Η ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΩΣ ΟΡΓΑΝΟΝ ΣΤΑΤΙΣΤΙΚΗΣ ΕΡΕΥΝΗΣ

(Περιληπτικῶς)

Ὑπὸ Ἀντισμηνάρχου Ι. Α. ΣΑΚΑΛΗ
Διευθυντοῦ Στατιστικῆς Γεν. Ἐπιτελείου Ἀεροπορίας

ΚΕΦΑΛΑΙΟΝ Ι. ΕΙΣΑΓΩΓΙΚΑΙ ΕΝΝΟΙΑΙ

Ἡ Στατιστικὴ Μέθοδος κατὰ τὴν ἐφαρμογήν της ἐπὶ δοθέντος πλήθους δεδομένων βοηθεῖ ἡμᾶς ν' ἀναλύσωμεν τὰ δεδομένα ταῦτα διττῶς. Ὅταν δηλαδὴ ἔχωμεν πρὸ ἡμῶν ἓν σύνολον μονάδων, τῶν ὁποίων ἐπιθυμοῦμεν νὰ μελετήσωμεν χαρακτηριστικά τινα, εἴτε πρόκειται περὶ ἀνθρώπων τῶν ὁποίων θέλομεν νὰ γνωρίσωμεν, ἐπὶ παραδείγματι, τὴν ἡλικίαν, εἴτε περὶ ἀντικειμένων διὰ τὸ βάρος αὐτῶν, δύο μέθοδοι δύνανται νὰ χρησιμοποιηθῶσι : Ἡ Μέθοδος τῆς Ἀπογραφῆς (Census) καὶ ἡ Μέθοδος τῆς Δειγματοληψίας (Sampling). Κατὰ τὴν πρώτην ἐξετάζομεν διαδοχικῶς ὅλους τοὺς ἀνθρώπους ἢ πράγματα τοῦ ὑπὸ ἔρευναν πληθυσμοῦ πρὸς προσδιορισμὸν τῆς ἡλικίας των ἢ τοῦ βάρους. Κατὰ τὴν δευτέραν περιοριζόμεθα εἰς τὴν λῆψιν μερικῶν μόνον μονάδων ἔκ τινος πολυαρίθμου κατὰ κανόνα πληθυσμοῦ, καὶ τὴν ἐξ αὐτοῦ συναγωγὴν συμπερασμάτων, τὰ ὁποῖα θὰ ἰσχύωσι δι' ὁλόκληρον τὸν πληθυσμόν. Ἐφαρμόζομεν τουτέστι τὴν γνωστὴν ἐκ τῆς Λογικῆς μέθοδον τῆς Ἐπαγωγῆς διὰ τὴν ἐξαγωγὴν ἐκ μερικῶν περιπτώσεων ἢ παρατηρήσεων γενικωτέρων τινῶν ἀληθειῶν, ἐξ οὗ καὶ ἡ ἑτέρα ὀνομασία τῆς Δειγματοληψίας ὡς Στατιστικῆς Ἐπαγωγῆς (Statistical Inference). Τὰ ἀποτελέσματα τῆς Δειγματοληψίας θὰ εἶναι προφανῶς καλὰ ἢ οὐ ἐφ' ὅσον τὸ ληφθὲν δεῖγμα θὰ εἶναι ἀντιπροσωπευτικὸν ἢ ὄχι τοῦ μελετωμένου συνόλου. Οὕτω ἐπὶ παραδείγματι κατὰ τὴν ἐξέτασιν τοῦ πληθυσμοῦ διαμερίσματός τινος ἐὰν λάβωμεν ὡς δεῖγμα μόνον τοὺς ἄγοντας ἡλικίαν 15 - 25 ἐτῶν, δὲν εἶναι δυνατὸν νὰ ἀποκομίσωμεν καμμίαν ἔνδειξιν ἐπὶ τῆς μέσης ἡλικίας τοῦ ὑπ' ὄψιν πληθυσμοῦ, ὡς ἐπίσης ἐὰν ἐκ τῶν προϊόντων ἐργοστασίου τινὸς ἐκλέξωμεν τὰ πλέον ἐλαφρὰ δὲν δυνάμεθα νὰ γνωρίσωμεν τὸ μέσον βάρος τῶν παραχθέντων ἀντικειμένων. Τὰ μνημονευθέντα παραδείγματα τονίζουσιν ἀφ' ἑαυτῶν ὅτι ἡ ἀπαίτησις τῆς ἀντιπροσωπευτικότητος τοῦ δείγματος τυγχάνει οὐσιώδης καὶ μόνον αὐτὴ δύναται νὰ μᾶς προσφέρῃ πληροφορίας ἐπὶ τοῦ ζητουμένου συνόλου, πρᾶγμα ὅπερ ἐπιτυγχάνεται βάσει Μεθόδων στηριζομένων κυρίως ἐπὶ τοῦ Τυχαίου καὶ τοῦ Λογισμοῦ τῶν Πιθανοτήτων. Μερικὰ ἀκόμη παραδείγματα θέλουσιν ἀποσαφηνίσει τὰς ἐκτεθείσας ἐννοίας. Εἰς τὰς Φυσικὰς Ἐπιστήμας συχνάκις ἀποβαίνει ἀδύνατον νὰ ἐπιτύχωμεν περαιτέρω δεδομένα καὶ υἱοθετοῦμεν τὴν δειγματοληψίαν π.χ. ὅταν ἓν πείραμα δὲν δύναται νὰ ἐκτελεσθῇ πέραν ὡρισμένου δοθέντος χρονικοῦ ὁρίου. Ὁμοίως ἀποτελέσματα ἐπιστημονικοῦ πειράματος, ἐπαναληφθέντος δεκάκις, δύνανται νὰ χρησιμεύσωσιν ὡς γενίκευσις τῶν ἀποτελεσμάτων, τὰ ὁποῖα λογικῶς ὑποτίθενται ὅτι εἶναι ἐπιτευκτὰ ἐὰν τὸ πείραμα ἤθελεν ἐκτελεσθῇ εἰς ἀπεριόριστον ἀριθμὸν χρονικῶν περιόδων. Εἰς τὴν Ἐκπαίδευσιν διὰ νὰ ἔχωμεν τὸν Μέσον Δείκτην Εὐφυΐας (Intelligent Quotient) τῶν μαθητῶν τῶν Δημοτικῶν Σχολείων, ἡ τυχὸν χρησιμοποίησις μεθόδου διαφόρου τῆς Δειγματοληψίας συνεπάγεται τὴν δαπανηρὰν συσσώρευσιν ἀπεράντου πλήθους δεδομένων. Ἔτι περαιτέρω ἔστω ὅτι θέλομεν νὰ λάβωμεν τὴν μέσην τιμὴν τοῦ ἄρτου εἰς τὴν Πόλιν τῶν Ἀθηνῶν. Προφανῶς τόσον ὁ παράγων τοῦ κόστους ὅσον καὶ ὁ ἀπαιτούμενος χρόνος διὰ πλήρη ἔρευναν τῶν ἑκατοντάδων ἀρτοποιείων καὶ ζαχαροπλαστείων ἀποτελοῦσιν ἀνασχετικοὺς παράγοντας, ἀλλ' ἐκ παραλλήλου καὶ ἐὰν λάβωμεν τιμὰς μόνον ἐκ τῶν κεντρικῶν καταστημάτων τὸ δεῖγμα μας θὰ ἦτο μεροληπτικόν. Διὰ νὰ ἐξασφαλίσωμεν ἀντιπροσωπευτικὰ δεδομένα εἶναι ἀνάγκη νὰ λάβωμεν τιμὰς δειγμάτων ἐκ τῶν καταστημάτων ὅλων τῶν ποικιλιῶν ἢ κατὰ τεχνικὴν ὁρολογίαν ἐκ τοῦ Συνόλου Πληθυσμοῦ, τοῦ πληθυσμοῦ νοουμένου ὑπὸ τὴν στατιστικήν του ἔννοιαν ὡς τοῦ συνόλου τῶν δεδομένων ἐξ ὧν λαμβάνεται τὸ δεῖγμα, ἐὰν τὸ ὅλον τούτου ἦτο ἐπιτευκτόν.

Διὰ νὰ ἐπιτύχωμεν ἓν ἀντιπροσωπευτικὸν δεῖγμα[1] δέον νὰ πληροῦνται αἱ ἑξῆς τέσσαρες συνθῆκαι :

α) Τὸ δεῖγμα νὰ ἐκλέγεται ἄνευ μεροληψίας ἢ προκαταλήψεως.

β) Τὰ μέρη τοῦ Δείγματος δέον νὰ εἶναι τελείως ἀνεξάρτητα ἀλλήλων.

γ) Ὑποκειμενικαὶ διαφοραὶ μεταξὺ τῶν περιοχῶν, ἐξ ὧν ἐκλέγονται τὰ δεδομένα, δέον νὰ μὴν ὑπάρχωσι.

δ) Δέον νὰ ὑφίστανται αἱ αὐταὶ δυνατότητες δι' ὅλα τὰ ποσὰ ἐν τῷ Δείγματι.

Οἱοσδήποτε ἔχει χύσει διαλελυμένον οὐΐσκυ εἰς ἕνα κύπελλον δοκιμῆς, ἀφοῦ ἐπῆρε μία ρουφηξιά, ἀσυναισθήτως ἔχει σχετικὴν πεῖραν τῆς δειγματοληψίας. Εἶναι εὔκολον ὅθεν ν' ἀντιληφθῇ τις ὅτι δὲν εἶναι ἀναγκαῖον νὰ εἶναι πτυχιοῦχος Μαθηματικὸς διὰ νὰ δύναται λογικῶς ν' ἀντιμετωπίζῃ τὰ δειγματολογικὰ θέματα κατὰ πρακτικὸν τρόπον. Παρὰ ταῦτα ὁ χειρισμὸς τοῦ προβλήματος τῆς δειγματοληψίας κατὰ τὸ πλεῖστον περιωρίσθη εἰς ἀρκετὰ στρατοσφαιρικὸν ἐπίπεδον. Περίπλοκοι τύποι καὶ ὅροι ἐπενοήθησαν μὲ ἀποτέλεσμα τὴν ἐνόχλησιν καὶ σύγχυσιν τοῦ μέσου ἀναγνώστου. Μία ἀρκούντως διαδεδομένη μεταξὺ τῶν Στατιστικῶν τάσις, ἀλλὰ δυστυχῶς παρατηρουμένη καὶ εἰς τὸν κύκλον τῶν Νομικῶν, Ἰατρῶν καὶ Οἰκονομολόγων, εἶναι νὰ χρησιμοποιῶσιν καθ' ὑπερβολὴν τὰ μαθηματικὰ καὶ τοὺς τεχνικοὺς ὅρους, πρᾶγμα ὅπερ δημιουργεῖ παρὰ τοῖς πολλοῖς τὴν εὔλογον ὑπόνοιαν ὅτι τοῦτο γίνεται ὄχι τόσον διὰ περισσοτέρας διευκρινίσεις ὅσον διὰ σύγχυσιν μᾶλλον καὶ δικαιολόγησιν τῶν ζητουμένων ἑκάστοτε ὑψηλῶν ἀμοιβῶν. Στατιστικοὶ ὅροι ὡς «Leptocurtic», «Homoscedasticity», «Interpenetrating Replicate Subsamples», καίτοι περιγραφικοὶ αὐτοὶ καθ' ἑαυτούς, θὰ ἠδύναντο νὰ ἔχωσιν ἁπλουστευθῇ ὥστε νὰ γίνωσι κτῆμα τῶν πάντων. Βεβαίως αἱ ἀρχαὶ τῆς Δειγματοληψίας στηρίζονται ἐπὶ τῆς Θεωρίας τῶν Πιθανοτήτων καὶ ἀπαιτοῦσι μίαν σχετικῶς ἐπαρκῆ μαθηματικὴν ὑποθεμελίωσιν, ἀλλὰ

---

1. Ἡ ἀντιπροσωπευτικὴ δειγματοληψία ἀπαντᾶται εἰς τὰς σφυγμομετρήσεις τῆς Κοινῆς Γνώμης τοῦ Ἰνστιτούτου GALLUP, ὅπου λαμβάνεται φροντὶς ὅπως ἐξασφαλισθῇ ἡ εἰς καταλλήλους ἀναλογίας ἀντιπροσώπευσις τῶν διαφόρων τάξεων τοῦ πληθυσμοῦ χωρὶς νὰ ἀφήνεται εἰς τὴν Τύχην ὁ προσδιορισμὸς τῶν ἀναλογιῶν αὐτῶν.

πέρα τούτων ἀνήκει εἰς τὴν Μαθηματικὴν Στατιστικὴν ἡ ἐξονυχιστικὴ διερεύνησις τοῦ μαθηματικοῦ μέρους.

Ἂν καὶ ἡ Δειγματοληψία κερδίζει σταθερῶς περισσότερον ἔδαφος εἰς τοὺς Ἰδιωτικοὺς Ὀργανισμοὺς καὶ τὰς Κρατικὰς Λειτουργίας, οὐχ ἧττον πολλοὶ παραμένουσιν εἰσέτι σκεπτικοὶ ἂν ὄχι τελείως δύσπιστοι ἐπὶ τῆς ἀκριβείας οἱωνδήποτε ἀριθμῶν ἐπιτυγχανομένων ἔκ τινος μερικοῦ μᾶλλον ἢ πλήρους ὑπολογισμοῦ. Ἔν τινι μέτρῳ τοῦτο ἀντικατοπτρίζει μίαν φυσικὴν ἀπροθυμίαν δι' ἀποδοχὴν μιᾶς ἀποδείξεως, παρεχομένης ἔκ τινος μικροῦ ἀριθμοῦ ὡς ὀρθοῦ ἔναντι τῆς συνολικῆς ὑπὸ μελέτην ὁμάδος. Ἀλλὰ καὶ πάλιν τοῦτο ὀφείλεται εἰς ἀπροσέκτους ἐφαρμογὰς τῶν μεθόδων δειγματοληψίας κατὰ τὸ παρελθόν.

Ὁ πολὺς κόσμος ἀσφαλῶς δὲν γνωρίζει ὅτι δι' ὡρισμένους τύπους μεθόδων δειγματοληψίας καθορίζεται ἐκ τῶν προτέρων ὁ βαθμὸς τῆς ἐπιθυμητῆς ἀκριβείας ἐκ τῆς ἐκτιμήσεως τοῦ δείγματος. Ἐπίσης εἶναι ἀρκετὰ ἐνδιαφέρον νὰ σημειώσωμεν ὅτι ἡ ἀκρίβεια συγκεκριμένου δείγματος τοιούτου τύπου δύναται νὰ ἐκτιμηθῇ ἐξ αὐτοῦ τούτου τοῦ δείγματος. Ἐπιπροσθέτως πολλοί, οἱ ὁποῖοι ἐπιμένουσιν ὅτι ὁ ἀκριβέστερος τρόπος εἶναι νὰ κάμωσι πλήρη καταμέτρησιν, παραβλέπουσι τὸ γεγονὸς ὅτι πολλὰς φορὰς ὑπάρχουσι πηγαὶ σφάλματος εἰς τὰ ἀρχικὰ δεδομένα καὶ ὅτι μία 100% καταμέτρησις δύναται νὰ ἀποβῇ τελείως ἐσφαλμένη ἂν ὄχι σχετικῶς ἀδύνατος πρὸς ἐπίτευξιν. Πράγματι πολλάκις τὸ δεῖγμα δύναται νὰ ἀποφέρῃ περισσότερον ἀκριβῆ ἀποτελέσματα ἢ μία ἀποπερατωθεῖσα πλήρης καταμέτρησις, δοθέντος ὅτι αἱ πηγαὶ τοῦ σφάλματος δύνανται νὰ ἐλεγχθῶσι πλέον ἀποτελεσματικῶς ὅταν σχετικῶς μικρὸς ἀριθμὸς δεδομένων πρόκειται νὰ ἐξετασθῇ. Ὡς παράδειγμα τούτου, ἔστω ἡ ἐκτίμησις τῆς ἀξίας τῶν ὑλικῶν ἀπογραφῆς 500.000 ἀναλωσίμων ὑλικῶν καὶ ἡ λῆψις δείγματος 60.000 ἐκ τοῦ πληθυσμοῦ. Ἡ ἐκτίμησις βάσει τοῦ δείγματος εὑρέθη ὅτι ἔφθασε κατὰ 8% κάτω τῆς συνολικῆς ἀξίας, ἥτις ἐλήφθη ἐκ τῆς πλήρους ἀπαριθμήσεως τῶν 500.000 ὑλικῶν. Ὁπωσδήποτε τόσον βασικὰ στοιχεῖα ἠλέγχθησαν προσεκτικῶς διὰ τὰ 60.000 ὑλικὰ τοῦ δείγματός μας, ὡς ἐπίσης ἐγένοντο καὶ πολλαὶ ἀναθεωρήσεις, ἐνῷ πάντα ταῦτα θὰ ἦσαν ἀδύνατα διὰ τὰ ὑπόλοιπα 440.000 ὑλικά. Ἐκ τῶν ἀνωτέρω συνάγεται ὅτι ὑφίσταται περισσοτέρα τοῦ δέοντος πιθανότης ὅτι ἡ ἐκτίμησις τοῦ δείγματος θὰ πλησιάζῃ τὴν ἀληθῆ τιμὴν παρ' ὅσον ἡ συνολικὴ ἀπαρίθμησις.

Ἐν τέλει ὑποσημειοῦμεν ὅτι ἀναμφιβόλως ἡ Δειγματοληψία προσφέρει πολυτίμους ὑπηρεσίας. Ἡ πεῖρα τοῦ παρελθόντος ὑποδεικνύει τὴν ἐξάπλωσίν της μὲ ταχύτατα βήματα. Μέχρι τοῦδε ἐχρησιμοποιήθη ἐλάχιστα καὶ δὴ ἐπιβοηθητικῶς, ἀλλὰ ἡ προοπτικὴ τοῦ μέλλοντος ἐπιβάλλει πλέον τὴν Δειγματοληψίαν ὡς τὸ μοναδικὸν ὄργανον ἐρεύνης, μὲ τὸ ὁποῖον καλοῦνται νὰ ἐξοικειωθῶσιν οἱ ἀσχολούμενοι εἰς τὰς Στατιστικὰς Ὑπηρεσίας. Δείγματα δύνανται νὰ ληφθῶσιν ἐκ τοῦ πληθυσμοῦ κατὰ τοὺς κατωτέρω τρόπους.

ΚΕΦΑΛΑΙΟΝ ΙΙ. ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (SIMPLE RANDOM SAMPLING)

Μία ἐπιλογὴ ἔχει γίνει κατὰ τύχην ὅταν ἑκάστη μονὰς τοῦ πληθυσμοῦ ἔχει τὴν αὐτὴν πιθανότητα ἐπιλογῆς ὅπως ἀκριβῶς καὶ εἰς τὴν περίπτωσιν τῆς κληρωτίδος, ἡ ὁποία περιέχει δέκα σφαιρίδια ἠριθμημένα ἀπὸ τοῦ 0 ἕως 9, ἐκ τῶν ὁποίων ἐξάγομεν μὲ κλειστοὺς ὀφθαλμοὺς ἓν σφαιρίδιον κατὰ τύχην. Ἐφ' ὅσον τὰ σφαιρίδια εἶναι ἀπολύτως ὅμοια, θὰ ἔχωσι βεβαίως τὴν αὐτὴν πιθανότητα νὰ ἐξαχθῶσιν ἐκ τῆς κληρωτίδος.

