



함수랑산악회

A/B 테스트를 비교에서 검증으로 바꾸기

통계적으로 다시 설계하는 상세페이지 A/B 테스트

이도윤

Software Engineer

발표자 소개



이도윤

- 함수랑 1기 클라이머
- ~ 2025. 데이원컴퍼니 프론트엔드 개발자
- 현재는 창업 준비 중

만드는 사람



만드는 사람

기획하고, 디자인하고, 개발하면서 프로젝트를 직접 만들어 봅니다.



디자인 전공

처음부터 관심사는 '무엇을 어떻게 만들 것인가'에 가까웠습니다.



창업 준비

퇴사 후 기획 공부와 개인 프로젝트를 병행하며 창업을 준비하고 있습니다.

STATISTICAL THINKING

AGENDA

01 A/B 테스트에서 생긴 의문

02 진짜 유저 데이터를 알려면

03 가설 검정과 오류

04 성공을 판단하기

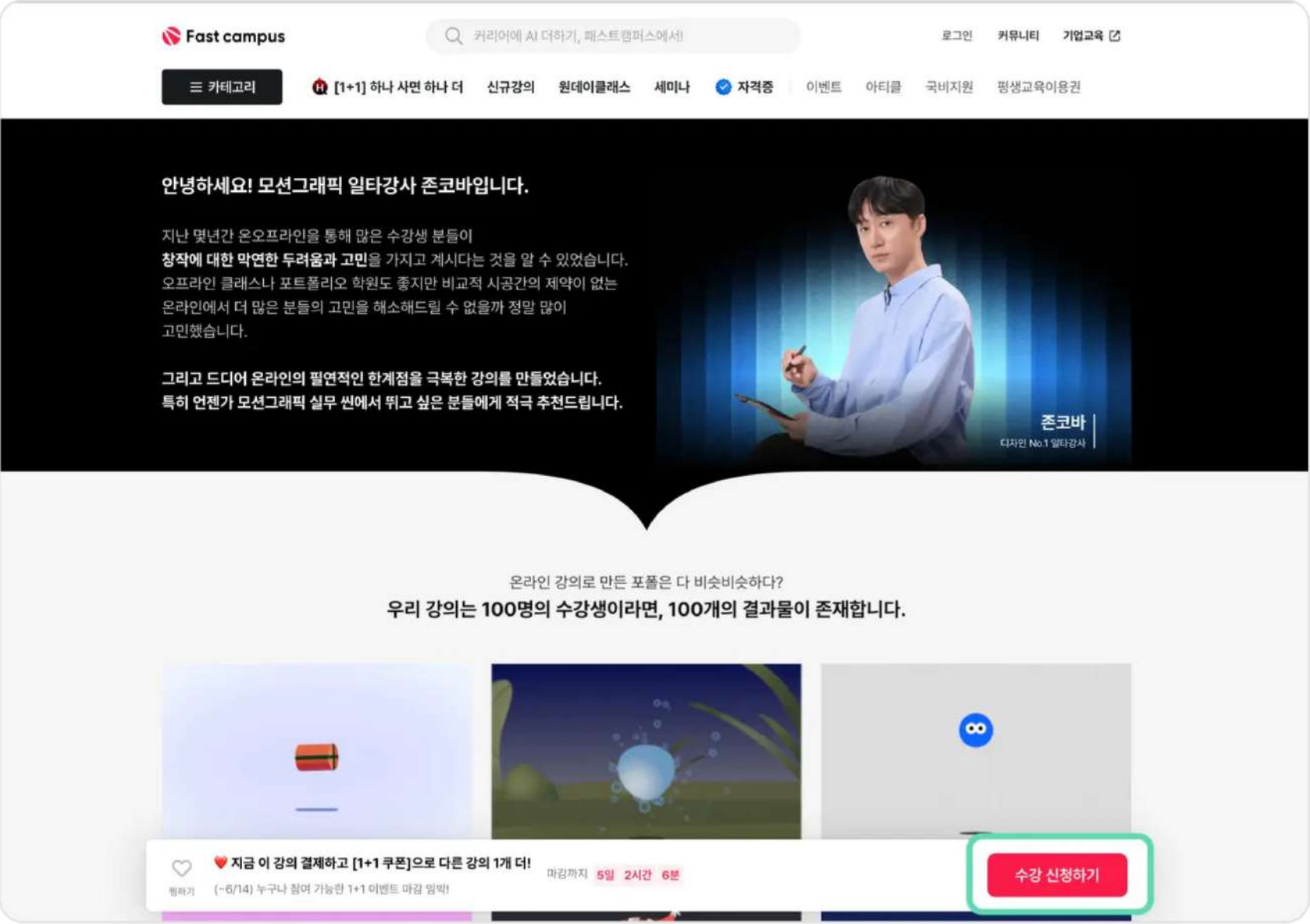
05 표본수와 기간 정하기

1. A/B 테스트에서 생긴 의문 - CASE

상세페이지 CTA 영역 개편 후 A/B 테스트를 진행했습니다

A · 기존

