

# Applied SPSS for Data Forecasting of Sale Quantity

Aung Cho, Aung Si Thu

University of Computer Studies, Maubin, Myanmar

**How to cite this paper:** Aung Cho | Aung Si Thu "Applied SPSS for Data Forecasting of Sale Quantity" Published in International

Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.694-697,

<https://doi.org/10.31142/ijtsrd26378>



IJTSRD26378

Copyright © 2019 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



The package is particularly useful for students and researchers in psychology, sociology, psychiatry, and other behavioral sciences, containing as it does an extensive range of both univariate and multivariate procedures much used in these disciplines. The text is not intended in any way to be an introduction to statistics and, indeed, we assume that most readers will have attended at least one statistics course and will be relatively familiar with concepts such as *linear regression, correlation, significance tests*, and *simple analysis of variance*. Our hope is that researchers and students with such a background will find this book a relatively self-contained means of using SPSS to analyze their data correctly.[2]

## 1.2. Regression Analyses

*Regression analysis* is about predicting the future (the unknown) based on data collected from the past (the known). Such an analysis determines a mathematical equation that can be used to figure out what will happen, within a certain range of probability:

- The analysis is performed based on a single dependent variable.
- The process takes into consideration the effect one or more dependent variables have on the dependent variable.
- It takes into account which independent variables have more effect than others.

Performing regression analysis is the process of looking for predictors and determining how well they predict a future outcome. When only one independent variable is taken into account, the procedure is called *simple regression*. If you use more than one independent variable, it's called a

## ABSTRACT

SPSS is powerful to analyze business and marketing data. This paper intends to support business and marketing leaders the benefits of data forecasting with applied SPSS. It showed the sale quantity forecasting based on unit price and advertising. As SPSS's background algorithms, it showed the regression algorithm for data forecasting and ANOVA algorithm for data significant. It includes one sample data was downloaded from Google and was analyzed and viewed. It used IBM SPSS statistics version 23 and PYTHON version 3.7.

**KEYWORDS:** SPSS is powerful to analyze business and marketing data, the regression algorithm for data forecasting

## 1. INTRODUCTION

Nowadays, businesses are competing with others not to lose their market places in local and external regions. To avoid the loss of market places they should use data science technology. This paper used SPSS integrated with Python software. It showed the regression algorithm for data forecasting and ANOVA algorithm for data significant that includes five tables, one graph and three data analytical views.

### 1.1. SPSS

SPSS, standing for Statistical Package for the Social Sciences, is a powerful, user-friendly software package for the manipulation and statistical analysis of data.

*multiple regression*. All dialog boxes in SPSS provide for multiple regression.[3]

### 1.3. Scatter grams

SPSS has plotted the relationship between two variables:

1. revision: How much revision an individual did for an exam (scored from 0 to 30 hours) by
2. score: Exam Score (marked out of 30). on to a chart called a scatter gram.

These two variables are not identical, but you would expect a strong positive relationship between them. When you have a 'positive relationship', the low values on one variable tend to go with low values on the other variable and high values on one variable tend to go with high values on the other. Here, we would expect that people doing only a small amount of revision for an exam should tend to score low on the exam, and people doing a lot of revision should tend to score high. When you have a 'negative relationship', the low values on one variable tend to go with high values on the other variable and high values on one variable tend to go with low values on the other.[4]

**2. Algorithm****REGRESSION Algorithms[1]****Notation Description**

$y_i$	Dependent variable for case $i$ with variance $\sigma^2/g_i$
$c_i$	Caseweight for case $i$ ; $c_i = 1$ if CASEWEIGHT is not specified
$g_i$	Regression weight for case $i$ ; $g_i = 1$ if REGWGT is not specified
$l$	Number of distinct cases
$w_i$	$c_i g_i$
$W$	$\sum_{i=1}^l w_i$
$P$	Number of independent variables
$C$	Sum of caseweights: $\sum_{i=1}^l c_i$
$x_{ki}$	The $k$ th independent variable for case $i$
$\bar{X}_k$	Sample mean for the $k$ th independent variable: $\bar{X}_k = \left( \sum_{i=1}^l w_i x_{ki} \right) / W$
$\bar{Y}$	Sample mean for the dependent variable: $\bar{Y} = \left( \sum_{i=1}^l w_i y_i \right) / W$
$h_i$	Leverage for case $i$
$\tilde{h}_i$	$\frac{h_i}{W} + h_i$
$S_{kj}$	Sample covariance for $X_k$ and $X_j$
$S_{yy}$	Sample variance for $Y$
$S_{ky}$	Sample covariance for $X_k$ and $Y$
$p^*$	Number of coefficients in the model. $p^* = p$ if the intercept is not included; otherwise $p^* = p + 1$
$R$	The sample correlation matrix for $X_1, \dots, X_p$ and $Y$

