

No.    date    /    /

title

# 혼자 공부 용어 노트


혼자 공부하는 데이터 분석 with 파이썬

혼자 공부하며 함께 만드는

---

혼공 용어 노트

# 목차

## 가나다 순

가설검정 statistical hypothesis test	21	머신러닝 machine learning	07
객체지향 API object-oriented API	18	명목형 데이터 nominal data	17
결정계수 coefficient of determination	23	모델 model	21
구글 코랩 Google Colab	07	모수검정 parametric test	20
기술통계 descriptive statistics	12	배열 array	08
깃허브 GitHub	07	분산 variance	13
넘파이 NumPy	07	분위수 quantile	12
데이터 과학 data science	06	불리언 배열 boolean array	11
데이터 마이닝 data mining	06	뷰티플수프 Beautiful Soup	10
데이터베이스 DB; database	06	비지도 학습 unsupervised learning	22
데이터 분석 data analysis	06	사이킷런 scikit-learn	21
데이터 정제 data cleaning	11	산점도 scatter plot	14
데이터프레임 DataFrame	08	상자 수염 그림 box-and-whisker plot	15
도수분포표 frequency table	14	서브플롯 subplot	17
딕셔너리 dictionary	09	선 그래프 line chart	17
람다 lambda	11	선형 회귀 linear regression	23
로그 스케일 log scale	15	순열검정 permutation test	21
로지스틱 회귀 logistic regression	23	스택 막대 그래프 stacked bar chart	19
리스트 list	09	스택 영역 그래프 stacked area chart	18
마커 marker	16	시리즈 series	08
막대 그래프 bar chart	17	신뢰구간 confidence interval	20
맷플롯립 matplotlib	14	역직렬화 deserialization	10

---

원 그래프 pie chart	19	피벗 테이블 pivot table	19
원소별 비교 element-wise comparison	11	훈련 training	22
웹 스크래핑 web scraping	10	히스토그램 histogram	14
이상치 outlier	16		
정규 표현식 regular expression	12		
정규분포 normal distribution	15		
중심극한정리 central limit theorem	20		
지도 학습 supervised learning	22		
중앙값 median	12		
직렬화 serialization	10		
최빈값 mode	13		
축 Axis	16		
컬러 막대 color bar	18		
컬러맵 color map	18		
판다스 pandas	07		
패키지 package	07		
평균 mean	12		
표본집단 sample	13		
표준점수 standard score	20		
표준정규분포 standard normal distribution	15		
표준편차 standard deviation	13		
피겨 figure	16		

# 목차

## ABC 순

API; Application Programming Interface		dictionary 딕셔너리	09
응용 프로그래밍 인터페이스	09	element-wise comparison 원소별 비교	11
array 배열	08	figure 피겨	16
Axis 축	16	frequency table 도수분포표	14
bar chart 막대 그래프	17	GitHub 깃허브	07
Beautiful Soup 뷰티풀수프	10	Google Colab 구글 코랩	07
boolean array 불리언 배열	11	histogram 히스토그램	14
box-and-whisker plot 상자 수염 그림	15	HTTP GET HTTP GET 방식	10
central limit theorem 중심극한정리	20	HTML; Hypertext Markup Language	09
coefficient of determination 결정계수	23	HTTP; HyperText Transfer Protocol	09
color bar 컬러 막대	18	JSON; JavaScript Object Notation	09
color map 컬러맵	18	lambda 람다	11
CSV; comma-separated values	07	linear regression 선형 회귀	23
confidence interval 신뢰구간	20	line chart 선 그래프	17
data analysis 데이터 분석	06	list 리스트	09
data cleaning 데이터 정제	11	log scale 로그 스케일	15
data mining 데이터 마이닝	06	logistic regression 로지스틱 회귀	23
data science 데이터 과학	06	machine learning 머신러닝	07
DataFrame 데이터프레임	08	marker 마커	16
DB; database 데이터베이스	06	matplotlib 맷플롯립	14
descriptive statistics 기술통계	12	mean 평균	12
deserialization 역직렬화	10	median 중앙값	12

mode 최빈값	13	series 시리즈	08
model 모델	21	stacked area chart 스택 영역 그래프	18
nominal data 명목형 데이터	17	stacked bar chart 스택 막대 그래프	19
normal distribution 정규분포	15	standard deviation 표준편차	13
NaN; Not A Number	11	standard normal distribution 표준정규분포	15
NumPy 넘파이	07	standard score 표준점수	20
object-oriented API 객체지향 API	18	statistical hypothesis test 가설검정	21
outlier 이상치	16	SQL; Structured Query Language	06
package 패키지	07	subplot 서브플롯	17
pandas 판다스	07	supervised learning 지도 학습	22
parametric test 모수검정	20	training 훈련	21
permutation test 순열검정	21	unsupervised learning 비지도 학습	22
pie chart 원 그래프	19	variance 분산	13
pivot table 피벗 테이블	19	web scraping 웹 스크래핑	10
quantile 분위수	12	XML; eXtensible Markup Language	10
regular expression 정규 표현식	12		
rcParams;			
Runtime Configuration Parameters	16		
sample 표본집단	13		
scatter plot 산점도	14		
scikit-learn 사이킷런	21		
serialization 직렬화	10		

