0.1 Introduction

Linear regression is the next step up after correlation. It is used when we 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). For example, you could use linear regression to understand whether exam performance can be predicted based on revision time; whether cigarette consumption 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.

0.2 Assumptions

When you choose to analyse your data using linear regression, part of the process involves checking to make sure that the data you want to analyse can actually be analysed using linear regression. You need to do this because it is only appropriate to use linear regression if your data "passes" six assumptions that are required for linear regression to give you a valid result. In practice, checking for these six assumptions just adds a little bit more time to your analysis, requiring you to click a few more buttons in SPSS Statistics when performing your analysis, as well as think a little bit more about your data, but it is not a difficult task.

Before we introduce you to these six assumptions, do not be surprised if, when analysing your own data using SPSS Statistics, one or more of these assumptions is violated (i.e., not met). This is not uncommon when working with real-world data rather than textbook examples, which often only show you how to carry out linear regression when everything goes well! However, dont worry. Even when your data fails certain assumptions, there is often a solution to overcome this. First, lets take a look at these six assumptions:

Assumption 1: Your two variables should be measured at the continuous level (i.e., they are either interval or ratio variables). Examples of continuous variables include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth. You can learn more about interval and ratio variables in our article: Types of Variable. Assumption 2: There needs to be a linear relationship between the two variables. Whilst there are a number of ways to check whether a linear relationship exists between your two variables, we suggest creating a scatterplot using SPSS Statistics where you can plot the dependent variable against your independent variable and then visually inspect the scatterplot to check for linearity. Your scatterplot may look something like one of the following:

Types of relationship

If the relationship displayed in your scatterplot is not linear, you will have to either run a non-linear regression analysis, perform a polynomial regression or "transform" your data, which you can do using SPSS Statistics. In our enhanced guides, we show you how to: (a) create a scatterplot to check for linearity when carrying out linear regression using SPSS Statistics; (b) interpret different scatterplot results; and (c) transform your data using SPSS Statistics if there is not a linear relationship between your two variables. Assumption 3: There should be no significant outliers. An outlier is an observed data point that has a dependent variable value that is very different to the value predicted by the regression equation. As such, an outlier will be a point on a scatterplot that is (vertically) far away from the regression line indicating that it has a large residual, as highlighted below:

1

Outliers in linear regression

The problem with outliers is that they can have a negative effect on the regression analysis (e.g., reduce the fit of the regression equation) that is used to predict the value of the dependent (outcome) variable based on the independent (predictor) variable. This will change the output that SPSS Statistics produces and reduce the predictive accuracy of your results. Fortunately, when using SPSS Statistics to run a linear regression on your data, you can easily include criteria to help you detect possible outliers. In our enhanced linear regression guide, we: (a) show you how to detect outliers using "casewise diagnostics", which is a simple process when using SPSS Statistics; and (b) discuss some of the options you have in order to deal with outliers. Assumption 4: You should have independence of observations, which you can easily check using the Durbin-Watson statistic, which is a simple test to run using SPSS Statistics. We explain how to interpret the result of the Durbin-Watson statistic in our enhanced linear regression guide. Assumption 5: Your data needs to show homoscedasticity, which is where the variances along the line of best fit remain similar as you move along the line. Whilst we explain more about what this means and how to assess the homoscedasticity of your data in our enhanced linear regression guide, take a look at the three scatterplots below, which provide three simple examples: two of data that fail the assumption (called heteroscedasticity) and one of data that meets this assumption (called homoscedasticity):

Homoscedasticity in linear regression

Whilst these help to illustrate the differences in data that meets or violates the assumption of homoscedasticity, real-world data can be a lot more messy and illustrate different patterns of heteroscedasticity. Therefore, in our enhanced linear regression guide, we explain: (a) some of the things you will need to consider when interpreting your data; and (b) possible ways to continue with your analysis if your data fails to meet this assumption. Assumption 6: Finally, you need to check that the residuals (errors) of the regression line are approximately normally distributed (we explain these terms in our enhanced linear regression guide). Two common methods to check this assumption include using either a histogram (with a superimposed normal curve) or a Normal P-P Plot. Again, in our enhanced linear regression guide, we: (a) show you how to check this assumption using SPSS Statistics, whether you use a histogram (with superimposed normal curve) or Normal P-P Plot; (b) explain how to interpret these diagrams; and (c) provide a possible solution if your data fails to meet this assumption. You can check assumptions 2, 3, 4, 5 and 6 using SPSS Statistics. Assumptions 2 should be checked first, before moving onto assumptions 3, 4, 5 and 6. We suggest testing the assumptions in this order because assumptions 3, 4, 5 and 6 require you to run the linear regression procedure in SPSS Statistics first, so it is easier to deal with these after checking assumption 2. Just remember that if you do not run the statistical tests on these assumptions correctly, the results you get when running a linear regression might not be valid. This is why we dedicate a number of sections of our enhanced linear regression guide to help you get this right. You can find out more about our enhanced content as a whole here, or more specifically, learn how we help with testing assumptions here.

In the section, Procedure, we illustrate the SPSS Statistics procedure to perform a linear regression assuming that no assumptions have been violated. First, we introduce the example that is used in this class. Example A salesperson for a large car brand wants to determine whether there is a relationship between an individual's income and the price they pay for a car. As such, the individual's "income" is the independent variable and the "price" they pay for a car is the dependent variable. The salesperson wants to use this information to determine which cars to offer potential customers in new areas where average income is known.

Setup in SPSS Statistics In SPSS Statistics, we created two variables so that we could enter our data: Income (the independent variable), and Price (the dependent variable). It can also be useful to create a third

variable, caseno, to act as a chronological case number. This third variable is used to make it easy for you to eliminate cases (e.g., significant outliers) that you have identified when checking for assumptions. However, we do not include it in the SPSS Statistics procedure that follows because we assume that you have already checked these assumptions. In our enhanced linear regression guide, we show you how to correctly enter data in SPSS Statistics to run a linear regression when you are also checking for assumptions. You can learn about our enhanced data setup content here. Alternately, we have a generic, "quick start" guide to show you how to enter data into SPSS Statistics, available here.

Test Procedure in SPSS Statistics The five steps below show you how to analyse your data using linear regression in SPSS Statistics when none of the six assumptions in the previous section, Assumptions, have been violated. At the end of these four steps, we show you how to interpret the results from your linear regression. If you are looking for help to make sure your data meets assumptions 2, 3, 4, 5 and 6, which are required when using linear regression and can be tested using SPSS Statistics, you can learn more in our enhanced guide here.

