Statistical Issues in Evaluating Hemostasis State of a Blood Sample

Kyungsook Kim[1]

[1]Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD 20993

Abstract

Hemostasis evaluations are commonly used to assess clinical conditions in trauma, cardiovascular surgery and it is also used in cardiology procedures to assess hemorrhage or thrombosis conditions before, during and following the procedure. Clotting time, clot stiffness/firmness, and platelet function are parameters for evaluating whether the new device is substantial equivalence to an already marketed device.

Although these parameters are semi-quantitative, if the units are different between devices due to using different techniques to measure these parameters, regression analysis is not an appropriate method to use. In this paper, a simulated data was used to determine whether ordinary regression analysis with rank-transferred data can be a possible analysis option in evaluating hemostasis state of a blood sample.

Key Words: Rank regression, Method Comparison, Passing-Bablok regression

1. Introduction

When a new/candidate device's outcome is quantitative, manufacturers need to establish the relationship of any new/candidate measurement procedure of measurand quantification with a comparative measurement procedure, ideally a reference measurement procedure. The goal of the comparison is either to establish or to show no significant bias between the two methods (CLSI EP09-A3). Linear regression models are used in method comparison study. An appropriate regression method should be used to estimate the slope β, the intercept α and their associated confidence intervals. Sufficient sample size is needed such that 95% confidence intervals on β and α are narrow enough to conclude that the slope β is near unity and the intercept α is near zero to within a clinical insignificant deviation δ. However, the classic linear regression analysis assumes that the linear relationship between dependent and independent variables, statistical independence, homoscedasticity and normality of the errors (Hocking, 1996), and assumptions may not be met by the data obtained in many studies. Especially, evaluating hemostasis parameters need to use different analysis than ordinary linear regression method since the assumptions are not met due to using different techniques to measure these parameters.

2. Rank-transformed linear regression

The rank regression is a simple technique which engages replacing the data with their corresponding ranks. Additionally, we simply fit a line through the (rank of the) points and therefore no assumptions are needed to employ this approach.

Let us consider a paired data (y_i, xi) for a simple linear regression model

$$yi = \alpha + \beta\, xi + \varepsilon i \qquad (2.1)$$

where x is called regressor (independent variable) and y is called response (dependent variable). α and β are the intercept and regression coefficients, respectively and εi is a random error term.

To use the rank transformation in model (2.1), the observations of the dependent variable y and the independent variable x are replaced by their corresponding ranks 1 to n. Let R(yi) is the rank corresponding to the i-th value of y and R(xi) is the rank corresponding to the i-th value of x. The average ranks are shared to the values if they are ties. Now the model (2.1) becomes

$$R(yi) = \alpha + \beta\, R(xi) + \varepsilon_i \qquad (2.2)$$

By minimizing the sum of square errors, $\sum_{i=1}^{n} \varepsilon_i^2$ the resultant rank regression parameter estimates are defined as

$$\beta = \frac{\sum R(x_i)R(y_i) - n(n+1)^2/4}{\sum \{R(x_i)\}^2 - n(n+1)^2/4} \qquad (2.3)$$

$$\alpha = \frac{(1-\hat{\beta})(n+1)}{2} \qquad (2.4)\ \text{(Rana, et al., 2013)}$$

It is seen that the rank regression parameters are the function of the rank of the observations.

3. Simulation Studies

Temperature Fahrenheit to Celsius conversion formula was used to examine the impact of various factors by using rank-transferred linear regression. Because the unit is different between two measures, the usual linear regression model from method comparison study is not appropriate, i.e., β=1 and α=0 cannot be shown. Can we use rank-based linear regression in this case?

3.1. Range of X

The data were simulated from the model:

$$y_i\,(^{o}F) = 1.8 * x_i\,(^{o}C) + 32 + \varepsilon_i,\ \varepsilon_i \sim N(0,1^2) \qquad (3.1)$$

To show differences due to distribution of X (^{o}C), a sample size (n=150) was selected and X was distributed ~ N(15, $1/2^2$) for the effect of sampling variability on model estimates. All simulations were performed 100 times.

