

ΚΕΦΑΛΑΙΟ 9

ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΑΚΡΙΒΕΙΑΣ ΤΩΝ ΔΗΜΟΓΡΑΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Τα δημογραφικά δεδομένα τα οποία προέρχονται από τις απογραφές πληθυσμού, τις καταγραφές της φυσικής και μεταναστευτικής κίνησης του πληθυσμού και τις διάφορες δειγματοληπτικές έρευνες είναι συχνά επιφορτισμένα με σφάλματα. Ακόμα και σε χώρες με παράδοση σε αξιόπιστα συστήματα συλλογής δημογραφικών στοιχείων, τα στοιχεία αυτά παρότι γνωστά για την ακρίβεια και την πληρότητα τους, εμπεριέχουν πηγές σφαλμάτων. Τα σφάλματα αυτά μπορεί να είναι τυχαία ή συστηματικά. Η ένταση αυτού του προβλήματος συνδέεται στενά τόσο και με την αξιοπιστία των συστημάτων συλλογής αυτών των στοιχείων όσο και με το εκπαιδευτικό επίπεδο του πληθυσμού από τον οποίο προέρχονται.

Για τον υπολογισμό των διαφόρων δημογραφικών μέτρων, όπως αυτά παρουσιάστηκαν στα προηγούμενα κεφάλαια, βασική προϋπόθεση αποτελεί η ύπαρξη πλήρων και ακριβών στατιστικών στοιχείων. Έτσι πριν την οποιαδήποτε αξιοποίηση των στοιχείων αυτών είναι απαραίτητος ο έλεγχος της ποιότητας τους και όσο αυτό είναι δυνατό, η διόρθωση τους.

Για τις ανάγκες μιας τέτοιας διερεύνησης δεν υπάρχει κάποια μοναδική, γενική και κοινά αποδεκτή, διαδικασία. Η αποτελεσματικότητα της οποιας διερεύνησης συνδέεται στενά με την εμπειρία του ερευνητή και την γνώση του όσο αφορά τις ιδιαιτερότητες που εμφανίζουν τόσο τα διάφορα δεδομένα όσο και τα αντίστοιχα συστήματα συλλογής τους. Στη συνέχεια παρατίθενται κάποιες βασικές κατευθυντήριες γραμμές για τη διαδικασία μιας τέτοιας διερεύνησης και γίνεται αναφορά σε κάποιες βασικές τεχνικές για τον έλεγχο της πληρότητας και της ακρίβειας των δημογραφικών δεδομένων που έχουν αναφερθεί στη δημογραφική βιβλιογραφία.

9.1 Πηγές σφαλμάτων

Τα διάφορα σφάλματα που εμπεριέχονται στα στατιστικά δεδομένα μπορεί να θεωρηθεί ότι χωρίζονται σε τρεις βασικές κατηγορίες:

1. Δειγματοληπτικά σφάλματα,
2. Σφάλματα κάλυψης,
3. Σφάλματα μη-απάντησης ή εσφαλμένης απάντησης.

Η πρώτη κατηγορία σφαλμάτων αφορά μόνο τις διάφορες δειγματοληπτικές έρευνες. Τα σφάλματα αυτής της κατηγορίας μπορεί να είναι τυχαία ή συστηματικά.

Τα τυχαία σφάλματα ενυπάρχουν σε όλες τις δειγματοληπτικές έρευνες και οφείλονται στο γεγονός ότι ένα τυχαίο δείγμα ποτέ δεν αποτελεί μια ακριβή μικρογραφία του πληθυσμού από τον οποίο προέρχεται αλλά η σύνθεση του εκφράζει πάντοτε κάποιες τυχαίες αποκλίσεις από εκείνη του πληθυσμού αναφοράς. Τα συστηματικά σφάλματα συνδέονται τόσο με τον σχεδιασμό του δείγματος στο οποίο εφαρμόζεται η έρευνα όσο και με την διαδικασία διεξαγωγής της έρευνας. Όσο καλύτερα σχεδιασμένο είναι το δείγμα, όσο περισσότερο δηλαδή η σύνθεση του προσεγγίζει εκείνη του πληθυσμού από τον οποίο προέρχεται τόσο καλύτερα αποτυπώνει τις ιδιότητες του πληθυσμού.

