ΚΟΝΟΜΙΚΑΙ
ΟΙΝΩΝΙΚΑΙ
ΤΕΧΝΙΚΑΙ

ΣΠΟΥΔΑΙ

ΤΡΙΜΗΝΙΑΙΑ
ΕΠΙΣΤΗΜΟΝΙΚΗ
ΕΠΙΘΕΩΡΗΣΙΣ

ΚΔΟΣΙΣ ΤΗΣ ΕΝ ΠΕΙΡΑΙΕΙ ΑΝΩΤΑΤΗΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΟΛΗΣ

ΟΙΚΟΝΟΜΙΚΟΝ ΕΤΟΣ 1972	ΙΟΥΛΙΟΣ - ΣΕΠΤΕΜΒΡΙΟΣ 1972	ΚΒ' ΤΟΜΟΣ	ΑΡΙΘΜ. ΤΕΥΧΟΥΣ 3

ΕΚΤΙΜΗΤΑΙ ΑΝΑΛΟΓΙΚΟΥ - ΤΥΠΟΥ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΠΟΛΥΩΝΥΜΙΚΗΣ ΤΙΝΟΣ ΚΑΤΑΝΟΜΗΣ

Τοῦ κ. ΔΗΜΗΤΡΙΟΥ Α. ΑΘΑΝΑΣΟΠΟΥΛΟΥ

Διδάκτορος τῆς Στατιστικῆς τοῦ Πανεπιστημίου Καλιφορνίας (Berkeley)

1. Εἰσαγωγὴ

Μία μέθοδος αὐξήσεως τῆς ἀκριβείας τῶν ἐκτιμήσεων αἱ ὁποῖαι προκύπτουν ἐκ μιᾶς δειγματοληπτικῆς ἐρεύνης εἶναι ἡ χρησιμοποίησις βοηθητικῶν πληροφοριῶν. Ἡ «ἀναλογική» ἐκτίμησις ὡς καὶ ἡ ἐκτίμησις «παλινδρομήσεως» τοῦ μέσου ἢ τοῦ συνόλου ἑνὸς ποσοτικοῦ χαρακτηριστικοῦ τοῦ πληθυσμοῦ ἀποτελοῦν παραδείγματα εἰς τὰ ὁποῖα γίνεται χρῆσις τῆς γνώσεως τοῦ μέσου ἢ τοῦ συνόλου μιᾶς ἄλλης μεταβλητῆς προκειμένου νὰ αὐξηθῇ ἡ ἀκρίβεια τῆς ἐκτιμήσεως τοῦ μέσου ἢ τοῦ συνόλου τοῦ ὑπὸ μελέτην χαρακτηριστικοῦ. Ἄλλαι μέθοδοι διὰ τὴν αὔξησιν τῆς ἀκριβείας τῶν ἐκτιμήσεων συνδέονται μὲ τὸν τρόπον ἐπιλογῆς τοῦ δείγματος ὡς π.χ. ἡ ἐπιλογὴ τῶν μονάδων τοῦ δείγματος μὲ πιθανότητα ἀνάλογον πρός τι μέγεθος αὐτῶν ἐκ τῶν προτέρων γνωστὸν κ.ο.κ.

Εἰς πλείστας ὅσας δειγματοληπτικὰς ἐρεύνας ὁ ἐρευνητὴς πέραν τῶν ποσοτικῶν χαρακτηριστικῶν - διὰ τὰ ὁποῖα ἔχει νόημα ὁ ὑπολογισμὸς ἑνὸς μέσου ἢ ἑνὸς συνόλου - μελετᾷ καὶ ποιοτικὰ τοιαῦτα (τὰς καλουμένας κατηγορικὰς μεταβλητὰς) ὡς π.χ. τὸ ἐπάγγελμα, τὸ φῦλον, ἡ οἰκογενειακὴ κατάστασις, ὁ τύπος τοῦ αἵματος κλπ. Εἰς τὰς περιπτώσεις αὐτὰς ἐνδιαφερόμεθα νὰ ἐκτιμήσωμεν τὸ διάνυσμα τῶν πιθανοτήτων ($q_0, q_1, q_2, \ldots q_s$) αἱ ὁποῖαι εἶναι αἱ παράμετροι μιᾶς πολυωνυμικῆς κατανομῆς ἀντιστοιχούσης εἰς τὴν κατανομὴν τοῦ πληθυσμοῦ ὡς πρὸς τὸ ὑπὸ μελέτην ποιοτικὸν χαρακτηριστικὸν (κατηγορικὴν μεταβλητήν).

Εἰς τὴν παροῦσαν ἐργασίαν ἐκτίθεται μία μέθοδος — ἐν ἀντιστοιχίᾳ πρὸς τὴν ἀναλογικὴν ἐκτίμησιν τοῦ μέσου ὅρου ἑνὸς πληθυσμοῦ — ἡ ὁποία ἐπιτρέπει τὴν βελτίωσιν τῆς ἀκριβείας τῆς κλασσικῆς μεθόδου ἐκτιμήσεως τοῦ διανύσματος ($q_0, q_1, q_2, \ldots . q_s$), ἤτοι, τῶν παραμέτρων τῆς πολυωνυμικῆς κατανομῆς ἑνὸς πληθυσμοῦ ὡς πρός τινα κατηγορικὴν μεταβλητὴν Μ. Εἰς τὸν ὑπολογισμὸν τῆς προτεινομένης «ἀναλογικοῦ — τύπου» ἐκτιμήσεως γίνεται χρῆσις πληροφοριῶν ἀναφερομένων εἰς μίαν κατηγορικὴν μεταβλητὴν κατὰ τὸ μᾶλλον ἢ ἧττον ἐντόνως συσχετισμένην πρὸς τὴν κυρίαν μεταβλητὴν Μ. Ἡ διανυσματικὴ παράμετρος ($p_0, p_1, p_2, \ldots p_r$) ἡ ὁποία καθορίζει τὴν πολυωνυμικὴν κατανομὴν τοῦ πληθυσμοῦ ὡς πρὸς τὴν βοηθητικὴν κατηγορικὴν μεταβλητὴν Α ὑποτίθεται γνωστή. Εἰς τὴν πρᾶξιν αἱ ἀνωτέρω ὑποθέσεις πληροῦνται εἰς τὰς ἑξῆς δύο κυρίως περιπτώσεις :

(α) Ὅταν ὡς βοηθητικὴ μεταβλητὴ Α χρησιμοποιῆται αὐτὴ αὕτη ἡ κατηγορικὴ μεταβλητὴ Μ εἰς προηγουμένην χρονικὴν στιγμήν, ὅτε π.χ. ἔλαβε χώραν ἀπογραφὴ τοῦ πληθυσμοῦ.

καὶ (β) Ὅταν εἴς τινα ἀπογραφὴν πληθυσμοῦ μερικὰ τῶν χαρακτηριστικῶν ἐκτιμῶνται δειγματοληπτικῶς.

