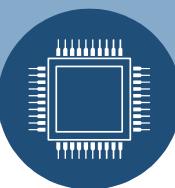


탐색적 데이터 분석



경남대학교 창의융합대학 교수 유현주





강의 진행 안내

- 수업 자료 공유 URL

- <https://github.com/comjoo94/>

- 수업에 필요한 강의 자료 다운로드 저장

- 소통을 위한 메일 주소: comjoo@uok.ac.kr

- comjoo@gmail.com



Contents



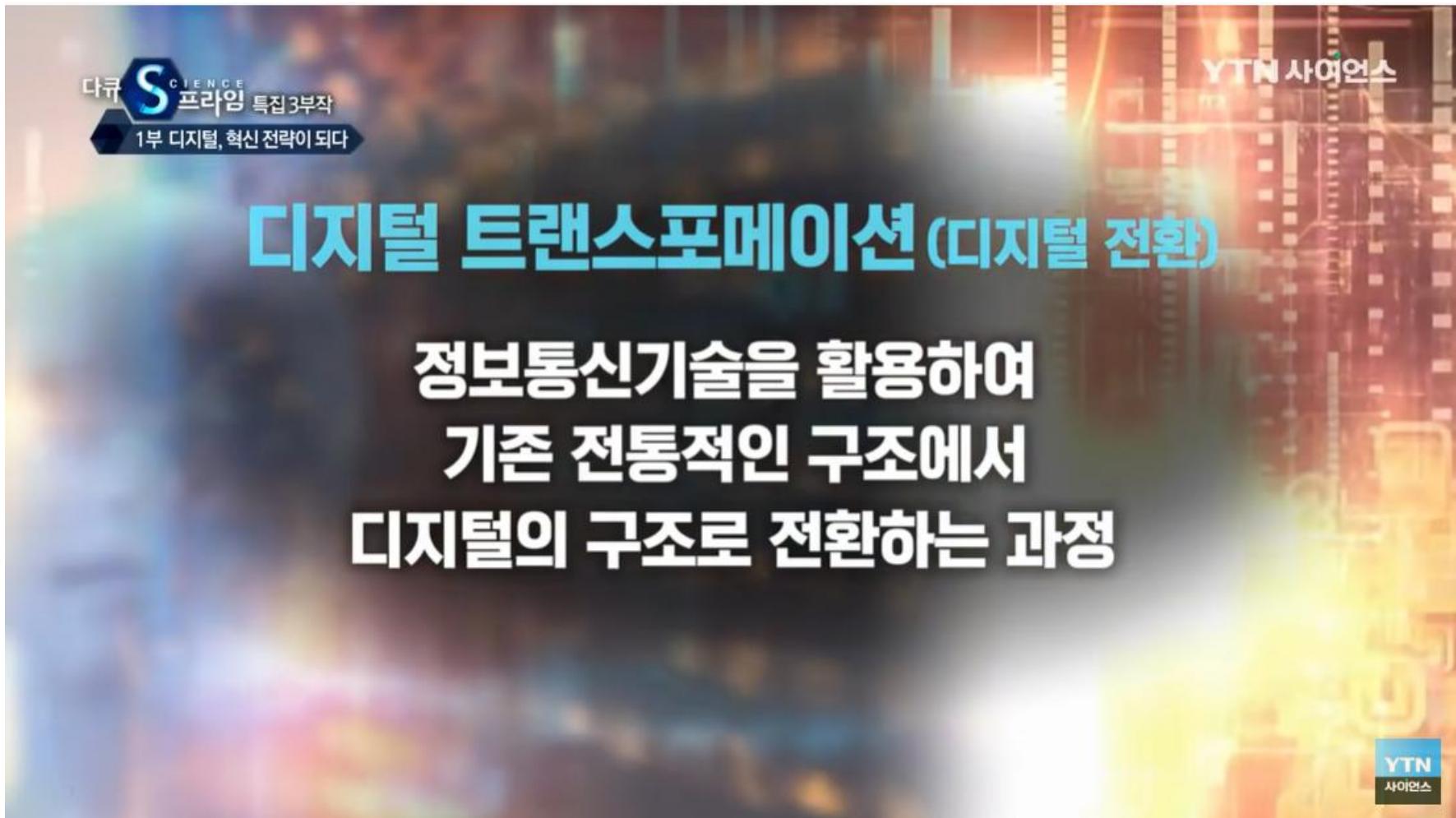
탐색적 데이터 분석

- 데이터 리터러시
- 데이터 수집
- 데이터 전처리
- 탐색적 데이터 분석
 - 피벗테이블
 - 기술 통계
 - 가설 검정
 - 상관 분석
 - 회귀 분석



데이터 리터러시







디지털대전환 시대



디지털 트랜스포메이션 시대
“데이터는 석유같은 자원이다.”

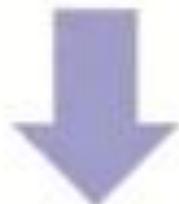




디지털대전환 시대

- 데이터 중심 애플리케이션(응용 소프트웨어) 개발

SW1.0 : Code In -> SW Out



SW2.0 : Data In -> SW Out

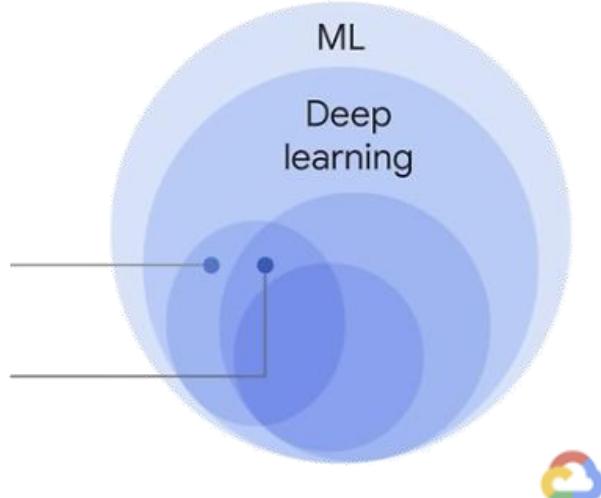
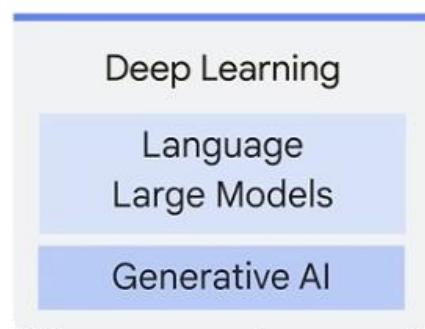


AX, beyond DX

- 인공지능 AI(Artificial Intelligence) Transformation



Large Language
Models (LLMs)
also intersects
with Generative AI





데이터를 먹고 사는 인공지능

- 인공지능 모델은 '코드'와 '데이터'로 이루어짐
 - 앤드류 응(Andrew Ng) 스탠퍼드대 교수(컴퓨터 과학자이자 인공지능 분야의 선구자)
- 인공지능 모델 구현 “코드”를 자동차에 비유한다면, “데이터”는 자동차를 움직이는 석유

Conventional model-centric approach:

$$\text{AI} = \text{Code} + \text{Data}$$

(algorithm/model)

Work on this

Data-centric approach:

$$\text{AI} = \text{Code} + \text{Data}$$

(algorithm/model)

Work on this

AI = Code + Data (출처: DeepLearningAI)



데이터를 먹고 사는 인공지능

● Improving the code vs. the data

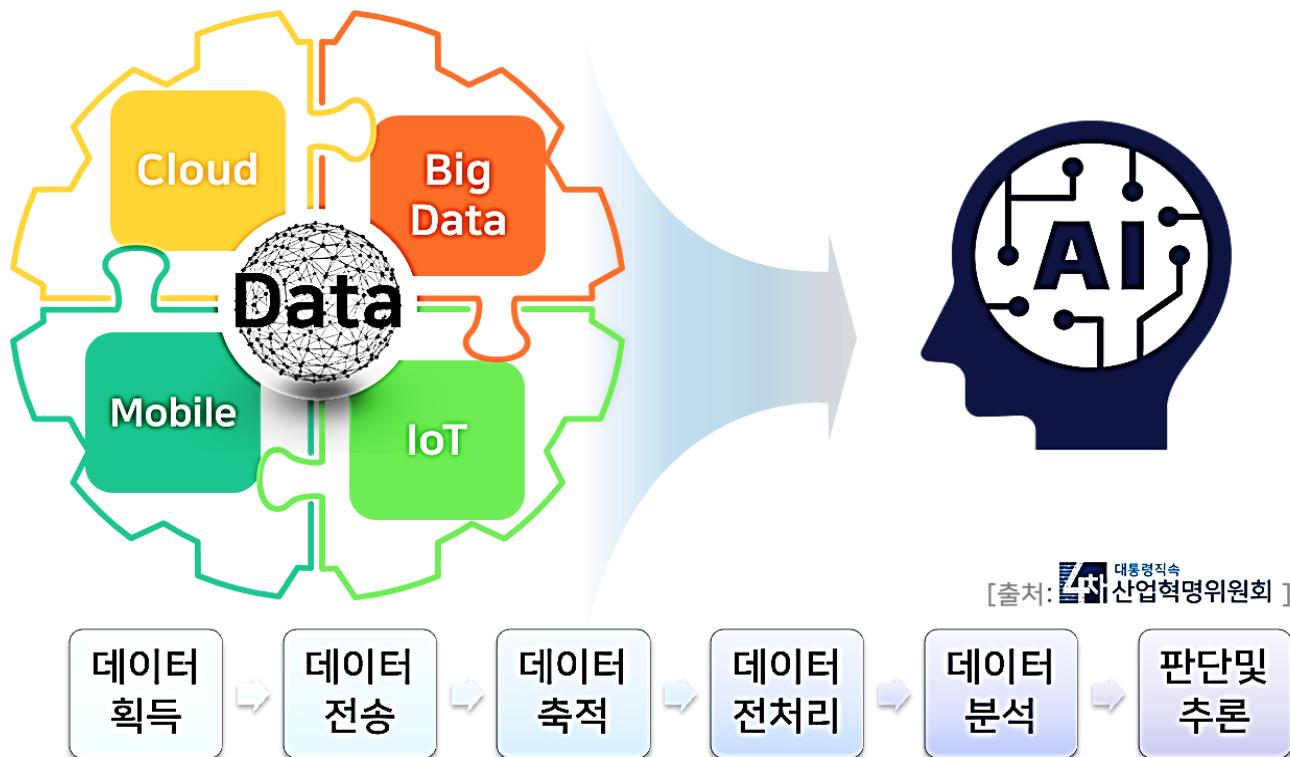
	강철 오류 검출 Steel defect detection	태양 패널 Solar panel	표면 검사 Surface inspection
기본 모델 성능 Baseline	76.2%	75.68%	85.05%
모델(코드) 기반 성능개선 Model-centric	+0% (76.2%)	+0.04% (75.72%)	+0.00% (85.05%)
데이터 기반 성능개선 Data-centric	+16.9% (93.1%)	+3.06% (78.74%)	+0.4% (85.45%)

- 인공지능 모델 성능 개선을 위한 모델(코드) 중심 개선과 데이터 중심 개선 성능 비교

인공지능대전환 시대의 석유 ≈ 데이터(Data)



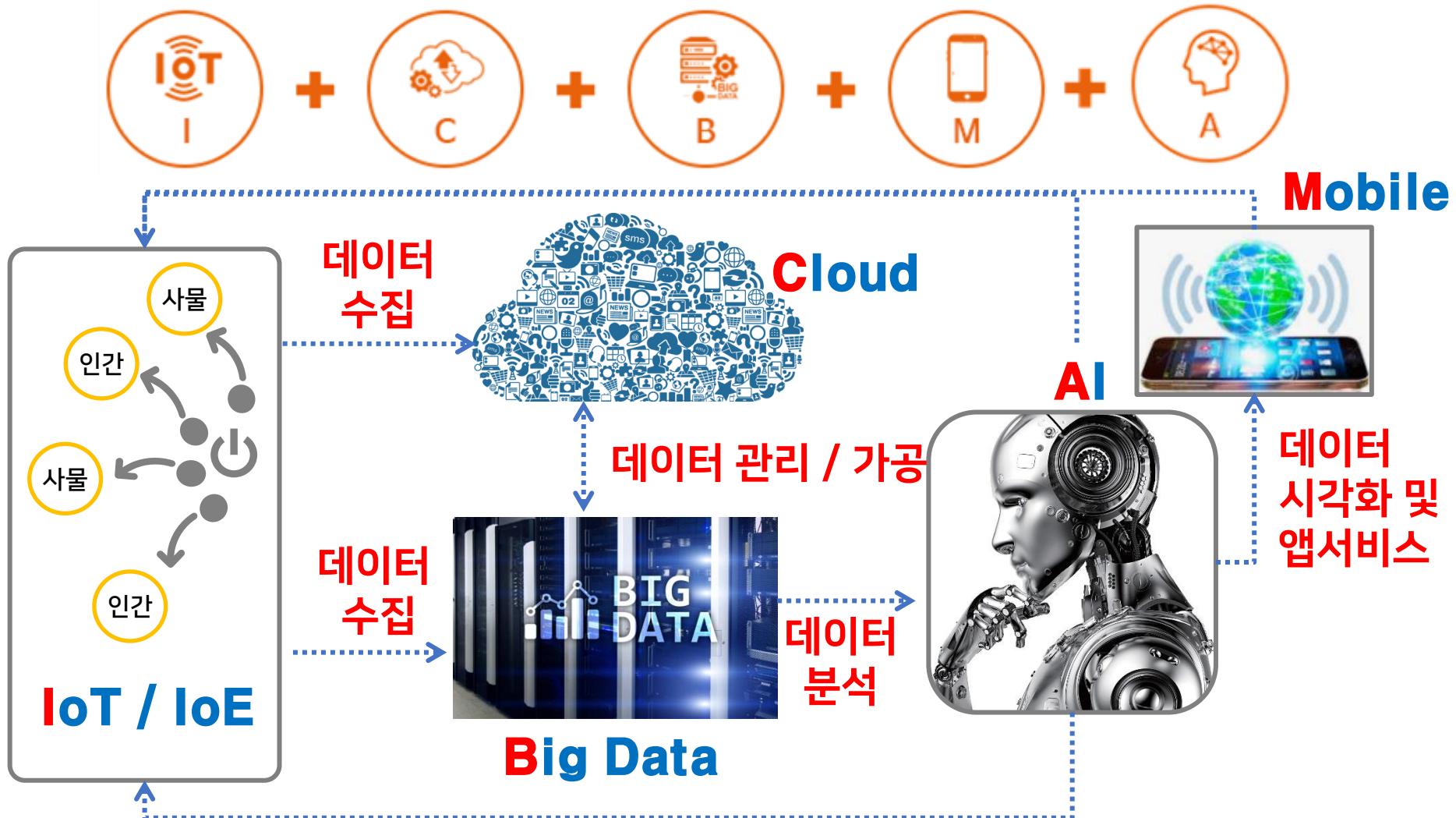
- 산업현장에서 갖추어야 할 필수 능력으로 데이터를 다루는 역량이 매우 중요
- 디지털대전환시대의 핵심 기술

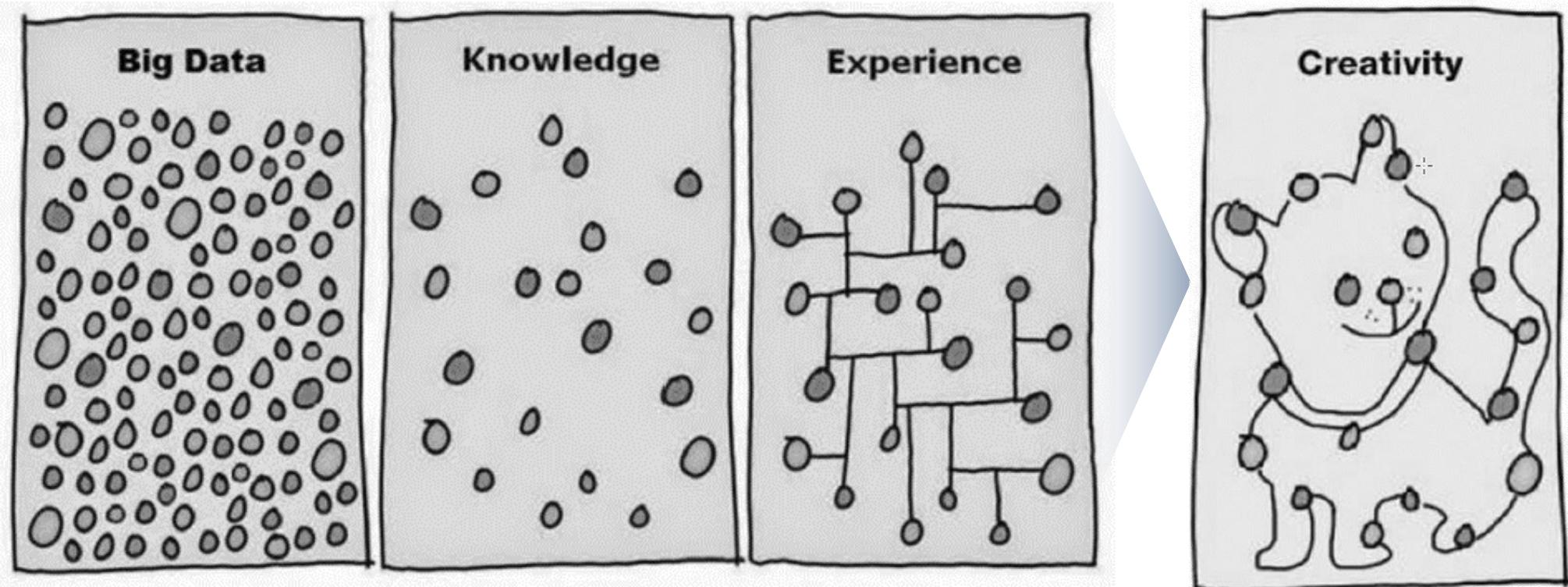


데이터를 다루는 핵심 기술: ICBMA



- 데이터를 다루는 4차산업혁명 핵심기술의 관계





[출처: 행복중심 4차산업혁명 전략 – GIST 발표자료 (2017.4.13)]



데이터 리터러시

● 데이터 리터러시 (Data Literacy)?

※ 리터러시(Literacy) : 글을 읽고 해독하고 활용하는 능력

- 데이터를 목적에 맞게 생성하고 데이터에 숨어있는 의미를 찾아 이해하고 해석된 결과를 일에 적용하고 소통하는 모든 행위
- 데이터의 숨겨진 의미를 읽어내고 인사이트를 얻는 일
- 분야의 구분 없이 향후 가장 중요한 개인의 역량으로서 **데이터 리터러시가 핵심으로 부각되고 있음**
- 데이터 분석 시대를 맞이하여 리더와 전문가 영역만의 능력에서 확장, 현대 산업 현장의 종사자 모두 데이터 활용 능력이 필요

데이터로 일하는 법

감에 의존하던 시대는 끝났다. 실패를 덜고, 기회비용을 줄이려면 데이터 중심의 업무를 익힐 필요가 있다.

BY 에디터 장혜정 | 2021.10.11

얼마 전 모 건설회사 신입사원 면접을 보고 온 김철수 씨는 데이터 활용 능력의 필요성을 절감했다. 면접 당일 회사에서 나눠준 A4용지 한 장에는 '2020년 인구주택총조사' 자료가 담겨 있었다. 수도권 인구 및 비율, 시도별 노후 주택의 숫자 등이 빼곡한 종이를 받아 들고 '까만 건 숫자요 하얀 건 종이'라며 난감해하던 김철수 씨는 '이 데이터를 보고 앞으로 건설 시장이 지향해야 하는 바와 그 이유를 설명하시오'란 면접관의 주문에 입 한번 떠보지 못하고 떨어졌다.

[출처] Singleplus

<https://m.thesingle.co.kr/article/712325/THESINGLE>



Data Literacy





데이터를 다루는 다양한 플랫폼 및 소프트웨어

● 데이터 분석 도구 선택 고려 사항

- 수집된 데이터를 분석에 적합한 형태로 표현하고 구성하는 데 제약이 적은 데이터 분석 도구 여부
- 처리 가능한 데이터의 양과 처리 속도를 비교하여 평가
- 다양한 연산, 집계 및 통계 기법, 분석기법 제공 여부
- 결과의 가독성과 이해도를 높일 수 있는 기능, 다양한 시각화 방법과 편리한 편집 기능 제공 여부
- 라이선스 취득 가능 여부 등...

- 데이터 수집 및 관리
- 데이터 전처리
- 데이터 분석
- 데이터 시각화



대표적인 데이터 문서 작성 프로그램: Excel



● Microsoft Excel

- 마이크로소프트 社 스프레드시트 프로그램
- 사무자동화(OA) 데이터 문서 작성 소프트웨어
- 계산 위주의 문서 작업에 탁월한 기능 제공
- 요약 집계 기능, 기초 통계 기능, 차트 기능 제공으로 편리
- 다른 데이터 분석 소프트웨어와 비교하면 제약사항이 다소 존재함
- 버전 업데이트로 데이터 전처리 및 분석을 위한 추가 기능 제공
- 데이터 분석 전문 외부 소프트웨어와 호환이 쉽고 편리함
- 전세계적으로 가장 대중적이며 사용자가 많은 데이터 관리 프로그램



Excel로 실습하는 데이터 분석 기초



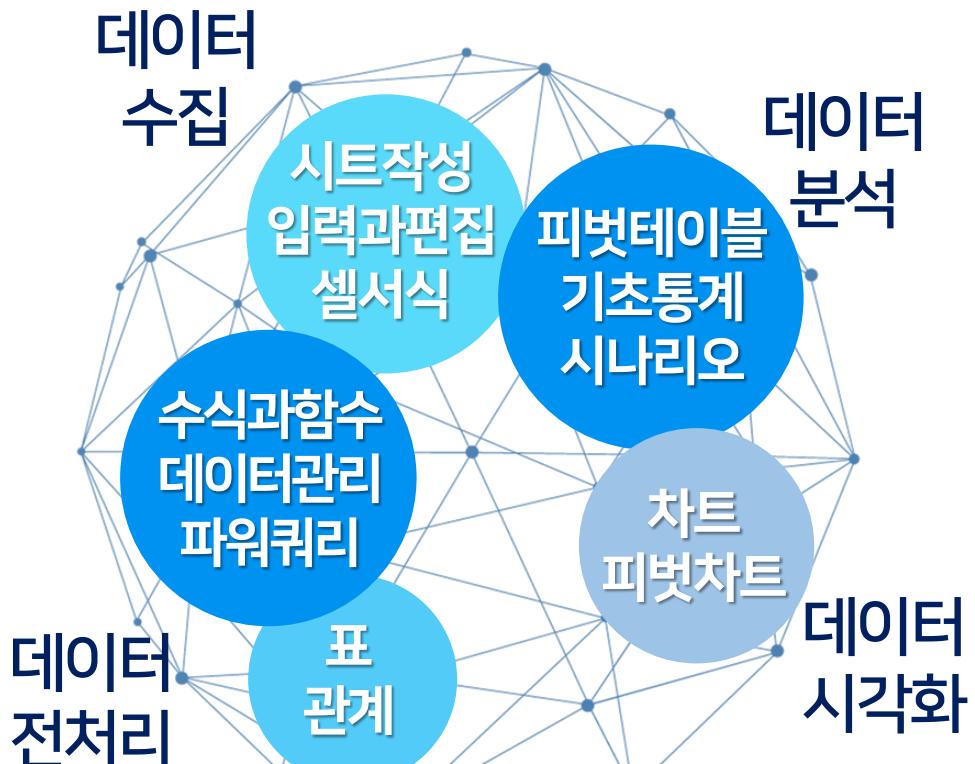
- Excel의 주요 기능

- 표 형식의 문서 작성
- 수식과 함수로 계산 값 산출
- 데이터 셋 정리 및 관리
- 값 비교 및 분포 파악을 위한 차트 작성
- 매크로 자동화 코딩 가능

- 데이터 리터러시

- 데이터 수집
- 데이터 전처리
- 데이터 분석
- 데이터 시각화

Excel + Data Literacy





대표적인 데이터 프로그래밍 도구: Python

- 데이터 과학 언어 생태계에서 파이썬을 선호
- 스크립트 작성과 프로세스 자동화, 웹 개발, 일반 애플리케이션 등 여러 소프트웨어 개발 영역에서 오랫동안 인기있는 범용 프로그래밍 언어
- 간단하면서도 직관적인 문법 구조
- 다양한 데이터 구조 제공
- 다른 언어 또는 데이터베이스와의 호환성이 좋음
- 다수의 라이브러리와 확장 패키지제공
- 최근 통계 라이브러리가 대거 추가
- 처리 속도가 비교적 빠르지 않음



- ① 단순 프로그래밍
- ② 풍부한 라이브러리
- ③ 느린 파이썬 그래도 좋다
- ④ 어플리케이션 개발

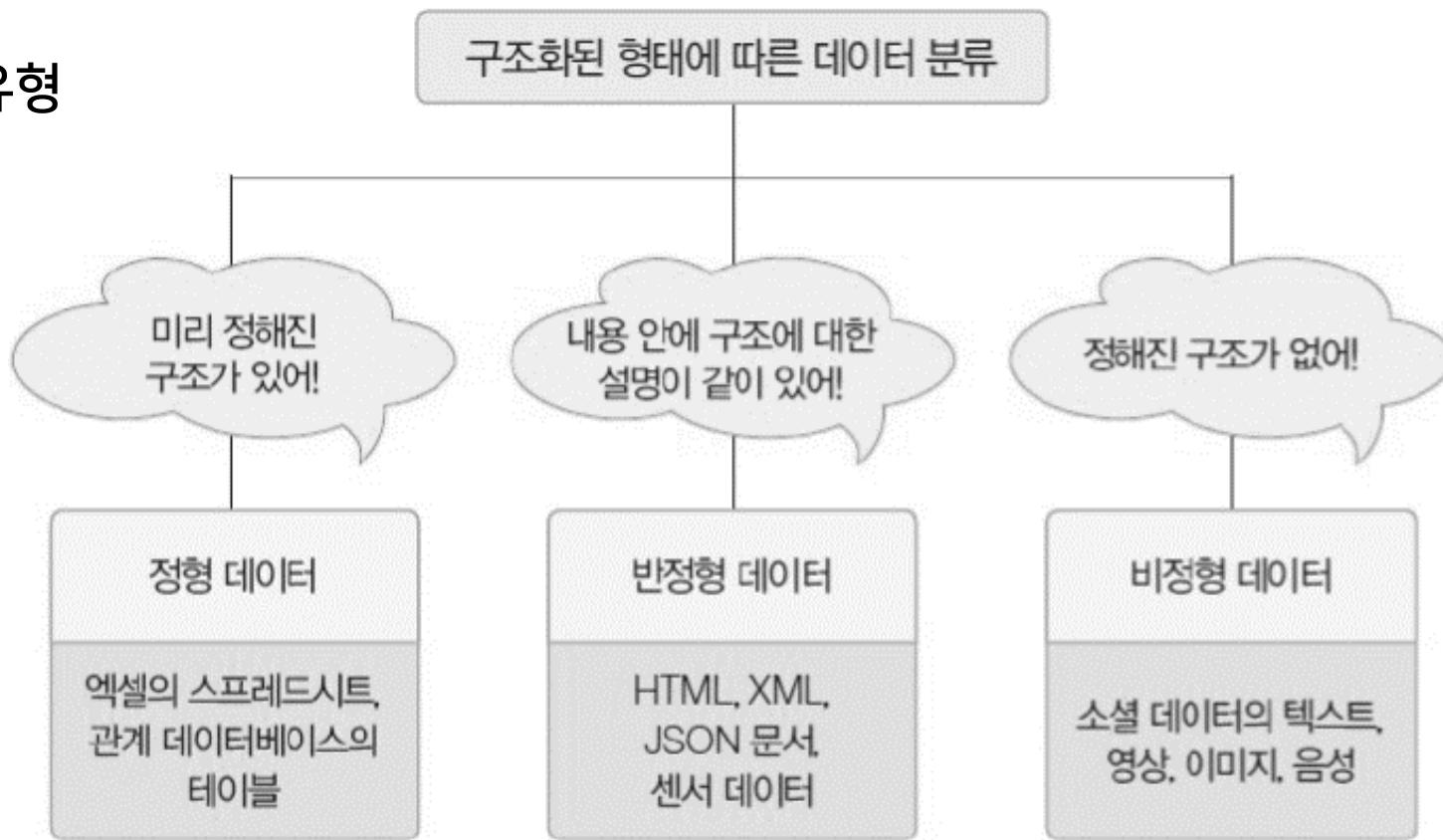


데이터

● 데이터 정의

- 자료(資料, Data, 데이터)는 수, 단어 등의 형태로 된 의미 단위
- 바탕이 되는 재료, 자료를 의미 있게 처리하여 정보(Information)로 활용

● 데이터 유형





데이터 유형

● 정형데이터

- 정형데이터(Structure data)는 미리 정해진 형식으로 구조화된 데이터
- 행과 열로 표현되는 데이터를 저장하는 엑셀 시트나 RDBMS 테이블이 정형 데이터
- 개체의 속성을 관측한 결과를 데이터시트로 표현한 것

변수명	항목ID	측정값	변수명
요소	1	11.29	변수
요소	2	11.68	변수
요소	3	11.89	변수
요소	4	12.16	변수
요소	5	12.76	변수



데이터 유형

● 반정형데이터

- 반정형데이터(Semi-structure data)는 특정한 형식에 따라 저장된 데이터이지만 정형데이터와 달리 형식에 대한 설명을 함께 제공
- 구조를 해석하는 파싱(일종의 번역) 과정이 필요하며 파일 형태로 저장
- XML, JSON 등이 대표적인 사례

XML 사례	JSON 사례
<교과정보> <교과명> 데이터분석 </교과명> <시간> 45 </시간> <담당교수> 유현주</담당교수> </교과정보>	{ "교과명" : "데이터분석" "시간" : 45 "담당교수" : "유현주" }



데이터 유형

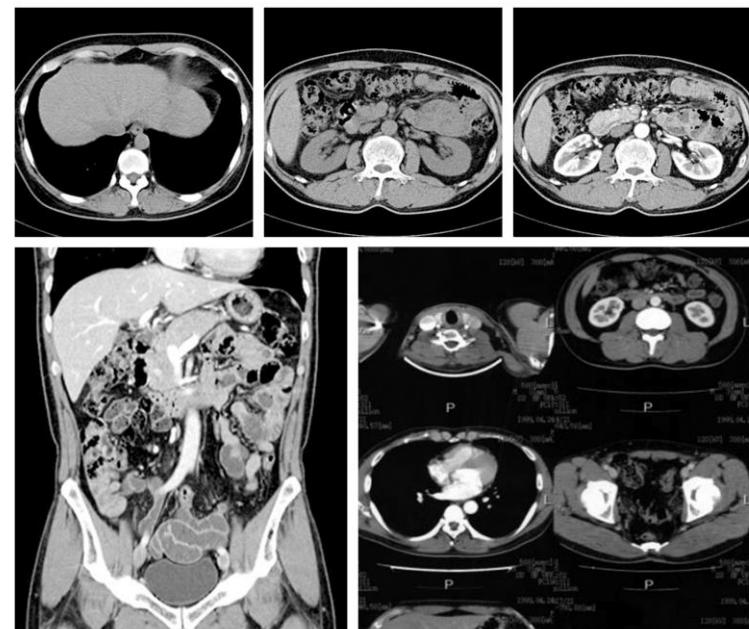
● 비정형데이터

- 비정형데이터(Unstructured data)는 일괄적으로 정해진 구조가 없고 각각 데이터 형식을 가지는 데이터
- 빅데이터의 대부분을 차지하는 텍스트, 영상, 이미지SNS등

```
twitter_analytics — python2.7 — 80x24

statuses_count":24096,"created_at":"Wed Nov 23 23:34:19 +0000 2011","utc_offset": -14400,"time_zone":"Eastern Time (US & Canada)","geo_enabled":true,"lang":"en","contributors_enabled":false,"is_translator":false,"profile_background_color":"191919","profile_background_image_url":"http://pbs.twimg.com/profile_background_images/738070196/237c3eb2ccbf14a8df528a80986a8676.jpeg","profile_background_image_url_https":"https://pbs.twimg.com/profile_background_images/738070196/237c3eb2ccbf14a8df528a80986a8676.jpeg","profile_background_tile":false,"profile_link_color":"009999","profile_sidebar_border_color":"FFFFFF","profile_sidebar_fill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image":true,"profile_image_url":"http://pbs.twimg.com/profile_images/456615890025340928/Qo_JEm96_normal.jpeg","profile_image_url_https":"https://pbs.twimg.com/profile_images/456615890025340928/Qo_JEm96_normal.jpeg","profile_banner_url":"https://pbs.twimg.com/profile_banners/419918470/1404682685","default_profile":false,"default_profile_image":false,"following":null,"follow_request_sent":null,"notifications":null,"geo":null,"coordinates":null,"place":null,"contributors":null,"retweet_count":34,"favorite_count":30,"entities":{"hashtags":[],"trends":[],"urls":[],"user_mentions":[],"symbols":[]},"favorited":false,"retweeted":false,"possibly_sensitive":false,"filter_level":"low","lang":"it"}, "retweet_count":0,"favorite_count":0,"entities":{"hashtags":[],"trends":[],"urls":[],"user_mentions":[{"screen_name":"vittoriozucconi","name":"Vittorio Zucconi","id":419918470,"id_str":"419918470","indices":[3,19]}],"symbols":[]}, "favorited":false,"retweeted":false,"possibly_sensitive":false,"filter_level":"medium","lang":"it"}
```

복부 CT





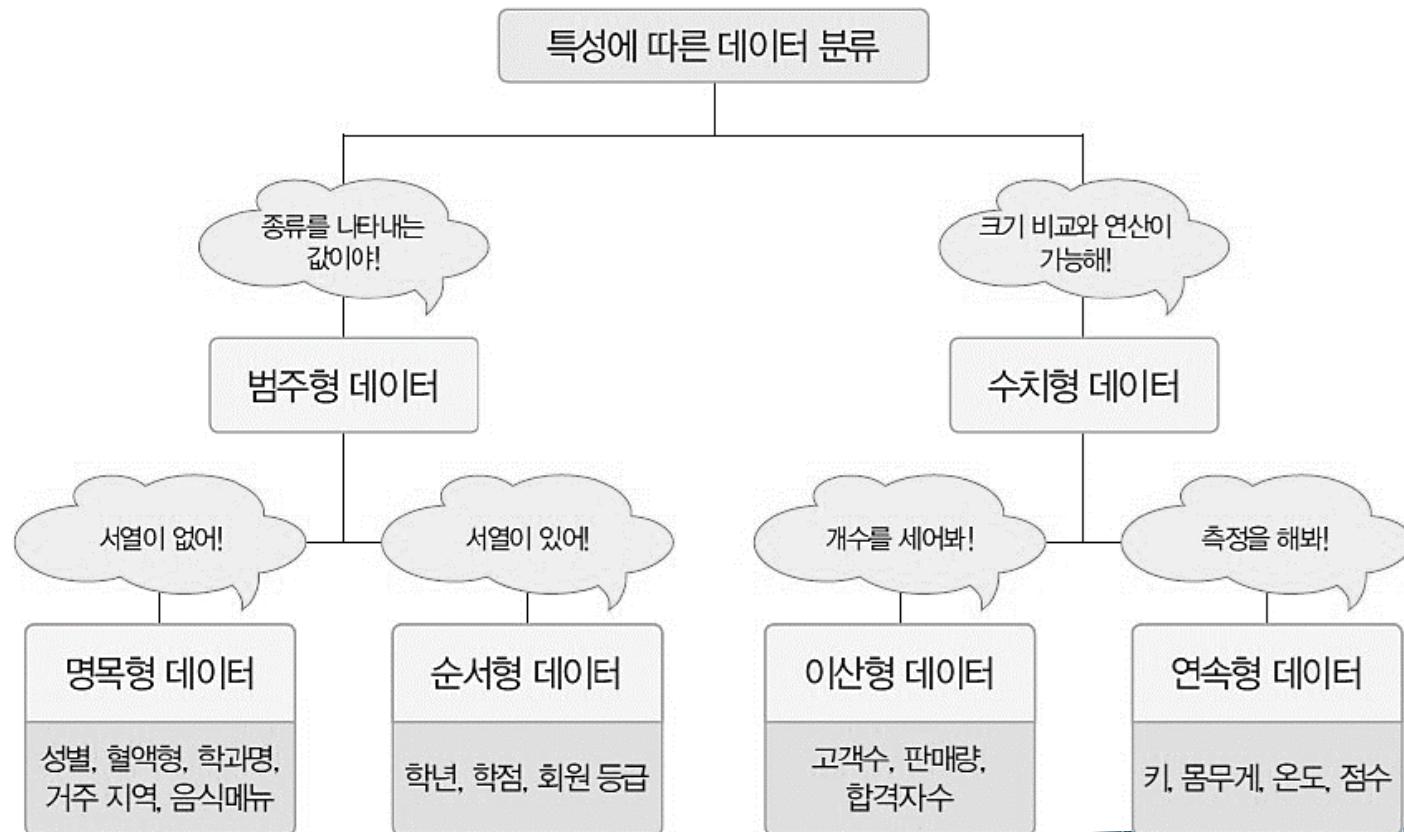
정형 데이터 종류

● 양적 데이터(quantitative data, 정량적 자료)

- 수치로 측정이 가능한 자료, 수치 계산하거나 대소 관계 비교 등의 작업 수행

● 질적 데이터(qualitative data, 정성적 자료)

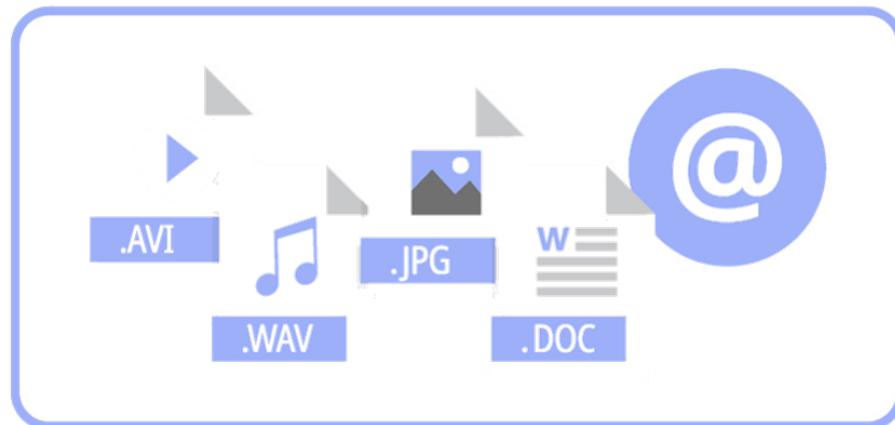
- 수치로 측정이 불가능한 자료, 자료를 범주로 나누어 주로 분류 등의 작업 수행





비정형 데이터 종류

- 오디오 또는 비디오 데이터, 보안 감시 데이터, 지리 공간 데이터, 이미지 및 날씨 데이터와 같은 리치 미디어
- 장치의 티커 또는 센서 데이터와 같은 사물 인터넷(IoT) 데이터
- 이메일, 문자 메시지, 인보이스, 레코드 및 생산성 애플리케이션 통신 데이터와 같은 텍스트 데이터
- 기계가 생성한 우주 탐사 또는 지진 보고서와 같은 과학적 데이터
- MRI, X선 및 CT 스캔과 같은 의료 데이터와 영상, 의사의 노트 및 처방전과 같은 의료 데이터 및 이미징
- 기타 다양한 형식





데이터 단위

- 신호를 나타내는 최소 단위 '비트(bit)'
- 정보의 기본단위 '바이트(Byte)'



비트 (binary digit, bit)	바이트 (Byte)
0과 1, 두 가지 값만 가질 수 있는 측정 단위	여덟 개의 비트로 구성된 데이터의 양을 나타내는 단위
0 OFF FALSE 1 ON TRUE	

● 데이터 단위

※ 이해를 돋기 위해

2진수 단위를 10진수로 대략 환산하여 →
단위 비교함

$$\approx 2^{10} = 1024 \rightarrow \text{약 } 1000$$

Yotta	Y	10^{24}	= 1 000 000 000 000 000 000 000 000
Zetta	Z	10^{21}	= 1 000 000 000 000 000 000 000 000
Exa	E	10^{18}	= 1 000 000 000 000 000 000 000 000
Peta	P	10^{15}	= 1 000 000 000 000 000 000 000 000
Tera	T	10^{12}	= 1 000 000 000 000
Giga	G	10^9	= 1 000 000 000
Mega	M	10^6	= 1 000 000
Kilo	K	10^3	= 1 000
Hecto	H	10^2	= 100
Deca	Da	10^1	= 10



데이터 규모

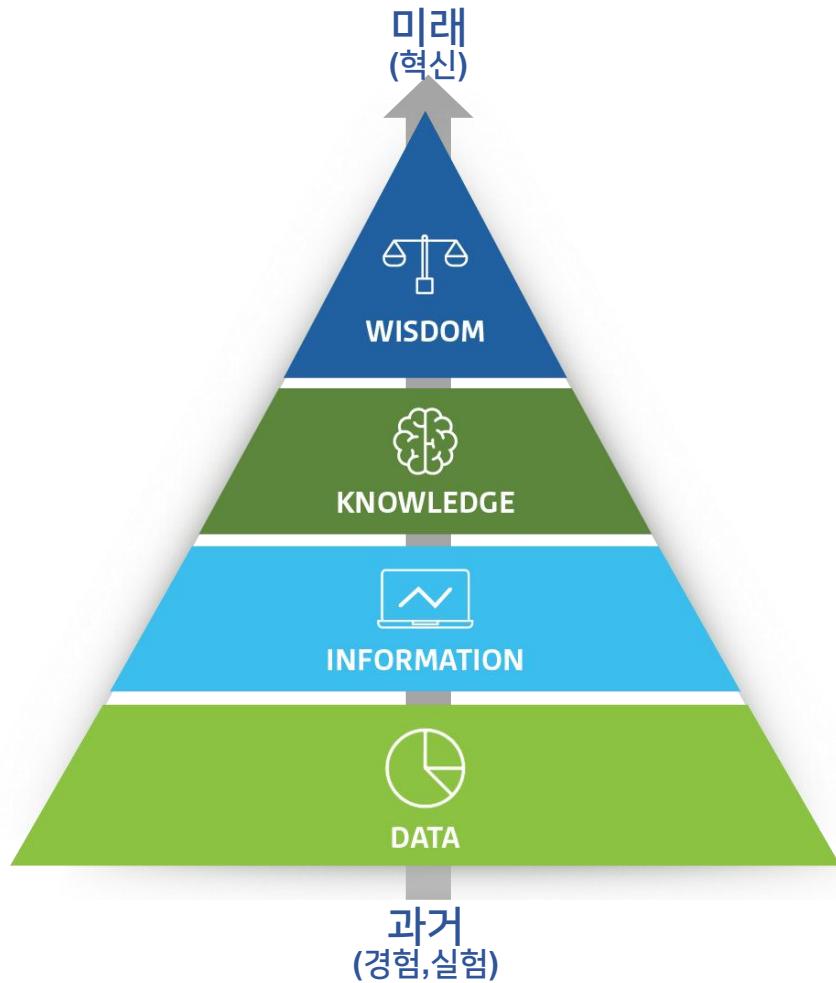
- 최근, 데이터는 기본적으로 Big Data를 말함
- 현실적으로 PC 단위 작업 데이터 규모는 Small Data라고 할 수 있음
- 서버 단위의 관계데이터베이스관리시스템(RDBMS) 등으로 관리하고 SQL로 운용하는 규모를 중간 데이터라고 말하기도 함
- 규모는 차이가 있으나 데이터 작업 방식은 Big Data 방식을 주로 적용
- 하나의 주제(맥락)로서 연관되는 데이터 묶음의 크기별로 구분