**Multiple R**

$$R = \sqrt{1 - r_{yy}}$$

**R Square**

$$R^2 = 1 - r_{yy}$$

**Adjusted R Square**

$$R^2_{adj} = R^2 - \frac{(1 - R^2)p}{C - p^*}$$

**Standard Error of Estimate**

Also known as the standard error of regression, this is simply the square root of the mean square residual from the ANOVA table, or  $\sqrt{(SS_e)/(C - p^*)}$ .

**F Change and Significance of F Change**

$$\Delta F = \begin{cases} \frac{\Delta R^2(C - p^*)}{q(1 - R^2_{current})} & \text{for the addition of } q \text{ independent variables} \\ \frac{\Delta R^2(C - p^* - q)}{q(R^2_{previous} - 1)} & \text{for the removal of } q \text{ independent variables} \end{cases}$$

the degrees of freedom for the addition are  $q$  and  $C - p^*$ , while the degrees of freedom for the removal are  $q$  and  $C - p^* - q$ .

**Residual Sum of Squares**

$$SS_e = r_{yy}(C - 1)S_{yy}$$

with degrees of freedom  $C - p^*$ .

**Sum of Squares Due to Regression**

$$SS_R = R^2(C - 1)S_{yy}$$

with degrees of freedom  $p$ .

**ANOVA Table**

Table 87-2  
ANOVA table

Analysis of Variance	df	Sum of Squares	Mean Square
Regression	$p$	$SS_R$	$(SS_R)/p$
Residual	$C - p^*$	$SS_e$	$(SS_e)/(C - p^*)$

**Regression Coefficient**

$$b_k = \frac{r_{yk}\sqrt{S_{yy}}}{\sqrt{S_{kk}}} \text{ for } k = 1, \dots, p$$

The standard error of  $b_k$  is computed as

$$\hat{\sigma}_{b_k} = \sqrt{\frac{r_{kk}r_{yy}S_{yy}}{S_{kk}(C - p^*)}}$$

**95% confidence interval for coefficient**

$$b_k \pm \hat{\sigma}_{b_k} t_{0.975, C - p^*}$$

If the model includes the intercept, the intercept is estimated as

$$b_0 = \bar{y} - \sum_{k=1}^p b_k \bar{X}_k$$

### 3. Testing

**Table-1: Sample Data**

Quantity Sold	Price	Advertising
8500	2.00	2,800.00
4700	5.00	200.00
5800	3.00	400.00
7400	2.00	500.00
6200	5.00	3,200.00
7300	3.00	1,800.00
5600	4.00	900.00

**SUMMARY OUTPUT**

**Table-2: Regression Values**

<i>Regression Statistics</i>	
<b>Multiple R</b>	<b>0.9807</b>
<b>R Square</b>	<b>0.9617</b>
<b>Adjusted R Square</b>	<b>0.9426</b>
<b>Standard Error</b>	<b>310.52</b>
<b>Observations</b>	<b>7</b>

#### A. Analytical view

##### As Table-2

R Square equals 0.962, which is a very good fit. 96% of the variation in Quantity Sold is explained by the independent variables Price and Advertising. The closer to 1, the better the regression line (read on) fits the data.

**Table-3: Performance-1**

ANOVA					
	df	SS	MS	F	Significance F
Regression	2	9694299.568	4847150	50.2685	0.001464128
Residual	4	385700.4318	96425.1		
Total	6	10080000			

**Table-4: Performance-2**

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	8536.21388	386.9117478	22.0624	2.5E-05	7461.974654	9610.45311	7461.974654	9610.45311
Price	-835.72235	99.65304469	-8.3863	0.00111	-1112.40356	-559.04114	-1112.40356	-559.04114
Advertising	0.5922285	0.10436803	5.67558	0.00476	0.3102515325	0.88094167	0.3102515325	0.88094167

#### B. Analytical view

##### As table-3 and 4,

##### Significance F and P-values

To check if your results are reliable (statistically significant), look at Significance F (0.001). If this value is less than 0.05, you're OK. If Significance F is greater than 0.05, it's probably better to stop using this set of independent variables. Delete a variable with a high P-value (greater than 0.05) and rerun the regression until Significance F drops below 0.05.

