

JumpData

Actionable Analytics

Key Driver Analysis Case Study

Key Driver Analysis



What is Key Driver Analysis?

A statistical technique that aims to discover which of a set of predictive variables have the greatest influence in determining a particular metric

Why choose it over a classical linear regression technique?

The predictive variables are often highly correlated with one another leading to results being inaccurate and misleading

Why choose JumpData over another data science company?

JumpData has developed a unique, easy-to-use tool that conducts three of the most respected Key Driver Analysis techniques extremely rapidly without requiring any other statistical software packages

The output from the tool is easy to interpret and its speed of execution means multiple analyses can be performed in a short space of time

What is Key Driver Analysis?

The term Key Driver Analysis refers to any one of many statistical techniques that are used to discover which of a set of **independent** variables cause the greatest fluctuations in a given **dependent** variable

By applying a suitable Key Driver technique, then ordering these variables in terms of a **measure of importance**, a researcher can better understand where an organisation should focus its attention if it wants to see the **greatest impact**

For example, supposing you had data containing your customer's satisfaction ratings on a number of factors, e.g., efficiency, value for money, customer service etc. and wished to know which of these most influenced your customers' general level of satisfaction with your organisation

Then Key Driver Analysis can be applied to this data to give accurate, unbiased measures of the importance of the factors (relative to each other) in determining satisfaction

The important factors are often called the **drivers of satisfaction**

Key Driver Analysis Methods

One of the main problems in assessing the relative importance of variables being used to predict an outcome is that they are often **highly correlated** with one another (multi-collinearity)

This can result in importance values which have been derived from simple linear regression / correlation analysis being **inaccurate and misleading**

Three of the most well-respected techniques to address this problem are

Shapley Value Analysis

Kruskal's Relative Importance Analysis

Ridge Regression

If applied correctly, each of these techniques will give similar results

Shapley's and Kruskal's methods



Both Shapley Value Analysis and Kruskal's Relative Importance Analysis are fairly similar in concept

For each independent variable, they attempt to **strip out its correlations** between the other independents to give a more accurate measure of importance. In both cases you must consider every single possible regression model

The **difference** between them comes in the measure of **correlation** used

Semi-partial correlations in the case of Shapley; **partial** correlations in the case of Kruskal

However this simple difference leads to the a huge difference in computation speeds. Shapley's can be fairly easily computed; Kruskal's is more time consuming

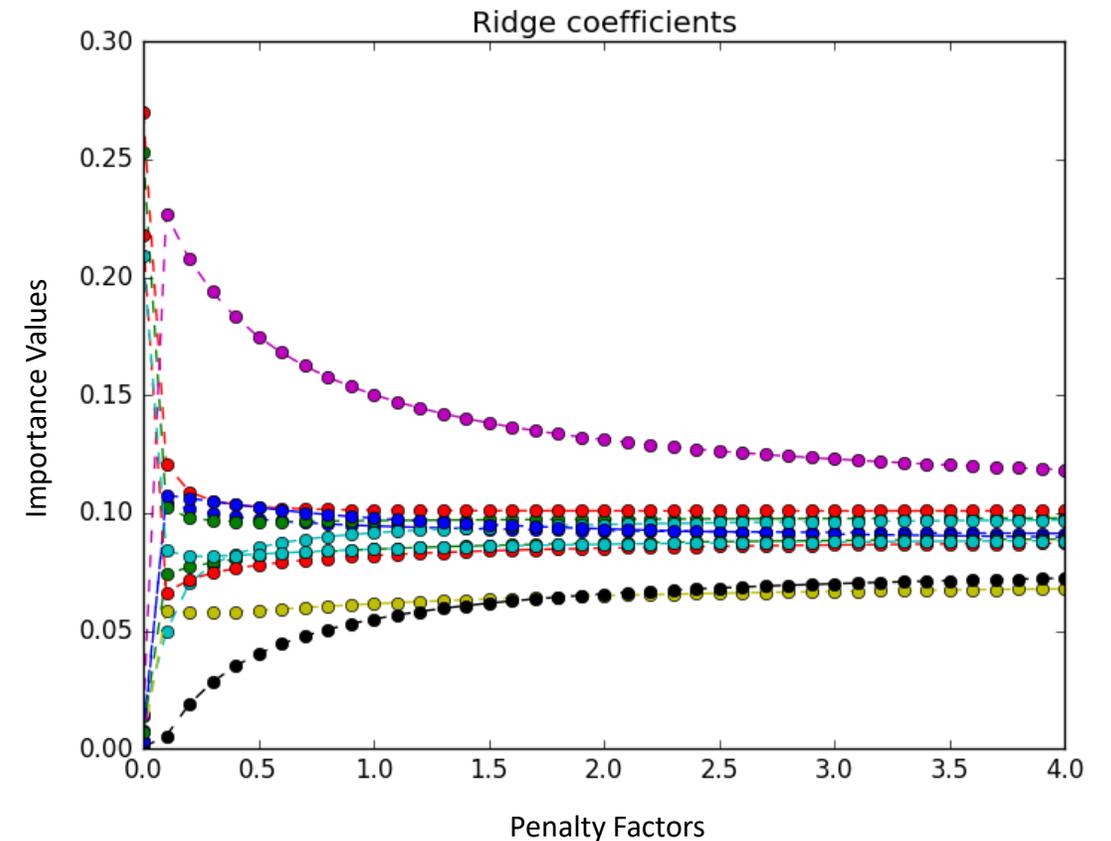
The JumpData Key Driver Tool addresses the issue with an innovative technique allowing **both** methods to be simultaneously constructed with **no** extra time-cost

