Στατιστικές Υποθέσεις ΙΙ

(Ερευνητικά Ερωτήματα / Υποθέσεις προς επιβεβαίωση)

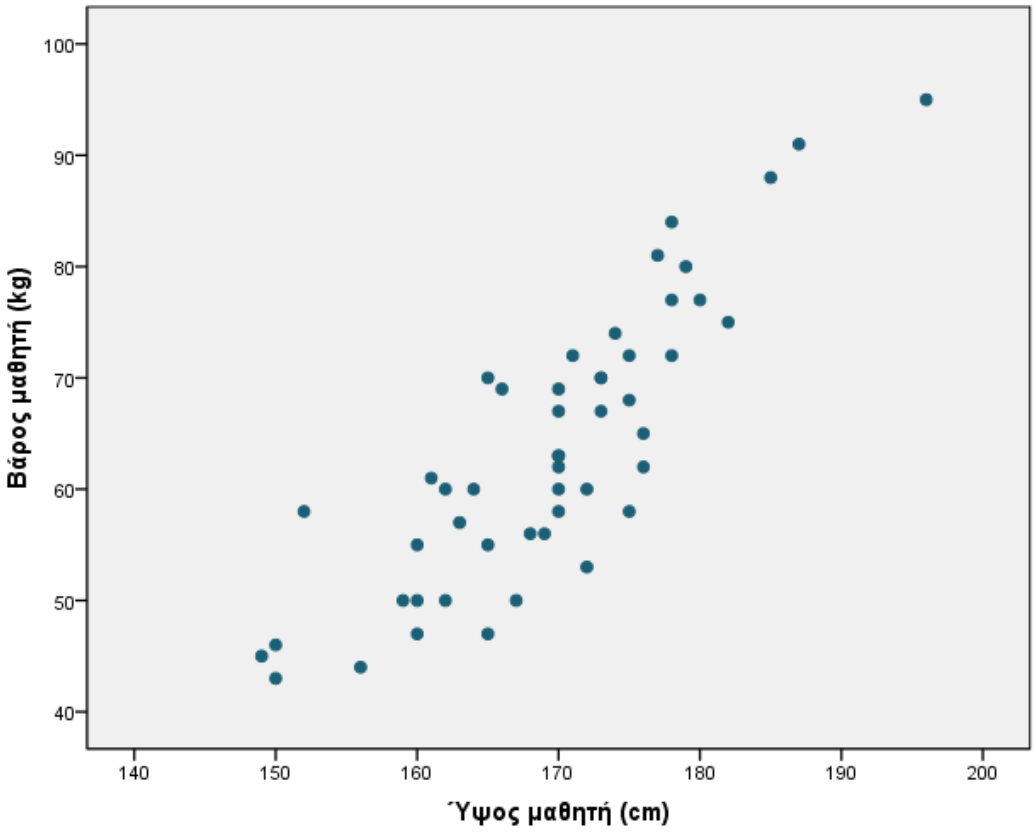
Περιπτώσεις

1. Ποιοτική με Ποσοτική
2. **<u>Ποσοτική με Ποσοτική</u>**
3. Ποιοτική με Ποιοτική

Παράδειγμα 1
(δυο ποσοτικά χαρακτ.)

- Θέλουμε να ελέγξουμε αν το ύψος έχει κάποια σχέση με το βάρος σε άτομα 15-18 ετών (ας πούμε σε μαθητές Λυκείου).
- Μετρήσαμε το ύψος και το βάρος σε 50 μαθητές
- Τα δεδομένα παρουσιάζονται σε ένα γράφημα

Διάγραμμα Διάχυσης (Διασποράς) Scatter



Σε όλες τις περιπτώσεις που ζητάμε συσχέτιση χαρακτηριστικών, ενδιαφερόμαστε να δώσουμε απάντηση στα ερωτήματα:

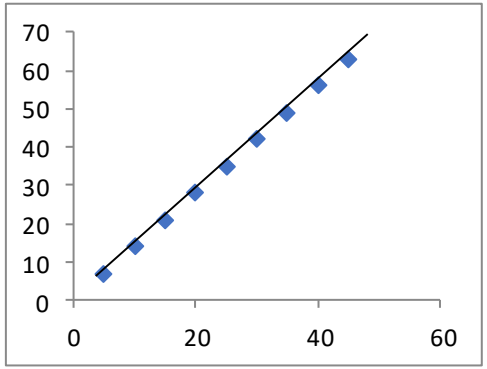
1. Υπάρχει σχέση;
2. Τι σχέση είναι αυτή; (π.χ. όταν αυξάνει το ένα, αυξάνει και το άλλο ή αντίστροφα;)
3. Ποιος είναι ο βαθμός αυτής της σχέσης;

ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ (CORRELATION COEFFICIENT) – **r**

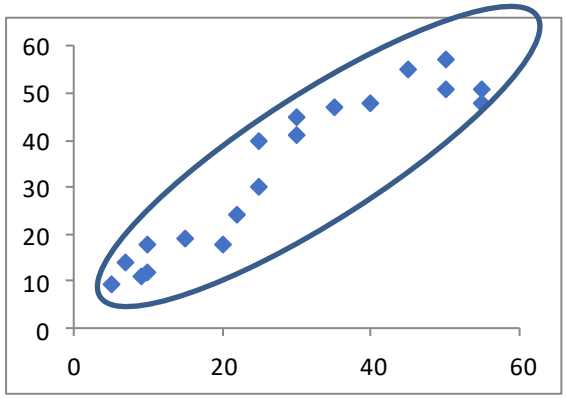
- Είναι ένας αριθμός ανάμεσα στο 0 και το 1. Όταν είναι μηδέν (0) δεν υπάρχει καμία σχέση.
- Όταν είναι κοντά στο 0 η σχέση είναι πολύ μικρή. Όσο ο αριθμός αυτός πλησιάζει στο 1, τόσο πιο ισχυρή είναι η σχέση αυτή.

(Υπάρχουν δύο ειδών τέτοιοι συντελεστές, του **Pearson** και του **Spearman**)

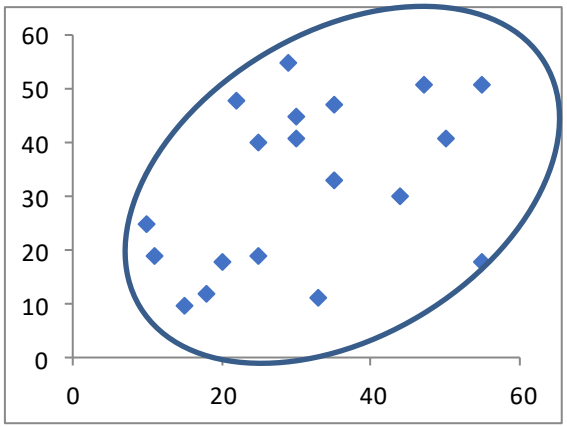
Ο συντελεστής συσχέτισης Pearson (r)



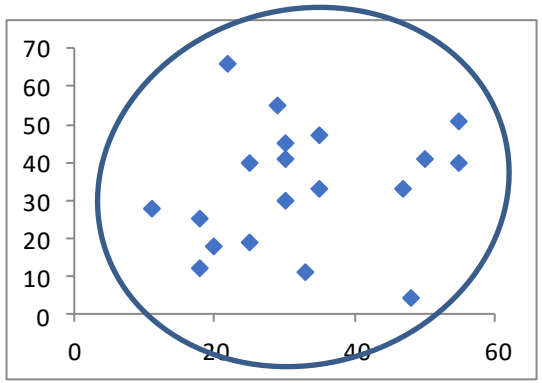
r=1



r=0,94



r=0,41



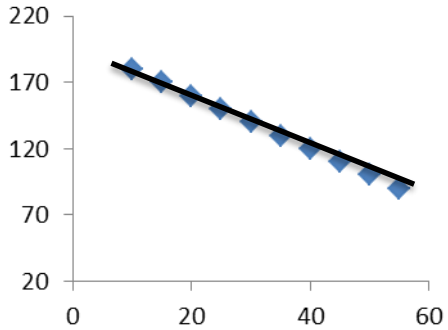
r=0,15

Θετική Συσχέτιση (r > 0)

