개념 정리

1. %ACV (All Commodity Volume)

(점포의 크기를 고려한) 제품을 판매하는 점포의 퍼센트.

1. GRP(Gross Ratings Points) = Reach \* Frequency

광고에 노출된 가구의 비율 \* 광고에 노출된 횟수

1. Lift factor: 판촉행사가 판매량을 증가시키는 정도.
2. 모수 해석.

**4-1 logy=α+βx**

**“x가 한 단위 변화 -> y는  100\***β% 변화”

**\*\* x가 더미변수 일 때**

더미 변수는 질적 변수이고, 변수의 차이는 측정할 수 없다. (1로 갔다고 해서 1이 증가한 개념이 아니다!) 이 경우는 e^β−1을 변화율로 따져야 한다. { β < 0.1일 때면, 근사해서 (100\* β)%로 볼 수 있나? }

“x가 1인 경우 -> y는 x가 0인 경우 보다 (e^β−1)% 변화”

**4-2 y=α+βlogx**

“x가 1% 변화 -> y는 0.01\*β 변화”

**4-3 logy=α+βlogx**

“x 1% 변화 -> y β% 변화”

**4-4 y=α+βx**

“x 한 단위 변화 -> y β만큼 변화

1. 도구 변수

오차항과 x의 상관성을 없애서 불편추정량을 만들기 위한 변수.

X와는 상관성이 있지만, 오차항과는 상관성이 없는 변수.

Y = 판매량 X = 가격이라면, 가격과는 상관성이 있고 판매량과는 관련이 없는 변수들

예를 들어 원가, 전날 가격, 다른 시장의 가격 등등…

1. 혼동행렬 (0 = 거짓, 1 = 참)

6-1 Hit Ratio

맞춘 것의 개수. (Tn+Tp)/N

6-2 F1 accuracy measure

Precision : 예측 “1” 중 실제 “1” 비율 (Tp)/(Tp+Fp)

Recall : 실제 “1” 중 예측 “1” 비율 (Tp)/(Tp+Fn)

F1: 2\*P\*R/(P+R) <- P와R의 조화평균.

6-3 민감도와 특이도(Sensitivity & Specificity)

Sensitivity: 실제 “1” 중 예측 “1” 비율 (Tp)/(Tp+Fn)

Specificity: 실제 “0” 중 예측 “0” 비율 (Tn)/(Tn+Fp)

예제 정리

1. 점포 수준의 스캐너 데이터는 분석 목적에 따라 Aggregated 된 데이터이다. 주로 시장 수준의 마케팅 활동 효과 측정과 최적화에 활용힌다. 반응 변수는 주로 양적이며, 가격결정, 프로모션, 광고, 수요예측 분석에 활용한다. 분석 단위 카데고리 브랜드의 판매량, 시장점유율, 분석기법 회귀 분석.cf

개별 소비자 수준의 데이터는 고객 이질성에 초점을 맞춰 고객별로 상이한 행동 패턴을 파악하고 타켓팅에 활용한다. 개별 소비자 수준의 데이터는 sparse한 특징이 있다. 개별 고객의 구매행동에 초점, 고객 이질성 고려 필요, 세 가지 유형의 분석, 구매타이밍, 구매량, 브랜드 선택

1. 교체 비용이 크면 대체가 어려워진다. 대체가 어려워지면 가격탄력성이 낮아진다. 예를 들어 내가 갤럭시폰 사용에 익숙해져 있고 아이폰의 인터페이스를 어렵게 느낀다면, 갤럭시폰의 가격이 높아지더라도 갤럭시를 사용할 것이다.(화는 나겠지만,) 이런 맥락

3-1

3-2 exp(r\*)

3-3 더미 변수인 디스플레이 변수를 제거하면 OLS추정량에 편향이 생긴다. 이는 디스플레이 변수와 가격 변수 사이에 공선성이 있기 때문이다. 디스플레이 변수를 제거하면 오차항에 디스플레이의 효과가 포함되고 오차항과 가격 변수가 상관성을 갖게 된다. 오차항과 독립변수가 상관성을 갖게되면 OLS추정의 기본 가정이 깨지고 추정량에 편향이 생긴다.

편향의 효과는 가격의 디스플레이변수의 회귀계수와 디스플레이 변수와 가격 변수의 상관관계의 방향에 따라 달라진다. 이 경우, 디스플레이를 하면 가격을 하므로 두 변수 사이에 음의 상관관계가 있고, 디스플레이를 하면 판매량이 증가할 것으로 회귀계수는 양수 있다. 회귀계수와 상관계수의 부호가 다르므로, 가격의 모회귀계수보다 더 낮게 추정하는 Downward bias가 발생한다. 즉 가격탄력성을 과대추정한다.

1. 월별 케이블 TV 가격은 거의 일정하기 때문에, 분산이 작고, 설명변수의 분산이 작으면 OLS 추정량의 분산이 커져서 계수 추정의 정확도가 낮아진다. 좀 더 직관적으로 설명하면, X의 변동성으로 Y의 변동성을 설명하는데, X의 변동성이 낮으면 Y를 잘 설명할 수 없다고 할 수 있다.
2. 훈련 오차율이 예측 오차율 보다 압도적으로 낮았다면, overfit을 의심해볼 수 있다. 모델이 overfit된 경우는 모델의 분산이 큰 경우이므로 람다수준을 1.5보다 크게 해서 모델의 유연성을 낮춰 편향을 늘리더라도 분산을 줄어야한다. (과적합 일때는 분산의 감소폭이 편향의 증가폭보다 훨씬 크므로) (bias-variance trade off)

람다를 통해, 모델의 유연성을 조절할 수 있고 모델의 유연성을 통해 분산-편향 절충을 이룰 수 있다. 모델의 유연성이 커질수록 편향은 줄어들지만, 분산은 커진다. 반면 유연성이 작아지면 분산은 작아지만 편향이 증가한다. 따라서 적절한 수준의 유연성 선택이 중요하다. 람다는 유연성의 수준을 선택할 수 있게 하는데, 예를 들어 람다가 0이면 모든 변수들이 수축되지 않고 선택되어 유연성이 매우 높아지고 훈련데이터를 과적합하여 모델의 분산이 커진다.

(참고로 훈련오차율의 크기가 크고 예측오차율과 비슷하면 underfit을 의심해 볼 수 있고 모델의 유연성을 늘려야한다. 유연성을 늘리기 위해서는 더 작은 람다 값을 사용해서 0의 수축하는 정도를 줄여야 한다. )

1. 단기적으로 소비자는 가격할인을 통해, 현재 상품 가격보다 준거 가격이 더 높다고 느낀다. 준거 가격을 높다고 느끼면 상대적으로 이익을 인식하게 되고 가격할인 판매증진에 효과적이다. 그러나 장기적으로는 가격 할인이 끝나고 나면, 준거 가격은 가격할인을 하기 전 보다 낮아지는 경향이 있다. 준거 가격이 낮아지면 소비자들은 상대적으로 손실을 인식하게 되고 같은 가격변화의 폭이라도 이익보다는 손실을 더 크게 느끼는 경향(prospect theory)이 있기 때문에 가격할인은 장기적인 판매증진에 부정적인 영향을 미친다.
2. B. 단기탄력성. B/(1-델타)