ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΣΥΣΧΕΤΙΣΗ

Καθηγητής Αντώνης Κ. Τραυλός (B.A., M.A., Ph.D.)
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
Σχολή Επιστημών Ανθρώπινης Κίνησης και Ποιότητας Ζωής
Τμήμα Οργάνωσης και Διαχείρισης Αθλητισμού

Πηγή Διάλεξης: Βασισμένη στο κεφάλαιο 11 του συγγράμματος Βαγενάς (2019, σελ 217-242), σε επιλεγμένους διαδικτυακούς χώρους και θεματικές ενότητες των προτεινόμενων συγγραμμάτων.

1

1

ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΣΥΣΧΕΤΙΣΗ

- Ένα μεγάλο μέρος της στατιστικής ασχολείται με αναλύσεις συσχέτισης, της οποίας βασική προϋπόθεση είναι η εξέταση της συνδιακύμανσης δύο τουλάχιστον μετρήσιμων ιδιοτήτων.
- Στη βασική της μορφή η συσχέτιση περιλαμβάνει δύο μόνο μεταβλητές της ίδιας ομάδας ατόμων ή περιπτώσεων, δηλαδή ασχολείται με τη **διερεύνηση της συνάφειας σε μια διμεταβλητή κατανομή**, και λέγεται **απλή συσχέτιση (simple correlation)**.

2

2

Απλή Γραμμική Συσχέτιση … συνέχεια

- Υπάρχουν όμως περιπτώσεις που η διερεύνηση ενός φαινομένου προϋποθέτει την εξέταση **τριών**, **τεσσάρων** ή **και περισσότερων μεταβλητών ταυτόχρονα**, οπότε η ανάλυση συσχέτισης στηρίζεται σε **τριμεταβλητές**, **τετραμεταβλητές** ή γενικά **πολυμεταβλητές κατανομές**, αντίστοιχα.
- Η διερεύνηση της συνάφειας σε πολυμεταβλητές κατανομές μπορεί να γίνει με **δύο** γενικές μεθόδους ανάλυσης.
- Η πρώτη μέθοδος λέγεται **πολλαπλή συσχέτιση** και στοχεύει στην ποσοτικοποίηση της συνάφειας μεταξύ μιας μεταβλητής και της αριθμητικής σύνθεσης μιας ομάδας 2, 3 ή και περισσότερων μεταβλητών.
- Η δεύτερη λέγεται **κανονική συσχέτιση** και στοχεύει στην ποσοτικοποίηση της συνάφειας μεταξύ δύο ομάδων μεταβλητών.

3

Συμβολικό παράδειγμα διάταξης δεδομένων πολυμεταβλητής κατανομής (από Βαγενάς, 2019)

Πίνακας 11.1 - Συμβολική διάταξη δεδομένων στην ανάλυση συσχέτισης

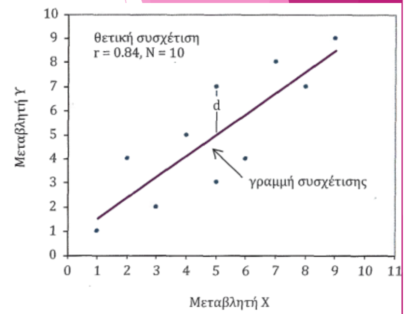
Άτομο (case)	Ομάδα μεταβλητών X				Ομάδα μεταβλητών Y			
	X_1	X_2	.	X_p	Y_1	Y_2	.	Y_q
1	X_{11}	X_{12}	.	X_{1p}	Y_{11}	Y_{12}	.	Y_{1q}
2	X_{21}	X_{22}	.	X_{2p}	Y_{21}	Y_{22}	.	Y_{2q}
3	X_{31}	X_{32}	.	X_{3p}	Y_{31}	Y_{32}	.	Y_{3q}
.	.	.	.	.	.	.	.	.
N	X_{N1}	X_{N2}	.	X_{Np}	Y_{N1}	Y_{N2}	.	Y_{Nq}

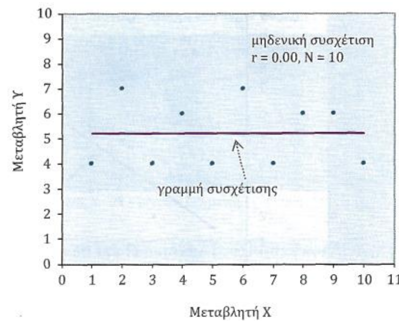
q = μεταβλητές X, p = μεταβλητές Y, N = μέγεθος δείγματος.

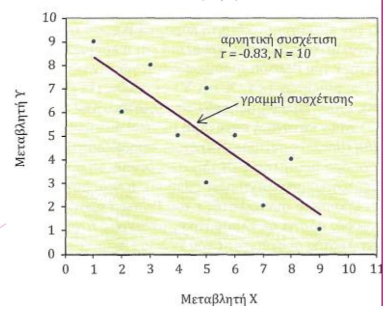
4

Βαθμός και Κατεύθυνση της Γραμμικής Συσχέτισης και Παραδοχές (από Βαγενάς, 2019)

- Η μορφή της συνάφειας δύο μεταβλητών (Χ και Υ) δίνεται στο διάγραμμα διασποράς ή απλώς διάγραμμα συσχέτισης (Σχ. 11.1, 11.2, 11.3). Το **διάγραμμα διασποράς** (scatter diagram, scatter plot) αποτελείται από όλα τα σημεία τομής των τεταγμένων των τιμών Χ και Υ κατά ζεύγη και από τη **<u>γραμμή συσχέτισης</u>** (correlation line) η οποία λέγεται και **γραμμή βέλτιστης προσαρμογής** (line of best fit).