Τὸ Τυχαῖον Δεῖγμα ἀποτελεῖ ἀντιπροσωπευτικὴν ὁμάδα τοῦ Συνόλου. Ὡς κριτήριον τῆς τυχαίας ἐπιλογῆς ἔχομεν ὅτι ἑκάστη μονὰς τοῦ συνόλου ἔχει τὴν αὐτὴν πιθανότητα νὰ ἐκλεγῇ κατὰ τὴν λῆψιν τοῦ δείγματος ὑπὸ τὴν προϋπόθεσιν ὅτι αἱ γενόμεναι ἐπιλογαὶ εἶναι στατιστικῶς ἀνεξάρτητοι. Τὸ ἰδιαίτερον γνώρισμα τῆς τυχαίας ἔκ τινος πληθυσμοῦ ἐπιλογῆς εἶναι ὅτι τὸ δεῖγμα δέον νὰ εἶναι ἀποδεκτὸν ὡς ἀντιπροσωπευτικὸν τοῦ συνόλου καὶ τὴν ἀντιπροσωπευτικότητα αὐτὴν ἐξασφαλίζει ὁ Νόμος τῶν Μεγάλων Ἀριθμῶν [2], ἀνεξαρτήτως τῆς ὑποκειμενικότητος τοῦ ἐρευνητοῦ. Κάθε μονὰς ἐκλεγομένη κατὰ τύχην ἀντιπροσωπεύει μόνον ἑαυτὴν καὶ μόνον τὸ σύνολον τῶν ληφθεισῶν μονάδων ἀπαρτίζει τὸ ἀντιπροσωπευτικὸν δεῖγμα.

Ὁ καλλίτερος τρόπος ἐξασφαλίσεως τῆς τυχαίας ἐπιλογῆς εἶναι ἡ χρησιμοποίησις τῶν Πινάκων Τυχαίων Ἀριθμῶν τῶν R. Fisher – F. Yates περιλαμβανόντων 300 ἐν συνόλῳ συμπλέγματα ἀριθμῶν. Ἡ βασικὴ ἀρχὴ τῶν Πινάκων αὐτῶν συνίσταται εἰς τὸ νὰ ὁρίσωμεν ἕνα καὶ μόνον ἀριθμὸν δι' ἑκάστην μονάδα τοῦ ἐρευνωμένου συνόλου, εἰς τρόπον ὥστε ἡ μονὰς αὐτὴ νὰ ἐπιλεγῇ εὐθὺς ὡς ὁ ἀντιπροσωπεύων αὐτὴν ἀριθμὸς ἐπιλεγῇ.

Παράδειγμα: Κατασκευαστὴς γλυκισμάτων ἔχει ἕνα κάλαθον διαφόρων εἰδῶν καρύων καὶ ἐπιθυμεῖ νὰ γνωρίζῃ πόσα κάρυα ὑπάρχουσι κατὰ λίτραν πρὶν ἢ προβῇ εἰς τὴν διὰ σοκολάτας ἐπικάλυψιν. Θὰ ἠδύνατο πράγματι νὰ προσλάβῃ μερικὰ μικρὰ παιδιὰ διὰ νὰ μετρήσωσι τὸν σωρὸν τῶν καρύων καὶ ἔπειτα νὰ διαιρέσῃ τὸν συνολικὸν ἀριθμὸν διὰ 500 διὰ νὰ ἐπιτύχῃ τὸν αἰτούμενον κατὰ λίτραν ἀριθμὸν ἢ νὰ ἐκλέξῃ μερικὰς ἑκατοντάδας καρύων κατὰ τύχην ἐκ τοῦ καλάθου, νὰ τὰς μετρήσῃ καὶ ζυγίσῃ καὶ διαιρέσῃ τὸν ἀριθμὸν διὰ τοῦ βάρους εἰς λίτρας. Ἐὰν τὰ κάρυα εἶχον ἀναμιχθῆ ἀκριβῶς πρὸ πάσης ἐκλογῆς τὸ Δεῖγμα θὰ μᾶς ἔδιδε ὁμοίως ἕνα ἀποδεκτὸν ἀποτέλεσμα. Τὸ σημαντικὸν πρᾶγμα, ὅπερ δέον νὰ ἐνθυμούμεθα, εἶναι ὅτι ἕκαστον κάρυον *πρέπει νὰ ἔχῃ τὴν αὐτὴν πιθανότητα νὰ ἐπιλεγῇ εἰς τὸ δεῖγμα καὶ αὐτὴ εἶναι καὶ ἡ πρώτη ἀπαίτησις τῆς Ἁπλῆς Τυχαίας Δειγματοληψίας* κατὰ τὰ προλεχθέντα.

Διὰ νὰ συνειδητοποιήσωμεν τὴν ἀνωτέρω ἀλήθειαν ἂς ἴδωμεν ἐπὶ παραδείγματι πῶς ἐξάγεται ἓν ἁπλοῦν τυχαῖον δεῖγμα 6 καρύων ἔκ τινος πληθυσμοῦ 18 καρύων, συγκειμένου ἐκ 3 καρύων, 6 λεπτοκαρύων καὶ 9 γαιοκαρύων, καὶ ἂς ἀριθμήσωμεν τὰ 18 κάρυα, προσδιορίζοντες ὑποθετικὰ βάρη δι' ἕκαστον εἰς χιλιοστὰ τοῦ γραμμαρίου.

---

2. Δυνάμει τοῦ Νόμου τῶν Μεγάλων Ἀριθμῶν ἀπεδείχθη θεωρητικῶς καὶ πρακτικῶς ὅτι ὅσον ὁ ἀριθμὸς τῶν παρατηρήσεων αὐξάνει τόσον παραλλήλως αὐξάνει καὶ ἡ ἀκρίβεια τῶν ἀποτελεσμάτων. Ὁ QUETELET εἶχε τὴν ὑπομονὴν νὰ ἐπαληθεύσῃ πειραματικῶς τὸν Νόμον διὰ τοῦ κλασσικοῦ παραδείγματος τῶν 40 βώλων ἐντὸς δοχείου, ἐξ ὧν οἱ 20 λευκοὶ καὶ ἕτεροι 20 μέλανες, καὶ τοὺς ὁποίους ἀνέσυρεν ἕνα πρὸς ἕνα καὶ τοὺς ἐπανέθετεν πάλιν εἰς τὸ δοχεῖον ὥστε νὰ μὴν ἀλλάζουν αἱ συνθῆκαι τοῦ πειράματος. Καίτοι αἱ πιθανότητες ἦσαν

ΚΑΡΥΑ		ΛΕΠΤΟΚΑΡΥΑ		ΓΑΙΟΚΑΡΥΑ	
Ἀριθμὸς Καρύου	Βάρος εἰς MG	Ἀριθμὸς Καρύου	Βάρος εἰς MG	Ἀριθμὸς Καρύου	Βάρος εἰς MG
1	55	4	27	10	8
2	67	5	32	11	12
3	43	6	24	12	8
		7	28	13	11
		8	31	14	7
		9	26	15	9
				16	7
				17	10
				18	9
Σύνολον Βάρους	165		168		81
Μέσος Ὅρος Βάρους	55		28		9
Μέσος Ὅρος Βάρους 18 Καρύων.	23				
Γενικὸν Σύνολον Βάρους 18 Καρύων	414				

Διὰ νὰ ἐπιτύχωμεν μίαν τυχαίαν ἐπιλογὴν 6 Καρύων ἐκ τῶν 18 τοιούτων χρησιμοποιοῦμεν τὸν ἀκόλουθον Πίνακα Τυχαίων Ἀριθμῶν, ἐκλέγοντες τοὺς πρώτους 6 διαφόρους ἀριθμοὺς μεταξὺ 01 καὶ 18.

ΠΙΝΑΞ ΤΥΧΑΙΩΝ ΑΡΙΘΜΩΝ

22	57	53	93
19	48	40	21
$\boxed{16}$	61	02	95
78	36	95	97
$\boxed{03}$	$\boxed{18}$	35	69
93	88	16	04
78	$\boxed{09}$	77	61
23	$\boxed{12}$	46	85
$\boxed{15}$	85	37	21
38	38	61	15

Ἐὰν ὁ ἀναγνώστης ἔχει τὴν δυνατότητα νὰ χρησιμοποιῇ Τυχαίους Ἀριθμούς, καλὸν θὰ εἶναί νὰ ἀναφέρεται μᾶλλον εἰς δημοσιευμένους Πίνακας παρὰ νὰ προσπαθῇ νὰ σταχυολογῇ τοιούτους ἐκ τῆς κεφαλῆς του.

αἱ αὐταὶ νὰ τραβήξωμεν ἕνα λευκὸν καὶ ἕνα μέλανα βῶλον, μολατοῦτα προέκυψαν διάφορα ἀποτελέσματα, τὰ ὁποῖα διορθώνονται μὲ τὸν πολλαπλασιασμὸν τῶν τραβηγμάτων.

'Απλοῦν Τυχαῖον Δεῖγμα

'Αριθμὸς Καρύων	Βάρος εἰς MG
16	7
3	43
15	9
18	9
9	26
12	8
Συνολικὸν Βάρος	102
Μέσον Βάρος	17 (ἐν συγκρίσει πρὸς τὸ 23 ὅλων τῶν 18 καρύων)

Αὐτὸ εἶναι ἓν δεῖγμα ἐκ 18.564 δυνατῶν διαφόρων δειγμάτων ἐξ 6, ἑκάστου ἐκ πληθυσμοῦ 18 καρύων. Ὅλα τὰ δυνατὰ δείγματα ἐξ ἑκάστων τῶν 6 καρύων δύνανται νὰ ταξινομηθῶσι συστηματικῶς κατά τινα τρόπον ὡς ἀκολούθως :

1 2 3 4 5 6	2 3 4 5 6 7	11 13 14 15 16 17
1 2 3 4 5 7	2 3 4 5 6 8	11 13 14 15 16 18
....	....	....
....	....	....
		11 14 15 16 17 18
1 3 4 5 6 7	2 4 5 6 7 8	12 13 14 15 16 17
1 14 15 16 17 18	2 14 15 16 17 18	13 14 15 16 17 18

'Αποδεικνύεται μαθηματικῶς ὅτι τὸ μέσον βάρος τῶν 18564 μέσων Τυχαίων Δειγμάτων ἀνέρχεται εἰς 23 MG, τὸ αὐτὸ δηλαδὴ ὡς ὁ μέσος ὅρος τῶν ἀρχικῶν καρύων εἰς τὸν πληθυσμόν μας.

Ἓν δυσάρεστον χαρακτηριστικὸν τῶν ἁπλῶν τυχαίων δειγμάτων εἶναι ὅτι καθ' ὅσον ἐπιλέγομεν ὁ μέσος ὅρος τῆς ἐπιλογῆς ἀπομακρύνεται ἀρκετὰ ἀπὸ τοῦ ἀληθοῦς μέσου. Οὕτω ἀντὶ τοῦ πραγματικῶς ἐξαχθέντος δείγματος (Μέσον Βάρος 17 MG) θὰ ἠδυνάμεθα νὰ ἔχωμεν ἓν ἐκ τῶν ἀκολούθων δειγμάτων :

ΔΕΙΓΜΑ Α		ΔΕΙΓΜΑ Β	
'Αριθμὸς Καρύων	Βάρος εἰς MG	'Αριθμὸς Καρύων	Βάρος εἰς MG
1	55	10	8
2	67	12	8
3	43	14	7
5	32	15	9
7	28	16	7
8	31	18	9
Συνολικὸν Βάρος	256		48
Μέσον Βάρος	42 2/3		8

Τῇ ἀληθείᾳ αἱ ἐκτιμήσεις τοῦ μέσου βάρους τοῦ πληθυσμοῦ μας τῶν 18 καρύων, ὅπως ἐπετεύχθησαν ἐξ ἑκατέρου ἐκ τῶν δύο τούτων δειγμάτων, εἶναι ὀλίγον πτωχαί.

Τὸ σημειωθὲν μειονέκτημα, τοῦ μικροῦ τουτέστι κινδύνου ἐπιλογῆς ἑνὸς τοιούτου λίαν πτωχοῦ δείγματος, ἐξουδετεροῦται διὰ τῆς Μεθόδου τῆς Στρωματοειδοῦς Δειγματοληψίας.

ΚΕΦΑΛΑΙΟΝ ΙΙΙ. ΣΤΡΩΜΑΤΟΕΙΔΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (STRATIFIED SAMPLING)

Μία μέθοδος διὰ νὰ βελτιώσωμεν τὴν ἐκ τῆς ἁπλῆς τυχαίας δειγματοληψίας ληφθεῖσαν ἐκτίμησιν θὰ ἦτο νὰ ταξινομήσωμεν τὸ πλῆθος τῶν 18 καρύων εἰς στρώματα, προφανῶς δὲ ὁ καλύτερος τρόπος θὰ ἦτο νὰ συνενώσωμεν τὰ κάρυα εἰς τὰς κεχωρισμένας ποικιλίας των (Κάρυα, Λεπτοκάρυα, Γαιοκάρυα), ὁπότε δυνάμεθα νὰ λάβωμεν ἓν τυχαῖον δεῖγμα ἐξ ἑκάστου στρώματος.

Κατὰ τὴν μέθοδον ταύτην δηλαδὴ ὁ ἐρευνητὴς διαιρεῖ τὸν ἀριθμὸν τῶν μονάδων, τὰς ὁποίας κρίνει ἀπαραίτητον νὰ περιληφθῶσιν εἰς τὸ δεῖγμα, εἰς τόσας ὑποομάδας, ὅσαι ὑποδιαιρέσεις ἢ στρώματα (Stratum) ὑπάρχουσιν εἰς τὸ σύνολον καὶ εἰς ἕκαστον τῶν ὑποδειγμάτων τούτων δίδει τὴν ἀνάλογον βαρύτητα. Οἴκοθεν νοεῖται ὅτι καὶ δι' ἑκάστην λῆψιν ὑποδείγματος ἐκ τοῦ ἀρχικοῦ συνόλου τηρεῖται ἡ διαδικασία τῆς τυχαίας ἐπιλογῆς.

Ἀναλογικὴ Δειγματοληψία: Διὰ νὰ λάβωμεν ἓν δεῖγμα ἀπὸ 6 ἐκλέγομεν ἓν ἐκ τῶν τριῶν καρύων, δύο ἐκ τῶν 6 λεπτοκαρύων καὶ τρία ἐκ τῶν 9 γαιοκαρύων, δηλ. τὸ 1/3 ἐκ τῶν καρύων ἑκάστου στρώματος. Χρησιμοποιοῦντες τὸν Περιληπτικὸν Πίνακα Τυχαίων Ἀριθμῶν, ὅπως συνηντήσαμεν τοῦτον εἰς προηγούμενον Κεφάλαιον, ἐκλέγομεν τὸν πρῶτον ἀριθμὸν μεταξὺ 01 καὶ 03, τοὺς πρώτους δύο μεταξὺ 04 καὶ 09 καὶ τοὺς πρώτους τρεῖς μεταξὺ 10 καὶ 18.

Ἀναλογικὸν Στρωματοειδὲς Δεῖγμα

Στρῶμα	Ἀριθμὸς Καρύων	Βάρος εἰς MG
Κάρυα	3	43
Λεπτοκάρυα	9	26
»	4	27
Γαιοκάρυα	16	7
»	15	9
»	18	9
Συνολικὸν Βάρος 6 Καρύων		121

Μέσος Ὅρος 121 : 6 ἴσον 20.16 (ἐν συγκρίσει πρὸς τὸν Μέσον Ὅρον 23 δι' ὅλα τὰ 18 Κάρυα). Δοθέντος ὅτι ἕκαστον κάρυον ἔχει τὴν αὐτὴν πιθανότητα ἐκλογῆς εἰς τὸ δεῖγμα μας (ἓν πρὸς τρία) δὲν εἶναι ἀπαραίτητον νὰ σταθμίσωμεν τὰ ἀποτελέσματα τῶν κεχωρισμένων στρωμάτων διὰ νὰ ἐπιτύχωμεν τὸ μέσον βάρος τῶν 6 Καρύων. Θὰ ἠδύνατό τις ἐκ διαισθήσεως νὰ ἀναμένῃ γενικῶς καλλιτέραν ἐκτίμησιν ἐκ τοῦ ἀνωτέρω στρωματοειδοῦς δείγματος παρὰ ἀπὸ τὸ Ἁπλοῦν Τυχαῖον τῶν 6 Καρύων. Τοῦτο ὀφείλεται εἰς τὸ γεγονὸς ὅτι εἴμεθα βέβαιοι ὅτι ἔχομεν Κάρυα ἐξ ἑκάστης ποικιλίας τοῦ στρωματοειδοῦς δείγματος, ἀποφευγομένων οὕτω δειγμάτων ἀκανονίστων ὡς τὰ 6 Γαιοκάρυα. Εἰς τὴν προκειμένην μέθοδον στρωματοειδοῦς δείγματος ὑφίστανται 3780 δυνατὰ δείγματα, ἐξ ἓξ ἕκαστον, ἐν συγκρίσει πρὸς τὰ 18564 δείγματα τῶν 6 ἐξ ἑκάστου δυνατῶν κατὰ τὴν Ἁπλῆν Τυχαίαν Δειγματοληψίαν.

Ἰσομεγέθη Δείγματα

Ἡ μέθοδος αὐτὴ ἀπαιτεῖ τὴν ἐπιλογὴν δύο Καρύων, δύο Λεπτοκαρύων καὶ δύο Γαιοκαρύων διὰ τὸ δεῖγμα μας τῶν ἕξ. Καὶ πάλιν χρησιμοποιοῦμεν τὸν Πίνακα τῶν Τυχαίων Ἀριθμῶν διὰ νὰ ἐπιλέξωμεν τὰ Κάρυα ἑκάστου στρώματος.

Στρῶμα	Ἀριθμὸς Καρύων	Βάρος εἰς MG	Συντελεστὴς Σταθμίσεως	Σταθμιζόμενα Σύνολα
Κάρυα	3	43	—	— 10.3
	2	67	—	— 4
Σύνολον		110	3/2	165
Λεπτοκάρυα	9	26	—	— 53.6
	4	27	—	— 2
Σύνολον		53	6/2	159
Γαιοκάρυα	16	7	—	— 16·9
	15	9	—	— 2
Σύνολον		16	9/2	72
Γενικὸν Σύνολον				396

Σταθμικὸς Μέσος 396 : 18 ἴσον 22 (ἐν συγκρίσει πρὸς τὰ 23 τοῦ Πλήθους μας).

Πράγματι ἐὰν ἐκλέξωμεν τὰ 2/3 τῶν καρύων τοῦ πλήθους, τὸ 1/3 τῶν λεπτοκαρύων καὶ τὰ 2/9 τῶν γαιοκαρύων θὰ πρέπῃ νὰ προσαρμόσωμεν τὰς ποικιλλούσας αὐτὰς ἀναλογίας διὰ νὰ ἐπιτυχωμεν ἓν ἀποδεκτὸν ἀποτέλεσμα. Ὁ καλούμενος Συντελεστὴς Σταθμίσεως εἶναι ὁ ἀντίστροφος τῆς ἀναλογίας δειγματοληψίας· π.χ. ἐπὶ τῶν 2/3 τῶν καρύων, ἅτινα ἐξελέξαμεν, ὁ Συντελεστής μας θὰ εἶναι 3/2.

