# ΜΗ ΠΑΡΑΜΕΤΡΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ

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

Σύμφωνα με Βαγενάς (2019) (σελ. 330 – 348)

1

## Γενικές έννοιες

- Στην αρχή των διαλέξεων επισημάνθηκε ότι, όταν τα αρχικά δεδομένα μιας έρευνας έχουν μετρηθεί στην **ονομαστική** ή στην **τακτική κλίμακα,** τότε η στατιστική τους ανάλυση πρέπει να γίνει με κάποια από τις μεθόδους της **μη παραμετρικής στατιστικής.**
- Ο λόγος είναι ότι για τις δύο αυτές κλίμακες μέτρησης **δεν ισχύουν** οι στατιστικές παραδοχές περί της **κανονικότητας της κατανομής** και περί της **ομοιογένειας της διασποράς**.

2

2

# Γενικές έννοιες … συνέχεια

- Επιπρόσθετα, στην ερευνητική πρακτική συμβαίνει μερικές φορές, τα αρχικά δεδομένα μιας έρευνας να έχουν μετρηθεί σε μία από τις δύο άλλες κλίμακες μέτρησης, δηλαδή στην **κλίμακα διαστήματος** ή στην **κλίμακα λόγου**, αλλά να υπάρχουν αμφιβολίες ως προς την ισχύ των δύο στατιστικών παραδοχών.
- Τότε και στις περιπτώσεις αυτές εφαρμόζεται η **μη παραμετρική στατιστική ανάλυση** με κατάληξη την αξιολόγηση της στατιστικής σημαντικότητας με τον κατάλληλο **μη παραμετρικό έλεγχο.**

3

3

# Γενικές έννοιες … συνέχεια

- Οι **μη παραμετρικές στατιστικές αναλύσεις** περιλαμβάνουν κατά κύριο λόγο μεθόδους: (α) <u>μη παραμετρικής συσχέτισης δύο μεταβλητών</u> και (β) μεθόδους <u>μη παραμετρικής σύγκρισης δύο διάμεσων</u>.
- Στο παρόν κεφάλαιο εξετάζονται οι εξής **συντελεστές συσχέτισης:**
  - **συντελεστής συνάφειας C** για δύο **k-κατηγορικές** μεταβλητές
  - **συντελεστής** Φ για δύο 2-κατηγορικές **(διχοτομικές)** μεταβλητές
  - **Συντελεστής Spearman ($r_s$)** για δύο τακτικές μεταβλητές (σειράς).

4

4

# Συντελεστής Συνάφειας C: Συσχέτιση 2 πολυκατηγορικών Μεταβλητών

- Ο **συντελεστής C** (contingency coefficient) είναι δείκτης συνάφειας 2 κατηγορικών **ονομαστικών (nominal) μεταβλητών** και βασίζεται στο στατιστικό $\chi^2$.
- Εφαρμόζεται σε **πίνακες συνάφειας r x c**, όπου r (rows) οι κατηγορίες της μιας μεταβλητής και c (columns) οι κατηγορίες της άλλης μεταβλητής.
- **Για να είναι έγκυρος** ο δείκτης C **πρέπει οι 2 μεταβλητές να είναι ονομαστικές**, δηλαδή να έχουν κατάταξη ονομαστικής κλίμακας. Επίσης καλό είναι ο αριθμός των κατηγοριών (k) σε κάθε μεταβλητή να είναι > 2, ο πίνακας συνάφειας να είναι συμμετρικός (δηλαδή να έχει ίσο αριθμό σειρών και στηλών, r = c) και το δείγμα επαρκώς μεγάλο (π.χ., N > 100).

5

5

- Ο συντελεστής συσχέτισης C είναι

$$C = \sqrt{\chi^2 /(\chi^2 + N)}$$

- Η στατιστική σημαντικότητά του C είναι αυτή του $\chi^2$ με βαθμούς ελευθερίας df = (r - 1)(c - 1), όπου N το δείγμα (αριθμός τιμών X, Y σε ζεύγη).
- Για να είναι **ο C συγκρίσιμος με τον r του Pearson**, γίνεται διόρθωση με βάση τη μεγίστη τιμή ($C_{max}$) που μπορεί να πάρει θεωρητικά:

$$C_{max} = \sqrt{(k-1)/k}$$

6

6

$$C_{max} = \sqrt{(k-1)/k}$$

- Η διόρθωση αυτή έχει **εφαρμογή μόνο σε συμμετρικούς πίνακες συνάφειας** (symmetric contingency tables), δηλαδή με ίδιο αριθμό κατηγοριών στις 2 μεταβλητές (c = r = k), και **δίνει τον C ως ποσοστό του $C_{max}$**:
- **διορθωμένος συντελεστής** (C') είναι $C' = C / C_{max}$

7

7

- Ο <u>δείκτης C έχει το μειονέκτημα</u> ότι ποτέ δεν φθάνει τις οριακές τιμές 0 ή 1 και αυτό βελτιώνεται μόνο σε διμεταβλητές με πολλές ονομαστικές κατηγορίες (π.χ., > 4 χ 4).
- Για παράδειγμα σε πίνακες 2x2 μπορεί να φθάσει **το πολύ 0.71**, ενώ σε πίνακες 4x4 **το πολύ 0.87**.
- Όμως, μετά από διόρθωση ως προς $C_{max}$ προσεγγίζει καλύτερα τα 2 όρια και είναι συγκρίσιμος με τον r του Pearson.

