

Korea R Conference 2021

게임 데이터로 인과 추론 분석하기

2021.11.19. NCSOFT 이은조



Introduction

인과 추론 분석이란?

- X 라는 사건이 Y 라는 현상에 미치는 효과가 얼마나 되는지 통계적 가설 검정을 통해 분석하는 기법
 - 부모의 재산이 자식의 대학 진학률에 미치는 영향은?
 - 초콜릿을 먹으면 다이어트에 효과가 있는가?
 - 최저 임금을 인상하면 실업률이 증가할까?
- 인과 추론 분석이 필요한 이유 - A/B 테스트를 할 수 없는 상황이 많음
 - 시스템의 한계
 - 공정성 문제
 - 윤리적인 이슈
- 현업에서의 인과 추론 분석 예시
 - A 프로모션이 매출에 얼마나 효과가 있었나?
 - B 쿠폰을 사용한 고객의 서비스 이용 지표는 어떻게 달라졌나?
 - C 콘텐츠를 이용하는 고객은 그렇지 않은 고객보다 게임 플레이를 더 열심히 할까?

Case Study #1

Case Study #1

PK 가 이탈율에 미치는 영향 분석

- PK (Player Killing) 란?
 - 온라인 게임에서 다른 캐릭터를 공격하여 죽이는 행위
- PK의 명과 암
 - 긴장감을 주는 흥미로운 콘텐츠, 경쟁심과 성장 욕구 자극
 - PK의 폭력성, 저레벨 캐릭터 유입에 부정적인 영향
- PK 를 당한 고객은 게임에서 이탈할까?

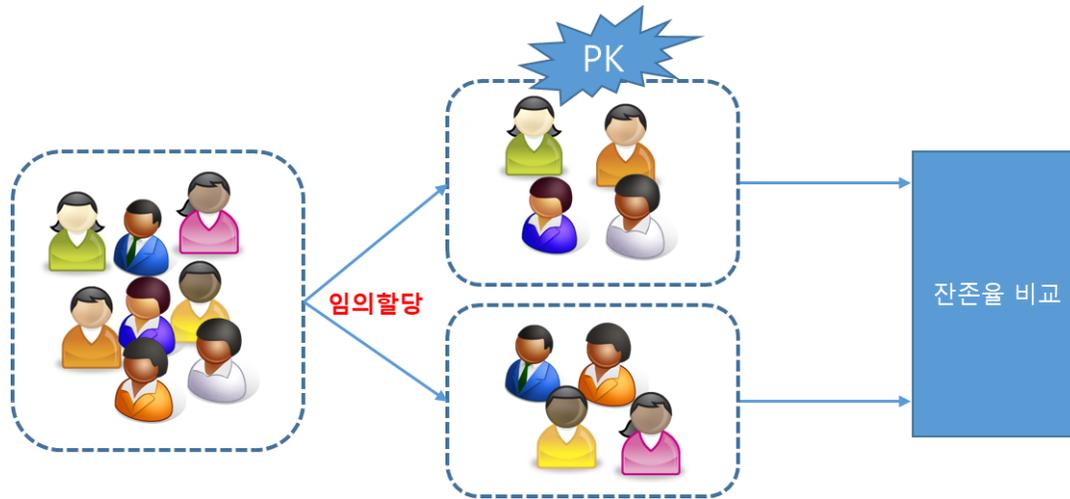


Case Study #1

PK 가 이탈율에 미치는 영향 분석

- 만약 A/B 테스트를 한다면?