Most of all P-values should be below 0.05. In our example, this is the case. (0.000, 0.001 and 0.005).

##### As table-4, Coefficients

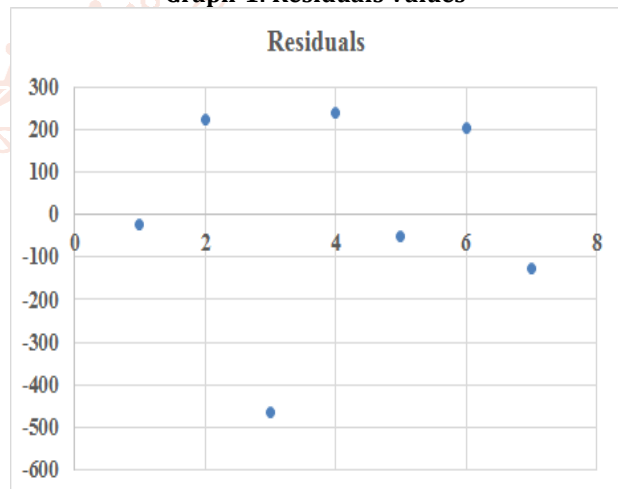
The regression line is:  $y = \text{Quantity Sold} = 8536.214 - 835.722 * \text{Price} + 0.592 * \text{Advertising}$ . In other words, for each unit increase in price, Quantity Sold decreases with 835.722 units. For each unit increase in Advertising, Quantity Sold increases with 0.592 units. This is valuable information.

You can also use these coefficients to do a forecast. For example, if price equals \$4 and Advertising equals \$3000, you might be able to achieve a Quantity Sold of  $8536.214 - 835.722 * 4 + 0.592 * 3000 = 6970$ .

**Table-5: RESIDUAL OUTPUT**

Observation	Predicted Quantity Sold	Residuals
1	8523.008967	-23.00896712
2	4476.047825	223.9521754
3	6265.938227	-465.9382265
4	7160.883427	239.1165726
5	6252.733311	-52.73331119
6	7095.05812	204.9418798
7	5726.330123	-126.3301229

**Graph-1: Residuals Values**



#### C. analytical Views

As table-5 and graph-1

##### Residuals

The residuals show you how far away the actual data points are from the predicted data points (using the equation). For example, the first data point equals 8500. Using the equation, the predicted data point equals  $8536.214 - 835.722 * 2 + 0.592 * 2800 = 8523.009$ , giving a residual of  $8500 - 8523.009 = -23.009$ .

#### 4. Conclusion

SPSS data analysis tools are valuable in social science, business and marketing fields. It is very good for presentation report by graphical design. It showed the sale quantity forecasting based on unit price and advertising before sale and then they can get their goal with right sale quantity and can avoid the waste of sale quantity and the lost of market places in local and global regions by using SPSS software.

#### References

- [1] IBM SPSS Statistics 24 Algorithms pdf book [book style]
- [2] A handbook of statistical analyses using SPSS / Sabine, Landau, Brian S. Everitt, ISBN 1-58488-369-3 [book style]
- [3] SPSS For Dummies®, 2nd Edition, ISBN: 978-0-470-48764-8 [book style]
- [4] SPSS for Social Scientists Robert L. Miller, Ciaran Acton, Deirdre A. Fullerton and John Maltby, ISBN 0-333-92286-7 [book style]

#### Author Profile



now with University of Computer Studies, Maubin.

**Aung Cho** received the B.A.(Eco) degree from Yangon University in 1987 and M.I.Sc.(Information Science) degree from University of Computer Studies, Yangon in 2001. After got Master degree, I served as a teacher at the software, information science and application departments of the computer universities. I am

#### Author Profile



computer universities. I am now with University of Computer Studies, Maubin, Myanmar.

**Aung Si Thu** received the B.Sc.(Hons)(Chemistry) degree from Magwe University in 2003 and M.I.Sc.(Information Science) degree from University of Computer Studies, Yangon in 2009. After got Master degree, I served as a teacher at the software, information science and hardware departments of the

