

番茄学院

# 课程：催收评分卡全流程开发

主讲老师：番茄风控老骑士—陈Sir



# 课程：催收评分卡全流程开发

适应对象	风控模型er、策略er，当时适用于想从事模型和了解模型的同学
课程亮点	风控老骑士亲自教授，项目数据+代码演练、注重模型的真实的业务应用场景
课程安排	2020年2月23日 AM9:30-AM12:00 PM2:30-5:30 （一天）
软件工具	三大软件全打击====》 <u>SAS跟PYTHON代码实操+提供R代码+数据演练</u>
内容安排	线上直播授课+答疑

# 一.学完本次课程你将收获

- 1.会用python/sas/R,这三种工具建模
- 2.能看懂基础代码，手动分箱，看懂模型的指标
- 3.了解贷后模块的相关知识（催收手段、策略、催收的激励方案）

## 二.课前提前思考

本次新型冠状病毒这次凶猛来袭，这只黑天鹅打乱了常规的金融秩序，贷后部门里之前做好的催收评分卡是否需要重新建模？

# 本次课程主要框架

40%

理论部分

## 第一部分：理论是实践的基础

- 第1章：简介——评分卡的意义
- 第2章：催收模型使用场景介绍
- 第3章：催收模型简要介绍
- 第4章：催收模型的开发流程
- 第5章：模型样本与数据维度
- 第6章：真实案例模拟场景

## 第二部分：实践、数据源代码演练

- 第7章：数据预处理
- 第8章：变量分箱
- 第9章：模型特征的选择
- 第10章：模型拟合
- 第11章：模型验证
- 第12章：模型的分数的转化

## 第三部分：项目实战、结合理论

- 第13章：催收手段介绍
- 第14章：催收模型的使用效果对比
- 第15章：催收模型的策略调整
- 第16章：催收模型的优化的方向
- 第17章：催收模块的其他内容
- 第18章：催收模型后续的内容

60%

实操部分

# 第一章：简介——评分卡的意义

信贷评分卡模型是量化信贷风险的一种工具，贷前评分卡可以作为审批策略供审批人员判断客户的风险程度一种工具，贷后评分卡也可以作为催收人员判断还款风险的一种工具。

信贷评分模型的计算方式主要是根据历史客户的人口统计的数据，如教育程度、收入水平、房产状态等信息，以及客户历史的逾期情况，结合人行PBOC等数据，运用特定算法，预测申请客户未来的信用风险。

在评分体系中，为了将客户风险加以量化和方便理解，最终将客户的信用风险依照违约概率分成不同的等级，其中A代表信用风险最低，字母越大风险越高。

# 第一章：简介——催收卡的分类

催收滚动率模型

还款率模型

失联模型

# 第一章：简介——催收卡的分类

未逾期人群或轻度逾期人群在变成轻度逾期或者重度逾期的概率

滚动率模型：

- M0到M1
- DPD1滚动到DPD10
- DPD1-4滚动到DPD5-20
- .....



本期目标设计  
---为了讨论方便在此将其称为催收卡

还款率模型： 预测客户在逾期后，经过催收，最终能催收来欠款的比例

失联模型： 在逾期阶段，对于尚能联系到的人群预测其未来失联的概率

区别在哪里？



# 第二章：催收模型简要介绍——滚动率模型

## 模型的目的：

用于提前催收，降低违约风险，提高还款率

根据逾期风险等级制定不同的催收策略

## 模型跑分的实际的运用场景：

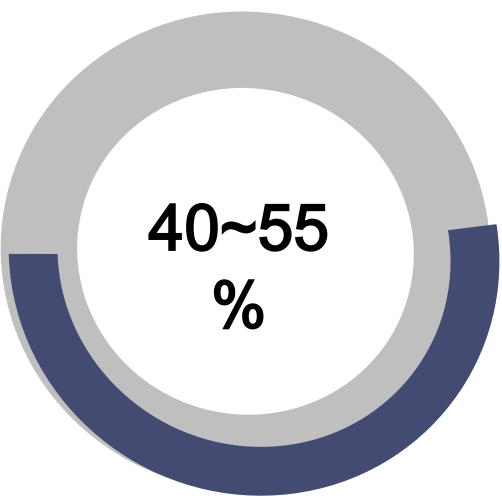
还款日前5-10日预测是否逾期

如M0-M1, M0-M1, M0-M3...

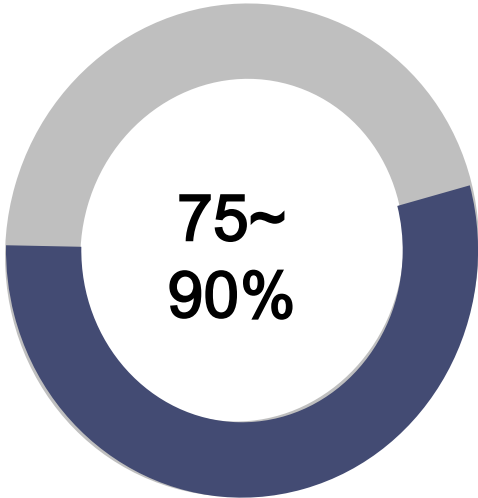
## 模型入参：

- 1.最大逾期天数
- 2.下P的次数
- 3.BP的次数
- 5.KPTP-RATE的比例
- 6.历史的总逾期天数
- .....

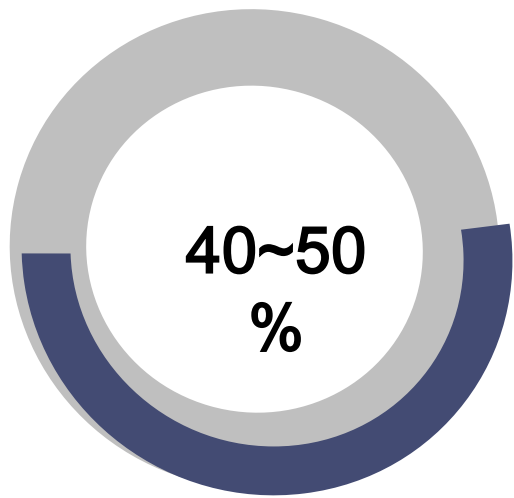
## 模型的指标的情况：



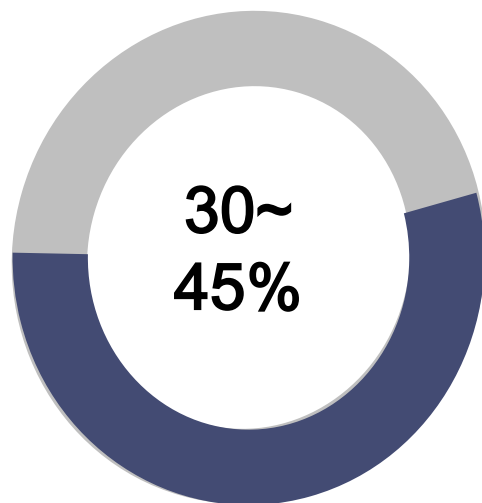
训练集KS值



训练集GINI



测试集KS值



验证集KS值



# 第二章：催收模型简要介绍——还款率模型

模型的目的：

用于精细化催收策略

如：IVR，预测式，预览式

模型跑分的实际的运用场景：

DPD 1-10模型

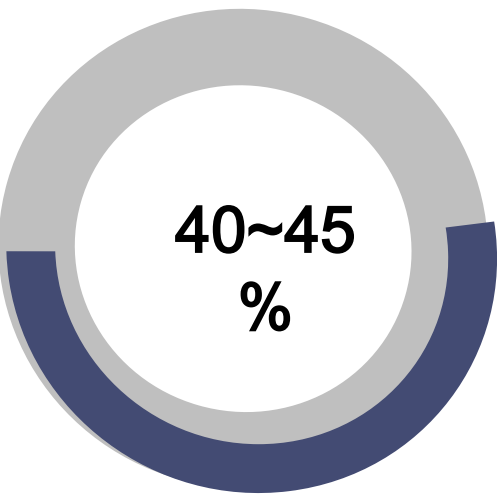
预测客户DPD 1-10的还款的概率

模型入参：

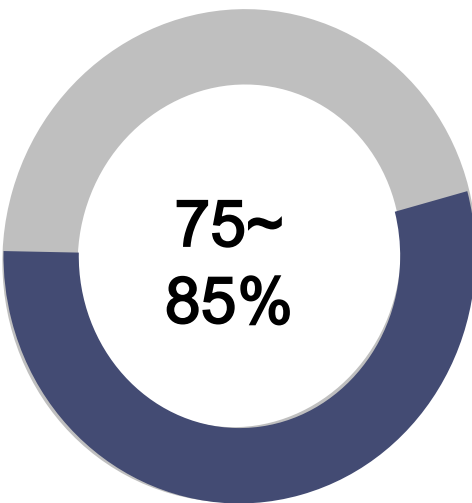
- 1. 通话总次数（近30天/60天/90天）
- 2. 客户所有未结清合同单逾期天数
- 3. 大于等于3的次数/所有还款期数
- 4. 历史有效联系人未接通次数
- 5. 本次合同较上期是否DPD>=10
- 6. 通话等于0的总次数（近15天）

.....

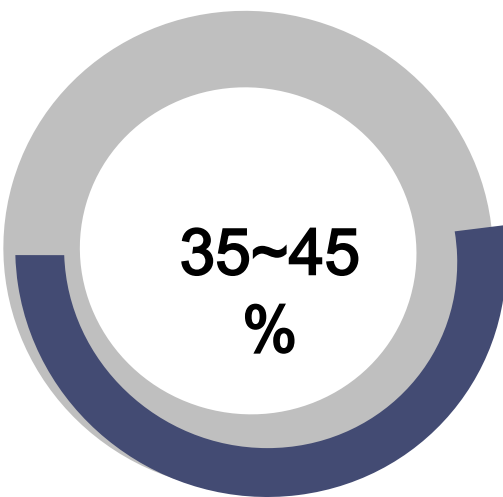
模型的指标的情况：



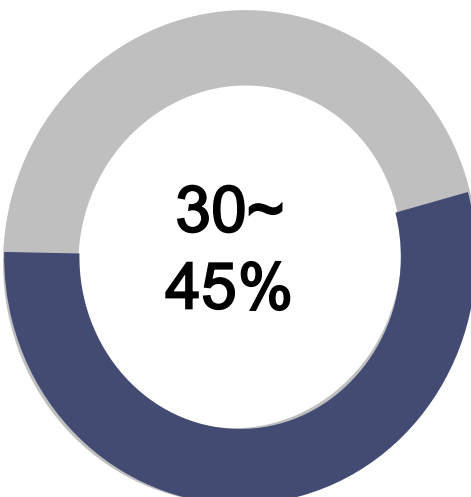
训练集KS值



训练集GINI



测试集KS值



验证集KS值

# 第二章：催收模型简要介绍——失联模型

### 模型的目的：

用于失联信息修复

用于提前委外

### 模型跑分的实际的运用场景：

在CPD=10时，预测未来5天后是否

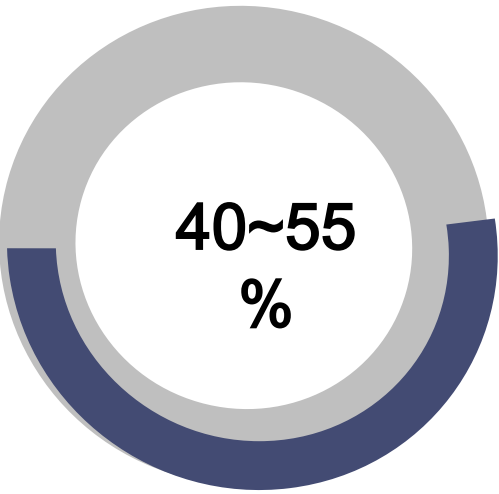
失联，即在CPD10的客户预测在

CPD16时是否失联

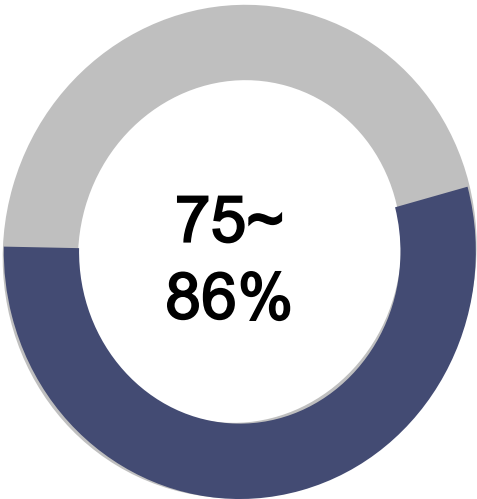
### 模型入参：

- 1.是否BP
- 2.当前与上次联络时间差
- 3.委外标识
- 4.与客户上次进入失联的时间差
- 5.年龄
- 6.最大贷款期数
- 7.联系人RPC数量
- 8.RPC是否有父亲，客户填写移动电话量
- 9.近一天本人电话催收次数
- 10.居住城市与银行卡城市是否一致
- 11.客户填写移动电话量
- 12.近一天本人电话催收次数
- 13.居住城市与银行卡城市是否一致
- .....