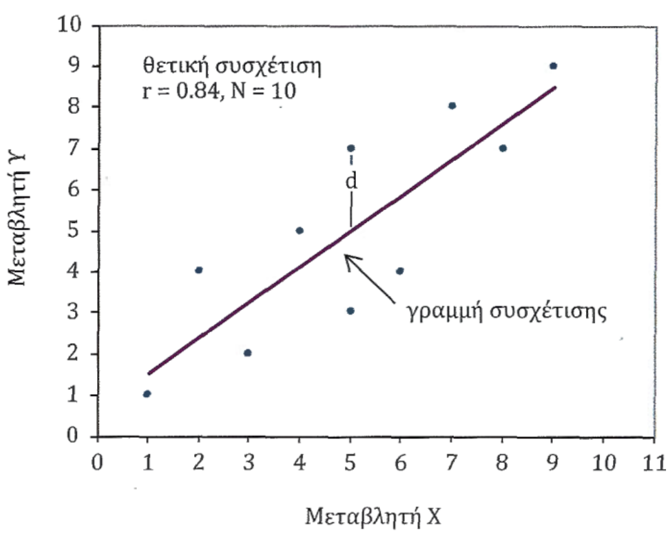


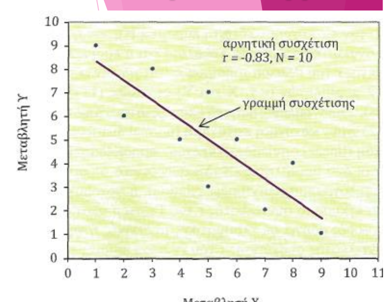


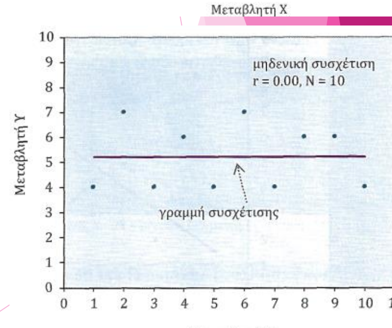
5

Παράδειγμα

X	Y
3	2
2	4
1	1
6	4
5	3
4	5
9	9
8	7
7	8
5	7





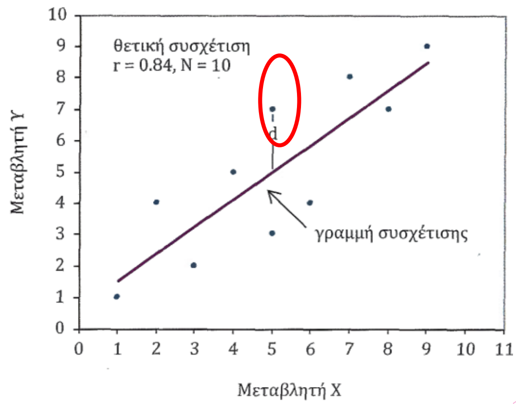


6

Παράδειγμα

- Τα σημεία τομής των τεταγμένων των τιμών της διμεταβλητής κατά ζεύγη δεν πέφτουν όλα πάνω στη γραμμή συσχέτισης, αλλά εμφανίζονται διασπαρμένα με ένα τρόπο που απεικονίζει **τον βαθμό** και **την κατεύθυνση** της συνάφειας για τη συγκεκριμένη ομάδα 10 ατόμων.
- Ο **βαθμός** της συσχέτισης (degree of linear correlation) αφορά **τις αποστάσεις (d)** των σημείων τομής των συντεταγμένων των ζευγών Χ-Υ από τη γραμμή συσχέτισης.

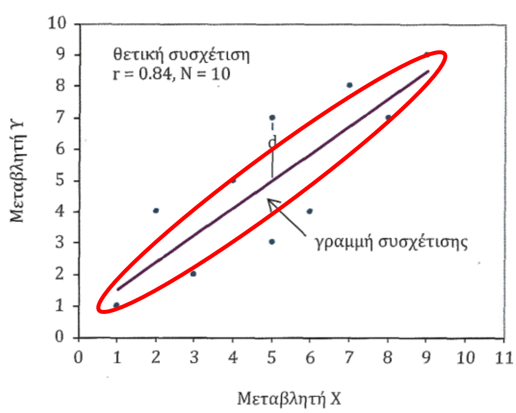
X	Y
3	2
2	4
1	1
6	4
5	3
4	5
9	9
8	7
7	8
5	7



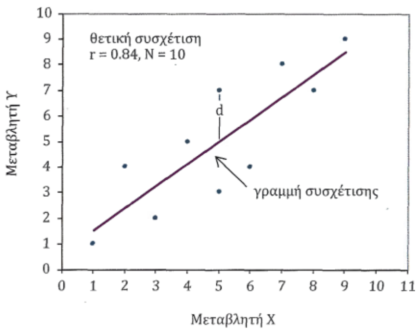
Παράδειγμα

- Η **κατεύθυνση** της συσχέτισης (direction of correlation) **αφορά την κλίση της γραμμής συσχέτισης** και δείχνει πόση μέση μεταβολή αντιστοιχεί στη μεταβλητή Υ σε σταθερή μεταβολή (ΔΧ) της μεταβλητής Χ.

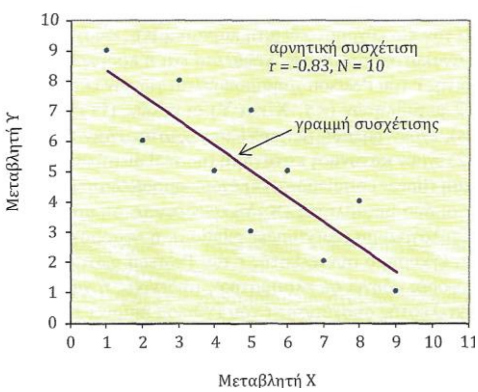
X	Y
3	2
2	4
1	1
6	4
5	3
4	5
9	9
8	7
7	8
5	7

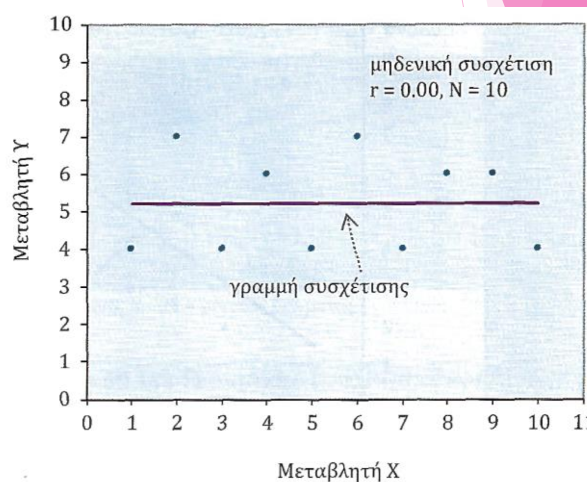


X	Y
3	2
2	4
1	1
6	4
5	3
4	5
9	9
8	7
7	8
5	7



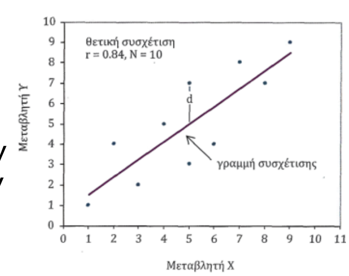
Έχουμε θετική, αρνητική και μηδενική συσχέτιση





9

Συσχέτιση και εξίσωση πρόβλεψης



- Η πλήρης ανάλυση της γραμμικής συσχέτισης περιλαμβάνει τον υπολογισμό του βαθμού συνάφειας (degree of correlation) των δύο ποσοτικών μεταβλητών και τον **προσδιορισμό της εξίσωσης πρόβλεψης** **<u>της μιας μεταβλητής (Y) από την άλλη (X)</u>**.
- Ο βαθμός συσχετισης (degree of correlation) ποσοτικοποιείται με τον συντελεστή συσχετισης r του Pearson (Pearson's correlation coefficient) που αποτελεί τον πιο εύχρηστο δείκτη παραμετρικής συσχέτισης.
- Η γραμμική εξίσωση που συνδέει τις δύο μεταβλητές (X και Y) είναι της μορφής

$$Y = a + bX$$

- Η υπάρχουσα συνάφεια συνοψίζεται γραφικά με ευθεία γραμμή της οποίας η κλίση εξαρτάται από την τιμή του **συντελεστή b** και η απόσταση του σημείου τομής της με τον άξονα Y από την τιμή της **σταθεράς a**.

