White paper

Multiple Regression

A tool to determine the impact in analysing the effectiveness of advertising spend.

adalyser.com
@adalyser

+44 (0) 333 666 7366

The complete TV advertising platform

Adalyser®

Introduction

In order to establish if the advertising mechanisms a company employ have an impact on the web hits or sales that a company may have, an analysis that correlates advertising spend (independent variable) with a sales (dependant variable) is useful.

For this we use regression analysis where we see how well a straight line fits through the data, indicating how related the two measures are.

As well as understanding the relationship between spend and response it is also important to understand if a significant amount of the response is explained by the independent variable (spend).

For this, we examine the relationship by analysing the variance in the data.

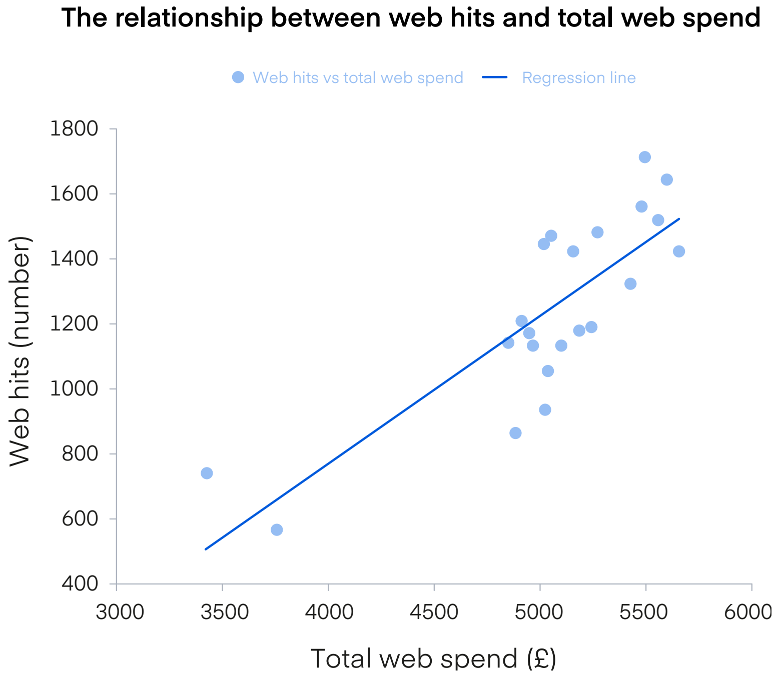
If a significant correlation is found, we can then use the model by which we performed the regression to predict the change in response that would result from a unit change in spend.

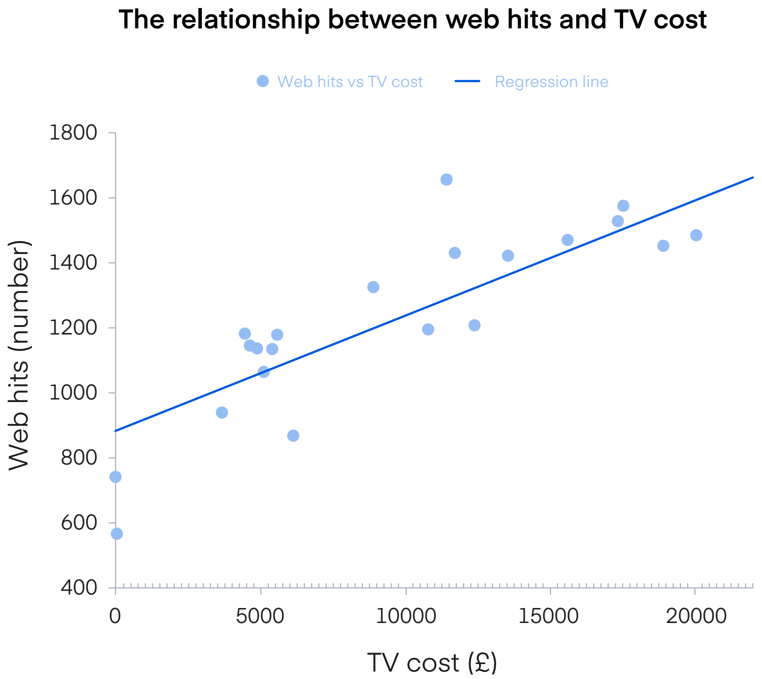
This document discusses multiple regression using a dataset consisting of a single dependent variable (web hits) and two independent variables (web spend and TV spend).

