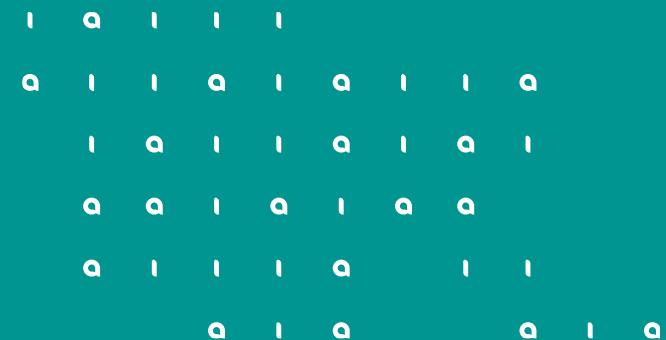


## [하나은행] Digital hana 路 금융서비스개발

# 은행의 업무와 데이터 분석 실제



# 목차



01	금융 산업과 은행 서비스	7p
02	비은행 금융과 은행 서비스	34p
03	금융업의 디지털 전환	50p
04	금융 데이터 분석을 위한 재무 기초	88p
05	데이터 분석 기본기	123p
06	데이터 획득과 데이터 전처리	159p
07	시각화와 기술 통계량을 활용한 사전 데이터 분석	185p
08	통계적 가설검정 방법론과 데이터 분석	214p

# 교과목 교육과정 구성

교과목	세부 주제	학습목표	학습내용	학습방법	학습시간
은행의 업무와 데이터 분석 실제	금융 산업과 은행 서비스	경제 메커니즘 상에서 금융산업의 위치와 역할을 이해하고, 그 중심에 있는 은행업의 역할과 기능을 이해한다.	<ul style="list-style-type: none"><li>금융업에 대한 이해</li><li>은행업 및 은행 금융상품</li><li>은행 서비스에 따른 업무 영역과 역할</li></ul>	이론	2.5
	비은행 금융과 은행 서비스	증권, 보험, 카드 등 비은행 서비스의 역할 및 기능을 이해하고, 은행 서비스와의 연관성을 이해 한다.	<ul style="list-style-type: none"><li>비은행 금융업의 이해</li><li>비은행 금융상품의 연관성과 판매 과정</li><li>고객 자산관리 서비스의 이해</li></ul>	이론	1.5
	금융업의 디지털 전환	ICT 기술의 발전에 따른 은행 및 금융업의 발전 과정을 복기하고, 비금융 서비스까지 결합된 혁신의 기회를 모색한다.	<ul style="list-style-type: none"><li>ICT 기술과 산업 발전의 과정</li><li>핀테크 금융 서비스 변화에 영향을 준 디지털 전환 양상</li><li>빅데이터 시대와 마이 데이터</li></ul>	이론	2
	금융 데이터 분석을 위한 재무 기초	금융 활동의 기본이 되는 재무 이론을, 금융 데이터 분석의 관점으로 연계하여 학습한다.	<ul style="list-style-type: none"><li>이자율과 수익률의 이해</li><li>통계적 데이터 분석과 재무 이론</li></ul>	이론	2

# 교과목 교육과정 구성

교과목	세부 주제	학습목표	학습내용	학습방법	학습시간
운행의 업무와 데이터 분석 실제	데이터 분석 기본기	데이터 처리에 기본이 되는 관계형 데이터 모델을 이해하고, 엑셀 기반으로 데이터 분석 프로세스 와 메커니즘을 이해한다.	<ul style="list-style-type: none"><li>데이터 분석 프로세스와 관계형 모델</li><li>엑셀 데이터 처리 기본기</li></ul>	이론/예제	2
	데이터 획득과 데이터 전처리	엑셀 기반으로 데이터를 획득하고, 원 데이터를 데이터 분석에 적합한 파생 데이터로 가공하는 방법을 학습한다.	<ul style="list-style-type: none"><li>데이터 획득법과 데이터 형태</li><li>파워쿼리를 활용한 데이터 입력</li><li>금융 데이터 전처리 방법론</li></ul>	이론/예제	2
	시각화와 기술 통계량을 활용한 사전 데이터 분석	차트를 통해 데이터를 시각화 하고, 기술 통계량을 통해 분포를 파악하여 빠르게 패턴을 찾는 방법을 학습한다.	<ul style="list-style-type: none"><li>시각화를 활용한 사전 데이터 분석</li><li>기술 통계량을 활용한 사전 데이터 분석</li></ul>	이론/예제	1.5
	통계적 가설검정 방법론과 데이터 분석	통계적 가설검정법을 사용하여 체계적으로 데이터를 분석하고, 찾아낸 패턴과 오류 가능성을 바탕으로 실제에 적용할 수 있는 인사이트를 함양한다.	<ul style="list-style-type: none"><li>통계적 가설 검정 방법론</li><li>데이터 분석 도구의 활용</li><li>상관관계와 회귀분석법</li></ul>	이론/예제	2.5
총계					16

## 교과목 교육과정 구성

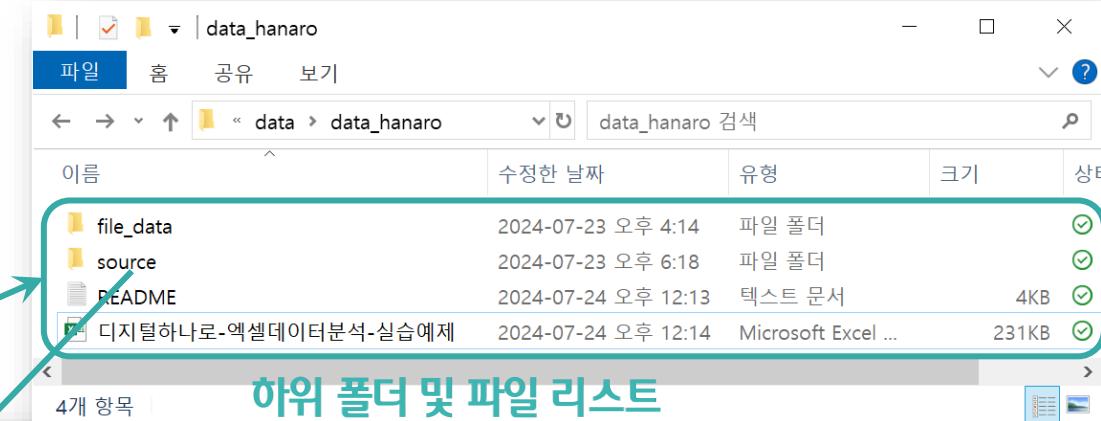
## 학습자료 사용 방법

## 강의 자료 :

[https://snowballproject.co.kr/static/files/dev\\_hanaro.pdf](https://snowballproject.co.kr/static/files/dev_hanaro.pdf)

## 실습 예제 :

[https://snowballproject.co.kr/static/files/data\\_hanaro.zip](https://snowballproject.co.kr/static/files/data_hanaro.zip)



폴더	폴더 설명	파일명	파일 설명
data_hanaro	최상위 폴더	README.txt	이 자료에 대한 설명 및 저작권
		*디지털하나로-엑셀데이터분석-실습예제.xlsx	메인 실습예제 파일
+ file_data	Ch.6 데이터 불러오기 예제 파일	APT서울매매지수.csv APT서울매매지수.json APT서울매매지수.txt APT서울매매지수.xlsx APT서울매매지수.xml xml_map.xml	CSV 파일 불러오기 예제 JSON 파일 불러오기 예제 TXT 파일 불러오기 예제 파일 불러오기 예제 엑셀 데이터 XML 파일 불러오기 예제 XML 데이터구조 정의 파일
+ source	사용한 원데이터	*디지털하나로-엑셀데이터분석-실습예제완성.xlsx	메인 실습예제 완성 파일 (참고용)
+ DB	Ch.6 DB 시연	*yt_clip_stat.xlsx	Ch.6 DB 접속 시연한 엑셀 파일 (데이터 갱신 불가)
+ KB부동산지수	실습예제 원데이터	Data-KB부동산통계-202406_월간시계열.xlsx	실습용 부동산 가격 자료 원데이터
+ 기타비실습	비실습자료 원데이터	12_04_06_E_과속방지턱정보.xlsx	Ch.6 과속방지턱 예제 원데이터
+ 네이버	Ch.6 웹데이터	네이버증권_시가총액상위50종목.xlsxm	Ch.6 웹데이터 크롤링 완성 파일
+ 통계이론	이론 수업용 데이터	KOSPI.xlsx  *시뮬레이션-가격변화시계열.xls  *중심극한정리.xlsx	Ch.8 KOSPI 지수 원데이터  Ch.8 가격 시계열 모델 시뮬레이션 예제 파일  Ch.8 중심극한정리 테스트 파일
+ 통계청	실습예제 원데이터	성_및_연령별_추계인구....xlsx	Ch.7 시각화 실습용 인구추계자료 원데이터

# ICT 기술을 적극 활용한 금융/경영 전문가

## (주)스노우볼프로젝트 대표이사 (現)

- 온라인 재무관리 컨텐츠 개발, 온/오프라인 재무 컨설팅
- (주)알파코 수석전문위원, 디지털 전환 교육 컨설팅/강사

## (주)아이로보투자자문 대표이사

- 온라인 투자자문, 투자일임 사업 개발, 자산관리 솔루션 디자인
- 핀테크산업협회 로보어드바이저 협의회장

## 세이크 칼리파 전문병원 (Sheikh Khalifa Specialty Hospital in UAE)

- Financial planning team, Strategy team, Manager
- 사업계획, 예산계획 수립 및 관리 집행, 병원장 비서 업무 및 Governance 조율

## 한화자산운용(주) 전략운용팀, Chief Manager

- 퀸트리서치에 기반한 투자전략 / 상품개발, 액티브퀀트 운용 (2014.07 기준 약 8,500억원)
- 솔루션 비즈니스 연구 : 위험배분 자산배분 모델, LDI 전략, 목표 중심 투자 등

## 한화자산운용(주) 퀸트리서치팀, 팀장

- 업계 최초 액티브퀀트 공모 주식형 펀드 출시(2007.05), 퀸트 룽숏펀드(2008.05)
- 솔루션 비즈니스 연구 : ETF 신상품 전략, Risk Parity 자산배분 모델, LDI 전략, 퇴직연금 DB형 전략 등

## 푸르덴셜자산운용(주) 데이터분석팀, 팀장

## 연세대 경영대학원 금융MBA (금융공학 트랙)

## 서울대 산업공학

▶ 강사 이진수



# 1. 금융 산업과 은행 서비스

---

# 학습목표

- ▶ 실물경제와 금융경제의 관계를 통한 금융업 이해
- ▶ 금융경제 중심인 은행업을 은행상품 관점으로 이해
- ▶ 은행 서비스에 따른 업무 영역과 역할을 이해

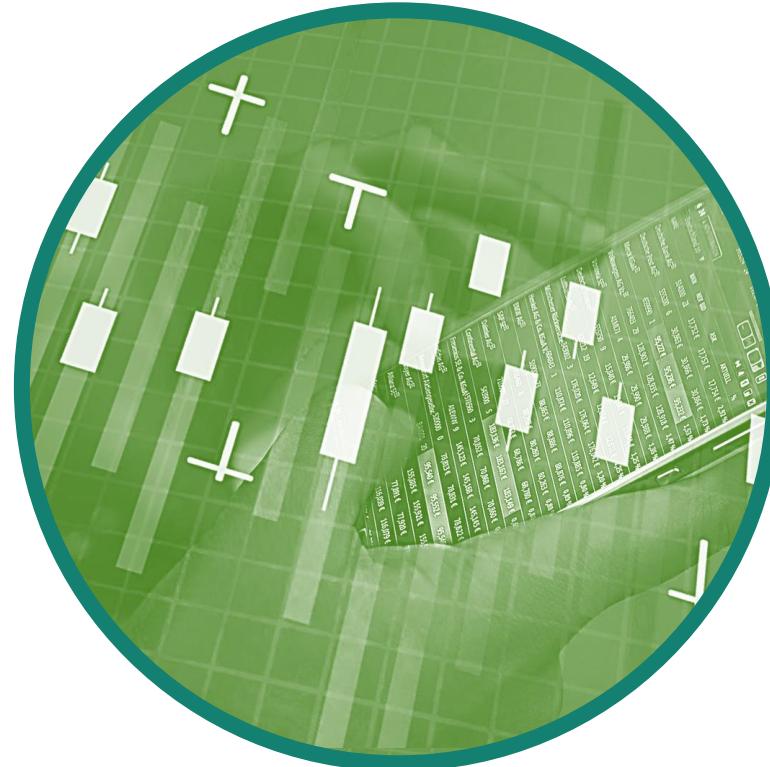
1. 금융의 역할

실물경제



돈의 대가는 재화/서비스

금융경제

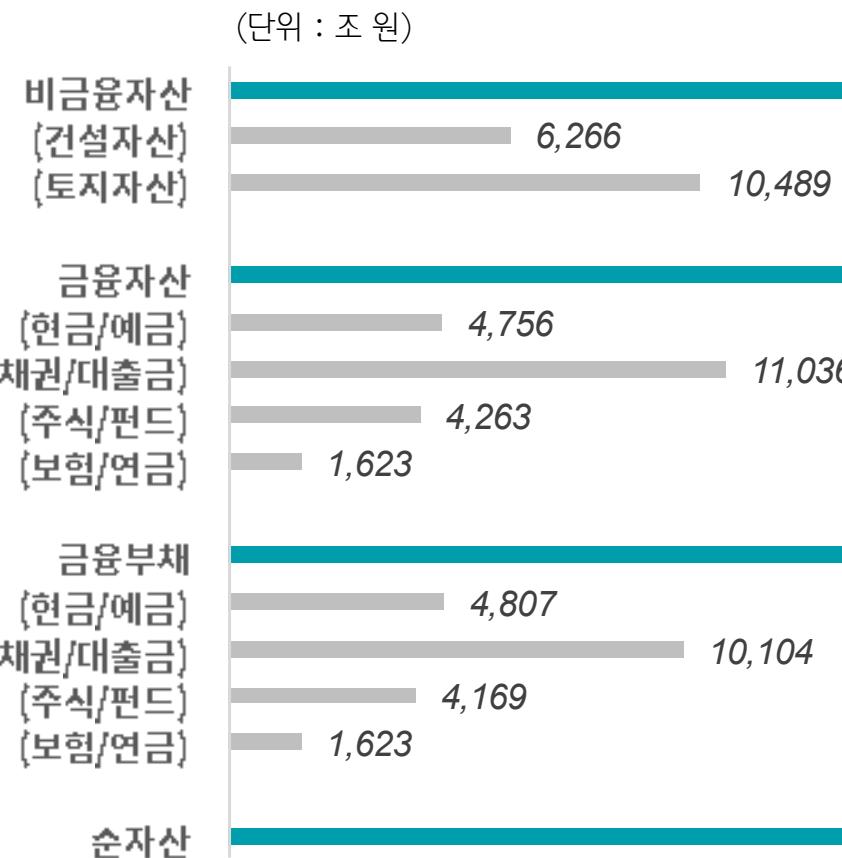


돈의 대가는 돈(이자/수익)

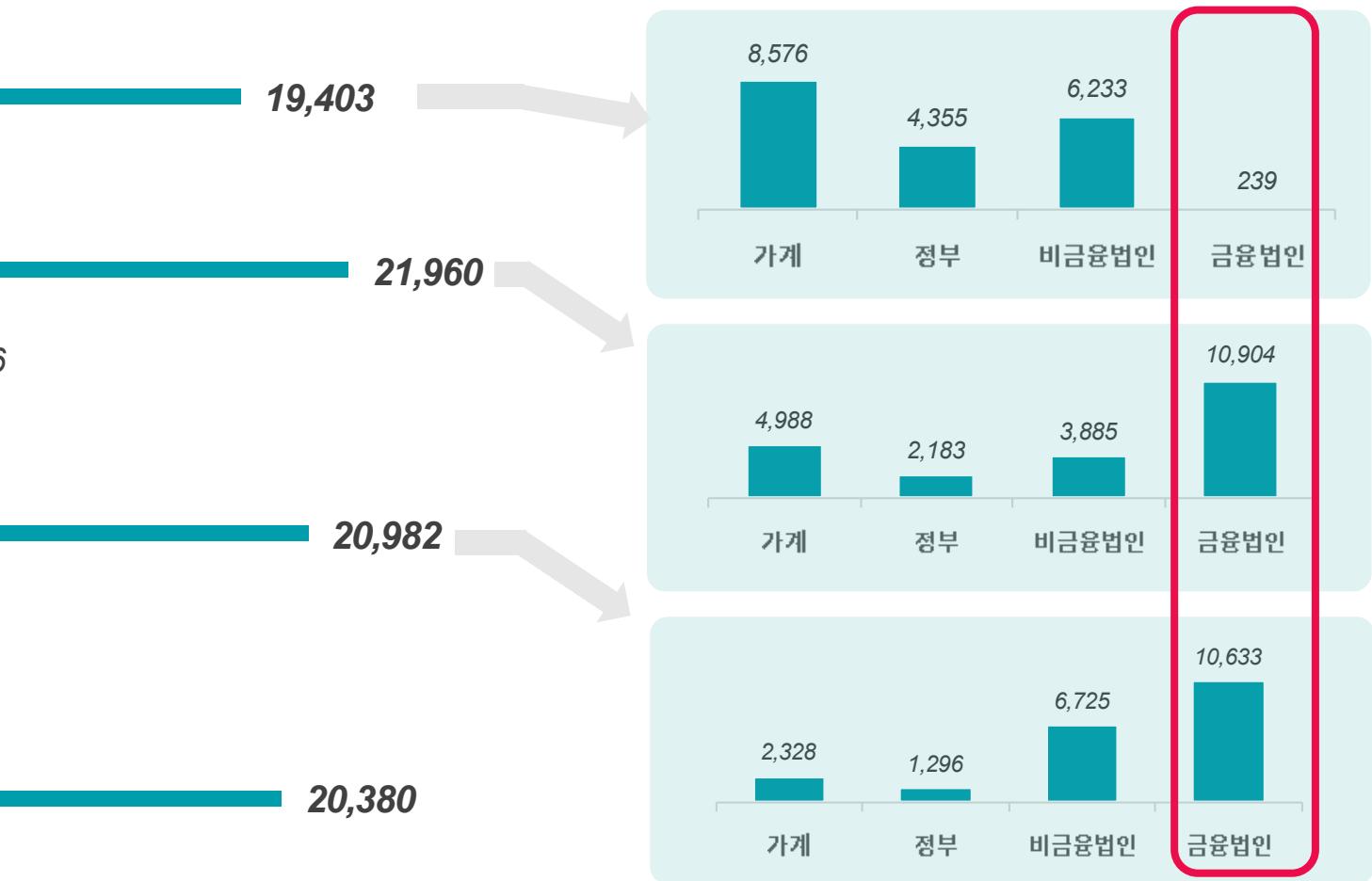
VS

## 1. 금융의 역할

# 국민 대차대조표



[자료 출처] 통계청, 국민 대차대조표 (2022년 기준 잠정치)

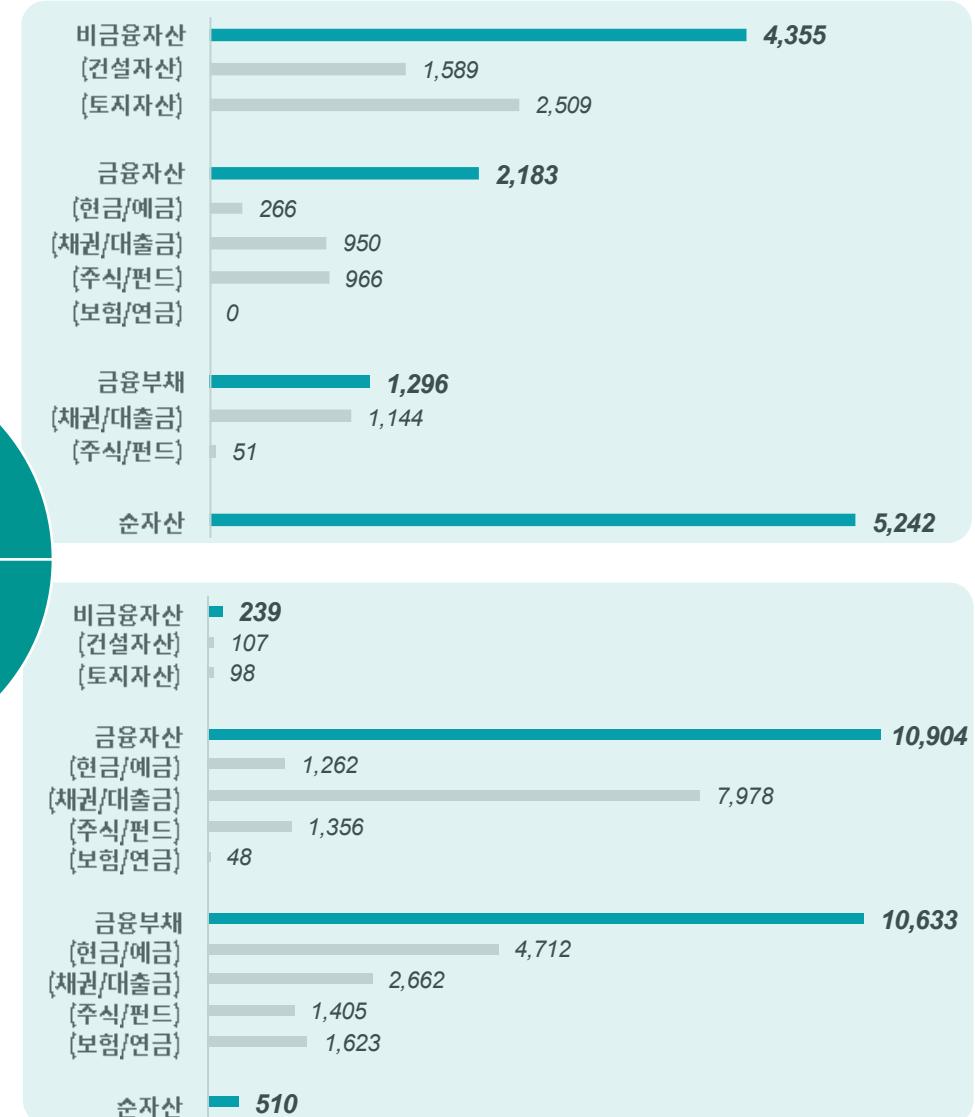
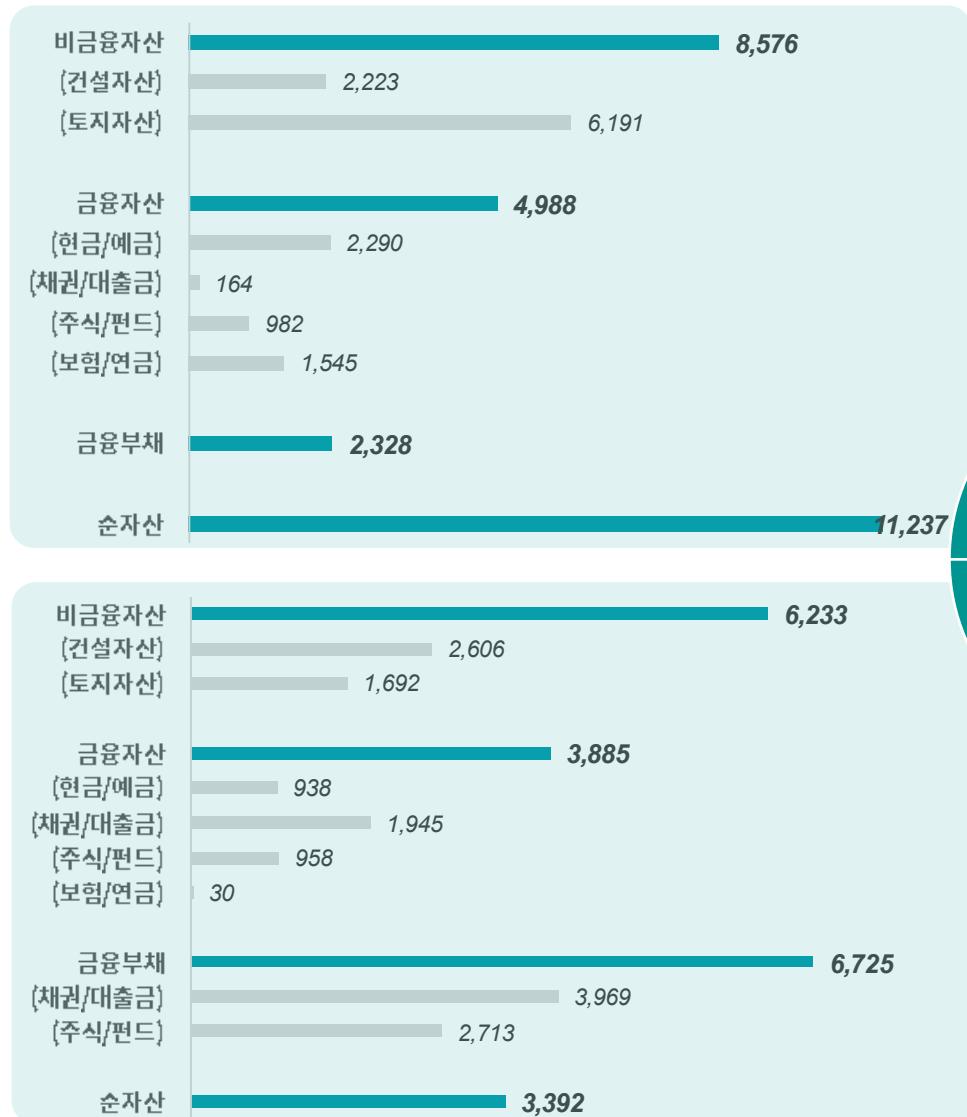


## 1. 금융 산업과 은행 서비스

### 1. 금융의 역할

(단위 : 조 원)

### 부문별



## 1. 금융의 역할

(단위 : 조 원)



## 1. 금융의 역할

# 금융기관의 역할

### 운용-조달 연결



- 저축/투자 (잉여자금) : 현재의 소득을 미래의 소비와 바꾸는 것
- 차입/출자 (부족자금) : 미래의 기대 소득을 현재의 지출과 바꾸는 것
- 잉여-부족 자금의 연결
- 현재-미래 지출의 연결

