Complex Data - final project outline

Anna Zaleska, Stanisław Wilczyński

21.05.2018

In our project we would like to examine **kidney** dataset from R package **PBImisc**. In this dataset we can find information about 334 patients' kidney condition after kidney transplant surgery. The condition is measured at 8 different time points after the transplant. For each patient we have a set of additional features like donor's age, recipient's age, if the patient is diabetic and the type of therapy that each patient underwent. Here we present the first 5 rows of the dataset divided to time measurements and mentioned additional features.

	MDRD7	MDRD30	MDRD3	MDRD6	MDRD12	MDRD24	MDRD36	MDRD60
1	46.00	71.00	65.00	71.00	65.00	70.00	76.00	72.00
2	44.00	58.00	20.00	78.00	62.00	87.00	72.00	47.00
3	6.00	36.00	37.80	42.00	45.00	46.00	46.00	33.00
4	8.00	39.00	55.70	52.00	60.00	73.00	75.00	91.00
5	36.00	79.00	64.20	64.00	64.00	79.00	70.00	63.00

Table 1: Time measurements

	recipient.age	donor.age	CIT	discrepancy.AB	discrepancy.DR	therapy	diabetes	bpl.drugs
1	26	40	25.00	2	1	tc	0	3
2	52	62	31.00	4	1	cm	1	4
3	50	50	20.00	2	1	cm	0	2
4	56	47	38.00	3	0	tc	0	3
5	67	18	26.00	1	1	ca	0	1

Table 2: Other features

In our analysis we would like to find and test some potential dependencies between kidney condition, time and the additional features eg.:

1. Are there any global trends of the condition in time?
2. Can we split the data using some factors (therapy, diabetes) and see different trends within each group (test for parallelism)?
3. We would like to find out which covariates impact the kidney condition (test for main effects).
4. We will test different levels of models complexity (eg. different covariance classes).
5. We will try to fit other models than linear (maybe quadratic curves).
6. Possibly some other things that will be covered within following weeks during the lecture.