Δημήτρης Βασιλόπουλος

έξι μαθήματα στατιστικής

ιατρικές εκδόσεις Λίτσας

6ο μάθημα
ΜΗ-ΠΑΡΑΜΕΤΡΙΚΕΣ ΔΟΚΙΜΑΣΙΕΣ

Στα προηγούμενα μαθήματα έγινε προσπάθεια να αναφερθούν, σε αδρές γραμμές, οι αρχές, οι ενδείξεις και οι περιορισμοί που διέπουν τη σύγκριση και τη συσχέτιση ποιοτικών και ποσοτικών χαρακτηριστικών. Αναφέρθηκε, ότι για τα ποιοτικά (αλλά και τα διατάξιμα χαρακτηριστικά–δες στη σελίδα 10) εφαρμόζονται οι δοκιμασίες x^2 και έγινε ειδική μνεία (σελ. 67) για τις προϋποθέσεις και τους περιορισμούς των δοκιμασιών αυτών. Για τα ποσοτικά χαρακτηριστικά, οι δοκιμασίες που εφαρμόζονται (t-test, δοκιμασία συσχετίσεως, κ.λπ.) στηρίζονται στην απαραίτητη προϋπόθεση ότι η κατανομή τους είναι ακριβώς, ή έστω κατά προσέγγιση, κανονική ή μπορεί να γίνει κανονική με κάποιο μετασχηματισμό (π.χ. λογαριθμικό μετασχηματισμό–δες στη σελίδα 55). Οι δοκιμασίες αυτές φέρονται ως **παραμετρικές**, γιατί η εφαρμογή τους προϋποθέτει την ύπαρξη διαφόρων παραμέτρων, όπως η μέση τιμή, η σταθερή απόκλιση κ.λπ. Σε αντιδιαστολή, όσες στατιστικές δοκιμασίες δεν "απαιτούν" τέτοιους υπολογισμούς, καλούνται **μη-παραμετρικές*** (non-parametric, distribution free).

Αρχές, ενδείξεις και περιορισμοί των μη-παραμετρικών δοκιμασιών

Οι μη-παραμετρικές δοκιμασίες έχουν αρκετές ενδείξεις αλλά και σημαντικούς περιορισμούς (Τριχόπουλος, 1971). Από το πλήθος των ενδείξεων και περιορισμών, επιλέγω:

1. Οι μη-παραμετρικές δοκιμασίες εφαρμόζονται σε ποσοτικά χαρακτηριστικά, σε περιπτώσεις που η κατανομή είναι είτε εκδήλως μη κανονική, είτε άγνωστη αλλά και ο αριθμός των παρατηρήσεων μικρός.
2. Οι μη-παραμετρικές δοκιμασίες εφαρμόζονται και σε διατάξιμα χαρακτηριστικά, ιδιαίτερα όταν το διατάξιμο αυτό χαρακτηριστικό μπορεί να λάβει πολλές

* Με την έννοια αυτή, οι δοκιμασίες x^2, που εφαρμόζονται σε ποιοτικά χαρακτηριστικά, είναι προφανώς "μη-παραμετρικές". Είναι όμως συνήθεια, οι δοκιμασίες x^2 να αναφέρονται ξεχωριστά, και υπό τον όρο "μη-παραμετρικές δοκιμασίες" να περιγράφονται μόνο οι δοκιμασίες εκείνες που εφαρμόζονται σε ποσοτικά χαρακτηριστικά χωρίς κανονική κατανομή ή σε διατάξιμα χαρακτηριστικά.

τιμές (π.χ. αριθμός δακτύλων (1-10), βαθμολόγηση μαθητών (1-20) κ.ο.κ.).

3. Οι μη-παραμετρικές δοκιμασίες εφαρμόζονται μόνο σε περιπτώσεις όπου οι αντίστοιχες παραμετρικές δοκιμασίες δεν είναι δυνατόν να εφαρμοσθούν. Άλλως, οι παραμετρικές δοκιμασίες διαθέτουν μεγαλύτερη ισχύ, μεγαλύτερη δηλαδή ικανότητα τεκμηριώσεως μιας διαφοράς ή μιας συσχετίσεως που ενδεχομένως υπάρχει.
4. Οι μη-παραμετρικές δοκιμασίες μπορούν να εφαρμοσθούν σε οποιαδήποτε κατανομή, αλλά πολλοί θεωρούν ότι καλό θα είναι: α) η κατανομή των δύο προς σύγκριση ομάδων να είναι σε αδρές γραμμές η ίδια και β) η κατανομή του χαρακτηριστικού να είναι συνεχής.
5. Η εφαρμογή των μη-παραμετρικών δοκιμασιών είναι εύκολη και πολύ απλούστερη από τις αντίστοιχες παραμετρικές (δεν απαιτείται καν η χρήση μικρού υπολογιστή). Ατυχώς όμως, μία από τις σημαντικές αδυναμίες αυτών των δοκιμασιών είναι ότι ο υπολογισμός των "ορίων αξιοπιστίας" που έχει μία διαφορά, είναι ιδιαίτερα δυσχερής. Τέλος, είναι ιδιαίτερα δύσκολη η χρήση τους επί μεγάλου αριθμού παρατηρήσεων.

