέντων πρωτογενών μονάδων είναι προσεγγιστικά ανάλογος του μεγέθους του στρώματος  $N_h$  στον πληθυσμό, όπως ορίστηκε παραπάνω.

Η πρωτογενής μονάδα  $i$  στο στρώμα  $h$  έχει πιθανότητα επιλογής ανάλογη του μεγέθους του πληθυσμού ως ακολούθως:

$$P_{hi} = \frac{N_{hi}}{N_h} \quad (1)$$

όπου:

$N_{hi}$ : είναι το πλήθος των νοικοκυριών στην πρωτογενή μονάδα  $i$  του στρώματος  $h$

### **2<sup>ο</sup> στάδιο δειγματοληψίας**

Στο 2<sup>ο</sup> στάδιο δειγματοληψίας, στην  $hi$  πρωτογενή μονάδα, επιλέχθηκε με ίσες πιθανότητες επιλογής ένα δείγμα  $n_{hi}$  από  $N_{hi}$  νοικοκυριά. Κάθε ένα από τα  $n_{hi}$  νοικοκυριά είχε την ίδια πιθανότητα να επιλεγεί, ίση με:

$$\frac{n_{hi}}{N_{hi}} \quad (2)$$

Ο συνολικός αριθμός νοικοκυριών προς διερεύνηση των  $a_h$  πρωτογενών

δειγματοληπτικών μονάδων είναι:  $n_h = \sum_{i=1}^{a_h} n_{hi}$

Μέσα σε κάθε πρωτογενή δειγματοληπτική μονάδα πραγματοποιήθηκε ο υπολογισμός του κλάσματος δειγματοληψίας  $\delta_{hi} = \frac{N_{hi}}{n_{hi}}$ , έτσι ώστε οι παρακάτω δύο επιθυμητές συνθήκες να ικανοποιούνται:

α) Η αναμενόμενη τιμή του κλάσματος  $\frac{n_h}{N_h}$  να είναι σταθερή σε κάθε στρώμα. Δηλαδή:

$$E\left(\frac{n_h}{N_h}\right) = \frac{1}{\lambda} = 2,11\% \quad (3)$$

και

β) Ο εκτιμητής του συνόλου σε επίπεδο στρώματος  $Y_h$  (για κάθε ένα χαρακτηριστικό) να είναι αυτοσταθμιζόμενος. Δηλαδή, η εκτίμηση των χαρακτηριστικών της έρευνας να προκύπτει ως το άθροισμα του γινομένου των τιμών των χαρακτηριστικών στα  $n_h$  επιλεγέντα νοικοκυριά επί το συντελεστή  $\lambda$ , που είναι ο ίδιος σε κάθε στρώμα.

Οι συνθήκες (α) και (β) ικανοποιούνται όταν:

$$\frac{1}{a_h} \cdot \frac{1}{P_{hi}} \cdot \frac{N_{hi}}{n_{hi}} = \lambda \quad (4)$$

$$\Rightarrow \frac{1}{a_h} \cdot \frac{1}{P_{hi}} \cdot \delta_{hi} = \lambda \Rightarrow \delta_{hi} = \frac{N_{hi}}{n_{hi}} = \lambda \cdot a_h \cdot P_{hi} \quad (5)$$

Από τις σχέσεις (1) και (5) προκύπτει

$$\begin{aligned} \frac{N_{hi}}{n_{hi}} &= \lambda \cdot a_h \cdot \frac{N_{hi}}{N_h} \Rightarrow n_{hi} = \frac{N_{hi} \cdot N_h}{\lambda \cdot a_h \cdot N_{hi}} \\ \Rightarrow n_{hi} &= \frac{N_h}{\lambda \cdot a_h} \end{aligned} \quad (6)$$

Από τη σχέση (3), προκύπτει ότι:

$$\frac{1}{\lambda} = \frac{n_h}{N_h} \Rightarrow \lambda = \frac{N_h}{n_h} \quad (7)$$

