**Information Value (IV) and Weight of Evidence (WOE) in R**

**Step 1 : Install and Load Package**

First you need to install 'Information' package and later you need to load the package in R.

*install.packages("Information")  
library(Information)*

**Step 2 : Import your data**

*#Read Data  
mydata <- read.csv("https://stats.idre.ucla.edu/stat/data/binary.csv")*

**Step 3 : Summarise Data**  
  
In this dataset, we have four variables and 400 observations. **The variable admit is a binary target or dependent variable.**

*summary(mydata)*

admit gre gpa rank

Min. :0.000 Min. :220 Min. :2.26 Min. :1.00

1st Qu.:0.000 1st Qu.:520 1st Qu.:3.13 1st Qu.:2.00

Median :0.000 Median :580 Median :3.40 Median :2.00

Mean :0.318 Mean :588 Mean :3.39 Mean :2.48

3rd Qu.:1.000 3rd Qu.:660 3rd Qu.:3.67 3rd Qu.:3.00

Max. :1.000 Max. :800 Max. :4.00 Max. :4.00

**Step 4 : Data Preparation**

Make sure your independent categorical variables are stored as factor in R. You can do it by using the following method -

*mydata$rank <- factor(mydata$rank)*

**Important Note :**The binary dependent variable has to be **numeric**before running IV and WOE as per this package. Do not make it factor.

**Step 5 : Compute Information Value and WOE**

In the first parameter, you need to define your data frame followed by your target variable. In the bins= parameter, you need to specify the number of groups you want to create it for WOE and IV.

*IV <- create\_infotables(data=mydata, y="admit", bins=10, parallel=FALSE)*

It takes all the variables except dependent variable as predictors from a dataset and run IV on them.

This function supports **parallel computing.**If you want to run you code in parallel computing mode, you can run the following code.

*IV <- create\_infotables(data=mydata, y="admit", bins=10,****parallel=TRUE****)*

You can add **ncore=** parameter to mention the number of cores to be used for parallel processing.

**Information Value in R**

In IV list,  the list Summary contains IV values of all the independent variables.

*IV\_Value = data.frame(IV$Summary)*

|  |
| --- |
| <https://2.bp.blogspot.com/-XFwD_q2DGZ8/WPNNRfij7NI/AAAAAAAAGL0/7uCfMskn0skP-UsCVuTYbWgyYf31UbpfQCLcB/s1600/IV%2BSummary.png> |
| Information Value Scores |
|  |

To get WOE table for variable **gre,**you need to call Tables list from IV list.

*print(IV$Tables$gre, row.names=FALSE)*

print(IV$Tables$gre, row.names=FALSE)

gre N Percent WOE IV

[220,420] 38 0.0950 -1.3748 0.128

[440,480] 40 0.1000 -0.0820 0.129

[500,500] 21 0.0525 -1.4860 0.209

[520,540] 51 0.1275 0.2440 0.217

[560,560] 24 0.0600 -0.3333 0.223

[580,600] 52 0.1300 -0.1376 0.225

[620,640] 51 0.1275 0.0721 0.226

[660,660] 24 0.0600 0.7653 0.264

[680,720] 53 0.1325 0.0150 0.265

[740,800] 46 0.1150 0.7653 0.339

To save it in a data frame, you can run the command below-

**gre = data.frame(IV$Tables$gre)**

**Plot WOE Scores**

To see trend of WOE variables, you can plot them by using **plot\_infotables**function.

*plot\_infotables(IV, "gre")*

|  |
| --- |
| [https://2.bp.blogspot.com/-jMbPo2yPFlE/WPNRUwqBksI/AAAAAAAAGME/2wSm48LQOIsdYzzHatiMr0xT3mbXDv9OQCLcB/s640/plot%2BWOE.png](https://2.bp.blogspot.com/-jMbPo2yPFlE/WPNRUwqBksI/AAAAAAAAGME/2wSm48LQOIsdYzzHatiMr0xT3mbXDv9OQCLcB/s1600/plot%2BWOE.png) |
| WOE Plot |

To generate multiple charts on one page, you can run the following command -

*plot\_infotables(IV, IV$Summary$Variable[1:3], same\_scale=FALSE)*

|  |
| --- |
| [https://3.bp.blogspot.com/-NxRJI3XcVgY/WPNRzs76YvI/AAAAAAAAGMI/s_alosPiMu8FRYbayf2-Xclkhb8Bkce3QCLcB/s640/MultiPlot%2BWOE.png](https://3.bp.blogspot.com/-NxRJI3XcVgY/WPNRzs76YvI/AAAAAAAAGMI/s_alosPiMu8FRYbayf2-Xclkhb8Bkce3QCLcB/s1600/MultiPlot%2BWOE.png) |
| MultiPlot WOE |