

Contents

I. CSS(Credit Scoring System) 일반

1. CSS모형 분류
2. 모형개발 방법론

II. 신용카드 심사전략

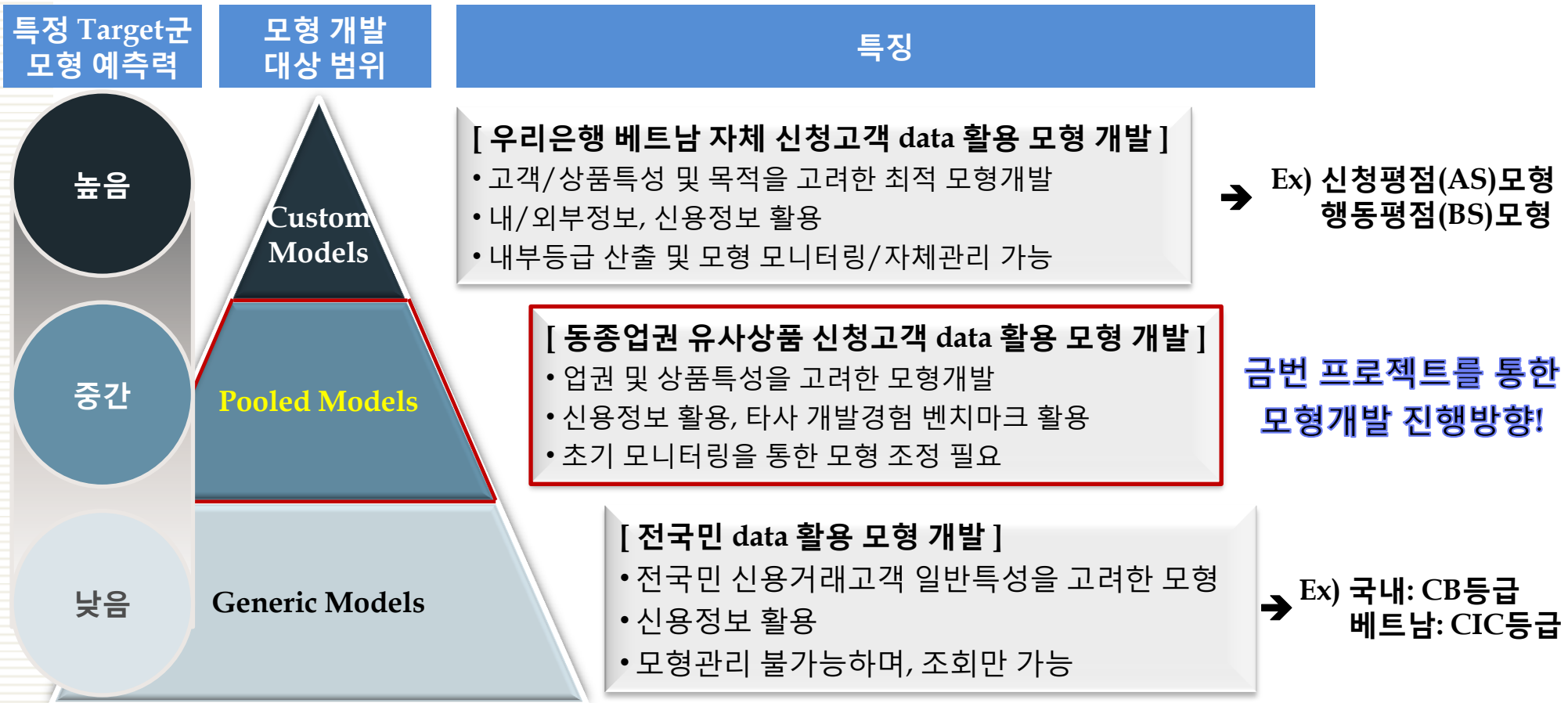
III. 여신 심사전략

I. CSS(Credit Scoring System) 일반

CSS모형 분류

CSS모형 분류 기준 - ① Target군 기준

○ 우리은행 베트남 신용평가모형 부재 → 동종업계의 유사상품 신청 정보를 활용하여 Pooled Data를 활용한 모형 개발



CSS모형 분류

CSS모형 분류 기준 - ② 목적 기준

- 목적에 따라 신청평점모형(Application Scoring, 이하 AS모형)과 행동평점(Behavior Scoring, 이하 BS모형)으로 구분됨
- 본 프로젝트의 개발모형은 AS모형임

구분	AS모형	BS모형
모형 정의	<ul style="list-style-type: none"> ▪ 고객의 대출(혹은 카드) 신청시점의 리스크 수준을 평가하여 불량 고객으로 전이될 가능성 예측 	<ul style="list-style-type: none"> ▪ 기존 고객에 대해서 매월 주기적으로 리스크 수준을 평가하여 불량 고객으로 전이될 가능성 예측
모형 가동 시점	<ul style="list-style-type: none"> ▪ 대출(혹은 카드) 신청시점(Online) 	<ul style="list-style-type: none"> ▪ 1회/月(Batch)
업무 활용 전략	<ul style="list-style-type: none"> ▪ 각종 심사전략 (Ex. 승인, 한도전략 등) 	<ul style="list-style-type: none"> ▪ 기존 고객 관리 전략 (Ex. 여신-기일관리전략 / 카드-한도상하향 전략)

모형개발방법론

모형개발 Process

○ 모형개발 Process는 개발 모집단 정의부터 최종 모형검증까지 총 10단계로 이루어짐

1	개발 모집단 정의	<ul style="list-style-type: none"> 모집단의 대표성 : 개발 데이터 최신성 확보 데이터 활용 가능기간 : 내부 및 외부 정보 변동 고려 	모형개발을 위한 요건 정의
2	Target(반응변수) 정의	<ul style="list-style-type: none"> vintage 분석 등을 통한 Bad 발생 증가율이 둔화되는 시점을 예측기간으로 정의. Bad/Good 구성비 및 수준 고려 	
3	모형Segment 구성	<ul style="list-style-type: none"> 세부모형은 전략적 목적에 부합해야 하며, 통계적 유의성이 있는 Segment로 분리 	
4	데이터 수집/추출	<ul style="list-style-type: none"> 원천별로 데이터 추출 및 검증 데이터 가공/Cleansing 후 후보항목 생성 	모형개발
5	항목 유의성 분석	<ul style="list-style-type: none"> 우불량 예측력 위주 후보항목 선택 Event율 및 Odds 기준 항목 범주화(Classing) 	
6	모형 적합	<ul style="list-style-type: none"> Logistic Regression 모형 적합 통해 최종 변수 선택 및 다중공선성 검정 선택된 최종 변수는 업무적 적용 타당성 확보 	
7	평점표 구성	<ul style="list-style-type: none"> 추정된 회귀계수(Estimate)를 업무에 활용할 수 있도록 배점 형태로 전환하여 평점표 구성 배점/평점은 PDO(Point to Double Odds) 체계로 구성 	
8	평점 통합	<ul style="list-style-type: none"> 세부모형 별 평점을 하나의 평점으로 통합 Event율 및 Odds 기준으로 통합평점 구성 	
9	등급화	<ul style="list-style-type: none"> 업무 활용성 및 PD Segment 일관성을 위해 현행과 동일한 등급체계로 구성 사전부도 회원은 별도 부도등급으로 구성 	
10	모형검증	<ul style="list-style-type: none"> 개발모형 안정성 측면에서 검증(PSI 지수 등) 개발모형 변별력 측면에서 검증(AR지수, KS통계량 등) 	모형검증

모형개발방법론

Target(반응변수) 정의

- Target(반응변수) 정의란 모형 개발 시 종속변수(예측하고자 하는 변수)에 대한 정의로 우량/판단미정/불량으로 구분함
- Target(반응변수) 정의를 위해서는 성과측정기간(Performance Period) 정의가 우선 결정되어야 하며, 불량성향이 가장 강한 요건부터 불량 → 판단미정 → 우량 순서로 정의함

■ 리스크 수준별 우불량 정의

우선순위	구분	우불량 정의
1	내부 부도 발생	불량
2	내부 연체 90일 이상 경험	불량
3	채무불이행 발생	불량
4	CB연체 90일 이상 경험	불량
5	60일<= 내부연체 <90일	판단미정
6	60일<= CB연체 <90일	판단미정
7	30일<= 내부연체 <60일	판단미정
8	30일<= CB연체 <60일	판단미정
9	15일<= 내부연체 <30일	우량
10	15일<= CB연체 <30일	우량
11	0일<= 내부연체 <15일	우량
12	0일< CB연체 <15일	우량
13	무연체	우량

불량

우량

■ 판단미정이란?

