ΑΝΑΓΚΑΙΑΙ ΚΑΙ ΙΚΑΝΑΙ ΣΥΝΘΗΚΑΙ ΔΙΑ ΤΗΝ ΙΣΧΥΝ ΒΑΣΙΚΩΝ ΤΙΝΩΝ ΙΔΙΟΤΗΤΩΝ ΤΟΥ ΔΕΙΚΤΟΥ ΠΡΟΣΔΙΟΡΙΣΜΟΥ

Τοῦ κ. ΔΗΜΗΤΡΙΟΥ ΑΘΑΝΑΣΟΠΟΥΛΟΥ
Ἐκτ. Καθηγητοῦ τῆς Α.Β.Σ.Π.

Ὁ δείκτης προσδιορισμοῦ R^2 ἀποτελεῖ ὡς γνωστὸν ἓν στατιστικὸν μέτρον τὸ ὁποῖον χρησιμοποιεῖται κατὰ τὴν μελέτην τοῦ τρόπου ἀλληλεξαρτήσεως δύο (ἢ περισσοτέρων) μεταβλητῶν (X,Y) καὶ — μεταξὺ ἄλλων — χαρακτηρίζει τὸν βαθμὸν προσαρμογῆς πρὸς τὰ ὑφιστάμενα δεδομένα — δηλαδὴ τὰ ζεύγη τῶν ἀριθμητικῶν τιμῶν (x_i, y_i), $i = 1, 2, \ldots, N$ ἢ τὸ ἀντίστοιχον «σημειακὸν νέφος» — μιᾶς καμπύλης (ἢ μιᾶς ἐπιφανείας) παλινδρομήσεως.

Ἡ κατανόησις τῆς σημασίας καὶ ἡ ἑρμηνεία τοῦ δείκτου R^2, ἰδιαιτέρως δὲ ἡ πρακτικὴ χρησιμότης αὐτοῦ, βασίζονται ἐν γένει ἐπὶ ὡρισμένων ἰδιοτήτων τοῦ ἐν λόγῳ μέτρου καὶ συγκεκριμένως εἰς τὸ ὅτι ὁ δείκτης R^2 εἶναι καθαρὸς — ἄνευ μονάδων μετρήσεως — ἀριθμὸς ὁ ὁποῖος

(α) λαμβάνει τιμὰς εἰς τὸ κλειστὸν διάστημα [0, 1], πληροῖ δηλαδὴ τὴν διπλῆν ἀνισότητα $0 \leq R^2 \leq 1$, καὶ

(β) εἶναι ἀνεξάρτητος τῶν χρησιμοποιουμένων κατὰ περίπτωσιν μονάδων μετρήσεως.

Αἱ ἀνωτέρω ὅμως ἰδιότητες, παρ' ὅτι εἰς τὴν πρᾶξιν θεωροῦνται κατὰ κανόνα ὡς ἀπορρέουσαι ἐκ τοῦ τρόπου ὁρισμοῦ τοῦ ἐν λόγῳ δείκτου καὶ κατὰ συνέπειαν ὡς αὐταπόδεκτοι καὶ γενικῶς ἀληθεῖς, ἰσχύουν, ὡς θὰ ἴδωμεν κατωτέρω, μόνον ὑπὸ ὡρισμένας — συγκεκριμένας ἑκάστοτε — συνθήκας καὶ ὡς ἐκ τούτου εἰς πολλὰς περιπτώσεις εἶναι δυνατὸν νὰ μὴ ἀληθεύουν.

Σκοπὸς τοῦ παρόντος ἄρθρου εἶναι ἡ διατύπωσις καὶ ἀπόδειξις τῶν ἐν λόγῳ συνθηκῶν — ἀναγκαίων καὶ ἱκανῶν διὰ τὴν ἰσχὺν τῶν ὡς ἄνω βασικῶν ἰδιοτήτων τοῦ δείκτου R^2 — ὡς καὶ ἡ διερεύνησις τῆς πρακτικῆς σημασίας αὐτῶν εἰς τὰς καθ' ἕκαστα ἐφαρμογάς.

Κατὰ τὴν παρουσίασιν τῶν σχετικῶν ἀποτελεσμάτων ἀναφερόμεθα κατὰ βάσιν — διὰ λόγους ἁπλότητος καὶ συντομίας — εἰς διμεταβλητοὺς στατιστικοὺς πληθυσμοὺς καὶ ἀντιστοίχως εἰς τὴν χρῆσιν τοῦ δείκτου R^2 ὡς μέτρου τοῦ βαθμοῦ προσαρμογῆς μιᾶς καμπύλης παλινδρομήσεως. Ἐξ ὅσων ὅμως ἀκολουθοῦν καθίσταται προφανὲς ὅτι ἡ γενίκευσις τῶν ἐν λόγῳ ἀποτελεσμάτων καὶ ἡ ἐφαρμογὴ αὐτῶν εἰς περιπτώσεις πολυμεταβλητῶν πληθυσμῶν καὶ τῶν προσαρμοζομένων ἀντιστοίχως ἐπιφανειῶν παλινδρομήσεως εἶναι ἄμεσος καὶ ἁπλῆ.

I. Δείκτης προσδιορισμοῦ ἢ προσαρμογῆς μιᾶς καμπύλης παλινδρομήσεως

Ὑποθέσωμεν ὅτι προκειμένου νὰ μελετηθῇ ὁ τρόπος ἀλληλεξαρτήσεως καὶ ὁ βαθμὸς συναφείας, δύο μεταβλητῶν (X, Y) καὶ εἰδικώτερον νὰ περιγραφῇ ἁπλᾶ καὶ συνοπτικὰ ἡ νομοτέλεια ἡ ὁποία διέπει ἐν γένει τὴν διαμόρφωσιν τῶν τιμῶν τῆς μιᾶς — καλουμένης ἐν προκειμένῳ ἐξηρτημένης μεταβλητῆς — ἐν σχέσει πρὸς τὴν διαμόρφωσιν τῶν τιμῶν τῆς ἄλλης — θεωρουμένης ἀντιστοίχως ὡς ἀνεξαρτήτου ἢ ἐπεξηγηματικῆς μεταβλητῆς — ἀποφασίζεται ὑπὸ τοῦ μελετητοῦ ἡ προσαρμογὴ πρὸς τὰ ἐμπειρικὰ δεδομένα — τὰ ζεύγη δηλαδὴ (x_i, y_i), $i=1, 2, \ldots, N$ τῶν τιμῶν τῶν ἐν λόγῳ μεταβλητῶν τὰ ὁποῖα ἀντιστοιχοῦν εἰς τὰς N ἐπὶ μέρους μονάδας τοῦ ὑπὸ ἔρευναν πληθυσμοῦ ἢ καλλίτερον τὰ σημεῖα τοῦ ἀντιστοίχου στικτοῦ διαγράμματος — μιᾶς ἁπλῆς ἐξισώσεως ἢ ἄλλως μιᾶς καμπύλης παλινδρομήσεως τῆς μορφῆς *

$$y = f(x,\alpha,\beta) \tag{1}$$

διὰ τῆς ὁποίας — πέραν τῶν ἄλλων — νὰ καθορίζεται μία συγκεκριμένη ποσοτικὴ διαδικασία ἐπιτρέπουσα ἐκ τῶν τιμῶν τῆς ἀνεξαρτήτου μεταβλητῆς X νὰ ὑπολογίζωνται μὲ ὅσον τὸ δυνατὸν καλλιτέραν προσέγγισιν αἱ ἀντίστοιχοι τιμαὶ τῆς ἐξηρτημένης τοιαύτης Y.

Ὁ προσδιορισμὸς — τῇ βοηθείᾳ πάντοτε τῶν δεδομένων τῆς παρατηρήσεως (x_i, y_i), $i=1, 2, \ldots, N$ — τῶν παραμέτρων α καὶ β τῆς ἐξισώσεως (1), ἐν ἄλλοις δηλαδὴ λόγοις ἡ ἐπιλογὴ μεταξὺ τῶν μελῶν τῆς ἐν λόγῳ οἰκογενείας καμπύλων μιᾶς συγκεκριμένης καμπύλης ἡ ὁποία νὰ διέρχεται ὅσον τὸ δυνατὸν «ἐγγύτερον» ἐκ τῶν σημείων (x_i, y_i), $i=1, 2, \ldots, N$ — συνοψίζουσα οὕτω κατὰ τὸν καλλίτερον ἐν προκειμένῳ δυνατὸν τρόπον τὸ ἀντίστοιχον σημειακὸν νέφος — γίνεται, ὡς γνωστὸν διὰ τῆς μεθόδου τῶν Ἐλαχίστων Τετραγώνων. Συγκεκριμένως, αἱ ζητούμεναι τιμαὶ $\hat{\alpha}$ καὶ $\hat{\beta}$ τῶν ὑπὸ προσδιορισμὸν παραμέτρων α καὶ β εὑρίσκονται δι' ἐπιλύσεως τοῦ συστήματος

$$\begin{aligned} &\frac{\partial}{\partial\alpha}\sum_{i=1}^{N}[\,y_i - f(x_i, \alpha, \beta)\,]^2 = 0 \\ &\frac{\partial}{\partial\beta}\sum_{i=1}^{N}[\,y_i - f(x_i, \alpha, \beta)\,]^2 = 0 \end{aligned} \tag{2}$$

* Ἡ ἐξίσωσις (1) εἶναι δυνατὸν νὰ περιλαμβάνῃ τρεῖς ἢ καὶ περισσοτέρας ἀκόμη παραμέτρους α, β,...