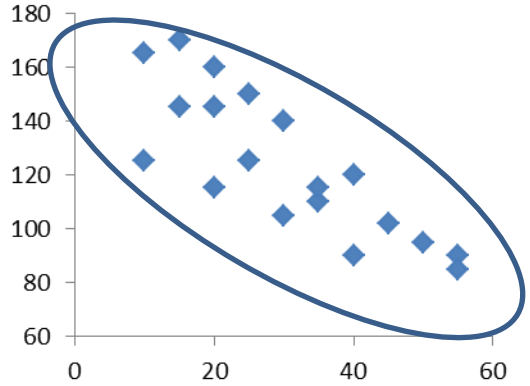
όταν αυξάνει το ένα χαρακτηριστικό αυξάνει και το άλλο

Η έλλειψη κατευθύνεται προς τα πάνω

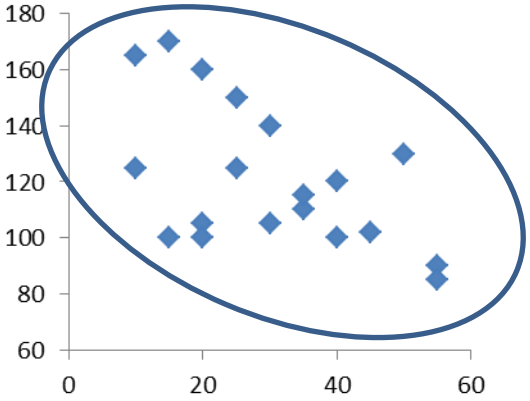
Ο συντελεστής συσχέτισης Pearson (r)



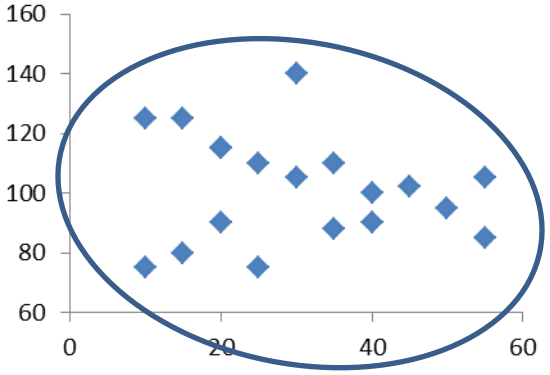
r= - 1



r= - 0,83



r= - 0,55



r= - 0,14

Αρνητική Συσχέτιση (r < 0)

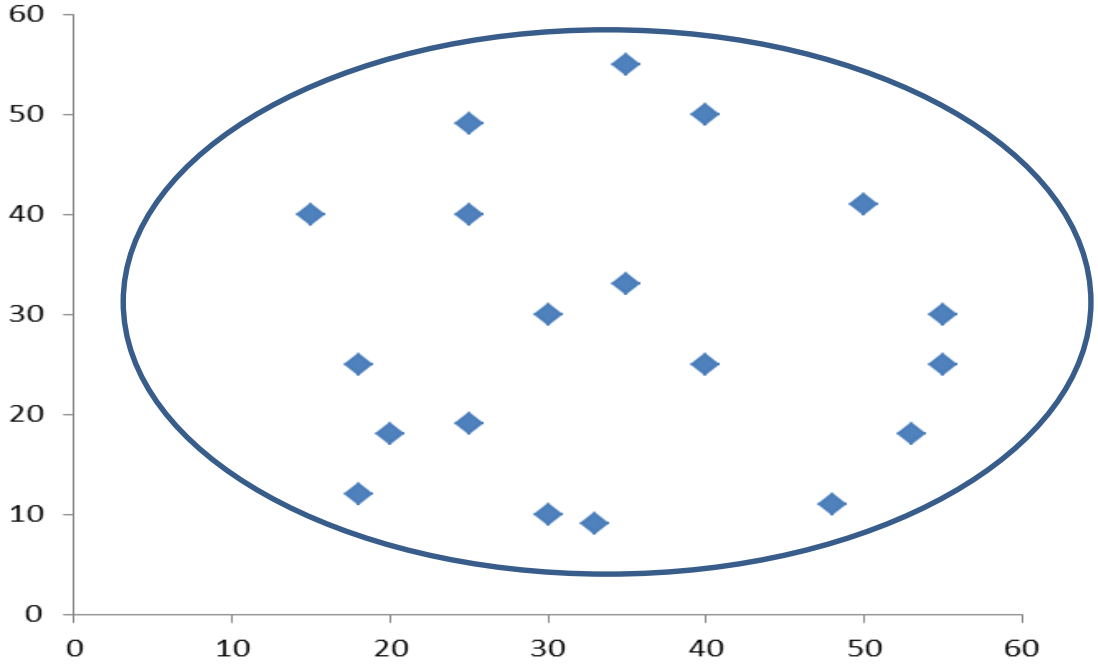
όταν αυξάνει το ένα χαρακτηριστικό μειώνεται το άλλο

