ΠΕΡΙ ΣΤΑΤΙΣΤΙΚΩΝ ΤΙΝΩΝ ΜΕΘΟΔΩΝ "ΕΛΕΥΘΕΡΩΝ ΚΑΤΑΝΟΜΗΣ,,

Ὑπὸ τοῦ κ. Ε. Α. ΖΟΥΛΙΑ

Γενικά. Ἡ «ἐκτίμησις» τῶν παραμέτρων τῶν πληθυσμῶν διὰ τυχαίας δειγματοληψίας καὶ ἡ εὕρεσις διαστημάτων ὡρισμένης ἐμπιστοσύνης τούτων, διὰ τῶν συνήθων στατιστικῶν μεθόδων(¹), προϋποθέτει (ἰδίᾳ εἰς τὰ δείγματα μικροῦ μεγέθους) γνῶσιν τῆς μορφῆς τῆς κατανομῆς τῆς παραμέτρου εἰς τὸν πληθυσμόν.

Π.χ. τὸ κριτήριον t τοῦ Student ἐφαρμόζεται ἐπὶ τῇ προϋποθέσει ὅτι ἡ κατανομὴ τῆς παραμέτρου εἰς τὸν πληθυσμὸν εἶναι κανονική.

Τὸ ἴδιον πρόβλημα παρουσιάζεται καὶ κατὰ τὴν κρίσιν στατιστικῶν ὑποθέσεων ἐπὶ παραμέτρου ἢ παραμέτρων ἐκ τῆς ἐκτιμήσεως τούτων ἐξ ἀντιστοίχων δειγμάτων.

Συμβαίνει ὅμως, οὐχὶ σπανίως, ἰδίως εἰς πειραματικὰς ἐρεύνας ἐπὶ νέων πεδίων τῆς ἐπιστήμης, νὰ ἀγνοεῖται ἡ μορφὴ τῆς κατανομῆς τῆς ἐρευνωμένης παραμέτρου καὶ τὸ «δεῖγμα» τῶν παρατηρήσεών μας νὰ εἶναι περιωρισμένου μεγέθους.

Διὰ τὴν ἀντιμετώπισιν τοιούτων περιπτώσεων παρετηρήθη τὴν τελευταίαν 20ετίαν μία ὤθησις τῶν Μαθηματικῶν - Στατιστικολόγων πρὸς τὴν ἀνάπτυξιν νέων μεθόδων, ἀνεξαρτήτων τῆς μορφῆς τῆς κατανομῆς τῶν παραμέτρων, διὰ τὴν ἐκτίμησιν τούτων ἢ τὴν κρίσιν στατιστικῶν ὑποθέσεων διὰ δειγματοληψίας.

Μερικαὶ ἐκ τῶν μεθόδων αὐτῶν, αἵτινες εἶναι γνωσταὶ ὑπὸ τὸ ὄνομα: «Μέθοδοι ἐλεύθεραι κατανομῆς» (Distribution - free Methods), θὰ ἐξετασθοῦν ἐν τοῖς ἑπομένοις, μὲ βάσιν κυρίως τὴν ἀκόλουθον βιβλιογραφίαν καὶ ἄρθρα:

1) S. S. Wilks: Mathematical statistics (1946).

2) A. M. Mood: «The theory of runs», Annals of Mathematical Statistics, vol. XI (1940) p. 367.

3) A. M. Mood: Introduction to Statistical Analysis (1950).

4) Dixon and Massey: Introduction to Statistical Analysis (1951).

5) A. C. Rosander: Elementary Principles of Statistics (1951).

6) A. C. Rosander: «The use of inversions as a test of random order», Journal of the American Statistical Association (1942) p. 352 - 358.

7) F. Swed and C. Eisenhart: «Tables for testing Randomness of

(¹) Τὸ κριτήριον τοῦ χ^2 εἶναι ἐφαρμόσιμον, ἀνεξαρτήτως τοῦ ἐὰν ἡ παράμετρος ἐν τῷ πληθυσμῷ ἔχει κανονικὴν ἢ μὴ κατανομήν· ἐν τούτοις εἰς τὴν ἐφαρμογήν του διὰ τὸν «ἔλεγχον τῆς καλῆς προσαρμογῆς» (goodness of fit) προϋποθέτει μόνον μεγάλα δείγματα, ἐνῷ ἀφ' ἑτέρου εἰς τὰς ἄλλας ἐφαρμογάς του προϋποθέτει μέγεθος δείγματος n=50 τουλάχιστον (ἴδε C. E. Weatherburn Mathematical Statistics (1952) σελὶς 185 καὶ C. U. Yule and M. C. Kendall «An introduction to the theory of Statistics (1950, σελὶς 469), ὡς καὶ ἄλλους τινὰς περιορισμούς.

grouping in a sequence of alternatives». Annals of Mathematical Statistics (1943) p.p. 66-87.

8) Johnson and Telley: Statistics (1950), volume II.

9) M. C. Kendall: ἓν ἄρθρον διὰ τὴν θεωρίαν τῶν Inversions (Biometrica (1938).

10) G. M. Dantzig: ἓν ἄρθρον διὰ τὴν θεωρίαν τῶν Inversions (Annals of Math. Statistics (1939) καὶ ἄλλα.

I. Ἐλεύθεραι κατανομῆς μέθοδοι βασιζόμεναι ἐπὶ διατεταγμένων δειγμάτων

Διατεταγμένον δεῖγμα (ordered sample).

Ἓν δεῖγμα μεγέθους n (ἀποτελούμενον ἐκ n στοιχείων) θὰ λέγεται «διατεταγμένον» ἐὰν κατατάξωμεν τὰ στοιχεῖα του οὐχὶ κατὰ τὴν χρονικὴν τάξιν καθ' ἣν ταῦτα ἐκλέγονται, ἀλλὰ κατὰ τάξιν ἀνιόντος μεγέθους.

Δυνάμεθα νὰ τὸ παραστήσωμεν διὰ τῶν τιμῶν $x_1, x_2, x_3 \ldots x_n$ τῆς μεταβλητῆς x, ὅπου $x_i \leqslant x_j$ ἐφ' ὅσον $i < j$

Ἐὰν ἡ μεταβλητὴ x εἶναι συνεχὴς μεταξὺ τῶν n τιμῶν τοῦ δείγματος, δὲν θὰ ἔχωμεν «συμπτώσεις», δηλαδὴ ἴσας τιμάς, διότι πιθ. $(x_i = x_j) = 0$ καὶ συνεπῶς θὰ ἔχωμεν $x_1 < x_2 < x_3 \ldots < x_n$.

Ὑποθέτομεν ὅτι ἡ μεταβλητή μας εἶναι συνεχὴς κατ' ἀρχὴν καὶ ἐν συνεχείᾳ θὰ ἴδωμεν πῶς πρέπει νὰ τροποποιήσωμεν τὰ συμπεράσματά μας εἰς περιπτώσεις «συμπτώσεων» ἐν τῷ δείγματι (δηλαδὴ εἰς περιπτώσεις ἀσυνεχῶν μεταβλητῶν).

Ὡς μέτρον θέσεως λαμβάνεται ἡ διάμεσος ν καὶ ὡς μέτρον διασπορᾶς ἡ ἐνδοτεταρτημοριακὴ ἀπόκλισις $Q_3 - Q_1$ (ἢ γενικώτερον ἡ ἀπόστασις δύο ἑτέρων ποσοστημορίων, π.χ. $\xi_{0.95} - \xi_{0.05}$, ὅπου $\xi_{0.95}$ καὶ $\xi_{0.05}$ εἶναι αἱ τιμαὶ τῆς μεταβλητῆς, αἵτινες εἶναι τοιαῦται ὥστε 95 % καὶ 5 % τῶν ὅλων τιμῶν τοῦ πληθυσμοῦ εἶναι μικρότεραι αὐτῶν ἀντιστοίχως).

Μία σημειακὴ ἐκτίμησις τῆς παραμέτρου ν εἶναι ἡ διάμεσος $\hat{\nu}$ τοῦ δείγματος, ἥτις εἶναι τὸ μέσον κατὰ μέγεθος στοιχεῖον εἰς τὸ διατεταγμένον δεῖγμα, ἐὰν τὸ n εἶναι περιττός, ἢ ὁ μέσος ὅρος τῶν δύο μέσων κατὰ μέγεθος στοιχείων, ὅταν τὸ n εἶναι ἄρτιον.

Διάστημα ἐμπιστοσύνης διὰ τὴν ν.

Διὰ νὰ εὕρωμεν ἓν διάστημα ἐμπιστοσύνης (confidence interval) διὰ τὴν διάμεσον ν, ἐργαζόμεθα ὡς ἑξῆς:

Ἡ πιθανότης ὅτι ἓν στοιχεῖον τοῦ δείγματος (οἱονδήποτε) θὰ ἐκλεγῇ μεταξὺ ἐκείνων ποὺ εὑρίσκονται ἀριστερὰ (ἢ δεξιὰ) τῆς διαμέσου ν τοῦ πληθυσμοῦ (δηλαδὴ θὰ εἶναι μικρότερον ἢ μεγαλύτερον τοῦ $\hat{\nu}$), εἶναι προφανῶς $\frac{1}{2}$.

Ὅλαι αἱ δυναταὶ περιπτώσεις ποὺ θὰ ἔχωμεν μετὰ τὴν ἐκλογὴν τῶν στοιχείων τοῦ τυχαίου δείγματος θὰ εἶναι: νὰ ἔχωμεν 0 ἢ 1 ἢ 2 ἢ . . . . . n στοιχεῖα μικρότερα τοῦ ν.

Ἡ πιθανότης νὰ ἔχωμεν 0 στοιχεῖα μικρότερα τοῦ ν (ἤ, ὅπερ τὸ αὐτό, καὶ τὰ n στοιχεῖα μεγαλύτερα τοῦ ν) εἶναι $\left(\frac{1}{2}\right)^n$, δεδομένου ὅτι αἱ τιμαὶ τοῦ δείγματος ὑποτίθενται ἀνεξάρτητοι ἀλλήλων.

Ἡ πιθανότης νὰ ἔχωμεν ἀκριβῶς 1 στοιχεῖον τοῦ δείγματος (ὡρισμένον κατὰ χρονικὴν τάξιν) μικρότερον τοῦ ν εἶναι $\left(\frac{1}{2}\right)^n$ ἀλλά, ἐπειδὴ τὸ γεγονὸς τοῦτο δύναται νὰ συμβῇ δι' οἱονδήποτε ἐκ τῶν n στοιχείων τοῦ δείγματος, ἡ πιθανότης νὰ ἔχωμεν ἀκριβῶς 1 στοιχεῖον τοῦ δείγματος (οἱονδήποτε κατὰ χρονικὴν τάξιν) μικρότερον τοῦ ν εἶναι $n\left(\frac{1}{2}\right)^n$

(Τοῦτο προφανῶς σημαίνει ὅτι τὸ ἐν λόγῳ στοιχεῖον θὰ εἶναι τὸ μικρότερον ὅλων εἰς τὸ διατεταγμένον δεῖγμα μας, ἤτοι τὸ x_1)

Ὅθεν: πιθανότης (ἵνα μόνον $x_1 < \nu$) $= n\left(\frac{1}{2}\right)^n$

Ὁμοίως ἡ πιθανότης νὰ ἔχωμεν ἀκριβῶς 2 στοιχεῖα (οἱονδήποτε ζεῦγος κατὰ χρονικὴν τάξιν ἐκλογῆς τῶν στοιχείων) μικρότερα τοῦ ν εἶναι $\frac{n(n-1)}{2}\left(\frac{1}{2}\right)^n$ ἤτοι: (1)

$$\text{πιθανότης (ἵνα μόνον } x_1 \text{ καὶ } x_2 < \nu) = \binom{n}{2}\left(\frac{1}{2}\right)^n$$

Καὶ γενικῶς

ἡ πιθανότης νὰ ἔχωμεν ἀκριβῶς i στοιχεῖα (οἱαδήποτε χρονικῶς) μικρότερα τοῦ ν εἶναι $\binom{n}{i}\left(\frac{1}{2}\right)^n$

ἤτοι πιθανότης (ἵνα μόνον $x_1, x_2, x_3 \ldots$ καὶ $x_i < \nu$) $= \binom{n}{i}\left(\frac{1}{2}\right)^n$

Ἐξ ἄλλου ἡ πιθανότης ἵνα τὸ x_r, δηλαδὴ τὸ κατέχον τὴν τάξιν r στοιχεῖον τοῦ διατεταγμένου δείγματος, εἶναι μεγαλύτερον τοῦ ν εἶναι ἡ κάτωθι:

$$\text{Πιθ. } (x_r > \nu) = \text{πιθ. (ἵνα μόνον } x_1 < \nu) + \text{πιθ. (ἵνα μόνον } x_1 \text{ καὶ } x_2 < \nu) + \ldots + \text{πιθ. (ἵνα μόνον } x_1, x_2, x_3 \ldots\ldots x_{r-1} < \nu)$$

διότι εἰς ὅλας τὰς ἀνωτέρω περιπτώσεις τὸ r κατὰ τάξιν στοιχεῖον τοῦ διατεταγμένου δείγματος θὰ εἶναι προφανῶς μεγαλύτερον τοῦ ν.

ὅθεν

$$\text{Πιθ } (x_r > \nu) = \sum_{i=0}^{r-1} \binom{n}{i}\left(\frac{1}{2}\right)^n \qquad (3)$$

Ὁμοίως εὑρίσκομεν ὅτι:

$$\text{Πιθ } (x_s < \nu) = \sum_{i=s}^{n} \binom{n}{i}\left(\frac{1}{2}\right)^n \qquad (4)$$

(1) Διὰ τοῦ συμβόλου $\binom{n}{2}$ ἐννοεῖται ὁ ἀριθμὸς τῶν συνδυασμῶν τῶν n πραγμάτων ἀνὰ 2.

Αἱ ἰσότητες (3) καὶ (4) εἶναι ἀληθεῖς δι' οἱανδήποτε τιμὴν τοῦ r καὶ s μικροτέραν τοῦ n.

Ἐὰν νῦν ὑποθέσωμεν ὅτι s καὶ r εἶναι μικρότερα τοῦ n καὶ ὅτι $r<s$ τότε

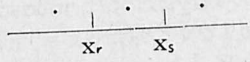
$$\text{πιθ.}\ (x_r > \nu) + \text{πιθ.}\ (x_s < \nu) + \text{πιθ.}\ (x_r < \nu < x_s) = 1 \qquad (5)$$

διότι, ἐφ' ὅσον $s>r$ καὶ κατὰ συνέπειαν $x_s > x_r$ τρία γεγονότα εἶναι δυνατά:

1) $x_r > \nu$ ὁπότε φυσικὰ καὶ $x_s > \nu$

2) $x_s < \nu$ ὅτε φυσικὰ καὶ $x_r < \nu$

3) $x_r < \nu < x_s$



Ἑπομένως ἡ (5) εἶναι ἀληθὴς καὶ ἐκ ταύτης ἔχομεν :

$$\text{πιθ.}\ (x_r < \nu < x_s) = 1 - \text{πιθ.}\ (x_r > \nu) - \text{πιθ.}\ (x_s < \nu)$$

ἀλλὰ

$$1 - \text{πιθ.}\ (x_r > \nu) - \text{πιθ.}\ (x_s < \nu) = 1 - \sum_{i=0}^{r-1} \binom{n}{i} \left(\frac{1}{2}\right)^n - \sum_{i=s}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n =$$

$$= \sum_{i=r}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n - \sum_{i=s}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \sum_{i=r}^{s-1} \binom{n}{i} \left(\frac{1}{2}\right)^n$$

Ὅθεν

$$\text{πιθ.}\ (x_r < \nu < x_s) = \sum_{i=r}^{s-i} \binom{n}{i} \left(\frac{1}{2}\right)^n \qquad (6)$$

Ἐὰν ἤδη τὸ δεξιὰ μέλος τῆς (6) εἶναι π.χ. >0.75 δι' ὡρισμένας τιμὰς τῶν r καὶ s, τὸ διάστημα x_r ἕως x_s ἀποτελεῖ ἓν διάστημα ἐμπιστοσύνης διὰ τὴν διάμεσον τοῦ πληθυσμοῦ ν μεγέθους ἴσου πρὸς τὴν ἀντίστοιχον πιθανότητα.

