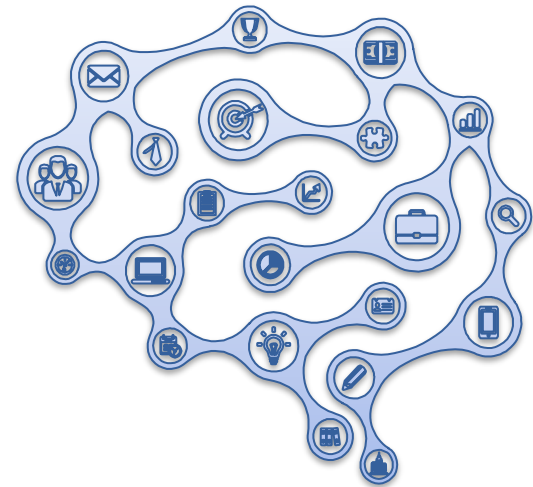


# 의료 Artificial Intelligence

통계적방식 인공지능 / 데이터 마이닝

2022.04.28



# 오늘 배울 내용 ...

1. 통계 기초 - 표준정규분포
2. 데이터 마이닝
3. 데이터 마이닝 실습
4. mblock 실습

어렵지 않다

쉬운 것도 아니다



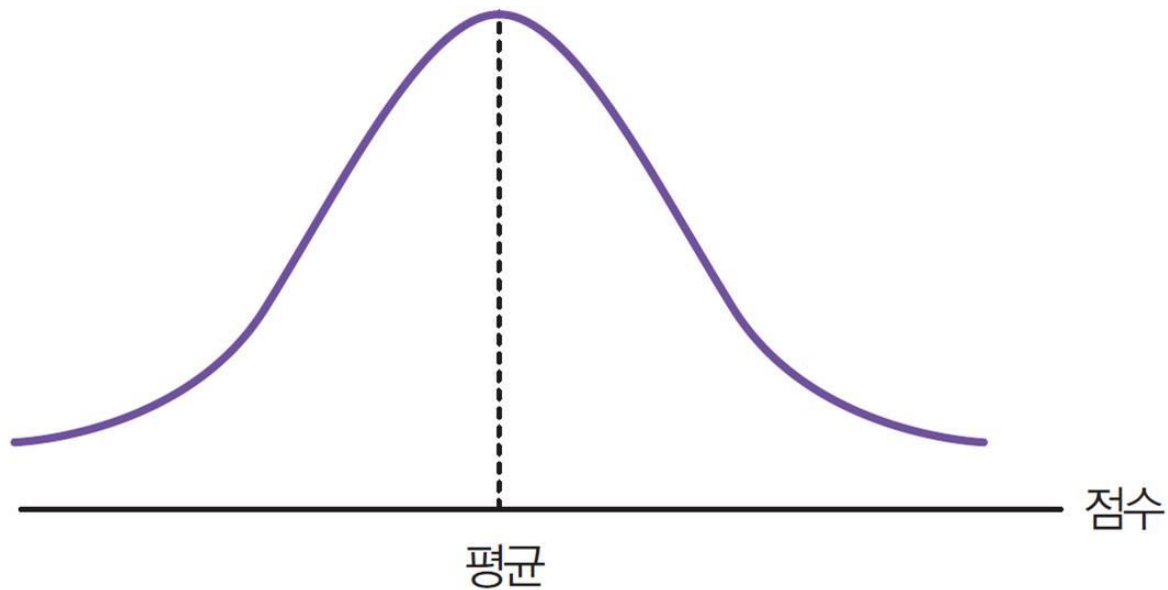
---

# 인공지능 이론

---

# 확률 통계 기초

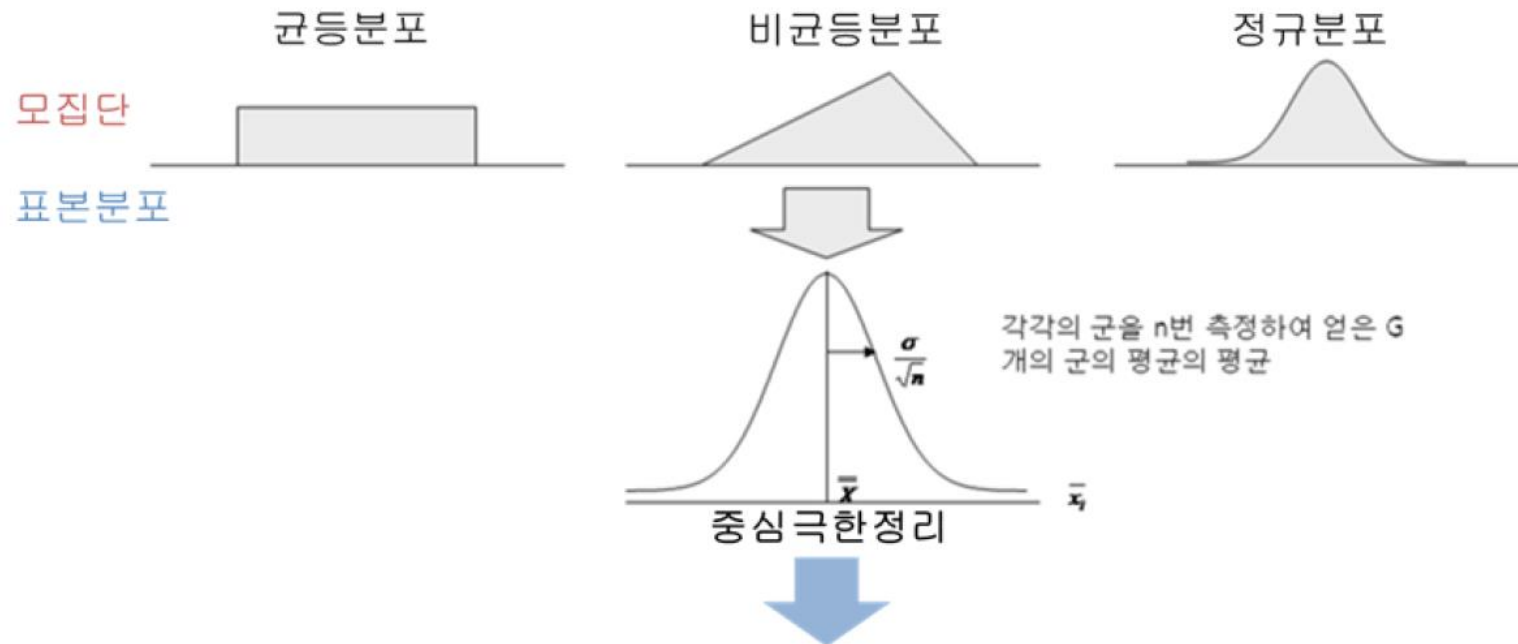
- Q : 현실에서의 여러가지 변수들의 분포를 잘 설명하는 확률밀도함수는 있을까?
- Q : 현실에서 일어나는 사건들에 대해 확률을 계산하는 공통적인 방법은 ?



