

# 『Marketing Science』

---

Week 7 : 준실험(Quasi-experiments) (CH.13)

1. 무작위 실험이 어려운 경우
2. 준실험이란
3. 이중차분법
4. 합성통제
5. Casual Impact

# 무작위 실험이 어려운 경우

무작위 대조 실험(Randomized Controlled Trial, RCT)가 가장 좋은 실험 방법이지만, 현실적으로/기술적으로 수행하기 어려울 때가 많음.

## 시장 전체의 변화나 리브랜딩

브랜드 전면 개편이나 전체 시장을 대상으로 한 변화는 모든 소비자가 동시에 변화에 노출되므로, 실험군과 대조군을 나누기 어려움

## 법적·윤리적 제약

가격 차별(같은 상품을 다른 소비자에게 서로 다른 가격에 판매)과 같은 실험은 법적 문제나 윤리적 논란으로 인해 실험이 어려움.

## 플랫폼 또는 채널 제약

어떤 디지털 플랫폼은 전체 사용자에게 동시에 알고리즘을 적용하기 때문에 통제군을 따로 만들 수 없음.

## 고비용 또는 실행상의 어려움

새로운 오프라인 매장 포맷 도입이나 대규모 캠페인 실행은 비용과 자원이 많이 들어, 무작위 실험을 설계하는 것이 어려움

## 장기 브랜딩 캠페인

장기간에 걸친 캠페인은 시간이 오래 걸리고, 소비자의 인식이나 행동에 누적된 영향을 미치기 때문에 실험적 통제가 어려움.

## 개인정보 보호 문제

맞춤형 마케팅 실험이 소비자 행동을 정밀하게 추적해야 할 때, 개인정보 보호 법규로 인해 실험 설계가 제한됨.

## SUTVA(Stable Unit Treatment Value Assumption) 위반

SUTVA 한 유닛(예: 사용자)의 결과가 다른 유닛의 처리(assigning treatment)에 영향을 받아선 안 된다”

예: 에어비앤비에서 개인화 알고리즘이 한 사용자에게 숙소를 추천하면, 그 숙소가 예약되어 다른 사용자에게는 노출될 수 없습니다.

이 경우 한 유닛(숙소)의 결과가 다른 유닛(사용자)의 처리에 영향을 받게 되므로 SUTVA가 깨집니다.

# 준실험(Quasi-experiments)

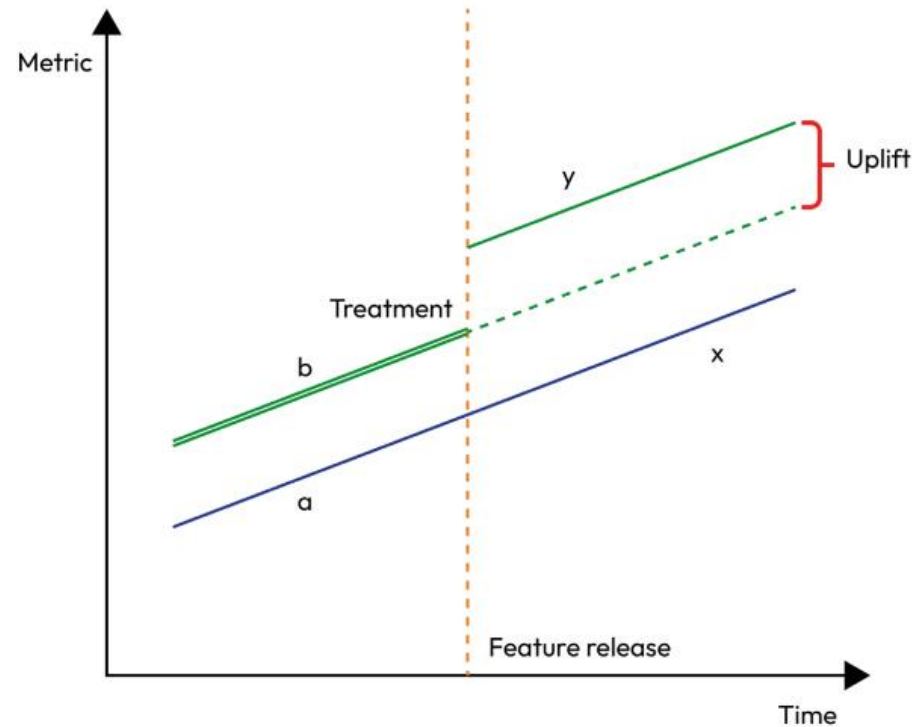
- 준실험, 또는 유사실험으로 불림
- 기존의 데이터를 활용하여 실험조건을 어느정도 통제하여 분석하는 방법
- 실제 실험과 유사하지만 참가자를 무작위로 그룹에 배정하지 않는 연구 방법론
- 내적타당도는 RCT와 비교했을 때 낮지만, RCT와 비슷한 특징을 많이 가지고 있기에 실험 처치전과 후에 대한 유의한 차이를 검증할 수 있음.