Ακόμα, όσο επιμελέστερη είναι η διεξαγωγή της έρευνας πεδίου τόσο περισσότερο ακριβής και αξιόπιστη είναι η πληροφορία η οποία συλλέγεται.

Η δεύτερη και η τρίτη κατηγορία σφαλμάτων αφορά τόσο τις δειγματοληπτικές όσο και τις ολικές έρευνες (απογραφή και καταγραφές φυσικής και μεταναστευτικής κίνησης του πληθυσμού), το σύνολο δηλαδή των βασικών πηγών συλλογής δημογραφικών δεδομένων.

Τα **σφάλματα κάλυψης** της δεύτερης κατηγορίας συνδέονται στις δειγματοληπτικές έρευνες με υποεκπροσώπηση ή έλλειψη εκπροσώπησης βασικών υποσυνόλων του πληθυσμού και τις καταγραφές φυσικής και μεταναστευτικής κίνησης με διαφυγές ή περισσότερο της μιας εκπροσωπήσεις συμβάντων ή ατόμων. Ειδικά στις καταγραφές φυσικής και μεταναστευτικής κίνησης ο αριθμός διπλοκαταγραφών είναι ασήμαντα χαμηλός αλλά οι διαφυγές αποτελούν πηγές έντονου συστηματικού σφάλματος. Στις καταγραφές φυσικής κίνησης ιδιαίτερα, μια συστηματική πηγή διαφυγών συνδέεται με τα νεογνά που πεθαίνουν σε μικρό χρονικό διάστημα μετά τη γέννηση τους και τα οποία δεν δηλώνονται ούτε σαν γέννηση ζώντος ούτε σαν μετέπειτα θάνατος, με αποτέλεσμα να διαφεύγουν τόσο από τις γεννήσεις όσο και από τους θανάτους της ηλικίας μηδέν.

Σοβαρή συνέπεια αυτού του προβλήματος είναι η εσφαλμένη εκτίμηση των μέτρων γεννητικότητας και νεογνικής θνησιμότητας που υπολογίζονται βάσει στοιχείων φυσικής κίνησης του πληθυσμού. Στα στοιχεία φυσικής κίνησης του Ελληνικού πληθυσμού αυτό το πρόβλημα εμφανίζεται έντονα.

Τέλος μια ακόμα πηγή συστηματικών αποκλίσεων των δημογραφικών στοιχείων από την πραγματικότητα αποτελούν τα σφάλματα περιεχομένου, ή απαντητικά σφάλματα της τρίτης κατηγορίας. Αυτά συνδέονται με την άρνηση των ατόμων να απαντήσουν σε συγκεκριμένα ερωτήματα ή ακόμα με τις, ακούσια ή εκούσια, λανθασμένες απαντήσεις τους σε διάφορα ερωτήματα που εμπεριέχονται στα δελτία, καταγραφών ή στα ερωτηματολόγια.

9.2 Ακρίβεια των κατά ηλικία δεδομένων

Το τυπικότερο ίσως συστηματικό απαντητικό σφάλμα το οποίο εμφανίζεται στα στοιχεία του συνόλου των πηγών δημογραφικών δεδομένων, συνδέεται με τις λανθασμένες δηλώσεις ηλικιών. Το πρόβλημα αυτό οδηγεί στην εμφάνιση του φαινομένου της συσσώρευσης ηλικιών στα δεδομένα, γνωστό στη διεθνή δημογραφική βιβλιογραφία σαν φαινόμενο “hearing”. Το τυπικότερο σφάλμα που παρατηρείται στις δηλώσεις ηλικιών τόσο των απογραφών όσο και των καταγραφών φυσικής και μεταναστευτικής κίνησης του πληθυσμού, ιδιαίτερα όσον αφορά άτομα ηλικιωμένα, είναι κύρια η προτίμηση ηλικιών που είναι πολλαπλάσια του πέντε και σε μικρότερη ένταση η προτίμηση σε ηλικίες που λήγουν σε δύο ή οκτώ. Τα προβλήματα αυτά εμφανίζονται με μεγάλη ένταση σε πληθυσμούς με χαμηλό εκπαιδευτικό επίπεδο, όπου εμφανίζεται ακόμα αναλφαβητισμός και αφορούν ειδικότερα ηλικιωμένα άτομα, τα οποία αγνοούν την ακριβή ημερομηνία γέννησης του και έτσι, κατά κανόνα, στρογγυλοποιούν την ηλικία τους κατά τη δήλωση, στην πλησιέστερη ηλικία που λήγει σε μηδέν ή πέντε. Εφ’ όσον δεν υπάρχει τρόπος εξακρίβωσης των ακριβών ηλικιών, δεν υπάρχει και κάποια μοναδική μέθοδος για την διόρθωση αυτού του προβλήματος.

