BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

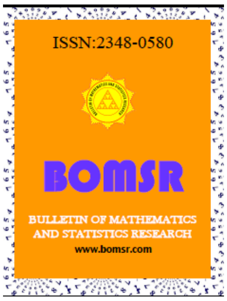
http://www.bomsr.com
Email:editorbomsr@gmail.com

ISSN:2348-0580

RESEARCH ARTICLE

ANALYSIS OF CONSUMER MARKETS – REGRESSION ANALYSIS APPROACH

KUMARASWAMY KANDUKURI
Research Scholar
Department of Statistics
Osmania University, Hyderabad



ISSN:2348-0580

BOMSR

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

www.bomsr.com

ABSTRACT

The aim of marketing is to meet and satisfy target customer’s needs and wants better than competitors. Consumer behavior is the study of how individuals, groups and organizations select, buy, use and dispose of goods, services, ideas or experiences to satisfy their needs and wants. One can analyze the relationship between the variables (factors) of consumer behavior and willingness to purchase. Regression analysis is a statistical technique for estimating the relationship among variables which have reason and result relation. This is proposed to find the relationship between the consumer behavior factors to the consumer willingness to buy a product. The coefficient of determination (**R^2**) is of the variation that occurs in the willingness to buy is described by all of the independent factors.

Keywords: Consumer behavior, Behavioral factors, Regression analysis and Coefficient of determination.

I. INTRODUCTION

Marketers must fully understand both the theory and reality of consumer behavior. Studying consumer’s provides clues for improving (or) introducing products (or) services, setting prices, devising channels, crafting messages and developing other marketing activities. Marketers are always looking for emerging trends that suggest new marketing opportunities.

Successful marketing requires that companies fully connect with their customers. Adopting a holistic marketing orientation means understanding consumers – gaining a 360–degree view of both, their daily lives and the changes that occur during their lifetimes. The researcher is invested to know deeper regarding their strategy of surviving and developing, especially in the strategy of promoting their products and services to the market which leads to customer motivation to come and shop. Gaining a thorough, in-depth consumer understanding helps to ensure that the right products are marketed to the right consumers in the right way.

II. Model Framework

Regression analysis used in the research where the dependent variables (factors) are hypothesized to influence the dependent variable (factor). For predictions, Multiple Regression, let us use more than one factor to make a prediction, while for explanation, multiple regression let us separate causal factors, analyzing each other's influence the factors, in this case, the influence of consumer behavior to willingness to buy (Samuel L. Baker: 2006).

Model development is a complex process requiring a broad range of inputs. The process generally includes the following components.

1. Inter disciplinary teams that combine substantial knowledge of the process with an understanding of analysis tools such as Multiple Regression that are needed for model development.
2. Ideal data for regression models have independent factors with a wide range and small correlations between independent factors.
3. Theory, experience, and historical knowledge are the foundation for analysis and problem solving.
4. An understanding based on experience that data contain both information and error. Analysis and modeling seek to enhance information and minimize error.

III. Multiple Regression Model

Let us consider the willingness to buy (Y) as a dependent (regress) factor and factors like Amount of Purchase on product (X_1), Discount rate (X_2), Age (X_3) and Promotion on the product (X_4) are independent (predictor) factors.

Therefore the model can be defined as some functions of X's,

i.e., $y = f(x) + \varepsilon$

More clearly, Multiple Regression is a procedure for determining the simultaneous linear effect of k-independent factors on a dependent factor. This is done by estimating the coefficients of the linear equations using the principle of least squares.

$$\mathbf{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$$

IV. Estimation of Parameters/Regression Coefficients

Let each of the predictor factors $\mathbf{X_1, X_2... X_K}$ have n levels. Then X_{ij} represents the i^{th} level of j^{th} predictor factor X_j. Observations, $Y_1, Y_2... Y_n$, recorded for each of these n-levels can be expressed in the following way.

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + ... + \beta_k X_{1k} + \varepsilon_1$$
$$Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + ... + \beta_k X_{2k} + \varepsilon_2$$
$$..$$
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i$$
$$...$$
$$Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + ... + \beta_k X_{nk} + \varepsilon_n$$

These systems of n - equations represented in matrix notation as follows,

Y = Xβ + ε i.e., $y_i = f(x_{ij}) + \varepsilon_i$; i = 1 to n, j = 1 to k

Where, $\mathbf{Y}=\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ Y_i \\ . \\ Y_n \end{bmatrix}$, $\mathbf{X}=\begin{bmatrix} 1 & X_{11} & X_{12} & .. & X_{1k} \\ 1 & X_{21} & X_{22} & .. & X_{2k} \\ .. & .. & .. & .. & .. \\ 1 & X_{I1} & X_{I2} & .. & X_{Ik} \\ .. & .. & .. & .. & .. \\ 1 & X_{n1} & X_{n2} & .. & X_{nk} \end{bmatrix}$, $\beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_k \end{bmatrix}$ and $\varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_k \end{bmatrix}$

The matrix **X** is referred to as the design matrix. It contains information about the levels of the predictor variables at the observations are obtained. The vector **β** contains all the regression coefficients. To obtain the regression model, **β** should be known. **β** is estimated using least squares principle.

