A SAS® Macro for Deming Regression

Allison M. Deal, UNC Lineberger Cancer Center, Chapel Hill, NC
Virginia W. Pate, Surgical Review Corporation, Raleigh, NC
Soumaya El Rouby, ITC, Piscataway, NJ

ABSTRACT

In method-comparison studies, regression analysis is often used to estimate the systematic difference between measurements from two different methods[1]. Ordinary linear regression should only be used if the measurements from one of the methods are without random error, which rarely occurs. The Deming method of regression analysis, which accounts for measurement error in both methods, is often more appropriate, and is requested by the U.S. Food and Drug Administration (FDA) in medical device submissions. Though PROC NLP and CALIS in SAS® can accommodate Deming regression models, these procedures are not widely used, and may not be easily understood by many applied statisticians. This paper will present a macro that uses multiple DATA steps and PROC MEANS statements to calculate the slope and intercept of the Deming regression line. Since it is difficult to derive a formula for the standard deviations of these estimates, the non-parametric jackknife method is employed to construct confidence intervals.
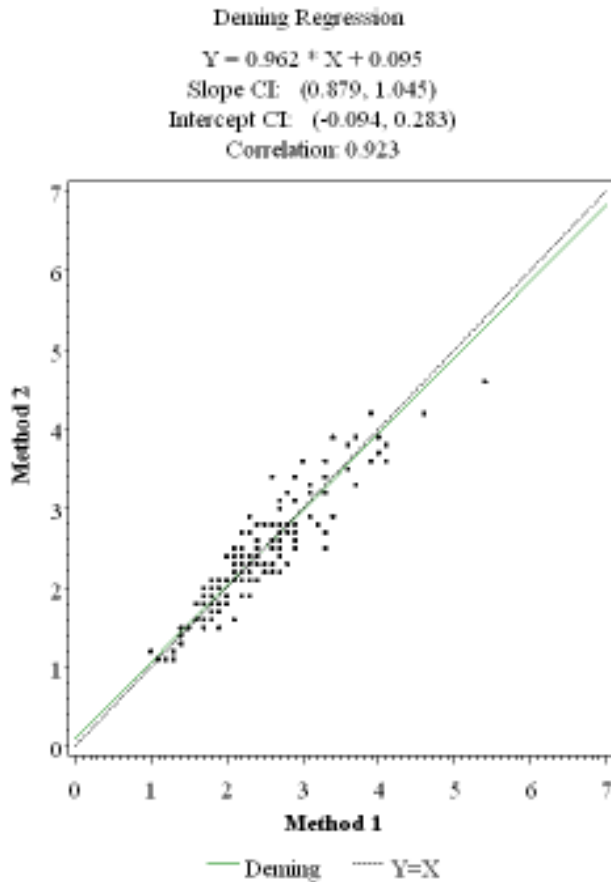
INTRODUCTION

In order to receive approval from the FDA for a new medical measurement device, it must be shown to measure the target value as accurately as the gold standard method[2]. A simple correlation statistic can be used to test the strength of the association between the two measurements; however, it is possible that the two methods of measurement could correlate well, but not produce similar results. Thus, a regression analysis is needed to show that the results produced by the new device are as good as the gold standard method. A regression line with a slope of one and an intercept of zero would indicate that the two methods give similar results throughout the entire range of measurements.

DEMING REGRESSION

Ordinary Least Squares regression cannot be used in the setting of method-comparison studies, since it assumes that only one variable has measurement error. Thus, in order to account for measurement error in both the independent and dependent variables, the FDA encourages the use of Deming regression. The Deming regression line is estimated by minimizing the sums of squared deviations in both the x and y directions at an angle determined by the ratio of the analytical standard deviations for the two methods. This ratio can be estimated if multiple measurements were taken with each method, but if only one measurement was taken with each method, it can be assumed to be equal to one. The following code uses equations set forth by Linnet[3] to estimate the Deming regression slope and intercept, assuming $\lambda = 1$. Standard errors of these estimates are found using the Jackknife method[4] and are used to calculate confidence intervals.

Figure 1 shows a Deming regression analysis comparing a device method to a gold standard laboratory method for patients receiving oral anticoagulants. (Method 1 is the lab method and method 2 is the device method.) Since the 95% confidence interval of the slope includes 1 and the 95% confidence interval of the intercept includes 0, it is reasonable to conclude that using the new device is unlikely to yield an erroneous result.

1

FIGURE 1: DEMING REGRESSION EXAMPLE

Deming Regression

Y = 0.962 * X + 0.095
Slope CI: (0.879, 1.045)
Intercept CI: (-0.094, 0.283)
Correlation: 0.923

SAS CODE

The %DEMING macro is created to estimate the Deming regression slope and intercept by using the equations set forth by Linnet and the values stored in the input dataset. This macro requires an input dataset that consists of three variables: an id variable, the measurement found by the new method, and the measurement found by the gold standard method.   Here, an example dataset is created with measurements from 200 subjects.

