TWIN STUDIES*: NOW THAT I HAVE YOUR ATTENTION, OR HOW ONE USERS GROUP KEPT ITS CROWD AFTER LUNCH

Lynne Davidson, Medical College of Virginia

ABSTRACT

The Virginia SAS Users Group holds quarterly meetings to promote SAS software usage and proficiency. The standard meeting format includes a business meeting and one or two speakers in the morning, a lunch break, and then small group sessions and/or a panel discussion after lunch. In order to deal with the tendency for members to leave before the after-lunch sessions, the program committee attempted an innovative panel presentation. This involved a problem presentation and several solution suggestions presented by Users Group members. The general interest of the topic and the applied solutions proved effective in maintaining a higher than average attendance in the afternoon session.

INTRODUCTION

The Virginia SAS Users Group (V ASUG) is a 2-year-old state-wide association of over 100 members which holds quarterly meetings to promote SAS software usage and proficiency. Standard meeting formats have been presentations by two speakers in the morning, followed by small group sessions or a panel discussion after lunch.

In an effort to bolster after-lunch attendance, the program committee decided to try a variation on the panel discussion. First, a "problem" which might lend itself to multiple methods of resolution was solicited from the membership. Next, volunteers were enlisted to solve the problem using a small sample data set from an existing data base. The panel was announced with the meeting agenda, and a brief introductory article explaining the presentation format was included in the V ASUG newsletter.

In this new after-lunch format, the audience was given a brief overview of the application which generated the problem, presented by the member whose problem was used. Then each volunteer presented their own particular solution. Solutions included the use of PROC TRANSPOSE, PROC SQL, and a macro application, in a variety of environments (UNIX®, MVS® and PC.)

The exercise generated lively discussions afterward about the merits/pitfalls of each methodology, and assets/limitations of different versions of SAS and/or different environments. It also provided an example for a hands-on walk through of PROC SQL for users unfamiliar with the SQL procedure.

OVERVIEW OF THE PROBLEM

The Medical College of Virginia, Department of Pediatric Cardiology, is conducting a longitudinal study of twins and their parents to assess determinants and indicators of hypertension. The twins are seen for initial assessment at 9.5 years of age, and every 18 months thereafter until age 17. Parents of the twins are seen twice, at the first two assessments. Data are collected on a large array of variables, including anthropometric measures, heart rate and blood pressure at rest and under physical and mental stress, non-invasive (ultrasound) measures of heart size, capacity, and function, pulmonary function, psychological and social scales, family history of heart-related conditions, health history, exercise habits, etc. A single family’s visit can generate over 2000 variables, about 500 per parent and 800 per twin.

Data are stored in 23 individual SAS data files on an IBM 3084Q under MVS/XA. Two files contain one-time variables (name, DOB, blood typing data, etc.), with one observation per individual (figure 1.) Twenty of the data sets hold developmental data, and contain one observation per individual, per visit (figure 2.) The last data set holds multiple heart measures at different levels of stress, with one observation per individual, per visit, per stress level (figure 3.)

One component of a twin study is comparing data between identical twins and between fraternal twins. Differences between twins can be used to determine whether a trait is largely hereditary or environmental. This type of twin comparison can show three primary results. For variables influenced largely by heredity (e.g., height) identical twins will show statistically less difference between twin 1 and twin 2 than will be seen in non-identical (fraternal) pairs. For variables influenced largely by common environmental factors
(e.g., common family diet) identical twin pairs and fraternal twin pairs both show similarities between twins. For variables influenced by individual environmental factors (e.g., 'fitness' - one twin plays soccer, the other is a bookworm) both identical and fraternal pairs can show significant differences between twins.

