

Part A

Introduction

This report details the steps taken to find the function that generated dependent variable values based on the values of an independent variable.

Methodology

Use SPSS to process the data. We first open both Data Set 1 and Data Set 2 into the environment. Data Set 1 contains ID numbers as well as the respective independent variables while Data Set 2 contained the respective dependent variables. Then, using the “Data” function we merge the sets and had the software add variables to the data. The environment now has an ID column going from 1 to 1709, and one independent variable (IV) column and one dependent variable column (DV). Before removing any missing values there were 1709 independent variables with IDs that ranged from 1 to 1709 and 1709 dependent variables with IDs ranging from 1 to 1709. Then, we reorder the data by using the “Variable view” function, and we set variable to numeric. Finally, we use the “Analyze” function and used a missing value analysis, selecting DV as the quantitative variable. This produced 1592 valid values for this data set, and 117 missing values. The SPSS software automatically uses a listwise deletion function while doing any statistical analysis so that the missing values are not taken into account. We then use the “Analyze” function again, selecting a linear regression with the DV column as the dependent variables, and IV as the independent variables. We then use the “Statistics” function to get estimates of confidence intervals (95%), model fit, R^2 change and Descriptives.

Results

Since $p=0.000$ we know that the association between the independent and dependent variable is highly significant. The model was found to be $DV=7.225IV + (-77.161)$. The 95% confidence interval for the intercept was given as -81.970 to -72.353. The 95% confidence interval for the slope was given as 7.186 to 7.265. The R^2 value was .988 and the correlational coefficient was found to be .994.

Table 1
Analysis of Variance Table
DV regressed on IV
(n=1592)

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1221432331.136	1	1221432331.136	128337.810	.000 ^b
	Residual	15132542.879	1590	9517.323		
	Total	1236564874.015	1591			

a. Dependent Variable: DV

b. Predictors: (Constant), IV

Conclusion

The association between independent variables and dependent variables in part A is highly significant ($p=0.000$).

Part B

Introduction

There are two files, one containing a subject identifier and value of the independent variable, and the other one containing the identifier and the value of the dependent variable. The objective is to recover the function that was used to generate the dependent variable value based on the value of the independent variable.

Methodology

We use the statistics package SPSS to process the data and solve the problem. The data files that were provided are merged using the Data function. At the beginning there were 1645 dependent, as well as independent variables with ID ranging from 1 to 1645. We use the Variable View function to reorder the data, setting variable to numeric. Next, we use the Analyze function to run a missing value analysis, selecting DV as the quantitative variable. The findings showed that there are 1525 valid values and 120 missing values. The SPSS software runs a listwise deletion function when the tests are run. We then run a linear regression analysis with the DV column as the dependent variables, and IV as the independent variables. It is important to note that there was no need for a transformation of the data as a scatter plot revealed the data to be linear. The R^2 value was found to be .955 which is very close to 1 and shows there to be no need for a transformation. We use the Statistics function for a good estimation of the confidence intervals (95%), model fit, R squared change and Descriptives.

Results

Table 2
Analysis of Variance Table
DV regressed on IV
(n = 1525)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	132615.664	1	132615.664	32473.029	.000 ^b
	Residual	6219.736	1523	4.084		
	Total	138835.399	1524			

a. Dependent Variable: DV

b. Predictors: (Constant), IV

The regression formula is given as $DV = .037IV + 37.48$. The 95% confidence interval for B showed to be 37.046 to 37.909. The 95% confidence interval for the slope showed to be .036 to .037. The R^2 value was found to be .955 and the correlational coefficient was .977.

Conclusion

The association between independent variables and dependent variables in part B is highly significant ($p=0.000$).