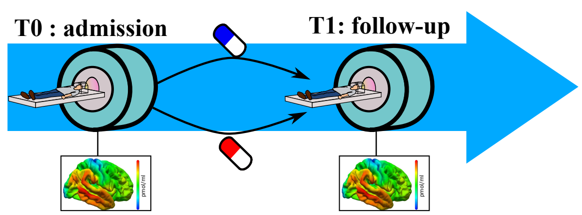
Exercises day 8

Basic Statistics for health researchers 2022

28 November 2022

Exercise A: what to adjust on?

In the lecture it was mentioned that using the change between baseline and follow-up provides a natural adjustment for certain but not all covariates (we assume that all covariates have a linear effect). Consider the following study:



The study aims at assessessing the impact of an antidepressive treatment (SSRI) on the brain serotonergic system. Patients were recruited, underwent baseline measurements, and were either given placebo or SSRI. A follow-up measurement was performed a week later. At each timepoint, a PET scan is performed to quantify the availability of serotonin receptors in the brain, which involves the injection of a radioactive contrast agent to the patient. A difference in change in PET signal between the two groups would be indicative of a treatment effect. However other factors may influence the PET signal:

- genetic polymorphisms (e.g. 5-HTTLPR)
- age (decline of 10% per decade)
- scanner type (binary variable, only 2 scanner types)
- radioactive dose (scan and patient dependent)

1. Which variable are "naturally" adjusted for when computed the change score? How would you test the treatment effect if there were no other variables to control for?

2. How would you control for the other variables?
What would be the benefit(s) of this adjustment?
(consider the case of a randomized study and an observational study)

3. In randomized experiment, adjusting for post-randomization variables is generally not recommended. Why? Is that problematic in this example?

Exercise B: analyzing a longitudinal study

In this exercise, we will reproduce the graphics and results presented during the lecture. A few extra-analyses will also be suggested. The exercise is divided in 3 independent parts:

- Part 1: descriptive statistics
- Part 2: comparing the change using t-tests
- Part 3: comparing the change using a mixed model

We recommend that you focus on 1-3, spending approximately 30 min for each part. The R syntax will be a bit more compliNote that most of the R code necessary to produce the results is in the R demo file.

The focus of today is more on the interpretation of the software output more

Key software output are included in the text and in the R demo file. to you so do not hesitate to ask for help!

To load the data in use:
(non R users should download the file `armd.txt` on the course webpage)

```
## requires the nlmeU package to be installed
data(armd.wide, package = "nlmeU")
```

The following code converts the data from the wide to the long format:

```
library(reshape2)
armd.long <- melt(armd.wide,
    measure.vars = paste0("visual",c(0,4,12,24,52)),
    id.var = c("subject","lesion","treat.f","miss.pat"),
    variable.name = "week",
    value.name = "visual")

armd.long$week <- factor(armd.long$week,
    level = paste0("visual",c(0,4,12,24,52)),
    labels = c(0,4,12,24,52))
```

You will also need to load the following packages:

```
library(LMMstar)
library(ggplot2)
```

Part 1: descriptive statistics

In this first part we will replicate the descriptive statistics presented during the lecture (slides 12-15).

1. We can display the dataset in the wide format using `str`. What is the meaning of the values in the columns `treat.f` and `miss.pat`?

```
str(armd.wide)
```

```
'data.frame':	240 obs. of  10 variables:
 $ subject : Factor w/ 240 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ lesion  : int  3 1 4 2 1 3 1 3 2 1 ...
 $ line0   : int  12 13 8 13 14 12 13 8 12 10 ...
 $ visual0 : int  59 65 40 67 70 59 64 39 59 49 ...
 $ visual4 : int  55 70 40 64 NA 53 68 37 58 51 ...
 $ visual12: int  45 65 37 64 NA 52 74 43 49 71 ...
 $ visual24: int  NA 65 17 64 NA 53 72 37 54 71 ...
 $ visual52: int  NA 55 NA 68 NA 42 65 37 58 NA ...
 $ treat.f : Factor w/ 2 levels "Placebo","Active": 2 2 1 1 2 2 1 1 2 1 ...
 $ miss.pat: Factor w/ 9 levels "----","---X",..: 4 1 2 1 9 1 1 1 1 2 ...
```

