

BULLETIN OF ENVIRONMENTAL SCIENCE & SUSTAINABLE MANAGEMENT



Website: http://journal.hibiscuspublisher.com/index.php/BESSM/index

Short Communication Testing for the Equality of Variance for the Four-parameter Logistic (4PL) Model for the Detection of Vibrio cholerae DNA with **Polystyrene-coacrylic Acid Composite Nanospheres**

Mohd Yunus Shukor¹*

¹Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia.

> *Corresponding author: Mohd Yunus Shukor, Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor, Malaysia

Email: mohdyunus@upm.edu.my

HISTORY

Received: 23rd Aug 2021 Received in revised form: 24th Oct 2021 Accepted: 12th Nov 2021

KEYWORDS

4-parameter logistic model Homogeneity of variance Vibrio cholera polystyrene-coacrylic acid composite nanosensor

ABSTRACT

When it comes to finding the best fit of nonlinear curves to acceptable models, linear regression with least squares is the most effective technique. Because residuals (the difference between observed and predicted data) must follow a normal distribution and the data must be free of outliers and uniform variance, statistical tests are used to identify the most appropriate model for a given situation (homoscedasticity). If all of these characteristics are satisfied, the system is said to be robust. In parametric nonlinear regression, one of the numerous assumptions is that the within-group variances of the groups are all the same, which is one of several assumptions (exhibit homoscedasticity). If the variances vary from one another (show heteroscedasticity), then the model is not statistically competent to describe the data as a whole. Data on the detection of Vibrio cholerae DNA with polystyrene-coacrylic acid composite nanospheres as modelled using the nonlinear four-parameter logistic (4PL) regression was preliminary check for homogeneity of variance using the Bartlett's and Levene's tests. It was found that the critical value of 2 was 28.869, according to Bartlett's test findings. Excel's CHIDIST function yielded a probability of 0.389 (not significant), suggesting that the variances of the residuals did not change significantly. The p-value for Levenes's test was 0.917, indicating that there were no distinct changes between the residual variances meaning that the use of the 4-PL model in fitting the data was adequate statistically.

INTRODUCTION

There are still many nations where cholera is a serious public health problem. 1 34 nations reported a total of 1,227,391 cases and 5654 fatalities in 2017, providing a global mortality rate of 0.5 percent [1.] in 2017. Cholera outbreaks in Latin America for the first time in more than a century, the first emergence of Vibrio cholerae O139 in Asian nations, and the first appearance of hybrid El Tor strains in Asia and Africa are all noteworthy events in the history of cholera epidemiology. As of May 2017, there has been 8,09,000 cholera infections and 9670 fatalities [5] in Haiti since the outbreak began in October 2010. The cholera pandemic in Yemen between April and July 2017 was the second largest in history after the Haitian cholera outbreak, with about 1700 deaths. It's necessary to have a precise approach for detecting this pathogen. Emerging detection technique that uses

biosensor systems, however, are not prone to statistical detection issues, as some do not utilize proper nonlinear regression tool when the calibration curves are obviously not linear. Furthermore, proper use of nonlinear regression approach requires the data to be normally distributed and the variance homogenous.

For this and other nonlinear regression models to be effective, the data must have a normal distribution, be devoid of outliers, and have a homogeneous variance, all of which may be additional statistical achieved bv doing analysis (homoscedasticity). The test is considered robust if all of these conditions are met. Nonlinear regression using least squares is the best strategy for finding the best fit of nonlinear curves to suitable models. Because residuals (the difference between observed and projected data) must follow a normal and the data

must be devoid of outliers and homogeneous variance, statistical tests are used to select the optimum model (homoscedasticity). It is said to be robust if all of these conditions are met. One of the many notions of nonlinear regression that is parametric-based is that the groups' within-group variances exhibit homoscedasticity or must be the same.

If the variances show heteroscedasticity, then they deviate from one another leading to the model unable to capture the data. Numerous tests to verify the equality of variances on residual data include Bartlett, Cochran Leven and Brown-Forsythe for three or more samples, where to test for homogeneity of two variances, the F-test is used. The Bartlett's test would be employed in this work to test for homogeneity of variance (homoscedasticity) or equality of variance of the residuals [1-4]. Based on this, data on the detection of Vibrio cholerae DNA with polystyrene-coacrylic acid composite nanospheres from a published result [5] as modelled using the nonlinear fourparameter logistic (4PL) regression was preliminary check for homogeneity of variance using the Bartlett's and Levene's tests.

MATERIALS AND METHODS

Acquisition of Data

Data on the detection of Vibrio cholerae DNA with polystyrenecoacrylic acid composite nanospheres from Figure 6b [5] as modelled using the four parameter logistic model (published elsewhere) was utilized in this study. The data were processed using the software Webplotdigitizer 2.5 [6] which digitizes the scanned figure into a comma separated value [7].

