评分模型开发方法

费埃哲信息技术（北京）有限公司

Confidential. This presentation is provided for the recipient only and cannot be reproduced or shared without Fair Isaac Corporation's express consent.

© 2011 Fair Isaac Corporation.

会议内容

一. 评分模型的基本概念

- 基本概念
- 生命周期中评分的应用

二. 申请评分模型开发方法

三. 数据驱动的统计模型开发

四. 专家经验模型

五. 混合式模型

© 2013 Fair Isaac Corporation. Confidential.

FICO

评分模型定义

- **评分模型（评分卡/模型/评分）**
  - » 风险评估工具，该评分被看作是信用风险的衡量标准(即还款概率)
- **特征项（变量）**
  - » 有关申请的疑问或信用报告的相关信息
    - 举例：年龄，当前地址居住时间，最近一次违约距今的时间等
- **属性（值）**
  - » **有关申请疑问和信用报告的解答**
    - 举例，年龄为32岁，当前地址居住时间为3年，最近一次违约距今的时间为45个月

$$\text{Score} = \sum_{i=1}^{N} w_i X_i$$

© 2013 Fair Isaac Corporation. Confidential.

FICO

评分卡实例

特征项

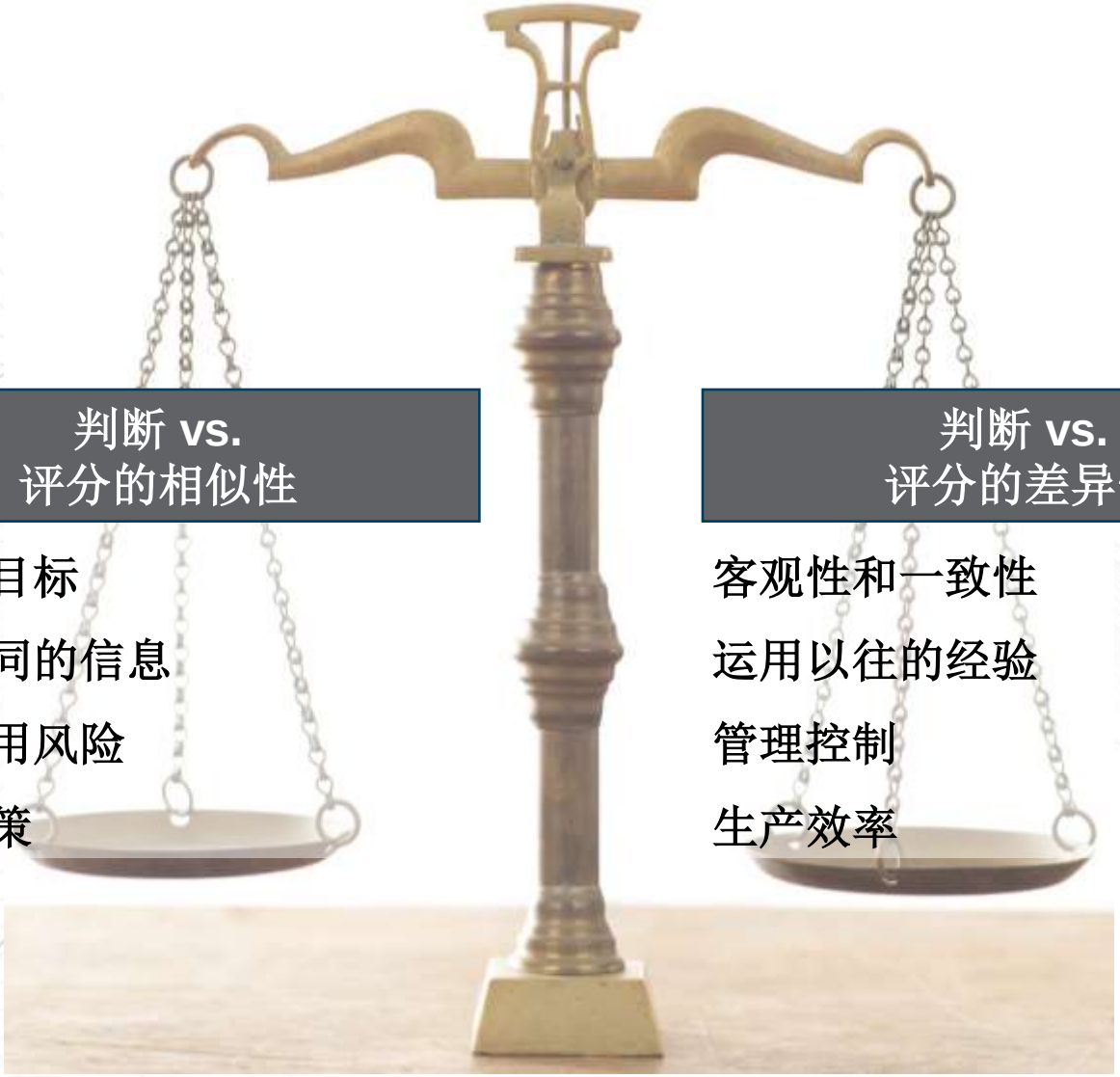
属性

	评分
在档时间	
0 to 5	6
5 to 10	10
10 to 15	28
15 to 20	35
20+	42
无相关信息	20
询问次数	
0	25
1	20
2	14
3+	10
无相关信息	14
逾期30天的次数	
0	30
1	20
2 to 3	14
4+	5
无相关信息	14
居住状态	
自有房产	40
租赁	15
与父母同住	20
雇主提供	22
无相关信息	20

	评分
使用	
小于25%	40
25%到50%	30
50%到75%	25
75%到90%	20
90%以上	13
无相关信息	25
工作时间	
小于6个月	16
6个月到3年	20
3年到5年	27
5年以上	38
无相关信息	20
客户情况	
有	30
没有	15
未进行调查	0
无相关信息	0
破产指示器	
有	10
没有	27
无相关信息	18

© 2013 Fair Isaac Corporation. Confidential.

评分系统的功能



判断 vs. 评分的相似性	判断 vs. 评分的差异性
相同的目标	客观性和一致性
基于相同的信息	运用以往的经验
评估信用风险	管理控制
作出决策	生产效率

© 2013 Fair Isaac Corporation. Confidential.

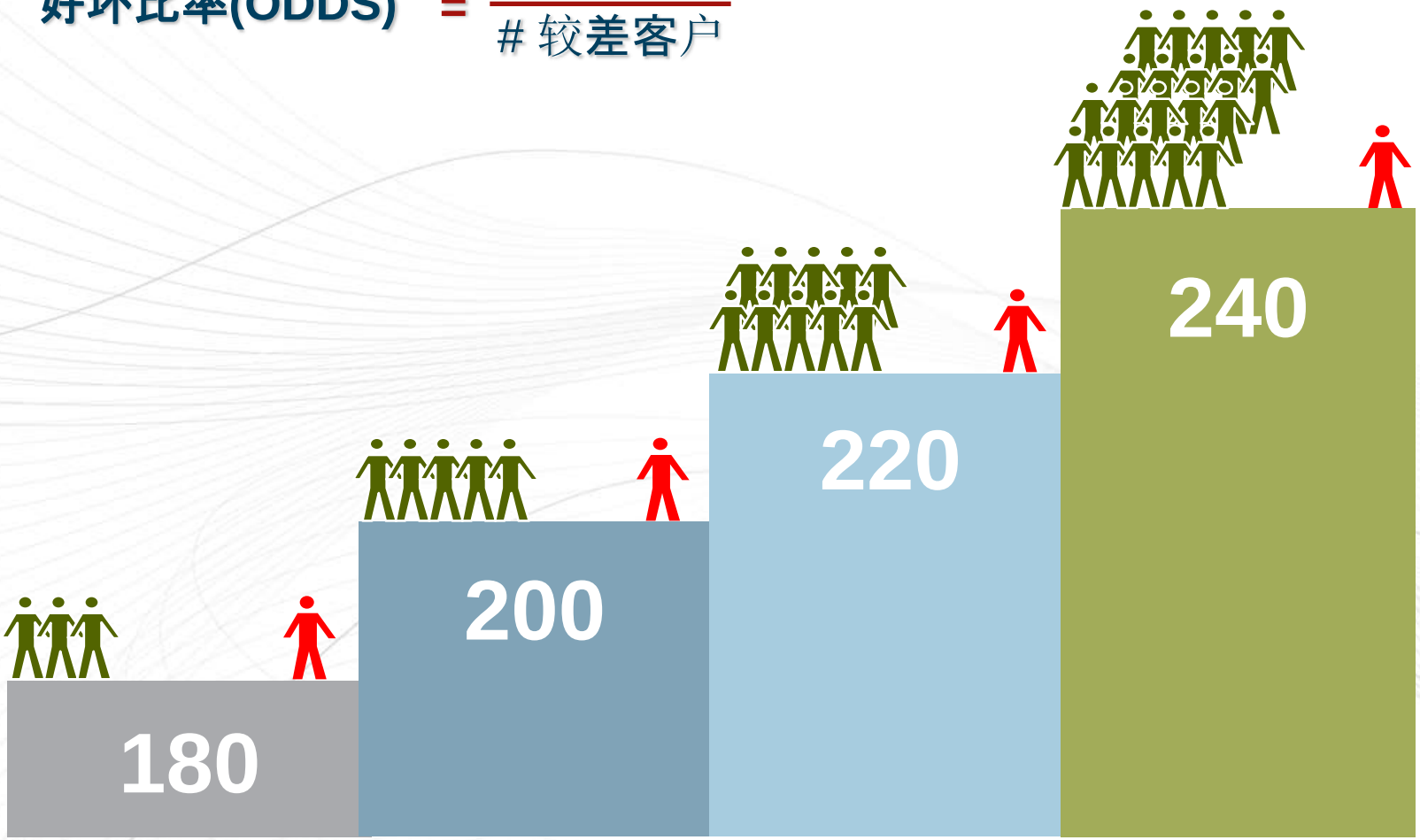
信贷风险评估判断 vs. 评分

特点	属性	判断 # 1	判断 # 2	信贷评分
存档的平均月数	120	+	+	47
已拖欠60天	2	-	-	30
已逾期月数	24	-	+	40
信用卡数量	3	+	-	45
询问数量	2	+	-	27
总计		+	-	189
决策		接受	拒绝	接受
偿还的可能性		?	?	29:1

© 2013 Fair Isaac Corporation. Confidential.

评分是为了划分风险等级

$$\text{好坏比率(ODDS)} = \frac{\text{\# 优良客户}}{\text{\# 较差客户}}$$



© 2013 Fair Isaac Corporation. Confidential.

会议内容

一. 评分模型的基本概念

- 评分的概念
- 生命周期中评分的应用

二. 申请评分模型开发方法

三. 数据驱动的统计模型开发

四. 专家经验模型

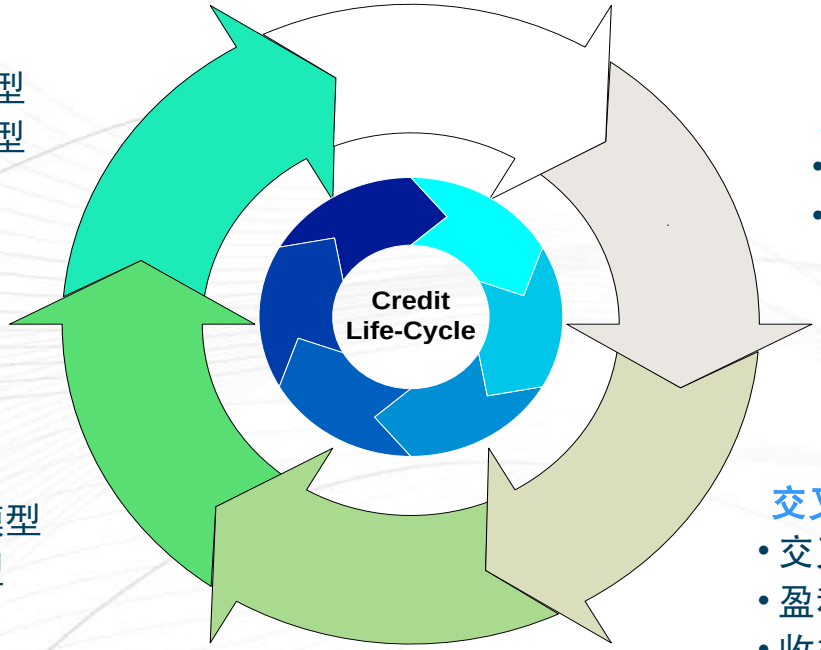
五. 混合式模型

© 2013 Fair Isaac Corporation. Confidential.