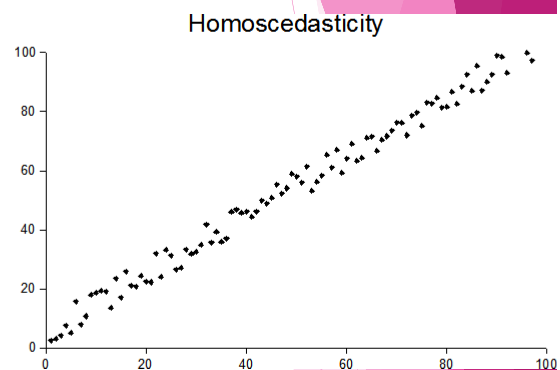
10

10

Συσχέτιση και τρόπος χρησιμοποίησης
(από Βαγενάς, 2019)

- Ο δείκτης r του Pearson χρησιμοποιείται για την εκτίμηση του βαθμού συνάφειας μεταξύ 2 μεταβλητών (π.χ. X και Y) οι οποίες

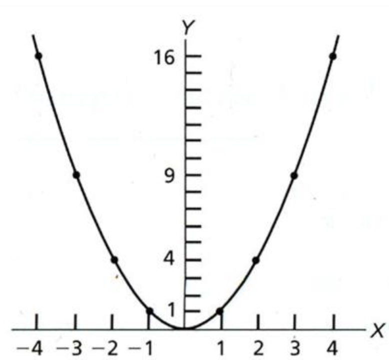
1. είναι **ποσοτικές** (quantitative variables), έχουν δηλαδή μετρηθεί τουλάχιστον στην διαστημική κλίμακα (interval scale),
2. έχουν **κανονική κατανομή** (normal distribution),
3. αποδίδουν **γραμμική συσχέτιση** (linear correlation),
4. έχουν **ομοσκεδαστικότητα** (homoskedasticity) (παρόμοια διασπορά μεταξύ X και Y) .



- Όταν οι δύο μεταβλητές (X, Y) δεν συνδέονται γραμμικά (non-linear relation), π.χ. καμπυλόγραμμα (curvilinear relation), η εφαρμογή της γραμμικής συσχέτισης δεν είναι έγκυρη.
- Για να καταλάβετε τι συμβαίνει όταν ο συντελεστής r του Pearson εφαρμόζεται σε δεδομένα που δεν συσχετίζονται γραμμικά, θεωρήστε τα στοιχεία στον Πίνακα 5.

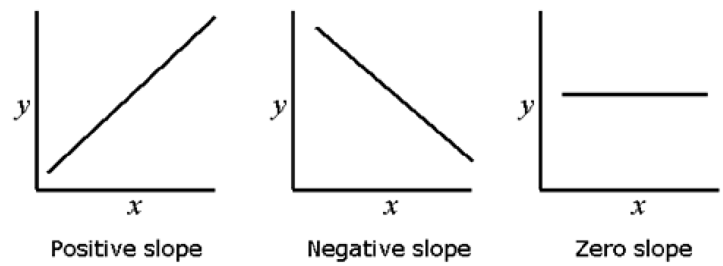
Πίνακας 5. Δεδομένα που παρουσιάζουν τέλεια μη γραμμική συσχέτιση που πηγάζει από τη σχέση $X = Y^2$

X	Y
4	16
3	9
2	4
1	1
0	0
−1	1
−2	4
−3	9
−4	16



Συσχέτιση και τρόπος χρησιμοποίησης

- Ο συντελεστής συσχέτισης r του Pearson παίρνει τιμές μεταξύ -1 και +1.
- Όταν 0 < r < 1, έχουμε **θετική συσχέτιση**,
- όταν-1 < r < 0 έχουμε **αρνητική συσχέτιση** και
- όταν r = 0 έχουμε **μηδενική συσχέτιση.**



13

Συντελεστής συσχέτισης Pearson r

- Ο συντελεστής συσχέτισης που χρησιμοποιείται συχνότερα είναι ο συντελεστής συσχέτισης του PEARSON, που συμβολίζεται από το r και αναπτύχθηκε από τον Άγγλο στατιστικό Karl Pearson (1857-1936).
- Για να καταλάβετε αυτόν τον συντελεστή, υποθέστε ότι υπάρχει μια θετική σχέση μεταξύ X και Y.
- Εάν ένα άτομο έχει μία τιμή στη μεταβλητή X που είναι επάνω από τη μέση τιμή του X (MX), αυτό το άτομο είναι πιθανό να έχει μία τιμή στη μεταβλητή Y που είναι επάνω από τη μέση τιμή του Y (MY).
- Ένα άτομο με μία τιμή X κάτω από MX είναι πιθανό να έχει τιμή Y κάτω από MY. Ομοίως, εάν η σχέση μεταξύ των μεταβλητών είναι αρνητική, κατόπιν ένα άτομο με μία τιμή X επάνω από MX είναι πιθανό να έχει μία τιμή Y κάτω από MY, και αντίστροφα.

14

14

- Ο Pearson ανέπτυξε τον συντελεστή συσχέτισης με αυτές τις σχέσεις στο μυαλό του.
- Ο συντελεστής περιλαμβάνει τον **υπολογισμό του αθροίσματος σταυρωτών γινομένων** (**sum of cross-products**) δηλαδή πολλαπλασιάζοντας τα δύο αποτελέσματα (X και Y) για κάθε άτομο και αθροίζοντας έπειτα αυτά τα γινόμενα στα άτομα n.
- Αυτό το άθροισμα διαιρείται συνέχεια με n-1.
- Στην ουσία, ο συντελεστής συσχέτισης είναι η μέση τιμή του αθροίσματος των αντίστοιχων γινομένων των αποτελεσμάτων (-τιμών).

Μέθοδος τυπικών τιμών z (standard score method)

- Λόγω της διαφοράς στις μετρήσεις για τις δύο μεταβλητές που συσχετίζονται, ο Pearson χρησιμοποίησε τις σταθερές τιμές z παρά τις αρχικές τιμές στην ανάπτυξη του συντελεστή συσχέτισης.
- Καθόρισε τον συντελεστή συσχέτισης ως εξής:

$$r_{XY} = \Sigma(z_X z_Y)/N\text{-}1$$

$$r = \frac{\sum (z_X z_Y)}{N}$$

Δουλεύοντας με ένα παράδειγμα!

- Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε τα στοιχεία στον πίνακα 1.
- Εάν χρησιμοποιήσαμε τον τύπο 1 ως υπολογιστικό τύπο, πρώτα υπολογίζουμε τη μέση τιμή και την τυπική απόκλιση των αποτελεσμάτων ΕΕ (Χ), M_X = 534.00 και s_X = 96.53.
- Κατόπιν υπολογίζουμε τη μέση τιμή και την τυπική απόκλιση των αποτελεσμάτων στο ΜΟΔΑ (Υ), M_Y = 51.47 και s_Y = 10.11.
- Στη συνέχεια μετατρέπουμε κάθε αρχική τιμή σε μία σταθερή τιμή z.
- Αυτό παρουσιάζεται παρακάτω για το πρώτο ζευγάρι των αποτελεσμάτων:

$z_X = (595 - 534)/96.53 = 0.63$ $z_Y = (68 - 51.47)/10.11 = 1.64$

Πίνακας 1. Αποτελέσματα ΕΕ και ΜΟΔΑ για 15 φοιτητές.

