Proc de Jour - Analysis for Linear Models with Continuous Response Data Using

Proc Mixed

GHSUG Presentation
April 23, 2010

by
David Borth
dmborth@gmail.com
http://ca.linkedin.com/in/davidborth

Outline

- What are mixed models – examples, math
- Basic syntax
- Class, model, random and repeated statements
- Simple repeated measures example with code
- Sample output
- G side vs. R side modeling
- Longitudinal model example with random coefficients
- BLUPs and EBLUPs
- Heterogeneous variances
- Conclusion

What Is Proc Mixed?

- An important Proc which significantly generalizes Proc GLM to handle linear mixed models
  - For continuous response data (not count data)

Examples of Mixed Models

- Split plot designs
  - e.g. fertilizer as main plot, varieties as subplots
- Components of variance models
  - e.g. measurement system study with different laboratories, days, analysts, etc.
- Repeated measures (often longitudinal studies)

What is a Mixed Model?

Generalize from General Linear Model:

$$y = X\beta + \varepsilon$$

$$\varepsilon \sim \mathrm{N}(0, \mathrm{I}\sigma^2)$$

To Linear Mixed Model:

$$y = X\beta + \mathrm{Zu} + \varepsilon$$

$$\mathrm{u} \sim N(0, G)$$

$$\varepsilon \sim \mathrm{N}(0, \mathrm{R})$$

$$\mathrm{Cov}(\mathrm{u}, \varepsilon) = 0$$

Allows more than one random effect
and allows that ε's have covariance structure

Proc Mixed - Syntax

PROC MIXED < Data=, Method=REML (*default*), other options > ;

BY variables ;

CLASS variables ;

ID variables ;

MODEL dependent = < fixed-effects > < / options > ;

RANDOM random-effects < / options > ;

REPEATED < repeated-effect > < / options > ;

PARMS (value-list) ... < / options > ;

PRIOR < distribution > < / options > ;

CONTRAST 'label' < fixed-effect values ... >
< | random-effect values ... > , ... < / options > ;

ESTIMATE 'label' < fixed-effect values ... >
< | random-effect values ... >< / options > ;

LSMEANS fixed-effects < / options > ;

MAKE 'table' **OUT=**SAS-data-set ;

WEIGHT variable ;

Class, Model ($X\beta$), Random (Zu) & Repeated (R) Statements

*Class as in Proc GLM defines variable as categorical:

Class A B C time individual ;

Model y=A B A*B x/ ddfm=kr; *ddfm = denominator df method;

Random intercept x/ type=un subject=individual; *un=unstructured;

Random C / type=vc; *vc=variance component;

Repeated time/subject=individual type=AR(1);