Εφαρμογή των μη-παραμετρικών δοκιμασιών

Οι μη-παραμετρικές δοκιμασίες έχουν ένδειξη: 1) σε περιπτώσεις ποσοτικών χαρακτηριστικών με μη-κανονική κατανομή (ή μικρό αριθμό παρατηρήσεων και άγνωστη κατανομή), και 2) σε περιπτώσεις διαταξίμων χαρακτηριστικών. Στη στατιστική πρακτική υπάρχει πλήθος μη-παραμετρικών δοκιμασιών, που όχι μόνο καλύπτουν ένα ευρύτατο φάσμα ενδείξεων, αλλά έχουν και ποικίλουσα δυσκολία, που αφορά είτε την πρακτική τους εφαρμογή, είτε την κατανόηση του θεωρητικού υποβάθρου στο οποίο στηρίζονται. Για τις εντελώς περιορισμένες φιλοδοξίες αυτού του μικρού βιβλίου, θεωρήθηκε επαρκής η παρουσίαση τριών μόνο απλών και καθιερωμένων στη στατιστική πρακτική, δοκιμασιών:

Για *σύγκριση* ποσοτικών (και διαταξίμων) χαρακτηριστικών:

1. Η "δοκιμασία Wilcoxon δύο δειγμάτων–για παρατηρήσεις χωρίς αντιστοιχία" που είναι το μη-παραμετρικό αντίστοιχο της παραμετρικής "δοκιμασίας Student-t-test".
2. Η "δοκιμασία Wilcoxon για παρατηρήσεις κατά ζεύγη" που αντιστοιχεί στη "δοκιμασία t για παρατηρήσεις κατά ζεύγη".

Για *συσχέτιση* ποσοτικών (και διαταξίμων) χαρακτηριστικών:

3. Η "δοκιμασία μη-παραμετρικού συντελεστή συσχετίσεως σειράς-Spearman" που είναι το μη-παραμετρικό ανάλογο της παραμετρικής "δοκιμασίας συσχετίσεως".

Μετά την παράθεση όλων αυτών των δοκιμασιών είναι ίσως εφικτό να συνοψίσει κανένας τις ενδείξεις των διαφόρων στατιστικών δοκιμασιών στον πίνακα 5.

1. Δοκιμασία Wilcoxon δύο δειγμάτων (για παρατηρήσεις χωρίς αντιστοιχία)*

Η μέθοδος αυτή εφαρμόζεται για σύγκριση (δηλαδή για αξιολόγηση τυχόν διαφοράς) δύο ανεξαρτήτων ομάδων παρατηρήσεων. Η δοκιμασία έχει ένδειξη σε σύγκριση διαταξίμων χαρακτηριστικών ή ποσοτικών χαρακτηριστικών που δεν έχουν κανονική κατανομή ή η κατανομή τους μας είναι άγνωστη και επομένως η κλασική (παραμετρική) "δοκιμασία t" δεν είναι δυνατόν να εφαρμοσθεί.

Εφαρμογή

1. Το **πρώτο βήμα** στην εφαρμογή αυτής της δοκιμασίας είναι η διάταξη όλων των παρατηρήσεων (και των δύο ομάδων) σε μία ενιαία (αύξουσα ή φθίνουσα) κλίμακα.
2. Το **δεύτερο βήμα** είναι η άθροιση των θέσεων που καταλαμβάνουν στην ενιαία κλίμακα οι παρατηρήσεις κάθε μιας ομάδας ξεχωριστά. Το άθροισμα των θέσεων που καταλαμβάνουν οι θέσεις της πρώτης ομάδας συμβολίζεται με T_1 και της δεύτερης ομάδας με T_2 (πρώτη

---

* (Wilcoxon test for two samples, Wilcoxon rank sum test, Mann-Whitney U test).

Πίνακας 5. Ενδείξεις των διαφόρων στατιστικών δοκιμασιών

Χαρακτηριστικό	Σύγκριση		Συσχέτιση
	Πρώτη επιλογή	Δεύτερη επιλογή	
Ποιοτικό	δοκιμασίες x^2	–	δοκιμασίες x^2
Ποσοτικό			
1. Κανονική κατανομή	δοκιμασία-t (δοκιμασία t κατά ζεύγη)	Μη-παραμετρικές δοκιμασίες	δοκιμασία συσχετίσεως
2. Μη κανονική (ή άγνωστη) κατανομή	δοκιμασία Wilcoxon (δοκιμασία Wilcoxon κατά ζεύγη)	δοκιμασίες x^2	δοκιμασία μη-παραμετρικού συντελεστή (Spearman)
Διατάξιμο	δοκιμασία Wilcoxon	δοκιμασίες x^2	δοκιμασία μη-παραμετρικού συντελεστή (Spearman)

ομάδα θεωρείται εκείνη που έχει το μικρότερο αριθμό παρατηρήσεων).

3. Το **τρίτο**, και τελευταίο, **βήμα** είναι (όπως άλλωστε σε όλες τις στατιστικές δοκιμασίες) η αναζήτηση τυχόν σημαντικότητας στους αντίστοιχους πίνακες (Πίνακας 6).

Αλλά και εδώ τα παραδείγματα είναι απολύτως αναγκαία.

Παράδειγμα 6.1. Συγκρίνονται οι τιμές μιας σφαιρίνης του πλάσματος σε δύο ομάδες ασθενών (ομάδα Α και ομάδα Β) από 10 ασθενείς η κάθε μία (Swinscow, 1976). Επειδή ο αριθμός των παρατηρήσεων είναι μικρός και η κατανομή των παρατηρήσεων είναι άγνωστη, ένδειξη έχει η μη-παραμετρική "δοκιμασία Wilcoxon δύο δειγμάτων". Οι τιμές της σφαιρίνης (g/l) έχουν ως εξής:

Ομάδα Α: 34, 45, 39, 27, 38, 39, 28, 40, 45, 42, και
Ομάδα Β: 36, 32, 29, 41, 38, 26, 31, 35, 30, 31.

1. **Πρώτο βήμα:** η κατάταξη των παρατηρήσεων σε μία ενιαία κλίμακα.
   Εδώ θα πρέπει να αναφερθούν, ως παρένθεση, δύο παρατηρήσεις:
   α. Αν δύο ή περισσότερες απόλυτες τιμές είναι ίδιες, οι τιμές αυτές "βαθμολογούνται" με τη μέση τιμή των θέσεων που θα καταλάμβαναν αν δεν υπήρχε ισοβαθμία (π.χ. στο παράδειγμα 6.1, η τιμή 38 παρατηρείται δύο φορές δηλαδή στις θέσεις 12 και 13, οπότε κάθε τιμή βαθμολογείται με 12.5 ((12 + 13) : 2 = 12.5).