在生命周期中应用不同的评分模型

新客户筛选
- 风险预筛选模型
- 响应模型

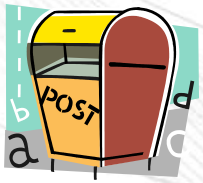
催收和回收
- 早期催收模型
- 晚期催收模型
- 回收模型



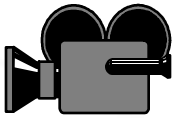
新客户审批
- 申请评分模型
- 申请欺诈模型

客户管理
- 行为评分模型
- 使用率模型
- 盈利模型
- 交易欺诈模型

交叉销售
- 交叉销售模型
- 盈利模型
- 收益评分模型

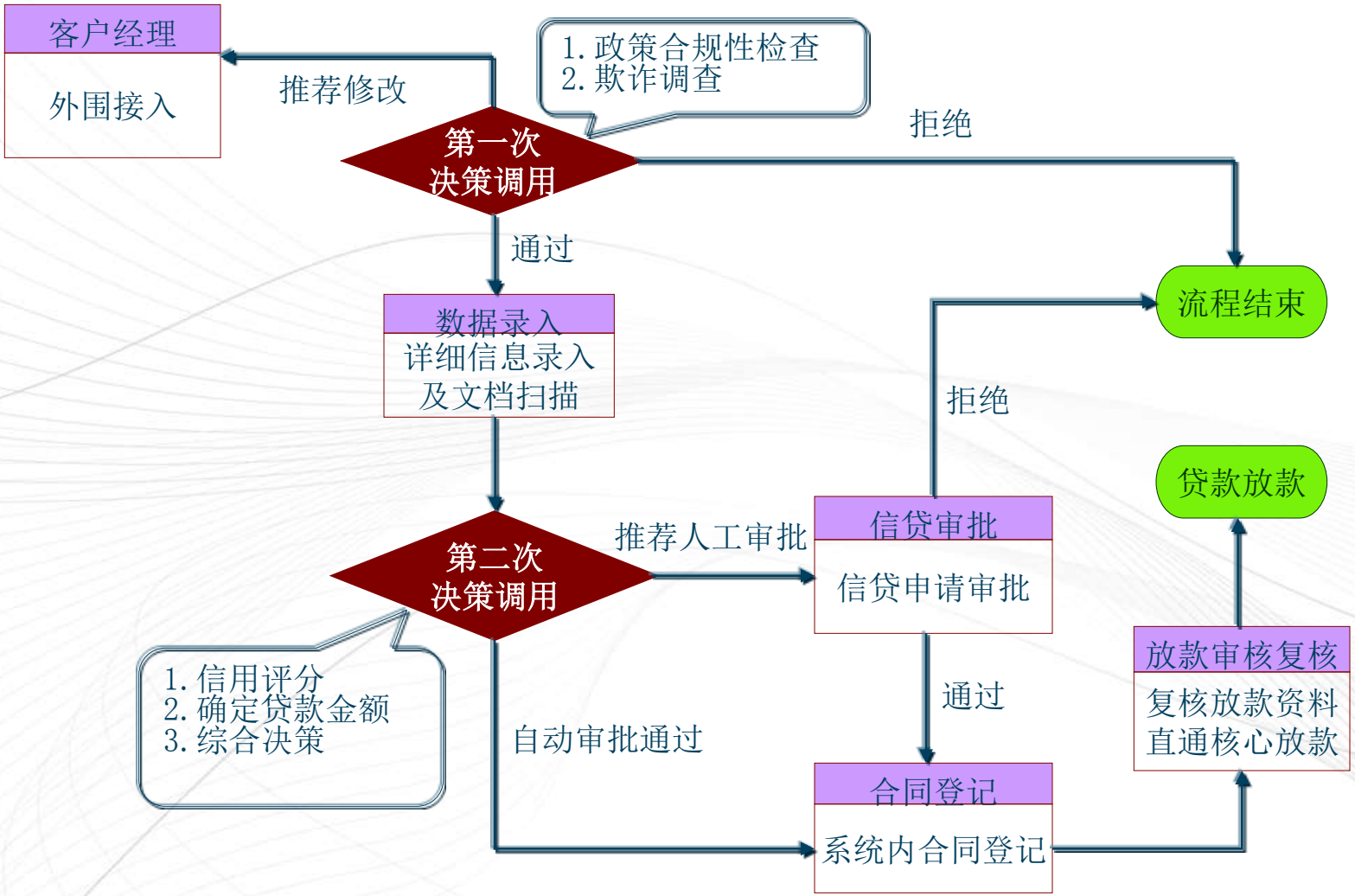


激活及客户挽留
- 提前还款模型
- 客户流失模型



© 2013 Fair Isaac Corporation. Confidential.

应用评分策略的审批流程示例



© 2013 Fair Isaac Corporation. Confidential.

设定最初的信用额度 策略举例– 银行卡

评分 (高分低风险)	还款能力 （收入、可支配收入、债务与收入的比率等等）		
	低	中等	高
低	拒绝	拒绝或先予以较低的信用额度	较低的信用额度
中等	较低的信用额度	中等的信用额度	中等到较高的信用额度
高	中等的信用额度	较高的信用额度	非常高的信用额度

© 2013 Fair Isaac Corporation. Confidential.

会议内容

一. 评分模型的基本概念

二. 申请评分模型开发方法

三. 数据驱动的统计模型开发

四. 专家经验模型

五. 混合式模型

© 2013 Fair Isaac Corporation. Confidential.

模型开发方法

专家模型

- 专家模式是适用于**即将开展，或刚开展，或已开展但数据量较少的业务**，不需要确保足够数量的“坏客户” 和表现期。

混合型模型

- 若在一个**合理的时间段**内能**有一定数量的坏客户**，可结合专家经验与所可用数据的统计分析，实施一种”混和”型的模型发展方案。

数据驱动的统计模型

- 若有**足够坏账户数**，且**各类信息**，如申请人信息、 债务或债项信息、还款信息、经营实体信息（如中小企业和小微产品）**收集较为完备**，则可开发基于数据的统计模型。

会议内容

一. 评分模型的基本概念

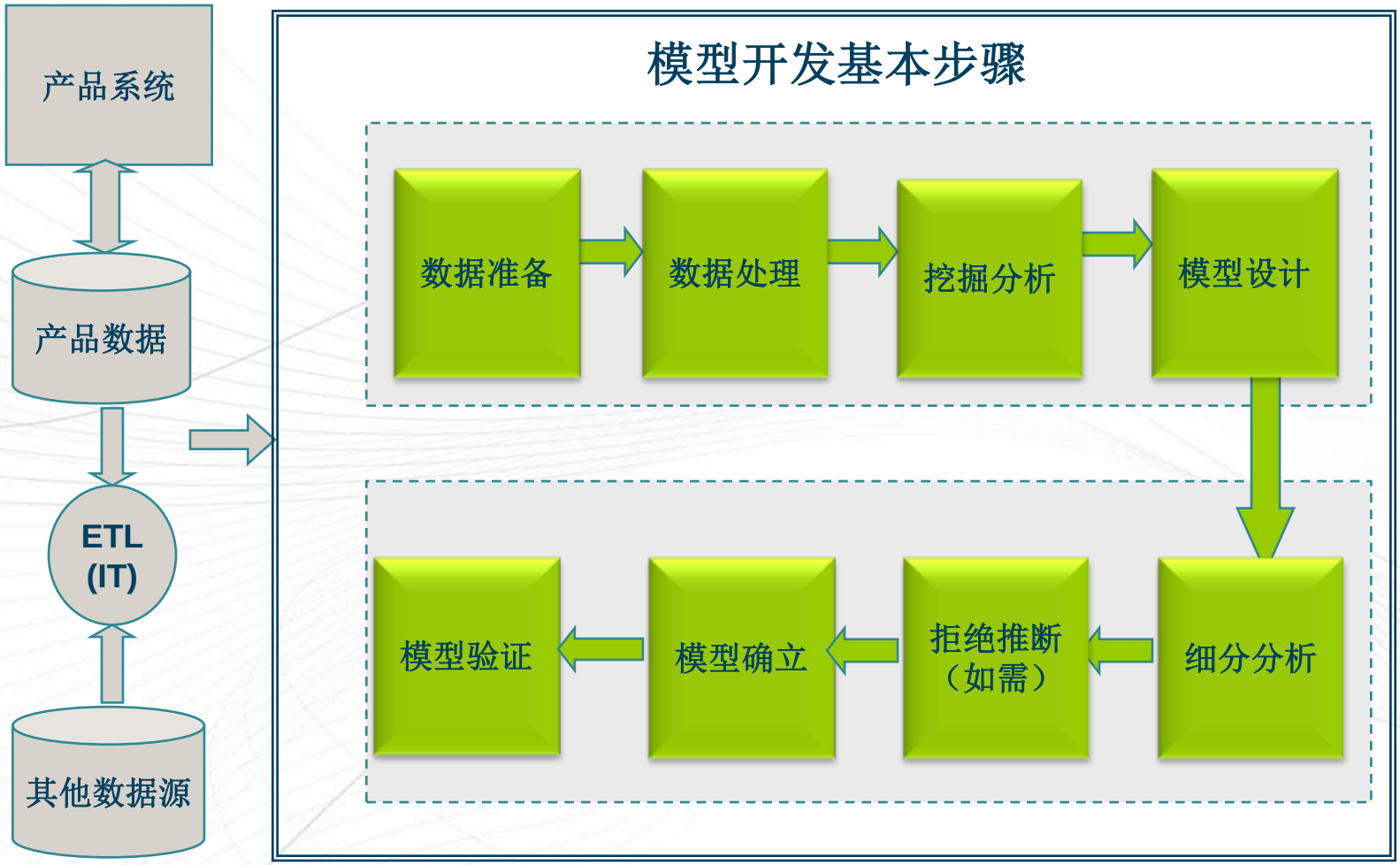
二. 申请评分模型开发方法

三. 数据驱动的统计模型开发

四. 专家经验模型

五. 混合式模型

统计模型开发的基本步骤



© 2013 Fair Isaac Corporation. Confidential.

数据分析和准备

提取		分析		汇总	
数据现状	理解数据	初步挖掘分析	产品重叠性分析	月度分析数据	生成汇总变量
• 理解银行产品结构和特点 • 理解银行数据现状 • 数据字典以及表结构 • 提供数据抽取模板 • 制定数据抽取方案	• 数据理解 • 变量及信息的匹配 • 数据时间跨度、来源及总量的检测 • 数据合理性、缺失等的检测	• 单项分析 • 多维分析 • 相关变量分析 • 不同地区、产品、时间段分析	• 数据表关联 • 生成内部关系变量 • 内部关系数据检测	• 数据对应与汇总 • 推导和计算项目所需变量 • 生成月度数据集市	• 项目所需数据信息的进一步汇总 • 滚动率分析 • **Vintage Analysis** 等

© 2013 Fair Isaac Corporation. Confidential.

数据集样例

» 申请变量-申请人

家庭净资产、年龄、性别、婚姻状况、居住状况、 教育、从业年限等。

» 申请变量-企业

» 企业成立年限、注册资本、销售收入、资产负债率、 行业类别等。

» 征信局变量

» 最大逾期月数
» 最近24个月的逾期次数等
» 过去12个月的信用卡审批查询次数
» 所有交易的全部余额
» 6个月银行卡的平均使用率
» 银行卡最大的信用额度
» 交易历史长度等

» 内部关系

» 存款日均余额
» 企业结算额
» 信用卡交易信息
» 在我行信贷关系年限等

© 2013 Fair Isaac Corporation. Confidential.

数据提取检查

- 数据的时间段检查
  - 检查是否是要求的时间段

- 数据关键变量检查
  - 根据数据提取要求及数据内容确定关键变量
  - 如：客户的年龄分布、客户的地域分布、客户持有的信贷产品数量分布等

© 2013 Fair Isaac Corporation. Confidential.

数据合并逻辑

- 根据数据处理目标确定何时进行数据合并
- 根据表间的关系确定表合并逻辑
- 检查各个需要合并的数据表的数据主键
  - 数据主键的唯一性检查

© 2013 Fair Isaac Corporation. Confidential.

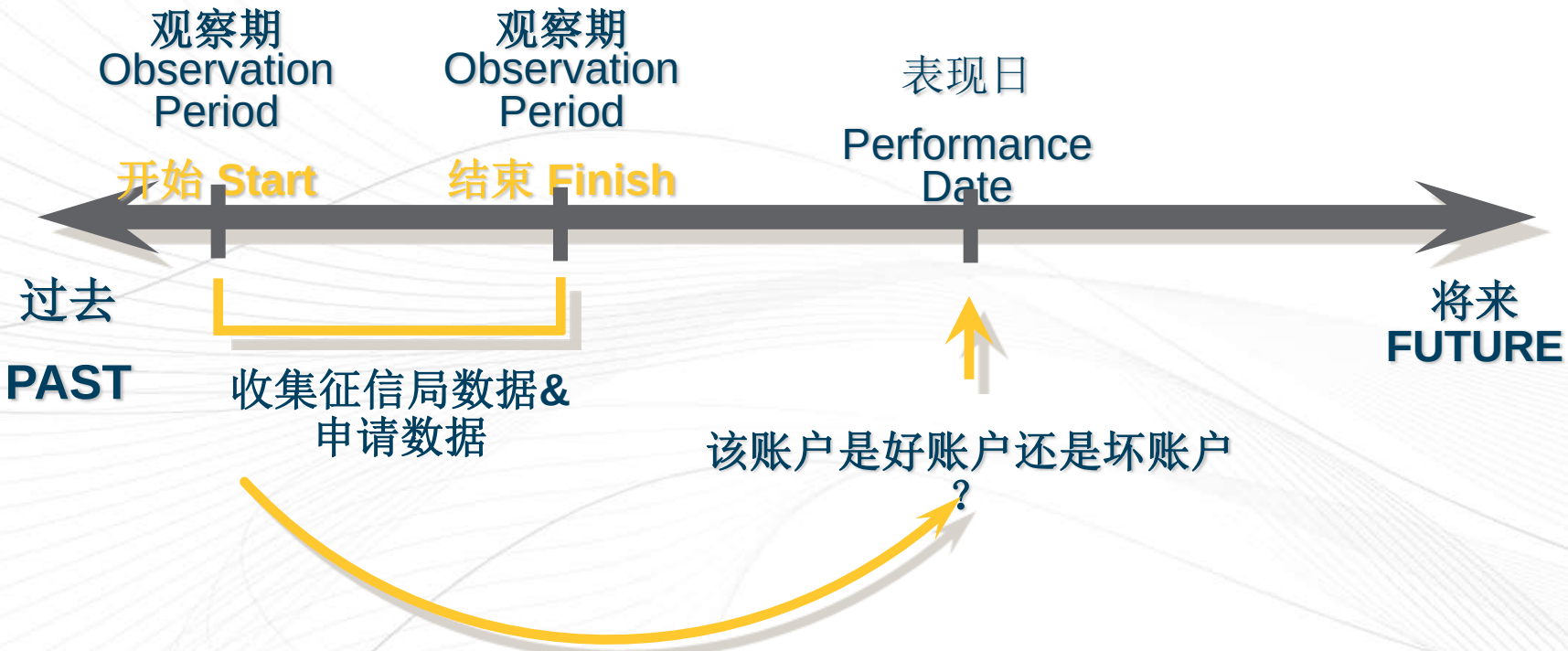
数据质量检查步骤

- **变量结构检查**
  - » 变量属性、长度、格式、描述
  - » 变量含义
- **变量分布检查**
  - » 频数分布
  - » 基本统计值
  - » Binning分布
- **变量逻辑检查**
  - » 重复记录
  - » 单个变量
  - » 变量间逻辑

实际数据情况复杂，需随需而变

© 2013 Fair Isaac Corporation. Confidential.