Η έλλειψη κατευθύνεται προς τα κάτω

Ο συντελεστής συσχέτισης Pearson (r)

Μηδενική Συσχέτιση
r = 0

Τα χαρακτηριστικά ΔΕΝ έχουν καμία σχέση



Η έλλειψη κατευθύνεται παράλληλα με τον άξονα των Χ

Εμπειρικά μπορούμε να λέμε:

1) Οταν $0<r\leq 0,25$, τότε είναι πολύ μικρή συσχέτιση
2) Οταν $0,25\leq r\leq 0,50$, τότε έχουμε ελαφρά συσχέτιση
3) Οταν $0,50\leq r\leq 0,75$, η συσχέτιση είναι σχετικά ισχυρή και
4) Οταν $0,75\leq r\leq 1$, τότε η συσχέτιση είναι πολύ ισχυρή.

Αντίστοιχα μπορούμε να πούμε και για τις αρνητικές τιμές του r

Στο παράδειγμα 1

Στους 50 μαθητές του Λυκείου, Βρήκαμε για το ύψος και το βάρος

r (Pearson) = 0,860

στο δείγμα λοιπόν φαίνεται μια <u>ισχυρή σχέση</u> ανάμεσα στο ύψος και το βάρος των μαθητών

(φαίνεται να είναι πολύ καλή η σχέση ύψους βάρους, αλλά δεν είναι τέλεια, δηλ 1, γιατί το βάρος δεν εξαρτάται μόνο από το ύψος, αλλά και από την κατανάλωση φαγητού ή άλλους 10 παράγοντες!!)

Ερώτηση:

- Ο συντελεστής συσχέτισης **r** που βρήκαμε στο δείγμα μας ήταν τυχαίος;
- Δηλαδή, μήπως ο συντελεστής συσχέτισης σε ολόκληρο τον πληθυσμό είναι 0;

ή τελικά με το δείγμα μας επιβεβαιώνουμε ότι ΥΠΑΡΧΕΙ πράγματι κάποια σχέση μεταξύ των χαρακτηριστικών και στον πληθυσμό;

Έχει νόημα αυτή η ερώτηση;

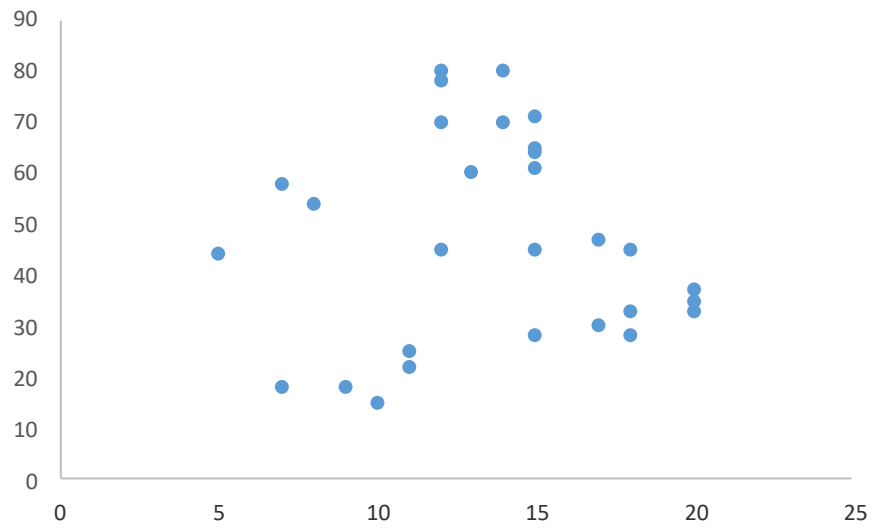
- Σε κάθε στατιστικό πρόβλημα, κάνουμε δειγματοληψία από ένα πληθυσμό
- Κατά τη δειγματοληψία, δεν είμαστε σίγουροι ποτέ ότι έχουμε πετύχει ακριβώς τις παραμέτρους του πληθυσμού που μας ενδιαφέρουν
- Π.χ. στη μέση τιμή, χρησιμοποιούμε το 95% Διάστημα Εμπιστοσύνης για να αυξήσουμε τη σιγουριά μας!