# 확률 통계 기초

중심극한 정리 : 표본의 크기가 충분히 크다면, 표본평균들의 분포는 정규분포를 이룬다

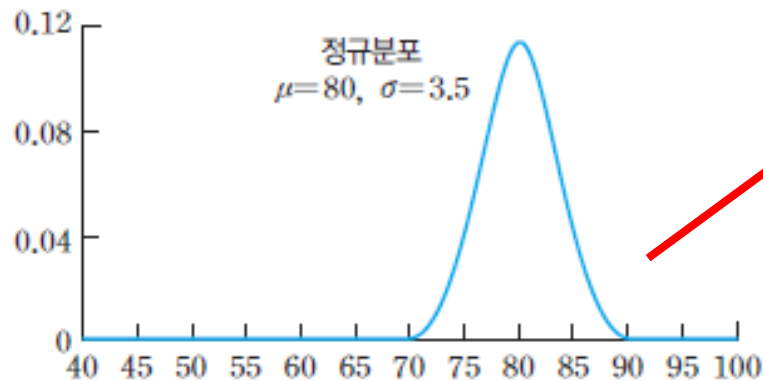
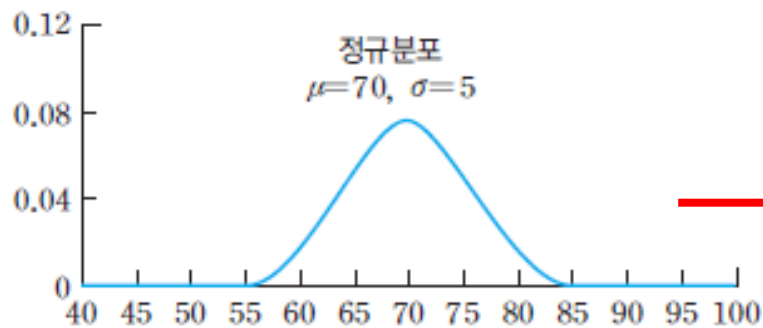
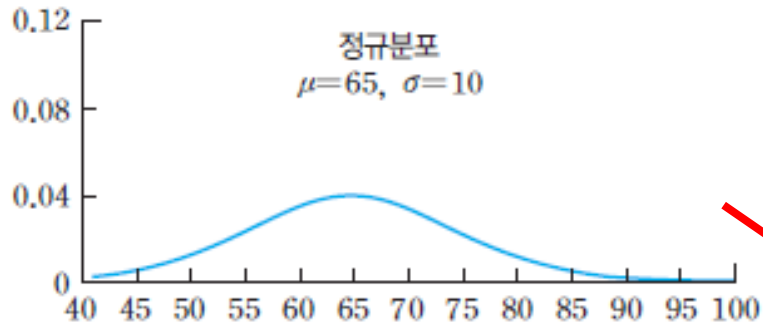
- 표본의 크기가 커질 수록 (보통 30 이상), 표본이 이루는 분포가 <모집단의 평균  $\mu$ , 표준편차가  $\sigma/\sqrt{n}$ 인 정규분포>에 근접
- 모집단이 어떤 분포를 가지고 있던지 간에 (모집단 분포가 모양이던 상관없이) 표본의 크기가 충분히 크면 성립
- 이를 이용하면 특정 사건(내가 수집한 표본의 평균)이 일어날 확률값을 계산할 수 있게 된다



“모집단 분포에 상관없이” 큰 표본들의 “표본평균의 분포”가 정규분포로 수렴한다는 점을 이용하여, Z값을 구해 확률값을 구할 수 있게 된다. 즉, 수학적 확률 판단(추정)을 할 수 있다!

# 확률 통계 기초

Z값으로 확률을 구하려면 표준화를 해야 한다.



표준화

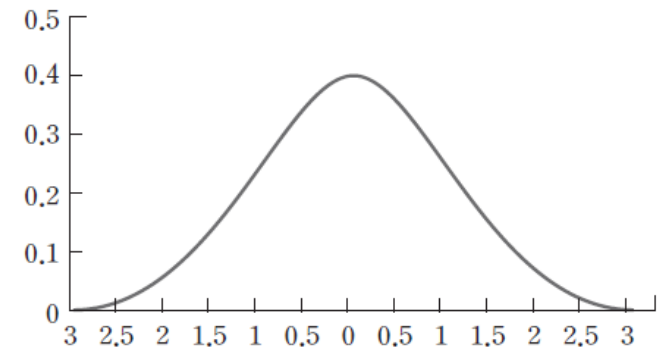
$$Z = \frac{X - \mu}{\sigma}$$

표준화

$$Z = \frac{X - \mu}{\sigma}$$

표준화

$$Z = \frac{X - \mu}{\sigma}$$



표준정규분포

$$\mu=0, \sigma=1$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$Z \sim N(0, 1)$$

# 확률 통계 기초

- 표준정규분포  $N(0,1)$  : 평균이 0, 표준편차가 1인 정규분포

## 정리

### 표준정규분포

연속확률변수  $X$ 가 정규분포  $N(m, \sigma^2)$ 을 따를 때, 확률변수  $Z = \frac{X - m}{\sigma}$ 은 표준정규분포  $N(0, 1)$ 을 따릅니다.

- 1 Z값 구하기 : 확률변수  $X$ 에 대해 Z값을 구하기

$$Z = \frac{X - \mu}{\sigma}$$

- 2 확률변수 구하기 : Z값에 대해 확률변수  $X$ 값을 구하기

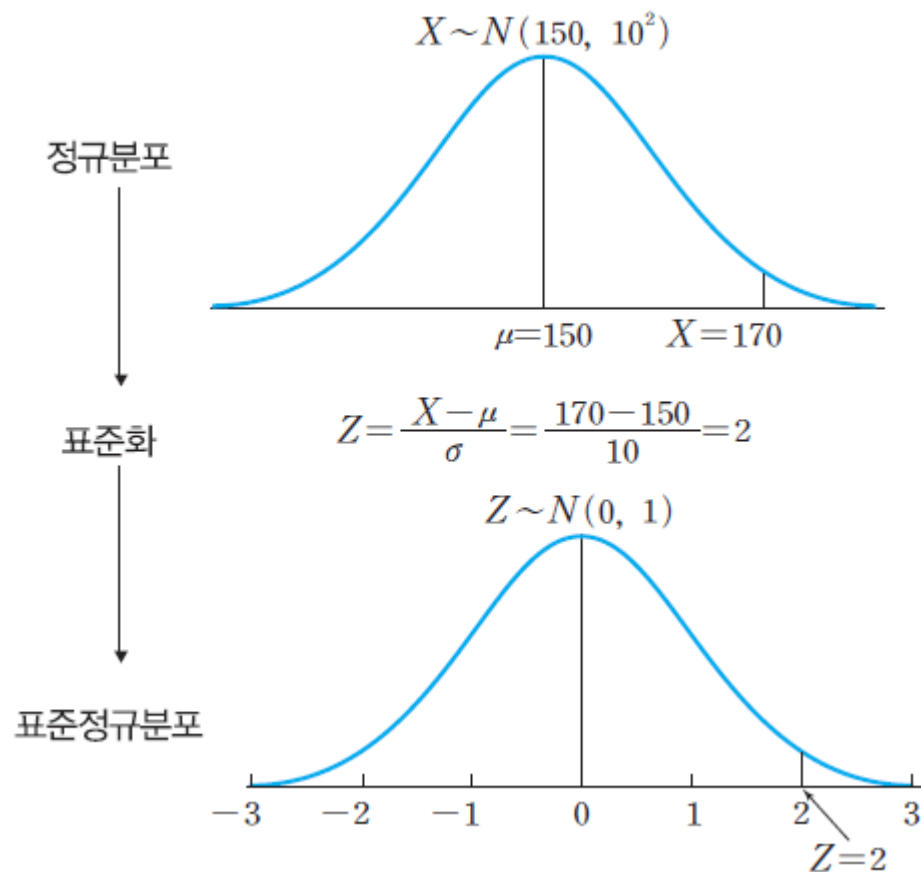
$$X = \mu + Z \times \sigma \qquad X = \bar{X} + Z \times \frac{\sigma}{\sqrt{n}}$$

# 확률 통계 기초

평균  $\mu=150$  이고, 표준편차  $\sigma=10$  인 정규분포  $N(150, 10^2)$ 에 대해  $X=170$ 에 대한  $Z$  점수를 계산하시오.