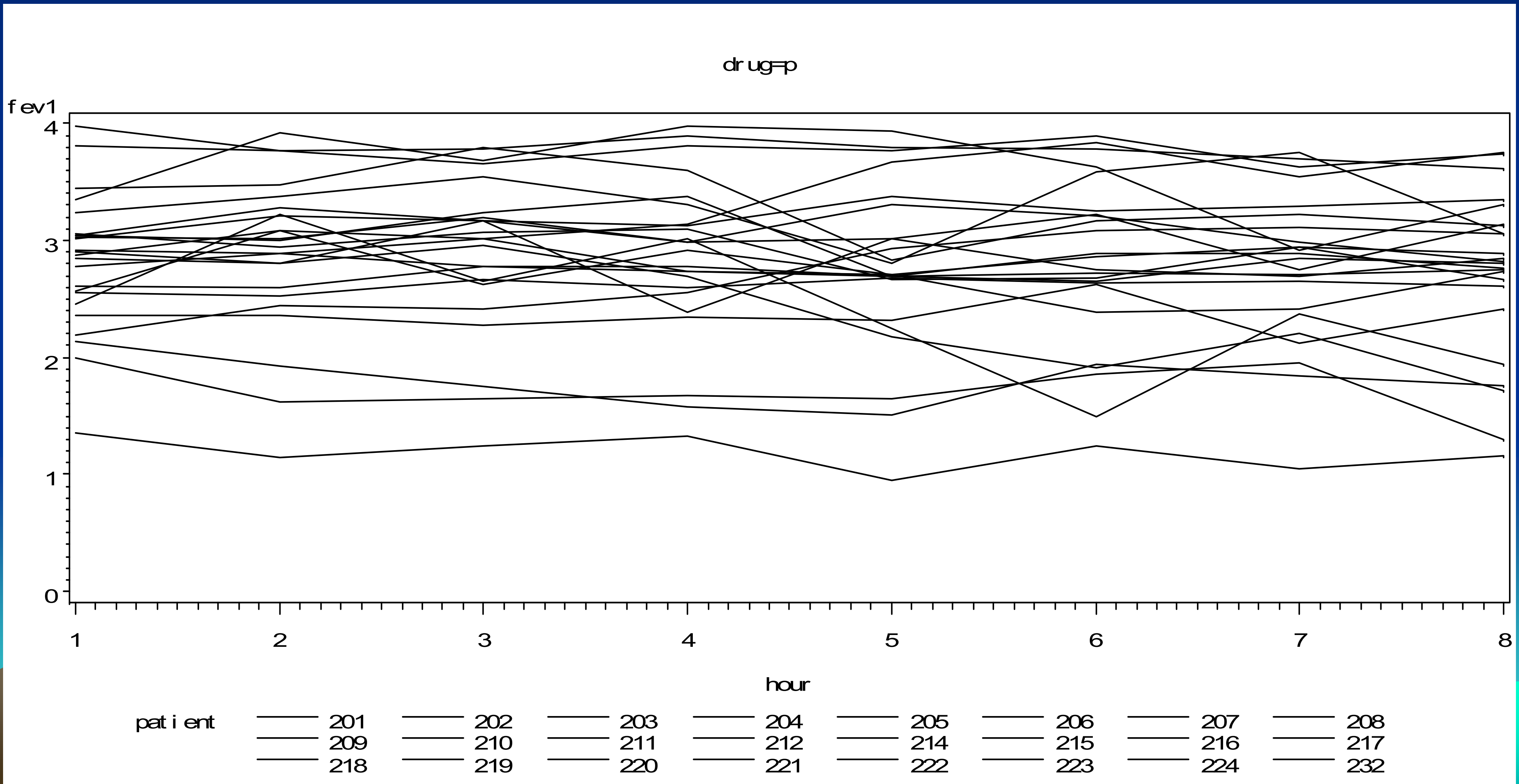
Many more options exist for model, random and repeated statements!

Many more types exist for the repeated statement!

Repeated Measures – Example 1

Effect of Three Drugs on Respiratory Ability of Asthma Patients

- **Respiratory ability (Fev1) was measured at one hour intervals**
- **3 x 24 patients randomly assigned to one of three drugs**
- **The drugs were placebo (P), standard (A) and experimental (C)**



Sample Code for Analysis

```
*Data set variables are patient drug hour fev1;
proc mixed data=fev1uni;
   class drug patient hour;
   model fev1 = drug hour drug*hour;
   random patient(drug);
   repeated hour / subject=patient(drug)
                           type=ar(1);
run;
```

SAS Output

The Mixed Procedure
Model Information

Data Set	**WORK.FEV1UNI**
Dependent Variable	**fev1**
Covariance Structures	**Variance Components, Autoregressive**
Subject Effect	**patient(drug)**
Estimation Method	**REML**
Residual Variance Method	**Profile**
Fixed Effects SE Method	**Model-Based**
Degrees of Freedom Method	**Containment**

