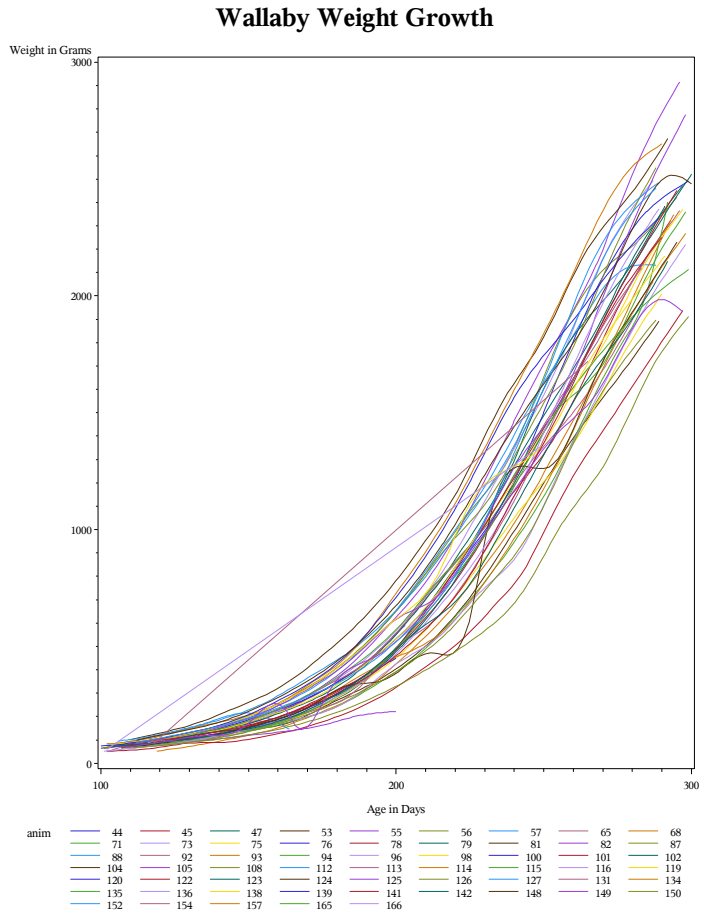
3/25/2010 36-402/608 ADA-II H. Seltman
Breakout #17 Results

This problem coems from http://www.statsci.org/data/oz/wallaby.html.

The data give growth measurements on Tammar wallabies (Macropus eugenii). We will focus on the pattern of change in weight in grams (original variable is tenths of grams) between ages 100 and 300 days. Other potentical explanatory variables are gender and location.

Here is some EDA using separate splines for each animal:



Question 1: What fixed effect model would you fit? What random effects would you consider including? Using your knowledge of biology and statistics, why would this analysis be harder for birth through death?

Here are SAS results for a rich fixed model with just a random intercept:

```
title "Handout 17 Wallaby Data";
data wallaby;
  infile "wallaby.dat" firstobs=2;
  input anim sex loca$ leng head ear arm leg pes tail weight age;
  grams = weight/10;
  male = 1;
  if sex=2 then male=0;
  drop leng head ear arm leg pes tail sex weight;
  if age<100 OR age>300 then delete;
  daysC = age-100;
  daysC2 = daysC*daysC;
  daysC3 = daysC*daysC2;
run;

proc print data=here.wallaby(obs=5);
run;

title2 "EDA";
proc freq;
  tables loca male;
run;
proc univariate;
  var age grams;
run;

title2 "Rich fixed effects + random intercept";
proc mixed covtest;
  class loca male;
  model grams = daysC|male daysC2|male daysC3|male loca;
  random int / subject=anim;
run;

### The log:
NOTE: Convergence criteria met.
NOTE: The PROCEDURE MIXED printed pages 3-4.
```

```
### Key results:
             Rich fixed effects + random intercept

                      The Mixed Procedure

                      Model Information
     Data Set                        HERE.WALLABY
     Dependent Variable              grams
     Covariance Structure            Variance Components  >> meaningless for RI only
     Subject Effect                  anim
     Estimation Method               REML
     Residual Variance Method        Profile           >>
     Fixed Effects SE Method         Model-Based       >> Highly technical info
     Degrees of Freedom Method       Containment       >>


                  Class Level Information
        Class    Levels    Values
        loca         12    "G" "H1" "H12" "H2" "H3" "H7"
                           "H8" "H9" "Ha" "Hb" "K" "W"
        male          2    0 1

                       Dimensions
              Covariance Parameters            2
              Columns in X                    24
              Columns in Z Per Subject         1
              Subjects                        59
              Max Obs Per Subject             16


                  Number of Observations
           Number of Observations Read           600
           Number of Observations Used           600
           Number of Observations Not Used         0
```

