

5 장

종합 실습 문제 답안 샘플

※ 다음 답안 샘플은 2026년 6월 25일부터 7월 7일 기간 동안 클로드 **Fable 5**로 테스트한 것입니다. 사용하는 모델과 환경에 따라 답은 달라질 수 있음을 참고하세요.

실습 데이터: StreamAI+_did_data.csv

-사용자 10,000명(실험군 A 5,000명, 대조군 B 5,000명) × 6개월(2025년 1~6월)의 사용자

-월 패널 60,000행.

-변수: user_id, month, is_treated, is_post, user_age, membership_years, avg_play_time(평균 일일 재생시간, 분).

-신규 AI 기능은 2025년 4월 1일에 A그룹에만 공개되었다.

답안 활용 시 유의사항: 생성형 AI 결과의 변동성에 관하여

이 장의 과제는 이 책의 실습 중 결과 변동이 가장 작은 편에 속한다. 2×2 집단 평균, 이중차분값, 회귀계수는 모두 닫힌 계산이며, 특히 상호작용항만 포함한 DiD 회귀는 포화(saturated) 모형이어서 네 계수가 2×2 평균표와 대수적으로 정확히 일치한다. 따라서 어떤 AI를 쓰든, R의 `lm()`이든 Python의 `statsmodels`이든 계수는 소수점 반올림 수준까지 같아야 하며, 다르다면 데이터 로딩이나 변수 정의의 오류를 의심해야 한다.

AI 간 차이가 정당하게 발생하는 지점은 세 곳이다. (1) 표준오차의 계산 방식(고전적 vs robust vs 사용자 클러스터)에 따라 t값과 p값이 달라진다(이 데이터에서는 어느 방식이든 $p < 0.001$ 로 결론 동일). (2) 평행추세의 확인 방법(시각적 판단 vs 형식적 검정 vs event-study)이 다를 수 있다. (3) ROI 계산에서 월의 정의(30일 vs 실제 일수)와 확대 대상의 해석에 따라 금액이 달라진다. 채점의 기준은 이 세 지점에서 자신의 선택을 명시했는지와, 이중차분의 논리(집단 차이와 시간 효과의 동시 제거)를 올바르게 적용했는지이다.

(※단계 1과 단계 2의 답안 샘플이 함께 제시되어 있습니다.)

■ 단계 1. 인과 추론 관점에서 질문 재정의

■ 단계 2. 집단별 평균 비교와 DiD 값 계산

[단계 1,2의 문제 1,2에 대한 답안 샘플]

이번 분석에서 처치(treatment)는 2025년 4월 1일에 A그룹 사용자에게만 신규 AI 기능이 공개된 것이다. 따라서 실험군(treated group)은 $is_treated = 1$ 인 A그룹, 대조군(control group)은 $is_treated = 0$ 인 B그룹이다. is_post 는 기능 도입 이후 여부를 나타내는 시점 변수로, 도입 전(1~3월)은 $is_post = 0$, 도입 후(4~6월)은 $is_post = 1$ 이다.

이 두 변수가 모두 필요한 이유는, 단순히 "A그룹이 B그룹보다 재생시간이 길다"는 사실만으로는 기능 효과를 판단할 수 없기 때문이다. 실제로 데이터에서도 도입 전부터 A그룹의 평균 재생시간이 B그룹보다 약 1.9분 높다. 따라서 집단간 원래 차이(선택 효과)와, 시간 경과에 따라 모두에게 공통으로 나타나는 변화(시간 효과)를 분리해야 한다. 누가 처치를 받았는가($is_treated$)와 언제가 처치 이후인가(is_post)가 함께 있어야만 "처치를 받은 집단이 처치 이후에 얼마나 추가로 변했는가"를 계산할 수 있다. DiD는 [실험군의 전후 변화 - 대조군의 전후 변화]를 구함으로써, 기능 도입이 없었다더라도 발생했을 공통 시간 효과를 제거하고 신규 AI 기능의 순수 효과를 추정한다.

