5

Η ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ (PEARSON's *r*)

Περίληψη

Σκοπός του κεφαλαίου είναι η εφαρμογή της ανάλυσης συσχέτισης (Pearson *r*) μέσω του PASW. Η ανάλυση συσχέτισης Pearson *r* χρησιμοποιείται για να εξεταστεί η σχέση μεταξύ δύο ή περισσότερων συνεχών (ή ποσοτικών) μεταβλητών.

Εισαγωγή

Παράδειγμα 1. Προπονητής κλασσικού αθλητισμού που μόλις έχει διδάξει το άλμα σε ύψος στους νέους αθλητές του θέλει να εξετάσει εάν υπάρχει συσχέτιση των επιδόσεών τους στο άλμα σε ύψος με τις επιδόσεις τους στο άλμα σε μήκος. Εάν υπάρχει, θέλει επίσης να βρει πόσο μεγάλες είναι αυτές οι σχέσεις.

Παράδειγμα 2. Διευθυντής γυμναστηρίου έχει δώσει ερωτηματολόγιο σε ασκούμενους στο οποίο σε κλίμακες 0-100 έχει μετρήσει τη γενική ικανοποίησή τους από τη συμμετοχή τους στο γυμναστήριό του καθώς και την ειδική ικανοποίησή τους από διάφορες υπηρεσίες (π.χ. εξοπλισμό και κόστος συμμετοχής). Θέλει να βρει ποιες από τις μεταβλητές ειδικής ικανοποίησης συσχετίζονται ισχυρότερα ή ασθενέστερα με τη γενική ικανοποίησή τους ώστε να δώσει την ανάλογη προσοχή στις υπηρεσίες, ή εξοπλισμό ή κόστος που σχετίζονται περισσότερο με τη γενική ικανοποίησή τους. Κάποιος ερευνητής του προτείνει να χρησιμοποιήσει αναλύσεις συσχέτισης Pearson-Product moment και να εξετάσει το μέγεθος της συσχέτισης κάθε μεταβλητής «ειδικής ικανοποίησης» με τη «γενική ικανοποίηση».

Σκοπός

Χρησιμοποιείται για να βρεθεί η σχέση μεταξύ δύο ή περισσότερων συνεχών (ή ποσοτικών) μεταβλητών.

Αρχικά εξετάζεται αν η σχέση είναι «στατιστικώς σημαντική», δηλαδή ότι οι συσχετίσεις μεταξύ των μεταβλητών δεν βρέθηκαν τυχαία στο δείγμα αλλά αντανακλούν συσχετίσεις μεταξύ των μεταβλητών σε επίπεδο πληθυσμού.

Στη συνέχεια εξετάζεται αν η συσχέτιση είναι θετική (όσο αυξάνονται οι τιμές στη μία μεταβλητή αυξάνονται και στην άλλη) ή αρνητική (όσο αυξάνονται οι τιμές στη μία μεταβλητή μειώνονται στην άλλη).

Κατόπιν παρατηρείται το μέγεθος της σχέσης. Αν η τάση για αύξηση των τιμών στη μια μεταβλητή συνοδεύεται από τάση για αύξηση των τιμών, η οποία εμφανίζεται μάλλον ομοιόμορφα σε όλο το δείγμα (μεγάλη συσχέτιση), ή αν η τάση για αύξηση των τιμών στη μια μεταβλητή εμφανίζει μικρό βαθμό ομοιομορφίας (μικρή συσχέτιση), ή αν η αύξηση των τιμών στη μία μεταβλητή δεν συνδέεται με αντίστοιχη αύξηση των τιμών στην άλλη μεταβλητή (μηδενική ή έλλειψη συσχέτισης).