Click Analyze ¿ Regression ¿ Linear... on the top menu, as shown below:

Linear Regression Menu in SPSS Statistics Published with written permission from SPSS Statistics, IBM Corporation. You will be presented with the Linear Regression dialogue box:

SPSS Statistics Linear Regression Dialogue Box

Transfer the independent variable, Income, into the Independent(s): box and the dependent variable, Price, into the Dependent: box. You can do this by either drag-and-dropping the variables or by using the appropriate SPSS Statistics Right Arrow Button buttons. You will end up with the following screen:

SPSS Statistics Linear Regression Dialogue Box with Variables Transferred Published with written permission from SPSS Statistics, IBM Corporation. You now need to check four of the assumptions discussed in the Assumptions section above: no significant outliers (assumption 3); independence of observations (assumption 4); homoscedasticity (assumption 5); and normal distribution of errors/residuals (assumptions 6). You can do this by using the and features, and then selecting the appropriate options within these two dialogue boxes. In our enhanced linear regression guide, we show you which options to select in order to test whether your data meets these four assumptions.

Click the button. This will generate the results.

Join the 10,000s of students, academics and professionals who rely on Laerd Statistics.TAKE THE TOUR PLANS and PRICING

Output of Linear Regression Analysis SPSS Statistics will generate quite a few tables of output for a linear regression. In this section, we show you only the three main tables required to understand your results from the linear regression procedure, assuming that no assumptions have been violated. A complete explanation of the output you have to interpret when checking your data for the six assumptions required to carry out linear regression is provided in our enhanced guide. This includes relevant scatterplots, histogram (with superimposed normal curve), Normal P-P Plot, casewise diagnostics and the Durbin-Watson statistic. Below, we focus on the results for the linear regression analysis only.

The first table of interest is the Model Summary table, as shown below:

Model Summary Table for Linear Regression Procedure in SPSS Statistics Published with written permission from SPSS Statistics, IBM Corporation. This table provides the R and R2 values. The R value represents the simple correlation and is 0.873 (the "R" Column), which indicates a high degree of correlation. The R2 value (the "R Square" column) indicates how much of the total variation in the dependent variable, Price, can be explained by the independent variable, Income. In this case, 76.2

The next table is the ANOVA table, which reports how well the regression equation fits the data (i.e., predicts the dependent variable) and is shown below:

ANOVA Table for Linear Regression Procedure in SPSS Statistics Published with written permission from SPSS Statistics, IBM Corporation. This table indicates that the regression model predicts the dependent

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 run. Here, p ¡ 0.0005, which is less than 0.05, and indicates that, overall, the regression model statistically significantly predicts the outcome variable (i.e., it is a good fit for the data).

The Coefficients table provides us with the necessary information to predict price from income, as well as determine whether income contributes statistically significantly to the model (by looking at the "Sig." column). Furthermore, we can use the values in the "B" column under the "Unstandardized Coefficients" column, as shown below:

Coefficients Table for Linear Regression in SPSS Statistics Published with written permission from SPSS Statistics, IBM Corporation. to present the regression equation as:

$$Price = 8287 + 0.564(Income)$$

If you are unsure how to interpret regression equations or how to use them to make predictions, we discuss this in our enhanced linear regression guide. We also show you how to write up the results from your assumptions tests and linear regression output if you need to report this in a dissertation/thesis, assignment or research report. We do this using the Harvard and APA styles. You can learn more about our enhanced content here.

Laerd StatisticsLoginCookies and Privacy Take the Tour Plans and Pricing SIGN UP Multiple Regression Analysis using SPSS Statistics

Introduction Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable). The variables we are using to predict the value of the dependent variable are called the independent variables (or sometimes, the predictor, explanatory or regressor variables).

For example, you could use multiple regression to understand whether exam performance can be predicted based on revision time, test anxiety, lecture attendance and gender. Alternately, you could use multiple regression to understand whether daily cigarette consumption can be predicted based on smoking duration, age when started smoking, smoker type, income and gender.

Multiple regression also allows you to determine the overall fit (variance explained) of the model and the relative contribution of each of the predictors to the total variance explained. For example, you might want to know how much of the variation in exam performance can be explained by revision time, test anxiety, lecture attendance and gender "as a whole", but also the "relative contribution" of each independent variable in explaining the variance.

This "quick start" guide shows you how to carry out multiple regression using SPSS Statistics, as well as interpret and report the results from this test. However, before we introduce you to this procedure, you need to understand the different assumptions that your data must meet in order for multiple regression to give you a valid result. We discuss these assumptions next.

SPSS Statisticstop and Assumptions When you choose to analyse your data using multiple regression, part of the process involves checking to make sure that the data you want to analyse can actually be analysed using multiple regression. You need to do this because it is only appropriate to use multiple regression if your data "passes" eight assumptions that are required for multiple regression to give you a valid result. In practice, checking for these eight assumptions just adds a little bit more time to your analysis, requiring you to click a few more buttons in SPSS Statistics when performing your analysis, as well as think a little bit more about your data, but it is not a difficult task.

Before we introduce you to these eight assumptions, do not be surprised if, when analysing your own data using SPSS Statistics, one or more of these assumptions is violated (i.e., not met). This is not uncommon when working with real-world data rather than textbook examples, which often only show you how to carry out multiple regression when everything goes well! However, dont worry. Even when your data fails certain assumptions, there is often a solution to overcome this. First, let's take a look at these eight assumptions:

Assumption 1: Your dependent variable should be measured on a continuous scale (i.e., it is either an interval or ratio variable). Examples of variables that meet this criterion include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth. You can learn more about interval and ratio variables in our article: Types of Variable. If your dependent variable was measured on an ordinal scale, you will need to carry out ordinal regression rather than multiple regression. Examples of ordinal variables include Likert items (e.g., a 7-point scale from "strongly agree" through to "strongly disagree"), amongst other ways of ranking categories (e.g., a 3-point scale explaining how much a customer liked a product, ranging from "Not very much" to "Yes, a lot"). You can access our SPSS Statistics guide on ordinal regression here. Assumption 2: You have two or more independent variables, which can be either continuous (i.e., an interval or ratio variable) or categorical (i.e., an ordinal or nominal variable). For examples of continuous and ordinal variables, see the bullet above. Examples of nominal variables include gender (e.g., 2 groups: male and female), ethnicity (e.g., 3 groups: Caucasian, African American and Hispanic), physical activity level (e.g., 4 groups: sedentary, low, moderate and high), profession (e.g., 5 groups: surgeon, doctor, nurse, dentist, therapist), and so forth. Again, you can learn more about variables in our article: Types of Variable. If one of your independent variables is dichotomous and considered a moderating variable, you might need to run a Dichotomous moderator analysis. Assumption 3: You should have independence of observations (i.e., independence of residuals), which you can easily check using the Durbin-Watson statistic, which is a simple test to run using SPSS Statistics. We explain how to interpret the result of the Durbin-Watson statistic, as well as showing you the SPSS Statistics procedure required, in our enhanced multiple regression guide. Assumption 4: There needs to be a linear relationship between (a) the dependent variable and each of your independent variables, and (b) the dependent variable and the independent variables collectively. Whilst there are a number of ways to check for these linear relationships, we suggest creating scatterplots and partial regression plots using SPSS Statistics, and then visually inspecting these scatterplots and partial regression plots to check for linearity. If the relationship displayed in your scatterplots and partial regression plots are not linear, you will have to either run a non-linear regression analysis or "transform" your data, which you can do using SPSS Statistics. In our enhanced multiple regression guide, we show you how to: (a) create scatterplots and partial regression plots to check for linearity when carrying out multiple regression using SPSS Statistics; (b) interpret different scatterplot and partial regression plot results; and (c) transform your data using SPSS Statistics if you do not have linear relationships between your variables. Assumption 5: Your data needs to show homoscedasticity, which is where the variances along the line of best fit remain similar as you move along the line. We explain more about what this means and how to assess the homoscedasticity of your data in our enhanced multiple regression guide. When you analyse your own data, you will need to plot the studentized residuals against the unstandardized predicted values. In our enhanced multiple regression guide, we explain: (a) how to test for homoscedasticity using SPSS Statistics; (b) some of the things you will need to consider when interpreting your data; and (c) possible ways to continue with your analysis if your data fails to meet this assumption. Assumption 6: Your data must not show multicollinearity, which occurs when you have two or more independent variables that are highly correlated with each other. This leads to problems with understanding which independent variable contributes to the variance explained in the dependent variable, as well as technical issues in calculating a multiple regression model. Therefore, in our enhanced multiple regression guide, we show you: (a) how to use SPSS Statistics to detect for multicollinearity through an inspection of correlation coefficients and Tolerance/VIF values; and (b) how to interpret these correlation