Φοιτητής	ΕΕ (Χ)	ΜΟΔΑ (Υ)
1	595	68
2	520	55
3	715	65
4	405	42
5	680	64
6	490	45
7	565	56
8	580	59
9	615	56
10	435	42
11	440	38
12	515	50
13	380	37
14	510	42
15	565	53
Σ	8,010	772
	$\overline{X} = 534.00$	$\overline{Y} = 51.47$
	$s_X = 96.53$	$s_Y = 10.11$

$\boxed{r_{XY} = (\Sigma z_X z_Y)/(n-1)}$ $z_X = (595 - 534)/96.53 = 0.63$ $z_Y = (68 - 51.47)/10.11 = 1.64$

Πίνακας 2. Δεδομένα για τον υπολογισμό του συντελεστή συσχέτισης r του Pearson χρησιμοποιώντας τον τύπο 1.

	X	Y	z_X	z_Y	$z_X z_Y$
	595	68	0.63	1.64	1.03
	520	55	−0.15	0.35	−0.05
	715	65	1.88	1.34	2.52
	405	42	−1.34	−0.94	1.26
	680	64	1.51	1.24	1.87
	490	45	−0.46	−0.64	0.29
	565	56	0.32	0.45	0.14
	580	59	0.48	0.74	0.36
	615	56	0.84	0.45	0.38
	435	42	−1.03	−0.94	0.97
	440	38	−0.97	−1.33	1.29
	515	50	−0.20	−0.15	0.03
	380	37	−1.60	−1.43	2.29
	510	42	−0.25	−0.94	0.24
	565	53	0.32	0.15	0.05
Σ	8,010	772	0.00	0.00	12.67

Στη συνέχεια το άθροισμα των αντίστοιχων γινομένων διαιρείται με n - 1, ως εξής:

- $r_{XY} = (\Sigma z_X z_Y)/(n-1)$
- $r_{XY} = 12.67/14$
- $r_{XY} = 0.90$

Πίνακας 2. Δεδομένα για τον υπολογισμό του συντελεστή συσχέτισης r του Pearson χρησιμοποιώντας τον τύπο 1.

	X	Y	z_X	z_Y	$z_X z_Y$
	595	68	0.63	1.64	1.03
	520	55	−0.15	0.35	−0.05
	715	65	1.88	1.34	2.52
	405	42	−1.34	−0.94	1.26
	680	64	1.51	1.24	1.87
	490	45	−0.46	−0.64	0.29
	565	56	0.32	0.45	0.14
	580	59	0.48	0.74	0.36
	615	56	0.84	0.45	0.38
	435	42	−1.03	−0.94	0.97
	440	38	−0.97	−1.33	1.29
	515	50	−0.20	−0.15	0.03
	380	37	−1.60	−1.43	2.29
	510	42	−0.25	−0.94	0.24
	565	53	0.32	0.15	0.05
Σ	8,010	772	0.00	0.00	12.67

19

Πόνος !!!!!!!!!!

- Η χρησιμοποίηση του τύπου ($r_{XY} = (\Sigma z_X z_Y)/(n-1)$) για τον υπολογισμό του συντελεστή συσχέτισης μεταξύ δύο μεταβλητών είναι σκληρή και επίπονη, επειδή κάθε αρχική τιμή πρέπει πρώτα να μετατραπεί σε μία τιμή z.
- Εντούτοις, με τη χρησιμοποίηση των σταθερών τιμών και των τυπικών αποκλίσεων, μπορούμε να παραγάγουμε τους καταλληλότερους τύπους για τον υπολογισμό του συντελεστή συσχέτισης.
- Ένας από αυτούς τους υπολογιστικούς τύπους είναι ο τύπος των τιμών απόκλισης.
- Ένας άλλος τύπος χρησιμοποιεί τις αρχικές τιμές των δεδομένων. Τέλος, θα περιγράψουμε πώς μπορούμε να υπολογίσουμε *το συντελεστή συσχέτισης r* χρησιμοποιώντας τη συνδιακύμανση (covariance).

20

20

Μέθοδος τιμών απόκλισης-1 (deviation score method -1)

- Το άθροισμα των σταυρωτών γινομένων διαιρείται με το πλήθος N του δείγματος καθώς και με τις τυπικές αποκλίσεις s_x και s_y των δύο μεταβλητών:

$$r = \frac{\sum xy}{(N-1)s_X s_Y}$$

- όπου
  - x, y οι αποκλίσεις (deviations) από τους αντίστοιχους μέσους,
  - N το μέγεθος του δείγματος,
  - s_x και s_y οι αμερόληπτες (με N-1) τυπικές αποκλίσεις.
- Ο τύπος αυτός είναι κατάλληλος για περιπτώσεις που οι τυπικές αποκλίσεις των μεταβλητών είναι ήδη υπολογισμένες.

21

21

Μέθοδος τιμών απόκλισης-2 (deviation score method -2)

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{12,332}{\sqrt{(130,460)(1,429.72)}} = 0.90$$

Πίνακας 3. Υπολογισμός συντελεστή συσχέτισης με τον τύπο των τιμών απόκλισης

X	Y	x	y	xy	x^2	y^2
595	68	61.0	16.53	1,008.33	3,721.0	273.24
520	55	− 14.0	3.53	− 49.42	196.0	12.46
715	65	181.0	13.53	2,448.93	32,761.0	183.06
405	42	−129.0	− 9.47	1,221.63	16,641.0	89.68
680	64	146.0	12.53	1,829.38	21,316.0	157.00
490	45	− 44.0	−6.47	284.68	1,936.0	41.86
565	56	31.0	4.53	140.43	961.0	20.52
580	59	46.0	7.53	346.38	2,116.0	56.70
615	56	81.0	4.53	366.93	6,561.0	20.52
435	42	− 99.0	− 9.47	937.53	9,801.0	89.68
440	38	− 94.0	−13.47	1,266.18	8,836.0	181.44
515	50	− 19.0	− 1.47	27.93	361.0	2.16
380	37	−154.0	−14.47	2,228.38	23,716.0	209.38
510	42	− 24.0	− 9.47	227.28	576.0	89.68
565	53	31.0	1.53	47.43	961.0	2.34
8,010	772	0.0	0.0	12,332.00	130,460.0	1,429.72

22

Μέθοδος αρχικών τιμών (raw score method)

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

$$r = \frac{15(424,580) - (8,010)(772)}{\sqrt{[15(4,407,800) - (8,010)^2][15(41,162) - (772)^2]}}$$

$$= 0.90$$

Πίνακας 4. Δεδομένα για τον υπολογισμό του συντελεστή συσχέτισης χρησιμοποιώντας τον τύπο των αρχικών τιμών

X	Y	XY	X^2	Y^2
595	68	40,460	354,025	4,624
520	55	28,600	270,400	3,025
715	65	46,475	511,225	4,225
405	42	17,010	164,025	1,764
680	64	43,520	462,400	4,096
490	45	22,050	240,100	2,025
565	56	31,640	319,225	3,136
580	59	34,220	336,400	3,481
615	56	34,440	378,225	3,136
435	42	18,270	189,225	1,764
440	38	16,720	193,600	1,444
515	50	25,750	265,225	2,500
380	37	14,060	144,400	1,369
510	42	21,420	260,100	1,764
565	53	29,945	319,225	2,809
8,010	772	424,580	4,407,800	41,162

23

23

Στατιστική Σημαντικότητα του r (από Βαγενάς, 2019)

Έμμεσος Έλεγχος της Στατιστικής Σημαντικότητας του r.