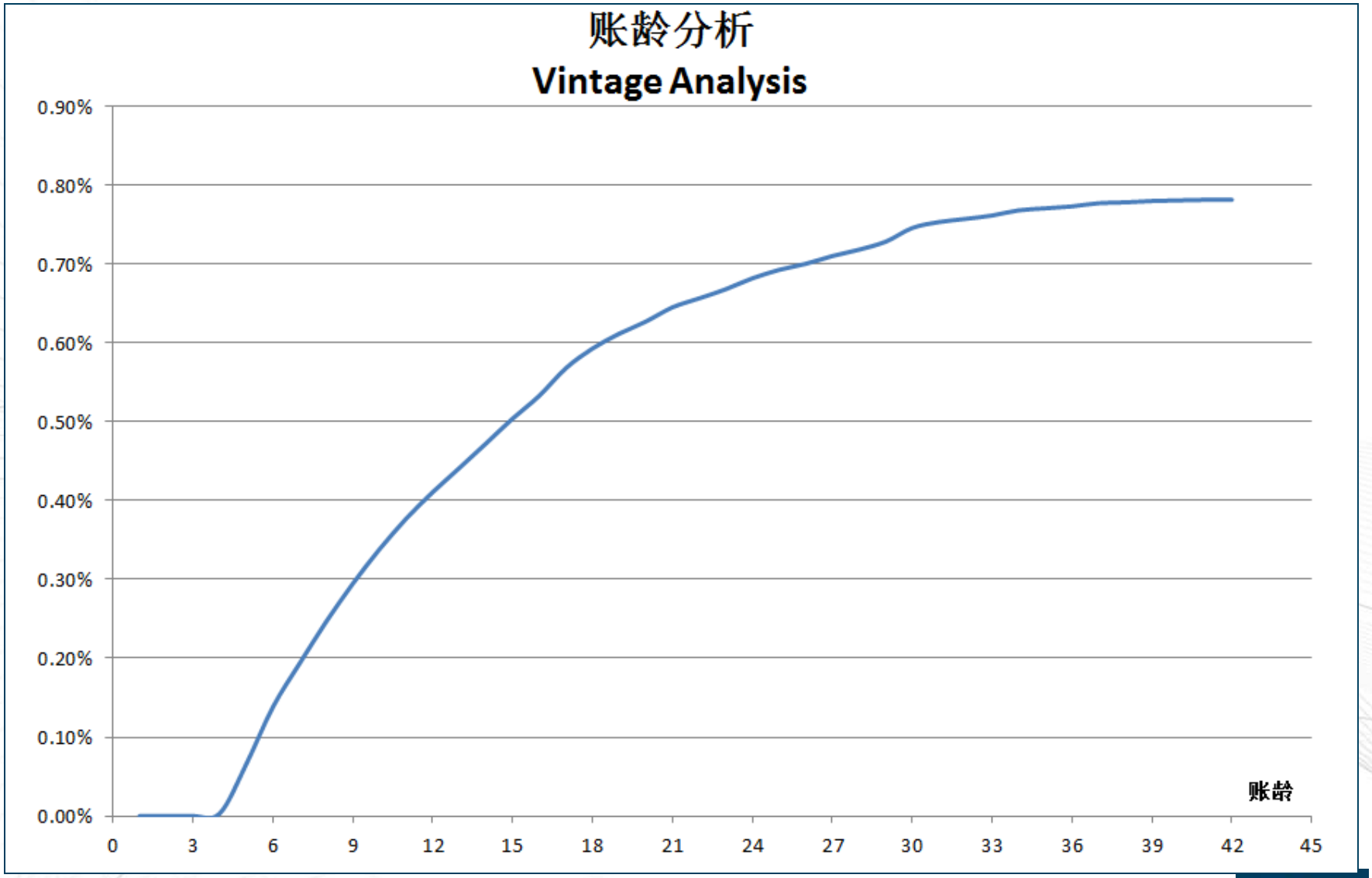
模型设计—样本窗口



appID	appAge	appIncome	appResidence	appTimeAddress	loanFate
'PI230000036-06...	MV00	2858.	'Owns Home'	22	'App rejected'
'PI230000153-06...	19.000	1197.	'Parents'	2	'Paid as agreed'
'PI230000377-07...	22.000	1417.	'Renting'	4	'App rejected'
'PI230000495-06...	39.000	3045.	'Renting'	3	'Charged-off'

© 2013 Fair Isaac Corporation. Confidential.

模型设计



© 2013 Fair Isaac Corporation. Confidential.

模型设计 — 滚动率分析

Roll Rate		Current Month Delinquency Status									
		No Amou	M0	M1	M2	M3	M4	M5	M6	M7	M8
Previous Month Delinquency Sraus	No Amount Due	96.10%	3.90%								
	M0	33.19%	61.43%	5.38%							
	M1	42.07%	19.27%	1.24%	37.41%						
	M2	38.10%	11.65%	0.55%	0.63%	49.08%					
	M3	47.22%	2.47%	0.46%	0.53%	2.06%	47.27%				
	M4	14.50%	2.42%	0.22%	0.53%	0.53%	1.20%	80.59%			
	M5	10.02%	1.88%	0.24%	0.20%	0.40%	0.40%	0.76%	86.11%		
	M6	8.42%	1.19%	0.19%	0.14%	0.19%	0.05%	0.29%	0.52%	89.01%	
	M7	5.77%	0.55%	0.16%	0.05%	0.22%	0.11%	0.22%	0.16%	0.55%	92.19%
	M8	2.01%	0.00%	0.05%	0.04%	0.01%	0.03%	0.07%	0.05%	0.07%	97.68%

	Roll Rate Analysis		
Delinquency Status	Better	No Change	Worse
No Amount Due	N/A	96.10%	3.90%
M0	33.19%	61.43%	5.38%
M1	61.35%	1.24%	37.41%
M2	50.29%	0.63%	49.08%
M3	50.67%	2.06%	47.27%
M4	18.21%	1.20%	80.59%
M5	13.13%	0.76%	86.11%
M6	10.47%	0.52%	89.01%
M7	7.26%	0.55%	92.19%
M8	2.32%	N/A	N/A

© 2013 Fair Isaac Corporation. Confidential.

模型设计关键要素

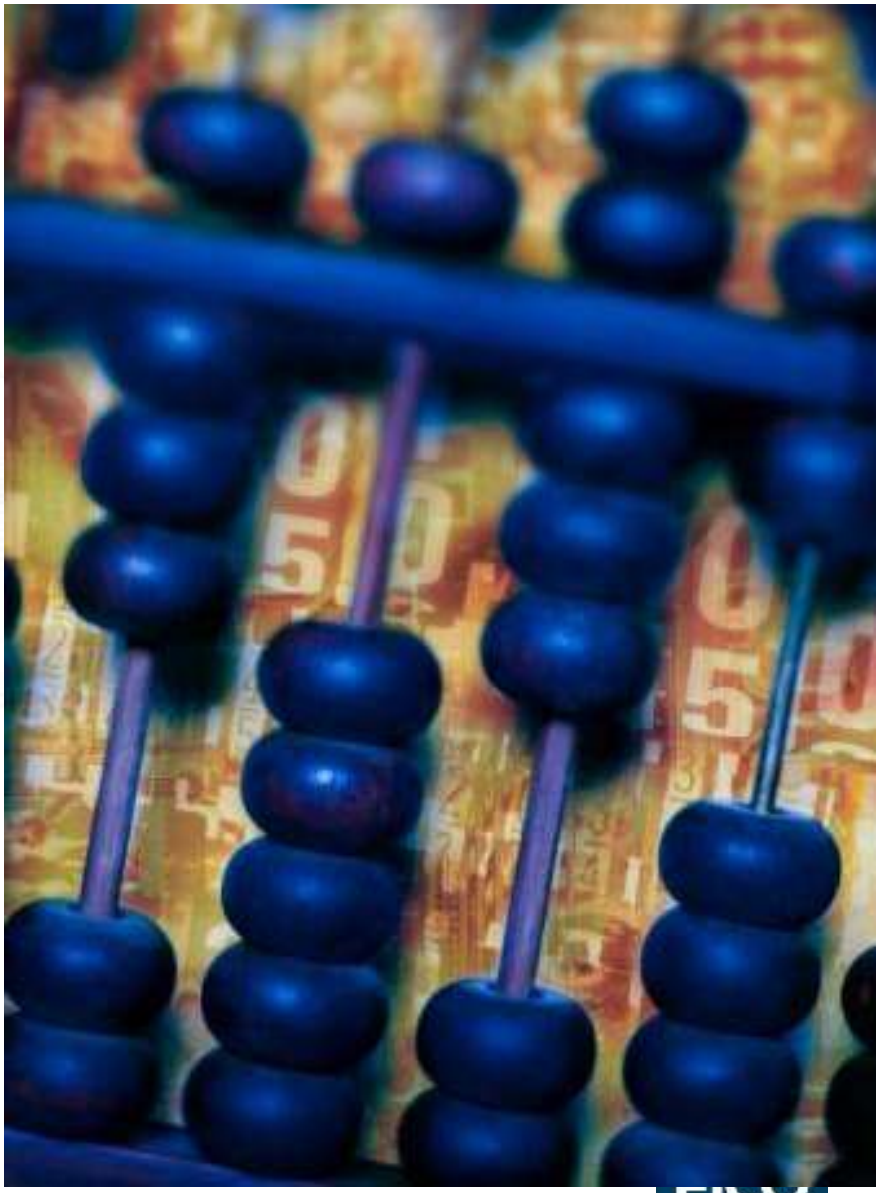
- 在表现窗口内定义好坏结果，评定帐户的风险水平
- 排除规则
  - » 申请欺诈 / 丢失或失窃
  - » 预批准 / 员工申请
- 好坏标识
  - » 好账户：从未逾期
  - » 不确定账户：逾期1-59天
  - » 坏账户：逾期60天以上

表现	分类
预批准	排除
丢失或失窃	排除
核销	坏帐户
90天以上逾期	坏帐户
60—89天逾期	坏帐户
30—59天逾期	不确定
1—29天逾期	不确定
从未逾期	好帐户

© 2013 Fair Isaac Corporation. Confidential.

抽样

- » 成本:
  - » 每个账户的征信局报告
  - » 大规模数据的处理成本
- » 时间:
  - » 程序的执行时间漫长
  - » 漫长的开发流程
- » 根据大样本定律 / 中央极限理论，我们可以用一部分的样本去代表整个数据群体。

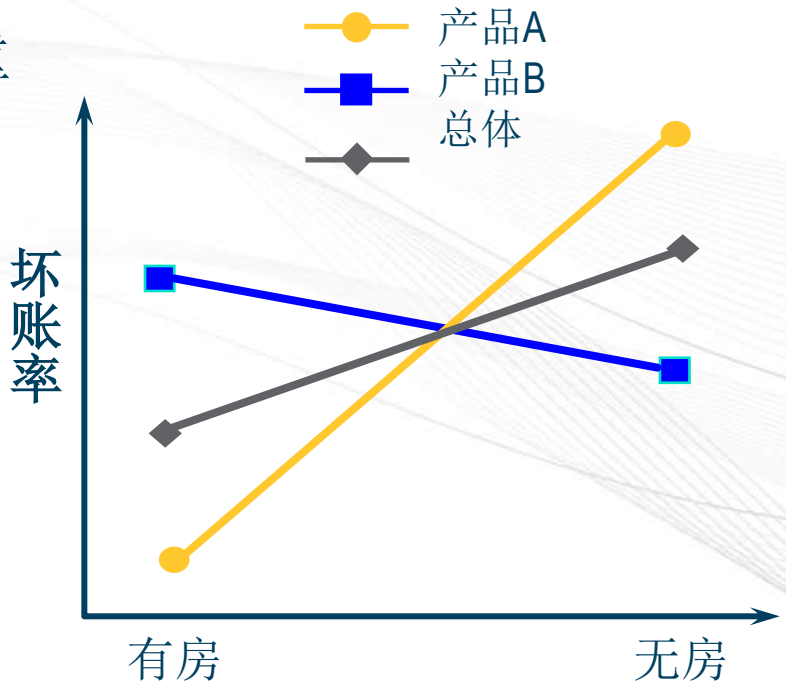


© 2013 Fair Isaac Corporation. Confidential.

细分分析

- 目的在于辨别最佳群体细分，使整个模型的预测能力最大化
- 不同人群行为模式存在差别

- 细分方案的业务方面的考量占很大比重
  - » 生成细分
    - » 启发型(Heuristic)
    - » 经验型(Empirical)

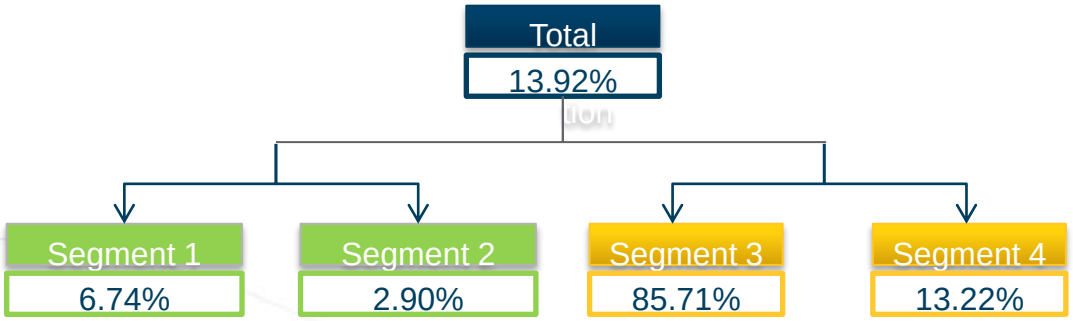


© 2013 Fair Isaac Corporation. Confidential.