γνωστοῦ ὡς συστήματος τῶν κανονικῶν ἐξισώσεων, ἡ συγκεκριμένη δὲ καμπύλη — μέλος τῆς οἰκογενείας (1) — ἡ ὁποία ἀντιστοιχεῖ εἰς τὰς οὕτω ὑπολογιζομένας τιμὰς $\hat{\alpha}$ καὶ $\hat{\beta}$, συμβολιζομένη ὑπὸ τῆς ἐξισώσεως

$$\hat{y} = f(x, \hat{\alpha}, \hat{\beta}) \tag{3}$$

ἀποτελεῖ τὴν ζητουμένην καμπύλην παλινδρομήσεως.

Ὁ βαθμὸς προσαρμογῆς τῆς οὕτω προσδιοριζομένης καμπύλης παλινδρομήσεως, δηλαδὴ τὸ πόσον ἱκανοποιητικὴ εἶναι ἡ προσέγγισις μὲ τὴν ὁποίαν ἡ ἐν λόγῳ καμπύλη συνοψίζει τὰ δεδομένα τῆς παρατηρήσεως καὶ κατὰ συνέπειαν τὸ «ποσὸν καλὰ» περιγράφει αὕτη τὸν τρόπον ἀλληλεξαρτήσεως τῶν συνεξεταζομένων μεταβλητῶν, ὡς καὶ ἡ ἀκρίβεια τῶν δι' αὐτῆς ἐξαγομένων συμπερασμάτων, χαρακτηρίζονται ὡς γνωστὸν ὑπὸ τοῦ Δείκτου Προσδιορισμοῦ ἢ ἄλλως Δείκτου Προσαρμογῆς τῆς καμπύλης ταύτης, ὁ ὁποῖος ὁρίζεται ἐκ τῆς σχέσεως:

$$R^2 = I - \frac{\sigma^2}{\sigma_y{}^2} \tag{4}$$

ὅπου σ^2 καὶ $\sigma_y{}^2$ συμβολίζουν ἀντιστοίχως τὸ καλούμενον μέσον τετραγωνικὸν σφάλμα περὶ τὴν καμπύλην παλινδρομήσεως καὶ τὴν — συνολικὴν — διακύμανσιν — τῶν τιμῶν — τῆς μεταβλητῆς Y καὶ ὁρίζονται ἐκ τῶν σχέσεων

$$\sigma^2 = \frac{I}{N} \sum_{i=1}^{N} [\, y_i - f(x_i, \hat{\alpha}, \hat{\beta})\,]^2 \tag{5}$$

καὶ

$$\sigma^2 = \frac{I}{N} \sum_{i=1}^{N} (\, y_i - \mu_y\,)^2 \tag{6}$$

τοῦ μ_y συμβολίζοντος τὸν μέσον ἀριθμητικὸν — τῶν παρατηρηθεισῶν τιμῶν — τῆς Y.

Ἡ σημασία τοῦ δείκτου R^2 ὡς μέτρου τοῦ βαθμοῦ προσαρμογῆς τῆς προσδιοριζομένης κατὰ περίπτωσιν καμπύλης παλινδρομήσεως καὶ εἰδικώτερον

(i) ἡ ἀξιολόγησις τῆς ἀριθμητικῆς τιμῆς αὐτοῦ καὶ ἡ ἐξαγωγὴ συμπερασμάτων ἀναφορικῶς πρὸς τὴν προσέγγισιν μὲ τὴν ὁποίαν ὡρισμένη καμπύλη συνοψίζει τὰ δεδομένα τῆς παρατηρήσεως ἢ ἄλλως «πόσον καλὰ» περιγράφει αὕτη τὸν τρόπον ἀλληλεξαρτήσεως τῶν συνεξεταζομένων μεταβλητῶν, καὶ

(ii) ἡ σύγκρισις τῶν τιμῶν δύο ἢ περισσοτέρων τοιούτων δεικτῶν — ἀναφερομένων εἰς διαφόρους καμπύλας—πρὸς ἐπιλογὴν τῆς καταλληλοτέρας ἑκάστοτε μορφῆς τῆς καμπύλης παλινδρομήσεως

βασίζονται, ὡς ἤδη ἐλέχθη, ἐπὶ τῶν ἀναφερομένων ἀνωτέρω ἰδιοτήτων του. Ἡ διερεύνησις τῶν ἐν λόγῳ ἰδιοτήτων καὶ ἰδιαιτέρως αἱ συνθῆκαι διὰ τὴν ἰσχὺν αὐτῶν εἰς τὰς καθ' ἕκαστα ἐφαρμογὰς θὰ μᾶς ἀπασχολήσουν λεπτομερῶς ἀμέσως κατωτέρω.

II. Βασικαὶ ἰδιότητες τοῦ δείκτου προσδιορισμοῦ

1η Ἰδιότης: «Ὁ δείκτης προσδιορισμοῦ R^2 εἶναι ἀριθμὸς καθαρὸς — ἄνευ μονάδων μετρήσεως — καὶ συνεπῶς πάντοτε συγκρίσιμος».

Ἡ ἀπόδειξις τῆς ἐν λόγῳ ἰδιότητος — ἰσχυούσης ἄνευ συνθηκῶν — εἶναι ἁπλουστάτη. Πράγματι, τόσον τὸ μέσον τετραγωνικὸν σφάλμα σ^2 ὑπολογιζόμενον ἐκ τοῦ τύπου (5), ὅσον καὶ ἡ διακύμανσις τῆς μεταβλητῆς Y — τύπος (6) — ἐκφράζονται εἰς τὰς μονάδας μετρήσεως — τετραγωνισμένας — τῆς ἐξηρτημένης μεταβλητῆς Y καὶ κατὰ συνέπειαν ὁ λόγος αὐτῶν $\sigma^2 : \sigma_y{}^2$ ὡς καὶ ὁ δείκτης R^2 — ὁριζόμενος ἐκ τοῦ τύπου (4) — εἶναι ἀριθμοὶ καθαροί.

Ἡ πρακτικὴ χρησιμότης τῆς ἐν λόγῳ ἰδιότητος συνίσταται, ὡς εἶναι εὐνόητον, εἰς τὸ ὅτι δι' αὐτῆς καθίσταται δυνατὴ ἡ σύγκρισις τιμῶν τοῦ δείκτου R^2 αἱ ὁποῖαι ἀναφέρονται εἰς διαφόρους διμεταβλητοὺς στατιστικοὺς πληθυσμοὺς ἢ ἄλλως εἰς μεταβλητὰς ἐκφραζομένας εἰς διαφόρους μονάδας μετρήσεως, πρᾶγμα τὸ ὁποῖον εἶναι, ὡς γνωστόν, ἀνέφικτον διὰ τὰ ἀντίστοιχα μέσα τετραγωνικὰ σφάλματα.

2α Ἰδιότης: «Ὁ δείκτης R^2 λαμβάνει τιμὰς εἰς τὸ κλειστὸν διάστημα $[0, 1]$, πληροῖ δηλαδὴ τὴν διπλῆν ἀνισότητα

$$0 \leqslant R^2 \leqslant 1 \qquad (7)$$

καὶ ὡς ἐκ τούτου ἡ ἀξιολόγησις τῶν τιμῶν αὐτοῦ καθίσταται ἁπλῆ καὶ ἀντικειμενική».

Συγκεκριμένως, ὡς θὰ ἀποδείξωμεν κατωτέρω, ἡ ἰσότης $R^2 = 1$ λαμβάνει χώραν ἐὰν — καὶ μόνον ἐὰν — μεταξὺ τῶν συνεξεταζομένων μεταβλητῶν (X, Y) ὑφίσταται συναρτησιακὴ σχέσις τῆς μορφῆς (1) ἐν ἄλλοις δηλαδὴ λόγοις ἐὰν ἡ κατὰ τὰ ὡς ἄνω προσδιοριζομένη καμπύλη παλινδρομήσεως (3) διέρχεται δι' ὅλων τῶν σημείων $(x_i, y_i), i = 1, 2, \ldots, N$ — πλήρης προσαρμογὴ — καὶ περιγράφει τὴν ὑφισταμένην μεταξὺ τῶν ἐν λόγῳ μεταβλητῶν σχέσιν κατὰ τρόπον ὥστε ἐκ τῶν τιμῶν τῆς ἀνεξαρτήτου μεταβλητῆς X νὰ ὑπολογίζωνται αἱ ἀντίστοιχοι τιμαὶ τῆς ἐξηρτημένης τοιαύτης Y μὲ ἀπόλυτον ἀκρίβειαν (ἄνευ σφάλματος).

