USE OF LOGISTIC REGRESSION TO PREDIT EFFICIENCY OF A COLLECTION CENTER IN A REVERSE SUPPLY CHAIN

KISHORE K. POCHAMPALLY

Department of Management Science, Southern New Hampshire University, 2500 North River Road, Manchester, NH, USA
E-mail: k.pochampally@snhu.edu

Abstract - This paper demonstrates with an illustrative example, how to predict the efficiency of a collection center in a reverse supply chain, using logistic regression. A model is built to predict the probability of on-time delivery based on the number of trucks owned by the collection center. Excel®[1] and Minitab®[1] screenshots are used to clarify the step-by-step approach.

Keywords - Collection Center, Reverse Supply Chain, Logistic Regression

I. INTRODUCTION

Today, there is much emphasis on reprocessing (recycle / remanufacture) of end-of-use products due to growing population, finite resources, and limited disposal capacities (Gungor and Gupta, 1999). Hence, reverse supply chains are designed and implemented to collect end-of-use products from consumers, reprocess them at recovery facilities, and then sell recycled goods and remanufactured products on the market (Alkhayyal and Gupta, 2018). An important driver for companies engaged in a reverse supply chain is that many end-of-use products represent a resource for recoverable value (Ilgin and Gupta, 2010). Furthermore, increasing environmental consciousness of consumers and regulations of governments offer the possibility to satisfy customer demand from remanufactured products instead of exclusively producing new products. This mode of demand satisfaction may offer a cost advantage over production of new products (Gupta, 2013). Another advantage is that remanufacturing lead times may be much shorter than production lead times (Gupta and Ilgin, 2018).

Implementation of a reverse supply chain requires at least three parties: collection centers where consumers return end-of-use products, recovery facilities where reprocessing (remanufacturing or recycling) is performed, and demand centers where customers buy reprocessed products (Pochampally et al., 2009).

Timely delivery of proper end-of-use products by collection centers to recovery facilities is critical to meet the ever-increasing demand for recycled goods and remanufactured products. To this end, this paper demonstrates with an illustrative example, how to predict the efficiency of a collection center using logistic regression. In this example, a model is built to predict the probability of on-time delivery based on the number of trucks owned by the collection center.

II. LOGISTIC REGRESSION

A logistic regression equation to predict the probability (p) of an event, based on the value of one explanatory (independent) variable x, can be written as follows in equation (1):

$$p = \frac{e^{a+bx}}{1+e^{a+bx}} \quad (1)$$

where e is the Euler's number (2.71828), and a and b are constants (Groebner et al., 2017).
A scatter plot between p and x, for given values of a and b, looks like an "S-curve" as shown in Figure 1.

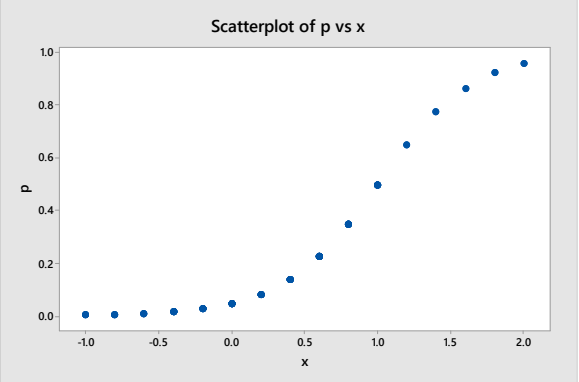


Figure 1. Scatter Plot of *p* and *x* in Equation 1.

The best values of a and b are those that best fit the existing data, viz., actual values of (x, p) pairs. Equation (1) can be transformed into equation (2), which then can be used to apply the least squares estimation method like in a linear regression problem.

$$Log\left(\frac{p}{1-p}\right) = a + bx \quad (2)$$

Since (p / 1 – p) represents the "odds" of the event of interest, we can also write equation (2) as equation (3):

$$Log(odds) = a + bx \quad (3)$$

An estimate of log(odds) for a given x, and for given a and b, is called the "fitted" log(odds) for that x, and can be calculated using the right-hand side of equation (2).

Actual log(odds) for a given (x, p) pair can be calculated using the left-hand side of equation (2).

A residual for a given x can be defined as the difference between the actual log(odds) and the fitted log(odds), as shown in equation (4) The example in the next section will clarify this.

