



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학 석사학위 논문

시계열 자료를 이용한
응급의료센터 수요 예측 모델링

아주대학교 대학원

의학과/의학전공

성진욱

시계열 자료를 이용한
응급의료센터 수요 예측 모델링

지도교수 박래웅

이 논문을 의학 석사학위 논문으로 제출함.

2010년 2월

아주대학교 대학원

의학과/의학전공

성진욱

성진옥의 이학 석사학위 논문을 인준함.

심사위원장 박 래 응 인

심사위원 노 재 성 인

심사위원 조 준 필 인

아주대학교 대학원

2009년 12월 21일

시계열 자료를 이용한 응급의료센터 수요 예측 모델링

본 연구는 응급의료센터 수요예측의 주요기준이 되는 ‘응급의료센터 일일 내원 환자 수’를 예측하는 시계열 모델을 개발하는 것이 목적이다.

‘응급의료센터 일일 내원 환자수’를 미리 예측함으로써 수요에 맞게 응급의료센터의 한정된 자원을 효율적으로 배분하고 시기 적절하게 활용할 수 있을 뿐 아니라 응급의료센터에 내원하는 환자들의 만족도나 치료 결과, 예후에도 좋은 영향력을 미칠 수 있을 것이다.

본 연구에 사용된 데이터는 병원정보시스템 데이터베이스로부터 응급의료센터 내원 환자 정보를 수집하였으며, 예측모델 개발을 위해서 2007~2008 년의 2 년간 데이터를, 모델평가를 위해서 3 개월간의 데이터를 가공하여 사용하였다.

예측모델구축에 사용한 변수는 날씨변수(월, 요일, 계절, 분기, 휴일, 추석 여부)와 날씨변수(평균기온, 최저기온, 최고기온, 일교차, 강수여부, 신적설여부, 풍속도, 상대습도, 황사여부)이다.

예측모델 구축을 위해 사용한 2 년간의 데이터에서는 응급의료센터에 내원한 전체 환자수가 169,375 명, 일일 평균 내원 환자수가 232 명이었으며, 동일 기간 동안의 응급의료센터 일일 내원 환자의 추이는 7 일 주기성과 계절적 추세를 보였다. 특히 ‘추석’변수에 갑작스런 상승추이를 보였으며, 상대적으로 ‘설’의 경우에는 약한 상승추이를 보여 본 연구의 선택변수에서는 제외하였다.

평균과 분산은 안정되어 차분이 필요하지 않았으나, 계절적 추세로 인해 1 차의 계절차분으로 자료의 정상성을 유지하였다.

응급의료센터의 일일 내원환자 수를 예측하는 모델로 본 연구에서는 시계열 모델 중 1) MA(Moving Average) 모델, 2) 단변량 - 계절형 ARIMA(Seasonal Auto-Regressive Integrated Moving Average) 모델, 3) 다변량 - 계절형 ARIMA(Seasonal Auto-Regressive Integrated Moving Average) 모델을 구축하였다.

그리고, 각 모델의 적합도 평가를 위해 1) 잔차분석, 2) AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion) 값을 비교·평가하였고, MAPE(Mean Absolute Percentage Error)를 통해 각 모델의 예측정확도를 평가하였다.

구축한 세 종류의 예측모델을 비교한 결과, 다변량 Seasonal ARIMA 모델이 응급의료센터 일일 내원 환자 수 예측에 가장 적합함을 알 수 있었고(AIC : 6703.7, BIC : 6749.5), 예측정확도는 MAPE(Mean absolute Percentage Error)가 7.4%로 오차가 가장 적어 본 연구의 목적에 부합하는 최종예측모델로 선정하였다. 그리고, 설명 변수는 추석여부, 계절, 평균기온, 강수여부를 채택하였다.

그러므로, 본 연구에서는 다변량 계절형 ARIMA 모델이 비교되었던 MA 모델과 단변량 계절형 ARIMA 모델에 비해 설명변수를 반영하고 있어 예측모델 설명력이 높고, 응급의료센터 일일 내원 환자 수에 대한 예측이 정확함을 확인할 수 있었다.

핵심어 : 수요예측모델, 응급의료센터 과밀화 현상, 시계열 분석, Seasonal Auto-Regressive Integrated Moving Average(ARIMA),

차 례

국문요약	i
차례	ii
그림차례	iii
표차례	iv
I. 서론	1
A. 연구배경	1
B. 연구목적	8
II. 연구대상 및 방법	9
A. 연구대상	9
B. 연구방법	12
1. 제안하는 응급의료센터 수요예측 모델	12
2. 시계열 예측모델 적용	17
3. SARIMA 을 활용한 수요예측 모델 구축	31
III. 결과	36
IV. 고찰	52
V. 결론	58
참고문헌	59
ABSTRACT	63

그림 차례

Fig. 1. Overview of Research Method	13
Fig. 2. Time Plots of Daily Emergency Department Patients(2007.01~2009.03)	39
Fig. 3. Transforms of Daily Emergency Department Patients using seasonal difference	40
Fig. 4. ACF Correlogram of Daily Emergency Department patients	41
Fig. 5. PACF Correlogram of Daily Emergency Department patients	42
Fig. 6. MA(2) Model Residual Analysis	44
Fig. 7. SARIMA(1,0,1)(0,1,1) ₇ Model Residual Analysis	45
Fig. 8. SARIMA(1,0,2)(0,1,1) ₇ Model Residual Analysis	46
Fig. 9. Observed and Predicted daily ED patients; MA (Moving Average) Model	50
Fig. 10. Observed and Predicted daily ED patients ; Uni-variate SARIMA (Seasonal Auto-regressive Integrated Moving Average) Model.....	50

Fig. 11. Observed and Predicted daily ED patients ; Multi-variate SARIMA

(Seasonal Auto-regressive Integrated Moving Average) Model 51



표 차례

Table 1. Previous Studies	5
Table 2. Definition of Variables	12
Table 3. Commentary of MAPE value	35
Table 4. The Results of Comparison between Training Date Set and Validation • Data Set by χ^2 test	37
Table 5. Residual Analysis of MA(2) Model, Univariate SARIMA Model, Multivariate SARIMA Model	43
Table 6. Multivariate Seasonal ARIMA Model Parameters	48
Table 7. Goodness of fits for Models(AIC, BIC)	49
Table 8. MAPE values of constructed Models	49

I. 서론

A. 연구배경

현대 의학이 발전하면서 인구의 고령화, 급박한 환경변화로 인해 발생하는 스트레스 압박, 자살·폭행 및 약물중독으로 인한 사회범죄의 증가, 차량 증가로 인한 교통사고 증가, 산업의 발달로 야기될 수 있는 대량재해 등으로 응급의료서비스에 대한 요구와 관심이 늘어나고 있다(정대유 등, 2006).

그로 인해 응급의료센터 내원 환자 수의 증가, 내원하는 환자들의 질병 형태의 다양화, 중증도 치료에 대한 요구의 증가, 응급의료센터의 인력부족(의사, 간호사, 의료기사, 당직인력, 행정 및 사무직원 등)과 입원병실 부족, 진료공간의 부족 등으로 인해 응급의료센터는 점점 더 과밀화(Overcrowding)해지고 있다(Tandberg 와 Qualls, 1994; 서동우 등, 2004; 제상모 등, 2005; 유인술 등, 2007).

응급의료센터 과밀화 현상은 응급의료센터의 수요와 공급의 변화를 초래하게 하였고, 응급의료센터 서비스에 대한 환자의 만족도 뿐만 아니라 환자의 치료 결과, 예후에도 직접적인 악영향을 미치고 있다(이은숙과 박금선, 1994; Sun 등, 2000; Schull 등, 2004; 최효선과 이경원, 2006).

응급의료센터 과밀화 현상을 해소하기 위한 방안으로는 국내의 응급의료에 대한 제도적 발전과 제공 체계의 확장, 응급의료센터의 인력 확충, 응급의료센터 병상 및 공간 확대, 응급의료센터의 검사 장비 확충, 특수검사를 위한 진료인력 보강, 단순 질환 치료를 위한 Walk in clinic 운영, Hallway 운영, 관찰병상운영 등이 추진(Derlet 와 Richards, 2000; Arnold 등, 2006)되고 있으나, 이러한

해소방안은 응급의료센터의 시설, 장비, 인력이 갖추어 지기 위해서 시간과 비용이 많이 필요하다.

그러므로, 응급의료센터의 과밀화를 효율적으로 해소하기 위해서는 ‘수요 예측’ 방법을 이용하여 개별 응급의료센터의 수요를 미리 예측하고 수요에 맞게 자원을 활용하는 것이 필요하다(Schweiger 등, 2007; Spenser 등, 2009).

즉, 개별 응급의료센터의 수요를 미리 예측함으로써 시기 적절하게 응급의료센터의 한정된 자원들을 효율적으로 활용하는 것이 과밀화 현상을 해소할 수 있는 효율적인 방안이다(정구영 등, 1997; Arnold 등, 1998; 유인술 등, 2007).

‘수요예측’이란 계획을 세우는 것을 전제로 하며, 불확실한 미래에 발생할 수 있는 다양한 것 중에서 가장 발생할 가능성이 높은 것을 미리 예측하기 위한 의사결정방법이다. 급박하게 변화하는 환경 속에서 수요를 미리 예측함으로써 의사결정과 관련된 불확실성과 위험을 줄이고, 경영자 또는 관리자들이 합리적인 의사결정을 할 수 있도록 하는 중요한 역할을 하는 것이다(이종원, 2006; 황성완 등, 2008; Lisa 등, 2009).

각 개별 응급의료센터의 수요예측 모델의 개발 및 적용은 한정된 응급의료센터의 자원을 적절하게 배분할 수 있도록 할 뿐만 아니라, 환자의 만족도나 치료 결과, 예후 등에도 좋은 영향을 미치게 된다(Asplin 등, 2006).

수요예측의 방법은 정치, 경제 등 다양한 분야에서 흔히 활용되고 있는 방법이며, 의료보건분야에서는 사용한 사례가 많지는 않으나 응급의료센터 수요예측에 관한 연구들이 (Table 1.)과 같이 몇몇 논문에서 수행되었다.

Spencer S. Jones 등은 3 개의 각각 다른 응급의료센터를 대상으로 SARIMA 모델(Seasonal Auto-Regressive Integrated Moving Average), Time Series Regression,

지수평활법, 인공신경망 모델, 선형회귀분석방법으로 응급의료센터 일일 내원 환자수를 예측하기 위한 모델을 구축하고, 구축된 다양한 예측모델을 비교하기 위한 연구를 수행하였다. 이 연구에서는 대부분의 예측방법이 단변량 예측 모델이었으며, 다변량 예측모델은 Time Series regression 으로 날씨변수(요일변수, 월, 휴일변수(휴일, 휴일 다음날, 평일) 등)와 날씨변수(일일 최고기온, 최저기온, 강우량(인치), 적설량(인치))를 포함한 예측모델과 날씨변수를 제외한 예측모델로 구축하여 비교하였다.

그 결과, 다중 선형 회귀분석법은 시계열 자료의 특성을 반영하고 있지 못하므로, 시계열자료인 응급의료센터 일일 내원환자수를 예측하기 위한 방법으로 Time Series Regression 이 가장 좋은 모델로 선정되었다. 또한, Time Series Regression 모델 간 비교에 있어서도 날씨변수를 제외한 Time Series regression model 이 응급의료센터 일일 환자수에 대해 더 적절하게 예측되었다고 연구를 통해 결론을 내렸다(Table 1.).

그러나, 이 연구는 3 개의 병원이 거의 동일한 지역에 위치해 있다는 점, 각각 병원의 특정요인들을 고려하지 않은 예측모델을 구축하였다는 제한점이 있었다(Spencer 등, 2008).

Yan Sun 등은 응급의료센터에 내원하는 일일 환자수에 영향을 미치는 지역적인 요인을 확인하고, 그 요인을 기반으로 한 예측 모델을 개발하는 것이었다. 또한, 자원은 환자 질환의 격심한 정도에 따라 달라질 수 있음을 가정하여 환자 질환의 격심한 정도를 3 가지로 나누어 ARIMA 를 이용하여 응급의료센터에 내원하는 일일 환자수를 예측한 연구이다(Table 1.).

이 연구에서 응급의료센터에 내원하는 일일 환자수에 영향을 미치는 지역적인 요인으로는 각각의 예측모델마다 조금씩 다르긴 했지만, PSI(Pollution

Standard Index)와 공휴일을 선택하였으며, 모델 구축시 독립변수 중 날씨변수로는 평균기온, 상대습도를 고려하였으나, 예측모델의 설명변수로는 채택되지 않았다 (Table 1).

이 연구의 제한점은 응급의료센터에 내원하는 일일 환자수에 영향을 미칠 수 있는 다른 개별병원들의 지역요인에 대해 고려하지 못했다는 점과 날씨변수로 평균기온과 상대습도만을 독립변수로 고려하였다는 점이였다(Yan 등, 2009).

국내에서는 이러한 응급의료센터 수요예측 연구가 많지 않았으며, 그 중 4 곳의 대학병원을 대상으로 날씨 및 날짜 변수가 응급실로 내원한 환자수에 미치는 영향에 대해 수행되었던 이장영 등의 연구가 있었다(Table 1).

이 연구에서는 월별 병원당 내원 환자수의 변화는 1 월이 가장 적었고, 그 이후로 증가하다 9 월에 정점에 오르고 다시 줄어드는 양상을 보였고, 요일별로는 일요일 92±13, 토요일 75±11, 월요일 71±10, 나머지 평일은 64~65 명의 분포를 보였는데, 토, 일, 월요일은 각각 다른 요일에 비해 유의한 차이가 있었다. 계절별로는 여름이 하루 평균 74±12 명으로 가장 많았고, 겨울이 65±13 명으로 가장 적었다.

그리고, 날씨 변수들은 일일 최고, 최저, 평균기온과 내원 환자수와의 상관 분석 결과 약한 상관관계를 보였으며, 비가 내린 날과 그렇지 않은 날의 내원 환자의 수에는 차이가 없었고, 눈이 쌓인 날과 그렇지 않은 날의 내원 환자의 수에도 차이가 없었다. 이 연구에서는 응급의료센터 일일 내원 환자에 영향을 주는 요인으로 날짜변수 중 월, 계절, 요일, 휴일여부가 연관이 있었고, 날씨변수 중에서는 기온 이외에는 의미 있는 변수가 없었음을 확인하였다(이장영 등, 2005).

Table 1. Previous Studies.

Author	Spencer at al, 2008	Yan at al,2009	Lee at al, 2005
Source	3 Emergency Dept.	1 Emergency Dept.	4 Emergency Dept.
Dependent Variables	Daily ED Patients.	1) Daily P1 group 2) Daily P2 group 3) Daily P3 group 4) Daily ED Patients	Daily ED Patients.
Independent Variables	* Calendar Variables - Weeks - Months - Holidays * Weather Variables - MAX. air temperature - MIN. air temperature - Daily precipitation - Daily snowfall	* Calendar Variables -Day of the Week -Month of the year, -Public Holiday * Weather Variables - Ambient temperature - Relative humidity * PSI : Pollution Standards Index	* Calendar Variables - Seasons - Months - Weeks - Holidays * Weather Variables - MAX. air temperature - MIN. air temperature - Average temperatures, -The Rainfall(>10mm), -The Snowfall)
	Forecasting Method	- Uni-variate SARIMA Method - Time Series Regression - Exponential smoothing - Artificial neural Network - Simple linear Regression model (Benchmark model).	- Multi-variate non-Seasonal ARIMA Model - Multi-variate Seasonal ARIMA Model
Final Forecasting Model	* Time Series regression - Without climatic Variables	*Multi-variate ARIMA Model	*Multiple linear Regression model
Result	<u>MAPE</u> Facility 1: 14.4% Facility 2: 9.41 % Facility 3: 9.13%	<u>MAPE</u> P1 : 18.2% P2: 7.7% P3: 7.2% All: 4.4%	<u>Calibrated R²: 62.6%</u>

또한, 이 연구에서는 선형회귀분석을 통해 미래의 응급의료센터에 내원하는 환자를 예측하는 회귀식을 3 가지의 변수만을 이용하여 만들었으나, Batel 등의 연구에서는 13 개의 변수를 이용한 복잡한 회귀식에 비해 예측력이 낮았다. 그리고, 우리나라의 특성상 설과 추석명절기간에 환자들이 폭주하는 현상으로 인해 회귀모델의 예측력을 떨어뜨리는 결과를 초래하기도 하였다(Batel 등, 2001; 이장영 등, 2005).