판단미정은 우량 또는 불량으로 판단하기 애매한 영역으로 모형개발 시 제외됨. 이를 통해, 모형 예측력 향상됨

단, 판단미정의 구성비가 너무 많을 경우 모집단의 대표성이 감소하므로 최대 10%를 넘지 않는 범위에서 정의함

모형개발방법론

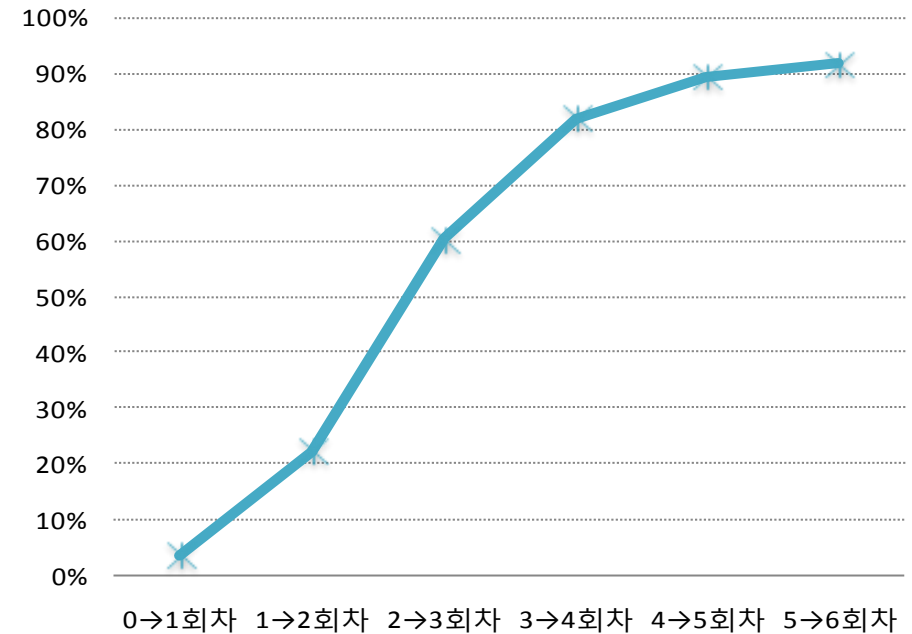
Target(반응변수) 정의 - Roll Rate 분석

○ Roll-rate분석은 M0시점 연체상태(무연체, 1회차연체, 2회차연체..)가 M1시점(익월) 어떻게 전이되는지 관찰하는 분석으로 익월 연체 정도가 한 단계 더 진행될 확률이 크다면 해당 연체 정도 이상을 불량으로 정의하는 것이 타당함

M0 연체회차	M1 연체회차						
	0	1	2	3	4	5	6
0	96.32%	3.68%	0.00%	0.00%	0.00%	0.00%	0.00%
1	43.65%	34.19%	22.16%	0.00%	0.00%	0.00%	0.00%
2	20.62%	10.43%	8.67%	60.29%	0.00%	0.00%	0.00%
3	8.73%	3.16%	1.87%	4.39%	81.86%	0.00%	0.00%
4	4.78%	1.40%	0.70%	1.14%	2.64%	89.34%	0.00%
5	4.60%	0.92%	0.26%	0.52%	0.54%	1.49%	91.67%
6	4.73%	0.50%	0.26%	0.31%	0.40%	0.60%	93.20%

0→1회차	1→2회차	2→3회차	3→4회차	4→5회차	5→6회차
3.7%	22.2%	60.3%	81.9%	89.3%	91.7%

Roll-rate Curve

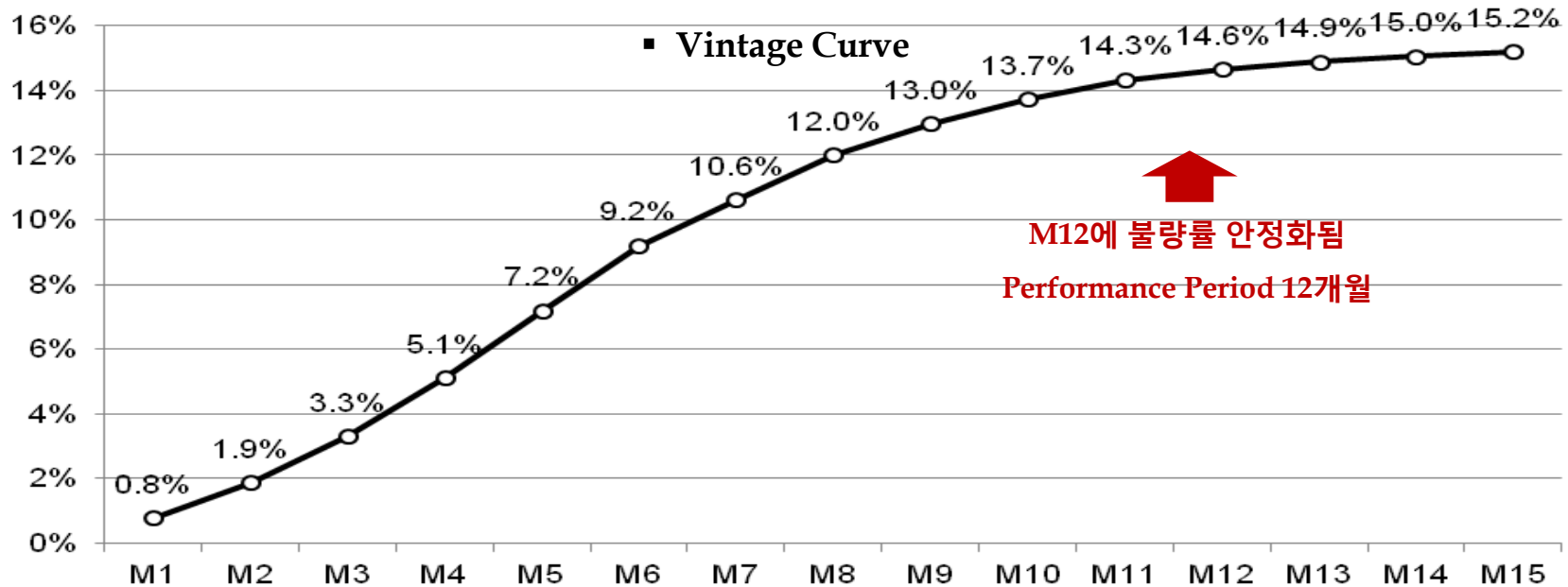


모형개발방법론

예측기간(Performance Period) 정의 - Vintage분석

- 예측기간 정의는 Vintage 분석을 수행하여 불량률 증가속도가 둔화되어 안정화되는 시점의 정의함
- 그러나 실제로는 Vintage Curve가 지속적으로 우상향하여 안정화되는 시점을 찾기 어렵고,
Basel II 의 영향(부도 정의와의 일관성)으로 인해 성과측정기간을 12개월로 고정하는 추세임

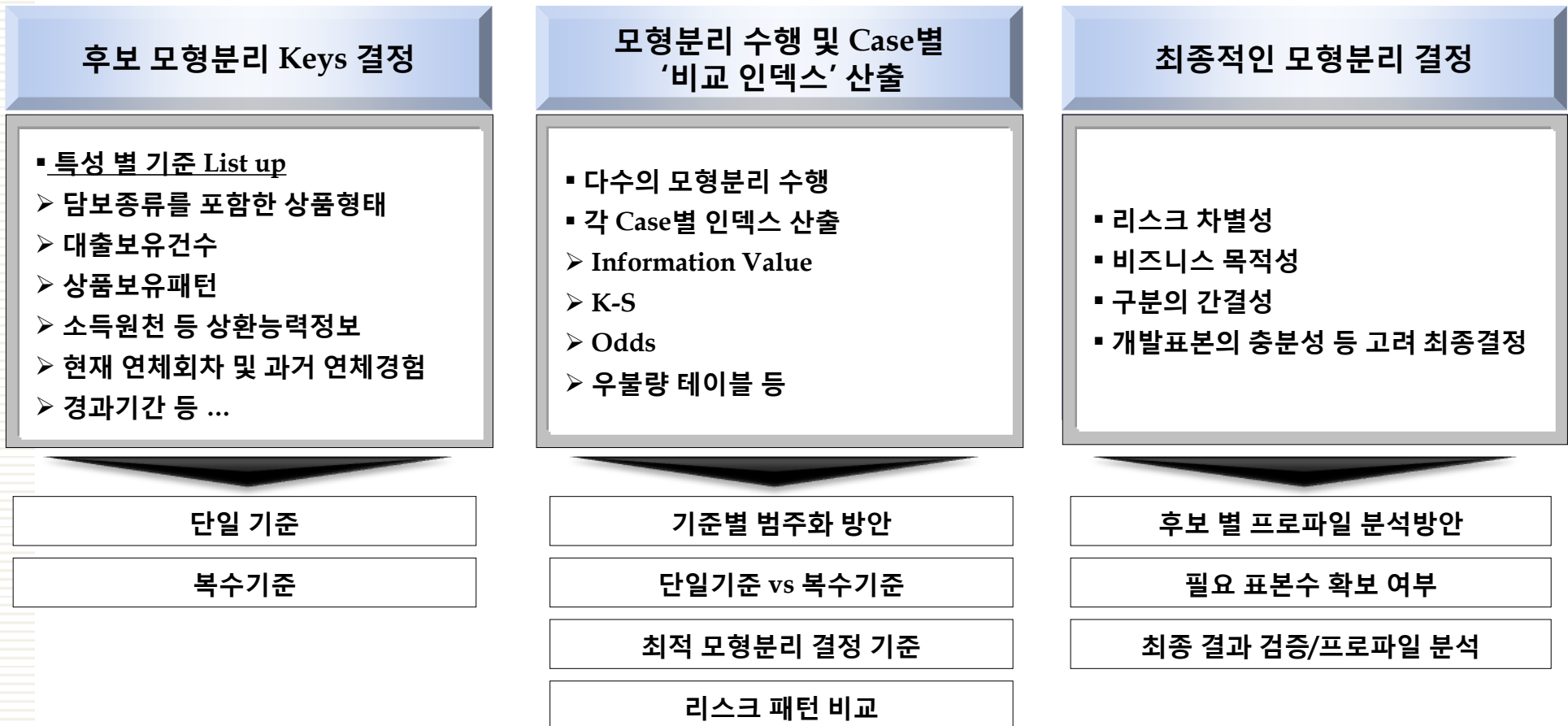
전체 신청건수	경과개월별 누적불량건수														
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
253,242	1,994	4,739	8,365	12,954	18,188	23,294	26,904	30,345	32,843	34,780	36,281	37,088	37,682	38,106	38,494



모형개발방법론

모형 세분화(Segmentation) 정의

- 신용평가모형은 신용평점의 예측력을 향상시키고, 전략 적용 시 신용평점 모형의 활용이 용이하도록 분리
- 전략적 관점, 운용적 관점, 리스크적 관점을 모두 고려함