Ridge Regression

This technique penalises the importance values in an attempt to neutralise the effect of multi-collinearity

In this case, the importance values returned depends on the penalising factor. So, it is common to return a set of importance values, then choose which one is most appropriate

To help choose the penalty factor, you can plot the importance values of the independents against the factors, then select the one where the graph appears to flatten



JumpData's Key Driver Tool



- Conducts **all three** analysis methods simultaneously
- Imports the data from .csv file so **no extra statistical software needed**
- Outputs all results to .csv file for ease of comparison / further analysis. Results shown as bar chart too
- Handles **20 or more** independent variables easily
- Executes rapidly
 - removing current processing restrictions on Kruskal's technique
 - allowing multiple analyses to be conducted, fast!
- Allows the user to inspect the Ridge Curve to choose a range of penalty functions

JumpData Key Driver Tool

Choose CSV input file

Choose File

Ridge

Ensure independent variable is in first column (see KeyDriverTestData.csv for example)

Run Ridge Regression and output scree plot

Min Ridge Parameter

Max Ridge Parameter

Ridge Step Value

Chosen Ridge Parameter

KD Output

Run Key Driver Analysis and output all results into .csv file in same location as input file
Analysis will be complete when bar chart appears



Expiry date: 08-11-2019 15:22

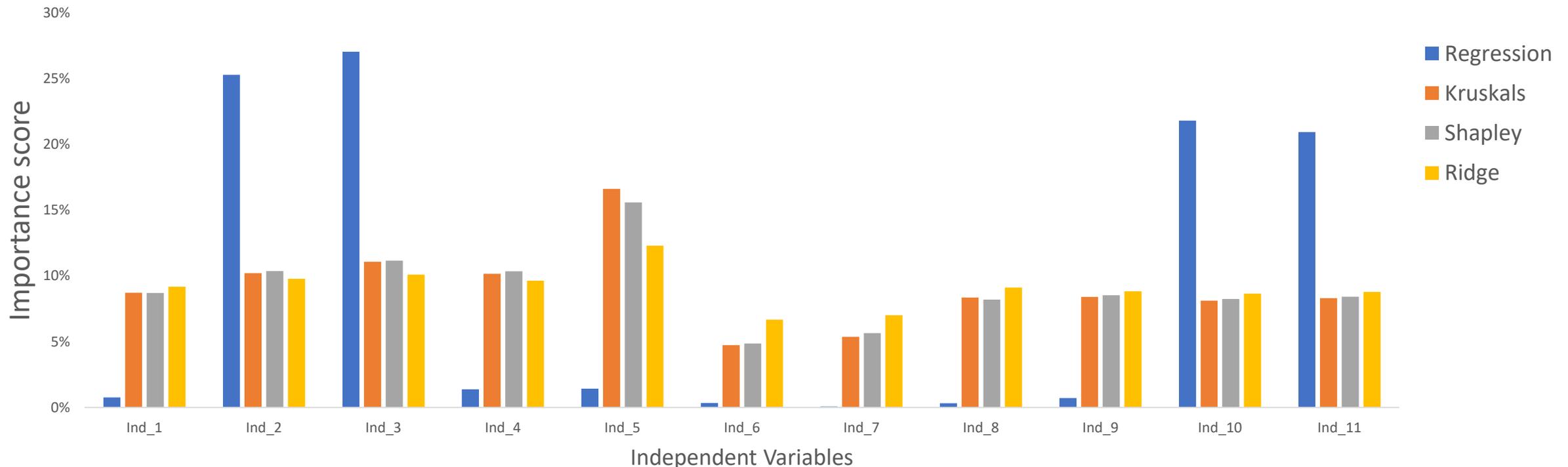
Key Driver Tool Output



Extreme Example - **highly correlated variables**

Linear regression importance scores completely misleading

Kruskal's and Shapley results similar to each other, Ridge close

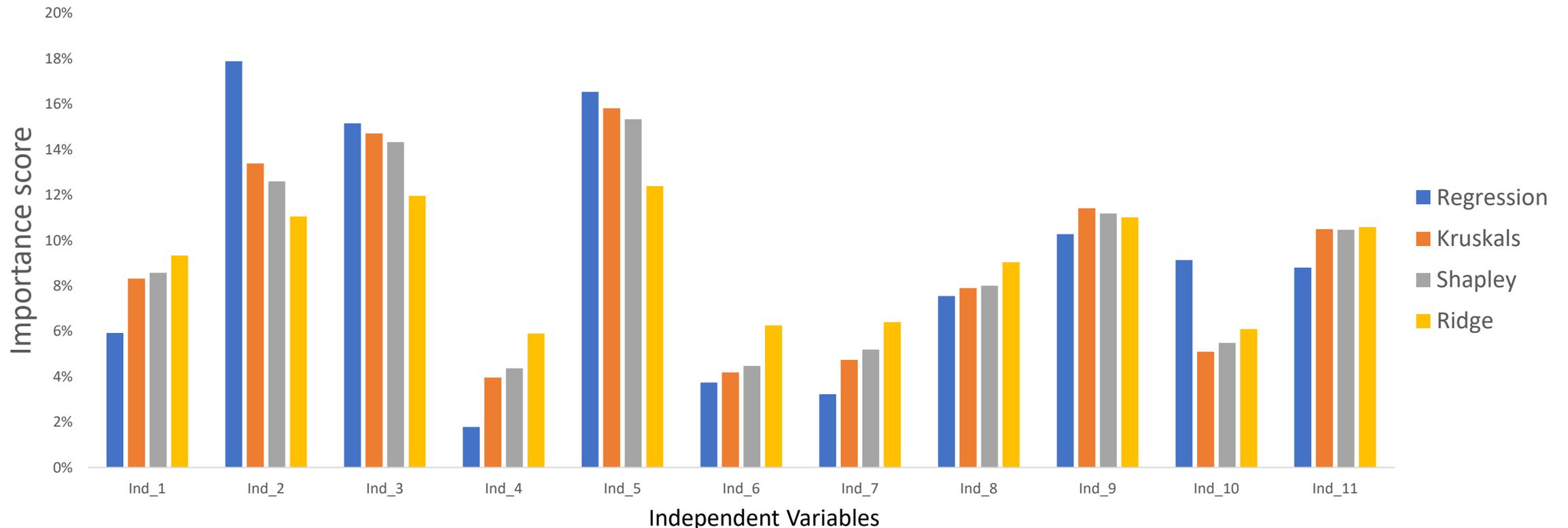


Key Driver Tool Output

Common Example – **loosely correlated variables**

Again linear regression importance are misleading – especially 2nd dependent variable

Kruskal's and Shapely results similar to each other, Ridge close





JumpData's Key Driver Tool

If you are new to Key Driver Analysis and want to get started quickly, or just want to have the edge over the competition, our tool is a cost effective way of achieving your objectives.

A tool requiring no additional stats software, coupled with the fast, simultaneous run-time of all three methods is, we believe, something **unique in this field**

For those who want to take a deeper dive in to how the analysis is performed, take a look at the blog post on the JumpData website

www.jumpdata.co.uk/post/key-driver-analysis-techniques

The tool can be purchased at a one-time cost of £495+VAT per installation

Support options are available. Get in touch if you'd like to know more