

# Tool Mastery | Linear Regression

ARTICLE OPTIONS



**Ozzie**  
Alteryx

11-03-2016 03:45 PM



***This article is part of the Tool Mastery Series, a compilation of Knowledge Base contributions to introduce diverse working examples for Designer Tools. Here we'll delve into uses of the Linear Regression Tool on our way to mastering the Alteryx Designer:***

*If you're using the Linear Regression Tool in v11.0+, be sure to familiarize yourself with the tool's redesign!*



Linear regression is a statistical approach that seeks to model the relationship between a dependent (target) variable and one or more predictor variables. It is one of the oldest forms of regression and its applications throughout history have been endless for modeling all kinds of phenomena. In linear regression, a line of best fit is calculated using the least squares method. This linear equation is then used to calculate projected values for the target variable given a set of new values for the predictor variables.

In Alteryx we have a linear regression tool that is actually an R based macro. It may look intimidating at first but it is actually very easy to configure. You simply select the target variable as well as the predictor variables in your dataset. You can also select whether or not you want to omit a model constant and/or use a weight variable for weighted least squares. After you have run your regression, two things will be outputted: The model object and a results report, which goes into statistical details of the generated least squares regression line.



The model object is outputted from the R tool inside the linear regression macro. This model object can be used as an input for three different tools in the predictive category (Nested Test, Score, and Stepwise).

- The Nested Test Tool does a model comparison to see if removing and adding variables to another Linear Regression tool significantly changes the predictive capabilities of the model.
- The Score Tool uses the regression line obtained from the Linear Regression tool and calculates the new values for the target variables based on the values from the predictor variables being fed into the tool.
- The Stepwise Tool is for determining the optimal predictor variables to include in your model out of all potential predictor variables.





SIGN IN

As mentioned above, the results output gives you a summary of how your model performed. For the sake of length here, we will not go into major detail of each and every component of this summary. If you still need further explanation of the statistics posted in the report we recommend seeking out additional resources to expand your knowledge (courses, Google, YouTube, [Alteryx's Udacity class](#), etc...) In the report we see:

- The numbers from a [boxplot of the residuals](#)
- Estimated model coefficients as well as their associated [P, t, and standard error values](#).
- RSE, [R-Squared](#), [F-statistic](#) for the regression line
- [Type II ANOVA Analysis](#) for the variables

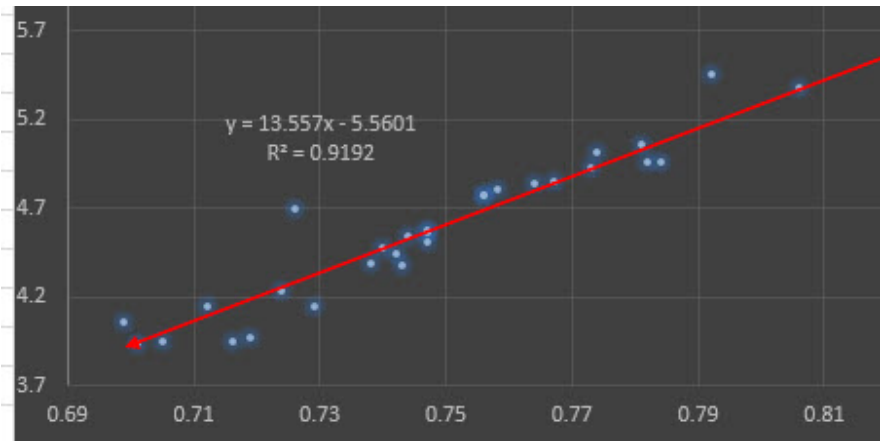
Now that we know what to expect from this tool, let's go into different ways we can utilize the Linear Regression Tool.

*Note: Before performing any predictive analysis it is imperative that the user be familiarized with the data they are feeding into the predictive tools. Data Investigation tools such as the [Field Summary Tool](#) make it easier to see what the data looks like to see if there are missing values, nulls, or a bunch of zeros that might cause the tools to error out. See [Troubleshooting the Predictive Tools](#) if you are getting errors from this tool.*

- **Simple linear regression**



This is for bivariate data (one predictor variable and one target variable). For Excel users that are used to highlighting two fields and making a scatterplot or using the regression tool analysis in Excel to obtain your regression line, this should look familiar to you.