Συνήθως λαμβάνομεν $s = n - r + 1$ οὕτως ὥστε τὰ x_r καὶ x_s νὰ εἶναι ἰσαπέχοντα ἐκ τῶν ἄκρων στοιχεῖα (κατὰ τάξιν).' Ἐὰν π.χ. ἔχωμεν ἓν δεῖγμα μεγέθους 6, τὸ $x_1, x_2, x_3, x_4, x_5, x_6$ (διατεταγμένον), θὰ ἔχωμεν ἐκ τῆς ἰσότητος (6)

$$\text{πιθ.}\ (x_1 < \nu < x_6) = \frac{6}{64} + \frac{15}{64} + \frac{20}{64} + \frac{15}{64} + \frac{6}{64} = \frac{31}{32} = \frac{97}{100} \ \text{περίπου.}$$

Λέγομεν ὅθεν ὅτι τὸ διάστημα x_1 ἕως x_6 ἀποτελεῖ ἓν 97% διάστημα ἐμπιστοσύνης διὰ τὸ ν.

Τοῦτο σημαίνει ὅτι τὸ μεταβλητὸν διάστημα x_1 ἕως x_6 (μεταβλητὸν διὰ διάφορα δείγματα μεγέθους 6) καλύπτει τὴν ἀληθῆ διάμεσον ν τοῦ πληθυσμοῦ εἰς τὰ 97 % τῶν περιπτώσεων (δηλαδὴ εἰς τὰ 97 % ἑνὸς μεγάλου πλήθους τοιούτων δειγμάτων $n = 6$).

Ἐὰν λοιπὸν θεωρήσωμεν ἓν ὡρισμένον τυχαῖον δεῖγμα (ὅτε φυσικὰ τὰ x_1 καὶ x_6 ἔχουν ὡρισμένας σταθερὰς τιμὰς) εἴμεθα δικαιολογημένοι νὰ ἔχωμεν ἕνα μεγάλον βαθμὸν ἐμπιστοσύνης (97 %) εἰς τὸ συμπέρασμα ὅτι τὸ ἄγνωστον ν τοῦ πληθυσμοῦ θὰ εὑρίσκεται μεταξὺ τῶν x_1 καὶ x_6 τοῦ δείγματος τούτου.

Πίνακες διὰ διαστήματα ἐμπιστοσύνης διὰ τὴν διάμεσον ἔχουν ὑπολογισθῆ
Nair, K. R. «Table of confidence interval for median in samples from any continuous population» Sankhyâ vol. 4 (1940) p.p. 551 - 558 ἀναδημο-

σιευόμεναι εἰς Introduction to Statistical analysis: W. Dixon and F. Massey ⟨New York Mc Graw - Hill 1951⟩ table 25 p. 360.

Οἱ πίνακες δίδουν διὰ διαστήματα ἐμπιστοσύνης $\geqslant 95\%$ καὶ $\geqslant 99\%$ τὰς μεγαλυτέρας τιμὰς τοῦ r ἵνα ἡ διάμεσος τοῦ πληθυσμοῦ εὑρίσκεται μεταξὺ τῶν x_r καὶ x_{n-r+1} εἰς ἓν δεῖγμα διατεταγμένον, διὰ τιμὰς τοῦ n ἀπὸ 6 ἕως 65. Π.χ. ἐὰν $n = 20$ ὁ πίναξ τοῦ Nair δεικνύει ὅτι $r = 6$ καὶ $n - r + 1 = 15$ καὶ ὅτι ἡ ἐμπιστοσύνη τοῦ διαστήματος x_6 ἕως x_{15}, διὰ τὸ ν, εἶναι 95,9%.

Δοκιμασίαι ὑποθέσεων (tests of Hypotheses) περὶ μιᾶς παραμέτρου

Ἔστω ν_0 δοθεὶς ἀριθμὸς ὅστις, κατὰ τὴν γνώμην μας, εἶναι ἐγγὺς τῆς ἀληθοῦς ἀλλ' ἀγνώστου τιμῆς τῆς διαμέσου τοῦ πληθυσμοῦ ν.

Διὰ νὰ κρίνωμεν τὴν ὑπόθεσιν μηδὲν (H_0) ὅτι $\nu = \nu_0$ εἰς ἓν ἐπίπεδον πιθανότητος π. χ. 5%, ἐργαζόμεθα ὡς ἑξῆς :

α) Εἰς περίπτωσιν ἀμφιπλεύρου (ἢ δύο ἄκρων) δοκιμασίας (two sided test) (δηλαδὴ ὅταν ὡς ἐναλλακτέον τῆς ὑποθέσεως μηδὲν(H_0) θεωροῦμεν ἀμφότερα τὰ γεγονότα $\nu < \nu_0$ ἢ $\nu > \nu_0$), ἀρκεῖ νὰ κατασκευάσωμεν ἓν διάστημα ἐμπιστοσύνης 95 % διὰ τὴν ἄγνωστον ν (ἀντίστοιχον πρὸς μέγεθος δείγματος n) καὶ ἐὰν π.χ. τοῦτο εἶναι τὸ x_λ ἕως $x_{n+1-\lambda}$ νὰ ἐκλέξωμεν ἓν τυχαῖον δεῖγμα μεγέθους n καὶ νὰ εὕρωμεν, εἰς τὸ ἀντίστοιχον διατεταγμένον δεῖγμα, τὰς τιμὰς τοῦ x_λ καὶ $x_{n+1-\lambda}$ καὶ ἂν $x_\lambda < \nu_0 < x_{n+1-\lambda}$, δὲν ἀπορρίπτομεν τὴν ὑπόθεσιν εἰς τὸ θεωρηθὲν ἐπίπεδον πιθανότητος 5% ἀλλ' ἀποφαινόμεθα ὅτι τὰ ἀποτελέσματα τοῦ δείγματος δὲν εὑρίσκονται εἰς ἀσυμφωνίαν μὲ τὴν ὑπόθεσιν $\nu = \nu_0$, ἐνῷ ἐὰν τὸ ν_0 εἶναι ἐκτὸς τοῦ διαστήματος x_λ ἕως $x_{n+1-\lambda}$, ἀπορρίπτομεν τὴν H_0 εἰς τὸ ἐπίπεδον 5 %.

β) Εἰς περίπτωσιν μονοπλεύρου (ἢ ἑνὸς ἄκρου) δοκιμασίας καθ' ἣν θέλομεν νὰ κρίνωμεν τὴν H_0 μὲ ἐναλλακτέαν περίπτωσιν μόνον τὴν $\nu_0 < \nu$ εἰς τὸ 5% π.χ. ἐπίπεδον σημαντικότητος, ἐργαζόμεθα ὡς ἑξῆς:

Εὑρίσκομεν ἕνα ἀκέραιον $r = r_0$ τοιοῦτον ὥστε τὸ δεξιὸν μέλος τῆς σχέσεως

(3) δηλαδὴ τῆς $\text{Πιθ}\,(x_r > \nu) = \sum_{i=0}^{r-1} \binom{n}{i} \left(\frac{1}{2}\right)^n$ νὰ ἔχῃ τιμὴν ἴσην ἢ μικροτέραν τοῦ 5% ἀλλὰ ὅσον τὸ δυνατὸν ἐγγύτερον ταύτης (μὲ ἄλλους λόγους εὑρίσκομεν τὴν μεγαλυτέραν τιμὴν τοῦ r δι' ἣν ἡ $\text{Πιθ}\,(x_r > \nu) \leqslant 5\%$).

Ἔστω ὅτι εὕρομεν τοιοῦτον ἀκέραιον τὸν $r = r_0$.

Λαμβάνομεν ἓν δεῖγμα τοῦ καθορισθέντος μεγέθους n· ἂν εὕρωμεν εἰς τὸ δεῖγμα αὐτὸ ὅτι $x_{r=r_0} > \nu_0$ ἡ ὑπόθεσις μηδὲν $\nu = \nu_0$ ἀπορρίπτεται, καθ' ὅσον ἡ πιθανότης νὰ συμβῇ. τοῦτο ἐκ τῆς διακυμάνσεως τῆς τυχαίας δειγματοληψίας εἶναι $\leqslant 5\%$.

Ἐὰν ὅμως $\nu_0 > x_{r=r_0}$ τότε δὲν δυνάμεθα νὰ ἀπορρίψωμεν τὴν ὑπόθεσιν μηδὲν $\nu = \nu_0$ εἰς τὸ 5 % ἐπίπεδον πιθανότητος (ἢ σημαντικότητος).

γ) Τέλος εἰς περίπτωσιν μονοπλεύρου δοκιμασίας μὲ ἐναλλακτέαν περίπτωσιν μόνον τὴν $\nu_0 > \nu$, ἐργαζόμεθα ὡς ἑξῆς:

Λαμβάνομεν τὴν σχέσιν (4), δηλαδὴ τὴν σχέσιν :

$$\Pi\iota\theta\ (x_s < \nu) = \sum_{i=s}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n$$

διὰ τὸ δοθὲν n καὶ εὑρίσκομεν ἕνα $s = s_0$ ὥστε :

Πιθ ($x_{s=s_0} < \nu$) $\leqslant 5\,\%$ (εἰς περίπτωσιν ἀνισότητος τὸ εὑρεθὲν $s = s_0$ δέον νὰ εἶναι τοιοῦτον ὥστε ἡ πιθ ($x_{s=s_0} < \nu$) νὰ εἶναι μὲν μικροτέρα τοῦ 5 % ἀλλὰ ὅσον τὸ δυνατὸν πλησιέστερον τούτου, δηλαδὴ εὑρίσκομεν τὴν μικροτέραν δυνατὴν τιμὴν τοῦ s δι' ἣν πιθ. ($x_s < \nu$) $\leqslant 5\,\%$).

Ἂν τώρα λάβωμεν ἓν τυχαῖον δεῖγμα καὶ εὕρωμεν $x_{s_0} < \nu_0$ ἀπορρίπτομεν τὴν ὑπόθεσιν μηδέν. Ἂν ὅμως $\nu_0 < x_{s_0}$ τότε δὲν ἀπορρίπτομεν τὴν ὑπόθεσιν $\nu = \nu_0$ εἰς τὸ 5 % ἐπίπεδον σημαντικότητος (ἢ πιθανότητος). Δηλαδὴ τότε λέγομεν ὅτι τὸ ἀποτέλεσμα τοῦ δείγματος δὲν εὑρίσκεται εἰς ἀσυμφωνίαν πρὸς τὴν ὑπόθεσιν $\nu = \nu_0$ εἰς τὸ 5 % ἐπίπεδον σημαντικότητος.

Περίπτωσις ἀσυνεχοῦς μεταβλητῆς

Εἰς τὴν περίπτωσιν ἀσυνεχοῦς μεταβλητῆς εἰς τὸ δεῖγμα μας δυνατὸν νὰ ἔχωμεν συμπτώσεις.

Τότε ἡ σχέσις (6), δηλαδὴ ἡ σχέσις

$$\Pi\iota\theta\ (x_r < \nu < x_s) = \sum_{i=r}^{s-1} \binom{n}{i} \left(\frac{1}{2}\right)^n$$

πρέπει νὰ ἀντικατασταθῇ διὰ τῆς σχέσεως

$$\Pi\iota\theta\ (x_r < \nu < x_s) \geq \sum_{i=r}^{s-1} \binom{n}{i} \left(\frac{1}{2}\right)^n \tag{7}$$

διότι ἂν μὲν εἰς τὸ δεῖγμα μας συμβῇ τὰ $x_1, x_2, x_3 \ldots x_n$ νὰ εἶναι διάφορα, ἔχομεν πάλιν τὴν (6) ἀμετάβλητον· ἂν ὅμως π. χ. (ἐπὶ δείγματος μεγέθους $n=6$) ἔχωμεν $x_1 = x_2$ καὶ $x_5 = x_6$, τότε τὸ διάστημα x_2 ἕως x_5 εἶναι τὸ ἴδιον μὲ τὸ διάστημα x_1 ἕως x_6, ἀλλὰ πιθ ($x_1 < \nu < x_6$) $= \frac{97}{100}$ ἐνῷ πιθ ($x_2 < \nu < x_5$) $= \frac{78}{100}$, ὡς εὐκόλως εὑρίσκομεν ἐκ τῆς παραστάσεως $\sum_{i=r}^{s-1} \binom{n}{i} \left(\frac{1}{2}\right)^n$ ἂν θέσωμεν $r=1$, $s=6$ ἢ $r=2$ καὶ $s=5$ καὶ $n=6$.

Ὅθεν εἰς τὴν περίπτωσιν ἀσυνεχοῦς μεταβλητῆς, καθ' ἣν ἐν τῷ δείγματι $x_1 = x_2$ καὶ $x_5 = x_6$, τὸ διάστημα x_2 ἕως x_5 ἀντιστοιχεῖ εἰς ἐπίπεδον πιθανότητος οὐχὶ πράγματι 78 % ἀλλὰ 97 % ἀφοῦ εἶναι τὸ ἴδιον μὲ τὸ διάστημα x_1 ἕως x_6.

Ἀναλόγως, συνεπῶς, τῆς παρουσιαζομένης ἐν τῷ δείγματι περιπτώσεως (ἀριθμοῦ συμπτώσεων κλπ.) δυνάμεθα διὰ καταλλήλων τροποποιήσεων τῶν ἐπιπέδων πιθανότητος νὰ ἔχωμεν ἀνάλογα συμπεράσματα πρὸς τὰ ἤδη ἐκτεθέντα διὰ τὴν περίπτωσιν τῶν συνεχῶν μεταβλητῶν.

Σύγκρισις δύο πληθυσμῶν.

Ἐὰν δύο πληθυσμοὶ ἔχωσι τὴν ἰδίαν διάμεσον, ὅτε λέγομεν ὅτι ἔχουσι τὴν αὐτὴν θέσιν, εἶναι δυνατὸν νὰ ἔχωσι κατανομὴν διαφόρου μορφῆς (σχήματος).

Ἐὰν ὅμως, ὅπως συνήθως συμβαίνει εἰς τὴν πρᾶξιν, ἐνδιαφερώμεθα νὰ κρίνωμεν ἁπλῶς ἐὰν ἔχωσι τὴν αὐτὴν διάμεσον, δηλαδὴ ἂν $\nu_1=\nu_2$, ἐργαζόμεθα ὡς ἑξῆς:

Ἔστω ὅτι $x_1, x_2, x_3 \ldots x_{n_1}$ εἶναι ἓν διατεταγμένον δεῖγμα μεγέθους n_1 ἐξ ἑνὸς πληθυσμοῦ A καὶ $y_1, y_2, y_3 \ldots y_{n_2}$ ἓν δεύτερον δεῖγμα μεγέθους n_2 ἐξ ἑνὸς πληθυσμοῦ B.

Σχηματίζομεν ἓν μικτὸν διατεταγμένον δεῖγμα ἐκ τῶν n_1 τιμῶν τοῦ πρώτου δείγματος καὶ τῶν n_2 τιμῶν τοῦ δευτέρου. Θὰ ἔχωμεν οὕτω ἓν τρίτον δεῖγμα $z_1, z_2, z_3 \ldots z_{n_1+n_2}$ μεγέθους n_1+n_2 στοιχείων. Ἔστω ὅτι ν_0 εἶναι ἡ διάμεσος εἰς τὸ μικτὸν δεῖγμα.

Μετροῦμεν τὸν ἀριθμὸν τῶν στοιχείων x, ἔστω m_1, τὰ ὁποῖα εἶναι μεγαλύτερα τοῦ ν_0 (εἰς τὸ μικτὸν διατεταγμένον δεῖγμα) καὶ τὸν ἀριθμὸν m_2 τῶν στοιχείων y ποὺ ἔχουν τὴν αὐτὴν ἰδιότητα.

Ἐὰν ἡ ὑπόθεσις μηδὲν H_0 $(\nu_1=\nu_2)$ εἶναι ἀληθής, θὰ ἀναμένωμεν (κατὰ μέσον ὅρον) ὅτι $m_1=\dfrac{n_1}{2}$ καὶ $m_2=\dfrac{n_2}{2}$

Ζητοῦμεν τὴν κατανομὴν τῶν m_1 καὶ m_2 ὅταν ἡ H_0 εἶναι ἀληθής.

Ἔστω ὅτι τὰ δείγματα τῶν δύο πληθυσμῶν κατὰ χρονικὴν τάξιν εἶναι:

δεῖγμα A: $\alpha_1, \alpha_2, \alpha_3 \ldots \alpha_{n_1}$

» B: $\beta_1, \beta_2, \beta_3 \ldots \beta_{n_2}$,

ὅπου ὑπετέθη ὅτι A ἐκαλέσαμεν τὸ μικρότερον δεῖγμα, δηλαδὴ ὅτι $n_1<n_2$ (εἰς περίπτωσιν καθ' ἣν $n_1\neq n_2$, ἐνῷ εἰς περίπτωσιν καθ' ἣν $n_1=n_2$ καλῶ A οἱονδήποτε ἐκ τῶν δύο δειγμάτων).

Περίπτωσις 1η. Ἔστω ὅτι $n_1+n_2=$ ἄρτιος.

Τότε τὸ ν_0 κεῖται μεταξὺ τῶν $\dfrac{n_1+n_2}{2}$ καὶ $\left(\dfrac{n_1+n_2}{2}+1\right)$ στοιχείων τοῦ μικτοῦ δείγματος.