9.3 Τεχνικές διερεύνησης της ακρίβειας των κατά ηλικία δεδομένων

Πολλοί πρακτικοί τρόποι της ακρίβειας των κατά ηλικία δημογραφικών δεδομένων έχουν παρουσιαστεί στη βιβλιογραφία. Ένας πολύ αποτελεσματικός, επαρκής και εξαιρετικά απλός τρόπος διερεύνησης του προβλήματος της συσσώρευσης ηλικιών σε κατά ηλικία δεδομένα είναι η απλή παρουσίαση του σε διάγραμμα. Ένα χαρακτηριστικό παράδειγμα εμπειρικών δεδομένων τα οποία εμφανίζουν πρόβλημα συσσώρευσης ηλικιών είναι εκείνα του Ελληνικού πληθυσμού. Στον Πίνακα 9.1 εμφανίζονται οι εμπειρικοί κατά ηλικία αριθμοί θανάτων του Ελληνικού πληθυσμού των δύο φύλων, των ετών 1960 και 1961 για τις ηλικίες από 45 και άνω. Γραφική απεικόνιση αυτών των δεδομένων δίνεται στα διαγράμματα 9.1 - 9.4. Στα διαγράμματα αυτά είναι εμφανές το πρόβλημα της συσσώρευσης ηλικιών, το οποίο εμφανίζεται εντονότατα στις ηλικίες που είναι πολλαπλάσια του πέντε και λιγότερο έντονα στις ηλικίες που λήγουν σε δύο ή οκτώ.

Ακόμα, είναι εμφανές ότι το πρόβλημα αυτό εμφανίζεται εντονότερο στις μεγάλες ηλικίες και ακόμα ότι είναι πολύ οξύτερο στα δεδομένα που αφορούν τον γυναικείο πληθυσμό.

Ένας τρόπος μέτρησης της έντασης με την οποία εμφανίζεται το πρόβλημα της συσσώρευσης ηλικιών σε κατά ηλικία δεδομένα είναι η άθροιση τους κατά ηλικίες που λήγουν στο ίδιο ψηφίο και ο υπολογισμός του κλάσματος:

$$\frac{\sum_{i=n_1}^{n_2} N_{5i}}{\frac{k}{n} \sum_{j=x_1}^{x_2} N_j}$$

όπου

N_i είναι ο αριθμός δηλώσεων σε ηλικία i ,

n ο συνολικός αριθμός ηλικιών,

k ο αριθμός ηλικιών που είναι πολλαπλάσια του πέντε.

Τα n_1 και n_2 επιλέγονται έτσι ώστε τα πενταπλάσια τους να είναι ίσα με τη μικρότερη αντιστοιχία μεγαλύτερη ηλικία πολλαπλάσιο του πέντε που εμφανίζεται στα δεδομένα.

Τέλος, x_1 και x_2 είναι η μικρότερη και μεγαλύτερη αντίστοιχα ηλικία που εμφανίζεται στα δεδομένα. Το κλάσμα αυτό πολλαπλασιαζόμενο με το 100 είναι γνωστό στη δημογραφική βιβλιογραφία σαν δείκτης του Wipple. Η τιμή του κυμαίνεται μεταξύ του 100 και 500. Η τιμή 100 αντιστοιχεί σε δεδομένα που είναι πλήρως απαλλαγμένα από συσσώρευση ενώ η τιμή 500 αντιστοιχεί σε δεδομένα που εμφανίζουν απόλυτη συσσώρευση σε ηλικίες που λήγουν σε μηδέν ή πέντε, όταν δηλαδή οι συχνότητες εμφάνισης των υπολοίπων ηλικιών είναι μηδενικές. Για παράδειγμα στα δεδομένα του Πίνακα 9.1, που αφορούν το διάστημα ηλικιών 45-84 ο δείκτης Wipple υπολογίζεται από:

$$\frac{\sum_{i=9}^{16} N_{5i}}{\frac{8}{40} \sum_{j=45}^{84} N_j} * 100$$

όπου N_i είναι ο αριθμός θανάτων στην ηλικία i .