Ομάδα Α	Ομάδα Β	Θέση	Ομάδα Α	Ομάδα Β	Θέση
–	26	1	–	36	11
27	–	2	–	38	12.5
28	–	3	38	–	12.5
–	29	4	39	–	14.5
–	30	5	39	–	14.5
–	31	6	40	–	16
–	32	7	–	41	17
–	33	8	42	–	18
34	–	9	45	–	19.5
–	35	10	45	–	19.5

β. Για την επαλήθευση της ορθής βαθμολογήσεως των "θέσεων", θα πρέπει να γνωρίζουμε ότι το άθροισμα αυτής της βαθμολογίας ισούται με $v(v+1):2$ (π.χ. στο παράδειγμα 6.1, το άθροισμα των θέσεων πρέπει να είναι ίσο προς $v(v+1):2 \rightarrow 20\ (20+1):2=210$).

2. **Δεύτερο βήμα:** η άθροιση των "θέσεων" που καταλαμβάνουν οι τιμές κάθε μιας ομάδας χωριστά. Έχουμε λοιπόν:

   Ομάδα Α: $2+3+9+12.5+14.5+14.5+16+18+19.5+19.5=128.5$ και

   Ομάδα Β: $1+4+5+6+7+8+10+11+12.5+17=81.5$.

3. **Τρίτο βήμα:** Η αναζήτηση (σε πίνακες) τυχόν σημαντικότητας (όπως άλλωστε σε όλες τις στατιστικές δοκιμασίες). Οι μη-παραμετρικές όμως δοκιμασίες παρουσιάζουν στο σημείο αυτό μία σημαντική ιδιομορφία: ενώ δηλαδή στις παραμετρικές δοκιμασίες για να είναι στατιστικά σημαντική μία τιμή πρέπει να ξεπερνάει ένα κάποιο όριο που θέτουν οι πίνακες, στις μη-παραμετρικές δοκιμασίες Wilcoxon, για να υπάρχει σημαντικότητα, η τιμή πρέπει να είναι μικρότερη εκείνης που αναφέρεται στον πίνακα. Εν προκειμένω, αξιολογείται το μικρότερο άθροισμα (εδώ της ομάδας $B=81.5$). Ανατρέχοντας στον πίνακα 6, βλέπουμε ότι με $n_1=10$ και $n_2=10$ (αριθμός ασθενών στην ομάδα Α και στην ομάδα Β, αντίστοιχα), χρειάζεται τιμή μικρότερη του 78 και για να τεκμηριωθεί σημαντικότητα σε επίπεδο 5% και τιμή μικρότερη του 71 για επίπεδο σημαντικότητας 1%. Στο παράδειγμά μας (6.1) το μικρότερο άθροισμα είναι 81.5 άρα δεν υπάρχει σημαντική διαφορά. Έτσι, η απάντηση στο ερώτημα που αρχικά ετέθη είναι: "οι τιμές της σφαιρίνης των δύο ομάδων που εξετάσθηκαν δεν διαφέρουν σημαντικά".

Παράδειγμα 6.2. Ένα ακόμη, αυτή τη φορά μη-ιατρικό, παράδειγμα: Σε μία επαρχιακή πόλη οι μαθητές ενός πολυκλαδικού (ομάδα Α) και ενός γενικού λυκείου (ομάδα Β), πέτυχαν στο μάθημα της Φυσικής των Πανελληνίων εξετάσεων την εξής βαθμολογία:

Ομάδα Α	Ομάδα Β	Θέση	Ομάδα Α	Ομάδα Β	Θέση
–	26	1	–	36	11
27	–	2	–	38	12.5
28	–	3	38	–	12.5
–	29	4	39	–	14.5
–	30	5	39	–	14.5
–	31	6	40	–	16
–	32	7	–	41	17
–	33	8	42	–	18
34	–	9	45	–	19.5
–	35	10	45	–	19.5

β. Για την επαλήθευση της ορθής βαθμολογήσεως των "θέσεων", θα πρέπει να γνωρίζουμε ότι το άθροισμα αυτής της βαθμολογίας ισούται με $\nu(\nu+1):2$ (π.χ. στο παράδειγμα 6.1, το άθροισμα των θέσεων πρέπει να είναι ίσο προς $\nu(\nu+1):2 \rightarrow 20\ (20+1):2=210$).

2. **Δεύτερο βήμα:** η άθροιση των "θέσεων" που καταλαμβάνουν οι τιμές κάθε μιας ομάδας χωριστά. Έχουμε λοιπόν:

Ομάδα Α: $2+3+9+12.5+14.5+14.5+16+18+19.5+19.5=128.5$ και

Ομάδα Β: $1+4+5+6+7+8+10+11+12.5+17=81.5$.