모형개발방법론

데이터 수집/추출

○ 정보영역별 취합 가능한 모든 데이터 수집 및 추출 → 모형 개발을 위한 분석항목리스트 생성

정보 구분	내 용	분석항목 리스트
1.1 고객정보	▪ 성별, 연령, 거주지 등 신청서 정보	▪ 나이, 성별, 현직장근무기간, 종합소득세 등
1.2 여신실적	▪ 대출실행일/상환내역/여신잔액 등	▪ 월말잔액, 최근6개월한도대출잔액평균 등
1.3 여신연체	▪ 여신연체, 특수채권 편입 및 상각 등	▪ 최근1개월최장연체일수 등
2.1 수신실적	▪ 상품(은행, 신탁, 공제)/평잔 등	▪ 최근3개월총수신평잔 등
3.1 카드실적	▪ 당행 카드 한도, 이용금액, 연체금액 등	▪ 최근6개월CA이용건수, 최근1개월한도소진율 등
4.1 부수거래	▪ CIF신규일, 급여이체횟수 등	▪ 최근3개월급여이체금액 등
5.1 신용정보	▪ 신용불량, 신용개설, 대출, 연체 등	▪ 최근1년내조회총건수, 미해제연체총금액 등
항목생성 기본원칙 1	내/외부 정보 Data Source별 추출 및 검증/정제 후 분석항목 리스트 작성	
항목생성 기본원칙 2	변수의 가/감/승/제를 통한 변수 변환 - 건수, 금액, 기간	
항목생성 기본원칙 3	2개 이상 변수를 결합하여 합성항목 생성	

모형개발방법론

항목 유의성 분석 - ① 변수 범주화

○ 정보영역별 취합 가능한 모든 데이터 수집 및 추출 → 모형 개발을 위한 분석항목리스트 생성

▪ 변수 범주화란?

일종의 변수 Transformation 작업으로 동질적 성격의 값들끼리는 묶고,
이질적 성격의 값들끼리는 나누는 작업을 의미함

(예1 : 연체일수 0, 연체일수 0~10일, 연체일수 11일~30일, 연체일수 30일 ~)

(예2 : 공무원 및 군인, 우수 고객군, 전문직 근무자, 자영업자, 기타)

▪ 변수 범주화의 목적

- 변수의 민감성을 줄이기 위함
- 변수를 설명하기가 쉬움
- 결측값 및 이상값 등을 처리하기 쉬움

▪ 변수 범주화 Process

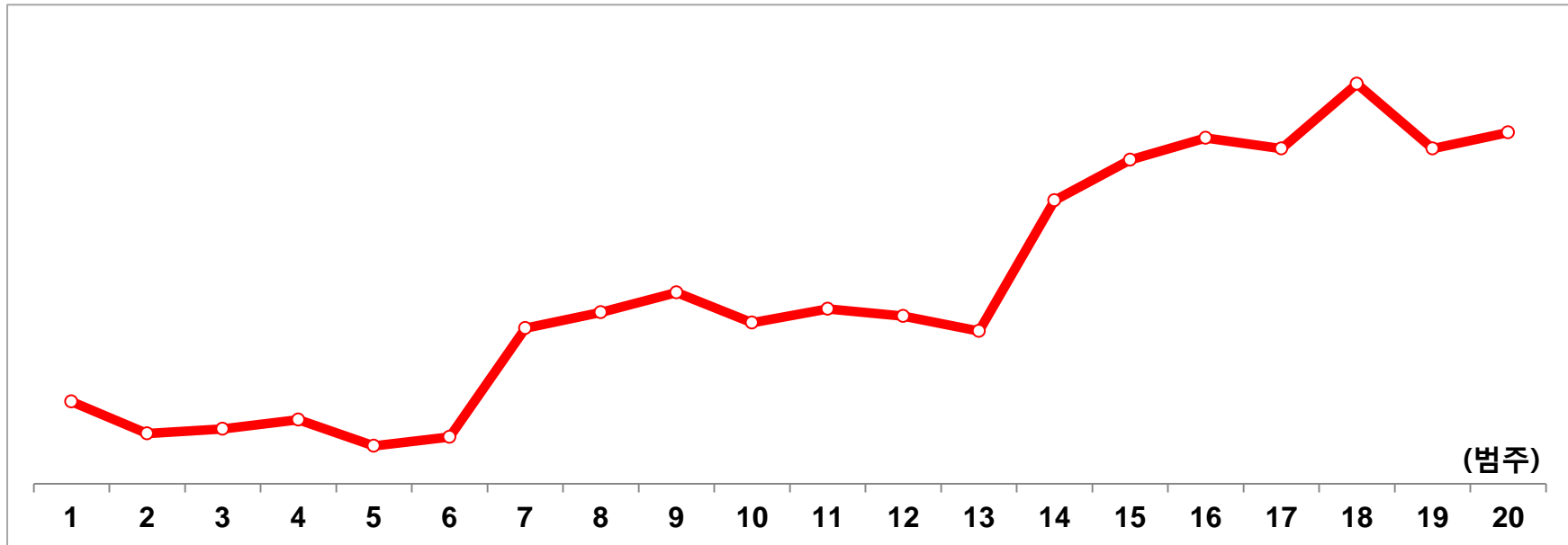
- Fine Classing → Coarse Classing → Recording

모형개발방법론

항목 유의성 분석 - ① Fine Classing

- Fine Classing 이란 분석항목을 구성비 5% 기준으로 최대 20개 구간으로 자동 범주화 하는 작업을 의미함
- 이 과정을 통해 우불량 판별력과 패턴을 찾아 1차적으로 분석항목을 선택하는 과정임

▪ Fine Classing 수행 후 구간별 불량률 추이



모형개발방법론

항목 유의성 분석 - ② Fine Classing을 통한 1차 유의항목 선정을 위한 지표 및 기준

○ 1차 항목 유의성 분석은 단변량 변수의 유의성은 변수의 변별력, 안정성, 불량률 서열화, 구간별 구성비, 비즈니스 목적 부합성을 종합적으로 고려하여 판단함

➔ 1차 항목 유의성 분석을 위한 주요 항목 및 판단 기준

1. 변별력 지표 : Information Value(I.V.) > 0.1 , K-S > 0.1 확인

IV = Score 구간/등급별 (% of 우량 - % of 불량) x ln(% of 우량 / % of 불량)의 합계

K-S = 항목 Classing별 | % of 누적우량 - % of 누적불량 | 의 최대값

2. 안정성 지표 : PSI(Population Stability Index(P.S.I.) < 0.1 확인

P.S.I. = Score 구간/등급별 (% of 개발 - % of 검증) x ln(% of 개발 / % of 검증)의 합계

3. 불량률 서열화 : 불량률 및 Odds의 서열화 유지여부 확인

불량률 = 불량건수/전체건수, Odds = 우량건수/불량건수

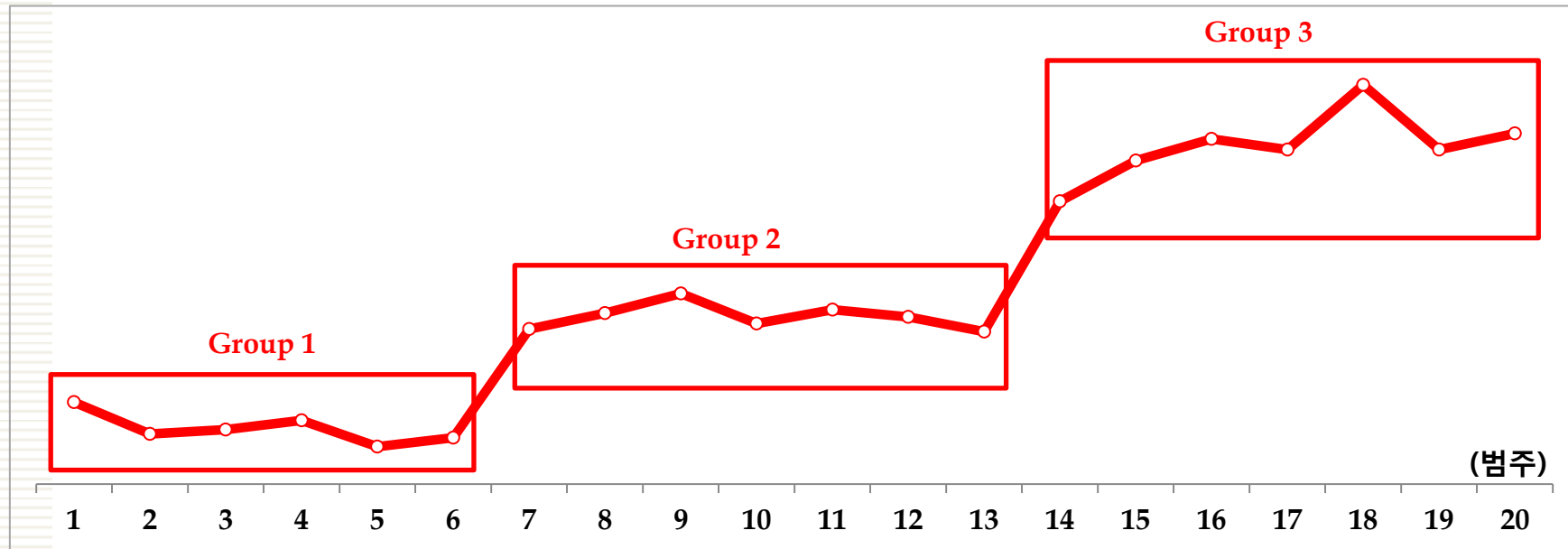
4. 구성비 : 구간별 최소 구성비 > 3% 확인

모형개발방법론

항목 유의성 분석 - ③ Coarse Classing

- Fine Classing 결과에 대하여 동질적인 불량률 패턴의 구간을 그룹화하는 작업을 의미함
- 단변량 변수의 유의성 분석 방법론에 따라 변별력, 안정성, 불량률 서열화, 구간별 구성비, 비즈니스 목적 부합성을 종합적으로 고려하여 Coarse Classing을 수행함