- Ο **έμμεσος έλεγχος σημαντικότητας** του r γίνεται για τον έλεγχο της υπόθεσης H_0: ρ = 0 (ότι δηλαδή η συσχέτιση των 2 μεταβλητών στον πληθυσμό είναι μηδέν) και μετά από μετασχηματισμό Fisher της τιμής του r σε t.

Άμεσος Έλεγχος της Στατιστικής Σημαντικότητας του r.

- Οι κρίσιμες τιμές r ισχύουν, σε αντιστοιχία με τις κρίσιμες τιμές t, και πάλι για τον έλεγχο της υπόθεσης H_0: ρ = 0, ο οποίος γίνεται άμεσα ως εξής:
- 1. Καθορίζουμε το επίπεδο πιθανότητας α (συνήθως 0.05 ή 0.01).
- 2. Υπολογίζουμε του βαθμούς ελευθερίας **df = N-2**.
- 3. Από τον Πίν. 21.Ε εντοπίζουμε την κρίσιμη τιμή (r_c).
- 4. Συγκρίνουμε την απόλυτη τιμή r με την κρίσιμη τιμή rc και
  - αν r > r_c η συσχετιση είναι στατιστικός σημαντική,
  - αν r < r_c η συσχέτιση είναι στατιστικώς μη σημαντική.

24

24

Πίνακας 4. Δεδομένα για τον υπολογισμό του συντελεστή συσχέτισης χρησιμοποιώντας τον τύπο των αρχικών τιμών

X	Y	XY	X^2	Y^2
595	68	40,460	354,025	4,624
520	55	28,600	270,400	3,025
715	65	46,475	511,225	4,225
405	42	17,010	164,025	1,764
680	64	43,520	462,400	4,096
490	45	22,050	240,100	2,025
565	56	31,640	319,225	3,136
580	59	34,220	336,400	3,481
615	56	34,440	378,225	3,136
435	42	18,270	189,225	1,764
440	38	16,720	193,600	1,444
515	50	25,750	265,225	2,500
380	37	14,060	144,400	1,369
510	42	21,420	260,100	1,764
565	53	29,945	319,225	2,809
8,010	772	424,580	4,407,800	41,162

$$r = \frac{15(424,580) - (8,010)(772)}{\sqrt{[15(4,407,800) - (8,010)^2][15(41,162) - (772)^2]}}$$

$$= 0.90$$

	Significance Level			
1-tailed	0.05	0.025	0.01	0.005
2-tailed	0.1	0.05	0.02	0.01
df				
1	0.988	0.997	0.9995	0.9999
2	0.9	0.95	0.98	0.99
3	0.805	0.878	0.934	0.959
4	0.729	0.811	0.882	0.917
5	0.669	0.754	0.833	0.874
6	0.622	0.707	0.789	0.834
7	0.582	0.666	0.75	0.798
8	0.549	0.632	0.716	0.765
9	0.521	0.602	0.685	0.735
10	0.497	0.576	0.658	0.708
11	0.476	0.553	0.634	0.684
12	0.458	0.532	0.612	0.661
13	0.441	0.514	0.592	0.641
14	0.426	0.497	0.574	0.628
15	0.412	0.482	0.558	0.606
16	0.4	0.468	0.542	0.59
17	0.389	0.456	0.528	0.575
18	0.378	0.444	0.516	0.561
19	0.369	0.433	0.503	0.549
20	0.36	0.423	0.492	0.537
21	0.352	0.413	0.482	0.526
22	0.344	0.404	0.472	0.515
23	0.337	0.396	0.462	0.505
24	0.33	0.388	0.453	0.495
25	0.323	0.381	0.445	0.487
26	0.317	0.374	0.437	0.479
27	0.311	0.367	0.43	0.471
28	0.306	0.361	0.423	0.463
29	0.301	0.355	0.416	0.456
30	0.296	0.349	0.409	0.449
35	0.275	0.325	0.381	0.418
40	0.257	0.304	0.358	0.393
45	0.243	0.288	0.338	0.372
50	0.231	0.273	0.322	0.354
60	0.211	0.25	0.295	0.325
70	0.195	0.232	0.274	0.302
80	0.183	0.217	0.256	0.284
90	0.173	0.205	0.242	0.267
100	0.164	0.195	0.23	0.254

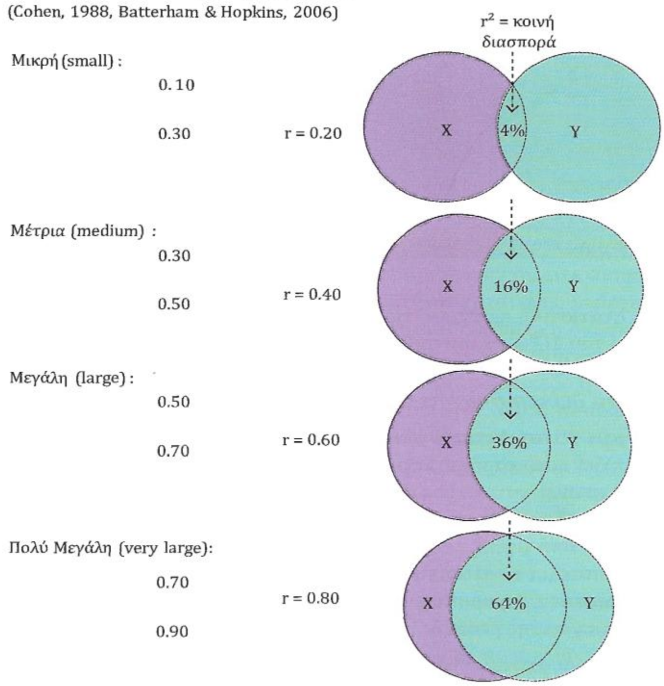
25

Ερμηνεία Γραμμικής Συσχέτισης

α. Συντελεστής προσδιορισμού r^2: Υπολογισμός και ερμηνεία

- Ο συντελεστής προσδιορισμού (r^2) δύο μεταβλητών δείχνει **το ποσοστό (%) της κοινής διασποράς** (common variance) των δύο μεταβλητών:
- (%) Συντελεστής Προσδιορισμού = 100 * r^2

Τυπικά Επίπεδα (Μεγέθη) Συσχέτισης (όρια) (Cohen, 1988, Batterham & Hopkins, 2006)



26

Χαρακτηριστικές συσχετίσεις και αντίστοιχες αξιολογήσεις (από Βαγενάς, 2019)

Συσχέτιση r	Κοινή Διασπορά r^2	%	Μέγεθος της Συσχέτισης (Effect Size)
0.0 - 0.1	0.00 - 0.01	0 - 1	Αμελητέα (trivial)
0.1 - 0.3	0.01 - 0.09	1 - 9	Μικρή (small)
0.3 - 0.5	0.09 - 0.25	9 - 25	Μέτρια (moderate)
0.5 - 0.7	0.25 - 0.49	25 - 49	Μεγάλη (large)
0.7 - 0.9	0.49 - 0.81	49 - 81	Πολύ μεγάλη (very large)
0.9 - 1	0.81 - 1.00	81 - 100	Σχεδόν τέλεια (nearly perfect)

Γενικά τιμές κοινής διασποράς $r^2 < 10\%$ έχουν συνήθως μικρή πρακτική αξία.