coefficients and Tolerance/VIF values so that you can determine whether your data meets or violates this assumption. Assumption 7: There should be no significant outliers, high leverage points or highly influential points. Outliers, leverage and influential points are different terms used to represent observations in your data set that are in some way unusual when you wish to perform a multiple regression analysis. These different classifications of unusual points reflect the different impact they have on the regression line. An observation can be classified as more than one type of unusual point. However, all these points can have a very negative effect on the regression equation that is used to predict the value of the dependent variable based on the independent variables. This can change the output that SPSS Statistics produces and reduce the predictive accuracy of your results as well as the statistical significance. Fortunately, when using SPSS Statistics to run multiple regression on your data, you can detect possible outliers, high leverage points and highly influential points. In our enhanced multiple regression guide, we: (a) show you how to detect outliers using "casewise diagnostics" and "studentized deleted residuals", which you can do using SPSS Statistics, and discuss some of the options you have in order to deal with outliers; (b) check for leverage points using SPSS Statistics and discuss what you should do if you have any; and (c) check for influential points in SPSS Statistics using a measure of influence known as Cook's Distance, before presenting some practical approaches in SPSS Statistics to deal with any influential points you might have. Assumption 8: Finally, you need to check that the residuals (errors) are approximately normally distributed (we explain these terms in our enhanced multiple regression guide). Two common methods to check this assumption include using: (a) a histogram (with a superimposed normal curve) and a Normal P-P Plot; or (b) a Normal Q-Q Plot of the studentized residuals. Again, in our enhanced multiple regression guide, we: (a) show you how to check this assumption using SPSS Statistics, whether you use a histogram (with superimposed normal curve) and Normal P-P Plot, or Normal Q-Q Plot; (b) explain how to interpret these diagrams; and (c) provide a possible solution if your data fails to meet this assumption. You can check assumptions 3, 4, 5, 6, 7 and 8 using SPSS Statistics. Assumptions 1 and 2 should be checked first, before moving onto assumptions 3, 4, 5, 6, 7 and 8. Just remember that if you do not run the statistical tests on these assumptions correctly, the results you get when running multiple regression might not be valid. This is why we dedicate a number of sections of our enhanced multiple regression guide to help you get this right. You can find out about our enhanced content as a whole here, or more specifically, learn how we help with testing assumptions here.

In the section, Procedure, we illustrate the SPSS Statistics procedure to perform a multiple regression assuming that no assumptions have been violated. First, we introduce the example that is used in this guide.

TAKE THE TOUR

PLANS and PRICING SPSS Statisticstop and Example A health researcher wants to be able to predict "VO2max", an indicator of fitness and health. Normally, to perform this procedure requires expensive laboratory equipment and necessitates that an individual exercise to their maximum (i.e., until they can longer continue exercising due to physical exhaustion). This can put off those individuals who are not very active/fit and those individuals who might be at higher risk of ill health (e.g., older unfit subjects). For these reasons, it has been desirable to find a way of predicting an individual's VO2max based on attributes that can be measured more easily and cheaply. To this end, a researcher recruited 100 participants to perform a maximum VO2max test, but also recorded their "age", "weight", "heart rate" and "gender". Heart rate is the average of the last 5 minutes of a 20 minute, much easier, lower workload cycling test. The researcher's goal is to be able to predict VO2max based on these four attributes: age, weight, heart rate and gender.

SPSS Statisticstop and Setup in SPSS Statistics In SPSS Statistics, we created six variables: (1) VO2max, which is the maximal aerobic capacity; (2) age, which is the participant's age; (3) weight, which is the participant's weight (technically, it is their 'mass'); (4) heart_rate, which is the participant's heart rate; (5) gender, which is the participant's gender; and (6) caseno, which is the case number. The caseno variable is used to make it easy for you to eliminate cases (e.g., "significant outliers", "high leverage points" and "highly influ-

ential points") that you have identified when checking for assumptions. In our enhanced multiple regression guide, we show you how to correctly enter data in SPSS Statistics to run a multiple regression when you are also checking for assumptions. You can learn about our enhanced data setup content here. Alternately, we have a generic, "quick start" guide to show you how to enter data into SPSS Statistics, available here.

SPSS Statisticstop and Test Procedure in SPSS Statistics The seven steps below show you how to analyse your data using multiple regression in SPSS Statistics when none of the eight assumptions in the previous section, Assumptions, have been violated. At the end of these seven steps, we show you how to interpret the results from your multiple regression. If you are looking for help to make sure your data meets assumptions 3, 4, 5, 6, 7 and 8, which are required when using multiple regression and can be tested using SPSS Statistics, you can learn more in our enhanced guide here.

Click Analyze ¿ Regression ¿ Linear... on the main menu, as shown below:

Published with written permission from SPSS Statistics, IBM Corporation. Note: Don't worry that you're selecting Analyze ¿ Regression ¿ Linear... on the main menu or that the dialogue boxes in the steps that follow have the title, Linear Regression. You have not made a mistake. You are in the correct place to carry out the multiple regression procedure. This is just the title that SPSS Statistics gives, even when running a multiple regression procedure.

You will be presented with the Linear Regression dialogue box below:

Published with written permission from SPSS Statistics, IBM Corporation. Transfer the dependent variable, VO2max, into the Dependent: box and the independent variables, age, weight, heart_rate and gender into the Independent(s): box, using the buttons, as shown below (all other boxes can be ignored):

Published with written permission from SPSS Statistics, IBM Corporation. Note: For a standard multiple regression you should ignore the and buttons as they are for sequential (hierarchical) multiple regression. The Method: option needs to be kept at the default value, which is . If, for whatever reason, is not selected, you need to change Method: back to . The method is the name given by SPSS Statistics to standard regression analysis.

