IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 4 Ver. V (Jul. - Aug.2016), PP 120-124
www.iosrjournals.org

Precision of Prediction in the Simple Calibration Model

Asma Ali Mohammedkhair[1], Ahamed Mohamed Abdalla Hamdi[2], Dina M.H. swidan[3]

[1, 2, 3] *Sudan University of Science & Technology, Faculty of science, Department of Statistics, Ministry of Education*

***Abstract**:* This paper considered statistical calibration models. It focused on Eiesinhart s calibration model for two tests X & Y where X is an exact but expensive and slow test and Y is less expensive but quick and cheap test. The objective was to investigate through a simulation experiment the effect of the degree of linear dependency between X & Y as well as sample size on confidence interval estimation of the forecasted value of X .It is shown that changes on these factors have no significant effect on the degree of confidence.

***Keyword*:** *Absolute calibration, comparative calibration, Jaundice, Inverse Predictions, Linear model, classical estimator*

I. Introduction

Statistical calibration has some similarities with scientific calibration which is the process whereby the scale of a measuring instrument is determined or adjusted on the basis of an informative or calibration experiment but it has a more complicated form .It is a problem of retrospection and some authors call it inverse regression rather than calibration. It is probably best explained by considering a typical univariate calibration problem.

Consider the problem of a chemist wishing to establish a calibration curve to use in measuring the amount of a certain chemical A in samples sent to an analytical Laboratory. There two method of measurement :an exact but expensive and slow method X and a less expensive and quick method Y . The problem is to find a model that relates Y to X so that measuret of X can be predicted from the measuret of Y . The known amounts of chemical A have been determined by an extremely accurate standard method that is slow and expensive (X). The resulting data constitutes the calibration experiment and is used to estimate the calibration curve f .This calibration curve is now ready for use in the second stage of the calibration process which involves prediction. In the second stage. Samples with unknown amounts of chemical A are analysed with the test method and the amount of chemical A predicted for each new sample. For a given sample, one or more measurements using the test method may he made [1]

In this paper the precision of the prediction of the out come of the standard treatment X from the nonstandard treatment Y , as reflected in the degree of confidence on the predicted value $\widehat{X}$ is investigated. The investigation covered different sample sizes, different degrees of linear correlation between X and Y as well as different population variances. Amonte carlo experiment, using a computer programme written by the researcher, is employed.

II. Theoretical Framework

2.1 Mathematical Formulation of the Univariate Calibration Problem

Let the true values associate with the standard and test method be designated by ξ and η respectively .We assume $\eta = f(\xi)$ and $f(\xi) = \beta_0 + \beta_1 \xi$, where β_0 and β_1 are the intercept and slope parameter respectively.

In the first stage of the calibration process, the calibration experiment , n pairs of observations (X_i, Y_i) are obtained where X_i and Y_i are observed values of ξ_i and η_i respectivel

$$Y_i = \eta_i + \varepsilon_i \qquad i = 1,2,\dots,n$$

$$X_i = \xi_i + \delta_i \qquad i = 1,2,\dots,n \qquad (1)$$

Where εi and δi are experimental errors. In absolute calibration problem $\delta i = 0$ for all i. Produces the following model

$$Y_i = \eta_i + \varepsilon_i = f(X_i) + \varepsilon_i \qquad i = 1,2,\ldots,n \qquad (2)$$

In the case of the linear calibration problem this becomes:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad\qquad i = 1,2,\ldots,n \quad (3)$$

The next assumption which is model is that the $\varepsilon i's$ are independent normal random variables with mean 0 and variance ${\sigma_1}^2$.