Next, the %DOIT macro is created. In Step 1 of this macro, the %DEMING macro is called on the entire input dataset in order to calculate the regression estimates. Then, the %DEMING macro is called on each jackknifed sample and the results are output to a dataset called JACK (Step 2). This dataset holds the information needed to calculate the standard errors (Step 3), which are then used to calculate the 95% confidence intervals for the regression estimates.

Finally, the %DOIT macro is called with the following arguments:
    dataset = the name of the dataset which contains the paired observations
    id = the name of the variable that contains the subject ids
    method1 = the name of the variable that contains the measurements from the first method
    method2 = the name of the variable that contains the measurements from the second method

The DATA step and PROC SQL statement after the %DOIT macro create a dataset named RESULTS, which contains the Deming regression results and their confidence intervals.

```sas
/*Create example dataset*/
data example;
      do myid=1 to 200;
            value=UNIFORM(0)*10;
            new=value + NORMAL(0);
            gold=value + NORMAL(1);
            output;
      end;
run;

/*Create %DEMING macro*/
%MACRO DEMING(dataset, id, method1, method2);

      **************************Deming Regression**************************;
      *****************************Linnet 1998****************************;
      data deming;
            set &dataset;
            keep &id &method1 &method2;
            rename &method2=x &method1=y &id=id;
      run;

      /*Use PROC MEANS to obtain the average measured value for each of the
        two measurement methods across all measured items*/
      proc means data=deming mean NOPRINT;
            var x y;
            output out=means;
      run;
      data _null_;
            set means;
            where _STAT_ IN ("MEAN");
            call symputx("x_mean",round(x,0.001));
            call symputx("y_mean",round(y,0.001));
      run;

      /*Calculate sum of squares using a DATA STEP to calculate the squared
        deviations and PROC MEANS to sum those deviations*/
      data deming1;
            set deming;
            u=(x-&x_mean);
            u2=u**2;
            q=(y-&y_mean);
            q2=q**2;
            p=u*q;
      run;
      proc means data=deming1 sum NOPRINT;
            var u2 q2 p;
            output out=means sum=u q p;
      run;
      data _null_;
            set means;
            call symputx("u",round(u,0.001));
            call symputx("q",round(q,0.001));
            call symputx("p",round(p,0.001));
      run;

      /*Use the sum of squares and means obtained above to calculate the
        Deming regression slope and intercept estimates, based on equations
        from Linnet 1998*/
      data deming2;
            b=((&q - &u) + ((&u-&q)**2 + 4*(&p)**2)**0.5)/(2*&p);
```

```
                a0=&y_mean - b*&x_mean;
                dummy=1;
        run;
%MEND DEMING;

/*Create %DOIT macro*/
%MACRO DOIT(dataset, id, method1, method2);

        **************************************************************************;
        * Step 1: DEMING REGRESSION ESTIMATES
        * Obtain the Deming regression slope and intercept estimates by calling
        * the %DEMING macro on the full dataset
        **************************************************************************;

        %DEMING(&dataset, &id, &method1, &method2)

        data estimates;
                set deming2;
                rename b = b_est
                        a0 = a0_est;
        run;


        **************************************************************************;
        * Step 2: JACKKNIFE PROCEDURE
        * Obtain the Deming regression slope and intercept estimates for each of
        * the N datasets created by removing the ith observation from the full
        * dataset and then calling the %DEMING macro on each of those N datasets
        * (based on Linnet 1990)
        **************************************************************************;

        /*Create a variable containing the number of observations in the full
          dataset*/
        proc means data=&dataset n NOPRINT;
                var &method1;
                output out=n;
        run;
        data _null_;
                set n;
                where _STAT_ IN ("N");
                call symputx("n",round(&method1,0.001));
        run;

        /*Number the N observations in the full dataset 1 thru N*/
        data jackorig;
                set &dataset;
                count+1;
        run;

        /*Create an empty dataset that will hold the results of the jackknife
          procedure*/
        data jack;
                set _null_;
        run;

        /*Call the %DEMING macro N times*/
        %DO i=1 %TO &n;

                /*On the ith iteration, remove the ith observation from the
                  original dataset*/
                data jack&i;
                        set jackorig;
```