SAS Output - Continued

```
                     Class Level Information

          Class        Levels    Values

          drug              3    a c p
          patient          24    201 202 203 204 205 206 207
                                 208 209 210 211 212 214 215
                                 216 217 218 219 220 221 222
                                 223 224 232
          hour              8    1 2 3 4 5 6 7 8


                         Dimensions

              Covariance Parameters             3
              Columns in X                     36
              Columns in Z                     72
              Subjects                          1
              Max Obs Per Subject             576


                   Number of Observations

            Number of Observations Read             576
            Number of Observations Used             576
            Number of Observations Not Used           0
```

SAS Output - Continued

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
patient(drug)		0.4145
AR(1)	patient(drug)	0.5420
Residual		0.08337

Fit Statistics

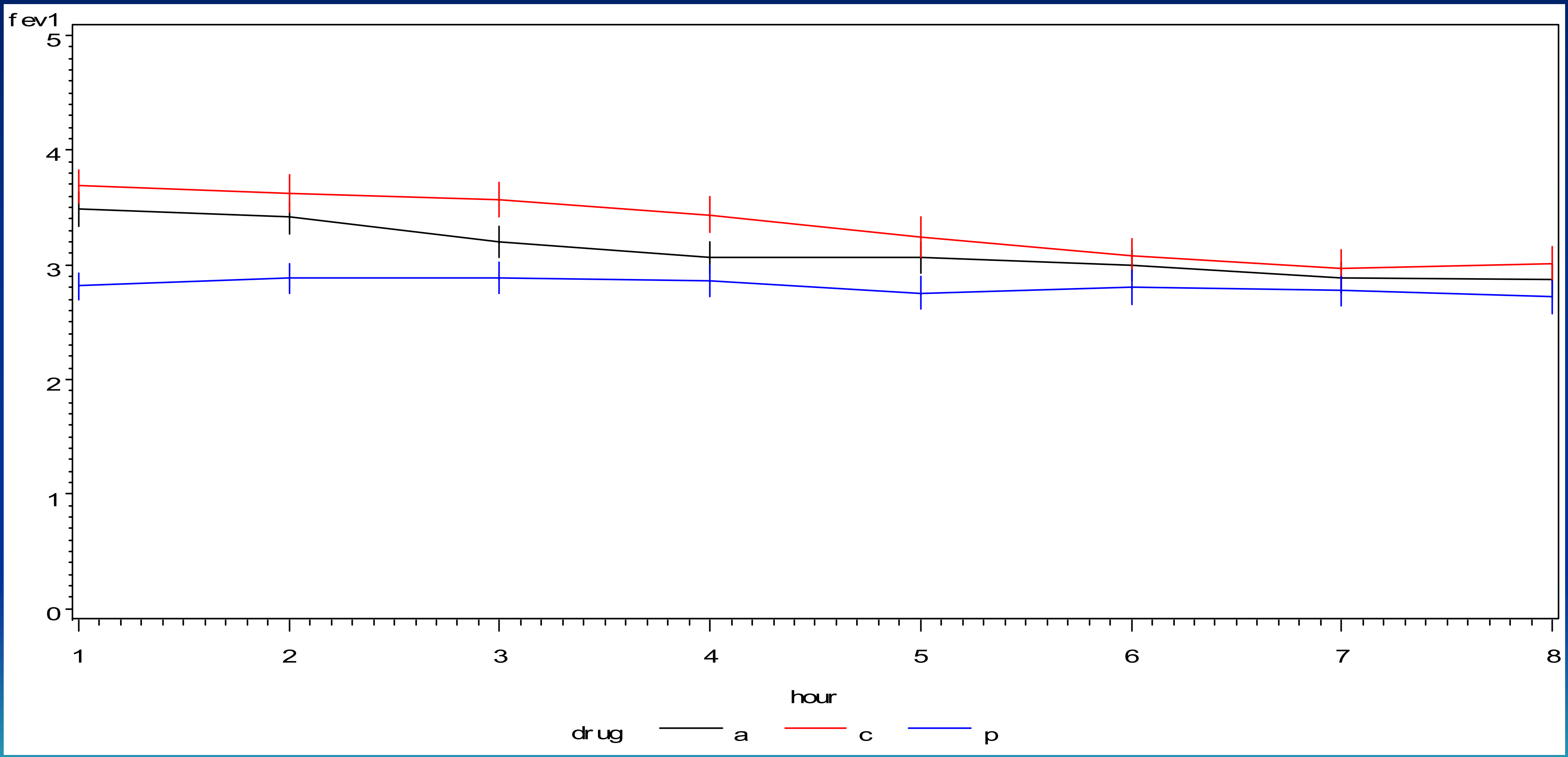
-2 Res Log Likelihood	296.0
AIC (smaller is better)	302.0
AICC (smaller is better)	302.1
BIC (smaller is better)	308.9

