

Vorbereitungen:

Laden Sie den Datensatz `zufriedenheit` von der Homepage herunter und lesen Sie ihn in R ein.

Aufgabe 1: Eigenständige Modellierung & Testen der Random effects assumption

In dieser Aufgabe sollen Sie eine Fragestellung anhand longitudinal erhobener Daten selbständig in R untersuchen. Unter anderem sollen Sie hierzu auch einen Hausman-Test durchführen, um die Random effects assumption zu testen. Verwenden Sie dafür die auf der Homepage bereitgestellte Funktion `hausman.test()`.

- a) Sehen Sie sich den Code der Funktion an und vollziehen Sie die einzelnen Schritte nach. An welcher Stelle im Code müssten Sie eine Änderung vornehmen, wenn Sie die Funktion auf ein `lmer`- oder `gam`-Modell anpassen wollten statt des `lme`-Modells (Sie sollen hier nichts im Code ändern, sondern nur erkennen, wo eine Änderung notwendig wäre)?
- b) Bearbeiten Sie die untenstehende Fragestellung in R. Wenn Sie mit der restlichen Modellierung fertig sind führen Sie für Ihr finales Modell einen Hausman-Test für alle metrischen Kovariablen durch. Achten Sie für das fixed effects Modell darauf, dass die Subjekt-Variable als `factor` kodiert ist! Zu welcher Testentscheidung kommen Sie?
Fitten Sie unabhängig von der Testentscheidung nach dem Hausman-Test ein Hybrid Modell zur Schätzung des Effekts von `Einkommen`. Vergleich Sie die Ergebnisse mit dem random effects Modell und dem fixed effects Modell.

Fragestellung:

In Ihrer späteren Arbeit als statistischer Berater kommt ein Kunde auf Sie zu, welcher von 2005 bis 2016 jährlich die Lebenszufriedenheit (Skala [0-100], als metrisch angenommen) von 50 Personen zusammen mit einigen potenziellen Einflussgrößen erfasst hat. Da er keine Ahnung von Statistik hat, gibt er Ihnen die Daten mit dem Auftrag zu untersuchen, ob das Geschlecht, die Anzahl der Freunde sowie das Einkommen [in Tsd. €] einen signifikanten Einfluss auf die Zufriedenheit haben. Untersuchen Sie die Fragestellung mithilfe gemischter Modelle in R. Achten Sie dabei darauf, dass das Modell korrekt spezifiziert ist und dass alle Annahmen erfüllt sind!

Wichtig: Versuchen Sie zunächst die Aufgabe so weit wie möglich allein zu lösen. Falls Sie nicht mehr wissen, welche Punkte Sie bei der Modellierung noch zu beachten haben sehen Sie sich die Hilfestellungen auf der nächsten Seite an (diese müssen nicht notwendigerweise in der angegebenen Reihenfolge bearbeitet werden!).

Hinweis: Um die Sache nicht weiter zu verkomplizieren nehmen wir an, dass a) sich keine Interaktionen im Modell befinden, dass b) die ϵ_{ij} nicht korrelieren und dass c) hier keine Modellierung glatter Effekte mittels Splines notwendig ist (damit würde man auch den Spezialfall des *Longitudinalen* LMMs verlassen).

Hilfestellungen Aufgabe 1:

- i) Weisen die Residuen Ihres aktuellen Modells eine Struktur auf? Solange dies der Fall ist ist das Modell nicht korrekt spezifiziert!
- ii) Was ist eine sinnvolle Annahme für die Art des Zeiteffektes? Scheint ein linearer Effekt hier sinnvoll?
- iii) Ist die Aufnahme eines Random Intercepts sinnvoll?
- iv) Ist die Aufnahme eines/mehrerer Random Slopes sinnvoll?
Hinweis: `lmList`
- v) Ist die Annahme korrelierter Random effects im Modell sinnvoll?
- vi) Ist die Random effects assumption bzgl. der Kovariablen erfüllt?