细分分析

» 细分分析的过程

- 建立初始细分方案
- 根据业务需要进行调整



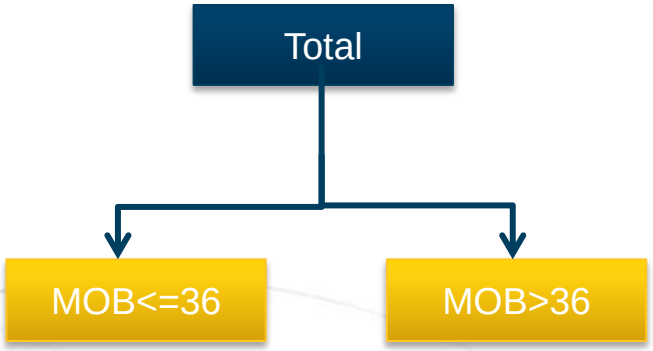
» 细分的评估

评估的标准非常重要，诸如：

- 样本的数量是否足够用以构建以细分为基础的稳健的评分模型
- 每个评分卡是否会对某一资产中足够数量的账户产生影响
- 细分方案是否合乎逻辑
- 细分方案是否符合业务要求

© 2013 Fair Isaac Corporation. Confidentia

细分分析—方案评估



方案2（MOB）	无细分 主模型(1个评分卡)			有细分 子模型(2个评分卡)		
	总体	估算	验证	总体	估算	验证
总体	63.14	63.20	63.57	65.02	65.48	64.15
MOB<=36		63.29	60.82		62.82	61.16
MOB>36		61.23	64.95		61.97	64.91

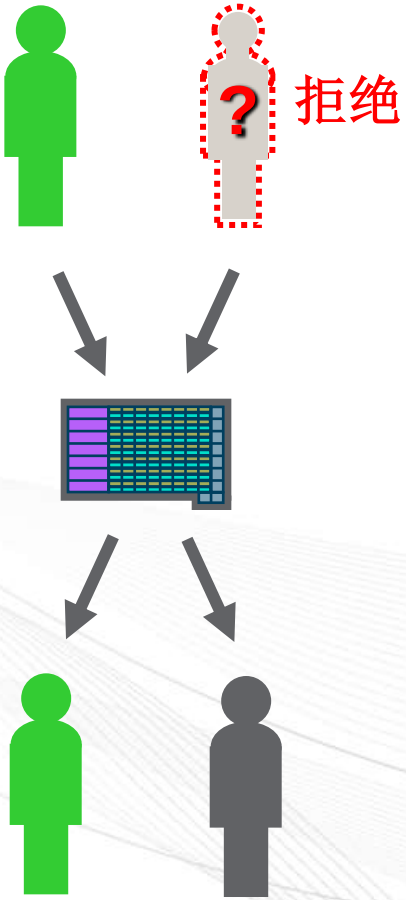
© 2013 Fair Isaac Corporation. Confidential.

拒绝推断的概述

» 理论上将先前拒绝的账户推断为好账户概率的流程

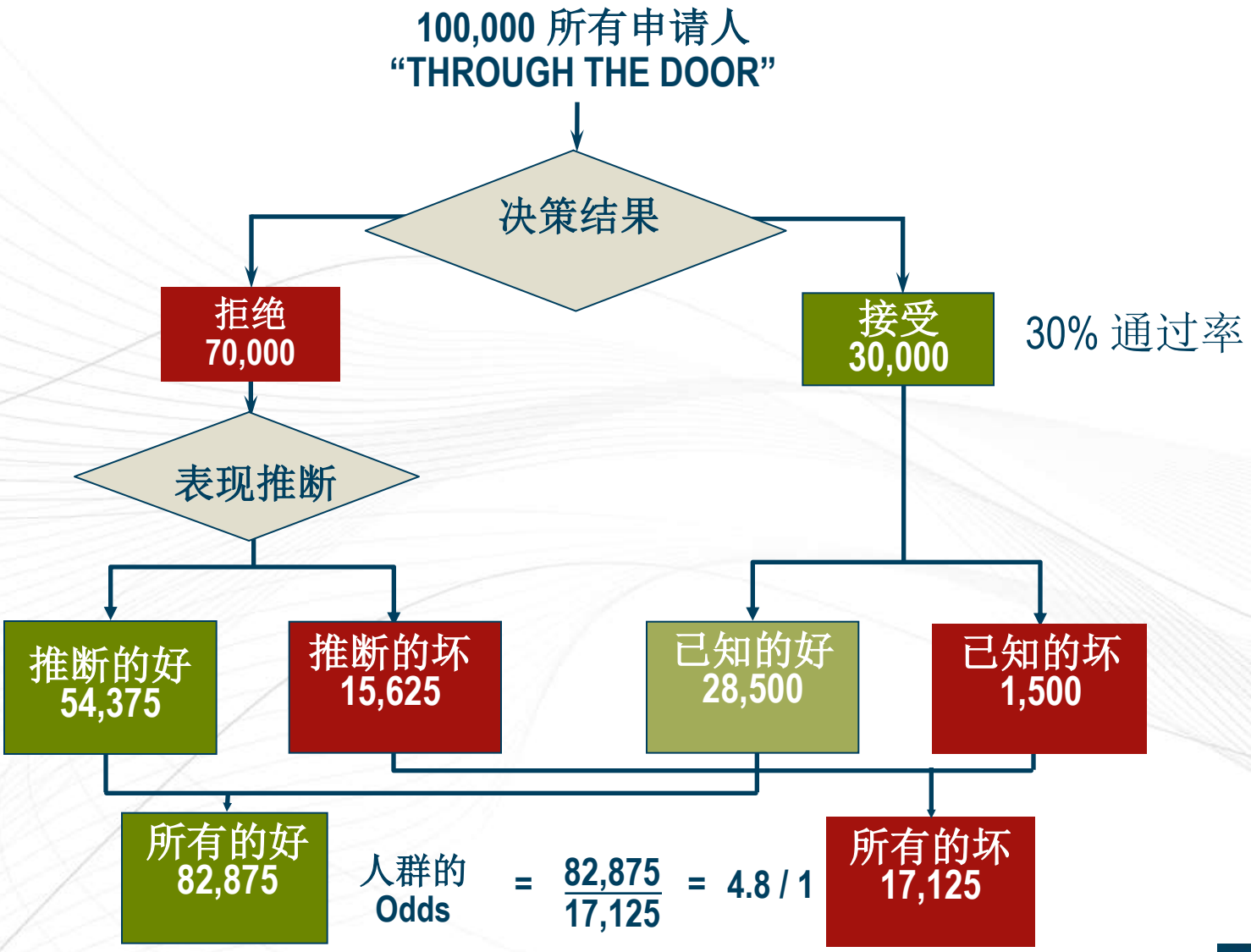
» 为什么进行拒绝推断？

- 拒绝是通过申请评分卡进行评估的申请群体的一部分
- 拒绝表现在模型开发是暂不可获得
- 拒绝推测是用于预测被拒绝的客户表现的统计技术
- 拒绝推测使我们可以创建“总体正常申请群体(TTD)”，从而避免在模型开发是出现统计偏差



© 2013 Fair Isaac Corporation. Confidential.

拒绝推测后的申请数据流 (所有好客户/所有坏客户)

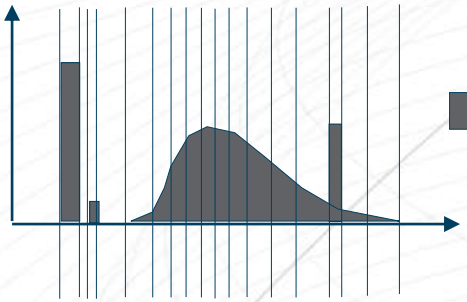


© 2013 Fair Isaac Corporation. Confidential.

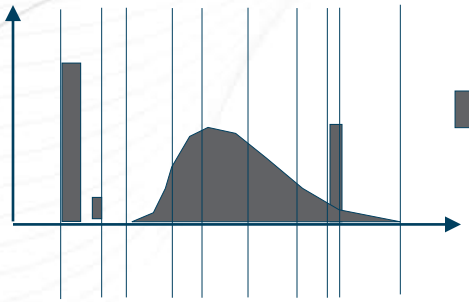
模型开发 — 特征变量分析（分箱）

预测变量 → *细分箱* → *粗分箱* → *分箱特征项* → *评分模型*

*细分箱**Fine bins***



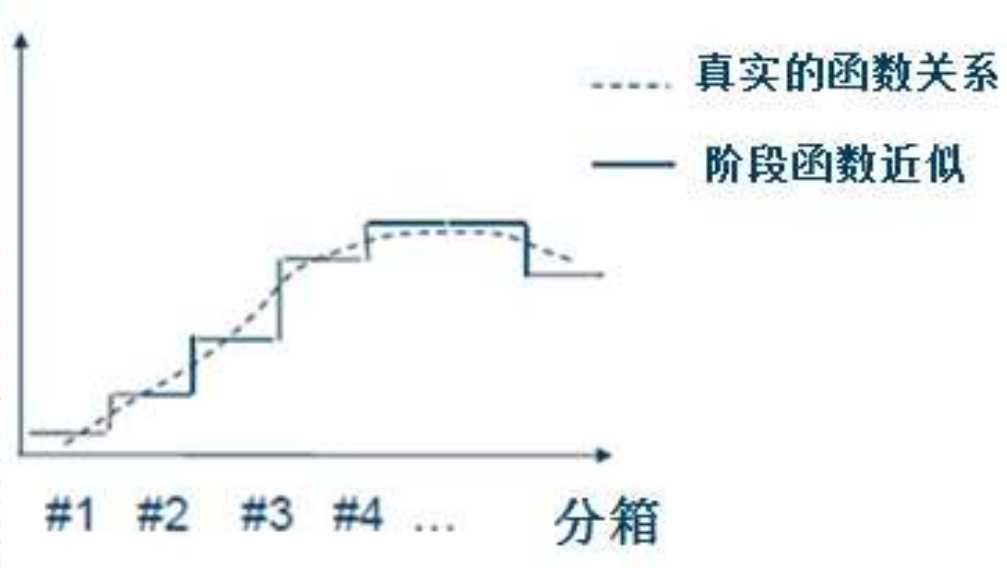
*粗分箱**Coarse bins***



逾期30天的次数	评分
0	30
1	20
2 to 3	14
4+	5
无相关信息	14

© 2013 Fair Isaac Corporation. Confidential.

模型开发 — 特征项分箱



» 易于解释，结合行业经验

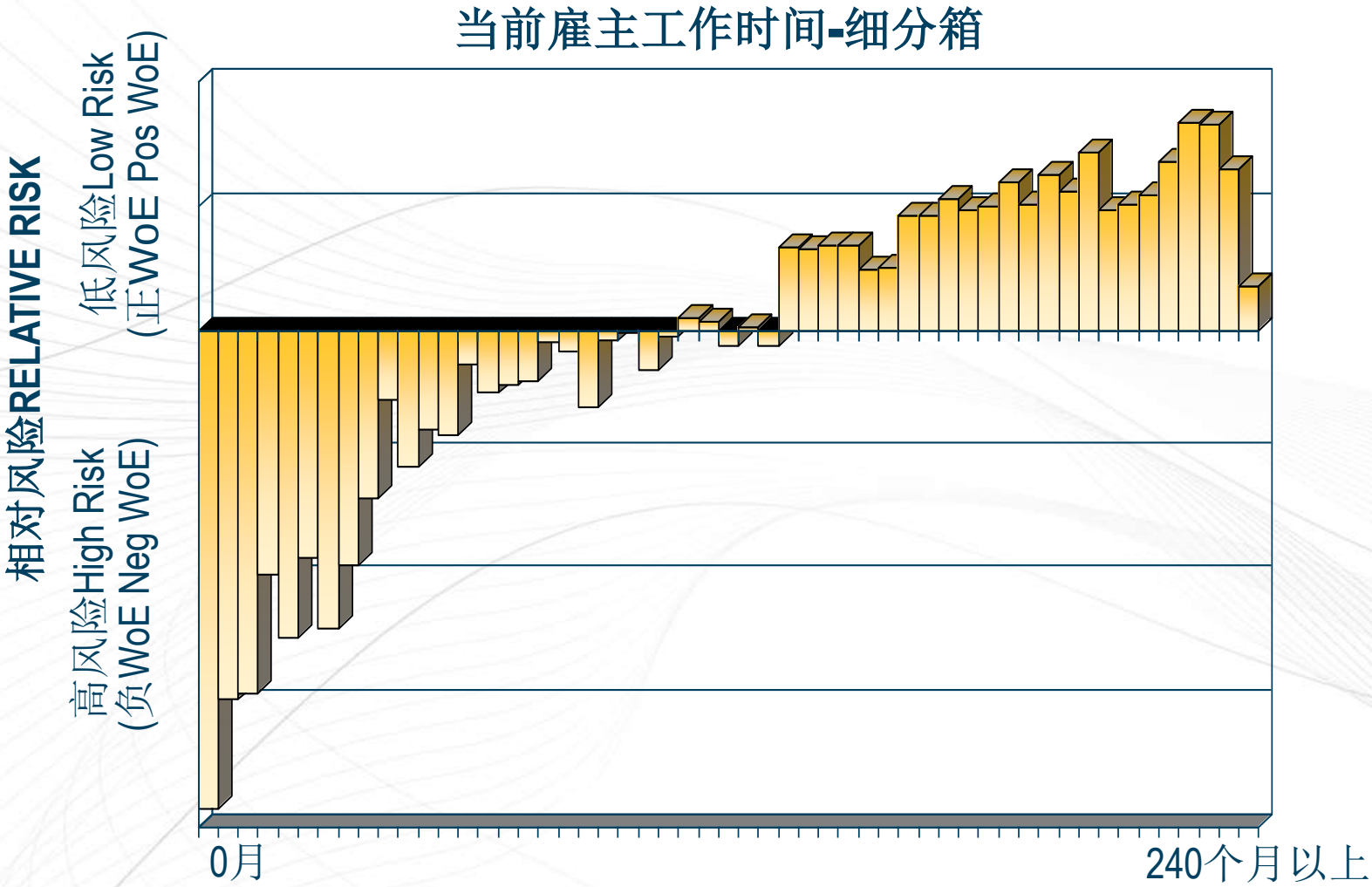
» 是一种复杂的数据变换手段，可以捕捉非线性关系

» 离散变量和连续变量的统一处理，并且可以处理缺失值和特殊值

© 2013 Fair Isaac Corporation. Confidential.