### 지불 수단 제공



- 결제/송금 : 현금이 아닌 지불 수단
- 환전 : 서로 다른 통화에 대한 지불
- 신용카드 : 단기 신용대출을 활용한 지불 수단

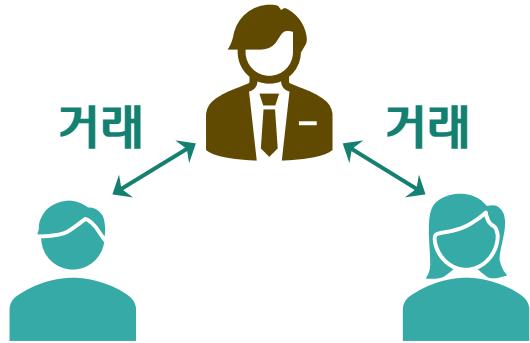
### 유동성 공급



- 유동성 : 거래량 창출
- 자금의 수요-공급에 대한 시장 조성자 역할
- 은행 : 예금-대출
- 증권 : 투자-출자/차입
- 딜러 또는 브로커 역할 수행

## 1. 금융의 역할

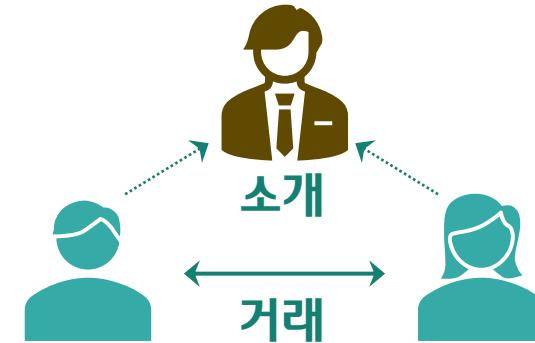
### 딜러



**Dealer | 자기매매**

- 자기 자본 필요
- 각각의 고객과 직접 거래
- 고객간의 관계 없음
- 거래 차익으로 수입 창출
- 예 : 도매상, 카지노 딜러, 중고차 딜러
- 2004년 이전의 펀드 판매 (미매각)

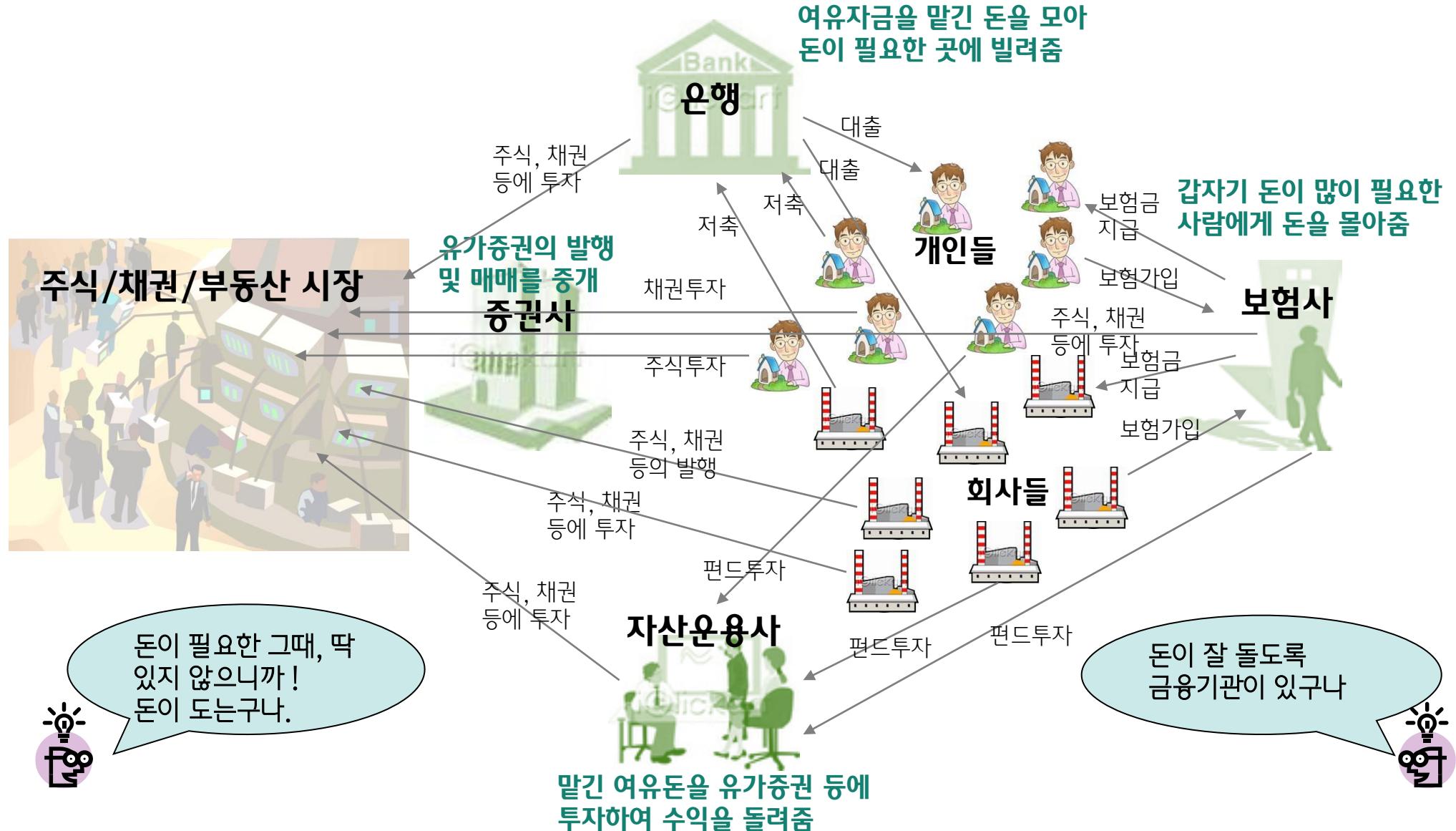
### 브로커



**Broker | 증개인**

- 자기 자본 불필요
- 고객간에 직접 거래
- 주식거래 : 고객-거래소 거래
- 증개 수수료로 수입 창출
- 예 : 부동산 증개인, 경매인
- 2004년 이후의 펀드 판매

## 1. 금융의 역할



## 1. 금융의 역할



# 신용(信用, Credit)

- **BIS 비율 / 예금자 보호 제도**
  - 은행이 망해도 내 예금을 돌려 받을 것이라는 믿음
- **RBC 비율 / 예금자 보호 제도 / 계약 이전 제도**
  - 보험사가 망해도 내 보험금이 유지될 것이라는 믿음
- **증권 예탁 제도 / 고객 계좌 구분 관리**
  - 증권사가 망해도 내 재산이 그대로 있을 것이라는 믿음
- **수탁-위탁 제도 / 펀드 계좌 구분 관리 / 계약 이전 제도**
  - 자산운용사가 망해도 내 재산이 유지될 것이라는 믿음

## 1. 금융의 역할

# 지급 결제 시스템



## 한은금융망(BOK-Wire)

- 거액결제 시스템
- 한국은행과의 거래, 금융기관간의 거래 최종 결제

금융결제원  
소액결제시스템

KRX  
KOREA EXCHANGE  
한국거래소

KSD  
한국예탁결제원  
증권결제시스템

CLS  
외환결제시스템

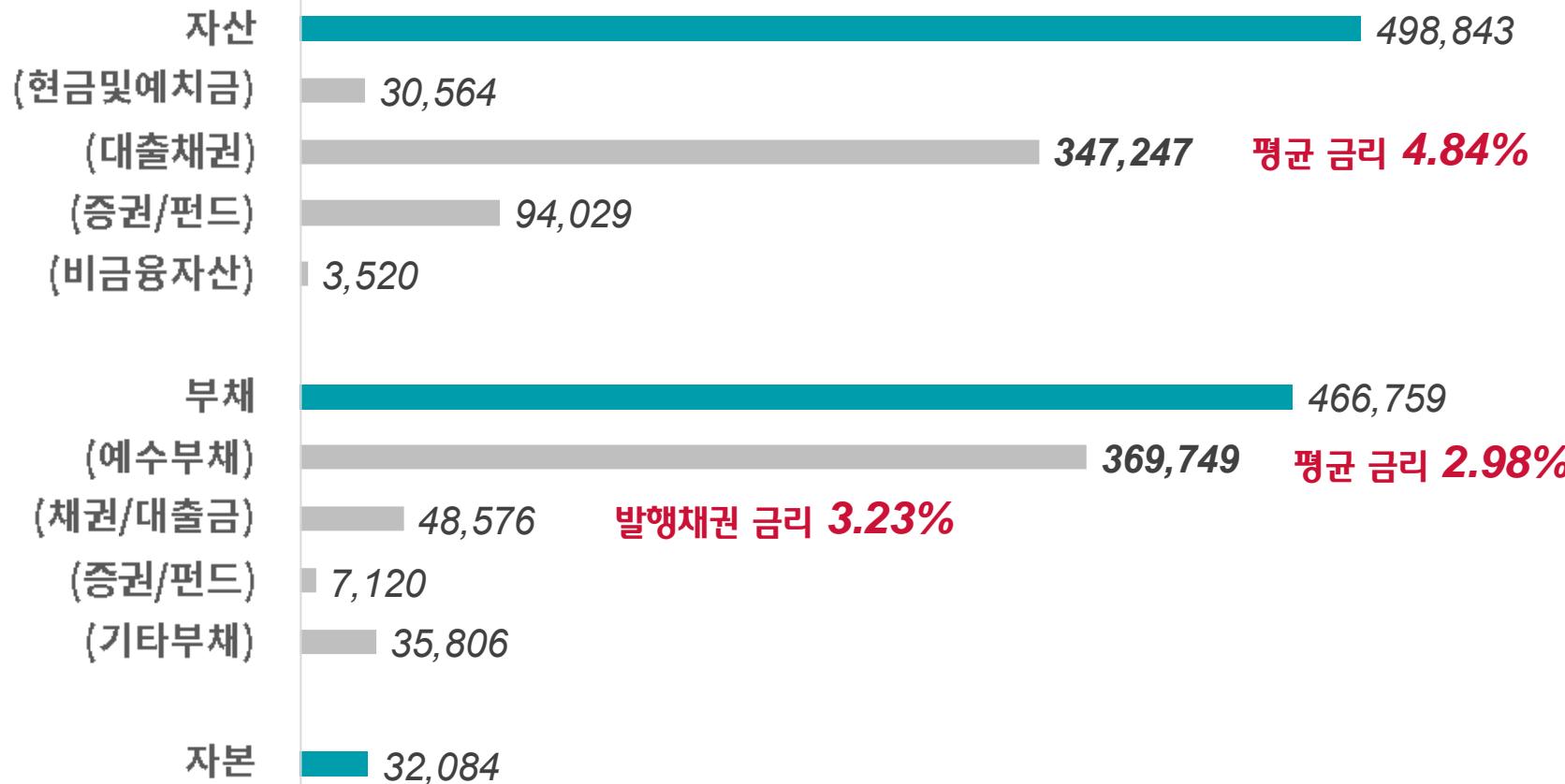
- 금융기관 고객들과의 거래 결제 시스템
- 은행공동망, 전자상거래, 지로, 신용카드, 소액결제...
- 주식, 채권 등 증권 소유권 이전 및 대금 결제 시스템
- 주식, 파생상품, 채권, RP...
- 외환시장의 외환 매도/매수 기관간의 동시결제 시스템
- CLS(Continuous Linked Settlement) 은행 연결

1. 금융 산업과 은행 서비스

2. 금융상품을 통해 보는 은행업

## 하나은행 재무상태표 (FY 2023)

(단위 : 십억 원)



운용

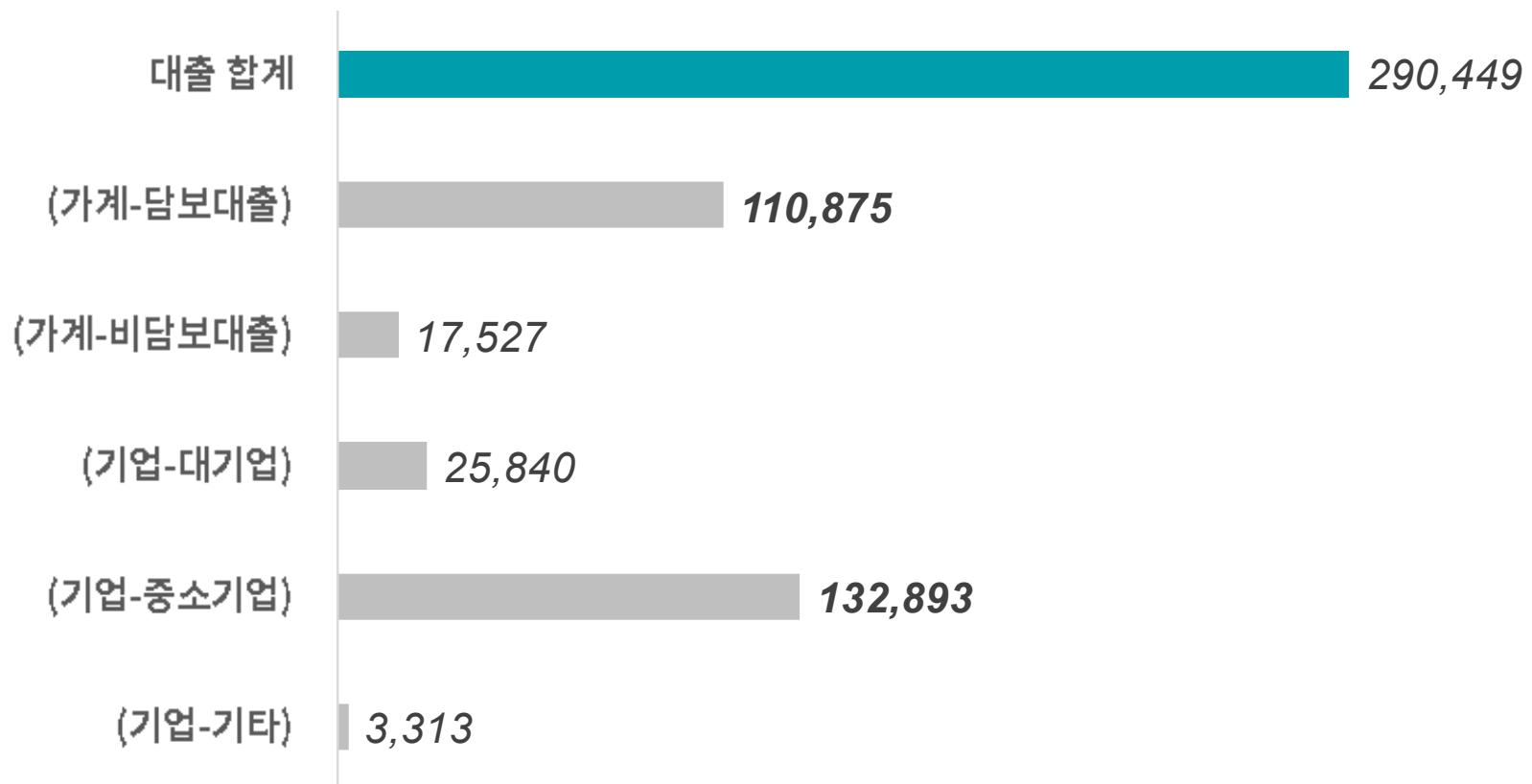
조달

1. 금융 산업과 은행 서비스

2. 금융상품을 통해 보는 은행업

## 대출채권 (국내)

(FY 2023)



## 1. 금융 산업과 은행 서비스

## 2. 금융상품을 통해 보는 은행업

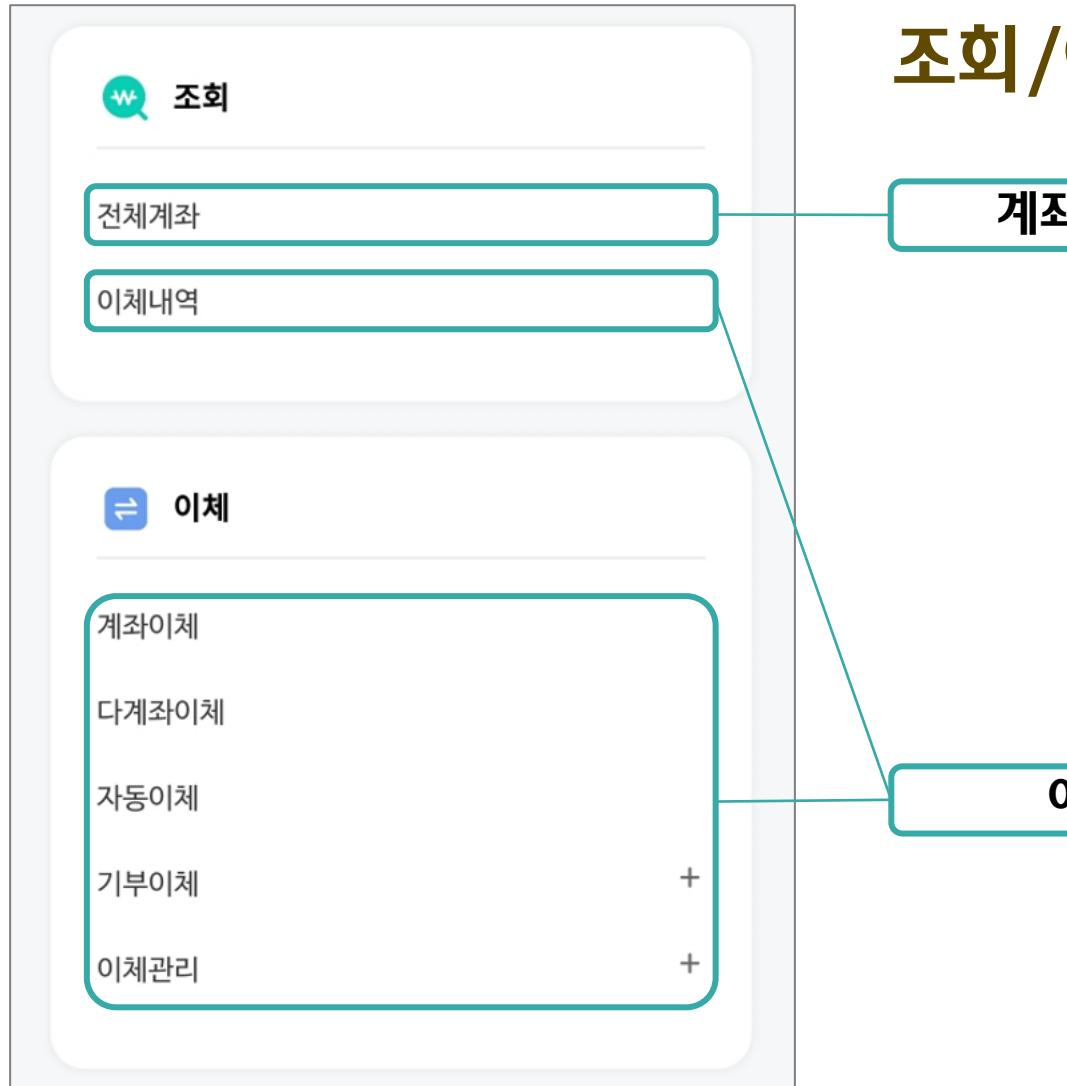
# 하나은행

조회	이체	공과금	외환	금융상품
계좌조회	계좌이체	My공과금	환율/외화예금 금리	추천상품
거래내역조회	심풀이체	지로/생활요금/기타	환전지갑	예금
거래증지/휴면계좌조회	다계좌이체	지방세	외화송금	대출
해지계좌 조회	이체 결과조회	국세/관세	외화계좌이체	펀드
세금우대한도조회	예약이체	범칙금/벌과금	외환매매 예약서비스	신탁
기일도래현황	자동이체	대학등록금	외화예금	퇴직연금/IRP
기타조회	기타이체서비스	4대 보험료	송금클럽	ISA
	계좌이동/계좌통합관리	아파트관리비		보험
	착오송금 조회/반환	지로/공과금예약납부조회		
	연금계좌이체	지역개발채권		
	간편 에스кро로			
예금	대출	펀드	신탁/ISA	마이하나
보험	카드	주택도시기금	전자통장	현금IC카드
비대면계좌개설	인증센터	보안센터	고객센터	하나라운지
퇴직연금/개인형IRP	연금상품 보유현황	골드바	오픈뱅킹	부동산

The image shows the main screen of the SmartBanking mobile application. At the top, there's a search bar with the placeholder "궁금한 것을 검색해 보세요". Below it, a summary section displays account details: "이진수 (Hana VIP)" with a balance of "1,234,567 원", and icons for "인증/보안", "영업점", and "고객센터". The main content area is divided into several sections:

- 회원**: Includes "최근 이용" (Recent Use) with items like "계좌이체", "공과금납부", "전체계좌", "マイ데이터", "지출관리", and "연금현황/매도".
- 회원**: Shows "이체" (Transfer) with options for "자산관리", "예적금", and "퇴직연금".
- 회원**: Shows "이체" (Transfer) with options for "전체계좌" and "이체내역".
- 회원**: Shows "이체" (Transfer) with options for "계좌이체", "다계좌이체", "자동이체", "기부이체", and "이체관리".
- 자산관리**: Shows "하나은행 자산" (One Bank Assets) with a balance of "1,234,567 원" and a blue button labeled "이체결제". It also lists "미래데이터" (Future Data), "마이데이터 설정", "지출관리", "대출똑똑캐어", "신용관리서비스", "국민연금 계약관리", "영업점 추천상품" (Recommended Branch Products), "골드클럽 매거진", and "오션뱅킹".
- 예적금**: Shows a summary of "입출금통장 가입" (Deposited and Withdrawn Accounts).

The image shows the Hana Q mobile application's main screen. At the top left is the logo '1Q' in red and teal, with the Korean text '하나원큐' (Hana Queue) below it. The top right features a search bar with the placeholder '궁금한 것을 검색해 보세요' (Search what you want to know). Below the search bar is a user profile section showing a profile icon, the name '이진수' (Lee Jin-su), and the status 'Hana VIP'. To the right is a '로그아웃' (Logout) button. A navigation bar at the bottom contains three items: '인증/보안' (Authentication/Security), '영업점' (Business Point), and '고객센터' (Customer Center). The main content area displays a '최근 이용' (Recent Use) section with six items, each with a star icon and text: '계좌이체' (Bank Transfer), '전체계좌' (Full Account), '지출관리' (Expense Management), '공과금납부' (Utility Payment), 'マイ데이터' (My Data), and '연금현황/매도' (Pension Status/Withdrawal).



