SPSS Trends(예측방법)의 활용

조신섭[1]

국민 총생산고, 물가지수, 종합주가지수, 강우량, 태양의 흑점 등과 같이 연도별, 계절별, 월별, 일별로 시간의 흐름에 따라 순서대로 관측되는 자료를 시계열자료(time series data)라고 부른다. 시계열자료는 대부분의 통계분석기법에서 다루는 자료들과는 달리 시간에 따라 관측된 관계로 시간의 영향을 받아 독립가정이 성립하지 않으므로 이에 맞는 분석방법을 필요로 한다.

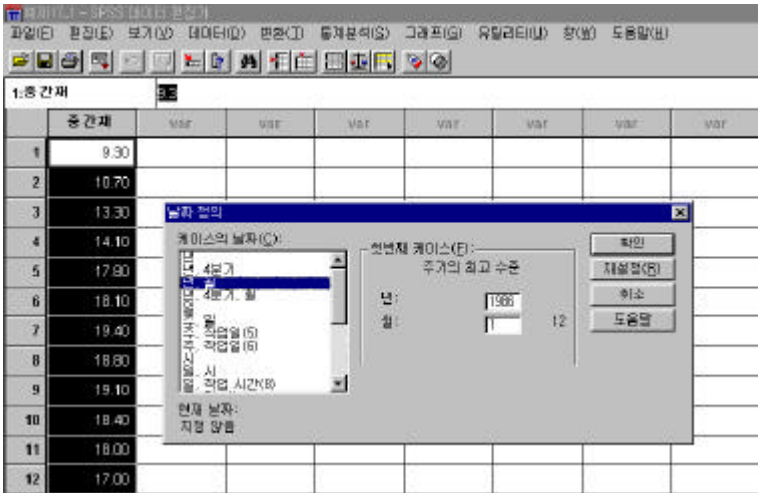
시계열자료의 분석을 위해 많이 사용되는 방법으로는 시간을 설명변수로 하는 추세분석(trend analysis), 평활법(smoothing method), 분해법(decomposition method) 및 Box-Jenkins에 의해 체계화된 ARIMA모형을 이용하는 방법들을 들 수 있다. SPSS에서는 통계분석의 시계열분석 절차를 이용하여 위에서 언급한 4가지 방법을 처리할 수 있도록 하고 있다.

시간변수의 생성 및 시계열그림 그리기

시계열분석의 특징은 시간에 따라 관측된 자료를 다루므로, 관측값과 함께 관측된 시점과 연관된 정보가 저장된다는 점이다. 관측시점에 대한 정보를 주지 않는 경우에는 관측순서에 따라 1부터 관측값의 개수에 해당되는 값들이 관측시점에 대한 값으로 사용된다. 다음의 예를 이용하여 시간변수를 생성하는 방법을 설명하기로 한다.

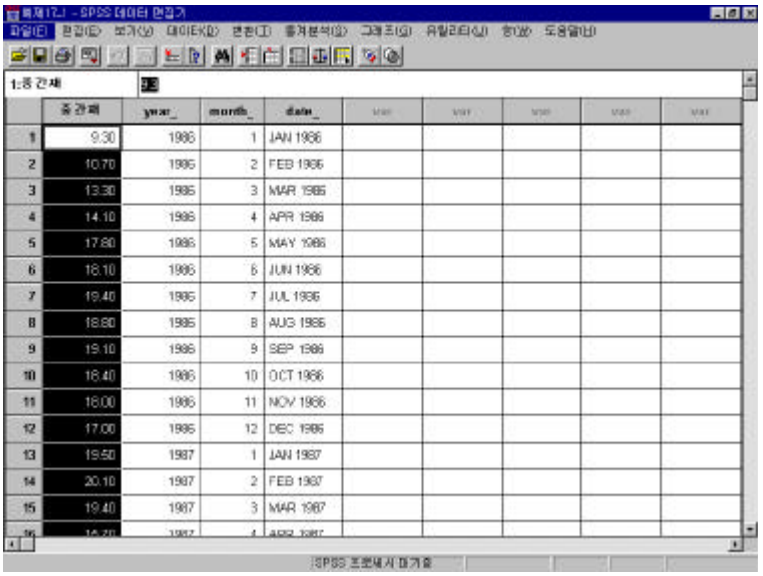
1) 서울대학교 계산통계학과 교수

예제 1 다음은 통계청에서 발표한 1986년 1월부터 1994년 4월까지의 중간재 출하지수 자료를 이용하여 시간변수를 생성하는 예이다. 데이터가 <그림 1.1>과 같이 첫 열에 입력되어 있는 상태에서 시간변수를 생성하여 보자.



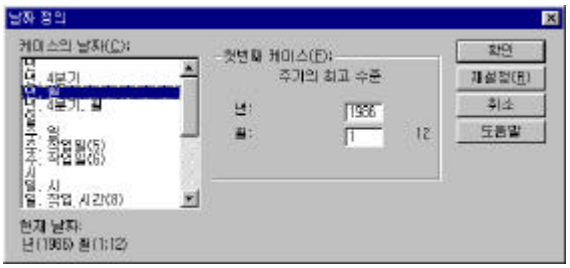
<그림 1.1> 날짜정의 대화상자를 이용한 시간변수의 생성

관측시점에 관한 정보를 저장한 시간변수를 생성하기 위해서는 데이터(D)-날짜정의(E) 절차를 이용한다. <그림 1.1>에 있는 날짜정의 대화상자에서 케이스의 날짜(C)에서 자료에 맞는 관측주기를 선택한다. 이 경우는 월별자료이므로 년, 월을 선택한다. 오른쪽의 첫번째 케이스(F)에서는 첫번째 관측시점인 1986년 1월을 입력한다. 확인 단추를 누르면 <그림 1.2>와 같이 year_, month_ 및 date_ 변수가 생성되어 각 열에 저장되어 있음을 볼 수 있다. 또한 <그림 1.3>과 같이 "현재 날짜"가 "년(1986) 월(1;12)"과 같이 변해 있음을 볼 수 있다. SPSS를 수행한 결과 생성되는 변수들은 year_ 등과 같이 변수명의 마지막이 "_"로 끝남을 알 수 있다.



	중간재	year_	month_	date_
1	9.30	1986	1	JAN 1986
2	10.70	1986	2	FEB 1986
3	13.30	1986	3	MAR 1986
4	14.10	1986	4	APR 1986
5	17.80	1986	5	MAY 1986
6	18.10	1986	6	JUN 1986
7	19.40	1986	7	JUL 1986
8	18.80	1986	8	AUG 1986
9	19.10	1986	9	SEP 1986
10	18.40	1986	10	OCT 1986
11	18.00	1986	11	NOV 1986
12	17.00	1986	12	DEC 1986
13	19.50	1987	1	JAN 1987
14	20.10	1987	2	FEB 1987
15	19.40	1987	3	MAR 1987

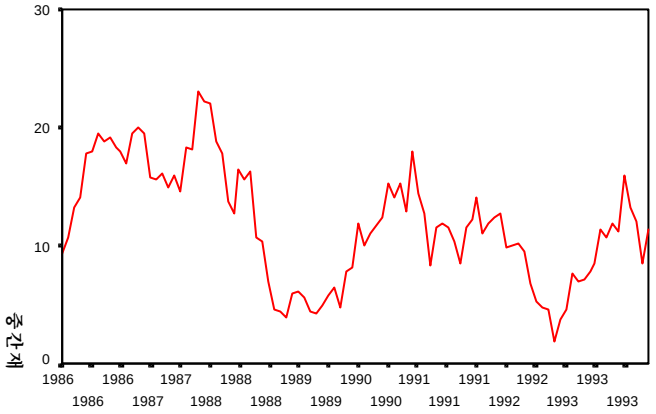
<그림 1.2> 시간변수가 생성되어 저장된 데이터 편집기창



<그림 1.3> 현재 날짜가 정의된 날짜정의 대화상자

시계열분석에서 가장 먼저 해야할 일은 시계열그림(time plot)을 그리는 것이다. 즉, 시계열자료들이 일정한 수준을 중심으로 움직이는지, 추세를 가지고 있는지 또는 계절성이 있는지 등을 시계열그림을 보고 판단하여 이에 적합한 방법을 적용하여야 한다. SPSS에서는 시계열그림을 순차도표라고 부르고 있다. 예제 1의 중간재 출하지수 자료를 이용하여 시계열그림을 그리는 방법에 대해 다시 설명하기로 한다.

시계열그림을 그리기 위해 그래프-순차도표 절차를 선택한 후 왼쪽의 변수목록에서 중간재를 변수(V)로 선택하고 시간축 설명(A)으로 year_를 선택하면 <그림 1.4>를 얻게된다.



<그림 1.4> year_ 변수를 시간축으로 하는 시계열그림

지수평활법(Exponential Smoothing Method)

시계열자료의 특징은 자료들이 시간의 흐름에 따라 관측되므로 독립이 아니라는 점이다. 또한 장기간에 걸쳐서 관측되므로 관측되는 기간 동안에 만약 시계열자료가 생성되던 시스템에 변화가 생긴다면 예측시 이를 반영하여야 한다는 점이다. 이러한 이유에서 예측값을 구할 때 최근의 자료에 더 큰 가중값을 주고 과거로 갈수록 가중값을 줄여나가는 방법을 이용한 예측방법이 지수평활법이다. 지수평활법은 그 계산법이 쉽고 또한 예측값의 계산을 위해 많은 자료를 저장할 필요가 없다는 장점으로 인해 많이 사용되어 온 예측방법이다.

