




PIAAC GREECE
Σχέδιο δειγματοληψίας
Κύριας Έρευνας (MS)

Ιωάννης Νικολαΐδης,
Ελληνική Στατιστική Αρχή
Προϊστάμενος του Τμήματος Μεθοδολογίας,
Ανάλυσης και Μελετών
e-mail: giannikol@statistics.gr



1. Ερευνώμενος πληθυσμός (πληθυσμός στόχος)

- Ενήλικα άτομα, ηλικίας 16 με 65 ετών, τα οποία διαμένουν σε ιδιωτικά νοικοκυριά («νοικοκυριά στόχος»)



2. Σχεδιασμός της κύριας έρευνας (MS)

2.1. Μέθοδος δειγματοληψίας

- Η έρευνα είναι δειγματοληπτική και εφαρμόζεται η πολυσταδιακή στρωματοποιημένη δειγματοληψία

Μονάδες έρευνας

- Πρωτογενής μονάδα έρευνας είναι η επιφάνεια (ένα ή περισσότερα οικοδομικά τετράγωνα ή ένας μικρός οικισμός)
- Δευτερογενής μονάδα έρευνας είναι το νοικοκυριό (με τουλάχιστον ένα άτομο ηλικίας 16-65 ετών: νοικοκυριό στόχος)
- Τελική μονάδα έρευνας είναι το άτομο ηλικίας 16 με 65 ετών



2.2. Στρωματοποίηση του πληθυσμού

- Για την αντιπροσωπευτική επιλογή των μονάδων της έρευνας, ο πληθυσμός κατανέμεται σε ομοιογενείς υπο-πληθυσμούς (στρώματα), με τη χρήση των παρακάτω μεταβλητών στρωμάτωσης



Μεταβλητές στρωμάτωσης:

α) Γεωγραφικές περιφέρειες (NUTS 2)

Χρησιμοποιούνται 15 γεωγραφικές Περιφέρειες, σύμφωνα με τον ακόλουθο πίνακα

Πίνακας 1.α : Γεωγραφικές Περιφέρειες

Περιφέρειες
Ανατολική Μακεδονία και Θράκη
Κεντρική Μακεδονία (εκτός του Πολεοδομικού Συγκροτήματος Θεσσαλονίκης)
Δυτική Μακεδονία
Θεσσαλία
Ήπειρος
Ιόνια Νησιά
Δυτική Ελλάδα
Κεντρική Ελλάδα
Πελοπόννησος
Αττική (εκτός της Περιφέρειας Πρωτεύουσας)
Πρώην Περιφέρεια Πρωτεύουσας
Πρώην Πολεοδομικό Συγκρότημα Θεσσαλονίκης



Μεταβλητές στρωμάτωσης:

α) Γεωγραφικές περιφέρειες (NUTS 2)

Χρησιμοποιούνται 15 γεωγραφικές Περιφέρειες, σύμφωνα με τον ακόλουθο πίνακα

Πίνακας 1.β : Γεωγραφικές Περιφέρειες

Περιφέρειες
Νησιά Βορείου Αιγαίου
Νησιά Νοτίου Αιγαίου
Κρήτη

Μεταβλητές στρωμάτωσης:

β) Βαθμοί Αστικότητας

Σε κάθε Περιφέρεια οι μονάδες της έρευνας κατανέμονται σε 3 στρώματα με βάση την αστικότητα των Δημοτικών / Τοπικών Κοινοτήτων, όπου αυτές ανήκουν.

Πίνακας 2 : Βαθμοί Αστικότητας

Αστικότητα	Κωδικός	Περιγραφή
Αστικές	1	Δημοτικές Κοινότητες (ΔΚ) με πληθυσμό μεγαλύτερο ή ίσο των 50.000 κατοίκων*
	2	ΔΚ με πληθυσμό 30.000-49.999 κατοίκων*
	3	ΔΚ με πληθυσμό 10.000-29.999 κατοίκων*
Ημιαστικές	4	Δημοτικές και Τοπικές Κοινότητες με πληθυσμό 5.000-9.999 κατοίκους*
	5	Δημ. /Τοπ. Κοιν. με πληθυσμό 2.000-4.999 κατοίκους*
Αγροτικές	6	Τοπικές Κοινότητες με πληθυσμό 1000-1999 κατοίκους*
	7	Τοπικές Κοινότητες με πληθυσμό 500-999 κατοίκους*
	8	Τοπικές Κοινότητες με πληθυσμό μέχρι 499 κατοίκους*