细分箱

» 细分箱显示趋势和噪音

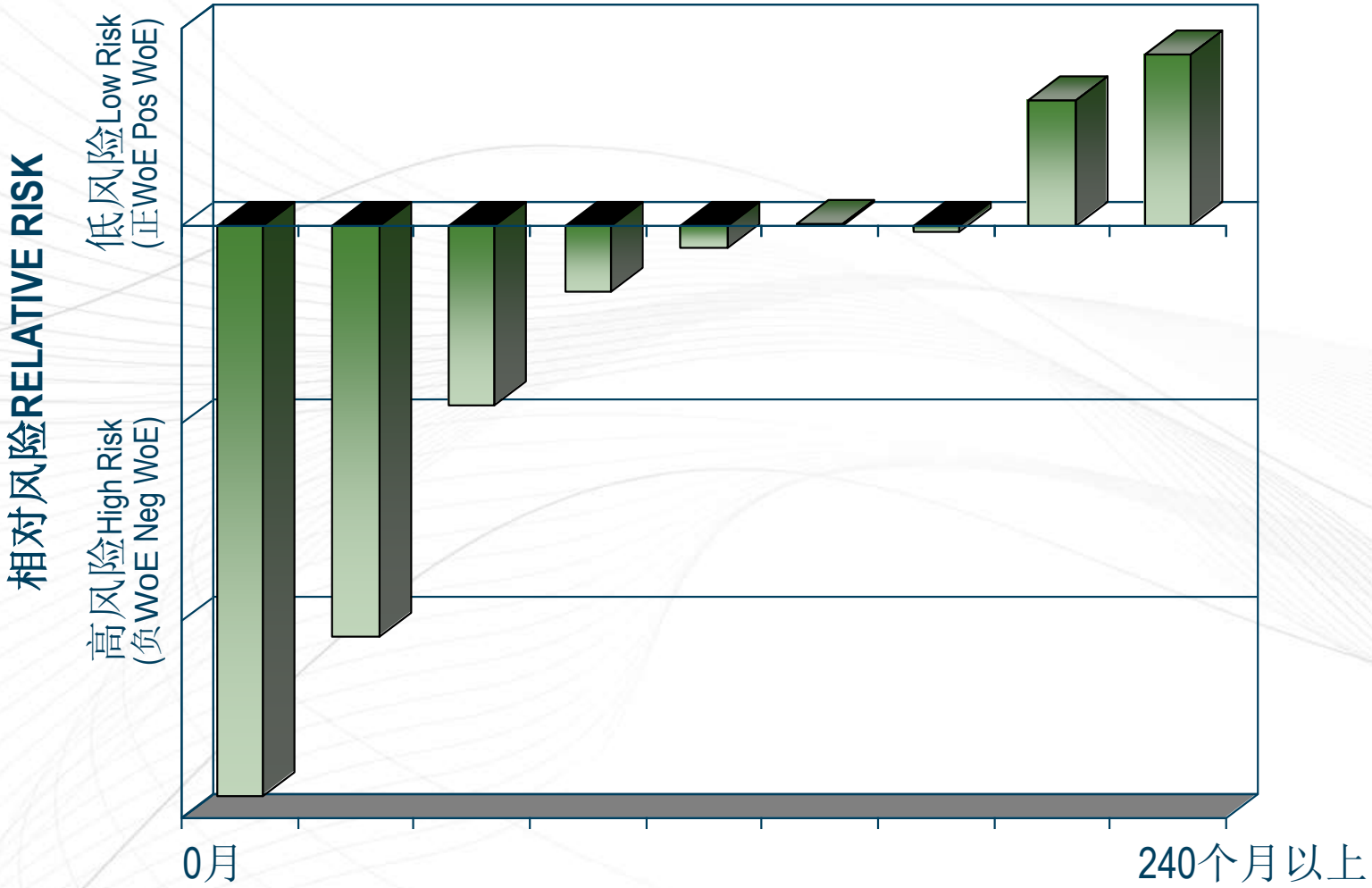


© 2013 Fair Isaac Corporation. Confidential.

粗分箱

» 粗分箱保留趋势并降低噪音

当前雇主工作时间-粗分箱



© 2013 Fair Isaac Corporation. Confidential.

例子

合并/拆分

上移/下移

Interactive Binner - edu: 教育

Index	Label	Constraint		# Bins	targetKGB (binary)								
		Operator	Value		r Good	r Bad	Total	% Good	% Bad	Odds	I.V.	WoE	WoE
0	硕士研究生	> bin #	1	1	488	10	3,626.9	9.8	2.7	381.667	0.089	1.268	
1	大学本科	> bin #	2	1	2,626	105	19,565.5	52.5	28.8	195.601	0.142	0.599	
2	专科	> bin #	3	2	1,198	115	8,989.5	24.0	31.6	81.475	0.021	-0.277	
3	高中	> bin #	4	1	465	93	3,535.1	9.3	25.5	39.105	0.164	-1.011	
4	初中、小学	> bin #	5	2	161	40	1,231.4	3.2	11.0	31.480	0.095	-1.228	
5	缺失、未知			2	62	1	460.5	1.2	0.3	484.905	0.015	1.507	
6	ELSE			1	0	0	0.0						

$$WOE = \ln\left(\frac{\%G}{\%B}\right)$$

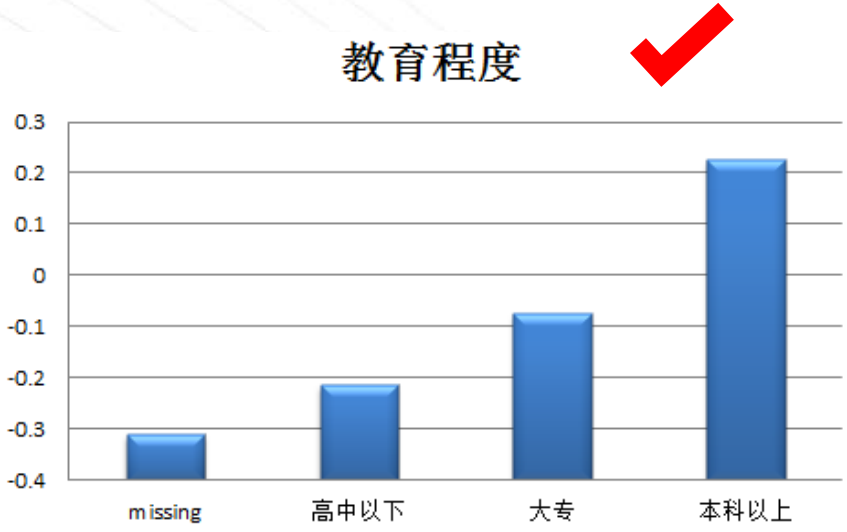
$$IV = \sum_i (\%G_i - \%B_i) \times \ln\left(\frac{\%G_i}{\%B_i}\right)$$

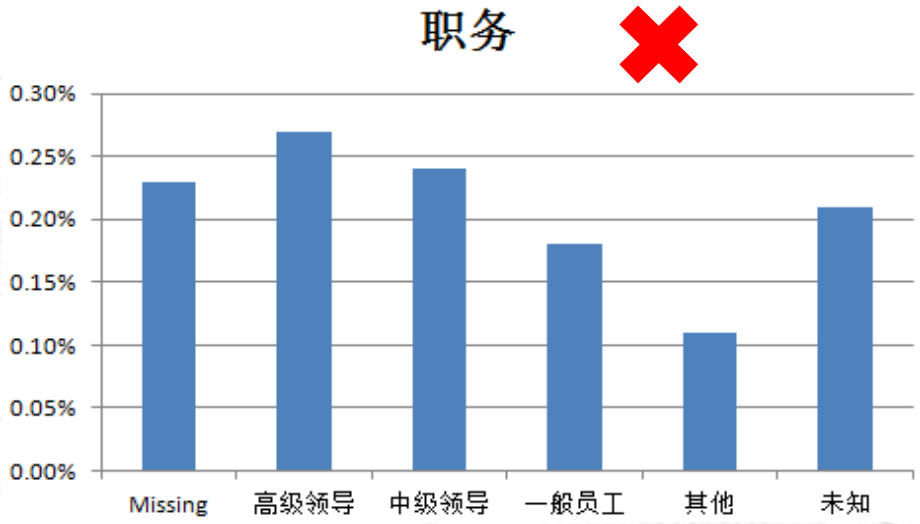
© 1996 Fair, Isaac and Co, Inc.

© 2013 Fair Isaac Corporation. Confidential.

粗分箱

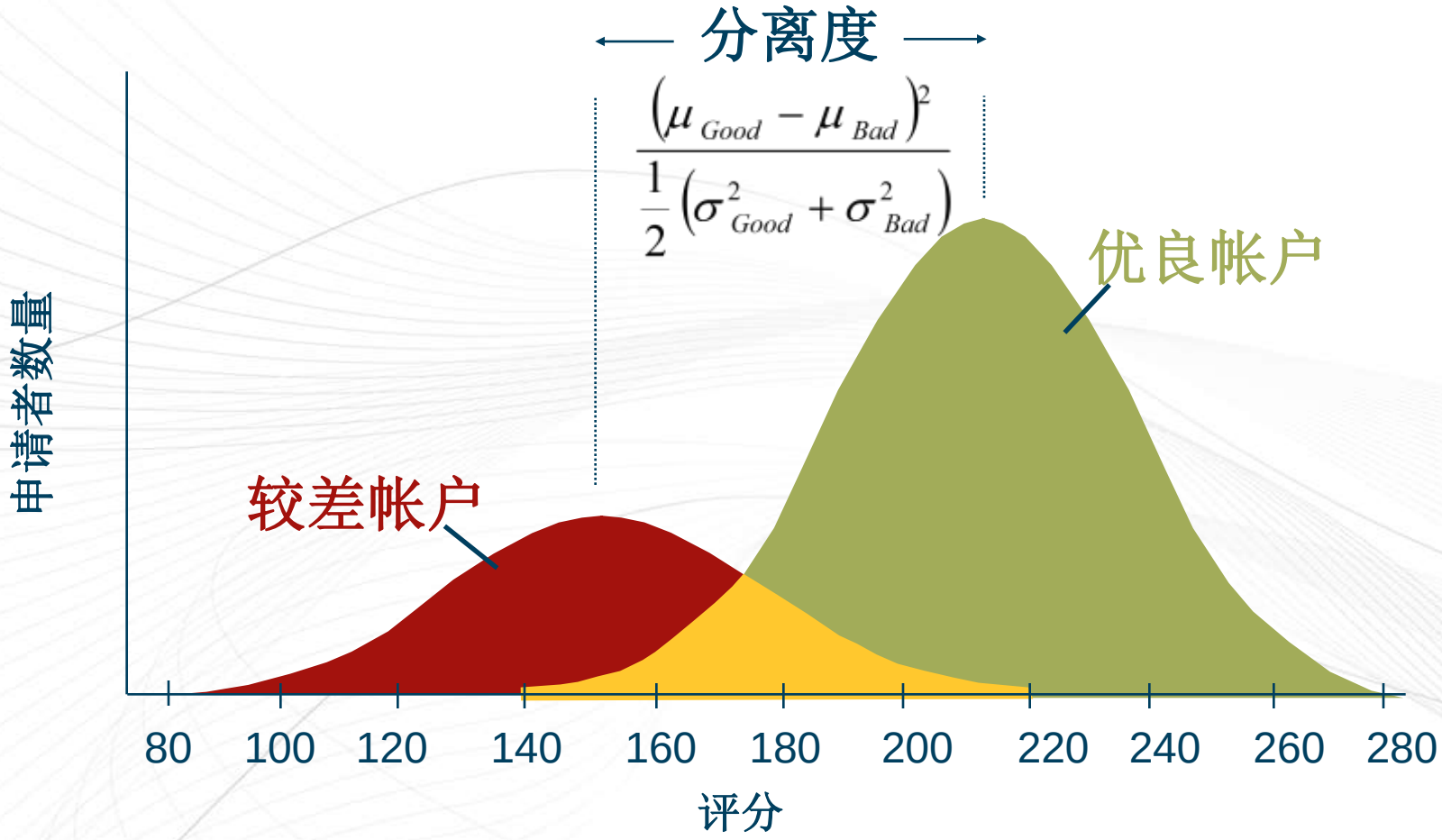
变量的风险趋势应当合理，业务可解释





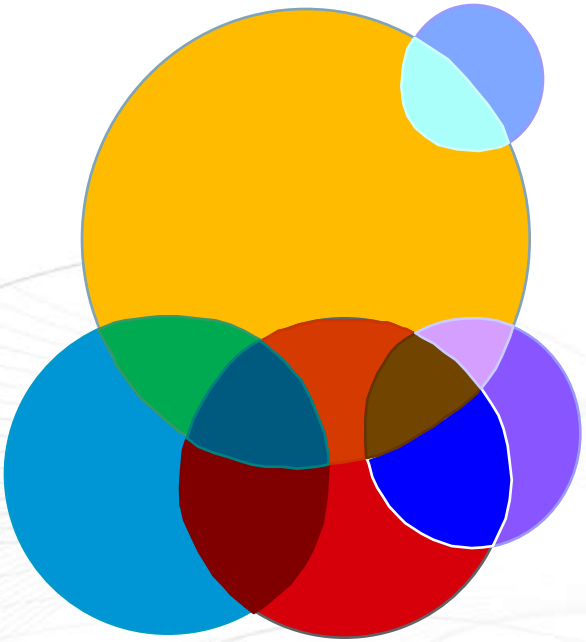
© 2013 Fair Isaac Corporation. Confidential.

模型开发 — 分离度 Divergence



© 2011 Fair Isaac Corporation. Confidential.

相互关系



相互关系

$$\rho = \frac{E(X_1X_2) - E(X_1)\,E(X_2)}{\sqrt{Var[X_1]\,Var[X_2]}}$$

© 2013 Fair Isaac Corporation. Confidential.

对每个特征进行评估

	信息价值
目前职位工作的时间	0.251
目前地址居住的时间	0.133
信用卡总数	0.042

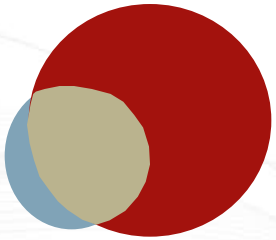
信息价值是单个特征相对的总体预测能力 .