3. **Τρίτο βήμα:** Η αναζήτηση (σε πίνακες) τυχόν σημαντικότητας (όπως άλλωστε σε όλες τις στατιστικές δοκιμασίες). Οι μη-παραμετρικές όμως δοκιμασίες παρουσιάζουν στο σημείο αυτό μία σημαντική ιδιομορφία: ενώ δηλαδή στις παραμετρικές δοκιμασίες για να είναι στατιστικά σημαντική μία τιμή πρέπει να ξεπερνάει ένα κάποιο όριο που θέτουν οι πίνακες, στις μη-παραμετρικές δοκιμασίες Wilcoxon, για να υπάρχει σημαντικότητα, η τιμή πρέπει να είναι μικρότερη εκείνης που αναφέρεται στον πίνακα. Εν προκειμένω, αξιολογείται το μικρότερο άθροισμα (εδώ της ομάδας $B=81.5$). Ανατρέχοντας στον πίνακα 6, βλέπουμε ότι με $n_1=10$ και $n_2=10$ (αριθμός ασθενών στην ομάδα Α και στην ομάδα Β, αντίστοιχα), χρειάζεται τιμή μικρότερη του 78 και για να τεκμηριωθεί σημαντικότητα σε επίπεδο 5% και τιμή μικρότερη του 71 για επίπεδο σημαντικότητας 1%. Στο παράδειγμά μας (6.1) το μικρότερο άθροισμα είναι 81.5 άρα δεν υπάρχει σημαντική διαφορά. Έτσι, η απάντηση στο ερώτημα που αρχικά ετέθη είναι: "οι τιμές της σφαιρίνης των δύο ομάδων που εξετάσθηκαν δεν διαφέρουν σημαντικά".

Παράδειγμα 6.2. Ένα ακόμη, αυτή τη φορά μη-ιατρικό, παράδειγμα: Σε μία επαρχιακή πόλη οι μαθητές ενός πολυκλαδικού (ομάδα Α) και ενός γενικού λυκείου (ομάδα Β), πέτυχαν στο μάθημα της Φυσικής των Πανελληνίων εξετάσεων την εξής βαθμολογία:

Πίνακας 6. "Δοκιμασία Wilcoxon δύο δειγμάτων". Επίπεδα σημαντικότητας 5% (πάνω) και 1% (κάτω) (το n_1 και n_2 αναφέρονται στον αριθμό των ατόμων κάθε ομάδας. Όταν οι αριθμοί αυτοί δεν είναι ίσοι το n_1 αναφέρεται στο μικρότερο αριθμό ατόμων).

Επίπεδο σημαντικότητας 5%

n_2 ↓ n_1 →	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4			10											
5		6	11	17										
6		7	12	18	26									
7		7	13	20	27	36								
8	3	8	14	21	29	38	49							
9	3	8	15	22	31	40	51	63						
10	3	9	15	23	32	42	53	65	78					
11	4	9	16	24	34	44	55	68	81	96				
12	4	10	17	26	35	46	58	71	85	99	115			
13	4	10	18	27	37	48	60	73	88	103	119	137		
14	4	11	19	28	38	50	63	76	91	106	123	141	160	
15	4	11	20	29	40	52	65	79	94	110	127	145	164	185
16	4	12	21	31	42	54	67	82	97	114	131	150	169	
17	5	12	21	32	43	56	70	84	100	117	135	154		
18	5	13	22	33	45	58	72	87	103	121	139			
19	5	13	23	34	46	60	74	90	107	124				
20	5	14	24	35	48	62	77	93	110					
21	6	14	25	37	50	64	79	95						
22	6	15	26	38	51	66	82							
23	6	15	27	39	53	68								
24	6	16	28	40	55									
25	6	16	28	42										
26	7	17	29											
27	7	17												
28	7													

Επίπεδο σημαντικότητας 1%

n_2 ↓ n_1 →	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5				15										
6			10	16	23									
7			10	17	24	32								
8			11	17	25	34	43							
9		6	11	18	26	35	45	56						
10		6	12	19	27	37	47	58	71					
11		6	12	20	28	38	49	61	74	87				
12		7	13	21	30	40	51	63	76	90	106			
13		7	14	22	31	41	53	65	79	93	109	125		
14		7	14	22	32	43	54	67	81	96	112	129	147	
15		8	15	23	33	44	56	70	84	99	115	133	151	171
16		8	15	24	34	46	58	72	86	102	119	137	155	
17		8	16	25	36	47	60	74	89	105	122	140		
18		8	16	26	37	49	62	76	92	108	125			
19	3	9	17	27	38	50	64	78	94	111				
20	3	9	18	28	39	52	66	81	97					
21	3	9	18	29	40	53	68	83						
22	3	10	19	29	42	55	70							
23	3	10	19	30	43	57								
24	3	10	20	31	44									
25	3	11	20	32										
26	3	11	21											
27	4	11												
28	4													

Ομάδα Α ($n_1=10$): 14, 6, 19, 9, 10, 7, 12, 13, 14, 11 και
Ομάδα Β ($n_2=14$): 18, 17, 8, 18, 7, 16, 10, 15, 18, 20, 19
17, 16, 18.

Είναι άραγε σημαντικά διαφορετική η επίδοση στη Φυσική των δύο αυτών ομάδων μαθητών;

1. **Πρώτο βήμα:** Διάταξη των παρατηρήσεων σε μία ενιαία κλίμακα:

Ομάδα Α	Ομάδα Β	Θέση	Ομάδα Α	Ομάδα Β	Θέση
6	–	1	–	15	13
7	–	2.5	–	16	14.5
–	7	2.5	–	16	14.5
–	8	4	–	17	16.5
9	–	5	–	17	16.5
10	–	6.5	–	18	19.5
–	10	6.5	–	18	19.5
11	–	8	–	18	19.5
12	–	9	–	18	19.5
13	–	10	19	–	22.5
14	–	11.5	–	19	22.5
14	–	11.5	–	20	24

2. **Δεύτερο βήμα:** Η άθροιση των "θέσεων" που καταλαμβάνουν οι τιμές κάθε μιας ομάδας χωριστά. Ομάδα A: $1+2.5+5+6.5+8+9+10+11.5+11.5+22.5=87.5$ και Ομάδα $B=2.5+4+6.5+13+14.5+14.5+16.5+1$ $6.5+19.5+19.5+19.5+19.5+22.5+24=212.5$. (Αν θέλουμε να επαληθεύσουμε την ορθή κατάταξή μας $212.5+87.5=300$ και $v\ (v+1):2=24\ (24+1):2=600:2=300$).

