

7 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.

답안 활용 시 유의사항: 생성형 AI 결과의 변동성에 관하여

본 해설의 수치는 첨부 데이터로 재현 가능한 값이다. 다만 시계열 예측 과제는 결과가 소프트웨어에 의존하는 정도가 특히 크다. 기술통계, ACF, 단위근 검정처럼 닫힌 계산은 어떤 AI를 쓰든 동일하게 나와야 하지만, 모형 차수의 선택, 예측치와 예측구간, 그리고 그로부터 파생되는 안전재고·발주량·품절량·비용은 사용한 라이브러리(R의 forecast/fable, Python의 statsmodels/Prophet)와 그 버전, 하이퍼파라미터, 심지어 난수에 따라 달라진다. 특히 Prophet의 예측구간은 몬테카를로 시뮬레이션으로 산출되므로 같은 코드를 두 번 실행해도 구간 경계가 수 개~수십 개 단위로 달라지는 것이 정상이다.

이 과제의 또 다른 특징은 오차의 연쇄다. 단계 2의 예측치가 조금 달라지면 단계 3의 리드타임 수요, 재발주점, 발주량, 품절량, 비용이 모두 연쇄적으로 달라진다. 따라서 단계 3의 숫자가 본 해설과 다른 것은 대부분 단계 2의 정당한 차이가 전파된 결과이며, 그 자체로 오답이 아니다.

채점의 기준은 숫자의 일치가 아니라 (1) 분석 절차의 논리(계절성 진단 → 모형 비교 → 예측 → 불확실성을 반영한 재고 계산)가 올바른지, (2) 각 단계에서 자신이 선택한 방법의 근거를 제시했는지, (3) "현 재고로는 성수기 진입 구간의 품절이 구조적으로 불가피하다"는 핵심 결론에 도달했는지이다. 다만 모든 차이가 허용되는 것은 아니며, 통계적으로 잘못된 설정(예: 안전재고에 원계열 표준편차를 무비판적으로 사용)은 각 단계의 주황색 상자에 명시된 기준에 따라 구분하여 평가한다.

■ 단계 0. 문제 개요와 접근 전략

이 문제는 수요 예측과 재고 최적화를 하나의 의사결정 흐름으로 연결한다. 분석의 최종 산출물은 "향후 8주 동안 언제, 얼마를 발주해야 품질을 막을 수 있는가"라는 운영 질문에 대한 답이다. 따라서 단계 1의 탐색과 단계 2의 예측은 그 자체가 목적이 아니라, 단계 3의 재고 계산에 투입될 수요의 기댓값과 불확실성을 추정하기 위한 과정이다.

데이터의 마지막 관측 시점이 봄(4월)이라는 점에 주목한다. 향후 8주(157~164주)는 여름 성수기로 진입하는 수요 급증 구간이며, 현 재고 1,200개는 이 구간의 한 주 수요에도 못 미친다. 즉 발주 타이밍이 조금만 늦어도 즉시 품질로 이어지는 구조다.

■ 단계 1. 데이터 탐색과 시계열 분해

[문제 1에 대한 답안 샘플]

1-1. 기술통계

프롬프트 예시

첨부한 글로우마켓 선크림 주별 판매 데이터의 기술통계량(평균, 중앙값, 표준편차, 변동계수, 최소/최대값과 그 시점)을 계산하고, 데이터의 전반적 특성을 요약해줘.

AI 결과

통계량	값
관측 수	156주
평균	966개
중앙값	938개
표준편차	412개

변동계수(CV)	0.43
최솟값	390개 (2024-01-01)
최댓값	1,811개 (2025-07-07)

〈표 1〉 주별 판매량 기술통계

평균(966)이 중앙값(938)보다 다소 크고 변동계수가 0.43에 이른다. 이는 평상시 수요 주위로 여름철 고수요 구간이 길게 분포하는, 우측으로 약간 치우친 분포임을 시사한다. 최솟값은 한겨울(1월 초), 최댓값은 한여름(7월 초)에 나타나 강한 연중 계절성을 뒷받침한다.

해설.

변동계수가 0.4를 넘는다는 것은 수요 변동이 평균의 40% 수준이라는 뜻으로, 안전재고 없이는 품질 위험이 큰 품목임을 일찍부터 알려 준다.

결과의 변동 가능성 (SI에 따라 다르게 나올 수 있는 부분)

이 문항의 수치는 닫힌 계산이므로 어떤 SI를 쓰든 동일해야 한다. 학생 답안의 기술통계가 위 표와 다르다면 변동성의 문제가 아니라 데이터 로딩 오류(인코딩, 열 지정, 결측 처리)일 가능성이 높으므로 확인이 필요하다. 단, 표준편차는 SI가 표본식($n-1$)과 모집단식(n) 중 무엇을 쓰는지에 따라 값이 약간 다를 수 있다(412.4 vs 411.1).