- 전체 유저를 두 집단에 임의 할당 후 한 집단에 대해서만 PK 수행
- PK를 강제하는 건 서비스 윤리에 위배되며 고객의 반발을 불러일으킬 수 있음



Case Study #1

PK 가 이탈율에 미치는 영향 분석

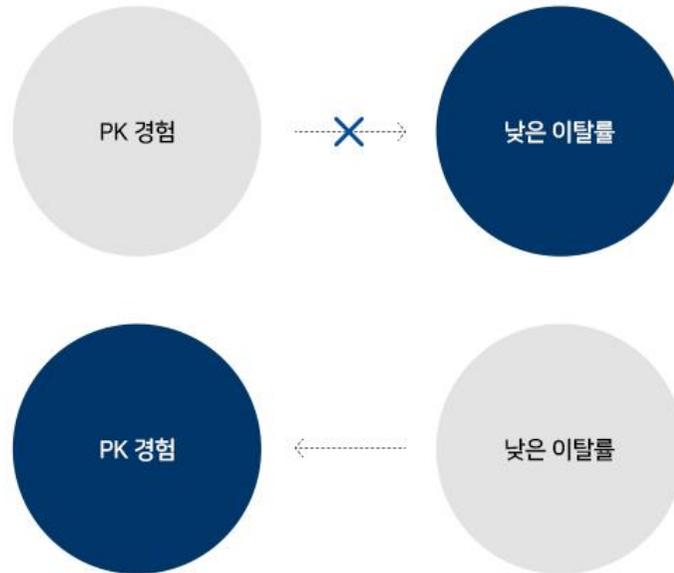
- PK 를 당한 집단과 미경험 집단의 이탈율 단순 비교
 - PK 경험 집단의 이탈율: 11.9%
 - PK 미경험 집단의 이탈율: 79.5%
- PK는 고객의 이탈을 크게 낮추는 효과가 있다? → **NO**

Case Study #1

PK 가 이탈율에 미치는 영향 분석

- 단순 비교의 문제점 #1 – 역인과 관계

- 지역별 경찰관 수와 범죄율 사이의 관계 – 경찰관이 많이 배치될수록 범죄율이 높아진다?
- 게임 플레이 시간이 높을수록 PK 를 당할 확률이 높아짐

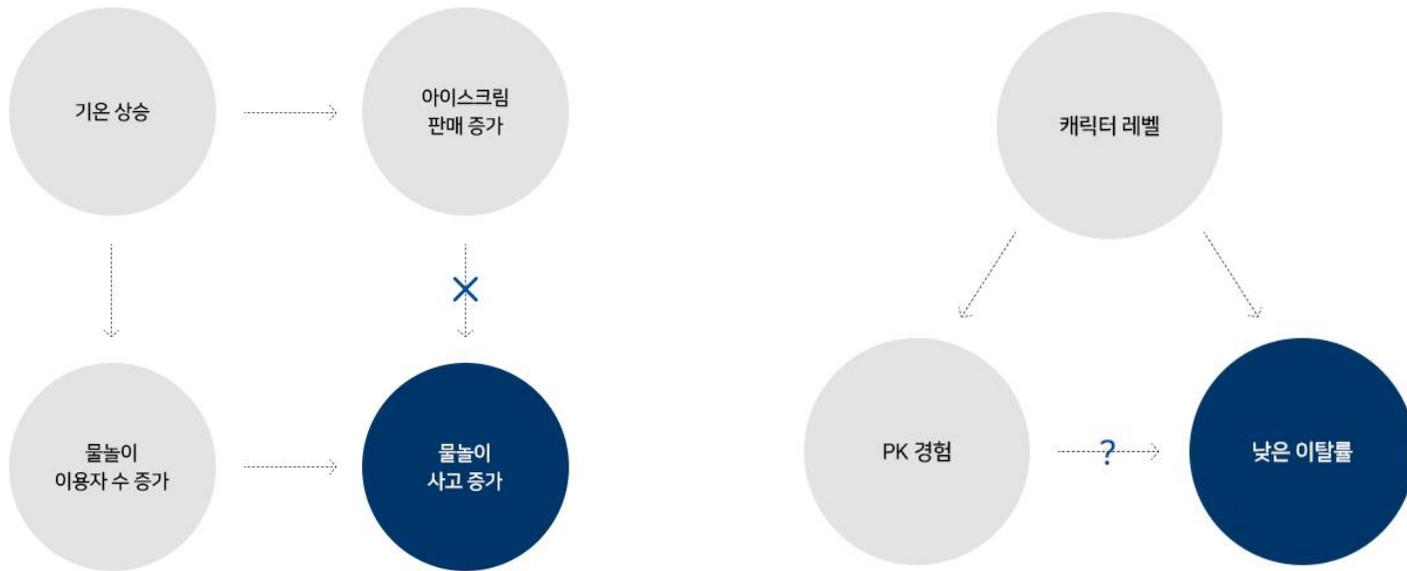


Case Study #1

PK 가 이탈율에 미치는 영향 분석

- 단순 비교의 문제점 #2 – 교란 요인 (Confounder)

