



지하철 경로 선택 A/B 테스트

시뮬레이션 완벽 학습 가 이드

초보자도 이해할 수 있는 상세 설명서

2025년 12월

 목차

1. 프로젝트 소개 - 이 프로젝트가 무엇인가요?

2. 핵심 개념 이해하기

 2.1 A/B 테스트란?

 2.2 로지스틱 회귀란?

 2.3 반복 측정 설계란?

3. 데이터 구조 살펴보기

4. 시뮬레이션 로직 상세 분석

5. 통계 분석 방법론

6. 시각화 결과 해석하기

7. 코드 구조 분석

8. 핵심 인사이트와 결론

9. 용어 사전

1. 프로젝트 소개

이 프로젝트가 무엇인가요?

이 프로젝트는 지하철 이용자들이 경로를 선택하는 행동을 분석하는 A/B 테스트 시뮬레이션입니다. 실제 사용자가 아닌 가상의 데이터(합성 데이터)를 생성하여 분석하는 연구입니다.

🤔 왜 이런 연구를 할까요?

현실에서 지하철을 탈 때, 우리는 항상 선택을 합니다:

- "빠르지만 사람이 많은 길로 갈까?"
- "느리지만 편하게 갈 수 있는 길로 갈까?"

이 프로젝트는 사용자에게 보여주는 **UI(화면 디자인)**가 다르면 선택도 달라지는지를 검증합니다.

📊 실험 설계 개요

| 항목 | 값 | 설명 |
|---------|--------------|---------------------------------|
| 총 사용자 수 | 100,000명 | 가상의 지하철 이용자 |
| 반복 횟수 | 5회 | 각 사용자가 5일 동안 선택 |
| 총 데이터 | 500,000건 | 100,000명 × 5회 |
| 그룹 | A와 B | A: 빠른 경로 강조 UI / B: 편한 경로 강조 UI |
| 경로 옵션 | Fast / Relax | Fast: 빠르지만 혼잡 / Relax: 느리지만 여유 |

💡 핵심 질문 3가지:

- **RQ1:** UI가 다르면 선택이 달라지나요? (A그룹 vs B그룹)
- **RQ2:** 사용자 특성(급함, 성격)이 선택에 영향을 주나요?
- **RQ3:** 반복 경험을 통해 행동이 변하나요? (학습 효과)

2. 핵심 개념 이해하기

2.1 A/B 테스트란?

A/B 테스트는 두 가지 버전을 비교하여 어느 것이 더 효과적인지 확인하는 실험 방법입니다.

⌚ 실생활 예시:

- 쇼핑몰에서 "구매하기" 버튼을 빨간색(A)으로 할지 파란색(B)으로 할지 테스트
- 이메일 제목을 "세일 중!"(A)으로 할지 "50% 할인"(B)으로 할지 테스트

📱 이 프로젝트에서의 A/B 테스트:

| 구분 | A그룹 (실험군) | B그룹 (대조군) |
|-------|--------------|--------------|
| UI 특징 | 빠른 경로 강조 | 편한 경로 강조 |
| 화면 순서 | Fast → Relax | Relax → Fast |
| 강조 정보 | ⌚ 시간 (큰 글씨) | 😊 쾌적도 (아이콘) |
| 색상 | 파란색, 주황색 | 연녹색 계열 |
| 버튼 문구 | "빠르게 가기" | "편하게 가기" |

중요: 사용자는 랜덤으로 A 또는 B 그룹에 배정됩니다. 이렇게 해야 공정한 비교가 가능합니다.

2.2 로지스틱 회귀란?

로지스틱 회귀는 "어떤 일이 일어날 확률"을 예측하는 통계 모델입니다.

이 프로젝트에서의 활용:

"이 사용자가 Fast Route를 선택할 확률은 몇 %일까?"를 계산합니다.