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
drug	2	69	3.64	0.0315
hour	7	483	17.03	<.0001
drug*hour	14	483	3.93	<.0001

Plot of Means by Treatment

Bars Show 1 Std. Error of Means



G Side vs. R Side Modeling

$$y \mid u \sim N(X\beta + \mathrm{Zu}, \mathrm{R})$$

$$\mathrm{y} \sim \mathrm{N}(\mathrm{X}\beta, V)$$

$$V = Var(y) = ZGZ' + R$$

$$u \sim N(0, G)$$

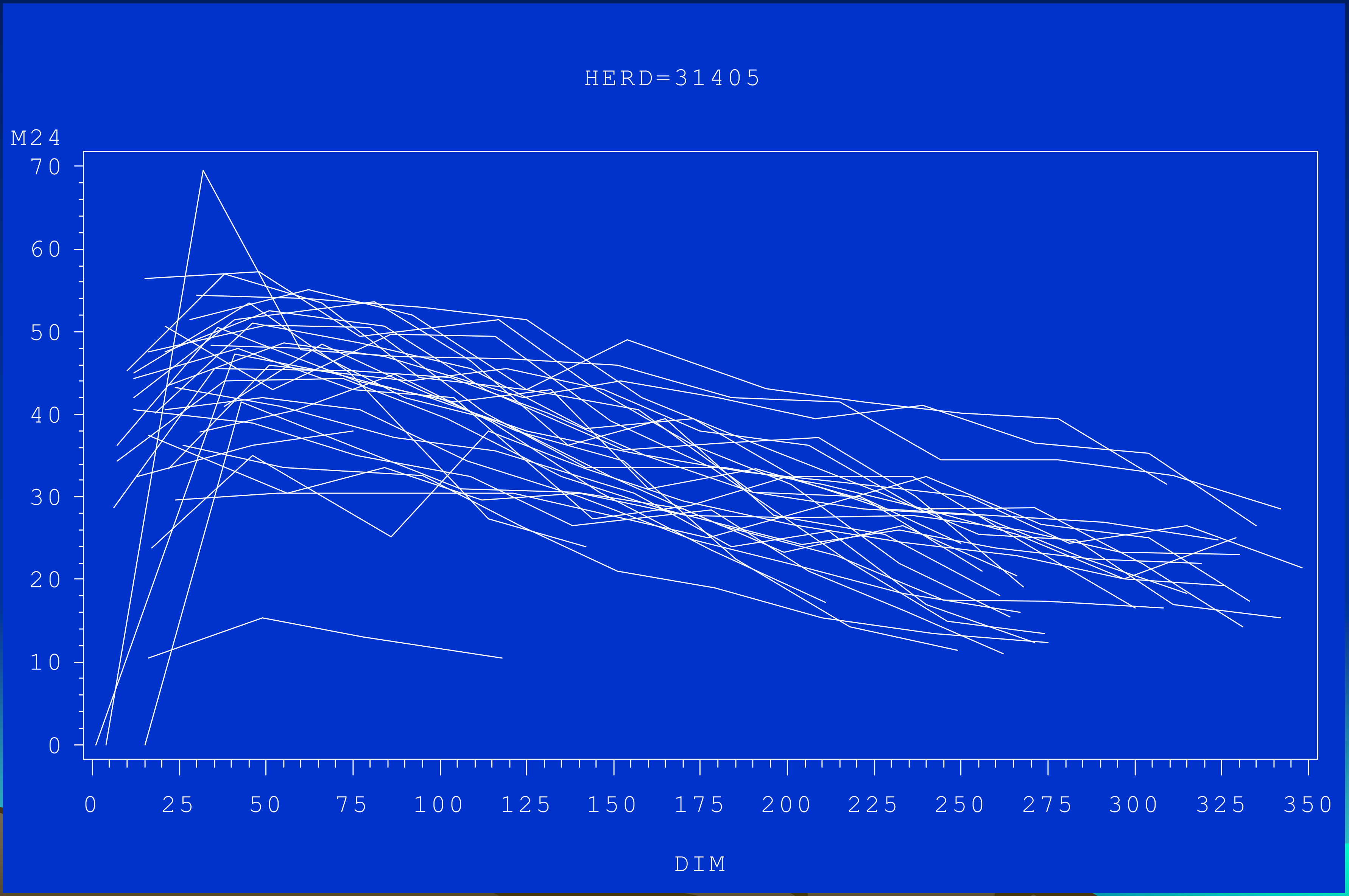
- Sometimes random effects (u) are not modeled directly, but instead are incorporated into the (modified) R matrix = ZGZ’+R
-Then in Proc Mixed there is a Repeated statement, but no Random statement
- There are a large number of type= choices on the repeated statement for R side modeling