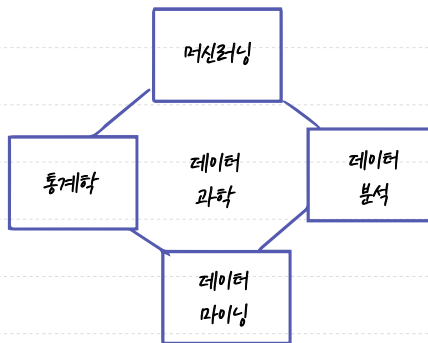
# 01 장 <sup>✓</sup> 데이터 분석을 시작하며

□ 데이터 분석      **data analysis**      [01장 025쪽]

데이터를 수집, 처리, 정제, 분석, 모델링 하여 의사 결정을 내리는 데 도움을 주는 작업. 통계학과 머신러닝의 기술을 사용하고 비즈니스 문제를 해결하기 위해 도메인 지식이 필요하다.

□ 데이터 과학      **data science**      [01장 025쪽]

데이터 분석, 머신러닝을 아우르는 포괄적인 개념. 데이터 분석에 비해 머신러닝 모델을 만들어 문제를 해결하는 것에 비중이 많다.



□ 데이터베이스      **DB; database**      [01장 029쪽]

데이터를 여러 사람이 공유하며 사용하기 위해 통합 관리하는 데이터의 집합.

□ SQL      **Structured Query Language**      [01장 029쪽]

데이터베이스에 데이터를 추가하거나 추출하기 위해 사용하는 언어.

□ 데이터 마이닝      **data mining**      [01장 034쪽]

의사 결정을 돕기 위해 데이터에서 패턴 혹은 지식을 추출하는 작업.

□ 머신러닝	<b>machine learning</b> <small>참고 용어</small> <b>딥러닝</b> (deep learning)	[01장 034쪽]
	<small>인공신경망을 기반으로 한 머신러닝 분야</small> 데이터에서 자동으로 규칙을 학습하여 문제를 해결하는 소프트웨어를 만드는 기술.	
□ 패키지	<b>package</b> <small>참고 용어</small> 라이브러리(library)	[01장 031쪽]
	파이썬에서 유용한 기능이 구현된 파일을 묶어서 배포하는 단위.	
□ 넘파이	<b>NumPy</b>	[01장 032쪽]
	파이썬의 대표적인 배열 패키지. 다차원 배열을 쉽게 다룰 수 있다. 하나의 배열엔 하나의 데이터 타입만 담을 수 있지만, 매우 효율적이고 성능이 높다. 파이썬의 다른 과학 패키지와 호환성이 높으며 기본 데이터 구조로 널리 사용된다.	
□ 판다스	<b>pandas</b>	[01장 032쪽]
	표 형식 데이터를 다루기 위해 편리한 도구를 다양하게 제공하는 강력한 데이터 분석 패키지. CSV 파일이나 엑셀 파일을 읽거나 쓸 수 있으며 데이터 분석을 위한 많은 기능을 제공한다.	
□ 구글 코랩	<b>Google Colab</b> <small>참고 용어</small> <b>노트북</b> (notebook)	[01장 039쪽]
	<small>코랩 파일</small> 웹 브라우저에서 무료로 파이썬 프로그램을 테스트하고 저장할 수 있는 서비스.	
□ 깃허브	<b>GitHub</b>	[01장 052쪽]
	버전 관리 도구인 깃(git)의 저장소를 무료로 호스팅 해 주는 웹사이트 중 하나. 여러 사람이 동시에 코드를 수정할 수 있고, 다양한 기능을 제공한다.	
□ 배열	<b>array</b>	[01장 070쪽]
	같은 종류의 데이터가 순서대로 나열된 데이터 구조. 나열된 축이 하나인 경우 1차원 배열, 축이 2개인 경우 2차원 배열이라고 한다.	

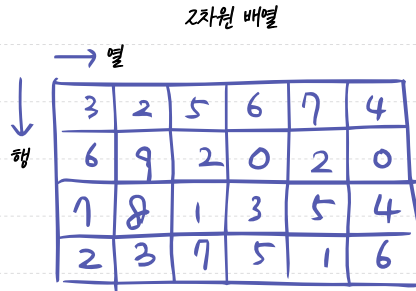


□ 데이터프레임

DataFrame

[01장 070쪽]

판다스의 핵심 데이터 구조로 행과 열로 구성된다. CSV 파일이나 엑셀 파일에서 데이터를 읽거나 저장할 수 있다.