그리고, 이장영 등의 연구에서는 요일별 내원환자수가 일요일이 가장 많았고, 화요일부터 금요일까지는 낮은 분포를 보이다가 토요일부터 올라가는 양상을 보였으나, Batel 등의 국외 연구에서는 월요일이 가장 높았고, 일요일로 갈수록 낮아지는 양상을 보였다. 이러한 차이의 원인으로서는 한 주에 7 일 모두 하루에 15.5 시간을 진료하는 외래형 응급실(Walk-in-Clinic)과 같은 우리나라와 다른 의료체계를 가지고 있었기 때문이었다(Batel 등, 2001).

이처럼 각 나라의 의료체계에 따라 응급의료센터의 요일별 내원환자의 추이가 달라질 수 있음을 이 연구를 통해 확인할 수 있었으며(김민정 등, 2005; 최효선과 이경원, 2006), 이 연구의 제한점으로는 선형회귀분석법을 이용한 응급의료센터 내원 환자수 예측이므로 연구기간을 연장하거나 병원 수를 늘리는 것에 따라서 통계 결과 및 회귀식이 달라질 수 있다는 점과 단일병원 연구의 오류를 줄이기 위해 4 개의 대학병원을 대상으로 하였지만, 특정시의 모든 응급환자를 대상으로 하지 못하였으므로 예측에 한계를 가진다는 점 등이었다(이장영 등, 2005).

위 선행연구의 고찰을 통해 발견한 관련연구들의 몇 가지 제한점을 요약하면, 다음과 같다.

- 1) 국외와 달리 국내연구에서는 응급의료센터 과밀화에 대한 연구는 많았으나 그에 비해 응급의료센터 ‘수요예측’에 관한 연구는 부족하였다는 점
- 2) 국내의 수요예측관련 연구 중 선형회귀 분석을 이용한 연구 외에 시계열 자료의 특징을 고려하여 시계열 모델을 이용한 연구가 없었다는 점
- 3) 응급의료센터의 수요예측 모델에 영향을 미치는 개별병원의 다양한 요인을 고려하지 못했다는 점이다.

그러므로, 본 연구에서는 이러한 제한점을 보완하기 위해 국내 단일기관 권역응급의료센터를 대상으로 시계열 모델 중 시계열 구성요소가 시간의 흐름에 따라 매우 빠르게 변동하는 경우에 효과적이며 어떤 형태를 가진 시계열 자료라도 분석이 가능한 장점을 가진 SARIMA 모델을 이용한 응급의료센터 수요예측의 기본이 되는 ‘응급의료센터 일일 내원 환자 수’를 예측하는 모델 개발이 필요하며, 계절성을 반영한 ARIMA 모델을 이용한 응급의료센터의 일일 내원 환자 수 예측 모델에 대해 단변량이 아닌 다변량 SARIMA 모델에 영향을 미치는 다양한 요인들을 고려할 필요가 있다.

B. 연구목적

본 연구는 국내 단일기관 응급의료센터를 대상으로 응급환자정보와 지역의 날씨정보, 휴일 등을 포함한 날씨변수를 이용하여 응급의료센터 과밀화 현상의 주요한 기준이 되는 ‘응급의료센터 일일 내원 환자 수’를 예측하는 시계열 모델을 개발하고, 구축된 예측모델을 비교·평가하는 것을 목적으로 수행하였다

본 연구의 목표는

- 시계열 자료의 특성을 고려하여 시계열 모델 중 MA(Moving Average)모델, 단변량 SARIMA 모델, 다변량 SARIMA 모델 구축,
- 구축된 시계열 모델의 예측정확도 상호 비교,
- 국내 단일기관 응급의료센터 내원 환자 수 예측에 가장 적합한 최종 예측 모델 선정이다.

위의 목표를 달성함으로써 향후 기대되는 효과는 다음과 같다.

첫째, 국외연구에서는 미처 반영하지 못한 국내의 지역적 특성이 반영된 응급의료센터 수요예측을 위한 시계열 모델을 개발할 수 있을 것이다.

둘째, 응급의료센터의 수요를 미리 예측함으로써 그 수요에 맞는 응급의료 서비스 및 응급의료자원을 효과적으로 공급하고 자원을 배분할 수 있을 것이다.

셋째, 응급의료센터의 한정된 자원을 효율적으로 활용함으로써 인해 기관의 시간-비용-노력 등의 경제적 효과를 기대할 수 있게 될 것이다.

II. 연구대상 및 방법

A. 연구대상

권역응급의료센터는 전국을 16개의 권역으로 나누고, 그 지역에서 가장 중추적 역할을 할 수 있는 병원 1개소씩을 응급의료의 중심병원으로 지정한 것으로 미흡한 응급의료체계를 개선하여 각종 재해나 사고로 생긴 응급환자보다 전문적이고 체계적으로 관리하기 위한 제도이며, 중증응급환자 치료의 중심일 뿐만 아니라, 응급환자 치료를 위한 정보의 제공, 응급의료인의 교육, 재해 중심병원의 역할을 수행하도록 한다.

1. 연구대상 선정의 타당성

본 연구의 대상이 된 응급의료센터는 권역응급의료센터로 지정되어 운영되고 있으며, 응급의료센터에 내원하는 연간 평균환자수는 84,688명(2007년 80,496명, 2008년 88,879명), 평균 일일 내원환자수는 232명으로 전국 권역응급의료센터 중에서 연간 내원 환자 수에 대해서는 상위권의 규모를 가지고 있다.

응급의료센터의 전담인력은 응급의학과 전문의(교수 및 강사), 외상 전문 외과 전문의(교수), 소아과 전문의(강사), 각 임상과의 교수 및 강사(당직체제)를 비롯한 응급의학과 전공의 13명, 인턴 8명, 몇 개 임상과의 파견 전공의, 2인 이상의 모든 임상진료과 전공의(당직 협진 체제)로 구성되어 있으며, 간호사는 응급간호팀장 1명, 파트장(수간호사) 3명 외 수십 명의 간호사들과 7명의 방사선기사, 6명의 응급구조사, 7명의 원무팀 직원,

2명의 행정요원, 9명의 청원경찰, 3명의 안내요원 등으로 이루어져 있다.

응급의료센터의 시설 및 장비는 별도의 권역응급의료센터 건물 내에 응급의료센터 중환자실, 응급병동 등의 시설을 갖추고 있으며, 응급의료센터 별도의 진단장비, 환자감시장치, 치료장비, 기타장비(혈액냉장고, 온장고(수액보관), 제빙기, 정수기 등)를 구비하고 있다.

2. 데이터 수집 및 가공

본 연구는 병원정보시스템(Hospital Information System(HIS))에서 추출한 데이터 집합을 이용한 후향적 연구로, 연구에 사용된 데이터는 병원정보시스템 데이터베이스(DataBase=DB)로부터 2007년 1월부터 2009년 3월까지 응급의료센터에 내원한 환자들의 정보를 병원 정보관리팀으로부터 개인식별정보(이름, 병록번호)를 제외하여 익명화된 데이터로 189,511건을 수집하였다.

2007년 1월 ~ 2008년 12월까지의 훈련 데이터 집합(Training Data Set)은 응급의료센터 수요예측 모델 개발(구축)에 사용하였고, 2009년 1월 ~ 2009년 3월까지의 평가 데이터 집합(Validation Data Set)은 예측 모델의 평가를 위하여 사용하였다.

연구에 사용된 데이터는 응급의료센터에 내원하여 전산 등록한 환자정보를 일별로 된 시계열 자료로 가공하였으며, 통상 동일한 환자가 응급의료센터를 여러 번 방문한 경우가 있을 수 있으나, 본 연구의 목적이 일일 내원 환자 수를 예측하는 것이므로 중복된 경우도 환자수의 합계에 포함시켜 자료를 가공하였다.

3. 데이터의 특징

본 연구에 사용된 변수들은 일일 내원 환자수, 월별, 요일별, 분기별, 휴일별, 추석여부, 계절별, 지역의 날씨 (평균기온, 최저기온, 최고기온, 일교차, 강수여부, 신적설여부, 풍속도, 상대습도, 황사여부) 등의 변수를 사용하였으며, 변수에 대한 설명은 (Table 2.)과 같다.

일일 내원 환자수는 내원 일시를 기준으로 오전 0시부터 다음날 오전 0시까지 응급의료센터에 내원한 환자수의 합계로 하였으며, 휴일별의 휴일은 공휴일과 일요일, 휴일 다음날은 휴일 다음날에 위치하는 정상근무일과 토요일을 함께 분류하였다

일일 내원 환자수의 시계열 도표에서 눈에 띄게 갑작스럽게 상승한 특이점 (outlier)은 확인한 결과 추석이였으며, 이는 특이점(outlier)이 아니라 의미 있는 값으로 추석여부를 변수로 처리하여 특이점을 보정하지 않았다.

날씨변수는 기상청 웹사이트로부터 수집하였으며, 강수여부는 기상청의 권고대로 일반인이 비가 충분히 내렸다고 생각하는 10mm이상을 기준으로 나누었고, 신적설은 눈이 쌓인 높이를 말하는 것으로 그 유무에 따라 변수를 나누었다.

Table 2. Definition of Variables

Variables	Explanation
Month	January, February, March, April, May, June, July, August, September, October, November, December
Day of the Week	Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
Quarters of the year	1Q, 2Q, 3Q, 4Q
Holiday	Weekdays, After Holiday, Holiday,
Chusuk	Chusuk (Y/N)
Seasons	Spring , Summer, Fall, Winter
Average-Temperature	Average-Temperature
Minimal-Temperature	Minimal-Temperature
Maximal-Temperature	Maximal-Temperature
Temperature Gap	Maximal-Minimal Temperature
RAIN	RAIN (Y(≥ 10 mm),N)
SNOW	SNOW (Y/N)
Air-Velocity	The speed of the wind
Relative Humidity	Relative Humidity
Yellow dust	Sandy dust Phenomena

B. 연구방법

1. 제안하는 응급의료센터 수요예측 모델

(A) 응급의료센터 수요예측모델

본 연구에서 제안하는 응급의료센터 수요예측모델에 관한 연구방법의 개괄은 (Fig. 1.)와 같다.

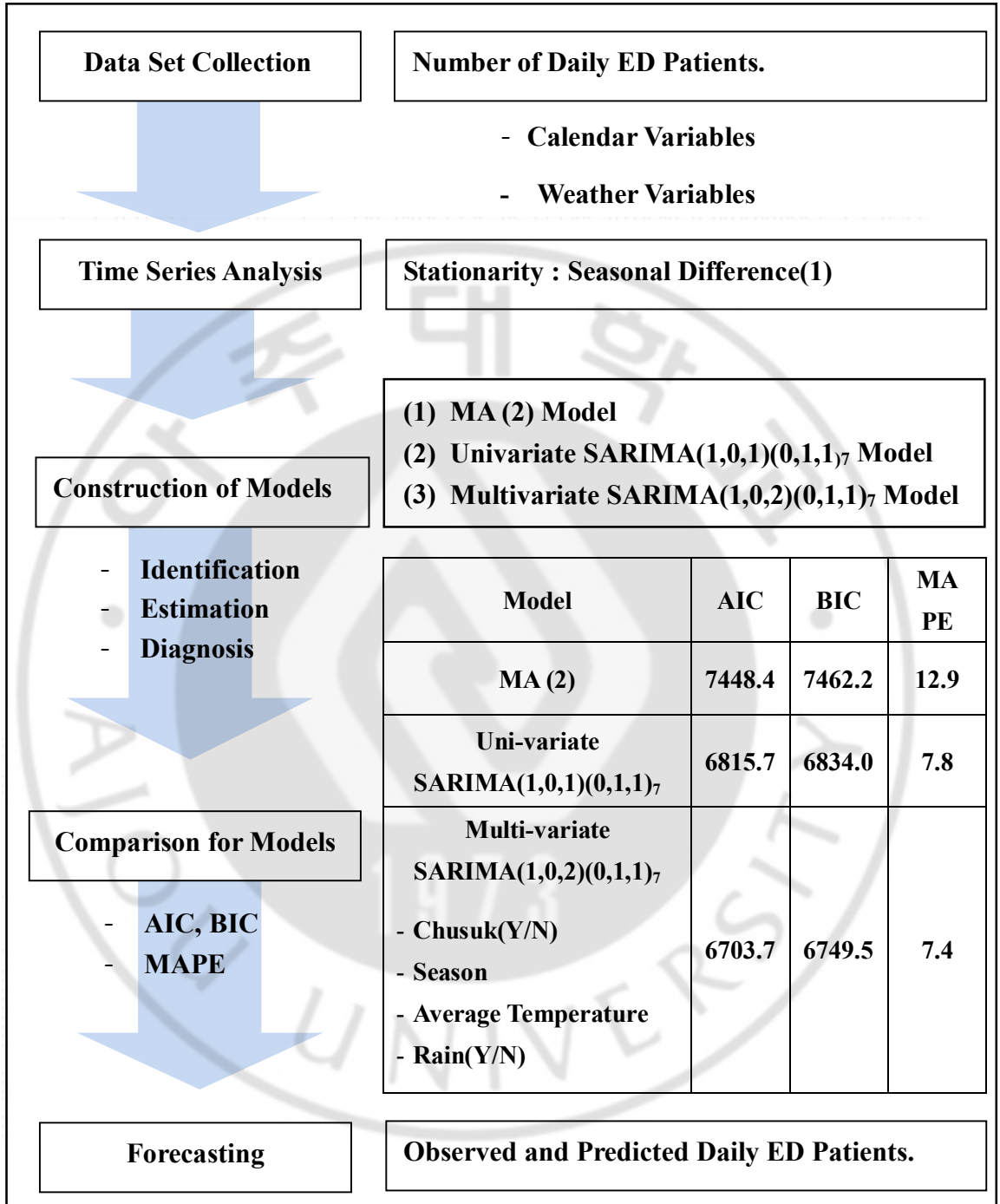


Fig. 1. Overview of Research Method. ED=Emergency Department, MA=Moving Average, SARIMA=Seasonal Auto-Regressive Integrated Moving Average, ACF=Auto-Correlation Function, PACF=Partial Auto-Correlation Function, AIC=Akaike Information Criterion, BIC=Bayesian Information Criterion, MAPE=Mean Absolute Percentage Error.

응급의료센터 과밀화 현상을 해소하기 위한 효과적인 방법으로 본 연구에서는 “수요예측”을 이용하고자 하며, “수요예측”이란 수요분석을 통하여 시장조사나 각종 예측조사의 결과를 종합해서 미래의 수요를 예측하는 것이다.

수요를 예측하는 방법으로는 예측대상이나 목적에 따라 여러 가지로 분류할 수 있지만, 일반적으로 정성적 예측방법 (Qualitative Technique)과 정량적 예측방법(Quantitative Technique)으로 구분할 수 있다.

