Example dataset

ACTG 320 (Hammer, Squires, et al., 1997) was a randomized, double-blind, placebo-controlled trial comparing a three-drug regimen (indinavir, lamivudine, and either zidovudine or stavudine) with a two-drug regimen (zidovudine and lamivudine) in HIV-infected adults with CD4 counts $\leq 200$ and at least 3 months of prior zidovudine therapy. The 1156 randomized patients were stratified according to their CD4 count ($\leq 50$ cells/mm$^3$ or $50-200$ cells/mm$^3$) at study entry. The primary endpoint was occurrence of an AIDS-defining event (according to the CDC definition) or death. In addition, blood specimens were collected at baseline and at weeks 4, 8, 24, and 40 during follow-up for analysis of CD4 counts and viral load. The ACTG 320 dataset available for purchase from the National Technical Information Service includes clinical endpoints and viral load for all patients but viral load data on only 198 patients who were randomly selected for a virology substudy.

Example dataset, continued

- includes the 198 patients who have RNA data
- variables are
  
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>1/0 treatment group indicator</td>
</tr>
<tr>
<td>strat</td>
<td>1/0 stratification group indicator</td>
</tr>
<tr>
<td>rna1</td>
<td>week 0 RNA</td>
</tr>
<tr>
<td>rna2</td>
<td>week 4 RNA</td>
</tr>
<tr>
<td>rna3</td>
<td>week 8 RNA</td>
</tr>
<tr>
<td>rna4</td>
<td>week 24 RNA</td>
</tr>
<tr>
<td>rna5</td>
<td>week 40 RNA</td>
</tr>
<tr>
<td>cd41</td>
<td>similar 5 cd4 values</td>
</tr>
<tr>
<td>cd42</td>
<td></td>
</tr>
<tr>
<td>cd43</td>
<td></td>
</tr>
<tr>
<td>cd44</td>
<td></td>
</tr>
<tr>
<td>cd45</td>
<td></td>
</tr>
<tr>
<td>obst</td>
<td>time at which clinical endpoint occurred, or last at which patient was observed and no clinical end</td>
</tr>
<tr>
<td>fail</td>
<td>1: clinical endpoint; 0: no clinical endpoint</td>
</tr>
</tbody>
</table>

- Note: no patient identifier

What we would like to do

- impute values for missing RNA and CD4 data
- calculate patient-specific rates of change of RNA by week and of CD4 by week
  
  – how will data file have to be laid out to do this?
Last-value-carried forward

- one (not terribly good) method of imputing missing values of longitudinal data
- may make sense if values are "missing at random"
  - that is, if the probability that a value is missing doesn’t depend on the value that would have been observed
  - not likely to be the case for this type of data

Options linesize = 72;

data actg320;
  infile '/group/ftp/pub/kcowles/datasets/combo1.dat' firstobs = 2;
  input trt strat rna1 rna2 rna3 rna4 cd41 cd42 cd43 cd44 cd45 obst fail;
  pid = _N_; * copy observation number into permanent variable;
run;

proc print data = actg320 (obs = 12);
title 'no arrays used';
run;

Arrays in SAS datasteps

- enable referencing a group of SAS variables by a single name and subscripts
- defined in array statements

array arrayname[number of items] name
  names or (list of values)

- exist during execution of data step
```plaintext
DATA ACTG320;
SET ACTG320;
ARRAY ARNA[5] RNA1 RNA2 RNA3 RNA4 RNA5; /* define array and make it a copy of existing variables in dataset*/
ARRAY ACD4[5] CD41 CD42 CD43 CD44 CD45;
RUN;
PROC PRINT DATA = ACTG320 (OBS = 12) ;
TITLE 'FIRST WAY OF DOING ARRAYS' ;
RUN;
PROC PRINT DATA = ACTG320 (OBS = 12) ;
TITLE 'SECOND WAY OF DOING ARRAYS' ;
RUN;
```

