

알고리즘 마케팅 4강

2023. 3. 16. (목)

서울과학기술대학교 데이터사이언스학과

김 종 대

오늘의 강의

- 3차시 Review
- 가격 관리
- 유통경로 관리
- 마케팅 커뮤니케이션 관리
- 서비스마케팅
- 인과관계와 실험
- 가설검정
- 회귀분석 기초

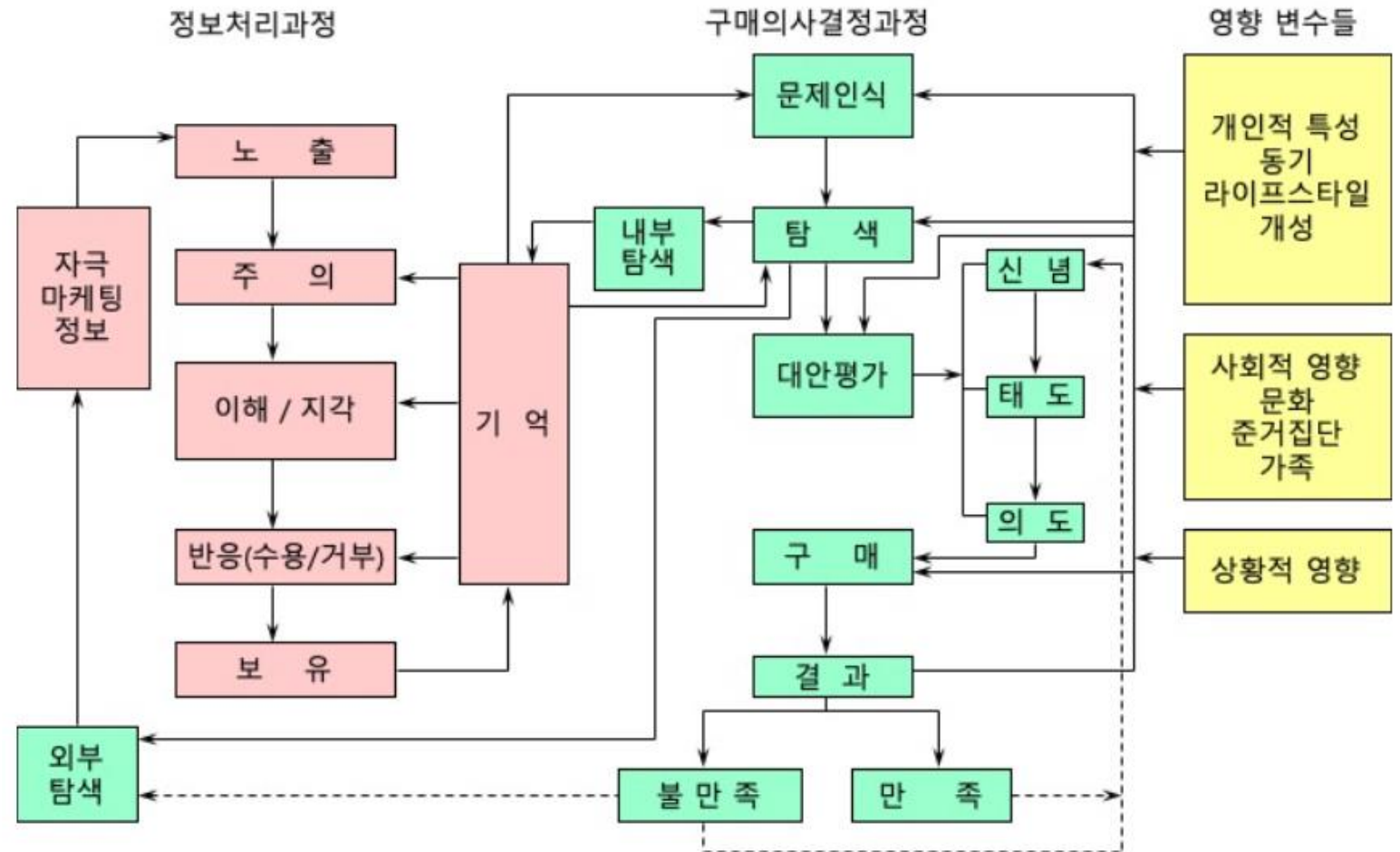
마케팅의 이해

- 마케팅의 주요 개념
 - 4P



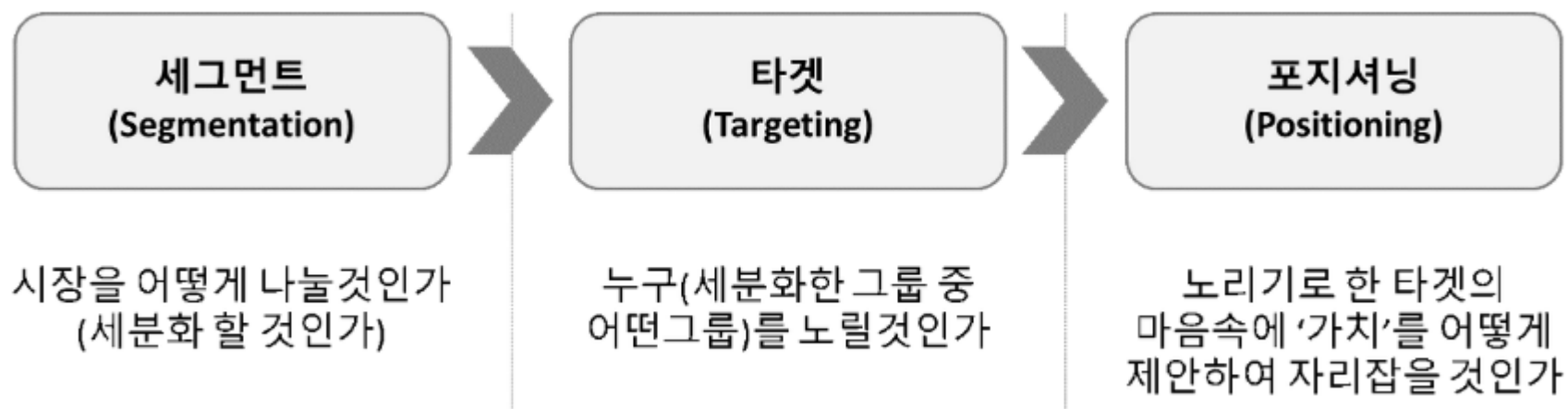
소비자행동 이해

- 소비자행동 개관
 - 소비자행동



시장세분화와 표적시장

- 시장세분화 및 표적시장선정의 개요

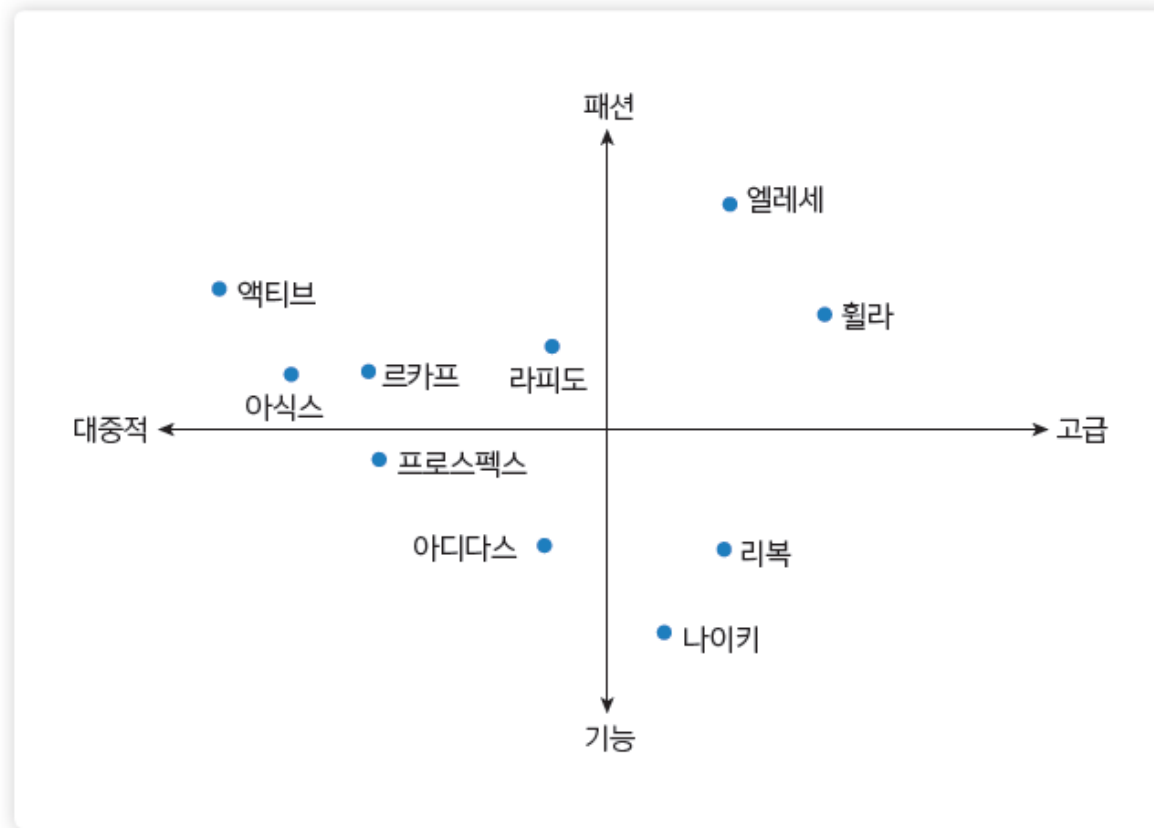


포지셔닝 전략

- 포지셔닝 과정
 - 사례: 지각도의 예

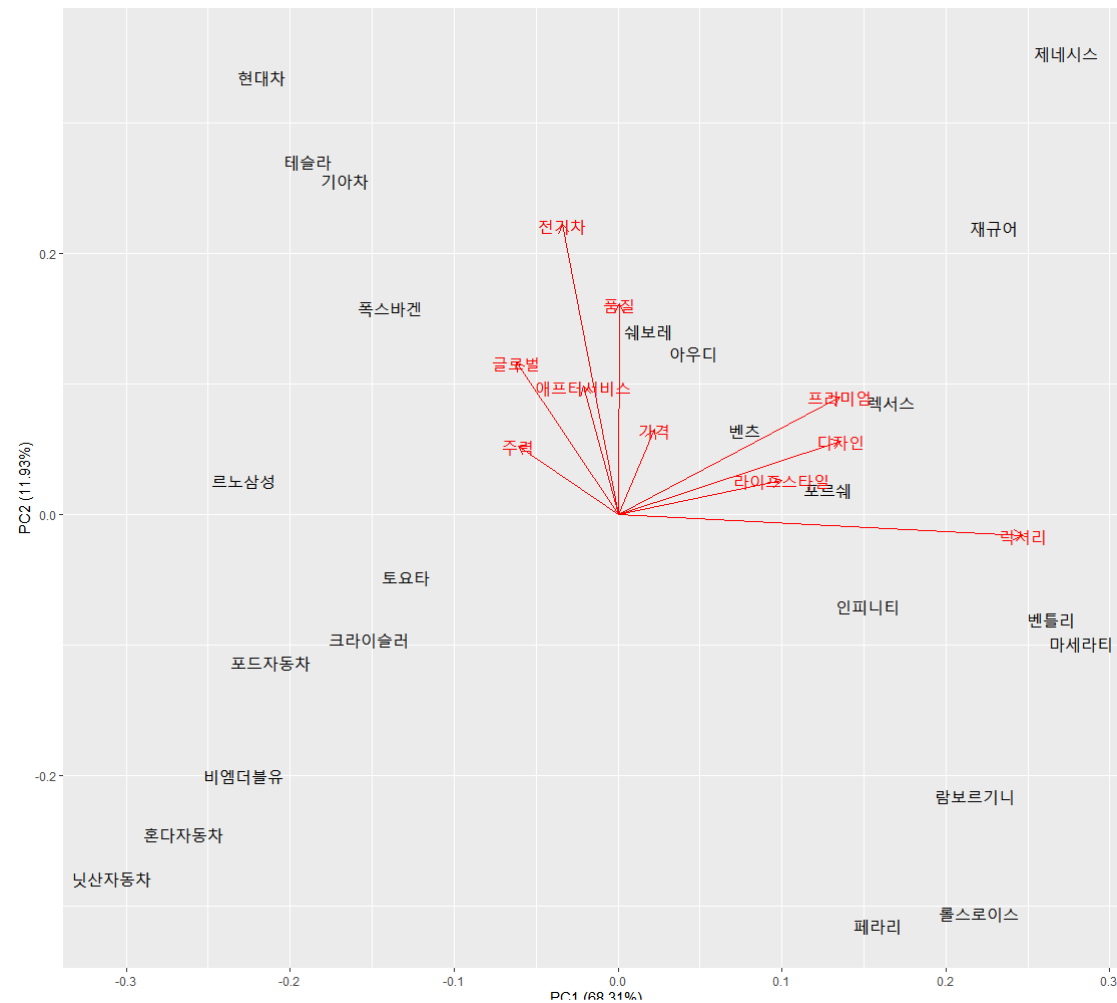
■ 그림 4-4 ■

스포츠화 시장의
지각도



포지셔닝 전략

- 포지셔닝 과정
 - 사례: 국내 자동차 시장 연구



브랜드 관리

- 브랜드 이해

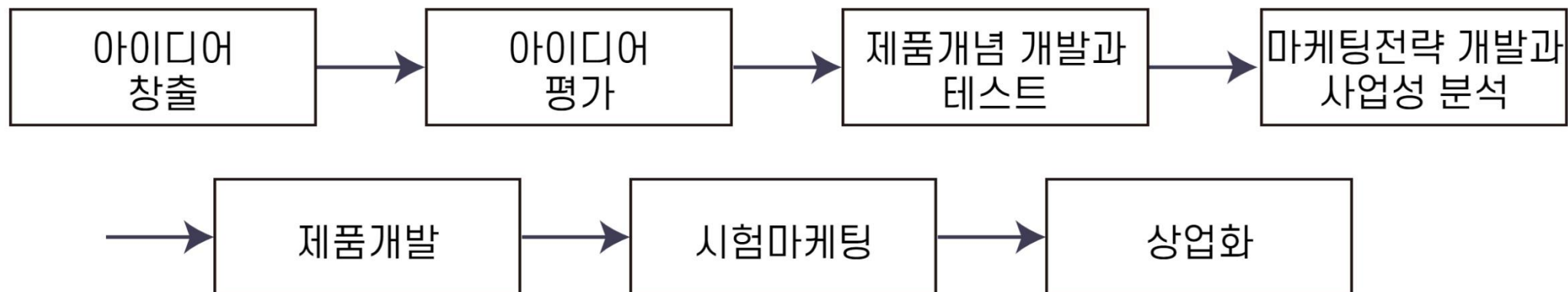
- 브랜드 (brand): “제품을 소비자에게 인지시키고 다른 경쟁자의 제품으로부터 차별화시킬 목적으로 사용하는 명칭, 말, 기호, 상징 또는 이들의 결합체”
- 브랜드 자산 (brand equity): “어떤 브랜드에 대해 축적된 지식 (knowledge)이 그 브랜드의 마케팅에 대한 소비자 반응을 차등화시키는 효과”
 - “해당 브랜드가 없을 때에 비해 높아진 고객의 브랜드 충성도 증가분, 시장점유율 증가분, 수익 증가분”
- 브랜드 가치 평가: 브랜드의 자산적 가치를 추정하는 과정

The Amazon logo, featuring the word "amazon" in a black, lowercase, sans-serif font, with a curved orange arrow underneath it pointing from the letter 'a' to the letter 'z'.The Samsung logo, consisting of the word "SAMSUNG" in a bold, blue, uppercase, sans-serif font.The Google logo, featuring the word "Google" in its multi-colored, lowercase, sans-serif font.