![TWIN.MASTER]

**TWIN.MASTER**

Family and Twin Number INDEX
Name
Date of Birth
Zygosity
Sex

Figure 1. One Record per Individual

![TWIN.ANTHROPOMETRY]

**TWIN.ANTHROPOMETRY**

Family and Twin Number INDEX
Visit INDEX
Height
Weight
Blood Pressure
Resting Heart Rate

Figure 2. One Record per Individual, per Visit

**THE PROBLEM**

Panelists were asked to create a data set which would allow them to answer the question, "Do identical twins more closely resemble each other in their reactions to stress than do fraternal twins?" They were asked to calculate the difference between heart rate under physical stress (squeezing a handgrip) and heart rate at rest, for each twin, and to calculate the difference between heart rate under mental stress (doing mental arithmetic) and heart rate at rest, for each twin. The mean difference between identical twins was to be compared to the mean difference between fraternal twins.

![TWIN.HEART_MEASURES]

**TWIN.HEART_MEASURES**

Family and Twin Number INDEX
Visit INDEX
Phase INDEX
Heart Rate
Heart Mass

Figure 3. One Record per Individual, per Visit, per Phase

The objective of the panel discussion was not to look at results, but to look at the various approaches which might be used to manipulate the data into a format in which it could be analyzed. The transition from storage format to final analysis format is shown in figure 4.

The panelists were given two sample data sets, MASTER and ECHO. These data sets are documented in figure 5.

**SOLUTIONS**

The Current Code - Program 1 shows code using the approach currently employed. One drawback, especially if the list of variables to be investigated is large, is the need to itemize variables in the retain statement and when setting initial variables to missing. Another is the chance for conceptual errors in the use of FIRST and LAST variables, which can lead to frequent rerunning of the program.

```
** Program 1 **
******************************************************
CURRENT PROCEDURE FOR CONVERTING INDIVIDUAL DATA TO COMBINED FILE DATA
******************************************************
Options pagesize=66;
proc sort data=sasuser.master; by fn in;
proc sort data=sasuser.echo; by fn in phase;
* Combine master records, which include zygosity
* variable, with ECHO dataset containing heart rate
* information for various stages of mental arithmetic
* or handgrip activity.
*;
  data one;
    merge sasuser.master(keep=fn in zyg)
      sasuser.echo (keep=fn in phase echohr);
    by fn in;
```

1013
### STORAGE FORMAT

<table>
<thead>
<tr>
<th>Family</th>
<th>Twin</th>
<th>Phase</th>
<th>Heart Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>A</td>
<td>70</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>B</td>
<td>80</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>A</td>
<td>65</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>B</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>A</td>
<td>58</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>B</td>
<td>74</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>A</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>B</td>
<td>81</td>
</tr>
</tbody>
</table>

### ANALYSIS FORMAT A
**Within an Individual**

<table>
<thead>
<tr>
<th>Family</th>
<th>Twin</th>
<th>Heart Rate A</th>
<th>Heart Rate B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>65</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>58</td>
<td>74</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>63</td>
<td>81</td>
</tr>
</tbody>
</table>

### ANALYSIS FORMAT B
**Within a Family**

<table>
<thead>
<tr>
<th>Family</th>
<th>Twin1</th>
<th>Twin2</th>
<th>Twin1 Heart Rate A</th>
<th>Twin2 Heart Rate B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70</td>
<td>65</td>
<td>80</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>58</td>
<td>63</td>
<td>74</td>
<td>81</td>
</tr>
</tbody>
</table>