Having established the calibration curve /line we proceed to the second stage of the calibration process. A sample is presented with a specific unknown value η and one or more measurements are made using the test method from which are obtained the

$$\grave{Y}_j = \eta_i + \grave{\varepsilon}_j = f(X_i) + \grave{\varepsilon}_j \qquad j = 1,2,\ldots,m \;\ldots(4)$$

$$= \beta_0 + \beta_1 \xi + \grave{\varepsilon}_j \qquad\qquad j = 1,2,\ldots,m\,..(5)$$

In the linear calibration problem,where $\grave{\varepsilon}_1, \grave{\varepsilon}_2, \ldots, \grave{\varepsilon}_m$ are independent normal random variables with mean 0 and variance ${\sigma_2}^2$. Whrer ${\sigma_1}^2 = {\sigma_2}^2 = \sigma^2$.

Given the data from first and second stages. Inferences are now made about the unknown ξ that corresponds to η for the sample being measured. For the linear model ξ is given by: [1]

$$\xi = \frac{(\eta - \beta_0)}{\beta_1} \ldots(6)$$

2.2 The Classical and Inverse Approaches to Calibration .

2.2.1 The Classical Estimator

Eisenhart (1939) set the stage for classical investigations of absolute calibration problems. His analysis and solution of the inverse estimation problem has come to be called classical. Eisenhart obtained his estimate of ξ by considering the regression of Y on X.

$$E(Y/X = x) = \beta_0 + \beta_1 x$$

The estimated regression line of Y on X is given by

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$= \bar{Y} = \frac{S_{xy}}{S_{xx}}(X - \bar{x}) \quad \ldots(7)$$

Where

$$S_{xy} = \sum_i (x_i - \bar{x})(Y_i - \bar{Y})$$

$$S_{xx} = \sum_i (x_i - \bar{x})^2$$

Eisenhart then inverted equation (2.6) to give an estimator of ξ, the unknown X, which has since become known as the classical estimator .Let it be denoted by ξ_c.Then

$$\xi_c = \bar{x} + \frac{S_{xx}}{S_{xy}}(\bar{\grave{Y}} - \bar{Y})$$

Where $\bar{\bar{Y}}$ is the mean of the m observations at the prediction stage.If one makes the assumption of normal errors in models (2.2)and(2.3) ,then ξ_c is the maximum likelihood estimator of ξ. Eisenhart also produced and interval estimate for ξ based on the t_distribuation with $(n-2) degrees\ of\ freedom.$

Feiller (1954) produced interval estimates for ξ identical to those of Eisenhart using a fiducial argument. Fieller showed that the calibration problem could be reduced to considering the ratio of the means of two normally distributed random variables.

The classical approach to interval estimation has caused consternation over the years because if the slope parameter β_1 is not significantly different from zero the interval is either the whole real line or even two disjoint semi –infinite lines. As a result of this problem,Berkson (1969) and Shulka (1972) obtained asymptotic expressions for the bias and mean square error($M.S.E$) of ξ_c conditional on the event $|\hat{\beta}_1| > 0.$ [1]

2.2.2 Inverse Predictions

At times, a regression model of Y on X is used to make a prediction of the value of X which gave rise to a new observation Y. This is known as an inverse prediction .We illustrate inverse predictions by two examples:

1. Abrade association analyst has regressed the selling price of a product (Y) on its cost (X) for the 15 member firms of the association .The selling price $Y_{h(new)}$ for another firm not belonging to the trade association is known ,and it is desired to estimated to estimate the cost $X_{h(new)}$ for this firm.
2. A regression analysis of the decrease in cholesterol level (Y) against dosage of a new drug (X) has been conducted, based on observation for 50 patients. A physician is treating a new patient for whom the cholesterol level should decrease by $Y_{h(new)}$.It is desired to estimate the appropriate dosage level decrease $Y_{h(new)}$.

The inverse prediction problem is also known as a calibration problem since it is applicable when expensive , and time-consuming measurements (X) based on n observations. The resulting regression model is then used to estimate for a new approximate measurement $Y_{h(new)}$ what is the precise measurement $X_{h(new)}$

In inverse prediction model (3) is assumed as before:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1,2,\dots,n$$

The estimated regression function based on n observations is obtained as usual:

$$\hat{Y}_i = b_0 + b_1 X_i \quad (8)$$

A new observations $Y_{h(new)}$ becomes available, and it is desired to estimate the level $X_{h(new)}$ which gave rise to this new observation .A natural point estimator is obtained by solving (2.7) for X ,given $Y_{h(new)}$:

$$\hat{X}_{h(new)} = \frac{Y_{h(new)} - b_0}{b_1} \qquad b_1 \neq 0$$

Where $\hat{X}_{h(new)}$ denotes the point estimator of the new level $X_{h(new)}$.

