

시계열 분석

[출생아 수 예측 모형1]

201552001 유승우

목 차

I. 분석 개요	1
II. 시계열 분석	2
1. 데이터 설명 및 전처리	2
2. ETS 모델 분석	3
III. 결론	6
[부록] R 코드	7

I. 분석 개요

한국의 출생율이 바닥을 치고 있다. 정부는 지난 19년간 점진적으로 예산을 늘려가며 약 230조를 투입하고, 정책의 방향성까지 바뀌가며 저출생 문제를 극복하고자 했지만, 여전히 2020년 출생율이 OECD 회원국 중 한국이 꼴찌이며, 평균 0.84명으로 유일하게 0명대를 기록하고 있다.

우리나라의 출생율 하락세는 계속해서 진행 중이므로, 이를 극복하기 위해 출생율 예측으로 앞으로의 정책의 방향성과 예산집행의 기초자료로 활용하기 위해 분석을 실시한다.



최근 19년 출생율 변화¹⁾



최근 저출생·고령화 예산 규모²⁾

1) <http://news.tf.co.kr/read/ptoday/1849187.htm>

2) <https://m.news.zum.com/articles/54972035>

II. 시계열 분석

1. 데이터 설명 및 전처리

KOSIS 국가 통계 포털의 월별 출생아 수³⁾

시군구/성/월별 출생

자료출신일: 2020-08-26 / 수록기간: 월, 년 1997.01 ~ 2019.12 / 자료문자처: 02-2012-9114, 042-4814

일괄선택 + 항목 [1/3] 시군구별 [1/306] 시점 [276/299] ☐ 새창보기 ☐ 주석정보

시점	전국 계 (명)
2019. 12	21,228
2019. 11	23,727
2019. 10	25,613
2019. 09	24,090
2019. 08	24,371
2019. 07	25,222
2019. 06	23,992
2019. 05	25,299
2019. 04	26,104
2019. 03	27,049
2019. 02	25,710
2019. 01	30,271
2018. 12	22,767
2018. 11	25,301
2018. 10	26,474
2018. 09	26,066
2018. 08	27,381
2018. 07	27,033
2018. 06	26,357
2018. 05	27,949
2018. 04	27,734
2018. 03	29,987
2018. 02	27,734
2018. 01	29,987

1997. 01 ~ 2019. 12 데이터를 사용

데이터 전처리

시점	전국
1 시점	계 (명)
2 1997. 01	63268
3 1997. 02	58144
4 1997. 03	62160
5 1997. 04	56949
6 1997. 05	55270
7 1997. 06	50978
8 1997. 07	53369
9 1997. 08	53254
10 1997. 09	53254
11 1997. 10	55461
12 1997. 11	57136
13 1997. 12	53848
14 1998. 01	55557
15 1998. 02	60347
16 1998. 03	56841
17 1998. 04	60238
18 1998. 05	54814
19 1998. 06	52046
20 1998. 07	49571

Showing 1 to 19 of 277 entries, 21

시점	전국
2 1997. 01	63268
3 1997. 02	58144
4 1997. 03	62160
5 1997. 04	56949
6 1997. 05	55270
7 1997. 06	50978
8 1997. 07	53369
9 1997. 08	53254
10 1997. 09	55461
11 1997. 10	57136
12 1997. 11	53848
13 1997. 12	55557
14 1998. 01	60347
15 1998. 02	56841
16 1998. 03	60238
17 1998. 04	54814
18 1998. 05	52046
19 1998. 06	49571
20 1998. 07	50151