시계열자료가 다음과 같은 추세모형을 따른다고 가정하자.

$$Z_t = \beta_0 + \varepsilon_t \qquad (2.1)$$

$\{Z_t, t=1,\ldots,n\}$이 관측되고 만일 Z_t들이 독립이라면 이들을 이용한 Z_{n+1}의 예측값으

로는 회귀분석에서와 같이 모든 관측값에 동일한 가중값 $1/n$을 부여하는 β_0의 추정량인 $\hat{\beta}_0=\sum_{t=1}^{n} Z_t/n$를 사용하면 될 것이다. 그러나 Z_t들이 시간에 따라 관측된 관계로 가장 최근의 관측값이 더 많은 정보를 가지고 있다면 $1/n$ 대신에 과거로 갈수록 지수적으로 줄어드는 형태의 가중값을 이용하는 것이 타당할 것이다. 이를 위해 $S_n^{(1)}$을 시점 n에서의 예측값이라고 하고 이들이 (2.2)와 같은 관계식을 만족시킨다고 하자.

$$\begin{aligned} S_n^{(1)} &= S_{n-1}^{(1)} + \alpha(Z_n - S_{n-1}^{(1)}) \\ &= \alpha Z_n + (1-\alpha) S_{n-1}^{(1)} \end{aligned} \qquad (2.2)$$

(2.2)는 다음과 같이 전개 될 수 있다.

$$\begin{aligned} S_n^{(1)} &= \alpha Z_n + (1-\alpha)\{\alpha Z_{n-1} + (1-\alpha) S_{n-2}^{(1)}\} \\ &= \alpha Z_n + \alpha(1-\alpha) Z_{n-1} + (1-\alpha)^2 \{\alpha Z_{n-2} + (1-\alpha) S_{n-3}^{(1)}\} \\ &\quad \vdots \\ &= \alpha \sum_{j=0}^{n-1} (1-\alpha)^j Z_{n-j} + (1-\alpha)^n S_0^{(1)} \end{aligned} \qquad (2.3)$$

(2.3)식에서 α는 평활상수(smoothing constant), $S_0^{(1)}$은 초기 평활값, 그리고 $S_n^{(1)}$은 단순지수평활(simple exponential smoothing)통계량이라 부른다. $S_n^{(1)}$은 시점 n까지의 자료들의 가중평균(weighted average)값으로서 가장 최근에 관측된 관측값 $Z_n, Z_{n-1}, Z_{n-2}, \cdots, Z_{n-j}, \cdots$의 순으로 각각 가중값 $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \ldots, \alpha(1-\alpha)^j, \ldots$이 주어지므로 과거의 자료일수록 가중값이 지수적으로 작아져 $S_n^{(1)}$에 작게 반영된다. 반면에 최근의 관측값일수록 가중값은 지수적으로 커져 $S_n^{(1)}$에 크게 반영되므로 $S_n^{(1)}$을 지수평활통계량이라 부른다.

(2.3)을 이용하여 예측값을 구하려면 초기 평활값인 $S_0^{(1)}$과 평활상수인 α가 필요하다. SPSS에서는 $S_0^{(1)}=\bar{Z}$를 사용하고 있으며, α의 값은 사용자가 입력할 수 있도록 하였다. α는 경험적으로 0.05와 0.3 사이의 값을 가진다고 알려져 있으며 이 값이 0에 가까우면 $S_n^{(1)}$

은 표본평균에 가까운 값을 갖고 1에 가까우면 최근의 관측값에 가까운 값을 갖는다. 즉 α의 값이 작을수록 그 평활의 정도가 커져서 예측값들은 부드럽게 변하는 경향을 보인다.

(2.1)은 시계열이 일정한 수준을 중심으로 변하는 경우에 적당한 모형이다. 만약 시계열이 추세를 가지고 증가하는 경우에는 다음과 같은 모형을 가정하는 것이 타당할 것이다.

$$Z_t = \beta_0 + \beta_1 t + \varepsilon_t \qquad (2.4)$$

(2.1)과 (2.4)는 시계열이 계절성을 가지고 있지 않은 경우에 적합한 모형이다. 만약 시계열이 계절성을 가지고 있으면서 일정한 수준을 중심으로 변하는 경우에는

$$Z_t = \beta_0 + I_t + \varepsilon_t \qquad (2.5)$$

와 같은 모형을 적합시키는 것이 타당할 것이다. (2.5)에서 I_t는 계절성분을 나타내는 항으로서 이와 같은 모형을 가법계절모형(additive seasonal model)이라고 부르며 시계열의 변동폭이 시간이 흐름에 따라 변하지 않는 경우에 적합한 모형이다. 시계열의 변동폭이 시간의 흐름에 따라 점차로 커지는 경우에는 (2.6)과 같이 계절성분이 곱의 형태로 주어지는 모형이 적합할 것이다.

$$Z_t = \beta_0 I_t + \varepsilon_t \qquad (2.6)$$

(2.6)과 같은 모형을 승법계절모형(multiplicative seasonal model)이라고 한다.

이들 모형 외에도 SPSS는 다음과 같은 모형들에 적합한 평활법을 제공하고 있다.

$$\begin{aligned}
&Z_t = \beta_0 + \beta_1 t + I_t + \varepsilon_t : \text{선형추세가 있는 가법계절모형} \\
&Z_t = (\beta_0 + \beta_1 t) I_t + \varepsilon_t : \text{선형추세가 있는 승법계절모형} \\
&Z_t = \beta_0 \beta_1^t + \varepsilon_t : \text{지수적으로 증가하는 추세가 있는 지수추세모형} \\
&Z_t = \beta_0 \beta_1^t + I_t + \varepsilon_t : \text{지수추세가 있는 가법계절모형}
\end{aligned} \qquad (2.7)$$

$Z_t = (\beta_0 \beta_1^t) I_t + \varepsilon_t$: 지수추세가 있는 승법계절모형

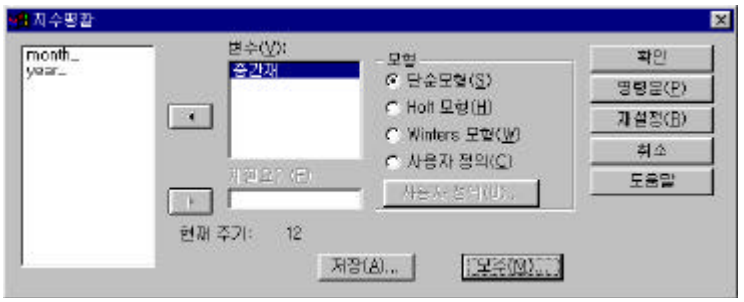
$Z_t = \beta_0 + \phi \beta_1 t + \varepsilon_t$: 변동폭이 줄어들고 추세가 있는 감소추세모형

$Z_t = \beta_0 + \phi \beta_1 t + I_t + \varepsilon_t$: 감소추세가 있는 가법계절모형

$Z_t = (\beta_0 + \phi \beta_1 t) I_t + \varepsilon_t$: 감소추세가 있는 승법계절모형

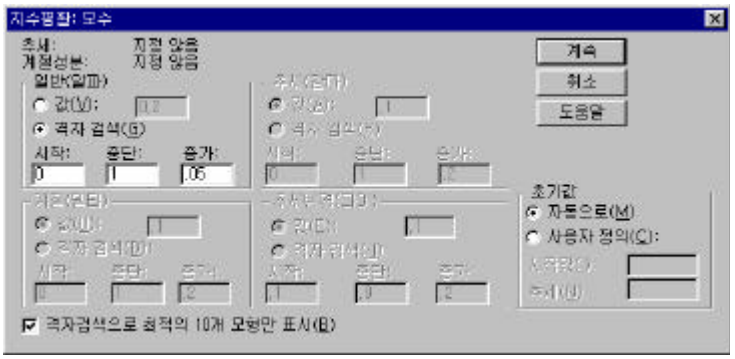
예제 2 예제 1의 자료를 이용하여 지수평활법에 의해 예측값을 구하는 방법에 대해 설명한다.

(1) 주메뉴에서 통계분석(S)-시계열분석(T)-지수평활(E) 절차를 선택한다. <그림 2.1> 지수평활 대화상자의 변수(V)에 중간재를 선택한다. 중간재 출하지수 자료의 시계열그림인 <그림 1.4>를 보면 일정한 수준을 중심으로 움직이는 것을 알 수 있어 식 (2.1)과 같은 추세와 계절성이 없는 모형이 적합함을 알 수 있다. 이러한 모형의 경우에는 단수지수평활법을 적용하며, SPSS에서는 이를 단순모형(S)이라고 부르고 있다.



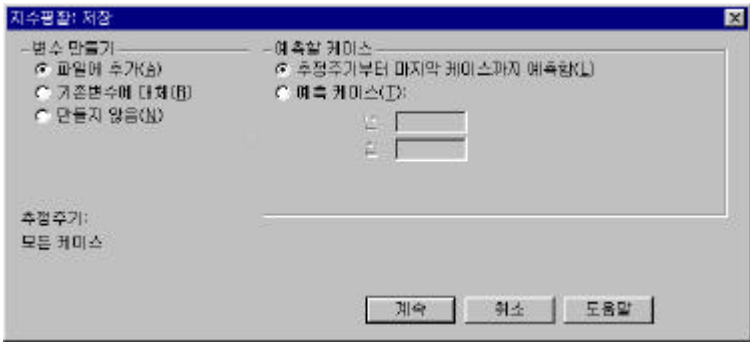
<그림 2.1> 지수평활 대화상자

(2) 모수 단추를 누르면 <그림 2.2>지수평활: 모수대화상자가 나타난다. 여기서 원하는 알파 값을 입력한다. 디폴트 값은 0.1이다. 만약 예측오차의 제곱합 SSE를 최소로 하는 알파값을 구하고자 하면 격자검색(G)를 선택한 후 적당한 시작, 중단 및 증가 값을 입력한다. 시작값으로는 0, 중단값으로는 1 그리고 증가값으로는 0.1이 디폴트로 사용된다. 또한 초기값을 자동으로(M) 할 것인지 사용자 정의(C)로 할 것인지를 결정한다. 특별한 이유가 없으면 디폴트인 자동으로(M)를 선택한 후 계속 단추를 누른다.



<그림 2.2> 지수평활 : 모수 대화상자

(3) 지수평활 대화상자에서 저장단추를 누르고 <그림 2.3> 지수평활 : 저장 대화상자에서 저장하기를 원하는 값과 변수형태를 지정하면 된다.



<그림 2.3> 지수평활 : 저장 대화상자

(4) 다음은 지수평활법을 수행한 결과의 출력이다. α=0.9일 때의 SSE=438.95 으로 가장 작으므로 예측값은 α=0.9를 이용하여 구한 값이라는 설명이 추가되어 있다.

