



Ποιοτική και ποσοτική ανάλυση ιατρικών δεδομένων

Κωνσταντίνος Τζιόμαλος

Επίκουρος Καθηγητής Παθολογίας ΑΠΘ

Α΄ Προπαιδευτική Παθολογική Κλινική, Νοσοκομείο ΑΧΕΠΑ

1^ο βήμα : καταγραφή δεδομένων

- ◆ Το πιο πρακτικό πρόγραμμα είναι το EXCEL
- ◆ Για νοσήματα, χρησιμοποιούμε συνήθως το 1 αν είναι παρόν και το 0 αν δεν είναι παρόν (ή εναλλακτικά το 1 και το 2 αντίστοιχα)
– π.χ. αρτηριακή υπέρταση : 1 = ο ασθενής είναι υπερτασικός, 2 = δεν είναι υπερτασικός
- ◆ Πάντα σημειώνουμε τι σημαίνουν οι κωδικοποιήσεις που χρησιμοποιούμε, και ιδιαίτερα όταν έχουμε πολλές κατηγορίες (π.χ. BMI < 20 = 1, BMI 20-25 = 2, BMI 25-30 = 3, BMI > 30 = 4)

Καταγραφή δεδομένων (II)

- ◆ Προσοχή στα δεκαδικά – χρησιμοποιούμε κόμμα (,) και όχι τελεία (.) – π.χ. 10,3 και όχι 10.3
- ◆ Σε μεγάλα αρχεία, βοηθάει η σταθεροποίηση της 1^{ης} γραμμής (από το παράθυρο προβολή - σταθεροποίηση παραθύρων) οπότε η 1^η γραμμή είναι πάντα ορατή
- ◆ Το EXCEL υπολογίζει μέσες τιμές και σταθερές αποκλίσεις (από την επιλογή f_x της γραμμής τύπων) και αυτό επιτρέπει την εύκολη δημιουργία πινάκων και τον έλεγχο της ορθής μεταφοράς των δεδομένων στο στατιστικό πρόγραμμα

2^ο βήμα : εισαγωγή δεδομένων σε στατιστικό πρόγραμμα

- ◆ Το SPSS (Statistical Package for the Social Sciences) είναι από τα πιο συχνά χρησιμοποιούμενα προγράμματα στατιστικής
- ◆ Έχει παρόμοια δομή με το EXCEL και τα δεδομένα μεταφέρονται από το ένα πρόγραμμα στο άλλο με απλό copy-paste
- ◆ Επιτρέπει μόνο αριθμούς – επομένως, όλες οι παράμετροι στο EXCEL πρέπει να είναι αριθμοί πριν τους μεταφέρουμε στο SPSS

3^ο βήμα : επεξεργασία δεδομένων

- ◆ Οι μεταβλητές διακρίνονται σε κατηγορικές και συνεχείς
- ◆ Κατηγορικές : φύλο, ύπαρξη υπέρτασης, διαβήτη κλπ, οικονομική τάξη, κατηγορία βάρους, κλπ
- ◆ Συνεχείς : ηλικία, βιοχημικές και αιματολογικές παράμετροι, αρτηριακή πίεση κλπ.

Έλεγχος κατανομής

- ◆ Το πρώτο βήμα στην στατιστική επεξεργασία είναι ο έλεγχος αν οι συνεχείς μεταβλητές ακολουθούν κανονική κατανομή
- ◆ Ο έλεγχος γίνεται με τη δοκιμασία Kolmogorov-Smirnov
 - ◆ analyze → nonparametric tests → 1-sample K-S
- ◆ Αν το p είναι μικρότερο του 0,05 η κατανομή δεν είναι κανονική
- ◆ Σε μικρά δείγματα ($n < 50$), η κατανομή σπάνια είναι κανονική

Αποτέλεσμα συγκρίσεων : p

- ◆ Αν $p < 0,05$, η διαφορά είναι στατιστικά σημαντική, δηλ. δεν οφείλεται στην τυχαία διακύμανση των αριθμών – όσο πιο μικρό το p, τόσο πιο σημαντική η διαφορά
- ◆ Αν το δείγμα είναι μεγάλο, κλινικά ασήμαντες διαφορές μπορεί να είναι στατιστικά σημαντικές
- ◆ Αν το δείγμα είναι μικρό, αριθμητικά μεγάλες διαφορές μπορεί να μην είναι στατιστικά σημαντικές (δεν είναι σαφές όμως αν αυτό οφείλεται στο μέγεθος του δείγματος ή σε τυχαία διακύμανση)