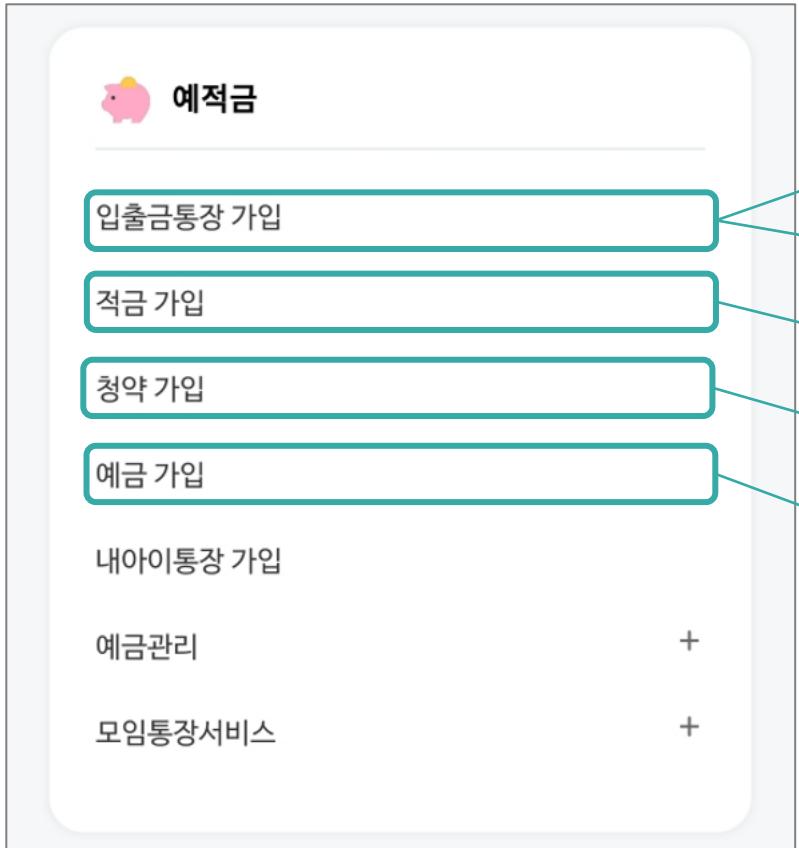
## 조회/이체

계좌 조회

하나은행에서 가입한 금융상품들의 현황

이체

내 계좌에서 다른 계좌로 송금 (지불)



## 저축 (예적금)

**수시 입출금 예금**

입출금에 제한이 없는 보통예금

**파킹 통장**

파킹 공간에 대한 우대금리 제공 수시 입출금식

**적금**

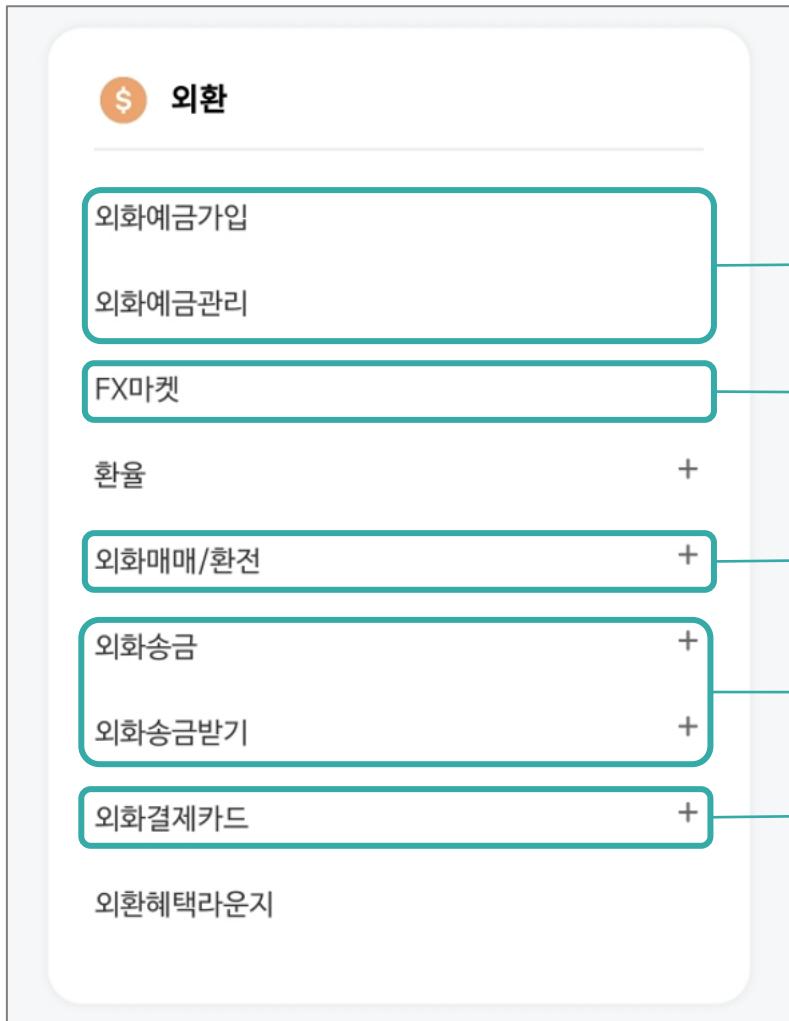
정해진 기간동안 매월 정기적 저축

**주택청약종합저축**

주택 청약 자격 획득을 위한 적금

**정기예금**

정해진 기간동안 목돈을 저축



## 저축 (외화)

**외화예금**

원화가 아닌 외화 기준 입출금, 정기예적금

**환전(투자)**

하나은행 원화 예금과 외화 예금 간의 이체 거래

**환전(현금)**

하나은행 원화 예금에서 현금 수령 예약

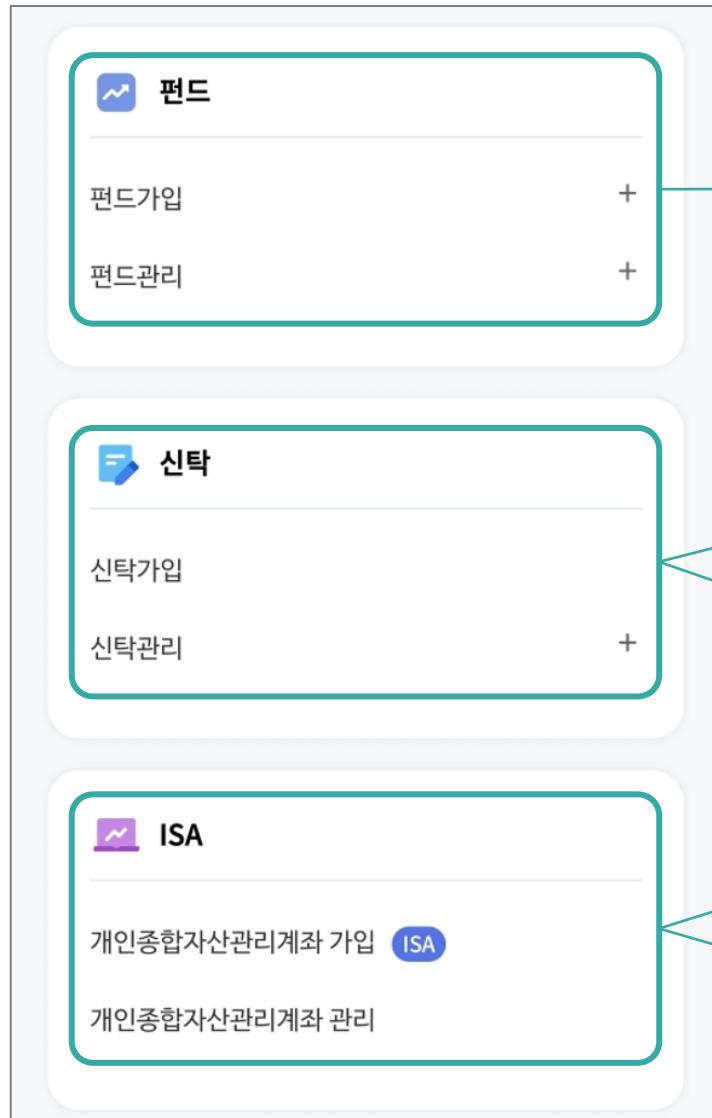
**외화 송금**

동일 통화 계좌간에 송금 주고 받기 (국내/국외)

**외화카드**

외화예금을 기초로 발행한 외화 결제 카드

## 2. 금융상품을 통해 보는 은행업



# 저축 (투자)

편드

## ‘자산운용사’의 실적배당 상품을 판매(대리점)

금전신탁

재산신탁

## 고객의 돈을 은행이 위탁 관리하는 실적배당 상품 (ELT, ETF, 자문형 EMP)

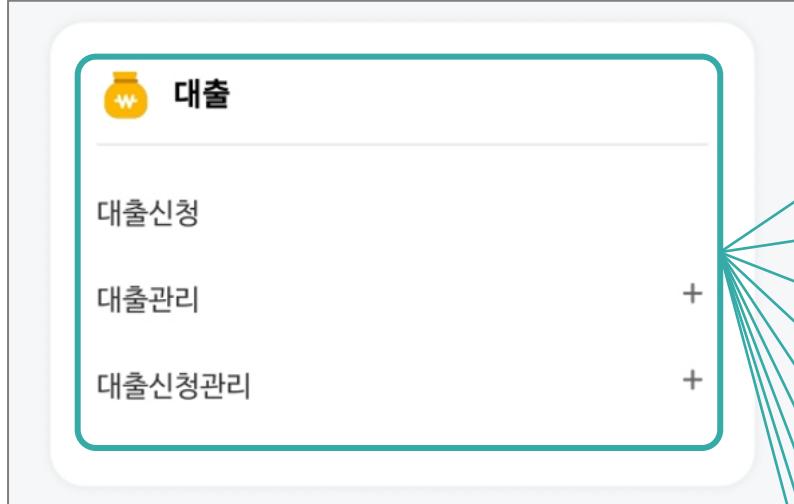
## 고객의 재산을 은행이 위탁 관리하는 실적배당 상품(금현물 신탁, 유언대용신탁)

신탁형 ISA

일인형 ISA

## 복수의 상품을 묶어서 운용 (세제 혜택) (예금, 펀드, ETF, ELS)

## 내 ISA 계좌 거래를 은행에 위임하여 운용 (펀드 포트폴리오)



## 상품가입 (대출)

**신용대출**

담보 없이 신용평가로 정해진 기간과 금액 대출

**마이너스통장**

금액 한도내에서 자유롭게 쓸 수 있는 신용대출

**주택담보대출**

주택을 담보로 한 대출 (모기지론)

**예금담보대출**

가입한 정기예금을 담보로 한 대출

**전세대출**

임차권(전월세보증금)을 담보로 한 대출

**정책자금대출**

서민금융, 주택도시기금 등의 대출상품(대리점)

**자동차 대출**

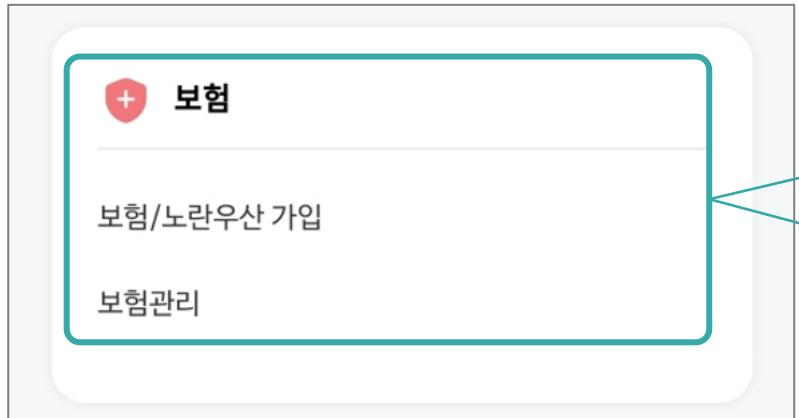
보증보험을 담보로 한 대출 (할부금융 유사)

**개인사업자 대출**

신용보증, 정부, 부동산 담보부 사업자금 대출

**대환대출**

기존(타 금융기관) 대출 갈아타기 대출



## 상품가입 (보험, 카드)

보험

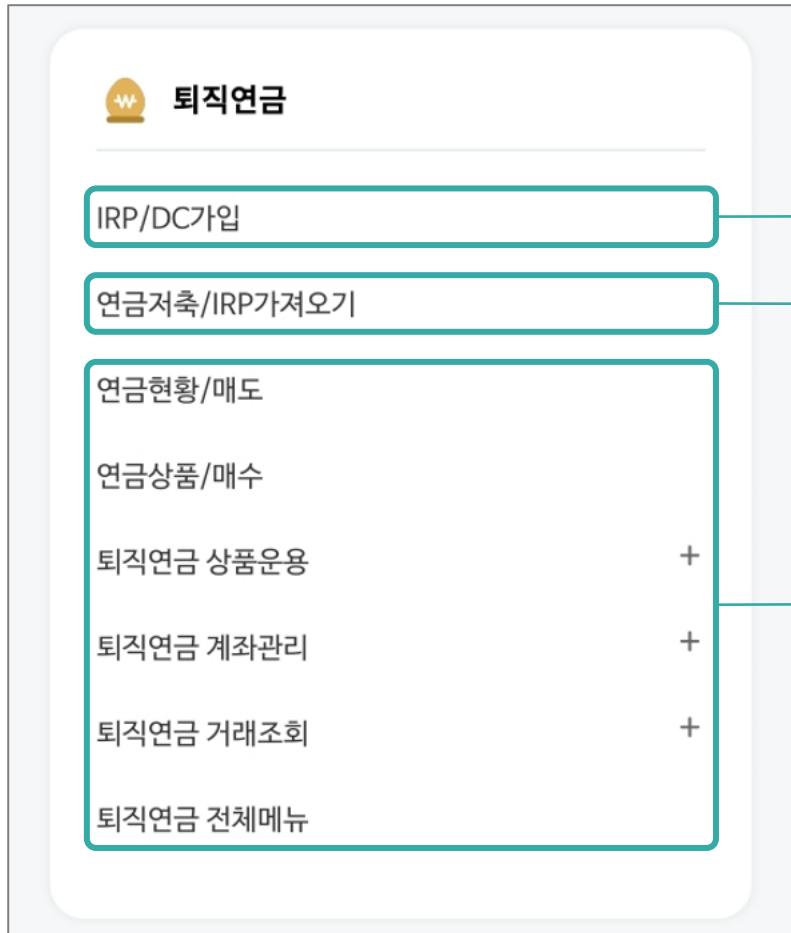
방카슈랑스; 생명/화재보험 보험상품 (대리점)

노란우산 공제

개인 사업자 폐업 보험 (대리점)

카드

신용카드, 체크카드 서비스 (하나카드 링크)



## 상품가입 (퇴직연금)

퇴직연금 가입

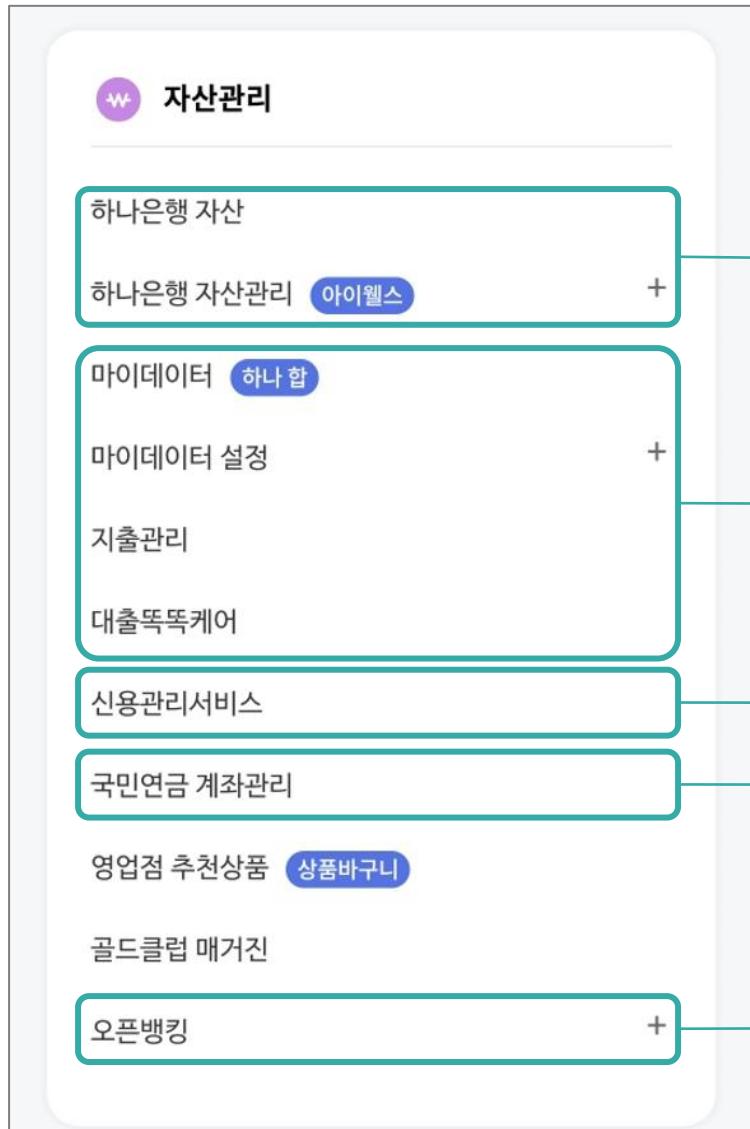
개인 퇴직연금 계좌 (IRP/DC형)

계좌 이전

타사 퇴직연금/개인연금 계좌 (IRP/연금저축)

계좌 운용

금융 상품 매매 (정기예금, 펀드, ETF)



## 자산관리 (상품 종합 관리)

온라인 자산관리 (은행 내 포트폴리오 분석 후 예적금+펀드+ELF 포트폴리오 제시)

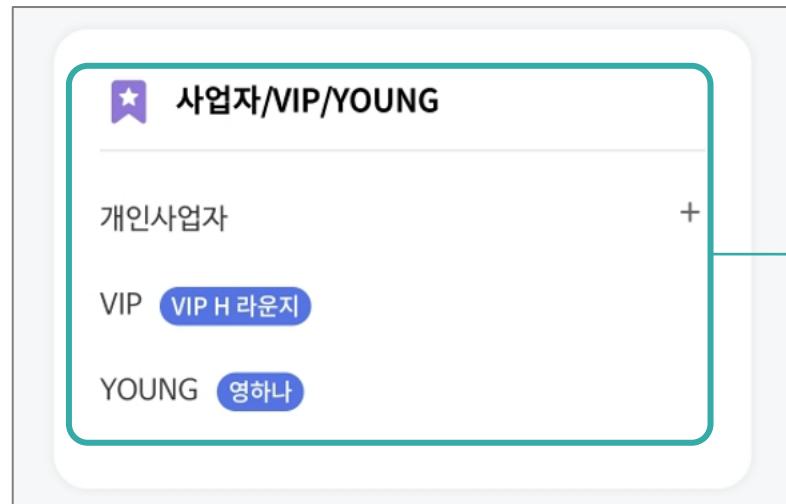
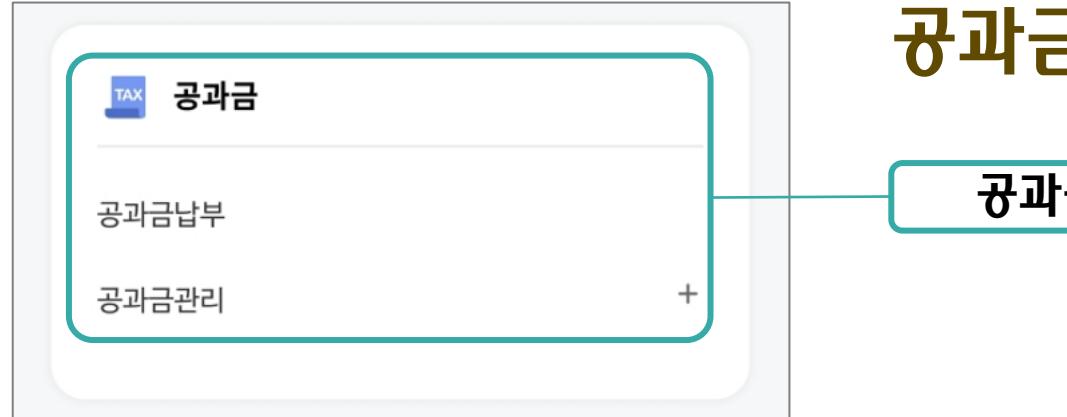
금융기관, 핀테크 서비스, 통신 서비스 정보 통합하여 자산/지출 현황 및 맞춤형 제안 제시

신용정보 업체 신용 기록 조회

국민연금 수령자 대상 고객 유치

**오픈뱅킹** - 다른 금융기관에서 가입한 금융상품들의 잔고 현황 및 이체

## 공과금 및 기타



## 3. 은행 업무 영역과 역할

# 하나은행 조직도



## 3. 은행 업무 영역과 역할

영업/지원

**영업지원**

- 각 부문별 영업 지원을 위한 마케팅 기획
  - 영업 지점 지원/관리

**온라인 신사업**

- 디지털 신사업 기획
- 디지털 마케팅 및 영업 기획

※ 마케팅/영업 관점의 고객 경험 분석/설계

**기업 고객(B2B)**

- 기업 고객 영업 기획/지원
- 기업여신/수신, 외환(수출입) 등 서비스
- 기업, 타금융회사 등과의 협업을 통한 사업

**일반 고객(B2C)**

- 대중(Mass) 고객 영업 기획/지원
- 지점 일반 창구 업무 중심 지원 업무
- 디지털 마케팅 및 디지털 영업 관리

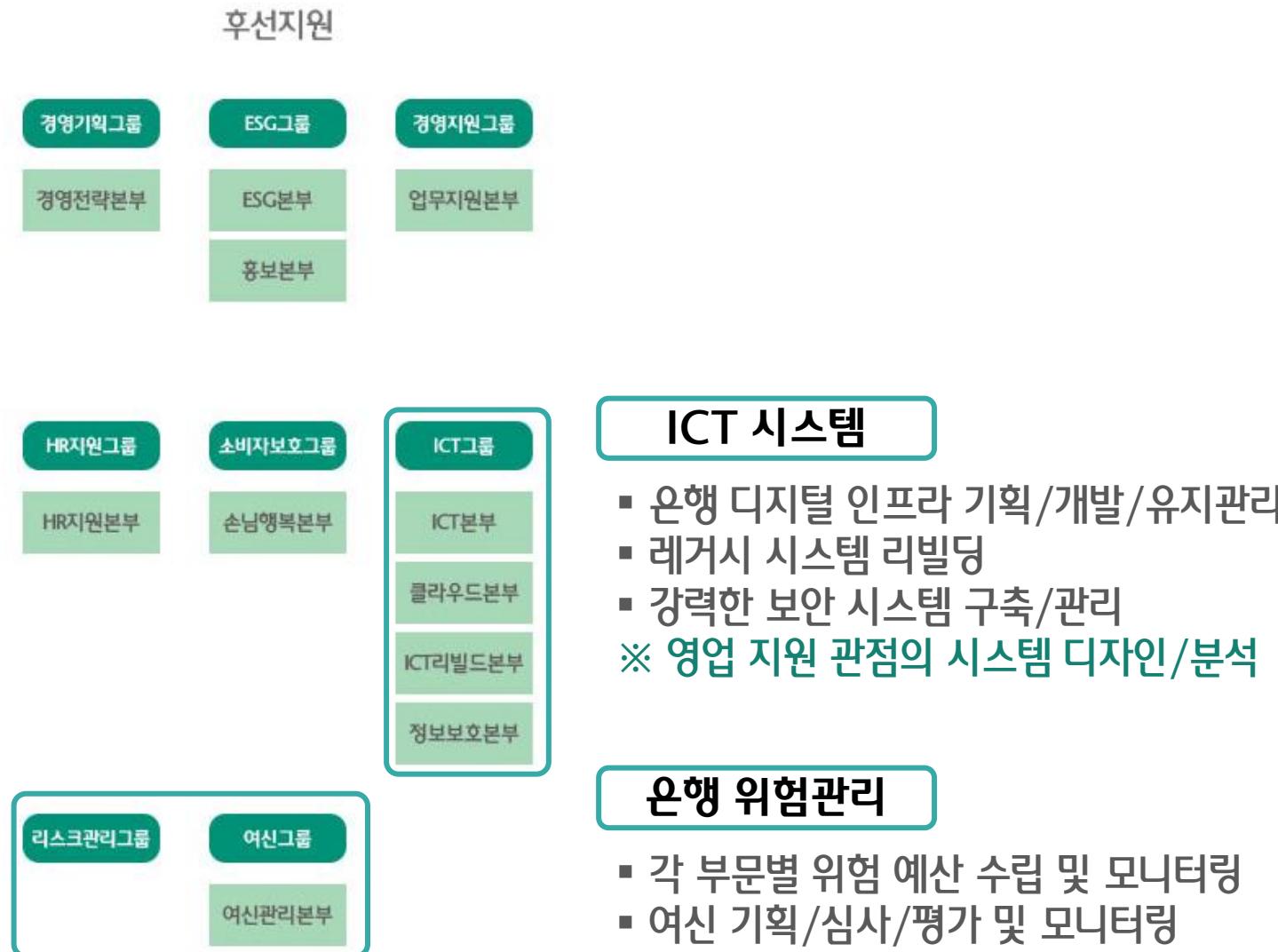
**종합 자산관리**

- 고액 자산가 PB 서비스 기획/지원
- 퇴직연금(DB/DC/IRP) 서비스 기획/지원
- 신탁상품 기획/관리/지원
- 펀드 등 투자상품 기획/관리/지원

**해외 영업**

- 해외 지점 영업 기획/지원
- 국내-해외 사업 연결 관리

### 3. 은행 업무 영역과 역할



## 핵심요약

- 경제는 실물 경제와 금융 경제가 두 축이 되어 작동되는 시스템
- 금융기관은 ‘딜러’ 또는 ‘브로커’가 되어 (1) 운용과 조달을 연결하고, (2) 지불 수단을 제공하며, (3) 경제 시스템에 유동성을 공급
- 금융 시스템 유지를 위해 ‘신용’이 가장 중요하며, 금융위원회와 한국은행이 필두가 되어 다양한 신용유지 장치가 존재
- 한은금융망을 중심으로 소액결제시스템, 증권결제시스템, 외환결제시스템을 통해 효율적으로 지급결제 수행
- 은행 영업의 중심은 수신(예금)-여신(대출) 영업임
- 은행의 고유 영업은 예금, 대출, 외환, 신탁 이지만, 고객 종합 자산관리를 위해 펀드, 보험 등의 판매와 금융상품과 컨설팅을 아우르는 자산관리 영업 수행

## 2. 비은행 금융과 은행 서비스

---

# 학습목표

- ▶ 증권, 보험, 자산운용, 여전업 등 비은행 금융업에 대해 이해
- ▶ 은행의 비은행 금융상품의 연관성과 판매 과정을 이해
- ▶ 종합 서비스로서 고객 자산관리 서비스를 이해

## 1. 비은행 금융업의 이해

# 상호저축은행

### 전국 저축은행 현황

서울(23)

DB, JT친애, KB, NH, OK, OSB, SBI, 대신, 더케이, 민국, HB, 스카이, 바로, 신한, 애큐온,  
예가람, 웰컴, 유안타, 다올, 조은, 키움예스, 푸른, 하나

인천/경기(19)

JT, 금화, 남양, 모아, 부림, 삼정, 상상인, 세람, 안국, 안양, 영진, 융창, 인성, 인천, 키움,  
페퍼, 평택, 한국투자, 한화

부산/경남(12)

BNK, DH, IBK, 고려, 국제, 동원제일, 솔브레인, 에스앤티, 우리, 조흥, 진주, 흥국

대구/경북/강원(11)

CK, 대백, 대아, 대원, 드림, 라온, 머스트삼일, 엠에스, 오성, 유니온, 참

광주/전남/전북/제주(7)

대한, 더블, 동양, 삼호, 센트럴, 스마트, 스타

대전/충남/충북(7)

대명, 상상인플러스, 아산, 우리금융, 오토, 청주, 한성

### 여신/수신 업무

- 은행과 동일 : 예금으로 조달하여 대출로 운용 (자산 138조)
- 중금리 예금/대출 취급
- 예금자 보호 대상 (최대 5천만원)