'Αρίστη Διάταξις (Optimum Allocation)

Ὅπως καὶ ἡ λέξις δηλοῖ, ἡ μέθοδος αὐτὴ παρέχει τὴν καλλιτέραν δυνατὴν ὑποδιαίρεσιν τοῦ συνολικοῦ μας δείγματος εἰς ἀτομικὰ στρώματα. Δηλαδὴ ἀνεπτύχθησαν τύποι, οἱ ὁποῖοι μᾶς λέγουν τὸ μέγεθος τοῦ δείγματος ἑκάστου στρώματος, ὅπερ θέλει ἀποφέρει τὰς πλέον ἀκριβεῖς ἐκτιμήσεις κατὰ μέσον ὅρον διὰ συγκεκριμένον συνολικὸν μέγεθος δείγματος. Καὶ εἰς τὴν περίπτωσιν αὐτὴν εἶναι περιττὸν νὰ προβῶμεν εἰς προσαρμογὴν διὰ ποικιλλούσας ἀναλογίας τοῦ δείγματος εἰς τὰ κατ' ἰδίαν στρώματα. 'Αφήνοντες κατὰ μέρος τοὺς μαθηματικοὺς τύπους παρατηροῦμεν ὅτι ἡ μέθοδος αὐτὴ μᾶς ὑποδεικνύει νὰ συλλέξωμεν τρία Κάρυα, δύο Λεπτοκάρυα καὶ ἓν Γαιοκάρυον. Χρησιμοποιοῦντες τὸν Πίνακα Τυχαίων 'Αριθμῶν διὰ νὰ ἐπιλέξωμεν τὰ δύο λεπτοκάρυα καὶ τὸ ἓν γαιοκάρυον εὕρομεν τὸν ἀριθμὸν 21 1/2 MG ὡς Σταθμικὸν Μέσον ἔναντι τοῦ 23 τοῦ Πλήθους.

Ἡ μέθοδος αὕτη ἐν τῇ Στρωματοειδεῖ Δειγματοληψίᾳ δύναται νὰ ἐπεκταθῇ εἰς καταστάσεις, ὅπου τὸ κόστος διεξαγωγῆς τῆς δειγματολογικῆς ἐρεύνης ποικίλλει ἀπὸ στρώματος εἰς στρῶμα. Κατ' ἀκολουθίαν ὁσάκις αἱ δυνατότητες τοῦ Προϋπολογισμοῦ δὲν ἐπιτρέπουσι τὴν διεξαγωγὴν μιᾶς ἐρεύνης, ὁ ἀποτελεσματικώτερος τρόπος εἶναι νὰ ὑποδιαιρέσωμεν τὸ συνολικὸν δεῖγμα εἰς στρώματα, ἐὰν δὲ ἔχῃ προκαθορισθῇ καὶ βαθμός τις ἀκριβείας δυνάμεθα νὰ προσδιορίσωμεν τὸ μέγεθος τοῦ δείγματος καὶ τὴν κατάτμησίν του κατὰ στρώματα οὕτως ὥστε νὰ ἐπιτύχωμεν τὸ ἐλάχιστον δυνατὸν κόστος μὲ τὴν μεγίστην ἐπιθυμητὴν ἀκρίβειαν.

ΚΕΦΑΛΑΙΟΝ IV. ΣΩΡΕΥΤΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (CLUSTER SAMPLING)

Ἕτερος τρόπος ἐπιλογῆς δείγματος 6 Καρύων ἐκ τῶν 18 τοῦ μικροσκοπικοῦ πληθυσμοῦ μας εἶναι νὰ συνενώσωμεν τὰ 18 Κάρυα εἰς ἀνομοιογενεῖς σωρούς. Πρὸς τοῦτο δημιουργοῦμεν τρεῖς Σωρούς, ἑκάστου περιέχοντος ἓν κάρυον, δύο λεπτοκάρυα καὶ τρία γαιοκάρυα ὡς κατωτέρω :

ΣΩΡΟΣ I			ΣΩΡΟΣ II			ΣΩΡΟΣ III		
'Αριθμὸς Καρύων	Ποικιλία	Βάρος MG	'Αριθμὸς Καρ.	Ποικιλία	Βάρος MG	'Αριθ. Καρ.	Ποικιλία	Βάρος MG
1	Κάρυον	55	2	Κάρυον	67	3	Κάρυον	43
4	Λεπτοκάρ.	27	6	Λεπτοκάρ.	24	8	Λεπτοκάρ.	31
5	»	32	7	»	28	9	»	26
10	Γαιοκάρ.	8	13	Γαιοκάρ.	11	16	Γαιοκάρ.	7
11	»	12	14	»	7	17	»	10
12	»	8	15	»	9	18	»	9
Σύνολον		142			146			126
Μέσος Ὅρος Σωροῦ		23 4)6			24 2)6			21

Χρησιμοποιοῦντες καὶ πάλιν τὸν μικρὸν Πίνακα Τυχαίων Ἀριθμῶν, ὁ πρῶτος ἀριθμὸς μεταξὺ 01 καὶ 03 πλησιάζει πρὸς τὸ 03, διὸ καὶ ἐκλέγεται ὁ Σωρὸς III ὡς δεῖγμα, ἔνθα τὸ μέσον βάρος εἶναι 21 MG ἔναντι τοῦ μέσου βάρους 23 τοῦ πληθυσμοῦ. Γενικῶς εἰπεῖν τὰ πλέον ἀκριβῆ ἀποτελέσματα κατὰ τὴν μέθοδον αὐτὴν ἀναμένονται ὁσάκις ἕκαστος Σωρὸς περιλαμβάνει ὅσον τὸ δυνατὸν ποικίλλον μῖγμα, ἐνῷ ἐκ παραλλήλου ὁ εἷς Σωρὸς ὁμοιάζει κατὰ τὸ δυνατὸν πρὸς τὸν ἕτερον. Ἡ αἰτιολογία τῆς κατὰ τὰ ἀνωτέρω ληφθείσης τόσον καλῆς ἐκτιμήσεως ἔγκειται εἰς τὸ γεγονὸς ὅτι τὰ κάρυα εἰς ἕκαστον σωρὸν ἦσαν ὅλα ἀναμίξ, ἐνῷ συγχρόνως ἕκαστος ἐκ τῶν τριῶν Σωρῶν ὡμοίαζε πρὸς τοὺς ἄλλους δύο κατὰ πολύ, τουτέστιν εἶχεν ἕκαστος Σωρὸς τὸν αὐτὸν ἀριθμὸν καρύων, λεπτοκαρύων καὶ γαιοκαρύων. Τὰ σημειωθέντα εἰς τὸ προηγούμενον κεφάλαιον κριτήρια διὰ μίαν καλὴν στρωματοειδῆ δειγματοληψίαν ἦσαν ἀκριβῶς τὰ ἀντίθετα, ἤτοι στρώματα ὅσον ἔνεστι ὁμογενῆ ἐσωτερικῶς καὶ ταῦτα δὲ διαφέροντα ἀλλήλων κατὰ τὸ δυνατὸν περισσότερον, τὸ δὲ παράδειγμα ἐπεβεβαίωσε ταῦτα, δηλ. ἕκαστον στρῶμα περιελάμβανε κεχωρισμένην ποικιλίαν καρύων καὶ αἱ τρεῖς δὲ ποικιλίαι ἦσαν τελείως διαφορετικαὶ ἡ μία ἀπὸ τὴν ἄλλην ἐν σχέσει πρὸς τὰ μέσα βάρη. Ἐνῷ λοιπὸν ἡ στρωματοποίησις, λογικῶς ἐκτελουμένη, σχεδὸν πάντοτε ἀποφέρει περισσότερον ἀκριβῆ ἐκτίμησιν ἢ τὸ Ἁπλοῦν Τυχαῖον Δεῖγμα, ἡ Σωρευτικὴ Δειγματοληψία γενικῶς ἀποδίδει ὀλιγώτερον ἀκριβῆ ἐκτίμησιν ἢ τὸ Ἁπλοῦν Τυχαῖον Δεῖγμα τοῦ αὐτοῦ μεγέθους. Ἡ αἰτία τούτου δέον νὰ ἀναζητηθῇ εἰς τὸ ὅτι εἰς πολλὰς ἐν τῇ πράξει καταστάσεις πρέπει νὰ ἀποδεχθῶμεν τὰ δείγματα ὅπως εἶναι, πολλάκις δὲ ἕκαστον ἐκ τούτων εἶναι σχετικῶς ὁμογενές, καίτοι διαφέρει ἀπὸ τῶν ἄλλων στωμμάτων. Κατὰ συνέπειαν ἡ χρῆσις τῆς Σωρευτικῆς Δειγματοληψίας κατὰ γενικὴν ἀναγνώρισιν ὑπαγορεύεται μᾶλλον ἐξ ὑπολογισμῶν κόστους καὶ διοικήσεως.

Ὑποτεθείσθω ἐπὶ παραδείγματι ὅτι μία Ἑταιρία Ξηρῶν Καρπῶν ἐπιθυμεῖ νὰ γνωρίζῃ τὴν μέσην ἐτησίαν κατανάλωσιν τῶν προϊόντων της κατ' οἰκογένειαν εἰς τὴν πόλιν τῆς Θεσσαλονίκης. Ἐὰν μὲν ὑπάρχῃ διαθέσιμος εἷς κατάλογος ὅλων τῶν οἰκογενειῶν τῆς πόλεως θὰ ἠδύνατο νὰ λάβῃ ἓν Τυχαῖον Δεῖγμα, ἔστω 2000 οἰκογενειῶν καὶ ἐκ τούτου νὰ πληροφορηθῇ τὴν μέσην κατανάλωσιν. Παρὰ ταῦτα ἁπλούστερον καὶ εὐθηνότερον διὰ τὴν Ἑταιρίαν εἶναι νὰ ἐκλέξῃ κατὰ τύχην περὶ τὰ 100 οἰκοδομικὰ τετράγωνα τῆς πόλεως μὲ μέσον ὅρον 40 οἰκογενείας εἰς ἕκαστον καὶ ἐν συνεχείᾳ νὰ προβῇ εἰς τελείαν καταγραφὴν τῶν τετραγώνων τούτων. Ὑποτίθεται γενικῶς ὅτι αἱ οἰκογένειαι εἰς ἕνα ἁπλοῦν τετράγωνον θὰ εἶναι σχεδὸν ὁμοιογενεῖς ἐξ ἀπόψεως εἰσοδήματος (ἐν συγκρίσει πρὸς ἕτερον τετράγωνον ἑτέρου τμήματος τῆς αὐτῆς πόλεως, κατοικουμένου λόγῳ θέσεως παρὰ τῶν λίαν εὐπόρων), ἀλλὰ καὶ οἱ ἔνοικοι τῶν τετραγώνων δὲν διαφέρουσι ἀπὸ πλευρᾶς γαστρονομικῶν ὀρέξεων. Οὕτω τὰ στρώματά μας θὰ εἶναι ὁμοιογενῆ καὶ τὸ δεῖγμα τῶν 100 τετραγώνων μὲ τὰς 4000 οἰκογενείας δύναται νὰ παράσχῃ εὐκόλως τόσον ἀκριβὲς ἀποτέλεσμα ὅσον καὶ τὸ εὐρέως διασπειρόμενον τυχαῖον δεῖγμα τῶν 2000 οἰκογενειῶν καὶ μὲ ἐλάχιστον μάλιστα συνολικὸν κόστος. Ἐνῷ λοιπὸν τὸ τυχαῖον δεῖγμα εἶναι πλέον ἀκριβὲς μέγεθος πρὸς μέγεθος εἰς τὸ παράδειγμά μας, τὸ μεγαλύτερον σωρευτικὸν δεῖγμα θὰ εἶναι πλέον ἀποτελεσματικὸν ἐξ ἀπόψεως κόστους. Εἰς τὴν σύγχρονον πρακτικὴν ἡ δειγματοληψία πλήρων στρωμάτων δὲν εἶναι καὶ τόσον συνήθης. Συνηθέστερον τὸ δεῖγμα ἐκλέγεται κατὰ δύο ἢ περισσότερα στάδια. Εἰς δειγματοληψίαν δύο σταδίων τὰ δείγματα ἐπιλέγονται κατὰ τύχην καὶ ἐξάγεται ἓν ὑπόδειγμα κατὰ τύχην ἐξ ἑκάστου στρώματος. Τὰ στρώματα εἰς μίαν τοιαύτην σχεδίασιν δειγμάτων καλοῦνται Ἀρχικαὶ Δειγματολογικαὶ Μονάδες, ἐνῷ τὰ στοιχεῖα τῶν ὑποδειγμάτων καλοῦνται Στοιχειώδεις Μονάδες. Παράδειγμα: ἐκλέγομεν 100 σάκκους διαφόρων εἰδῶν καρύων, βάρους μιᾶς λίτρας, ἔκ τινος πληθυσμοῦ 1000 τοιούτων σάκκων. Οἱ σάκκοι ἢ στρώματα εἶναι αἱ Ἀρχικαὶ Δειγματολογικαὶ Μονάδες καὶ τὰ κάρυα αἱ Στοιχειώδεις Μονάδες.

Ἡ Σωρευτικὴ Δειγματοληψία συχνάκις εἶναι κατορθωτὴ κατὰ τὴν ἑτοιμασίαν δειγματολογικῶν ὑπολογισμῶν ἐκ δεδομένων, περιεχομένων εἰς μέγαν ὄγκον Διατρήτων Δελτίων. Ἐὰν αἱ Καρτέλλαι ἐναποτίθενται εἰς ἀριθμόν τινα ἑρμαρίων, ἕκαστον ἑρμάριον δύναται νὰ θεωρηθῇ ὡς Σωρός. Κατόπιν συλλέγομεν πολλὰ ἑρμάρια κατὰ τύχην καὶ εἴτε ὅλα εἴτε καθοριζόμενον κλάσμα ἐκ τῶν Δελτίων ἑκάστου ἑρμαρίου δύναται νὰ ἐπιλεγῇ διὰ τὸ τελικὸν δεῖγμα.

ΚΕΦΑΛΑΙΟΝ V. ΕΤΕΡΑΙ ΜΕΘΟΔΟΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

1. **Συστηματικὴ Δειγματοληψία (Systematic or Patterned or Serial Sampling).**

Εἰς τὸ παράδειγμά μας τοῦ πληθυσμοῦ τῶν 18 Καρύων θὰ ἠδυνάμεθα νὰ ἐκλέξωμεν ἓν δεῖγμα 6 καρύων ὡς ἀκολούθως: Πρῶτον ἐκλέγομεν κατὰ τύχην ἕνα ἀριθμὸν μεταξὺ 1 καὶ 3, ἔστω π.χ. τὸν ἀριθμὸν 2. Κατόπιν περιλαμβάνομεν εἰς τὸ δεῖγμα μας τὸν ἀριθμὸν καρύου 02 καὶ κάθε τρίτον κάρυον μετέπειτα. Οὕτω τὸ δεῖγμα μας θὰ συνίσταται ἐκ τῶν καρύων 02, 05, 08, 11, 14 καὶ 17.

Ἡ μέθοδος αὐτὴ ἀπαντᾶται συχνάκις καθόσον εἶναι ἁπλῆ, ἄμεσος καὶ ἀνέξοδος. Ὅταν ἔχωμεν διαθέσιμον ἕνα Πίνακα ὀνομάτων ἢ πραγμάτων ἡ συστηματικὴ δειγματοληψία ἀποτελεῖ τὴν ἀποδοτικωτέραν μέθοδον. Διὰ τὴν ἐπιλογὴν δείγματος Διατρήτων Δελτίων ἐξ ἀριθμοῦ ἑρμαρίων, πλήρων ἐκ Καρτελλῶν, χρησιμοποιεῖται πολλάκις εἰς Κανὼν διὰ νὰ ἐκλέξωμεν μίαν Καρτέλλαν, π.χ. κατ' ἴντσαν.

2. **Διπλῆ Δειγματοληψία (Double or Two - Phased Sampling).**

Κατὰ ταύτην ἐκλέγομεν ἓν μέγα δεῖγμα καὶ αἱ πληροφορίαι τοῦ δείγματος τούτου χρησιμεύουσιν ὡς βάσις διὰ τὴν ἐκλογὴν ἑνὸς μικροῦ δείγματος χάριν ἐξονυχιστικωτέρας μελέτης. Παράδειγμα: Ἓν ἁπλοῦν σύντομον ἐρωτηματολόγιον χρησιμοποιεῖται διὰ νὰ ἐπιτύχωμεν δεδομένα ἐπὶ τῆς ἐπαγγελματικῆς κατανομῆς μεγάλου πληθυσμοῦ ἐκ δείγματος 10000 προσώπων καὶ ἓν ἀρκετὰ περιεκτικὸν ἐρωτηματολόγιον ἀποστέλλεται τότε εἰς ἓν δεῖγμα 1000 ἐκ τῶν 10000 αὐτῶν προσώπων. Χρησιμοποιοῦντες τὰς ἀναλογίας ἑκάστης ἐπαγγελματικῆς ὁμάδος εἰς τὸ μέγα δεῖγμα εἴμεθα εἰς θέσιν νὰ ἐκλέξωμεν τὸ δεῖγμα τῶν 1000.

3. **Ἀνεξάρτητα Ὑποδείγματα (Interpenetrating Replicate Subsamples).**

Τὰ ὑπο-δείγματα ταῦτα συνιστῶνται κατὰ τὴν μέτρησιν τοῦ βαθμοῦ τῆς συμφωνίας τῶν ἀποτελεσμάτων διαδοχικῶν δειγμάτων καὶ διὰ νὰ ἐπιτύχωμεν ἐντυπωσιακοὺς ὑπολογισμούς. Ἓν ὑπό-δειγμα δύναται νὰ χρησιμοποιηθῇ διὰ τὸν σκοπὸν τοῦτον πρὸ τοῦ ὑπολογισμοῦ τῆς τελικῆς ἐκτιμήσεως ἐκ τοῦ συνολικοῦ δείγματος. Ἡ περαιτέρω ἐνέργειά μας εἶναι νὰ ἐλέγξωμεν ἕνα ἀπαριθμητὴν ἔναντι ἑτέρου δίδοντες εἰς ἕκαστον κεχωρισμένον ὑπό-δειγμα διὰ νὰ ἐργασθῇ μ' αὐτό. Παράδειγμα : Ὑποτεθείσθω ὅτι ἔχομεν 1000 σάκκους μιᾶς λίτρας ἐκ καρύων ἠριθμημένους ἀπὸ τοῦ 1 ἕως τὸ 1000 διαδοχικῶς. Δυνάμεθα νὰ ἐκλέξωμεν 10 διαφορετικὰ δείγματα τῶν 10 σάκκων ἕκαστον κατὰ τύχην καὶ νὰ ἐκλέξωμεν 10 κάρυα ἐξ ἑκάστου σάκκου κατὰ τύχην πάλιν. Ἐὰν θέσωμεν τὰ 100 κάρυα ἀπὸ τὸ πρῶτον ὑπό-δειγμα τῶν 10 σάκκων εἰς τὸ πρῶτον τῶν 10 καλάθων, τὰ 100 κάρυα τοῦ δευτέρου ὑποδείγματος τῶν 10 σάκκων εἰς τὸ δεύτερον ἐκ τῶν 10 καλάθων κ.ο.κ. τὰ 10 καλάθια τῶν 100 καρύων ἕκαστον θέλουσιν ἀντιπροσωπεύσει δέκα ἀνεξάρτητα ὑπο-δείγματα.

4. Ἀνάλυσις Ἀκολουθίας (Sequential Analysis).