$\hat{X}_{h(new)}$ is, indeed the maximum likelihood estimator of $X_{h(new)}$ for regression model(3).

It can be shown that approximate $1-\alpha$ confidence limits for $X_{h(new)}$ are [3]:

$$\hat{X}_{h(new)} \pm t(1-\alpha/2; n-2)s(\hat{X}_{h(new)}) \qquad (9)$$

Where :

$$s^2(\hat{X}_{h(new)}) = \frac{MSE}{b_1^{\ 2}}\left[\ 1+\frac{1}{\text{n}}+\frac{(\hat{X}_{h(new)}-\overline{X})^2}{\sum(X_i-\overline{X})^2}\ \right]$$

III. Application Aspect

3.1 Amonte Carlo Experiment:

The experiment consisted of first determining 20 observation of a standard treatment X .The observations actually used were the exact percentages of the bilirubin in the blood of Jaundice patients and were as follows :6.3, 10, 3.5, 12.5, 20, 17.2, 18.3, 5.8, 9.3, 13, 11.7, 8.9, 22.3, 10.4, 23, 4.3 ,7.5, 8, 19 and 5.

The next step was to decide on "true" values for the regression parameters β_0 and β_1 as well as the error variance σ^2. Based on some studies values for β_0 are chosen as 1.22, 3.75 and 5.79 and the corresponding values for β_1 0 .001, 0 .02, and 0.999.

The population variances used are 3.4, 15.3 and 2.22. This provides three levels(small, medium and large) for each of the parameters β_0 and β_1 and σ^2.

Finally using Eisenhart's simple calibration model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \text{i} = 1,2,3,\dots,\text{n}$$

The correspond values for the "nonstandard treatment" Y are generated for each pair (β_0 , β_1) with ε_i drown from normal distribution. This yielded a bivariate population for the bivariate variable (X , Y).The process of generation the population is explained in the next section.

3.2 Generation of the Population of (X , Y):

The following re followed:

1. Select a value 1.22 *for* β_0 & 0.001 for β_1 and a given value of X say X_1 and define:

$$Y' = \beta_0 + \beta_1 X_1$$

2. From the normal distribution with mean zero and variance $\sigma^2 = 3.4$ select a random number ε and obtain a value of Y as:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

3. Repeat step(2) 500 times. This yields 500 values of Y corresponding to X_1.
4. Repeat step $(1) \to (3)$ for other values of X .
5. Repeat step $(1) \to (4)$ with $\beta_0 = 3.75$ & $\beta_1 = 0.02$ & σ^2 in step(2) equal to 15.3
6. Repeat step $(1) \to (4)$ with $\beta_0 = 5.79$ & $\beta_1 = 0.999$ & $\sigma^2 = 2.22$.

This yields 10,000 values (or population) of the pairs (X,Y) for each of the models:

$$(1)\quad \text{E}(\text{Y}_\text{i}) = 1.22 + 0.001 X_i$$

$$(2)\quad \text{E}(\text{Y}_\text{i}) = 3.75 + 0.02 X_i$$

$$(3)\quad \text{E}(\text{Y}_\text{i}) = 5.79 + 0.999 X_i$$

$$i = 1,2,\dots,500$$

3.3 Sampling and Confidence Intervals:

1) Take a given population.
2) Choose a given sample size. (
3) Select a random sample of the given size.
4) Estimate β_0 & β_1 by $\widehat{\beta}_0$ & $\widehat{\beta}_1$ $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X$ and from this find the predicted value