Example 2: Relationship Between Health Traits and Milk, Fat and Protein Yields.

A Longitudinal Analysis using Random Coefficients.

- Production measurements from a single lactation per cow - 1292 cows in 53 herds
- Average of 8 production measurements per lactation
- Approximately 10,000 measurements of 24 hour milk, fat and protein yield, overall

Longitudinal Data
24 Hour Milk Yields vs. Days in Milk



Statistical Modeling - Wood's Model with Random Coefficients

$$\ell n(M24_{ij}) = a_{0i} + a_{1i}DIM_{j(i)} + a_{2i}\ell n(DIM_{j(i)}) + \varepsilon_{ij}$$

$$a_{0i} = \beta_{00} + \sum_{k=1}^{K} \beta_{0k} x_{ki} + \delta_{0i}$$

$$a_{1i} = \beta_{10} + \sum_{k=1}^{K} \beta_{1k} x_{ki} + \delta_{1i}$$

$$a_{2i} = \beta_{20} + \sum_{k=1}^{K} \beta_{2k} x_{ki} + \delta_{2i}$$

- Conceptually, coefficients for each cow are modeled as functions of the model fixed effects
- Each coefficient has random residuals at the cow level which are most likely not independent

See J. D. Singer – Use of SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models and Individual Growth Models. J. Ed. Behav. Stats, 1998, 24, No.4 323-355.

