

9. ΣΥΣΧΕΤΙΣΙΣ

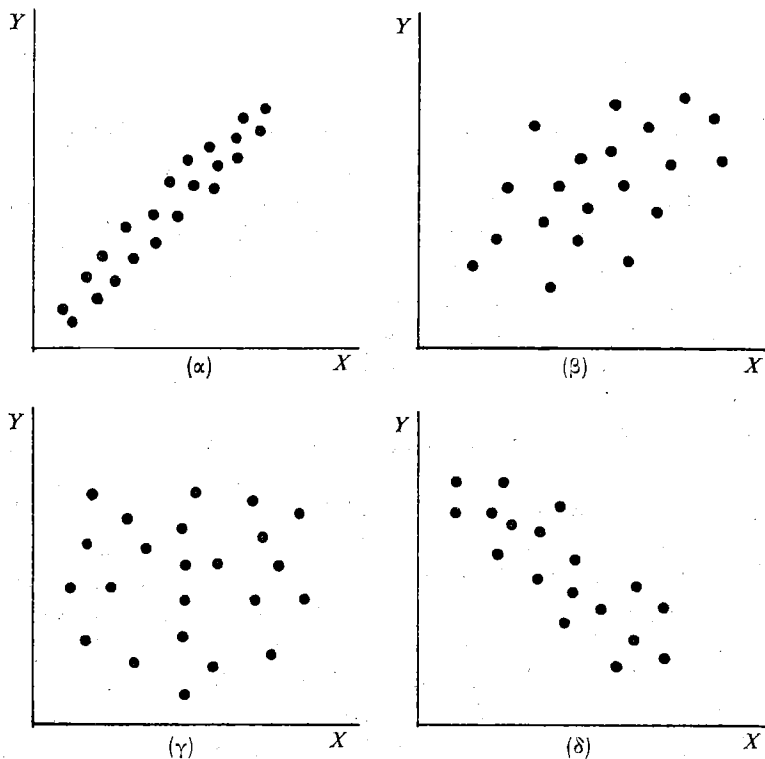
ΟΡΙΣΜΟΣ

Εἰς διάφορα πειράματα συμβαίνει συχνάκις νὰ θέλωμεν νὰ γνωρίζωμεν τὴν ἔκτασιν συσχετισμοῦ δύο μεταβλητῶν. Ὑπάρχουν διάφοροι μέθοδοι μετρήσεως τοιούτων σχέσεων, αἱ ὁποῖαι ἐξαρτῶνται ἀπὸ τὴν φύσιν τῶν μεταβλητῶν καὶ ἀπὸ τὸ εἶδος τῶν κατανομῶν. Μία μέθοδος, πολὺ διαδεδομένη, χρησιμοποιεῖται εἰς ἐκείνας τὰς περιπτώσεις εἰς τὰς ὁποίας αἱ δύο μεταβληταὶ εἶναι ἀπαριθμηταὶ καὶ κατανέμονται κανονικῶς. Πληρουμένων τῶν ἀνωτέρω συνθηκῶν, θεωρητικῶς ἡ ἀναφερθεῖσα μέθοδος εἶναι ἡ πλέον ἱκανοποιητικὴ καὶ ἡ πλέον χρήσιμος διὰ τὴν εὑρεσιν τῆς **συσχετίσεως** (correlation).

Θεωροῦμεν τὴν βαθμολογίαν ἑνὸς τμήματος εἰς τὰ Μαθηματικά καὶ εἰς τὰ Φυσικά. Ἐὰν ὁ πρῶτος μαθητὴς εἰς τὰ Μαθηματικά συμβῇ νὰ εἶναι καὶ πρῶτος εἰς τὰ Φυσικά, ὁ δεύτερος εἰς τὰ Μαθηματικά τύχη πάλιν νὰ εἶναι καὶ δεύτερος εἰς τὰ Φυσικά καὶ κατὰ τὸν αὐτὸν τρόπον ὁ τελευταῖος εἰς τὸ ἓνα μάθημα νὰ εἶναι καὶ τελευταῖος εἰς τὸ ἄλλο, τότε λέγομεν ὅτι ὑπάρχει **πλήρης θετικὴ συσχέτις** μεταξύ Μαθηματικῶν καὶ Φυσικῶν καὶ ἡ τιμὴ τοῦ **συντελεστοῦ συσχέτισεως** (coefficient of correlation), ὁ ὁποῖος συμβολίζεται μὲ r , ἰσοῦται μὲ τὴν μονάδα. Ἐὰν ὅμως συμβῇ ὁ πρῶτος εἰς τὰ Μαθηματικά νὰ εἶναι τελευταῖος εἰς τὰ Φυσικά καὶ μὲ τὸν αὐτὸν τρόπον καταλήξωμεν εἰς τὸν τελευταῖον μαθητὴν εἰς τὰ Μαθηματικά ὁ ὁποῖος θὰ εἶναι πρῶτος εἰς τὰ Φυσικά, τότε θὰ λέγωμεν ὅτι ὑπάρχει **πλήρης ἀρνητικὴ συσχέτις** μεταξύ τῶν δύο μαθημάτων καὶ ὁ συντελεστὴς συσχέτισεως r θὰ ἰσοῦται μὲ -1 . Ὅλοι αἱ ἄλλαι περιπτώσεις μᾶς δίδουν τιμὰς τοῦ συντελεστοῦ συσχέτισεως κυμαινομένας εἰς τὸ διάστημα $(-1, +1)$. Μεταξὺ τῆς πλήρους θετικῆς συσχέτισεως καὶ τῆς ἀρνητικῆς τοιαύτης ὑπάρχει ἀπεριόριστος ἀριθμὸς δυνατῶν τιμῶν.

Τέσσερις διαφορετικαὶ περιπτώσεις παρουσιάζονται εἰς τὸ σχ. 37. Ἡ περίπτωσις (α) παρουσιάζει τὴν γραφικὴν παράστασιν μιᾶς ὑψηλῆς θετικῆς συσχέτισεως. Εἰς τὸ σχῆμα (β) ὑπάρχει μὲν θετικὴ συσχέτις, ἀλλὰ ἡ τιμὴ τῆς εἶναι μικρά. Τὸ σχῆμα (γ) παρουσιάζει μίαν κατάστασιν κατὰ τὸ μᾶλλον ἢ ἥττον τυχαίαν. Δὲν ὑπάρχει οἰαδήποτε σχέσις κατὰ τὴν ὁποίαν δυνάμεθα νὰ προγνώσωμεν τὸ X ἐκ τοῦ Y ἢ ἀντιστρόφως. Ἡ τελευταία μορφή (δ) μᾶς παρουσιάζει μίαν σχετικῶς μεγάλην ἀρνητικὴν συσχέτισιν.

Θεωροῦμεν δύο κατανομάς, τῶν ὁποίων ὁ ἀριθμητικὸς μέσος εἶναι $\bar{X}$ καὶ $\bar{Y}$ καὶ ἡ τυπικὴ ἀπόκλισίς των εἶναι s_x καὶ s_y ἀντιστοίχως. Ὑποθέτομεν ἐπίσης ὅτι ἕκαστον στοιχεῖον ἐκ τῶν X συνδέεται μὲ ἓνα μόνον στοιχεῖον ἐκ τῶν Y . Τοῦτο ὑποδηλοῖ ὅτι ἑκάστη τῶν κατανομῶν ἔχει τὸ αὐτὸ πλῆθος τιμῶν N .



Σχ. 37

A

Ὁ πίναξ 44 παρουσιάζει τὴν βαθμολογίαν $N = 20$ μαθητῶν εἰς γραπτὰς ἐξετάσεις εἰς τὴν Γεωγραφίαν (μεταβλητὴ X) καὶ εἰς τὴν Ἱστορίαν (μεταβλητὴ Y). Οἱ βαθμοὶ εἶναι ἀναγεγραμμένοι εἰς τὰς στήλας 1 καὶ 4. Ἡ στήλη 2 προκύπτει, ἐὰν ἐξ ἑκάστης τιμῆς X ἀφαιρέσωμεν τὸν ἀριθμητικὸν μέσον $\bar{X} = 11$. Ὁμοίως προκύπτει καὶ ἡ στήλη 5. Αἱ στήλαι 3 καὶ 6 περιλαμβάνουν τὰ τετράγωνα τῶν ἀλγεβρικῶν διαφορῶν $X - \bar{X}$ καὶ $Y - \bar{Y}$. Τέλος ἡ ἑβδόμη στήλη λαμβάνεται ἐκ τοῦ πολλαπλασιασμοῦ τῶν ἀντιστοίχων τιμῶν τῶν στηλῶν 2 καὶ 5.

Βάσει τῶν ἀναγραφέντων στοιχείων τοῦ πίνακος 44 δυνάμεθα νὰ εὐρω-

μεν τὸν συντελεστὴν συσχέτισεως r ἐκ τοῦ τύπου

$$9.1 \quad r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}. \quad \Delta\iota' \text{ ἀντικαταστάσεως λαμβάνομεν}$$

$$r = \frac{294}{\sqrt{372 \cdot 452}} = 0,717$$

Ἡ εὑρεθεῖσα τιμὴ $r = 0,717$ εἶναι ἀρκούντως μεγάλη καὶ δυνάμεθα νὰ συμπεράνωμεν ὅτι ὑπάρχει συσχέτισις μεταξὺ τῶν μεταβλητῶν X καὶ Y .

Βαθμοὶ Γεωγραφίας	$\bar{X} = 11$		Βαθμοὶ Ἱστορίας	$\bar{Y} = 10$		
X	$X - \bar{X} = x$	$(X - \bar{X})^2 = x^2$	Y	$Y - \bar{Y} = y$	$(Y - \bar{Y})^2 = y^2$	xy
8	-3	9	2	-8	64	+ 24
19	+ 8	64	15	+ 5	25	+ 40
13	+ 2	4	12	+ 2	4	+ 4
15	+ 4	16	19	+ 9	81	+ 36
2	-9	81	7	-3	9	+ 27
10	-1	1	11	+ 1	1	- 1
13	+ 2	4	10	0	0	0
12	+ 1	1	17	+ 7	49	+ 7
18	+ 7	49	15	+ 5	25	+ 35
12	+ 1	1	13	+ 3	9	+ 3
17	+ 6	36	14	+ 4	16	+ 24
11	0	0	12	+ 2	4	0
8	-3	9	10	0	0	0
10	-1	1	5	-5	25	+ 5
6	-5	25	8	-2	4	+ 10
12	+ 1	1	10	0	0	0
10	-1	1	4	-6	36	+ 6
7	-4	16	2	-8	64	+ 32
13	+ 2	4	10	0	0	0
4	-7	49	4	-6	36	+ 42
$\Sigma X = 220$ $\bar{X} = 11$	$\Sigma x = 0$	$\Sigma x^2 = 372$	$\Sigma Y = 200$ $\bar{Y} = 10$	$\Sigma y = 0$	$\Sigma y^2 = 452$	$\Sigma xy = 294$

Πίναξ 44

Ἡ ἐφαρμογὴ τοῦ ἀνωτέρω τύπου συνιστᾶται εἰς τὰς περιπτώσεις ἐκείνας κατὰ τὰς ὁποίας,

- (α) δὲν ἔχομεν ὁμαδοποιημένα δεδομένα
- (β) οἱ ἀριθμητικοὶ μέσοι $\bar{X}$ καὶ $\bar{Y}$ εἶναι ἀκέραιοι
- (γ) τὸ $N \leq 100$.

