

Application of Deming Regression in Molecular Diagnostics using a SAS® Macro

Merlin M Njoya and Parichehr Hemyari, Roche Molecular Solutions, Pleasanton, California USA

ABSTRACT

In Molecular Diagnostics, method comparison studies are conducted to estimate possible systematic difference between measures from a new investigational assay and an approved assay. The U.S. Food and Drug Administration (FDA) recommends in such case the use of Deming regression analysis to establish that the new assay measures the target value as accurately as the gold standard method¹. While extensive literature is available on the estimation of the Deming Regression parameters using SAS, there is limited information on how to construct confidence interval of these parameters when there is correlation among data due to repeated measurements. This paper will present a SAS macro that efficiently calculates the slope and intercept of the Deming regression and also performs the bootstrap² method to estimate the confidence interval around these parameters.

INTRODUCTION

Molecular Diagnostics aims to provide new assays for screening, diagnostic and monitoring of patients. In the process of the validation of a new assay, method comparison studies are conducted to establish that the new assay measures the target value as accurately as the reference method. For quantitative assays, method comparisons include linear regressions and the U.S. Food and Drug Administration (FDA) recommends Deming regression analysis or similar analysis that account for random errors from both the new assay and reference method¹. This paper presents a SAS macro that efficiently calculates the slope and intercept of the Deming regression and also performs method the bootstrap² method to estimate the confidence interval around these parameters.

DEMING REGRESSION THEORY

Every measurement method is subject to some random measurement error. As an example (Linnet K et al. 1998)², when measuring an analyte concentration on a given sample, the measured value from the reference test method (x_i) would deviate from the target value (X_i) by some small “random” amount (ε_i). Similarly the measured value from the new test method (y_i) would deviate from the target value (Y_i) by some small “random” amount (σ_i) and the following relations exist:

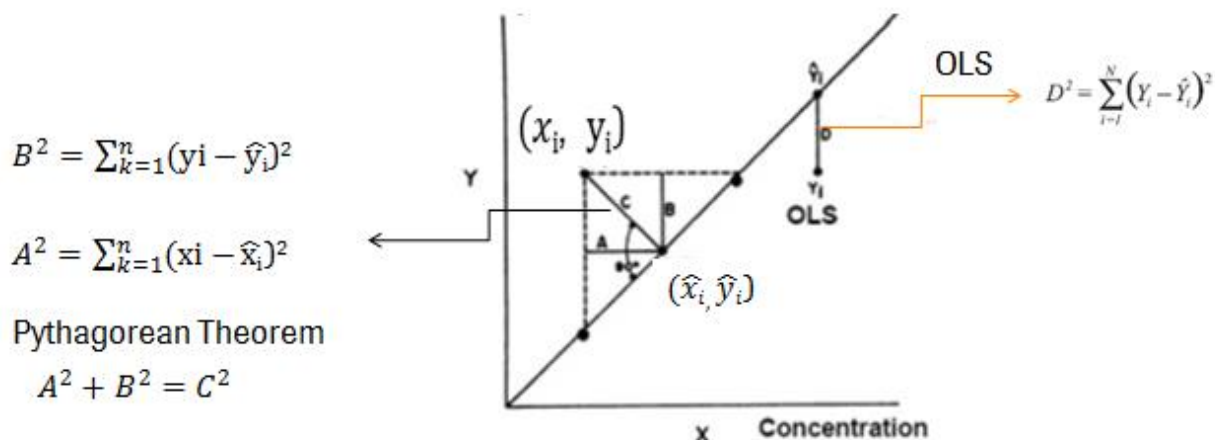
$$x_i = X_i + \varepsilon_i$$

$$y_i = Y_i + \sigma_i$$

Assuming a linear relationship between the target values of the two methods $Y_i = \alpha_0 + \beta X_i$, the Deming regression procedure takes into account the measurement errors in both the reference test method and the new test method and provides an unbiased estimate of the slope β (Linnet K. 1998)³.

In other words, while the Ordinary Least Square (OLS) regression makes the assumption that the x variable (reference test method in this example) is error-free and tries to find the line of best fit to minimize the residual errors (D) in the y variable (the new test method), the Deming regression tries to find the line of best fit to minimize the residual errors (A and B) on both X and Y (Figure 1).

Figure 1. Deming regression versus OLS regression



Assuming constant analytical errors, the unweighted form of Deming regression analysis is appropriate and the slope and intercept estimates are given by the following equations (Linnet K. 1993)⁴:

$$\hat{\beta} = [(\lambda q - u) + [(u - \lambda q)^2 + 4\lambda p^2]^{0.5}] / 2\lambda p \quad (1)$$

$$\hat{\alpha}_0 = y_m - \hat{\beta}x_m$$

$$\text{Where } \begin{cases} x_m = \sum x_i / N \\ y_m = \sum y_i / N \\ u = \sum (x_i - x_m)^2 \\ q = \sum (y_i - y_m)^2 \\ p = \sum (x_i - x_m)(y_i - y_m) \end{cases} \quad (2)$$

With N the total number of samples measured and λ the analytical errors ratio generally assigned a default value of one for two measurement methods with similar precision. It can also be estimated based on the analytical standard deviation (SDa) obtained from the reproducibility study of each measurement method as SD_{ax}^2 / SD_{ay}^2 .