□ 시리즈

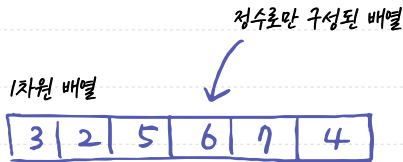
series

1차원 배열과 매우 비슷하다.

[01장 070쪽]

판다스의 또 다른 데이터 구조. 시리즈에 담긴 데이터는 모두 정수이거나 모두 문자열인 것처럼 동일한 종류여야 한다.

파이썬의 리스트와 다름 주의!!



□ CSV

comma-separated values

[01장 063쪽]

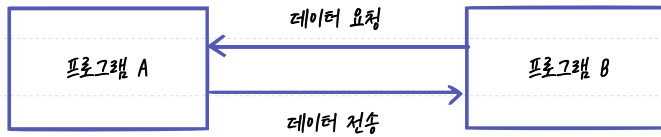
콤마(,)로 구분된 텍스트 파일. 한 줄이 하나의 레코드이며, 레코드는 콤마로 구분된 여러 필드 혹은 열로 구성된다.

데이터가 표 형태여야 하므로 레코드에 있는 필드 개수는 모두 동일해야 한다.

# 02장 <sup>④</sup> 데이터 수집하기

□ API Application Programming Interface [02장 085쪽]

프로그램 간 데이터를 전달하기 위해 정한 규칙. 공개 API를 사용하면 데이터 수집 과정을 자동화할 수 있다. *API 인증키 필요!*



□ HTTP HyperText Transfer Protocol [02장 085쪽]

인터넷에서 웹 페이지를 전송하는 기본 통신 규약. *프로토콜(protocol)*

□ HTML Hypertext Markup Language [02장 086쪽]

웹 브라우저가 화면에 표시할 수 있는 문서의 한 종류이자 웹 페이지를 위한 표준 언어. *마크업(markup) 언어라고 부르며 <div>와 같은 태그(tag) 형태로 구성된다.*

□ JSON JavaScript Object Notation [02장 088쪽]

데이터 전달 포맷. 원래는 자바스크립트 언어를 위해 만들어졌지만 현재는 범용적인 포맷으로 사용한다. *사람이 읽고 쓰기에 편리하고, 프로그램이 읽고 분석하기에도 좋아서!*

□ 딕셔너리 dictionary [02장 089쪽]

중괄호 안에 키(key)와 값(value)을 쌍으로 가지고 있는 자료 형태.

□ 리스트 list [02장 089쪽]

대괄호 안에 여러 종류의 데이터를 쉼표로 구분하여 일렬로 나열한 자료형. *리스트 안에는 정수나 문자열을 섞어 쓸 수 있다.*

□ 직렬화	<b>serialization</b> <span style="float: right;">[02장 091쪽]</span> 프로그램 상의 객체를 저장하거나 읽을 수 있는 형태로 변환하는 것.
□ 역직렬화	<b>deserialization</b> <span style="float: right;">[02장 091쪽]</span> 직렬화된 정보를 다시 프로그램에서 실행 가능한 객체로 변환하는 것.
□ XML	<b>eXtensible Markup Language</b> <span style="float: right;">[02장 119쪽]</span> 컴퓨터와 사람이 모두 읽고 쓰기 편한 문서 포맷. 엘리먼트들이 계층 구조를 이루면서 정보를 표현한다. <pre style="background-color: #f0f0f0; padding: 10px; margin-top: 10px;"> &lt;book&gt;   &lt;name&gt;혼자 공부하는 데이터 분석&lt;/name&gt;   &lt;author&gt;박해선&lt;/author&gt;   &lt;year&gt;2022&lt;/year&gt; &lt;/book&gt; </pre>
□ HTTP GET 방식	<b>HTTP GET</b> <span style="background-color: #e0e0e0; padding: 2px;">참고 용어</span> HTTP <span style="border: 1px solid blue; border-radius: 50%; padding: 2px;">POST</span> <span style="float: right;">[02장 104쪽]</span> 브라우저가 웹 서버에 요청을 할 때 URL로 파라미터 값이나 데이터를 전달하는 방식. <div style="margin-top: 10px; background-color: #f0f0f0; padding: 10px;"> <p style="text-align: center;">URL 뒤에 파라미터나 데이터를 붙이지 않고, HTTP 프로토콜 안에 있는 별도의 공간에 데이터를 실어 보낸다.</p> <p>http://data4library.kr/api/loanItemSrch?startDt=2021-04-01&amp;endDt=2021-04-30 → ? 기호 뒤에 파라미터 값이 줄줄이 소시지...</p> </div>
□ 웹 스크래핑	<b>web scraping</b> <span style="background-color: #e0e0e0; padding: 2px;">참고 용어</span> 웹 크롤링(web crawling) <span style="float: right;">[02장 119쪽]</span> 웹사이트에서 필요한 데이터를 추출하는 기술.
□ 뷰티풀수프	<b>Beautiful Soup</b> <span style="float: right;">[02장 125쪽]</span> HTML 문서를 파싱하는데 사용하는 대표적인 파이썬 패키지. requests 패키지로 가져온 HTML에서 원하는 태그나 텍스트를 찾는 기능을 제공한다.