Παράδειγμα
(πως μπορούμε να βγάλουμε λάθος συμπεράσματα)

Ολόκληρος ο πληθυσμός (μόνο 30 άτομα για ευκολία!)

X	Y
5	44
7	18
7	58
8	54
9	18
10	15
12	80
12	78
12	70
14	70
18	45
15	71
17	47
17	30
11	25

X	Y
18	28
11	22
15	45
14	80
12	45
18	33
20	37
20	35
20	33
15	28
15	65
15	64
15	61
13	60

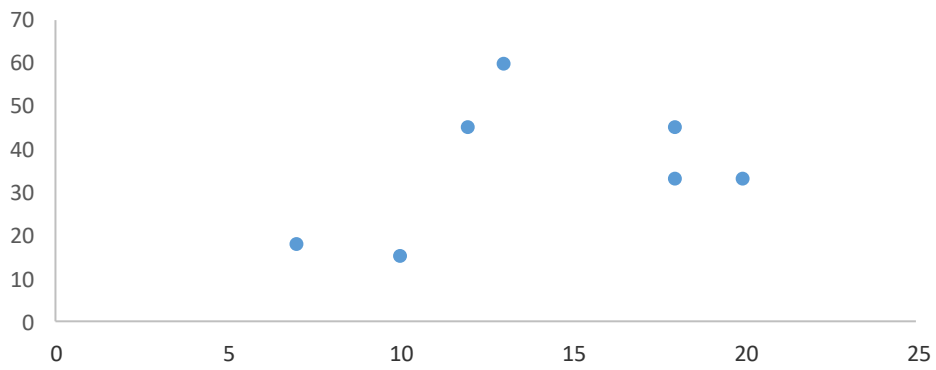


Ο συντελεστής συσχέτισης του πληθυσμού είναι 0

Ας πάρουμε τώρα ένα δείγμα 7 ατόμων (τα κίτρινα)

X	Y
5	44
7	18
7	58
8	54
9	18
10	15
12	80
12	78
12	70
14	70
18	45
15	71
17	47
17	30
11	25

X	Y
18	28
11	22
15	45
14	80
12	45
18	33
20	37
20	35
20	33
15	28
15	65
15	64
15	61
13	60



Υπολογίζω το συντελεστή συσχέτισης του δείγματος και βγήκε θετικός r=0,38

Πήρα λοιπόν ένα δείγμα από έναν πληθυσμό, για τον οποίο οι μεταβλητές Χ και Υ δεν είχαν ΚΑΜΙΑ σχέση, και έβγαλα στο δείγμα ότι έχουν συσχέτιση 0,38

Οπότε χρειαζόμαστε ένα p

- Που μας δείχνει τι πιθανότητα υπάρχει ο συντελεστής συσχέτισης σε ολόκληρο τον πληθυσμό να είναι μηδενικός
- Δηλαδή τι πιθανότητα έχει, να βρήκαμε εμείς κάποια σχέση στο δείγμα μας (μικρή ή μεγάλη), ενώ στον πληθυσμό να μην υπάρχει σχέση.
- Οπότε….

…για να βγάλουμε συμπέρασμα για ΟΛΟΚΛΗΡΟ τον πληθυσμό….

- Μαζί με τον συντελεστή συσχέτισης **r**, ψάχνουμε στην απάντηση για τον αριθμό **SIG (Τιμή p)**
- Συγκρίνουμε το **p** με το 0,05 (ή 5%)
- Αν **p<0,05** τότε οι ερωτήσεις μας έχουν σχέση μεταξύ τους
- αν **p>0,05** τότε οι ερωτήσεις δεν έχουν καμία σχέση και τα ευρήματα του Δείγματος ήταν τυχαία και όχι σημαντικά

Στο δείγμα με τα 7 άτομα

Το r=0,38 που βρήκαμε, αντιστοιχεί σε

$p=0,401 > 0,05$

(βρίσκεται με τη βοήθεια υπολογιστή)

Φυσικά το αποτέλεσμα ήταν τυχαίο, μη σημαντικό, όπως αναμενόταν άλλωστε, αφού πήραμε δείγμα από πληθυσμό που είχε μηδενική συσχέτιση.