B

Ἐνας ἄλλος τύπος, ὁ ὁποῖος μᾶς δίδει τὸν συντελεστὴν συσχέτισεως, ὅταν εἶναι γνωστὰ αἱ τυπικαὶ ἀποκλίσεις s_x καὶ s_y , εἶναι

$$9.2 \quad r = \frac{\Sigma xy}{(N-1)s_x s_y} *$$

Ἄν καὶ δὲν μᾶς δίδονται αἱ τυπικαὶ ἀποκλίσεις s_x καὶ s_y , θὰ γίνῃ ἐφαρμογὴ τοῦ τύπου 9.2 μὲ τὰ δεδομένα τοῦ πίνακος 44. Δι' ἀντικαταστάσεως λαμβάνομεν

$$r = \frac{\Sigma xy}{(N-1)s_x s_y}$$

$$r = \frac{\Sigma xy}{(N-1) \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N-1}} \sqrt{\frac{\Sigma y^2 - \frac{(\Sigma y)^2}{N}}{N-1}}}$$

$$r = \frac{294}{19 \sqrt{\frac{372 - \frac{0^2}{20}}{19}} \sqrt{\frac{452 - \frac{0^2}{20}}{19}}} = \frac{294}{\sqrt{168144}} = 0,717$$

Ἡ τιμὴ 0,717 εἶναι ἡ αὐτὴ μὲ τὴν εὑρεθεῖσαν ἐκ τοῦ τύπου 9.1. Θεωρεῖται σκόπιμον νὰ τονισθῇ διὰ μίαν εἰσέτι φορὰν ὅτι ἡ ἐφαρμογὴ τῶν τύπων 9.1 καὶ 9.2 προϋποθέτει τὴν ὑπαρξίν ἀκεραίων ἀριθμητικῶν μέσων.

*Εἰς τὴν πραγματικότητα ὁ τύπος 9.1 εἶναι ὁ αὐτὸς μὲ τὸν 9.2. Ὑποθέτοντες πάντοτε ὅτι $\bar{X}$ καὶ $\bar{Y}$ εἶναι ἀκέραιοι ἀριθμοί, πρᾶγμα τὸ ὁποῖον συμβαίνει ὅσον ἀφορᾷ τὰ δεδομένα τοῦ πίνακος 44, συνεπάγεται ὅτι $\Sigma x = \Sigma (X - \bar{X}) = 0$ καὶ $\Sigma y = \Sigma (Y - \bar{Y}) = 0$. Τοῦτο ἀπλοποιεῖ ταυτοχρόνως ἀμφοτέρους τοὺς τύπους.

$$\begin{aligned} r &= \frac{\Sigma xy}{(N-1)s_x s_y} = \frac{\Sigma xy}{(N-1) \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N-1}} \sqrt{\frac{\Sigma y^2 - \frac{(\Sigma y)^2}{N}}{N-1}}} \\ &= \frac{\Sigma xy}{(N-1) \sqrt{\frac{\Sigma x^2}{N-1}} \sqrt{\frac{\Sigma y^2}{N-1}}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \end{aligned}$$

Γ

Ἐὰν οἱ ἀριθμητικοὶ μέσοι εἰς οἰαδήποτε δεδομένα δὲν εἶναι ἀκέραιοι ἀριθμοὶ ἢ ἐὰν θέλωμεν νὰ χρησιμοποιήσωμεν ὑποθετικοὺς μέσους χάριν εὐκολίας τῶν ὑπολογιστικῶν πράξεων, γίνεται χρῆσις τοῦ τύπου

$$9.3 \quad r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{N}}{(N-1) s_x s_y}$$

Με ἀρχικὰ δεδομένα τὰς δύο βαθμολογίας τοῦ πίνακος 44 δημιουργοῦμεν τὸν πίνακα 45.

Βαθμοὶ Γεωγραφίας X	$\bar{X}_v = 12$ $X - \bar{X}_v = x$	x^2	Βαθμοὶ Ἱστορίας Y	$\bar{Y}_v = 9$ $Y - \bar{Y}_v = y$	y^2	xy
8	-4	16	2	-7	49	+ 28
19	+ 7	49	15	+ 6	36	+ 42
13	+ 1	1	12	+ 3	9	+ 3
15	+ 3	9	19	+ 10	100	+ 30
2	-10	100	7	-2	4	+ 20
10	-2	4	11	+ 2	4	- 4
13	+ 1	1	10	+ 1	1	+ 1
12	0	0	17	+ 8	64	0
18	+ 6	36	15	+ 6	36	+ 36
12	0	0	13	+ 4	16	0
17	+ 5	25	14	+ 5	25	+ 25
11	-1	1	12	+ 3	9	- 3
8	-4	16	10	+ 1	1	- 4
10	-2	4	5	-4	16	+ 8
6	-6	36	8	-1	1	+ 6
12	0	0	10	+ 1	1	0
10	-2	4	4	-5	25	+ 10
7	-5	25	2	-7	49	+ 35
13	+ 1	1	10	+ 1	1	+ 1
4	-8	64	4	-5	25	+ 40
$\Sigma X = 220$	$\Sigma x = -20$	$\Sigma x^2 = 392$	$\Sigma Y = 200$	$\Sigma y = + 20$	$\Sigma y^2 = 472$	$\Sigma xy = 274$

Πίναξ 45

Διὰ τὴν δημιουργίαν τῶν στηλῶν 2 καὶ 5 ἔχουν ληφθῇ ὡς ὑποθετικοὶ μέσοι τὸ $\bar{X}_y = 12$ καὶ $\bar{Y}_y = 9$. Ὅλαι αἱ στήλαι τοῦ πίνακος 45 εὐρέθησαν κατὰ τὸν ἴδιον τρόπον ὡς αἱ στήλαι τοῦ πίνακος 44, εἰς ἀντιστοιχίαν μίαν πρὸς μίαν.

Δι' ἀντικαταστάσεως τῶν διαφορῶν τιμῶν τοῦ πίνακος 45 εἰς τὸν τύπον

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{N}}{(N-1) s_x s_y}$$

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{N}}{(N-1) \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N-1}} \sqrt{\frac{\Sigma y^2 - \frac{(\Sigma y)^2}{N}}{N-1}}} \quad \text{λαμβάνομεν}$$

$$r = \frac{274 - \frac{20 \cdot 20}{20}}{19 \sqrt{\frac{392 - \frac{(-20)^2}{20}}{19}} \sqrt{\frac{472 - \frac{20^2}{20}}{19}}} = \frac{294}{\sqrt{168144}} = 0,717$$

Τὸ $r = 0,717$, τὸ ὁποῖον εὐρέθη, εἶναι ἡ τιμὴ ἡ ὁποία προέκυψεν ἐκ τῆς ἐφαρμογῆς τῶν τύπων 9.1 καὶ 9.2.

Δ

Ὅταν ὑπάρχῃ δυνατότης χρησιμοποίησεως ὑπολογιστικῆς μηχανῆς, ἐφαρμόζεται ὁ τύπος 9.4, ἀνεξαρτήτως τοῦ ἐὰν οἱ ἀριθμητικοὶ μέσοι εἶναι ἀκέραιοι ἀριθμοὶ ἢ ὄχι.

$$9_4. \quad r = \frac{N \Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N \Sigma X^2 - (\Sigma X)^2] [N \Sigma Y^2 - (\Sigma Y)^2]}} *$$

Διὰ τὴν ἐφαρμογὴν τοῦ τύπου 9.4 δημιουργοῦμεν τὸν πίνακα 46 ὁ ὁποῖος προκύπτει ἐκ τῶν δοθέντων στοιχείων τοῦ πίνακος 44. Ἡ δευτέρα καὶ ἡ τετάρτη στήλη περιλαμβάνουν τὰ τετράγωνα τῶν ἀντιστοιχῶν τιμῶν τῶν X καὶ Y καὶ ἡ πέμπτη στήλη τὰ γινόμενα τῶν ἀντιστοιχῶν τιμῶν τῶν στηλῶν 1 καὶ 3.

* Ὁ τύπος αὐτὸς προκύπτει ἐκ τοῦ 9.2.

Δι' ἀντικαταστάσεως εἰς τὸν τύπον 9.4

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2] [N\Sigma Y^2 - (\Sigma Y)^2]}}$$

λαμβάνομεν

$$r = \frac{20 \cdot 2494 - 220 \cdot 200}{\sqrt{[20 \cdot 2792 - 220^2] [20 \cdot 2452 - 200^2]}} = 0,717$$

Βαθμοὶ Γεωγραφίας X	X ²	Βαθμοὶ Ἱστορίας Y	Y ²	XY
8	64	2	4	16
19	361	15	225	285
13	169	12	144	156
15	225	19	361	285
2	4	7	49	14
10	100	11	121	110
13	169	10	100	130
12	144	17	289	204
18	324	15	225	270
12	144	13	169	156
17	289	14	196	238
11	121	12	144	132
8	64	10	100	80
10	100	5	25	50
6	36	8	64	48
12	144	10	100	120
10	100	4	16	40
7	49	2	4	14
13	169	10	100	130
4	16	4	16	16
ΣX = 220 X̄ = 11	ΣX ² = 2792	ΣY = 200 Ȳ = 10	ΣY ² = 2452	ΣXY = 2494

ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΕΩΣ ΔΡ' ΟΜΑΔΟΠΟΙΗΜΕΝΑ
ΔΕΔΟΜΕΝΑ

Ἡ εὕρεσις τοῦ συντελεστοῦ συσχετίσεως εἶναι ἐπίπονος, ἰδιαιτέρως ὅταν τὸ πλῆθος τῶν δεδομένων εἶναι μεγάλο. Ὅπως κατὰ τὴν εὕρεσιν τοῦ ἀριθμητικοῦ μέσου καὶ τῆς τυπικῆς ἀποκλίσεως ἐγένετο ὁμαδοποιήσις τῶν ἀρχικῶν στοιχείων διὰ μέγαλον N , οὕτω καὶ κατὰ τὴν εὕρεσιν τοῦ r θὰ ἀκο-

X	Y	X	Y	X	Y	X	Y
20	20	14	16	11	12	8	10
20	18	14	15	11	9	8	11
20	17	14	13	11	13	8	5
19	16	14	13	11	8	8	8
19	18	14	13	11	10	8	9
19	14	13	13	11	12	7	9
18	15	13	15	11	13	7	6
18	18	13	16	10	9	6	5
18	15	13	14	10	10	6	5
18	17	12	14	10	12	6	5
17	14	12	15	10	11	6	6
17	18	12	15	10	12	6	6
16	19	12	11	10	13	5	3
16	14	12	12	9	9	5	5
16	14	12	11	9	9	5	4
16	14	12	13	9	11	4	7
16	14	12	14	9	7	4	6
16	14	12	15	9	6	4	4
15	14	12	11	9	7	3	3
15	14	12	10	8	10	3	3

Πίναξ 47

λουθήσωμεν τὴν ἰδίαν προεργασίαν. Τοῦτο συνιστᾶται εἰς τὰς περιπτώσεις κατὰ τὰς ὁποίας τὸ πλῆθος τῶν δοθέντων στοιχείων εἶναι μεγαλύτερον τοῦ 100.

Ἀνεξαρτήτως ὅμως τοῦ ἀριθμοῦ 100, χάριν παραδείγματος, θεωροῦμεν ὅτι ὁ πίναξ 47 παρουσιάζει τοὺς βαθμοὺς 80 μαθητῶν εἰς τὰ Ἀρχαῖα (στῆλαι X) καὶ τοὺς βαθμοὺς τῶν ἰδίων μαθητῶν εἰς τὰ Λατινικά (στῆλαι Y).