*Διευκρίνιση: Δημοτικές και Τοπικές Κοινότητες των οποίων ο πολυπληθέστερος οικισμός αποτελείται από τον αναγραφόμενο πληθυσμό, στους παραπάνω βαθμούς Αστικότητας



Μεταβλητές στρωμάτωσης:

γ) Αστικές Περιοχές Αθήνας & Θεσσαλονίκης

- Η Πρώην Περιφέρεια Πρωτεύουσας χωρίστηκε σε 48 στρώματα ίσου περίπου μεγέθους (ίσος αριθμός νοικοκυριών) με τα οικοδομικά τετράγωνα των Δήμων ή Δημοτικών Κοινοτήτων και λαμβάνοντας υπόψη κοινωνικοοικονομικά κριτήρια
- Το Πρώην Πολεοδομικό Συγκρότημα Θεσσαλονίκης χωρίστηκε σε 11 ίσου μεγέθους στρώματα
- Το πλήθος των τελικών στρωμάτων, τα οποία προέκυψαν από την εφαρμογή όλων των παραπάνω μεταβλητών στρωματοποίησης του πληθυσμού, ανέρχεται στα 151



2.3.1 Κλάσμα δειγματοληψίας Κύριας Έρευνας (MS)

Η έρευνα θα διενεργηθεί σε δείγμα 10.364 «νοικοκυριών στόχων» (κλάσμα δειγματοληψίας $1/\lambda = 3,09 \text{ ‰}$) προκειμένου να εξασφαλίσουμε 5.000 συνεντεύξεις που είναι ο στόχος της έρευνας

Ακολουθεί πίνακας με την κατανομή του δείγματος ανά Περιφέρεια και βαθμό αστικότητας.

2.3.1 Κατανομή θεωρητικού δείγματος νοικοκυριών (eligible & μη-eligible) ανά Περιφέρεια

Region Code	Region Description	Theoretical Sample of Eligible HHs	Percentage of Theor. Sample Elig. HHs (%)	Theoretical Sample of All HHs (eligible+non eligible)	Percentage of Theor. Sample of all HHs (%)
5	Greater Athens Area (Urban region)	3283	31.67	3955	30.90
6	Greater Thessaloniki Area (Urban region)	844	8.14	1013	7.91
11	Eastern Macedonia & Thrace	550	5.31	708	5.53
12	Central Macedonia	936	9.03	1201	9.38
13	Western Macedonia	251	2.42	322	2.52
14	Thessaly	636	6.14	802	6.27
21	Epirus	308	2.97	397	3.11
22	Ionian Islands	199	1.92	250	1.95
23	Western Greece	608	5.86	750	5.86
24	Central Greece	480	4.63	618	4.83
25	Peloponnesus	514	4.96	660	5.15
32	Attica	622	6.00	725	5.67
41	Islands of Northern Aegean	221	2.13	288	2.25
42	Islands of Southern Aegean	303	2.92	361	2.82
43	Crete	611	5.89	748	5.84
	TOTAL	10364	100.00	12799	100.00

2.3.1 Κατανομή θεωρητικού δείγματος νοικοκυριών (eligible & μη-eligible) ανά Αστικότητα

Region Code	URBANISATION	Theoretical Sample of Eligible HHs	Th. Sample Eligible HHs %	Theoretical Sample of All HHs	Th. Sample of all HHs %
1	Urban Regions	6892	66.5	8224	64.3
2	Semi-Urban Regions	1375	13.3	1674	13.1
3	Rural Regions	2097	20.2	2902	22.7
	TOTAL	10364	100.0	12800	100.0



2.4. Ακρίβεια αποτελεσμάτων (MS)

- Σε σύνολο Χώρας και χωριστά για κάθε φύλο και ομάδες ηλικιών, για τα χαρακτηριστικά με συχνότητα εμφάνισης $\geq 8\%$ ($p \geq 8\%$), το σχετικό σφάλμα δειγματοληψίας είναι $\leq 10\%$ ($CV \leq 10\%$)