구분	데이터 크기
작은 데이터	노트북 컴퓨터 메모리에 저장할 수 있는 크기 : 10 GB 이하
중간 데이터	서버 컴퓨터 메모리에 저장할 수 있는 크기 : 10 GB – 5 TB
매우 큰 데이터	컴퓨터 한대 메모리에 저장될 수 없는 크기 : 5 TB 이상



데이터, 정보, 지식, 지혜 (DIKW)

- DIKW(Data, Information, knowledge, Wisedom) 피라미드
- 문헌 정보 관리, 정보 시스템, 지식 관리 영역에서 이용되는 체계도



단계	설명
지혜 (Wisdom)	<ul style="list-style-type: none">• 지식에 유연성을 더하고, 상황이나 맥락에 맞게 규칙을 적용하는 것• 근본원리에 대한 깊은 이해를 바탕으로 도출되는 창의적 아이디어 <p>ex) A마트의 다른 상품들도 B마트보다 쌀 것이라고 예측</p>
지식 (Knowledge)	<ul style="list-style-type: none">• 정보를 일반화하고 체계화하여 즉시 적용 및 활용 가능한 형태• 상호 연결된 정보 패턴을 이해하여 이를 토대로 예측한 결과물 <p>ex) 연필을 살 때는 A마트로 가는 것이 유리</p>
정보 (Information)	<ul style="list-style-type: none">• 데이터 중 사용자가 필요로 하는 데이터• 사용자의 필요에 의해 정제 및 가공된 데이터 <p>ex) A마트의 연필가격이 더 저렴함</p>
데이터 (Data)	<ul style="list-style-type: none">• 관찰, 측정을 통해서 수집된 사실이나 값, 수치, 문자 등 가공되지 않은 원본 데이터 <p>ex) A마트 펜 500원, 연필 200원, 라면 3000원.. B마트 연필 300원</p>



데이터 기반 업무 프로세스





데이터 분석 사례

- 통계청 공식 블로그의 사례 [https://blog.naver.com/hi_nso/]



데이터 준비





데이터 준비

- 오픈(공공) 데이터파일 / API
- 웹 / 미디어(SNS등)
- 데이터베이스
- 클라우드
- 사물인터넷





데이터 수집 기술

- 서비스 활용에 필요한 데이터를 시스템의 내부 혹은 외부에서 주기성을 갖고 필요한 형태로 수집하는 활동
- 수집 데이터의 위치에 따라 구축하려는 시스템 내부 혹은 외부의 데이터로 구분
- 외부 데이터는 주로 인터넷 기반으로 수집
- 대표적인 데이터 수집 기술

수집 기술	설명
FTP(File Transfer Protocol)	데이터 파일 전송(웹 환경에서 파일 전송도 포함)
Open API(Application Programming Interface)	실시간으로 데이터를 수신할 수 있도록 공개된 API
웹 스크래핑(Web Scraping)	웹 페이지에서 데이터를 추출
웹 크롤링(Web Crawling)	웹 사이트 URL을 탐색해서 데이터 복사 수집
스트리밍(Streaming)	센서데이터 및 미디어 데이터를 실시간으로 수집



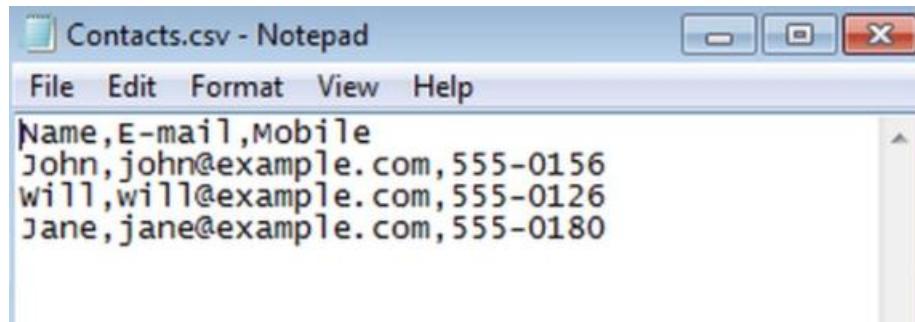
공개(Open) 데이터 파일 활용하기

- 외부 서버에 저장된 데이터 파일 → 공공데이터개방 활용
 - 정부, 지방자치단체 및 공공기관이 보유·관리하는 공공데이터를 이용자가 자유롭게 활용할 수 있도록 다양한 형태로 개방·제공함을 법률로 제정
- 공공 데이터 주요 사이트
 - AI허브 <https://www.aihub.or.kr/>
 - 한국지능정보사회진흥원이 운영하는 AI 통합 플랫폼
 - 공공데이터포털 <https://www.data.go.kr/>
 - 행정안전부에서 운영하는 공공데이터 통합제공 시스템
 - 캐글(Kaggle) <https://www.kaggle.com/>
 - 2010 설립, 2017년 3월 구글이 인수한 글로벌 데이터 분석 경쟁 대회 및 커뮤니티 플랫폼
 - 경남 빅데이터 허브 플랫폼 <https://bigdata.gyeongnam.go.kr/index.gn>



데이터 파일 형식

- 공공 데이터는 TXT, CSV, XLS, XLSX 파일 형식으로 제공되는 경우가 많음
- TXT 파일 ← 비정형데이터
 - 기본 문자나 수치로 구성된 구조화되지 않은 데이터 파일
- CSV(Comma-Separated Value) 파일 ← 정형데이터
 - 값이 쉼표로 구분되어 있는 데이터 파일
 - 각 값은 쉼표로 구분되고 <Enter>키로 행 단위를 구분함



- XLS, XLSX 파일 (Excel 통합 문서) ← 정형데이터
 - Excel 프로그램에서 편집할 수 있는 통합문서파일



Excel 데이터 준비

- 데이터 입력
 - 키보드로 데이터 입력
 - 복사 및 붙여넣기 등 편집 기능을 사용하여 데이터 입력
- 데이터 수집
 - [데이터]메뉴탭 → [데이터 가져오기 및 변환]그룹 → [데이터 가져오기]메뉴
 - 외부 데이터 파일을 읽어 들임
 - 외부 서버에 저장되어 있는 공개된 데이터 파일을 검색하여 다운로드 저장
 - TXT, CSV, PDF, XML, JSON 파일 형식 지원
 - 웹 페이지 크롤링
 - 웹 페이지 URL에 접속하여 웹 상의 페이지 표를 가져오는 기능
예시) <https://finance.naver.com/> 의 웹 페이지 표를 워크시트의 데이터 표로 로드



데이터 입력 편의를 위한 메뉴: 데이터 유효성 검사

- 동일 속성의 데이터 입력 시 오류를 줄이고 자동화 시트 작성을 위한 도구
 - 설정 탭에서 데이터 유효 조건의 정수, 목록, 텍스트길이 등의 제한 대상을 지정하여 유효한 데이터 여부 확인
 - 유효한 데이터 대상을 설명하는 설명 메시지 탭의 옵션과 유효하지 않은 데이터 입력을 불가하거나 허용할 수 있는 오류메시지 탭의 옵션을 지정하여 활용함
- 특정 셀 범위 데이터를 목록으로 입력 / 정수(숫자)의 입력 범위 시작과 끝 설정
- 언어 입력기(IME) 모드 탭
 - 데이터 입력 시작시 키보드에서 한/영을 누르지 않아도 영문/한글 모드로 입력을 시작하는 기능
 - 셀 범위를 선택하고, [데이터 유효성 검사] 대화상자에서 [IME 모드] 탭을 클릭후 입력기 보드에서 [영문]을 선택
 - 다른 셀의 입력 상태와 상관없이 지정 셀 입력 시작시 영문 또는 한글로 지정되나, 한영키로 입력모드를 변경하면 변경됨 (강제 제한은 아님)



데이터 입력 편의를 위한 메뉴: 데이터 유효성 검사

- 데이터 입력할 셀 선택 → [데이터]메뉴 탭 → [데이터유효성검사]메뉴

1

봉사자 명단						
	이름	아이디	생년월일	성별	담당	담당명
3	박주원					청소
4	정성민					빨래
5	주정환					요리
6	성미란					설비
7	이정미					
8	김주화					
9						
10						

2

데이터 유효성

설정 설명 메시지 오류 메시지 IME 모드

유효성 조건

제한 대상(A):

- 모든 값
- 모든 값
- 정수
- 소수점
- 목록
- 날짜
- 시간
- 텍스트 길이

사용자 지정

변경 내용을 설정이 같은 모든 셀에 적용(P)

모두 지우기(C) 확인 취소

- 데이터 유효성 검사의 잘못된 데이터

F X BT BU BV BW BX BY BZ LA



외부 데이터 파일 다운로드

- 외부 공공 데이터 셋 사이트: 경남 빅데이터 허브 플랫폼 접속

<https://bigdata.gyeongnam.go.kr/index.gn>

- ‘경남빅데이터허브플랫폼’ → [경남통계] → [통계DB]

The screenshot shows the main navigation menu of the Gyeongnam Big Data Hub Platform. The 'Statistics' section is highlighted with a red box. Below the menu, there is a sidebar with various survey categories and a large 'KOSIS' logo.

- 소개·참여
- 빅데이터
- 공공데이터
- 데이터분석
- 경남통계**
- 빅데이터센터

한눈에 보는 경남
통계DB
통계보고서

• 주민등록인구	• 토지 및 기후
• 장래인구추계	• 노동
• 경남의 사회지표	• 사업체
• 지역내 출생산(GDP)	• 농업·어업·수산업
• 사업체조사	• 전기·가스·수도
• 광업 제조업조사	• 유통·금융·보험 및 기타 서비스
• 경기종합지수	• 주택·건설
• 부서별로	• 보건·관광 및 사회보장
	• 환경
	• 교육 및 문화
	• 재정
	• 소득 및 지출
	• 공공행정 및 사업

KOSIS

통계조회 및 다운로드

▶ 통계작성 기관별 > 경상남도 > 경상남도기본통계 > 경상남도 통계연보

▣ 토지 및 기후

- ▣ 행정구역 수록기간 년:2000~2020
- ▣ 토지지목별 현황 수록기간 년:2000~2020
- ▣ 일기일수 수록기간 년:2000~2020
- ▣ 기상개황 수록기간 년:2000~2020
- ▣ 강수량 수록기간 년:2000~2020**
- ▣ 해안선 및 도서 수록기간 년:2000~2019

▣ 인구

- ▣ 노동
- ▣ 사업체
- ▣ 노리스시어



외부 데이터 파일 다운로드

- 경남 빅데이터 허브 플랫폼 → 통계청 데이터 사이트 접속
- 예시) ‘토지및기후’ → ‘강수량’ → [다운로드]
- 파일 형식 선택 → Excel 파일 형식일때는 “셀 병합” 옵션 해제 필요
- TXT 파일 다운로드

The screenshot shows the 'Korea Big Data Hub' interface. In the top navigation bar, '통계DB조회' and '경상남도' are selected. On the right, there's a logo for '통계청'.

The main content area displays a table for 'Precipitation' data from 2000 to 2020. The table has columns for Month (월), 5월 (May), 6월 (June), and 7월 (July). The data values range from 24.4 to 116.4.

A modal window titled '다운로드' (Download) is open over the table. It contains several options:

- 파일형태 (File Type):** Options include EXCEL(xlsx), EXCEL xls, CSV, TXT, and SDMX(2.0). The 'EXCEL(xlsx)' option is selected. A checkbox for '셀 병합' (Cell Merge) is checked. Other checkboxes for '빈셀 부호(-)' (Blank cell symbol (-)), '통계부호' (Statistical symbols), and '코드포함' (Include code) are unchecked.
- 시점정렬 (Perspective Sort):** Options include '오름차순' (Ascending) and '내림차순' (Descending). '오름차순' is selected.
- 소수점 (Decimal Point):** Options include '수록자료형식과 동일' (Same as recorded data) and '조회화면과 동일' (Same as query screen). '조회화면과 동일' is selected.

At the bottom of the modal is a large red-bordered '다운로드' (Download) button.

Excel의 외부 데이터 관련 메뉴: 텍스트 파일 가져오기



● 텍스트 파일 가져오기 → 파워쿼리 표로 로드

The screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected. In the 'Get Data' section of the ribbon, the 'From Text/CSV' button is highlighted with a red box and labeled '1'. Below it, the 'Text/CSV' option under 'From File' is also highlighted with a red box and labeled '2'. The 'Import Data' dialog box is open, displaying the file '강수량_20220816203601.txt' in the 'Name' field. The 'Load' button at the bottom right of the dialog is highlighted with a red box. The Power Query Editor window is visible in the background, showing the imported data in a table format.

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10
행정구역(시군)별(1)	행정구역(시군)별(2)	2020	2020	2020	2020	2020	2020	2020	2020
평균	소계	1869.5	104.7	54.8	39.7	60.9	90.9	244.4	563.4
동부	김해시	1743.0	104.0	67.9	41.4	56.7	58.5	248.5	536.2
동부	밀양시	1503.0	85.6	63.3	34.4	47.5	47.0	195.7	437.3
동부	양산시	1826.0	150.0	65.4	41.1	61.0	52.6	228.6	499.4
중부	청원시	1717.0	106.7	59.4	36.8	59.1	66.7	247.6	501.7
중부	거제시	2208.0	122.3	59.7	55.3	79.1	178.7	341.2	629.0
중부	통영시	2006.0	111.8	66.0	59.1	70.6	125.6	358.1	582.2
중부	고성군	2051.0	116.5	66.4	42.3	75.2	119.5	341.4	651.4
중부	합천군	1620.0	100.7	46.1	32.8	57.6	52.2	176.8	497.1
중부	청녕군	1497.0	85.1	44.7	30.6	44.2	45.4	181.9	465.3
서부	진주시	1799.0	104.5	46.6	40.6	62.7	88.8	208.8	549.1
서부	사천시	1817.0	100.8	54.1	38.8	69.9	116.4	275.1	592.0
서부	남해군	2132.0	118.5	73.7	53.7	84.4	164.7	357.9	651.4
서부	하동군	2217.0	112.5	49.9	47.6	69.1	142.1	279.9	635.6
서부	의령군	1824.0	97.6	42.0	31.4	52.8	70.8	190.6	577.5
서부	산청군	2045.0	89.3	42.4	36.9	59.6	94.5	203.7	603.5
서부	합성군	2043.0	104.3	52.3	33.4	53.6	68.8	211.1	620.6
서부	거창군	1868.0	91.3	47.5	29.1	47.7	75.1	176.4	603.2



웹 스크래핑(Web scraping)

- 웹(Web)상에 존재하는 콘텐츠를 수집하는 작업
 - 웹 페이지를 구성하는 HTML 태그 정보를 기반으로 원하는 내용 부분 수집
 - 브라우저를 직접 동작하는 것처럼 프로그래밍하여 데이터 수집
 - 해당 사이트에서 제공하는 서비스의 오픈 API 활용
 - 서버와 브라우저 간의 Network를 분석하여 필요한 데이터 수집

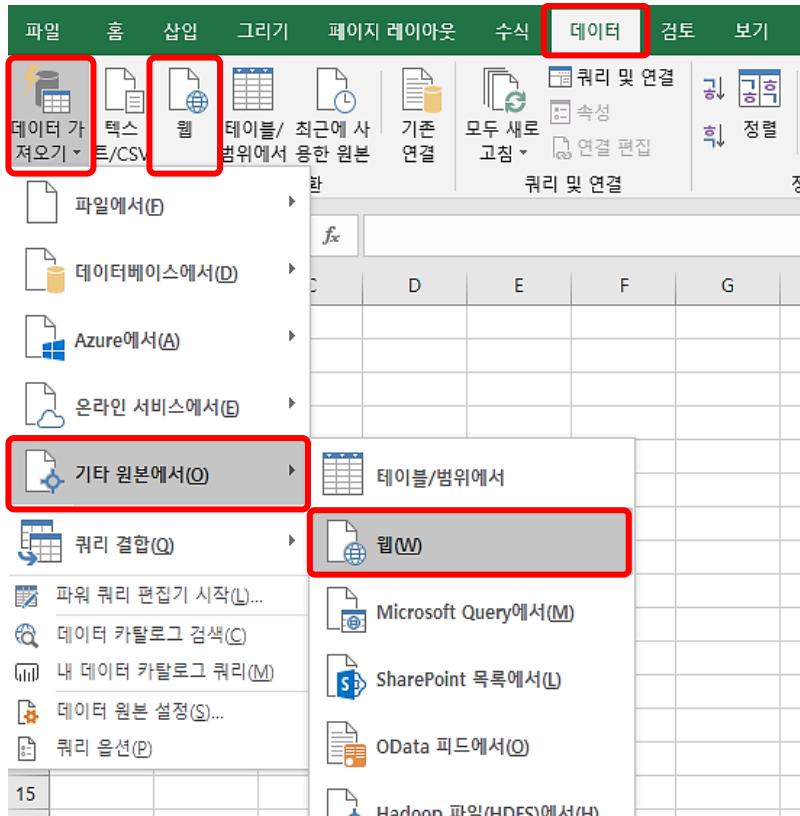
- 코딩 없이 실행하는 Excel 웹 스크래핑
 - 외부 데이터 가져오기 메뉴로 공개된 웹 페이지 HTML 태그의 <Table> 읽기 가능
 - [데이터]메뉴탭 → [데이터 가져오기 및 변환]그룹 → [웹]메뉴 → [파워쿼리] 대화상자 실행
 - 해당 웹 페이지의 표(Table)을 목록으로 표시하여 데이터 표로 읽어 가져오기



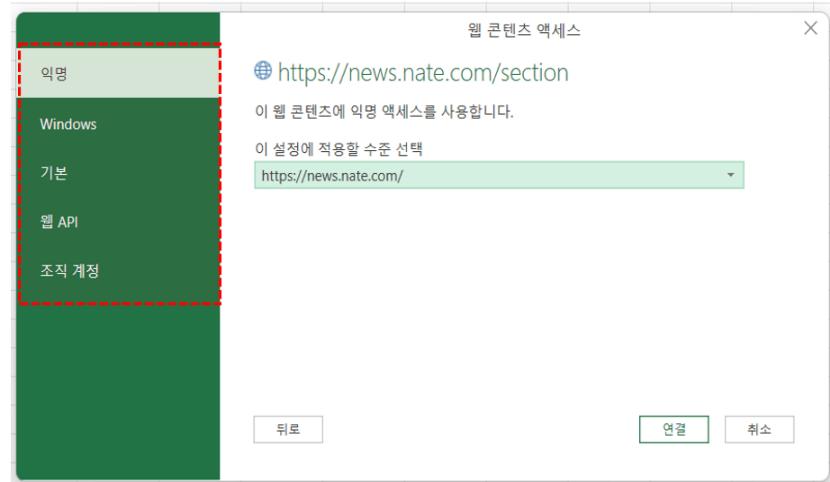
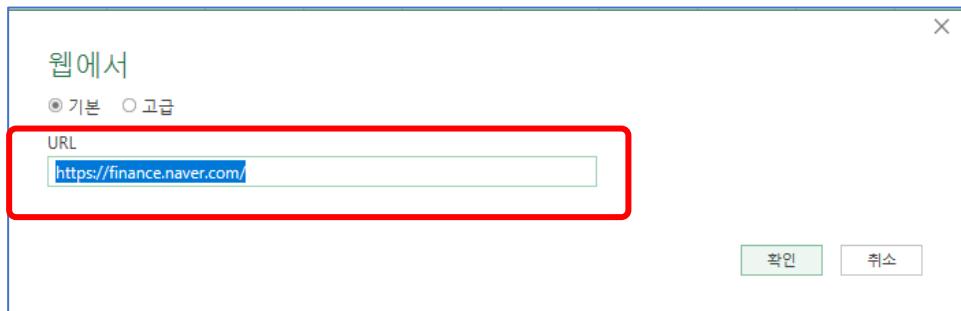
Excel의 외부 데이터 관련 메뉴: 웹 가져오기

● 웹페이지 표를 워크시트에 가져오기

- [데이터]메뉴 탭 → [데이터 가져오기 및 변환]그룹 → [데이터 가져오기] → [웹]



※ URL입력: <https://finance.naver.com/>





웹 표 가져오기

- 웹페이지에 구성된 표 목록 중 원하는 표를 선택
- 워크시트의 데이터 표로 가져오기(로드) → [파워쿼리]기반 데이터 표
- [데이터] → [데이터 가져오기] → [기타원본에서] → [웹]메뉴를 실행 → 파워쿼리의 [탐색창] 실행

The screenshot shows the Microsoft Power Query Editor interface. On the left, the 'Search' window displays a list of web pages from 'https://finance.naver.com/[13]'. Item 1 is highlighted with a red dashed box and points to the '거래상위 TOP 종목 종목명에 대한 현재가...' item. On the right, the main editor window shows a table titled '거래상위 TOP 종목 종목명에 대한 현재가...' with the following data:

종목명	현재가	전일대비	등락률
쌍방울	860	상승 198	29.91%
KODEX 200선물인버스2X	2680	상승 10	0.37%
링링	2565	상승 590	29.87%
캘시스	2435	상승 380	18.49%
한창	1530	상승 5	0.33%
아미노로직스	2845	상승 315	12.45%
아이오케이	1060	상승 177	20.05%
미아이스캐피탈	773	상승 53	7.36%
이루온	2820	상승 395	16.29%
대한전선	2080	하락 45	-2.12%
KODEX 코스닥150선물인버스	4660	하락 40	-0.85%
유진로보ット	4495	하락 5	-0.11%
삼성전자	65200	하락 100	-0.15%
삼성 인버스 2X WTI원유 선물	90	하락 5	-5.26%
제니텍스	3255	상승 230	7.60%

A red arrow labeled 2 points to the 'Load' button at the bottom of the editor.

데이터 전처리





데이터 테이블

- 데이터 테이블 ← 정형데이터
- 행과 열로 구성

	A	B	C	D	E	F	G	H	I
1	2022년도 교양 관련 수업 성적 집계								
3	작성자: 우영우			2022년 9월 1일 목요일					
5	소속학과	학년	학번	성명	중간	기말	과제	출석	태도
6	컴퓨터	2	2021512512	이재환	52.1	46.8	44.4	83.3	50
7	의류디자인	3	2022430131	김기태	55.9	57.3	100.0	95.6	100
8	경영정보	3	2021212934	배정이	59.0	58.2	100.0	100.0	100
9	스포츠건강과학	4	2018410341	박도연	19.9	32.3	33.3	74.4	50
10	컴퓨터	3	2022512705	박주현	69.0	76.8	100.0	100.0	100
11	경영정보	3	2022212312	채예정	51.7	52.3	100.0	94.4	100
12	공공인재	4	2018123958	김태훈	93.9	84.6	100.0	100.0	100
13	식품영양	3	2020321368	김민주	86.6	76.8	100.0	97.8	100
14	식품영양	2	2021321140	정승환	54.8	60.9	100.0	90.0	100
15	공공인재	3	2020123550	박준규	85.1	75.0	100.0	98.9	100
16	산업디자인	1	2020524713	강성연	55.6	19.6	88.9	96.7	100
17	스포츠건강과학	3	2018410823	임진규	89.3	70.9	100.0	100.0	100
18	산업디자인	1	2020524642	이민형	68.6	75.5	88.9	86.7	100
19	도시환경	1	2023560029	권지원	55.9	41.4	100.0	100.0	100
20	산업디자인	4	2018524437	주현진	52.1	23.6	66.7	91.1	100
21	도시환경	3	2020560464	박성호	71.3	67.7	100.0	98.9	100



정형데이터 행과 열

- 변수(Variable, 열, 컬럼, 피처, 특징): 키, 몸무게, 성별
- 관측치(Observation, 행, 로우, 레코드): 값을 측정한 단위, 각각의 사람
- 값(Value): 152 cm, 80 kg, 여성

1 변수(열, 컬럼, 특징, 피처)

Name	Sex	Age	Grade	Absence	Blood type	Height	weight
김길동	남자	23	3	유	O	165.3	68.2
이미린	여자	22	2	무	AB	170.1	57
홍길동	남자	24	4	무	B	175	80.1
김철수	남자	23	3	무	AB	182	85.7
손세수	여자	20	1	유	A	168	49.5
박미희	여자	21	2	무	O	162	52
강수진	여자	22	1	무	O	155.2	45.3

4 변수값

3
관측치

2
행,로우
데이터양



정리된 깔끔한 데이터

- 제프리크(Jeff Leek)의 데이터분석스타일기초(The Elements of Data Analytic Style)
저서에서 데이터 분석이 용이한 특징을 가지는 데이터 → 깔끔한 데이터(Tidy data)
- 깔끔한 데이터는 데이터를 조작하고, 모형화 하고, 시각화가 용이
- 깔끔한 데이터는 특정한 구조를 갖추고 있는데 변수는 열(column)이고, 관측점은 행(row)이며, 관측 단위에 대한 형태는 테이블(table)로 구성
- 깔끔한 데이터(tidy data)는 결국 데이터분석을 쉽게 할 수 있는 데이터라 할 수 있음

* Tidy data

NO	변수	값
김길동	A	1
이미린	A	4
홍길동	A	5
김철수	A	3
손세수	A	2
김길동	B	7
이미린	B	3
홍길동	B	6
김철수	B	5
손세수	B	3

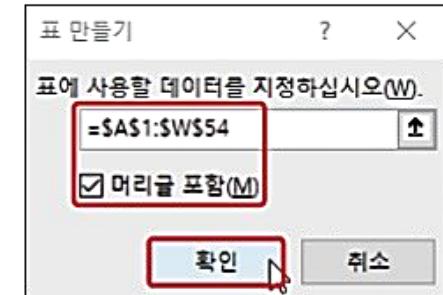
* Messy data

NO	A	B
김길동	1	7
이미린	4	3
홍길동	5	6
김철수	3	5
손세수	2	3

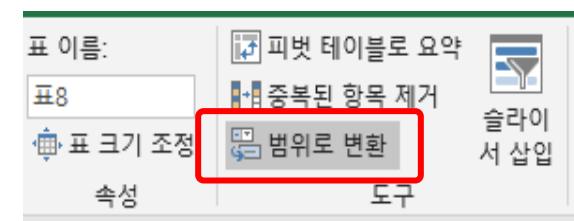


데이터 표(Table)

- Excel의 표(테이블) 기능을 사용하면 셀 주소를 사용하지 않고 참조 가능
 - 구조적 참조: 데이터 추가 삭제 수정 등의 작업 시 데이터 범위가 자동 지정
 - 필터링이나 간단한 함수 연산 기능 포함
 - 즉각적인 데이터 확인이나 연산 가능
- [삽입]메뉴탭 → [표]그룹 → [표]메뉴
 - [표 만들기]창
 - “머리글 포함” 옵션 확인



- [표도구][디자인]탭 → [도구]그룹 → [범위로변환]메뉴
 - 데이터 표를 일반 시트의 범위로 변환하여 표기능 제거 가능





데이터 표(Table)

- 적용된 데이터 영역의 열 제목 부분에 생기는 □를 이용하여 데이터를 선별

Screenshot of Microsoft Excel showing the 'Table' feature being used to filter data.

The screenshot shows an Excel spreadsheet titled "데이터분석기초_04주차_2차시실습_완.xlsx". The ribbon menu is visible, with the "Data" tab selected. A red box highlights the dropdown arrow next to the column header "소속학과" (Department). Another red box highlights the "Design" tab in the ribbon, which is currently active. A third red box highlights the "New Style" button in the bottom right corner of the ribbon.

The main content area displays a table with columns: 소속학과, 학년, 학번, 성명, 중간, 기말, 과제, 출석, 태도. The first row contains dropdown arrows indicating it is a table. The data rows show various student records with their respective department, year, ID, name, and scores.

On the right side of the ribbon, a large preview area shows various table styles. A red box highlights the "New Style" button at the bottom of this preview area. The preview area is divided into sections: "Bright", "Middle", and "Dark".

	A	B	C	D	E	F	G	H	I
1	소속학과	학년	학번	성명	중간	기말	과제	출석	태도
2	컴퓨터	2	2021512512	이○환	52.1	46.8	44.4	83.3	50
3	의류디자인	3	2022430131	김○태	55.9	57.3	100.0	95.6	100
4	경영정보	3	2021212934	배○이	59.0	58.2	100.0	100.0	100
5	스포츠건강과학	4	2018410341	박○연	19.9	32.3	33.3	74.4	50
6	컴퓨터	3	2022512705	박○현	69.0	76.8	100.0	100.0	100
7	경영정보	3	2022212312	채○정	51.7	52.3	100.0	94.4	100
8	공공인재	4	2018123958	김○훈	93.9	84.6	100.0	100.0	100
9	식품영양	3	2020321368	김○주	86.6	76.8	100.0	97.8	100
10	식품영양	2	2021321140	정○환	54.8	60.9	100.0	90.0	100
11	공공인재	3	2020123550	박○규	85.1	75.0	100.0	98.9	100
12	산업디자인	1	2020524713	강○연	55.6	19.6	88.9	96.7	100
13	스포츠건강과학	3	2018410823	임○규	89.3	70.9	100.0	100.0	100
14	산업디자인	1	2020524642	이○형	68.6	75.5	88.9	86.7	100
15	도시환경	1	2023560029	권○원	55.9	41.4	100.0	100.0	100
16	산업디자인	4	2018524437	주○진	52.1	23.6	66.7	91.1	100
17	도시환경	3	2020560464	박○호	71.3	67.7	100.0	98.9	100
18									
19									
20									
21									
22									
23									



데이터 표(Table)

- 요약행을 표시하여 간단한 요약 실행 가능
- 요약행을 표시한 후 다시 해제하면 데이터를 추가한 후 다시 요약행 표시 가능

표 스타일 옵션

14	산업디자인	1	2020524642	이○형	68.6	75.5	88.9	86.7	100
15	도시환경	1	2023560029	권○원	55.9	41.4	100.0	100.0	100
16	산업디자인	4	2018524437	주○진	52.1	23.6	66.7	91.1	100
17	도시환경	3	2020560464	박○호	71.3	67.7	100.0	98.9	100
18	요약			16	63.8	57.5	88.9	94.2	93.75
19									
20									
21									
22									
23									
24									
25									

- 요약행 해제 후 데이터 추가

14	산업디자인	1	2020524642	이○형	68.6	75.5	88.9	86.7	100
15	도시환경	1	2023560029	권○원	55.9	41.4	100.0	100.0	100
16	산업디자인	4	2018524437	주○진	52.1	23.6	66.7	91.1	100
17	도시환경	3	2020560464	박○호	71.3				
18	컴퓨터	4	2022112233	유현					
19									
14	산업디자인	1	2020524642	이○형	68.6	75.5	88.9	86.7	100
15	도시환경	1	2023560029	권○원	55.9	41.4	100.0	100.0	100
16	산업디자인	4	2018524437	주○진	52.1	23.6	66.7	91.1	100
17	도시환경	3	2020560464	박○호	71.3	67.7	100.0	98.9	100
18	컴퓨터	4	2022112233	유현주	80.0	80.0	100.0	95.0	100
19	요약			17	64.7	58.8	89.5	94.3	94.1176
20									



데이터 표

- Excel의 표(테이블) 기능을 사용하면 셀 주소를 사용하지 않고 참조 가능
→ 구조적 참조

구조적참조	설명	사용 예
표	표의 데이터 범위만 참조합니다.	=판매대장
표[#모두]	표의 전체 범위를 참조합니다.	=판매대장[#모두]
표[#머리글]	표의 머리글 범위만 참조합니다.	=판매대장[#머리글]
표[#데이터]	표의 데이터 범위만 참조합니다.	=판매대장[#데이터]
표[#요약]	표의 요약 행 범위만 참조합니다.	=판매대장[#요약]
표[머리글]	표의 해당 머리글을 사용하는 열의 데이터 범위를 참조합니다.	=판매대장[판매]
표[[머리글1]:[머리글2]]	표의 머리글1을 사용하는 열 데이터 범위부터 머리글2를 사용하는 열 데이터 범위까지 참조합니다.	=판매대장[[제품]:[판매]]



데이터 표의 구조적 참조

● 표 상태에서 계산 필드 추가

- 수식 입력 등호로 시작 → 계산할 필드를 대괄호 “[]”로 불러옴
- 필드(열) 전체에 수식 바로 적용

No	제품	단가	수량	판매
1	1 컬러레이저복사기 XI-3200	1,176,000	3	=[단가]*[수량]
2	2 바코드 Z-350	48,300	3	=[단가]*[수량]
3	3 잉크젯팩시밀리 FX-1050	47,400	3	=[단가]*[수량]
4	4 프리미엄복사지A4 2500매	17,800	9	=[단가]*[수량]
5	5 바코드 BCD-100 Plus	86,500	7	=[단가]*[수량]
6	6 고급복사지A4 500매	3,500	2	=[단가]*[수량]
7	7 바코드 Z-350	46,300	7	=[단가]*[수량]
8	8 바코드 BCD-100 Plus	104,500	8	=[단가]*[수량]
9	9 잉크젯복합기 AP-3300	79,800	1	=[단가]*[수량]
10	10 잉크젯복합기 AP-3200	89,300	8	=[단가]*[수량]

데이터 표(Table)

● 표의 활용

● 슬라이서 추가

The screenshot shows a Microsoft Excel spreadsheet with a table of student data. On the right side, there are two slicer panes.

Slicer Panes:

- 학년 (Grade) Slicer:** Contains four items: 1, 2, 3, and 4. The item "1" is highlighted in green, indicating it is selected.
- 소속학과 (Major) Slicer:** Contains eight items: 경영정보, 공공인재, 도시환경, 산업디자인, 스포츠건강과학, 식품영양, 의류디자인, and 컴퓨터. The items "경영정보", "공공인재", "도시환경", "산업디자인", "스포츠건강과학", "식품영양", and "의류디자인" are highlighted in orange, while "컴퓨터" is highlighted in green, indicating it is selected.

Table Data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	소속학과	학년	학번	성명	중간	기말	과제	출석	태도	합계	등급	소속학과	학년	소속학과	학년	소속학과	학년
2	컴퓨터	2	2021512512	이○환	52.1	46.8	44.4	83.3	50	276.7		경영정보	1	경영정보	1	경영정보	1
3	의류디자인	3	2022430131	김○태	55.9	57.3	100.0	95.6	100	408.8		공공인재	2	공공인재	2	공공인재	2
4	경영정보	3	2021212934	배○이	59.0	58.2	100.0	100.0	100	417.2		도시환경	3	도시환경	3	도시환경	3
5	스포츠건강과학	4	2018410341	박○연	19.9	32.3	33.3	74.4	50	210.0		산업디자인	4	산업디자인	4	산업디자인	4
6	컴퓨터	3	2022512705	박○현	69.0	76.8	100.0	100.0	100	445.8		스포츠건강과학		스포츠건강과학		스포츠건강과학	
7	경영정보	3	2022212312	채○정	51.7	52.3	100.0	94.4	100	398.4		식품영양		식품영양		식품영양	
8	공공인재	4	2018123958	김○훈	93.9	84.6	100.0	100.0	100	478.4	우수	의류디자인		의류디자인		의류디자인	
9	식품영양	3	2020321368	김○주	86.6	76.8	100.0	97.8	100	461.2	우수	컴퓨터		컴퓨터		컴퓨터	
10	식품영양	2	2021321140	정○환	54.8	60.9	100.0	90.0	100	405.7		도시환경		도시환경		도시환경	
11	공공인재	3	2020123550	박○규	85.1	75.0	100.0	98.9	100	459.0	우수	산업디자인		산업디자인		산업디자인	
12	산업디자인	1	2020524713	강○연	55.6	19.6	88.9	96.7	100	360.7		도시환경		도시환경		도시환경	
13	스포츠건강과학	3	2018410823	임○규	89.3	70.9	100.0	100.0	100	460.2	우수	컴퓨터		컴퓨터		컴퓨터	
14	산업디자인	1	2020524642	이○형	68.6	75.5	88.9	86.7	100	419.6		의류디자인		의류디자인		의류디자인	
15	도시환경	1	2023560029	권○원	55.9	41.4	100.0	100.0	100	397.3		컴퓨터		컴퓨터		컴퓨터	
16	산업디자인	4	2018524437	주○진	52.1	23.6	66.7	91.1	100	333.5		의류디자인		의류디자인		의류디자인	
17	도시환경	3	2020560464	박○호	71.3	67.7	100.0	98.9	100	437.9		컴퓨터		컴퓨터		컴퓨터	
18	컴퓨터	4	2022112233	유현주	80.0	80.0	100.0	95.0	100	455.0	우수	의류디자인		의류디자인		의류디자인	
19	요약				17	64.7	58.8	89.5	94.3	94.12		컴퓨터		컴퓨터		컴퓨터	
20																	

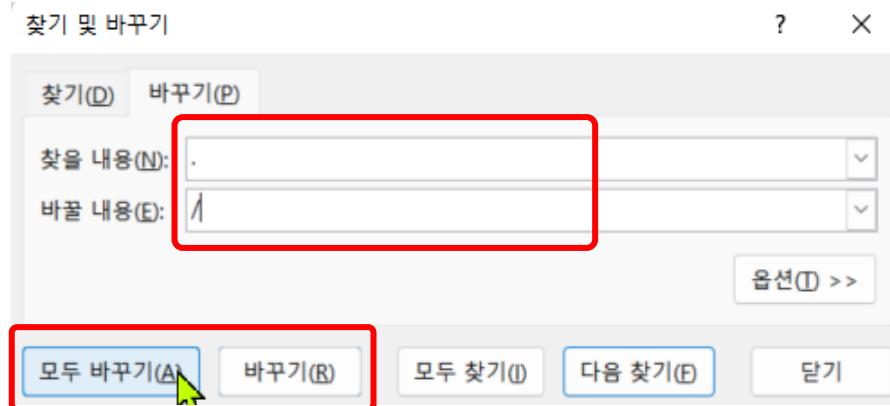
슬라이서 삽입





Excel 데이터 형식 및 값 정리

- 데이터 구성 파악을 위한 데이터 속성 기반 값의 정리 필요
- [홈]메뉴탭 → [편집]그룹 → [찾기및선택] → [바꾸기]메뉴



A	B	C	D	E	F	G	H	I
고객	담당	주문일	제품	판가	수량	수량	할인율	판매
S&C무역 (주)	오서윤	2017.01.01	컬러레이저복사기 XI-3200	1,176	3	3	15%	2,998,800
S&C무역 (주)	오서윤	2017.01.01	바	텍스트 형식으로 저장된 숫자	3	3	0%	144,900
S&C무역 (주)	오서윤	2017.01.01	잉크젯프	숫자로 변환(C)	3	3	0%	142,200
드림씨푸드 (주)	박현우	2017.01.02	프리미엄	이 오류에 대한 도움말(H)	9	9	0%	160,200
드림씨푸드 (주)	박현우	2017.01.02	바코드	오류 무시(I)	7	7	0%	605,500
자이언트무역 (주)	정시우	2017.01.05	고급복	수식 입력줄에서 편집(E)	2	2	0%	7,000
자이언트무역 (주)	정시우	2017.01.05	바	오류 검사 옵션(O)...	7	7	0%	324,100
자이언트무역 (주)	정시우	2017.01.05	바코드 BCD 100 1183	10,500	8	8	0%	836,000
진왕통상 (주)	오서윤	2017.01.05	잉크젯복합기 AP-3300	79,800	1	1	0%	79,800
진왕통상 (주)	오서윤	2017.01.05	잉크젯복합기 AP-3200	89,300	8	8	0%	714,400
진왕통상 (주)	오서윤	2017.01.05	고급복사기 A4 500매	4,100	7	7	0%	28,700
삼양트레이드 (주)	김민준	2017.01.06	잉크젯복합기 AP-3200	79,500	2	2	0%	159,000
삼양트레이드 (주)	김민준	2017.01.06	레이저복합기 L200	165,300	3	3	0%	495,900
삼양트레이드 (주)	김민준	2017.01.06	고급복사기 A4 500매	3,600	8	8	0%	28,800

- 텍스트데이터 숫자로 변환



데이터 표 형태 정리

● 데이터 표 구성 정리 작업

	A	B	C	D	E	F	G	H	
1	거래처별 매출채권 관리대장								
2									
3	고유번호	협상일	만기일	송장번호	선적일	통화	금액	결재일	
4	820-028152778	2021.04.23	2021.10.20	202102K007	2021.02.25	USD	16,136.40	11월 30일	
5				202103K007	2021.03.21	USD	3,570.00	11월 30일	
6				202104K001	2021.04.11	USD	16,136.40	11월 30일	
7							35,842.80		
8	820-028467749	2021.07.09	2020.01.05	202105K006	2021.05.23	USD	3,570.00	02월 28일	
9				202107K001	2021.07.11	USD	3,570.00	02월 28일	
10				202107K013	2021.07.21	USD	10,200.00	02월 28일	
11							17,340+		
12	820-028793887	2021.09.19	2020.03.18	202108K021	2021.08.31	USD	225,000.00	04월 30일	
13				202108K021	2021.08.31	USD	80,504.61	04월 30일	
14							305,504.61		
15	820-028816970	2021.09.25	2020.03.24	202108K021	2021.08.31	USD	80,504.61	04월 30일	
16				202108K002	2021.08.17	USD	4,284.00	04월 30일	
17							84,788.61		
18	276-71-107565	2021.01.07	2021.07.05	202112Q012	2021.12.15	EUR	347.13	08월 31일	
19				202112Q011	2021.12.27	EUR	13,559.79	08월 31일	
20							13,906.92		
21	FET1-475-000083	2021.01.25	2021.07.23	202112Q013	2021.01.17	EUR	22,292.50	08월 31일	
22				202101Q011	2021.01.31	EUR	14,077.48	08월 31일	
23				202102Q012	2021.02.05	EUR	153.98	08월 31일	
24							36,523.96		
25	276-71-110840	2021.03.08	2021.09.04	202102Q013	2021.02.16	EUR	189.00	10월 31일	
26				202102Q014	2021.02.26	EUR	415.80	10월 31일	
27				202102Q011	2021.02.28	EUR	7,153.56	10월 31일	
28				202102Q015	2021.02.28	EUR	7,851.75	10월 31일	
29							15,610.11		
30	FET1-475-000332	2021.03.31	2021.09.27	202103Q012	2021.03.12	EUR	119.80	10월 31일	
				202103Q011	2021.03.31	EUR			