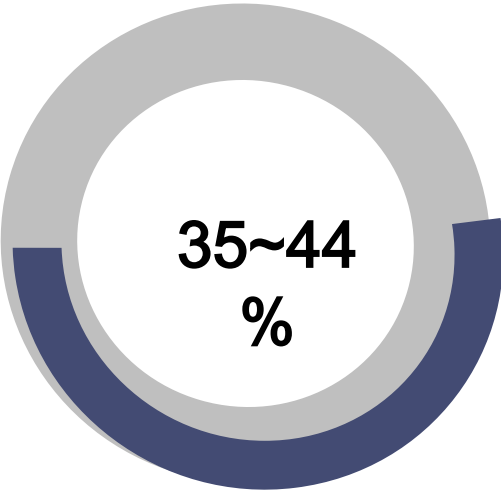
### 模型的指标的情况：



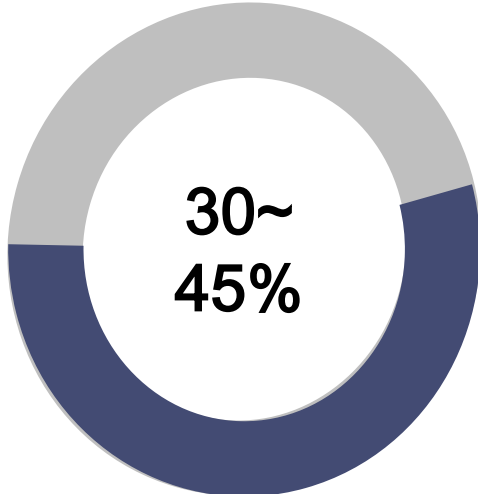
训练集KS值



训练集GINI



测试集KS值



验证集KS值

## 第三章：催收模型使用场景介绍

催收滚动率模型

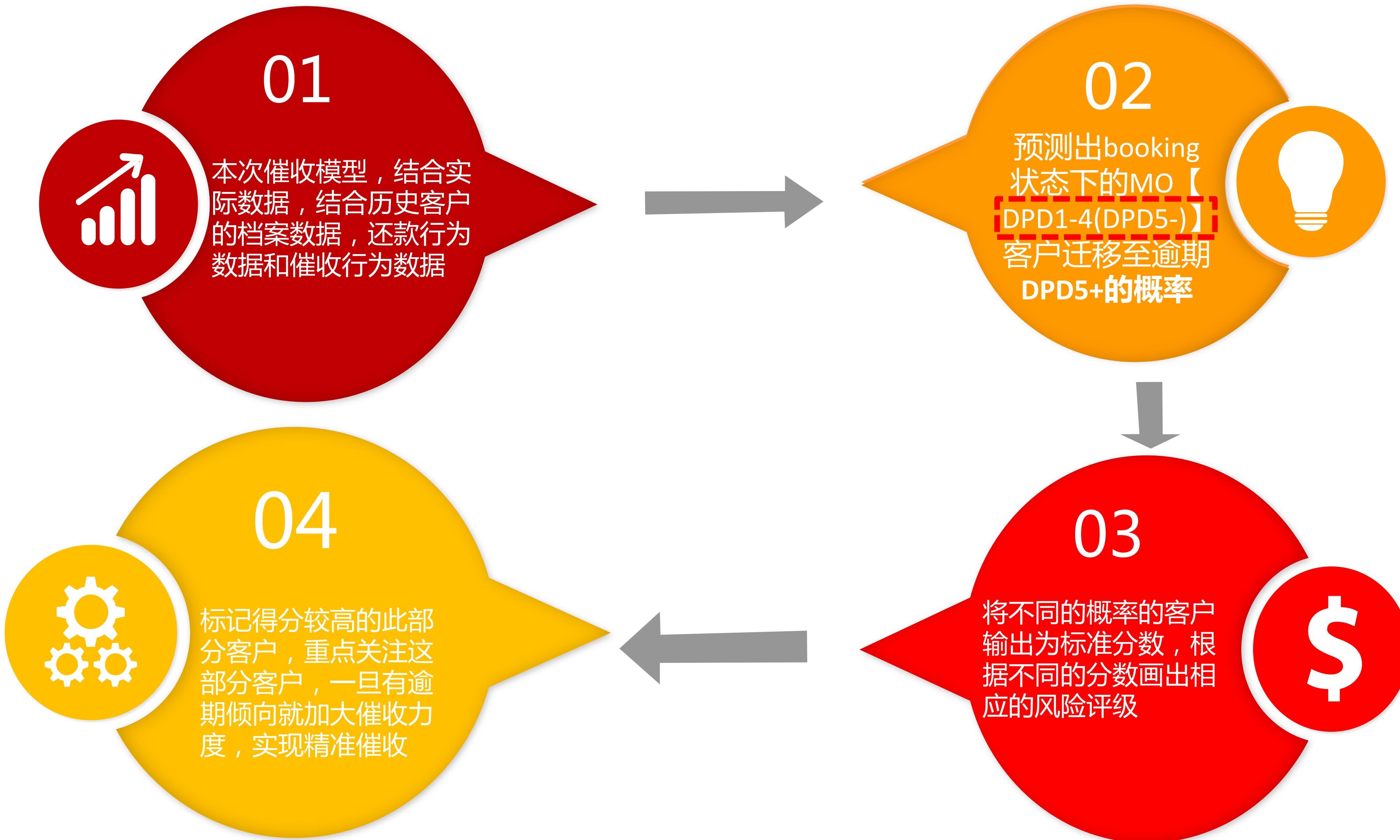


本期目标设计--为了讨论方便在此将其称为催收卡

还款率模型

失联模型

# 第三章：催收模型使用场景介绍



# 第三章：催收模型使用场景介绍—常见问题

- 本次催收评分卡内容课程：数据+代码+案例+贷后模块的常见内容

贷后  
模块

滚动  
模型

还款  
模型

失联  
模型

催收  
策略

委外  
催收

人工  
催收

1

C卡、B卡跟A卡的区别在哪里？

2

C卡的样本怎么选择？

3

C卡一般会生成多少样本？

4

C卡的观察期跟表现期怎么定义？

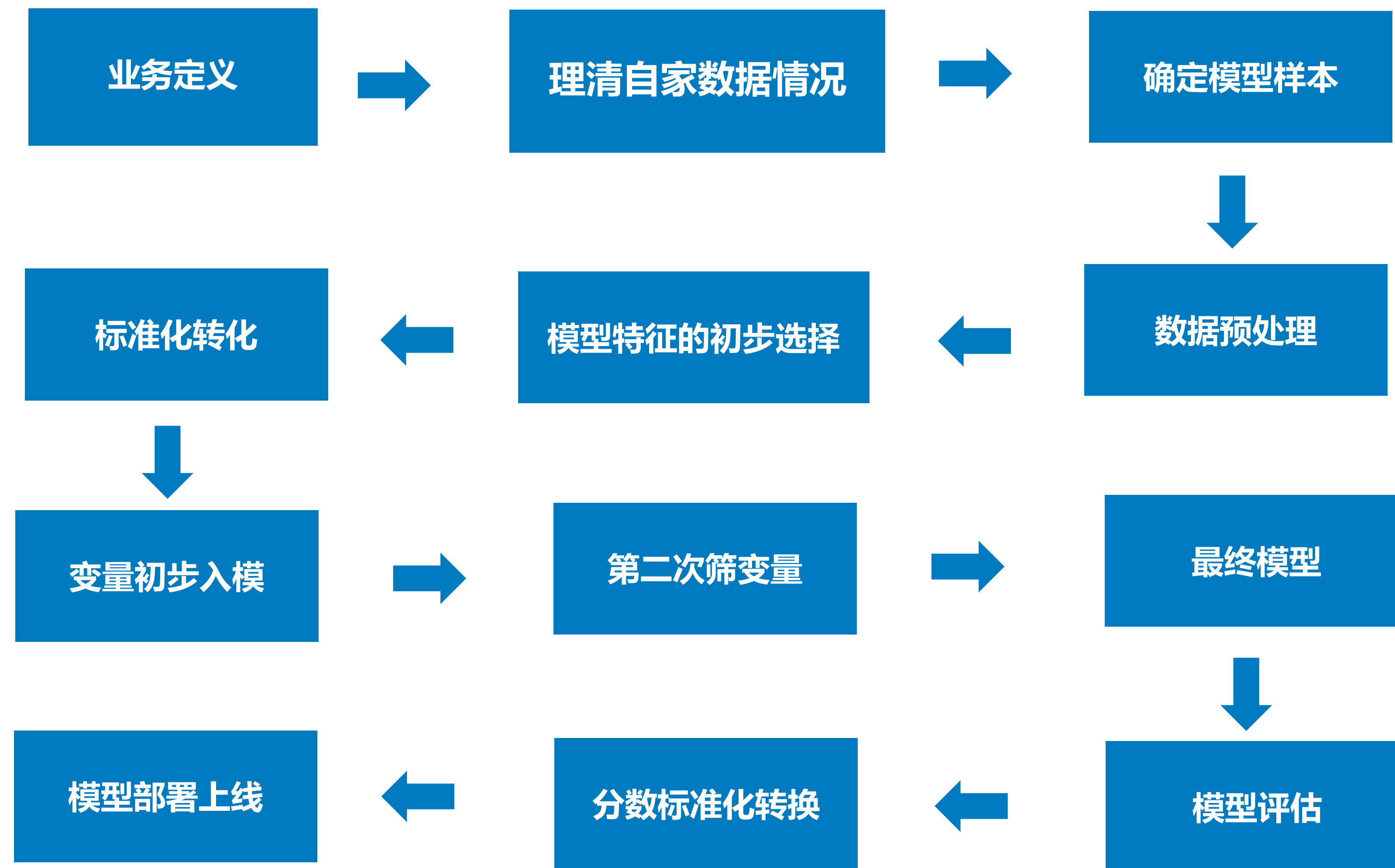
5

C卡开发的步骤有哪些？

6

C卡开发完如何结合策略同步上线？

## 第四章：催收模型的开发流程



➤ 谨记一点，模型是通过客群历史数据去预测未来的风险

# 第五章：模型样本与数据维度

## 滚动率模型

### □原则

- ✓ 根据实际的业务场景去倒推历史建模数据

### □选择窗口

- ✓ 建模样本选择窗口
- ✓ 观察期和表现期窗口



# 第五章：模型样本与数据维度

## 专有词汇说明

贷后管理	DPD	即合同的最早的一笔分期的逾期日期到现在的时间间隔（ Days past due ）
	CPD	客户逾期天数，与DPD相似。贷后管理的专有名词。历史经验设定逾期金额在50元以上的客户，才有价值通过人工进行催收。所以CPD是指贷后管理中，逾期金额在50元以上的客户的逾期天数。CPD的值取决于最早一期未还清的时间点。
	M1、M2、M3	逾期阶段，以30天为一个阶段。M1：DPD1-30；M2：DPD31-60；M3：DPD61-90
	Outbound/Inbound	电话呼出/电话呼入
	RPC	Right Public Contact，指有效的联系人，通过电话催收可以找到客户本人或直属亲属。
	PTP	Promise To Pay，通过电话催收，客户承诺在一定期限内归还一定数额的欠款，称之为承诺还款。值得注意的是，只有在RPC有效标识之后，才可以有PTP标识。
	KPTP	Kept Promise,K_PTP，客户按照约定还款。
	BP	Broken Promise，BP，承诺到期内，客户未按约定还款。
	RPC Ratio	联系RPC合同数/总合同数
	PTP Ratio	承诺还款合同数/联系到RPC的合同数
	KPTP Ratio	实际还款合同数/承诺还款合同数
	电话催收	催收人员通过电话方式联系欠款人督促其还款的操作称为电话催收，简称电催
	委外催收	通过将债务委托给其他公司或个人进行债务追讨，并按照一定比例支付追讨费用的催收方式
	上门催收	去到债务人在居住地址、工作地址等进行债务追讨
	法务催收	债权人通过起诉的方式追讨债务的行为称之为法务催收，又可简称为法催、诉讼催收等

# 第五章：模型样本与数据维度

计算机 > File (D:) > 催收卡 > 催收卡数据					
搜索 催收卡数据					
组织 包含到库中 共享 新建文件夹					
收藏夹	名称	修改日期	类型	大小	
下载	C卡 代码	2019/2/2 10:43	文件夹		
桌面	BP_INCALL	2019/1/25 18:09	Microsoft Excel 工作表	407 KB	
最近访问的位置	CONTACT_ALL	2019/1/18 13:56	Microsoft Excel 工作表	807 KB	
2345Downloads	CONTACT_ALL整理	2019/1/18 13:56	Microsoft Excel 工作表	933 KB	
	GB_fix	2019/1/25 18:09	Microsoft Excel 工作表	182 KB	
库	KPTP_ALL整理	2019/4/27 7:36	Microsoft Excel 工作表	5,754 KB	
视频	PTP_ALL整理	2019/4/30 7:19	Microsoft Excel 工作表	6,367 KB	
图片	催收字段	2019/1/18 15:42	Microsoft Excel 工作表	1,180 KB	
文档	放款金额	2019/2/1 14:17	Microsoft Excel 工作表	225 KB	
音乐	还款信息	2019/2/1 14:15	Microsoft Excel 逗号分隔值...	1,113 KB	
优酷影视库	客户档案数据 update	2019/1/25 18:09	Microsoft Excel 工作表	941 KB	
	历史放款数据Segment--新旧LOS	2019/2/12 14:56	Microsoft Excel 工作表	842 KB	
计算机	欠款信息	2019/4/30 7:19	Microsoft Excel 逗号分隔值...	805 KB	
本地磁盘 (C:)	申请日期	2019/3/20 16:12	Microsoft Excel 工作表	188 KB	
File (D:)	逾期次数	2019/1/25 18:09	Microsoft Excel 工作表	744 KB	
文档 (E:)	逾期天数	2019/4/7 21:22	Microsoft Excel 逗号分隔值...	884 KB	
	逾期天数0422	2019/4/22 13:54	Microsoft Excel 逗号分隔值...	1,030 KB	
网络					
17 个对象					
个性设置，点我看看					
CH S ? 18:07 2019/6/29					

## 第六章：案例模拟场景

### 案例模拟场景:

近期由于公司催收业务的回款越来越难，领导想对催收业务进行改革创新，开发一张客户滚动率模型来预测客户是否处在当前不逾期的情况，在未来就有可能逾期的风险，想对这部分客户重点关注。

### 实际应用场景：

此评分卡会在客户还款日10天前开始跑评分，想要在客群没有逾期的情况下M0预测其滚动到dpd5+的概率，对于评分特别低的客户就是我们的重点关注客户，进行催收强策略的干预。

# 第六章：真实案例模拟场景

每日定点评分卡

客群为DPD5-

7.1

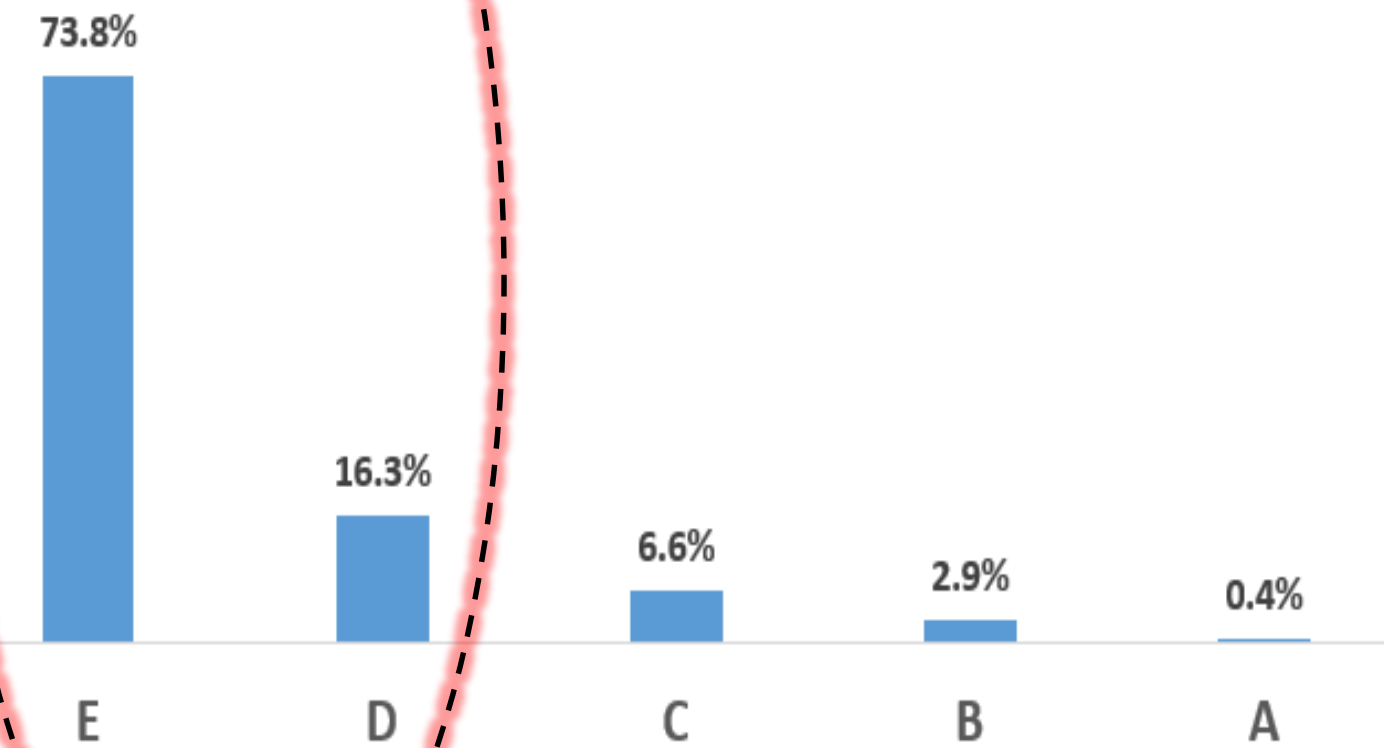
7.10

A1

A2

A3

.....