Click the button. You will be presented with the Linear Regression: Statistics dialogue box, as shown below:

Published with written permission from SPSS Statistics, IBM Corporation. In addition to the options that are selected by default, select Confidence intervals in the Regression Coefficients area leaving the Level(

Published with written permission from SPSS Statistics, IBM Corporation. Click the button. You will be returned to the Linear Regression dialogue box.

Click the button. This will generate the output.

Join the 10,000s of students, academics and professionals who rely on Laerd Statistics.TAKE THE TOUR PLANS and PRICING SPSS Statisticstop and Interpreting and Reporting the Output of Multiple Regression Analysis SPSS Statistics will generate quite a few tables of output for a multiple regression analysis. In this section, we show you only the three main tables required to understand your results from the multiple regression procedure, assuming that no assumptions have been violated. A complete explanation of the output you have to interpret when checking your data for the eight assumptions required to carry out multiple regression is provided in our enhanced guide. This includes relevant scatterplots and partial regression plots, histogram (with superimposed normal curve), Normal P-P Plot and Normal Q-Q Plot, correlation coefficients and Tolerance/VIF values, casewise diagnostics and studentized deleted residuals.

However, in this "quick start" guide, we focus only on the three main tables you need to understand your multiple regression results, assuming that your data has already met the eight assumptions required for multiple regression to give you a valid result:

Determining how well the model fits The first table of interest is the Model Summary table. This table provides the R, R2, adjusted R2, and the standard error of the estimate, which can be used to determine how

well a regression model fits the data:

Multiple Regression Analysis using SPSS Statistics

Introduction Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable). The variables we are using to predict the value of the dependent variable are called the independent variables (or sometimes, the predictor, explanatory or regressor variables).

For example, you could use multiple regression to understand whether exam performance can be predicted based on revision time, test anxiety, lecture attendance and gender. Alternately, you could use multiple regression to understand whether daily cigarette consumption can be predicted based on smoking duration, age when started smoking, smoker type, income and gender.

Multiple regression also allows you to determine the overall fit (variance explained) of the model and the relative contribution of each of the predictors to the total variance explained. For example, you might want to know how much of the variation in exam performance can be explained by revision time, test anxiety, lecture attendance and gender "as a whole", but also the "relative contribution" of each independent variable in explaining the variance.

This "quick start" guide shows you how to carry out multiple regression using SPSS Statistics, as well as interpret and report the results from this test. However, before we introduce you to this procedure, you need to understand the different assumptions that your data must meet in order for multiple regression to give you a valid result. We discuss these assumptions next.

SPSS Statisticstop and Assumptions When you choose to analyse your data using multiple regression, part of the process involves checking to make sure that the data you want to analyse can actually be analysed using multiple regression. You need to do this because it is only appropriate to use multiple regression if your data "passes" eight assumptions that are required for multiple regression to give you a valid result. In practice, checking for these eight assumptions just adds a little bit more time to your analysis, requiring you to click a few more buttons in SPSS Statistics when performing your analysis, as well as think a little bit more about your data, but it is not a difficult task.

Before we introduce you to these eight assumptions, do not be surprised if, when analysing your own data using SPSS Statistics, one or more of these assumptions is violated (i.e., not met). This is not uncommon when working with real-world data rather than textbook examples, which often only show you how to carry out multiple regression when everything goes well! However, dont worry. Even when your data fails certain assumptions, there is often a solution to overcome this. First, let's take a look at these eight assumptions:

Assumption 1: Your dependent variable should be measured on a continuous scale (i.e., it is either an interval or ratio variable). Examples of variables that meet this criterion include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth. You can learn more about interval and ratio variables in our article: Types of Variable. If your dependent variable was measured on an ordinal scale, you will need to carry out ordinal regression rather than multiple regression. Examples of ordinal variables include Likert items (e.g., a 7-point scale from "strongly agree" through to "strongly disagree"), amongst other ways of ranking categories (e.g., a 3-point scale explaining how much a customer liked a product, ranging from "Not very much" to "Yes, a lot"). You can access our SPSS Statistics guide on ordinal regression here. Assumption 2: You have two or more independent variables, which can be either continuous (i.e., an interval or ratio variable) or categorical (i.e., an ordinal or nominal variable). For examples of continuous and ordinal variables, see the bullet above. Examples of nominal variables include gender (e.g., 2 groups: male and female), ethnicity (e.g., 3 groups: Caucasian, African American and Hispanic), physical activity level (e.g., 4 groups: sedentary, low, moderate and high), profession (e.g., 5 groups: surgeon,

doctor, nurse, dentist, therapist), and so forth. Again, you can learn more about variables in our article: Types of Variable. If one of your independent variables is dichotomous and considered a moderating variable, you might need to run a Dichotomous moderator analysis. Assumption 3: You should have independence of observations (i.e., independence of residuals), which you can easily check using the Durbin-Watson statistic, which is a simple test to run using SPSS Statistics. We explain how to interpret the result of the Durbin-Watson statistic, as well as showing you the SPSS Statistics procedure required, in our enhanced multiple regression guide. Assumption 4: There needs to be a linear relationship between (a) the dependent variable and each of your independent variables, and (b) the dependent variable and the independent variables collectively. Whilst there are a number of ways to check for these linear relationships, we suggest creating scatterplots and partial regression plots using SPSS Statistics, and then visually inspecting these scatterplots and partial regression plots to check for linearity. If the relationship displayed in your scatterplots and partial regression plots are not linear, you will have to either run a non-linear regression analysis or "transform" your data, which you can do using SPSS Statistics. In our enhanced multiple regression guide, we show you how to: (a) create scatterplots and partial regression plots to check for linearity when carrying out multiple regression using SPSS Statistics; (b) interpret different scatterplot and partial regression plot results; and (c) transform your data using SPSS Statistics if you do not have linear relationships between your variables. Assumption 5: Your data needs to show homoscedasticity, which is where the variances along the line of best fit remain similar as you move along the line. We explain more about what this means and how to assess the homoscedasticity of your data in our enhanced multiple regression guide. When you analyse your own data, you will need to plot the studentized residuals against the unstandardized predicted values. In our enhanced multiple regression guide, we explain: (a) how to test for homoscedasticity using SPSS Statistics; (b) some of the things you will need to consider when interpreting your data; and (c) possible ways to continue with your analysis if your data fails to meet this assumption. Assumption 6: Your data must not show multicollinearity, which occurs when you have two or more independent variables that are highly correlated with each other. This leads to problems with understanding which independent variable contributes to the variance explained in the dependent variable, as well as technical issues in calculating a multiple regression model. Therefore, in our enhanced multiple regression guide, we show you: (a) how to use SPSS Statistics to detect for multicollinearity through an inspection of correlation coefficients and Tolerance/VIF values; and (b) how to interpret these correlation coefficients and Tolerance/VIF values so that you can determine whether your data meets or violates this assumption. Assumption 7: There should be no significant outliers, high leverage points or highly influential points. Outliers, leverage and influential points are different terms used to represent observations in your data set that are in some way unusual when you wish to perform a multiple regression analysis. These different classifications of unusual points reflect the different impact they have on the regression line. An observation can be classified as more than one type of unusual point. However, all these points can have a very negative effect on the regression equation that is used to predict the value of the dependent variable based on the independent variables. This can change the output that SPSS Statistics produces and reduce the predictive accuracy of your results as well as the statistical significance. Fortunately, when using SPSS Statistics to run multiple regression on your data, you can detect possible outliers, high leverage points and highly influential points. In our enhanced multiple regression guide, we: (a) show you how to detect outliers using "casewise diagnostics" and "studentized deleted residuals", which you can do using SPSS Statistics, and discuss some of the options you have in order to deal with outliers; (b) check for leverage points using SPSS Statistics and discuss what you should do if you have any; and (c) check for influential points in SPSS Statistics using a measure of influence known as Cook's Distance, before presenting some practical approaches in SPSS Statistics to deal with any influential points you might have. Assumption 8: Finally, you need to check that the residuals (errors) are approximately normally distributed (we explain these terms in our enhanced multiple regression guide). Two common methods to check this assumption include using: (a) a histogram (with a superimposed

