

# 출생자 및 시설물 분석

## 학습 목표

- 매년 시설물 데이터를 이용하여 시설물 수 예측을 수행해 본다.
- [경제-건축분야]
- 시설물 데이터 수
- 데이터 출처 :
- 데이터 분석 코드
  - [HTML코드 \(https://ldjwj.github.io/dataAnalysis/01\\_12\\_population\\_analysis.html\)](https://ldjwj.github.io/dataAnalysis/01_12_population_analysis.html)

## 학습 내용

In [ ]:

## 목차

- [01. 데이터 준비 및 라이브러리 임포트](#)
- [02. 출생자 예측 모델 만들기](#)
- [03. 2030년도 출생자수 예측](#)

## 01. 데이터 준비 및 라이브러리 임포트

[목차로 이동하기](#)

In [3]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [11]:

```
dat = pd.read_csv("./data/arch/1926_2021.csv")
birth_dat = pd.read_csv("./data/arch/birth_data_1970_2021.csv")
dat.shape, birth_dat.shape
```

Out[11]:

((97, 11), (52, 7))

In [12]:

```
dat
```

Out[12]:

|     | 준공<br>년<br>도 | 교<br>량 | 터<br>널 | 지하<br>차도 | 인<br>원 | 출생아수<br>(명) | 자연증가<br>건수(명) | 조출생률<br>(천명당) | 자연증가<br>율(천명<br>당) | 합계출<br>산율<br>(명) | 출생<br>성비<br>(명) |
|-----|--------------|--------|--------|----------|--------|-------------|---------------|---------------|--------------------|------------------|-----------------|
| 0   | 1926.0       | 1.0    | 1.0    | 0.0      | 8.0    | NaN         | NaN           | NaN           | NaN                | NaN              | NaN             |
| 1   | 1927.0       | 2.0    | 0.0    | 0.0      | 8.0    | NaN         | NaN           | NaN           | NaN                | NaN              | NaN             |
| 2   | 1928.0       | 1.0    | 0.0    | 0.0      | 4.0    | NaN         | NaN           | NaN           | NaN                | NaN              | NaN             |
| 3   | 1929.0       | 0.0    | 0.0    | 0.0      | 0.0    | NaN         | NaN           | NaN           | NaN                | NaN              | NaN             |
| 4   | 1930.0       | 0.0    | 0.0    | 0.0      | 0.0    | NaN         | NaN           | NaN           | NaN                | NaN              | NaN             |
| ... | ...          | ...    | ...    | ...      | ...    | ...         | ...           | ...           | ...                | ...              | ...             |
| 92  | 2018.0       | 613.0  | 93.0   | 16.0     | 2824.0 | 326822.0    | 28002.0       | 6.4           | 0.5                | 0.977            | 105.4           |
| 93  | 2019.0       | 283.0  | 54.0   | 10.0     | 1348.0 | 302676.0    | 7566.0        | 5.9           | 0.1                | 0.918            | 105.5           |
| 94  | 2020.0       | 215.0  | 83.0   | 8.0      | 1192.0 | 272337.0    | -32611.0      | 5.3           | -0.6               | 0.837            | 104.8           |
| 95  | 2021.0       | 187.0  | 40.0   | 12.0     | 908.0  | NaN         | NaN           | NaN           | NaN                | NaN              | NaN             |
| 96  | NaN          | NaN    | NaN    | NaN      | NaN    | NaN         | NaN           | NaN           | NaN                | NaN              | NaN             |

97 rows × 11 columns

In [13]:

```
birth_dat.head()
```

Out[13]:

|   | 기본항목<br>별 | 출생아수<br>(명) | 자연증가건수<br>(명) | 조출생률(천명<br>당) | 자연증가율(천명<br>당) | 합계출산율<br>(명) | 출생성비<br>(명) |
|---|-----------|-------------|---------------|---------------|----------------|--------------|-------------|
| 0 | 1970      | 1006645     | 748056        | 31.2          | 23.2           | 4.53         | 109.5       |
| 1 | 1971      | 1024773     | 787245        | 31.2          | 23.9           | 4.54         | 109.0       |
| 2 | 1972      | 952780      | 742709        | 28.4          | 22.2           | 4.12         | 109.5       |
| 3 | 1973      | 965521      | 698061        | 28.3          | 20.5           | 4.07         | 104.6       |
| 4 | 1974      | 922823      | 674016        | 26.6          | 19.4           | 3.77         | 109.4       |

2030년도 출생자 수는 어떻게 될까?