Fastcampus detail page

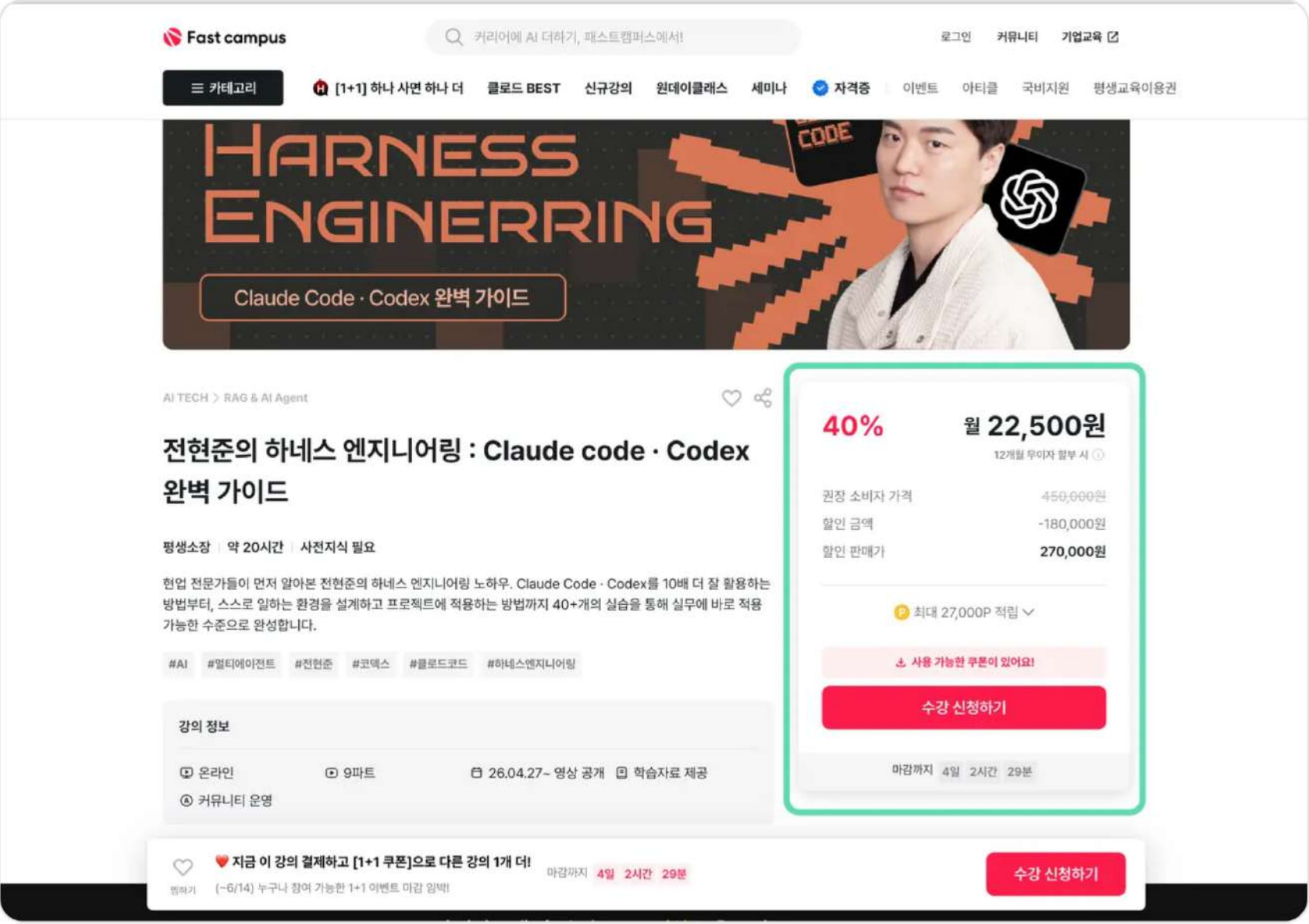


기존 상세페이지 A

하단 플로팅바에 구매 CTA가 집중되어 있었습니다.

B · 개편

Fastcampus detail page



개편 상세페이지 B

상단 상품 정보 섹션과 CTA 버튼을 추가했습니다.

1. A/B 테스트에서 생긴 의문 - SETUP & RESULT

당시에는 '수치가 올랐으니 성공'이라고 생각했습니다

OBSERVED RESULT

참여율 +5.4% · 스크롤 +6.8% · 구매 전환율 +1.6% 향상

01

SPLIT

사용자 분리

A/B 그룹을 5:5로 나눔

NEXT.js

02

METRICS

지표 측정

참여율, 스크롤 유지율, 구매 전환율 확인



03

DURATION

기간 설정

가격 정책에 맞춰 2주간 진행



수치가 올랐다면 성공일까



UI가 바뀌어서 초반에만 사람들이 흥미를 보였을 수도 있지 않을까?



CTA 버튼이 두 개 되었으니 자연스럽게 높아진 것처럼 보인 건 아닐까?