8

8

- Ο βαθμός συσχέτισης μεταξύ **2 ονομαστικών μεταβλητών** εκτιμάται εναλλακτικά με τον δείκτη

$$\text{Cramer V} = \sqrt{\chi^2 / \text{N}(\text{k}-1)}$$

- όπου **k** ο αριθμός σειρών ή στηλών (**όποιος είναι μικρότερος**). Αν, όμως, οι 2 μεταβλητές έχουν κατηγορίες σε σειρά (ordered categories), όπως π.χ., η τάξη Α, Β, Γ, Δ, με το επίπεδο $1^ο$, $2^ο$, $3^ο$, $4^ο$, τότε κατάλληλος είναι ο **Kendall-tau-b**, του οποίου εναλλακτική προέκταση σε 2-μεταβλητές με πολλές κατηγορίες σε σειρά είναι ο δείκτης **Spearman**.

9

9

# Παράδειγμα 15.7 - Συντελεστής C: ανάλυση του παραδείγματος 15.4

**Πίνακας 15.4** - Ανάλυση $\chi^2$ σε διπλή ταξινόμηση 3x3: Έλεγχος διαφοράς μεταξύ πραγματικής (f) και αναμενόμενης (θεωρητικής, F) κατανομής

| Κατηγορίες | Συχνότητες | | Αποκλίσεις | Τετράγωνα | Πηλίκα |
|---|---|---|---|---|---|
| j | f | F | f - F | $(f-F)^2$ | $(f-F)^2/F$ |
| 1η (1, 1) | 100 | 111 | -11 | 121 | 1.09 |
| 2η (1, 2) | 150 | 120 | 30 | 900 | 7.50 |
| 3η (1, 3) | 50 | 69 | -19 | 361 | 5.23 |
| 4η (2, 1) | 50 | 92.5 | -42.5 | 1806 | 19.53 |
| 5η (2, 2) | 120 | 100 | 20 | 400 | 4.00 |
| 6η (2, 3) | 80 | 57.5 | 22.5 | 506 | 8.80 |
| 7η (3, 1) | 220 | 166.5 | 53.5 | 2862 | 17.19 |
| 8η (3, 2) | 130 | 180 | -50 | 2500 | 13.89 |
| 9η (3, 3) | 100 | 103.5 | -3.5 | 12.25 | 0.12 |
| Σ = | 1000 = N | 1000 = N | 0 | $\chi^2$ = | 77.35 |

f = πραγματικές συχνότητες, F = θεωρητικές συχνότητες (από τιν Πίν. 15.3)

Στο Παράδειγμα 15.4 εξετάστηκε στο επίπεδο πιθανότητας α = 0.01 αν η κατανομή μιας πολοκατηγορικής ονομαστικής 2-μεταβλητής (X, Y) σε δείγμα N = 1000 αθλητών απείχε σημαντικά από την αντίστοιχη θεωρητική κατανομή.
Οι μεταβλητές X και Y είχαν 3 μόνο κατηγορίες η κάθε μία και έτσι η ταξινόμηση έγινε σε πίνακα συνάφειας 3x3 .

10

**Πίνακας 15.4** - Ανάλυση $\chi^2$ σε διπλή ταξινόμηση 3x3: Έλεγχος διαφοράς μεταξύ πραγματικής (f) και αναμενόμενης (θεωρητικής, F) κατανομής

| Κατηγορίες | Συχνότητες | | Αποκλίσεις | Τετράγωνα | Πηλίκα |
|---|---|---|---|---|---|
| j | f | F | f - F | $(f-F)^2$ | $(f-F)^2/F$ |
| 1η (1, 1) | 100 | 111 | -11 | 121 | 1.09 |
| 2η (1, 2) | 150 | 120 | 30 | 900 | 7.50 |
| 3η (1, 3) | 50 | 69 | -19 | 361 | 5.23 |
| 4η (2, 1) | 50 | 92.5 | -42.5 | 1806 | 19.53 |
| 5η (2, 2) | 120 | 100 | 20 | 400 | 4.00 |
| 6η (2, 3) | 80 | 57.5 | 22.5 | 506 | 8.80 |
| 7η (3, 1) | 220 | 166.5 | 53.5 | 2862 | 17.19 |
| 8η (3, 2) | 130 | 180 | -50 | 2500 | 13.89 |
| 9η (3, 3) | 100 | 103.5 | -3.5 | 12.25 | 0.12 |
| Σ = | 1000 = N | 1000 = N | 0 | $\chi^2$ = | 77.35 |

f = πραγματικές συχνότητες, F = θεωρητικές συχνότητες (από τιν Πίν. 15.3)

Η ανάλυση έδωσε $\chi^2 = 77.35$ στατιστικώς σημαντικό στο $\alpha = 0.01$. Έτσι ο βαθμός της συνάφειας των 2 τρικατηγορικών μεταβλητών Χ & Υ με βάση **τον συντελεστή C** είναι:

$$\mathbf{C = \sqrt{77.35/(77.35+1000)} = \sqrt{0.0718} = 0.27.}$$

Με k = 3, η <u>μεγίστη τιμή C</u> είναι

$$\mathbf{C_{max} = \sqrt{(3-1)/3} = \sqrt{(2)/3} = 0.82}$$

και ο <u>διορθωμένος C</u> είναι

$$\mathbf{C' = C / C_{max} = 0.27/0.82 = 0.33}$$

(συγκρίσιμος με τον r του Pearson).

11

- Ο συντελεστής C = 0.27 δείχνει μικρή συσχέτιση (κλίμακα Cohen, 1988, σελ. 222): **0.10 = μικρή, 0.29 = μέτρια, 0.45 = μεγάλη**).
- Το ίδιο δείχνει και ο Cramer $V = \sqrt{\chi^2/N(k-1)} = \sqrt{77.35/1000(3-1)} = \sqrt{77.35/2000} = 0.20$ (βλ. Παράδειγμα 15.4).
- Η στατιστική σημαντικότητα του C = 0.27 είναι ίδια με αυτήν του $\chi^2 = 77.351$.
- **Συμπέρασμα**: Η συσχέτιση μεταξύ της μεταβλητής άθλημα (Χ) και της μεταβλητής σωματικό μέλος τραυματισμού (Υ) είναι στατιστικώς σημαντική (p<0.01) αλλά ουσιαστικώς μικρή (0.27).