7.2

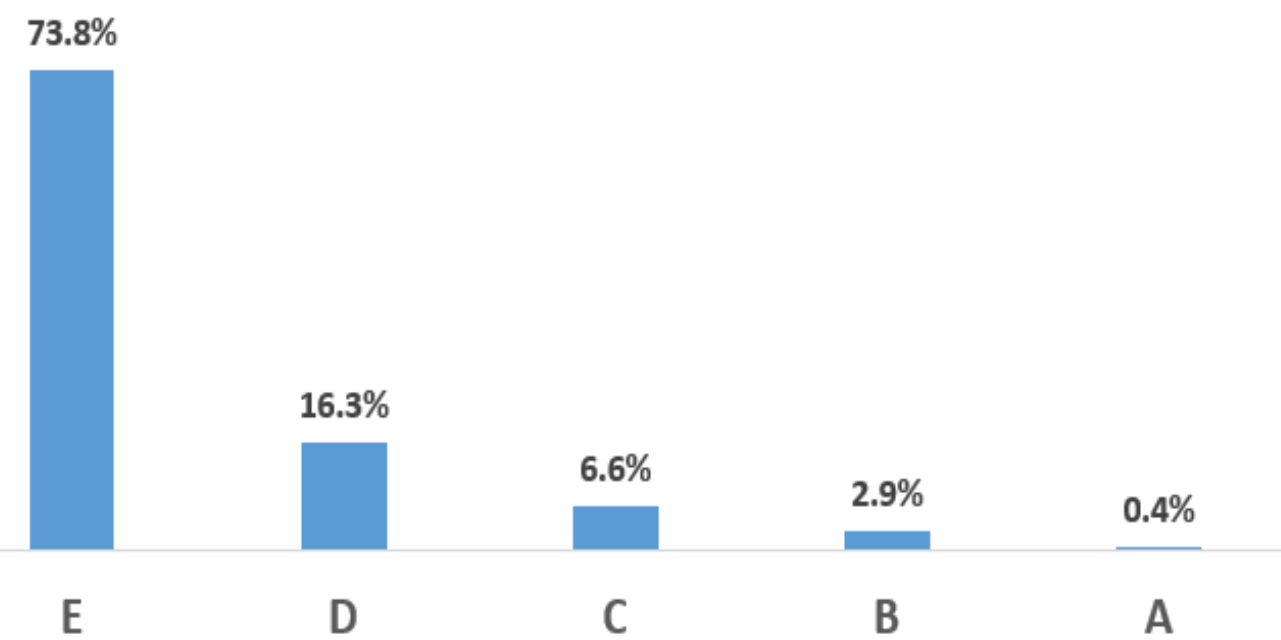
7.11

B1

B2

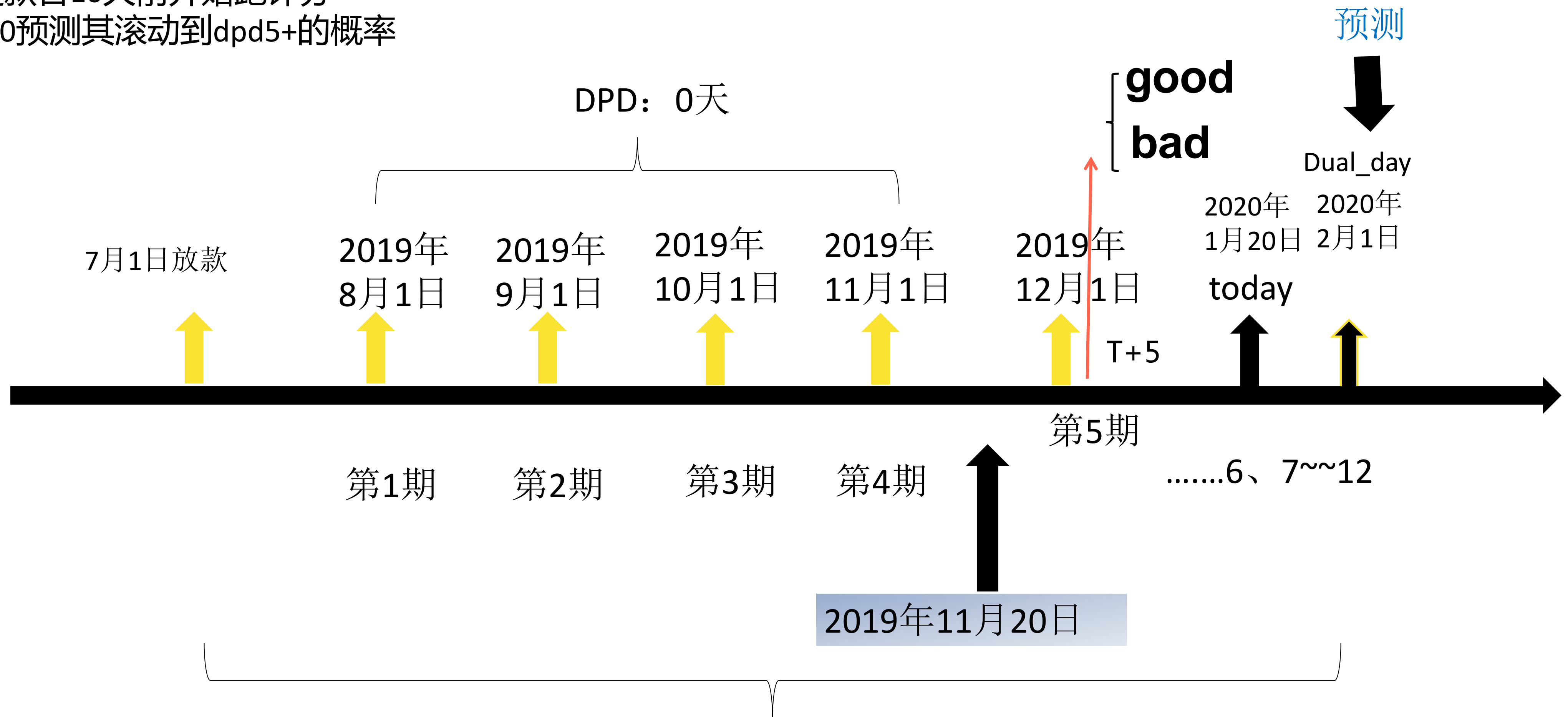
B3

.....



## 第六章：案例模拟场景

A客户：1.还款日10天前开始跑评分  
2.M0预测其滚动到dpd5+的概率

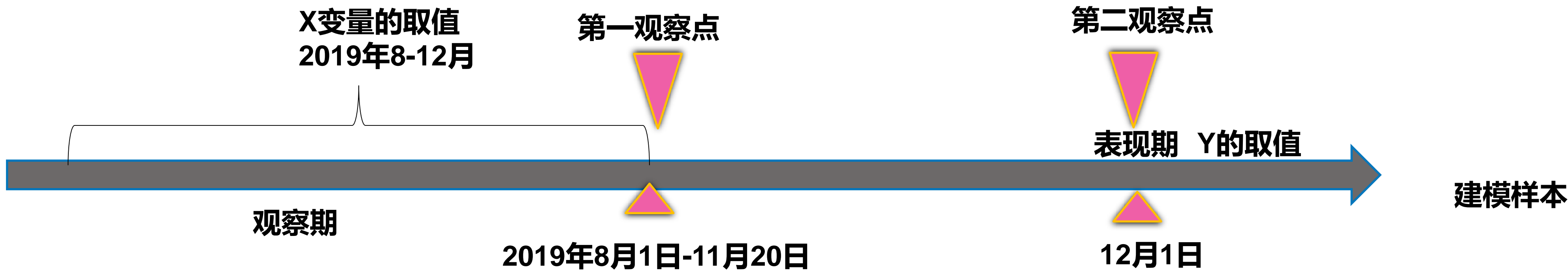


DPD:截止在理应还款日当天是否逾期  
MO：当前没有逾期

12期产品

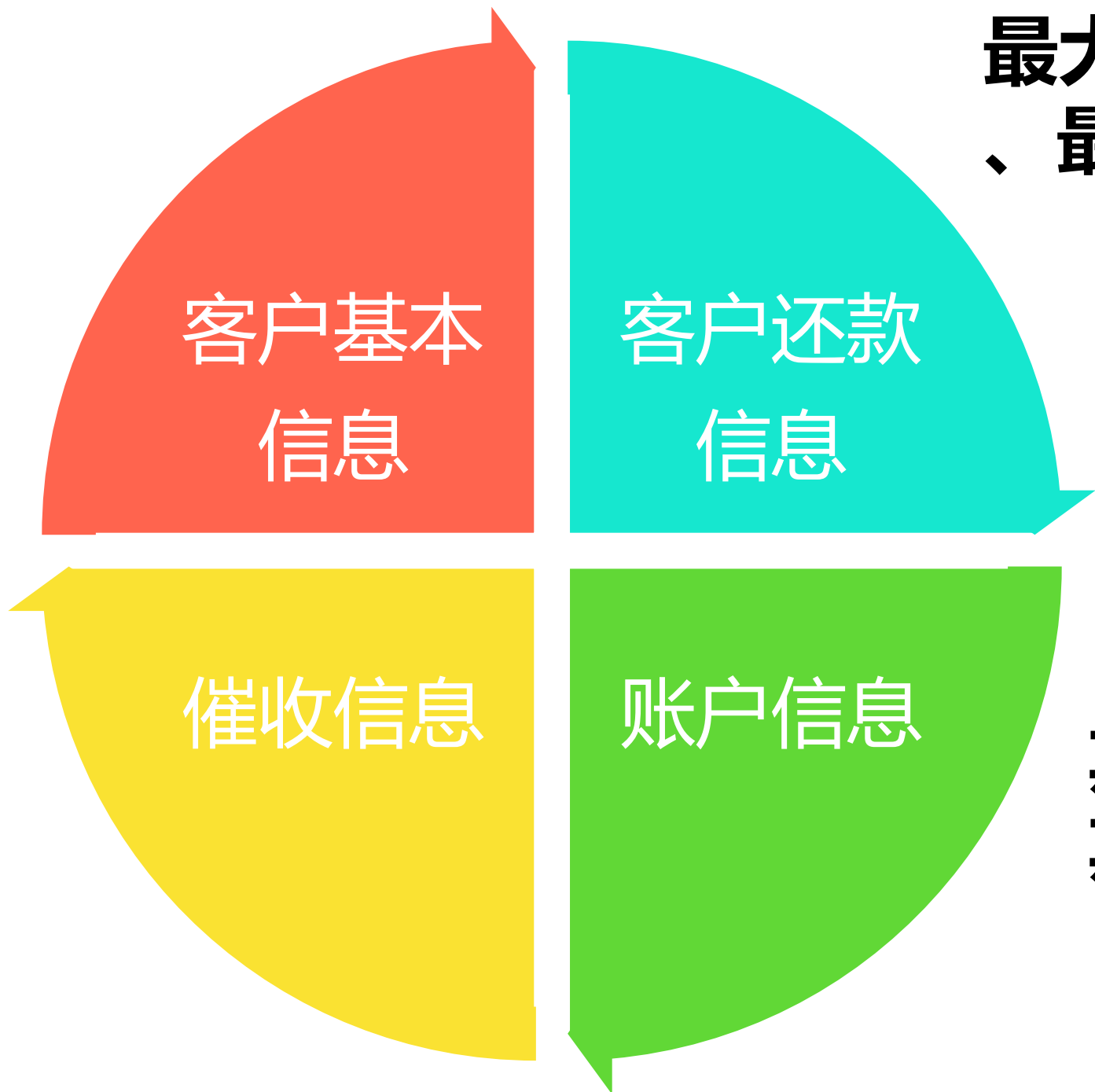


# 第六章：案例模拟场景



行业、年龄、性别、职业等

下P的次数  
BP的次数  
KPTP-RATE的比例  
历史的总逾期天数



最大逾期天数、历史最大逾期天数、逾期次数、最近一次还款日期、最近一次还款金额等

期款、分期期数、分期金额、贷款用途等

## 第六章：案例模拟场景

实际场景问题处理：

**问题1.客户C是在5月20日这天，不逾期的，但是在20日之前，他都是属于逾期的情况，也就是说c是在19日才还上上一期的欠款，那这种情况还能不能把c当成目标样本？**

**问题2：按照12月1日~12月30日作为一个时间窗口去卡客群的时候，发现客群的样本量不够，这个情况下应该怎么处理？**



# 第六章：案例模拟场景

LOAN_NO1	G/B	近一个月的窗口 的客群情况	数量	占比		LOAN_NO1	G/B	近三个月的窗口 的客群情况	数量	占比
'1000355895	GOOD	GOOD	9534	79.36%		'1000347106	GOOD	GOOD	10485	80.70%
'1000361702	GOOD	BAD	2479	20.64%		'1000355895	GOOD	BAD	2507	19.30%
'1000372336	GOOD	TOTAL	12013			'1000361702	GOOD	TOTAL	12992	
'1000448173	GOOD					'1000372336	GOOD			
'1000450601	GOOD					'1000448173	GOOD			
'1000451426	GOOD					'1000450601	GOOD			
'1000458465	GOOD					'1000451426	GOOD			
'1000497424	GOOD					'1000458465	GOOD			
'1000354693	GOOD					'1000497424	GOOD			
'1000354774	GOOD					'1000354693	GOOD			
'1000355015	GOOD					'1000354774	GOOD			
'1000356081	BAD					'1000355015	GOOD			
'1000356475	GOOD					'1000356081	BAD			
'1000357424	BAD					'1000356475	GOOD			
'1000357834	GOOD					'1000357424	BAD			
'1000358795	GOOD					'1000357834	GOOD			
'1000359864	GOOD					'1000358795	GOOD			
'1000360016	GOOD					'1000359864	GOOD			
'1000360113	GOOD					'1000360016	GOOD			
'1000360570	GOOD					'1000360113	GOOD			
'1000360860	GOOD					'1000360570	GOOD			
'1000361743	GOOD					'1000360860	GOOD			
'1000363096	GOOD					'1000361743	GOOD			
'1000363595	GOOD					'1000363096	GOOD			
'1000367740	GOOD					'1000363595	GOOD			
'1000368685	GOOD					'1000367740	GOOD			
'1000369093	GOOD					'1000368685	GOOD			
'1000370920	GOOD					'1000369093	GOOD			
'1000372996	GOOD					'1000370920	GOOD			
'1000375051	GOOD					'1000372996	GOOD			
'1000375766	GOOD					'1000375051	GOOD			
'1000375921	GOOD					'1000375766	GOOD			
'1000397224	BAD					'1000375921	GOOD			
'1000398212	GOOD					'1000392946	GOOD			
'1000399126	GOOD					'1000397224	BAD			
'1000399416	GOOD					'1000397255	GOOD			
'1000402580	GOOD					'1000397955	GOOD			

# 第七章：数据预处理

## 常见的需要处理的数据预处理：

**数据清洗数据整合：**数据整合的意思就是把本来一个客户多条的数据整合成一条。

**数据清洗：**包含的方面很多，譬如日期格式的清洗，公司名字的清洗，地址的清洗等等，在这里就介绍常用的几类清洗。

### 日期格式的清洗

譬如把“2016年4月6号”转化为“20160406”

### 字符/数值变量的清洗

比如学历里，大学本科跟本科属于同个意思，都是需要清洗为本科

# 第七章：数据预处理

## 1.缺失值

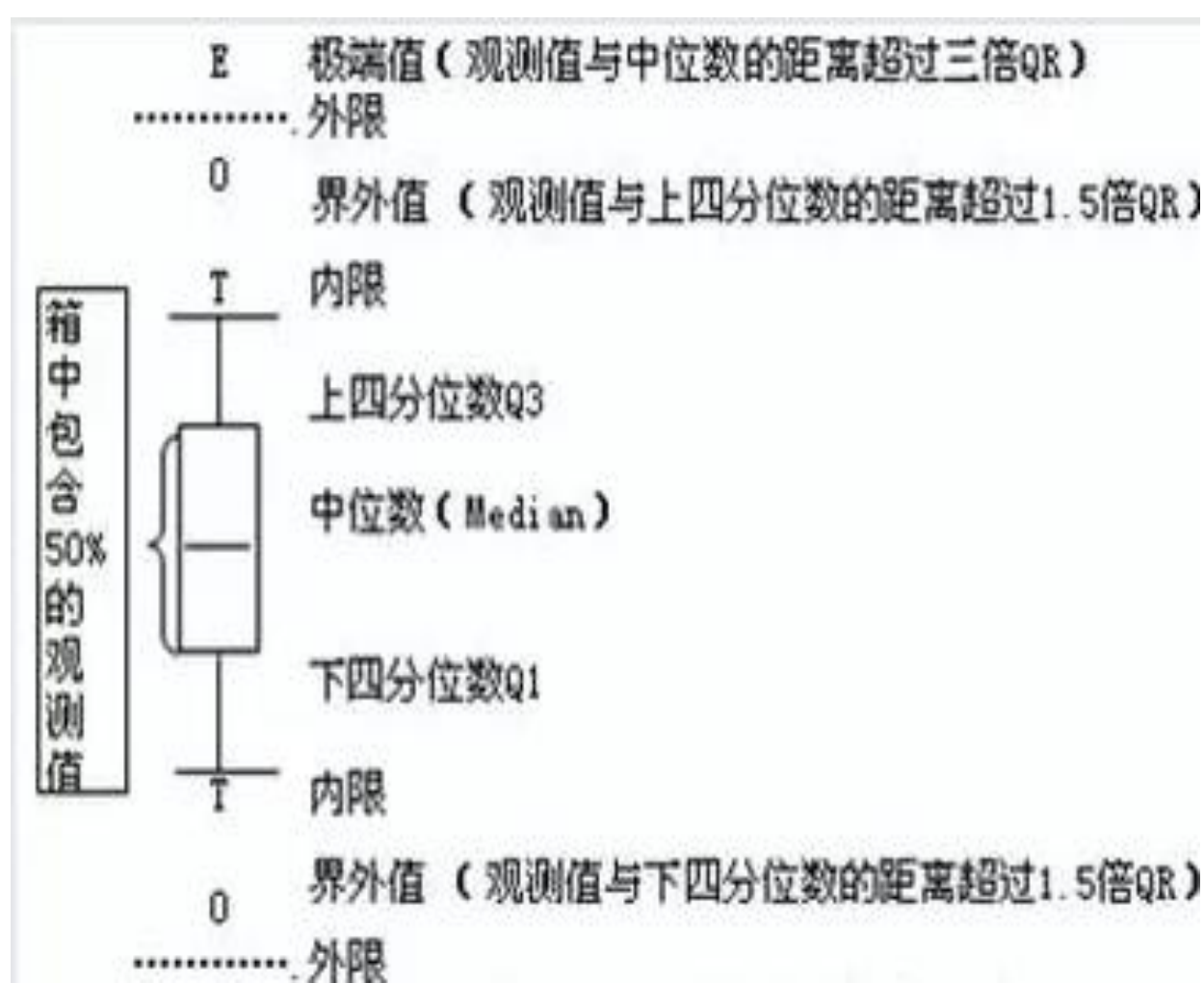
一般情况下，某个变量的缺失比例不要超过70%

## 2.离散值的单一程度

对某些离散值太多的变量，考虑剔除或者将类合并

## 3.异常值

异常值是偏离常规数值范围，特别严重的值，比如年龄大于100岁



# 第七章：数据预处理

## 特征工程的处理

### 常规方法：

**作差：**最近这笔贷款距上次的贷款有多长的时间差

**求和：**例如过去一段时间内的贷款次数的总和

**比例：**例如申请贷款的月还款本息与月收入的占比

**频率：**例如过去一段时间内的消费次数

**平均：**例如过去一段时间内平均每次信用卡取现额度

### 时间切片法：

时间切片内的逾期的次数

时间切片内下P的个数

时间切片内的联系的次数

**特征组合法：** 大专已婚... 本科未婚...

# 第七章：数据预处理

数据库	变量概要	具体变量
催收行为数据	BP情况	最近一期的BP次数
		最近二期的BP次数
		最近三期的BP次数
		总的BP次数
	PTP情况	最近一期的PTP
		最近二期的PTP
		最近三期的PTP
		总的PTP次数
	concontact次数情况	最近一期的concontact次数
		最近二期的concontact次数
		最近三期的concontact次数
		总的 concontact次数
	KPTP次数情况	最近一期的KPTP
		最近二期的KPTP
		最近三期的KPTP
		总的KPTP次数
	失联天数情况	最近一期的失联天数
		最近二期的失联天数
		最近三期的失联天数
		总的失联天数
	有效联系人数量情况	有效的直系亲属
		有效好友
		有效联系人的总数
		平均每次通话时长
	通话时长	总通话时长
	回电次数	平均每期通话次数
		总回电次数
	催收信息数量	最近一期催收信息数量
		总的催收信息数量
	催收函数量	总的催收函数量
		总的经手人数量