	A	B	C	D	E	F	G	H	
1	거래처별 매출채권 관리대장								
2									
3	고유번호	협상일	만기일	송장번호	선적일	통화	금액	결재일	
4	820-028152778	2021.04.23	2021.10.20	202102K007	2021.02.25	USD	16,136.40	11월 30일	
5	820-028152778	2021.04.23	2021.10.20	202103K007	2021.03.21	USD	3,570.00	11월 30일	
6	820-028152778	2021.04.23	2021.10.20	202104K001	2021.04.11	USD	16,136.40	11월 30일	
7	820-028467749	2021.07.09	2020.01.05	202105K006	2021.05.23	USD	3,570.00	02월 28일	
8	820-028467749	2021.07.09	2020.01.05	202107K001	2021.07.11	USD	3,570.00	02월 28일	
9	820-028467749	2021.07.09	2020.01.05	202107K013	2021.07.21	USD	+200.00	02월 28일	
10	820-028793887	2021.09.19	2020.03.18	202108K021	2021.08.31	USD	225,000.00	04월 30일	
11	820-028793887	2021.09.19	2020.03.18	202108K021	2021.08.31	USD	80,504.61	04월 30일	
12	820-028816970	2021.09.25	2020.03.24	202108K021	2021.08.31	USD	80,504.61	04월 30일	
13	820-028816970	2021.09.25	2020.03.24	202108K002	2021.08.17	USD	4,284.00	04월 30일	
14	276-71-107565	2021.01.07	2021.07.05	202112Q012	2021.12.15	EUR	347.13	08월 31일	
15	276-71-107565	2021.01.07	2021.07.05	202112Q011	2021.12.27	EUR	13,559.79	08월 31일	
16	FET1-475-000083	2021.01.25	2021.07.23	202112Q013	2021.01.17	EUR	22,292.50	08월 31일	
17	FET1-475-000083	2021.01.25	2021.07.23	202101Q011	2021.01.31	EUR	14,077.48	08월 31일	
18	FET1-475-000083	2021.01.25	2021.07.23	202102Q012	2021.02.05	EUR	153.98	08월 31일	
19	276-71-110840	2021.03.08	2021.09.04	202102Q013	2021.02.16	EUR	189.00	10월 31일	
20	276-71-110840	2021.03.08	2021.09.04	202102Q014	2021.02.26	EUR	415.80	10월 31일	
21	276-71-110840	2021.03.08	2021.09.04	202102Q011	2021.02.28	EUR	7,153.56	10월 31일	
22	276-71-110840	2021.03.08	2021.09.04	202102Q015	2021.02.28	EUR	7,851.75	10월 31일	
23	FET1-475-000332	2021.03.31	2021.09.27	202103Q012	2021.03.12	EUR	119.80	10월 31일	
24	FET1-475-000332	2021.03.31	2021.09.27	202103Q011	2021.03.21	EUR	26,133.61	10월 31일	
25	FET1-475-000373	2021.04.14	2021.10.11	202104Q011	2021.04.11	EUR	23,867.56	11월 30일	
26	FET1-475-000373	2021.04.14	2021.10.11	202104Q012	2021.04.18	EUR	12,342.72	11월 30일	
27	FET1-475-000373	2021.04.14	2021.10.11	202104Q013	2021.04.25	EUR	12,198.63	11월 30일	
28	FET1-475-000605	2021.06.09	2021.12.06	202105Q011	2021.05.07	EUR	1,821.60	01월 31일	
29	FET1-475-000605	2021.06.09	2021.12.06	202104Q115	2021.05.21	EUR	327.60	01월 31일	
30	FET1-475-000605	2021.06.09	2021.12.06	202104Q116	2021.05.21	EUR	163.96	01월 31일	





데이터 표 형태 정리

- 불필요한 행 삭제 → "금액"의 소계 행 삭제
 - 데이터(값) 셀 범위 지정: "송장번호"열 데이터 선택
 - [이동옵션]으로 빈 셀만 선택
 - 셀 [삭제]메뉴 옵션 [행전체] 삭제
- 빈셀에 값 입력 → "고유번호", "협상일", "만기일"의 빈 셀의 데이터 입력
 - 데이터 범위 지정: "고유번호", "협상일", "만기일" 열 데이터 선택
 - [이동옵션]으로 빈 셀만 선택
 - 선택 셀의 상위 행 값 입력: =상위위치셀주소 입력
 - <Ctrl>키 + <Enter>키로 일괄 입력
 - 수식을 제거하기 위하여 데이터 영역 선택 후 [복사] → [선택하여붙여넣기]의 [값] 붙여넣기



테이블을 구성하는 필드와 레코드

- 필드(Field)

- 표의 열(Column)에 해당하는 데이터의 가장 작은 단위이며, 값의 속성을 표현
- 데이터베이스관리시스템에서는 속성(Attribute)라고 일컫기도 함

- 레코드(Record)

- 표의 행(Row)로 표현되며 논리적으로 연관된 필드의 집합
- 데이터베이스관리시스템에서는 튜플(Tuple)이라고 일컫기도 함

- 키(Key)

- 테이블의 레코드를 식별(유일한 레코드로 인식)할 수 있도록 부여된 필드 또는 필드 집합
- 키가 되는 필드는 값이 누락될 수 없음(값이 존재하지 않는 상태를 null이라 함)
→ null값을 허용하지 않음

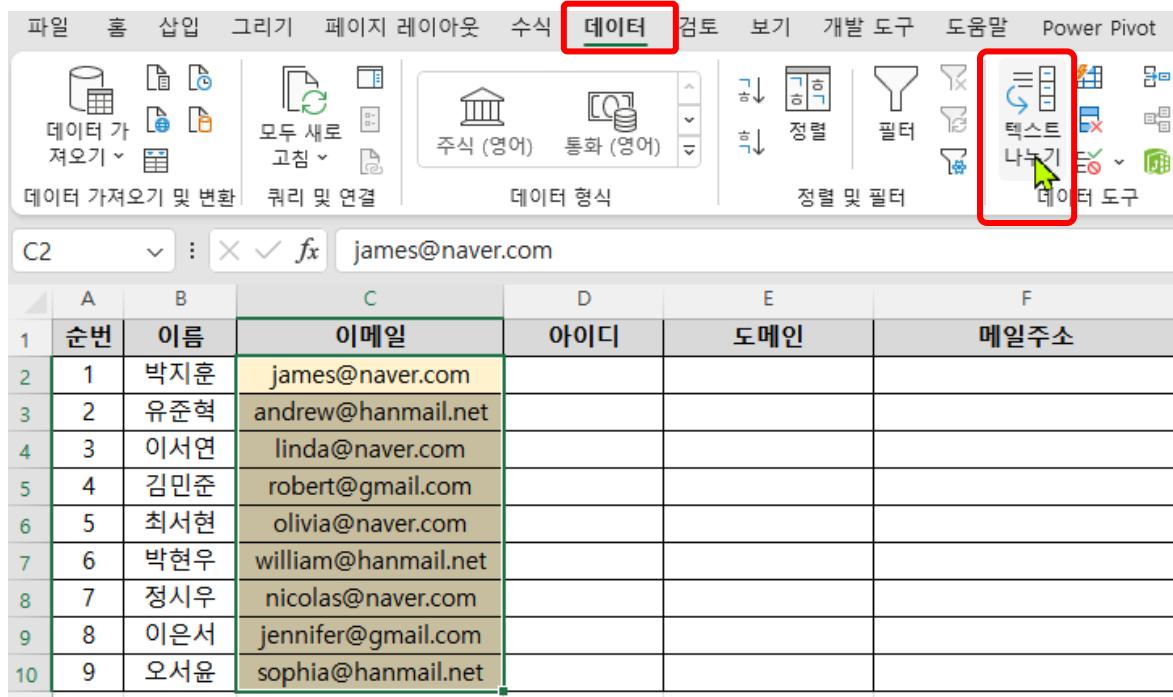


데이터 필드 정리

- 필드 분리 → 텍스트 나누기

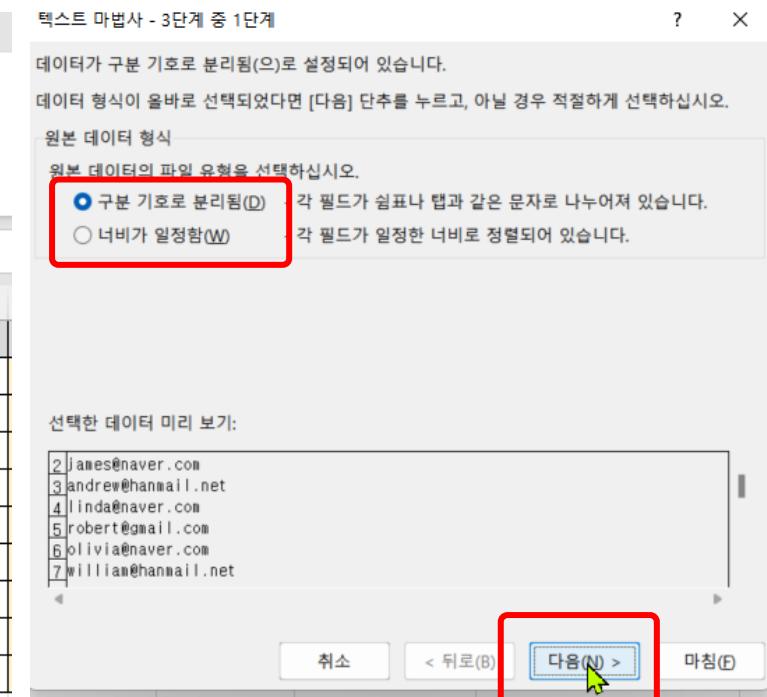
- [데이터] 메뉴 탭 → [데이터 도구] 그룹 → [텍스트나누기] 메뉴

- 구분 기호 있는 열 데이터 나누기
- 구분 기호 없는 열 데이터 나누기



The screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected (highlighted with a red box). Below the ribbon, there is a table with columns A through F. Column A is labeled '순번' (Index), column B is '이름' (Name), column C is '이메일' (Email), column D is '아이디' (ID), column E is '도메인' (Domain), and column F is '메일주소' (Email Address). The data in column C is being processed by the Text-to-Columns feature.

순번	이름	이메일	아이디	도메인	메일주소
1	박지훈	james@naver.com			
2	유준혁	andrew@hanmail.net			
3	이서연	linda@naver.com			
4	김민준	robert@gmail.com			
5	최서현	olivia@naver.com			
6	박현우	william@hanmail.net			
7	정시우	nicolas@naver.com			
8	이은서	jennifer@gmail.com			
9	오서윤	sophia@hanmail.net			





텍스트 나누기

● 구분 기호로 분리

● 구분 기호 옵션 선택 및 구분 문자 입력

텍스트 마법사 - 3단계 중 2단계

데이터의 구분 기호를 설정합니다. 미리 보기 상자에서 적용된 텍스트를 볼 수 있습니다.

구분 기호

- 탭(T)
- 세미콜론(M)
- 쉬프(S)
- 공백(S)
- 기타(Q):

연속된 구분 기호를 하나로 처리(R)

텍스트 한정자(Q):

데이터 미리 보기(P)

james	naver.com
andrew	hanmail.net
linda	naver.com
robert	gmail.com
olivia	naver.com
william	hanmail.net

취소 < 뒤로(B) 다음(N) > 마침(F)

텍스트 마법사 - 3단계 중 3단계

각 열을 선택하여 데이터 서식을 지정합니다.

열 데이터 서식

- 일반(G)
- 텍스트(I)
- 날짜(D):
- 열 가져오지 않음(건너뜀)(I)

일반]을 선택하면 숫자 값은 숫자로, 날짜 값은 날짜로, 모든 나머지 값은 텍스트로 변환됩니다.

고급(A)...

대상(E): \$C\$2

데이터 미리 보기(P)

텍스트	일반
james	naver.com
andrew	hanmail.net
linda	naver.com
robert	gmail.com
olivia	naver.com
william	hanmail.net

취소 < 뒤로(B) 다음(N) > 마침(F)



텍스트 나누기

● 구분 기호 없는 필드 나누기

- 너비가 일정함 옵션

- 데이터 너비를 눈금선으로 분리 → 분리선 삭제

- 열 데이터 서식 옵션: 날짜 형식 지정하여 날짜 데이터로 변환 가능

The screenshot shows two windows of the 'Text Wizard' (텍스트 마법사) for splitting text into columns.

Step 2: Setting field widths (3rd step of 3)

Message: 각 필드의 너비(열 구분선)를 지정합니다.
화살표가 있는 선에서 열이 나누어집니다.
구분선을 넣으려면 원하는 위치를 마우스로 클릭합니다.
구분선을 삭제하려면 구분선을 마우스로 두 번 클릭합니다.
구분선을 옮기려면 선을 마우스로 클릭한 상태에서 끕니다.

Step 3: Setting column formats (3rd step of 3)

Message: 각 열을 선택하여 데이터 서식을 지정합니다.

선택한 열의 서식은 아래와 같습니다.

열 데이터 서식
<input type="radio"/> 일반(G)
<input type="radio"/> 텍스트(I)
<input checked="" type="radio"/> 날짜(D): 년월일 ▼
<input type="radio"/> 열 가져오지 않음(건너뜀)(I)

[일반]을 선택하면 숫자 값은 숫자로, 날짜 값은 날짜로, 모든 나머지 값은 텍스트로 변환됩니다.

대상(E): \$B\$2

Preview Window:

데이터 미리 보기(P)

날짜
20220101
20220101
20220101
20220102
20220102
20220105

Buttons at the bottom: 취소, < 뒤로(B), 다음(N) >, 마침(E)



데이터 필드 합치기

- 데이터 필드 값 합치기 → 셀 값 연결 연산자 “&” 또는 =CONCAT()함수 사용
- 셀 값 연결 결과는 수식 결과이므로 수식 결과를 값으로 변환하여 필요
- 연산 결과 셀 범위를 지정한 후 [복사] → [선택하여붙여넣기] → [값 및 숫자서식]

The screenshot shows a Microsoft Excel spreadsheet with three columns: D (아이디), E (도메인), and F (메일주소). The formula `=D2&"@"&E2` is entered into cell F2. A red box highlights this formula. The formula bar also displays the formula. The '선택하여 붙여넣기' (Paste Special) dialog box is open, with the '값 및 숫자 서식(U)' (Values and Number Format) option selected. A red box highlights this option. The '선택하여 붙여넣기(S)' (Paste Special) button at the bottom of the dialog box is also highlighted with a red box and a green arrow pointing to it.

D	E	F
아이디	도메인	메일주소
james	naver.com	=D2&"@"&E2
andrew	hanmail.net	
linda	naver.com	
robert	gmail.com	
olivia	naver.com	
william	hanmail.net	

선택하여 붙여넣기

붙여넣기

- 모두(A)
- 수식(E)
- 값(V)
- 서식(F)
- 주석 및 메모(C)
- 유효성 검사(N)

연산

- 없음(O)
- 더하기(D)
- 빼기(S)

내용 있는 셀만 붙여넣기(B) 행/열 바꿈(E)

선택하여 붙여넣기(S) >

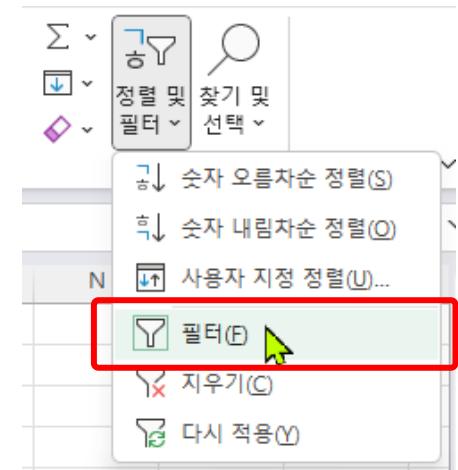
확인 취소



데이터 레코드(행) 정리

- 분석을 위한 기초 조건을 제시하여 원하는 레코드 셋을 추출 → 필터
- [홈]탭 → [편집]그룹 → [정렬및필터]메뉴 → [필터]메뉴

A	B	C	D	E	F	G
일련번호	판매일	제품명	거래처명	단가	수량	금액
1	1 09월 01			54,700	58	3,172,600
2	2 09월 01			96,700	18	1,740,600
3	3 09월 01			37,700	13	490,100
4	4 09월 02			96,700	28	2,707,600
5	5 09월 02			54,700	33	1,805,100
6	6 09월 03			13,700	18	246,600
7	7 09월 03			96,700	13	1,257,100
8	8 09월 03			13,700	13	178,100
9	9 09월 03			57,700	28	1,615,600
10	10 09월 04			37,700	8	301,600
11	11 09월 04			54,700	18	984,600
12	12 09월 07			32,700	18	588,600
13	13 09월 07			96,700	45	4,351,500
14	14 09월 07			13,700	38	520,600
15	15 09월 07			84,700	13	1,101,100
16	16 09월 08			32,700	33	1,079,100
17	17 09월 08			57,700	18	1,038,600
18	18 09월 08			84,700	18	1,524,600
19	19 09월 08			96,700	43	4,158,100
20	20 09월 08			37,700	28	1,055,600
21	21 09월 09일	알파인 맥주	동일상사	13,700	53	726,100
22	09월 09일	대익 푸자 치즈	동일상사	54,700	29	2,133,300



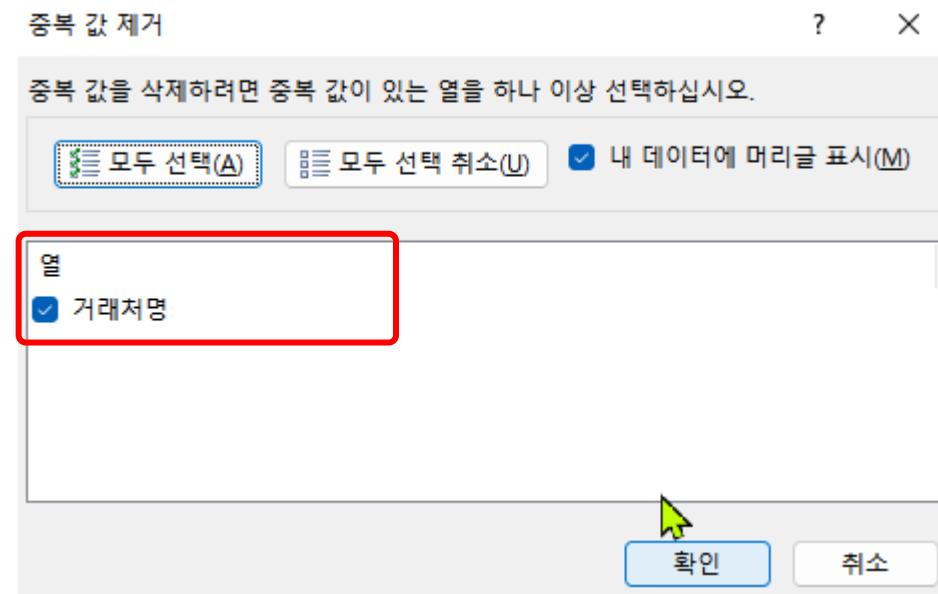


중복된 항목 제거

- 데이터 필드 중 범주형 데이터의 중복되는 값을 제거하여 범주(카테고리) 기준을 작성할 수 있음
- [데이터]메뉴탭 → [데이터도구]그룹 → [중복된 항목 제거] 메뉴



I	J	K	L	M	N
거래처명	거래처명				
대세상사	대세상사				
동일상사	동일상사				
삼보상사	삼보상사				
대세상사	원정유통				
원정유통	우성유통				
동일상사					
우성유통					
원정유통					
대세상사					
삼보상사					
삼보상사					
동일상사					
대세상사					
대세상사					





고급 필터로 데이터 레코드 추출

- 조건을 별도로 표시한 후 원하는 필드만 따로 추출 → 고급 필터
 - 필터 메뉴 실행하기 전에 미리 조건을 따로 필드명과 조건 값을 표시해야 함
 - 전체 필드가 아닌 원하는 필드만 추출하기 위하여 원하는 필드명만 미리 표시해두기도 함
- [데이터]메뉴탭 → [정렬및필터]그룹 → [고급]메뉴
 - 고급 필터 대화상자의 결과 옵션은 [다른 장소에 복사]를 선택

I
추출조건
제품명
게살 통조림

A	B	C	D	E	F	G	
1	일련번호	판매일	제품명	거래처명	단가	수량	금액
2	1	09월 01일	대일 포장 치즈	대세상사	54,700	58	3,172,600
3	2	09월 01일	게살 통조림	동일상사	96,700	18	1,740,600
4	3	09월 01일	유미 멸치 가루	삼보상사	37,700	13	500,100
5	4	09월 02일	게살 통조림	대세상사			
6	5	09월 02일	대일 포장 치즈	원정유통			
7	6	09월 03일	알파인 맥주	동일상사			
8	7	09월 03일	게살 통조림	우성유통			
9	8	09월 03일	알파인 맥주	원정유통			
10	9	09월 03일	우미 코코넛 쿠키	대세상사			
11	10	09월 04일	유미 멸치 가루	삼보상사			
12	11	09월 04일	대일 포장 치즈	삼보상사			
13	12	09월 07일	콜롬비아산 원두커피	동일상사			
14	13	09월 07일	게살 통조림	대세상사			
15	14	09월 07일	알파인 맥주	대세상사			
16	15	09월 07일	진미 트로피컬 캔디	원정유통			
17	16	09월 08일	콜롬비아산 원두커피	동일상사	32,700	33	1,079,100

고급 필터

추출조건
제품명
게살 통조림

필요한필드
판매일 거래처명 수량 금액

결과
 현재 위치에 필터(E)
 다른 장소에 복사(Q)

목록 범위(L): \$A\$1:\$G\$46
 조건 범위(C): \$I\$2:\$I\$3
 복사 위치(I): \$I\$7:\$L\$7

동일한 레코드는 하나만(R)

확인 취소



파워 쿼리의 이해

- Microsoft 社에서 제공하는 데이터 변환 및 데이터 편집 엔진 추가 기능
- Microsoft 社 정책으로 오피스 윈도우버전만 가능
(Mac버전 불가 → 2021년 7월부터 일부 제한적 제공)
- Office 2010 이후 개발된 “COM 추가 기능”으로 **2016버전 이상 데이터 메뉴에 포함**
(2010 Professional Plus, 2013 Home/Personal 버전은 추가 설치 필요)
- 데이터의 **ETL**(추출(Extract), 변환(Transform), 로드(Load)) 처리를 수행
- 데이터 변환과 가공을 위한 도구이자 프로그램

The screenshot shows a Microsoft Excel window titled "통합 문서1 - Excel". The ribbon is visible at the top, with the "데이터" (Data) tab selected. In the "데이터" tab's ribbon group, the "파일에서(F)" (From File) dropdown is open, showing options like "파일에서(F)", "데이터베이스에서(D)", "Azure에서(A)", "온라인 서비스에서(I)", "기타 원본에서(O)", and "쿼리 결합(Q)". Below this, a red box highlights the "파워 쿼리 편집기 시작(L)...". Another red box highlights the "Power Query 편집기" tab in the Power Query ribbon tab's dropdown menu. The main workspace shows a grid of columns D through J. A large red box highlights the "변환" (Transform) tab in the Power Query ribbon tab's dropdown menu.



파워 쿼리의 시작

- 파워 쿼리 핵심 목표는 원활한 데이터 분석을 위하여
“원시 데이터를 분석 기능 적용이 가능하도록 효율적으로 가공하는 것”
- 분석 데이터의 형태

제목	제목	제목	제목
값	값	값	값
값	값	값	값
값	값	값	값

첫 번째 행에는 각 열의 제목이 입력됩니다. 엑셀에서 제목은 ‘머리글’이라고 하며, 이 경우 열을 대표하는 제목이라 하여 ‘열 머리글’이라고 합니다.

두 번째 행부터 입력된 값이 표의 데이터입니다.

- 분석 데이터 형태가 아닌 표를 분석 데이터 형태로 가공할 때 편리한 기능 제공
- 파워 쿼리를 실행하기 위하여 데이터 범위를 표(Table)로 만드는 작업이 필요
- 데이터 표의 구조를 기반으로 파워 쿼리 기능 실행



파워쿼리의 표

- 파워쿼리는 데이터 관련 처리를 편리하게 해주는 추가 기능
- 외부 데이터 가져오기를 실행하면 쿼리 표로 로드됨
 - 쿼리를 시트로 로드한 후 시트의 표에서 쿼리만 삭제 가능

X | ☺ | 거래상위 TOP 종목 종목명에 대한 현재가, 전일대비, 등락률로 구분되어 있습니다 - Power Query 편집기

파일 홈 변환 열 추가 보기

닫기 및 로드 ▾ 속성 고급 편집기 열 관리 행 관리 정렬 데이터 형식: 텍스트 ▾ 첫 행을 머리글로 사용 ▾ 결합 매개 변수 관리 데이터 원본 설정 새 원본 ▾ 최근 원본 ▾ 새 쿼리

닫기 미리 보기 새로 고침 관리 쿼리

닫기 및 로드 ▾

= Table.TransformColumnTypes(Data0,{{"종목명", type text}, {"현재가", type number}, {"전일대비", type number}, {"등락률", type number}})

쿼리 설정

A	B	C	D	E
종목명	현재가	전일대비	등락률	
KODEX 200선물인버스2X	2840	하락 15	-0.53%	
CBI	674	상승 18	2.74%	
보성파워텍	6220	상승 190	3.15%	
조광IL	1650	상승 65	4.10%	
국전약품	9480	상승 1,380	17.04%	
신한 인버스 2X WTI원유 선...	125	상승 10	8.70%	
메가엠디	4130	상승 950	29.87%	
에이프로젠	1535	상승 25	1.66%	
삼성 인버스 2X WTI원유 선...	135	상승 10	8.00%	
휴림로봇	3325	상승 115	3.58%	
프로미천	2835	상승 470	19.87%	
KODEX 레버리지	16530	상승 95	0.58%	
KODEX 인버스	4655	하락 5	-0.11%	
아이비김영	2640	상승 175	7.10%	
KODEX 코스닥150선물인버스	4655	하락 25	-0.53%	

쿼리 및 연결

쿼리 | 연결

거래상위 TOP 종목 종목명...
15개의 행이 로드되었습니다.



파워쿼리의 표

- 쿼리: 데이터 구조적 편집 편의를 위한 가상의 임시 표
- 쿼리 표를 워크 시트의 기본 표로 변환하기 위하여 쿼리 삭제 가능
 - [쿼리]메뉴 탭 → [삭제]메뉴 또는 쿼리 작업 창에서 단축메뉴의 [삭제]메뉴
 - 쿼리를 삭제하면 [쿼리]메뉴탭이 표시되지 않음

The screenshot shows the Microsoft Excel ribbon with the 'Query & Power BI' tab selected (highlighted by a red box). In the main area, a floating 'Query Remove' dialog box is displayed, also with a red box highlighting its 'Delete' button. The Power Query ribbon's 'Home' tab is also highlighted with a red box.

Query Remove Dialog Content:

'거래상위 TOP 종목 종목명에 대한 현재가, 전일대비, 등락률로 구분되어 있습니다 (2)'을(를) 삭제하시겠습니까? 이 쿼리를 삭제하면 이 쿼리에 의해 로드된 데이터를 새로 고칠 수 없습니다.

Buttons in Dialog:

- 삭제 (highlighted with a red box)
- 취소

Power Query Ribbon Buttons:

- 삭제 (highlighted with a red box)
- 새로고침
- 로드
- 대상
- 재사용
- 병합
- 추가
- 결합
- 공유
- 데이터 카탈로그에 보내기
- Power BI



파워 쿼리의 실행

● 데이터 범위를 표로 삽입

- [삽입]메뉴탭 → [표]
- [표도구][디자인] → [표이름]지정을 권유: 쿼리를 구분하기 위한 이름

● 지정된 표를 선택 → [데이터]메뉴탭 → [테이블/범위에서]메뉴 실행 → 파워쿼리 실행

The screenshot shows a Microsoft Excel interface with the following details:

- Menu Bar:** 파일 (File), 홈 (Home), 삽입 (Insert), 그리기 (Draw), 페이지 레이아웃 (Page Layout), 수식 (Formulas), **데이터 (Data)**, 검토 (Review), 보기 (View), 개발 도구 (Developer Tools), 도움말 (Help).
- Toolbars:** 텍스트/CSV에서 (Text/CSV from), 최근에 사용한 원본 (Recent Sources), 웹 (Web), 기존 연결 (Existing Connection), **테이블/범위에서 (Table/Range)** (highlighted with a red box), 모두 새로 고침 (All Refresh), 속성 (Properties), 연결 편집 (Edit Connection), 주식 (영어) (Stock (English)), 통화 (Currency).
- Cells:** D10 is selected.
- Buttons:** 대세상사 (Large Supplier) button.
- Data Table:** A table with columns: 일련번호 (Serial Number), 판매일 (Sales Date), 제품명 (Product Name), 거래처명 (Supplier Name), 단가 (Unit Price), 수량 (Quantity), 금액 (Amount). The data rows are as follows:

일련번호	판매일	제품명	거래처명	단가	수량	금액
1	09월 01일	대일 포장 치즈	대세상사	54,700	58	3,172,600
2	09월 01일	게살 통조림	동일상사	96,700	18	1,740,600
3	09월 01일	유미 멜치 가루	삼보상사	37,700	13	490,100
4	09월 02일	게살 통조림	대세상사	96,700	28	2,707,600
5	09월 02일	대일 포장 치즈	원정유통	54,700	33	1,805,100
6	09월 03일	알파인 맥주	동일상사	13,700	18	246,600
7	09월 03일	게살 통조림	우성유통	96,700	13	1,257,100
8	09월 03일	알파인 맥주	원정유통	13,700	13	178,100
9	09월 03일	우미 코코넛 쿠키	대세상사	57,700	28	1,615,600
10	09월 04일	유미 멜치 가루	삼보상사	37,700	8	301,600
11	09월 04일	대일 포장 치즈	산부산사	54,700	18	984,600



파워 쿼리의 실행

● 파워 쿼리 창

- [파일] / [홈] / [변환] / [열추가] / [보기] 메뉴 탭으로 구성

데이터세트1 - Power Query 편집기

파일 **홈** 변환 열 추가 보기

닫기 및 로드 미리 보기 새로 고침 속성 고급 편집기 관리 열 선택 제거 행 유지 제거 행 감소 정렬 열 분할 그룹화 첫 행을 머리글로 사용 1 2 값 바꾸기 결합 매개 변수 관리 데이터 원본 설정 새 원본 최근 원본 데이터 입력 새 쿼리

쿼리 설정

속성
이름: 데이터세트1
모든 속성

적용된 단계
원본
변경된 유형

1	2	3	4
1	2022-09-01 오전 12:00:00	대일 포장 치즈	대세상사
2	2022-09-01 오전 12:00:00	게살 통조림	동일상사
3	2022-09-01 오전 12:00:00	유미 멸치 가루	삼보상사
4	2022-09-02 오전 12:00:00	게살 통조림	대세상사
5	2022-09-02 오전 12:00:00	대일 포장 치즈	원정유통
6	2022-09-03 오전 12:00:00	알파인 맥주	동일상사
7	2022-09-03 오전 12:00:00	게살 통조림	무성유통
8	2022-09-03 오전 12:00:00	알파인 맥주	원정유통
9	2022-09-03 오전 12:00:00	우미 코코넛 쿠키	대세상사
10	2022-09-04 오전 12:00:00	우미 면류 기름	사부사파

10



파워 쿼리의 시작

● [홈]메뉴탭

The screenshot shows the Power Query ribbon with the 'Home' tab highlighted by a red box. The ribbon includes tabs for File, Home, Transform, Add, and View. Under the Home tab, there are sections for File (Close & Load, Save As, Refresh, etc.), Transform (Select, Filter, Sort, Group, Unpivot, etc.), and View (Change, Pivot, etc.). A yellow arrow points to the 'Close & Load' icon in the File section.

● [변환]메뉴탭

The screenshot shows the Power Query ribbon with the 'Transform' tab highlighted by a red box. The ribbon includes tabs for File, Home, Transform, Add, and View. Under the Transform tab, there are sections for Group (Group Rows, First Row as Headers), Pivot (Pivot Row, Pivot Column, Pivot Both), and Text (Text, Number, Date, Time, Duration).

● [열추가]메뉴탭

The screenshot shows the Power Query ribbon with the 'Add' tab highlighted by a red box. The ribbon includes tabs for File, Home, Transform, Add, and View. Under the Add tab, there are sections for General (New User, User Definition, Index, Union, Append, etc.), From Text (Text, CSV, JSON, etc.), From Numbers (Sum, Product, etc.), and From Dates (Date, Time, Duration).



파워 쿼리 표

- 파워 쿼리 창의 [홈]메뉴탭 → [닫기및로드]메뉴 → 워크시트 표 생성
- 시트의 쿼리 및 연결 작업 창 표시

The screenshot shows the Microsoft Power Query interface. On the left is a table of sales data with columns: 일련번호 (Row ID), 판매일 (Sale Date), 제품명 (Product Name), 거래처명 (Seller Name), 단가 (Unit Price), 수량 (Quantity), and 금액 (Amount). The table has 16 rows of data from September 1 to 7, 2022.

In the center, a '쿼리 및 연결' (Query & Connect) dialog box is open, showing a single data set named '데이터세트1' (Dataset1) with 45 rows loaded. A red box highlights this area.

To the right, a larger '쿼리 및 연결' (Query & Connect) window is open, showing the same dataset. A red box highlights the '편집' (Edit) option in the ribbon menu, which is also highlighted with a yellow arrow. The ribbon menu includes options like 복사 (Copy), 투여보기 (Preview), 이름 바꾸기 (Change Name), 새로 고침 (Refresh), 다음으로 로드... (Load Next...), 복제 (Copy), 참조 (Reference), 병합 (Merge), 추가 (Add), 연결 파일 내보내기... (Export Connection File...), 그룹으로 이동 (Move to Group), 위로 이동 (Move Up), 아래로 이동 (Move Down), 미리 보기 표시 (Show Preview), and 속성... (Properties...).

At the top right, there is a small window titled '닫기 및 로드' (Close & Load) with a '닫기' (Close) button.



파워 쿼리로 데이터 가공 셀 병합 문제 해결

● 데이터 범위를 표로 삽입

- ‘머릿글 포함’ 옵션 해제
- [표 디자인] - 표 스타일 옵션에서 '머릿글 행'을 체크 해제 → ‘표스타일을 없음’으로 기존 범위와 동일한 디자인 표

담당부서	직원정보	매출		
		목표	실적	매출이익
영업1팀	정희엘(강남)	5,680,000	6,192,000	433,500
	박단비(영등포)	6,310,000	7,131,000	998,400
	정진하(구로)	7,080,000	8,426,000	674,100
	김병민(용산)			54,400
	이제우(종로)			219,300
	김준용(암구정)			-49,800
영업2팀	박정화(을지로)			662,700
	이서우(서초)			824,200
	최리(잠실)	5,760,000	5,300,000	795,000
	김교은(뚝섬)	7,360,000	7,581,000	227,500
	최효율(신촌)	6,050,000	6,292,000	-629,200
	이유림(홍대)	7,270,000	7,198,000	647,900

담당부서	직원정보	매출		
		목표	실적	매출이익
영업1팀	정희엘(강남)	5,680,000	6,192,000	433,500
	박단비(영등포)	6,310,000	7,131,000	998,400
	정진하(구로)	7,080,000	8,426,000	674,100
	김병민(용산)	5,440,000	5,440,000	54,400
	이제우(종로)	7,860,000	7,310,000	219,300
	김준용(암구정)	5,990,000	4,972,000	-49,800
영업2팀	박정화(을지로)	5,160,000	5,522,000	662,700
	이서우(서초)	6,630,000	7,492,000	824,200
	최리(잠실)	5,760,000	5,300,000	795,000
	김교은(뚝섬)	7,360,000	7,581,000	227,500
	최효율(신촌)	6,050,000	6,292,000	-629,200
	이유림(홍대)	7,270,000	7,198,000	647,900



파워 쿼리로 데이터 가공 셀 병합 문제 해결

- [데이터] - [테이블 범위에서]를 클릭하면 파워쿼리 편집기가 실행
- 파워쿼리 편집기가 실행되면서 기본적인 데이터 가공 단계가 추가
 - 기존 단계가 있다면, 단계 왼쪽에 있는 [X] 버튼을 클릭해서 기존 단계를 모두 제거

		매출		
	담당부서	직원정보	목표	실적
5	영업1팀	정희열(강남)	5,680,000	6,192,000
6		박단비(영등포)	6,310,000	7,131,000
7		정진하(구로)	7,080,000	8,426,000
8		김병민(용산)	5,440,000	5,440,000
9		이제우(종로)	7,860,000	7,310,000

	L1: Column1	L1: Column4	L1: Column5
1	null	null	null
2	목표	실적	매출이익
3			
4	5,680,000	6,282,000	433,500
5	6,310,000	7,131,000	998,400
6	7,080,000	8,426,000	674,100
7	5,440,000	5,440,000	54,400
8	7,860,000	7,310,000	219,300
9	null	null	null
10	이제우(종로)	7,310,000	219,300
11	null	7,310,000	219,300
12	null	7,310,000	219,300
13	null	7,310,000	219,300
14	null	7,310,000	219,300

쿼리 설정

- 작성된 단계
- 필터



파워 쿼리로 데이터 가공 셀 병합 문제 해결

- [홈] - [행 제거] - [상위 행 제거]를 선택
 - 제거할 행 개수로 1로 입력하여 첫번째 행 제거
- 표의 좌측 상단의 [표 모양 아이콘]을 클릭 - [상위 행 제거]를 선택도 같은 작업

The screenshot shows the Microsoft Power Query Editor interface. On the left, there is a table with five columns and 14 rows of data. A red dashed box highlights the 'Remove Top Rows' dialog box, which has a title '상위 행 제거' and a note '맨 위에서부터 제거할 행 수를 지정합니다.' Below is a text input field with the value '1'. At the bottom right of the dialog are '확인' (Confirm) and '취소' (Cancel) buttons. To the right of the table, the ribbon menu is open, and a red arrow points to the '상위 행 제거...' (Remove Top Rows...) option under the '행' (Row) section of the '행 제거' (Delete Row) dropdown.



파워 쿼리로 데이터 가공 셀 병합 문제 해결

● 데이터 첫 행을 머릿글로 승격

- [홈] - [첫 행을 머리글로 사용]을 클릭
- 표 좌측 상단 [표 모양 아이콘]을 클릭 - [첫 행을 머리글로 사용]을 클릭

The screenshot shows the Microsoft Power Query Editor interface. The ribbon menu at the top has '데이터 헤더 텍스트' (Data Header Text) selected. A red box highlights the '첫 행을 머리글로 사용' (Use First Row as Headers) button in the ribbon. On the left, there's a table preview showing data from '서울 담당부서' (Seoul Department) to '이유원(충남)'). On the right, the '쿼리 설정' (Query Settings) pane is open, showing '속성' (Properties) with '이름' (Name) set to '표1' (Table1), and '작용된 단계' (Applied Steps) with '변경된 유형' (Changed Type) highlighted.

서울 담당부서	제작원정보	목표	실적	잔율이익
영업1팀	한희열(강남)	5680000	6192000	433500
	null 박단비(영등포)	6310000	7131000	998400
	null 한진희(구로)	7080000	8426000	674100
	null 김병민(용산)	5440000	5440000	54400
	null 이재우(중로)	7860000	7310000	219300
	null 김준룡(강구점)	5950000	4972000	-49800
	null 박정화(용지로)	5160000	5522000	662700
영업2팀	null 이세우(서초)	6630000	7492000	824200
	null 홍민(잠실)	5760000	5300000	795000
	null 김교은(목동)	7360000	7581000	227500
	null 최효문(신촌)	6050000	6292000	-629200
	null 이유원(충남)	7270000	7198000	647900



파워 쿼리로 데이터 가공 셀 병합 문제 해결

● 병합된 셀 채우기

- 범위가 표로 변환되면서 셀 병합이 해제되면, 병합된 범위의 첫번째 셀에만 값이 입력되고 나머지 셀은 빈칸이 반환
- 표의 '담당부서' 열을 선택 후, [변환] - [채우기] - [아래로]를 클릭

The screenshot shows the Microsoft Power Query Editor interface. On the left, there is a preview of a table with columns '담당부서' and '직원정보'. The '담당부서' column contains values like '영업1팀', '영업2팀', etc., with many rows having 'null' or empty values. A blue arrow points from this preview to the main editor area. In the main editor, the 'Transform' ribbon tab is selected. The 'Fill' dropdown menu is open, showing options: '행/열 바꿈' (Row/Column Swap), '데이터 형식' (Data Type), '이름 바꾸기' (Rename), '채우기' (Fill), and '아래로' (Down). The '채우기' option is highlighted with a red box. Below the ribbon, the formula bar shows the command: `- Table.FillDown(#"변경된 유형", {"담당부서"})`. The bottom half of the screen displays the transformed table, where the '담당부서' column now contains consistent values ('영업1팀', '영업2팀', etc.) across all rows, indicating that the 'Down' fill operation was successful.