그것이 알고싶다 파싱(parsing)과 파서(parser)

입력 데이터를 받아 데이터 구조를 만드는 소프트웨어 라이브러리를 파서라고 하는데, 이런 과정을 파싱이라고 한다. 예를 들어 HTML 문서를 가져오기 위해 파이썬 내장 파서인 html.parser를 사용할 수 있다.

□ 람다 **lambda** [02장 140쪽]  
함수 이름 없이 한 줄로 쓰는 함수. 코드를 간결하게 작성할 수 있어 함수를 간단하고 빠르게 구현할 수 있다.

# 03장 데이터 정제하기

□ 데이터 정제 **data cleaning** [03장 154쪽]  
**참고 용어** 데이터 랭글링(data wrangling), 데이터 먼징(data munging)  
데이터에서 손상되거나 부정확한 부분을 수정하고, 불필요한 데이터를 삭제하거나 불완전한 값을 교체하는 등의 작업.

□ 불리언 배열 **boolean array** [03장 156쪽]  
True 또는 False로 이루어진 배열.

□ NaN **Not A Number** [03장 155쪽]  
판다스에서 누락된 값을 표시할 때 사용한다.

□ 원소별 비교 **element-wise comparison** [03장 157쪽]  
판다스의 데이터프레임과 인덱스를 하나의 값과 비교하면 데이터프레임과 인덱스에 있는 모든 원소와 비교한다. 비교 결과는 불리언 배열로 반환한다.

□ 정규 표현식

regular expression

[03장 196쪽]

문자열에서 패턴을 찾고 대체하기 위한 규칙의 모음. 정규 표현식을 사용하면 복잡한 패턴을 가진 문자열을 쉽게 검색할 수 있다.

# 04장 <sup>✓</sup> 데이터 요약하기

□ 기술통계

descriptive statistics

[04장 223쪽]

데이터 내용을 요약하여 설명하는 방법. 정량적인 수치로 전체 데이터의 특징을 요약하거나 이해하기 쉬운 간단한 그래프를 포함한다.

**그것이 알고싶다** 대표적인 기술통계량

평균, 중앙값, 최솟값, 최댓값, 분위수, 분산, 표준편차, 최빈값.

□ 평균

mean

[04장 226쪽]

데이터값을 모두 더한 후 데이터 개수로 나눈 값.

□ 중앙값

median

[04장 229쪽]

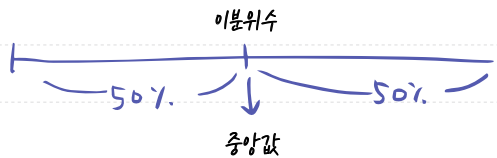
전체 데이터를 크기 순서대로 일렬로 늘어 놓았을 때 중간에 위치한 값. 전체 데이터 개수가 짝수일 때는 중간의 두 데이터의 평균을 계산한다.

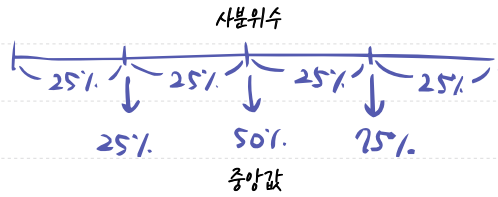
□ 분위수

quantile

[04장 231쪽]

순서대로 나열된 데이터를 일정한 간격으로 나누는 기준점.





그것이 알고싶다 보간(interpolation)

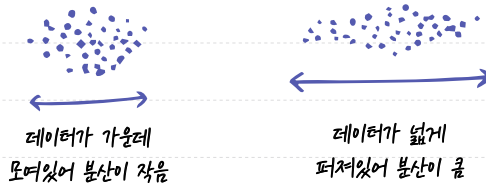
두 지점 사이에 놓인 특정 위치의 값을 구하는 방법.

□ 분산

variance

[04장 235쪽]

데이터가 평균에서 얼마나 멀리 퍼져 있는지 나타내는 값. 데이터를 평균에서 뺀 다음 제곱한 후 전체 데이터 개수로 나누어 구한다.



□ 표준편차

standard deviation

[04장 237쪽]

분산의 제곱근으로 데이터의 분포 정도를 나타내는 값.

