International Journal of Economics, Commerce and Management
United Kingdom Vol. VI, Issue 4, April 2018
http://ijecm.co.uk/ ISSN 2348 0386

A STUDY OF RELATIONSHIPS BETWEEN PRODUCTION LEVELS AND SALES OF CARBONATED SOFT DRINKS (CASE OF COCACOLA, ILORIN PLANT, NIGERIA)

Awoyemi Samson Oyebode
Department of Training Technology
Agricultural and Rural Management Training Institute (ARMTI), Ilorin, Nigeria
bodesok2012@gmail.com

Adekunle Johnson Ade
Institute of Basic and Applied Sciences
Kwara State Polytechnic, Ilorin, Kwara State, Nigeria
adekunlejade@yahoo.com

Fadayomi Akinwumi Festus
Department of Statistics, Federal School of Statistics, IBADAN Campus, Ibadan, Nigeria

Abstract

This study is relatively gentle but reasonably comprehensive introduction to the principle and technique of practical multivariate regression analysis applying the techniques to the production levels and sales of a Bottling Company. Secondary data was obtained from Nigerian Bottling Company Plc, Ilorin Plant, and it includes the amount of carbonated soft drinks consumed at both dry and rainy season tagged ($Y^1 = Y_1\ Y_2$) also quantity produced at the same period of the years (tagged $X^1 = X_3\ X_4$) respectively. The result of analysis shows a negative correlation -0.58 and 0.08 between sales during the dry season and the rainy season which suggest that sales during these two seasons does not follow the same pattern. On the other hand, correlation between sales during dry season and quantity produced during the same period is between 0.75 and 0.9388 which shows a very high positive correlation and this suggest that the sales of carbonated drink during the dry season and the quantity produced during the same period are related.

Keywords: Consumption, Carbonated drinks, Product level, Sales, Nigeria

INTRODUCTION

Much of scientific studies are directed toward discovering the form of relationship between variables and predicting the value of a variable from some functional relationship. This is one of the most important areas of applied statistics.

The term Regression was first used by Sir Francies Galton (1822 – 1911) to describe the phenomenon by which offspring of unusually tall parents tend to be shorter than their parents and the offspring of the unusually short parents tend to be taller. He characterized this phenomenon as a tendency for offspring to revert back or regress towards the mean or average height of the general population.

A statistical tool, which helps to predict one variable from the other variable or variables, on the basis of, assumed nature of the association or relationship between the variables is known as regression analysis. The variable being predicted is usually referred to as the unknown or dependent variable because its values are assumed to be dependent on the values of the other variable or variables referred to as the independent variables or predictor variables. On the other hand the values of the dependent variable are determined by the value of the independent variables. In other words, the dependent variables are a function of the independent variable. If a linear or linearizable form is assumed, then this is usually represented explicitly by an equation $Y = \beta o + \beta_1 X$, where Y is the value of the dependent variable, X is the value of independent variable, and βo and β_1 are constants. Each different straight line has different values of βo and $\beta_{1.}$ Bo is the value of Y, the dependent variable when X equals Zero, β_1 is the slope or gradient of the line. The gradient of a line is the change in Y for unit increase in the value of X. In each case, we say that the dependent variable is regressed on the independent variable. The variables involved are assumed to be

Correlation is said to exist when the two groups or series of items vary together directly or inversely. That is when in a given group of individuals' measures of two characteristics of each individual are obtained. It may frequently be observed that the two measures of each individual have a tendency to occupy almost the same relative position in their respective distributions. When this kind of phenomenon is observed we say that the two characteristics are mutually related or correlated.

