

Übungszettel Multi-Level-Modelle

M.Psy.205, Dozent: Peter Zezula

Joshua Driesen (joshua.driesen@stud.uni-goettingen.de)

German

Links

[Übungszettel als PDF-Datei zum Drucken](#)

Hinweise zur Bearbeitung

1. Bitte beantworten Sie die Fragen in einer .Rmd Datei. Sie können Sie über `Datei > Neue Datei > R Markdown...` eine neue R Markdown Datei erstellen. Den Text unter dem *Setup Chunk* (ab Zeile 11) können Sie löschen. [Unter diesem Link](#) können Sie auch unsere Vorlage-Datei herunterladen (Rechtsklick > Speichern unter...).
2. Informationen, die Sie für die Bearbeitung benötigen, finden Sie auf der [Website der Veranstaltung](#)
3. Zögern Sie nicht, im Internet nach Lösungen zu suchen. Das effektive Suchen nach Lösungen für R-Probleme im Internet ist tatsächlich eine sehr nützliche Fähigkeit, auch Profis arbeiten auf diese Weise. Die beste Anlaufstelle dafür ist der [R-Bereich der Programmiererplattform Stackoverflow](#)
4. Auf der Website von R Studio finden Sie sehr [hilfreiche Übersichtszettel](#) zu vielen verschiedenen R-bezogenen Themen. Ein guter Anfang ist der [Base R Cheat Sheet](#)

Ressourcen

Das Thema Multi-Level-Modelle wird im Field, Kapitel 19 sehr gut verständlich und in adäquater Tiefe behandelt. Starke Empfehlung!

Allgemeines

In Sheet 07 haben Sie Multi-Level-Ansätze bereits im Kontext von Messwiederholungen kennengelernt. Dabei wurden verschiedene Messzeitpunkte auf Level 1 innerhalb verschiedener Personen bzw. Länder auf Level 2 genested/verschachtelt begriffen. Da sich Datenpunkte der gleichen Person etc. vermutlich ähnlicher sind als Datenpunkte unterschiedlicher Personen, ist die Annahme der Unabhängigkeit der Datenpunkte (a.k.a. keine Autokorrelation), die z.B. normale Regressionsverfahren beinhalten, verletzt. Multi-Level-Modelle ermöglichen es, die durch die hierarchische Struktur eines Datensatz entstehende Autokorrelation zu berücksichtigen. Abgesehen von den bereits bekannten Anwendungen auf Personen und Länder, werden am häufigsten Kliniken (wie in den Aufgaben in diesem Sheet) und Schulklassen auf Level 2 berücksichtigt.

1 Daten und Pakete

Laden Sie zunächst das tidyverse und das Paket nlme. Lesen Sie dann den Datensatz <https://pzezula.pages.gwdg.de/data/Tabstopp-getrennt> ein. Dieser enthält Daten zu der Lebensqualität von Personen, die sich für eine Schönheitsoperation entschieden haben:

Variable	Erklärung
particnu	Teilnehmernummer
Post_QoL	Lebensqualität nach Eingriff bzw. Wartezeit
Base_QoL	Baseline-Lebensqualität vor Beginn
Clinic	Codenummer dafür, zu welcher der zehn untersuchten Kliniken der Patient gehörte
Surgery	1 = Operation erfolgt, 0 = Wartelistenkontrollgruppe
Reason	1 = Es bestand ein körperlicher Grund für den Eingriff, 0 = Der Eingriff erfolgte zur Aussehensverbesserung
Age	Alter der Probanden
Gender	1 = Männlich, 0 = Weiblich
BDI	Standardisierter Fragebogen zur Erhebung depressiver Störungen
_Text-Variablen	Wiederholungen von Surgery, Reason und Gender in Worten, für etwaige Plotlabels

Uns interessiert vor allem, ob die Lebensqualität nach einer Schönheitsoperation tatsächlich höher ist als nach einer Zeit auf der Warteliste. Baseline-Lebensqualität, Alter, Geschlecht und Depressivität wurden als Kovariate erhoben. Clinic als Level-2-Variable erlaubt uns, für etwaige Variabilität in der OP-Güte zu kontrollieren, und Reason wird uns später eine differenziertere Aussage zum OP-Effekt erlauben.

2 - Fixed Effects only

Erzeugen Sie zunächst ein Modell ohne Random Effects, in dem Sie nur die Kovariaten als Prädiktoren verwenden. Nutzen Sie hierfür anstelle des `lm()`-Befehls den `nlme::gls()`-Befehl (Die Syntax ist die gleiche, sie müssen nur zusätzlich `method = "ML"` spezifizieren). Dies hat den Vorteil, dass die so erzeugten Modelle mittels `anova()` direkt mit Modellen mit Random Effects verglichen werden können.

