

시계열 분석을 활용한 적혈구제제 수급 예측

목차

I.서론

- 1) 자료에 관한 배경 및 분석 목적

II.본론

- 1) 데이터 소개
- 2) 데이터 변환
- 3) 백색잡음검정 방법 소개

III.본론

- 1) 분해법
- 2) Holt Winters 가법계절지수평활법
- 3) 계절형 ARIMA

IV.본론

- 1) 모형 선택 및 해석

V.결론

VI.참고문헌

VII.부록

I.서론

1) 자료에 관한 배경 및 분석 목적

현재 전 세계적으로 코로나 팬데믹 영향으로 혈액 부족 상황이 더욱 심각해졌으며 우리나라 또한 그러한 상황이 지속되고 있다. 헌혈은 수혈이 필요한 환자에게 도움

이 될 수 있는 유일한 수단이며 혈액은 현재 인공적으로 만들거나 대체할 물질이 존재하지 않아 중요성이 매우 크다. 외상환자에서 발생하는 쇼크의 가장 흔한 원인은 저혈량 쇼크이고, 이는 외상 직후 조기 사망의 주된 원인으로 보고된다. 각종 혈액제제의 수혈은 이러한 상황을 대처할 수 있는 매우 중요한 수단이다. 이에 월별 헌혈 건수에서 적혈구제제공급량을, 월별 중증 외상 환자 수에서 적혈구제제수요량을 도출하여 공급량과 수요량의 비율을 예측하고 모형을 해석하여 특정 기간이나 시점에서 비율의 개형을 파악해 미리 혈액 부족 현상을 대비할 수 있을 것으로 기대한다. 또한 헌혈 횟수와 중증 환자 수의 변화를 분석하여 헌혈 횟수가 줄어 들고 있으며 이에 대한 심각성을 통계적으로 유의하게 보여주거나 어떤 시점에 중증 환자 수가 증가하는지 등 해석 및 추가분석에 활용할 수 있다. 이를 통해 혈액의 수급을 늘리거나 효율적으로 수혈을 최소화할 수 있는 대책의 필요성을 강조하는 것을 분석의 목적으로 한다.

사용되는 시계열 자료는 월별 적혈구제제수요량이다. 시각적으로도 추세성분과 계절성분이 나타나며 여러 가지 분석 방법의 비교를 보여주기 적합하여 자세하게 소개하고자 한다. 월별 적혈구제제공급량의 경우 분석과정은 생략하며 월별 적혈구제제수요량과 같은 분석 방법을 진행했음을 알린다.

여러 가지 분석 방법 중 분해법과 Holt Winters 계절지수평활법, 더하여 계절형 ARIMA를 사용하여 비교해보고자 한다. 각 방법은 장단점이 존재하며 이는 본문에서 소개한다. 각 분석 방법은 차분에 의해 데이터 포인트가 달라 RMSE로 비교하며 계절형 ARIMA에서 차수를 정할 때, AIC는 관측치의 수가 적을 때 너무 많은 예측 변수를 선택하는 경향이 있으므로 이를 보완한 AICc¹⁾를 사용하여 비교한다.

II.본론

1) 데이터 소개

첫 번째 데이터는 대한적십자사, 「혈액정보통계」의 2016년 1월부터 2020년 12월까지의 월별 헌혈통계자료다. <부록1>을 살펴보면 헌혈 횟수는 점차 감소하고 있음을 볼 수 있다. 또한 뚜렷하지는 않으나 계절성을 가지는 것을 보아 시계열 분석에 적합할 것으로 판단한다.

두 번째 데이터는 국립중앙의료원의 「응급의료현황통계」의 2016년 1월부터 2020년 12월까지의 중증외상환자 수의 자료다. <부록1>을 살펴보면 월별 헌혈통계와는 다르게 증가하는 추세를 보인다. 이는 혈액의 수요량에 대한 공급량의 비율이 줄어 들고 있다고 볼 수 있다. 또한 뚜렷하게 계절성이 보이며 2020년에 급격히 줄어드는 것을 고려하여 시계열 분석에 적합할 것으로 판단한다.

1) $AICc = AIC + \frac{2(k+2)(k+3)}{T-k-3}$, T : 관측치 수, k : 모수의 수

2) 데이터 변환

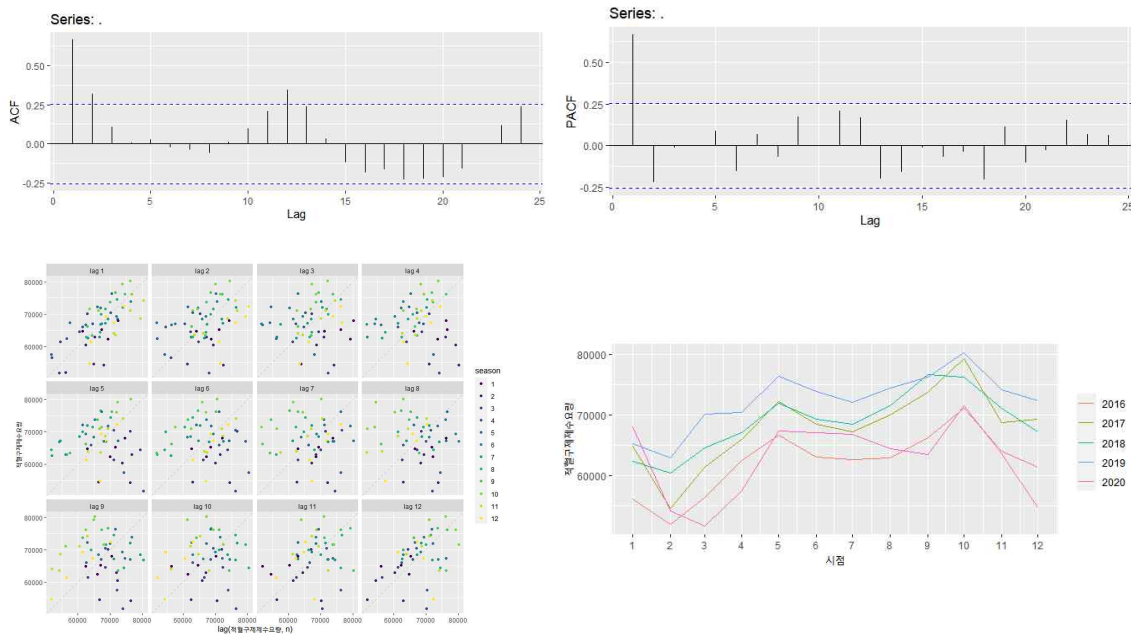
먼저 n 년도의 적혈구제제공급량을 도출하기 위해 $n-1$ 년도의 헌혈실적 중 사용된 적혈구제제의 비율을 n 년도의 헌혈 건수에 곱하여 도출했다. n 년도의 적혈구제제수요량은 이경원(2012)을 참고하여 중증 외상 환자 수 합계에 환자당 평균 수혈량을 곱하여 도출했다. 이를 나타낸 그래프는 <부록2>와 같다

3) 백색잡음검정 방법 소개

모형을 적합한 후 백색잡음임을 확인하기 위해 네 가지의 절차를 시행한다. 등분산의 가정을 만족하는지 확인하기 위해 Modified Levene's test를 사용한다. 등평균임을 확인하고 평균이 0인지를 확인하기 위해 Anova test와 t-test를 사용한다. 마지막으로 모형이 적합한지 확인하고 정상성을 따르는지 확인하기 위해 Portmanteau test를 수행한다. 이때 Ljung-Box Q 통계량을 사용한다.