- 아이스크림 판매량과 물놀이 사고와의 관계 – 아이스크림 판매가 증가할수록 물놀이 사고가 증가한다?
- 캐릭터 레벨이 PK 와 이탈율에 모두 영향을 미칠 수 있음

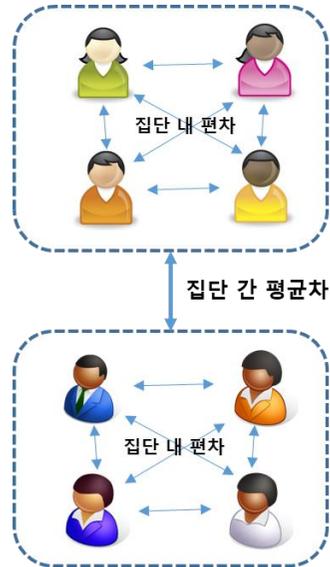


Case Study #1

PK 가 이탈율에 미치는 영향 분석

● 단순 비교의 문제점 #3 – 통계적 유의성

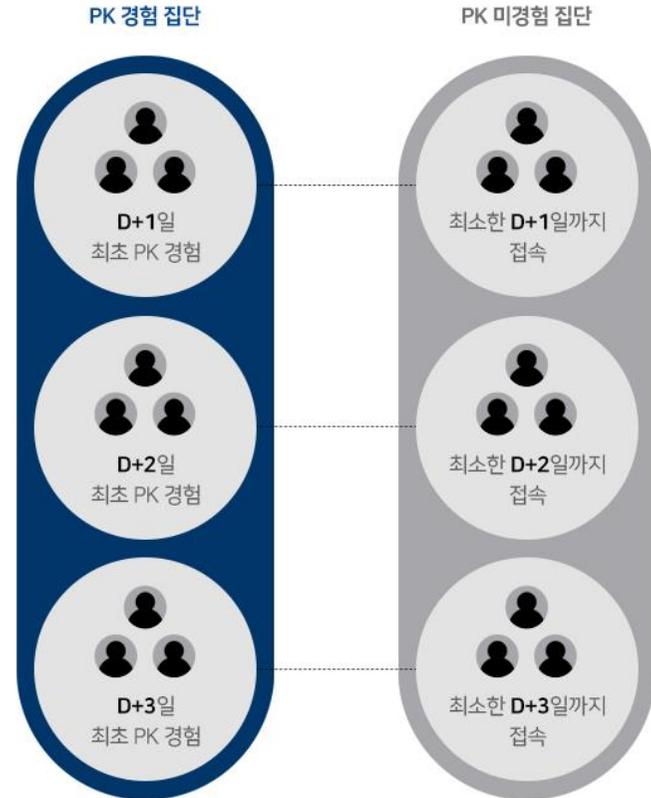
- 이탈율 차이가 통계적으로 볼 때 충분히 크다고 볼 수 있는가?
- 통계적 유의성: 내가 관측한 집단 간의 평균차가 통계적으로 볼 때 충분히 의미가 있을만큼 큰가?
- 충분한 의미가 있다? → 다른 관측 데이터를 수집하여 분석했을 때도 **일관된 결과**가 나올 것으로 기대되는가?