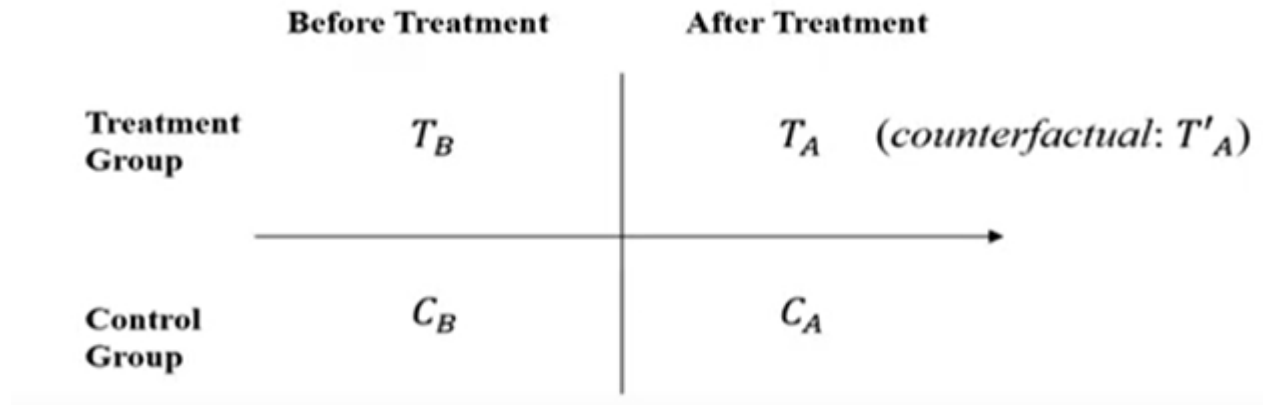
## 준실험의 종류

1. Difference in Difference (DID) (이중차분법)
2. Synthetic Control(합성 통제)
3. Causal Impact

## 이중차분법(Difference-in-Differences, DiD)

- 정책 변화나 특정 개입의 효과를 추정하기 위해 사용되는 준실험적 방법론
- 실험군과 대조군 모두 시간에 따라 결과가 어떻게 변하는지를 비교하여 개입의 순수한 효과(causal effect)를 추정하는 방식





- 두 가지 변화 : 시간에 따른 변화, Treatment에 따른 변화
- 1종 차분 :  $T_A - T_B$ 는 treatment에 따른 실험군의 시간에 따른 변화를,  $C_A - C_B$ 는 대조군의 시간에 따른 변화
- 2종 차분:  $T_A - C_A$ 는 treatment 이후에 실험군과 대조군의 차이를,  $T_B - C_B$ 는 treatment 시점 이전에 실험군과 대조군의 차이를 의미

