International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321- 4767 P-ISSN: 2321-4759
www.ijmsi.org Volume 9 Issue 5 || May, 2021 || PP 07-10

A Note on the Regression Control Chart

Claude R. Superville, PhD, FRSS, FIMA
JHJ School of Business, Texas Southern University, USA

Marion Smith, PhD
JHJ School of Business, Texas Southern University, USA

***ABSTRACT**: This article provides the mathematical background for the Least Squares method, that serves as the basis for the development of the Simple Linear Regression equation. The Regression Control Chart uses the predicted value y from Simple Linear Regression as the chart's center line. An example using recent incidents of COVID in a large metropolitan area, is used to illustrate the use of the Regression Control Chart.*

***KEYWORDS**: Simple Linear Regression, Regression Control Chart, in control*

Date of Submission: 08-05-2021 Date of Acceptance: 22-05-2021

I. INTRODUCTION

Non-uniformity or variation is a fact of life in all manufacturing and service industries. A manufacturing employee, who takes reading on the flow pressure produced by an oil pump on an hourly basis, will find that the readings are never the same. This non-uniformity is referred to as variation. Decreases in variation of the level of product or service result in improvements in quality of conformance.

A control chart is a plot of observations from a process over time. Pegels (1995) suggests that "control charting involves the charting of statistics on a chart in such a way that deviations from a standard can be quickly observed, and action can be taken to correct the undesirable variation." Typically, a process can be summarized by an average value ($\bar{x}$) and a measure of variation, the range (R).

If a product is to consistently meet the customers' fitness for use criterion, generally, it should be produced by a stable process. A stable process is better able to meet customers' needs for a uniform, high quality product. The purpose of the control chart is process stability, through the reduction of process variability. This is done by distinguishing between common cause and special cause variation.

Common causes are small, uncontrollable influences that are an inherent part of the process. They cannot be removed from the process without basic changes in the process that usually require management action. Special causes are larger, unusual influences that can be removed from the process. A process that is operating with only common cause variation present is said to be in statistical control. The control chart is used to determine whether a process is in-control.

II. MATHEMATICS OF THE REGRESSION CONTROL CHART

The Regression Control Chart is based on monitoring and control of an independent y variable from Simple Linear Regression as the center line (CL) with upper control limits (UCL) and lower control limits (LCL). Simple Linear Regression attempts to find the best fit line through a set of observations. The method used to determine the coefficients of the regression equation is the Least Squares method. With this method, the sum of the squared residuals from the regression line are minimized.

The Least Squares Regression (LSR) equation is:

$$\hat{y} = \beta0 + \beta1x \qquad (1)$$

where β0 = estimate of the y-intercept and β1 = estimate of the slope.

To minimize the sum of the squared residuals:

$ei = yi - \hat{y}i$, for i = 1, 2, 3 … n ( )

$$\text{minimize } \Sigma\, ei^2 = \Sigma\, (yi - \hat{y}i)^2 \qquad (2)$$
$$= \Sigma\, [yi - (\beta0 + \beta1x_i)]^2$$

$E(\beta0,\beta1) = \Sigma\, [yi - (\beta0 + \beta1x_i)]^2$ as a function of β0 and β1.

To minimize E(β0,β1), set the partial derivatives equal to zero and find solutions:

$\partial E(\beta0,\beta1) / \partial\beta0) = 0$

$\partial E(\beta0,\beta1) / \partial\beta1) = 0$

In expanded form:

$E(\beta0,\beta1) = \Sigma\, [yi - (\beta0 + \beta1x_i)]^2$

$\quad = [y1 - \beta0 - \beta1x1]^2 + [y2 - \beta0 - \beta1x2]^2 + \dots + [yn - \beta0 - \beta1xn]^2$

If $\partial E(\beta0,\beta1) / \partial\beta0) = 0$ and solve for $\beta0$:

$\partial E(\beta0,\beta1) / \partial\beta0) = 2(y1 - \beta0 - \beta1x1)(-1) + 2(y2 - \beta0 - \beta1x2)(-1) + \dots + 2(yn - \beta0 - \beta1xn)(-1) = 0$

$\quad = -2\Sigma(yi - \beta0 - \beta1xi) = 0$

$\Sigma yi - \Sigma\beta0 - \Sigma\beta1xi = 0$

$\Sigma yi - n\beta0 - \beta1\Sigma xi = 0$

$\beta0 = \frac{\Sigma yi}{n} - \beta1\frac{\Sigma xi}{n}$

$\beta0 = \bar{y} - \beta1\bar{x}$ (3)