In [14]:

```
birth_dat.corr()
```

Out[14]:

|                | 기본항목<br>별 | 출생아수<br>(명) | 자연증가건<br>수(명) | 조출생률(천<br>명당) | 자연증가율<br>(천명당) | 합계출산<br>율(명) | 출생성비<br>(명) |
|----------------|-----------|-------------|---------------|---------------|----------------|--------------|-------------|
| 기본항목별          | 1.000000  | -0.956769   | -0.956755     | -0.956306     | -0.958636      | -0.861605    | -0.395079   |
| 출생아수(명)        | -0.956769 | 1.000000    | 0.996674      | 0.979651      | 0.988320       | 0.897343     | 0.364093    |
| 자연증가건<br>수(명)  | -0.956755 | 0.996674    | 1.000000      | 0.967377      | 0.981922       | 0.873741     | 0.397090    |
| 조출생률(천<br>명당)  | -0.956306 | 0.979651    | 0.967377      | 1.000000      | 0.996787       | 0.961482     | 0.272847    |
| 자연증가율<br>(천명당) | -0.958636 | 0.988320    | 0.981922      | 0.996787      | 1.000000       | 0.946340     | 0.305694    |
| 합계출산율<br>(명)   | -0.861605 | 0.897343    | 0.873741      | 0.961482      | 0.946340       | 1.000000     | 0.107513    |
| 출생성비(명)        | -0.395079 | 0.364093    | 0.397090      | 0.272847      | 0.305694       | 0.107513     | 1.000000    |

In [15]:

birth\_dat

Out[15]:

|    | 기본항목<br>별 | 출생아수<br>(명) | 자연증가건수<br>(명) | 조출생률(천명<br>당) | 자연증가율(천명<br>당) | 합계출산율<br>(명) | 출생성비<br>(명) |
|----|-----------|-------------|---------------|---------------|----------------|--------------|-------------|
| 0  | 1970      | 1006645     | 748056        | 31.2          | 23.2           | 4.530        | 109.5       |
| 1  | 1971      | 1024773     | 787245        | 31.2          | 23.9           | 4.540        | 109.0       |
| 2  | 1972      | 952780      | 742709        | 28.4          | 22.2           | 4.120        | 109.5       |
| 3  | 1973      | 965521      | 698061        | 28.3          | 20.5           | 4.070        | 104.6       |
| 4  | 1974      | 922823      | 674016        | 26.6          | 19.4           | 3.770        | 109.4       |
| 5  | 1975      | 874030      | 603373        | 24.8          | 17.1           | 3.430        | 112.4       |
| 6  | 1976      | 796331      | 529474        | 22.2          | 14.8           | 3.000        | 110.7       |
| 7  | 1977      | 825339      | 576085        | 22.7          | 15.8           | 2.990        | 104.2       |
| 8  | 1978      | 750728      | 498430        | 20.3          | 13.5           | 2.640        | 111.3       |
| 9  | 1979      | 862669      | 622683        | 23.0          | 16.6           | 2.900        | 106.4       |
| 10 | 1980      | 862835      | 585551        | 22.6          | 15.4           | 2.820        | 105.3       |
| 11 | 1981      | 867409      | 629928        | 22.4          | 16.3           | 2.570        | 107.1       |
| 12 | 1982      | 848312      | 602545        | 21.6          | 15.3           | 2.390        | 106.8       |
| 13 | 1983      | 769155      | 514592        | 19.3          | 12.9           | 2.060        | 107.3       |
| 14 | 1984      | 674793      | 438348        | 16.7          | 10.8           | 1.740        | 108.3       |
| 15 | 1985      | 655489      | 415071        | 16.1          | 10.2           | 1.660        | 109.4       |
| 16 | 1986      | 636019      | 396763        | 15.4          | 9.6            | 1.580        | 111.7       |
| 17 | 1987      | 623831      | 380327        | 15.0          | 9.1            | 1.530        | 108.8       |
| 18 | 1988      | 633092      | 397313        | 15.1          | 9.5            | 1.550        | 113.2       |
| 19 | 1989      | 639431      | 402613        | 15.1          | 9.5            | 1.560        | 111.8       |
| 20 | 1990      | 649738      | 408122        | 15.2          | 9.5            | 1.570        | 116.5       |
| 21 | 1991      | 709275      | 467005        | 16.4          | 10.8           | 1.710        | 112.4       |
| 22 | 1992      | 730678      | 494516        | 16.7          | 11.3           | 1.760        | 113.6       |
| 23 | 1993      | 715826      | 481569        | 16.0          | 10.8           | 1.654        | 115.3       |
| 24 | 1994      | 721185      | 478746        | 16.0          | 10.6           | 1.656        | 115.2       |
| 25 | 1995      | 715020      | 472182        | 15.7          | 10.3           | 1.634        | 113.2       |
| 26 | 1996      | 691226      | 450077        | 15.0          | 9.8            | 1.574        | 111.5       |
| 27 | 1997      | 675394      | 430701        | 14.5          | 9.3            | 1.537        | 108.2       |
| 28 | 1998      | 641594      | 395769        | 13.7          | 8.4            | 1.464        | 110.1       |
| 29 | 1999      | 620668      | 372934        | 13.2          | 7.9            | 1.425        | 109.5       |
| 30 | 2000      | 640089      | 391349        | 13.5          | 8.2            | 1.480        | 110.1       |
| 31 | 2001      | 559934      | 316121        | 11.7          | 6.6            | 1.309        | 109.0       |
| 32 | 2002      | 496911      | 249387        | 10.3          | 5.2            | 1.178        | 109.9       |
| 33 | 2003      | 495036      | 248573        | 10.2          | 5.1            | 1.191        | 108.6       |

