

시계열분석및실습

HW_FINAL

정병철 교수님

통계학과 2022580033 김나경

통계학과 2023580036 주희찬

환경공학부 2022890064 전아현

2025-12-23

인천국제공항 월별 입국자 수에 대한 시계열 분석: 계절형 ARIMA 모형을 중심으로

김나경 · 주희찬 · 전아현
서울시립대학교 통계학과 · 환경공학부

Abstract

본 연구는 인천국제공항의 월별 입국자 수 데이터를 이용하여 입국자 수요의 시계열적 특성을 분석하고 변동성을 설명할 수 있는 최적의 시계열 모형을 구축하는 것을 목적으로 한다. 입국자 수 데이터는 장기적인 증가 추세와 뚜렷한 계절성을 보임에 따라, 분산 안정화를 위한 로그 변환과 계절 차분 및 비계절 차분을 수행하였다. Box-Jenkins 방법론에 따라 모형을 식별한 결과, $ARIMA(0,1,3)X(0,1,1)_{12}$ 모형이 가장 우수한 적합도를 보였으며, 잔차 분석 결과 해당 모형의 잔차는 백색잡음 과정을 따르는 것으로 확인되었다. 본 연구의 결과는 공항 운영 계획 및 정책 수립을 위한 기초 자료로 활용될 수 있다.

I 서론

1.1 연구 목적 및 배경

1990년대 중반 이후 우리나라의 해외여행 자유화가 본격화되고, 경제 성장과 항공 운송 인프라의 확대에 의해 인천국제공항의 입국자 수는 장기적으로 증가하는 추세를 보여왔다. 특히 휴가철, 명절, 연말·연초 등 특정 시기에 집중되는 국제 이동 패턴은 입국자 수의 뚜렷한 계절적 변동성을 형성하는 핵심 요인으로 작용해 왔다. 이러한 계절적 요인들은 공항의 인력 배치, 시설 운영 계획, 항공 편성 등 다양한 운영 및 정책 의사결정에서 중요한 고려 요소가 된다.

입국자 수 시계열 자료는 추세와 계절성이 동시에 존재하는 복합적 구조를 가지고 있어, 이를 단순한 비계절형 모형으로 설명하기에는 한계가 있다. 따라서 계절차분과 비계절차분을 활용하여 데이터의 불안정성을 제거하고, 계절성과 비계절성을 모두 반영할 수 있는 계절형 ARIMA 모형을 적용하는 것이 적절하다.

이에 1993년부터 2009년까지의 인천국제공항 월별 입국자 수 자료를 활용하여 시계열적 특성을 분석하고 여러 계절형 ARIMA 모형의 적합성과 예측력을 비교하여 최적 모형을 도출하는 것을 목적으로, 과거 인천공항 수요의 계절적 변동 패턴을 정량적으로 파악하여 시계열 분석 기반의 수요 예측 모형 구축에 기초 자료를 제공하고자 한다.

II 데이터 소개 및 수집 과정

2.1 데이터 소개 및 수집 과정

본 연구에서 사용한 데이터는 1993년부터 2009년까지의 인천국제공항 월별 입국자 수 시계열 자료이다. 해당 기간은 우리나라의 해외여행 자유화가 본격화된 시기를 포함하며, 입국자 수의 장기적인 변동 추이를 파악하기에 적합한 기간이다.

	A	B
1	date	x
2	1993.01	162117
3	1993.02	161719
4	1993.03	194579
5	1993.04	184128
6	1993.05	189356
7	1993.06	205476
8	1993.07	201984
9	1993.08	266352
10	1993.09	227534
11	1993.1	311032
12	1993.11	238421
13	1993.12	225796
14	1994.01	178274
15	1994.02	206077

Fig. 1. 데이터 원본 (상위 15개)