Ἐξ ἄλλου — ὡς θὰ ἀποδειχθῇ ἐπίσης — ὁ δείκτης R^2 λαμβάνει τὴν τιμὴν μηδὲν ($R^2 = 0$), ἐὰν — καὶ μόνον ἐὰν — ἡ διὰ τῆς μεθόδου τῶν ἐλαχίστων τετραγώνων ἐπιλεγομένη γραμμὴ παλινδρομήσεως εἶναι ἡ παράλληλος πρὸς τὸν ἄξονα τῶν x εὐθεῖα

$$\hat{y} = \mu_y = \frac{1}{N} \sum_{i=1}^{N} y_i \qquad (8)$$

ἐν ἄλλοις δηλαδὴ λόγοις ἐὰν ἡ διαμόρφωσις τῶν τιμῶν τῆς Y οὐδόλως ἐπηρρεάζεται ἢ σχετίζεται πρὸς τὴν διαμόρφωσιν τῶν τιμῶν τῆς X, ἤτοι εἰς τὴν περίπτωσιν μὴ σ υ σ χ ε τ ι σ μ έ ν ω ν — στατιστικῶς — μεταβλητῶν.

Ἐξυπακούεται ἐκ τῶν ἀνωτέρω ὅτι εἰς πᾶσαν ἄλλην περίπτωσιν ἡ τιμὴ τοῦ R^2 εὑρίσκεται μεταξὺ 0 καὶ 1, πλησιάζει δὲ πρὸς τὸ 1 ὅσον ἐντονωτέρα εἶναι ἡ μεταξὺ τῶν ἐν λόγῳ μεταβλητῶν συνάφεια καὶ ἡ μορφὴ τῶν καμπύλων τῆς οἰκογενείας (1), καταλληλοτέρα διὰ τὴν περιγραφὴν τῆς ἀντιστοίχου μεταξὺ αὐτῶν σχέσεως.

Ἡ ἀπόδειξις καὶ τῆς παρούσης ἰδιότητος εἶναι κατὰ τὸ μᾶλλον ἢ ἧττον ἁπλῆ. Ἀντιθέτως ὅμως πρὸς τὴν προηγουμένην τοιαύτην ἡ ἐν λόγῳ ἰδιότης δ ὲ ν ε ἶ ν α ι κ α θ ο λ ι κ ῆ ς ἰ σ χ ύ ο ς, ἀληθεύει δὲ μ ό ν ο ν ὑπὸ τὴν ἑξῆς προϋπόθεσιν (συνθήκην):

«Ἡ οἰκογένεια τῶν καμπυλων $y=f(x,\alpha,\beta)$ μεταξὺ τῶν μελῶν τῆς ὁποίας ἀναζητεῖται — διὰ τῆς μεθόδου τῶν ἐλαχίστων τετραγώνων — ἡ συγκεκριμένη καμπύλη παλινδρομήσεως $\hat{y} = f(x, \hat{\alpha}, \hat{\beta})$ δέον νὰ περιλαμβάνῃ ὡς μέλη αὐτῆς ἁπάσας τὰς παραλλήλους πρὸς τὸν ἄξονα τῶν x εὐθείας, ἐν ἄλλοις δηλαδὴ λόγοις ἡ ἐξίσωσις $y = f(x, \alpha, \beta)$ νὰ δύναται διὰ καταλλήλους τιμὰς τῶν παραμέτρων α καὶ β νὰ λάβῃ τὴν μορφὴν $y = c$ ὅπου c μία αὐθαίρετος σταθερά».

Ὑπὸ τὴν ἀνωτέρω προϋπόθεσιν προβαίνομεν τώρα εἰς τὴν ἀπόδειξιν τῆς διπλῆς ἀνισότητος $0 \leqslant R^2 \leqslant 1$.

Πρὸς τὸν σκοπὸν αὐτὸν ἀρκεῖ νὰ ἀποδείξωμεν — ὑπὸ ὑπὸ τὴν ὡς ἄνω συνθήκην — ὅτι ἰσχύει πάντοτε ἡ ἀνισότης

$$\sigma^2 \leqslant \sigma_y{}^2 \tag{9}$$

καθ' ὅσον ἐκ τῆς (9) προκύπτει ἀμέσως ἡ σχέσις *

$$\frac{\sigma^2}{\sigma_y{}^2} \leqslant 1 \quad \text{ἢ} \quad 1 - \frac{\sigma^2}{\sigma_y{}^2} \geqslant 0 \quad \text{ἤτοι} \quad R^2 \geqslant 0$$

ἐνῷ παραλλήλως ἐκ τῆς προφανοῦς σχέσεως $\frac{\sigma^2}{\sigma_y{}^2} \geqslant 0$ θὰ ἔχωμεν ὅτι

$$-\frac{\sigma^2}{\sigma_y{}^2} \leqslant 0 \quad \text{ἢ} \quad 1 - \frac{\sigma^2}{\sigma_y{}^2} \leqslant 1 \quad \text{ἤτοι} \quad R^2 \leqslant 1$$

Πρὸς ἀπόδειξιν τῆς ἀνισότητος (9) ἐργαζόμεθα ὡς ἑξῆς: Ὑποθέσωμεν πρὸς στιγμὴν ὅτι ἡ ἐπιλογὴ τῆς γραμμῆς παλινδρομήσεως π ε ρ ι ο ρ ί ζ ε τ α ι μόνον μεταξὺ τῶν παραλλήλων πρὸς τὸν ἄξονα τῶν x εὐθειῶν $y=c$. Τὸ μ.τ.σ. περὶ οἱανδήποτε τῶν εὐθειῶν αὐτῶν εἶναι ὡς γνωστὸν ἡ ποσότης

* Ἐκτὸς ἐὰν $\sigma_y{}^2 = 0$ πρᾶγμα τὸ ὁποῖον σημαίνει ὅτι $y_i = \mu_y$, $i = 1,2,\ldots,N$ καὶ κατὰ συνέπειαν ὅτι εὑρισκόμεθα εἰς τὴν ὁριακὴν περίπτωσιν κατὰ τὴν ὁποίαν ἅπαντα τὰ σημεῖα (x_i, y_i), $i = 1,2,\ldots,N$ εὑρίσκονται ἐπὶ τῆς παραλλήλου πρὸς τὸν ἄξονα τῶν χ εὐθείας $y = \mu_y$ καὶ αἱ μεταβληταὶ X καὶ Y εἶναι προφανῶς στατιστικῶς ἀσυσχέτιστοι.

$$\frac{1}{N}\sum_i (y_i - c)^2$$

Θέτοντες τὴν πρώτην παράγωγον τῆς ἐν λόγῳ ποσότητος — ἢ ἁπλούστερον τοῦ ἀριθμητοῦ αὐτῆς — ὡς πρὸς c ἴση πρὸς μηδὲν προκύπτει ἡ ἰσότης

$$(-2)\sum_i (y_i - c) = 0$$

ἐκ τῆς ὁποίας λαμβάνομεν

$$\hat{c} = \frac{1}{N}\sum_i y_i = \mu_y$$

Οὕτως, ἡ ἐπιλεγομένη ἐν προκειμένῳ διὰ τῆς μεθόδου τῶν ἐλαχίστων τετραγώνων γραμμὴ παλινδρομήσεως εἶναι ἡ παράλληλος πρὸς τὸν ἄξονα τῶν x εὐθεῖα

$$\hat{y} = \mu_y$$

τὸ δὲ μ.τ.σ. περὶ αὐτὴν

$$\sigma^2 = \frac{1}{N}\sum_i (y_i - \mu_y)^2$$

εἶναι ἴσον πρὸς τὴν διακύμανσιν $\sigma_y{}^2$ τῶν τιμῶν τῆς Y. Εἰς τὴν πραγματικότητα ὅμως ἡ γραμμὴ παλινδρομήσεως (3) ἐπιλέγεται μεταξὺ τῶν μελῶν τῆς οἰκογενείας (1), ἡ ὁποία πέραν τῶν εὐθειῶν $y=c$ — αἱ ὁποῖαι ἀποτελοῦν μέλη αὐτῆς συμφώνως πρὸς τὴν τεθεῖσαν συνθήκην—περιλαμβάνει ἐν γένει καὶ ἄλλας καμπύλας. Κατὰ συνέπειαν,τὸ μ.τ.σ. σ^2 περὶ τὴν ἐπιλεγεῖσαν καμπύλην παλινδρομήσεως (3) εἶναι ἀ δ ύ ν α τ ο ν νὰ ὑ π ε ρ β α ί ν η τὴν διακύμανσιν $\sigma_y{}^2$, διότι ἐν τοιαύτῃ περιπτώσει ἡ μέθοδος τῶν ἐλαχίστων τετραγώνων θὰ ὁδήγει εἰς τὴν ἐπιλογὴν ὡς γραμμῆς παλινδρομήσεως τῆς εὐθείας $y=\mu_y$ τὸ μ.τ.σ. τῆς ὁποίας εἶναι, ὡς εἴδομεν ἀνωτέρω, ἴσον πρὸς $\sigma_y{}^2$. Οὕτω συνάγεται τὸ συμπέρασμα ὅτι τὸ μ.τ.σ. σ^2 περὶ τὴν γραμμὴν παλινδρομήσεως (3), εἶναι πάντοτε μικρότερον ἢ τὸ πολὺ ἴσον πρὸς τὴν διακύμανσιν $\sigma_y{}^2$, ἤτοι ἡ ἀνισότης $\sigma^2 \leqslant \sigma_y{}^2$ τὴν ὁποίαν ἐχρησιμοποιήσαμεν ἀνωτέρω πρὸς ἀπόδειξιν τῆς διπλῆς ἀνισότητος $0 \leqslant R^2 \leqslant 1$.