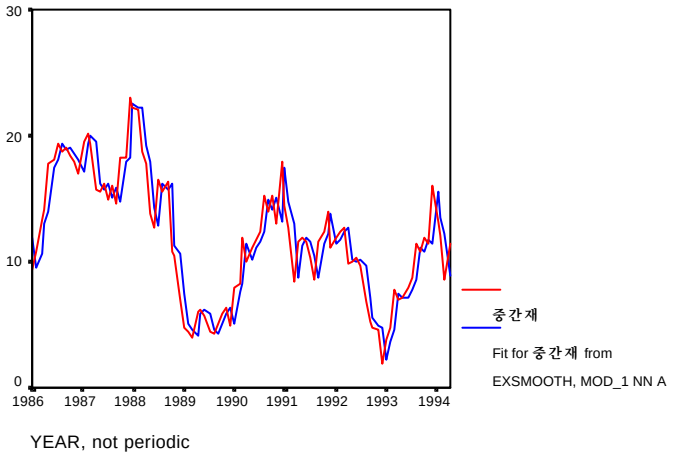
지수 평활

```
MODEL:  MOD_1.

Results of EXSMOOTH procedure for Variable 품간재
MODEL= NN (No trend, no seasonality)

 Initial values:       Series            Trend
                    11,90700         Not used

DFE = 99.

The 10 smallest SSE's are:      Alpha             SSE
                              ,9000000       488,96840
                              ,8000000       444,51477
                             1,000000        444,92645
                              ,7000000       461,85825
                              ,6000000       498,85464
                              ,5000000       544,54972
                              ,4000000       627,12528
                              ,3000000       766,40299
                              ,2000000      1021,09197
                              ,1000000      1525,78624

The following new variables are being created:

  NAME          LABEL

  FIT_1         Fit for 품간재 from EXSMOOTH, MOD_1 NN A ,90
  ERR_1         Error for 품간재 from EXSMOOTH, MOD_1 NN A ,90
```

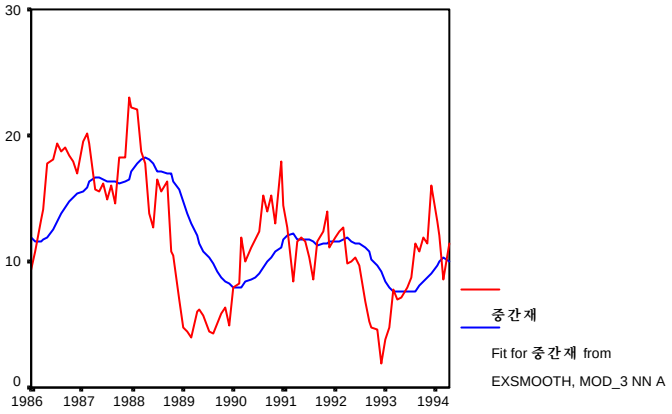
<그림 2.4>를 보면 예측값과 예측오차가 원래의 데이터 편집기 창에 추가되어 있음을 알 수 있다. 이제 원래의 관측값과 예측값의 시계열그림을 겹쳐 그려 예측이 어느 정도로 좋은지를 확인해 보기로 한다. α=0.9를 이용하여 예측값을 구했으므로 예측값은 최근의 값과 비슷한 값을 갖게되어 <그림 2.5>에서 보는 것처럼 예측값이 관측값과 거의 유사함을 볼 수 있다. 참고로 α=0.1일 때의 예측값의 시계열그림을 보면 α=0.9인 경우와는 달리 매우 부드러운 곡선으로 나타남을 볼 수 있다.

	중간재	year_	month_	date_	fit_1	err_1	var	var
1	9.30	1986	1	JAN 1986	11.90700	-2.60700		
2	10.70	1986	2	FEB 1986	9.58070	1.13930		
3	13.30	1986	3	MAR 1986	10.58607	2.71393		
4	14.10	1986	4	APR 1986	13.02861	1.07139		
5	17.80	1986	5	MAY 1986	13.99286	3.80714		
6	18.10	1986	6	JUN 1986	17.41929	.68071		
7	19.40	1986	7	JUL 1986	18.03193	1.36807		
8	18.80	1986	8	AUG 1986	19.28319	-.48319		
9	19.10	1986	9	SEP 1986	18.84632	.25368		
10	18.40	1986	10	OCT 1986	19.07463	-.67463		
11	18.00	1986	11	NOV 1986	18.46746	-.46746		
12	17.00	1986	12	DEC 1986	18.04675	-1.04675		
13	19.50	1987	1	JAN 1987	17.10467	2.39533		
14	20.10	1987	2	FEB 1987	19.26047	.83953		
15	19.40	1987	3	MAR 1987	20.01605	-.61605		
16	16.70	1987	4	APR 1987	19.45160	-2.75160		

<그림 2.4> 예측값과 예측오차가 추가로 저장된 데이터편집기 창

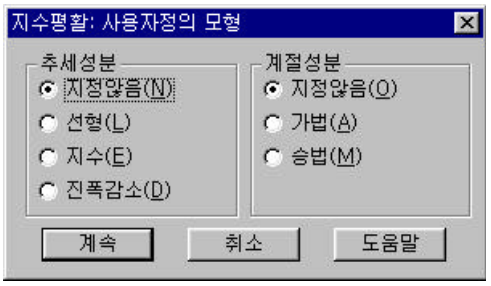


<그림 2.5> α=0.9일 때의 관측값과 예측값의 시계열그림



<그림 2.6> α=0.1일 때의 관측값과 예측값의 시계열그림

식 (2.4)와 같은 선형추세모형을 적합시키려면 <그림 2.1>에서 Holt 모형(H)을 선택하고 식(2.6)과 같은 승법계절모형을 적합시키려면 Winters 모형(W)을 선택하면 된다. 가법계절모형 및 식 (2.7)에서 정의된 다른 모형들을 적합시키려면 사용자 정의(C)를 선택한 후 사용자 정의 대화상자인 <그림 2.7>에서 추세성분과 계절성분을 적당히 조합하여 선택하면 된다. 예를 들어 지수추세가 있는 승법계절모형을 선택하려면 추세성분에서 지수(E)를 계절성분에서 승법(M)을 선택하면 된다.



<그림 2.7> 사용자정의 모형 대화상자

자기회귀오차모형(Autoregressive Error Model)

추세모형이나 회귀모형을 이용하여 자료를 분석할 경우에는 일반적으로 오차항이 서로 독립이고 동일한 분포를 따른다는 가정을 필요로 한다. 그러나 반응변수(또는 종속변수)와 설명변수(또는 독립변수)들이 시계열자료인 경우에는 오차항들이 시간에 따른 자기상관(autocorrelation) 관계를 갖는 경우가 대부분이다. 따라서 시계열자료들을 이용한 회귀분석시에는 오차항들이 서로 독립인지 여부를 판단하기 위해 시간에 따른 오차들의 시계열그림을 그려보거나 Durbin-Watson(DW) 통계량을 이용하여 독립성검정을 시행한다. 만일 오차들이 서로 독립이 아니라면 오차들 사이의 공분산을 추정하여 일반화최소제곱법(generalized least squares method)을 이용하여 해결할 수 있다. 그러나 오차들이 시간에 따라 특별한 형태의 모형을 따른다면 이를 모형적합 시에 반영하는 것이 좀더 일반적인 해결법이다.

다음과 같이 오차들이 자기회귀(autoregressive) 모형을 따르는 회귀모형을 생각해 보자.

$$\begin{aligned} Z_t &= \beta_0 + \sum_{i=1}^{k} \beta_i x_{it} + u_t, \quad t = 1, 2, \ldots, n \\ u_t &= \phi_1 u_{t-1} + \cdots + \phi_p u_{t-p} + \varepsilon_t \end{aligned} \tag{3.1}$$

단, ε_t는 오차로서 서로 독립이고 평균이 0, 분산이 σ^2인 확률변수이다. 식(2.8)은 오차들이 차수가 p인 자기회귀모형, AR(p), 를 따르므로 자기회귀오차모형이라고 부르며 오차들이 자기상관관계를 갖고 있는 경우에 적합한 모형이다.

SPSS에서는 식(3.1)모형 중에서 $p=1$인 경우의 자기회귀오차모형을 자기회귀(U) 부메뉴를 이용하여 적합시킬 수 있다. p가 2 이상인 경우는 회귀모형과 ARIMA모형을 이용하여 적합시킬 수 있으나 효율적인 방법은 아니다. $p=1$인 경우의 자기회귀오차모형을 적합시키기 위해 SPSS에서는 최대가능도방법과 최소제곱법에 기초한 Cochrane-Orcutt 방법과 Prais-Winsten 방법을 이용할 수 있다.

오차들이 독립인지 여부를 검정하는데 사용되는 DW 검정통계량은 다음과 같이 정의된다.

$$DW = \frac{\sum_{t=1}^{n} (\hat{u}_{t+1} - \hat{u}_t)^2}{\sum_{t=1}^{n} \hat{u}_t^2} \tag{3.2}$$

단, $\hat{u}_t$는 일반적인 회귀모형을 적합시킨 후의 잔차들이다. DW 검정통계량은

$$u_t = \phi u_{t-1} + \varepsilon_t \qquad (3.3)$$

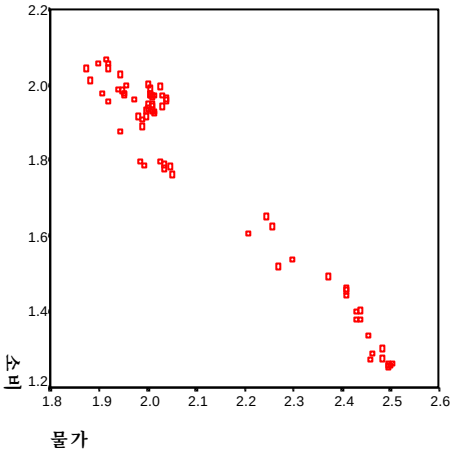
에서 오차들 사이의 1차 자기상관관계의 존재여부의 검정, 즉 $\phi = 0$인지 여부의 검정에 이용된다. 양의 자기상관이 강할수록 ϕ는 1에 가까운 값을 가질 것이므로 분자가 작아져서 DW 검정통계량값은 0에 가까울 것이며, 음의 자기상관이 강할수록 ϕ는 -1에 가까운 값을 가지게 되어 DW 검정통계량의 값은 4에 가까운 값을 가질 것이다. 자기상관이 존재하지 않으면 DW 검정통계량은 2에 가까운 값을 가질 것이다. DW 검정통계량의 유의성검정을 위해서는 DW검정을 위한 표가 필요하며 대부분의 회귀분석 교재들에서 제공되고 있다.

식(3.3)의 DW 검정통계량은 식(3.1)과 같은 모형에서 1차 자기상관이 존재하는지의 여부만을 검정할 수 있으며 고차의 자기상관의 존재여부를 알아보기 위해서는 일반화된 DW 검정법을 사용하여야 하나 SPSS에서는 이를 제공하지 않으므로 더 이상은 언급하지 않기로 한다. 만일 DW 검정결과 오차들 사이에 자기상관관계가 없다면 자기회귀오차모형 대신에 일반적인 회귀모형을 적합시키면 될 것이다.

이제 AREG절차를 이용하여 자기회귀오차모형을 적합시키는 방법에 대해 설명하기로 한다. 반응변수와 설명변수들의 관측값이 데이터편집기에 저장된 상태에서 먼저 회귀분석을 이용하여 오차들이 독립인지 여부를 알아본 후 자기상관관계가 있는지를 알아본다. 이를 위해서는 회귀분석에서와 같이 통계분석(S)-회귀분석(R)-선형(L) 절차를 선택한 후 통계량 대화상자에서 Durbin-Watson(U)를 선택하면 된다.

예제 3 SPSS의 Trends Chapter9.sav에 저장되어 있는 소비, 수입, 물가 자료를 이용하여 자기회귀오차모형을 적합시키는 방법을 설명하기로 한다. 관측값의 개수는 69개이다. 소비가 물가에 의해 어떻게 영향을 받는지를 알아보고자 한다.

(1) 먼저 물가자료와 소비자료의 산점도를 그려본다.



<그림 3.1> 물가와 소비의 산점도

<그림 3.1>로부터 물가가 증가하면 소비가 감소하는 경향이 있으며 두 변수 사이에는 선형관계가 있음을 알 수 있다. 따라서 물가자료를 반응변수, 소비자료를 설명변수로 하는 선형모형이 적합하다는 것을 알 수 있다.