정성적 예측방법은 주로 경영자, 종업원 등 조직 내외의 인간이 가진 경험이나 이들의 견해에 의존하여 미래의 결과를 주관적으로 예측하는 방법으로 신제품 및 서비스 도입 시와 같이 과거 자료와 시간이 불충분할 때 흔히 사용되는 방법이고, 정량적 예측방법은 관측된 과거자료에 대한 통계적 분석을 통하여 미래의 수요를 예측하는 방법으로 과거에 대한 자료가 충분하고, 그 자료가 수치로 나타낼 수 있을 때 사용되는 방법이다.

또한, 정량적 예측기법은 다시 인과형 예측기법(Causal Forecasting Method)과 시계열 예측 기법(Time Series Analysis)으로 분류될 수 있으며, 인과형 예측기법은 수요에 영향을 주는 환경 요인들을 파악하고 수요와 이 요인들간의 인과관계를 파악함으로써 수요를 예측하는 방법이고, 시계열 예측기법은 과거의 수요를 분석하여 시간에 따른 수요의 패턴을 파악하고 과거의 수요패턴이 미래에도 지속된다는 기본적인 가정하에 미래의 수요를 예측하는 방법이다.

이러한 수요예측기법에는 공통적 특징이 있는데, 효과적으로 수요를 예측하는데 있어 공통적인 특징을 이해하는 것이 매우 중요하다

수요예측기법의 공통적 특징은 다음과 같다.

- 1) “수요예측기법들은 일반적으로 과거에 존재했던 인과관계가 미래에도 계속될 것이라는 가정하에 이루어진다.”

즉, 시계열 예측 기법의 경우 과거의 패턴이 미래에도 그대로 지속된다는 것을 전제로 하며, 실제로는 과거의 패턴이 계속되는 것이 아니므로 안정성이 기본적으로 가정되어야 한다.

- 2) “예측은 언제나 정확할 수 없으며, 오차가 발생된다.”

이러한 예측 값과 실제 값과의 차이를 최소화시키는 것이 수요예측의 목적이다.

- 3) “개별적으로 예측하기 보다는 집단화, 그룹화하여 예측하는 것이 오차를 줄일 수 있다.”

집단화, 그룹화해서 예측하게 되면, 개체수가 증가한 만큼 신뢰성 있는 계획을 할 수 있게 되므로 예측오차(Forecasting error)를 줄일 수 있다.

- 4) “예측의 정확성은 예측기간이 증가할수록 감소하게 된다.”
일반적으로 단기예측이 장기예측보다 정확성이 높다.

수요예측기법은 이러한 공통적인 특징을 바탕으로 관측된 과거의 시계열로부터 시간이 흘러도 변하지 않고, 계속되어 온 현상이 미래에도 계속 되리라는 가정하에 예측모델을 개발하고, 그 모델을 통해 미래를 예측하는 것이다. 모든 수요예측기법들에는 오차가 발생할 수 밖에 없으므로, 예측모델을 구축하는 과정에서 오차를 최소화하는 것이 수요예측의 목적이 된다.

본 연구에서는 수요예측의 방법 중 “시계열 예측기법”을 이용하여 사회·문화·경제·기술적 환경에 따라 급격히 변화하는 응급의료서비스에 대한 요구를 파악하고, 미래의 응급의료서비스에 대한 미래의 수요를 예측하고자 한다.

“응급의료센터의 수요예측”은 응급의료센터에 내원하는 환자수를 기준으로 전반적인 응급의료센터의 수요를 예측할 수 있다는 가정하에 단일기관의 응급의료센터를 내원하는 일일 환자 수를 예측하는 모델을 구축하고, 예측모델을 비교·평가하는 연구를 수행하고자 한다.

2. 시계열 예측모델 적용

(A) 시계열 모델의 이론적 고찰

(1) 시계열 자료의 특성과 구성

시계열(Time Series)은 어떠한 현상에 대하여 시간의 흐름에 따라 일정한 간격으로 이들을 관측하여 기록한 자료를 말하며, 일반적으로 자료는 횡단면자료(Cross-Sectional Data)와 종단면 자료(Longitudinal Data)로 구분하는데, 시계열 자료는 종단면 자료에 해당된다.

시계열 자료 즉, 종단면 자료는 횡단면 자료에 비해 관측값들의 독립성이 가정되지 않고 관측값들 사이의 상호연관성이 정보획득을 위한 중요한 원리가 된다. 즉, 시계열 자료는 어떠한 현상에 관한 시간의 변화를 나타내는 것이므로 어느 한 시점에서 관측된 시계열 자료는 그 이전까지의 자료들에 영향을 받게 된다(정동빈과 원태연, 2001).

시계열은 시계열의 특성에 따라 하나의 통일된 시간단위로 관찰하여 기록되어야 하며, 시계열의 특성을 파악하기 위해서는 시계열 도표를 통해 시계열의 구성요인을 알아보아야 한다.

시계열 자료의 주요 구성요소는 다음과 같다.

- ① 추세변동(T ; Trend Movement) : 장기간에 걸쳐 수요가 일정하게 증가 또는 감소하는 추세의 형태를 말하며, 기술 및 환경의 변화로 인해 시계열 자료에 영향을 주는 장기 변동 요인이다.
- ② 순환변동(C ; Cyclical Fluctuation) : 1 년 이상의 주기를 가지고 순환하는 시계열의 구성요소로 추세를 중심으로 반복된

상승과 하강의 형태를 보이는 변동이다. 수요의 추세가 경기 순환과 같이 장기적인 경제변동에 의하여 결정되고, 순환의 기간이 변동적이다.

③ 계절변동(S ; Seasonal Variation) : 주로 1년을 주기로 발생하는 시계열의 변동요인으로 추세나 순환변동에 비해 상대적으로 단기변동에 해당하며, 월 효과나 분기효과 등이 대표적이다.

④ 불규칙변동(I ; Irregular Movement) : 추세, 계절, 순환변동으로 설명되지 않는 변동으로 지진, 전쟁, 홍수, 파업과 같은 사전에 예상할 수 없는 사건에 의한 변동과 파악되지 않는 요인에 의해서 발생하는 우연적 변동이 있다. 단기간에만 일어나고, 측정이나 예측이 어려운 오차변동을 말한다.

(2) 시계열 모델과 확률과정(Stochastic Process)

과거의 사실이나 현상은 미래에 영향을 미치며, 년, 월, 주, 일 등의 시간간격에 따라 관측된 과거의 자료를 토대로 추세나 경향을 분석하여 일정한 법칙을 발견하고 미래에도 그러한 법칙이 지속될 것이라는 가정하에 모델을 구축하고, 예측모델을 통해 미래의 수요를 예측하는 것이 시계열 모델(Time Series Model)의 원리이다.

① 시계열과 확률과정

시계열은 시간의 흐름에 따라 변하는 현상을 일정한 시간 간격으로 관측하여 얻어진 일련의 자료들이며, 각 시점 t 에서의 시계열을 Y_t 로 '확률변수'라고 하고, 관측된 실제값은 확률변수 Y_t 의 모집단으로부터 각 시점에서 얻어진 표본으로 이를 실현값

(realization)이라고 한다(이덕기, 1999).

불규칙하게 변동하는 시계열 모델의 특징을 파악하기 위해서는 확률의 개념을 도입하여야 하며, 특히 시점마다 얻어진 값을 확률변수의 실현값으로 간주해서, 시계열을 만드는 확률적인 구조를 모델화하는 것 즉, 확률변수들의 모임인 $\{Y_n, t=1, 2, \dots, n\}$ 를 확률과정(Stochastic Process)이라 하며, 특히 t 가 시간을 나타낼 때의 확률과정을 시계열 모델이라고 한다.

따라서 이러한 시계열 모델 $\{Y_t\}$ 의 확률적 특성은 Y_t 들의 결합분포함수에 의해서 결정되고, 시계열의 확률적 특성을 구하기 위해서는 평균이나 분산 또는 자기상관함수를 통해 시계열의 분포를 알 수 있다. 특히 시계열이 정규분포로 안정되어 있다면, 평균과 분산만으로도 분포의 특성을 정확하게 알 수 있다.

② 평균 및 자기상관함수

시계열모델의 확률적 특성을 알기 위해서는 시계열의 평균 및 분산, 자기상관함수(Auto-Correlation Function)을 구하여야 한다.

평균(means)은

$$\mu_t = E(Y_t), \quad t=0,1,2,\dots,n \text{ 이고,}$$

자기공분산함수(Autocovariance Function)는

$$\begin{aligned} \gamma_{t,s} &= \text{Cov}(Y_t, Y_s) \\ &= E(Y_t - \mu_t)(Y_s - \mu_s), \quad t,s = 0,1,2,\dots,n, \end{aligned}$$

자기상관함수(Auto-Correlation Function)는

$$\begin{aligned} \rho_{t,s} &= \text{Corr}(Y_t, Y_s) \\ &= \gamma_{t,s} / (\gamma_{t,t} \gamma_{s,s})^{1/2}, \quad t,s = 0,1,2,\dots,n \text{ 이다.} \end{aligned}$$

③ 정상성(Stationarity)

표본으로부터 모집단을 추론하는 것처럼 시계열 모델에서도 주어진 시계열 Y_t 를 어떤 확률과정에 의해 실현값을 생성하는 과정을 추론하게 된다. 이를 추론하기 위해서는 ‘정상성’의 가정이 필요하다.

‘정상성(Stationarity)’이란 시계열의 확률적 특징이 시간에 따라 변하지 않는다는 것이다.

즉, 정상확률과정은 시계열의 평균과 분산이 시점 t 에 관계없이 일정하고, 정상적인 확률과정일 경우에는 다음과 같은 특징이 있다.

(1) 모든 t 에 대하여 평균이 일정하다.

$$E(Y_t) = \mu.$$

(2) 모든 t 에 대하여 분산이 일정하며, 상수이다.

$$\text{Var}(Y_t) = \gamma(0) < \infty.$$

(3) 모든 t 에 대하여 두 시점 사이의 자기 공분산은 시차에만 의존한다.

$$\text{Cov}(Y_t, Y_s) = \gamma(|t - s|).$$

④ 자기상관함수(ACF)와 편자기상관함수(PACF)

자기 공분산값은 변수를 측정하는 스케일에 따라 달라지기 때문에 대부분 자기상관계수 ρ_k 를 계산하게 된다.

‘자기상관함수(Auto-Correlation Function)’는 동일한 변수에 대해 서로 다른 시점에 관찰된 값들 사이에 존재하는 상호작용관계를 나타내는 것으로 동일한 변수에서 얻어진 연속적인 관측값들 사이의

상호연관관계를 나타내는 척도이다(이덕기, 1999).

$$\rho_{\kappa} = \text{Corr}(Y_t, Y_{t-\kappa}) = \frac{\text{Cov}(Y_t, Y_{t-\kappa})}{\text{Var}(Y_t)} = \frac{\sum_{t=1}^{T-\kappa} (Y_t - \bar{Y})(Y_{t+\kappa} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2} = \frac{\gamma_{\kappa}}{\gamma_0}$$

‘편자기상관함수(Partial Auto-Correlation Function)’는 k 이외의 모든 시차를 갖는 관측치($Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$)들로부터의 영향력을 배제한 가운데 특정의 두 관측치 Y_t 와 Y_{t-k} 가 얼마나 관련이 있는지를 나타내는 척도이다(이덕기, 1999).

즉, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ 의 영향을 제거시킨 후 Y_t 와 Y_{t-k} 의 순수한 상관계수를 의미한다.

정상시계열 $\{Y_t\}$ 에서 시차가 κ 인 편자기상관함수는 회귀식

$$Y_t = \phi_{\kappa 1} Y_{t-1} + \phi_{\kappa 2} Y_{t-2} + \dots + \phi_{\kappa \kappa} Y_{t-\kappa} + a_t \text{에서}$$

시차가 κ 번째 회귀계수를 편자기상관계수 $\phi_{\kappa \kappa}$ 로 나타낸다.

$$\phi_{\kappa \kappa} = \text{Corr}[Y_t, Y_{t-\kappa} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-\kappa+1}]$$

(3) 시계열 예측방법

다양한 수요예측기법 중 본 연구에 사용한 예측방법은 시계열 예측방법이며, 시계열 예측방법에는 이동평균(Moving Average)법, 지수평활법(Exponential Smoothing Model), ARIMA(Auto-Regressive Integrated Moving Average)모델이 있다.

① 이동평균법(Moving Average)

이동평균은 시계열 자료의 일정기간 즉, 기준 시점과 인접한 n 개의 관측값들의 평균을 의미하며, 이동평균법이란 이동평균을 이용하여 불규칙변동이나 계절변동을 제거하는 것으로 평활법 중 한 가지이다.

단순이동평균법은 “ 시계열 모델 중 가장 단순한 예측방법으로 과거 시계열 자료(연도별, 월별, 분기별자료)를 연속적으로 산술 평균하여 수요를 예측하는 것”을 말하며, 이것을 식으로 표현하면 다음과 같다(우주리, 2007).

$$F_{n+1} = \frac{1}{m}(X_n + X_{n-1} + X_{n-m+1})$$

(F_{n+1} = 현재 시점이 n 일 때 다음 시점 n+1 의 예측값,

X_n : 시점 n 의 실제 값, m= 이동평균기간)

이동평균법은 불규칙변동이나 계절변동을 간단히 제거할 수 있는 장점이 있으나, 긴 기간의 이동평균이나 반복되는 이동평균은 많은 의미 있는 자료를 잃게 할 수 있다.

② 지수평활법(Exponential Smoothing Model)

과거의 관측값들에 대해 동일한 가중치를 부여하는 이동 평균법과는 달리 최근의 관측치가 미래예측에 더 많은 영향을 줄 것이라는 가정하에 과거보다 현재의 관측값에 많은 가중치를 주어

수요의 변화를 평활시키는 방법이다. 즉, 미래 시계열을 예측하기 위하여 과거의 시계열 값들을 단순 또는 가중 평균하여 시계열에 포함된 확률오차를 평준화시키는 것이다.

③ ARIMA(AutoRegressive Integrated Moving Average)모델

ARIMA 모델은 이동평균법, 지수평활법과 달리 확률적인 접근을 통해 예측하는 것으로, 시계열의 추세를 제거하여 정상성을 확보한 후 시계열을 확률에 의해서 모델을 구축하고, 예측하는 방법이다(이주호, 2006).

ARIMA의 일반적인 표현은 다음과 같다.

$$ARIMA(p, d, q)(P, D, Q)_s$$

ARIMA 모델은 자기회귀(AR)부분, 차분 부분, 이동평균부분으로 크게 3 가지 부분으로 구성되어 있으며, 여기서 (p, d, q) 부분은 비계절 부분을, (P, D, Q) 부분은 계절 부분을, s 는 계절의 기간을 나타낸다.

비계절적 부분은 비계절적 자기회귀(AR) 차수 = p , 비계절적 차분 (Difference) 차수 = d , 그리고 비계절적 이동평균(MA) 차수 = q 로 표현되며, 계절성을 가지는 시계열 자료의 경우에는 계절적 자기회귀(Seasonal AR) 차수 = P , 계절적 차분(Seasonal Difference) 차수 = D , 그리고 계절적 이동평균(Seasonal MA) 차수 Q 로 표현된다.

