The SAS ICC9 Macro

Ellen Hertzmark and Donna Spiegelman

October 4, 2010

Abstract

The %ICC9 macro is a SAS version 9 macro that computes reliability coefficients (intraclass correlation coefficients) and their 95% confidence intervals. These quantities can be calculated after first adjusting for fixed effects.

Keywords: SAS, macro, PROC MIXED, ICC, intraclass correlation coefficient, reliability, coefficient of variation, variance partition, fixed effects

Contents

1 Description

The %ICC9 macro uses `PROC MIXED` to compute intraclass correlation coefficients (ICCs, or reliability coefficients) and their 95% confidence intervals, as well as the 95% confidence intervals of the within-subject coefficient of variation (Hankinson, et al., 1995). These quantities can be calculated after first adjusing for fixed effects.

2 Invocation and Details

To call %ICC9, your program must know where to look for it. The most efficient way is to include teh following statement in your program.

```
options mautosource sasautos='/usr/local/channing/sasautos';
```

You type

```
%icc9(
  data          name of input data set
                REQUIRED
  varlist       list of variables (standard sas list, no commas)
                REQUIRED
  byvar         BY variables, if any
                OPTIONAL
  subject       name of the subject variable
                REQUIRED
  where         use like a where statement to restrict the analysis data set.
                OPTIONAL
  maxdec=4      maximum number of decimal places in printed output
                OPTIONAL
  outdat        name of a data set to which the results of %ICC9 will be
```

```
              output
              OPTIONAL
  noprint=F   print the output.
              noprint= T means 'do not print the output'
              You might want to use this if you are making an output
              data set (i.e. outdat has a value).
              OPTIONAL
  model       fixed effect variables for which you want to adjust
              before computing ICC.
              OPTIONAL
```

For the two parameters with default settings (*MAXDEC* and *NOPRINT*), the default values are given after the "=".

As with all macros, do not include in your macro call the parameters you intend to leave unspecified or those for which you want to use the default values.

3 Examples

Below we give 3 examples of %ICC9 using data from the pooling project. In contrast to the previous version of the macro, %ICC8, we now use a more efficient estimating method and can run examples on large datasets.

The variable of interest is

```
  FIBER   energy-adjusted fiber intake
```

In order to get a single variable with all the data, we use the usual Channing method:

```
data fiber ;
   merge n84_ant n86_ant n90_ant n94_ant n98_ant;
   by id;
   fiber=aofib84a;  time=1;  output;
   fiber=aofib86a;  time=2;  output;
   fiber=aofib90a;  time=3;  output;
   fiber=aofib94a;  time=4;  output;
```

```
fiber=aofib98a;  time=5;  output;
run;
```

3.1 Example 1. Basic run

Here is an example of a basic call to %ICC9.

```
title2 'basic macro run';
%icc9(data=fiber, subject=id, varlist=fiber);
```

Here are the results.

```
------------------------------------------------------------------------------

/udd/stleh/doctn/examples.icc  Program exicc9   02AUG2010   14:53    stle 1
basic macro run



Number of subjects with more than 1 measurement:  94660
These subjects have 391904 measurements (4.1 on average)

order of vbl=1.0000 VARNAME=FIBER label=' '

                                        Estimated           Estimated
                                      coefficient of      coefficient of
               #         #             reliability,        within-subject
Obs     subjects    observations      ICC (95% CI)       variance (95% CI)

 1        104360        401604      0.50 (0.50, 0.51)    0.22 (0.22, 0.22)

------------------------------------------------------------------------------
```

The titles tell you how many subjects have at least 2 measurements, how many measurements these subjects have in total, and the average number per subject. Three titles (title1, title2, title3) are available for user titles.

The output shows the number of subjects, the number of total measurements, the reliability coefficient (ICC) and its 95% confidence interval, the

coefficient of within-subject variation and its 95% confidence interval.

3.2 Example 2. Using the WHERE parameter

The ICC calculation assumes an exchangeable (compound symmetry) variance-covariance matrix for FIBER. As part of an attempt to see whether the ICC falls off as the time interval between the two measurements gets longer, we did an analysis using only the first and last periods (1 and 5). To get this subset we used the *WHERE* parameter.

The macro call is

```
title2 'using WHERE parameter';
%icc9(data=fiber, subject=id, varlist=fiber,
where=time in (1, 5));
```

The results are

```
------------------------------------------------------------------------------

/udd/stleh/doctn/examples.icc  Program exicc9   02AUG2010   14:53    stle 2
using WHERE parameter

Analysis run on subset WHERE time in (1, 5).

Number of subjects with more than 1 measurement:  65997
These subjects have 131994 measurements (2 on average)

order of vbl=1.0000 VARNAME=FIBER label=' '

                                        Estimated            Estimated
                                      coefficient of       coefficient of
              #         #              reliability,         within-subject
Obs     subjects    observations       ICC (95% CI)        variance (95% CI)

 1         98407        164404      0.34 (0.33, 0.34)    0.25 (0.25, 0.26)

------------------------------------------------------------------------------
```

When the *WHERE* parameter is used, the macro shows it in `title4`.

3.3 Example 3. Using the MODEL parameter to adjust for the fixed effect of time

Since we saw that the ICC when we only used the 2 outermost time periods was lower than the ICC using all the time periods (.34 (.33, .34), compared to .50 (.50, .51)), we decided to adjust for the fixed effect of time.
The macro call is

```
title2 'using MODEL parameter';
%icc9(data=fiber, subject=id, varlist=fiber,
model=time);
```

The results are

```
------------------------------------------------------------------------------

/udd/stleh/doctn/examples.icc  Program exicc9   02AUG2010   14:53    stle 3
using MODEL parameter


Adjusting for:  time
Number of subjects with more than 1 measurement:  94660
These subjects have 391904 measurements (4.1 on average)

order of vbl=1.0000 VARNAME=FIBER label=' '

                                        Estimated           Estimated
                                      coefficient of      coefficient of
               #         #             reliability,        within-subject
Obs     subjects    observations      ICC (95% CI)       variance (95% CI)

 1        104360        401604      0.53 (0.53, 0.54)    0.24 (0.24, 0.24)
------------------------------------------------------------------------------
```

Adjusting for time increased the point estimate of the ICC a bit.
When the *MODEL* parameter is used, the macro gives the fixed effects in `title5`.

4 Computational Methods

All computations use all observations, not just those from subjects with repeated measures. The number of subjects with repeated measures, and the number of observations they have are given because some users find this information useful.
Since maximum likelihood estimation is used, it is possible for some of the variance components to be negative.

Since %ICC9 uses a more efficient estimating method than previous versions of the ICC macro, it can be used on large datasets. In fact, all 3 examples ran in 11.1 seconds on cilea13.

5 References

The reference for the computation of the confidence intervals for the ICC and the within-subject coefficient of variation is

Hankinson SE, Manson JE, Spiegelman D, Willett WC, Longcope C, Speizer FE. Reproduciblity of plasma hormone levels in postmenopausal women over a two to three year period. *Cancer Epidemiology, Biomarkers and Prevention*, 1995; 4:649-654.

6 Credits

Written by Ellen Hertzmark and Donna Spiegelman for the Channing Laboratory (with help from Peter Gaccione, Edmonde Kabagambe, Eric Tchetgen, and Sally Skinner). Questions can be directed to Ellen Hertzmark, `stleh@channing.harvard.edu`, (617) 432-4597.