수식 (간단히 설명):

$$P(\text{Fast 선택}) = \text{sigmoid}(\beta_0 + \beta_1 \times \text{시간압박} + \beta_2 \times \text{성격} + \beta_3 \times \text{시간차이} + \beta_4 \times \text{이전선택} + \text{노이즈})$$

여기서 각 β (베타)는 "영향력의 크기"를 나타냅니다:

| 계수 | 값 | 의미 |
|------------------|-------|-------------------------------|
| β_0 (A그룹) | +0.3 | A그룹은 기본적으로 Fast 선택 확률이 높음 |
| β_0 (B그룹) | -0.2 | B그룹은 기본적으로 Fast 선택 확률이 낮음 |
| β_1 (시간압박) | +1.2 | 급할수록 Fast 선택 확률 증가 |
| β_2 (성격) | +0.9 | 효율 지향 성격일수록 Fast 선택 증가 |
| β_3 (시간차이) | +0.16 | 시간 절약 폭이 클수록 Fast 선택 증가 |
| β_4 (이전선택) | -0.7 | 이전에 Fast 선택 → 다음에 Relax 확률 증가 |

2.3 반복 측정 설계란?

반복 측정 설계는 같은 사람을 여러 번 측정하는 실험 방법입니다.

왜 반복 측정을 할까요?

- 한 번만 측정하면 우연한 결과일 수 있음
- 시간에 따른 변화(학습 효과)를 볼 수 있음
- 더 정확한 분석이 가능함

이 프로젝트: 각 사용자가 5일 동안 매일 경로를 선택합니다.

Trial 1 → Trial 2 → Trial 3 → Trial 4 → Trial 5

3. 데이터 구조 살펴보기

3.1 config.py - 모든 설정의 중심

config.py 파일은 시뮬레이션의 모든 설정값을 한 곳에서 관리합니다. 마치 요리의 레시피처럼, 이 값들을 바꾸면 결과가 달라집니다.

| 설정 | 값 | 설명 |
|-----------------------|---------|----------------------|
| NUM_USERS | 100,000 | 시뮬레이션할 사용자 수 |
| NUM_TRIALS | 5 | 각 사용자의 반복 측정 횟수 |
| RANDOM_SEED | 42 | 재현성을 위한 랜덤 시드 |
| FAST_TIME_MEAN | 25분 | Fast Route 평균 소요 시간 |
| RELAX_TIME_MEAN | 36분 | Relax Route 평균 소요 시간 |
| BASE_CONGESTION_FAST | 85% | Fast Route 기본 혼잡도 |
| BASE_CONGESTION_RELAX | 45% | Relax Route 기본 혼잡도 |

3.2 데이터 스키마 (각 열의 의미)

| 컬럼명 | 타입 | 설명 |
|------------------|----|--------------------------|
| user_id | 정수 | 사용자 고유 번호 (1~100,000) |
| assigned_group | 문자 | A 또는 B (배정된 그룹) |
| personality_type | 문자 | 효율지향/편안함지향/중립 |
| trial_number | 정수 | 몇 번째 측정인지 (1~5) |
| time_pressure | 정수 | 시간 압박 (0=급함, 1=보통, 2=여유) |

| | | |
|--------------------|----|-----------------------|
| route_time_fast | 실수 | Fast Route 소요 시간 (분) |
| route_time_relax | 실수 | Relax Route 소요 시간 (분) |
| congestion_fast | 실수 | Fast Route 혼잡도 (%) |
| congestion_relax | 실수 | Relax Route 혼잡도 (%) |
| selected_route | 문자 | 선택한 경로 (Fast/Relax) |
| satisfaction_score | 실수 | 만족도 점수 (0~5) |
| previous_choice | 문자 | 이전 회차 선택 (학습 효과용) |

3.3 사용자 특성 분포

성격 유형 (**Personality Type**):

| 유형 | 비율 | 특징 |
|---------------------|-----|------------------------|
| efficiency-oriented | 55% | 효율을 중시, 빠른 경로 선호 경향 |
| comfort-oriented | 35% | 편안함을 중시, 여유로운 경로 선호 경향 |
| neutral | 10% | 중립적, 상황에 따라 유연하게 선택 |

4. 시뮬레이션 로직 상세 분석

4.1 선택 확률은 어떻게 결정되나요?

각 사용자의 Fast Route 선택 확률은 여러 요소의 합산으로 결정됩니다.

📊 예시로 이해하기:

상황: A그룹에 배정된 효율지향 성격의 사용자가 급한 상황에서 선택

| 요소 | 계산 | 결과 |
|---------------------|------------------|-------|
| 기본값 (A그룹) | $\beta_0 = +0.3$ | +0.3 |
| 시간압박 (급함=2) | 1.2×2 | +2.4 |
| 성격 (효율지향=1) | 0.9×1 | +0.9 |
| 시간차이 (11분) | 0.16×11 | +1.76 |
| 이전선택 (첫회차) | 0 | 0 |
| 합계 (logit) | - | +5.36 |

$\text{logit} = 5.36 \rightarrow \text{sigmoid}(5.36) \approx 99.5\% \text{ 확률로 Fast 선택!}$

4.2 동적 혼잡도 시스템

이 프로젝트의 독특한 점은 혼잡도가 고정되지 않고 변한다는 것입니다.

