

银行个人信用评分体系模型

论文复现课程报告

第 18 组

---

## 摘要

近年来，随着经济的逐步发展以及个人征信系统的完善，个人或企业的信用行为已经逐渐被各行各业所重视。针对这个课题，我们小组通过阅读三篇会议论文，并综合三篇论文的方法，结合现有的数据，利用 WOE 值筛选出对是否逾期有一定预测能力的变量，将部分分类型变量转化为连续型变量，并运用机器学习对逾期行为进行预测，从而构建了一套完整的个人信用风险评估模型。变量筛选结果显示，客户渠道、工作城市、收入、出生城市、贷记卡账户数、首张贷记卡发卡月份、合同金额和贷款余额对是否逾期存在着一定程度上的影响，其余因素影响较小。针对训练集和测试集存在一些选取数据上的差异，剔除了出生城市和合同金额。实验结果表明，此模型能以较高的准确度预测出用户的逾期行为，并对其做出评级，可解决个人信用评估问题。

**关键字:** 信用评估；WOE；机器学习；

---

## 目录

目录.....	3
一、 数据获取.....	4
二、 表现变量定义.....	4
1.不良贷款.....	4
2.贷款好、坏、不确定客户的定义.....	4
三、 模型流程图.....	4
四、 变量筛选方法.....	5
(一) 证据权重 WOE.....	5
(二) IV 的计算公式.....	5
(三) 指标剔除原则.....	6
五、 模型计算.....	6
(一)通过 report_id 进行匹配，形成文件 train（大表）.xlsx.....	6
(二)计算变量的 IV 值，筛选出预测能力一般和预测能力强的变量.....	6
(三)8 个变量的相关性分析.....	7
先对 8 个变量的缺失值进行替换，之后进行计算 8 个变量的相关性。.....	7
(四)评分卡模型预测.....	8
1.传统评分卡模型中参数的计算.....	8
2. 机器学习模型预测.....	11
3. 评价指标.....	11
六、 模型检验.....	12
(一) 传统评分卡模型逾期预测.....	12
(二) 机器学习模型逾期预测.....	13
(三) 模型选择.....	16
七、 信用规则.....	16
八、 数据对比.....	18
(一) 训练集和测试集变量对比分析.....	18
九、 附录.....	23
附录 1 变量分组部分.....	23
1.1 train(大表) 文件中数据的分组.....	23
1.2 contest_ext_crd_cd_ln(train)中的数据分组.....	27
1.3 contest_ext_crd_cd_lnd (train) 文件中数据的分组.....	30
附录 2 变量剔除部分.....	35

## 一、数据获取

我们获取并拿到了上海地区浦发银行的信用卡违约数据，我们挑选了两万人的优质数据，参考论文以及其他资料处理数据，运用传统评分卡模型与机器学习模型实现了信贷违约预测。

## 二、表现变量定义

### 1. 不良贷款

我国自 2002 年全面实行贷款五级分类制度，该制度按照贷款的风险程度，将银行信贷资产分为五类：正常、关注、次级、可疑、损失。不良贷款主要指次级、可疑和损失类贷款，关注贷款为不确定贷款。

### 2. 贷款好、坏、不确定客户的定义

只要有一个贷款定义为不良贷款，该客户即被定义为贷款坏客户；所有贷款定义为正常贷款，该客户即被定义为贷款好客户；如果有一个贷款为不确定，排除在模型开发样本之外。

$Y1=1$ ，表示贷款“坏”客户；  $Y1=0$ ，表示贷款“好”客户。

## 三、模型流程图

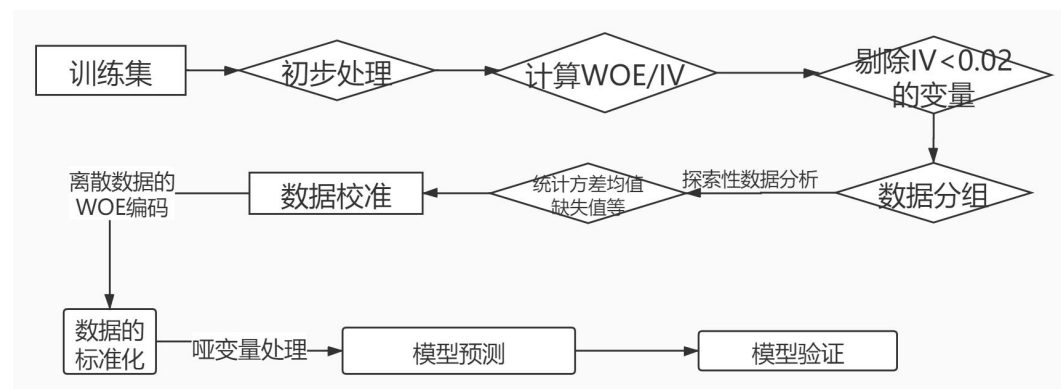


图 1 数据处理流程图

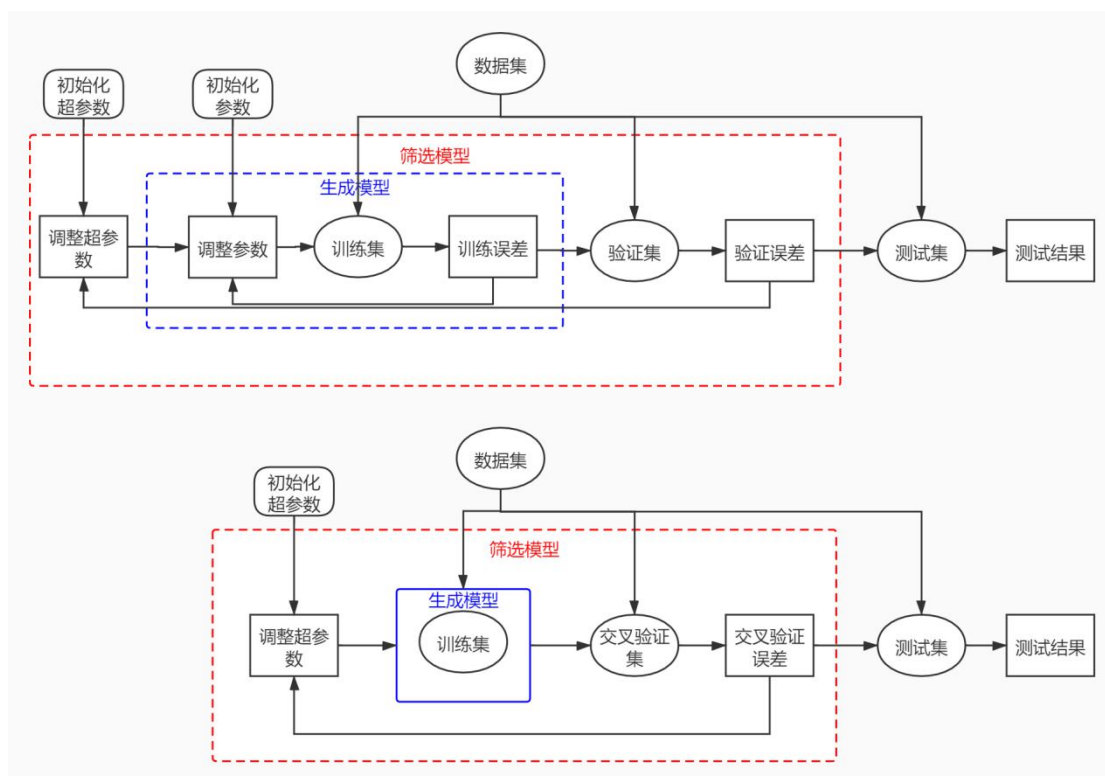


图 2 模型预测流程图

## 四、变量筛选方法

### (一) 证据权重 WOE

WOE 表示的实际上是“当前分组中坏客户占所有坏客户的比例”和“当前分组中好客户占所有坏客户的比例”的差异。

$$WOE_i = \ln \frac{py_i}{pn_i}$$

其中：  $py_i$  为第  $i$  个分组中的坏样本占所有坏样本的比例，  $pn_i$  为第  $i$  个指标中的好样本占所有好样本的比例；

注：将模型目标变量  $Y$  为 1 记为逾期用户（坏样本），对于目标变量为 0 记为未逾期用户（好样本）

### (二) IV 的计算公式

---

$IV$  的计算基于 WOE，可以看成对 WOE 的加权求和

第  $i$  个分组的  $IV_i$  值计算：

$$IV_i = (py_i - pn_i) \times WOE_i$$

总体的  $IV$ ：

$$IV = \sum_{i=1}^n IV_i$$

### (三) 指标剔除原则

若 IV 信息量取值小于 0.02，认为该指标对因变量没有预测能力，应该被剔除；

若 IV 信息量取值在 0.02 与 0.1 之间，认为该指标对因变量有较弱的预测能力；

若 IV 信息量取值在 0.1 与 0.3 之间，认为该指标对因变量的预测能力一般；

若 IV 信息量取值大于 0.3，认为该指标对因变量有较强的预测能力。

实际应用时，可以保留 IV 值大于 0.1 的指标。

## 五、模型计算

### (一) 申请评分卡评分

#### (一) 通过 report\_id 进行匹配，形成文件 train（大表）.xlsx

文件 train（大表）.xlsx 中含有 41 个变量，具体见文件。

#### (二) 计算变量的 IV 值，筛选出预测能力一般和预测能力强的变量

对 train（大表）.xlsx 中的 41 个变量计算 WOE 值和 IV 值，经过计算筛选出具有预测能力的 8 个变量。具体计算过程见：附录（变量剔除部分）.docx

表 1 评分模型最终选择的变量列表

变量符号	变量	IV 值	预测能力
X1	AGENT	0.125338	一般
X2	work_province	1.11385	较强
X3	SALARY	0.239494	一般
X4	born_provence	0.439003	较强
X5	share_credit_limit_amount	0.413962	较强
X6	FIRST_LOANCARD_OPEN_MONTH	0.198915	一般
X7	CREDIT_LIMIT	0.118648	一般
X8	BALANCE	0.107451	一般

(三)8 个变量的相关性分析

先对 8 个变量的缺失值进行替换，之后进行计算 8 个变量的相关性。

表 2 8 个变量的相关系数表

Correlations								
		SMEAN(AGENT)	SMEAN(work_province)	SMEAN(SALARY)	SMEAN(born_provence)	SMEAN(LOANCARD_OPEN_MONTH)	SMEAN(CREDIT_LIMIT)	SMEAN(BALANCE)
SMEAN(AGENT)	Pearson Correlation	1	.010	.130**	.022**	.081**	.035**	.067**
	Sig. (2-tailed)		.087	.000	.000	.000	.000	.000
	N	30000	30000	30000	30000	30000	30000	30000
SMEAN(work_province)	Pearson Correlation	.010	1	.039**	.338**	-.012*	.014*	.025**
	Sig. (2-tailed)	.087		.000	.000	.031	.017	.000
	N	30000	30000	30000	30000	30000	30000	30000
SMEAN(SALARY)	Pearson Correlation	.130**	.039**	1	.076**	.117**	.107**	.155**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
	N	30000	30000	30000	30000	30000	30000	30000

N		30000	30000	30000	30000	30000	30000	30000	30000
SMEAN(born_province)	Pearson Correlation	.022**	.338**	.076**	1	.070**	.088**	.077**	.083**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000
N		30000	30000	30000	30000	30000	30000	30000	30000
SMEAN(LOANCARD_COUNT)	Pearson Correlation	.081**	-.012*	.117**	.070**	1	.431**	.320**	.287**
	Sig. (2-tailed)	.000	.031	.000	.000		.000	.000	.000
N		30000	30000	30000	30000	30000	30000	30000	30000
SMEAN(FIRST_LOANCARD_OPEN_MONTH)	Pearson Correlation	.035**	.014*	.107**	.088**	.431**	1	.217**	.191**
	Sig. (2-tailed)	.000	.017	.000	.000	.000		.000	.000
N		30000	30000	30000	30000	30000	30000	30000	30000
SMEAN(CREDIT_LIMIT)	Pearson Correlation	.067**	.025**	.155**	.077**	.320**	.217**	1	.714**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000
N		30000	30000	30000	30000	30000	30000	30000	30000
SMEAN(BALANCE)	Pearson Correlation	.048**	.044**	.118**	.083**	.287**	.191**	.714**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	
N		30000	30000	30000	30000	30000	30000	30000	30000

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

#### (四) 评分卡模型预测

##### 1. 传统评分卡模型中参数的计算

- 根据参考文献，得到评分对照表



表 3 评分对照表

评分	Odds	评分	Odds
510	10.33	660	330.70
520	13.02	670	416.66
530	16.41	680	524.96
540	20.67	690	661.41
550	26.04	700	833.32
560	32.81	710	1049.92
570	41.34	720	1322.82
580	52.08	730	1666.65
590	65.62	740	2099.84
600	82.68	750	2645.63
610	104.17	760	3333.29
620	131.24	770	4199.68
630	165.35	780	5291.27
<b>640</b>	<b>208.33</b>	790	6666.58
650	262.48	800	8399.37

●  $A$ 、 $B$  的计算

$$Q = \frac{\text{未逾期客户}}{\text{逾期客户}} = \frac{28125}{1875} = 15, \text{ 逾期比}$$

对应的评分为  $P=530$ .