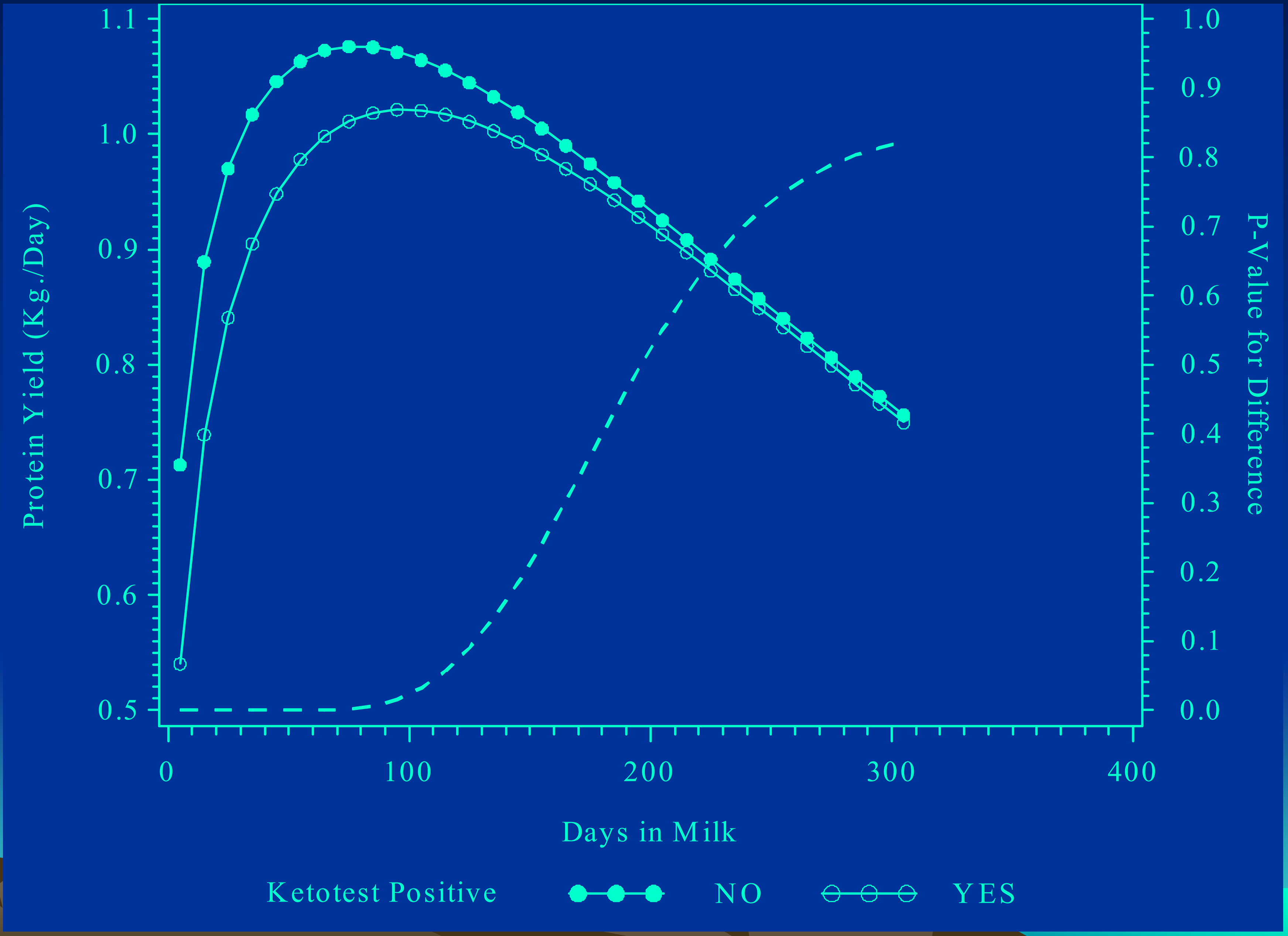
Fixed Effects in Model

1. DIM (Days in Milk) and ln(DIM)
2. Health trait effects and parity(lactation number): treated as categorical/indicator variables
3. Seasonality terms
4. Herds

Random Effects in Model

1. The Wood's model coefficients for each cow
   - Covariance expected between residuals of coefficients
2. The within cow residuals
   - auto-correlated in time for each cow
   - AR(1) model fitted

Protein Yield vs. DIM for Ketotest +ve and –ve Fitted Curves and p-values



SAS Code for Dairy Example

```
proc mixed data=c.logtrans2 method=MIVQUE0 cl;

class herd cow lact_cat scc_ls_rnd ketotest DA
   lameness ketosis RP milk_fever;

model lp24=dim ldim ketotest dim*ketotest
   ldim*ketotest herd dim*lact_cat ldim*lact_cat
   ketosis*dim ketosis*ldim dim*SCC_ls_rnd
   ldim*SCC_ls_rnd lameness*dim lameness*ldim
   rp*dim rp*ldim dim*DA ldim*DA sintm costm
  / ddfm=kr residuals influence(effect=cow(herd));
```