Ιδιότητες του δείκτη συσχέτισης r (Pearson)

- **Συσχέτιση & Αίτιο-Αιτιατό** (δεν παρέχει καμία πληροφορία για την αιτία που προκαλεί τη σχέση μεταξύ των 2 μεταβλητών).
- **Συσχέτιση & Αριθμητικοί Μετασχηματισμοί** - Οι αρχικές τιμές των μεταβλητών που συμμετέχουν σε μια ανάλυση απλής (γραμμικής) συσχέτισης μπορούν να υποστούν οποιουσδήποτε γραμμικούς αριθμητικούς μετασχηματισμούς (π.χ. X+C, XC ή/και Y+C, YC) **χωρίς αυτό να αλλοιώνει τον βαθμό συσχέτισης**, δηλαδή την τιμή του r. Αυτό έχει τεράστια πρακτική σημασία, καθότι σε πολλές περιπτώσεις είτε τίθεται η ανάγκη μετασχηματισμού των αρχικών τιμών X & Y, είτε οι τιμές αυτές είναι ήδη μετασχηματισμένες σε τυπικές τιμές (z, T κ.λπ.).
- **Συσχέτιση & Δειγματική Διασπορά** - Οι τιμές του συντελεστή συσχέτισης r διαφέρουν από δείγμα σε δείγμα ιδιαίτερα για μικρά δειγματικά πλήθη. Το **μέγεθος του δείγματος N**, καθώς και το **μέγεθος του συντελεστή συσχέτισης r** σε σχέση με τη **στατιστική σημαντικότητα** απαιτούν ιδιαίτερη προσοχή κατά την ερμηνεία τους.

Ομοιογένεια ομάδας δεδομένων και μέγεθος συντελεστή συσχέτισης

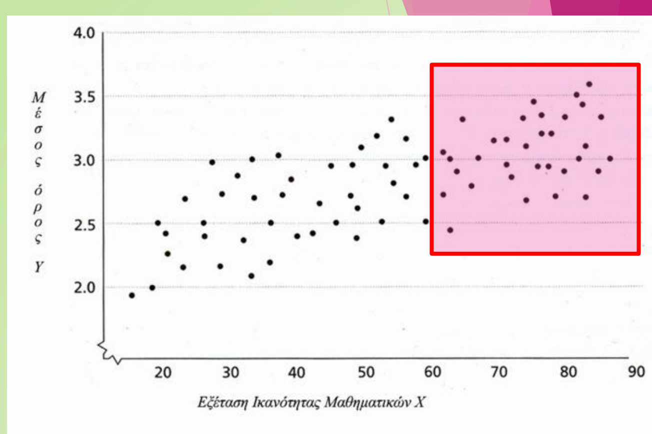
- Ένας παράγοντας που έχει επιπτώσεις στο μέγεθος του συντελεστή συσχέτισης r του Pearson είναι η ομοιογένεια της ομάδας για την οποία οι δύο μεταβλητές συσχετίζονται.
- **Όταν μια ομάδα είναι ομοιογενής** στη μία ή και στις δύο μεταβλητές, η διασπορά (και ως εκ τούτου η τυπική απόκλιση) τείνει προς μηδέν.
- Παρατηρείστε στον τύπο ($r_{XY} = s_{XY}/s_X s_Y$), πως όταν συμβαίνει αυτό, διαιρούμε με μηδέν, και ο τύπος εφαρμόζεται χωρίς να υπάρχει νόημα.
- Στην ουσία, η μεταβλητή έχει μειωθεί σε μια σταθερά.
- Δεδομένου ότι μια εξεταζόμενη ομάδα γίνεται όλο και περισσότερο ομοιογενής, τότε ο συντελεστής συσχέτισης μειώνεται.

29

29

Ας εξετάσουμε το παράδειγμα!

- Υποθέστε ότι ένας ερευνητής ενδιαφέρεται, για τη σχέση μεταξύ των αποτελεσμάτων μιας εξέτασης στην **ικανότητα μαθηματικών** και το **μέσο όρο των βαθμών του πρώτου εξαμήνου**, των 75 πρωτοετών φοιτητών/τριών με στόχο τον διαχωρισμό των καλύτερων φοιτητών/τριών.
- Αυτά τα στοιχεία απεικονίζονται στο σχήμα. Ο συντελεστής συσχέτισης για αυτά τα στοιχεία είναι 0.70.
- Ας υποθέσουμε ότι οι φοιτητές/τριες θα επιλεγούν με κριτήριο την επίδοση 60 στα μαθηματικά. Παρατηρούμε ότι οι 34 φοιτητές/τριες που σημειώνουν βαθμό πάνω από 60 δημιουργούν μία ομοιογενή υποομάδα. Ο συντελεστής συσχέτισης για αυτήν την ομοιογενή υποομάδα είναι 0.40.



30

30

Μερική Συσχέτιση ($r_{YX.Z}$) (από Βαγενάς, 2019)

- Όταν η ανάλυση συσχέτισης αφορά πολλές μεταβλητές, **τίθεται ζήτημα πιθανών επιδράσεων τρίτων μεταβλητών** στις υπό διερεύνηση συσχετίσεις.
- Στις περιπτώσεις αυτές για να υπολογίσουμε την πραγματική σχέση μεταξύ 2 μεταβλητών (X, Y) υπολογίζουμε μια εκδοχή τής συσχέτισης r (zero order correlation), στην οποία έχει αφαιρεθεί η συνεπίδραση (covariance) μιας 3ης μεταβλητής (π.χ., Z).
- Η εκδοχή αυτή είναι η **μερική συσχέτιση (partial correlation)**.
- Η μερική συσχέτιση της X με την Y με συμμεταβλητή τη Z είναι:

$$r_{yx.z} = \frac{r_{yx} - (rxz)(ryz)}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}}$$

- όπου r_{XY}, r_{XZ}, r_{YZ} οι συσχετίσεις μεταξύ των 3 μεταβλητών. Η σημαντικότητά της ελέγχεται με βαθμούς ελευθερίας df = N —3.
- **Δουλεύουμε στο Παράδειγμα 11.4α**.