在此基础上, 评分值增加 30, 可以使未逾期/逾期比翻番

于是得到  $B = 43.2809$

$$A = P + B \cdot \ln(Q) = 530 + 43.2809 \cdot \ln(15) = 647.207$$

● 将变量的 woe 值进行转换变量,  $X_i$  变量的值  $x_{ij}$ , 按照

$$x'_{ij} = \frac{x_{ij}}{\max_j\{x_{ij}\} - \min_j\{x_{ij}\}} \text{ 进行转换}$$

- 得到回归模型

表 4 8 个变量的 WOE 值关于 Y 的回归模型参数计算表

Coefficients <sup>a</sup>					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	.102	.002		64.143	.000
x1	.082	.011	.043	7.685	.000
x2	.125	.005	.156	26.389	.000
x3	.121	.010	.069	12.044	.000
x4	.082	.004	.116	19.488	.000
x5	.024	.004	.040	6.224	.000
x6	.062	.006	.059	9.572	.000
x7	.023	.008	.024	2.908	.004
x8	.008	.005	.012	1.501	.133

a. Dependent Variable: Y

从表 4 可以看出，回归方程为

$$Y = 0.102 + 0.082 \cdot X1 + 0.125 \cdot X2 + 0.121 \cdot X3 + 0.082 \cdot X4 + 0.024 \cdot X5 + 0.062 \cdot X6 + 0.023 \cdot X7 + 0.008 \cdot X8$$

## (2) 计算信用卡打分

计算行为评分卡（信用卡部分打分），得到文件：contest\_ext\_crd\_cd\_ind  
（数据处理）train（打分）（缺失处理）

	A	B	C	D	E	F	G	H	I	J
	report_id	credit_limit_amount	used_credit_limit_amount	latest6_month_used_avg_amount	used_highest_amount	scheduled_payment_amount	信用卡已使用天数	最近还款日到应还款日的天数	行为评分贷款部分	state
	8787	0.15	-0.04	0.2	0.09	0.14	0.2	-9.09	758.857	0
	8787	0.15	0.16	-0.66	0.18	-0.5	-0.26	42.58	808.857	0
	9410	0.04	0.15	-0.62	-0.21	-0.5	0.58	42.58	809.227	0
	9410	0.04	-0.04	0.2	0.09	0.14	0.2	4.85	772.687	0
	9410	0.15	0.16	-0.82	0.04	-0.76	0.07	42.58	808.627	0
	9410	-0.12	0.05	-0.31	-0.21	-0.54	-0.34	42.58	808.317	0
	9410	-0.12	-0.04	0.2	0.09	0.14	0.2	-9.09	758.857	0

## 2. 机器学习模型预测

我们参照论文，添加了 SVM/KNN/决策树/随机森林/GBDT/Adaboost 以及现在较为流行的 Xgboost 以及 LightGBM，运用网格搜索法调整参数从而提升模型准确度，并在训练集上运用交叉验证得出效果最好的模型，在测试集中预测并进行不同模型的对比分析。

## 3. 评价指标

我们参照了论文后，使用了精度、精确率、召回率、F1 值、AUC 值以及 KS 值评估预测效果。

精度

Accuracy

定义： $(TP+TN)/(TP+FN+FP+TN)$

即所有分类正确的样本占全部样本的比例

精确率

Precision、查准率

定义： $(TP)/(TP+FP)$

即预测是正例的结果中，确实是正例的比例

召回率

Recall、查全率

定义： $(TP)/(TP+FN)$

即所有正例的样本中，被找出的比例

---

AUC

Area Under ROC Curve

定义：ROC 曲线下的面积

KS：KS 值越大，表示模型能够将正、负客户区分开的程度越大。通常来讲，KS>0.2 即表示模型有较好的预测准确性。

ks 值	含义
> 0.3	模型预测性较好
0.2~0.3	模型可用
0~0.2	模型预测能力较差
< 0	模型错误

## 六、模型检验

### （一）传统评分卡模型逾期预测

通过统计得到，逻辑回归模型预测 Y 和 Y 的对比汇总表，见表 6

表 6 模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.81	0.97
Recall	0.53	0.99
F1-score	0.64	0.98
Accuracy	96.07%	
AUC	0.7595	
KS	0.49	

从表 6 可以看出，逾期预测的正确率为 81%，未逾期预测的正确率为 97%，针对

---

Y=1 的预测准确度仍有待提升。

## （二）机器学习模型逾期预测

表 7 SVM 模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.97	0.97
Recall	0.49	1.00
F1-score	0.65	0.98
准确率	96.52%	
AUC	0.7456	
KS	0.49	

表 7 KNN 模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.91	0.97
Recall	0.51	1.00
F1-score	0.65	0.98
准确率	96.53%	
AUC	0.7519	
KS	0.504	

表 8 决策树模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.95	0.96

Recall	0.48	1.00
F1-score	0.64	0.98
准确率	96.42%	
AUC	0.7416	
KS	0.4831	

表 9 随机森林模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.98	0.97
Recall	0.49	1.00
F1-score	0.65	0.98
准确率	96.55%	
AUC	0.7458	
KS	0.4916	

表 10 Xgboost 模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.97	0.97
Recall	0.50	1.00
F1-score	0.66	0.98
准确率	96.8%	
AUC	0.75	
KS	0.51	

表 11 GBDT 模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.99	0.96
Recall	0.49	1.00
F1-score	0.65	0.98
准确率	96.73%	
AUC	0.7398	
KS	0.4797	

表 12 Adaboost 模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.95	0.97
Recall	0.51	1.00
F1-score	0.66	0.98
准确率	96.71%	
AUC	0.749	
KS	0.5005	

表 13 LightGBM 模型预测 Y 和 Y 的对比汇总表

	Y=1	Y=0
Precision	0.90	0.97
Recall	0.51	0.99
F1-score	0.66	0.98
准确率	96.69%	
AUC	74.01%	
KS	0.48	

### （三）模型选择

在银行信贷违约预测中我们参考了论文，采用了传统评分卡模型与机器学习方法，由表 6-表 13 可知，在传统的评分卡模型中，整体准确度尚可，不过对于违约（Y=1）人群预测的准确度有待提升，综合 KS/AUC 以及准确度来看，xgboost/GBDT/adaboost 以及 lightGBM 较 KNN/SVM 有了明显提升，在训练时长方面 LightGBM 略占优势，准确度方面 xgboost 略占优势，在不同的情况下建议使用不同集成模型，要求训练时长建议使用 LightGBM，准确度要求更高使用 xgboost

## 七、信用规则

表 15 变量的属性打分排序表

序号	变量	属性	逾期所占比例
1	SALARY	1	25.77%
2	work_province	东北地区	15.84%
3	work_province	华北地区	14.98%
4	born_province	华北地区	16.71%
5	SALARY	2	16.36%
6	FIRST_LOANCARD_OPEN_MONTH	2017 年	13.38%
7	FIRST_LOANCARD_OPEN_MONTH	2016 年	11.06%
8	AGENT	rongzhijia	15.72%
9	SALARY	3	12.40%
10	CREDIT_LIMIT	0-1000	13.80%
11	FIRST_LOANCARD_OPEN_MONTH	2015 年	8.03%
12	LOANCARD_COUNT	0-3	9.67%
13	CREDIT_LIMIT	1001-5000	9.82%
14	AGENT	APP	11.65%
15	work_province	西北地区	6.67%



16	FIRST_LOANCARD_OPEN_MONTH	2014 年	6.97%
17	CREDIT_LIMIT	5001-10000	7.74%
18	BALANCE	0	9.01%
19	LOANCARD_COUNT	4-6	6.96%
20	BALANCE	1-5000	8.13%
21	CREDIT_LIMIT	10001-20000	6.75%
22	FIRST_LOANCARD_OPEN_MONTH	2013 年	6.38%
23	BALANCE	5001-10000	6.98%
24	BALANCE	10001-20000	5.67%
25	CREDIT_LIMIT	20001-40000	5.84%
26	BALANCE	20001-50000	4.95%
27	BALANCE	50000 以上	3.84%
28	CREDIT_LIMIT	40001-60000	5.20%
29	FIRST_LOANCARD_OPEN_MONTH	2012 年	4.44%
30	LOANCARD_COUNT	10-15	4.29%
31	born_province	东北、华中地区	5.59%
32	LOANCARD_COUNT	7-9	4.08%
33	LOANCARD_COUNT	16-116	3.98%
34	CREDIT_LIMIT	60000 以上	3.89%
35	FIRST_LOANCARD_OPEN_MONTH	2010-2011 年	4.47%
36	AGENT	wechat	7.24%
37	FIRST_LOANCARD_OPEN_MONTH	2008-2009 年	3.41%
38	work_province	西南地区	4.31%
39	born_province	华东地区	3.41%
40	SALARY	4	6.94%

41	FIRST_LOANCARD_OPEN_MONTH	1998-2007 年	2.11%
42	work_province	华东地区	3.59%
43	SALARY	5	5.58%
44	SALARY	6	4.84%
45	AGENT	其他	2.84%
46	SALARY	7	4.17%
47	work_province	华南地区	1.85%
48	work_province	华中地区	0.33%