Coarse Classing 수행 후 구간별 불량률 추이



모형개발방법론

항목 유의성 분석 - ④ Recoding

- Coarse Classing 후 구간의 불량률 (또는 Odds)의 크기 순서에 따라 정수 값을 부여
- 이렇게 부여된 정수 값은 Logistic Regression 모형 적합 시 Dummy Variable로 사용됨

Recording 과정

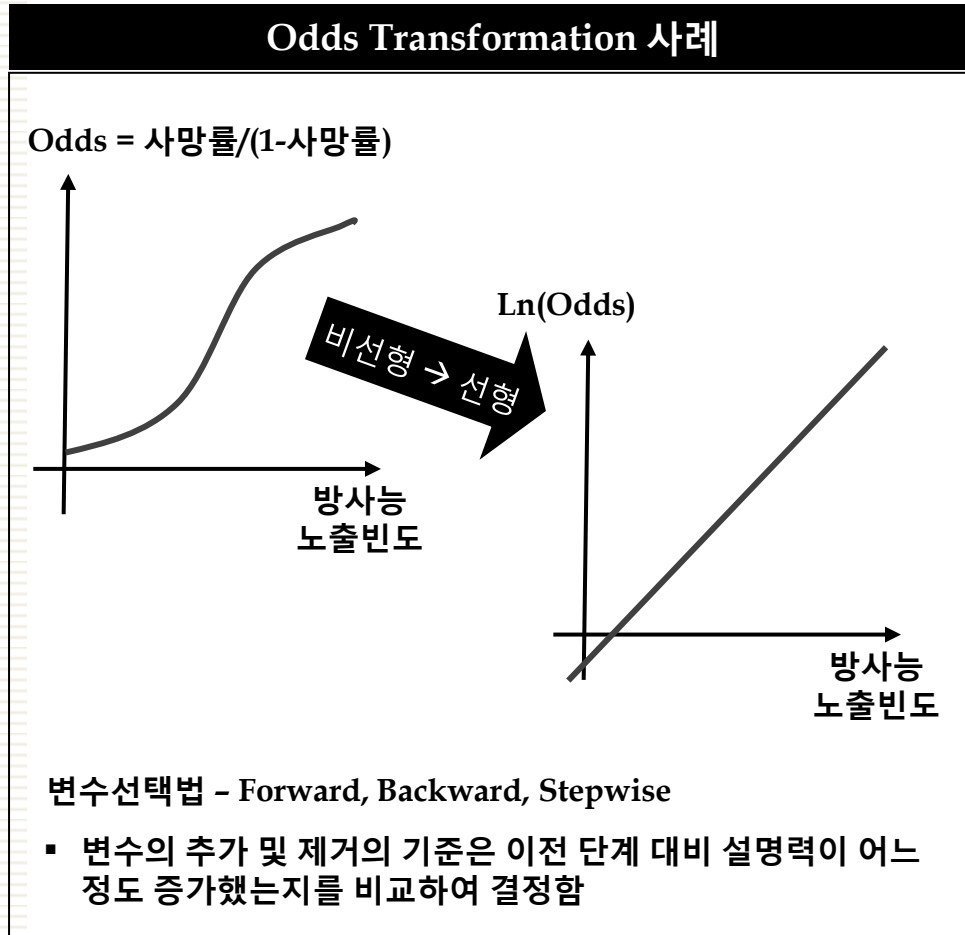
Coarse Classing	Recoding
Group 1	2
Group 2	1
Group 3	0

Dummy Variable 변환

	Dummy 1	Dummy 2
Recode = 0	0	0
Recode = 1	1	0
Recode = 2	0	1

모형개발방법론

모형적합 - ① Logistic Regression의 이해



1 반응변수가 연속형이 아닌 이항형인 경우

- $Odds = p / (1-p)$
 - ✓ 사망률 = 0.1일 때 $Odds = 0.1 / 0.9$
 - ✓ 사망률 = 0.8일 때 $Odds = 0.8 / 0.2$

2 Odds Transformation

- $Odds \rightarrow \ln(Odds)$ 로 변환
 - ✓ 설명변수와 종속변수의 관계가 비선형 → 선형으로 변환
 - ✓ Odds의 범위 = $(0, \infty) \rightarrow \ln(Odds)$ 의 범위 = $(-\infty, \infty)$

3 회귀적합식

- $\ln(Odds) = B_0 + B_1 * \text{혈압(Blood Pressure)} + e$

4 모형의 적합성 검정 - $-2\log L$, AIC Schwartz Criterion

- $MSR(=SSR/1) / MSE(=SSE/(n-2))$

5 회귀계수 유의성 검정 - Wald의 Chi-square 검정

- 귀무가설 : $B_1 = 0$ vs 대립가설 : $B_1 \neq 0$

모형개발방법론

모형적합 - ② 변수 선택 방법

단계	주요 내용	고려 사항
1 1차 변수 선택	<ul style="list-style-type: none"> Fine Classing 수행 후 단변량 변수의 변별력, 안정성, 불량을 서열화, 구간별 구성비, 비즈니스 목적 부합성을 종합적으로 고려하여 판단함 	1차 항목 유의성 분석
2 2차 변수 선택	<ul style="list-style-type: none"> 1차 선택된 변수에 대해 Coarse Classing 수행 이후 정보 성격이 유사한 특성항목 Pool별로 Partial Logistic Regression을 이용하여 유의한 변수 선택(Stepwise Selection 이용하여 P-Value < 0.05 인 경우 선택) 	정보 영역별 상관성 고려
3 3차 변수 선택	<ul style="list-style-type: none"> 2차 선택된 변수 중 추정치 부호가 음(-)의 값인 경우 변수간의 다중공선성이 의심되며, 이 경우 범주 재조정 후 특성항목 Pool별로 Partial Logistic Regression을 이용하여 정보 영역별 최종 변수 선택함 	
4 4차 변수 선택	<ul style="list-style-type: none"> 정보 영역별로 선택된 변수를 종합하여 전체 영역에 대하여 Logistic Regression을 이용하여 유의한 변수 선택(Stepwise Selection 이용하여 P-Value < 0.05 인 경우 선택) 	전체 영역 상관성 고려
5 최종 변수 선택 (모형적합)	<ul style="list-style-type: none"> 추정치 부호가 음(-)의 값인 범주 재조정하거나 삭제하며, 비즈니스적인 의미를 고려하여 최종 변수 선택을 완료함 	

모형개발방법론

모형적합 - ③ Scorecard의 이해

- Logistic Regression의 결과값 해석(SAS OUTPUT)
- Logistic Regression의 Parameter Estimate 값은 PDO 개념을 통해 Score로 변환됨
(즉, Logistic Regression의 결과값을 기반으로 항목별 배점(Score) 및 최종 모형(Scorecard)를 산출함)

➤ 회귀분석 결과 해석

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.7185	0.1128	40.6013	<.0001
연체일수 1	1	1.4159	0.1601	78.2328	<.0001
연체일수 2	1	2.7865	0.2164	165.7368	<.0001

➤ 회귀분석 결과 적용

$$\ln(\text{Odds}) = -0.72 + 1.42 * \text{연체일수1} + 2.79 * \text{연체일수2} + e$$

Recode = 0

$$-0.72 = -0.72 + 1.42 * \underline{0} + 2.79 * \underline{0}$$

Recode = 1

$$0.70 = -0.72 + 1.42 * \underline{1} + 2.79 * \underline{0}$$

Recode = 2

$$2.07 = -0.72 + 1.42 * \underline{0} + 2.79 * \underline{1}$$

I. CSS(Credit Scoring System) 일반

모형개발방법론

모형적합 - ③ Scorecard의 이해

○ Logistic Regression의 결과값 해석(SAS OUTPUT)

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept	1	-1.7412	0.0512	1156.6690	<.0001	
r_FIRST_CONT_PRD	1	0.1703	0.0310	30.2667	<.0001	
r_FIRST_CONT_PRD	1	0.1940	0.0378	26.3778	<.0001	
r_FIRST_CONT_PRD	1	0.2915	0.0366	63.4347	<.0001	
r_FIRST_CONT_PRD	1	0.5320	0.0349	232.2365	<.0001	
r_FIRST_CONT_PRD	1	0.8105	0.0464	305.3544	<.0001	
r_FIRST_CONT_PRD	1	1.1013	0.0729	22.1111	<.0001	
r_FIRST_CONT_PRD	1	0.0425	0.0337	1.5849	0.2101	
r_FIRST_CONT_PRD	1	0.2303	0.0460	25.0000	<.0001	
r_FIRST_CONT_PRD	1	0.3858	0.0552	48.9000	<.0001	
r_FIRST_CONT_PRD	1	0.7385	0.0206	1281.1111	<.0001	
r_FIRST_CONT_PRD	1	0.2134	0.0386	30.5556	<.0001	
r_MAX2_ARR_ILSU	2	1.0093	0.0345	854.1111	<.0001	
r_I21140100_det	1	0.8349	0.0326	656.8889	<.0001	
r_I21140100_det	2	1.2924	0.0315	1678.7130	<.0001	
r_I21140100_det	3	1.6992	0.0354	2301.1821	<.0001	
r_I24000105	1	0.2642	0.0268	97.4557	<.0001	
r_I24000105	1	0.7671	0.0281	744.6920	<.0001	
r_I24000105	1	0.4926	0.0240	421.0053	<.0001	
r_I24000105	1	0.1902	0.0340	31.2200	<.0001	
r_I24000105	1	0.6489	0.0330	385.9061	<.0001	
r_I24000105	1	0.5606	0.0353	251.6865	<.0001	
r_I24000105	1	1.3861	0.0322	1857.3466	<.0001	
r_I24000105	1	0.0270	0.0343	0.6199	0.4311	
r_PE0000074	2	0.4045	0.0412	96.2111	<.0001	

2

추정치(Estimate) 값의 부호 : 추정치의 부호가 양(+)의 값을 가져야 함. 이미 개별 항목은 불량률 순으로 순서화되어 순차적인 (ordinary)변수로 리코딩되었기 때문에 부호가 음(-)으로 나온다면 이는 변수 상호간의 높은 상관관계로 인한 것임. Recoding이 '0'인 구간 값은 추정치도 0이 되므로 배점이 0점이 됨.