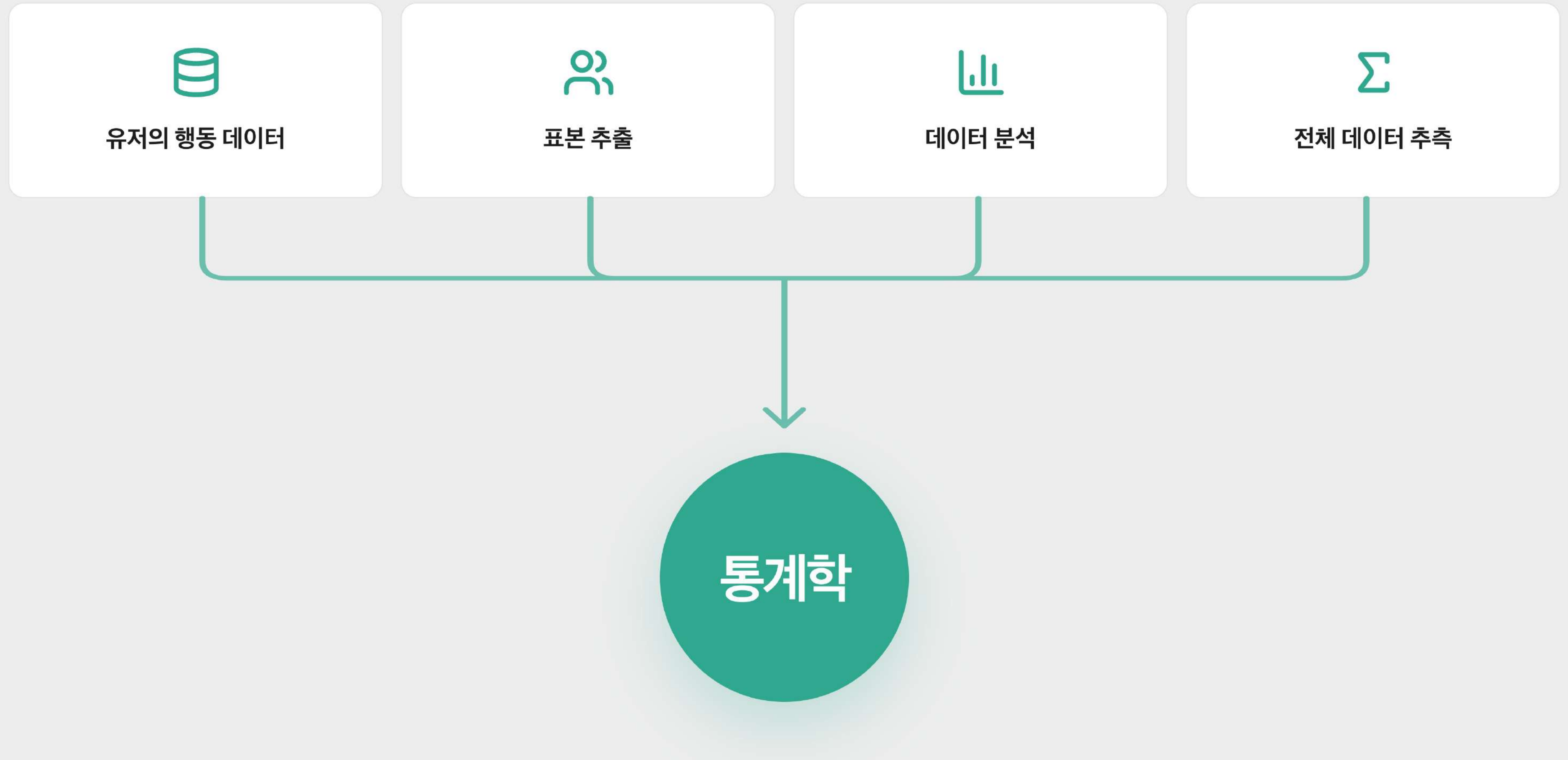
이번 일부 데이터로 전체 사용자에게도 효과가 있다고 말할 수 있을까?

