

Δειγματοληπτικές έρευνες ηλεκτρονικού ταχυδρομείου (email surveys) και μεροληψία μη απόκρισης

Νικόλαος Φαρμάκης και Ελένη Γοργογιάννη

Τμήμα Μαθηματικών, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Περίληψη

Οι δειγματοληπτικές έρευνες ηλεκτρονικού ταχυδρομείου (ΔΕΗΤ, email surveys) εφαρμόζονται ολοένα και περισσότερο τα τελευταία χρόνια, χάρη στην ραγδαία εξέλιξη του διαδικτύου και την ολοένα αυξανόμενη πρόσβαση των χρηστών κάθε ηλικίας σε αυτό. Έχουν ήδη αναπτυχθεί και διάφορες μορφές λογισμικού κατάλληλες για: 1. την κατάρτιση των ερωτηματολογίων, 2. την προώθηση τους στο διαδίκτυο και 3. την αυτόματη εξαγωγή κάποιων πρώτων αποτελεσμάτων κύρια περιγραφικής στατιστικής (συχνότητες, μέσες τιμές). Στις έρευνες αυτές εμφανίζονται και αρκετά προβλήματα αφού κάθε τεχνική δειγματοληψίας έχει τόσο πλεονεκτήματα όσο και μειονεκτήματα. Βασικό ζήτημα θεωρείται αυτό της μη απόκρισης μερικών μελών του εκάστοτε δείγματος της έρευνας, ειδικά όταν το κλάσμα των μη αποκρινόμενων (non-response rate) είναι σχετικά υψηλό. Αυτό το γεγονός εισάγει μεροληψία (non-response bias) στα αποτελέσματα της έρευνας με συνέπεια την εξαγωγή εσφαλμένων αποτελεσμάτων και συμπερασμάτων (non-response error). Στην παρούσα εργασία γίνεται προσπάθεια εκτίμησης αυτής της μεροληψίας και αντιμετώπισής της με την αύξηση της ακτίνας του διαστήματος εμπιστοσύνης. Παρουσιάζεται μία έρευνα ηλεκτρονικού ταχυδρομείου με πληθυσμό στόχο τα μέλη της πανεπιστημιακής κοινότητας του Α.Π.Θ. και τη συμπεριφορά τους σε σχέση με τους χώρους κοινωνικής δικτύωσης.

1. Εισαγωγή

Οι δειγματοληπτικές έρευνες αποτελούν σήμερα μια από τις πιο δημοφιλείς μεθόδους συλλογής και επεξεργασίας διαφόρων ειδών δεδομένων με εφαρμογή σε αρκετούς επιστημονικούς κλάδους. Υπάρχουν πολλές μέθοδοι διεξαγωγής σωστής και αξιόπιστης δειγματοληπτικής έρευνας, Φαρμάκης (2009^α, 2009^β). Αυτή που θα επιλεγεί εξαρτάται από τους στόχους που έχουν τεθεί για τη έρευνα αλλά και από τους περιορισμούς (χρονικούς ή χρηματικούς) του περιβάλλοντός της. Οι συχνά εφαρμοζόμενες δειγματοληπτικές τεχνικές σήμερα είναι οι ακόλουθες:

- με ταχυδρομική αποστολή του ερωτηματολογίου στον πιθανό ερωτώμενο

- με προσωπική συνέντευξη σε δημόσιο χώρο όπως π.χ. ένας κεντρικός δρόμος
- με προσωπική συνέντευξη με κατ' οίκον επισκέψεις του ερευνητή
- με τηλεφωνική συνέντευξη σε σταθερά ή κινητά τηλέφωνα
- με ηλεκτρονικό ταχυδρομείο (ΔΕΗΤ, email surveys)
- μέσω κάποιας ιστοσελίδας στην οποία είναι ήδη αναρτημένο το ερωτηματολόγιο και τέλος με συνδυασμούς των μεθόδων που αναφέρθηκαν προηγούμενα.

Οι δειγματοληπτικές έρευνες χρησιμοποιούνται στην έρευνα διαφόρων καταστάσεων και φυσικά αποτελούν ένα από τα καλύτερα εργαλεία επικοινωνίας απαντώντας σε προβλήματα που αναφύονται καθημερινά. Ένα από τα προβλήματα που ενδιαφέρει αυτή τη χρονική περίοδο είναι αυτό των χώρων κοινωνικής δικτύωσης και ειδικά η συχνότητα με την οποία ασχολούνται οι άνθρωποι μ' αυτά καθώς (το κυριότερο) και επίδραση αυτών στην κοινωνία. Ένα κομμάτι της Ελληνικής κοινωνίας που πρωταγωνιστεί σε τέτοια φαινόμενα είναι τα μέλη των Πανεπιστημιακών Κοινοτήτων και αυτό επιχειρείται να εξεταστεί εδώ.