从表 11 可以看出，SALARY 取值为 1 的得分最低，为-5.67；work\_province 取值为华中地区的得分最高，为 16.31。

## 八、数据对比

### (一) 训练集和测试集变量对比分析

#### (1) 做出 AGENT 变量的测试集和训练集数据对比图

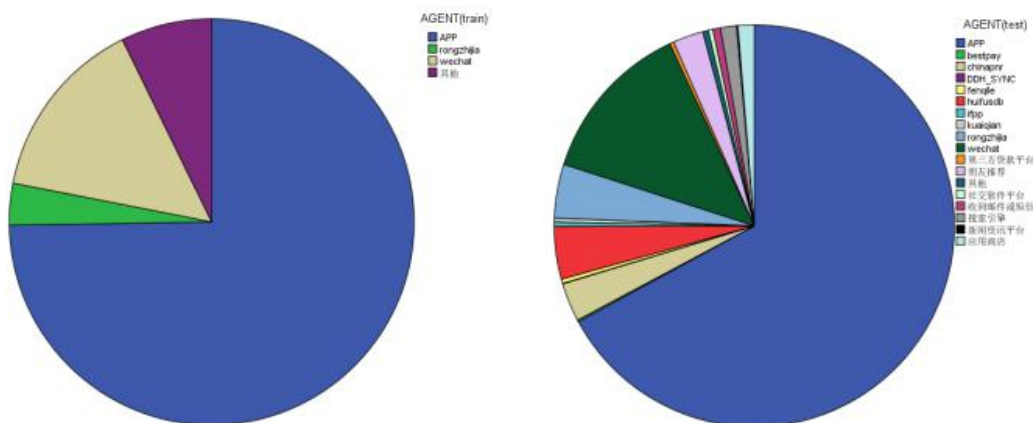


图 11 AGENT 变量的测试集和训练集数据对比图

从图 11 可以看出，测试集和训练集的 AGENT 变量分布基本一致，可以使用训练集的模型打分进行计算。

(2) 做出 work\_province 变量的测试集和训练集数据对比图

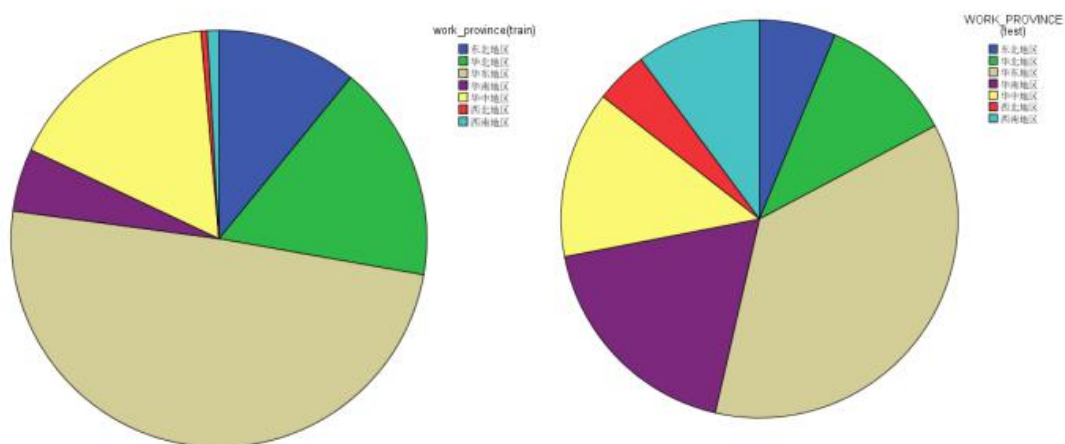


图 12 work\_province 变量的测试集和训练集数据对比图

从图 12 可以看出，测试集和训练集的 work\_province 变量分布基本一致，可以使用训练集的模型打分进行计算。

(3) 做出 SALARY 变量的测试集和训练集数据对比图

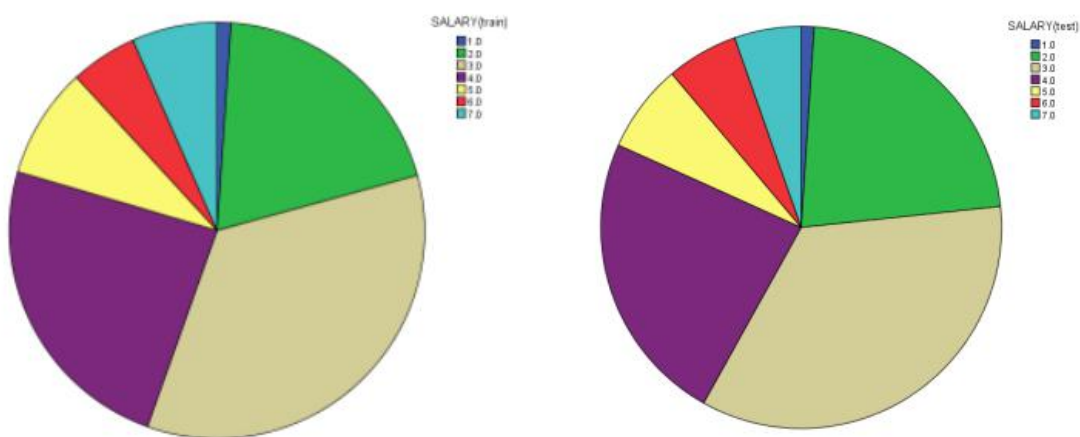


图 13 SALARY 变量的测试集和训练集数据对比图

从图 13 可以看出，测试集和训练集的 SALARY 变量分布基本一致，可以使用训练集的模型打分进行计算。

(4) 做出 born\_province 变量的测试集和训练集数据对比图

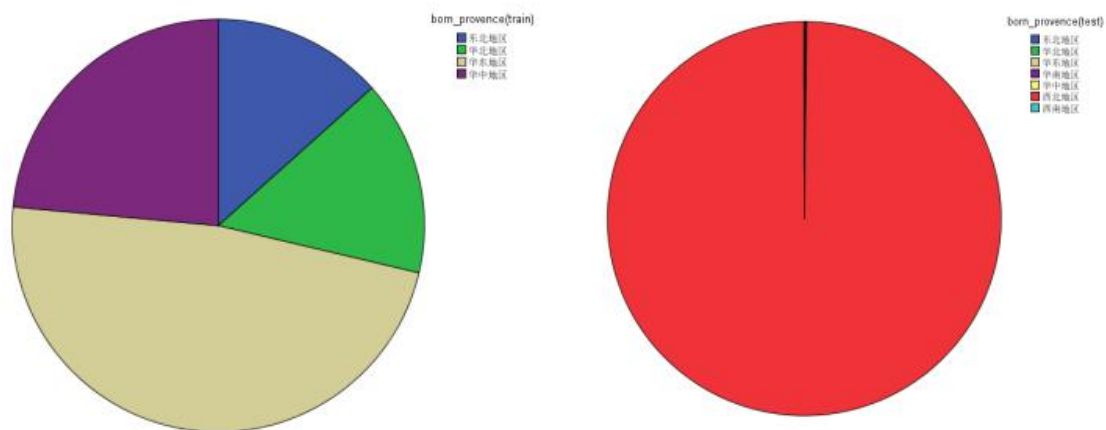


图 14 born\_province 变量的测试集和训练集数据对比图

从图 14 可以看出，测试集和训练集的 SALARY 变量分布完全不同，不能使用该变量对测试集打分进行计算。

#### (5) 做出 LOANCARD\_COUNT 变量的测试集和训练集数据对比图

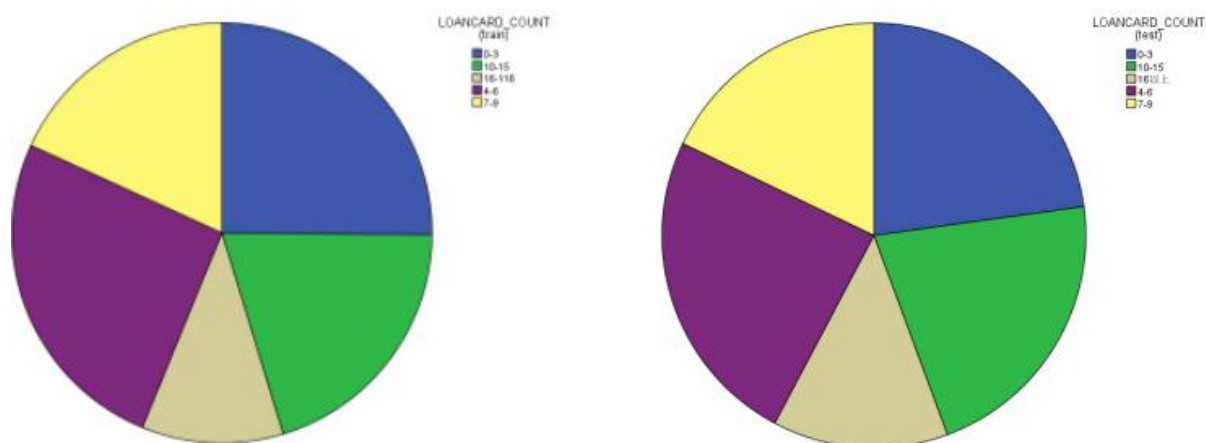


图 15 LOANCARD\_COUNT 变量的测试集和训练集数据对比图

从图 15 可以看出，测试集和训练集的 LOANCARD\_COUNT 变量分布基本一致，可以使用训练集的模型打分进行计算。

#### (6) 做出 FIRST\_LOANCARD\_OPEN\_MONTH 变量的测试集和训练集数据对比图

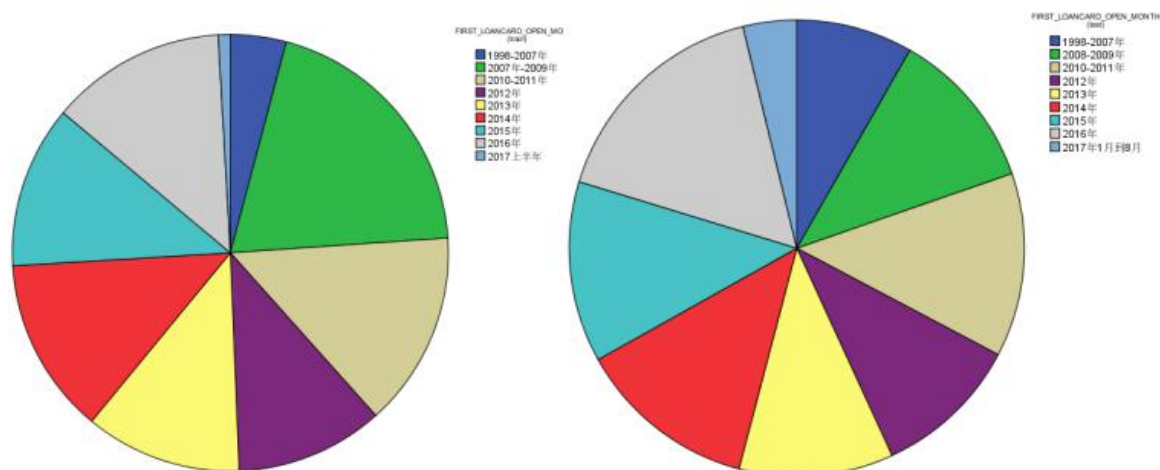


图 16 FIRST\_LOANCARD\_OPEN\_MONTH 变量的测试集和训练集数据对比图

从图 16 可以看出，测试集和训练集的 FIRST\_LOANCARD\_OPEN\_MONTH 变量分布基本一致，可以使用训练集的模型打分进行计算。

#### (7) 做出 CREDIT\_LIMIT 变量的测试集和训练集数据对比图

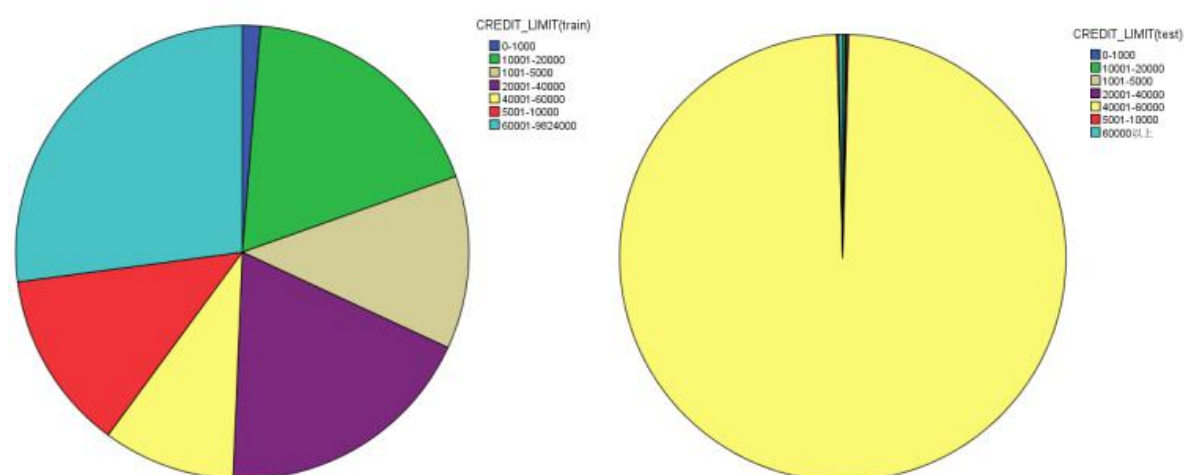


图 17 CREDIT\_LIMIT 变量的测试集和训练集数据对比图

从图 17 可以看出，测试集和训练集的 CREDIT\_LIMIT 变量分布完全不同，不能使用该变量对测试集打分进行计算。

#### (8) 做出 BALANCE 变量的测试集和训练集数据对比图

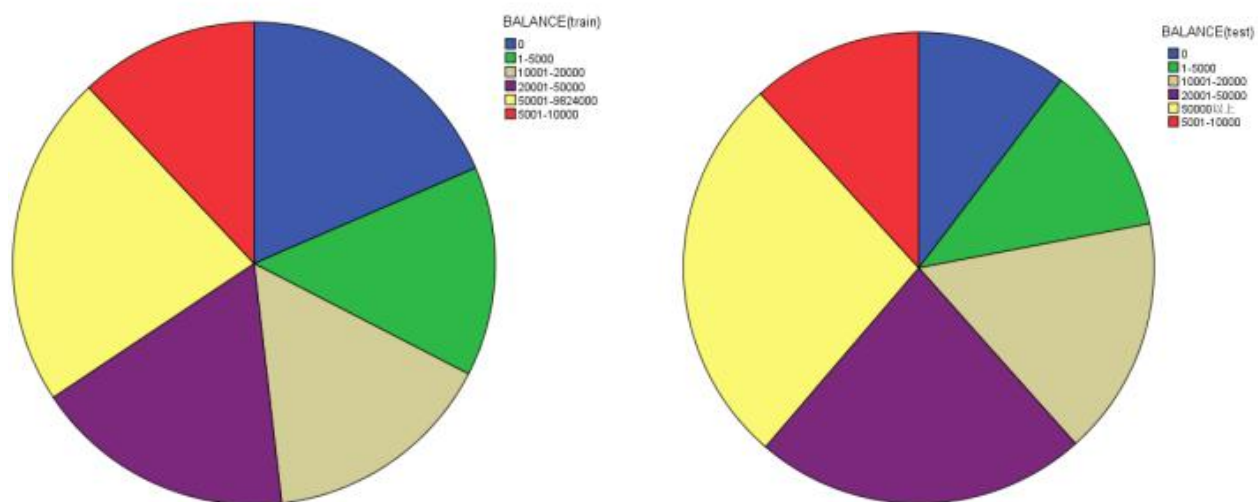


图 18 BALANCE 变量的测试集和训练集数据对比图

从图 18 可以看出，测试集和训练集的 BALANCE 变量分布基本一致，可以使用训练集的模型打分进行计算。

## 九、附录

### 附录 1 变量分组部分

#### 1.1 train(大表) 文件中数据的分组

##### ❖ 新增 SEX 变量

通过用户身份证号可以提取一个客户的 SEX 信息。在 30000 个客户中，男客户 21170 人，女客户 8830 人；

用户身份证号	SEX 中分组数据
第 17 位表示性别，偶数为女	0
第 17 位表示性别，奇数为男	1

##### ❖ 新增 born\_province（身份证所在省份）变量

通过用户身份证号可以提取一个客户的所在省份信息。身份证号码中第 1、2 位表示所在省份（自治区）的代码

##### ❖ born\_province、work\_province 变量的数据分组

通过对 27742 个客户的 work\_province 的邮政编码数据进行处理，按照省份进行修正数据，得到汇总数据见下：

work_province	计数	出生省份	计数	出生省份	计数
辽宁	1540	黑龙江	27	河南	182
浙江	4790	江苏	2089	河北	43
江西	4412	安徽	942	广东	138
湖南	3230	福建	3512	广西	70
吉林	1651	山东	1621	四川	133
山西	1454	湖北	1906	云南	2

born\_province、work\_province 的数据进行重新分组汇总，将 11（北京）、12（天津）、13（河北）、14（山西）、15（内蒙古）的数据，修正为“华北地区”；21（辽宁）、22（吉林）、23（黑龙江）的数据，修正为“东北地区”；37（山东）、32（江苏）、34（安徽）、33（浙江）、35（福建）、31（上海）的数据，修正为“华东地区”；42（湖北）、43（湖南）、41（河南）、36（江西）的数据，修正为“华中地区”；44（广东）、45（广西）、46（海南）的数据，修正为“华南地

区” ;51(四川)、53(云南)、52(贵州)、54(西藏)、50(重庆)的数据,修正为“西南地区”;64(宁夏)、65(新疆)、63(青海)、61(陕西)、62(甘肃)的数据,修正为“”;

