

## 변수란 무엇인가?

- 변수 (Variable)
  - 하나의 개념을 대표하는 상징으로서 그것의 특성이 갖는 값이나 강도 또는 크기의 차이를 나타낼 수 있는 것
- 무슨 뜻인지 이해가 되나요?
- 그냥 쉽게 말하겠습니다. 일단, 변수란 변하는 숫자입니다.
  - 여기서 변하는 것은 변수의 값이 변한다는 것이겠지요.
    - 예를 들어 안유인 주임의 회사에서 고객을 상대로 상품과 서비스에 대한 만족도를 조사했습니다. 여기서 변수명은 '고객만족도' 이고, 조사한 이 변수의 값은 고객마다 다르기 때문에 값이 변할 것입니다. 그래서 "변수" 입니다.
    - 변수의 반대말은 상수 입니다. 변하지 않고 항상 그대로 있는 숫자라는 뜻이지요.

## 변수란 무엇인가?

### 설문조사지의 예

문 항	내 용	전 혀 그 렇 지 않 다			보 통 그 렇 다			매 우 그 렇 다
1	나는 이러한 경험이 즐겁다	①	②	③	④	⑤	⑥	⑦
2	나는 이러한 경험이 행복하다	①	②	③	④	⑤	⑥	⑦
3	나는 이러한 경험에 만족한다	①	②	③	④	⑤	⑥	⑦
4	나는 다음번에 다른 호텔에 머물겠다	①	②	③	④	⑤	⑥	⑦
5	나는 이 호텔에 다시는 방문하지 않겠다	①	②	③	④	⑤	⑥	⑦
6	나는 앞으로도 계속 이 호텔에 방문할 것이다	①	②	③	④	⑤	⑥	⑦

### 설문조사 결과를 코딩한 예

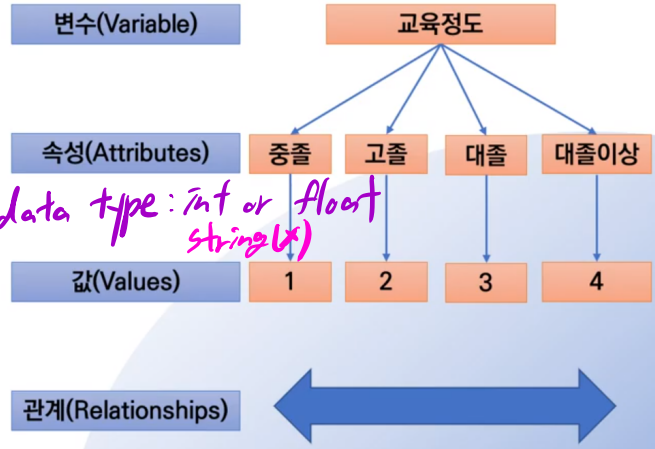
	A	B	C	D	E	F	G	H	I
1	id	v1	v2	v3	v4	v5	v6	v7	v8
2	1	1	7	1	6	6	5	Male	50
3	2	4	2	4	4	6	6	Male	21
4	3	2	2	2	7	4	4	Female	23
5	4	5	1	3	7	7	7	Male	24
6	5	5	5	6	7	7	7	Female	23
7	6	5	3	2	1	6	6	Female	26
8	7	1	1	1	7	7	7	Male	26
9	8	2	2	2	6	6	6	Female	29
10	9	2	2	1	7	6	6	Male	29
11	10	2	1	3	4	3	3	Male	33
12	11	5	5	5	3	3	3	Female	25
13	12	5	2	2	3	5	3	Male	26
14	13	5	5	6	3	3	4	Female	25
15	14	4	4	4	2	6	2	Male	36
16	15	5	5	4	6	6	5	Male	31
17	16	3	3	4	4	3	3	Male	29
18	17	3	2	3	5	5	5	Female	29

관찰값

# 변수란 무엇인가?

• 변수의 예를 들면 우측과 같습니다.