$$\widehat{X} = \frac{\widehat{Y} - \widehat{\beta}_0}{\widehat{\beta}_1}$$

5) Calculate the error of forecast $(\widehat{X} - X)$ and its square $(\widehat{X} - X)^2$ (
6) Calculate a 95% confidence interval for $\widehat{X}$ using (2.8)
7) If the true value of X falls in the interval put 1 if not put zero.
8) Repeat step (II) to (VII) 1000 times.
9) Count the proportion of time X fall in the calculate interval and also mean of $(\widehat{X} - X)$ and $(\widehat{X} - X)^2$.
10) Repeat steps (II) to (VIIII) other sample size in (II)
11) Repeat steps (I) to (X) for other population.

3.4 Analysis of Results:

Table (4.1)

Population parameters	Samples size	The percentage of times when the period has contained the true value of $X_{h(new)}$	Mean Mse
$\widehat{\beta}_0 = 1.22$	$n = 25$	100%	0.023
$\beta_1 = .001$	$n = 50$	100%	0.0199
$\sigma^2 = 3.4$	$n = 100$	100%	0.2712
$\beta_0 = 3.75$	$n = 25$	99.72%	0.0017
$\beta_1 = .02$	$n = 50$	99.93%	0.779
$\sigma^2 = 15.3$	$n = 100$	99.99%	0.0039
$\beta_0 = 5.79$	$n = 25$	99.31%	0.1817
$\beta_1 = .999$	$n = 50$	99.20%	0.1616
$\sigma^2 = 2.22$	$n = 100$	99.06%	0.1583

Table (4.1) summarizes the result of the simulation experiment. From the table we see that in all cases the actual degree of confidence is much large than the stated degree of confidence i.e. 95%

However the difference between the two decreases with increase in sample size through slightly.

IV. Conclusions

In this paper the calibration regression has been discussed where the data of variable Y has been generated having the confidence interval and the value of MSE.

The objective was to investigate through a simulation experiment the effect of the degree of linear dependency between X & Y as well as sample size on confidence interval estimation of the forecasted value of X .It is shown that changes on these factors have no significant effect on the degree of confidence.

Acknowledgement

I would take this opportunity to thank my research supervisor **Dr. Ahamed Mohamed Abdalla Hamdi**, and special thanks to our great teacher and my idle role **Prof. Zainelabdian A.El Beshir**, professor in alneelain university department of statistic for their support and guidance without which this research would not have been possible.

References

[1]. Christine Osborne . (1991). Statistical Calibration: A Review. School of Mathematical Sciences, University of Bath, Bath BA2 7A Y, England
[2]. Aitchison, J. & Dunsmore, I.R. (1975). Statistical Prediction Analysis. Cambridge University Press. Ali, M.A. & Singh, N. (1981).
[3]. John Neter &William Wasserman & Michael H. Kunter. Applied linear regression models. Richard D. Irwin, ink(1983) ,172-173
[4]. An alternative estimator in inverse linear regression. J. Statist. Comp. 14, 1-15. Barnett, V.D. (1966).
[5]. Discussion of a paper by P. Sprent. J. R. Statist. Soc. B 28, 291-292. Barnett, V.D. (1969).
[6]. Simultaneous pairwise linear structural relationships. Biometrics 25, 129-142. Berkson, J. (1969).
[7]. Estimation of a linear function for a calibration line: consideration of a recent proposal. Technometrics 11, 649-660. Breiman, L. & Friedman, J.H. (1985).
[8]. Estimating optimal transformations for multiple regression and correlation. J. Am. Statist. Assoc. 80, 580-597. Brown, G.H. (1979).
[9]. An optimisation criterion for linear inverse estimation. Technometrics 21, 575-579. Brown, P.J. (1982).
[10]. Multivariate calibration (with discussion). J. R. Statist. Soc. B 44, 287-321. Brown, P.J. & Sundberg, R. (1987)
[11]. Confidence and conflict in multivariate calibration. J. R. Statist. Soc. B 49, 46-57. Brown, P.J. & Sundberg