Επιλογή στατιστικής δοκιμασίας

- ◆ Αν η κατανομή είναι κανονική, χρησιμοποιούμε παραμετρικές δοκιμασίες
- ◆ Αν η κατανομή είναι μη κανονική, χρησιμοποιούμε μη παραμετρικές δοκιμασίες
- ◆ Αν κάποιες μεταβλητές έχουν κανονική και κάποιες μη κανονική κατανομή, χρησιμοποιούμε μη παραμετρικές δοκιμασίες
- ◆ Εναλλακτική επιλογή είναι η μετατροπή των μεταβλητών ώστε να ακολουθούν κανονική κατανομή (π.χ. λογαριθμική)

Σύγκριση συνεχών μεταβλητών μεταξύ 2 ομάδων

- ◆ Παράδειγμα : διαφέρει το βάρος μεταξύ ανδρών και γυναικών?
- ◆ Παραμετρική δοκιμασία : independent samples t-test
 - ◆ analyze → parametric tests → independent samples t-test
- ◆ Μη παραμετρική δοκιμασία : Mann-Whitney test
 - ◆ analyze → nonparametric tests → 2 independent samples

Σύγκριση συνεχών μεταβλητών μεταξύ > 2 ομάδων (I)

- ◆ Παράδειγμα : διαφέρει η LDL χοληστερόλη μεταξύ ατόμων με φυσιολογικό βάρος, υπέρβαρων και παχύσαρκων?
- ◆ Παραμετρική δοκιμασία : One way analysis of variance (ANOVA)
 - ◆ analyze \rightarrow parametric tests \rightarrow one way ANOVA
- ◆ Μη παραμετρική δοκιμασία : Kruskal-Wallis test
 - ◆ analyze \rightarrow nonparametric tests \rightarrow K independent samples

Σύγκριση συνεχών μεταβλητών μεταξύ > 2 ομάδων (II)

- ◆ Αν βρεθεί σημαντική διαφορά μεταξύ των ομάδων στην one way ANOVA, το επόμενο βήμα είναι η πραγματοποίηση post-hoc tests με διάφορες μεθόδους (π.χ. Holm-Sidak, Bonferroni) για να δούμε ποια ζεύγη διαφέρουν μεταξύ τους (άτομα με φυσιολογικό βάρος και υπέρβαροι, άτομα με φυσιολογικό βάρος και παχύσαρκοι, ή υπέρβαροι και παχύσαρκοι)
- ◆ Στην περίπτωση των post-hoc tests, το επίπεδο στατιστικής σημαντικότητας δεν είναι 0,05 αλλά δίνεται από τον τύπο:
$$p = (1 - 0,95^a) / a \text{ (} a = \text{αριθμός συγκρίσεων)}$$

Σύγκριση συνεχών μεταβλητών μεταξύ > 2 ομάδων (III)

- ◆ Αντίθετα, η δοκιμασία Kruskal-Wallis παρέχει μόνο τη συνολική σύγκριση των ομάδων (άτομα με φυσιολογικό βάρος, υπέρβαροι, παχύσαρκοι) και όχι ανά ζεύγη
- ◆ Αν η δοκιμασία Kruskal-Wallis δείξει σημαντική διαφορά μεταξύ των ομάδων, για να δούμε ποια ζεύγη διαφέρουν μεταξύ τους (άτομα με φυσιολογικό βάρος και υπέρβαροι, άτομα με φυσιολογικό βάρος και παχύσαρκοι, ή υπέρβαροι και παχύσαρκοι), θα μπορούσε να εφαρμοστεί η δοκιμασία Mann-Whitney

Σύγκριση συνεχών μεταβλητών στην ίδια ομάδα πριν και μετά από μια παρέμβαση

- ◆ Παράδειγμα : ελαττώνονται τα τριγλυκερίδια με την απώλεια βάρους?
- ◆ Παραμετρική δοκιμασία : paired samples t-test
 - ◆ analyze → parametric tests → paired samples t-test
- ◆ Μη παραμετρική δοκιμασία : Wilcoxon test
 - ◆ analyze → nonparametric tests → 2 related samples

Σύγκριση συνεχών μεταβλητών μεταξύ δύο ομάδων πριν και μετά από μια παρέμβαση (I)