- 우리에게 교육정도라는 변수명이 있습니다.
- 이 변수는 하위에 4가지 수준의 속성을 갖고 있습니다. 바로 중졸/고졸/대졸/대졸이상입니다. 물론 여기서 속성은 연구자의 의지에 따라 조금씩 바뀔 수는 있습니다.
- 이렇게 4가지의 속성을 그대로 가지고 분석하기는 어렵습니다. 우리의 통계프로그램은 숫자만을 인식하고, 분석의 기초는 수학적인 것이기 때문에 각 속성에 값을 부여합니다.
- 부여된 값들 사이에는 어떠한 관계가 있을 수도 있고 없을 수도 있습니다.



← 변수

교육정도
중졸 → 1
고졸 → 2
대졸 → 3
대졸 이상 → 4

X (처럼, 변수의 타입과 레깅 변수에 들어있는 데이터의 타입이 일치한다!!!)

## 01. 데이터

데이터(Data)란 무엇일까요? 사전적 정의를 살펴보면 다음과 같습니다.

1. **재료, 자료, 논거**라는 뜻인 Datum의 복수형이다.
2. 컴퓨터 용어로는 정보를 작성하기 위해 필요한 자료를 뜻한다.
3. **컴퓨터에 입력하는 기호, 숫자, 문자를 말하며** 그 자체는 단순한 사실에 불과하지만 컴퓨터에 의해서 일정한 프로그램에 따라 처리되어 특정한 목적에 소용되는 정보를 만들어낸다.

즉, 데이터는 현실 세계에서 '수집한 객관적 사실' 뿐만 아니라 추론이나 예측을 위한 근거로 기능하는 존재를 말합니다. 세상에 존재하는 모든 존재를 데이터로 정의할 수도 있습니다. 하지만 조금 엄밀히 말하면 원주율이라 불리는 '3.141592653589...'와 같은 숫자는 데이터가 될 수 없습니다. 왜냐하면 수집된 객관적 사실이기보다는, 태초부터 존재하는 그러니까 신이 만든 파라미터(Parameter)값에 해당합니다. 우리가 저 숫자를 통해서 추론이나 예측을 하는데 사용하지 않기 때문입니다.

예를들면 사람들의 신용카드 기록을 통해 다음달 매출을 예측 하거나, 현재 날씨를 보고 내일 날씨를 예상하는 것. 그리고 몇몇 사람들에게 설문조사하여 여론을 조사하는 행위들이 데이터를 통해 예측이나 추론하는 것에 해당합니다. 데이터가 추론이나 예측을 위한 근거로 기능하고 있기 때문입니다.

## 02. 데이터의 종류

데이터는 어떻게 구분할 수 있을까요? 데이터는 크게 정성적 데이터와 정량적 데이터로 구분할 수 있습니다.

1. **정성적 데이터(비정형 데이터): 언어, 문자 등 비정형 데이터와 신호(Signal), 이미지, 동영상 등의 데이터** ex. '오늘 날씨가 좋다', 음악
2. **정량적 데이터(정형 데이터): 숫자, 도형, 기호 등 정형 데이터** ex. 나이, 몸무게

정량적 데이터(정형 데이터)는 우리에게 조금 더 익숙한 데이터 형태입니다. 엑셀 파일 안에 숫자형태로 수집되어 입력된 모습을 떠올리면 가장 유사합니다. 그리고 정성적 데이터(비정형 데이터)는 문자로 이루어진 글(소설, 뉴스기사 등), 그리고 사진, 동영상, 음악, 소리 등이 대표적입니다.

## 03. 데이터의 역할

그리고 데이터는 지식을 형성하는데 매우 중요한 역할을 합니다. 잠시 언급하자면 데이터를 가공해서 우리는 정보(Information)를 만들어내고 정보에서 지식(Knowledge)을 만들어냅니다. 그러니 지식을 형성하기 위한 근본적인 존재가 바로 데이터라고 할 수 있습니다.

우리가 데이터 분석을 하는 이유도 바로 정보, 지식을 만들기 위한 것입니다. 지식의 종류는 다음과 같습니다.