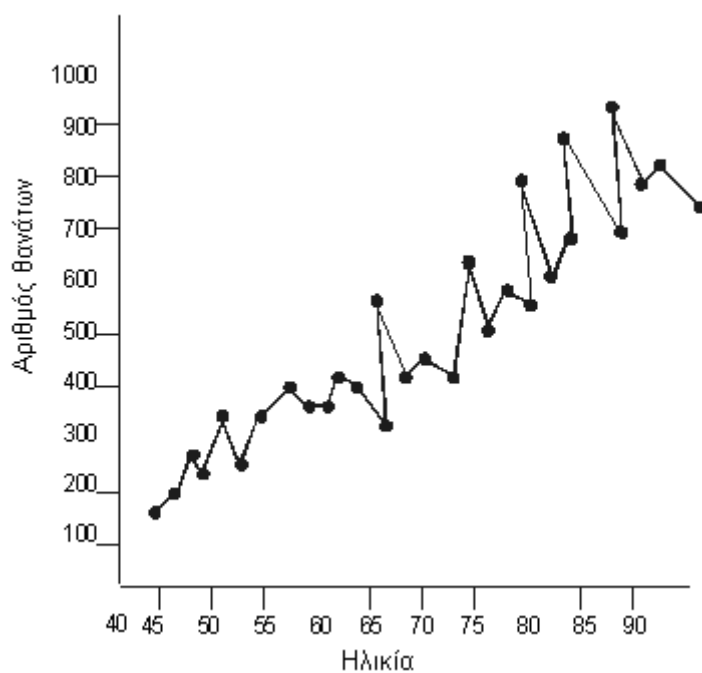
Στον Πίνακα 9.2 δίνονται οι τιμές του δείκτη για τα δύο επιλεγμένα έτη και τα δύο φύλα.

Πίνακας 9.2 Δείκτης Wipple

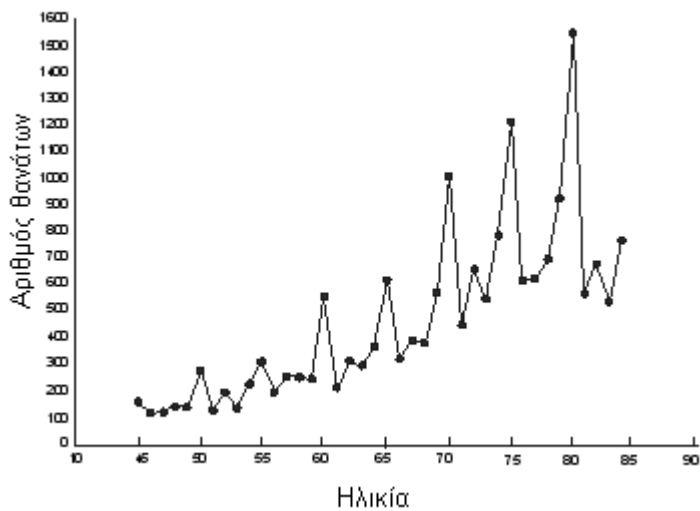
	1960	1961
ΑΝΔΡΕΣ	117,8	109,1
ΓΥΝΑΙΚΕΣ	153,2	123,7
ΣΥΝΟΛΟ	134,7	116,1

Παρατηρούμε ότι οι τιμές του δείκτη είναι πολύ υψηλότερες στον γυναικείο πληθυσμό. Ακόμα το 1960 η συσσώρευση ήταν πολύ εντονότερη από εκείνη του έτους 1961.

