Repeated Measurements in Clinical Studies

Veronika Ročková[1] , *Emmanuel Lesaffre*[1 2]

[1]Dept. of Biostatistics
Erasmus University, Rotterdam, The Netherlands

[2]L-BioStat
KU Leuven, Belgium

Exercise Session One

Erasmus MC
University Medical Center Rotterdam

Outline

Outline

Outline

Outline

Erasmus MC
University Medical Center Rotterdam

Outline

1. Course Materials
2. Goals of This Exercise Session
3. The Data
4. The Data Exploration
5. Data Visualization
6. "Simple Methods"
7. Conclusions

Erasmus MC
University Medical Center Rotterdam

Outline

Outline

Materials for the Session

- The course materials contain 5 sets of exercises, each has two parts
  - Part one: *guided tour through an exercise with a complete solution*
  - Part two: *just an assignment without a solution*
- All `sas` files for today are located in the directory `//Day one`
- We will focus on **Part two**, leaving Part one as a home exercise
- Please find
  - **Exercise one Part two**
  - `exercise1_part2_code.sas`

Goals of This Exercise Session

(1) Getting familiar with the SAS system (loading and plotting data, summary statistics etc.)

(2) Understanding the complexity of repeated measurements data

(3) Performing data exploration to get a better feel for the data

(4) Trying out simple methods to analyze longitudinal data

(5) Understanding the limitations of such simple approaches

The Hipfracture Data

- 60 patients recovering after a hip fracture operation
- 5 examinations at days 1, 3, 5, 8 and 12 after the operation
- OUTCOME
  Mini-Mental State Examination score (MMSE): number of correct answers to 30 questions
- BASELINE CHARACTERISTICS
  - age: continuous
  - neuro-status: binary
    (0/1 for neuro-psychiatric/non-neuro-psychiatric before the operation)

The Objective of the Investigator

How does the MMSE score **evolve** over time and how this **evolution** depends on the neuro-status and age?

Getting the Feel for the Data

⇝ Data exploration is a necessary pre-modeling step!

- *Check for missing values*
- *Check for outliers*
- *Computing simple summary statistics (mean, standard deviation)*
- *Deviations from model assumptions (normality, homoscedasticity)*

  ⋮

Erasmus MC
University Medical Center Rotterdam

Getting Started with SAS

- Scripting language to perform statistical analysis
- Many different ways to load the data

```
data hipfracture;
input NEURO MMSE1 MMSE3 MMSE5 MMSE8 MMSE12 IDNR AGE;
cards;
0       28      28      28      26      25      1       74
0       25      25      23      27      .       2       67
0       26      29      29      27      .       3       67
....
0       24      20      26      29      25      59      83
0       26      28      25      30      27      60      77
;
run;
```

- This is a horizontal data format

Summary Statistics

- How many neuro-psychiatric patients are in the data?

```
proc freq data=hipfracture;
 table NEURO;
run;
```

- Compute means of the responses and age over time for the two neuro groups

```
proc sort data=hipfracture;
by neuro;
run;

proc means data=hipfracture Q1 median mean  std Q3 maxdec=2 n;
by NEURO;
var  MMSE1 MMSE3 MMSE5 MMSE8 MMSE12  AGE;
run;
```

Plot Histograms

- Check for symmetry/skewness in the MMSE distribution.
- Normality in the responses would be nice.

```
proc univariate data=hipfracture;
var NEURO MMSE1 MMSE3 MMSE5 MMSE8 MMSE12  AGE;
histogram  MMSE1 MMSE3 MMSE5 MMSE8 MMSE12 / normal;
run;
```

- Separate histograms for the two NEURO groups?

Summary Statistics

- *How many missing observations there are in the data for each response variable?*
- *Can we already tell something about the effect of NEURO on the mean evolution of MMSE over time?*
- *How about the effect of AGE (a continuous variable)?*

Summary Statistics

- Categorize AGE based on quartiles and get the summary statistics of MMSE in the 4 categories

```
data hipfracture;
set hipfracture;
agecat=.;
 if (AGE<72) then agecat=1;
 if (AGE>=72) and  (AGE<80) then agecat=2;
 if (AGE>=80) and (AGE<84) then agecat=3;
 if (AGE>=84) then agecat=4;
run;

proc sort data=hipfracture;
by agecat;
run;

proc means data=hipfracture Q1 median mean  std Q3 maxdec=2 ;
by agecat;
var NEURO MMSE1 MMSE3 MMSE5 MMSE8 MMSE12  AGE;
run;
```