$$\text{DID estimator} = (T_A - T_B) - (C_A - C_B)$$

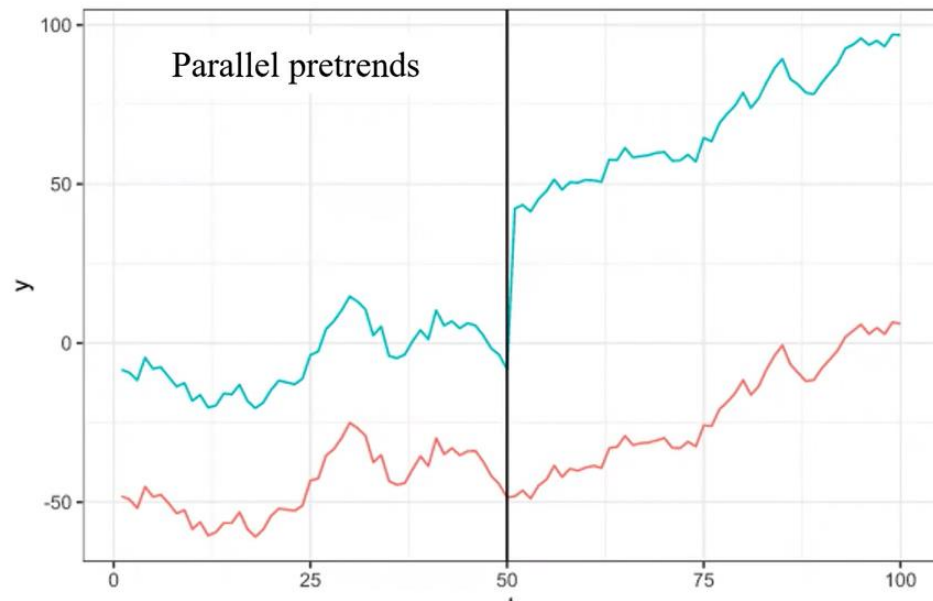
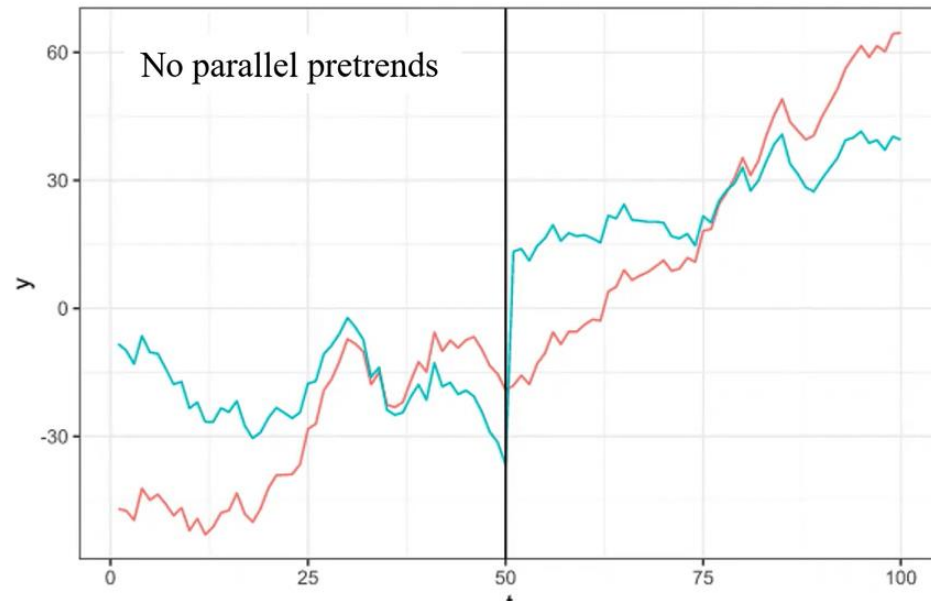
- DID는 1종차분값 에서 2종 차분 값을 빼는 방식
- 실험군과 대조군 각각의 treatment 시점 전후의 차이를 구하고, 이것을 다시 실험군의 변화에서 대조군의 변화의 차이를 뺀다.