NIGERIAN BOTTLING COMPANY HISTORICAL BACKGROUND/PERSPECTIVE

Coca-cola, the product that gave the world its best-known taste was developed under modest circumstances in Atlanta, Georgia on May 8^{th}, 1886. Dr. John Styth Pemberton first produced what was to become the coca-cola syrup (NBC periodicals May 1993 edition), in a three-legged brass pot with a boat oar in his backyard. He carried the jug of the new product down the street to Jacob's pharmacy where it was placed on sale for 5 cents a glass as a soda fountain drink. Carbonated the new product down the street to Jacob's pharmacy where it was placed on sale for 5 cents a glass as a soda fountain drink. Carbonated water was teamed with the syrup to produce a drink that was at once "delicious and refreshing". Dr. Pemberton's partner and bookkeeper Frank M. Robinson suggested the name "coca-cola" (NBC periodicals May 1993 edition), and designed the flowing script that distinguishes the famous trade mark.

COCA-COLA IN NIGERIA

Coca-cola first came to Nigeria in 1953 when Nigerian Bottling Company (NBC) set up its first plant in Lagos. It was to be the beginning of an exciting story of growth and development particularly in the recent past. Nigerian Bottling Company is today Nigeria's number one bottler of soft drinks, selling more than 10,000,000 bottles per day. Currently, there are 20 plants in various parts of the Federation. Fanta is by far the number one best seller in the orange segment and, sprite one of the most widely sold lemon-lime drinks in Nigeria.

WORLD MARKETING AND ADVERTISING

Pemberton had an appreciation of advertising. The point of sale sign – "Drink Coca-cola 5C" was fastened to the awning of Jacob's Pharmacy 1886. The first newspaper advertisement appeared in the Atlanta Journal, not only promising refreshment, but also informing the public where to get it.

One problem Pemberton had though been getting people to try the product. Sure if they could be enticed to sample it, they would become regular customers, he formalized this idea by means of a free ticket; good for one drink and redeemable at soda fountain. Since the slogan "Delicious and Refreshing" was first introduced in 1886, coca-cola advertising has continued to reflect the refreshing and delightful things in life. Throughout the years, slogans have created lasting impressions,

1929 – "The pause that refreshes"

1936 – "It's the refreshing thing to do"

1959 – "Be really refreshed"

1963 – "Things go better with coke"

"It's the Real Thing" first introduced in 1942 and was re-introduced in 1969.

1976 – "Coke adds life"

1982 – "Have a coke and a smile" took the market by storm followed by "Coke is it"

1989 – You can't Beat the Feeling etc.

NBC'S CONTRIBUTION TO THE NIGERIAN ECONOMY

The success of coca-cola in Nigeria has brought with it the development of other industries such as Delta Glass Company, Ugheli which supplies the millions of bottles required to keep a large bottling company in operation, crown products factories, Ijebu-Ode and Kano which manufacture the metal crowns to seal the bottles, Benin Plastics company which makes the plastic crates for carrying the bottles. Altogether NBC employee are over 6,000 Nigerians in all fields of operation and has about 150,000 customers (dealers) all over the Federation.

NBC is also the largest manufacturer of carbon-dioxide (CO_2) used to carbonate coca-cola.

OBJECTIVES OF THE STUDY

The objectives of this study are as follows:

1. To build a regression equation for predicting sales given the production levels.
2. To determine the relationship between sales and production levels.
3. To measure the degree of relationship between sales and quantity produced.

SCOPE AND COVERAGE

The scope of this study is limited to Nigerian Bottling Company Plc. Ilorin plant, Kwara State.

RESEARCH METHODOLOGY

The Data

The data used is a secondary data, and was obtained from Nigerian Bottling Co. Plc, Ilorin plant, Kwara State. The production data was collected from the warehouse in the production department and the sales data from their sales department. The data collected was between 2009 and 2014, a period of regular and consistent production.

The data collected include the amount of soft drink consumption during the dry season (tagged Y_1) the amount during the raining season (Y_2) and the quantity of soft drink produced during the dry season (X_3) those produced during the raining season (X_4). Measurements are in bottles and the period of dry season is taken to be between October – March and the raining season is period between April – September each year.