Μετὰ τὴν ἀπόδειξιν τῆς διπλῆς ἀνισότητος (7) εἶναι χρήσιμον νὰ σχολιάσωμεν δι' ὀλίγων τὴν σημασίαν τῶν διαφόρων τιμῶν τοῦ δείκτου R^2 καὶ νὰ διευκρινισθοῦν τὰ ἑξῆς :

Κατ' ἀρχὴν εἶναι προφανὲς ὅτι ὅσον μ ε γ α λ υ τ έ ρ α εἶναι ἡ τιμὴ τοῦ R^2 — ἡ ὁποία σημειωτέον ἐξαρτᾶται οὐσιαστικῶς μόνον ἐκ τοῦ μ.τ.σ. σ^2 — τόσον μ ι κ ρ ο τ έ ρ α εἶναι ἡ σχετικὴ διασπορὰ τῶν σημείων (x_i, y_i), $i=1,2,\ldots, N$ περὶ τὴν ἐπιλεγεῖσαν γραμμὴν παλινδρομήσεως καὶ κατὰ συνέπειαν τόσον μ ε γ α λ ύ τ ε ρ ο ς ὁ βαθμὸς προσαρμογῆς τῆς ἐν λόγῳ καμπύλης καὶ ἀ κ ρ ι β έ σ τ ε ρ α τὰ δι' αὐτῆς ἐξαγόμενα συμπεράσματα.

Εἰδικώτερον, ἡ ἰσότης $R^2 = 1$ λαμβάνει προφανῶς χώραν μόνον ἐὰν ἡ ἐπιλεγεῖσα καμπύλη παλινδρομήσεως (3) διέρχεται δι' ὅλων τῶν σημείων (x_i, y_i) τοῦ ἀντιστοίχου στικτοῦ διαγράμματος, ἐὰν δηλαδὴ ἡ προσαρμογὴ αὐτῆς εἶναι πλήρης. Πράγματι, εἰς μίαν τοιαύτην περίπτωσιν — καὶ μόνον τότε — αἱ ἐξ ὑπολογισμοῦ — οἱονεὶ — τιμαὶ τῆς Y ταυτίζονται πρὸς τὰς ἀντιστοίχους ἐμπειρικὰς τοιαύτας, τὸ μ.τ.σ. σ^2 μηδενίζεται καὶ ὁ δείκτης R^2 — ὡς προκύπτει ἐκ τοῦ τύπου (4) — γίνεται ἴσος πρὸς τὴν μονάδα.

Ἀντιστρόφως, ἡ ἰσότης $R^2 = 0$ λαμβάνει χώραν ἐὰν — καὶ μόνον ἐὰν — ἡ ἐπιλεγομένη διὰ τῆς μεθόδου τῶν ἐλαχίστων τετραγώνων γραμμὴ παλινδρομήσεως εἶναι ἡ παράλληλος πρὸς τὸν ἄξονα τῶν x εὐθεῖα $y = \mu_y$ — ὅτε τὸ ἀντίστοιχον μ.τ.σ. ἰσοῦται, ὡς εἴδομεν, πρὸς τὴν διακύμανσιν σ_y^2 — ἤτοι εἰς τὴν περίπτωσιν μὴ συσχετισμένων μεταβλητῶν.

Θὰ ἀποδείξωμεν τώρα — δι' ἑνὸς ἁπλοῦ ἀριθμητικοῦ παραδείγματος — ὅτι ἐὰν ἡ οἰκογένεια τῶν καμπύλων $y = f(x, \alpha, \beta)$ — ἐκ τῆς ὁποίας ἐπιλέγεται ἡ γραμμὴ παλινδρομήσεως — δὲν περιλαμβάνει τὰς παραλλήλους πρὸς τὸν ἄξονα τῶν x εὐθείας $y = c$ (δὲν πληροῦται δηλαδὴ ἡ τεθεῖσα συνθήκη) εἶναι δυνατὸν ὁ δείκτης R^2 νὰ λάβῃ καὶ ἀρνητικὰς τιμάς.

Ὑποθέσωμεν ὅτι τὰ ἐμπειρικὰ δεδομένα συνίστανται ἐκ τῶν ἀριθμητικῶν ζευγῶν τῶν δύο πρώτων στηλῶν τοῦ κατωτέρου πίνακος (1).

Πίναξ 1

x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	9	1	9	81
2	8	4	16	64
3	10	9	30	100
4	5	16	20	25
10	32	30	75	270

Πρὸς τούτοις, ὑποθέσωμεν ὅτι ὁ μελετητὴς ἀπεφάσισε νὰ ἐπιλέξῃ τὴν γραμμὴν παλινδρομήσεως μεταξὺ τῶν μελῶν τῆς οἰκογενείας $y = \beta x$ ἡ ὁποία προφανῶς δὲν περιλαμβάνει εὐθείας παραλλήλους πρὸς τὸν ἄξονα τῶν x. Τὸ σύστημα τῶν κανονικῶν ἐξισώσεων (2) συνίσταται ἐν προκειμένῳ ἐκ μόνης τῆς ἐξισώσεως

$$\beta \Sigma x_i^2 = \sum_i x_i y_i$$

Ὑπολογίζοντες ἐκ τῶν δεδομένων τῆς παρατηρήσεως τὰ ἀθροίσματα

$$\sum_i x_i^2 \text{ καὶ } \sum_i x_i y_i$$

— ὡς ἐμφαίνεται ἀντιστοίχως εἰς τὴν τρίτην καὶ τετάρτην στήλην τοῦ πίνακος (1) — καὶ ἀντικαθιστῶντες τὰς τιμὰς αὐτῶν εἰς τὴν ὡς ἄνω ἐξίσωσιν λαμβάνομεν $\hat{\beta}=2{,}5$. Οὕτως, ἡ ἐπιλεγεῖσα ἐν προκειμένῳ γραμμὴ παλινδρομήσεως εἶναι ἡ εὐθεῖα $\hat{y}=2{,}5x$.

Τὸ μ.τ.σ. περὶ αὐτὴν ὑπολογιζόμενον ἐκ τοῦ τύπου (5), ὁ ὁποῖος εἰς τὴν παρούσαν συγκεκριμένην περίπτωσιν λαμβάνει τὴν μορφὴν

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta})^2 = \frac{1}{N}\sum_{i=1}^{4}(y_i - 2{,}5x_i)^2$$

εἶναι $\sigma^2 = 20{,}625$.

Ἐξ ἄλλου, ἡ διακύμανσις σ_y^2 ὑπολογιζομένη ἐκ τοῦ τύπου (6) εἶναι $\sigma_y^2 = 3{,}5$. Κατὰ συνέπειαν, ὁ δείκτης R^2 ὑπολογιζόμενος ἐκ τοῦ τύπου (4) λαμβάνει ἐν προκειμένῳ τὴν ἀρνητικὴν τιμὴν

$$R^2 = 1 - \frac{20{,}625}{3{,}5} = -4{,}89$$

Ἐνταῦθα δέον νὰ τονισθῇ ὅτι ἡ ὡς ἄνω συνθήκη οὐδόλως ἀποτελεῖ δέσμευσιν εἰς τὴν διαδικασίαν ἐπιλογῆς τῆς καταλλήλου ἑκάστοτε γραμμῆς παλινδρομήσεως. Ἀντιθέτως, αὕτη εἶναι ἀπολύτως ἐπιβεβλημένη πρὸς ἀντιμετώπισιν τῶν περιπτώσεων ὅπου ἡ διαμόρφωσις τῶν τιμῶν τῆς Υ δὲν σχετίζεται καθ' οἱονδήποτε τρόπον πρὸς τὴν διαμόρφωσιν τῶν τιμῶν τῆς Χ καὶ κατὰ συνέπειαν μεταβαλλομένης τῆς Χ ἡ μεταβλητὴ Υ λαμβάνει εἴτε τὴν αὐτὴν πάντοτε τιμὴν εἴτε γενικώτερον τιμὰς κυμαινομένας κατὰ μὴ συστηματικὸν τρόπον πέριξ μιᾶς εὐθείας παραλλήλου πρὸς τὸν ἄξονα τῶν x (περίπτωσις ἀσυσχετίστων μεταβλητῶν).