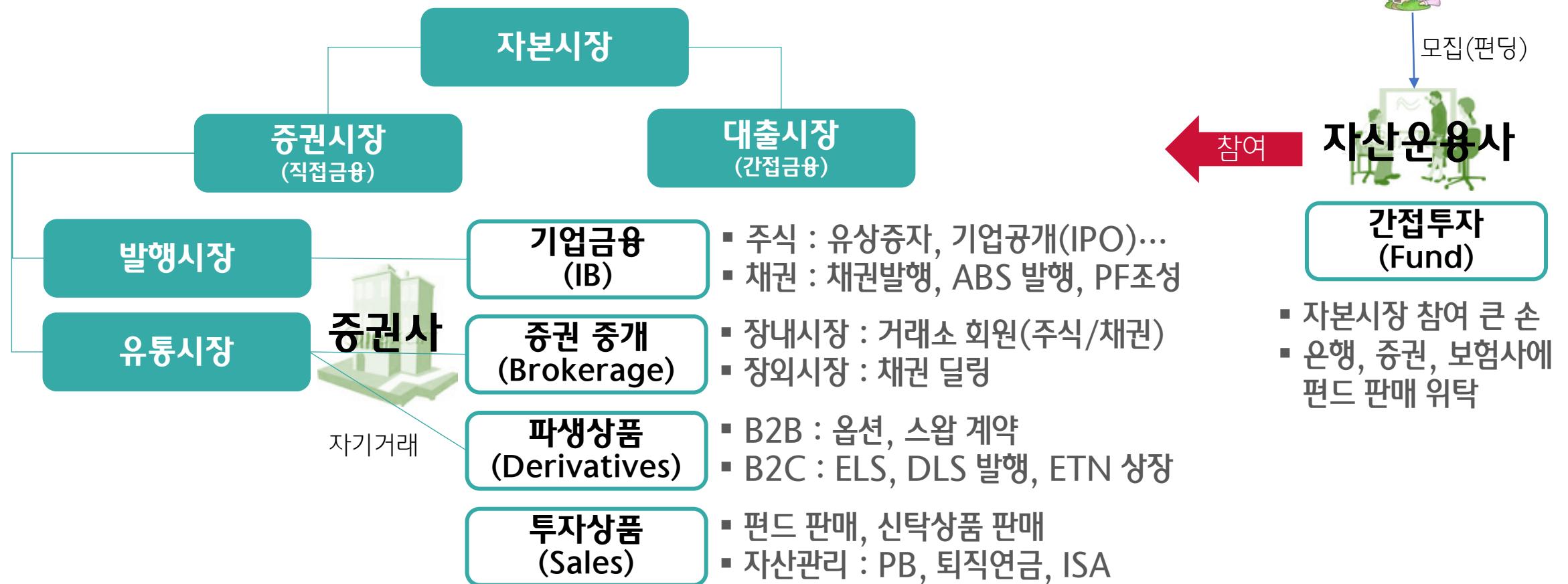
### 서민금융 표방

- 권역 제한 : 6개 권역 내에서만 영업
- 서민금융 : 개인 및 중소기업 50% 이상 대출

[자료 출처] 저축은행중앙회 (<https://fsb.or.kr>)

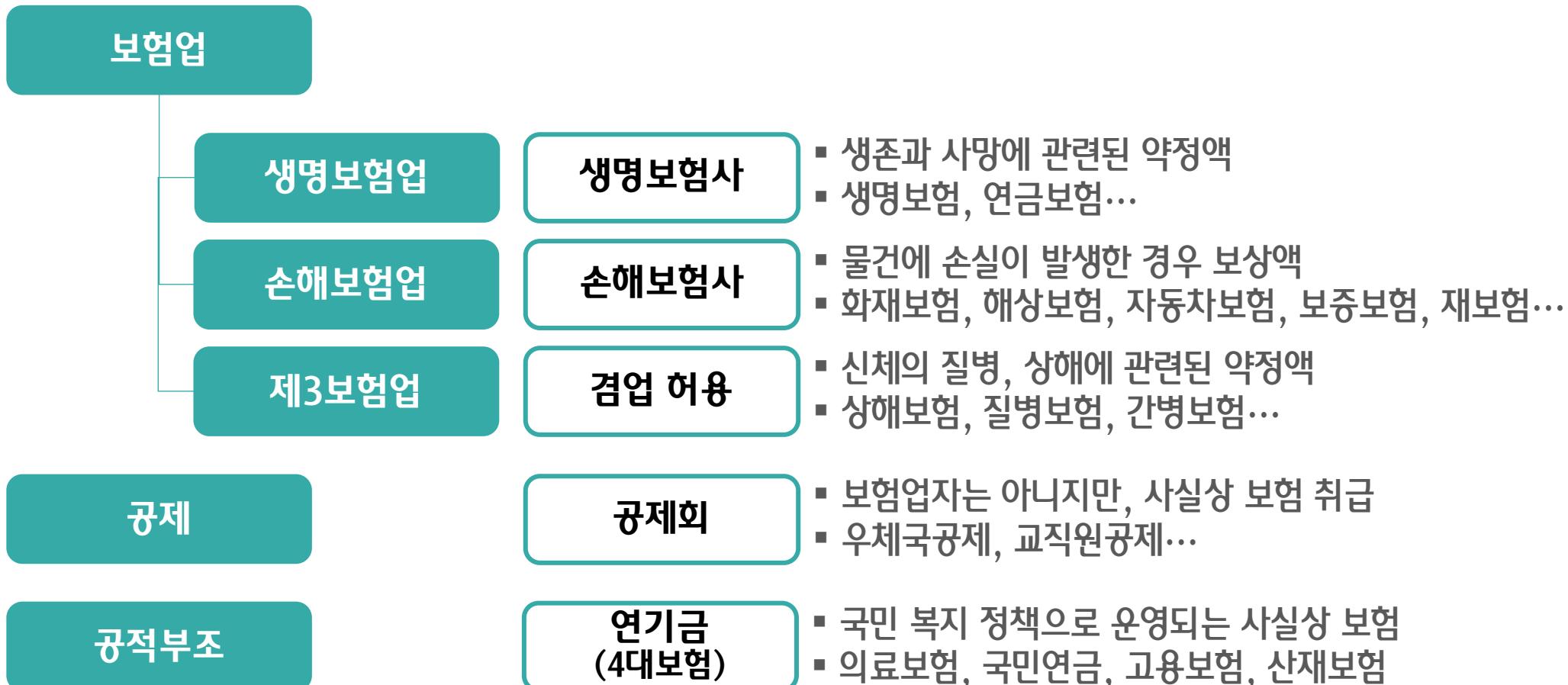
## 1. 비은행 금융업의 이해

# 자본시장과 증권투자업



1. 비은행 금융업의 이해

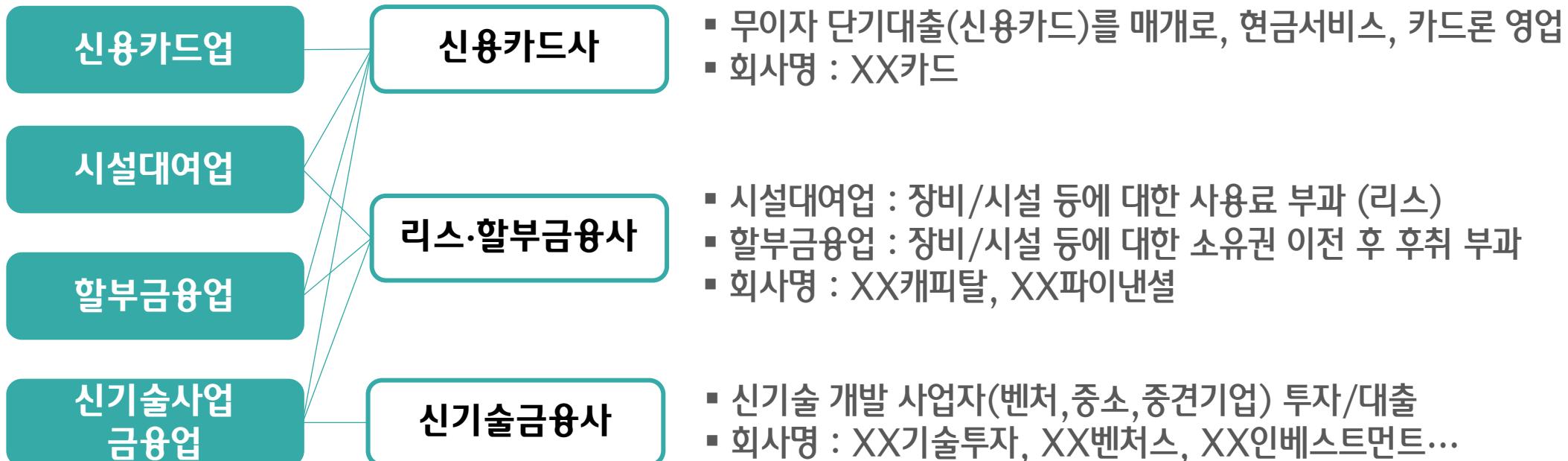
## 보험업과 보험상품



1. 비은행 금융업의 이해

## 여신전문금융업

수신(예금) 없이 자금 조달하여 다양한 형태의 대출 영업



## 2. 비은행 금융상품의 연관성과 판매 과정

# 분리에서 통합으로

## 제조와 판매의 분리 활성화

### 금융사 고유사업

- 은행 : 수신(예금), 여신(대출)
- 증권 : 유가증권 거래 주선, 중개
- 보험 : 보험상품 개발, 판매
- 투자신탁(현 자산운용) : 펀드 개발, 판매

### 오픈 아키텍처

- 타 금융업권의 (유사)금융상품 판매/제조 순차적 허용
- 은행 : 펀드, 보험(방카슈랑스), 투자증권(ELT)...
- 증권 : CMA(수시입출금식 예금), 펀드, 보험(방카슈랑스)
- 보험 : 저축성보험, 펀드(변액보험)
- 자산관리 : 퇴직연금, ISA, 종합자산관리 서비스

## 금산분리 원칙 완화 기조

### 금산분리

- 금융자본과 산업자본의 분리 (겸업 금지)
- 실제로는 은행(지주사)과 산업의 분리
- 제한 지분률 : 4%

### 핀테크부터 완화

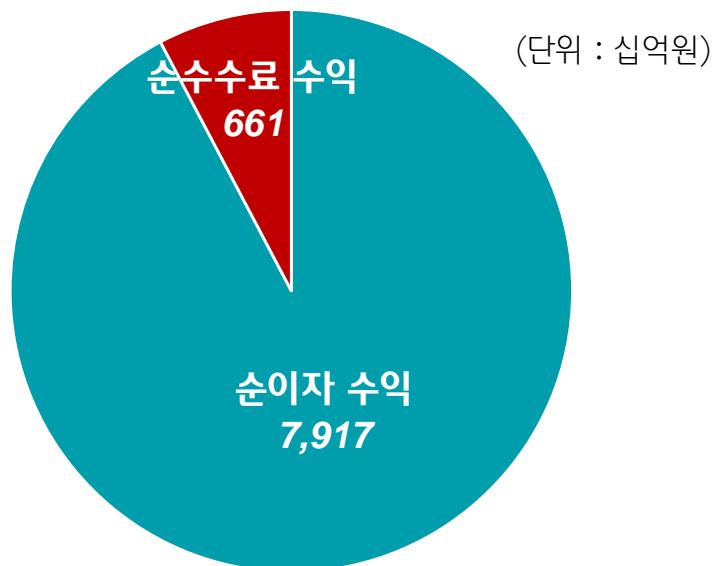
- 산업자본의 은행자본 소유 허용 : 카카오, 토스
- 은행자본의 산업자본 소유 허용 요청
- 금융/비금융 통합 디지털 플랫폼 지향
- 금융지주사의 통합앱 기획/관리/운영 허용 검토

## 2. 비은행 금융상품의 연관성과 판매 과정

# 수익모델의 다각화

전통 사업모델 비중 축소, 자본부담 낮은 거래기반 수익 확대

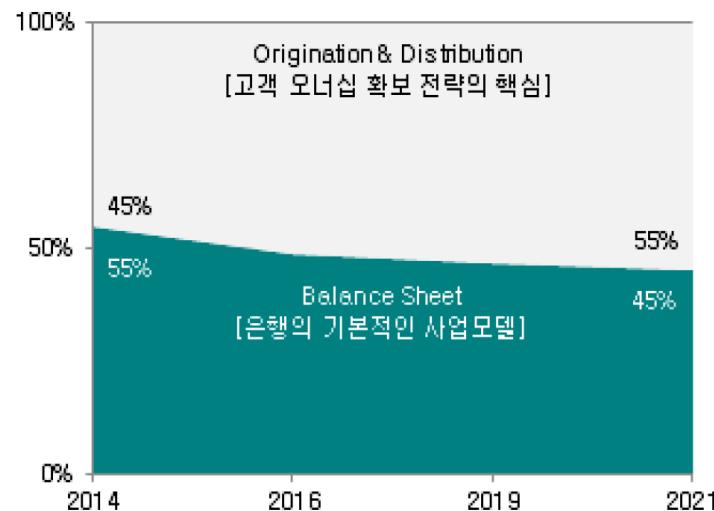
## 하나은행 2023년 수익



[자료 출처] 하나금융그룹, 2023년 결산 재무제표

## 글로벌 은행 수익 비중

### ■ 사업모델별 영업수익 기여도



자료 : McKinsey, S&P Global

[자료 출처] 심윤보, “은행업, 미래지향적 사업모델 구축이 필요”,  
하나금융경영연구소 금융경영브리프, 2021.01.

## 2. 비은행 금융상품의 연관성과 판매 과정



# 은행의 수수료 수입 상품

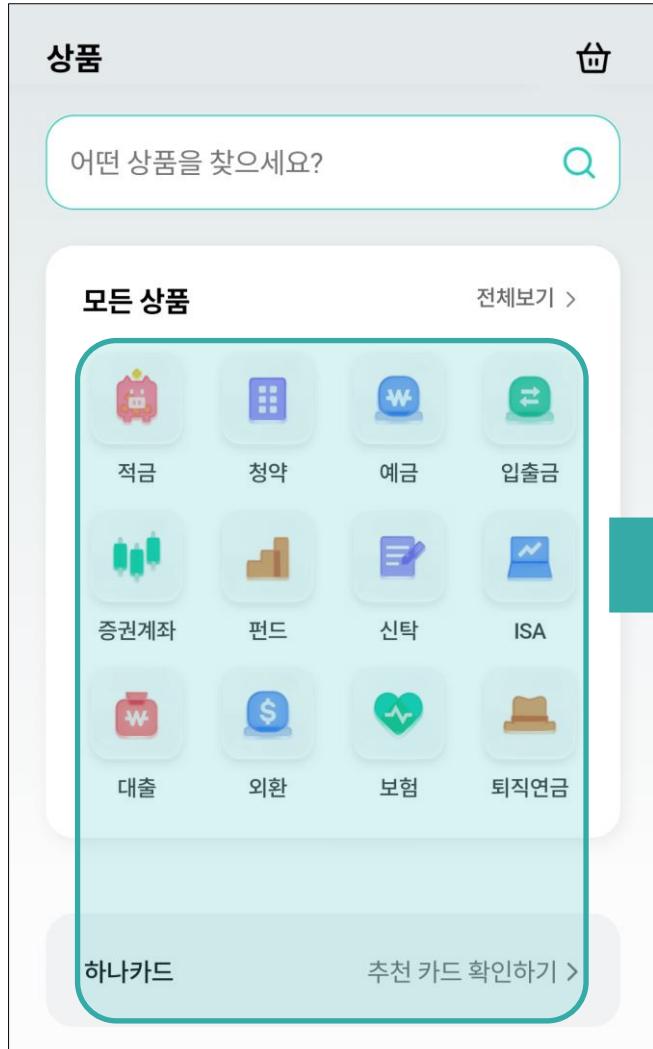
## 저축 포트폴리오

- 금리상품(예적금)을 제외한 대부분 수수료 또는 고객 유치 사업
- 청약 : 주택도시보증공사 상품의 판매 (고객유치)
- 증권계좌 : 증권사 입출금 계좌(예탁계좌) 연결 (고객유치), 투자 불가
- 신탁 : 자체 상품이나 투자 운용 위탁 다수 (증권사 상품, 자문계약)
- ISA : 예적금은 자사상품이나 투자 상품은 주로 펀드/ETF

## 위험관리 포트폴리오

- 보험 : 저축성 보험 중심 판매. 보험 계약 유지관리 취약
- 퇴직연금 : 자체 상품이나 투자 상품은 판매 (펀드/ETF), 예적금도 다수 판매

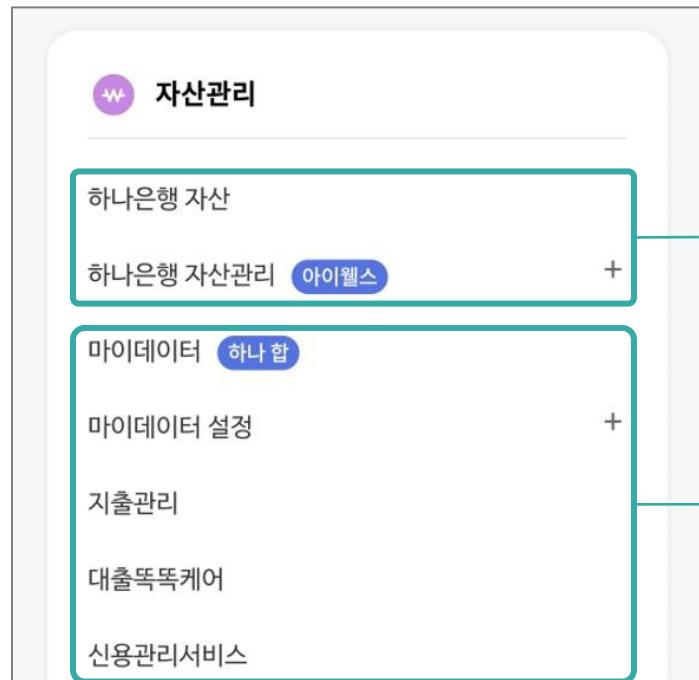
## 2. 비은행 금융상품의 연관성과 판매 과정



# 어떻게 엮을 것인가?

자산관리 서비스 영역

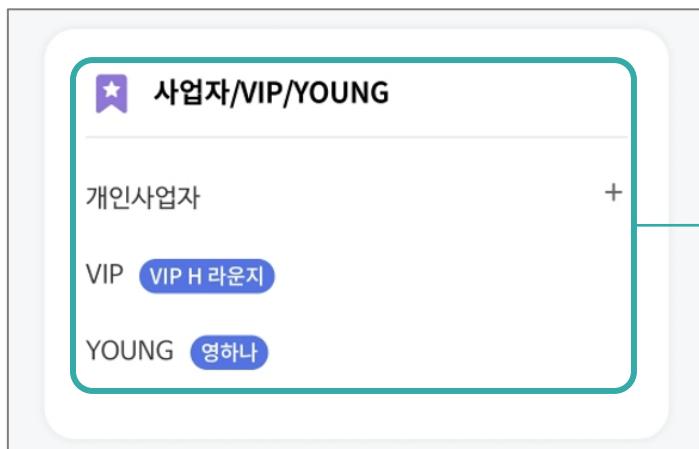
## 3. 고객 자산관리 서비스의 이해

**자산관리 서비스****로보어드바이저**

보유 재무 데이터 기준에서 포트폴리오 제시

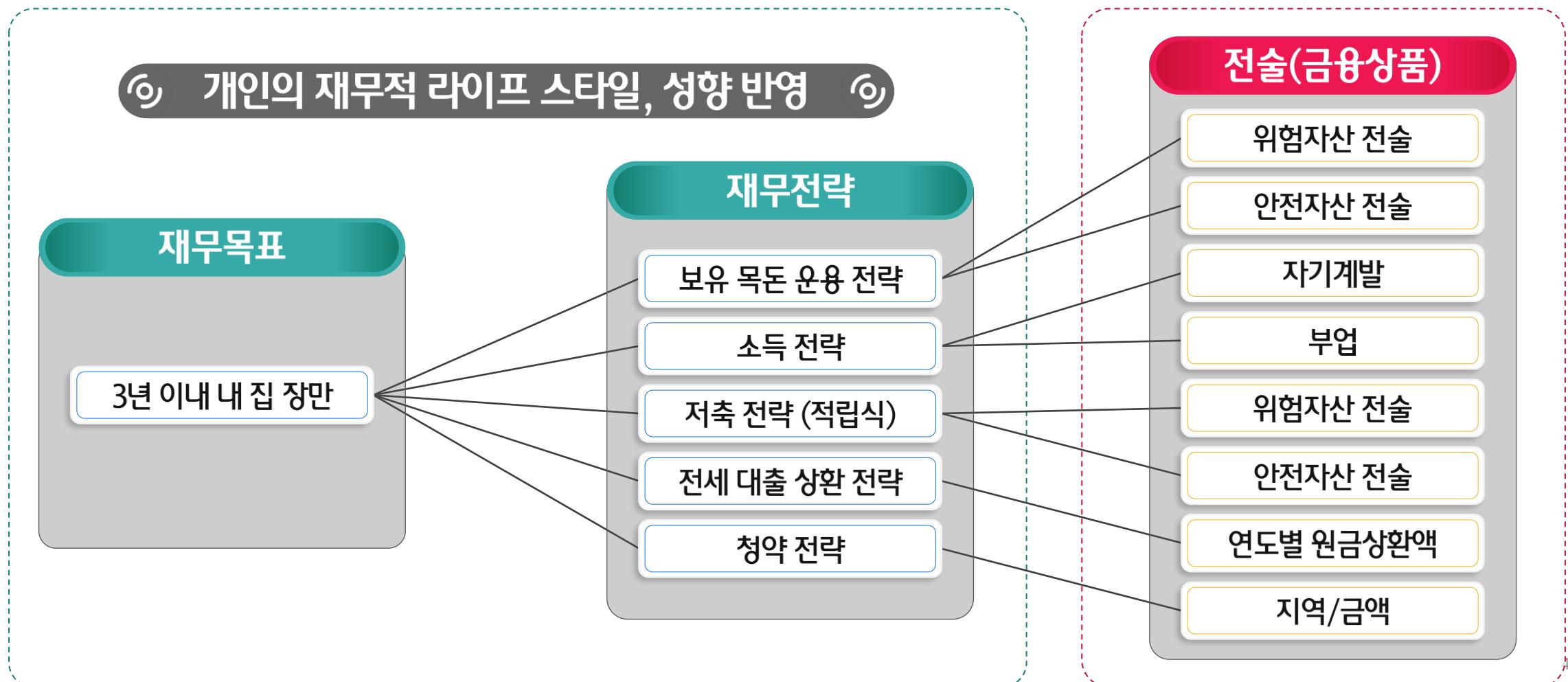
**마이 데이터**

고객의 재무 데이터 소스

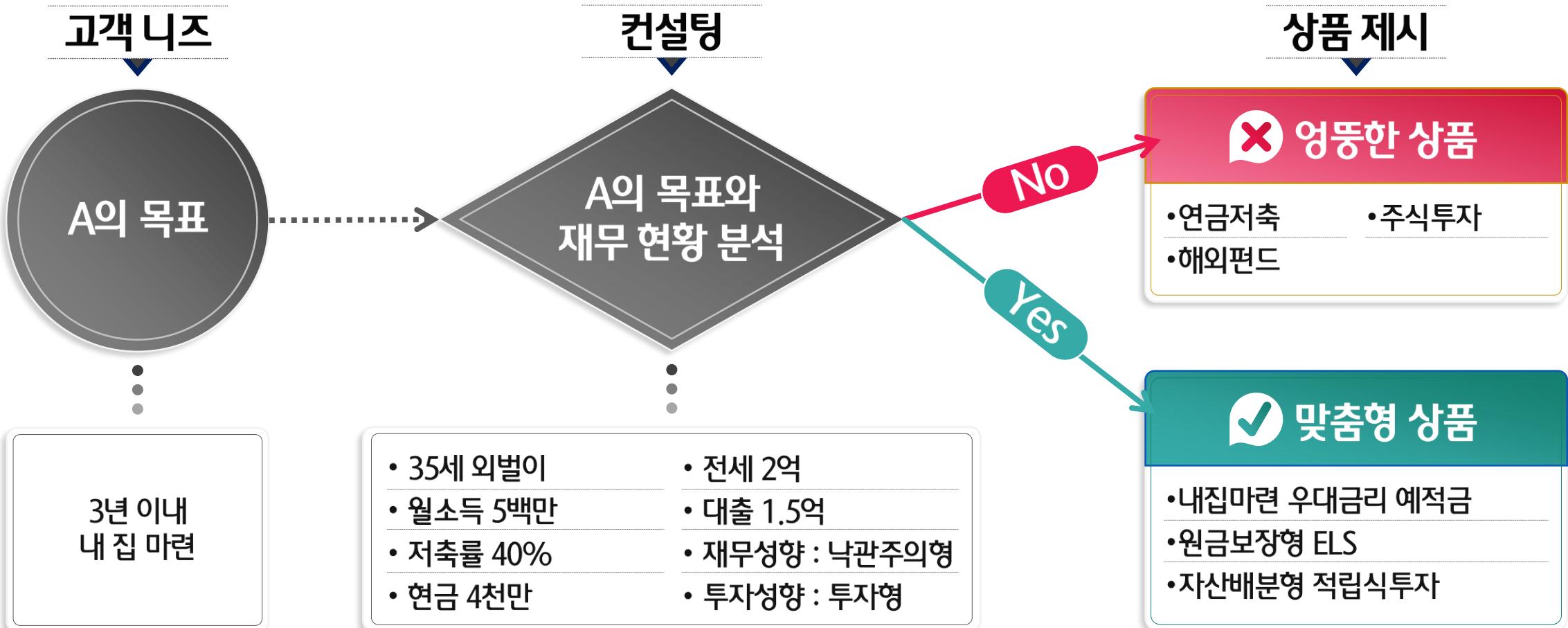
**맞춤형 서비스**

개인 사업자 : 사업 관리 지원  
 고액 자산가 : 전문가 서비스 안내 (오프라인)  
 청년층 : 재산 형성, 내 집 마련 정보 제공

## 고객에 대한 종합적 이해 필요



## 컨설팅 서비스 부재의 문제



## 재무 주치의 서비스

### 자기 진단 Self Diagnosis

- 연령대 대비 재무 상황/목표 진단
- 포트폴리오 적정성 진단
- 재무 성향 진단



### 실행 관리 Do & See

- 계획 대비 실천력 평가
- 포트폴리오 성과 평가
- 재무 상황/목표 변화 여부 평가

### 전략 제안 Suitable Strategy

- 재무 목표 수립
- 목표에 따른 전략 제안
- 재무 전략 수립

- 실행 계획 수립
- 실행 벤치마크 설정
- 포트폴리오 대상 상품 가입

### 플랜 실행 Action Plan

## 디지털 자산관리 포탈



## 핵심요약

- 저축은행은 비교적 높은 금리로 은행과 같은 수신-여신업 영위
- 증권사는 자본시장의 발행시장과 유통시장에서 (1) 기업금융, (2) 증권증개, (3) 파생상품 발행, (4) 투자상품 판매 등을 영위
- 자산운용사는 펀드를 조성하여 자본시장에 자금을 공급하며 수익 추구
- 현대판 상부상조인 보험사는 낮은 확률의 큰 손해 가능성이 있는 이벤트를 대비하는 (1) 생명보험, (2) 손해보험, (3) 제3보험 상품 판매
- 여신전문금융사는 (1) 신용카드업, (2) 시설대여업, (3) 할부금융업, (4) 신기술사업금융업을 영위
- 오픈 아키텍처와 플랫폼 비즈니스는 자산관리 서비스 강화를 요구