□ 최빈값

mode

[04장 239쪽]

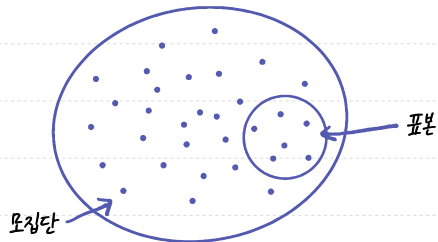
데이터에서 가장 많이 등장하는 값.

□ 표본집단

sample

[04장 247쪽]

모집단 전체 데이터 중에서 수집한 일부 데이터.



□ 맷플롯립

matplotlib

[04장 255쪽]

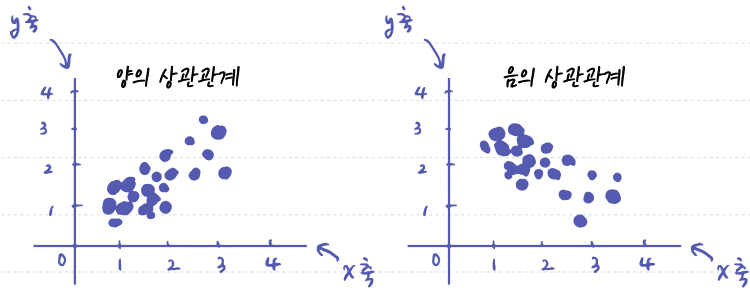
파이썬의 대표적인 그래프 패키지입니다. 산점도, 히스토그램, 상자 수염 그림을 비롯하여 막대 그래프, 선 그래프 등 많은 종류의 그래프를 지원한다.

□ 산점도

scatter plot

[04장 255쪽]

데이터를 2차원 평면 또는 3차원 공간에 점으로 표시하는 그래프.

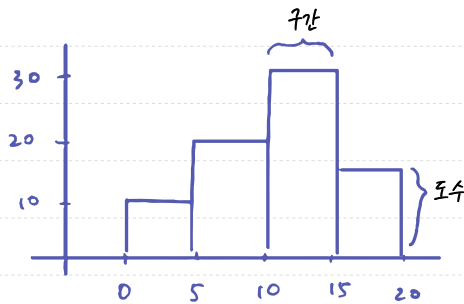


□ 히스토그램

histogram

[04장 261쪽]

데이터를 일정 구간으로 나누어 구간에 속한 데이터 개수를 막대로 표현하는 그래프. 데이터의 분포를 확인할 수 있다.



□ 도수분포표

frequency table

[04장 262쪽]

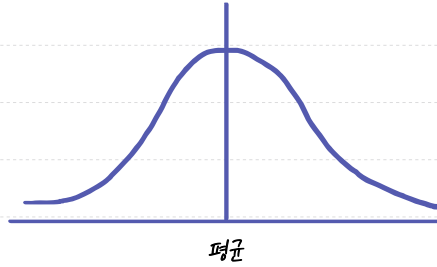
히스토그램에 나오는 구간과 도수를 표로 요약한 것.

□ 정규분포

normal distribution

[04장 264쪽]

종 모양처럼 가운데가 볼록하고 평균을 중심으로 대칭인 분포.



□ 표준정규분포

standard normal distribution

[04장 264쪽]

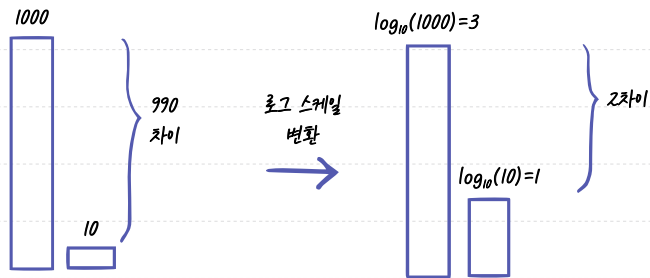
평균이 0이고 표준편차가 1인 정규분포.

□ 로그 스케일

log scale

[04장 265쪽]

그래프 축에 로그 함수를 사용하여 눈금 수치를 표현하는 것.

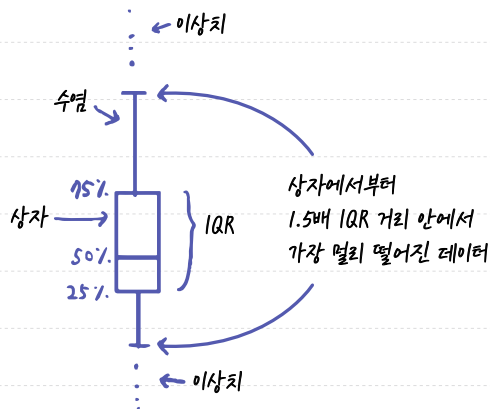


□ 상자 수염 그림

box-and-whisker plot

[04장 269쪽]

사분위수, 최솟값, 최댓값을 사용해 여러 특성의 분포를 비교할 수 있는 그래프.