Ἐξ ἄλλου, αἱ πλεῖσται τῶν χρησιμοποιουμένων εἰς τὴν πρᾶξιν μορφαὶ καμπύλων παλινδρομήσεως πληροῦν τὴν ὡς ἄνω συνθήκην. Συγκεκριμένως ἡ εὐθεῖα $y=\alpha+\beta x$ ὡς καὶ αἱ καμπύλαι πολυωνυμικῆς μορφῆς δευτέρου ἢ ἀνωτέρου βαθμοῦ, $y=\alpha+\beta x+\gamma x^2$, $y=\alpha+\beta x+\gamma x^2+\delta x^3$ κ.λ.π. ὑπὸ τὴν προϋπόθεσιν ὅτι περιλαμβάνουν τὸν γνωστὸν ὡς σταθερὸν ὅρον — παράμετρος α — πληροῦν προφανῶς τὴν ὡς ἄνω συνθήκην καθ' ὅσον διὰ μηδενισμοῦ τῶν ὑπολοίπων παραμέτρων των λαμβάνουν τὴν μορφὴν $y=c$. Ὁμοίως αἱ καμπύλαι ὑπερβολικῆς μορφῆς $1/y=\alpha+\beta x$, $y=\alpha+\beta/x$ κλπ. (διὰ $\beta=0$), ἐκθετικῆς μορφῆς $y=\alpha\beta^x$ (διὰ $\beta=1$), γεωμετρικῆς μορφῆς $y=\alpha x^\beta$ (διὰ $\beta=0$), ἡ γενικευμένη ἐκθετικὴ καμπύλη $y=k+\alpha\beta^x$ (διὰ $\alpha=0$, $\beta=1$), ἡ λογιστικὴ καμπύλη

$$y = \frac{k}{1+e^{\alpha+\beta^x}}$$

(διὰ $\alpha=\beta=0$) περιλαμβάνουν προφανῶς — ὡς ὑποοικογένειαν — τὰς παραλλήλους πρὸς τὸν ἄξονα τῶν x εὐθείας $y=c$. Ἐκ τῶν ἀνωτέρω βεβαίως καθίσταται προφανὲς ὅτι εἰς οὐδεμίαν περίπτωσιν ἡ καμπύλη παλινδρομήσεως θὰ πρέπει νὰ ἀναζητῆται μεταξὺ οἰκογενειῶν μὲ ἐξίσωσιν τῆς μορφῆς $y=\beta\chi$ ἢ $y=\beta x+\gamma x^2$ ἢ $y=\beta x+\gamma x^2+\delta x^3$ ἢ ἄλλην πολυωνυμικὴν μορφὴν ἄνευ σταθεροῦ ὅρου ὡς ἐπίσης μεταξὺ τῶν ἐκθετικῶν καμπύλων $y=\beta^x$, τῶν γεωμετρικῶν τοιούτων $y=x^\beta$ ἢ διαφόρων ἄλλων καμπύλων ὡς π.χ. $y=\beta\log x$, $y=\beta\log x+\gamma(\log x)^2$, $\sqrt{y}=\beta\chi$ κ.ο.κ. αἱ ὁποῖαι προφανῶς δὲν περιλαμβάνουν τὰς εὐθείας $y=c$, καθ' ὅσον εἶναι δυνατὸν νὰ ὁδηγηθῶμεν εἰς τὸ ἄτομον τοῦ προηγουμένου παραδείγματος, ἀρνητικὰς δηλαδὴ τιμὰς τοῦ δείκτου R^2 ἄνευ οὐδεμιᾶς πρακτικῆς ἀξίας.

3η Ἰδιότης. «Ὁ δείκτης R^2 εἶναι ἀνεξάρτητος τῶν χρησιμοποιουμένων μονάδων μετρήσεως, ὡς ἐπίσης καὶ τῆς λαμβανομένης ἑκάστοτε ὡς ἀρχῆς αὐτῶν. Ἐν ἄλλοις δηλαδὴ λόγοις ἡ τιμὴ τοῦ ἐν λόγῳ δείκτου παραμένει ἐν γένει ἀναλλοίωτος εἰς γραμμικοὺς μετασχηματισμοὺς τῶν μεταβλητῶν».

Ἡ ἰδιότης αὕτη ἐπιτρέπουσα τὴν ἑρμηνείαν καὶ κατανόησιν τῆς σημασίας τοῦ ἐν λόγῳ δείκτου ἄνευ οἱωνδήποτε προσθέτων πληροφοριῶν — ἀναφερομένων εἰς τὰς μονάδας μετρήσεως, τὴν ἀρχὴν αὐτῶν κ.λ.π. — ἰσχύει ὅπως καὶ ἡ προηγουμένη τοιαύτη ὑπὸ ὡρισμένας μόνον συνθήκας καὶ συγκεκριμένως τὰς ἑξῆς:

Διὰ τὴν ἀνεξαρτησίαν τοῦ δείκτου R^2 ἐκ τῶν μονάδων μετρήσεως θὰ πρέπει «Ἡ οἰκογένεια τῶν καμπύλων $y=f(x,\alpha,\beta)$ — μεταξὺ τῶν μελῶν τῆς ὁποίας ἀναζητεῖται ἡ συγκεκριμένη καμπύλη παλινδρομήσεως $\hat{y}=f(x,\hat{\alpha},\hat{\beta})$ — νὰ εἶναι τοιαύτη ὥστε, ἐὰν ἡ καμπύλη $y=f(x,\alpha',\beta')$ ἀποτελεῖ μέλος αὐτῆς καὶ ἡ καμπύλη $y=\lambda f(x,\alpha',\beta')$, ὅπου λ αὐθαίρετος σταθερά, νὰ εἶναι ἐπίσης μέλος τῆς οἰκογενείας».

Γενικώτερον ὅμως ἵνα ὁ δείκτης R^2 εἶναι ἀνεξάρτητος — πέραν τῶν μονάδων μετρήσεως — καὶ τῆς λαμβανομένης ὡς ἀρχῆς αὐτῶν, δέον ὅπως:

«καὶ ἡ καμπύλη $y=k+\lambda f(x,\alpha',\beta')$, ὅπου k, λ αὐθαίρετοι σταθεραί, νὰ ἀποτελῇ ὡσαύτως μέλος τῆς ἀρχικῆς οἰκογενείας».

Ἐνταῦθα δέον νὰ σημειωθῇ ὅτι δι' ὡρισμένας οἰκογενείας καμπύλων * ἡ ἐν λόγῳ συνθήκη δὲν πληροῦται μὲ ἀποτέλεσμα ἡ τιμὴ τοῦ δείκτου R^2 εἰς τὰς περιπτώσεις αὐτὰς νὰ ποικίλλῃ ἀναλόγως τῶν χρησιμοποιουμένων μονάδων μετρήσεως καὶ τῆς λαμβανομένης ἑκάστοτε ὡς ἀρχῆς (τῶν μετρήσεων).

Διὰ τὸν λόγον αὐτὸν πρὸ τῆς προσαρμογῆς — πρὸς τὰ ἑκάστοτε ἐμπειρικὰ δεδομένα — μιᾶς καμπύλης παλινδρομήσεως θὰ πρέπει νὰ ἐλέγχεται κατὰ πόσον ἡ ἀντίστοιχος ἐξίσωσις πληροῖ τὴν ὡς ἄνω συνθήκην καθ' ὅσον ἐν ἐναντίᾳ περιπτώσει ὁ δείκτης προσδιορισμοῦ R^2 τυγχάνει, ὡς εἶναι εὐνόητον, λίαν περιωρισμένης — ἂν ὄχι ἄνευ οὐδεμιᾶς σημασίας — καὶ ἡ χρῆσις αὐτοῦ εἰς τὴν πρᾶξιν καθίσταται ἐξαιρετικῶς δυσχερής.

---

* Ἴδε σχετικὸν παράδειγμα κατωτέρω.

Ἡ ἀπόδειξις τῆς ὡς ἄνω ἰδιότητος τοῦ δείκτου R^2 ὑπὸ τὴν γενικὴν μορφὴν αὐτῆς — ἤτοι τὸ ἀναλλοίωτον αὐτοῦ εἰς ἀλλαγὴν τῶν μονάδων μετρήσεως καὶ ταυτόχρονον μεταφορὰν τῆς ἀρχῆς τῶν ἀξόνων — δίδεται κατωτέρω. Διὰ τὴν ἀπόδειξιν τῆς μερικῆς περιπτώσεως—ἤτοι τῆς ἀνεξαρτηρίας αὐτοῦ ἐκ τῶν μονάδων μετρήσεως — ἀρκεῖ εἰς ὅτι ἀκολουθεῖ νὰ ὑποτεθῇ ὅτι $k=0$.