Ἡ μορφὴ καὶ ὁ τρόπος ὑπολογισμοῦ τοῦ «ἀναλογικοῦ—τύπου» ἐκτιμητοῦ $\overline{q}' = (\overline{q}_1, \overline{q}_2, \ldots\ldots \overline{q}_s)$ τῆς διανυσματικῆς παραμέτρου $q' = (q_1, q_2, \ldots q_s)$ πραγματεύονται εἰς τὸ κεφάλαιον 2. Εἰς τὸ κεφάλαιον 3 ἀποδεικνύεται ὅτι ὁ προτεινόμενος ἐκτιμητὴς $\overline{q}$ εἶναι ἀμερόληπτος καὶ ὑπολογίζεται ἡ μήτρα τῶν διακυμάνσεων – συνδιακυμάνσεων αὐτοῦ μὲ προσέγγισιν* τῆς τάξεως $0(n^{-2})$ ὑπὸ τὴν προϋπόθεσιν μεγάλου δείγματος. Ἡ δειγματοληπτικὴ ἐκτίμησις τῆς ἀνωτέρω μήτρας διακυμάνσεων – συνδιακυμάνσεων δίδεται εἰς τὸ κεφάλαιον 4. Ἡ ἐν λόγῳ ἐκτίμησις εἶναι ἐλαφρῶς μεροληπτικὴ (element-wise)*. Ἡ μεροληψία εἶναι τῆς τάξεως $\frac{1}{n^2}$ καὶ διὰ μεγάλα δείγματα εἶναι προφανῶς ἀμελητέα. Εἰς τὸ κεφάλαιον 5 συγκρίνεται ἡ ἀκρίβεια τοῦ προτεινομένου ἐκτιμητοῦ μὲ ἐκείνην τοῦ κλασσικοῦ ἁπλοῦ ἀμερολήπτου ἐκτιμητοῦ καὶ ἀποδεικνύεται ὅτι εἶναι πάντοτε μεγαλυτέρα. Φυσικά, ἡ αὔξησις τῆς ἀκριβείας εἶναι μεγάλη ἐὰν ἡ ὑπὸ μελέτην κατηγορικὴ μεταβλητὴ Μ καὶ ἡ βοηθητικὴ τοιαύτη Α παρουσιάζουν ἔντονον συσχέτισιν. Εἰς τὸ κεφάλαιον 6 ἀποδεικνύεται ὅτι ὁ ἐκτιμητὴς $\overline{q}$ εἶναι συνεπὴς (consistant), ἡ δὲ ἀσυμπτωτικὴ κατανομὴ αὐτοῦ (ἐὰν $n \to \infty$) εἶναι ἡ κανονικὴ τοιαύτη.

Τέλος, εἰς τὸ κεφάλαιον 7 δίδεται ἓν ἀριθμητικὸν παράδειγμα μὲ πραγματικὰ δεδομένα ἐκ τοῦ ὁποίου καταφαίνεται ἐναργέστερον τόσον ὁ τρόπος ἐφαρμογῆς τῶν βασικῶν συμπερασμάτων καὶ τύπων ὅσον καὶ ἡ πραγματικὴ σημασία τῆς προτεινομένης μεθόδου.

Δέον νὰ τονισθῇ ὅτι τὰ ἐξαγόμενα ἐνταῦθα συμπεράσματα διὰ πολυωνυμικὰς κατανομὰς ὡς πρός τινα κατηγορικὴν μεταβλητὴν δύνανται νὰ ἐφαρμοσθοῦν καὶ εἰς τὴν περίπτωσιν κατανομῶν ὡς πρός τινα συνεχῆ μεταβλητὴν (π.χ. τὸ εἰσόδημα κλπ.) ἐὰν ἔχει γίνει — ὡς συνήθως — ταξινόμησις τῶν μονάδων τοῦ πληθυσμοῦ ἢ τοῦ δείγματος εἰς ἕνα πεπερασμένον ἀριθμὸν τάξεων. Τοῦτο, δεδομένου ὅτι ἡ ὁμαδοποίησις συνεχῶν κατανομῶν ἀποτελεῖ τὸν κανόνα εἰς τὰς πάσης φύσεως στατιστικὰς ἐρεύνας, διευρύνει οὐσιωδῶς τὰς δυνατότητας ἐφαρμογῆς τῆς προτεινομένης διὰ τοῦ παρόντος ἄρθρου μεθόδου.

---

*) Ἐνταῦθα ἡ προσέγγισις νοεῖται στοιχεῖον πρὸς στοιχεῖον (element - wise).

2. Ὁ προτεινόμενος ἀναλογικοῦ - τύπου ἐκτιμητὴς

Ἔστωσαν A_i, $i = 0, 1, 2, \ldots, r$ καὶ M_j, $j = 0, 1, 2, \ldots s$ αἱ κατηγορίαι (τάξεις) εἰς τὰς ὁποίας ταξινομοῦνται αἱ μονάδες τοῦ πληθυσμοῦ ὡς πρὸς τὴν βοηθητικὴν (A) καὶ κυρίαν (M) κατηγορικὴν μεταβλητήν. Ἔστωσαν ἀκόμη $(p_0, p_1, p_2, \ldots, p_r)$ καὶ $(q_0, q_1, q_2, \ldots q_s)$ αἱ διανυσματικαὶ παράμετροι αἱ ὁποῖαι καθορίζουν τὰς ἀντιστοίχους πολυωνυμικὰς κατανομὰς τοῦ πληθυσμοῦ. Δεδομένου ὅτι $\sum_{j=0}^{s} q_j = 1$ θὰ ἀσχοληθῶμεν κατωτέρω μὲ τὴν ἐκτίμησιν τῶν s ἐκ τῶν ὡς ἄνω παραμέτρων, ἤτοι μὲ τὴν ἐκτίμησιν τῆς διανυσματικῆς παραμέτρου $q' = (q_1, q_2, \ldots, q_s)$.

Ἡ διανυσματικὴ παράμετρος $P' = (p_0, p_1, p_2, \ldots, p_r)$ ὑποτίθεται, ὡς ἤδη ἐλέχθη, γνωστὴ καὶ καλούμεθα νὰ ἐκτιμήσωμεν τὴν διανυσματικὴν παράμετρον $q' = (q_1, q_2, \ldots, q_s)$. Πρὸς τούτοις ἐπιλέγεται ἐκ τοῦ πληθυσμοῦ ἁπλοῦν τυχαῖον δεῖγμα μεγέθους n καὶ ἐξ ἑκάστης μονάδος αὐτοῦ συλλέγεται ἡ πληροφορία περὶ τῶν κατηγοριῶν A_i καὶ M_j εἰς τὰς ὁποίας ἡ ἐν λόγῳ μονὰς ἀνήκει.

Ἐὰν π_{ij} εἶναι ἡ δεσμευμένη πιθανότης μία μονὰς τοῦ πληθυσμοῦ u νὰ ἀνήκῃ ὡς πρὸς τὴν μεταβλητὴν M εἰς τὴν κατηγορίαν M_j δοθέντος ὅτι ὡς πρὸς τὴν βοηθητικὴν μεταβλητὴν A ἀνήκει εἰς τὴν κατηγορίαν A_i δηλαδὴ ἐὰν $\pi_{ij} = \Pr(u \varepsilon M_j \,/\, u \varepsilon A_i)$ δυνάμεθα νὰ γράψωμεν $q_j = \sum_{i=0}^{r} p_i \pi_{ij}$ δι' ὅλα τὰ $j = 0, 1, 2, \ldots, s$. (1)

Χρησιμοποιοῦντες τὰς ἀνωτέρω σχέσεις λαμβάνομεν:

$$q = \Pi' \cdot p \tag{2}$$

ὅπου q εἶναι ἡ διανυσματικὴ παράμετρος q' γραμμένη ὡς διάνυσμα - στήλη (ἐναλλακτὴ τοῦ διανύσματος - γραμμῆς $q' = (q_1, q_2, \ldots, q_s)$) καὶ Π' εἶναι ἡ ἐναλλακτὴ τῆς $(r+1) \times s$ μήτρας $\Pi = (\pi_{ij})$ τῶν δεσμευμένων πιθανοτήτων π_{ij}, $i = 0, 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$.