집단	도입 전 ($is_post=0$)	도입 후 ($is_post=1$)	변화량
대조군 B ($is_treated=0$)	29.6425분	32.0160분	+2.3735분
실험군 A ($is_treated=1$)	31.5652분	38.9985분	+7.4333분

〈표 1〉 집단 × 시점별 평균 일일 재생시간 (2×2)

$$DiD = (38.9985 - 31.5652) - (32.0160 - 29.6425) = 7.4333 - 2.3735 = 5.0598$$

즉, 신규 AI 기능으로 인해 순수하게 추가된 평균 일일 재생시간은 약 5.06분이다. 단순히 실험군의 전후 차이인 +7.43분만 보면 효과가 매우 커 보이지만, 같은 기간 대조군도 +2.37분 증가했다. 이는 계절성, 콘텐츠 변화, 전반적 플랫폼 성장, 마케팅 등 기능 외의 공통 변화가 있었음을 보여준다. 따라서 대조군의 변화량을 빼주어야만 기능 도입으로 추가 발생한 증가분을 추정할 수 있다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

〈표 1〉의 네 평균과 DiD 값은 닫힌 계산이므로 어떤 AI를 쓰든 동일해야 하며, 반올림 자릿수(5.06 vs 5.0598)만 다를 수 있다. 값이 다르면 is_post의 월 구분(1~3월 vs 4~6월)이나 집계 단위(사용자-월 관측치 평균 vs 사용자별 평균의 평균 — 이 데이터는 균형 패널이라 둘이 일치)의 오류를 확인해야 한다.

■ 단계 3. 평행 추세(Parallel Trends) 가정 점검 계획

[문제 1,2에 대한 답안 샘플]

회귀식은 다음과 같다.

$$Y_{it} = \beta_0 + \beta_1 \cdot is_treated_i + \beta_2 \cdot is_post_t + \beta_3 \cdot (is_treated_i \times is_post_t) + \epsilon_{it}$$

여기서 Y_{it} 는 사용자 i 의 시점 t 에서의 평균 일일 재생시간(avg_play_time)이다. 실제 데이터로 추정한 계수는 다음과 같다.

계수	추정치	의미
β_0 (절편)	29.6425	기준집단, 즉 대조군(B)의 도입 전 평균 재생시간
β_1 (is_treated)	1.9226	도입 전 시점에서 실험군(A)이 대조군(B)보다 평균 1.92분 더 많이 재생 (집단 고정 차이)
β_2 (is_post)	2.3735	대조군의 도입 전후 증가분. AI 기능과 무관한 공통 시간 효과
β_3 (상호작용)	5.0599	이중차분 추정치. AI 기능 도입으로 실험군의 재생시간이 대조군 대비 약 5.06분 추가 증가

〈표 2〉 DiD 회귀 추정 결과

상호작용항 β_3 는 통계적으로 유의하다(사용자 수준 클러스터 표준오차 0.032, $p < 0.001$). 비즈니스 관점에서 이는 다음을 의미한다. AI 기능이 단순한 상관관계 수준이 아니라 실제로 사용자 재생시간을 늘리는 데 기여했으며, 사용자 참여도 향상에 효과적인 기능으로 평가할 수 있고, 전사 확대·유사 기능 투자·개인화 기능 강화 등의 의사 결정에 실증적 근거를 제공한다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