Ὑποθέσωμεν ὅτι ἡ ἐξηρτημένη μεταβλητὴ Y μετρουμένη μὲ δύο διαφόρους μονάδας μετρήσεως καὶ μὲ δύο διαφορετικὰς ἀρχὰς τῶν μετρήσεων λαμβάνει ἀντιστοίχως τὰς τιμὰς y_i καὶ z_i ὅπου $i=1,2,\ldots, N$ αἱ ἐπὶ μέρους μονάδες τοῦ ὑπὸ ἔρευναν συνόλου (πληθυσμοῦ). Ἐὰν λ εἶναι ὁ λόγος τῶν δύο μονάδων μετρήσεως καὶ k ἡ ἀπόστασις τῶν δύο σημείων—ἀρχῶν τῶν μετρήσεων, θὰ ἔχωμεν ὡς γνωστὸν ὅτι

$$z_i = k+\lambda y_i \quad \text{διὰ } i=1,2,\ldots,N$$

Ὑπὸ τὴν προϋπόθεσιν λοιπὸν τῆς δευτέρας — τῆς γενικωτέρας — ἐκ τῶν ὡς ἄνω συνθηκῶν θὰ ἀποδείξωμεν ὅτι ὁ δείκτης R^2 λαμβάνει τὴν αὐτὴν ἀκριβῶς τιμὴν ἀνεξαρτήτως ἐὰν διὰ τὸν ὑπολογισμὸν αὐτοῦ χρησιμοποιηθοῦν αἱ τιμαὶ y_i ἢ αἱ τιμαὶ z_i, $i=1,2,\ldots,N$. Πρὸς τοῦτο ἐργαζόμεθα ὡς ἑξῆς:

Ὑποθέσωμεν ὅτι χρησιμοποιουμένων τῶν τιμῶν y_i, $i=1,2,\ldots,N$ ἡ ἐπιλεγομένη καμπύλη παλινδρομήσεως ἔχει ἐξίσωσιν

$$\hat{y} = f(x, \hat{\alpha}, \hat{\beta})$$

ἐνῷ χρησιμοποιουμένων τῶν τιμῶν z_i, $i=1,2,\ldots, N$ ἐπιλέγεται ἀντιστοίχως ἡ καμπύλη

$$\hat{z} = f(x, \alpha', \beta')$$

Ἔστω ἀκόμη ὅτι τὸ μ.τ.σ. περὶ τὴν καμπύλην $\hat{y} = f(x, \hat{\alpha}, \hat{\beta})$ εἶναι σ^2 ἤτοι

$$\frac{1}{N}\sum_i [\, y_i - f(x_i, \hat{\alpha}, \hat{\beta})]^2 = \sigma^2$$

Θὰ δείξωμεν ὅτι

$$f(x, \alpha', \beta') = k + \lambda f(x, \hat{\alpha}, \hat{\beta}) \tag{10}$$

Τὸ μ.τ.σ. περὶ τὴν καμπύλην $\hat{z} = k + \lambda f(x, \hat{\alpha}, \hat{\beta})$ εἶναι

$$\frac{1}{N}\sum_i [\, z_i - k - \lambda f(x_i, \hat{\alpha}, \hat{\beta})]^2 = \frac{1}{N}\sum_i [\, \lambda y_i - \lambda f(x_i, \hat{\alpha}, \hat{\beta})]^2 = \lambda^2\sigma^2$$

Κατὰ συνέπειαν, ἐὰν ὑποθέσωμεν ὅτι δὲν ἀληθεύει ἡ σχέσις (10) θὰ πρέπει τὸ μ.τ.σ. περὶ τὴν καμπύλην $\hat{z}= f(\chi, \alpha',\beta')$ νὰ εἶναι μικρότερον τοῦ $\lambda^2\ \sigma^2$ ἤτοι

$$\frac{1}{N}\sum_i [z_i - f(x, \alpha', \beta')]^2 < \lambda^2\sigma^2$$

Τοῦτο ὅμως θὰ ὁδήγει εἰς τὸ ἄτοπον τὸ μ.τ.σ. περὶ τὴν καμπύλην

$$\hat{y} = \frac{1}{\lambda}[f(x, \alpha', \beta') - k]$$

νὰ εἶναι μικρότερον τοῦ σ^2 καὶ κατὰ συνέπειαν ἡ καμπύλη $\hat{y} = f(x, \hat{\alpha}, \hat{\beta})$ νὰ μὴ εἶναι, ὡς ὑπετέθη, ἡ καμπύλη παλινδρομήσεως ἐλαχίστου μ.τ.σ. Πράγματι τὰ μ.τ.σ. περὶ τὴν καμπύλην $\hat{y} = \frac{1}{\lambda}[f(x, \alpha', \beta') - k]$ εἶναι

$$\frac{1}{N}\sum_i [y_i - \frac{1}{\lambda}(f(x_i, \alpha', \beta') - k)]^2 = \frac{1}{N}\sum_i [\frac{1}{\lambda}z_i - \frac{1}{\lambda}f(x_i, \alpha', \beta')]^2 =$$

$$= \frac{1}{\lambda^2}\frac{1}{N}\sum_i [z_i - f(x_i, \alpha', \beta')]^2 < \sigma^2$$

Ἐκ τῶν ἀνωτέρω συνάγεται ὅτι ἡ σχέσις (10) ἀ λ η θ ε ύ ε ι καὶ κατὰ συνέπειαν τὸ μ.τ.σ. περὶ τὴν καμπύλην $\hat{z} = f(x, \alpha', \beta')$ εἶναι ἴσον πρὸς $\lambda^2\sigma^2$ ἤτοι

$$\frac{1}{N}\sum_i [z_i - f(x_i\ \alpha', \beta')]^2 = \lambda^2\sigma^2$$

Ἐξ ἄλλου, ἐὰν ἡ διακύμανσις τῶν τιμῶν y_i, $i=1,2,\ldots,N$ τῆς μεταβλητῆς Y εἶναι $\sigma_y{}^2$ ἡ διακύμανσις τῶν τιμῶν z_i, $i = 1,2,\ldots,N$ εἶναι ὡς γνωστὸν — λόγῳ τῆς σχέσεως $z_i = k + \lambda y_i$ — ἴση πρὸς $\lambda^2\sigma^2$. Οὕτως, ὁ δείκτης R^2 χρησιμοποιουμένων τῶν τιμῶν z_i, $i = 1,2,\ldots,N$ εἶναι

$$R^2 = 1 - \frac{\lambda^2\sigma^2}{\lambda^2\sigma_y{}^2} = 1 - \frac{\sigma^2}{\sigma_y{}^2}$$

ἤτοι ἴσος πρὸς τὸν δείκτην R^2 ὑπολογιζομένου τῇ βοηθείᾳ τῶν τιμῶν y_i, $i = 1, 2,\ldots,N$ καὶ κατὰ συνέπειαν ὁ ἐν λόγῳ δείκτης εἶναι ἀνεξάρτητος τῶν χρησιμοποιουμένων κατὰ περίπτωσιν μονάδων μετρήσεως, ὡς ἐπίσης τοῦ σημείου τὸ ὁποῖον λαμβάνεται ὡς ἀρχὴ τῶν μετρήσεων.

Παραθέτομεν τώρα μερικὰς — ἐκ τῶν συνήθως χρησιμοποιουμένων εἰς τὴν πρᾶξιν — ἐξισώσεις ἢ ἄλλως οἰκογενείας καμπύλων αἱ ὁποῖαι δ ὲ ν π λ η ρ ο ῦ ν τὴν ὡς ἄνω συνθήκην — ἤτοι ἐὰν περιλαμβάνουν ὡς μέλος των μίαν καμπύλην $y = f(x,\alpha',\beta')$ νὰ περιλαμβάνουν καὶ τὴν $y = k + \lambda f(x,\alpha',\beta')$ — μὲ ἄμεσον συνέπειαν ἡ τιμὴ τοῦ δείκτου R^2 — ἡ ὁποία ἀντιστοιχεῖ εἰς τὴν ἐπιλεγομένην μεταξὺ αὐτῶν καμπύλην παλινδρομήσεως — νὰ ποικίλλῃ, ὡς θὰ ἴδωμεν, καὶ εἰς ἀντίστοιχα ἀριθμητικὰ παραδείγματα, ἀναλόγως τῶν χρησιμοποιουμένων μονάδων καὶ τῆς ἀρχῆς τῶν μετρήσεων. Κατ' ἀρχὴν ἡ ὡς ἄνω συνθήκη δὲν πληροῦται ὑπὸ τῆς οἰκογενείας τῶν εὐθειῶν

$$y = \beta x$$

καὶ γενικώτερον ὑπὸ τῶν οἰκογενειῶν τῶν καμπύλων πολυωνυμικῆς μορφῆς

$$y = \beta x + \gamma x^2, \quad y = \beta x + \gamma x^2 + \delta x^3, \quad \text{κ.ο.κ.}$$

καθ' ὅσον αἱ ἐν λόγῳ οἰκογένειαι δὲν περιλαμβάνουν ὡς μέλος των τὰς εὐθείας $y=\alpha+\beta x$ καὶ τὰς καμπύλας $y = \alpha + \beta x + \gamma x^2$, $y^2 = \alpha + \beta x + \gamma x^2 + \delta x^3$ κ.ο.κ.