Στην παρούσα εργασία θα αναφερθούμε μόνο στις δειγματοληπτικές έρευνες ηλεκτρονικού ταχυδρομείου, (email surveys) τονίζοντας τα πλεονεκτήματα όσο και τα μειονεκτήματα τους. Οι έρευνες αυτές είναι αποτελεσματικές στις περιπτώσεις όπου ο πληθυσμός στόχος είναι αυστηρά καθορισμένος (το προσωπικό μιας εταιρείας ή το σύνολο των μελών της πανεπιστημιακής κοινότητας) και αρκετά εξοικειωμένος με την χρήση του ηλεκτρονικού υπολογιστή. Η συλλογή των δεδομένων είναι αρκετά γρήγορη, υλοποιείται σε πραγματικό χρόνο (real time), με πολύ χαμηλό κόστος και οι έρευνες αυτές είναι εύκολα διαχειρίσιμες. Εντοπίζονται εύκολα τα άτομα που δεν απάντησαν αρχικά προκειμένου να τους αποσταλεί ξανά το ερωτηματολόγιο. Παρατηρείται ότι οι έρευνες αυτές έχουν μεγαλύτερη απήχηση σε άτομα που ανήκουν στις μικρότερες ηλικιακές ομάδες. Συχνά τίθεται το θέμα της ειλικρίνειας των απαντήσεων του ερωτώμενου αλλά και της ασφαλείας του προγράμματος που «φιλοξενεί» το ερωτηματολόγιο. Σημαντικότερο ίσως μειονέκτημα αυτών των ερευνών είναι ότι πολλές φορές οι ερωτώμενοι αλλάζουν συχνά την ηλεκτρονική τους διεύθυνση ή διατηρούν παραπάνω από έναν ενεργό λογαριασμό ηλεκτρονικού ταχυδρομείου. Αυτό το γεγονός συντελεί στην εμφάνιση μεροληψίας λόγω μη απόκρισης ή ακόμα και στην υπερεκπροσώπηση μελών του δείγματος σε περίπτωση που ο ερωτώμενος λάβει σε όλους τους λογαριασμούς ηλεκτρονικού ταχυδρομείου το υπό συμπλήρωση ερωτηματολόγιο.

Υπήρξαν πολλές προσπάθειες να μετρηθεί η επίδραση της μεροληψίας αυτού του είδους στις τιμές των εκτιμητών. Μία μέθοδος αποτίμησης της μεροληψίας (ίσως η πιο δημοφιλής) είναι αυτή που συνδέθηκε με την τιμή του ποσοστού απόκρισης με βάση το μέγεθος του πληθυσμού στόχου. Η μεροληψία αυτή μετριέται με βάση την τυπική απόκλιση S_x της εκτιμώμενης τυχαίας μεταβλητής X . Ο σχετικός τύπος εκτίμησης και απόδοσης της μεροληψίας είναι, Bethlehem, J., (2010),

$$B = s_x \cdot \sqrt{p^{-1} - 1} \quad (1.1)$$

και πρακτικά δίνει μια αυξημένη ακτίνα στο σχετικό διάστημα εμπιστοσύνης. Εφαρμόζεται με τη χρήση των εκτιμητών s_x και \hat{p} των παραμέτρων S_x και p αντίστοιχα. Σαφήνεια δίνει το ακόλουθο παράδειγμα.

Παράδειγμα: Σε μία ΔΕΗΤ έχουμε πληθυσμό μεγέθους $N = 20000$ άτομα και απάντησαν στην έρευνα $n = 400$ άτομα. Η πρόσκληση είχε αποσταλεί σε 1000 άτομα. Από το δείγμα προέκυψαν για την τ.μ. $X =$ οικογενειακό εισόδημα $\bar{x} = 45000$ € και $s_x = 6000$. Μεταξύ ποιών ορίων κυμαίνεται το μέσο εισόδημα;