표준화 식  $Z = \frac{X - \mu}{\sigma}$  을 이용하면

$$Z = \frac{170 - 150}{10} \text{ 이므로 점수는 } 2.$$





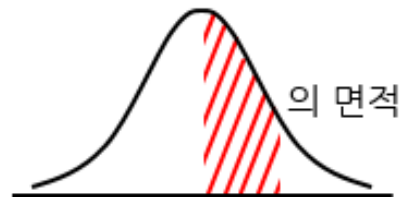
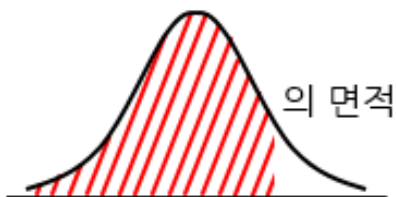
# 확률 통계 기초

## 3 Z값에 대한 확률 구하기

Z값	0.00	0.01	...	0.08	...
...					
2.0	0.4772				
...					
2.5				0.4951	

$$P(0 \leq Z \leq 2) = 0.4772$$

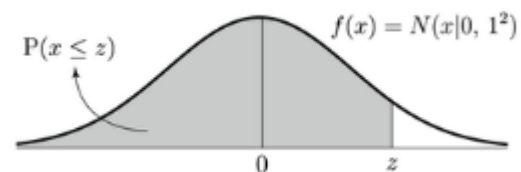
$$P(0 \leq Z \leq 2.58) = 0.4951$$



Z	0.00	0.01	0.02	0.03
0.0	0.5000	0.5040	0.5080	0.5120
0.1	0.5398	0.5438	0.5478	0.5517
0.2	0.5793	0.5832	0.5871	0.5910
0.3	0.6179	0.6217	0.6255	0.6293
0.4	0.6554	0.6591	0.6628	0.6664
0.5	0.6915	0.6950	0.6985	0.7019
0.6	0.7257	0.7291	0.7324	0.7357
0.7	0.7580	0.7611	0.7642	0.7673
0.8	0.7881	0.7910	0.7939	0.7967
0.9	0.8159	0.8186	0.8212	0.8238

Z	0.00	0.01	0.02	0.03
0.0	0.0000	0.0040	0.0080	0.0120
0.1	0.0398	0.0438	0.0478	0.0517
0.2	0.0793	0.0832	0.0871	0.0910
0.3	0.1179	0.1217	0.1255	0.1293
0.4	0.1554	0.1591	0.1628	0.1664
0.5	0.1915	0.1950	0.1985	0.2019
0.6	0.2257	0.2291	0.2324	0.2357
0.7	0.2580	0.2611	0.2642	0.2673
0.8	0.2881	0.2910	0.2939	0.2967
0.9	0.3159	0.3186	0.3212	0.3238

# 확률 통계 기초

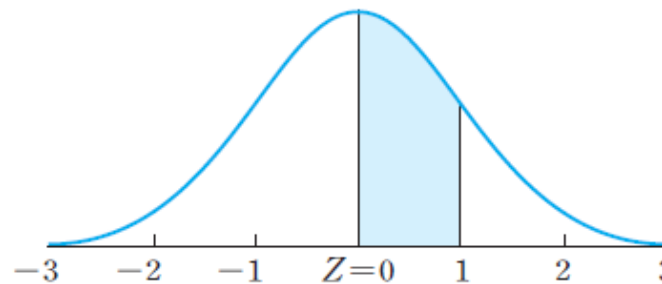
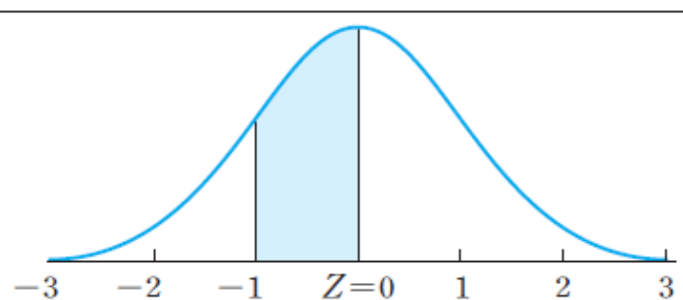


$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857

# 확률 통계 기초

## Case1

$$P(-1 \leq Z \leq 0) = P(0 \leq Z \leq 1)$$

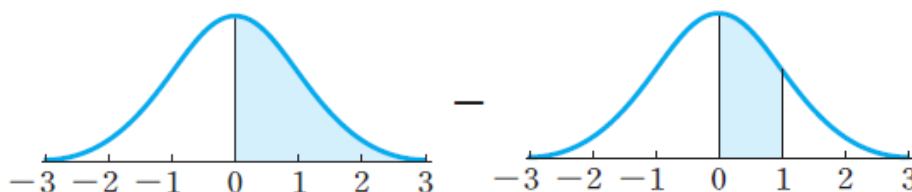
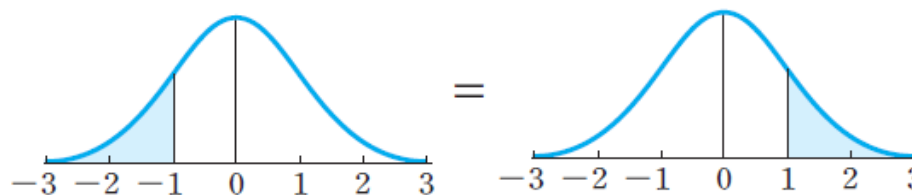


## Case2

$$P(Z \leq -1) = P(Z \geq 1)$$

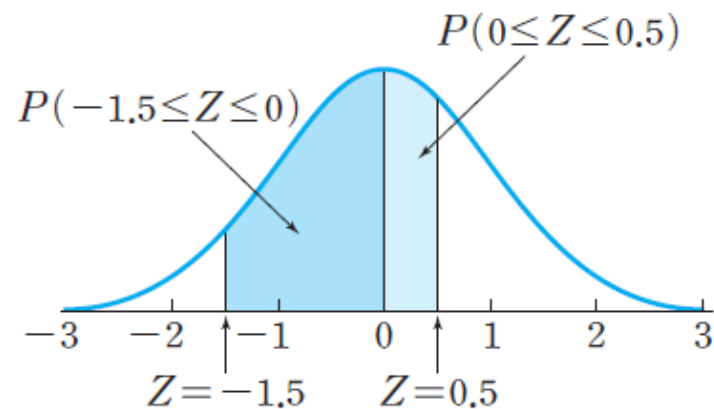
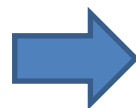
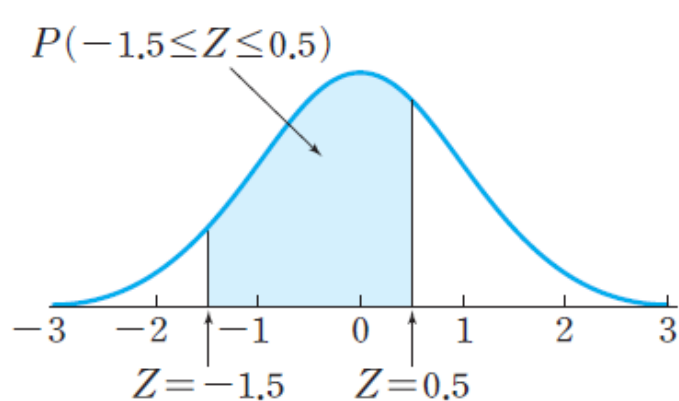
[계산]

$$P(Z \geq 1) = 0.5 - P(0 \leq Z \leq 1)$$



# 확률 통계 기초

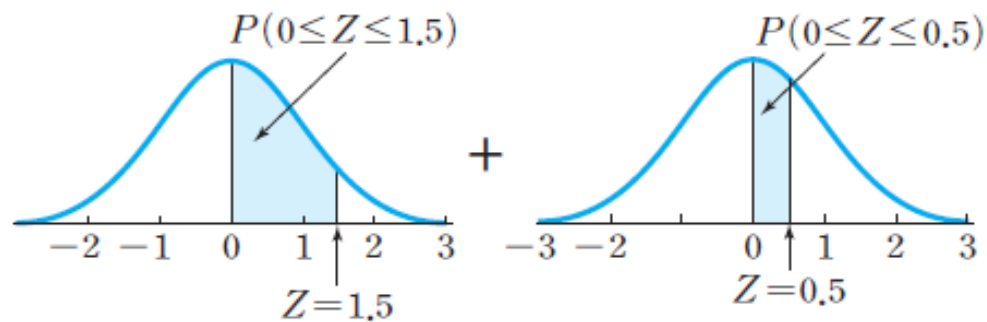
## Case3



[계산]

$$P(-1.5 \leq Z \leq 0) \\ = P(0 \leq Z \leq 1.5)$$

$$\therefore P(-1.5 \leq Z \leq 0.5) = \\ P(0 \leq Z \leq 1.5) + P(0 \leq Z \leq 0.5)$$