ARIMA 과정이란, 시계열의 패턴을 정확하게 반영해 줄 수 있는 차수 p,d,q 와 P,D,Q 를 적절하게 찾아내는 것이며, 이러한 개념식을 수학적으로 표현하면 다음과 같다.

$$\phi(B)\Phi(B^s)(1-B^s)^D(1-B)^d \hat{y}_t = \theta(B)\Theta(B^s)e_t, \hat{y}_t = y_t - \mu$$

t: 시차, y_t : 종속변수 또는 차분변수, μ : 종속변수의 평균

d=비계절적 차분횟수, D: 계절적 차분횟수

B= 후향연산자, $BX_t = X_{t-1}$,

$\phi(B)$: 비계절적 AR 모델

$\Phi(B^s)$: 계절적 AR 모델

$\theta(B)$: 비계절적 MA 모델

$\Theta(B^s)$: 계절적 MA 모델

e_t : 오차항 (백색잡음, white noise)

(B) 시계열 모델의 적용

시계열 모델을 구축하는 과정은

- (1) 시계열 모델 예비 단계 (정상화),
- (2) 모델 식별 단계,
- (3) 모수 추정 단계,
- (4) 모델 진단 단계
- (5) 예측에 의해 이루어진다.

(1) 시계열 모델 예비 단계,

시계열 모델 구축을 위한 예비단계로 시계열 도표를 통해서 시계열이 시간이 흐름에 따라 증가하거나 감소하는 추세가 있는지, 분산이 커지거나 작아지는지, 일정한 현상이 규칙적으로 나타나는 계절성이 있는지, 특정시점에서 시계열이 갑자기 크게 변하는지, 그리고 그 변화가 바로 소멸되는지 아니면 오래 지속되는지 등을 살펴보아야 한다(정동빈과 원태연, 2003).

만약 이러한 현상들이 나타난다면, 비정상 시계열을 적절한 변환을 통해서 정상시계열로 변환한 후에 모델식별을 하여야 한다.

시계열의 적절한 변환방법으로는 ① 차분을 통해 추세를 제거하는 방법과 ② 분산안정화 변환방법이 있다.

① 추세성의 제거

시계열 도표를 통해 추세가 있는 경우에는 추세가 큰 의미를 부여하지 않는다면, 차분을 통해서 추세를 제거하여 정상시계열로

변환한다.

그러나, 차분을 수행할 때 주의해야 할 점은 ‘과다차분’이며 일반적으로 선형추세의 경우는 1차 차분을 통해서 추세성분이 제거된다.

과다차분 여부를 확인하기 위해서는 차분을 한 후 차분된 시계열의 시계열도표와 자기상관함수도표를 다시 그려보고 시계열의 정상성 여부를 판단해야 한다.

이처럼 이미 정상화된 시계열에 대하여 다시 차분을 실시하는 것을 과다 차분이라 하며, 과다차분은 시계열의 분산을 증가시키거나 자기상관함수를 복잡하게 만든다(이덕기, 1999).

② 분산안정화 변환

시계열 도표에서 시계열의 분산이 시간의 흐름에 따라 변한다면, 비정상 시계열이며, 비정상 시계열은 분산안정화변환을 통해 분산이 일정하게 되도록 정상시계열로 변환한다.

일반적으로 많이 이용되는 분산안정화 변환으로는 로그변환과 루트변환 그리고 Box-Cox변환 등이 있다.

(2) 모델 식별 단계

모델식별이란 $ARIMA(p,d,q)(P,D,Q)_s$ 모델의 차수 p,d,q 와 P,D,Q 를 결정하는 것이다.

모델식별을 위해서 일반적으로 사용하는 도구로는.

① 자기상관함수 (ACF : Auto-Correlation Function)

② 편자기상관함수 (PACF : Partial Auto-Correlation Function) 가 있다

즉, 주어진 시계열의 자기상관함수와 편자기상관함수의 도표를 그린

후 이미 알려져 있는 각 시계열 모델의 자기상관함수와 편자기상관함수의 형태 비교를 통해서 주어진 시계열의 모델을 식별하게 된다.

자기상관함수와 편자기상관함수는 혼합모델의 경우, 모델식별이 어려우므로 이러한 문제점을 보완하기 위해서는 아카이케 정보판단기준(AIC)과 베이저언 정보 판단기준(BIC)을 사용한다.

③ 아카이케 정보판단기준(Akaike Information Criterion)

AIC를 수식으로 나타내면 다음과 같다.

$$“ AIC=n \ln \hat{\sigma}_a^2 + 2(p+q) ”$$

여기서 n 은 시계열의 관측값들의 개수이고, $\hat{\sigma}_a^2$ 은 σ_a^2 의 최대우도(Maximum Likelihood)추정량이며, $(p+q)$ 는 설정된 모델의 모수의 수로 모델을 과다 적합하여 모수를 너무 많이 포함시키는 것을 견제하는 일종의 벌칙함수의 역할을 한다(이덕기, 1999).”

모델식별 방법은 복수의 잠정모델에서 AIC를 기준으로 최소값을 가지는 모델을 선정한다.

④ 베이저언 정보판단기준(BIC= Bayesian Information Criterion)

AIC와 유사한 정보판단기준으로 BIC 통계량 식은 다음과 같다

$$“ BIC= n \ln \hat{\sigma}_a^2 + 2(p+q) \ln n ”$$

여기서 n 은 시계열의 관측값들의 개수이고, $\hat{\sigma}_a^2$ 은 σ_a^2 의 최대우도(Maximum Likelihood)추정량이며, $(p+q)$ 는 설정된 모델의 모수의 수이다(이덕기, 1999).”

모델식별방법은 AIC와 같이 여러 개의 잠정모델 중 작은 BIC값을

가질수록 모델이 잘 적합되었음을 의미한다.

(3) 모수 추정 단계

모델 식별 단계에서 모델이 식별되면, 다음 단계로 식별된 모델의 모수를 추정하여야 한다.

식별된 ARIMA 모델의 모수를 추정하는 방법으로는 크게

- ① 조건부 최소제곱법(CLS: Conditional Least Squares Estimation),
- ② 비조건부 최소제곱법(ULS: Un-conditional Least Squares Estimation),
- ③ 최대우도추정법(ML: Maximum Likelihood Estimation) 등이 있다.

① 조건부 최소제곱추정(CLS: Conditional Least Squares Estimation)법

: 조건부 최소제곱법으로 모수 추정을 간편하게 하기 위하여 처음 몇 개의 관측값에 조건화하여 잔차제곱합을 최소화하는 방법이다. 계산이 비교적 간단하나, 처음 몇 관측값이 갖는 정보가 효율적으로 이용되지 못한다.

② 비조건부 최소제곱추정(ULS: Unconditional Least Squares Estimation)법

: 조건부 최소제곱법과 같은 조건이 없이 잔차제곱합을 최소화하는 방법이다. 이는 자료의 수가 크지 않을 경우나 계절 시계열 모델에서는 추정값에 영향을 미치게 되므로 초기값을 지정하지 않고, 오차제곱합을 최소화시키기 위해 비조건부 최소제곱법을 사용한다.

③ 최대우도추정(ML: Maximum Likelihood Estimation)법

: 최대우도추정법은 시계열의 결합밀도함수인 우도함수를 최대화시키는 모수를 구하는 방법이다.

이를 위해서는 백색잡음의 분포를 가정해야 하며, 일반적으로 백색잡음의 분포는 평균이 0, 분산이 σ_a^2 인 정규분포로 가정한다.

(4) 모델 진단 단계

모델 진단 단계는 추정된 모델의 적합성을 검토하는 단계이다.

ARIMA 모델에서는 잔차의 백색잡음 여부를 확인하는 ‘잔차분석 (Residual Analysis)’을 통해서 추정된 모델이 타당한지의 여부를 판단한다.

시계열 모델의 오차항 a_t 즉, 모델의 잔차는 다음의 백색잡음 성질을 만족해야 한다(노형진, 2007).

- i) 오차들은 서로 독립적어서, 자기상관계수 $\rho_k = \text{Corr}(Y_t, Y_{t-k}) = 0$ 이고,
- ii) 평균은 $E(a_t) = 0$, 분산은 $V(a_t) = \sigma_a^2$ 이며,
- iii) 정규분포를 한다. $a_t \sim N(0, \sigma_a^2)$

따라서, 잔차들의 시계열 도표와 자기상관함수도표에서 잔차들의 평균 0을 중심으로 랜덤하게 움직이고, 특정한 패턴이 보이지 않으며, 분산이 일정하고 또한 잔차들의 자기상관함수가 모든 시차에서 신뢰한계내에 존재한다면, 이것은 잔차들의 자기상관 함수가 0 이라는 귀무가설($H_0 : \rho_k=0$)을 채택하게 되어 잔차들이 통계적으로 서로 독립적이라는 것을 의미한다

(5) 예측

시계열 분석의 목적은 시계열 모델의 구축과 예측에 있으며, 연속성의 원리(The Principle of Continuity)에 입각하여 미래를 예측하는 방법이다.

따라서, 시계열 분석을 통한 예측에서는 관측된 과거의 자료들을 분석하여 일정한 규칙을 발견해서 이를 모델화하여 추정하고, 이 추정된 모델을 사용하여 미래에 관측될 값을 예측하게 된다

이러한 예측모델은 과거에 일어났던 일이 예기치 못한 사건에 의하여 방해받지 않는 한 미래에도 재연될 것이라는 가정이므로 예측모델을 구축하는 과정에서 오차가 발생할 수 밖에 없으며, 이러한 오차는 확률분포에 의해 구하게 된다.

또한, 시계열 자료들을 통해 분석한 결과가 미래에도 지속되고, 일정한 규칙성을 가지고, 시계열 자료의 타당성과 신뢰성을 확보한다면, 예측의 정확성을 높일 수 있다.

3. SARIMA 를 활용한 수요예측 모델 구축

(A) SARIMA 모델

SARIMA 모델의 시계열 자료들은 일정한 시간간격을 두고 동일한 현상이 반복되는 경향을 보이는 계절성을 가지고 있다.

이러한 반복적인 현상이 계속 일어나는 시간 간격을 계절주기라고 하며, 이러한 특징을 지닌 시계열을 계절 시계열이라고 한다.

SARIMA 모델은 비계절적 부분과 계절적 부분, 계절주기를 포함한 ARIMA (p,d,q)(P,D,Q)_s로 표현된다.

SARIMA 모델은 가장 활용도가 높은 시계열 모델 방법론으로 크게 자기회귀부분, 차분, 이동평균부분으로 구성되어 있다(허명희, 2009).

(1) 자기회귀부분

자기회귀과정은 AR(p)과정이라고 하며, 시계열의 현재값은 과거의 영향을 받아 실현된다는 가정하에 이루어지는 것이고, P 차 AR 모델은 다음과 같이 표현된다.

$$AR(p); Y_t = \phi Y_{t-1} + \phi Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t.$$

여기서 a_t 는 평균이 0 이고 분산이 σ^2 인 정규분포를 독립적으로 따르는 잡음(오차)항이다.

(2) 차분

보통의 시계열 자료는 추세를 가지고 있거나 변동이 불규칙하므로 정상성 가정을 충족시키지 못하므로, 비정상 시계열 자료를 분석하기 위해서는 차분을 통하여 추세를 제거하여야 한다.

또한 계절성을 띠고 있는 계절 시계열은 ‘계절차분(Seasonal difference)’을 통해 추세를 제거하여 시계열을 정상화시킨다.

(3) 이동평균부분

이동평균과정은 시계열을 백색잡음 과정의 누적평균으로 표현 되도록 한 확률과정이며, q 차 MA 모델은 다음과 같이 표현된다.

$$MA(q) : z_t = a_t - \theta a_{t-1} - \theta a_{t-2} - \dots - \theta a_{t-q}.$$

그리고, $\theta_1, \theta_2, \dots, \theta_q$ 등은 자료로부터 추정되어야 할 모수이다(허명희, 2009).

여기서 a_t 는 평균이 0 이고 분산이 σ^2 인 정규분포를 따르는 잡음(오차)항이다.

(B) SARIMA 모델의 적용

(1) 시계열 예비단계

관측된 시계열의 특성을 파악하기 위해서는 시계열도표과 자기상관함수도표를 그려서 시계열의 추세성분이 있는지, 계절성분이 있는지, 아니면 분산이 변하고 있는지 등을 알아보아야 하며, 보통의 시계열 자료는 추세를 가지고 있거나 변동이 불규칙한 비정상 시계열이므로 적절한 변환을 통해서 정상적인 시계열로 변환하는 것이 필요하다.

본 연구에서는 응급의료센터의 일일 내원환자수의 시계열 도표와 편자기상관함수(PACF), 자기 상관함수(ACF) 도표에서 별다른 추세는 보이지 않았으나, 계절적 추세로 인해 비계절적 차분인 $d=0$, 계절적 차분인 $D=1$ 로 설정하여 정상적인 시계열로 변환하였다.

(2) 모델식별단계

정상화된 시계열에 대하여 자기상관함수와 편자기상관함수 도표를 통해서 잠정적인 모델을 설정한다.

편자기상관함수(PACF:partial Auto-Correlation Function)와 자기상관함수(ACF:Auto-Correlation Function) 도표에서는 PACF에서 1번째의 상관계수가 유의하게 돌출되어 있었으므로 AR(1) Model를 잠정모델로 결정할 수 있으며, ACF에서 1번째에서 유의하게 돌출되어 있었으므로 MA(1) Model를 잠정모델로 결정할 수 있었다.

비계절적 자기회귀를 나타내는 $p=1$ 로, 계절적 자기회귀를 나타내는 $P=0$ 로 설정, 비계절적 이동평균을 나타내는 $q=1$ 로, 계절적 이동 평균을

나타내는 $Q=1$ 로 설정하여 단변량 SARIMA 모델을 식별하였다.

이와 같이 모델식별 단계에서는 ARMA모델의 식별을 AR모델과 MA모델의 절충모델을 여러 개 시도해 본 뒤, 적합도가 좋으며 최대한결한 모델을 선택하여 결정하여야 한다.

즉, “순수 AR모델 중에는 차수가 p 인 것이 좋게 나타났고, 순수한 MA모델 중에는 차수가 q 인 것이 좋게 나타났다면, 다수의 ARMA(k, l) 모델에서는 잠정모델에 대한 차수를 $1 \leq k \leq p-1, 1 \leq l \leq q-1$ 으로 고려해 볼 수 있다(허명희, 2009).”

본 연구에서는 최종 후보 모델로 1) 이동평균 모델 : MA (2) Model과 SPSS version 15.0의 시계열 모델 자동판별 Modelar 모듈을 통해 2) 단변량 Seasonal ARIMA 모델 : SARIMA(1,0,1)(0,1,1)₇, 3) 다변량 Seasonal ARIMA 모델 : SARIMA(1,0,2)(0,1,1)₇ 을 구축하였다.

(3) 모델추정 및 진단단계

구축된 수요예측 모델의 추정(Estimation)은 최대우도추정법(Maximum Likelihood)에 의해 이루어졌으며, 모델 진단을 위하여 예측모델의 잔차를 분석하였다.

그리하여 구축된 수요 예측모델 중 최종모델을 선정하기 위하여 적합모델의 엔트로피를 측정한 것인 AIC(AKaike Information Criterion)와 BIC(Bayesian Information Criterion)기준 값을 비교하였으며, 그 값이 작은 모델을 최적 예측모델로 선택하였다.

(4) 예측

수요예측모델을 평가하기 위하여 응급의료센터 일일 내원환자 수의 예측값과 관측값의 차이인 예측오차의 상대적인 오차의 크기를 나타내는 MAPE (Mean Absolute Percentage Error)를 비교하여 예측정확도를 평가하였다.