Ὁ προτεινόμενος «ἀναλογικοῦ - τύπου» ἐκτιμητὴς τῆς διανυσματικῆς παραμέτρου q, δίδεται ὑπὸ τῆς σχέσεως

$$\bar{q} = \widehat{\Pi}' \cdot p \tag{3}$$

ὅπου $\widehat{\Pi} = (\widehat{\pi}_{ij})$ εἶναι ἡ στοιχεῖον πρὸς στοιχεῖον (element - wise) ἐκτίμησις μεγίστης πιθανοφανείας (maximum likelihood estimator) τῆς μήτρας Π ἥτις ἐπιτυγχάνεται ἐκ τοῦ δείγματος ὡς ἀκολούθως:

Αἱ μονάδες τοῦ δείγματος ταξινομοῦνται ὡς πρὸς τὰς κατηγορικὰς μεταβλητὰς A καὶ M εἰς ἕνα πίνακα μὲ $(r+1)$ γραμμὰς καὶ $(s+1)$ στήλας.

Ἐὰν n_{ij} εἶναι ὁ ἀριθμὸς τῶν μονάδων τοῦ δείγματος αἱ ὁποῖαι ἀνήκουν εἰς τὸ φατνίον (cell) (A_i, M_j) $i = 0, 1, 2, \ldots r$, $j = 0, 1, 2, \ldots, s$

καὶ $n_{i\cdot} = \sum_{j=0}^{s} n_{ij}$, αἱ ἐκτιμήσεις μεγίστης πιθανοφανείας (M.L.E.) τῶν δεσμευμένων πιθανοτήτων ($\pi_{i0}, \pi_{i1}, \pi_{i2}, \ldots\ldots, \pi_{is}$), $i = 0, 1, 2, \ldots\ldots, r$ ἐπιτυγχάνονται εὐκόλως ἐκ τῆς συγκατανομῆς (joint - distribution) τῶν τυχαίων μεταβλητῶν ($n_{i0}, n_{i1}, n_{i2}, \ldots\ldots, n_{is}$), (δεσμευμένης ἐπὶ τοῦ $n_{i.}$), ἡ ὁποία εἶναι πολυωνυμικὴ μὲ παραμέτρους π_{i0}, π_{i1} $\pi_{i2}, \ldots\ldots \pi_{is}$ ἱκανοποιούσας τὴν συνθήκην $\sum_{j=0}^{s} \pi_{ij} = 1$ καὶ δίδονται ὑπὸ τῶν σχέσεων

$$\widehat{\pi}_{ij} = \frac{n_{ij}}{\sum_{j=0}^{s} n_{ij}} = \frac{n_{ij}}{n_{i\cdot}} \tag{4}$$

δι' ὅλα τὰ i, j.

3. Ἡ μήτρα τῶν διακυμάνσεων - συνδιακυμάνσεων

Λαμβάνοντες ὑπ' ὄψιν ὅτι ἡ ἀναμενομένη τιμὴ

$$E(\widehat{\pi}_{ij}) = E\left(E\left[\frac{n_{ij}}{n_{i\cdot}} / n_{i\cdot}\right]\right) = E(\pi_{ij}) = \pi_{ij}$$

προκύπτει ὅτι ὁ ἀναλογικὸς ἐκτιμητὴς $\bar{q}$ ὡς οὗτος ὁρίζεται ἐκ τῆς σχέσεως (3) εἶναι ἀμερόληπτος. Ἡ μήτρα διακυμάνσεων – συνδιακυμάνσεων τοῦ ἐκτιμητοῦ $\bar{q}$ προκύπτει ἐκ τοῦ ἑξῆς θεωρήματος.

Θ ε ώ ρ η μ α 1. Ἐὰν τὸ μέγεθος τοῦ δείγματος n εἶναι ἀρκετὰ μεγάλον οὕτως ὥστε

$$\left| \frac{n_{i.} - np_i}{np_i} \right| < 1, \quad i = 0, 1, 2, \ldots\ldots, r \tag{5}$$

τότε

$$\operatorname{Var}(\bar{q}_j) = \frac{1}{n} \sum_i p_i \pi_{ij} (1 - \pi_{ij}) + 0(n^{-2}), \; j = 1, 2, \ldots s \tag{6}$$

καὶ

$$\operatorname{Cov}(\bar{q}_j, \bar{q}_k) = -\frac{1}{n} \sum_i p_i \pi_{ij} \pi_{ik} - 0(n^{-2}), \; k \neq j = 1, 2, \ldots s \tag{7}$$

Ἀπόδειξις: Ἐκ τῆς σχέσεως (3) λαμβάνομεν $\bar{q}_j = \sum_i p_i \widehat{\pi}_{ij}$

Συνεπῶς

$$\operatorname{Var}(\bar{q}_j) = E(\operatorname{Var}[\bar{q}_j / n_{0\cdot}, n_{1\cdot}, \ldots, n_{r\cdot}]) + \operatorname{Var}(E[\bar{q}_j / n_{0\cdot}, n_{1\cdot}, \ldots, n_{r\cdot}]) =$$
$$= E\left(\sum_i p_i^2 \frac{\pi_{ij}(1 - \pi_{ij})}{n_{i\cdot}}\right) + \operatorname{Var}(q_j) = \sum_i p_i^2 \pi_{ij}(1 - \pi_{ij}) E\left(\frac{1}{n_{i\cdot}}\right)$$

Ἀναπτύσσοντες τὸ $\frac{1}{n_{i\cdot}}$ εἰς τὴν περιοχὴν τοῦ np_i εἰς σειρὰν Taylor λαμβάνομεν

$$E\left(\frac{1}{n_{i\cdot}}\right)=\frac{1}{np_i}+\frac{1-p_i}{n^2p_i^2}-o\left(n^{-2}\right) \tag{8}$$

ὡς ἐπίσης τὴν σχέσιν

$$\operatorname{Var}\left(\bar{q}_j\right)=\frac{1}{n}\sum_i p_i\,\pi_{ij}\left(1-\pi_{ij}\right)+\frac{1}{n^2}\sum_i\left(1-p_i\right)\pi_{ij}\left(1-\pi_{ij}\right)-o\left(n^{-2}\right)$$

τὸ ὁποῖον ἀποδεικνύει τὴν ἀλήθειαν τῆς σχέσεως (6).

Ἐργαζόμενοι ὁμοίως ἐπιτυγχάνομεν τὴν σχέσιν

$$\operatorname{Cov}\left(\bar{q}_j,\bar{q}_k\right)=-\frac{1}{n}\sum_i p_i\,\pi_{ij}\,\pi_{ik}-\frac{1}{n^2}\sum_i\left(1-p_i\right)\pi_{ij}\,\pi_{ik}+o\left(n^{-2}\right)$$

ἡ ὁποία ἀποδεικνύει τὴν ἀλήθειαν τῆς σχέσεως (7).

