Applied Science Periodical ISSN 0972-5504
Volume - XXV, No. 1, February 2023
Refereed and Peer-Reviewed Quarterly Periodical
Journal website : www.internationaljournalsiwan.com

A Case Study on Sales of Different Products of a Superstore

by **Gayatri Behere,** *Associate Professor,*
Department of Statistics,
Institute of Science, Nagpur - 440001, India
Sagar Jangale and Krishna Gupta, *P.G. Students,*
Department of Statistics,
Institute of Science, Nagpur - 440001, India

(Received: January 15, 2023; Accepted: February 20, 2023;
Published Online: February 28, 2023)

Abstract :

This study is done to predict the sales of products in superstore, for the companies to help them in planning future demand for the products in particular areas, cities, regions, etc.

In this case, superstore data of US has been used for the study to understand the customers and their needs.

Certain visualizations are done in order to understand the behaviors of the sales data according to the situation and customer's interest. Data patterns and trends are observed to draw the conclusions on the sales. As the major motto of a retailer is to make profits by selling the product, there is a need for him to understand the data variations with the change in time, climate, regions and customer's interest. Thus to make his work easier, he will use the resulted visualizations formed out of the sales data. These visualizations help him better understand the change in sales which he can use in order to control the inventory in the superstore and earn profits by reducing loses.

Hence this paper provides efficient ways of analyzing the sales data of a superstore, finding the reasons for the increase and decrease in the sales, controlling product imports and attaining a profitable business.

Introduction :

Regression analysis is a predictive modeling technique which estimates the relationship between two or more variables. Regression analysis focuses on the relationship between a dependent (target) variable and independent variable(s) (predictors). Here, dependent variable is assumed to be the effect of the independent variable(s). The value of predictors is used to estimate or predict the likely-value of the target variable.

To do this, we first try to assume a mathematical relationship between the target and the predictor(s). The relationship can be a straight line (linear regression) or a polynomial curve (polynomial regression) or a non-linear relationship (non-linear regression). This can be done through various ways. The simplest and most popular way is to create a scatter plot of the target variable and predictor variable.

Once the type of relationship is established, we try to find the most-likely values of the coefficients in the mathematical formula.

Regression analysis comprises of the entire process of identifying the target and predictors, finding the relationship, estimating the coefficients, finding the predicted values of target, and finally evaluating the accuracy of the fitted relationship.

Coefficient of Determination:

Now, we look at the *R*-squared value of the model, which is also called the "Coefficient of Determination". This statistic calculates the percentage of variation in target variable explained by the model. The below illustration captures the explained vs. unexplained variation in data

R-squared is calculated using the following formula:

$$R^2 = \frac{Explained\ Variance}{Total\ Variance} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