Test for equality of variance

Bartlett's test

$$\chi^{2} = \frac{(N-k)\ln(S_{p}^{2}) - \sum_{i=1}^{k} (N_{i}-1)\ln(S_{i}^{2})}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^{k} \left(\frac{1}{N_{i}-1}\right) - \frac{1}{N-k}\right)}$$

Where the variance pooled estimate are;

 $N = \sum_{i=1}^{k} (n_i)$ And

 $S_p^2 = \frac{1}{N-k} \sum_{i} (n_i - 1) S_i^2$

The distribution of the test statistics is roughly as follows; χ^2_{k-1}

$$\chi^2 > \chi^2_{k-1,\alpha} \qquad \qquad \chi^2_{k-1,\alpha}$$

 S_i^2

 S_p^2

Levene's test

The test statistic, W, is defined as follows:

$$W = \frac{(N-k)}{(k-1)} \frac{\sum_{i=1}^{k} N_i (Z_i - Z_i)^2}{\sum_{i=1}^{k} N_i (Z_{ij} - Z_i)^2}$$
(Eqn. 2)

Where,

k is the variety of distinct categories into which the instances in the sample may be divided,

N is the total number of cases in all groups,

 Y_{ij} is the value of the measured variable for the *j*-th case from the *i*-th group,

 N_i is the number of cases in the *i*-th group,

$$Z_{ij} = \begin{cases} |Y_{ij} - \overline{Y}_{i.}|, & \overline{Y}_{i.} \text{ is a mean of the } i\text{-th group,} \\ |Y_{ij} - \overline{Y}_{i.}|, & \overline{Y}_{i.} \text{ is a median of the } i\text{-th group.} \end{cases}$$

Levene's test statistic is represented by W has N-k and k-ldegrees of freedom. The selected level of significance or alpha is usually 0.05 or 0.01. The Leven test statistics is roughly Fdistributed and is the significance of the outcome w of W tested against the quantile of the F-distribution, which is F with k-1 and *N-k* degrees of freedom.

RESULTS AND DISCUSSION

As concentrations grow, the dispersion of the points around the centre line shows signs of likely heteroscedasticity, which may be interpreted as a sign of equal variance or homoscedasticity in a residual graph. Variances are not uniform under these circumstances. The purpose of the Bartlett and Levene tests is to corroborate this finding (Massart et al., 1997; Miller and Miller, 2000). Homoscedasticity (homoscedasticity) and equality of residual variance may be tested using the Bartlett tests [8]. It was found that the critical value of 2 was 28.869, according to Bartlett's test findings. Excel's CHIDIST function yielded a probability of 0.389 (not significant), suggesting that the variances of the residuals did not change significantly. The pvalue for Levenes's test was 0.917, indicating that there were no distinct changes between the residual variances.

Because of the robustness of Levene's test for homogeneity of variances (Levene, 1960), it is particularly well suited for detecting if data from three or more samples are homoscedastic in nature. As a result, it's much less vulnerable to non-normality than the Bartlett's test (Snedecor & Cochrane, 1989), which must be employed only when there is sufficient evidence that the test outcomes come from a normal distribution (Snedecor & Cochrane, 1989). If we want to know if the variances in a set of analytical data are homogeneous or not, the Levene test is what we should use.

In regression models, homoscedasticity assumption means "identical variance" is crucial. The word "homoscedasticity" refers to a scenario wherein the "noise" or error term are subject to random disturbances in the connection between the dependent and the independent variables and is equal for all independent variable values. When the amount of the error term varies across values of an independent variable, heteroscedasticity (violation of homoscedasticity) is evident.

2

The consequence of breaching the homoscedasticity premise is proportional to the extent of heteroscedasticity. A bias in the standard errors leads to heteroscedasticity and is a more significant concern. Despite the relevance of the standard error in performing significance tests and determining confidence intervals, incorrect standard errors may result in incorrect conclusions about the significance of the regression coefficients. Incorrect standard errors may result in incorrect conclusions about the importance of the regression coefficients. In order to adjust for this prejudice, several statistical tools give the option of robust standard errors. Weighted least squares regression handles this issue as well, however it necessitates the use of a number of extra constraints. It is also possible to cope with heteroscedasticity by applying one of the variance stabilising transformations to convert the dependent variable, which is yet another option. Generally, nevertheless, considering the resilient nature of OLS regression, the violation of the homoscedasticity assumption must be rather strong in order to effect significant problems.

To conclude, the test for homogeneity of variance for the four-parameter logistic (4PL) model for the detection of Vibrio cholerae DNA using polystyrene-coacrylic acid composite nanospheres was found to be satisfactory. There is a solution if the conclusions drawn did not conform to the assumption of homogeneity of variance, which can also be ameliorated by data transformations using a variety of methodology for this study. For example, size data can be log transformed, count data can be subjected to square-root transformation while proportional data can be subjected to Arcsine transformation.

REFERENCES

- 1. Snedecor GW, Cochran WG. Statistical methods. 7th ed. Ames Iowa: Iowa State University Press; 1980.
- 2 Motulsky HL Ransnas LA. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. FASEB J Off Publ Fed Am Soc Exp Biol. 1987;1(5):365-74.
- López S, Prieto M, Dijkstra J, Dhanoa MS, France J. Statistical 3. evaluation of mathematical models for microbial growth. Int J Food Microbiol. 2004;96(3):289-300.
- Mcdonald JH, Dunn KW. Statistical tests for measures of 4. colocalization in biological microscopy. J Microsc. 2013;252(3):295-302.
- Rahman M, Heng LY, Futra D, Ling TL. Ultrasensitive Biosensor 5. for the Detection of Vibrio cholerae DNA with Polystyrene-coacrylic Acid Composite Nanospheres. Nanoscale Res Lett. 2017 Aug 1;12(1):474.
- WebPlotDigitizer. Rohatgi Α. http://arohatgi.info/WebPlotDigitizer/app/ Accessed June 2 2014.; 2015
- 7. Khare KS, Phelan Jr FR. Quantitative comparison of atomistic simulations with experiment for a cross-linked epoxy: A specific volume-cooling rate analysis. Macromolecules. 2018;51(2):564-75.
- Jarque CM, Bera AK. Efficient tests for normality, 8. homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. Econ Lett. 1981;7(4):313-8.