Exercise 2: Complete, No, and Partial Pooling

Henrik Singmann

January 2019 (updated: 2019-01-14)

Task Description

One finding reported in the literature is that the weather influences ones subjective well-being. File `weather_wb.csv` contains simulated (i.e., fictitious) data from a study in which a random sample of 60 individuals (column `id`) were asked to report their subjective well being (column `swb`) at the end of each of 45 consecutive days. Note, the start date for each participant can be a different day and participants can come from different locations/climates. Subjective well-being was a composite score constructed from four individual items. Participants had to rate how "happy", "joyful", "sad", and "angry" they felt on a scale from 0 to 10, and the last two items were reverse coded for the composite score. In addition, the data also contains the daily sun hours (column `sunh`) at the location the participant was during each day.

Your task is to analyse this data using no-pooling, at least one complete-pooling approach (ideally two different complete pooling approaches), and partial pooling using a linear mixed models. The research goal is to answer the question if there is a relationship between the daily sun hours and the reported subjective well-being.

Do the different approaches agree in their conclusions? If not, what you think is the correct approach and why do you think the different approaches might provide a different answer? If you can, please add one figure for each of your analyses approaches to the results.

Solution

Load Data and Packages

Read data and load some packages.

```
library("tidyverse")
library("cowplot")
theme_set(theme_bw(base_size = 12) +
            theme(legend.position="bottom",
                  panel.grid.major.x = element_blank(),
                  panel.grid.minor.x = element_blank()))
w <- read_csv("weather_wb.csv")
```

Descriptive Statistics

Then let's calculate some descriptives statistics. First, number of observations and participants

```
w %>%
  group_by(id) %>%
  summarise(n = n()) %>%
  do(psych::describe(.$n))
```

```
##    vars  n mean sd median trimmed mad min max range skew kurtosis se
## X1    1 60   45  0     45      45   0  45  45     0  NaN      NaN  0
```

Second, descriptive statistics where data is aggregated within participants:

```
w %>%
  group_by(id) %>%
  summarise(mean_sunh = mean(sunh),
            mean_swb = mean(swb)) %>%
  select(-id) %>%
  do(psych::describe(.))
```

```
##           vars  n mean   sd median trimmed  mad  min  max range  skew
## mean_sunh    1 60 5.09 2.27   4.97    5.08 3.12 1.14 8.84  7.71  0.03
## mean_swb     2 60 7.01 0.96   7.09    7.06 0.92 4.32 8.79  4.47 -0.50
##           kurtosis   se
## mean_sunh    -1.38 0.29
## mean_swb     -0.10 0.12
```

This shows quite a considerable range in mean daily sun hours as well as mean subjective well-being.

No Pooling

```
# ...
```

Complete Pooling

```
# ...
```

Partial Pooling

```
library("lme4")
# ...
```