12

12

# Συντελεστής Φ: Συσχέτιση 2 Δικατηγορικών Μεταβλητών

- Ο **συντελεστής συσχέτισης Φ** (phi correlation coefficient) αποτελεί υπο-περίπτωση του συντελεστή C, **εφαρμόζεται σε πίνακες συνάφειας 2x2** και ποσοτικοποιεί τη σχέση **δύο 2-κατηγορικών ονομαστικών μεταβλητών**.
- **Βασική προϋπόθεση της εγκυρότητας του συντελεστή Φ** είναι κάθε διχοτομική μεταβλητή να είναι **αληθής διχοτομική** (true dichotomy): η κατάταξη σε 2 μόνο κατηγορικές τιμές να είναι πρωτογενής και γνήσια.

13

13

# Συντελεστής Φ: Συσχέτιση 2 Δικατηγορικών Μεταβλητών

- Ο συντελεστής συσχέτισης Φ είναι

$$\Phi = \sqrt{\chi^2/\mathrm{N}}$$

- Η στατιστική σημαντικότητα του δείκτη Φ είναι ίδια με του $\chi^2$.
- **Για την περαιτέρω διερεύνηση Πινάκων συνάφειας 2x2** μπορούμε να χρησιμοποιήσουμε τους δείκτες Risk Ratio και Odds Ratio (θα μιλήσουμε παρακάτω).

14

14

# Π.χ., 15.8 - Συντελεστής Φ: ανάλυση π.χ., 15.5.

- Στο Παράδειγμα 15.5 εξετάστηκε στο $\alpha = 0.01$ αν η πραγματική (observed) κατανομή της 2-κατηγορικής 2-μεταβλητής X (φύλο) και Y (τραυματισμός ή όχι) σε δείγμα $N = 100$ αθλητών διέφερε σημαντικά από τη θεωρητική (αναμενόμενη, expected).
- Η τιμή $\chi^2 = 36.95$ αποδείχτηκε στατιστικώς σημαντική και ερμηνεύτηκε ως σημαντική διαφορά μεταξύ της πραγματικής και της αναμενόμενης κατανομής και επομένως ως σημαντική εξάρτηση μεταξύ φύλου και τραυματισμού.
- Συντελεστής $\Phi = \sqrt{36.95/100} = 0.6079 \approx 0.61$ ή εναλλακτικά
- <u>Η στατιστική σημαντικότητα του $\Phi = 0.61$ είναι αυτή του $\chi^2 = 36.95$, δηλαδή στο επίπεδο πιθανότητας $\alpha = 0.01$</u>.
- Η συσχέτιση μεταξύ φύλου (X) και εμφάνισης τραυματισμού ή όχι (Y) είναι στατιστικώς σημαντική ($p < 0.01$) και μεγάλη (0.61 ).

15

# Συντελεστής Spearman ($r_s$): Συσχέτιση 2 Διατακτικών Μεταβλητών

- Ο **δείκτης Spearman** εκτιμά τον βαθμό συσχέτισης είτε
  1. μεταξύ **2 διατακτικών (ordinal) μεταβλητών** (δεδομένα **σειράς**, **ranks**), είτε
  2. (ii) <u>μεταξύ 2 ποσοτικών (quantitative) μεταβλητών που δεν πληρούν τις προϋποθέσεις της παραμετρικής συσχέτισης Pearson</u>.
- Συχνά, συμβαίνει να παρατηρούνται **ίσες σειρές (tied ranks)** σε δύο ή και περισσότερα υποκείμενα (subjects, cases).

16

# Συντελεστής Spearman ($r_s$): Συσχέτιση 2 Διατακτικών Μεταβλητών

- Όταν δεν υπάρχουν ίσες σειρές ο Spearman είναι

$$r_s = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

- Όταν υπάρχουν ίσες σειρές (ισοβαθμίες, ties, T) ο Spearman είναι

$$r_s = \frac{\Pi - \sum T\chi - \sum Ty - \sum d^2}{\sqrt{(\Pi - 2*\sum T\chi)*(\Pi - 2*\sum Ty)}}$$

όπου **d** η διαφορά σειρών κατά-ζεύγη, **N** το δείγμα (π.χ., άτομα), **Π** = ($N^3$- N)/6, **τ** = ο αριθμός ίσων σειρών, **T** = (τ3 - τ)/12 η διόρθωση για κάθε ισοβαθμία, **∑T** άθροισμα των T για κάθε μεταβλητή (X, Y).

17

## Ο Spearman παίρνει τιμές από +1 (τέλεια θετική συσχέτιση) έως -1 (τέλεια αρνητική συσχέτιση).

- Για N = 2, ο $r_s$ μπορεί να είναι είτε +1, είτε -1, ενώ για N = 3, οι δυνατές τιμές του $r_s$ είναι -1 , -0.50, +0.50, +1.
- Ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή συσχέτισης Spearman γίνεται με προκαθορισμένη πιθανότητα α:
  - για N από 1 έως 30 άμεσα με βάση κρίσιμες τιμές $r_s$ (Πίν. 21.Κ) &
  - για N > 10 έμμεσα με μετασχηματισμό του σε t

$$t = r_s\sqrt{N-2} \,/\, \sqrt{1-r_s^{\,2}}$$

- και σύγκριση του t με το κρίσιμο $t_c$, που αντιστοιχεί σε df = N - 2 βαθμούς ελευθερίας (Πίν. 21.Δ).

