



Using SPSS for Regression

Your assignment is to use regression to predict credit card spending by replicating the information contained herein SPSS. Copy the SPSS results into WORD and answer the questions in sections I and II in one or two sentences. Please label your analyses beginning with Numeral I.

I. Simple Regression

- Age
- Dollar amount person is expected to charge

II. Multiple Regression

- Age, Income, Home Value
- Dollar amount person is expected to charge

All of the SPSS output shown in the following pages is based on a sample of 1000 customer records and variables drawn from the Bank Credit Card Dataset file of customer information. The variables used are:

- Charge: Amount person is expected to charge
- Homeval: Value of a person's home in thousands of dollars
- Income: Person's annual income in thousands of dollars
- Age: (in years)

I. Simple Regression

Independent variable (Age)/ Dependent variable (Charge)

Simple regression allows for the prediction of a dependent variable (Y) from an independent variable (X). Both variables are interval or ratio scale (Scale variables in SPSS).

The Bank Credit Card Dataset includes the variables AGE indicating the person's age and CHARGE indicating the amount a person is expected to charge.

Using the Five Step Process, determine if a person's age predicts the amount a person is expected to charge on a credit card. Please include in your write up in WORD:

- The SPSS output
- Apply decision rule to draw a conclusion and how you could use this conclusion as a marketing manager. What other statistical analysis could you conduct that would add insight to this analysis?

To test the claim that a person's (AGE) predicts the amount a person is expected to charge on a credit card (CHARGE):

1. From the pull-down menus, select **Analyze / Regression / Linear**
2. In the dialog box that appears, select CHARGE for the dependent variable (click on CHARGE, then click on the arrow to move it to Dependent:).
3. Select AGE for the independent variable (click on AGE, then click on the arrow to move it to Independent(s):).
4. Select the **Statistics** button. The statistics button allows you to select a number of statistics to report. The default selection is Estimates under Regression Coefficients and Model Fit. If these are not selected, please select them. Select "Continue."
5. The **Plots** button allows you to produce a number of plots for testing the assumptions of regression. For this example, no plots are required.
6. The **Save** button allows you to save a number of calculated items produced from the regression analysis. These are mainly used for checking regression assumptions and outliers. For this example, none of these items need to be saved.
7. The **Options** button allows you to change the stepping method criteria. The default is Use probability of F Entry: .05 and Removal: .10. For this example, no changes to the default values are required.
8. The **Style** button allows you to format the background and text within tables. For this example, you do not need to this.
9. Click "OK." The Output window will display four tables. The first table, **Variable Entered/Removed** provides the independent variable(s) and dependent variable in the model and the method for entering the independent variables in the model. The second table, **Model Summary** provides R Square (coefficient of determination) of the model.

This is a measure of the variance of the dependent variable explained by the independent variable. The third table, **ANOVA**, tests the hypothesis if the two variables are related. The important number to review is the significance (Sig). If the significance is less than .05, the two variables are related and the independent variable predicts the dependent variable. If the significance is greater than .05, then the two variables are not related. The fourth table, **Coefficients**, gives you the coefficients for the regression equation. Use the unstandardized B (beta) coefficients. Notice that the significance (Sig) is less than .05 for a person's age. This means that AGE is related to the dependent variable (CHARGE). In this case, a one unit change in AGE will result in a 17.217 increase in the amount a person is expected to charge on a credit card.

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Person's Age ^b	.	Enter

a. Dependent Variable: Amount person is expected to charge

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.319 ^a	.102	.101	883.66629

a. Predictors: (Constant), Person's Age

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	88470051.95	1	88470051.95	113.297	.000 ^b
	Residual	779304384.3	998	780866.117		
	Total	867774436.3	999			

a. Dependent Variable: Amount person is expected to charge

b. Predictors: (Constant), Person's Age

Coefficients^a

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	1436.690	85.469		16.809	.000
	Person's Age	17.217	1.618	.319	10.644	.000

a. Dependent Variable: Amount person is expected to charge

II. Multiple Regression

Independent variables (Age, Income, Home Value)/ Dependent variable (Charge)

Multiple regression is an extension of simple linear regression. Multiple regression allows for the prediction of a dependent variable (Y) from several independent variables (X). All the variables are interval or ratio scale (Scale variables in SPSS).