Now if $\partial E(\beta0,\beta1) / \partial\beta1) = 0$, let $\beta0 = \bar{y} - \beta1\bar{x}$ and solve for $\beta1$:

$\partial E(\beta0,\beta1) / \partial\beta1) = -2\Sigma\, xi\, (yi - \beta0 - \beta1xi) = -2\Sigma\, xi\, (yi - (\bar{y} - \beta1\bar{x}) - \beta1xi) = -2\Sigma(xiyi - xi\bar{y} - \beta1xi\bar{x}) - \beta1xi^2) = 0$

$\quad = \Sigma(xiyi - xi\bar{y} - \beta1xi\bar{x}) - \beta1xi^2) = 0$

$\quad \beta1\, \Sigma\, (xi^2 - xi\bar{x}) = \Sigma(xiyi - xi\bar{y})$

$\quad \beta1 = \Sigma(xiyi - xi\bar{y}) / \Sigma(xi^2 - xi\bar{x})$

$\quad \beta1 = [\Sigma(xiyi - \bar{y}\Sigma xi)] / [\Sigma xi^2 - \bar{x}\Sigma xi]$

$\quad \beta1 = [\Sigma(xiyi - ni\bar{x}\bar{y}) / [\Sigma xi^2 - n\bar{x}^2]$ (4)

In summary:

$\beta0 = \bar{y} - \beta1\bar{x}$

$\beta1 = [\Sigma(xiyi - ni\bar{x}\bar{y}) / [\Sigma xi^2 - n\bar{x}^2]$

and the Least Squares Regression line is:

$\hat{y} = \beta0 + \beta1x$

The idea of a regression control chart was initiated by DiPaola (1945). Wallis and Roberts (1956) and Mandel (1969) suggest a chart consisting of a regression line with confidence intervals as control limits but clearly state a preference for the use of tolerance limits as control limits. It is assumed that the y values are normally and independently distributed with a mean value estimated by the regression line. The standard error, s_e, independent of the values of x, is estimated from the deviations of the actual y values from the predicted y values from the regression line.

The Regression Control Chart is of the form:

$\hat{y} \pm ks_e$

where k is a constant (typically k= 2 or 3). Control limits set to k=3 is equivalent to a 99.7% confidence interval.

III. AN APPLICATION OF THE REGRESSION CONTROL CHART

The cumulative number of cases of COVID in Harris County, Texas is retrieved from USA Facts (2021) for a period of 347 days from February 1, 2020 to January 2, 2021. Harris County, Texas includes the Houston metropolitan area. There are no cases of COVID in the first 43 days until March 4, 2021 so there are eliminated from the analysis. Figure 1 shows the cumulative number of COVID cases in Harris County, Texas over a period of 304 days from March 5, 2020 to January 2, 2021. The upward trend indicates a positive growth rate in the number of COVID cases.

Table 1 shows the Regression output for the cumulative number of COVID cases. The prediction equation for the cumulative number of COVID cases is:

$\hat{y} = -78689.3 + 840.46x$

with $R^2 = 97.7\%$, F= 6360.51 and $p < 0.0$ indicates that the model is good fit to the data.

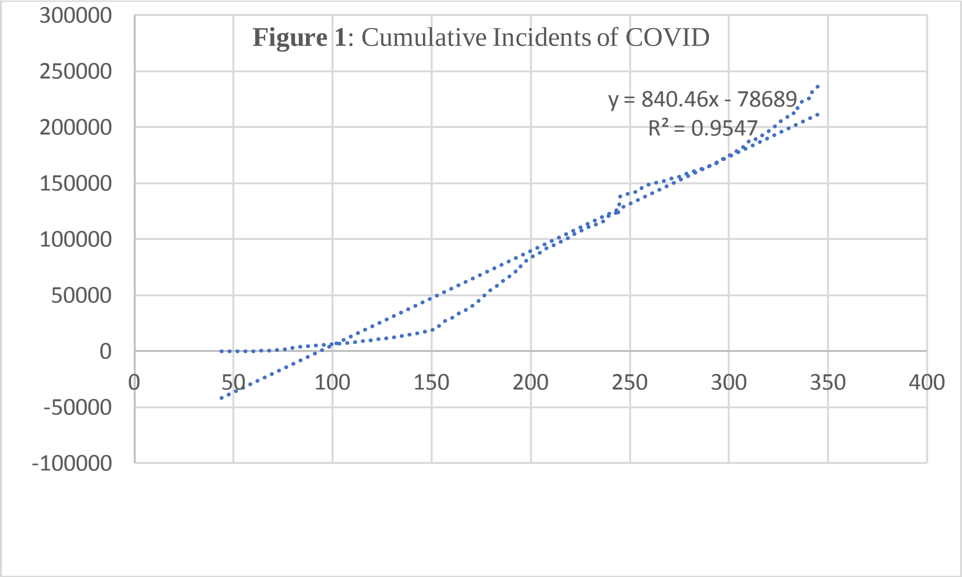
Figure 1: Cumulative Incidents of COVID

Table 1: Regression Output for Number of COVID Cases

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.977073
R Square	0.954672
Adjusted R Square	0.954522
Standard Error	16124.67
Observations	304