1

P-Value : 해당 항목의 추정치 (Estimates)의 유의성(significance)을 나타내는 척도로서 일반적으로 5%(0.05) 미만이어야 해당 항목이 유의적이라고 판단함.

3

변수의 타당성 : 통계적으로 선택된 결과를 바탕으로 비즈니스 로직 또는 타행자료 분석경험 등을 통해 변수를 추가 및 교체할 수 있음. 그러나, 이는 위의 2가지 기준(P-Value, Estimate값의 부호)을 부합시키는 경우에 한해서이며, 기준을 부합시키지 않는 경우의 인위적인 추가는 불가능함.

모형개발방법론

모형적합 - ④ Scorecard 수정 방법

Analysis of Maximum Likelihood Estimates		
Parameter	Estimate	Pr > ChiSq
Intercept: TARGET=0	-1.4205	<.0001
미해제연체총건수 1	0.5832	<.0001
미해제연체총건수 2	1.1006	<.0001
미해제연체총건수 3	1.6598	<.0001
미해제최장연체일수 1	0.8700	<.0001
미해제최장연체일수 2	0.9490	<.0001
최근1개월최고회수금액 1	0.4568	<.0001
최근1개월최고회수금액 2	0.6456	<.0001
최근3개월내총할부이용금액 1	0.1596	0.0074
최근3개월내총할부이용금액 2	-0.1447	0.1472
최근3개월최고CA잔액 1	0.0845	0.5653
최근3개월최고CA잔액 2	0.1648	0.0146
최근1년내미상환대출총건수 1	0.1241	0.0254
최근1년내미상환대출총건수 2	-0.1190	0.3071

최종변수 포함여부		
포함여부	Check List	사후조치
포함	▪ 제외사유 없음	▪ 없음
포함	▪ 제외사유 없음	▪ 없음
포함	▪ 제외사유 없음	▪ 없음
제거	▪ P-Value ≥ 0.05 ▪ Estimate < 0	▪ Correlation 확인 후 높은 변수 제거
범주 수정 후 포함	▪ P-Value ≥ 0.05	▪ 유의하지 않은 구간 결합 후 재적합
제거	▪ P-Value ≥ 0.05 ▪ Estimate < 0	▪ Correlation 확인 후 높은 변수 제거

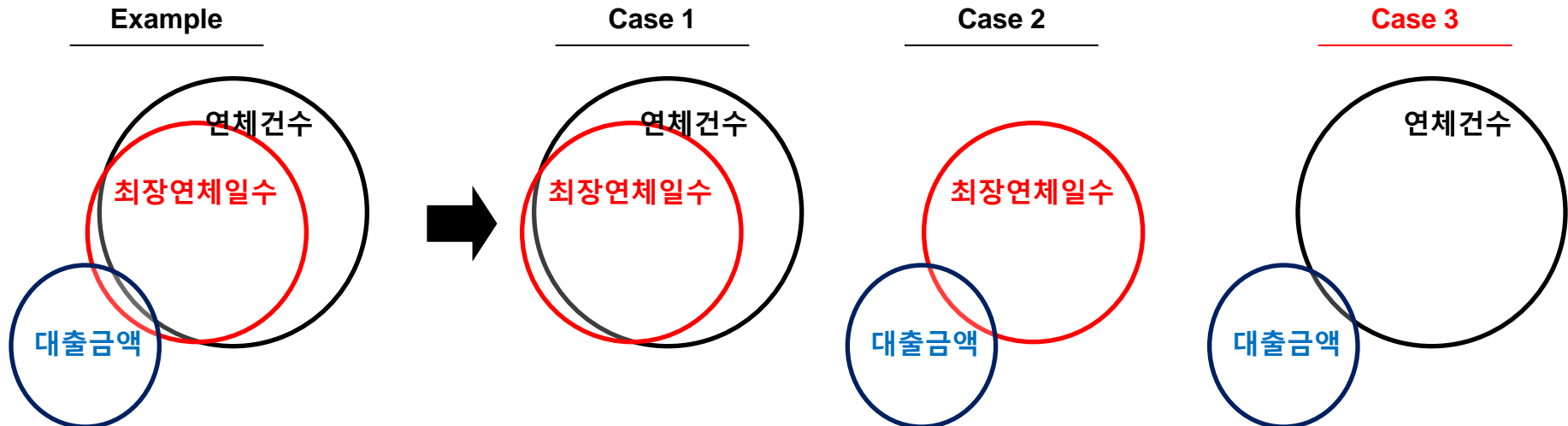
모형개발방법론

모형적합 - ④ Scorecard 수정 방법

○ 상관관계(다중공선성) 제거 - 최종 선택된 평가항목 간의 상관관계는 낮아야 하며 그 값의 수준이 유의적이어야 함

단일변수 예측력		변수간 Correlation		
변수명	I.V	연체건수	최장연체일수	대출금액
연체건수	0.713	1.0	0.8	0.3
최장연체일수	0.607		1.0	0.4
대출금액	0.253			1.0

상관관계가 높다 = 중복부분이 많다

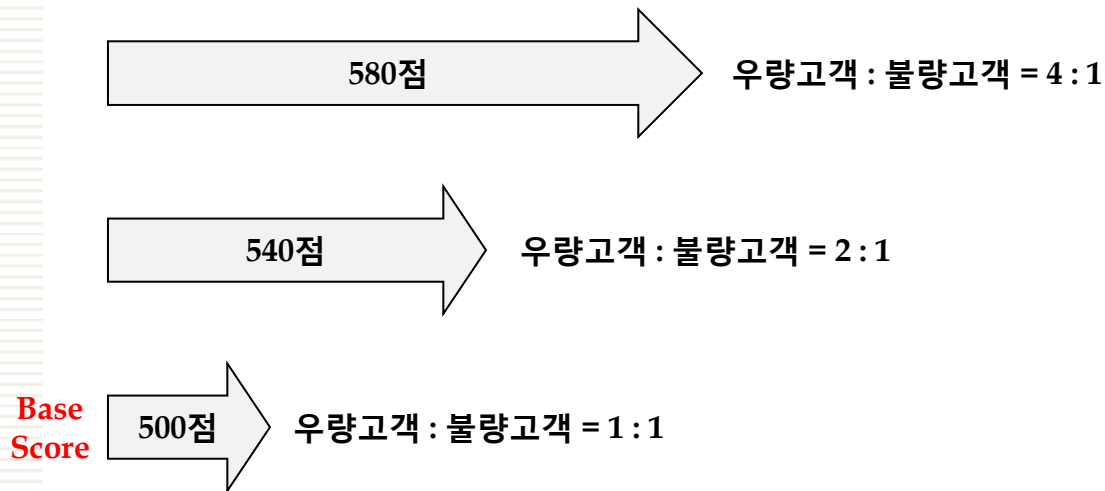


모형개발방법론

평점표 구성 - ① PDO를 통한 Score 산출

- PDO(Point to Double Odds)란 Odds가 2배씩 증가하는 점수 구간을 의미함
- 신용평가 모형에서 Score는 Base Score(일반적으로 Odds=1인 점수를 의미)와 PDO 개념을 통해 산출됨

➤ Score 산출 예시 : PDO = 40, Base Score = 500인 경우

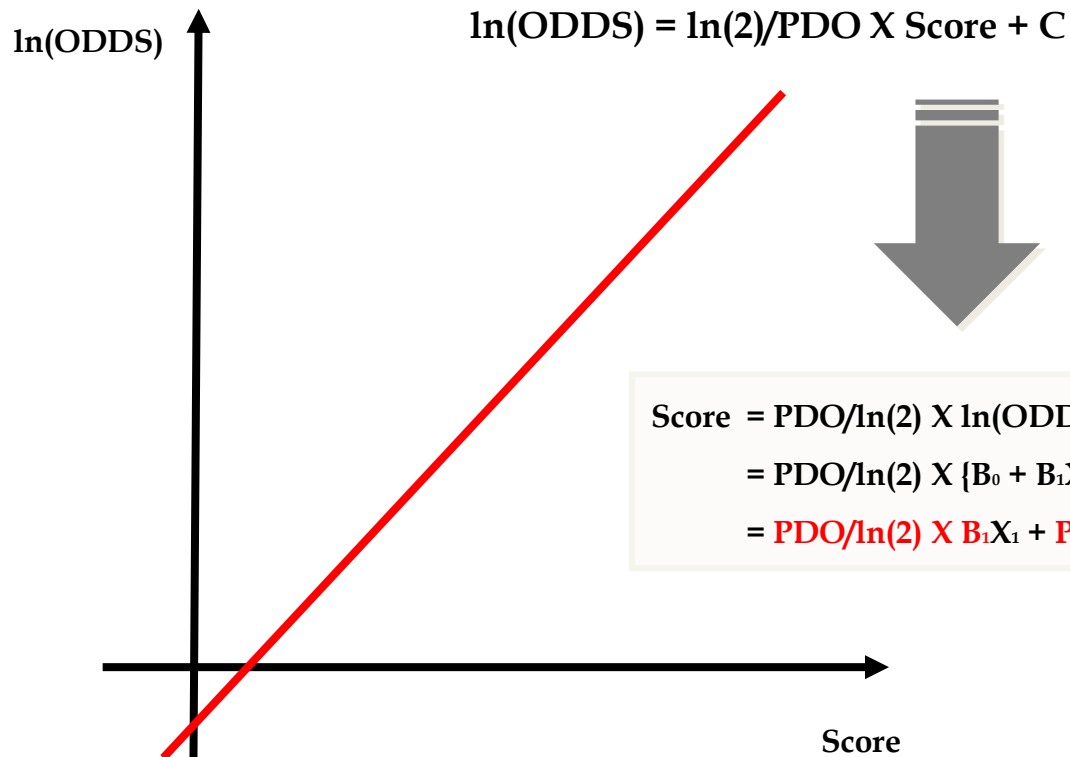


✓ Score의 의미

신용평점모형은 500점에서 불량고객1명 대비 우량고객 1명이 되도록 하여, 40점씩 증가(감소)할 때 2배(1/2배) 되도록 구성됨

모형개발방법론

평점표 구성 - ② Score to log Odds Relationships



$$\begin{aligned} \text{Score} &= \text{PDO}/\ln(2) \times \ln(\text{ODDS}) \\ &= \text{PDO}/\ln(2) \times \{B_0 + B_1X_1 + \dots + B_pX_p\} \\ &= \text{PDO}/\ln(2) \times B_1X_1 + \text{PDO}/\ln(2) \times B_2X_2 + \dots + \text{PDO}/\ln(2) \times B_pX_p \end{aligned}$$