Απάντηση: Το διάστημα εμπιστοσύνης 100 (1- α)% είναι:

$$\left\{ \bar{x} \pm z_{\alpha/2} \cdot s_B / \sqrt{n} \right\}$$

όπου πρέπει να προσδιορίσουμε την τυπική απόκλιση s_B έτσι ώστε να ενσωματώνεται και η μεροληψία. Έτσι από $s_x = 6000$ και με βάση την (1.1) μεταβαίνουμε σε

$$s_B = B = s_x \cdot \sqrt{p^{-1} - 1} = 6000 \cdot \sqrt{0.4^{-1} - 1} = 7348.$$

Άρα το διάστημα εμπιστοσύνης (δ.ε.) 95% είναι πλέον:

$$\left\{ 45000 \pm (1.96 \cdot 7348) / \sqrt{400} \right\} = \{ 45000 \pm 720 \} = \{ 44280, 45720 \} \text{ €}.$$

2. Ερευνητικό μέρος και αποτελέσματα

Από τις 31/3/-6/4/2011, ύστερα από έγκριση της Πρυτανείας του Α.Π.Θ. πραγματοποιήθηκε ΔΕΗΤ με θέμα: «Η Διάδοση και η χρήση των χώρων κοινωνικής δικτύωσης (social media) στα μέλη της Πανεπιστημιακής κοινότητας». Το σχετικό μήνυμα ηλεκτρονικού ταχυδρομείου (email) έλαβαν όλα τα μέλη Δ.Ε.Π. καθώς και οι ενεργοί φοιτητές του Α.Π.Θ. με email διεύθυνση που τους δίνεται από το Α.Π.Θ. ενώ την προώθηση του ανέλαβε και διεκπεραίωσε το NOC (Network Operation Center) του Α.Π.Θ. Οι φοιτητές δεν χρησιμοποιούν συχνά την διεύθυνση email αυτή.

Από τα 1210 ερωτηματολόγια που συμπληρώθηκαν συνολικά μόνο τα 1177 ήταν πλήρως συμπληρωμένα και αυτά αποτέλεσαν το υπό εξέταση δείγμα μας. Η απόκριση των μελών της Πανεπιστημιακής κοινότητας ήταν ικανοποιητική, ξεπερνώντας το 10%. Με βάση την προηγούμενη παράγραφο θεωρείται ότι τα άτομα που πήραν το ερωτηματολόγιο ενώ είναι ενεργά μέλη είναι περίπου 12000. Άρα σε κάθε εκτίμηση μέσης τιμής θα μπορούμε να έχουμε και διάστημα εμπιστοσύνης με ακτίνα επαυξημένη κατά τη διόρθωση μεροληψίας μη απόκρισης της (1.1), ίση με $\sqrt{p^{-1} - 1}$. Συγκεκριμένα εδώ είναι $p=0.1$ και άρα η διόρθωση είναι $\sqrt{p^{-1} - 1} = 3$, τριπλασιασμός της ακτίνας.

Η συμμετοχή των ερωτώμενων στην έρευνα φθίνει με την πάροδο των ημερών (Πίνακας 1). Αυτό πιθανόν να οφείλεται στο γεγονός ότι το ερωτηματολόγιο προωθήθηκε από το NOC

ημέρα Πέμπτη και ακολούθησε το Σαββατοκύριακο κατά το οποίο όλο και λιγότερα μέλη της Πανεπιστημιακής κοινότητάς είδαν την προσωπική ηλεκτρονική αλληλογραφία που ενδεχομένως να είχαν λάβει μέχρι τότε.

Επιχειρήθηκε προσέγγιση της κατανομής που ακολουθεί ο ρυθμός απόκρισης στην έρευνα. Ελέγχθηκε αν η μορφή της κατανομής είναι η εκθετική, οπότε μέση τιμή και τυπική απόκλιση είναι ίσες. Θεωρητικά είναι $EX=\sigma_x=(VarX)^{0.5}$, Φαρμάκης, Ν., (2001). Από τον πίνακα 1 προκύπτουν κοντινές τιμές για τη μέση τιμή και την τυπική απόκλιση, και μια προσέγγιση της σ.π.π. με την εκθετική μορφή θα ήταν ενδιαφέρονσα. Όμως, το εγχείρημα αυτό δεν πέτυχε λόγω της μικρής έστω διαφοράς μεταξύ των \bar{x} και s . Κρίθηκε προτιμητέα προσέγγιση αυτής της συνάρτησης με ένα πολυώνυμο με τη χρήση μιας νεότερης και ευκολότερης μεθόδου, με τη συμβολή του *συντελεστή μεταβλητότητας (ΣΜ, Αγγλιστί Cv)*, πολύτιμου εργαλείου στη μελέτη των κατανομών μέσα από δειγματοληψία