신제품 개발과 관리

- 신제품 개발 과정

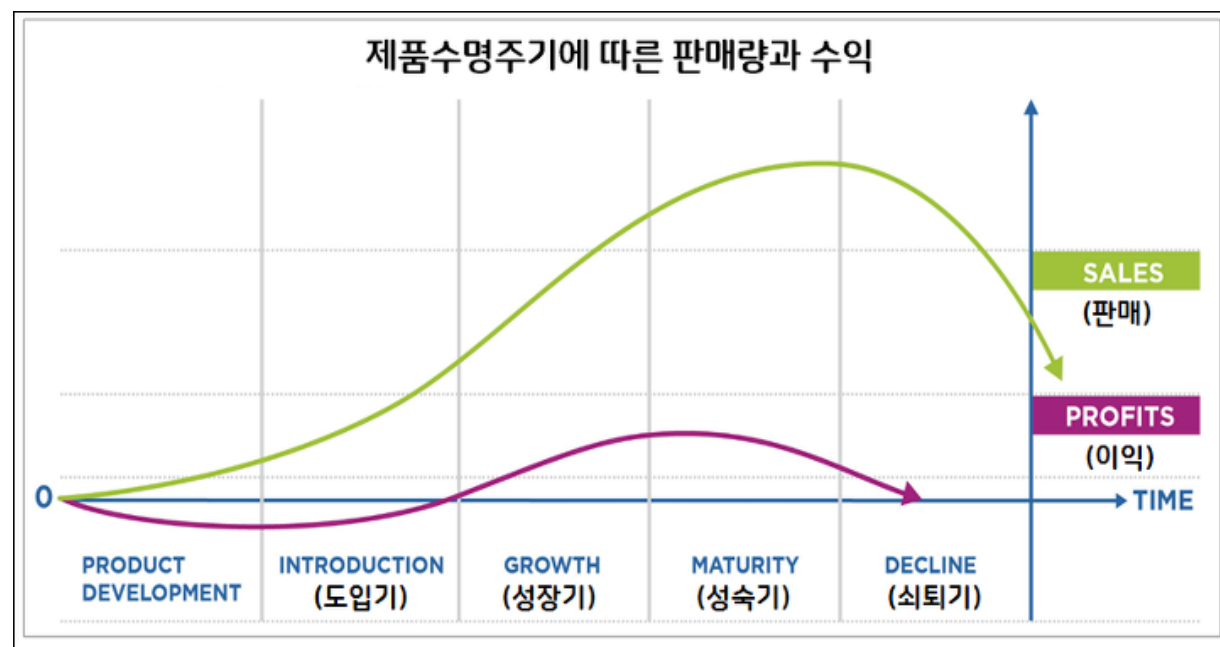
- 진행 단계별로 면밀한 조사와 체계적 검토를 거쳐 개발을 진행



신제품 개발과 관리

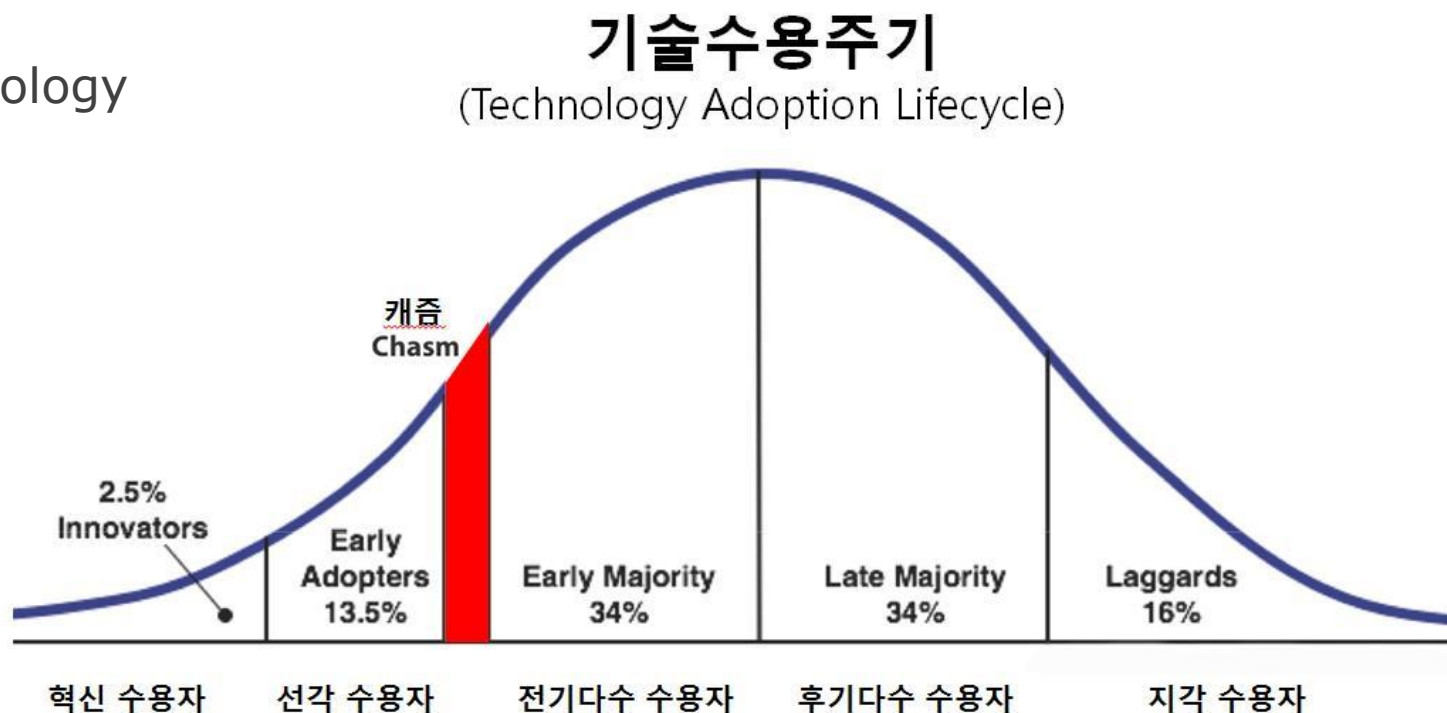
- 제품수명주기 관리

- 제품수명주기(product life cycle): 제품 출시부터 시장 퇴출까지의 과정을 시간별 매출과 수익의 변화로서 집약한 것
 - 시장의 반응이 좋으면 매출이 증가
 - 그러나 시장 내 소비자의 수요가 한정되어 있기 때문에 어느 수준부터 정체 및 감소



신제품 개발과 관리

- 신제품 수용시기와 수용자
 - Rogers의 기술수용주기(Technology Adaption Life-Cycle)



가격관리

- 가격의 이해

- 가격(price): “소비자가 제품이나 서비스의 소유나 사용으로 얻는 편익의 대가로 내놓아야 하는 반대급부의 합계”
- 마케팅 믹스 중 유일하게 수익을 창출하는 변수
- 기업의 재무적 성과에 직접적이고 막대한 영향
- 다른 마케팅 믹스들에 비해 비교적 유연하게 변경 가능



가격관리

- 기본적 가격의사결정

- 고객가치(customer value)

- 총고객가치 = 고객의 비용 + 순고객가치(net value)

- 가격탄력성(price elasticity)

- 비탄력적(inelastic): 가격 변화에 비해 수요량 변화가 작을 때

- 예: 필수재 / 대체재가 적을 경우

- 탄력적(elastic): 가격 변화에 비해 수요량 변화가 클 때

- 예: 사치재 / 대체재가 많을 경우

$$\begin{aligned}\epsilon &= \frac{\text{수요량 변화율}}{\text{가격 변화율}} \\ &= \frac{\Delta Q/Q}{\Delta P/P} \\ &= \frac{\Delta Q}{\Delta P} \times \frac{P}{Q}\end{aligned}$$

가격관리

- 기본적 가격의사결정
 - 고객가치중심 가격결정
 - 원가중심 가격결정
 - 원가가산 가격결정
 - 목표이익 가격결정
 - 경쟁자 중심 가격결정

가격관리

- 신제품 가격의사결정

- 시장침투가격정책(penetration pricing): 신제품 도입 초기부터 저가격을 고수, 시장에 빠르게 침투하여 시장점유율을 빠르게 끌어올리는 가격결정방식
 - 대다수 소비자가 가격에 민감할 때 효과적
 - 잠재 경쟁자의 시장 진입을 막는 효과도 존재
- 초기고가가격정책(skimming pricing): 신제품 도입 초기에 혁신제품을 구입하기 위해 얼마든 지불하려는 고객들로 구성된 세분시장을 타겟으로 삼아 고가격을 매겨 수익을 극대화
 - 제품의 품질, 디자인 등이 매우 우수해야 효과적
- 중립가격정책(neutral pricing): 신제품 도입시 합리적 또는 합당한 가격을 매기는 방식