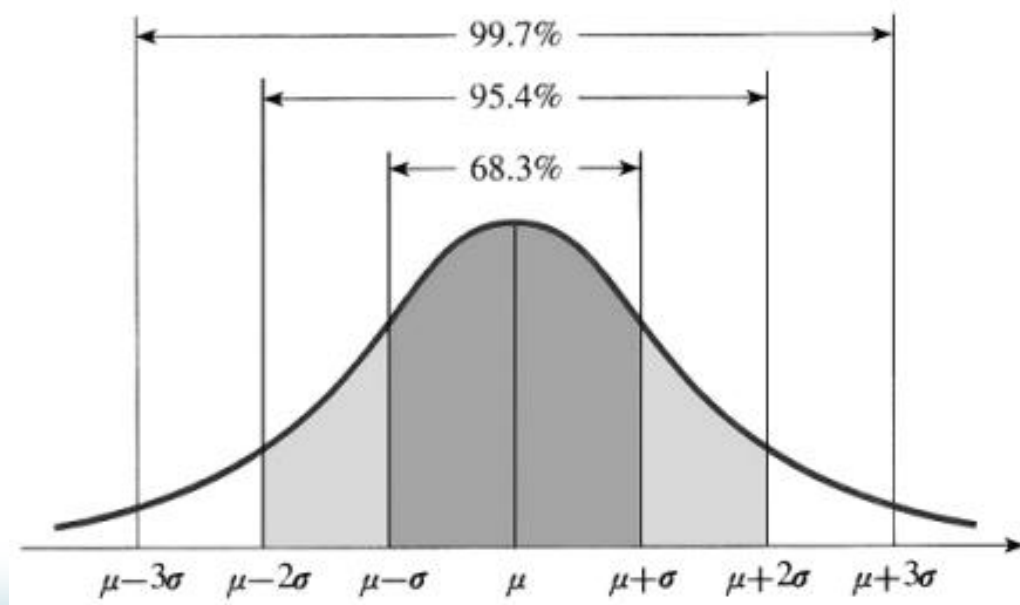
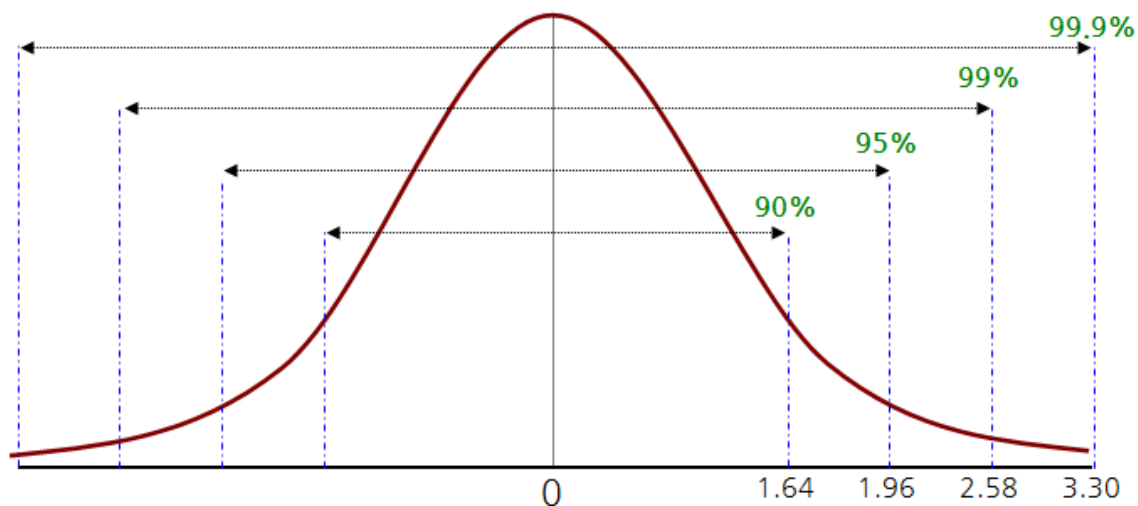
# 확률 통계 기초

한 회사의 건전지 수명은 평균이 110시간, 표준편차가 10인 정규분포를 따른다

- 1) 건전지의 수명이 90시간 이하일 확률은
  
  
  
  
  
  
  
  
  
  
- 2) 건전지 수명이 100시간 이상, 110시간 이하일 확률은

# 확률 통계 기초

## 주요 확률



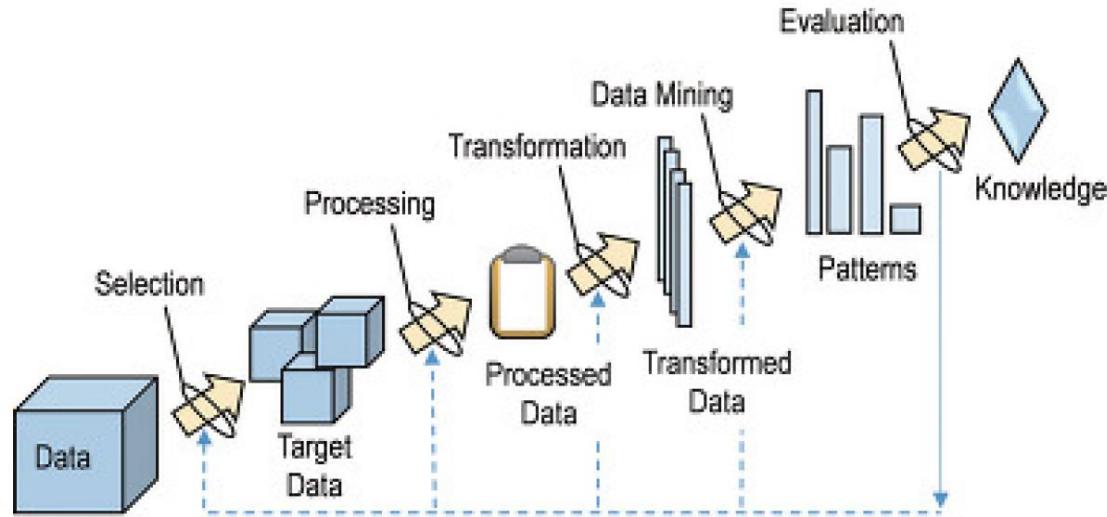
# 데이터 마이닝

데이터 마이닝(data mining)은 대규모로 저장된 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 분석하여 가치있는 정보를 추출하는 과정

- 데이터를 탐색하고 모델을 구축하는데 주안점을 둠
- 자료의 수집, 분류, 가공 등의 전처리 작업이 필요함



# 데이터 마이닝 과정



1. 데이터 정제(Data Cleaning) : 불필요하거나 일치하지 않는 데이터를 제거
2. 데이터 통합(Data Integration) : 다수의 데이터 소스들을 결합
3. 데이터 선택(Data Selection) : 필요한 데이터들을 데이터베이스로부터 검색
4. 데이터 변환(Data Transformation) : 요약이나 집계 등을 수행해 데이터마이닝을 위한 적합한 형태로 데이터 가공
5. 데이터 마이닝(Data Mining) : 지능적 방법들을 적용하여 데이터 패턴이나 지식을 추출
6. 데이터 검증(Data Evaluation) : 데이터 마이닝으로 찾아낸 패턴이나 지식을 검증
7. 데이터 시각화(Data Presentation) : 발견한 패턴이나 지식을 사용자에게 효과적으로 보여주기 위해 시각화



# 데이터 형태

- 데이터의 형태 - 구조 유무

- 정형(structured) 데이터

- 일정한 구조 보유
    - 데이터베이스 테이블(table, relation)
    - 시장바구니 데이터(market basket data)
      - 매출별 구매 항목 목록에 대한 데이터
      - 행(row)이 항목(item)의 리스트 구성

- 비정형(unstructured) 데이터

- 구조가 일정하지 않은 데이터
    - 텍스트(text) 데이터 : 신문기사, SNS 메시지 등
    - 스트림(stream) 데이터 : 지속적으로 관측되어 생성되는 데이터
    - 서열(sequence) 데이터 : 염기 서열, 아미노산 서열 데이터
    - 클릭(click) 데이터 : 홈페이지 방문자들의 순차적인 클릭
    - 시스템 로그(log) 데이터
    - 그래프(graph) 데이터

