

*Modelling – R Models as customised HTML reporting*

2017

Summary

*Modelling licence holders have the ability to create models in R and import them into Market Insight as well as creating their own RMarkdown files to display html outputs alongside R results.*

This feature is available to Modelling licence holders through the Modelling Environment and is primarily aimed at experienced R users.

For more information and help with R visit their official website - <https://www.r-project.org/>

The following example will illustrate the functionality of this feature using the Market Insight Training System. It will use Logistic Regression in R, to build a model that can then generate an expression in Market Insight, which will return the probability of a Customer having a Policy Product D or not.

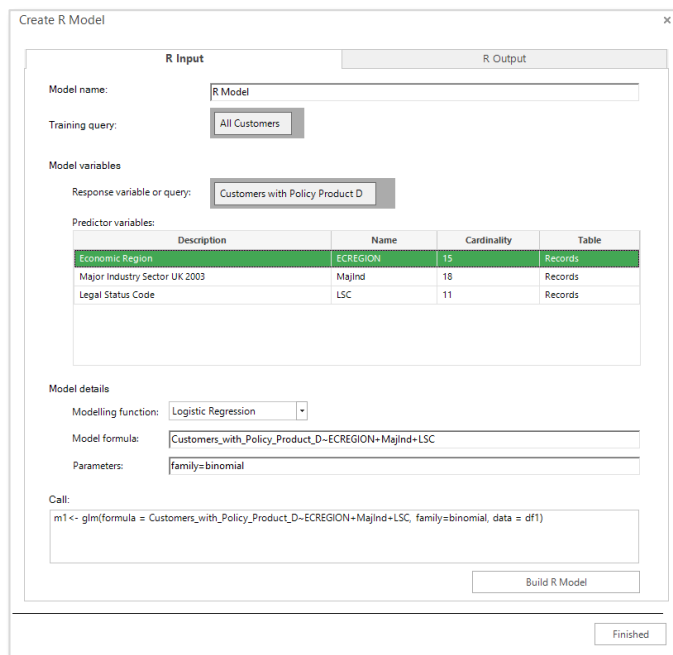
- Drag the **Modelling Environment** tool onto the workspace
- Create a **Customer** level selection of Customers with any **Policy Product Type** of **Product D** and drag it on to the selections tab as your **Analysis** group
- Create a **Customer** level selection and set a **Random limit** of **100,000**. Drag this on the selections tab and set it as the **Base** group.
- Click onto the **Dimensions** tab and then from the **System Explorer** drag and drop the variables **Economic Region**, **Major Industry Sector** and **Legal Status Code**

If you wish to compare the R Model with the equivalent Profile and Decision Tree build those now and create the respective Virtual Variables.

Id	Description	Ordering	Cardinality	Table	Omit Unclassified	Use Dimension
1	Emp Here Range	Ascending	14	Site	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
2	Major Industry Category	Nominal	13	Site	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3	Legal Status	Nominal	200	Site	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

- To prepare your environment to be transferred to R click on the **R** icon

- The **Create R Model** window is pre-populated using the information from your Modelling environment
- Change the Model Name to **R Model**



**Training Query** – this identifies the records to be used to build the model

**Response variable or Query** – this would be a numeric value which could also be a 0 or 1 to denote a yes/no query response

**Predictor variables** – variables that the user has determined to be used in the building of the model

The above information will form columns in a Data Grid which will be uploaded to R.

The information contained in the lower part of the window entitled **Model details** is as follows:

**Modelling functions** – a list of available functions are listed here including the ability to call out to user defined functions in R

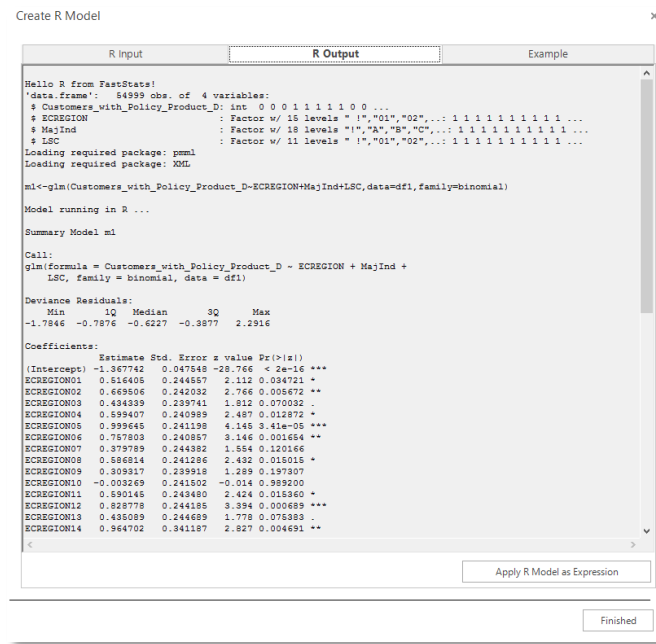
**Model formula** – Discoverer will automatically generate the model formula, however experienced users have the ability to edit the formula in this window