#	A: 담당부서	B: 직원정보	C: 목표	D: 실적
1	영업1팀	김희율(남)	5680000	6292000
2	영업1팀	박단비(여성)	6310000	7131000
3	영업1팀	한진하(구조)	7080000	8426000
4	영업1팀	김별민(윤산)	5440000	5440000
5	영업1팀	이제우(솔로)	7860000	7310000
6	영업1팀	김준룡(압구정)	5990000	4972000
7	영업2팀	곽정화(용자로)	5160000	5522000
8	영업2팀	이서무(서초)	6630000	7492000
9	영업2팀	최리(잠실)	5760000	5300000
10	영업2팀	김교은(뚝섬)	7360000	7581000
11	영업2팀	김효운(신촌)	6050000	6292000
12	영업2팀	이유월(송파)	7270000	7198000



계산 열 자동 추가

- 가공된 데이터에 계산이 필요한 열을 자동으로 추가: ‘매출이익률’
- 매출이익률은 매출이익 ÷ 실적으로 계산
- [매출이익] 열을 선택 후, [열추가] - [표준] - [나누기]를 선택

	123 목표	123 실적	123 매출이익
1 영업1팀	5680000	6192000	433500
2 영업1팀	6310000	7131000	998400
3 영업1팀	7080000	8426000	674100
4 영업1팀	5440000	5440000	54400
5 영업1팀	7860000	7310000	219300
6 영업1팀	5990000	4972000	-49800
7 영업2팀	5160000	5522000	662700
8 영업2팀	6630000	7492000	824200
9 영업2팀	5760000	5300000	795000
10 영업2팀	7360000	7581000	227500
11 영업2팀	6050000	6292000	-629200
12 영업2팀	이유립(총대)		

- 나눠줄 범위로
[열의 값에서]를 선택 후,
[실적] 범위를 선택



계산 열 자동 추가

- 가공된 데이터에 계산이 필요한 열을 자동으로 추가: '달성을'
- "실적" 열을 선택 후, [열추가] - [표준] - [나누기] - [목표] 열을 선택 후 [확인]을 누르면 달성을이 계산
- 각 머리글을 더블클릭하여 '매출이익률'과 '달성을'로 변경
- 각 머리글 왼쪽에 있는 데이터 형식 아이콘을 클릭, '백분율'로 데이터형식 변경

	작명정보	액표	실적	매출이익	1.2 달성률
1	전화영(김남)	5680000	6192000	433500	
2	박단비(김종호)	6310000	7331000	998400	
3	절진현(구로)	7080000	8426000	674100	
4	김별민(용산)	5440000	5440000	54400	
5	이제우(홍토)	7860000	7310000	219300	
6	김준룡(인구철)	5990000	6972000	-49800	
7	박경화(문지호)	5160000	5522000	662700	
8	이서우(서호)	6630000	7492000	824200	
9	최관(강실)	5760000	5300000	795000	
10	김교근(죽설)	7380000	7381000	227500	
11	최효운(신촌)	6050000	6292000	-629200	
12	이유람(송대)	7270000	7188200	647900	

나누기
열의 각 값을 나누는데 사용할 숫자를 입력합니다.

기준 목표

확인 취소

액	1.2 매출이익률	1.2 달성률
433500	0.07000969	1.090140845
998400	0.140008414	1.130110935
674100	0.080002374	1.190112994
54400	0.01	1
219300	0.03	0.930025445
-49800	-0.01001609	0.830050083
662700	0.120010866	1.070155039
824200	0.110010678	1.130015083

% 매출이익률	1.2 달성률
500	7.00%
400	14.00%
100	8.00%
400	1.00%
300	3.00%
500	-1.00%
700	12.00%
200	11.00%
500	15.00%

1.2 10진수
\$ 통화
정수
% 백분율
날짜/시간
날짜
시간
날짜/시간/표준 시간대
기간

데이터 분석 기능을 적용할 수 있는 올바른 데이터 구조



올바른 데이터 구조 - 3가지 규칙

첫번째

1행은 머릿글, 1열은 ID(대표값)으로 입력합니다.

두번째

병합된 셀은 없어야 합니다.

세번째

합계/소계가 Raw Data에 포함되면 안 됩니다.

네번째

각 필드는

정해진 데이터 형태로만

입력되어야 합니다.

(숫자면 숫자, 날짜면 날짜, 등)

직원명	1분기	2분기	3분기	4분기
정희엔	4,900	4,800	3,800	4,400
박단비	7,700	6,400원	7,100	4,600
정진하	7,800	7,000	5,600	7,900
김병민	6,700	6,100	6,100	5,200
이제우	4,300	7,900원	3,400	4,900
김준용	5,900	4,900	7,800	4,900
박정화	4,400	3,300	6,000	3,200
이서우	4,000	4,100	7,700	5,600
최리	5,200	4,200	4,000	6,100
김교은	4,000	4,600	3,600	7,900
최효윤	4,900	4,100	5,100	6,500



표의 가공

● 표의 종류

- 분석을 위한 기본 표(테이블) 형태
- 요약 및 정리된 크로스-탭 형태
- 특정 양식이 적용되었거나 입력의 편의를 위해 편집된 양식 모양 템플릿 형태

● 기본 표 형태로 구성해야 분석 기능을 적용할 수 있음

제목	제목	제목	제목
제목	집계 값	집계 값	집계 값
제목	집계 값	집계 값	집계 값
제목	집계 값	집계 값	집계 값

첫 번째 행의 값은 열 머리글입니다.

열 머리글과 행 머리글이 교차(Cross)하는 위치에 테이블의 값을 집계한 요약 값이 들어갑니다.

첫 번째 열의 값은 행의 제목이라는 뜻에서 '행 머리글'이라고 합니다.



제목	제목	제목	제목
값	값	값	값
값	값	값	값
값	값	값	값

첫 번째 행에는 각 열의 제목이 입력됩니다. 엑셀에서 제목은 '머리글'이라고 하며, 이 경우 열을 대표하는 제목이라 하여 '열 머리글'이라고 합니다.

두 번째 행부터 입력된 값이 표의 데이터입니다.



파워 쿼리의 활용

- 파워쿼리에서 데이터의 표시 방향과 표의 구조를 변경 가능

- 행/열 바꿈 기능

- 피벗 열 기능

- 분석 데이터 표 형태로 변환

- 생산내역 단위 기준 값이 열 제목으로 표시된 가로세로 템플릿 표 형식

	A	B	C	D	E	F	G	H	I	J	K	L	
1	+ 직원 정보			생산 내역(단위 : mm)									
2	부서	사번	성명	250	255	260	265	270	275	280	285	총계	
3	생산1팀	S-0328	강영광	3	2	5	5	3	4	2	2	26	
4	생산1팀	S-0324	문분홍	3	2	4	7	4	4	4	3	31	
5	생산1팀	S-0341	박영재	2	2	6	7	4	4	2	3	30	
6	생산1팀	S-0331	강다래	5	4	4	3	7	4	1	1	29	
7	생산1팀	S-0384	안민주	3	4	4	8	8	3	4	3	37	
8	생산1팀	S-0361	박단비	2	1	5	6	6	4	2	3	29	
9	생산2팀	S-0390	이보석	4	4	4	6	5	3	2	2	30	
10	생산2팀	S-0376	전솔	2	3	5	8	8	4	3	2	35	
11	생산2팀	S-0323	김민지	5	1	8	7	7	5	5	1	39	
12	생산2팀	S-0322	성보람	5	5	7	5	7	1	2	3	35	
13	생산2팀	S-0317	장소리	1	2	7	6	6	5	5	2	34	
14	생산2팀	S-0325	정공주	2	2	3	8	5	4	5	1	30	
15													

파워 쿼리의 활용



- 속성 값으로 변환할 열을 선택
 - 열머리글을 클릭 → <Shift>키 + 마지막 열머리글 클릭

= Table.TransformColumnTypes(원본, {{"부서", type text}, {"사번", type text}})

	1	2	3	4	5	6	7	8	9	10	11	12
1	5		3		4		2		1	2	3	2
2	7		4		4		4		3		3	
3	7		4		4		2		3		3	
4	3		7		4		1		1		1	
5	8		8		3		4		3		3	
6	6		6		4		2		3		3	
7	6		5		3		2		2		2	
8	8		8		4		3		2		2	
9	7		7		5		5		1		1	
10	5		7		1		2		3		3	
11	6		6		5		5		2		2	
12	8		5		4		5		1		1	

- [변환]메뉴 탭 → [열 피벗 해제]메뉴



파워 쿼리의 활용

● [열 피벗 해제]로 세로 형태의 표로 변환

- 변환된 열 머리글 이름 변경
- 데이터 형식 변경

	A ^B _C 사번	A ^B _C 설명	A ^B _C 단위	1 ² ₃ 생산량	
1	S-0328	강영광	250		3
2	S-0328	강영광	255		2
3	S-0328	강영광	260		5
4	S-0328	강영광	265		5
5	S-0328	강영광	270		3
6	S-0328	강영광	275		4
7	S-0328	강영광	280		2
8	S-0328	강영광	285		2
9	S-0324	문분홍	250		3
10	S-0324	문분홍	255		2
11	S-0324	문분홍	260		4
12	S-0324	문분홍	265		7
13	S-0324	문분홍	270		4
14	S-0324	문분홍	275		4
15	S-0324	문분홍	280		4
16	S-0324	문분홍	285		3
17	S-0324	박연재	250		2

The screenshot shows a Power Query editor interface. A context menu is open over the last row of the table, specifically over the '단위' column cell for row 17. The menu is titled 'A^B_C 단위' and includes options like '1.2 10진수', '\$ 통화', '1²₃ 정수', '% 백분율', etc. The '1.2 10진수' option is highlighted with a red box.



파워 쿼리의 활용

● 세로 형태의 분석 가능한 표 모양 변환 결과

	A	B	C	D	E	F
1	부서	사번	성명	단위	생산량	
2	생산1팀	S-0328	강영광	250	3	
3	생산1팀	S-0328	강영광	255	2	
4	생산1팀	S-0328	강영광	260	5	
5	생산1팀	S-0328	강영광	265	5	
6	생산1팀	S-0328	강영광	270	3	
7	생산1팀	S-0328	강영광	275	4	
8	생산1팀	S-0328	강영광	280	2	
9	생산1팀	S-0328	강영광	285	2	
10	생산1팀	S-0324	문분홍	250	3	
11	생산1팀	S-0324	문분홍	255	2	
12	생산1팀	S-0324	문분홍	260	4	
13	생산1팀	S-0324	문분홍	265	7	
14	생산1팀	S-0324	문분홍	270	4	
15	생산1팀	S-0324	문분홍	275	4	
16	생산1팀	S-0324	문분홍	280	4	

쿼리 및 연결

쿼리 | 연결

5개 쿼리

- 데이터세트1
연결 전용입니다.
- 영업매출1
12개의 행이 로드되었습니다.
- 영업매출2
12개의 행이 로드되었습니다.
- 병합1
12개의 행이 로드되었습니다.
- 생산내역
96개의 행이 로드되었습니다.



표 합치기

- 파워 쿼리는 여러 개의 표를 하나의 표로 합칠 수 있는 병합과 추가 기능 제공
- [추가]는 표에 데이터를 바로 추가하여 하나의 표처럼 사용할 수 있는 기능
- 표 2개 이상을 아래로 계속 추가하여 하나의 표로 만듦

The screenshot shows the Microsoft Power Query Editor interface with two main windows and several toolbars.

Left Window (Main View):

- Toolbar:** 파일, 풀 (highlighted), 변환, 열 추가, 보기.
- Left pane:** 데이터세트1 (영업매출1) is selected.
- Table:** 영업1팀, 영업2팀, 영업3팀, 영업4팀, 영업5팀, 영업6팀, 영업7팀, 영업8팀, 영업9팀, 영업10팀, 영업11팀, 영업12팀.
- Bottom pane:** 추가1 (highlighted).

Right Window (Query Settings):

- Toolbar:** 데이터 형식: 텍스트, 첫 행을 머리글로 사용, 그룹화, 분할, 결합 (highlighted), 열 관리, 행 감소, 정렬, 변환.
- Left pane:** 추가 (2개의 테이블 선택).
- Center pane:** 두 번째 테이블 영업매출2 (highlighted).
- Right pane:** 쿼리 설정 (속성, 이름: 추가1, 모든 속성, 적용된 단계: 원본).

Bottom Table:

A ^B 담당부서	A ^B 직원정원	1 ² 목표	1 ² 실적	1 ² 매출이익
영업1팀	정희열(강남)	5500000	5300000	-200000
영업1팀	박단비(영등포)	6000000	5800000	-200000
영업1팀	정진하(구로)	7080000	8426000	1346000
영업1팀	김병민(용산)	5440000	5440000	0
영업1팀	이재우(종로)	7860000	7310000	-550000
영업1팀	김준용(아그로)	6000000	6070000	70000
영업2팀	박정화	5500000	5300000	-200000
영업2팀	이서우	6630000	7492000	824000
영업2팀	최리(점)	5760000	5300000	-460000
영업2팀	김교은	7360000	7581000	227000
영업2팀	최효윤	6050000	6292000	-242000
영업2팀	이유임	7270000	7198000	-72000
영업2팀	정희열(강남)	5810000	4997000	-813000
영업1팀	박단비(영등포)	6770000	6093000	-677000
영업1팀	정진하(구로)	7700000	8085000	385000
영업1팀	김병민(용산)	6440000	5925000	-515000
영업1팀	이재우(종로)	5420000	5637000	217000
영업2팀	김준용(압구정)	7430000	8099000	669000
영업2팀	박정화(을지로)	5360000	5628000	268000
영업2팀	이서우(서초)	6950000	7159000	209000
영업2팀	최리(잠실)	5280000	5228000	-52000
영업2팀	김교은(뚝섬)	6670000	6070000	-600000



표 병합

- [병합]은 두 표를 기준 필드 값을 기반으로 하나의 표로 합치는 기능
 - 두 표를 비교하거나 하나로 병합 해 새로운 표를 만들 때 두 표를 키 필드(값)으로 연결해 하나로 병합한 쿼리 생성 가능
 - 두 표가 서로 다른 정보를 가지고 있으나 공통의 키 필드로 연결하여 합치는 기능
 - 연결 종류

조인	설명	다이어그램
왼쪽 외부	첫 번째 쿼리와 두 번째 쿼리를 연결할 때, 첫 번째 쿼리의 모든 키 값은 다 사용하고, 두 번째 쿼리에만 존재하는 키 값은 제외합니다.	
오른쪽 외부	첫 번째 쿼리와 두 번째 쿼리를 연결할 때, 두 번째 쿼리의 모든 키 값은 다 사용하고, 첫 번째 쿼리에만 존재하는 키 값은 제외합니다.	
완전 외부	첫 번째, 두 번째 쿼리의 모든 키 값을 사용해 연결합니다.	
내부	첫 번째와 두 번째 쿼리에 모두 존재하는 키 값으로만 연결합니다. 이렇게 하면 중복된 데이터를 확인할 수 있습니다.	
왼쪽 앤티	첫 번째 쿼리에만 존재하는 키 값으로만 연결합니다. 첫 번째 쿼리와 두 번째 쿼리에 모두 존재하는 키 값은 제외됩니다. 이렇게 하면 첫 번째 쿼리에만 존재하는 값을 확인할 수 있습니다.	
오른쪽 앤티	두 번째 쿼리에만 존재하는 키 값으로만 연결합니다. 두 번째 쿼리와 첫 번째 쿼리에 모두 존재하는 키 값은 제외됩니다. 이렇게 하면 두 번째 쿼리에만 존재하는 값을 확인할 수 있습니다.	



표 병합

● 먼저, 표를 쿼리로 등록하여 로드

Two screenshots of Microsoft Power BI Data Editor showing the process of loading tables into queries.

Screenshot 1: A screenshot of the Power BI Data Editor interface. The ribbon shows '파일', '홀드', '변환', '열 추가', and '보기'. The '홀드' tab is selected. The main area displays a table with columns '품번' (Product ID), '제품' (Product), and '미월' (Month). A red box highlights the '쿼리 설정' (Query Settings) dialog box, which is open over the table. The dialog shows '속성' (Properties) with '이름' (Name) set to '입고' (Import) and '모든 속성' (All Properties) checked. Under '적용된 단계' (Applied Steps), '원본' (Original) and '변경된 유형' (Changed Type) are listed. The table below shows the imported data.

A	B	C	D	E
1	품번	제품	출고	
2	P00151	클라리에저복사기 xi...	15	15
3	P00879	라미저복합기 L200	20	20
4	P00944	도트 TIC-7A	69	65
5	P01553	잉크젯복합기 AP-5500	20	20
6	P02063	링제본기 ST-200X	40	40
7	P02126	와이어제본기 WC-51...	61	55
8	P02378	도트 TIC-1A	35	35
9	P02466	오피스 Z-03	25	25
10	P02615	고급복사기 A4 1000매	25	25
11	P03280	무한레이저복합기 L...	15	15
12	P03412	RF OA-300	35	35
13	P03418	문서세단기 SCUT-1000	20	20
14	P03440	와이어제본기 WC-55...	20	20

Screenshot 2: Another screenshot of the Power BI Data Editor interface, showing a different table with columns '품번', '제품', and '출고'. A red box highlights the '쿼리 설정' (Query Settings) dialog box, which is open over the table. The dialog shows '속성' (Properties) with '이름' (Name) set to '출고' (Export) and '모든 속성' (All Properties) checked. Under '적용된 단계' (Applied Steps), '원본' (Original) and '변경된 유형' (Changed Type) are listed. The table below shows the exported data.

A	B	C	D	E
1	품번	제품	출고	
2	P03412	RF OA-300	29	
3	P07941	RF OA-400	34	
4	P02615	고급복사기 A4 1000매	26	
5	P09121	고급복사기 A4 2500매	81	
6	P02378	도트 TIC-1A	56	
7	P00944	도트 TIC-7A	123	
8	P00879	레이저복합기 L200	11	
9	P03984	레이저복합기 L500	9	
10	P06000	레이저복합기 L800	114	
11	P08111	레이저복합기 L950	11	
12	P07535	링제본기 ST-100	10	
13	P02063	링제본기 ST-200X	20	
14	P08292	무한레이저복합기 L500C	15	
15	P03418	문서세단기 SCUT-1000	5	
16	P09742	바코드 BCD-100 Plus	10	
17	P05409	바코드 BCD-200 Plus	107	
18	P05049	바코드 Z-750	106	
19	P07875	복사기 A4 5000매	8	
20	P08393	복사기 A4 500매	26	
21	P04785	열제본기 TB-8200	22	
22	P02126	와이어제본기 WC-5100	102	
23	P03440	와이어제본기 WC-5500	35	
24	P08686	잉크젯복합기 AP-3200	8	
25	P07777	이그제브차기 A4 2200	13	

Right Panel (Common for both screenshots):

- 쿼리 및 연결** (Queries and Connections)
 - 9개 쿼리** (9 queries)
 - 데이터세트1** (Dataset 1): 연결 전용입니다. (Connected only)
 - 영업매출1** (Sales 1): 12개의 행이 로드되었습니다. (12 rows loaded)
 - 영업매출2** (Sales 2): 12개의 행이 로드되었습니다. (12 rows loaded)
 - 생산내역** (Production History): 96개의 행이 로드되었습니다. (96 rows loaded)
 - 지역강수량** (Regional Rainfall): 216개의 행이 로드되었습니다. (216 rows loaded)
 - 영업매출2.2** (Sales 2.2): 연결 전용입니다. (Connected only)
 - 표추가결과** (Table Add Result): 24개의 행이 로드되었습니다. (24 rows loaded)
- 입고** (Import): 44개의 행이 로드되었습니다. (44 rows loaded)
- 출고** (Export): 37개의 행이 로드되었습니다. (37 rows loaded)



표 병합

● 파워 쿼리의 [병합]메뉴 실행

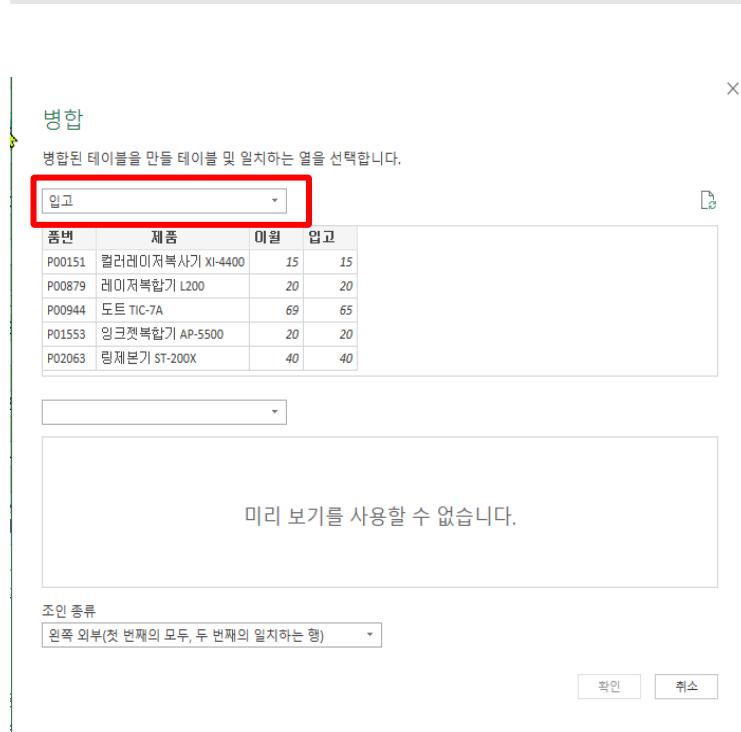




표 병합

● “입고” 쿼리에 “출고” 쿼리 [왼쪽외부] 조인으로 병합

萤幕截图展示了在Microsoft Access中通过“[왼쪽외부] 조인”(Left Outer Join)将“출고”(Sales)查询与“입고”(Purchase)查询合并的过程。

萤幕1: 显示了“출고”(Sales)表的结构，包含7行数据。右侧有一个带有“펼침버튼클릭”(Click the expand button)注释的箭头指向“Table”按钮。

	A ^B 품번	A ^B 제품	1 ² 미월	1 ² 입고	출고	1 ² 출고
1	P00151	컬러레이저복사기 XI...	15	15	15	15
2	P00879	레이저복합기 L200	20	20	20	20
3	P00944	도트 TIC-7A	69	65	65	65
4	P01553	잉크젯복합기 AP-5500	20	20	20	20
5	P02063	링제본기 ST-200X	40	40	40	40
6	P02126	와이어제본기 WC-51...	61	55	55	55
7	P02378	도트 TIC-1A	35	35	35	35

萤幕2: 显示了“출고”(Sales)表的筛选器界面。右侧有一个带有“펼침버튼클릭”(Click the expand button)注释的箭头指向“Table”按钮。

筛选器界面显示了以下设置：

- 筛选方式：◎ 확장 (Expansion)
- 选择项：
 - (모든 열 선택) (Select all columns)
 - 품번 (Item No.)
 - 제품 (Product)
 - 출고 (Sales)
- 复选框：□ 원래 열 이름을 접두사로 사용 (Use original column name as prefix)
- 按钮：확인 (OK) 和 취소 (Cancel)

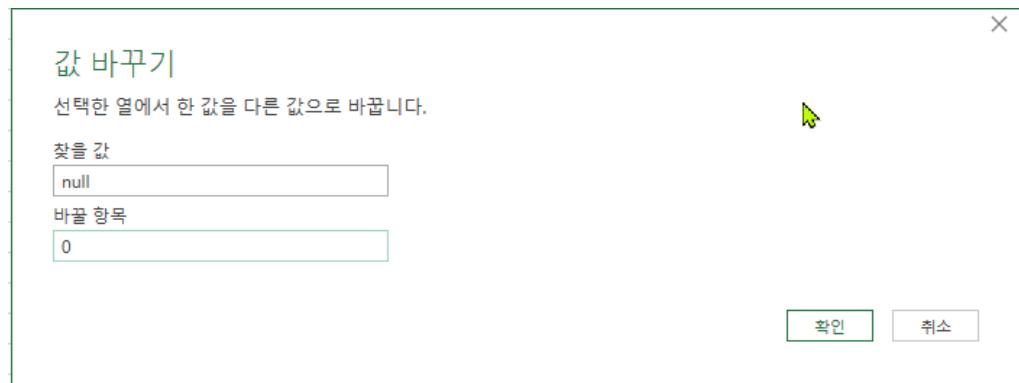
萤幕3: 显示了合并后的结果表，包含15行数据。右侧有一个带有“펼침버튼클릭”(Click the expand button)注释的箭头指向“Table”按钮。

	A ^B 품번	A ^B 제품	1 ² 미월	1 ² 입고	1 ² 출고	1 ² 출고.1
1	P00151	컬러레이저복사기 XI...	15	15	15	18
2	P03412	RF OA-300		35	35	29
3	P00879	레이저복합기 L200	20	20	20	11
4	P07941	RF OA-400		40	40	34
5	P00944	도트 TIC-7A	69	65	65	123
6	P02615	고급복사기 A4 1000매	25	25	25	26
7	P01553	잉크젯복합기 AP-5500	20	20	20	34
8	P02063	링제본기 ST-200X	40	40	40	20
9	P02378	도트 TIC-1A	35	35	35	56
10	P02126	와이어제본기 WC-51...	61	55	55	102
11	P02466	오피스 Z-03	25	25	25	null
12	P03984	레이저복합기 L500	45	45	45	9
13	P06000	레이저복합기 L800	68	60	60	114
14	P03280	무한레이저복합기 L...	15	15	15	null
15	P08111	레이저복합기 L950	10	10	10	11



데이터 전처리

- 병합된 “출고” 열 값 정리
 - null값 결측치 → 0으로 변환



- “이월”, “입고”, “출고”로 “잔고” 계산 열 추가
 - [잔고] = [이월] + [입고] - [출고]

파일 허 변포 열 추가 보기

예제의 사용자 사용자 지정 열 조건 열 인덱스 열 열 복제

제작자 지정 열 수식 호출

열 복제

열 병합

인덱스 열

열 추출

통계 표준 지수

구문 분석

텍스트에서

숫자에서

날짜 및 시간에서

날짜 시간 간간

날짜 및 시간에서

사용자 지정 열

다른 열에서 계산된 열을 추가합니다.

새 열 이름
잔고

사용자 지정 열 수식
= [이월]+[입고]-[출고]

사용 가능한 열

품번 제품 이월 입고 출고

<<삽입

파워 쿼리 수식에 대해 알아보기

✓ 구문 오류가 검색되지 않았습니다.

확인 취소

	A _C 품번	A _C 제품	1 ² 3 이월	1 ² 3 입고	1.2 출고	1.2 잔고
1	P00151	컬러레이저복사기 XI...	15	15	18	12
2	P03412	RF OA-300	35	35	29	41
3	P00879	레이저복합기 L200	20	20	11	29
4	PU7941	RF OA-400	40	40	34	46
5	P00944	도트 TIC-7A	69	65	123	11
6	P02615	고급복사기 A4 1000매	25	25	26	24
7	P01553	잉크젯복합기 AP-5500	20	20	34	6
8	P02063	링제본기 ST-200X	40	40	20	60
9	P02378	도트 TIC-1A	35	35	56	14
10	P02126	와이어제본기 WC-51...	61	55	102	14
11	P02466	오피스 Z-03	25	25	0	50



데이터 전처리

- 데이터 바이닝 (Data Binning: 구간화)
 - 제시된 기준에 따라 데이터 값을 특정한 bin(구간, group)으로 묶는 작업
 - 주어진 간격 bin에 속하는 데이터 값이 해당 간격을 대푯값(주로 중심값)으로 대체되는 방식으로 작업하기도 함
- 파워쿼리의 [열추가]탭 → [조건열]을 이용한 수치 데이터의 구간별 등급 표시 예시

The screenshot shows the Microsoft Power Query Editor interface. On the left, the ribbon has the 'File' tab selected, followed by 'Home', 'Transform', 'Add Column' (which is highlighted with a red box), and 'View'. Below the ribbon are buttons for 'Example Column', 'User Column', 'User Column Settings', 'Function Catalog', and 'General'. The main area shows the 'Add Column' dialog box for creating a 'Conditional Column'. The dialog has a title '조건 열 추가' and a subtitle '다른 열 또는 값에서 계산된 조건 열을 추가합니다.'. It includes a 'New column name' input field containing '잔고별등급' (highlighted with a red box). Below this is a table with three rows of conditions. The first row has '조건' set to '잔고', '运算자' to '보다 크거나 같음', '값' to 'ABC 123' and '75', and '결과' to 'ABC 123' and 'A'. The second row has '조건' set to '잔고', '运算자' to '보다 크거나 같음', '값' to 'ABC 123' and '50', and '결과' to 'ABC 123' and 'B'. The third row has '조건' set to '잔고', '运算자' to '보다 크거나 같음', '값' to 'ABC 123' and '25', and '결과' to 'ABC 123' and 'C'. At the bottom of the dialog, there is a 'More' button and two buttons on the right labeled '확인' (Confirm) and '취소' (Cancel). A dropdown menu for 'Other' is open at the bottom left, showing options like 'ABC 123', '값 입력' (Value Input), '열 선택' (Column Selection), and '매개 변수' (Parameter).



데이터 전처리 결과

- 파워쿼리를 이용한 전처리 결과
- 결측치 처리 / 계산열 추가 / 값 구간화 작업

A	B	C	D	E	F	G
1	품번 ▾ 제품	이월 ▾	입고 ▾	출고 ▾	잔고 ▾	잔고별등급 ▾
2	P00151 컬러레이저복사기 XI-4400	15	15	18	12	D
3	P03412 RF OA-300	35	35	29	41	C
4	P00879 레이저복합기 L200	20	20	11	29	C
5	P07941 RF OA-400	40	40	34	46	C
6	P00944 도트 TIC-7A	69	65	123	11	D
7	P02615 고급복사지A4 1000매	25	25	26	24	D
8	P01553 잉크젯복합기 AP-5500	20	20	34	6	D
9	P02063 링제본기 ST-200X	40	40	20	60	B
10	P02378 도트 TIC-1A	35	35	56	14	D
11	P02126 와이어제본기 WC-5100	61	55	102	14	D
12	P02466 오피스 Z-03	25	25	0	50	B
13	P03984 레이저복합기 L500	45	45	9	81	A
14	P06000 레이저복합기 L800	68	60	114	14	D
15	P03280 무한레이저복합기 L800C	15	15	0	30	C
16	P08111 레이저복합기 L950	10	10	11	9	D
17	P07535 링제본기 ST-100	15	15	10	20	D
18	P03418 문서세단기 SCUT-1000	20	20	5	35	C
19	P03440 와이어제본기 WC-5500	20	20	35	5	D
20	P08292 무한레이저복합기 L500C	15	15	15	15	D
21	P04174 잉크젯복합기 AP-4900	20	20	14	26	C
22	P04459 프리미엄복사지A4 2500매	45	45	21	69	B
23	P05409 바코드 BCD-200 Plus	60	60	107	13	D
24	P04785 열제본기 TB-8200	22	20	22	20	D
25	P05049 바코드 Z-750	61	55	106	10	D
26	P07875 복사지A4 5000매	10	10	8	12	D
27	P05057 잉크젯팩시밀리 FX-1000	53	45	83	15	D
28	P08393 복사지A4 500매	21	20	26	15	D
29	P05424 잉크젯팩시밀리 FX-2000	30	30	24	36	C
30	P05488 흑백레이저복사기 TIE-5000	65	65	30	100	A



데이터 정규화의 이해

- 관계형데이터베이스(RDBMS)에서 데이터를 설계하는 과정
- 데이터 테이블을 최대한 작고 잘 조직되도록 관계를 만들어 주는 과정이 데이터 정규화
- 정규화 되어 있지 않은 표 ← 데이터 중복이 존재

거래처명ID	거래처명	거래처위치	담당자	직통번호	물품분류	물품명	공급가	담당자2	직통번호2	물품분류2	물품명2	공급가2	담당자3	직통번호3	물품분류3	물품명3	공급가3
1	남양유업	서울 신촌	유재석	02-123-1234	병음료	코카콜라	\$100.00	유재석	02-123-1234	병음료	환타	\$120.00	노홍철	02-123-1235	캔음료	스프라이트	\$105.00
2	동서식품	서울 신림	박명수	02-123-1236	병음료	에비앙	\$80.00	박명수	02-123-1236	캔음료	페리에	\$90.00	이경규	02-123-1237	병음료	코카콜라	\$105.00
3	롯데푸드	서울 영등포	정준하	02-123-1238	케잌	딸기케잌	\$250.00	박명수	02-123-1236	과자	죠리퐁	\$50.00					
4	링그레	서울 논현	하동훈	02-123-1239	타르트	레몬타르트	\$210.00	이광수	02-123-1240	샌드위치용	상추	\$20.00					
5	오뚜기	서울 용산	정형돈	02-123-1241	샐러드용	양파	\$30.00	황광희	02-123-1242	샐러드용	비트	\$60.00					

담당자	직통번호	물품분류	물품명	공급가			
유재석	02-123-1234	병음료	코카콜라	\$100.00			
박명수	02-123-1236	병음료	에비앙	\$80.00			
정준하	02-123-1238						
담당자2	직통번호2	물품분류2	물품명2	공급가2			
유재석	02-123-1234	병음료	환타	\$120.00			
박명수	02-123-1236	캔음료	페리에	\$90.00			
박명수	02-123-1236	과자					
담당자3	직통번호3	물품분류3	물품명3	공급가3			
이광수	02-123-1240	샌드위	노홍철	02-123-1235	캔음료	스프라이트	\$105.00
황광희	02-123-1242	샐러드	이경규	02-123-1237	병음료	코카콜라	\$105.00



데이터 정규화의 이해

- 정규화 되어 있지 않은 표는 분석 기능을 적용하기가 불편할 수 있음
- 정규화 되어 있지 않은 상태의 문제점
 - 열의 순서가 변경되면 표를 인식하는데 문제가 발생
 - 데이터 정렬 문제 발생
 - 데이터를 추가 삽입 문제 발생

거래처명ID	거래처명	거래처위치	담당자	직통번호	물품분류	물품명	공급가	담당자2	직통번호2	물품분류2	물품명2	공급가2	담당자3	직통번호3	물품분류3	물품명3	공급가3
1 남양유업	서울 신촌	유재석	02-123-1234	병음료	코카콜라	\$100.00	유재석	02-123-1234	병음료	환타	\$120.00	노홍철	02-123-1235	캔음료	스프라이트	\$105.00	
2 동서식품	서울 신림	박명수	02-123-1236	병음료	에비앙	\$80.00	박명수	02-123-1236	캔음료	페리에	\$90.00	이경규	02-123-1237	병음료	코카콜라	\$105.00	
3 롯데푸드	서울 영등포	정준하	02-123-1238	케일	딸기케잌	\$250.00	박명수	02-123-1236	과자	죠리퐁	\$50.00						
4 빙그레	서울 논현	하동훈	02-123-1239	타르트	레몬타르트	\$210.00	이광수	02-123-1240	샌드위치용	상추	\$20.00						
5 오뚜기	서울 용산	정형돈	02-123-1241	샐러드용	양파	\$30.00	황광희	02-123-1242	샐러드용	비트	\$60.00						

- 1차 정규화 →
 - 중복 제거 필드 단위로 편집

거래처명	거래처위치	담당자	직통번호	물품분류	물품명	공급가
남양유업	서울 신촌	노홍철	02-123-1235	캔음료	스프라이트	\$ 105.00
남양유업	서울 신촌	유재석	02-123-1234	병음료	코카콜라	\$ 100.00
남양유업	서울 신촌	유재석	02-123-1234	캔음료	페리에	\$ 120.00
동서식품	서울 신림	박명수	02-123-1236	병음료	에비앙	\$ 80.00
동서식품	서울 신림	이경규	02-123-1237	병음료	코카콜라	\$ 105.00
동서식품	서울 신림	박명수	02-123-1236	캔음료	페리에	\$ 90.00
롯데푸드	서울 영등포	정준하	02-123-1238	케잌	딸기케잌	\$ 250.00
빙그레	서울 논현	하동훈	02-123-1239	타르트	레몬타르트	\$ 210.00
빙그레	서울 논현	이광수	02-123-1240	샌드위치용	상추	\$ 20.00
오뚜기	서울 용산	황광희	02-123-1242	샐러드용	비트	\$ 60.00
오뚜기	서울 용산	정형돈	02-123-1241	샐러드용	양파	\$ 30.00



데이터 표 정규화

- 1차 정규화 결과의 필드 중 서로 종속되어 있는 데이터 존재
- “물품명”과 “물품분류 ”는 서로 종속

거래처명	거래처위치	담당자	직통번호	물품분류	물품명	공급가
남양유업	서울 신촌	노홍철	02-123-1235	캔음료	스프라이트	\$ 105.00
남양유업	서울 신촌	유재석	02-123-1234	병음료	코카콜라	\$ 100.00
남양유업	서울 신촌	유재석	02-123-1234	캔음료	페리에	\$ 120.00
동서식품	서울 신림	박명수	02-123-1236	병음료	에비앙	\$ 80.00
동서식품	서울 신림	이경규	02-123-1237	병음료	코카콜라	\$ 105.00
동서식품	서울 신림	박명수	02-123-1236	캔음료	페리에	\$ 90.00
롯데푸드	서울 영등포	정준하	02-123-1238	케잌	딸기케잌	\$ 250.00
빙그레	서울 논현	하동훈	02-123-1239	타르트	레몬타르트	\$ 210.00
빙그레	서울 논현	이광수	02-123-1240	샌드위치용	상추	\$ 20.00
오뚜기	서울 용산	황광희	02-123-1242	샐러드용	비트	\$ 60.00
오뚜기	서울 용산	정형돈	02-123-1241	샐러드용	양파	\$ 30.00

- 2차, 3차 정규화 → 종속되는 필드는 표를 분리하고 ID로 키 필드를 부여함

거래처명	거래처위치	담당자	직통번호	물품명ID	공급가	물품분류ID	물품분류	물품명ID	물품분류ID	물품명
1 남양유업	서울 신촌	유재석	02-123-1234	1	\$ 100.00	1	병음료	1	코카콜라	
2 남양유업	서울 신촌	유재석	02-123-1234	2	\$ 120.00	2	캔음료	2	환타	
3 남양유업	서울 신촌	노홍철	02-123-1235	3	\$ 105.00	3	케잌	3	스프라이트	
4 동서식품	서울 신림	박명수	02-123-1236	4	\$ 80.00	4	타르트	4	에비앙	
5 동서식품	서울 신림	박명수	02-123-1236	5	\$ 90.00	5	샌드위치용	5	페리에	
6 동서식품	서울 신림	이경규	02-123-1237	6	\$ 105.00	6	샐러드용	6	딸기케잌	
7 롯데푸드	서울 영등포	정준하	02-123-1238	7	\$ 250.00	7		7	레몬타르트	
8 빙그레	서울 논현	하동훈	02-123-1239	8	\$ 210.00	8		8	상추	
9 빙그레	서울 논현	이광수	02-123-1240	9	\$ 20.00	9		9	양파	
10 오뚜기	서울 용산	정형돈	02-123-1241	10	\$ 30.00	10		10	비트	
11 오뚜기	서울 용산	황광희	02-123-1242							



데이터 표 정규화 예시

● 병합 셀 제거와 데이터 필드 정리

- 예제) 결제 방법의 표시를 한 열에 합치기
- IF(), IFERROR() 함수 활용

1 거래번호	고객	제품	단가	수량	판매	결제여부	결제방법	
							카드	현금
3 N-0705	S&C무역	컬러레이저복사기 XJ-3200	1,176,000	3	2,998,800	입금	○	
4 N-0706	드림씨퍼드	프리미엄복사지A4 2500매	17,800	9	160,200	입금	○	
5 N-0707	자이언트무역	고급복사지A4 500매	3,500	2	7,000	입금	○	
6 N-0708	진왕통상	잉크젯복합기 AP-3300	79,800	1	79,800	입금	○	
7 N-0709	삼영트레이드	잉크젯복합기 AP-3200	79,500	2	159,000	입금	○	
8 N-0710	자이언트무역	정제본기 ST-100	127,800	4	511,200	입금		○
9 N-0711	동남무역	흑백레이저복사기 TLE-5000	597,900	3	1,794,015	입금	○	
10 N-0712	한남상사	프리미엄복사지A4 2500매	16,800	5	84,000	입금	○	
11 N-0713	금화트레이드	정제본기 ST-100	161,900	9	1,384,245	미입금		
12 N-0714	칠성무역	비코드 BCD-200 Plus	96,900	6	581,400	미입금		
13 N-0715	뉴넷트워크	복사지A4 1000매	5,400	3	16,200			
14 N-0716	신성백화점	프리미엄복사지A4 2500매	20,700	2	41,400			
15 N-0717	시선무역	복사지A4 500매	3,200	8	25,600			
16 N-0718	네트워크통상	RF OA-300	48,400	10	484,000			



1 거래번호	고객	제품	단가	수량	판매	결제여부	결제방법		J	K
							카드	현금		
3 N-0705	S&C무역	컬러레이저복사기 XJ-3200	1,176,000	3	2,998,800	입금	○		카드	
4 N-0706	드림씨퍼드	프리미엄복사지A4 2500매	17,800	9	160,200	입금	○		카드	
5 N-0707	자이언트무역	고급복사지A4 500매	3,500	2	7,000	입금	○		카드	
6 N-0708	진왕통상	잉크젯복합기 AP-3300	79,800	1	79,800	입금	○		카드	
7 N-0709	삼영트레이드	잉크젯복합기 AP-3200	79,500	2	159,000	입금	○		카드	
8 N-0710	자이언트무역	정제본기 ST-100	127,800	4	511,200	입금		○	현금	
9 N-0711	동남무역	흑백레이저복사기 TLE-5000	597,900	3	1,794,015	입금	○		카드	
10 N-0712	한남상사	프리미엄복사지A4 2500매	16,800	5	84,000	입금	○		카드	
11 N-0713	금화트레이드	정제본기 ST-100	161,900	9	1,384,245	미입금				
12 N-0714	칠성무역	비코드 BCD-200 Plus	96,900	6	581,400	미입금				
13 N-0715	뉴넷트워크	복사지A4 1000매	5,400	3	16,200	입금	○		카드	
14 N-0716	신성백화점	프리미엄복사지A4 2500매	20,700	2	41,400	미입금				
15 N-0717	시선무역	복사지A4 500매	3,200	8	25,600	입금	○		현금	
16 N-0718	네트워크통상	RF OA-300	48,400	10	484,000	입금	○		카드	



데이터 표 정규화 예시

- 데이터 종속이 발생 → 테이블 분리가 필요

A 1 2	B 고객	C 제품	D 단가	E 수량	F 판매	G 결제여부	결제방법	
							H 카드	I 현금
N-0705	S&C무역 (주)	컬러레이저복사기 XI-3200	1,176,000	3	2,998,000	입금	o	
N-0706	드림씨푸드 (주)	프리미엄복사지 A4 2500매	17,800	9	160,200	입금	o	
N-0707	자이언트무역 (주)	고급복사지 A4 500매	3,500	2	7,000	입금	o	
N-0708	진왕통상 (주)	잉크젯복합기 AP-3300	79,800	1	79,800	입금	o	
N-0709	삼양트레이드 (주)	잉크젯복합기 AP-3200	79,500	2	159,000	입금	100,000	59,000
N-0710	자이언트무역 (주)	링제본기 ST-100	127,800	4	511,200	입금		o
N-0711	동남무역 (주)	흑백레이저복사기 TLE-5000	597,900	3	1,704,015	입금	1,500,000	204,015
N-0712	한남상사 (주)	프리미엄복사지 A4 2500매	16,800	5	84,000	입금	o	
N-0713	금화트레이드 (주)	링제본기 ST-100	161,900	9	1,384,245	미입금		
N-0714	칠성무역 (주)	바코드 BCD-200 Plus	96,900	6	581,400	미입금		
N-0715	뉴럴네트워크 (주)	복사지 A4 1000매	5,400	3	16,200	입금	o	
N-0716	신성백화점 (주)	프리미엄복사지 A4 2500매	20,700	2	41,400	미입금		
N-0717	사선무역 (주)	복사지 A4 500매	3,200	8	25,600	입금		o
N-0718	네트워크통상 (주)	RF OA-300	48,400	10	484,000	입금	o	



A 1 2	B 고객	C 제품	D 단가	E 수량	F 판매	G 결제여부	H 결제방법
N-0705	S&C무역 (주)	컬러레이저복사기 XI-3200	1,176,000	3	2,998,000	입금	카드
N-0706	드림씨푸드 (주)	프리미엄복사지 A4 2500매	17,800	9	160,200	입금	카드
N-0707	자이언트무역 (주)	고급복사지 A4 500매	3,500	2	7,000	입금	카드
N-0708	진왕통상 (주)	잉크젯복합기 AP-3300	79,800	1	79,800	입금	카드
N-0709	삼양트레이드 (주)	잉크젯복합기 AP-3200	79,500	2	159,000	입금	카드(100,000), 현금(59,000)
N-0710	자이언트무역 (주)	링제본기 ST-100	127,800	4	511,200	입금	현금
N-0711	동남무역 (주)	흑백레이저복사기 TLE-5000	597,900	3	1,704,015	입금	카드(1,500,000), 현금(204,015)
N-0712	한남상사 (주)	프리미엄복사지 A4 2500매	16,800	5	84,000	입금	카드
N-0713	금화트레이드 (주)	링제본기 ST-100	161,900	9	1,384,245	미입금	
N-0714	칠성무역 (주)	바코드 BCD-200 Plus	96,900	6	581,400	미입금	
N-0715	뉴럴네트워크 (주)	복사지 A4 1000매	5,400	3	16,200	입금	카드
N-0716	신성백화점 (주)	프리미엄복사지 A4 2500매	20,700	2	41,400	미입금	
N-0717	사선무역 (주)	복사지 A4 500매	3,200	8	25,600	입금	현금
N-0718	네트워크통상 (주)	RF OA-300	48,400	10	484,000	입금	카드



데이터 표 정규화 예시

- 데이터 종속이 발생 → 테이블 분리가 필요
- 두 개의 표로 분리
- ‘거래번호’ 필드가 두 표의 관계 생성을 위한 키 필드

A	B	C	D	E	F	G	
1	거래번호	고객	제품	단가	수량	판매	결제여부
2	N-0705	S&C무역 (주)	컬러레이저복사기 XI-3200	1,176,000	3	2,998,800	입금
3	N-0706	드림씨푸드 (주)	프리미엄복사지 A4 2500매	17,800	9	160,200	입금
4	N-0707	자이언트무역 (주)	고급복사지 A4 500매	3,500	2	7,000	입금
5	N-0708	진왕통상 (주)	잉크젯복합기 AP-3300	79,800	1	79,800	입금
6	N-0709	삼양드레이드 (주)	잉크젯복합기 AP-3200	79,500	2	159,000	입금
7	N-0710	자이언트무역 (주)	링제본기 ST-100	127,800	4	511,200	입금
8	N-0711	동남무역 (주)	흑백레이저복사기 TLE-5000	597,900	3	1,794,015	입금
9	N-0712	한남상사 (주)	프리미엄복사지 A4 2500매	16,800	5	84,000	입금
10	N-0713	금화트레이드 (주)	링제본기 ST-100	161,900	9	1,384,245	미입금
11	N-0714	칠성무역 (주)	바코드 8CD-200 Plus	96,900	6	581,400	미입금
12	N-0715	뉴탈네트워크 (주)	복사지 A4 1000매	5,400	3	16,200	입금
13	N-0716	신성백화점 (주)	프리미엄복사지 A4 2500매	20,700	2	41,400	미입금
14	N-0717	사선무역 (주)	복사지 A4 500매	3,200	8	25,600	입금
15	N-0718	네트워크통상 (주)	RF OA-300	48,400	10	484,000	입금

A	B	C	
1	거래번호	결제방법	입금액
2	N-0705	카드	2,998,800
3	N-0706	카드	160,200
4	N-0707	카드	7,000
5	N-0708	카드	79,800
6	N-0709	카드	100,000
7	N-0709	현금	59,000
8	N-0710	현금	511,200
9	N-0711	카드	1,500,000
10	N-0711	현금	204,015
11	N-0712	카드	84,000
12	N-0715	카드	16,200
13	N-0717	현금	25,600
14	N-0718	카드	484,000



데이터 표의 관계 생성

- 표의 정규화로 분리된 표의 관계 생성
- 관계(Relationship)의 이해
- 표를 수평적으로 연결하는 기술

Two tables are shown side-by-side. The first table has columns '사번' (번호), '이름' (Name), and '직위' (Position). The second table has columns '사번' (번호) and '부서' (Department). A blue dashed line connects the '사번' column of the first table to the '사번' column of the second table. A blue arrow labeled '연결' (Join) points from the first table to the second.

