

## **Filling missing data using interpolation methods: Study on the effect of fitting distribution**

### **Abstract**

The presence of missing values in statistical survey data is an important issue to deal with. These data usually contained missing values due to many factors such as machine failures, changes in the siting monitors, routine maintenance and human error. Incomplete data set usually cause bias due to differences between observed and unobserved data. Therefore, it is important to ensure that the data analyzed are of high quality. A straightforward approach to deal with this problem is to ignore the missing data and to discard those incomplete cases from the data set. This approach is generally not valid for time-series prediction, in which the value of a system typically depends on the historical time data of the system. One approach that commonly used for the treatment of this missing item is adoption of imputation technique. This paper discusses three interpolation methods that are linear, quadratic and cubic. A total of 8577 observations of PM<sub>10</sub> data for a year were used to compare between the three methods when fitting the Gamma distribution. The goodness-of-fit were obtained using three performance indicators that are mean absolute error (MAE), root mean squared error (RMSE) and coefficient of determination ( $R^2$ ). The results shows that the linear interpolation method provides a very good fit to the data.

**Keywords;** Gamma Distribution, Goodness-of-Fit, Missing Values