Residuals and Least Squares Criterion

If $\overset{\wedge}{\beta}$ is a (k+1x1) vector of elements of**β**, then the estimated model may be written as

$$\mathbf{Y}=\mathbf{X}\overset{\wedge}{\beta}+\varepsilon$$

$$\Rightarrow \varepsilon=\mathbf{Y}-\mathbf{X}\overset{\wedge}{\beta}$$

To determine the least squares estimation, we write the sum of squares of the residuals (as a functions of **β**) as:

$$\mathbf{S}(\overset{\wedge}{\beta})=\sum_{I=1}^{N} e_i^2=\mathbf{e'e}=(\mathbf{Y}-\mathbf{X}\overset{\wedge}{\beta})^{'}(\mathbf{Y}-\mathbf{X}\overset{\wedge}{\beta})$$

$$=\mathbf{Y'Y}-\mathbf{Y'X}\overset{\wedge}{\beta}-\overset{\wedge}{\beta}\mathbf{X'Y}+\overset{\wedge}{\beta}'\mathbf{X'X}\overset{\wedge}{\beta}$$

The minimum of $\mathbf{S}(\overset{\wedge}{\beta})$ is obtained by setting the first derivative of equal to zero.

We obtain, $$\frac{\partial \mathbf{S}}{\partial \overset{\wedge}{\beta}}=-2\mathbf{X'Y}+2\mathbf{X'X}\overset{\wedge}{\beta}$$

$$\Rightarrow -2\mathbf{X'Y}+2\mathbf{X'X}\overset{\wedge}{\beta}=0$$

Which gives normal equations, $\mathbf{X'X}\overset{\wedge}{\beta}=\mathbf{X'Y}$

Solving this for $\overset{\wedge}{\beta}$, we have $\overset{\wedge}{\beta}$ = **(X'X)$^{-1}$X'Y**, provided the inverse of **(X'X)** exist.

And $E(Y/X_1X_2...X_k)=\mathbf{X}\beta; \sin ceE(\varepsilon)=0$

$\therefore$The fitted model can be written as

$$\overset{\wedge}{\mathbf{Y}}=\mathbf{X}\overset{\wedge}{\beta}$$

$$\Rightarrow \overset{\wedge}{\mathbf{Y}}=X(X'X)^{-1}X'Y$$

$$\Rightarrow \overset{\wedge}{\mathbf{Y}}=\overset{\wedge}{\beta_0}+\overset{\wedge}{\beta_1}X_1+\overset{\wedge}{\beta_2}X_2+..........+\overset{\wedge}{\beta_k}X_k$$

V. Properties/Axioms

1. The least square estimator is unbiased

$$E(\hat{\beta}) = E[\beta + (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon]$$

$$\Rightarrow E(\hat{\beta}) = \beta + (\mathbf{X'X})^{-1}\mathbf{X'}E(\varepsilon)$$

$$\Rightarrow E(\hat{\beta}) = \beta, \; provided \varepsilon \approx N(0, \sigma^2\mathbf{I})$$

2. Variance – Covariance matrix of $\hat{\beta}$:

$$C = V(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta})(\hat{\beta} - E(\hat{\beta})^{'}]$$

$$\Rightarrow C = V(\hat{\beta}) = \sigma^2(\mathbf{X'X})^{-1}$$

Which is symmetric and positive definite and

$$V(\hat{\beta}_j) = \sigma^2((\mathbf{X'X})^{-1})_{ij}$$

$$COV(\hat{\beta}_j, \hat{\beta}_h) = \sigma^2((\mathbf{X'X})^{-1})_{jh}$$

3. The positive square root of the C_{ij} represents the estimated standard deviation of the j^{th} regression coefficient $\hat{\beta}_j$, and is called the estimated standard error of β_j

$$Se(\beta_j) = \sqrt{C_{jj}}$$

4. If $\hat{\beta}$ is the Ordinary Least Square estimator of β in the classical linear regression model (**Y; Xβ, σ^{2}I**), and if β^* is any other linear unbiased estimator of β, then $V(\gamma\beta^*) \geq V(\gamma\hat{\beta})$, where γ is any constant vector of the appropriate order. **(Gauss-Markov Theorem)**

5. $\hat{\beta}$ is best linear unbiased estimator (BLUE) for β.