normal curve) and a Normal P-P Plot; or (b) a Normal Q-Q Plot of the studentized residuals. Again, in our enhanced multiple regression guide, we: (a) show you how to check this assumption using SPSS Statistics, whether you use a histogram (with superimposed normal curve) and Normal P-P Plot, or Normal Q-Q Plot; (b) explain how to interpret these diagrams; and (c) provide a possible solution if your data fails to meet this assumption. You can check assumptions 3, 4, 5, 6, 7 and 8 using SPSS Statistics. Assumptions 1 and 2 should be checked first, before moving onto assumptions 3, 4, 5, 6, 7 and 8. Just remember that if you do not run the statistical tests on these assumptions correctly, the results you get when running multiple regression might not be valid. This is why we dedicate a number of sections of our enhanced multiple regression guide to help you get this right. You can find out about our enhanced content as a whole here, or more specifically, learn how we help with testing assumptions here.

In the section, Procedure, we illustrate the SPSS Statistics procedure to perform a multiple regression assuming that no assumptions have been violated. First, we introduce the example that is used in this guide.

TAKE THE TOUR

PLANS and PRICING SPSS Statisticstop and Example A health researcher wants to be able to predict "VO2max", an indicator of fitness and health. Normally, to perform this procedure requires expensive laboratory equipment and necessitates that an individual exercise to their maximum (i.e., until they can longer continue exercising due to physical exhaustion). This can put off those individuals who are not very active/fit and those individuals who might be at higher risk of ill health (e.g., older unfit subjects). For these reasons, it has been desirable to find a way of predicting an individual's VO2max based on attributes that can be measured more easily and cheaply. To this end, a researcher recruited 100 participants to perform a maximum VO2max test, but also recorded their "age", "weight", "heart rate" and "gender". Heart rate is the average of the last 5 minutes of a 20 minute, much easier, lower workload cycling test. The researcher's goal is to be able to predict VO2max based on these four attributes: age, weight, heart rate and gender.

SPSS Statisticstop and Setup in SPSS Statistics In SPSS Statistics, we created six variables: (1) VO2max, which is the maximal aerobic capacity; (2) age, which is the participant's age; (3) weight, which is the participant's weight (technically, it is their 'mass'); (4) heart_rate, which is the participant's heart rate; (5) gender, which is the participant's gender; and (6) caseno, which is the case number. The caseno variable is used to make it easy for you to eliminate cases (e.g., "significant outliers", "high leverage points" and "highly influential points") that you have identified when checking for assumptions. In our enhanced multiple regression guide, we show you how to correctly enter data in SPSS Statistics to run a multiple regression when you are also checking for assumptions. You can learn about our enhanced data setup content here. Alternately, we have a generic, "quick start" guide to show you how to enter data into SPSS Statistics, available here.

SPSS Statisticstop and Test Procedure in SPSS Statistics The seven steps below show you how to analyse your data using multiple regression in SPSS Statistics when none of the eight assumptions in the previous section, Assumptions, have been violated. At the end of these seven steps, we show you how to interpret the results from your multiple regression. If you are looking for help to make sure your data meets assumptions 3, 4, 5, 6, 7 and 8, which are required when using multiple regression and can be tested using SPSS Statistics, you can learn more in our enhanced guide here.

Click Analyze ¿ Regression ¿ Linear... on the main menu, as shown below:

Published with written permission from SPSS Statistics, IBM Corporation. Note: Don't worry that you're selecting Analyze ¿ Regression ¿ Linear... on the main menu or that the dialogue boxes in the steps that follow have the title, Linear Regression. You have not made a mistake. You are in the correct place to carry out the multiple regression procedure. This is just the title that SPSS Statistics gives, even when running a multiple regression procedure.

You will be presented with the Linear Regression dialogue box below:

Published with written permission from SPSS Statistics, IBM Corporation. Transfer the dependent variable,

VO2max, into the Dependent: box and the independent variables, age, weight, heart_rate and gender into the Independent(s): box, using the buttons, as shown below (all other boxes can be ignored):

Published with written permission from SPSS Statistics, IBM Corporation. Note: For a standard multiple regression you should ignore the and buttons as they are for sequential (hierarchical) multiple regression. The Method: option needs to be kept at the default value, which is . If, for whatever reason, is not selected, you need to change Method: back to . The method is the name given by SPSS Statistics to standard regression analysis.

Click the button. You will be presented with the Linear Regression: Statistics dialogue box, as shown below:

Published with written permission from SPSS Statistics, IBM Corporation. In addition to the options that are selected by default, select Confidence intervals in the Regression Coefficients area leaving the Level(

Published with written permission from SPSS Statistics, IBM Corporation. Click the button. You will be returned to the Linear Regression dialogue box.

Click the button. This will generate the output.

Join the 10,000s of students, academics and professionals who rely on Laerd Statistics.TAKE THE TOUR PLANS and PRICING SPSS Statisticstop and Interpreting and Reporting the Output of Multiple Regression Analysis SPSS Statistics will generate quite a few tables of output for a multiple regression analysis. In this section, we show you only the three main tables required to understand your results from the multiple regression procedure, assuming that no assumptions have been violated. A complete explanation of the output you have to interpret when checking your data for the eight assumptions required to carry out multiple regression is provided in our enhanced guide. This includes relevant scatterplots and partial regression plots, histogram (with superimposed normal curve), Normal P-P Plot and Normal Q-Q Plot, correlation coefficients and Tolerance/VIF values, casewise diagnostics and studentized deleted residuals.