**Figure 4. Storage & Analysis Formats**

```plaintext
proc sort; by fn in phase;
* Convert multiple records to a single record containing:
  * resting heart rates (rhr), mental arithmetic heart rates (mhr), and handgrip heart rates (hhr),
* within twin 1 and 2.
*;
data two;
  set one; by fn in phase;
  retain rhr1 rhr2 mhr1 mhr2 hhr1 hhr2;
  if first.fn then do;
    rhr1=.; rhr2=.; mhr1=.; mhr2=.;
    hhr1=.; hhr2=.; end;
  if ln='0001' then do;
    if phase='RE1' then rhr1=echohr;
    if phase='MA1' then mhr1=echohr;
    end;
  if ln='0002' then do;
    if phase='RE1' then rhr2=echohr;
    if phase='MA1' then mhr2=echohr;
    end;
  if last.fn then output;
  * Calculate difference from resting heart rate for:
  * mental arithmetic heart rate and handgrip heart rate.
*;
data three;
  set two;
  madiff=abs((mhr1-rhr1)-(mhr2-rhr2));
  hdiff=abs((hhr1-rhr1)-(hhr2-rhr2));
*;  * Run PROC TTEST on the two differences, by:
*  zygosity.
*;
proc ttest;
  var madiff hdiff;
  class zyg;
run;
```

**PROC TRANSPOSE** - The first alternate solution used PROC TRANSPOSE in a TSO environment. This code is less tedious to write, but needs familiarity and/or frequent PROC PRINTS, LABEL and NAME statements to be able to transpose the correct variables from the correct groups.

```plaintext
* Libname OLD;
  data master ;
    set OLD.master(keep=fn zyg);
    by fn;
    if first.fn;
  data echo;
    set OLD.echo;
    if echohr;
  proc sort:
    by fn in phase echohr;
    *
      Move all heart rate readings to a single record, for each individual.
    *
    proc transpose out=echo;
      id phase;
      var echohr;
      by fn in;
    *
      Create a data set with only the differences
      between heart rates, two records per individual.
    *
    data echo;
      set;
      delta=MA1-RE1;
      type='MENTAL EXERCISE';
      output;
      delta=HG1-RE1;
      type='HANDGRIP';
      output;
      drop HG1 MA1 RE1 NAME_;;
  proc sort;
    by fn type in;
    *
      Create a data set with differences in heart rate for both individuals, by type.
    *
    proc transpose out=echo prefix=twin;
      id in;
      var delta;
      by fn type;
    data echo;
      merge echo (in=IN1) master;
```

1014
Figure 5. PROC DATASET'S Master and Echo

**Program 3**

• Subset datasets, keeping only necessary variables.
•; data master;
set old.master(keep=fn zyg);
if in='0001';
*; data echo;

**Program 4**

data master;
set old.master(keep=fn zyg);
by fn;
if in;
diff=abs(twin1-twin2);
proc sort;
by type zyg;
proc ttest;
var diff;
class zyg;
array type (2)
i j;
retain A1-A6;
if first.fn then do;
do i=1 to ~;

**Program 3**

• Subset datasets, keeping only necessary variables.
•; data master;
set old.master(keep=fn in zyg);
if in="0001";
*; data echo;
Finding volunteers to provide the alternate solutions proved to be the more difficult part of the exercise. We found that initial uncertainties about the nature of the application made several potential volunteers hesitate or decline to participate. We also found that some volunteers later withdrew from the exercise due to conflicting time demands, forcing us to find replacements on short notice.

Initially we tried to find a problem which would force the panelists to develop a specific data structure. This led to some confusion among the panelists as to the real objective of the exercise and left us using a very poor statistic not normally used in twin analyses. A clearly defined, succinct statement of the objective as "creating a specific data structure with specific variables" rather than "solving a problem" would allow more consistency in end results (the same final data structure from all panelists) as well as make it easier for volunteers to commit to the exercise. We also believe that seeing this first exercise will clarify for Users Group members what exactly is expected, and assist them to be more willing to participate in future problem panels.

CONCLUSION

The Virginia SAS Users Group found this "problem-solving" panel to be an effective method for promoting audience interest/participation in what had in the past been a less well attended part of their quarterly meetings.

AUTHOR

Lynne Davidson
Pediatric Cardiology
Medical College of Virginia
P.O. Box 26
Richmond, Virginia 23229
(804) 786-1936

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries.
* indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

* Supported by the National Institutes of Health, National Heart, Lung and Blood Institute (ROI HL31010).