MAPE (Mean Absolute Percentage Error)를 구하는 공식은 다음과 같으며, 여기서 X 는 실제값(관측값), F 는 예측값, 그리고 n 은 관측값의 기간 수를 각각 나타내며, 계산된 MAPE 값은 (Table 3.)과 같이 해석될 수 있다(Lewis, 1982; 이충기, 2003).

$$MAPE = \frac{1}{n} \sum \left| \frac{X_t - F_t}{X_t} \right| * 100$$

Table 3. Commentary of MAPE value.

MAPE	Commentary
$0\% \leq MAPE < 10\%$	Very accurate prediction
$10\% \leq MAPE < 20\%$	Relatively accurate predictions
$20\% \leq MAPE < 50\%$	A relatively reasonable prediction
$MAPE \geq 50\%$	Inaccurate forecasts

III. 결과

A. 연구결과

응급의료센터 환자정보를 분석한 결과, 2007년~2008년도 2년간 응급의료센터에 내원한 전체 환자수는 169,375명이었으며, 연간 평균환자수는 84,688명 (2007년 80,496명, 2008년 88,879명), 평균 일일 내원환자수는 232명이었다. 본 연구에 사용된 훈련 데이터집합(Training Data Set)과 평가 데이터 집합(Validation Data Set)간의 유의한 차이가 있는지에 대해 알아보기 위하여 카이제곱검정을 하였으며, 두 Data Set 간의 성별, 연령대, 응급진료결과, 진료구역에 대한 카이제곱검정 결과는 (Table 4.)과 같으며, 유의확률 P-value가 0.05보다 커 두 군간의 유의한 차이는 없었음을 확인하였다.

Table 4. The results of Comparison between Training Data Set and Validation Data Set by χ^2 test.

	Characters	Training Data Set	Validation Data Set	P-Value
Sex	Male	47.90%	47.10%	0.876
	Female	52.10%	52.90%	
Age Group	Geriatric (Age>65)	34.00%	31.90%	0.627
	Adult (16≤Age<65)	37.00%	42.00%	
	Pediatric (Age<16)	28.90%	26.10%	
ER care Result	Discharge	25.60%	26.90%	0.237
	Others	3.60%	0%	
	Death	9.00%	5.90%	
	Admission	25.60%	31.10%	
	Transfer	15.70%	17.60%	
	Cancel	20.50%	18.50%	
Emergency Zone	1 st Zone (New Pt area)	15.10%	18.50%	0.071
	2 nd Zone (Med.-Adult)	14.20%	14.30%	
	2 nd Zone (Surgery-Adult)	13.30%	18.50%	
	3 rd Zone (Child area)	12.30%	14.30%	
	9 th Zone (CPR,ICU)	16.60%	18.50%	
	Delivery	14.50%	5.90%	
	Waiting Room for Admission	9.90%	10.10%	

The results of Comparison between Training Data Set and Validation Data Set by χ^2 test is non-significant (p-value > 0.05).

(A) 시계열의 사전조정

응급의료센터의 일일 내원하는 환자수의 시계열 도표는 (Fig. 2.)과 같이 나타났으며, 응급의료센터 일일 환자수는 일요일에 증가하였다가 화요일부터 금요일까지는 낮게 분포하는 양상을 보이는 7 일을 주기로 반복됨을 확인하였다.

또한, 시계열 도표에서 시간이 흐름에 따라 증가하거나 감소하는 추세는 있지 않았으며, 분산이 커지거나 작아지는 양상 등의 별다른 추세가 보이지 않아 차분이 필요하지는 않았지만, 7 일을 주기로 반복되는 계절적 추세가 있어 1차 계절적 차분을 통해 정상성을 유지하였다.

모델구축을 위한 2년간의 데이터를 계절적 1차 차분을 한 시계열 도표는 (Fig. 3.)와 같이 나타났다

계절적 차분을 한 시계열 도표에서 평균과 분산의 변화를 뚜렷하게 볼 수 없어 정상시계열임을 확인할 수 있었다.

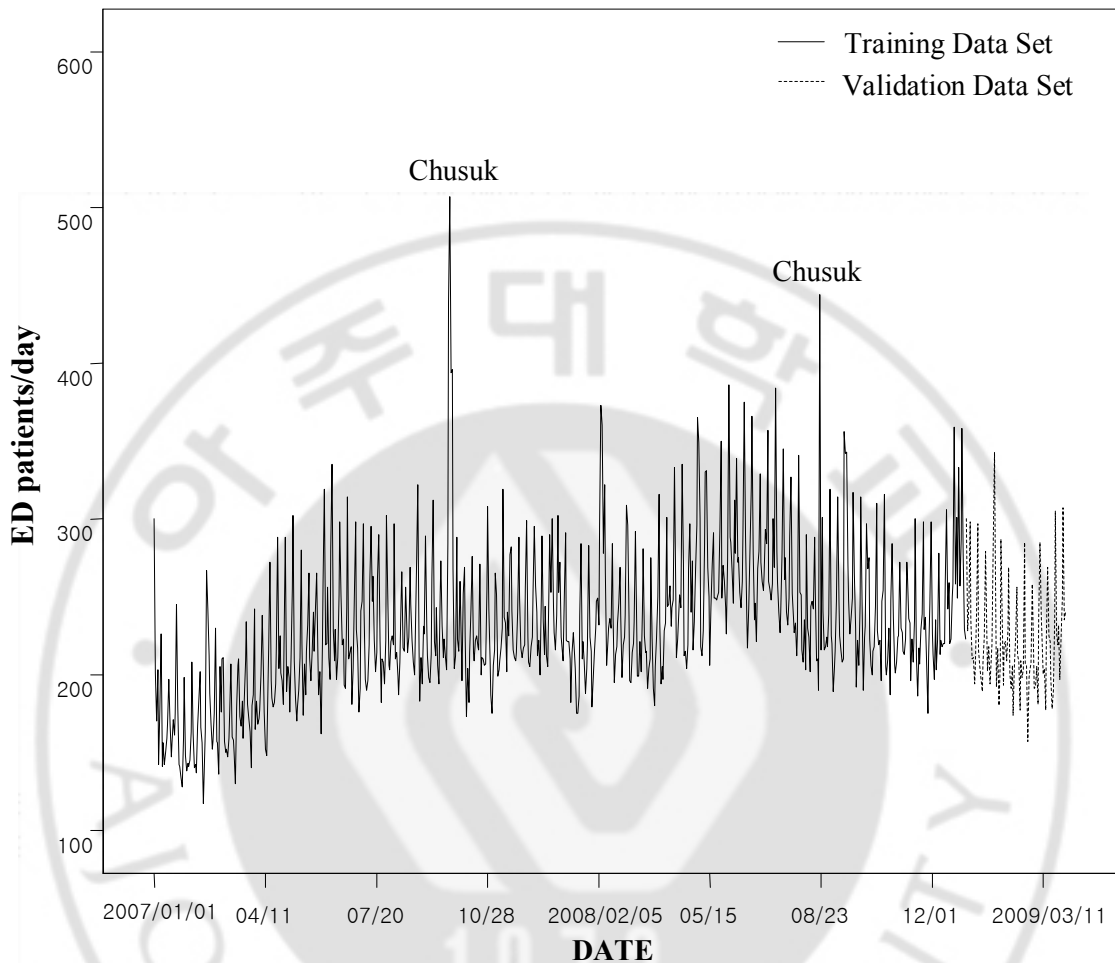


Fig. 2. Time Plots of Daily Emergency Department Patients(2007.01~2009.03). During the period from January 2007 to March 2009, Total of 189,511 ED patients visited and average number of daily patients was 231. The Sequencing graph showed 7-day periodicity and seasonal trend. In particular, there was a sharp increase in the number of patients in Chusuk.

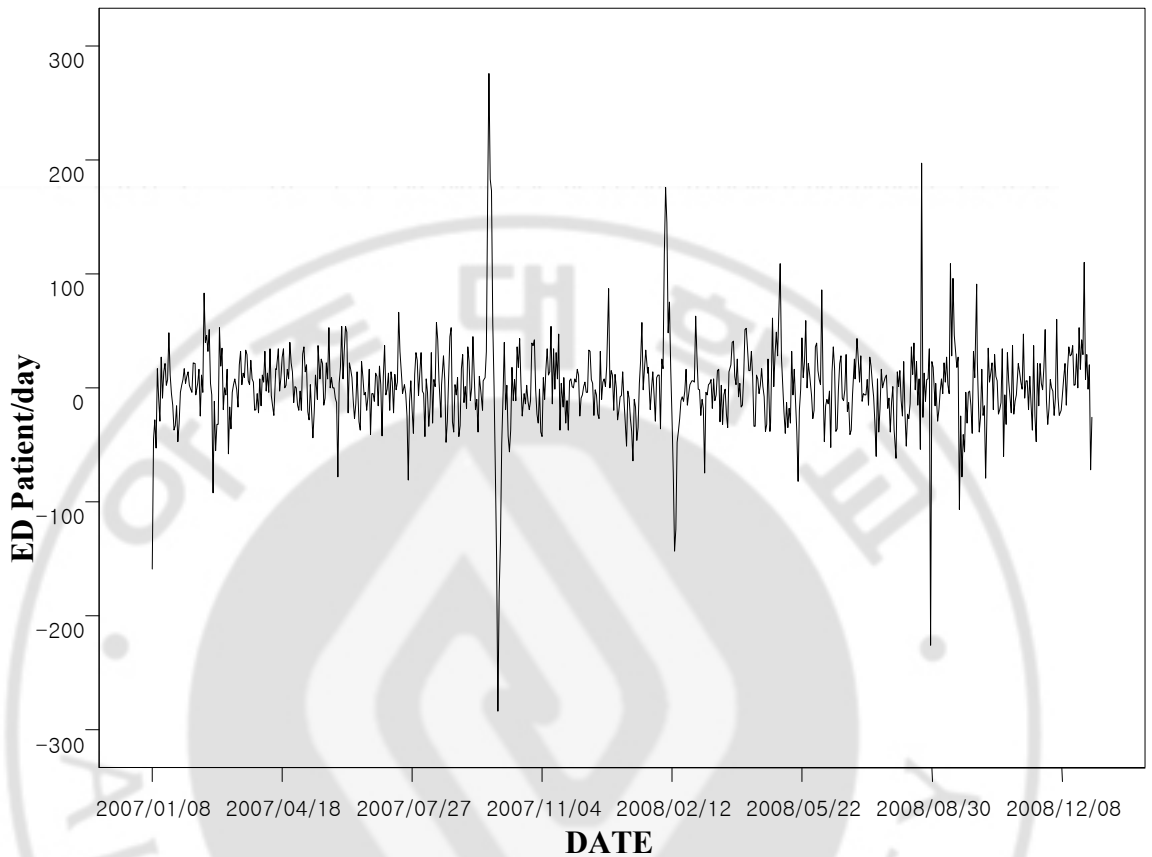


Fig. 3. Transforms of Daily Emergency Department Patients using seasonal difference.

(B) 모델 식별

모델 식별을 위해서 편자기상관함수(PACF: Partial Auto-Correlation Function)와 자기상관함수 (ACF : Auto-Correlation Function)를 이용하고자 그려진 상관함수 도표는 (Fig. 4.)과 (Fig. 5.)같으며, PACF에서 1번째의 상관계수가 유의하게 돌출되어 있었고, AR(1) Model과 ACF에서 1번째에서 유의하게 돌출되어 있었으므로 MA(1) Model을 잠정모델로 결정할 수 있음을 확인하였다.

그리고, 본 연구에서 최종 후보 모델로 1) 이동평균 모델 : MA (2) Model과 SPSS의 시계열모델의 자동관별 Modelar 모듈을 통해 2) 단변량 Seasonal ARIMA 모델:SARIMA(1,0,1)(0,1,1)₇, 3) 다변량 Seasonal ARIMA모델: SARIMA (1,0,2)(0,1,1)₇ 구축하였다.

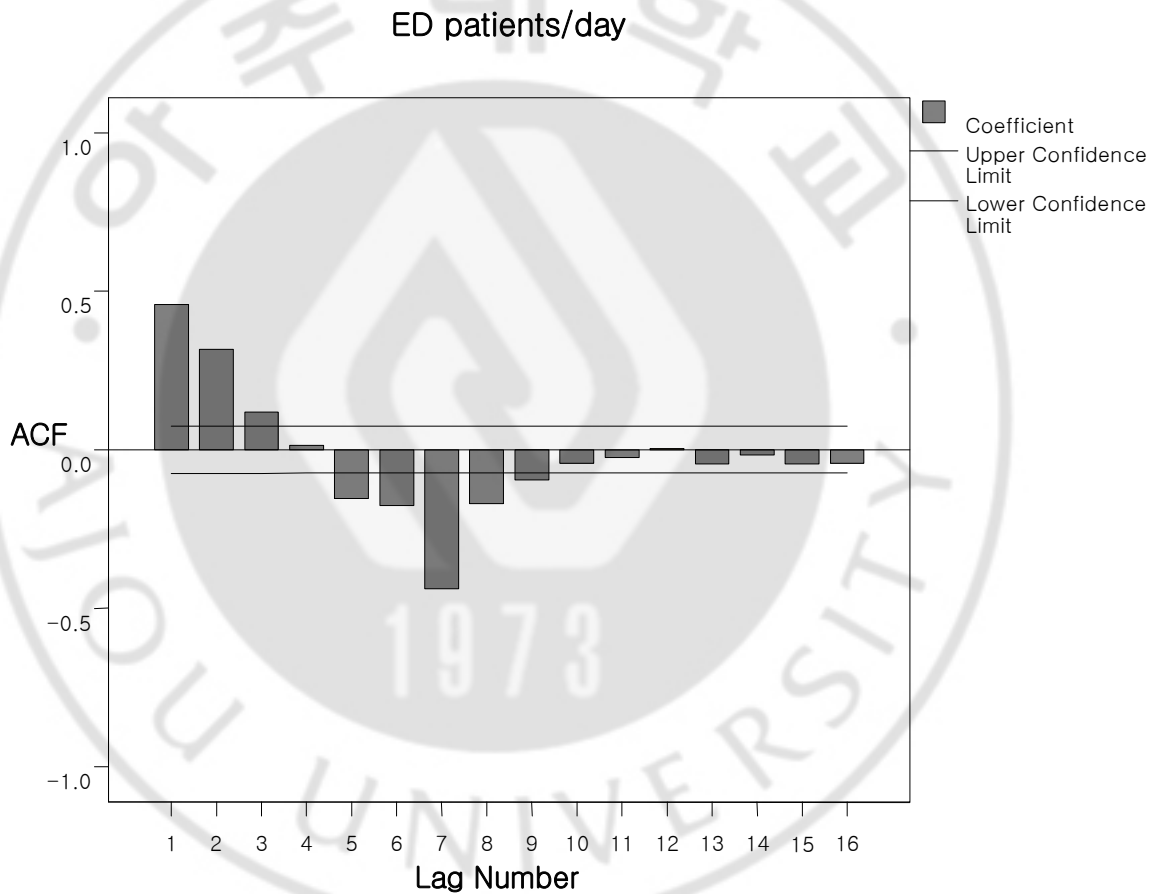


Fig. 4. ACF Correlogram of Daily Emergency Department Patients.