However, in this "quick start" guide, we focus only on the three main tables you need to understand your multiple regression results, assuming that your data has already met the eight assumptions required for multiple regression to give you a valid result:

Determining how well the model fits The first table of interest is the Model Summary table. This table provides the R, R2, adjusted R2, and the standard error of the estimate, which can be used to determine how well a regression model fits the data:

Published with written permission from SPSS Statistics, IBM Corporation.

The "R" column represents the value of R, the multiple correlation coefficient. R can be considered to be one measure of the quality of the prediction of the dependent variable; in this case, VO2max. A value of 0.760, in this example, indicates a good level of prediction. The "R Square" column represents the R2 value (also called the coefficient of determination), which is the proportion of variance in the dependent variable that can be explained by the independent variables (technically, it is the proportion of variation accounted for by the regression model above and beyond the mean model). You can see from our value of 0.577 that our independent variables explain 57.7

Statistical significance The F-ratio in the ANOVA table (see below) tests whether the overall regression model is a good fit for the data. The table shows that the independent variables statistically significantly predict the dependent variable, $F(4, 95) = 32.393$, p ¡ .0005 (i.e., the regression model is a good fit of the data).

Published with written permission from SPSS Statistics, IBM Corporation.

Estimated model coefficients The general form of the equation to predict VO2max from age, weight, heart_rate, gender, is:

predicted VO2max = 87.83 (0.165 x age) (0.385 x weight) (0.118 x heart_rate) + (13.208 x gender)

This is obtained from the Coefficients table, as shown below:

Published with written permission from SPSS Statistics, IBM Corporation. Unstandardized coefficients indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant. Consider the effect of age in this example. The unstandardized coefficient, B1, for age is equal to -0.165 (see Coefficients table). This means that for each one year increase in age, there is a decrease in VO2max of 0.165 ml/min/kg.

Statistical significance of the independent variables You can test for the statistical significance of each of the independent variables. This tests whether the unstandardized (or standardized) coefficients are equal to 0 (zero) in the population. If p ¡ .05, you can conclude that the coefficients are statistically significantly different to 0 (zero). The t-value and corresponding p-value are located in the "t" and "Sig." columns, respectively, as highlighted below: Laerd StatisticsLoginCookies and Privacy Take the Tour Plans and Pricing SIGN UP Multiple Regression Analysis using SPSS Statistics

Introduction Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable). The variables we are using to predict the value of the dependent variable are called the independent variables (or sometimes, the predictor, explanatory or regressor variables).

For example, you could use multiple regression to understand whether exam performance can be predicted based on revision time, test anxiety, lecture attendance and gender. Alternately, you could use multiple regression to understand whether daily cigarette consumption can be predicted based on smoking duration, age when started smoking, smoker type, income and gender.

Multiple regression also allows you to determine the overall fit (variance explained) of the model and the relative contribution of each of the predictors to the total variance explained. For example, you might want to know how much of the variation in exam performance can be explained by revision time, test anxiety, lecture attendance and gender "as a whole", but also the "relative contribution" of each independent variable in explaining the variance.

This "quick start" guide shows you how to carry out multiple regression using SPSS Statistics, as well as interpret and report the results from this test. However, before we introduce you to this procedure, you need to understand the different assumptions that your data must meet in order for multiple regression to give you a valid result. We discuss these assumptions next.

SPSS Statisticstop and Assumptions When you choose to analyse your data using multiple regression, part of the process involves checking to make sure that the data you want to analyse can actually be analysed using multiple regression. You need to do this because it is only appropriate to use multiple regression if your data "passes" eight assumptions that are required for multiple regression to give you a valid result. In practice, checking for these eight assumptions just adds a little bit more time to your analysis, requiring you to click a few more buttons in SPSS Statistics when performing your analysis, as well as think a little bit more about your data, but it is not a difficult task.

Before we introduce you to these eight assumptions, do not be surprised if, when analysing your own data using SPSS Statistics, one or more of these assumptions is violated (i.e., not met). This is not uncommon when working with real-world data rather than textbook examples, which often only show you how to carry out multiple regression when everything goes well! However, dont worry. Even when your data fails certain assumptions, there is often a solution to overcome this. First, let's take a look at these eight assumptions:

Assumption 1: Your dependent variable should be measured on a continuous scale (i.e., it is either an interval or ratio variable). Examples of variables that meet this criterion include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth. You can learn more about interval and ratio variables in our article: Types of Variable. If your dependent variable was measured on an ordinal scale, you will need to carry out ordinal regression rather

than multiple regression. Examples of ordinal variables include Likert items (e.g., a 7-point scale from "strongly agree" through to "strongly disagree"), amongst other ways of ranking categories (e.g., a 3-point scale explaining how much a customer liked a product, ranging from "Not very much" to "Yes, a lot"). You can access our SPSS Statistics guide on ordinal regression here. Assumption 2: You have two or more independent variables, which can be either continuous (i.e., an interval or ratio variable) or categorical (i.e., an ordinal or nominal variable). For examples of continuous and ordinal variables, see the bullet above. Examples of nominal variables include gender (e.g., 2 groups: male and female), ethnicity (e.g., 3 groups: Caucasian, African American and Hispanic), physical activity level (e.g., 4 groups: sedentary, low, moderate and high), profession (e.g., 5 groups: surgeon, doctor, nurse, dentist, therapist), and so forth. Again, you can learn more about variables in our article: Types of Variable. If one of your independent variables is dichotomous and considered a moderating variable, you might need to run a Dichotomous moderator analysis. Assumption 3: You should have independence of observations (i.e., independence of residuals), which you can easily check using the Durbin-Watson statistic, which is a simple test to run using SPSS Statistics. We explain how to interpret the result of the Durbin-Watson statistic, as well as showing you the SPSS Statistics procedure required, in our enhanced multiple regression guide. Assumption 4: There needs to be a linear relationship between (a) the dependent variable and each of your independent variables, and (b) the dependent variable and the independent variables collectively. Whilst there are a number of ways to check for these linear relationships, we suggest creating scatterplots and partial regression plots using SPSS Statistics, and then visually inspecting these scatterplots and partial regression plots to check for linearity. If the relationship displayed in your scatterplots and partial regression plots are not linear, you will have to either run a non-linear regression analysis or "transform" your data, which you can do using SPSS Statistics. In our enhanced multiple regression guide, we show you how to: (a) create scatterplots and partial regression plots to check for linearity when carrying out multiple regression using SPSS Statistics; (b) interpret different scatterplot and partial regression plot results; and (c) transform your data using SPSS Statistics if you do not have linear relationships between your variables. Assumption 5: Your data needs to show homoscedasticity, which is where the variances along the line of best fit remain similar as you move along the line. We explain more about what this means and how to assess the homoscedasticity of your data in our enhanced multiple regression guide. When you analyse your own data, you will need to plot the studentized residuals against the unstandardized predicted values. In our enhanced multiple regression guide, we explain: (a) how to test for homoscedasticity using SPSS Statistics; (b) some of the things you will need to consider when interpreting your data; and (c) possible ways to continue with your analysis if your data fails to meet this assumption. Assumption 6: Your data must not show multicollinearity, which occurs when you have two or more independent variables that are highly correlated with each other. This leads to problems with understanding which independent variable contributes to the variance explained in the dependent variable, as well as technical issues in calculating a multiple regression model. Therefore, in our enhanced multiple regression guide, we show you: (a) how to use SPSS Statistics to detect for multicollinearity through an inspection of correlation coefficients and Tolerance/VIF values; and (b) how to interpret these correlation coefficients and Tolerance/VIF values so that you can determine whether your data meets or violates this assumption. Assumption 7: There should be no significant outliers, high leverage points or highly influential points. Outliers, leverage and influential points are different terms used to represent observations in your data set that are in some way unusual when you wish to perform a multiple regression analysis. These different classifications of unusual points reflect the different impact they have on the regression line. An observation can be classified as more than one type of unusual point. However, all these points can have a very negative effect on the regression equation that is used to predict the value of the dependent variable based on the independent variables. This can change the output that SPSS Statistics produces and reduce the predictive accuracy of your results as well as the statistical significance. Fortunately, when using SPSS Statistics to run multiple regression on your data, you can detect possible outliers, high leverage points and highly influential