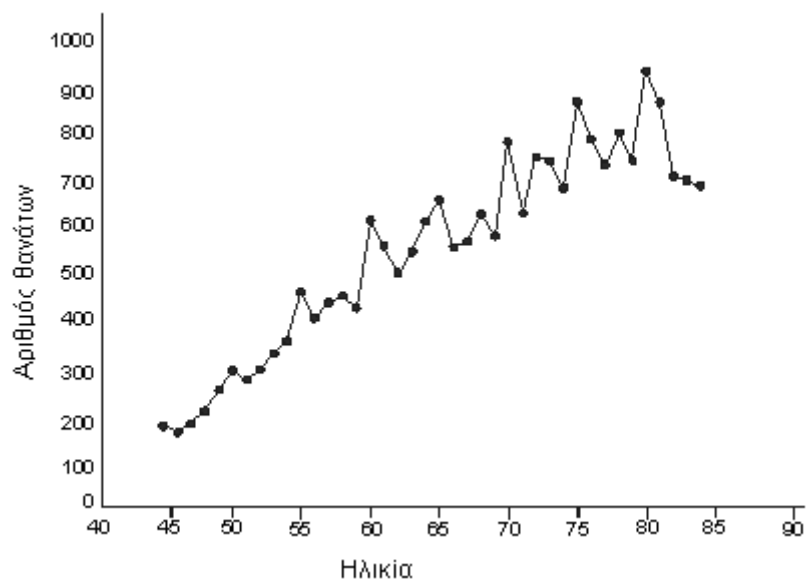
Διάγραμμα 9.1 Θάνατοι κατά ηλικία, του Ελληνικού ανδρικού πληθυσμού του έτους 1960



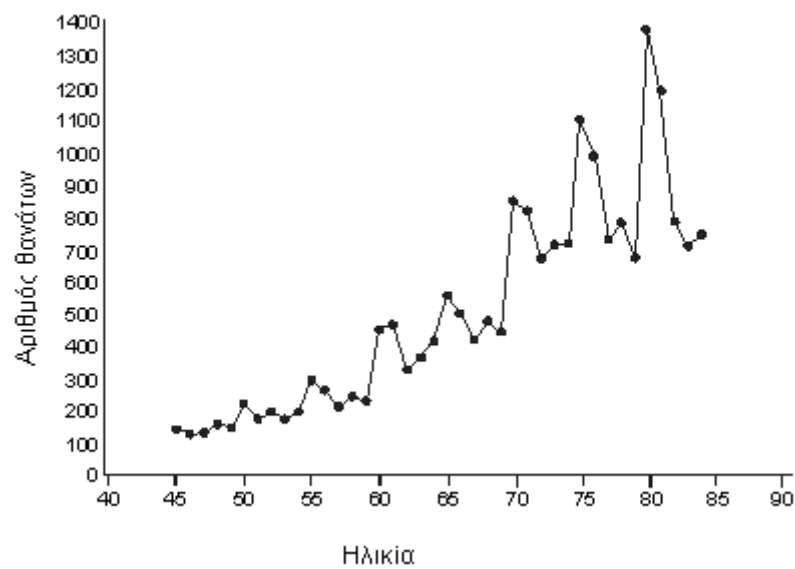
Διάγραμμα 9.2 Θάνατοι κατά ηλικία, του Ελληνικού γυναικείου πληθυσμού του έτους 1960



Διάγραμμα 9.3 Θάνατοι κατά ηλικία, του Ελληνικού ανδρικού πληθυσμού του έτους 1961



Διάγραμμα 9.4 Θάνατοι κατά ηλικία, του Ελληνικού γυναικείου πληθυσμού του έτους 1961



Πίνακας 9.1: Θάνατοι κατά ηλικία του Ελληνικού πληθυσμού ηλικιών 45-84, των ετών 1960 και 1961

