

10. ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

ΕΙΣΑΓΩΓΗ

Ένα πείραμα συχνάκις συνίσταται από επαναλαμβανομένης ανεξαρτή-
τους δοκιμάς, εκάστη τῶν ὁποίων ἔχει μόνον δύο ἐκβάσεις, τὴν ἐπιτυχίαν ἢ
τὴν ἀποτυχίαν. Τὰ πειράματα αὐτὰ εἶναι γνωστὰ ὡς διωνυμικά. Ἡ ἐπι-
τυχία ἢ ἀποτυχία εἰσοδοχῆς εἰς τὸ Πανεπιστήμιον καὶ ἡ ρίψις ἐνὸς κέρματος
εἶναι δύο διωνυμικά πειράματα.

Διὰ τὰ εἶναι ἓνα πείραμα διωνυμικόν, πρέπει νὰ ἰσχύουν αἱ κάτωθι
συνθῆκαι.

- (α) Τὸ πείραμα νὰ ἀποτελῇται ἀπὸ επαναλαμβανομένης δοκιμάς.
- (β) Αἱ επαναλαμβανόμεναι δοκιμαὶ νὰ εἶναι ἀνεξάρτητοι.
- (γ) Ἐκάστη δοκιμὴ νὰ ταξινομῇται ὡς «ἐπιτυχία» ἢ «ἀποτυχία».
- (δ) Ἡ πιθανότης ἐπιτυχίας νὰ παραμένῃ σταθερὰ ἀπὸ δοκιμῆς εἰς δοκιμὴν.

Ἄς φαντασθῶμεν N ἀνεξάρτητα γεγονότα (δοκιμάς) μὲ πιθανότητα
ἐπιτυχίας ἐκάστου γεγονότος p καὶ ἀποτυχίας $q = 1 - p$. Πρὸς ὑπολο-
γισμὸν τῆς πιθανότητος νὰ ἔχωμεν f ἐπιτυχίας εἰς τὰς N δοκιμάς, ὅπου f δύ-
ναται νὰ λάβῃ τὰς τιμὰς $0, 1, 2, \dots, N$, πειραματίζομαι μὲ δοχεῖον τὸ ὅποion
περιέχει 1 λευκὸν καὶ 3 μαύρους κύβους. Ὁ ἀριθμὸς τῶν ἄσπρων κύβων (1)
διαιρούμενος διὰ τοῦ πλῆθους ὅλων τῶν κύβων (4) μᾶς δίδει τὴν πιθανότητα
ἐπιτυχίας $p = \frac{1}{4}$ ἐξαγωγῆς λευκοῦ κύβου καὶ ὁ λόγος τῶν μαύρων κύβων (3)
διὰ τοῦ ἀριθμοῦ ὅλων τῶν κύβων μᾶς δίδει τὴν πιθανότητα ἀποτυχίας $q = \frac{3}{4}$.

Ἐξάγομεν ἐκ τοῦ δοχείου ἓνα κύβον, βλέπομεν τὸ χρῶμα του καὶ τὸν
τοποθετοῦμεν πάλιν εἰς τὸ δοχεῖον. Τοῦτο επαναλαμβάνεται μέχρις ὅτου
ἐξαχθοῦν N κύβοι. Ζητεῖται ἡ πιθανότης νὰ γίνῃ ἐξαγωγή f ἄσπρων κύβων
ἐκ τῶν N δοκιμῶν. Διὰ τὴν εὑρεσιν τῆς πιθανότητος καταγράφομεν ὅλας
τὰς δυνατὰς περιπτώσεις ἐξαγωγῆς λευκοῦ καὶ μαύρου κύβου, ὅταν $N = 4$
(πίν. 54). Ἐκάστη τετράδα ἐξαγωγῶν περιλαμβάνει f λευκοὺς κύβους
($f = 0, 1, 2, 3, 4$).