SAS CODES FOR DEMING REGRESSION PARAMETER ESTIMATES

Deming regression analysis can be performed in SAS using the NLP procedure, the OPTMODEL procedure as well as the CALIS procedure. A deep understanding of these procedures is required in order to accommodate them to the Deming regression models. In contrast, the %Deming_Linnet SAS macro presented below is straightforward for the computation of the Deming regression estimates (slope and intercept) in the comparison of a new measurement method to a reference method.

The %Deming_Linnet macro requires the following five arguments:

1. **Datain**: the SAS input data set containing at least the variable with the reference method results (Xvar) and the variable with the new method results (Yvar).
2. **Dataout**: the SAS output data set that will contain the slope and intercept estimates.
3. **Xvar**: the name of the variable with the reference method results.
4. **Yvar**: the name of the variable with the new method results.

5. **Lambda**: the error ratio set by default to 1.

Using the SQL Procedure, %Deming_Linnet computes the quantities in equation (2) and uses them to estimate the Deming regression slope and intercept in equation (1).

```
%macro Deming_Linnet(datain = , dataout = , Xvar=, Yvar= , lambda=1) ;
  /** Parameter estimates using all samples in datain dataset **/

  /* computation of the quantities in equation (2) */
  proc sql;
    create table distance as
      select
        mean(howfar.x_mean) as x_mean,
        mean(howfar.y_mean) as y_mean,
        sum(howfar.xxi) as xxi_sum,
        sum(howfar.yyi) as yyi_sum,
        sum(howfar.xyi) as xyi_sum
      from (
        select
          x, mean(x) as x_mean,
          y, mean(y) as y_mean,
          (x - calculated x_mean)**2 as xxi,
          (y - calculated y_mean)**2 as yyi,
          ((x - calculated x_mean) * (y - calculated y_mean)) as xyi
        from &datain) howfar;
  quit;

  /* computation of the quantities in equation (1) */
  proc sql;
    create table &dataout as
      select
        &lambda as lambda,
        ( (calculated lambda)*yyi_sum - xxi_sum + sqrt((xxi_sum -
        (calculated lambda)*yyi_sum)**2
        + 4*(calculated lambda)*xyi_sum**2)) / (2*(calculated
        lambda)*xyi_sum) as slope,
        (y_mean - calculated slope*x_mean) as intercp
      from distance;
  quit;

%mend Deming_Linnet;
```

SAS CODES FOR BOOTSTRAPPED DEMING REGRESSION

Once the %Deming_Linnet macro generates the Deming regression slope and intercept estimates, one will be interested in making inferences about the comparison of the new measurement method with the reference method. Calculating confidence intervals around these estimates can be useful.

The %Deming_bootstrap SAS macros presented below computes the confidence intervals of the Deming regression slope and intercept estimates using the Bootstrap re-sampling method. Therefore it can accommodate designs with correlation among data due to repeated measurements for same subject.

The %Deming_bootstrap macro requires the following nine arguments:

1. **Datain**: the SAS input data set containing at least the variable with the reference method results (Xvar), the variable with the new method results (Yvar) and the subject ID variable (subject).
2. **Dataout**: the SAS output data set that will contain the bootstrapped slope and intercept estimates as well as their corresponding confidence interval (CI) limits.
3. **Xvar**: the name of the variable with the reference method results.

4. Yvar: the name of the variable with the new method results.
5. Lambda: the error ratio set by default to 1.
6. num_rep: the number of replicates for the Bootstrap re-sampling set to 1000 by default.
7. seedBoot: the seed for the random sampling.
8. alpha: the significance level for the two sided confidence intervals set to 0.05 by default.
9. subject: the name of the variable indicating the subject ID or sample ID. Useful when there are repeated measurements on same subjects or samples.

```

%macro deming_bootstrap(datain = , dataout = , Xvar=, Yvar= , lambda=1 ,
                        num_rep=1000, seedBoot=48941, alpha=0.05, subject=);

  /** Compute BOOTSTRAPPED Confidence Intervals for the Deming Regression slope
      and Intercept **/

  /* Re-sampling to create the input datasets for the bootstrap */
  proc surveyselect data=&datain out=resampled
    seed=&seedBoot
    method=urs
    samprate=1
    outhits
    reps=&num_rep;
    strata &subject;
  run;

  proc sort data=resampled;
    by replicate &subject;
  run;

  /* Run %deming_Linnet on the first re-sampled dataset;
     This will create the first instance of the dataset where all the
     bootstrap estimates will be appended */
  data boot_1;
    set resampled;
    where Replicate eq 1;
  run;

  %deming_Linnet (datain =boot_1 , dataout =dem_par, Xvar= &Xvar,
                Yvar=&Yvar , lambda=&lambda);

  /* Perform the bootstrap with %bootit */
  %macro bootit;
    %do rep=2 %to &num_rep;

      *subset dataset to i-th re-sampled dataset (i>1);
      data boot_&rep;
        set resampled;
        where Replicate eq &rep;
      run;

      %deming_Linnet (datain =boot_&rep , dataout =dem_par_&rep, Xvar=
                    &Xvar, Yvar=&Yvar , lambda=&lambda);
      proc append base=dem_par data=dem_par_&rep;
      run;

    %end;
  %mend bootit;