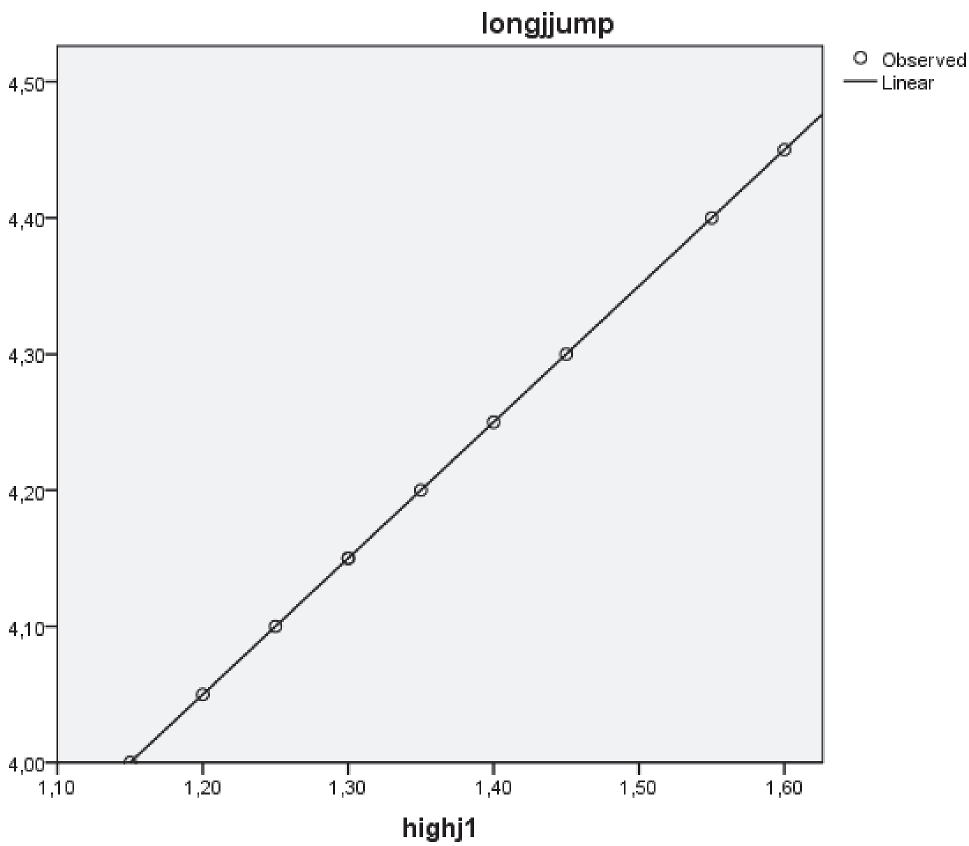
Στον Πίνακα 1 και στα Γραφήματα 1-4 δίνονται παραδείγματα με διαφορετικό μέγεθος συσχέτισης μιας μεταβλητής (άλμα σε μήκος: «longjump») με μια δεύτερη μεταβλητή (άλμα σε ύψος: «highj») όταν η δεύτερη παίρνει διαφορετικές τιμές. Στην συσχέτιση της longjump με την highj1 υπάρχει τέλεια σχέση, όσο ακριβώς αυξάνονται οι τιμές στη μία μεταβλητή τόσο ακριβώς αυξάνονται και στην δεύτερη μεταβλητή. Η τέλεια σχέση απεικονίζεται στο γράφημα με ευθεία γραμμή. Στη συσχέτιση με την highj2 υπάρχει υψηλός βαθμός συσχέτισης. Στο γράφημα δημιουργείται ένα σύννεφο τιμών που βρίσκονται κοντά στη νοητή ευθεία της τέλειας συσχέτισης. Στη συσχέτιση με την highj3 υπάρχει μέτριος βαθμός συσχέτισης. Στη συσχέτιση με την highj4 η συσχέτιση είναι μηδενική.

Είδος μεταβλητών: και οι δύο μεταβλητές είναι συνεχούς τύπου (ή ποσοτικές).