18

18

- Ο έλεγχος της στατιστικής σημαντικότητας του Spearman είναι όμοιος με αυτόν του Pearson.
- Ο Spearman είναι 91% πιο ισχυρός ως προς τον Pearson όταν δεν υπάρχουν ίσες σειρές.

| Πλήθος | Κριτής 1 | Κριτής2 | Διαφορά | Τετράγωνα |
|---|---|---|---|---|
| N | $R_1$ | $R_2$ | $R_1 - R_2 = d$ | $d^2$ |
| 1 | 3 | 1 | 2 | 4 |
| 2 | 2 | 3 | -1 | 1 |
| 3 | 6 | 6 | 0 | 0 |
| 4 | 4 | 4 | 0 | 0 |
| 5 | 5 | 5 | 0 | 0 |
| 6 | 10 | 9 | 1 | 1 |
| 7 | 9 | 10 | -1 | 1 |
| 8 | 8 | 7 | 1 | 1 |
| 9 | 7 | 8 | -1 | 1 |
| 10 | 1 | 2 | -1 | 1 |
| Σ = | - | - | 0 | $\Sigma d^2 = 10$ |

Οι σειρές $R_1$ & $R_2$ αποτελούν εκτιμήσεις των 2 κριτών.

19

## Π.χ., 15.9 - Συσχέτιση Spearman: χωρίς ίσες σειρές.


| Πλήθος | Κριτής 1 | Κριτής2 | Διαφορά | Τετράγωνα |
|---|---|---|---|---|
| N | $R_1$ | $R_2$ | $R_1 - R_2 = d$ | $d^2$ |
| 1 | 3 | 1 | 2 | 4 |
| 2 | 2 | 3 | -1 | 1 |
| 3 | 6 | 6 | 0 | 0 |
| 4 | 4 | 4 | 0 | 0 |
| 5 | 5 | 5 | 0 | 0 |
| 6 | 10 | 9 | 1 | 1 |
| 7 | 9 | 10 | -1 | 1 |
| 8 | 8 | 7 | 1 | 1 |
| 9 | 7 | 8 | -1 | 1 |
| 10 | 1 | 2 | -1 | 1 |
| Σ = | - | - | 0 | $\Sigma d^2 = 10$ |

Οι σειρές $R_1$ & $R_2$ αποτελούν εκτιμήσεις των 2 κριτών.


- Μια ομάδα 10 ατόμων αξιολογήθηκε από 2 διαφορετικούς κριτές σε μια κινητική δεξιότητα και πήρε, αντίστοιχα, τις σειρές $R_1$ και $R_2$.
- Ζητήθηκε να ελεγχθεί αν η συνάφεια μεταξύ των 2 σειρών αξιολόγησης είναι στατιστικά σημαντική στο επίπεδο πιθανότητας $\alpha = 0.01$.
- Οι διαφορές των σειρών κατά ζεύγη και τα τετράγωνά τους δίνονται στις στήλες d και $d^2$, αντίστοιχα.
- $\sum d = 0$ και $\sum d^2 = 10$

20

20

| Πλήθος | Κριτής 1 | Κριτής2 | Διαφορά | Τετράγωνα |
|---|---|---|---|---|
| N | $R_1$ | $R_2$ | $R_1 - R_2 = d$ | $d^2$ |
| 1 | 3 | 1 | 2 | 4 |
| 2 | 2 | 3 | -1 | 1 |
| 3 | 6 | 6 | 0 | 0 |
| 4 | 4 | 4 | 0 | 0 |
| 5 | 5 | 5 | 0 | 0 |
| 6 | 10 | 9 | 1 | 1 |
| 7 | 9 | 10 | -1 | 1 |
| 8 | 8 | 7 | 1 | 1 |
| 9 | 7 | 8 | -1 | 1 |
| 10 | 1 | 2 | -1 | 1 |
| Σ = | - | - | 0 | $\Sigma d^2 = 10$ |

Οι σειρές $R_1$ & $R_2$ αποτελούν εκτιμήσεις των 2 κριτών.

Ο συντελεστής συσχέτισης Spearman των σειρών $R_1$ & $R_2$ είναι:

$$r_s = 1 - \frac{6\sum d^2}{N(N^2-1)} =$$

$$= 1 - \frac{6(10)}{10(10^2-1)} = 1 - 0.0606 =$$

$$0.939$$

21

21

Πίνακας 21.Κ – Κρίσιμες τιμές τού συντελεστή συσχέτισης $r_s$ του **Spearman**

| Έλεγχος μονόπλευρος → | | επίπεδο στατιστικής σημαντικότητας 0.05 | 0.025 | 0.01 | 0.005 |
|---|---|---|---|---|---|
| N↓ | 5 | 0.900 | 1.000 | 1.000 | -- |
| | 6 | 0.829 | 0.886 | 0.943 | 1.000 |
| | 7 | 0.714 | 0.786 | 0.893 | 0.929 |
| | 8 | 0.643 | 0.738 | 0.833 | 0.881 |
| | 9 | 0.600 | 0.683 | 0.783 | 0.833 |
| | 10 | 0.564 | 0.648 | 0.746 | $\boxed{0.824}$ |
| | 11 | 0.520 | 0.620 | 0.745 | 0.815 |
| | 12 | 0.506 | 0.591 | 0.712 | 0.777 |
| | 13 | 0.475 | 0.566 | 0.671 | 0.744 |
| | 14 | 0.546 | 0.544 | 0.645 | 0.715 |
| | 15 | 0.440 | 0.524 | 0.622 | 0.688 |
| | 16 | 0.425 | 0.506 | 0.601 | 0.665 |
| | 17 | 0.411 | 0.490 | 0.582 | 0.644 |
| | 18 | 0.399 | 0.475 | 0.564 | 0.625 |
| | 19 | 0.388 | 0.462 | 0.548 | 0.607 |
| | 20 | 0.377 | 0.450 | 0.534 | 0.591 |
| | 21 | 0.368 | 0.439 | 0.520 | 0.576 |
| | 22 | 0.359 | 0.428 | 0.509 | 0.562 |
| | 23 | 0.351 | 0.418 | 0.496 | 0.549 |
| | 24 | 0.343 | 0.409 | 0.485 | 0.537 |
| | 25 | 0.336 | 0.400 | 0.475 | 0.526 |
| | 26 | 0.329 | 0.392 | 0.465 | 0.515 |
| | 27 | 0.323 | 0.384 | 0.456 | 0.505 |
| | 28 | 0.317 | 0.377 | 0.448 | 0.496 |
| | 29 | 0.311 | 0.370 | 0.440 | 0.487 |
| | 30 | 0.306 | 0.360 | 0.432 | 0.478 |
| δίπλευρος → | | 0.10 | 0.05 | 0.02 | 0.01 |

