

ChatGPT 실무적용 SCM/물류 데이터분석 정요표 및 개정내용

8월 1일 이전 출고한 POD 도서의 일부 오타와 POD 도서 출판 후 독자의 소중한 의견을 반영하여 수정한 내용을 아래에 기술합니다. 최신 내용의 반영이 가능한 POD 도서의 장점을 활용하고자 합니다.

P6 중간, 초보자~~로~~ → 초보자도

(수정 후) 초보자도 파이썬 코드를 자연스럽게 이해하도록 설명

P75 하단, 도표 2-59 → 2-57

(수정 후) 도표 2-57. 중복되어 나타나는 물동 PSI

P142 하단, 생산법인의 **공급능력지수**는 → 생산법인의 **계획 공급능력**은

(수정 후) 생산법인의 계획 공급능력은 90이다.

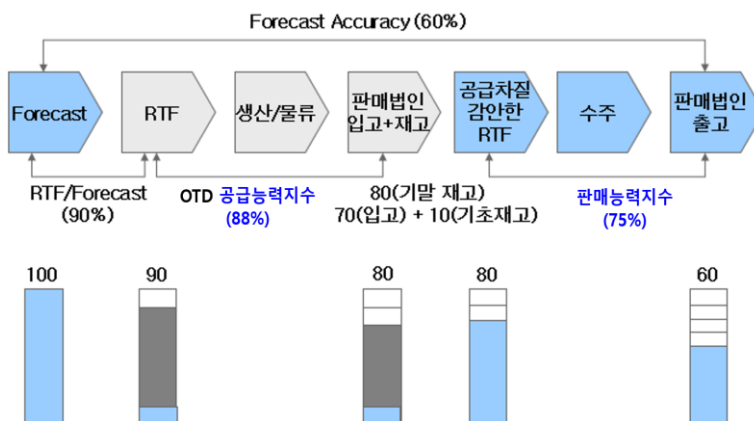
내용 추가, **OTD 공급능력지수**는 88%이다.

(수정 후) 입고되었다. OTD 공급능력지수는 88%이다.

도표 8-1. 공급능력지수(80%) → OTD 공급능력지수(88%)

(수정 후)

도표 8-1. 수요예측 정확도 산출 과정



P147 하단, 48.25% → 48.28%

(수정 후) 48.28%라고 답하였다.

P150 상단, (최솟값 뒤) 맞춤표 . → 따옴표,

(수정 후) 도표 8-11. 모델별 주차별 최솟값, 최댓값 계산 파이썬 코드

P169 하단 내용 삭제, ~~공급리드타임 정보는 SUPPLY.xlsx에 저장되어 있음~~

(수정 후) 적음($0.10 < 0.34$)

P175 상단 내용 추가, 분포의 첨도인 0 보다 → 분포의 **과도** 첨도인 0 보다

(수정 후) 분포의 과도 첨도인 0 보다 낮다.

p188 하단 변수명 **X** 에서 **Z** 로 변경,

(수정 후) $2 \times P(Z \geq |Z\text{-통계량}|) = 2 \times P(Z \geq 28.62) = 2 \times (1 - P(Z \leq 28.62))$,

$P(Z \leq 28.62)$

[방법 1. 수작업/엑셀]

1) 가설수립

$H_0: \mu = 1,500$ vs $H_1: \mu \neq 1,500$ (μ 는 인당 일 평균 입출고량)

2) 유의 수준결정: 5%

3) 검정통계량 계산

모 표준편차는 알려져 있고, $n(91\text{일})$ 이 충분히 크므로(일반적으로 $n > 30$)

검정통계량은 $Z = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}}$ 이고, 근사적으로 정규분포를 따른다.

$Z_0 = \frac{(\bar{X} - \mu_0)}{\sigma/\sqrt{n}} = \frac{(1,650 - 1,500)}{50/\sqrt{91}} = 28.62$, $|Z_0| \geq Z_{0.025} = 1.96$ 이므로

여기서 $Z_{0.025} = \text{NORM.S.INV}(0.975)$ 표준정규분포에서 상위 2.5%(100% - 97.5%)에 해당하는 Z 값을 반환
유의수준 5%하에서 H_0 을 기각한다.

4) p-값 계산

양측검정에서의 $p\text{-value} = 2 \times P(Z \geq |Z\text{-통계량}|) = 2 \times P(Z \geq 28.62) = 2 \times (1 - P(Z \leq 28.62)) = 2 \times (1 - 1) = 0$

여기서, $P(Z \leq 28.62) = \text{NORM.S.DIST}(28.62, 1)$ 표준정규분포에서 Z 값 이하의 누적 확률을 반환

5) p-값과 유의수준을 비교해서 통계적 의사결정

$p\text{-value} = 0 < 0.05$ 이므로 유의수준 5%하에서 H_0 을 기각한다.

P279 하단 내용 삭제, **아런 경우**

(수정 후) 선형모형이 아니다. 다항식을 변수 치환을

P295 중간 내용 삭제, **종속**변수만으로 모형을 구성 → 변수만으로 모형을 구성

(수정 후) 시차 정보가 없는 변수만으로 모형을 구성한다면

P299 중간 내용 추가, 생각하지 → 생각하지만

(수정 후) SCFI가 상승한다고 생각하지만,

p307 중간 [2-3-3 → [2-4-5

(수정 후) [2-4-5 잔차와 독립변수의 상관성 검정]

P311 상단 도표 내용 수정, **경기선행지수** → 물가, 실질 → **미국** 실질

(수정 후) (미국 물가와 미국 실질 GDP 사이의 상관관계수 0.99)