2. 진짜 유저 데이터를 알려면 - MOTIVATION

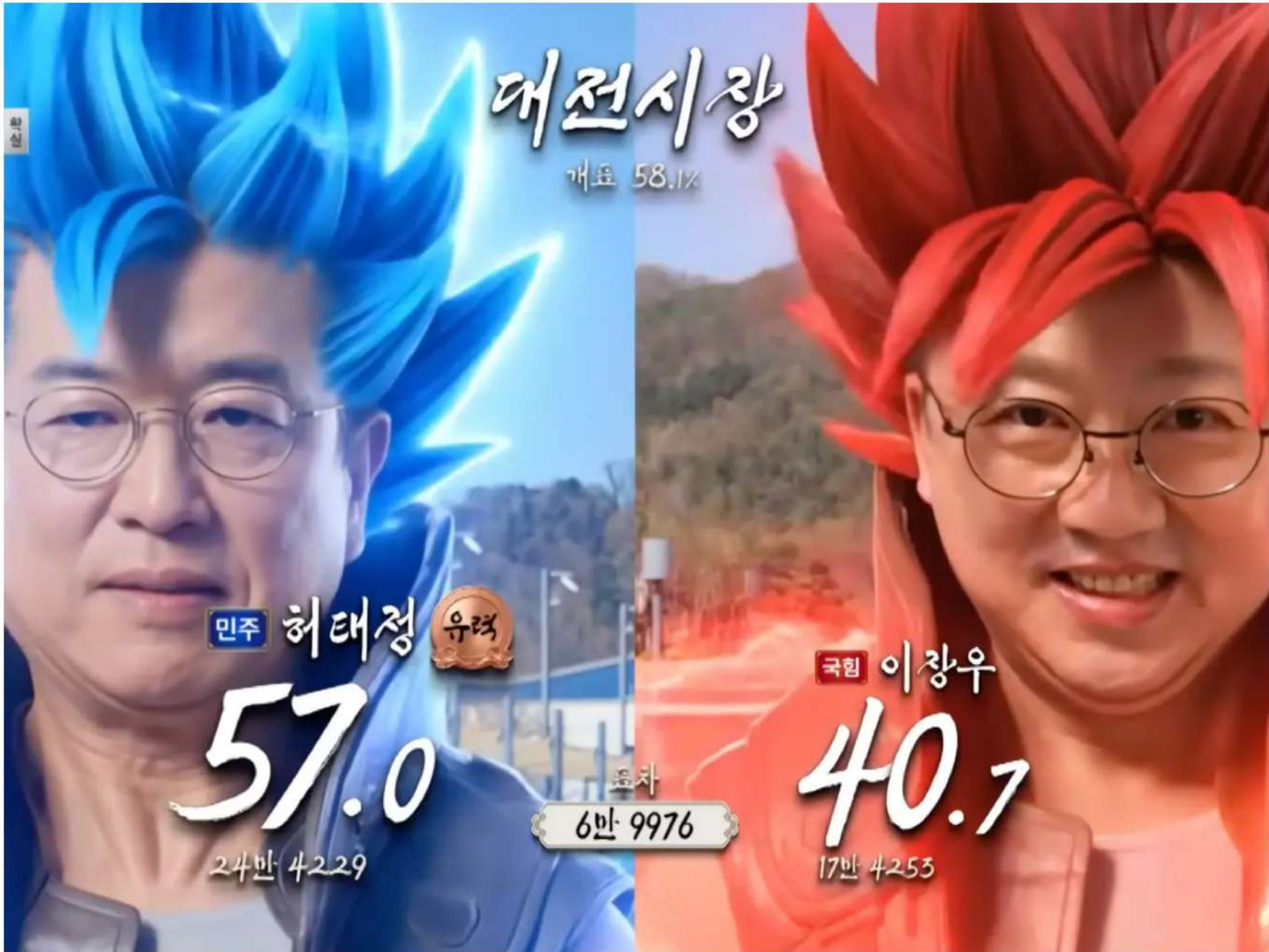
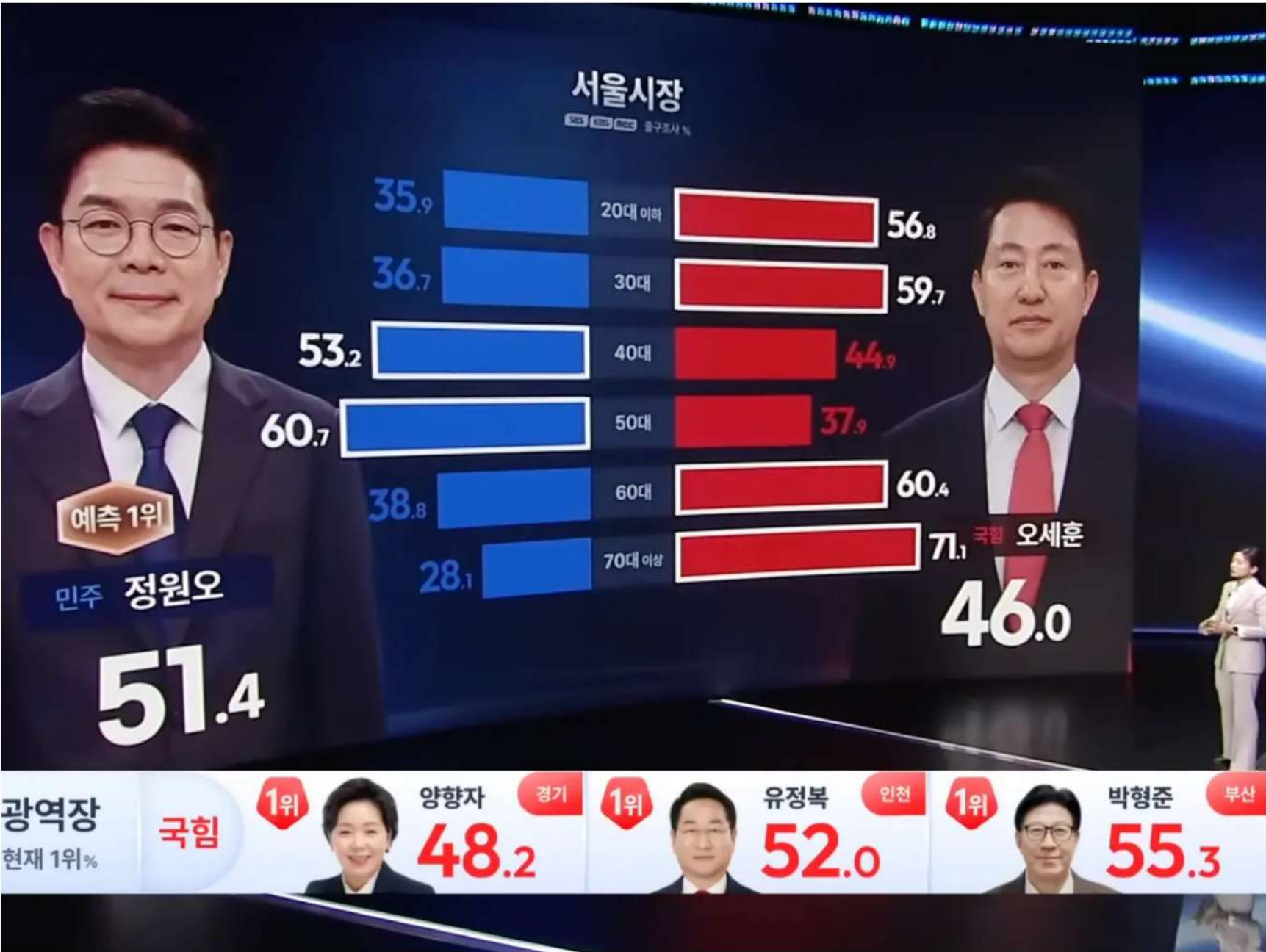
내 생각은 의미가 없다