Μία σημαντικὴ μέθοδος ἀποδοχῆς τῆς δειγματοληψίας εἰς ἐργασίαν ἐλέγχου ποιότητος σημειοῦται ἐν ἁδραῖς γραμμαῖς καὶ αὐτὴ εἶναι ἡ Ἀνάλυσις Ἀκολουθίας, ἡ ὁποία ἀνεπτύχθη κατὰ τὸ ἔτος 1945. Διαφέρει τῶν ἄλλων τύπων δειγματοληψίας κατὰ τὸ ὅτι διαδοχικαὶ μονάδες ἐξετάζονται καὶ ἡ ἀπόφασις πρὸς ἀποδοχὴν ἑνὸς συνολικοῦ μεριδίου (κλήρου), ἀπορρίψεως τούτου, ἢ ἐξετάσεως τῆς ἑπομένης μονάδος λαμβάνει χώραν ἀφοῦ ἐξετασθῇ ἑκάστη διαδοχικὴ μονάς. Ἡ μέθοδος αὐτὴ ἀπεδείχθη λυσιτελὴς καὶ οἰκονομικὴ διὰ τὴν ἐπαλήθευσιν τῶν Διατρήτων Δελτίων IBM. Ἓν ὑψηλὸν ἐπίπεδον ἀκριβείας ἐπιτυγχάνεται ἀκόμη καὶ ὅταν πολὺ μικρὰ ἀναλογία Καρτελλῶν ἔχει ἐπαληθευθῆ.

5. Μικταὶ Μέθοδοι Δειγματοληψίας (Combined Sampling Methods).

Εἰς πολλὰς ἐρεύνας χρησιμοποιεῖται συνδυασμὸς μεθόδων δειγματοληψίας μᾶλλον ἢ μία μόνον μέθοδος. Ἐπὶ παραδείγματι εἰς τὴν Πολεμικὴν Ἀεροπορίαν μία δειγματολογικὴ ἔρευνα συνεπάγεται τὴν διευθέτησιν ἀριθμοῦ τινος στρωμάτων (Ἀρχηγεῖα), ἑκάστου περιλαμβάνοντος στρώματά τινα (Ὑποδιοικήσεις). Μεθ' ὃ ἐκλέγομεν Σωροὺς ἐξ ἑκάστου στρώματος μὲ ἀρίστην διάταξιν, ἐξάγομεν ἀναλογικὸν δεῖγμα τῶν Τμημάτων (Πτέρυγες) ἐξ ἑκάστου σωροῦ, ἐκλέγομεν τυχαῖον δεῖγμα ὑπομονάδων (Μοῖραι) ἐξ ἑκάστου τμήματος καὶ τέλος ἐξάγομεν συστηματικὸν δεῖγμα στοιχειωδῶν μονάδων (Ἀεροσκάφη) ἐξ ἑκάστης ὑπομονάδος.

ΚΕΦΑΛΑΙΟΝ VI. ΔΙΑΔΙΚΑΣΙΑΙ ΕΚΤΙΜΗΣΕΩΣ

Γενικά:

Ὑποτεθείσθω ὅτι ἐζητήθη νὰ ἐκτιμήσωμεν τὸ σύνολον τῶν Ὡρῶν Πτήσεως τῶν 10 Μοιρῶν τῶν Μαχίμων Πτερύγων τοῦ 28 Α.Τ.Α Πρὸς τοῦτο ἀποφασίζομεν νὰ λάβωμεν τὴν ἐκτίμησίν μας ἔκ τινος τυχαίου δείγματος τριῶν Μοιρῶν. Ὑποθετικὰ δεδομένα διὰ τοὺς μῆνας Ἰούλιον καὶ Αὔγουστον καὶ τῶν 10 Μοιρῶν καταγράφονται κατωτέρω διὰ νὰ καταδείξωμεν τὰ ἐπιτυγχανόμενα ἐκ διαφόρων διαδικασιῶν ἐκτιμήσεως ἀποτελέσματα.

KATANOMH ΩΡΩΝ ΠΤΗΣΕΩΣ ΜΑΧΙΜΩΝ ΜΟΙΡΩΝ ΕΒΑ (ΠΛΗΘΥΣΜΟΣ)

ΑΡΙΘΜΟΣ ΜΟΙΡΩΝ	ΩΡΑΙ ΠΤΗΣΕΩΣ	
	ΙΟΥΛΙΟΥ (Ψ)	ΑΥΓΟΥΣΤΟΥ (Χ)
ΜΟΙΡΑ Α′	1179	1227
» Β′	714	885
» Γ′	1076	1145
» Δ′	763	843
» Ε′	874	913
» ΣΤ′	1191	1284
» Ζ′	483	651
» Η′	745	890
» Θ′	1272	1370
» Ι′	703	792
Σύνολον Ὡρῶν Πτήσεως Μονάδων 28 ΑΤΑ	9000	10000
Μέσος Ὅρος Ὡρῶν	900	1000

ΚΑΤΑΝΟΜΗ ΩΡΩΝ ΠΤΗΣΕΩΣ ΤΡΙΩΝ ΜΟΙΡΩΝ
(ΑΠΛΟΥΝ ΤΥΧΑΙΟΝ ΔΕΙΓΜΑ)

ΕΠΙΛΕΓΕΙΣΑΙ ΜΟΙΡΑΙ	ΩΡΑΙ ΠΤΗΣΕΩΣ	
	ΙΟΥΛΙΟΥ (Ψ)	ΑΥΓΟΥΣΤΟΥ (Χ)
ΜΟΙΡΑ Β′	714	885
» Ε′	874	913
» Η′	745	890
Σύνολον Δείγματος	2.333	2.688
Μέσος Ὅρος Ὡρῶν Πτήσεως Δείγματος	778	896

Ἁπλῆ Ἀμερόληπτος Ἐκτίμησις (Blowup or Simple Unbiased Estimate).

Ἡ πλέον προφανὴς ἐκτίμησις τοῦ συνολικοῦ χρόνου πτήσεως μ. Αὐγούστου εἶναι ἡ Ἁπλῆ Ἀμερόληπτος Ἐκτίμησις, ἡ ὁποία ἐπιτυγχάνεται διὰ τῆς ἐφαρμογῆς τοῦ ἀντιστρόφου τῆς ἀναλογίας δείγματος πρὸς τὸ σύνολον δείγματος μ. Αὐγούστου. Ἐφ' ὅσον ἐπελέγησαν τρεῖς ἐξ ὅλων τῶν Μοιρῶν διὰ τὸ δεῖγμα μας ἡ προκειμένη ἐκτίμησις θὰ εἶναι 2688 × 10 : 3 ἴσον 8960 ὧραι ἔναντι τοῦ συνολικοῦ χρόνου πτήσεως τοῦ μηνὸς Αὐγούστου ὅλων τῶν Μοιρῶν, ὅστις εἶναι 10000 ὧραι. Ὡς πρὸς τὸν Ἰούλιον θὰ ἔχωμεν 2333 × 10 : 3 = 7777.

Ἀναλογικὴ Ἐκτίμησις (Ratio Estimate).

Ἡ μέθοδος αὕτη ἀποφέρει περισσότερον ἀκριβῆ ἀποτελέσματα ἢ ἡ προηγουμένη. Ἐὰν ἔχωμεν διαθέσιμα δεδομένα ὡρῶν πτήσεως ὅλων τῶν Μοιρῶν τοῦ προηγουμένου μηνὸς ἡ ἀναλογικὴ ἐκτίμησις χρησιμοποιεῖται ὡς ἑξῆς :

$$\begin{array}{c}\text{Ἀναλογικὴ Ἐκτίμησις}\\ \text{Συνολικοῦ χρόνου πτή-}\\ \text{σεως μηνὸς Αὐγούστου}\end{array} = \frac{\begin{array}{c}\text{Χρόνος Πτήσεως Μοι-}\\ \text{ρῶν Δείγματος μηνὸς}\\ \text{Αὐγούστου}\end{array}}{\begin{array}{c}\text{Χρόνος Πτήσεως τῶν}\\ \text{Μοιρῶν Δείγματος μη-}\\ \text{νὸς Ἰουλίου}\end{array}} \times \begin{array}{c}\text{Χρόνον Πτήσεως ὅλων}\\ \text{τῶν Μοιρῶν μηνὸς}\\ \text{Ἰουλίου}\end{array}$$

$$\text{ἢ } \frac{2688}{2333} \times 9000 = 10.369 \text{ Ὧραι.}$$

Ἐκτίμησις Διαφορᾶς (Difference Estimate).

$$\begin{array}{c}\text{Ἐκτίμησις Δια-}\\ \text{φορᾶς Συνόλου}\\ \text{Ὡρῶν Πτήσεως}\\ \text{μ. Αὐγούστου}\end{array} = \begin{array}{c}\text{Ἀμερόληπτος Ἐκ-}\\ \text{τίμησις Συνόλου}\\ \text{Ὡρῶν Πτήσεως}\\ \text{μ. Αὐγούστου}\end{array} + \left(\begin{array}{c}\text{Ὧραι Πτήσεως}\\ \text{ὅλων τῶν Μοι-}\\ \text{ρῶν μ. Ἰουλίου}\end{array} - \begin{array}{c}\text{Ἀμερόληπτος}\\ \text{Ἐκτίμησις Συνό-}\\ \text{λου Ὡρῶν Πτή-}\\ \text{σεως μ. Ἰουλίου}\end{array} \right)$$

$$\text{ἢ } 8960 + (9000 - 7777) = 10.183 \text{ Ὧραι}$$

Ἐκτίμησις Γραμμικῆς Παλινδρομήσεως (Linear Regression Estimate).

Ἡ ἐκτίμησις γραμμικῆς παλινδρομήσεως εἶναι παρομοία τῆς προηγουμένης, μὲ τὴν διαφορὰν ὅτι ὁ ἐν τῇ παρενθέσει ὅρος πολλαπλασιάζεται ἐπὶ β, ὅστις εἶναι τὸ κοινὸν εἶδος τοῦ ἀπαντωμένου ἐν τῇ Ἁπλῇ Γραμμικῇ Συσχετίσει Συντελεστοῦ Συσχετίσεως (ἐν τῇ ἐξισώσει πάλινδρομήσεως $\psi = \alpha + \beta\chi$). Οὕτως ἡ ἐξίσωσις τῆς Ἐκτιμήσεως Γραμμικῆς Παλινδρομήσεως εἶναι $X'' = X' + \beta(\Psi - \Psi')$, ἔνθα :

X'' εἶναι ἡ ἐκτίμησις γραμμικῆς παλινδρομήσεως μηνὸς Αὐγούστου
X' εἶναι ἡ Ἀμερόληπτος Ἐκτίμησις τοῦ μηνὸς Αὐγούστου
Ψ εἶναι τὸ πραγματικὸν σύνολον τοῦ μηνὸς Ἰουλίου
Ψ' εἶναι ἡ Ἀμερόληπτος Ἐκτίμησις τοῦ μηνὸς Ἰουλίου

Εἰς τὸ παράδειγμά μας $X'' = 8960 + 0{,}8959(9000 - 7777) = 10.056$ Ὧραι. Διὰ τὸν ἐπιθυμοῦντα νὰ ἐλέγξῃ τὴν ἀπάντησιν αὐτὴν ὑπενθυμίζομεν ὅτι

$$\beta = \frac{\Sigma(X_\iota - \bar{X}) - (\Psi_\iota - \bar{\Psi})}{\Sigma(\Psi_\iota - \bar{\Psi})^2} \quad \text{ἔνθα}$$

X_i εἶναι ὁ χρόνος πτήσεως τῆς Μοίρας i μηνὸς Αὐγούστου
Ψ_i » » » i » Ἰουλίου
$\overline{X}$ εἶναι ὁ μέσος χρόνος πτήσεως ὅλων τῶν Μοιρῶν μ. Αὐγούστου
$\overline{\Psi}$ εἶναι ὁ μέσος χρόνος πτήσεως ὅλων τῶν Μοιρῶν μ. Ἰουλίου
Σ εἶναι τὸ σύμβολον τῆς ἀθροίσεως τῶν γινομένων ἢ τετραγώνων καὶ τῶν 10 Μοιρῶν.

Κανονικῶς τὸ β θὰ ἠδύνατο νὰ ὑπολογισθῇ ἐκ τῶν δεδομένων τοῦ δείγματος διὰ τοὺς δύο διαδοχικοὺς μῆνας, δοθέντος ὅτι τὰ δεδομένα ὅλων τῶν Μοιρῶν δὲν θὰ ἦσαν διαθέσιμα διὰ τὸν τελευταῖον μῆνα.

Σύνοψις:

Ὡς ἴδωμεν εἰς τὰς προηγουμένας παραγράφους εἶναι ἀναγκαῖον ὄχι μόνον νὰ ἐπιδιώκωμεν τὴν πλέον ἀποτελεσματικὴν μέθοδον δειγματοληψίας ἀλλὰ νὰ ἐκλέγωμεν καὶ μίαν κατάλληλον διαδικασίαν ἐκτιμήσεως εἰς συγκεκριμμένον πρόβλημα δειγματοληψίας.

Ἐπανερχόμενοι καὶ πάλιν ἐπὶ τοῦ παραδείγματός μας βλέπομεν ὅτι προέκυψαν τὰ ἀκόλουθα ἀποτελέσματα :

	ΩΡΑΙ ΠΤΗΣΕΩΣ	% ΣΦΑΛΜΑΤΟΣ
Πραγματικὸν Σύνολον Ὡρῶν Αὐγούστου	10.000	—
Ἀμερόληπτος Ἐκτίμησις	8.960	10.40 %
Ἀναλογικὴ Ἐκτίμησις	10.369	3.69 %
Ἐκτίμησις Διαφορᾶς	10.183	1.83 %
Ἐκτίμησις Γραμ. Παλινδρομήσεως	10.056	0.56 %

Ὅταν τὰ δεδομένα διαδοχικῶν περιόδων ἢ κεχωρισμένων μονάδων συσχετίζωνται ἡ Ἀναλογικὴ Ἐκτίμησις πολλάκις ὑπερτερεῖ τῆς Ἀμερολήπτου τοιαύτης. Ὁμοίως αἱ Ἐκτιμήσεις Διαφορᾶς καὶ Γραμμικῆς Παλινδρομήσεως συνήθως συνεπάγονται καλύτερα ἀποτελέσματα ὑπὸ τοιαύτας περιστάσεις, καίτοι οἱ ὑπολογισμοὶ ἰδίᾳ τῆς Ἐκτιμήσεως Παλινδρομήσεως παρουσιάζονται πλέον ἐπίπονοι.

ΚΕΦΑΛΑΙΟΝ VII. ΣΦΑΛΜΑΤΑ ΕΚ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΙ ΜΗ

Γενικά:

Εἰς τοὺς ἐκ τῶν ἀναγνωστῶν μεμψιμοίρους διὰ τὴν ἀποτελεσματικότητα τῆς δειγματοληψίας ὡς μέσου περιγραφικοῦ μιᾶς ὁλότητος δεδομένων ἐξ ἴσου ἂν ὄχι καλλίτερον συνιστωμένου ἔναντι τῆς πλήρους ἀπαριθμήσεως τοῦ Συνόλου κρίνεται σκόπιμον νὰ ὑπομνησθῶσι τὰ ἀκόλουθα :

α) Ὁσάκις ἡ δειγματοληψία ἐνεργῆται οὕτως ὥστε ἑκάστη μονὰς τοῦ στατιστικοῦ πληθυσμοῦ νὰ ἔχῃ τὴν αὐτὴν τύχην νὰ ἐπιλεγῇ καὶ ἡ πιθανότης τῆς ἐπιλογῆς εἶναι γνωστή, τότε τὰ σφάλματα τῆς δειγματοληψίας δύνανται νὰ ἐλεγχθῶσιν ἱκανοποιητικῶς καὶ ἡ τοιαύτη δειγματοληψία ὁρολογεῖται ὡς Δειγματοληψία Πιθανότητος (Probability Sampling).

β) Εἶναι ἐξ ἴσου ἀληθὲς ὅτι ὑφίσταται ἀριθμός τις μεθόδων δειγματοληψίας οὐχὶ ἐκ πιθανότητος, ὅπου τὸ σφάλμα τῆς δειγματοληψίας δὲν δύναται νὰ ἐλεγχθῇ

Παρὰ ταῦτα ἱκανοὶ δειγματολῆπται ἀποφεύγουσι τὸν τύπον αὐτὸν δειγματοληψιῶν ἐξαιρέσει εἰδικῶν περιπτώσεων καὶ μόνον ὁσάκις ἡ χρησιμοποίησις δεδομένων μὲ ἄγνωστα σφάλματα καὶ μεροληψίας κατὰ τὴν δειγματοληψίαν δὲν εἶναι σημαντική.

γ) Οἱονδήποτε σύστημα συλλογῆς καὶ προωθήσεως δεδομένων ὑπόκειται εἰς σωρείαν σφαλμάτων, τὰ δὲ μὴ δειγματολογικὰ σφάλματα εἰς τοιοῦτον σύστημα πολλάκις εἶναι μεγαλυτέρα ἢ τὰ ἐκ δειγματοληψίας τοιαῦτα. Προσέτι ὁ βαθμὸς σφάλματος μιᾶς ἐκτιμήσεως ἑνὸς δείγματος δύναται νὰ μετρηθῇ ἐπὶ δειγμάτων πιθανότητος διὰ τῶν ὑπαρχόντων μέσων τῆς Στατιστικῆς Μεθοδολογίας, ἐνῷ τὰ μὴ ἐκ δειγματοληψίας σφάλματα γενικῶς δὲν εἶναι μετρήσιμα (ἐκτὸς ἐὰν χρησιμοποιηθῶσιν ἔλεγχοι δείγματος). Ἀλλὰ καὶ τὰ ἐκ δειγματοληψίας σφάλματα εἶναι δυνατὸν νὰ ἀποφευχθῶσιν ἐὰν προβῶμεν εἰς πλήρη ἀπαρίθμησιν. Ὁπωσδήποτε μία καθολικὴ ἀκρίβεια συχνάκις δύναται νὰ ἐπαυξηθῇ ἐφ' ὅσον ἐνεργήσωμεν μίαν δειγματολογικὴν ἔρευναν παρὰ νὰ κάμωμεν πλήρη ἀπαρίθμησιν. Τὸ μικρότερον πεδίον τοῦ σχεδιασθέντος δείγματος ἐπιτρέπει εἰς τοῦτο νὰ εἶναι περισσότερον ἐκλεκτικόν, πλέον ὀρθὸν καὶ ἱκανὸν ἂν συγκεντροῦται εἰς μεγαλύτερον βαθμὸν πρὸς μείωσιν τῶν μὴ ἐκ δειγματοληψίας σφαλμάτων. Τὸ καθαρὸν προϊὸν μιᾶς δειγματολογικῆς ἐρεύνης ἐν συγκρίσει πρὸς τὴν 100% ἀπαρίθμησιν πολλάκις παρέχει ἀκριβεστέραν ἀπάντησιν, ἐπιτυγχανόμενον μὲ ὀλιγώτερον προσωπικόν, ὀλιγωτέραν γραφικὴν ἐργασίαν, ἠλαττωμένον κόστος καὶ τέλος εἰς βραχύτερον χρονικὸν διάστημα.

Κατωτέρω ἐπισημαίνομεν τοὺς πλέον σοβαροὺς τύπους σφαλμάτων, τοὺς ὁποίους δέον νὰ ἔχωμεν ὑπ' ὄψιν μας κατὰ τὴν συλλογὴν τῶν δεδομένων.