- 문제는 DID는 시간에 의해 변하지 않는 요인을 제거하는 반면, 시간에 의해 변하는 요인은 제거하지 못함
- 즉,  $T_A - T_B$ 에는 시간의 흐름에 따라 발생하는 효과와 treatment의 효과가 함께 포함됨
- 반면  $C_A - C_B$ 에는 treatment가 없기 때문에 시간의 흐름에 따른 변화만 포함.
- 즉, 시간에 따른 변화가 실험군과 대조군에서 동일하다면(평행추세검증을 만족한다면) 아래의 DID 계산을 통해서 인과관계만 추려낼 수 있다.

$$\begin{aligned}\text{DID estimator} &= (T_A - T_B) - (C_A - C_B) \\ &= T_A - [T_B + (C_A - C_B)] \\ &\quad \text{Inferred counterfactual}\end{aligned}$$

- Inferred counterfactual : treatment가 없었을 때 자연스럽게 나타나는 효과

평행 추세 검증 -> 도표로 확인한다.



사전기간 회귀 (Pre-Trend Test)

추세의 기울기 비교 (Slope Test)

그러나 실험군이 여러개이고, Treatment 시점이 다르다면, 이렇게 두그룹으로 나눠서 도표로 보기 힘들다.



# 이중 차분법 - 평행 추세 검증

Relative time model : Treatment가 다르게 적용된 것을 같게 맞춰줌

도시	정책시행날짜	lag2	lag1	시행	lead1	lead2
뉴욕	3/23	3/21	3/22	3/23	3/24	3/25
캘리포니아	3/22	3/20	3/21	3/22	3/23	3/24
텍사스	4/2	3/31	4/1	4/2	4/3	4/4

시행 - 코로나 시기 도시를 lockdown한 날짜.