가격관리

- 가격다변화 전략

- 고객별 가격다변화

- 예: 놀이공원 등은 어린이, 노인에게 낮은 가격을 제시
 - 예: 구매자의 노력에 따른 할인혜택 부여(쿠폰 발급)

- 입지별 가격다변화

- 예: 공연장, 야구장, 극장 좌석 위치에 따른 차별화

- 시간별 가격다변화

- 예: 성수기와 비수기의 항공기 티켓 가격 차별화



가격관리

- 가격 조정

- 심리적 가격조정

- 단수가격: 100,000원 vs. 99,900원
 - 준거가격: 소비자의 비교 기준이 되는 가격

- 프로모션 가격조정: 소비자의 구매유발효과를 강화하기 위한 가격 전략
 - 예: 할인, 리스/렌탈 등

ONLY ONLINE 6월 29일 9시

999특가
TIME SALE

9시를 잡아라!
단 하루 진행되는 선착순 한정특가 세일!

선착순 10명
~~268,000원~~
99,900원
시그니처 플러스
(올리브 블라썸 선착순 한정)

유통경로 관리

- 유통경로의 의의와 중요성

- 유통 경로(marketing channel; distribution channel)
 - “마케터가 제품이나 서비스를 소비자에게 전달하여 소비되도록 하는 과정에 참여하는 개인이나 조직의 집합”
 - 중간상: “생산된 제품이 소비자로 전달되는 과정에서 해당 제품의 판매 및 구매에 직접적으로 관련된 서비스를 제공하는 도매상, 소매상 등의 사업체”
- 유통경로 관리의 중요성
 - 경쟁자가 단기간에 모방하지 못한다는 차별화 요소
 - 전략적 관리를 통해 유통비용 감축 가능

유통경로 관리

- 유통경로의 의의와 중요성

- 사례: 쇼루밍족

- 구경은 오프라인 매장에서 하고 실제 구매는 온라인/모바일 경로로 하는 고객
 - 오프라인 매장 직원의 “공짜 노동” 문제 등이 발생
 - 역쇼루밍족: 반대로 온라인을 통해 제품 정보를 꼼꼼하게 확인한 후 실제 구매는 오프라인 매장에서 하는 고객



마케팅 커뮤니케이션 관리

- 마케팅 커뮤니케이션의 이해
 - 마케팅 커뮤니케이션(marketing communication)
 - 마케팅 과정에서 전개되는 커뮤니케이션; 촉진(promotion)
 - 마케팅 목적을 달성하기 위해 발신자(마케터)와 수신자(고객)가 상호작용(메시지 전달 및 피드백)하는 영역 또는 그 총체적인 과정
 - 메시지: 발신자가 수신자에게 전달하는 상징들의 집합 (예: 광고)
 - 매체: 메시지가 전달되는 커뮤니케이션 채널 (예: 대면, 전화, 메일, 인터넷, 구전 등)



마케팅 커뮤니케이션 관리

- IMC의 이해

- Integrated Marketing Communications: “광고, PR, 판매 촉진, 인적 판매 등 다양한 커뮤니케이션 기법들의 직접적 역할과 제품, 가격, 유통경로 등의 간접적 커뮤니케이션 역할을 전체적으로 고려하여 표적시장을 대상으로 전개하는 종합적 마케팅 커뮤니케이션 과정”
 - 마케터의 일방적인 메시지 전달이 아닌, 고객과의 상호작용을 강조하는 개념
 - 관계마케팅(relationship marketing)
 - 브랜드 관리(brand management)
 - 여러 커뮤니케이션 채널을 독립적으로 전개하는 것이 아니라 통합하여 관리

마케팅 커뮤니케이션 관리

- 광고의 개념과 기능

- 광고: “제품, 서비스 또는 아이디어를 제시하거나 촉진하기 위해 마케터가 일체의 비용을 지불하고 전개하는 비대면적 커뮤니케이션”
- 정보전달기능: 표적 청중이 인지하고 지식을 형성할 수 있도록 촉진
 - 예: 가격 인하 광고
- 설득기능: 표적 청중이 호감을 형성하고 제품을 선호하도록 설계
 - 예: 고급 승용차로서의 포지셔닝 광고
- 상기기능: 표적 청중이 스스로 도달한 구매준비단계를 다시금 환기시키는 기능



마케팅 커뮤니케이션 관리

- 광고 매체의 유형과 선택

- 유형: TV, 라디오, 신문, 인터넷, 모바일, 옥외(전광판 등), 일대일(전화, 카탈로그, DM 등)
- 매체 유형 선정의 고려 요인: 매체의 효과를 측정
 - 도달범위(Reach): 일정기간 동안 광고에 한 번 이상 노출된 표적청중의 수나 비율
 - 노출횟수(Frequency): 일정기간 동안 표적청중에게 노출된 평균횟수
 - 총도달률(GRPs; Gross Rating Points): 도달범위 X 노출횟수
 - 청중의 품질수준: 매체의 주된 청중이 마케팅의 표적청중과 얼마나 합치하는가?
 - 청중의 주의수준: 매체의 주된 청중이 게재되는 광고에 얼마나 주의를 많이 기울이는가?

마케팅 커뮤니케이션 관리

- 광고 매체의 유형과 선택
 - 사례: 실제 광고 효과 대비 고평가 매체 조사 결과

[도표7] 실제 광고 효과 대비 고평가 매체

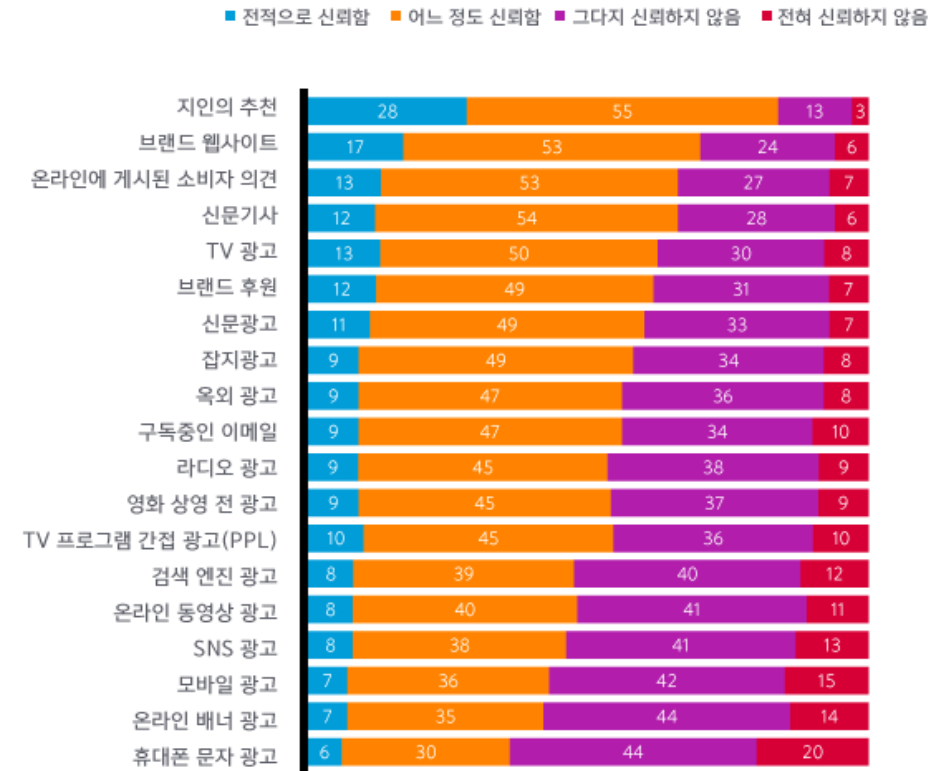


마케팅 커뮤니케이션 관리

■ 광고 매체의 유형과 선택

■ 사례: 매체 유형에 따른 광고의 이미지

다음 유형의 광고를 얼마나 신뢰하십니까?