Πίνακας 1. Ρυθμός Απόκρισης ανά 24ώρο

	Frequency	Percent	Cumulative Percent
Πρώτη μέρα	571	48,5	48,5
Δεύτερη μέρα	320	27,2	75,7
Τρίτη μέρα	92	7,8	83,5
Τέταρτη μέρα	60	5,1	88,6
Πέμπτη μέρα	76	6,5	95,1
Έκτη μέρα	50	4,2	99,3
Έβδομη μέρα	8	,7	100,0
Σύνολα	1177	100,0	

Πίνακας 2. Υπολογισμοί παραμέτρων για εύρεση σ.π.π.

$$W=1, m=3.5, \bar{x} = 1.5926, s=1.4803, Cv=0.9295, k=-0.1257, h=-0.9935$$

a_i	b_i	x_i	n_i	x'_i	$n_i \cdot x'_i$	$n_i \cdot x_i^2$
0	1	0,5	571	-3	-1713	5139
1	2	1,5	320	-2	-640	1280
2	3	2,5	92	-1	-92	92
3	4	3,5	60	0	0	0
4	5	4,5	76	1	76	76
5	6	5,5	50	2	100	200
6	7	6,5	8	3	24	72
$n=1177$					$T_1=-2245$	$T_2=6859$

Σημείωση: Στους Πίνακες 2 και 3, αφού θεωρήσουμε το πεδίο τιμών της σ.π.π. $f(x)$ το $(0, 7]$, το διαιρούμε σε 7 ισομήκη διαστήματα για να βρούμε τη δειγματική μέση τιμή και την αντίστοιχη διασπορά, με βάση τη μέθοδο που εμφανίζεται στο σύγγραμμα Φαρμάκης (2001). Οι παράμετροι είναι: W =εύρος τάξης, m =προσωρινή μέση τιμή, a_i, b_i, n_i = τα άκρα της i -τάξης και η συχνότητα παρατηρήσεων σ' αυτήν. Το x' προκύπτει από αλλαγή αρχής μέτρησης (νέα αρχή= m) και αλλαγή μονάδων μέτρησης (κανονικοποίηση) με βάση τη σχέση $x'_i = (x_i - m)/W$, $i=1,2,\dots,7$ και όπου x_i είναι το κέντρο της i -τάξης, ήτοι $x_i = (a_i + b_i)/2$. Τα T_1 και T_2 είναι τα αθροίσματα των στηλών τους. Τα k, h τέλος είναι ο εκθέτης και ο συντελεστής στη σχέση

$$f(x) = \begin{cases} h \cdot \{1 - (x/b)^k\}, & \alpha \nu x \in (0, b] \\ 0 & , \alpha \nu x \notin (0, b] \end{cases}$$

που πρωτοεμφανίζεται στην εργασία Farmakis και Makris (2011) και αφορά τις φθίνουσες σ.π.π. Η τυχαία μεταβλητή που χρησιμοποιούμε εδώ είναι ο χρόνος που είναι κάτι το συνεχές. Στην περίπτωση μας έγινε διαίρεση της όλης χρονικής διάρκειας των 7 ημερών σε 7 τάξεις και επιτυγχάνεται η ομαδοποίηση των δειγματικών δεδομένων (Φαρμάκης, 2001). Η φύση του τύπου της σ.π.π. μας υποχρεώνει να χρησιμοποιούμε στους υπολογισμούς μας το ημιανοιχτό αριστερά διάστημα $(0, 7]$ αντί του $[0, 7]$. Άλλοι τύποι σ.π.π. αντιμετωπίζονται με κατάλληλη πολυωνυμική μορφή (Farmakis, 2003).

Εφαρμόστηκε η μέθοδος που προτάθηκε στην εργασία των Farmakis, N., Makris, G., (2011) και από τα δεδομένα του Πίνακα 2 προκύπτει η παρακάτω σ.π.π.:

$$f(x) = \begin{cases} -0.9936 \cdot \{1 - (x/7)^{-0.1257}\}, & \alpha \nu x \in (0, 7] \\ 0 & , \alpha \nu x \notin (0, 7] \end{cases} \quad (2.1)$$

Αυτό σημαίνει ότι προσεγγίσαμε την σ.π.π. με μία πολυωνυμική συνάρτηση. Η (2.1) δίνει πολύ κοντινές τιμές για τα n_i , $i=1,2,\dots,7$, παρόλο που η σχετικά μεγάλη τιμή του μεγέθους n του δείγματος αποδυναμώνει κάπως τη δοκιμασία X^2 για να κριθεί η προσαρμογή των δεδομένων στην (2.1).