Showing 1 to 19 of 276 entries, 21

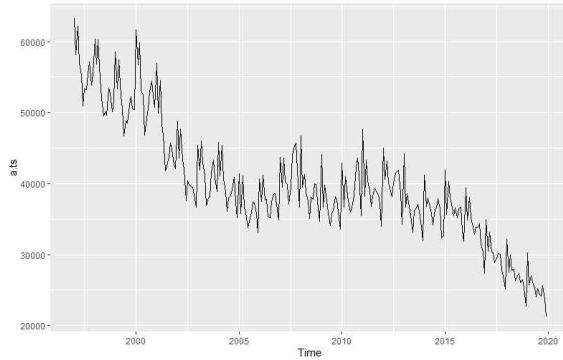
```
> str(a)
'data.frame': 276 obs. of 2 variables:
 $ 시점: chr "1997. 01" "1997. 02" "1997. 03" "1997. 04" ...
 $ 전국: int 63268 58144 62160 56949 55270 50978 53369 53254 55461 57136 ...
> a$time = as.Date(a$시점, start=c(1997,1), freq=12)
> a$time
  Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1997 63268 58144 62160 56949 55270 50978 53369 53254 55461 57136 53848 55557
1998 60347 56841 60238 54814 52046 49571 50151 49634 53484 52709 50164 51595
1999 58607 53305 57495 52386 50516 46765 48780 48502 51056 52224 50559 50473
2000 61644 56723 59878 53058 52492 46774 48377 50623 52913 54429 52403 50775
2001 56949 49939 54549 48309 46802 41809 42404 43830 45673 44814 42753 42103
2002 48825 43579 47753 43453 41977 37575 40361 39751 39615 39458 37921 36643
2003 45423 41909 45972 43040 41132 36879 38008 38054 41645 43345 40752 38877
2004 45778 41059 45422 40611 38933 36086 38031 38324 39238 40855 37432 35189
2005 41354 35810 41052 36605 35983 33725 34658 36003 37413 37283 35737 33084
2006 40692 37522 41209 37666 37740 35347 35154 37457 38471 38556 37110 34835
2007 43704 39485 43578 40290 39913 37112 39522 44041 44871 45722 41865 36719
2008 46747 39495 41318 38783 37598 35035 38042 37787 40037 39886 36473 34691
2009 44149 36681 39834 37539 35864 34122 35779 36169 38187 37737 35172 33616
2010 42936 36639 40975 38336 37005 36028 37106 38392 42294 43625 41318 35517
2011 47577 38160 43260 40376 38958 36766 38168 39360 38963 38359 37328 33990
2012 44984 40609 43209 40094 39388 38170 40127 41455 41735 41889 38579 34311
2013 44247 36645 38540 36738 35626 33154 36188 36400 37063 36055 33827 31972
2014 41229 36754 38021 37183 35749 34171 36390 36539 37879 36450 32379 32691
2015 41914 35709 40229 38072 36534 35520 36612 35207 36444 36702 33467 31910
2016 39405 34830 38131 35147 34341 32849 33920 33897 34375 31592 30366 27390
2017 34834 30499 33196 30337 30303 28892 29418 30135 30085 27857 27068 25147
2018 32198 27575 29987 27734 27949 26357 27033 27381 26066 26474 25301 22767
2019 30271 25710 27049 26104 25299 23992 25222 24371 24090 25613 23727 21228
```

첫 열 삭제 및 두 번째 열 숫자형 변환 후 시계열 데이터 생성

3) https://kosis.kr/statisticsList/statisticsListIndex.do?vwcd=MT_ZTITLE&menuId=M_01_01#content-group

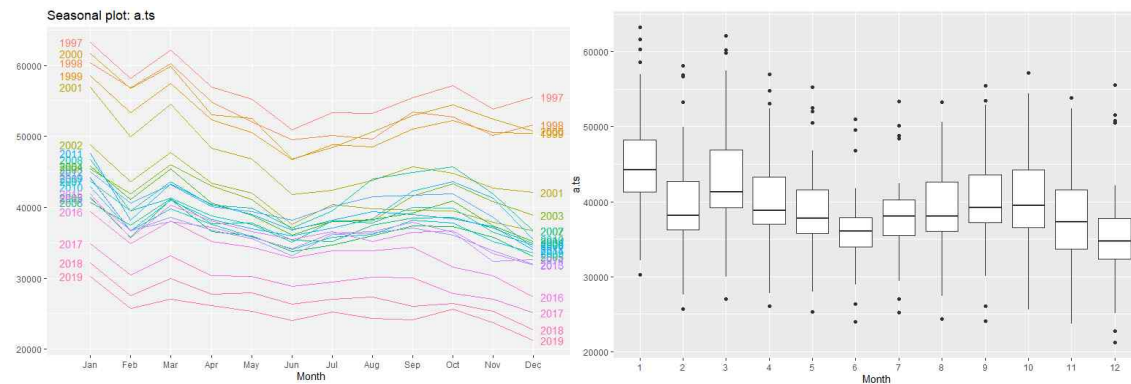
2. ETS 모델 분석

전체적인 출생아 수 분포



출생아 수 분포에서 출생아 수가 감소하는 추세와 계절성이 있는 것으로 보인다.