Table 1: Estimates of rank-transferred slope (β) and intercept (α) to data simulated for various range of X

X ~	Rank (β^) (5^{th}, 95^{th})*	Rank (α^) (5^{th}, 95^{th})*
N(15, $1/2^2$)	0.63 (0.55, 0.71)	27.97 (21.61, 34.04)
N(15,1^2)	0.85 (0.816, 0.88)	11.3 (8.85, 13.92)
N(15,2^2)	0.955 (0.94, 0.96)	3.39 (2.73, 4.16)
N(15,5^2)	0.992 (0.989, 0.993)	0.63 (0.50, 0.77)
N(15,10^2)	0.9975 (0.997, 0.998)	0.18 (0.15, 0.23)

* Based on100 simulations, 5^{th} and 95^{th} percentile

As the range of X increases, e.g., (13.5 ^oC – 16 ^oC) to (-10^oC – 40 ^oC), the estimate of rank-based slope is close to 1. The following figures (Figures 1-2) are shown the scatter plot of mean of 15 ^oC and variance of $(1/2)^2$ for the original data and rank-transferred data. Figures 3-4 are shown the scatter plot of mean of 15 ^oC and variance of 10^2 for the original data and rank-transferred data.

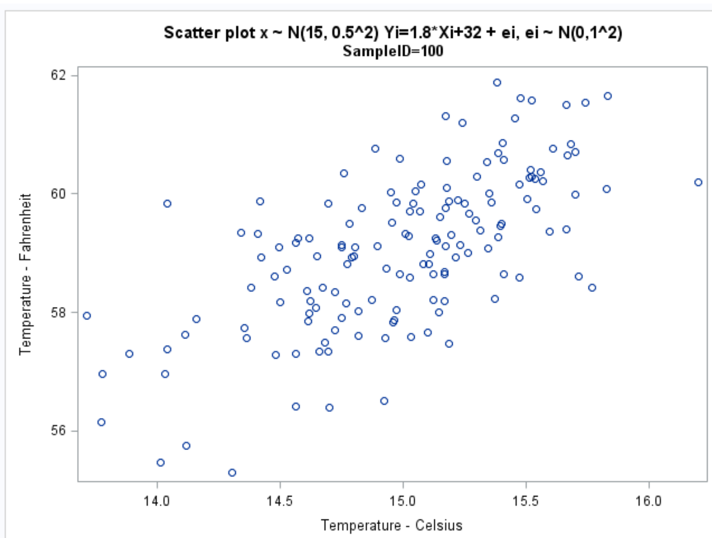

Figure 1: Scatter plot of mean of 15 ^oC and variance of $(1/2)^2$

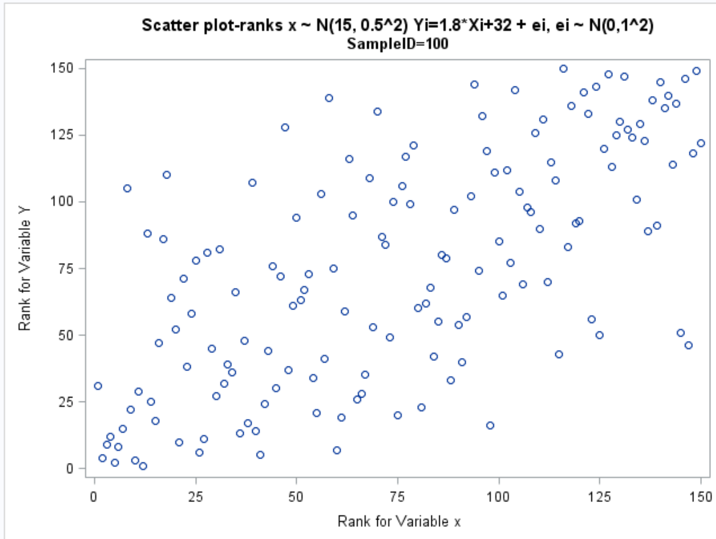

Figure 2: Scatter plot of rank-transferred data with mean of 15 °C and variance of $(1/2)^2$