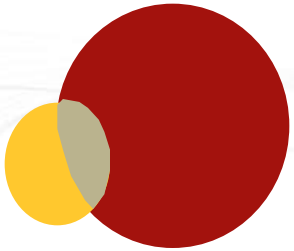
信息价值更多，相对的，特征的预测行更强.

© 2013 Fair Isaac Corporation. Confidential.

举例说明: 特征集的选择

	边际贡献 Marginal Contribution	分离度 Divergence
目前职位工作的时间 目前地址居住的时间	0.012	0.263
目前职位工作的时间 信用卡总数	0.029	0.280





“目前职位工作的时间”这一特征的信息价值为0.251

© 2013 Fair Isaac Corporation. Confidential.

Moder Builder 7 模型训练 — 模型建立

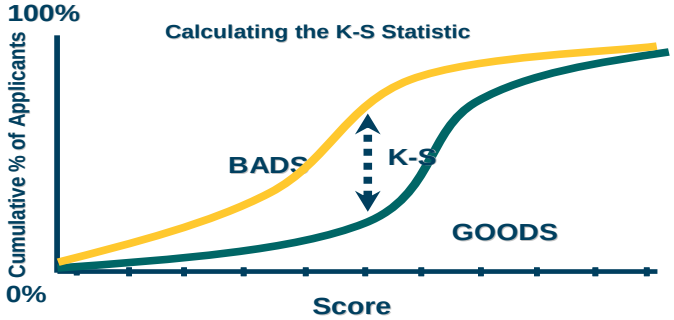
Scorecard Model Editor: Variables - AutoLoans_KGB.mb

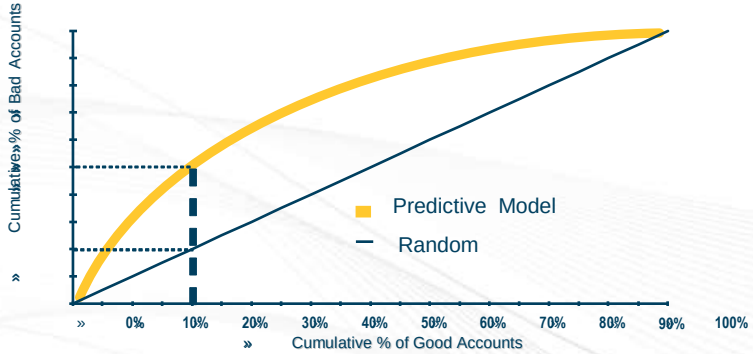
Initial	In	Tier	Variable Name/Bin Label	Variable Description	Co...t	Marginal Contribution (Ver. 1) Training	Test	Training
	☑	1	cbFICO	FICO score at time of application		0.892	1.179	
	☑	1	appOcc	Applicant occupation		0.324	0.155	
	☑	1	dealLoanToVal	Loan to value ratio		0.191	0.156	
	☑	1	appIncome	Applicant income		0.147	-0.001	
	☑	1	cbUtilizn	Net credit utilization		0.112	0.316	
	☑	1	appFinanceCo	Finance company loan reference		0.078	-0.025	
	☑	1	appChkSv	Checking and sav...count reference		0.072	0.032	
	☑	1	cbInq5Mos	# of inquiries in last 5 months		0.064	0.051	
	☑	1	cbTimeFile	Time in CB file (months)		0.045	0.112	
	☑	1	cbMosInq	Months since last inquiry		0.044	0.060	
		1	cbMinOpen	Months since newest trade opened		0.027		
		1	appTimeAddress	Time at address (months)		0.010		
		1	cb90Ever	# of times ever 90+ days past due		0.009		
		1	cbMosDlq	Months since last delinquency		0.008		
		1	dealNewUsed	New/used vehicle		0.006		
		1	cbPctGood	% trade lines in good standing		0.006		
		1	appAge	Applicant age		0.006		
		1	appResidence	Residential status		0.002		

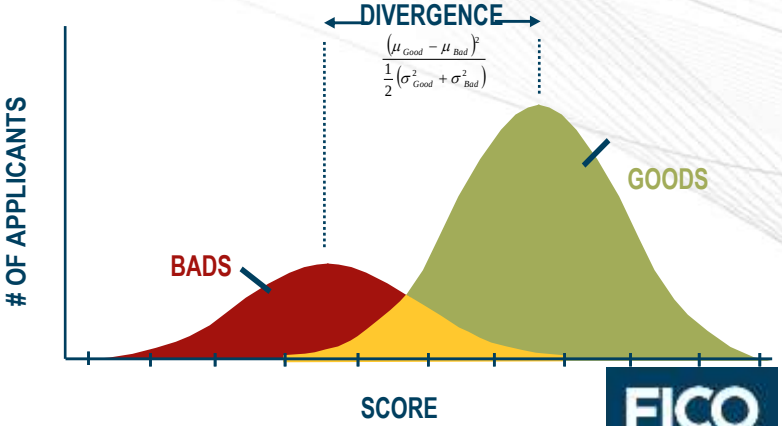
© 2013 Fair Isaac Corporation. Confidential.

评分模型的开发－模型的衡量和评估

- » 衡量模型排序能力－区分好坏的能力
  - » KS 值
    - 好坏群体累计分布的最大差异
    - KS越高，排序能力越强
  - » Gini(ROC)系数
    - 坏账户数的累积分布与随机分布的差值
    - 高值意味着好/坏之间的差异较大
  - » Divergence 分离度
    - 衡量“好”和“坏”群体的信息值差异
    - 分离度越高，排序能力越强
  - » 评分分布(Performance Chart)
    - 评分卡分数分布
    - 排序能力
    - 好坏比（Good:Bad odds）
    - 坏账率
    - 风险等级分布



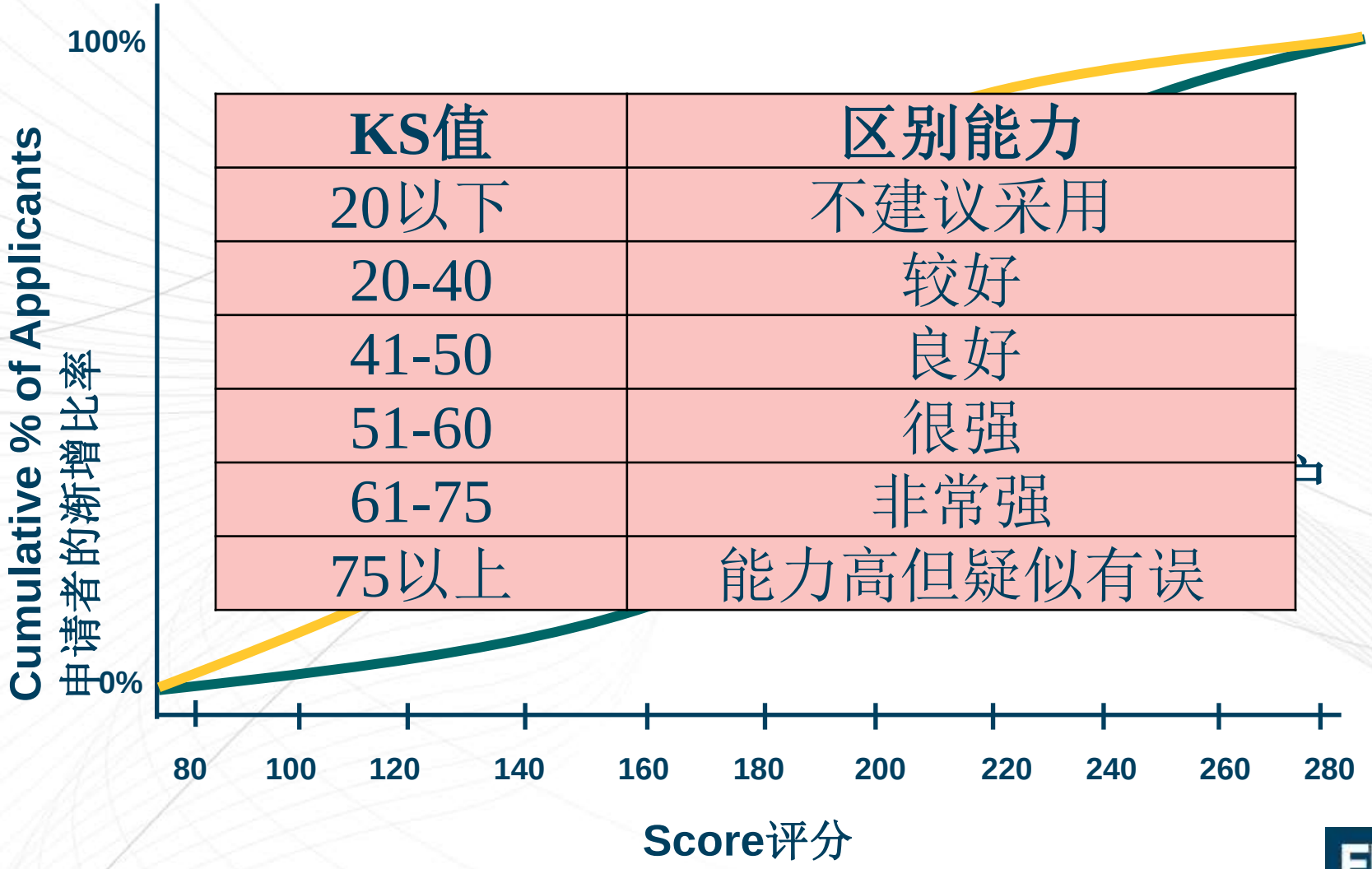




© 2013 Fair Isaac Corporation. Confidential.

评分模型的开发－模型的衡量和评估

计算K-S统计值



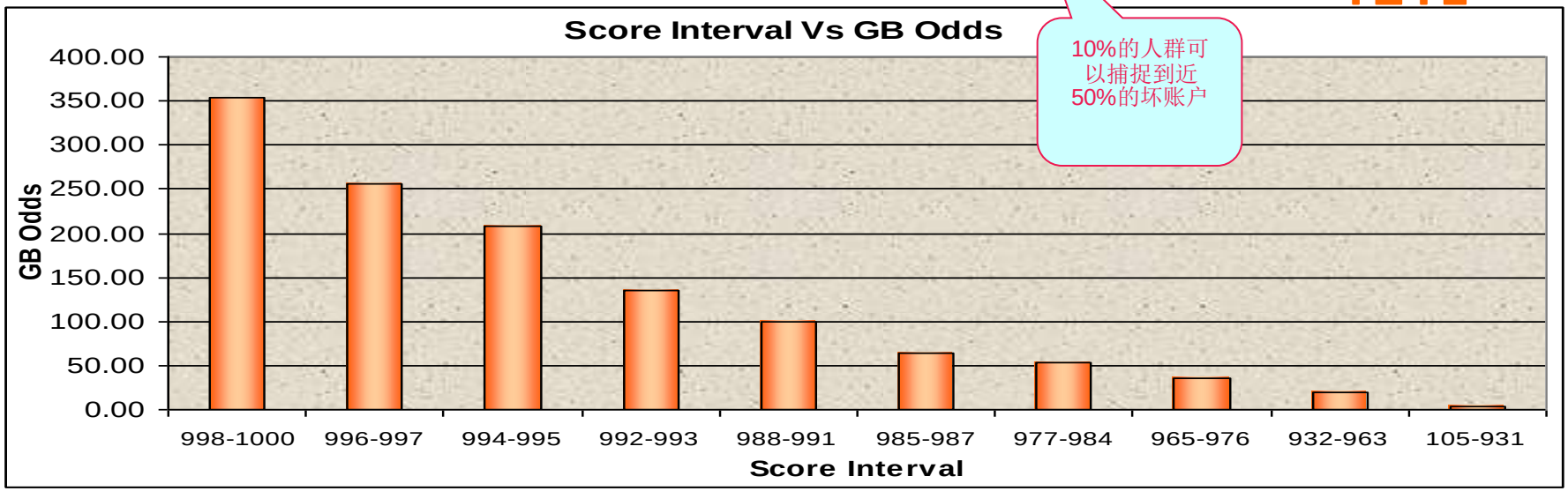
KS值	区别能力
20以下	不建议采用
20-40	较好
41-50	良好
51-60	很强
61-75	非常强
75以上	能力高但疑似有误

© 2013 Fair Isaac Corporation. Confidential.

模型的预测和排序能力 举例

- KS 值　- Gini系数　- 分离度(Divergency)　- 增益图

SCORE INTERVAL	TOTAL RECORDS		PERFORMANCE			CUMULATIVE			Decum	BAD RATE		ODDS	
	Total	Cumulative	Good	Ind	Bad	Good	Ind	Bad	Bad	Interval	Cumulative	Interval	Cumulative
998-1000	12,293	5.8%	11,660	600	33	7.0%	1.4%	0.8%	100.0%	0.27%	0.27%	353.33	353.33
996-997	21,374	15.8%	19,910	1,386	78	18.9%	4.7%	2.5%	99.2%	0.36%	0.33%	255.26	284.41
994-995	31,602	30.5%	28,713	2,751	138	36.1%	11.2%	5.7%	97.5%	0.44%	0.38%	208.07	242.10
992-993	23,525	41.5%	20,890	2,480	155	48.6%	17.0%	9.2%	94.3%	0.66%	0.45%	134.77	200.92
988-991	23,533	52.6%	20,046	3,287	200	60.7%	24.7%	13.7%	90.8%	0.85%	0.54%	100.23	167.58
985-987	16,449	60.2%	13,079	3,165	205	68.5%	32.2%	18.4%	86.3%	1.25%	0.63%	63.80	141.28
977-984	22,871	70.9%	17,668	4,875	328	79.1%	43.7%	25.9%	81.6%	1.43%	0.75%	53.87	116.07
965-976	20,708	80.6%	14,393	5,900	415	87.7%	57.5%	35.3%	74.1%	2.00%	0.90%	34.68	94.30
932-963	20,306	90.1%	12,536	7,100	670	95.2%	74.3%	50.6%	64.7%	3.30%	1.15%	18.71	71.51
105-931	21,091	100.0%	7,979	10,939	2,173	100.0%	100.0%	100.0%	**49.4%**	10.30%	2.06%	3.67	37.97
TOTALS	213,752	**100.0%**	166,874	42,483	4,395					**2.1%**		**37.97**	



© 2013 Fair Isaac Corporation. Confidential.