Ⅲ.본론

아래는 적혈구제제수요량의 ACF, PACF, lag plot, 연도별 그래프를 나타낸다



ACF의 경우 시차 1, 12에서 유의하게 나타난 후 지수적 감소를 보여주며 PACF에서는 시차 1에서 유의하게 나타난다. 이처럼 주기적인 개형을 가질 수 있음을 보여주며 lag plot과 연도별 그래프에서도 계절 성분이 존재할 수 있음을 암시한다.

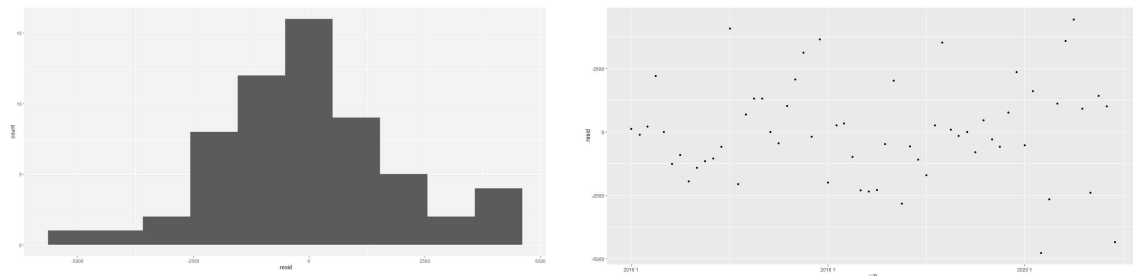
1) 분해법

분해법은 시계열을 분석하는 전통적인 방법으로 시계열이 체계적성분과 불규칙성분으로 구성되어 있으며 체계적 성분은 대표적으로 추세 성분과 계절 성분으로 분해할 수 있다고 보는 것이다. 이 방법은 이론적으로는 미흡하지만, 직관적이며 이해하기 쉽다는 장점이 있다. 보통 각 성분이 서로 독립이라는 가정하에 순서대로 분해하나 이러한 가정을 취할 수 없어 한 번에 진행했다. 계절 성분을 나타내기 위해 지시함수를 사용했고 제약 조건으로 1월에 해당하는 계수를 0으로 두었다. 추세 성분에서는 1차, 2차, 3차 항을 순서대로 추가하며 비교하였다. 또한 2020년 1월 31일 부2)로 평균 수준이 변화하여 $t \geq 2020.02$ 일 때 '1' 값을 가지는 변수를 추가하였다. 시간에 따라 변동 폭이 변하는 모습은 보이지 않아 변환의 필요성이 없으며 가법 모형을 사용했다.

각 모형을 적합한 결과는 <부록3-1>, <부록3-2>, <부록3-3>과 같다. Adjusted R-squared가 2차 항까지 추가했을 때 가장 높았으며 계절 성분을 제외한 나머지 계수가 유의하므로 이를 택했다. 적합된 모형은 아래와 같다.

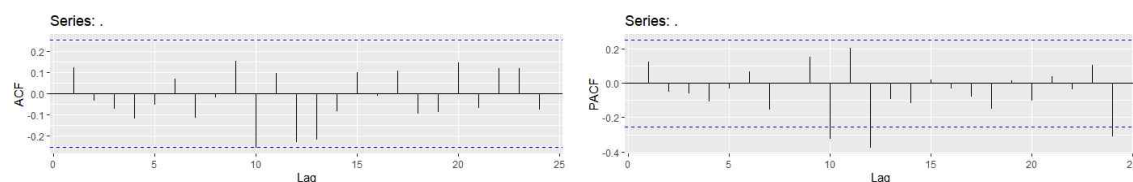
$$\begin{aligned} \widehat{\text{적혈구제제수요량}} = & 55591.672 - 11549.582I(t \geq 2020.02) + 438.365t - 3.567t^2 - 4457.234I(t = 2\text{월}) \\ & - 652.734I(t = 3\text{월}) + 2936.029I(t = 4\text{월}) + 8901.431I(t = 5\text{월}) + 6143.613I(t = 6\text{월}) \\ & + 5034.34I(t = 7\text{월}) + 6045.706I(t = 8\text{월}) + 8430.041I(t = 9\text{월}) + 12630.826I(t = 10\text{월}) \\ & + 5077.804I(t = 11\text{월}) + 1592.527I(t = 12\text{월}) + \hat{\epsilon}_t \end{aligned}$$

1-1) 잔차 확인



잔차의 히스토그램과 분포를 보면 정규성을 만족한다고 가정할 수 있으며 무작위성 또한 가진다고 볼 수 있다.

1-2) 분해법 적합 후 ACF, PACF



2) 세계보건기구(WHO)는 2020년 1월 31일, 국제적 공중보건 비상사태를 선포

ACF의 경우 전부 범위 안에 있으며 PACF의 경우 시차 10, 12에서 살짝 벗어났지만 유의한 특징이 보이지 않아 백색잡음인지에 대한 검정을 진행한다.

1-3) 백색잡음검정(등분산, 등평균, 잔차의 평균이 0인지 검정, 포트멘토 검정)

잔차의 등분산 검정결과는 <부록4-1>과 같고 p-value = 0.7778로 유의 수준 0.05 하 잔차는 등분산이라는 귀무가설을 기각할 충분한 근거가 없다. 잔차의 등평균 검정결과는 <부록4-2>와 같고 p-value = 0.664로 유의 수준 0.05 하 잔차가 등평균이라는 귀무가설을 기각할 충분한 근거가 없다. 또한 잔차의 평균이 0인지에 대한 검정 또한 <부록4-3>과 같고 p-value = 0.664로 유의 수준 0.05 하 잔차의 평균이 0이라는 귀무가설을 기각할 충분한 근거가 없다. 시차에 따른 포트멘토 검정은 부록<4-4>와 같고 유의수준 0.05하 p-value가 전부 유의하지 않아 모형은 적합하며 등평균, 등분산, 잔차의 평균은 0이라고 할 수 있으며 백색잡음이라고 할 수 있다.

<부록5>은 적합 후 예측값과 실제값을 함께 나타낸 그래프이다. 코로나 이후에 살짝 벗어난 부분이 있으나 전체적으로 잘 적합 된 것으로 보인다.

2) Holt Winters 가법계절지수평활법

시간이 경과 함에 따라 시계열이 생성되는 시스템 자체에 변화가 올 수 있다. 이 같은 경우 모든 자료를 동일하게 취급하여 구한 예측값보다는 시스템이 변화한 시점에 가까운 최근의 자료에 더 큰 비중을 주는 예측법을 사용하는 것이 합리적일 것이다. 이러한 방법에 지수평활법이 사용되고, 더하여 Winters의 가법계절지수평활법은 시계열이 추세 성분과 계절 성분 및 불규칙성분들의 합으로 구성되어 있다고 보고 각 성분을 평활법에 의해 추정 후 이를 이용하여 예측값을 구하는 방법이다.