SAS Code for Dairy Example – Cont'd

```
random intercept dim ldim/ type=un
                                    subject=cow(herd);
*type= un allows covariance among the random
   coefficients (default would not);

repeated testnum/ subject=cow(herd)
                      type=ar(1);
```

SAS Code for Dairy Example – Cont'd

lsmeans ketotest/at (dim ldim) =(**5** 1.6094)

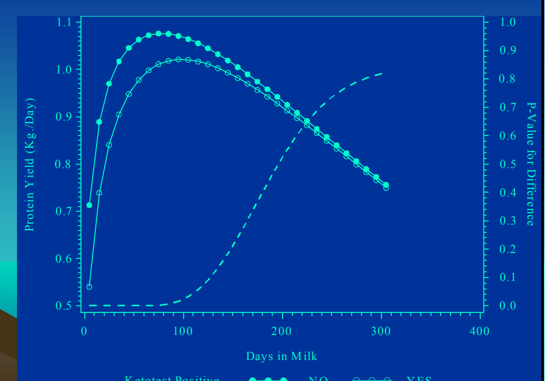
…………………………………………………….;

lsmeans ketotest/diff at (dim ldim) =(30**5 5.7203**);

lsmeans ketotest/diff at (dim ldim) =(**5** 1.6094);

…………………………………………………….;

lsmeans ketotest/diff at (dim ldim) =(30**5 5.7203**);



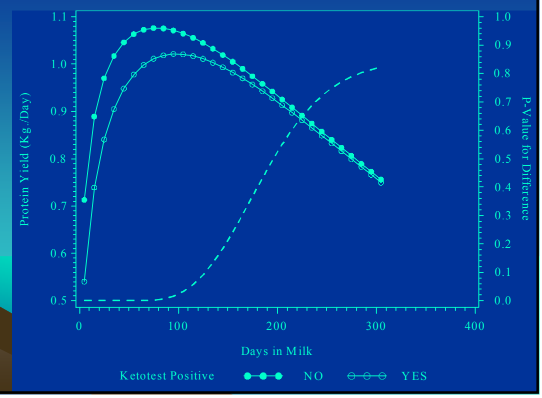
SAS Code for Dairy Example – Final

contrast 'ketotest' ketotest **1 -1**,dim*ketotest **1 -1**, ldim*ketotest **1 -1**;

The contrast statement provides an overall test of whether the curves are different for ketotest +ve or –ve.

estimate 'ketotest average' ketotest -**1 1** ketotest*dim -**155 155** ketotest*ldim -**2.445123643 2.445123643**;

The estimate statement allows us to test whether the curves are different on average over a lactation



BLUPs and EBLUPs

- For random factors, Proc Mixed will generate **shrinkage estimates** called Estimated Best Linear Unbiased Predictors (EBLUPs)
- If variances known exactly they are called BLUPs

Heterogeneous Variances (Group=,Local=)

- Group=<groupname1,..> is a potential option on both the random and repeated statements
- Results in different variance or covariance estimates for different levels of <groupname1,..>
- Local= is an option on the repeated statement Local=EXP(x) fits $Var(\varepsilon_{\mathrm{i}}) = \sigma^2 \exp(x_i \gamma)$
- Local=POM(*POM-data-set*) fits a power of the mean (predicted) value model for the variance

Conclusion

- Proc Mixed is a powerful procedure for linear mixed models with a continuous response
- Be careful with denominator degrees of freedom: model …./ddfm= ; See SUGI paper 262-26
- Be careful with random … / type= ; for random coefficient models
- For non-linear models see Proc Nlmixed
- For count data see Proc Glimmix