3. **Τρίτο βήμα:** Η αναζήτηση τυχόν σημαντικότητας. Στον πίνακα 6 για $n_1=10$ και $n_2=14$, χρειαζόμαστε το μικρότερο άθροισμα θέσεων να είναι μικρότερο από 91 για σημαντικότητα σε επίπεδο 5% και μικρότερο από 81 για επίπεδο σημαντικότητας 1%. Εδώ, το μικρότερο άθροισμα είναι 87.5 άρα "η επίδοση στη φυσική των δύο ομάδων μαθητών διαφέρει σημαντικά ($P < 0.05$)".

Παράδειγμα 6.3. Σε μία πρώτη κλινική εφαρμογή ενός νέου σκευάσματος εναντίον της ημικρανίας, το φάρμακο χορηγήθηκε σε 12 ασθενείς, ενώ άλλοι 10 ασθενείς πήραν placebo (Swinscow, 1976). Ο αριθμός των ημικρανικών κρίσεων που παρατηρήθηκε σε ένα εξάμηνο ήταν: ασθενείς με το φάρμακο (ομάδα B): 8, 6, 0, 3, 14, 5, 11, 2, 1, 9, 12, 4 και ασθενείς με placebo (ομάδα A): 7, 10, 4, 12, 2, 8, 8, 6, 0, 5. Υπάρχει άραγε διαφορά στη συχνότητα των κρίσεων ανάμεσα στις δύο ομάδες;

1. **Πρώτο βήμα:** Η διάταξη των παρατηρήσεων σε μία ενιαία κλίμακα.

Ομάδα placebo	Ομάδα με φάρμακο	Θέση	Ομάδα placebo	Ομάδα με φάρμακο	Θέση
0	–	1.5	–	6	11.5
–	0	1.5	7	–	13
–	1	3	8	–	15
2	–	4.5	8	–	15
–	2	4.5	–	8	15
–	3	6	–	9	17
4	–	7.5	10	–	18
–	4	7.5	–	11	19
5	–	9.5	12	–	20.5
–	5	9.5	–	12	20.5
6	–	11.5	–	14	22

2. **Δεύτερο βήμα:** Η άθροιση των θέσεων που καταλαμβάνει κάθε ομάδα. Έτσι, έχουμε ομάδα placebo $(T_1)=$ $1.5+4.5+7.5+9.5+11.5+13+15+15+18+20.5=116$ και ομάδα με φάρμακο $(T_2)=$ $1.5+3+4.5+6+7.5+9.5+11.5+15+17+19+20.5+22=137$

3. **Τρίτο βήμα:** Η αναζήτηση τυχόν σημαντικότητας. Στον πίνακα 6, για $n_1=10$ και $n_2=12$, για να τεκμηριωθεί σημαντικότητα χρειάζεται το μικρότερο άθροισμα (εδώ είναι 116) να είναι μικρότερο του 85 για επίπεδο 5%,

και μικρότερο του 76 για σημαντικότητα σε επίπεδο 1%. Είναι επομένως προφανές ότι "η συχνότητα των ημικρανικών κρίσεων ανάμεσα στις δύο ομάδες δεν έχει σημαντική διαφορά".

Δοκιμασία Wilcoxon για παρατηρήσεις κατά ζεύγη*

Η δοκιμασία αυτή, που είναι το μη-παραμετρικό ανάλογο της "δοκιμασίας t κατά ζεύγη", εφαρμόζεται για να αξιολογηθεί η διαφορά μεταξύ δύο ομάδων όταν οι παρατηρήσεις εμφανίζουν ατομική αντιστοιχία (σελ. 49). Η δοκιμασία Wilcoxon κατά ζεύγη χρησιμοποιείται κυρίως για την αξιολόγηση ποσοτικών χαρακτηριστικών όταν οι παρατηρήσεις δεν έχουν κανονική κατανομή (ή όταν η κατανομή είναι άγνωστη και ο αριθμός των ζευγών μικρός) αλλά και, σπανιότερα ίσως, στην αξιολόγηση διαταξίμων χαρακτηριστικών. Είναι βέβαια ευνόητο, και έχει ήδη τονισθεί, ότι η κλασική παραμετρική "δοκιμασία-t κατά ζεύγη" είναι ισχυρότερη της Wilcoxon κατά ζεύγη (όπως άλλωστε όλες οι παραμετρικές δοκιμασίες είναι ισχυρότερες των αντιστοίχων μη-παραμετρικών) και επομένως, όπου είναι δυνατόν να εφαρμοσθεί μπορεί να αναδείξει σημαντικές διαφορές που η "Wilcoxon κατά ζεύγη" δεν μπορεί να τεκμηριώσει. Πάντως, και στη μη-

* (Wilcoxon matched -pairs test, Wilcoxon signed-ranks test, Wilcoxon test for pair differences).

παραμετρική αυτή δοκιμασία ισχύουν οι επιφυλάξεις που έχουν αναφερθεί (σελ. 52) για την παραμετρική δοκιμασία-t κατά ζεύγη".

Εφαρμογή

Αν θεωρήσουμε δύο ομάδες παρατηρήσεων (Α και Β) με ατομική, ανά ζεύγη, αντιστοιχία.