시점 $n+l$ 에서 예측되는 값은 아래와 같이 표현된다.

$$Z_{n+l} = T_{n+l} + S_{n+l} + I_{n+l}$$

순서대로 추세 성분, 계절 성분, 불규칙성분에 해당한다. 이때 추세 성분은 선형추세를 가정하면 아래와 같은 관계를 갖는다.

$$T_{n+l} = T_n + \beta_{1,n}l$$

또한 s 개의 가법계절성분들은 아래와 같은 관계를 갖는다고 가정한다.

$$S_i = S_{i+s} = S_{i+2s} = \dots \quad (i = 1, 2, \dots, s), \quad \sum_{i=0}^s S_i = 0$$

이때 미래값 Z_{n+l} 의 예측값은 아래와 같다.

$$\begin{aligned}\hat{Z}_n(l) &= \hat{T}_n + \hat{\beta}_{1,n}l + \hat{S}_{n+l-s}, l=1,2,\dots,s \\ \hat{Z}_n(l) &= \hat{T}_n + \hat{\beta}_{1,n}l + \hat{S}_{n+l-2s}, l=s+1,s+2,\dots,2s \\ &\vdots\end{aligned}$$

갱신되는 방법은 아래와 같이 진행된다.

$$\begin{aligned}\hat{T}_{n+1} &= \omega_1(Z_{n+1} - \hat{S}_{n+1-s}) + (1-\omega_1)(\hat{T}_n + \hat{\beta}_{1,n}) \\ \hat{\beta}_{1,n+1} &= \omega_2(\hat{T}_{n+1} - \hat{T}_n) + (1-\omega_2)\hat{\beta}_{1,n} \\ \hat{S}_{1,n+1} &= \omega_3(Z_{n+1} - \hat{T}_{n+1}) + (1-\omega_3)\hat{S}_{n+1-s}\end{aligned}$$

$\omega_1, \omega_2, \omega_3$ 는 서로 다른 평활상수들이다. 갱신 식을 사용하려면 초기 평활 값이 필요한데 이들을 구하는 방법은 먼저 아래와 같은 회귀모형을 적합시키고 각 최소제곱법으로 초기 평활 값을 구한다.

$$Z_t = \beta_0 + \beta_{1t} + \sum_{i=0}^{s-1} \delta_i (IND_{ti} - IND_{ts}) + \epsilon_t$$

$$\hat{T}_0 = \hat{\beta}_0$$

$$\hat{\beta}_{1,0} = \hat{\beta}_1$$

$$\hat{S}_0 = -\sum_{j=0}^{s-1} \hat{\delta}_j$$

위의 과정을 거쳐 가지는 평활상수와 초기 평활 값은 다음과 같다.

$$\omega_1 = 0.7936727$$

$$\omega_2 = 0.02390908$$

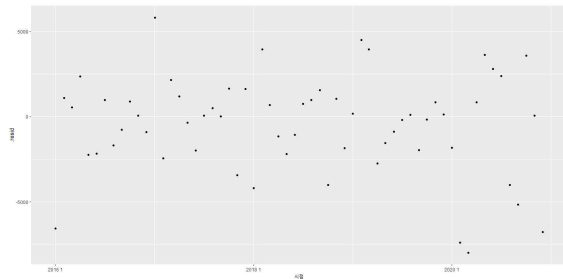
$$\omega_3 = 0.0001000013$$

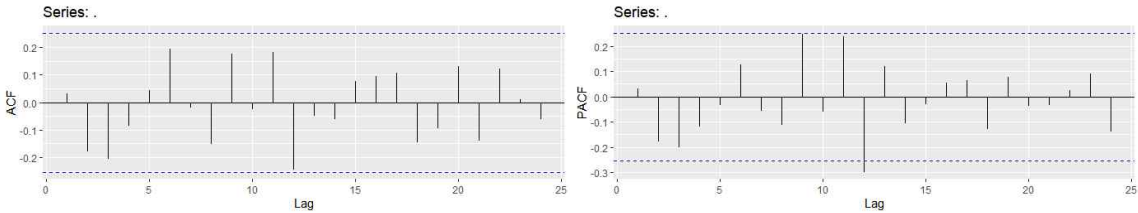
$$\hat{T}_0 = 65453.84$$

$$\hat{\beta}_{1,0} = 400.7694$$

$$\hat{S}_0 = -385.6391$$

2-1) 잔차 체크





ACF의 경우 전부 범위 안에 있으며 PACF의 경우 시차 12에서 미세하게 벗어났지만 유의한 특징이 보이지 않아 백색잡음인지에 대한 검정을 진행한다.

2-3) 백색잡음검정(등분산, 등평균, 잔차의 평균이 0인지 검정, 포트멘토 검정)

잔차의 등분산 검정결과는 <부록6-1>과 같고 p-value = 0.3971로 유의 수준 0.05 하 잔차는 등분산이라는 귀무가설을 기각할 충분한 근거가 없다. 잔차의 등평균 검정결과는 <부록6-2>와 같고 p-value = 0.544로 유의 수준 0.05 하 잔차가 등평균이라는 귀무가설을 기각할 충분한 근거가 없다. 또한 잔차의 평균이 0인지에 대한 검정 또한 <부록6-3>과 같고 p-value = 0.2275로 유의 수준 0.05 하 잔차의 평균이 0이라는 귀무가설을 기각할 충분한 근거가 없다. 시차에 따른 포트멘토 검정은 <부록6-4>와 같고 유의수준 0.05하 p-value가 전부 유의하지 않아 모형은 적합하며 잔차는 등평균, 등분산, 평균은 0이라고 할 수 있으며 백색잡음이라고 할 수 있다.

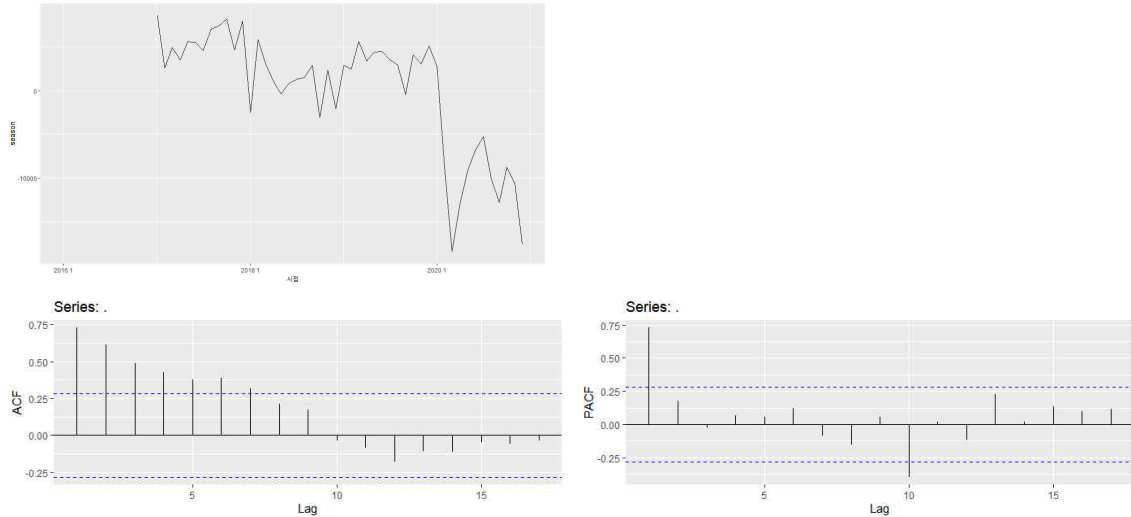
<부록7>은 적합 후 예측값과 실제값을 함께 나타낸 그래프이다. 코로나 이후부터 적합이 잘 안되는 모습이 보인다. 이는 코로나 전후로 양상이 변화하였으며 계절지 수평활법만을 사용해서는 이를 고려하지 못해 생긴 결과로 판단된다.