1-2. 시각화와 추세·계절 패턴

원계열에는 (i) 매년 7월 초를 정점으로 12월 말 저점을 찍는 1년 주기 계절성, (ii) 2년차 후반부터 가팔라지는 상승 추세, (iii) 2024-02-12의 단발성 이상치(판매 1,141개, 인접 5주 평균 610개의 약 1.9배)가 함께 나타난다. 추세 가속은 문제에서 언급된 SNS 마케팅 강화 시점과 일치한다.

[문제 2에 대한 답안 샘플]

프롬프트 예시

이 주별 데이터를 STL로 분해해줘. 주별 데이터의 연간 계절성을 보려면 주기를 어떻게 설정해야 하는지 먼저 설명하고, 추세·계절·잔차 성분을 각각 그려줘. 계절성 강도와 추세 강도도 계산해줘.

AI 결과

주별 데이터에서 1년 주기는 약 52주이므로 period=52로 설정한다. STL 분해 결과는 다음과 같다.

성분	요약
추세	약 869개 → 1,121개로 상승 (약 +29%), 2년차 후반부터 가속
계절	진폭 약 1,417개, 7월 초 정점 · 12월 하순 저점
잔차	표준편차 약 31개 (이상치 시점 제외 시)
계절성 강도 F_s	0.99
추세 강도 F_t	0.83

〈표 2〉 STL 분해 요약

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

STL은 robust 옵션과 계절 창(seasonal window) 설정에 따라 성분 추정치가 조금씩 달라진다. 예컨대 robust 옵션을 켜면 추세 종점이 1,121개 대신 1,125개로, 계절 진폭이 1,417개 대신 1,431개로, 추세 강도가 0.83 대신 0.84로 나올 수 있다. 이 수준의 차이는 모두 정답 범위다. $F_s \approx 0.99$ (매우 강한 계절성), $F_t \approx 0.8$ 대(강한 추세)라는 정성적 판정에 도달했는지를 본다.

이상치 진단도 분해 옵션에 따라 다른 시점이 지목될 수 있으므로, 잔차 진단에만 의존하지 말고 원자료를 직접 확인하는 것이 안전하다. 이 데이터에서 원자료 기준의 명백한 단발성 이상치는 2024-02-12(1,141개, 인접 주 대비 약 2배) 하나다. 겨울 비수기의 고립된 스파이크이므로 일회성 프로모션이나 기록 오류로 해석하는 것이 자연스럽다.

■ 단계 2. 모형 구축과 비교

[문제 1에 대한 답안 샘플]

프롬프트 예시

SARIMA 적용을 위해 정상성을 점검하려고 해. ADF·KPSS 검정을 수행하고, 그 결과만으로 계절 차분 필요성을 판단할 수 있는지 설명해줘. ACF도 함께 보고 d, D, s를 결정해줘.

AI 결과

검정	원계열(로그)	계절차분 후
ADF p값	< 0.001	< 0.001
KPSS p값	> 0.10	0.075

〈표 3〉 단위근·정상성 검정 결과

ADF·KPSS는 모두 원계열을 "정상"으로 판정한다. 그러나 이 두 검정은 단위근(레벨의 추세적 비정상성)만 검정할 뿐 계절적 비정상성은 탐지하지 못한다. 실제 ACF를 보면 lag 1에서 0.96, lag 26에서 -0.82, lag 52에서 0.64로 뚜렷한 계절 자기상관이 남아 있다. 따라서 검정 결과만 믿지 않고 ACF의 계절 구조를 근거로 일반차분 $d=1$, 계절차분 $D=1$, 계절주기 $s=52$ 로 결정한다.

해설.

이 문항의 핵심은 "검정의 한계를 이해하는가"이다. ADF/KPSS의 통과가 곧 계절차분 불필요를 의미하지 않는다. 계절성이 강한 데이터에서는 ACF의 계절 lag(26, 52 등) 패턴을 반드시 함께 확인해야 한다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

ADF·KPSS의 p값은 시차(lag) 선택 기준(AIC/BIC, 고정 시차)에 따라 소폭 달라질 수 있으나 판정은 동일해야 한다.

ACF 값(0.96/-0.82/0.64)은 닫힌 계산이므로 AI 간 차이가 없어야 한다.