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	7598.18708672	
1	2	7357.02282769	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	anim	11806	2939.27	4.02	<.0001
Residual		11624	710.23	16.37	<.0001

Fit Statistics

-2 Res Log Likelihood	7357.0
AIC (smaller is better)	7361.0
AICC (smaller is better)	7361.0
BIC (smaller is better)	7365.2

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
daysC	1	534	24.07	<.0001
male	1	534	0.04	0.8344
daysC*male	1	534	1.90	0.1681
daysC2	1	534	89.82	<.0001
daysC2*male	1	534	5.47	0.0197
daysC3	1	534	2.25	0.1345
daysC3*male	1	534	8.91	0.0030
loca	11	534	0.74	0.6991

Question 2: Explain everything except "highly technical" and AICC.

```
/* With more than one random effect (here, random int. and slope) use
   TYPE=UN(STRUCTURED) to allow correlated random effects. */
title2 "Rich fixed effects + random intercept + random time";
  proc mixed covtest;
  class loca male;
  model grams = daysC|male daysC2|male daysC3|male loca;
  random int daysC/ subject=anim type=UN;
run;

### The log:
NOTE: Convergence criteria met.

### Selected results:
                      Model Information
     Covariance Structure         Unstructured

                          Dimensions
               Covariance Parameters             4
               Columns in X                     24
               Columns in Z Per Subject          2

                    Iteration History
                Convergence criteria met.

                  Covariance Parameter Estimates
                                    Standard        Z
  Cov Parm    Subject   Estimate      Error    Value       Pr Z
  UN(1,1)     anim       1293.45     553.72     2.34     0.0097 << rand. int.
  UN(2,1)     anim      -53.1466    14.6732    -3.62     0.0003 << cov
  UN(2,2)     anim        2.2014     0.4606     4.78     <.0001 << rand. slope
  Residual                3950.47     250.82    15.75     <.0001

                     Fit Statistics
               BIC (smaller is better)        6832.4
```

Question 3: Compare the models. Calculate the estimated correlation of the intercept and slope: UN(2,2)/sqrt(UN(1,1))/sqrt(2,2).

```
title2 "Rich fixed effects + random intercept + random time and T^2";
proc mixed covtest;
  class loca male;
  model grams = daysC|male daysC2|male daysC3|male loca;
  random int daysC daysC2/ subject=anim type=UN;
run;

### The log:
WARNING: Did not converge.
```

Question 4: What does this code model?

```
                    Dimensions
          Covariance Parameters             7
          Columns in X                     24
          Columns in Z Per Subject          3

                  Iteration History
  Iteration    Evaluations    -2 Res Log Like       Criterion
          0              1      7598.18708672
        ...
         50              1      6728.73773631      0.00001680
            WARNING: Did not converge.

              Covariance Parameter Values
                   At Last Iteration
           Cov Parm     Subject    Estimate
           UN(1,1)      anim         350.13
           UN(2,1)      anim        -52.6824
           UN(2,2)      anim          3.2288
           UN(3,1)      anim          0.1444
           UN(3,2)      anim         -0.01005
           UN(3,3)      anim        0.000077
           Residual                  3174.17
```

The usual next step is to let the computer try harder to converge at the maximum of the likelihood. We can add these to options to the MIXED statement: MAXITER=200 MAXFUNC=600. Since this still doesn't converge (with 4 times as many iterations), we can conclude that this is probably a bad model. The very small value of the estimated variance of the curvature, UN(3,3) also suggests that this is a bad model, i.e., there is essentially no animal-to-animal variation in the curvature.

Now we try the AR(1) serial correlation model. Because there is unequal spacing, we use the spherical-power correlation structure for the R matrix, which reduces to AR(1) in the case of equal spacing.

```
title2 "Rich fixed effects + RI + random time + spatial(pow)";
proc mixed covtest;
  class loca male;
  model grams = daysC|male daysC2|male daysC3|male loca;
  random int daysC / subject=anim type=UN;
  repeated / subject=anim type=sp(pow)(daysC);
run;

### The log:
NOTE: Convergence criteria met.
NOTE: Estimated G matrix is not positive definite.

### The results:
                          Dimensions
               Covariance Parameters             4
               Columns in X                     24
               Columns in Z Per Subject          2

                   Convergence criteria met.

                 Covariance Parameter Estimates

                                      Standard        Z
  Cov Parm    Subject   Estimate      Error    Value       Pr Z
  UN(1,1)     anim       1293.45     553.72     2.34     0.0097
  UN(2,1)     anim      -53.1466    14.6732    -3.62     0.0003
  UN(2,2)     anim        2.2014     0.4606     4.78     <.0001
  Residual               3950.47     250.82    15.75     <.0001

                         Fit Statistics
               BIC (smaller is better)        6832.4
```

Question 5: What does "not positive definite" mean and what does that mean for an estimated variance-covariance matrix?

