Linear Mixed Models

Ina Conrado, SAS Institute

Overview of Linear Mixed Models

What are they?

A *mixed linear model:*

- is a generalization of the standard linear model used in the GLM procedure.
- provides flexibility of modeling not only the means of your data, but their variances and covariances as well.
- useful when we need to analyze data that is non-independent, hierarchical, longitudinal, or correlated.

Uses of Linear Mixed Models

- The experimental units on which the data are measured can be grouped into clusters, and the data from a common cluster are correlated.

- Repeated measurements are taken on the same experimental unit, and these repeated measurements are correlated or exhibit variability that changes.

Uses in Longitudinal Studies

- Longitudinal studies: are widely used in a variety of fields.
- Longitudinal survey data: has multiple observations per subject across waves and subjects are grouped into clusters.
- Mixed-effects models: are able to account for probability weights, clusters, or strata.

Advantages of Mixed Models

Mixed Models are capable of:

- handling data that may violate the assumptions of standard methods
- handling missing data (missing at random)
- handling uneven spacing of repeated measures

Fixed Effects vs. Random Effects

Fixed effects: describes how population means differ across subject characteristics

Random effects: capture the variability among subjects or other units

Random Effects

- Random effects allow us to account for:
  - correlation of observations within the same subject
  - correlation of subjects within the same cluster or strata

Important Concepts

PROC MIXED using SAS

PROC MIXED:

- provides easy accessibility to numerous mixed linear models that are useful in many common statistical analyses.

- fits the specified mixed linear model and produces appropriate statistics.

Features of PROC MIXED used in SAS

- covariance structures, including variance components, compound symmetry, unstructured, AR(1), Toeplitz, spatial, general linear, and factor analytic
- GLM-type grammar, by using MODEL, RANDOM, and REPEATED statements for model specification and CONTRAST, ESTIMATE, and LSMEANS statements for inferences
- appropriate standard errors for all specified estimable linear combinations of fixed and random effects, and corresponding *t* and *F* tests
- subject and group effects that enable blocking and heterogeneity, respectively
- REML and ML estimation methods implemented with a Newton-Raphson algorithm
- capacity to handle unbalanced data
- ability to create a SAS data set corresponding to any table

Primary Assumptions in PROC MIXED used in SAS

- The data are normally distributed (Gaussian).
- The means (expected values) of the data are linear in terms of a certain set of parameters.
- The variances and covariances of the data are in terms of a different set of parameters, and they exhibit a structure matching one of those available in PROC MIXED.