Ὁ ἀριθμὸς τῶν τετράδων αἱ ὁποῖαι περιέχουν f ἐπιτυχίας (ἐξαγωγή
λευκοῦ κύβου) δίδεται ἐκ τοῦ τύπου

$$10.1 \quad C_f^N = \frac{N(N-1)(N-2)\dots(N-f+1)}{1 \cdot 2 \cdot 3 \dots f}$$

Δοκιμαί	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Λήψεις	M	M	M	M	Λ	M	M	Λ	Λ	M	Λ	M	Λ	Λ	Λ	Λ
	M	M	M	Λ	M	M	Λ	Λ	M	Λ	M	Λ	M	Λ	Λ	Λ
	M	M	Λ	M	M	Λ	Λ	M	Λ	M	M	Λ	Λ	M	Λ	Λ
	M	Λ	M	M	M	Λ	M	M	M	Λ	Λ	Λ	Λ	Λ	M	Λ
Έμφάνισης λευκοῦ κύβου	0	1	1	1	1	2	2	2	2	2	2	3	3	3	3	4
	1		4					6					4			1

Πίναξ 54

Πράγματι, εάν θέλωμεν νὰ εὐρώμεν τὸν ἀριθμὸν τῶν τετράδων αἱ ὁποῖαι περιέχουν 2 λευκοὺς κύβους, θὰ ἀντικαταστήσωμεν τὰς τιμὰς $N = 4$, $f = 2$ εἰς τὸν τύπον 10.1

$$C_2^4 = \frac{4 \cdot 3}{1 \cdot 2} = 6$$

Ὁ ἀριθμὸς 6 εἶναι ὁ ἀριθμὸς ὁ ὁποῖος ἐμφαίνεται εἰς τὸν πίνακα 54.

Ἡ πιθανότης ἐμφανίσεως f λευκῶν κύβων εἰς ἐκάστην τετράδα δίδεται ὑπὸ τοῦ τύπου

$$p^f q^{N-f}$$

Ἐπειδὴ ὑπάρχουν ἰσοπίθανοι τετράδες ἐξαγωγῶν, συνεπάγεται ὅτι ἡ πιθανότης ἐμφανίσεως f λευκῶν κύβων ἐκ τῶν N δίδεται ὑπὸ τοῦ τύπου

$$C_f^N p^f q^{N-f}$$

Τὸ σύμβολον f τῶν ἐπιτυχιῶν καλεῖται διωνυμικὴ τυχαία μεταβλητή. Ἡ κατανομή τῆς τυχαίας μεταβλητῆς f , ὅπου f ὁ ἀριθμὸς τῶν δυνατῶν ἐπιτυχιῶν, πλησιάζει πολὺ τὴν κανονικὴν κατανομήν, νοουμένου ὅτι ἡ τιμὴ τοῦ p δὲν πλησιάζει τὸ μηδὲν καὶ τὴν μονάδα. Ἡ κατανομή αὕτη καλεῖται διωνυμικὴ (binomial distribution).

Ὁ ἀριθμητικὸς μέσος καὶ ἡ διασπορὰ τῆς διωνυμικῆς κατανομῆς δίδονται ἐκ τῶν τύπων

$$10.2 \quad \mu = Np$$

$$10.3 \quad \sigma^2 = Np q,$$

ή δὲ λοξότης καὶ κύρτωσις ἐκ τῶν τύπων

$$10.4 \quad \gamma_1 = \frac{p-q}{\sqrt{Npq}}$$

$$10.5 \quad \gamma_2 = \frac{1-6pq}{Npq}$$

ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Ἡ κατανομή δειγματοληψίας ποσοστῶν εἶναι δυνατόν νὰ προσεγγισθῇ πειραματικῶς ἢ θεωρητικῶς. Εἰς ἓνα μέγαν δοχεῖον ὑπάρχουν κέρματα τῶν 50 καὶ 100 μίλς. Ὑποθέτομεν ὅτι τὰ ποσοστὰ εἶναι ἀντιστοίχως p καὶ $q = 1 - p$. Ἐξάγομεν ἓνα ἀριθμὸν δειγμάτων πλήθους N καὶ καταγράφομεν κάθε φοράν τὰ ποσοστὰ ἐμφάνισεως κέρματος τῶν 50 μίλς καὶ σχηματίζομεν τοιοῦτοτρόπως τὴν κατανομὴν συχνότητος αὐτῶν τῶν ποσοστῶν. Ἡ κατανομή αὕτη καλεῖται **πειραματικὴ κατανομή δειγματοληψίας τῶν ποσοστῶν** (experimental sampling distribution of proportions).