ED patients/day

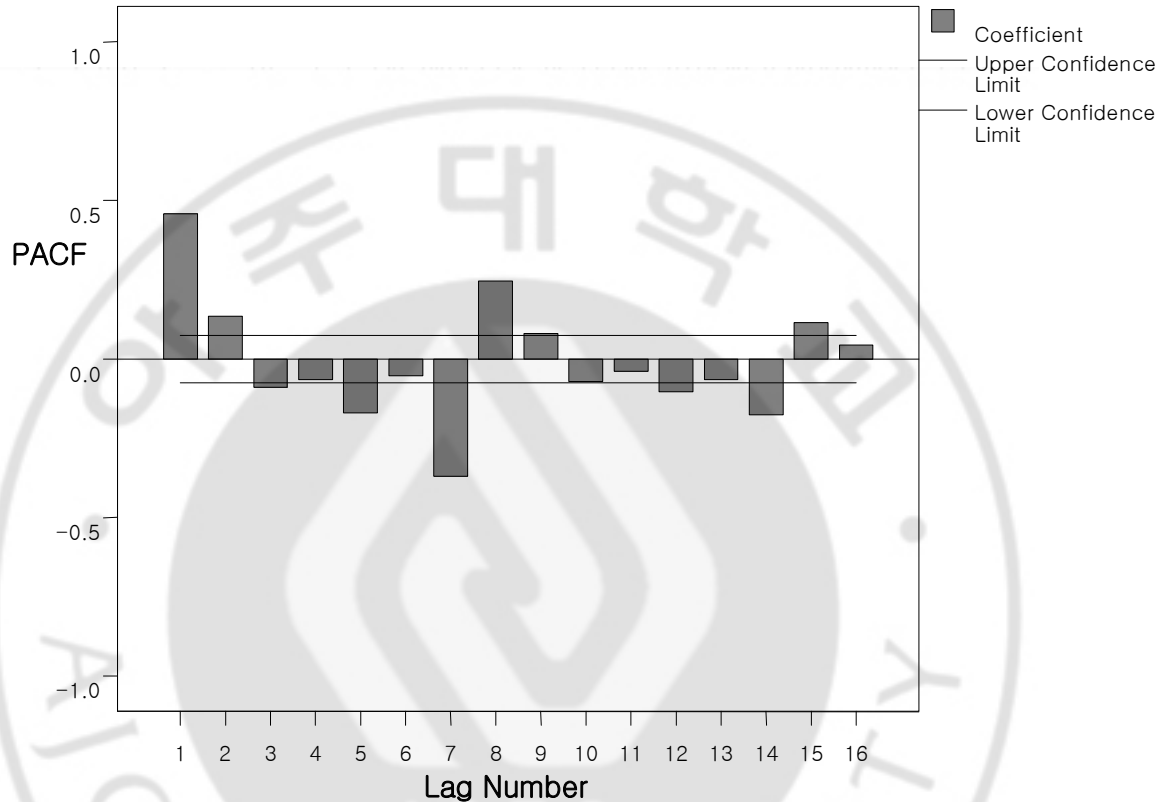


Fig. 5. PACF Correlogram of Daily Emergency Department Patients.

(C) 모델추정 및 진단단계

구축된 모델은 SPSS Time Series Modeler에 의해 자동판별된 모델이며, 모델을 판별하는 알고리즘은 가장 좋은 모델을 BIC를 기준으로 적용하여 선택하게 된다.

(1) 잔차분석

구축된 모델을 진단하기 위해서는 일반적으로 많이 사용되는 잔차 분석 (Residual Analysis)을 이용하였으며, 세 모델의 잔차분석 결과는 (Table 5.)와 같다.

Table 5. Residual Analysis of MA(2) Model, Univariate SARIMA Model, Multivariate SARIMA Model.

	MA(2) Model	Univariate SARIMA Model	Multivariate SARIMA Model
White noise	X	O	O
P-value	0.000	0.164	0.063

Two SARIMA Model are satisfied with the nature of White noise and MA(2) Model is dissatisfied with it.

① 이동평균 MA모델 - MA(2)

시차 2를 이용한 이동평균법으로 구축된 예측모델의 잔차에 대한 분석은 (Fig. 6.)와 같으며, 여기서 유의확률 $0.000 < \text{유의수준 } 0.005$ 이므로 가설 $H_0 : \rho(1)=\rho(2)=\rho(3)=\dots=\rho(m)=0$ 는 기각된다.

따라서 이 추정된 모델의 모수들은 유의하지 않으며, 적합된 모델에 의해서 추정된 잔차(관측값-적합된 값)들은 모델에 대한 정보를 가지고 있어 백색잡음(White noise)의 성질을 만족하지 못하였다. 즉, 모든 m차의 자기상관계수가 (Fig. 6.)와 같이 신뢰구간 내에 존재하지 않으므로 이 예측 모델은 잘 적합되지 못하였음을 알 수 있었다.

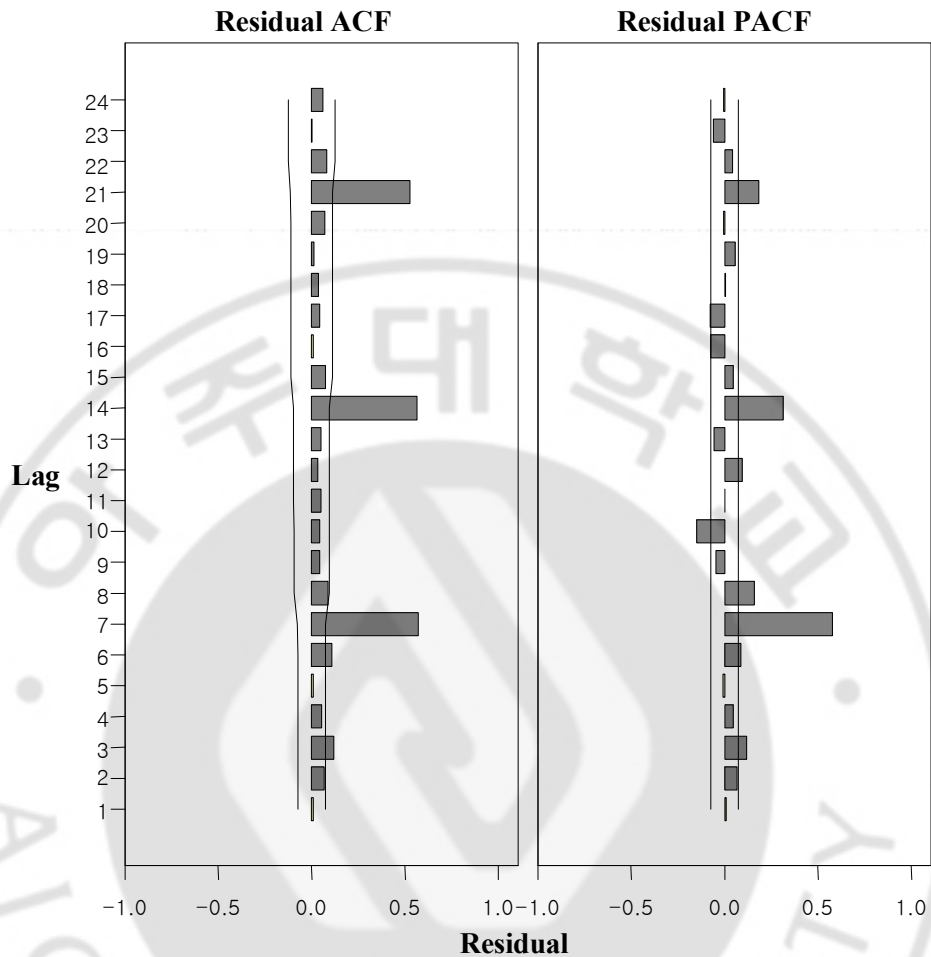


Fig. 6 . MA(2) Model Residual Analysis.

② 단변량 Seasonal ARIMA 모델 - ARIMA(1,0,1)(0,1,1)₇

Seasonal ARIMA 모델 ARIMA(1,0,1)(0,1,1)₇의 잔차에 대한 분석은 Fig. 7.과 같이 나타났으며, 유의확률 $0.164 > \text{유의수준} 0.005$ 이므로 가설 $H_0 : \rho(1)=\rho(2)=\rho(3)=\dots=\rho(m)=0$ 는 기각되지 않는다.

따라서 이 잔차는 백색잡음(White noise)이라고 생각할 수 있으며, 모든 m 차의 자기상관계수가 신뢰한계에 있으므로 이 잔차들의 자기상관함수가 0이라는 귀무가설을 채택하게 되고, 잔차들은 통계적으로

자기상관이 없이 독립적이라는 것을 알 수 있었다. 따라서 이 예측모델은 잘 적합되었음을 알 수 있었다.

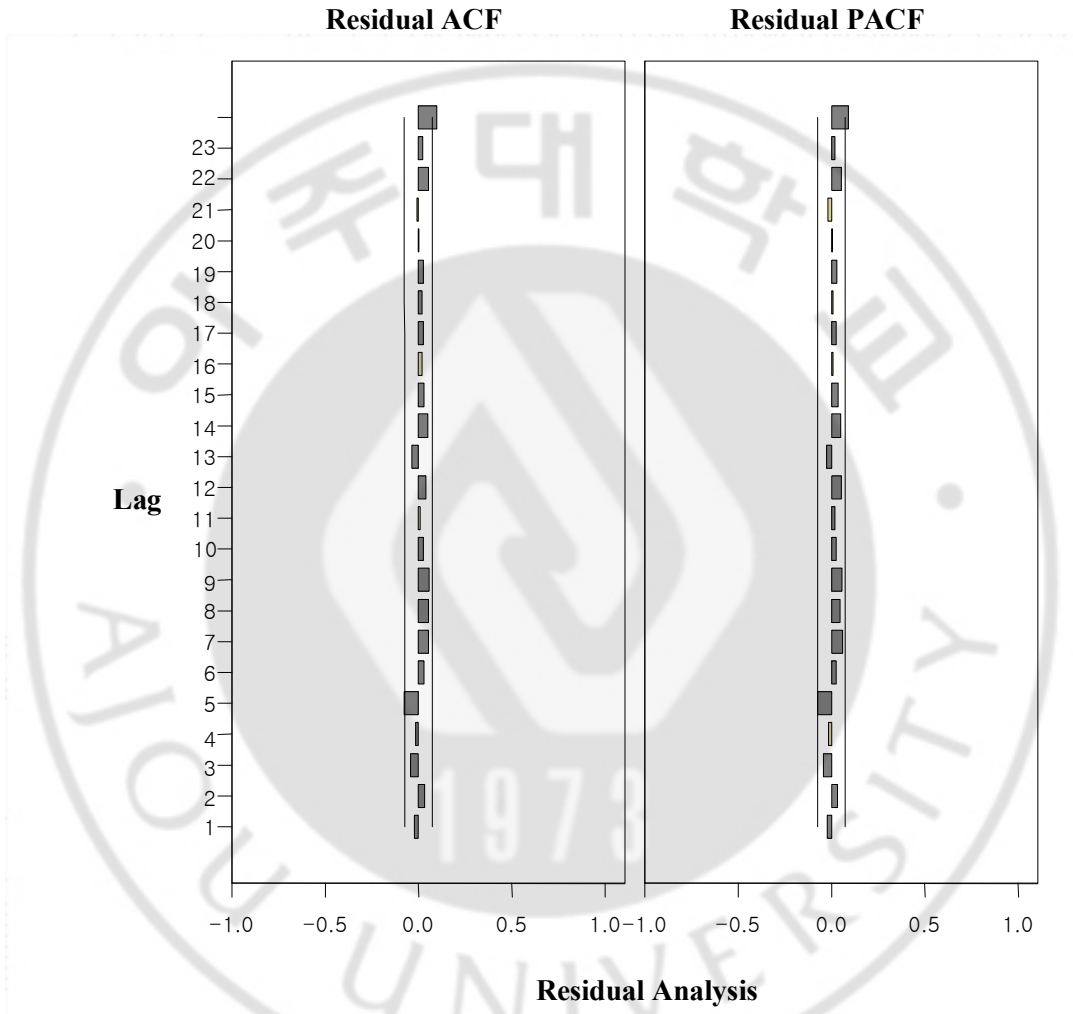


Fig. 7. SARIMA(1,0,1)(0,1,1)₇ Model Residual Analysis.

③ 다변량 Seasonal ARIMA 모델 - SARIMA(1,0,2)(0,1,1)₇

기상변수와 날짜변수를 고려한 Seasonal ARIMA(1,0,2)(0,1,1)₇의 잔차에 대한 분석은 (Fig. 8.)와 같이 나타났으며, 유의확률 0.063 > 유의수준 0.005이므로 가설 $H_0 : \rho(1)=\rho(2)=\rho(3)=\dots=\rho(m)=0$ 는 기각되지 않는다.

따라서 이 잔차는 백색잡음(White noise)이라고 생각할 수 있으며, 모든 m 차의 자기상관계수가 신뢰한계에 있으므로 이 잔차들의 자기상관함수가 0이라는 귀무가설을 채택하게 되고, 잔차들은 통계적으로 자기상관이 없이 독립적이라는 것을 알 수 있었다. 따라서 이 예측모델은 잘 적합되었음을 알 수 있었다.

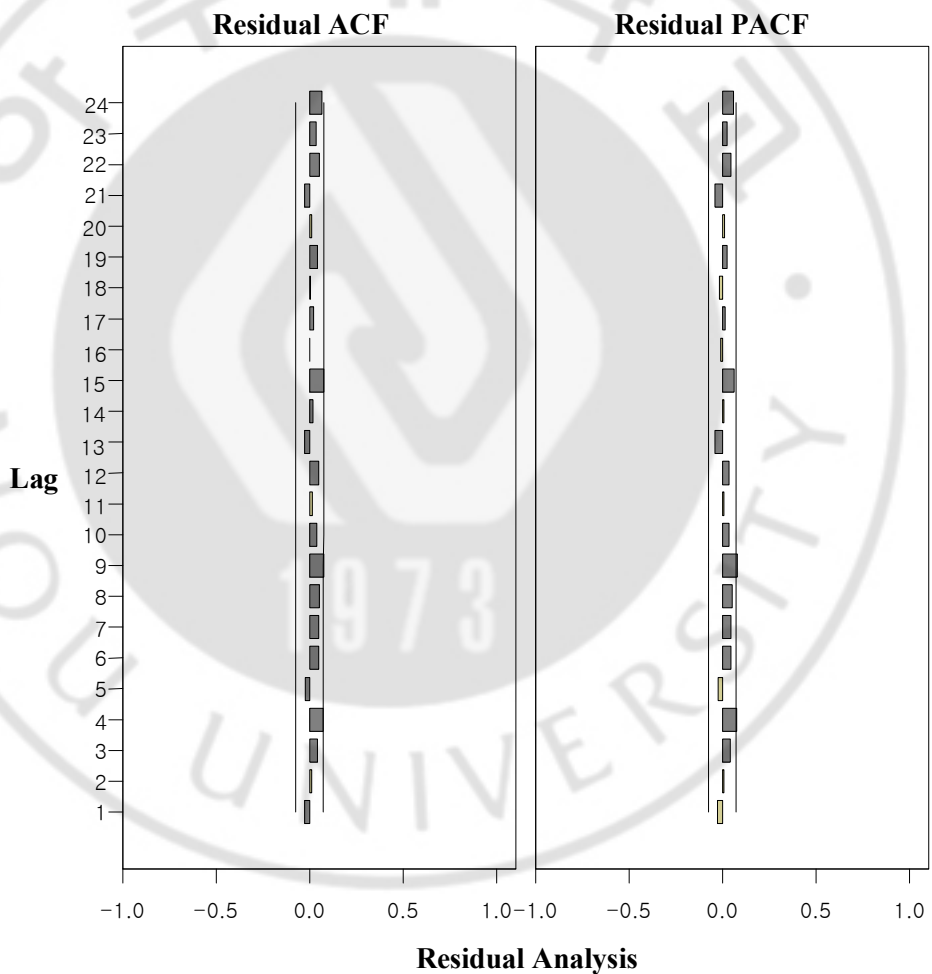


Fig.8. SARIMA(1,0,2)(0,1,1)₇ Model Residual Analysis

(2) 모델 추정(비교)

세 개의 예측모델의 모수 추정은 최우추정(ML:Maximum Likelihood Estimation)법에 의해 이루어졌으며, 다변량 계절형 ARIMA 모델의 모수추정 결과는 (Table 6.)와 같이 나타났다.