Ἀφοῦ ἐκλέξωμεν τὸ διάστημα τάξεως, ἐν προκειμένῳ δύο μονάδας, σχη-

ματίζομεν τὸν πίνακα 48. Ὁ πίναξ αὐτὸς δημιουργεῖται ὡς ἐξῆς. Ὁ πρῶτος μαθητὴς ἔλαβεν 20 εἰς τὰ Ἀρχαῖα καὶ 20 εἰς τὰ Λατινικά. Εἰς τὴν τομὴν τῆς στήλης καὶ τῆς γραμμῆς τῶν διαστημάτων 19-20 τοποθετοῦμεν τὸν ὑπ' ἀριθμὸν 1 μαθητὴν. Ὁ δεύτερος μαθητὴς, ὁ ὁποῖος ἐβαθμολογήθη μὲ 20 εἰς τὰ Ἀρχαῖα καὶ 18 εἰς τὰ Λατινικά, τοποθετεῖται εἰς τὴν τομὴν τῆς στήλης 19-20 καὶ τῆς γραμμῆς 17-18. Ὁ ἐνδέκατος μαθητὴς, ὁ ὁποῖος ἔλαβεν 17 εἰς τὰ Ἀρχαῖα καὶ 14 εἰς τὰ Λατινικά, θὰ τοποθετηθῇ εἰς τὴν τομὴν τῆς στήλης 17-18 καὶ τῆς γραμμῆς 13-14. Μὲ τὸν ἴδιον τρόπον τοποθετοῦμεν καὶ τοὺς ὑπολοίπους μαθητάς.

Ἡ στήλη καὶ ἡ γραμμὴ f δημιουργοῦνται ἐκ τῆς ἀριθμήσεως τῶν περιπτώσεων τῶν εὐρισκομένων εἰς τὰς ἀντιστοίχους γραμμὰς καὶ στήλας.

Ἡτοι ἡ στήλη 3-4 μᾶς δίδει $f_{3-4} = 3 + 1 + 1 = 5$,

ἡ στήλη 5-6 $f_{5-6} = 2 + 6 = 8$.

Ὁμοίως καὶ αἱ ἄλλαι στήλαι. Διὰ προσθέσεως τῶν περιπτώσεων τῶν εὐρισκομένων εἰς τὴν γραμμὴν 3-4 ἔχομεν $f_{3-4} = 3 + 2 = 5$,

εἰς τὴν γραμμὴν 5-6 $f_{5-6} = 1 + 6 + 2 + 1 = 10$ κ.ο.κ.

Ἡ μεγαλύτερα συχνότης εἰς τὴν στήλην f εἶναι τὸ 20. Δεξιὰ τοῦ 20 τοποθετοῦμεν τὸ 0 καὶ ἄνωθεν τοῦ 0 διαδοχικῶς τὰ $-1, -2, -3, \dots$ καὶ κάτωθεν τοῦ μηδενὸς τὰ $+1, +2, +3, \dots$. Τοιοῦτοτρόπως ἐδημιουργήθη ἡ στήλη y . Ἡ γραμμὴ f ἔχει μεγαλύτεραν συχνότητα τὸ 18. Κάτωθεν αὐτοῦ τοποθετοῦμεν τὸ 0 καὶ δεξιὰ του τὰ $+1, +2, +3, \dots$. Ἀριστερὰ τοῦ μηδενὸς διαδοχικῶς γράφομεν τὰ $-1, -2, -3, \dots$. Οὕτω ἔχομεν τὴν γραμμὴν x . Ἡ ἐκλογή τῆς τοποθετήσεως τοῦ μηδενὸς εἶναι τυχαία. Προτιμᾶται ἐκείνη ἡ θέσις εἰς τὴν ὁποίαν αἱ πράξεις γίνονται εὐκολώτεροι. Ἡ θέσις αὕτη ἀντιστοιχεῖ εἰς τὴν μεγαλύτεραν συχνότητα.

Ἡ γραμμὴ fx δημιουργεῖται διὰ τοῦ πολλαπλασιασμοῦ τῶν ἀντιστοίχων τιμῶν τῶν γραμμῶν f καὶ x , ἡ δὲ γραμμὴ fx^2 διὰ τοῦ πολλαπλασιασμοῦ τῶν ἀντιστοίχων ἀριθμῶν τῶν γραμμῶν x καὶ fx . Κατὰ τὸν ἴδιον τρόπον γίνονται καὶ αἱ στήλαι fy καὶ fy^2 .

Οἱ ἀριθμοὶ οἱ ὁποῖοι κεῖνται ἐντὸς τῶν διαφόρων κύκλων, τῶν ὁποίων τὸ ἄθροισμα δίδει $\Sigma f_0xy = 326$, εὐρίσκονται διὰ τοῦ πολλαπλασιασμοῦ τριῶν ἀριθμῶν, δηλαδὴ τῆς συχνότητος εἰς ἕκαστον κελλίον καὶ τῶν δύο ἀριθμῶν τῆς γραμμῆς x καὶ στήλης y , οἱ ὁποῖοι ἀντιστοιχοῦν εἰς τὸ κελλίον. Ἡτοι

$$60 = 3(-4) (-5)$$

$$30 = 2(-3) (-5)$$

$$16 = 1(-4) (-4) \text{ κ.λ.π.}$$

Εἰς τὸν πίνακα 48 ὑπάρχουν ὅλα τὰ στοιχεῖα διὰ τὴν εὑρεσιν τοῦ συντελεστοῦ συσχετίσεως.

$\begin{array}{c} X \\ Y \end{array}$	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	f	y	fy	fy^2
3-4	$\begin{array}{c} (60) \\ \square \end{array}$	$\begin{array}{c} (30) \\ \square \end{array}$								5	-5	-25	125
5-6	$\begin{array}{c} (16) \\ \end{array}$	$\begin{array}{c} (72) \\ \square \end{array}$	$\begin{array}{c} (16) \\ \square \end{array}$	$\begin{array}{c} (4) \\ \end{array}$						10	-4	-40	160
7-8	$\begin{array}{c} (12) \\ \end{array}$		$\begin{array}{c} (6) \\ \end{array}$	$\begin{array}{c} (6) \\ \square \end{array}$	$\begin{array}{c} \\ \square \end{array}$					5	-3	-15	45
9-10			$\begin{array}{c} (16) \\ \square \end{array}$	$\begin{array}{c} (8) \\ \square \end{array}$	$\begin{array}{c} \square \end{array}$					11	-2	-22	44
11-12			$\begin{array}{c} (2) \\ \end{array}$	$\begin{array}{c} (4) \\ \square \end{array}$	$\begin{array}{c} \square \end{array}$					11	-1	-11	11
13-14				$\begin{array}{c} \\ \square \end{array}$	$\begin{array}{c} \square \end{array}$	$\begin{array}{c} \square \end{array}$				20	0	0	0
15-16					$\begin{array}{c} \square \end{array}$	$\begin{array}{c} (4) \\ \square \end{array}$		$\begin{array}{c} (6) \\ \square \end{array}$	$\begin{array}{c} (4) \\ \end{array}$	10	+1	+10	10
17-18								$\begin{array}{c} (18) \\ \square \end{array}$	$\begin{array}{c} (24) \\ \square \end{array}$	6	+2	+12	24
19-20							$\begin{array}{c} (6) \\ \end{array}$		$\begin{array}{c} (12) \\ \end{array}$	2	+3	+6	18
f	5	8	8	12	18	9	8	6	6	$\Sigma f = 80$		$\Sigma fy = -85$	$\Sigma fy^2 = 437$
x	-4	-3	-2	-1	0	+1	+2	+3	+4				
fx	-20	-24	-16	-12	0	+9	+16	+18	+24	$\Sigma fx = -5$			
fx^2	80	72	32	12	0	9	32	54	96	$\Sigma fx^2 = 387$			

Ἀντικαθιστῶμεν εἰς τὸν τύπον

$$9.5 \quad r = \frac{\frac{\sum f_0 xy}{N} - \left(\frac{\sum fx}{N}\right) \left(\frac{\sum fy}{N}\right)}{\sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2} \sqrt{\frac{\sum fy^2}{N} - \left(\frac{\sum fy}{N}\right)^2}}$$

τὰς εὐρεθείσας τιμὰς

$$\begin{aligned}\sum f_0 xy &= 326 \\ \sum fx &= -5 \\ \sum fx^2 &= 387 \\ \sum fy &= -85 \\ \sum fy^2 &= 437 \\ N &= 80\end{aligned}$$

καὶ εὐρίσκομεν

$$r = \frac{\frac{326}{80} - \left(\frac{-5}{80}\right) \left(\frac{-85}{80}\right)}{\sqrt{\frac{387}{80} - \left(\frac{-5}{80}\right)^2} \sqrt{\frac{437}{80} - \left(\frac{-85}{80}\right)^2}} = \frac{25655}{29291,33} = 0,876$$

Ἕνας ἄλλος τύπος, ὁ ὁποῖος ἐφαρμόζεται διὰ τὴν εὐρεσιν τοῦ συντελεστοῦ συσχέτισεως, ὅταν τὰ δεδομένα μας εἶναι ὁμαδοποιημένα, εἶναι ὁ 9.6

$$9.6 \quad r = \frac{A + B - C}{2 \sqrt{A B}}$$

ὅπου $A = \sum fx^2 - \frac{(\sum fx)^2}{N} = 387 - \frac{(-5)^2}{80} = 386,6875$

$$B = \sum fy^2 - \frac{(\sum fy)^2}{N} = 437 - \frac{(-85)^2}{80} = 346,6875$$

$$C = \sum fd^2 - \frac{(\sum fd)^2}{N} = 92 - \frac{0}{80} = 92$$

Τὰ A καὶ B εὐρίσκονται διὰ τῆς ἀντικαταστάσεως τῶν

$$\sum fx = -5, \quad \sum fx^2 = 387, \quad \sum fy = -85, \quad \sum fy^2 = 437$$

τὰ ὁποῖα ἐλήφθησαν ἐκ τοῦ πίνακος 48 καὶ τὸ C δι' ἀντικαταστάσεως τῶν $\sum fd = 0$ καὶ $\sum fd^2 = 92$, τὰ ὁποῖα ἐλήφθησαν ἐκ τοῦ πίνακος 49.

Οἱ ἀριθμοὶ οἱ εὐρισκόμενοι ἐντὸς τῶν κελλίων τοῦ πίνακος 49 παρουσιάζουν τὰς ἰδίας συχνότητας τοῦ πίνακος 48. Διαγωνίως προστιθέμενοι μᾶς δίδουν τὴν στήλην f . Δεξιὰ τῆς μεγαλυτέρας συχνότητος τοποθετοῦμεν

X \ Y	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20
3-4		2							
5-6	3		1						
7-8	1	6		1					
9-10	1	2	1						
11-12		4	4	3					
13-14			4	6	1				
15-16			1	5	5	7			
17-18				3	4	2	1		
19-20						3	3	1	

f	d	fd	fd^2
1	-3	-3	9
4	-2	-8	16
21	-1	-21	21
29	0	0	0
18	+1	+18	18
7	+2	+14	28
$\Sigma f = 80$		$\Sigma fd = 0$	$\Sigma fd^2 = 92$

τὸ μηδὲν καὶ διαδοχικῶς τοὺς ἀριθμοὺς $-1, -2, -3$ καὶ $+1, +2$ ἄνωθεν καὶ κάτωθεν αὐτοῦ ἀντιστοιχῶς. Ἡ στήλη fd εἶναι τὸ ἀποτέλεσμα τοῦ πολλαπλασιασμοῦ τῶν ἀντιστοιχῶν τιμῶν τῶν στηλῶν f καὶ d . Τέλος ἡ στήλη fd^2 προέκυψεν ἐκ τοῦ πολλαπλασιασμοῦ τῶν ἀντιστοιχῶν ἀριθμῶν τῶν στηλῶν d καὶ fd .