🔄 동작 원리:

1. Trial 1: 기본 혼잡도 (Fast=85%, Relax=45%)로 시작
2. 많은 사람이 Fast를 선택하면 → Fast 혼잡도 증가!
3. Trial 2~5: 이전 선택 비율에 따라 혼잡도 조정

4. 혼잡도가 높아지면 → 다음에는 Relax 선택 증가

이것이 바로 "학습 효과"입니다. 사용자들이 경험을 통해 행동을 조정합니다.

4.3 만족도 점수 생성 로직

| 조건 | 점수 변화 | 예시 |
|--------------|-------------|---------------|
| 기본 점수 | +3.0 | 모든 사용자의 시작점 |
| 성향-선택 일치 | +1.5 ~ +2.0 | 효율지향이 Fast 선택 |
| 급한데 Relax 선택 | -1.0 | 시간 압박과 불일치 |
| 랜덤 노이즈 | ±0.5 | 개인차 반영 |

5. 통계 분석 방법론

5.1 기초 검정

① Two-Proportion Z-Test (두 비율 검정)

"A그룹과 B그룹의 Fast 선택 비율이 정말 다른가?"를 확인합니다.

| 항목 | 결과 |
|--------------|----------------|
| A그룹 Fast 선택률 | 74.04% |
| B그룹 Fast 선택률 | 68.33% |
| 차이 | 5.71%p |
| z-통계량 | 44.60 |
| p-value | < 0.001 (거의 0) |

해석: p-value가 0.05보다 훨씬 작으므로, 두 그룹의 차이는 우연이 아닙니다. A그룹 UI가 Fast 선택을 더 유도한다고 결론 내릴 수 있습니다.

② Chi-square Test (카이제곱 검정)

"그룹(A/B)과 선택(Fast/Relax)이 서로 관련이 있는가?"를 확인합니다.

| | Fast 선택 | Relax 선택 | 합계 |
|-----|---------|----------|---------|
| A그룹 | 184,851 | 64,819 | 249,670 |
| B그룹 | 171,038 | 79,292 | 250,330 |
| 합계 | 355,889 | 144,111 | 500,000 |

$$\chi^2 = 1,988.49, p < 0.001 \rightarrow \text{그룹과 선택은 독립이 아닙니다 (관련 있음)}$$

③ Effect Size (효과 크기, Cohen's h)

통계적으로 유의미해도, 실제로 얼마나 큰 차이인가?를 확인합니다.

| Cohen's h 값 | 해석 |
|-------------|----------------|
| < 0.2 | small (작은 효과) |
| 0.2 ~ 0.5 | medium (중간 효과) |
| > 0.5 | large (큰 효과) |

이 프로젝트의 Cohen's h = **0.126 (small)**

효과 크기는 작지만, 500,000건 데이터에서 일관되게 나타나므로 실용적으로 의미 있습니다.

5.2 GEE 분석

GEE(Generalized Estimating Equations)는 반복 측정 데이터를 분석하는 고급 통계 기법입니다.

🧐 왜 GEE를 사용할까요?

같은 사람의 5번 선택은 서로 연관되어 있습니다 (오늘 Fast 선택한 사람은 내일도 Fast 선택할 가능성 높음). 일반 통계 방법은 이 "상관관계"를 무시 하지만, GEE는 이를 고려합니다.

📊 GEE 분석 결과:

| 변수 | 계수 (β) | 해석 |
|---------------|----------------|-----------------------|
| group (A=1) | +0.33 | A그룹은 Fast 선택 확률 더 높음 |
| time_pressure | +0.94 | 급할수록 Fast 선택 확률 증가 |
| personality | +0.60 | 효율지향일수록 Fast 선택 확률 증가 |

| | | |
|-----------------|--------|-----------------------------|
| trial_index | -0.40 | Trial이 지날수록 Fast 선택 감소 (학습) |
| time_diff | +0.13 | 시간 절약이 클수록 Fast 선택 증가 |
| congestion_diff | -0.009 | 혼잡도 차이가 클수록 Fast 선택 감소 |

5.3 FDR 보정

여러 개의 통계 검정을 동시에 하면 "우연히 유의미"하게 나올 확률이 높아집니다. FDR(False Discovery Rate) 보정은 이 문제를 해결합니다.

