Contents

1 Output of Linear Regression Analysis

Linear regression is used when you want to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable). The variable we are using to predict the other variable's value is called the independent variable (or sometimes, the predictor variable).

$$\hat{y} = b_o + b_1 x$$

- b_o is the intercept estimate
- b_1 is the slope estimate
- $\hat{y}$ is the ***fitted*** y-value, given a specified value of the indepenent variable x.
- The fitted value and the observed value of y often differ. This difference is known as the ***residual***

For example, you could use linear regression to understand whether exam performance can be predicted based on revision time; whether cigarette consumptions can be predicted based on smoking duration; and so forth. If you have two or more independent variables, rather than just one, you need to use ***multiple regression***.

SPSS will generate quite a few tables of output for a linear regression procedure. Only the three main tables required to understand your results from the linear regression procedure, assuming that no assumptions have been violated.

This includes relevant scatterplots, histogram (with superimposed normal curve) and Normal P-P Plot, and case-wise diagnostics and Durbin-Watson statistic tables. Below, we focus on the results for the linear regression analysis only.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.873[a]	.762	.749	874.779

a. Predictors: (Constant), Income

Figure 1: Model Summary table

The first table of interest is the ***Model Summary*** table. This table provides the R and R^2 value. The R value is 0.873, which represents the simple correlation. It indicates a high degree of correlation. The R^2 value indicates how much of the dependent variable, ***price*** (Not evident on output), can be explained by the independent variable,***income***. In this case, 76.2% can be explained, which is very large.

The next table is the ANOVA table. This table indicates that the regression model predicts the outcome variable significantly well. How do we know this? Look at the ***Regression*** row and go to the **Sig.** column. This indicates the statistical significance of the regression model that was applied. Here,the p-value is $p < 0.0005$, which is less than 0.05, and indicates that, overall, the model applied can statistically significantly predict the outcome variable.

The next table again, ***Coefficients***, provides us with information on each predictor variable. This gives us the information we need to predict price from income. We can see that both the

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.418E7	1	4.418E7	57.737	.000[a]
	Residual	1.377E7	18	765238.393		
	Total	5.796E7	19			

a. Predictors: (Constant), Income

b. Dependent Variable: Price

Figure 2: ANOVA Table

constant and income contribute significantly to the model (by looking at the Sig. column). By looking at the B column under the Unstandardized Coefficients column, we can present the regression equation as:

$\hat{Price}$ = ***8287 + 0.564(Income)***

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8286.786	1852.256		4.474	.000
	Income	.564	.074	.873	7.598	.000

a. Dependent Variable: Price

Figure 3: Coefficients Table

2 Multiple Linear Regression

2.1 What is Multiple Linear Regression

Multiple regression is a statistical technique that allows us to predict a numeric value on the response variable on the basis of the observed values on several other independent variables.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots$$

- $\hat{y}$ is the ***fitted value*** for the dependent variable Y, given a linear combination of values for the independent valriables.
- x_i is the value for independent variable X_i. (For Example, x_1 is the value for independent variable X_1.)
- b_o is the constant regression estimate (commonly known as the **Intercept Estimate** in the case of simple linear regression).
- b_i is the regression estimate for Independent Variable X_1 (commonly known as the **Slope Estimate** in the case of simple linear regression).

2.1.1 Simple Example

Suppose we were interested in predicting how much an individual enjoys their job. Independent Variables such as salary, extent of academic qualifications, age, sex, number of years in full-time employment and socioeconomic status might all contribute towards ***job satisfaction***.

If we collected data on all of these variables, perhaps by surveying a few hundred members of the public, we would be able to see how many and which of these variables gave rise to the most accurate prediction of job satisfaction. We might find that job satisfaction is most accurately predicted by type of occupation, salary and years in full-time employment, with the other variables not helping us to predict job satisfaction.