Ὅθεν ἐὰν $m_1=0$ τὸ m_2 πρέπει νὰ εἶναι $\dfrac{n_1+n_2}{2}$

» $m_1=1$ » » $\dfrac{n_1+n_2}{2}-1$

$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$

καὶ ἐὰν $m_1=n_1$ » » $\dfrac{n_1+n_2}{2}-n_1$

διότι εἰς τὸ μικτὸν διατεταγμένον δεῖγμα ἡ διάμεσος ν_0 διαιρεῖ τὰ (n_1+n_2) στοιχεῖα τοῦ δείγματος εἰς δύο ἴσα μέρη καὶ συνεπῶς πάντοτε πρέπει νὰ ἔχωμεν $\dfrac{n_1+n_2}{2}$ στοιχεῖα (x ἢ y) μεγαλύτερα τοῦ ν_0,

θὰ ἔχωμεν:

$$\text{πιθ}\left(m_1=0,\ m_2=\frac{n_1+n_2}{2}\right)=\frac{\dbinom{n_1}{0}\dbinom{n_2}{\frac{n_1+n_2}{2}}}{\dbinom{n_1+n_2}{\frac{n_1+n_2}{2}}}$$

διότι τὰ $\frac{n_1+n_2}{2}$ στοιχεῖα τοῦ μικτοῦ δείγματος, τὰ ὁποῖα ἠμπορεῖ νὰ εἶναι μεγαλύτερα τοῦ ν_0 ἠμποροῦν νὰ ἐκλεγοῦν κατὰ $\binom{n_1+n_2}{\frac{n_1+n_2}{2}}$ ἐν συνόλῳ τρόπους μεταξὺ τῶν $\alpha_1, \alpha_2 \dots \alpha_{n_1}, \beta_1, \beta_2 \dots \beta_{n_2}$ στοιχείων.

Αἱ εὐνοϊκαὶ δὲ περιπτώσεις ἵνα ἔχωμεν 0 ἐκ τῶν $\alpha_1, \alpha_2 \dots \alpha_{n_1}$ καὶ $\frac{n_1+n_2}{2}$ ἐκ τῶν β εἶναι τὸ γινόμενον $\binom{n_1}{0}\binom{n_2}{\frac{n_1+n_2}{2}}$,

Ὁμοίως:

$$\text{πιθ}\left(m_1=1,\ m_2=\frac{n_1+n_2}{2}-1\right)=\frac{\binom{n_1}{1}\binom{n_2}{\frac{n_1+n_2}{2}-1}}{\binom{n_1+n_2}{\frac{n_1+n_2}{2}}}$$

καὶ γενικῶς

$$\text{πιθ}\,(m_1=m_1, m_2=m_2)=\frac{\binom{n_1}{m_1}\binom{n_2}{m_2}}{\binom{n_1+n_2}{\frac{n_1+n_2}{2}}}=\frac{\binom{n_1}{m_1}\binom{n_2}{m_2}}{\binom{n_1+n_2}{\alpha_1}} \qquad (8)$$

ὅπου $\frac{n_1+n_2}{2}=\alpha_1$ καὶ $m_1+m_2=\alpha_1$ · ἢ $m_2=\frac{n_1+n_2}{2}-m_1$

Περίπτωσις 2α: $n_1+n_2=$ περιττὸς

Εἰς τὴν περίπτωσιν ταύτην ν_0 εἶναι τὸ $\frac{n_1+n_2+1}{2}$ κατὰ τάξιν στοιχεῖον τοῦ μικτοῦ δείγματος, δηλαδὴ τὸ $z_{\frac{n_1+n_2+1}{2}}$

Ἑπομένως:

Ἐὰν $m_1=o$ πρέπει $m_2=\frac{n_1+n_2-1}{2}$

Ἐὰν $m_1=1$ » $m_2=\frac{n_1+n_2-1}{2}-1$

⋮ ⋮ ⋮

καὶ ἐὰν $m_1=n_1$ » $m_2=\frac{n_1+n_2-1}{2}-n_1$

διότι εἰς τὸ μικτὸν δεῖγμα θὰ ἔχωμεν τώρα πάντοτε $\frac{n_1+n_2-1}{2}$ στοιχεῖα μεγαλύτερα τοῦ ν_0.

Καὶ

$$\text{πιθ}\left(m_1=o,\ m_2=\frac{n_1+n_2-1}{2}\right)=\frac{\binom{n_1}{0}\binom{n_2}{\frac{n_1+n_2-1}{2}}}{\binom{n_1+n_2}{\frac{n_1+n_2-1}{2}}}$$

Ὁμοίως

$$\pi\iota\theta\left(m_1=1,\ m_2=\frac{n_1+n_2-1}{2}-1\right)=\frac{\binom{n_1}{1}\binom{n_2}{\frac{n_1+n_2-1}{2}-1}}{\binom{n_1+n_2}{\frac{n_1+n_2-1}{2}}}$$

Καὶ γενικῶς

$$\pi\iota\theta\left(\mathbf{m}_1=m_1,\ m_2=m_2\right)=\frac{\binom{n_1}{m_1}\binom{n_2}{m_2}}{\binom{n_1+n_2}{\frac{n_1+n_2-1}{2}}}=\frac{\binom{n_1}{m_1}\binom{n_2}{m_2}}{\binom{n_1+n_2}{\alpha_1-\frac{1}{2}}} \tag{9}$$

ὅπου $m_1+m_2=\dfrac{n_1+n_2-1}{2}=\dfrac{n_1+n_2}{2}-\dfrac{1}{2}=\alpha_1-\dfrac{1}{2}$

Οὕτω ἡ μηδενικὴ ὑπόθεσις $\nu_1=\nu_2$, δηλαδὴ ὅτι οἱ δύο πληθυσμοὶ ἔχουν τὴν ἰδίαν διάμεσον, δύναται νὰ κριθῇ ἐκ τοῦ μεγέθους τῆς πιθανότητος (8) ἢ (9) ἀναλόγως τοῦ ἀριθμοῦ n_1+n_2 (δηλαδὴ ἂν $n_1+n_2=$ περιττὸς ἢ ἄρτιος) καὶ τῆς πιθανότητος περιπτώσεων ἔτι περισσότερον ἀπεχουσῶν τῶν ἀναμενομένων τιμῶν (expected values) τῶν m_1 καὶ m_2.

Παράδειγμα 1ον

Δύο δείγματα, Α μεγέθους 5 καὶ Β μεγέθους 7, ἀναμιγνυόμενα ἀποτελοῦν ἓν μικτὸν διατεταγμένον δεῖγμα ἐκ 12 στοιχείων, εἰς ὃ εὑρίσκομεν ὅτι ὑπάρχουν 1 στοιχεῖον ἐκ τοῦ Α καὶ 5 στοιχεῖα ἐκ τοῦ Β δείγματος μεγαλύτερα τοῦ ν_0 (διαμέσου τοῦ μικτοῦ διατεταγμένου δείγματος).

Ζητεῖται νὰ κριθῇ ἡ H_0 ($\nu_1=\nu_2$) εἰς τὸ 5 % ἐπίπεδον σημαντικότητος.

Ἐὰν τὰ δύο δείγματα προήρχοντο ἀπὸ πληθυσμοὺς μὲ τὴν ἰδίαν διάμεσον (δηλαδὴ ἐὰν H_0 ἀληθής), τότε ἐκ τῆς (8) εὑρίσκομεν

$$\pi\iota\theta\ (m_1=1, m_2=5)=\frac{\binom{5}{1}\binom{7}{5}}{\binom{12}{6}}=\frac{105}{924}$$

$$\pi\iota\theta\ (m_1=0, m_2=6)=\frac{7}{924}$$

$$\pi\iota\theta\ (m_1=5, m_2=1)=\frac{7}{924}$$

$$\pi\iota\theta\ (m_1=4, m_2=2)=\frac{105}{924}$$

Ὅθεν, ἡ πιθανότης νὰ ἔχωμεν «κατὰ τύχην» ἓν γεγονὸς ὡς τὸ ἐμφανισθὲν εἰς τὸ μικτὸν διατεταγμένον δεῖγμα ἢ χειρότερον (ἢ σπανιώτερον, δυνάμεθα μᾶλλον νὰ εἴπωμεν) εἶναι τὸ ἄθροισμα τῶν ὡς ἄνω πιθανοτήτων, ἤτοι περίπου $^{24}/_{100}$. Ἡ πιθανότης αὕτη εἶναι μεγαλυτέρα τοῦ 5 % καὶ ἑπομένως ἡ παρουσιασθεῖσα εἰς τὸ μικτὸν διατεταγμένον δεῖγμα διαφορὰ (ἀπόκλισις) τιμῶν τῶν $m_1=1$ καὶ $m_2=5$ ἐκ τῶν ἀναμενομένων $m_1=2{,}5$ καὶ $m_2=3{,}5$ δὲν εἶναι στατιστικῶς σημαντικὴ εἰς τὸ 5 % ἐπίπεδον σημαντικότητος καὶ δὲν δυνάμεθα νὰ ἀπορρίψωμεν τὴν ὑπόθεσιν μηδὲν (δηλαδὴ ὅτι $\nu_1=\nu_2$).

'Εὰν ὅμως εἰς τὸ ἴδιον παράδειγμα εἴχομεν ὅτι εἰς τὸ μικτὸν διατεταγμένον δεῖγμα καὶ τὰ 6 στοιχεῖα, τὰ μεγαλύτερα τῆς διαμέσου του ν_0 προήρχοντο ἀπὸ τὸ Β δεῖγμα, τότε περιπτώσεις τόσον σπάνιαι ὡς ἡ $m_1=0$ καὶ $m_2=6$ εἶναι μόνον ἡ $m_1=5$ καὶ $m_2=1$ (διότι ἡ ἀναμενομένη τιμὴ τοῦ $m_1=2{,}5$ καὶ τοῦ $m_2=3{,}5$, ὅταν δὲ $m_1=0$ καὶ $m_2=6$ ἔχομεν ἀπόκλισιν ἀπὸ τὰς ἀναμενομένας τιμὰς 2,5, ἀλλὰ καὶ εἰς τὴν περίπτωσιν $m_1=5$ καὶ $m_2=1$ πάλιν ἡ ἀπόκλισις ἀπὸ τὰς «ἀναμενομένας» τιμὰς εἶναι 2,5).

Ἔχομεν: $\pi\iota\theta(m_1=0, m_2=6)=\frac{7}{924}$ (ἐκ τῆς (8) ἐφ' ὅσον $n_1+n_2=$ ἄρτιος)

καὶ $\pi\iota\theta\,(m_1=5,\ m_2=1)=\frac{7}{924}$.

Ἄρα $\pi\iota\theta(m_1=0,\ m_2=6$ ἢ $m_1=5$ καὶ $m_2=1)=\frac{14}{924}$, ἤτοι περίπου 1,5%.

Ἄρα, εἰς τοιαύτην περίπτωσιν τὸ ἀποτέλεσμα θὰ ἦτο στατιστικῶς σημαντικὸν καὶ ἀπορρίπτομεν τὴν ὑπόθεσιν μηδὲν $(\nu_1=\nu_2)$ εἰς τὸ 5% ἐπίπεδον σημαντικότητος, ἐφ' ὅσον γεγονότα ὡς τὸ ἐμφανισθὲν ἢ χειρότερα ἔχουν πιθανότητα 1,5%, δηλαδὴ μικροτέραν τοῦ 5%, νὰ ὀφείλωνται εἰς τὴν τύχην.

2. 'Ελεύθεραι κατανομῆς μέθοδοι βασιζόμεναι ἐπὶ τῆς θεωρίας τῶν «διαδρομῶν» ἢ ὁμοειδῶν ὁμάδων (runs).

'Ορισμοί. Ἔστω ὅτι ἔχομεν μίαν ἀκολουθίαν Ν στοιχείων ἀποτελουμένην ἐξ ἐμφανίσεων κατὰ οἱανδήποτε τάξιν n γεγονότων ἀμοιβαίως ἀποκλειομένων (mutually exclusive) ὅπου $N \geqslant n$.

Μία τοιαύτη ἀκολουθία π.χ. εἶναι ἡ ΑΓΓΓΓΒΒΑΑΒΒΒΒΒΒ (10)

ἧς ὁ ἀριθμὸς τῶν στοιχείων $N=15$ καὶ τὰ διάφορα γεγονότα $n=3$ (τὰ Α, Β καὶ Γ).

Καλοῦμεν διαδρομὴν (run) κάθε ἀκολουθίαν ἑνὸς ἢ περισσοτέρων ὁμοίων γεγονότων, περιλαμβανομένην εἰς τὴν ἀρχικὴν ἀκολουθίαν τῶν Ν στοιχείων, τῆς ὁποίας προηγοῦνται ἢ ἕπονται (ἢ προηγοῦνται καὶ ἕπονται) διαφορετικὰ πρὸς ταύτην γεγονότα (ἢ οὐδὲν γεγονός). Π.χ. εἰς τὴν ἀκολουθίαν (10) ἔχομεν 5 διαδρομάς, τὰς Α, ΓΓΓΓ, ΒΒ, ΑΑ, ΒΒΒΒΒΒ.

Εἰς τὴν ἀκολουθίαν ΕΕΕΕΕΕΕΕ, εἰς ἣν $N=8$ καὶ $n=1$, ἔχομεν μίαν μόνον διαδρομὴν (αὐτὴν ταύτην τὴν ἀκολουθίαν).

Ὁ ἀριθμὸς τῶν στοιχείων ἑκάστης διαδρομῆς καλεῖται μῆκος της. Οὕτω εἰς τὸ ἀνωτέρω παράδειγμα (10) ἔχομεν 2 διαδρομὰς διὰ τὸ γεγονὸς Α, μήκους 1 καὶ 2 στοιχείων ἀντιστοίχως, μίαν διαδρομὴν διὰ τὸ γεγονὸς Γ, μήκους 4, καὶ 2 διαδρομὰς διὰ τὸ γεγονὸς Β, μήκους 2 καὶ 6 στοιχείων ἀντιστοίχως.

Ἡ ἁπλουστέρα περίπτωσις εἶναι ἐκείνη καθ' ἣν ἔχομεν ἀκολουθίας μὲ δύο μόνον διάφορα γεγονότα ἢ ἐναλλακτέα.

Ἀκολουθίαι δύο ἐναλλακτέων, καὶ συνεπῶς καὶ διαδρομαὶ δύο ἐναλλακτέων, παρουσιάζονται συχνότατα εἰς τὴν πρᾶξιν.

Ἔχομεν π.χ. ἀκολουθίας «κορωνῶν» καὶ «γραμμάτων» κατὰ τὴν ρῖψιν ἑνὸς νομίσματος, διαδρομὰς ἀρρένων καὶ θηλέων εἰς σειρὰν γεννήσεων κατὰ χρονολογικὴν τάξιν, ἀκολουθίας ὑγιῶν καὶ ἀσθενῶν ἀτόμων κατὰ μίαν ὁμαδικὴν ἐξέτασιν κατὰ χρονολογικὴν τάξιν ἐξετάσεως, κενὰ καὶ πλήρη καθίσματα εἰς μίαν σειρὰν καθισμάτων ἑνὸς θεάτρου κατὰ τὴν διάρκειαν μιᾶς παραστάσεως, ἀκολουθίας μονῶν καὶ ζυγῶν ἀριθμῶν κατὰ τὰς διαδοχικὰς κληρώσεις μιᾶς ρουλέττας κλπ.

Ἐπίσης εἶναι δυνατὸν νὰ συγκρίνωμεν τὰ στοιχεῖα ἑνὸς δείγματος (κατὰ χρονικὴν τάξιν ἐμφανίσεώς των) μὲ ἕνα δοθέντα σταθερὸν ἀριθμὸν Κ καὶ ἐὰν ἓν στοιχεῖον εἶναι μεγαλύτερον τοῦ Κ νὰ σημειώνωμεν εἰς τὴν θέσιν του ἕνα (+), ἐὰν δὲ μικρότερον ἕνα (—). Θὰ ἔχωμεν οὕτω διαδρομὰς θετικῶν καὶ ἀρνητικῶν σημείων.

Ἐνδιαφέροντα στατιστικὰ προβλήματα δύνανται νὰ ἀναλυθῶσι δι' ἁπλῆς ἀπαριθμήσεως τῶν διαδρομῶν μιᾶς ἀκολουθίας ἐκ δύο ἐναλλακτέων.

Τινὰ τῶν προβλημάτων τούτων θὰ θεωρήσωμεν.

Ἔστω κατ' ἀρχὴν ἕνας πληθυσμὸς μιᾶς συνεχοῦς μεταβλητῆς.

Ἂς λάβωμεν δύο τυχαῖα δείγματα μεγέθους n_1 καὶ n_2 ἐκ τοῦ πληθυσμοῦ τούτου καὶ ἂς καλέσωμεν Α καὶ Β τὰ δείγματα ἀντιστοίχως.