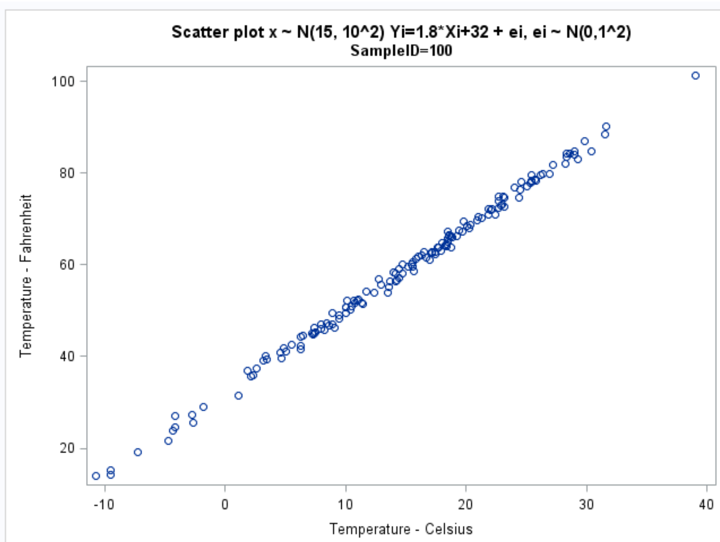

Figure 3: Scatter plot of mean of 15 °C and variance of $(10)^2$

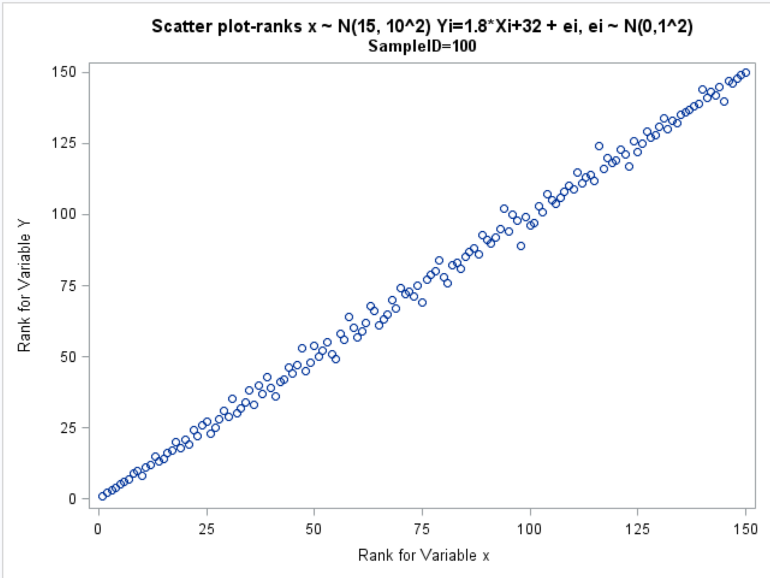

Figure 4: Scatter plot of rank-transferred data with mean of 15 °C and variance of $(10)^2$

3.2. Sample Size

The data were simulated from the model :
yi (°F)= 1.8*xi (°C)+32+εi, εi ~ $N(0,1^2)$, xi~ $N(15, 5^2)$

To examine differences due to different sample sizes, three sample sizes (50, 100, and 300) were selected and X was distributed ~ $N(15, 5^2)$ for the effect of sampling variability on model estimates. All simulations were performed with 100 repeats.

Table 2: Estimates of rank-transferred slope (β) and intercept (α) to data simulated for sample size of 50, 100, and 300

Sample Size	Rank (β^) (5^{th}, 95^{th})*	Rank (α^) (5^{th}, 95^{th})*
50	0.99 (0.987, 0.994)	0.24 (0.15, 0.34)
100	0.991 (0.988, 0.994)	0.43 (0.31, 0.56)
300	0.992 (0.99, 0.993)	1.18 (0.99, 1.42)

* Based on100 simulations, 5^{th} and 95^{th} percentile

The estimate of rank-based slope is close to 1 regardless of sample sizes if the random error of y is small and range of X is reasonably wide. Intercept is a function of slope and sample size. Given that slope is fixed, an intercept increases as sample size increases. The following figures (Figures 5-6) are shown the scatter plot of mean of 15 °C and variance of 5^2 for the original data and rank-transferred data with sample size of 50 and Figures 7-8 are sample size of 300, respectively.