Δι' ἀντικαταστάσεως τῶν A, B καὶ C εἰς τὸν τύπον

$$r = \frac{A + B - C}{2\sqrt{AB}} \quad \text{λαμβάνομεν}$$

$$r = \frac{386,6875 + 346,6875 - 92}{2\sqrt{386,6875 \cdot 346,6875}} = \frac{641,375}{732,283} = 0,876$$

Ἡ τιμὴ $r = 0,876$ εἶναι ἡ αὐτὴ ὡς καὶ προηγουμένως.

ΕΛΕΓΧΟΣ ΣΗΜΑΝΤΙΚΟΤΗΤΟΣ ΤΟΥ r

Ἐνα πρόβλημα, τὸ ὁποῖον πολὺ συχνὰ ἀπαντᾶται εἰς ψυχολογικὰς ἐρεῦνας, εἶναι τὸ τοῦ ἐλέγχου τῆς συσχετίσεως μεταξὺ διαφόρων ζευγῶν παρατηρήσεων ἢ μετρήσεων.

Θεωροῦμεν τὴν μηδενικὴν ὑπόθεσιν ὡς ἰσχύουσιν, ὅτι δηλ. ἡ τιμὴ τοῦ συντελεστοῦ συσχετίσεως ἰσοῦται μὲ τὸ μηδέν. Ἡ οἰαδήποτε τιμὴ τοῦ r διάφορος τοῦ μηδενὸς θεωρεῖται ὅτι εἶναι ἀποτέλεσμα τύχης.

Δυνάμεθα κατὰ διαφόρους τρόπους νὰ ἐλέγξωμεν τὴν σημαντικότητα ἢ μὴ τοῦ συντελεστοῦ συσχετίσεως.

- (α) Ὁ πίναξ IV μᾶς δίδει κρισίμους τιμὰς τοῦ r εἰς διάφορα ἐπίπεδα, πέραν τῶν ὁποίων ὑπάρχει σημαντικότης τοῦ συντελεστοῦ συσχετίσεως. Ἐκ τοῦ ἰδίου πίνακος παρατηροῦμεν ὅτι διὰ μικρὸν πλῆθος δεδομένων, τὸ ὁποῖον συνεπάγεται μικρὸν ἀριθμὸν βαθμῶν ἐλευθεριῶν, μία μεγάλη τιμὴ τοῦ r εἶναι ἀναγκαία διὰ σημαντικότητα εἰς τὰ δοθέντα ἐπίπεδα. Τὰ δεδομένα τοῦ πίνακος 44 ἔχουν δώσει $r = 0,717$. Ἐπειδὴ $r = 0,717 > 0,444$ εἰς τὸ ἐπίπεδον 5% διὰ ἀμφίπλευρον ἔλεγχον καὶ διὰ $N - 2 = 20 - 2 = 18$ βαθμοὺς ἐλευθερίας, συμπεραίνομεν ὅτι ἡ εὐρεθεῖσα τιμὴ τοῦ συντελεστοῦ συσχετίσεως εἶναι στατιστικῶς σημαντική.

Ἐὰν ἕνα οἰονδήποτε ἄλλο παράδειγμα ἔδιδε $r = 0,717$ μὲ 5 βαθμοὺς ἐλευθερίας, τότε θὰ εἴχομεν $r = 0,717 < 0,754$ εἰς τὸ ἐπίπεδον 5%. Ἡ ἀνισότης αὕτη μᾶς λέγει ὅτι ἡ τιμὴ τοῦ $r = 0,717$ δὲν εἶναι στατιστικῶς σημαντικὴ διὰ 5 β.ε. Γενικῶς ὁ μικρὸς ἀριθμὸς τῶν δεδομένων χρειάζεται μεγάλην τιμὴν τοῦ r διὰ σημαντικότητα.

- (β) Ἀντὶ τῆς ἀρχικῆς τιμῆς τοῦ $r = 0,717$, δυνάμεθα νὰ ἐλέγξωμεν, τὴν σημαντικότητα τοῦ συντελεστοῦ συσχέτισεως καὶ ἐκ τῆς εὐρέσεως τοῦ t , τὸ ὁποῖον δίδεται ἐκ τοῦ τύπου

$$9.7 \quad t = r \sqrt{\frac{N-2}{1-r^2}}$$

$$t = 0,717 \sqrt{\frac{20-2}{1-0,717^2}} = 4,364$$

Ἐκ τοῦ πίνακος III λαμβάνομεν

$$t = 4,364 > 2,101 \text{ εἰς τὸ ἐπίπεδον } 5\%$$

διὰ $N-2 = 18$ βαθμοὺς ἐλευθερίας καὶ διὰ ἀμφίπλευρον ἐλεγχον.

Ἡ ἀνωτέρω ἀνισότης καταρρίπτει τὴν ἰσχὺν τῆς μηδενικῆς ὑποθέσεως.

Ἡ τιμὴ $r = 0,717$ εἶναι στατιστικῶς σημαντικὴ.

- (γ) Ἕνας ἄλλος τρόπος ἐλέγχου τῆς σημαντικότητος ἢ μὴ τοῦ συντελεστοῦ συσχέτισεως εἶναι ὁ τύπος

$$9.8 \quad T_{\sigma} \cdot r = \frac{1-r^2}{\sqrt{N-1}}$$

ὁ ὁποῖος μᾶς δίδει τὸ **τυπικὸν σφάλμα** (standard error) τοῦ r . Ὁ τύπος αὐτὸς ἐφαρμόζεται εἰς ἐκείνας τὰς περιπτώσεις κατὰ τὰς ὁποίας τὸ N εἶναι μεγάλος ἀριθμὸς ($N > 50$) καὶ ἡ τιμὴ τοῦ r εἶναι μικρὴ. Ὁ συντελεστὴς συσχέτισεως θὰ διαφέρει σημαντικῶς ἀπὸ τὸ μηδέν, ἐὰν τὸ $r > 2 T_{\sigma} \cdot r$.

Δι' ἀντικαταστάσεως τοῦ $r = 0,717$ καὶ $N = 20$ λαμβάνομεν

$$T_{\sigma} \cdot r = \frac{1-0,717^2}{\sqrt{20-1}} = 0,111. \text{ Εἶναι προφανὲς ὅτι ἰσχύει}$$

$0,717 > 2 \cdot 0,111$. Ἐκ τούτου προκύπτει ὅτι τὸ r εἶναι στατιστικῶς σημαντικόν.

- (δ) Μία διαφορετικὴ προσέγγισις ἐλέγχου τῆς σημαντικότητος τοῦ r εἶναι ἡ ἐφαρμογὴ τοῦ ἐξῆς κριτηρίου.

- (i) Εἰς τὸ ἐπίπεδον 5%

ἀπορρίπτομεν τὴν H_0 ὅταν $z < -1,96$ ἢ $z > 1,96$

δεχόμεθα τὴν H_0 ὅταν $-1,96 \leq z \leq 1,96$

$$\text{ὅπου} \quad z = r \sqrt{N-1}$$

(ii) Είς τὸ ἐπίπεδον 1%

ἀπορρίπτομεν τὴν H_0 ὅταν $z < -2,58$ ἢ $z > 2,58$

δεχόμεθα τὴν H_0 ὅταν $-2,58 \leq z \leq 2,58$

Ἡ ἐφαρμογὴ τοῦ κριτηρίου ἰσχύει καὶ εἰς ἐκεῖνας τὰς περιπτώσεις κατὰ τὰς ὁποίας τὸ N εἶναι σχετικῶς μικρόν.

Θέτοντες $r = 0,717$ καὶ $N = 20$ εἰς τὸν δοθέντα τύπον λαμβάνομεν

$$z = 0,717 \cdot \sqrt{20-1} = 3,125$$

Ἡ εὐρεθεῖσα τιμὴ 3,125 εἶναι μεγαλυτέρα τοῦ 2,58, συνεπῶς ἀπορρίπτομεν τὴν μηδενικὴν ὑπόθεσιν εἰς τὸ ἐπίπεδον 1%. Ἡ τιμὴ $r = 0,717$, εἶναι ὀριστικῶς πλέον στατιστικῶς σημαντικὴ.

ΕΛΕΓΧΟΣ ΣΗΜΑΝΤΙΚΟΤΗΤΟΣ ΤΗΣ ΔΙΑΦΟΡΑΣ $r_1 - r_2$ ΔΥΟ ΑΝΕΞΑΡΤΗΤΩΝ ΔΕΙΓΜΑΤΩΝ

ὑποθέτομεν ὅτι ἔχομε δύο συντελεστὰς συσχέτισεως $r_1 = 0,740$ καὶ $r_2 = 0,530$, οἱ ὅποιοι προέκυψαν ἐκ δύο γραπτῶν ἐξετάσεων εἰς τὰ Ἀγγλικά καὶ Γαλλικά ἐνὸς τμήματος τῆς Ε' τάξεως πληθικοῦ ἀριθμοῦ $N_1 = 43$ καὶ ἐνὸς τμήματος τῆς Γ' τάξεως πληθικοῦ ἀριθμοῦ $N_2 = 33$. Εἶναι στατιστικῶς σημαντικὴ ἡ διαφορὰ τῶν δύο συντελεστῶν συσχέτισεων;

Ἡ μηδενικὴ ὑπόθεσις H_0 θεωρεῖ ὅτι οἱ συντελεσταὶ εἶναι ἴσοι καὶ κατ' ἐπέκτασιν ὅτι τὰ δύο τυχαῖα δείγματα (ἓνα τμήμα Ε' τάξεως καὶ ἓνα τμήμα Γ' τάξεως) δύνανται νὰ θεωρηθοῦν ὡς προερχόμενα ἐκ τοῦ ἰδίου πληθυσμοῦ. Ἡ ὑπάρχουσα διαφορὰ τῶν r_1 καὶ r_2 ὀφείλεται μόνον εἰς τὴν τύχην.

Ἡ σημαντικότης ἢ μὴ τῆς διαφορᾶς ἐλέγχεται διὰ τῆς μετατροπῆς τῶν r_1 καὶ r_2 εἰς τὰς ἀντιστοίχους τιμὰς z_{r_1} καὶ z_{r_2} τῇ βοήθειᾳ τῶν μετασχηματισμῶν τοῦ Fisher. Ἡ μετατροπὴ τῶν r_1 καὶ r_2 εἰς z_{r_1} καὶ z_{r_2} δίδεται ἐκ τοῦ πίνακος V.

Ἡ τιμὴ $r_1 = 0,740$ μετατρέπεται εἰς $z_{r_1} = 0,950$ καὶ

ἡ τιμὴ $r_2 = 0,530$ μετατρέπεται εἰς $z_{r_2} = 0,590$

Τὰς μετασχηματισθείσας τιμὰς μετὰ τῶν πληθικῶν ἀριθμῶν $N_1 = 43$ καὶ $N_2 = 33$ ἀντικαθιστῶμεν εἰς τὸν τύπον

$$9.9 \quad z = \frac{|z_{r_1} - z_{r_2}|}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}}$$

$$z = \frac{|0,950 - 0,590|}{\sqrt{\frac{1}{43-3} + \frac{1}{33-3}}} = 1,49$$

Ἐπειδὴ $1,49 < 1,96$, συμπεραίνομεν ὅτι ἡ διαφορά τῶν δύο συντελεστῶν συσχέτισεων δὲν εἶναι σημαντικὴ οὔτε εἰς τὸ ἐπίπεδον 5% οὔτε εἰς τὸ ἐπίπεδον 1%. Τὸ H_0 δὲν καταρρίπτεται.