Ἐὰν ὁ πληθυσμὸς μας εἶναι πεπερασμένος, πλήθους N_π στοιχείων καὶ N εἶναι ὁ πληθικὸς ἀριθμὸς τῶν δειγμάτων, τότε ἡ κατανομή συχνότητος τῶν ποσοστῶν ἐμφάνισεως τοῦ κέρματος τῶν 50 μίλς, ὅλων τῶν συνδυασμῶν τῶν N_π στοιχείων ἀνὰ N , μᾶς δίδει τὴν **θεωρητικὴν κατανομὴν δειγματοληψίας**, ἡ ὁποία κατανέμεται κανονικῶς μὲ ἀριθμητικὸν μέσον $\mu_p = p$. Τὸ τυπικὸν σφάλμα (τυπικὴ ἀπόκλισις) τῆς θεωρητικῆς κατανομῆς δειγματοληψίας τῶν ποσοστῶν, ὅταν τὸ $N_\pi < 30$, δίδεται ἐκ τοῦ τύπου

$$10.6 \quad \sigma_p = \sqrt{\frac{pq}{N} \cdot \frac{N_\pi - N}{N_\pi - 1}}$$

Ὅταν τὸ N_π τεῖνῃ εἰς μέγαν ἀριθμὸν, τότε ὁ ἀνωτέρω τύπος γράφεται

$$10.7 \quad \sigma_p = \sqrt{\frac{pq}{N}}$$

Ὑπάρχουν περιπτώσεις κατὰ τὰς ὁποίας καλούμεθα νὰ ἀποφασίσωμεν, ἐὰν αἱ παρατηρηθεῖσαι διαφοραὶ ποσοστῶν διαφέρουν σημαντικῶς. Πρὸς τοῦτο ἐὰν ὀρίσωμεν ὡς p_1 καὶ p_2 τὰ ποσοστὰ ἐπιτυχίας δύο πληθυσμῶν καὶ ὡς $\frac{f_1}{N_1}$, $\frac{f_2}{N_2}$ τὰ ποσοστὰ τῶν δύο δειγμάτων πλήθους N_1 καὶ N_2 , τὰ ὁποῖα ἐλήφθησαν ἐκ τῶν πληθυσμῶν, τότε ἡ κατανομή δειγματοληψίας τῶν διαφορῶν

$\frac{f_1}{N_1} - \frac{f_2}{N_2}$ κατά προσέγγισιν πλησιάζει τὴν κανονικὴν κατανομὴν, τῆς ὁποίας ὁ ἀριθμητικὸς μέσος ἰσοῦται μετὰ τὴν διαφορὰν $\mu_{p_1-p_2} = p_1 - p_2$ καὶ ἡ τυπικὴ ἀπόκλισις μετὰ

$$s_{p_1-p_2} = \sqrt{\frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}}$$

Πλέον συγκεκριμένως: Ὑποθέτομεν ὅτι τὰ p_1 καὶ p_2 εἶναι τὰ ποσοστὰ ἐπιτυχίας ὅλων τῶν μαθητῶν καὶ μαθητριῶν Σχολῶν Μέσης Παιδείας τῆς Κύπρου εἰς τὰ Μαθηματικά. Ἐὰν λάβωμεν ἓνα μεγάλον ἀριθμὸν δειγμάτων πλήθους N_1 καὶ N_2 διὰ τοὺς μαθητὰς καὶ μαθητρίας ἀντιστοίχως, τότε προφανῶς τὰ ποσοστὰ ἐπιτυχίας $\frac{f_1}{N_1}$ καὶ $\frac{f_2}{N_2}$ δὲν θὰ εἶναι ὅλα τὰ αὐτά. Ἡ κατανομὴ τῶν διαφορῶν $\frac{f_1}{N_1} - \frac{f_2}{N_2}$ μᾶς δίδει τὴν κατανομὴν δειγματοληψίας τῶν διαφορῶν τῶν ποσοστῶν.

Εἰς τὰς περιπτώσεις κατὰ τὰς ὁποίας τὸ πλῆθος τῶν δεδομένων εἶναι ἴσον ἢ μικρότερον τοῦ 25 καὶ ἰσχύει $p = q = \frac{1}{2}$, χρησιμοποιεῖται διαφορετικὴ προσέγγισις διὰ τὸν ἔλεγχον τῆς μηδενικῆς ὑποθέσεως H_0 , ὅπως φαίνεται εἰς τὸ κατωτέρω παράδειγμα.