Case Study #1

PK 가 이탈율에 미치는 영향 분석

- 역인과 관계 고려한 데이터 샘플링
 - PK 경험 집단을 최초 PK 당한 일자별로 분류 후,
 - 최소한 해당 일자까지 접속한 이력이 있는 PK 미경험 집단 샘플링
- 역인과 효과 제거 전/후 PK 경험 집단과 미경험 집단 이탈율 차이 비교
 - 역인과 효과 제거 전: 11.9% vs. 79.5%
 - 역인과 효과 제거 후: **13.9% vs. 24.3%**



Case Study #1

PK 가 이탈율에 미치는 영향 분석

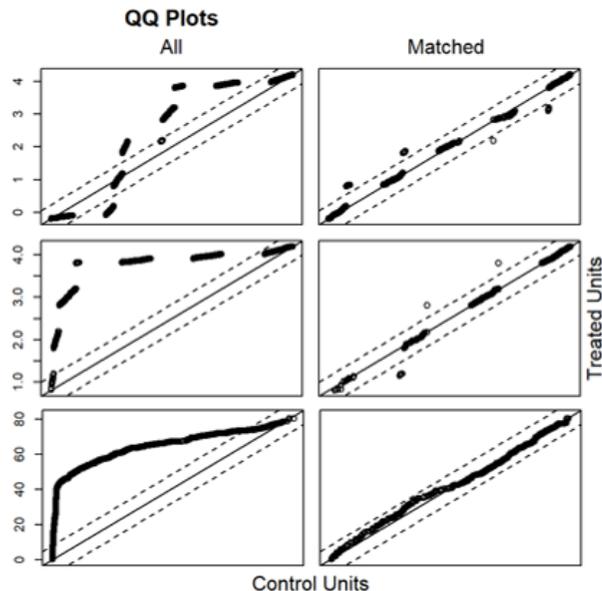
- 교란 요인 제거를 위해 성향 점수 매칭 (Propensity Score Matching, PSM) 기법 적용
 - 실험군 (PK 경험 집단) 과 대조군 (PK 미경험 집단) 에 속한 개체들의 여러 가지 특징 비교
 - 사전 조건이 비슷한 실험군/대조군 쌍을 선별하여 **교란 요인 통제**
 - 과학 실험에서 실험 결과에 영향을 주는 환경 조건을 통제하는 작업과 비슷함

종속변수: 인과효과 측정 변수

독립변수: 통제 조건들

```
library(MatchIt)
```

```
mat <- matchit(formula = treat ~ ., data) # 성향점수매칭 모델 생성  
plot(mat) # 매칭 결과 시각화  
matching.data <- match.data(mat) # 매칭된 데이터 추출
```



Case Study #1

PK 가 이탈율에 미치는 영향 분석

- 회귀 분석을 통해 인과 효과 추정 및 통계적 유의성 확인
 - PK 경험이 평균적으로 이탈율을 약 5%p 낮추는 경향이 있음

```
> model <- glm(is.churn ~ treat, data = matching.set, family = 'binomial')
> summary(model)
```

```
Call:
glm(formula = is.churn ~ treat, family = "binomial", data = matching.set)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7437 -0.7437 -0.6484 -0.6484  1.8236
```

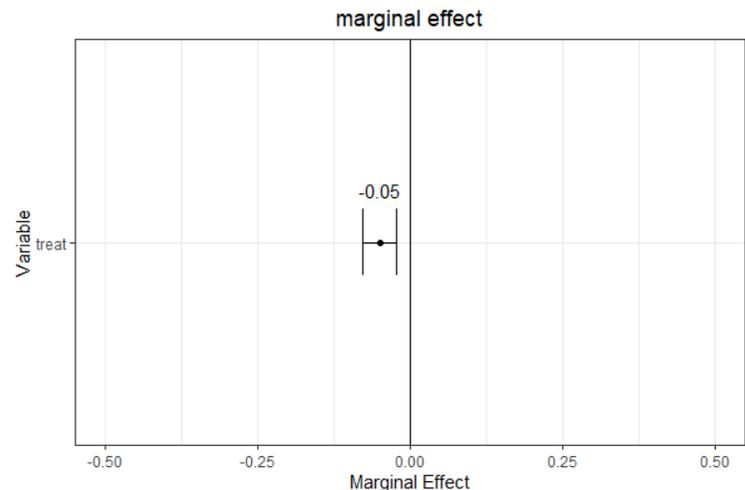
```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.14406    0.05583  -20.492 < 2e-16 ***
treat       -0.30852    0.08267   -3.732  0.00019 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 3651 on 3501 degrees of freedom
Residual deviance: 3637 on 3500 degrees of freedom
AIC: 3641
```

```
Number of Fisher Scoring iterations: 4
```

```
> library(margins)
> margins(model) %>% summary
  factor    AME    SE      z      p  lower  upper
treat -0.0520 0.0139 -3.7477 0.0002 -0.0791 -0.0248
```