Ἐκ τοῦ Θεωρήματος I προκύπτει τὸ ἀκόλουθον :

Π ό ρ ι σ μ α : Ἡ μήτρα διακυμάνσεων–συνδιακυμάνσεων τοῦ ἐκτιμητοῦ $\bar{q}$ κατὰ προσέγγισιν (element - wise) τῆς τάξεως $0\left(n^{-2}\right)$ δίδεται ὑπὸ τῆς σχέσεως

$$V\left(\bar{q}\right)=\frac{1}{n}\cdot\left(v_{jk}\right) \tag{9}$$

ὅπου

$$v_{ij}=\sum_i p_i\,\pi_{ij}\left(1-\pi_{ij}\right),\ j=1,2,\ldots\ldots,s$$

καὶ

$$v_{jk}=-\sum_i p_i\,\pi_{ij}\,\pi_{ik},\ k\neq j=1,2,\ldots,s$$

Τὰ στοιχεῖα τῆς κυρίας διαγωνίου τῆς μήτρας εἰς τὴν σχέσιν (9) εἶναι προφανῶς ἐλαφρῶς μικρότερα τῶν πραγματικῶν διακυμάνσεων τῶν ἐκτιμήσεων τῶν παραμέτρων q_j, $j=1,2,\ldots,s$. Αἱ ἀπόλυτοι τιμαὶ τῶν γενομένων σφαλμάτων εἶναι πάντοτε μικρότεραι τοῦ ἀριθμοῦ $\frac{r}{4n^2}$ καὶ συνεπῶς δι' ἀρκούντως μέγα μέγεθος δείγματος n ταῦτα εἶναι ἀμελητέα. Τὸ σχετικὸν μέγεθος τῶν ἐν λόγῳ σφαλμάτων e_j, ἤτοι ὁ λόγος τοῦ ἀπολύτου σφάλματος πρὸς τὴν ἀντίστοιχον ἀρχικὴν τιμὴν τῆς διακυμάνσεως $\operatorname{Var}\left(\bar{q}_j\right)$, τὸ ὁποῖον παρουσιάζει πολὺ μεγαλύτερον ἐνδιαφέρον ἐν σχέσει πρὸς τὸ ἀπόλυτον σφάλμα, ἀποδεικνύεται εὐκόλως ὅτι εἶναι τῆς τάξεως

$$e_j = 0\,(n^{-1})$$

πάντοτε δὲ μικρότερον τοῦ ἑξῆς ἀνωτέρω φράγματος

$$e_j < \frac{r}{4n\; p_i\; \pi_{ij}\,(1-\pi_{ij})} \tag{10}$$

Ἐκ τῆς σχέσεως (10), ὑποθέτοντες $.05 \leqslant \pi_{ij} \leqslant .95$ δι' ὅλα τὰ ζεύγη (i, j) (σύνηθες εἰς τὴν πρᾶξιν) καὶ λαμβάνοντες ὑπ' ὄψιν ὅτι συνήθως τὸ r ἔχει τὴν τιμὴν 3, 4, ἢ 5 εὑρίσκομεν ὅτι ἐὰν τὸ μέγεθος τοῦ δείγματος $n = 200$ τὰ σχετικὰ σφάλματα e_j εἶναι οὐσιωδῶς μικρότερα τοῦ 10% ἐνῷ ἐὰν τὸ μέγεθος τοῦ δείγματος $n = 1{,}000$ τὰ e_j εἶναι πολὺ μικρότερα τοῦ 2%. Κατὰ παρόμοιον τρόπον εὑρίσκεται ὅτι τὰ σχετικὰ σφάλματα τῶν ἐκτὸς τῆς διαγωνίου στοιχείων (ἤτοι τῶν συνδιακυμάνσεων) εἶναι τῆς αὐτῆς τάξεως μὲ τὰ ἀνωτέρω.

Εἰς τὴν πρᾶξιν τὰ ἀνωτέρω σφάλματα εἶναι πολὺ μικρότερα ἐκείνων ποὺ προκύπτουν ἐκ τῶν ἀνωτέρω θεωρητικῶν ὑπολογισμῶν. Ἂς θεωρήσωμεν πρὸς τούτοις τὸ ἑξῆς παράδειγμα :

Ἔστω $r = s = 2,\quad n = 100,\quad p' = (.6, .3, .1)$

$$\text{καὶ}\quad \Pi = \begin{bmatrix} .2 & .2 \\ .7 & .1 \\ .1 & .8 \end{bmatrix}$$

Ἐφαρμόζοντες τὴν σχέσιν (9) λαμβάνομεν

$$V(\bar{q}) = \frac{1}{100}\begin{bmatrix} .16800 & -.05300 \\ -.05300 & .13900 \end{bmatrix}$$

ἐνῷ ἀντιστοίχως ἡ μήτρα τῶν διακυμάνσεων – συνδιακυμάνσεων, ἡ ὁποία περιλαμβάνει τοὺς ὅρους τάξεως $\frac{1}{n^2}$ (αἱ οὕτω διδόμεναι διακυμάνσεις καὶ συνδιακυμάνσεις εἶναι μεγαλύτεραι τῶν πραγματικῶν τοιούτων) εἶναι

$$W(\bar{q}) = \frac{1}{100}\begin{bmatrix} .16800 + .00292 & -.5300 - .00137 \\ -.05300 - .00137 & .13900 + .00271 \end{bmatrix}$$

καὶ συνεπῶς τὰ σχετικὰ σφάλματα διὰ τὰ διαγώνια στοιχεῖα εἶναι $e_1 = 1.7\%$ $e_2 = 1.9\%$ καὶ διὰ τὰ ἐκτὸς τῆς διαγωνίου στοιχεῖα $e_{12} = 2.5\%$.

Εἰς τὸ κεφάλαιον 7 δίδεται ἓν ἄλλο λεπτομερέστερον παράδειγμα ἀναφορικῶς πρὸς δυωνυμικὰς κατανομάς.

Τέλος, διὰ νὰ ἀποκτήσῃ ὁ ἀναγνώστης μίαν ἰδέαν ἐπὶ τοῦ πόσον μεγάλον πρέπει νὰ εἶναι τὸ μέγεθος τοῦ δείγματος n ὥστε νὰ ἱκανοποιῆται ὁ περιορισμὸς (5) τοῦ θεωρήματος I ἐφαρμόζομεν τὴν ἀνισότητα τοῦ Chebychev εἰς τὴν τυχαίαν μεταβλητὴν $\frac{n_{i\cdot}}{n}$

Ἔχομεν οὕτω

$$\Pr\left(\left|\frac{n_{i\cdot}-np_i}{np_i}\right|<1\right)=\Pr\left(\left|\frac{n_{i\cdot}}{n}-p_i\right|<p_i\right)\geqslant 1-\frac{1-p_i}{np_i}$$

καὶ ὁ κατωτέρω πίναξ 1 ἐμφανίζει διὰ διαφόρους τιμὰς τῆς p_i τὸ ἀπαιτούμενον μέγεθος τοῦ δείγματος ὥστε νὰ ἰσχύῃ ὁ περιορισμὸς (5) μὲ πιθανότητα ἀντιστοίχως .90, .95 καὶ .99. Ὁ πίναξ δὲν περιλαμβάνει τιμὰς τοῦ p_i μεγαλυτέρας τοῦ .50 διότι εἰς αὐτὴν τὴν περίπτωσιν ὁ περιορισμὸς (5) ἰσχύει πάντοτε.