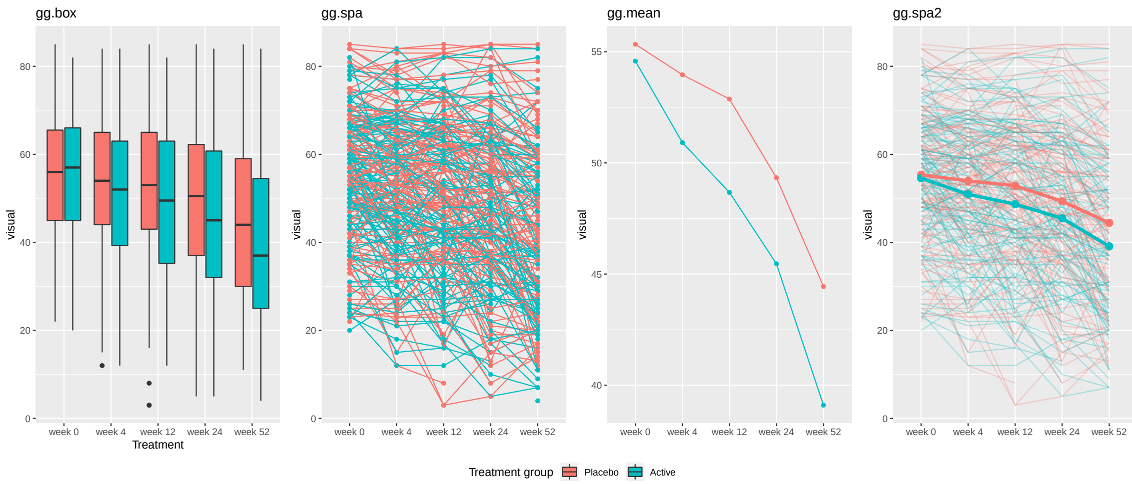
The `summarize` function can be used to compute summary statistics per group. Its first argument is a formula where the outcome is on the left hand side and the grouping variable(s) on the right-hand side, separated with `+`.

2. What information does the following software output provides?
   How would you do proceed to compute the mean and variance per time, regardless to the treatment group?

```
armd.s <- summarize(visual ~ week + treat.f, na.rm = TRUE,
      data = armd.long)
armd.s
```

```
   outcome week treat.f observed missing     mean       sd min    q1 median    q3 max
1   visual    0 Placebo      119       0 55.33613 15.00129  22 45.00   56.0 65.50  85
2   visual    4 Placebo      117       2 53.96581 15.90973  12 44.00   54.0 65.00  84
3   visual   12 Placebo      117       2 52.87179 17.20091   3 43.00   53.0 65.00  85
4   visual   24 Placebo      112       7 49.33036 18.51242   5 37.00   50.5 62.25  85
5   visual   52 Placebo      105      14 44.43810 18.53683  11 30.00   44.0 59.00  85
6   visual    0  Active      121       0 54.57851 14.82270  20 45.00   57.0 66.00  82
7   visual    4  Active      114       7 50.91228 15.81114  12 39.25   52.0 63.00  84
8   visual   12  Active      110      11 48.67273 17.47665  12 35.25   49.5 63.00  82
9   visual   24  Active      102      19 45.46078 18.08050   5 32.00   45.0 60.75  84
10  visual   52  Active       90      31 39.10000 18.40069   4 25.00   37.0 54.50  84
```

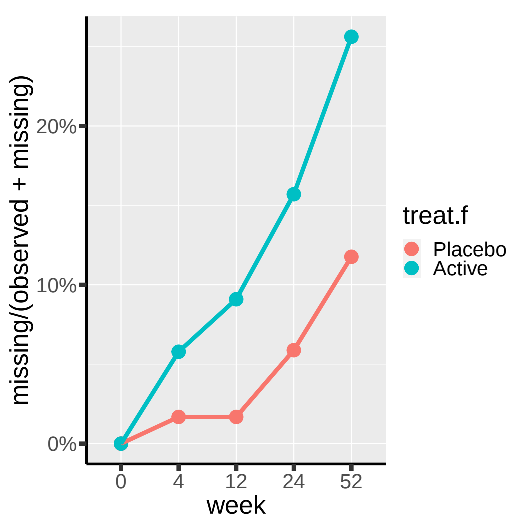
3. Discuss which of the following graphical representation (line 44-71 of the R demo file) you find the most useful to summarize the data? What information is missing?

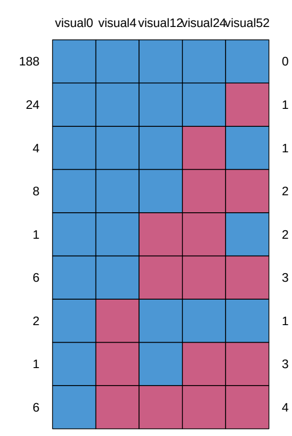


4. What type of information is provided by the following figures? Should we be worried?

```
## left panel
gg.NA <- ggplot(armd.s , aes(x = week, y = missing/(observed+missing),
    color = treat.f, group = treat.f))
gg.NA <- gg.NA + geom_point(size = 6) + geom_line(size = 2)
gg.NA <- gg.NA + scale_y_continuous(labels = scales::percent)
gg.NA

## right panel
library(mice)
md.pattern(armd.wide[,paste0("visual",c(0,4,12,24,52))])
```





Part 2: Univariate approach

5. What are the following lines of code achieving?

```
test <- is.na(armd.wide$visual0)+is.na(armd.wide$visual52)
armd.wideCC <- armd.wide[test==0,]
armd.wideCC$change <- armd.wideCC$visual52 - armd.wideCC$visual0
```

6. Assess the treatment effect by comparing the change between the two groups using a t-test. Extract the estimated effect, its confidence interval, and p-value.
   How does this analysis compares with the summary statistics computed in question 2?