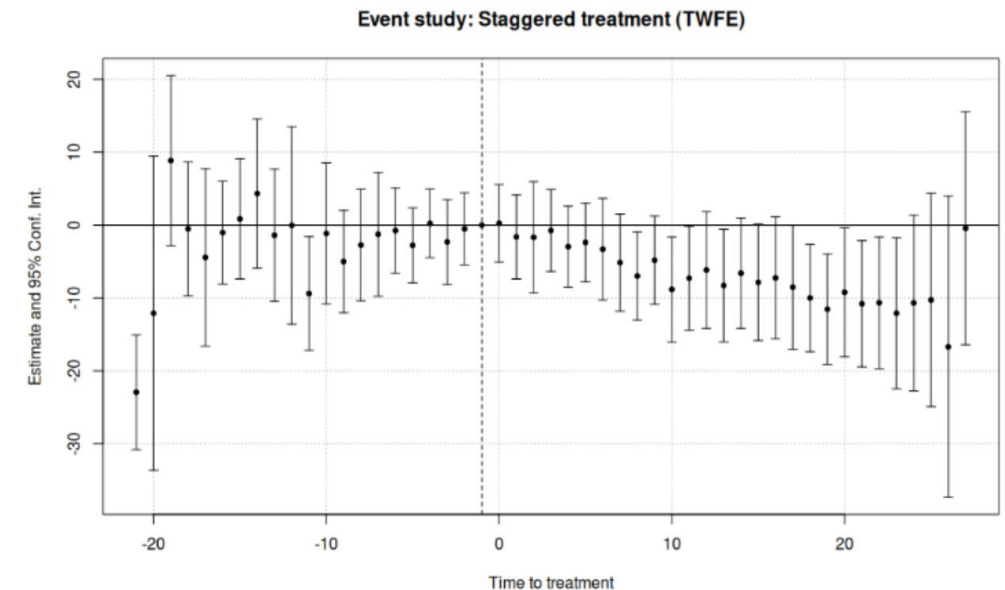
## Regression Coefficients Plot

- 계수들의 분포를 시각화하여 평행추세검증을 만족하는지 확인할 수 있음
- X축: 시간 (예: 처리 이전, 처리 시점, 처리 이후 연/월)
- Y축: 회귀 계수 (통제군과 실험군의 결과 차이)
- 점과 선: 각 시점에서의 추정 계수 (점)와 신뢰구간 (선)
- 정책이 시행되기 전에 집단들의 패턴이 비슷하다면 특별한 효과가 나타나지 않기 때문에 정책이 시행되기 이전의 기간에는 효과가 관찰되지 않아야함.

- 오른쪽 표에서는 lag 부분에는 0에 가깝지만 lead 부분은 0과 멀어지는 것이 관찰 됨.

→ 집단들이 평행추세 검증을 만족한다고 볼 수 있음

- 여기서 시간이 지날 수록 정책의 효과가 강해지는 것을 관찰할 수 있음



# 이중 차분법 예시

이중 차분법에서 가장 중요한 것은 적절한 Control그룹을 찾는 것.

A라는 웹사이트에서 대규모 UI/UX 개편을 했다. 실무자들은 UI/UX 개편에 대한 효과를 알고 싶어한다.

그런데 UI/UX개편과 동시에 대규모 프로모션을 집행하였다.

프로모션의 효과를 제외하고 UI/UX의 효과만 보기 위한 방법은?

Ex) 무신사

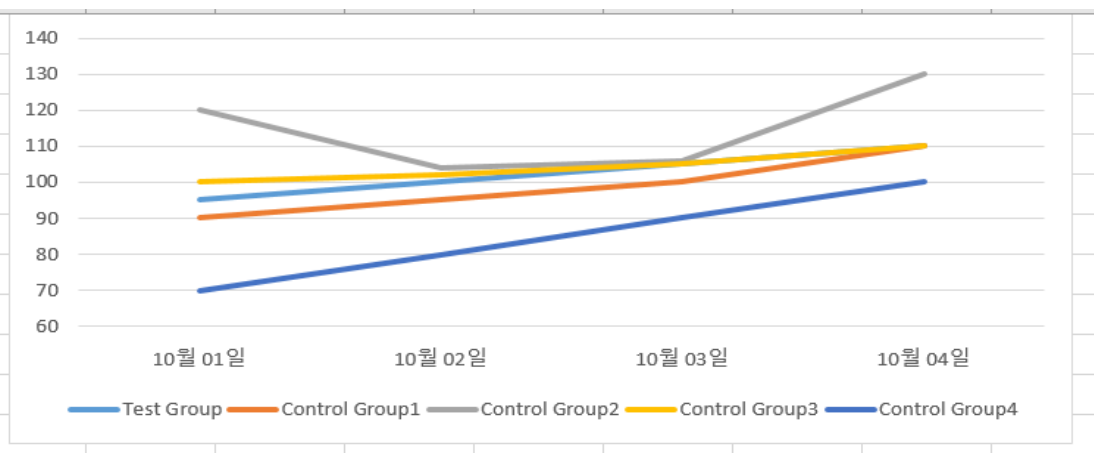
# Synthetic Control(합성 통제)

- 합성통제법은 처리(정책, 개입 등)를 받은 실제 대상(unit)에 대해, 비슷한 특성을 가진 여러 통제 그룹의 가중합(synthetic control)을 만들어서 가상의 대조군을 구성하고, 이와 비교함으로써 Treatment 효과를 추정하는 방법
- 적절한 대조군을 선택하기 어렵거나, 평행추세를 만족하지 않는 경우 사용
- 가중치를 구하는 것이 핵심
- Python package : synthcontrol, SyntheticControlMethods