Π ί ν α ξ 1

Ἀπαιτούμενον μέγεθος δείγματος n ἵνα ἰσχύῃ ὁ περιορισμὸς (5) διὰ διαφόρους τιμὰς τῆς παραμέτρου p_i

p_i	ΠΙΘΑΝΟΤΗΣ		
	.90	.95	.99
.05	190	380	1900
.10	90	180	900
.20	40	80	400
.30	23	47	233
.40	15	30	150

4. Δειγματοληπτικὴ ἐκτίμησις τῆς μήτρας διακυμάνσεων—συνδιακυμάνσεων

Ἡ δειγματοληπτικὴ ἐκτίμησις τῆς μήτρας τῶν διακυμάνσεων—συνδιακυμάνσεων $V(\bar{q})$ δίδεται ὑπὸ τοῦ τύπου

$$\hat{V}(\bar{q})=\frac{1}{n}\cdot(\hat{v}_{jk}) \tag{11}$$

ὅπου

$$\hat{v}_{jj} = \sum_i p_i \hat{\pi}_{ij} (1 - \hat{\pi}_{ij}) \qquad j = 1, 2, \ldots, s$$

καὶ

$$\hat{v}_{jk} = -\sum_i p_i \hat{\pi}_{ij} \hat{\pi}_{ik} \qquad k \neq j = 1, 2, \ldots, s$$

Αἱ ἀνωτέρω ἐκτιμήσεις εἶναι ἐλαφρῶς μεροληπτικαί. Ἡ μεροληψία (bias) στοιχεῖον πρὸς στοιχεῖον (element - wise) εἶναι τῆς τάξεως $\frac{1}{n^2}$ καὶ συνεπῶς διὰ μεγάλον μέγεθος δείγματος ἀσήμαντος.

Πράγματι, ἀποδεικνύεται εὐκόλως ὅτι δι' ἀρκούντως μέγα μέγεθος δείγματος n, ἔχομεν

$$E\left(\frac{1}{n}\hat{v}_{jj}\right) = \frac{1}{n} v_{jj} - \frac{1}{n^2}\sum_i \pi_{ij}(1 - \pi_{ij}) - o(n^{-2}) =$$

$$= \frac{1}{n} v_{jj} - 0(n^{-2}) \tag{12}$$

καὶ

$$E\left(\frac{1}{n}\hat{v}_{jk}\right) = \frac{1}{n} v_{jk} + \frac{1}{n^2}\sum_i \pi_{ij}\pi_{ik} + o(n^{-2}) =$$

$$= \frac{1}{n} v_{jk} + 0(n^{-2}) \tag{13}$$

Εἶναι προφανὲς ἐκ τῆς σχέσεως (11) ὅτι τὰ διαγώνια στοιχεῖα τῆς μήτρας διακυμάνσεων—συνδιακυμάνσεων $V(\bar{q})$ (ἤτοι αἱ διακυμάνσεις), ἐλαφρῶς ὑποεκτιμῶνται. Αἱ ἐν λόγῳ μεροληψίαι ἔχουσαι ὡς κύριον ὅρον

$$\frac{1}{n^2}\sum_i \pi_{ij}(1 - \pi_{ij}), \quad j = 1, 2, \ldots, s$$

ὁ ὁποῖος εἶναι πάντοτε μικρότερος τοῦ ἀριθμοῦ $\frac{r+1}{4n^2}$, εἶναι ἀσήμαντοι ἐὰν τὸ μέγεθος δείγματος n εἶναι μεγάλον. Ἐργαζόμενοι ὡς καὶ εἰς τὴν περίπτωσιν τοῦ σχετικοῦ σφάλματος e_j (ἴδε σχέσιν (10)) βλέπομεν εὐκόλως ὅτι τὸ σχετικὸν μέγεθος τῶν μεροληψιῶν b_j (ὁ λόγος τῆς μεροληψίας πρὸς τὴν ὑπὸ ἐκτίμησιν διακύμανσιν) εἶναι ἐπίσης τῆς τάξεως $\frac{1}{n}$ ἐνῷ αἱ ἀπόλυτοι τιμαὶ τῶν μεροληψιῶν b_j δὲν ὑπερβαίνουν τὸ δεύτερον μέλος τῆς ἀνισότητος

$$\left| b_j \right| < \frac{(r+1)}{4n \sum_i p_i \pi_{ij} (1-\pi_{ij})} \tag{14}$$

Χρησιμοποιοῦντες τὰ δεδομένα τοῦ παραδείγματος τοῦ κεφαλαίου 3 (μὲ $n = 100$) καὶ ἐφαρμόζοντες τὰς σχέσεις (12) καὶ (13) (δηλ. περιλαμβάνοντες τὸν ὅρον τῆς τάξεως $\frac{1}{n^2}$) λαμβάνομεν

$$E\left(\hat{V}(\bar{q})\right) = \frac{1}{100}\begin{bmatrix} .1680 - .0046 & -.0530 + .0019 \\ -.0530 + .0019 & .1390 - .0041 \end{bmatrix}$$

ἐκ τῆς ὁποίας προκύπτει ὅτι τὸ σχετικὸν μέγεθος τῶν μεροληψιῶν εἶναι διὰ μὲν τὰ διαγώνια στοιχεῖα

$$|b_1| = 2.7\,\%, \qquad |b_2| = 2.9\,\%$$

διὰ δὲ τὰ ἐκτὸς διαγωνίου στοιχεῖα (ἐκτιμήσεις συνδιακυμάνσεων)

$$|b_{12}| = 3.5\,\%.$$

Ἓν λεπτομερέστερον ἀριθμητικὸν παράδειγμα δίδεται εἰς τὸ κεφάλαιον 7 ἀναφορικῶς πρὸς τὴν δυωνυμικὴν κατανομήν.

5. Αὔξησις τῆς ἀποτελεσματικότητος (ἀκριβείας)

Εἰς τὸ παρὸν κεφάλαιον συγκρίνομεν τὴν ἀκρίβειαν τοῦ προτεινομένου εἰς τὴν παροῦσαν ἐργασίαν ἀναλογικοῦ–τύπου ἐκτιμητοῦ $\bar{q}$ μὲ ἐκείνην τοῦ κλασσικοῦ ἀμερολήπτου ἐκτιμητοῦ

$$\hat{q}' = \frac{1}{n}\,(n_{\cdot 1}, n_{\cdot 2}, \ldots n_{\cdot s})$$

– εἰς τὸν ὑπολογισμὸν τοῦ ὁποίου δὲν γίνεται χρῆσις βοηθητικῶν πληροφοριῶν – προκειμένου νὰ καταστῇ δυνατὴ ἡ ἀξιολόγησις τοῦ ἐπιτυγχανομένου κέρδους εἰς ἀκρίβειαν (ἀποτελεσματικότητα).