III 사용 모형 및 방법론

3.1 데이터 전처리 및 변환

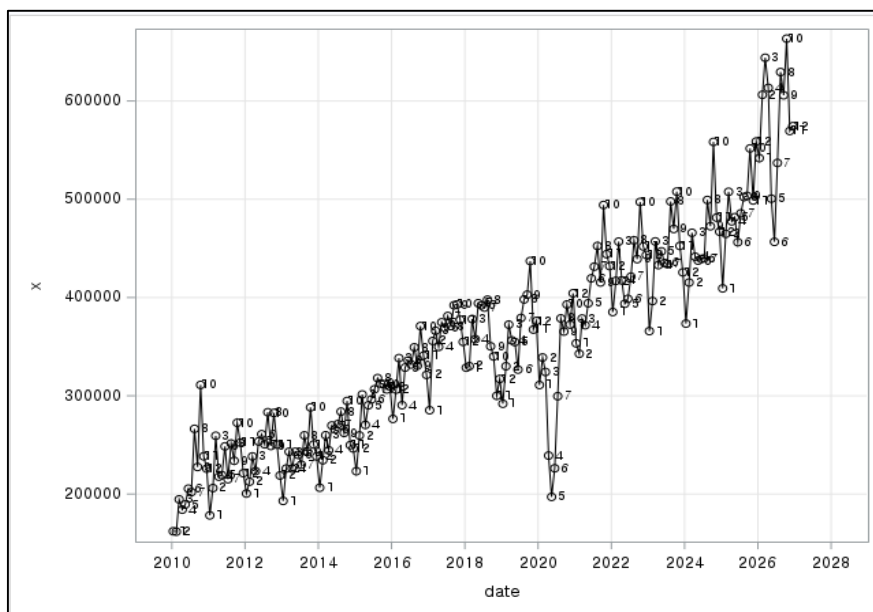


Fig. 2. 원계열(X_t) 시도표

원계열 x_t 에 대한 시도표 분석 결과, 시간이 지날수록 분산이 커지는 이분산성과 강한 계절성 변동이 확인되었다. 시계열의 정상성을 확보하기 위해 다음과 같은 변환을 수행하였다.

1. 분산 안정화: 자연로그 변환을 통해 $\log(x_t)$ 생성하여 이분산성 완화
2. 추세 제거: 1차 차분(∇)을 수행하여 시계열의 평균을 0 중심으로 안정화
3. 계절성 제거: 1차 차분 후 남아있는 12개월 주기의 계절성 제거를 위해 계절 차분(∇_{12}) 진행