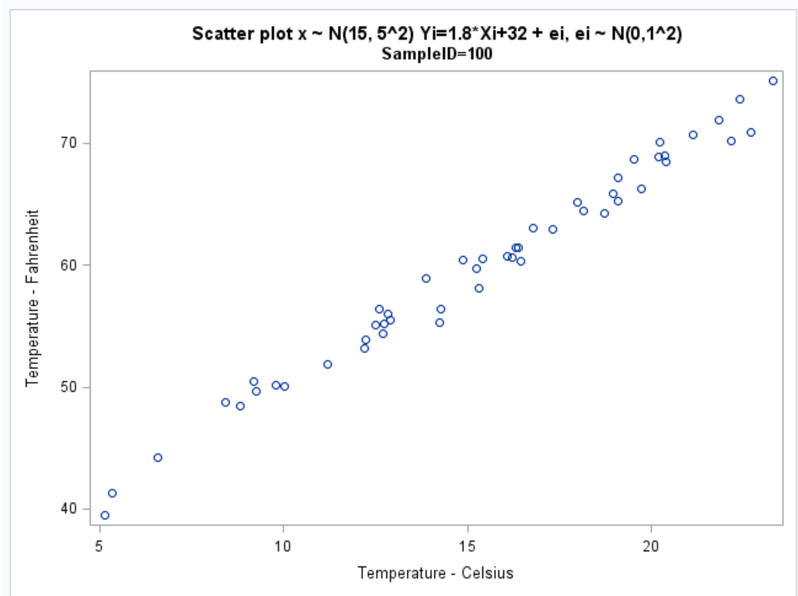

Figure 5: Scatter plot of mean of 15 ^{o}C and variance of 5^2 with Sample Size 50

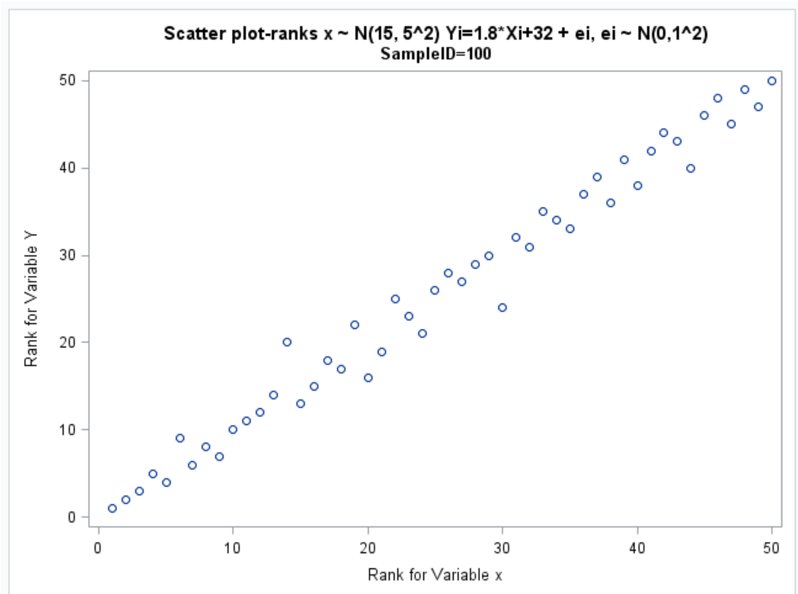

Figure 6: Scatter plot of rank-transferred data with mean of 15 ^{o}C and variance of $(5)^2$ with Sample Size 50