Εἶναι γνωστὸν ὅτι ἡ ἀποτελεσματικότης (ἀκρίβεια) ἑνὸς ἀμερολήπτου ἐκτιμητοῦ $\hat{\vartheta}' = (\hat{\vartheta}_1, \hat{\vartheta}_2, \ldots\ldots \hat{\vartheta}_s)$ τῆς διανυσματικῆς παραμέτρου $\vartheta' = (\vartheta_1, \vartheta_2, \ldots\ldots \vartheta_s)$ (δηλαδὴ ἡ ἀπὸ κοινοῦ ἀποτελεσματικότης – joint efficiency – τῶν στοιχείων τοῦ ἐκτιμητοῦ $\hat{\vartheta}_1, \hat{\vartheta}_2, \ldots\ldots \hat{\vartheta}_s$) εἶναι ἀντιστρόφως ἀνάλογος τοῦ τετραγώνου τοῦ ὄγκου τοῦ ἐλλειψοειδοῦς συγκεντρώσεως (concentration ellipsoid) τῆς συγκατανομῆς τῶν $\hat{\vartheta}_1, \hat{\vartheta}_2, \ldots\ldots \hat{\vartheta}_s$, καὶ εἰς τελευταίαν ἀνάλυσιν τῆς γενικευμένης διακυμάνσεως τοῦ $\hat{\vartheta}$, δηλαδὴ τῆς ὁριζούσης τῆς μήτρας τῶν διακυμάνσεων – συνδιακυμάνσεων $V(\hat{\vartheta})$. Κατὰ συνέπειαν ἡ σχετικὴ ἀποτελεσματικότης (ἀκρίβεια) τοῦ ἀναλογικοῦ–τύπου ἐκτιμητοῦ $\bar{q}$ ἐν σχέσει πρὸς τὸν κλασσικὸν ἐκτιμητὴν $\hat{q}$ δίδεται ὑπὸ τῆς σχέσεως

$$e = \frac{|V(\hat{q})|}{|V(\bar{q})|} \tag{15}$$

ὅπου $V(\hat{q})$ εἶναι ἡ μήτρα διακυμάνσεων–συνδιακυμάνσεων τοῦ $\hat{q}$. Κατωτέρω ἀποδεικνύεται ὅτι ὁ ἀναλογικοῦ – τύπου ἐκτιμητὴς $\bar{q}$ εἶναι πάντοτε ἀκριβέστερος τοῦ κλασσικοῦ ἐκτιμητοῦ $\hat{q}$, ἤτοι $e \geqslant 1$.

Θεώρημα 2. Δι' ἀρκούντως μέγα μέγεθος δείγματος n ἡ μήτρα $V(\hat{q}) - V(\bar{q})$, εἶναι ὁριστικῶς μὴ ἀρνητικὴ (non-negative definite).

Ἀπόδειξις: Ἐκ τῆς σχέσεως $q_j = \sum_i p_i \pi_{ij}$, $j = 1, 2, \ldots, s$ (ἴδε (1)), λαμβάνομεν

$$\sum_i p_i \pi_{ij} (1 - \pi_{ij}) = q_j - \sum_i p_i \pi^2_{ij} = q_j (1 - q_j) - \sum_i p_i (\pi_{ij} - q_j)^2$$

καὶ

$$\sum_i p_i \pi_{ij} \pi_{ik} = q_j q_k + \sum_i p_i (\pi_{ij} - q_j)(\pi_{ik} - q_k)$$

Συνεπῶς

$$V(\hat{q}) - V(\bar{q}) = \frac{1}{n}(d_{jk})$$

ὅπου

$$d_{jj} = q_j (1 - q_j) - \sum_i p_i \pi_{ij} (1 - \pi_{ij}) = \sum_i p_i (\pi_{ij} - q_j)^2, \quad j = 1, 2, \ldots, s$$

καὶ

$$d_{jk} = - q_j q_k + \sum_i p_i \pi_{ij} \pi_{ik} = \sum_i p_i (\pi_{ij} - q_j)(\pi_{ik} - q_k), \quad k \neq j = 1, 2, \ldots, s$$

Οὕτω τελικῶς λαμβάνομεν $V(\hat{q}) - V(\bar{q}) = \frac{1}{n} T'T$

ὅπου $T = (t_{ij})$ εἶναι ἡ μήτρα τάξεως $(r + 1) \times s$ μὲ $t_{ij} = \sqrt{p_i}\,(\pi_{ij} - q_j)$, $i = 0, 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$.

Ἐκ τῆς τελευταίας σχέσεως προκύπτει ἡ ἀλήθεια τοῦ θεωρήματος δεδομένου ὅτι ἡ μήτρα $T'T$ εἶναι πάντοτε ὁριστικῶς μὴ ἀρνητική.

Ἐκ τοῦ ἀνωτέρω θεωρήματος προκύπτει ὅτι $\left| V(\hat{q}) \right| \geqslant \left| V(\bar{q}) \right|$ πρᾶγμα τὸ ὁποῖον συνεπάγεται τὴν $e \geqslant 1$. Οἱ ἐκτιμηταὶ $\bar{q}$ καὶ $\hat{q}$ εἶναι ἐξ ἴσου ἀποτελεσματικοὶ εἰς τὴν περίπτωσιν καθ' ἣν τὰ στοιχεῖα ἑκάστης στήλης τῆς μήτρας Π εἶναι ἴσα μεταξύ των. Ἐν ἄλλοις λόγοις ἐὰν ἡ πιθανότης μία μονὰς τοῦ πληθυσμοῦ νὰ ἀνήκη εἰς τὴν κατηγορίαν M_j δὲν ἐπηρεάζεται ἐκ τῆς κατηγορίας εἰς τὴν ὁποίαν ἀνήκει ἡ μονὰς συμφώνως πρὸς τὴν βοηθητικὴν μεταβλητὴν Α δηλαδὴ ἐὰν ἡ ὑπὸ μελέτην μεταβλητὴ Μ καὶ ἡ βοηθητικὴ μεταβλητὴ Α εἶναι στοχαστικῶς ἀνεξάρτητοι.

Πράγματι, ἐὰν $\pi_{0j} = \pi_{1j} = \pi_{2j} = \ldots = \pi_{rj} = \pi_j$ διὰ $j = 1, 2, \ldots s$

τότε $q_j = \sum\limits_i p_i \pi_{ij} = \pi_j \quad j = 1, 2, \ldots s$

καὶ ὅλα τὰ στοιχεῖα τῆς μήτρας $V(\hat{q}) - V(\bar{q})$ εἶναι μηδέν.

Ἀντιθέτως ἡ μείωσις τῶν διακυμάνσεων τῶν στοιχείων τοῦ ἀναλογικοῦ τύπου ἐκτιμητοῦ $\bar{q}$ γίνεται μεγίστη ὅταν τὰ διαγώνια στοιχεῖα

$\sum\limits_i p_i (\pi_{ij} - q_j)^2, \quad j = 1, 2, \ldots s$ μεγιστοποιοῦνται.

Δεδομένου ὅτι $\sum\limits_i p_i (\pi_{ij} - q_j)^2$ εἶναι ἡ διακύμανσις τῶν στοιχείων τῆς $j-th$ στήλης τῆς μήτρας Π, λαμβανομένων μὲ συντελεστὰς σταθμίσεως p_i, $i = 0, 1, 2, . r$, ἡ χρῆσις τοῦ ἀναλογικοῦ – τύπου ἐκτιμητοῦ $\bar{q}$ ἀντὶ τοῦ ἁπλοῦ ἀμερολήπτου ἐκτιμητοῦ $\hat{q}$ συνιστᾶται ἰδιαιτέρως εἰς τὴν περίπτωσιν καθ' ἣν τὰ στοιχεῖα ἑκάστης στήλης τῆς μήτρας Π παρουσιάζουν μεγάλην μεταβλητικότητα.

Χρησιμοποιοῦντες καὶ πάλιν τὰ δεδομένα τοῦ παραδείγματος τοῦ κεφαλαίου 3 ἔχομεν

$$(q_0 = .43 \qquad q_1 = .34 \qquad q_2 = .23)$$

ἐπειδὴ δὲ

$$\left| V(\hat{q}) \right| = \frac{1}{n} \cdot q_0 \cdot q_1 \cdot q_2$$

εὑρίσκομεν

$$\left| V(\hat{q}) \right| = 336 \times 10^{-6}$$

Ἐξ ἄλλου, εὐκόλως ὑπολογίζεται ὅτι

$$\left| V(\bar{q}) \right| = 205 \times 10^{-6}$$

καὶ συνεπῶς συμφώνως πρὸς τὴν σχέσιν (15) ἔχομεν $e = 1.64$ ἤτοι εἰς τὴν προκειμένην περίπτωσιν ὁ ἀναλογικοῦ–τύπου ἐκτιμητὴς εἶναι κατὰ 64% ἀκριβέστερος τοῦ ἁπλοῦ ἀμερολήπτου ἐκτιμητοῦ $\hat{q}$.