#### ❖ 对 MARRY\_STATUS 变量的分组处理

通过对 30000 个客户的婚姻情况进行处理,把离婚、离异、丧偶、其他都修正为离异或丧偶,修正后,婚姻情况有离异及丧偶、未婚、已婚三个指标。

#### ❖ 对 EDU\_LEVEL 变量的分组处理

通过对 26942 个客户的教育情况进行处理,把博士研究生、硕士及以上、硕士研究生都修正为硕士及以上,专科、专科及以下、高中、初中都修正为专科及以下,修正后,教育情况有硕士及以上、本科、专科及以下三个指标。

#### ❖ 对 AGENT 变量的分组处理

保留 APP、rongzhijia、Wechat,将 chinapnr、fenqile、huifusdb、ifpp 的数据,修正为“其他”。

#### ❖ 增加新变量观察期

观察点设置为 2017/11/14,观察期为 LOAN\_DATE 到观察点的时间天数,之后将数据进行分组,分组为 121-180 天,181-240 天,241-300 天,301-360 天。

#### ❖ HOUSE\_LOAN\_COUNT 中的数据分组处理

HOUSE_LOAN_COUNT 中的原数据	HOUSE_LOAN_COUNT 中分组数据
0	0
1	1
2	2
3、4、5、6、7	3-7

#### ❖ COMMERCIAL\_LOAN\_COUNT 中的数据分组处理

COMMERCIAL_LOAN_COUNT 中的原数据	COMMERCIAL_LOAN_COUNT 中分组数据
0	0
1、2、3	1-3

#### ❖ OTHER\_LOAN\_COUNT 中的数据分组处理

OTHER_LOAN_COUNT 中的原数据	OTHER_LOAN_COUNT 中分组数据
0	0
1	1
2	2
3	3
4-6 的数据	4-6
7-12 的数据	7-12
13-24 的数据	13-24



25-342 的数据	25-342
------------	--------

❖ FIRST\_LOAN\_OPEN\_MONTH 中的数据分组处理

FIRST_LOAN_OPEN_MONTH 中的原数据	FIRST_LOAN_OPEN_MONTH 中分组数据
1994-2007 年的数据	1994-2007 年
2008 年-2011 年 1 月到 12 月的数据	2008-2011 年
2012 年、2013 年 1 月到 12 月的数据	2012-2013 年
2014 年 1 月到 12 月的数据	2014 年
2015 年 1 月到 12 月的数据	2015 年
2016 年 1 月到 6 月的数据	2016 上半年
2016 年 7 月到 12 月的数据	2016 下半年
2017 年 1 月到 6 月的数据	2017 上半年

❖ LOANCARD\_COUNT 中的数据分组处理

LOANCARD_COUNT 中的原数据	LOANCARD_COUNT 中分组数据
0-3 的数据	0-3
4-6 的数据	4-6
7-9 的数据	7-9
10-15 的数据	10-15
16-116 的数据	16-116

❖ FIRST\_LOANCARD\_OPEN\_MONTH 中的数据分组处理

FIRST_LOANCARD_OPEN_MONTH 中的原数据	FIRST_LOANCARD_OPEN_MONTH 中分组数据
1998-2007 年的数据	1998-2007 年
2008 年-2009 年 1 月到 12 月的数据	2008-2009 年
2010 年-2011 年 1 月到 12 月的数据	2010-2011 年
2012 年 1 月到 12 月的数据	2012 年
2013 年 1 月到 12 月的数据	2013 年
2014 年 1 月到 12 月的数据	2014 年
2015 年 1 月到 12 月的数据	2015 年
2016 年 1 月到 12 月的数据	2016 年
2017 年 1 月到 5 月的数据	2017 上半年

❖ STANDARD\_LOANCARD\_COUN 中的数据分组处理

STANDARD_LOANCARD_COUN 中的原数据	STANDARD_LOANCARD_COUN 中分组数据
0	0
1	1
2	2
3	3
4-12 的数据	4-12

❖ FIRST\_SL\_OPEN\_MONTH 中的数据分组处理

FIRST_SL_OPEN_MONTH 中的原数据	FIRST_SL_OPEN_MONTH 中分组数据
1993-2007 年的数据	1993-2007 年

2008 年-2012 年 1 月到 12 月的数据	2008-2012 年
2013、2014 年 1 月到 12 月的数据	2013-2014 年
2015 年 1 月到 12 月的数据	2015 年
2016 年 1 月到 12 月的数据	2016 年

❖ FINANCE\_CORP\_COUNT 中的数据分组处理

FINANCE_CORP_COUNT 中的原数据	FINANCE_CORP_COUNT 中分组数据
1	1
2	2
3	3
4	4
5	5
6	6
7-15 的数据	7-15

❖ FINANCE\_ORG\_COUNT 中的数据分组处理

FINANCE_ORG_COUNT 中的原数据	FINANCE_ORG_COUNT 中分组数据
1	1
2	2
3	3
4	4
5	5
6	6
7-15 的数据	7-15

❖ ACCOUNT\_COUNT 中的数据分组处理

ACCOUNT_COUNT 中的原数据	ACCOUNT_COUNT 中分组数据
1	1
2	2
3	3
4	4
5	5
6-8 的数据	6-8
9-105 的数据	9-105

❖ CREDIT\_LIMIT 中的数据分组处理

CREDIT_LIMIT 中的原数据	CREDIT_LIMIT 中分组数据
0-1000 的数据	0-1000
1001-5000 的数据	1001-5000
5001-10000 的数据	5001-10000
10001-20000 的数据	10001-20000
20001-40000 的数据	20001-40000
40001-60000 的数据	40001-60000
60001-9824000 的数据	60001-9824000

❖ MAX\_CREDIT\_LIMIT\_PER\_ORG 中的数据分组处理

MAX_CREDIT_LIMIT_PER_ORG 中的原数据	MAX_CREDIT_LIMIT_PER_ORG 中 分组数据
0	0
1-10000 的数据	1-10000
10001-500000 的数据	10001-500000

❖ MIN\_CREDIT\_LIMIT\_PER\_ORG 中的数据的分组处理

MIN_CREDIT_LIMIT_PER_ORG 中的原数据	MIN_CREDIT_LIMIT_PER_ORG 中 分组数据
0	0
1-10000 的数据	1-10000
10001-500000 的数据	10001-500000

❖ BALANCE 中的数据的分组处理

BALANCE 中的原数据	BALANCE 中分组数据
0	0
1-5000 的数据	1-5000
5001-10000 的数据	5001-10000
10001-20000 的数据	10001-20000
20001-50000 的数据	20001-50000
50001-9824000 的数据	50001-9824000

❖ USED\_CREDIT\_LIMIT 中的数据的分组处理

USED_CREDIT_LIMIT 中的原数据	USED_CREDIT_LIMIT 中分组数 据
0	0
1-5000 的数据	1-5000
5001-10000 的数据	5001-10000
10001-20000 的数据	10001-20000
20001-50000 的数据	20001-50000
100001-368124 的数据	100001-368124

❖ LATEST\_6M\_USED\_AVG\_AMOUNT 中的数据的分组处理

LATEST_6M_USED_AVG_AMOUNT 中的原数据	❖ LATEST_6M_USED_AVG_AMOU NT 中分组数据
0	❖ 0
1-500 的数据	❖ 1-500
501-1000 的数据	❖ 501-1000
1001-2000 的数据	❖ 1001-2000
2001-5000 的数据	❖ 2001-5000
5001-345186 的数据	❖ 5001-345186

## 1.2 contest\_ext\_crd\_cd\_ln(train) 中的数据的分组

finance_org 中的原数据	finance_org 中分组数据
A、B、C 开头的的数据	A、B、C 类贷款机构
D、E、F 开头的的数据	D、E、F 类贷款机构
G、H、I 开头的的数据	G、H、I 类贷款机构
J、K、L 开头的的数据	J、K、L 类贷款机构
M、N、O 开头的的数据	M、N、O 类贷款机构
P、Q、R 开头的的数据	P、Q、R 类贷款机构
S、T、U 开头的的数据	S、T、U 类贷款机构
V、W、X、Y、Z 开头的的数据	V、W、X、Y、Z 类贷款机构

❖ 增加新变量总贷款天数、已还款天数、最近还款日到应还款日的天数  
先用 end\_date 减去 open\_date 计算变量总贷款天数，再根据时间段进行分类替换，对于贷款期限少于 90 天的数据进行剔除。

end_date 减去 open_date 计算变量	贷款总天数
90-180 天的数据	90 至 180 天
181-270 天的数据	181 至 270 天
271-360 天的数据	271 至 360 天
361-540 天的数据	361 至 540 天
541-720 天的数据	541 至 720 天
721-1080 天的数据	721 至 1080 天
1081-10958 天的数据	1080 天以上

用 recent\_pay\_date 减去 open\_date 计算变量已还款天数，再根据时间段进行分类替换。

recent_pay_date 减去 open_date 计算变量	已还款天数
0-90 天的数据	0 至 90 天
91-180 天的数据	91-180 天
181-360 天的数据	181-360 天
361-540 天的数据	361-540 天
541-720 天的数据	541-720 天
721-6636 天的数据	720 天以上

用 scheduled\_payment\_date 减去 recent\_pay\_date 计算变量最近还款日到应还款日的天数，再根据时间段进行分类替换。

scheduled_payment_date 减去 recent_pay_date 计算变量	最近还款日到应还款日的天数
-30 到 0 天的数据	-30 到 0 天
1-90 天的数据	1-90 天
91-180 天的数据	91-180 天
181-270 天的数据	181-270 天
271-360 天的数据	271-360 天
361-1581 天的数据	361 天以上

❖ 变量 payment\_cyc 中的数据的分组处理

payment_cyc 中的原数据	payment_cyc 中分组数据
1-6 期的数据	1-6 期
7-12 期的数据	7-12 期
13-18 期的数据	13-18 期
19-24 期的数据	19-24 期
25-36 期的数据	25-36 期
37-361 期的数据	36 期以上

❖ 变量 credit\_limit\_amount 中的数据的分组处理

credit_limit_amount 中的原数据	payment_cyc 中分组数据
0-1000 的数据	0-1000
1001-2000 的数据	1001-2000
2001-3000 的数据	2001-3000
3001-4000 的数据	3001-4000
4001-6000 的数据	4001-6000
6001-10000 的数据	6001-10000
10001-9800000 的数据	10000 以上

❖ 变量 balance 中的数据的分组处理

balance 中的原数据	balance 中分组数据
0-300 的数据	0-300
301-600 的数据	301-600
601-1000 的数据	601-1000
1001-2000 的数据	1001-2000
2001-5000 的数据	2001-5000
5001-10000 的数据	5001-10000
10001-9800000 的数据	10000 以上

❖ 变量 remain\_payment\_cyc 中的数据的分组处理

remain_payment_cyc 中的原数据	remain_payment_cyc 中分组数据
0-3 的数据	0 至 3
3-6 的数据	3 至 6
7-9 的数据	7 至 9
10-12 的数据	10 至 12
13-24 的数据	13 至 24
25-360 的数据	25 至 360

❖ 变量 scheduled\_payment\_amount 中的数据的分组处理

scheduled_payment_amount 中的原数据	scheduled_payment_amoun
--------------------------------	-------------------------

	t 中分组数据
0-100 的数据	0-100
101-300 的数据	101-300
301-500 的数据	301-500
501-800 的数据	501-800
801-1200 的数据	801-1200
1201-3000 的数据	1201-3000
3003-30000 的数据	3000 以上

❖ 新增变量本月应还款与本月实际还款的差额

本月应还款与本月实际还款的差额 = scheduled\_payment\_amount - actual\_payment\_amount

本月应还款与本月实际还款的差额中的原数据	本月应还款与本月实际还款的差额中分组数据
-435190 到-1 的数据	小于 0
等于 0 的数据	0
1-9999	1-9999
10000-30000 的数据	10000-30000

❖ 变量 curr\_overdue\_amount 中的数据的分组处理

curr_overdue_amount 中的原数据	curr_overdue_amount 中分组数据
0 的数据	0
1-30000 的数据	大于 0

### 1.3 contest\_ext\_crd\_cd\_lnd (train) 文件中数据的分组

❖ 变量 finance\_org 中的数据的分组处理

finance_org 中的原数据	finance_org 中分组数据
A、B、C 开头的的数据	A、B、C 类银行
D、E、F 开头的的数据	D、E、F 类银行
G、H、I 开头的的数据	G、H、I 类银行
J、K、L 开头的的数据	J、K、L 类银行
M、N、O 开头的的数据	M、N、O 类银行
P、Q、R 开头的的数据	P、Q、R 类银行
S、T、U 开头的的数据	S、T、U 类银行
V、W、X、Y、Z 开头的的数据	V、W、X、Y、Z 类银行

❖ 变量 credit\_limit\_amount 中的数据的分组处理

credit_limit_amount 中的原数据	credit_limit_amount 中分组数据
---------------------------	---------------------------

0	0
1-2000 的数据	1-2000
2001-4000 的数据	2001-4000
4001-6000 的数据	4001-6000
6001-10000 的数据	6001-10000
10001-20000 的数据	10001-20000
20001-50000 的数据	20001-50000
50001-544325 的数据	50000 以上