Η αντίστοιχη συνάρτηση κατανομής (σ.κ.) της μεταβλητής X = Χρόνος απόκρισης στην ΔΕΗΤ, με βάση τη σ.π.π. στην (2.1) είναι:

$$F(x) = \begin{cases} 0 & , \alpha \nu x < 0 \\ -0.9936 \cdot \{x - 8.0065 \cdot (x/7)^{0.8743}\} & , \alpha \nu x \in (0, 7] \\ 1 & , \alpha \nu x > 7 \end{cases} \quad (2.2)$$

Επειδή, ο παραπάνω υπολογισμός οριακά μόνο επιβεβαιώνεται με τη διαδικασία του X^2 υπολογίσαμε και πάλι την κατανομή που ακολουθεί ο ρυθμός απόκρισης στην παρούσα έρευνα με τον ορισμό που χρησιμοποιεί η American Association for Public Opinion Research

(2006) στο “Standard definitions: Final dispositions of case codes and outcome rates for surveys” και αναφέρουν οι The-Hua Shin και Xitao Fan (2008).

Οι παραπάνω ορίζουν ως **ελάχιστο ρυθμό απόκρισης** το παρακάτω κλάσμα:

$$RR1 = \frac{I}{(I + P) + (R + NC + O) + (UH + UO)}$$

όπου $RR1$ = ο ελάχιστος ρυθμός απόκρισης, I = το πλήθος των ολοκληρωμένων ερωτηματολογίων, P = το πλήθος των μερικά συμπληρωμένων ερωτηματολογίων, R = το πλήθος ατόμων που αρνήθηκαν να συμπληρώσουν το ερωτηματολόγιο, NC = το πλήθος των ερωτώμενων που δεν μπόρεσαν να εντοπιστούν, O = άλλο, UH = το πλήθος των δειγματικών μονάδων για τις οποίες δεν έχει η ερευνητική ομάδα κάποια πληροφορία για τον εντοπισμό τους (για την περίπτωση μας λογαριασμοί email που δεν είναι ενεργοί) και τέλος, UO = το πλήθος των δειγματικών μονάδων που δεν είναι γνωστές στην ερευνητική ομάδα για κάποιο άλλο λόγο.

Με βάση τον Πίνακα 3 υπολογίστηκε η σ.π.π. για την κατανομή του ελάχιστου ρυθμού απόκρισης ανά 24ώρο. Ο νέος πίνακας που προκύπτει είναι ο ακόλουθος.

Πίνακας 3. Υπολογισμοί παραμέτρων για εύρεση σ.π.π.

$$W=1, m=3.5, \bar{x}=1.739, s=1.6700, C_v=0.9576, k=-0.1877, h=-0.6182$$

a_i	b_i	x_i	n_i	x'	$n_i x_i'$	$n_i \cdot x_i'^2$
0	1	0,5	19	-3	-57	171
1	2	1,5	11	-2	-22	44
2	3	2,5	3	-1	-3	3
3	4	3,5	2	0	0	0
4	5	4,5	3	1	3	3
5	6	5,5	2	2	4	8
6	7	6,5	1	3	3	9
$n=41$					$T_1 = -72$	$T_2 = 238$

Με εκ νέου εφαρμογή της μεθόδου Farmakis, N. και Makris, G., (2011) καταλήγουμε στην προσέγγιση της σ.π.π., που ακολουθεί ο ελάχιστος ανά 24ώρο ρυθμός απόκρισης. Αυτή δίνεται από την επόμενη σχέση:

$$f(x) = \begin{cases} -0.6178 \cdot \{1 - (x/7)^{-0.1878}\}, & \alpha \nu x \in (0, 7] \\ 0 & , \alpha \nu x \notin (0, 7] \end{cases} \quad (2.1)'$$

Με την ολοκλήρωση της (2.1)' προκύπτει η αθροιστική σ. κ. της που δίνεται από την επόμενη εξίσωση

$$F(x) = \begin{cases} 0 & , \alpha \nu x < 0 \\ -0.6178 \cdot x + 5.3246 \cdot (x/7)^{0.8122} & , \alpha \nu x \in (0, 7] \\ 1 & , \alpha \nu x > 7 \end{cases} \quad (2.2)'$$