Multivariate Regression

Generally, it is assumed that if x takes values Xo then Y has a multivariate normal distribution with mean βX_0 and dispersion matrix $\sum$. Thus the rows of Y are assumed to be independent observations from such normal distributions, and their means are determined by the corresponding rows of X. Hence the model for the observed data can be written in the form.

$Y = X\beta + E$. -- *

Where y is the nxp matrix whose ith row is yi,

X is the n x (q + 1) matrix whose ith row is X_i,

B is the (q + 1) x p matrix of unknown parameters and

E is an nxp matrix of random variables whose rows are independent observation from a multivariate normal distribution with means zero and dispersion matrix ∑. Montgomery, D.C and Peak E.A. (1982) said that assumption of multivariate normality is only necessary if hypotheses are to be tested or confidence region are to be constructed.

First let us relate the model * to the familiar multiple regression model for a single dependent variable. Then the multiple regression model for this would be

$Y_1 = \beta_{o1} + \beta_{11} X_{i1} + \beta_{21} X_{i2} + e_{i1}$ (i = 1, ----, n).

Where

Y i1 is the ith observation on y1,

X i1 is the ith observation on X_1 and so on

eil is the deviation (departure, or error) of the ith observation yil from its mean.

$B_{01} + \beta_{11} X_{i1} + \beta_{21} X_{i2}$

In general, linear model from this set of n equations can be written compactly as

$Y^{(1)} + X\beta_1 + e_i$

Where y $^{(1)}$ is the nx1 vector of values of y_1

X is n x (q + 1) matrix of explanatory variable value

β_1 is the (q + 1) x 1 vector of parameter values

e_1 is the n x 1 vector of random variables whose elements are independent observations.

From a univariate normal distribution with mean 0 and variance δ^2

It could be written a $y^{(j)} = X^{\beta}{}_j + e_j$ -----------**

One final simplification can be made in analogous fashion to meet in multiple regression analysis, and this involves a small modification to the parameters of the model. The multiple regression model (**) is the matrix statement of

$Y_{ij} = \beta_{oj} + \beta_{1j}Xil + \beta_{2j} X_{i2} + \text{-------------} + \beta_{qj} X_{iq} + e_{ij}$ (i = 1 ---n)

For the regression of the jth dependent variable on the explanatory variables. Fisher, R. A. (1938) said, it is often more convenient computationally to reparameterize this equation, by measuring each of the explanatory variable about its means (i.e. by mean centering of all columns of the X matrix except the first). Thus the ith individual's value on the kth explanatory variable becomes $X_{ik} – \bar{X}_k$ where $\bar{X}_k$ is the sample means of this variable. The multiple regression model then becomes.

$Y_{ij} = \beta^*_{oj} + \beta_{1j} (X_{i1} - X_1) + \beta_{2j}(X_{i2} - X_2) + \text{--------} + \beta_{qj} (X_{iq} - X_q) + e_{ij}$ (i = 1, --- n) for (k=1,2,3---------q)

Where the coefficient β_{ij}, β_{2j} ------------β_{qj} are as before but,

$B^*_{oj} = \beta_o j + \beta_{1j}\bar{X}_i + \beta_{2j}, \bar{x}_2 + \text{-----------} \beta_{qj}\bar{x}_q$

One implication, of this reparametization is that the first column of X is orthogonal to the other columns, which not only has computational benefits but which also ensures that the estimate of β_{oj} is uncorrelated with the estimates of the other β_{kj} (K=1, ----, q). Since the same X is used in all the separate regression ** for j = I, ----, p, it follows that mean centering X also has corresponding benefits in the multivariate regression model.

One obvious practical objective of multivariate regression according to Anderson T. W. (1984) is to develop a predictive model for sales from the knowledge of quantity produced. The correlation matrix shows this to be a sensible objective, as each sale has suitably large correlations with each quantity produced (suggesting that quantity produced should be reasonably effective explanatory variables).