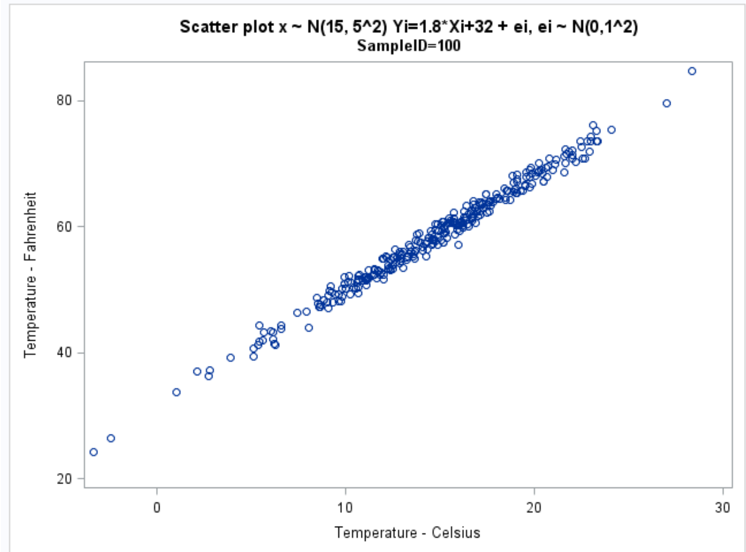

Figure 7: Scatter plot of mean of 15 ^{o}C and variance of $(5)^2$ with Sample Size 300

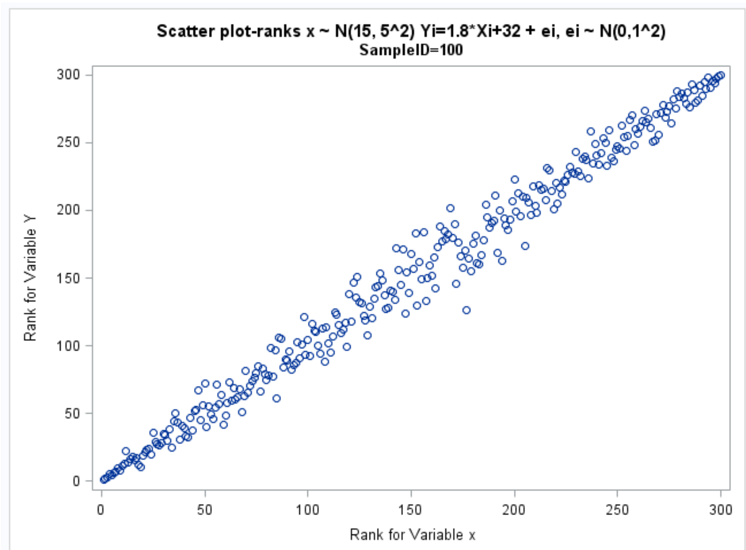

Figure 8: Scatter plot of rank-transferred data with mean of 15 ^{o}C and variance of $(5)^2$ with Sample Size 300

3.3. Random Error

The data were simulated from the model:

yi (^{o}F)= 1.8*xi (^{o}C)+32+εi, xi ~ N(15,5^2), εi~ N(0, 3^2) to εi~ N(0, 20^2),

To show rank-based slope estimate differences due to distribution of εi, a sample size (*n*=150), xi ~ N(15, 5^2) with various random error of y were evaluated (e.g., εi~ N(0, 3^2) to εi~ N(0, 20^2)). All simulations were performed with 100 repeats.

Table 3: Estimates of rank-transferred slope (β) and intercept (α) to data simulated for various εi			
E ~	Rank (β^) (5^{th}, 95^{th})*	Rank (α^) (5^{th}, 95^{th})*	%CV
$N(0,3^2)$	0.937 (0.92, 0.95)	4.73 (3.76, 5.87)	5.1%
$N(0, 6^2)$	0.80 (0.76, 0.85)	14.88 (11.4, 18.0)	10.2%
$N(0,9^2)$	0.669 (0.60, 0.74)	24.97 (19.4, 30.1)	15.3%
$N(0,10^2)$	0.629 (0.55, 0.71)	27.97 (21.6, 34.0)	16.9%
$N(0,12^2)$	0.559 (0.47, 0.66)	33.27 (25.96, 40.1)	20.3%
$N(0,15^2)$	0.475 (0.37, 0.59)	39.65 (31.28, 47.28)	25.4%
$N(0,20^2)$	0.374 (0.26, 0.50)	47.28 (37.76, 55.81)	33.8%

* Based on100 simulations, 5^{th} and 95^{th} percentile

Random error variability (SD or %CV) of y is the major influence factor for estimated rank-based slope. For a same relationship existing, if Y has random error of 15 %CV, the rank-based slope became 0.669. Percent CV was calculated at mean (59^{o}F) of the distribution.