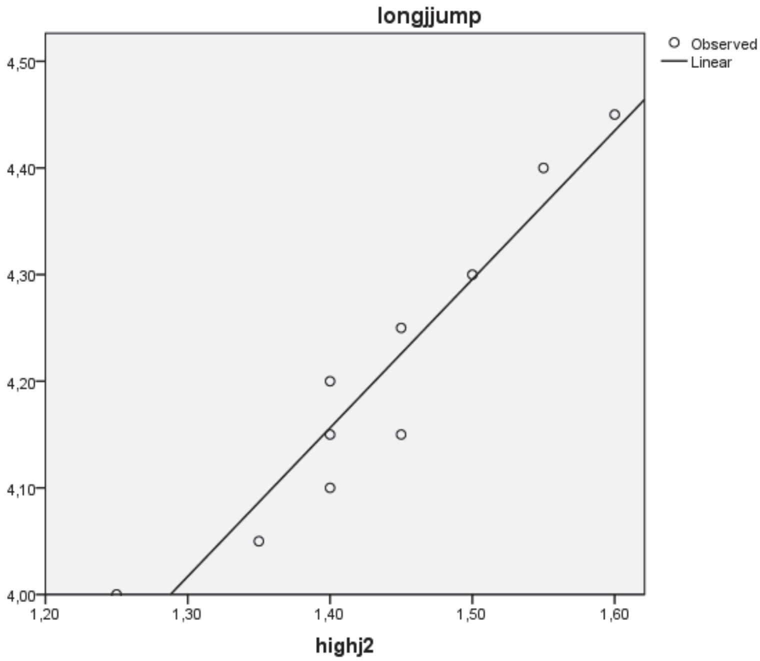
Πίνακας 1. Παραδείγματα τιμών σε μία μεταβλητή άλμα σε μήκος και σε τρεις μεταβλητές άλμα σε ύψος

Longjump	Highj1	Highj2	Highj3	Highj4
4,00	1,15	1,25	1,30	1,40
4,05	1,20	1,35	1,30	1,35
4,10	1,25	1,40	1,35	1,45
4,15	1,30	1,45	1,35	1,40
4,15	1,30	1,40	1,45	1,30
4,20	1,35	1,40	1,40	1,45
4,25	1,40	1,45	1,35	1,35
4,30	1,45	1,50	1,40	1,45
4,40	1,55	1,55	1,45	1,35
4,45	1,60	1,60	1,40	1,40

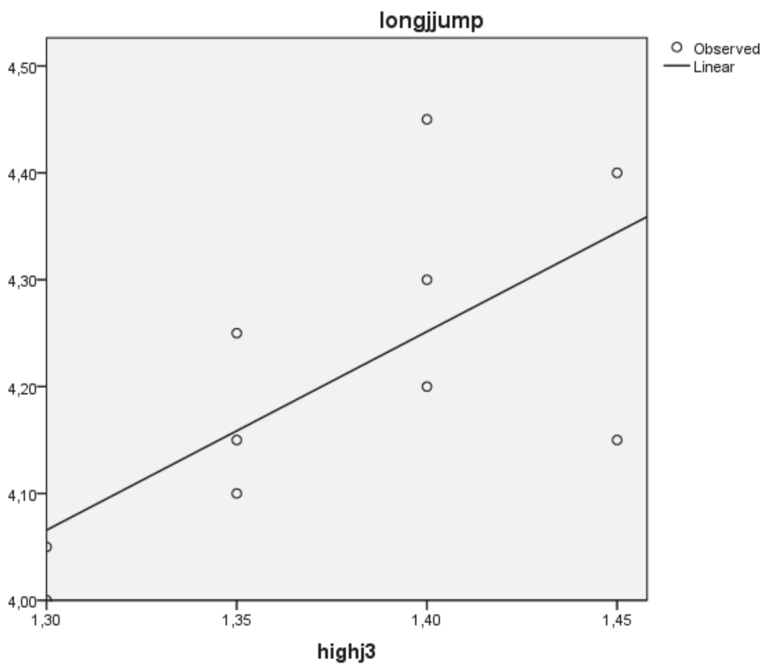
Γράφημα 1. Τέλεια θετική συσχέτιση ($r = 1$)