모형개발방법론

평점표 구성 - ③ 평점표 작성 (PDO = 40기준)

항목명	Recoding	항목구간설명	Estimate	배점	배점 전환 Logic
소진율	3	~ 30%	1.213	70	$1.213 * (40/0.69315) = 70$
	2	31% ~ 50%	0.866	50	$0.866 * (40/0.69315) = 50$
	1	51% ~ 70%	0.346	20	$0.346 * (40/0.69315) = 20$
	0	71% ~	0.000	0	$0.000 * (40/0.69315) = 0$
Debt-to-Income Ratio	2	~ 20%	2.599	150	$2.599 * (40/0.69315) = 150$
	1	21% ~ 40%	1.039	60	$1.039 * (40/0.69315) = 60$
	0	41% ~	0.000	0	$0.000 * (40/0.69315) = 0$
최근 12개월 동안 연체경험횟수	2	0건	1.733	100	$1.733 * (40/0.69315) = 100$
	1	1건	0.690	40	$0.690 * (40/0.69315) = 40$
	0	2건 이상	0.000	0	$0.000 * (40/0.69315) = 0$

(단, $\log(2) = 0.69315$)

I. CSS(Credit Scoring System) 일반

모형개발방법론

평점표 구성 - ④ Scorecard의 이해

○ Scorecard의 일반적인 형태

정보영역	항목설명	Cat	CODE 설명	배점	p-value	개발%	TODDS	불량률
기본배점				219				
내부 수신정보	최근 6개월간 요구불평잔합	1	월 100만원 초과	57	0.05	25.9%	133.33	0.7%
		2	월 20만원 초과 100만원 이하	12	0.53	20.1%	39.64	2.2%
		3	월 20만원 이하	-		54.1%	15.33	5.2%
내부 수신정보	최근 6개월간 저축성평잔합	1	월 20만원 초과	80	0.00	36.1%	138.75	0.7%
		2	월 20만원 이하	-		63.9%	16.05	5.0%
내부 수신정보	최근 3개월간 급여이체액합	1	월 200만원 초과	39	0.24	18.3%	122.00	0.8%
		2	월 200만원 이하	-		81.7%	20.48	4.0%
내부 대출정보	최근 3개월간 평균 대출평잔	1	0원 초과 2,000만원 이하	53	0.01	30.8%	75.89	1.2%
		2	0원, 2,000만원 초과	-		69.2%	18.62	4.4%
KFB 채무불이행정보	(3년내)등록삭제건중최근삭제일로부터의경과일수	1	없음	93	0.00	90.6%	33.53	2.6%
		2	1년 초과	43	0.09	6.7%	6.88	10.2%
		3	1년 이하	-		2.7%	3.62	15.4%
KFB 개설정보	미해지신용카드총건수	1	6건 초과	110	0.00	29.7%	79.12	1.1%
		2	5건 ~ 6건	58	0.00	21.5%	34.26	2.5%
		3	2건 ~ 4건	44	0.01	36.3%	19.05	4.3%
		4	1건 이하	-		12.5%	9.54	8.0%
KFB 대출정보	미상환주택담보제외대출총건수	1	1건 이하	42	0.00	72.1%	32.78	2.7%
		2	1건 초과	-		27.9%	14.29	5.5%
CB 연체정보	1개월이상 경험 연체 총 건수(해제후1년경과건제외)	1	無	65	0.00	92.0%	35.51	2.5%
		2	有	-		8.0%	3.19	14.5%
CB 연체정보	최근3개월내 경험한 최장 연체경험기간	1	無	93	0.00	84.8%	40.13	2.2%
		2	15일 이하	63	0.01	7.2%	13.44	5.0%
		3	15일 초과	-		7.9%	3.03	14.9%

모형개발방법론

평점 통합 - ① Scaling

▪ 평점표 Scaling 이란?

평점표 Scaling은 개별 평점표 별 Score를 Odds 혹은 불량률 기준으로 Mapping하는 표준화 작업을 의미함

➤ Odds 기준 Scaling

- 동일한 평점에 동일한 Odds를 부여하는 방법
- 우불량 정의에서 판단미정(Indeterminate)의 비중이 약 10% 미만인 경우에만 활용 가능

➤ 불량률 기준 Scaling

- 동일한 평점에 동일한 불량률을 부여하는 방법
- 모든 평점 구간에서 동일하게 불량률을 부여할 수 없으며 구성비가 집중된 구간을 중심으로 불량률을 맞추어 평점을 조정함
- 구성비가 작은 평점 구간에서는 불량률의 차이가 클 수 있으며 모형의 개수가 많을수록 불량률의 차이도 커지는 경향이 있음

I. CSS(Credit Scoring System) 일반

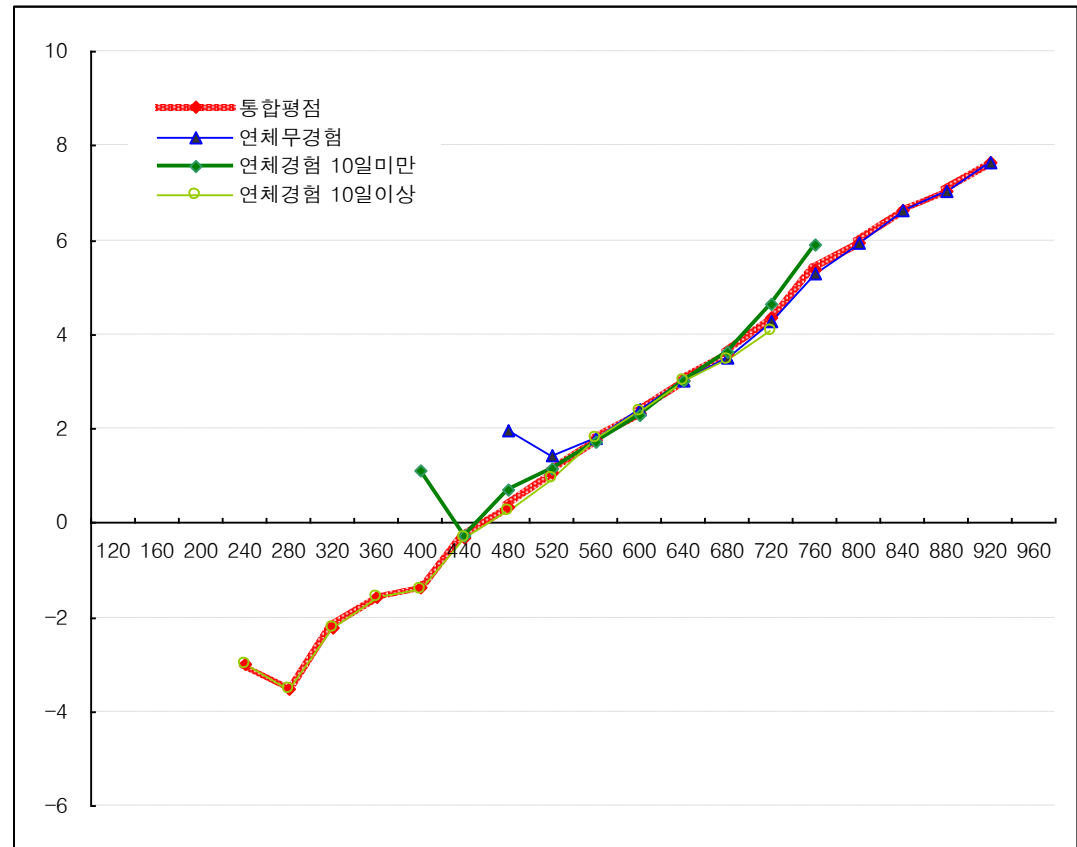
모형개발방법론

평점 통합 - ② Scaling 사례 (Odds 기준 Scaling)

평점구간별 TODDS

SCR_40	통합평점	연체무경험	연체경험 10일미만	연체경험 10일이상
120				
160	0.00			0.00
200	0.00			0.00
240	0.05			0.05
280	0.03			0.03
320	0.11			0.11
360	0.21			0.21
400	0.25		3.00	0.24
440	0.74		0.75	0.74
480	1.42	7.00	1.98	1.29
520	2.90	4.17	3.09	2.62
560	5.80	6.07	5.51	5.97
600	10.29	11.07	9.69	10.45
640	20.18	20.10	20.16	20.57
680	35.33	33.18	38.15	31.40
720	78.45	70.80	102.75	57.83
760	211.59	201.42	368.57	
800	383.55	382.63		
840	761.03	761.03		
880	1133.50	1133.50		
920	2106.00	2106.00		
960				
전체	26.13	28.13		18.09