□ 이상치

outlier

[04장 269쪽]

관측 범위에서 아주 많이 벗어난 값 혹은 데이터에 내재한 패턴을 크게 벗어난 값.

# 05장 ✓ 데이터 시각화하기

□ 피겨

figure

[05장 283쪽]

맷플롯립의 그래프 요소를 모두 담고 있는 최상위 객체. 맷플롯립으로 그래프를 그릴 때 자동으로 피겨가 생성되고 그래프가 그려진 후 삭제된다. 명시적으로 피겨 객체를 만들면 다양한 옵션을 제어할 수 있다.

□ rcParams

Runtime Configuration Parameters

[05장 291쪽]

맷플롯립 그래프의 기본값을 관리하는 객체. 객체에 담긴 값만 출력하는 것뿐만 아니라 새로운 값으로 바꿀 수도 있다. 이후에 그려지는 모든 그래프에 바뀐 설정이 적용된다.

□ 축

Axis

[05장 294쪽]

두 개 이상의 Axis 객체로 이루어진  
Axes 객체는 그래프가 그려질 영역을 표현한다.

그래프에서 데이터 좌표를 표현한다. 맷플롯립에서는 **Axis** 클래스로 축 객체를 다룬다.

□ 마커

marker

[05장 294쪽]

맷플롯립의 기본 마커는 동그라미를 나타내는 'o'

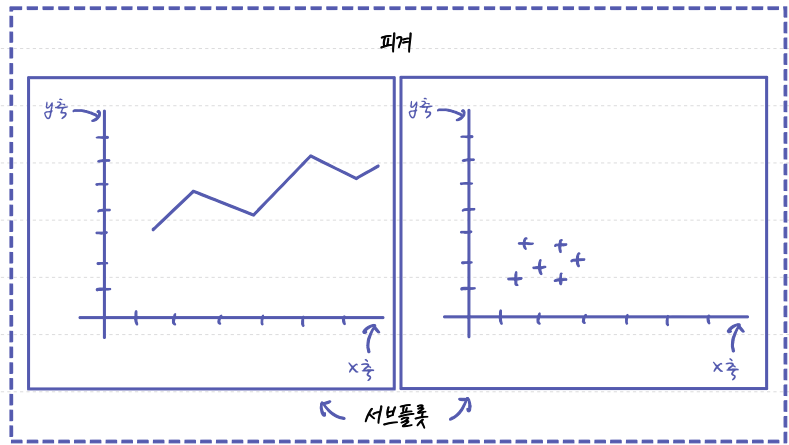
그래프에 **데이터 포인트**를 표시하는 방법. rcParams 객체나 scatter() 함수의 marker 매개변수로 마커를 바꿀 수 있다.

□ 서브플롯

subplot

[05장 294쪽]

피겨 안에 포함된 그래프 영역. 보통 Axes 객체를 말하며 여러 개의 서브플롯을 포함하는 피겨를 만들 수 있다.



□ 명목형 데이터

nominal data 참고 용어 순서형 데이터 (ordinal data)

[05장 307쪽]

성별, 국가와 같이 순서를 매길 수 없는 데이터.

성적 등급처럼 순서가 있는 레이어(1, 2, 3...).

□ 선 그래프

line chart

[05장 308쪽]

각 데이터 포인트를 직선으로 연결한 그래프. 시간별 추세를 표현하기 좋다.

□ 막대 그래프

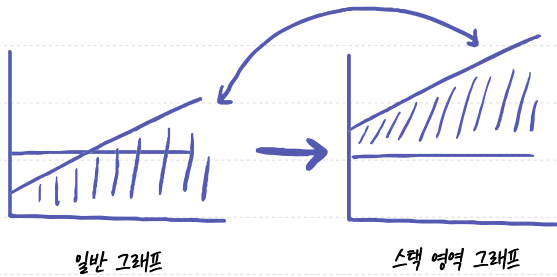
bar chart

[05장 314쪽]

범주형 또는 명목형 데이터에 대한 값의 빈도를 막대 길이로 표현한 그래프.

# 06장 <sup>✓</sup>복잡한 데이터 표현하기

- 객체지향 API      **object-oriented API**      [06장 329쪽]  
명시적으로 피겨 객체와 서브플롯 객체를 만들고, 해당 객체의 메서드를 사용하여 맷플롯립 그래프를 그리는 방법.
- 컬러맵      **color map**      [06장 343쪽]  
맷플롯립에서 그래프를 그리는 데 사용하기 위해 사전에 정의한 색상 리스트.
- 컬러 막대      **color bar**      [06장 343쪽]  
컬러맵의 색깔이 어떤 값에 대응하는지 참조 정보를 제공하는 막대.
- 스택 영역 그래프      **stacked area chart**      [06장 356쪽]  
여러 개의 선 그래프를 y축 방향으로 쌓은 그래프. 선 아래로 색상이 채워져 영역 형태로 표현된다.