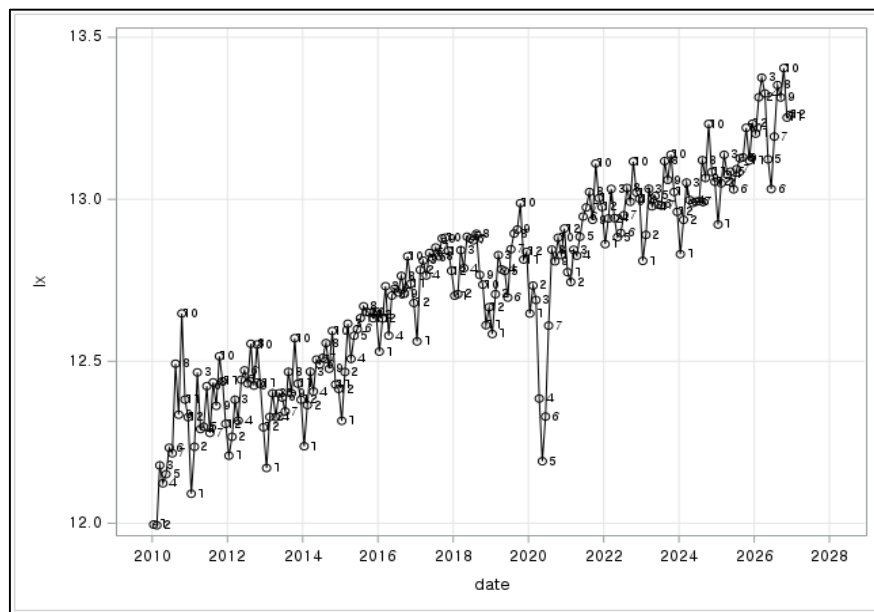


Fig. 3. 로그변환 $\log(x_t)$ 에 대한 시도표

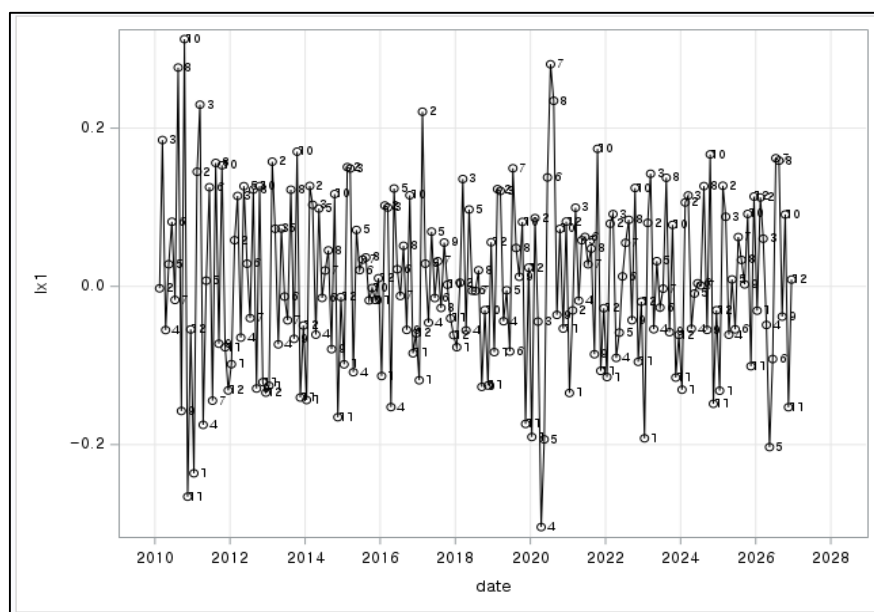


Fig. 4. 로그변환 계열에 추세차분을 실시한 $\nabla \log(x_t)$ 에 대한 시도표

- 1차 차분 결과 시계열의 평균이 0을 중심으로 안정화되었다.
- 매년 주기적으로 상승과 하락을 반복하는 강한 계절성이 보임 → 계절차분 필요

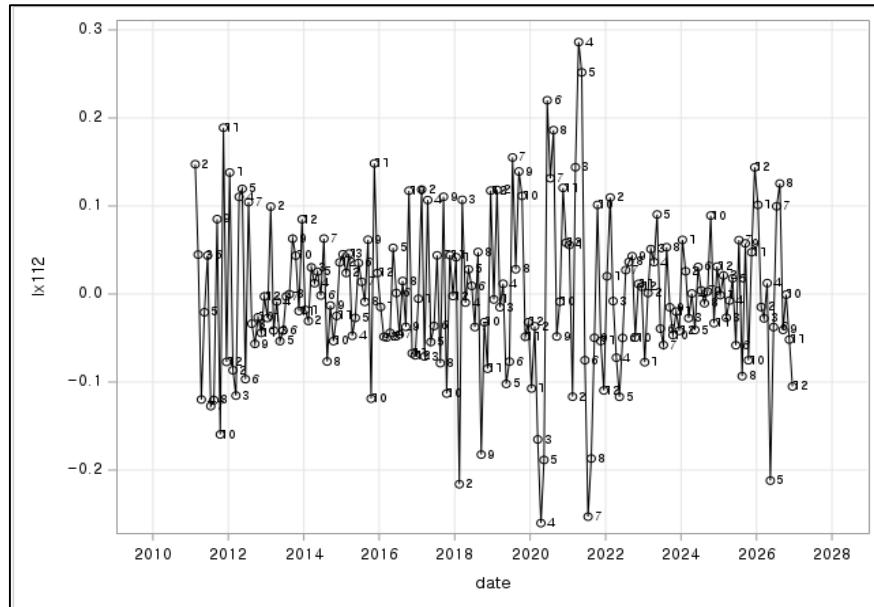


Fig. 5. $\nabla \log(X_t)$ 에 대한 시도표

특별한 추세가 존재하지 않고 전반적으로 분산이 안정화된 모습이다.