계수는 일치해야 하고, 표준오차는 달라질 수 있다. 상호작용항만 포함한 DiD 회귀는 포화 모형이므로 네 계수가 〈표 1〉의 2×2 평균에서 대수적으로 유도된다($\beta_{\square} = \text{DiD 값}$). 따라서 계수가 본 답안과 다르면 오류다. 반면 표준오차는 계산 방식에 따라 달라진다: 이 데이터에서 고전적·robust 표준오차는 0.034, 사용자 클러스터 표준오차는 0.032이다. 같은 사용자가 6개월 반복 관측되는 패널이므로 사용자 수준 클러스터가 방법론적 정석이나, 이 데이터에서는 어느 방식이든 $p < 0.001$ 로 결론이 같으므로 어떤 표준오차를 보고했든 정답으로 인정한다(클러스터의 필요성을 언급한 답안은 우수 답안). 공변량(user_age, membership_years)을 추가해도 $\beta_{\square}$ 는 5.0599로 사실상 변하지 않는데, 이는 처치가 사용자 특성과 무관하게 배정되었음을 시사하는 강건성 확인으로 가산 요인이다.

오답 기준: (1) 상호작용항 없이 is_treated와 is_post만 넣은 회귀에서 is_treated 계수를 기능 효과로 해석한 경우, (2) 도입 후 두 집단의 단순 차이(약 6.98분)나 실험군의 전후 차이(7.43분)를 인과효과로 보고한 경우는 DiD의 논리(선택 효과·시간 효과의 제거)를 위반한 것이므로 틀린 것으로 처리한다.

■ 단계 4. 시각화를 통한 평행추세 검정

[문제 1,2에 대한 답안 샘플]

도입 전 1~3월의 실제 월별 평균 재생시간은 다음과 같다.

월	대조군 B	실험군 A	격차(A-B)
1월	28.8602	30.7328	1.8726
2월	29.6286	31.5808	1.9522
3월	30.4387	32.3819	1.9432

〈표 3〉 도입 전 월별 평균 재생시간

두 집단 모두 1월에서 3월로 갈수록 재생시간이 꾸준히 증가하고, 집단 간 격차도 약 1.9분 내외로 거의 일정하게 유지된다. 따라서 도입 전 그래프는 수준(level) 차이는 있으나 기울기(slope)는 대체로 유사한, 거의 평행한 형태를 보인다.

질문 1 답.

DiD 분석이 타당하려면 도입 전 두 집단의 그래프가 유사한 기울기를 유지해야 한다. 두 집단의 절대 수준이 꼭 같을 필요는 없지만, 시간에 따라 증가하거나 감소하는 속도는 비슷해야 한다. 즉 선 두 개가 떨어져 있더라도 평행하게 움직여야 평행추세 가정이 대체로 성립한다고 볼 수 있다. 수준 차이는 β_0 이 흡수하지만, 기울기 차이는 어떤 항도 흡수하지 못하고 β_1 에 섞여 들어가기 때문이다.

질문 2 답.

만약 도입 전부터 실험군(A)의 재생시간이 대조군(B)보다 훨씬 가파르게 상승하고 있었다면, 실험군은 원래부터 더 빠르게 성장하는 집단이었을 수 있다. 이 경우 도입 후 관찰된 추가 증가분에는 AI 기능의 효과뿐 아니라 원래 존재하던 상승 추세까지 포함되므로, 단계 3의 β_1 는 순수한 기능 효과보다 크게 계산되어 과대평가된다.

실제 데이터 해석.

이 데이터에서는 도입 전 두 집단의 평균 재생시간이 비슷한 속도로 증가하고 격차도 거의 일정하다. 시각적 판단을 보완하기 위해 도입 전 기간(1~3월)만으로 $is_treated \times month$ 상호작용을 추정하면 기울기 차이는 월 0.035분에 불과하고 통계적으로 유의하지 않다($p = 0.216$). 따라서 평행추세 가정은 시각적으로도, 형식적 검정으로도 지지되며, 이 점이 DiD 결과의 신뢰성을 높여준다. 참고로 도입 후 3개월(4~6월)의 집단 간 격차도 7.03분, 6.93분, 6.99분으로 안정적이어서 효과가 일시적 반짝임이 아니라 지속적임을 보여준다.