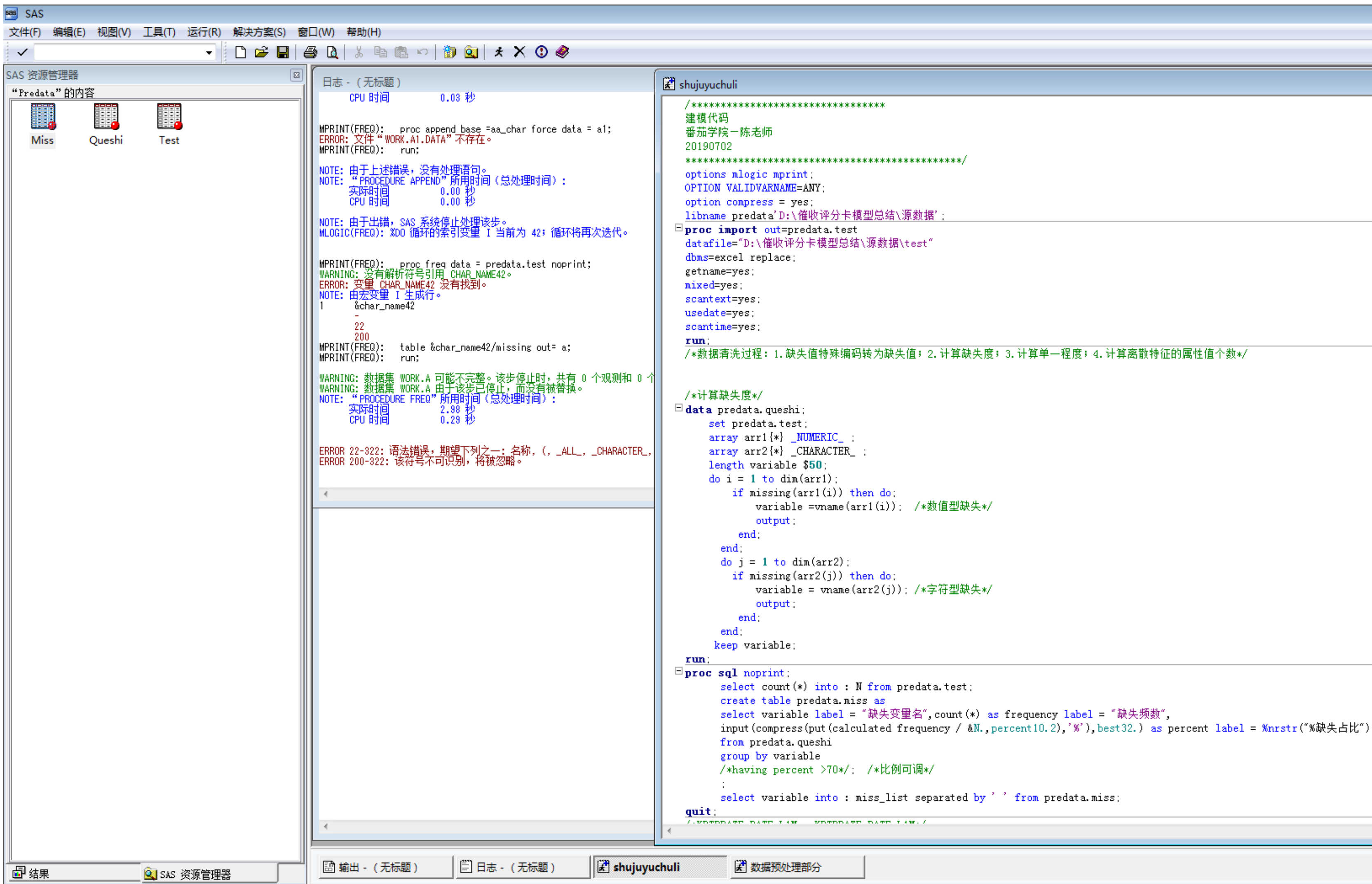
# 第七章：数据预处理

数据库	变量概要	具体变量
历史还款数据	逾期金额	最近一期逾期金额
		最近二期逾期金额
		最近三期逾期金融
		历史最大逾期金额
		历史合计逾期金额
	逾期天数	最近一期逾期天数
		最近二期逾期天数
		最近三期逾期天数
		历史最大逾期天数
		历史合计逾期天数
	逾期细分情况	已还三期，近3个月内曾经最大逾期天数
		已还六期，近6个月内曾经最大逾期天数
	提前还款时间情况	已还三期，近3个月内总逾期天数
		已还六期，近6个月内总逾期天数
		最近一期还款离扣款的累积天数
		最近二期还款离扣款的累积天数
		最近三期还款离扣款的累积天数
		历史最短的提前还款天数
		平均提前多少天还款



# 第七章：数据预处理（实操—sas&python）

## Sas实操，代码演示



The screenshot displays the SAS software interface. On the left, the 'SAS 资源管理器' (SAS Resource Manager) shows a project named 'Predata' with three datasets: 'Miss', 'Queshi', and 'Test'. The main window is divided into two panes. The left pane, titled '日志 - (无标题)' (Log - (Untitled)), shows the execution log. It contains several messages, including errors and warnings. The right pane, titled 'shujuyuchuli' (Data Processing), contains SAS code for data preprocessing. The code includes comments in Chinese, variable definitions, and data manipulation steps.

```
*****
建模代码
番茄学院—陈老师
20190702
*****/
options mlogic mprint;
OPTION VALIDVARNAME=ANY;
option compress = yes;
libname predata 'D:\催收评分卡模型总结\源数据';
proc import out=predata.test
datafile='D:\催收评分卡模型总结\源数据\test'
dbms=excel replace;
getname=yes;
mixed=yes;
scantext=yes;
usedate=yes;
scantime=yes;
run;
/*数据清洗过程：1. 缺失值特殊编码转为缺失值；2. 计算缺失度；3. 计算单一程度；4. 计算离散特征的属性值个数*/

/*计算缺失度*/
data predata.queshi;
set predata.test;
array arr1{*} _NUMERIC_ ;
array arr2{*} _CHARACTER_ ;
length variable $50;
do i = 1 to dim(arr1);
    if missing(arr1(i)) then do;
        variable = vname(arr1(i)); /*数值型缺失*/
        output;
    end;
end;
do j = 1 to dim(arr2);
    if missing(arr2(j)) then do;
        variable = vname(arr2(j)); /*字符型缺失*/
        output;
    end;
end;
keep variable;
run;
proc sql noprint;
select count(*) into : N from predata.test;
create table predata.miss as
select variable label = "缺失变量名", count(*) as frequency label = "缺失频数",
input(compress(put(calculated frequency / &N, percent10.2), '%'), best32.) as percent label = %nrstr("%缺失占比")
from predata.queshi
group by variable
/*having percent >70*/; /*比例可调*/
;
select variable into : miss_list separated by ' ' from predata.miss;
quit;
```

The log window shows the following messages:

```
MPRINT(FREQ): proc append base =aa_char force data = a1;
ERROR: 文件 "WORK.A1.DATA" 不存在。
MPRINT(FREQ): run;

NOTE: 由于上述错误，没有处理语句。
NOTE: "PROCEDURE APPEND" 所用时间 (总处理时间):
      实际时间      0.00 秒
      CPU 时间      0.00 秒

NOTE: 由于出错，SAS 系统停止处理该步。
MLOGIC(FREQ): %DO 循环的索引变量 1 当前为 42；循环将再次迭代。

MPRINT(FREQ): proc freq data = predata.test noprint;
WARNING: 没有解析符号引用 CHAR_NAME42。
ERROR: 变量 CHAR_NAME42 没有找到。
NOTE: 由宏变量 I 生成行。
1      &char_name42
      -
      22
      200
MPRINT(FREQ): table &char_name42/missing out= a;
MPRINT(FREQ): run;

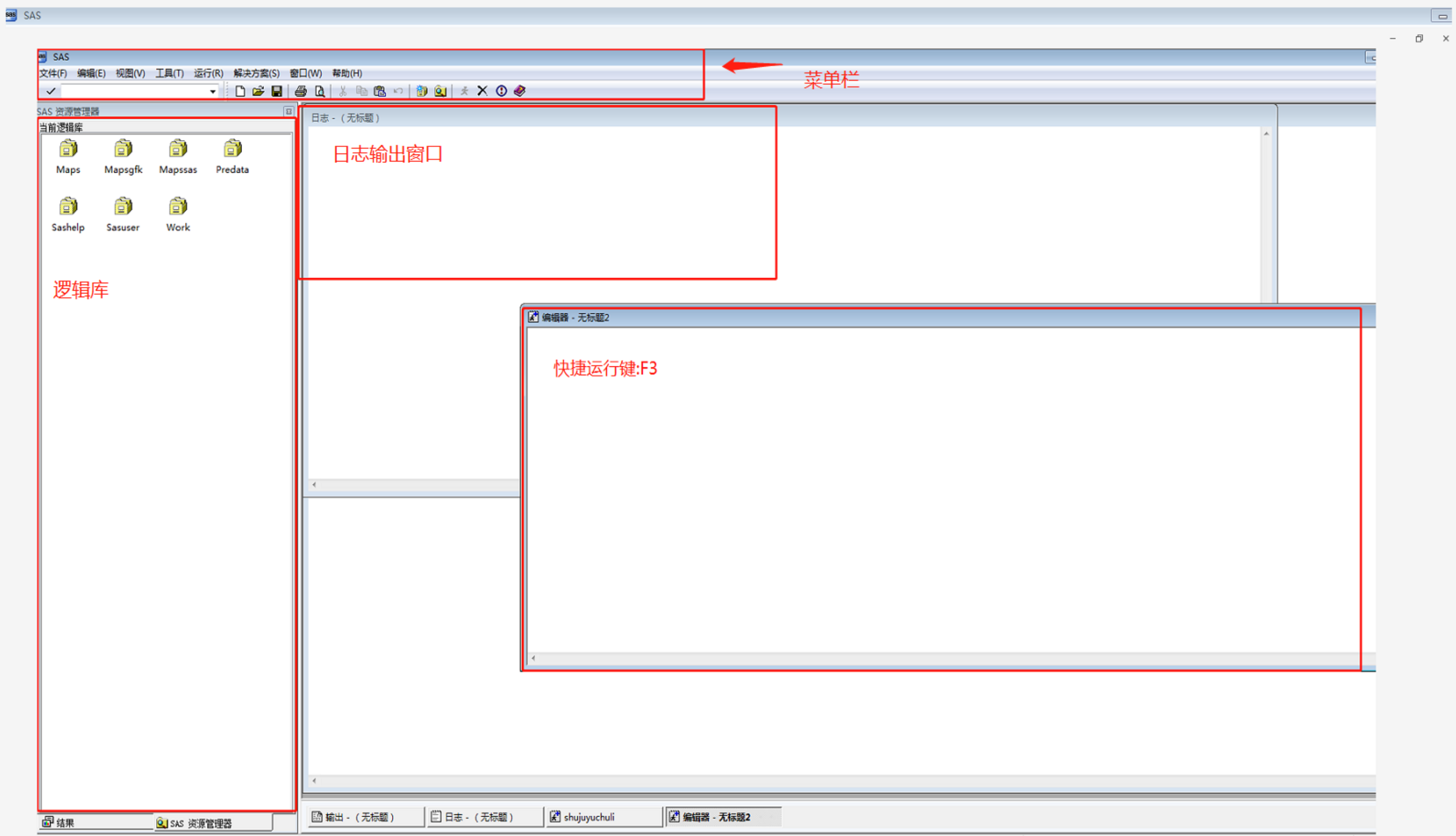
WARNING: 数据集 WORK.A 可能不完整。该步停止时，共有 0 个观测和 0 个
WARNING: 数据集 WORK.A 由于该步已停止，而没有被替换。
NOTE: "PROCEDURE FREQ" 所用时间 (总处理时间):
      实际时间      2.98 秒
      CPU 时间      0.29 秒

ERROR 22-322: 语法错误，期望下列之一：名称，(，_ALL_，_CHARACTER_，
ERROR 200-322: 该符号不可识别，将被忽略。
```



# 第七章：数据预处理—SAS

## Sas的介绍



# 第七章：数据预处理—python

## python的介绍

```
148 ## 对变量进行缺失值分析
149 missing = train_X.isnull().sum()
150 missing = missing[missing > 0]
151 missing.sort_values(inplace=True)
152 missing.plot.bar()
153
154 ## 只有两列变量有值，本方法中不对缺失值做填充处理，用区别于原始分布的数值代替
155 train_X['KPTP_RATE_L1M'].fillna(-999, inplace=True)
156 train_X['KPTP_RATE_L2M'].fillna(-999, inplace = True)
157 train_X['KPTP_RATE_L3M'].fillna(-999, inplace = True)
158 train_X['KPTP_RATE_L4M'].fillna(-999, inplace = True)
159 train_X['KPTP_RATE_L5M'].fillna(-999, inplace = True)
160 train_X['KPTP_RATE_L6M'].fillna(-999, inplace = True)
161 train_X['KPTP_RATE_L7M'].fillna(-999, inplace = True)
162 train_X['KPTP_RATE_L8M'].fillna(-999, inplace = True)
163 train_X['KPTP_RATE_L9M'].fillna(-999, inplace = True)
164 train_X['KPTP_RATE_L10M'].fillna(-999, inplace = True)
165 train_X['KPTP_RATE_L11M'].fillna(-999, inplace = True)
166 train_X['KPTP_RATE_L12M'].fillna(-999, inplace = True)
167 train_X['KPTP_RATE_SUM'].fillna(-999, inplace = True)
168 train_X['DPD_L3M_MAX'].fillna(-999, inplace = True)
169 train_X['DPD_L3M_SUM'].fillna(-999, inplace = True)
170 train_X['DPD_L6M_SUM'].fillna(-999, inplace = True)
171 train_X['DPD_L6M_MAX'].fillna(-999, inplace = True)
172
173
174
175
176 ## 观察异常值
177
178
179 sns.boxplot(train['PERIOD_PERCEN'])
180 sns.boxplot(train['b'])
181 sns.boxplot(train['c'])
182 sns.boxplot(train['d'])
183 sns.boxplot(train['e'])
184 sns.boxplot(train['f'])
185
186 ## 箱型图识别异常值标准
187 outile_var = [
188 'PERIOD_PERCEN',
189 'CONTACT_SUM',
190 'CONTACT_L1M',
191 'CONTACT_L2M',
192 'CONTACT_L3M'
```

```
238 ### a是否为异常值
239 for col in outile_var:
240     qua_U = train[col].quantile(0.75)
241     qua_L = train[col].quantile(0.25)
242     IQR = qua_U - qua_L
243
244     mean_values = train[col].mean()
245     std_values = train[col].std()
246
247     qua_U1 = train[col].quantile(0.95)
248
249     outile1 = qua_U + 1.5 * IQR
250     outile2 = qua_U1
251     outile3 = mean_values + 3 * std_values
252
253     median = train[col].median()
254     train_X.loc[((train_X[col] > outile1) & (train_X[col] > outile2) & (train_X[col] > outile3)), col] = median
255
256
```