출처: Nielsen Global Trust in Advertising Report Q1, 2015

마케팅 커뮤니케이션 관리

- 광고 성과의 측정
 - 커뮤니케이션 효과의 측정
 - 브랜드 인지도 제고 효과의 측정
 - 소비자의 태도 및 행동의 변화 측정
 - 브랜드 이미지의 변화 측정
 - 재무 효과의 측정: 매출액, 시장점유율 등의 변화 관찰
- 광고의 효과를 (계량적으로) 어떻게 측정할 것인가?
 - 예: 온라인 광고의 전환율(conversion rate)

마케팅 커뮤니케이션 관리

- PR의 개념과 유형

- PR(Public Relations): “이미지 쇄신, 호의적 평판의 유도, 비호의적 평판의 완화 등을 목적으로 기업과 직간접적으로 관련된 여러 유형의 집단들과 좋은 관계를 구축하고 유지해 나가기 위해 벌이는 비대면적 커뮤니케이션 활동”
 - 예: 자동차 론칭쇼 등 대대적 이벤트, 사건사고에 대한 해명 등 입장표명, 구전 관리
- 홍보(publicity): 언론매체를 통하여 벌이는 비대면적 커뮤니케이션 활동
 - 예: 대언론 보도자료 배포, 기자간담회, 언론인터뷰 등
- 후원(sponsorship): 외부의 개인이나 조직에 대한 직간접적 지원을 통해 바람직한 공중관계 조성
 - 예: 스포츠, 문화 등에 관한 이벤트성 후원활동

마케팅 커뮤니케이션 관리

■ 판매촉진 관리

- 판매촉진(sales promotion): “표적고객의 행동을 촉발할 목적으로 전개하는 인센티브 위주의 마케팅 커뮤니케이션 활동”
 - 마일리지 제도: 단골고객에 대한 보상으로 사용량, 빈도에 따라 혜택 제공
 - 쿠폰: 정해진 기간에 기재된 조건만큼 할인 판매한다는 약속을 명시한 각종 증빙자료
 - 할인: 기본가격을 일시적으로 낮추는 판매촉진 전략
 - 잦은 할인은 소비자들의 지각된 준거가격 자체를 낮추는 역효과

모든 상품에 사용 가능한
20만원 할인쿠폰팩

100,000원 200만원 이상 결제 시	50,000원 100만원 이상 결제 시
30,000원 60만원 이상 결제 시	5% 최대 2만원 할인

서비스마케팅

- 서비스마케팅의 의의

- 서비스(Service): “판매 목적으로 제공되거나 제품 판매와 연계하여 제공되는 모든 활동, 편익, 만족”
 - 무형성(intangibility)
 - 비분리성(inseparability): 생산과 소비가 동시에 이루어지는 경향
 - 이질성(heterogeneity): 서비스 제공자와 상황에 따라 매우 다양하게 전개
 - 소멸성(perishability): 제품과 달리 재고의 형태로 보관하는 것이 거의 불가능
- 제품(product)와 근본적으로 다르기 때문에 다른 관점에서 마케팅 관리 필요

서비스마케팅

- 서비스 관리

- 7P의 관점에서 마케팅 믹스를 설계하고 추진
- 5. 물리적 증거(physical evidence): “서비스가 전달되고 서비스 조직과 고객의 상호작용이 이루어지는 실체적 환경”
 - 물리적 환경(예: 주차장), 유형적 요소(예: 종업원의 유니폼)
- 6. 프로세스(process): “서비스가 전달되는 절차나 활동의 흐름”
 - 대부분의 서비스는 결과물이라기보다는 일련의 과정이자 흐름
 - 고객은 서비스를 제공받는 총체적인 경험을 중시
- 7. 사람(people): 서비스 종업원의 관리



서비스마케팅

- 내부마케팅

- 서비스 경쟁에서 종업원은 그 자체가 서비스이며, 고객의 눈에 비치는 기업 그 자체
- 내부고객(internal product): 조직의 종업원
 - 종업원을 최초의 고객으로 보고 그들에게 서비스 마인드나 고객지향적 사고를 심어주며 더 좋은 서비스 성과를 낼 수 있도록 동기부여하는 활동
 - 외부고객과 내부고객의 상호작용을 중시하는 관점
- 주요 내용
 - 서비스 문화의 확립과 유지
 - 내부 브랜딩: 내부고객의 브랜드 내재화 활동

인과관계

- 인과관계(causality)의 개념
 - “어떠한 변수(원인변수)가 다른 변수(결과변수)의 값에 영향을 미치는 관계”
 - “X가 원인이 되어 결과 Y가 나타난다는 의미”
- 대부분의 연구들은 복잡한 변수들의 구성과 다양한 인과관계를 가설로 설정하고 있기 때문에 이러한 인과관계를 확률적인 가능성으로 표현하는 것이 필요

인과관계

- 인과관계의 필요조건 (NOT 충분조건)
 - 조건 1. 동시에 발생하는 변동(concomitant variation)
 - 원인변수와 결과변수 간의 인과관계가 자료분석을 통해 얻은 관계와 일치해야 한다.
 - 조건 2. 원인변수가 아닌 다른 변수에 의한 변동(variation due to other possible causal factor)
 - 연구자가 미처 생각하지 못했던 원인변수가 존재하며, 이에 따라 우연히 실제로 존재하지 않는 인과관계가 존재하는 것 같이 보일 수 있다는 가능성을 고려해야 한다.
 - 조건 3. 시간 순서에 따른 변동(time order of occurrence)
 - 원인변수의 값이 변화하였다면, 변화한 시점이나 그 이후에 결과변수의 값이 변화해야 한다.

인과관계

- 인과관계의 필요조건 (NOT 충분조건)
 - 조건 1. 동시에 발생하는 변동(concomitant variation)
 - 원인변수와 결과변수 간의 인과관계가 자료분석을 통해 얻은 관계와 일치해야 한다.
 - 주의사항: 인과관계가 성립하기 위한 필요조건이지 충분조건이 아니다.
 - 인과관계가 아니라 단순한 상관관계일 가능성
 - 독립 관계(independent relationship): 두 변수 사이에 어떠한 관계도 존재하지 않는 것으로, X의 수준이 Y의 수준에 전혀 영향을 미치지 않는 상태
 - X에 따른 Y의 확률 분포가 서로 동일한 상태

인과관계

- 인과관계의 필요조건 (NOT 충분조건)
 - 조건 2. 원인변수가 아닌 다른 변수에 의한 변동(variation due to other possible causal factor)
 - 연구자가 미처 생각하지 못했던 원인변수가 존재하며, 이에 따라 우연히 실제로 존재하지 않는 인과관계가 존재하는 것 같이 보일 수 있다는 가능성을 고려해야 한다.
 - 예: 광고비(X) \rightarrow 매출액(Y)
 - 다른 고려 가능한 원인변수들: 소비자의 실질 소득 증가, 제품의 품질 향상, 가격 경쟁력 향상, 경쟁회사 또는 제품의 상대적인 부진, 유통구조의 효율화, 서비스 종업원의 친절도 향상
 - 이러한 변수들을 충분히 통제된 상태에서 분석이 진행되어야 한다.