Στο παράδειγμα με το ύψος/βάρος

Correlations

		Ύψος μαθητή (cm)	Βάρος μαθητή (kg)
Ύψος μαθητή (cm)	Pearson Correlation	1	,860
	Sig. (2-tailed)		**,000**
	N	50	50
Βάρος μαθητή (kg)	Pearson Correlation	,860	1
	Sig. (2-tailed)	,000	
	N	50	50

Το r (Pearson)

Τιμή p

<u>Συμπέρασμα</u>: Αφού p<<0,05 άρα υπάρχει μια θετική συσχέτιση ανάμεσα στο ύψος και το βάρος

Αυτό σημαίνει ότι όσο μεγαλώνει το ύψος, τόσο μεγαλώνει και το βάρος

Παράδειγμα 2

Ας πούμε ότι στο αρχείο μας μετρήσαμε και το βαθμό του Λυκείου των 50 μαθητών. Ας ελέγξουμε αν το ΥΨΟΣ έχει σχέση με τον ΒΑΘΜΟ που πήραν οι μαθητές στην Β Λυκείου

Βρήκαμε **r=-0.147** (το αρνητικό πρόσημο θα εξηγηθεί στη συνέχεια)

Το p(SIG) είναι 0,308 δηλαδή **p=0,308 > 0,05**

Αυτό σημαίνει πως **δεν** υπάρχει σχέση ανάμεσα στο Ύψος και τη Βαθμολογία, και το αποτέλεσμα που βρέθηκε ήταν τυχαίο, όπως αναμενόταν άλλωστε!

Θετική ή Αρνητική Συσχέτιση

- Πολλές φορές ο συντελεστής συσχέτισης μπορεί να είναι αρνητικός αριθμός.
- Τότε η συσχέτιση ανάμεσα στις ερωτήσεις λέγεται Αρνητική Συσχέτιση.
- Στην περίπτωση αυτή οι ερωτήσεις είναι αντιστρόφως ανάλογες, δηλαδή όταν αυξάνει η μία, ή άλλη γενικά φαίνεται να μειώνεται και αντίστροφα, όταν μειώνεται η μία, η άλλη γενικά αυξάνεται.

- Αν αξιολογούσαμε π.χ. το γεγονός ότι η συσχέτιση ύψους και βαθμολογίας ήταν αρνητικός αριθμός, αυτό θα σήμαινε ότι…
- Οι πιο ψηλοί τείνουν να παίρνουν μικρούς βαθμούς και οι πιο κοντοί παίρνουν μεγάλους βαθμούς!!!

- *(όμως το αποτέλεσμα έδωσε $p>0,05$ οπότε δεν αξιολογείται φυσικά)*

Από τι εξαρτάται το p;

- Όσο πιο μεγάλος είναι ο συντελεστής συσχέτισης που υπολογίζουμε στο δείγμα μας, τόσο πιο μικρό βγαίνει το p
- Όμως για να βγει το p σημαντικό ($p<0,05$), πόσος πρέπει να είναι ο συντελεστής που υπολογίσαμε;
- Αυτό εξαρτάται από το πόσο μεγάλο είναι το δείγμα μας δηλαδή από το ***n***

Πίνακας σημαντικότητας του r.

n	r-κρισιμο	n	r-κρισιμο
10	0.632	25	0,396
11	0.602	26	0,388
12	0.576	27	0,381
13	0.553	28	0,374
14	0.532	29	0,367
15	0.514	30	0,361
16	0.497	40	0,312
17	0.482	50	0,279
18	0.468	60	0,254
19	0.456	70	0,236
20	0.444	80	0,221
21	0.433	90	0,209
22	0.423	100	0,198
23	0.413	400	0,100
24	0.404	500	0,089

Ανάλογα με το μέγεθος δείγματος που έχουμε, αν βρούμε το r πιο μεγάλο από το r-κρίσιμο, τότε είμαστε βέβαιοι ότι το $p<0,05$, άρα ο συντελεστής συσχέτισης είναι ΣΗΜΑΝΤΙΚΟΣ

Ο πίνακας είναι περίληψη φυσικά γιατί βλέπουμε μόνο κάποια n, όχι όλες τις τιμές n, (πχ που είναι το n=35q)