□ 피벗 테이블

pivot table

[06장 357쪽]

테이블 형태의 데이터를 평균, 합 등의 방식으로 집계하여 만든 요약표.

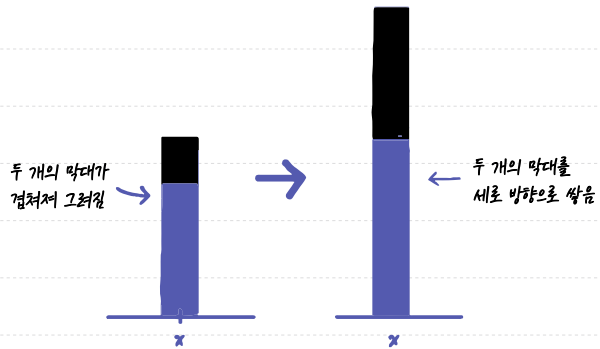
출판사	발행년도	대용건수		2021	2022
A출판사	2020	10	피벗 테이블 →	A출판사	30
B출판사	2021	20		B출판사	20
A출판사	2021	30			
	⋮				

□ 스택 막대 그래프

stacked bar chart

[06장 362쪽]

여러 개의 막대 그래프를 y축 방향으로 쌓은 그래프. 막대 위에 막대가 누적되듯이 표현된다.



□ 원 그래프

pie chart

[06장 366쪽]

전체 데이터에 대한 비율을 부채꼴로 나타낸 그래프.

# 07장 <sup>✓</sup> 검증하고 예측하기

□ 모수검정      parametric test      [07장 383쪽]

통계학에서 모집단에 대한 파라미터를 예측하는 방법.

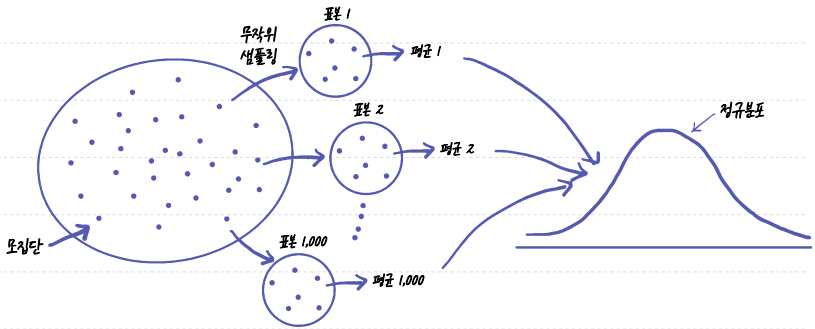
□ 표준점수      standard score      [07장 383쪽]

정규분포 상에서 데이터 포인트가 원점에서 얼마나 떨어져 있는지 표준편차 비율로 나타내는 점수.

↓  
데이터 포인트에서 평균을 빼고  
표준편차로 나누어 구한다.

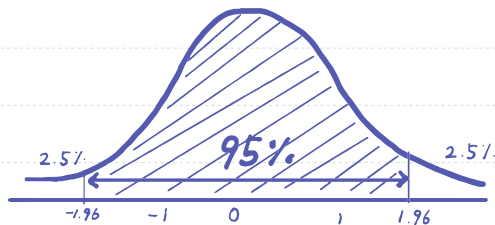
□ 중심극한정리      central limit theorem      [07장 388쪽]

동일한 분포에서 무작위로  $n$ 개의 샘플을 추출하여 평균을 반복해서 구하면 이 평균의 분포가 정규분포에 가까워진다는 개념.



□ 신뢰구간      confidence interval      [07장 393쪽]

알려지지 않은 모집단의 파라미터를 추정할 범위. 일반적으로 95% 신뢰구간을 사용하며 종종 90%나 99%의 신뢰구간을 사용한다.



□ 가설검정

statistical hypothesis test

[07장 397쪽]

표본의 통계치를 기반으로 모집단에 대한 가설을 받아들일지 기각할지 결정하는 방법.

