



ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ
Σχολή Επιστημών Υγείας
Πανεπιστήμιο Θεσσαλίας

Ιατρική και Υπολογιστική Στατιστική
Εργαστήριο Βιομαθηματικών

Μπρότσης Θεόδωρος, MSc, PhD Candidate
Ακαδημαϊκός Υπότροφος

Σημειώσεις Στατιστικής

Ιατρική και Υπολογιστική Στατιστική



A. Είδη μεταβλητών

- **Ποιοτικές:** Τα δεδομένα ανήκουν σε κατηγορίες ή ομάδες (2 ή και περισσότερες). Η σειρά κατηγοριοποίησης μπορεί να παίζει (=ordinal data, π.χ., βαρύτητα νόσου ή πια<μέτρια<σοβαρή) ή να μην παίζει ρόλο (=nominal data, π.χ., καπνιστής-μη καπνιστής). Στην πράξη στο SPSS χρησιμοποιείται το “nominal” στο variable view.
- **Ποσοτικές:** Τα δεδομένα λαμβάνουν αριθμητικές τιμές είτε διακριτές (discrete data, π.χ., απόλυτος αριθμός γεννήσεων στη Λάρισα το 2013) είτε συνεχείς (continuous data, π.χ., τιμή χοληστερόλης στο αίμα). Στο SPSS η αντίστοιχη επιλογή και για τα δύο είναι το “Scale”.

B. Ανάλυση σχέσης 2 μεταβλητών

Μία **ποιοτική (με 2 κατηγορίες ή ομάδες)** και μια **ποσοτική: «Φάρμακο»** (Α ή Β) και **«τιμή χοληστερόλης στο αίμα»**

- ✓ Μέτρηση σε **ΔΙΑΦΟΡΕΤΙΚΑ άτομα**. Κάποια έλαβαν το Α και κάποια το Β και μετρήθηκε η τιμή της χοληστερόλης στο αίμα.

Ερώτημα: Υπάρχει διαφορά (στον πληθυσμό, στην πραγματικότητα) μεταξύ φαρμάκου Α και Β ως προς την αποτελεσματικότητα τους στη ρύθμιση των τιμών χοληστερόλης στο αίμα;

Απάντηση: Κάνω **t-test για μη-ζευγαρωτές παρατηρήσεις(2-sample t-test)** ή το αντίστοιχο μη παραμετρικό (**Mann-Whitney**), βάσει του αν η ποσοτική μεταβλητή είναι κανονικά κατανομημένη στον πληθυσμό (υπάρχουν μεταβλητές γνωστές για τη στρεβλή κατανομή τους, π.χ. CRP, γλυκόζη αίματος) ή και το μέγεθος δείγματος. **ΠΡΑΚΤΙΚΑ:** Όσο μεγαλύτερο μέγεθος δείγματος έχω (πάνω από 30(;)) ανά κατηγορία της ποιοτικής μεταβλητής), τόσο με μεγαλύτερη ασφάλεια κάνω παραμετρικό test.

- ✓ Μέτρηση σε **ΙΔΙΑ άτομα**. Τα άτομα αρχικά έλαβαν το φάρμακο Α, μετρήθηκε η χοληστερόλη και μετά (σε χρονικό διάστημα ικανό ώστε να έχει περάσει η επίδραση του Α) έλαβαν το φάρμακο Β και μετρήθηκε η χοληστερόλη.



Ερώτημα: Υπάρχει διαφορά (στον πληθυσμό, στην πραγματικότητα) μεταξύ φαρμάκου A και B ως προς την αποτελεσματικότητα τους στη ρύθμιση των τιμών χοληστερόλης στο αίμα;

Απάντηση: Κάνω **t-test για ζευγαρωτές παρατηρήσεις** (paired t-test) ή το αντίστοιχο μη παραμετρικό (**Wilcoxon**). Η ίδια λογική υπάρχει και εδώ για την επιλογή μεταξύ παραμετρικού και μη παραμετρικού τεστ.

Μία **ποιοτική (με 3 ή περισσότερες κατηγορίες)** και μία **ποσοτική**: «**Φάρμακο**» (A ή B) και «**τιμή χοληστερόλης στο αίμα**»

- ✓ Μέτρηση σε **ΔΙΑΦΟΡΕΤΙΚΑ άτομα**. Κάποια έλαβαν το A, κάποια το B και κάποια το Γ και μετρήθηκε η τιμή της χοληστερόλης στο αίμα.

Ερώτημα: Υπάρχει διαφορά (στον πληθυσμό, στην πραγματικότητα) μεταξύ φαρμάκου A, B και Γ ως προς την αποτελεσματικότητα τους στη ρύθμιση των τιμών χοληστερόλης στο αίμα;

- ✓ Απάντηση: Κάνω **ANOVA** ή το αντίστοιχο μη παραμετρικό (**Kruskal Wallis**). Το αν διαφέρει το A από το B, το A από το Γ ή το B από το Γ δε μας το διευκρινίζει η ANOVA ή το Kruskal Wallis. Κάνω post-hoc t-test ή Mann-Whitney.