评分模型的验证

为什么要进行验证？

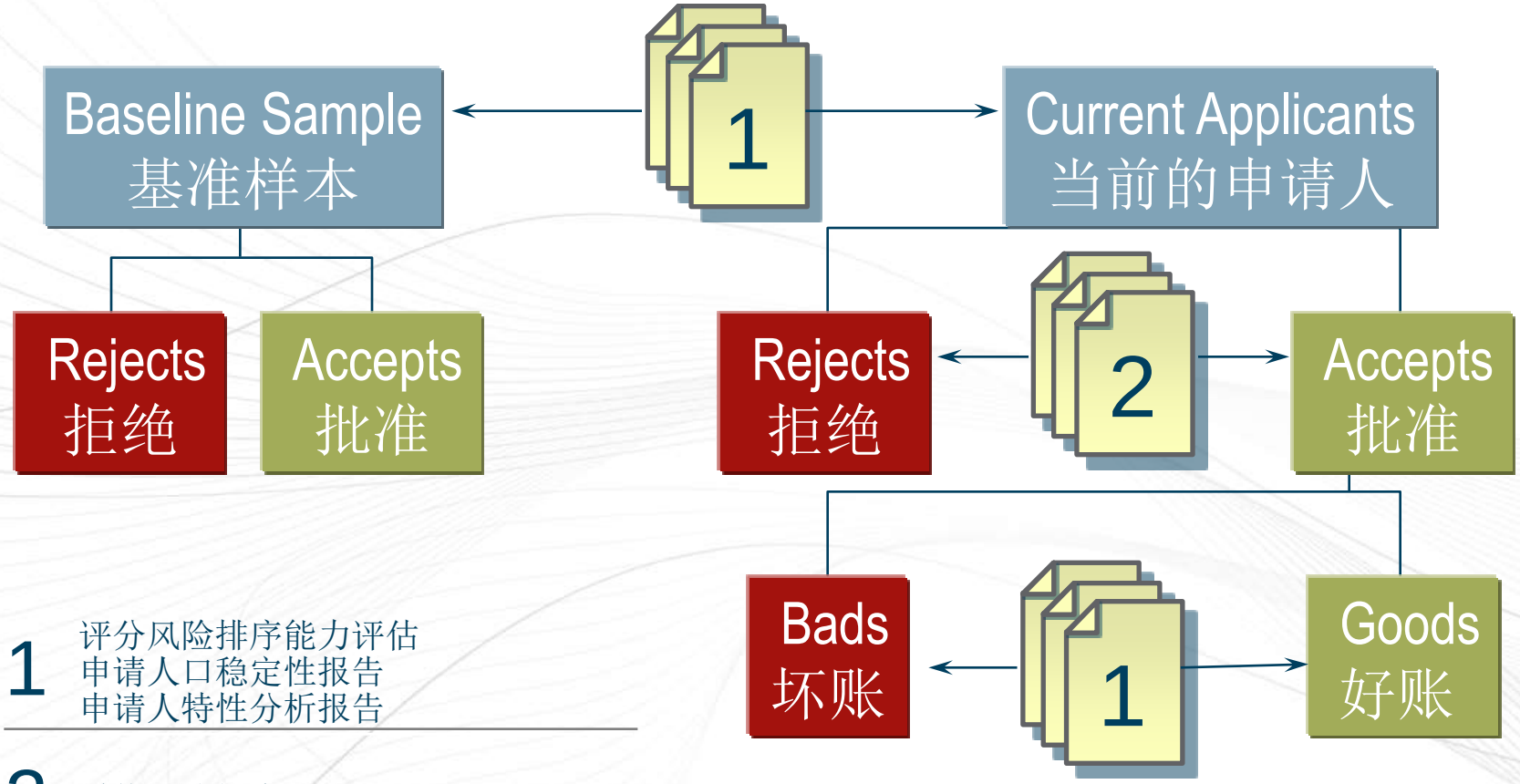
- » 确认无过拟合 — 建模稳健性
- » 识别模型针对更近期群体有效

验证的类型

- » 预留样本校验
  - » 开发样本中随机留出的一部分
- » 跨时间校验
  - » 使用来自不同时期的群体测试模型的稳健性

© 2013 Fair Isaac Corporation. Confidential.

评分卡跟踪报告

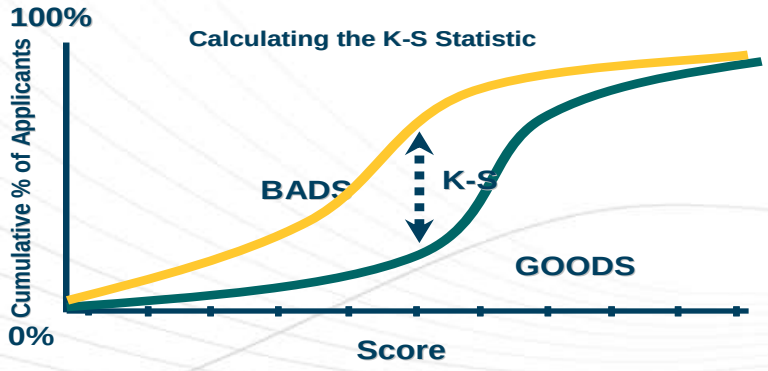


1 评分风险排序能力评估
申请人口稳定性报告
申请人特性分析报告

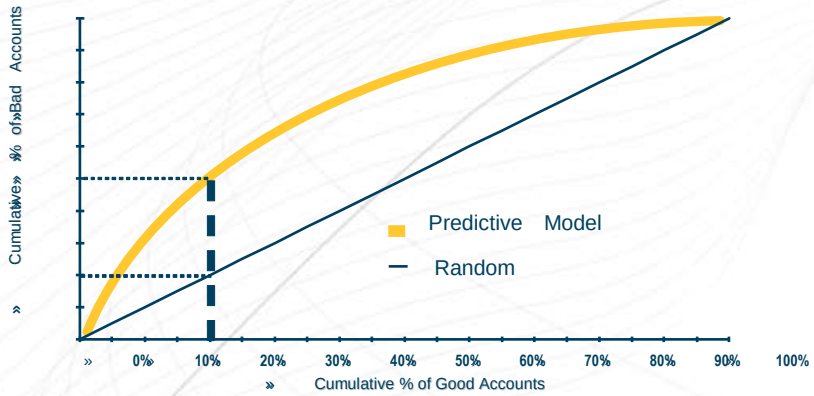
2 最终评分报告

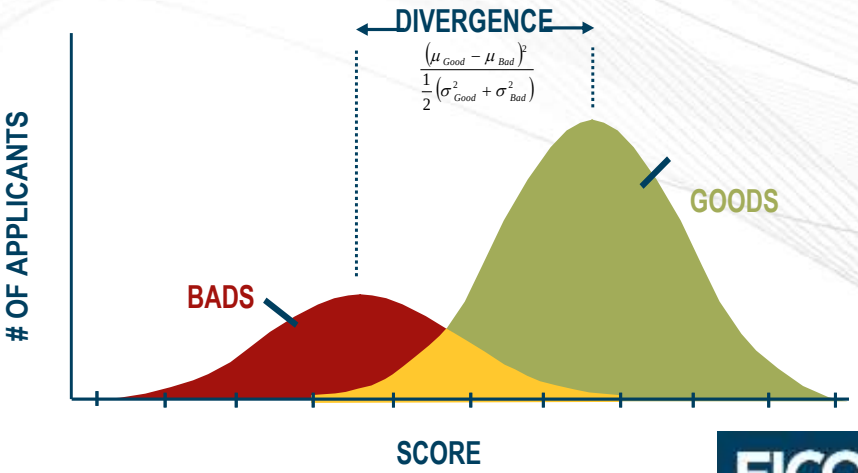
© 2013 Fair Isaac Corporation. Confidential.

评分卡风险排序能力评估



KS值	区别能力
20以下	不建议采用
20-40	较好
41-50	良好
51-60	很强
61-75	非常强
75以上	能力高但疑似有误





© 2013 Fair Isaac Corporation. Confidential.

申请人口稳定性报告

无担保的个人贷款产品组合

申请的时间：2004年4月至6月

SCORE INTERVAL 评分区间	BASELINE 基线 #	BASELINE 基线 %	CURRENT 当前人口 #	CURRENT 当前人口 %	SHIFT 变动率 %
129 – 179	31	10.3	400	20.0	9.7
180 – 189	39	13.0	380	19.0	6.0
190 – 199	48	16.0	369	18.5	2.5
200 – 209	59	19.7	304	15.2	-4.5
210 – 219	49	16.4	205	10.4	-6.0
220 – 229	40	13.3	188	9.4	-3.9
230 – 278	34	11.3	154	7.7	-3.6
	300	100%	2,000	100%	

Population Stability Index: 0.16

申请人口稳定性指数：0.16

© 2013 Fair Isaac Corporation. Confidential.

FICO

稳定性指数标准 PSI

PSI 范围	后续行动
小于0.10	没有变化或者很少变化
0.10 至 0.25	有变化，继续监控后续变化
大于0.25	发生大变化，进行特征项分析

© 2013 Fair Isaac Corporation. Confidential.

申请人特性分析

无担保的个人贷款产品组合

申请的时间：2004年4月至6月

CHARACTERISTIC: Residential Status 特性：居住状况

ATTRIBUTES 属性	BASELINE 基线 #	BASELINE 基线 %	CURRENT 当前申请人 #	CURRENT 当前申请人 %	SHIFT 变动率 %	POINTS 分数	SCORE DIFF. 评分差异
自有/购买	146	48.7	102	5.1	-43.6	41	-17.9
与父母同住	36	12.0	590	29.5	17.5	33	5.8
租房	103	34.3	1102	55.1	20.8	25	5.2
未填	15	5.0	206	10.3	5.3	25	1.3
	300	100%	2,000	100%			
						总体评分差异	-5.6

© 2013 Fair Isaac Corporation. Confidential.

FICO

最终评分报告

- 将批准率显示为评分的功能
- 报告每一个评分区间上的被批准和被拒绝的申请人数
- 衡量当前的操作是否与评分卡、当前的批准率以及潜在的批准率相一致
- 季度出版，年度汇总

© 2013 Fair Isaac Corporation. Confidential.

最终评分报告（续）

SCORE RANGE 评分范围	# APPLICANTS 申请人数量	# ACCEPTED 被接受的申请人数	# REJECTED 被拒绝的申请人数	# OVERRIDES 干预人数	% OVERRIDE 干预率	POTENTIAL ACC RATE 潜在的批准率
170以下	211	6	205	6	3%	100%
170 - 179	422	11	411	11	3%	93%
180 - 189	397	14	383	14	4%	79%
190 -199	213	12	201	12	6%	66%
200 - 209	242	201	41	41	17%	59%
210 - 219	295	255	40	40	14%	51%
220 - 229	337	301	36	36	11%	42%
230 - 239	276	255	21	21	8%	31%
240 - 249	441	423	18	18	4%	21%
250 和250以上	213	207	6	6	3%	7%
低于取舍分数线	1,243	43	1,200	43	3%	
高于取舍分数线	1,804	1,642	162	162	9%	
合计	3,047	1,685	1,362			
批准率		55%				

© 2013 Fair Isaac Corporation. Confidential.