1. **Πρώτο βήμα:** είναι ο υπολογισμός της διαφοράς που έχουν οι τιμές κάθε ατόμου, δηλαδή οι τιμές κάθε ζεύγους παρατηρήσεων. Οι διαφορές αυτές μπορεί να είναι θετικές, αρνητικές ή ίσες με μηδέν. Όπου η διαφορά ισούται με μηδέν, τα ζεύγη αυτά των παρατηρήσεων εξαιρούνται από την περαιτέρω επεξεργασία.
2. **Δεύτερο βήμα:** είναι η διάταξη κατ' αύξουσα κλίμακα των *απολύτων τιμών* που έχουν οι διαφορές. Η διάταξη γίνεται έτσι ώστε η μικρότερη απόλυτη τιμή να καταλαμβάνει την 1η θέση, ή αμέσως μεγαλύτερη τη 2η θέση, κ.ο.κ. (Όπως και στη δοκιμασία Wilcoxon δύο δειγμάτων, όταν κάποιες διαφορές έχουν την ίδια τιμή λαμβάνουν την ίδια θέση και "βαθμολογούνται" με το μέσο όρο των τιμών που θα έπαιρναν αν δεν ήταν ίσες).
3. **Τρίτο βήμα:** είναι να σημανθούν οι αντίστοιχες θέσεις με το θετικό ή αρνητικό πρόσημο (+ ή −) που είχε κάθε διαφορά. (Έτσι, οι θέσεις είναι τώρα "σημασμένες" με θετικό ή αρνητικό πρόσημο και γι' αυτό η δοκιμασία καλείται "δοκιμασία των σημασμένων

θέσεων-signed rank test). Αφού τεθεί το πρόσημο ακολουθεί η άθροιση των θέσεων που έχουν θετικό πρόσημο ξεχωριστά και των θέσεων με αρνητικό πρόσημο ξεχωριστά.

4. **Τέταρτο βήμα:** είναι η αναζήτηση σημαντικότητας από τους αντίστοιχους πίνακες (Πίνακας 7). Όπως και στην προηγούμενη δοκιμασία (Wilcoxon δύο δειγμάτων), έτσι και εδώ αξιολογείται το άθροισμα που έχει τη μικρότερη απόλυτη τιμή και επί τη βάσει αυτού αναζητείται η τυχόν σημαντικότητα.

Ένα απλό σχετικά παράδειγμα (Swinscow 1976, με τροποποιήσεις) θα κάνει ίσως τα πράγματα πιο κατανοητά.

Παράδειγμα 6.4. Υπολογίσθηκε σε 10 ασθενείς η τιμή μιας σφαιρίνης (g/l) πριν και μετά τη χορήγηση ενός νέου αντιρευματικού σκευάσματος. Τα αποτελέσματα είχαν ως εξής: Ασθενής Α: 38 (πριν)–45 (μετά), Ασθενής Β=26-28, Ασθενής Γ=29-27, Ασθενής Δ=41-38, Ασθενής Ε=36-40, Ασθενής ΣΤ=31-42, Ασθενής Ζ=32-39, Ασθενής Η= 30-39, Ασθενής Θ=35-34, και Ασθενής Ι=33-45. Υπάρχει άραγε διαφορά στην τιμή της σφαιρίνης πριν και μετά τη θεραπεία;

1. **Πρώτο βήμα:** Ο υπολογισμός των διαφορών στον πίνακα που ακολουθεί.

Ασθενής	Τιμές σφαιρίνης (g/l)			Απόλυτη τιμή διαφοράς (4)	Θέση (5)	"Σημασμένη" θέση (6)
	Πριν τη θεραπεία (1)	Μετά τη θεραπεία (2)	Διαφορά (3)			
Α	38	45	−7	7	6.5	−6.5
Β	26	28	−2	2	2.5	−2.5
Γ	29	27	2	2	2.5	2.5
Δ	41	38	3	3	4	4
Ε	36	40	−4	4	5	−5
ΣΤ	31	42	−11	11	9	−9
Ζ	32	39	−7	7	6.5	−6.5
Η	30	39	−9	9	8	−8
Θ	35	34	1	1	1	1
Ι	33	45	−12	12	10	−10

2. **Δεύτερο βήμα:** η διάταξη των απολύτων τιμών που έχουν οι διαφορές κατά αύξουσα κλίμακα (Πίνακας-στήλη 5).
3. **Τρίτο βήμα** είναι ο χαρακτηρισμός με πρόσημο (+ ή −) της θέσεως που έχει η απόλυτη τιμή της διαφοράς (Πίνακας-στήλη 6) και η άθροιση ξεχωριστά των θέσεων που έχουν θετικό και των θέσεων που έχουν αρνητικό πρόσημο. Η άθροιση αυτή στο παράδειγμα 6.4 έχει ως εξής: Θέσεις με θετικό πρόσημο $=2.5+4+1=7.5$ και θέσεις με αρνητικό πρόσημο $=6.5+2.5+5+9+6.5+8+10=47.5$.

Πίνακας 7. "Δοκιμασία Wilcoxon για παρατηρήσεις κατά ζεύγη". (Επίπεδα σημαντικότητας 5% και 1%).

Αριθμός ζευγών	Επίπεδο 5%	Επίπεδο 1%
7	2	0
8	2	0
9	6	2
10	8	3
11	11	5
12	14	7
13	17	10
14	21	13
15	25	16
16	30	19

4. **Τέταρτο βήμα:** η αναζήτηση σημαντικότητας. Αξιολογείται και εδώ (όπως στη Wilcoxon δύο δειγμάτων) το μικρότερο, κατά απόλυτη έννοια, άθροισμα, που στο παράδειγμα είναι 7.5. Ανατρέχοντας στον πίνακα 7, βλέπουμε ότι για 10 ζεύγη παρατηρήσεων, τιμές μικρότερες του 8 σημαίνουν σημαντικότητα σε επίπεδο 5% και τιμές μικρότερες του 3 σε επίπεδο 1%. Η τιμή επομένως (7.5) που βρέθηκε υποδηλώνει ότι "η διαφορά στις τιμές της σφαιρίνης πριν και μετά τη θεραπεία είναι σημαντική ($P < 0.05$)".

Δοκιμασία μη-παραμετρικού συντελεστή συσχετίσεως σειράς-Spearman

Η δοκιμασία αυτή είναι η αντίστοιχη της παραμετρικής δοκιμασίας συσχετίσεως (δες στη σελίδα 79) και εφαρμόζεται στις περιπτώσεις εκείνες που απαιτείται η συσχέτιση δύο ποσοτικών χαρακτηριστικών που η κατανομή

τους είτε εμφανώς δεν είναι κανονική, είτε μας είναι άγνωστη.