6. Ἀσυμπτωτικὴ κατανομὴ τοῦ ἐκτιμητοῦ

Συμβολίζοντες τὸ διάνυσμα-γραμμὴ ($\pi_{i1}, \pi_{i2}, \ldots, \pi_{is}$) $i = 0, 1, 2, \ldots\ldots, r$ (δηλαδὴ τὴν i γραμμὴν τῆς μήτρας Π) μὲ π'_i αἱ σχέσεις (2) καὶ (3) γράφονται ὡς ἑξῆς

$$q = \sum_i p_i \pi_i \tag{16}$$

καὶ

$$\bar{q} = \sum_i p_i \hat{\pi}_i \tag{17}$$

Ἐκ τῆς τελευταίας σχέσεως προκύπτει ὅτι ὁ «ἀναλογικοῦ – τύπου» ἐκτιμητὴς $\bar{q}$ εἶναι γραμμικὸς συνδυασμὸς ἐκτιμητῶν μεγίστης πιθανοφανείας καὶ κατὰ συνέπειαν καὶ ὁ ἐκτιμητὴς οὗτος εἶναι συνεπὴς (consistent). Ἡ ἀσυμπτωτικὴ κατανομὴ ($n \rightarrow \infty$) τοῦ ἐκτιμητοῦ $\bar{q}$ δίδεται ἐκ τοῦ κατωτέρω θεωρήματος.

Θεώρημα 3. Ἡ ἀσυμπτωτικὴ κατανομὴ ($n \rightarrow \infty$) τοῦ διανύσματος $\sqrt{n}\,(\bar{q} - q)$ εἶναι κανονικὴ μὲ μέσον διάνυσμα 0 καὶ μήτρα διακυμάνσεων–συνδιακυμάνσεων τὴν

$$W = (w_{jk}) \tag{18}$$

ὅπου

$$w_{jj} = \sum_i p_i \pi_{ij} (1 - \pi_{ij}) = v_{jj}$$

καὶ

$$w_{jk} = - \sum_i p_i \pi_{ij} \pi_{ik} = v_{jk}$$

Ἀπόδειξις: Ἐκ τῶν γνωστῶν ἰδιοτήτων καὶ σχέσεων διὰ τὴν πολυωνυμικὴν κατανομὴν προκύπτει ὅτι τὸ διάνυσμα μὲ στοιχεῖα $\sqrt{n_i}\,(\hat{\pi}_i - \pi_i)$ $i = 0, 1, 2, \ldots\ldots, r$ ἔχει ἀσυμπτωτικὴν κατανομὴν ἡ ὁποία εἶναι κανονικὴ μὲ μέσον διάνυσμα 0 καὶ μήτρα διακυμάνσεων—συνδιακυμάνσεων τὴν

$$W_i = \begin{bmatrix} \pi_{i1}(1-\pi_{i1}) & -\pi_{i1}\pi_{i2} & \dots & -\pi_{i1}\pi_{is} \\ \dots & \dots & \dots & \\ -\pi_{ij}\pi_{i1} & \pi_{ij}(1-\pi_{ij}) & \dots & -\pi_{ij}\pi_{is} \\ \dots & \dots & \dots & \\ -\pi_{is}\pi_{i1} & -\pi_{is}\pi_{ij} & \dots & \pi_{is}(1-\pi_{is}) \end{bmatrix}$$

Εἶναι ἐπίσης γνωστὸν ὅτι διὰ $n \rightarrow \infty$

$$\text{plim} \frac{n_{i\cdot}}{np_i} = 1$$

καὶ συνεπῶς τὸ διάνυσμα $\sqrt{n_{i\cdot}}\ (\hat{\pi}_i - \pi_i)$ ἔχει τὴν ἰδίαν ἀσυμπτωτικὴν κατανομὴν μὲ τὸ διάνυσμα $\sqrt{np_i}\ (\hat{\pi}_i - \pi_i)$. Ἐξ αὐτοῦ προκύπτει ὅτι τὸ διάνυσμα $\sqrt{n} \cdot p_i\ (\hat{\pi}_i - \pi_i)$ ἔχει ἀσυμπτωτικὴν κατανομὴν τὴν κανονικὴν τοιαύτην μὲ μέσον διάνυσμα 0 καὶ μήτρα διακυμάνσεων—συνδιακυμάνσεων τὴν $p_i w_i$.

Δι' ἐφαρμογῆς τῶν σχέσεων (16) καὶ (17) προκύπτει ὅτι

$$\sqrt{n}\ (\bar{q} - q) = \sum_i \sqrt{n}\ p_i\ (\hat{\pi}_i - \pi_i)$$

καὶ συνεπῶς τὸ διάνυσμα $\sqrt{n}\ (\bar{q} - q)$ ὡς ἄθροισμα διανυσμάτων κατανεμομένων ἀσυμπτωτικῶς κανονικῶς ἔχει καὶ τοῦτο ὡς ἀσυμπτωτικὴν κατανομὴν τὴν κανονικὴν μὲ μέσον διάνυσμα 0 καὶ μήτρα διακυμάνσεων—συνδιακυμάνσεων τὸ ἄθροισμα $\sum_i p_i w_i = w$, δεδομένου ὅτι τὰ διανύσματα $\hat{\pi}_i - \pi_i$ $i = 0,1,2,\dots.r$ εἶναι ἀσυσχέτιστα.

7. Ἓν παράδειγμα (πραγματικὰ δεδομένα)

Τὰ δεδομένα τοῦ κατωτέρω παραδείγματος προέρχονται ἐκ δειγματοληπτικῆς τινος ἐρεύνης ἡ ὁποία διεξήχθη κατὰ τὰ ἔτη 1965 καὶ 1966 εἰς τὴν πολιτείαν τῆς Καλιφορνίας (USA).

Τὸ ἔτος 1965 δεδομένα 11.000 περίπου ἀτόμων ἡλικίας 16 ἐτῶν καὶ ἄνω περιελήφθησαν εἰς τὴν ἔρευναν καὶ ὁ πίναξ I δίδει τὴν κατανομὴν αὐτῶν ὡς πρὸς τὰς συνθήκας ἀπασχολήσεως.

Ἓν ἔτος ἀργότερον (1966) ἐν ὑπό-δειγμα 2.000 περίπου ἐκ τῶν ἀνωτέρω προσώπων ἠρευνήθη ἐκ νέου καὶ εἰς τὸν πίνακα 2 δίδεται ἡ ταξινόμησις τοῦ ἐν λόγῳ δείγματος τόσον ὡς πρὸς τὴν κατάστασιν ἀπασχολήσεως αὐτῶν τοῦ ἔτους 1965 (βοηθητικὴ μεταβλητὴ A) ὅσον καὶ ὡς πρὸς τὴν τοιαύτην τοῦ 1966 (κυρία μεταβλητὴ M).

