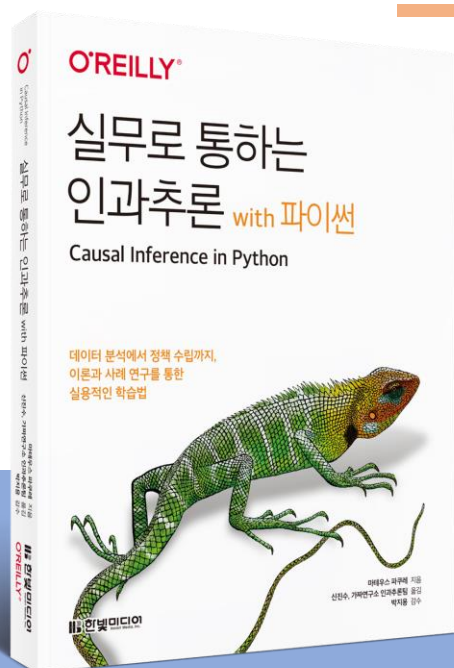


『실무로 통하는 인과추론 with 파이썬』 특강

실무자가 바라본 고객 신용한도와 채무불이행률의 인과적 관계



정호재 (Jeong Hojae)

롯데캐피탈 신용분석2팀 (2022.01~)

- 신용평점시스템 (Credit Scoring System, CSS) 개발
- 상품 전략 고도화 (대출승인/한도/금리 정책변경)
- Default율 변동에 따른 Risk 관리



- 개인신용대출 도메인 지식
- 금융회사에서의 회귀분석
- 예제 소개

내 신용점수

대출 추천

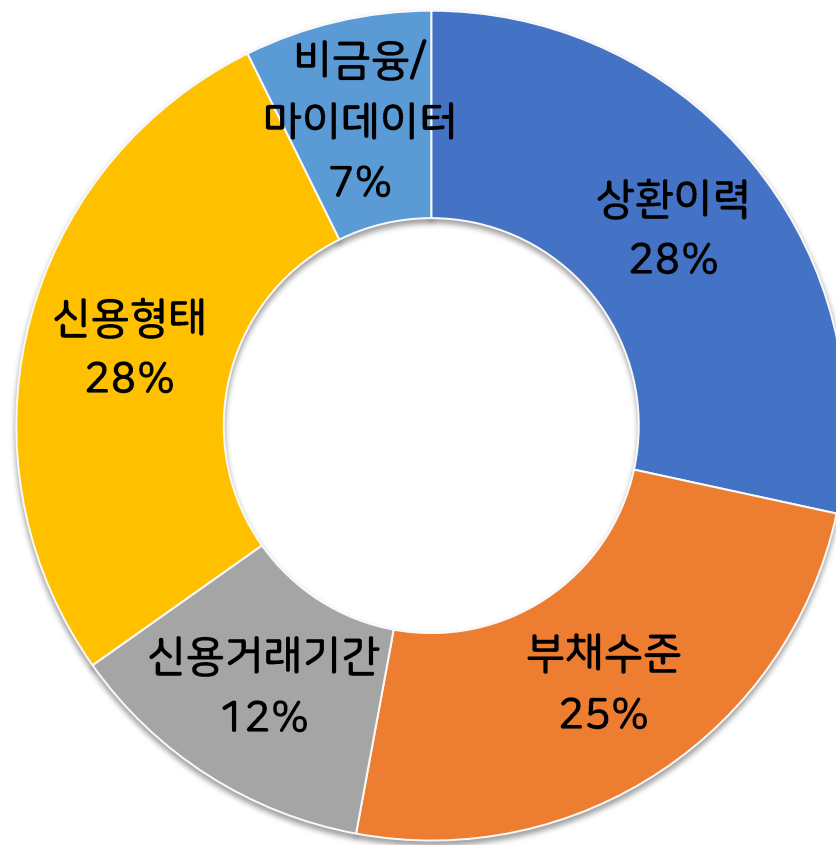


상위 %



상위 %

- 신용평점시스템 (CSS) 활용 변수 - NICE Score 기준



※ NICE 평가정보(주) - 신용등급체계공시

- 채무불이행(Default)