Σφάλματα Ἀναφορῶν καὶ Προωθήσεως

Ἓν ἀρκετὰ ἐν χρήσει στατιστικὸν παράδοξον εἶναι τὸ γεγονὸς ὅτι ἡ μέση ἡλικία τῶν ἄνω τῶν 40 ἐτῶν γυναικῶν ἐμφανίζεται πολὺ κάτω τῶν 40 ὁσάκις ἀναφέρεται παρὰ τῶν Κυριῶν. Παρομοίως εἷς καλλιεργητὴς μπανανῶν δύναται ν' ἀναφέρῃ τὸ ἐπάγγελμά του ὡς κιτροκαλλιεργητοῦ καὶ ὁ συρράπτων τὰς ὀπὰς τῶν κομβιοδόχων εἰς τὰ πέτα τῶν ζακετῶν ὡς ράπτου. Ἄλλος πάλιν δύναται νὰ ἀναφέρῃ διαφόρους ἀριθμοὺς δι' ἓν καὶ τὸ αὐτὸ πρᾶγμα, π.χ. τὴν ἡλικίαν του ὡς αὕτη γράφεται ἐπὶ μιᾶς αἰτήσεώς του δι' Ἀσφάλειαν Ζωῆς καὶ ὅπως ἀναφέρεται καὶ εἰς τὸ Ἀσφαλιστήριον Συμβόλαιον.

Βασικὰ δεδομένα δύνανται νὰ καταγραφῶσιν ἀνακριβῶς, νὰ ἀντιγραφῶσι ἐσφαλμένως, νὰ κωδικοποιηθῶσι μὴ ὀρθῶς, νὰ χειρισθῶσι μὴ καταλλήλως ἢ νὰ παραλειφθῶσι τελείως. Οὕτω ἐκ μιᾶς ἐρεύνης διαταχθείσης διὰ τὴν συλλογὴν τοῦ χρόνου πτήσεως καὶ τῶν χορηγηθέντων καυσίμων βάσει ἀρχικῶν πηγῶν προέκυψεν ὅτι ἀεροσκάφη τινὰ ἐπέταξαν 20, 30 ἢ περισσοτέρας ὥρας χωρὶς νὰ καταναλώσωσιν οὐδὲ σταγόνα καυσίμων καὶ τὸ κατόρθωμα αὐτὸ δὲν ἐπετεύχθη διὰ τῆς χρησιμοποιήσεως τυχὸν νέου βελτιωμένου τύπου ἀναμικτῆρος ἀλλ' ἁπλῶς διὰ τῆς παραλείψεως τῆς ἐγγραφῆς τῆς χορηγήσεως τῶν καυσίμων εἰς τὸ οἰκεῖον ἔντυπον.

Σφάλματα μὴ Ἀποκρίσεως

Πολλαὶ σοβαραὶ μεροληψίαι ἠμποροῦν νὰ προέλθωσι διὰ τῆς χρήσεως τῶν ἀναφερθέντων δεδομένων ὑπὸ τῶν ἀποκρινομένων εἰς ἕνα ἐρωτηματολόγιον ἢ τῶν ἐνοίκων κατὰ μίαν οἰκογενειακὴν ἔρευναν, ἀδιαφόρως ἐὰν ἀνάγωνται εἰς ἓν δεῖγμα ἢ εἰς τὸν συνολικὸν πληθυσμόν. Μία ἡμερησία ἀπὸ οἰκίας εἰς οἰκίαν ἔρευνα, διεξαχθεῖσα ἐντὸς μιᾶς ἑβδομάδος θὰ ἀπέφερεν ἀναμφιβόλως πολὺ πτωχὴν ἐκτίμησιν τῶν οἰκοκυρῶν, ἐνῷ μία ἀπογευματινὴ συνέχισις διὰ τὰς μὴ εὑρισκομένας οἴκοι θὰ ἦτο ἀναγκαία διὰ νὰ ὑπερνικήσωμεν τὴν προφανῆ αὐτὴν μεροληψίαν. Εἶναι πάντοτε ἐπικίνδυνον νὰ παραδεχθῶμεν ὅτι οἱ μὴ ἀποκρινόμενοι ἢ οἱ ἀπουσιάζοντες κατέχουσι τὰ αὐτὰ χαρακτηριστικὰ καὶ εἰς τὸν αὐτὸν βαθμὸν ὅπως οἱ ἀποκρινόμενοι καὶ οἱ κατ' οἶκον, χωρὶς νὰ ἐλέγξωμεν τὸ δεῖγμα διὰ νὰ ἐξασφαλίσωμεν τὸ κῦρος τῆς ὑποθέσεως αὐτῆς.

Σφάλματα κατὰ τὴν ἐκλογὴν Δείγματος

Μία ἐπίσης λίαν κοινὴ πηγὴ σφάλματος εἶναι ἡ μὴ πρέπουσα καὶ ἀνεπιστημονικὴ ἐκλογὴ τῶν μονάδων τοῦ δείγματος ὡς ἐπὶ παραδείγματι τὰ ἐπιλεγόμενα χάριν εὐκολίας χονδροειδῆ δείγματα. Ὁ Δειγματολήπτης ὁ ὁποῖος καταλήγει εἰς τὸ Μπὰρ τῶν Ζόες καὶ Γρὶλλ ὡς τὸ μέρος ὅπου πρόκειται νὰ ἐρευνήσῃ τὰ τῆς ποτοαπαγορεύσεως, ἀσφαλῶς θὰ εἶναι ὁ δράστης ἐλαφρᾶς μεροληπτικῆς, χονδροειδοῦς δειγματοληψίας. Δείγματα Κρίσεως ἂν καὶ πολλάκις ἀποφέροντα καλὰ ἀποτελέσματα εἶναι οὐχ ἧττον ἐπικίνδυνα, διὰ τὸν ἁπλούστατον λόγον ὅτι ἀκόμη καὶ ἡ κρίσις τῶν εἰδημόνων ποικίλλει σημαντικῶς κατὰ τὴν ἐκλογὴν τῶν ἀντιπροσωπευτικῶν δειγμάτων. Ὁμοίως Ἀναλογικὰ Δείγματα (Quota Samples) πάσχουσιν ἀπὸ τὸ ἐλάττωμα ὅτι ἐξαρτῶνται περισσότερον ἐκ τῆς ὑποκειμενικῆς διαθέσεως τοῦ Δειγματολήπτου ἢ ἐπὶ τῆς τυχαίας δειγματοληψίας. Ἓν τοιοῦτον δεῖγμα π.χ. δύναται νὰ προσδιορίσῃ εἰδικῶς πῶς ὁ Δειγματολήπτης θὰ ἐπιτύχῃ δεδομένα ἐξ ἀρρένων ἡλικίας 35 - 40 ἐτῶν μὲ εἰσόδημα δρχ. 5000 - 7000. Καίτοι τοῦτο περιορίζει τοῦτον ἔν τινι μέτρῳ ἀπὸ τοῦ νὰ κάμῃ μίαν ἐξ ὁλοκλήρου μὴ ἀντιπροσωπευτικὴν ἐκλογὴν δύναται μολαταῦτα νὰ ἐκλέξῃ ὥστε νὰ ἐνεργήσῃ τὰς συνεντεύξεις του κατὰ τὴν διάρκειαν Ἀθλητικῶν Ἀγώνων εἰς τὸ Στάδιον.

Σφάλματα Δειγματοληψίας

Τὸ ἀποτέλεσμα μιᾶς δειγματολογικῆς ἐρεύνης σχεδὸν πάντοτε δὲν συμφωνεῖ τελείως πρὸς τὸ ἐπιτυγχανόμενον τοιοῦτον ἐκ μιᾶς πλήρους ἀπαριθμήσεως κατὰ τὸν αὐτὸν τρόπον. Ἡ μεταξὺ αὐτῶν διαφορὰ ὀφείλεται εἰς τὸ σφάλμα δειγματοληψίας. Ἐὰν χρησιμοποιηθῇ ἓν σχεδιασθὲν δεῖγμα πιθανότητος εἶναι δυνατὸν νὰ προκαθορίσωμεν τὸ μέγεθος τοῦ ἀπαιτουμένου δείγματος διὰ νὰ ἔχωμεν εἰδικῶς προσδιωρισμένον βαθμὸν ἀκριβείας. Αὐτὴ ἡ σπουδαιοτάτη καὶ λίαν εὔστοχος ἰδιότης τοῦ δείγματος πιθανότητος συνήθως καθιστᾶ αὐτὸ δυνατὸν ὥστε τὸ ἀναμενόμενον σφάλμα δείγματος νὰ κρατῆται ἐντὸς τῶν ἐπιθυμητῶν ὁρίων.

ΚΕΦΑΛΑΙΟΝ VIII.

ΚΡΙΤΗΡΙΑ ΣΗΜΑΝΤΙΚΟΤΗΤΟΣ ΚΑΙ ΑΞΙΟΠΙΣΤΙΑΣ ΔΕΙΓΜΑΤΩΝ

(Κανονικὰ Σφάλματα)

Ἐκ τῶν ἐκτεθέντων εἰς τὰ προηγούμενα Κεφάλαια προκύπτει ὅτι, ἐπειδὴ τὰ στοιχεῖα ἑνὸς πληθυσμοῦ διαφέρουσι μεταξύ των, ἡ δὲ τύχη ἢ ἀνεξέλεγκτα αἴτια ὑπεισῆλθον κατὰ τὸν καταρτισμὸν ἑνὸς δείγματος, τὸ ἐκ τοῦ τελευταίου τούτου προκύψαν στατιστικὸν ἀποτέλεσμα διαφέρει βεβαίως ἐκείνου ὅπερ θὰ ἐπετυγχάνετο ἐκ τοῦ συνολικοῦ πληθυσμοῦ. Τὸ ἀποτέλεσμα ὅθεν αὐτὸ δὲν δύναται νὰ ληφθῇ εἰμὴ ὡς ἐκτίμησις τοῦ ἀληθοῦς ἀποτελέσματος. Ἐκ παραλλήλου ὅμως μᾶς χρειάζεται νὰ γνωρίζωμεν κατὰ ποῖον μέτρον δυνάμεθα νὰ τὸ θεωρήσωμεν ὡς μίαν καλὴν ἐκτίμησιν τῆς ζητουμένης τιμῆς διὰ νὰ μὴ ἀποδώσωμεν πραγματικὴν σημασίαν εἰς γεγονότα κατὰ τὸ πλεῖστον τυχαίας μορφῆς ἢ νὰ ἑρμηνεύσωμεν ἓν γεγονὸς σημασίας ὡς ἁπλῆν σύμπτωσιν.

Εἰς τὸ κεφάλαιον λοιπὸν αὐτὸ θὰ ἀσχοληθῶμεν μὲ ὡρισμένας μεθόδους, αἱ ὁποῖαι θὰ μᾶς ἐπιτρέψωσι νὰ ἀναλύσωμεν τὰ ἐκ τῶν παρατηρήσεων ἐξαγόμενα στατιστικὰ ἀποτελέσματα. Πρὸς τοῦτο ὅμως θὰ προτάξωμεν στοιχεῖα τινὰ τῆς Θεωρίας τῶν Σφαλμάτων τῆς Δειγματοληψίας καὶ τὴν Λογικήν, ἐφ' ἧς αὕτη ἐρείδεται.

Ἔστω ἐπὶ παραδείγματι ὅτι ἐπιθυμοῦμεν νὰ μελετήσωμεν χαρακτηριστικόν τι καὶ δὴ τὸ μέσον ἀνάστημα ἐνηλίκων, ἡλικίας 18 ἐτῶν. Δοθέντος ὅτι εἶναι ἀνέφικτον νὰ διέλθωσιν ἐκ τοῦ ἀναστημομέτρου ὅλοι οἱ ἄγοντες τὴν ἡλικίαν τῶν 18 ἐτῶν λαμβάνομεν κατὰ τύχην ἕνα ἀρκετὰ μεγάλον ἀριθμὸν ἐξ αὐτῶν διὰ νὰ ἀπαρτίσωμεν τὸ δεῖγμα μας. Ὑπολογίζομεν τότε τὸν Μέσον τῶν ἀναστημάτων τοῦ δείγματος αὐτοῦ, ὁ ὁποῖος συνιστᾷ μίαν ἐκτίμησιν τοῦ ζητουμένου μέσου ἀναστήματος. Ἐὰν ἡ ἐμπιστοσύνη μας ἐπὶ τοῦ ἀριθμοῦ αὐτοῦ εἶναι κατὰ πολὺ μεγαλυτέρα παρ' ὅσον εἶναι σημαντικὸς ὁ ἀριθμὸς τοῦ δείγματος, τὸ γεγονὸς τοῦτο μετριάζεται καθ' ὅσον δὲν γνωρίζομεν ἐὰν τὸ θεωρούμενον δεῖγμα εἶναι ὄντως ἀντιπροσωπευτικὸν τοῦ ὑπ' ὄψιν πληθυσμοῦ. Ἡ τύχη θὰ συνέτεινεν ὥστε ἡ ἀναλογία τῶν Νέων ὑψηλοῦ ἀναστήματος ἢ μικροῦ τοιούτου νὰ εἶναι μὴ ὁμαλή, ὁπότε ὁ παρατηρηθεὶς Μέσος θὰ διαφέρῃ προφανῶς ἀρκετὰ ἐκ τοῦ ἀληθοῦς Μέσου. Ἐκεῖνο τὸ ὁποῖον ἐνδιαφέρει ἤδη νὰ μάθωμεν εἶναι ἡ μεγίστη ἀπόκλισις, ἥτις σημειοῦται μεταξὺ τοῦ παρατηρηθέντος Μέσου καὶ τοῦ πραγματικοῦ Μέσου ἤ, ἐν ἄλλοις λόγοις, μὲ ποίαν ἐμπιστοσύνην δυνάμεθα νὰ πιστοποιήσωμεν ὅτι ὁ Μέσος τοῦ δείγματος εὑρίσκεται εἰς δοθεῖσαν ἀπόστασιν ἀπὸ τοῦ Μέσου τοῦ πληθυσμοῦ. Παρατηροῦμεν λοιπὸν ὅτι ἐὰν ἐκλέξωμεν ἀνεξαρτήτως τοῦ πρώτου καὶ ἓν δεύτερον δεῖγμα, ἔχον τὸν αὐτὸν ἀριθμὸν μονάδων, ἡ κατανομὴ τῶν ἀναστημάτων τοῦ δείγματος τούτου θὰ διαφέρῃ πιθανῶς ἀπὸ τοῦ πρώτου. Ὁ Μέσος τῶν ἀναστημάτων τοῦ δευτέρου δείγματος θὰ διαφέρῃ σχεδὸν βεβαίως ἀπὸ τοῦ Μέσου τοῦ πρώτου δείγματος, πρᾶγμα ὅπερ ἀποδεικνύει ὅτι οὐδὲν ἐκ τῶν δύο ἀποτελεσμάτων δύναται νὰ γίνῃ δεκτὸν διὰ τὴν ἀκριβῆ ἀπεικόνισιν τοῦ συνολικοῦ πληθυσμοῦ. Ἓν τρίτον δεῖγμα ἐξαγόμενον κατὰ τὸν αὐτὸν τρόπον ἀποφέρει ἕνα νέον Μέσον τῶν ἀναστημάτων καὶ ἐν συνεχείᾳ, ἐὰν ἐξακολουθήσωμεν νὰ ἐξάγωμεν δείγματα ἐξ ἴσου πολυπληθῆ ἐκ τοῦ συνόλου πληθυσμοῦ, θὰ σχηματίσωμεν ἕνα σεβαστὸν ἀριθμὸν ἀποτελεσμάτων, ὁπότε οἱ ἐπιτυγχανόμενοι Μέσοι θὰ κατανέμωνται ἐπὶ εὐρυτάτης κλίμακος, τοῦ πλείστου ὅμως ἐξ αὐτῶν συγκεντρουμένου πέριξ μιᾶς κεντρικῆς τιμῆς, ἑτέρων ὀλιγαρίθμων βεβαίως τιμῶν ἐμφανιζουσῶν ἀπόκλισιν ἀπὸ τῆς Μέσης Τιμῆς, λίαν δὲ σημαντικήν. Ἡ μνημονευθεῖσα Σειρὰ τῶν Μέσων μεγάλου ἀριθμοῦ δειγμάτων τῶν αὐτῶν μονάδων καλεῖται Κατανομὴ τοῦ Μέσου (Mean Distribution) τῆς οἰκογενείας τῶν δειγμάτων καὶ ἡ Κατανομὴ αὐτὴ ἱκανοποιεῖ τὸν Κανονικὸν Νόμον.

Ἡ ἐμπιστοσύνη, τὴν ὁποίαν δυνάμεθα νὰ ἔχωμεν ἐπί τινος Μέσου ἑνὸς μοναδικοῦ δείγματος, ἐξαρτᾶται συνεπῶς ἐκ τῆς διασπορᾶς τῆς Κατανομῆς αὐτῆς τοῦ Μέσου. Πράγματι ἐὰν οἱ Μέσοι τῆς οἰκογενείας τῶν δειγμάτων ἐξαπλοῦνται ἐπὶ μεγάλου διαστήματος δὲν ἠμποροῦμεν καθόλου νὰ ἔχωμεν ἐμπιστοσύνην ἐπὶ τοῦ ληφθέντος ἐξ ἑνὸς μόνον δείγματος Μέσου, διότι εἶναι δυνατὸν οὗτος νὰ συνιστᾷ μίαν ἐκ τῶν τιμῶν αἵτινες ἐμφανίζουσιν ἰσχυρὰν ἀπόκλισιν. Ἐὰν ἀντιθέτως ἡ διασπορὰ τῶν Μέσων εἶναι πτωχὴ ἡ ληφθεῖσα τιμὴ δὲν θὰ διαφέρῃ ἐκ τοῦ ἀληθοῦς Μέσου εἰμὴ κατ' ἐλάχιστον, ὁπότε θὰ εἶναι δυνατὸν νὰ ἔχωμεν βασίμως ὑψηλὴν ἐμπιστοσύνην ἐπὶ τῆς τιμῆς του. Διὰ νὰ προσδιορίσωμεν ὅθεν τὸν βαθμὸν τῆς ἐπιτευχθείσης ἐμπιστοσύνης θὰ πρέπῃ νὰ μετρήσωμεν τὴν διασπορὰν τῆς Κατανομῆς τοῦ Μέσου καὶ ἡ μέτρησις αὐτὴ παρέχεται διὰ τῆς Σταθερᾶς Ἀποκλίσεώς της σ, ἡ ὁποία καὶ ὁρολογεῖται ἐν προκειμένῳ ὡς Κανονικὸν Σφάλμα[3].