3) 계절형 ARIMA

우리가 접하는 시계열들은 그 구성 성분들이 확률적이거나 다른 성분들과 상관성이 있는 경우가 많다. 이 같은 경우에는 확률적인 분석 방법인 ARIMA모형을 이용한 분석이 활용될 수 있다. 이때 계절형을 가진다고 판단할 수 있을 때 계절형 ARIMA모형을 사용할 수 있다. 하지만 차분이 필요할 경우 데이터 소실의 단점이 존재한다. 원 시계열 자료는 추세 성분과 계절 성분이 존재한다고 볼 수 있다. 이 성분들은 비정상 시계열 요소이므로 차분을 통해 제거할 수 있다. 이때 1차 차분은 추세 성분을, 계절 차분은 계절 성분을 제거하는 데에 사용될 수 있다. 하지만 계절 차분의 경우 $(1 - B^{12})$ 항에 1차 차분의 $(1 - B)$ 항이 포함되어 있어 계절 차분으로도 비정

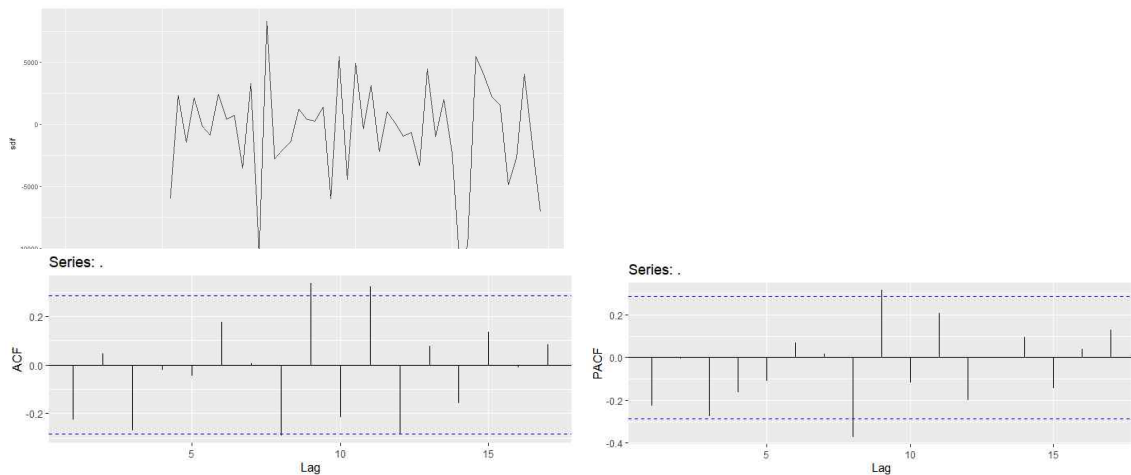
상성 문제가 해결될 수 있다. 따라서 간결성의 원칙에 따라 계절 차분을 먼저 진행 후 정상성을 만족할 수 없다면 이후 추가로 1차 차분을 진행한다.

3-1) 계절 차분 후 그래프 및 ACF, PACF



계절 차분 후 그래프에서 시간에 따라 감소하는 추세가 보여 확률적 추세를 갖는다고 판단할 수 있고, ACF 그래프에서도 천천히 0으로 수렴하며 PACF에서도 시차 1에서 유의하게 나타나 추가 적으로 1차 차분이 필요하다.

3-2) 계절 차분 + 1차 차분 그래프 및 ACF, PACF



그래프 해석에 앞서 계절 차분 후 분산(5110.043)과 계절 차분 + 1차 차분 후의 분산(6882.162)을 비교했을 때 후자의 경우가 더 낮아 차분을 하는 것에 문제가 없다고 판단할 수 있다.

차분 후 그래프의 경우 추세나 계절성이 보이지 않아 정상성을 가지는 것처럼 보인다. ACF의 경우 시차 9, 11, 12에서 미세하게 벗어난 모습을 보이고 PACF의 경우 시차 8, 9에서 조금 벗어난다. 큰 특징은 보이지 않아 백색잡음검정을 진행했다.

3-3) 차분 후 백색잡음검정

(등분산, 등평균, 잔차의 평균이 0인지 검정, 포트멘토 검정)

잔차의 등분산 검정 결과는 <부록8-1>과 같고 p-value = 0.2565로 유의 수준 0.05 하 잔차는 등분산이라는 귀무가설을 기각할 충분한 근거가 없다. 잔차의 등평균 검정결과는 <부록8-2>와 같고 p-value = 0.232로 유의 수준 0.05 하 잔차가 등평균이라는 귀무가설을 기각할 충분한 근거가 없다. 또한 잔차의 평균이 0인지에 대한 검정 또한 <부록8-3>과 같고 p-value = 0.3661로 유의 수준 0.05 하 잔차의 평균이 0이라는 귀무가설을 기각할 충분한 근거가 없다. 등평균, 등분산, 잔차의 평균은 0이라고 할 수 있으나 시차에 따른 포트멘토 검정은 <부록8-4>와 같고 유의 수준 0.05하 시차는 11에서부터 p-value가 유의하게 나타나 모형이 적합하다는 충분한 근거가 없으며 정상성을 따르지 않아 잔차가 백색잡음이라고 할 수 없다.

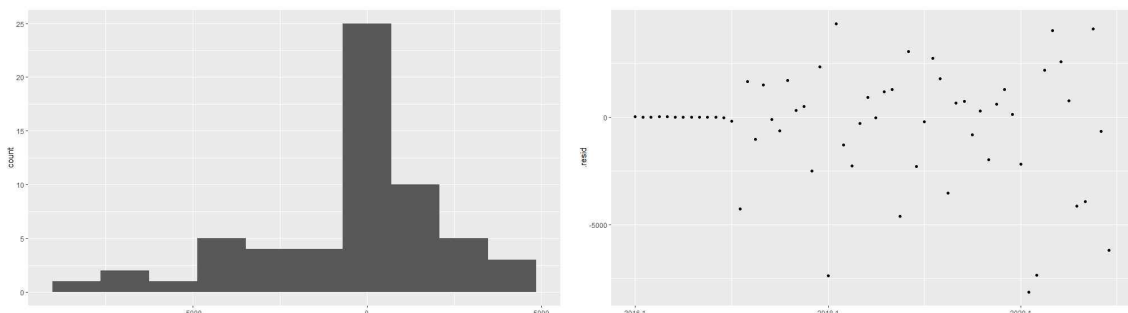
3-4) 계절형 ARIMA 모형 적합

차분 만을 이용해서는 잔차가 백색잡음을 따르지 않아 여러 가지 후보군을 가지고 계절형 ARIMA를 적합하였다. 간결성의 원칙에 따라 시차 3 이상은 고려하지 않고 범위는 넘지 않았으나 범위 선에 가까운 시차 1과 3을 고려하였다. 또한 평균이 0인지에 대한 검정 결과를 바탕으로 상수항이 없는 계절형 ARIMA 모형을 적합하였다.

후보군에 따른 AICc 값은 <부록9>과 같다. AICc가 가장 낮은 모형인 SARIMA(0,1,0)(0,1,1)₁₂을 택했고 적합된 모형식은 아래와 같다.