Case Study #2

Case Study #2

이벤트 쿠폰 효과 분석

- 분석 배경

- 출석 이벤트를 통해 28일 연속으로 게임에 출석한 고객에게 특별한 쿠폰을 지급하는 이벤트 수행
- 해당 쿠폰이 이후 게임 플레이에 어떤 영향을 미쳤는지 효과 분석

이벤트 기간 : 4월 21일(수) 정기점검 후 ~ 5월 26일(수) 정기점검 전

데일리 혜택 · TJ's Coupon

4주 동안 진행되는 크로니클V의 특별한 데일리 혜택을 누리보세요!

421 아테나로 <TJ's 마스터 쿠폰> 데일리 혜택을 오픈할 수 있습니다
매일 출석하고 7종의 TJ's 쿠폰을 모두 획득하세요!

데일리 혜택	추가 혜택 주요 보상				
 x500 아인하사드의 은총	1일차  TJ's 쿠폰 상자 (이벤트) TJ's 쿠폰 - 클래스/ 장비 복구 / 유료 장비 복구	7일차  TJ's 쿠폰 - 아가시온	14일차  TJ's 쿠폰 - 아가시온 카드 변경	21일차  TJ's 쿠폰 - 클래스 카드 변경	28일차  TJ's 마스터 쿠폰

* 자세한 내용은 공식 홈페이지 이벤트 안내를 참고해 주시기 바랍니다.

Case Study #2

이벤트 쿠폰 효과 분석

- 쿠폰을 지급받은 고객의 이벤트 전/후 게임 플레이 변화를 단순 비교 분석할 경우 → 일평균 플레이 시간 소폭 감소
 - 이벤트 쿠폰 효과는 부정적이다? → **NO**
- 다양한 요인으로 인한 영향력 통제 필요
 - 시계열 효과 및 다른 이벤트로 인한 영향 통제 → DID (이중 차분)
 - 쿠폰 지급 고객 집단의 특성 통제 → RD (회귀 단절)

Case Study #2

이벤트 쿠폰 효과 분석

- 이중 차분 (Difference in Difference, DID)