Πρὸς διάκρισιν, τὰ στοιχεῖα τοῦ δείγματος Α ἂς τὰ καλέσωμεν α_i καὶ τοῦ Β, β_i, δηλαδή :

δεῖγμα Α: $\alpha_1, \alpha_2, \alpha_3 \ldots \alpha_{n_1}$

» Β: $\beta_1, \beta_2, \beta_3 \cdots \beta_{n_2}$

ἔνθα τὰ στοιχεῖα εἶναι διατεταγμένα κατὰ τὴν χρονικὴν σειρὰν ἐκλογῆς των.

Καὶ ἔστω

$$\left.\begin{array}{l} X: x_1, x_2 \cdots\cdots x_{n_1} \\ \Psi: y_1, y_2 \cdots\cdots y_{n_2} \end{array}\right\}$$ τὰ ἴδια δείγματα (Α=Χ) καὶ (Β=Υ) διατεταγμένα κατὰ μέγεθος.

Καὶ ἔστω τέλος

Ζ: $z_1, z_2, \ldots z_{n_1+n_2}$ τὸ μικτὸν διατεταγμένον κατὰ μέγεθος δεῖγμα ἐκ τῶν Α καὶ Β ληφθέντων ὁμοῦ.

Τὸ μικτὸν δεῖγμα Ζ θὰ σχηματισθῇ ἀπὸ κάποιαν διάταξιν τῶν x_i καὶ y_i.

Π. χ. μιὰ δυνατὴ διάταξις τοῦ Ζ εἶναι $x_1\, x_2\, y_1\, y_2 \ldots y_{n_2}\, x_3\, x_4\, x_5 \ldots x_{n_1}$ (11)

Θὰ ζητήσωμεν νὰ εὕρωμεν τὴν πιθανότητα ἵνα τὸ Ζ ἔχῃ τὴν διάταξιν (11).

Εἶναι προφανὲς ὅτι εἰς τὴν (11) εἰς τὰς θέσεις τῶν y_j δυνάμεθα νὰ θέσωμεν οἱανδήποτε μετάθεσιν τῶν β_j διότι κάθε β δύναται νὰ εἶναι y_1 ἢ y_2 κλπ. ὅλαι αὐταὶ αἱ μεταθέσεις εἶναι $n_2!$ Ὁμοίως εἰς τὴν θέσιν τῶν x_i εἰς τὴν (11) δυνάμεθα νὰ θέσωμεν κάθε μετάθεσιν τῶν α_i καὶ ὁ ἀριθμὸς αὐτῶν τῶν μεταθέσεων θὰ εἶναι n_1 !

Ἕνας ὡρισμένος τρόπος σχηματισμοῦ τοῦ μικτοῦ δείγματος Ζ, π.χ. ὁ (11), δύναται νὰ ἀντιστοιχῇ εἰς $(n_1!)(n_2!)$ μεταθέσεις τῶν n_1+n_2 στοιχείων α_i καὶ β_j.

Ἀλλὰ ὅλαι αἱ δυναταὶ μεταθέσεις τῶν (n_1+n_2) στοιχείων α_i καὶ β_j εἶναι $(n_1+n_2)!$ Ἑπομένως ἡ πιθανότης σχηματισμοῦ τοῦ Ζ κατὰ τὴν διάταξιν (11) εἶναι : $\dfrac{(n_1!)\cdot(n_2!)}{(n_1+n_2)!} = \dfrac{1}{\binom{n_1+n_2}{n_1}}$.

Ἐὰν θεωρήσωμεν ἕνα ἄλλον ὡρισμένον τρόπον σχηματισμοῦ τοῦ Ζ, εὑρίσκομεν πάλιν τὴν ἰδίαν πιθανότητα, δηλαδὴ ὅλοι οἱ τρόποι σχηματισμοῦ τοῦ Z εἶναι ἐξ ἴσου πιθανοί.

Ἐξ ἄλλου ἐφ' ὅσον ἡ πιθανότης κάθε ὡρισμένου τρόπου σχηματισμοῦ τοῦ Z εἶναι $\frac{1}{\binom{n_1+n_2}{n_1}}$ ἕπεται ὅτι ὅλοι οἱ διάφοροι τρόποι σχηματισμοῦ τοῦ Z εἶναι $\binom{n_1+n_2}{n_1}$ (διότι ἡ πιθανότης σχηματισμοῦ οἱουδήποτε ἐξ αὐτῶν ἀδιαφόρως θὰ ἰσοῦται μὲ τὴν μονάδα).

Συγκρίσεις πληθυσμῶν ἢ δοκιμασία ὑποθέσεων ὅτι δύο δείγματα προέρχονται ἀπὸ τὸν ἴδιον πληθυσμόν.

Ἂς ὑποθέσωμεν ἤδη ὅτι δύο ἀκολουθίαι στοιχείων ὡς αἱ ἀνωτέρω, ἀλλὰ διὰ τὰς ὁποίας δὲν γνωρίζομεν ἐὰν προέρχωνται ἐκ τοῦ αὐτοῦ πληθυσμοῦ, ἀναμιγνύονται εἰς ἓν δεῖγμα Z διατεταγμένον κατὰ μέγεθος καὶ ὅτι καταμετρῶμεν τὸν ἀριθμὸν τῶν διαδρομῶν τῶν x καὶ y, ἔστω δὲ οὗτος u.

Εἶναι φανερὸν ὅτι ἐὰν τὰ δύο δείγματα (δηλαδὴ αἱ δύο ἀκολουθίαι στοιχείων) προέρχωνται ἀπὸ τὸν ἴδιον πληθυσμὸν τὰ x καὶ y εἰς τὸ Z εἶναι κατ' ἀρχὴν (ἢ κατὰ μέσον ὅρον) καλῶς ἀναμεμιγμένα καὶ τὸ u θὰ εἶναι μεγάλον.

Ἐὰν ἀντιθέτως τὰ δείγματα ἀνήκουν εἰς πληθυσμοὺς τελείως ξεχωριστούς, δηλαδὴ ὅταν τὸ εὖρος τῆς μεταβλητῆς εἰς τὸν ἕνα πληθυσμὸν δὲν ἔχει οὐδὲν κοινὸν σημεῖον μὲ τὸ εὖρος τῆς μεταβλητῆς τοῦ δευτέρου πληθυσμοῦ, τὸ u θὰ εἶναι ἴσον μὲ 2.

Καὶ γενικῶς διαφοραὶ μεταξὺ τῶν δύο πληθυσμῶν θὰ τείνωσι νὰ ἐλαττώσωσι τὸ u. Οὕτω π.χ. οἱ δύο πληθυσμοὶ δυνατὸν νὰ ἔχωσι τὴν αὐτὴν διάμεσον, ἀλλὰ εἰς τὸν πληθυσμὸν τῶν x δυνατὸν νὰ ἔχωμεν πολὺ μικρὰν διασπορὰν (δηλαδὴ ὅλα τὰ x νὰ εἶναι πολὺ πλησίον τοῦ μέσου ἢ τῆς διαμέσου), ἐνῷ εἰς τὸν πληθυσμὸν τῶν y νὰ ἔχωμεν μεγάλην διασποράν.

Τότε εἰς τὰ δύο ἄκρα τοῦ Z θὰ ἔχωμεν (κατὰ μέσον ὅρον) μακρὰς διαδρομὰς στοιχείων y καὶ συνεπῶς τὸ u θὰ εἶναι ἐν γένει ἠλαττωμένον.

Οὕτω τὸ κριτήριον τὸ βασιζόμενον εἰς τὸν ἀριθμὸν τῶν διαδρομῶν συνίσταται εἰς τὸ νὰ μετρήσωμεν τὸ u εἰς τὸ μικτὸν δεῖγμα Z καὶ νὰ ἀπορρίψωμεν ἢ μὴ τὴν H_0 (δηλαδὴ ὅτι τὰ δείγματα ἀνήκουν εἰς τὸν ἴδιον πληθυσμὸν ἢ εἰς δύο ὁμοίους πληθυσμούς), ἐὰν τὸ u εἶναι μικρότερον ἑνὸς ὡρισμένου ἀριθμοῦ u_0.

Πρὸς τοῦτο δέον νὰ γνωρίζωμεν τὴν κατανομὴν πιθανότητος τοῦ u.

Κατανομὴ πιθανότητος τοῦ u.

Εἴδομεν ὅτι ὅλοι οἱ διάφοροι τρόποι σχηματισμοῦ τοῦ Z εἶναι $\binom{n_1+n_2}{n_1}$ καὶ ἔχουν τὴν ἰδίαν πιθανότητα νὰ συμβοῦν ὅταν ἡ H_0 εἶναι ἀληθής.

Περίπτωσις 1η ἔστω ὅτι τὸ u = ἄρτιον, ἔστω 2K (ἔνθα K = ἀκέραιος).

Θὰ ἔχωμεν K διαδρομὰς διὰ τὰ x καὶ K διαδρομὰς διὰ τὰ y, διότι ἂν τὸ πρῶτον στοιχεῖον τοῦ Z εἶναι τὸ x_1, διὰ νὰ ἔχωμεν ζυγὸν u πρέπει νὰ τελειώσωμεν τὸν σχηματισμὸν τοῦ Z μὲ τὸ $y_{\bar{n}_2}$ καὶ συνεπῶς ἔχομεν K διαδρομὰς τῶν x καὶ K διαδρομὰς τῶν y.

Τὸ ἴδιον ἰσχύει ὅταν $z_1=y_1$ ὅτε $z_{n_1+n_2}=x_{n_1}$

Διὰ νὰ ἔχωμεν K διαδρομὰς τῶν x πρέπει τὰ n_1 στοιχεῖα x νὰ διαιρῶνται ὑπὸ τῶν y εἰς K ὁμάδας καὶ ὁμοίως διὰ νὰ ἔχωμεν K διαδρομὰς τῶν y πρέπει τὰ n_2 στοιχεῖα y νὰ διαιρῶνται ὑπὸ τῶν x εἰς K ὁμάδας.

Ἡ διαίρεσις τῶν n_1 στοιχείων x εἰς K ὁμάδας δύναται νὰ γίνῃ κατὰ πολλοὺς ἐν γένει τρόπους· διὰ νὰ εὕρωμεν πόσοι εἶναι οἱ τρόποι μιᾶς τοιαύτης διαιρέσεως θεωροῦμεν τὴν ταυτότητα:

$$\frac{1-x^{n_1}}{1-x}=1+x+x^2+\ldots\ldots+x^{n_1-1}$$

ἢ τὴν
$$\frac{x(1-x^{n_1})}{1-x}=x+x^2+x^3+\ldots\ldots+x^{n_1}$$

ἢ τὴν
$$\left(x+x^2+x^3+\ldots+x^{n_1}\right)^K\equiv x^K\left(1-x^{n_1}\right)^K\frac{1}{(1-x)^K}\qquad(11α)$$

ἀλλὰ

$$\frac{1}{(1-x)^K}\equiv 1+\binom{K}{1}x+\binom{K+1}{2}x^2+\binom{K+2}{3}x^3+\ldots\equiv\sum_{i=o}^{\infty}\binom{K+i-1}{i}x^i\quad(12)$$

ὅθεν ἡ (11α) γίνεται

$$(x+x^2+\ldots+x^{n_1})^K\equiv x^K(1-x^{n_1})^K\sum_{i=o}^{\infty}\binom{K+i-1}{i}x^i$$

Ὁ συντελεστὴς τοῦ ὅρου x^{n_1-K} εἰς τὸ γινόμενον $(1-x^{n_1})\sum_{i=o}^{\infty}\binom{K+i-1}{i}x^i$

εἶναι $\binom{K+(n_1-K)-1}{n_1-K}=\binom{n_1-1}{n_1-K}$

Ὁ συντελεστὴς οὗτος μᾶς δίδει τὸν ἀριθμὸν τῶν τρόπων καθ' οὓς δυνάμεθα νὰ ἔχωμεν ἐκ «K διατεταγμένων» μερῶν (μεγαλυτέρων τοῦ μηδενὸς) τοῦ n_1 ἄθροισμα ἴσον πρὸς τὸ n_1, ἤτοι κατὰ πόσους τρόπους δυνάμεθα νὰ ἔχωμεν K διαδρομὰς τῶν x εἰς τὸ δεῖγμα Z.

Ὁμοίως ὁ ἀριθμὸς τῶν τρόπων καθ' οὓς δυνάμεθα νὰ ἔχωμεν K διαδρομὰς τῶν y εἰς τὸ Z θὰ εἶναι $\binom{n_2-1}{K-1}$. Ἐπειδὴ δὲ κάθε τρόπος καθ' ὃν ἔχομεν K διαδρομὰς x δύναται νὰ συνδυασθῇ μὲ οἱονδήποτε τρόπον καθ' ὃν ἔχομεν K διαδρομὰς τῶν y, προκύπτει ὅτι ὅλοι οἱ δυνατοὶ τρόποι καθ' οὓς δυνάμεθα νὰ σχηματίσωμεν τὸ Z ὥστε νὰ ἔχωμεν K διαδρομὰς τῶν x καὶ K διαδρομὰς τῶν y, εἶναι $\binom{n_1-1}{K-1}\cdot\binom{n_2-1}{K-1}$

Ὁ ἀριθμὸς αὐτὸς πρέπει τέλος νὰ πολλαπλασιασθῇ ἐπὶ 2, διότι δυνατὸν τὸ Z νὰ ἀρχίζῃ ἀπὸ x_1 καὶ νὰ τελειώνῃ εἰς y_{n_2} ἢ νὰ ἀρχίζῃ ἀπὸ y_1 καὶ νὰ τελειώνῃ εἰς x_{n_1}.

Θὰ ἔχωμεν οὕτω ἐν γένει $2\binom{n_1-1}{K-1}\binom{n_2-1}{K-1}$ τρόπους.

Ἡ πιθανότης ἑκάστης διατάξεως Z (ἢ ἑκάστου τρόπου σχηματισμοῦ τοῦ Z) εὕρομεν ὅτι ἦτο $\frac{1}{\binom{n_1+n_2}{n_1}}$, συνεπῶς

$$\Pi\iota\vartheta(u=2K)=\frac{2\binom{n_1-1}{K-1}\binom{n_2-1}{K-1}}{\binom{n_1+n_2}{n_1}} \tag{14}$$

Δευτέρα περίπτωσις: Ἔστω $u=2K+1$.

Ἀναλόγως εὑρίσκομεν:

$$\Pi\iota\vartheta\,(u=2K+1)=\frac{\binom{n_1-1}{K}\binom{n_2-1}{K-1}+\binom{n_1-1}{K-1}\binom{n_2-1}{K}}{\binom{n_1+n_2}{n_1}} \tag{15}$$

Αἱ (14) καὶ (15) μᾶς δίδουν τὴν ζητουμένην κατανομὴν πιθανότητος τοῦ ἀριθμοῦ τῶν διαδρομῶν u εἰς τὸ μικτὸν δεῖγμα Z [1].

Διὰ νὰ κρίνωμεν ὅθεν τὴν μηδενικὴν ὑπόθεσιν (ὅτι τὰ δύο δείγματα προέρχονται ἀπὸ τὸν αὐτὸν πληθυσμὸν) εἰς ἓν ἐπίπεδον σημαντικότητος π.χ. 5 %, εὑρίσκομεν τὸν μεγαλύτερον ἀκέραιον u_0 τοιοῦτον ὥστε ἡ πιθανότης νὰ ἔχωμεν 0 ἢ 1 ἢ 2 ἢ... u_0 διαδρομὰς νὰ εἶναι μικροτέρα τοῦ 5 %.

Ἐὰν τότε σχηματίσωμεν ἓν δεῖγμα Z ἐκ δύο τυχαίων δειγμάτων καὶ εὕρωμεν $u<u_0$ ἀπορρίπτομεν τὴν H_0 εἰς τὸ 5 % ἐπίπεδον σημαντικότητος. Ἐὰν u παρατηρηθὲν εἶναι μεγαλύτερον τοῦ κ_0 δὲν δυνάμεθα νὰ ἀπορρίψωμεν τὴν H_0 εἰς τὸ 5 % ἐπίπεδον.

Ἀναλόγως ἐργαζόμεθα καὶ ὅταν κρίνωμεν τὴν H_0 εἰς ἄλλο ἐπίπεδον σημαντικότητος π.χ. εἰς τὸ 2,5 % κλπ.