❖ 增加新变量信用卡已使用天数、最近还款日到应还款日的天数

用 recent\_pay\_date 减去 open\_date 计算变量信用卡已使用天数，再根据时间段进行分类替换。

recent_pay_date 减去 open_date 计算变量	信用卡已使用天数
0	0
1-90 天的数据	1-90 天
91-180 天的数据	91-180 天
181-360 天的数据	181-360 天
361-540 天的数据	361-540 天
541-720 天的数据	541-720 天
721-900 天的数据	721-900 天
901-1080 天的数据	901-1080 天
1081-1440 天的数据	1081-1440 天
1441-1800 天的数据	1441-1800 天
1801-2160 天的数据	1801-2160 天
2161-9012 天的数据	2160 天以上

用 scheduled\_payment\_date 减去 recent\_pay\_date 计算变量最近还款日到应还款日的天数，再根据时间段进行分类替换。

scheduled_payment_date 减去 recent_pay_date 计算变量	最近还款日到应还款日的天数
-30 到 0 天的数据	-30 到 0 天
1-30 天的数据	1-30 天
31-90 天的数据	31-90 天
91-180 天的数据	91-180 天
181-270 天的数据	181-270 天
271-360 天的数据	271-360 天
361-540 天的数据	361-540 天
541-900 天的数据	541-900 天
901-1260 天的数据	901-1260 天
1261-6880 天的数据	1261 天以上

❖ 变量 share\_credit\_limit\_amount 中的数据分组处理

share_credit_limit_amount 中的原数据	share_credit_limit_amount 中分组数据
0	0
1-2000 的数据	1-2000
2001-4000 的数据	2001-4000
3001-4000 的数据	3001-4000
4001-6000 的数据	4001-6000
6001-10000 的数据	6001-10000
10001-20000 的数据	10001-20000
20001-50000 的数据	20001-50000
50001-620000 的数据	50000 以上

❖ 变量 used\_credit\_limit\_amount 中的数据分组处理

used_credit_limit_amount 中的原数据	used_credit_limit_amount 中分组数据
0	0
1-2000 的数据	1-2000
2001-4000 的数据	2001-4000
3001-4000 的数据	3001-4000
4001-6000 的数据	4001-6000
6001-10000 的数据	6001-10000
10001-20000 的数据	10001-20000
20001-629032 的数据	20000 以上

❖ 变量 latest6\_month\_used\_avg\_amount 中的数据分组处理

latest6_month_used_avg_amount 中的原数据	latest6_month_used_avg_amount 中分组数据
0	0
1-1000 的数据	1-1000
1001-2000 的数据	1001-2000
2001-3000 的数据	2001-3000
3001-4000 的数据	3001-4000
4001-6000 的数据	4001-6000
6001-10000 的数据	6001-10000
10001-20000 的数据	10001-20000
20001-629032 的数据	20000 以上

❖ 变量 used\_highest\_amount 中的数据分组处理

used_highest_amount 中的原数据	used_highest_amount 中分组数据
0	0
1-1000 的数据	1-1000
1001-2000 的数据	1001-2000



2001-4000 的数据	2001-4000
4001-6000 的数据	4001-6000
6001-10000 的数据	6001-10000
10001-20000 的数据	10001-20000
20001-30000 的数据	20001-30000
30001-1007202 的数据	30000 以上

❖ 变量 `scheduled_payment_amount` 中的数据分组处理

<code>scheduled_payment_amount</code> 中的原数据	<code>scheduled_payment_amount</code> 中分组数据
0	0
1-200 的数据	1-200
201-500 的数据	201-500
501-1000 的数据	501-1000
1001-2000 的数据	1001-2000
2001-4000 的数据	2001-4000
4001-199664 期的数据	4000 以上

❖ 新增变量实际还款额与应还款额之差

用 `scheduled_payment_amount` 减去 `actual_payment_amount` 计算变量实际还款额与应还款额之差，再根据差额分布行分组替换。

<code>scheduled_payment_amount</code> 减去 <code>actual_payment_amount</code> 计算变量	实际还款额与应还款额之差
-263927 到-10000 的数据	-263927 到-10000
-9999 到-5000 的数据	-9999 到-5000
-4999 到-1000 的数据	-4999 到-1000
-999 到-1 的数据	-999 到-1
0	0
1-86469 的数据	大于 1

❖ 变量 `curr_overdue_cyc` 中的数据分组处理

<code>curr_overdue_cyc</code> 中的原数据	<code>curr_overdue_cyc</code> 中分组数据
0	0
1	1
2-40 的数据	2-40

❖ 变量 `curr_overdue_amount` 中的数据分组处理

<code>curr_overdue_amount</code> 中的原数据	<code>curr_overdue_cyc</code> 中分组数据
0	0
1-1000 的数据	1-1000
1001-86469 的数据	大于 1000

(2) `contest_ext_crd_is_creditcue` 文件中数据的预处理

❖ 变量 `OTHER_LOAN_COUNT` 中的数据分组处理

OTHER_LOAN_COUNT 中的原数据	OTHER_LOAN_COUNT 中分组数据
0、1、2 的数据	0-2
1-1000 的数据	1-1000
1001-86469 的数据	大于 1000

(3) `contest_ext_crd_lnd_ovd(train)` 文件中数据的预处理

❖ 匹配因变量 Y

根据 `contest_basic_train` 中的 `report_id` 进行匹配，利用 `VLOOKUP` 函数的得到 `contest_ext_crd_lnd_ovd(train)` 文件中对应 `id` 下的 Y 的值。

(4) `contest_ext_crd_is_creditcue` 文件中数据的预处理

❖ 变量 `MONTH_DW` 中的数据分组处理

MONTH_DW 中的原数据	MONTH_DW 中分组数据
1994-2007 年的数据	1994-2007 年
2008 年-2011 年 1 月到 12 月的数据	2008-2011 年
2012 年、2013 年 1 月到 12 月的数据	2012-2013 年
2014 年 1 月到 12 月的数据	2014 年
2015 年 1 月到 12 月的数据	2015 年
2016 年 1 月到 12 月的数据	2016 年
2017 年 1 月到 5 月的数据	2017 上半年

❖ 变量 `AMOUNT` 中的数据分组处理

AMOUNT 中的原数据	AMOUNT 中分组数据
0-50 的数据	0-50
50-100 的数据	50-100
101-200 的数据	101-200
201-500 的数据	201-500
501-1000 的数据	501-1000
1001-1500 的数据	1001-1500
1501-2000 的数据	1501-2000
2001-5000 的数据	2001-5000
5000 以上的数据	5000 以上

## 附录 2 变量剔除部分

SEX	逾期客户	$B_i$	无逾期客户	$P_i$	$WE$ 值	IV 值
男	1440	0.768	19730	0.7015	0.090553	0.00602
女	435	0.232	8395	0.298489	-0.25200	0.0167549
总和	28125		1875			0.0227757

SEX 变量的 IV 值为 0.0227757，认为该指标对是否逾期有较弱的预测能力。

AGENT	逾期客户	$B_i$	无逾期客户	$P_i$	$WE$ 值	IV 值
APP	774	0.82604	5870	0.732377	0.12035	0.0112723
rongzhijia	47	0.05216	252	0.031441	0.46711	0.00874376
wechat	97	0.103522	1242	0.154959	-0.40338	0.0207489
其他	19	0.0202775	651	0.0812227	-1.3877	0.0845727
总和	937		8015			0.125338

AGENT 变量 IV 值为 0.125338，该指标对是否逾期该指标对因变量的预测能力一般。

IS_LOCAL	逾期客户	$B_i$	无逾期客户	$P_i$	$WE$ 值	IV 值
本地籍	1231	0.656533	16275	0.578667	0.12625	0.00983042
非本地籍	644	0.343467	11850	0.421333	-0.20433	0.0159108
总和	1875		28125			0.02574122

IS\_LOCAL 变量 IV 值为 0.02574122，该指标对是否逾期有较弱的预测能力。

work\_province 变量下的总的逾期客户为 1732，总的无逾期客户为 26010

work_province	逾期客户	$B_i$	无逾期客户	$P_i$	$WE$ 值	IV 值
东北地区	480	0.277136	2551	0.0980777	1.0387	0.185997
华北地区	700	0.404157	3972	0.15271	0.97326	0.244723
华东地区	492	0.284065	13201	0.507536	-0.58036	0.129695

华南地区	25	0.0144342	1329	0.0510957	-1.2641	0.0463439
华中地区	15	0.00866051	4591	0.176509	-3.0146	0.505996
西北地区	9	0.0051963	126	0.00484429	0.0701469	0.000024927
西南地区	11	0.00635104	240	0.00922722	-0.373539	0.00107437
总和						1.11385

**work\_province 变量 IV 值为 1.11385，认为该指标对是否逾期有较强的预测能力**

EDU\_LEVEL 变量下的总的逾期客户为 1514，总的无逾期客户为 23226

EDU_LEVEL	逾期客户	$B_i$	无逾期客户	$P_i$	$WE$ 值	IV 值
硕士及以上	6	0.003963	204	0.008783	-0.79584	0.0039362
本科	284	0.187583	5699	0.245372	-0.26855	0.0155195
专科及以下	1208	0.797886	17050	0.734091	0.083333	0.0053163
其他	16	0.010568	273	0.0117541	-0.10637	0.0001262
总和						0.0247981

EDU\_LEVEL 变量 IV 值为 0.0247981，认为该指标对是否逾期有较弱的预测能力

MARRY\_STATUS 变量下的总的逾期客户为 1875，总的无逾期客户为 28125

MARRY_STATUS	逾期客户	$B_i$	无逾期客户	$P_i$	$WE$ 值	IV 值
未婚	674	0.359467	8892	0.31616	0.12837	0.00555941
已婚	1108	0.590933	18091	0.643236	-0.084808	0.00443564
离异或丧偶	93	0.0496	1142	0.00406044	0.20011	0.00180013
总和						0.0117952

MARRY\_STATUS 变量 IV 值为 0.0117952，认为该指标对是否逾期没有预测能力，应该被剔除

SALARY 变量下的总的逾期客户为 930，总的无逾期客户为 7934

SALARY	逾期客户	$B_i$	无逾期客户	$P_i$	$WE$ 值	IV 值
1	25	0.0268817	72	0.00907487	1.0859	0.0193371
2	289	0.310753	1477	0.186161	0.51239	0.0638392
3	379	0.407527	2678	0.337535	0.18844	0.0131892
4	147	0.158065	1970	0.248298	-0.45163	0.0407522

5	42	0.0451613	711	0.0896143	-0.68527	0.0304625
6	23	0.0247312	452	0.05697	-0.83446	0.026902
7	25	0.0268817	574	0.0723469	-0.99003	0.0450117
总和						0.239494

**SALARY 变量 IV 值为 0.239494，认为该指标对是否逾期有较强的预测能力。**

HAS\_FUND 变量下的总的逾期客户为 1875，总的无逾期客户为 28123

HAS_FUND	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>P<sub>i</sub></i>	<i>WE</i> 值	IV 值
0	1048	0.558933	15290	0.543683	0.027664	0.000422
1	827	0.441067	12833	0.456317	-0.033991	0.0005184
总和						0.0009402

HAS\_FUND 变量 IV 值为 0.0009402，认为该指标对是否逾期没有预测能力，应该被剔除

born\_provence 变量下的总的逾期客户为 1875，总的无逾期客户为 28125

born_provence	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>P<sub>i</sub></i>	<i>WE</i> 值	IV 值
东北、华中地区	619	0.330133	10463	0.372018	-0.11945	0.005003
华北地区	767	0.409067	3822	0.135893	1.1020	0.301039
华东地区	489	0.2068	13840	0.479289	-0.60855	0.132961
总和						0.439003

**born\_provence 变量 IV 值为 0.439003，认为该指标对是否逾期有较强的预测能力**

观察期变量下的总的逾期客户为 1875，总的无逾期客户为 28125

观察期	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>P<sub>i</sub></i>	<i>WE</i> 值	IV 值
121-180 天	802	0.427733	15667	0.557049	-0.26415	0.0341591
181-240 天	870	0.464	9309	0.330987	0.33781	0.0449328
241-300 天	168	0.0896	2612	0.0928711	-0.035857	0.000117
301-360 天	35	0.018667	537	0.0190933	-0.022600	0.0000095
总和						0.072188

观察期变量 IV 值为 0.072188，认为该指标对是否逾期有较弱的预测能力

HOUSE\_LOAN\_COUNT 变量下的总的逾期客户为 1874，总的无逾期客户为 28097

HOUSE_LOAN_COUNT	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>p<sub>i</sub></i>	<i>WE</i> 值	IV 值
0	1708	0.911419	24499	0.871944	0.044278	0.0017479
1	156	0.0832444	3301	0.117486	-0.34454	0.0117975
2	8	0.0042689	254	0.00904011	-0.75031	0.0035798
3-7	2	0.00106724	43	0.00153041	-0.36047	0.0001670
总和						0.0172922

HOUSE\_LOAN\_COUNT 变量 IV 值为 0.0172922，认为该指标对是否逾期没有预测能力，应该被剔除；

COMMERCIAL\_LOAN\_COUNT 变量下的总的逾期客户为 1874，总的无逾期客户为 28097

COMMERCIAL_LOAN_COUNT	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>p<sub>i</sub></i>	<i>WE</i> 值	IV 值
0	1870	0.997866	27993	0.996299	0.0015716	0.0000025
1-3	4	0.00213447	101	0.00359469	-0.52124	0.0007611
总和						0.0007636