(2) 물가자료를 반응변수, 소비자료를 설명변수로 하는 선형모형을 적합시킨다. 적합 결과는 다음과 같다.

모형 요약[b]

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.977[a]	.955	.954	5.80E-02	.237

a. 예측값: (상수), 물가
b. 종속변수: 소비

분산분석[b]

모형		제곱합	자유도	평균제곱	F	유의확률
1	선형회귀분석	2.833	1	2.833	1421.520	.000[a]
	잔차	.134	67	1.993E-03		
	합계	2.966	68			

a. 예측값: (상수), 소비
b. 종속변수: 물가

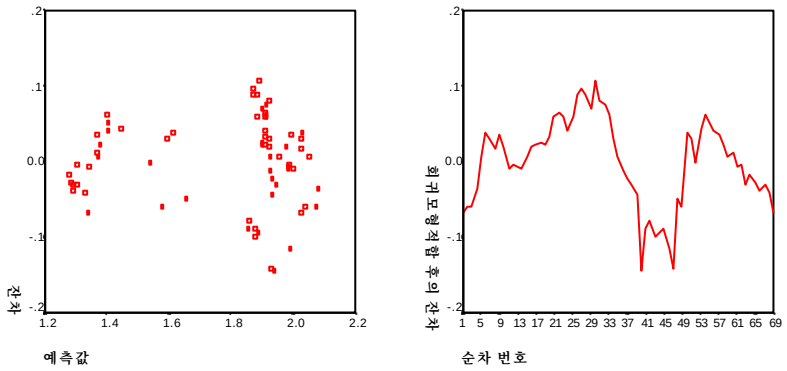
계수[a]

모형		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	4.458	.072		62.241	.000
	물가	-1.269	.034	-.977	-37.703	.000

a. 종속변수: 소비

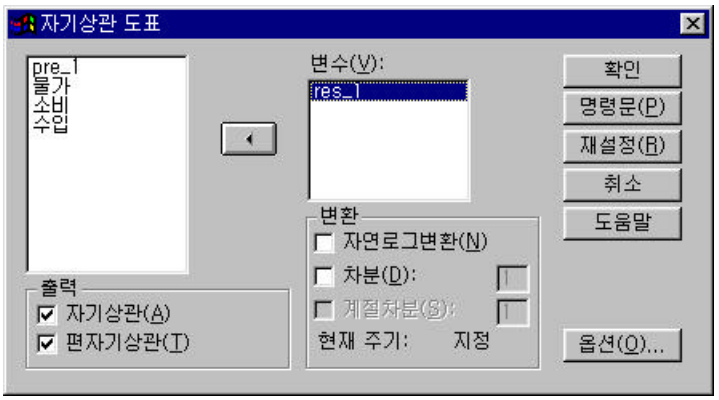
분산분석과 계수의 추정량에 대한 요약으로부터 적합된 모형이 유의함을 알 수 있다. 그러나 모형요약의 Durbin-Watson검정통계량의 값은 0.237로서 잔차들 사이에 양의 자기상관이 존재함을 알 수 있다.

다음은 잔차와 예측값과의 산점도와 잔차의 시계열그림이다. 시간에 따른 독립성의 문제 외에는 크게 문제가 되지 않는 것 같다. 잔차의 시계열그림은 강한 자기상관의 가능성을 보여주고 있다.



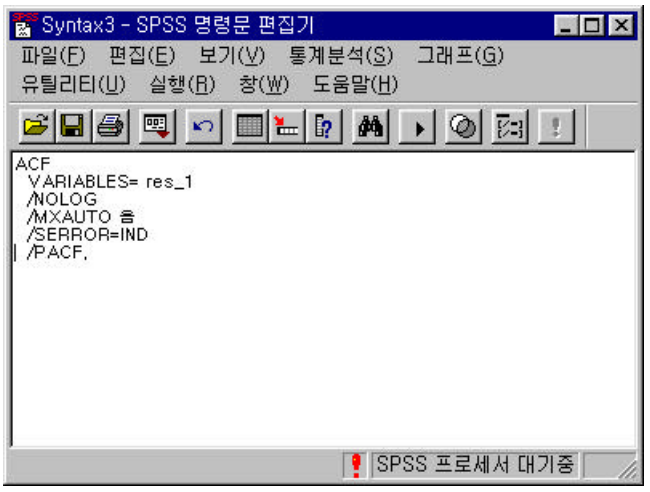
<그림 3.2> 잔차들의 그림

(3) 잔차들의 자기상관여부의 판단을 위해서는 다음과 같이 잔차들의 자기상관(ACF)도표를 이용하여도 된다. 그래프(G)-시계열도표(T)-자기상관(A) 절차를 선택하면 <그림 3.3>과 같이 자기상관도표 대화상자가 나타난다. 왼쪽의 변수목록에서 회귀모형을 적합한 결과의 잔차인 res_1를 선택한 후 확인 단추를 누른다.



<그림 3.3> 자기상관도표 대화상자

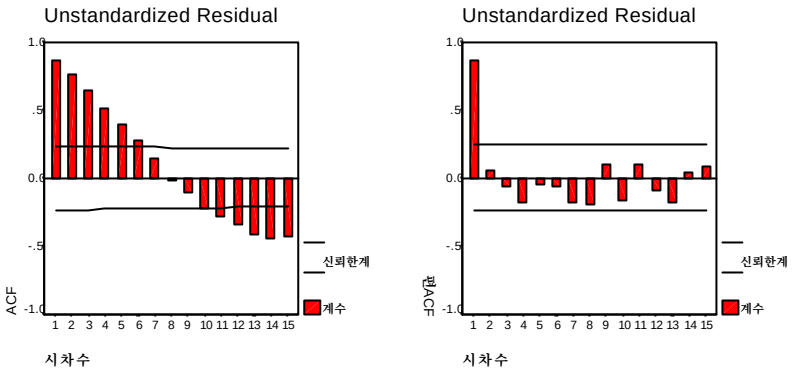
그러나 SPSS을 한글화하면서 버그(bug)가 발생하였다. 확인 단추를 누르면 에러메시지가 출력되는 경우가 있는데, 이는 " /MAXAUTO"의 디폴트값이 "음"으로 정의되기 때문이다. 이를 해결하는 방법은 다음과 같다. <그림 3.3>에서 명령문 단추를 누르면 <그림 3.4>와 같이 SPSS 명령문 편집기가 나타난다. 여기서 네번째 줄 "/MAXAUTO 음"의 "음"을 "/MAXAUTO 15"와 같이 원하는 차수의 값으로 바꾸어 준 후 실행(R) 단추를 눌러 실행시키면 된다.



<그림 3.4> SPSS 명령문 편집기

실행결과 다음과 같이 acf와 pacf도표가 출력된다. acf와 pacf로부터 잔차들이 AR(1)모형을 따름을 알 수 있다.

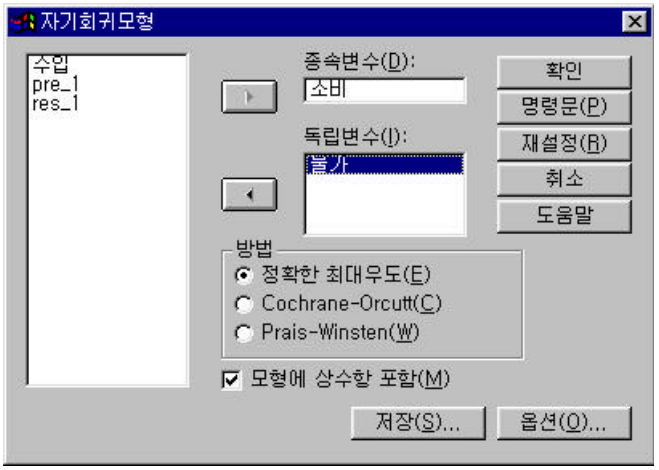
```
MODEL: MOD_1.

Autocorrelations:   RES_1   Unstandardized Residual

     Auto-  Stand.
Lag  Corr.   Err.  -1  -.75  -.5  -.25   0   .25  .5   .75   1   Box-Ljung  Prob.
                   +----+----+----+----+----+----+----+----+----+
  1   .861   .118                .    [****.**********          58.480   .000
  2   .758   .117                .    [****.**********          95.416   .000
  3   .651   .116                .    [****.********           126.886   .000
  4   .515   .115                .    [****.*****              146.859   .000
  5   .396   .114                .    [****.***                158.880   .000
  6   .280   .113                .    [****.*                  164.961   .000
  7   .189   .112                .    [***.                    166.482   .000
  8  -.012   .112                .    *    .                   166.494   .000
  9  -.101   .111                .  **[    .                   167.824   .000
 10  -.221   .110                .****[    .                   171.887   .000
 11  -.279   .109                **.***[   .                   177.984   .000
 12  -.385   .108                ***.***[  .                   187.624   .000
 13  -.417   .107                ****.***[ .                   202.859   .000
 14  -.446   .106                *****.***[.                   220.607   .000
 15  -.487   .105                *****.***[.                   287.948   .000

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  69     Computable first lags:  68

Partial Autocorrelations:   RES_1   Unstandardized Residual

   Pr-Aut-  Stand.
Lag  Corr.   Err.  -1  -.75  -.5  -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+----+
  1   .861   .120                .    [****.**********
  2   .062   .120                .    [*   .
  3  -.058   .120                .   *[    .
  4  -.176   .120                .****[    .
  5  -.051   .120                .   *[    .
  6  -.068   .120                .   *[    .
  7  -.176   .120                .****[    .
  8  -.198   .120                .****[    .
  9   .098   .120                .    [**  .
 10  -.165   .120                . ***[    .
 11   .099   .120                .    [**  .
 12  -.089   .120                .  **[    .
 13  -.182   .120                .****[    .
 14   .087   .120                .    [*   .
 15   .085   .120                .    [**  .

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  69     Computable first lags:  68
```



<그림3.5> 회귀모형 적합 후의 잔차들의 ACF와 PACF

<그림 3.5>를 보면 잔차들의 acf와 pacf가 텍스트형태와 그래프의 두가지 형태로 출력됨을 볼 수 있다. 어느 것을 사용하여도 무방하다.

(4) 이제 오차가 AR(1) 모형을 따르는 자기회귀오차모형을 적합시키기로 한다. 이를 위해서 통계분석(S)-시계열분석(T)-자기회귀(U) 절차를 선택하면 <그림 3.6> 자기회귀모형 대화상자가 나타난다.



<그림 3.6> 자기회귀모형 대화상자

종속변수(D)로 소비를 독립변수(I)로 물가를 선택한 후 저장과 옵션 단추를 이용하여 원하는 형태의 모형을 선택한다. 다음은 출력물 중에서 분산분석표와 계수추정부분만을 발췌한 결과이다.