사번	이름	직위
1	A	부장
2	B	과장
3	C	과장
4	D	대리

사번	부서
1	가
2	가
3	나
4	나

The resulting joined table has four columns: '사번' (번호), '부서' (Department), '이름' (Name), and '직위' (Position). It contains four rows, each combining a department from the first table with a name and position from the second table.

사번	부서	이름	직위
1	가	A	부장
2	가	B	과장
3	나	C	과장
4	나	D	대리

이 표가 실제로 생성되는 것은 아니지만, 관계로 표를 연결하면 이런 표를 사용하는 것처럼 작업할 수 있습니다.

관계	설명
1:1	연결할 표의 키 열에 값이 하나씩 있는 경우. 즉, 키 열에 중복된 값이 없고 고유한 값만 존재
1:N	연결할 표에는 키 열에 고유한 값만 존재하고 다른 표에는 키 열에 중복된 값이 존재하는 경우. 즉, 키 값 하나가 다른 표에서 여러 개 사용되고 있다는 의미

This diagram shows two tables being joined horizontally. The left table has columns '사번' (번호), '이름' (Name), and '직위' (Position). The right table has columns '사번' (번호) and '부서' (Department). A blue dashed line connects the '사번' column of the left table to the '사번' column of the right table. A blue arrow labeled '연결' (Join) points from the left table to the right table.

Below the tables, a note states: "원쪽 표를 기준으로 연결한 경우, 원쪽 표의 5번 사번은 오른쪽 표에는 없지만 부서만 빙 셀인 것처럼 사용할 수 있습니다. 하지만 오른쪽 표에만 있는 3번 사번은 연결된 표에서는 사용할 수 없습니다."

사번	이름	직위
1	A	부장
2	B	과장
4	C	과장
5	D	대리

사번	부서
1	가
2	가
3	나
4	나

사번	부서	이름	직위
1	가	A	부장
2	가	B	과장
4	나	C	과장
5	나	D	대리

사번	부서
1	가
1	나
2	나
3	다
3	가
4	라

사번	부서	이름	직위
1	가	A	부장
1	나	A	부장
2	나	B	과장
3	다	C	과장
3	가	C	과장
4	라	D	대리

1 : 다 관계에서는 행이 더 많은 표에 맞춰 구성된 표에 데이터를 연결해 사용할 수 있습니다.



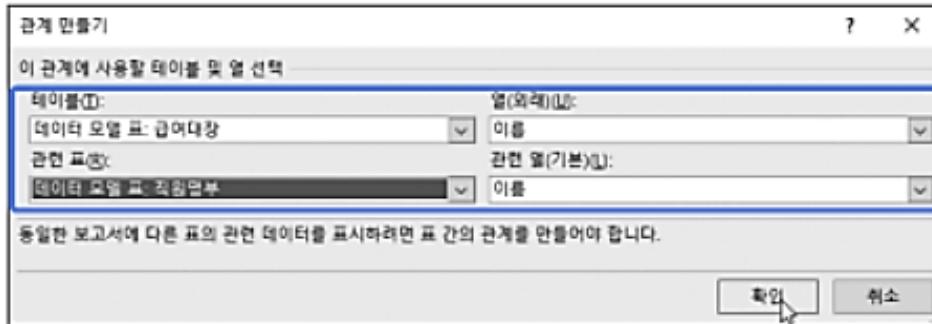
데이터 표의 관계 생성

- 1:1 관계 설정
 - [이름]열이 키
 - 키값은 중복을 허용하지 않는다.
- [데이터]탭-[데이터 도구]그룹-[관계] 명령

급여대장				
이름	기본급	급여총액	공제총액	월급액
구현상	5,100,000	6,444,000	497,650	5,946,350
청다정	3,800,000	4,438,000	322,800	4,115,200
유예찬	3,600,000	3,950,000	295,900	3,654,100
전보람	3,600,000	3,950,000	295,900	3,654,100
임선정	3,200,000	3,566,000	242,100	3,323,900
재우희	3,200,000	3,350,000	242,100	3,107,900
홍다림	3,100,000	3,335,000	226,650	3,106,350
남영재	2,800,000	2,900,000	188,300	2,711,700
김민경	2,500,000	2,668,000	147,950	2,520,050
서현디	2,500,000	2,668,000	147,950	2,520,050
황용기	2,600,000	2,680,000	161,400	2,518,600
심여로	2,600,000	2,680,000	161,400	2,518,600
손예지	2,500,000	2,580,000	147,950	2,432,050
화연우	2,400,000	2,530,000	134,500	2,395,500
김영애	2,400,000	2,480,000	134,500	2,345,500
김영재	2,400,000	2,400,000	134,500	2,345,500
배영자	2,400,000	2,400,000	134,500	2,345,500
김상아	2,400,000	2,400,000	134,500	2,345,500
임선정	2,400,000	2,400,000	134,500	2,345,500
남영재	2,400,000	2,400,000	134,500	2,345,500
배영한	2,400,000	2,400,000	134,500	2,345,500
김은희	2,400,000	2,400,000	134,500	2,345,500

직원명부				
이름	부서	직위	오봉	
구현상	영업부	부장	2호봉	
전보람	영업부	과장	1호봉	
홍다림	영업부	대리	1호봉	
김민경	영업부	사원	3호봉	
화연우	영업부	사원	1호봉	
김영애	영업부	사원	2호봉	
서현디	영업부	사원	3호봉	
황용기	영업부	사원	2호봉	
심여로	영업부	사원	2호봉	
손예지	영업부	과장	2호봉	
화연우	영업부	대리	2호봉	
임선정	영업부	사원	3호봉	
김영재	영업부	사원	1호봉	
배영자	영업부	사원	3호봉	
김상아	영업부	사원	2호봉	
임선정	영업부	주임	2호봉	
배영한	영업부	사원	3호봉	
김은희	영업부	사원	2호봉	

A	B	C	D	E	F	G	H	I	J	K	L	M
이름	부서	직위	오봉	이름	부서	직위	오봉	이름	부서	직위	오봉	이름
구현상	영업부	부장	2호봉	구현상	영업부	부장	2호봉	전보람	영업부	과장	1호봉	전보람
청다정	영업부	과장	1호봉	청다정	영업부	과장	1호봉	홍다림	영업부	대리	1호봉	홍다림
유예찬	영업부	대리	1호봉	유예찬	영업부	대리	1호봉	김민경	영업부	사원	3호봉	김민경
전보람	영업부	사원	3호봉	전보람	영업부	사원	3호봉	화연우	영업부	사원	1호봉	화연우
임선정	영업부	사원	3호봉	임선정	영업부	사원	3호봉	김영애	영업부	사원	2호봉	김영애
심여로	영업부	사원	3호봉	심여로	영업부	사원	3호봉	서현디	영업부	사원	3호봉	서현디
손예지	영업부	과장	2호봉	손예지	영업부	과장	2호봉	황용기	영업부	사원	2호봉	황용기
화연우	영업부	대리	2호봉	화연우	영업부	대리	2호봉	임선정	영업부	사원	3호봉	임선정
임선정	영업부	사원	3호봉	임선정	영업부	사원	3호봉	김영재	영업부	사원	1호봉	김영재
배영자	영업부	사원	3호봉	배영자	영업부	사원	3호봉	화연우	영업부	주임	2호봉	화연우
김상아	영업부	사원	2호봉	김상아	영업부	사원	2호봉	임선정	영업부	사원	3호봉	임선정
임선정	영업부	주임	2호봉	임선정	영업부	주임	2호봉	배영한	영업부	사원	3호봉	배영한
배영한	영업부	사원	3호봉	배영한	영업부	사원	3호봉	김은희	영업부	사원	2호봉	김은희
김은희	영업부	사원	2호봉	김은희	영업부	사원	2호봉					



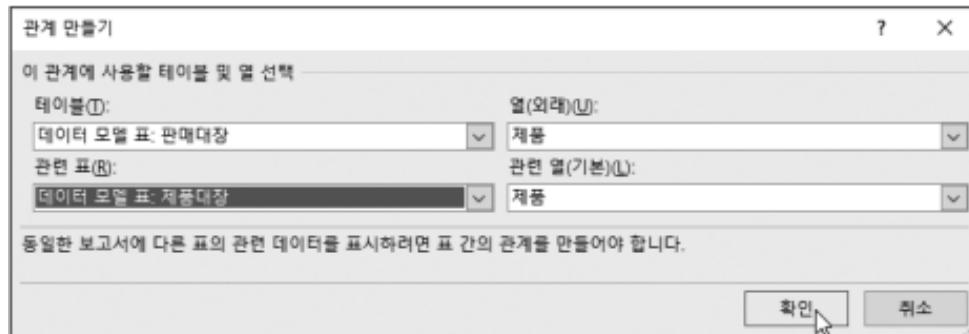


데이터 표의 관계 생성

● 1:N 관계 설정

- 한 테이블은 고유한 키 값을 갖고,
다른 테이블은 중복된 키 값을 가짐

● [데이터]탭-[데이터 도구]그룹-[관계] 명령

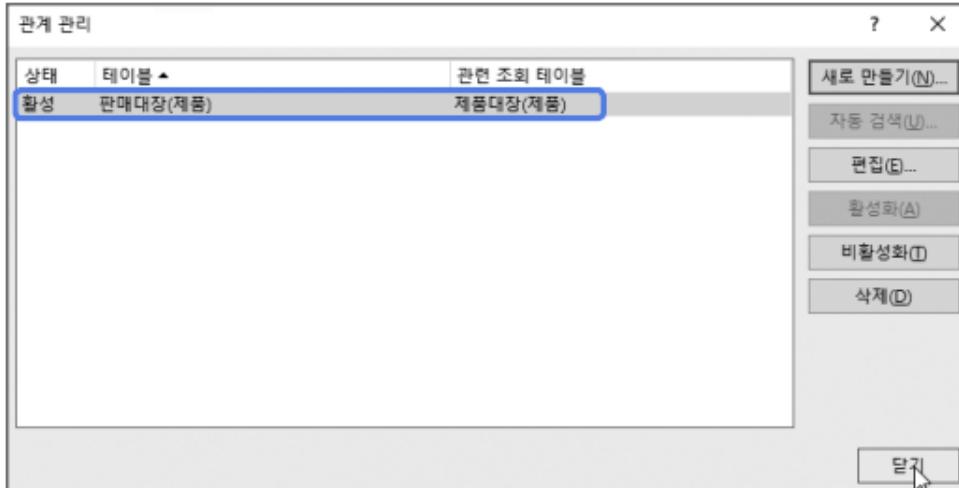


테이블	열(외래)	비교
판매대장	제품	피벗 테이블 보고서의 기본이 되는 표

관련 표	관련 열	비교
제품대장	제품	기본 표에 추가할 데이터를 갖고 있는 표

판매대장

A	B	C	D	E	F	
1						
2						
3						
4	거래번호	고객	제품	단가	수량	전액
5	N-0705	SBC 푸 앤 드	바코드 Z-350	48,300	3	144,900
6	N-0706	도광복부드	바코드 BCD-100 Plus	86,500	7	605,500
7	N-0707	자이언트푸드	바코드 Z-350	46,300	7	324,100
8	N-0710	자이언트무역	바코드 BCD-100 Plus	104,500	8	836,000
9	N-0712	한남살사	김제분기 ST-100	127,800	4	511,200
10	N-0713	금화트레이드	김제분기 ST-100	90,300	7	632,100
11	N-0714	칠성후대	바코드 BCD-200 Plus	161,900	9	1,457,100
12	N-0715	동원상사	김제분기 TB-8200	177,000	10	1,869,100
13	N-0720	뉴밀네트워크	바코드 Z-350	59,800	2	119,600
14	N-0722	필드유통	바코드 BCD-100 Plus	99,800	4	399,200
15	N-0724	진주	바코드 BCD-200 Plus	113,700	8	909,600
16	N-0726	신풀유통	바코드 BCD-100 Plus	103,600	9	932,400
17	N-0730	스타벅커피	바코드 Z-350	47,300	3	141,900
18	N-0734	증평통상	바코드 BCD-200 Plus	111,800	3	335,400
19	N-0735	반디살사	김제분기 ST-100	164,700	8	1,251,720
20	N-0738	나래백화점	김제분기 ST-100	133,500	8	1,064,600
21	N-0741	스타벅커피	바코드 Z-350	54,800	6	328,800
22	N-0742	길가운교	바코드 BCD-200 Plus	112,700	10	1,070,600





파워 피벗의 이해

- 관계형 데이터 표를 기반으로 피벗 테이블의 기능 적용 확장을 위한 추가 기능
- 표를 데이터모델에 추가한 후 파워 피벗 실행
- 파워 피벗 추가 기능 표시

The screenshot shows two windows from Microsoft Excel:

- Microsoft Office 추가 기능을 보고 관리합니다.** (Manage Microsoft Office add-ins): This window lists various add-ins like LogiOptionsExcelAddin, Microsoft Power Pivot for Excel, MySQL For Excel, etc. A red box highlights the "추가 기능" (Add-in) category in the left sidebar. A red arrow points from the "Power Pivot" entry in this list to the "Power Pivot for Excel" entry in the next window.
- COM 추가 기능** (Add-in for COM): This window shows a list of available add-ins. The "Microsoft Power Pivot for Excel" checkbox is checked and highlighted with a blue selection bar. A red dashed box encloses this list. A red arrow points from the "Power Pivot" entry in the previous window to this one.

At the bottom of the screen, the Excel ribbon is visible, showing the "파일" (File) tab selected. The "Power Pivot" tab is also visible in the ribbon, indicating it is currently active.



파워 피벗 화면

- [파워 피벗]메뉴 탭 → [데이터모델]메뉴

The screenshot shows the Microsoft Power Pivot ribbon interface. The 'Data Model' tab is highlighted with a red box. The ribbon tabs shown are 파일 (File), 툴 (Tools), 디자인 (Design) (selected), and 고급 (Advanced). The main menu area includes options like 불여넣기 (Import), 데이터베이스 서비스 (Database Services), 데이터 원본 연결 (Data Source Connection), 새롭고침 (New Refresh), 피벗 테이블 (Pivot Table), and various data format and filter settings. On the right side, there's a 'Data Model' ribbon tab with options like 데이터 보기 (Data View), 다이어그램 뷰 (Diagram View), and KPI 만들기 (Create KPI). The bottom navigation bar also has a red box around the 'tbl구매내역' tab, which is currently selected. The main content area displays a table of purchase data with columns for 날짜 (Date), 제품 (Product), 구매가격 (Purchase Price), 구매수량 (Purchase Quantity), 구매금액 (Purchase Amount), and 열 추가 (Add Column). The table contains 12 rows of data from 2020-01 to 2020-12.

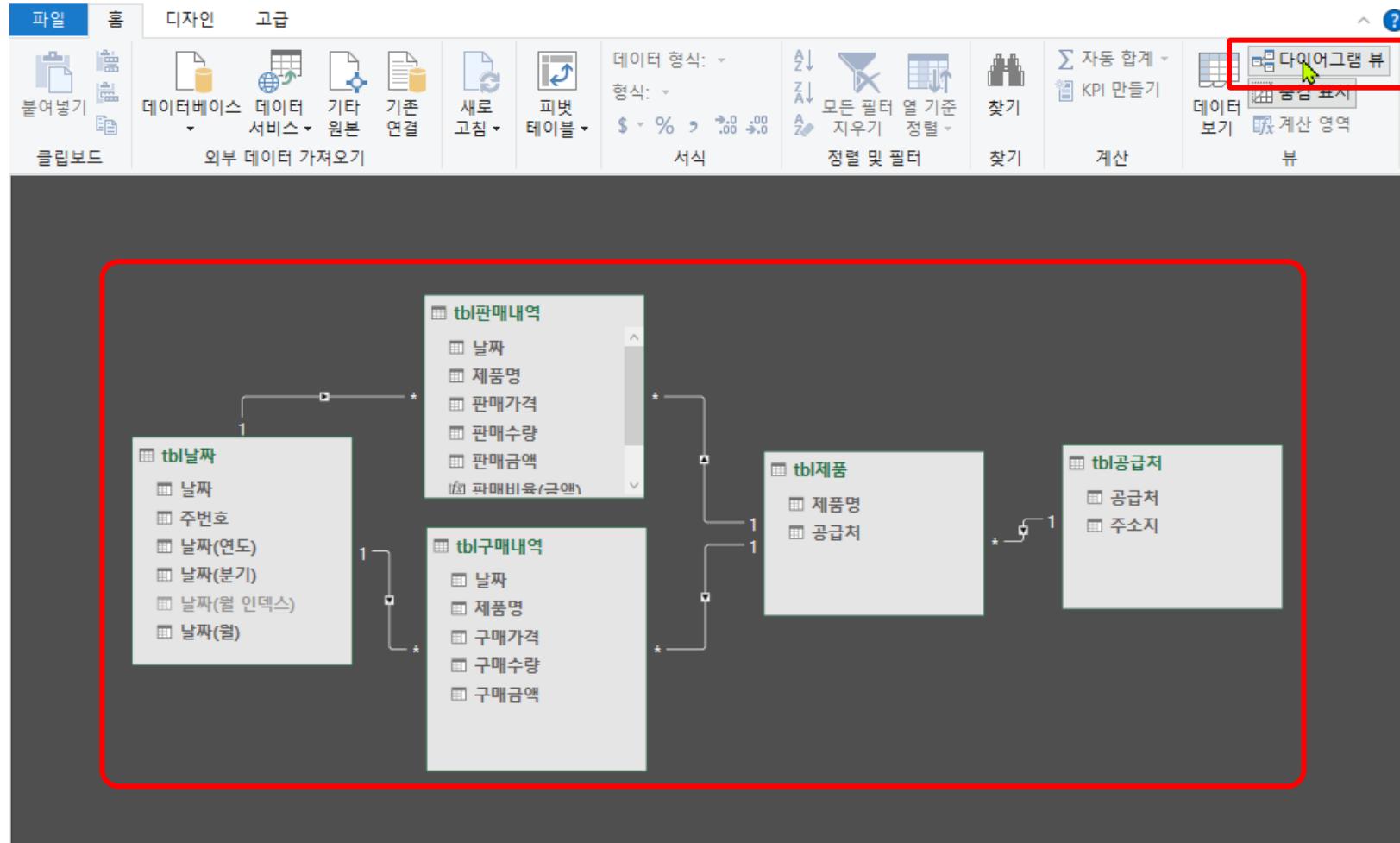
	날짜	제품	구매가격	구매수량	구매금액	열 추가
1	2020-01-01	75인치 UHD TV	1471000	120	176,520,000	
2	2020-01-02	55인치 UHD TV	830000	110	91,300,000	
3	2020-01-03	에스프레소 머신	450000	100	45,000,000	
4	2020-01-04	에스프레소 머신	365000	50	18,250,000	
5	2020-01-05	복합식 가전제품	135000	100	13,500,000	
6	2020-01-06	75인치 UHD TV	1803000	100	180,300,000	
7	2020-01-07	와플 메이커	30000	140	4,200,000	
8	2020-01-08	와플 메이커	26000	135	3,510,000	
9	2020-01-09	55인치 UHD TV	784000	60	47,040,000	
10	2020-01-10	초음파 분말기	41000	90	3,690,000	
11	2020-01-11	초음파 분말기	32000	140	4,480,000	
12	2020-01-12	50인치 UHD TV	652000	135	88,020,000	

tbl구매내역 | tbl판매내역 | tbl제품 | tbl공급처 | tbl날짜 |



파워 피벗 표

- [다이어그램뷰]메뉴 → 표의 관계 표시

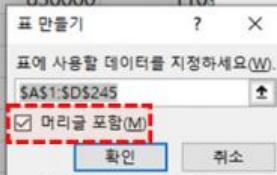




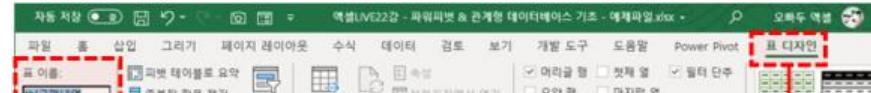
데이터 모델링

● 데이터 표 생성

A	B	C	D	E
1 날짜	제품명	구매가격	구매수량	
2 01월 01일	75인치 UHD 4K HDR	1471000	120	
3 01월 01일	55인치 UHD 4K TV	830000	110	
4 01월 02일	에스프레소 머신			
5 01월 03일	에스프레소 머신			
6 01월 04일	복합식 가습기			
7 01월 04일	75인치 UHD 4K HDR			
8 01월 04일	와플 메이커 3 in 1	26000	135	
9 01월 05일	와플 메이커 3 in 1			
10 01월 05일	55인치 UHD 4K TV	784000	60	
11 01월 05일	초음파 분리세척 가습기	41000	90	
12 01월 05일	초음파 분리세척 가습기	32000	140	
13 01월 06일	50인치 UHD TV HDR	652000	135	



● 표 이름 정하기



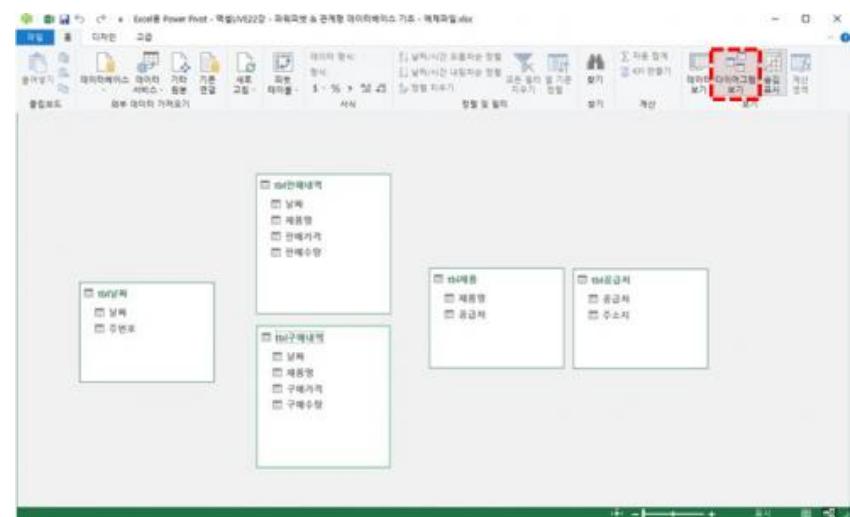
The screenshot shows the Microsoft Excel ribbon with the 'Tables' tab selected. A red box highlights the 'New Table' button in the 'Tables' group. The main area displays a table with columns labeled A through G, containing data from rows 1 to 11.

A	B	C	D	E	F	G
1 날짜	제품명	구매가격	구매수량			
2 01월 01일	75인치 UHD 4K HDR	1471000	120			
3 01월 01일	55인치 UHD 4K TV	830000	110			
4 01월 02일	에스프레소 머신	450000	100			
5 01월 03일	에스프레소 머신	365000	50			
6 01월 04일	복합식 가습기	135000	100			
7 01월 04일	75인치 UHD 4K HDR	1803000	100			
8 01월 04일	와플 메이커 3 in 1	30000	140			
9 01월 05일	와플 메이커 3 in 1	26000	135			
10 01월 05일	55인치 UHD 4K TV	784000	60			
11 01월 05일	초음파 분리세척 가습기	41000	90			

● 표의 관계 생성 → 데이터모델에 추가

● 데이터모델에 추가 → 다이어그램에서 필드 연결

- 1. tbl날짜[날짜] (1) - tbl판매내역[날짜] (*)
- 2. tbl날짜[날짜] (1) - tbl구매내역[날짜] (*)
- 3. tbl제품[제품명] (1) - tbl판매내역[제품명] (*)
- 4. tbl제품[제품명] (1) - tbl구매내역[제품명] (*)
- 5. tbl공급처[공급처] (1) - tbl제품[공급처] (*)



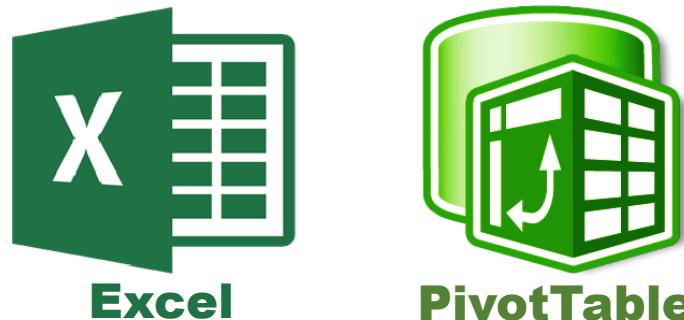
피벗 테이블





피벗테이블

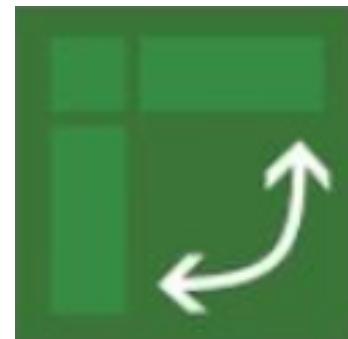
- 데이터 요약 분석 도구
- 데이터 비교, 패턴 및 추세를 파악하는 데 사용할 수 있음
- ‘피벗’은 ‘회전 축’이라는 의미를 가지며 요약을 위한 기준 데이터 필드를 축으로 설정하고 설정된 기준으로 요약 계산하여 표를 작성해 줌
- 합계, 평균, 기타 통계 계산으로 요약하며 유용한 정보에 집중할 수 있도록 계산 결과를 정렬 또는 재정렬(피벗)함





기준열과 계산열

- 데이터 테이블의 열(Column: 필드)의 파악
- 범주형 데이터 필드를 기준으로 요약 집계 계산 실행
- 피벗테이블의 페이지, 행, 열 영역에 기준이 되는 열을 표시
- 요약 계산할 열을 계산 영역에 표시
- 기준 필드의 위치를 자유롭게 위치 변경 회전하여 피벗테이블 형태를 변경가능





피벗테이블 작성

● 피벗테이블 작성을 위한 데이터 확인

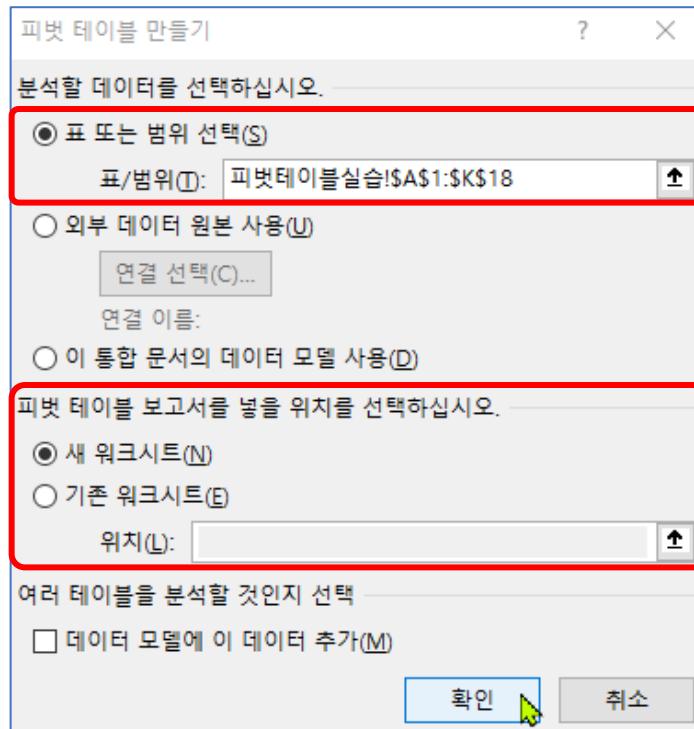
	A 소속학과	B 학년	C 학번	D 성명	E 중간	F 기말	G 과제	H 출석	I 태도	J 합계	K 등급
1	컴퓨터	2	2021512512	이○환	52.1	46.8	44.4	83.3	50	276.7	
2	의류디자인	3	2022430131	김○태	55.9	57.3	100.0	95.6	100	408.8	
3	경영정보	3	2021212934	배○이	59.0	58.2	100.0	100.0	100	417.2	
4	스포츠건강과학	4	2018410341	박○연	19.9	32.3	33.3	74.4	50	210.0	
5	컴퓨터	3	2022512705	박○현	69.0	76.8	100.0	100.0	100	445.8	
6	경영정보	3	2022212312	채○정	51.7	52.3	100.0	94.4	100	398.4	
7	공공인재	4	2018123958	김○훈	93.9	84.6	100.0	100.0	100	478.4	Good
8	식품영양	3	2020321368	김○주	86.6	76.8	100.0	97.8	100	461.2	Good
9	식품영양	2	2021321140	정○환	54.8	60.9	100.0	90.0	100	405.7	
10	공공인재	3	2020123550	박○규	85.1	75.0	100.0	98.9	100	459.0	Good
11	산업디자인	1	2020524713	강○연	55.6	19.6	88.9	96.7	100	360.7	
12	스포츠건강과학	3	2018410823	임○규	89.3	70.9	100.0	100.0	100	460.2	Good
13	산업디자인	1	2020524642	이○형	68.6	75.5	88.9	86.7	100	419.6	
14	도시환경	1	2023560029	권○원	55.9	41.4	100.0	100.0	100	397.3	
15	산업디자인	4	2018524437	주○진	52.1	23.6	66.7	91.1	100	333.5	
16	도시환경	3	2020560464	박○호	71.3	67.7	100.0	98.9	100	437.9	
17	컴퓨터	4	2022112233	유현주	80.0	80.0	100.0	95.0	100	455.0	Good

- 범주형 필드 → 기준 필드로 지정 가능
- 개별 데이터 필드 → 계산 필드로 지정 가능



피벗테이블 작성

- [삽입]메뉴탭 → [피벗테이블]메뉴



→ 데이터영역 지정

→ 피벗테이블 작성 위치

- 데이터 시트 왼쪽에 새 워크시트가 생성되고 피벗테이블 영역이 설정됨



피벗테이블 작성

● 피벗테이블영역과 필드목록창

The screenshot shows the Microsoft Excel ribbon with the '분석' (Analysis) tab selected. On the left, a PivotTable is displayed with the message: '보고서를 작성하려면 피벗 테이블 필드 목록에서 필드를 선택하세요.' (Select fields from the PivotTable field list to create a report). A red box highlights this area with the label '← 피벗테이블영역'. On the right, the 'PivotTable Fields' pane is open, showing a list of fields: 소속학과, 학년, 학번, 성명, 중간, 기말, 과제, 출석, 태도, 합계, 등급. Below this is a 'Filter' section with '행' (Row) and '열' (Column) dropdowns. A red box highlights this pane with the label '필드목록→'. At the top right of the ribbon, there is a '필드 목록' (Field List) button, which is also highlighted with a red box and labeled '필드목록창표시→'.



피벗테이블 작성

- 행(기준필드), 값(계산필드) 지정
- 값 계산필드 요약 계산
 - 수치 데이터의 기본 계산 함수는 합계
 - 문자 데이터의 기본 계산 함수는 개수

피벗 테이블 필드

보고서에 추가할 필드 선택:

검색

행 레이블

개수 : 성명 합계 : 기말

	A	B	C
1			
2			
3	경영정보	2	110.45
4	공공인재	2	159.55
5	도시환경	2	109.09
6	산업디자인	3	118.64
7	스포츠건강과학	2	103.18
8	식품영양	2	137.73
9	의류디자인	1	57.27
10	컴퓨터	3	203.64
11	총합계	17	999.55

피벗테이블

행기준 →

요약계산

행

소속학과

값

개수 : 성명

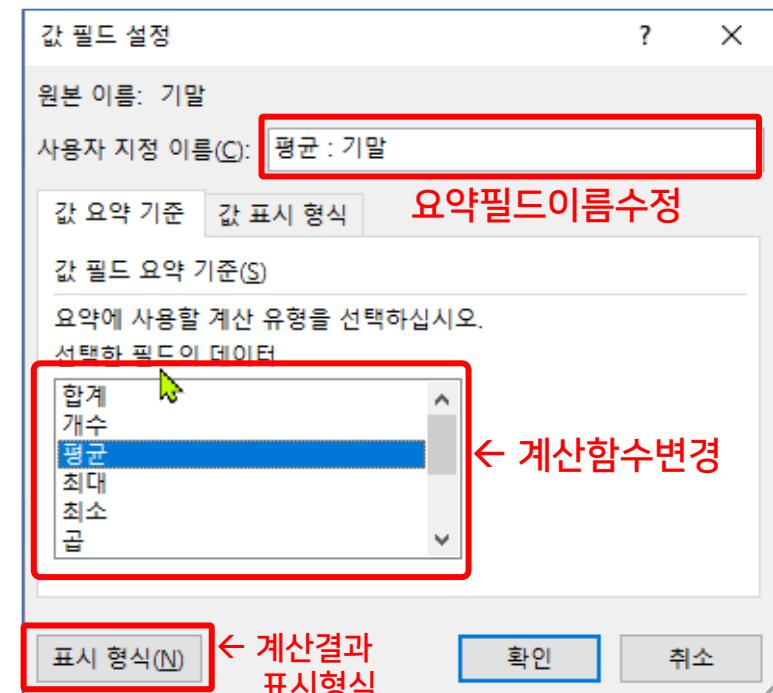
합계 : 기말



피벗테이블 작성

- 계산 필드 편집 → 필드 선택 후 [값 필드 설정]메뉴

값필드설정→



- 계산 필드를 제거: 단축메뉴 또는 필드이름 마우스드래그드랍으로 필드 삭제



피벗테이블 작성

● 소속학과별 인원수와 기말성적평균 요약집계 결과

	A	B	C	D	E
1					
2					
3	소속학과	인원수	중간평균	기말평균	기말최고점
4	경영정보	2	55.4	55.2	58.18
5	공공인재	2	89.5	79.8	84.55
6	도시환경	2	63.6	54.5	67.73
7	산업디자인	3	58.8	39.5	75.45
8	스포츠건강과학	2	54.6	51.6	70.91
9	식품영양	2	70.7	68.9	76.82
10	의류디자인	1	55.9	57.3	57.27
11	컴퓨터	3	67.0	67.9	80
12					
13					
14					
15					
16					
17					
18					
19					
20					

피벗 테이블 필드

보고서에 추가할 필드 선택:

검색

소속학과

인원수

중간평균

기말평균

기말최고점

아래 영역 사이에 필드를 끌어 놓으십시오.

필터

열

값

행

소속학과

인원수

중간평균

기말평균

기말최고점

구성된 데이터 중 소속학과별 인원수,
중간/기말 평균과 기말 최대값 데이터 파악

소속학과	인원수	중간평균	기말평균	기말최고점
경영정보	2	55.4	55.2	58.18
공공인재	2	89.5	79.8	84.55
도시환경	2	63.6	54.5	67.73
산업디자인	3	58.8	39.5	75.45
스포츠건강과학	2	54.6	51.6	70.91
식품영양	2	70.7	68.9	76.82
의류디자인	1	55.9	57.3	57.27
컴퓨터	3	67.0	67.9	80

※ '기말'필드를 계속 추가하여 계산함수를
최대값으로 변경하여 다양한 의미를
파악할 수 있음



피벗테이블 편집

- [피벗테이블도구][분석]메뉴탭 → [피벗테이블] → [옵션]메뉴

The screenshot shows the Microsoft Excel ribbon with the "Analysis" tab selected. The ribbon tabs include File, Home, Insert, Draw, Page Layout, Styles, Data, Review, Developer, Add-Ins, Power Pivot, Analysis, Design, and Help. The "Analysis" tab is highlighted with a red box.

Below the ribbon, the "PivotTable Options" dialog box is displayed. It contains the following sections:

- General**:
 - PivotTable Name: PivotTable1
 - Layout & Structure tab is selected (highlighted with a red box).
 - Show summary rows (unchecked)
 - Show multiple consolidation ranges (checked)
- Consolidation**:
 - Show summary rows (unchecked)
 - Show multiple consolidation ranges (checked)
- Filter**:
 - Filtered pages included in subtotal calculations (unchecked)
 - Filter applied to one field (unchecked)
- Calculated**:
 - Use calculated fields and items (checked)

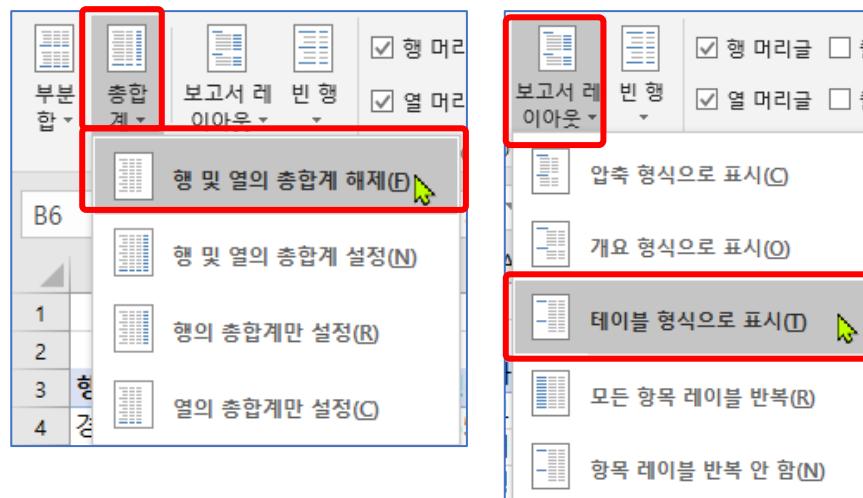
At the bottom of the dialog box are "OK" and "Cancel" buttons.