3.2 $\nabla \log(X_t)$ 에 대한 식별과정

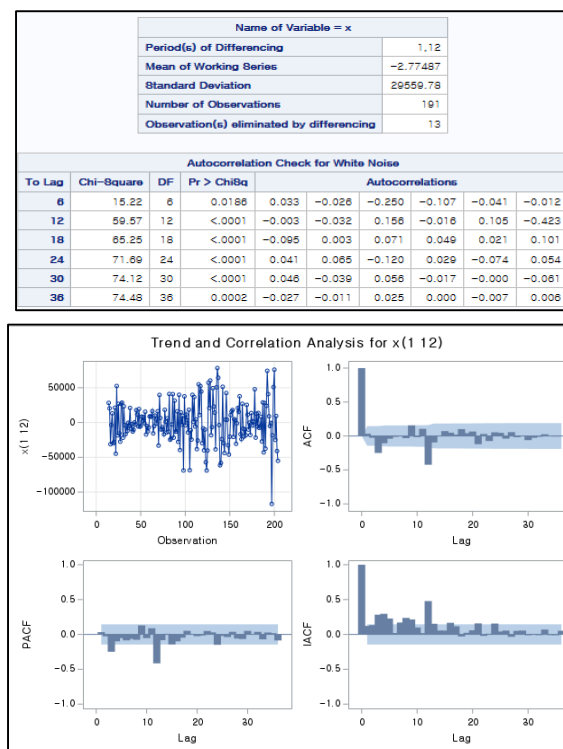


Fig. 6. $\nabla \log(X_t)$ 에 대한 식별과정

$\nabla \nabla_{12} \log(X_t)$ 에 대한 포토만토검정 결과 모든 K에서 무상관이라는 귀무가설을 기각한다.

1) 계절 PART 식별 : ACF - 시차 12에서 절단 / PACF - 시차 12에서 절단

→ 후보모형 : AR(1), MA(1)

2) 비계절 PART 식별 : ACF - 시차 3에서 절단 / PACF - 시차 1에서 절단

→ 후보모형 : AR(1), MA(3)

→ 후보모형 : 평균수준이 0인 ARIMA (0,1,3)×(1,1,0)₁₂, ARIMA (0,1,3)×(0,1,1)₁₂,
ARIMA (1,1,0)×(1,1,0)₁₂, ARIMA (1,1,0)×(0,1,1)₁₂

IV 모형 진단 결과

4.1 각 모형에 대한 CLS 추정결과

ARIMA (0,1,3)×(1,1,0)₁₂ 모형에 대한 추정결과

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0,06523	0,07045	0,93	0,3557	1
MA1,2	0,12717	0,07016	1,81	0,0715	2
MA1,3	0,29224	0,07047	4,15	<,0001	3
AR1,1	-0,49851	0,07134	-6,99	<,0001	12

Variance Estimate

6,4985E8

Std Error Estimate

25492,14

AIC

4421,812

SBC

4434,821

Number of Residuals

191

ARIMA (0,1,3)×(0,1,1)₁₂ 모형에 대한 추정결과

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0,12183	0,06906	1,76	0,0794	1
MA1,2	0,11086	0,06939	1,60	0,1118	2
MA1,3	0,33837	0,06991	4,84	<,0001	3
MA2,1	0,65610	0,06300	10,41	<,0001	12

Variance Estimate

6,0209E8

Std Error Estimate

24537,56

AIC

4407,233

SBC

4420,242

Number of Residuals

191

ARIMA (1,1,0)×(1,1,0)₁₂ 모형에 대한 추정결과

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
AR1,1	0,05176	0,07340	0,71	0,4816	1
AR2,1	-0,51495	0,07133	-7,22	<,0001	12