Now we drop the random intercept and verify that the BIC is best and that the G matrix is valid. Finally we switch to the PROC MIXED option METHOD=ML, then use backward selection with BIC to drop un-needed terms. **Remember not to drop terms that are significant when combined with other terms in an interaction!!**

Here is our best model (finally, back to REML):

```
title2 "REML: Final model with solution and residual plots";
/* Save diagnostics to a pdf file: */
ods graphics on / imagename="ResNoRI" imagefmt = pdf;
proc mixed covtest method=REML
plots=studentpanel(conditional);
  class male;
  model grams = daysC daysC2|male daysC3|male / solution;
  random daysC/ subject=anim;
  repeated / subject=anim type=sp(pow)(daysC);
run;
ods graphics off;

NOTE: Convergence criteria met.
```

Dimensions

Covariance Parameters	3
Columns in X	10
Columns in Z Per Subject	1

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
daysC	anim	1.3256	0.2981	4.45	<.0001
SP(POW)	anim	0.9813	0.003016	325.34	<.0001
Residual		6884.47	969.01	7.10	<.0001

Fit Statistics

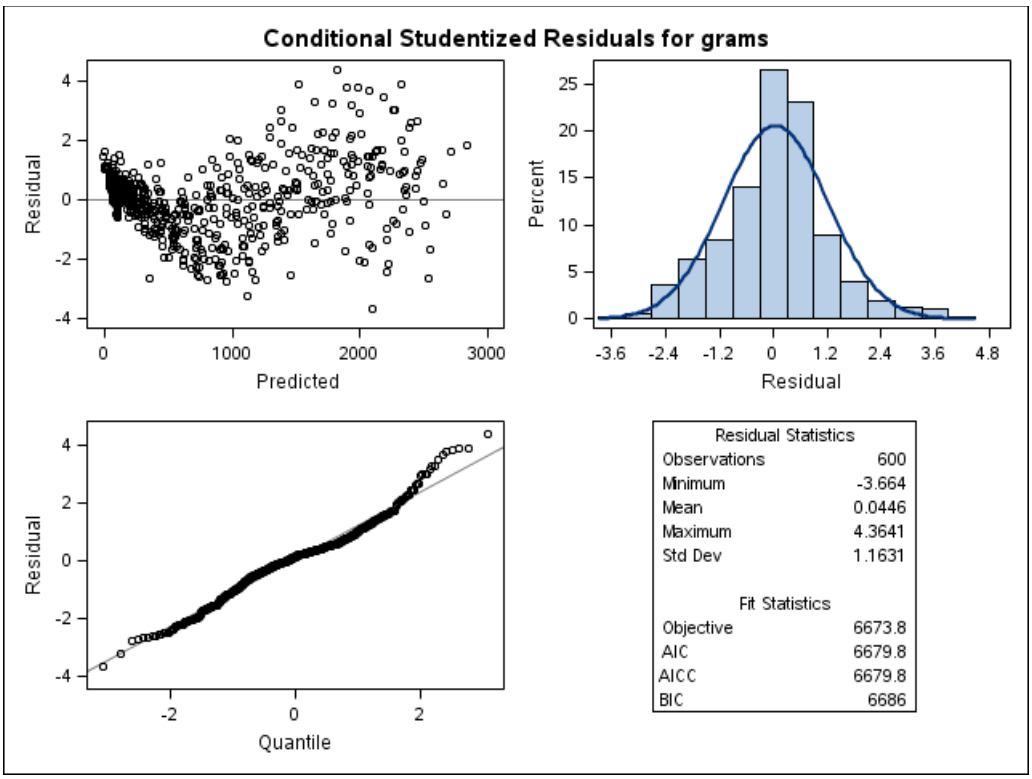
BIC (smaller is better)	6686.0

Solution for Fixed Effects

Effect	male	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		100.81	16.9723	535	5.94	<.0001
daysC		-3.6295	0.5736	58	-6.33	<.0001
daysC2		0.08086	0.007185	535	11.26	<.0001
male	0	0.2440	20.1932	535	0.01	0.9904
male	1	0	.	.	.	.
daysC2*male	0	0.02009	0.005269	535	3.81	0.0002
daysC2*male	1	0	.	.	.	.
daysC3		-8.79E-6	0.000025	535	-0.35	0.7241
daysC3*male	0	-0.00011	0.000023	535	-4.95	<.0001
daysC3*male	1	0	.	.	.	.

Question 6: What do all the estimated parameter mean?

And here are the diagnostics:



Question 7: Can you say "Oh, shit!"?