To the left of the dialog box, the "PivotTable Options" ribbon is visible, showing the "PivotTable" tab selected (highlighted with a red box). Other tabs include "Options" (selected), "Report Filter Page", and "GetPivotData".



피벗테이블 편집

- [피벗테이블도구][디자인]메뉴탭 → [레이아웃]그룹 → [총합계]메뉴
- [피벗테이블도구][디자인]메뉴탭 → [레이아웃]그룹 → [보고서레이아웃]메뉴



※ 요약 완성

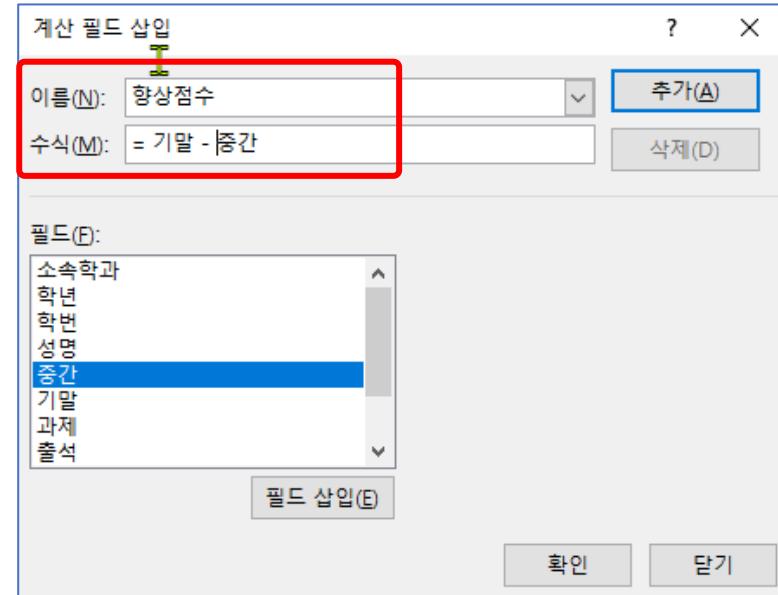
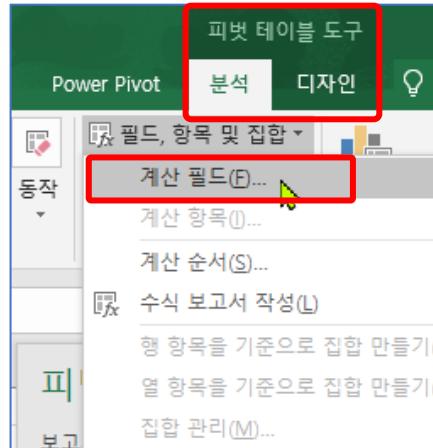
소속학과	인원수	중간평균	기말평균
경영정보	2	55.4	55.2
공공인재	2	89.5	79.8
도시환경	2	63.6	54.5
산업디자인	3	58.8	39.5
스포츠건강과학	2	54.6	51.6
식품영양	2	70.7	68.9
의류디자인	1	55.9	57.3
컴퓨터	3	67.0	67.9



계산필드 추가하기

- [피벗테이블도구][분석]메뉴탭 → [계산]그룹 → [필드, 항목및집합] → [계산필드]

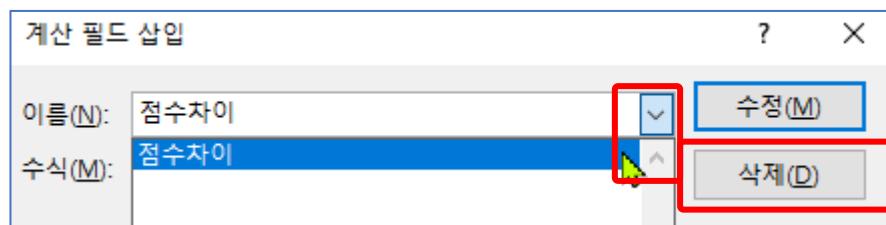
행 레이블	인원수	중간평균	기말평균
경영정보	2	55.4	55.2
공공인재	2	89.5	79.8
도시환경	2	63.6	54.5
산업디자인	3	58.8	39.5
스포츠건강과학	2	54.6	51.6
식품영양	2	70.7	68.9
의류디자인	1	55.9	57.3
컴퓨터	3	67.0	67.9
총합계	17	64.7	58.8



행 레이블	인원수	중간평균	기말평균	향상점수평균
경영정보	2	55.4	55.2	-0.3
공공인재	2	89.5	79.8	-19.4
도시환경	2	63.6	54.5	-18.1
산업디자인	3	58.8	39.5	+57.6
스포츠건강과학	2	54.6	51.6	-6.0
식품영양	2	70.7	68.9	-1.8
의류디자인	1	55.9	57.3	1.4
컴퓨터	3	67.0	67.9	2.6

계산필드 추가 후
함수를 평균으로 설정

※계산 필드 완전 삭제 : [계산필드]메뉴 → 목록버튼 → [삭제]버튼





기준 필드의 그룹 설정

- 기준 필드 데이터가 수치데이터이면 다양한 요약 집계를 위하여 그룹 설정 가능

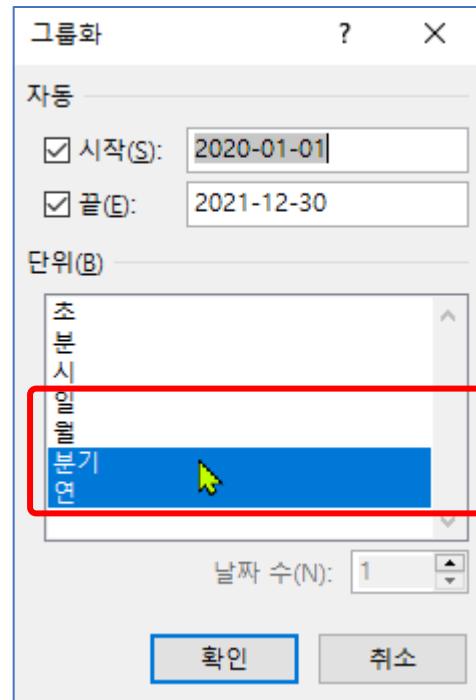
A	B	C	D	E	F	G	H	I	J
거래번호	고객	담당	주문일	분류	제품	단가	수량	할인율	판매액
2 N-0705	S&C무역 (주)	오서윤	2020-01-01	복사기	컬러레이저복사기 XI-3200	1,176,000	3	15%	2,998,800
3 N-0705	S&C무역 (주)	오서윤	2020-01-01	바코드스캐너	바코드 Z-350	48,300	3	0%	144,900
4 N-0705	S&C무역 (주)	오서윤	2020-01-01	팩스	잉크젯팩시밀리 FX-1050	47,400	3	0%	142,200
5 N-0706	드림씨푸드 (주)	박현우	2020-01-04	복사용지	프리미엄복사지A4 2500매	17,800	9	0%	160,200
6 N-0706	드림씨푸드 (주)	박현우	2020-01-04	바코드스캐너	바코드 BCD-100 Plus	86,500	7	0%	605,500
7 N-0707	자이언트무역 (주)	정시우	2020-01-05	복사용지	고급복사지A4 500매	3,500	2	0%	7,000
8 N-0707	자이언트무역 (주)	정시우	2020-01-05	바코드스캐너	바코드 Z-350	46,300	7	0%	324,100
9 N-0707	자이언트무역 (주)	정시우	2020-01-05	바코드스캐너	바코드 BCD-100 Plus	104,500	8	0%	836,000
10 N-0708	진왕통상 (주)	오서윤	2020-01-05	복합기	잉크젯복합기 AP-3300	79,800	1	0%	79,800
11 N-0709	지와트사 (주)	오서윤	2020-01-05	복합기	이그제보복합기 AP-2200	80,200	0	0%	711,400
2681 N-1749	동화상사 (주)	오서윤	2021-12-29	복사기	흑백레이저복사기 TLE-9000	896,500	2	5%	1,703,350
2682 N-1750	한성트레이딩 (주)	이서연	2021-12-29	복합기	레이저복합기 L950	504,400	4	10%	1,815,840
2683 N-1750	한성트레이딩 (주)	이서연	2021-12-29	팩스	잉크젯팩시밀리 FX-1050	51,000	8	5%	387,600
2684 N-1751	신화식품 (주)	박현우	2021-12-29	출퇴근기록기	지문인식 FPIN-1000+	130,000	8	5%	988,000
2685 N-1751	신화식품 (주)	박현우	2021-12-29	복합기	레이저복합기 L500	286,200	5	5%	1,359,450
2686 N-1752	용도교역 (주)	박현우	2021-12-29	복합기	무한잉크젯복합기 AP-3300W	101,000	4	0%	404,000
2687 N-1752	용도교역 (주)	박현우	2021-12-29	복사용지	프리미엄복사지A4 5000매	29,700	8	0%	237,600
2688 N-1753	노을백화점 (주)	유준혁	2021-12-29	문서세단기	오피스 Z-05C	120,500	10	5%	1,144,750
2689 N-1753	노을백화점 (주)	유준혁	2021-12-29	문서세단기	문서세단기 SCUT-1500B	686,800	1	5%	652,460
2690 N-1753	노을백화점 (주)	유준혁	2021-12-29	출퇴근기록기	지문인식 FPIN-2000F	172,800	8	5%	1,313,280



기준 필드의 그룹 설정

- 날짜 데이터의 그룹 설정
- 피벗테이블 작성 후 기준 열 데이터를 선택한 후 단축메뉴 [그룹]메뉴 실행

The screenshot shows the Microsoft Excel ribbon with the 'Home' tab selected. In the 'Layout' group, the '필드' button is active, displaying a dropdown menu. The '그룹(G)...' option is highlighted with a red box and has a yellow cursor arrow pointing to it. Other options in the menu include '그룹 해제(U)...', '이동(M)', '연 제거(V)', '필드 설정(N)...', '피벗 테이블 옵션(O)...', and '필드 목록 숨기기(D)'.



※ 날짜데이터의 연/분기/월/일 단위로
그룹 설정 가능
클릭하여 설정, 다시 클릭하면 설정 해제



기준 필드의 그룹 설정

● 날짜데이터 기준의 그룹 설정 결과

The screenshot shows the 'Pivot Field' settings dialog box from Power BI. On the left, there is a preview table with columns A and B. Column A contains years (2020년, 2021년) and quarters (1사분기, 2사분기, 3사분기, 4사분기). Column B contains values (4,737, 8,338, 1,017, 1,119, 1,259, 1,342, 1,178, 1,569, 2,150, 3,441, 13,075). Row 3 is highlighted with a green border and labeled '행 레이블' (Row Label) and '합계 : 수량' (Sum: Quantity). The '검색' (Search) field is empty. In the '보고서에 추가할 필드 선택:' (Select fields to add to the report) section, '수량' (Quantity) is checked. Below it, there is a note: '아래 영역 사이에 필드를 끌어 놓으십시오.' (Drag fields here between these areas). The '필터' (Filter) section shows '행' (Row) set to '연' (Year) and '값' (Value) set to '합계 : 수량' (Sum: Quantity). At the bottom, there is a checkbox for '나중에 레이아웃 업데이트' (Update layout later) and a '업데이트' (Update) button.

	A	B
1		
2		
3	행 레이블	합계 : 수량
4	2020년	4,737
5	1사분기	1,017
6	2사분기	1,119
7	3사분기	1,259
8	4사분기	1,342
9	2021년	8,338
10	1사분기	1,178
11	2사분기	1,569
12	3사분기	2,150
13	4사분기	3,441
14	총합계	13,075
15		
16		
17		
18		
19		
20		
21		

연도와 분기별 '수량'데이터의 합계를
요약 집계하여 파악함



기준 필드의 그룹 설정

- 일반 수치데이터를 기준으로 지정할 때 그룹 설정
- [그룹화] 대화상자의 시작값, 끝값, 단위를 입력

The screenshot shows a Microsoft Excel spreadsheet with a pivot table. The pivot table has '단가' (Unit Price) in row 3 and '합계 : 수량' (Sum: Quantity) in column B. The data rows show price ranges from 0-99999 to 1800000-1899999, with corresponding quantities like 8420, 2774, etc. A red box highlights the '단가' cell in row 3.

A 'PivotTable Fields' dialog box is open, showing the 'Gruppiere' (Group) settings. It has '자동' (Automatic) selected. Under 'Gruppiere', there are fields for '시작(S)' (Start) set to 0, '끝(E)' (End) set to 2000000, and '단위(B)' (Unit) set to 100000. A red box highlights this group settings dialog.

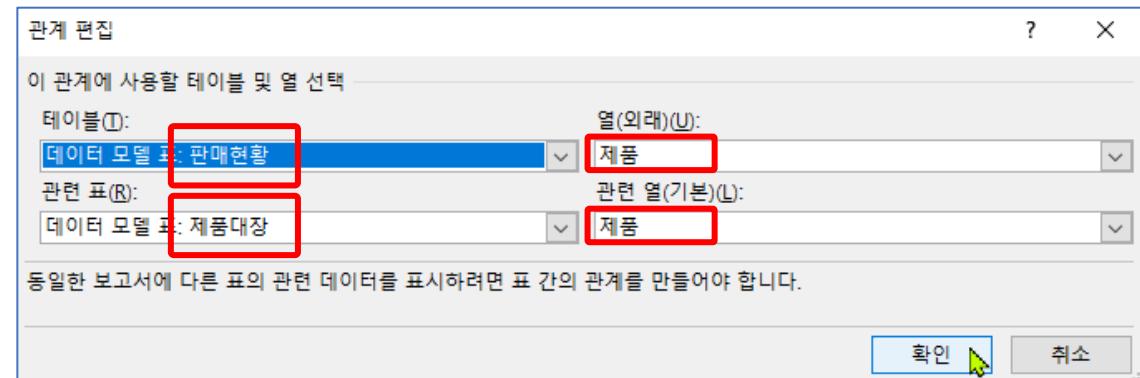
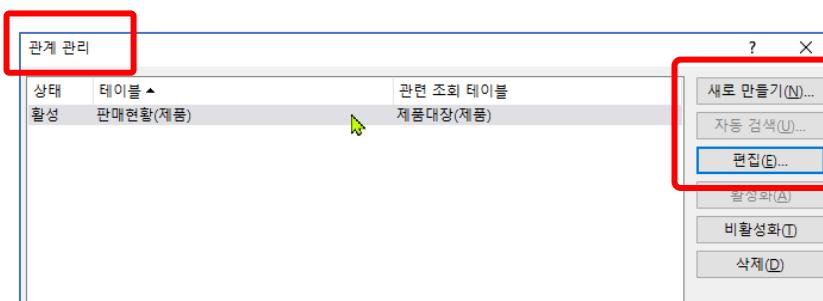
In the bottom right corner of the dialog, a blue arrow points to a callout box containing the following text:

'단가' 데이터를 십만 단위로
구분하여 해당 단가 그룹의
'수량' 데이터 합계 파악



관계 설정된 표들의 요약 집계

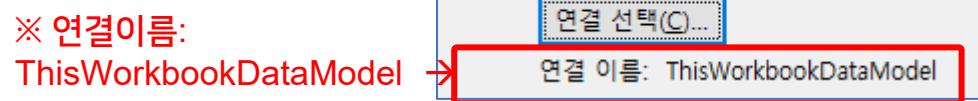
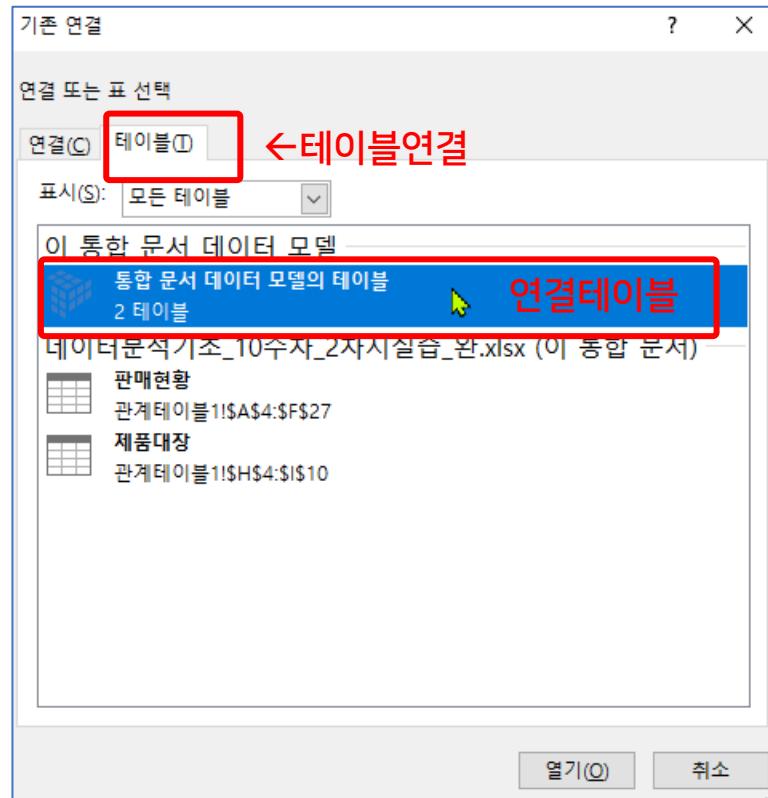
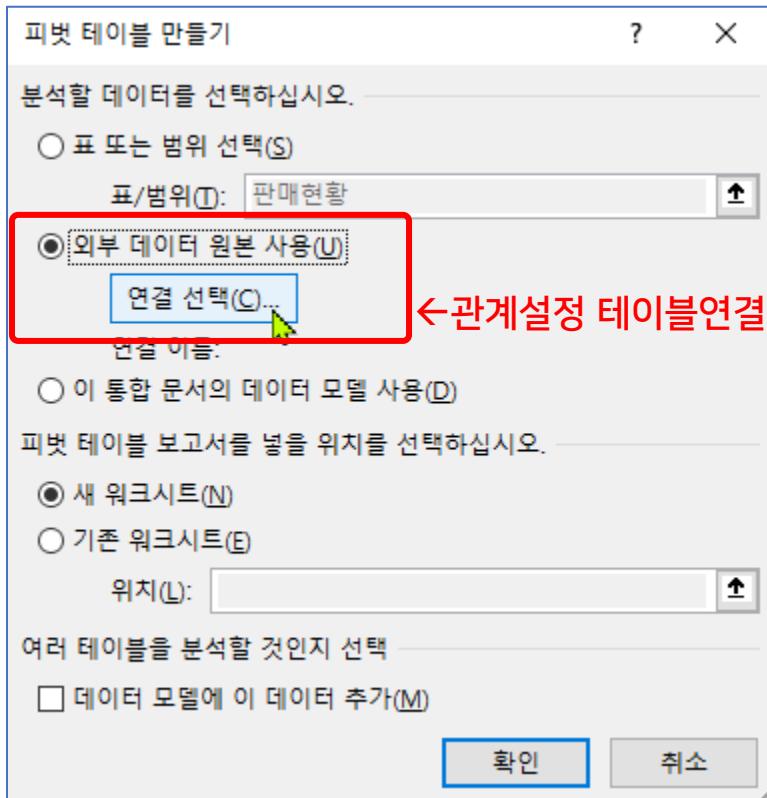
- 데이터 모델의 피벗테이블 작성
- 관계 설정된 두 테이블을 기반으로 요약 집계
 - 분리된 2개 이상의 표
 - [데이터]메뉴탭 → [데이터도구]그룹 → [관계]메뉴





관계 설정된 표들의 요약 집계

- [삽입]메뉴탭 → [피벗테이블]메뉴





관계 설정된 표들의 요약 집계

● 관계 데이터 모델의 피벗테이블 필드 목록

The screenshot illustrates the configuration of a PivotTable in Excel based on a relational data model.

Left Side: A screenshot of an Excel spreadsheet with a PivotTable named "피벗 테이블1". The user has selected the range A3:D27. A callout bubble points to the PivotTable Fields dialog box.

PivotTable Fields Dialog (Left): Shows the "활성" tab selected. Under "보고서에 추가할 필드 선택:", the "모두" button is highlighted. The "판매현황" category is expanded, showing fields: 거래번호, 고객, 제품, 단가, 수량, and 판매. These last four fields are highlighted with a red box.

PivotTable Fields Dialog (Right): Shows the "활성" tab selected. Under "보고서에 추가할 필드 선택:", the "모두" button is highlighted. The "제품대장" category is expanded, showing fields: 분류 (highlighted with a red box) and 제품. The "판매현황" category is also expanded, showing fields: 거래번호, 고객, 제품, 단가, 수량 (highlighted with a red box), and 판매 (highlighted with a red box). The "수량" and "판매" fields are checked.

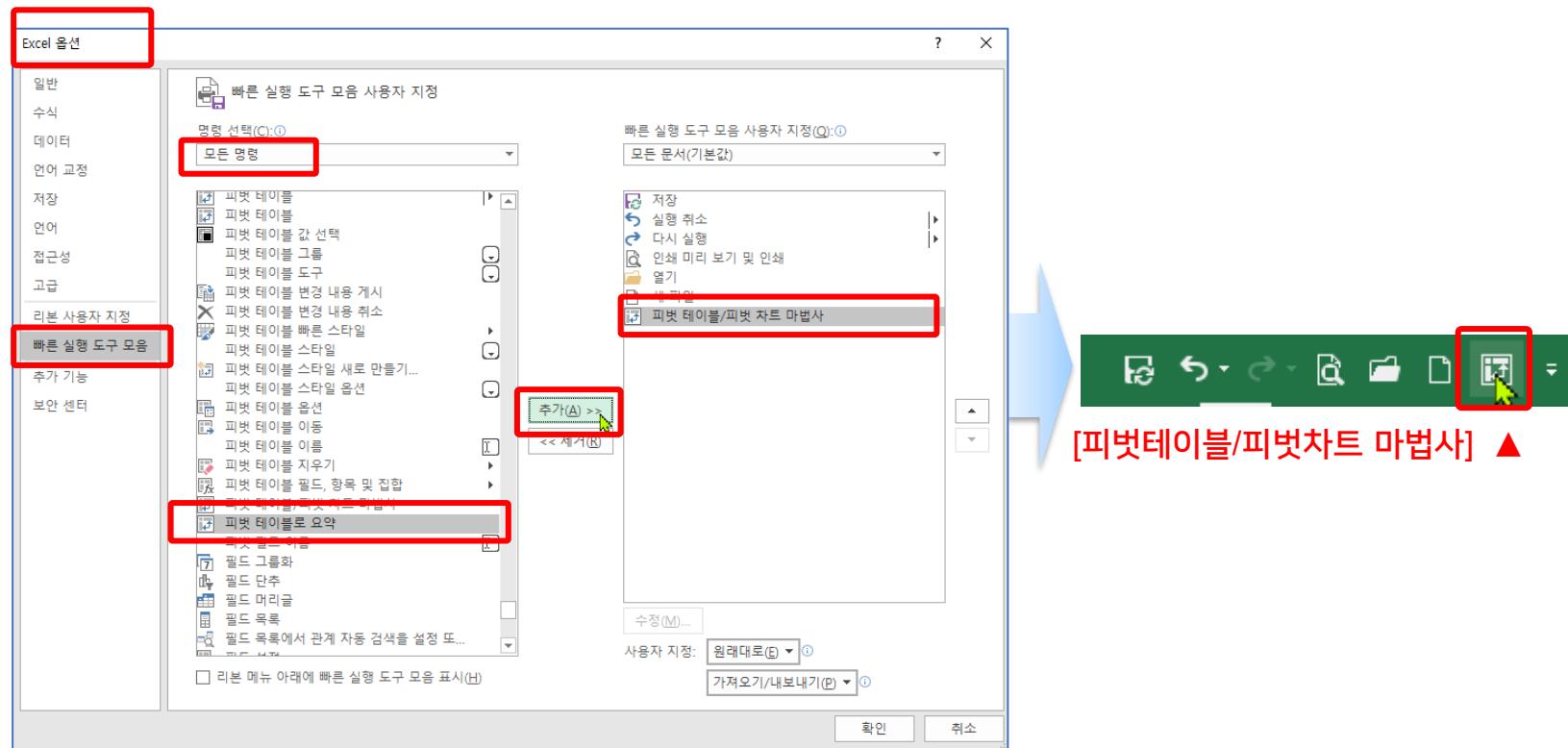
Resulting PivotTable: An arrow points from the dialogs to a screenshot of the final PivotTable. The PivotTable has "행 레이블" set to "제품" and "열" set to "수량" and "판매". The data shows sales for different products:

제품	합계: 수량	합계: 판매
바코드스캐너	92	8256450
제본기	50	7156465
총합계	142	15412915



피벗테이블 / 피벗차트 마법사 활용

- 집계표 형식의 분리된 표를 하나로 취합하기 위한 피벗테이블마법사 기능 제공
- 피벗테이블마법사 메뉴
 - 기본 메뉴 탭에 표시 없음 → 빠른실행도구모음에 등록하여 사용가능
 - [파일] → [옵션] → [빠른실행도구모음] → [모든명령] → [피벗테이블/피벗차트마법사]





피벗테이블 / 피벗차트 마법사 활용

- 피벗테이블 / 피벗차트 마법사를 적용할 수 있는 표

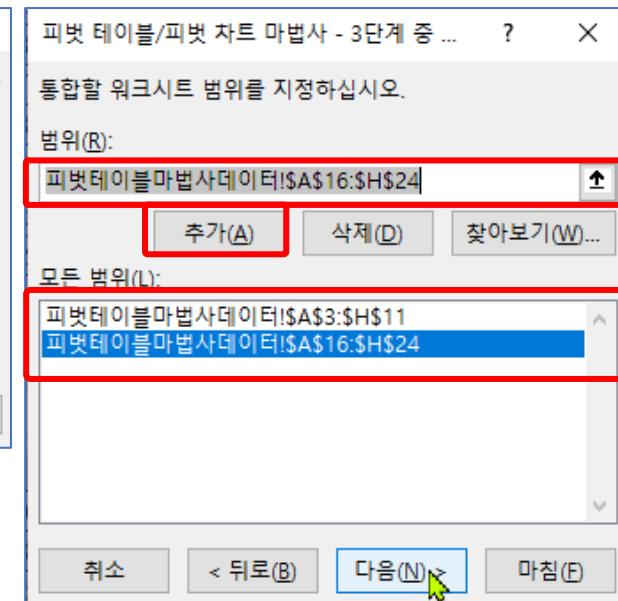
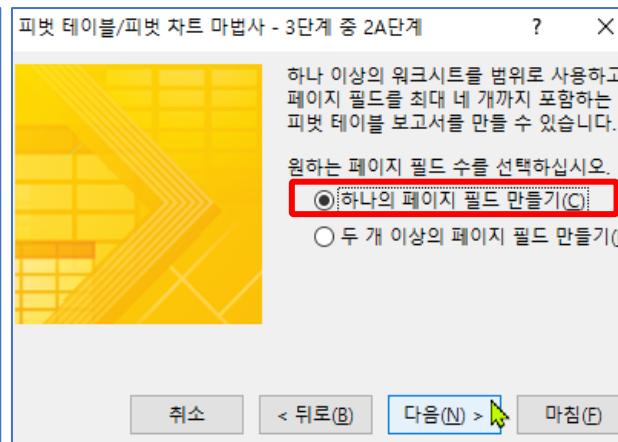
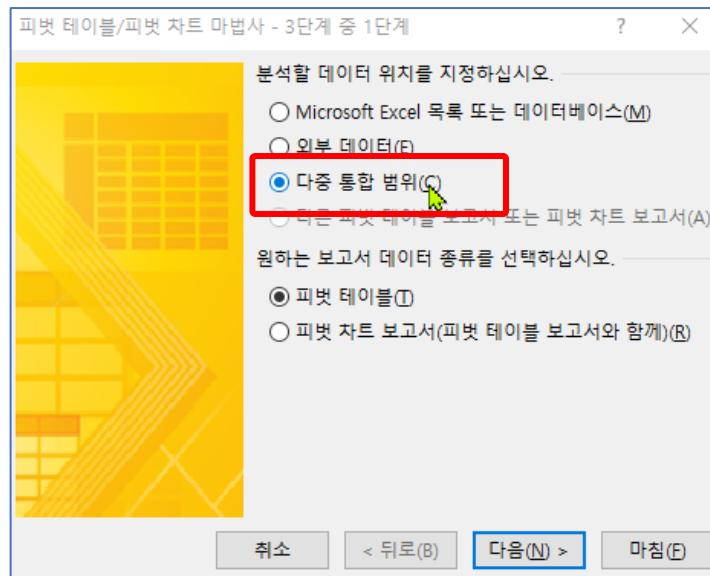
	A	B	C	D	E	F	G	H
1	2021년도 현황							
2	분류	서울	광역시	경기	경상	전라	강원	충청
3	복사기	75,900,910	73,974,380	32,874,325	12,109,425	4,045,350	2,774,400	
4	복합기	46,712,555	29,993,660	20,779,810	3,208,020	5,180,080	1,585,055	1,147,400
5	문서세단기	34,178,570	9,482,745	9,435,700	2,561,010	1,237,000	199,500	173,800
6	제본기	22,712,965	15,782,860	10,081,880	662,400	483,000	552,600	1,341,000
7	바코드스캐너	17,831,785	12,191,190	8,313,985	720,195	1,078,800	1,933,400	2,524,300
8	출퇴근기록기	15,599,485	11,081,020	7,303,900	437,915	336,100	731,500	462,900
9	복사용지	10,078,595	7,009,060	4,907,285	200,340	509,000	774,200	355,000
10	팩스	3,528,525	4,623,295	2,513,795	747,800	988,800	93,900	
11	지역합계	226,543,390	164,138,210	96,210,680	20,647,105	13,858,130	8,644,555	6,004,400
12								
13								
14	2022년도 현황							
15	분류	서울	광역시	경기	경상	전라	강원	충청
16	복합기	192,241,205	113,612,315	69,101,610	12,410,175	5,501,520	7,236,015	5,819,370
17	복사기	182,558,085	124,444,225	63,600,845	13,286,230	15,505,020	4,456,740	
18	문서세단기	65,091,060	39,793,325	29,146,555	6,040,075	37,145	1,797,970	1,216,380
19	출퇴근기록기	38,239,795	17,857,855	11,398,995	768,580	497,990	53,900	1,774,860
20	제본기	28,954,110	23,367,325	9,976,935	2,930,700			
21	바코드스캐너	26,451,455	15,166,080	10,032,660	1,826,565	98,700	1,176,365	1,901,055
22	팩스	22,095,585	12,271,245	5,342,195	2,669,800		1,516,930	220,590
23	복사용지	16,040,240	8,176,665	4,471,640	1,735,320	314,700	1,244,535	679,095
24	지역합계	571,671,535	354,689,035	203,071,435	41,667,445	21,955,075	17,482,455	11,611,350
25								

- 행제목과 열제목의 순서가 일치하지 않더라도 같은 항목끼리 통합



피벗테이블 / 피벗차트 마법사 활용

- 마법사 1단계 → [다중 통합 범위]로 지정
- 마법사 2단계 → [하나의 페이지 필드 만들기] 지정
- 마법사 3단계 → 데이터 범위 지정하여 추가



- 데이터 범위를 지정시 총합은 제외하고 지정



피벗테이블 / 피벗차트 마법사 활용

● 통합 결과

	A	B	C	D	E	F	G	H	I
1	페이지1	(모두)							
2									
3	합계 : 값	열 레이블							
4	행 레이블	강원	경기	경상	광역시	서울	전라	충청	총합계
5	문서세단기	4,358,980	38,582,255	6,213,875	49,276,070	99,269,630	236,645	2,453,380	200,390,835
6	바코드스캐너	1,896,560	18,346,645	4,350,865	27,357,270	44,283,240	2,032,100	2,979,855	101,246,535
7	복사기	16,566,165	96,475,170	13,286,230	198,418,605	258,458,995	18,279,420	4,045,350	605,529,935
8	복사용지	1,444,875	9,378,925	2,090,320	15,185,725	26,118,835	1,088,900	1,188,095	56,495,675
9	복합기	10,444,035	89,881,420	13,557,575	143,605,975	238,953,760	7,086,575	10,999,450	514,528,790
10	제본기	662,400	20,058,815	4,271,700	39,150,185	51,667,075	552,600	483,000	116,845,775
11	출퇴근기록기	491,815	18,702,895	1,231,480	28,938,875	53,839,280	1,229,490	2,110,960	106,544,795
12	팩스	2,264,730	7,855,990	2,669,800	16,894,540	25,624,110	93,900	1,209,390	56,612,460
13	총합계	38,129,560	299,282,115	47,671,845	518,827,245	798,214,925	30,599,630	25,469,480	1,758,194,800
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									

피벗 테이블 필드

보고서에 추가할 필드 선택:

검색

행 세로 막대형 값 페이지1

기타 테이블...

아래 영역 사이에 필드를 끌어 놓으십시오.

필터 열

페이지1 세로 막대형

행 값

나중에 레이아웃 업데이트 업데이트

탐색적 데이터 분석(EDA)

Exploratory Data Analysis



데이터 분석

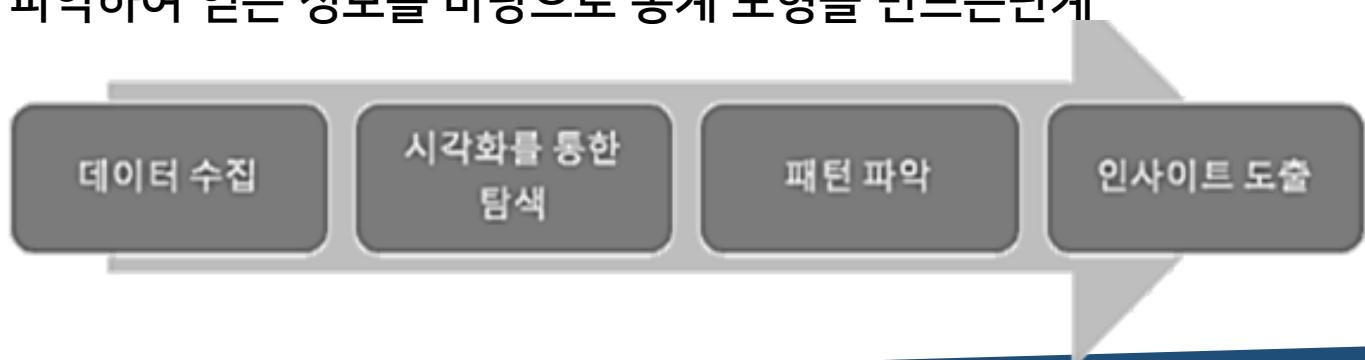
● CDA(Confirmatory Data Analysis) : 확증적 데이터 분석

- 목적을 가지고 데이터를 확보하여 분석하는 방법 (연역적 방법)
- 관측된 형태나 효과의 재현성평가, 유의성 검정, 신뢰구간 추정 등 통계적 추론하는 단계
- 보통은 설문조사, 논문에 대한 내용을 입증하는데 많이 사용



● EDA(Exploratory Data Analysis) : 탐색적 데이터 분석

- 쌓여있는 데이터를 기반으로 가설을 세워 데이터를 분석하는 방법 (귀납적 방법)
- 데이터의 구조와 특징을 파악하여 얻은 정보를 바탕으로 통계 모형을 만드는 단계
- 빅데이터 분석에 사용됨



데이터 탐색적 분석



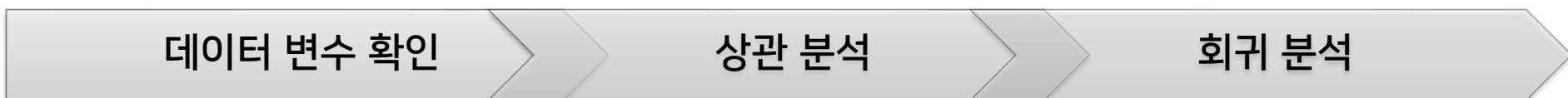
- 데이터 탐색적 분석 (Exploratory Data Analysis: EDA)
- 데이터의 속성 및 특징 이해
- 데이터 변수들간의 관계 파악
- 데이터의 패턴 파악 및 잘못된 자료들을 탐색
- 데이터를 분석의 가설을 정형화하고 세분화하고 분석의 방향 및 방법 제시
- 본격적인 데이터 분석에 앞서 데이터의 주요 특성을 요약 또는 시각화
- EDA는 본격적인 데이터 분석 모델링을 하기 전 데이터 이해를 돋기 위해 실시
- EDA를 위해 통계 기법을 사용



탐색적 데이터 분석(EDA)

● 데이터 탐색(EDA)

- 데이터 분석의 초기 단계에서 데이터의 특성을 살펴보는 과정
- 데이터 세트를 분석 및 조사하고 주요 특성을 요약하여 데이터의 패턴 발견, 이상 징후 발견, 가설을 테스트, 가정을 확인하는 작업이라 할 수 있음
- 통계학을 기반으로 데이터의 특성을 발견할 수 있음



● 통계학(統計, Statistics)

- 한데 몰아서 어림잡아 계산(표준국어대사전), 현상을 통계에 의하여 관찰 · 연구하는 학문
- 불확실성에 대한 논리를 부여하는 학문, 경험과학의 한 분야이자 대부분 학문의 기초이며, 다양한 정의가 존재하고 축약하면 자료를 연구하는 학문, 데이터를 분석하는 학문
- 기술통계학(**Descriptive Statistics**) : 데이터를 수집, 정리, 요약하여 데이터 의미를 기술(설명)
- 추론통계학(**Inferention Statistics**) : 표본 자료에서 얻은 정보를 이용하여 전체 집단(단위)에 대한 정보 및 불확실한 사실에 대해 예측하는 방법과 이론을 제시



통계(統計, Statistics)

● 표준국어대사전

- 한데 몰아서 어림잡아 계산함
- 어떤 현상을 종합적으로 한눈에 알아보기 쉽게 일정한 체계에 따라 숫자로 나타냄

● 통계학

- 표준국어대사전: 현상을 통계에 의하여 관찰 · 연구하는 학문
- 불확실성에 대한 논리를 부여하는 학문, 경험과학의 한 분야이자 대부분 학문의 기초
- 다양한 정의가 존재하고 축약하면 자료를 연구하는 학문, 데이터를 분석하는 학문

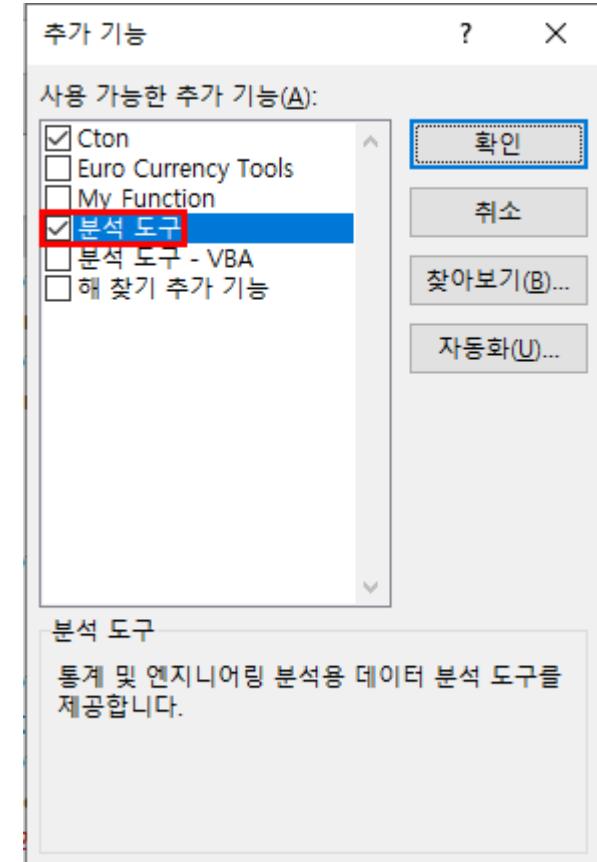
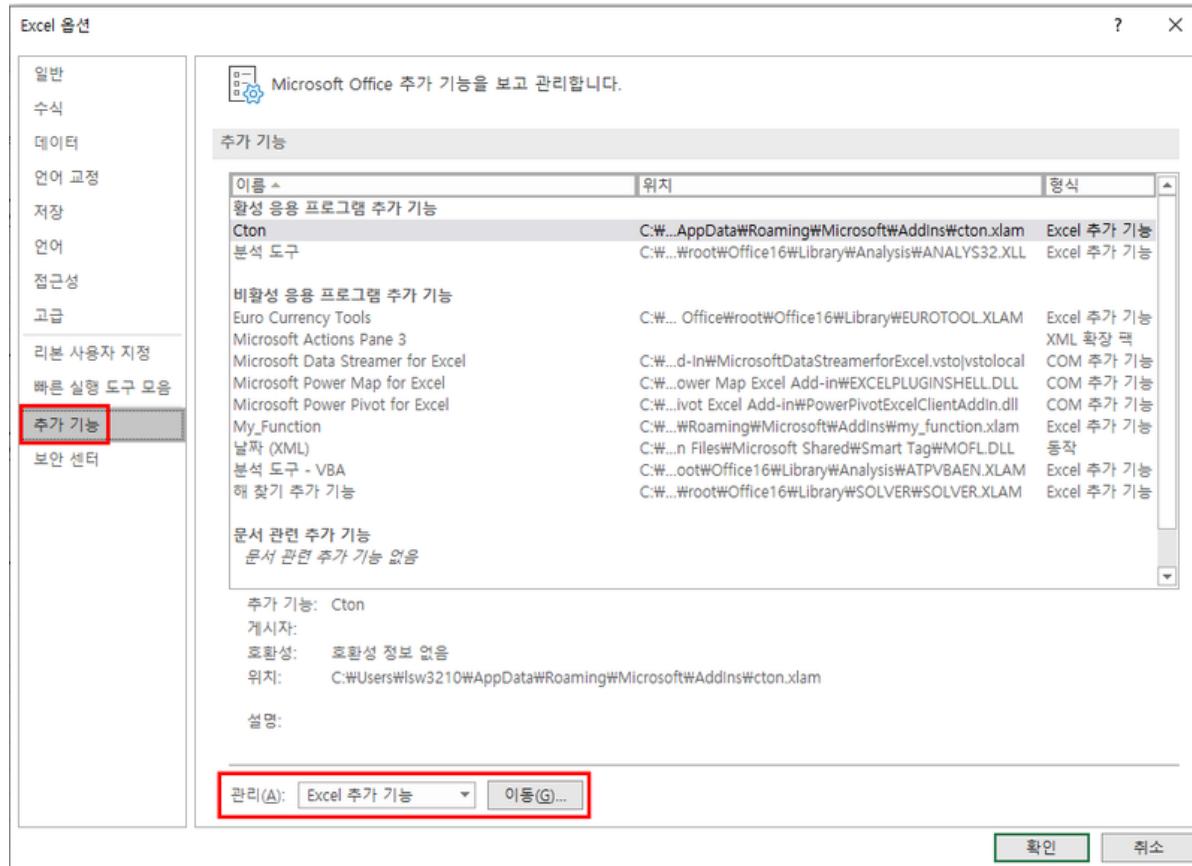
● 기술통계 : 데이터들을 수집, 정리, 요약