Συνοπτικῶς ἡ ἔκφρασις ὅτι δὲν δυνάμεθα νὰ ἔχωμεν ἐμπιστοσύνην εἰς τὸν Μέσον ἑνὸς δείγματος σημαίνει ὅτι οἱ Μέσοι τῶν δειγμάτων τῶν αὐτῶν μονάδων ἔχουσι μεγάλην διασπορὰν ἢ ὅτι ὁ Μέσος ἔχει ἓν ἰσχυρὸν Κανονικὸν Σφάλμα (Erreur-type). Παρίσταται λοιπὸν ἐπιτακτικὴ ἡ ἀνάγκη ὅπως ἐν συνεχείᾳ τῆς μελέτης μας ὁρίσωμεν ποσοτικῶς τὸν βαθμὸν τῆς ἐμπιστοσύνης, τὸν ὁποῖον εἴμεθα εἰς θέσιν νὰ ἔχωμεν εἰς ἀποτελέσματα ἐξαχθέντα ἐκ γενονότων παρατηρηθέντων ἐκ μοναδικοῦ τινος δείγματος, πρᾶγμα ὅπερ συνεπάγεται τὸν ἐννοιολογικὸν προσδιορισμὸν τοῦ Ἐπιπέδου Σημαντικότητος (Seuil de signification). Ὁ βαθμὸς τῆς ἐμπιστοσύνης τὸν ὁποῖον δύναταί τις νὰ προσδώσῃ εἰς δοθεῖσαν διαπίστωσιν μόνον μέσῳ τῶν Πιθανοτήτων εἶναι δυνατός. Ὑποτεθείσθω ἐπὶ παραδείγματι ὅτι σάκκος τις περιέχει 100 σφαίρας, ἐξ ὧν 95 φέρουσι μίαν μάρκαν. Διαπιστοῦμεν μὲ πολὺ μεγάλην ἐμπιστοσύνην ὅτι ἐὰν ἐξαγάγωμεν κατὰ τύχην μίαν σφαίραν ἐκ τοῦ σάκκου, αὕτη θὰ εἶναι μαρκαρισμένη καὶ ὁ βαθμὸς τῆς ἐμπιστοσύνης αὐτῆς καλεῖται ἐπίπεδον σημαντικότητος 5%. Τὸ κριτήριον τοῦτο δικαιολογεῖται ἐκ τοῦ γε-

---

3. Κατὰ τὴν Θεωρίαν τῶν Πιθανοτήτων τὸ μέσον Σφάλμα εἶναι ἐπὶ ἀπολύτων μὲν ἀριθμῶν $\sigma = \sqrt{n.\ p.\ q.}$, ἐπὶ ἀναλογιῶν δὲ $\sigma = \sqrt{\frac{p.\ q.}{n}}$. Κατὰ τὸν Bernoulli τὸ ἁπλοῦν Μέσον Σφάλμα δηλοῖ ὅτι εἰς $^2/_3$ τῶν παρατηρήσεων ὁ ἀριθμὸς τῶν συμβάντων εἶναι ἴσος πρὸς τὴν εὑρεθεῖσαν ἀναλογίαν μὲ τὴν προσθήκην ἢ ἀφαίρεσιν τοῦ σ, ἐφ' ὅσον δὲν ἐπηρεάζεται ὁ ἀριθμὸς τούτων ἐξ ἄλλων αἰτιῶν. Περαιτέρω διὰ νὰ περιορισθῇ ἡ ὅλη ἔκτασις τῶν ἐκ τυχαίων ἐπιδράσεων προκαλουμένων δυνατοτήτων διακυμάνσεως λαμβάνομεν τὸ τριπλάσιον τοῦ Μέσου Σφάλματος.

γονότος ὅτι ἐὰν ἐπαναλάβωμεν πολλὰς φορὰς τὴν ἐν λόγῳ πρᾶξιν (ἀφ' οὗ βεβαίως ἀντικαταστήσωμεν κάθε φορὰν τὴν ἐξαγομένην ἐκ τοῦ σάκκου σφαῖραν) ἡ διαπίστωσίς μας θὰ εἶναι ἐσφαλμένη μόνον 5 φορὰς ἐπὶ τῶν 100 ἤ, καλλίτερον, ἡ πιθανότης τοῦ σφάλματός μας θὰ εἶναι 0.05. Ἀλλὰ καὶ ἕτερα ἐπίπεδα σημαντικότητος δύνανται νὰ ὁρισθῶσιν, ὅπως λ.χ. εἰς τὴν περίπτωσιν ἐξαγωγῆς ἑνὸς παιγνιοχάρτου ἐκ δέσμης τριάκοντα δύο τοιούτων δυνάμεθα νὰ βεβαιώσωμεν ἐκ τῶν προτέρων ὅτι τοῦτο θὰ εἶναι σπαθὶ μὲ ἐπίπεδον σημαντικότητος 75 %. Πράγματι ἡ πιθανότης τοιαύτης ἐξαγωγῆς εἶναι 8 : 32 = 25 καὶ ἡ διαπίστωσίς μας θὰ εἶναι ἐσφαλμένη τρεῖς φορὰς ἐπὶ τεσσάρων. Περαιτέρω βαίνοντες ἐπὶ ρίψεως ἑνὸς κύβου δυνάμεθα νὰ προκαθορίσωμεν ὅτι θὰ ἐξαχθῇ ἄρτιος ἀριθμὸς μὲ ἐπίπεδον σημαντικότητος 50 %, τῆς πιθανότητος τῆς τοιαύτης ἐξαγωγῆς οὔσης 3 : 6 = 0.5.

ΠΟΡΙΣΜΑΤΑ

Τὸ ὁριζόμενον ἐπίπεδον σημαντικότητος συνδέεται ἀρνητικῶς πρὸς τὸν βαθμὸν ἐμπιστοσύνης.

Οὕτω ἐπίπεδον σημαντικότητος 1 % σημαίνει ὑψηλὸν βαθμὸν ἐμπιστοσύνης, καθόσον ἀντιστοιχεῖ εἰς διαπίστωσιν, ἡ ὁποία θὰ εἶναι ἐσφαλμένη μόνον μίαν φορὰν ἐπὶ τῶν 100.

Ἡ ἔννοια τοῦ ἐπιπέδου σημαντικότητος δύναται νὰ συνδεθῇ εὐκόλως πρὸς τὴν Κανονικὴν Κατανομήν. Ὡς γνωστὸν τὰ 99 % τῶν μονάδων μιᾶς Στατιστικῆς Σειρᾶς τοῦ Κανονικοῦ Νόμου παρουσιάζουσιν ἀπόκλισιν κάτω τῶν 2.58 σ. Ἐὰν λάβωμεν κατὰ τύχην μονάδα τινὰ τῆς Σειρᾶς δυνάμεθα νὰ πιστοποιήσωμεν μὲ ἐπίπεδον σημαντικότητος 1 % ὅτι ἡ ἀπόκλισίς της ἔναντι τοῦ Μέσου θὰ εἶναι κάτω τῶν 2.58 σ. Ὁμοίως ἡ ἀπόκλισις αὐτὴ εἶναι κατωτέρα τοῦ 1,96 σ εἰς ἐπίπεδον σημαντικότητος 5 %, κατωτέρα τῶν 2.33 σ εἰς ἐπίπεδον 2 % καὶ κατωτέρα τῆς σ εἰς ἐπίπεδον 33 %.

Παραδείγματα Κατανομῆς συχνότητος εἰδικοῦ χαρακτηριστικοῦ ὁμάδος Δειγμάτων.

1) Ἔστωσαν ἐκ τοῦ Ληξιαρχείου Θεσσαλονίκης 700 γεννήσεις, ἐξ ὧν 362 ἀρρένων καὶ 338 θηλέων. Ἡ συχνότης γεννήσεως ἄρρενος εἰς τὸ δεῖγμα αὐτὸ θὰ εἶναι

$$p' = \frac{\chi}{\eta} = \frac{362}{700} = 0,517 \qquad q = \frac{338}{700} = 0,482$$

$$\sigma_{p'} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{362 \cdot 338}{700}} = 0.018.$$

Συνεπῶς εἰς ἐπίπεδον σημαντικότητος 5% δυνάμεθα νὰ πιστοποιήσωμεν ὅτι ἡ συχνότης γεννήσεως ἄρρενος μεταξὺ τῶν κάτωθι ὁρίων θὰ εἶναι :

$$0,517-1.96 \cdot 0,018=0,482 \quad \text{καὶ} \quad 0,517+1.96 \cdot 0,018=0,552$$

2) κατὰ τὴν μέτρησιν ἀριθμοῦ τινος μεταλλικῶν ράβδων εὑρέθη ὅτι αὗται διαφέρουσι μεταξύ των λόγῳ ἐλαττώματος κατασκευῆς κατὰ τὸ 1/4 mm. Αἱ γενόμεναι μετρήσεις εἴτε εἶναι πλήρεις ἀριθμοὶ χιλιοστομέτρων εἴτε συνεπάγονται ἐπιπρόσθετον κλάσμα ὡς 1/4, 1/2, 3/4. Κανονικῶς, ἐὰν συνενώσωμεν τὰ μέτρα

τοῦ αὐτοῦ κλάσματος, ἢ τὰ μὴ ἔχοντα τοιοῦτον, θὰ προκύψωσι τέσσαρες ὁμάδες αἰσθητῶς ἴσαι. Ἔστω λοιπὸν ὅτι ἐπὶ 1041 μετρήσεων μόνον 203 ἔχουσι ἐπιπρόσθετον κλάσμα 3/4. Φυσικὸν ὅθεν νὰ διερωτηθῶμεν μήπως ἡ παρατηρηθεῖσα διαφορὰ δύναται νὰ ἀποδοθῇ εἰς τὴν τύχην. Ὑποθέτομεν ὅτι αἱ 1041 μετρήσεις συνιστῶσιν ἓν τυχαῖον δεῖγμα ἑνὸς ἀπεριορίστου πληθυσμοῦ μετρήσεων, διὰ τὸ ὁποῖον ἡ πιθανότης τοῦ χαρακτηριστικοῦ 3/4 εἶναι $p=0.25$. Ἐπανερχόμεθα πάλιν εἰς τὸ ἐρώτημα περὶ τοῦ ποία εἶναι ἡ πιθανότης ὥστε ἐκ μιᾶς κατὰ τύχην ἐκ τοῦ πληθυσμοῦ τούτου ἐκλογῆς νὰ λάβωμεν ἓν δεῖγμα, ὅπερ νὰ παρέχῃ τοιαύτην διαφορὰν μεταξὺ τῆς θεωρητικῆς καὶ τῆς παρατηρηθείσης πιθανότητος.

Ἐφ' ὅσον $p=0.25$ καὶ τὸ Κανονικὸν Σφάλμα $\sigma_p = \sqrt{\dfrac{0{,}25 \cdot 0{,}75}{1041}} = 0.0134$

ἡ παρατηρηθεῖσα συχνότης εἶναι $203 : 1041 = 0.195$ καὶ ἡ διαφορὰ $0.25 - 0.195 = 0.055$. Ἐὰν συγκρίνωμεν τὴν διαφορὰν ταύτην πρὸς τὸ Κανονικὸν Σφάλμα θὰ ἔχωμεν τὴν ἀνηγμένην ἀπόκλισιν $0.055 : 0.0134 = 4.1$.

Δοθέντος ὅτι ἡ Κατανομὴ τοῦ p' εἶναι σχεδὸν κανονική, ἐὰν ἀναζητήσωμεν εἰς ἕνα Πίνακα τοῦ Κανονικοῦ Νόμου τὴν πιθανότητα μιᾶς τοιαύτης ἀνηγμένης ἀποκλίσεως θὰ εὕρωμεν ὅτι περιλαμβάνεται μεταξὺ 0.0001 καὶ 0.00001. Εἶναι κατ' ἀκολουθίαν δύσκολον νὰ πιστεύσωμεν ὅτι ἡ διαφορὰ ὀφείλεται εἰς τὴν τύχην, καθόσον ἐν τοιαύτῃ περιπτώσει θὰ ἀντιστοιχῇ εἰς περίπτωσιν ἐμφανιζομένην ἅπαξ ἐπὶ 10.000. Θὰ συμπεράνωμεν λοιπὸν ὅτι συχνότης τόσον μικρὰ τοῦ χαρακτηριστικοῦ 3/4 ὀφείλεται εἴτε εἰς ἐλάττωμα τῆς συσκευῆς μετρήσεως εἴτε εἰς τάσιν τοῦ χειριστοῦ νὰ ἀντικαθιστᾷ ἓν μέτρον, ὅπερ θὰ ἔδει νὰ προσδιορισθῇ διὰ 3/4, διὰ πλήρους μέτρου ἢ ἔχοντος ἐπιπρόσθετον κλάσμα 1/2.

ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΣ ΜΗΔΕΝ (NULL HYPOTHESIS)

Τὰ ἀνωτέρω παραδείγματα συχνάκις ἀπαντῶνται ἐν τῇ Στατιστικῇ ὑπὸ τὸ ὄνομα Ἔλεγχος Ὑποθέσεως Μηδέν. Βάσει τῶν ἐκ τοῦ δείγματος προκυπτόντων δεδομένων διαμορφοῦμεν τὴν Ὑπόθεσιν Μηδὲν ὡς ἀκολούθως: τοῦ δείγματος ἐξαχθέντος κατὰ τύχην ἔκ τινος πληθυσμοῦ προσδιορισθέντος ὑπολογίζομεν τὸ ἀποτέλεσμα ὅπερ δέον νὰ ἀναμένωμεν. Ἐὰν τὸ παρατηρηθὲν ἀποτέλεσμα διαφέρει ἀπὸ τοῦ τελευταίου τούτου, ἐξετάζομεν μήπως ἡ διαφορὰ αὐτὴ δύναται νὰ ἀποδοθῇ εἰς διακυμάνσεις ὀφειλομένας εἰς τὴν τυχαίαν δειγματοληψίαν. Πρὸς τοῦτο ὑπολογίζεται ἡ διαφορὰ μεταξὺ τῆς ὀφειλομένης εἰς τὴν ὑπόθεσιν τιμῆς καὶ τῆς παρατηθείσης τιμῆς καὶ ἀναζητοῦμεν τὴν πιθανότητα ὅπως ἡ διαφορὰ αὐτὴ ὑπερπηδηθῇ εἰς παρόμοια δείγματα διὰ τῆς ἀποδοχῆς τῆς ἀληθοῦς ὑποθέσεως. Ἡ ὑπόθεσις αὕτη τότε ἀπορρίπτεται ἢ γίνεται ἀποδεκτὴ ἀναλόγως τῆς τιμῆς αὐτῆς τῆς πιθανότητος. Ἐὰν ἡ πιθανότης εἶναι πτωχὴ ἐκλέγομεν μίαν ἐκ τῶν δύο ἀκολούθων ἀποφάσεων:

α) ἀπορρίπτομεν τὴν ὑπόθεσιν, δεχόμενοι ὅτι εἶναι ἀδύνατον νὰ ὑπέστη τὸ δεῖγμα κάτι τοῦ ὁποίου ἡ πιθανότης θὰ ἦτο ἰσχνὴ ἐὰν ἡ ὑπόθεσις ἦτο ἀληθής.

β) Δεχόμεθα τὴν ὑπόθεσιν, θεωροῦντες ὅτι εἶναι λογικὸν νὰ δεχθῶμεν ὅτι συνέβη εἰς τὸ δεῖγμα κάτι τὸ ἀρκετὰ σπάνιον.

Ἀνηγμένη Ἀπόκλισις	Πιθανότης	Ἀνηγμένη Ἀπόκλισις	Πιθανότης
0	1	1.881	0.06
0.126	0.9	1.960	0.05
0.253	0.8	2.054	0.04
0.385	0.7	2.170	0.03
0.524	0.6	2.326	0.02
0.674	0.5	2.576	0.01
0.842	0.4	3.291	0.001
1.036	0.3	3.891	0.0001
1.282	0.2	4.417	0.00001
1.645	0.1	4.892	0.000001
1.695	0.09	5.327	0.0000001
1.751	0.08	5.731	0.00000001
1.812	0.07	6.109	0.000000001

Ἐὰν ἡ πιθανότης εἶναι ἀρκετὰ ἰσχνή, μικροτέρα τοῦ 2% ἢ 1%, θὰ προτιμήσωμεν συνήθως τὴν πρώτην κατάστασιν. Παρὰ ταῦτα ὅταν ἡ πιθανότης εἶναι ἴση ἢ ἀνωτέρα τοῦ 5 % πολλάκις εἶναι δυνατὸν νὰ στραφῶμεν πρὸς τὴν δευτέραν.

Ὅθεν τὸ ἐπίπεδον σημαντικότητος διὰ τοῦ ὁποίου ἀπορρίπτομεν τὴν ὑπόθεσιν ἐξαρτᾶται ἐκ τῆς πιθανότητος ἑνὸς σφάλματος, ἀνωτέρου τοῦ ὑπολογισθέντος εἰς τὴν περίπτωσιν τῆς ὑποθέσεως. Ἐὰν ἡ πιθανότης αὐτὴ εἶναι 2% δυνάμεθα ἐπὶ παραδείγματι νὰ ἀπορρίψωμεν τὴν ὑπόθεσιν μὲ ἐπίπεδον σημαντικότητος 2%, ἀλλὰ πρὸ τῆς κατηγορηματικῆς ἀπορρίψεως θὰ ἐξετάσωμεν μήπως ἠμποροῦμεν νὰ τὸ πράξωμεν εἰς τὸ ἐπίπεδον 5 % ἢ εἰς τὸ τοιοῦτον 1 %. Ὁμολογουμένως ἡ ἀπόφασις τῆς ἀπορρίψεως ἢ ἀποδοχῆς ἐξαρτᾶται ὀλίγον καὶ ἐκ τῆς ἰδιοσυστασίας τοῦ Στατιστικοῦ ὡς καὶ τῶν πρακτικῶν ἀποτελεσμάτων τῆς ἀποφάσεώς του. Θὰ ἦτο παρακινδυνευμένον νὰ καθορίσωμεν a priori κανόνας. Ἡ ἐπιλογὴ τοῦ καταλλήλου ἐπιπέδου πρέπει νὰ ἀποφασίζεται ἀντικειμενικῶς καὶ ἀναλόγως τῆς συγκεκριμένης ὑπὸ ἐξέτασιν καταστάσεως μὲ ποιάν τινα δόσιν φρονήσεως. Ἐκ τῶν ἀνωτέρω ὑπογραμμισθέντων συνάγεται ὅτι ἡ χρησιμοποιηθεῖσα μέθοδος ὁμοιάζει πρὸς τὴν τοιαύτην τῶν Μαθηματικῶν κατὰ τὴν ἔμμεσον ἀπόδειξιν. Οὕτω εἰς τὸ προηγούμενον πρόβλημα ἐξ ἐπόψεως ἀριθμοῦ ἐξηγγείλαμεν τὴν πρόγνωσιν ὅτι τὰ κλάσματα $^3/_4$ εἶχον παραμεληθῆ. Ἡ πρώτη μας σκέψις ἦτο νὰ κάμωμεν τὴν ἀντίθετον ὑπόθεσιν καὶ βάσει ταύτης εὕρομεν ὅτι ἡ πιθανότης μιᾶς τόσον μεγάλης διαφορᾶς ἦτο μικροτέρα τοῦ 0.0001, πρᾶγμα ὅπερ ἤρχετο εἰς ἀντίθεσιν πρὸς τὴν ὑπόθεσίν μας, ἥτις θὰ ἔπρεπε πλέον νὰ θεωρηθῇ ὡς ἀνακριβής.

Σημειοῦμεν ἐπίσης ὅτι ἐκ τοῦ γεγονότος τοῦ ἐπιπέδου σημαντικότητος ὁδηγούμεθα πολλάκις νὰ ἀπορρίπτωμεν ἀληθεῖς ὑποθέσεις. Ἐὰν ἐπὶ παραδείγματι χρησιμοποιήσωμεν τὸ ἐπίπεδον 1 %, ἡ ἀναλογία των θὰ εἶναι 1 %. Ὁμοίως ἡ ἐλάττωσις τοῦ ἐπιπέδου σημαντικότητος συνεπάγεται τὴν ἀποδοχὴν μεγαλυτέρου ἀριθμοῦ σφαλερῶν ὑποθέσεων.

Ἐκτίμησις τῶν Παραμέτρων ἑνὸς Πληθυσμοῦ βάσει Δείγματος.

Βάσει ἀποτελεσμάτων, ληφθέντων ἐκ δείγματος ἐξαχθέντος τυχαίως ἔκ τινος πληθυσμοῦ, δυνάμεθα νὰ τὰ ἐπεκτείνωμεν ἐπὶ τοῦ πληθυσμοῦ τούτου. Ὑποτίθεται ὅτι ὑπελογίσθησαν διὰ τὸ δεῖγμα ὁ $\overline{X}_\delta$ καὶ ἡ σ_δ. Αἱ καλλίτεραι ἐκτιμήσεις τῶν $\overline{X}_\pi$ καὶ σ_π τοῦ πληθυσμοῦ δίδονται ὑπὸ τῶν τύπων:

$$\overline{X}_\pi = \overline{X}_\delta \qquad \boxed{\sigma_\pi = \sqrt{\frac{n}{n-1}}\,\sigma_\delta}$$

ἔνθα n εἶναι αἱ μονάδες τοῦ δείγματος.