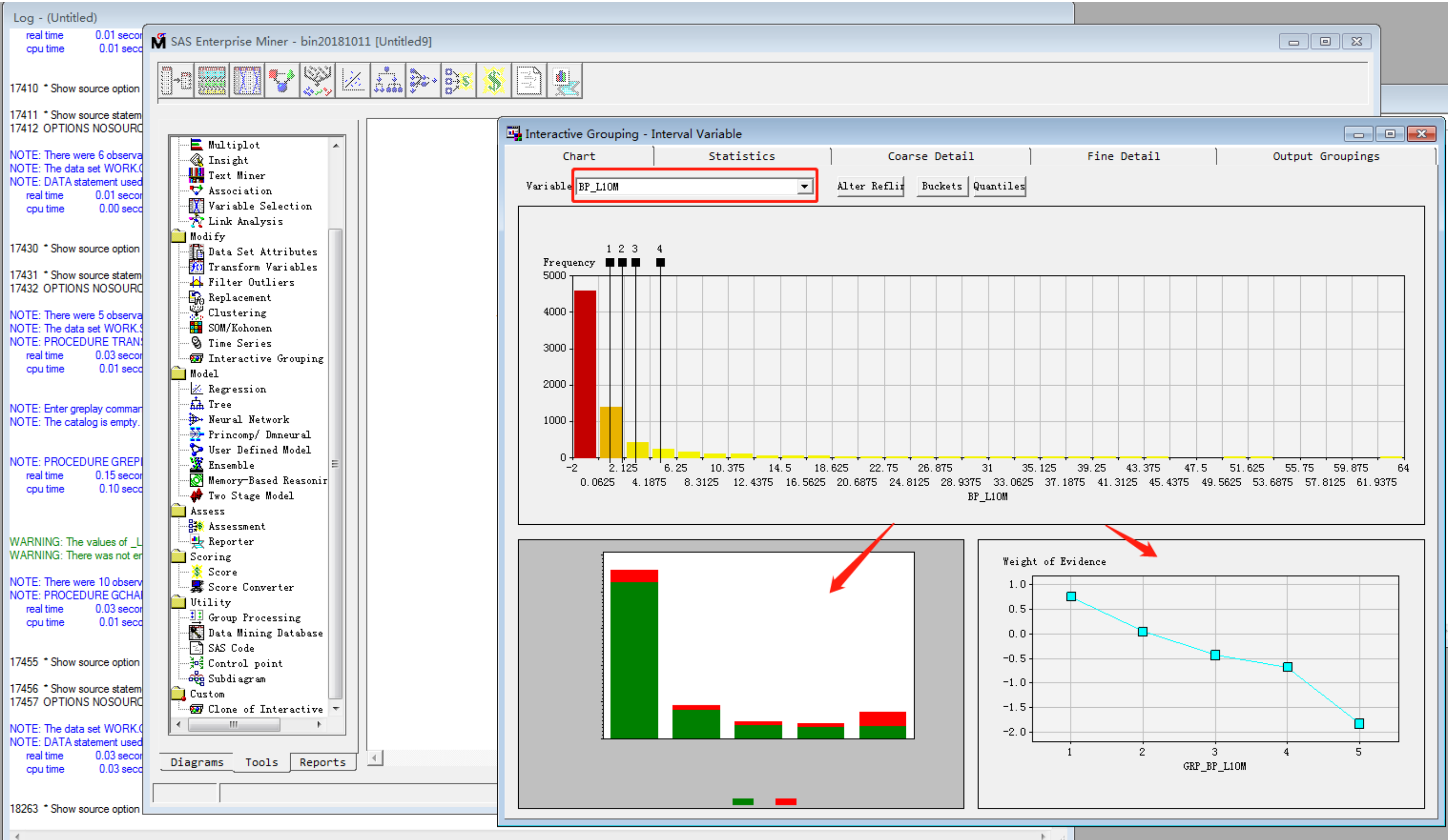
# 上午课程回顾

难点：

- 1.模型的目标Y变量的定义
- 2.用Y预测Y
- 3.模型的策略的使用方式

# 第八章：变量分箱（实操）

Sas :



Python :

'PTP_L5M':	PTP_L5M	chi_square	expected_pos_cnt	pos_cnt
0	0	171.939498	576.965265	262
1	13	121.003686	444.168190	676
2	15	159.955925	7.458867	42
3	18	87.667569	4.387569	24
4	52	119.632610	7.020110	36,
'PTP_L6M':	PTP_L6M	chi_square	expected_pos_cnt	pos_cnt
0	0	170.872306	542.449726	238
1	14	87.087200	473.857404	677
2	15	102.278830	5.118830	28
3	19	161.247564	8.336380	45
4	55	170.360517	10.237660	52,
'PTP_L7M':	PTP_L7M	chi_square	expected_pos_cnt	pos_cnt
0	0	154.739540	477.952468	206
1	19	69.553354	541.864717	736
2	20	78.286135	4.095064	22
3	26	131.437226	8.190128	41
4	72	93.007577	7.897623	35,
'PTP_L8M':	PTP_L8M	chi_square	expected_pos_cnt	pos_cnt
0	0	129.703735	421.937843	188
1	22	51.763330	602.413163	779
2	24	55.278476	3.363803	17
3	32	71.221464	6.873857	29
4	80	86.128564	5.411335	27,
'PTP_L9M':	PTP_L9M	chi_square	expected_pos_cnt	pos_cnt
0	0	110.746661	385.667276	179
1	25	40.845551	641.170018	803
2	32	50.276029	5.850091	23
3	40	71.712495	4.680073	23
4	89	33.332541	2.632541	12,
'PTP_SUM':	PTP_SUM	chi_square	expected_pos_cnt	pos_cnt
0	0	81.789910	330.383912	166
1	41	27.186659	701.864717	840
2	42	24.918761	0.731261	5
3	49	38.887965	3.071298	14
4	118	30.927978	3.948812	15}

# 第八章：变量分箱-WOE证据权重

WOE(weight of evidence, 证据权重)：  
一种有监督的编码方式,将预测类别的集中度的属性作为编码的数值

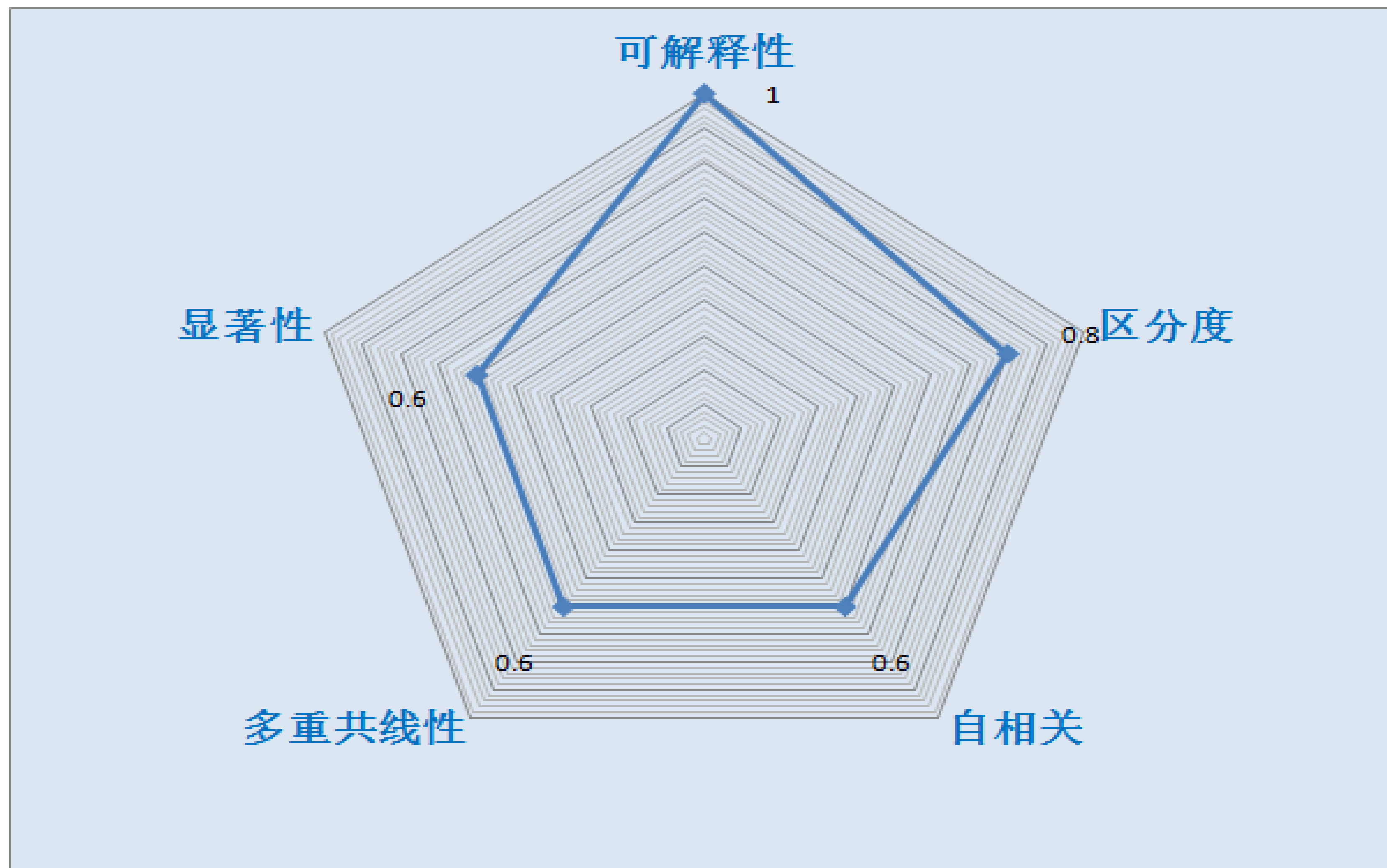
优势：将特征的值规范到相近的尺度上。  
(经验上讲,WOE的绝对值波动范围在0.1~3之间)。  
具有业务含义。

缺点：需要每箱中同时包含好、坏两个类别。

	Good	Bad	Good Percent	Bad Percent
Group 1	$G_1$	$B_1$	$G_1/G_{total}$	$B_1/B_{total}$
Group 2	$G_2$	$B_2$	$G_2/G_{total}$	$B_2/B_{total}$
...	...	...	...	...
Group N	$G_N$	$B_N$	$G_N/G_{total}$	$B_N/B_{total}$
Total	$G_{total} = \sum G_i$	$B_{total} = \sum B_i$		

$$WOE = \sum \log(\frac{G_i/G_{total}}{B_i/B_{total}})$$

## 第九章：模型特征的选择（实操）



# 第九章：模型特征的选择——自相关

范围	强度	相关性
R=1	完全	正相关
0.9<R<+1	较高	
0.5<R<0.9	适中	
0<R<0.5	较弱	
R=0	完全不相关	
-0.5<R<0	较弱	负相关
-0.9<R<-0.5	适中	
-1<R<-0.9	较高	
R=-1	完全	

如果自变量之间的相关系数超过0.5以上，可能会对模型产生影响，一般情况下会将相关性较强的变量剔除



# 第九章：模型特征的选择——共线性

## 方差膨胀系数与多重共线性

多重共线性是指自变量之间存在线性相关关系，即一个自变量可以是其他一个或几个自变量的线性组合。

检验方法主要有：**容忍度**(Tolerance:R)和**方差膨胀系数**(Variance inflation factor , VIF)。其中最常用的是VIF，计算公式为：

$$VIF = \frac{1}{1 - R_i^2}$$



R	R方	VIF
0.9	0.81	5.263158
0.7	0.49	1.960784
0.5	0.25	1.333333
0.4	0.16	1.190476
0.2	0.04	1.041667

方差膨胀系数是容忍度的倒数，VIF越大，表示自变量的容忍度越小，越有共线性问题。

## 第九章：模型特征的选择——共线性

③

Collinearity Diagnostics			-----Proportion of Variation-----			
Number	Eigenvalue	Condition Index	Intercept	x1	x2	x3
1	3.99037	1.00000	0.00015078	0.00024594	0.00017472	0.00027254
2	0.00501	28.21596	0.09477	0.18137	0.16968	0.23098
3	0.00329	34.80401	0.06637	0.48355	0.07204	0.55264
4	0.00132	54.90612	0.83872	0.33484	0.75811	0.21611

【结果解释】

① 为模型总体的假设检验结果。在这里，模型总体的假设检验如下。

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0, \text{ vs } H_1: \beta_i \neq 0, 1 \leq i \leq 3$$

从模型的总体检验结果得知，模型总体拟合数据很好 ( $F=14.07$ ,  $P<0.0001$ ,  $R^2=0.7251$ )。

② 为参数估计和参数假设检验结果。在这里，参数的假设检验如下。

$$H_0: \beta_i = 0, \text{ vs } H_1: \beta_i \neq 0, 1 \leq i \leq 3$$

从参数估计检验结果中得知，变量 x1 ( $P=0.0090$ ) 和 x2 ( $P=0.0334$ ) 的偏回归系数与零有显著性差异，变量 x3 的偏回归系数与零没有显著性差异。

③ 为共线性诊断。观察第 2 列条件指数，条件数 54.90612，大于 10，因此数据存在严重多重共线性。从最后两列可以看到，变量 x2 和 x3 的方差比例值很大，因此这两个变量有共线性。如果一个模型中同时包含这两个变量，得到的结果就会很不稳定，甚至产生误导。

下面用主成分回归分析法分析这个数据。

(2) 用主成分分析求主成分和主成分得分变量

【SAS 输出结果】

The PRINCOMP Procedure  
Observations  
Variables

③ Simple Statistics

Mean 49.  
Std 3.

② Correlation

x1  
x1 1.000  
x2 0.123  
x3 0.578

③ Eigenvalues

Eigenvalue  
1 1.7631  
2 0.6849  
3 0.3519

④ Eigenvectors

z1  
x1 0.  
x2 0.  
x3 0.