Θεωροῦμεν 20 φοιτητάς, 10 ἐκ τῶν ὁποίων ἔχουν ληφθῇ τυχαίως διὰ νὰ ἀποτελέσουν τὴν ομάδα Α καὶ οἱ ὑπόλοιποι 10 τὴν ομάδα Β. Ἐπὶ πλέον ὑποθέτομεν ὅτι διὰ τὴν λύσιν ἑνὸς προβλήματος ὑπάρχουν δύο τρόποι. Ἀμφότεραι αἱ ομάδες ἐδιδάχθησαν νὰ λύουν τὸ πρόβλημα καὶ μετὰ τοὺς δύο. Ἀλλὰ ἡ ὁμάς Α ἐδιδάχθη πρῶτον νὰ τὸ λύῃ μετὰ τὸν τρόπον Α, ἡ δὲ ὁμάς Β μετὰ τὸν τρόπον Β.

Μετὰ παρέλευσιν ἑνὸς μηνὸς ἐδόθη παρόμοιον πρόβλημα καὶ εἰς τὰς δύο ομάδας πρὸς λύσιν. Ὡς H_0 θεωροῦμεν ὅτι δὲν ὑπάρχει διαφορὰ μεταξὺ τῆς πιθανότητος p_1 λύσεως τοῦ προβλήματος μετὰ τὸν τρόπον Α καὶ p_2 λύσεως τοῦ προβλήματος μετὰ τὸν τρόπον Β. Ὡς ἐναντίαν ὑπόθεσιν θεωροῦμεν τὴν $H_1: p_1 > p_2$. Αἱ διαφοραὶ μεταξὺ τῶν συχνοτήτων, αἱ ὁποῖαι δυνατόν νὰ παρουσιασθοῦν, πιστεύομεν ὅτι ὀφείλονται εἰς τὴν τύχην.

Ὁ κατωτέρω πίναξ μᾶς δίδει τὰ δεδομένα μετὰ τὴν λύσιν τοῦ προβλήματος.

	1ος τρόπος λύσεως	2ος τρόπος λύσεως	Σύνολον
Συχνότης	16	4	$N = 20$

Διὰ τὴν ἀπόρριψιν ἢ μὴ τῆς H_0 καταφεύγομεν εἰς τὸν πίνακα VII. Ὁ πίναξ VII μᾶς δίδει τὰς τιμὰς τῶν περιοχῶν ἀπορρίψεως ὧν τῶν x , ὅπου $x = 4$ εἶναι ὁ μικρότερος ἀριθμὸς ἐκ τῶν παρατηρηθεισῶν συχνοτήτων καὶ $N = 20$ εἶναι τὸ πλῆθος ὧν τῶν στοιχείων.

Διὰ μονόπλευρον ἔλεγχον εἰς τὸ ἐπίπεδον 1% καὶ μὲ τὰ στοιχεῖα $x = 4$, $N = 20$, μᾶς δίδεται ἡ τιμὴ $p = 0,006 < 1\% = 0,01$. Ἐπειδὴ $p < 0,01$, συμπεραίνομεν τὴν ἀπόρριψιν τῆς H_0 ὑπὲρ τῆς H_1 . Δηλ. μετὰ ἀπὸ ὠρισμένην περίοδον τὰ ἄτομα χρησιμοποιοῦν συχνότερον τὸν τρόπον λύσεως τὸν ὁποῖον ἐδιδάχθησαν πρῶτον.

Διὰ ἀμφίπλευρον ἔλεγχον αἱ διάφοροι τιμαὶ τοῦ πίνακος VII πολλαπλασιάζονται ἐπὶ δύο. Οὕτω διὰ $N = 15$ καὶ $x = 3$ λαμβάνεται ἡ τιμὴ

$$p = 2 \cdot 0,018 = 0,036$$

ΕΛΕΓΧΟΣ ΠΟΣΟΣΤΩΝ

Ἐνας καθηγητὴς τῶν Φυσικῶν ἰσχυρίζεται ὅτι τοῦλάχιστον 70% τῶν μαθητῶν τοὺς ὁποίους διδάσκει λαμβάνουν εἰς τὰς τελικὰς γραπτὰς ἐξετάσεις βαθμοὺς μεγαλυτέρους τοῦ 15. Καλοῦμεθα νὰ ἐλέγξωμεν τὴν ἀλήθειαν τοῦ ἰσχυρισμοῦ του.

Συμβολίζομεν μὲ $p = 0,70$ τὴν πιθανότητα ἐπιτυχίας συμφώνως πρὸς τὸν ἰσχυρισμὸν τοῦ καθηγητοῦ. Δίδεται δὲ πρὸς ἔλεγχον ἡ $H_0: p \leq 0,70$ ἐναντίον τῆς $H_1: p > 0,70$. Χρησιμοποιεῖται μονόπλευρος ἔλεγχος καὶ ὄχι ἀμφίπλευρος, διότι ὁ καθηγητὴς ἰσχυρίζεται ὅτι «ἔχει τοῦλάχιστον 70% ἐπιτυχίαν».