Η $F(x)$ στη (2.2)' χρησιμοποιήθηκε για τον προσδιορισμό των θεωρητικών τιμών των συχνοτήτων για τη δοκιμασία του X^2 . Η τιμή που προέκυψε είναι $X^2 = 1.8322$. Ενώ η κρίσιμη τιμή για 3 βαθμούς ελευθερίας (β.ε.) και στάθμη σημαντικότητας (σ.σ.) $\alpha=0.05$ είναι $X^2 = 7.81473 > 1.8322$. Άρα, η παραπάνω κατανομή στατιστικά γίνεται αποδεκτή, αφού τα δεδομένα προσαρμόζονται καλύτερα σε αυτήν, έναντι της πρώτης που απορρίπτεται. Οι β.ε. από 6 έγιναν 3 για να προκύψουν θεωρητικές συχνότητες μεγαλύτερες του 5 (Φαρμάκης, 2001).

Πιο κάτω παρατίθενται τα αποτελέσματα που προέκυψαν ύστερα από την στατιστική ανάλυση των απαντήσεων που συγκεντρώθηκαν με την βοήθεια του Lime Survey για το οποίο θα γίνει λόγος στη συνέχεια. Η επεξεργασία των αποτελεσμάτων έγινε αρχικά με το Microsoft Excel 2007 και στη συνέχεια με το στατιστικό πακέτο PASW Statistics έκδοση 18. Παρατηρούμε λοιπόν ότι:

(α) Η συμμετοχή των γυναικών στην έρευνα ήταν ιδιαίτερα αυξημένη. Από τα 1177 άτομα του δείγματος τα 658 (ποσοστό 55.9%) είναι γυναίκες ενώ τα υπόλοιπα 519 (ποσοστό 44.1%) είναι άνδρες.

(β) Στην ερώτηση σχετικά με την ηλικία των ερωτώμενων παρατηρούμε ότι 714 άτομα του δείγματος, ποσοστό περί το 60.7% έχουν ηλικία ανάμεσα στα 18 έως 24 έτη. Ακόμη, 213 άτομα εμφανίζονται στην δεύτερη ηλικιακή ομάδα αυτή των 25 έως 30 ετών (ποσοστό 18.1%). Οι υπόλοιπες ηλικιακές ομάδες παρουσιάζουν ολοένα και μικρότερα ποσοστά εκπροσώπησης γεγονός που ίσως να οφείλεται στην μικρότερη εξοικείωση τους με την χρήση του Η/Υ και τις νέες τεχνολογίες.

(γ) Σε επόμενη ερώτηση ο ερωτώμενος καλείται να δηλώσει ποιους χώρους κοινωνικής δικτύωσης (social media) γνωρίζει. Το Facebook είναι ο πιο αναγνωρίσιμος χώρος κοινωνικής δικτύωσης. Το 97.4% του δείγματος ισχυρίζεται ότι γνωρίζει το Facebook. Δεύτερο σε αναγνωρισιμότητα εμφανίζεται το Skype με ποσοστό της τάξης του 93.3% και τρίτο με ποσοστό 89.2% εμφανίζεται το MSN. Τελευταίο εμφανίζεται το Twitter με ποσοστό 80.5%, γεγονός που ίσως να οφείλεται στην πρόσφατη εμφάνιση του έναντι των υπολοίπων χώρων κοινωνικής δικτύωσης.

(δ) Στην πορεία ο συμμετέχων δηλώνει την ύπαρξη ή μη κάποιου ενεργού προσωπικού προφίλ στους προαναφερθέντες χώρους κοινωνικής δικτύωσης. Η συντριπτική πλειονότητα του δείγματος δηλώνει «Ναι» στην κατοχή ενεργού προφίλ σε ποσοστό που αγγίζει το 88%. Μόνο το 11.6% δηλώνει «Όχι» στην κατοχή ενεργού προφίλ και το υπόλοιπο 0.4% αρνείται να απαντήσει στο ερώτημα.

(ε) Ένα ακόμη σημαντικό ερώτημα που τίθεται συχνά και στο οποίο έχει ιδιαίτερη σημασία η άποψη των μελών της Πανεπιστημιακής κοινότητας είναι το πώς αυτά αξιολογούν τους χώρους κοινωνικής δικτύωσης. Έτσι σε σχετική ερώτηση το 16.2% των ερωτηθέντων απαντά «Θετικά» και το 58.5% «Μάλλον θετικά». «Αρνητικά» απαντά το 4.7% των ερωτηθέντων και «Μάλλον αρνητικά» το 14.9% αυτών. Το 4.4% του δείγματος δηλώνει ότι «Δεν έχει γνώμη» και το υπόλοιπο 1.4% «Δε γνωρίζω/Δεν απαντώ».