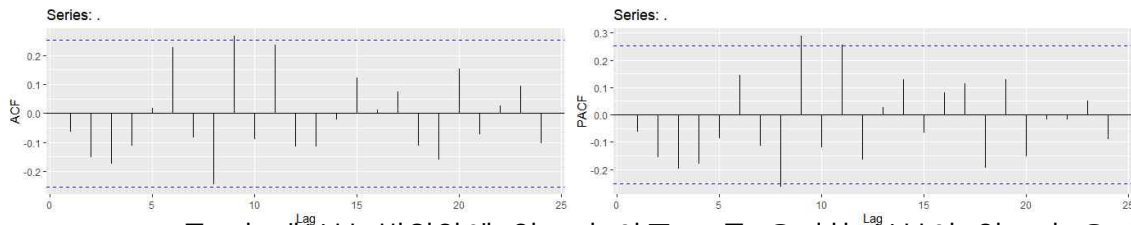
$$(1 - B)(1 - B^{12})\ln Z_t = (1 + 0.9991B^{12})\varepsilon_t$$

3-5) 잔차 확인



잔차의 히스토그램과 분포를 보면 정규성을 만족한다고 가정하기 애매하나 0에 몰려있어 이는 감안하고 진행하며 무작위성은 가진다고 볼 수 있다.

3-6) 계절형 ARIMA 적합 후 ACF, PACF



ACF, PACF 둘 다 대부분 범위안에 있으며 아주 조금 유의한 부분이 있으나 유의한 특징이 보이지 않아 백색잡음인지에 대한 검정을 진행한다.

3-7) 계절형 ARIMA 적합 후 백색잡음검정

(등분산, 등평균, 잔차의 평균이 0인지 검정, 포트멘토 검정)

잔차의 등분산 검정결과는 <부록10-1>과 같고 $p\text{-value} = 0.9546$ 로 유의 수준 0.05 하 잔차는 등분산이라는 귀무가설을 기각할 충분한 근거가 없다. 잔차의 등평균 검정결과는 <부록10-2>와 같고 $p\text{-value} = 0.889$ 로 유의 수준 0.05 하 잔차가 등평균이라는 귀무가설을 기각할 충분한 근거가 없다. 또한 잔차의 평균이 0인지에 대한 검정 또한 <부록10-3>과 같고 $p\text{-value} = 0.2313$ 로 유의 수준 0.05 하 잔차의 평균이 0이라는 귀무가설을 기각할 충분한 근거가 없다. 시차에 따른 포트멘토 검정은 <부록10-4>와 같고 유의수준 0.05하 $p\text{-value}$ 가 전부 유의하지 않아 모형은 적합하며 잔차는 등평균, 등분산, 평균은 0이라고 할 수 있으며 백색잡음이라고 할 수 있다.

<부록11>은 적합 후 예측값과 실제값을 함께 나타낸 그래프이다. 계절지수평활법과 같이 코로나 이후에 적합에 어려움이 있음을 확인할 수 있다. 이또한 코로나 전후로 양상이 변화하였으며 계절형 ARIMA만을 사용해서는 이를 고려하지 못해 생긴 결과로 판단된다.

IV.본론

1) 모형 선택 및 해석

분해법을 이용한 경우 RMSE는 1928.095, Holt Winters 가법계절지수평활법을 활용할 때 RMSE는 2887.734, 계절형 ARIMA를 이용한 경우 RMSE는 2681.106로 나타났다. 또한 예측값과 실제값을 비교한 그래프에서도 분해법을 이용한 경우가 눈에 보이게 잘 적합 되었음을 볼 수 있다. 이러한 까닭은 평균 수준을 코로나의 영향에 따라 고려할 수 있어서 나타난 결과로 볼 수 있다. 또한 시각적으로도 뚜렷하게 추세 성분과 계절 성분이 존재함을 볼 수 있어 분해법이 적절한 분석 방법으로 가장 좋은 결과를 도출했다고 판단할 수 있다. 또한 분해법을 사용했기 때문에 추세 성분과 계절 성분을 나누어 직관적인 해석이 가능하다. 물론 제약 조건과 상수항에 의해 조금 복잡하게 해석될 수 있으나 간단히, 연말부터 시작해서 연초에 낮은 수요량을 가지고 봄과 가을에 수요량이 늘어나는 것으로 해석할 수 있다.

2) 적혈구제제공급량

적혈구제제수요량과 같은 방법으로 진행했을 때 최종모형은 <부록12>과 같았다. <부록13>은 예측된 '적혈구제제공급량 - 적혈구제제수요량' 과 실제값을 비교한 그래프이다. 크게 변하는 구간을 제외하고서는 잘 적합 되었음을 볼 수 있다.

V.결론

계절적인 요소가 나타나는 시계열 자료를 분해법, Holt Winters 가법계절지수평활법, 계절형 ARIMA를 이용하여 여러 가지 분석 방법으로 비교해보았다. 분해법의 경우 전통적인 방법으로 이론적으로 미흡하나 결과적으로 가장 좋은 결과를 도출해냈다. 분해법의 경우 직관적인 이해와 해석이 가능하여 이를 세심히 해석해본다면 더 다채로운 분석 결과를 도출할 수 있을 것이다. 계절지수평활법과 계절형 ARIMA의 경우 코로나의 영향을 고려할 수 없었고 이 부분이 크게 작용한 것으로 보인다. 또한 ARIMA의 경우 차분이 필요해 데이터 소실의 단점이 존재하기도 한다. 이에 계절지수평활법과 계절형 ARIMA를 사용할 때 이 점을 추가로 고려하여 모형을 적합해보는 것을 제안한다.

종합하여 두 가지 방향을 제시할 수 있다. 첫 번째는 각 분석 방법에 장단점이 존재하는데 이를 정확히 비교하기 위해 코로나의 영향을 고려할 수 있는 방법, 즉 전이함수를 활용하는 방향으로 진행된다면 더 세심한 비교가 가능할 것이다. 두 번째는 적혈구제제공급량과 적혈구제제수요량으로 변환되는 과정에서 정확히 중증외상 환자에게 공급되는 비율과 필요한 비율을 계산할 수 있다면 좀 더 즉각적으로 헌혈이 필요한 때를 예측할 수 있다.

본 보고서에서는 생략했으나 적혈구제제수요량의 경우 시각적으로도 계절 성분이 보이지 않기 때문인지 계절 성분의 계수에서 유의하지 않은 것들이 많았다. 그럼에도 불구하고 계절지수평활법과 계절형 ARIMA보다는 낮은 RMSE값을 도출했었다. 또한 코로나의 영향을 사용하지 않을 때 더 좋은 결과가 나타났음을 첨언 하며 보고서를 마친다.

VI.참고문헌

이경원. (2012). 응급의료센터에서 수혈을 시행한 성인 외상환자에서 사망 예측 인자. 대한외상학회지, 25(3), 79-86.