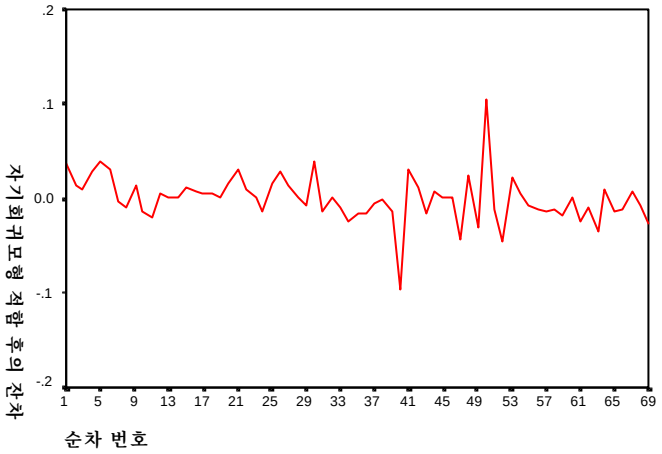
```
MODEL:  MOD_2

Model Description:

Variable:   소비
Regressors: 물가

FINAL PARAMETERS:

Number of residuals  69
Standard error       ,02545518
Log likelihood       155,55096
AIC                  -305,10193
SBC                  -298,89961

            Analysis of Variance:

               DF  Adj. Sum of Squares    Residual Variance

Residuals      66            ,04448981            ,00064797

           Variables in the Model:

                       B          SEB      T-RATIO   APPROX. PROB.

AR1            ,9667144   ,03864885    28,729495         ,0000000
물가          -,9496926   ,08585896   -11,114158         ,0000000
CONSTANT      3,7888821   ,19566689    19,108149         ,0000000
```

적합된 모형은 다음과 같으며 모든 모수들이 유의함을 알 수 있다.

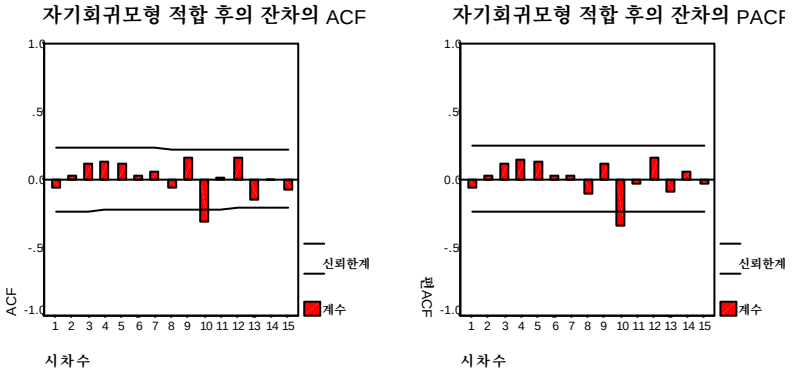
$$\text{소비}_t = 3.74 - 0.95\text{물가}_t + u_t$$

$$u_t = .97u_{t-1} + \varepsilon_t$$

마지막으로 자기회귀오차모형을 적합한 후의 잔차인 er_1의 시계열그림과 acf 및 pacf도표를 확인해 보기로 한다. <그림 3.7>과 <그림 3.8>을 보면 lag 10에서 acf와 pacf가 신뢰한계를 벗어나는 것 이외에는 특별히 의심할 만한 점이 없어 모형이 잘 적합된 것같다.

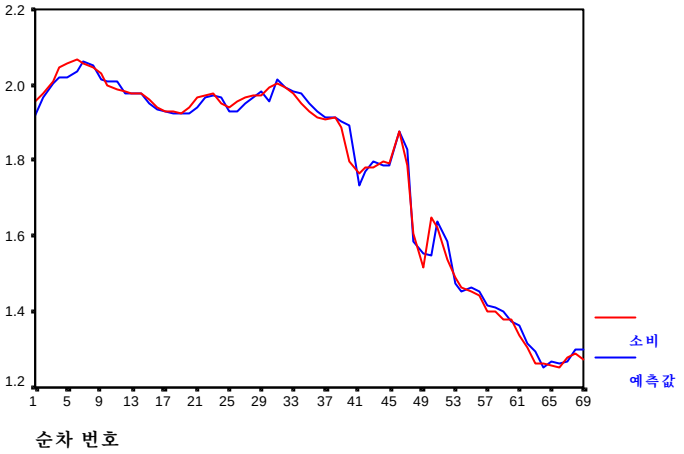


<그림 3.7> 자기회귀모형 적합 후의 잔차의 시계열그림



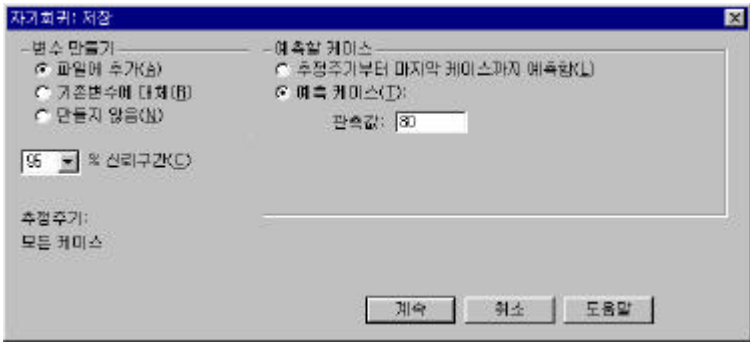
<그림 3.8> 자기회귀모형 적합 후의 잔차들의 ACF와 PACF

관측값과 적합된 모형을 이용한 관측된 기간 동안의 예측값의 시계열그림인 <그림 3.9>를 보면 예측이 잘 되었음을 볼 수 있다.



<그림 3.9> 자기회귀모형 적합 후의 관측값과 예측값의 시계열그림

미래의 예측값을 구하기 위해서는 설명변수인 물가의 예측값이 있어야 하며, 지수평활법 또는 ARIMA모형을 이용한 일변량시계열자료의 예측을 이용하면 될 것이다. 일단 예측하고자 하는 미래의 시점까지의 물가의 예측값이 구해지면 이들 예측값을 설명변수로 하는 자기회귀오차모형을 적합시키되 <그림 3.6>에서 저장 단추를 눌러 나타나는 <그림 3.10>의 저장 대화상자에서 예측케이스를 선택하고 원하는 미래의 시점의 값을 관측값으로 입력하고 이를 수행하면 된다.



<그림 3.10> 자기회귀 : 저장 대화상자

자기회귀누적이동평균모형

(Autoregressive Integrated Moving Average Model)

자기회귀누적이동평균(ARIMA)모형은 Box와 Jenkins가 3단계 모형적합과정을 제안한 이후로 시계열자료의 분석에 가장 널리 사용되는 모형이다. ARIMA모형의 기본 형태는 시계열이 차분(differencing)을 필요로 하느냐의 여부에 따라 다르다. 시계열이 차분을 하지 않아도 정상(stationary)인 경우에는 ARMA(p,q)모형을 적합시키고, 차분을 필요로 하는 경우에는 차분을 하여 정상화한 후 ARMA(p,q)모형을 적합시킨다. 즉, 원 시계열 Z_t를 d번 차분한 시계열 $W_t=(1-B)^dZ_t$가 정상이고 ARMA(p,q)모형을 따른다면 Z_t는 ARIMA(p,d,q)모형을 따른다고 하며 다음과 같이 표현한다.

$$(1-\phi_1B-\cdots-\phi_pB^p)(1-B)^dZ_t=\delta+\varepsilon_t-\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q},\quad \varepsilon_t\sim \text{i.i.d. } N(0,\sigma^2) \tag{4.1}$$

단, ε_t는 오차항으로서 서로 독립이고 평균이 0, 분산이 σ^2인 정규확률변수이며 백색잡음(white noise)이라고 부른다. 즉, 현 시점에서의 관측값 Z_t는 과거의 관측값 $Z_{t-1}, Z_{t-2}, \cdots, Z_{t-p}$와 백색잡음 $\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q}$들의 선형함수로 표현 가능하다는 것이 ARIMA모형의 기본 개념이다. 식(4.1)에서 $d=0$인 경우의 모형을 ARMA(p,q)모형, $q=0$인 경우는 AR(p)모형, $p=0$인 경우는 MA(q)모형이라고 부른다.

식(4.1)모형은 다음과 같은 연산자(operator)를 이용하여 표현할 수도 있다.

$$\phi(B)(1-B)^dZ_t=\delta+\theta(B)\varepsilon_t$$

$\phi(B)=1-\phi_1B-\cdots-\phi_pB^p$: AR 연산자

$\theta(B)=1-\theta_1B-\cdots-\theta_qB^q$: MA 연산자

시계열이 주기가 s인 계절성을 가지는 경우에는 아래와 같은 계절형 연산자들을 이용하여

$\Phi(B^s)=1-\Phi_1 B^s-\cdots-\Phi_P B^{Ps}$: 계절형 AR 연산자

$\Theta(B^s)=1-\Theta_1 B^s-\cdots-\Theta_Q B^{Qs}$: 계절형 MA 연산자

다음과 같은 계절형 ARIMA$(p,d,q)(P,D,Q)s$ 모형을 적합시킬 수 있다.

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Z_t=\delta+\theta(B)\Theta(B^s)\varepsilon_t$$

ARIMA모형의 적합은 다음과 같이 일반적인 모형적합의 3단계와 동일하게 이루어진다.

모형의 식별(model identification)

모형의 추정(model estimation)

모형의 진단(model diagnostic checking)

위의 3단계를 거쳐 최종적으로 선택된 모형을 이용하여 예측(forecasting)에 이용하면 된다.

① 모형의 식별 : 시계열그림(time series plot), 자기상관함수(autocorrelation function : acf), 부분자기상관함수(partial autocorrelation function : pacf) 등을 이용하여 시계열의 정상(stationary) 여부를 판단하고 AR 차수 p와 MA 차수 q 등을 결정한다. 이 단계에서는 하나 이상의 잠정적인 모형을 선택할 수 있다. 만약 시계열자료가 정상이 아니라고 판단되면 적절한 차수의 차분(differencing)을 시행하거나 로그변환 등의 분산안정화 변환을 실시하여 정상화시킨다.

② 모형의 추정 : 조건부최소제곱법(conditional least squares method), 비조건부최소제곱법(unconditional least squares method) 및 최대가능도법(maximum likelihood method) 등을 이용하여 잠정적으로 선택된 모형의 모수들을 추정한다.

③ 모형의 진단 : 추정단계에서 출력된 AIC, SBC 등의 적합성검정통계량과 잔차분석을 통해 잠정적으로 추정된 모형의 적합 여부를 판단한다. 만약 잠정모형이 적절하지 못

하다고 판단되면 새로운 모형을 식별하여 다시 추정을 한다. 잔차분석에서는 주로 잔차들의 시계열그림, racf와 rpacf 등을 이용하여 잔차의 백색잡음 여부를 판단한다.

모형의 식별단계에서는 다음과 같은 acf와 pacf의 이론적인 특성을 참고하면 모형의 차수를 정하는데 도움이 된다.