평점구간별 TODDS



→ 동일 평점대에 동일한 Odds 수준 확인

모형개발방법론

등급화(Grading)

○ 등급을 구성하는 접근 방법은 다음과 같음

정보 구분	내 용
구성비에 따른 등급 정의	<ul style="list-style-type: none"> ▪ 평점대별 구성비만을 고려한 등급정의로 특정 등급에 집중화 되는 현상 방지 가능 ▪ 정규분포에 가까운 등급을 정의할 수 있음 ▪ 등급이 구성비에 종속되어 등급별 불량률 차별화가 다소 약화되는 경향이 있음
불량률에 따른 등급 정의	<ul style="list-style-type: none"> ▪ 등급 정의 시 불량률 수준을 감안하여 특정 등급에 특정 불량률의 의미 부여 ▪ 운영 시 등급에 대해 직관적인 해석이 가능함 ▪ 등급이 불량률에 종속되어 특정 등급에 구성비가 집중되는 현상 발생가능
분석적 방법에 의한 등급 정의	<ul style="list-style-type: none"> ▪ 통계적 분석방법인 의사결정나무(Decision Tree) 분석에 따라 리스크를 최적화하여 등급 정의 가능 ▪ 통계적 분석 결과에 따라 운영의 편의를 위해 특정 개수의 등급으로 정의하기 어려운 경향이 있음 ▪ 특정 등급에 구성비가 집중되는 현상이 발생할 수 있음
전략적인 등급 정의	<ul style="list-style-type: none"> ▪ 평점대별 구성비와 불량률을 적절히 감안하고, 현업의 경험을 반영하여 전략적인 등급 정의 가능 ▪ 일반적으로 선호되는 방식임 ▪ 분석자의 자의성이 반영될 수 있는 가능성이 존재하며, 등급정의 시 타당한 근거자료를 필요로 함

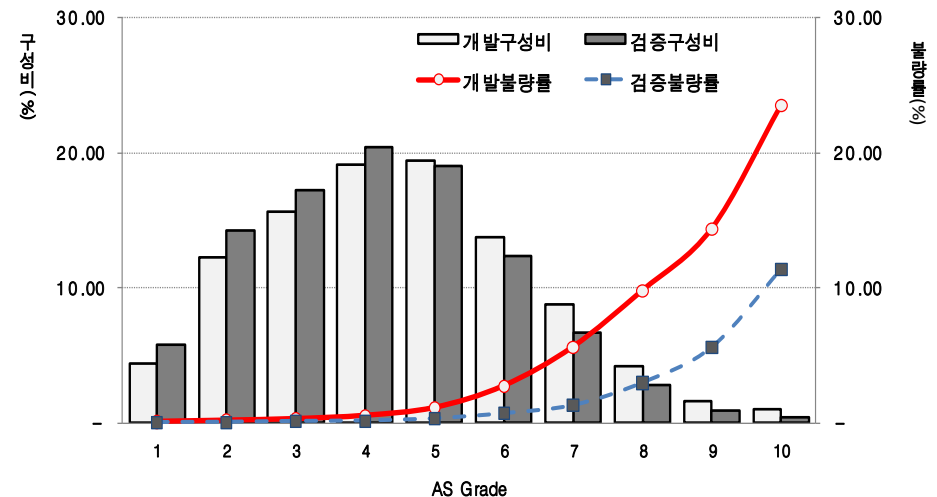
모형개발방법론

등급화(Grading)

○ 일반적으로 등급화는 아래의 내용을 고려하여 정의함

- ① 등급 구간별 **불량률의 분명한 서열화**가 존재
- ② 등급의 구성비는 한 등급에 **너무 작거나 크지 않게 구성 (3~20%)**
- ③ 등급의 분포는 운용 편의성을 고려하여 **정규분포 형태**로 구성
- ④ 개발 검증 시 등급별 **불량률이 역전되지 않도록** 구성
- ⑤ 향후 안정성, 변별력 및 서열화에 신용평가모형 운용기준에 부합하는지 지속적인 모니터링 필요

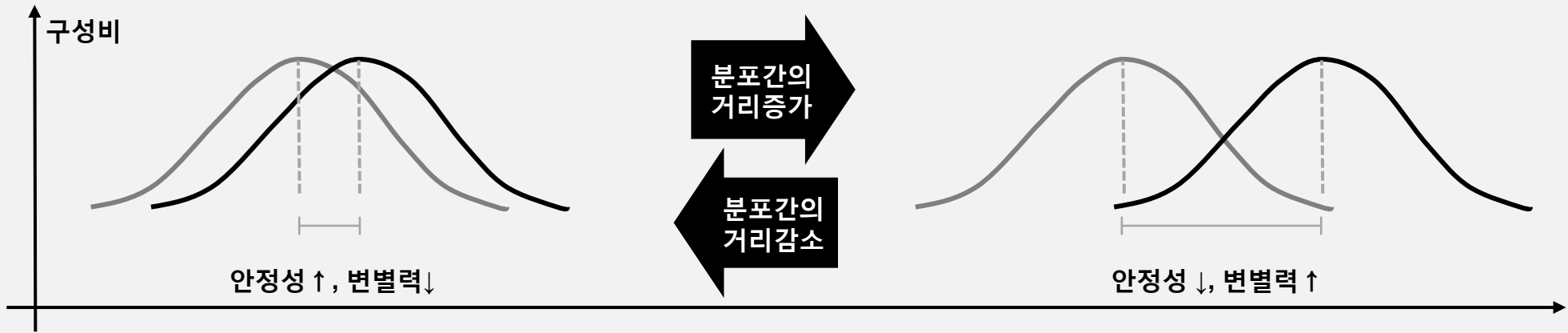
등급	평점구간
10등급	530점미만
9등급	530점이상 600점미만
8등급	600점이상 650점미만
7등급	650점이상 700점미만
6등급	700점이상 730점미만
5등급	730점이상 750점미만
4등급	750점이상 780점미만
3등급	780점이상 800점미만
2등급	800점이상 820점미만
1등급	820점이상



모형검증방법론

모형 검증 방법론 개요

안정성 및 변별력 측정 지표의 기본 개념



검증 보고서 구분

Front-end Report(안정성 검증 보고서)
모형평점 및 등급과 평가항목의 안정성 검증

Back-end Report(변별력 검증 보고서)
모형평점 및 등급과 평가항목의 변별력과
변별력의 안정성 검증

검증지표 / 검증보고서

- PSI(Population Stability Index)
- CAR(Characteristic Analysis Report)
- K-S Statistic(Kolmogorov-Smirnov Statistic)
- AR(Accuracy Ratio)
- AUROC(Area Under Receiver Operating Characteristic)
- SDR(Score Distribution Report)
- CDR(Characteristic Distribution Report)

모형검증방법론

안정성 검증 지표의 이해 - PSI, CAR

A PSI - Population Stability Index

- 모형 개발시점 대비 운영시점 간 평가대상의 **Score 구간/등급 분포 변화 정도**를 측정하는 지표
- PSI가 불안정한 경우 CAR 분석을 통해 평가항목의 변동성 확인

산출 공식 (판단 기준)	<ul style="list-style-type: none"> ▪ 모형의 안정성 측정 지표 ▪ Score 구간/등급별 (% of 개발시점 - % of 검증시점) x $\ln(\% \text{ of 개발시점} / \% \text{ of 검증시점})$의 합계 <ul style="list-style-type: none"> ✓ $PSI < 0.10$ - Green(안정) ✓ $0.10 \leq PSI < 0.25$ - Yellow(다소 불안정) ✓ $PSI \geq 0.25$ - Red(불안정)
------------------	---

B CAR - Characteristic Analysis Report

- 모형 개발시점 대비 운영시점 간 평가대상의 **평가항목 Classing 분포 변화 정도**를 측정하는 지표
- 양(+)의 값이 나오면 우량고객, 음(-)의 값이 나오면 불량고객의 비중이 증가

산출 공식 (판단 기준)	<ul style="list-style-type: none"> ▪ 평가항목의 안정성 측정 지표 ▪ 평가항목 Classing별 (% of 검증시점 - % of 개발시점) x 배점의 합계 <ul style="list-style-type: none"> ✓ $CAR \leq 10$ - No Action Required ✓ $CAR > 10$ - 전산 코딩 및 Logic 체크 or Data Quality 및 Biz 환경 변화 체크
------------------	---