Residual = Actual log(odds) – Fitted log(odds) (4)

The best values of a and b are the ones that minimize the sum of squares of all the residuals.

III. EXAMPLE

In this example, the following data, as shown in Table 1, is collected for eleven collection centers: number of trucks owned (x) and proportion of on-time deliveries (p).

A scatter plot, shown in Figure 2, is plotted for the data in Table 1. Evidently, the plot shows an "S-curve" which motivates us to fit a logistic regression model as in equation (1).

In order to obtain the best fit (best values of a and b) for the data in this example, we must start with some random initial values of a and b. Let the initial values be $a = 1$ and $b = 2$.

Then, for each collection center, the actual log(odds), fitted log(odds), and residuals, are calculated and shown in Table 2. For example, for collection center 1, the calculations are illustrated below:

Actual log(odds) = log (p / (1 – p)) = log (0.225 / (1 – 0.225)) = log (0.29) = -1.24.

Fitted log(odds) = $a + bx = 1 + 2*20 = 41$.

Residual = Actual log(odds) – Fitted log(odds) = -1.24 – 41 = -42.24.

Collection Center	x	p
1	20	0.225
2	40	0.236
3	60	0.398
4	80	0.628
5	100	0.678
6	120	0.795
7	140	0.853
8	160	0.86
9	180	0.921
10	200	0.94
11	240	0.968

Table 1. Data

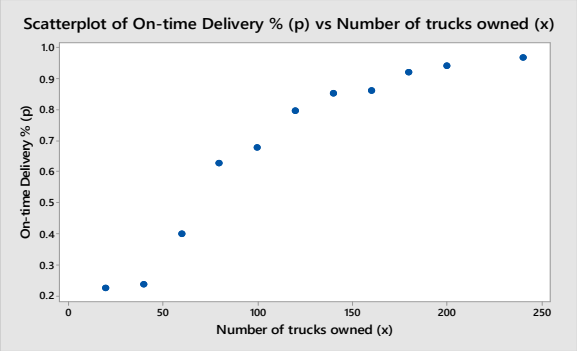


Figure 2. Scatter Plot of Data

Collection Center	Actual Log (odds)	Fitted Log (odds)	Residual
1	-1.24	41	-42.24
2	-1.17	81	-82.17
3	-0.41	121	-121.4
4	0.524	161	-160.5
5	0.745	201	-200.3
6	1.355	241	-239.6
7	1.758	281	-279.2
8	1.815	321	-319.2
9	2.456	361	-358.5
10	2.752	401	-398.2
11	3.409	481	-477.6

Table 2. Residuals

The sum of squares of all of the residuals is then calculated as shown below:

$$(-42.24)^2 + (-82.17)^2 + \ldots (-477.6)^2 = 841{,}666$$

The Solver tool in Excel® can then be used as shown in the screenshot in Figure 3, to obtain the best values of a and b (J1 and J2 cells, respectively, in the screenshot), which minimize the sum of squares of residuals (G13 cell in the screenshot).

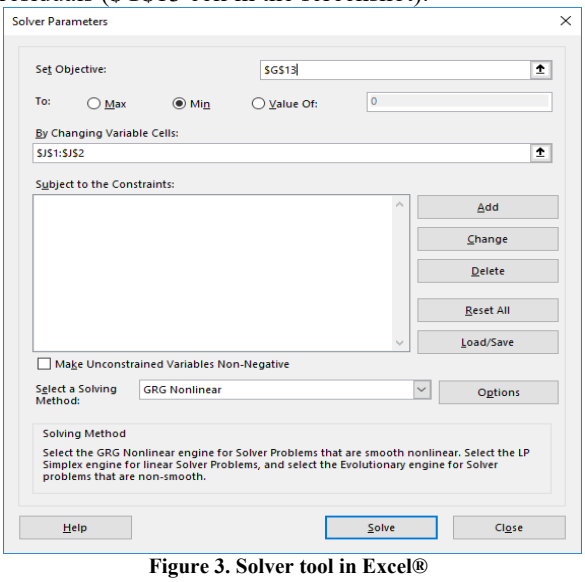


Figure 3. Solver tool in Excel®

The best values of a and b, as given by the Solver tool, are -1.603 and 0.022, respectively, and the corresponding sum of squares of residuals is 0.655. Hence, the logistic regression model for this example is as shown in equation (5).

$$p = \frac{e^{-1.603+0.022x}}{1+e^{-1.603+0.022x}} \quad (5)$$

Equation (5) can be used to predict the on-time delivery performance (p) of a collection center, for a given number of trucks (x) owned by the collection center. For example, if $x = 90$ trucks, $p = 0.60$.