Ἐὰν θελήσωμεν νὰ ἐκφράσωμεν τὴν σ τοῦ Μέσου τοῦ πληθυσμοῦ βάσει τῶν παρατηρηθέντων ἀποτελεσμάτων θὰ πρέπῃ νὰ περιορίσωμεν τὴν σ_π μεταξὺ τῶν ἰσοτήτων:

$$\sigma_{\overline{X}} = \frac{\sigma_\pi}{\sqrt{n}} \text{ καὶ } \sigma_\pi = \sqrt{\frac{n}{n-1}}\,\sigma_\delta$$

ὁπότε προκύπτει ὅτι $\sigma_{\overline{X}} = \frac{\sigma_\delta}{\sqrt{n-1}}$

Ἡ ἐκτίμησις τοῦ Μέσου τοῦ Πληθυσμοῦ δύναται νὰ καθορισθῇ διὰ τοῦ προσδιορισμοῦ τοῦ βαθμοῦ ἐμπιστοσύνης, τὸν ὁποῖον δυνάμεθα νὰ τοῦ ἀποδώσωμεν. Οὕτω θὰ ἔχωμεν τὸν τύπον:

$$\boxed{\overline{X} = \overline{X}_\delta \pm \frac{\sigma_\delta}{\sqrt{n-1}}}$$

ὅστις δεικνύει ὅτι ὑπάρχει μία πιθανότης περίπου 2/3 ὅτι ὁ μέσος τοῦ Πληθυσμοῦ θὰ περιλαμβάνηται εἰς τὸ διάστημα:

$$\overline{X}_\delta - \frac{\sigma_\delta}{\sqrt{n-1}} \quad \text{καὶ} \quad \overline{X}_\delta + \frac{\sigma_\delta}{\sqrt{n-1}}$$

Ἀλλ' εἶναι προφανὲς ὅτι ἔχομεν τὴν εὐχέρειαν νὰ ὁρίσωμεν καὶ ἄλλα διαστήματα, εἰς τὰ ὁποῖα νὰ διαπιστώσωμεν μὲ συγκεκριμένους βαθμοὺς πιθανότητος ποῦ εὑρίσκεται ὁ Μέσος, τὸ δὲ κατὰ τοιοῦτον τρόπον ὁριζόμενον διάστημα ὁρολογεῖται ὡς Διάστημα Ἐμπιστοσύνης (confidence interval) καὶ τὰ ὅριά του Ὅρια Ἐμπιστοσύνης (confidence limits).

Παραδείγματα

Ἔστω δεῖγμα 65 μηνῶν μὲ $\overline{X}_\delta = 77$cm καὶ $\sigma_\delta = 9.6$ cm $n = 65$. Ἡ σ του πληθυσμοῦ θὰ εἶναι

$$\sigma_{\overline{X}} = \frac{9.6}{\sqrt{64}} = 1{,}2 \text{ cm}$$

καὶ εἰς τὸ ἐπίπεδον σημαντικότητος 5% τὸ διάστημα ἐμπιστοσύνης θὰ ἔχῃ ὡς ὅρια $77+1.96\times12 = 79.35$ cm καὶ $77-1.96\times12 = 77.65$ cm.

Μὲ πιθανότητα 98 % ἢ ἐπίπεδον σημαντικότητος 2 % πιστοποιοῦμεν ὅτι ὁ Μέσος τοῦ Πληθυσμοῦ θὰ περιλαμβάνεται μεταξὺ τῶν ὁρίων $77 + 2.33\times1{,}2 =$

$=79.6$ cm καὶ $77-2.33\times1.2=74.4$ cm καὶ τέλος τὸ διάστημα ἐμπιστοσύνης εἰς ἐπίπεδον 0.1 % περιορίζεται εἰς $77+3.3\times1.2=80.96$ cm καὶ $77-3.3\times1.2=$ $=73.04$ cm. Χρησιμοποιοῦντες τὸ διάστημα τοῦτο ἀποδεικνύομεν μίαν ὑπόθεσιν, γενομένην ἐπὶ τοῦ Μέσου τοῦ Πληθυσμοῦ. Ἐὰν ὑποθέσωμεν ὅτι ὁ Μέσος εἶναι 79, τοῦ εἰς τὴν τιμὴν αὐτὴν ἀντιστοιχοῦντος διαστήματος ἐμπιστοσύνης ὄντος εἰς ἐπίπεδον 5 %, ἀποφαινόμεθα ὅτι ἡ ὑπόθεσίς μας βασίζεται ἐπὶ τοῦ ἐπιπέδου 5 %. Ἐὰν τὰ διαστήματα ἐμπιστοσύνης δὲν εἶναι ἐκ τῶν προτέρων γνωστὰ ἀρκεῖ νὰ σχηματίσωμεν τὴν σχέσιν τῆς ληφθείσης ἐμπιστοσύνης, διαιροῦντες τὴν διαφορὰν μεταξὺ τοῦ ὑποθετικοῦ Μέσου καὶ τοῦ παρατηρηθέντος Μέσου διὰ τῆς Σταθερᾶς Ἀποκλίσεως καὶ περαιτέρω ἀναζητοῦντες τὴν πιθανότητα μιᾶς Ἀποκλίσεως ἀνωτέρας βάσει ἑνὸς Πίνακος. $79-77=2$ cm $2:1.2=$ $=1.7$, ὁπότε ἡ ὑπόθεσίς μας δικαιολογεῖται περίπου εἰς τὸ ἐπίπεδον 9 %.

Κανονικὰ Σφάλματα τῶν Παραμέτρων Κεντρικῆς Τάσεως

Γενικῶς Κανονικὸν Σφάλμα μιᾶς στατιστικῆς ποσότητος καλεῖται ἡ Σταθερὰ Ἀπόκλισις τῆς Κατανομῆς αὐτῆς τῆς ποσότητος εἴς τινα οἰκογένειαν δειγμάτων

Ἀριθμητικοῦ Μέσου $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Διαμέσου $\sigma_{me} = \dfrac{5}{4} \cdot \sigma_{\bar{x}}$

Τεταρτημορίων $\sigma_Q = 0{,}787\ \sigma_{\bar{x}}$

Σταθερᾶς Ἀποκλίσεως $\sigma_\sigma = 0{,}707\ \sigma_{\bar{x}}$

Διὰ μεγάλα Δείγματα ἐξαχθέντα κατὰ τύχην ἐκ πληθυσμῶν αἰσθητῶς κανονικῶν, οἱ ἀνωτέρω τύποι δύνανται νὰ χρησιμοποιηθῶσιν ἐπωφελῶς, καθόσον ὑποτίθεται ὅτι τὰ εἰς τὴν δειγματοληψίαν ὀφειλόμενα σφάλματα κατανέμονται κανονικῶς. Τοῦτο ὅμως δὲν συμβαίνει εἰς Μικρὰ Δείγματα ($n<30$).

ΚΡΙΤΗΡΙΟΝ χ^2 PEARSON

Μέχρι τοῦδε ἐξητάσαμεν τὸν προσδιοριζόμενον βαθμὸν ἐμπιστοσύνης εἰς μίαν στατιστικὴν ποσότητα ὁρισθεῖσαν βάσει ἑνὸς δείγματος. Ἤδη πρόκειται νὰ ἐρευνήσωμεν ἐὰν τὸ δεῖγμα εἶναι ὄντως ἀντιπροσωπευτικὸν τοῦ πληθυσμοῦ ἐξ οὗ ἐξήχθη καὶ ἡ σχετικὴ μέθοδος ἐπενοήθη παρὰ τοῦ K. Pearson διὰ τοῦ κριτηρίου χ^2, ὅπερ εἶναι τὸ ἄθροισμα διὰ τὸ σύνολον N παρατηρήσεων τῶν ληφθεισῶν ποσοτήτων ἐὰν διαιρέσωμεν τὸ τετράγωνον τῆς διαφορᾶς θεωρητικῶν - πραγματικῶν παρατηρήσεων διὰ τῆς θεωρητικῆς τοιαύτης:

$$\boxed{\chi^2 = \sum \left(\frac{(f_o - f)^2}{f} \right)}$$

Ἔνθα: f_o αἱ παρατηρηθεῖσαι τιμαί, f αἱ θεωρητικαὶ τιμαί.

Αἱ τιμαὶ τοῦ χ^2 αἱ ἀντιστοιχοῦσαι εἰς βαθμοὺς ἐλευθερίας 1—30 καὶ τὰς πλέον συνήθεις τιμὰς τῆς πιθανότητος P ἔχουσιν ὑπολογισθῆ παρὰ τοῦ Pearson εἰς ἴδιον Πίνακα. Διὰ νὰ χρησιμοποιήσωμεν τὸν Πίνακα ζητοῦμεν εἰς τὴν γραμμήν, ἥτις ἀντιστοιχεῖ εἰς τὸν ἀριθμὸν τῶν βαθμῶν ἐλευθερίας, τὴν τιμὴν χ^2, ἥτις εὑρέθη διὰ τὸ δεῖγμα. Ἡ ἀντίστοιχος τιμὴ τοῦ P εἰς τὴν ...

Πίνακος μᾶς δίδει τὴν πιθανότητα διὰ μίαν τιμὴν τοῦ χ^2 ἴσην ἢ ἀνωτέραν ἐκείνης ἥτις ἐδόθη ἐκ τοῦ δείγματος. Γενικῶς τὸ P προσδιορίζεται διὰ μιᾶς κατὰ προσέγγισιν παρεμβολῆς, δοθέντος ὅτι ἡ παρατηρηθεῖσα τιμὴ τοῦ χ^2 σπανίως θὰ εὑρεθῇ εἰς τὸν Πίνακα.

Παρατηρήσεις ἐπὶ τοῦ Κριτηρίου : α) Τὸ συνολικὸν N πρέπει νὰ εἶναι ἀρκούντως μέγα, τουλάχιστον δὲ 50. β) Τὸ δυναμικὸν (ἀριθμὸς) (effectif) ἑκάστης τάξεως πρέπει νὰ εἶναι κατὰ τὸ δυνατὸν ἀνώτερον τοῦ 5. γ) Ὁ ἀριθμὸς τῶν τάξεων πρέπει νὰ εἶναι κάτω τῶν 20. δ) Τιμαὶ τοῦ P γειτονικαὶ πρὸς τὸ 1 ἢ 0 παρέχουσιν ἱκανὰς ἐνδείξεις ἐπὶ τοῦ χαρακτῆρος τοῦ δείγματος. ε) Τὸ Κριτήριον χρησιμεύει διὰ τὴν σύγκρισιν δύο ἢ πλειόνων μεταβλητῶν˙. Ἐ ὰ ν τ ὸ P ε ἶ ν α ι μ ι κ ρ ό τ ε ρ ο ν τ ο ῦ 0,02 τ ὸ δ ε ῖ γ μ α γ ε ν ι κ ῶ ς δ έ ο ν ν ὰ θ ε ω ρ η θ ῇ ὡ ς δ ε ῖ γ μ α μ ὴ ἐ ξ α χ θ ὲ ν ἐ κ τ ύ χ η ς. Ἐ ὰ ν τ ὸ P κ ε ῖ τ α ι μ ε τ α ξ ὺ 0.02 κ α ὶ 0.05 τ ὸ ὀ λ ι γ ώ τ ε ρ ο ν ὅ π ε ρ δ υ ν ά μ ε θ α ν ὰ ε ἴ π ω μ ε ν ε ἶ ν α ι ὅ τ ι τ ὸ δ ε ῖ γ μ α ε ἶ ν α ι ἀ ρ κ ε τ ὰ ὕ π ο π τ ο ν. στ) Τὸ κριτήριον εἶναι τὸ μέσον διὰ νὰ κρίνωμεν τὴν τιμὴν μιᾶς ὑποθέσεως, ἀφορώσης τὸν πληθυσμὸν οὗτινος γνωρίζομεν ἓν δεῖγμα. Ἡ ὑπόθεσις θὰ εἶναι τοιαύτη ὥστε νὰ εἶναι δυνατὸς ὁ ὑπολογισμὸς εἴτε τοῦ χ^2 εἴτε στατιστικῆς τινος ποσότητος κατανεμομένης ὅπως τὸ χ^2 ὡς καὶ τοῦ ἀριθμοῦ τῶν βαθμῶν ἐλευθερίας. Ὁ Πιναξ θὰ μᾶς δώσῃ τὴν πιθανότητα P ὅπως εἰς ἓν τυχαῖον δεῖγμα ἡ τιμὴ τοῦ χ^2 ἢ τῆς στατιστικῆς ποσότητος εἶναι μεγαλυτέρα ἀπὸ τὴν παρατηρηθεῖσαν τιμήν. Ἐ ὰ ν ἡ π ι θ α ν ό τ η ς ε ἶ ν α ι π ο λ ὺ μ ι κ ρ ὰ θ ὰ σ υ μ π ε ρ ά ν ω μ ὲ ν ὅ τ ι ἡ ὑ π ό θ ε σ ι ς π ρ έ π ε ι ν ὰ ἀ π ο ρ ρ ι φ θ ῇ. ἄλλως τὸ κριτήριον δὲν θὰ παράσχῃ ἀκριβὲς συμπέρασμα καὶ ἡ ὑπόθεσις θὰ δύναται νὰ θεωρηθῇ ὡς πιθανὴ ἀφοῦ τὸ κριτήριον δὲν θὰ μᾶς δίδῃ καμμίαν ἀπόδειξιν ἐναντίον της. ζ) Γενικῶς ἐὰν ἡ δειχθεῖσα τιμὴ τοῦ P εἶναι μικροτέρα εἰδικῆς τινος τιμῆς, συνήθως τοῦ 0,05 ἢ 0,01, αἱ διαφοραὶ εἶναι ἀρκετὰ μεγάλαι διὰ νὰ ἀποδοθῶσιν εἰς τὴν τύχην. Τὸ ἀποδεκτὸν ὅριον τιμῆς τοῦ P 0,05 ἢ 0,01 καλεῖται Σ η μ ε ῖ ο ν Ἐ μ π ι σ τ ο σ ύ ν η ς (fiducial point). Μ ὲ ἄ λ λ ο υ ς λ ό γ ο υ ς ἐ ὰ ν τ ὸ κ ρ ι τ ή ρ ι ο ν χ^2 δ ε ί ξ ῃ ὅ τ ι ἡ δ ι α φ ο ρ ὰ μ ε τ α ξ ὺ π ρ α γ μ α τ ι κ ῶ ν κ α ὶ ἀ ν α μ ε ν ο μ έ ν ω ν σ υ χ ν ο τ ή τ ω ν ε ἶ ν α ι ἀ ρ κ ε τ ὰ μ ε γ ά λ η δ ι ὰ ν ὰ ἀ π ο δ ο θ ῇ ε ἰ ς τ ὴ ν τ ύ χ η ν, δ η λ. τ ὸ P ε ἶ ν α ι μ ι κ ρ ό τ ε ρ ο ν τ ο ῦ ἐ κ λ ε γ έ ν τ ο ς σ η μ ε ί ο υ ἐ μ π ι σ τ ο σ ύ ν η ς 0.01 ἢ 0,05, ἡ ὑ π ό θ ε σ ι ς δ ύ ν α τ α ι ν ὰ λ ε χ θ ῇ ὅ τ ι ε ἶ ν α ι ἐ σ φ α λ μ έ ν η.

Παράδειγμα

Εἰς πληθυσμὸν ἐνηλίκων ἀρρένων πέντε μεγάλων πόλεων βάσει δειγμάτων τυχαίων λαμβάνομεν τὸν ἀριθμὸν τῶν ἐγγάμων καὶ τῶν ἀγάμων:

Πόλεις	Α	Β	Γ	Δ	Ε	Σύνολα
Ἔγγαμοι	130	165	151	172	146	764
Ἄγαμοι	39	56	47	67	43	252
Σύνολα	169	221	198	239	189	1016

Ἐρωτᾶται ἐὰν τὰ δεδομένα αὐτὰ δεικνύουσι μίαν σημαντικὴν διακύμανσιν τῆς τάσεως τῶν ἀνδρῶν νὰ νυμφεύωνται ἀναλόγως τῶν διαφόρων πόλεων. Ἂς λάβωμεν τὴν ὑπόθεσιν ὅτι αἱ διακυμάνσεις αὐταὶ εἶναι ἀσήμαντοι. Τότε τὰ ἄτομα ἑκάστης πόλεως συνιστῶσιν ἓν ἁπλοῦν δεῖγμα ἑνὸς πληθυσμοῦ, εἰς τὸν ὁποῖον κατὰ τὰ σύνολα τῆς τελευταίας στήλης ἡ σχέσις τῶν ἀγάμων ἔναντι τοῦ συνολικοῦ πλυθυσμοῦ εἶναι 1/4. Βασιζόμενοι εἰς τὴν ὑπόθεσιν αὐτὴν ἀναπαριστῶμεν τὰ θεωρητικὰ δεδομένα ἑκάστης πόλεως ὡς κατωτέρω:

Πόλεις	Α	Β	Γ	Δ	Ε	Σύνολον
Ἔγγαμοι	127	166	149	180	142	764
Ἄγαμοι	42	55	49	59	47	252
Σύνολα	169	221	198	239	189	1016

Ἐκ τῶν ἀριθμῶν τῶν δύο Πινάκων ὑπολογίζομεν τὰς ἀποκλίσεις τῶν θεωρητικῶν ἀπὸ τῶν πραγματικῶν δεδομένων, ὑψοῦμεν εἰς τὸ τετράγωνον κα. διαιροῦμεν διὰ τῶν θεωρητικῶν δεδομένων διὰ νὰ εὕρωμεν τὴν τιμὴν χ^2 ἰ

$$\chi^2 = \frac{3^2}{127} + \frac{(-1)^2}{166} + \frac{2^2}{149} + \frac{(-8)^2}{180} + \frac{4^2}{142} + \frac{(-3)^2}{42}$$

$$+ \frac{1^2}{55} + \frac{(-2)^2}{49} + \frac{8^2}{59} + \frac{(-4)^2}{47} = 2{,}307 \qquad n = 5 \qquad n - 1 = 4$$

Ὁ Πίναξ δεικνύει ὅτι ἡ πιθανότης νὰ ὑπερβάλωμεν τὴν τιμὴν 2.307 διὰ χ^2 περιλαμβάνεται μεταξὺ 0.50 καὶ 0.70, ἀλλ' ἡ τιμὴ αὐτὴ δὲν εἶναι σημαντικὴ καὶ τὸ Κριτήριον χ^2 οὐδεμίαν ἀπόδειξιν παρέχει ἐναντίον τῆς ὑποθέσεώς μας.

Ὁσάκις εἰς Κατανομὴν Συχνότητος τὰ δεδομένα ἔχουσι προσαρμοσθῇ πρὸς τὸν Κανονικὸν Νόμον τῆς Κατανομῆς, ὁ ὑπολογισμὸς τῶν Βαθμῶν διαφέρει καὶ οὕτως ἀντὶ n–1 θὰ ἔχωμεν n–3, δοθέντος ὅτι αἱ $\bar{X}$, σ, N ἔχουσιν ὑπολογισθῇ.

Τέλος διὰ νὰ ἀποφανθῶμεν ἐὰν τὸ χ^2 εἶναι στατιστικῶς σημαντικὸν πρέπει νὰ εὕρωμεν τιμὴν τούτου ὑπερβαίνουσαν τὸν ἀριθμὸν 6 ἢ κατ' ἄλλους τὸν ἀριθμὸν 9.