Accordingly we designate sales at dry season and sales during rainy season by y1 and y2 respectively, while the quantity produced at these seasons are denoted by X_3 and X_4 respectively. We consider the prediction of y^1 = (y1,y2). From $X^1 = (1,X_3,X_4)$ and we write y_i, X_i for the ith individual values of y, x respectively (I = 1, ----, n). The data matrix y is given by the first two columns of appendix 1.

Partial Correlation

Morrison, D. F (1976) described the two ways of calculating a partial correlation. In his work on multivariate statistical method, he argue that since correlation is a measure of linear dependence then removal of the effect of some variables on the correlation between others must infact be the removal of linear effect. Hence, the partial correlation between X_i and X_j on fixing the variables in X_2 can be obtained equivalently by calculating the correlation between the two sets of residuals from the regression of X_i and X_j respectively on variables in X_2. Alternatively, the partial correlation can be built up using a sequence of recurrence relationship denoted by $r_{ij.k}$ the (first order) partial correlation between X_i and X_j fixing X_k; by $r_{ij.kl}$ the (second order), partial correlation between X_i and X_j on fixing both X_k and X_L and so on.

$r_{ij.k} = (r_{ij} - r_{ik} \text{ X } r_{jk}) / \sqrt{(1 - r^2_{ik}) (1-r^2_{jk})}$ and

$r_{ij.kl} = (r_{ij.k} - r_{il.k} \text{ x } r_{j1.k}) / \sqrt{(1 - r^2_{il.k}) (1-r^2_{jl.k})}$

This pattern is continued for higher order partial correlations, and enables a partial correlation of qth order to be obtained from those of (q-1) th order.

ANALYSIS

Test of hypothesis can be described as the procedure used for deciding on an appropriate line of action in accordance with a decision criterion derived from the evaluation of sample results. In many cases a statistical model such as $Y_i = \beta_0 + \beta_1 X_{il} + \beta_2 X_{i3}$ ----- $\beta_n X_{in}$ is formulated for the purpose of describing how variables are related (and to either reject or accept any hypothesis). Here we observed Y as the response variable first (in this case sales). B_o, β_1 --- β_3 are parameter with unknown values X_1 and X_2 are information contributing variables that are measured without error. In this case X_1 and X_2 represent the quantity produced during the dry season and the raining seasons respectively.

$$R = \begin{pmatrix} 1 & 0.6 & 0.93 & 0.7 \\ 0.6 & 1 & 0.73 & 0.84 \\ 0.93 & 0.73 & 1 & 0.79 \\ 0.7 & 0.84 & 0.79 & 1 \end{pmatrix}$$

Fitting the Regression Model

The first stage in a multivariate regression analysis is the fitting of model to the observed data. This requires the estimation of the unknown parameters β and ∑ which can be done by maximum livelihood if normally is assumed for the error matrix E or by least square if no distribution assumptions are made.

$B = (x^1 x)^{-1} x^1 y$

$$X^{1}X = \begin{pmatrix} 30 & 11052376 & 10857212 \\ 11052376 & 4.176 \times 10^{12} & 4.08 \times 10^{12} \\ 10857212 & 4.08 \times 10^{12} & 4.028 \times 10^{12} \end{pmatrix}$$

$$X^{1}y = \begin{pmatrix} 10758335 & 10493415 \\ 4.05 \times 10^{12} & 3.95 \times 10^{12} \\ 3.96 \times 10^{12} & 3.89 \times 10^{12} \end{pmatrix}$$

$$X^{1}X = \begin{pmatrix} 0.0904 & -1.15 \times 10^{-7} & -1.27 \times 10^{-7} \\ -1.15 \times 10^{-7} & 1.53 \times 10^{-12} & -1.24 \times 10^{-12} \\ -1.27 \times 10^{-7} & -1.24 \times 10^{-12} & 1.6 \times 10^{-12} \end{pmatrix}$$

$$\beta = (x^1x)^{-1} \quad (x^1 y)$$

$$\beta = \begin{pmatrix} 3883.8 & 324.72 \\ 0.489 & 0.0132 \\ -0.052 & -0.0067 \end{pmatrix}$$

Thus the appropriate regression equation.