```

```

%bootit;

/* Get the median value of the bootstrap estimates and their CIs */
proc univariate data=dem_par;
var intercp slope;
output median=int_med slp_med pctlpre=int_slp_
pctlpts=%sysevalf(100*(alpha./2)), %sysevalf(100*(1-alpha./2))
out=&dataout;
run;

%mend deming_bootstrap;

```

PRACTICAL EXAMPLE OF THE USE OF BOOTSTRAPPED DEMING REGRESSION SAS MACRO

Viral load testing to determine the amount of circulating Human immunodeficiency virus (HIV) is crucial for medical decision making and monitoring of HIV infected people. We present data from a method comparison study of a new HIV-1 assay that runs on a fully automated system as compared to a reference test method.

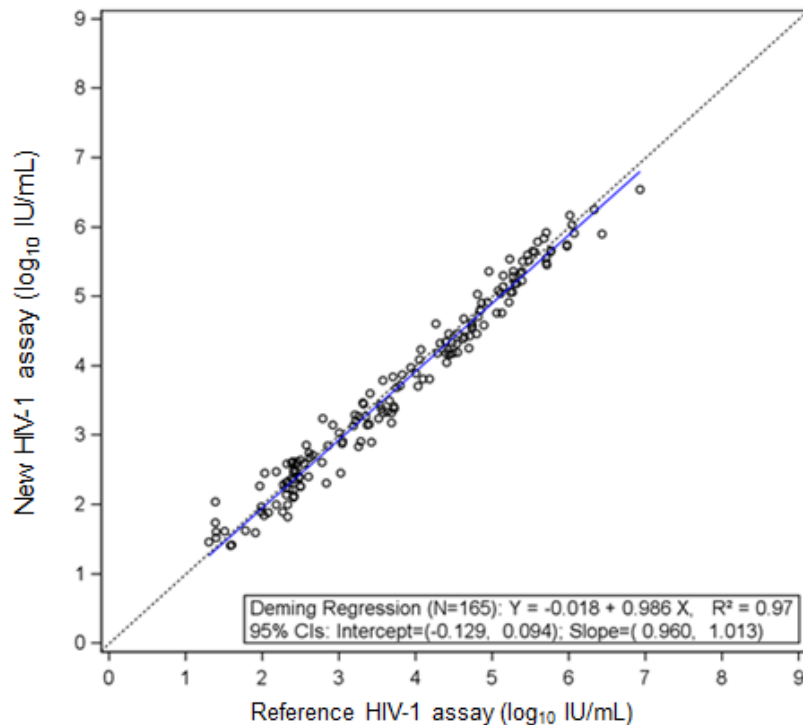
```

%deming_bootstrap(datain =TestRslt, dataout=boot_ci_dem , Xvar=old_test,
Yvar=new_test , lambda=1 , num_rep=1000, seedBoot=48941,
alpha=0.05, subject=subjid);

```

You run this macro and will get the Deming regression intercept and the slope as well as their corresponding bootstrapped 95% confidence intervals. Results can be presented as in Figure 2 by a scatter plot showing the fitted Deming regression line, the identity line (methods equal) and the regression equation with the confidence intervals for the intercept and the slope.

Figure 2. Deming regression for two HIV assays comparison



CONCLUSION

When comparing two measurement methods that both have measurement error, Deming regression is one of the appropriate methods that should be used. You can easily use the %Deming_Linnet SAS macro presented in this paper to calculate the Deming regression parameters (intercept and slope). For inferences, the code of the %Deming_bootstrap SAS macro can be used to perform the bootstrap method and generate the confidence intervals around the Deming regression estimates. This macro can accommodate designs with correlation among data due to repeated measurements.

REFERENCES

1. FDA Guidance: Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices. Available at <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070890.pdf>
2. Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, Vol. 82, No. 397.
3. Linnet K. (1998). 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. (1993). Evaluation of regression procedures for methods comparison studies. *Clinical Chemistry* Mar 1993, 39 (3) 424-432;
5. SAS Enterprise Guide v7.13. Copyright © 2016 by SAS Institute Inc., Cary, NC, USA. All rights reserved. Product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
6. SAS Institute Inc. 2011. SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

The authors thank Enrique Marino and Jesse Canchola for their valuable comments as well as for testing and validating the macro.

CONTACT INFORMATION

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

Merlin M. Njoya
Roche Molecular Solutions
4300 Hacienda Drive
Pleasanton, California, 94588
Email: merlin.njoya@roche.com
Web: <http://usdiagnostics.roche.com/en/index.html>

Parichehr Hemyari
Roche Molecular Solutions
4300 Hacienda Drive
Pleasanton, California, 94588
Email: pari.hemyari@roche.com
Web: <http://usdiagnostics.roche.com/en/index.html>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

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