同行中，多个变量同时

## 第九章：模型特征的选择——IV值

IV	预测能力
<0.02	无预测能力
0.03~0.09	低
0.1~0.2	中
>0.2	强

可根据IV值来调整分箱结构并重新计算WOE和IV值，但并不完全是IV值越大越好。因为还需要考虑具体分成的组数。



# 第十章：模型拟合（实操）

模型算法：逻辑回归（最小二乘法） 向前、向后、逐步

逻辑回归的推导：

$$P = \exp(z) / [1 + \exp(z)]$$

$$\Pr\{y = 1\} = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r)}$$

$$p = \frac{1}{1 + e^{-z}}$$

$$\ln\left(\frac{p}{1-p}\right) = z \quad \Rightarrow \quad z = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

$$\ln(odds) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$\nu_1 = WOE(x_1) = \delta_{11} w_{11} + \delta_{12} w_{12} + \delta_{13} w_{13}$$

$$\nu_2 = WOE(x_2) = \delta_{21} w_{21} + \delta_{22} w_{22} + \delta_{23} w_{23} + \delta_{24} w_{24}$$

$$\nu_3 = WOE(x_3) = \delta_{31} w_{31} + \delta_{32} w_{32}$$

$$\begin{aligned} \ln(odds) = & \beta_0 + (\beta_1 w_{11}) \delta_{11} + (\beta_1 w_{12}) \delta_{12} + (\beta_1 w_{13}) \delta_{13} + \\ & (\beta_2 w_{21}) \delta_{21} + (\beta_2 w_{22}) \delta_{22} + (\beta_2 w_{23}) \delta_{23} + \\ & (\beta_2 w_{24}) \delta_{24} + (\beta_3 w_{31}) \delta_{31} + (\beta_3 w_{32}) \delta_{32} \end{aligned}$$

# 第十章：模型拟合（实操）

Enter：所有X一次性全部进入

Forward：X一个一个进，每次进入P-value最小的X，直到未进入的X都不significant

Backward：所有的X先一次性进入，然后一个一个剔除，每次剔除P-value最大的X，直到保留的X全都significant

Stepwise：X一个一个进，但是进入新的X以后，会重新审查所有已进入X的P-value，如果进入新的X导致原来的X的P-value从significant变成不significant，则把原来的X剔除。

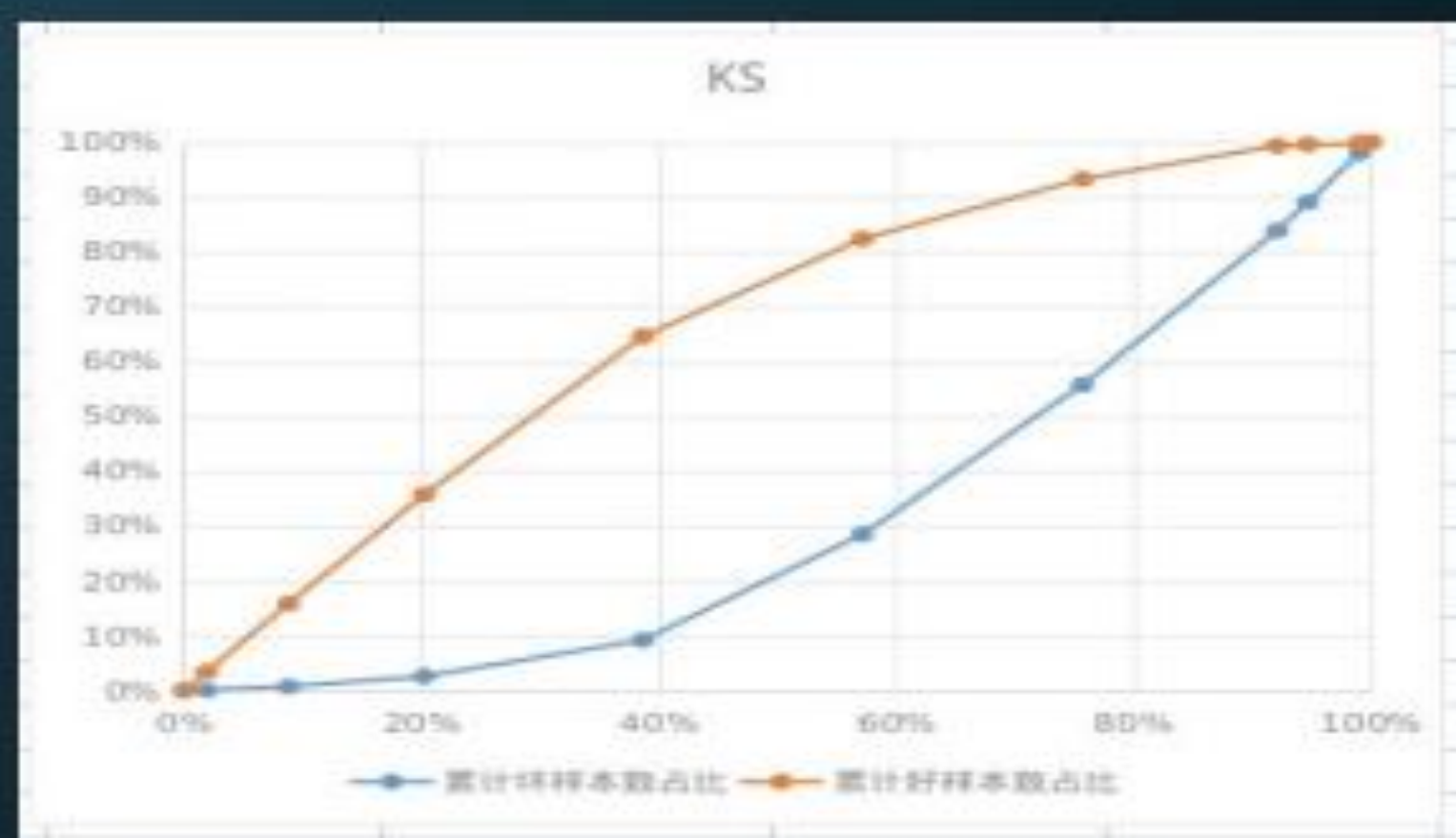
总地来说，forward只进不出，backward只出不进，stepwise又进又出。有文献提到一般来讲backward的效果往往是这三种方法之中最好的

## 逻辑回归模型

基本原理：最小化{预测精度损失 + 模型复杂度}

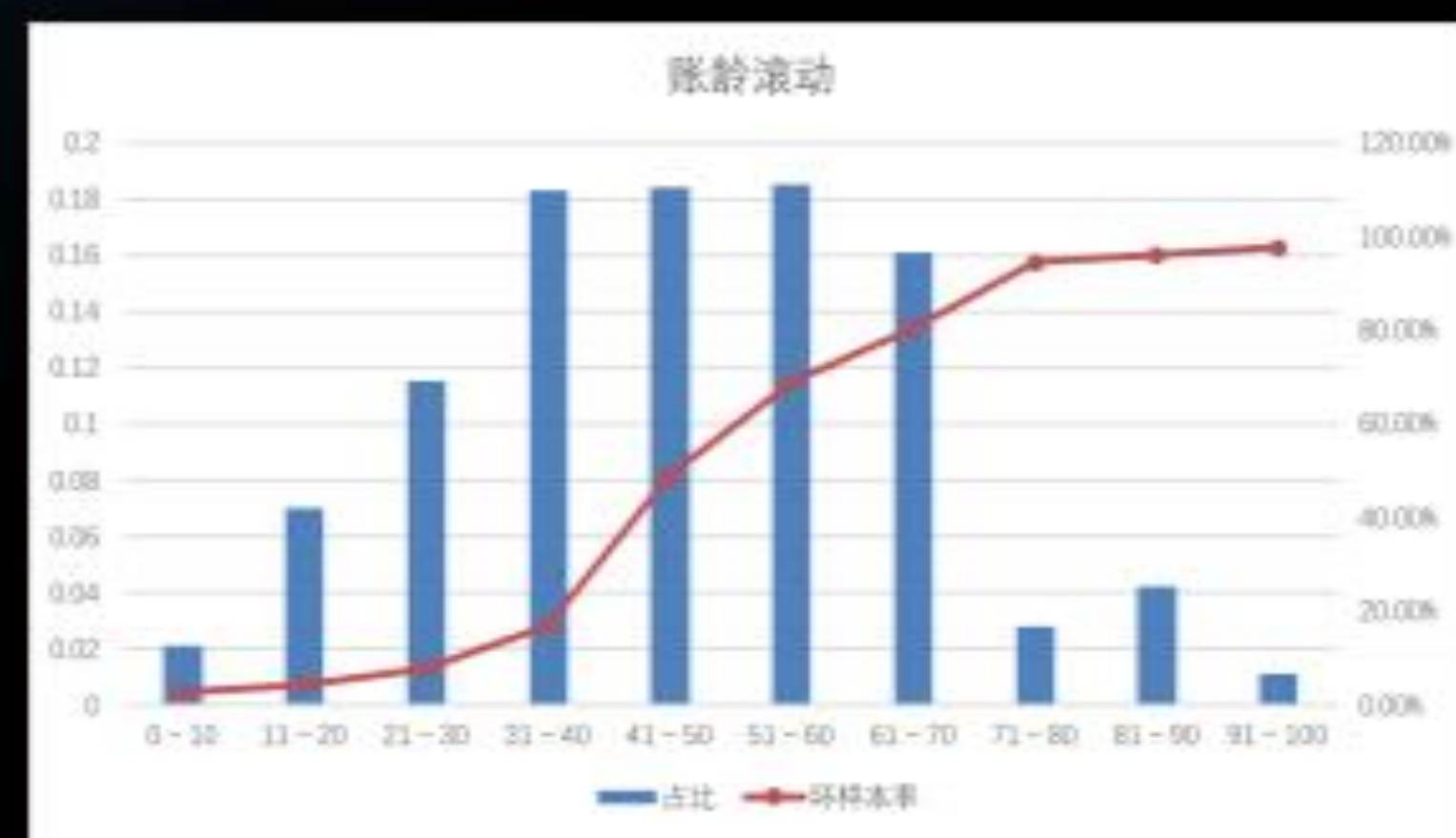
特点：属于广义线性回归模型的一种。

优点：可解释性高，稳定性高



## 基于逻辑回归的账龄滚动模型的性能

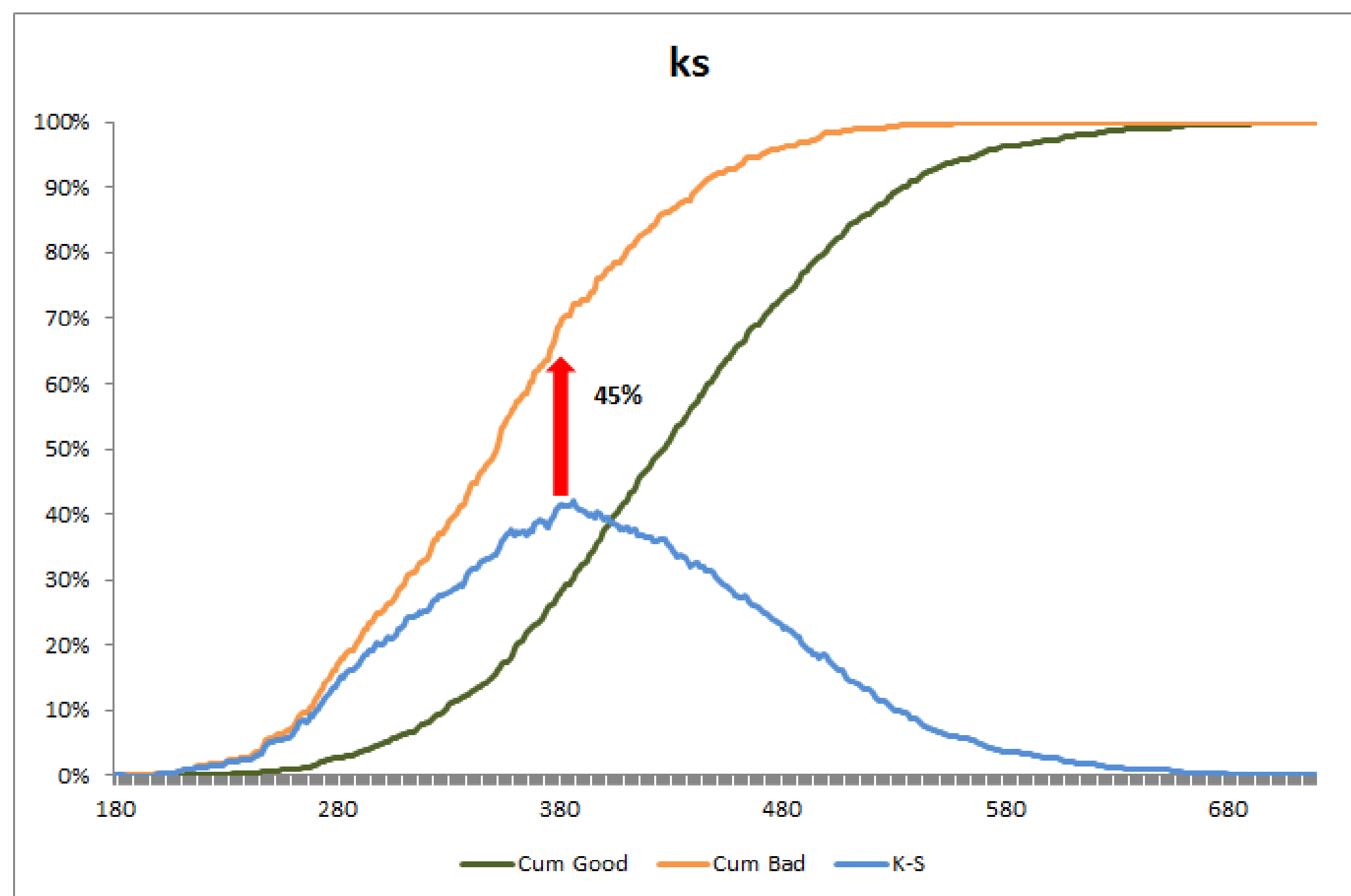
- KS：60% +
- AR：70% +



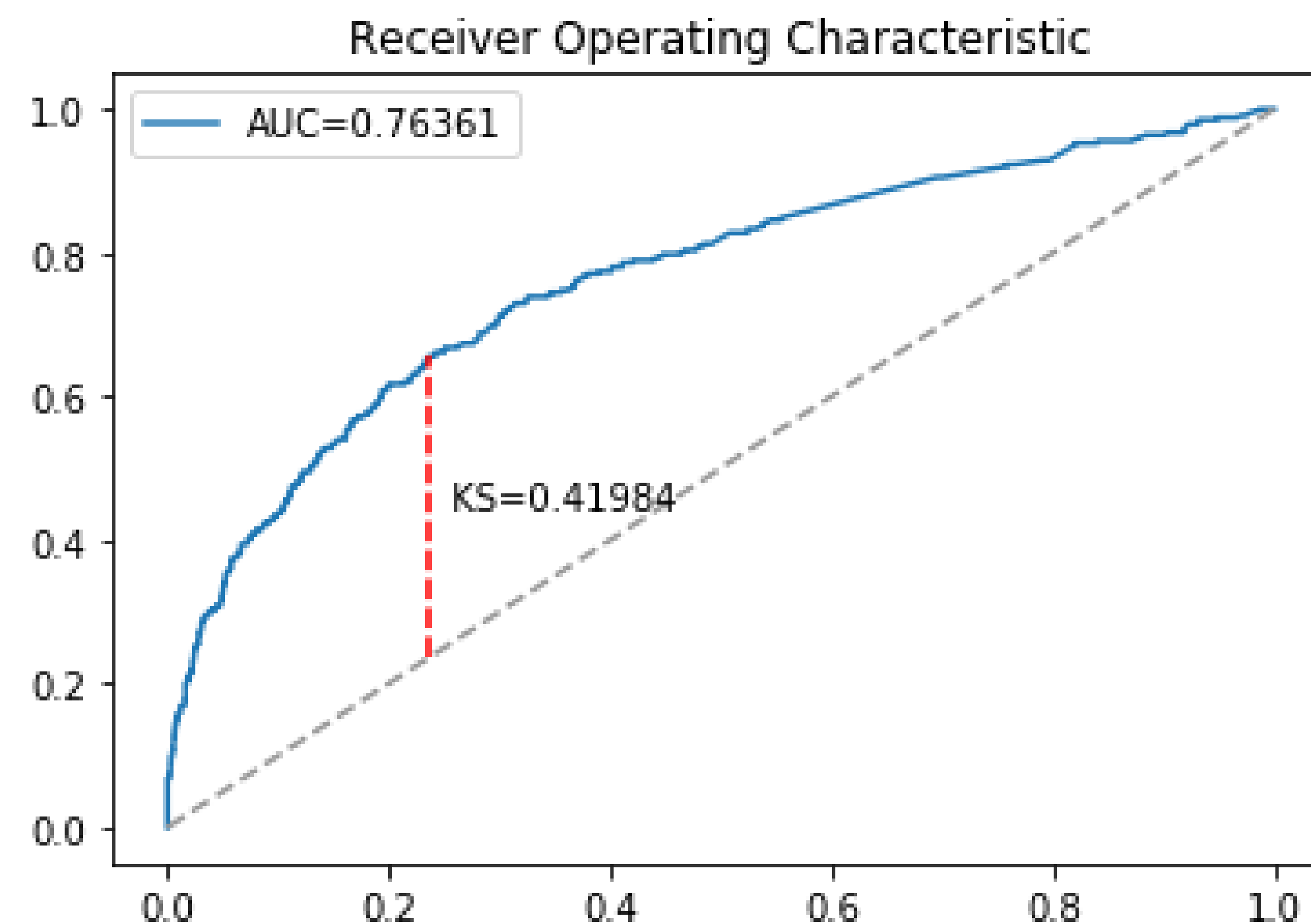
# 第十一章：模型验证

风控模型	IV	information value 信息值。一般取值区间(0,1)。该值用来表示某个变量的预测能力，越大越好。通常IV值0.3以上的，预测能力较高。 IV=SUM((B_P-G_P)*LN(B_P/G_P))
	K-S	klmogrov-smirnov，这是一个区分度指标。所谓区分度，是指模型对于好坏客户的辨识能力，区分力越强，模型准确度越高，误判的几率越低。K-S值越大越好，一般0.6以上用户解释能力很高。 KS=Max(RETAIN_BAD_P-RETAIN_GOOD_P)
	PSI	population stability index，稳定度指标，越低越稳定。用于比较当前客群与模型开发样本客群差异程度，评价模型的效果是否符合预期。 PSI = SUM( ( VALID_BAD_P-TRAIN_BAD_P ) * LN(VALID_BAD_P/TRAIN_BAD_P))
	Lift	模型提升度，表示使用模型比未使用的区分效果提升能力
	AUC	Area Under Curve,定义为ROC曲线下面积，通常大于0.5小于1。体现模型预测精准度指标之一
	GINI	同KS指标一样，都是体现模型区分能力的指标
	AR	Accuracy Rate,AR=2AUC-1，表征模型的区分能力，同Gini指标计算结果一致