Δέον νὰ σημειώσωμεν ὅτι εἰς τὸ πρόβλημα τῆς κρίσεως τῆς ὑποθέσεως ὅτι δύο δείγματα προέρχονται ἀπὸ τὸν ἴδιον (ἢ ὁμοίους) πληθυσμοὺς διὰ τῆς μεθόδου τῶν διαδρομῶν, ἐφαρμόζομεν μονόπλευρον (ἢ ἑνὸς ἄκρου) δοκιμασίαν, διότι πράγματι ἀπορρίπτομεν τὴν H_0 ὅταν ἔχωμεν ὀλιγωτέρας διαδρομὰς τοῦ u_0 καὶ δὲν ἐνδιαφερόμεθα νὰ τὴν ἀπορρίψωμεν ὅταν ἔχωμεν περισσοτέρας διαδρομὰς ἑνὸς ἀριθμοῦ π.χ. u_0.

Εἰς ἄλλου ὅμως εἴδους προβλήματα, ἀναλυόμενα διὰ τῆς μεθόδου τῶν διαδρομῶν, μᾶς ἐνδιαφέρει νὰ μὴ ἔχωμεν οὔτε πάρα πολλὰς οὔτε πολὺ ὀλίγας τοιαύτας. Τέλος εἰς ἄλλας περιπτώσεις, ὡς θὰ ἴδωμεν, ἐνδιαφερόμεθα μόνον νὰ μὴ ἔχωμεν πάρα πολλὰς διαδρομάς. Θὰ ἔχωμεν δηλαδὴ ἐν γένει «δύο ἄκρων» ἢ «ἑνὸς ἄκρου» δοκιμασίαν ἀναλόγως τῆς περιπτώσεως.

---

(1) Ἡ κατανομὴ αὕτη εὑρέθη ὑπὸ τοῦ W. L. Stevens καὶ ὑπὸ τῶν Wald καὶ Wolfowitz, ἴδε W. L. Stevens: «Distribution of groups in a sequence of alternatives» Annals of Eugenics Vol. IX (1939) καὶ A. Wald and J. Wolfowitz «On a test of whether two samples are from the same population», Annals of Mathematical Statistics Vol. XI (1940). Ἴδε ὁμοίως Mathematical Statistics (1946) S. S. Wilks, σελὶς 200 – 207. Ἐπέκτασιν τῆς θεωρίας τῶν runs εἰς περίπτωσιν περισσοτέρων ἐναλλακτέων ἴδε Mood: The theory of runs (Annals of M. Stat. Vol. XI 1940 p. 367).

Τὸ κριτήριον τῶν διαδρομῶν εἶναι «εὐαίσθητον» ὄχι μόνον κατὰ τὴν θέσιν ἀλλὰ καὶ κατὰ τὴν μορφὴν τῶν κατανομῶν. Ὅταν κρίνωμεν ἐὰν δύο δείγματα προέρχονται ἀπὸ ὁμοίους πληθυσμοὺς διὰ τῆς μεθόδου τῶν διαδρομῶν (runs), κρίνομεν ἐὰν οἱ πληθυσμοὶ ἔχωσι τὴν αὐτὴν θέσιν καὶ τὴν αὐτὴν κατανομὴν συχνότητος συγχρόνως.

Αἱ σχέσεις (14) καὶ (15) ὑπολογίζονται ἀπ' εὐθείας διὰ μικρὰς τιμὰς τῶν n_1 καὶ n_2. Ὅταν τὰ n_1 καὶ n_2 εἶναι μεγάλα, ὁ ἀπ' εὐθείας ὑπολογισμὸς τοῦ u_0 τῇ βοηθείᾳ τῶν (14) ἢ (15) εἶναι κοπιώδης. Ὅταν n_1 καὶ n_2 εἶναι μεγάλα ἡ κατανομὴ τοῦ u εἶναι κατὰ προσέγγισιν κανονική. Ἡ προσέγγισις εἶναι ἀρκετὰ καλὴ πρακτικῶς ὅταν n_1 καὶ n_2 εἶναι ἀμφότερα μεγαλύτερα τοῦ 10.

Ἀναφέρομεν ἄνευ ἀποδείξεως ὅτι $E(u) = \dfrac{2n_1n_2}{n_1+n_2} + 1$

καὶ $\sigma^2_u = \dfrac{2n_1n_2\,(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$. Ἔνθα $E(u)$ = ἀναμενομένη τιμὴ τοῦ u (expected value) ἢ μέσος ἀριθμητικὸς τοῦ u.

Ἔχουν ἤδη ὑπολογισθῆ πίνακες, οἵτινες δίδουν ἑτοίμους, διὰ τιμὰς τῶν n_1 καὶ n_2 ἀπὸ 2 ἕως 100, τὰς τιμὰς $u_{0\cdot025}$ καὶ $u_{0\cdot975}$ τοιαύτας ὥστε, ἐὰν εἰς τὸ δεῖγμα z εὕρωμεν $u\delta \leqslant u_{0\cdot025}$ τοῦ πίνακος, διὰ δοθέντα n_1 καὶ n_2, τοῦτο θὰ συμβαίνῃ ἕνεκα τυχαίας διακυμάνσεως μὲ πιθανότητα $\leqslant 2{,}5\%$. Ὁμοίως ὅταν $u\delta \geqslant u_{0\cdot975}$. (Ἴδε Eisenhart C. and Swed F. «Tables for testing randomness of grouping in a sequence of alternatives» Annals of Mathematical Statistics, vol. XIV (1943) p. 66 ἢ Dixon and Massey: Introduction to Statistical Analysis (1951) p. 325-326 (πίνακες 11), ὅπου οἱ ἀνωτέρω πίνακες ἀναδημοσιεύονται).

Παράδειγμα

Προκειμένου νὰ κριθῇ ἡ ἀποδοτικὴ ἀξία μιᾶς νέας ποικιλίας σπόρου σίτου Β ἐν συγκρίσει πρὸς μίαν ἄλλην Α, ἤδη χρησιμοποιουμένην, ἐκτελεῖται τὸ ἑξῆς πείραμα:

Μίαν ἔκτασιν ἐδάφους (τῆς αὐτῆς ποιότητος) ἐκ 16 στρεμμάτων τὴν χωρίζομεν εἰς 16 ἴσα μέρη, ἐκτάσεως ἑνὸς στρέμματος. Εἶτα «τυχαίως» ἐκλέγομεν 8 μέρη ἐκ τούτων, εἰς ἃ σπείρομεν τὴν ποικιλίαν Β, ἐνῷ εἰς τὰ ἕτερα 8 σπείρομεν τὴν ποικιλίαν Α (μὲ ἴσας ποσότητας σπόρων).

Κατὰ τὴν ἐποχὴν τῆς συγκομιδῆς μετρῶμεν τὰς στρεμματικὰς ἀποδόσεις εἰς τὰ 16 μέρη κεχωρισμένως καὶ ἔστω ὅτι αἱ στρεμματικαὶ ἀποδόσεις, κατατασσόμεναι κατὰ μέγεθος καὶ κατ' εἶδος σπόρου, εἶναι αἱ κάτωθι εἰς κιλά:

Ποικιλία Α	110,5	111,3	112,7	113,1	114,6	117,1	120,2	120,6
Ποικιλία Β	111,6	111,8	113,2	114,8	115,2	116,0	121,0	121,4

Εἶναι δυνατὸν νὰ θεωρηθῇ ὅτι ἡ ποικιλία Β τοῦ σίτου εἶναι ἀποδοτικωτέρα τῆς ποικιλίας Α εἰς τὸ 2,5% ἐπίπεδον σημαντικότητος;

Λύσις: Ἡ ὑπόθεσις μηδὲν (H_0) ἐνταῦθα εἶναι ὅτι τὰ δύο εἴδη σπόρων εἶναι τῆς αὐτῆς ἀποδόσεως, δηλαδὴ ὅτι αἱ διαφοραὶ εἰς τὰς ἀποδόσεις ὀφείλονται εἰς τὰς διακυμάνσεις τῆς τυχαίας δειγματοληψίας.

7. 8. 46

Σχηματίζω τὸ μικτὸν διατεταγμένον δεῖγμα, εἰς ὃ ὑποσημειώνω κάτωθεν ἑκάστου ἀριθμοῦ τὸ ψηφίον Α ἢ Β, ἐφ' ὅσον ὁ ἀριθμὸς δηλοῖ ἀπόδοσιν τῆς Α ἢ τῆς Β ποικιλίας. Θὰ ἔχω:

110,5	111,3	111,6	111,8	112,7	113,1	113,2	114,6
Α	Α	Β	Β	Α	Α	Β	Α
114,8	115,2	116,0	117,1	120,2	120,6	121,0	121,4
Β	Β	Β	Α	Α	Α	Β	Β

Ἔχομεν εἰς τὴν ἀκολουθίαν ταύτην $n_1=8$, $n_2=8$ καὶ $u_0=8$.

Ἐκ τῶν προαναφερθέντων πινάκων ἔχω $u_{0 \cdot 025}=4$.

Τὸ παρατηρηθὲν $u_0=8$ εἶναι μεγαλύτερον τοῦ 4 καὶ συνεπῶς δὲν ἀπορρίπτω τὴν H_0 (ὅτι αἱ ποικιλίαι σπόρων εἶναι τῆς ἰδίας ἀποδόσεως) εἰς τὸ 2,5% ἐπίπεδον σημαντικότητος.

Εἰς τὸ παράδειγμα τοῦτο ἔχομεν μονόπλευρον δοκιμασίαν (ἢ ἑνὸς ἄκρου), διότι οἱ σπόροι θὰ ἦσαν διαφορετικῆς ἀποδόσεως ἐὰν ὑπῆρχον εἰς τὸ διατεταγμένον δεῖγμα πολὺ ὀλίγαι διαδρομαί.

Δοκιμασία τοῦ «τυχαίου» (randomness) τῆς διατάξεως μιᾶς μακρᾶς ἀκολουθίας.

Ἓν ἄλλο πρόβλημα τὸ ὁποῖον δυνάμεθα νὰ ἐξετάσωμεν διὰ τῶν διαδρομῶν εἶναι τὸ νὰ κρίνωμεν τὴν ὑπόθεσιν τῆς «τυχαίας» διατάξεως μιᾶς μακρᾶς ἀκολουθίας δυὸ ἐναλλακτέων.

Μία μακρὰ σειρὰ δύο ἐναλλακτέων (ἢ γενικώτερον ὁσωνδήποτε ἐναλλακτέων) λέγομεν ὅτι ἔχει «τυχαίαν διάταξιν», ὅταν ἡ γνῶσις ἑνὸς στοιχείου τῆς σειρᾶς οὐδεμίαν πληροφορίαν μᾶς παρέχει περὶ τοῦ εἴδους τῶν ἄλλων.

Π.χ. ἡ σειρὰ ΑΒΑΒΑΒΑΒΑΒ... δὲν εἶναι «τυχαία» (τουλάχιστον εἰς τὸ τμῆμα τῶν 10 πρώτων στοιχείων της), διότι, ὡς παρατηροῦμεν, μετὰ ἕκαστον Α ἔχομεν ἓν Β. Ἀναλύοντες τὸ ὡς ἄνω πρόβλημα διὰ τῆς μεθόδου τοῦ ἀριθμοῦ τῶν διαδρομῶν, χρησιμοποιοῦμεν ἀμφίπλευρον (ἢ δύο ἄκρων) κριτήριον· δηλαδὴ ἀπορρίπτομεν τὸ «τυχαῖον» τῆς διατάξεως ἐὰν ἔχωμεν ἢ π ο λ ὺ ὀ λ ί γ α ς ἢ π ά ρ α π ο λ λ ὰ ς διαδρομάς.

Πράγματι, π ο λ ὺ ὀ λ ί γ α ι διαδρομαὶ δύνανται νὰ θεωρηθῶσιν ὡς μία τάσις ὁμοίων γεγονότων νὰ παρουσιάζωνται ὁμαδικῶς. Οὕτω ἔχομεν ἓν εἶδος πληροφορίας ἀπὸ ἓν ὡρισμένον στοιχεῖον τῆς ἀκολουθίας διὰ τὰ ἑπόμενα ἢ τὰ προηγούμενα. Π ά ρ α π ο λ λ α ὶ ἐξ ἄλλου διαδρομαὶ δύνανται ἐπίσης νὰ θεωρηθῶσιν ὡς ἔνδειξις ὑπάρξεως μιᾶς «κυματοειδοῦς μεταβολῆς» τοῦ εἴδους τῶν στοιχείων τῆς ἀκολουθίας (wave-like variation) καὶ συνεπῶς καὶ πάλιν ἔχομεν ἕνα βαθμὸν πληροφορίας ἀπὸ ἓν στοιχεῖον τῆς ἀκολουθίας διὰ τὰ ἄλλα.

Ἐν τούτοις ὑπάρχουν ὡρισμέναι περιπτώσεις, κατὰ τὰς ὁποίας κρίνομεν τὸ τυχαῖον τῆς ἀκολουθίας διὰ μονοπλεύρου κριτηρίου (πολὺ ὀλίγων ἢ πάρα πολλῶν διαδρομῶν) ὅταν γνωρίζωμεν τὸ πιθανὸν ἐναλλακτέον τοῦ μὴ τυχαίου.

Παραδείγματα (1)

1ον) Ὡς μίαν περίπτωσιν ὅπου δυνάμεθα νὰ κρίνωμεν τὸ τυχαῖον τῆς διατάξεως μιᾶς μακρᾶς σειρᾶς δύο ἐναλλακτέων, ὅταν ὑποψιαζόμεθα ὡς ἐναλλακτέον τοῦ τυχαίου τὴν ὁμαδοποίησιν τῶν ὁμοίων στοιχείων, δυνάμεθα νὰ θεωρήσωμεν τὰς κατὰ τάξιν θέσεις ὑγιῶν καὶ προσβεβλημένων ὑπό τινος ἀσθενείας φυτῶν εἰς μίαν μακρὰν σειρὰν φυτῶν ἐπὶ τοῦ ἐδάφους.

Ἔχομεν ἐνταῦθα τὴν εὔλογον ὑποψίαν ὅτι ἡ ἀσθένεια τῶν φυτῶν εἶναι μεταδοτικὴ καὶ συνεπῶς ὅτι ἐὰν κάπου, εἰς τὴν μακρὰν σειρὰν τῶν φυτῶν, ὑπάρχει ἕν προσβεβλημένον ὑπὸ τῆς ἀσθενείας φυτόν, τότε καὶ τὰ γειτονικὰ πρὸς αὐτὸ θὰ εἶναι ἐπίσης προσβεβλημένα.

Πληθυσμός μας ἐνταῦθα εἶναι ὁλόκληρος ἡ μακρὰ σειρὰ τῶν φυτῶν ἐπὶ τοῦ ἐδάφους. Ἔστω ὅτι λαμβάνομεν ἓν τυχαῖον διάστημα 25 στοιχείων εἰς τὴν μακρὰν σειρὰν (π.χ. ἀπὸ τὸ 51ον ἕως τοῦ 75ου στοιχείου τῆς σειρᾶς) καὶ τὸ ἀποτέλεσμα εἶναι τὸ ἑξῆς :

ΥΥΥΥΥΥΥΥΥΥΑΥΑΑΑΑΥΥΥΥΥΥΥΥΥ (ἔνθα Υ = ὑγιὲς καὶ Α = ἀσθενὲς φυτόν).

Ἔχομεν $n_1 = 20$, $n_2 = 5$ καὶ $u = 5$. Ἐκ τῶν πινάκων $u_{0,025} = 5$, ἑπομένως τὸ γεγονὸς εἶναι σημαντικὸν στατιστικῶς εἰς τὸ 2,5 % ἐπίπεδον σημαντικότητος καὶ ἀπορρίπτομεν τὴν H_0 (H_0 = ὅτι διάταξις εἶναι τυχαία), δηλαδὴ ἀποφαινόμεθα ὅτι καὶ εἰς τὴν μακρὰν σειρὰν τῶν χιλιάδων φυτῶν ἡ ἐναλλαγὴ τῶν Υ καὶ Α δὲν εἶναι τυχαία, διότι ἂν ἦτο τυχαία, ὁπουδήποτε καὶ ἐὰν ἐλαμβάνομεν τυχαῖα διαστήματα 25 διαδοχικῶν στοιχείων, θὰ εὑρίσκομεν εἰς τὰ 97,5 % τοῦ ἀριθμοῦ τῶν τυχαίων τούτων 25άδων ἀριθμὸν διαδρομῶν u μεγαλύτερον τοῦ 5.