도표 15-28. ChatGPT가 처음 제안한 다중회귀모형의 점검 결과 요약

No	점검 항목		분석 결과
2.1	모형의 적합성		- 회귀식에서 분산분석 F 값의 p-value가 0.05보다 작아서 회귀식 유의함 - 훈련데이터에 대한 조정 결정계수 0.915로 높음
2.2	회귀계수의 크기와 부호가 실제 상황과 일치하는지		- (수요)중국발 컨테이너 물동량, (공급) 명목 선풍량 독립변수에 미 포함 - 용선료, 유가, 미국 경기선행지수가 SCFI와 음의 상관관계
2.3	회귀계수의 유의성과 다중 공선성		- 다수의 변수가 p-value가 0.05보다 커서 유의하지 않음 - 여러 변수가 매우 높은 VIF값을 가지고 있어, 다중 공선성 발생 (미국 물가와 미국 실질 GDP 사이의 상관관계수 0.99)
2.4	잔차분석 (전체데이터)	기술통계 분석	- 잔차의 평균은 0에 가깝지 않고, 표준편차가 52.16으로 변동성이 큼
		잔차의 정규성 검정	- 잔차의 정규분포를 따른 것으로 간주할 수 있음
		잔차와 예측치의 독립성 검정	- 잔차가 독립적이라고 할 수 있음
		잔차의 등분산성 검정	- 잔차가 등분산성을 만족함
		잔차와 독립변수의 상관성 검정	- 대부분의 독립변수들이 잔차와 낮은 상관관계를 가지고 있음
2.5	실제 데이터에 적용하여 검증		- 훈련데이터에 과적합되어있고 테스트데이터를 적절히 설명하지 못함
2.6	시계열 데이터 검증		- 시계열 데이터 분석 필요(17장과 18장)

P314 도표 하단 내용 2 개 수정, 표준편차 → 표준오차

(수정 후 코드) `print("잔차의 표준오차:", standard_error_of_regression)` # 결과출력

(수정 후 그래프)

```
잔차의 표준오차: 105.48522822654937
OLS Regression Results
Dep. Variable: SCFI_C      R-squared: 0.737
Model: OLS                Adj. R-squared: 0.730
Method: Least Squares     F-statistic: 120.3
Date: Fri, 17 May 2024    Prob (F-statistic): 4.88e-14
Time: 07:40:46           Log-Likelihood: -272.47
No. Observations: 45      AIC: 548.9
Df Residuals: 43          BIC: 552.5
Df Model: 1
Covariance Type: nonrobust

```

	coef	std err	t	P> t	[0.025	0.975]
const	-2.504e+04	2358.018	-10.617	0.000	-2.98e+04	-2.03e+04
CLI_USA	258.6228	23.583	10.966	0.000	211.063	306.183

```
Omnibus: 0.501    Durbin-Watson: 0.610
Prob(Omnibus): 0.778    Jarque-Bera (JB): 0.600
Skew: -0.223    Prob(JB): 0.741
Kurtosis: 2.653    Cond. No. 1.50e+04
```

P363 3 째줄 아래 문장 추가,

(수정 후) ~ 요청한다. 이는 테스트데이터를 훈련 단계에서 반영하겠다는 의미이다.

P363 도표 17-27 위 마지막 문장 삭제,

(수정 후) 훈련데이터만으로 모형을 개발하였으나 좋은 평가결과를 보여준다.

P366 마지막 문장 삭제 후, 다음 문장으로 대체

(수정 후) 데이터의 특성과 분석의 목적에 따라 달라질 수 있다.

도표 17-31 과 같이 훈련데이터에 적절하지 않은 나쁜 결과를 초래할 수 있다.

P367 마지막 문장에 다음에 문장 추가,

(수정 후) ~ 적합하지 않다. ChatGPT 요청 시, “테스트데이터” 대신에 “훈련데이터”에 대한 RMSE 가 가장 작은 값을 갖는 평활계수를 추천함으로 문장을 수정하거나 문장을 삭제하고 실험하여 결과를 비교할 필요가 있다.

P375 하단 내용 추가, 모형에서 사용하는

(수정 후)

✓ 자기회귀(AR): 'p'라고 표시되며, 모형에서 사용하는 관측치의 개수이다.

P377. 7 째줄 아래 문장 추가,

(수정 후) ~ 요청한다. 이는 테스트데이터를 훈련 단계에서 반영하겠다는 의미이다.

p395 도표 18-14 에서 X 삭제, S 추가. ARIMAX → SARIMA

(수정 후) 도표 18-14. SARIMA(1, 1, 1)(1, 1, 1, 13) 결과 요약(코랩)

아래 페이지 모두 ARIMAX, SARIMA → SARIMAX 로 수정합니다.

1) p398 의 8 번째 줄

이번에 ARIMAX 모형을 적용한 → (수정 후) 이번에 SARIMAX 모형을 적용한

2) p399 첫 줄 도표 18-18 과 마지막 줄 도표 18-19

(수정 후) 도표 18-18. SARIMAX(1,1,1)(1,1,1,2)의

(수정 후) 도표 18-19... 튜닝 후, SARIMAX 의 예측결과(코랩)

3) p400 의 6 번째 줄

개발한 모형인 ARIMAX → (수정 후) 개발한 모형인 SARIMAX

P404. 마지막 문장에 다음에 문장 추가,

(수정 후) ~ 지정할 수도 있다. ARIMA, SARIMA, SARIMAX 등 모형 개발을 위해 ChatGPT 요청 시, “테스트데이터” 대신에 “훈련데이터”에 대한 RMSE 가 가장 작은 값을 갖는 평활계수를 추천함으로 문장을 수정하거나 문장을 삭제하고 실험하여 결과를 비교할 필요가 있다.