The Bank Credit Card Dataset includes the variables AGE indicating the person's age, HOMEEVAL indicating the value of a person's home in thousands of dollars, INCOME indicating a person's annual income in thousands of dollars, and CHARGE indicating the amount a person is expected to charge.

Using the Five Step Process, determine if a person's age, value of a person's home, and a person's annual income predicts the amount a person is expected to charge on a credit card. Please include in your write up in WORD:

- The SPSS output
- Apply decision rule to draw a conclusion and how you could use this conclusion as a marketing manager. What other statistical analysis could you conduct that would add insight to this analysis?

To test the claim that a person's age (AGE), value of a person's home (HOMEEVAL), and a person's annual income (INCOME) predict the amount a person is expected to charge on a credit card (CHARGE):

1. From the pull-down menus, select **Analyze / Regression / Linear**
2. In the dialog box that appears, select CHARGE for the dependent variable (click on CHARGE, then click on the arrow to move it to Dependent:).
3. Select AGE, HOMEEVAL, and INCOME for the independent variables (click on AGE, HOMEEVAL, and INCOME then click on the arrow to move them to Independent(s):).
4. Select the **Statistics** button. The statistics button allows you to select a number of statistics to report. The default selection is Estimates under Regression Coefficients and Model Fit. If these are not selected, please select them. Select "Continue."
5. The **Plots** button allows you to produce a number of plots for testing the assumptions of regression. For this example, no plots are required.
6. The **Save** button allows you to save a number of calculated items produced from the regression analysis. These are mainly used for checking regression assumptions and outliers. For this example, none of these items need to be saved.
7. The **Options** button allows you to change the stepping method criteria. The default is Use probability of F Entry: .05 and Removal: .10. For this example, no changes to the default values are required.

8. The **Style** button allows you to format the background and text within tables. For this example, you do not need to this.
9. Click “OK.” The Output window will display four tables. The first table, **Variable Entered/Removed** provides the independent variable(s) and dependent variable in the model and the method for entering the independent variables in the model. The second table, **Model Summary** provides adjusted R Square (coefficient of determination) of the model. In multiple regression, we use adjusted R square instead of R square. Adjusted R square accounts for the effect of multiple independent variables in the model, and it is a measure of the variance of the dependent variable explained by the independents variable. The third table, **ANOVA**, tests the hypothesis if the independent and dependent variables are related. The important number to review is the significance (Sig). If the significance is less than .05, then there is a linear relationship between the independent variables and the dependent variable. If the significance is greater than .05, then the independent variables and the dependent are not linearly related. If this is the case, there is no need to test whether individual variables in the **Coefficients** table are significant or not. The fourth table, **Coefficients**, gives you the coefficients for the regression equation. Use the unstandardized B (beta) coefficients. Notice that the significance (Sig) is less than .05 for a person’s age. This means that AGE is related to the dependent variable CHARGE. In this case, a one unit change in AGE will result in a 12.695 increase in the amount a person is expected to charge on a credit card. It is similar for HOMEEVAL and INCOME. A unit change in HOMEEVAL results in a 1.429 increase in the amount a person is expected to charge on a credit card. And, a unit change in INCOME results in a 12.072 increase in the amount a person is expected to charge on a credit card

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Person's annual income, Person's Age, Value of a person's home ^b	.	Enter

a. Dependent Variable: Amount person is expected to charge

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.579 ^a	.335	.333	761.08157

a. Predictors: (Constant), Person's annual income, Person's Age, Value of a person's home

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	290846258.7	3	96948752.89	167.371	.000 ^b
	Residual	576928177.6	996	579245.158		
	Total	867774436.3	999			

a. Dependent Variable: Amount person is expected to charge

b. Predictors: (Constant), Person's annual income, Person's Age, Value of a person's home

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	191.245	99.500		1.922	.055
	Person's Age	12.695	1.414	.235	8.977	.000
	Value of a person's home in thousands of dollars	1.429	.191	.198	7.476	.000
	Person's annual income in thousands of dollars	12.072	.779	.412	15.497	.000

a. Dependent Variable: Amount person is expected to charge