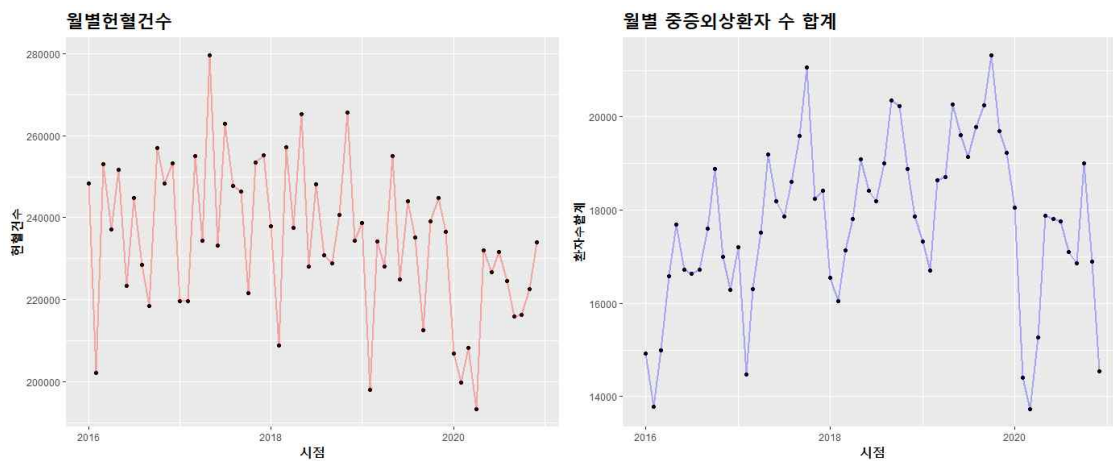
김승우, 이평연, 권상욱 and 김종훈. (2022). 시계열 모델 기반의 계절성에 특화된 S-ARIMA

모델을 사용한 리튬이온 배터리의 노화 예측 및 분석. 전력전자학회 논문지, 27(4), 316-324.

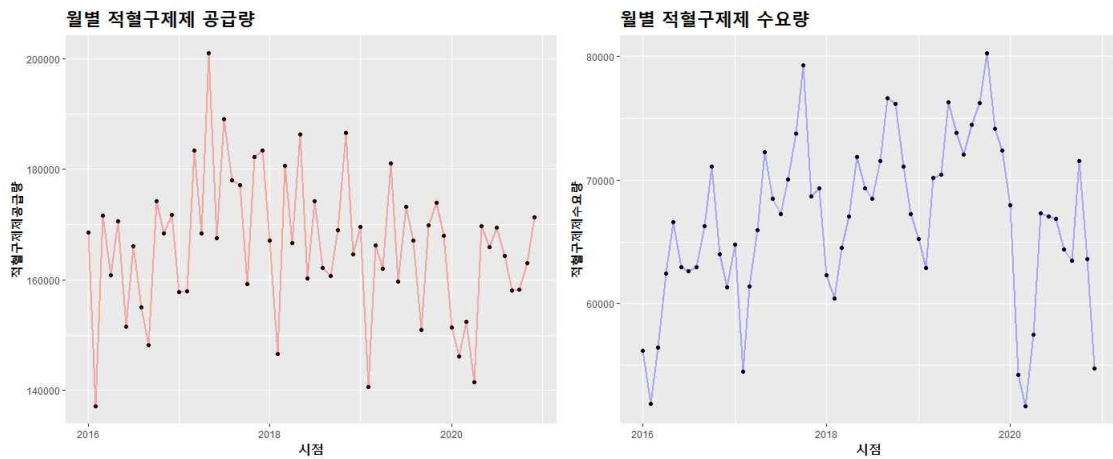
이정현, 송규원, 임윤정 and 조민상. (2022). 수돗물 수질 이상 탐지를 위한 수질 시계열 예측 모형 개발. 정보과학회 컴퓨팅의 실제 논문지, 28(9), 465-473.

VII.부록

<부록1> 월별 헌혈건수, 월별 중증외상환자 수 합계



<부록2> 월별 적혈구제제공급량, 월별 적혈구제제수요량



<부록3-1>아래 모형식 적합 결과

$$\begin{aligned} \widehat{\text{적혈구제제수요량}} = & \hat{\beta}_0 + \beta_1 I(t \geq 2020.02) + \hat{\beta}_2 t + \hat{\beta}_3 I(t = 2\text{월}) + \hat{\beta}_4 I(t = 3\text{월}) + \hat{\beta}_5 I(t = 4\text{월}) \\ & + \hat{\beta}_6 I(t = 5\text{월}) + \hat{\beta}_7 I(t = 6\text{월}) + \hat{\beta}_8 I(t = 7\text{월}) + \hat{\beta}_9 I(t = 8\text{월}) + \hat{\beta}_{10} I(t = 9\text{월}) \\ & + \hat{\beta}_{11} I(t = 10\text{월}) + \hat{\beta}_{12} I(t = 11\text{월}) + \hat{\beta}_{13} I(t = 12\text{월}) + \hat{\epsilon}_t \end{aligned}$$

	$\hat{\beta}$	S.E.	P-value
(Intercept)	56793.62	1195.43	2e-16 ***
$\hat{\beta}_1 \cdot I(t \geq 2020.2)$	-14117.54	1195.43	2e-16 ***
$\hat{\beta}_2$	260.03	23.92	2.69e-14 ***
$\hat{\beta}_3$	-3947.21	1476.95	0.010380 *
$\hat{\beta}_4$	-153.41	1475.20	0.917628
$\hat{\beta}_5$	3417.52	1473.84	0.024903 *
$\hat{\beta}_6$	9357.96	1472.87	8.53e-08 ***
$\hat{\beta}_7$	6568.04	1472.29	5.23e-05 ***
$\hat{\beta}_8$	5419.53	1472.09	0.000607 ***
$\hat{\beta}_9$	6384.53	1472.29	7.82e-05 ***
$\hat{\beta}_{10}$	8715.37	1472.87	3.85e-07 ***
$\hat{\beta}_{11}$	12855.52	1473.84	2.61e-11 ***
$\hat{\beta}_{12}$	5234.73	1475.20	0.000905 ***
$\hat{\beta}_{13}$	1674.56	1476.95	0.262754
R-squared	0.9063	Adjusted R-squared	0.8798

<부록3-2>아래 모형식 적합 결과

$$\begin{aligned}
\widehat{\text{적혈구제제수요량}} = & \hat{\beta}_0 + \hat{\beta}_1 I(t \geq 2020.02) + \hat{\beta}_{21}t + \hat{\beta}_{22}t^2 + \hat{\beta}_3 I(t = 2\text{월}) + \hat{\beta}_4 I(t = 3\text{월}) + \hat{\beta}_5 I(t = 4\text{월}) \\
& + \hat{\beta}_6 I(t = 5\text{월}) + \hat{\beta}_7 I(t = 6\text{월}) + \hat{\beta}_8 I(t = 7\text{월}) + \hat{\beta}_9 I(t = 8\text{월}) + \hat{\beta}_{10} I(t = 9\text{월}) \\
& + \hat{\beta}_{11} I(t = 10\text{월}) + \hat{\beta}_{12} I(t = 11\text{월}) + \hat{\beta}_{13} I(t = 12\text{월}) + \hat{\epsilon}_t
\end{aligned}$$

	$\hat{\beta}$	S.E.	P-value
(Intercept)	55591.672	1276.278	2e-16
$\hat{\beta}_1 \cdot I(t \geq 2020.2)$	-11549.582	1561.795	2.65e-09
$\hat{\beta}_{21}$	438.365	85.310	5.79e-06
$\hat{\beta}_{22}$	-3.567	1.643	0.035257
$\hat{\beta}_3$	-4457.234	1440.021	0.003378
$\hat{\beta}_4$	-652.734	1437.569	0.651972
$\hat{\beta}_5$	2936.029	1434.986	0.046623
$\hat{\beta}_6$	8901.431	1432.329	1.50e-07
$\hat{\beta}_7$	6143.613	1429.678	9.14e-05
$\hat{\beta}_8$	5034.340	1427.133	0.000979
$\hat{\beta}_9$	6045.706	1424.821	0.000109
$\hat{\beta}_{10}$	8430.041	1422.891	4.04e-07