### First Way of Doing Arrays

<table>
<thead>
<tr>
<th>Obs</th>
<th>TRT</th>
<th>STRAT</th>
<th>RNA1</th>
<th>RNA2</th>
<th>RNA3</th>
<th>RNA4</th>
<th>RNA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4.24790</td>
<td>3.27323</td>
<td>4.05660</td>
<td>3.98290</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5.56951</td>
<td>5.10036</td>
<td>4.96781</td>
<td>5.41695</td>
<td>5.01041</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4.96314</td>
<td>5.48520</td>
<td>4.60326</td>
<td>5.29003</td>
<td>5.96755</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3.91666</td>
<td>3.53046</td>
<td>3.96881</td>
<td>2.03342</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>3.36286</td>
<td>2.29667</td>
<td>2.44560</td>
<td>3.55835</td>
<td>4.20850</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>4.65297</td>
<td>3.03342</td>
<td>2.71517</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>5.09670</td>
<td>2.87157</td>
<td>3.89856</td>
<td>2.14613</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>5.10824</td>
<td>5.14251</td>
<td>5.04452</td>
<td>2.37291</td>
<td>2.26007</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>5.52962</td>
<td>3.28625</td>
<td>2.88081</td>
<td>5.36319</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>4.97237</td>
<td>2.48001</td>
<td>2.13672</td>
<td>2.23045</td>
<td>2.30320</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>5.71936</td>
<td>3.48994</td>
<td>2.80895</td>
<td>2.28285</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>5.85708</td>
<td>3.18949</td>
<td>2.90309</td>
<td>2.02531</td>
<td>2.69984</td>
</tr>
</tbody>
</table>

### Second Way of Doing Arrays

<table>
<thead>
<tr>
<th>Obs</th>
<th>TRT</th>
<th>STRAT</th>
<th>RNA1</th>
<th>RNA2</th>
<th>RNA3</th>
<th>RNA4</th>
<th>RNA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4.24790</td>
<td>3.27323</td>
<td>4.05660</td>
<td>3.98290</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5.56951</td>
<td>5.10036</td>
<td>4.96781</td>
<td>5.41695</td>
<td>5.01041</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4.96314</td>
<td>5.48520</td>
<td>4.60326</td>
<td>5.29003</td>
<td>5.96755</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3.91666</td>
<td>3.53046</td>
<td>3.96881</td>
<td>2.03342</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>3.36286</td>
<td>2.29667</td>
<td>2.44560</td>
<td>3.55835</td>
<td>4.20850</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>4.65297</td>
<td>3.03342</td>
<td>2.71517</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>5.09670</td>
<td>2.87157</td>
<td>3.89856</td>
<td>2.14613</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>5.10824</td>
<td>5.14251</td>
<td>5.04452</td>
<td>2.37291</td>
<td>2.26007</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>5.52962</td>
<td>3.28625</td>
<td>2.88081</td>
<td>5.36319</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>4.97237</td>
<td>2.48001</td>
<td>2.13672</td>
<td>2.23045</td>
<td>2.30320</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>5.71936</td>
<td>3.48994</td>
<td>2.80895</td>
<td>2.28285</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>5.85708</td>
<td>3.18949</td>
<td>2.90309</td>
<td>2.02531</td>
<td>2.69984</td>
</tr>
</tbody>
</table>

---

**Data actg320 ;
Set actg320 ;
Array arna{s} RNA1 RNA2 RNA3 RNA4 RNA5 ;
Array acd4{s} CD41 CD42 CD43 CD44 CD45 ;
Run ;
Proc print data = actg320 (obs = 12) ;
Title 'First way of doing arrays' ;
Run ;
Proc print data = actg320 (obs = 12) ;
Title 'Second way of doing arrays' ;
Run ;
```
Do loops in SAS data steps

- enable coding a task once and having SAS execute it repeatedly
- framework
  
  ```sas
  do <...> ;
  .
  .
  .
  end ;
  ```

- in simplest form, do statement includes a "loop-counter" such as
  
  ```sas
do i = 1 to 5 ;
  ```

```sas
18:26 Sunday, June 29, 2003

data actg320lvcf ;
set actg320 ;
array arna[5] rna1 - rna5 ;

* set up last-value-carried-forward ;
do i = 2 to 5 ;
if arna[i] eq . then arna[i] = arna[i-1] ;
if acd4[i] eq . then acd4[i] = acd4[i-1] ;
end ;
run ;