모형	ACF	PACF
AR(p)	지수적으로 감소하거나 소멸하는 sine 함수 형태	p 시차 이후에는 0으로의 절단 형태
MA(q)	q 시차 후에는 0으로의 절단형태	지수적으로 감소하거나 소멸하는 sine 함수 형태
ARMA(p,q)	시차 (q-p) 이후로는 소멸하는 형태	시차 (p-q) 이후로는 소멸하는 형태

모형의 진단단계에서 잔차분석시, 잔차의 백색잡음 여부의 검정을 위해서는 다음과 같은 Portmanteau Q-통계량이 χ^2-분포를 따름을 이용한다.

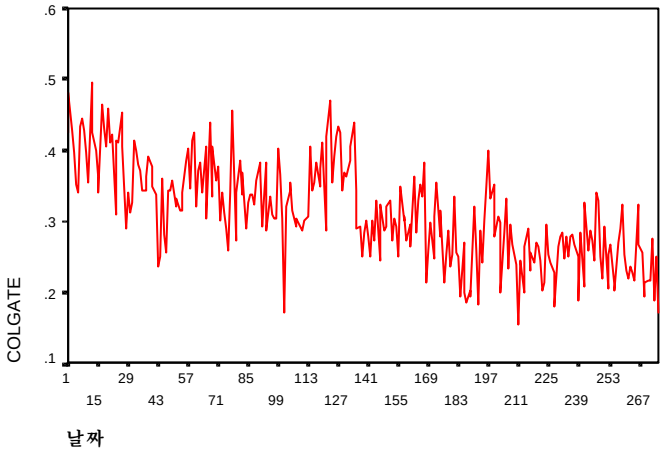
$$\chi^2_K = n(n+2)\sum_{k=1}^{K}\frac{r_k^2}{(n-k)}, \quad r_k = \sum_{t=1}^{n-k} a_t a_{t+k} / \sum_{t=1}^{n} a_t^2$$

단, a_t는 잔차이고 n은 시계열의 관측개수이며 r_k는 잔차의 sacf이다. 이 값을 이용하여 잔차가 백색잡음이라는 가설검정의 유의확률이 작은 경우 잔차가 백색잡음이 아니라고 판단한다. ARIMA모형에 대한 자세한 내용은 조신섭, 손영숙(1999)을 참고하기 바란다.

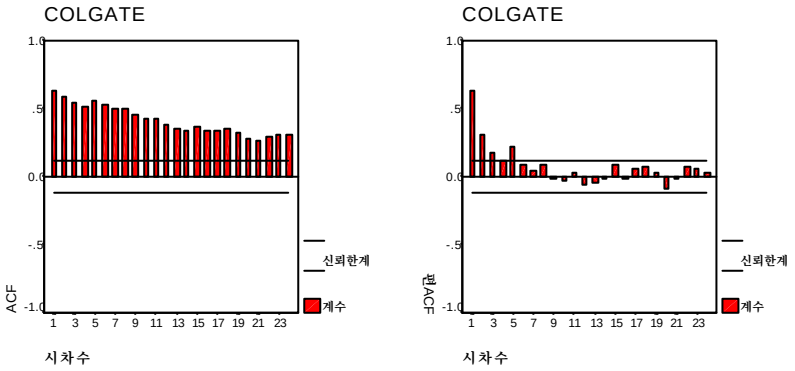
예제 3 SPSS의 Trends Chapter10.sav에 저장되어 있는 colgate와 crest의 시장점유율자료를 이용하여 ARIMA모형을 적합시키는 방법에 대해 설명하기로 한다. 이 자료는 1958년부터 1963년까지 276주 동안의 colgate와 crest의 시장점유율 경쟁시의 자료이다. $t=125$에서 crest가 미국치과협회의 승인을 받은 사실을 광고에 이용하면서 시장점유율에 어떠한 변화가 생겼는지에 대한 연구가 많이 이루어 진 자료이다. 여기서는 colgate의 자료만을 이용하여 ARIMA모형을 적합시켜보기로 한다.

(1) 모형의 식별

ARIMA모형 적합의 첫 단계인 모형의 식별단계에서는 먼저 시계열그림을 그려보아 정상 여부를 판단한다. <그림 4.1>은 colgate의 시장점유율의 시계열그림이다. 감소하는 추세를 보여주므로 정상이 아니라고 생각된다.

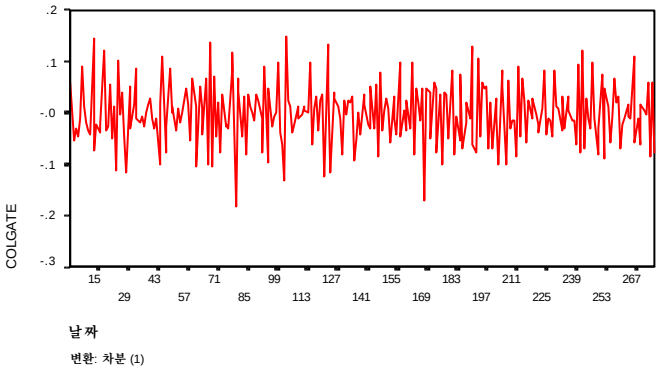


<그림 4.1> colgate 자료의 시계열그림



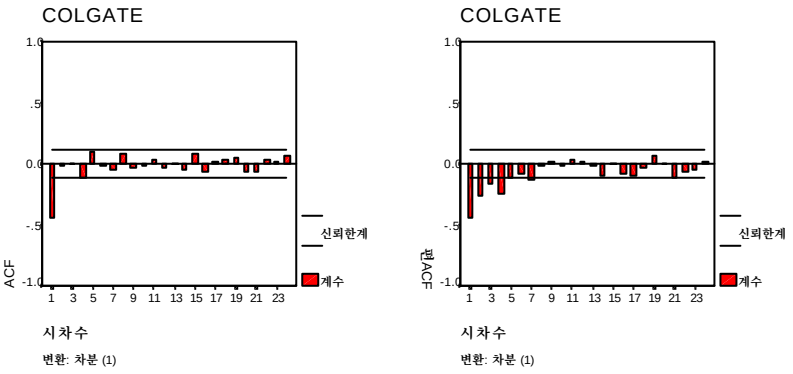
<그림 4.2> colgate 자료의 ACF와 PACF

참고로 1차 차분하기 전의 colgate 자료의 acf와 pacf도표를 그려보면 <그림 4.2>와 같이 acf는 서서히 감소하는 경향을 보여주므로 역시 정상이 아니라는 것을 확인할 수 있다. 따라서 원 시계열을 1차 차분을 한 후의 시계열그림을 그려보기로 한다. 1차 차분을 한 후의 시계열그림을 그리기 위해서는 순차도표 대화상자의 변환에서 차분을 선택하면 된다.



<그림 4.3> 1차 차분한 후의 colgate 자료의 시계열그림

<그림 4.3>을 보면 특별히 증가하거나 감소하는 추세는 보이지 않으므로 정상이라고 생각된다. 1차 차분된 colgate 자료의 acf와 pacf를 그리기 위해서는 <그림 4.4> 자기상관도표 대화상자의 변환에서 차분을 선택하면 된다. 다음은 1차 차분한 후의 acf와 pacf이다.

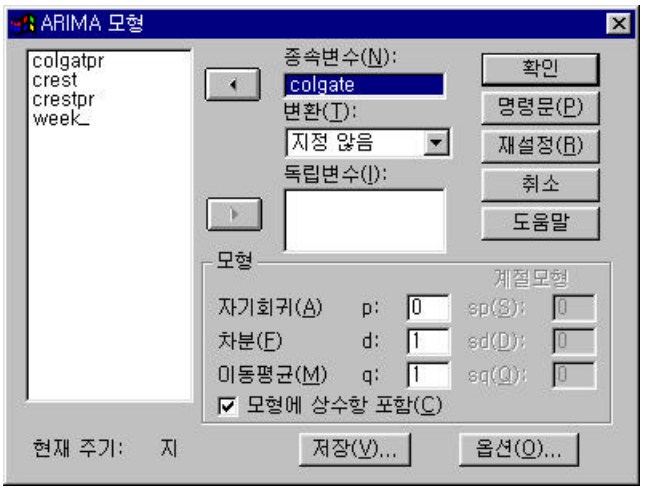


<그림 4.5> 1차 차분한 후의 colgate 자료의 ACF와 PACF

1차 차분한 자료의 acf는 시차 1에서만 유의하고 pacf는 시차 1부터 4까지 유의하되 줄어드는 경향을 보이므로 MA(1)모형이 적당하다고 생각된다. 따라서 ARIMA(0,1,1)모형을 적합시키기로 한다.

(2) 모형의 적합 및 진단

ARIMA(0,1,1)모형을 적합시키기 위해 통계분석(S)-시계열분석(T)-ARIMA 모형(I) 절차를 선택한다. 다음과 같은 ARIMA모형 대화상자에서 종속변수로 colgate를 선택하고 모형에서 차분(F) 차수로 1을, 이동평균(M) 차수로 1을 입력하고 이를 수행하면 된다.

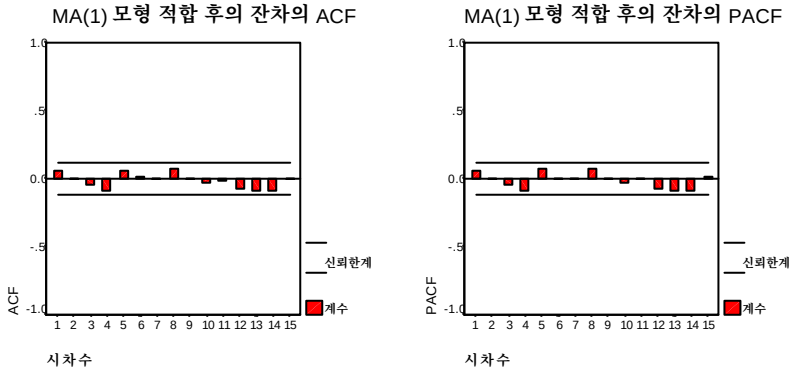


<그림 4.6> ARIMA모형 대화상자

다음은 ARIMA(0,1,1)모형을 적합한 결과이다.

```
FINAL PARAMETERS:

Number of residuals  275
Standard error       .04766722
Log likelihood       447.8062
AIC                  -890.61289
SBC                  -888.87885

           Analysis of Variance:

              DF  Adj. Sum of Squares    Residual Variance

Residuals     273           .62285178           .00227216

           Variables in the Model:

                      B          SEB      T-RATIO   APPROX. PROB.

MA1           .77287748   .03892742    19.8414777       .00000000
CONSTANT    -.00075112   .00066251    -1.1337836       .25789979
```

MA1 계수는 유의하고 상수의 유의확률은 0.2579로서 크게 유의하지는 않은 것 같다. ARIMA(0,1,1) 모형을 적합한 후의 잔차의 acf와 pacf를 구해보기로 한다.