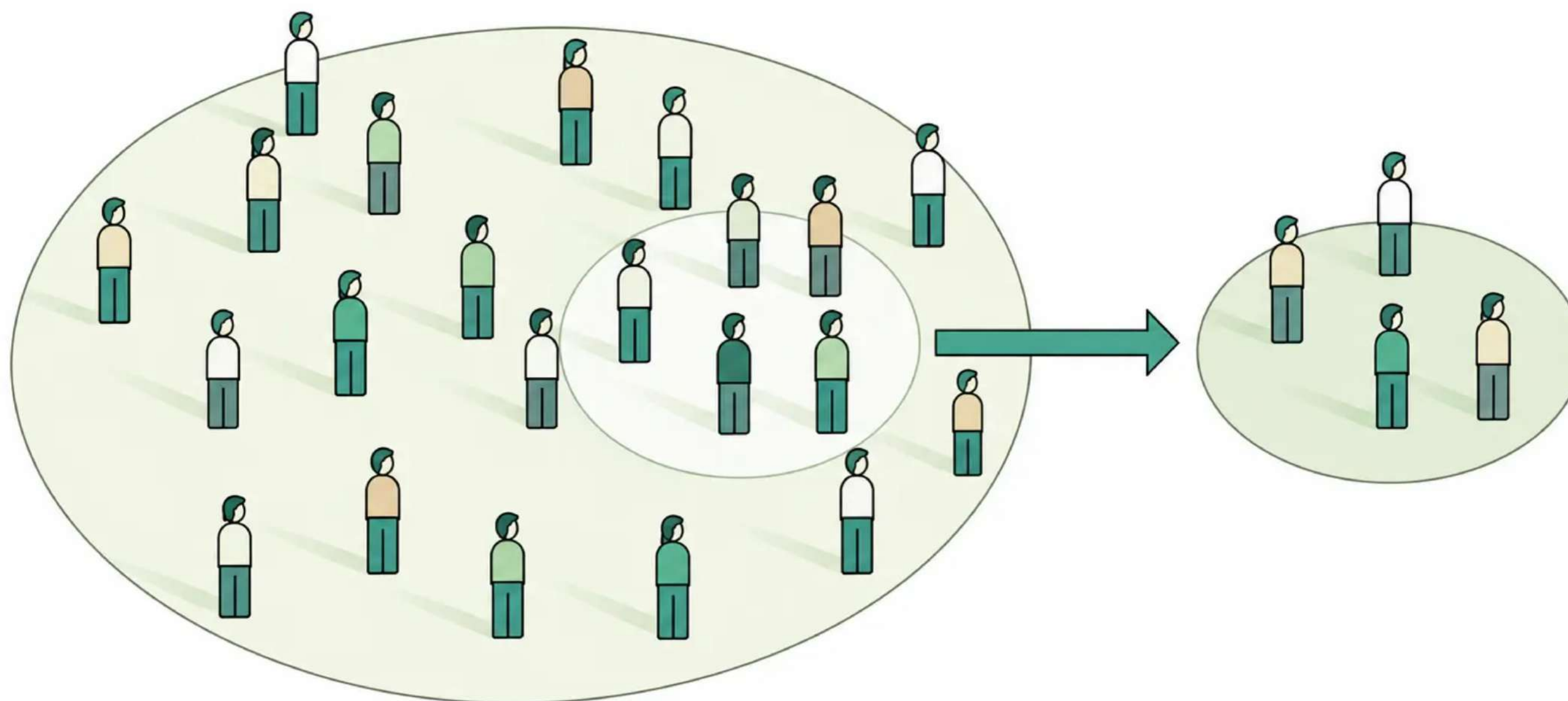
진짜 유저의 생각을 들여다보려면



선거 출구조사와 통계



A/B 테스트도 결국 일부 데이터로 전체를 추정하는 일입니다

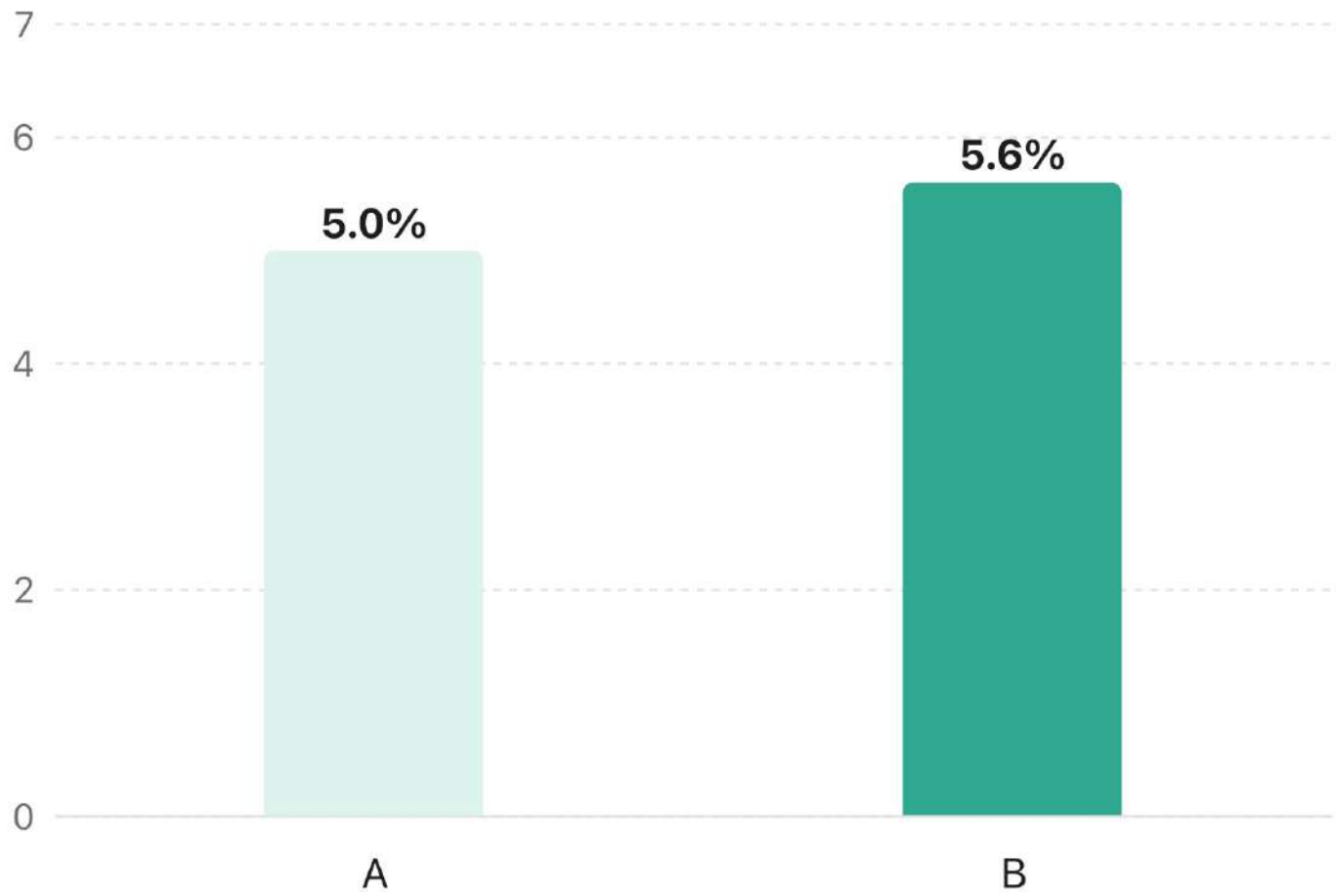


이 결과는 우연일까?



이번에관측된 차이가 정말
의미 있는 차이일까,
아니면 우연히 나온 차이일까?

관측된 구매 전환율 차이



A 구매 전환율

5.0%

500 / 10,000

B 구매 전환율

5.6%

560 / 10,000

관측 차이

+0.6%

우연성 검정 대상

Section 3

가설 검정과 오류

관측된 차이가 정말 의미 있는지 판단하기 위해,
먼저 가설과 오류의 개념을 정리합니다

가설 검정이란? 귀무가설 vs 대립가설



H_0

귀무가설

기존에 사실처럼 인정되고 있던 가설



H_1

대립가설

실험에서 검정하고 싶은 새로운 주장

A/B 테스트에서의 가설

H0

귀무가설

A/B 테스트에서는,

- 개편 전/후 페이지의 구매 전환율은 같다.

H1

대립가설

A/B 테스트에서는,

- 개편 후의 페이지가 더 구매 전환율이 높다.