VI. Hypothesis Test

The validity of the fitted regression can be checked by hypothesis test of multiple linear regressions.

1. Test for significance of Regression: This test is used to check if a linear statistical relationship exists between the response factor and at least one of the predictor factors, carried out by analysis of variance.

The Null Hypothesis (H_0): $\beta_1 = \beta_2 = \beta_3 = \ldots\ldots\ldots\ldots = \beta_k = 0$

Alternative Hypothesis (H_1): $\beta_j \neq 0$, for at least one j

Analysis of variance is used in multiple regressions to determine

i) The amount of explained variability, RSS, and the amount of unexplained variability, ESS.

ii) An estimate of error variance.

iii) The components of variability explained by each of the independent factors give the variability by the other factor.

$$TSS = RSS + ESS$$

$$Where$$

$$ESS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$RSS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

The statistic $F_0 = \frac{MSR}{MSE} = \frac{(\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 / k)}{(\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 / [n-k-1])} \sim F_{\alpha,k,n-k-1}$

Reject H_0, if $F_0 > F_{\alpha,k,n-k-1}$

Remark: The error mean square is an estimate of the variance σ^2, of the random error term ε_i,

$$\hat{\sigma}^2 = S^2_{Y/X} = \frac{ESS}{n-k-1}$$

And $\hat{\sigma} = S_{Y/X} = \sqrt{\frac{ESS}{n-k-1}}$ is the standard error of estimate. This is the variance in **Y** after the effect of all the independent factors has been removed.

2. Tests on individual Regression coefficients: The 't' test is used to check the significance of individual regression coefficients in the multiple linear regression model. Adding a significant factor to a regression model makes the model more effective, while and adding an unimportant factor may take the model worse. The hypothesis statements to tests the significance of particular regression coefficients β_j :

$$H_0 : \beta_j = 0 \text{ V/s } H_1 : \beta_j \neq 0$$

The test statistic for this test is based on the t – distribution,

$$t_0 = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)} \sim t_{\alpha/2,(n-k-1)}$$

Reject H$_0$, if absolute value of t_0 is greater than $t_{\alpha/2,(n-k-1)}$, and we conclude that the factor X_j has a significant effect and should be included in the multiple regression model.

Remarks

(i) A 100(1-α)% confidence interval for the slope coefficient follow the standard form

$$\hat{\beta}_j \pm t_{\alpha/2,(n-k-1)} \sqrt{C_{jj}} \text{ (Or) } \hat{\beta}_j \pm t_{\alpha/2,(n-k-1)} Se(\hat{\beta}_j)$$

(ii) Confidence interval for fitted values $(\hat{Y}_i)$

$$\hat{Y}_i \pm t_{\alpha/2,(n-k-1)} \sqrt{\hat{\sigma}^2 X_i^{'} (X'X)^{-1} X_i} \text{ Where } X_i = (1 \quad X_{i1} \quad X_{i2} \quad .. \quad .. \quad X_{ik})^{'}$$

VII. Discussion

Measure of model accuracy: The coefficient of multiple determination $(\mathbf{R}^2)$ is similar to the coefficient of determination used in the case of simple linear regression, and is defined as the ratio of total explained variability divided by total variability.

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

Using $\mathbf{R}^2$ to compare models with the same set of observed **Y'**s is always useful because a higher $\mathbf{R}^2$ implies a smaller ESS.

A number of analysts prefer to use $R^2{}_a$ adjusted for degrees of freedom to compare equations with different independent factors. $R^2{}_a$ is defined as

$$R^2{}_a = 1 - \frac{S^2{}_{Y/X}}{S^2{}_Y} = 1 - \frac{MSE}{MST} = 1 - \frac{\left(\frac{ESS}{n-k-1}\right)}{\left(\frac{TSS}{n-1}\right)}$$

$$\Rightarrow R^2{}_a = 1 - \frac{ESS(n-1)}{TSS(n-k-1)}$$

$$\Rightarrow R^2{}_a = 1 - (1 - R^2)\frac{(n-1)}{(n-k-1)}$$

$R^2{}_a$ is thus adjusted for degrees of freedom and will be slightly smaller than R^2. An important advantage is that $R^2{}_a$ will increase only if $S^2{}_{Y/X}$ decreases. In contrast, R^2 increases whenever ESS decreases and ESS decreases even when non-significant factors are added. For that reason $R^2{}_a$ is more useful statistic for comparing regression models. The difference between $R^2{}_a$ and R^2 is small and they approach each other as the sample size increases. Thus the general discussion of R^2 applies to $R^2{}_a$.