2ον) Ὡς μίαν περίπτωσιν, ὅπου πάρα πολλαὶ διαδρομαὶ δύνανται νὰ εἶναι τὸ πιθανὸν ἐναλλακτέον τοῦ τυχαίου τῆς διατάξεως μιᾶς ἀκολουθίας, δυνάμεθα νὰ θεωρήσωμεν τὸ ἑξῆς :

Εἰς ἓν μπὰρ-ἑστιατόριον (ὅπου οἱ πελάται κάθηνται εἰς μίαν μακρὰν σειρὰν καθισμάτων πυκνῶς τοποθετημένων τὸ ἓν παραπλεύρως τοῦ ἄλλου, διὰ τὴν ἐξυπηρέτησιν ὅσον τὸ δυνατὸν περισσοτέρων ἀτόμων πρὸ τοῦ μακροῦ πάγκου (counter) σερβιρίσματος ροφημάτων καὶ φαγητῶν), ἡμισείαν ὥραν πρὸ τῆς συνήθους ὥρας τοῦ μεσημβρινοῦ γεύματος, μετρῶμεν καὶ καταγράφομεν κατὰ τὴν ἰδίαν σειρὰν (ὅπως τὰ βλέπομεν ἐν τῷ ἑστιατορίῳ) τὰ κενὰ καὶ τὰ πλήρη καθίσματα τοῦ ἑστιατορίου.

Ἔστω ὅτι ἓν τμῆμα τῆς μακρᾶς σειρᾶς μᾶς παρέσχε τὴν ἑξῆς ἀκολουθίαν ΚΠΚΚΠΚΚΚΠΚΠΚΠΚΠΚ (ἔνθα Κ = κενὸν καὶ Π = πλῆρες κάθισμα). Δυνάμεθα νὰ ἀπορρίψωμεν τὴν ὑπόθεσιν H_0 περὶ τυχαίας διατάξεως κενῶν καὶ πλήρων καθισμάτων (τὴν ὥραν ἐκείνην), εἰς τὴν μακρὰν σειρὰν καθισμάτων τοῦ ἑστιατορίου εἰς τὸ 2,5 % ἐπίπεδον σημαντικότητος ;

Ἔχομεν $n_1 = 10$, $n_2 = 6$ καὶ $u = 13$ καὶ ἐκ τῶν πινάκων εὑρίσκομεν

(1) Τὰ παραδείγματα 1 καὶ 2 ἀναφέρονται μὲ μικρὰς ἀριθμητικὰς παραλλαγὰς εἰς τὸ προαναφερθὲν ἄρθρον τῶν C. Eisenhart καὶ F. Swed εἰς τὸ Annals of Math. Statistics (1943) p. p. 66-87.

$u_{0,975} = 12$. Ὅθεν ἀπορρίπτομεν τὴν ὑπόθεσιν τυχαίας διατάξεως εἰς τὸ 2,5 % ἐπίπεδον σημαντικότητος.

Εἰς τὸ ὡς ἄνω παράδειγμα ἔχομεν τὴν εὔλογον ὑποψίαν ὅτι οἱ πελάται προτιμοῦν, χάριν τῆς ἀνέσεώς των καὶ τῆς ἐλευθερίας κινήσεών των κατὰ τὴν ὥραν τοῦ φαγητοῦ των, νὰ κάθηνται εἰς καθίσματα τοιαῦτα, ὥστε τὰ γειτονικά των νὰ εἶναι κενά, διὰ νὰ μὴν ἐνοχλῶνται ἀπὸ τὰς κινήσεις τῶν παραπλεύρως καθημένων (ἐφ' ὅσον δύνανται νὰ πράξουν τοῦτο), δηλαδὴ ἀναμένομεν ὡς ἐναλλακτέον τοῦ τυχαίου μίαν τάσιν πρὸς κυματοειδῆ μεταβολὴν τῶν κενῶν καὶ πλήρων καθισμάτων (παρόμοιον φαινόμενον παρατηρεῖται καὶ εἰς τὸν κινηματογράφον ὅταν οὗτος δὲν εἶναι τελείως πλήρης θεατῶν, διότι συνήθως προτιμῶμεν θέσεις ἐλευθέρας τόσον ἐμπρὸς ὅσον καὶ παραπλεύρως μας διὰ τὴν ἄνετον παρακολούθησιν τῆς ταινίας).

3ον) Ὡς περίπτωσιν, τέλος, ἀμφιπλεύρου κριτηρίου διὰ τὴν H_0 (περὶ «τυχαίας» διατάξεως τῶν στοιχείων μιᾶς ἀκολουθίας) δυνάμεθα νὰ θεωρήσωμεν τὸ κάτωθι παράδειγμα :

Εἰς τὰς πρώτας 30 κληρώσεις μιᾶς ρουλέττας τὰ ἀποτελέσματα ἦσαν :

ΜΜΚΚΜ ΜΚΜΜΚ ΚΜΚΜΚ ΚΜΜΚΚ ΜΚΜΜΜ ΚΚΜΚΚ (Μ=μαῦρο Κ=κόκκινο).

Ζητεῖται νὰ κριθῇ ἡ ὑπόθεσις ὅτι ἡ ρουλέττα δὲν παρουσιάζει, οὔτε τάσιν κυματοειδοῦς μεταβολῆς κληρώσεων ἐρυθρῶν καὶ μαύρων, οὔτε τάσιν ὁμαδοποιήσεως τῶν ἐρυθρῶν καὶ μαύρων κατὰ τὰς κληρώσεις της, δηλαδὴ ὅτι ἡ ρουλέττα κατὰ τὰς κληρώσεις της παρουσιάζει τυχαίαν διάταξιν χρώματος εἰς τὸ 5 % ἐπίπεδον σημαντικότητος.

Ἔχομεν $n_1 = 15$, $n_2 = 15$ καὶ $u = 18$.

Ἐκ τῶν πινάκων τῶν Eisenhart - Swed ἔχομεν $u_{0,025} = 10$ καὶ $u_{0,975} = 21$

Ἡ παρατηρηθεῖσα τιμὴ τοῦ $u = 18$ εὑρίσκεται μεταξὺ τοῦ $u_{0,025} = 10$ καὶ τοῦ $u_{0,975} = 21$. Ἑπομένως δὲν ἀπορρίπτομεν τὴν ὑπόθεσιν H_0 περὶ τυχαίας μεταβολῆς τῶν μαύρων καὶ ἐρυθρῶν κατὰ τὰς κληρώσεις τῆς ρουλέττας εἰς τὸ 5 % ἐπίπεδον σημαντικότητος.

ΠΛΗΘΥΣΜΟΙ ΔΥΟ ΜΕΤΑΒΛΗΤΩΝ

Δοκιμασία ἀνεξαρτησίας τῶν δύο μεταβλητῶν δι' ἐλευθέρων κατανομῆς μεθόδων. Μέθοδος τῶν ἀντιστροφῶν.

Ἔστω ὅτι ἔχομεν ἓν δεῖγμα ἐκ n στοιχείων ἐξ ἑνὸς πληθυσμοῦ δύο μεταβλητῶν x καὶ y (ὅπου αἱ μεταβληταὶ x καὶ y ὑποτίθενται συνεχεῖς).

Εἰς ἕκαστον ἐκ τῶν n στοιχείων ἀντιστοιχεῖ ἓν ζεῦγος τιμῶν (x_i, y_i) καὶ συνεπῶς ἔχομεν n ζεύγη ἀντιστοίχων τιμῶν x καὶ y, ἐξ ὧν ἐνδέχεται νὰ προκύπτῃ μία συσχέτισις μεταξὺ τῶν δύο μεταβλητῶν (ἐν τῷ δείγματι). Εἶναι ὅμως ἡ συσχέτισις αὕτη στατιστικῶς σημαντικὴ (εἰς ἓν ἐπίπεδον σημαντικότητος α %) ἢ μήπως ὀφείλεται, ἡ παρουσιασθεῖσα ἐν τῷ δείγματι συσχέτισις, εἰς τὰς διακυμάνσεις τῆς τυχαίας δειγματοληψίας ;

Πῶς θὰ κρίνωμεν τοῦτο, ἰδίᾳ ὅταν οὐδεμίαν ἔχομεν γνῶσιν ἐπὶ τῆς μορφῆς τῶν κατανομῶν συχνότητος τῶν μεταβλητῶν εἰς τὸν πληθυσμόν ;

Δηλαδὴ ἂν H_0 εἶναι ὅτι αἱ δύο μεταβληταὶ εἶναι εἰς τὸν πληθυσμὸν ἀνεξάρτητοι, δυνάμεθα νὰ ἀπορρίψωμεν τὴν H_0 ἐκ τῶν δεδομένων τοῦ δείγματος εἰς ἓν ἐπίπεδον σημαντικότητος α %;

Ἐκ τῶν διαφόρων ἐλευθέρων κατανομῆς μεθόδων, δι' ὧν δυνάμεθα νὰ ἀναλύσωμεν τὸ πρόβλημα τοῦτο, θὰ ἀναφέρωμεν ἐνταῦθα μίαν μόνον, τὴν μέθοδον τὴν βασιζομένην ἐπὶ τοῦ ἀριθμοῦ τῶν «ἀντιστροφῶν» τῶν δύο σειρῶν τιμῶν τῶν x καὶ y εἰς τὸ δεῖγμα.

Ὁρισμοί. Εἰς τὴν τυχοῦσαν μετάθεσιν π.χ. τῶν πέντε πρώτων (κατὰ φυσικὴν σειρὰν) ἀκεραίων ἀριθμῶν, ἔστω τὴν 3, 5, 4, 1, 2 καλοῦμεν ἀντιστροφὴν (inversion), ἀντίστοιχον πρὸς ἓν στοιχεῖον της π.χ. τὸ 3, κάθε ἀριθμὸν μικρότερον τοῦ 3, εὑρισκόμενον ἐν τῇ μεταθέσει πρὸς τὰ δεξιὰ αὐτοῦ (ἢ ἐὰν ἡ μετάθεσις εἶναι γεγραμμένη εἰς μιὰν κάθετον στήλην, κάθε ἀριθμὸν μικρότερον τοῦ 3 εὑρισκόμενον κάτωθι αὐτοῦ).

Οὕτω εἰς τὴν ἄνω μετάθεσιν ἔχομεν δύο ἀντιστροφὰς ἀντιστοίχους πρὸς τὸ 3: τοὺς ἀριθμοὺς 1 καὶ 2, τρεῖς ἀντιστροφὰς ἀντιστοίχους πρὸς τὸ 5, δύο ἀντιστροφὰς ἀντιστοίχους πρὸς τὸ 4 καὶ μηδὲν ἀντιστροφὰς ἀντιστοίχους πρὸς τὸ 1 ἢ 2.

Ἀριθμὸν ἀντιστροφῶν ὁλοκλήρου τῆς σειρᾶς 3, 5, 4, 1, 2 (κατὰ τὴν σειρὰν ταύτην γεγραμμένην) καλοῦμεν τὸ σύνολον τῶν ἀντιστροφῶν ὅλων τῶν στοιχείων της καὶ τὸ παριστῶμεν διὰ λ.

Οὕτω εἰς τὴν 3, 5, 4, 1, 2 ἔχομεν $\lambda = 2+3+2+0+0 = 7$ ἀντιστροφάς.

Οἱ ὁρισμοὶ αὐτονοήτως ἰσχύουν καὶ ὅταν ἀντὶ τῶν μεταθέσεων τῶν 5 πρώτων ἀριθμῶν (κατὰ φυσικὴν τάξιν) ἔχομεν μεταθέσεις τῶν n πρώτων κατὰ φυσικὴν σειρὰν ἀκεραίων ἀριθμῶν ἢ γενικώτερον μεταθέσεις n διαφόρων κατὰ μέγεθος ἀριθμῶν ἢ n τιμῶν μιᾶς συνεχοῦς μεταβλητῆς.

Εἶναι προφανὲς ὅτι εἰς τὴν μετάθεσιν 1, 2, 3, 4... n ἔχομεν ἀριθμὸν ἀντιστροφῶν τῆς σειρᾶς μηδὲν (δηλαδὴ τὸν ἐλάχιστον δυνατὸν ἀριθμόν), ἐνῷ εἰς τὴν μετάθεσιν: $n, n-1, n-2, \ldots, 3, 2, 1$ θὰ ἔχωμεν $\frac{n(n-1)}{2}$ ἀντιστροφὰς (καὶ ὁ ἀριθμὸς οὗτος εἶναι ὁ μέγιστος δυνατὸς ἀριθμὸς ἀντιστροφῶν διὰ σειρὰς n ἀριθμῶν).

Ὅθεν τὸ εὖρος τῶν τιμῶν τῆς μεταβλητῆς λ εἶναι ἀπὸ 0 ἕως $\frac{n(n-1)}{2}$ διὰ κάθε n.

Ἐξ ἄλλου τὸ λ δύναται νὰ λάβῃ οἱανδήποτε ἀκεραίαν τιμὴν μεταξὺ 0 καὶ τοῦ $\frac{n(n-1)}{2}$, διότι ἐὰν εἰς τὴν σειρὰν 1, 2, 3 .... n, ἥτις ἔχει 0 διαδρομάς, ἐναλλάξωμεν τὰς θέσεις τοῦ n καὶ τοῦ ἀμέσως προηγουμένου του στοιχείου (n–1), θὰ ἔχωμεν τὴν διάταξιν $1, 2, 3 \ldots n-2, n, n-1$ εἰς ἣν $\lambda=1$.

Ἐὰν νῦν εἰς τὴν νέαν διάταξιν ἐναλλάξωμεν τὰς θέσεις τοῦ n καὶ τοῦ ἀμέσως προηγουμένου του (n–2), θὰ ἔχωμεν τὴν διάταξιν :

$$1, 2, 3, \ldots (n-3), n, (n-2), (n-1) \text{ εἰς } \lambda=2.$$

Προχωροῦντες κατὰ τὸν αὐτὸν τρόπον, φθάνομεν εἰς τὴν διάταξιν :

$$n, 1, 2, 3, \ldots (n-1) \text{ εἰς ἣν } \lambda=n-1. \quad (20)$$

τοῦ λ διατρέξαντος ἤδη διαδοχικῶς ὅλας τὰς ἀκεραίας τιμὰς ἀπὸ τοῦ 0 ἕως τοῦ $(n-1)$.

Ἤδη συνεχίζομεν τὰς ἰδίας ἐναλλαγὰς θέσεων εἰς τὴν (20) μεταξὺ τοῦ στοιχείου $(n-1)$ καὶ τοῦ ἀμέσως προηγουμένου του μέχρι τῆς διατάξεως:

$$n,\ n-1,\ 1,\ 2,\ 3 \ldots .\ n-2. \qquad (21)$$

Συνεχίζοντες κατὰ τὸν ἴδιον τρόπον ἐναλλαγὰς τοῦ στοιχείου $(n-2)$ καὶ εἶτα ὁμοίως διὰ τὸ $(n-3)$ κλπ. θὰ φθάσωμεν εἰς τὴν διάταξιν:

$$n,\ (n-1),\ (n-2) \ldots\ldots 3,\ 2,\ 1, \text{ ἥτις ἔχει } \lambda = \frac{n\,(n-1)}{2}$$

Οὕτω τὸ λ διέτρεξε διαδοχικῶς ὅλας τὰς ἀκεραίας τιμὰς μεταξὺ τοῦ 0 καὶ τοῦ $\dfrac{n\,(n-1)}{2}$

Θὰ δείξωμεν ἤδη ὅτι εἰς κάθε διάταξιν τῶν n ἀριθμῶν, ποὺ ἔχει λ_0 ἀντιστροφάς, ἀντιστοιχεῖ μία διάταξις μὲ $\dfrac{n\,(n-1)}{2} - \lambda_0$ ἀντιστροφάς, δηλαδὴ ὅτι διὰ τιμὰς τοῦ λ ἰσαπεχούσας ἀπὸ τὰς ἀκραίας τιμὰς 0 καὶ $\dfrac{n\,(n-1)}{2}$ ἔχομεν ἴσας συχνότητας εἰς τὴν κατανομὴν συχνότητος τοῦ λ.