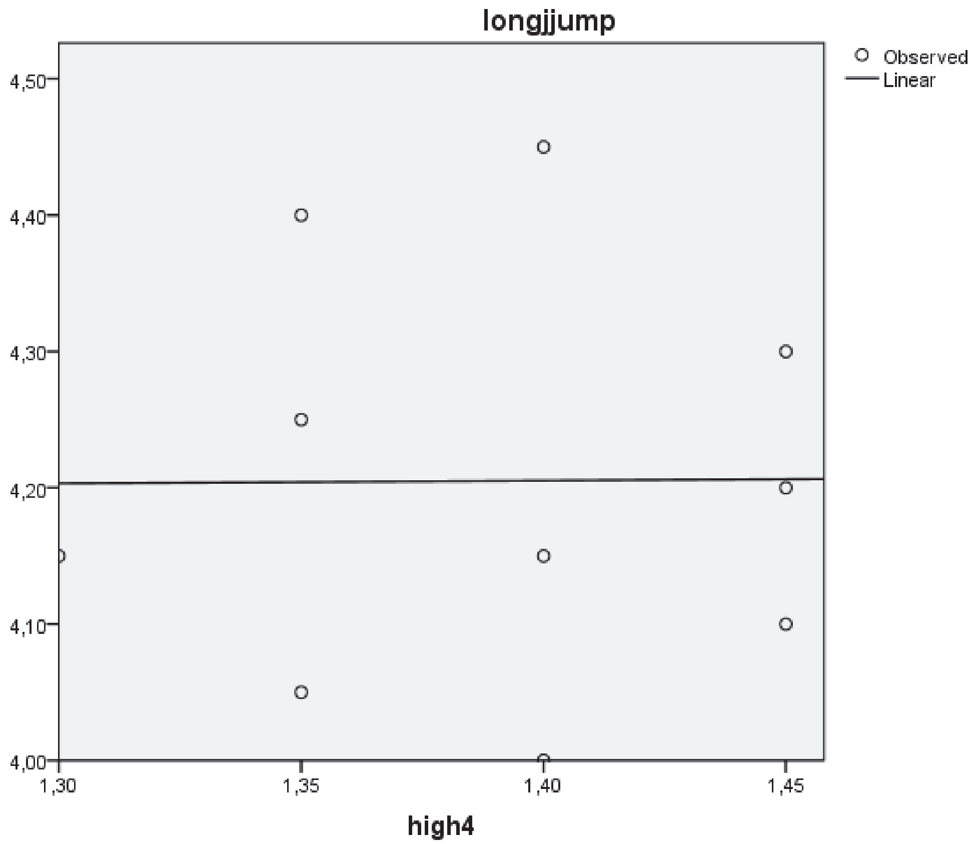
Γράφημα 2. Υψηλή θετική συσχέτιση (r = .95, p < .001)



Γράφημα 3. Θετική συσχέτιση μετρίου μεγέθους (r = .68, p < .05)



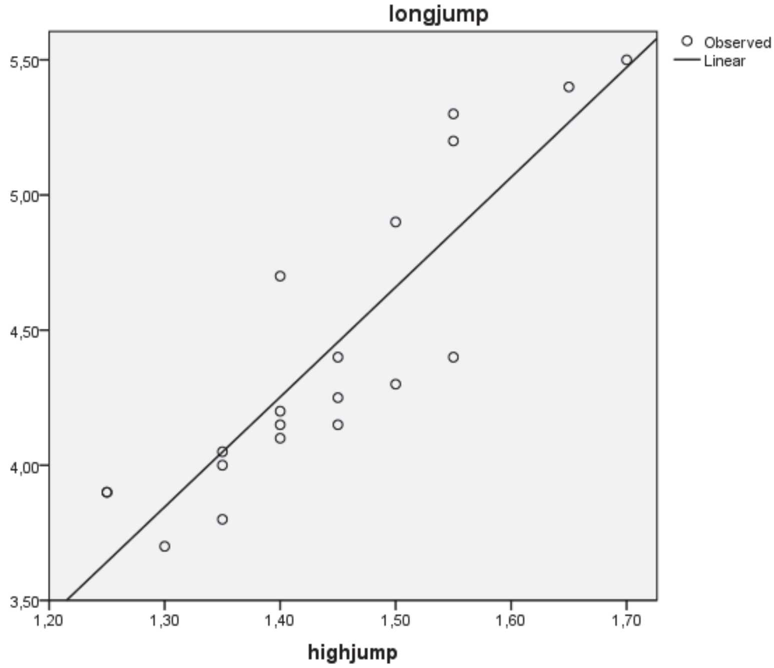
Γράφημα 4. Μηδενική συσχέτιση ($r = 0$)