차분 차수의 선택은 갈릴 수 있다. ADF가 정상을 지지하므로 $d=0$ 으로 두고 계절차분($D=1$)만 적용하는 답안, 또는 R의 `auto.arima()`나 `fable`이 자동 선택한 다른 차수(예: $d=0$, $D=1$ 에 상수항 포함)를 쓴 답안도 계절 자기상관을 인식하고

D=1을 적용했다면 정답으로 인정한다. 반면 계절차분 없이(D=0) 비계절 ARIMA만 적합한 답안은 계절성 진단에 실패한 것이므로 감점한다.

[문제 2에 대한 답안 샘플]

프롬프트 예시

두 가지 예측 모델을 만들어줘. (1) 로그 변환 후 SARIMA(1,1,1)(0,1,1)[52], (2) Prophet(연간 계절성, 승법 모드, 95% 구간). 각 모형의 설정 의도를 간단히 설명해줘.

AI 결과

SARIMA(1,1,1)(0,1,1)[52] — 로그 변환으로 분산을 안정화한 뒤, 일반차분과 계절차분을 한 번씩 적용한다. 계절 MA(1) 항이 매년 반복되는 충격을 흡수한다.

Prophet — yearly_seasonality=True, seasonality_mode="multiplicative"(수요 수준이 커질수록 계절 변동폭도 커지는 구조 반영), changepoint_prior_scale=0.05(추세 변화에 과적합하지 않도록), interval_width=0.95.

[문제 3에 대한 답안 샘플]

프롬프트 예시

두 모델을 교차검증으로 비교해줘. 컷오프 124·132·140·148주, 예측 구간 8주, 평가지표 MAPE.

AI 결과

컷오프	SARIMA MAPE	Prophet MAPE
124주	5.0%	2.4%
132주	4.8%	4.8%
140주	10.1%	8.2%

148주	7.4%	7.0%
평균	6.8%	5.6%

〈표 4〉 롤링 오리진 교차검증 결과 (예측 구간 8주)

평균에서 Prophet이 약 1.2%p 낮은 MAPE를 보인다. 따라서 Prophet을 최종 예측 모형으로 채택한다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

교차검증 MAPE는 이 과제에서 변동이 큰 수치다. 모형 차수·하이퍼파라미터가 다르면 당연히 달라지고, 같은 설정이라도 R의 forecast/fable과 Python의 statsmodels/Prophet은 추정 알고리즘이 달라 소수점 단위의 차이가 난다. 컷오프를 다르게 잡으면 순위 자체가 뒤집혀 SARIMA가 더 낮은 MAPE를 보이는 결과도 나올 수 있다.

따라서 학생이 SARIMA를 최종 모형으로 채택했더라도, (1) 동일한 컷오프·구간·지표로 두 모형을 공정하게 비교했고, (2) 자신의 비교 결과에 근거해 선택했다면 정답이다. 감점 대상은 모형 선택 자체가 아니라 비교 절차의 오류다: 검증 없이 한 모형만 적합, 표본 내 적합도(R^2 , AIC)만으로 예측력을 비교, 학습 데이터에 겹치는 구간으로 평가하는 경우 등이다.

[문제 4에 대한 답안 샘플]

프롬프트 예시

채택한 Prophet 모형으로 전체 데이터를 학습해 향후 8주(157~164주)를 예측하고, 95% 예측구간과 함께 표로 정리해줘. 비교를 위해 SARIMA 예측치도 함께 보여줘.

AI 결과

주차	날짜	Prophet 예측(개)	95% 구간	(참고) SARIMA
157	2026-04-27	1,274	1,145 ~ 1,405	1,280
158	2026-05-04	1,351	1,229 ~ 1,475	1,441

159	2026-05-11	1,426	1,296 ~ 1,558	1,500
160	2026-05-18	1,504	1,384 ~ 1,633	1,653
161	2026-05-25	1,582	1,450 ~ 1,707	1,673
162	2026-06-01	1,650	1,523 ~ 1,783	1,696
163	2026-06-08	1,709	1,575 ~ 1,844	1,816
164	2026-06-15	1,771	1,643 ~ 1,886	1,944
합계		12,267		

〈표 5〉 향후 8주 수요 예측

해설. 예측치는 여름 성수기 진입에 맞춰 1,274개에서 1,771개로 꾸준히 상승하며, 8주 누적 수요는 약 12,267개다. 예측구간이 시간이 갈수록 넓어지는 것은 먼 미래일수록 불확실성이 커지는 자연스러운 결과다. 두 모형의 예측 방향은 동일하나 SARIMA가 후반부에서 더 가파른 상승을 제시한다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

Prophet의 점예측은 같은 버전·같은 설정이면 재현되지만, 95% 예측구간은 몬테카를로 시뮬레이션으로 산출되므로 실행마다 경계가 달라진다. 실제로 동일 코드를 재실행하면 157주 구간이 1,145~1,405 대신 1,153~1,406으로 나오는 식이다. 구간 경계가 수 개~수십 개 다른 것은 완전히 정상이며, 점예측도 Prophet·statsmodels 버전에 따라 소폭 이동할 수 있다.