인과관계

- 인과관계의 필요조건 (NOT 충분조건)
 - 사례: Simpson's paradox (p.209)
 - 연구자가 미처 생각하지 못한 원인변수의 효과와 표본선택 바이어스가 결합하여 결과가 왜곡

	제품 A 구매	제품 A 비구매	합 계	구매비율
교육수준 높음	20	20	40	0.5
교육수준 낮음	16	24	40	0.4
합계	36	44	80	

인과관계

<고소득층>	제품 A 구매	제품 A 비구매	합 계	구매비율
교육수준 높음	18	12	30	0.6
교육수준 낮음	7	3	10	0.7
합계	25	15	40	

<저소득층>	제품 A 구매	제품 A 비구매	합 계	구매비율
교육수준 높음	2	8	10	0.2
교육수준 낮음	9	21	30	0.3
합계	11	29	40	

인과관계

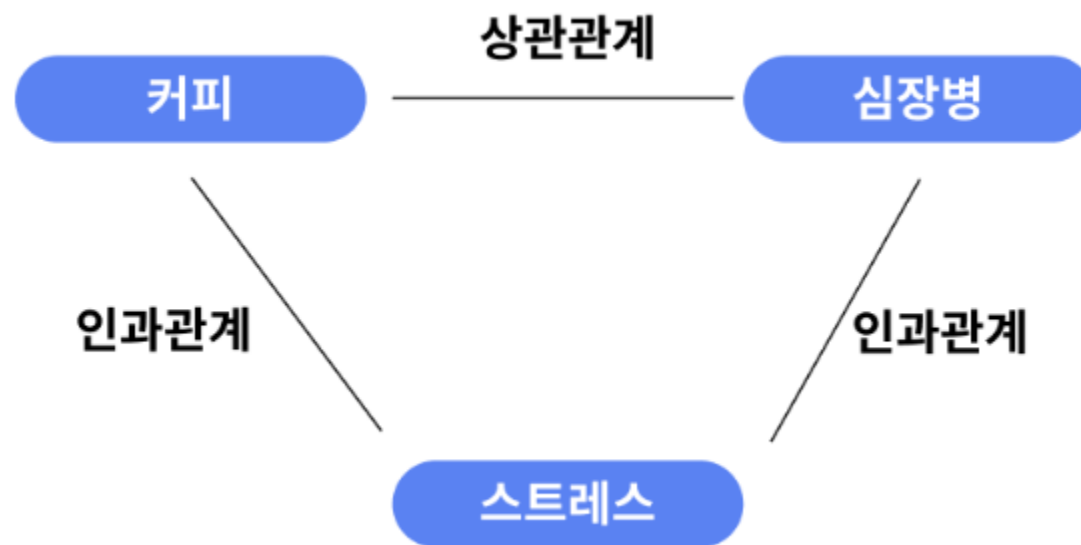
- 인과관계의 필요조건 (NOT 충분조건)
 - 사례: Simpson's paradox (p.209)
 - 소득수준의 영향력을 연구자가 미처 생각하지 못했더라도, 랜덤화 과정을 통해 표본을 잘 선택하였다면(즉, 표본선택 바이어스가 없었다면), 소득수준이 골고루 분포되어 다음과 같은 결과가 나와야 한다.
 - 교육수준이 높은 경우: 고소득층과 저소득층의 평균 0.4
 - 교육수준이 낮은 경우: 고소득층과 저소득층의 평균 0.5
 - 그러나 실제 분석 결과는 반대의 결과가 도출
 - 이는 표본에 가중치가 부여되었기 때문: 교육수준이 높은 경우 75%가 고소득층

인과관계

- 인과관계의 필요조건 (NOT 충분조건)
 - 조건 3. 시간 순서에 따른 변동(time order of occurrence)
 - 원인변수의 값이 변화하였다면, 변화한 시점이나 그 이후에 결과변수의 값이 변화해야 한다.
 - 만약 결과변수가 원인변수가 변하기 전에 미리 움직이게 되면 이를 인과관계라고 할 수 없다.
 - 예: 광고비(X) → 매출액(Y)
 - 광고비 증가 이후 매출액이 증가된 상황에서, 광고비를 일정 기간 동안 증가시키지 않고 일정 수준을 유지하였는데도 매출액이 계속 증가하였다면?
 - “광고비”와 “매출액”의 인과관계를 재검토할 필요성 → “광고비”가 아닌 “브랜드 인지도”가 정확한 원인변수일 가능성

인과관계

- 인과관계의 필요조건 (NOT 충분조건)
 - 사례: “커피를 마시면 심장병 발병 확률이 감소한다?”



인과관계

- 인과관계에 관한 주요 용어
 - 독립변수(independent variable), 설명변수(explanatory variable), 예측변수(predictor)
 - 종속변수(dependent variable), 반응변수(response variable)
 - 혼란변수(confounding variable): 종속변수에 영향을 미치는 독립변수를 제외한 모든 다른 변수
 - 종속변수의 결과에 영향을 주기 때문에 연구결과를 왜곡
 - “원인변수가 아닌 다른 변수에 의한 변동”은 혼란변수의 존재 가능성을 검토하는 과정
 - 실험(experiment): 연구자가 혼란변수를 통제된 상태에서 독립변수를 조작하면서 종속변수의 값이 어떻게 변하는지를 관찰하는 과정

인과관계

- 인과관계에 관한 주요 용어
 - 실험설계(experimental design):
 - (1) 실험 단위(연구 대상이 되는 개체)를 어떻게 정할 것인지?
 - 예: 소비자 단위, 기업 단위, 부서 단위 등
 - (2) 실험 단위를 비슷한 집단으로 어떻게 구분할 것인지?
 - 예: 소비자의 소득 수준, 교육 수준 등 다른 여건이 고르게 섞여있는 집단으로 구분
 - (3) 독립변수를 어떻게 정의하고 조작할 것인지?
 - (4) 종속변수를 어떻게 측정할 것인지?
 - (5) 혼란변수를 어떻게 통제할 것인지?

인과관계

- 내적 타당성(internal validity)
 - 연구가 연구 내적으로 제대로 진행, 분석된 것이 맞는가?
 - 연구 결과에 영향을 주는 다른 요인이 있는지 검토
 - 독립변수를 제외한 다른 변수들, 즉, 혼란변수가 종속변수에 어떠한 영향을 미치는지 검토
- 1. 특정사건의 영향(history effect)
 - 연구 기간 동안 종속변수에 영향을 주는 특정한 사건이 발생하여 환경이 바뀌었을 경우
 - 예: COVID-19로 인한 마케팅/심리학 연구의 어려움
 - 실험실 연구가 아닌 현장 연구(field experiment)에서 더 심각

인과관계

- 내적 타당성(internal validity)
 - 2. 시간에 따른 성숙 효과(maturation effect)
 - 시간이 지남에 따라 자연적으로 연구 참여자의 특성이 변화하여 결과에 영향을 미치는 효과
 - 예: 실험 응답자가 시간이 지남에 따라 피로해하거나 지루해하는 경우
 - 3. 학습 효과(learning effect; carry-over effect)
 - 처음 측정한 값의 영향을 받아 두 번째 측정값이 다르게 나타나는 효과
 - 결과변수의 변화가 원인변수의 변화에 의한 것인지 학습효과에 의한 것인지 구분하기 어려움.

인과관계

- 내적 타당성(internal validity)
 - 4. 측정도구의 변화
 - 5. 회귀효과
 - 첫 번째 측정에서 매우 높은 점수 또는 매우 낮은 점수를 매긴 참여자가 두 번째 측정에서는 평균으로 되돌아가고자 하는 회귀 경향을 가진다.
 - 6. 표본선택 바이어스(selection bias)
 - 예: 두 그룹의 비교 연구에서 실험그룹과 대조그룹 사이에 이미 차이가 존재하는 경우
 - 7. 연구참여자의 탈락
 - 8. 변수들 사이의 예상치 못한 상호작용

인과관계

- 혼란변수의 통제
 - 연구에 관심 대상이 되는 처리효과(treatment effect)에 혼란변수의 효과가 섞이지 않도록 구분
 - 즉, 어떤 효과가 발견되면, 이러한 효과가 처리효과에 의한 것인지 또는 혼란변수에 의한 효과인지 구분할 수 있는 실험설계를 하는 것
- 랜덤화(randomization)
- 매칭(matching)
- 통계분석에 의한 보정: 공분산분석
- 실험설계에 의한 통제

인과관계

- 혼란변수의 통제
 - 랜덤화(randomization)
 - 난수표(random number)를 이용해 실험그룹이나 대조그룹으로 할당
 - 연구자가 미처 생각하지 못한 혼란변수가 존재하는 경우라도 이에 대한 효과를 희석
 - 발견되지 않은 체계적인 차이가 완벽하게 제거된다기보다는, 왜곡 가능성을 확률적으로 줄여주는 방법
 - 오차(error)의 증가: 혼란변수에 대한 효과를 처리효과와 분리시켜 오차에 포함
 - 실험에서 오차가 크다는 것은 혼란변수의 효과가 커서 실험의 정확도가 떨어질 수도 있다는 가능성을 제기하는 것