$Y_1 = 3883.5 + 0.489 (x_3 – 368413) – 0.052 (x_4 – 361907) + e_1$

$Y_2 = 324.72 + 0.0132 (x_3 – 368413) – 0.0067 (x_4 – 361907) + e_2$

Testing for the Usefulness of the Mode

To test for the usefulness of the model we use the multiple regression equivalent of r2, the coefficient of determination. Thus we define the sample multiple coefficient of determination.

$R^2r = \sum (y_i – y)^2 / \sum (yi – y)^2$ = explained variability / total variability.

$R^2r = 2.5206E11/2.83271E12$

$= 0.8898$

This implies that 89% of the variation of the sales during the dry season (Y_1) about their mean can be explained by the model.

Similarly for Y_2

$Rr^2 = 3.46702 \times 10^{12} / 4.8032 \times 10^{12}$

$= 0.7218$

This which implies that 72% of the variation of the sales during the raining season (Y_2) about their mean can be explained by the model.

Hypothesis Testing

The test that all population canonical correlations are zero is a test of no association between the two sets of variables Y_1 and Y_2, which in turn is equivalent to the test, that $\sum_{12} = 0$. If the sample and population covariance matrixes are partitioned in the usual way containing p_1 and p_2 elements respectively and we wish to test whether the variates in Y_1 are independent of those in

Y_2. The sample statistics Y and S can be partitioned and the hypothesis of interest thus becomes.

Ho : $\sum_{12} = 0$ Vs H_1: $\sum_{12} \neq 0$

However, to make the connection with canonical correlation this test statistics must be re-written denoting the canonical correlations by R_1 R_2, ------- R_2 (where S = min (p, q) for p, q, elements in Y_1 and Y_2 respectively. Then the likelihood ratio test statistics ^ is given by $-2\log_e \wedge = - n \sum \log_e (1-R^2_i)$

Test That All Population Correlation are Zero Mutual Independence of All Variables

Here we wish to test the null hypothesis H_o: p = 0 Vs H_1 : p ≠ 0

The likelihood ratio test statistic λ for this situation may be obtained from a generalization of

$-2\log_e \lambda - n\log_e / R /$

Since Ho specified that all off- diagonal elements of P are zero, there are ½ p(p-1) constraints in the statements of Ho. Thus if Ho is true $-2\log_e\lambda$ of above has an asymptotic X_2 distribution on ½p(p-1) degrees of freedom.

Box (1949) showed that the X^2 approximation is improved if n is replaced by n^1.

n1 = n = 1/6 (2P + 11)

Here in our data

P=4 n=3 0 so that ½p(p-1) = 6

$n^1 = 26.83$

According to Krzanowski (1988), the determinant of a matrix equals the products of its eigen values so

/ R / = (0.89) X 0.32X0.62X0.1208

= 0.02133

Such that $-n\log_e / R / = - 30\log_e^{0.02133}$

= 115.429

The 95% point of the $X^2_6 = 12.592$

So the evidence against the null hypothesis of mutual independence is over whelming.

Partial Correlation

Since it has been established that there is association between variables, it is also of interest to find out the extent of their interdependency through the use of partial correlation analysis. To do this, the 2^{nd} order partial correlation is employed using the first order partial correlation.