제 1종 오류, 제 2종 오류의 개념

False Positive

제 1종 오류

귀무가설이 실제로는 참인데,
검정 결과 귀무가설이 기각 됨

- 실제로는 사실이 아닌 대립가설을 효과가 있다 인정
- 위험한 오류

더 심각한 오류

False Negative

제 2종 오류

대립가설이 실제로는 참인데,
검정 결과 대립가설이 인정되지 못함

- 실제로 사실인 대립가설이 인정되지 못함
- 상대적으로 덜 치명적인 오류

A/B 테스트에서의 오류

False Positive

제 1종 오류

귀무가설이 실제로는 참인데,
검정 결과 귀무가설이 기각 됨

- 개편 페이지가 실제로는 구매 전환율을 높이지 않았는데, 효과가 있는것으로 인정

더 심각한 오류

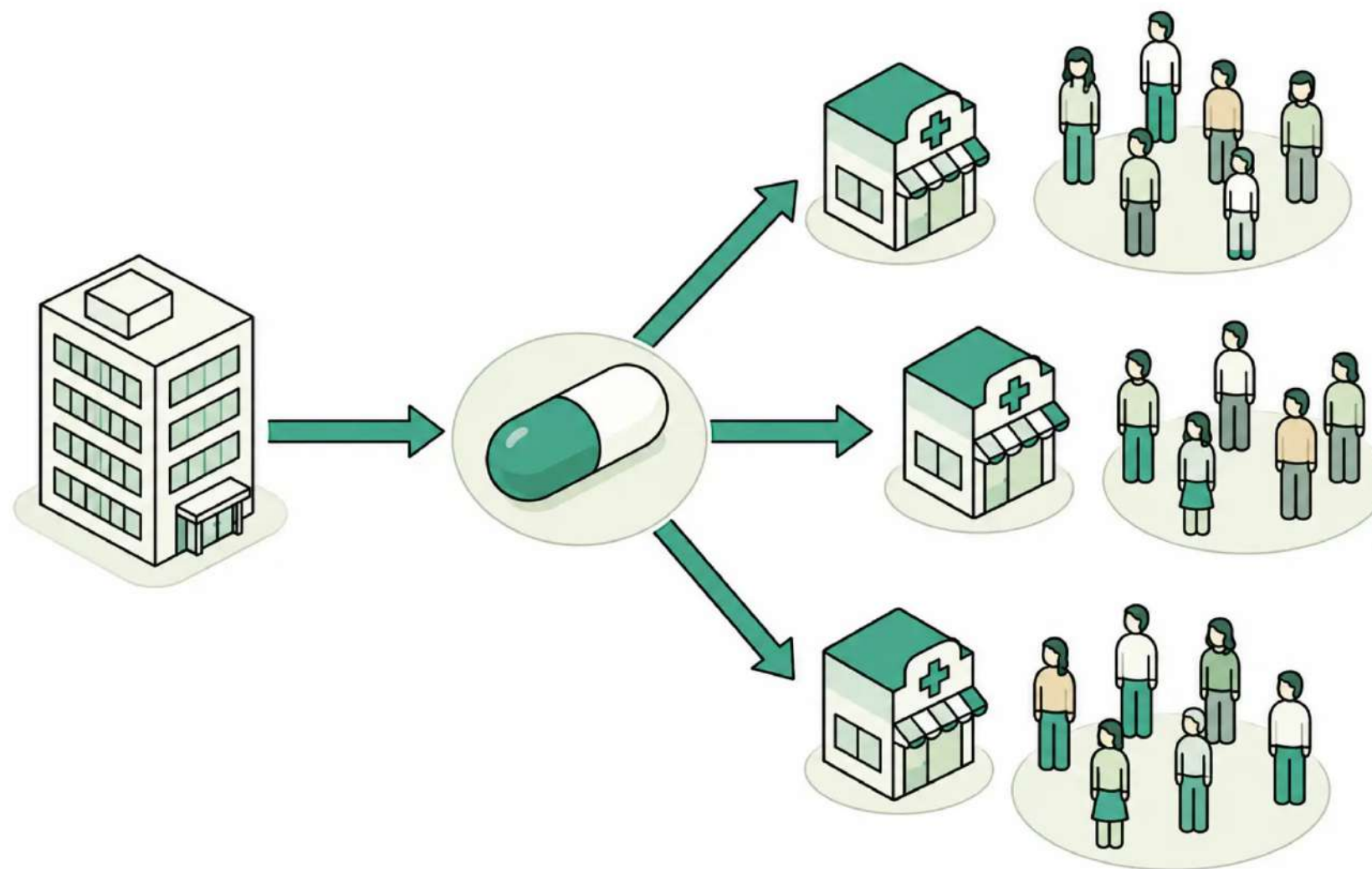
False Negative

제 2종 오류

대립가설이 실제로는 참인데,
검정 결과 대립가설이 인정되지 못함

- 개편 페이지가 실제로는 구매 전환율을 높였는데, 효과가 없는것으로 판정

새로운 감기약 예시



귀무가설: 새로운 감기약은 감기에 효과가 없다.

대립가설: 새로운 감기약은 감기에 효과가 있다.

다시 A/B 테스트 예시로

False Positive

제 1종 오류

귀무가설이 실제로는 참인데,
검정 결과 귀무가설이 기각 됨

- 개편 페이지가 실제로는 구매 전환율을 높이지 않았는데, 효과가 있는것으로 인정

더 심각한 오류

False Negative

제 2종 오류

대립가설이 실제로는 참인데,
검정 결과 대립가설이 인정되지 못함

- 개편 페이지가 실제로는 구매 전환율을 높였는데, 효과가 없는것으로 판정

Section 4

성공을 판단하기

유의수준과 p-value를 비교해
우연으로 보기 어려운 차이인지 판단합니다

4. 성공을 판단하기 - SIGNIFICANCE LEVEL

유의수준 = 공항검색대 스캐너



테스트의 뽀뽀한 정도

0.01

뽀뽀함

0.05

0.1

느슨함

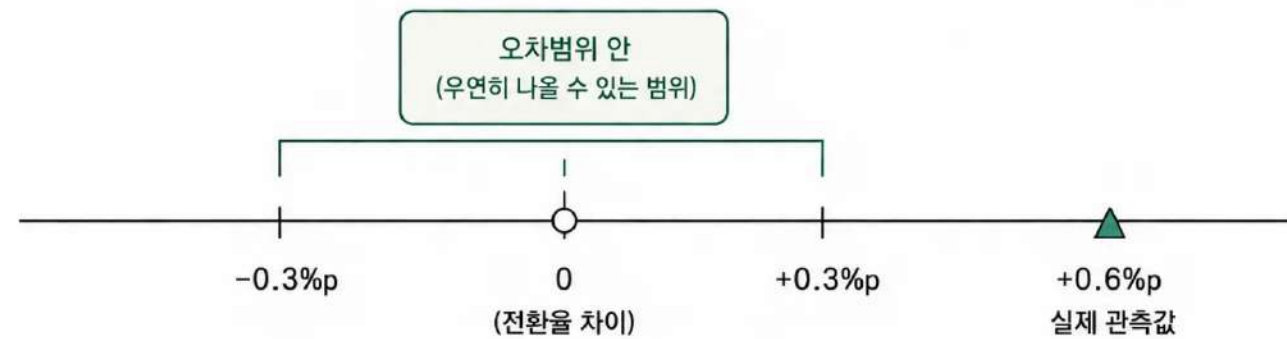
유의수준 5%의 의미

심각한 제 1종 오류를 범하는 경우를 전체의 5%
까지만 허용하겠다.

A/B 테스트로 말하면

전환율이 실제로는 같은데도, 개편 페이지가 더 좋
다고 잘못 판단하는 비율을 정하는 것입니다.

p-value에 대해서



p-value = probability value

이 결과가 우연히도 나타날 수 있는 값인가?

그 정도 현상은 우연히 발생 가능한 일인지 판단하기 위한 수치

p-value 를 구하는 법



측정

실제 A/B 테스트 데이터를 측정합니다.



가정

귀무가설이 맞아 전환율이 같다고 가정합니다.



계산

이번 차이가 오차범위를 넘는 드문 결과인지 계산합니다.

우연이라고 보기 어렵고, 개편 페이지가 영향을 줬다고 볼 근거가 있다.