Λαμβάνομεν δείγμα $N = 150$ μαθητῶν, ἐξ ἐκείνων τοὺς ὁποίους διδάσκει, καὶ εὐρίσκομεν ὅτι $X = 112$ μαθηταὶ ἔλαβον βαθμοὺς μεγαλυτέρους τοῦ 15 εἰς τὴν Φυσικὴν.

Ἀντικαθιστῶμεν εἰς τὸν τύπον

$$10. \quad z = \frac{X - Np}{\sqrt{Npq}} \quad \begin{array}{l} \text{ὅπου } X = 112, \text{ περιπτώσεις ἐπιτυχίας εἰς τὸ δείγμα,} \\ N = 150, \text{ πληθικὸς ἀρ. δείγματος,} \\ p = 0,70, \text{ ποσοστὸν ἐπιτυχίας ἰσχυρισμοῦ,} \\ q = 1 - p = 0,30 \end{array}$$

$$z = \frac{112 - 150 \cdot 0,70}{\sqrt{150 \cdot 0,70 \cdot 0,30}}$$

$$z = \frac{7}{\sqrt{31,50}} = \frac{7}{5,61} = 1,247$$

Ἐπειδὴ $z = 1,247 < 1,64$ εἰς τὸ ἐπίπεδον 5%, δεχόμεθα τὴν μηδενικὴν ὑπόθεσιν, δηλ. ὁ ἰσχυρισμὸς τοῦ καθηγητοῦ τῶν Φυσικῶν δὲν ἀληθεύει. Ἐὰν ἴσχυεν ἡ ἀνισότης $z > 1,64$, τότε θὰ ἐδεχόμεθα αὐτὸ τὸ ὁποῖον ἰσχυρίζεται.

ΕΛΕΓΧΟΣ ΔΙΑΦΟΡΑΣ ΔΥΟ ΑΝΕΞΑΡΤΗΤΩΝ ΠΟΣΟΣΤΩΝ

Ἐκ δύο πληθυσμῶν, Ἀρρένων καὶ Θηλέων, λαμβάνομεν δύο δείγματα πλήθους N_1 καὶ N_2 . Προφανῶς τὰ δύο δείγματα εἶναι ἀνεξάρτητα. Νοούμενου ὅμως ὅτι f_1 εἶναι ἡ συχνότης ἐπιτυχίας εἰς τὸ δείγμα N_1 καὶ f_2 εἶναι ἡ συχνότης ἐπιτυχίας εἰς τὸ δείγμα N_2 , τὰ ἀντίστοιχα ποσοστὰ ἐπιτυχίας εἶναι

$$p_1 = \frac{f_1}{N_1} \quad \text{καὶ} \quad p_2 = \frac{f_2}{N_2}$$

Τὸ ἐρώτημα τὸ ὁποῖον τίθεται εἶναι κατὰ πόσον ἡ διαφορὰ $p_1 - p_2$ εἶναι στατιστικῶς σημαντικὴ ἢ μήπως εἶναι ἀποτέλεσμα τύχης;

Τὸ τυπικὸν σφάλμα τῆς διαφορᾶς δύο ποσοστῶν, τὰ ὁποῖα βασίζονται εἰς ἀνεξάρτητα δείγματα, ὑπολογίζεται ἐκ τοῦ τύπου

$$10_{.10} \quad s_{p_1 - p_2} = \sqrt{pq \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

$$\delta\text{που} \quad p = \frac{f_1 + f_2}{N_1 + N_2} \quad \text{καὶ} \quad q = 1 - p$$

Ἐλέγχοντες τὴν διαφορὰν $p_1 - p_2$ διαιροῦμεν αὐτὴν διὰ τοῦ τυπικοῦ σφάλματος

$$10_{.11} \quad z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

Ἡ ἐφαρμογὴ τοῦ τύπου προϋποθέτει ὅτι τὸ πλῆθος τῶν δειγμάτων N_1 καὶ N_2 εἶναι μέγας ἀριθμὸς καὶ ἡ τιμὴ τοῦ p δὲν πλησιάζει πολὺ τὸ μηδέν ἢ τὴν μονάδα.