Ηλικία	Ανδρες	Γυναίκες	Σύνολο	Ηλικία	Ανδρες	Γυναίκες	Σύνολο
45	193	157	350	45	171	151	322
46	171	115	286	46	158	136	294
47	198	121	319	47	178	139	317
48	240	142	382	48	205	163	368
49	197	137	334	49	250	155	405
50	313	276	589	50	291	226	517
51	212	127	339	51	272	181	453
52	293	195	488	52	294	201	495
53	340	136	476	53	327	180	507
54	317	226	543	54	357	198	555
55	422	311	733	55	462	296	758
56	384	196	580	56	404	269	673
57	382	256	638	57	437	217	654
58	436	251	687	58	453	248	701
59	422	244	666	59	426	233	659
60	584	555	1139	60	616	453	1069
61	368	213	581	61	562	466	1028
62	456	316	772	62	502	329	831
63	488	296	784	63	548	369	917
64	495	369	864	64	613	418	1031
65	659	618	1277	65	661	552	1213
66	490	323	813	66	560	496	1056
67	518	391	909	67	572	416	988
68	551	382	933	68	631	475	1106
69	534	577	1111	69	583	441	1024
70	772	1013	1785	70	787	848	1635
71	548	448	996	71	632	816	1448
72	691	661	1352	72	754	669	1423
73	616	550	1166	73	747	711	1458
74	708	788	1496	74	688	715	1403
75	881	1218	2099	75	873	1098	1971
76	681	620	1301	76	793	985	1778
77	679	625	1304	77	737	724	1461
78	819	701	1520	78	807	774	1581
79	691	931	1622	79	748	665	1413
80	953	1553	2506	80	943	1374	2317
81	617	571	1188	81	872	1187	2059
82	730	683	1413	82	713	782	1495
83	635	540	1175	83	703	702	1405
84	595	775	1370	84	692	738	1430

9.4 Τεχνικές διόρθωσης της ανακρίβειας των κατά ηλικία δεδομένων

Διάφορες μέθοδοι έχουν προταθεί στην βιβλιογραφία για την εκτίμηση των σωστών με αγνώστων κατά ηλικία μεγεθών και την εξάλειψη της εσφαλμένης συγκέντρωσης των δηλωθέντων ηλικιών κύρια σε πολλαπλάσια του πέντε.

Μερικές προτάσεις βασίζονται στην ιδέα της εφαρμογής κάποιας καμπύλης στα εμπειρικά δεδομένα ή τη χρήση κάποιας τεχνικής εξομάλυνσης.

Αλλά η επιλογή μίας καμπύλης για την περιγραφή της κατά ηλικίας σύνθεσης του πληθυσμού εμπεριέχει αυθαιρεσία. Άλλες μέθοδοι που εναλλακτικά μπορούν να εφαρμοστούν σε εμπειρικά δεδομένα επιφορτισμένα με το πρόβλημα της συσσώρευσης ηλικιών είναι διάφορες τεχνικές εξομάλυνσης. Αλλά τόσο η εφαρμογή μίας καμπύλης όσο και η εφαρμογή οποιασδήποτε μεθόδου εξομάλυνσης δεν θεωρούνται ιδιαίτερα κατάλληλες διαδικασίες εφόσον τα δεδομένα εμπεριέχουν συστηματικά σφάλματα που πρέπει να διορθωθούν αλλά και τυχαίες, άρρυθμης μορφής διακυμάνσεις οι οποίες αντανακλούν την πραγματικότητα και έτσι δεν είναι σκόπιμο να εξομαλυνθούν.

Έτσι η πρακτική λύση που υιοθετείται από τις στατιστικές υπηρεσίες των διαφόρων χωρών είναι η παρουσίαση των εμπειρικών στοιχείων κατά ομάδες ηλικιών, επιλεγμένες έτσι ώστε τα σύνολα των ομάδων να αντιστοιχούν στα αντίστοιχα σύνολα των πραγματικών μα άγνωστων στοιχείων, ενώ για τις ηλικίες από 85 και άνω, όπου το πρόβλημα της συσσώρευσης ηλικιών είναι το σοβαρότερο εφ' όσον η άγνοια της ακριβούς ηλικίας είναι περισσότερο εκτεταμένη, τα στοιχεία δίνονται αθροιστικά, σαν ένας μόνον αριθμός.

Μια αποτελεσματική μέθοδος εξάλειψης του φαινομένου της λανθασμένης συσσώρευσης ηλικιών σε δεδομένα θνησιμότητας έχει προταθεί πρόσφατα από τους Κωστάκη και Lanke (1994).¹⁸

¹⁸ Kostaki, A., J: "Degrouping mortality data" Technical report, Department of Statistics, Athens University of Economics, 1994.