Variance Estimate

6,8883E8

Std Error Estimate

26245,55

AIC

4430,97

SBC

4437,475

Number of Residuals

191

ARIMA (1,1,0)×(0,1,1)₁₂ 모형에 대한 추정결과

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0,62235	0,06504	9,57	<,0001	12
AR1,1	0,01263	0,07276	0,17	0,8624	1

Variance Estimate

6,5324E8

Std Error Estimate

25558,59

AIC

4420,838

SBC

4427,343

Number of Residuals

191

Table. 1. 각 모형에 대한 CLS 추정결과

	<u>ARIMA</u> <u>(0,1,3)×(1,1,0)₁₂</u>	<u>ARIMA</u> <u>(0,1,3)×(0,1,1)₁₂</u>	<u>ARIMA</u> <u>(1,1,0)×(1,1,0)₁₂</u>	<u>ARIMA</u> <u>(1,1,0)×(0,1,1)₁₂</u>
AIC	4421.812	4407.233	4430.97	4420.838
SBC	4434.821	4420.242	4437.475	4427.343

Table. 2. 최적 모형 선정

후보모형들 중 AIC와 BIC가 가장 작고 통계적으로 유의한 ARIMA (0,1,3)×(0,1,1)₁₂모형을 최종 모형으로 선택한다.