Ὁμοίως, ἡ ἐν λόγῳ συνθήκη δὲν πληροῦται ὑπὸ τῆς οἰκογενείας τῶν καμπύλων ἐκθετικῆς καὶ γεωμετρικῆς μορφῆς, ἤτοι ὑπὸ τῶν ἐξισώσεων $y=\alpha\beta^x$ καὶ $y=\alpha x^\beta$ ἢ ἄλλως $\log y=A+Bx$ καὶ $\log y = A+\beta \log x$ ὅπου $A=\log a$ καὶ $B=\log\beta$, καθ' ὅσον αἱ ἐν λόγῳ οἰκογένειαι δὲν περιλαμβάνουν μεταξὺ τῶν μελῶν των τὰς καμπύλας $y=k+\alpha\beta^x$ (γενικευμένη ἐκθετικὴ καμπύλη) καὶ $y=k+\alpha x^\beta$ (γενικευμένη γεωμετρικὴ καμπύλη). Πέραν τῶν ἀνωτέρω, τὸ αὐτὸ πρόβλημα παρουσιάζεται μὲ τὴν καλουμένην λογιστικὴν καμπύλην

$$y = \frac{c}{1 + e^{\alpha+\beta x}}$$

καθ' ὅσον ἡ ἐν λόγῳ οἰκογένεια δὲν περιλαμβάνει τὴν καμπύλην

$$y = k + \frac{c}{1 + e^{\alpha+\beta x}}$$

ὡς ἐπίσης μὲ ὡρισμένας ἄλλας οἰκογενείας καμπύλων ὡς π.χ. αἱ

$$y = \beta \log x \qquad y = \beta \log x + \gamma (\log x)^2$$

$$\sqrt{y} = \beta x \qquad \sqrt{y} = \beta x + \gamma x^2$$

$$y = \frac{\beta}{x} \qquad y = \frac{\beta}{y+\delta x} \qquad \text{κοκ.}$$

Πρὸς πληρεστέραν κατανόησιν τῶν ἀνωτέρω παραθέτομεν ἐνδεικτικῶς δύο ἁπλᾶ ἀριθμητικὰ παραδείγματα :

(A) Χρησιμοποιοῦντες ἀντὶ τῶν τιμῶν τῆς μεταβλητῆς Y αἱ ὁποῖαι παρατίθενται εἰς τὸν ἀνωτέρω πίνακα (1) τὰς τιμὰς $z_i = y_i + I$ λαμβάνομεν τὰ ἑξῆς :

$\sum\limits_i x_i^2 = 30$, $\sum\limits_i x_i z_i = 85$, $\beta = \dfrac{\Sigma x_i z_i}{\Sigma x_i^2} = \dfrac{85}{30} = 2{,}833$ καὶ μ.τ.σ. περὶ τὴν εὐθεῖαν $\hat{z} = 2{,}833\ x$ τὴν ποσότητα

$$1/4\,[\,(10 - 2{,}833)^2 + (9 - 5{,}666)^2 + (11 - 8{,}499)^2 + (6 - 11{,}332)^2\,] = 24{,}5$$

Ἐξ ἄλλου ἡ διακύμανσις τῶν τιμῶν $z_i = y_i+1$ εἶναι ὡς γνωστὸν ἡ αὐτὴ μὲ τὴν ὑπολογισθεῖσαν διακύμανσιν τῶν τιμῶν y_i, ἤτοι 3,5. Οὕτως ἡ τιμὴ τοῦ δείκτου R^2 γίνεται ἐν προκειμένῳ

$$R^2 = 1 - \frac{24,5}{3,5} = - \frac{21}{3,5} = -6$$

διάφορος τῆς ὑπολογισθείσης προηγουμένης τιμῆς -4,89

(B) Ὑποθέσωμεν ὅτι τὰ ἐμπειρικὰ δεδομένα συνίστανται ἐκ τῶν ἀριθμητικῶν ζευγῶν (x_i,y_i), $i=1,2,\ldots,5$ τῶν δύο πρώτων στηλῶν τοῦ κατωτέρω πίνακος 2 καὶ ἀκόμη ὅτι ὁ μελετητὴς ἀπεφάσισε νὰ προσαρμόσῃ πρὸς αὐτὰ μίαν ἐκθετικὴν καμπύλην τῆς μορφῆς $y=\alpha\beta^x$ ἢ ἄλλως $\log y = A + Bx$ ὅπου $A=\log\alpha$ καὶ $B=\log\beta$.

Πίναξ 2

x_i	y_i	$y_i' = \log y_i$	x_i^2	$x_i y_i'$	$y_i'^2$
1	11.800	4,07188	1	4,07188	16,58118
2	19.200	4,28330	4	8,56660	18,34666
3	43.600	4,63849	9	13,91547	21,51104
4	92.600	4,96661	16	19,86644	24,66712
5	148.400	5,17143	25	25,85715	26,74338
15	—	23,13171	55	72,27754	107,84938

Τὸ σύστημα τῶν κανονικῶν ἐξισώσεων τὸ ὁποῖον ἐν προκειμένῳ λαμβάνει τὴν μορφὴν

$$\Sigma y_i' = AN + B\Sigma x_i$$

$$\Sigma x_i y_i' = A\Sigma x_i + B\Sigma x_i^2$$

εἰς τὴν συγκεκριμένην περίπτωσιν γίνεται

$$23,13171 = 5A + 15B$$

$$72,27754 = 15A + 55B$$

ἡ ἐπίλυσις δὲ αὐτοῦ δίδει

$$\hat{A} = 3,76162 \quad \text{καὶ} \quad \hat{B} = 0,28824$$

ἐκ τῶν ὁποίων δι' ἀντιλογαριθμήσεως λαμβάνομεν

$$\hat{\alpha} = 5.776 \quad \text{καὶ} \quad \hat{\beta} = 1,94$$

Οὕτως, ἡ ζητουμένη καμπύλη παλινδρομήσεως ἔχει ἐξίσωσιν

$$\hat{y} = 5.776 . 1,94^x \quad \text{ἢ} \quad \log\hat{y} = 3,762 + 0,288_x .$$

Τὸ μ.τ.σ. περὶ τὴν ὡς ἄνω καμπύλην ὑπολογιζόμενον ὡς γνωστὸν ἐκ τοῦ τύπου

$$\sigma^2 = \frac{1}{N} (\Sigma y_i'^2 - \hat{A}\, \Sigma y_i' - \hat{B}\Sigma x_i y_i')$$

εἶναι $\sigma^2 = 0,0522$.

Ἐξ ἄλλου ἡ διακύμανσις τῶν τιμῶν y_i', $i=1,2,\ldots,N$ (ὅπου $y_i' = \log y)_i$ εἶναι

$$\sigma^2_{y'} = 0,1701$$

καὶ κατὰ συνέπειαν ὁ δείκτης προσδιορισμοῦ R^2 ὁριζόμενος ἐκ τοῦ τύπου (4) εἶναι

$$R^2 = 1 - \frac{0,0522}{0,1701} = 0,69 \text{ (ἢ } 6/\%)$$

Ὑποθέσωμεν τώρα ὅτι ἐπὶ τῆς μεταβλητῆς Y ἐφαρμόζεται ὁ γραμμικὸς μετασχηματισμὸς

$$Z = \frac{Y - 10.000}{100}$$

ἐν ἄλλοις δηλαδὴ λόγοις ὅτι ὡς ἀρχὴ τῶν μετρήσεων λαμβάνεται τὸ 10.000 καὶ ὡς μονὰς τὸ 100. Ἐφαρμοζομένης τῆς ὡς ἄνω διαδικασίας ἐπὶ τῶν νέων δεδομένων — τὰ ὁποῖα παρατίθενται κατωτέρω εἰς τὸν πίνακα 2α — λαμβάνονται ἀντιστοίχως τὰ ἑξῆς ἀποτελέσματα :

Π ί ν α ξ 2α

x_i	z_i	$z_i = \log z_i$	x_i^2	$x_i z_i'$	$z_i'^2$
1	18	1,25527	1	1,25527	1,57503
2	92	1,96379	4	3,92758	3,85729
3	336	2,52634	9	7,57902	6,38068
4	826	2,91698	16	11,66792	8,50888
5	1.384	3,14114	25	15,70570	9,86588
15	—	11,80352	55	40,13549	30,18776

Τὸ σύστημα τῶν κανονικῶν ἐξισώσεων γίνεται ἐν προκειμένῳ

$$11{,}80352 = 5A + 15B$$

$$40{,}13549 = 15A + 55B$$

ἐξ αὐτοῦ δὲ προκύπτει ὅτι

$$\hat{A} = 0{,}94323 \qquad \hat{B} = 0{,}47249$$

ἀντιστοίχως δὲ

$$\hat{\alpha} = 8{,}78 \qquad \hat{\beta} = 2{,}96.$$

Οὕτως, ἡ ἐξίσωσις τῆς καμπύλης παλινδρομήσεως γίνεται ἐν προκειμένῳ

$$\hat{y} = 8{,}78 \cdot 2{,}96^x \quad \text{ἢ} \quad \log\hat{y} = 0{,}943 + 0{,}472x$$

τὸ δὲ μ.τ.σ. περὶ αὐτὴν γίνεται ἀντιστοίχως $\sigma^2 = 0{,}1129$.