Ἐπιθυμοῦντες νὰ ἀποφύγωμεν τὰ ὅρια ἐντὸς τῶν ὁποίων πρέπει νὰ εὐρίσκωνται τὰ N_1 καὶ N_2 καὶ τὸ p , ἀναφέρομεν ἓνα πρακτικὸν τρόπον ὃ ὁποῖος εἶναι ἀπαραίτητος προϋπόθεσις διὰ τὴν ἐφαρμογὴν τοῦ τύπου 10.11. Ἐὰν τὸ γινόμενον τῆς μικροτέρας τιμῆς ἐκ τῶν p καὶ q ἐπὶ τὴν μικροτέραν τιμὴν ἐκ τῶν N_1 καὶ N_2 εἶναι ἀριθμὸς μεγαλύτερος τοῦ 5, τότε ἡ ἐφαρμογὴ τοῦ τύπου εἶναι ἐπιτρεπτή.

Δι' ἀμφίπλευρον ἔλεγχον ἡ μηδενικὴ ὑπόθεσις H_0 ἀπορρίπτεται, ὅταν $z < -1,96$ ἢ $z > 1,96$ εἰς τὸ ἐπίπεδον 5%. Διὰ τὴν ἀπόρριψιν τῆς H_0 εἰς τὸ ἐπίπεδον 1% μᾶς χρειάζονται τιμαὶ $z < -2,58$ ἢ $z > 2,58$.

Διὰ μονόπλευρον ἔλεγχον χρειάζονται τιμαὶ τοῦ $|z| > 1,64$ ἢ 2,33 εἰς τὸ ἐπίπεδον 5% καὶ 1% ἀντιστοίχως.

Ὑποθέτομεν ὅτι ἀπὸ ἓνα δῆγμα $N_1 = 100$ μαθητῶν εἰσήχθησαν εἰς τὸ

πανεπιστήμιον $f_1 = 65$ καὶ ἐξ ἑτέρου δείγματος $N_2 = 120$ μαθητριῶν εἰσῆχθησαν $f_2 = 30$. Ἡ διαφορὰ τῶν ποσοστῶν εἰσαγωγῆς μαθητῶν καὶ μαθητριῶν εἶναι στατιστικῶς σημαντικὴ ἢ αἱ διαφοραὶ αἱ ὁποῖαι ὑπάρχουν εἶναι ἀποτέλεσμα τὸ ὁποῖον ὀφείλεται εἰς τὴν δειγματοληψίαν;

Τὰ ποσοστὰ διὰ τοὺς μαθητὰς καὶ τὰς μαθητρίδας εἶναι

$$p_1 = \frac{f_1}{N_1} = \frac{65}{100} = 0,65 \quad \text{καὶ}$$

$$p_2 = \frac{f_2}{N_2} = \frac{30}{120} = 0,25$$

Αἱ τιμαὶ τῶν p καὶ q εὐρίσκονται ἐκ τῶν τύπων

$$p = \frac{f_1 + f_2}{N_1 + N_2} = \frac{65 + 30}{100 + 120} = \frac{95}{220} = 0,43$$

$$q = 1 - p = 1 - 0,43 = 0,57$$

Πρὶν γίνῃ ἡ ἀντικατάστασις εἰς τὸν τύπον 10.11 ὀφείλομεν νὰ ἐλέγξωμεν ἐὰν μᾶς ἐπιτρέπεται ἡ ἐφαρμογὴ του.

$$\begin{aligned} \text{Ἐπειδὴ} \quad p &= 0,43 < 0,57 = q \quad \text{καὶ} \\ N_1 &= 100 < 120 = N_2 \end{aligned}$$

πολλαπλασιάζομεν τὰς ἀντιστοίχους μικροτέρας τιμὰς

$$p \cdot N_1 = 0,43 \cdot 100 = 43 > 5$$

Τὸ γινόμενον τοῦτο εἶναι μεγαλύτερον τοῦ 5 καὶ ὥς ἐκ τούτου ὁ τύπος 10.11 δύναται νὰ ἐφαρμοσθῇ.

$$z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

$$z = \frac{0,65 - 0,25}{\sqrt{0,43 \cdot 0,57 \left(\frac{1}{100} + \frac{1}{120} \right)}} = \frac{0,40}{\sqrt{0,0045}} = \frac{0,40}{0,067} = 5,97$$

$$\begin{aligned} \text{Ἐπειδὴ} \quad z &= 5,97 > 1,96 \quad \text{εἰς τὸ ἐπίπεδον } 5\% \\ z &= 5,97 > 2,58 \quad \text{εἰς τὸ ἐπίπεδον } 1\%, \end{aligned}$$