<그림 4.7> MA(1) 모형 적합 후의 colgate 자료의 ACF와 PACF

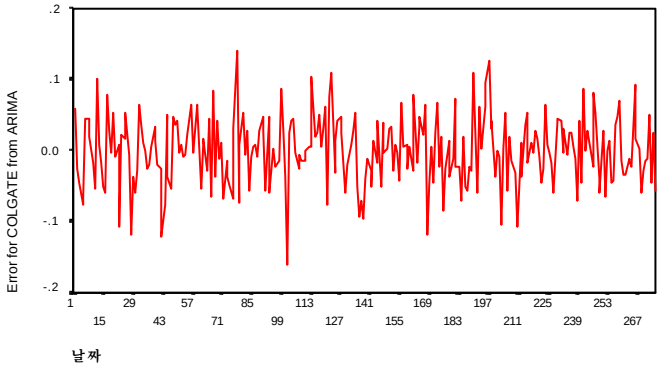
<그림 4.7I>을 보면 acf와 pacf가 모두 신뢰한계안에 있으므로 잔차들이 백색잡음이라고 생각되어 모형이 잘 적합된 것 같다. 잔차의 백색잡음 여부의 검정을 위해서는 텍스트로 출력되는 acf 도표의 오른쪽 끝의 Box-Ljung Prob.를 참고하면 된다. 이 유의확률이 크면 잔차들이 백색잡음이라는 가설을 기각하지 못한다.

Lag	Auto-Corr.	Stand. Err.	-1 -.75 -.5 -.25 0 .25 .5 .75 1	Box-Ljung	Prob.
1	.060	.060	. [*.	1.010	.315
2	-.002	.060	. * .	1.011	.603
3	-.052	.060	.*[.	1.777	.620
4	-.099	.060	**[.	4.521	.340
5	.052	.060	. [*.	5.279	.383
6	.006	.059	. * .	5.291	.507
7	.000	.059	. * .	5.291	.625
8	.072	.059	. [*.	6.785	.560
9	-.009	.059	. * .	6.810	.657
10	-.084	.059	.*[.	7.188	.712
11	-.014	.059	. * .	7.197	.788
12	-.088	.059	**[.	9.208	.686
13	-.088	.059	**[.	11.470	.571
14	-.098	.059	**[.	14.021	.448
15	-.001	.058	. * .	14.021	.524

Plot Symbols: Autocorrelations * Two Standard Error Limits .

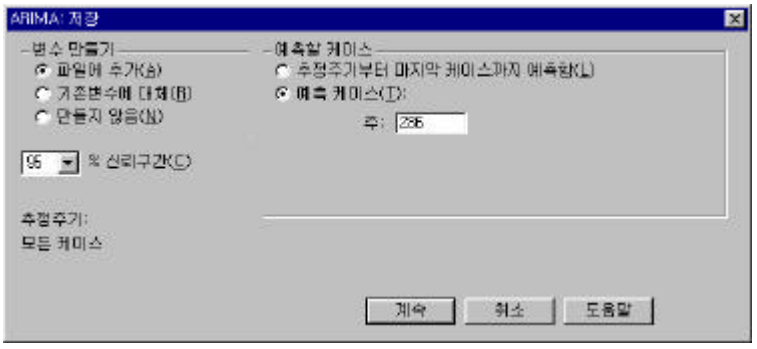
Total cases: 276 Computable first lags: 274

이제 적합단계의 마지막으로 잔차의 시계열그림을 그려보면 <그림 4.8>과 같다. 특별한 경향이 보이지 않으므로 모형이 잘 적합되었다고 생각된다.



<그림 4.8> ARIMA(0,1,1)모형 적합 후의 잔차의 시계열그림

이제 ARIMA(0,1,1)모형을 이용하여 277 시점부터 286 시점까지의 예측값을 구하기로 한다. 저장 단추를 누르면 <그림 4.9> ARIMA: 저장 대화상자가 나타나고 예측할 케이스에서 예측 케이스(T)를 체크하고 286을 입력한 후 이를 수행하면 된다. 예측값 "fit_1", 예측오차 "err_1", 95% 예측하한 "lcl_1", 95% 예측상한 "ucl_1" 등의 변수가 데이터 편집기 창에 저장되어 있음을 확인할 수 있을 것이다.



<그림 4.9> ARIAM : 저장 대화상자

분해법(Decomposition method)과 X11 ARIMA 계절조정(Seasonal Adjustment)

시계열을 분석하는 방법 중에서 가장 오래된 방법으로는 20세기초 경제학자들이 경기변동(business cycle)을 예측하려고 시도한데서 비롯된 분해법(decomposition method)을 들 수 있다. 분해법에서는 시계열이 체계적성분(systematic component)과 불규칙성분(irregular component)으로 구성되어 있으며, 체계적성분은 추세성분(trend component), 계절성분(seasonal component)과 순환성분(cyclical component) 등으로 분해할 수 있다고 생각한다. 이 방법은 이론적으로는 미흡하나 주로 경험에 의존하고 직관적이며 이해하기 쉬워 전통적으로 계절형 시계열을 분석하는데 많이 사용되어 왔으며, 특히 계절성분이 뚜렷한 경우 많이 쓰인다. 분해법은 시계열을 구성하고 있는 성분들이 결정적(deterministic)이며 서로 독립이라는 가정 하

에서 출발하고 있다.

분해법에서 가정하고 있는 시계열변동 성분들은 다음과 같다.

① 불규칙성분(또는 우연성분) : I_t

시간과 관계없이 랜덤한 원인에 의해 나타나는 변동으로 일반적으로 설명하기 어려운 여러 가지 복합적인 원인에 의한 변동을 의미한다. 이 성분은 예측할 수도 없고 관심의 대상도 아니다.

② 추세성분(또는 경향변동, 장기변동) : T_t

시간이 경과함에 따라 증가하거나 감소하는 등의 어떤 추세를 가지고 움직이는 장기적인 변동을 의미하며 보통 직선 또는 이차식 등과 같은 다항식으로 설명할 수 있다. 경우에 따라서는 지수추세, S-곡선 등의 형태도 갖는다.

③ 계절성분 : S_t

기온, 강수량, 월별 효과 등과 같이 1년, 1개월, 1주 등의 일정한 주기를 가지고 규칙적으로 반복되는 변동 성분을 의미하며 보통 1년 이내의 주기적인 변동을 의미한다. 삼각함수 또는 지시함수(indicator function)들의 선형결합을 이용하여 설명할 수 있다.

④ 순환성분 : C_t

계절성분과 유사하게 변화하나 그 변화의 주기가 길 때의 변동을 의미하며 경기변동(business cycle)이라고도 부르기도 한다. 단, 그 주기는 일정하지 않다. 경기나 어떤 기업의 호황과 불황 등을 나타낼 때 사용되기도 하며, 주로 GNP, 건축허가건수, 이자율 등과 같은 시계열에 공통적으로 나타난다고 본다.

분해법의 기본 가정은 시계열이 위에서 설명한 여러 가지 성분들로 구성되어 있다는 것이며 그 결합 방식에 따라 다음과 같이 두 가지 형태의 모형을 가정하고 있다.

① 가법모형(additive model) : $Z_t = T_t + S_t + C_t + I_t$

② 승법모형(multiplicative model) : $Z_t = T_t \cdot S_t \cdot C_t \cdot I_t$

가법모형은 계절성분의 진폭이 시계열의 수준에 관계없이 일정할 때 주로 사용되며, 시계열의 수준에 따라 계절성분의 진폭이 달라질 때는 승법모형을 사용한다. 승법모형의 경우 로그변환을 해 주면 다음과 같은 로그가법모형

$$\ln Z_t = \ln T_t + \ln S_t + \ln C_t + \ln I_t$$

이 되어 결국 가법모형

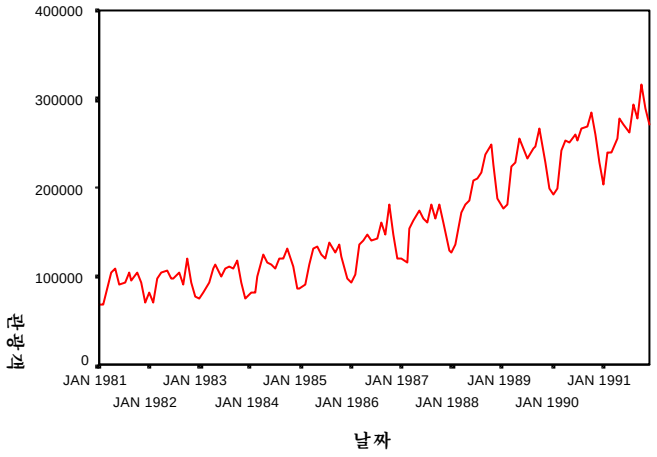
$$Z_t^* = T_t^* + S_t^* + C_t^* + I_t^*$$

과 같이 됨을 알 수 있다. 실제로 많은 경제 관련 자료들이 시간이 흐름에 따라 점차로 일정한 추세를 가지고 증가하며 또한 변동폭도 증가하는 경향이 있는 관계로 주로 승법모형이 사용되거나 로그변환 후 가법모형을 사용한다. 컴퓨터의 사용이 보편화되기 이전에는 계산상의 편이성 때문에 승법모형을 사용하였으나 X11 ARIMA와 같은 계절조정용 프로그램에서는 두 모형을 다 수용하고 있다.

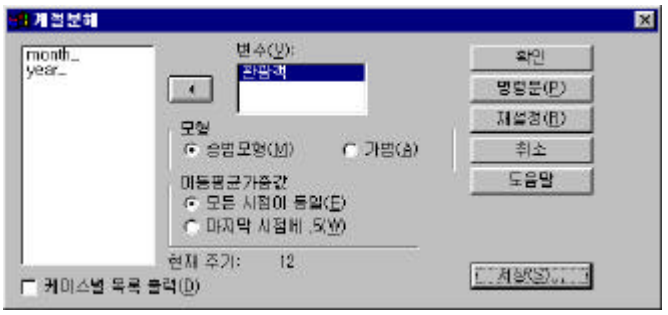
먼저 계절분해(S) 절차를 이용한 분해법에 대해 설명한 후 X11 ARIMA 모형(X) 절차를 이용한 계절조정방법에 대해 설명하기로 한다.

예제 5 한국관광공사에서 조사한 1981년 1월부터 1991년 12월까지 우리나라에 입국한 관광객의 시계열자료를 각 성분별로 분해하는 방법에 대해 설명하기로 한다. 관광객자료의 시계열그림을 그려보면 <그림 5.1>과 같다. 시간이 흐름에 따라 점차로 변동의 폭이 커지는 경향을 보여주므로 승법모형이 적당한 것 같다. 또는 위에서 설명한 것처럼 로그변환을 한 후에 가법모형을 적용하여도 될 것이다. 여기서는 승법모형을 적용하여 보기로 한다. 이를 위해 통계분석(S)-시계열분석(T)-계절분해(S) 절차를 선택하면 <그림 5.2>의 계절분해 대화상

자가 나타난다. 디폴트로 승법모형이 선택된다. 하단의 이동평균가중값에서는 자료의 개수가 짝수이냐 홀수이냐에 따라 다른 형태의 가중값을 줄 수 있다. 디폴트는 짝수를 가정한다. 이 경우에는 $n=12$로 짝수이므로 모든 시점이 동일(E)을 디폴트로 선택한다.