Minitab® can be used to reproduce the above result in just a few steps, as follows:

i. Arrange the worksheet as in Table 3. Notice that the word 'Actual' is removed from 'Actual log(odds)' because, otherwise, Minitab® would use the same label in the output which is a regression equation for fitted log(odds).
ii. Stat -> Regression -> Fitted Line Plot (This opens the dialog box shown in Figure 4).
iii. Click on OK. This results in the graph and equation shown in Figure 5 (The minor difference from the Excel output is due to rounding error). A P-value of 0.000 for *Regression* is also produced in the session window. Since this value is less than the standard level of significance (0.05), it means that the logistic regression model is statistically significant.

Collection Center	x	Log (odds)
1	20	-1.24
2	40	-1.17
3	60	-0.41
4	80	0.524
5	100	0.745
6	120	1.355
7	140	1.758
8	160	1.815
9	180	2.456
10	200	2.752
11	240	3.409

Table 3. Minitab® Worksheet

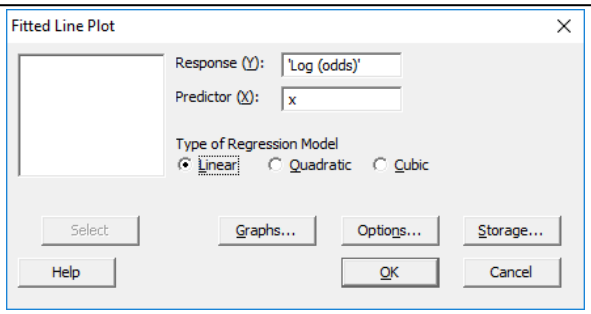


Figure 4. Minitab® Dialog Box

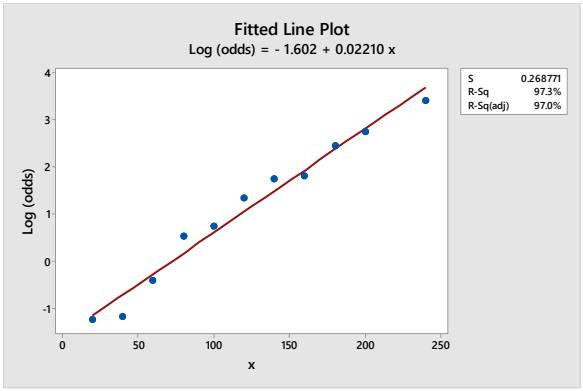


Figure 5. Minitab® output

REFERENCES

[1] Alkhayyal, B. A. and Gupta, S. M., "A Linear Physical Programming Approach for Evaluating Collection Centers for End-Of-Life Products",Applications of Management Science, Emerald Publishing Limited, United Kingdom, Vol. 19, 65-79, 2018.

[2] Groebner, D., Shannon, P., and Fry, P.,"Business Statistics: A Decision-Making Approach",Pearson, United States of America, 10th edition, ISBN: 978-0134496498, 2017.

[3] Gungor, A., and Gupta, S. M., "Issues in environmentally conscious manufacturing and product recovery: A survey", Computers and Industrial Engineering, Vol. 36, No. 4, 811 - 853, 1999.

[4] Gupta, S. M. and Ilgin, M. A., "Multiple Criteria Decision Making Applications in Environmentally Conscious Manufacturing and Product Recovery", CRC Press, Boca Raton, Florida, ISBN: 978-1498700658, 2018.

[5] Gupta, S. M., "Reverse Supply Chains: Issues and Analysis", CRC Press, Boca Raton, Florida, ISBN: 978-1439899021, 2013.

[6] Ilgin, M. A. and Gupta, S. M., "Environmentally Conscious Manufacturing and Product Recovery (ECMPRO): A Review of the State of the Art", Journal of Environmental Management, Vol. 91, No. 3, 563-591, 2010.

[7] Pochampally, K. K., Nukala, S. and Gupta, S. M., "Strategic Planning Models for Reverse and Closed-loop Supply Chains", CRC Press, Boca Raton, Florida, ISBN: 9781420054781, 2009.

★★★