**XLSTAT results for Logistic regression**

XLSTAT can display the classification table (also called the confusion matrix) used to calculate the percentage of well-classified observations for a given cutoff point. Typically, for a cutoff value of 0.5, if the probability is less than 0.5, the observation is considered as being assigned to class 0, otherwise it is assigned to class 1.

The ROC curve can also be displayed. The ROC curve (Receiver Operating Characteristics) displays the performance of a model and enables a comparison to be made with other models. The terms used come from signal detection theory.

The proportion of well-classified positive events is called the sensitivity. The specificity is the proportion of well-classified negative events.

Results for logistic regression in XLSTAT

* **Correspondence between the categories of the response variable and the probabilities:**This table shows which categories of the dependent variable have been assigned probabilities 0 and 1.
* **Summary of the variables selection:**Where a selection method has been chosen, XLSTAT displays the selection summary. For a stepwise selection, the statistics corresponding to the different steps are displayed. Where the best model for a number of variables varying from p to q has been selected, the best model for each number or variables is displayed with the corresponding statistics and the best model for the criterion chosen is displayed in bold.
* **Goodness of fit coefficients:**This table displays a series of statistics for the independent model (corresponding to the case where the linear combination of explanatory variables reduces to a constant) and for the adjusted model.
  + Observations: The total number of observations taken into account (sum of the weights of the observations);
  + Sum of weights: The total number of observations taken into account (sum of the weights of the observations multiplied by the weights in the regression);
  + DF: Degrees of freedom;
  + -2 Log(Like.): The logarithm of the likelihood function associated with the model;
  + R² (McFadden): Coefficient, like the R², between 0 and 1 which measures how well the model is adjusted. This coefficient is equal to 1 minus the ratio of the likelihood of the adjusted model to the likelihood of the independent model;
  + R²(Cox and Snell): Coefficient, like the R², between 0 and 1 which measures how well the model is adjusted. This coefficient is equal to 1 minus the ratio of the likelihood of the adjusted model to the likelihood of the independent model raised to the power 2/Sw, where Sw is the sum of weights.
  + R²(Nagelkerke): Coefficient, like the R², between 0 and 1 which measures how well the model is adjusted. This coefficient is equal to ratio of the R² of Cox and Snell, divided by 1 minus the likelihood of the independent model raised to the power 2/Sw;
  + AIC: Akaike’s Information Criterion;
  + SBC: Schwarz’s Bayesian Criterion.
* **Test of the null hypothesis H0: Y=p0:**The H0 hypothesis corresponds to the independent model which gives probability p0 whatever the values of the explanatory variables. We seek to check if the adjusted model is significantly more powerful than this model. Three tests are available: the likelihood ratio test (-2 Log(Like.)), the Score test and the Wald test. The three statistics follow a Chi² distribution whose degrees of freedom are shown.
* **Type III analysis:**This table is only useful if there is more than one explanatory variable. Here, the adjusted model is tested against a test model where the variable in the row of the table in question has been removed. If the probability Pr > LR is less than a significance threshold which has been set (typically 0.05), then the contribution of the variable to the adjustment of the model is significant. Otherwise, it can be removed from the model.
* **Model parameters:**
  + Binary case: The parameter estimate, corresponding standard deviation, Wald's Chi2, the corresponding p-value and the confidence interval are displayed for the constant and each variable of the model. If the corresponding option has been activated, the "profile likelihood" intervals are also displayed.
  + Multinomial case: In the multinomial case, (J-1)\*(p+1) parameters are obtained, where J is the number of categories and p is the number of variables in the model. Thus, for each explanatory variable and for each category of the response variable (except for the reference category), the parameter estimate, corresponding standard deviation, Wald's Chi2, the corresponding p-value and the confidence interval are displayed. The odds-ratios with corresponding confidence interval are also displayed.

*Note: For PCR logistic regression, the first table of the model parameters corresponds to the parameters of the model which use the principal components which have been selected. This table is difficult to interpret. For this reason, a transformation is carried out to obtain model parameters which correspond to the initial variables.*

* **Model equation:**The equation of the model is then displayed to make it easier to read, or to re-use the model.
* **Table of standardized coefficients:**The table of standardized coefficients is used to compare the relative weights of the variables. The higher the absolute value of a coefficient, the more important the weight of the corresponding variable. When the confidence interval around standardized coefficients has value 0 (this can easily be seen on the chart of standardized coefficients), the weight of a variable in the model is not significant.
* **Predictions and residuals table:**The predictions and residuals table shows, for each observation, its weight, the value of the qualitative explanatory variable, if there is only one, the observed value of the dependent variable, the model's prediction, the same values divided by the weights, the standardized residuals and a confidence interval.
* **Classification table:**Activate this option to display the table showing the percentage of well-classified observations for both categories. If a validation sample has been extracted, this table is also displayed for the validation data.
* **ROC curve:**The ROC curve is used to evaluate the performance of the model by means of the area under the curve (AUC) and to compare several models together (see the description section for more details).
* **Comparison of the categories of the qualitative variables:**If one or more explanatory qualitative variables have been selected, the results of the equality tests for the parameters taken in pairs from the different qualitative variable categories are displayed.
* **Probability analysis table:** If only one quantitative variable has been selected, the probability analysis table allows you to see to which value of the explanatory variable corresponds a given probability of success.

The model equation for output= 1 is:

Log(P(outcome<=1)/P(outcome>1))= 4.251 - 4.528 \* Guided - 2.172\*Expertise

we can say that for one unit change in the variable Guided, the log of the ratio of the two probabilities, P(Small output)/P(bigger output)), will be decreased by 4.528.

Therefore, we can say that, in general, if you have a higher value in Guided ( which in our case 1) then the probability of saying that you are likely or gain less success will get smaller. So, it is more probable to gain more success

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | |  | | --- | | retVal.sort(function(a, b){ | |  | var aCount = wordCount[a] || -1; | |  | var bCount = wordCount[b] || -1; | |  | return bCount - aCount; | |  | }); | |
|  |  |
|  |  |