3. Διαδικασία επιλογής του δείγματος

1ο στάδιο δειγματοληψίας (επιλογή μονάδων επιφάνειας)

- Έστω ότι συμβολίζουμε με h το στρώμα $h = 1, \dots, 151$. Από κάθε στρώμα επιλέγονται n_h πρωτογενείς μονάδες (επιφάνειες)
- Ο αριθμός των πρωτογενών μονάδων σε κάθε στρώμα είναι τέτοιος, ώστε σε κάθε επιφάνεια να ερευνώνται 13-20 νοικοκυριά
- Σε κάθε στρώμα, η πρωτογενής μονάδα του δείγματος επιλέγεται με πιθανότητα ανάλογη του μεγέθους της (αριθμός νοικοκυριών «στόχος», σύμφωνα με την Απογραφή Πληθυσμού 2011)
- Ο συνολικός αριθμός των πρωτογενών μονάδων του δείγματος ανέρχεται σε 775 επιφάνειες, για το MS.



3. Διαδικασία επιλογής του δείγματος

1ο στάδιο δειγματοληψίας (συνέχεια)

- Εάν X_{hi} είναι ο αριθμός των νοικοκυριών «στόχος» της μονάδας του δείγματος τάξης i ($i=1, 2, \dots, n_h$) και X_h ο αριθμός των νοικοκυριών του στρώματος h , σύμφωνα με την Απογραφή Πληθυσμού έτους 2011, τότε η πιθανότητα επιλογής της μονάδας επιφάνειας είναι :

$$P_{hi} = \frac{X_{hi}}{X_h} \quad (1)$$



3. Διαδικασία επιλογής του δείγματος

2ο στάδιο δειγματοληψίας (επιλογή νοικοκυριών)

- Σε κάθε επιλεγείσα πρωτογενή δειγματοληπτική μονάδα (μονάδα επιφάνειας) επιλέγεται δείγμα νοικοκυριών, από ενημερωμένο κατάλογο νοικοκυριών (πλαίσιο δειγματοληψίας). Στην πραγματικότητα, στο δεύτερο στάδιο επιλέγεται δείγμα κατοικιών. Εντούτοις, στις περισσότερες περιπτώσεις, υπάρχει μία προς μία αντιστοίχιση μεταξύ νοικοκυριού και κατοικίας. Εάν η επιλεγείσα κατοικία αποτελείται από περισσότερα του ενός νοικοκυριά, τότε ερευνώνται όλα τα νοικοκυριά.

3. Διαδικασία επιλογής του δείγματος: Υπολογισμός νοικοκυριών δείγματος

Έστω M_{hi} ο αριθμός των κατοικιών του ενημερωμένου καταλόγου-πλασίου της έρευνας στην i επιλεγείσα μονάδα επιφανείας στο στρώμα h . Από αυτά, ένα συστηματικό δείγμα m_{hi} κατοικιών επιλέγεται με ίσες πιθανότητες. Για τον καθορισμό του μεγέθους του δείγματος νοικοκυριών-κατοικιών m_{hi} ακολουθούμε τα εξής βήματα:

- Αρχικά υπολογίζεται το διάστημα δειγματοληψίας $\delta_{hi} = \frac{M_{hi}}{m_{hi}}$

- Κατόπιν: $m_{hi} = \frac{M_{hi}}{\delta_{hi}}$
Από το δείγμα των m_{hi} νοικοκυριών, μόνο σε όσα περιλαμβάνουν τουλάχιστον ένα άτομο ηλικία 16-65 ετών, θα πραγματοποιηθεί η έρευνα.



3. Διαδικασία επιλογής του δείγματος: Υπολογισμός νοικοκυριών δείγματος (διάστημα δειγματοληψίας)

- Το διάστημα δειγματοληψίας υπολογίζεται ως εξής

$$\delta_{hi} = \frac{M_{hi}}{m_{hi}} = \lambda \cdot n_h \cdot P_{hi} \quad (2)$$

Το λ είναι ο αντίστροφος αριθμός του κλάσματος δειγματοληψίας των νοικοκυριών, όπου είναι **σταθερός** και ίσος με 322.99 και n_h είναι ο αριθμός των μονάδων επιφανείας του στρώματος h .