Evaluating Hemostasis parameters which are in different units or different methods, rank-based regression analysis can be used if the random error of Y is small (e.g., %CV <5%). If same order of X does not observed in Y, the interpretation of rank-based slope cannot be made as ordinary linear regression slope does.

As random error variances are increases (e.g., 3^2 to 20^2), the estimate of rank-based slope is farther away from 1. The following figures (Figures 9-10) are shown the scatter plot of mean of 15 ^{o}C and variance of $(5)^2$ with εi ~ $N(0,3^2)$ for the original data and rank-transferred data. Figures 11-12 are shown the scatter plot with $N(0,20^2)$, respectively.

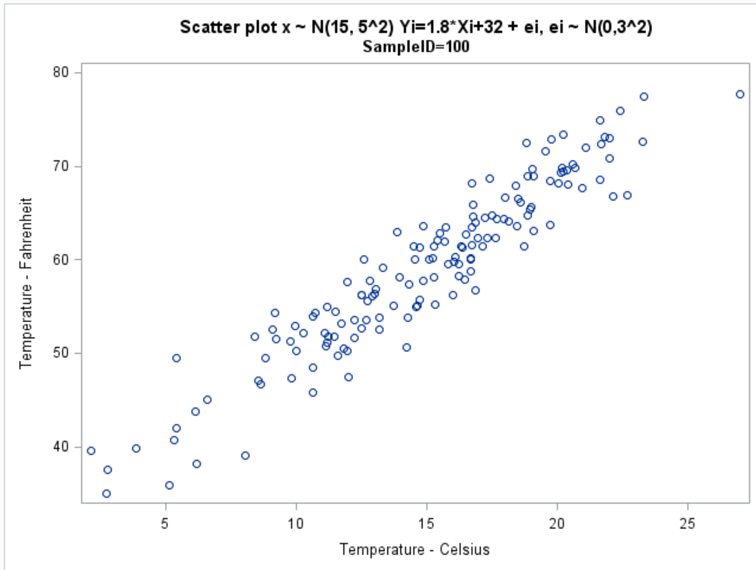

Figure 9: Scatter plot of mean of 15 °C and variance of 5^2 with εi ~ N(0,3^2)

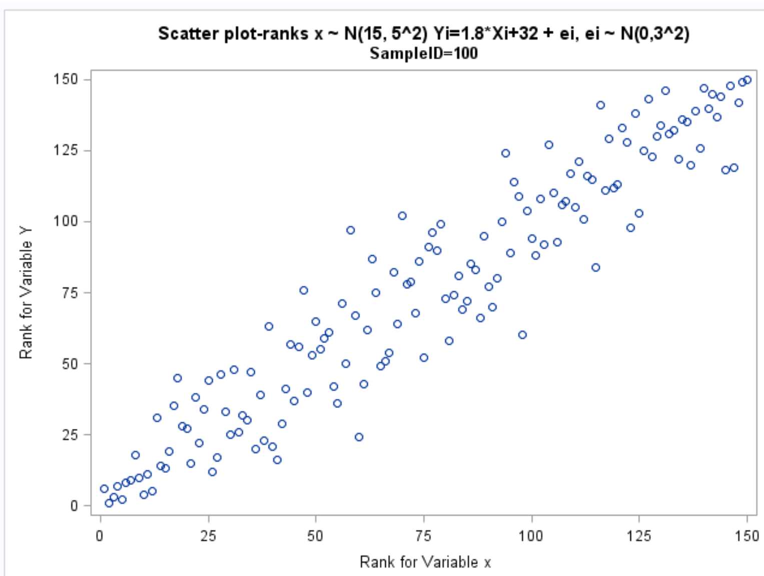

Figure 10: Scatter plot of rank-transferred data with mean of 15 °C and variance of 5^2 with εi ~ N(0,3^2)