3. Το λογισμικό Lime Survey

Η έρευνα έγινε με το Lime Survey, ένα λογισμικό που βοηθάει το χρήστη να δημιουργήσει ταυτόχρονα ένα μεγάλο πλήθος διαφορετικών ερωτηματολογίων και να διεξάγονται ταυτόχρονα οι αντίστοιχες δειγματοληπτικές έρευνες.

Ένα ακόμα δυνατό πλεονέκτημα του λογισμικού αυτού είναι ότι επιτρέπει στον ερευνητή να διαχειρίζεται τουλάχιστον είκοσι (20) διαφορετικά είδη ερωτήσεων αλλά και να εισάγει έτοιμες ερωτήσεις από κάποιο άλλο πρόγραμμα. Το πρόγραμμα παρέχει και μεγάλο εύρος έτοιμων ερωτήσεων για τον σχεδιασμό ολοκληρωμένης ΔΕΗΤ. Είναι γνωστό ότι από τον τρόπο διατύπωσης μιας ερώτησης επηρεάζονται πολλές φορές οι απαντήσεις που θα λάβει ο ερευνητής από τον ερωτώμενο και κατ' επέκταση και τα συμπεράσματα που θα βγουν έπειτα από την στατιστική ανάλυση και επεξεργασία των δεδομένων του. Δεν υπάρχει περιορισμός στο πλήθος των ερωτήσεων που θα απαρτίζουν ένα ερωτηματολόγιο. Το λογισμικό Lime Survey επιτρέπει την δημιουργία ερωτηματολογίου σε αρκετές διαφορετικές γλώσσες, αλλά και την άνετη παρακολούθηση της πορείας μιας έρευνας με την προβολή στην οθόνη του Η/Υ του επικεφαλής της έρευνας των τελευταίων απαντήσεων αλλά και στατιστικών αποτελεσμάτων σε πινάκες (Γοργογιάννη, 2011).

4. Συζήτηση - Συμπεράσματα

Έρευνες όπως η παρούσα, που ερευνούν διάφορα κοινωνικά, οικονομικά ή και άλλα φαινόμενα μέσα από τη χρήση του διαδικτύου και παράλληλα εκτιμούν παραμέτρους της διαδικασίας, είναι πολύ επίκαιρες και πληθαίνουν. Το διαδίκτυο αναδεικνύεται ολοένα και περισσότερο καλός και ευέλικτος συνδετικός μηχανισμός της κοινωνίας και των εκφάνσεών της. Βοηθάει στη συλλογή δεδομένων με σχετική ευκολία. Σημασία έχει και η αξιοποίηση των δεδομένων που συγκεντρώθηκαν. Μπορούμε να κάνουμε στατιστικές αναλύσεις και να έχουμε συμπεράσματα. Εδώ έγινε η χρήση και του συντελεστή C_v ώστε να έχουμε με μια απλή και γρήγορη διαδικασία εκτίμηση για τη σ.π.π. μιας τμ X . Η διαδικασία αυτή μάλιστα είναι και αυτοελεγχόμενη, πράγμα που της προσδίδει μεγάλη αξία και χρησιμότητα.

Ο C_v βοηθά να έχουμε αποτελέσματα εκτιμήσεων υπό μορφή συναρτήσεων πιθανότητας ή και κατανομής που συνδυάζουν συντομία και ένταση πληροφορίας. Αποτυπώθηκε η σχέση της Πανεπιστημιακής Κοινότητας με τους χώρους κοινωνικής δικτύωσης και το πως τους αξιολογεί και τους αξιοποιεί.