会议内容

一. 评分模型的基本概念

二. 申请评分模型开发方法

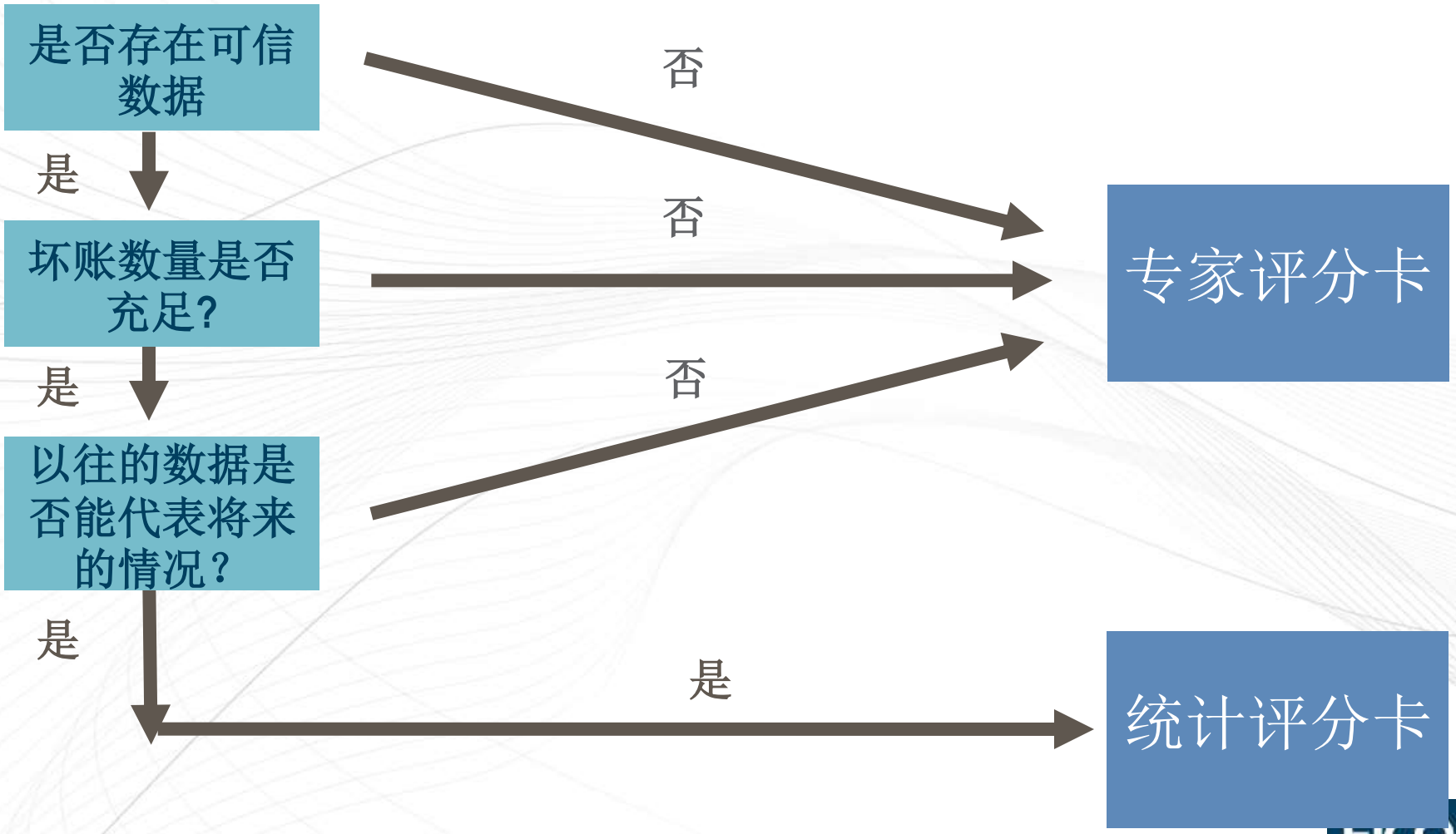
三. 数据驱动的统计模型开发

四. 专家经验模型

五. 混合式模型

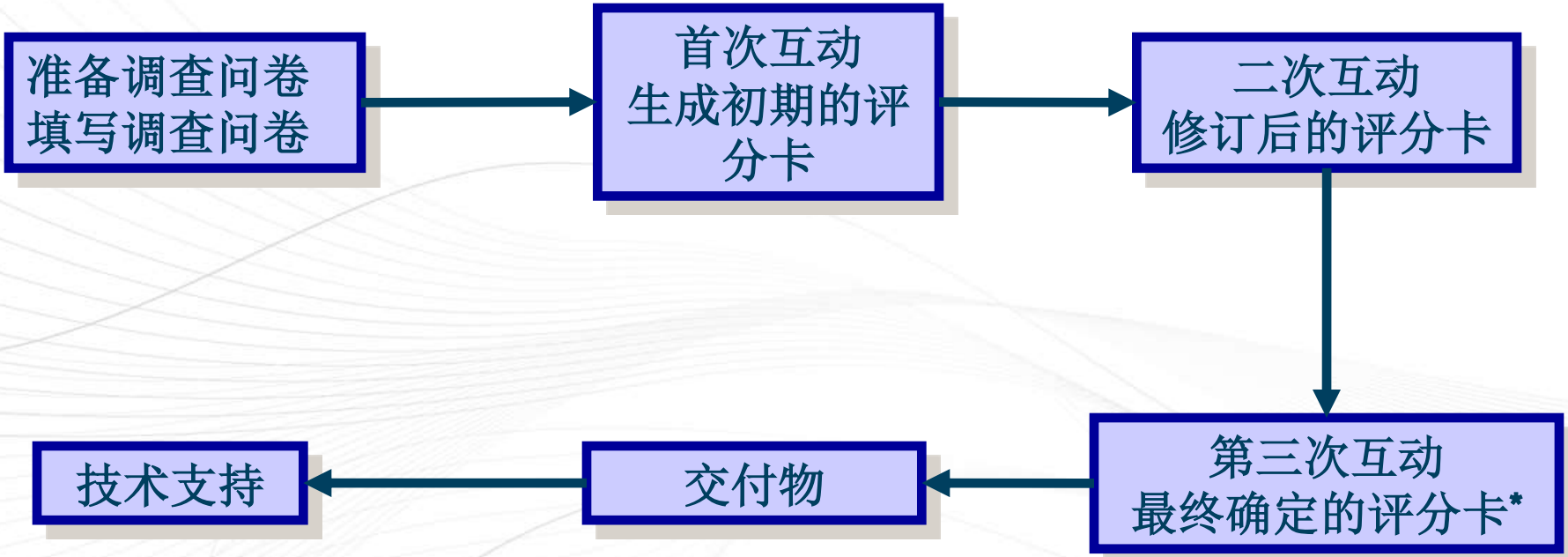
© 2013 Fair Isaac Corporation. Confidential.

专家 vs. 统计评分卡开发



© 2013 Fair Isaac Corporation. Confidential.

专家模型流程



*平均来讲，三次迭代后便可最终生成稳定及可靠的评分卡。根据开发环境的不同，迭代的次数也可能会有所不同。

© 2013 Fair Isaac Corporation. Confidential.

专家模型流程

重要性程度：1-无意义，3-一般，5-很重要

变量序号	变量类别	变量	变量说明	变量备注	意见	重要性排序
1	信贷申请信息	PRODNAME	贷款品种	二手商用房、二手住房、一手商用房、一手住房、保障性住房、车库、房易贷、信易贷、综合消费贷款		
2	信贷申请信息	APPTTERM	申请期限	申请的贷款期限（月）		
3	信贷申请信息	ASSUKIND	担保方式	抵押、质押、法人保证、自然人保证、信用		
4	信贷申请信息	RETUKIND	还款方式	等额本息、等额本金、分期付息一次还本、到期一次还本付息		
5	信贷申请信息	RETUPERI	还款频率	月、季、年、天		
6	信贷申请信息	LOANUSE	借款用途	购房、装修、消费等		
7	信贷申请信息	HOUSEBUILDYEAR	房龄	根据房屋建筑年限计算得到		
8	个人基本信息	SEXSIGN	性别	男、女		
9	个人基本信息	BIRTDATE	年龄	客户年龄，由出生日期计算		
10	个人基本信息	EDUCSIGN	学历	高中及以下，专科，本科，硕士，博士及以上		
11	个人基本信息	MARRSIGN	婚姻状况	已婚、单身 （未婚，已婚，丧偶，离异，其他）		

米勒法则（1956）：人脑同时只能处理7±2个信息团。

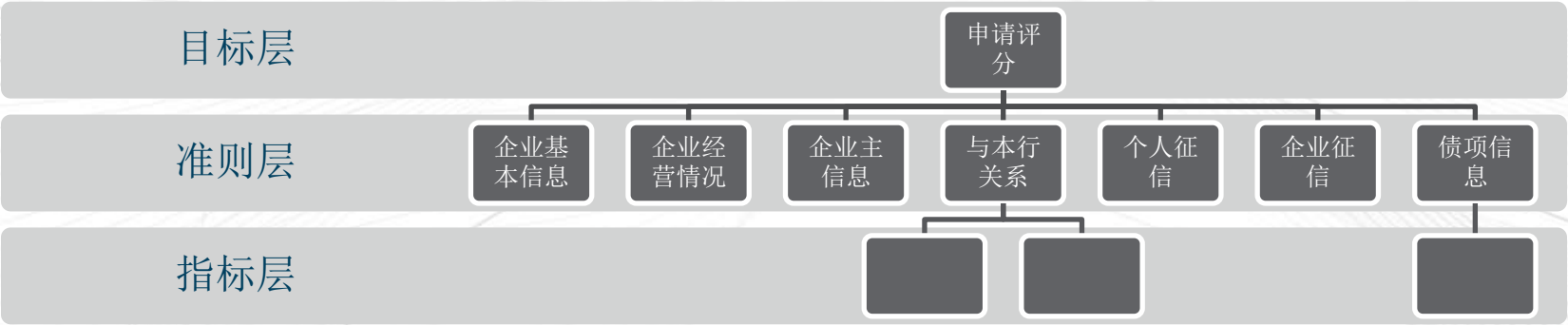
一致性检查：分成6组，检查组间的重要性排序

© 2013 Fair Isaac Corporation. Confidential.

评分卡指标权重设定

» 采用层次分析法（AHP）方法确认指标权重

层次分析法是美国运筹学家匹兹堡大学教授萨蒂（T.L.Saaty）于上世纪70年代初，应用网络体统理论和多目标综合评价方法，提出的一种层次权重决策分析方法。这种方法的特点是在对复杂的决策问题的本质、影响因素及其内在关系等进行深入分析的基础上，利于较少的定量信息使决策的思维过程数学化，从而为多目标、多准则或无结构特性的复杂决策问题提供简便的决策方法。



© 2013 Fair Isaac Corporation. Confidential.

评分卡指标权重设定

层次结构反映了因素之间的关系，但各评价因素在目标衡量中所占的比重并不一定相同。层次分析法主要是对每一层次中各因素相对重要性给出的判断，这些判断通过引入合适的标度用数值表示出来，写成判断矩阵。

C_K	A_1	A_2	……	A_n
A_1	a_{11}	a_{12}	……	a_{1n}
A_2	a_{21}	a_{22}	……	a_{2n}
⋮	⋮	⋮	⋮	⋮
A_n	a_{n1}	a_{n2}	……	a_{nn}

1	表示两个元素相比，具有同样重要性
3	表示两个元素相比，一个元素比另一个元素稍微重要
5	表示两个元素相比，一个元素比另一个元素明显重要
7	表示两个元素相比，一个元素比另一个元素强烈重要
9	表示两个元素相比，一个元素比另一个元素极端重要

© 2013 Fair Isaac Corporation. Confidential.

评分卡指标权重设定

层次单排序是指根据判断矩阵计算对于上一层某元素而言，本层次与之有联系的元素重要性次序的权值。根据线性代数有关理论，正互反矩阵 A 的特征值可作为衡量同一层次中每个因素对上一目标的影响中所占的比重，A 的最大特征值与权向量的“和法”计算公式是：

$$W_i = \frac{1}{n}\sum_{j=1}^{n}\frac{a_{ij}}{\sum_{k=1}^{n} a_{ki}}$$

在特殊情况下，判断矩阵可以具有传递性和一致性。一般情况下，并不要求判断矩阵严格满足这一性质。但从人类认识规律看，一个正确的判断矩阵重要性排序是有一定逻辑规律的，例如若 A 比 B 重要，B 又比 C 重要，则从逻辑上讲，A 应该比 C 明显重要，若两两比较时出现 A 比 C 重要的结果，则该判断矩阵违反了一致性准则，在逻辑上是不合理的。因此在实际中要求判断矩阵满足大体上的一致性，需进行一致性检验。只有通过检验，才能说明判断矩阵在逻辑上是合理的才能继续对结果进行分析。一致性检验的步骤如下：

1）计算一致性指标 CI（consistency index）：

$$C.I. = \frac{\lambda_{\max} - n}{n-1}$$

2）查表确定相应的平均随机一致性指标 RI（random index）。萨蒂给出了 RI 的值（表 6-2）。

表 6- 2 平均随机一致性指标表

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.12	1.24	1.32	1.41	1.45

3）计算一致性比例 CR（consistency ratio）：

$$C.R. = \frac{C.I.}{R.I.}$$

4）判断一致性。当 CR < 0.1 时，认为判断矩阵的一致性是可以接受的，否则需要对该判断矩阵进行重新修正。

© 2013 Fair Isaac Corporation. Confidential.

层次分析法示例

» 判断矩阵

	家庭资产负债率	房屋状况	企业主本行业从业年限	户籍	婚姻状况
家庭资产负债率	1	2	2	9	9
房屋状况	1/2	1	1	6	8
企业主本行业从业年限	1/2	1	1	5	7
户籍	1/9	1/6	1/5	1	1
婚姻状况	1/9	1/8	1/7	1	1

» 一致性检查

λmax	5.0350
CI	0.0087
RI	0.12
CR	0.0728
< 0.1?	是

» 权重

指标	权重
家庭资产负债率	43%
房屋状况	25%
企业主本行业从业年限	24%
户籍	4%
婚姻状况	4%

© 2013 Fair Isaac Corporation. Confidential.

会议内容

一. 评分模型的基本概念

二. 申请评分模型开发方法

三. 数据驱动的统计模型开发

四. 专家经验模型

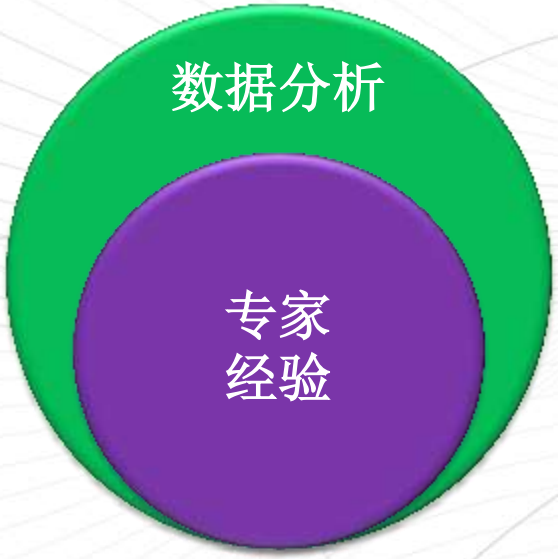
五. 混合式模型

© 2013 Fair Isaac Corporation. Confidential.

评分模型的开发 - “混合式”方法开发模型

© 2013 Fair Isaac Corporation. Confidential.

评分模型的开发 - “混合式”方法开发模型



© 2013 Fair Isaac Corporation. Confidential.

评分模型的开发 - “混合式”方法开发模型

- 数据无错误
  - 缺失比例高
    - 数据用于分析关键变量的分布，为模型变量binning做参考
  - 缺失比例尚可
    - 坏样本太少
      - 数据用于分析关键变量的分布，为模型变量binning做参考
      - 数据用于分析关键变量的WOE风险趋势，为变量Binning及评分分值做参考
    - 坏样本接近满足要求
      - 基于数据产生初始数据驱动模型，作为专家经验调整的基础

© 2013 Fair Isaac Corporation. Confidential.

评分模型的开发 - “混合式”方法开发模型

变量分箱	分值		参考分值		
	校准	未校准	最小	最大	中性
截距项					
	621	0			
教育水平（Education）			-21	26	0
硕士研究生	26	1.021			
大学本科	13	0.387			
专科	-7	-0.219			
高中	-19	-0.628			
初中、小学	-21	-0.808			
年龄（Age）			-31	18	0
(-, 25)	-31	-0.680			
[25, 30)	-23	-0.561			
[30, 35)	2	0.143			
[35, 40)	10	0.312			
[40, 45)	18	0.373			
[45, 50)	6	0.257			
[50, +)	1	0.038			

© 2013 Fair Isaac Corporation. Confidential.

谢谢！

Confidential. This presentation is provided for the recipient only and cannot be reproduced or shared without Fair Isaac Corporation's express consent.

© 2011 Fair Isaac Corporation.