### 3. 금융업의 디지털 전환

---

# 학습목표

- ▶ ICT 기술이 산업을 변화시키고 발전시킨 과정을 고찰
- ▶ 핀테크로 대표되는 금융 서비스 변화에 영향을 준 디지털 전환 양상
- ▶ 빅블러, 마이 데이터 등의 환경 변화에 따른 금융업의 과제를 논의

## 1. ICT 기술의 영향과 발전

생산자는 컴퓨터로 인해 변해왔고, 소비자와 거리 축소

소품종 대량 생산

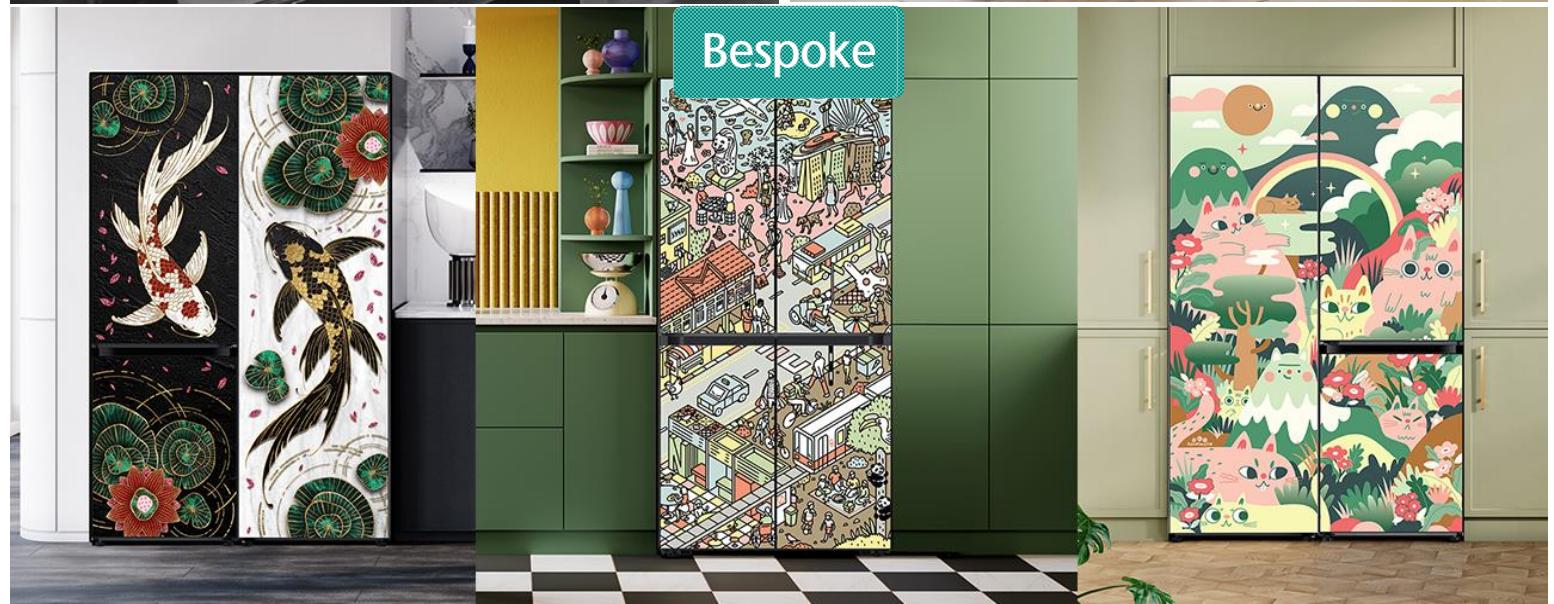


다품종 소량 생산



소비자 맞춤형 생산

- ✓ 취향 : 쇼핑, 유튜브
- ✓ 의료/재무/교육 : 진단?



## 1. ICT 기술의 영향과 발전

수단/도구로써의 인터넷

## 1996년 말, 과내 컴퓨터 동아리 창설

**“컴퓨터과외 받고 再起 자신감”**

서울대서 강의듣는 실직자모임 「이무기클럽」

매주 금요일 오후 6시면 서울공대 36동 공학연구소 지하1층에 마련된 작은 강의실에 모이는 「이무기 클럽」 회원 12명. 어린 서울대생들로부터 특별 과외수업을 받는 이들은 더 이상 「고개숙인 아버지」가 아니다. 「인터넷과 관련된 용어 10가지」의 뜻을 조사해 전자우편으로 보내달라고 했는데 안하신 분이 있네요.』 이진수(李鎮洙·27·서울대 산업공학과 4)군이 웃으며 말하자 『오늘 밤으로 마칠게요』라는 대답이 돌아왔다.

서울공대가 학부생·대학원

생들과 똑같은 내용의 강의를 들을 수 있도록 실직자들에게 개방한 「특별청강」을 듣는 수강생중 일부가 만든 「이무기 클럽」. 30년경력 기장출신의 유근무(柳根茂·61)씨, 무역회사에서 퇴직한 정해옥(丁海玉·34)씨 등 30~60대의 실직자들이 컴퓨터동아리 「용틀임」 회원들을 만난 건 지난 4월. 『용틀임 학생들에게 배우는 우리는 이무기 정도는 되지 않을까 싶어 그렇게 정했죠. 열심히 해서 새 길을 찾게 되면 승천하는 용이 되지 않을까요.』(클럽 총무 손남수·

나 컴퓨터 태. 『을 가을 나 만들 의 목소리 <木 chosun.com



◇서울대 공대의 컴퓨터 동아리 「용틀임」 회원들이 모임인 「이무기 클럽」 터를 가르치고 있다. <李基龍기자·krlee@chosun.com>

W Monroe St  
of Former Union Station Concourse

209 S LaSalle St  
Chicago, Illinois  
Google Street View - Aug 2021

1998년, 금융업과의 만남

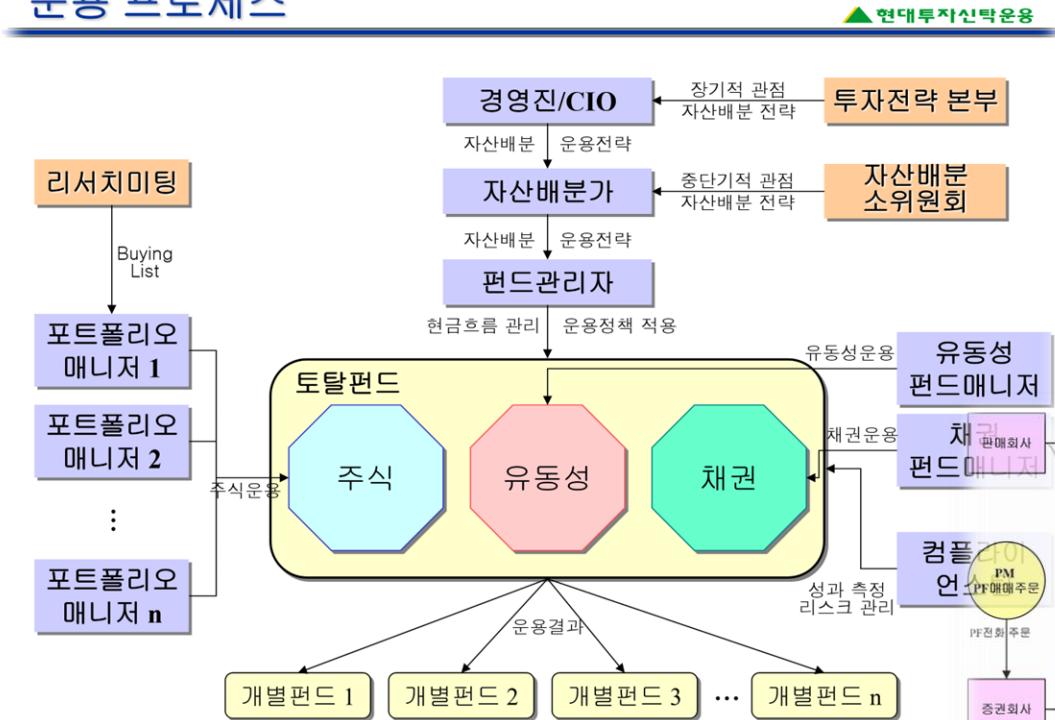


## 1. ICT 기술의 영향과 발전

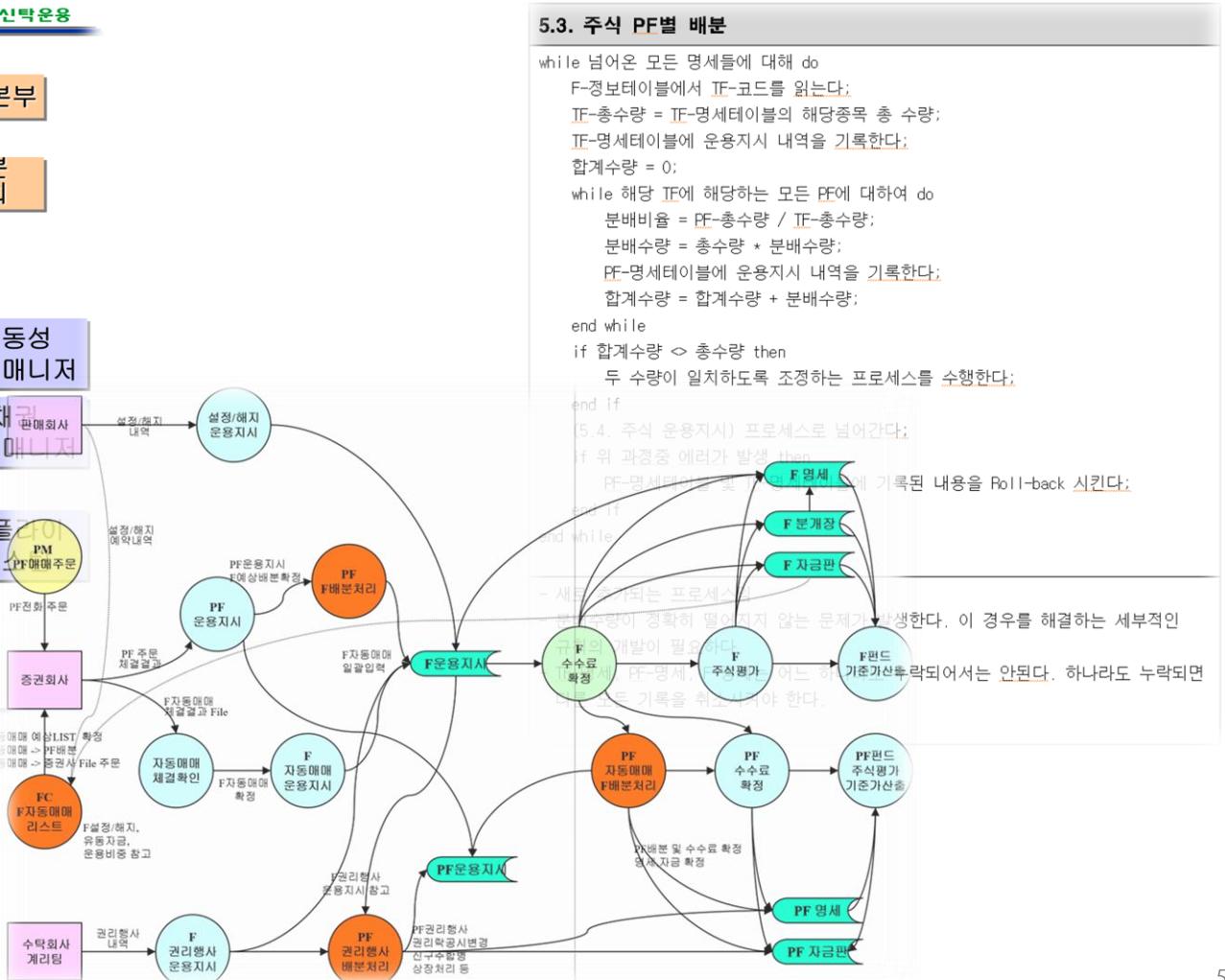
## 업무 프로세스 개선

# 입사하자마자 미들맨 역할로 믿을만 자리매김

## 운용 프로세스



## 2000년, 멀티매니저 시스템 구현





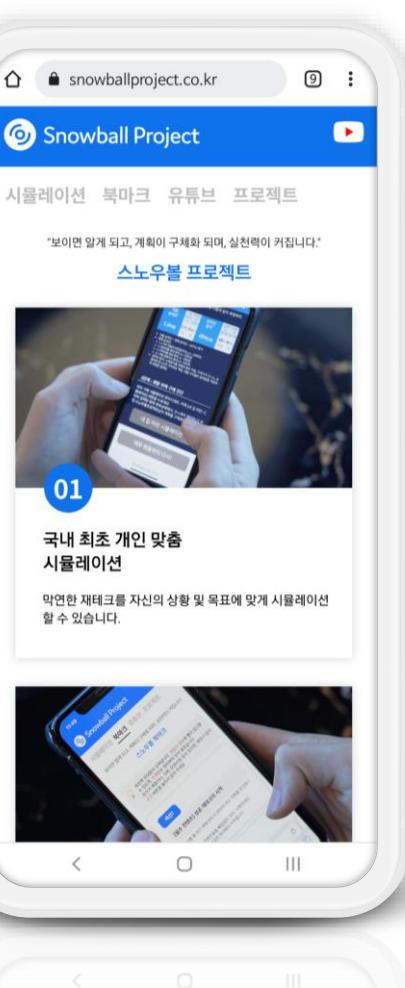
### 3. 금융업의 디지털 전환

#### 1. ICT 기술의 영향과 발전

고객 맞춤 개인화 서비스

## 대중을 위한 재테크 주치의 서비스를 꿈꾸다!

### 웹 앱



### 유튜브 채널



### 비대면 재무 컨설팅



1. ICT 기술의 영향과 발전

데이터와 인사이트를 연결해  
현업으로부터의 변화

고객과의 관계를 증진하고

CX를 중심으로 조직을 구축하는 방법

“더 나은 의사결정, 획기적인 효율성, 세분화된 개인화로 고객 경험 개선”  
을 위한 비즈니스 프로세스 변화

디지털화  
(Digitalization)

디지털 데이터를 사용하여 작업 방식 간소화

전자화된 정보를 사용하여 기존의 작업 방식을 더 간단하고 효율적으로 만드는 과정

정보전자화  
(Digitization)

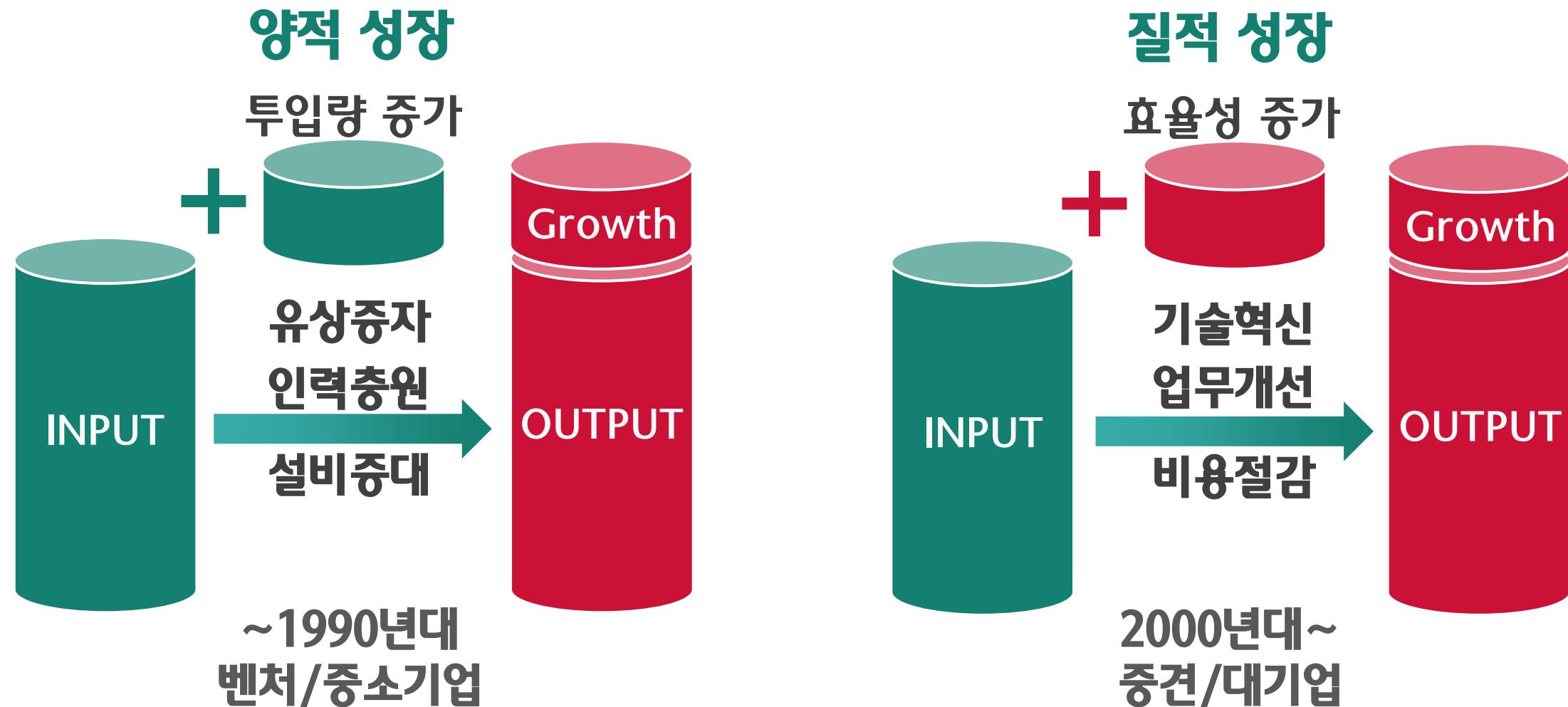
아날로그에서 디지털로의 이동

종이 문서를 통한 장부 기록에서 모든 종이 기록을 디지털 컴퓨터 파일로 변환

ICT 환경의 제공

## 1. ICT 기술의 영향과 발전

## 질적 성장 관점의 경영 환경



## 1. ICT 기술의 영향과 발전



**혁신(Innovation)은 어디서 오는가?**

**정보** ‘예상하지 못한 성공’에서  
가장 많은 혁신이 나온다!

**AI기술** ‘지식에 기초한 혁신’은  
무엇보다도 경영관리가 중요!

59 p

1. ICT 기술의 영향과 발전

## 대한민국 성장과 비즈니스 성공의 비결

영어는 일자리의 필수 무기



네트워킹은 사업 기회를 찾는 중요 원천



## 1. ICT 기술의 영향과 발전

# 디지털 리터러시(Digital Literacy)는 두 가지 모두에 해당



## 1. ICT 기술의 영향과 발전

# 디지털 리터러시는 현업에서의 변화를 이야기하는 것!

## 업무/서비스 개선 가능성 미인식

# 배운 신기술을 어디에 적용할까 고민



## 1. ICT 기술의 영향과 발전

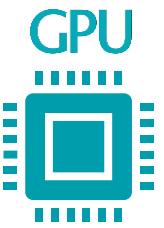
## 4차 산업혁명, 쏟아지는 용어들...