$R_{ij\cdot kl} = R_{ij.k} - R_{il.k} (R_{ji.k}) / \sqrt{} (1 - R^2_{il.k})(1-R^2_{jl.k})$

$R_{13.4} = 0.861$ $R_{12.3} = 0.3141$

$R_{12.4} = 0.861$ $R_{14.3} = -0.154$

$R_{23.4} = 0.2$ $R_{24.3} = 0.628$

$R_{13.2} = 0.8998$

$R_{14.2} = 0.439$

$R_{34.2} = 0.477$

$R_{12.34} = R_{12.3} – (R_{14.3}) (R_{24.3}) / \sqrt{} (1-R^2_{14.3}) (1-R^2_{24.3})$

$= 0.3141 – (0.154) (0.628) / \sqrt{} 1 – (0.154)^2\ 1 – (0.628)^2$

$= -0.217388/0.768929912 = -0.283$

$R_{13.34} = R_{13.2} – (R_{14.2}) (R_{34.2}) / \sqrt{} (1-R^2_{14.2}) (1-R^2_{34.2})$

$= 0.0097954/0.383467793 = 0.0255$

$R_{23.14} = R_{23.4} – (R_{13.4}) (R_{12.4}) / \sqrt{} (1-R^2_{13.4}) (1-R^2_{12.4})$

$= 0.189668/0.508568333$

$= 0.373$

$R_{13.24} = R_{13.4} – (R_{23.4}) (R_{12.4}) / \sqrt{} (1-R^2_{23.4}) (1-R^2_{12.4})$

$= 0.8586/0.979725349 = 0.876$

$R_{24.13} = R_{24.3} – (R_{12.3}) (R_{14.3}) / \sqrt{} (1-R^2_{12.3}) (1-R^2_{14.3})$

$= 0.5796286/0.938064487$

$= 0.6178$

$R_{34.12} = R_{34.2} – (R_{13.2}) (R_{14.2}) / \sqrt{} (1-R^2_{13.2}) (1-R^2_{14.2})$

$= 0.08198/0.392 = 0.209$

SUMMARY

As discussed in introduction multivariate regression analysis helps to predict values of variables from the value of other variables, on the basis of assumed nature of the association or relationship between the variables. It was also discovered the regression equations/models for predicting future sales as follows:-

$Y_1 = 388.5 + 0.0489 (x_3 – 368413) – 0.052 (x_4 – 361907) + e_1$ and

$Y_2 = 324.72 + 0.0132 (x_3 – 368413) – 0.0067 (x_4 – 361907) + e_2$

It was discovered that Y_1 could explain about 89% of the variation of the sales during the dry season about their mean. This which means that Y_1 is 89% accurate in all cases, and that Y_2 also could explain 72% of the variation of the sales during the raining season about their mean.

Similarly, confidence interval was employed to determine the confident correlation coefficient and the following assertions were made:

(i) That we are 95% confident, that correlation coefficient between sales during the dry season and the raining season is between -0.58 and 0.08. This is a high negative correlation, and it suggest that the sales of soft drinks during these two season does not follow similar pattern

(ii) That at 95% confidence the correlation coefficient between sales during the dry season and quantity produced during the same period is between 0.75 and 0.9388. This which implies a very high positive correlation and it suggest that the sales of soft drink during the dry season and the quantity produced during the same period are related.

CONCLUSION

From above tests and measurement it was observed that there exist relationships between sales and quantity produced it also revealed that association between sales during the dry season and the quantity produced at the same period appears to be significantly higher than any of the other variables under consideration. This may be as a result of market target at such period when demand for the products tends to be higher.

REFERENCES

Anderson, T. W. (1984) An Introduction to Multivariate Statistical Analysis, John Wiley, New York.

Birkhead, M. (2007) Multivariate Modeling Cambridge University Press. 53 – 110.

Fisher, R. A. (1938) The Statistical Utilization of Multiple Measurement, Arm Eugen.

Galton, F. (1908) Introduction to Multivariate Analysis.

Krzanowski, W. J. (1988) Principles of Multivariate Analysis (A User's Perspective) Oxford Science Publisher.

Montgomery, D. C. and Peak, E. A. (1982) Introduction to Linear Regression Analysis, John Wiley, New York.

Morrison, D. F. (1976) Multivariate Statistical Methods, Second edition Mc Graw-Hill, New York.

Nigeria Bottling Company Plc Periodicals (May 2000).