Όπως έχει ήδη αναφερθεί πρόβλημα συσχετίσεως ανακύπτει όταν έχουμε δύο (ή περισσότερες) σειρές παρατηρήσεων (διαφορετικών χαρακτηριστικών) και θέλουμε να διαπιστώσουμε αν οι μεταβολές του ενός χαρακτηριστικού έχουν κάποια σχέση (συσχέτιση) με τις μεταβολές του άλλου. Όπως είναι γνωστό από την παραμετρική δοκιμασία, σε κάθε τιμή της μιας σειράς παρατηρήσεων (X) αντιστοιχεί μία τιμή της άλλης σειράς (Y) και το πρόβλημα είναι, αν οι τιμές της μιας σειράς και οι αντίστοιχες της άλλης παρουσιάζουν κάποια συσχέτιση.

Η ερμηνεία των αποτελεσμάτων είναι αντίστοιχη της παραμετρικής "δοκιμασίας συσχετίσεως". Όπως δηλαδή και στην παραμετρική δοκιμασία, οι τιμές του μη-παραμετρικού συντελεστή συσχετίσεως σειράς (του Spearman) κυμαίνονται από -1 ως $+1$. Θετικές τιμές του συντελεστή αυτού σημαίνουν θετική συσχέτιση, οι αρνητικές αρνητική συσχέτιση, ενώ τιμές γύρω από το μηδέν υποδηλώνουν απουσία συσχετίσεως (δες στη σελίδα 87).

Εφαρμογή

1. **Πρώτο βήμα:** είναι η διάταξη (κατά τα γνωστά) των τιμών X κατ' ανιούσα κλίμακα και η βαθμολόγηση της θέσεως που κατέχει κάθε τιμή. Κατά την ίδια κλίμακα,

όπως στις τιμές X (δηλ. ανιούσα), κατατάσσονται και οι τιμές Y και "βαθμολογείται" και πάλι η θέση κάθε τιμής. Έτσι, για κάθε παρατήρηση υπάρχει η τιμή X (και η "θέση" της ανάμεσα στις τιμές X) και μία τιμή Y (και η "θέση" της ανάμεσα στις τιμές Y). Όπως είναι ευνόητο, αν η θέση κάθε τιμής X είναι η ίδια με τη θέση της αντίστοιχης τιμής Y, τότε υπάρχει πλήρης θετική συσχέτιση (συντελεστής συσχετίσεως σειράς = +1).

2. **Δεύτερο βήμα.** Για να υπολογισθεί ο συντελεστής συσχετίσεως σειράς (του Spearman) θα πρέπει να υπολογισθεί η διαφορά (Δ) μεταξύ της θέσεως που έχει μία τιμή X (στην κλίμακα των τιμών X) και της θέσεως που κατέχει η αντίστοιχη τιμή Y (στην κλίμακα των τιμών Y). Κάθε διαφορά (Δ) υψώνεται στο τετράγωνο (Δ^2) και λαμβάνεται το άθροισμα των τετραγώνων των διαφορών αυτών ($\Sigma\Delta^2$).

   Ο υπολογισμός του "μη–παραμετρικού συντελεστή συσχετίσεως σειράς του Spearman" (R) γίνεται με την εφαρμογή ενός απλού μαθηματικού τύπου που είναι:

$$R = 1 - \frac{6 \cdot \Sigma\Delta^2}{N \cdot (N^2 - 1)}$$

όπου $\Sigma\Delta^2$ = το άθροισμα των τετραγώνων των διαφορών και N = το πλήθος των παρατηρήσεων (των ζευγών).

3. **Τρίτο βήμα** είναι η αναζήτηση σημαντικότητας, η διερεύνηση δηλαδή αν ο συντελεστής R τεκμηριώνει κάποια σημαντική συσχέτιση. Από τον αντίστοιχο πίνακα (8) καθορίζεται αν υπάρχει σημαντική συσχέτιση ανάμεσα στα δύο χαρακτηριστικά, καθώς και το επίπεδο αυτής της σημαντικότητας.

Παράδειγμα 6.5. Σε μία πρώτη προσπάθεια ανιχνεύσεως τυχόν συσχετίσεως ανάμεσα στο βαθμό της Ανατομίας (Αν) και στο βαθμό της παθολογικής Ανατομίας (ΠΑ) αξιολογήθηκε η βαθμολογία "τυχαίου δείγματος" 12 φοιτητών. Τα δεδομένα έχουν ως εξής: Φοιτητής Α: Αν = 8 – ΠΑ 10, Φοιτητής Β: Αν = 2 – ΠΑ = 2, Φοιτητής Γ: Αν = 7 – ΠΑ = 6, Φοιτητής Δ: Αν = 6 – ΠΑ = 8, Φοιτητής Ε: Αν = 9 – ΠΑ = 9, Φοιτητής ΣΤ: Αν = 4 – ΠΑ = 5, Φοιτητής Ζ: Αν = 6 – ΠΑ = 3, Φοιτητής Η: Αν = 8 – ΠΑ = 7, Φοιτητής Θ: Αν = 5 – ΠΑ = 3, Φοιτητής Ι: Αν = 7 – ΠΑ = 9, Φοιτητής ΙΑ: Αν = 3 – ΠΑ = 4, Φοιτητής ΙΒ: Αν = 5–ΠΑ = 6. Το ερώτημα είναι: υπάρχει άραγε κάποια σχέση ανάμεσα στο βαθμό της Ανατομίας και στο βαθμό της παθολογικής Ανατομίας;

1. **Πρώτο βήμα:** η κατάταξη των τιμών X και των αντιστοίχων τιμών Y και η βαθμολόγηση της σειράς τους (στήλη 1-4 του πίνακα)