- 추론통계 : 표본 자료에서 얻은 정보를 이용하여 전체 집단(단위)에 대한 정보 및 불확실한 사실에 대해 예측하는 방법과 이론을 제시



통계 기반 Excel 데이터 분석 기능

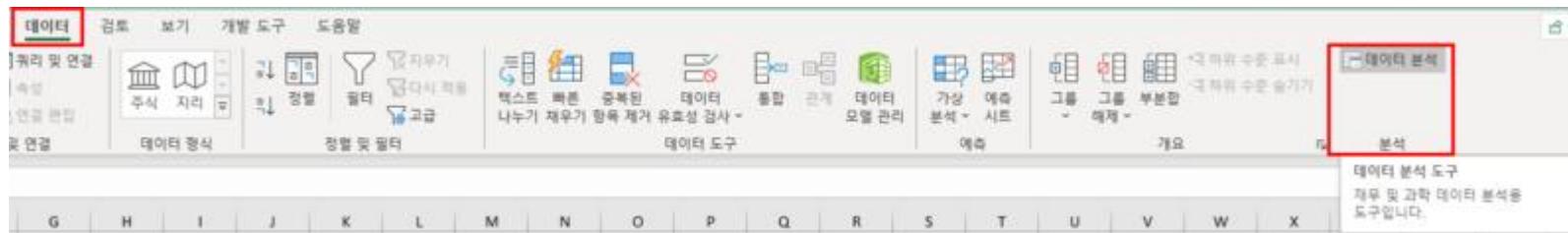
- Excel 추가 기능 → 데이터 분석 도구 설치
- [파일] → [옵션] → [추가기능] → [엑셀 추가 기능] 이동



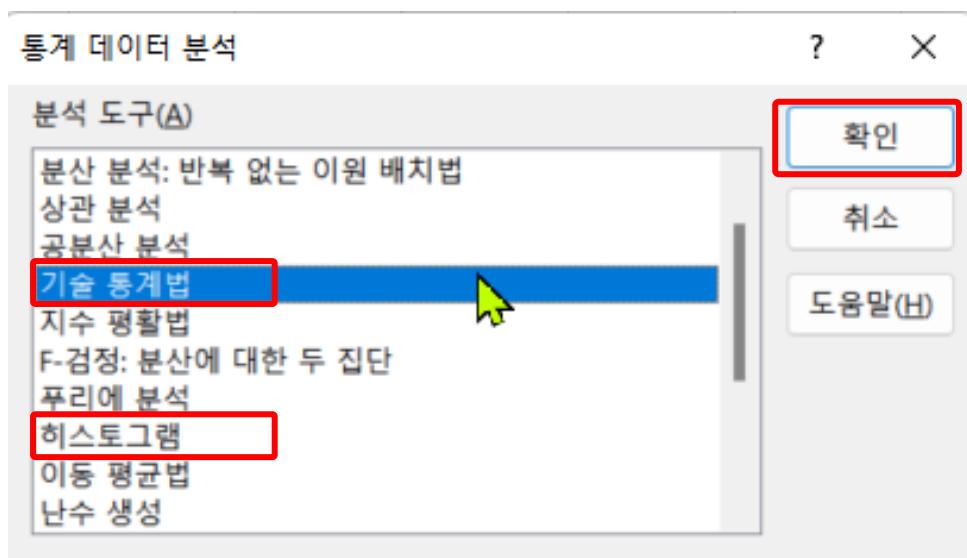


통계 기반 Excel 데이터 분석 기능

- [데이터]탭 → 추가된 [데이터 분석] 메뉴



- 기술 통계법 / 히스토그램

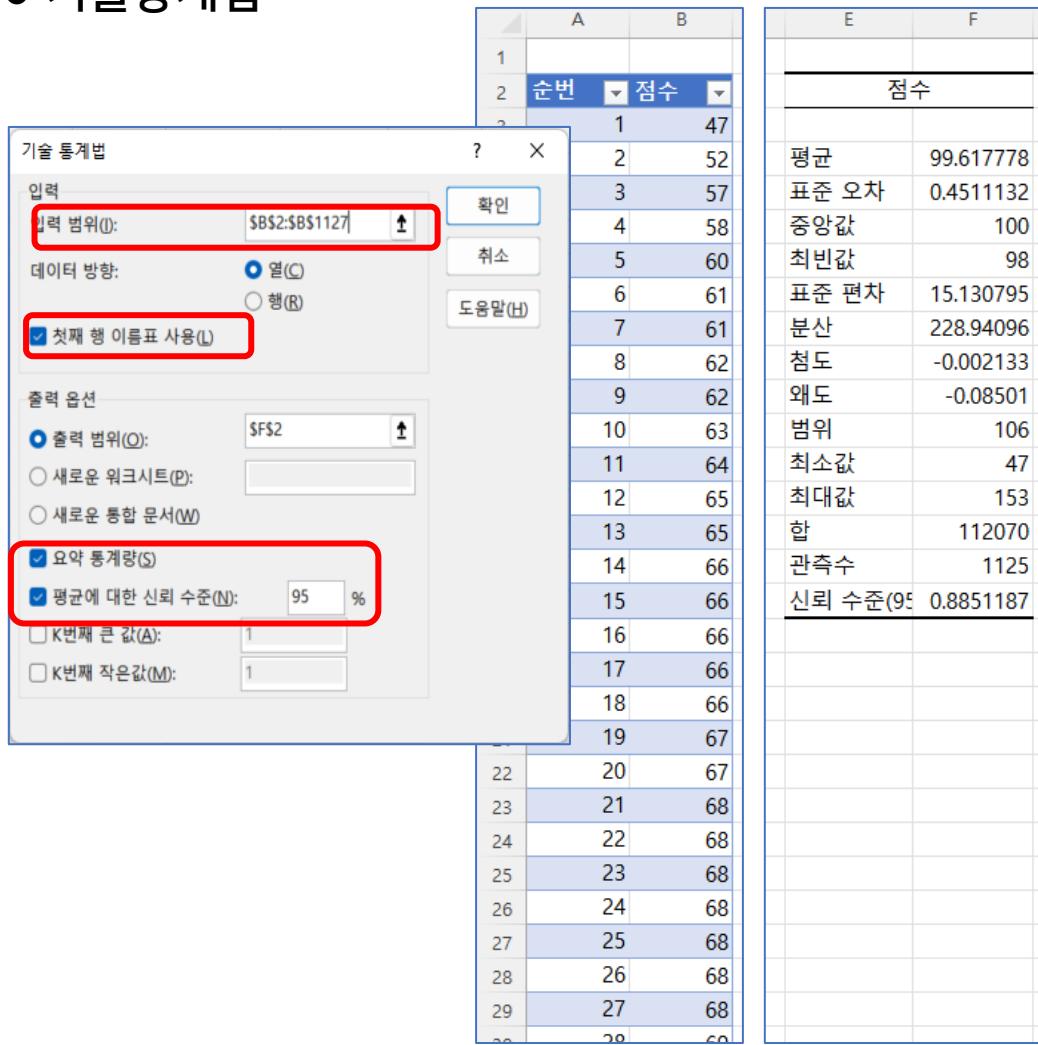




기술 통계: 기술 통계 실습

- [데이터]메뉴탭 → [데이터 분석]메뉴

- 기술통계법



- 중심성

- ✓ 중앙값은 100이고 최빈값이 98로 100과 98 사이의 값에 집중화 경향을 보임

- 변동성

- ✓ 최댓값이 153이고, 최솟값은 47로 범위가 106인 것을 알 수 있음, 또한 사람들 간의 점수 편차는 15.13 정도로 차이가 남

- 정규성

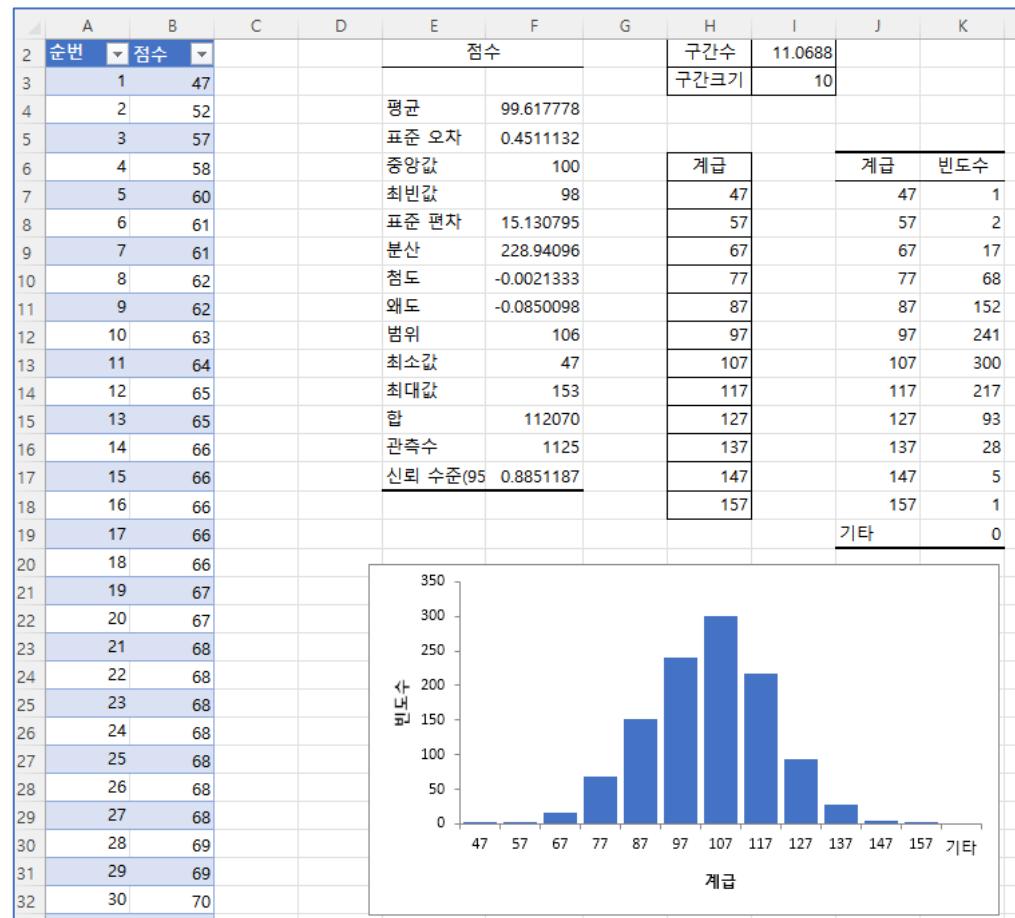
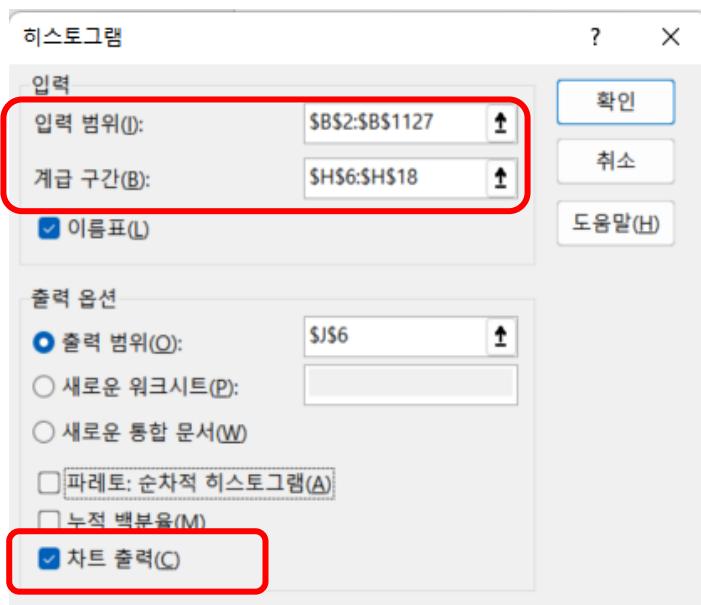
- ✓ 왜도값은 -0.085로 0보다 작기 때문에 부적비대칭으로 좌측으로 긴 꼬리를 가진 형태임
- ✓ 또한 첨도는 -0.002로 0보다 작기 때문에 가운데에 데이터들이 정규분포보다 적게 몰려있음



기술 통계: 히스토그램(도수분포표)

- 계급을 위한 구간수와 구간크기 구하기

- 구간수 정하기
- 구간크기 = (최대값-최소값) / 구간수
- 계급 경계값: 최소값부터 구간크기를 순서대로 더하여 구함
- 계급 구간 지정





빅데이터시대의 기술 통계와 추론 통계

● 기술 통계

- 기초통계로 데이터 요약 집계하여 데이터를 설명함

● 추론 통계

- 가설 검정: 모집단에서 샘플링한 표본으로 모집단의 특성을 추론하고 그 결과가 신뢰성이 있는지 검정하는 과정

● 빅데이터 시대

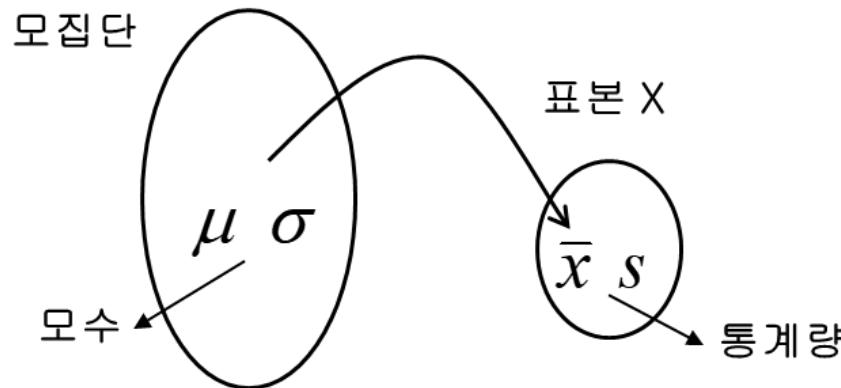
- 모집단과 표본집단을 구분하기 보다는 **데이터 전체를 표본으로 하고 수집하지 못한 현실 세계 전체 데이터나 미래에 대한 데이터를 모집단으로 볼 수 있음**

예) 통신 회사의 고객 데이터 전체는 데이터 과학자의 샘플링 단계가 없더라도 표본 집단으로 보고, 수집하지 못한 다른 회사의 고객데이터나 다음 달 고객 데이터는 모집단이라고 볼 수 있음

통계 기초 이론: 모집단과 표본



모집단	표본
관심의 대상이 되는 모든 개체의 관측값이나 측정값의 집합	모집단에서 실제로 추출한 관측값이나 측정값



구분	모수	통계량
대상	모집단의 특성	표본에서 계산한 특성
표시	그리스, 로마자로 표시	알파벳으로 표시
평균	μ	x
표준편차	σ	s

통계 기초 이론: 기술통계 (Descriptive Statistics)



- 수집한 데이터를 요약 묘사 설명하는 통계 기법

※ “Descriptive : 묘사하는, 그려서 설명하는”

- 기술통계(Descriptive Statistics) 기법

- 수집한 데이터를 대표하는 값이 무엇인지 찾기 → 대표값
- 수집한 데이터가 어떻게 퍼져 있는지를 설명하기 → 데이터 분포

- 대푯값

- 주어진 자료를 대표하는 특정 값
- 대푯값은 자료의 중심적인 경향이나 자료분포의 중심의 위치를 나타냄
- 평균 / 중앙값 / 최빈값 등

- 데이터 분포

- 주어진 자료의 퍼짐 정도를 표현하는 값 또는 시각화 차트
- 분산 / 표준편차 / 사분위값 / 왜도 / 첨도 / 도수분포표 / 히스토그램

통계 기초 이론: 기술통계 (Descriptive Statistics)

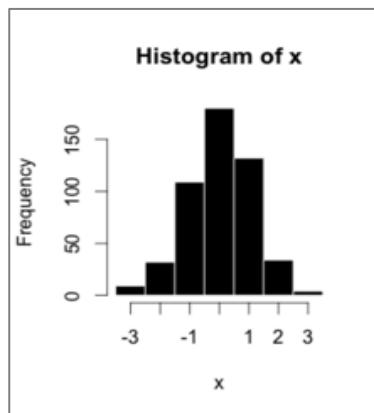


● 수치적인 탐색 : 기술통계

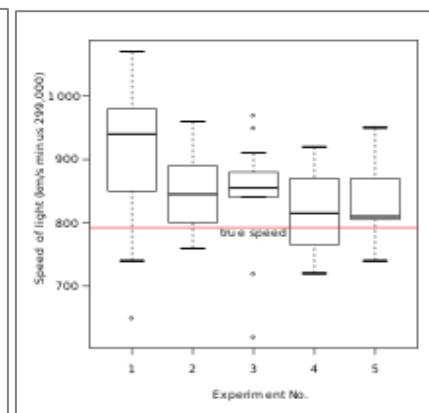
- 대표값 → 평균(Mean), 최대값(Max), 최소값(Min), 중앙값(Median), 최빈값(Mode)
- 데이터 분포 → 표준편차(Standard Deviation), 분산(Variance), 사분위수범위(Interquartile Range), 첨도(Kurtosis), 왜도(Skewness)

● 시각화(그래프) 탐색 : 분포

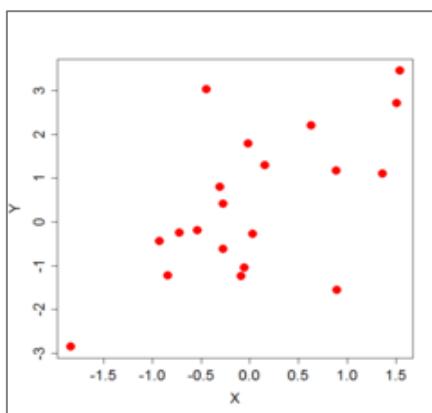
- 히스토그램(Histogram)
- 상자수염그림(Box plots)
- 산점도(Scatter plots)



- 히스토그램(Histogram) -



- 상자수염그림(Box plots) -



- 산점도(Scatter plots) -



통계 기초 이론: 대푯값

● 평균

- 주로 산술평균을 사용

- 산술평균: 데이터 모음의 값을 모두 더한 후 데이터 개수로 나눈 값
- 산술평균 외에 용도에 따라 기하평균, 조화평균, 가중평균 등을 사용하기도 함
- 데이터 중 극단적인 값 또는 이상한 값이 섞여 있거나 최대값이나 최소값의 크기가 크게 차이 나면 평균값이 데이터 모음의 대푯값으로 왜곡된 의미를 가질 수 있음

● 중앙값

- 수집된 데이터 모음의 값을 크기별로 나열하였을 때 가장 중앙에 위치한 값
 - 데이터 개수가 짝수개이면 중앙 위치의 두 데이터의 평균으로 표시하기도 함

● 최빈값

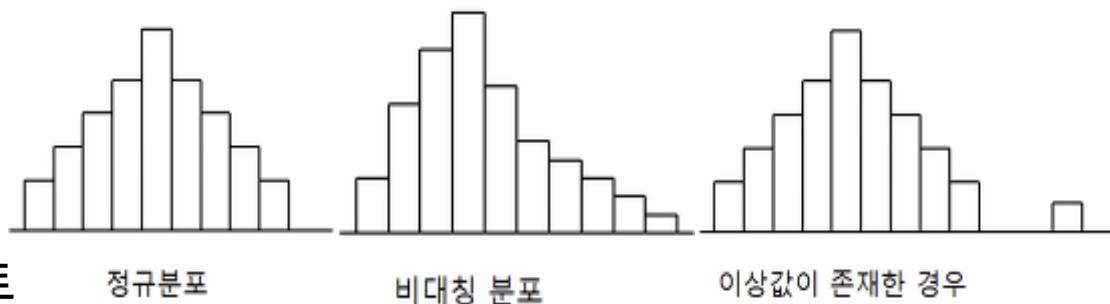
- 수집된 데이터 모음에서 가장 자주 발생한 값(빈도수가 가장 큰 값)
 - 주로 대소관계가 의미 없는 자료(명목형)에서 많이 사용됨



통계 기초 이론: 데이터 분포

● 도수분포표(Frequency table)

- 자료의 분포를 구간으로 나누고 각 구간에 속하는 값이 몇 개인지 빈도를 나타낸 표
- 전체 분포를 요약해서 파악할 수 있음
 - 자료의 개수, 자료 중 최대값과 최소값을 구함
 - 구간(계급수)의 개수를 결정: 자료의 개수나 분포에 따라 다름
 - 각 구간에 5개 이상의 값이 들어가는 것을 추천 (일반적으로 5~15구간 추천)
 - 구간의 폭(계급의 폭)을 구함: 구간폭 = (최대값 – 최소값) / 구간수
 - 정수, 짝수, 5의배수 등의 사용을 추천
 - 구간의 경계값을 구함: 최소값에서부터 구간폭을 더해서 구간의 경계값을 구함
 - 구간별 값의 개수(도수)를 표시



● 히스토그램(Histogram)

- 도수분포를 막대그래프로 시각화한 차트
- 그래프 모양(종모양, 비대칭 종모양 등)으로 데이터 분포를 파악할 수 있음



통계 기초 이론: 기술 통계 값의 의미

- 중심성 : 데이터가 어느 부분에 집중되는가?
 - 평균, 중앙값, 최빈값
- 변동성 : 데이터의 중심으로부터 얼마나 떨어져있는가?
 - 분산, 표준편차, 범위(최대값-최소값), 사분위값
- 정규성: 데이터의 분포 모양이 정규 분포인가?
 - 왜도, 첨도



통계 기초 이론: 변수 유형

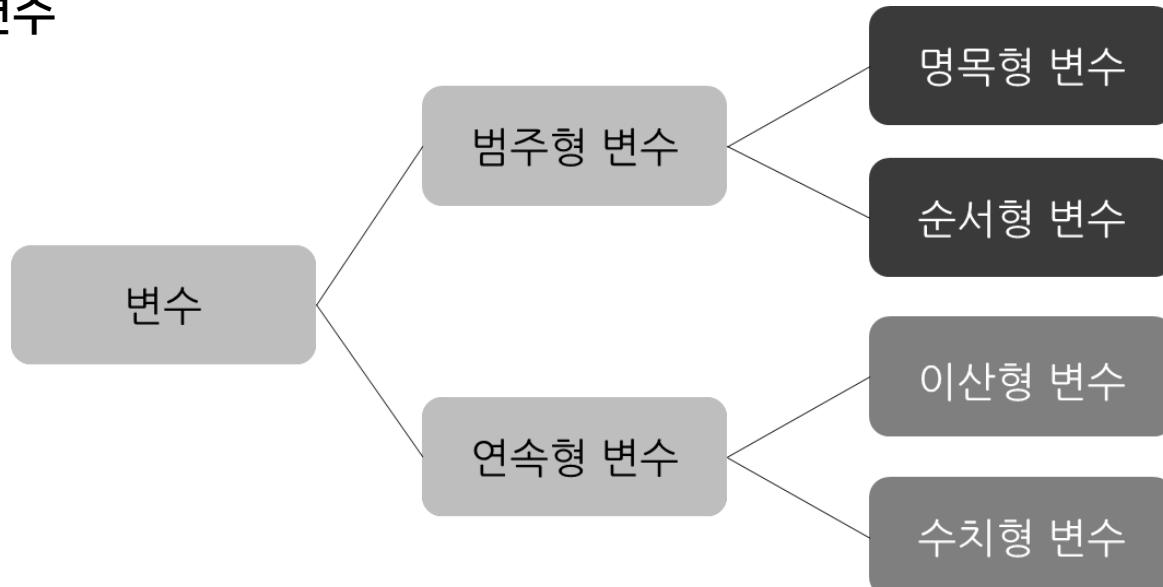
● 범주형 변수

- 연속형 데이터와 범주형 데이터의 가장 큰 차이는 가감승제가 가능하지만 의미가 없다는 것
 - 예를 들어 성별을 1(남자), 2(여자)로 구분하여 이들의 평균을 구하면 1.5가 됨

● 명목변수, 서열변수

● 연속형 변수

- 사칙연산이 가능. 예로부터 자연현상에 대해 과학적 수치를 부여하는 과정을 “수치화”라고 함
- 등간, 비율변수





통계 기초 이론: 범주형 변수 분석

- 범주형 변수의 분석의 첫 단계는 해당 변수의 **빈도(Frequency)**
- 전체데이터 빈도를 계산하는 것부터 시작
- **빈도분석**
 - 빈도는 대상 자료에서 반복하는 횟수를 기록하는 것으로 빈도분석을 통해 전반적 분포를 살펴보면서 전체 데이터의 구성을 파악 과정
 - 가장 빈도가 높은 데이터와 가장 빈도가 낮은 데이터를 기록하고 데이터 분석을 시작해야 함
- **교차분석**
 - 교차분석은 범주형 변수간 빈도분석을 확장하여 빈도를 교차시킨 **분할표(Contingency Table)**를 만들어 분석하는 방법
 - 분할표는 **교차분석표(Cross table)** 라고도 함



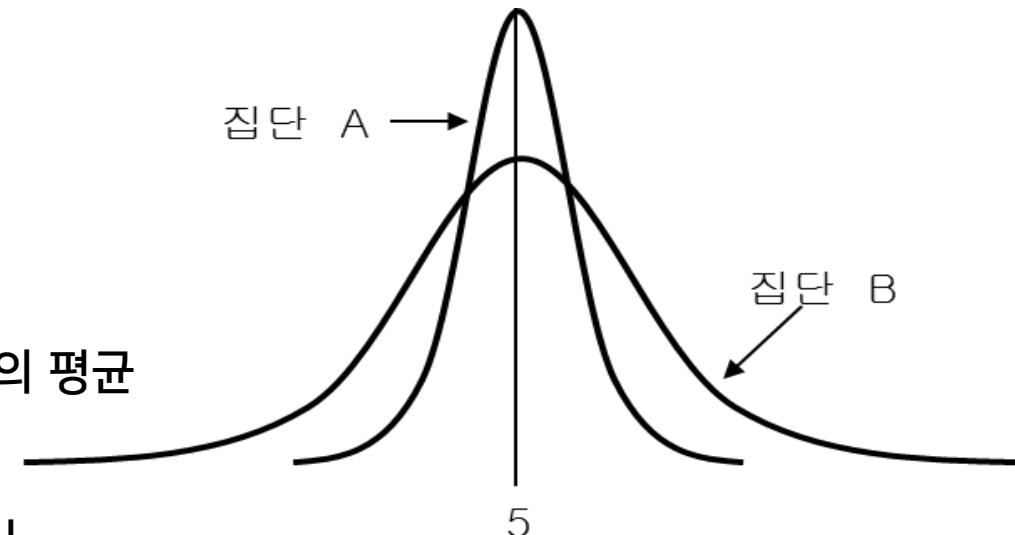
통계 기초 이론: 연속형 변수 분석

● 데이터 중심 이해

- 데이터의 중심은 여러 자료들의 비교하였을 경우 중앙에 위치하는 값으로 자료의 특성을 보여줌
- 평균, 중위수

● 데이터 퍼짐 정도 이해

- **분산**: 데이터 퍼짐 정도
- **표준편차**: 데이터의 퍼짐 정도를 동일 한 기준을 적용하기 위하여 편차 제곱합의 평균
- **범위**: 데이터의 흩어진 범위
- **사분위**: 데이터의 분포가 좌우대칭이 아니거나 이상치가 있는 경우 평균은 극단적으로 치우친 대표성이 없는 값에 의해 영향을 받음
 - 이 경우 자료를 나열하여(시각화) 전체 데이터를 파악할 수 있음
 - 자료를 순서대로 나열했을 때 50%에 위치하는 수가 중위수(Q2)이고 25%에 위치하는 수가 Q1





통계 기초 이론: 변수 종류에 따른 통계 분석 종류

● 종속변수와 독립변수의 종류에 따라 통계 분석법의 종류

Y 종속변수 (반응변수)	X 독립변수(설명변수)	통계분석법	귀무가설
연속형	범주형(2개 범주)	T-검정, paired T-검정	집단 간 평균이 동일
연속형	범주형(3개 이상)	분산분석(ANOVA)	집단 간 평균이 동일
연속형	연속형	회귀분석	회귀 계수 = 0
연속형	혼합(수치형+범주형)	공분산분석(ANCOVA)	
범주형	범주형	χ^2 검정/로짓분석	집단 간 연관성이 없음/ 회귀 계수 = 0
범주형	혼합(수치형+범주형)	로짓분석	
생존시간	혼합(수치형+범주형)	생존분석	



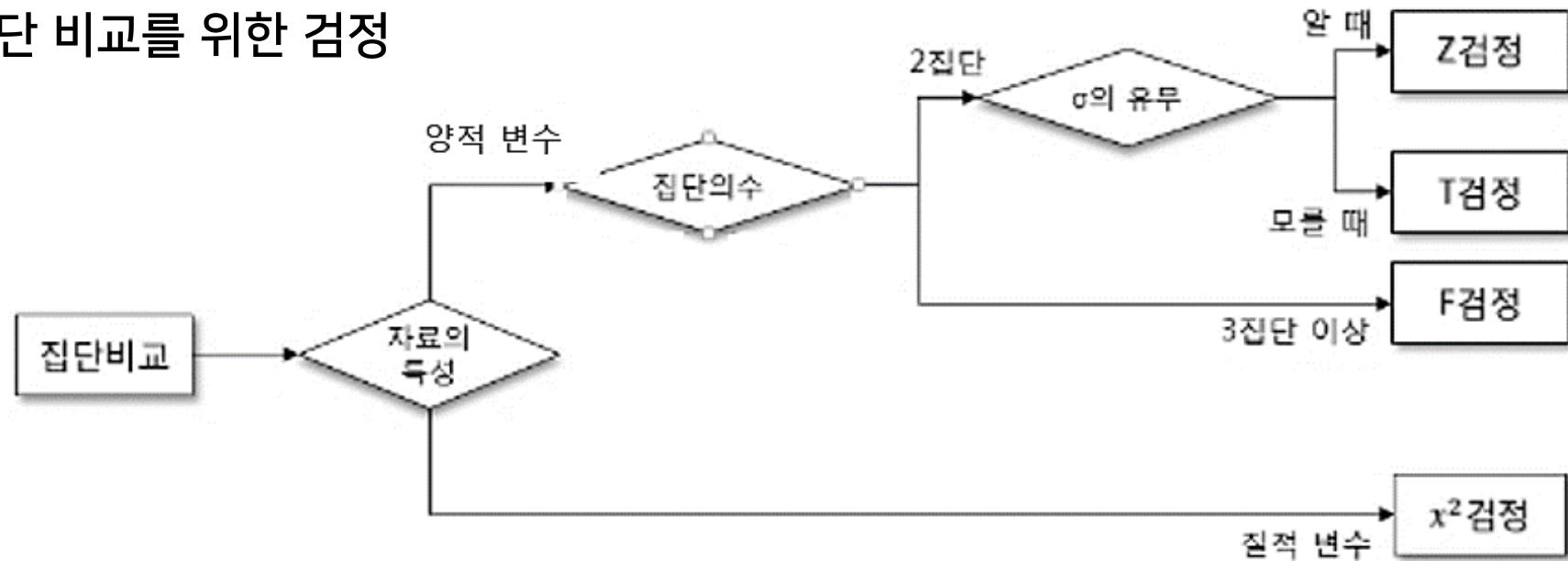
통계 기초 이론: 분석 모형 선택 기준

- 분석하고자 하는 내용(주제) 즉, “가설”을 명확히 규명
- 가설에 맞는 데이터의 변수 척도를 파악 → 변수 척도에 따라 적용할 모형이 정해짐
- 가설 검정의 구분
 - 차이 검정: 집단 간 평균 차이
 - 관계 검정: 함수적 관계 규명
- 종속 변수의 종류가 연속형과 이산형에 따라 분석 모형을 결정
 - 종속변수 Y 가 연속형일 때
 - 차이 : T-test, ANOVA(분산분석)
 - 관계 : Regression (회귀분석)
 - 종속변수 Y 가 이산형일 때
 - 차이 : Chi-square Independence Test (카이제곱 독립성검정)
 - 관계 : Logistic Regression (로지스틱 회귀분석)



통계 기초 이론: 차이 분석

- 집단 비교를 위한 검정



집단	대상
단일 표본 비교	하나의 집단(단일 표본 - one sample)중 관심 있는 연속형 변수의 모평균이 어떤 특정 값과 같은지 알아보고자 할 때
두 집단 간 평균비교	독립 표본 T : 서로 독립인 두 표본에 의한 모평균 비교 대응 표본 T : 대응하는(Paired) 쌍에 대한 차의 모평균 검정
3 이상 집단간 평균비교	일원배치분산분석 (One-way ANOVA)

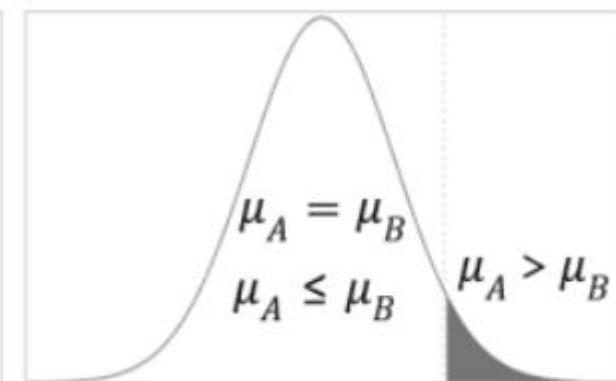
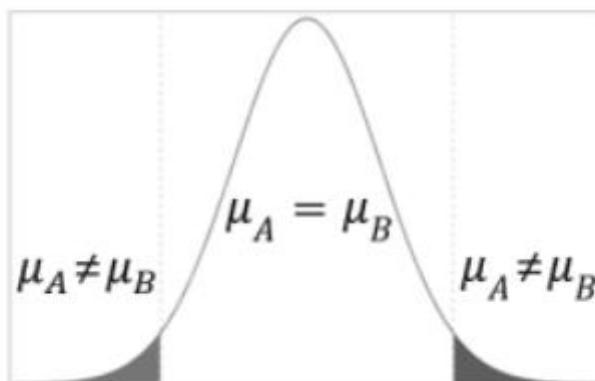
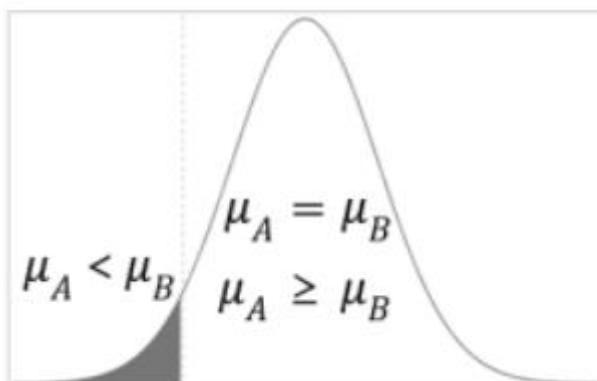


통계 기초 이론: Z-검정

- Z-검정: 분석(종속) 변수가 양적 변수, 정규 분포, 등분산, 모집단의 분산을 알고 있는 경우 단일 표본의 평균 변화를 분석하거나 두 표본의 평균 차이 분석
- 단순히 차이의 존재 여부를 떠나 **두 집단의 비교가 통계적으로 의미가 있는가를 검정**
- 표본 크기가 크면 “중심극한정리”에 의해 정규분포를 따른다고 보기 때문에 Z-검정 사용
- Z-검정 기본 가설

- 영 가설 (H_0) : 표본의 평균 차이는 없다.
- 대립가설 (H_1) : 표본의 평균 차이는 있다.

$$Z_{stat} = \frac{\bar{X} - \mu}{\frac{\sigma_X}{\sqrt{n}}}$$





통계 기초 이론: T-검정

- T-검정은 두 집단의 평균을 비교하는 통계적 검정 방법
- 모집단을 대표하는 표본으로부터 추정된 분산이나 표준편차를 가지고 검정하는 방법으로 “두 모집단의 평균간의 차이는 없다”라는 귀무가설과 “두 모집단의 평균 간에 차이가 있다”라는 대립가설 중에 하나를 선택할 수 있도록 하는 통계적 검정방법
- 단순히 차이의 존재 여부를 떠나 **두 집단의 비교**가 통계적으로 의미가 있는가를 검정
- 즉, 두 모집단의 차이가 우연에 의해서인지 아닌지를 검정하는 방법
- T분석의 기본 가설

- ✓ 가설
 - 영 가설 : 집단간의 평균 차이는 없다.
 - 대립가설 : 집단간의 평균 차이는 있다.

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

표준오차(SE)

\bar{X} : 두 집단 차이의 평균
 μ : 모집단의 평균
 S : 두 집단 차이의 표준편차

- 30개 이하의 비교적 적은 수의 표본에 대해 활용 → 30개 이상이면 정규분포의 Z검정
- 모집단의 표준편차를 알 수 없을 때 사용



추론 통계의 기초 : 일원분산분석(ANOVA)

- 일원분산분석(ANOVA): 단일 변수 세 표본 이상의 평균을 비교할 때 사용
- 각 표본은 서로 독립적이고, 정규분포를 따르며, 등분산인 경우 사용
- 기본 가설

- 영 가설 : 모든 집단의 평균은 같다.
- 대립가설 : 평균이 서로 다른 집단이 존재한다.

$$(x_{ij} - \bar{x}) = (\bar{x}_i - \bar{\bar{x}}) + (x_{ij} - \bar{x}_i)$$

여기서, x_{ij} : 개별 데이터,

\bar{x}_i : 그룹의 평균,

$\bar{\bar{x}}$: 전체 데이터의 평균

$$\sum_i \sum_j (x_{ij} - \bar{x})^2 = \sum_i \sum_j (\bar{x}_i - \bar{\bar{x}})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

총제곱합(SST)=집단간제곱합(SSA)+집단내제곱합(SSW)

● 분산분석표

요인	제곱합	자유도 (df)	평균제곱	F-value	p
집단간	SSA	r-1	집단간제곱합/자유도	집단간평균제곱/집단내평균제곱	
집단내	SSW	N-r	집단내제곱합/자유도		
합계	SST	N-1			

r: 집단수, N: 자료수



통계 기초 이론: 가설 수립

- 가설(Hypothesis)은 모수에 대한 예상, 주장, 또는 단순한 추측
 - 예) '외계인은 존재한다, 그 사람은 유죄이다' 처럼 아직은 하나의 추측인 사실을 가설이라 함
- 통계적 가설 검정은 가설에 대해 증거를 수집하고 과학적으로 증명하는 과정
- 가설 검정의 첫 단계는 가설 수립부터 시작
 - 검정하고자 하는 모집단의 모수(조사하고자 하는 자료의 평균, 분산, 표준편차, 상관계수)에 대해 항상 귀무가설과 대립가설 두 가지로 수립

영(귀무)가설(null hypothesis)	대립가설(alternative hypothesis)
H_0	H_1
기각하기를 희망하여 형식화한 가설, 기존에 받아들이던 가설, 모수에 관한 귀무가설은 항상 모수의 정확한 값을 지정하도록 진술. 반면 대립가설에서는 여러 개의 값의 가능성이 허용.	표본을 통해 입증하고자 하는 새로운 가설, 모수에 대한 관심의 영역 중에서 귀무가설로 지정되지 않은 모든 경우를 포괄적으로 지정.
외계인은 존재하지 않는다. 그 사람은 무죄이다	외계인은 존재한다 그 사람은 유죄이다

통계 기초 이론: 가설 검정 단계



구분	사례1	구분	사례2
가설 수립	<p>외계인이 존재할까? 외계인이 존재한다는 확실한 증거 수집 전까지는 외계인은 없다라고 한다.</p> <p>H_0 : 외계인 = 0. 외계인은 0 명이다. 외계인은 없다 H_1 : 외계인 ≠ 0. 외계인은 0 명이 아니다. 외계인은 있다</p>	가설 수립	<p>그 사람은 무죄일까? 유죄일까? 법정에서는 유죄 판결을 받기 전까지 모든 사람들은 무죄</p> <p>H_0 : 기존에 받아들이던 가설 죄=0. 죄가 없다. 무죄 H_1 : H_0를 기각하기를 바라는 가설 죄≠0. 죄가 있다. 유죄</p>
가설 검정	<p>외계인이 있다라는 증거가 많이 있다 ⇒ 외계인은 존재한다.</p> <p>외계인이 있다라는 증거가 조금밖에 없다 ⇒ 외계인은 존재한다는 증거가 부족하다</p>	가설 검정	<p>무죄가 아니라는 증거가 많이 있다(유죄) 무죄가 아니라는 증거가 조금밖에 없다(증거 불충분으로 무죄)</p>
증거 수집	<p>외계인은 존재하는가? 외계인이 존재한다는 객관적인 증거가 97% 있다. 증거가 95%보다 많으므로 외계인은 존재한다</p>	증거 수집	<p>그 사람은 무죄인가? 유죄인가? 유죄라는 객관적인 증거가 80% 있다 증거가 95%보다 적으므로 증거불충분으로 무죄</p>



통계 기초 이론: 오류의 이해

- 유죄라는 객관적인 증거와 외계인이 존재한다는 객관적인 증거는 얼마나 필요할까?
일반적으로 95% 정도가 필요
- 그럼 객관적인 증거가 95%보다 많으면 실제로 외계인은 존재할까(H_1)?
- 아니면 객관적인 증거가 95%보다 적으면 실제로 무죄일까(H_0)?
- 반드시 그런 것은 아니며 이것을 오류라 하며, 통계적 오류는 다음과 같이 구분

구분	영가설진실(H_0)	대립가설진실(H_1)
영가설선택(H_0)	옳은 판단 신뢰수준($1-\alpha$)	2종 오류 유의수준(β)
대립가설선택(H_1)	1종 오류 유의수준(α)	옳은 판단 신뢰수준($1-\beta$)



통계 기초 이론: 유의확률과 유의수준

● 유의확률(Significance Probability), p-value(피-밸류) 또는 P-값(피-값)

- 귀무가설이 옳다고 가정했을 때, 통계치가 관측될 확률

● 유의수준(Level of Significance)

- 제 1종 오류의 위험성을 부담하는 최대 확률로 α (알파)라고 부르며 0.05의 값으로 주로 사용
- 유의수준(α)는 항상 0.05로 고정되어 있을까? 그렇지 않다.
- 유의수준은 일반적으로 0.05를 사용하는데, 연구자의 기준에 따라서 변할 수 있다.
- 유의수준은 분석 전에 미리 결정을 하여야 함.