<그림 5.1> 관광객 자료의 시계열그림

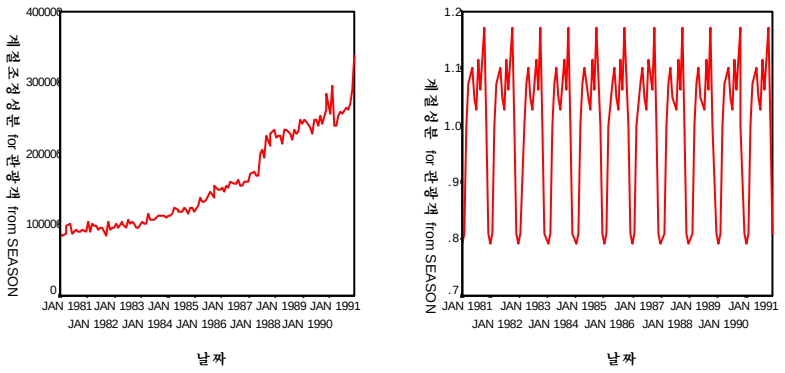


<그림 5.2> 계절분해 대화상자

출력결과는 다음과 같으며 불규칙성분, 계절조정계열, 계절성분 및 추세 · 순환성분이 생성되어 데이터편집기에 저장된다.

```
MODEL:  MOD_1.

Results of SEASON procedure for variable 관광객
Multiplicative Model.  Equal weighted MA method.  Period = 12.

                Seasonal
                   index
      Period     (* 100)
         1        79,010
         2        80,789
         3        99,722
         4       107,405
         5       109,952
         6       104,941
         7       102,795
         8       111,424
         9       106,205
        10       117,146
        11        99,382
        12        80,729

The following new variables are being created:

  Name         Label

  ERR_1        Error for 관광객 from SEASON, MOD_1  MUL EQU 12
  SAS_1        Seas adj ser for 관광객 from SEASON, MOD_1  MUL EQU 12
  SAF_1        Seas factors for 관광객 from SEASON, MOD_1  MUL EQU 12
  STC_1        Trend-cycle for 관광객 from SEASON, MOD_1  MUL EQU 1
```

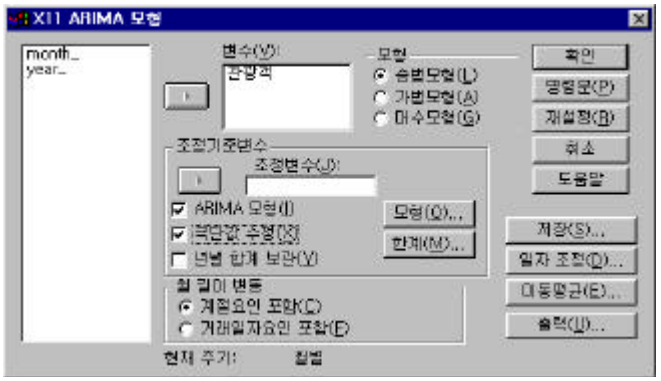
<그림 5.3>은 계절조정계열과 계절성분의 시계열그림이다. 원계열의 시계열그림 <5.1>과 비교해 보면 계절성분이 원계열에서 효과적으로 분해되었음을 알 수 있다.



<그림 5.3> 계절조정계열과 계절성분의 시계열그림

X11 ARIMA법의 계산과정은 사전조정 계산과정, 시계열 구성요인 분해과정 및 통계적 평가과정의 3단계로 구성되어 있다. X11 ARIMA법을 국내의 시계열 자료에 적용하는데 있어서의 문제점으로는 개별 시계열을 잘 나타낸다고 생각되는 ARIMA모형이 자동적으로 적합되고 이를 이용하여 양 끝값들을 미리 예측하여 확장시키는 과정에서 프로그램내에 이미 저장되어 있는 3개의 ARIMA모형만을 고려한다는 점이며, 특히 이 모형들은 캐나다의 174개의 주요 경제 관련 시계열들을 분석하여 가장 잘 적합되는 12개의 모형 중에서 선택된 관계로 국내의 시계열을 분석하는데는 적합하지 않을 수 있다는 점이다. 또한 경계활동일수를 사전에 조정하는 과정에서 미국 또는 캐나다와 우리의 문화차이, 예를들어, 명절효과 또는 토요일 휴무제 등이 다른데서 오는 차이가 고려되지 않았다는 점이다. 이러한 점들을 보완하여 미국 상무부에서는 최근 X12 ARIMA법을 개발하여 공개하였다. X12 ARIMA법에서는 X11 ARIMA법에 RegARIMA라는 예측 및 사전조정과정과 사후진단과정을 추가한 방법이다. X12 ARIMA에 대해서는 조신섭 · 손영숙(1999)을 참고하면 될 것이다.

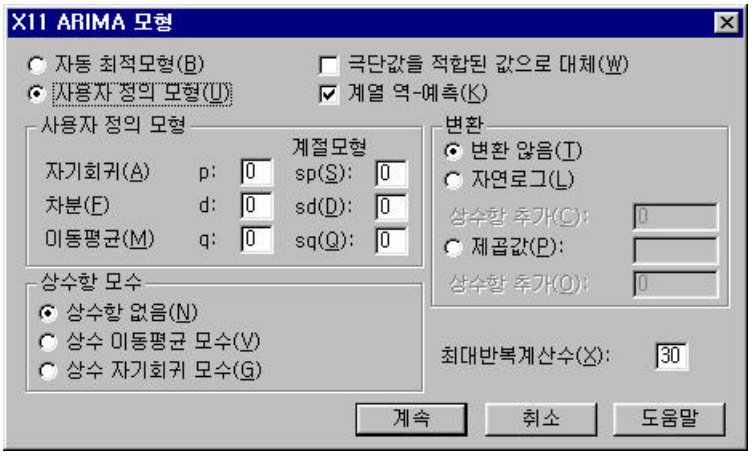
통계분석(S)-시계열분석(I)-X11 ARIMA 모형(X) 절차를 선택하면 <그림 5.4> X11 ARIMA 모형 대화상자가 나타난다. X11 ARIMA법에서는 선택해야할 옵션들이 너무 많으므로 디폴트로 제공되는 가장 기본적인 옵션만을 이용하여 계절조정을 하기로 한다.



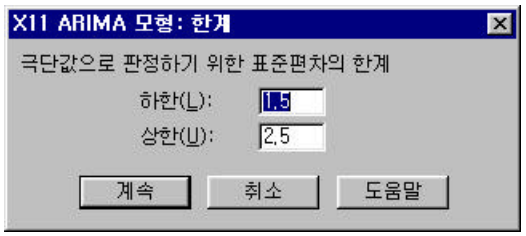
<그림 5.4> X11 ARIMA 모형 대화상자

ARIMA 모형(I)을 체크한 후 모형 단추를 누르면 <그림 5.5>와 같이 X11 ARIMA 모형 대

화상자가 나타난다. 여기서 사용자 정의 모형(U)을 선택한 후 원하는 형태의 모형을 선택한다. 디폴트는 자동 최적모형(B)으로 프로그램에 내장되어 있는 모형 중에서 가장 적합이 잘 되는 모형을 자동으로 선택한다. 극단값 수정(X)을 체크한 후 한계 단추를 누르면 <그림 5.6>과 같이 극단값으로 판정하기 위한 기준을 선택할 수 있다. 계절조정을 시행하면 많은 결과물이 출력된다. 이들 중에서 원하는 결과만을 출력하려면 출력 단추를 눌러서 선택하면 된다.



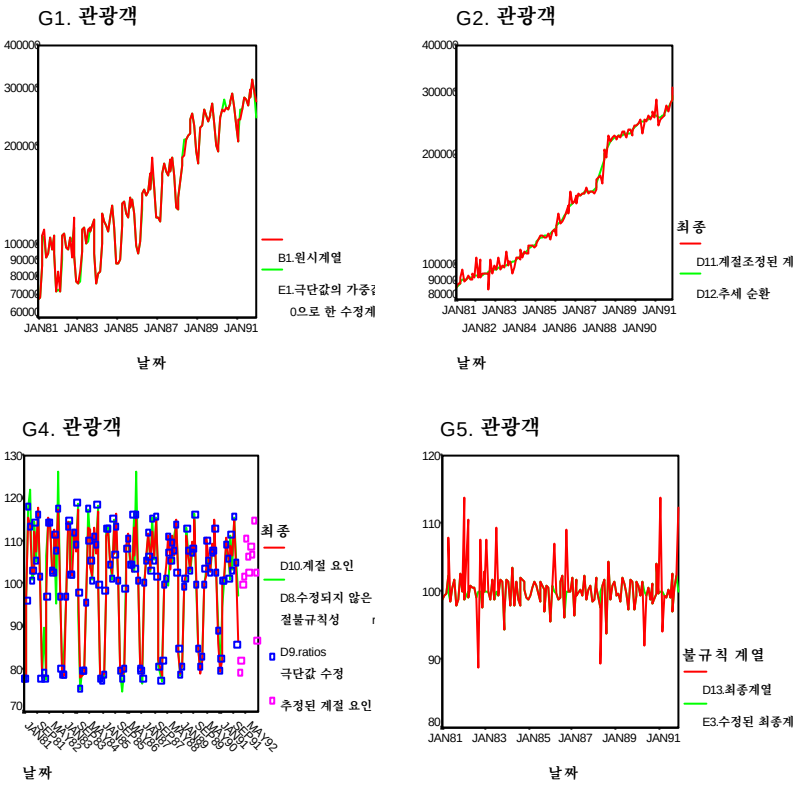
<그림 5.5> ARIMA 모형 선택 대화상자



<그림 5.6> 극단값 판정기준 선택 대화상자

다음은 승법모형을 가정하고 계절조정을 시행한 결과의 출력물 중에서 원계열(G1), 계절조정계열과 추세 · 순환 성분을 겹쳐그린 계열(G2), 계절성분(G4)과 불규칙계열(G5)의 시계열 그림들이다. <그림 5.4>와 비교하면 단순한 분해법에 의한 각 성분들 이외에도 여러 가지가

추가되어 있음을 알 수 있다. 물론 <그림 5.2>에서 저장 단추를 선택하여 나타나는 대화창에서 원하는 계열 또는 성분을 데이터 편집기 창에 출력시킨 후 이를 이용하여 그림을 그려도 동일한 결과를 얻을 수 있다.



<그림 5.7> X11 ARIMA에 의해 출력된 여러 가지 성분의 시계열그림