Erzeugen Sie nun ein weiteres Modell ohne Random Effects, in dem neben den Kovariaten auch Surgery als Prädiktor verwendet wird. Wenn Sie möchten, versuchen Sie, das neue Modell mithilfe des `update()`-Befehls aus dem ersten Modell herzuleiten. Wenn das nicht klappt, können Sie das Modell aber auch wie gewohnt aufstellen.

Vergleichen Sie die beiden Modelle mittels `anova()`. Betrachten Sie zusätzlich die `summary()` des zweiten Modells, und dort insbesondere den Surgery-Prädiktor.

3 - Random Effects

Es ist gut vorstellbar, dass die Qualität der durchgeführten Operationen zwischen den untersuchten Kliniken schwankt, z.B. aufgrund der persönlichen Fähigkeiten der dort angestellten Chirurgen. Erstellen Sie ein Modell mit den gleichen Prädiktoren wie das zweite Modell aus der Aufgabe zuvor. Erlauben Sie hier jedoch, dass der Effekt von Surgery zwischen den verschiedenen Kliniken schwankt. Versuchen Sie, sich die Syntax hierfür mittels `?nlme::lme()` selbst zu erarbeiten. Ansonsten finden Sie in "sheet_repeated_measure" eine Erklärung.

Zur Klärung, ob Random Effects angemessen sind, wird häufig mittels `nlme::intervals(Modelname)` das 95%-Konfidenzintervall der Streuung der Effekte untersucht. Führen Sie diese Untersuchung durch. Sprechen die Ergebnisse eher für oder gegen Random Effects? Warum?

Nutzen Sie wieder eine `anova()`, um zu prüfen, ob das neue Modell signifikant mehr Varianz aufklärt als das letzte. Betrachten Sie auch wieder die `summary()`. Was fällt Ihnen auf?

4 - Interaktion 1

Bisher haben wir die Information, aus welchen Gründen die untersuchten Personen eine Schönheitsoperation anstreben, ignoriert. Es ist jedoch gut denkbar, dass eine Schönheitsoperation, die ein körperliches Problem behandeln soll (etwa eine Brustverkleinerung zur Behandlung von Rückenschmerzen), einen anderen Einfluss auf die Lebensqualität hat, als eine Schönheitsoperation, die lediglich der Verbesserung des subjektiven Aussehens dient. Erweitern Sie Ihr Modell aus der vorigen Aufgabe daher um eine Interaktion von Surgery und Reason. Führen Sie wie gehabt einen Modellvergleich mit dem vorigen Modell durch, und betrachten Sie die `summary` des neuen Modells.

5 - Interaktion 2

Das Ergebnis der vorigen Aufgabe legt nahe, dass der Effekt einer Schönheitsoperation von den Gründen für die OP abhängt. Erzeugen Sie zur näheren Untersuchung zwei weitere Modelle mit Random Effects, nur diesmal getrennt für die zwei möglichen Gründe. Die Interaktion von Surgery und Reason sollte nicht im Modell auftauchen, da es ja pro Modell nur einen Wert von Reason gibt. Versuchen Sie dies zunächst, indem Sie sich in die `subset`-Option von `nlme::lme()` einlesen. Wenn dies nicht klappt, können Sie auch mit `dplyr::filter()` zunächst zwei neue Datensätze erstellen. Betrachten Sie beide Modelle, und interpretieren Sie sie inhaltlich.

6 - Random Effects vs Interaktion

Sowohl bei Random Effects, als auch bei Interaktionen schwankt die Wirkung einer Variablen in Abhängigkeit von der Ausprägung einer anderen Variable. In dem hier verwendeten Beispiel war der Effekt von Surgery sowohl unterschiedlich für verschiedene Kliniken, als auch für verschiedene Gründe für die OP. Können Sie sich vorstellen, was der inhaltliche Unterschied ist? Bei welcher Fragestellung wäre eine Interaktion zwischen Klinik und Surgery möglicherweise die interessantere Herangehensweise? Und warum erlauben wir zwar dem Effekt von Surgery, zwischen den Kliniken zu schwanken, nicht aber den Effekten von Alter, Geschlecht, Depressivität etc.?

7 - Plot

Überlegen Sie sich eine angemessene Visualisierung für die Erkenntnisse aus dieser Auswertung!

Literatur

Anmerkung: Diese Übungszettel basieren zum Teil auf Aufgaben aus dem Lehrbuch *Discovering Statistics Using R* (Field, Miles & Field, 2012). Sie wurden für den Zweck dieser Übung modifiziert, und der verwendete R-Code wurde aktualisiert.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: SAGE Publications Ltd.

English

Links

[Exercise sheet in PDF](#)

Some hints

1. Please give your answers in a .Rmd file. You may generate one from scratch using the file menu: 'File > new file > R Markdown ...' Delete the text below *Setup Chunk* (starting from line 11). Alternatively you may use this [sample Rmd](#) by downloading it.
2. You may find the informations useful that you can find on the [front page of this course](#).
3. Don't hesitate to google for solutions. Effective web searches to find solutions for R-problems is a very useful ability, professionals do that too ... A really good starting point might be the R area of the programmers platform [Stackoverflow](#)
4. You can find very useful [cheat sheets](#) for various R-related topics. A good starting point is the [Base R Cheat Sheet](#).