Από τις σχέσεις (6) και (7), έχουμε ότι:

$$n_{hi} = \frac{n_h}{a_h} \quad (8)$$

## Εκτίμηση των χαρακτηριστικών της έρευνας

Έστω ότι  $h$  είναι το τελικό στρώμα των νοικοκυριών (Τελικό στρώμα = Περιφέρεια x Βαθμός αστικότητας), τότε αυτό θα λάβει τις ακόλουθες τιμές :  $h = 1, 2, \dots, H$  (όπου  $H = 147$ ). Σε κάθε ένα από τα τελικά στρώματα (έστω  $h$ ), εάν η στατιστική πληροφορία συλλέχθηκε από ένα δείγμα  $n'_h$  νοικοκυριών, ο αναγωγικός συντελεστής του  $j$  νοικοκυριού που ανήκει στη  $i$  πρωτογενή μονάδα, ορίζεται ως:

$$w_{hij} = \frac{N_h}{a_h \cdot N_{hi}} \cdot \frac{N_{hi}}{n_{hi}} \cdot \frac{1}{r_h} \cdot t_{hij} = \frac{N_h}{a_h \cdot n_{hi}} \cdot \frac{1}{r_h} \cdot t_{hij} \quad (9)$$

Από τις σχέσεις (8) και (9), ισχύει:

$$w_{hij} = \frac{N_h}{a_h \cdot \frac{n_h}{a_h}} \cdot \frac{1}{r_h} \cdot t_{hij} \Rightarrow w_{hij} = \frac{N_h}{n_h} \cdot \frac{1}{r_h} \cdot t_{hij} \quad (10)$$

όπου:

$N_h$ : μέγεθος το πληθυσμού στο στρώμα  $h$  σύμφωνα με την Απογραφή Πληθυσμού Κατοικιών 2011 και τις προβολές πληθυσμού

$n_h$ : αρχικό δείγμα στο στρώμα  $h$

$\frac{N_h}{n_h}$ : αντίστροφη πιθανότητα επιλογής του επιλεγέντος δείγματος νοικοκυριών στο

στρώμα  $h$ , έτσι ώστε ο εκτιμητής του συνόλου  $Y_h$ , (για κάθε χαρακτηριστικό) στο στρώμα να είναι αυτοσταθμιζόμενος,

$r_h = \frac{n'_h}{n_h}$  είναι το αντίστροφο ποσοστό ανταπόκρισης στο στρώμα  $h$

$t_{hij}$ : συντελεστής προσαρμογής των αναγωγικών συντελεστών των νοικοκυριών, ούτως ώστε η δειγματική κατανομή τους να συμφωνεί με εκείνη του πληθυσμού σε επίπεδο Μεγάλης Γεωγραφικής Περιοχής (NUTS 1). Οι τάξεις οι οποίες χρησιμοποιήθηκαν ήταν το σύνολο των νοικοκυριών κατά τάξη μεγέθους νοικοκυριού (1, 2, 3, 4+ μέλη) και το σύνολο των ατόμων κατά φύλο και τάξη ηλικίας (0-14, 15-24, 25-34, 35-44, 45-54, 55-

64, 65-74, 75+), σύμφωνα με τα στοιχεία των προβολών πληθυσμού και των στοιχείων της Απογραφής Πληθυσμού Κατοικιών 2011.

Έστω ότι  $w_{hijk} (>0)$  είναι ο αναγωγικός συντελεστής που αντιστοιχεί στο άτομο  $k$  ( $k = 1, \dots, m_{hij}$ ) που ανήκει στο  $j$  νοικοκυριό ( $j = 1, \dots, n_{hi}$ ) που ερευνήθηκε και το οποίο ανήκει στην πρωτογενή δειγματοληπτική μονάδα  $i$ , του στρώματος  $h$ .

Έστω ότι  $y_{hijk}$  είναι το χαρακτηριστικό  $y$  του ατόμου  $k$  του νοικοκυριού  $j$ , που ανήκει στην πρωτογενή δειγματοληπτική  $hi$  (ένα ή περισσότερα οικοδομικά τετράγωνα). Επιπλέον,  $Y$  είναι το χαρακτηριστικό στο σύνολο του πληθυσμού, το οποίο προκύπτει από την άθροιση των χαρακτηριστικού  $y$  όλων των ατόμων που ανήκουν σε όλα τα στρώματα  $h$ . Συνεπώς, η μορφή του εκτιμητή της διασταδιακής έρευνας είναι:

$$Y = \sum_{h=1}^H \sum_{i=1}^{a_h} \sum_{j=1}^{n_{hi}} \sum_{k=1}^{m_{hij}} w_{hijk} \cdot y_{hijk} \quad (11)$$