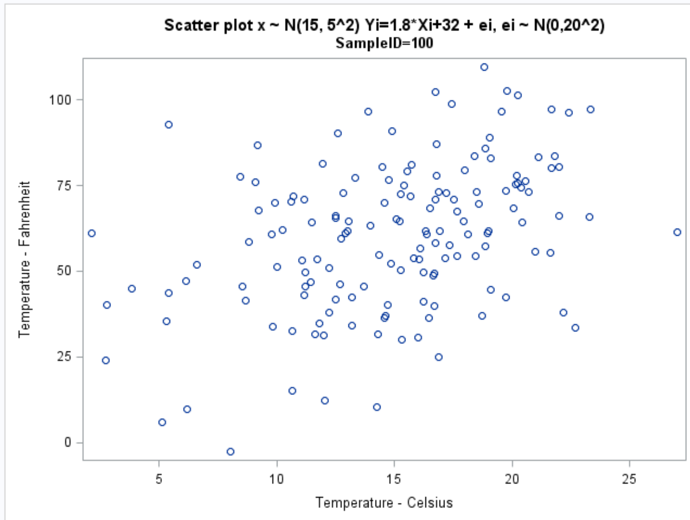

Figure 11: Scatter plot of mean of 15 °C and variance of 5^2 with εi ~ N(0,20^2)

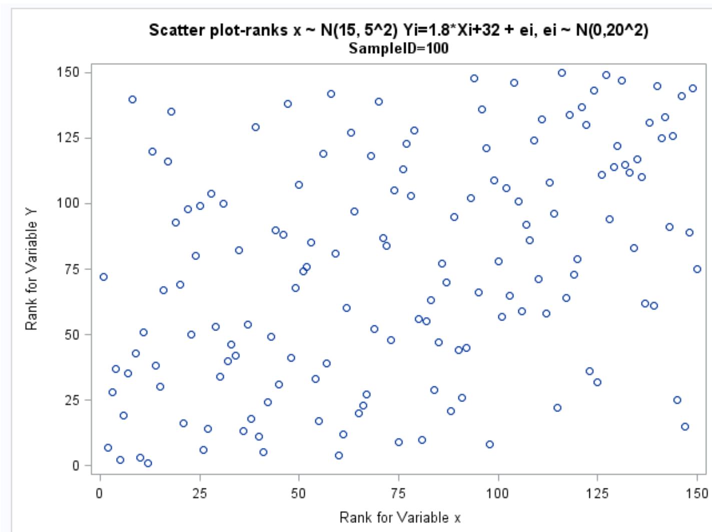

Figure 12: Scatter plot of rank-transferred data with mean of 15 °C and variance of 5^2 with εi ~ N(0,20^2)

3.4. Passing-Bablok Regression

Passing-Bablok regression was applied to the rank-transferred data for various random error distributions (i.e., Table 3 data). The data were simulated from the model :
yi (°F)= 1.8*xi (°C)+32+εi, xi ~ N(15,5^2), εi~ N(0, 3^2) to εi~ N(0, 20^2),

To show rank-based slope estimate differences due to distribution of εi, a sample size (*n*=150), xi ~ N(15, 5^2) with various random error of y were evaluated (e.g., εi~ N(0, 3^2) to εi~ N(0, 20^2)). All simulations were performed with 100 repeats.

Table 4: Estimates of rank-transferred OLS slope and Passing-Bablok slope to data simulated for various εi

Mean= 59 °F	%CV	OLS Rank-based Slope β^(5th, 95th)	Passing-Bablok slope with ranked data
SD=1	1.7%	0.992 (0.989, 0.993)	1.0 (0.99, 1.01)
SD=3	5.1%	0.937 (0.92, 0.95)	1.0 (0.96, 1.04)
SD=6	10.2%	0.803 (0.76, 0.85)	1.0 (0.93, 1.07)
SD=9	15.2%	0.669 (0.6, 0.74)	1.0 (0.91, 1.12)
SD=12	20.3%	0.559 (0.47, 0.66)	1.0 (0.90, 1.15)
SD=15	25.4%	0.475 (0.37, 0.59)	1.02 (0.88, 1.18)

If Passing-Bablok regression is used with rank transformed data, the slope is not affected by the size of random error.

