

미국의 연도별 인구수 변화

년 도 (x)	1900	1905	1910	1915	1920	1925	1930	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980	1985	1990
인 구 수 (y)	76.1	83.8	92.4	100.5	106.5	115.8	123.1	127.3	132.5	133.4	151.9	165.1	180.0	193.5	204.0	215.5	227.2	237.9	249.0

In [10]:

연습문제 7 p343, node (15)

데이터를 보고 물음에 답하라!

[조건]

1. 회귀직선식과 상관계수를 구하여라

2. 비선형 모델을 찾기 위해, log(인구수) 대 연도(1900~)을 그리고 회귀직선과 상관계수를 구하여라

3. 각각의 모델을 이용해 2100년의 인구수를 예측하고 인구수가 300(millions)에 이르는 시기를 예측하라

4. (3)의 정답에 대한 정확성을 논평하라(???)

```
import numpy as np
```

```
from sklearn.linear_model import LinearRegression
```

```
X = np.array([1900, 1905, 1910, 1915, 1920, 1925, 1930, 1935, 1940, 1945, 1950, 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990]).reshape(1, -1)
```

```
Y = np.array([76.1, 83.8, 92.4, 100.5, 106.5, 115.8, 123.1, 127.3, 132.5, 133.4, 151.9, 165.1, 180.0, 193.5, 204.0, 215.5, 227.2, 237.9, 249.0])
```

```
log_Y = np.log(Y)
```

```
model = LinearRegression()
```

```
model.fit(X, Y)
```

```
print("1. 회귀직선식과 상관계수를 구하여라")
```

```
print("Intercept: ", round((model.intercept_), 4))
```

```
print("Coefficient: ", round((model.coef_[0]), 4))
```

```
correlation_coefficient = np.corrcoef(X.reshape(1, -1), Y)[0][1]
```

```
print("상관계수 : ", round((correlation_coefficient), 4))
```

```
population_2100_modelA = model.predict(np.array([[2100]]))[0]
```

```
year_modelA = (300 - model.intercept_) / model.coef_[0]
```

```
model.fit(X, log_Y)
```

```
print("\n2. 비선형 모델을 찾기 위해, log(인구수) 대 연도(1900~)을 그리고 회귀직선과 상관계수를 구하여라")
```

```
print("Intercept: ", round((model.intercept_), 4))
```

```
print("Coefficient: ", round((model.coef_[0]), 4))
```

```
correlation_coefficient = np.corrcoef(X.reshape(1, -1), Y)[0][1]
```

```
print("상관계수 : ", round((correlation_coefficient), 4))
```

```
population_2100_modelB = model.predict(np.array([[2100]]))[0]
```

```
year_modelB = (300 - model.intercept_) / model.coef_[0]
```

```
print("\n3. 각각의 모델을 이용해 2100년의 인구수를 예측하고 인구수가 300(millions)에 이르는 시기를 예측하라")
```

```
print(f"2100년의 인구수 예측 (모델A) {population_2100_modelA}")
```

```
print(f"300 million에 인구수가 도달하는 시점 (모델A) : {year_modelA}\n")
```

```
print(f"2100년의 인구수 예측 (모델B) {population_2100_modelB}")
```

```
print(f"300 million에 인구수가 도달하는 시점 (모델B) : {year_modelB}")
```

1. 회귀직선식과 상관계수를 구하여라

Intercept: -3593.284

Coefficient: 1.9264

상관계수 : 0.989

2. 비선형 모델을 찾기 위해, $\log(\text{인구수})$ 대 $\text{년도}(1900\sim)$ 을 그리고 회귀직선과 상관계수를 구하여라

Intercept: -20.2833

Coefficient: 0.013

상관계수 : 0.989

3. 각각의 모델을 이용해 2100년의 인구수를 예측하고 인구수가 300(millions)에 이르는 시기를 예측하라

2100년의 인구수 예측 (모델 A) 452.0528070175437

300 millian에 인구수가 도달하는 시점 (모델 A) : 2021.0669204568223

2100년의 인구수 예측 (모델 B) 6.983727241844431

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js