Αυτό σημαίνει **αυτοσταθμιζόμενο** δείγμα νοικοκυριών, δηλαδή στα νοικοκυριά του δείγματος δίνεται ίση πιθανότητα επιλογής (ίση ευκαιρία να επιλεγούν).



3. Διαδικασία επιλογής του δείγματος

3ο στάδιο δειγματοληψίας (επιλογή ατόμων)

- Σε κάθε νοικοκυριό «στόχος» του δείγματος επιλέγεται ένα άτομο, με ίσες πιθανότητες, από κατάλογο μελών του νοικοκυριού με ηλικίες 16-65 ετών. Η πιθανότητα επιλογής δίνεται από:

$$P_{hijk} = \frac{1}{M_{hij}}$$

Όπου M_{hij} είναι το πλήθος των «eligible» ατόμων στο j νοικοκυριό, στρώματος h , μονάδας επιφανείας i



4. Στάθμιση των ατόμων και νοικοκυριών του δείγματος

Για την εκτίμηση των χαρακτηριστικών της έρευνας, τα στοιχεία κάθε ατόμου σταθμίζονται με κατάλληλο συντελεστή (αναγωγικό συντελεστή), που προκύπτει ως το γινόμενο των ακόλουθων τριών παραγόντων (α , β , γ):

4. Στάθμιση των ατόμων και νοικοκυριών του δείγματος: Παράγοντας α.

α. Αντίστροφη πιθανότητα επιλογής του ατόμου ($hijk$)

$$DW_{hijk} = \frac{1}{n_h \cdot P_{hi}} \cdot \frac{M_{hi}}{m_{hi}} \cdot \frac{M_{hij}}{m_{hij}} \quad (3)$$

όπου

n_h : το μέγεθος του δείγματος των πρωτογενών μονάδων στο στρώμα h

P_{hi} : η πιθανότητα επιλογής της πρωτογενούς μονάδας επιφάνειας hi

M_{hi} : ο αριθμός των νοικοκυριών στον ενημερωμένο κατάλογο-πλαίσιο της πρωτογενούς μονάδας επιφάνειας hi

m_{hi} : ο αριθμός των επιλεγμένων νοικοκυριών στην πρωτογενή μονάδα επιφάνειας hi

M_{hij} : ο αριθμός των μελών ηλικίας άνω των 6 μηνών του επιλεγμένου νοικοκυριού hij .

m_{hij} : ο αριθμός των επιλεγμένων ατόμων στο νοικοκυριό hij ($m_{hij}=1$)

Κάθε παράγοντας του ανωτέρω γινομένου αντιστοιχεί στις αντίστροφες πιθανότητες επιλογής κάθε μονάδας δειγματοληψίας (επιφάνεια, νοικοκυριό, άτομο). Το γινόμενο των δύο πρώτων παραγόντων είναι το αντίστροφο της δεσμευμένης πιθανότητας επιλογής του νοικοκυριού που είναι σταθερό και ίσο με 322,99 επειδή το δείγμα είναι αυτοσταθμιζόμενο.



4. Στάθμιση των ατόμων και νοικοκυριών του δείγματος: Παράγοντας β.

β. Αντίστροφος του ποσοστού απόκρισης των νοικοκυριών του στρώματος h

Το ποσοστό απόκρισης των νοικοκυριών του δείγματος στο στρώμα h :

$$r_h = \frac{m'_h}{m_h}$$

Όπου:

m_h : αριθμός νοικοκυριών αρχικού δείγματος

m'_h : αριθμός νοικοκυριών που αποκρίθηκαν στην έρευνα

4. Στάθμιση των ατόμων και νοικοκυριών του δείγματος: Παράγοντας γ.