결과: 모든 변수가 FDR 보정 후에도 유의미함 ($p < 0.05$)
이 결과들은 우연이 아니라 실제 효과입니다!

6. 시각화 결과 해석하기

이 프로젝트에서 생성된 8개의 차트를 하나씩 살펴봅니다.



차트 1: A/B 그룹 비교

내용: A그룹(74.04%)과 B그룹(68.33%)의 Fast 선택률 비교

핵심: A그룹이 약 5.7%p 더 높은 Fast 선택률을 보임

의미: "빠른 경로 강조" UI가 실제로 Fast 선택을 유도함



차트 2: Personality 유형별 분석

| Personality | A그룹 | B그룹 |
|---------------------|-------|-------|
| efficiency-oriented | 82.5% | 77.3% |
| neutral | 72.4% | 66.7% |
| comfort-oriented | 61.4% | 54.8% |

의미: 모든 성격 유형에서 A그룹이 더 높은 Fast 선택률 → UI 효과는 보편적!



차트 3: Trial별 변화 (학습 효과)

| Trial | Fast 선택률 | 변화 |
|-------|----------|-----------------|
| 1 | 92.2% | 초기: 대부분 Fast 선택 |
| 2 | 74.8% | 혼잡 경험 후 감소 |
| 3 | 57.1% | 최저점 (혼잡 피크) |
| 4 | 66.1% | 균형점 찾기 |

| | | |
|---|-------|-----|
| 5 | 65.7% | 안정화 |
|---|-------|-----|

학습 효과 확인: Trial 1에서 대부분 Fast를 선택했지만, 혼잡을 경험하면서 점차 Relax 선택이 증가 → 균형점에서 안정화

차트 4: Time Pressure × Personality 히트맵

| | 여유(0) | 보통(1) | 급함(2) |
|-------|-------|-------|-------|
| 효율지향 | 73.7% | 79.1% | 94.5% |
| 중립 | 62.5% | 68.2% | 91.4% |
| 편안함지향 | 49.6% | 56.5% | 83.7% |

핵심 인사이트: "급함(2)" 상태에서는 모든 성격 유형이 80% 이상 Fast 선택!
시간 압박이 성격보다 더 강력한 영향을 미침.

차트 5: GEE 계수 시각화

각 변수의 영향력 크기를 막대 그래프로 표시합니다.

파란색(양수): Fast 선택 확률 증가

빨간색(음수): Fast 선택 확률 감소

가장 큰 영향: time_pressure (+0.94) → 급함이 가장 중요한 요인!

차트 6: 만족도 분포

결과: A그룹과 B그룹의 만족도 분포가 거의 동일

의미: UI가 달라도 최종 만족도에는 큰 차이 없음 → 사용자는 자신의 성향에 맞는 선택을 하면 만족함



차트 7: 혼잡도 동적 변화

Trial 1: Fast 혼잡도 85% → Trial 2: Fast 혼잡도 196% (급증!)

많은 사람이 Fast를 선택하면서 혼잡도가 폭발적으로 증가. 이후 사용자들이 Relax로 분산되면서 점차 안정화.



차트 8: 학습 효과 분석

초기 | **Trial (1회차)**에서 **Fast** 선택 → 후기 | **Trial (2회차)**에서:

- Fast 유지: 60.32%
- Relax로 전환: 39.68%

초기 | **Trial**에서 **Relax** 선택 → 후기 | **Trial**에서:

- Fast로 전환: 80.77%
- Relax 유지: 19.23%

해석: Relax 경험자는 다음에 Fast 시도, Fast 경험자 중 일부는 Relax로 전환

7. 코드 구조 분석

7.1 프로젝트 파일 구조

| 파일명 | 역할 | 주요 기능 |
|--------------------|--------|---------------------------------|
| config.py | 설정 관리 | 모든 파라미터를 중앙에서 관리 |
| generate_users.py | 사용자 생성 | 100,000명 사용자 기본 정보 생성 |
| simulate_trials.py | 시뮬레이션 | 5회 반복 측정 데이터 생성 |
| basic_tests.py | 기초 통계 | Z-Test, Chi-square, Effect Size |
| mixed_models.py | 고급 통계 | GEE, FDR 보정 |
| visualization.py | 시각화 | 8개 차트 생성 |
| app.py | 대시보드 | Streamlit 웹 인터페이스 |