|    | 기본항목<br>별 | 출생아수<br>(명) | 자연증가건수<br>(명) | 조출생률(천명<br>당) | 자연증가율(천명<br>당) | 합계출산율<br>(명) | 출생성비<br>(명) |
|----|-----------|-------------|---------------|---------------|----------------|--------------|-------------|
| 34 | 2004      | 476958      | 230738        | 9.8           | 4.8            | 1.164        | 108.2       |
| 35 | 2005      | 438707      | 192833        | 9.0           | 4.0            | 1.085        | 107.8       |
| 36 | 2006      | 451759      | 207597        | 9.2           | 4.2            | 1.132        | 107.6       |
| 37 | 2007      | 496822      | 250340        | 10.1          | 5.1            | 1.259        | 106.2       |
| 38 | 2008      | 465892      | 219779        | 9.4           | 4.4            | 1.192        | 106.4       |
| 39 | 2009      | 444849      | 197907        | 9.0           | 4.0            | 1.149        | 106.4       |
| 40 | 2010      | 470171      | 214766        | 9.4           | 4.3            | 1.226        | 106.9       |
| 41 | 2011      | 471265      | 213869        | 9.4           | 4.3            | 1.244        | 105.7       |
| 42 | 2012      | 484550      | 217329        | 9.6           | 4.3            | 1.297        | 105.7       |
| 43 | 2013      | 436455      | 170198        | 8.6           | 3.4            | 1.187        | 105.3       |
| 44 | 2014      | 435435      | 167743        | 8.6           | 3.3            | 1.205        | 105.3       |
| 45 | 2015      | 438420      | 162525        | 8.6           | 3.2            | 1.239        | 105.3       |
| 46 | 2016      | 406243      | 125416        | 7.9           | 2.5            | 1.172        | 105.0       |
| 47 | 2017      | 357771      | 72237         | 7.0           | 1.4            | 1.052        | 106.3       |
| 48 | 2018      | 326822      | 28002         | 6.4           | 0.5            | 0.977        | 105.4       |
| 49 | 2019      | 302676      | 7566          | 5.9           | 0.1            | 0.918        | 105.5       |
| 50 | 2020      | 272337      | -32611        | 5.3           | -0.6           | 0.837        | 104.8       |
| 51 | 2021      | 260562      | -57200        | 5.1           | -1.1           | 0.808        | 105.1       |

- 방법1 : 년도와 출생자명수 데이터를 이용한 선형 모델을 구해서 2030년 출생자수를 예측하자.

In [16]:

```
birth_dat.columns
```

Out[16]:

```
Index(['기본항목별', '출생아수(명)', '자연증가건수(명)', '조출생률(천명당)', '자연증가율(천명당)', '합계출산율(명)', '출생성비(명)'],
      dtype='object')
```

In [20]:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [23]:

```
birth_dat.shape
```

Out[23]:

```
(52, 7)
```

In [41]:

```
# 데이터 나누기
sel = ['기본항목별']
X = birth_dat[sel]
y = birth_dat['출생아수(명)']

X_train = X[0:40]
X_test = X[41:]
y_train = y[0:40]
y_test = y[41:]
```

In [42]:

X\_test

Out[42]:

| 기본항목별 |      |
|-------|------|
| 41    | 2011 |
| 42    | 2012 |
| 43    | 2013 |
| 44    | 2014 |
| 45    | 2015 |
| 46    | 2016 |
| 47    | 2017 |
| 48    | 2018 |
| 49    | 2019 |
| 50    | 2020 |
| 51    | 2021 |

In [43]:

```
# 모델 생성 및 학습
model = LinearRegression()
model.fit(X_train, y_train)
```

Out[43]:

LinearRegression()

**2011~2022년까지 출생자 수, 예측 후, 평가해 보기**

In [44]:

```
pred = model.predict(X_test)
pred
```

Out[44]:

```
array([424196.94896811, 411334.52101313, 398472.09305816, 385609.66510319,
       372747.23714822, 359884.80919325, 347022.38123827, 334159.9532833 ,
       321297.52532833, 308435.09737336, 295572.66941838])
```

In [45]:

```
from sklearn.metrics import mean_absolute_error
mean_absolute_error(pred, y_test)
```

Out[45]:

```
38903.59906191277
```

- 약 전체 오차가 38903명정도 차이가 난다.

In [46]:

```
X_test
```

Out[46]:

| 기본항목별 |      |
|-------|------|
| 41    | 2011 |
| 42    | 2012 |
| 43    | 2013 |
| 44    | 2014 |
| 45    | 2015 |
| 46    | 2016 |
| 47    | 2017 |
| 48    | 2018 |
| 49    | 2019 |
| 50    | 2020 |
| 51    | 2021 |

### 03. 2030년도 출생자수 예측

[목차로 이동하기](#)

In [47]:

```
### 2030년도 예측수행해 보기
pred_2030 = model.predict([[2030]])
pred_2030
```

```
C:\Users\Wtotofriend\Anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X
does not have valid feature names, but LinearRegression was fitted with feature name
s
  warnings.warn(
```

Out[47]:

```
array([179810.81782364])
```

- 결과 : 2030년도의 출생자수는 179810명으로 예상된다.