Διὰ σημαντικότητα πρέπει τὸ $z > 1,96$ εἰς τὸ ἐπίπεδον 5%
καὶ $z > 2,58$ εἰς τὸ ἐπίπεδον 1%.

ΕΛΕΓΧΟΣ ΣΗΜΑΝΤΙΚΟΤΗΤΟΣ ΤΗΣ ΔΙΑΦΟΡΑΣ ΔΥΟ ΣΥΝΤΕΛΕΣΤΩΝ ΣΥΣΧΕΤΙΣΕΩΝ ΣΥΣΧΕΤΙΣΜΕΝΩΝ ΔΕΙΓΜΑΤΩΝ

Θεωροῦμεν ἓνα δεῖγμα N ἀτόμων εἰς τὰ ὁποῖα ἐδόθησαν 3 tests. Μὲ τὰ ἀποτελέσματα τὰ ὁποῖα ἐλήφθησαν δυνάμεθα νὰ εὗρωμεν τρεῖς συντελεστές συσχέτισεως r_{12} , r_{23} , r_{13} .

Διὰ νὰ ἐλέγξωμεν τὴν σημαντικότητα ἢ μὴ τῆς διαφορᾶς δύο ἐκ τῶν τριῶν συντελεστῶν συσχέτισεως, τῶν r_{12} καὶ r_{13} , ὑπολογίζομεν τὸ t ἐκ τοῦ τύπου

$$9_{10} \quad t = \frac{(r_{12} - r_{13}) \sqrt{(N-3)(1+r_{23})}}{\sqrt{2(1-r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})}}$$

μὲ $N-3$ βαθμοὺς ἐλευθερίας.

Ἐὰν ἡ τιμὴ ἢ ὁποία θὰ εὑρεθῇ ἐκ τῆς ἐφαρμογῆς τοῦ ἀνωτέρω τύπου εἶναι μεγαλυτέρα ἀπὸ τὴν τιμὴν ἢ ὁποία δίδεται ἐκ τοῦ πίνακος III διὰ $N-3$ βαθμοὺς ἐλευθερίας, τότε ἡ διαφορά $r_{12} - r_{13}$ εἶναι στατιστικῶς σημαντικὴ.

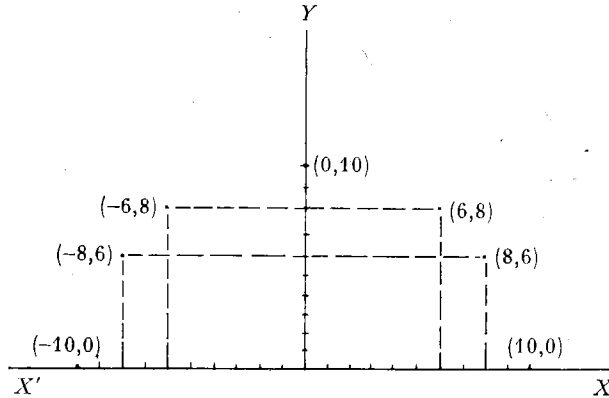
ΠΕΡΙΟΡΙΣΜΟΙ ΕΙΣ ΤΗΝ ΧΡΗΣΙΝ ΣΥΝΤΕΛΕΣΤΟΥ ΣΥΣΧΕΤΙΣΕΩΣ

Ὅταν ὁ συντελεστὴς συσχέτισεως πλησιάζῃ εἴτε τὸ $+1$ εἴτε τὸ -1 , ὅπως ἔχομεν ἀναφέρει, ὑπάρχει μεγάλη σχέσις μεταξὺ τῶν δύο μεταβλητῶν. Ὅταν ὅμως τὸ r πλησιάζῃ τὸ μηδέν, τότε ἀποφαινόμεθα ὅτι δὲν ὑπάρχει οἰαδήποτε συσχέτισις μεταξὺ τῶν μεταβλητῶν. Τοῦτο ὅμως δὲν συμβαίνει πάντοτε. Ἐξαιροῦνται αἱ περιπτώσεις κατὰ τὰς ὁποίας ἡ γραφικὴ παράστασις τῶν διατεταγμένων ζευγῶν (X , Y) δὲν μᾶς δίδει μίαν κατὰ προσέγγισιν εὐθεῖαν γραμμὴν. Πράγματι ὁ συντελεστὴς συσχέτισεως τοῦ πίνακος 50 (σημειακὸν διάγραμμα εἰς σχ. 38) ἰσοῦται μὲ τὸ μηδέν. Ἐν τούτοις ὑπάρχει πλήρης συσχέτισις μεταξὺ τῶν μεταβλητῶν X καὶ Y , διότι τὰ δοθέντα σημεῖα κεῖνται ἐπὶ περιφερείας, τῆς ὁποίας ἡ ἀκτίς ἰσοῦται μὲ 10. Δι' οἰανδήποτε τιμὴν τοῦ X δυνάμεθα νὰ εὗρωμεν εὐκόλως τὴν τιμὴν τοῦ Y καὶ ἀντιστρόφως. Ἐκ τῆς ἐξισώσεως $X^2 + Y^2 = 10^2$ βλέπομεν ὅτι ἡ συσχέτισις αὐτοῦ τοῦ εἴδους οὐδεμίαν σχέσιν ἔχει μὲ ὅσα ἔχουν ἀναφερθῇ. Παρατηροῦμεν ἐκ τῶν στοι-

χείων τοῦ πίνακος 50 ὅτι, ὅταν τὸ X αὐξάνεται ἀπὸ τὸ -10 ἕως τὸ 0 , τὸ Y ἐπίσης αὐξάνεται, ἀλλ' ὅταν τὸ X λαμβάνη τὰς τιμὰς μεταξὺ τοῦ 0 καὶ τοῦ $+10$, τὸ Y ἐλαττωῖται ἀπὸ τὸ 10 ἕως τὸ 0 . Ὡστε πρέπει νὰ εἴμεθα προσεκτικοὶ εἰς τὴν ἐφαρμογὴν τοῦ συντελεστοῦ συσχέτισεως, ἡ ὁποία ἔχει σχέσιν μόνον μὲ τὰ διαγράμματα ἐκεῖνα τὰ ὁποῖα εἶναι ἀποτέλεσμα δεδομένων τὰ ὁποῖα ἀπεικονίζονται εἰς τὸ σημειακὸν διάγραμμα μὲ μίαν κατὰ προσέγγισιν εὐθεῖαν γραμμὴν.

X	Y
-10	0
-8	6
-6	8
0	10
6	8
8	6
10	0

Πίναξ 50



Σχ. 38

ΔΙΑΤΑΚΤΙΚΗ ΣΥΣΧΕΤΙΣΙΣ*

Εἰς μερικὰς περιπτώσεις ἀντὶ τῆς χρησιμοποιήσεως τῶν μετρήσεων εἰς τὴν εὑρεσιν τοῦ συντελεστοῦ συσχέτισεως — ἡ μέτρησις εἶναι πολὺ δύσκολος εἰς ψυχολογικὰς μεταβλητάς — προτιμᾶται ἡ σειρὰ ἐπιδόσεως ἢ κατὰ-ξέως τῶν ἀτόμων. Ἄν καὶ κατ' αὐτὸν τὸν τρόπον ἔχομεν ἀπώλειαν πληροφοριῶν, ἐντούτοις προκύπτουν καὶ πλεονεκτήματα ὡς

- (α) ἡ ἐξουδετέρωσις τῆς δυσκολίας εἰς τὴν μέτρησιν ψυχολογικῶν καταστάσεων,
- (β) ὁ περιορισμὸς τῶν ἀπαραιτήτων ὑποθέσεων διὰ τὴν εὑρεσιν τοῦ συντελεστοῦ συσχέτισεως καὶ
- (γ) ἡ ἀπλοποίησις τῶν ἀριθμητικῶν πράξεων. Ὁ τρόπος αὐτὸς γενικῶς ἐφαρμόζεται, ὅταν ὁ πληθικὸς ἀριθμὸς τῶν στοιχείων εἶναι μικρὸς ἢ ὅταν ζητῆται ταχύτερος ὑπολογισμὸς τοῦ συντελεστοῦ συσχέτισεως.

A

Ὑποθέτομεν ὅτι 10 μαθηταὶ ἔχουν βαθμοὺς εἰς τὰ Γαλλικὰ καὶ Ἀγγλικά οἱ ὁποῖοι ἐμφαίνονται εἰς τὸν πίνακα 51 (στῆλαι 1 καὶ 2). Ἡ τρίτη καὶ ἡ

* Μολονότι ἡ διατακτικὴ συσχέτισις εἶναι μέρος τῆς ἀπαραμετρικῆς στατιστικῆς, ἐν τούτοις τοποθετεῖται ἐδῶ πρὸς συμπλήρωσιν τοῦ κεφαλαίου «συσχέτισις».

τετάρτη στήλη περιλαμβάνει τὴν σειρὰν κατατάξεως τῶν μαθητῶν εἰς τὰ Γαλλικὰ καὶ Ἀγγλικὰ ἀντιστοίχως. Ἡ πέμπτη στήλη περιλαμβάνει τὰς ἀλγεβρικὰς διαφορὰς τῶν ἀντιστοιχῶν τιμῶν τῶν στηλῶν 3 καὶ 4. Τέλος ἡ ἕκτη στήλη περιλαμβάνει τὰ τετράγωνα τῶν τιμῶν τῆς πέμπτης στήλης.

Δι' ἀντικαταστάσεως εἰς τὸν τύπον

$$9_{.11} \quad \rho = 1 - \frac{6 \Sigma d^2}{N(N^2 - 1)}$$

τῶν τιμῶν $\Sigma d^2 = 94$

$N = 10$

λαμβάνομεν

$$\rho = 1 - \frac{6 \cdot 94}{10(10^2 - 1)} = 0,430$$

Βαθμοὶ Γαλλικῶν	Βαθμοὶ Ἀγγλικῶν	Σειρὰ ἐπιδόσεως εἰς τὰ Γαλλικὰ X	Σειρὰ ἐπιδόσεως εἰς τὰ Ἀγγλικὰ Y	X - Y = d	d ²
12	15	8	4	+ 4	16
18	17	2	2	0	0
14	9	6	10	- 4	16
16	10	4	9	- 5	25
13	16	7	3	+ 4	16
17	14	3	5	- 2	4
15	12	5	7	- 2	4
19	20	1	1	0	0
11	13	9	6	+ 3	9
10	11	10	8	+ 2	4
				$\Sigma d = 0$	$\Sigma d^2 = 94$

Πίναξ 51

Ἡ ἔννοια τοῦ **διατακτικοῦ συντελεστοῦ συσχέτισεως** ρ (rank correlation coefficient) εἶναι ἡ αὐτὴ με τοῦ r . Διὰ πολλοὺς λόγους ὅμως δὲν πρέπει νὰ χρησιμοποιῶνται εὐρέως. Ἡ ἐφαρμογὴ τοῦ τύπου 9_{.11} εἶναι εὐκόλος, ὅταν ὁ ἀριθμὸς τῶν περιπτώσεων εἶναι μικρὸς καὶ ὡς ἐκ τούτου ἡ ἀξιοπιστία του εἶναι μικρά. Ὁ Spearman, ὁ ὁποῖος εἰσήγαγε τὸν ὁμώνυμον τύπον, τοῦ ὁποῖου φέρει καὶ τὸ ὄνομα, ὑπέθεσεν ὅτι ἡ κατανομὴ εἶναι ὀρθογών-

νιος, ἐνῶ γνωρίζομεν ὅτι σχεδὸν εἰς ὅλας τὰς περιπτώσεις μετρήσεων ἔχομεν κανονικὴν κατανομήν. Εἰς τὴν ὀρθογώνιον κατανομήν τὰ διάφορα διαστήματα μεταξὺ τῶν διαφορῶν τιμῶν εἶναι ἴσα, ἐνῶ εἰς τὴν πραγματικότητα τοῦτο δὲν συμβαίνει. Εἶναι γνωστὸν ὅτι εἰς τὰ δύο ἄκρα τῆς κατανομῆς τὰ διαστήματα μεταξὺ τῶν διαφορῶν τιμῶν εἶναι σχετικῶς μεγάλα, ἐνῶ περὶ τὸ μέσον πολὺ μικρά, καὶ τοῦτο λόγῳ τοῦ μεγάλου πλήθους τῶν περιπτώσεων. Ὅπου εἶναι δυνατὴ ἡ εὗρεσις τοῦ r πρέπει νὰ παραγνωρίζεται τὸ ρ .