Covariance Structures

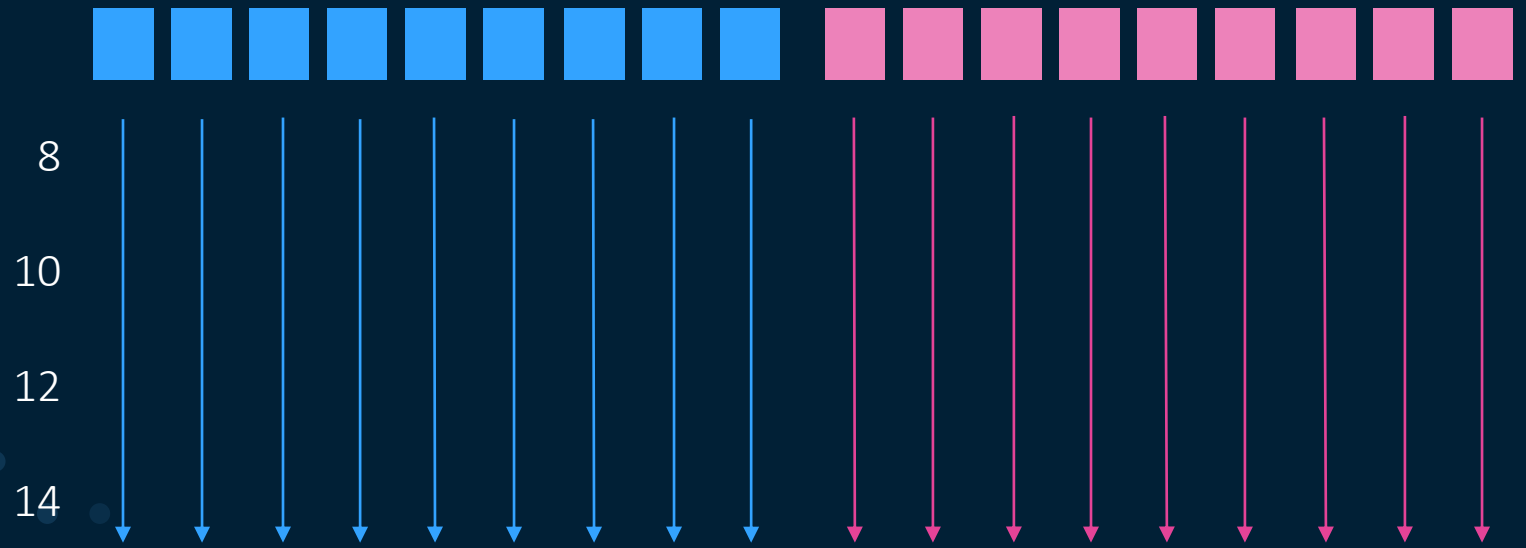
Structure	Description	# of Parameters	{i,j}th element
AR(1)	Autoregressive(1)	2	$\sigma_{ij} = \sigma^2 \rho^{\lvert i-j \rvert}$
CS	Compound Symmetry	2	$\sigma_{ij} = \sigma_1 + \sigma^2 1(i=j)$
UN	Unstructured	t(t+1)/2	$\sigma_{ij} = \sigma_{ij}$
TOEP	Toeplitz	t	$\sigma_{ij} = \sigma_{\lvert i-j \rvert+1}$
VC	Variance Components	q	$\sigma_{ij} = \sigma_k^2 1(i=j)$ and i corresponds to the kth effect
ARH(1)	Heterogeneous AR(1)	t+1	$\sigma_{ij} = \sigma_i \sigma_j \rho^{\lvert i-j \rvert}$
CSH	Heterogeneous CS	t+1	$\sigma_{ij} = \sigma_i \sigma_j [\rho 1(i \neq j) + 1(i=j)]$
TOEPH	Heterogeneous TOEP	2t-1	$\sigma_{ij} = \sigma_i \sigma_j \rho_{\lvert i-j \rvert}$

Case Study

Evaluation of Growth in Males and Females over Time

Generating Dataset of Interest

- Dataset contains growth measurement for 11 girls and 16 boys at the ages of 8,10, 12, and 14.



PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

Gender: different intercept for girls and boys

PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

Age: Overall linear growth trend

PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

Age*Gender: Different Slopes over Time

PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

Takes advantage of the assumption that observations are ordered similarly for each subject

PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

Specifies an unstructured block design for each subject=Person

PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

Estimation by maximum likelihood

PROC MIXED using SAS

Model Building

```
proc mixed data=WORK.MIXEDMODELS_SASDAG method=ml plots=(residualPanel)
        alpha=0.05;
    class Person Gender;
    model y=Age Gender Age*Gender / solution;
    repeated / type=UN subject=Person r;
run;
```

Requests the display of the fixed-effects solution vector.

Mixed Models using R and Python

Common Packages to fit Mixed Models

Some of the most common R packages for fitting linear mixed models are:

- nlme (lme)
- lme4 (lmer)

Option in Python:

- Statsmodels

Other packages in SAS:

- proc glimmix
- proc nlmixed

Conclusions

Further Reading

- https://support.sas.com/resources/papers/proceedings/proceedings/sugi25/25/aa/25p020.pdf
- https://go.documentation.sas.com/?activeCdc=pgmsascdc&cdcId=sasstudiocdc&cdcVersion=5.2&docsetId=statug&docsetTarget=statug_mixed_toc.htm&locale=en
- https://go.documentation.sas.com/?docsetId=casactstat&docsetTarget=casactstat_mixed_example01.htm&docsetVersion=8.3&locale=en
- https://support.sas.com/resources/papers/proceedings/proceedings/sugi30/198-30.pdf

Thanks!

 Ina.Conrado@sas.com

 linkedin.com/in/ina-conrado/

sas.com