Ressources

Multi-Level-Models are explained in Field's chapter 19, we highly recommend his accessible work!

general remarks

In Sheet 07, you've already encountered multi-level concepts in the context of repeated measures designs. There, we interpreted different time points on level 1 as nested within people (or countries) on level 2. Since different data points of one person are probobaly more similar to each other than to data points of a different person, the simple-regression assumption of independency is most likely violated (i.e., there's autocorrelation). Multi-level-models allow you to include the dependency structure resulting from the hierarchical structure of your data in your analysis. Apart from people and countires, the most frequent level 2 entities you'll encounter are clinics or schools.

1 Data and packages

Load the `tidyverse` and the `nlme`-package. Read in the data at https://pzezula.pages.gwdg.de/data/cosmetic_surgery (separated by tabs). This dataframe contains data on the quality of life of people who've decided to undergo cosmetic surgery:

Variable	explanation
particnu	participant number
Post_QoL	Quality of life after surgery/after the waiting time
Base_QoL	Baseline quality of life at the outset of the study
Clinic	Code number identifying in which of the ten clinics studied the surgery was performed
Surgery	1 = surgery performed, 0 = waiting list control
Reason	1 = there was a physicl原因 for the surgery, 0 = the surgery was performed for purely aesthetical reason

Variable	explanation
Age	age of the subject
Gender	1 = male, 0 = female
BDI	standardized depression questionnaire
_Text-Variables	Repetition of surgery, reason and gender as strings, for use as plot labels

We're mostly interested in whether quality of life really is higher after cosmetic surgery as compared to a waiting list control. Baseline quality of life, age, gender and depression were measured as covariates. Using clinic as a level 2 variable allows us, to control for possible differences in the quality of surgeries. The variable Reason will enable a more detailed analysis of surgery effects later on.

2 - Fixed Effects only

First, create a model without random effects, using only the covariates as predictors. Instead of `lm()`, use the `nlme::gls()` command - the syntax is the same, you'll just have to add `method = "ML"`. Creating a fixed-effects-model this way has the advantage of being able to compare this model with a random-effects-model using the `anova()` command. Now, create another model without random effects, using surgery as well as the covariates as predictors. If you want, try creating this new model from the old one using the `update()` command. Should that not work, you can just as well create it the old-fashioned way. Compare these two models with `anova()`. Additionally, check out the `summary()` of the second model, especially the surgery predictor.

3 - Random Effects

It's possible that there's considerable variability in the quality of the surgeries performed depending on the clinic, e.g. due to differing skills of the surgeons.

Create a model with the same predictors as the second model from the task before. For this new model however, allow the surgery effect to differ between the different clinics. Try figuring out the syntax for this yourself, using `?nlme::lme()`. If that doesn't work, you'll find an explanation in "sheet_repeated_measure". To make sure that random effects are warranted, there's often an analysis of the 95% confidence intervals of the standard deviation of the effects. You get these using `nlme::intervals(model_name)`. Perform this command. Do the results support allowing random effects or not? Why? Use another `anova()` to analyse whether the new model's able to explain significantly more variance than the last one. Also check out the `summar()` again. What do you notice?

4 - Interaction 1

Up until now, we haven't considered the reasons for getting cosmetic surgery. However, it's definitely possible that cosmetic surgery that is supposed to help with a physical problem (e.g., a breast reduction to help with back pain) might have a different effect on quality of life than a surgery for purely aesthetic reasons. Expand your model from the last task to include the interaction between surgery and reason. Again, perform an ANOVA and look at the summary.

5 - Interaction 2

The last task's result indicates that a surgery's effect depends on the reasons for the surgery. To investigate this further, create two new models with random effects, one for each reason. These models should not

include the interaction term from before, because there's always only one reason per model. First, try to solve this task using the `subset()`-option of `nlme::lme()`. If that doesn't work, you can create two separate dataframes using `dplyr::filter()`. Look at both models and interpret them.

6 - Random Effects vs Interaction

Both with random effects and interactions, the effect of one variable varies depending on the value of another variable. In our example, surgery's effect was affected both by different clinics and by different reasons. Can you explain the theoretical difference between these? What kind of question might have warranted looking at an *interaction* between clinic and surgery? And why did we allow surgery's effect to vary between clinics, but not the effects of age, gender or depression?

7 - Plot

Create a fitting visualization for the conclusions of this analysis!

Literature

Annotation: This exercise sheet bases in part on exercises, that you can find in the textbook *Discovering Statistics Using R* (Field, Miles & Field, 2012). They were modified for the purpose of this sheet and the R-code was actualized.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: SAGE Publications Ltd.