Τὰ στοιχεῖα τὰ ὁποῖα ἐδόθησαν εἰς τὸν πίνακα 51 παρουσιάζουν βαθμοὺς, ἕκαστος τῶν ὁποίων ἐδόθη μόνον μίαν φοράν. Ὅμως τοῦτο δὲν συμβαίνει εἰς τὴν πρᾶξιν.

B

Θεωροῦμεν τὰ στοιχεῖα τοῦ πίνακος 44. Εἰς τὴν Γεωγραφίαν ἔχουν βαθμολογηθῇ 2 μαθηταὶ μὲ τὸν βαθμὸν 8,
3 μαθηταὶ μὲ τὸν βαθμὸν 13,
3 μαθηταὶ μὲ τὸν βαθμὸν 10 κ.ο.κ.

Τὸ αὐτὸ συμβαίνει καὶ μὲ τοὺς βαθμοὺς τῆς Ἱστορίας. Εἰς αὐτὰς τὰς περιπτώσεις, κατὰ τὴν τοποθέτησιν τῶν μαθητῶν ἀναλόγως τῆς ἐπιδόσεως τῶν, παρουσιάζονται δεσμοί. Τοῦτο παρουσιάζει μίαν ἐπὶ πλέον ἀνακρίβειαν εἰς τὸν τύπον ὁ ὁποῖος δίδει τὸ ρ . Διὰ τὸν ὑπολογισμὸν τοῦ ρ , καὶ πρὸς μεγαλυτέραν εὐκολίαν μας, τοποθετοῦμεν τοὺς βαθμοὺς τῆς Γεωγραφίας (πίν. 52) διατακτικῶς. Ἀρχίζομεν ἀπὸ τὸν μεγαλύτερον, ὁ ὁποῖος εἶναι τὸ 19, καὶ τελειώνομεν εἰς τὸν μικρότερον ἀριθμὸν, ἰὸ 2.

Ὁ βαθμὸς 13 εἰς τὴν Γεωγραφίαν ἐδόθη εἰς τρεῖς περιπτώσεις καὶ ἐκάστη κανονικῶς πρέπει νὰ καταλάβῃ τὴν 5ην θέσιν. Ὅμως ἕκαστος ἐξ αὐτῶν τῶν βαθμῶν θὰ διαταχθῇ εἰς τὴν 6ην θέσιν, ἐπειδὴ οἱ μαθηταὶ οἱ ὁποῖοι ἔλαβον τὸν βαθμὸν 13 θὰ κατελάμβανον τὴν 5ην, 6ην καὶ 7ην θέσιν $\left(\frac{5+6+7}{3} = 6\text{η θέσις} \right)$. Θὰ ἀκολουθήσῃ ἡ 8η θέσις διὰ τὸν μαθητὴν μὲ βαθμὸν 12. Τὸ 12 ἐδόθη πάλιν τρεῖς φορές. Ἐκ τούτου τὸ 12 θὰ ἡδύνατο νὰ ἀντιστοιχῇ εἰς τὴν 8ην, 9ην καὶ 10ην θέσιν. Λόγῳ ὅμως τοῦ δεσμοῦ τοποθετοῦμεν τὸ 12 εἰς τὴν 9ην θέσιν $\left(\frac{8+9+10}{3} = 9\text{η θέσις} \right)$. Ἡ αὐτὴ ἐργασία ἐπαναλαμβάνεται καὶ διὰ τοὺς ὑπολοίπους δεσμοὺς τῶν βαθμῶν τῆς Γεωγραφίας καὶ τῆς Ἱστορίας.

Ἡ τρίτη καὶ ἡ τετάρτη στήλη περιλαμβάνει τὴν σειρὰν ἐπιδόσεως τῶν μαθητῶν ἀντιστοιχῶς εἰς τὴν Γεωγραφίαν καὶ Ἱστορίαν. Ἡ πέμπτη καὶ ἡ ἕκτη στήλη περιέχουν τὰς ἀλγεβρικὰς διαφορὰς τῶν ἀντιστοιχῶν τιμῶν X καὶ Y καὶ τὰ τετράγωνά των.

Δι' ἀντικαταστάσεως εἰς τὸν τύπον 9.₁₁ τῶν στοιχείων $\Sigma d^2 = 287,50$,

$N = 20$ (πίν. 52) λαμβάνομεν

$$e = 1 - \frac{6 \Sigma d^2}{N(N^2 - 1)} = 1 - \frac{6 \cdot 287,50}{20(20^2 - 1)} = 0,784$$

Ἡ τιμὴ αὐτὴ εἶναι διαφορετικὴ ἀπὸ τὴν $r = 0,717$. Διὰ τὸν λόγον αὐτὸν πρέπει νὰ ἀποφεύγεται ἡ εὕρεσις τοῦ e , ὅπου εἶναι δυνατὸς ὁ ὑπολογισμὸς τοῦ r . Τὸ e ὑπολογίζεται μόνον ὅταν δὲν εἶναι δυνατόν νὰ γίνη ἄλλως, ὅπως ὅταν μᾶς δοθῇ ἡ σειρὰ ἐπιδόσεως ἐκάστου ἀτόμου εἰς τὰς ὑπὸ ἐξέτασιν δύο μεταβλητάς ἢ ὅταν δὲν δυνάμεθα νὰ βαθμολογήσωμεν ἀντικειμενικῶς, ὡς συμβαίνει εἰς ψυχολογικὰς μεταβλητάς, π.χ. εἰς τὴν ἀντίδρασιν μιᾶς ομάδος

Βαθμοὶ Γεωγραφίας	Βαθμοὶ Ἱστορίας	Σειρὰ ἐπιδό- σεως εἰς τὴν Γεωγραφίαν X	Σειρὰ ἐπιδό- σεως εἰς τὴν Ἱστορίαν Y	$X - Y = d$	d^2
19	15	1	3,5	- 2,5	6,25
18	15	2	3,5	- 1,5	2,25
17	14	3	5	- 2	4,00
15	19	4	1	+ 3	9,00
13	10	6	11,5	- 5,5	30,25
13	10	6	11,5	- 5,5	30,25
13	12	6	7,5	- 1,5	2,25
12	17	9	2	+ 7	49,00
12	13	9	6	+ 3	9,00
12	10	9	11,5	- 2,5	6,25
11	12	11	7,5	+ 3,5	12,25
10	11	13	9	+ 4	16,00
10	5	13	16	- 3	9,00
10	4	13	17,5	- 4,5	20,25
8	2	15½	19,5	- 4	16,00
8	10	15½	11,5	+ 4	16,00
7	2	17	19,5	- 2,5	6,25
6	8	18	14	+ 4	16,00
4	4	19	17,5	+ 1,5	2,25
2	7	20	15	+ 5	25,00
				$\Sigma d = 0$	$\Sigma d^2 = 287,50$

Πίναξ 52

ατόμων εις ἓνα φαινόμενον, εις τὴν τοποθέτησίν των ἀναλόγως τῆς ἐργατικότητός των, τῆς κοινωνικῆς θέσεώς των, κ.ο.κ.

ΕΛΕΓΧΟΣ ΣΗΜΑΝΤΙΚΟΤΗΤΟΣ ΤΟΥ ρ

Διὰ νὰ ἐλέγξωμεν τὴν σημαντικότητα τῆς εὐρεθείσης τιμῆς $\rho = 0,784$, ἀνατρέχουμεν εἰς τὸν πίνακα VI ὁ ὁποῖος μᾶς δίδει εἰς δύο ἐπίπεδα σημαντικότητος κρισίμους τιμὰς τοῦ ρ διὰ διαφόρους τιμὰς τοῦ N . Ἐπειδὴ $N = 20$, τότε $\rho = 0,784 > 0,534$ εἰς τὸ ἐπίπεδον 1%. Συνεπῶς, ἡ τιμὴ $\rho = 0,784$ προεξοφλεῖ στατιστικῶς σημαντικὴν θετικὴν συσχέτισιν μεταξὺ τῶν δύο μεταβλητῶν.

Διὰ $N > 10$ δυνάμεθα νὰ ἐλέγξωμεν τὴν σημαντικότητα τῆς τιμῆς τοῦ ρ χρησιμοποιοῦντες καὶ τὴν τιμὴν τοῦ t , ἡ ὁποία δίδεται ἐκ τοῦ τύπου

$$9.12 \quad t = \rho \sqrt{\frac{N-2}{1-\rho^2}}$$

Ἀντικαθιστῶμεν τὰς τιμὰς $N = 20$
 $\rho = 0,784$ καὶ λαμβάνομεν

$$t = 0,784 \sqrt{\frac{20-2}{1-0,784^2}} = 5,358$$

Ἐπειδὴ $t = 5,358 > 2,878$ δι' ἀμφίπλευρον ἔλεγχον εἰς τὸ ἐπίπεδον 1% καὶ διὰ $N-2 = 18$ βαθμοὺς ἐλευθερίας, συμπεραίνομεν ὅτι ἡ τιμὴ τοῦ ρ εἶναι στατιστικῶς σημαντικὴ.

ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΡΙΚΗΣ ΣΥΣΧΕΤΙΣΕΩΣ

Εἶναι γνωστὸν ὅτι ἡ εὐφυΐα καὶ ἡ ἱκανότης τοῦ χειρίζεσθαι αὐξάνεται μέχρι μιᾶς ὀρισμένης ἡλικίας. Δηλ. τὸ παιδί μέσης εὐφυΐας τῶν 12 ἐτῶν εἶναι εὐφύτερον ἀπὸ ἐκεῖνο τῆς ἡλικίας τῶν 6 ἐτῶν. Τὸ αὐτὸ συμβαίνει καὶ μετὰ τὴν ἱκανότητα νὰ χειρίζεται κάποιος κάτι, κατὰ τὴν νεαράν τοῦλάχιστον ἡλικίαν. Ἐὰν δοθοῦν δύο tests τῆς ἀνωτέρω μορφῆς, ἀσφαλῶς θὰ συσχετίζωνται, διότι ἀμφότερα εἶναι συνάρτησις τῆς ἡλικίας. Ὁμοίως ἐὰν ἔχωμεν τοὺς βαθμοὺς ἑνὸς τμήματος ἐκ $N = 40$ μαθητῶν εἰς τὰ Ἀρχαῖα Ἑλληνικὰ (μεταβλητὴ x) καὶ εἰς τὰ Λατινικὰ (μεταβλητὴ y). Ἡ συσχέτισις μεταξὺ αὐτῶν ἔγκειται εἰς τὸ γεγονός ὅτι εἰς ἀμφότερα τὰ μαθήματα ἐπιδρᾷ ἡ εὐφυΐα (μεταβλητὴ w) τοῦ ατόμου. Εἰς αὐτάς τὰς περιπτώσεις ἐπιθυμοῦμεν νὰ εὗρωμεν τὸν συντελεστὴν συσχέτισεως μεταξὺ τῶν μεταβλητῶν x καὶ y ἀπομεμονωμένων ἐκ τῆς ἐπιδράσεως τῆς μεταβλητῆς w . Ἡ συσχέτισις αὕτη εὐρίσκεται διὰ τῆς ἐφαρμογῆς τοῦ τύπου

$$9.13 \quad r_{xy \cdot w} = \frac{r_{xy} - r_{xw} r_{yw}}{\sqrt{(1 - r_{xw}^2)(1 - r_{yw}^2)}}$$

όπου $r_{xy \cdot w}$ είναι ο **συντελεστής μερικής συσχέτισης** (partial correlation coefficient) και r_{xy} , r_{xw} , r_{yw} είναι οι συντελεστές συσχέτισης των μεταβλητών x , y , w λαμβανομένων ανά δύο.