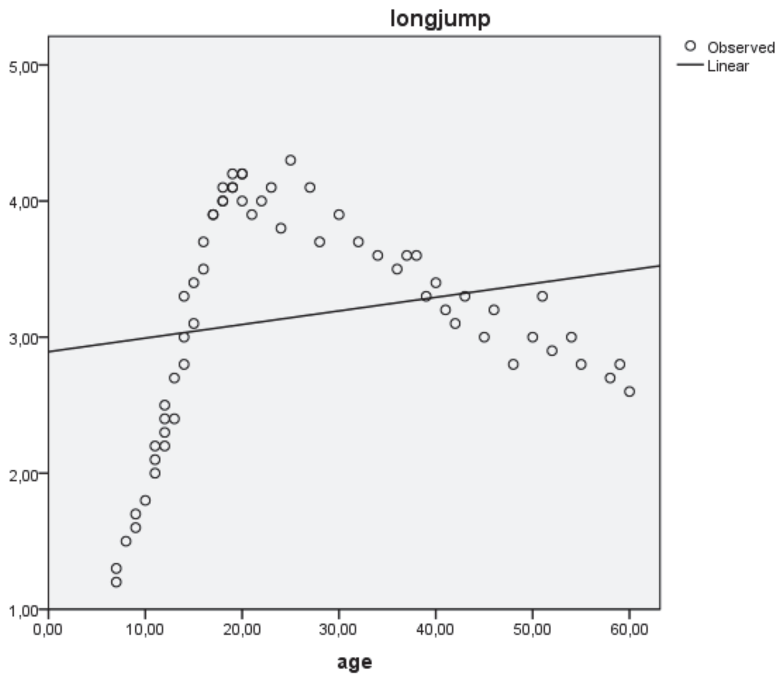
Προϋποθέσεις

- Θα πρέπει οι κατανομές και των δύο μεταβλητών να ακολουθούν την κανονική κατανομή. Σε δείγματα με μεγάλο αριθμό ατόμων ή παρατηρήσεων αυτό συνήθως δεν είναι πρόβλημα.
- Θα πρέπει η σχέση των μεταβλητών να είναι γραμμική. Παράδειγμα γραμμικής σχέσης μεταξύ δύο μεταβλητών (άλμα σε μήκος και άλμα σε ύψος, τα δεδομένα δεν είναι πραγματικά) απεικονίζεται στο Γράφημα 5. Παράδειγμα μη γραμμικής σχέσης μεταξύ δύο μεταβλητών (άλμα σε μήκος και ηλικία, τα δεδομένα δεν είναι πραγματικά) απεικονίζεται στο Γράφημα 6.

Γράφημα 5. Παράδειγμα γραμμικής θετικής συσχέτισης μεταξύ δύο μεταβλητών



Γράφημα 6. Παράδειγμα μη γραμμικής θετικής συσχέτισης μεταξύ δύο μεταβλητών



Προσοχή: Όταν υπάρχει σχέση μεταξύ δυο μεταβλητών αυτό δεν πρέπει να λαμβάνεται ως απόδειξη ύπαρξης *αιτιώδους σχέσης – causal relationship*. Για παράδειγμα αν η τιμή μιας μεταβλητής αυξάνεται ή μειώνεται όσο και η τιμή της άλλης, αυτό δεν σημαίνει ότι η μια μεταβλητή μπορεί να προβλέψει την άλλη. Η σχέση έχει να κάνει με τη μεταβλητότητα (variability) δυο μεταβλητών, δηλαδή αν κάθε μεταβολή των τιμών της μιας μεταβλητής συνοδεύεται από συστηματική μεταβολή στις τιμές της άλλης, αλλά αυτό δεν είναι απαραίτητα αιτιώδης σχέση. Για παράδειγμα, μπορεί να υπάρχει κάποια άλλη τρίτη μεταβλητή που επηρεάζει και τις δύο μεταβλητές που συσχετίζονται μεταξύ τους, όπως η ηλικία στην διάρκεια της ανθρώπινης ανάπτυξης η οποία επηρεάζει το ύψος και την ταχύτητα επίλυσης του ίδιου μαθηματικού προβλήματος. Αν εξετάσουμε τη σχέση του ύψους και της ταχύτητας επίλυσης του ίδιου μαθηματικού προβλήματος έχοντας δείγμα 180 μαθητές, 30 μαθητές από κάθε τάξη από την Δ΄ Δημοτικού ως την Γ΄ Γυμνασίου, το πιο πιθανό είναι να βρούμε στατιστικά σημαντική συσχέτιση μεταξύ ύψους και ταχύτητας επίλυσης του μαθηματικού προβλήματος. Ωστόσο αυτό δε σημαίνει ότι το ύψος επηρεάζει την ταχύτητα επίλυσης του μαθηματικού προβλήματος ή το αντίστροφο, αφού η μεταβλητή ηλικία επηρεάζει και το ύψος και την ταχύτητα επίλυσης του προβλήματος.