Erasmus MC
University Medical Center Rotterdam

Conversion to the Long Format

- We proceed by visualizing MMSE evolution over time
- It is easier to convert the data in the **long format**
- We use a SAS macro `makelong` (alternative ways in Part One of Exercise 1)

```
proc sort data=hipfracture;
by IDNR;
run;

%makelong(data=hipfracture,
out=hip_long,
id=IDNR,
root=MMSE,
copy=NEURO AGE agecat,
measurement=TIMECLSS)

proc print data=hip_long
;
run;
```

Plotting the Mean Evolution

- **Mean** MMSE evolution over time, overall and in relation to AGE and NEURO

```
proc gplot data=hip_long;
goptions reset=all ftext=swiss;
plot MMSE*TIMECLSS ;
symbol c=blue i=std1mjt l=1 w=2;
run; quit;

proc gplot data=hip_long;
plot MMSE*TIMECLSS=NEURO ;
symbol1 c=red i=std1mjt l=1 w=2;
symbol2 c=blue i=std1mjt l=1 w=2;
run; quit;
```

- *Does MMSE decrease over time? Is the behavior of MMSE different in the two NEURO groups?*

Plotting the Individual Evolution

- **Individual** MMSE evolution over time (separately based on NEURO)

```
proc sort data=hip_long;
by NEURO;
run;

proc sgplot data=hip_long;
by NEURO;
series x=timeclss y=MMSE /group=IDNR
                    lineattrs=(color=green pattern=1);
run;
```

- *Are the individual curves **wiggly**? Does the wigglyness differ in the two NEURO groups?*
- *Which NEURO group has a higher "within-subject" variability?*

Erasmus MC
University Medical Center Rotterdam

Exploring the Correlation

- Repeated measurements are likely to be correlated!
- Pearson correlation matrix (separately for the two NEURO groups)
- Again we use the horizontal data format

```
proc sort data=hipfracture;
by neuro;
run;

proc corr data=hipfracture;
by NEURO;
var MMSE1 MMSE3 MMSE5 MMSE8 MMSE12;
run;
```

- *Are there higher correlations in the NEURO group with less wiggly individual profiles?*

Erasmus MC
University Medical Center Rotterdam

Simple Methods for Longitudinal Data

(a) Separate analysis for every day

⇝ *What is an appropriate model for the analysis?*

```
proc glm data = hipfracture;
model MMSE1 = AGE NEURO;
run;
proc glm data = hipfracture;
model MMSE3 = AGE NEURO;
run;
proc glm data = hipfracture;
model MMSE5 = AGE NEURO;
run;
proc glm data = hipfracture;
model MMSE8 = AGE NEURO;
run;
proc glm data = hipfracture;
model MMSE12 = AGE NEURO;
run;
```

⇝ *What are the conclusions and disadvantages of this approach?*

(b) **Analysis of Increments**

⇝ *Linear regression outcome: difference MMSE12-MMSE1*

```
data hipfracture;
set hipfracture;
increment=MMSE12-MMSE1;
run;

proc glm data = hipfracture;
model increment = AGE NEURO;
run;
```

⇝ *What is the conclusion?*

⇝ *What are the disadvantages of this approach?*

Erasmus MC
University Medical Center Rotterdam

(c) **Analysis of an Endpoint with an Additional Covariate**

- *Linear regression outcome: MMSE12*
- *Covariates: NEURO, AGE, MMSE1*

```
proc glm data = hipfracture;
model MMSE12 =MMSE1 AGE NEURO;
run;
```

- *Are the results consistent with analysis of increments?*
- *How does the interpretation of the model change as compared to the analysis of increments?*

Summary

(1) We learned simple SAS commands for data exploration

- proc freq
- proc univariate
- proc means
- proc corr

(2) We learned how to convert data into a long format

(3) We learned how to plot mean and individual evolutions

- proc sgplot

(4) We learned simple methods for repeated measurements

- proc glm

(5) We learned that **more elaborate methods are needed to make an efficient use of the data!**

Erasmus MC
University Medical Center Rotterdam

Thank you for your attention!

Questions?

v.rockova@erasmusmc.nl

Erasmus MC
University Medical Center Rotterdam