8주 누적 수요가 12,267개와 다르면(예: 12,100 또는 12,400) 이후 단계 3의 모든 숫자가 함께 이동한다. 채점 시에는 누적 수요의 절대값이 아니라 "성수기 진입으로 주당 1,300~1,800개 수준까지 상승"이라는 패턴의 포착 여부와, 학생 자신의 예측치와 재고 계산 간의 내적 일관성을 본다.

■ 단계 3. 재고 최적화

[문제 1에 대한 답안 샘플]

프롬프트 예시

서비스수준 95%, 리드타임 4주 조건에서 안전재고와 재발주점을 계산해줘. 수요 표준편차는 채택 모형의 표본 내 예측 오차(RMSE)를 사용하고, 계산 과정을 보여줘.

AI 결과

서비스수준 95% → 안전계수 $z = 1.645$. 주별 수요 표준편차는 Prophet 표본 내 예측오차 RMSE인 $\sigma = 64.5$ 개를 사용하고, 리드타임 $L = 4$ 주를 적용한다.

$$\text{안전재고(SS)} = z \cdot \sigma \cdot \sqrt{L} = 1.645 \times 64.5 \times \sqrt{4} \approx 212\text{개}$$

$$\text{리드타임 수요}(157\sim 160\text{주}) = 1,274 + 1,351 + 1,426 + 1,504 = 5,555\text{개}$$

$$\text{재발주점(ROP)} = \text{리드타임 수요} + \text{SS} = 5,555 + 212 \approx 5,767\text{개}$$

해설.

리드타임이 4주이므로 발주 후 도착까지의 수요(리드타임 수요)는 단순히 주평균의 4배가 아니라 해당 4주의 예측치 합으로 계산해야 한다. 수요가 상승 추세이므로 이 구분이 특히 중요하다. 현재고 1,200개는 ROP(5,767개)를 크게 밑돌아 즉시 발주가 필요한 상태다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

이 단계에서 답안이 가장 크게 갈리는 지점은 σ 의 선택이며, 여기에는 허용되는 차이와 허용되지 않는 차이가 있다. 일부 AI는 안전재고 공식에 원계열의 표준편차(412개)를 그대로 대입한다. 이 경우 $\text{SS} = 1.645 \times 412 \times \sqrt{4} \approx 1,357$ 개로 본 해설(212개)의 6배가 넘는다. 그러나 원계열의 변동 대부분은 계절성과 추세, 즉 예측 모형이 이미 설명하는 예측 가능한 변동이다. 안전재고가 대비해야 할 대상은 예측하고 남은 불확실성이므로, 예측 모형을 사용하는 상황에서는 예측오차의 표준편차(RMSE 64.5)를 쓰는 것이 개념적으로 올바르다.

채점 기준: 원계열 표준편차를 아무 논거 없이 대입해 $\text{SS} \approx 1,357$ 개를 보고한 답안은 안전재고의 개념(예측 불가능한 변동에 대한 완충)을 오해한 것이므로 감점한다. 반면 σ 를 표본 내 RMSE 대신 교차검증 예측오차나 잔차 표준편차로 정의한 답안은 오히려 더 보수적이고 타당한 대안이므로 정답으로 인정한다(이 경우 SS는 212개보다 다소 커진다). 리

드타임 수요를 예측치 합 대신 과거 주평균 $\times 4 \approx 3,864$ 개로 계산한 답안도 상승 추세와 성수기 진입을 무시한 과소추정이므로 감점한다.

z값을 1.64 또는 1.65로 반올림하거나, $\sqrt{L}$ 공식 대신 4주 예측오차의 합산 분산을 직접 계산한 경우의 소수점 차이는 모두 허용한다.

[문제 2에 대한 답안 샘플]

프롬프트 예시

현재고 1,200개에서 시작해 8주간 주별 발주, 입고, 수요, 기말재고·품절량을 시뮬레이션해줘. 발주는 목표재고수준까지 채우는 방식으로 하고, 품절이 발생하는 주차를 표시해줘.