VIII. Results and Conclusions

The data was collected from the customers who visited 12 shopping malls during Aashadam Season Sale in Hyderabad, Telangana. A simple questionnaire is handover to the customers and their responses were received. The Avg. Sales (**Y** in k.g's) of each shop is influenced by the mean of four factors, Amount spent on Purchase (X_1 in 00's), Discount rate (X_2 in percentage), Age of the customers (X_3) and amount spent on promotion activity (X_4 in 000's) by the shop.

Table 1: Pearson's Correlations

	Avg Sales	Purchase mount	Discount	Age	Promotion
Avg Sales	1.000	.697	.753	.594	.793
Purchase	.697	1.000	.907	.809	.423
amount	.753	.907	1.000	.688	.371
Discount	.594	.809	.688	1.000	.504
Age	.793	.423	.371	.504	1.000
Promotion					

From the table we conclude that there exist a good amount of association between the dependent factor and independent factors. The most of the customers attracted by the promotional tool (r_{YX4}=0.793) and Discount given on sales (r_{YX2}=.753).

Table 2. Result of Coefficient Correlation of Multiple determination Test (R^2)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.942*	.887	.822	.5213

According to Table 2 which describes the criteria for the correlation between independent factors and dependent factor, it can be interpreted as independent factors have a highly correlation with decision for 94.2% and from the coefficient of determination (R^2) is equal to 0.822 that showed 82.2% of the variation that occurs in the willingness to buy is described by all of the independent factors, cultural factors, social factors, personal factors and psychological factors. While the remaining 17.8% are explained by the other factors which is not described in the research.

Table 3. Regression Analysis

Coefficients[a]

	Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1(Constant)	.016	1.855		.009	.993	-4.370	4.402
purchaseamount	-.023	.084	-.106	-.279	.789	-.222	.175
discount	8.126	3.602	.696	2.256	.059	-.391	16.643
age	-.040	.075	-.123	-.532	.611	-.217	.137
promotion	.710	.163	.642	4.354	.003	.324	1.095

a. Dependent Variable: Avg. Sales

From the above table, the regression equation is made as follows:

$Y = 0.016 - 0.023X_1 + 8.126X_2 - 0.04X_3 + 0.71X_4$ And estimated regression equation is $\hat{\mathbf{Y}} = -0.106X_1 + 0.696X_2 - 0.123X_3 + 0.642X_4$ and tabulated t value for two tailed t-test is 1.80 for 11 degrees of freedom. The t test, partially determines the influence of independent of factors towards the dependent factor i.e., Discount rate and Promotion factors are significant on dependent factor willingness to buy (Y) and remaining factors are not significant.

Confidence interval for regression coefficients:

1. C.I for β_0 : (-4.370, 4.402) 2. C.I for β_1 : (-0.222, 0.175) 3. C.I for β_2 : (-0.391, 16.643)

4. C.I for β_3 : (-0.217, 0.137) 5. C.I for β_4 : (0.324, 1.095)

Analysis of Variance

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14.875	4	3.719	13.688	.002[a]
	Residual	1.902	7	0.272		
	Total	16.777	11			

a. Predictors: (Constant), promotion, discount, age, purchase amount

b. Dependent Variable: Avg. Sales

From ANOVA test or F test table above, it is shown that the value of F count is 13.688 with probability 0.002. Since the F count is 13.688 > F table $(F_{0.05,4,7} = 4.12)$

We observe that the independent factors are highly significant with dependent factor then H_1 is accepted.

References

[1]. Gianie Abdu, Purwanto (2013), Analysis of consumer behavior affecting consumer willingness to buy in 7-Eleven convenience store, Universal journal of Management 1 (2): 69-75.

[2]. Baker, S.L. (2006), Multiple Regression Theory.

[3]. Philip Kotler and Kewin Lane Keller (2006), Marketing Management, 12th edition, New Delhi; Prentice-Hall of India Pvt Ltd.

[4]. M.G. Abott, ECON 351* -- Note 4: Statistical Properties of OLS Estimators.

[5]. Norman R. Draper and Harry Smith: Applied Regression Analysis, Wiley Series.