- Ο συντελεστής συσχέτισης Spearman των σειρών $R_1$ & $R_2$ είναι:

$$r_s = 0.939$$

- Με N = 10 και δίπλευρο έλεγχο στο α = 0.01 (Πίν. 21.Κ) η κρίσιμη τιμή Spearman είναι 0.824, οπότε, επειδή 0.939 > 0.824, η συσχέτιση είναι στατιστικά σημαντική.
- Ως μέγεθος είναι πολύ μεγάλη (Cohen, 1988, 0.50 = μεγάλη συσχέτιση.

22

22

# Π.χ., 15.10 - Συσχέτιση Spearman με ίσες σειρές (ties)

**Πίνακας 15.8** - Υπολογισμός συντελεστή συσχέτισης Spearman
(2 δευτερογενώς διατακτικές μεταβλητές: χρόνος συμμετοχής & πόντοι)

| Πλήθος | Χρόνος | Πόντοι | Σειρές (ranks) | | Διαφορά | Τετράγωνα |
|---|---|---|---|---|---|---|
| N | X | Y | $R_1$ | $R_2$ | $R_1 - R_2 = d$ | $d^2$ |
| 1 | 20 | 16 | 1 | 1 | 0 | 0 |
| 2 | 18 | 12 | 3 | 4 | -1 | 1 |
| 3 | 18 | 10 | 3 | 5 | -2 | 4 |
| 4 | 18 | 14 | 3 | 2.5 | 0.5 | 0.25 |
| 5 | 12 | 6 | 5 | 8.5 | -3.5 | 12.25 |
| 6 | 4 | 14 | 6.5 | 2.5 | 4 | 16 |
| 7 | 4 | 8 | 6.5 | 7 | -0.5 | 0.25 |
| 8 | 3 | 9 | 8 | 6 | 2 | 4 |
| 9 | 2 | 6 | 9 | 8.5 | 0.5 | 0.25 |
| 10 | 1 | 5 | 10 | 10 | 0 | 0 |
| Σ = | 100 | 100 | -- | -- | 0 | $\Sigma d^2 = 38$ |

Οι τιμές $R_1$ & $R_2$ προέκυψαν από μετασχηματισμό των τιμών X & Y, σε σειρές.

- Σε αγώνα μπάσκετ διαπιστώθηκε ότι οι 10 παίχτες της μιας ομάδας είχαν τους συνολικούς χρόνους συμμετοχής (Χ) και τους πόντους (Υ) που δίνονται στον Πίν. 15.8.
- Τίθεται το **ερευνητικό ερώτημα** αν η σειριακή συνάφεια μεταξύ χρόνου συμμετοχής και πόντων είναι στατιστικά σημαντική στο επίπεδο πιθανότητας $a = 0.05$.

23

23

# Π.χ., 15.10 - Συσχέτιση Spearman με ίσες σειρές (ties)

**Πίνακας 15.8** - Υπολογισμός συντελεστή συσχέτισης Spearman
(2 δευτερογενώς διατακτικές μεταβλητές: χρόνος συμμετοχής & πόντοι)

| Πλήθος | Χρόνος | Πόντοι | Σειρές (ranks) | | Διαφορά | Τετράγωνα |
|---|---|---|---|---|---|---|
| N | X | Y | $R_1$ | $R_2$ | $R_1 - R_2 = d$ | $d^2$ |
| 1 | 20 | 16 | 1 | 1 | 0 | 0 |
| 2 | 18 | 12 | 3 | 4 | -1 | 1 |
| 3 | 18 | 10 | 3 | 5 | -2 | 4 |
| 4 | 18 | 14 | 3 | 2.5 | 0.5 | 0.25 |
| 5 | 12 | 6 | 5 | 8.5 | -3.5 | 12.25 |
| 6 | 4 | 14 | 6.5 | 2.5 | 4 | 16 |
| 7 | 4 | 8 | 6.5 | 7 | -0.5 | 0.25 |
| 8 | 3 | 9 | 8 | 6 | 2 | 4 |
| 9 | 2 | 6 | 9 | 8.5 | 0.5 | 0.25 |
| 10 | 1 | 5 | 10 | 10 | 0 | 0 |
| Σ = | 100 | 100 | -- | -- | 0 | $\Sigma d^2 = 38$ |

Οι τιμές $R_1$ & $R_2$ προέκυψαν από μετασχηματισμό των τιμών X & Y, σε σειρές.

- Οι τιμές Χ και Υ μετασχηματίστηκαν σε σειρές $R_1$ και $R_2$ αντίστοιχα, που με τα ενδιάμεσα στατιστικά, δίνονται στον Πίν. 15.8.
- Σε κάθε μεταβλητή, όπου οι τιμές ήταν ίσες δόθηκε η μέση σειρά. Για παράδειγμα, οι παίχτες 2, 3 & 4 είχαν ίδιο χρόνο συμμετοχής (18 min) και κάλυψαν τις θέσεις 2, 3 & 4 με τη μέση σειρά τους $(2+3+4)/3 = 3$.