- 반정형(semi-structured) 데이터

- 구조화되어 있지만 관계형 데이터베이스의 테이블과 같은 형태로 저장되기 곤란한 데이터
    - XML(eXtensible Markup Language), JSON(JavaScript Object Notation) 등으로 표현

# 데이터 형태

## 데이터 설명과 데이터가 구조화 된 데이터

### XML 데이터



```
<?xml version="1.0" encoding="iso-8859-8" standalone="yes" ?>
<CURRENCIES>
  <LAST_UPDATE>2004-07-29</LAST_UPDATE>
  <CURRENCY>
    <NAME>dollar</NAME>
    <UNIT>1</UNIT>
    <CURRENCYCODE>USD</CURRENCYCODE>
    <COUNTRY>USA</COUNTRY>
    <RATE>4.527</RATE>
    <CHANGE>0.044</CHANGE>
  </CURRENCY>
  <CURRENCY>
    <NAME>euro</NAME>
    <UNIT>1</UNIT>
    <CURRENCYCODE>EUR</CURRENCYCODE>
    <COUNTRY>European Monetary Union</COUNTRY>
    <RATE>5.4417</RATE>
    <CHANGE>-0.013</CHANGE>
  </CURRENCY>
</CURRENCIES>
```

### JSON 데이터



```
{ "users": [
  {
    "firstName": "Ray",
    "lastName": "Villalobos",
    "joined": {
      "month": "January",
      "day": 12,
      "year": 2012
    }
  },
  {
    "firstName": "John",
    "lastName": "Jones",
    "joined": {
      "month": "April",
      "day": 28,
      "year": 2010
    }
  }
]
}
```

# 데이터 마이닝 모델

---

- **classification** : 분류
- **regression** (a.k.a. value estimation) : 회귀 (value 추정)
- **similarity matching** : 유사도 매칭
- **clustering** : 군집화
- **co-occurrence grouping** (a.k.a. association rule discovery) : 동시발생 (관계 규칙 발견)
- **profiling** (a.k.a. behavior description) : 프로파일링 (행동 묘사)
- **link prediction** : 연관성 예측 (ex. recommendation)
- **data reduction** : 데이터 사이즈 줄이기 (불필요한 데이터 제거, 형태 변환 등)
- **causal modeling** : 인과관계 모델링

---

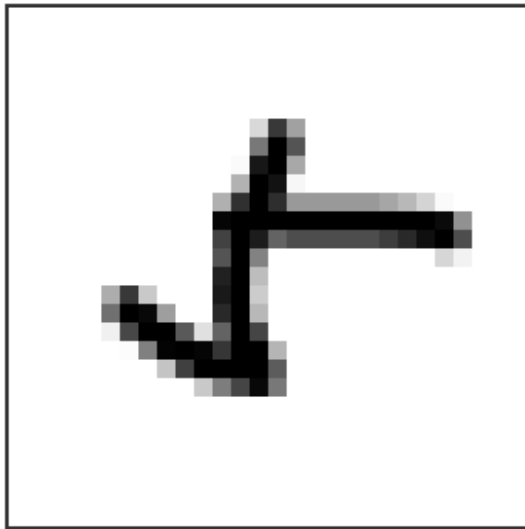
# 인공지능 실습

---

# 인공지능 실습

<http://myselfph.de/neuralNet.html>

Draw a digit in the box below and click the "recognize" button.



5

☒ Display Preprocessing

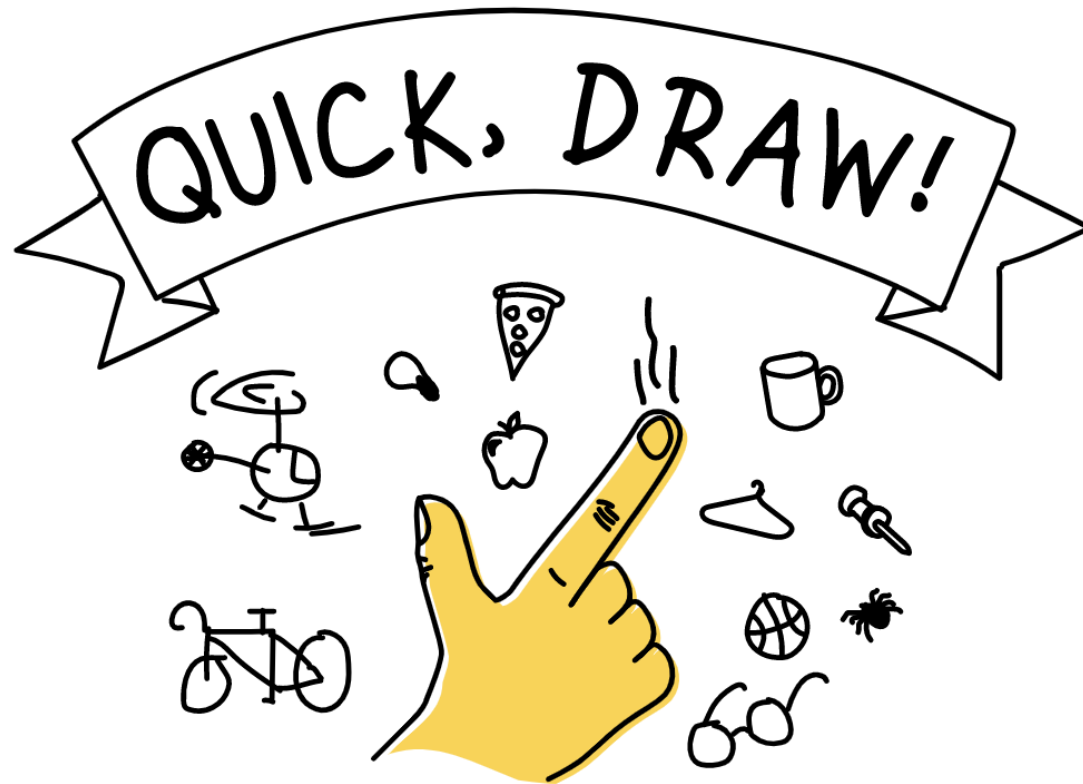
☒ Scale Stroke Width

clear

recognize

# 인공지능 실습

<https://quickdraw.withgoogle.com>



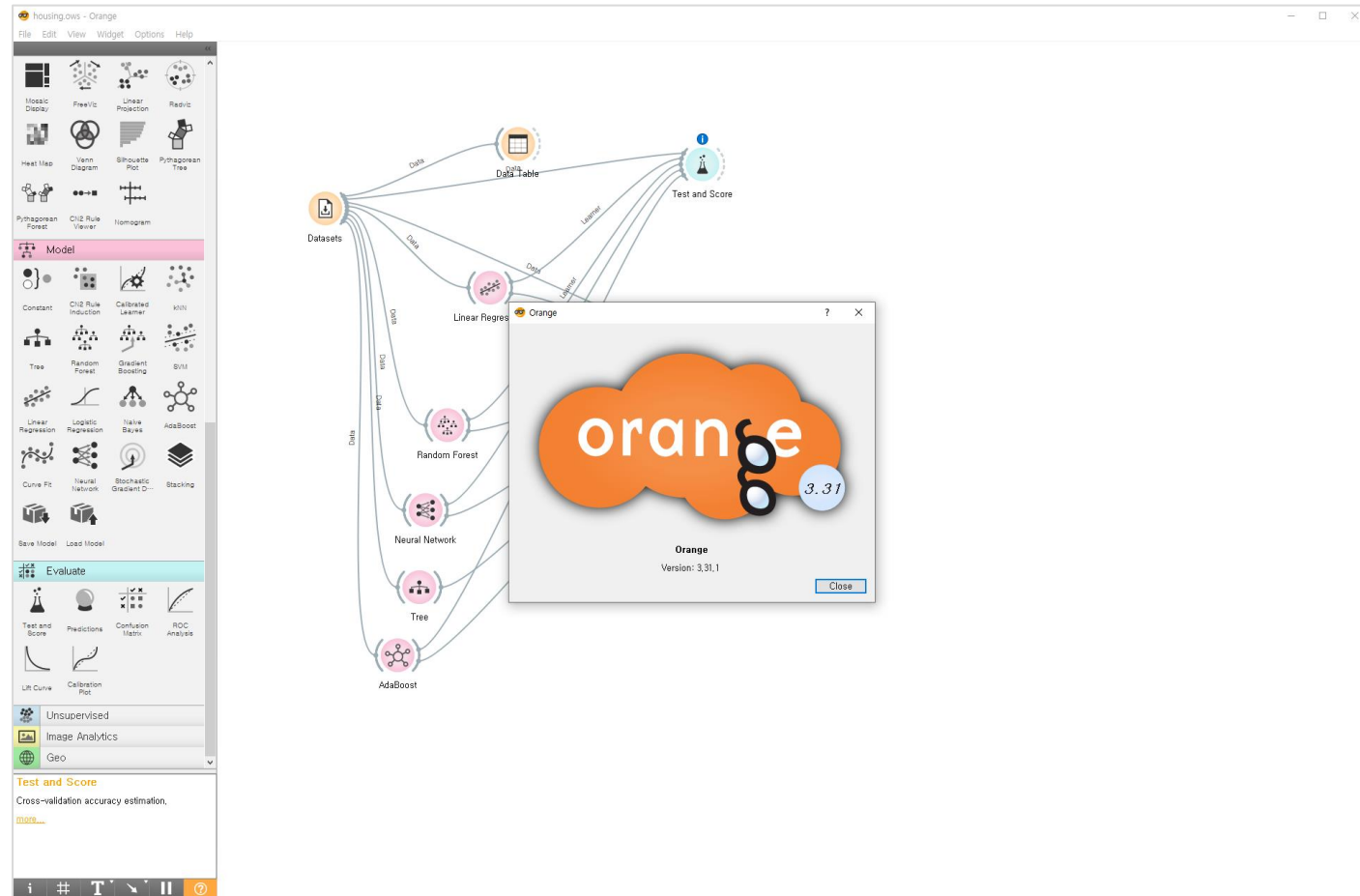
머신 러닝 기술이 학습을 통해 낙서를 인식할 수 있을까요?