A/B 테스트의 p-value

p-value = 0.03

기존 페이지와 개편 페이지의 구매 전환율이 실제로는 같다

그럼에도 개편 후 페이지인 B가 A보다 0.6% 높게 나오는 경우가 얼마나 드문지

4. 성공을 판단하기 - DECISION RULE

유의수준과 p-value를 비교합니다

유의수준 = 0.05

p-value = 0.03

SIGNIFICANT

$$0.03 < 0.05$$

귀무가설을 기각함
개편 페이지는 효과가 있다

- 전환율이 같다는 가정 하에서, 이번 결과를 설명하기 어려울만큼 드문 결과가 나왔다.
- 개편 페이지가 구매 전환율을 높였다고 볼 근거가 있다.

p-value = 0.1

NOT SIGNIFICANT

$$0.05 < 0.1$$

귀무가설을 기각하지 못함
개편 페이지의 효과는 아직 알 수 없음

- 이번 정도의 결과는 우연히도 발생 가능하다.
- 이 결과로는 개편 페이지가 효과가 있다고 말하기 어렵다.

용어 다시보기



귀무가설	_____	기존에 인정되고 있던 가설
대립가설	_____	새로 주장하고 싶은 가설
유의수준	_____	테스트의 뽀뽀한 정도
p-value	_____	이 결과가 우연히도 나타날 수 있는 값인가?를 표현한 수치

끝없는 아이디어션

쓸데없는 디자인 아이디어 목록

- | | | | |
|-----------------|---------------|----------------|---------------|
| - 구멍 난 컵 | - 진동 알람 포크 | - 자동 정리 책상 | - 자동 쓰레기통 |
| - 거꾸로 우산 | - 놀이하는 핸드폰 | - 잠 잘 오는 음악 박스 | - 냄새 나는 지우개 |
| - 투명한 돌 | - 재채기 방지 스프레이 | - 기억 식재 지우개 | - 자동 세차 우산 |
| - 붙는 신발 | - 옷감 감지 스피커 | - 기분 맞춤 향초 | - 기분 따라 변하는 옷 |
| - 바람 저장 병 | - 눈물 흡수 티슈 | - 스스로 굴러가는 공 | - 말하는 식물 |
| - 소리 나는 양말 | - 집중 향 초 | - 만능 리모컨 | - 냄새 나는 책상 |
| - 타이머 치약 | - 생각 정리 노트 | - 투명한 신발 | - 냄새 나는 책상 |
| - 자동 잡는 옷 | - 춤추는 화분 | - 온도 조절 컵 | - 냄새 나는 책상 |
| - 달콤한 공기 | - 말하는 벽시계 | - 바람 방향 조절기 | - 냄새 나는 책상 |
| - 그림자 램프 | - 투명한 우산 | - 소리 증폭 안경 | - 냄새 나는 책상 |
| - 점하는 모자 | - 바다 향수 | - 자동 번역 말피 | - 냄새 나는 책상 |
| - 물 먹는 수건 | - 냄새 지우개 | - 햄버거 알림 | - 냄새 나는 책상 |
| - 점자는 인경 | - 고인 정리 상자 | - 무지개 우유 | - 냄새 나는 책상 |
| - 냄새 색연필 | - 스트레스 측정 팔찌 | - 생각하는 의자 | - 냄새 나는 책상 |
| - 자동 넘기 책 | - 에너지 충전 쿠션 | - 마음 읽는 거울 | - 냄새 나는 책상 |
| - 미끄럼 방지 바나나 껍질 | - 소리 캔버스 | - 하늘색 토스터기 | - 냄새 나는 책상 |
| - 햇빛 저장 박스 | - 꿈 기록 베개 | - 소리 먹는 패널 | - 냄새 나는 책상 |
| - 날씨 예측 풍선 | - 시간 멈춤 버튼 | - 시간 여행 티켓 | - 냄새 나는 책상 |
| - 모양 바꾸는 의자 | - 무게 없는 가방 | - 투명 양토 | - 냄새 나는 책상 |
| - 소리 반사 거울 | - 색 바꾸는 조명 | - 꿈 꾸게 하는 인대 | - 냄새 나는 책상 |

(2시간을 향해 달려가는 회의)



오늘 날 잡으셨네

Section 5

표본수와 기간 정하기

검증에 필요한 표본 수를 계산하고, 실제 테스트 기간을 역산합니다

표본수: 전체 데이터를 추정하기 위해 뽑은 샘플

표본수를 정하려면 네 가지 값이 필요합니다



더 미세한 변화를 감지하려면, 더 많은 표본이 필요합니다

기존 전환율	차이	목표 전환율	필요한 표본
5.0%	+1.0%	6.0%	그룹당 약 8,000명
5.0%	+0.6%	5.6%	그룹당 약 21,000명
5.0%	+0.3%	5.3%	그룹당 약 84,000명
5.0%	+0.1%	5.1%	그룹당 약 760,000명

통계학을 공부하고 나니 A/B 테스트는 이렇게 개선할 수 있었습니다

처음 진행한 A/B 테스트

개선된 A/B 테스트

1. 명확한 가설 세우기

명확한 귀무가설, 대립가설 부재

>>>

가설을 먼저 정의

귀무가설: 개편 전/후 페이지의 구매 전환율은 같다.

대립가설: 개편 후 상세페이지가 더 구매 전환율이 높다.

2. 유의수준과 p-value 비교

수치가 상승했으니, 테스트는 성공했다

>>>

판정 기준으로 성공 판단

유의수준을 정하고 실제 데이터 결과로 p-value를 계산해 비교

3. 최소 표본수 정하기

필요한 표본수 알 수 없음. 테스트 기간 2주 지정

>>>

표본 수로 기간 역산

최소 기간 14일 -> 5일. 비용 절감



감사합니다!

발표 자료는 함수랑산악회 노선을 통해 확인하실 수 있습니다.

함바! 🙌