24

24

| Πλήθος | Χρόνος | Πόντοι | Σειρές (ranks) | | Διαφορά | Τετράγωνα |
|---|---|---|---|---|---|---|
| N | X | Y | $R_1$ | $R_2$ | $R_1 - R_2 = d$ | $d^2$ |
| 1 | 20 | 16 | 1 | 1 | 0 | 0 |
| 2 | 18 | 12 | 3 | 4 | -1 | 1 |
| 3 | 18 | 10 | 3 | 5 | -2 | 4 |
| 4 | 18 | 14 | 3 | 2.5 | 0.5 | 0.25 |
| 5 | 12 | 6 | 5 | 8.5 | -3.5 | 12.25 |
| 6 | 4 | 14 | 6.5 | 2.5 | 4 | 16 |
| 7 | 4 | 8 | 6.5 | 7 | -0.5 | 0.25 |
| 8 | 3 | 9 | 8 | 6 | 2 | 4 |
| 9 | 2 | 6 | 9 | 8.5 | 0.5 | 0.25 |
| 10 | 1 | 5 | 10 | 10 | 0 | 0 |
| Σ = | 100 | 100 | -- | -- | 0 | $\Sigma d^2 = 38$ |

Οι τιμές $R_1$ & $R_2$ προέκυψαν από μετασχηματισμό των τιμών X & Y, σε σειρές.

- Στη μεταβλητή **X έχουμε 2 τιμές με ισοβαθμίες**: την 18 με τ = 3 επαναλήψεις και την 4 με τ = 2 επαναλήψεις, οπότε $\Sigma T_X = (3^3 - 3) / 12 + (2^3 - 2) / 12 = 2.5$.
- Στη μεταβλητή Y έχουμε 2 τιμές με ισοβαθμίες: την 14 με τ = 2 επαναλήψεις και την 6 με τ = 2 επαναλήψεις, οπότε $\Sigma T_Y = (23 - 2) /12 + (23 - 2) /12 = 1$.
- Με **Π** $= (N^3 - N) / 6 = (10^3 - 10) / 6 =$ **165**, **$\Sigma T_X = 2.5$** και **$\Sigma T_Y = 1$**,
- ο συντελεστής συσχέτισης Spearman των σειρών $R_1$ και $R_2$ είναι

$$r_s = \frac{\Pi - \Sigma T\chi - \Sigma T y - \Sigma d^2}{\sqrt{(\Pi - 2 * \Sigma T\chi) * (\Pi - 2 * \Sigma Ty)}}$$

$$r_s = \frac{\Pi - \Sigma T\chi - \Sigma T y - \Sigma d^2}{\sqrt{(\Pi - 2 * \Sigma T\chi) * (\Pi - 2 * \Sigma Ty)}} = \frac{165 - 2.5 - 1 - 38}{\sqrt{(165 - 2 * 2.5) * (165 - 2 * 1)}} = \frac{123.50}{161.49} = 0.765$$






| Πλήθος | Χρόνος | Πόντοι | Σειρές (ranks) | | Διαφορά | Τετράγωνα |
|---|---|---|---|---|---|---|
| N | X | Y | $R_1$ | $R_2$ | $R_1 - R_2 = d$ | $d^2$ |
| 1 | 20 | 16 | 1 | 1 | 0 | 0 |
| 2 | 18 | 12 | 3 | 4 | -1 | 1 |
| 3 | 18 | 10 | 3 | 5 | -2 | 4 |
| 4 | 18 | 14 | 3 | 2.5 | 0.5 | 0.25 |
| 5 | 12 | 6 | 5 | 8.5 | -3.5 | 12.25 |
| 6 | 4 | 14 | 6.5 | 2.5 | 4 | 16 |
| 7 | 4 | 8 | 6.5 | 7 | -0.5 | 0.25 |
| 8 | 3 | 9 | 8 | 6 | 2 | 4 |
| 9 | 2 | 6 | 9 | 8.5 | 0.5 | 0.25 |
| 10 | 1 | 5 | 10 | 10 | 0 | 0 |
| Σ = | 100 | 100 | -- | -- | 0 | $\Sigma d^2 = 38$ |

Οι τιμές $R_1$ & $R_2$ προέκυψαν από μετασχηματισμό των τιμών X & Y, σε σειρές.


- Με N = 10 και δίπλευρο έλεγχο στο α = 0.05 (Πίν 21.Κ) η κρίσιμη τιμή του $r_s$ είναι 0.648, οπότε, επειδή $0.765 > 0.648$, η συσχέτιση είναι στατιστικώς σημαντική.
- Ως μέγεθος είναι μεγάλη (> 0.50, Cohen, 1988).

**νακας 21.Κ** – Κρίσιμες τιμές τού συντελεστή συσχέτισης $r_s$ του **Spearman**

| Έλεγχος μονόπλευρος → | | επίπεδο στατιστικής σημαντικότητας 0.05 | 0.025 | 0.01 | 0.005 |
|---|---|---|---|---|---|
| N ↓ | 5 | 0.900 | 1.000 | 1.000 | -- |
| | 6 | 0.829 | 0.886 | 0.943 | 1.000 |
| | 7 | 0.714 | 0.786 | 0.893 | 0.929 |
| | 8 | 0.643 | 0.738 | 0.833 | 0.881 |
| | 9 | 0.600 | 0.683 | 0.783 | 0.833 |
| | 10 | 0.564 | $\boxed{0.648}$ | 0.746 | 0.824 |
| | 11 | 0.520 | 0.620 | 0.745 | 0.815 |
| | 12 | 0.506 | 0.591 | 0.712 | 0.777 |
| | 13 | 0.475 | 0.566 | 0.671 | 0.744 |
| | 14 | 0.546 | 0.544 | 0.645 | 0.715 |
| | 15 | 0.440 | 0.524 | 0.622 | 0.688 |
| | 16 | 0.425 | 0.506 | 0.601 | 0.665 |
| | 17 | 0.411 | 0.490 | 0.582 | 0.644 |
| | 18 | 0.399 | 0.475 | 0.564 | 0.625 |
| | 19 | 0.388 | 0.462 | 0.548 | 0.607 |
| | 20 | 0.377 | 0.450 | 0.534 | 0.591 |
| | 21 | 0.368 | 0.439 | 0.520 | 0.576 |
| | 22 | 0.359 | 0.428 | 0.509 | 0.562 |
| | 23 | 0.351 | 0.418 | 0.496 | 0.549 |
| | 24 | 0.343 | 0.409 | 0.485 | 0.537 |
| | 25 | 0.336 | 0.400 | 0.475 | 0.526 |
| | 26 | 0.329 | 0.392 | 0.465 | 0.515 |
| | 27 | 0.323 | 0.384 | 0.456 | 0.505 |
| | 28 | 0.317 | 0.377 | 0.448 | 0.496 |
| | 29 | 0.311 | 0.370 | 0.440 | 0.487 |
| | 30 | 0.306 | 0.360 | 0.432 | 0.478 |
| δίπλευρος → | | 0.10 | 0.05 | 0.02 | 0.01 |