4

```
                    where count NE &i;
            run;

            /*Call the %DEMING macro on the subsetted dataset*/
            %DEMING(jack&i, &id, &method1, &method2)

            /*Add the Deming regression slope and intercept estimates for the
              ith iteration as the ith observation in the dataset JACK and
              define the variable JACK equal to the value of i*/
            data jack;
                    set jack deming2(in=a);
                    if a then jack = &i;
                    dummy = 1;
            run;

    %END;

    ***********************************************************************;
    * Step 3: CONFIDENCE INTERVALS
    * Using results from the jackknife procedure and the Deming regression
    * slope and intercept estimates, calculate the standard errors of the
    * Deming regression slope and intercept estimates and use those to
    * produce 95% confidence intervals
    ***********************************************************************;

    /*Calculate the jackknifed estimators of the slope and intercept using a
      a DATA STEP to calculate the estimators for the ith iteration and PROC
      MEANS to obtain the means of those estimators*/
    data jackcalc;
          merge jack estimates;
          by dummy;
          jackb=(&n * b_est) - ((&n-1)*b);    /*See Equation 3a*/
          jacka=(&n * a0_est) - ((&n-1)*a0); /*See Equation 3a*/
          call symputx("b_est", b_est);
          call symputx("a0_est", a0_est);
    run;
    proc means data=jackcalc mean NOPRINT;
          var jackb jacka;
          output out=jackcalc1 mean=jackbmean jackamean;
    run;

    data jackcalc1;
          set jackcalc1 (drop= _type_ _freq_);
          dummy=1;
    run;

    /*Calculate the variances of the Deming regression slope and intercept
      estimates using a DATA STEP to calculate the variances for the ith
      ith iteration and PROC MEANS to obtain the sum of those variances*/
    data jackcalc2;
          merge jackcalc jackcalc1;
          by dummy;
          diff_b2=(jackb- jackbmean)**2/(&n-1);
          diff_a2=(jacka- jackamean)**2/(&n-1);
    run;
    proc means data=jackcalc2 sum NOPRINT;
          var diff_b2 diff_a2;
          output out=variance sum=sumb suma;
    run;

    /*Calculate the standard errors and 95% confidence intervals for the
```

```
        Deming regression slope and intercept estimates*/
    data variance2;
        set variance;
        se_b=SQRT(sumb/&n);
        se_a=SQRT(suma/&n);
        t=TINV(.975,&n-1);
        b_lower= &b_est - TINV(.975,&n-1)* se_b;
        b_upper= &b_est + TINV(.975,&n-1)* se_b;
        a_lower= &a0_est - TINV(.975,&n-1)* se_a;
        a_upper= &a0_est + TINV(.975,&n-1)* se_a;
    run;

%MEND DOIT;


/*Call the %DOIT macro*/
%DOIT(example, myid, new, gold);

/*Create dataset RESULTS that stores the Deming regression estimates and CIs*/
data result;
    merge estimates variance2;
    drop _TYPE_ _FREQ_ dummy sumb suma;
    slope= STRIP(ROUND(b_est, 0.01)) || " (" ||
        STRIP(ROUND(b_lower, 0.001)) || ", " ||
        STRIP(ROUND(b_upper, 0.001)) || ")";
    int= STRIP(ROUND(a0_est, 0.01)) || " (" ||
        STRIP(ROUND(a_lower, 0.001)) || ", " ||
        STRIP(ROUND(a_upper, 0.001)) || ")";
run;

PROC SQL;
    CREATE TABLE results AS SELECT slope, int, b_est, se_b, b_lower,
    b_upper, a0_est, se_a, a_lower, a_upper
    FROM result;
QUIT;
```

CONCLUSION

Deming regression is an important statistical method that should be used when comparing two measurements that both have measurement error. Since this type of analysis is recommended by the FDA for medical device submissions, it is important for applied statisticians to be able to perform this analysis. While SAS has the capability to produce estimates and confidence intervals for Deming regression models, the PROCs used to perform these calculations are not widely used, and may not be easily understood by many applied statisticians. The code presented in this paper is intended to provide a more straightforward method of calculating Deming regression estimates and calculating the confidence intervals for those estimates by using the jackknife method.

REFERENCES

1. Linnet, K. Evaluation of Regression Procedures for Methods Comparison Studies. *Clinical Chemistry 39:3, 424-432* (1993).

2. Guidance for Industry and FDA Staff: Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices: January 30, 2008.
   http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079632.htm

3. Linnet, K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clinical Chemistry 44:5 1024–1031* (1998).

4. Linnet, K. Estimation of the Linear Relationship between the Measurements of Two Methods with Proportional Errors. *Statistics in Medicine, Vol. 9, 1463-1473* (1990).

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Allison M. Deal, MS
UNC Lineberger Comprehensive Cancer Center
CB #7295
Chapel Hill, NC 27599-7295
Email: Allison_Deal@med.unc.edu
Phone: 919-966-1387
Fax: 919-966-4244

Virginia W. Pate, MS
Surgical Review Corporation
4800 Falls of Neuse Rd
Suite 160
Raleigh, NC 27609
Email: Virginia.Pate@surgicalreview.org
Phone: 919-792-1710
Fax: 919-882-1801

Soumaya El Rouby, Ph.D.
Sr. Manager, Clinical Affairs
ITC
20 Corporate Place South
Piscataway, NJ 08854
Email: selrouby@itcmed.com
Phone: 1-800-631-5945 ext 4221
Fax: 732-635-0144