4.2 ARIMA (0,1,3)×(0,1,1)₁₂ 모형진단

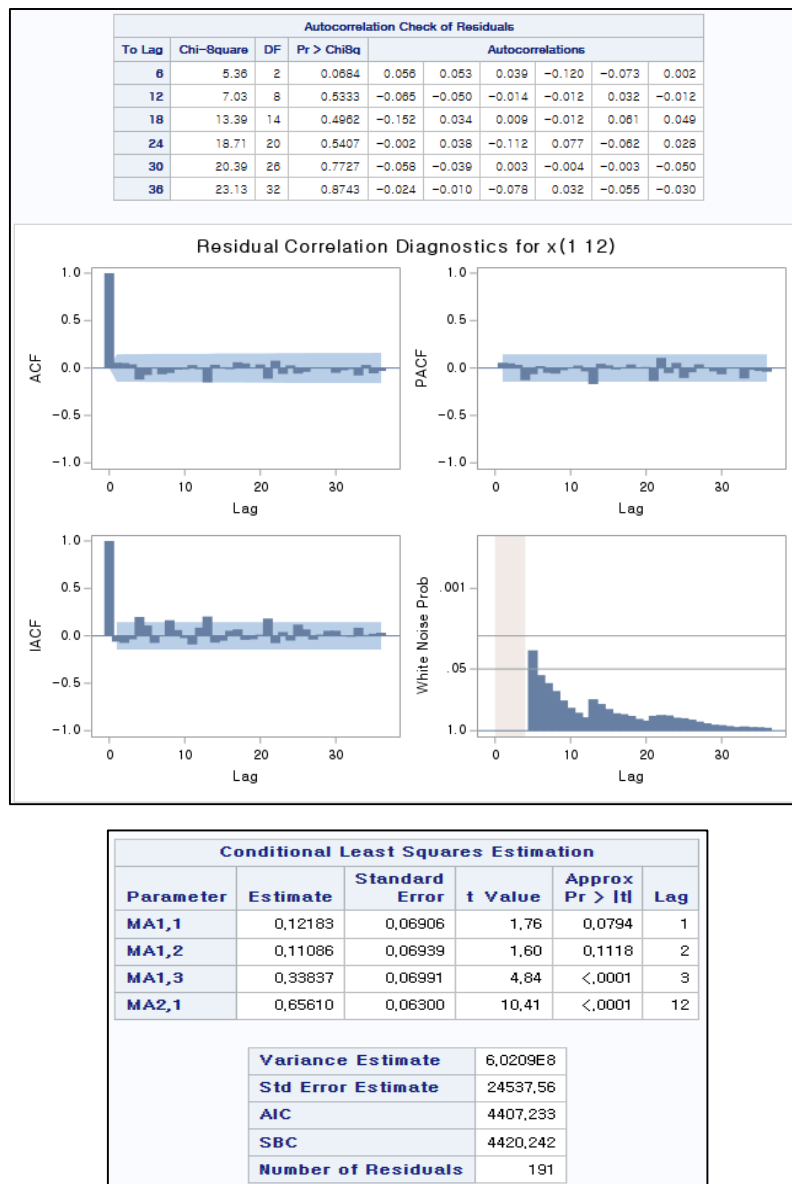


Fig. 7. RIMA (0,1,3)×(0,1,1)₁₂ 모형진단

- 잔차에 대한 포토만토 검정결과 모든 차수에서 자기상관이 존재하지 않는다.
- 잔차의 SACF와 SPACF도 유의한 차수의 자기상관이 존재하지 않는다.
- 따라서 현재 고려한 ARIMA (0,1,3)×(0,1,1)₁₂ 모형의 적합은 적절하다.

4.3 ARIMA (0,1,3)×(0,1,1)₁₂ 모형의 추정결과

$W_t = (1 - B)(1 - B^{12}) \log X_t$ 라 하면,

$$W_t = (1 - 0.12183B - 0.11086B^2 - 0.33837B^3)(1 - 0.65610B^{12}) \epsilon_t$$

$$W_t = \epsilon_t - 0.12183 \epsilon_{t-1} - 0.11086 \epsilon_{t-2} - 0.33837 \epsilon_{t-3} - 0.65610 \epsilon_{t-12} + (\text{Interaction Terms}) \dots$$

- W_t : 1차 및 12차 계절 차분이 적용된 로그 변환 시계열
- ϵ_t : 현 시점의 백색소음(White Noise) 오차
- $-0.12183\epsilon_{t-1}$: 현 시점의 값이 바로 전 시점(t-1)의 오차에 의해 영향을 받는 비계절성 MA(1) 효과
- $-0.11086\epsilon_{t-2}$: 현 시점의 값이 2시점 전(t-2)의 오차에 의해 영향을 받는 비계절성 효과 MA(2) 효과
- $-0.33837 \epsilon_{t-3}$: 현 시점의 값이 3시점 전(t-3)의 오차에 의해 영향을 받는 비계절성 MA(3) 효과
- $-0.65610 \epsilon_{t-12}$: 현 시점의 값이 1년 전 동일(t-12)의 오차에 의해 영향을 받는 계절성 MA(1) 효과

V 결론

인천국제공항의 월별 입국자 수 시계열 자료를 바탕으로 계절형 ARIMA 모형을 이용하여 입국자 수요의 변동 패턴을 실증적으로 분석하였으며 주요 분석 결과는 다음과 같다.

1) 입국자 수 시계열은 뚜렷한 상승 추세와 계절적 변동성을 보였으며, 이를 안정화하기 위해 로그 변환 및 계절 차분을 수행하였다. 다양한 후보 모형을 비교·분석한 결과, AIC와 SBC 기준에서 가장 우수한 성능을 보인 ARIMA(0,1,3)×(0,1,1)₁₂ 모형을 최적 모형으로 선정하였다

2) 선정된 모형에 대한 잔차 분석 결과, 포토만토 검정과 잔차의 ACF 및 PACF 분석에서 유의한 자기상관이 발견되지 않아 잔차가 백색잡음과정을 따르는 것으로 확인되었으며, 이를 통해 모형의 통계적 적합성을 검증하였다.

3) 본 연구에서 도출된 최적 시계열 모형은 향후 인천국제공항의 입국자 수요를 예측하는 데 있어 신뢰할 수 있는 기초 자료로 활용될 수 있다. 특히 계절적 변동성을 고려한 수요 예측은 공항의 효율적인 인력 배치, 시설 운영 계획 수립 등 실무적 의사결정에 유용한 시사점을 제공할 것으로 기대된다.