The scatter plot with Passing-Bablok slope [Figure13. (εi ~ $N(0, 3^2)$) and Figure14. εi ~ $N(0, 15^2)$] were following:

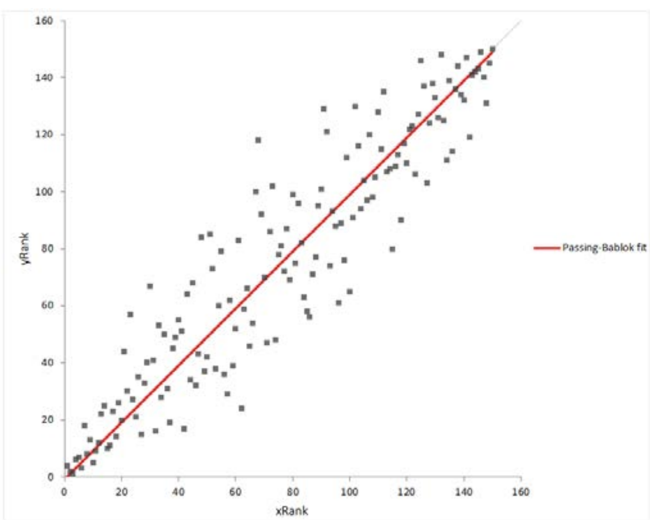

Figure 13: Passing-Bablok regression with scatter plot of rank-transferred data with mean of 15 °C and variance of 5^2 with εi ~ $N(0,3^2)$

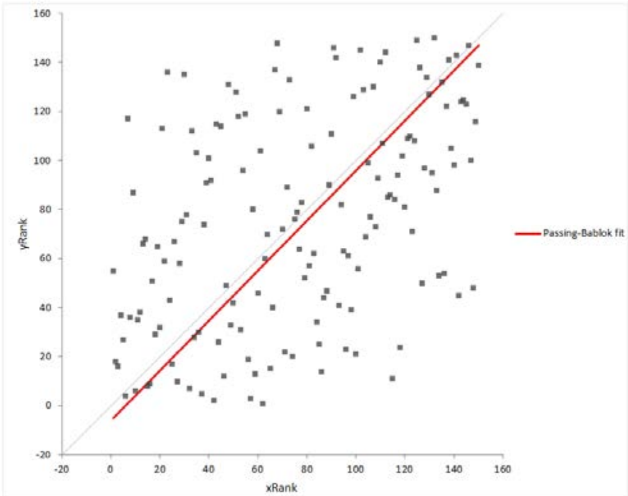

Figure 14: Passing-Bablok regression with scatter plot of rank-transferred data with mean of 15 ^oC and variance of 5^2 with $\varepsilon i \sim N(0,15^2)$

4. Summary

Evaluating hemostasis parameters need to use different analysis than ordinary linear regression method since the assumptions are not met due to using different techniques to measure these parameters. To examine the impact of various factors by using rank-transferred linear regression, temperature Fahrenheit to Celsius conversion formula with various factors were simulated.

Random error variability (SD or %CV) of Y is the major influence factor for estimated rank-based slope. For a same relationship existing, if Y has random error of 15 %CV, the rank-based slope became 0.669 (%CV was calculated at mean (59^{o}F)). As the range of X increases, e.g., (13.5 ^{o}C – 16 ^{o}C) to (-10^{o}C – 40 ^{o}C), the estimate of rank-based slope is close to 1. The estimate of rank-based slope is close to 1 regardless of sample sizes if the random error of Y is small and range of X is reasonably wide. If Passing-Bablok regression is used with rank transformed data, the slope is not affected by the size of random error. The data obtained in evaluating Hemostasis parameters which do not usually meet the classic linear regression assumptions. In this paper I proposed Passing-Bablok rank-based regression as an alternative analysis.

Acknowledgements

I would like to express my appreciation to Drs. Meijuan Li, Branch Chief of DSB II, DBS and Bipasa Biswas, Team Leader for their helpful discussion.

References

1. CLSI EP09-A3: Measurement Procedure Comparison and Bias Estimation Using Patient Samples; Approved Guideline-Third Edition
2. Hocking, R. R. 1993. Methods and applications of linear models: regression and the analysis of variance, *John Wiley & Sons, Inc.*
3. Rana, S. H. Midi, and A. Fitrianto. 2013. Statistical Significance of Rank Regression, *Applied Math, Sciences,* 7(82):4067-4072