COMMERCIAL\_LOAN\_COUNT 变量 IV 值为 0.0007636，认为该指标对是否逾期没有预测能力，应该被剔除；

OTHER\_LOAN\_COUNT 变量下的总的逾期客户为 1874，总的无逾期客户为 28097

OTHER_LOAN_COUNT	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>p<sub>i</sub></i>	<i>WE</i> 值	IV 值
0	352	0.187834	3731	0.13279	0.34679	0.0190884
1	250	0.133404	3296	0.117308	0.12858	0.002070
2	237	0.126467	2825	0.100545	0.22938	0.005946
3	177	0.0944504	2411	0.085099	0.095940	0.0008290
4-6	335	0.178762	5020	0.178667	0.0005329	0.00000005
7-12	303	0.161686	5213	0.185536	-0.13759	0.00328147
13-24	158	0.0843116	3759	0.133787	-0.46173	0.0228438
25-342	62	0.0330843	1842	0.0655586	-0.68389	0.0222087
总和						0.0762675

OTHER\_LOAN\_COUNT 变量 IV 值为 0.0762675，认为该指标对是否逾期有较弱的预测能

力

FIRST\_LOAN\_OPEN\_MONTH 变量下的总的逾期客户为 1535，总的无逾期客户为 24695

FIRST_LOAN_OPEN_MONTH	逾期客户	<i>B<sub>1</sub></i>	无逾期客户	<i>p<sub>1</sub></i>	<i>WE</i> 值	IV 值
1994-2007 年	57	0.0371336	1517	0.0614294	-0.50337	0.0122298
2008-2011 年	119	0.0775244	2679	0.108483	-0.33600	0.0104024
2012-2013 年	101	0.065798	1990	0.0805831	-0.20270	0.0029969
2014 年	75	0.0488599	1445	0.0585139	-0.18031	0.0017407
2015 年	272	0.177199	4944	0.200202	-0.12206	0.0028078
2016 上半年	333	0.216938	4829	0.195546	0.10382	0.0022209
2016 下半年	353	0.229967	4633	0.187609	0.20358	0.0086233
2017 上半年	225	0.14658	2658	0.107633	0.30884	0.0120284
总和						0.0530501

FIRST\_LOAN\_OPEN\_MONTH 变量 IV 值为 0.0530501，认为该指标对是否逾期有较弱的预测能力

LOANCARD\_COUNT 变量下的总的逾期客户为 1874，总的无逾期客户为 28097

LOANCARD_COUNT	逾期客户	<i>B<sub>1</sub></i>	无逾期客户	<i>p<sub>1</sub></i>	<i>WE</i> 值	IV 值
0-3	728	0.388474	6804	0.242161	0.47262	0.0691507
4-6	535	0.285486	7156	0.254689	0.11415	0.0035154
7-9	222	0.118463	5218	0.185714	-0.44960	0.0302362
10-15	259	0.138207	5781	0.205752	-0.39792	0.026877
16-116	130	0.0693073	3138	0.111685	-0.47622	0.0201508
总和						0.14993

LOANCARD\_COUNT 变量 IV 值为 0.14993，认为该指标对是否逾期预测能力一般

FIRST\_LOANCARD\_OPEN\_MONTH 变量下的总的逾期客户为 1862，总的无逾期客户为 27964

FIRST_LOANCARD_OPEN_MONTH	逾期客户	<i>B<sub>1</sub></i>	无逾期客户	<i>p<sub>1</sub></i>	<i>WE</i> 值	IV 值
1998-2007 年	26	0.0139635	1205	0.0430911	-1.1269	0.0328231

2007-2009 年	201	0.107948	5700	0.2033834	-0.63565	0.0609493
2010-2011 年	193	0.103652	4122	0.147404	-0.35214	0.0154066
2012 年	190	0.102041	3085	0.11032	-0.078016	0.0006459
2013 年	221	0.11869	3244	0.116006	0.022867	0.0000614
2014 年	272	0.146079	3629	0.129774	0.11836	0.0019297
2015 年	288	0.154672	3299	0.117973	0.27085	0.0099401
2016 年	435	0.23362	3499	0.125125	0.62438	0.0677419
2017 上半年	36	0.019334	231	0.00826062	0.85037	0.0094165
总和						0.198915

**FIRST\_LOANCARD\_OPEN\_MONTH 变量 IV 值为 0.198915，认为该指标对是否逾期预测能力一般**

STANDARD\_LOANCARD\_COUN 变量下的总的逾期客户为 1874，总的无逾期客户为 28097

STANDARD_LOANCARD_COUN	逾期客户	<i>B<sub>1</sub></i>	无逾期客户	<i>p<sub>1</sub></i>	<i>WE</i> 值	IV 值
0	1741	0.929029	25092	0.0893049	0.039498	0.0014211
1	99	0.0528282	2065	0.0734954	-0.33018	0.0068239
2	24	0.0128068	577	0.020536	-0.47220	0.0036497
3	7	0.00373533	241	0.00857743	-0.83130	0.0040252
4-12	3	0.00160085	122	0.0043421	-0.99782	0.0027353
总和						0.0186552

STANDARD\_LOANCARD\_COUN 变量 IV 值为 0.0186552，认为该指标对是否逾期没有预测能力，应该被剔除；

FIRST\_SL\_OPEN\_MONTH 变量下的总的逾期客户为 258，总的无逾期客户为 5335

FIRST_SL_OPEN_MONTH	逾期客户	<i>B<sub>1</sub></i>	无逾期客户	<i>p<sub>1</sub></i>	<i>WE</i> 值	IV 值
1993-2007 年	20	0.0775194	639	0.119775	-0.43509	0.0183849
2008-2012 年	85	0.329457	2178	0.408247	-0.21443	0.0168947
2013-2014 年	90	0.348837	1300	0.243674	0.35877	0.03773
2015 年	26	0.100775	544	0.101968	-0.017768	0.000014
2016 年	37	0.143411	574	0.107591	0.28737	0.0102936



总和						0.0833171
----	--	--	--	--	--	-----------

FIRST\_SL\_OPEN\_MONTH 变量 IV 值为 0.0833171，认为该指标对是否逾期有较弱的预测能力

TYPE\_DW 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

TYPE_DW	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>p<sub>i</sub></i>	<i>W<sub>i</sub></i> 值	IV 值
未结清贷款信息汇总	1392	0.742796	23273	0.828899	-0.10968	0.0094435
未销户贷记卡信息汇总、未销户准贷记卡信息汇总	485	0.258805	4801	0.170994	0.41444	0.0363927
总和						0.0458362

TYPE\_DW 变量 IV 值为 0.0458362，认为该指标对是否逾期有较弱的预测能力

FINANCE\_CORP\_COUNT 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

FINANCE_CORP_COUNT	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>p<sub>i</sub></i>	<i>W<sub>i</sub></i> 值	IV 值
1	559	0.298292	7062	0.251523	0.17054	0.0079762
2	484	0.258271	6592	0.234783	0.095348	0.0022396
3	333	0.177685	5219	0.185882	-0.045043	0.0003688
4	196	0.104589	3644	0.129786	-0.21786	0.0054387
5	141	0.0752401	2437	0.086797	-0.14289	0.0016513
6	75	0.0400213	1549	0.0551697	-0.32100	0.0048626
7-15	86	0.0458911	1574	0.0560601	-0.20015	0.0020354
总和						0.0245725

FINANCE\_CORP\_COUNT 变量 IV 值为 0.0245725，认为该指标对是否逾期有较弱的预测能力

FINANCE\_ORG\_COUNT 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

FINANCE_ORG_COUNT	逾期客户	<i>B<sub>i</sub></i>	无逾期客户	<i>p<sub>i</sub></i>	<i>W<sub>i</sub></i> 值	IV 值
1	556	0.296692	7021	0.250062	0.17098	0.0079728
2	480	0.256137	6509	0.231827	0.099721	0.0024242

3	334	0.178228	5185	0.184671	-0.035509	0.0002288
4	197	0.105123	3631	0.129323	-0.20718	0.0050139
5	119	0.0635005	2407	0.0857285	-0.30014	0.0066715
6	88	0.0469584	1566	0.0557753	-0.17207	0.0015171
7-15	100	0.0533618	1758	0.0626135	-0.15989	0.0014792
总和						0.0253074

FINANCE\_ORG\_COUNT 变量 IV 值为 0.0253074，认为该指标对是否逾期有较弱的预测能力

ACCOUNT\_COUNT 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

ACCOUNT_C OUNT	逾期客户	<i>B?</i>	无 逾 期 客 户	<i>Pr</i>	<i>WE</i> 值	IV 值
1	433	0.231057	5569	0.198347	0.15264	0.0049928
2	393	0.209712	4918	0.175161	0.18003	0.0062201
3	290	0.154749	4041	0.143926	0.072509	0.000785
4	208	0.110993	3070	0.109342	0.014981	0.0000247
5	135	0.0720384	2315	0.0824518	-0.13501	0.0014060
6-8	220	0.117396	4189	0.149197	-0.23971	0.0076231
9-105	191	0.101921	3975	0.1411575	-0.32863	0.0130315
总和						0.034083

ACCOUNT\_COUNT 变量 IV 值为 0.034083，认为该指标对是否逾期有较弱的预测能力

CREDIT\_LIMIT 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

CREDIT_LIMI T	逾期客户	<i>B?</i>	无 逾 期 客 户	<i>Pr</i>	<i>WE</i> 值	IV 值
0-1000	53	0.0282818	331	0.011789	0.87505	0.014432
1001-5000	363	0.193703	3333	0.118709	0.48965	0.0367209
5001-10000	297	0.158485	3539	0.126046	0.22901	0.007429
10001-20000	370	0.197439	5108	0.181928	0.081815	0.001269
20001-40000	328	0.175027	5286	0.188268	-0.072928	0.000966
40001-60000	147	0.0784418	2681	0.0954847	-0.19664	0.0033518
60001-9824000	316	0.168623	7799	0.277772	-0.49913	0.0544797
总和						0.118648

**CREDIT\_LIMIT 变量 IV 值为 0.118648，认为该指标对是否逾期预测能力一般**

MAX\_CREDIT\_LIMIT\_PER\_ORG 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

MAX_CREDIT_LIMIT_PER_ORG	逾期客户	$B_i$	无逾期客户	$p_i$	$WE$ 值	IV 值
0	1393	0.74333	23278	0.829077	-0.10917	0.009361
1-10000	243	0.129669	1963	0.06999169	0.61771	0.0369107
10001-500000	238	0.127001	2836	0.101008	0.22900	0.0059523
总和						0.0522244

MAX\_CREDIT\_LIMIT\_PER\_ORG 变量 IV 值为 0.0522244，认为该指标对是否逾期有较弱的预测能力

MIN\_CREDIT\_LIMIT\_PER\_ORG 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

MIN_CREDIT_LIMIT_PER_ORG	逾期客户	$B_i$	无逾期客户	$p_i$	$WE$ 值	IV 值
0	1426	0.760939	23681	0.843431	-0.10292	0.0084904
1-10000	380	0.202775	3733	0.132956	0.42208	0.0294692
10001-500000	68	0.036286	663	0.0236136	0.42961	0.0054442
总和						0.0434037

MIN\_CREDIT\_LIMIT\_PER\_ORG 变量 IV 值为 0.0434037，认为该指标对是否逾期有较弱的预测能力

BALANCE 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

BALANCE	逾期客户	$B_i$	无逾期客户	$p_i$	$WE$ 值	IV 值
0	500	0.266809	5047	0.179756	0.39493	0.0343803
1-5000	342	0.182497	3864	0.137622	0.28223	0.0126652
5001-10000	251	0.133938	3345	0.119137	0.11711	0.0017334
10001-20000	265	0.141409	4408	0.156997	-0.10457	0.0016301
20001-50000	259	0.138207	4976	0.177227	-0.24868	0.0097034
50001-9824000	257	0.13714	6437	0.229262	-0.51387	0.0473387
总和						0.107451

**BALANCE 变量 IV 值为 0.107451，认为该指标对是否逾期的预测能力一般**

USED\_CREDIT\_LIMIT 变量下的总的逾期客户为 1874，总的无逾期客户为 28077

USED_CREDIT_LIMIT	逾期客户	$B_i$	无逾期客户	$p_i$	$WE$ 值	IV 值
0	1475	0.787086	23660	0.842683	-0.068252	0.003795

1-5000	118	0.0629669	940	0.0334794	0.63168	0.0186267
5001-10000	62	0.0330843	710	0.0252876	0.26874	0.0020953
10001-20000	124	0.0661686	1429	0.0508958	0.26243	0.004008
20001-50000	89	0.047492	1137	0.0404958	0.15936	0.001115
100001-368124	6	0.0032017	201	0.0071589	-0.80467	0.0031842
总和						0.0328238

USED\_CREDIT\_LIMIT 变量 IV 值为 0.0328238，认为该指标对是否逾期有较弱的预测能力

LATEST_6M_USED_AVG_AMOUNT	逾期客户	<del>BI</del>	无逾期客户	<del>PI</del>	<del>WE</del> 值	IV 值
0	259	0.138207	2625	0.0934929	0.39087	0.017477
1-500	377	0.201174	4274	0.152224	0.27882	0.013648
501-1000	224	0.11953	3416	0.121665	-0.017704	0.0000378
1001-2000	309	0.164145	4721	0.168145	-0.019559	0.000064
2001-5000	314	0.167556	6766	0.24098	-0.36340	0.0266821
5001-345186	391	0.208645	6275	0.223493	-0.068746	0.0010207
总和	1874		28077			0.0589296