월별 출생아 수 분포



주로 1~3월의 출생아 수가 많았고, 6~7월, 12월이 적게 나타났다

train, test 분리

```
> train <- window(a.ts, end=c(2017,12))
> test <- window(a.ts, start=c(2018,1))
> train
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1997	63268	58144	62160	56949	55270	50978	53369	53254	55461	57136	53848	55557
1998	60347	56841	60238	54814	52046	49571	50151	49634	53484	52709	50164	51595
1999	58607	53305	57495	52386	50516	46765	48780	48502	51056	52224	50559	50473
2000	61644	56723	59878	53058	52492	46774	48377	50623	52913	54429	52403	50775
2001	56949	49939	54549	48309	46802	41809	42404	43830	45673	44814	42753	42103
2002	48825	43579	47753	43453	41977	37575	40361	39751	39615	39458	37921	36643
2003	45423	41909	45972	43040	41132	36879	38008	38054	41645	43345	40752	38877
2004	45778	41059	45422	40611	38033	36086	38031	38324	39238	40855	37432	35189
2005	41354	35810	41052	36605	35983	33725	34658	36003	37413	37283	35737	33084
2006	40692	37522	41209	37666	37740	35347	35154	37457	38471	38556	37110	34835
2007	43704	39485	43578	40290	39913	37112	39522	44041	44871	45722	41865	36719
2008	46747	39495	41318	38783	37598	35035	38042	37787	40037	39886	36473	34691
2009	44149	36681	39834	37539	35864	34122	35779	36169	38187	37737	35172	33616
2010	42936	36639	40975	38336	37005	36028	37106	38392	42294	43625	41318	35517
2011	47577	38160	43260	40376	38958	36766	38168	39360	38963	38359	37232	33990
2012	44984	40609	43209	40094	39388	38170	40127	41455	41735	41889	38579	34311
2013	44247	36645	38540	36738	35626	33154	36188	36400	37063	36055	33827	31972
2014	41229	36754	38021	37183	35749	34171	36390	36539	37879	36450	32379	32691
2015	41914	35709	40329	38072	36534	35520	36612	35207	36444	36702	33467	31910
2016	39405	34830	38131	35147	34341	32849	33920	33897	34375	31592	30366	27390
2017	34834	30499	33196	30337	30303	28892	29418	30135	30085	27857	27068	25147

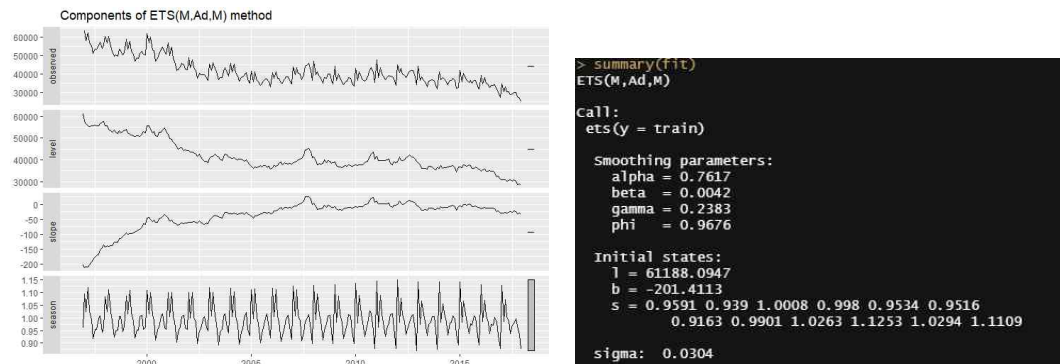
```
> test
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	32198	27575	29987	27734	27949	26357	27033	27381	26066	26474	25301	22767
2019	30271	25710	27049	26104	25299	23992	25222	24371	24090	25613	23727	21228

train 데이터는 1997. 01 ~ 2017. 12 기간의 출생아 수이다.

test 데이터는 2018. 01 ~ 2019. 12 기간의 출생아 수이다.