συμπεραίνομεν ὅτι αἱ διαφοραὶ τῶν ποσοστῶν ἐπιτυχίας μεταξύ μαθητριῶν

καὶ μαθητῶν δὲν ὀφείλονται μόνον εἰς τὴν δειγματοληψίαν, ἀλλὰ καὶ εἰς ἄλλους λόγους, οἱ ὅποιοι ὑπείσρχονται καὶ δίδουν εἰς τοὺς μαθητὰς μεγαλυτέραν ἐπιτυχίαν ἀπὸ ὅσων εἰς τὰς μαθητρίας.

Δι' ἀνεξάρτητα δείγματα μόνον δυνάμεθα νὰ ἐλέγξωμεν τὴν σημαντικότητα τῆς διαφορᾶς μεταξὺ ποσοστῶν καὶ διὰ τῆς ἐφαρμογῆς τοῦ κριτηρίου χ^2 (βλ. σελ. 174).

ΕΛΕΓΧΟΣ ΔΙΑΦΟΡΑΣ ΔΥΟ ΣΥΣΧΕΤΙΣΜΕΝΩΝ ΠΟΣΟΣΤΩΝ

Ὑπάρχουν ἔρευναι εἰς τὰς ὁποίας εἶναι δυνατόν νὰ ζητηθῇ ὁ ἔλεγχος τῆς διαφορᾶς δύο **συσχετισμένων ποσοστῶν** (correlated proportions). Τοῦτο σημαίνει ὅτι διὰ κάθε στοιχεῖον τοῦ δείγματος πρέπει νὰ ληφθοῦν δύο τιμαί. Προφανῶς θὰ ὑπάρχῃ κάποια συσχέτισις μεταξὺ τῶν διαφόρων ζευγῶν καὶ ὥς ἐκ τούτου ὀφείλομεν νὰ λάβωμεν τοῦτο ὑπ' ὄψιν εἰς τὸν ἔλεγχον τῆς διαφορᾶς.

Ὑποθέτομεν ὅτι ἓνα δεῖγμα πλήθους N ἀπαντᾷ εἰς δύο ἐρωτήσεις ἐνὸς "true-false" test. Συμβολίζομεν μὲ A τὸν ἀριθμὸν τῶν μαθητῶν οἱ ὅποιοι ἀπῆντησαν ὀρθῶς, εἰς ἀμφοτέρας τὰς ἐρωτήσεις καὶ μὲ Δ τὸν ἀριθμὸν τῶν μαθητῶν οἱ ὅποιοι ἔχουν δώσει μὴ ὀρθὰς ἀπαντήσεις. Μὲ B καὶ Γ συμβολίζομεν τοὺς ἀριθμοὺς τῶν μαθητῶν οἱ ὅποιοι ἔχουν ἀπαντήσῃ ὀρθῶς εἰς μίαν ἐκ τῶν δύο ἐρωτήσεων.

Σχηματίζομεν τὸν πίνακα συχνότητων

		Ἐρώτησις 1		
Ἐρώτησις 2		Ἐπιτυχία	Ἀποτυχία	
	Ἐπιτυχία	A	B	$A + B$
	Ἀποτυχία	Γ	Δ	$\Gamma + \Delta$
		$A + \Gamma$	$B + \Delta$	N

Πίναξ 56

καὶ τὸν πίνακα τῶν ποσοστῶν.

		Ἑρώτησις 1		
		Ἐπιτυχία	Ἀποτυχία	
Ἑρώτησις 2	Ἐπιτυχία	α	β	p_2
	Ἀποτυχία	γ	δ	q_2
		p_1	q_2	1,00

$$\alpha = \frac{A}{N}$$

$$\beta = \frac{B}{N}$$

$$p_2 = \frac{A+B}{N}$$

κ. ο. κ.