If you have multiple independent variables, the most straightforward approach is to look at the effect of each independent variable on the dependant variable separately and then look at the effect both independent variables have together.

The value of doing the separate analyses is that it contextualises the combined analysis that takes both into account.

Firstly we plot the numbers on a simple scatter chart to see what kind of correlation we have. The dependant measure (in this case web hits) is always plotted on the Y axis.





A regression line is then drawn to best fit the data. A positive correlation means that both variables rise together. A negative correlation means that as one rises, the other falls. Both of these are positive meaning that the more your spend the more web hits you get which is intuitive but not very useful. Lets look at the regression figures.

Web spend alone

The Pearson correlation coefficient between web spend and web hits is 0.778. A value of 1.0 would mean a perfect correlation and a value of 0.0 would mean no correlation at all. The square of the correlation coefficient (R Square) explains the proportion of variation of one variable 'explained' by the other. So in the case of web spend, 0.778 squared is 0.6053. In other words there is a 60.53% dependence of web hits on web spend.

We now need to know if the correlation between web hits and web spend is a significant one; i.e. is one significantly contingent on the other?

In order to test this, an analysis of the variance (ANOVA) is carried out.

Tabel 1: ANOVA (Web spend only)

Model	Sum of squares	df	Mean square	F	Sig (p)
Regression	1999743.427	1	1999743.427	44.399	0.000
Residual	**1306175.993**	**29**	**45040.551**		
Total	3305919.419	30			

In table 1 we can see that the significance (statistically called "the p value") is 0.000. The lower the p value the more significant the result. If $p < 0.05$ we can accept that these two variables are related.

A significant value in the ANOVA means that we can believe what the regression analysis is telling us (whether there was a strong or weak regression) i.e. the method by which we can use our current data to predict outcomes in web hits if we spend £x on the web.

The second part of the ANOVA analysis (table 2) provides the necessary values to create a regression model equation.

The web only model

The coefficients table (table 2) gives us the values we need in order to write the regression equation:

Web hits = Slope × independent variable + intercept or y = MX+C The equation of a line in geometry

The intercept (c) is the value at the intersection and is -704.415 (the value of web hits if spend is 0). The slope (mx) is 4.696 (in the column labelled B for Web spend) and the unit is hits / £. This means for every unit change in TV spend (£) 4.696 web hits hits are achieved.

Conversely, a one unit change in web hits would cost an increase of 21p in web spend.

Tabel 2: Coefficients (Web spend only)

Model	B (Intercept)	Unstandardised co-efficients Std Error	Standardised co-efficicnets Beta	F	Sig (p)
Constant	-704.415	359.24		1.961	0.060
Web spent	4.696	0.705	0.778	6.663	0.000

Testing the Web only model

The regression equation can be used to test how good your model is. For example, if an TV ad cost £300, how many web hits might you get?

Web hits = 0.489 * 300 + (1356.15)

Web hits = 1502.85 for a £300 TV spend

This can be compared with the observed data to see what the margin of error exists. The more data points you have in order to do your regression, the stronger the model will be.

TV spend alone

TV spend is based on a CPT (Cost per Thousand) of £3.

