

관찰, 측정을 통해 수집된 사실이나 값 수치 문자 등 가공되지 않은 원자료: 데이터

데이터가 정리 및 분석되어 의미나 가치를 갖게 된 것: 정보

- 경험이나 이해를 통해 체계화되어 문제해결이나 판단에 활용될 수 있는 형태: 지식
 - 통찰력+아이디어 결합해 해결책 도출: 지혜



데이터를 읽고 분석하며 목적에 맞게 해석 및 활용할 수 있는 능력: 데이터 리터러시

- 하위역량 5개: 수집, 관리, 가공 및 분석, 시각화, 해석

빅데이터 3V: 양, 종류, 속도

데이터 큰 분류 2개: 정형 비정형

- 표: 정형
- 사진 영상 등: 비정형

정형 데이터 변수 두 종류: 범주형, 수치형

- 범주형 두종류: 명목형, 순서형
- 수치형 변수: 정수형, 연속형

해결하고 싶은 문제가 무엇인지 명확해야 한다: O, X

직접 조사 전 고려사항 3개: 대상, 항목, 시기

관심있는 조사대상을 하나씩 무작위 선택하여 추출하는 방식: 무작위 추출, random sampling

한 질문에는 한 내용만이 아니어도 된다: O, X

유도 질문을 통하여 원하는 결과를 도출해도 된다: O, X

예에 부정적, 아니요에 긍정적 답변을 달아 집중을 유도해도 된다: O, X

민감한 질문은 뒤에 배치해서 설문조사를 하는 사람의 거부감을 낮춘다: O, X

최대한 많은 질문을 해야 하기에 최대한 많은 양의 질문을 설문지에 포함한다: O, X

설문조사 질문 형태: 자유기술, 양자택일, 다중택일, 체크리스트, 리커트 척도

데이터를 올바르게 분석 및 해석하려면 (데이터 클리닝)이 필요함

- 데이터의 (편향) 이 없도록 함



오타 발견시 빨리 찾도록 해주는 엑셀 기능: 데이터- 데이터 유효성 검사

엑셀에서 중복된 데이터 삭제할 수 있는 기능: 데이터 - 중복 값 제거

데이터가 누락되거나 비어 있는 값: 결측치

- 처리 방식 2가지: 삭제, 대체

데이터 범위에서 벌어난 아주 작거나 큰 데이터 값을 보이는 데이터: 이상치

- 처리 방법: 삭제, 대체
- (이상치를 제외한) 최대 ,최솟값으로 (대체)를 한다.

데이터를 요약하고 설명하는 도구: 통계

- 데이터의 전반적인 (경향)을 파악 가능

알고자 하는 대상 전체: 모집단

- 바탕으로 계산된 거로가: 모수

모집단을 적절히 대표하는 모집단의 일부" 표본

- 바탕으로 계산된 결과: 통계량

데이터 자체의 성질을 기술하고 요약하는 통계량: 기술통계량

- 구하는 분석방법: 기술통계 분석
- 대표값 종류: 최빈값, 중앙값, 평균값
- 데이터의 퍼짐 정도 나타내는 값: 분산, 표준편차
- 데이터의 범위 나타내는 값: 최대값, 최소값

분산을 제공근 한 값: 표준편차

표준편차가 같으면 변동성이 같다: O, X

표준편차를 평균으로 나눈 비율, 상대적 변동성 측정방법: 변동계수, CV

주어진 데이터 집합에서 가장 크고 작은 값: 최대값, 최소값

- 범위: 최대값 - 최소값

기술통계량	
최빈값	=MODE(셀범위)
중앙값	=MEDIAN(셀범위)
평균	=AVERAGE(셀범위)
분산	=VAR.P(셀범위) → 도 =VAR.S(셀범위) → 표
표준편차	=STDEV.P(셀범위) → 도 =STDEV.S(셀범위) → 표
최대값	=MAX(셀범위)
최소값	=MIN(셀범위)

중심축을 기준으로 우리가 보고 싶은 항목을 배치, 정렬해주는 테이블: 피벗테이블

데이터 안에서 보고싶은 정보만 따로 뽑아서 살펴볼 수 있는 기능: 필터

필터 / 열

행 / 값

두 변수가 얼마나 관계성이 있는지 개념: 상관관계

- 같거나 다른 (방향) 으로 변하는지 살펴봄
- 시각화: 산점도
- 방향에 따른 값: 양, 음

방향의 일관성: 힘

상관관계 정도를 나타내는 수치: 상관계수

- 범위: -1 ~ +1
- 방향: 부호
- 힘: 절대값

절대값이 1일수록 힘이 세고 0일수록 약하다: **O**, X

상관관계가 존재하면 무조건 인과관계가 있다: **O**, **X**

강한 상관관계의 기준: 0.6

데이터 전반적 경향을 나타내는 선: 추세선

함수: COLLEL(셀범위1, 셀범위2)