모수 추정 결과, 다변량 모델구축시 고려하였던 독립변수(월, 일, 요일, 계절, 분기, 휴일, 추석여부, 평균기온, 최저기온, 최고기온, 일교차, 강수여부, 신적설여부, 풍속도, 상대습도, 황사여부) 중 예측모델이 채택한 설명변수는 추석여부, 계절별, 평균기온, 강수여부임을 확인하였다.

Table 6. Multivariate Seasonal ARIMA Model Parameters.

Variables	Transformation		Estimate	SE	t	Sig.	
ED patient/Day	NO	Constant	0.72	0.23	3.10	0.00	
		AR	Lag 1	0.49	0.04	13.59	0.00
		MA	Lag 2	-0.12	0.04	-2.85	0.00
		Seasonal Difference	1				
		MA, Seasonal	Lag 1	0.90	0.02	48.77	0.00
Chusuk(Y/N)	NO	Delay	1.00				
		Numerator	Lag 0	58.60	16.36	3.58	0.00
		Seasonal Difference	1				
Seasons	NO	Numerator	Lag 0	-12.00	4.67	-2.57	0.01
			Lag 1	-13.67	4.67	-2.93	0.00
		Seasonal Difference	1.00				
AVG-Temperature	NO	Numerator	Lag 0	0.99	0.27	3.75	0.00
		Seasonal Difference	1				
RAIN(Y/N)	NO	Delay	4.00				
		Numerator	Lag 0	9.12	3.19	2.86	0.00
			Lag 2	-10.33	3.20	-3.23	0.00
		Seasonal Difference	1				

Adopted 4 Predictors by Multivariate SARIMA(Seasonal Autoregressive Integrated Moving Average) Model were Season, Chusuk(Y/N), Average Temperature, Rain(Y/N).

또한, 예측모델들을 비교하기 위해서는 모델간의 AIC(Akaike Information Criterion)와 BIC(Bayesian Information Criterion)를 비교하였으며, 그 결과는 (Table 7.)과 같이 나타났다.

AIC 및 BIC 값이 적을수록 잘 적합된 모델로 판단할 수 있으므로, 본 연구에서는 가장 적합한 모델로 다변량 계절형 ARIMA 모델, 즉 Seasonal ARIMA(1,0,2)(0,1,1)₇ 모델을 채택하였다.

Table 7. Goodness of fits for Models (AIC, BIC)

Model	AIC	BIC
MA(2)	7448.4	7462.2
Univariate SARIMA(1,0,1)(0,1,1) ₇	6815.7	6834
Multivariate SARIMA(1,0,2)(0,1,1)₇	6703.7^a	6749.5^a

^aMultivariate SARIMA Model is The Best Model.

- SARIMA(Seasonal Autoregressive Integrated Moving Average)
- AIC(Akaike Information Criterion)
- BIC(Bayesian Information Criterion)

(D) 모델평가

예측모델을 평가하기 위한 모델의 예측정확도는 실제 관측값과 예측값의 차이인 예측오차의 상대적인 크기를 나타내는 MAPE(Mean Absolute Percentage Error)를 비교하였으며, 그 결과는 (Table 8.)과 같았다.

Table 8. MAPE values of constructed models

Model	MAPE
MA(2)	12.909
Univariate SARIMA(1,0,1)(0,1,1) ₇	7.788
Multivariate SARIMA(1,0,2)(0,1,1)₇	7.372^b

^bMultivariate SARIMA Model is The Best Model.

- SARIMA(Seasonal Auto-Regressive Integrated Moving Average)
- MAPE(Mean Absolute Percentage Error)

(E) 예측

세 예측모델의 예측기간 3 개월의 예측값과 관측값을 비교한 도표를 그려본 결과, 1) 이동평균(MA) 모델에 관한 비교표는 (Fig. 9), 2) 단변량 Seasonal ARIMA 모델에 관한 예측 비교는 (Fig. 10), 3) 다변량 Seasonal ARIMA 모델에 관한 예측 비교는 (Fig. 11.)과 같았다.

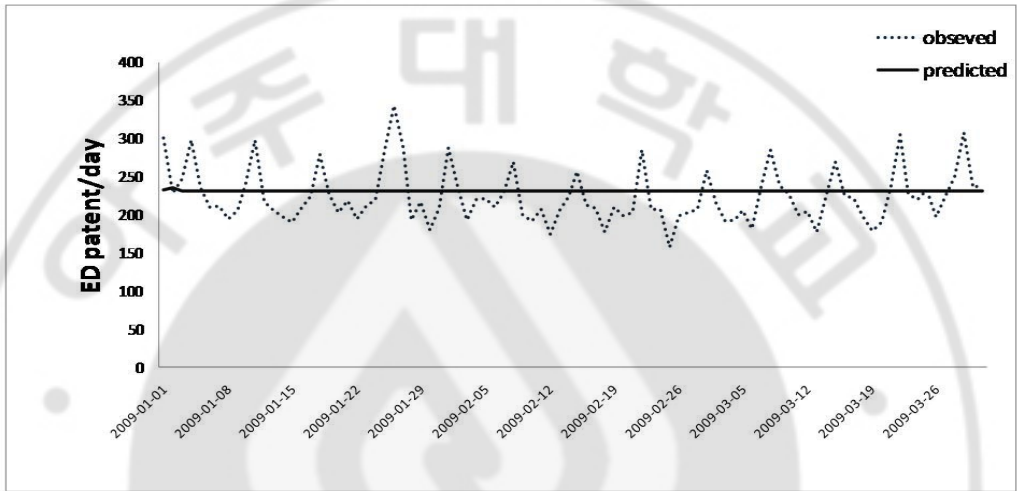


Fig. 9. Observed and predicted daily attendances at an ED ; MA(Moving Average) Model.

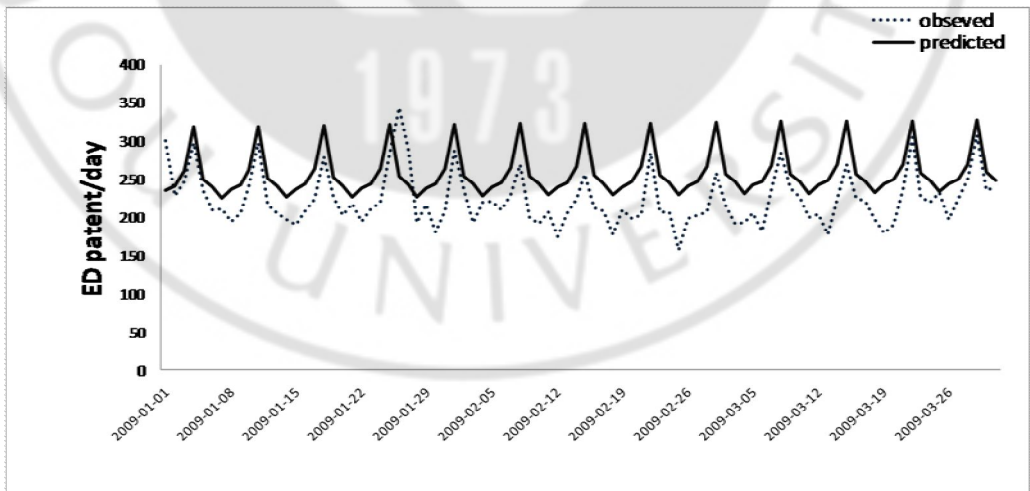


Fig. 10. Observed and predicted daily ED patients ; Uni-variate SARIMA(Seasonal Autoregressive Integrated Moving Average) Model.

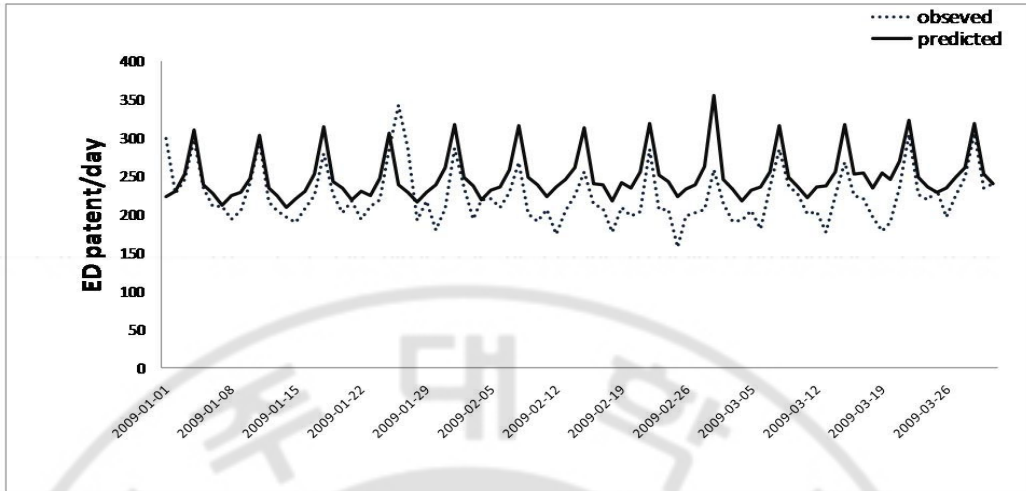


Fig. 11. Observed and predicted daily ED patients ; Multivariate SARIMA (Seasonal Autoregressive Integrated Moving Average) Model.



IV. 고찰

본 연구의 대상이 된 응급의료센터는 전담인력과 응급의료센터 별도의 진단 장비, 환자 감시장치, 치료장비, 기타 장비 등을 구비하고 있었으며, E-ICU, Emergency 병동 등의 시설을 구비하고 있는 권역응급의료센터로 2007 년 ~2008 년도 2 년간 응급의료센터에 내원한 전체 환자수는 169,375 명, 연간 평균환자수는 84,688 명 (2007 년 80,496 명, 2008 년 88,879 명), 평균 일일 내원환자수는 232 명으로 전국에서 상위권의 규모를 차지하고 있어 응급의료센터 일일 내원 환자 수 예측을 위한 적절한 연구대상이었다.

본 연구에 사용된 훈련 데이터집합(Training Data Set)과 평가 데이터집합 (Validation Data Set)간의 유의한 차이를 확인하기 위하여 카이제곱검정을 한 결과, 두 Data Set 간의 성별, 연령대, 응급진료결과, 진료구역의 유의한 차이는 없었으며, 본 연구기간 동안의 응급의료센터 일일 내원 환자 추이에 영향을 줄 수 있는 특별한 요인은 없이 국내 단일기관 응급의료센터의 일상적인 일일 내원 환자추이에 대한 수요예측을 할 수 있는 데이터 집합이었다.

응급의료센터 일일 내원 환자 수의 추이는 주말인 토요일부터 증가하여 일요일이 가장 많았고, 휴일 다음날인 월요일에 조금 감소하다가, 화요일부터 금요일까지는 낮은 분포로 7 일을 주기로 변동하는 양상과 계절적 추세를 보였다. 그리고, 7 일의 주기성을 벗어나는 특이점으로 ‘추석’과 ‘설’의 경우가 있었으며, ‘추석’에는 갑작스럽게 상승하는 양상을 보이는데 반해 ‘설’의 경우에는

상대적으로 약한 상승 추이를 보여 본 연구에서는 추석여부만을 변수로 선택하여 예측모델에 반영하였다.

본 연구에서 사용한 다변량 예측모델의 독립변수로는 날짜변수(월, 일, 요일, 계절, 분기, 휴일, 추석여부)와 날씨변수(평균기온, 최저기온, 최고기온, 일교차, 강수여부, 신적설여부, 풍속도, 상대습도, 황사여부)를 사용하였으며, 이 중 다변량 예측모델의 설명변수로 ‘추석여부’, ‘강수여부’, ‘계절’, ‘평균기온’을 채택하였다.

본 연구에서 보여준 응급의료센터의 일일 환자 추이는 이장영 등의 연구에서 요일별로는 일요일 92 ± 13 , 토요일 75 ± 11 , 월요일 71 ± 10 , 나머지 평일은 64~65 일의 분포와 거의 비슷하게 나타났으며, 국내 응급의료센터 일일 내원 환자수에 대한 수요예측모델에 있어 요일, 휴일, ‘추석’과 ‘설’에 대해 고려하여야 함을 알 수 있었다.

또한, 각 개별병원의 사회 문화적 배경, 지역적 특성에 따라 응급의료센터 일일 내원 환자 수에 미치는 영향의 차이가 있음을 국내에서 이루어진 이장영 등의 연구에서는 휴일 중 특히 추석과 설, 일요일과 같은 공휴일에 응급의료센터 일일 내원 환자수가 많았으나, Batel 등의 연구에서는 국내의 의료체계와 달리 한 주에 7 일 모두 하루에 15.5 시간을 진료하는 외래형 응급실(Walk-in-Clinic) 운영으로 인해 요일별 내원 환자수가 월요일이 가장 높았고, 일요일로 갈수록 낮아지는 양상을 통해 알 수 있었다.

그리고, 휴일 및 휴일 다음날(월요일)에 응급의료센터를 내원하는 환자수가 증가하는 양상은 적정인력 배치에 있어 주말 및 공휴일에 대해 고려하여야 함을 시사하는 것이다.

날씨변수에 있어서는 이장영 등의 연구에서는 일일 최고, 최저, 평균기온과 내원 환자수와의 상관분석 결과 약한 상관관계를 보였으며, 비가 내린 날과 그렇지 않은 날의 내원 환자의 수에는 차이가 없었고, 눈이 쌓인 날과 그렇지 않은 날의 내원 환자의 수에도 차이가 없었으나, 본 연구에서는 일일 내원 환자수가 강수여부에 의해 영향을 받는 것을 확인할 수 있었다.

선행연구를 통해서도 응급의료센터 일일 내원 환자 수에 영향을 미치는 대표적인 날씨변수로는 기온이 있었으며, 이는 기온이 상승하면 사람들의 활동이 증가됨으로 인해 사고발생이 증가할 수 있기 때문이다.

또한, 본 연구에서 채택한 날씨변수 중 강수여부는 비가 내리지 않는 날에 비해 비가 내리는 날에는 비 응급환자들의 내원이 줄어들었기 때문인 것으로 보인다.

Spencer 등의 연구에서와 같이 통상적으로 응급환자들의 응급의료센터 방문에는 날씨의 영향력이 미미할 수 있을 것이나, 국내 현실상 응급의료센터를 방문하는 환자들은 (1) 실제 응급상황으로 인한 진료를 받기 위해서, (2) 휴일 또는 야간에 진료를 받기 위해서, (3) 진료의뢰서가 없이 외래진료를 받기 위해서 응급의료센터를 방문하고 있으며, 황성완과 이해종의 연구를 통해서 본 바와 같이 초진 환자 4273 명 중 2276(53.3%) 명이 응급실을 경유해서 진료를 보는 것으로 나타났다(황성완과 이해종, 2008). 즉, 응급의료센터를 방문하는 비 응급환자들을 고려한다면, 국내 응급의료센터 수요예측모델에 있어서는 날씨변수를 고려할 필요가 있다.

위에서 언급한 것과 같이 국내 응급의료센터 일일 내원 환자수에 대한 수요예측모델에 있어서 고려되어야 할 지역적인 특성에 대해 살펴본 결과,

국외에서 이루어진 응급의료센터 수요예측에 관한 연구들이 국내의 의료체계와 사회문화적 배경, 지역적 요인 등을 고려하지 못한 점이 있었으므로, 국내 각 개별 단일병원의 특성을 반영한 수요예측 모델 개발의 필요성을 본 연구를 통해 확보할 수 있었다.