- 실험군과 대조군에 대해 각각 이벤트 전/후의 차이를 측정 후 둘 간의 차이를 비교

$$\text{쿠폰 지급 효과} = (C - A) - (D - B)$$

	쿠폰 지급 고객	쿠폰 미지급 고객
이벤트 이전 플레이 시간	A	B
이벤트 이후 플레이 시간	C	D

- 회귀 분석에서 '상호 작용 항 (interaction term)' 과 같은 개념

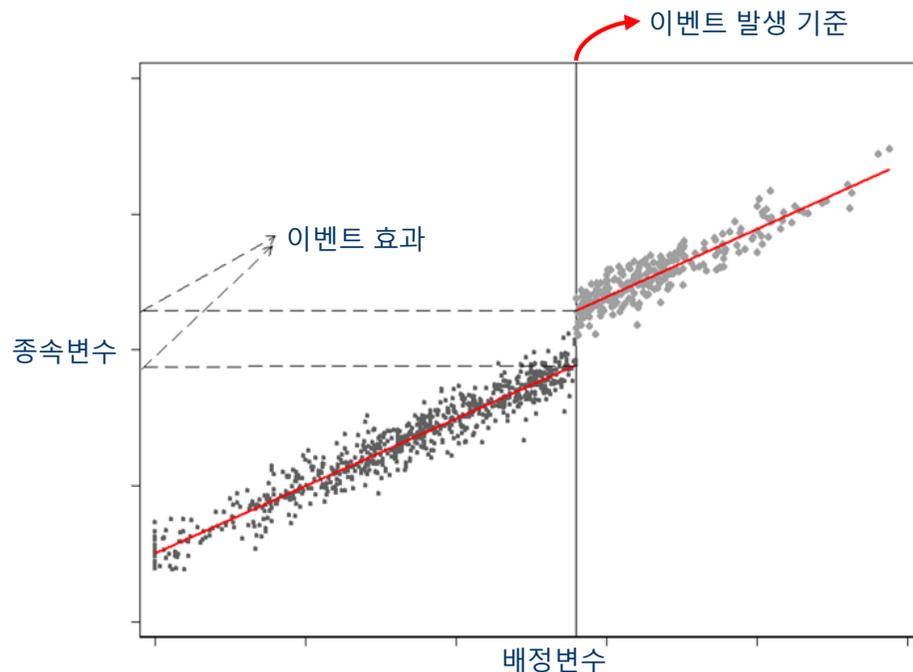
- 이중 차분 활용 시 주의점

- 실험군과 대조군이 주어진 사건 (여기서는 쿠폰 발급) 을 제외하면 모두 동일한 특성을 갖고 있어야 함 (매칭 기법과 같이 사용하면 좋음)

Case Study #2

이벤트 쿠폰 효과 분석

- 회귀 단절 모형 (Regression Discontinuity Design, RDD)
 - 인과 효과 측정 대상이 되는 이벤트가 특정 변수 (배정변수) 에 의해 결정
 - 배정 변수가 효과 측정 대상이 되는 종속변수와 상관 관계
- 쿠폰 이벤트의 특징
 - 이벤트 기간동안 게임 접속 일수가 28일 이상이 되어야 함
 - 게임 접속 일수는 플레이 시간 및 결제 금액과 상관성 높음
- 명문대학 입학이 졸업 후 연봉에 미치는 영향은?
 - 입학 커트라인 근처의 합격자와 불합격자 연봉 비교
- 쿠폰 이벤트가 플레이에 미치는 영향은?
 - 출석 일수 28일 근처의 쿠폰 지급자와 미지급자 비교
 - $\hat{y} = \alpha + \beta T_i + \gamma C_i + \varepsilon$
 - T_i : 쿠폰 지급 여부 (처치변수, $C_i \geq 28$ 이면 1 아니면 0)
 - C_i : 출석 일수 (배정변수)