Μία **ποιοτική** με άλλη μία **ποιοτική**: «**Φάρμακο**» (A ή B) και «**Επίτευξη Θεραπείας**» (ΝΑΙ ή ΟΧΙ) (ή όσες κατηγορίες και αν υπάρχουν)

- ✓ Ερώτημα: Υπάρχει σχέση μεταξύ της επιλογής Φαρμάκου και επίτευξης θεραπείας;;; Ή ισοδύναμα, υπάρχει διαφορά στο ποσοστό επίτευξης θεραπείας μεταξύ φαρμάκου A και B;
- ✓ Απάντηση: Κάνω **χ^2** . Για να δω ποσοτικά ποιο φάρμακο υπερτερεί, υπολογίζω το Odds ratio.

Μία **ποσοτική** με μία άλλη **ποσοτική**: «**Ηλικία**» και «**τιμή χοληστερόλης στο αίμα**»

- ✓ Ερώτημα: Υπάρχει γραμμική σχέση μεταξύ της ηλικίας και της τιμής χοληστερόλης;



- ✓ Απάντηση: Υπολογίζω **r coefficient** και κάνω **simple linear regression**. Ο συντελεστής (κλίση) που παίρνω από τη γραμμική παλινδρόμηση (έστω στατιστικά σημαντικός) μου λέει πόσο αυξάνεται (ή μειώνεται) κατά μέσο όρο η τιμή της χοληστερόλης με κάθε χρόνο (μία μονάδα μέτρησης) αύξηση της ηλικίας.

Γ. Ανάλυση σχέσης περισσότερων από 2 μεταβλητών

Μία **ποσοτική** με περισσότερες της 1 μεταβλητές (**ποσοτικές** ή **ποιοτικές**)

- ✓ Παράδειγμα με 2 **ποσοτικές** ανεξάρτητες μεταβλητές: Συνεχίζοντας με το παραπάνω παράδειγμα, **Multivariable regression** σε αυτή την περίπτωση κάνω, όταν θέλω να δω πώς μεταβάλλεται η εξαρτώμενη μεταβλητή (τιμή χοληστερόλης) με τη μεταβολή της ηλικίας «ανεξαρτήτως» της πιθανής επίδρασης, π.χ., του βάρους. Ο συντελεστής (κλίση) που παίρνω από τη γραμμική παλινδρόμηση μου λέει πόσο αυξάνεται (ή μειώνεται) κατά μέσο όρο η τιμή της χοληστερόλης με κάθε χρόνο αύξηση της ηλικίας όταν περιλαμβάνεται το βάρος στο μοντέλο παλινδρόμησης. ΠΡΑΚΤΙΚΑ είναι σα να παίρνω άτομα του ίδιου βάρους και να βλέπω την «καθαρή» πλέον επίδραση της ηλικίας στην τιμή της χοληστερόλης.
- ✓ Παράδειγμα με 2 **ποιοτικές** ανεξάρτητες μεταβλητές: Ηλικία και βάρος σαν κατηγορικές μεταβλητές (άνω ή κάτω των 40 ετών, παχύσαρκος ή φυσιολογικού βάρους) και σχέση με την τιμή χοληστερόλης. Τότε κάνω είτε πάλι **Multivariable linear regression** είτε **two-way ANOVA χωρίς αλληλεπίδραση**. Ο συντελεστής (κλίση) που παίρνω από τη γραμμική παλινδρόμηση μου λέει πόσο αυξάνεται (ή μειώνεται) κατά μέσο όρο η τιμή της χοληστερόλης για άτομα άνω των 40 ετών σε σχέση με άτομα κάτω των 40 ετών (η αντιστοιχία με μια ποσοτική μεταβλητή είναι ότι η διαφορά μεταξύ των 2 κατηγοριών, άνω των 40 έναντι των κάτω των 40, είναι σα μεταβολή μίας μονάδας μέτρησης) όταν περιλαμβάνεται το βάρος στο μοντέλο παλινδρόμησης. ΠΡΑΚΤΙΚΑ είναι σα να παίρνω άτομα της ίδιας κατηγορίας βάρους και να βλέπω την «καθαρή» πλέον επίδραση της κατηγορικής ηλικίας στην τιμή της χοληστερόλης.
- ✓ Αν θέλω να εξετάσω αν υπάρχει αλληλεπίδραση των 2 ανεξάρτητων μεταβλητών, δηλ αν η σχέση μεταξύ ηλικίας (είτε ως ποσοτικής είτε ως ποιοτικής) και τιμής χοληστερόλης είναι διαφορετική σε διαφορετικά επίπεδα βάρους (είτε ως ποσοτικής είτε ως ποιοτικής) τότε κάνω **Multivariable linear regression με αλληλεπίδραση** είτε **two-way ANOVA με αλληλεπίδραση**. Για παράδειγμα, στο output στο SPSS εξετάζω αν



το p -value που βρίσκεται δίπλα από το interaction coefficient (π.χ., AGE*WEIGHT) είναι μικρότερο από το 0.05(δλδ υπάρχει αλληλεπίδραση) ή μεγαλύτερο από το 0.05(δλδ δεν υπάρχει αλληλεπίδραση). ΠΡΑΚΤΙΚΑ, όταν δεν περιλαμβάνεται αλληλεπίδραση στο μοντέλο, τότε λέω ότι η σχέση μεταξύ ηλικίας και χοληστερόλης είναι η ίδια σε όλα τα επίπεδα του βάρους. Αν υπάρχει αλληλεπίδραση τότε πρέπει να ξέρω το επίπεδο του βάρους για να πω ποια είναι ποσοτικά η σχέση της ηλικίας με τη χοληστερόλη.