: 채무자가 이자나 원리금 상환을 계약에 정해진 대로 이행할 수 없게 된 상황

- 신용정보의 이용 및 보호에 관한 법률 제 18조 2항

- 신용정보의 이용 및 보호에 관한 법률 시행령 제 15조 4항

- 일반신용정보관리규약 <별표1> 신용정보관리기준



금융위원회
Financial Services Commission



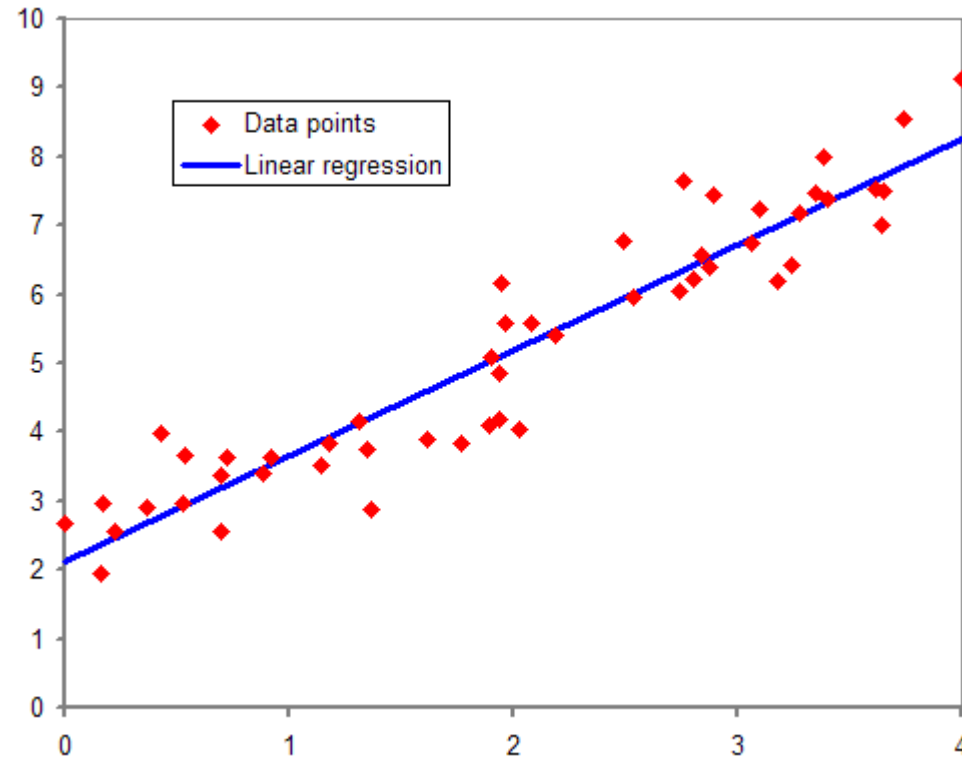
한국신용정보원



- 채무불이행 판단
 - 연체, 부도, 대위변제 및 대지급과 관련된 정보
 - 신용질서 문란행위와 관련된 정보
 - 법원의 파산선고 · 면책 · 복권 결정 및 회생 · 간이회생 · 개인회생의 결정과 관련된 정보
 - 체납 관련 정보
 - 유사한 형태의 불이익정보로서 금융위원회가 정하여 고시하는 신용정보

- 회귀분석

: 관측된 독립 변수와 종속 변수 간의 관계를 모델링 및 분석하여 예측과 인과추론을 수행하는 통계적 기법



- 회귀분석 장점 (MA Iqbal. (2020))
 - 로버스트 하고 효율적이다
 - 해석이 쉽고 속도가 빠르다
 - 변수 간의 관계를 찾는데 사용 가능하다

- 금융회사에서 회귀분석을 선호하는 이유 (Bücker et al. (2021))
 - 감사를 위한 모형 해석이 필요
 - 지속적인 모니터링이 필요
 - 고객에 심사 결과를 설명해야 할 의무 존재

- 목적
 - 신용 한도가 채무불이행에 미치는 영향을 추정
 - $Default_i = \beta_0 + \beta_1 line_i + e_i$

- 데이터 소개

	wage	educ	Exper	Married	credit_score1	credit_score2	credit_limit	default
0	950.0	11	16	1	500.0	518.0	3200.0	0
1	780.0	11	7	1	414.0	429.0	1700.0	0
2	1230.0	14	9	1	586.0	571.0	4200.0	0
3	1040.0	15	8	1	379.0	411.0	1500.0	0
4	1000.0	16	1	1	379.0	518.0	1800.0	0

- 회귀 분석으로 신용한도와 채무불이행 관계 추정
 - 신용한도 (처치 : T)
 - 채무불이행 (결과 : Y)

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.2192	0.004	59.715	0.000	0.212	0.226
credit_limit	-2.402e-05	1.16e-06	-20.689	0.000	-2.63e-05	-2.17e-05