7. Why do we get a (slightly) different p.value when using the `lm` function compared to the `t.test`?

```
e.lm <- lm(change ~ treat.f, data = armd.wideCC)
summary(e.lm)$coef
```

```
                Estimate Std. Error   t value     Pr(>|t|)
(Intercept)   -11.180952   1.557168 -7.180312 1.466539e-11
treat.fActive  -4.296825   2.292089 -1.874633 6.235402e-02
```

8. Repeat this analysis considering another timepoint (e.g. 24 weeks).
   What are the limitations of this approach?

Part 3: Multivariate approach

To start with we restrict the analysis to the first and last endpoint:

```
armd.long52 <- armd.long[armd.long$week %in% c("0","52"),]
armd.long52$week <- droplevels(armd.long52$week)
```

9. What is the interpretation of coefficients from the following mixed model?
   Do the results match those of the t-test?
   Can you deduce from the coefficients the estimated average vision at which timepoint?

```
e052.lmm <- lmm(visual ~ treat.f*week,
  repetition = ~week|subject,
  data = armd.long52[armd.long52$subject %in% armd.wideCC$subject,])
model.tables(e052.lmm)
```

```
                       estimate       se       df      lower      upper      p.value
(Intercept)           55.619048 1.452203 193.0400  52.754826 58.4832695 0.000000e+00
treat.fActive         -1.041270 2.137585 193.0400  -5.257290  3.1747506 6.267228e-01
week52               -11.180952 1.557168 192.9844 -14.252206 -8.1096988 1.466849e-11
treat.fActive:week52  -4.296825 2.292089 192.9844  -8.817588  0.2239375 6.235414e-02
```

10. Contrast the estimated treatment effect to the one of the following mixed models. Which one is the most reliable?

```
e52.lmm <- lmm(visual ~ treat.f*week,
  repetition = ~week|subject,
  data = armd.long52)
model.tables(e52.lmm)
```

```
                        estimate       se       df      lower      upper      p.value
(Intercept)           55.3361345 1.366936 238.0491  52.643299 58.0289704 0.000000e+00
treat.fActive         -0.7576221 1.925135 238.0490  -4.550098  3.0348538 6.942712e-01
week52               -11.0948777 1.550149 196.0714 -14.151983 -8.0377727 1.608180e-11
treat.fActive:week52  -4.3831236 2.274691 197.7356  -8.868891  0.1026443 5.542433e-02
```

```
e.lmm <- lmm(visual ~ treat.f*week,
  repetition = ~week|subject,
  data = armd.long)
model.tables(e.lmm)
```

	estimate	se	df	lower	upper	p.value
(Intercept)	55.3361345	1.366936	238.0191	52.643297	58.02897213	0.000000e+00
treat.fActive	-0.7576221	1.925135	238.0200	-4.550100	3.03485623	6.942712e-01
week4	-1.2812792	0.764694	231.3334	-2.787934	0.22537572	9.517842e-02
week12	-2.3516584	1.091400	219.6983	-4.502611	-0.20070566	3.227167e-02
week24	-6.0200224	1.318454	212.4899	-8.618947	-3.42109743	8.414486e-06
week52	-11.3109451	1.598782	192.6856	-14.464305	-8.15758503	2.701706e-11
treat.fActive:week4	-2.2042232	1.087419	231.9888	-4.346702	-0.06174429	4.380391e-02
treat.fActive:week12	-3.5079396	1.560344	222.4007	-6.582891	-0.43298809	2.554512e-02
treat.fActive:week24	-3.0695747	1.895345	216.4638	-6.805269	0.66611980	1.067885e-01
treat.fActive:week52	-4.8662683	2.317422	198.7570	-9.436157	-0.29637910	3.700270e-02

11. Create a numeric time variable `week.num` indicating the number of weeks since baseline.
    Fit a mixed model including in the mean structure the categorical time variable and an interaction between the continuous time variable and the treatment variable.
    What is the estimated treatment effect in this new model?

	estimate	se	df	lower	upper	p.val
(Intercept)	54.95416667	0.96083065	239.0200	53.0613893	56.846944015	0.000000e+
week4	-2.20654872	0.55201346	242.6356	-3.2938989	-1.119198564	8.505743e-
week12	-3.58487586	0.81927366	258.5491	-5.1981745	-1.971577178	1.757722e-
week24	-6.56331226	1.05848300	279.3102	-8.6469293	-4.479695245	2.015522e-
week52	-11.60066367	1.53164482	203.2552	-14.6206139	-8.580713446	1.248779e-
week.num:treat.fActive	-0.08299735	0.04090117	187.3855	-0.1636833	-0.002311424	4.385081e-