Φοιτητής	Ανατομία (Χ)		Παθ. Ανατομία (Υ)		Διαφορά (Δ)	$Δ^2$
	Βαθμός (1)	Σειρά (2)	Βαθμός (3)	Σειρά (4)	Σειρά (5)	(6)
Α	8	10.5	10	12	−1.5	2.25
Β	2	1	2	1	0	0
Γ	7	8.5	6	6.5	2	4
Δ	6	6.5	8	9	−2.5	6.25
Ε	9	12	9	10.5	1.5	2.25
ΣΤ	4	3	5	5	−2	4
Ζ	6	6.5	3	2.5	4	16
Η	8	10.5	7	8	2.5	6.25
Θ	5	4.5	3	2.5	2	4
Ι	7	8.5	9	10.5	−2	4
ΙΑ	3	2	4	4	−2	4
ΙΒ	5	4.5	6	6.5	−2	4
					$ΣΔ=0$	$ΣΔ^2=57$

2. **Δεύτερο βήμα:** Ο συντελεστής συσχετίσεως σειράς του Spearman είναι $R=1-\frac{6\cdot ΣΔ^2}{N\,(N^2-1)}$

Στο παράδειγμά μας

$$R=1-\frac{6\cdot 57}{12\,(12^2-1)}=1-\frac{342}{12\cdot 143}=1-\frac{342}{1716}=1-0.20=+0.80$$

$(R=0.80)$.

Πίνακας 8. Τιμές συντελεστή συσχετίσεως σειράς του Spearman. Η υπέρβαση μιας τιμής υποδηλώνει σημαντικότητα στο αντίστοιχο επίπεδο.

Βαθμοί ελευθερίας	Πιθανότητα (P)	
	5°/o	1°/o
5	1,000	—
6	0,886	1,000
7	0,786	0,929
8	0,738	0,881
9	0,683	0,833
10	0,649	0,794
11	0,619	0,772
12	0,591	0,751
13	0,566	0,729
14	0,544	0,708
15	0,524	0,687
16	0,506	0,665
17	0,490	0,644
18	0,475	0,625
19	0,462	0,607
20	0,450	0,591
21	0,438	0,576
22	0,428	0,562
23	0,418	0,549
24	0,409	0,537
25	0,400	0,526
26	0,392	0,515
27	0,384	0,505
28	0,377	0,496
29	0,370	0,487
30	0,364	0,478
31	0,358	0,470
32	0,352	0,463
33	0,347	0,455
34	0,341	0,448
35	0,336	0,442
40	0,314	0,413
45	0,296	0,388
50	0,280	0,368
60	0,255	0,335
70	0,236	0,310
80	0,221	0,290
90	0,208	0,273
100	0,197	0,259
150	0,161	0,211
200	0,139	0,183

3. **Τρίτο βήμα:** Ανατρέχουμε στον πίνακα 8 και με αριθμό παρατηρήσεων (ζεύγη) $N = 12$, παρατηρούμε ότι χρειάζεται τιμή R μεγαλύτερη (όπως και στην παραμετρική δοκιμασία) του 0.59 για επίπεδο σημαντικότητας 5% και μεγαλύτερη του 0.75 για επίπεδο σημαντικότητας 1%. Η τιμή μας R είναι 0.80 και επομένως: "στο δείγμα των φοιτητών που μελετήθηκε, υπάρχει ισχυρή συσχέτιση ($P < 0.01$) ανάμεσα στην επίδοση τους στην Ανατομία και στην επίδοσή τους στην Παθολογική Ανατομία".

Κλείνοντας το μικρό αυτό κεφάλαιο των μη-παραμετρικών δοκιμασιών, θα πρέπει και πάλι να τονισθεί, ότι οι τρεις μη-παραμετρικές δοκιμασίες που αναφέρθηκαν θα πρέπει να θεωρηθούν μόνο ως μία μικρή επιπλέον δυνατότητα επιλύσεως απλών προβλημάτων της βιοϊατρικής έρευνας και όχι ως επαρκές εφόδιο για την επίλυση των ποικίλων προβλημάτων που αναφύονται στην καθημερινή στατιστική πρακτική.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Armitage P.: Statistical Methods in Medical Research, Blackwell Scientific Publ., Oxford, 1971.

Cohran W.G.: Biometrics **10**:417, 1954. Όπως αναφέρεται από τον Swinscow (1976).

Colton T.: Statistics in Medicine, Little Brown, Boston, 1974.

Fisher R.A., Yates F.: Statistical tables for Biological, Agricultural and Medical Research, 6th edit., Oliver and Boyd, Edinburgh, 1963.

Hayslett H.T.: Statistics, Allen, London, 1968.

Hill A.B.: Principles of Medical Statistics, 9th Edit., Lancet, London, 1971.

Hill A.B.: Statistical Methods in Clinical and Preventive Medicine, Livingstone, Edinburgh, 1962.

Καραγιάννης Μ.Ι.: Επεξεργασία, αξιολόγηση και παρουσίαση αναλυτικών δεδομένων, Αθήνα, 1978.

Leaverton P.E.: Στοιχεία Βιοστατιστικής, Λίτσας, Αθήνα, 1983.

Snedecor W.G., Cohran W.G.: Statistical Methods, 6th Edit., Iowa State Univ. Press, Ames Iowa, 1967.

Swinscow T.D.V.: Statistics at Square One. Brit Med J, London, 1976.

Τριχόπουλος Δ.: Στοιχεία Ιατρικής Στατιστικής, Παρισιάνος, Αθήνα, 1971.

Ψωΐνος Δ.Π.: Εφαρμοσμένη Στατιστική, Εκδόσεις Ζήτη, Θεσσαλονίκη, 1989.

Yates F.: Journal of the Royal Statistical Society, Suppl. 1:217, 1934. Όπως αναφέρεται από τον Swinscow (1976).