points. In our enhanced multiple regression guide, we: (a) show you how to detect outliers using "casewise diagnostics" and "studentized deleted residuals", which you can do using SPSS Statistics, and discuss some of the options you have in order to deal with outliers; (b) check for leverage points using SPSS Statistics and discuss what you should do if you have any; and (c) check for influential points in SPSS Statistics using a measure of influence known as Cook's Distance, before presenting some practical approaches in SPSS Statistics to deal with any influential points you might have. Assumption 8: Finally, you need to check that the residuals (errors) are approximately normally distributed (we explain these terms in our enhanced multiple regression guide). Two common methods to check this assumption include using: (a) a histogram (with a superimposed normal curve) and a Normal P-P Plot; or (b) a Normal Q-Q Plot of the studentized residuals. Again, in our enhanced multiple regression guide, we: (a) show you how to check this assumption using SPSS Statistics, whether you use a histogram (with superimposed normal curve) and Normal P-P Plot, or Normal Q-Q Plot; (b) explain how to interpret these diagrams; and (c) provide a possible solution if your data fails to meet this assumption. You can check assumptions 3, 4, 5, 6, 7 and 8 using SPSS Statistics. Assumptions 1 and 2 should be checked first, before moving onto assumptions 3, 4, 5, 6, 7 and 8. Just remember that if you do not run the statistical tests on these assumptions correctly, the results you get when running multiple regression might not be valid. This is why we dedicate a number of sections of our enhanced multiple regression guide to help you get this right. You can find out about our enhanced content as a whole here, or more specifically, learn how we help with testing assumptions here.

In the section, Procedure, we illustrate the SPSS Statistics procedure to perform a multiple regression assuming that no assumptions have been violated. First, we introduce the example that is used in this guide.

TAKE THE TOUR

PLANS and PRICING SPSS Statisticstop and Example A health researcher wants to be able to predict "VO2max", an indicator of fitness and health. Normally, to perform this procedure requires expensive laboratory equipment and necessitates that an individual exercise to their maximum (i.e., until they can longer continue exercising due to physical exhaustion). This can put off those individuals who are not very active/fit and those individuals who might be at higher risk of ill health (e.g., older unfit subjects). For these reasons, it has been desirable to find a way of predicting an individual's VO2max based on attributes that can be measured more easily and cheaply. To this end, a researcher recruited 100 participants to perform a maximum VO2max test, but also recorded their "age", "weight", "heart rate" and "gender". Heart rate is the average of the last 5 minutes of a 20 minute, much easier, lower workload cycling test. The researcher's goal is to be able to predict VO2max based on these four attributes: age, weight, heart rate and gender.

SPSS Statisticstop and Setup in SPSS Statistics In SPSS Statistics, we created six variables: (1) VO2max, which is the maximal aerobic capacity; (2) age, which is the participant's age; (3) weight, which is the participant's weight (technically, it is their 'mass'); (4) heart_rate, which is the participant's heart rate; (5) gender, which is the participant's gender; and (6) caseno, which is the case number. The caseno variable is used to make it easy for you to eliminate cases (e.g., "significant outliers", "high leverage points" and "highly influential points") that you have identified when checking for assumptions. In our enhanced multiple regression guide, we show you how to correctly enter data in SPSS Statistics to run a multiple regression when you are also checking for assumptions. You can learn about our enhanced data setup content here. Alternately, we have a generic, "quick start" guide to show you how to enter data into SPSS Statistics, available here.

0.3 Test Procedure in SPSS Statistics

The seven steps below show you how to analyse your data using multiple regression in SPSS Statistics when none of the eight assumptions in the previous section, Assumptions, have been violated. At the end of these seven steps, we show you how to interpret the results from your multiple regression. If you are looking for

help to make sure your data meets assumptions 3, 4, 5, 6, 7 and 8, which are required when using multiple regression and can be tested using SPSS Statistics, you can learn more in our enhanced guide here.

- Step 1 Click Analyze ¿ Regression ¿ Linear... on the main menu, as shown below:

  Published with written permission from SPSS Statistics, IBM Corporation. Note: Don't worry that you're selecting Analyze ¿ Regression ¿ Linear... on the main menu or that the dialogue boxes in the steps that follow have the title, Linear Regression. You have not made a mistake. You are in the correct place to carry out the multiple regression procedure. This is just the title that SPSS Statistics gives, even when running a multiple regression procedure.

- Step 2 You will be presented with the Linear Regression dialogue box below:

  Transfer the dependent variable, VO2max, into the Dependent: box and the independent variables, age, weight, heart_rate and gender into the Independent(s): box, using the buttons, as shown below (all other boxes can be ignored):

  Note: For a standard multiple regression you should ignore the and buttons as they are for sequential (hierarchical) multiple regression. The Method: option needs to be kept at the default value, which is . If, for whatever reason, is not selected, you need to change Method: back to . The method is the name given by SPSS Statistics to standard regression analysis.

- Step 3 Click the button. You will be presented with the Linear Regression: Statistics dialogue box, as shown below:

  In addition to the options that are selected by default, select Confidence intervals in the Regression Coefficients area leaving the Level(%): option at "95". You will end up with the following screen:

- Step 4 Click the button. You will be returned to the Linear Regression dialogue box.

  Click the button. This will generate the output.

0.4   Interpreting and Reporting the Output of Multiple Regression Analysis

SPSS Statistics will generate quite a few tables of output for a multiple regression analysis. In this section, we show you only the three main tables required to understand your results from the multiple regression procedure, assuming that no assumptions have been violated. A complete explanation of the output you have to interpret when checking your data for the eight assumptions required to carry out multiple regression is provided in our enhanced guide. This includes relevant scatterplots and partial regression plots, histogram (with superimposed normal curve), Normal P-P Plot and Normal Q-Q Plot, correlation coefficients and Tolerance/VIF values, casewise diagnostics and studentized deleted residuals.