**Parameters** – extra parameters are required for some of the model building methods, experienced users may need to edit this especially if they have defined their own function

**Call** – the information that is sent to R based upon the settings made above

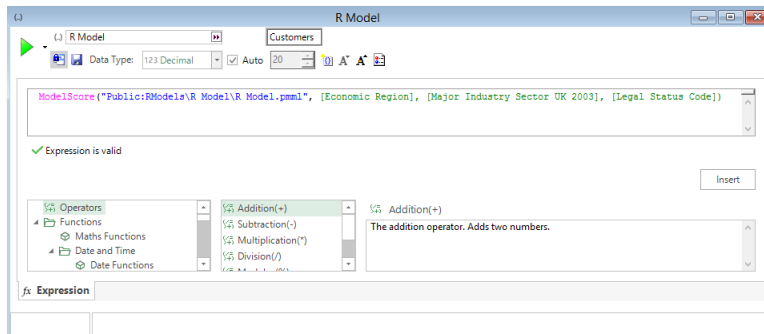
- Click the button **Build R Model**

This will export the data frame out and then bring the data back in, as shown in the tabbed window titled R Output.



A summary of the model activity is displayed and a pmml file is written into the Public directory and then accessed by:

- Click the button **Apply R Model as Expression**



Based upon the pmml file a probability score is calculated on whether we think the Customer Data has Policy Product D. The score will be between 0 and 1.

To make a comparison of the R model with say, a Profile model or Decision Tree model in the Market Insight Model Report tool, the expression will need to be converted into a virtual variable.

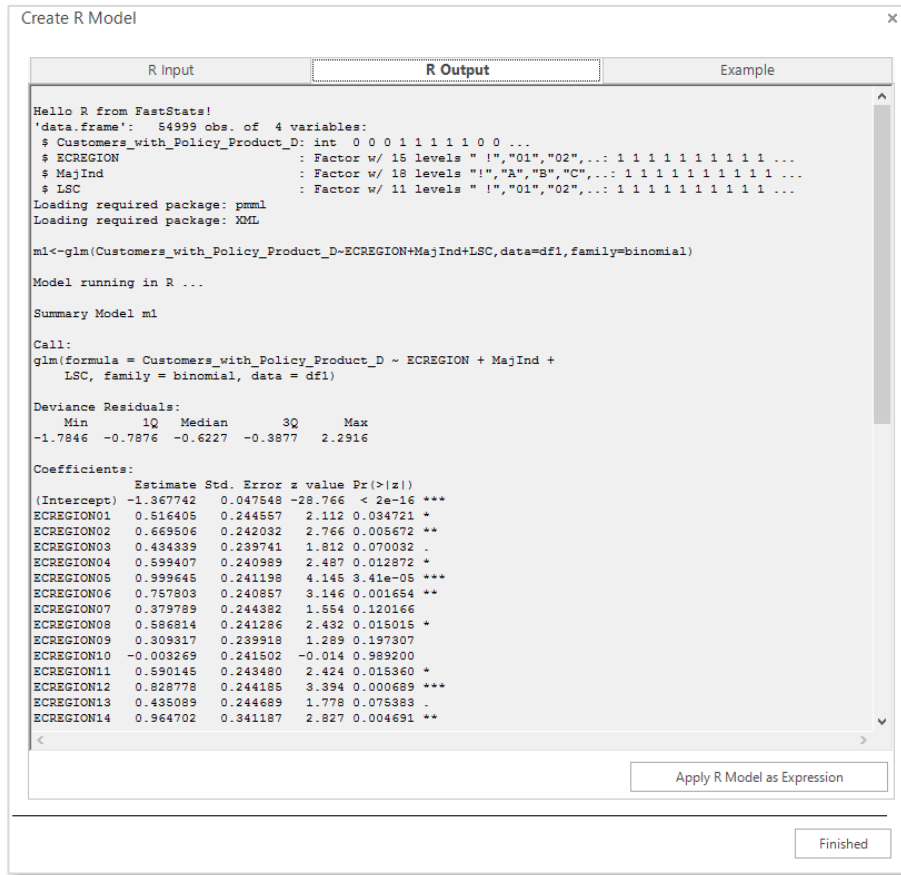
- Use the **Calculate Expression** wizard to create a numeric virtual variable of the model

The Model Report tool can only use Selector variables so the numeric virtual variable will need to be converted.