본 연구에서는 응급의료센터 일일 내원 환자수를 예측하기 위하여 1) MA 모델, 2) 단변량 계절형 ARIMA 모델, 3) 다변량 계절형 ARIMA 모델, 세가지 형태의 예측모델을 구축하였으며, 예측모델을 비교해 본 결과 MA(2) 모델을 제외한 두 개의 SARIMA 모델에서는 단변량 모델과 다변량 모델간의 차이가 있긴 했지만 추정된 모델의 잔차가 신뢰한계내에 포함되어 있어 두 모델 모두 예측모델이 잘 적합되었음을 확인할 수 있었다

이러한 결과는 MA 모델이 SARIMA 모델과 달리 시계열의 자기상관이라는 특성과 계절성을 반영하지 못했기 때문이다.

따라서, 국내 응급의료센터 선행연구 중 다중선형회귀분석법을 이용한 예측의 제한점인 1) 연구기간을 연장하거나 병원 수를 늘리는 것에 따라서 통계 결과 및 회귀식이 달라질 수 있다는 점과 2) 단일병원 연구의 오류를 줄이기 위해 4 개의 대학병원을 대상으로 하였지만, 특정시의 모든 응급환자를 대상으로 하지 못하였으므로 예측에 한계를 가진다는 점을 단일기관을 대상으로 시계열의 구성요소가 시간의 흐름에 따라 매우 빠르게 변동하는 경우에 효과적이며 어떤 형태를 가진 시계열 자료라도 분석이 가능한 장점을 가진 SARIMA 모델을 이용하여 수요예측모델을 구축함으로써 보완할 수 있었다.

또한, 예측모델의 예측정확도 MAPE 값에 있어서도 1) MA 모델, 2) 단변량 계절형 ARIMA 모델, 3) 다변량 계절형 ARIMA 모델 순으로 12.9%, 7.78%, 7.37%

으로 나타났으며, MA 모델 외의 두 모델의 MAPE 가 10% 미만이어서 정확하게 예측되었음을 확인할 수 있었다.

따라서, SARIMA 모델이 시계열의 자기상관이라는 특성과 계절성을 고려하여 예측모델을 구축하였기 때문에 예측정확도가 MA 모델에 비해 정확함을 알 수 있었고, 단변량 SARIMA 모델과 다변량 SARIMA 모델의 예측정확도는 예측기간 3 개월간의 관측값과 예측값의 시계열도표를 통해 최종예측모델로 선정한 다변량 SARIMA 모델이 더 정확하게 예측되었음을 확인할 수 있었다.

이는 다변량 SARIMA 모델이 단변량 SARIMA 모델과 달리 예측모델의 설명변수들이 설명력을 높여 주었기 때문이다.

본 연구에서는 예측모델의 예측기간을 3 개월로 정하였으나, 수요예측의 일반적인 특징에 따르면, 예측기간을 단축하였을 때 예측오차가 더 작아진다는 특징이 있으므로, 추후 연구진행시에는 예측모델 구축에 있어 예측기간을 1 개월이나 1 주로 정하는 등의 예측기간에 대한 세심한 고려도 필요할 것이다.

이처럼 수요예측은 아무리 정교한 기법을 사용한다고 해도 오차가 존재하기 마련이며, 완벽하게 실제값과 일치할 수 없으므로 하나의 기법을 이용한 예측 절대치를 사용하기 보다는 대안적인 방법들을 병행하여 활용함으로써 예측정확도를 높일 수 있을 것이다.

본 연구의 제한점으로는 다양한 수요예측기법에 대해 고려하지 못한 점과 지역적 특성을 반영한 좀 더 다양한 변수에 대해 고려하지 못한 점이 있었으며, 추후 연구에서는 응급의료센터 개별병원의 날씨변수, 날짜변수, 추석 여부, 계절 외의 다른 요인에 대한 연구들도 이루어져야 할 것이다.

다른 요인으로 고려해 볼 수 있는 것으로는 최근 SARS, 신종플루 등과 같은 신종전염병의 유행, 즉 어느 특정시점에 어떤 지역에서만 발생하는 특정질환의 발생 유무가 있을 수 있으며, 이러한 질환들로 인해 응급의료센터의 수요는 급격히 증가할 것이므로 이에 따른 응급의료센터의 수요예측을 통한 효율적인 자원활용이 필요할 것이다.

또한, 본 연구에서 사용한 예측방법인 다변량 계절형 ARIMA model 은 시계열의 구성요소가 시간의 흐름에 따라 매우 빠르게 변동하는 경우에 효과적이며 어떤 형태를 가진 시계열 자료라도 분석이 가능한 장점을 가지고 있으며, 설명변수를 통해 수요예측모델을 잘 설명할 수 있어 새롭게 발생하는 요인들을 잘 반영한 수요예측 모델을 개발할 수 있을 것으로 기대한다.

그리고, 최근에는 국가응급환자진료정보망(NEDIS: National Emergency Department Information System) 및 병원정보시스템(HIS : Hospital Information System), 전자의무기록(EMR: Electronic Medical Record)의 사용의 확대에 의해 환자 정보를 이용한 수요예측 모델 개발에도 촉진제가 될 것이라 기대하며, 향후에는 후향적 연구가 아닌 전향적 연구를 통해 수요예측모델에 대한 평가뿐 아니라 수요예측모델 도입으로 인한 인력배분에 대한 평가나 과밀화 개선에 대한 평가, 의료진의 반응을 평가하는 연구가 필요할 것이다.

V. 결론

본 연구는 국내 단일기관의 응급의료센터의 환자정보 및 지역의 날씨정보, 휴일을 포함한 날씨정보를 이용하여 응급의료센터 일일 내원 환자 수 예측 모델을 개발하고자 1) MA(2) 모델, 2) 단변량 SARIMA (1,0,1)(0,1,1)₇ 모델, 3) 다변량 SARIMA(1,0,2)(0,1,1)₇ 모델을 구축한 후 1) 잔차 분석과 2) AIC, BIC 값을 통해 모델을 비교하고, MAPE를 통해 예측정확도를 평가한 결과, MA 모델을 제외한 두 SARIMA 모델이 잘 적합되었음을 확인할 수 있었다.

이는 MA(2)모델이 응급의료센터 일일 내원 환자 수의 계절성과 자기 상관이라는 특징을 반영하지 못했기 때문이었으며, 단변량 SARIMA모델에 비해 다변량 SARIMA (1,0,2)(0,1,1)₇ 모델의 AIC와 BIC 값이 6703.7, 6749.5로 적어 가장 적합한 최종 모델로 선정하였다. 그리고, 다변량 SARIMA 모델의 설명변수로는 추석여부, 계절별, 평균기온, 강수여부를 채택하였고, 이 중 추석여부는 국내 응급의료센터의 사회 문화적 특성을 반영한 변수였으며, 강수여부는 국내 응급의료센터를 방문하는 환자 중 비응급환자들에 의한 영향으로 판단하였다.

다변량 SARIMA 모델의 예측정확도는 MAPE가 7.4%로 MAPE가 10% 미만이므로 정확하게 예측하였음을 확인할 수 있었다.

그러므로, 본 연구에서는 다변량 계절형 ARIMA모델이 비교되었던 MA모델 과 단변량 계절형ARIMA 모델에 비해 설명변수를 반영하고 있어 예측모델 설명력이 높고, 응급의료센터 일일 내원 환자 수에 대한 예측이 정확함을 확인할 수 있었다.

참고문헌

1. 김민정, 이대로, 김갑득: 주말이 응급의료센터에 내원하는 환자의 예후에 미치는 영향. *대한응급의학회지* 16: 356-62, 2005
2. 노형진: SPSS/Excel 에 의한 재미있는 시계열 분석. *도서출판 효산*, 2007
3. 서동우, 임경수, 문이상, 손유동, 조민우, 김원, 이상일: 응급의료관리료제도 시행 전후의 응급환자 내원 양상의 변화와 효과. *대한응급의학회지* 15: 228-232, 2004
4. 우주리: 시계열 모델을 이용한 방한 일본인 관광객 수요예측에 관한 연구 (석사학위 논문). *경희대학교 일반대학원*, 2007
5. 유인술, 허탁, 신상도, 박정배, 양혁준, 유연호: 응급실 과밀화 해소방안 마련을 위한 기초연구. 보건복지부 연구용역과제 최종보고서. *충남대학교*, 2007
6. 이덕기: 예측방법의 이해. *SPSS 아카데미*, 1999
7. 이은숙, 박금선: 응급의료서비스에 대한 만족도 분석. *대한응급의학회지* 5: 336-346, 1994
8. 이장영, 민진홍, 박정수, 정성필, 박준석, 정시경, 양영모: 요일 및 날씨 특성이 응급센터 내원 환자 수에 미치는 영향. *대한응급의학지* 16: 287-291, 2005
9. 이종원: 경제예측론. *도서출판 해남*, 2006

10. 이주호, 박영환: ARIMA 모델을 이용한 모든 교통량 예측에 관한 연구.
산업기술연구소 14: 81-96, 2002
11. 이충기: 관광응용경제학. 일신사, 2003
12. 정구영, 임경수, 민용일, 이삼범, 김세경: 응급환자의 현황과 응급의료의 실태
- 응급의료센터를 중심으로 -. 대한응급의학회지 8: 442-459, 1997
13. 정대유, 조수형, 김성중 : 광주 전남지역 응급실 진료환경 및 현황에 대한
보고. 대한응급의학회지 17: 116-123, 2006
14. 정동빈, 원태연: SPSS 를 활용한 시계열 자료와 단순화 분석 I. SPSS 아카데미,
2001
15. 정동빈, 원태연: SPSS 를 활용한 시계열 자료와 단순화 분석 II. SPSS 아카데미,
2003
16. 제상모, 최영환, 박유선, 조영순, 김승호 : 응급의학전문의를 얼마나 필요한가.
대한응급의학회지 16: 613-619, 2005
17. 최효선, 이경원: 전미 응급의료센터 과밀화 지수. 대한응급의학회지 17: 377-
84, 2006
18. 황성완, 이해중: 초진환자 재방문 예측모형 개발. 대한의료정보학회지 14:
137-145, 2008

19. 허명회: SPSS 시계열 분석. *SPSS*, 2009
20. Arnold JL, Song HS, Chung JM: The recent development of emergency medicine in seoul Korea. *Ann Emerg Med* 32: 730-735, 1998
21. Asplin BR, Flottesmesch TJ, Gordon BR: Developing models for patient flow and daily surge capacity research. *Acad Emerg Med* 13: 1109-1113, 2006
22. Batel H, Tencj J, Memillan S, Adams J, Mehler PS: Predicting patient visits to an urgent care clinic using calendar variables. *Acad Emerg Med* 8: 48-53, 2001
23. Derlet RW, Richards JR: Overcrowding in the nation's emergency departments: complex causes and disturbing effects. *Ann Emerg Med* 35: 63-68, 2000
24. Korea Meteorological Administration website. <http://www.kma.go.kr>
25. Lisa M. Schweigler, Jeffrey S. Desmond, Melissa L. McCarthy, Kyle J. Bukowski, Edward L. Ionides, John G. Younger: Forecasting Models of Emergency Department Crowding. *Acad Emerg Med* 16:301-308, 2009
26. Lewis CD; Industrial and Business Forecasting Methods, London. *Butterworth*, 1982
27. Schweiger L, Younger J, Ionides E, Desmond J: Autoregression models can reliably forecast emergency department occupancy levels 12 hours in advance. *Acad Emerg Med* 14(5 Suppl. 1):S82, 2007

28. Schull MJ, Vermeulen M, Slaughter G, Morrison L, Daly P: Emergency department crowding and thrombolysis delays in acute myocardial infarction. *Ann Emerg Med* 44: 577-585, 2004
29. Spencer S. Jones, MStat, Alun Thomas, PhD, R. Scott Evans, PhD, Shari J. Welch, MD, Peter J. Haug, MD, Gregory L. Snow, PhD: Forecasting Daily Patient Volumes in the Emergency Department. *Acad Emerg Med* 15: 159-170, 2008
30. Spencer S. Jones, R. Scott Evans, Todd L. Allen, Alun Thomas, Peter J. Haug, Shari J. Welch, Gregory L. Snow: A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of Biomedical Informatics* 42: 123–139, 2009
31. Sun BC, Adams J, Orav EJ, Rucker DW, Brennan TA, Burstin HR: Determinants of patient satisfaction and willingness to return with emergency care. *Ann Emerg Med* 35: 426-434, 2000
32. Tandberg D, Qualls C: Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med* 23:299–306, 1994
33. Yan Sun, Bee Hoon Heng, Yian Tay Seow, Eillyne Seow: Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emerg Med* 9: 1, 2009

- ABSTRACT -

Time Series Forecasting Modeling for Demand of Emergency Department

Jin Ok Sung

Department of Medical Sciences

The Graduate School, Ajou University

(Supervised by Associate Professor Rae Woong Park)

The purpose of this study is to develop and evaluate Models to forecast the number of patients who are admitted to an Emergency Department of a Korean hospital in a day as a solution to solve overcrowding issue of Emergency Department(ED).

Forecasting the number of Daily ED patients is the basis of demand forecast of an Emergency Department and it will be able to distribute and utilize restricted resources of an Emergency Department effectively according to the demand forecast.

The data used in the study was collected from the Hospital Information System(HIS) database. In order to develop a Forecasting Model, two-year data from 2007 to 2008 were used and for Model Validation, three-month data were processed for use.

In establishing a Forecasting Model, Calendar variables (month, day of the week, season, quarter, holiday, and Chusuk) and Weather variables (average temperature, minimal

temperature, maximal temperature, temperature gap, rain(Y,N), snow(Y,N), air velocity, relative humidity, yellow dust) were used as variables.

In the two-year database that were used for establishing a forecast model, the whole number of patients admitted to the Emergency Department was 169,375 and average number of daily patients was 232. Also, as for the transition of daily patients during the same period, they showed 7-day periodicity and seasonal trend. In particular, there was a sharp increase in the number of patients in Chusuk. However, as for New Year, there was relatively weak increase and hence New Year was excluded as a selective variable in this study.

Since average and variance was stable, difference was not required, but primary seasonal difference was used to maintain the stability of data caused by the seasonal transition.

As for the time series models to forecast the number of patients admitted in an Emergency Department in a day, 1) MA(Moving Average) model; 2) Univariate-Seasonal ARIMA(Seasonal Auto- Regressive Integrated Moving Average) model; and 3) Multivariate-Seasonal ARIMA(Seasonal Auto- Regressive Integrated Moving Average) model were set up.

In order to test the goodness of fit, 1) residual analysis was done; 2) AIC(Akaike Information Criterion) and BIC(Bayesian Information Criterion) were compared ; and forecast accuracy was assessed for each model through MAPE(Mean Absolute Percentage Error).

As a result to compare existing three forecast models, it was found that multivariate seasonal ARIMA model was the most suitable for forecasting the number of patients who are

admitted to an Emergency Department in a day(AIC : 6703.7, BIC : 6749.5). Since its MAPE(Mean absolute Percentage Error) was 7.4%, which was the smallest among the models, it was selected as the final Forecasting Model that is fit for the purpose of this study. Chusuk, season, average temperature, and rain were selected as explanatory variables. —

Unlike MA model and univariate seasonal ARIMA model, this study applied explanatory variables in multivariate seasonal ARIMA model. In that sense, it has high reliability and accuracy on forecasting the number of patients admitted in Emergency Department in a day.

Key words : Forecasting Model, Time Series Model, Auto- Regressive Integrated Moving Average(ARIMA), Emergency Department Overcrowding.