◆ Παραδείγματα :

- ◆ το σωματικό βάρος ελαττώνεται περισσότερο με την μεσογειακή διαίτα ή με διαίτα χαμηλής περιεκτικότητας σε λίπος? (2 ξεχωριστές ομάδες)
- ◆ το σωματικό βάρος ελαττώνεται περισσότερο με την μεσογειακή διαίτα στους άνδρες ή στις γυναίκες (μία ομάδα με 2 υποομάδες)

◆ Παραμετρική δοκιμασία : GLM repeated measures

- ◆ analyze → general linear model → repeated measures

Σύγκριση συνεχών μεταβλητών μεταξύ δύο ομάδων πριν και μετά από μια παρέμβαση (II)

- ◆ Η δοκιμασία GLM repeated measures παρέχει:
 - ◆ p για την αλλαγή της παραμέτρου στη διάρκεια του χρόνου σε κάθε ομάδα χωριστά (μεταβλήθηκε το βάρος στα άτομα που έλαβαν μεσογειακή δίαιτα? μεταβλήθηκε το βάρος στα άτομα που έλαβαν δίαιτα χαμηλής περιεκτικότητας σε λίπος?)
 - ◆ p για την σύγκριση της αλλαγής της παραμέτρου μεταξύ των 2 ομάδων (η μεταβολή του βάρους διαφέρει μεταξύ των ατόμων που έλαβαν μεσογειακή δίαιτα και των ατόμων που έλαβαν δίαιτα χαμηλής περιεκτικότητας σε λίπος?)

Σύγκριση συνεχών μεταβλητών μεταξύ δύο ομάδων πριν και μετά από μια παρέμβαση (III)

- ◆ Εναλλακτικές δοκιμασίες (λιγότερο δόκιμες)
 - ◆ Paired samples t-test (ή Wilcoxon test) σε κάθε ομάδα χωριστά
 - ◆ Independent samples t-test (ή Mann-Whitney test) μεταξύ των ομάδων μετά την παρέμβαση
 - ◆ Independent samples t-test (ή Mann-Whitney test) μεταξύ των μεταβολών της παραμέτρου (απόλυτων ή ποσοστιαίων) κατά τη διάρκεια της μελέτης μεταξύ των 2 ομάδων

Σύγκριση κατηγορικών μεταβλητών μεταξύ 2 ή περισσότερων ομάδων (I)

- ◆ Παράδειγμα : διαφέρει η συχνότητα της αρτηριακής υπέρτασης μεταξύ ανδρών και γυναικών?
- ◆ Δοκιμασία χ^2 (chi-square test)
 - ◆ analyze → descriptive statistics → crosstabs
- ◆ Σε μικρά δείγματα, εναλλακτική επιλογή είναι η δοκιμασία Fisher

Σύγκριση κατηγορικών μεταβλητών μεταξύ 2 ή περισσότερων ομάδων (II)

- ◆ Σε πίνακα 2x2 (π.χ. διαφέρει η συχνότητα της στεφανιαίας νόσου μεταξύ διαβητικών και μη διαβητικών ασθενών?) χρησιμοποιείται η διόρθωση κατά Yates
- ◆ Σε συγκρίσεις μεταξύ > 2 ομάδων δεν είναι δυνατό να εκτιμηθεί μεταξύ ποιων ομάδων διαφέρει η μεταβλητή (π.χ. διαφέρει η συχνότητα της στεφανιαίας νόσου μεταξύ ατόμων με φυσιολογικό βάρος, υπέρβαρων και παχύσαρκων?)

Συσχέτιση συνεχών μεταβλητών

- ◆ Παράδειγμα : σχετίζεται η γλυκόζη αίματος με την περίμετρο της μέσης?
- ◆ Παραμετρική δοκιμασία : Pearson correlation
 - ◆ analyze → correlate → bivariate και επιλογή Pearson
- ◆ Μη παραμετρική δοκιμασία : Spearman correlation
 - ◆ analyze → correlate → bivariate και επιλογή Spearman
- ◆ Παρέχει εκτός από το p και τον συντελεστή συσχέτισης r (εύρος 0-1 : όσο πιο μεγάλος, τόσο πιο ισχυρή η συσχέτιση)



**Πρέπει να βασανίσεις τα δεδομένα
μέχρι να ομολογήσουν !**