- Use the **Numeric Banding** wizard to convert the numeric virtual variable to a selector virtual variable

## R Markdown Files

This feature is also available to Modelling licence holders through the Modelling Environment and is primarily aimed at experienced R users.



The screenshot shows a window titled "Create R Model" with three tabs: "R Input", "R Output", and "Example". The "R Output" tab is active, displaying the following text:

```
Hello R from FastStats!  
'data.frame': 54999 obs. of 4 variables:  
 $ Customers_with_Policy_Product_D: int 0 0 0 1 1 1 1 1 0 0 ...  
 $ ECREGION : Factor w/ 15 levels "1","01","02",...: 1 1 1 1 1 1 1 1 1 1 ...  
 $ MajInd : Factor w/ 18 levels "!", "A", "B", "C",...: 1 1 1 1 1 1 1 1 1 1 ...  
 $ LSC : Factor w/ 11 levels "1", "01", "02",...: 1 1 1 1 1 1 1 1 1 1 ...  
Loading required package: pmml  
Loading required package: XML  
  
ml<-glm(Customers_with_Policy_Product_D~ECREGION+MajInd+LSC,data=df1,family=binomial)  
  
Model running in R ...  
  
Summary Model ml  
  
Call:  
glm(formula = Customers_with_Policy_Product_D ~ ECREGION + MajInd +  
LSC, family = binomial, data = df1)  
  
Deviance Residuals:  
 Min 1Q Median 3Q Max  
-1.7846 -0.7876 -0.6227 -0.3877 2.2916  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -1.367742 0.047548 -28.766 < 2e-16 ***  
ECREGION01 0.516405 0.244567 2.112 0.034721 **  
ECREGION02 0.669506 0.242032 2.766 0.005672 **  
ECREGION03 0.434339 0.239741 1.812 0.070032 .  
ECREGION04 0.599407 0.240989 2.487 0.012872 *  
ECREGION05 0.999645 0.241198 4.145 3.41e-05 ***  
ECREGION06 0.757803 0.240857 3.146 0.001654 **  
ECREGION07 0.379789 0.244982 1.554 0.120166 .  
ECREGION08 0.586814 0.241286 2.432 0.015015 *  
ECREGION09 0.309317 0.239918 1.289 0.197307 .  
ECREGION10 -0.003269 0.241502 -0.014 0.989200 .  
ECREGION11 0.590145 0.243480 2.424 0.015360 *  
ECREGION12 0.828778 0.244185 3.394 0.000689 ***  
ECREGION13 0.435059 0.244689 1.778 0.075383 .  
ECREGION14 0.964702 0.341187 2.827 0.004691 **
```

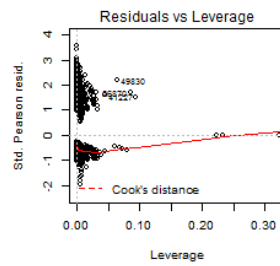
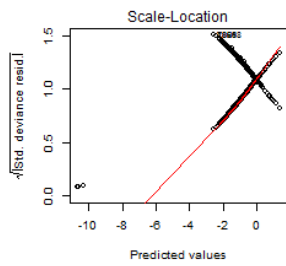
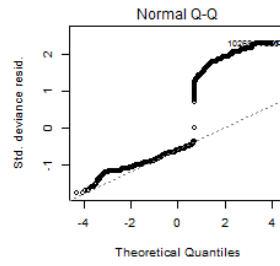
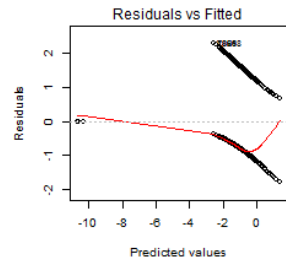
At the bottom of the window, there is a button labeled "Apply R Model as Expression" and a "Finished" button.

R users can create their own RMarkdown files (.Rmd) which will generate an HTML page with embedded R code. This code can refer to the dataframe used to train the model and the model itself. These files are placed in the Market Insight Public/RModels directory.

The file Example.Rmd has been created to display 4 plot graphs based upon the R model generated:

	R Input	R Output			Example
## 4		1	!	!	!
## 5		1	!	!	!
## 6		1	!	!	!

Use m1 to refer to the model that has been created.



Finished