Πράγματι ἀρκεῖ νὰ παρατηρήσωμεν ὅτι γενικῶς εἰς τὸ σύστημα ἀριθμήσεως κάθε ἀκέραιος ἀριθμὸς K ἔχει $(K-1)$ ἀκεραίους μικροτέρους του. Ἂν δὲ $K \leqslant n$ εἰς μίαν τυχοῦσαν διάταξιν τῶν n ἀριθμῶν, π.χ. τὴν

$$\alpha_1,\ \alpha_2,\ \alpha_3 \ldots\ \alpha_\rho \ldots\ \alpha_{n-1},\ \alpha_n \qquad (22)$$

(ὅπου τὰ α_i εἶναι οἱ ἀριθμοὶ $1, 2, \ldots\ldots n$ κατά τινα τάξιν), ὁ K θὰ εἶναι ἓν ἐκ τῶν στοιχείων ταύτης ἔστω τὸ α_ρ (δηλαδὴ $K = \alpha_\rho$). Οἱ $(K-1)$ ἢ $\alpha_{\rho}-1$ ἀριθμοὶ ποὺ εἶναι εἰς τὸ σύστημα ἀριθμήσεως μικρότεροι τοῦ α_ρ θὰ εὑρίσκωνται εἰς τὴν (1) εἴτε ἀριστερὰ τοῦ α_ρ ἢ δεξιά του ἢ μερικοὶ ἀριστερά του καὶ οἱ ὑπόλοιποι δεξιά του. Ἐν πάσῃ περιπτώσει, ἂν τὸ πλῆθος ἐκείνων ποὺ εὑρίσκονται ἀριστερά του καλέσωμεν K_ρ καὶ ἐ κ ε ί ν ω ν π ο ὺ ε ὑ ρ ί σ κ ο ν τ α ι δ ε ξ ι ά τ ο υ λ_ρ θὰ ἔχωμεν πάντοτε

$$\alpha_\rho - 1 = K_\rho + \lambda_\rho \quad (\text{ὅπου } K_\rho \geqslant 0 \text{ καὶ } \lambda_\rho \geqslant 0)$$

Τὸ αὐτὸ ἰσχύει δι' ὅλα τὰ στοιχεῖα τῆς (22) καὶ συνεπῶς θὰ εἶναι

$$\begin{aligned}
\alpha_1 - 1 &= K_1 + \lambda_1 \\
\alpha_2 - 1 &= K_2 + \lambda_2 \\
\alpha_3 - 1 &= K_3 + \lambda_3 \\
&\vdots \\
\alpha_n - 1 &= K_n + \lambda_n
\end{aligned}$$

$$\text{ἄρα } \sum_{i=1}^{n} (\alpha_i\ ) - n = \sum_{i=1}^{n} K_i + \sum_{i=1}^{n} \lambda_i$$

$$\text{ἢ } \frac{n(n+1)}{2} - n = \sum_{i=1}^{n} K_i + \lambda. \qquad (23)$$

Ἂν ἤδη θεωρήσωμεν τὴν ἀκριβῶς κατ' ἀντίστροφον τάξιν τῆς (22) διάταξιν

$$\alpha_n,\ \alpha_{n-1},\ \alpha_{n-2} \ldots\ldots \alpha_3,\ \alpha_2,\ \alpha_1 \qquad (24)$$

θὰ ἔχωμεν διὰ κάθε στοιχεῖον της, π.χ. τὸ $α_ρ$, ὅτι οἱ ἀριθμοὶ ποὺ εἶναι μικρότεροι τοῦ $α_ρ$ καὶ εὑρίσκονται δεξιά του εἶναι ἀκριβῶς οἱ ἀριθμοὶ ποὺ εἰς τὴν (22) εἶναι μικρότεροι τοῦ $α_ρ$ καὶ εὑρίσκονται ἀριστερά του, δηλαδὴ εἰς τὴν (24) θὰ εἶναι : $\sum_{i=1}^{n} \lambda'_i = \sum_{i=1}^{n} K_i$ ἀλλὰ $\sum_{i=1}^{n} \lambda'_i = \lambda'$, ὅπου λ' εἶναι τὸ πλῆθος τῶν ἀντιστροφῶν τῆς (24), ἄρα θὰ ἔχωμεν ἕνεκα τῆς (23) $\frac{n(n-1)}{2} - \lambda = \lambda'$.

Ἑπομένως, ἐπειδὴ εἰς κάθε διάταξιν (22) ἀντιστοιχεῖ μία μόνη διάταξις (24), ἀπεδείχθη ὅτι εἰς κάθε διάταξιν μὲ λ_0 ἀντιστροφὰς θὰ ἀντιστοιχῇ μία διάταξις μὲ $\lambda'_0 = \frac{n(n-1)}{2} - \lambda_0$ ἀντιστροφάς, ἤτοι ὅτι αἱ τοιαῦται διατάξεις εἶναι ἰσάριθμοι.

Ἀπεδείχθη οὕτω ὅτι ἡ κατανομὴ συχνότητος τοῦ λ εἶναι μία κατανομὴ συμμετρικὴ ὡς πρὸς τὰς ἀκραίας τιμὰς τοῦ εὕρους της καὶ συνεπῶς ὅτι ἡ μέση τιμὴ τοῦ λ θὰ εὑρίσκεται ἀκριβῶς εἰς τὸ μέσον τοῦ εὕρους της, δηλαδὴ θὰ εἶναι

$$\bar{\lambda} = \frac{n(n-1)}{4} \tag{25}$$

Γενικῶς ἡ κατανομὴ συχνότητος τοῦ ἀριθμοῦ τῶν ἀντιστροφῶν λ διὰ κάθε n δύναται νὰ εὑρεθῇ ὡς ἀκολούθως:

Ἔστω ὅτι γνωρίζομεν τελείως τὴν κατανομὴν συχνότητος τοῦ λ διὰ $n=n_0-1$ καὶ τὸν ἀριθμὸν τῶν περιπτώσεων καθ' ἃς $\lambda=\lambda_0-1$ διὰ $n=n_0$ καὶ θέλομεν νὰ εὕρωμεν τὸν ἀριθμὸν τῶν περιπτώσεων καθ' ἃς $\lambda=\lambda_0$ διὰ $n=n_0$.

Θὰ δείξωμεν ὅτι :

$$\text{Περιπτώσεις } (\lambda=\lambda_0,\ n=n_0) = \text{Περιπτ.}(\lambda=\lambda_0,\ n=n_0-1)+$$
$$\text{Περιπτ. } (\lambda=\lambda_0-1,\ n=n_0) - \text{Περιπτ. } (\lambda=\lambda_0-n_0,\ n=n_0-1) \tag{26}$$

ὅπου ἐνδεχομένως τινὲς ἐκ τῶν προσθετέων τοῦ δευτέρου μέλους εἶναι ἴσοι μὲ μηδέν.

Ἀπόδειξις : Ἐὰν λάβω ὅ λ α ς τὰς διατάξεις καθ' ἃς $\lambda=\lambda_0$ διὰ $n=n_0-1$ (ἐὰν ὑπάρχωσι τοιαῦται) καὶ τοποθετήσω τὸν ἀριθμὸν n_0 εἰς τὸ τέλος ἑκάστης ἐξ αὐτῶν θὰ ἔχω ὅ λ α ς τὰς διατάξεις ποὺ ἔχουν $\lambda=\lambda_0$, $n=n_0$ κ α ὶ τὸ n_0 ὡ ς τ ε λ ε υ τ α ῖ ο ν σ τ ο ι χ ε ῖ ο ν τ ω ν. Ἔστω ὁ ἀριθμὸς τούτων K_1.

Ἐὰν ἐπίσης λάβω ὅ λ α ς τὰς διατάξεις καθ' ἃς $\lambda=\lambda_0-1$ διὰ $n=n_0$ π λ ὴ ν ἐκείνων ἐξ αὐτῶν ποὺ ἔχουν τὸ n_0 εἰς τὴν πρώτην θέσιν (ἐὰν ὑπάρχωσι τοιαῦται) καὶ ἐναλλάξω τὴν θέσιν τοῦ n_0 καὶ τ ο ῦ π ρ ο η γ ο υ μ έ ν ο υ τοῦ n_0 στοιχείου, θὰ εὕρω διατάξεις μὲ $n=n_0$ καὶ $\lambda=\lambda_0$ δ ι α φ ό ρ ο υ ς ἐ κ ε ί ν ω ν τ ὰ ς ὁ π ο ί α ς ε ὕ ρ ο μ ε ν π ρ ο η γ ο υ μ έ ν ω ς (διότι αἱ προηγούμεναι ἔχουν τελευταῖον στοιχεῖον των τὸ n_0) καὶ ἔστω ὁ ἀριθμὸς τούτων K_2.

Ἤδη λέγω ὅτι : Περιπτώσεις $(\lambda=\lambda_0,\ n=n_0) = K_1+K_2$ διότι ἂν φαντασθῶ τυχοῦσαν τοιαύτην διάταξιν μὲ $\lambda=\lambda_0$ καὶ $n=n_0$ ἡ διάταξις αὕτη θὰ ἔχῃ τὸ n_0 ἢ εἰς τὸ τέλος της, ὅτε ἀφαιρῶν τὸ n_0 ἔχω μίαν διάταξιν ἐκ τῶν K_1, καὶ συνεπῶς τὴν ἔχω ἤδη μετρήσει εἰς τὰς K_1, ἢ θὰ ἔχῃ τὸ n_0 εἰς ἄλλην τινὰ θέσιν καὶ ἂν ἐναλλάξωμεν εἰς ταύτην τὴν θέσιν τοῦ n_0 μὲ τὸ ἀμέσως ἑπόμενον στοιχεῖον θὰ προκύψῃ μία διάταξις μὲ $\lambda=\lambda_0-1$ καὶ $n=n_0$, δηλαδὴ μία ἐκ τῶν K_2 ἣν ἐπίσης ἔχομεν ἤδη μετρήσει.

Ὁ ἀριθμὸς K_1 εἶναι ἐξ ὑποθέσεως γνωστός, ἐφ' ὅσον ὑπετέθη ὅτι γνωρίζομεν τελείως τὴν κατανομὴν τοῦ λ διὰ $n=n_0-1$. Ὁ ἀριθμὸς ὅμως K_2 ἰσοῦται μὲ τὸν ἀριθμὸν τῶν διατάξεων αἵτινες ἔχουν $\lambda=\lambda_0-1$ καὶ $n=n_0$, ἔστω K_3, πλὴν ἐκείνων ἐξ αὐτῶν ποὺ ἔχουν τὸ n_0 εἰς τὴν πρώτην θέσιν (ἐὰν ὑφίστανται τοιαῦται).

Ἐὰν τοιαῦται περιπτώσεις ὑφίστανται καὶ ὑποτεθῇ ὅτι ὁ ἀριθμός των εἶναι K_4 τότε θὰ ὑφίστανται ἰσάριθμοι διατάξεις μὲ $n=n_0-1$ καὶ $\lambda=\lambda_0-n_0$. Διότι, ἂν ἀφαιρέσω ἀπὸ τὰς διατάξεις $\lambda=\lambda_0-1$ καὶ $n=n_0$ τὸ n_0 ποὺ κατέχει τὴν πρώτην θέσιν, θὰ ἔχω περιπτώσεις μὲ $n=n_0-1$ καὶ $\lambda=\lambda_0-1-(n_0-1)=\lambda_0-n_0$ ἀντιστροφὰς καὶ ἀντιστρόφως οἱαδήποτε διάταξις μὲ $n=n_0-1$ καὶ $\lambda=\lambda_0-n_0$ γίνεται διάταξις μὲ $\lambda=\lambda_0-1$ καὶ $n=n_0$ ἂν θέσω τὸ n_0 ὡς πρῶτον στοιχεῖον πρὸ τῶν (n_0-1) στοιχείων της.

Ἄρα θὰ εἶναι $K_2 = K_3 - K_4$ καὶ ἐπειδὴ οἱ ἀριθμοὶ K_3 καὶ K_4 εἶναι ἀριθμοὶ ἐξ ὑποθέσεως γνωστοὶ ὑπελογίσθη καὶ τὸ K_2, δηλαδὴ ἐδείχθη ἡ (26) [1].

Ἤδη παρατηροῦμεν ὅτι ὅταν $n = n_0 - 1 = 2$ ἡ κατανομὴ τοῦ λ εἶναι ἡ:

	συχνότης
$\lambda = 0$	1
$\lambda = 1$	1

(διότι ὅταν $n = 2$ αἱ δυναταὶ διατάξεις εἶναι 1,2 καὶ 2,1 μὲ ἀντιστροφὰς 0 καὶ 1 ἀντιστοίχως).

Ἐπίσης παρατηροῦμεν ὅτι δι' οἱονδήποτε n ὁ ἀριθμὸς τῶν περιπτώσεων καθ' ἃς ἔχομεν 0 ἀντιστροφὰς εἶναι 1 (δηλαδὴ ἡ κατάταξις 1, 2, 3 . . . . n). Οὕτω δυνάμεθα νὰ κατασκευάσωμεν πίνακας κατανομῆς συχνότητος τοῦ λ διὰ διαφόρους τιμὰς τοῦ λ. Ἔχομεν π.χ. τὸν πίνακα I (ἴδε τέλος παρούσης ἐργασίας) ὅταν $n = 10$.

Σημείωσις. Ἕνας πρακτικὸς τρόπος κατασκευῆς τῆς κατανομῆς συχνότητος τοῦ ἀριθμοῦ τῶν ἀντιστροφῶν λ διὰ $n=2, 3, 4 \ldots$ κλπ. εἶναι ὁ ἑξῆς: διὰ $n = 2$ ἔχομεν, ὡς ἐλέχθη, τὴν κατανομὴν $\lambda = 0$ συχνότης 1

$\lambda = 1$ » 1

Ἐὰν ἤδη γράψωμεν τρεῖς φορὰς τὰς συχνότητας τῆς περιπτώσεως $n=2$ ὡς κάτωθι καὶ προσθέσωμεν, ἔχομεν τὴν κατανομὴν συχνότητος τοῦ λ διὰ $n = 3$ $(\lambda = 0, 1, 2, 3)$

$$\left.\begin{array}{cccc} 1 & 1 & & \\ & 1 & 1 & \\ & & 1 & 1 \\ \hline 1, & 2, & 2, & 1 \end{array}\right\}$$

δηλαδὴ γράφομεν κάτωθι τῆς 1, 1 τὰ ἴδια ψηφία ἀλλὰ μίαν θέσιν δεξιώτερον, ὁμοίως τὴν τρίτην φορὰν γράφομεν 1, 1 μίαν εἰσέτι θέσιν δεξιώτερον καὶ εἶτα προσθέτομεν.

Ὁμοίως ἐὰν γράψωμεν τὴν κατανομὴν συχνότητος τοῦ λ διὰ $n = 3$ (δηλαδὴ τὴν εὑρεθεῖσαν 1, 2, 2, 1) τέσσαρας φορὰς κατ' ἀνάλογον τρόπον

---

(1) Μέθοδοι εὑρέσεως τῆς κατανομῆς τοῦ λ ἀνεπτύχθησαν ὑπὸ τοῦ M. C. Kendall καὶ ἄλλων. Ἡ ἀναπτυχθεῖσα ἐνταῦθα στοιχειώδης μέθοδος ἀποτελεῖ ἀτομικὴν ἐργασίαν τοῦ γράφοντος.

καὶ προσθέσωμεν, θὰ εὕρωμεν τὴν κατανομὴν τοῦ λ διὰ $n = 4$ ὡς κάτωθι :

$$
(\lambda = 0, 1, 2, 3, 4, 5, 6) \qquad
\begin{array}{ccccccc}
1 & 2 & 2 & 1 & & & \\
 & 1 & 2 & 2 & 1 & & \\
 & & 1 & 2 & 2 & 1 & \\
 & & & 1 & 2 & 2 & 1 \\
\hline
1 & 3 & 5 & 6 & 5 & 3 & 1
\end{array}
$$

Ὁμοίως ἡ κατανομὴ τοῦ λ διὰ $n = 5$ εὑρίσκεται ἐκ τοῦ ἀθροίσματος τῶν

$$
(\lambda = 0, 1, 2, \ldots, 10) \qquad
\begin{array}{ccccccccccc}
1 & 3 & 5 & 6 & 5 & 3 & 1 & & & & \\
 & 1 & 3 & 5 & 6 & 5 & 3 & 1 & & & \\
 & & 1 & 3 & 5 & 6 & 5 & 3 & 1 & & \\
 & & & 1 & 3 & 5 & 6 & 5 & 3 & 1 & \\
 & & & & 1 & 3 & 5 & 6 & 5 & 3 & 1 \\
\hline
1, & 4, & 9, & 15, & 20, & 22, & 20, & 15, & 9, & 4, & 1
\end{array}
$$

καὶ οὕτω καθεξῆς [1].

Ἡ κατανομὴ τοῦ λ προσεγγίζει τὴν κανονικὴν κατανομὴν πολὺ ταχέως, δηλαδὴ ὅταν τὸ n εἶναι μετρίως μεγάλον (π.χ. > 10).