인과관계

- 혼란변수의 통제
 - 랜덤화(randomization)
 - 랜덤 샘플링(random sampling)과의 비교
 - 랜덤화: 표본을 설정한 후에 선택된 실험 단위를 어떻게 실험그룹과 대조그룹으로 나눌 것인가의 문제 → 그룹 간 차이가 나지 않게 하는 것
 - 랜덤 샘플링: 모집단으로부터 표본을 선정하는 것에 초점 → 그룹 간 차이보다는 모집단의 특성을 잘 대표할 수 있도록 표본을 선정하는 것이 목표

인과관계

- 외적 타당성(external validity)
 - 연구 결과를 일반화(generalization)할 수 있는가?
 - 다른 그룹이나 장소, 시점에 적용하더라도 동일한 결과가 도출되는가?
 - 내적 타당성이 없는 결과에 대해 외적 타당성을 검토하는 것은 무의미
- 1. 모집단에 대한 타당성
 - 연구 분석에 사용되는 표본자료가 연구 대상이 되는 모집단의 특성을 충분히 반영하고 있는가?
 - 랜덤 샘플링에 관한 문제
 - 예: 선거 관련 여론조사

인과관계

- 외적 타당성(external validity)
 - 2. 환경에 의한 타당성
 - 연구 결과가 연구 환경이 아닌 좀더 현실적인 다른 환경에도 적용될 수 있는가?
 - 예: 신입사원을 대상으로 한 영어회화 프로그램 성과의 연구 결과가 다른 직원들에게도 확장될 수 있는가? → 신입사원이기 때문에 더 열심히 한 것이라면?
 - 호선효과(Hawthorne effect): 연구참여자가 특별한 대우를 받으며 지속적으로 관찰되고 있다는 생각 때문에 평상시와는 다르게 열심히 연구에 참여하여 나타나는 효과
 - 예: 밝은 작업환경에서 작업한 그룹과 어두운 작업 환경에서 작업한 그룹 모두에서 생산성이 향상되었을 경우

가설검정

- 가설검정의 원리

- 기본적으로 모집단을 대표하는 표본의 “표본평균”에 대한 분포를 파악하여 문제를 해결
- 어떠한 가설을 참이라고 가정하고, 일어날 가능성이 희박한 표본평균 값에 대한 수준을 정해 놓고, 표본 자료를 통하여 계산된 표본평균이 그 수준을 벗어나면, 참이라고 가정한 그 가설을 기각하게 되는 원칙

- 절차

- (1) 귀무가설 및 대립가설 설정과 유의 수준 결정 → (2) 검정통계량 결정 → (3) 기각역 결정 → (4) 검정통계량의 계산 → (5) 통계적 의사결정

가설검정

- 가설검정의 절차
 - (1) 귀무가설 및 대립가설 설정과 유의 수준 결정
 - 연구가설을 통계적 가설로서 구체화: 모집단의 특성을 나타내는 모수(parameter)를 이용하여 표현
 - 귀무가설(null hypothesis) \leftrightarrow 대립가설(alternative hypothesis)
 - 일반적으로 연구의 목표는 귀무가설(예: 효과가 없다)을 기각하는 것

가설검정

- 가설검정의 절차
 - (1) 귀무가설 및 대립가설 설정과 유의 수준 결정
 - 제1종 오류(Type 1 error): True인 귀무가설을 기각하는 오류
 - 제2종 오류(Type 2 error): False인 귀무가설을 기각하지 않고 받아들이는 오류

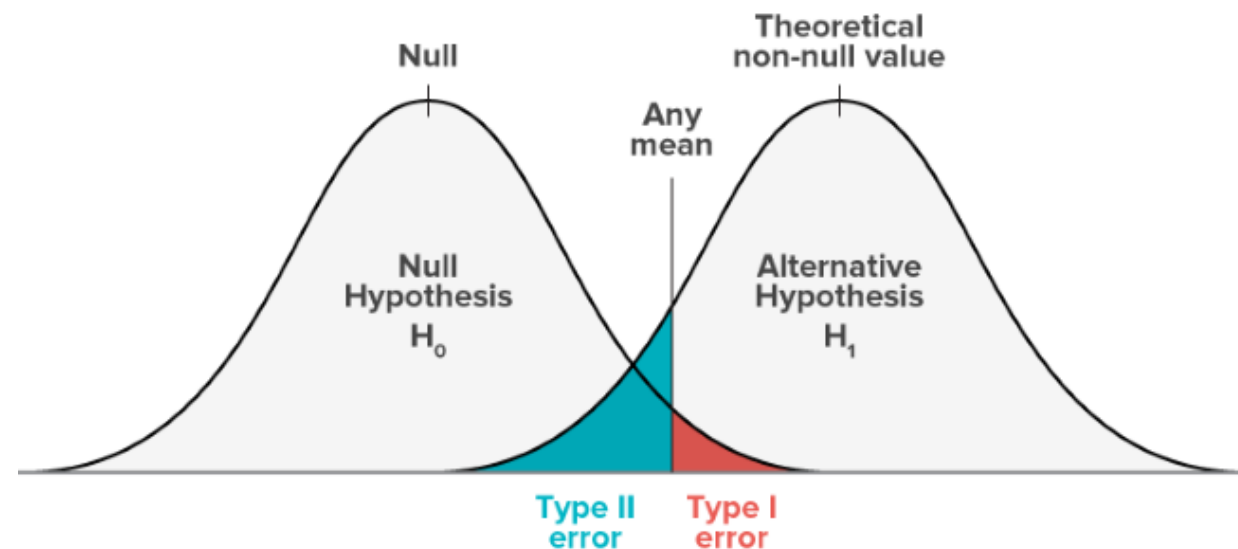
통계적 결정 \ 실제상황	H_0 가 사실 (참)	H_0 가 허위 (거짓)
	H_0 채택	H_0 기각
H_0 채택	옳은결정 확률 = $1 - \alpha$	제II종 오류 확률 = β
H_0 기각	제I종 오류 확률 = α	옳은결정 확률 = $1 - \beta$

유의수준

검정력

가설검정

- 가설검정의 절차
 - (1) 귀무가설 및 대립가설 설정과 유의수준 결정
 - 유의수준: 제1종 오류를 허용하는 최대 확률 (α)
 - 일어날 가능성이 희박하다고 생각하는 확률 수준으로, 귀무가설을 기각하는 기준
 - 가설검정 이전에 연구자가 유의수준을 미리 결정



가설검정

- 가설검정의 절차

- (2) 검정통계량 결정

- 검정통계량: 수집된 자료로부터 계산된 통계량
 - 검정통계량 분포: 귀무가설이 참이라는 가정 하의 표본이 따르는 확률 분포 (예: 표준정규분포, t 분포)

- (3) 기각역 결정

- (4) 검정통계량의 계산

- (5) 통계적 의사결정



가설검정

- P-value를 이용한 가설검정
 - P-value: 검정통계량 값이 귀무가설의 가정으로부터 얼마만큼 벗어나는지의 정도
 - 귀무가설이 True라고 가정하였을 때, 표본을 통해 계산된 검정통계량 값보다 귀무가설을 기각하는 방향으로 더 심하게 검정통계량 값이 관측될 확률
 - 예: 검정통계량 Z 는 0으로부터 멀리 떨어져 있을수록 (Z 의 절대값이 클수록) 귀무가설을 기각하는 방향에 가까워지는 것
 - P-value가 사전에 정한 유의수준보다 작을 경우, 검정통계량이 기각역에 위치 → 가설을 기각
 - 즉, 연구자는 p-value와 유의수준을 단순 비교함으로써 귀무가설 기각 여부를 판단 가능

상관관계분석

- 상관관계분석: 두 변수 X, Y 의 연관도, 즉, 선형관계에 대한 정도

- 공분산(covariance)

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

- 변수 간 선형관계의 방향과 유무를 판단할 수 있으나, 선형관계의 정도는 파악할 수 없음.
 - 예: 측정단위에 따른 공분산 값의 변화

- 상관계수(correlation coefficient)

$$Corr(X, Y) = \rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

- X, Y 의 표준편차로 나누어 준 형태로서 언제나 $-1 \leq \rho(X, Y) \leq 1$

상관관계분석

- 상관관계분석: 두 변수 X, Y 의 연관도, 즉, 선형관계에 대한 정도
 - 상관관계의 계수가 절대값 1이라면?
 - X, Y 는 완벽한 선형관계를 가지고 있는 것 $\rightarrow Y = aX + b$ 꼴로 표현 가능 (100% 예측 가능)
- 공분산 또는 상관관계가 0이라면 X, Y 는 무조건 독립인가?
 - 예: X, Y 가 비선형관계일 경우
 - 즉, X, Y 가 서로 독립이면 $Corr(X, Y) = \rho(X, Y) = 0$ 은 항상 성립하지만, 반대는 성립하지 않을 수 있다.