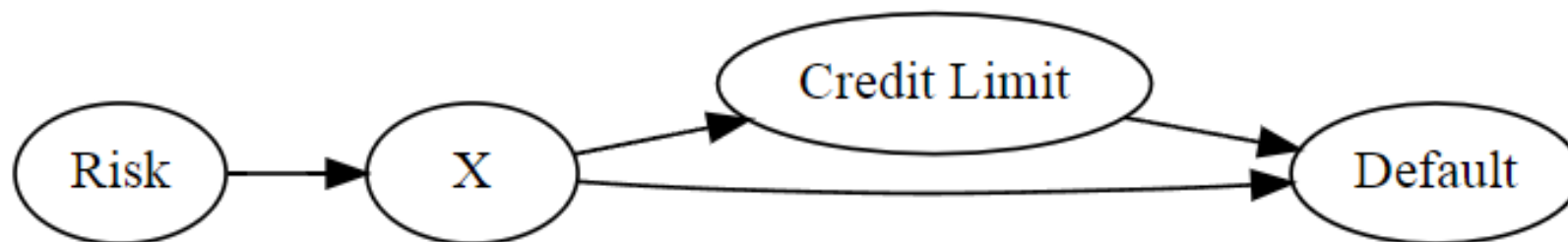
- 결과 : 신용한도 1달러가 증가할 때 채무불이행률은 0.00002402 만큼 감소한다
신용한도 1000달러를 높여주면 채무불이행률은 2.402% 감소한다
- 결론
 - 많은 사람들에게 돈을 많이 빌려줘서 채무불이행 리스크를 없애자!

예제 소개 (1)

- 왜 이런 결과가 나왔을까?
 - 교란으로 인해 상식과 반대되는 결과가 나왔던 것!
 - 신용도가 높은 고객일 수록 한도나 금리에 우대를 더 많이 받기 때문에 교란을 줄 수 있는 변수들을 잘 통제하여 분석을 진행합니다
 - $Default_i = \beta_0 + \beta_1 line_i + \theta X_i + e_i$

- 회귀 모형에 교란 요인을 포함하여 신용한도와 채무불이행 관계 추정
 - 신용한도 (처치 : T)
 - 채무불이행 (결과 : Y)
 - 신용도1, 신용도2, 임금 (교란 요인 : X)

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.4037	0.009	46.939	0.000	0.387	0.421
credit_limit	3.063e-06	1.54e-06	1.987	0.047	4.16e-08	6.08e-06
wage	-8.822e-05	6.07e-06	-14.541	0.000	-0.000	-7.63e-05
credit_score1	-4.175e-05	1.83e-05	-2.278	0.023	-7.77e-05	-5.82e-06
credit_score2	-0.0003	1.52e-05	-20.055	0.000	-0.000	-0.000



예제 소개 (1)

- $E[y|t, X] : Default_i = \beta_0 + \beta_1 line_i + \theta X_i + e_i$ ← 회귀추정식
- $\frac{\partial}{\partial t} E[y|t, X] = \beta_1$ ← 처치가 결과에 대한 효과
- 교란 요인을 포함하여 추정한 회귀식에서 처치변수에 대한 매개변수 값은
모델의 다른 모든 변수가 고정된 상태에서 처치효과를 나타내는 것
- 따라서 앞의 결과에선 동일 임금, 동일 신용점수에서
신용한도 1000달러를 높여주면 채무불이행률은 0.3063% 증가한다

- FWL 스타일의 직교화를 사용한 편향 제거

1. 편향 제거 단계 : 처치 T 를 교란 요인 X 에 회귀하여 처치 잔차 $\tilde{T} = T - \hat{T}$ 를 구합니다

2. 잡음 제거 단계 : 결과 Y 를 교란 요인 X 에 회귀하여 결과 잔차 $\tilde{Y} = Y - \hat{Y}$ 를 구합니다

3. 결과 모델 단계 : 결과 잔차 \tilde{Y} 를 처치 잔차 \tilde{T} 에 회귀하여 T 가 Y 에 미치는 인과효과 추정값을 구합니다

- FWL 스타일의 직교화를 사용한 편향 제거

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.1421	0.005	30.458	0.000	0.133	0.151
credit_limit	3.063e-06	1.54e-06	1.987	0.047	4.17e-08	6.08e-06

- 다중회귀분석 결과랑 동일한 값 도출
- 대용량 데이터에서 FWL 정리를 사용하여 데이터 차원을 줄여 계산 시간을 줄일 수 있다

- 서민 · 소상공인에 대한 신속 신용회복지원
 - 지원대상
 - 소액연체자(2천만원 이하)
 - '2021.09.01 ~ '2024.01.31 까지 불이행한 채무를 가진자
 - 2024.05.31까지 연체금액을 전액 상환한 자
 - 대상인원
 - 개인 : 약 298만명 (전체 연체발생자의 98%)
 - 개인사업자 : 약 31만명
 - 시행일자
 - 2024.03.12 ~

- 서민 · 소상공인에 대한 신속 신용회복지원

연체이력정보가 삭제되면?