$\hat{\beta}_{11}$	12630.826	1421.514	1.84e-11
$\hat{\beta}_{12}$	5077.804	1420.888	0.000854
$\hat{\beta}_{13}$	1592.527	1421.229	0.268433
R-squared	0.9151	Adjusted R-squared	0.8887

<부록3-3>아래 모형식 적합 결과

$$\begin{aligned} \widehat{\text{적혈구제제수요량}} = & \hat{\beta}_0 + \beta_1 I(t \geq 2020.02) + \hat{\beta}_{21}t + \hat{\beta}_{22}t^2 + \hat{\beta}_{23}t^3 + \hat{\beta}_3 I(t = 2\text{월}) + \hat{\beta}_4 I(t = 3\text{월}) + \hat{\beta}_5 I(t = 4\text{월}) \\ & + \hat{\beta}_6 I(t = 5\text{월}) + \hat{\beta}_7 I(t = 6\text{월}) + \hat{\beta}_8 I(t = 7\text{월}) + \hat{\beta}_9 I(t = 8\text{월}) + \hat{\beta}_{10} I(t = 9\text{월}) \\ & + \hat{\beta}_{11} I(t = 10\text{월}) + \hat{\beta}_{12} I(t = 11\text{월}) + \hat{\beta}_{13} I(t = 12\text{월}) + \hat{\epsilon}_t \end{aligned}$$

	$\hat{\beta}$	S.E.	P-value
(Intercept)	5.549e+04	1.543e+03	2e-16
$\hat{\beta}_1 \cdot I(t \geq 2020.2)$	-1.170e+04	1.995e+03	5.34e-07
$\hat{\beta}_{21}$	4.600e+02	1.936e+02	0.021933
$\hat{\beta}_{22}$	-4.522e+00	7.848e+00	0.567424
$\hat{\beta}_{23}$	1.177e-02	9.445e-02	0.901427
$\hat{\beta}_3$	-4.433e+03	1.469e+03	0.004229
$\hat{\beta}_4$	-6.342e+02	1.461e+03	0.666359
$\hat{\beta}_5$	2.949e+03	1.454e+03	0.048717
$\hat{\beta}_6$	8.908e+03	1.449e+03	2.05e-07
$\hat{\beta}_7$	6.144e+03	1.446e+03	0.000109
$\hat{\beta}_8$	5.029e+03	1.444e+03	0.001134
$\hat{\beta}_9$	6.033e+03	1.444e+03	0.000137
$\hat{\beta}_{10}$	8.411e+03	1.447e+03	6.36e-07
$\hat{\beta}_{11}$	1.260e+04	1.453e+03	4.44e-11
$\hat{\beta}_{12}$	5.043e+03	1.463e+03	0.001259
$\hat{\beta}_{13}$	1.550e+03	1.478e+03	0.300055
R-squared	0.9152	Adjusted R-squared	0.8863

<부록4-1>등분산 검정결과

t	df	P-value
0.28351	58	0.7778

<부록4-2>등평균 검정결과

	df	F value	P-value
group	1	0.191	0.664
Residuals	58		

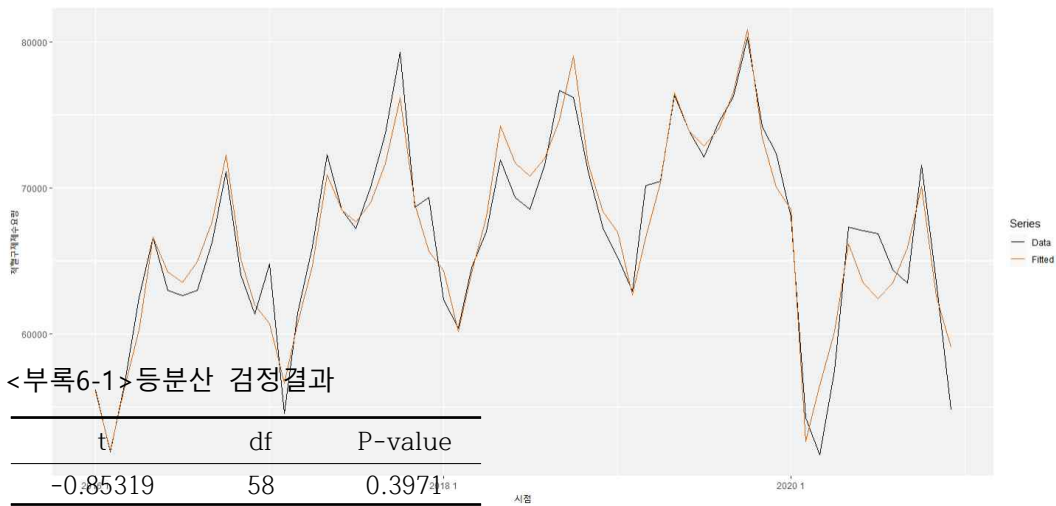
<부록4-3>잔차의 평균이 0인지 검정결과

t	df	P-value
-1.3963e-16	59	1

<부록4-4> Ljung-box test

lag	P-value
1	0.32426
2	0.59335
3	0.71012
4	0.67392
5	0.76994
6	0.82643
7	0.80153
8	0.87153
9	0.78099
10	0.39756
11	0.42749
12	0.22544
13	0.12012
14	0.14128
15	0.15465
16	0.19928
17	0.20605
18	0.21865
19	0.23653
20	0.20305
21	0.23084
22	0.22179
23	0.21081

<부록5>분해법 적합 후 예측값과 실제값 비교 그래프



<부록6-2> 등평균 검정결과

	df	F value	P-value
group	1	0.372	0.544
Residuals	58		

<부록6-3> 잔차의 평균이 0인지 검정 결과

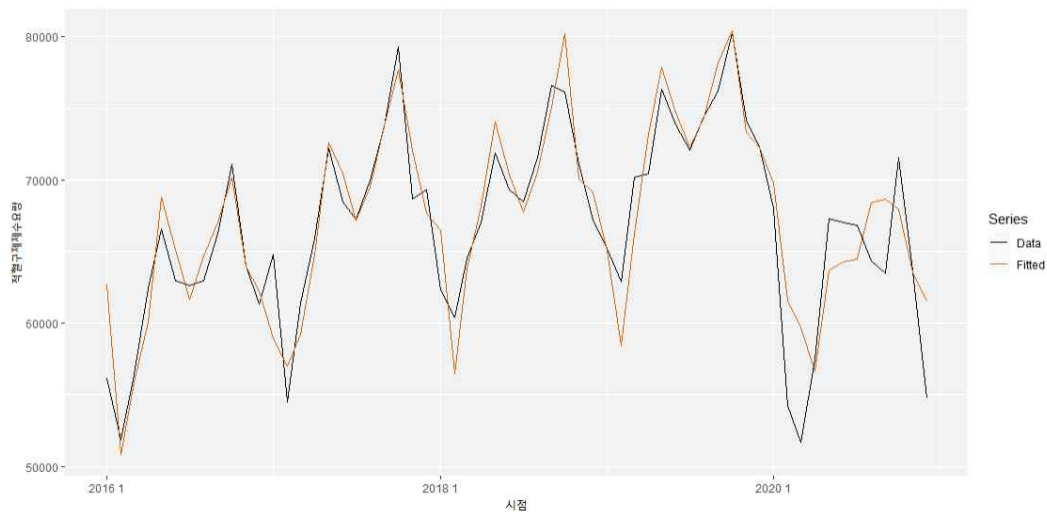
t	df	P-value
-1.2194	59	0.2275

<부록6-4> Ljung-box test

lag	P-value
1	0.80655
2	0.35223
3	0.18147
4	0.25271
5	0.36028
6	0.23439
7	0.32596
8	0.28693
9	0.21615
10	0.28523
11	0.20672
12	0.08394
13	0.11169
14	0.14015
15	0.16445
16	0.18091