Ἀναφέρεται ἄνευ ἀποδείξεως ὅτι

$$\sigma^2_\lambda = \frac{n(n-1)(2n+5)}{72}$$

Ὅθεν ὅταν τὸ n εἶναι > 10 δυνάμεθα, διὰ τὰ ἐπὶ πιθανοτήτων βασιζόμενα συμπεράσματά μας, νὰ χρησιμοποιήσωμεν πίνακας τῶν ἐμβαδῶν κανονικῆς κατανομῆς μὲ μέσον $\bar{\lambda} = \frac{n(n-1)}{4}$ καὶ $\sigma^2_\lambda = \frac{n(n-1)(2n+5)}{72}$

Δέον νὰ σημειώσωμεν ὅτι διὰ n ἀριθμοὺς τὸ σύνολον τῶν δυνατῶν μεταθέσεων εἶναι $n!$, ὅθεν τὸ ἄθροισμα τῶν συχνοτήτων τοῦ πίνακος I διὰ $n = 10$ εἶναι $10!$

Ἐὰν ἤδη ὑποτεθῇ ὅτι ὅλαι αἱ μεταθέσεις εἶναι ἐξ ἴσου πιθαναί, ἡ πιθανότης ἵνα εἰς ἓν τυχαῖον δεῖγμα ἐκ $n_0 = 10$ ἀριθμῶν ἔχωμεν διάταξιν εἰς ἣν $\lambda \leqslant \lambda_0$ ἢ $\lambda \geqslant \frac{n_0(n_0-1)}{2} - \lambda_0$ δύναται νὰ εὑρεθῇ ἐκ τοῦ πίνακος I ἐὰν λάβωμεν τὸ ἄθροισμα τῶν συχνοτήτων ὅλων τῶν περιπτώσεων καθ' ἃς

$$\lambda \leqslant \lambda_0 \text{ ἢ } \lambda \geqslant \frac{n_0(n_0-1)}{2} - \lambda_0$$

καὶ διαιρέσωμεν διὰ $n_0! = 10!$ (τὸ λ_0 ὑποτίθεται μικρότερον τοῦ $\bar{\lambda} = 22{,}5$).

Δυνάμεθα οὕτω νὰ σχηματίσωμεν τὸν πίνακα II ὅστις μᾶς δίδει τοιούτου εἴδους πιθανότητας ὑπὸ δεκαδικὴν μορφήν.

Ἃς ἐπανέλθωμεν νῦν εἰς τὴν περίπτωσιν δύο σειρῶν τιμῶν x_i καὶ y_i ἡ ὁποία μᾶς δίδει ἓν τυχαῖον δεῖγμα ἐξ ἑνὸς πληθυσμοῦ δύο μεταβλητῶν.

Θὰ ἔχωμεν n ζεύγη ἀντιστοίχων τιμῶν τῶν x καὶ y.

---

(1) Τοιοῦτοι πίνακες ὡς ὁ I καὶ II τοῦ τέλους τῆς παρούσης διὰ τιμὰ τοῦ n ἀπὸ 2 ἕως 10 ἔχουν ἤδη ὑπολογισθῆ. Ἴδε A. C. Rosander: Elementary principles of Statistics (1951).

Κατατάσσομεν τὰ xi κατὰ σειρὰν κατιόντος μεγέθους (κάτωθεν δὲ ἑκάστου xi τοποθετοῦμεν τὰ ἀντίστοιχα yi).

Ἀντικαθιστῶμεν εἶτα τὰ πραγματικὰ xi μεγέθη δι' ἀριθμῶν ἐκφραζόντων τὴν σειρὰν κατατάξεώς των κατὰ μέγεθος (διὰ τοῦ 1 τὸ μεγαλύτερον ὅλων κλπ.).

Θὰ ἔχωμεν οὕτω διὰ τὰ x τὴν διάταξιν

$$1 \quad 2 \quad 3 \ldots\ldots n \qquad (28)$$

Ἐὰν ἤδη ἀντικαταστήσωμεν καὶ τὰ yi (ὡς εἶναι ταῦτα γεγραμμένα κάτωθι τῶν ἀντιστοίχων xi) διὰ τῶν ἀριθμῶν ποὺ ἐκφράζουν τὴν σειρὰν κατατάξεώς των κατὰ μέγεθος θὰ ἔχωμεν εἰς τὴν θέσιν τῶν yi μίαν διάταξιν τῶν n ἀριθμῶν ἀντίστοιχον τῆς (28) ἥτις ἐν γένει θὰ εἶναι διάφορος τῆς (28).

Ἂν ἡ διάταξις τῶν y εἶναι ἀκριβῶς ἡ ἰδία μὲ τὴν 1, 2, 3 . . . n τῶν x τότε λέγομεν ὅτι τὰ x καὶ y παρουσιάζουν μίαν τελείαν θετικὴν συσχέτισιν.

Ἂν ἡ ἀντίστοιχος πρὸς τὴν (28) διάταξις τῶν y εἶναι ἡ n, (n − 1). . .3, 2, 1 τότε ἔχομεν μίαν τελείαν ἀρνητικὴν συσχέτισιν.

Παρατηροῦμεν ὅτι εἰς τὴν πρώτην περίπτωσιν ἔχομεν διὰ τὴν κατάταξιν τῶν y : $\lambda = o$, ἐνῷ εἰς τὴν δευτέραν περίπτωσιν $\lambda = \dfrac{n(n-1)}{2}$.

Γενικῶς παραδεχόμεθα ὅτι δύο κατατάξεις τῶν y κάτωθι τῶν x ποὺ ἔχουν τὸ αὐτὸ λ θὰ παρέχωσι τὸν αὐτὸν βαθμὸν συσχετίσεως τῶν x καὶ y. Διὰ τὴν μέτρησιν τῆς συσχετίσεως τῶν x καὶ y ἐκ τοῦ ἀριθμοῦ τῶν ἀντιστροφῶν χρησιμοποιοῦμεν τὸν τύπον :

$$r' = 1 - \frac{4\lambda}{n(n-1)} \qquad (29)$$

Ὁ τύπος (29) μᾶς παρέχει ἕνα συντελεστὴν συσχετίσεως r' βασιζόμενον εἰς τὸν ἀριθμὸν τῶν ἀντιστροφῶν τῶν διατάξεων τῶν y, ὅστις εὑρίσκεται ἐν ἁρμονίᾳ πρὸς τοὺς συνήθεις συντελεστὰς συσχετίσεως, διότι ὅταν ἡ σειρὰ τῶν y εἶναι ἡ 1, 2, 3 . . . . n (ὅτε $\lambda = o$) ὁ τύπος (29) μᾶς δίδει $r' = 1$.

Ὅταν δὲ ἡ σειρὰ τῶν y εἶναι n, n − 1, ... 3, 2, 1 τότε $\lambda = \dfrac{n(n-1)}{2}$ καὶ ὁ τύπος (29) δίδει $r' = -1$.

Εἰς ὅλας τὰς ἄλλας περιπτώσεις $-1 < r' < 1$.

Κρίσις τῆς ὑποθέσεως H_0 περὶ ἀνεξαρτησίας τῶν δύο μεταβλητῶν x καὶ y τοῦ πληθυσμοῦ. (Δοκιμασία δύο ἄκρων).

Ἔστω ἤδη ὅτι εἰς ἓν «τυχαῖον» δεῖγμα μεγέθους ($n = n_0$) εὕρομεν $\lambda = \lambda_0$, εἰς τὴν κατὰ τὸν περιγραφέντα τρόπον σχηματισθεῖσαν διάταξιν τῶν y. Ἐὰν $H_0 =$ ὅτι: αἱ μεταβληταὶ x καὶ y εἶναι ἀνεξάρτητοι εἰς τὸν πληθυσμόν, αἱ διατάξεις τῶν y αἱ ἀντίστοιχοι πρὸς τὴν (28) τῶν x εἶναι ἐξ ἴσου πιθαναί.

Ἡ ἐμφάνισις διατάξεως μὲ $\lambda = \lambda_0$ μᾶς ἐπιτρέπει νὰ ὑπολογίσωμεν τὴν πιθανότητα νὰ ὀφείλεται τοῦτο (ἢ αἱ σπανιώτεραι περιπτώσεις) εἰς τὴν τύχην (ἐὰν H_0 ἀληθής, ὅτε Ε (λ) ἢ $\bar{\lambda} = \dfrac{n_0(n_0-1)}{4}$).

Ἐὰν ἐκ τοῦ πίνακος II εὕρωμεν ὅτι πιθ $\left(\lambda \leqslant \lambda_0 \text{ ἢ } \lambda \geqslant \dfrac{n_0(n_0-1)}{2} - \lambda_0\right)$

εἶναι μικροτέρα ἢ ἴση πρὸς 5 % διὰ τὸ δοθὲν n_0 ἀπορρίπτομεν τὴν H_0 εἰς τὸ 5 % ἐπίπεδον σημαντικότητος.

Ἐὰν ὅμως πιθ $\left(\lambda \leqslant \lambda_0 \text{ ἢ } \lambda \geqslant \frac{n_0(n_0-1)}{2} - \lambda_0\right) > 5\,\%$ δὲν δυνάμεθα νὰ ἀπορρίψωμεν τὴν H_0 (περὶ ἀνεξαρτησίας τῶν x καὶ y) εἰς τὸ θεωρηθὲν ἐπίπεδον σημαντικότητος.

Κατὰ τὸν ὑπολογισμὸν τῆς πιθ $\left(\lambda \leqslant \lambda_0 \text{ καὶ } \lambda \geqslant \frac{n(n-1)}{2} - \lambda_0\right)$ λαμβάνομεν ὡς λ_0 τὸν μικρότερον τῶν δύο ἀριθμῶν λ_0 καὶ $\frac{n(n-1)}{2} - \lambda_0$. Ἂν π.χ. ἔχω $n = 5$ καὶ εὕρω μίαν διάταξιν τῶν y μὲ $\lambda = 7$, ἐπειδὴ $\frac{5\times 4}{2} - 7 = 3$ λαμβάνω τὴν πιθ. $(\lambda \leqslant 3 \text{ καὶ } \lambda \geqslant 7)$.

Παράδειγμα. Ἔστω ὅτι εἰς ἓν «τυχαῖον» δεῖγμα ἐκ 10 μαθητῶν ἑνὸς σχολείου ἡ σειρὰ ἐπιδόσεως τῶν 10 μαθητῶν συγκρινομένων μεταξύ των εἰς τὰ Μαθηματικὰ καὶ τὴν Ἱστορίαν ἔχει ὡς ἑξῆς :

Μαθηταί :	Α	Β	Γ	Δ	Ε	Ζ	Η	Θ	Ι	Κ
Μαθηματικά :	3ος	1ος	6ος	10ος	7ος	2ος	8ος	5ος	9ος	4ος
Ἱστορία :	9ος	8ος	3ος	2ος	1ος	10ος	5ος	7ος	6ος	4ος

Δυνάμεθα νὰ ἀπορρίψωμεν τὴν ὑπόθεσιν ὅτι ἡ ἐπίδοσις τῶν μαθητῶν εἰς ὁλόκληρον τὸ σχολεῖον εἰς τὰ Μαθηματικὰ εἶναι ἀνεξάρτητος τῆς τοιαύτης εἰς τὴν Ἱστορίαν εἰς τὸ 5 % ἐπίπεδον σημαντικότητος ;

Κατατάσσω τοὺς μαθητὰς κατὰ σειρὰν ἐπιδόσεως εἰς τὰ Μαθηματικὰ καὶ κάτωθεν αὐτῆς γράφω τὴν σειρὰν ἐπιδόσεώς των εἰς τὴν Ἱστορίαν.

Μαθηταί :	Β	Ζ	Α	Κ	Θ	Γ	Ε	Η	Ι	Δ
Μαθηματικά :	1	2	3	4	5	6	7	8	9	10
Ἱστορία :	8	10	9	4	7	3	1	5	6	2

ἡ διάταξις 8, 10, 9, 4, 7, 3, 1, 5, 6, 2 ἔχει $\lambda = 34$ ἀντιστροφάς. Ἓν τοιοῦτον γεγονὸς ἢ καὶ σπανιώτερα (ἐὰν H_0 ἦτο ἀληθὲς μὲ ἐναλλακτέα $r' > 0$ ἢ $r' < 0$) συμβαίνουσιν ἕνεκα τυχαίας διακυμάνσεως τῆς δειγματοληψίας μὲ πιθανότητα $< \frac{5}{100}$. Πράγματι εὑρίσκομεν ἐκ τοῦ πίνακος ΙΙ ὅτι :

$$\text{πιθ. } (\lambda \geqslant 34 \text{ ἢ } \lambda \leqslant 11) = 2 \times 0{,}023 = 4{,}6\,\%$$

Ἄρα τὸ ἀποτέλεσμα εἶναι στατιστικῶς σημαντικὸν εἰς τὸ 5 % ἐπίπεδον σημαντικότητος καὶ ἀπορρίπτομεν τὴν ὑπόθεσιν ὅτι ἡ ἐπίδοσις τῶν μαθητῶν εἰς τὰ Μαθηματικὰ εἶναι ἀνεξάρτητος τῆς ἐπιδόσεώς των εἰς τὴν Ἱστορίαν εἰς τὸ σχολεῖον τοῦτο εἰς τὸ 5 % ἐπίπεδον σημαντικότητος.

Αἱ ἐλεύθεραι κατανομῆς μέθοδοι ἰσχύουν ἄνευ οὐδεμιᾶς ὑποθέσεως ἐπὶ τῆς μορφῆς τῆς κατανομῆς συχνότητος τῶν παραμέτρων. Ἡ θεωρία των ἀπαιτεῖ τὰς βασικὰς γνώσεις τῆς θεωρίας τῶν πιθανοτήτων καὶ τῆς συνδυαστικῆς, ἀλλὰ ἡ ἐφαρμογή των, ὅταν ἰδίως ἔχωμεν εἰς τὴν διάθεσίν μας τοὺς σχετικοὺς πίνακας τῶν διαφόρων μεθόδων, εἶναι λίαν ἁπλῆ καὶ οὐδένα ἀπαιτεῖ ὑπολογισμὸν ἀλλὰ ἁπλᾶς ἀπαριθμήσεις. Ἐπιφυλασσόμεθα νὰ περιγράψωμεν μερικὰς εἰσέτι στατιστικὰς μεθόδους «ἐλευθέρας κατανομῆς» εἰς ἄλλην εὐκαιρίαν, ἰδίᾳ τὴν δοκιμασίαν τῶν σημείων (Sign test), ἥτις ἔχει πλείστας πρακτικὰς ἐφαρμογὰς εἰς τὴν Γεωργίαν καὶ λοιπὰς ἐπιστήμας.

ΠΙΝΑΞ I

Κατανομῆς συχνότητος τοῦ ἀριθμοῦ τῶν ἀντιστροφῶν διὰ n = 10.

λ	συχνότης	λ	συχνότης	λ	συχνότης	λ	συχνότης
0	1	12	47043	24	243694	36	13640
1	9	13	64889	25	230131	37	8095
2	44	14	86054	26	211089	38	4489
3	155	15	110010	27	187959	39	2298
4	440	16	135853	28	162337	40	1068
5	1068	17	162337	29	135853	41	440
6	2298	18	187959	30	110010	42	155
7	4489	19	211089	31	86054	43	44
8	8095	20	230131	32	64889	44	9
9	13640	21	243694	33	47043	45	1
10	21670	22	250749	34	32683	Σύνολον	10!
11	32683	23	250749	35	21670		(3628800)

ΠΙΝΑΞ II

Πιθανότητος ἐμφανίσεως λ ἀντιστροφῶν ἢ ὀλιγωτέρων (διὰ τιμὰς τοῦ λ μικροτέρας τοῦ $\bar{\lambda}$) καὶ τῆς πιθανότητος ἐμφανίσεως λ ἀντιστροφῶν ἢ περισσοτέρων (διὰ τιμὰς τοῦ λ μεγαλυτέρας τοῦ $\bar{\lambda}$) ὅταν n = 10.

λ	πιθανότης διὰ n=10	λ	πιθανότης διὰ n=10	λ	πιθανότης διὰ n=10	λ	πιθανότης διὰ n=10
0	0,0000003	12	0,036	24	0,431	36	0,008
1	0,000003	13	0,054	25	0,364	37	0,005
2	0,000015	14	0,078	26	0,300	38	0,002
3	0,00006	15	0,108	27	0,242	39	0,001
4	0,0002	16	0,146	28	0,190	40	0,0005
5	0,0005	17	0,190	29	0,146	41	0,0002
6	0,001	18	0,242	30	0,108	42	0,00006
7	0,002	19	0,300	31	0,078	43	0,000015
8	0,005	20	0,364	32	0,054	44	0,000003
9	0,008	21	0,431	33	0,036	45	0,0000003
10	0,014	22	0,500	34	0,023		
11	0,023	23	0,500	35	0,014		

Σημείωσις. Ἐπειδὴ $\bar{\lambda} = \frac{n(n-1)}{4} = 22{,}5$, ὁ πίναξ διὰ τιμὰς τοῦ $\lambda \leqslant 22$ δεικνύει τὴν πιθανότητα νὰ ἔχωμεν λ ἢ ὀλιγωτέρας ἀντιστροφάς, ἐνῷ διὰ τιμὰς τοῦ $\lambda \geqslant 23$ δεικνύει τὴν πιθανότητα νὰ ἔχωμεν λ ἢ περισσοτέρας ἀντιστροφάς. Προκειμένου περὶ ἀμφιπλεύρου κριτηρίου πρέπει νὰ λαμβάνωμεν τὸ ἄθροισμα τῶν πιθανοτήτων δύο ἰσαπεχουσῶν τιμῶν τοῦ λ ἀπὸ τὰς ἀκραίας 0 καὶ 45.