## 1. ICT 기술의 영향과 발전

## 컴퓨터 능력 활용의 범주

## H/W 및 기반 기술



응용 프로그램(App) 개발

## 기술 인프라 응용 기반기술



## 서비스 응용 기술

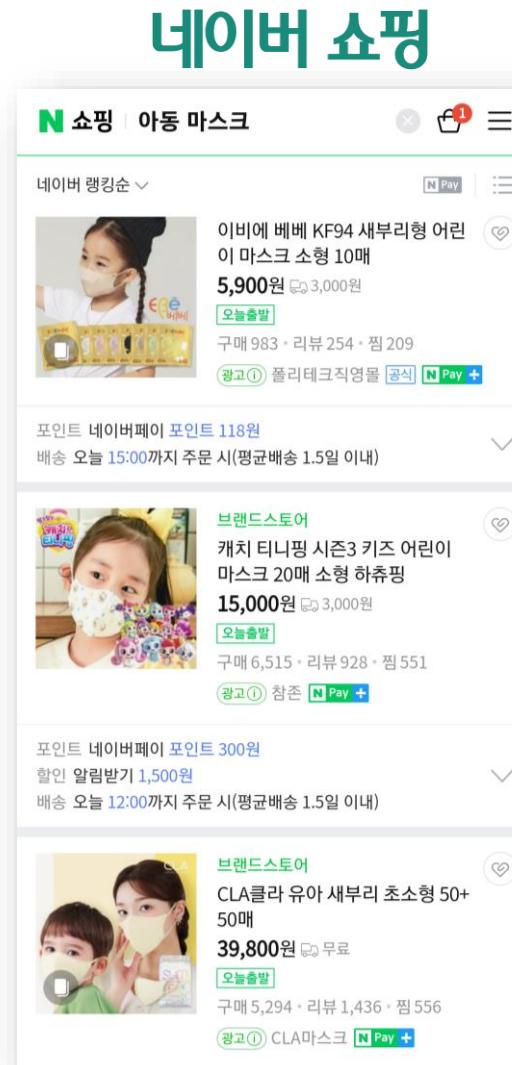


## 2. 금융업의 디지털 전환과 펀테크

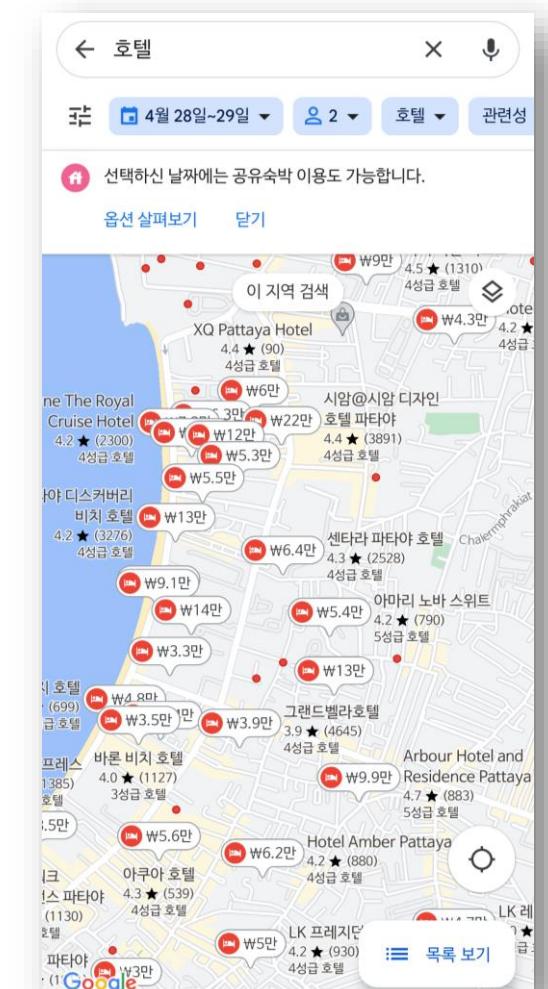
# 성공적 플랫폼이 제공하는 것

- ✓ 쉬운 접근
- ✓ 가격 비교
- ✓ 빠른 정보 획득
- ✓ 경험의 공유
- ✓ 거래 안정성

**금융 서비스는 ???**

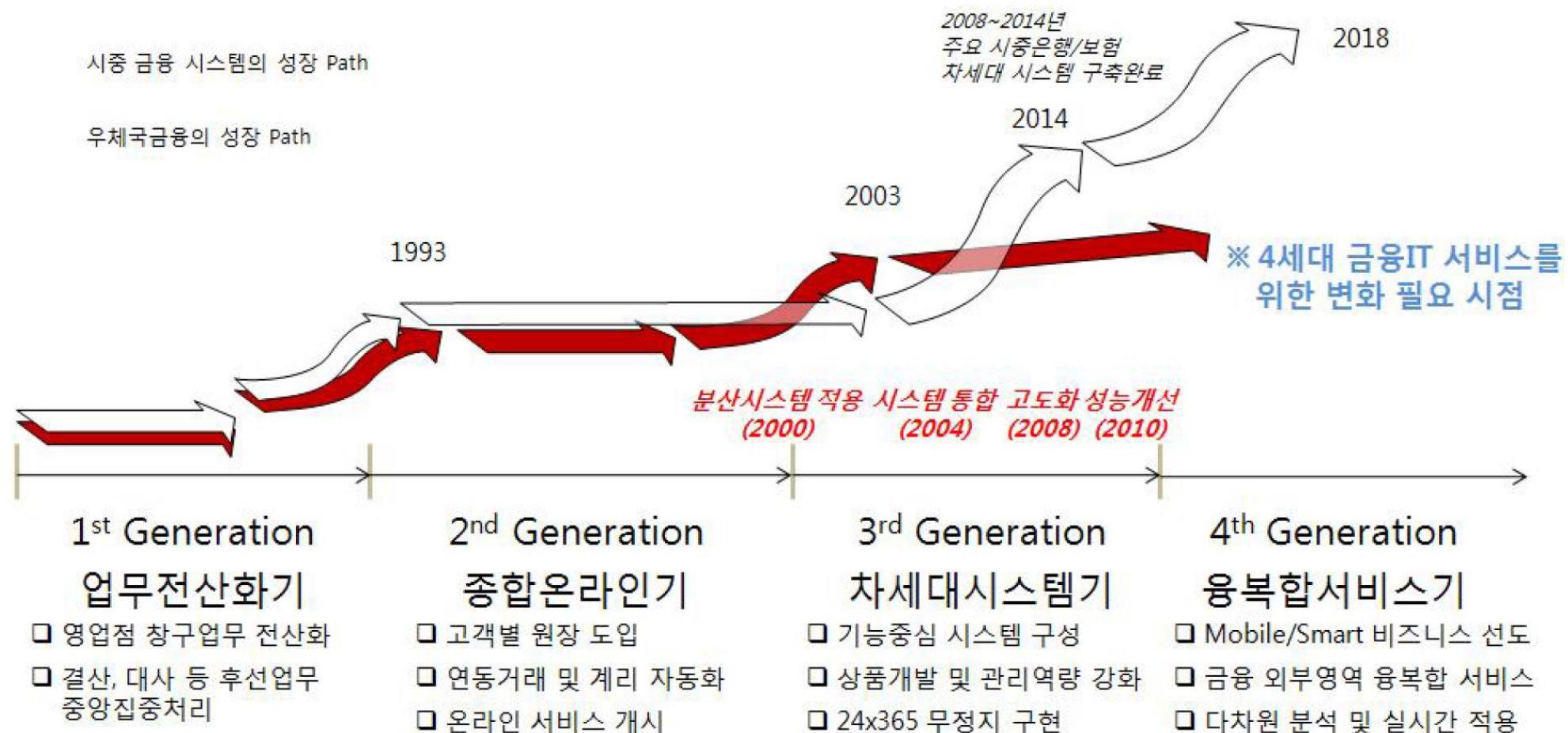


**구글 지도**



## 2. 금융업의 디지털 전환과 펀테크

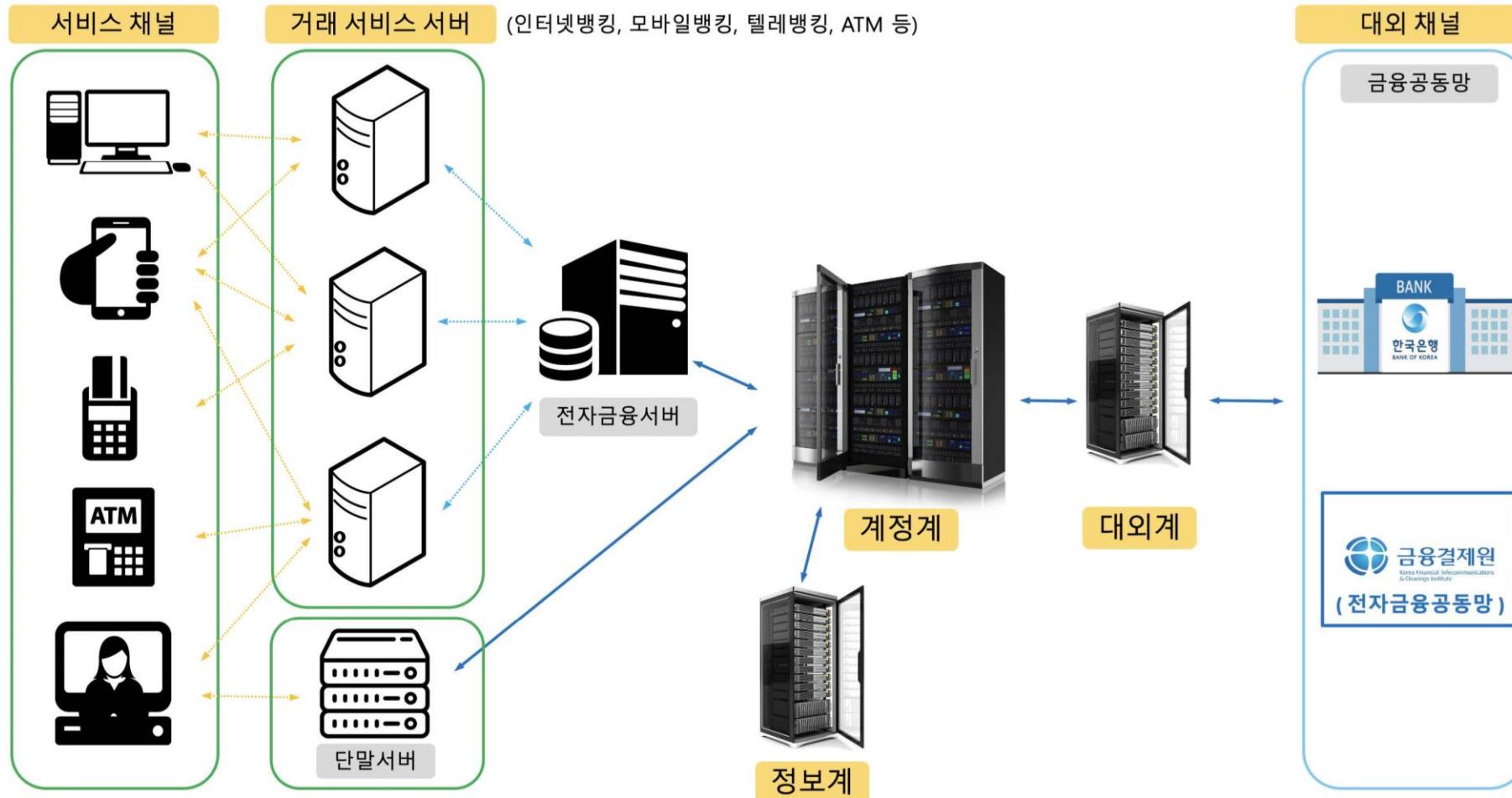
## 금융 ICT의 진화



[자료 출처] 이용수, 금융산업 환경변화에 대응한 우체국금융 IT 전략, 우정경영연구소, 2021.

## 2. 금융업의 디지털 전환과 펀테크

## 금융기관 전산시스템 개념도



### 3. 금융업의 디지털 전환

## 2. 금융업의 디지털 전환과 펀테크

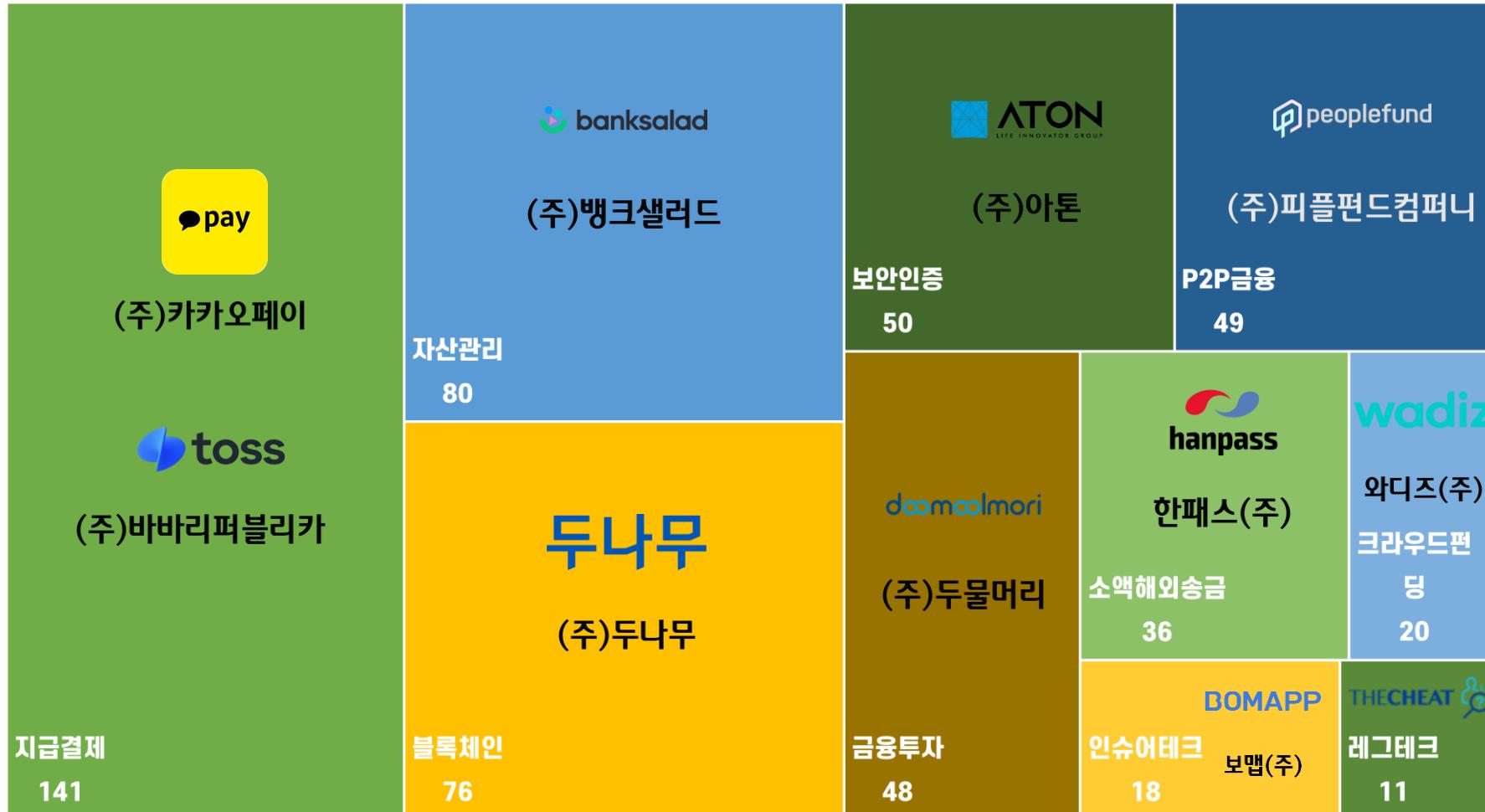
[자료 출처] 산업은행, 차세대 정보시스템 구축 제안요청서, 2016.

# 은행 차세대 시스템 단위업무 구성도



## 2. 금융업의 디지털 전환과 펀테크

## 펀테크 기업들의 기존 금융사 위협?



[자료 출처] 펀테크산업협회, 2023.12 현재 (분과별 중복 기업 존재)

## 펀테크 기업의 기존 금융권 진입



kakao



pay 증권

pay 손해보험

## 2. 금융업의 디지털 전환과 펀테크

## 기존 은행들의 디지털 전환 양상

### 프론트 오피스

고객과 거래가 이루어지는 영역

양상	기술	효과	예시
금융서비스 개인화 및 맞춤화	빅데이터	개인 맞춤화로 수요 창출	고객 금융 데이터를 기반으로 고객별 맞춤형 서비스 제공
AI 도입 본격화	머신러닝	새로운 비즈니스 모델을 통한 수익 창출	전통적 분류방식의 한계를 극복하기 위해 머신러닝을 활용한 ETF 출시
플랫폼화	오픈API	새로운 비즈니스 모델을 통한 수익 창출	여러 제휴사에 자사의 오픈 플랫폼을 제공하여 새로운 금융서비스 개발
플랫폼화	디지털 플랫폼	멀티 채널을 통한 고객 접점 강화	온라인 플랫폼을 중심으로 멀티 채널 제공 및 서비스 통합. 이를 통해 고객과의 접점 강화

## 2. 금융업의 디지털 전환과 펀테크

## 기존 은행들의 디지털 전환 양상

### 미들 오피스

영업지원, 위험관리 영역

양상	기술	효과	예시
핀테크 기업과의 제휴	디지털 플랫폼	포트폴리오 통합관리 및 모니터링	다수 금융사로의 포트폴리오 분산관리 중인 초부유층 투자자 대상 통합 플랫폼 제공
핀테크 기업과의 제휴	머신러닝	리서치 역량 강화 및 서비스 비용 감소	고객 포트폴리오 통합 관리 및 모니터링이 가능한 투자 관리 플랫폼 펀테크 스타트업에서 제공받음
핀테크 기업과의 제휴	빅데이터 머신러닝	리서치 역량 강화	시장 이벤트가 금융시장에 미치는 영향력에 관한 인공지능 금융분석 서비스 제공
레그텍(RegTech) 투자 확대	머신러닝	서비스 비용 감소 및 업무 효율화	AI 및 빅데이터 기반 약관심사 솔루션 도입, 수만 건의 문서를 자동으로 해독. 규정 위반사항 진단

## 2. 금융업의 디지털 전환과 펀테크

## 기존 은행들의 디지털 전환 양상

### 백 오피스

거래 사후관리, 회계/인사 등 경영지원 영역

양상	기술	효과	예시
핀테크 기업과의 제휴	머신러닝	자금세탁 방지 이행 효율화	핀테크 기술을 활용하여 자금세탁 및 테러자금 등을 감지, 예방하는 솔루션 제공 받음
핀테크 기업과의 제휴	머신러닝	내부 통제시 비용 및 시간 감소	핀테크 AI 기술을 활용하여 직원의 이메일, 메신저, 전화 등을 다각도로 모니터링 하고, 이를 기반으로 내부자 거래 등을 포착하는 서비스 제공 받음
내부통제 효율화	빅데이터	내부 통제시 비용 및 시간 절감	내부 직원의 인터넷, 이메일, 전화통화 기록 등을 분석하여 비리 혐의 포착
핀테크 기업과의 제휴	블록체인 오픈 API	청산결제 서비스 제공 탈중앙화	블록체인 기반 청산결제 서비스 활용 계획 오픈 API를 통한 클라우드 활용 청산결제 상용화

2. 금융업의 디지털 전환과 펀테크

## 고객의 니즈와 구매 사이클 변화

Know Me

내가 누구인지, 무엇이 필요한지 알아 주세요

Advise Me

필요한 시점에 적합한 정보와 조언을 주세요

Compare Me

내가 잘 하고 있는지 다른 사람과 비교해 주세요

Educate Me

수익성, 리스크, 세금... 좀 가르쳐주세요

Let Me Choose

무엇을 선택할 수 있는지 제시하고 비교해주세요

Make Me Easy

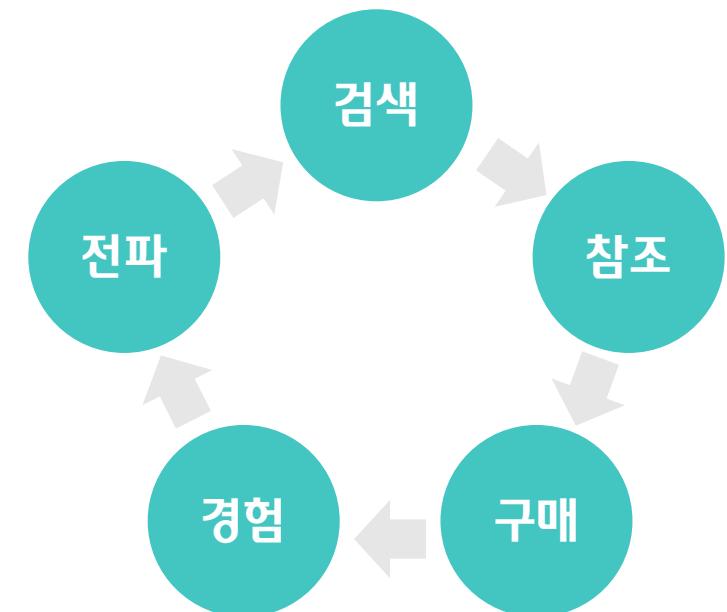
거래를 편하고 쉽게 처리해 주세요

Make Me Fun

지루하지 않도록 즐겁게 해 주세요

Protect Me

나의 거래와 데이터를 안전하게 지켜주세요



2. 금융업의 디지털 전환과 펀테크



### 3. 금융업의 과제



## 3. 금융업의 과제

## 빅뱅러와 규제 장벽 완화

### Exhibit F. 핀테크 유형별 '금융 혁신' 접근 방법

  
기존의 방식에서 탈피하여 새로운 관점으로 접근

고객 중심 철학 기반 불편요소 선제적 발견 및 해결

[사업모델] 디지털 플랫폼 기반  
고객 대상 서비스 제공 (제조-유통)  
예. Revolut, N26, Toss



오프라인 지점 및 전문인력  
기반으로한 레거시 금융사업에서 시작



디지털/테크 기반의 비금융 사업을  
영위하다 고객의 니즈와 연계된  
금융 영역으로 진출

기존 사업 생태계와 밀접하게  
연계된 금융 혁신 시도

[사업모델] 오프라인 고객 점점을  
기반으로 쉽게 접할 수 있는  
서비스로부터 확장 (예. 결제)  
예. Google, Facebook, Amazon, Naver

[사업모델] 기존 금융사업에 디지털/테크  
요소 접목한 토탈 금융 솔루션 제공  
예. Goldman Sachs, JP Morgan, 기존 금융사

## 3. 금융업의 과제

## 디지털 경제 시대 뱅킹서비스 방향성

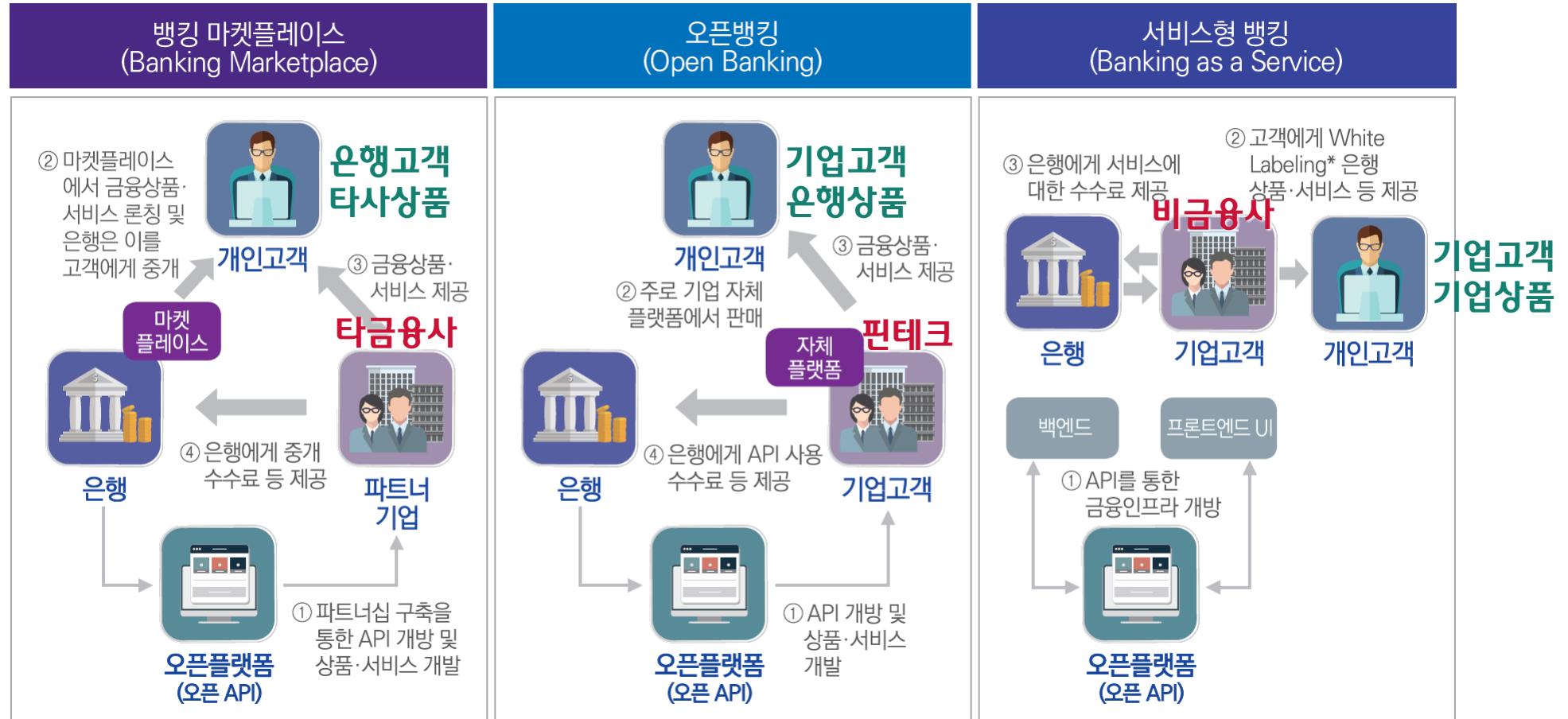
- 맞춤화된 금융 서비스 제공 및 데이터 역량을 극대화할 수 있는 디지털 신기술 도입
  - 데이터 전문가 및 디지털 인프라 확보 필요
  - 고객 데이터·오픈 API 활용 등 미래 수익을 창출할 수 있는 방안 검토
  - 마켓플레이스 등 자사 디지털 전략·여건에 적합한 플랫폼화 전략 수립
- 데이터 기반 신사업 기회  
“데이터 역량·기술 자산화”
- 비금융과 연계  
“전략적·기능적 파트너십 확대”
- 금융 서비스 리번들링  
“원스톱 금융서비스”
- 뱅킹의 플랫폼화  
“오픈플랫폼 활성화”
- 데이터 기반 기능적 관점의  
“금융서비스 초기인화”



[자료 출처] 삼정KPMG 경제연구원,  
은행산업에 펼쳐지는 디지털 혁명과 금융 패권의 미래, 2021

## 3. 금융업의 과제

# 오픈 API 기반 뱅킹 플랫폼 (B2B)



[자료 출처] 삼정KPMG 경제연구원, 은행산업에 펼쳐지는 디지털 혁명과 금융 패권의 미래, 2021

## 3. 금융업의 과제

# 데이터 기반 가치 제공

고객 유형에 따른 맞춤형 서비스

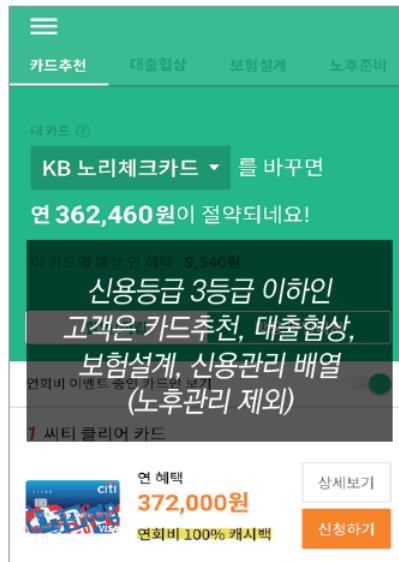
개인 맞춤화

초개인화

고유한 개인으로서의 맞춤형 서비스



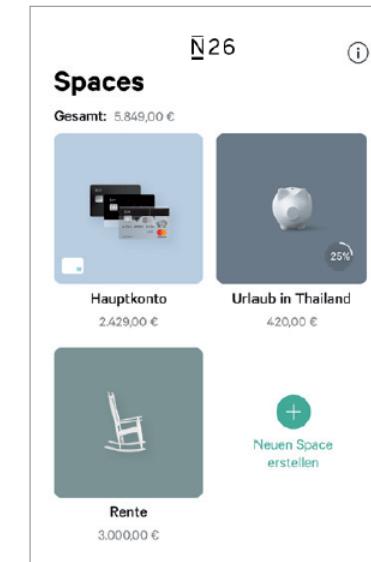
고객에 따라 달라지는 메뉴



고객 등급별 대출상품 Listing



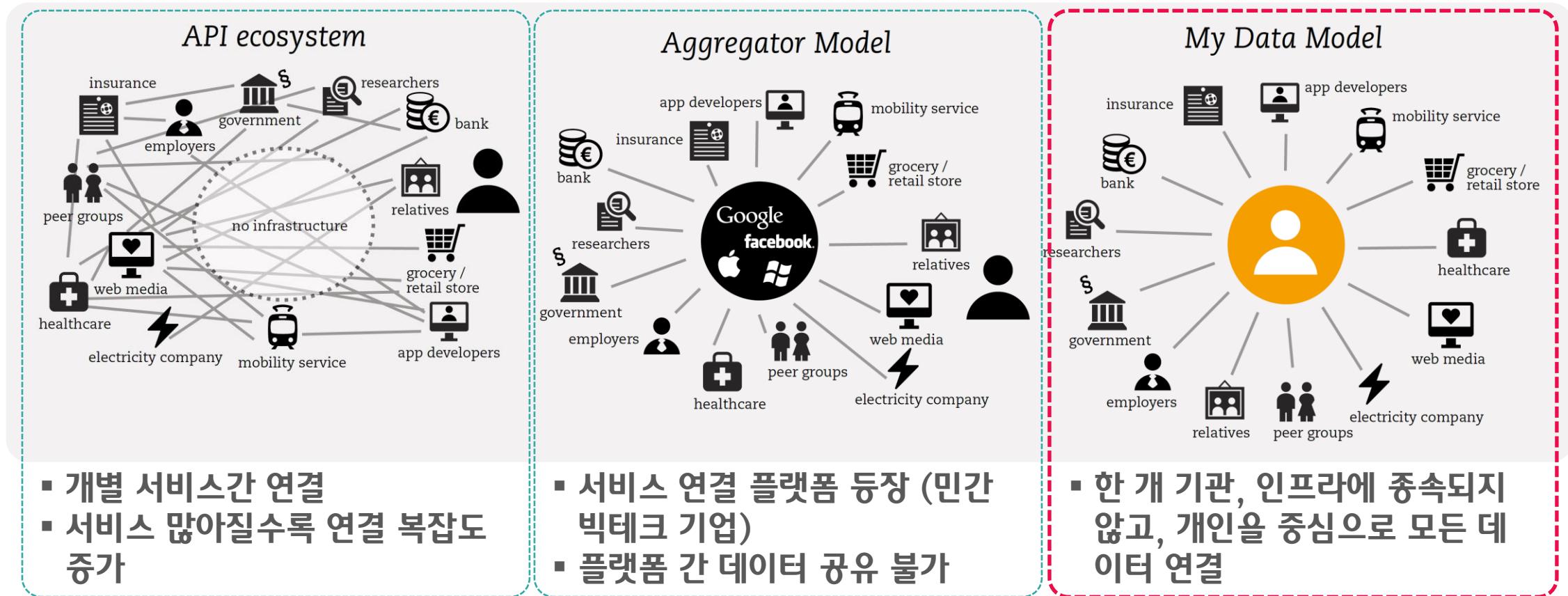
소비 내역 기반 목적성 자금 범주화



## 3. 금융업의 과제

# 마이데이터 개념과 도입 배경

## 효율성과 개인의 데이터 통제 최적 모델



### 3. 금융업의 과제

[자료 출처] 마이데이터 종합포털 (mydatacenter.or.kr)

## 마이데이터란

마이데이터의 제도적 기반은 [개인신용정보 전송요구권](#) (신용정보법 제33조의2) 입니다.