Σημείωση. Ανάλογα με το είδος των μεταβλητών έχουμε και τον ανάλογο δείκτη συσχέτισης. Εμείς θα χρησιμοποιήσουμε τον **Pearson *r* ή Correlation Coefficient *r*.**

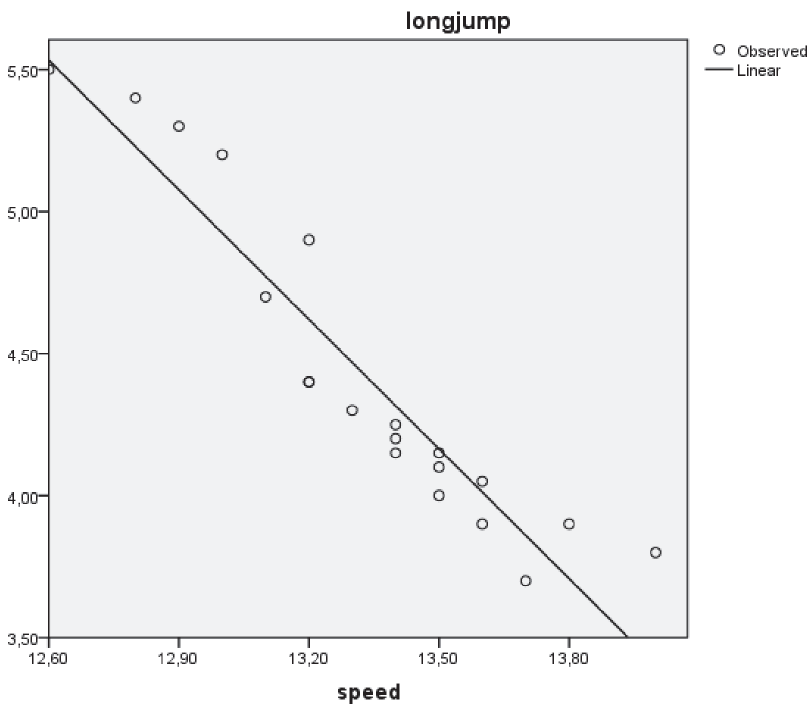
Σημεία Κλειδιά

- **Το πρόσημο και μέγεθος του συντελεστή συσχέτισης (*r*)**. Ο συντελεστής συσχέτισης (*r*) αποτελείται από: (α) Ένα **θετικό** ή **αρνητικό** πρόσημο (συνήθως το θετικό παραλείπεται). Θετικό πρόσημο σημαίνει ότι όσο αυξάνονται οι τιμές στη μία μεταβλητή αυξάνονται οι τιμές και στη δεύτερη μεταβλητή (παράδειγμα Γράφημα 5). Αρνητικό πρόσημο σημαίνει ότι όσο αυξάνονται οι τιμές στη μία μεταβλητή μειώνονται οι τιμές στη δεύτερη μεταβλητή (παράδειγμα Γράφημα 7). (β) Την αριθμητική τιμή που κυμαίνεται από **–1,00 έως 1,00.** Το 0 δείχνει πλήρη έλλειψη συσχέτισης, το +1 δηλώνει την τέλεια θετική σχέση και το –1 δηλώνει την τέλεια αρνητική σχέση. Συσχετίσεις μεταξύ –0,40 έως 0,40 θεωρούνται μικρές. Συσχετίσεις μεταξύ 0,40 έως 0,70 (ή –0,40 έως

–0,70) θεωρούνται μέτριες. Συσχετίσεις μεγαλύτερες του 0,70 (ή μεταξύ –0,70 και –1) θεωρούνται μεγάλες.