# 第十一章：模型验证



M	N	O	P
MAX K-S	Score	Cum Good	Cum Bad
45.05%	397	16.59%	61.63%



IV/KS/PSI/Lift/AUC → 0.02~0.05/0.3+/0.1-/0.3+/0.7+



## 第十二章：模型的分数的转化——生成评分卡

1.评分卡定义：

已知坏客户的概率为： $p$ ，已知好客户的概率为： $1-P$ ，好坏比为 $ODDS=(1-P)/P$

评分卡设定的分值刻度可以通过分值表示为：比率对数的线性表达式，即

$$\text{Score}=A+B*\ln(\text{odds})$$

其中A跟B均为常数

2.评分卡转换

设定 $ODDS=X1$ ,分数为 $P0$

设定 $ODDS$ 每增加1倍，增加的分数 $PDO$

当 $ODDS=2X1$ ,分数为 $P0+PDO$ ,于是可以得到：

$$\begin{cases} P0=A+B*\ln(X1) \\ P0+pdo=A+B*\ln(2X1) \end{cases}$$

$$\begin{cases} B=PDO/\ln(2) \\ A=P0-B*\ln(x) \end{cases}$$

# 第十二章：模型的分数的转化——生成评分卡

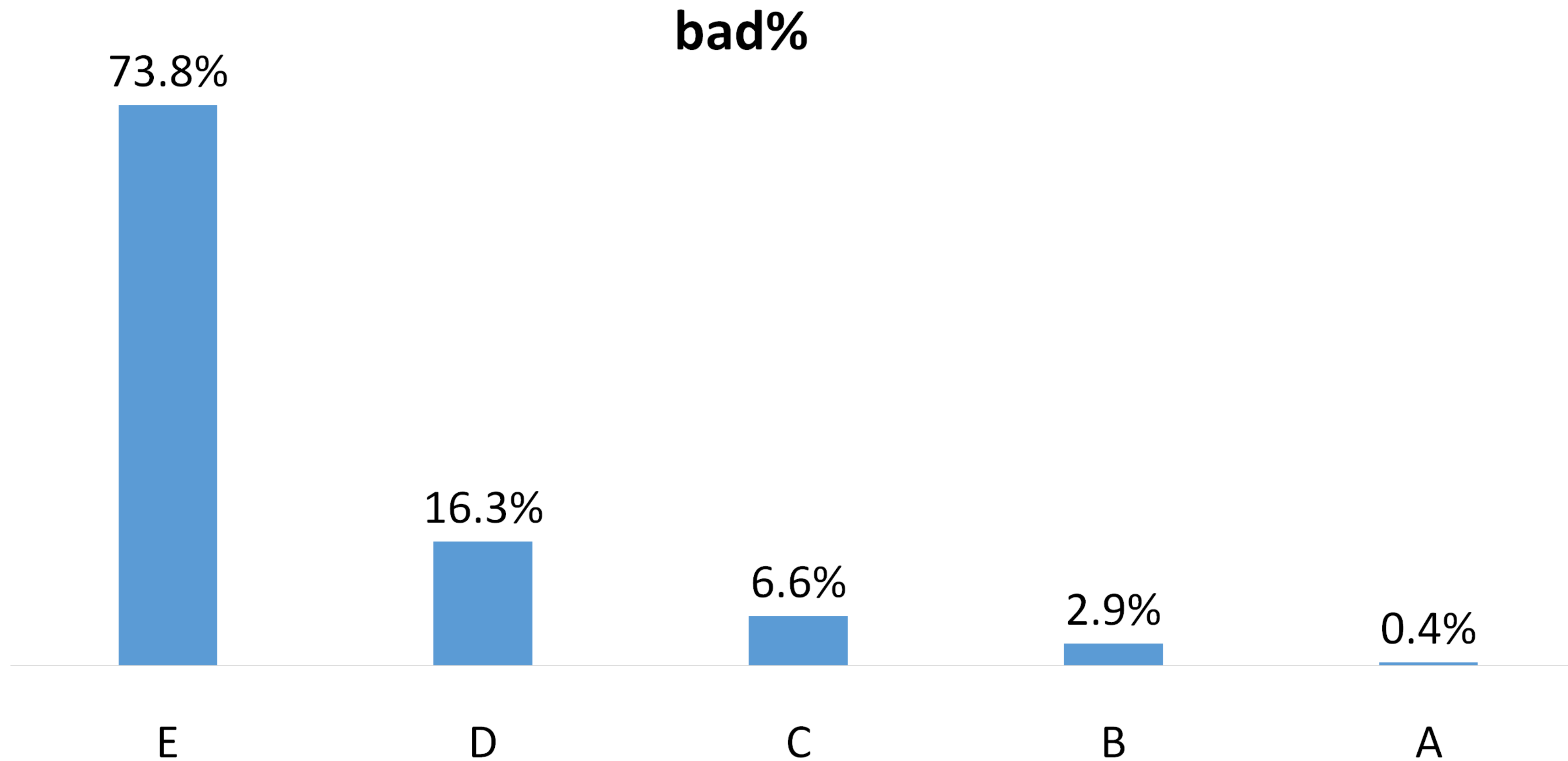
ODDS	SCORE	违约率
1 : 400	660	0.20%
1 : 240	640	0.41%
1 : 120	620	0.82%
1 : 60	600	1.63%
1 : 30	580	3.22%
1 : 15	560	6.25%
1 : 7.5	540	11.76%

# 第十二章：模型的分数的转化（代码中的实现与实操）

socre bin	Low	High	Score Interval	Range	Cum BadRate	Marginal Bad Rate	Cum Good	Cum Bad	Good	Bad	Good%	Bad%	Bad2%	bad	good	G/B ODDS	WoE	Informati on Value	Cumulati ve Good %	Cumulative Bad %	GINI Value
E	0	390	0-390	390	40.00%	40.00%	933	622	933	622	15.4%	59.8%	40.0%	40.0%	60.0%	0.26	(1.36)	0.60	15.4%	59.8%	459.6
D	391	455	391-455	64	26.52%	12.82%	2266	818	1333	196	22.0%	18.8%	12.8%	12.8%	87.2%	1.17	0.15	0.00	37.3%	78.7%	496.5
C	456	482	456-482	26	21.21%	5.68%	3262	878	996	60	16.4%	5.8%	5.7%	5.7%	94.3%	2.84	1.05	0.11	53.7%	84.4%	262.7
B	483	497	483-497	14	16.65%	5.33%	4842	967	1580	89	26.0%	8.6%	5.3%	5.3%	94.7%	3.04	1.11	0.19	79.8%	93.0%	571.2
A	498	550	498-550	52	14.63%	5.61%	6071	1040	1229	73	20.2%	7.0%	5.6%	5.6%	94.4%	2.88	1.06	0.14	100.0%	100.0%	630.9
	Total					14.63%			6071	1040	100.0%	100.0%						1.05			51.58%
	Total GINI				51.58%																
	Total KS				45.05%																

## 第十二章：模型的分数的转化（实操）

模型训练集的分组情况



风险等级A为最低，E为最高



# 第十二章：模型的分数的转化

序号	客户	贷款号	扣款卡银行	扣款日	剩余本金	进件评分	期款（元）	CALL_LEVEL	针对**行是否能扣款	自雇/授薪	发出日期
1	向**	*****	深圳市平安银行	23	4163.18	3	4230	A			2019/6/21
2	向**	*****	中国农业银行	23	4416.21	7	4487	C			2019/6/21
3	周**	*****	中国民生银行	23	1738.09	0	2246	E			2019/6/21
4	黄**	*****	中国工商银行总行清算中心	23	3518.01	0	3568	E			2019/6/21
5	潘**	*****	深圳市平安银行	23	2814.93	0	2855	B			2019/6/21
6	陈**	*****	深圳市平安银行	23	2814.93	0	2855	A			2019/6/21
7	杨**	*****	中国工商银行	23	13104.31	0	4522	D			2019/6/21
8	吴**	*****	深圳市平安银行	23	31220.43	5	10704	E			2019/6/21
9	颜**	*****	中国工商银行总行清算中心	23	11174.03	7	3868	A			2019/6/21
10	张**	*****	中国工商银行	23	8326.96	0	2855	A			2019/6/21
11	林**	*****	中国工商银行总行清算中心	23	26395.53	5	6944	C			2019/6/21
12	于**	*****	中国建设银行	23	7070.27	5	2339	D			2019/6/21
13	陈**	*****	深圳市农村商业银行	23	11314.41	4	3743	A			2019/6/21
14	黄**	*****	深圳市平安银行	23	42323.3	4	11004	D			2019/6/21
15	唐**	*****	招商银行	23	13778.98	0	3568	C			2019/6/21
16	徐**	*****	深圳市平安银行	23	29519.67	5	7736	D			2019/6/21
17	周**	*****	中国银行总行	23	24614.93	5	4405	E			2019/6/21
18	聂**	*****	中国工商银行	23	20880.41	2	4678	E			2019/6/21
19	梁**	*****	中国工商银行	23	14586.51	5	2568	D			2019/6/21
20	刘**	*****	深圳市平安银行	23	18343.09	0	3211	A			2019/6/21
21	张**	*****	中国建设银行总行	23	25279.87	6	4524	E			2019/6/21
22	董**	*****	中国农业银行资金清算中心	23	13039.42	6	2321	A			2019/6/21
23	罗**	*****	中国农业银行资金清算中心	23	15830.68	2	2848	A			2019/6/21
24	黄**	*****	中国工商银行总行清算中心	23	17295.18	7	3095	E			2019/6/21
25	李**	*****	中国工商银行总行清算中心	23	13081.85	7	2341	D			2019/6/21
26	李**	*****	深圳市平安银行	23	46342.12	8	8293	D			2019/6/21
27	官**	*****	中国工商银行总行清算中心	23	9672.32	6	2032	E			2019/6/21
28	胡**	*****	中国光大银行	23	15072.17	5	2321	C			2019/6/21
29	胡**	*****	中国工商银行	23	24170.77	4	5078	E			2019/6/21
30	刘**	*****	中国工商银行总行清算中心	23	38674.47	6	5992	D			2019/6/21
31	杜**	*****	深圳市平安银行	23	17419.34	6	2699	D			2019/6/21
32	李**	*****	深圳市平安银行	23	19339.65	1	3743	D			2019/6/21
33	曾**	*****	中信实业银行	23	21930.67	5	4063	A			2019/6/21
34	郑**	*****	平安银行	23	22757.5	5	3095	A			2019/6/21
35	谭**	*****	中国农业银行资金清算中心	23	39991.98	4	5476	E			2019/6/21
36	章**	*****	平安银行	23	17707.07	7	2441	D			2019/6/21
37	赵**	*****	中国农业银行资金清算中心	23	14755.07	6	2034	C			2019/6/21
38	胡**	*****	深圳市平安银行	23	52160.22	5	7142	E			2019/6/21
39	谢**	*****	中国工商银行总行清算中心	23	50421.83	6	6904	E			2019/6/21
40	黄**	*****	平安银行	23	53034.67	7	7262	C			2019/6/21
41	王**	*****	兴业银行总行	23	27535.69	3	4210	C			2019/6/21
42	熊**	*****	招商银行总行	23	19023.96	6	2321	C			2019/6/21
43	廖**	*****	中国农业银行	23	12243.31	8	1872	B			2019/6/21
44	刘**	*****	平安银行	23	64557.34	5	7936	E			2019/6/21
45	谢**	*****	平安银行	23	48417.98	6	5952	C			2019/6/21
46	黄**	*****	中国银行总行	23	20661.07	8	2540	B			2019/6/21
47	封**	*****	平安银行	23	65525.47	4	8055	B			2019/6/21
48	刘**	*****	平安银行	23	33571.4	4	4127	B			2019/6/21
49	周**	*****	中国民生银行	23	13767.89	7	2105	D			2019/6/21
50	赖**	*****	中国工商银行总行清算中心	23	79728.54	7	9801	B			2019/6/21
51	高**	*****	中国农业银行	23	18359.07	5	3047	C			2019/6/21
52	邱**	*****	中国工商银行总行清算中心	23	22693.32	7	2515	B			2019/6/21
53	毛**	*****	深圳市平安银行	23	33671.09	5	3668	E			2019/6/21
54	钟**	*****	中国银行	23	32046.23	2	4444	B			2019/6/21
55	候**	*****	平安银行	23	16835.55	6	1834	E			2019/6/21
56	谢**	*****	中国民生银行总行	23	18410.49	7	2289	C			2019/6/21
57	邱**	*****	深圳市平安银行	23	57795.98	4	5913	E			2019/6/21
58	祝**	*****	平安银行	23	86042.58	6	8297	C			2019/6/21
59	聂**	*****	北京银行股份有限公司	23	128424.42	5	12384	E			2019/6/21
60	彭**	*****	交通银行	23	98704.87	4	9283	E			2019/6/21
61	郑**	*****	中国农业银行	23	1994.53	6	2427	D			2019/6/21
62	刘**	*****	中国工商银行	23	3734.8	4	3820	B			2019/6/21
63	刘**	*****	平安银行	23	42850.53	6	4678	D			2019/6/21
64	张**	*****	中国工商银行	23	2493.33	6	3034	A			2019/6/21
65	逢**	*****	中国工商银行	23	36427.82	4	3977	D			2019/6/21
66	杜**	*****	中国工商银行总行清算中心	23	26492.22	5	2321	E			2019/6/21
67	杨**	*****	中国工商银行	23	2742.35	5	3337	A			2019/6/21
68	蔡**	*****	深圳市平安银行	23	123282.84	6	11104	C			2019/6/21

# 十三章：催收手段介绍

- 1.自动代扣
- 2.短信提醒
- 3.IVR、智能语音
- 4.预测式跟预览式外呼(主动与被动)
- 5.上门催收
- 6.法务催收、律师函
- 7.委外催收
- 8.法律诉讼、网络仲裁

# 十三章：催收手段介绍

## 法律诉讼

### 01 审判阶段

---

立案——法院审理——法院判决

### 02 诉讼保全

---

申请（与起诉状同时递交）——法院受理——法院保全

### 03 执行阶段

---

立案——法院查财产——法院处置财产——执行完毕结案



# 十三章：催收手段介绍

对比项目		诉讼	线下仲裁	线上仲裁
流程时效		1.简易程序：90天； 2.普通程序：180天—360天	60天	30天
执行效力	裁定方式	可上诉	一裁终局	
	审理人员	法院指定审理人，当事人无权选择	当事人同时指定仲裁员	
	递件立案	律师代理	专人	
	执行机构	我行所在地区级法院	借款人常住地或财产所在地的中级法院	
	判决效力	具有法律效力	具有法律效力	
	核销执行	执行程序一致	执行程序一致	
费用成本	各项费用	1.律师费（风险代理，按标的5%） 2.受理费 3.执行费 4.保全费	1.仲裁受理费+处理费 2.执行费 3.保全费	1.系统接入费（包年） 2.仲裁受理费+处理费 3.执行费 4.保全费