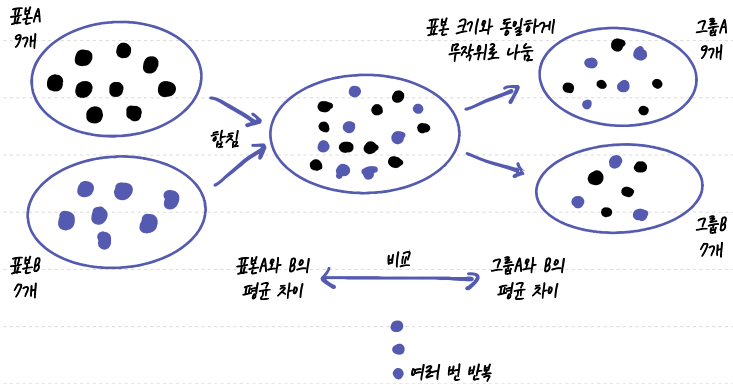
- 영가설(귀무가설): 통계적으로 의미가 없다고 예상되는 가설.
- 대립가설: 표본 사이에 통계적인 차이가 있다고 보는 가설

□ 순열검정

permutation test

[07장 402쪽]

정규분포에 대한 가정을 하지 않고 수행할 수 있는 검정 방법.



□ 모델

model

[07장 412쪽]

좁은 의미로 보면 머신러닝으로 훈련된 프로그램 객체를 의미. 데이터에서 유용한 패턴을 학습하고, 새로운 데이터에 대한 예측을 만들 수 있다.

□ 사이킷런

scikit-learn

[07장 412쪽]

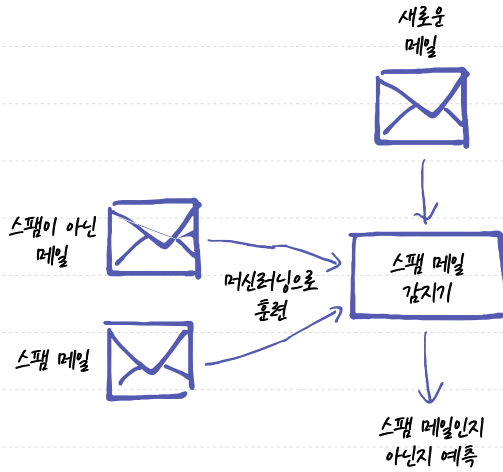
파이썬의 대표적인 머신러닝 라이브러리.

□ 훈련

training

[07장 413쪽]

머신러닝 알고리즘이 데이터에서 규칙을 찾는 과정 또는 모델에 데이터를 전달하여 데이터의 패턴을 학습시키는 과정.

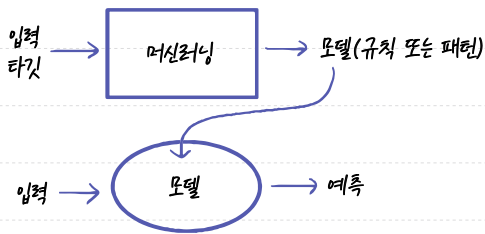
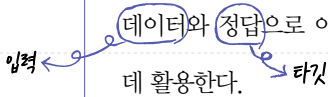


□ 지도 학습

supervised learning

[07장 412쪽]

데이터에 있는 각 샘플에 대한 정답을 알고 있는 경우 사용하는 머신러닝 알고리즘. 데이터와 정답으로 이루어진 훈련 데이터가 필요하며 새로운 데이터를 예측하는데 활용한다.



□ 비지도 학습

unsupervised learning

[07장 412쪽]

타깃 데이터 없이 입력 데이터만 있을 때 사용하는 알고리즘.

대표적으로 비슷한 샘플끼리 그룹으로 묶는 '군집 알고리즘'이 있다.

□ 결정계수

coefficient of determination

[07장 418쪽]

회귀 모델에서 모델을 평가하는 한 가지 방법. 0과 1 사이의 값을 가진다.

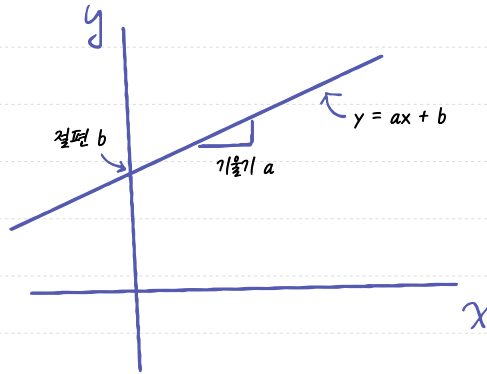
↙  
↘  
1에 가까울수록 완벽하다.

□ 선형 회귀

linear regression

[07장 420쪽]

연속적인 타깃 값에 적용하는 회귀 알고리즘. 선형 함수를 사용해 모델을 만든다.



□ 로지스틱 회귀

logistic regression

[07장 421쪽]

타깃이 실수가 아닌 어떤 종류의 카테고리(범주)일 사용하는 분류 알고리즘. 선형 함수를 사용하여 패턴을 학습하고 출력 결과에 로지스틱 함수를 적용한다.



**MEMO**

Handwriting practice area consisting of horizontal dashed lines.

혼자  
공부하는  
사람들을 위한  
용어 노트