AI 결과

주차	입고	발주	수요	기말재고	품절량
157	0	11,279	1,274	0	74
158	0	0	1,351	0	1,351
159	0	0	1,426	0	1,426
160	0	0	1,504	0	1,504
161	11,279	0	1,582	9,697	0
162	0	0	1,650	8,047	0
163	0	0	1,709	6,338	0
164	0	0	1,771	4,567	0

〈표 6〉 8주 발주·재고 시뮬레이션 (발주량 = 8주 누적 예측 12,267 + SS 212 - 현재고 1,200)

품절 발생 주차: 157~160주, 누적 부족 4,355개.

해설. 현 재고 1,200개로는 157주 수요(1,274개)조차 감당하지 못하고, 발주분은 4주 뒤인 161주에야 도착하므

로 그 사이 157~160주 내내 품절이 이어진다. 이는 발주 시스템의 실패가 아니라 초기 재고가 위험 수준까지 방치된 결과다.

결과의 변동 가능성 (SI에 따라 다르게 나올 수 있는 부분)

시뮬레이션의 구체적 숫자는 발주 정책의 정의에 따라 달라진다. 목표재고수준을 무엇으로 잡는지(8주 누적 + SS, ROP 기준, 주기적 재검토 등), 발주를 매주 허용하는지 1회만 하는지, 품절을 이월 주문(backorder)으로 볼지 판매 상실로 볼지에 따라 발주량과 기말재고가 모두 달라진다. 학생의 단계 2 예측치가 다르면 여기서도 숫자가 연쇄적으로 이동한다.

그러나 어떤 정책을 쓰든 바뀌지 않는 구조적 결론이 있다: 현재고 1,200개 < 첫 주 수요이고 리드타임이 4주이므로, 157주에 즉시 발주하더라도 157~160주의 품절(약 4,300~4,400개 부족)은 피할 수 없다. 채점은 이 구조적 결론에 도달했는지, 그리고 자신의 예측치·정책 정의와 시뮬레이션 숫자가 내적으로 일관되는지를 본다. 정확한 발주량(11,279개)의 일치는 요구하지 않는다.

[문제 3에 대한 답안 샘플]

프롬프트 예시

이 시뮬레이션 결과를 바탕으로 품절로 인한 기회손실과 재고유지비를 추정하고, 품절을 줄이기 위한 실무적 대응 방안을 제시해줘.

AI 결과

비용 추정. 기회손실은 누적 부족 4,355개 × 단위 마진($42,000 - 18,000 = 24,000$ 원) $\approx$ 약 1억 452만 원, 8주 재고유지비는 기말재고 합계 28,649개 × 400원 $\approx$ 약 1,146만 원이다. 품절 비용이 유지비의 9배를 넘어, 이 품목의 비용 구조가 품절 회피 쪽으로 크게 기울어 있음을 보여준다.

대응 방안.

- ① 긴급·분할 입고 — 정규 발주와 별도로 항공특송 등으로 일부 물량을 조기 확보해 157~160주 공백을 메운다.
- ② 서비스수준 상향 — 임계비율 $Cu/(Cu+Co) = 24,000/(24,000+400) \approx 0.984$ 로, 뉴스벤더 관점의 최적 서비스수준은 약 98.4%($z \approx 2.13$)에 이른다. 품절비용(마진 24,000원)이 유지비(주당 400원)를 압도하므로 95%보다 높은 서비스수준이 경제적으로 정당화된다.
- ③ 리드타임 단축 — 공급사 다변화·근거리 소싱으로 L을 줄이면 리드타임 수요와 안전재고가 동시에 감소한다.
- ④ 수요 평활 — 성수기 진입 전 사전 프로모션 분산, 사전예약 등

으로 피크 부하를 낮춘다.

결과의 변동 가능성 (AI에 따라 다르게 나올 수 있는 부분)

기회손실의 크기는 손실 단가의 정의에 따라 달라진다. 공헌마진(24,000원) 기준이면 약 1.05억 원, 판매가(42,000원) 기준이면 약 1.83억 원이다. 상실된 것은 매출이 아니라 이익이므로 마진 기준이 표준이나, 판매가 기준을 쓰고 "매출 손실"로 명확히 표기한 답안은 인정한다. 재고유지비도 기말재고 기준과 평균재고 기준($(\text{기초} + \text{기말})/2$)이 모두 통용되므로 어느 쪽이든 계산 기준을 명시했다면 정답이다.

임계비율에서 도출되는 최적 서비스수준을 97.5%($z=1.96$)나 98%로 제시한 답안도, 0.984라는 임계비율을 올바르게 계산하고 "95%보다 높여야 한다"는 방향에 도달했다면 정답으로 인정한다. 정확한 z 값보다 품질비용과 유지비의 비대칭이 서비스수준 결정을 좌우한다는 원리의 이해를 평가한다.