여러분의 그림으로 머신 러닝의 학습을 도와주세요. Google은 머신 러닝 연구를 위해 세계 최대의 낙서 데이터 세트를 오픈소스로 공유합니다

# 인공지능 실습

<https://orangedatamining.com/>

<https://orangedatamining.com/download/#windows>



# 인공지능 실습

The screenshot displays the Orange3 data mining software interface. On the left is a vertical panel titled "위젯" (Widget) which contains various tool categories: Data, Transform, Visualize, Model, Evaluate, Unsupervised, Image Analytics, and Text Mining. Each category lists specific widgets with corresponding icons. The main area on the right is a large, empty workspace for building workflows. A red rectangular box is drawn over the workspace, containing the text "위젯 편집" (Widget Edit).

위젯

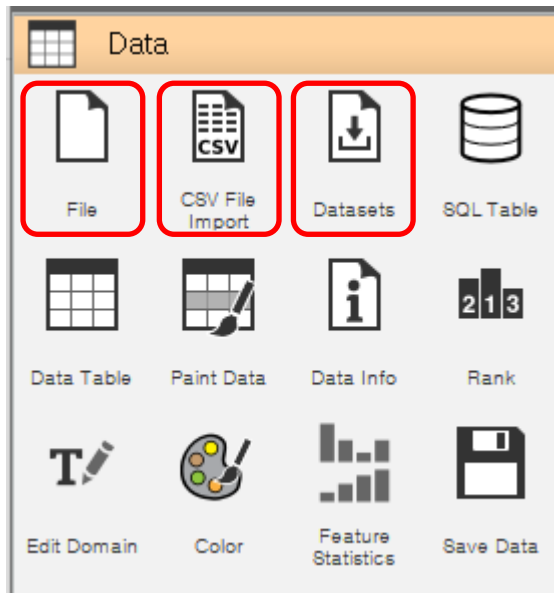
위젯 편집



# 인공지능 실습

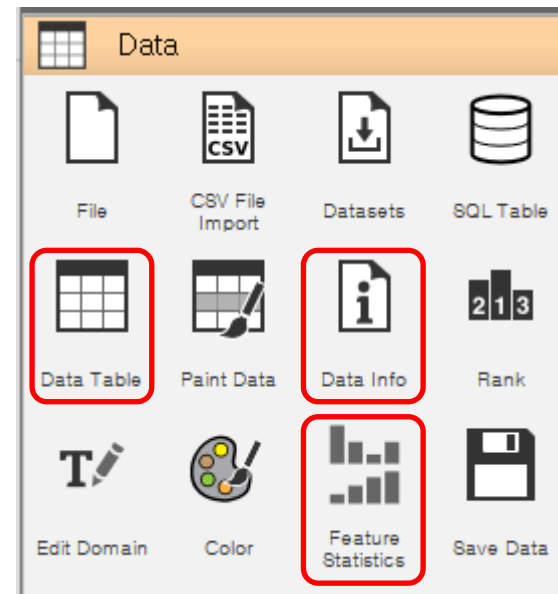
## 1. 데이터 입력

- Datasets
- File / CSV File



## 2. 데이터 보기

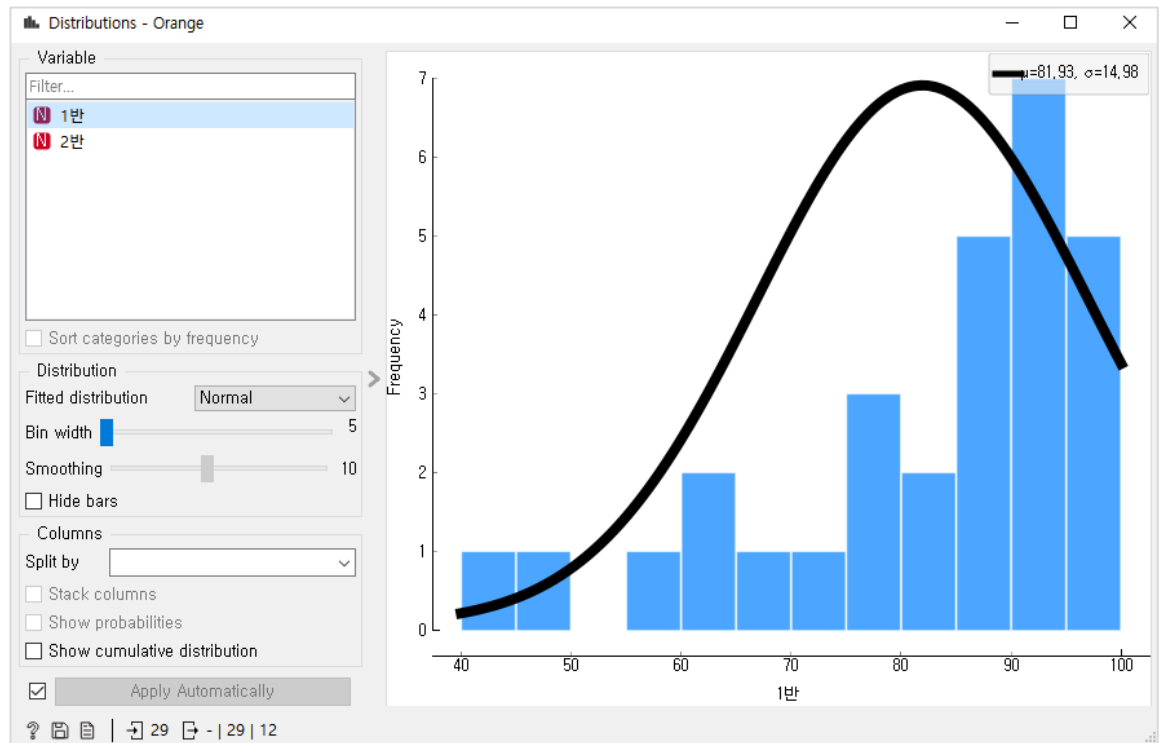
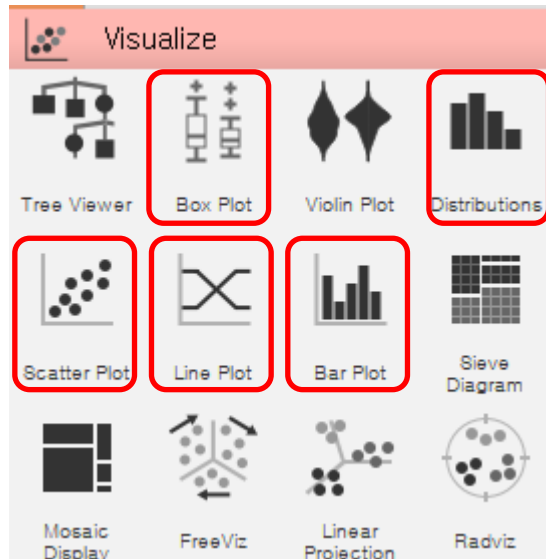
- Data Table
- Data Info
- Feature Statistics