In this case, instead of seeing the equation in  $y=mx+b$  format, your slope (predictor variable coefficient) and intercept are going to appear here in the Results Output.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-5.56	0.5707	-9.743	1.70e-10 ***
OPS	13.56	0.7599	17.842	< 2.2e-16 ***



Tips: The assumptions you're making before running this tool is that your variables have some correlation and that a change in the input variable can be expected to cause a uniform change in the output variable. It might be a good idea to perform an [Association Analysis](#) to see if this is the case; if the two variables are not correlated at all then linear regression might not hold any value. Also, ask yourself whether one predictor variable is enough to accurately predict the target variable.

- **Multiple Linear Regression**

This when you are feeding in more than one predictor variable in your linear regression. All that changes here in terms of the output will be a longer list of coefficients for your variables, as well as the list for Type II ANOVA analysis. Unlike simple linear regression, there are a lot more things to watch out for in multiple linear regression.

Tips: Adjusted  $R^2$  is more important than  $R^2$ ; the more variables you throw into the inputs without substantively improving the  $R^2$ , the smaller the adjusted  $R^2$  gets. Also be mindful of [multicollinearity](#) with the predictor variables you choose. A simplified model can often do better than one with too many input variables.



- **Exponential Regression and Power Regression**

The great thing is we can also use the linear regression tool to perform exponential regression and power regression. So for all those relationships that are more suited to be fitted with an exponential curve, like bank balance after compound interest (exponential growth), drug in your body over time (exponential decay) or even as something as complex as the [initial mass function](#) of stars (power law), we can still perform linear regression. How's that, you may ask? Well, we can transform the target variables and/or predictor variables to make the relationships become linear.

For [exponential relationships](#), we can just simply take the natural log of your target variable in the formula tool before the Linear Regression tool. What is great is that the Score tool has an option that asks whether or not the target variables were transformed via the natural log. This means that you won't have to do raise the outputs to the power of  $e$  after the Score Tool.

For [power regression](#) we take the natural log of both the target variable and predictor variable.

Hopefully this article has given you the courage and power to try linear regression on your data. If you're still feeling uneasy or are more of a visual learner like myself, check out these two resources:

- A [sample tutorial](#) from our YouTube channel.
- [Predictive analytics starter kit](#) from the Alteryx gallery, as it provides you with a full walkthrough of an example of using linear regression.

If you enjoyed this topic and want to learn more about predictive capabilities within Alteryx, check out the [One Stop Shop for Predictive Resources](#) or our Community [Live Training](#) video on [Regression Modeling](#)!

***By now, you should have expert-level proficiency with the Linear Regression Tool! If you can think of a use case we left out, feel free to use the comments section below! Consider yourself a Tool Master already? Let us know at [Community@alteryx.com](mailto:Community@alteryx.com) if you'd like your creative tool uses to be featured in the Tool Mastery Series.***

***Stay tuned with our latest posts every [Tool Tuesday](#) by following [Alteryx](#) on Twitter! If you want to master all the Designer tools, consider [subscribing](#) for email notifications.***



SIGN IN

COMMENTS



fredrick\_berning  
Atom

11-05-2016 12:04 PM



Is it possible to adjust the precision of Coefficients?



0 STARS



Ozzie  
Alteryx

11-07-2016 08:44 AM



Hi Fredrick,

You could adjust the precision but that would involve going into the linear regression macro and changing some of the R code.

-Ozzie



0 STARS



Using Alteryx Designer 32-bit?  
Click for details about future support

Recommendations

Linear Regression | Alteryx 11.0

Tool Mastery | Association Analysis

Tool Mastery | Join

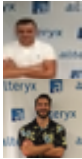
Tool Mastery | RegEx



Contributors

SIGN IN

 Ozzie



MattD