Θεωρούμεν ότι έχουμε υπολογίσει τους τρεις συντελεστές συσχέτισης κατά τους γνωστούς τρόπους και ότι

$$r_{xy} = 0,72$$

$$r_{xw} = 0,50$$

$$r_{yw} = 0,60$$

Η τιμή του $r_{xy \cdot w}$ απομονώνει την μεταβλητήν w εκ τής επιδράσεώς της επί των x και y . Αντικαθιστώμεν εις τόν τύπον 9.13 τὰς δοθείσας 3 τιμάς των συντελεστών συσχέτισεων και έχουμε

$$r_{xy \cdot w} = \frac{r_{xy} - r_{xw} \cdot r_{yw}}{\sqrt{(1 - r_{xw}^2)(1 - r_{yw}^2)}}$$

$$r_{xy \cdot w} = \frac{0,72 - 0,50 \cdot 0,60}{\sqrt{(1 - 0,50^2)(1 - 0,60^2)}} = 0,606$$

Η ανωτέρω μέθοδος δύναται νά επεκταθῇ και εις 4 ἢ περισσότερας μεταβλητάς. Πρὸς περισσότεραν εὐκολίαν ἀντὶ των μεταβλητών w , x , y , z θέτομεν ἀντιστοίχως τοὺς ἀριθμοὺς 1, 2, 3, 4. Διὰ 4 μεταβλητάς ὁ τύπος 9.13 γράφεται

$$9.14 \quad r_{12 \cdot 34} = \frac{r_{12 \cdot 4} - r_{13 \cdot 4} \cdot r_{23 \cdot 4}}{\sqrt{(1 - r_{13 \cdot 4}^2)(1 - r_{23 \cdot 4}^2)}} \quad *$$

Τὸ $r_{12 \cdot 34}$ εἶναι ὁ συντελεστής συσχέτισης των μεταβλητών 1 καὶ 2 ἀπομεμονωμένων ἐκ των μεταβλητών 3 καὶ 4. Ἐκ τοῦ τύπου 9.14 φαίνεται ὅτι εἶναι ἀπαραίτητοι οἱ ὑπολογισμοὶ των συντελεστών μερικής συσχέτισεως $r_{12 \cdot 4}$, $r_{13 \cdot 4}$, $r_{23 \cdot 4}$, ἕκαστος των ὁποίων θεωρεῖ τὴν αὐτὴν μεταβλητὴν ὡς μὴ ἐπηρεάζουσιν τὰς ἄλλας δύο.

Ὁ ἔλεγχος τῆς τιμῆς $r_{xy \cdot w} = 0,606$, ἐὰν εἶναι στατιστικῶς σημαντικὴ ἢ ὄχι, καθορίζεται ἐκ τῆς εὐρέσεως τοῦ t , ὁ ὁποῖος δίδεται ἐκ τοῦ τύπου 9.15 μὲ $N - p - 2$ βαθμοὺς ἐλευθερίας.

* Ὁ τύπος 9.14 εἶναι ἰσοδύναμος μὲ τὸν

$$r_{12 \cdot 34} = \frac{r_{12 \cdot 3} - r_{14 \cdot 3} r_{24 \cdot 3}}{\sqrt{(1 - r_{14 \cdot 3}^2)(1 - r_{24 \cdot 3}^2)}}$$

$$9_{.15} \quad t = \frac{r_{xy \cdot w}}{\sqrt{\frac{1 - (r_{xy \cdot w})^2}{N - p - 2}}}$$

όπου $r_{xy \cdot w} = 0,606$ ό συντελεστής μερικῆς συσχετίσεως,
 $N = 40$ τὸ πλῆθος τῶν ἀτόμων,
 $p = 1$ τὸ πλῆθος τῶν μεταβλητῶν αἱ ὁποῖαι ἐθεωρήθησαν σταθεραὶ.

Δι' ἀντικαταστάσεως εἰς τὸν $9_{.15}$ λαμβάνομεν

$$t = \frac{0,606}{\sqrt{\frac{1 - (0,606)^2}{40 - 1 - 2}}} = 4,634$$

Ἐπειδὴ $t = 4,634 > 2,027$ διὰ $N - p - 2 = 37$ βαθμοὺς ἐλευθερίας εἰς τὸ ἐπίπεδον 5% καὶ διὰ ἀμφίπλευρον ἔλεγχον, ἐξάγομεν τὸ συμπέρασμα ὅτι ἡ τιμὴ $r_{xy \cdot w} = 0,606$ εἶναι στατιστικῶς σημαντικὴ.

ΕΥΘΕΙΑ ΠΡΟΣΑΡΜΟΓΗΣ

Οἷονδήποτε ζεύγος παρατηρήσεων ἢ μετρήσεων δύναται νὰ ἀπεικονισθῇ ἐπὶ τετραγωνισμένου φύλλου χάρτου δι' ἐνὸς σημείου. Τὰ σημεῖα τὰ ὁποῖα ἐμφαίνονται εἰς τὸ σχῆμα 39 ἀντιπροσωπεύουν τὰ ζεύγη τιμῶν τὰ ὁποῖα εἶναι ἀναγεγραμμένα εἰς τὸν πίνακα 53*. Π.χ. τὸ σημεῖον Α ἀντιπροσωπεύει τὸ ζεύγος τῶν τιμῶν $(X, Y) = (123, 72)$, τὸ σημεῖον Β ἀντιστοιχεῖ εἰς τὸ ζεύγος $(X, Y) = (118, 73)$ καὶ κατὰ τὸν ἴδιον τρόπον τὰ ὑπόλοιπα σημεῖα. Γενικὴ ἀρχὴ τῆς παραστάσεως τῶν σημείων εἶναι ἡ τοποθέτησις των εἰς τὰ σημεῖα τομῆς τῶν ἀντιστοιχῶν τετμημένων (X) καὶ τῶν τεταγμένων (Y). Τὸ διάγραμμα τοῦτο καλεῖται **διάγραμμα διασπορᾶς** (scatter diagram).

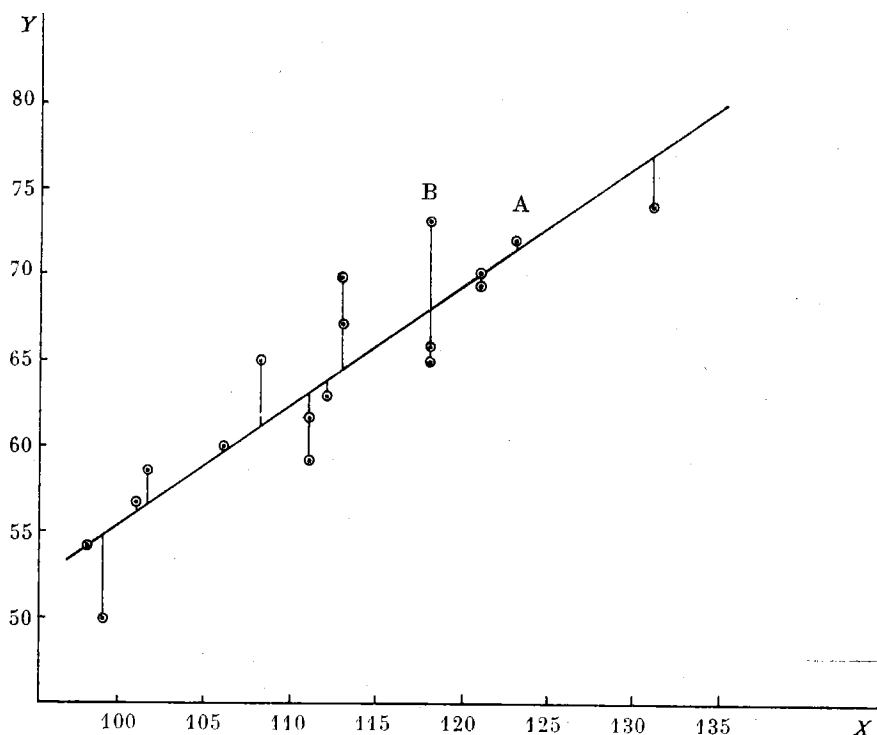
Ἐκ τοῦ σχήματος 39 φαίνεται καθαρὰ ὅτι τὰ διάφορα σημεῖα δὲν κεῖνται ἐπὶ εὐθείας γραμμῆς καὶ εἶναι ἀδύνατον νὰ ὑπάρξῃ εἰς τὴν πρᾶξιν τοιαύτη σύμπτωσης. Δυνάμεθα ὅμως τὰ διάφορα σημεῖα νὰ τὰ ἀντικαταστήσωμεν διὰ μιᾶς εὐθείας γραμμῆς, ἡ ὁποία κατὰ προσέγγισιν νὰ ἀντιπροσωπεύῃ τὸν μέσον ὅρον τῶν τιμῶν X καὶ Y . Κανεῖς ἀσφαλῶς δὲν πρέπει νὰ παίρῃ τὸν κανόνα καὶ νὰ φέρῃ μίαν εὐθεῖαν γραμμὴν «μὲ τὸ μάτι» καὶ νὰ λέγῃ ὅτι εἶναι ἡ ζητούμενη εὐθεῖα.

Ζητεῖται λοιπὸν νὰ εὑρεθῇ ἡ ἐξίσωσις τῆς εὐθείας γραμμῆς ἡ ὁποία

* Ἡ πρώτη στήλη τοῦ πίνακος 53 περιλαμβάνει τὸν δείκτην εὐφυΐας 18 μαθητῶν καὶ ἡ τρίτη στήλη τὴν βαθμολογίαν τῶν ἰδίων μαθητῶν εἰς τὰ Φυσικὰ.

προσαρμόζει ακριβέστερον εἰς τὰ δοθέντα στοιχεῖα. "Όταν εἶναι γνωστή ἡ εὐθεΐα, ἡ ὁποία καλεῖται **εὐθεΐα προσαρμογῆς** (linear regression), δυνάμεθα νὰ προγνώσωμεν κατὰ τὸ δυνατόν τὴν ἐπίδοσιν κάποιου μαθητοῦ εἰς τὰ Φυσικὰ, ὅταν μᾶς εἶναι γνωστὸς ὁ δείκτης εὐφυΐας του καὶ ἀντιστρόφως.