# 인공지능 실습

## 3. 데이터 차트

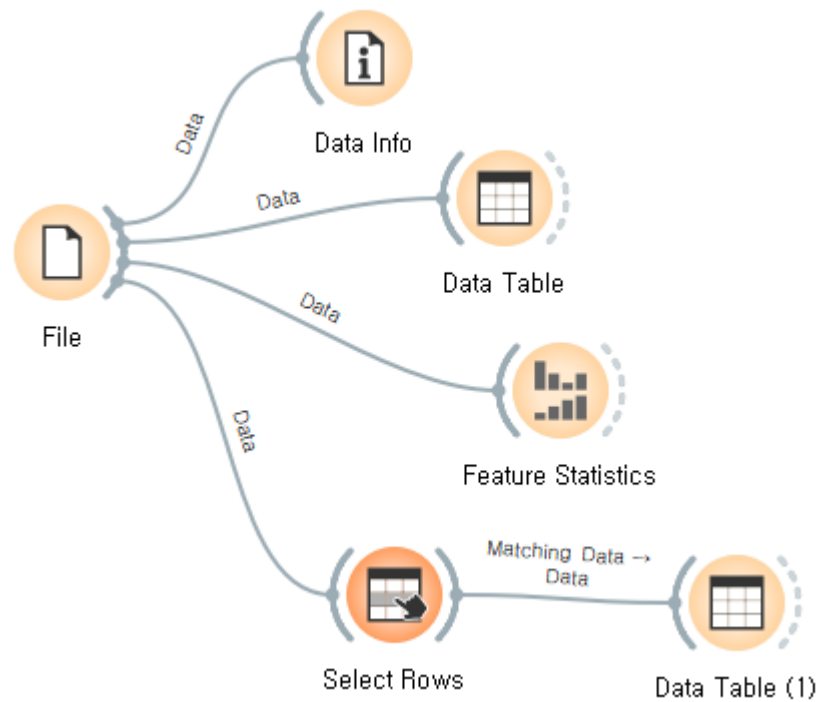
### - Visualize 위젯



# 인공지능 실습

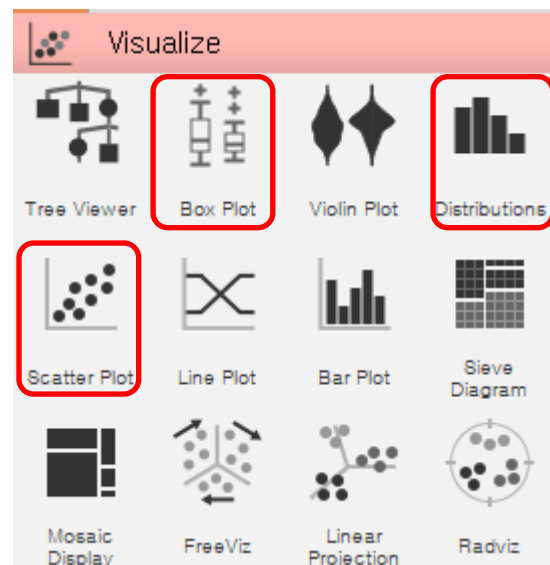
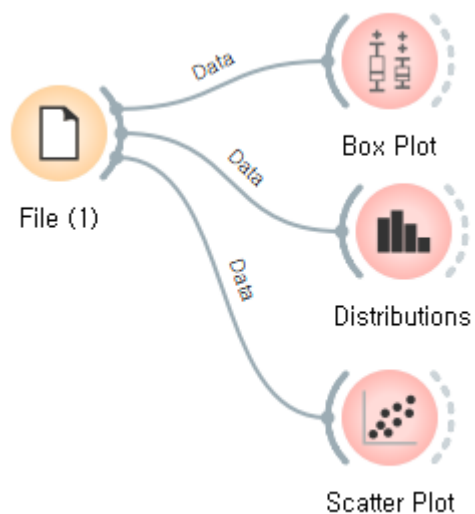
## 실습 1

- 데이터 불러오기 : math\_test.tab



# 인공지능 실습

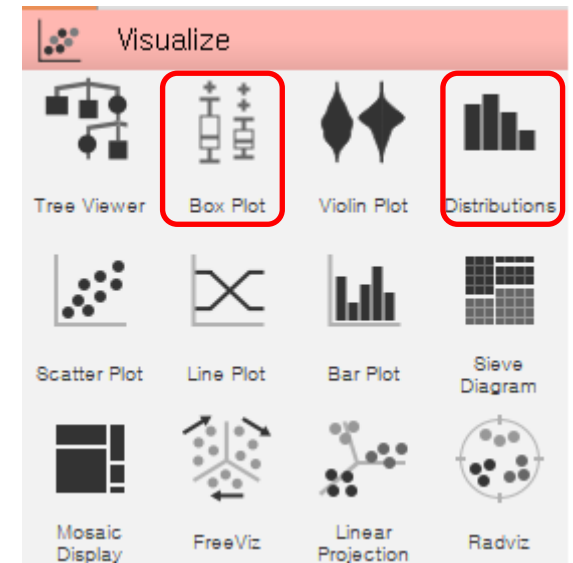
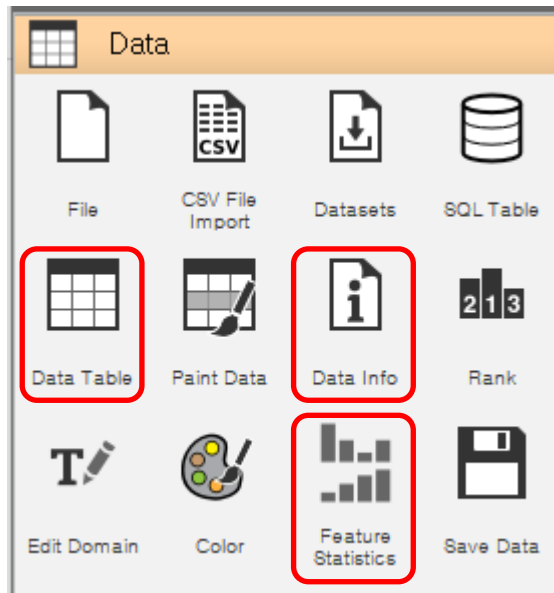
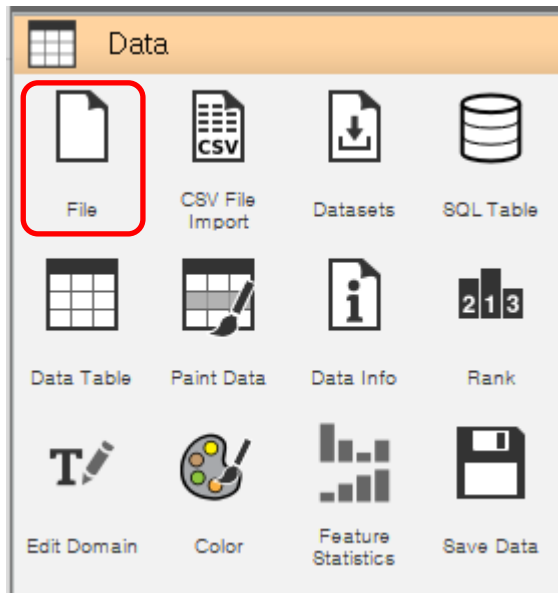
## - 차트 그리기



# 인공지능 실습

## 실습 2

- 데이터 불러오기 : '시도별 출생 성비' 읽어 오기
- 데이터 보기 : Data Table, Data Info, Feature Statistics
- 데이터 차트 : Box plot, Distributions