	Significance Level			
1-tailed	**0.05**	**0.025**	**0.01**	**0.005**
2-tailed	**0.1**	**0.05**	**0.02**	**0.01**
df				
1	0.988	0.997	0.9995	0.9999
2	0.9	0.95	0.98	0.99
3	0.805	0.878	0.934	0.959
4	0.729	0.811	0.882	0.917
5	0.669	0.754	0.833	0.874
6	0.622	0.707	0.789	0.834
7	0.582	0.666	0.75	0.798
8	0.549	0.632	0.716	0.765
9	0.521	0.602	0.685	0.735
10	0.497	0.576	0.658	0.708
11	0.476	0.553	0.634	0.684
12	0.458	0.532	0.612	0.661
13	0.441	0.514	0.592	0.641
14	0.426	0.497	0.574	0.628
15	0.412	0.482	0.558	0.606
16	0.4	0.468	0.542	0.59
17	0.389	0.456	0.528	0.575
18	0.378	0.444	0.516	0.561
19	0.369	0.433	0.503	0.549
20	0.36	0.423	0.492	0.537
21	0.352	0.413	0.482	0.526
22	0.344	0.404	0.472	0.515
23	0.337	0.396	0.462	0.505
24	0.33	0.388	0.453	0.495
25	0.323	0.381	0.445	0.487
26	0.317	0.374	0.437	0.479
27	0.311	0.367	0.43	0.471
28	0.306	0.361	0.423	0.463
29	0.301	0.355	0.416	0.456
30	0.296	0.349	0.409	0.449
35	0.275	0.325	0.381	0.418
40	0.257	0.304	0.358	0.393
45	0.243	0.288	0.338	0.372
50	0.231	0.273	0.322	0.354
60	0.211	0.25	0.295	0.325
70	0.195	0.232	0.274	0.302
80	0.183	0.217	0.256	0.284
90	0.173	0.205	0.242	0.267
100	0.164	0.195	0.23	0.254

Μερική Συσχέτιση ($r_{YX.Z}$) (π.χ., 11.4α)

- Σε έρευνα N = 28 παιδιών - εφήβων ηλικίας 8-18 ετών η γνώση (X), η δύναμη (Y) και η ηλικία (Z) έχουν συσχέτιση r_{YX} = 0.50, r_{xz} = 0.60, και r_{YZ} = 0.80, αντίστοιχα.
- Η μερική συσχέτιση της γνώσης (X) με τη δύναμη (Y) χωρίς τη συνεπίδραση της ηλικίας (Z) στη X και στην Y είναι
- $r_{yx.z} = \frac{r_{yx} - (rxz)(ryz)}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}} = \frac{0.50 - (0.60)(0.80)}{\sqrt{1-(0.60)^2} * \sqrt{1-(0.80)^2}} = \frac{0.50 - 0.48}{0.8 * 0.6} = \frac{0.02}{0.48} = 0.042$
- Με **επίπεδο α = 0.05** και **βαθμούς ελευθερίας df** = N-3 = 28 - 3 = 25.
- Η κρίσιμη τιμή είναι r_c = 0.381. Συμπεραίνουμε ότι η μερική συσχέτιση 0.04 της γνώσης (X) με τη δύναμη (Y) είναι **στατιστικώς μη σημαντική**, ενώ η απλή συσχέτιση (0.50) είναι στατιστικώς σημαντική και θα μπορούσε να οδηγήσει στο παράδοξο συμπέρασμα "ότι η γνώση συνεπιδρά με τη μυϊκή δύναμη".

	Significance Level			
1-tailed	**0.05**	**0.025**	**0.01**	**0.005**
2-tailed	**0.1**	**0.05**	**0.02**	**0.01**
df				
1	0.988	0.997	0.9995	0.9999
2	0.9	0.95	0.98	0.99
3	0.805	0.878	0.934	0.959
4	0.729	0.811	0.882	0.917
5	0.669	0.754	0.833	0.874
6	0.622	0.707	0.789	0.834
7	0.582	0.666	0.75	0.798
8	0.549	0.632	0.716	0.765
9	0.521	0.602	0.685	0.735
10	0.497	0.576	0.658	0.708
11	0.476	0.553	0.634	0.684
12	0.458	0.532	0.612	0.661
13	0.441	0.514	0.592	0.641
14	0.426	0.497	0.574	0.628
15	0.412	0.482	0.558	0.606
16	0.4	0.468	0.542	0.59
17	0.389	0.456	0.528	0.575
18	0.378	0.444	0.516	0.561
19	0.369	0.433	0.503	0.549
20	0.36	0.423	0.492	0.537
21	0.352	0.413	0.482	0.526
22	0.344	0.404	0.472	0.515
23	0.337	0.396	0.462	0.505
24	0.33	0.388	0.453	0.495
25	0.323	0.381	0.445	0.487
26	0.317	0.374	0.437	0.479
27	0.311	0.367	0.43	0.471
28	0.306	0.361	0.423	0.463
29	0.301	0.355	0.416	0.456
30	0.296	0.349	0.409	0.449
35	0.275	0.325	0.381	0.418
40	0.257	0.304	0.358	0.393
45	0.243	0.288	0.338	0.372
50	0.231	0.273	0.322	0.354
60	0.211	0.25	0.295	0.325
70	0.195	0.232	0.274	0.302
80	0.183	0.217	0.256	0.284
90	0.173	0.205	0.242	0.267
100	0.164	0.195	0.23	0.254

Ημι-Μερική Συσχέτιση ($r_{Y(X.Z)}$) (π.χ., 11.4β)

- Η ημι-μερική συσχέτιση (part ή semipartial correlation) είναι η συσχέτιση μεταξύ 2 μεταβλητών (π.χ. X, Y) μετά την αφαίρεση της συνεπίδρασης μιας 3ης μεταβλητής (π.χ. Z) μόνο από την μία εκ των 2 αρχικών μεταβλητών (π.χ. από την X).
- Η ημι-μερική συσχέτιση της X με την Y με συμμεταβλητή τής X τη Z είναι

$$r_{Y(X.Z)} = \frac{r_{YX} - (r_{YZ})(r_{XZ})}{\sqrt{1 - r_{XZ}^2}}$$

- όπου r_{XY}, r_{XZ}, r_{YZ} οι συσχετίσεις μεταξύ των 3 μεταβλητών.
- Η σημαντικότητά της ελέγχεται με **βαθμούς ελευθερίας df** = N −3.
- **Δουλεύουμε στο Παράδειγμα 11.4β**.

	Significance Level			
1-tailed	0.05	0.025	0.01	0.005
2-tailed	0.1	0.05	0.02	0.01
df				
1	0.988	0.997	0.9995	0.9999
2	0.9	0.95	0.98	0.99
3	0.805	0.878	0.934	0.959
4	0.729	0.811	0.882	0.917
5	0.669	0.754	0.833	0.874
6	0.622	0.707	0.789	0.834
7	0.582	0.666	0.75	0.798
8	0.549	0.632	0.716	0.765
9	0.521	0.602	0.685	0.735
10	0.497	0.576	0.658	0.708
11	0.476	0.553	0.634	0.684
12	0.458	0.532	0.612	0.661
13	0.441	0.514	0.592	0.641
14	0.426	0.497	0.574	0.628
15	0.412	0.482	0.558	0.606
16	0.4	0.468	0.542	0.59
17	0.389	0.456	0.528	0.575
18	0.378	0.444	0.516	0.561
19	0.369	0.433	0.503	0.549
20	0.36	0.423	0.492	0.537
21	0.352	0.413	0.482	0.526
22	0.344	0.404	0.472	0.515
23	0.337	0.396	0.462	0.505
24	0.33	0.388	0.453	0.495
25	0.323	0.381	0.445	0.487
26	0.317	0.374	0.437	0.479
27	0.311	0.367	0.43	0.471
28	0.306	0.361	0.423	0.463
29	0.301	0.355	0.416	0.456
30	0.296	0.349	0.409	0.449
35	0.275	0.325	0.381	0.418
40	0.257	0.304	0.358	0.393
45	0.243	0.288	0.338	0.372
50	0.231	0.273	0.322	0.354
60	0.211	0.25	0.295	0.325
70	0.195	0.232	0.274	0.302
80	0.183	0.217	0.256	0.284
90	0.173	0.205	0.242	0.267
100	0.164	0.195	0.23	0.254