Tabel 3: ANOVA (TV spend only)

Model	Sum of squares	df	Mean square	F	Sig (p)
Regression	2076193.626	1	2076193.626	48.962	0.000
Residual	**1229725.794**	**29**	**42404.338**		
Total	3305919.419	30			

Tabel 4: Coefficients (TV spend only)

Model	B (Intercept)	Unstandardised co-efficients Std Error	Standardised co-efficicnets Beta	F	Sig (p)
Constant	1356.154	58.774		23.074	0.000
TV spent	**0.489**	**0.70**	**0.792**	**6.997**	**0.000**

The Pearson Correlation between web hits and TV spend is 0.792 so it would seem that web hits are a better relation (closer to 1) to TV spend than web spend. The R square is 0.628 meaning that there is a 62.8% dependence of web hits on TV spend. To test the significance ANOVA is carried out.

As the p value (Sig) is < 0.05, it can be accepted that the relationship between TV spend and web hits is significant.

The TV only model

With TV alone the slope is 0.489 hits per £ spent on TV adverts. The intercept (c) is the value at the intersection and is -1356.15 (the value of web hits if spend is 0). The slope (mx) is 0.489 (in the column labelled B for TV spend) and the unit is hits / £. This means for every unit change in TV spend (£) 0.489 hits are achieved. Multiplied to sensible numbers, for every £10 spent, 4.89 web hits are achieved.

Conversely, one unit change in web hits would require a £2.05 spend on TV advertising.

Testing the TV only model

The regression equation can be used to test how good your model is. For example, if an TV ad cost £300, how many web hits might you get?

Web hits = 0.489 * 300 + (1356.15)

Web hits = 1502.85 for a £300 TV spend

This can be compared with the observed data to see what margin of error exists. Of course the analysis will be more exact when all of the variables are taken into account. To take both independent variables into account we run a similar exercise but input both TV and web spend into the model.

With both variables applied the Pearson correlation for each variable does not change however the R square becomes 0.788: i.e. there is a 78.8% dependence of web hits on both TV spend and web spend (up from the individual variables alone).

As before the ANOVA is performed to test the the significance of the relationship.

Tabel 5: ANOVA

Model	Sum of squares	df	Mean square	F	Sig (p)
Regression	2605100.568	2	1302550.284	52.041	0.000
Residual	**700818.852**	**28**	**25029.245**		
Total	3305919.419	30			

Tabel 6: Coefficients

Model	B (Intercept)	Unstandardised co-efficients Std Error	Standardised co-efficicnets Beta	F	Sig (p)
Constant	-16.921	302.089		-0.056	0.956
TV spent	**3.20**	**0.065**	**0.519**	**4.918**	**0.000**
Web spend	2.927	0.637	0.485	4.597	0.000

As before the p value (Sig) is < 0.05, it can therefore be accepted that the relationship between TV and we spend on web hits is significant.

The model

The slope for web spend is 2.927 hits per £ when both independent variables are included. If TV spend is held constant, every £ spent on the web will result in 2.9 more hits. Conversely, every hit costs 0.34p.

The units for the variables in a multiple regression should be comparable. If they are not, the Beta coefficients standardise all of the variables in the regression, including the dependent and independent variables, putting all the variables on the same scale.

They tell us the number of standard deviations that the outcome will change as a result of one standard deviation change in the predictor.

Therefore, you can compare the magnitude of the coefficients to see which one has more effect. Larger betas will give you larger t values and lower p values.

The regression equation for both web and TV spend can now be modelled:

Web hits = Slope web spend × Google spend + Slope TV spend × TV spend + intercept

If we now substitute in web and TV spend of £600 and £300 respectively we get:

Web hits = (2.927 * 600) + (3.20 * 300) + (-16.921)

Web hits = 2699.28

Conclusion

In conclusion, a regression analysis requires steps to not only investigate the correlation between variables, but to test the significance of this correlation.

Once the correlation has been proven to provide a useful model then a regression equation can be built to predict the value of the dependant variable given arbitrary values of the independent variable.

adalyser.com
@adalyser

+44 (0) 333 666 7366

The complete TV advertising platform

Talk to us

By phone

+44 (0) 333 666 7366

By email

sales@adalyser.com

Or pay us a visit

Manchester Business Park, 3000 Aviator Way, Manchester, M22 5TG, United Kingdom