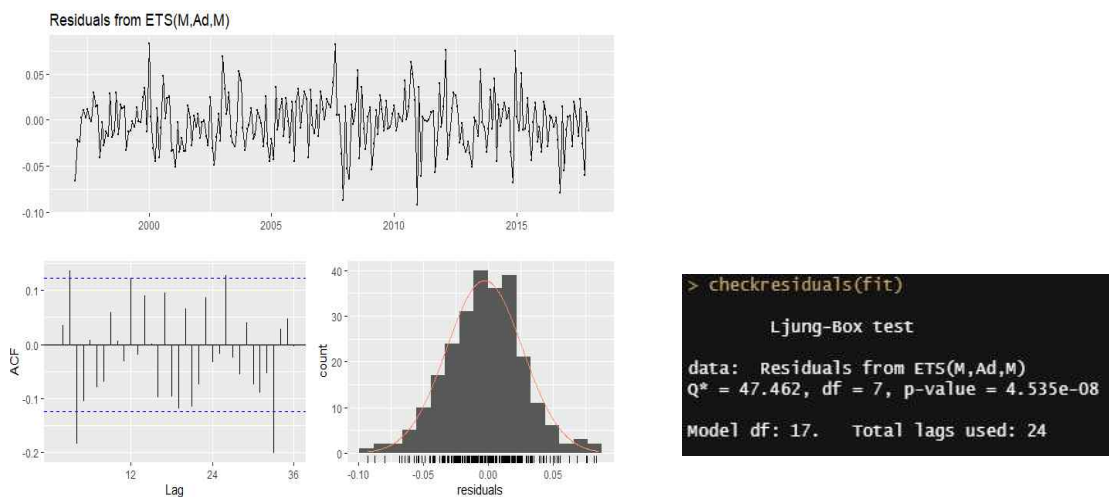
train데이터 ETS 모델 적합



적합 결과 Multiplicative Holt-Winters damped with multiplicative error model이 적합되었다.

alpha가 0.7617로 level에서는 큰 변동이 있고, gamma가 0.2383으로 계절에서는 약간의 변동이 있으며, beta가 0.0042로 기울기는 일정하게 나타났다.

잔차의 독립성 가정 검정



H_0 : 잔차는 독립이다.

H_1 : 잔차는 독립이 아니다.

p-value가 4.535e-08으로 귀무가설을 기각하여 잔차는 독립이 아니라는 충분한 근거가 있다.

검정 결과 분포가 정규성을 띄는 것으로 보이고, 분산도 크지 않게 보이지만, 독립성 가정이 위반되었다. 독립이 아니어도 예측값에는 큰 문제가 없지만, 예측 구간이 좁아질 수 있어 신뢰성이 떨어진다.