Ἐξ ἄλλου ἡ διακύμανσις τῶν $z_i' = \log z_i$ εἶναι $\sigma^2_{z'} = 0{,}4646$ καὶ κατὰ συνέπειαν ἡ τιμὴ τοῦ δείκτου R^2 εἶναι ἐν προκειμένῳ

$$R^2 = 1 - \frac{0{,}1129}{0{,}4646} = 0{,}76 \ (\text{ἢ } 76\%)$$

ἤτοι σημαντικῶς διάφορος τῆς εὑρεθείσης προηγουμένως 0,69.

III. Ὁ συντελεστὴς γραμμικῆς συσχετίσεως καὶ αἱ ἐφαρμογαί του

Τὸ τετράγωνον τοῦ συντελεστοῦ — γραμμικῆς — συσχετίσεως δύο μεταβλητῶν X καὶ Y, ὁριζομένου ἐκ τῆς σχέσεως

$$\rho = \frac{Cov(X,Y,)}{\sigma_x \, \sigma_y} \qquad (11)$$

ἰσοῦται, ὡς γνωστόν, πρὸς τὸν δείκτην προσδιορισμοῦ R^2 ὁ ὁποῖος ἀναφέρεται ἀντιστοίχως εἰς τὴν εὐθεῖαν παλινδρομήσεως

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \qquad (12)$$

ἐπιλεγομένην διὰ τῆς μεθόδου τῶν ἐλαχίστων τετραγώνων ἐκ τῆς διπαραμετρικῆς * οἰκογενείας

$$y = \alpha + \beta x \qquad (13)$$

---

* καὶ ὄχι ἐκ τῆς $y = \beta \chi$.

Οὕτω, δοθέντος ὅτι ἡ ἐξίσωσις (13) ἱκανοποιεῖ τὰς τεθείσας ὡς ἄνω συνθήκας, συνάγεται τὸ συμπέρασμα ὅτι ὁ συντελεστὴς συσχετίσεως ρ εἶναι καθαρὸς ἀριθμὸς ὁ ὁποῖος πληροῖ πάντοτε τὴν διπλῆν ἀνισότητα $-1 \leqslant \rho \leqslant +1$, ἐνῶ παραλλήλως εἶναι ἀνεξάρτητος τῶν χρησιμοποιουμένων μονάδων μετρήσεως, ὡς καὶ τῆς λαμβανομένης ἑκάστοτε ὡς ἀρχῆς αὐτῶν (ἀναλλοίωτος εἰς γραμμικοὺς μετασχηματισμοὺς τῶν μεταβλητῶν).

Πολλάκις ὅμως εἰς τὴν πρᾶξιν ὁ ἐν λόγῳ συντελεστὴς χρησιμοποιεῖται καὶ ὡς μέτρον τῆς συναφείας δύο μὴ ποσοτικῶν ἀλλὰ ποιοτικῶν μεταβλητῶν — αἱ ὁποῖαι ὡς γνωστὸν ἐπιτρέπουν μόνον ἱεράρχησιν τῶν μονάδων τοῦ ὑπὸ ἔρευναν πληθυσμοῦ — λαμβανομένων ὡς τιμῶν τῶν μεταβλητῶν τῶν ἀντιστοίχων βαθμῶν ἢ ἄλλως κωδικαρίθμων.

Ἐν προκειμένῳ θὰ πρέπει νὰ τονισθῇ ὅτι ἡ τιμὴ τοῦ ἐν λόγῳ δείκτου εἶναι δυνατὸν νὰ ποικίλλῃ ἀναλόγως τοῦ χρησιμοποιουμένου συστήματος ἱεραρχήσεως ἢ ἄλλως κωδικογραφήσεως, ἐκτὸς ἐὰν οἱ χρησιμοποιούμενοι εἰς διαφόρους περιπτώσεις πρὸς ἱεράρχησιν τῶν μονάδων τοῦ πληθυσμοῦ βαθμοὶ (κωδικάριθμοι) εὑρίσκονται ἐν γραμμικῇ ἐξαρτήσει.

Πρὸς πληρεστέραν κατανόησιν τῶν ἀνωτέρω παραθέτομεν ἓν ἁπλοῦν ἀριθμητικὸν παράδειγμα. Ὑποθέσωμεν ὅτι 5 τροχαῖα ἀτυχήματα ἱεραρχοῦνται ἐξ ἀπόψεως σοβαρότητος αὐτῶν καὶ ἐμπειρίας τοῦ ὁδηγοῦ προκειμένου νὰ διαπιστωθῇ ἡ τυχὸν ὑφισταμένη μεταξὺ ἐμπειρίας τοῦ ὁδηγοῦ — μεταβλητὴ X — καὶ σοβαρότητος τοῦ ἀτυχήματος — μεταβλητὴ Y — σχέσις. Εἰς τοὺς κατωτέρω πίνακας 3, 3α καὶ 3β παρατίθενται τρεῖς διάφοροι τρόποι κωδικογραφήσεως (βαθμολογίας)* τῶν μεταβλητῶν καὶ ὑπολογίζονται οἱ ἀντίστοιχοι συντελεσταὶ συσχετίσεως ἐφαρμοζομένου τοῦ γνωστοῦ τύπου.

$$\rho = \frac{\Sigma x_i y_i - \dfrac{\Sigma x_i \Sigma y_i}{N}}{\sqrt{\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{N}}\sqrt{\Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{N}}} \tag{14}$$

Πίναξ 3

x_i	y_i	x_i^2	$x_i y_i$	y_i
1	5	1	5	25
2	3	4	6	9
3	4	9	12	16
4	1	16	4	1
5	2	25	10	4
15	15	55	37	55

* Ὅσον μεγαλυτέρα ἡ ἐμπειρία καὶ ἡ σοβαρότης τοῦ ἀτυχήματος, τόσον μεγαλύτερος ὁ ἀντίστοιχος βαθμὸς εἰς ἁπάσας τὰς περιπτώσεις.

Π ί ν α ξ 3α

x_i'	y_i'	$x_i'^2$	$x_i'y_i'$	$y_i'^2$
3	8	9	24	64
5	6	25	30	36
7	7	49	49	49
9	4	81	36	16
11	5	121	55	25
35	30	285	194	190

Π ί ν α ξ 3β

x_i''	y_i''	$x_i''^2$	$x_i''y_i''$	$y_i''^2$
2	8	4	16	64
5	5	25	25	25
7	7	49	49	49
8	1	64	8	1
10	3	100	30	9
32	24	242	128	148

Ἐκ τῶν δεδομένων τοῦ πίνακος (3) δι' ἐφαρμογῆς τοῦ τύπου (14) λαμβάνομεν

$$\rho = -0{,}8$$

Ἐκ τοῦ πίνακος (3α) ἔχομεν ὁμοίως

$$\rho = -0{,}8$$

ἐνῶ ἐκ τοῦ πίνακος (3β) λαμβάνομεν

$$\rho = -0{,}73$$

Ἡ ἰσότης τῶν συντελεστῶν συσχετίσεως οἱ ὁποῖοι προκύπτουν ἐκ τῶν δύο πρώτων πινάκων ὀφείλεται εἰς τὸ ὅτι οἱ βαθμοὶ (κώδικες) τοῦ πίνακος (3α) εὑρίσκονται ἐξ ἐκείνων τοῦ πίνακος (3) ἐκ τοῦ γραμμικῶν σχέσεων

$$x' = 2x+1 \quad \text{καὶ} \quad y' = y+3$$

Ἀντιθέτως ἡ διαφορὰ τοῦ ἐκ τοῦ πίνακος (3β) ἀποτελέσματος ὀφείλεται εἰς τὸ ὅτι οἱ βαθμοὶ x'' καὶ y'' — παρ' ὅτι διατηροῦν τὴν αὐτὴν ἱεράρχησιν—δὲν ἀποτελοῦν γραμμικοὺς μετασχηματισμοὺς τῶν x καὶ y ἢ χ' καὶ ψ'.

Διὰ τὸν λόγον αὐτὸν ἡ χρῆσις τοῦ συντελεστοῦ ρ θὰ πρέπει ἐν προκειμένῳ νὰ ἀποφεύγηται διότι τὸ σχετικὸν ἀποτέλεσμα εἶναι κατὰ βάσιν ἄνευ σημασίας.