33

Ημι-Μερική Συσχέτιση ($r_{Y(X.Z)}$) (από Βαγενάς, 2019)

- Σε έρευνα N = 28 παιδιών και εφήβων ηλικίας 8 – 18 ετών η γνώση (X), η γνώση (X), η δύναμη (Y) και η ηλικία (Z) είχαν συσχετίσεις

$$r_{YX} = 0.50,\ r_{XZ} = 0.60,\ r_{YZ} = 0.80$$

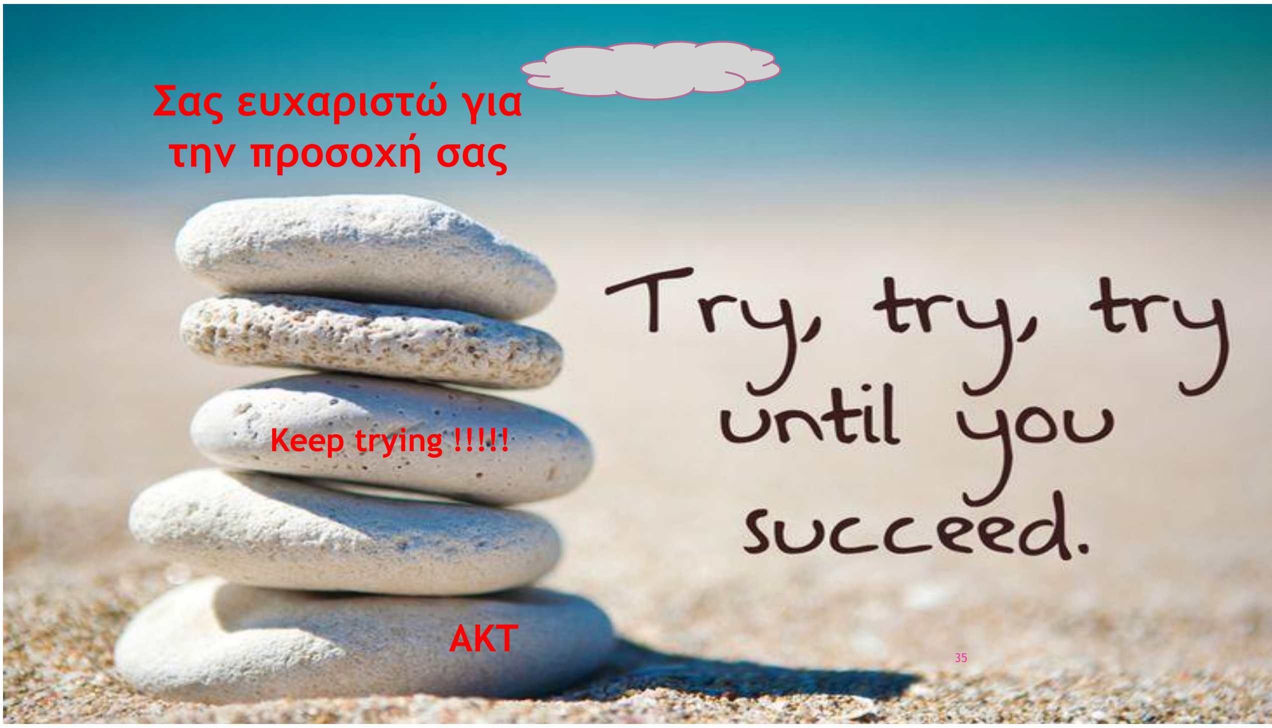
- Η ημι-μερική συσχέτιση της γνώσης (X) με τη δύναμη (Y) χωρίς την συνεπίδραση της ηλικίας (Z) στη X μόνο (όχι και στην Y) είναι

$$r_{Y(X.Z)} = \frac{r_{YX} - (r_{YZ})(r_{XZ})}{\sqrt{1 - r_{XZ}^2}} = \frac{0.50 - (0.80)(0.60)}{\sqrt{1-(0.60)^2}} = \frac{0.50 - 0.48}{\sqrt{1-0.36}} = \frac{0.02}{0.80} = 0.025$$

- Με επίπεδο α = 0.05 και βαθμούς ελευθερίας df = N - 3 = 28 - 3 = 25.
- Άρα η (ημι-μερική) συσχέτιση 0.025 της γνώσης (X) με τη δύναμη (Y) (δηλαδή χωρίς την επίδραση της ηλικίας στη γνώση) είναι **στατιστικώς μη σημαντική**.

	Significance Level			
1-tailed	0.05	0.025	0.01	0.005
2-tailed	0.1	0.05	0.02	0.01
df				
1	0.988	0.997	0.9995	0.9999
2	0.9	0.95	0.98	0.99
3	0.805	0.878	0.934	0.959
4	0.729	0.811	0.882	0.917
5	0.669	0.754	0.833	0.874
6	0.622	0.707	0.789	0.834
7	0.582	0.666	0.75	0.798
8	0.549	0.632	0.716	0.765
9	0.521	0.602	0.685	0.735
10	0.497	0.576	0.658	0.708
11	0.476	0.553	0.634	0.684
12	0.458	0.532	0.612	0.661
13	0.441	0.514	0.592	0.641
14	0.426	0.497	0.574	0.628
15	0.412	0.482	0.558	0.606
16	0.4	0.468	0.542	0.59
17	0.389	0.456	0.528	0.575
18	0.378	0.444	0.516	0.561
19	0.369	0.433	0.503	0.549
20	0.36	0.423	0.492	0.537
21	0.352	0.413	0.482	0.526
22	0.344	0.404	0.472	0.515
23	0.337	0.396	0.462	0.505
24	0.33	0.388	0.453	0.495
25	0.323	0.381	0.445	0.487
26	0.317	0.374	0.437	0.479
27	0.311	0.367	0.43	0.471
28	0.306	0.361	0.423	0.463
29	0.301	0.355	0.416	0.456
30	0.296	0.349	0.409	0.449
35	0.275	0.325	0.381	0.418
40	0.257	0.304	0.358	0.393
45	0.243	0.288	0.338	0.372
50	0.231	0.273	0.322	0.354
60	0.211	0.25	0.295	0.325
70	0.195	0.232	0.274	0.302
80	0.183	0.217	0.256	0.284
90	0.173	0.205	0.242	0.267
100	0.164	0.195	0.23	0.254

34


Σας ευχαριστώ για
την προσοχή σας
Keep trying !!!!!
AKT
Try, try, try
until you
succeed.
35

35