LMM - Background

Τα Linear mixed models είναι μια επέκταση των απλών γραμμικών μοντέλων όπου επιτρέπεται η συμμετοχή τόσο σταθερών όρων (fixed effects) όσο και τυχαίων όρων (random effects)

Η χρησιμότητά τους έχει να κάνει με τη μοντελοποίηση δεδομένων στα οποία η υπόθεση της ανεξαρτησίας δεν ισχύει. Τέτοια δεδομένα μπορούν να προκύψουν από ιεραρχικές δομές (hierarchical structures). Για παράδειγμα, δεδομένα από μαθητές που συμμετέχουν σε μια μελέτη θα μπορούσαν να συλλεγούν από τάξεις οι οποίες ανήκουν σε σχολεία ενώ κατ’ αντιστοιχία δεδομένα ασθενών θα μπορούσαν να συλλεγούν από τους θεράποντες ιατρούς, οι οποίοι είναι σε κλινικές εντός νοσηλευτικών ιδρυμάτων.

Όταν υπάρχουν πολλαπλά επίπεδα στις ιεραρχικές δομές, όπως ασθενείς από τον ίδιο γιατρό, τότε η μεταβλητότητα στα δεδομένα μπορεί να θεωρηθεί ότι προκύπτει τόσο από την εντός της ομάδας μεταβλητότητα (within group) όσο και από τη μεταξύ των ομάδων (between group) μεταβλητότητα. Με λίγα λόγια, παρατηρήσεις ασθενών στο ίδιο επίπεδο (πχ. κάτω από τον ίδιο γιατρό) δεν μπορούν να θεωρηθούν ανεξάρτητες, μιας και οι ασθενείς του ίδιου γιατρού έχουν παρόμοια συμπεριφορά.

Οι παρατηρήσεις στο υψηλότερο επίπεδο είναι ανεξάρτητες μεταξύ τους.

Υπάρχουν διάφοροι τρόποι για να χειριστούμε ιεραρχικά δεδομένα.

* Ένας απλός τρόπος θα ήταν να τα ομαδοποιήσουμε (aggregate). Για παράδειγμα, αν υποθέσουμε ότι παρατηρήσεις από 5 ασθενείς έχουν ληφθεί από κάθε γιατρό, τότε αντί να χρησιμοποιήσουμε τα ατομικά δεδομένα των ασθενών (τα οποία δεν είναι ανεξάρτητα) θα μπορούσαμε να πάρουμε ένα μέσο όρο των παρατηρήσεων των ασθενών ανά γιατρό. Αυτά τα συγκεντρωτικά δεδομένα τότε θα ήταν ανεξάρτητα μεταξύ τους.
* Είναι γεγονός ότι τα συγκεντρωτικά δεδομένα θα δώσουν συνεπείς εκτιμήτριες και τυπικά σφάλματα για τις υπό εκτίμηση παραμέτρους. Το πρόβλημα είναι ότι με αυτή τη προσέγγιση δεν αξιοποιούμε τα δεδομένα στο σύνολό τους. Στο παράδειγμα μας, κάθε γιατρός αντί να προσφέρει 5 παρατηρήσεις, μια για κάθε ασθενή, θα προσφέρει μόνον μια παρατήρηση που θα είναι ο μέσος όρος αυτών.
* Μια άλλη προσέγγιση θα ήταν να αναλύσουμε τα δεδομένα ανά ομάδα (πχ. αν γιατρό). Αυτό θα μας υποχρέωνε να ‘τρέξουμε’ πολλά διαφορετικά γραμμικά μοντέλα, ένα ανά ομάδα. Για άλλη μια φορά θα είχαμε μια ανάλυση η οποία θα μας έδινε χρήσιμες πληροφορίες αλλά δε θα χρησιμοποιούσε όλο το ‘πλούτο’ των δεδομένων μας (richness) μιας και για κάθε μοντέλο που θα προσαρμόζαμε δε θα χρησιμοποιούσε τα δεδομένα των ασθενών των υπόλοιπων γιατρών, κάνοντας κάθε φορά την ανάλυση μας σε μικρό πλήθος παρατηρήσεων.
* Τα Linear Mixed Models (αναφέρονται και ως multilevel models) μπορούν να θεωρηθούν ότι είναι ένας χρυσός συμβιβασμός μεταξύ των πιο πάνω προσεγγίσεων. Η προσαρμογή πολλών μοντέλων απαιτεί την εκτίμηση πολλών παραμέτρων με τη χρήση λίγων δεδομένων ανά εκτίμηση (noisy estimates). Από την άλλη μεριά η χρήση συγκεντρωτικών δεδομένων μας αναγκάζει να χάσουμε όλες τις διαφορές που ενδέχεται να υπάρχουν ανά ομάδα (πχ. ανά γιατρό). Τα LMMs είναι κάπου στη μέση.