결과의 변동 가능성 (시에 따라 다르게 나올 수 있는 부분)

평행추세의 확인 방법은 시에 따라 다르게 제시될 수 있다: (1) 도입 전 월별 평균의 시각적 비교(본 답안의 기본), (2) 도입 전 기간의 $is_treated \times month$ 기울기 검정(본 답안의 보완, $p = 0.216$), (3) 월별 더미와의 상호작용을 모두 추정하는 event-study 그림. 세 접근 모두 정답이며, 사전 기간이 3개월뿐이라는 한계(검정력이 낮아 '시각 실패 = 가정 성립의 증명'이 아님)를 지적한 답안은 우수 답안이다. 시각적 판단만 제시했더라도 문제의 요구(시각화를 통한 검정)에 부합하므로 감점하지 않는다.

■ 단계 5. 전략적 의사결정과 ROI 계산

[답안 샘플]

앞선 분석에서 신규 AI 기능은 사용자 1인당 일일 평균 재생시간을 약 5.06분 증가시키는 인과적 효과를 보였다. 주어진 조건은 재생시간 10분당 추가 광고 수익 100원, 유저당 월간 AI 운영 비용 400원이다. 월을 30일로 두고 계산하면 다음과 같다.

항목	계산	값
① 유저 1인당 월간 추가 재생시간	$5.0598\text{분} \times 30\text{일}$	151.794분
② 유저 1인당 월간 추가 광고 수익	$151.794 \div 10 \times 100\text{원}$	1,517.94원
③ 유저 1인당 월간 기대 순이익	$1,517.94 - 400\text{원}$	1,117.94원
④ 전체 확대 시 월간 기대 순이익	$1,117.94\text{원} \times 10,000\text{명}$	약 11,179,400원

〈표 4〉 ROI 계산

즉 전사 확대 시 월간 약 1,118만 원의 기대 순이익을 얻을 수 있다. 유저당 수익(1,517.94원)이 운영비(400원)의 약 3.8배이므로, 효과 추정치가 다소 보수적으로 조정되더라도 흑자 구조는 유지된다.

최종 요약.

대조군 전후 변화 +2.37분, 실험군 전후 변화 +7.43분, 이중차분 추정치 +5.06분($\beta_3 = 5.0599$, $p < 0.001$), 전사 확대 시 월간 기대 순이익 약 1,118만 원. 도입 전 추세가 시각적·통계적으로 평행하고 공변량 추가에도 추정치가 불변이므로, 이 결과는 단순 상관이 아니라 비교적 신뢰할 수 있는 인과효과 추정치로 해석할 수 있다. 따라서 신규 AI 기능은 사용자 참여와 수익성 모두에서 전사 확대 가치가 있다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

ROI 금액은 두 가지 선택에 따라 달라진다. 첫째, 월의 정의: 30일 기준이면 1인당 순이익 1,117.94원, 평균 월 길이 (30.44일) 기준이면 약 1,140원, 4~6월의 실제 일수를 쓰면 월별로 조금씩 다르다. 어느 기준이든 명시했다면 정답이

다. 둘째, 확대 대상의 해석: '전체 사용자 10,000명 기준'이면 월 약 1,118만 원이고, '아직 기능이 없는 대조군 5,000명에게 추가 확대'로 해석하면 신규 발생 순이익은 그 절반인 약 559만 원이다(실험군 5,000명의 이익은 이미 실현 중이므로). 문제의 기본 해석은 전자이나, 후자의 관점을 구분해 제시한 답안은 증분(incremental) 개념을 정확히 이해한 것이므로 우수 답안으로 평가할 수 있다.

한 가지 유의할 점은 이 계산이 '효과의 지속'을 가정한다는 것이다. 이 데이터에서는 도입 후 3개월간 격차가 약 7.0분으로 안정적이어서 가정이 지지되지만, 실제 서비스에서는 신기능 효과가 시간이 지나며 감소(novelty effect)할 수 있음을 한계로 지적한 답안은 가산 요인이다.