# Synthetic Control(합성 통제) 예시

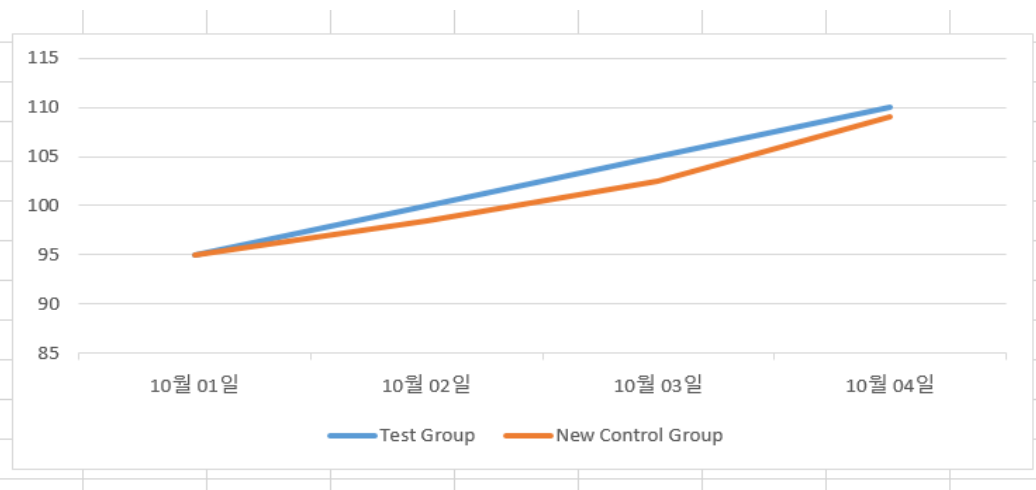
## 가중치 적용 전

	Pre Treatment Period				Treatment
	10월 01일	10월 02일	10월 03일	10월 04일	10월 05일
Test Group	95	100	105	110	
Control Group1	90	95	100	110	
Control Group2	120	104	106	130	
Control Group3	100	102	105	110	
Control Group4	70	80	90	100	