고객이

본인에 관한  
개인신용정보를



금융회사로부터

- 신용정보 제공·이용자  
([금융회사](#) 등)
- 공공기관
- 본인신용정보관리회사  
([마이데이터사업자](#))



마이데이터사업자에게

- 신용정보주체 본인(고객)
- 본인신용정보관리회사
- 신용정보 제공·이용자
- 개인(사업자)
- 신용평가회사 등

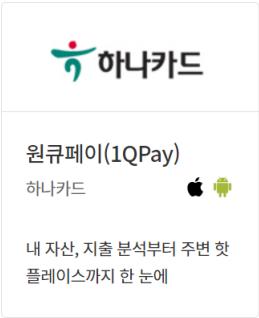
‘전송하도록 요구할 수 있는 권리’

## 3. 금융업의 과제

[자료 출처] 마이데이터 종합포털 (mydatacenter.or.kr), 2024년 7월 현재

# 마이데이터 범위와 서비스 현황

35

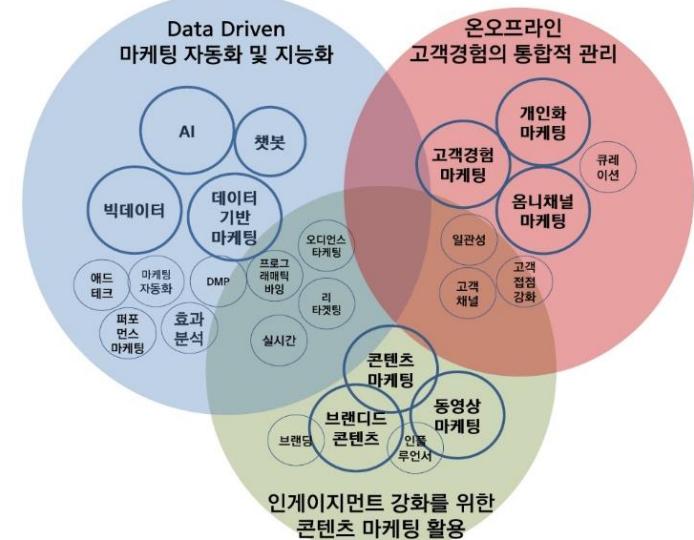
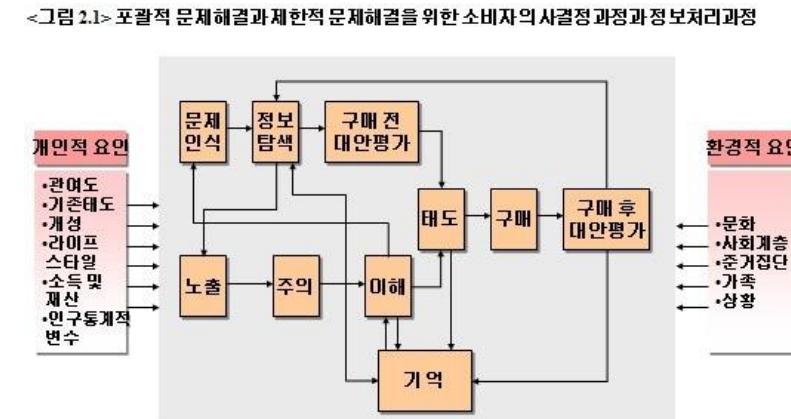
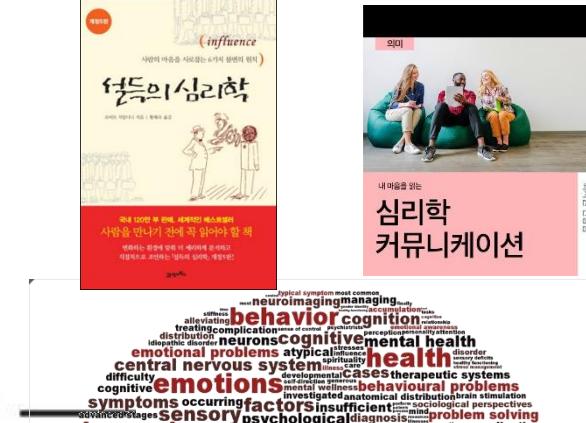
은행  
11보험  
4증권  
10여신  
10저축은행  
2

30

핀테크  
23신용평가  
2통신  
3유통  
1IT  
1

### 3. 금융업의 과제

# 인간에 대한 이해를 바탕으로 하였는가?



### 3. 금융업의 과제

## 넛지(Nudge); 행동 경제학 활용 재무관리의 중요성을 알지만 실천하기는 어렵다!

현재의 금융 서비스

5%



- 이성적, 반성적 사고 활동
- 의식된 행동, 변형 가능
- 결정 속도 느림, 멀티 태스킹 어려움
- 사용에 피로감과 부담감이 있음

미래의 금융 서비스

95%

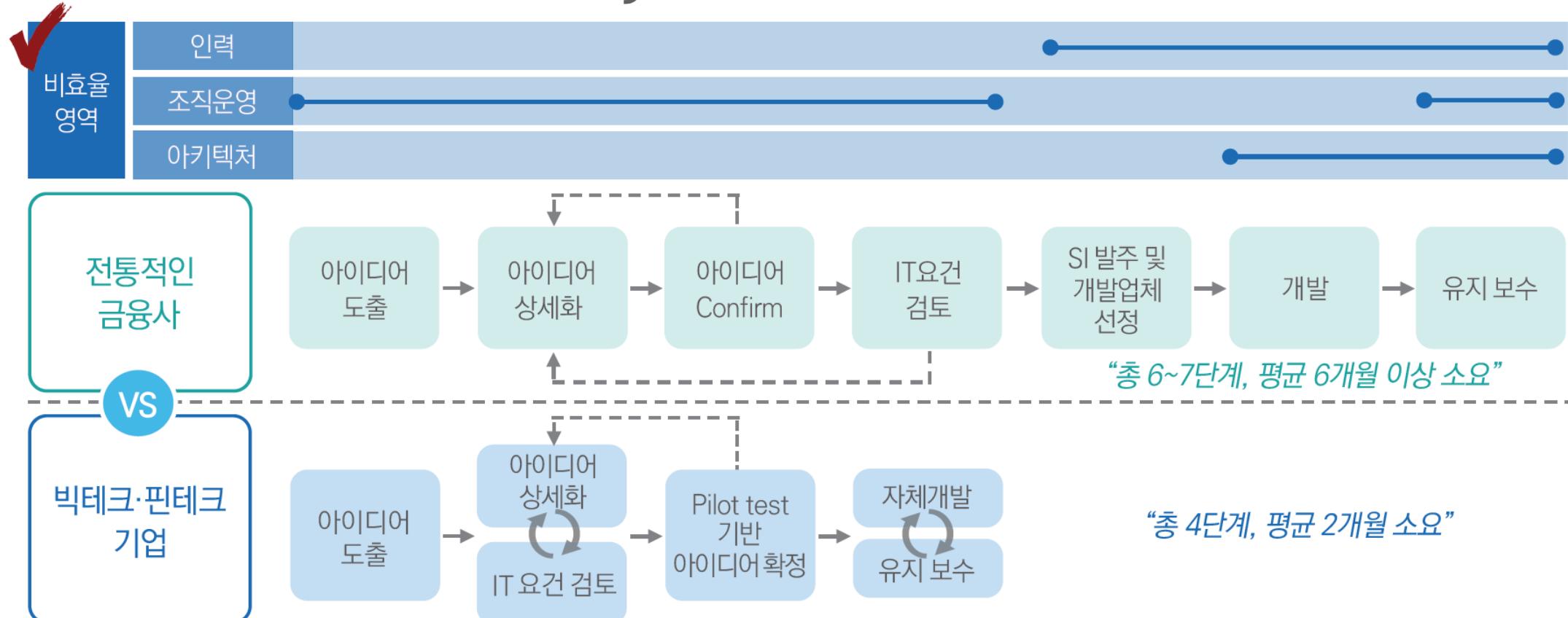


- 휴리스틱 : 본능적, 습관적 행동
- 거의 의식되지 않고 자동 적용
- 결정 속도가 빠름
- 사용에 피로감이 거의 없음

## 3. 금융업의 과제

# 은행의 애자일 조직 운영과 DevOps 개발자 양성

## 금융업계 주요 Player간 신규 서비스 개발 프로세스 비교



[자료 출처] 삼정KPMG 경제연구원, 은행산업에 펼쳐지는 디지털 혁명과 금융 패권의 미래, 2021

## 핵심요약

- ICT 기술은 ‘소비자 맞춤형 생산’ 단계로 이끌었고, ‘디지털 전환’의 성공은 소비자 접점에서의 정보 활용에 달려 있음
- ‘디지털 리터러시’는 현업 지식과 기술이 분리될 수 없음을 의미
- 현재 금융사 전산 시스템은 ‘계정계’, ‘채널계’, ‘대외계’, ‘정보계’ 등으로 구성되어 있으며, ‘정보계’와 ‘대외계’의 중요성이 커지고 있음
- 핀테크 기업들의 침투로 기존 금융권은 위협을 느끼다 협업 체계로 가는 중
- 빅테크 기업의 금융권 진출이 가장 큰 위협이며, ‘빅블러’ 추세에 오픈 API를 통한 ‘마켓플레이스’, ‘오픈뱅킹’, ‘서비스형 뱅킹’으로 확장 추세
- 맞춤형, 초개인화 서비스를 위한 ‘マイ데이터’ 활용 방안이 필요하며, 고객 중심 서비스 개발을 위한 애자일 개발 조직 운영이 요구됨

## 4. 금융 데이터 분석을 위한 재무 기초

---

# 학습목표

- ▶ 금융에서 다루는 핵심 숫자인 금리(이자율)와 수익률을 정확히 이해
- ▶ 통계 모델인 주요 재무이론을 통계적 데이터 분석 관점에서 이해

## 1. 금리(이자율)와 수익률

빌라



보유 자금 1억원

매매가격 1.2억원

임대보증금 2천만원

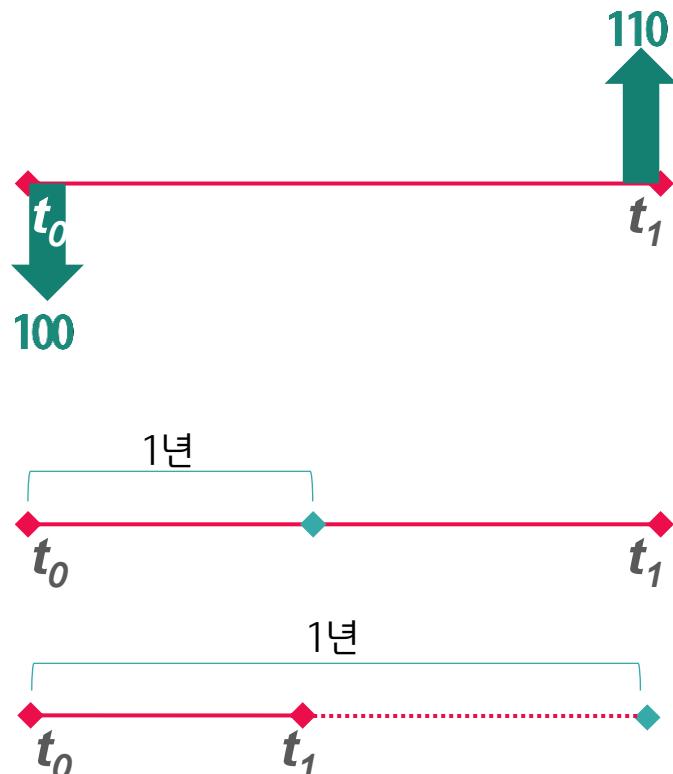
월세 연간 480만원

현금흐름 수익률 연 4.8%

무위험 수익률 연 2.5%

## 1. 금리(이자율)와 수익률

## 수익률 계산 방법



기간 개념 없음

### 수익률

$$r = (p_1 - p_0) / p_0$$

$$= p_1 / p_0 - 1$$

$$= 110 / 100 - 1 = 10\%$$

### 연환산 수익률

기간은 무조건 연 단위

Simple Interest

### 단리

$$r_s = r / (t_1 - t_0)$$

산술평균 : 연수( $n$ )로 나눔

Compound Interest

### 복리

$$r_c = (1 + r)^{1/(t_1-t_0)} - 1$$

기하평균 : 연수( $n$ )의  $n$ 제곱근

### 예시

$(t_1 - t_0)$	2년	6개월(0.5년)
단리	5.0%	20%
복리	4.9%	21%

## 1. 금리(이자율)와 수익률

## 기간 수익률 연결 구조 (시계열 분석)

**복리는 구간 수익률을 곱하나, 로그 수익률은 더하는 것**

$$r_{total} = (1 + r_1) \times (1 + r_2) \times (1 + r_3) \times \dots \times (1 + r_n) - 1$$

$$1 + r_{total} = (1 + r_1) \times (1 + r_2) \times (1 + r_3) \times \dots \times (1 + r_n)$$

where,  $r_1 = p_1/p_0 - 1$ ,  $r_2 = p_2/p_1 - 1$ ,  $r_3 = p_3/p_2 - 1$ , ...,  $r_n = \boxed{p_n/p_0 - 1}$

**산술수익률  
(AR)**

$$\ln(1 + r_{total}) = \ln[(1 + r_1) \times (1 + r_2) \times (1 + r_3) \times \dots \times (1 + r_n)]$$

$$LR = \ln(1+AR)$$

$$= \ln(1 + r_1) + \ln(1 + r_2) + \ln(1 + r_3) + \dots + \ln(1 + r_n) \quad AR = e^{LR} - 1$$

$$= \ln(p_1/p_0) + \ln(p_2/p_1) + \ln(p_3/p_2) + \dots + \ln(p_n/p_{n-1})$$

$$= \boxed{\ln(p_n/p_0)}$$

**로그수익률  
(LR)**

## 1. 금리(이자율)와 수익률

## 포트폴리오 수익률 가중 구조 (동시대 분석)

**동시대(Cross-section)에 대한 분석은 산술 수익률 사용**

구간 t1, 구간 t2 전체 수익률

$$\begin{aligned} r_{total} &= (1 + r_{t1}) \times (1 + r_{t2}) - 1 \\ &= (1 + 1) \times (1 - 0.5) - 1 \\ &= 2 \times 0.5 - 1 = 0 \end{aligned}$$

종목 p1, 종목 p2 분산 투자시 전체 수익률

$$\begin{aligned} r_{total} &= w_{p1} \times r_{p1} + w_{p2} \times r_{p2} \\ &= 0.5 \times 1 + 0.5 \times (-0.5) \\ &= 0.5 - 0.25 = 0.25 \end{aligned}$$

기간	수익률
t1	100%
t2	- 50%
Total	<b>0%</b>

포트폴리오	투자비중	수익률
p1	50%	100%
p2	50%	- 50%
Total	100%	<b>25%</b>

## 1. 금리(이자율)와 수익률

수익률 계산은 돈 계산			
	현재	1년 뒤	(기대)수익률
예금	<b>1천만원</b> 이자율 : 연 3% 중도해지시연 1%	<b>1천만원</b> $+ 30\text{만원}$	<b>연 3%</b>
채권	<b>1천만원</b> 3개월마다 이자 1% 지급 만기(1년)에 액면가 1천만원 상환	<b>1천만원</b> $+ 40\text{만원}$	<b>연 4%</b>
주식	<b>1천만원</b> 시가총액 1천억원, 순자산 8백억원 매출액 2천억원, 영업이익 1백억원	<b>1천만원</b> $+ 100\text{만원}$	<b>연 10%</b>

## 1. 금리(이자율)와 수익률

## 수익률 계산은 돈 계산

	현재	1년 뒤	(기대)수익률
예금	<b>300만원</b> 이자율 : 연 3% 중도해지시연 1%	<b>300만원</b> $+ 9\text{만원}$	<b>연 3%</b>
채권	<b>300만원</b> 3개월마다 이자 1% 지급 만기(1년)에 액면가 300만원 상환	<b>300만원</b> $+ 12\text{만원} + \alpha$	<b>연 4% + \alpha</b>
주식	<b>300만원</b> 시가총액 1천억원, 순자산 8백억원 매출액 2천억원, 영업이익 1백억원	<b>300만원</b> $+ 30\text{만원}$	<b>연 10%</b>

## 1. 금리(이자율)와 수익률

## 시간에 내재된 돈

	1년 전	현재	1년 뒤
국채	100만원	100만원	100만원
1년 전 발행 1년 뒤 만기		+이자 5만원	+이자 5만원

**이자** 남에게 돈을 빌려 쓴 대가로 치르는 돈

[필수 조건] 원금을 돌려주는 날과 이자를 지급하는 날 포함 (정확한 기간)

**이자율** 원금에 대한 이자의 비율

=이자/원금 = 5만원/100만원 = 5% (1년에)

## 1. 금리(이자율)와 수익률

# 시간에 내재된 돈

이자

남에게 돈을 빌려 쓴 대가로 치르는 돈

[필수 조건] 원금을 돌려주는 날과 이자를 지급하는 날 포함(정확한 기간)

이자율

원금에 대한 이자의 비율

=이자/원금 = 5만원/100만원 = 5% (1년에)

= 금리

## 대가의 균형점

수요 : 빌리는 사람(차입자) 입장

지금 당장 돈을 쓸 수 있게 해 주는 것 (투자 기회 또는 근시안적 소비)

공급 : 빌려주는 사람(대여자) 입장

지금 돈을 쓰는 것을 포기한 대가 (소비 이연 또는 저축)

돈을 제대로 못 돌려 받을 위험에 대한 대가 (부도, 상환 유예 가능성)

## 1. 금리(이자율)와 수익률

## 과거와 미래의 수익률

	1년 전	현재	1년 뒤
국채	<b>100만원</b>	<b>100만원</b>	<b>100만원</b>
1년 전 발행 1년 뒤 만기		<b>+수익 5만원</b>	<b>+수익 5만원</b>

**수익**      경제활동의 **대가로 얻은 돈**

이자는 수익의 한 종류

**수익률**      **자본금(원금)에 대한 수익의 비율**

$$= \text{수익}/\text{원금} = 5\text{만원}/100\text{만원} = 5\% \text{ (1년에)}$$

1. 금리(이자율)와 수익률

## 과거와 미래의 수익률

	1년 전	현재	1년 뒤
<b>주체</b>	<b>100만원</b>	<b>110만원</b>	<b>???</b>
	+수익 10만원	+수익 ???	

**수익률**      실현된 **과거**의 수익률  
= 수익/원금 = 10만원/100만원 = 10%

## 1. 금리(이자율)와 수익률

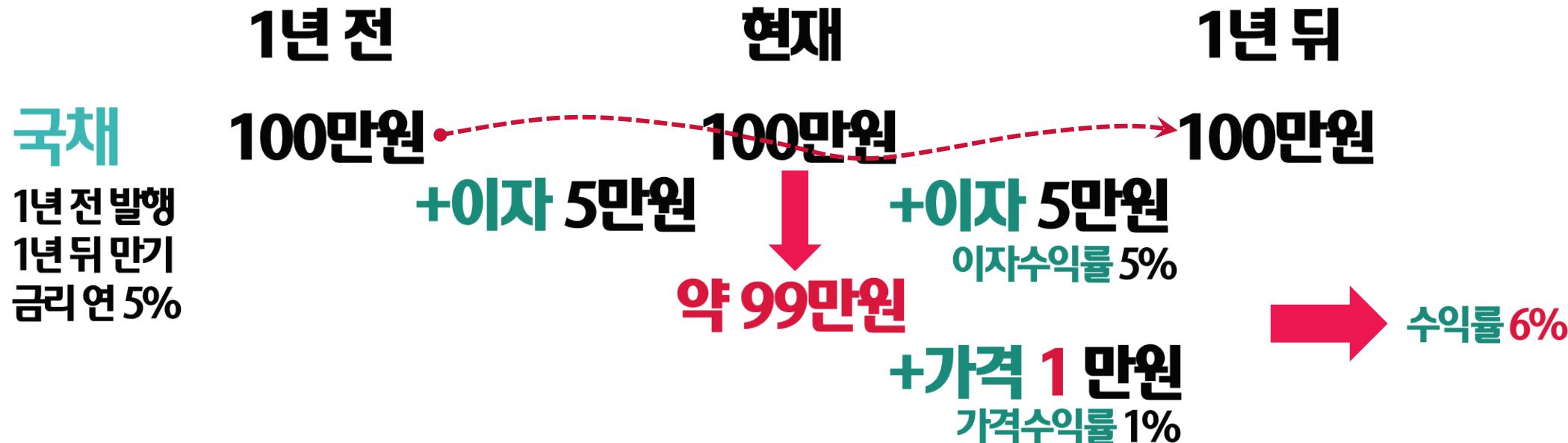
## 과거와 미래의 수익률

	1년 전	현재	1년 뒤
주식	100만원	110만원 수익 10만원	121만원 기대수익 11만원
수익률	실현된 과거의 수익률	= 수익/원금 = 10만원/100만원 = 10%	
기대수익률	예상(기대)하는 미래의 수익률	= 기대수익/원금 = 11만원/110만원 = 10%	
제시수익률	기대수익률의 예시		

기대수익률을 이해하기 어려우니까, 구체화 시켜 표현한 것!  
기대수익률은 각자 다른 것. 절대 제시수익률로 실현되지 않음!