17	0.18751
18	0.16071
19	0.17252
20	0.15756
21	0.13627
22	0.12767
23	0.15996

<부록7>Holt Winters 가법계절지수평활법 적합 후 예측값과 실제값 비교 그래프



<부록8-1>등분산 검정결과

t	df	P-value
-1.1497	44	0.2565

<부록8-2>등평균 검정결과

	df	Sum Sq	Mean Sq	F value	P-value
grp	1	24691326	24691326	1.466	0.232
Residuals	44	741103405	16843259		

<부록8-3>잔차의 평균이 0인지 검정결과

t	df	P-value
-0.9129	46	0.3661

<부록8-4>L-jung box test

lag	P-value
1	0.10821
2	0.26109

3	0.08971
4	0.16341
5	0.24990
6	0.21366
7	0.30277
8	0.10097
9	0.01683
10	0.01037
11	0.00167
12	0.00044
13	0.00069
14	0.00067
15	0.00072
16	0.00122
17	0.00168
18	0.00259
19	0.00257
20	0.00208
21	0.00164
22	0.00116
23	0.00157

<부록9> 후보군에 따른 AICc

SARIMA(0,1,0)(0,1,1)	910.2109
SARIMA(0,1,1)(0,1,1)	912.1053
SARIMA(1,1,0)(0,1,1)	912.2262
SARIMA(0,1,0)(1,1,1)	912.3115
SARIMA(0,1,0)(1,1,0)	914.3007
SARIMA(0,1,1)(1,1,1)	914.3937
SARIMA(1,1,0)(1,1,1)	914.5132
SARIMA(0,1,3)(0,1,1)	914.5186
SARIMA(0,1,1)(1,1,0)	915.2430
SARIMA(1,1,0)(1,1,0)	915.7941
SARIMA(0,1,3)(1,1,1)	916.8754
SARIMA(0,1,3)(1,1,0)	917.0562
SARIMA(1,1,3)(0,1,1)	917.1541
SARIMA(0,1,1)(0,1,0)	918.9167
SARIMA(1,1,0)(0,1,0)	919.1494
SARIMA(0,1,0)(0,1,0)	919.3656
SARIMA(1,1,3)(1,1,1)	919.4496
SARIMA(1,1,3)(1,1,0)	919.5656

SARIMA(0,1,3)(1,1,0)	920.7457
SARIMA(1,1,3)(0,1,0)	922.1904

<부록10-1> 등분산 검정결과

t	df	P-value
0.057197	58	0.9546

<부록10-2> 등평균 검정결과

	df	F value	P-value
group	1	0.02	0.889
Residuals	58		

<부록10-3> 잔차의 평균이 0인지 검정결과

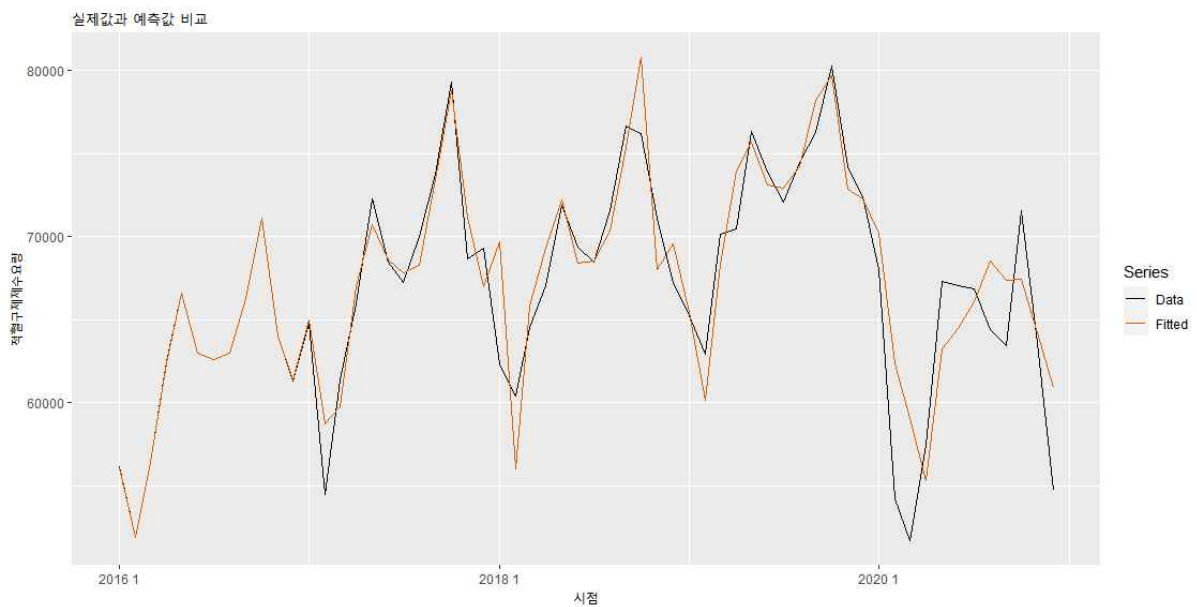
t	df	P-value
-1.2095	59	0.2313

<부록10-4> Ljung-box test

lag	P-value
1	0.62571
2	0.44515
3	0.33211
4	0.38190
5	0.52102
6	0.29016
7	0.35324
8	0.18306
9	0.07563
10	0.09759
11	0.05425
12	0.06360
13	0.07323
14	0.10128
15	0.10991
16	0.14552
17	0.17510
18	0.19084

19	0.17726
20	0.16744
21	0.19685
22	0.23902
23	0.26316

<부록11>계절형 ARIMA 적합 후 예측값과 실제값 비교 그래프



<부록12>적혈구제제공급량 최종 모형식

$$\begin{aligned}
 \widehat{\text{적혈구제제공급량}} = & 177400 - 11790t + 1762t^2 - 103.2t^3 + 2.872t^4 - 0.03836t^5 + 0.0001978t^6 \\
 & + 1762I(t = 2\text{월}) - 103.2I(t = 3\text{월}) + 225I(t = 4\text{월}) \\
 & + 22530I(t = 5\text{월}) + 1975I(t = 6\text{월}) + 15430I(t = 7\text{월}) \\
 & + 6173I(t = 8\text{월}) - 398.4I(t = 9\text{월}) + 6242I(t = 10\text{월}) \\
 & + 14290I(t = 11\text{월}) + 10420I(t = 12\text{월}) + \hat{\epsilon}_t
 \end{aligned}$$

<부록13> 적혈구제제공급량 - 적혈구제제수요량 예측값과 실제값 비교 그래프