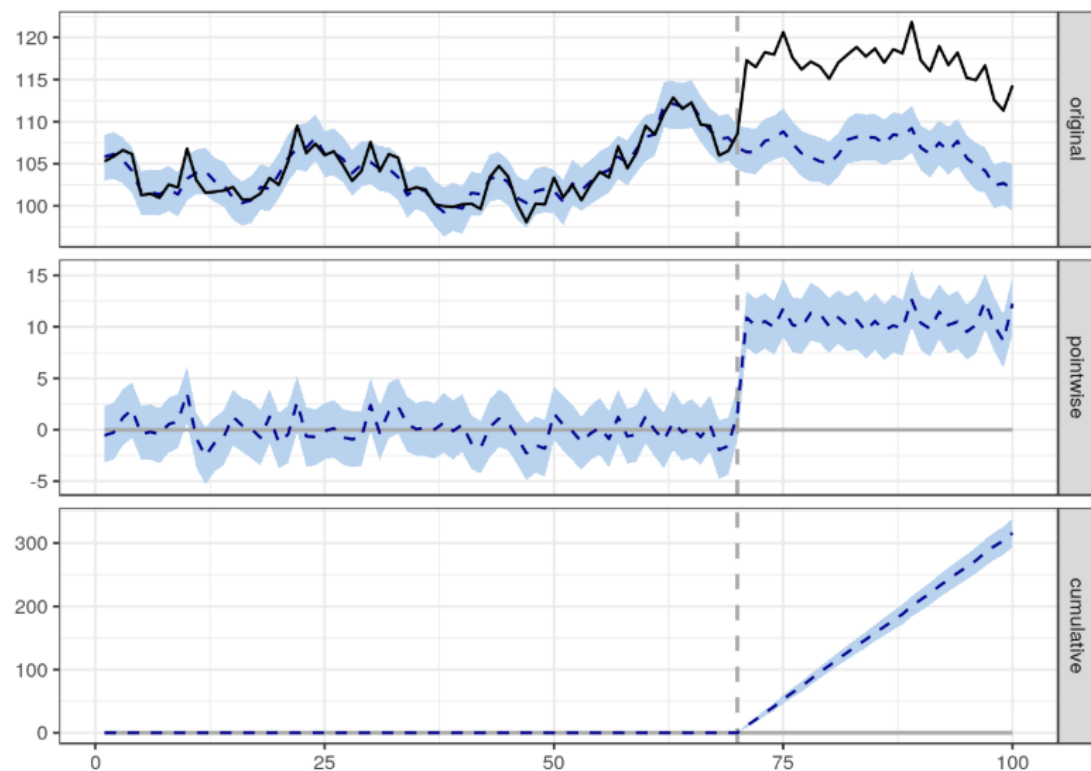
## 가중치 적용하여 가상의 대조군 구성

가중치		10월 01일	10월 02일	10월 03일	10월 04일
0.2	Control Group1	18	19	20	22
0	Control Group2	0	0	0	0
0.7	Control Group3	70	71.4	73.5	77
0.1	Control Group4	7	8	9	10
	New Control Gro	95	98.4	102.5	109



- 시계열 데이터에 대한 인과효과 추정
  - 실제 관측된 결과와 개입이 없었을 때를 예측한 값의 차이를 treatment의 효과 (casual impact)로 판단
  - 빈도주의를 기반으로 한 ARIMA와 다른 Bayesian 접근법을 사용
  - 추세, 계절성, 여러 시계열 데이터간의 상관관계 등을 반영
  - 즉, 참고할 만한 다른 시계열 데이터(공변량 변수)가 필요
  - 보통 대조군이 없을 때 사용
- 
- Package : CausalImpact (R) , pycausalImpact (python, R패키지를 python으로 바꾼 비공식 버전)
  - <https://google.github.io/CausalImpact/CausalImpact.html>

# Causal Impact



	Average	Cumulative
Actual	110.2	3306.0
Prediction (s.d.)	100.3 (0.5)	3009.0 (15.0)
95% CI	[99.3, 101.2]	[2979.0, 3035.4]
Absolute effect	9.9	297.0
95% CI	[9.0, 10.9]	[270.6, 327.0]
Relative effect (%)	+9.8%	+9.9%
95% CI	[+9.0%, +10.9%]	
Probability of causal effect:	~100%	

Probability of causal effect : 개입이 원인으로 작용할 확률, 100%에 가까울 수록 개입이 실제로 인과효과를 가질 확률이 높다고 해석  
Bayesian에서는 p-value 대신에 posterior 확률을 사용

**분석 주제 :** 비용을 많이 들인 TVCF(PPL)의 단기적인 효과 측정 (유입 및 매출)

## 분석 과정

- Treatment 전후 (TVCF/PPL) 동시에 홈페이지 유입과 매출에 영향을 주는 요소들을 통제
- 주로 각 매체사들의 비용, 프로모션 여부, 인스타 광고 여부
- Causal Impact 알고리즘을 활용하여 Treatment 전후 차이 비교

## 결과

- TVCF가 유입에는 효과가 없었지만, 매출에는 효과가 있었음
- 해석 : TVCF가 객단이 높은 상품을 광고 -> 다른 제품 사려고 들어온 고객이 TVCF광고를 보고 객단이 높은 상품으로 갈아탐
- 분석 결과에 따라 TVCF/PPL 집행 방향 수정

인과추론의 데이터 과학

- <https://www.youtube.com/watch?v=yCeaZ9Ktk7g>
- [https://www.youtube.com/watch?v=0C4\\_OikuQZc](https://www.youtube.com/watch?v=0C4_OikuQZc)

<https://docs.iza.org/dp13524.pdf>

[https://lost-stats.github.io/Model\\_Estimation/Research\\_Design/event\\_study.html](https://lost-stats.github.io/Model_Estimation/Research_Design/event_study.html)

<https://google.github.io/CausallImpact/CausallImpact.html>



# 감사합니다.

---