proc print data = actg320lvcf (obs = 12) ;
title 'after last value carried forward' ;
run ;
```
data actg320lvf;
set actg320;
array aweek[5] (0, 4, 8, 12, 24); * define array and assign numeric values to its elements;
array arna[5] rna1 - rna5;
* set up better last-value-carried-forward;
* do not carry values forward to times later than obst;
do i = 2 to 5;
  if (arna[i] eq . and obst > aweek[i]) then arna[i] = arna[i-1];
  if (acd4[i] eq . and obst > aweek[i]) then acd4[i] = acd4[i-1];
end;
drop i; * drop loop counter;
run;

Using arrays to create a dataset with one record per week per patient

data actg320oneper;
set actg320; * might do this with actg320lvf instead;
array aweek[5] (0, 4, 8, 12, 24); * define array and assign numeric values to its elements;
array arna[5] rna1 - rna5;
do i = 1 to 5;
  week = aweek[i];
  rna = arna[i];
  cd4 = acd4[i];
  output;
end;
run;
proc print data = actg320oneper (obs = 25);
title 'actg320oneper';
run;
data actg320oneper;
set actg320oneper (keep = pid trt strat rna cd4 week);
run;

proc print data = actg320oneper (obs = 25);
title 'actg320oneper after keep';
run;

proc reg data = actg320oneper;
model rna = week;
by pid;
where pid < 5;
run;

The REG Procedure
Model: MODEL1
Dependent Variable: rna

Analysis of Variance

Source DF     SS     MS  F Value Pr > F
Model        1  0.00000676 0.00000676 0.00  0.9965
Error        2  0.54488  0.27244
Corrected Total  3  0.54488

Root MSE      0.52196 R-Square     0.0000
Dependent Mean 3.89016 Adj R-Sq -0.5000
Coeff Var      13.41736

Parameter Estimates

Variable DF      Estimate     Error  t Value Pr > |t|
### PID 2

The REG Procedure
Model: MODEL1
Dependent Variable: rna

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.06870</td>
<td>0.06870</td>
<td>0.96</td>
<td>0.3986</td>
</tr>
<tr>
<td>Error</td>
<td>3</td>
<td>0.21385</td>
<td>0.07128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>4</td>
<td>0.28254</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.26699 R-Square 0.2431
Dependent Mean 5.21301 Adj R-Sq -0.0091
Coeff Var 5.12156

Parameter Estimates

| Variable | DF | Parameter | Standard Error | t Value | Pr > |t| |
|----------|----|-----------|----------------|---------|------|---|
| Intercept| 1  | 5.34963   | 0.18337        | 29.17   | <.0001|
| week     | 1  | -0.01423  | 0.01450        | -0.98   | 0.3986|

### PID 3

The REG Procedure
Model: MODEL1
Dependent Variable: rna

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.49062</td>
<td>0.49062</td>
<td>2.53</td>
<td>0.2097</td>
</tr>
<tr>
<td>Error</td>
<td>3</td>
<td>0.58104</td>
<td>0.19368</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>4</td>
<td>1.07166</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.44009 R-Square 0.4578
Dependent Mean 5.26184 Adj R-Sq 0.2771
Coeff Var 8.36384

Parameter Estimates

| Variable | DF | Parameter | Standard Error | t Value | Pr > |t| |
|----------|----|-----------|----------------|---------|------|---|
| Intercept| 1  | 4.89673   | 0.30226        | 16.20   | 0.0005|
| week     | 1  | 0.03803   | 0.02390        | 1.59    | 0.2097|

### PID 4

The REG Procedure
Model: MODEL1
Dependent Variable: rna

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>1.35792</td>
<td>1.35792</td>
<td>2.44</td>
<td>0.2584</td>
</tr>
<tr>
<td>Error</td>
<td>2</td>
<td>1.11145</td>
<td>0.55573</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>3</td>
<td>2.46937</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.74547 R-Square 0.5499
Dependent Mean 3.36234 Adj R-Sq 0.3249
Coeff Var 22.17118

Parameter Estimates

| Variable | DF | Parameter | Standard Error | t Value | Pr > |t| |
|----------|----|-----------|----------------|---------|------|---|
| Intercept| 1  | 4.14404   | 0.62370        | 6.64    | 0.0219|
| week     | 1  | -0.13028  | 0.08335        | -1.56   | 0.2584|
The Mixed Procedure

Model Information

Data Set WORK.ACTG320ONEPER
Dependent Variable rna
Covariance Structure Variance Components
Subject Effect pid