## Πίνακες Συνάφειας 2 x 2: **Λόγος κινδύνου** (Risk Ratio) και **Λόγος σχετικής πιθανότητας** (Odds Ratio)

- Στα παραδείγματα 15.5 και 15.8 είδαμε το στατιστικό $\chi^2$ και το δείκτη $\Phi = \sqrt{\chi^2/\mathrm{N}}$ για τη διερεύνηση της συνάφειας μεταξύ 2 δι-κατηγορικών μεταβλητών, στο πλαίσιο της ανάλυσης ενός 2x2 πίνακα συνάφειας (2x2 contingency table).
- Οι πίνακες συνάφειας 2x2 χρησιμοποιούνται σε διάφορους τομείς έρευνας, αλλά κυρίως σε **πειραματικές έρευνες προοπτικής** (prospective studies) και σε **επιδημιολογικές έρευνες αναδρομικές** (retrospective studies), στις οποίες εξετάζεται η σχέση μεταξύ ενός υποθετικού «αιτίου» (π.χ. κάπνισμα) και της εμφάνισης μιας ασθένειας (π.χ. υπέρταση).

27

27

Δύο βασικά στατιστικά μεγέθη που εφαρμόζονται σε πίνακες συνάφειας 2x2 είναι (α) ο λόγος κινδύνων (risk ratio) και ο λόγος σχετικών πιθανοτήτων (odds ratio).

- **Παράδειγμα.** Ας υποθέσουμε ότι σε 100 αθλητές αξιολογήθηκε αν έχουν ή όχι ασυμμετρία δύναμης (> 15%) στους τετρακέφαλους (ανεξάρτητη μεταβλητή Χ) και αν έχουν υποστεί ή όχι κάκωση στα πόδια (εξαρτημένη μεταβλητή Υ).
- Τα αποτελέσματα των αξιολογήσεων αυτών δίνονται στον ακόλουθο πίνακα 2x2 μετά από ανάλυση των δεδομένων μέσω «crosstabulation» στο SPSS. Σε κάθε κελί ο αντίστοιχος αριθμός των αθλητών (f).

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ | | |
|---|---|---|---|---|
| | | ΝΑΙ | ΟΧΙ | Total |
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

28

28

## Λόγος κινδύνων (risk ratio) και Λόγος σχετικών πιθανοτήτων (odds ratio)

- Τέθηκαν πριν τη συλλογή των δεδομένων τα **ερευνητικά ερωτήματα** (research questions):
- (i) υπάρχει **σημαντική σχέση** μεταξύ ασυμμετρίας (ή όχι) και κάκωσης (ή όχι) στο επίπεδο α = 0.01;
- (ii) ποιος είναι ο **σχετικός κίνδυνος κάκωσης (risk ratio)**;
- (iii) ποια είναι η **σχετική πιθανότητα (odds ratio)** κάκωσης σε ασύμμετρους και συμμετρικούς αθλητές;

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ | | Total |
|---|---|---|---|---|
| | | ΝΑΙ | ΟΧΙ | |
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

29

29

## Ερευνητικό ερώτημα 1°: Συσχέτιση της ασυμμετρίας με την κάκωση

- Υπολογίζουμε το $\chi^2$ με τον γνωστό τύπο ως

$\boldsymbol{\chi^2}$ = N (ad - bc)$^2$ / [(a+b)*(a+c)*(d+b)*(d+c)] = 100 (42 * 23 - 20 * 15)2 / [62 * 57 * 43 * 38] = **7.68**,

- που με βαθμούς ελευθερίας df = 1 = (σειρές -1) * (στήλες -1) = (2 -1) * (2 -1) είναι στατιστικώς σημαντικό, καθότι είναι μεγαλύτερο από το κρίσιμο $\boldsymbol{\chi_c^2 = 6.63}$ (Πίν. 21.Θ, α = 0.01).

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ | | Total |
|---|---|---|---|---|
| | | ΝΑΙ | ΟΧΙ | |
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

30

30

## Ερευνητικό ερώτημα 1°: Συσχέτιση της ασυμμετρίας με την κάκωση

- Άρα η συσχέτιση μεταξύ των 2 μεταβλητών είναι στατιστικώς σημαντική.
- Η συσχέτιση αυτή είναι $\Phi = \sqrt{\chi^2/\mathrm{N}} = \sqrt{7.68/100} = 0.277$, δηλαδή **<u>σχεδόν μέτρια συσχέτιση</u>** (κλίμακα Cohen, 1988, 0.10 = μικρή, 0.30 = μέτρια, 0.50 = μεγάλη).

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ ΝΑΙ | ΑΣΥΜΜΕΤΡΙΑ ΟΧΙ | Total |
|---|---|---|---|---|
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

31

31

## Ερευνητικό ερώτημα 2°: Σχετικός κίνδυνος (RR) κάκωσης

- Ο δείκτης αυτός χρησιμοποιείται κυρίως σε **μελέτες προοπτικής** (prospective studies).
- Στο παράδειγμά μας, η ομάδα των ασύμμετρων αθλητών και η ομάδα των συμμετρικών αθλητών εντοπίζονται και στη συνέχεια καταγράφεται ποιοι τραυματίζονται (σε προπόνηση ή αγώνα) για ένα διάστημα π.χ. ενός έτους.