Π ί ν α ξ 1

Συνθῆκαι ἀπασχολήσεως		Ποσοστὸν
Ἀπασχολούμενοι	(E)	53.8
Ἄνεργοι	(U)	2.5
Μὴ ἐνεργὸς πληθυσμὸς	(R)	43.7

Ὑποθέτομεν εἰς τὸ παρὸν παράδειγμα ὅτι τὸ δεῖγμα τοῦ ἔτους 1965 εἶναι ὁ συνολικὸς πληθυσμὸς καὶ συνεπῶς τὸ πρόβλημά μας εἶναι νὰ ἐκτιμήσωμεν τὴν κατανομὴν (q_0, q_1, q_2) τοῦ ἐν λόγῳ πληθυσμοῦ ὡς πρὸς τὴν κατάστασιν ἀπασχολήσεως αὐτοῦ τοῦ ἔτους 1966.

Π ί ν α ξ 2

1965 \ 1966	E	U	R	Σύνολον
E	1,023	14	62	1,099
U	23	11	7	41
R	126	10	794	930
Σύνολον	1,772	35	863	2,070

Χρησιμοποιοῦντες τὸν συμβολισμὸν τὸν ὁποῖον ἠκολουθήσαμεν μέχρι τώρα καὶ ἀντιστοιχοῦντες τὰς κατηγορίας Ἀπασχολουμένων, Ἀνέργων καὶ Μὴ ἐνεργοῦ πληθυσμοῦ πρὸς τοὺς ἀριθμοὺς 0, 1, 2 ἔχομεν

$$p' = (p_0 = .538, \; p_1 = .025, \; p_2 = .437)$$

Ἐκ τοῦ Πίνακος 2 εὑρίσκομεν τὰς ἐκτιμήσεις τῶν δεσμευμένων πιθανοτήτων π_{ij}, $i, j = 0, 1, 2$ (ἴδε (4)) αἱ ὁποῖαι δίδονται εἰς τὸν πίνακα 3.

Πίναξ 3

Ἐκτιμήσεις τῶν δεσμευμένων πιθανοτήτων π_{ij}

1965 \ 1966	0	1	2	Σύνολον
0	.931	.013	.056	1,000
1	.561	.268	.171	1,000
2	.135	.011	.854	1,000

Δι' ἐφαρμογῆς τῆς σχέσεως (3) εὑρίσκομεν τὰς ἀκολούθους «ἀναλογικοῦ–τύπου» ἐκτιμήσεις

$$(\bar{q}_0 = .574 \quad \bar{q}_1 = .019 \quad \bar{q}_2 = .407)$$

ἐνῷ αἱ κλασσικαὶ ἁπλαῖ ἀμερόληπτοι ἐκτιμήσεις ὡς προκύπτουν ἐκ τοῦ πίνακος 2 (ὁριζόντια σύνολα) εἶναι

$$(\hat{q}_0 = .566 \quad \hat{q}_1 = .017 \quad \hat{q}_2 = .417)$$

Διὰ νὰ ἀποκτήσωμεν μίαν ἰδέαν τοῦ ἐπιτευχθέντος κέρδους εἰς ἀκρίβειαν (ἀποτελεσματικότητα) ἐκ τῆς χρησιμοποιήσεως τοῦ «ἀναλογικοῦ – τύπου» ἐκτιμητοῦ $\bar{q}$ ἔναντι τοῦ κλασσικοῦ ἐκτιμητοῦ $\hat{q}$ ἀντικαθιστοῦμεν τὰς μήτρας διακυμάνσεων–συνδιακυμάνσεων $V(\hat{q})$ καὶ $V(\bar{q})$ μὲ τὰς δειγματοληπτικὰς ἐκτιμήσεις των.

Ἐπειδὴ

$$\left| \hat{V}(\hat{q}) \right| = \frac{1}{n} \hat{q}_0 \cdot \hat{q}_1 \cdot \hat{q}_2 = \frac{.004012}{2070}$$

καὶ

$$\left| \hat{V}(\bar{q}) \right| = \frac{.001405}{2070} \quad (\text{ἴδε } (11))$$

λαμβάνομεν

$$e = \frac{.004012}{.001405} = 2.86$$

ἐκ τῆς ὁποίας καθίσταται προφανὲς ὅτι ὁ «ἀναλογικοῦ–τύπου» ἐκτιμητὴς $\bar{q}$

εἶναι περίπου τρεῖς φορὰς ἀκριβέστερος τοῦ κλασσικοῦ ἐκτιμητοῦ $\hat{q}$.

Χρησιμοποιοῦντες ἐπίσης ἀντὶ τῶν δεσμευμένων πιθανοτήτων π_{ij} τὰς ἐκτιμήσεις αὐτῶν $\hat{\pi}_{ij}$, ὑπολογίζομεν κατωτέρω τὸ μέγεθος τοῦ σφάλματος εἰς τὰ στοιχεῖα τῆς μήτρας διακυμάνσεων – συνδιακυμάνσεων $V(\bar{q})$ — ὡς αὕτη δίδεται ὑπὸ τῆς σχέσεως (9) — ὡς ἐπίσης καὶ τὸ μέγεθος τῶν μεροληψιῶν (biases) εἰς τὰ στοιχεῖα τῆς ἐκτιμήσεως τῆς $\hat{V}(\bar{q})$ ὡς αὕτη ὁρίζεται ὑπὸ τῆς σχέσεως (11).

Ἓν ἀνώτερον φράγμα τῶν σχετικῶν σφαλμάτων e_j (εἰς τὰ στοιχεῖα τῆς διαγωνίου τῆς μήτρας $V(\bar{q})$) εἶναι

$$e_j < \frac{1}{n} \cdot \frac{\sum_i (1 - p_i)\, \pi_{ij}\, (1 - \pi_{ij})}{\sum_i p_i\, \pi_{ij}\, (1 - \pi_{ij})}$$

Ἐκ τῆς ἀνωτέρω σχέσεως προκύπτει ὅτι $e_1 < .6\,\%$ καὶ $e_2 < .1\,\%$ τὰ ὁποῖα εἶναι προφανῶς ἀμελητέαι ποσότητες.

Ὁμοίως ἓν ἀνώτερον φρᾶγμα τοῦ σχετικοῦ μεγέθους μεροληψιῶν τῶν στοιχείων τῆς διαγωνίου τῆς μήτρας $\hat{V}(\bar{q})$ δίδεται ὑπὸ τῆς σχέσεως

$$\left| b_j \right| < \frac{1}{n} \cdot \frac{\sum_i \pi_{ij}\, (1 - \pi_{ij})}{\sum_i p_i\, \pi_{ij}\, (1 - \pi_{ij})}$$

ἐκ τῆς ἐφαρμογῆς τῆς ὁποίας προκύπτει ὅτι $|b_1| < .7\,\%$ καὶ $|b_2| < .2\,\%$ ποσότητες ἐπίσης ἀμελητέαι. Διὰ τὰ ἐκτὸς τῆς κυρίας διαγωνίου στοιχεῖα τῶν μητρῶν $V(\bar{q})$ καὶ $\hat{V}(\bar{q})$ προκύπτουν εὐκόλως παρόμοια συμπεράσματα

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Cochran, W. G., Sampling Techniques, Second Edition, J. Wiley (1963).

2. Hansen, M. H., Hurwitz, W. N., and Madow, S. G., Sample Survey Methods and Theory. Vol. I and II. J. Wiley (1953).

3. California State Department of Public Health. Human Population Laboratory Survey: Alameda County Population 1965. Series A, No. 7. April 1966.