7.2 개발 일정 (DAY 1-6)

| 일차 | 주요 작업 | 산출물 |
|-------|----------------|---------------------------------|
| DAY 1 | 환경 설정, 사용자 생성 | config.py, users_base.csv |
| DAY 2 | 반복 측정 시뮬레이션 | trials_data_partial.csv |
| DAY 3 | 선택 모델링, 데이터 완성 | synthetic_data.parquet |
| DAY 4 | 통계 분석 구현 | basic_tests.py, mixed_models.py |
| DAY 5 | 시각화 및 리포트 | visualization.py, 8개 차트 |
| DAY 6 | 대시보드 및 마무리 | app.py (Streamlit) |

8. 핵심 인사이트와 결론

8.1 연구 질문 답변

| 연구 질문 | 결과 | 결론 |
|------------------|---------------------------------|--|
| RQ1: UI가 선택에 영향? | A그룹 74%, B그룹 68% $p < 0.001$ | <input checked="" type="checkbox"/> 예, 유의미한 영향 |
| RQ2: 사용자 특성 영향? | 시간압박, 성격 모두 유의미 | <input checked="" type="checkbox"/> 예, 주요 예측 요인 |
| RQ3: 학습 효과 존재? | Trial 1: 92% → Trial 5: 66% | <input checked="" type="checkbox"/> 예, 확실한 학습 효과 |

8.2 실무 적용 제안

1. Dynamic UI 전략

- 출퇴근 러시아워: A안 (빠른 경로 강조)
- 주말/여가 시간: B안 (편안함 강조)

2. 개인화 추천

- 효율지향 사용자 → Fast Route 우선 표시
- 편안함지향 사용자 → Relax Route 우선 표시

3. 혼잡도 분산 전략

- 실시간 혼잡도 정보 제공
- Relax Route 선택 시 인센티브 (마일리지 등)

8.3 한계점과 향후 연구

| | |
|-----|----------|
| 한계점 | 향후 연구 방향 |
|-----|----------|

| | |
|-----------|------------------|
| 시뮬레이션 데이터 | 실제 사용자 A/B 테스트 |
| 2개 경로만 제공 | 3개 이상 경로 (다중 선택) |
| 단순 학습 모델 | 비선형 학습 효과 모델링 |
| 개인 특성 제한적 | 더 다양한 변수 추가 |

9. 용어 사전

| 용어 | 설명 |
|-------------|---|
| A/B 테스트 | 두 가지 버전을 비교하여 효과를 측정하는 실험 |
| 로지스틱 회귀 | 결과가 0 또는 1인 이진 분류를 위한 통계 모델 |
| p-value | 결과가 우연히 발생할 확률. 0.05 미만이면 유의미 |
| Effect Size | 효과의 실질적 크기. 통계적 유의성과 별개 |
| GEE | Generalized Estimating Equations. 반복 측정 데이터 분석용 |
| FDR | False Discovery Rate. 다중 검정 시 오류 보정 방법 |
| Cohen's h | 두 비율 간 효과 크기 측정 지표 |
| 반복 측정 | 같은 대상을 여러 번 측정하는 연구 설계 |
| 합성 데이터 | 실제 데이터가 아닌 시뮬레이션으로 생성된 데이터 |
| 혼잡도 | 경로의 블빔 정도 (%). 100%가 기준 |
| logit | 로그 오즈. 로지스틱 회귀의 중간 계산값 |
| sigmoid | S자 형태 함수. logit을 확률(0~1)로 변환 |
| Trial | 반복 측정의 각 회차 (이 프로젝트에서는 1~5) |
| 랜덤 시드 | 재현 가능한 랜덤 숫자 생성을 위한 초기값 |

 참고 자료

- Kohavi, R. "Trustworthy Online Controlled Experiments" (A/B 테스트 바이블)
- Kahneman, D. "Thinking, Fast and Slow" (행동경제학)
- Statsmodels Documentation (GEE 분석)

———— 문서 끝 ———