Ο ρυθμός ανταπόκρισης σε ΔΕΗΤ είναι σημαντικότερος. Βοηθάει να εκτιμηθεί η μεροληψία της έρευνας και άρα η αξιοπιστία της και να γίνουν διορθωτικές παρεμβάσεις στα αποτελέσματα, ώστε να μειώνονται οι επιπτώσεις της μεροληψίας. Οι διορθωτικές επεμβάσεις φαίνονται (αποκτούν νόημα) όταν από την εκτίμηση σε σημείο πάμε στην εκτίμηση σε διάστημα με τα διαστήματα εμπιστοσύνης (δ.ε.). Εκεί κάθε δ.ε. με ακτίνα r αποκτά, λόγω μεροληψίας μη απόκρισης, ακτίνα $r' = r \cdot \sqrt{p^{-1} - 1}$, όπως προαναφέρουμε και δείξαμε και με παράδειγμα στην εισαγωγή. Με ποσοστό απόκρισης της τάξης 10% ($p=0.10$) η ακτίνα στο δ.ε. διορθώνεται με τριπλασιασμό. Αν θέλουμε δηλαδή να δώσουμε δ.ε. στο ποσοστό 80,5% αυτών που χρησιμοποιούν το twitter υπολογίζουμε αρχικά την ακτίνα του δ.ε. χωρίς να παίρνουμε υπόψη μας τη μεροληψία μη απάντησης. Είναι $r=0.0226$. Μετά την πολλαπλασιάζουμε με το $3=\sqrt{p^{-1} - 1}$ και γίνεται $r'=0.0678$. Έτσι το δ.ε. 95% από $\{0.7824 - 0.8276\}$ γίνεται $\{0.7372 - 0.8728\}$. Αυτή η διαχείριση μπορεί να γίνει για όλες τις μέσες τιμές του υπό μελέτη προβλήματος, αν αποφασίσει ο ερευνητής να τις δει ως μέσα δ.ε. $100(1-\alpha)\%$.

Τέλος, οι ΔΕΗΤ παρουσιάζουν χαμηλό κόστος και αυξημένη ταχύτητα στη αποστολή του ερωτηματολογίου, με προϋπόθεση ότι δεν υπάρχουν τεχνολογικά προβλήματα.

Email Surveys and non-response bias

Nikolaos Farmakis, Eleni Gorgogianni

*Department of Mathematics
Aristotle University of Thessaloniki, Greece*

Abstract

The surveys supported by the email services (email surveys) become more and more popular and their use is extended the last few years. The reason of this phenomenon is the extension of use of web by the people and the easy access on it. As a support to the above we have a development of software on: 1. Questionnaire preparation, 2. Promotion of questionnaires to the internet and 3. The mining of the first simple findings of the survey basically of a descriptive statistics nature (mean value, variance, standard deviation, Coefficient of Variation at least). During an email survey some problems are faced. An important case is the rate of non-response persons, especially when this rate is very high. In this case a kind of bias is introduced and the results include important errors (non-response bias causes, non-response error). In the present paper we try to estimate the above mentioned bias and some movements are proposed in order to face the related non-response errors, basically by adopting a bigger radius of the related confidence interval. Also an email survey has taken place at Aristoteleion

University of Thessaloniki (A.U.Th.) with target population its staff members and its students. The main goal was to estimate their behaviour with the social media. A kind of software was used named *Lime Survey*. This kind of software helps the user to prepare more than one questionnaire and permits the running of many email surveys conducted by one user in the same time. Some interesting results are presented. In all these results a correction of the non-response bias can be applied, providing more reliable confidence intervals for the suitable theoretic values of the estimated parameters.

Βιβλιογραφία

- Γοργογιάννη, Ε. (2011). *Ειδικές Μορφές Δειγματοληψίας*, Διπλωματική Διατριβή, Α.Π.Θ.
- Φαρμάκης, Ν. (2001). *Στατιστική, Περιληπτική θεωρία-Ασκήσεις*, Α & Π Χριστοδουλίδη, Θεσσαλονίκη.
- Φαρμάκης, Ν. (2009^α). *Εισαγωγή στη Δειγματοληψία*, Α & Π Χριστοδουλίδη, Θεσσαλονίκη.
- Φαρμάκης, Ν. (2009^β). *Δημοσκοπήσεις και Δεοντολογία*, Α & Π Χριστοδουλίδη, Θεσσαλονίκη.
- Bethlehem, J. (2010). Selection Bias in Web Surveys, *International Statistical Review*, 78(2), 161-188.
- Farmakis, N. (2003). Estimation of Coefficient of Variation: Scaling of Symmetric Continuous Distributions, *Statistics in Transition*, 6(1), 83-96.
- Farmakis, N. and Makris G. (2011). Web-sampling: Probabilities of Specific Information Achievement, *Proc. of Applied Stochastic Models 7 Data Analysis (ASMDA)*, Rome, Italy.
- Shih, T. H., and Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field methods*, 20(3), 249-271.

Ευχαριστίες: Ευχαριστούμε τους κριτές για τις εύστοχες παρατηρήσεις τους, που μας βοήθησαν να βελτιώσουμε σημαντικά μορφή αλλά και περιεχόμενο στην εργασία αυτή, που κινείται σε χώρους δειγματοληπτικής έρευνας σχετικά σπάνιους μέχρι τώρα και αναδυόμενους.