

# 다중회귀분석을 활용한 공공자전거 이용에 날씨가 미치는 영향 분석

도시데이터코딩 7팀: 김상우, 정홍, 제태성, 이강신



## 1. 연구 배경 및 목적





## 2. 연구 방법

### ● 연구절차

공공데이터 구득

다중회귀분석을  
위한  
테이블 구축

다중회귀분석을  
통한  
영향 분석

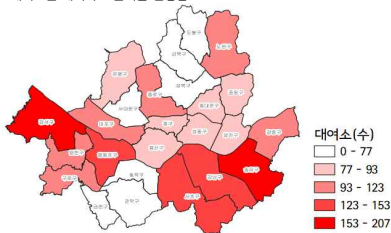


## 2. 연구 방법

## 분석자료

## 공공자전거 이용정보 데이터

- ◆ 서울 열린데이터 광장 - 2021년 서울시 공공자전거 이용정보(일별)
- ◆ 대여일자, 대여소번호, 대여소명, 성별, 연령대코드 등의 자료가 포함
- ◆ 대여 일자에 대한 이용량을 다중회귀분석의 종속변수(Y)로 사용되며, 종속변수에 log를 변환함
- ◆ 분석은 2021년의 대여소를 기준으로 진행하며 2021년 이후 신설된 대여소는 제외하고 분석을 진행함



## 기상정보 데이터

- ◆ 일시, 평균기온(°C), 최저기온(°C), 최저기온 시각(hhmm), 최고기온(°C), 최고기온 시각(hhmm), 일강수량(mm), 최대 순간 풍속(m/s), 최대 순간풍속 시각(hhmm), 평균 풍속(m/s), 최대 순간 풍속 풍향(deg)
- ◆ 공공자전거 데이터의 대여 일자와 매칭하여 **평균기온(°C), 최저기온(°C), 최고기온(°C), 일강수량(mm), 최대 순간 풍속(m/s), 평균 풍속(m/s)**이 독립변수(x)로 사용되며 계산된 지표들의 단위 등이 서로 상이하고 다양하므로 지표 측정의 단위를 표준화하여 그 값을 사용함
- ◆ 표준화된 지표값은 -2에서 2 사이에서 변화하며 그 식은 아래와 같음.

$$X = \frac{(\text{해당지표값} - \text{해당지표의 평균값})}{\text{해당지표의 표준편차}}$$



## 2. 연구 방법

## 다중 회귀 분석

## 다중회귀모형

- ◆ 회귀분석이란 종속변수와 독립변수 간의 함수적 관계를 파악하는 통계적 기법을 말하며, 회귀분석은 회귀모형 구축에 의한 자료의 기술과 모수추정, **적합된 모형을 통한 예측을 목적으로 함.**
- ◆ 본 연구에서는 **6가지 기상 데이터**에 따른 공공자전거의 이용량을 예측하는 것을 목적으로 하기 때문에 **다중회귀 분석**을 적용하였다.
- ◆ 식 (1)은 자료의 개수가  $n$ 개일 때, 독립변수  $x_1, x_2, x_3$ 를 이용하여 종속변수  $y$ 를 추정하기 위한 다중 회귀모형을 나타낸 것이다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

## 최소제곱법

- ◆ 여기서  $\beta_0, \beta_1, \beta_2, \beta_3$ 는 회귀계수를 나타내며,  $\varepsilon_i$ 는 오차항을 나타낸다. **회귀계수를 추정하기 위해서** 주로 최소제곱법(least square method)의 원리를 적용함
- ◆ 최소제곱법은 식 (1)을 회귀모형으로 가정하였을 때, 회귀모형에 의한 모든 독립변수와 종속변수 간의 관계에서 나타나는 **오차항  $\varepsilon_i$ 의 제곱의 합, 즉, 식 (2)의 값이 최소가 되도록 하는 회귀계수를 추정하는 방법임**

$$\sum \varepsilon_i^2 = [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i})]^2, \quad i = 1, 2, \dots, n \quad (2)$$

- ◆ 본 연구에서는 기상 데이터와 공공자전거 이용량 간의 상관관계를 예측하기 위하여 **다중선형회귀분석**을 적용함
- ◆ **다중선형회귀분석**은 종속변수와 각각의 독립변수를 설명하는 회귀계수가 선형을 이루고 있는 모형에 대한 분석을 말하며, 독립변수와 종속변수의 관계가 1차 식으로 구성된 다중선형회귀모형을 적용할 경우 각각의 **회귀계수는 종속변수를 예측하기 위한 독립변수의 영향력**을 의미함



## 3. 연구 범위



### 공간적 범위

- ◆ 본 연구는 서울시의 공공자전거 따릉이의 이용량과 기상 데이터와의 상관관계를 분석하는 연구이고 각 대여소별 대여량을 시군구 단위로 합산하여 분석을 진행하기 때문에 연구의 공간적 범위는 **서울특별시의 자치구** 단위로 한다.