Estimation Method REML
Residual Variance Method Profile
Fixed Effects SE Method Model-Based
Degrees of Freedom Method Containment

Dimensions

Covariance Parameters 3
Columns in X 4
Columns in Z Per Subject 2
Subjects 24
Max Obs Per Subject 5
Observations Used 103
Observations Not Used 17
Total Observations 120

Iteration History

Iteration Evaluations -2 Res Log Like Criterion
0 1 318.66503481

Solution for Fixed Effects

      Std
Effect  Estimate  Error DF  t Value Pr > |t|
Intercept  5.1153  0.2444  21  20.93 <.0001
trt -0.1875  0.2885  55  -0.65 0.5186
strat -1.0555  0.2900  55  -3.64 0.0006
week -0.06458  0.01740 55  -3.71 0.0012

Solution for Random Effects

      Std Err
Effect Subject Estimate Pred DF  t Value Pr > |t|
Intercept  1   0.05687  0.3731  55   0.15 0.8815
week  1   0.02040  0.4434  55   0.46 0.6472
week  2   0.2085  0.3672  55   0.57 0.5725
week  3   0.04076  0.03281 55  1.24 0.2194
week  4   0.08236  0.3672  55   0.22 0.8234
week  5   0.06641  0.03281 55  2.02 0.0478
week  6   0.01654  0.3682  55  -1.67 1.0033
week  7   0.0463  0.3722  55  -1.31 0.1968
week  8   0.05163  0.03295 55  1.57 0.1229
week  9   0.02646  0.3748  55  -0.07 0.9440
week  10  0.02388  0.05000 55  -0.48 0.6348
week  11  0.1119  0.3707  55   0.30 0.7640
week  12  0.02876  0.04426 55  -0.65 0.5185
week  13  0.03262  0.3672  55  -0.09 0.9295
week  14  0.04778  0.03281 55  -1.46 0.1510

The Mixed Procedure

Fit Statistics

-2 Res Log Likelihood 299.3
AIC (smaller is better) 305.3
AICC (smaller is better) 305.5
BIC (smaller is better) 308.8

The Mixed Procedure

Solution for Random Effects

      Std Err
Effect Subject Estimate Pred DF  t Value Pr > |t|
Intercept  1   0.05687  0.3731  55   0.15 0.8815
week  1   0.02040  0.4434  55   0.46 0.6472
week  2   0.2085  0.3672  55   0.57 0.5725
week  3   0.04076  0.03281 55  1.24 0.2194
week  4   0.08236  0.3672  55   0.22 0.8234
week  5   0.06641  0.03281 55  2.02 0.0478
week  6   0.01654  0.3682  55  -1.67 1.0033
week  7   0.0463  0.3722  55  -1.31 0.1968
week  8   0.05163  0.03295 55  1.57 0.1229
week  9   0.02646  0.3748  55  -0.07 0.9440
week  10  0.02388  0.05000 55  -0.48 0.6348
week  11  0.1119  0.3707  55   0.30 0.7640
week  12  0.02876  0.04426 55  -0.65 0.5185
week  13  0.03262  0.3672  55  -0.09 0.9295
week  14  0.04778  0.03281 55  -1.46 0.1510

The Mixed Procedure

Convergence criteria met.

Estimated G Matrix

Row  Effect Subject  Col1  Col2
0   Intercept  1   0.2391
1  week   1   0.003092

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>pid</td>
<td>0.2391</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>0.6756</td>
</tr>
</tbody>
</table>

The Mixed Procedure

actg320oneper after keep 13
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>21</td>
<td>0.7732</td>
<td>0.3701</td>
<td>55</td>
<td>2.09</td>
</tr>
<tr>
<td>week</td>
<td>21</td>
<td>0.07491</td>
<td>0.04424</td>
<td>55</td>
<td>1.69</td>
</tr>
<tr>
<td>Intercept</td>
<td>22</td>
<td>-0.4401</td>
<td>0.3748</td>
<td>55</td>
<td>-1.17</td>
</tr>
<tr>
<td>week</td>
<td>22</td>
<td>-0.02374</td>
<td>0.05000</td>
<td>55</td>
<td>-0.47</td>
</tr>
<tr>
<td>Intercept</td>
<td>23</td>
<td>-0.2525</td>
<td>0.3693</td>
<td>55</td>
<td>-0.68</td>
</tr>
<tr>
<td>week</td>
<td>23</td>
<td>-0.00515</td>
<td>0.03284</td>
<td>55</td>
<td>-0.16</td>
</tr>
<tr>
<td>Intercept</td>
<td>24</td>
<td>0.02324</td>
<td>0.3672</td>
<td>55</td>
<td>0.06</td>
</tr>
<tr>
<td>week</td>
<td>24</td>
<td>-0.03077</td>
<td>0.03281</td>
<td>55</td>
<td>-0.94</td>
</tr>
</tbody>
</table>

### Type 3 Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num</th>
<th>Den</th>
<th>Num</th>
<th>Den</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>1</td>
<td>55</td>
<td>0.42</td>
<td>0.5186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>strat</td>
<td>1</td>
<td>55</td>
<td>13.25</td>
<td>0.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>1</td>
<td>23</td>
<td>13.77</td>
<td>0.0012</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>