0.5   Determining how well the model fits

The first table of interest is the Model Summary table. This table provides the R, R2, adjusted R2, and the standard error of the estimate, which can be used to determine how well a regression model fits the data:

The "R" column represents the value of R, the multiple correlation coefficient. R can be considered to be one measure of the quality of the prediction of the dependent variable; in this case, VO2max. A value of 0.760, in this example, indicates a good level of prediction. The "R Square" column represents the R^2 value (also called the coefficient of determination), which is the proportion of variance in the dependent variable that can be explained by the independent variables (technically, it is the proportion of variation accounted for

by the regression model above and beyond the mean model). You can see from our value of 0.577 that our independent variables explain 57.7% of the variability of our dependent variable, VO2max. However, you also need to be able to interpret "Adjusted R Square" (adj. R2) to accurately report your data. We explain the reasons for this, as well as the output, in our enhanced multiple regression guide. Statistical significance The F-ratio in the ANOVA table (see below) tests whether the overall regression model is a good fit for the data. The table shows that the independent variables statistically significantly predict the dependent variable, $F_{(4, 95)} = 32.393$, p ¡ .0005 (i.e., the regression model is a good fit of the data).

Estimated model coefficients The general form of the equation to predict VO2max from age, weight, heart_rate, gender, is:

predicted VO2max = 87.83 (0.165 x age) (0.385 x weight) (0.118 x heart_rate) + (13.208 x gender)

This is obtained from the Coefficients table, as shown below:

Unstandardized coefficients indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant. Consider the effect of age in this example. The unstandardized coefficient, B1, for age is equal to -0.165 (see Coefficients table). This means that for each one year increase in age, there is a decrease in VO2max of 0.165 ml/min/kg.

0.6   Statistical significance of the independent variables

You can test for the statistical significance of each of the independent variables. This tests whether the unstandardized (or standardized) coefficients are equal to 0 (zero) in the population. If p ¡ .05, you can conclude that the coefficients are statistically significantly different to 0 (zero). The t-value and corresponding p-value are located in the "t" and "Sig." columns, respectively, as highlighted below:

You can see from the "Sig." column that all independent variable coefficients are statistically significantly different from 0 (zero). Although the intercept, B0, is tested for statistical significance, this is rarely an important or interesting finding.

0.7   Putting it all together

You could write up the results as follows:

General A multiple regression was run to predict VO2max from gender, age, weight and heart rate. These variables statistically significantly predicted VO2max, $F_{(4, 95)} = 32.393$, p ¡ .0005, R2 = .577. All four variables added statistically significantly to the prediction, p ¡ .05.

If you are unsure how to interpret regression equations or how to use them to make predictions, we discuss this in our enhanced multiple regression guide. We also show you how to write up the results from your assumptions tests and multiple regression output if you need to report this in a dissertation/thesis, assignment or research report. We do this using the Harvard and APA styles. You can learn more about our enhanced content here. You can see from the "Sig." column that all independent variable coefficients are statistically significantly different from 0 (zero). Although the intercept, B0, is tested for statistical significance, this is rarely an important or interesting finding.

Putting it all together You could write up the results as follows:

General A multiple regression was run to predict VO2max from gender, age, weight and heart rate. These variables statistically significantly predicted VO2max, $F_{(4, 95)} = 32.393$, p ¡ .0005, R2 = .577. All four variables added statistically significantly to the prediction, p ¡ .05.

If you are unsure how to interpret regression equations or how to use them to make predictions, we discuss this in our enhanced multiple regression guide. We also show you how to write up the results from your assumptions tests and multiple regression output if you need to report this in a dissertation/thesis, assignment

or research report. We do this using the Harvard and APA styles. You can learn more about our enhanced content here.

Join the 10,000s of students, academics and professionals who rely on Laerd Statistics.TAKE THE TOUR PLANS and PRICING 1 Home About Us Contact Us Terms and Conditions Privacy and Cookies 2013 Lund Research Ltd

The "R" column represents the value of R, the multiple correlation coefficient. R can be considered to be one measure of the quality of the prediction of the dependent variable; in this case, VO2max. A value of 0.760, in this example, indicates a good level of prediction. The "R Square" column represents the R2 value (also called the coefficient of determination), which is the proportion of variance in the dependent variable that can be explained by the independent variables (technically, it is the proportion of variation accounted for by the regression model above and beyond the mean model). You can see from our value of 0.577 that our independent variables explain 57.7% of the variability of our dependent variable, VO2max. However, you also need to be able to interpret "Adjusted R Square" (adj. R2) to accurately report your data. We explain the reasons for this, as well as the output, in our enhanced multiple regression guide.

Statistical significance The F-ratio in the ANOVA table (see below) tests whether the overall regression model is a good fit for the data. The table shows that the independent variables statistically significantly predict the dependent variable, $F_{(4, 95)} = 32.393$, p ¡ .0005 (i.e., the regression model is a good fit of the data).

Published with written permission from SPSS Statistics, IBM Corporation.

Estimated model coefficients The general form of the equation to predict VO2max from age, weight, heart_rate, gender, is:

predicted VO2max = 87.83 (0.165 x age) (0.385 x weight) (0.118 x heart_rate) + (13.208 x gender)

This is obtained from the Coefficients table, as shown below:

Published with written permission from SPSS Statistics, IBM Corporation. Unstandardized coefficients indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant. Consider the effect of age in this example. The unstandardized coefficient, B1, for age is equal to -0.165 (see Coefficients table). This means that for each one year increase in age, there is a decrease in VO2max of 0.165 ml/min/kg.

Statistical significance of the independent variables You can test for the statistical significance of each of the independent variables. This tests whether the unstandardized (or standardized) coefficients are equal to 0 (zero) in the population. If p ¡ .05, you can conclude that the coefficients are statistically significantly different to 0 (zero). The t-value and corresponding p-value are located in the "t" and "Sig." columns, respectively, as highlighted below:

You can see from the "Sig." column that all independent variable coefficients are statistically significantly different from 0 (zero). Although the intercept, B0, is tested for statistical significance, this is rarely an important or interesting finding.

Putting it all together You could write up the results as follows:

General A multiple regression was run to predict VO2max from gender, age, weight and heart rate. These variables statistically significantly predicted VO2max, $F_{(4, 95)} = 32.393$, p ¡ .0005, R2 = .577. All four variables added statistically significantly to the prediction, p ¡ .05.

If you are unsure how to interpret regression equations or how to use them to make predictions, we discuss this in our enhanced multiple regression guide. We also show you how to write up the results from your assumptions tests and multiple regression output if you need to report this in a dissertation/thesis, assignment or research report. We do this using the Harvard and APA styles. You can learn more about our enhanced content here.