### 시간적 범위

- ◆ 본 연구의 시간적 범위는 서울특별시의 공공자전거인 따릉이 서비스가 안정적으로 도입되기 시작한 2021년을 시간적 범위로 한다.

2021



## 4. 연구 과정

# Colab Notebooks



## 5. 결과 분석



날씨(기후) 변수의 기초 통계량

	평균기온(°C)	최저기온(°C)	최고기온(°C)	일강수량(mm)	최대순간풍속(m/s)	평균풍속(m/s)
count	8771.000000	8771.000000	8771.000000	8771.000000	8771.000000	8771.000000
mean	13.725698	9.425356	18.740531	3.071258	7.100878	1.616771
std	10.398509	10.769162	10.532885	10.067497	2.608120	0.730427
min	-17.100000	-22.600000	-12.700000	0.000000	1.900000	0.200000
25%	6.400000	1.700000	11.200000	0.000000	5.300000	1.100000
50%	14.000000	9.400000	19.400000	0.000000	6.600000	1.500000
75%	22.700000	19.000000	27.700000	0.500000	8.400000	2.000000
max	33.000000	29.500000	39.200000	104.500000	45.000000	6.800000





## 5. 결과 분석

## 날씨(기후) 변수의 분석통계량

OLS Regression Results

Dep. Variable:	log_rental(Y)	R-squared:	0.435
Model:	OLS	Adj. R-squared:	0.435
Method:	Least Squares	F-statistic:	1126.
Date:	Thu, 09 Jun 2022	Prob (F-statistic):	0.00
Time:	17:17:00	Log-Likelihood:	-8795.7
No. Observations:	8771	AIC:	1.761e+04
Df Residuals:	8764	BIC:	1.766e+04
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	7.8329	0.007	1111.707	0.000	7.819	7.847
평균기온	0.1932	0.121	1.595	0.111	-0.044	0.431
최저기온	-0.1943	0.067	-2.893	0.004	-0.326	-0.063
최고기온	0.4930	0.065	7.570	0.000	0.365	0.621
일강수량	-0.2880	0.007	-39.204	0.000	-0.302	-0.274
최대순간풍속	-0.1225	0.010	-11.856	0.000	-0.143	-0.102
평균풍속	0.0528	0.010	5.070	0.000	0.032	0.073

Omnibus:	4379.977	Durbin-Watson:	0.790
Prob(Omnibus):	0.000	Jarque-Bera (JB):	134939.660
Skew:	-1.796	Prob(JB):	0.00
Kurtosis:	21.877	Cond. No.	36.9

- ◆ 날씨가 전체 자전거 이용에 미치는 영향을 분석한 결과 온도 등의 날씨 변수가 자전거 이용에 영향을 미치는 것으로 나타났다. 모형의 수정된 R제곱값은 0.435로, 분석한 변수들이 약 43.5% 설명력을 가짐을 알 수 있었다.
- ◆ 기온의 경우 평균기온과 최저기온 최고기온으로 나누어 확인하였는데, 평균기온 변수는 p값이 0.111으로 채택되지 않았다. 이외의 최저기온과 최고기온은 매우 유의한 변수로 채택되었다.  
최저기온의 경우 음의 값을 가지는데, 이는 최저온도 내에서 비교적 온도가 덜 낮으면 자전거 이용이 증가하고 온도가 더 낮으면 이용이 감소하는 것으로 해석할 수 있다. 최고기온은 날씨 변수 중 자전거 이용수에 가장 큰 영향을 미치는 변수임을 확인했다.
- ◆ 상관계수를 해석해보면 최저기온과 반대로 비교적 온도가 낮아지면 자전거 이용수가 감소하며 온도가 높아지면 자전거 이용수가 증가하는 것으로 해석할 수 있었다.
- ◆ 강수량은 음의 값으로 나타났는데, 이는 비가 많이 오면 자전거 이용이 감소하고 비가 적게 오면 자전거 이용이 증가함을 뜻한다. 풍속의 경우에는 평균풍속과 최대순간풍속의 상관계수가 다른 영향을 미침을 알 수 있다. 먼저 최대순간풍속은 음의 값이 나왔는데, 이것은 강한 돌풍 등의 바람이 불면 자전거 이용이 감소한다는 것으로 해석할 수 있다.
- ◆ 평균풍속은 양의 값으로 나타났으며 이것은 평균 바람의 세기가 강해질수록 자전거 이용 수가 증가하는 것으로 이해할 수 있었다.



## 6. 결론

본 연구에서 도출한 결과는 날씨가 자전거 이용에 영향을 미친다는 것을 보여주고 있다. 결과적으로 온도와 강수량, 그리고 바람(풍속)이 영향을 미치는 것으로 나타났다. 그러나 겨울기간에 대한 적설량 변수가 분석대상에 포함되지 않아 향후 이를 포함한 분석도 수행되어야 공공자전거 이용 수와 날씨의 관계를 더 명확히 볼 수 있을 것이다. 본 연구결과는 앞으로 자전거 정책을 수립하고 시행하는데 기초적인 자료가 될 수 있을 것으로 판단된다.



감사합니다