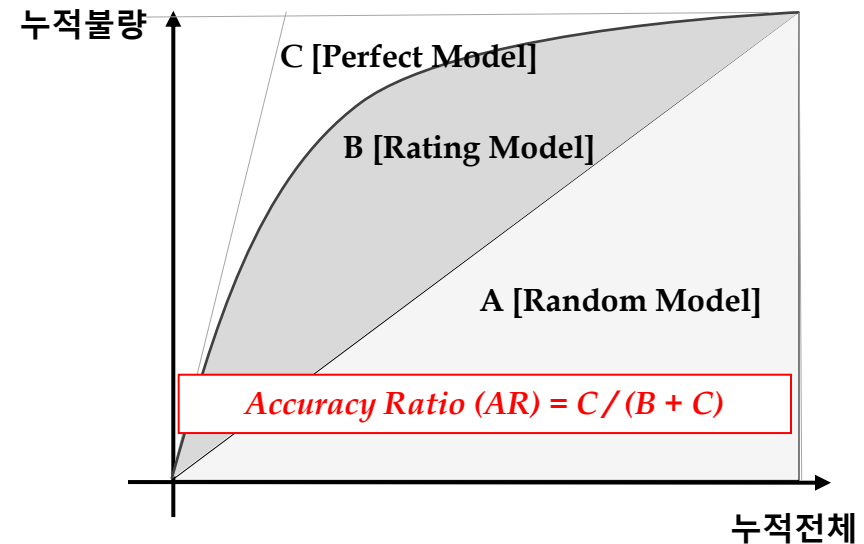
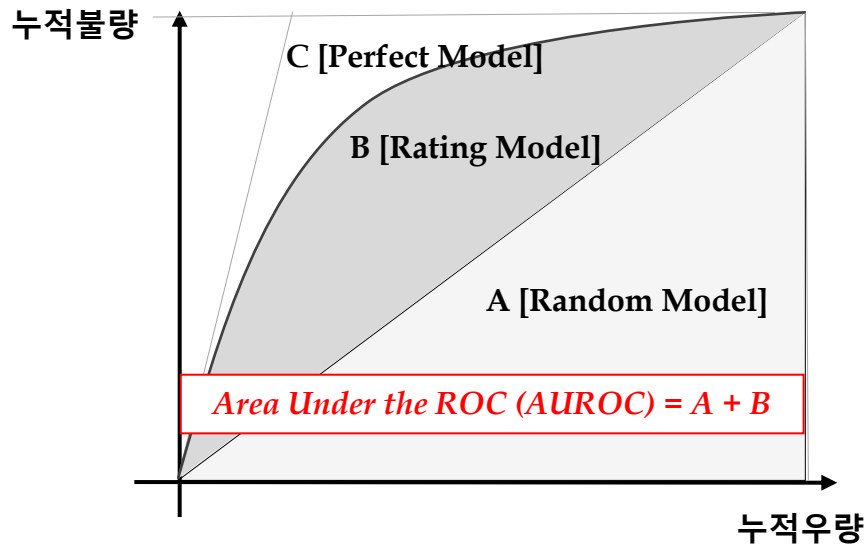
모형검증방법론

변별력 검증 지표의 이해 - AUROC, AR, K-S Statistic, I.V.

C AUROC - Area Under the ROC

- 평가대상을 상대적으로 서열화할 경우, 불량계좌가 하위 구간에 얼마나 집중되어 있는지를 평가함
- AUROC와 AR(Gini)는 판단미정이 없을 경우 정확한 선형관계에 있음[$AR = 2 \times AUROC - 1$]

산출 공식 (판단 기준)	<ul style="list-style-type: none"> 모형(평점/등급) 및 항목 Classing별 $0.5 \times \% \text{ of 우량}_i \times \% \text{ of 불량}_i + (1 - \% \text{ of 누적우량}_i) \times \% \text{ of 불량}_i$의 합계 ✓ [AS] AUROC ≥ 0.65 - Good ✓ [BS] AUROC ≥ 0.70 - Good <p>주) % of 누적우량은 불량구간에서 우량구간 순으로 누적된 % of 우량</p>
------------------	---



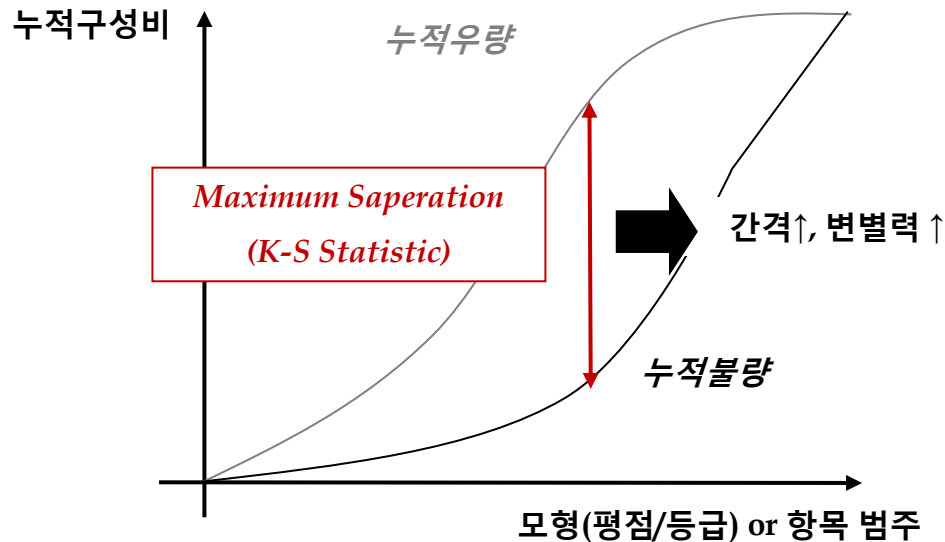
모형검증방법론

변별력 검증 지표의 이해 - AUROC, AR, K-S Statistic, I.V.

D K-S Statistic - Kolmogorov-Smirnov Statistic

- 모형 및 평가항목의 변별력이 극대화되는 지점을 측정하여 평가하는 지표
- 우량집단과 불량집단 간의 누적분포 차이의 최대값

산출 공식 (판단 기준)	<ul style="list-style-type: none"> 모형(평점/등급) 및 항목 Bin별 $\% \text{ of 누적우량}_i - \% \text{ of 누적불량}_i$의 최대값 ✓ $K-S < 30$ - the scorecard's probably not worth using ✓ $30 \leq K-S < 40$ - Fair ✓ $K-S \geq 40$ - Good
------------------	--



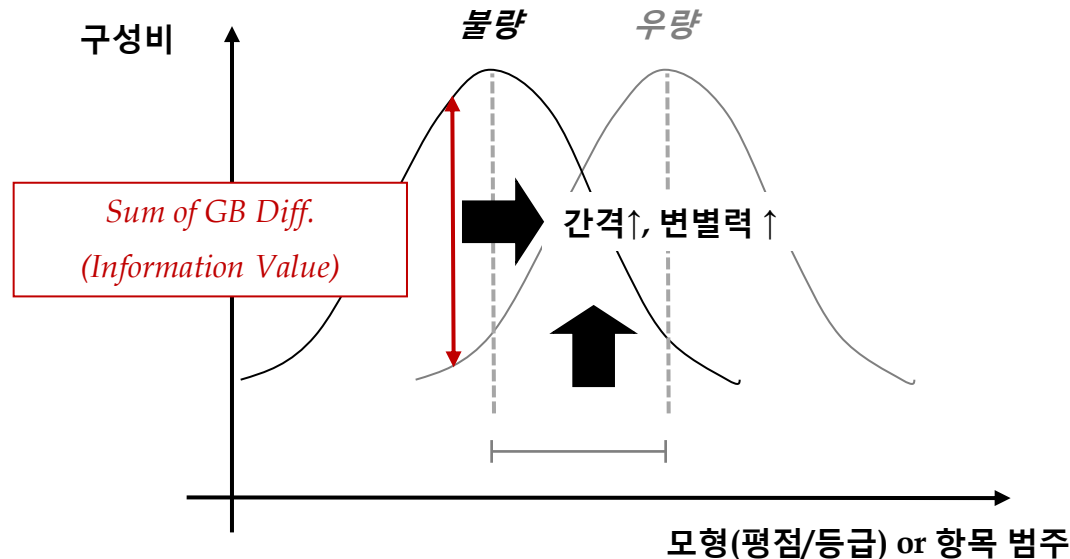
모형검증방법론

변별력 검증 지표의 이해 - AUROC, AR, K-S Statistic, I.V.

E I.V. - Information Value

- 우량집단과 불량집단의 Score 구간/등급 분포의 차이를 측정하는 지표
- 우량집단과 불량집단 간의 분포 차이의 합계

산출 공식 (판단 기준)	<ul style="list-style-type: none"> ▪ 모형(평점/등급) 및 항목 Bin별 (% of 우량 - % of 불량) x $\ln(\% \text{ of 우량} / \% \text{ of 불량})$의 합계 ✓ $IV < 0.5$ - the scorecard's probably not worth using ✓ $0.5 \leq IV < 1.0$ - Fair ✓ $IV \geq 1.0$ - Good
------------------	---



모형검증방법론

변별력 검증 지표의 이해 - AUROC, AR, K-S Statistic, I.V.

F SDR - Score Distribution Report

- 개발시점의 Score/등급별 불량률과 Ln(Odds)에 대한 서열화가 검증시점에도 잘 유지되는지 분석
- 계량적인 Guide-line은 없으며 과거 경험기준(Empirical Study)에 근거한 직관적 판단

산출 공식 (판단 기준)	<ul style="list-style-type: none"> ▪ Score 구간/등급별 불량률과 Ln(Odds)의 추세 비교 <ul style="list-style-type: none"> ✓ 검증시점의 불량률 및 Ln(Odds)가 선형증가 추세를 보이며 개발시점과 교차하지 않음 - Green ✓ 검증시점의 불량률 및 Ln(Odds)가 선형증가 추세를 보이나 개발시점과 교차함 - Yellow ✓ 검증시점의 불량률 및 Ln(Odds)가 선형증가 추세를 보이지 않음 - Red
------------------	---

G CDR - Characteristic Distribution Report

- 개발시점의 평가항목 Classing별 불량률과 Ln(Odds)에 대한 서열화가 검증시점에도 유지되는지 분석
- 계량적인 Guide-line은 없으며 과거 경험기준(Empirical Study)에 근거한 직관적 판단

산출 공식 (판단 기준)	<ul style="list-style-type: none"> ▪ 항목 Classing별 불량률과 Ln(Odds)의 추세 비교 <ul style="list-style-type: none"> ✓ 검증시점의 불량률 및 Ln(Odds)가 선형증가 추세를 보이며 개발시점과 교차하지 않음 - Green ✓ 검증시점의 불량률 및 Ln(Odds)가 선형증가 추세를 보이나 개발시점과 교차함 - Yellow ✓ 검증시점의 불량률 및 Ln(Odds)가 선형증가 추세를 보이지 않음 - Red
------------------	---