## 2.2 An estimation of quality on a control sample.

*Control (test) sample* is called sample, which is not used for building a tree, but used for an estimation of quality of the constructed tree. Two parameters are calculated: for the recognition problem the relative number of mistakes and for the regression analysis problem the variance on control sample. Since this sample does not participate in the tree construction, these parameters reflect the «true» unknown error more objectively. The bigger the control sample size, the higher the degree of approximation.

For a recognition problem, under the condition of independency of the observations, the frequency of mistakes belongs to binomial distribution. Therefore, knowing the number of errors on the control sample, it is possible to find a confidence interval to which, with the given probability, the unknown value of misclassification error belongs. In work [5] are given diagrams in which it is possible to define a confidence interval for the given control sample size and number of errors in the control.