LATEST\_6M\_USED\_AVG\_AMOUNT 变量 IV 值为 0.0589296，认为该指标对是否逾期有较弱的预测能力

贷款逾期笔数	逾期客户	无逾期客户	IV 值
0	864	13924	
1	251	3201	
2	65	723	
3	24	214	
4	6	111	
5-53	9	211	
总和	1219	18384	0.0190801

贷款逾期笔数变量 IV 值为 0.0190801，认为该指标对是否逾期没有预测能力，应该被剔除；

贷款逾期月份数	逾期客户	无逾期客户	IV 值
0	864	13924	
1	132	1720	
2	80	810	
3	31	444	
4	26	289	

5	20	224	
6	14	151	
7-57	52	822	
总和	1219	18384	0.0181986

贷款逾期月份数变量 IV 值为 0.0181986，认为该指标对是否逾期没有预测能力，应该被剔除；

贷款单月最高逾期总额	逾期客户	无逾期客户	IV 值
0	963	14979	
1-100	13	141	
101-500	74	594	
501-1000	35	516	
1001-5000	90	1378	
5001-4009542	44	776	
总和	1219	18384	0.0206489

贷款单月最高逾期总额变量 IV 值为 0.0206489，认为该指标对是否逾期有较弱的预测能力

最大贷款时长	逾期客户	无逾期客户	IV 值
0	864	13924	
1	263	3658	
2	32	304	
3	22	208	
4-6	24	132	
7	14	158	
总和	1219	18384	0.0256722

最大贷款时长变量 IV 值为 0.0256722，认为该指标对是否逾期有较弱的预测能力

贷记卡逾期账户数	逾期客户	无逾期客户	IV 值
0	101	1077	
1	540	6748	
2	279	4131	
3	135	2654	
4	80	1635	
5	35	932	
6	26	607	

7	14	292	
8-17	9	308	
总和	1219	18384	0.0654731

贷记卡逾期账户数变量 IV 值为 0.0654731，认为该指标对是否逾期有较弱的预测能力

贷记卡逾期月份数	逾期客户	无逾期客户	IV 值
0	495	5764	
1	69	684	
2	45	342	
3-5	44	379	
6-10	26	219	
11-50	13	190	
总和	1219	18384	0.0587859

贷记卡逾期月份数变量 IV 值为 0.0587859，认为该指标对是否逾期有较弱的预测能力

贷记卡单月最高逾期总额	逾期客户	无逾期客户	IV 值
0	540	6126	
1-100	53	1168	
101-200	44	713	
201-400	59	967	
401-1000	113	1912	
1001-2000	102	1923	
2001-4000	106	2022	
4001-6000	57	1018	
6001-10000	55	979	
10001-403678	90	1556	
总和	1219	18384	0.0568487

贷记卡单月最高逾期总额变量 IV 值为 0.0568487，认为该指标对是否逾期有较弱的预测能力

最大贷记卡时长	逾期客户	无逾期客户	IV 值
0	106	1109	
1	600	9780	
2	211	3697	

3	87	1432	
4	64	692	
5	43	429	
6	22	244	
7	86	1001	
总和	1219	18384	0.033117

最大贷记卡时长变量 IV 值为 0.033117，认为该指标对是否逾期有较弱的预测能力

准贷记卡 60 天以上透支账 户数	逾期客户	无逾期客 户	IV 值
0	1207	18041	
1	11	338	
2	1	5	
总和	1219	18384	0.00734698

准贷记卡 60 天以上透支账户数变量 IV 值为 0.00734698，认为该指标对是否逾期有较弱的预测能力

准贷记卡 60 天以上透支月 份数	逾期客户	无逾期客 户	IV 值
0	1182	17804	
1	9	150	
2	4	87	
3	2	76	
4	4	49	
5	5	38	
6-59	13	180	
总和	1219	18384	0.0086823

准贷记卡 60 天以上透支月份数变量 IV 值为 0.0086823，认为该指标对是否逾期有较弱的预测能力

准贷记卡 60 天以上透支单 月最高透支余 额	逾期客户	无逾期客 户	IV 值

0	1183	17862	
1-1000	2	163	
1001-221409	34	359	
总和	1219	18384	0.0151738

准贷记卡 60 天以上透支单月最高透支余额变量 IV 值为 0.0151738，认为该指标对是否逾期没有预测能力

最大准贷记卡 时长	逾期客户	无逾期客 户	IV 值
0	1182	17804	
3	13	232	
4	2	97	
5	4	68	
6	5	46	
7	13	137	
总和	1219	18384	0.0108146

最大准贷记卡时长变量 IV 值为 0.0108146，认为该指标对是否逾期没有预测能力

贷款	逾期客户	无逾期客 户	IV 值
呆账	4	43	
逾期	38	155	
正常	1493	24501	
总和	1535	24699	0.0260918

贷款变量 IV 值为 0.0260918，认为该指标对是否逾期有较弱的预测能力

贷记卡	逾期客户	无逾期客 户	IV 值
冻结	29	359	
正常	1811	27241	
止付、呆账	32	451	
总和	1872	28051	0.000590873

贷记卡变量 IV 值为 0.000590873，认为该指标对是否逾期有较弱的预测能力

finance\_org 变量下的总的坏样本为 361，总的好样本为 152059

finance_org	坏样本	<i>B?</i>	好样本	<i>P?</i>	<i>W?</i> 值	IV 值
-------------	-----	-----------	-----	-----------	-------------	------



A、B、C类贷款机构	46	0.12424	16254	0.106893	0.17569	0.00360717
D、E、F类贷款机构	26	0.0720222	16178	0.106393	-0.39017	0.0134103
G、H、I类贷款机构	26	0.0720222	16045	0.105518	-0.38191	0.0127925
J、K、L类贷款机构	29	0.0803324	17263	0.113528	-0.34588	0.0114818
M、N、O类贷款机构	61	0.16975	20620	0.133238	0.23762	0.00849176
P、Q、R类贷款机构	66	0.182825	18806	0.123676	0.38087	0.0231199
S、T、U类贷款机构	31	0.0858726	15855	0.104269	-0.19411	0.00357082
V、W、X、Y、Z类贷款机构	76	0.210526	30606	0.201277	0.044928	0.000415546
总和						0.0768897

finance\_org 变量 IV 值为 0.0768897，认为该指标对是否逾期有较弱的预测能力

type\_dw 变量下的总的坏样本为 361，总的好样本为 151698

type_dw	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
个人经营性贷款	44	0.121884	15803	0.103927	0.15938	0.00286198
个人汽车贷款、个人商用房（包括商住两用）贷款	3	0.00831025	2272	0.0149416	-0.5866	0.00389031
个人消费贷款	237	0.65651	110804	0.728691	-0.10431	0.00752937
个人住房贷款	20	0.0554017	3267	0.0214851	0.94725	0.0321275
个人住房公积金贷款	1	0.00277008	231	0.0015915	0.60073	0.000751473
农户贷款	10	0.0277008	777	0.00510986	1.6903	0.0381853
其他贷款	46	0.127424	18544	0.121953	0.043886	0.000240106
总和						0.0855861

type\_dw 变量 IV 值为 0.0855861，认为该指标对是否逾期有较弱的预测能力

guarantee\_type 变量下的总的坏样本为 361，总的好样本为 151698

guarantee_type	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
保证	15	0.0415512	2700	0.0177985	0.84781	0.0201349
抵押担保	17	0.0470914	4207	0.0277327	0.52948	0.01025
农户联保	3	0.00831025	173	0.00114042	1.9861	0.0142399
其他担保	7	0.0193906	5258	0.034661	-0.58083	0.00886945

信用/免担保	301	0.833795	123942	0.817031	0.020310	0.000340478
质押（含保证金）担保、组合（含保证）担保	3	0.00831025	914	0.00602513	0.32155	0.000734782
组合（不含保证）担保	15	0.0415512	14504	0.095611	-0.83336	0.0450513
总和						0.0996237

guarantee\_type 变量 IV 值为 0.0996237，认为该指标对是否逾期有较弱的预测能力

payment\_rating 变量下的总的坏样本为 361，总的好样本为 151698

payment_rating	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
按半年归还	1	0.00277008	15	0.0000988807	3.3327	0.00890237
按季归还	4	0.0110803	686	0.00452214	0.89619	0.00587736
按年归还、按其他方式归还	1	0.0027008	7143	0.047087	-2.8331	0.125555
按月归还	335	0.927978	130862	0.862648	0.073001	0.00476913
不定期归还	9	0.0249307	11124	0.0733299	-1.0789	0.0522162
一次性归还	11	0.0304709	1868	0.0123139	0.90604	0.016451
总和						0.213771

payment\_rating 变量 IV 值为 0.213771，认为该指标对是否逾期预测能力一般

payment\_cyc 变量下的总的坏样本为 352，总的好样本为 132109

payment_cyc	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
1-6 期	46	0.130682	27676	0.209494	-0.47193	0.0371935
7-12 期	185	0.525568	73873	0.559182	-0.061995	0.00208391
13-18 期	21	0.0596591	5133	0.0388543	0.42883	0.00892169
19-24 期	43	0.122159	10699	0.00892169	0.41105	0.016924
25-36 期	26	0.0738636	5982	0.0452808	0.48934	0.0139867
36 期以上	31	0.0880682	8746	0.0662029	0.28539	0.00624006
总和						0.0853498

payment\_cyc 变量 IV 值为 0.0853498，认为该指标对是否逾期有较弱的预测能力

class5\_state 变量下的总的坏样本为 307，总的好样本为 151698

class5_state	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
正常	209	0.680782	151428	0.99822	-0.38273	0.12494
关注、可疑、次级	97	0.315961	168	0.00110746	5.6535	1.78004
未知	1	0.00325733	102	0.00672389	1.5778	0.00407859
总和						1.90561

class5\_state 变量 IV 值为 1.90561，认为该指标对是否逾期有较强的预测能力

credit\_limit\_amount 变量下的总的坏样本为 361，总的好样本为 151698

credit_limit_amount	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
0-1000	88	0.243767	23400	0.155254	0.45761	0.0409627
1001-2000	48	0.132964	26729	0.173799	-0.28153	0.0121721
2001-3000	45	0.124654	18026	0.118828	0.047861	0.000278817
3001-4000	20	0.0554017	9862	0.0650107	-0.15994	0.00153691
4001-6000	54	0.149584	21748	0.143364	0.042476	0.00026423
6001-10000	15	0.0415512	19401	0.127892	-1.1243	0.0970698
10000 以上	91	0.252078	32532	0.214452	0.16165	0.00608207
总和						0.158367

credit\_limit\_amount 变量 IV 值为 0.158367，认为该指标对是否逾期预测能力一般

balance 变量下的总的坏样本为 307，总的好样本为 151698

balance	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
0-300	26	0.0846906	11857	0.0781619	0.080222	0.000523744
301-600	45	0.14658	11672	0.0769423	0.64451	0.0448823
601-1000	52	0.169381	16571	0.109237	0.43863	0.0263813
1001-2000	54	0.175896	26323	0.173522	0.013585	0.0000322421
2001-5000	62	0.201954	36431	0.240155	-0.17324	0.0066179
5001-10000	16	0.0521173	20798	0.137101	-0.96722	0.0821987
10000 以上	52	0.169381	28046	0.18488	-0.087558	0.0013571
总和						0.161993

balance 变量 IV 值为 0.161993，认为该指标对是否逾期预测能力一般

remain\_payment\_cyc 变量下的总的坏样本为 301，总的好样本为 132109

remain_payment_cyc	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
0 至 3	61	0.202658	32521	0.246168	-0.19450	0.00846251

3 至 6	52	0.172757	22522	0.17048	0.013268	0.0000302122
7 至 9	64	0.212625	21446	0.162336	0.26986	0.0135711
10 至 12	52	0.172757	32147	0.243337	-0.34256	0.0241776
13 至 24	39	0.129568	11829	0.0895397	0.36952	0.0147915
25 至 360	33	0.109635	11644	0.0881393	0.21823	0.00469098
总和						0.0657238

remain\_payment\_cyc 变量 IV 值为 0.0657238，认为该指标对是否逾期有较弱的预测能力

scheduled\_payment\_amount 变量下的总的坏样本为 307，总的好样本为 151698

scheduled_payment_amount	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
0-100	77	0.250814	60233	0.397059	-0.45937	0.0671804
101-300	107	0.348534	29554	0.194821	0.58165	0.0894077
301-500	59	0.192182	16016	0.105578	0.59899	0.0518753
501-800	14	0.0456026	11901	0.0784519	-0.54252	0.0178214
801-1200	13	0.0423453	11626	0.0766391	-0.59325	0.0203448
1201-3000	18	0.0586319	17997	0.118637	-0.70479	0.042291
3000 以上	19	0.0618893	4371	0.0288138	0.76449	0.0252859
总和						0.314207

scheduled\_payment\_amount 变量 IV 值为 0.314207，认为该指标对是否逾期有较强的预测能力

贷款总天数变量下的总的坏样本为 361，总的好样本为 151698

贷款总天数	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
90 至 180 天	18	0.0498615	16244	0.107081	-0.76343	0.0437352
181 至 270 天	38	0.105263	14718	0.0970217	0.081529	0.000671914
271 至 360 天	39	0.108033	12533	0.0826181	0.26821	0.0068166
361 至 540 天	156	0.432133	68570	0.452017	-0.044985	0.000894469
541 至 720 天	17	0.0470914	5526	0.0364276	0.25676	0.00273806
721 至 1080 天	43	0.119114	10625	0.0700405	0.53100	0.026058
1080 天以上	50	0.138504	23482	0.154794	-0.11120	0.00181143
总和						0.0827257