상관관계분석

- 상관관계의 검정
 - 피어슨 상관계수(Pearson's Correlation Coefficient)
 - 상관계수가 t 분포를 따른다는 가정 하에 검정
 - X, Y 가 이변량 정규분포를 따라야 한다는 가정이 필요
 - 스피어만 상관계수(Spearman's Correlation Coefficient)
 - 비모수적 접근으로, X, Y 에 대한 분포 가정이 불필요
 - 편상관계수 (Partial Correlation Coefficient)
 - 나머지 변수들에 대한 효과를 제거하고 X, Y 의 상관관계를 분석

선형회귀분석

- 단순선형회귀분석
 - 두 변수 사이에 존재하는 상호의존관계를 함수 관계로 표현하여 연관성을 검정하는 통계적 분석
- $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
 - Y_i : i번째 관측값에 대한 종속변수의 값
 - X_i : i번째 관측값에 대한 독립변수의 값
 - β_0, β_1 : 회귀 계수(regression coefficient)
 - ε_i : Y_i 의 오차항을 나타내는 확률변수

선형회귀분석

- 단순선형회귀분석

- 최소제곱법(least square method)을 통한 추정

- $\min[\sum_{i=1}^I \varepsilon_i^2] = \min \left[\sum_{i=1}^I (Y_i - \hat{Y}_i)^2 \right] = \min \left[\sum_{i=1}^I (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2 \right]$

- 위 식을 최소화하는 $\hat{\beta}_0, \hat{\beta}_1$ 의 값을 구하는 것
 - 결과로서 추정된 $\hat{\beta}_0, \hat{\beta}_1$ 의 값은 독립변수가 취하는 범위에 제약을 받으며, 따라서 종속변수의 값을 추정할 때 데이터의 범위를 벗어나는 값을 독립변수의 값으로 사용하면 안 된다.
 - 참고: 데이터 수가 적을수록 예측구간이 넓어지는 성질

선형회귀분석

- 분산의 분해

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{\text{SST (Total sum of squares)}} = \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{SSE (Error sum of squares)}} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{\text{SSR (Regression sum of squares)}}$$

- 결정계수(coefficient of determination): $R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$
 - 독립변수의 수가 증가할수록 자연스럽게 커지는 경향
- 수정된 결정계수(adjusted R^2): $= 1 - \frac{\frac{SSE}{n-2}}{\frac{SST}{n-1}}$

선형회귀분석

- 회귀계수에 대한 t검정 ($\widehat{\beta}_1$)
 - 평균: $E(\widehat{\beta}_1) = \beta_1$
 - 분산: $Var(\widehat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$
- 가설: $H_0: \beta_1 = 0$ vs. $H_0: \beta_1 \neq 0$
- 검정통계량:

$$Z = \frac{\widehat{\beta}_1 - \beta_1}{\sigma / \sqrt{\sum (x_i - \bar{x})^2}} \rightarrow t = \frac{\widehat{\beta}_1 - \beta_1}{s / \sqrt{\sum (x_i - \bar{x})^2}}$$

선형회귀분석

- 다중선형회귀분석
 - Y와 p개의 독립변수 X_1, \dots, X_p 사이의 관계를 분석하는 통계적 방법론
 - 주요 이슈: 다중공선성(multicollinearity), 모형 선택(model selection)
- $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$
 - 단순회귀분석에서 독립변수 여러 개가 추가하여 확장된 형태
 - 검정 방법 등에서 근본적인 차이는 없음

선형회귀분석

- 선형회귀분석의 기본 가정
 - 선형성(linearity)
 - 정규성(normality)
 - 외생성(exogeneity)
 - 조건부 독립성(conditional independence)
 - 등분산성(equal-variance)

선형회귀분석

- 선형회귀분석의 기본 가정
 - 선형성: X, Y 의 관계를 선형으로 가정
 - 검정: Q-Q plot 등 시각화
 - 해결 방안: 다른 형태의 모형을 고려
 - 정규성: $\varepsilon_i \sim N(0, \sigma^2)$
 - 검정: Q-Q plot, Shapiro-Wilks test, Kolmogorove-Smirnov test
 - 해결 방안: Y 등의 데이터 분포에 맞는 모형을 고려 (예: 로지스틱 회귀분석)

선형회귀분석

- 선형회귀분석의 기본 가정
 - 외생성(exogeneity) / 내생성(endogeneity)

$$\text{Cov}(\varepsilon_i, X_i) = 0$$

$$\text{Cov}(\varepsilon_i, X_i) = E[\varepsilon_i X_i] - E[\varepsilon_i]E[X_i] = 0$$

$$E[\varepsilon_i | X_i] = 0 \rightarrow E[E[\varepsilon_i | X_i]] = E[\varepsilon_i] = 0$$

- 예: 임금(y)을 설명하는 모형에서 설명변수(x)로서 교육연수를 활용 → 이 경우, 모형에서 누락된 능력, 경험, 열의(동기부여) 등이 오차에 포함되는데, x와 강한 연관성을 가질 수도 있음.
- 해결 방안: 도구 변수(instrumental variable), 누락 변수(omitted variable), Heckman's approach

선형회귀분석

- 선형회귀분석의 기본 가정
 - 조건부 독립성(conditional independence) / 자기상관성(auto-correlation): $Cov(\varepsilon_i, \varepsilon_j | X) = 0$
 - 예: 시계열 데이터, 패널 데이터
 - 검정: Durbin-Watson test
 - 해결 방안: 시계열 모형 또는 패널 분석 모형으로 대체
 - 등분산성: $Var(\varepsilon_i) = \sigma^2, Cov(\varepsilon_i) = \sigma^2 I$
 - 검정: Bartlett test, Fligner test, Levene test
 - 해결 방안: 가중최소제곱법(weighted least square method), 변수 변환을 통한 분산 안정화(variance stabilization)

선형회귀분석

- 다중공선성(multicollinearity)
 - 독립변수 간의 강한 상관관계로 인해 회귀분석의 결과가 왜곡되는 경우
 - 예: X_1, X_2 모두 Y 에 유의한 영향을 미치는 변수이지만, X_1, X_2 가 강한 상관관계를 가지고 있어 Y 와의 관계가 제대로 포착되지 않는 경우
 - 다중공선성의 검정: 분산확대인자(Variance inflation factor; VFI)
 - 특정 독립변수를 대상으로 나머지 독립변수로 다중회귀분석을 실시한 후 나타나는 결정계수(R^2)의 값을 기준으로 판단

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

- 즉, 결정계수가 높으면 다른 독립변수에 의해 충분히 설명된다는 의미이므로 높은 다중공선성
- 일반적으로 VFI 값이 5보다 높으면 문제가 의심되고, 10 이상이면 문제가 심각한 상황으로 판단

선형회귀분석

- 모형의 선택(Model selection)
 - 1. 어떤 형태의 함수를 가정하여 모형을 구성할 것인가
 - 2. 어떤 독립변수를 회귀모형에 포함시킬 것인가 (Variable selection)
 - (1) Forward selection
 - (2) Backward elimination
 - (3) Stepwise selection
 - (1)과 (2)의 혼합 형태로, (1)로 한 개 변수를 추가한 뒤, (2)에 의해 제거될 것이 있는지 판단
 - (4) All possible regression
 - (5) Best subset regression

Lecture 5 Introduction

- 통계학 기초 / 인과추론(Causal Inference) / 알고리즘 마케팅 개관
 - 로지스틱 / 프로빗 회귀분석
 - 분산분석
 - 인과추론
 - 알고리즘 마케팅 개관
 - 알고리즘 마케팅 분석 방법의 개요