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ ΝΑΙ | ΑΣΥΜΜΕΤΡΙΑ ΟΧΙ | Total |
|---|---|---|---|---|
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

32

32

## Ερευνητικό ερώτημα 2°: Σχετικός κίνδυνος (RR) κάκωσης

- **Κίνδυνος (risk) κάκωσης (R)**:

  Ασύμμετροι αθλητές: $R_1 = a / (a + c) = 42 / 57 = 0.737$

  Συμμετρικοί αθλητές: $R_2 = b / (b + d) = 20 / 43 = 0.465$
- **Λόγος Κινδύνων (risk ratio) κάκωσης**:
- **Risk Ratio (RR)**: (κίνδυνος κάκωσης σε ασύμμετρους αθλητές) / (κίνδυνος κάκωσης σε συμμετρικούς αθλητές) = $R_1 / R_2 = 0.737 / 0.465 = 1.58$

**Ερμηνεία του Λόγου Κινδύνων (RR)**

Ο κίνδυνος κάκωσης στους ασύμμετρους αθλητές είναι 1.58 φορές μεγαλύτερος απ' ότι στους συμμετρικούς αθλητές.

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ | | |
|---|---|---|---|---|
| | | ΝΑΙ | ΟΧΙ | Total |
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |



## Ερευνητικό ερώτημα 3°: Σχετική πιθανότητα (OR) κάκωσης

- Ο δείκτης αυτός χρησιμοποιείται σε μελέτες αναδρομής (retrospective studies).
- Στο παράδειγμά μας, εντοπίζονται οι 2 ομάδες αθλητών (ασύμμετροι, συμμετρικοί) και **καταγράφονται αναδρομικά** οι τραυματισμοί που ήδη υπέστησαν οι αθλητές για ένα διάστημα π.χ. ενός έτους.

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ | | |
|---|---|---|---|---|
| | | ΝΑΙ | ΟΧΙ | Total |
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

## Ερευνητικό ερώτημα 3°: Σχετική πιθανότητα (OR) κάκωσης

- **Σχετική πιθανότητα (odds) κάκωσης (Ο)**:
  Ασύμμετροι αθλητές: $O_1 = a / b = 42 / 20 = 2.1$
  Συμμετρικοί αθλητές: $O_2 = c / d = 15 / 23 = 0.652$.
- **Λόγος Σχετικών Πιθανοτήτων (odds ratio) τραυματισμού**:
- **Odds Ratio (OR)** = (σχετική πιθανότητα κάκωσης σε ασύμμετρους αθλητές) / (σχετική πιθανότητα κάκωσης σε συμμετρικούς αθλητές) $= O_1 / O_2 = 2.1/ 0.652 =$ **3.22**

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ ΝΑΙ | ΑΣΥΜΜΕΤΡΙΑ ΟΧΙ | Total |
|---|---|---|---|---|
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

35

35

## Ερευνητικό ερώτημα 3°: Σχετική πιθανότητα (OR) κάκωσης

- **Λόγος Σχετικών Πιθανοτήτων (odds ratio) τραυματισμού**:
- **Odds Ratio (OR)** = (σχετική πιθανότητα κάκωσης σε ασύμμετρους αθλητές) / (σχετική πιθανότητα κάκωσης σε συμμετρικούς αθλητές) $= O_1 / O_2 = 2.1/ 0.652 =$ **3.22**
- **<u>Ερμηνεία του Λόγου Σχετικών Πιθανοτήτων</u> (odds ratio)**
- **<u>Η σχετική πιθανότητα κάκωσης στους ασύμμετρους αθλητές είναι 3.22 φορές μεγαλύτερη απ' ότι στους συμμετρικούς αθλητές</u>.**

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ ΝΑΙ | ΑΣΥΜΜΕΤΡΙΑ ΟΧΙ | Total |
|---|---|---|---|---|
| ΚΑΚΩΣΗ | ΝΑΙ | a 42 | b 20 | 62 |
| | ΟΧΙ | c 15 | d 23 | 38 |
| Total | | 57 | 43 | 100 |

36

36

## Εναλλακτική Προσέγγιση: Κίνδυνος κάκωσης ανά ομάδα (cohort)

- Στην ερευνητική πρακτική κάθε ομάδα που αξιολογείται προοπτικά (σε βάθος χρόνου) είναι μια cohort. Έτσι, στο παράδειγμα μας, έχουμε 2 cohorts: ομάδα ασύμμετρων αθλητών και ομάδα συμμετρικών αθλητών.
- Η εκτίμηση του **λόγου της σχετικής πιθανότητας (odds ratio)** μπορεί να υπολογισθεί από τους σχετικούς κινδύνους της κάθε ομάδας (cohort).
- **Ασύμμετροι: (42 / 62) / (15 / 38) = 0.677 / 0.395 = 1.716.**
- **Συμμετρικοί: (20 / 62) / (23 / 38) = 0.323 / 0.605 = 0.533.**

**Λόγος σχετικών πιθανοτήτων (Odds Ratio) για κάκωση ή όχι**: 1.72 / 0.53 = 3.22.

**ΚΑΚΩΣΗ * ΑΣΥΜΜΕΤΡΙΑ Crosstabulation**

| | | ΑΣΥΜΜΕΤΡΙΑ | | | | |
|---|---|---|---|---|---|---|
| | | ΝΑΙ | | ΟΧΙ | | Total |
| ΚΑΚΩΣΗ | ΝΑΙ | a | 42 | b | 20 | 62 |
| | ΟΧΙ | c | 15 | d | 23 | 38 |
| Total | | | 57 | | 43 | 100 |

37


38