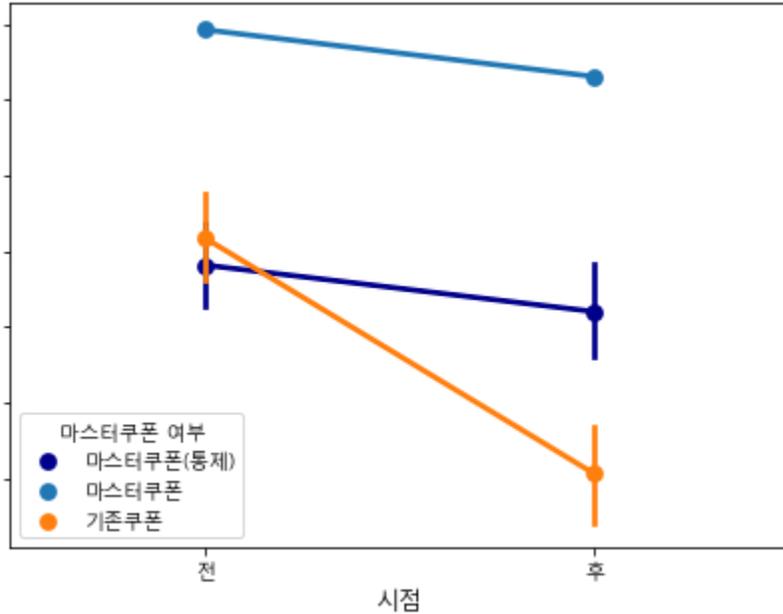
Case Study #2

이벤트 쿠폰 효과 분석

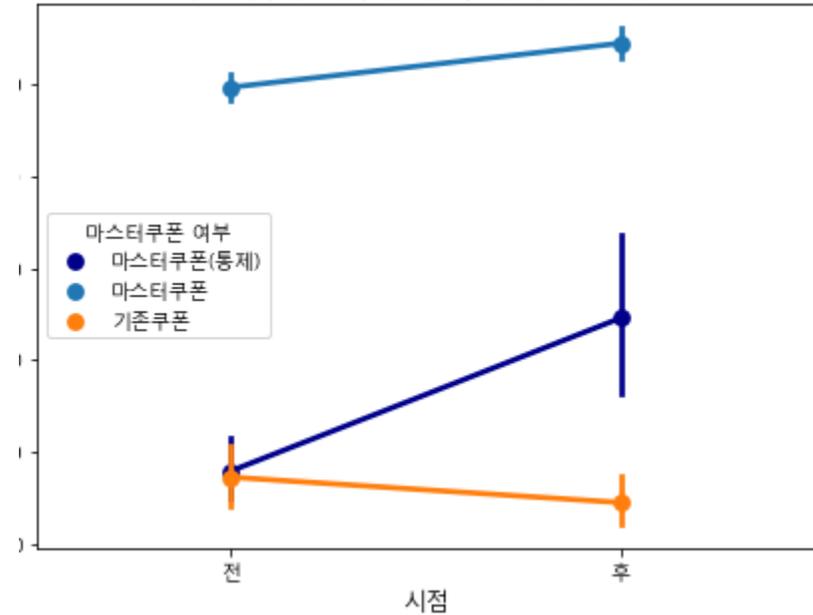
- 쿠폰 효과 분석 결과 #1 - 이중 차분

- 기존 대비 일 평균 플레이 시간 및 결제 금액에서 모두 긍정적인 효과가 있었음

마스터쿠폰 여부별 전/후 일별 플레이타임 변화



마스터쿠폰 여부별 전/후 결제 금액 변화



Case Study #2

이벤트 쿠폰 효과 분석

- 쿠폰 효과 분석 결과 #2 - 회귀 단절 모형

- 기존 대비 일 평균 플레이 시간 및 결제 금액에서 모두 긍정적인 효과가 있었음

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.094*** (0.012)	-0.071*** (0.015)	-0.072** (0.029)	-0.048 (0.034)	-0.098*** (0.014)	-0.070*** (0.017)
출석 횟수	0.023** (0.011)	0.044*** (0.014)	0.080** (0.027)	0.083*** (0.031)	0.057 (0.041)	0.078* (0.042)
	0.306*** (0.041)	0.305*** (0.041)				
이벤트쿠폰 획득 여부	2.559*** (0.304)	1.954*** (0.384)	2.001*** (0.768)	1.327 (0.897)	2.895*** (0.357)	1.955*** (0.448)
	0.018*** (0.002)	-0.083** (0.039)	0.017*** (0.002)	-0.081 (0.068)	0.029 (0.027)	-0.083* (0.049)
		0.101** (0.039)		0.098 (0.068)		0.162*** (0.059)
	-0.205*** (0.003)	-0.205*** (0.003)	-0.213*** (0.004)	-0.213*** (0.004)	-0.183*** (0.006)	-0.183*** (0.006)

Summary & Conclusions

Summary & Conclusions

- 인과 추론 분석은 현업에서 이제 반드시 알아야 하는 스킬이 되고 있다
 - A/B 테스트를 이용할 수 있는 상황은 제한적임
 - 이벤트나 프로모션의 효과 분석 시 상관을 인과로 오해하면 잘못된 결론을 도출할 수 있음
- 인과 추론 분석의 핵심은 적절한 대조군 선정과 모형 설계
 - 성향 점수 매칭 (Propensity Score Matching, PSM)
 - 이중 차분 (Difference In Difference, DID)
 - 회귀 단절 모형 (Regression Discontinuity Design, RDD)
- 인과 추론 분석할 때는 **재현성 (Reproducibility)** 을 항상 신경 써야 함
 - 분석 결과를 선불리 일반화해선 안됨



END OF DOCUMENT