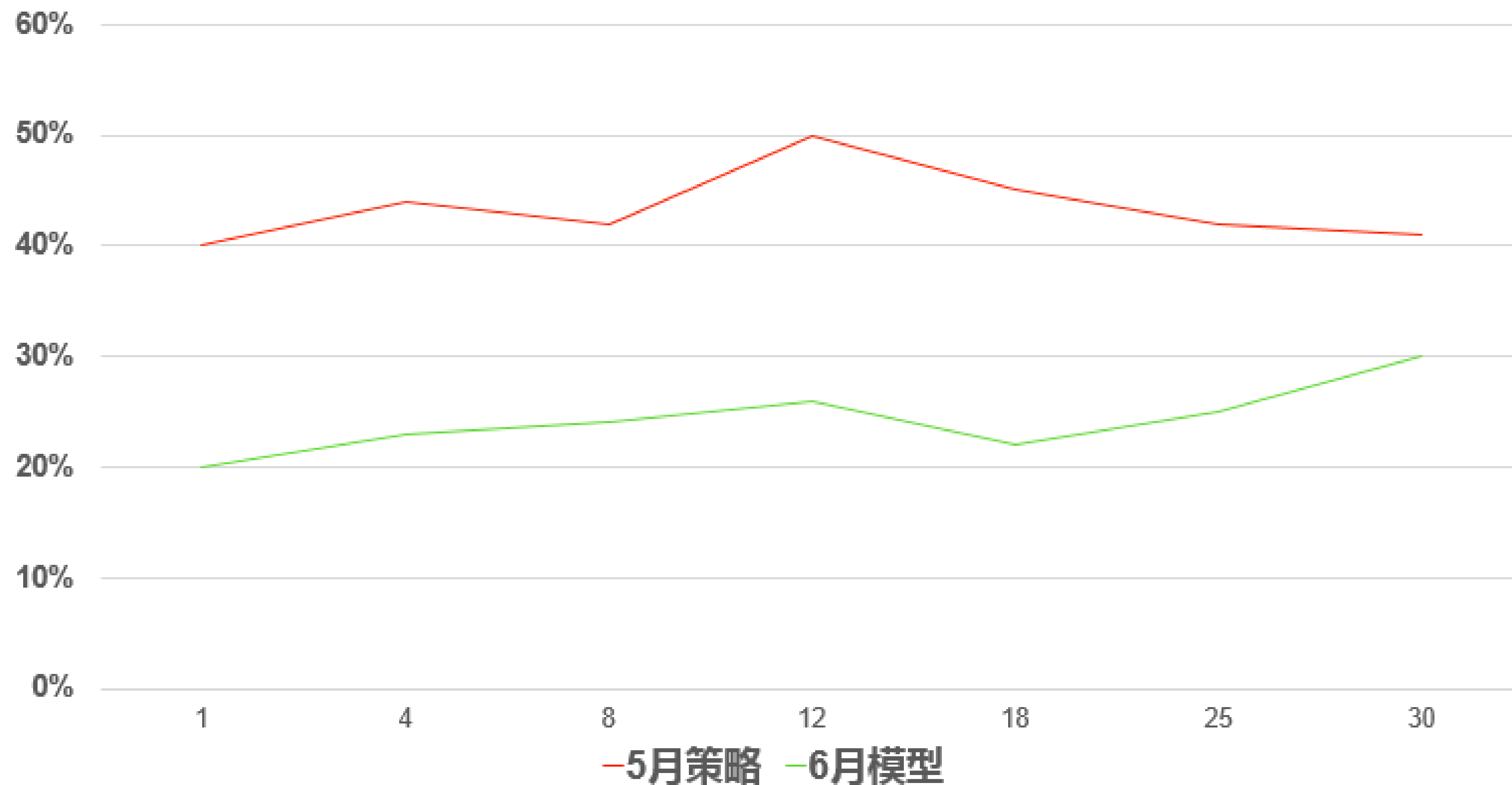
# 十四章：催收模型的使用效果对比

Call List等级定义：

风险程度	排序	维度	说明
A	HIGH	ABS项目	最近一期逾期天数≥15天，或
			最近一期BP次数≥8次，或
			最近一期PTP次数≥8次，或
	MIDDLE		最近一期Contact次数≥8次
			未命中高、低等级的其他情况
	LOW		最近一期逾期天数≤5天，且
			最近一期BP次数≤3次，且
			最近一期PTP次数≤3次，且
			最近一期Contact次数≤3次
B	HIGH	其他	最近一期逾期天数≥15天，或
			最近一期BP次数≥8次，或
			最近一期PTP次数≥8次，或
	MIDDLE		最近一期Contact次数≥8次
			未命中高、低等级的其他情况
	LOW		最近一期逾期天数≤5天，且
			最近一期BP次数≤3次，且
			最近一期PTP次数≤3次，且
			最近一期Contact次数≤3次

# 十四章：催收模型的使用效果对比

策略与模型入催率比较



# 十五章：催收模型的策略调整

## 策略设计-举例

评分	模型评分等级	分组	比例	策略
I	高风险	测试组1	20%	DPD1-3IVR催收，DPD4手工外呼
H		测试组2	30%	DPD1-15自动外呼，DPD16委外
G		对照组	50%	DPD1-30手工外呼
F	中风险	测试组1	20%	DPD1-3IVR催收，DPD4手工外呼
E		测试组2	30%	DOD1-15自动外呼，DPD16手工外呼
D		对照组	50%	DPD1-7自动外呼，DPD8手工外呼
C	低风险	测试组1	20%	DPD1-4IVR发送短信，DPD5自动外呼
B		测试组2	30%	DPD1-4IVR催收，DPD自动外呼
A		对照组	50%	DPD1-30自动外呼



# 十五章：催收模型的策略调整



一般来说对照组都会保留原先的策略，而测试组是用来跟对照组比对策略优劣程度的，属于冠军挑战者的一种方法。所以测试的都只用占比较少的客户做测试比对，万一测试的策略不好也不用担心会影响到实际的业务数据。可以理解为测试都是在小范围进行调试

2019年7月3日 下午11:11

那为什么一组20% 一组30%，而不是25%和25%呢

2019年7月3日 下午11:18

其实还是那个意思，在每个风险等级里，你会看到策略的强到弱的趋势分别是：对照组》测试组2》测试组1。所以你可以看到从占比上看，我都是希望从强到弱进行优化，所以策略最强的占比都是最高，中间的次之，最弱的占比最少

评分	评分等级	分组	比例	策略
I	高风险	测试组1	20%	DEP1-S178策略，DEP4手工外呼
II		测试组2	30%	DEP1-15自动外呼，DEP16委外
III		对照组	50%	DEP1-30手工外呼
IV	中风险	测试组1	20%	DEP1-S178策略，DEP4手工外呼
V		测试组2	30%	DEP1-15自动外呼，DEP16手工强催



# 十六章：催收模型的优化的方向

MODLE_I \ MODLE_II									
	风险等级A	风险等级B	风险等级C	风险等级D	风险等级E	风险等级F	风险等级G	风险等级H	Total
风险等级A	98.00%	97.80%	94.50%	93.50%	84.50%	74.60%	73.50%	72.30%	86%
风险等级B	97.50%	96.80%	91.50%	92.50%	76.00%	70.20%	64.30%	63.20%	82%
风险等级C	97.50%	94.00%	89.70%	85.20%	74.00%	69.80%	64.00%	62.00%	80%
风险等级D	96.00%	95.00%	88.00%	84.50%	73.00%	67.40%	64.20%	55.00%	78%
风险等级E	95.30%	94.50%	89.60%	83.50%	72.00%	66.30%	63.00%	54.00%	77%
风险等级F	95.50%	92.50%	89.80%	82.80%	72.30%	65.30%	61.00%	53.20%	77%
风险等级G	96.50%	91.90%	88.90%	80.80%	73.50%	62.30%	51.20%	53.60%	75%
风险等级H	92.50%	90.50%	87.50%	78.80%	74.80%	62.9%	52.00%	49.00%	75%
Total	96%	94%	90%	85%	75%	68%	62%	58%	79%

风险等级	催回率
风险等级A	95%+
风险等级B	90~95%
风险等级C	85~90%
风险等级D	75~85%
风险等级E	70~75%
风险等级F	65~70%
风险等级G	55~65%
风险等级H	55%-

# 十七章：催收模块的其他内容

- 1.员工效能、绩效指标
- 2.贷后回收率指标
- 3.贷后迁徙率指标
- 4.贷后催收绩效指标



# 十七章：催收模块的其他内容

过程指标：员工效能指标

常用指标	定义
合同接通率	当天接通的合同数/当天外呼的合同数
接通率	接通数/外呼次数
RPC比率	联系RPC合同数/接通合同数
PTP比率	承诺还款合同数/联系到RPC的合同数
KPTP比率	实际还款合同数/承诺还款合同数
跟P比率	跟p的合同量/有效Ptp的合同量
工时利用率	接通电话的时长/工作时间（8小时）

# 十七章：催收模块的其他内容

结果指标：员工绩效指标

常用指标	定义
回收合同量	考核周期内的回收合同数量
合同回收率	回收合同量/分配合同量
本金回收率	本金回收金额/分配合同的本金逾期总金额
回收金额	考核周期内的回收总金额
金额回收率	回收金额/分配合同的逾期总金额
本金回收率	本金回收金额/分配合同的本金逾期总金额

# 十七章：催收模块的其他内容

## 关键指标：贷后回收率指标

常用指标	定义
M1合同回收率	分母：月初1号到月末逾期阶段M1的合同量，分子：月初1号到月末回收阶段为M1的合同量
M1金额回收率	分母：月初1号到月末逾期阶段M1的总逾期金额，分子：月初1号到月末回收阶段为M1的回收金额
M2合同回收率	分母：月初1号到月末逾期阶段M2的合同量，分子：月初1号到月末回收阶段为M2的合同量
M2金额回收率	分母：月初1号到月末逾期阶段M2的总逾期金额，分子：月初1号到月末回收阶段为M2的回收金额
Mn合同回收率	分母：月初1号到月末逾期阶段Mn的合同量，分子：月初1号到月末回收阶段为Mn的合同量
Mn金额回收率	分母：月初1号到月末逾期阶段Mn的总逾期金额，分子：月初1号到月末回收阶段为Mn的回收金额



# 十七章：催收模块的其他内容

## 回款率绩效激励

### 3. 近半年数据参考：

201805-201810	M1	M2	M3	外访（2人）	外包
回收金额（个人）	785,310.29	42,731.35	24,061.59	-	-
回收金额（团队）	10,209,033.71	213,656.77	120,307.93	90,817.80	408,656.72

### 4. 激励标准：

#### 第一部分：特别激励

参与人	第一档	第二档	第三档
M1 组员	回收金额 $\geq 80$ 万*0.3%	回收金额 $\geq 90$ 万*0.5%	回收金额 $\geq 110$ 万*1.0%
M2 组员	回收金额 $\geq 4$ 万*3.0%	回收金额 $\geq 4.5$ 万*5.0%	回收金额 $\geq 6$ 万*10.0%
M3 组员	回收金额 $\geq 2$ 万*5.0%	回收金额 $\geq 2.5$ 万*7.0%	回收金额 $\geq 4$ 万*12.0%
外访组员	回收金额 $\geq 10$ 万*2.0%	回收金额 $\geq 12$ 万*4.0%	回收金额 $\geq 15$ 万*7.0%
外包组员	回收金额 $\geq 35$ 万*0.5%	回收金额 $\geq 40$ 万*1.0%	回收金额 $\geq 50$ 万*2.0%
组长	M1 组回收金额 $\geq 900$ 万*0.05%	M1 组回收金额 $\geq 1050$ 万*0.07%	M1 组回收金额 $\geq 1300$ 万*0.1%
副经理/ 经理	M1 组回收金额 $\geq 900$ 万*0.05%	M1 组回收金额 $\geq 1050$ 万*0.07%	M1 组回收金额 $\geq 1300$ 万*0.1%
	M2 组回收金额 $\geq 16$ 万*1.0%	M2 组回收金额 $\geq 18$ 万*1.2%	M2 组回收金额 $\geq 23$ 万*1.5%
	M3 组回收金额 $\geq 8$ 万*2.0%	M3 组回收金额 $\geq 9$ 万*2.4%	M3 组回收金额 $\geq 11$ 万*3.0%
	外访组回收金额 $\geq 10$ 万*1.0%	外访组回收金额 $\geq 12$ 万*1.5%	外访组回收金额 $\geq 15$ 万*2.0%
	外包组回收金额 $\geq 35$ 万*0.5%	外包组回收金额 $\geq 40$ 万*0.7%	外包组回收金额 $\geq 50$ 万*1.5%
预计	1000 万	1150 万	1400 万

#### 第二部分：竞争激励（前提条件为该员工可以拿到第一部分特别激励）

名额/组别奖励金额	M1 组奖励金额	M2/M3 组奖励金额
第一名	5000 元	5000 元
第二名	4000 元	4000 元
第三名	3000 元	/
第四名	2000 元	/
第五名	1000 元	/

#### （二）委外催收

1. 工作内容：公司委托外部催收机构对公司逾期贷款进行催收，并根据催收回款金额支付一定比例的佣金。

2. 考核标准：为激励外部催收机构加大对公司案件的催收力度，佣金支付标准在现有基础上有增加，但新的支付标准暂定执行3个月的时间，并以11月30日为第一个检视点，检视提高佣金支付标准后的效果，已确定后续的激励方案。

3. 佣金支付标准：

# 十七章：催收模块的其他内容

入催率、出催率

常用指标	定义
DPD1入催率	分母：当月应还款合同数，分子：当月逾期DPD1的合同数
DPD31入催率	分母：当月累计逾期CPD30天的合同数，分子：当月逾期DPD31的合同数
DPD61入催率	分母：当月累计逾期DPD60天的合同数，分子：当月逾期DPD61的合同数
出催率	分母：当月逾期合同数，分子：当月DPD0合同数



# 十七章：催收模块的其他内容

## 迁移率/催收绩效指标

常用指标	定义
M0 → M1	分母：上月末M0待收余额，分子：本月末M1待收余额
M1 → M2	分母：上月末M1待收余额，分子：本月末M2待收余额
M2 → M3	分母：上月末M2待收余额，分子：本月末M3待收余额
M3 → M4	分母：上月末M3待收余额，分子：本月末M4待收余额
M4 → M5	分母：上月末M4待收余额，分子：本月末M5待收余额
M5 → M6	分母：上月末M5待收余额，分子：本月末M6待收余额
Mn-1 → Mn	分母：上月末Mn-1待收余额，分子：本月末Mn待收余额

催收绩效简单来说，催收绩效 = 1 - 迁徙率

# 十七章：催收模块的其他内容

逾期阶段	19-1	19-2
M0	623,791,619	750,057,514
M1	16,561,476	11,159,895
M2	4,719,028	3,439,872
M3	1,925,360	2,284,796
M4	887,330	1,341,182
M5	731,527	807,867
M6	540,643	696,529
WO ( 新增坏账 )	593,092	524,134
在贷余额 ( M0-M6 )	649,156,982	769,787,655
不良金额	2,159,500	2,845,578
不良率	0.33%	0.37%
M3+	4,084,860	5,130,375
M3+%		0.67%
coincident		

迁徙率	
M0-1	1.8%
M1-2	20.8%
M2-3	48.4%
M3-4	69.7%
M4-5	91.0%
M5-6	95.2%
M6-7	96.9%

M0-7	0.1%
M1-7	5.9%
M2-7	28.3%
M3-7	58.5%



# 十七章：催收模块的其他内容

## 迁移率/催收绩效案例：

➤ 目标及对应奖励标准 9月-12月：

组别	Flow Rate	第一档	Flow Rate	第二档	Flow Rate	第三档	Flow Rate	第四档	Flow Rate	第五档
Current-M1	30.00%	6,000	32.00%	4,000	33.00%	2,500	34.00%	1,500	35.00%	1,000
M1-M2	7.00%	6,000	7.40%	4,000	7.60%	2,500	7.80%	1,500	8.00%	1,000
M2-M3	65.00%	6,000	69.00%	4,000	72.00%	2,500	75.00%	1,500	78.00%	1,000
M3-CO	65.00%	6,000	67.00%	4,000	69.00%	2,500	71.00%	1,500	73.00%	1,000
费用	24,000 元		16,000 元		10,000 元		6,000 元		4,000 元	

➤ 奖励办法：

- 1、以上均为催收团队考核目标，奖金为团队经费形式；
- 2、催收团队所有组别完成至第一档时最高奖励合计 24,000 元/月，各组分开考核，例如 M1→M2 最终 Flow Rate 为 7.8%，则对应奖励 1,500 元，以此类推，按最终完成值对应档位奖励，第五档为考核最低标准；
- 3、以下为近三个月数据仅供参考：

Flow Rate	6月	7月	8月	平均
Current-M1	35.52%	35.94%	32.14%	34.53%
M1-M2	6.99%	8.01%	9.01%	8.00%
M2-M3	85.46%	81.36%	67.23%	78.02%
M3-CO	87.52%	69.25%	62.85%	73.21%

催收绩效简单来说，催收绩效 = 1 - 迁徙率

# 十八章：催收模型后续的内容

- 1.模型的线上跑分跟线下跑分
2. 模型的上线和部署
- 3.线下部署+SMG3部署
- 3.机器学习类的使用(XGB /集成算法)



# THANK YOU