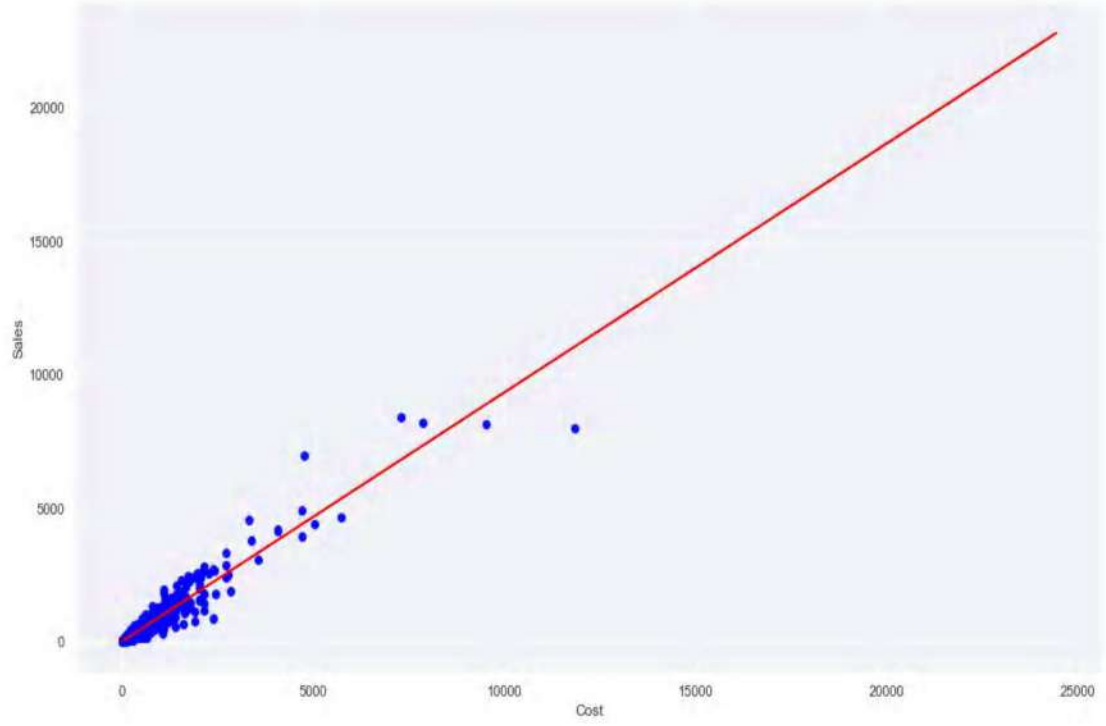
R-squared is always between 0 and 100%. As a guideline, the more the *R*-squared, the better is the model. The objective is not to maximize the *R*-squared, since the stability and applicability of the model are equally important. Next, check the adjusted *R*-squared value. Ideally, the *R*-squared and adjusted *R*-squared values need to be in close proximity of each other. If this is not the case, then the analyst may have over fitted the model and may need to remove the insignificant variables from the model.

Residual Analysis:

We can also evaluate a regression model based on various summary statistics on error or residuals. Root Mean Square Error (RMSE): Where we find average of squared residuals as per the given formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}$$

Simple Linear Regression Best Fitted Line:



Sample Dataset:

Sr. No.	Order Date	Ship Mode	Segment	Country	City	State	Region	Category	Sub-Category	Sales	Quantity	Discount	Profit	Profit/Loss
1	06-01-13	Second Class	Consumer	United States	Henderson	Kentucky	South	Furniture	Bookcases	261.96	2	0	41.9136	1
2	13-01-13	Second Class	Consumer	United States	Henderson	Kentucky	South	Furniture	Chairs	731.94	3	0	219.582	1
3	13-01-13	Second Class	Corporate	United States	Los Angeles	California	West	Office Supplies	Labels	14.62	2	0	6.8714	1
4	13-01-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	South	Furniture	Tables	957.5775	5	0.45	-383.031	0
5	18-01-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	South	Office Supplies	Storage	22.368	2	0.2	2.5164	1
6	19-01-13	Standard Class	Consumer	United States	Los Angeles	California	West	Furniture	Furnishings	48.86	7	0	14.1694	1
7	19-01-13	Standard Class	Consumer	United States	Los Angeles	California	West	Office Supplies	Art	7.28	4	0	1.9656	1
8	19-01-13	Standard Class	Consumer	United States	Los Angeles	California	West	Technology	Phones	907.152	6	0.2	90.7152	1
9	19-01-13	Standard Class	Consumer	United States	Los Angeles	California	West	Office Supplies	Binders	18.504	3	0.2	5.7825	1
10	20-01-13	Standard Class	Consumer	United States	Los Angeles	California	West	Office Supplies	Appliances	114.9	5	0	34.47	1
11	20-01-13	Standard Class	Consumer	United States	Los Angeles	California	West	Furniture	Tables	1706.184	9	0.2	85.3092	1
12	23-01-13	Standard Class	Consumer	United States	Los Angeles	California	West	Technology	Phones	911.424	4	0.2	68.3568	1
13	27-01-13	Standard Class	Consumer	United States	Concord	North Carolina	South	Office Supplies	Paper	15.552	3	0.2	5.4432	1
14	27-01-13	Standard Class	Consumer	United States	Seattle	Washington	West	Office Supplies	Binders	407.976	3	0.2	132.5922	1
15	27-01-13	Standard Class	Home Office	United States	Fort Worth	Texas	Central	Office Supplies	Appliances	68.81	5	0.8	-123.858	0
16	31-01-13	Standard Class	Home Office	United States	Fort Worth	Texas	Central	Office Supplies	Binders	2.544	3	0.8	-3.816	0
17	02-02-13	Standard Class	Consumer	United States	Madison	Wisconsin	Central	Office Supplies	Storage	665.88	6	0	13.3176	1
18	03-02-13	Second Class	Consumer	United States	West Jordan	Utah	West	Office Supplies	Storage	55.5	2	0	9.99	1
19	03-02-13	Second Class	Consumer	United States	San Francisco	California	West	Office Supplies	Art	8.56	2	0	2.4824	1
20	04-02-13	Second Class	Consumer	United States	San Francisco	California	West	Technology	Phones	213.48	3	0.2	16.011	1
21	04-02-13	Second Class	Consumer	United States	San Francisco	California	West	Office Supplies	Binders	22.72	4	0.2	7.384	1
22	04-02-13	Standard Class	Corporate	United States	Fremont	Nebraska	Central	Office Supplies	Art	19.46	7	0	5.0596	1
23	08-02-13	Standard Class	Corporate	United States	Fremont	Nebraska	Central	Office Supplies	Appliances	60.34	7	0	15.6884	1
24	12-02-13	Second Class	Consumer	United States	Philadelphia	Pennsylvania	East	Furniture	Chairs	71.372	2	0.3	-1.0196	0
25	14-02-13	Standard Class	Consumer	United States	Orem	Utah	West	Furniture	Tables	1044.63	3	0	240.2649	1
26	14-02-13	Second Class	Consumer	United States	Los Angeles	California	West	Office Supplies	Binders	11.648	2	0.2	4.2224	1
27	14-02-13	Second Class	Consumer	United States	Los Angeles	California	West	Technology	Accessories	90.57	3	0	11.7741	1
28	15-02-13	Standard Class	Consumer	United States	Philadelphia	Pennsylvania	East	Furniture	Bookcases	3083.43	7	0.5	-1665.0522	0
29	20-02-13	Standard Class	Consumer	United States	Philadelphia	Pennsylvania	East	Office Supplies	Binders	9.618	2	0.7	-7.0532	0
30	22-02-13	Standard Class	Consumer	United States	Philadelphia	Pennsylvania	East	Furniture	Furnishings	124.2	3	0.2	15.525	1

Simple Linear Regression:

- We have the Root Mean Square Error (RMSE) of 28.4.
- Variance score: 0.90.
- *R*-square: 0.89.
- Cost has a significant effect on the Sales of the Products.

Multiple Linear Regressions:

- Variance Score is found to be 1.00 in Sales for variables Category, Quantity, Cost and Discount etc.
- Root Mean Square Error is found to be 2.25.
- Adjusted *R*-square is 88%.
- Our study also shows that variables like Category, Sub-Category, Quantity, Cost and Discount has significant impact on Sales of the Products.

References:

1. Jump up to: a b "Grocery". Oxford Learner's Dictionary. Retrieved 13 July 2020.
2. "Grocery Store". Merriam-Webster Dictionary. Retrieved 13 July 2020.
3. Meyer, Zlati (5 April 2017): "Why 'Grocerants' are the new trend, taking bite out of restaurants". USA Today. Retrieved 6 April 2017. The phenomenon is growing fast enough both in prevalence and sophistication that the food industry has coined a name for these combination grocery stores and eateries - the 'grocerant'.
4. Vadini, Ettore (28 February 2018): Public Space and an Interdisciplinary Approach to Design. ISBN 9788868129958.
5. "Opening of the Astor market, New York City, 1915". Library of Congress. 1915.