- **Το επίπεδο εμπιστοσύνης – λάθος τύπου Ι (α).** Δηλώνει το επίπεδο εμπιστοσύνης στο οποίο θεωρείται ότι η συσχέτιση που προέκυψε στο δείγμα αντανακλά συσχέτιση που υπάρχει μεταξύ δύο μεταβλητών στον πληθυσμό από τον οποίο επιλέχθηκε το δείγμα. Όπως αναφέρθηκε στην ενότητα «έλεγχος υποθέσεων» σε προηγούμενο κεφάλαιο, το επίπεδο εμπιστοσύνης προσδιορίζει και το λάθος τύπου Ι (α). Έτσι, αν το επίπεδο εμπιστοσύνης είναι 95% η πιθανότητα για λάθος τύπου Ι (α) είναι μικρότερη από 5% ή $p < 0{,}05$. Συνήθως οι συντελεστές συσχέτισης που συναντώνται στη βιβλιογραφία συνοδεύονται με την πιθανότητα λάθους τύπου Ι (α) για $p < 0{,}05$ ή/και $p < 0{,}01$, ή/και $p < 0{,}001$.
- Ο συντελεστής συσχέτισης υψωμένος στο τετράγωνο δηλώνει το ποσοστό διακύμανσης μεταξύ των δύο μεταβλητών. Αν για παράδειγμα δύο μεταβλητές συσχετίζονται μεταξύ τους με $r = 0{,}70$ τότε 49% της διακύμανσης της μίας μεταβλητής είναι ταυτόχρονα διακύμανση και της δεύτερης μεταβλητής.

Γράφημα 7. Παράδειγμα αρνητικής συσχέτισης μεταξύ δύο μεταβλητών



Παράδειγμα στη Φυσική Αγωγή

Υπάρχει σχέση μεταξύ της απόδοσης στο άλμα σε μήκος και της ταχύτητας που θα αναπτύξει ο αθλητής για να εκτελέσει το άλμα; (τα παρακάτω δεδομένα είναι υποθετικά).

Υπολογισμός του βαθμού σχέσης μεταξύ (**Product Moment correlation *r*** ή ***Pearson's r***)

Άτομο	**Άλμα σε μήκος X**	**Ταχύτητα Ψ**	**χ**	**ψ**	**χ^2**	**ψ^2**	**χψ**
1	4,10	13,5	-0,10	+0,1	0,01	0,01	-0,01
2	4,20	13,4	0	0	0	0	0
3	4,15	13,4	-0,05	0	0,0025	0	0
4	4,30	13,3	+0,10	-0,1	0,01	0,01	-0,01
5	4,00	13,5	-0,20	+0,1	0,04	0,01	-0,02
6	4,25	13,4	+0,05	0	0,0025	0	0
7	4,40	13,2	+0,20	-0,2	0,04	0,04	-0,04
8	4,15	13,5	-0,05	+0,1	0,0025	0,01	-0,005
9	4,05	13,6	-0,15	+0,2	0,0225	0,04	-0,03
10	4,40	13,2	+0,20	-0,2	0,04	0,04	-0,04
	$\Sigma X = 42{,}00$ $MX = 4{,}2$	$\Sigma \Psi = 134$ $M\Psi = 13{,}4$	$\Sigma\chi=0$	$\Sigma\psi=0$	$\Sigma\chi^2 = 0{,}17$	$\Sigma\psi^2 = 0{,}16$	$\Sigma\chi\psi = -0{,}155$

- Βήμα 1. Υπολογίζουμε το μέσο όρο των τιμών της κάθε μεταβλητής. Για το άλμα σε μήκος είναι 4,2 και για την ταχύτητα είναι 13,4.
- Βήμα 2. Για την κάθε τιμή ζητάμε τη διαφορά της από το μέσο όρο των τιμών της μεταβλητής. Για παράδειγμα, για το πρώτο άτομο η διαφορά $\chi = 4{,}10 - 4{,}20 = -0{,}10$ για το άλμα σε μήκος και $\psi = 13{,}5 - 13{,}4 = +0{,}10$ για την ταχύτητα.
- Βήμα 3. Για κάθε τιμή υπολογίζουμε τα τετράγωνα των διαφορών που πήραμε από το προηγούμενο βήμα.

