

## QUALITY CONTROL OF BINNING

Maslyonkova M.V. (Russian Federation) Email: Maslyonkova426@scientifictext.ru

Maslyonkova Marina Valeryevna – student,  
DEPARTMENT OF DATA ANALYSIS, DECISION-MAKING THEORY AND FINANCIAL TECHNOLOGY,  
FINANCIAL UNIVERSITY UNDER THE GOVERNMENT OF THE RUSSIAN FEDERATION, MOSCOW

**Abstract:** binning is a categorization process to transform a continuous variable into a small set of groups or bins, which is widely used in credit scoring. The article analyzes one of the non-standard condition for a good binning algorithm. It is based on the construction of logistic regression model for the Weight of Evidence (WOE) values. The proposed method can be useful, for example, when it is unclear whether to allocate a bin with missing values to a separate group or not. In addition, it is easy to implement in practice.

**Keywords:** binning, credit scoring, logistic regression.

## ПРОВЕРКА КАЧЕСТВА ПРОЦЕДУРЫ БИННИНГА Маслѐнкова М.В. (Российская Федерация)

Маслѐнкова Марина Валерьевна – студент,  
Департамент анализа данных, принятия решений и финансовых технологий,  
Финансовый университет при Правительстве Российской Федерации, г. Москва

**Аннотация:** биннинг – это процесс категоризации, суть которого заключается в разбиении переменных определенным образом на группы, или бины. Он широко используется в кредитном скоринге. В статье анализируется одно из нестандартных условий проведения качественной процедуры биннинга. Оно основано на построении логистической регрессии по значениям Weight of Evidence (WOE). Предложенный метод может быть полезен, например, когда неясно, выделять ли бин с пропущенными значениями в отдельную группу или нет. К тому же, он легко реализуем на практике.

**Ключевые слова:** биннинг, кредитный скоринг, логистическая регрессия.

Математически WOE определяется как логарифм отношения доли «хороших» наблюдений к доле «плохих» наблюдений. Под «хорошими» и «плохими» наблюдениями понимается клиент, который не стал или стал дефолтным за период наблюдения соответственно. Этот показатель измеряет статистическую значимость каждого бина. Построим парную логистическую регрессию с одной объясняющей переменной WOE(x):

$$\pi(WOE(x)) = \frac{e^{\beta_0 + \beta_1 WOE(x)}}{1 + e^{\beta_0 + \beta_1 WOE(x)}}$$

Когда логистическая регрессия содержит одну предопределенную переменную, трансформированную к WOE, оценка коэффициентов по методу максимального правдоподобия может быть записана в явном виде:

$$\beta_0 = \ln\left(\frac{b}{g}\right), \beta_1 = 1 \quad (1)$$

Поскольку логистическая функция правдоподобия является вогнутой, то существует единственный максимум этой функции, а значит и единственное решение [1]. Сформулируем правило проверки правильности биннинга. Чтобы биннинг считался правильно проведенным, необходимо проверить, что оценки коэффициентов в логистической регрессии, где объясняющей переменной служит WOE переменной  $x$ , а объясняемой – переменная  $y$ , принимающая значение или **0**, или **1**, удовлетворяют равенству (1).

Для того чтобы показать практическую применимость описанного выше метода верификации биннинга, попытаемся решить вопрос для переменной, численно равной отношению ссуды к величине залога (LTV, от англ. Loan to Value): что делать с группой, содержащей пропущенные значения. Согласно правилу (1):

$$\beta_0 = \ln\left(\frac{713}{49101}\right) \approx -4.2321; \beta_1 = 1$$

Вариант с выделенными отдельно пропущенными значениями дает неточную оценку параметров. Если присоединить пропущенные значения к группе с наибольшим количеством дефолтов, оценки полностью совпадают в вычисленных.

Итак, описанный подход, основанный на построении логистической регрессии по значениям WOE полезен при проведении биннинга и довольно легко реализуем на практике.

*Список литературы / References*

1. *Гопин Цзэн*. Необходимое условие для хорошего алгоритма биннинга в кредитном скоринге. HIKARI Ltd. Прикладные Математические Науки. Т. 8, 2014. Вып. 65. [Электронный ресурс]. Режим доступа: <http://dx.doi.org/10.12988/ams.2014/> (дата обращения: 13.04.2017).

*Список литературы на английском языке / References in English*

1. *Guoping Zeng*. A necessary condition for a good binning algorithm in credit scoring. HIKARI Ltd. Applied Mathematical Sciences. Vol. 8, 2014. № 65. [Electronic resource]. URL: <http://dx.doi.org/10.12988/ams.2014/> (date of access: 13.04.2017).