Δηλαδή…

- Αν είχαμε 25 μετρήσεις, και βρούμε το r=0,45 τότε είναι σημαντικό (p<0,05), αφού είναι πάνω από το όριο που είναι 0,396 (από τον πίνακα)
- Αν είχαμε 40 μετρήσεις, και βρούμε το r=0,27 τότε ΔΕΝ είναι σημαντικό (p>0,05), αφού είναι κάτω από το όριο που είναι 0,312 (από τον πίνακα)

Ο συντελεστής συσχέτισης Spearman

- Έχει παρόμοιες ιδιότητες με αυτόν του Pearson
- Χρησιμοποιείται συχνά όταν έχουμε ποιοτικά διατάξιμα χαρακτηριστικά (ή απαντήσεις στην κλίμακα 1-2-3-4-5 – καθόλου, λίγο, μέτρια, πολύ, πάρα πολύ)
- αλλά μπορεί να χρησιμοποιηθεί και στα ποσοτικά, κυρίως σε περιπτώσεις που δεν έχουμε συμμετρικές κατανομές

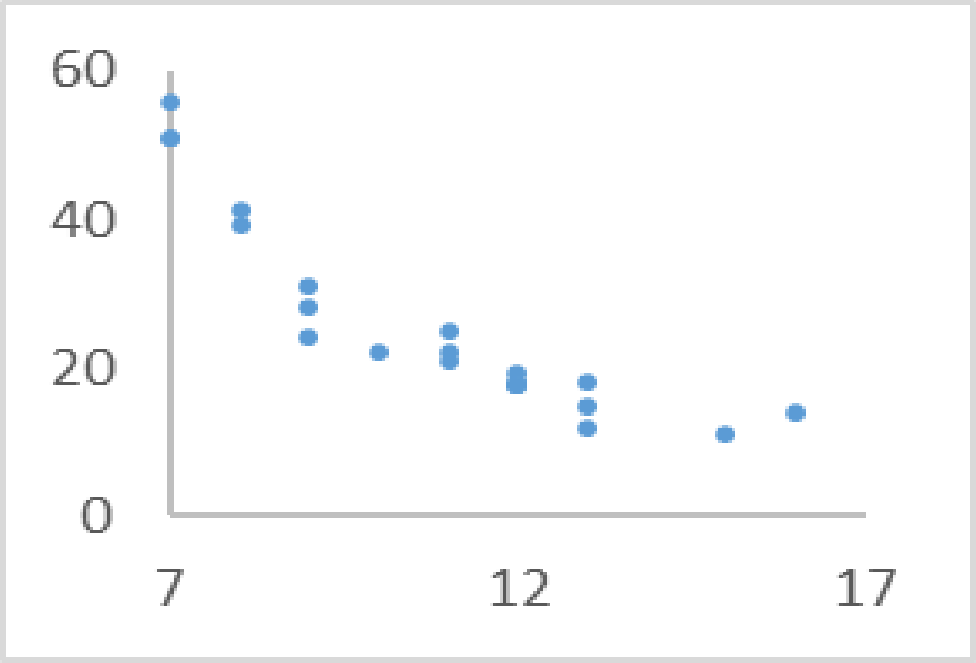
Ο συντελεστής συσχέτισης του Spearman στο παράδειγμα 1 (βάρος-ύψος)

Correlations

			Ύψος μαθητή (cm)	Βάρος μαθητή (kg)
Spearman's rho	Ύψος μαθητή (cm)	Correlation Coefficient	1,000	**,849**
		Sig. (2-tailed)	.	,000
		N	50	50
	Βάρος μαθητή (kg)	Correlation Coefficient	,849	1,000
		Sig. (2-tailed)	,000	.
		N	50	50

Όπως βλέπετε ο συντελεστής που βρέθηκε (0,849) είναι πολύ κοντά στον συντελεστή του Pearson (0,860) που είχαμε βρει πιο πρίν

Ερώτηση 1:



Σε μια έρευνα πήραμε το παραπάνω διάγραμμα. Πόσο πιστεύετε πως ήταν ο συντελεστής συσχέτισης του Pearson, που βρήκαμε;

Α) +0,7 Β) -0,9 Γ) 0,1 Δ) -0,25

Ερώτηση 2:

- Αν χρησιμοποιήσαμε 20 μετρήσεις στο προηγούμενο διάγραμμα, υπάρχει τελικά σχέση ανάμεσα στις δύο μεταβλητές που μελετήσαμε;

(χρησιμοποιήστε τον πίνακα με τα κρίσιμα r)

……

……