비교	결과
$P < \alpha = 0.05$ (오류 5% 이하, 95% 이상 진실)	H1 선택
$P \geq \alpha = 0.05$ (오류 5% 이상, 95% 이하 진실)	H0 선택



통계 기초 이론: 통계 결과 해석

- 결과 해석은 유의수준과 유의확률을 비교하여 결정
- 유의수준은 제1종 오류의 최대 허용 한계
 - 유의수준 α 값이 작아지면(오류가 작아진다) 영가설이 틀렸다는 결론을 내리기 어려움
 - 반대로 유의수준 α 값이 커지면 귀무가설이 틀렸다는 결론을 내리기 쉬워짐
- 유의확률(검정통계량)은 P 값이나 P-value라고 함
- 영가설(H_0)이 맞을 경우, 대립가설 쪽의 값이 나올 확률이 얼마나 되는지를 나타내는 값으로 결론적으로 통계는 유의확률 P와 유의수준 α 를 비교하여 영가설과 대립가설을 선택하는 과정
- 대부분 통계는 새로운 가설을 선택하는게 목적, 주로 판정 기준을 이렇게 표현

$P < 0.05$ 이하 기준 이면 새로운 대립가설(H_1)을 선택



통계 기초 이론: 통계 절차 요약

- ① 통계분석방법 선정
- ② 분석하고자 하는 목적에 따른 귀무가설(영가설)과 대립가설 설정
- ③ 분석도구(Excel, SPSS, R, Python programming 등) 검정통계량 실행 및 확인
- ④ 유의수준(α) 결정 : 0.1, 0.05, 0.01
- ⑤ 유의확률(P) 확인
- ⑥ 유의확률과 유의수준 비교 ($< \alpha$)
- ⑦ 귀무가설 과 대립가설 선택
- ⑧ 분석 결론

상관 분석





통계 기초 이론: 상관분석

- 두 연속형 변수 사이 상관관계가 존재하는지를 파악하고, 상관관계의 정도를 확인 하는 것이 상관분석(Correlation analysis)이라 함
- 상관분석에서는 관련성을 파악하는 지표로 상관계수(Correlation coefficient)라는 통계학적 관점에서 선형적 상관도를 확인하여 정도를 파악
- 상관분석은 간단한 분석이지만 머신러닝의 기반이 됨

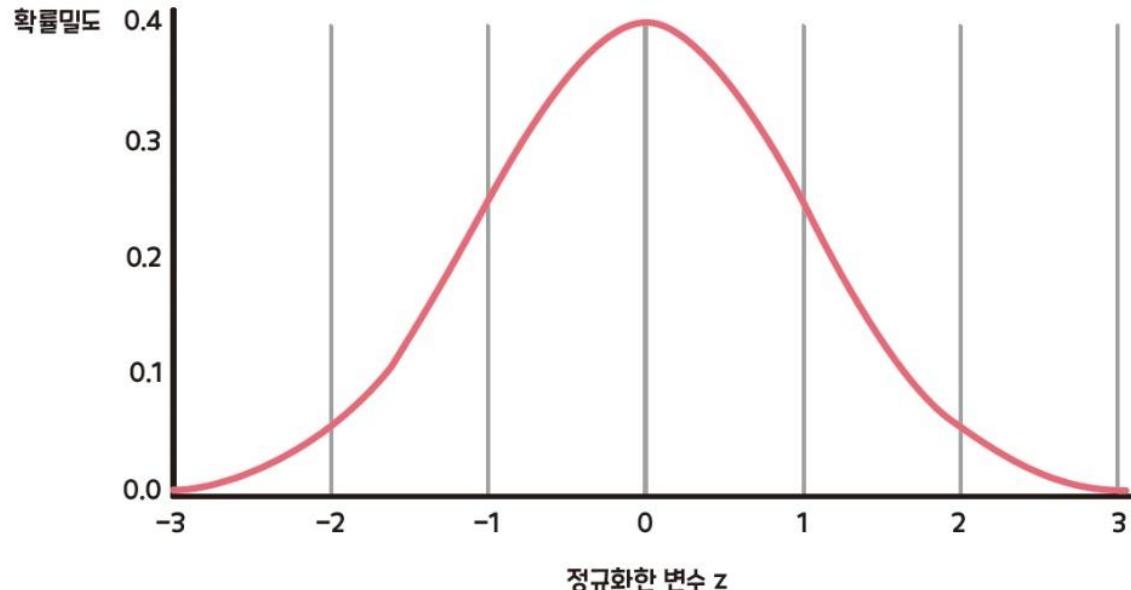
- ① 산점도(Scatter) 두 변수 상관 파악
 - ② 상관계수 확인
 - ③ 의사결정



상관관계 분석의 개념

● 상관관계 분석의 세 가지 방법

- 피어슨 상관분석(Pearson correlation analysis) : 가장 일반적인 상관분석 방법.
- 스피어만 상관분석(Spearman correlation analysis) : 두 변수가 정규성을 보이지 않을 때 사용하기 적합한 방법.
- 켄달 상관분석(Kendall correlation analysis) : 스피어만 상관분석과 비슷하나 표본 데이터가 적고 동점이 많을 때 사용하기 적합한 방법.



✓ 하나 더 알기: 상관관계 분석의 의미

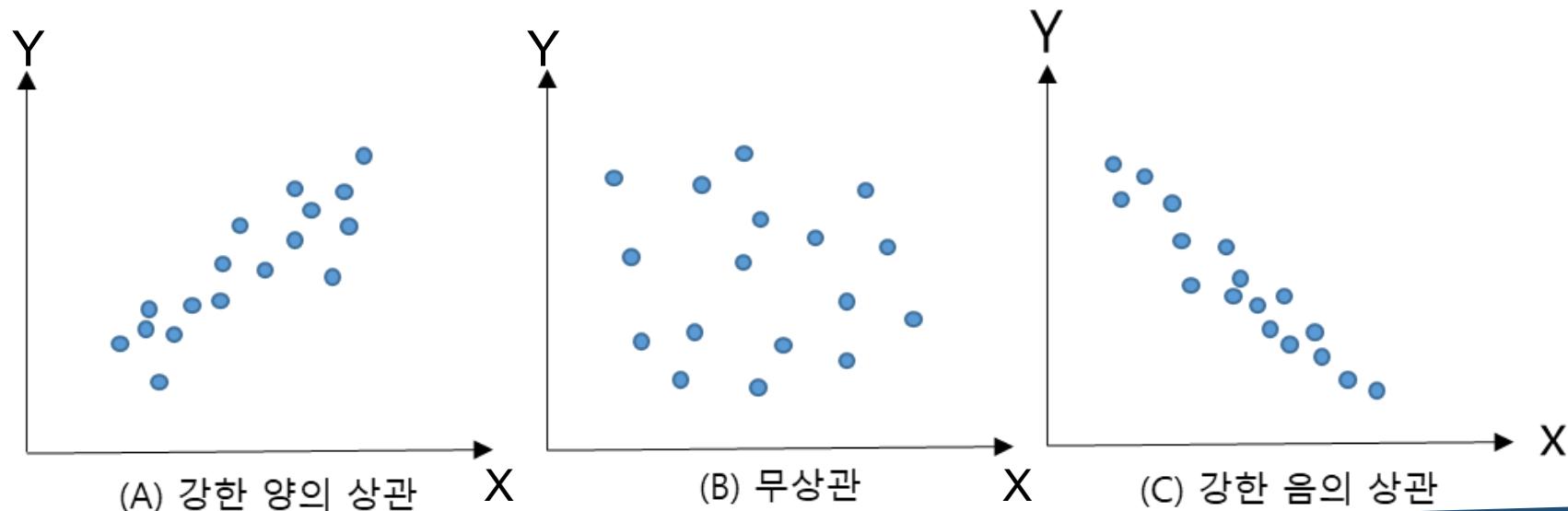
상관계수만 가지고 두 변수 사이의 상관성이 있는지 없는지 판단할 수는 없음.

두 변수에 선형 상관관계가 아닌 다른 상관관계가 있을 수 있음.



통계 기초 이론: 상관분석

상관	상관계수
양의 상관	+0.1 ~ +0.3 이면, 약한 양의 상관관계 +0.3 ~ +0.7 이면, 뚜렷한 양의 상관관계 +0.7 ~ +1.0 이면, 강한 양의 상관관계 – 그림A
무상관	-0.1 ~ +0.1 이면, 없다고 할 수 있는 상관관계 - 그림 B
음의 상관	-1.0 ~ -0.7 이면, 강한 음의 상관관계 – 그림C -0.7 ~ -0.3 이면, 뚜렷한 음의 상관관계 -0.3 ~ -0.1 이면, 약한 음의 상관관계



회귀분석

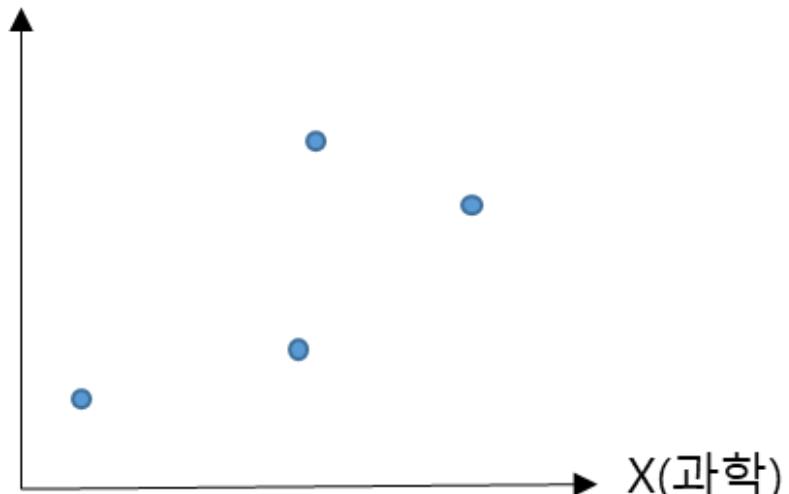




통계 기초 이론: 회귀분석

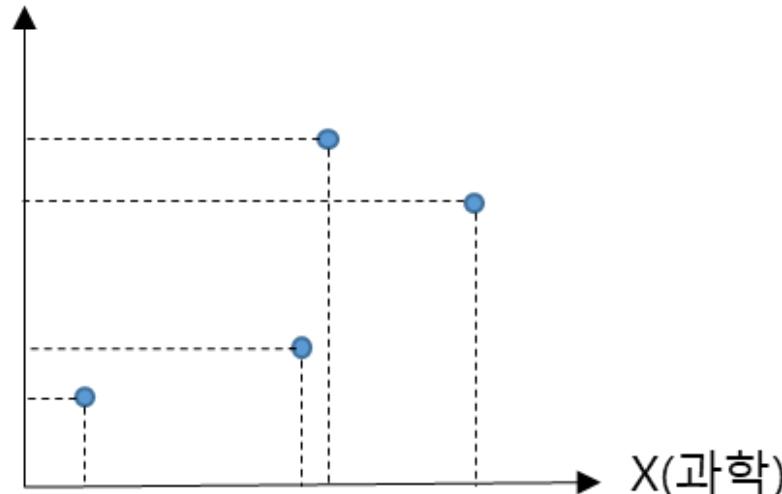
- 상관분석에서는 두 연속형 변수 X (과학)와 Y (수학)의 상관 정도만 알 수 있고 인과관계는 알 수 없었음
- 회귀분석에서는 두 연속형 변수 X 와 Y 를 **독립변수**와 **종속변수**라고 하는 **인과관계**로 설명
- ‘과학 점수가 좋으면 수학점수가 좋을까요?’ 와 같이 간단 하지만 미래를 예측할 수 있는 머신러닝의 초기 모델이 됨

Y (수학)



상관분석

Y (수학)



회귀분석

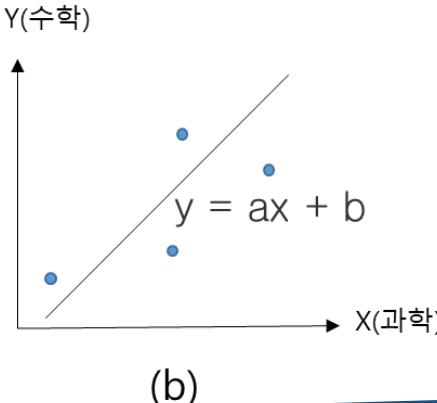
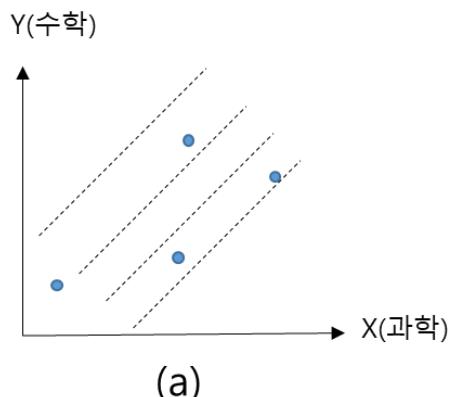


통계 기초 이론: 회귀분석

- 선형회귀분석(Linear Regression Analysis)은 쌍으로 관찰된 연속형 변수들 사이의 관계에 있어서 한 변수를 원인으로 하고 다른 변수들을 결과로 하는 분석
- 독립변수와 종속변수 사이 선형식을 구하고 그 식을 이용하여 변수값 들이 주어졌을 때 종속변수의 변수 값을 예측하는 분석방법

X	Y
독립변수, 설명변수, 원인변수	종속변수, 반응변수, 결과변수 머신러닝(클래스, 라벨)
다른 변수에 영향을 주는 원인	다른 변수에 영향을 받는 결과

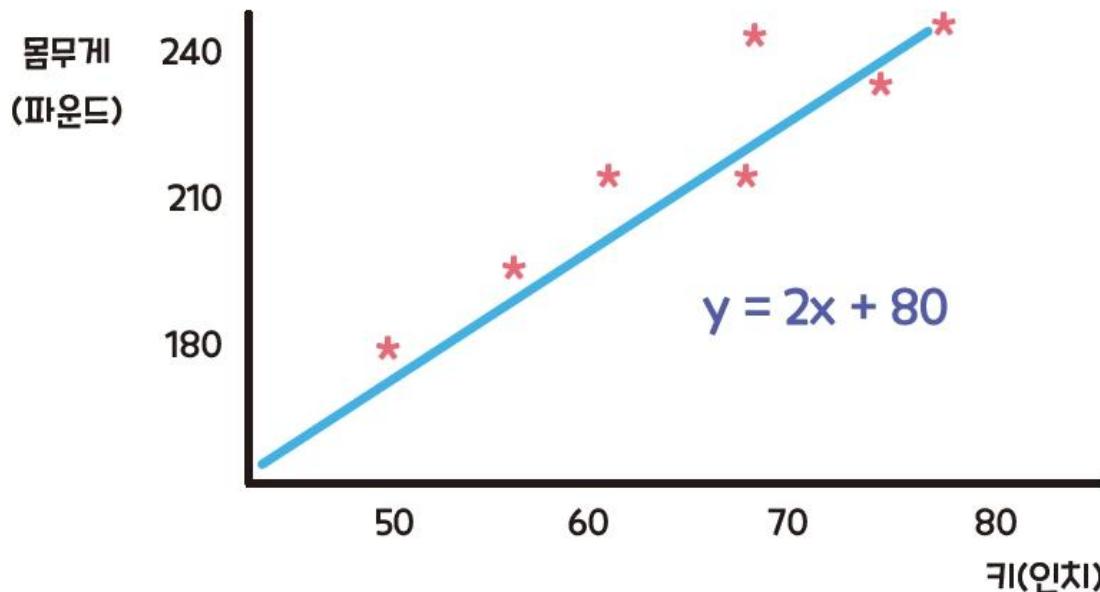
- x변수와 y변수 간의 관계를 $y = ax + b$ 와 같은 하나의 선형 관계식으로 표현
- $y = ax + b$ 인 회귀식에서 독립변수 x가 하나인 것이기에 단순선형회귀분석이라함





통계 기초 이론: 회귀분석

- 선형 회귀분석의 모형
- 선형 회귀분석(Linear regression analysis) : 두 개 또는 그 이상의 변수 간 인과관계의 패턴을 원래 모습과 가장 가깝게 추정하는 분석 방법.
 - 함수 $y = 2x + 80$ 의 그래프, 이를 선형 회귀분석의 모형(Model)이라고 부름.
 - 선형 회귀분석은 x 변수가 원인, y 변수가 결과로 인과관계여야 한다는 조건 있음.

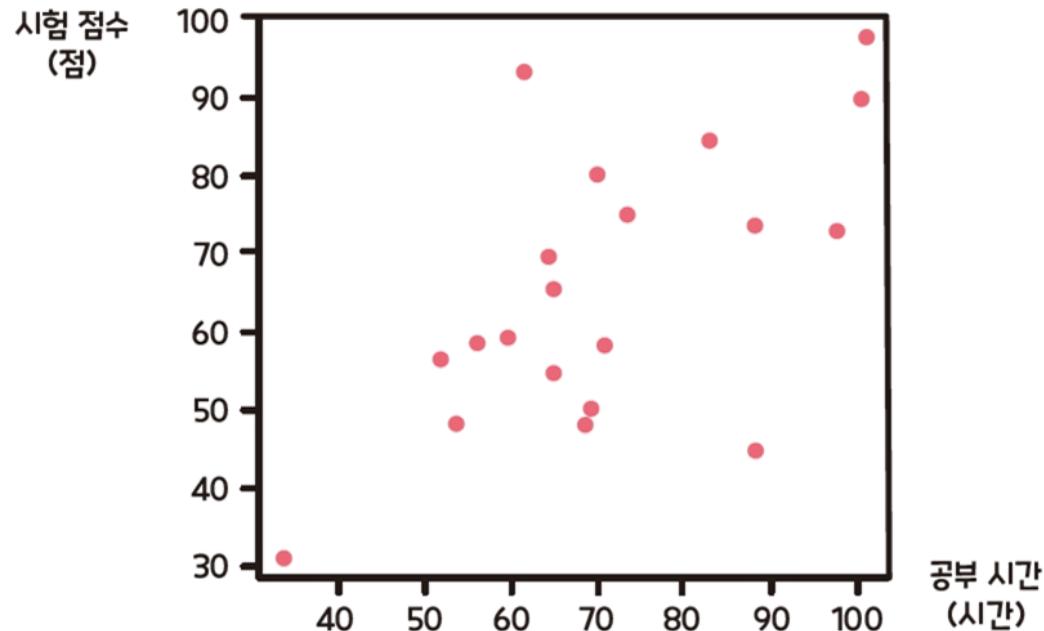




통계 기초 이론: 회귀분석

● 선형 회귀분석의 모형

- 회귀분석에서 원인인 x 변수는 독립변수(Independent variable), 결과인 y 변수는 종속변수 (Dependent variable)라고 부름.
- 어떠한 결과의 원인이 되는 독립변수가 한 개일 때 단순 선형 회귀분석(Simple linear regression analysis), 두 개 이상이면 다중 선형 회귀분석(Multiple linear regression analysis).
- 공부 시간과 시험 점수의 인과관계를 분석해보면 종속변수는 시험 점수이며 독립변수는 공부 시간만 고려.



통계 기초 이론: 회귀분석



- 선형 회귀분석의 모형

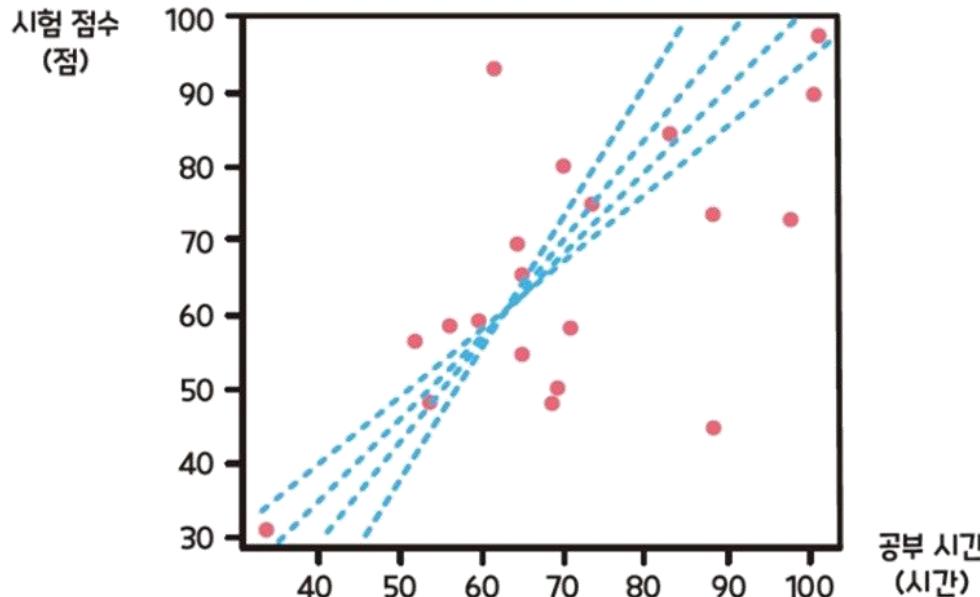
- 점들을 직선 하나로 표현하고자 함.
- 직선을 일차함수 $y = mx + b$ 로 표현한다면 계수 m 은 이 직선의 기울기이므로 양수일 것
- 일반적인 선형 회귀모형은 다음 식과 같이 독립변수 x_i 앞에 계수 β_i 가 붙음.

마지막 ε 은 오차 항

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \varepsilon$$

$\beta_i : x_i$ 의 계수 오차

- 데이터를 표현하는 직선은 여러 개 존재할 수 있음.

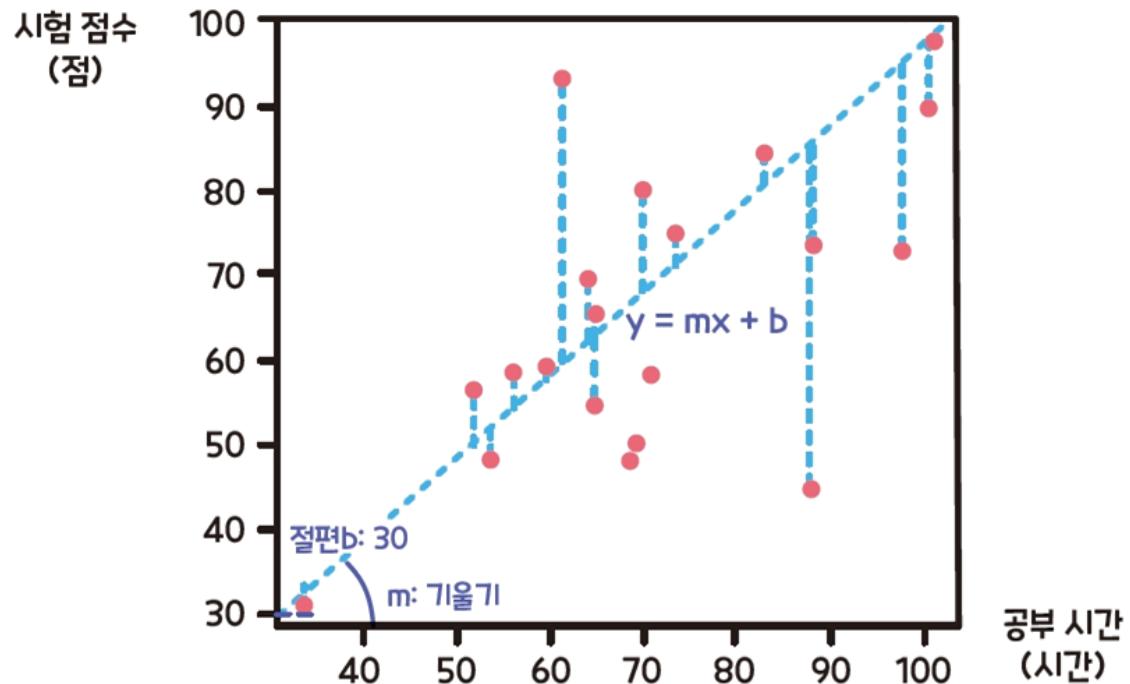




통계 기초 이론: 회귀분석

● 선형 회귀분석의 모형

- 최대한 많은 점과 거리가 가까운 직선이 좋은 직선.
- 점에서 직선까지 y 축과 평행한 선분을 그렸을 때 모든 선분 길이의 합을 최소로 하는 직선 찾는 것.
- 점이 가리키는 값과 직선이 예측하는 값의 차를 잔차(Residual)라고 부름.





통계 기초 이론: 회귀분석

● 선형 회귀분석의 모형

✓ 하나 더 알기: 더미 변수

더미 변수(Dummy variable)는 독립변수를 0과 1로 변환하여 '예'와 '아니오'로 나타낼 수 있는 변수.

더미 변수를 여러 개 두면 '예'와 '아니오'만으로 결과를 세 가지 이상으로 구분할 수 있음.

어린이, 청소년, 성인으로 구분.

- ① 데이터가 어린이일 때 '어린이' 더미 변수는 1이고 '청소년' 더미 변수는 0.
- ② 청소년일 때 '어린이' 더미 변수는 0이고 '청소년' 더미 변수는 1.
- ③ 어린이도 청소년도 아닌 성인은 두 더미 변수에 모두 0을 대입하여 표현할 수 있음.

따라서 구분하고자 하는 데이터의 종류가 N개일 때 더미 변수 N-1개를 선언하면 됨.

	어린이 더미 변수	청소년 더미 변수
어린이	1	0
청소년	0	1
성인	0	0



통계 기초 이론: 회귀분석

- 선형 회귀분석의 모형

- 결정계수

- 선형 회귀분석에서 모형이 데이터의 패턴을 얼마나 효과적으로 보여주는지 수치화한 값
- 결정계수 R²(R square)를 다음 수식과 같이 정의.

$$R^2 = \frac{(Q - Q_e)}{Q}$$

Q = 전체 데이터의 편차 제곱의 합

Q_e = 전체 데이터의 잔차 제곱의 합

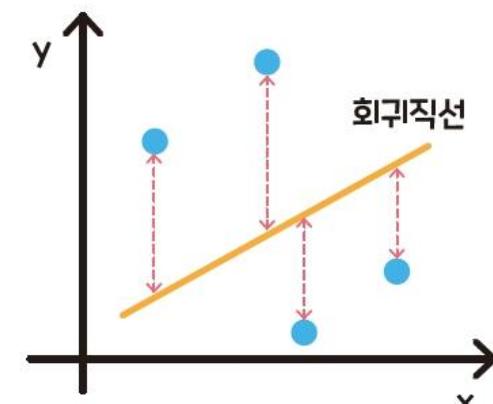
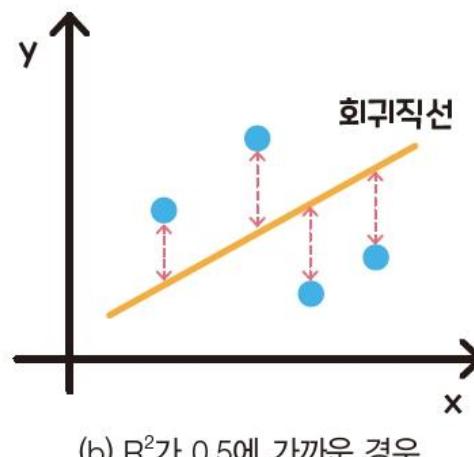
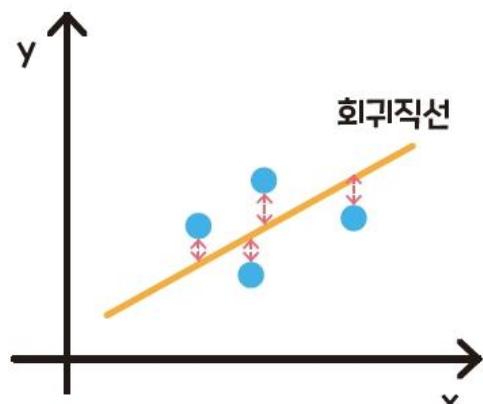


통계 기초 이론: 회귀분석

● 선형 회귀분석의 모형

• 결정계수

- 편차(Deviation)는 평균과 실제 값의 차이.
- 결정계수 R^2 는 0 이상 1 이하의 값으로 계산됨.
- R^2 값이 1에 가까울 때 잔차가 작고 예측의 정밀도가 높음.
반면 R^2 값이 0에 가까울 때 잔차가 커 예측의 정밀도가 낮음.





통계 기초 이론: 회귀분석

● 선형 회귀분석의 모형

● 수정된 결정계수

- 결정계수 R^2 는 독립변수의 개수가 많을수록 커지는 경향을 보임.
- 이러한 문제를 해결하기 위해 다중 선형 회귀분석에서
수정된 결정계수 adj. R^2 (adjusted R square)로 설명력을 나타냄.

$$adj. R^2 = \frac{(n - 1)}{(n - p - 1)(1 - R^2)}$$

R^2 = 결정계수 p = 독립변수의 개수 n = 표본 수



통계 기초 이론: 회귀분석

- 선형 회귀분석의 해석

- 통계적 가설검정

- 모집단에 대한 추측을 하고 표본의 정보를 기준으로 가설이 타당한지 판정하는 방법.
- 통계적 가설에는 두 종류가 있음.

통계학에서 처음부터 거짓일 것으로 기대하는 가설인 귀무가설(Null hypothesis),
입증하고자 하는 가설인 대립가설(Alternative hypothesis).

- 실험 결과를 보고 귀무가설을 채택하거나 대립가설을 채택하는 기준을 세워 두어야 하고, 그 기준을 **유의수준(Significance level)**이라고 함.
- 귀무가설이 참일 때 실제 결과가 실험 결과와 같을 확률은 p-값(p-value, 유의확률)이라고 부름.



통계 기초 이론: 회귀분석

II. 선형 회귀분석의 해석

● 선형 회귀분석 과정

- 선형 회귀분석을 수행하고 해석하는 과정은 크게 세 단계
 - ① 결과인 종속변수를 y 로 두고, 원인이 되는 독립변수를 x_i 로 둠.
 - ② 설명력 R^2 또는 $adj.R^2$ 값을 확인. 결정계수가 0.6 또는 0.4 이상이면 해당 회귀모형이 설명력을 갖추었다고 인정.
 - ③ 각 독립변수의 p -값이 유의수준보다 작은지 확인. p -값이 유의수준 이상인 변수를 제외하고 남은 독립변수가 결과에 영향을 주는 원인.



Excel의 회귀분석

- 회귀분석(단순회귀분석)

- 두 변수중 독립 변수('x')의 알려진 값을 기반으로 종속 변수('y')의 값을 예측
- 선형회귀식 : $y = a + bx$ (a 는 절편, b 는 기울기)
- 결정계수 (R^2) : 독립변수가 종속변수를 얼마나 잘 설명하는가를 나타내는 수치
 - 상관계수의 제곱 : 0 ~ 1 사이의 값으로 1에 가까울 수록 잘 설명하고 있다하고 해석함

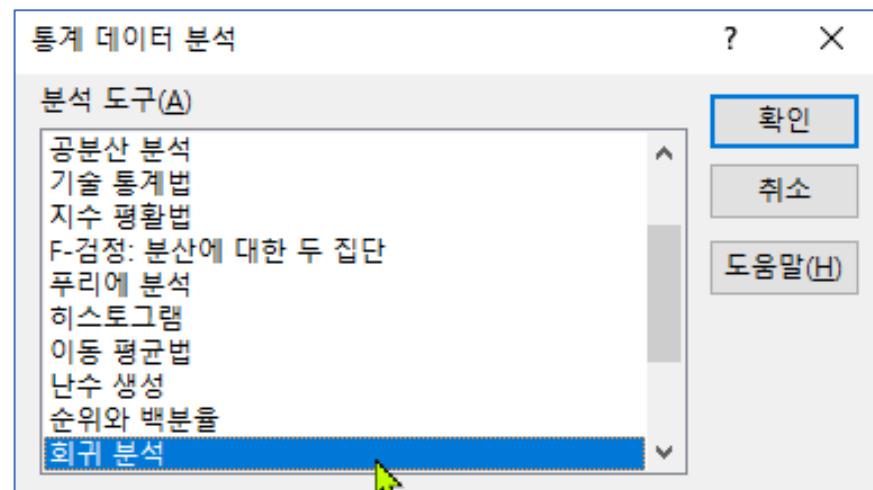
- 분석할 데이터는 반드시 연속된 '숫자'이어야 함

- 적절한 가설(설명변수로 결과변수를 설명하는 가설)을 수립

- 상관계수의 높음이 두 변수의 인과관계가 있다고 해석하면 안됨

- Excel의 메뉴

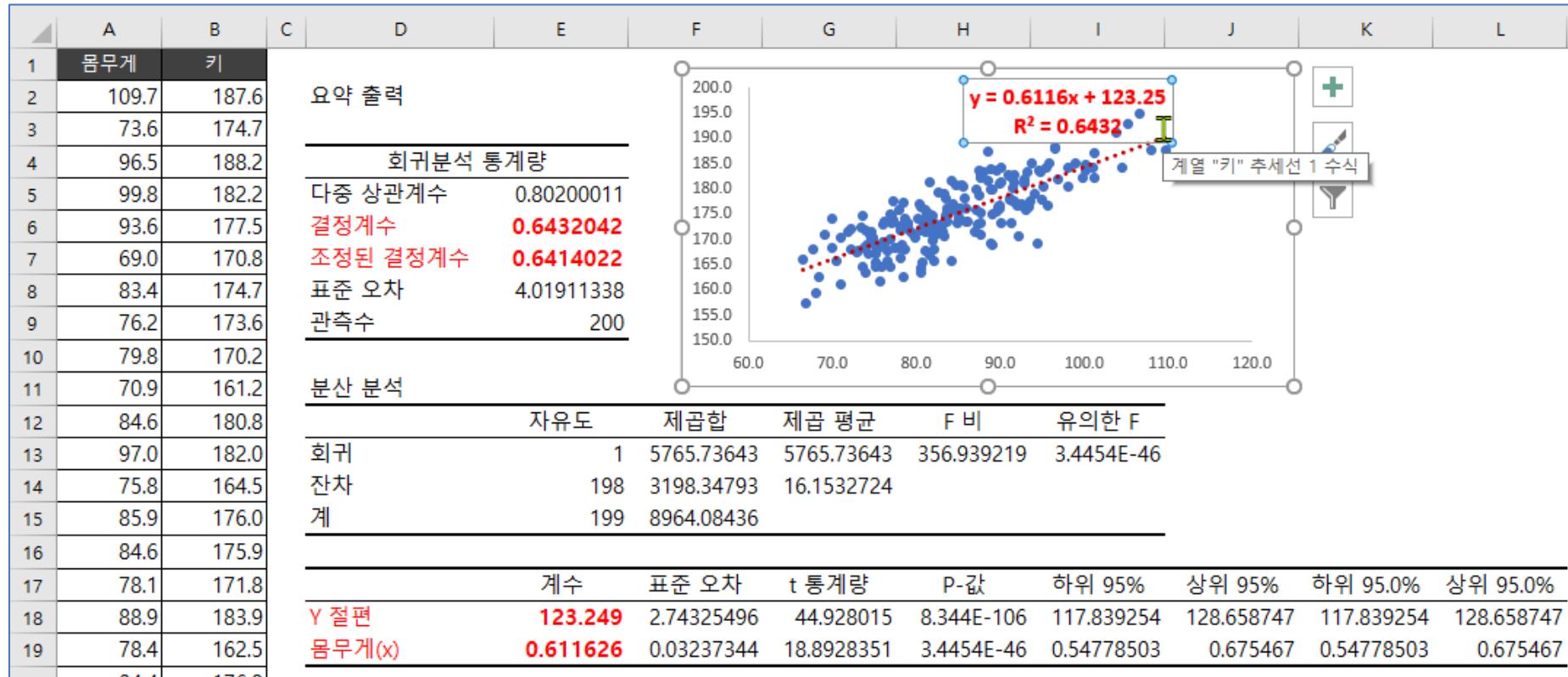
- [데이터]메뉴탭 → [데이터분석]메뉴
- 분산형차트의 추세선과 수식표시
- 상관관계 함수: CORREL(), LINEST()등





분산형 차트와 추세선 그리고 회귀식의 표시

- [데이터분석]메뉴의 [회귀분석] 결과와 분산형 차트의 추세선 식과 결정계수





통계 기초 이론: 로지스틱 회귀분석

● 로지스틱 회귀모형

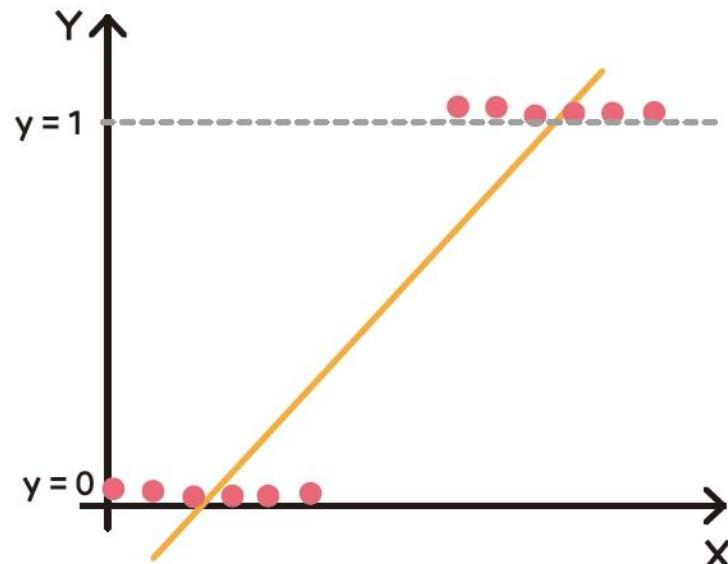
- 로지스틱 회귀분석(Logistic regression analysis)은 결과인 종속변수에 미치는 요인들을 독립변수로 두고 각 독립변수의 영향을 설명.
- 로지스틱 회귀분석의 종속변수는 범위에 제한이 있음.
로지스틱 회귀분석의 종속변수는 0에서 1사이의 값.

통계 기초 이론: 로지스틱 회귀분석

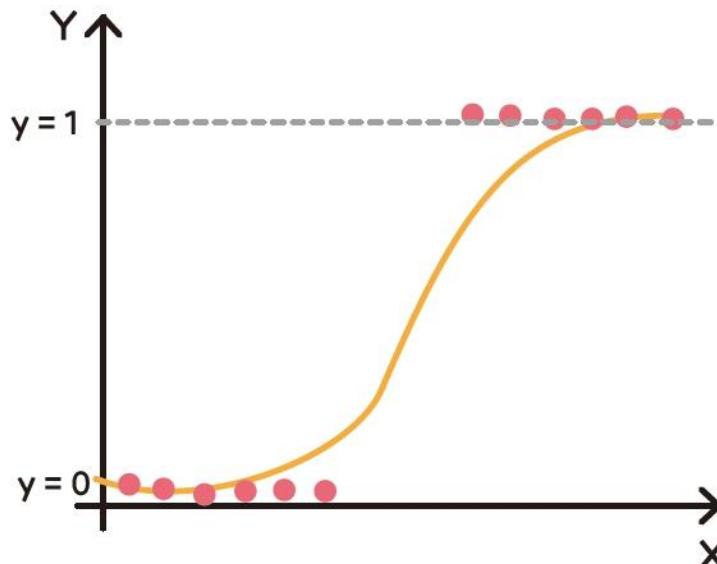


● 로지스틱 회귀모형

- 선형 회귀분석과 로지스틱 회귀분석의 모형.
- 왼쪽의 선형 회귀분석은 표본 데이터를 직선으로 그룹화.
- 반면 오른쪽의 로지스틱 회귀분석은 표본 데이터가 0에서 1사이의 값으로 그룹화되어 있음.



(a) 선형 회귀분석



(b) 로지스틱 회귀분석



통계 기초 이론: 로지스틱 회귀분석

● 로지스틱 회귀모형

- 로지스틱 회귀분석 결과 오즈비를 얻음. [오즈비\(Odds Ratio, OR\)](#)는 우리말로 [승산비](#).
- 사건이 발생할 확률을 p라고 할 때, 오즈비를 수식으로 나타내면 다음과 같음.

$$OR = \frac{\text{사건이 발생할 확률}}{\text{사건이 발생하지 않을 확률}} = \frac{p}{1-p}$$

- 대학 합격이라는 사건에서 합격을 1, 불합격을 0으로 정했을 때 합격할 확률이 0.80라면 오즈비는 4. 이는 대학에 합격할 확률이 불합격할 확률보다 4배 높다는 뜻.

$$OR = \frac{p}{1-p} = \frac{0.8}{1-0.8} = 4$$



통계 기초 이론: 로지스틱 회귀분석

● 로지스틱 회귀분석의 해석

● 로지스틱 회귀분석의 해석 과정은 다음과 같음.

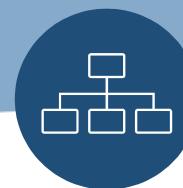
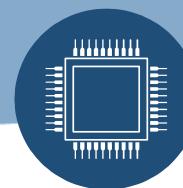
- 첫째, 선형 회귀분석과 마찬가지로 각 독립변수의 p-값을 확인함. p-값이 유의수준보다 작은 독립변수를 통계적으로 유의한 변수라고 판단.
- 둘째, 오즈비를 구해 각 독립변수가 종속변수를 1로 만들 확률을 비교.
- 사과 가격이 사과 판매 여부에 미치는 영향을 분석해보기. 종속변수는 사과 판매 여부로, 사과가 판매되면 1이고 판매되지 않으면 0. 독립변수는 사과 가격.

사과 판매 여부	1	1	1	1	1	1	1	1	1
가격(원)	1,500	2,000	5,000	3,000	3,500	2,500	4,000	4,500	3,000
사과 판매 여부	0	0	0	0	0	0	0	0	
가격(원)	4,500	4,000	4,500	5,500	6,500	5,000	3,500	7,000	



Thank you!

See you next time.



담당교수 : 유 현 주
comjoo@uok.ac.kr