γ. Διορθωτικός συντελεστής t_{hijk}

- Διορθωτικός συντελεστής, που προσαρμόζει τις σταθμίσεις του δείγματος των ατόμων κατά φύλο, ομάδες ηλικιών και επίπεδο εκπαίδευσης, έτσι ώστε η εκτίμηση της κατανομής των ατόμων να συμφωνεί με την αντίστοιχη κατανομή του πληθυσμού στο σύνολο Χώρας.
- Οι κατανομές είναι 10 και ορίζονται από το φύλο και την ηλικιακή ομάδα (2 φύλα x 8 ηλικιακές ομάδες). Οι ηλικιακές ομάδες ορίζονται από τα χρονικά διαστήματα των ετών: 16-24, 25-34, 35-44, 45-54, 55-65.
- Η κατανομή του πληθυσμού των ατόμων κατά φύλο και ηλικιακή ομάδα υπολογίζεται με προβολή για την περίοδο αναφοράς της έρευνας, με βάση τα στοιχεία της φυσικής κίνησης πληθυσμού (Απογραφή Πληθυσμού 2011, Γεννήσεις, Θάνατοι, Μετανάστευση)
- Επίπεδο Εκπαίδευσης (6 κατηγορίες).



4. Στάθμιση των ατόμων του δείγματος: Υπολογισμός Αναγωγικού Συντελεστή

Κατόπιν των ανωτέρω, ο αναγωγικός συντελεστής κάθε ατόμου του δείγματος υπολογίζεται από τη σχέση:

$$w_{hijk} = DW_{hijk} \cdot \frac{1}{r_h} \cdot t_{hijk} \quad (4)$$



4.1 Αναμενόμενα προβλήματα & Αντιμετωπίσεις

- A. Η μη απόκριση θα αντιμετωπιστεί με την επιλογή μεγαλύτερου δείγματος
- B. Η μη απόκριση επιδρά στην ακρίβεια των αποτελεσμάτων (δημιουργία μεροληψίας, αύξηση σφάλματος δειγματοληψίας)
- C. Μελέτη χρειάζεται για την επίδραση της μη απόκρισης επί της ακρίβειας των αποτελεσμάτων.



4.1 Προβλήματα & Αντιμετωπίσεις (συνέχεια)

Συγκεκριμένα, η μη απόκριση επιδρά στην ακρίβεια των αποτελεσμάτων ως εξής:

- Δημιουργία μεροληψίας: το αρχικό επιλεγμένο νοικοκυριό ενδεχομένως να μην έχει τα ίδια χαρακτηριστικά (μέγεθος νοικοκυριού, σύνθεση μελών νοικοκυριού κ.ά.) με το νέο νοικοκυριό (από αντικατάσταση)
- Αύξηση σφάλματος δειγματοληψίας: λόγω επιπλέον σταθμίσεων στα άτομα του δείγματος, οι οποίες επιβάλλονται από την ανάγκη για μείωση της μεροληψίας.
Οι επιπλέον αυτές σταθμίσεις έχουν τυχαίο χαρακτήρα και ως εκ τούτου αυξάνουν τη διακύμανση των αποτελεσμάτων και συνεπώς το σφάλμα δειγματοληψίας.



4.1 Προβλήματα & Αντιμετωπίσεις (συνέχεια)

- Η μελέτη για την επίδραση της μη απόκρισης στην ακρίβεια των αποτελεσμάτων περιλαμβάνει τα εξής:
 - Εκτίμηση της μεροληψίας (έλεγχος της διαφοράς μεταξύ αποκρινόμενων και μη αποκρινόμενων νοικοκυριών π.χ. με βάση τη διαθέσιμη μεταβλητή «μέγεθος νοικοκυριού»)
 - Υπολογισμός του ποσοστού αύξησης του σφάλματος δειγματοληψίας λόγω των επιπρόσθετων σταθμίσεων που εφαρμόστηκαν για τη μείωση της μεροληψίας
 - Υπολογισμός του intra-cluster correlation των βασικών χαρακτηριστικών των ατόμων του δείγματος, που δηλώνει το βαθμό ομοιογένειας των ατόμων ως προς το συγκεκριμένο χαρακτηριστικό εντός των μονάδων επιφανείας. Στόχος είναι το intra-cluster correlation να είναι $\leq 5\%$, προκειμένου να εμφανίζεται καλή διασπορά του δείγματος.



Ευχαριστούμε πολύ
για την παρακολούθηση!