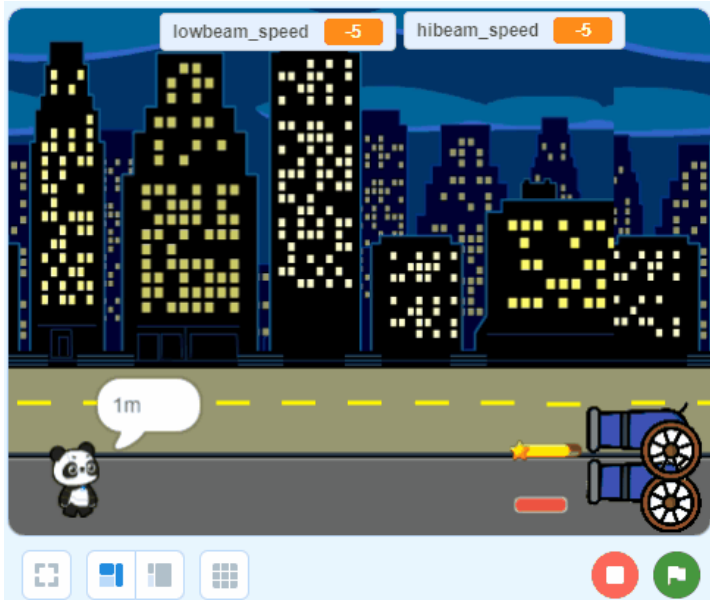
---

# MBlock 실습

---

# 게임 만들기

## 총알 피하기 게임 (최종)

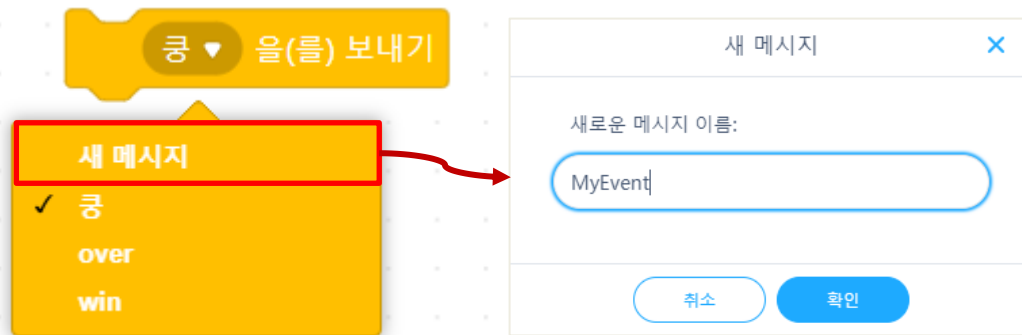


## 1단계 : 분해

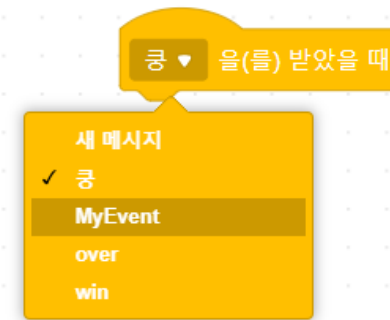
1. 펜더 객체
  - 달리기 모션
  - 점프 : 1단, 2단
  - 숙이기
2. 하단/상단 포탄 객체
  - 포탄 생성
  - 좌로 이동
  - 모양 변함
3. 배경
  - 횡스트롤
4. 대포 객체
  - 거리가 30미터이면 모양 변경
5. 게임 룰
  - 펜더가 포탄에 닿으면 수명이 줄어 듦
  - 수명이 3개 이상 줄어 들면 게임 끝
  - 펜더는 달리면서 이동 거리를 계산해서 말함
  - 거리가 30미터 이면 게임 클리어
  - 게임이 끝나면 'Game Over'
  - 게임 클리어시에는 'You win the Game' 출력

# 게임 만들기 - 사용자 이벤트 만들기

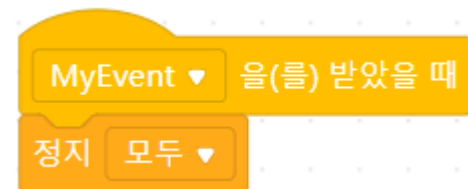
## - 이벤트 정의



## - 이벤트 반응 로직 만들기



## - 이벤트 발생 시키기





# 게임 만들기

## - 생명 관리

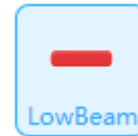
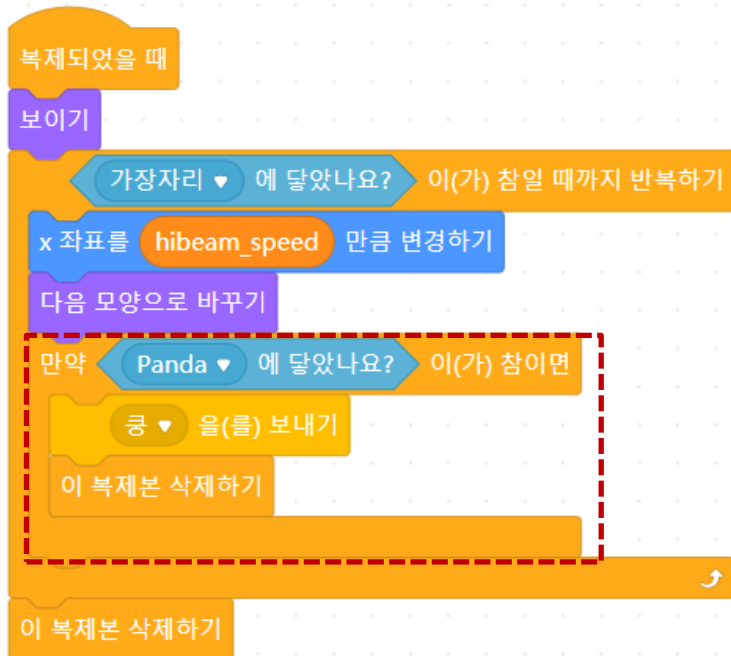
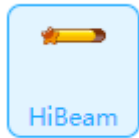


생명



# 게임 만들기

## - 포탄 맞춤 처리



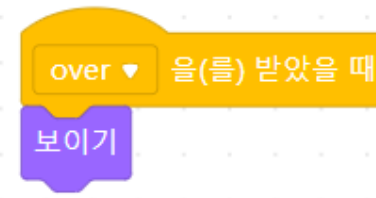
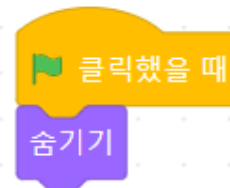
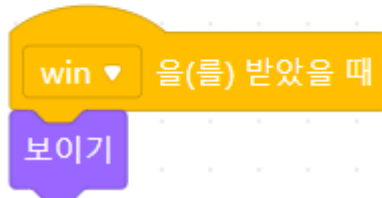
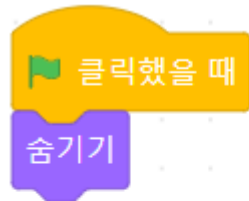
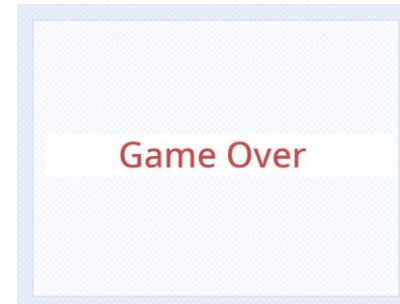
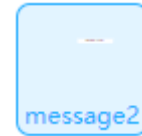
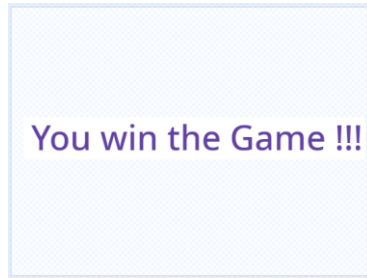
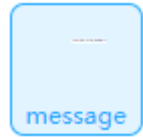
# 게임 만들기

## - 메세지



# 게임 만들기

## - 메시지 스프라이트



# 게임 만들기

## - 게임 종료