- 실험 환경 구축
 - 전국민 5천만명 중 약 300만명이 대상 (6%)
 - 동일한 환경 구축을 위해 회귀모형에 포함하지 않을 exper변수에서 채무불이행률이 높은 3집단의 고객 3000명의 신용점수(credit_score1, credit_score2)를 39점 높이기로 함

<단위:건,%>

exper/default	Frequency		Percentage	
	0	1	0	1
21	538	112	82.769	17.231
4	1,275	262	82.954	17.046
3	50	10	83.333	16.667
18	1,329	254	83.955	16.045
6	2,186	412	84.142	15.858
17	2,363	445	84.152	15.848

- 원본 데이터의 회귀 결과

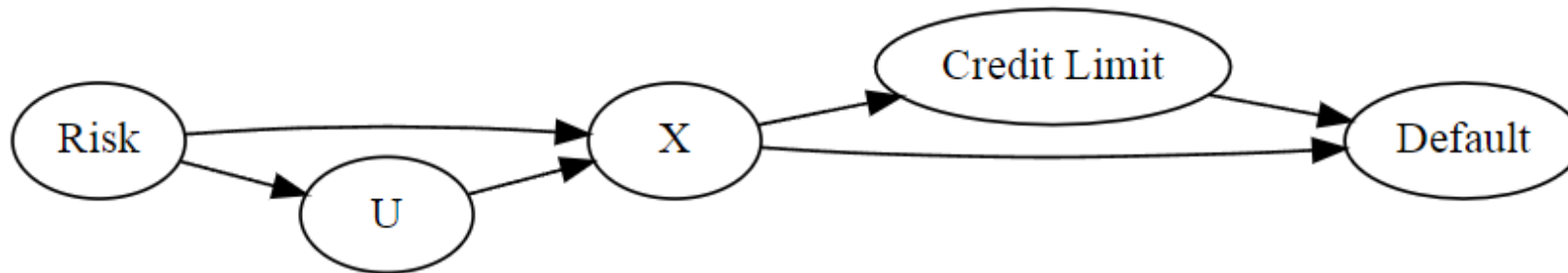
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.4037	0.009	46.939	0.000	0.387	0.421
credit_limit	3.063e-06	1.54e-06	1.987	0.047	4.16e-08	6.08e-06
wage	-8.822e-05	6.07e-06	-14.541	0.000	-0.000	-7.63e-05
credit_score1	-4.175e-05	1.83e-05	-2.278	0.023	-7.77e-05	-5.82e-06
credit_score2	-0.0003	1.52e-05	-20.055	0.000	-0.000	-0.000

- 신용사면 (신용점수를 높인 데이터)의 회귀 결과

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.4009	0.009	46.622	0.000	0.384	0.418
credit_limit	2.939e-06	1.54e-06	1.907	0.056	-8.11e-08	5.96e-06
wage	-9.005e-05	6.04e-06	-14.900	0.000	-0.000	-7.82e-05
credit_score1	-3.414e-05	1.82e-05	-1.876	0.061	-6.98e-05	1.53e-06
credit_score2	-0.0003	1.51e-05	-19.809	0.000	-0.000	-0.000

예제 소개 (2)

- 변수 통제를 완벽하게 했다면 신용한도와 채무불이행간의 관계는 동일
- 신용사면이라는 미지의 변수로 인한 신용한도와 채무불이행간의 관계에 변화 발생
- 어떤 고객이 사면 받았는지 알 수 없는 상황에서 인과관계 변화를 근거로 대응 정책을 수립



- 적은 데이터로 새로운 모델을 적용하기엔 리스크가 크니 신용한도를 조정하여 정부 정책에 대응
- 회귀 결과 : $3.063e-06 \rightarrow 2.939e-06$ (신용한도가 채무불이행에 미치는 영향 감소)
- 현 판단기준 변경시, 금융사의 정책 및 실적 등 큰 영향이 발생 위험이 있으니 기존에 활용했던 신용한도와 채무불이행 관계를 동일 적용하는 대신 두 영향의 비인 4%로 한도를 높이자고 제안
- 신용사면으로 인한 영향이 불확실한 상황에서 고객 한도를 높이자는 주장은 받아들여지기 힘들것군요

