Exercises for course day 6

Statistical analysis of repeated measurements and clustered data

This exercise guides you through different strategies for handling missing data and makes you aware of their statistical consequences. The first part of the exercise considers a randomized clinical trial where missing data occur due to an incomplete study design. The second part of the exercise considers the calcium study from the lecture where missing data is due to dropout.

Exercise 1: Misssing by design and consequences of naive handling.

A randomized controlled clinical trial was performed to evaluate the effect of topical zink sulfate on epidermal wound healing. Thirty healthy volunteers had suction-blister wounds induced on each buttock and were randomized to receive two out of the following three treatments:

A. Active treatment with a zink shower gel.

B. Placebo treatment with a gel identical to A execpt from containing zink sulfate.

C. Control treatment with demineralized water.

Several clinical outcomes were evaluated in the trial, but in this exercise we will only be concerned with the pain sensation experienced for each wound just after application of the treatments. This was assesssed by VAS-scores (0-100mm visual analogue scale) after each treatment application and summarized by area under the curve for the whole treatment course.

REF: Larsen et al, *Suction-blister wound healing studied in a double-blind randomized controlled trial with or without zink sulfate*, Wound Repair and Regeneration.

1. Load the data from `vasscoresW` (wide format) and make some quick descriptive statistics to get an overview of its contents.
   - How many subjects were allocated to each of the groups `AB`, `AC`, and `BC`?
   - How many observations of `vasA`, `vasB`, and `vasC` are missing?
   - What *type* of missing data are we dealing with?
   - Would it make sense to perform a complete case analysis on these data?
2. Consider two simple analyses to compare the response to treatment A and C:

   I. A paired t-test based on the data from allocation group `AC`.

   II. A two-sample t-test based on the data from allocation groups `AB` and `BC`.

What are the shortcomings of the t-tests? Could we make an alternative analysis that applies to all of the data at the same time?

3. Transform data to the long format and make spaghettiplots to illustrate it.

4. Make an analysis of the data based on an apropriate mixed model.

   - Estimate the mean difference between treatments A and C with a 95% CI.
   - Estimate the mean responses for treatments A, B, and C with 95% CIs.
   - What are the estimated SDs of the responses and the correlations between them?

5. Make an imputed dataset where missing data have been replaced by *mean value predictions*. **Hint:** Using a copy of the wide data, insert the treatment group means from question 1 in place of the missing data for each allocation group in turn.

   - Compute sample means, SDs and correlations for the imputed data. How do they compare to those of the original data.
   - Using a paired t-test, estimate the mean difference between treatments A and C with a 95% CI. How does the result compare with the mixed model in question 4?
   - Make spaghettiplots to illustrate the imputed data. How does it look?

Exercise 2: Analysis of a randomized baseline follow-up study with dropout using multiple imputations.

Run the demo `calcium2.R` step by step to reproduce the multiple imputations from scenario A presented in the lecture. Try to understand what the code is doing .

The program is organized in seven steps:

Step 1. Read in the data and transform it to the long format.

Step 2. Make descriptive statistics comparing completers and non-completers.

Step 3. Make $m = 50$ multiply imputed datasets.

Step 4. Transform the imputed data to the long format.

Step 5. Make descriptive statistics to check that the imputations are sensible.

Step 6. Evaluate the treatment effect in each of the imputed datasets and save the results.

Step 7. Combine the results into a final estimate with a 95% CI using Rubins formulae.