Ἡ μέθοδος προσαρμογῆς τῆς εὐθείας γραμμῆς εἰς ἓνα σύνολον σημείων καλεῖται **μέθοδος τῶν ἐλαχίστων τετραγώνων** (method of least squares). Ἡ μέθοδος τῶν ἐλαχίστων τετραγώνων, ὅπως χρησιμοποιεῖται ἐδῶ, μᾶς καλεῖ νὰ εὕρωμεν τὸ μικρότερον ἄθροισμα τῶν τετραγώνων τῶν καθέτων



Σχ. 39

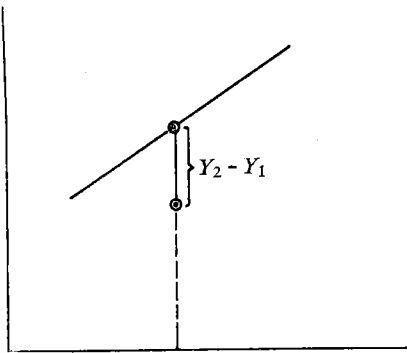
πρὸς τὸν ἄξονα τῶν X ἀποστάσεων ὅλων τῶν δοθέντων σημείων ἀπὸ μίαν εὐθεΐαν γραμμὴν, ἡ ὁποία εἶναι ἡ εὐθεΐα προσαρμογῆς. "Αλλως ζητεῖται νὰ εὕρωμεν εὐθεΐαν, ἐκ τῆς ὁποίας τὸ ἄθροισμα τῶν τετραγώνων τῶν ἀποστάσεων $\Sigma (Y_2 - Y_1)^2$ νὰ εἶναι ἐλάχιστον. Τὸ Y_2 εἶναι ἡ τεταγμένη τοῦ σημείου τομῆς τῆς καθέτου εἰς τὸν ἄξονα τῶν X , ἡ ὁποία διέρχεται διὰ τινος δοθέντος σημείου καὶ τῆς ζητουμένης εὐθείας γραμμῆς, τὸ δὲ Y_1 εἶναι τὸ διάστημα τῆς καθέτου, τὸ ὁποῖον περιέχεται μεταξύ τοῦ προαναφερθέντος δοθέντος σημείου καὶ τοῦ ἄξονος X . (σχ. 40). "Όταν χρησιμοποιηθῇ ἡ μέθοδος

τῶν ἐλαχίστων τετραγώνων διὰ τὴν εὕρεσιν τῆς εὐθείας προσαρμογῆς, ἐλαττώ-
νονται εἰς τὸ ἐλάχιστον τὰ λάθη προγνώσεως.

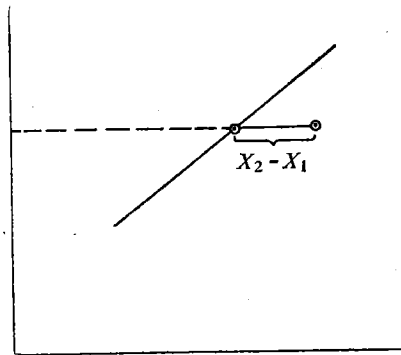
Ἡ εὐθεῖα γραμμὴ ἡ ὁποία ἱκανοποιεῖ τὸ κριτήριον τῶν ἐλαχίστων τε-
τραγώνων δίδεται ὑπὸ τῆς ἐξισώσεως

$$9.16 \quad Y = \beta X + \alpha$$

εἰς τὴν ὁποίαν τὸ α εἶναι μία σταθερὰ καὶ ἐκφράζει τὴν ἀπόστασιν τῆς ἀρχῆς
τῶν ἄξόνων ἀπὸ τῆς τομῆς τῆς εὐθείας μετὰ τὸν ἄξονα τῶν τεταγμένων. Τὸ α
εἶναι ἡ τιμὴ τοῦ Y ὅταν $X = 0$. Τὸ β εἶναι ἡ κλίσις τῆς εὐθείας γραμμῆς.



Σχ. 40



Σχ. 41

Αἱ τιμαὶ τῶν α καὶ β δίδονται ὑπὸ τῶν τύπων

$$9.17 \quad \beta = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}$$

$$9.18 \quad \alpha = \frac{\Sigma Y - \beta \Sigma X}{N}$$

Τὸ $\Sigma XY = 130806$ εἶναι τὸ ἄθροισμα τῶν τιμῶν τῆς στήλης 5 τοῦ πί-
νακος 53. Αἱ τιμαὶ τῆς προαναφερθείσης στήλης προέκυψαν ἐκ τοῦ πολλα-
πλασιασμοῦ τῶν ἀντιστοιχούντων δεδομένων τῶν στηλῶν 1 καὶ 3. Τὸ ἄθροισμα
τῶν στοιχείων τῶν στηλῶν 1 καὶ 3 μᾶς δίδει τὸ $\Sigma X = 2024$ καὶ $\Sigma Y = 1155$.
Τὸ $\Sigma X^2 = 228978$ καὶ $\Sigma Y^2 = 74885$ ἐκφράζουν τὰ ἄθροίσματα τῶν τετρα-
γώνων τῶν τιμῶν τῶν στηλῶν 1 καὶ 3. Τέλος τὸ N ἐκφράζει τὸν ἀριθμὸν
τῶν μετρήσεων ἢ παρατηρήσεων.

Δι' αντικαταστάσεως αὐτῶν τῶν δεδομένων εἰς τοὺς τύπους 9.17 καὶ 9.18 λαμβάνομεν

$$\beta = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\Sigma X^2 - \frac{(\Sigma X)^2}{N}} = \frac{130806 - \frac{2024 \cdot 1155}{18}}{228978 - \frac{(2024)^2}{18}} = 0,671$$

X	X^2	Y	Y^2	XY	Y_e	X_e
123	15129	72	5184	8856	71	122
118	13924	73	5329	8614	68	123
121	14641	69	4761	8349	70	118
118	13924	66	4356	7788	68	115
99	9801	50	2500	4950	55	95
111	12321	62	3844	6882	63	110
121	14641	70	4900	8470	70	120
108	11664	65	4225	7020	61	113
98	9604	54	2916	5292	54	100
131	17161	74	5476	9694	77	124
106	11236	60	3600	6360	60	107
113	12769	67	4489	7571	65	116
111	12321	59	3481	6549	63	106
118	13924	65	4225	7670	68	113
112	12544	63	3969	7056	64	111
101	10201	57	3249	5757	57	104
102	10404	59	3481	6018	57	106
113	12769	70	4900	7910	65	120
$\Sigma X = 2024$	$\Sigma X^2 = 228978$	$\Sigma Y = 1155$	$\Sigma Y^2 = 74885$	$\Sigma XY = 130806$		

Πίναξ 53

$$\alpha = \frac{\Sigma Y - \beta \Sigma X}{N} = \frac{1155 - 0,671 \cdot 2024}{18} = -11,26$$

Ἀντικαθιστῶμεν τὰς δύο εὐρεθείσας τιμὰς εἰς τὸν τύπον 9.16

$$Y = \beta X + \alpha \quad \text{καὶ εὐρίσκομεν τὴν ἐξίσωσιν}$$

$$9.19 \quad Y = 0,671X - 11,26$$

ἡ ὁποία εἶναι ἡ ἐξίσωσις τῆς εὐθείας προσαρμογῆς διὰ τὰ δοθέντα στοιχεῖα τοῦ πίνακος 53.

Ἐκ τῆς εὐθείας προσαρμογῆς δυνάμεθα δι' ἐκάστην τιμὴν τοῦ X (δείκτης εὐφυΐας) νὰ εὐρίσκωμεν τὴν ἀντίστοιχον τιμὴν τοῦ Y (βαθμὸς Φυσικῶν).

Ἡ στήλη 6 (πίναξ 53) περιέχει τὰς ἀντιστοίχους τιμὰς Y_e εἰς τὴν πλησιεστέραν ἀκεραίαν μονάδα ἐκ τῶν δοθεισῶν τιμῶν τοῦ X .

$$\begin{aligned}\text{π.χ. διὰ } X = 123 \text{ λαμβάνομεν } Y &= 0,671X - 11,26 \\ Y &= 0,671 \cdot 123 - 11,26 \\ Y &= 71,273 \simeq 71\end{aligned}$$

$$\begin{aligned}\text{διὰ } X = 118 \text{ λαμβάνομεν } Y &= 0,671 \cdot 118 - 11,26 \\ Y &= 67,918 \simeq 68\end{aligned}$$

Ὅμοίως καὶ αἱ ἄλλαι τιμαὶ τῆς στήλης Y_e .

Ἐὰν μᾶς δοθοῦν βαθμοὶ Φυσικῶν καὶ μᾶς ζητηθῇ νὰ εὐρώμεν τοὺς ἀντιστοίχους δείκτας εὐφυΐας, θὰ ἀκολουθηθῇ ἡ ἰδίᾳ πορεία. Πρέπει ὅμως νὰ τονισθῇ ὅτι ἡ εὐθεῖα προσαρμογῆς θὰ ἔχῃ διαφορετικὴν ἐξίσωσιν. Ἡ θέσις αὐτῆς τῆς εὐθείας εἶναι τοιαύτη, ὥστε νὰ μᾶς δίδεται τὸ μικρότερον ἄθροισμα τῶν τετραγώνων τῶν ἀποστάσεων τῶν σημείων τῶν κειμένων ἐπὶ τῶν καθέτων πρὸς τὸν ἄξονα τῶν Y καὶ τῆς εὐθείας (σχ. 41).

Αἱ δύο εὐθεῖαι προσαρμογῆς συμπίπτουν μόνον εἰς ἐκείνας τὰς περιπτώσεις, κατὰ τὰς ὁποίας ὅλα τὰ σημεῖα κεῖνται ἐπὶ μιᾷ εὐθείας. Ὅταν δὲν ὑπάρχῃ οἰαδήποτε σχέσις μεταξύ τῶν μεταβλητῶν X καὶ Y , δηλ. εἶναι ἀνεξάρτητοι, αἱ εὐθεῖαι προσαρμογῆς τέμνονται καθέτως.

Ὁ τύπος ὁ ὁποῖος δίδει τὴν εὐθεῖαν προσαρμογῆς, ὅταν εἶναι γνωστὸν τὸ Y καὶ ζητεῖται τὸ X , εἶναι

$$9.20 \quad X = \beta Y + \alpha \quad \text{ὅπου}$$

$$9.21 \quad \beta = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}$$

$$9.22 \quad \alpha = \frac{\Sigma X - \beta \Sigma Y}{N}$$

Δι' ἀντικαταστάσεως εἰς τοὺς ἀνωτέρω τύπους ἔχομεν

$$\beta = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}} = \frac{130806 - \frac{2024 \cdot 1155}{18}}{74885 - \frac{(1155)^2}{18}} = 1,207$$

$$\alpha = \frac{\Sigma X - \beta \Sigma Y}{N} = \frac{2024 - 1,207 \cdot 1155}{18} = 35$$

Με τήν εύρεσιν τῶν τιμῶν $\alpha = 35$ καὶ $\beta = 1,207$ ὁ τύπος 9.20 γίνεται

$$X = 1,207Y + 35$$

Ἐκ τῆς εύρεθείσης ἐξισώσεως δυνάμεθα νὰ εύρωμεν τιμὰς τοῦ X διὰ δοθείσας τιμὰς τοῦ Y .

$$\text{Διὰ } Y = 72 \text{ ἔπεται } Y = 1,207 \cdot 72 + 35$$

$$Y = 121,904 \simeq 122$$

Κατὰ τὸν ἴδιον τρόπον εύρίσκομεν καὶ ὅλας τὰς ἄλλας τιμὰς τῆς στήλης 7 τοῦ πίνακος 53.