Πίναξ 57

Τὸ τυπικὸν σφάλμα τῶν διαφορῶν δύο συσχετισμένων ποσοστῶν δίδεται ὑπὸ τοῦ τύπου

$$10_{.12} \quad s_{p_1 - p_2} = \sqrt{\frac{\beta + \gamma}{N}}$$

Διὰ $B + \Gamma > 20$ ὁ ἔλεγχος σημαντικότητος τῶν διαφορῶν παρέχεται ὑπὸ τοῦ τύπου

$$10_{.13} \quad z = \frac{p_1 - p_2}{\sqrt{\frac{\beta + \gamma}{N}}}$$

Ἀπορρίπτεται ἡ μηδενικὴ ὑπόθεσις H_0 δι' ἀμφίπλευρον ἔλεγχον, ὅταν

$$|z| > 1,96 \quad \text{εἰς τὸ ἐπίπεδον } 5\% \text{ καὶ}$$

$$|z| > 2,58 \quad \text{εἰς τὸ ἐπίπεδον } 1\%$$

Ἡ ἀντικατάστασις τῶν p_1 , p_2 , β καὶ γ μᾶς ἀπλοποιεῖ τὸν τύπον 10.13 εἰς τὸν

$$10_{.14} \quad z = \frac{\Gamma - B}{\sqrt{B + \Gamma}}$$

Πρὸς ἐφαρμογὴν τῶν λεχθέντων θεωροῦμεν τὸν πίνακα 58, εἰς τὸν ὁποῖον εἶναι ἀναγεγραμμένα αἱ συχνότητες ἐπιτυχίας ἢ ἀποτυχίας 300 μαθητῶν, οἱ ὅποιοι ἔλαβον μέρος εἰς τὰς ἐξετάσεις τῶν Ἀγγλικῶν καὶ Γαλλικῶν.

		Ἀγγλικά		
		Ἐπιτυχία	Ἀποτυχία	
Γαλλικά	Ἐπιτυχία	$A = 60$	$B = 90$	150
	Ἀποτυχία	$\Gamma = 120$	$\Delta = 30$	150
		180	120	$N = 300$

Πίναξ 58

Δι' ἐφαρμογῆς τοῦ τύπου 10.14 λαμβάνομεν

$$z = \frac{\Gamma - B}{\sqrt{B + \Gamma}} = \frac{120 - 90}{\sqrt{120 + 90}} = \frac{30}{14,4914} = 2,0702$$

Ἐπειδὴ $z = 2,0702 > 1,96$ εἰς τὸ ἐπίπεδον 5%, δι' ἀμφίπλευρον ἔλεγχον, συμπεραίνομεν τὴν ἀπόρριψιν τῆς H_0 .

Διὰ τὴν ἐφαρμογὴν τοῦ τύπου 10.13 εἶναι ἀπαραίτητος ὁ πίναξ τῶν ποσοστῶν. Μετατρέπομεν τὸν πίνακα 58 τῶν συχνοτήτων εἰς τὸν πίνακα 59 τῶν ποσοστῶν.

		Ἀγγλικά		
		Ἐπιτυχία	Ἀποτυχία	
Γαλλικά	Ἐπιτυχία	$\alpha = 0,20$	$\beta = 0,30$	$p_2 = 0,50$
	Ἀποτυχία	$\gamma = 0,40$	$\delta = 0,10$	$q_2 = 0,50$
		$p_1 = 0,60$	$q_1 = 0,40$	1,00

Πίναξ 59

$$\alpha = \frac{A}{N} = \frac{60}{300} = 0,20$$

$$\beta = \frac{B}{N} = \frac{90}{300} = 0,30$$

$$p_2 = \frac{A + B}{N} = \frac{150}{300} = 0,50$$

κ.λ.π.

Δι' αντικαταστάσεως εἰς τὸν τύπον 10.13

$$z = \frac{p_1 - p_2}{\sqrt{\frac{\beta + \gamma}{N}}} \quad \text{λαμβάνομεν}$$

$$z = \frac{0,60 - 0,50}{\sqrt{\frac{0,30 + 0,40}{300}}} = \frac{0,10}{0,0483} = 2,07021$$

Ἡ μικρὴ διαφορὰ ἣ ὁποία προκύπτει μεταξύ τῶν δύο τύπων ὀφείλεται εἰς τὴν κατὰ προσέγγισιν εὐρεσιν τῆς τετραγωνικῆς ρίζης, ἀλλ' ἀνεξαρτήτως τούτου καὶ εἰς αὐτὴν τὴν περίπτωσιν.

$$\text{τὸ } z = 2,07021 > 1,96 \quad \text{εἰς τὸ ἐπίπεδον } 5\%$$

Συνεπῶς ἡ διαφορὰ τῶν ποσοστῶν ἐπιτυχίας εἰς τὰ Ἀγγλικά καὶ Γαλλικά εἶναι στατιστικῶς σημαντικὴ.