예측 결과

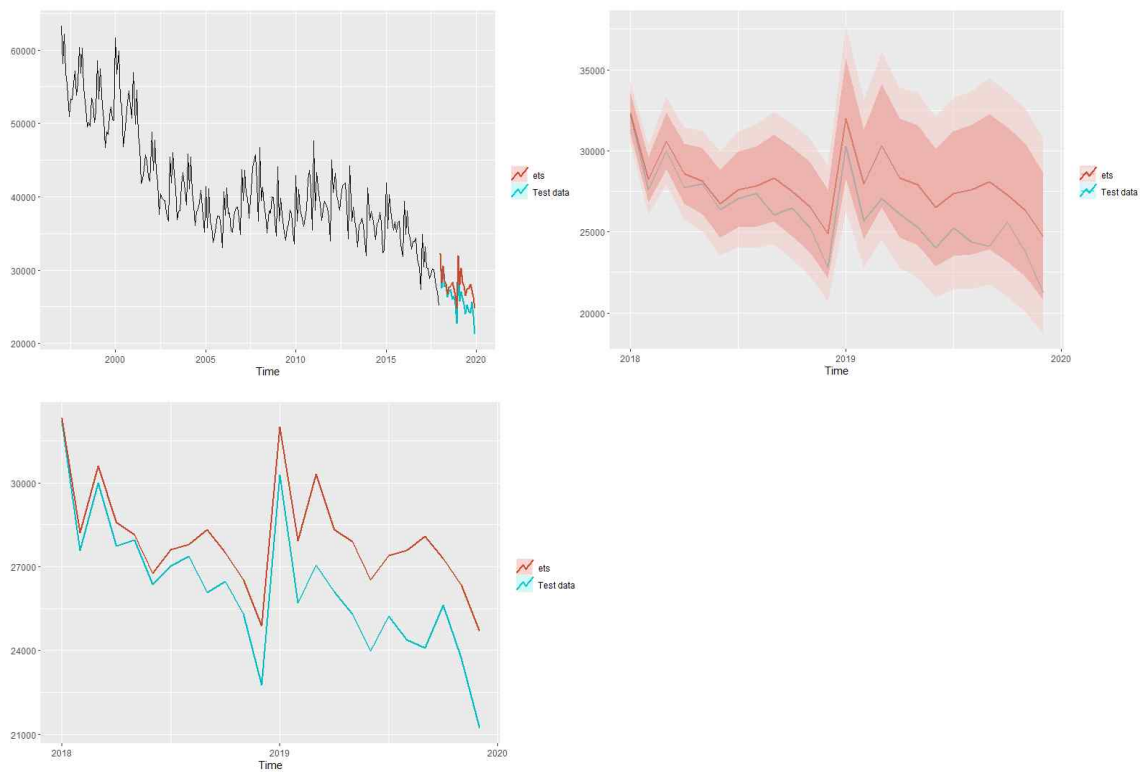
```
> accuracy(fc, test)
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-127.0947	1234.696	936.6066	-0.3492013	2.280223	0.3490140	0.03718538	NA
Test set	-1754.1348	2077.466	1754.1348	-7.0094822	7.009482	0.6536549	0.59952176	0.8657656

```
>
```

모형의 예측 결과 중 RMSE가 2077로 나타났고, MASE가 0.653으로 1보다 작게 나타났다.

예측 결과 그래프



예측이 test data와 약간의 차이가 있지만, 형태가 비슷하게 나타났고, 신뢰구간 안에 포함되어있는 것을 확인할 수 있다.

Ⅲ. 결론

분석 결과 출생아 수가 앞으로도 계속 줄어든 것으로 예측되며, 예측된 출생아 수를 이용하여 예산을 측정해 지원을 늘리는 등 저출생 문제에 대한 대책이 강구되어야 할 필요가 있다고 보인다.

[부록]

R 코드

```
library(tidyverse)
library(forecast)

# 데이터 불러오기
a <- read.csv("C:/Data/출생아 수.csv")
a <- a[-1,]
a$전국 <- as.integer(a$전국)
str(a)
summary(a)

# 시계열데이터 변환 & 분포확인
a.ts <- ts(a$전국,start=c(1997,1), freq=12)
a.ts

autoplot(a.ts)

ggseasonplot(a.ts, year.labels=TRUE,year.labels.left = TRUE)

ggsubseriesplot(a.ts)

data.frame(a.ts=as.numeric(a.ts), mon=as.factor(cycle(a.ts))) %>%
  ggplot() +
  geom_boxplot(aes(x=mon,y=a.ts)) +
  labs(x="Month")

# train & test 나누기
train <- window(a.ts , end=c(2017,12))
test <- window(a.ts, start=c(2018,1))

# ETS 모델 적합 & 가정 검정
fit <- ets(train)
summary(fit)

autoplot(fit)
checkresiduals(fit)
```

```

# 예측
fc <- forecast(fit, h = length(test))
accuracy(fc, test)

autoplot(train) +
  autolayer(test, series="Test data", size=1) +
  autolayer(fc, series="ets", size=1,PI=FALSE) +
  labs(y= NULL, color=NULL)

autoplot(test, series="Test data", size=1) +
  autolayer(fc, series="ets", size=1, alpha=0.5) +
  labs(y= NULL, color=NULL)

autoplot(test, series="Test data", size=1) +
  autolayer(fc, series="ets", size=1, PI = FALSE) +
  labs(y= NULL, color=NULL)

```