ΠΕΡΙΠΤΩΣΕΙΣ ΜΙΚΡΩΝ ΔΕΙΓΜΑΤΩΝ

Προκειμένου περὶ μικρῶν δειγμάτων ($n < 30$) ὁ Student ἀντικατέστησε τὴν ἀνηγμένην ἀπόκλισιν διὰ μιᾶς νέας στατιστικῆς μεταβλητῆς:

$$\boxed{t = \frac{|\bar{X}_\delta - \bar{X}_\pi|}{S}\sqrt{n}}$$

Ἔνθα
$\bar{X}_\pi$ ὁ Μέσος του Πληθυσμοῦ
$\bar{X}_\delta$ ὁ Μέσος τοῦ Δείγματος

καὶ S εἶναι μία ποσότης, τῆς ὁποίας τὸ τετράγωνον S^2 καλεῖται Ἀκριβὴς Ἐκτίμησις τῆς Διακυμάνσεως τοῦ Πληθυσμοῦ ἐκ τοῦ Δείγματος

$$\boxed{S^2 = \frac{\Sigma\,(x_1 - \bar{X}_\delta)^2}{n - 1}}$$

ὅπου $n - 1$ εἶναι οἱ βαθμοὶ ἐλευθερίας τοῦ δείγματος.

Ἡ κατανομὴ τῆς μεταβλητῆς t τοῦ Student ἐμελετήθη καὶ ἀνευρέθησαν Πίνακες κριτικῶν τιμῶν τοῦ t διὰ τὰς διαφόρους τιμὰς καὶ τὰ οἰκεῖα ἐπίπεδα σημαντικότητος. Ἡ χρῆσις τοῦ Πίνακος t εἶναι ἀνάλογος πρὸς τὸν τοιοῦτον τῶν Πινάκων τοῦ Κανονικοῦ Νόμου ἢ τοῦ χ^2. Παραδείγματά τινα θὰ ἐξηγήσωσι τὸν μηχανισμόν :

1) Ὑποθέσωμεν ὅτι ἔχομεν ἓν τυχαῖον δεῖγμα ἔκ τινος πληθυσμοῦ μιᾶς μεγάλης πόλεως, περιλαμβάνον 9 ἄτομα, τῶν ὁποίων τὸ μέσον ἀνάστημα εἶναι 171 ἑκατοστόμετρα καὶ ἡ ἀκριβὴς ἐκτίμησις τῆς διακυμάνσεως τοῦ ἄρρενος πληθυσμοῦ αὐτῆς τῆς πόλεως ἔστω ὅτι εἶναι 81 cm. Κατὰ ποῖον μέτρον δυνάμεθα νὰ πιστεύσωμεν ὅτι τὸ μέσον ἀνάστημα τῶν ἀτόμων τῆς πόλεως ταύτης θὰ εἶναι 173 cm ; Διατάσσομεν τὰ δεδομένα μας ὡς ἑξῆς :

$$\bar{X}_\delta = 171 \quad \bar{X}_\pi = 173 \quad n-1 = 8 \quad S = \sqrt{81} = 9 \text{ cm}.$$

Ἐν συνεχείᾳ κατὰ τὴν ὑπόθεσιν ὅτι $\bar{X}_\pi = 173$ σχηματίζομεν διὰ τὴν μεταβλητὴν t τὴν τιμὴν

$$t = \frac{171-173}{9}\sqrt{9} = 0.666$$

Ὁ Πίναξ t δεικνύει ὅτι διὰ $n-1 = 8$ ἡ πιθανότης ὅτι ἡ τιμὴ αὐτὴ τοῦ t θὰ ὑπερβαίνηται ἀριθμητικῶς εἶναι μεγαλυτέρα τοῦ 0.50. Ἡ τιμὴ ὅθεν αὐτὴ δὲν εἶναι σημαντικὴ καὶ τὸ κριτήριον οὐδεμίαν παρέχει ἀπόδειξιν ἔναντι τῆς Ὑποθέσεως ὅτι ὁ Μέσος τῶν ἀναστημάτων τῆς πόλεως αὐτῆς εἶναι 173 cm. Ἐὰν εἴχομεν κάμει τὴν ὑπόθεσιν ὅτι τὸ μέσον ἀνάστημα εἶναι 177 cm θὰ εὑρίσκομεν διὰ t τὴν τιμὴν

$$t = \frac{171-177}{9}\sqrt{9} = 2$$

Ὁ Πίναξ δεικνύει ὅτι διὰ $n-1 = 8$ ἡ πιθανότης ὅπως τοιαύτη τιμὴ τοῦ t ὑπερβληθῇ εἶναι μεγαλυτέρα τοῦ 0.05, ἀλλὰ καὶ ἀκόμη ἡ νέα αὐτὴ τιμὴ δὲν εἶναι σημαντική. Κατὰ συνέπειαν ὑπάρχουσιν ἀρκετὰ μεγάλα ὅρια διὰ τὸν ὑποθετικὸν Μέσον τῶν ἀναστημάτων, τοιούτων ὥστε διὰ τὰ δεδομένα τοῦ δείγματος ἡ τιμὴ τοῦ t νὰ μὴν εἶναι σημαντική. Ἄς προχωρήσωμεν ὅθεν εἰς τὸν προσδιορισμὸν τοῦ διαστήματος ἐμπιστοσύνης μὲ ἐπίπεδον σημαντικότητος τὸ 5 %. Ὁ Πίναξ μᾶς δεικνύει ὅτι διὰ $P = 0.05$ καὶ $n-1 = 8$ ἡ τιμὴ $t = 2{,}31$, ὁπότε θὰ ἔχωμεν :

$$\frac{|171-\bar{X}_\pi|}{9}\sqrt{9} < 2.31 \text{ ἐξ οὗ } [\,171-\bar{X}_\pi\,] < 3\ 2.31$$

καὶ $171 - 3 \times 2.31 < \bar{X}_\pi < 171 + 3 \times 2.31$.

Τὰ ὅρια τοῦ διαστήματος ἐμπιστοσύνης εἰς τὸ ἐπίπεδον 5 % εἶναι 164.07 καὶ 177.93 cm.

2) Ὑποθέσωμεν ὅτι ἔχομεν ὁρίσει εἰς πέντε ἀσθενεῖς ἓν φάρμακον καὶ ἀκολούθως κατεγράψαμεν δι' ἕκαστον ἐξ αὐτῶν τὴν αὔξησιν τῆς πιέσεως τοῦ αἵματος ὡς 6, 2, –2, 3, –4. Τὰ ληφθέντα ἀποτελέσματα δεικνύουσιν ἆραγε μίαν ἐνέργειαν τοῦ φαρμάκου; Ἐνταῦθα ἡ πρὸς ἔλεγχον Ὑπόθεσις Μηδὲν μᾶς λέγει ὅτι τὰ ἀποτελέσματα ἐξήχθησαν κατὰ τύχην ἔκ τινος πληθυσμοῦ κανο-

νικοῦ, ἔχοντος ὡς Μέσον τὸ μηδέν. Ὑπολογίζοντες τὸν Μέσον τοῦ δείγματος ἔχομεν :

$$\bar{X}_\delta = \frac{6+2+(-2)+3+(-4)}{5} = 1 \qquad t = \frac{1}{4}\sqrt{5} = 0{,}559$$

$$S^2 = \frac{5^2+1^2+(-3)^2+2^2+(-5)^2}{4} = \frac{64}{4} = 16$$

Ὁ Πίναξ μᾶς δεικνύει ὅτι ἡ πιθανότης νὰ ὑπερβῶμεν αὐτὴν τὴν τιμὴν εἶναι ἀνωτέρα τοῦ 50 % καὶ ἡ τιμὴ τοῦ t δὲν εἶναι σημαντική, ἡ δὲ Ὑπόθεσις Μηδὲν δὲν εἶναι συνεπῶς ἀνίσχυρος, δηλ. εἶναι δυνατὸν αἱ διακυμάνσεις τῆς πιέσεως τοῦ αἵματος μὴ ὀφείλωνται εἰς τὸ φάρμακον.

Σημαντικότης τῆς διαφορᾶς δύο $\bar{X}$.

Ἐὰν δύο μικρὰ ἀνεξάρτητα δείγματα n_1 n_2 μὲ Μέσους ἀντιστοίχως $\bar{X}_1$ καὶ $\bar{X}_2$ εἶναι γνωστά, δυνάμεθα καὶ πάλιν νὰ χρησιμοποιήσωμεν τὴν Κατανομὴν t διὰ νὰ ἀποφασίσωμεν ἐὰν ἡ διαφορὰ τῶν Μέσων των εἶναι σημαντικὴ ἢ ἐὰν τὰ δύο δείγματα δέον νὰ θεωρηθῶσιν ἐξαχθέντα ἐκ τοῦ αὐτοῦ κανονικοῦ πληθυσμοῦ. Ἂς λάβωμεν τὴν ὑπόθεσιν ὅτι τὰ δύο δείγματα προέρχονται ἐκ τοῦ αὐτοῦ πληθυσμοῦ. Τότε ἡ ἔκφρασις τῆς χρησιμοποιητέας στατιστικῆς μεταβλητῆς λαμβάνει τὴν ἀκόλουθον μορφήν :

$$\boxed{t = \frac{|\bar{X}_1 - \bar{X}_2|}{S}\sqrt{\frac{n_1 \times n_2}{n_1 + n_2}}}$$

$$\boxed{S^2 = \frac{\Sigma\,(X_1 - \bar{X}_1)^2 + \Sigma\,(X_2 - \bar{X}_2)^2}{(n_1 + n_2) - 2}}$$

Παραδείγματα

1) Ἔστω ἓν δεῖγμα 9 ἀτόμων μὲ μέσον ἀνάστημα 171 cm καὶ ἄθροισμα τῶν τετραγώνων τῶν ἀποκλίσεων ἀπὸ τοῦ Μέσου 60 καὶ ἕτερον δεῖγμα 8 ἀτόμων μὲ μέσον ἀνάστημα 173 cm καὶ ἄθροισμα τετραγώνων ἀποκλίσεων ἀπὸ Μέσου 90. Δυνάμεθα νὰ θεωρήσωμεν τὰ δύο αὐτὰ δείγματα ὡς ἐξαχθέντα ἐκ τοῦ αὐτοῦ πληθυσμοῦ ; Δεχόμενοι τὴν ὑπόθεσιν ταύτην θὰ ἔχωμεν :

$$S^2 = \frac{60+90}{17-2} = 10$$

$$t = \frac{|\,171 - 173\,|}{\sqrt{10}}\sqrt{\frac{8\times 9}{8+9}} = 1{,}30$$

Ὁ Πίναξ t δεικνύει ὅτι διὰ $n-2=15$ ἡ τιμὴ 1,30 δὲν εἶναι σημαντική. Τὸ κριτήριον ὅθεν οὐδεμίαν παρέχει ἀπόδειξιν ἐναντίον τῆς ὑποθέσεως, ἥτις δύναται νὰ θεωρηθῇ ὡς ἰσχύουσα.

2) Δύο Ἐργοστάσια παράγουσι τὸ αὐτὸ προϊόν. Λαμβάνομεν κατὰ τύχην 8 ἐργάτας ἐκ τοῦ Ἐργοστασίου Α καὶ 10 ἐκ τοῦ Ἐργοστασίου Β, σημειοῦντες τὸν ἀριθμὸν τῶν παρ᾽ ἑκάστου ἐξ αὐτῶν παραχθέντων τεμαχίων εἰς δοθέντα χρόνον. Τὰ ἀποτελέσματα αὐτὰ καταγράφονται κατωτέρω :

7. 8. 140

Ἐργοστάσιον Α			Ἐργοστάσιον Β		
Τεμάχια	$X-\bar{X}_A$	$(X-\bar{X}_A)^2$	Τεμάχια	$X-\bar{X}_B$	$(X-\bar{X}_B)^2$
12	—1	1	13	—1	1
11	—2	4	15	—1	1
13	0	0	12	—2	4
15	—2	4	17	—3	9
10	—3	9	11	—3	9
17	—4	16	18	—4	16
12	—1	1	12	—2	4
14	1	1	16	2	4
			14	—0	0
			12	—2	4
103		36	140		52
$\bar{X}_A = 13$			$\bar{X}_B = 14$		

Δύναταί τις νὰ συμπεράνῃ ὅτι ἡ παραγωγικότης τοῦ Ἐργοστασίου Β εἶναι ἀνωτέρα τῆς τοιαύτης τοῦ Ἐργοσταοίου Α;

Ἐπὶ τῇ ὑποθέσει ὅτι αἱ παραγωγικότητες τῶν δύο Ἐργοστασίων εἶναι συγκρίσιμοι ἡ ἀκριβὴς ἐκτίμησις τῆς διακυμάνσεως καὶ ἡ τιμὴ t θὰ ἔχωσιν ὡς ἀκολούθως:

$$S^2 = \frac{36+52}{(8+10-2)} = 5{,}50$$

$$t = \frac{|\,13-14\,|}{\sqrt{5{,}50}} \sqrt{\frac{8\times10}{8+18}} = 0{,}89$$

Ὁ Πίναξ t δεικνύει ὅτι διὰ n—2=16 ἡ τιμὴ 0,89 δὲν εἶναι σημαντικὴ καὶ οὐδεμία λοιπὸν προσάγεται ἀπόδειξις ἐναντίον τῆς ὑποθέσεως περὶ τῆς ἰσότη-τος ἐν τῇ παραγωγικότητι τῶν δύο Ἐργοστασίων.

Ἔλεγχος t δύο Δειγμάτων βασιζόμενος ἐπὶ τοῦ Εὔρους.

Ὁ ἔλεγχος τῆς σημαντικότητος τῆς διαφορᾶς τῶν Μέσων διὰ τῆς χρησιμο-ποιήσεως ὡς ὑπολογισμοῦ τῆς σ, ὅπως βασίζεται αὕτη ἐπὶ τοῦ εὔρους τοῦ δεί-γματος, καθιερώθη κατὰ πρῶτον ὑπὸ τοῦ Lord (1947) ὡς ἓν εὔκολον ὑποκατά-στατον τοῦ ἐλέγχου t τοῦ Student εἰς τὴν περίπτωσιν κατὰ τὴν ὁποίαν τὰ δύο δείγματα εἶναι ἰσομεγέθη. Οὕτως ὁ Lord δίδει ἓξ ἐπίπεδα σημαντικότητος διὰ τὸν λόγον $\boxed{\kappa = \dfrac{|\,\bar{X}_1 - \bar{X}_2\,|}{\dfrac{\omega_1 + \omega_2}{2}}}$ ἔνθα ω_1 καὶ ω_2 εἶναι τὰ εὖρη τῶν δειγμάτων μεγέ-θους n ἐκ δύο κανονικῶν πληθυσμῶν ἐχόντων κοινὴν σ. Ἡ ὑπόθεσις μηδὲν δηλοῖ ὅτι οἱ Μέσοι τοῦ πληθυσμοῦ μ_1 μ_2 εἶναι ἴσοι.

Ὅταν ὅμως τὰ δείγματα εἶναι διαφόρου μεγέθους, δηλ. τὸ $n_1 \neq n_2$, χρησιμοποιοῦμεν τὸν λόγον $\boxed{\kappa = \dfrac{|\,\overline{X}_1 - \overline{X}_2\,|}{\omega_1 + \omega_2}}$, σχετικὸς δὲ Πίναξ δίδει τὴν τιμὴν κ ἀντιστοίχως πρὸς τέσσαρα διάφορα ἐπίπεδα σημαντικότητος α, χρησιμοποιούμενα εἰς ἕνα διπλῆς οὐρᾶς ἔλεγχον.

Παράδειγμα:

Δώδεκα παιδία ἡλικίας 12 ἐτῶν ἕως 12 ἐτῶν καὶ 3 μηνῶν ἐπελέγησαν κατὰ τύχην ἔκ τινος Σχολείου ἵνα ὑποβληθῶσιν εἰς δίαιταν διὰ παστεριωμένου γάλακτος ἐπὶ τετράμηνον. Τὸ ἀποκτηθὲν βάρος εἰς οὐγγίας μετὰ τὸ πέρας τοῦ τετραμήνου ἦτο

7, 17, 53, –2, 27, 41, 37, 35, 10, 12, 9, 38.

Ἕτερα ὀκτὼ παιδία τῆς αὐτῆς ἡλικίας ἐλήφθησαν ὡς ἔλεγχος ἄνευ διαίτης μὲ ἀποτέλεσμα εἰς βάρος κατὰ τὸ τέλος τοῦ 4μήνου

10, 0, 29, 11, – 21, 25, 19, – 19

Ἐρωτᾶται ἂν ἡ διὰ παστεριωμένου γάλακτος δίαιτα ἐπέφερε μεγαλυτέραν αὔξησιν βάρους μετὰ τὴν λῆξιν τοῦ 4μήνου.

$\overline{X}_1 = 23{,}669$	$\overline{X}_2 = 6{,}750$
$\Omega_1 = 55$	$\Omega_2 = 50$
$\Sigma\,(X_1 - \overline{X}_1)^2 = 3202{,}7$	$\Sigma\,(X_2 - \overline{X}_2)^2 = 2485{,}5$
$n_1 = 12 \quad \sigma_1 = 16{,}34$	$n_2 = 8 \quad \sigma_2 = 17{,}60$

Λόγος Ἐλέγχου κατὰ LORD $\quad u = \dfrac{23{,}667 - 6{,}750}{50 + 55} = 0{,}161$

Ἔλεγχος Student $\quad t = \dfrac{23{,}667 - 6{,}750}{\sqrt{\dfrac{3{,}202{,}7 + 2485{,}5}{20 - 2}}}\;\sqrt{\dfrac{12 \times 8}{12 + 8}} = 2{,}090$

Ἄρα ὁ μὲν Λόγος τοῦ LORD μᾶς δίδει 0,161 καὶ ἐν ἀναφορᾷ εἰς τὸν εἰδικὸν Πίνακα μὲ $n_1 = 12$ καὶ $n_2 = 8$ βλέπομεν ὅτι τὸ u πίπτει ἀκριβῶς κάτω τοῦ ἐπιπέδου σημαντικότητος $2\,^1/_2\,^0/_0$, ἐνῷ κατὰ τὸν Ἔλεγχον t τοῦ Student ἔχομεν εἰς $n - 2 = 18$ τιμὴν 2,090, ἥτις δὲν φθάνει ἀκριβῶς τὸ ἐπίπεδον $2\,^1/_2\,^0/_0$.

Κατ' ἀμφοτέρους τοὺς Ἐλέγχους δεικνύεται ὅτι ὑπάρχει μία αὔξησις εἰς τὸ μέσον βάρος τῶν ὑποβληθέντων εἰς τὴν δίαιταν παστεριωμένου γάλακτος μαθητῶν καὶ ὅτι αὕτη περίπου εἶναι σημαντικὴ ἐν τῷ ἐπιπέδῳ $2\,^1/_2\,^0/_0$.

ΒΙΒΛΙΟΓΡΑΦΙΑ

F. Mills, Statistical Methods (1938).

R. A. Fisher, Statistical Methods for Research Workers (1946).

H. Arkin - R. Colton, An Outline of Statistical Methods (1950).

Ε. Μαργαρίτη, Στατιστικὴ (1952).

F. Croxton - D. Gowden, Applied General Statistics (1953).

Κ. Ἀθανασιάδου, Στατιστικὴ (1953).

A. Monjallon, Introduction à la Methode Statistique (1954).

M. Slonim, Sampling in a Nutshell (1955).

P. G. Moore, The two sample t-test based on range εἰς Biometrica (Δεκ. 1957).