- Βήμα 4. Για κάθε άτομο υπολογίζουμε το γινόμενο των διαφορών χψ που πήραμε από το βήμα 2.
- Βήμα 5. Υπολογίζουμε τα αθροίσματα των τετραγώνων των διαφορών ($\Sigma\chi^2$ και $\Sigma\psi^2$).
- Βήμα 6. Υπολογίζουμε τα αθροίσματα των γινομένων των διαφορών $\Sigma\chi\psi$.
- Βήμα 7. Χρησιμοποιούμε τους παρακάτω τύπους για τον υπολογισμό της συσχέτισης.

$$r = \Sigma\chi\psi \,/\, (\sqrt{(\Sigma\chi^2)\,(\Sigma\psi^2)} = –0{,}155 \,/\, \sqrt{0{,}0272} = –0{,}155 \,/\, 0{,}16492 = –0{,}94$$

Υπολογισμός Pearson Product Moment Correlation

$$r = [N\,\Sigma\chi\psi - (\Sigma\chi)(\Sigma\psi)] \,/\, \sqrt{[N\,\Sigma\chi^2 - (\Sigma\chi)^2]\,[N\,\Sigma\psi^2 - (\Sigma\psi)^2]}$$

$$r = 10\,(-0{,}155) - 0 \,/\, \sqrt{[10\,(0{,}17) - 0]\,[10\,(0{,}16) - 0]}$$

$$r = –1{,}55 \,/\, \sqrt{[1{,}7]\,[1{,}6]} = –1{,}55 \,/\, 1{,}6492 = –0{,}94$$

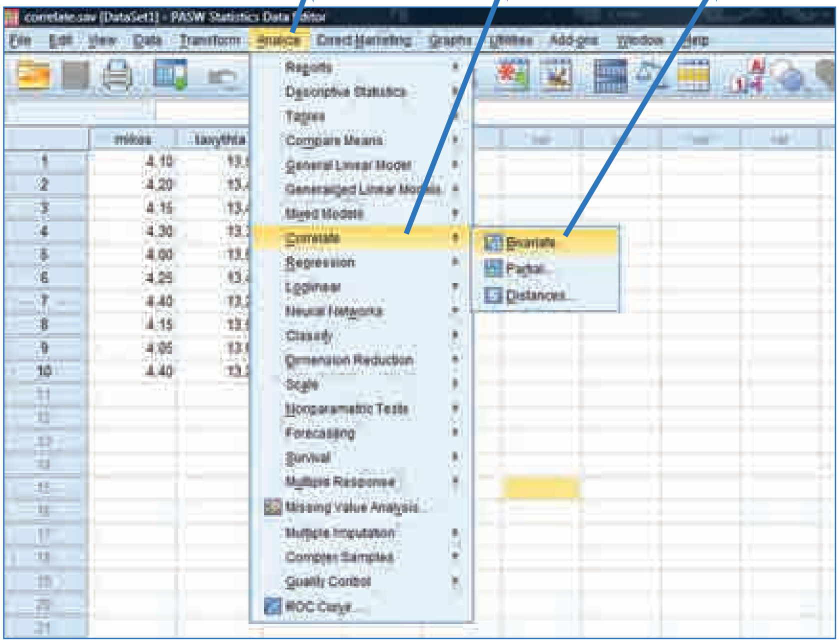
Ονομασία μεταβλητών:
Mhkos : άλμα σε μήκος
Taxythta : ταχύτητα

Υπολογισμός με τη χρήση του PASW

Το SYNTAX της ανάλυσης είναι:

```
DATASET ACTIVATE DataSet1.
CORRELATIONS
  /VARIABLES=mikos taxythta
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

1) Επιλέγουμε την εντολή ***Analyze*** μετά ***Correlate*** και μετά ***Bivariate***



2) Μαυρίζουμε τις μεταβλητές *mikos, taxythta* και επιλέγουμε το ***βελάκι*** για να τις περάσουμε δεξιά στο Variables: μετά επιλέγουμε το ***OK*** για να τρέξουμε την ανάλυση.