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.65E+12	1.65E+12	6360.508	6.1E-205
Residual	302	7.85E+10	2.6E+08		
Total	303	1.73E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.7%*	*Upper 99.7%*
Intercept	-78689.3	2258.299	-34.8445	7.9E-108	-83133.3	-74245.3	-85446.2	-71932.5
X	840.4642	10.53837	79.75279	6.1E-205	819.7263	861.2022	808.9333	871.9952

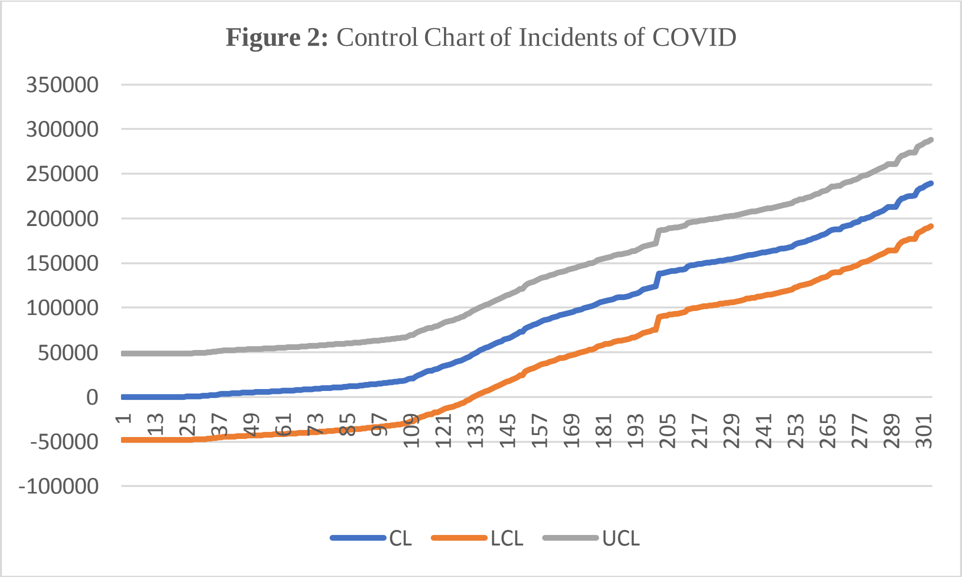

Figure 2: Control Chart of Incidents of COVID

Figure 2 displays a Regression Control Chart on the predicted y values with:

Center Line (CL): $\hat{y} = -78689.3+ 840.46x$, for $x_i = 1, 2, 3, \ldots, 304$
LCL: $\hat{y} + 3(16124.67)$
UCL: $\hat{y} - 3(16124.67)$
where $s_e = 16124.67$ from Table 1. There are no points exceeding the control limits indicating that the process is in control. There are no days when an unusual number of COVID cases occurred.

IV. CONCLUSIONS

The Regression Control Chart uses the predicted value y from Simple Linear Regression as the chart's center line. This article provides the mathematical derivation for the Least Squares method, used to determine the coefficients of the predicted y equation used in the Regression Control Chart. An application using recent incidents of COVID in a large metropolitan area, is used to illustrate the use of the Regression Control Chart.

REFERENCES

[1]. DiPaola, P.P. (1945). Use of Correlation in Quality Control, Industrial Quality Control, 2(1), 10-14.
[2]. Mandel, B.J. (1969). "The Regression Control Chart," Journal of Quality Technology, 1(1), 1-9.
[3]. Pegels, C.C. (1995). Total Quality Management, Danvers, MA.: Boyd and Frasier.
[4]. Wallis, W.A, and Roberts, H.V. (1956). Statistics: A New Approach, The Free press, Chicago, IL, 549-553.
[5]. USA Facts (2021). USA Facts Our Nation, in Numbers, https://usafacts.org/

Claude R. Superville, et. al. "A Note on the Regression Control Chart." *International Journal of Mathematics and Statistics Invention (IJMSI)*, vol. 09(05), 2021, pp. 07-10.