贷款总天数变量 IV 值为 0.0827257，认为该指标对是否逾期有较弱的的预测能力

已还款天数变量下的总的坏样本为 307，总的好样本为 151698

已还款天数	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
0 至 90 天	121	0.394137	79592	0.524674	-0.28608	0.037344
91-180 天	62	0.201954	30604	0.201743	0.0010476	0.0000022153
181-360 天	73	0.237785	28731	0.189396	0.22753	0.0110098
361-540 天	25	0.0814332	6017	0.0396643	0.71933	0.0300457
541-720 天	7	0.0228013	2840	0.0187214	0.19715	0.000804352
720 天以上	19	0.0618893	3914	0.0258013	0.87492	0.0315742
总和						0.110778

已还款天数变量 IV 值为 0.110778，认为该指标对是否逾期的预测能力一般

最近还款日到应还款日的天数变量下的总的坏样本为 307，总的好样本为 151698

最近还款日到应还款日	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
-30 到 0 天	42	0.136808	103541	0.682547	-1.6073	0.877141
1-90 天	259	0.843648	45059	0.297031	1.0439	0.570613
91-180 天	2	0.00651466	1954	0.0128809	-0.68169	0.00433976
181-270 天	2	0.00651466	582	0.00383657	0.52948	0.00141798
271-360 天	1	0.00325733	347	0.00228744	0.35347	0.000342831
361 天以上	1	0.00325733	215	0.00141729	0.83216	0.00153121
总和						1.45539

最近还款日到应还款日变量 IV 值为 1.45539，认为该指标对是否逾期有较强的预测能力

本月应还款与本月实际还款的差额变量下的总的坏样本为 307，总的好样本为 151698

本月应还款与本月实际还款的差额	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
小于 0	32	0.104235	6760	0.0445622	0.84976	0.050707
0	24	0.0781759	141888	0.935332	-2.4819	2.12741
1-9999	246	0.801303	3025	0.0199409	3.6935	2.88593
10000-30000	5	0.0162866	25	0.00016480	4.5934	0.0740535
总和						5.1381

本月应还款与本月实际还款的差额变量 IV 值为 5.1381，认为该指标对是否逾期有较强的预测能力

(6) contest\_ext\_crd\_cd\_lnd 中的变量剔除步骤

finance\_org 变量下的总的坏样本为 361，总的好样本为 152059

finance_org	坏样本	$B_i$	好样本	$p_i$	$WE$ 值	IV 值
A、B、C 类贷款机构	46	0.12424	16254	0.106893	0.17569	0.00360717
D、E、F 类贷款机构	26	0.0720222	16178	0.106393	-0.39017	0.0134103
G、H、I 类贷款机构	26	0.0720222	16045	0.105518	-0.38191	0.0127925
J、K、L 类贷款机构	29	0.0803324	17263	0.113528	-0.34588	0.0114818
M、N、O 类贷款机构	61	0.16975	20620	0.133238	0.23762	0.00849176
P、Q、R 类贷款机构	66	0.182825	18806	0.123676	0.38087	0.0231199
S、T、U 类贷款机构	31	0.0858726	15855	0.104269	-0.19411	0.00357082
V、W、X、Y、Z 类贷款机构	76	0.210526	30606	0.201277	0.044928	0.000415546
总和						0.0768897

finance\_org 变量 IV 值为 0.0768897，认为该指标对是否逾期有较弱的预测能力

type\_dw 变量下的总的坏样本为 361，总的好样本为 151698

type_dw	坏样本	$B_i$	好样本	$p_i$	$WE$ 值	IV 值
个人经营性贷款	44	0.121884	15803	0.103927	0.15938	0.00286198
个人汽车贷款、个人商用房（包括商住两用）贷款	3	0.00831025	2272	0.0149416	-0.5866	0.00389031
个人消费贷款	237	0.65651	110804	0.728691	-0.10431	0.00752937
个人住房贷款	20	0.0554017	3267	0.0214851	0.94725	0.0321275
个人住房公积金贷款	1	0.00277008	231	0.0015915	0.60073	0.000751473
农户贷款	10	0.0277008	777	0.00510986	1.6903	0.0381853
其他贷款	46	0.127424	18544	0.121953	0.043886	0.000240106
总和						0.0855861

type\_dw 变量 IV 值为 0.0855861，认为该指标对是否逾期有较弱的预测能力

guarantee\_type 变量下的总的坏样本为 361，总的好样本为 151698

guarantee_type	坏样本	$B_i$	好样本	$p_i$	$WE$ 值	IV 值
保证	15	0.0415512	2700	0.0177985	0.84781	0.0201349
抵押担保	17	0.0470914	4207	0.0277327	0.52948	0.01025
农户联保	3	0.00831025	173	0.00114042	1.9861	0.0142399

其他担保	7	0.0193906	5258	0.034661	-0.58083	0.00886945
信用/免担保	301	0.833795	123942	0.817031	0.020310	0.000340478
质押（含保证金）担保、组合（含保证）担保	3	0.00831025	914	0.00602513	0.32155	0.000734782
组合（不含保证）担保	15	0.0415512	14504	0.095611	-0.83336	0.0450513
总和						0.0996237

guarantee\_type 变量 IV 值为 0.0996237，认为该指标对是否逾期有较弱的预测能力

payment\_rating 变量下的总的坏样本为 361，总的好样本为 151698

payment_rating	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
按半年归还	1	0.00277008	15	0.0000988807	3.3327	0.00890237
按季归还	4	0.0110803	686	0.00452214	0.89619	0.00587736
按年归还、按其他方式归还	1	0.0027008	7143	0.047087	-2.8331	0.125555
按月归还	335	0.927978	130862	0.862648	0.073001	0.00476913
不定期归还	9	0.0249307	11124	0.0733299	-1.0789	0.0522162
一次性归还	11	0.0304709	1868	0.0123139	0.90604	0.016451
总和						0.213771

payment\_rating 变量 IV 值为 0.213771，认为该指标对是否逾期预测能力一般

payment\_cyc 变量下的总的坏样本为 352，总的好样本为 132109

payment_cyc	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
1-6 期	46	0.130682	27676	0.209494	-0.47193	0.0371935
7-12 期	185	0.525568	73873	0.559182	-0.061995	0.00208391
13-18 期	21	0.0596591	5133	0.0388543	0.42883	0.00892169
19-24 期	43	0.122159	10699	0.00892169	0.41105	0.016924
25-36 期	26	0.0738636	5982	0.0452808	0.48934	0.0139867
36 期以上	31	0.0880682	8746	0.0662029	0.28539	0.00624006
总和						0.0853498

payment\_cyc 变量 IV 值为 0.0853498，认为该指标对是否逾期有较弱的预测能力

class5\_state 变量下的总的坏样本为 307，总的好样本为 151698

class5_state	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
正常	209	0.680782	151428	0.99822	-0.38273	0.12494
关注、可疑、次级	97	0.315961	168	0.00110746	5.6535	1.78004
未知	1	0.00325733	102	0.00672389	1.5778	0.00407859
总和						1.90561

class5\_state 变量 IV 值为 1.90561，认为该指标对是否逾期有较强的预测能力

credit\_limit\_amount 变量下的总的坏样本为 361，总的好样本为 151698

credit_limit_amount	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
0-1000	88	0.243767	23400	0.155254	0.45761	0.0409627
1001-2000	48	0.132964	26729	0.173799	-0.28153	0.0121721
2001-3000	45	0.124654	18026	0.118828	0.047861	0.000278817
3001-4000	20	0.0554017	9862	0.0650107	-0.15994	0.00153691
4001-6000	54	0.149584	21748	0.143364	0.042476	0.00026423
6001-10000	15	0.0415512	19401	0.127892	-1.1243	0.0970698
10000 以上	91	0.252078	32532	0.214452	0.16165	0.00608207
总和						0.158367

credit\_limit\_amount 变量 IV 值为 0.158367，认为该指标对是否逾期预测能力一般

balance 变量下的总的坏样本为 307，总的好样本为 151698

balance	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
0-300	26	0.0846906	11857	0.0781619	0.080222	0.000523744
301-600	45	0.14658	11672	0.0769423	0.64451	0.0448823
601-1000	52	0.169381	16571	0.109237	0.43863	0.0263813
1001-2000	54	0.175896	26323	0.173522	0.013585	0.0000322421
2001-5000	62	0.201954	36431	0.240155	-0.17324	0.0066179
5001-10000	16	0.0521173	20798	0.137101	-0.96722	0.0821987
10000 以上	52	0.169381	28046	0.18488	-0.087558	0.0013571
总和						0.161993

balance 变量 IV 值为 0.161993，认为该指标对是否逾期预测能力一般

remain\_payment\_cyc 变量下的总的坏样本为 301，总的好样本为 132109

remain_payment_cyc	坏样本	$B_i$	好样本	$P_i$	$WE$ 值	IV 值
0 至 3	61	0.202658	32521	0.246168	-0.19450	0.00846251



3 至 6	52	0.172757	22522	0.17048	0.013268	0.0000302122
7 至 9	64	0.212625	21446	0.162336	0.26986	0.0135711
10 至 12	52	0.172757	32147	0.243337	-0.34256	0.0241776
13 至 24	39	0.129568	11829	0.0895397	0.36952	0.0147915
25 至 360	33	0.109635	11644	0.0881393	0.21823	0.00469098
总和						0.0657238

remain\_payment\_cyc 变量 IV 值为 0.0657238，认为该指标对是否逾期有较弱的预测能力

scheduled\_payment\_amount 变量下的总的坏样本为 307，总的好样本为 151698

scheduled_payment_amount	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
0-100	77	0.250814	60233	0.397059	-0.45937	0.0671804
101-300	107	0.348534	29554	0.194821	0.58165	0.0894077
301-500	59	0.192182	16016	0.105578	0.59899	0.0518753
501-800	14	0.0456026	11901	0.0784519	-0.54252	0.0178214
801-1200	13	0.0423453	11626	0.0766391	-0.59325	0.0203448
1201-3000	18	0.0586319	17997	0.118637	-0.70479	0.042291
3000 以上	19	0.0618893	4371	0.0288138	0.76449	0.0252859
总和						0.314207

scheduled\_payment\_amount 变量 IV 值为 0.314207，认为该指标对是否逾期有较强的预测能力

贷款总天数变量下的总的坏样本为 361，总的好样本为 151698

贷款总天数	坏样本	$B_i$	好样本	$P_i$	$W_i$ 值	IV 值
90 至 180 天	18	0.0498615	16244	0.107081	-0.76343	0.0437352
181 至 270 天	38	0.105263	14718	0.0970217	0.081529	0.000671914
271 至 360 天	39	0.108033	12533	0.0826181	0.26821	0.0068166
361 至 540 天	156	0.432133	68570	0.452017	-0.044985	0.000894469
541 至 720 天	17	0.0470914	5526	0.0364276	0.25676	0.00273806
721 至 1080 天	43	0.119114	10625	0.0700405	0.53100	0.026058
1080 天以上	50	0.138504	23482	0.154794	-0.11120	0.00181143
总和						0.0827257

贷款总天数变量 IV 值为 0.0827257，认为该指标对是否逾期有较弱的的预测能力

已还款天数变量下的总的坏样本为 307，总的好样本为 151698

已还款天数	坏样本	$B_i$	好样本	$p_i$	$W_i$ 值	IV 值
0 至 90 天	121	0.394137	79592	0.524674	-0.28608	0.037344
91-180 天	62	0.201954	30604	0.201743	0.0010476	0.0000022153
181-360 天	73	0.237785	28731	0.189396	0.22753	0.0110098
361-540 天	25	0.0814332	6017	0.0396643	0.71933	0.0300457
541-720 天	7	0.0228013	2840	0.0187214	0.19715	0.000804352
720 天以上	19	0.0618893	3914	0.0258013	0.87492	0.0315742
总和						0.110778

已还款天数变量 IV 值为 0.110778，认为该指标对是否逾期的预测能力一般

最近还款日到应还款日的天数变量下的总的坏样本为 307，总的好样本为 151698

最近还款日到应还款日	坏样本	$B_i$	好样本	$p_i$	$W_i$ 值	IV 值
-30 到 0 天	42	0.136808	103541	0.682547	-1.6073	0.877141
1-90 天	259	0.843648	45059	0.297031	1.0439	0.570613
91-180 天	2	0.00651466	1954	0.0128809	-0.68169	0.00433976
181-270 天	2	0.00651466	582	0.00383657	0.52948	0.00141798
271-360 天	1	0.00325733	347	0.00228744	0.35347	0.000342831
361 天以上	1	0.00325733	215	0.00141729	0.83216	0.00153121
总和						1.45539

最近还款日到应还款日变量 IV 值为 1.45539，认为该指标对是否逾期有较强的预测能力

本月应还款与本月实际还款的差额变量下的总的坏样本为 307，总的好样本为 151698

本月应还款与本月实际还款的差额	坏样本	$B_i$	好样本	$p_i$	$W_i$ 值	IV 值
小于 0	32	0.104235	6760	0.0445622	0.84976	0.050707
0	24	0.0781759	141888	0.935332	-2.4819	2.12741
1-9999	246	0.801303	3025	0.0199409	3.6935	2.88593
10000-30000	5	0.0162866	25	0.00016480	4.5934	0.0740535
总和						5.1381

本月应还款与本月实际还款的差额变量 IV 值为 5.1381，认为该指标对是否逾期有较强的预测能力