- 고객들의 신용점수구간별로 신용한도와 채무불이행 관계의 변화를 살펴봅시다
- 신용점수(credit_score1)별 고객 구성비를 고려하여 5구간으로 나눠서 회귀분석을 진행해봅시다
 - 1구간 : 300점 미만
 - 2구간 : 300점 이상 400점 미만
 - 3구간 : 400점 이상 500점 미만
 - 4구간 : 500점 이상 600점 미만
 - 5구간 : 600점 이상

<단위:건,%>

	Frequency		Percentage	
	변경 전	변경 후	변경 전	변경 후
1구간	2,915	2,816	5.83	5.632
2구간	11,311	10,992	22.622	21.984
3구간	15,198	15,226	30.396	30.452
4구간	13,854	14,058	27.708	28.116
5구간	6,722	6,908	13.444	13.816

- 신용점수 구간별 신용한도와 채무불이행 관계 변화 확인

		coef	std err	t	P> t	[0.025	0.975]	
1구간 (300점 미만)	변경 전	1.085e-05	8.11e-06	1.339	0.181	-5.05e-06	2.68e-05	5.9% 증가
	변경 후	1.153e-05	8.29e-06	1.391	0.164	-4.72e-06	2.78e-05	
2구간 (400점 미만)	변경 전	5.987e-06	3.7e-06	1.619	0.105	-1.26e-06	1.32e-05	15.2% 증가
	변경 후	7.059e-06	3.76e-06	1.877	0.061	-3.13e-07	1.44e-05	
3구간 (500점 미만)	변경 전	3.507e-06	2.91e-06	1.206	0.228	-2.19e-06	9.21e-06	18.6% 감소
	변경 후	2.856e-06	2.9e-06	0.985	0.325	-2.83e-06	8.54e-06	
4구간 (600점 미만)	변경 전	2.351e-06	2.79e-06	0.844	0.399	-3.11e-06	7.81e-06	15.5% 감소
	변경 후	1.987e-06	2.78e-06	0.715	0.474	-3.46e-06	7.43e-06	
5구간 (600점 이상)	변경 전	-3.45e-07	2.97e-06	-0.116	0.907	-6.17e-06	5.48e-06	-
	변경 후	-2.28e-07	2.95e-06	-0.077	0.938	-6.02e-06	5.56e-06	

예제 소개 (2)

- 400점 미만인 고객군 (1, 2구간)에서는 신용한도가 채무불이행에 미치는 영향 증가,
400점 이상 600점 미만 고객군 (3, 4구간)에서는 영향 감소
- 400점 미만인 고객군 대상으로 한도 축소를 제안
- 400점 ~ 500점 중 관계 역전이 일어난 점수 추가 검토 필요
- 한도 축소는 실적 및 대출 승인율 등 요인을 고려하여 구체화 된 수치로 대응 정책 제안

- 회귀분석으로 변수간의 인과 관계를 파악할 수 있다
- 회귀분석시에는 변수간의 관계를 잘 파악하여 정확한 실험 설계가 중요하다
- 회귀분석을 사용하여 변화에 대응하기 위한 정책을 수립할 수 있다

- NICE 평가정보(주) - 신용등급체계공시 - https://www.niceinfo.co.kr/creditrating/cb_info_4_1.nice
- 신용정보법 - <https://www.law.go.kr/%EB%B2%95%EB%A0%B9/%EC%8B%A0%EC%9A%A9%EC%A0%95%EB%B3%B4%EC%9D%98%20%EC%9D%B4%EC%9A%A9%20%EB%B0%8F%20%EB%B3%B4%ED%98%B8%EC%97%90%20%EA%B4%80%ED%95%9C%20%EB%B2%95%EB%A5%A0>
- 일반신용정보관리규약 - https://www.kcredit.or.kr:1441/archive/creditInfoRule.do?menuNo=410&hpBoardSn=CREDIT_RULES
- 서민 · 소상공인에 대한 신속 신용회복지원 - <https://eiec.kdi.re.kr/policy/materialView.do?num=249004>
- MA Iqbal. (2020). [Application of Regression Techniques with their Advantages and Disadvantages](#), Elektron Magazine 4 (1), 11-17
- Michael Bücker, Gero Szepannek, Alicja Gosiewska, Przemysław Biecek. (2021). [Transparency, auditability, and explainability of machine learning models in credit scoring](#), Journal of the Operational Research Society, 73 (1) 2021, s. 70-90.

감사합니다