## 1. 금리(이자율)와 수익률

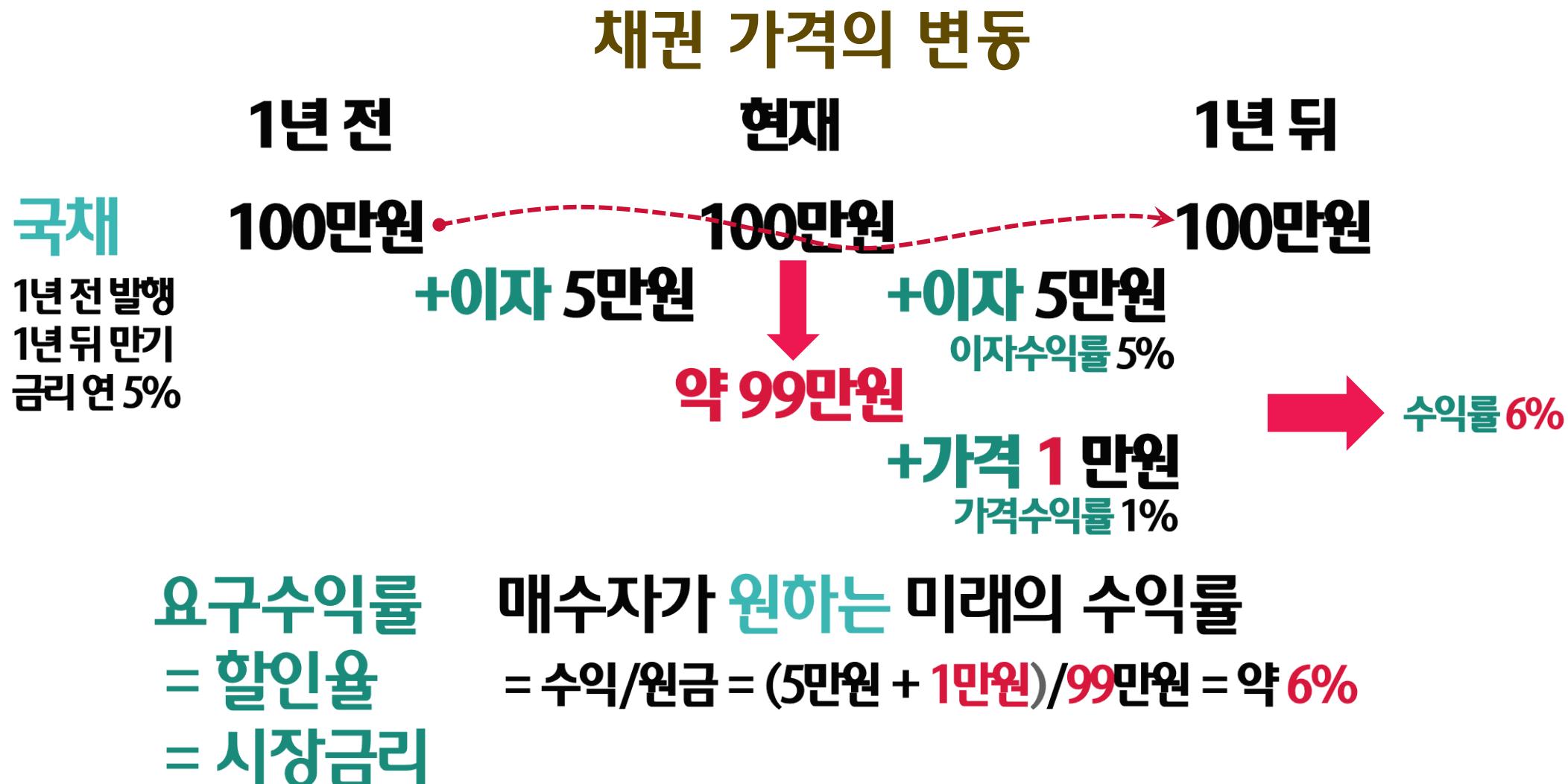
## 채권 가격의 변동



**국채**  
오늘 발행  
1년 뒤 만기  
금리 연 6%

**100만 원**  
**+이자 6만 원**  
**수익률 6%**

## 1. 금리(이자율)와 수익률



1. 금리(이자율)와 수익률

## 가격에 대한 태도

### 기대수익률 평가 수익률

현재 가격에서 미래를 예측하는 수익률

### 요구수익률 매수 수익률

살 의향이 있는 기대수익률  
• 가격 변화의 원인

기대수익률 > 요구수익률  
기대수익률 < 요구수익률  
기대수익률 = 요구수익률

현재 가격 상승  
현재 가격 하락  
가격 균형

1. 금리(이자율)와 수익률

## 가격에 대한 태도

기대수익률 > 요구수익률

현재 가격 **상승**

기대수익률 < 요구수익률

현재 가격 **하락**

기대수익률 = 요구수익률

가격 **균형**

$$\text{요구수익률} = \text{금리} + \text{위험대가}$$

금리 하락

현재 가격 **상승**

금리 상승

현재 가격 **하락**

위험 하락

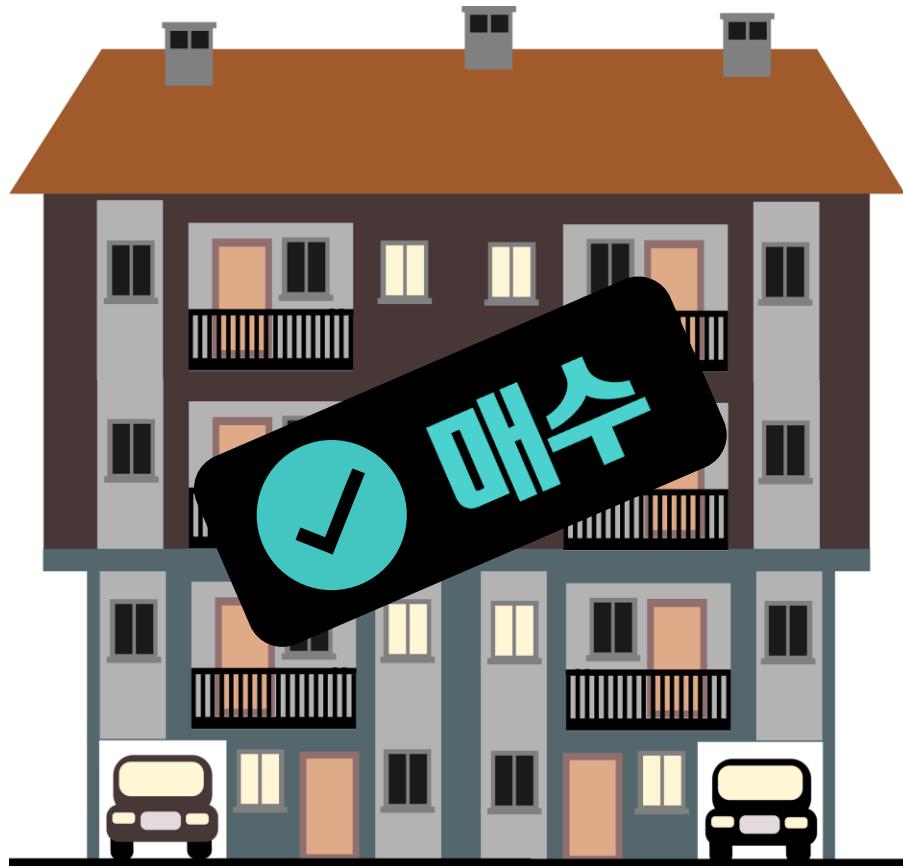
현재 가격 **상승**

위험 상승

현재 가격 **하락**

## 1. 금리(이자율)와 수익률

빌라



보유 자금 1억원

매매가격 1.2억원

임대보증금 2천만원

월세 연간 480만원

기대 수익률 연 5.0%

무위험 수익률 연 2.5%

요구 수익률 연 4.5%

가격변동 위험 (위험 프리미엄) 2.0% 감안

## 1. 금리(이자율)와 수익률

빌라



보유 자금 1억원

매매가격 1.2억원

임대보증금 2천만원

월세 연간 480만원

기대 수익률 연 4.5%

무위험 수익률 연 3.5%

요구 수익률 연 5.5%

가격변동 위험 (위험 프리미엄) 2.0% 감안

1. 금리(이자율)와 수익률

**수익률 = 돈** (원금과 무관한…)

**(1) 남의 돈을 빌려 쓴 대가**

**이자(율) = 금리**

**(2) 정해져 있거나 과거의 수익 반영**

**수익률**

**(3) 알 수 없는 미래의 예상 수익률**

**기대수익률**

**(4) 자산을 구매할 의향이 있는 수익률**

**요구수익률 = 할인율**

**(5) 현재 자산 가격의 균형점**

**기대수익률 = 요구수익률**

2. 재무 이론의 통계적 해석

## 회색인 세상에서 유능한 사람?

### 분석적 능력

대상의 회색 정도를 규명

### 직관력 (실행력; 설득력)

회색을 흰색 또는 검은색으로 판단



### 태도(Attitude)

타인으로부터의 협조와 타인에 대한 설득에 영향

## 2. 재무 이론의 통계적 해석

# 데이터 분석 프로세스와 필요 역량

## 데이터 획득



- 텍스트 파일 : CSV, TXT
- Open API : JSON, XML
- 내부 데이터베이스 : SQL
- 웹 (공공 데이터) : XLS, 텍스트
- 웹 페이지 : HTML(크롤링, 스크레이핑)

## 처리/분석 도구



- 데이터 입력/출력
- 데이터 가공/정제 : 셀, 시트, 함수
- 데이터 시각화 : 차트, 피벗 테이블
- 데이터 분석 : 함수, 데이터분석 도구, 해찾기 도구 등
- 보고서 출력 : XLS, PDF

## 분석 방법론



- 시각적 방법
- 통계적 방법 : 기술통계, 회귀분석, 판별분석 등
- 머신 러닝 : 클러스터링, 딥 러닝 등

2. 재무 이론의 통계적 해석

## 데이터 안에 숨겨진 패턴 찾기 (시각화)

### 실마리 찾기

#### 차트 (Charts)



### 설득하기

#### 인포그래픽 (Infographics)



## 2. 재무 이론의 통계적 해석

# 데이터 안에 숨겨진 패턴 찾기 (모델링)

## 결정론적 세계관

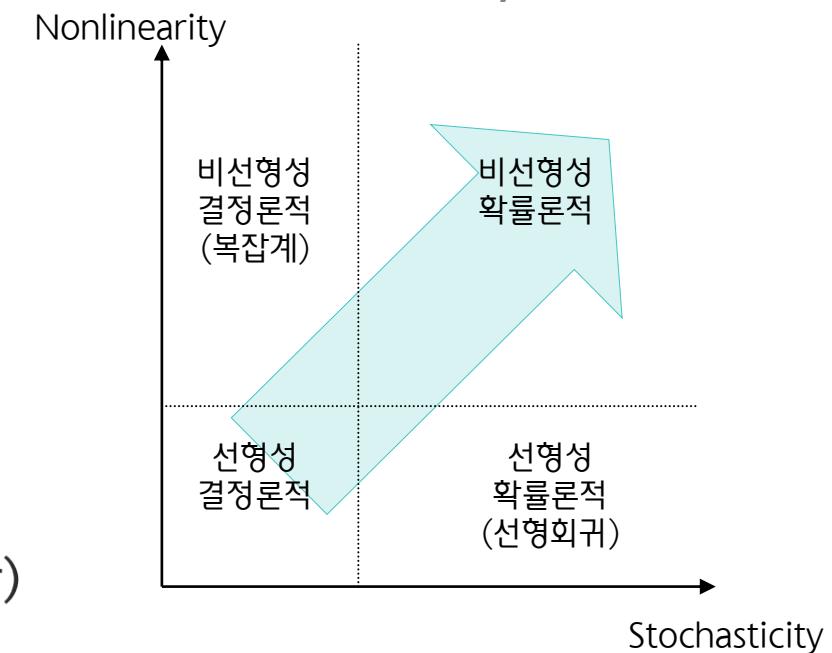
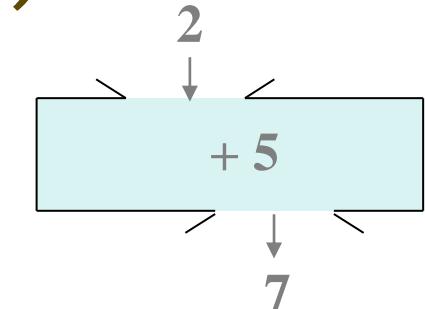
- 함수식 :  $F = ma$  로 대표되는 뉴튼 역학의 세계관
- 모든 걸 알 수 있음 : 10m 높이에서 떨어뜨린 물체의 1초 후 위치
- 최적해 도출 : 수요곡선과 공급곡선이 만나는 점

## 확률론적 세계관

- 불확정성의 원리 : 결정론적 세계관의 붕괴
- 가격 결정 : 미래에 일어날 수 있는 현금흐름의 현재가치? (확률론적)
- 의미 : 어느 값이 나타날 가능성이 어느 정도다 (높은 확률도 보장 안됨!)
- 확률적 근사해 도출 : **회귀분석**

## 복잡계 이론 (결정론적 + 확률)

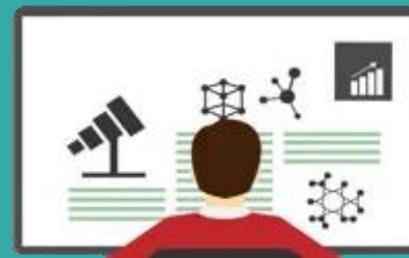
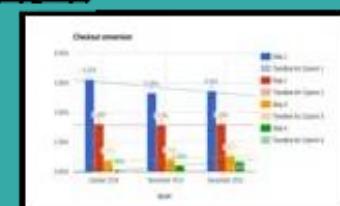
- 비선형성 : 나비효과 / 네트워크의 특성
- 의미 : 세상이 복잡해서 함수를 찾지 못하는 것 (조건 조금 변하면 결과 크게 변화)
- 근사 최적해 도출 : 기계학습 – **신경망(Neural Network) 딥러닝**



2. 재무 이론의 통계적 해석

## 데이터 안에 숨겨진 패턴 찾기 (통계 분석) 함수식을 “명시적”으로 찾는 방법

틀릴 가능성과 원인도 알 수 있음

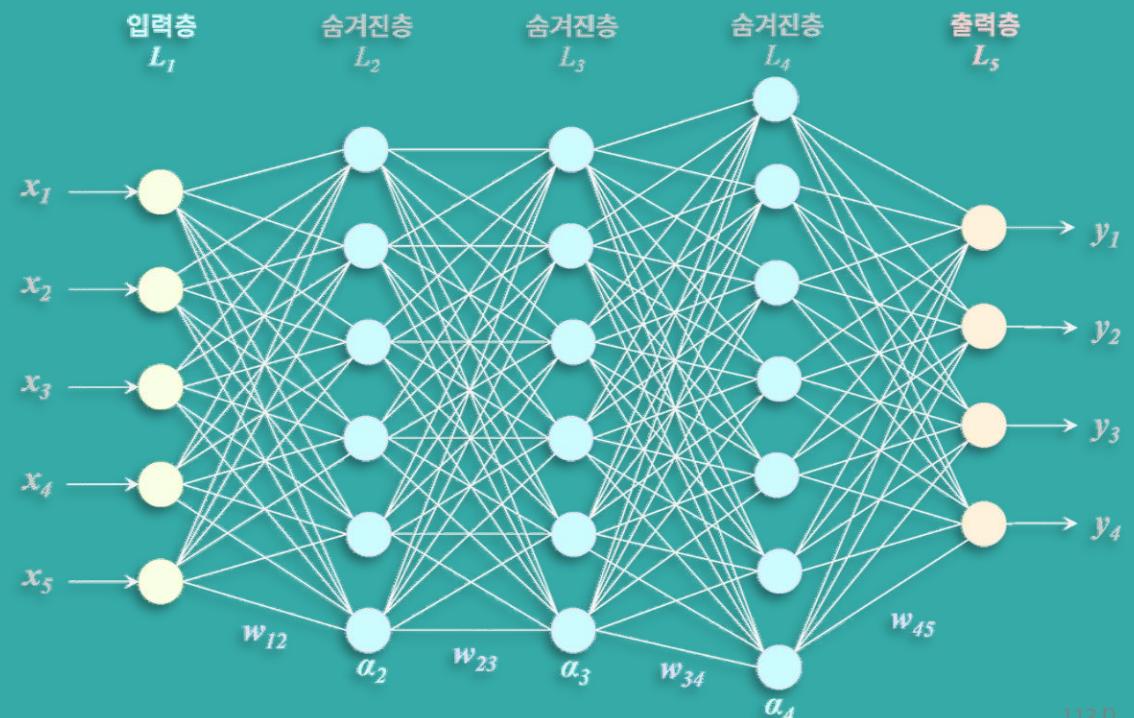
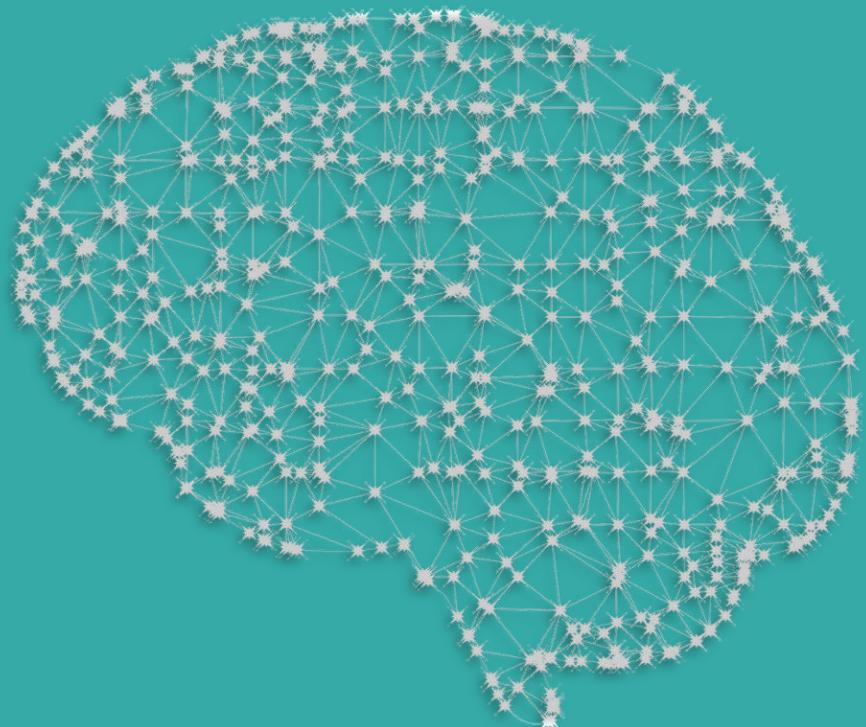


## 2. 재무 이론의 통계적 해석

# 데이터 안에 숨겨진 패턴 찾기 (기계학습;AI)

## 함수식을 “내재적”으로 찾는 방법

틀릴 가능성은 추정하나 틀린 원인을 알기 어려움



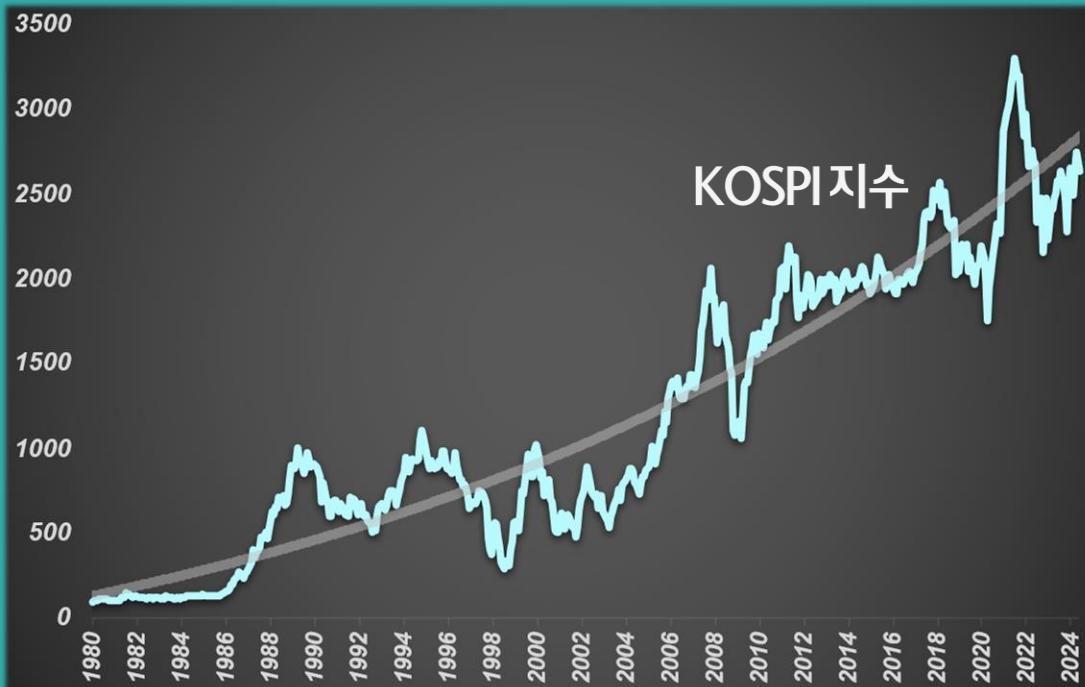
## 2. 재무 이론의 통계적 해석

# 과학적으로 관찰되는 시장의 그래프는?

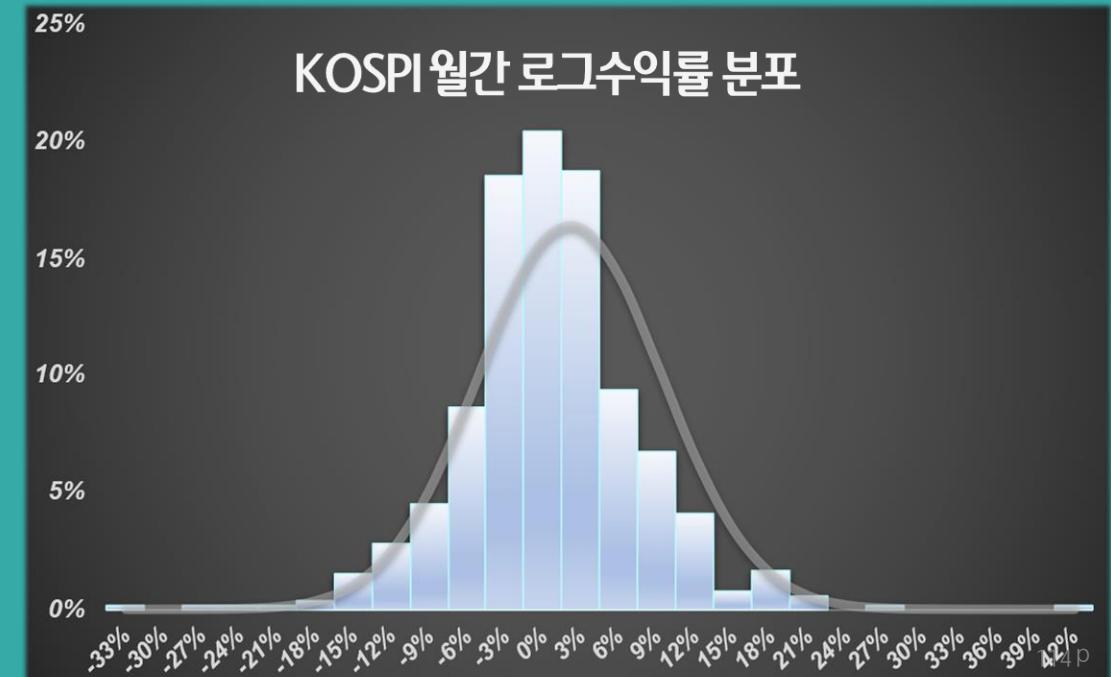
## 가격 추이

## 수익률 분포

- 스토리 텔링에 용이
- 과거만 설명할 뿐 (1개의 경로)
- 이 그림에 기반한 모델 없음



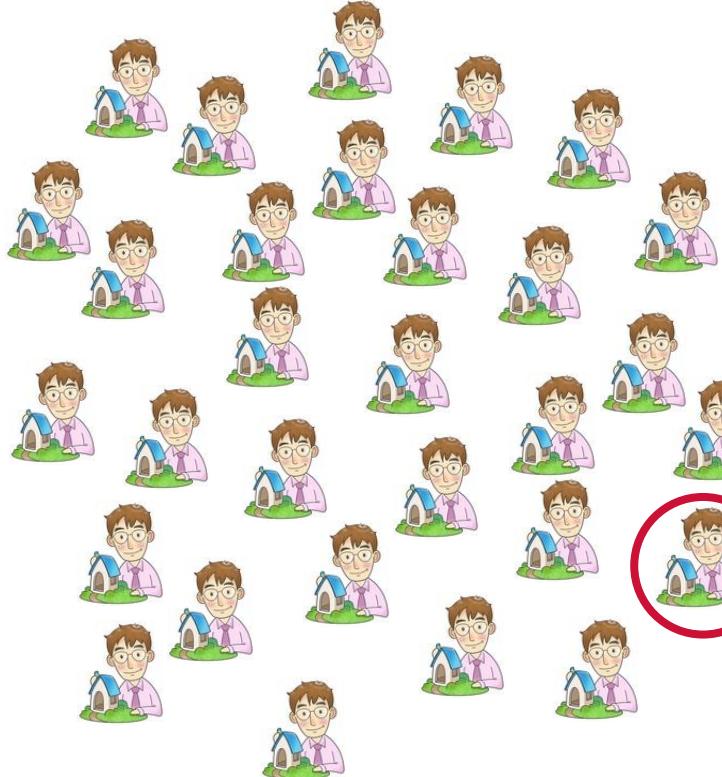
- 스토리 텔링 거의 불가능
- 미래에도 이와 유사한 분포
- Stochastics 모델의 기초



## 2. 재무 이론의 통계적 해석

# 대수의 법칙 보험업

가입자들 (100명)



보험료 수입 1천만원( $100 \times 10\text{만원}$ )

보험료 10만원

발생확률 1.0%

보험금 지급 8백만원( $1 \times 8\text{백만원}$ )



(경비+수익) 2백만원

## 2. 재무 이론의 통계적 해석

## 프로스펙트 이론

# 우리가 인식하는 확률은?



Prospect Theory: An Analysis of Decision under Risk

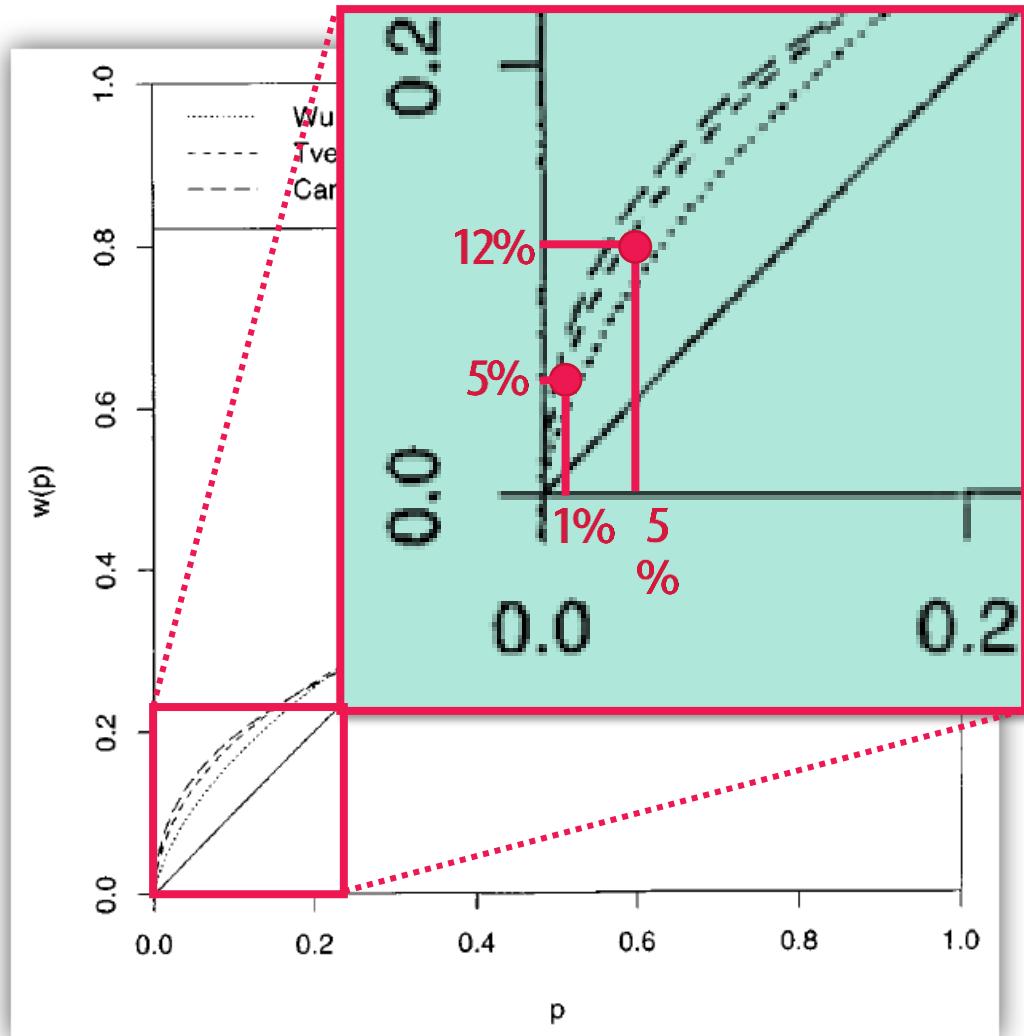
Daniel Kahneman and Amos Tversky

Econometrica  
Vol. 47, No. 2 (Mar., 1979), pp. 263-292 (29 pages)

Published By: The Econometric Society



## The Weighting Function



## 무작위 분포 중심으로 설계된 시계열 모형

$$dS_t = \mu S_t dt + \sigma S_t dw_t$$

Drift

Diffusion

연기의 중심방향 (Drift)

기(技) : 바람 (편더멘탈)

연기의 무작위 퍼짐 (Diffusion)

운(運) : 무작위 움직임

기(技) : 확산 속도 (위험관리)

# 합리적인 평균-분산 모형의 진실은?



Harry Markowitz, 1959.

기대수익률 14%

$$\min_{x_i, x_j} v = \sum_{i \in I} x_i^2 \sigma_{ii} + \sum_{i \in I} \sum_{j \in J} x_i x_j \sigma_{ij} \quad (48.1)$$

$$\text{subject to } \sum_{i \in I} r_i x_i \geq k \quad (48.2)$$

$$\sum_{i \in I} x_i = 1 \quad (48.3)$$

$$x_i \geq 0 \quad \forall i \in I, \quad (48.4)$$

where  $x_i$  = proportion of investment in security  $i$

$\sigma_{ii}$  = variance of rate of return of security  $i$

$\sigma_{ij}$  = covariance of rate of return of security  $i$  and  $j$

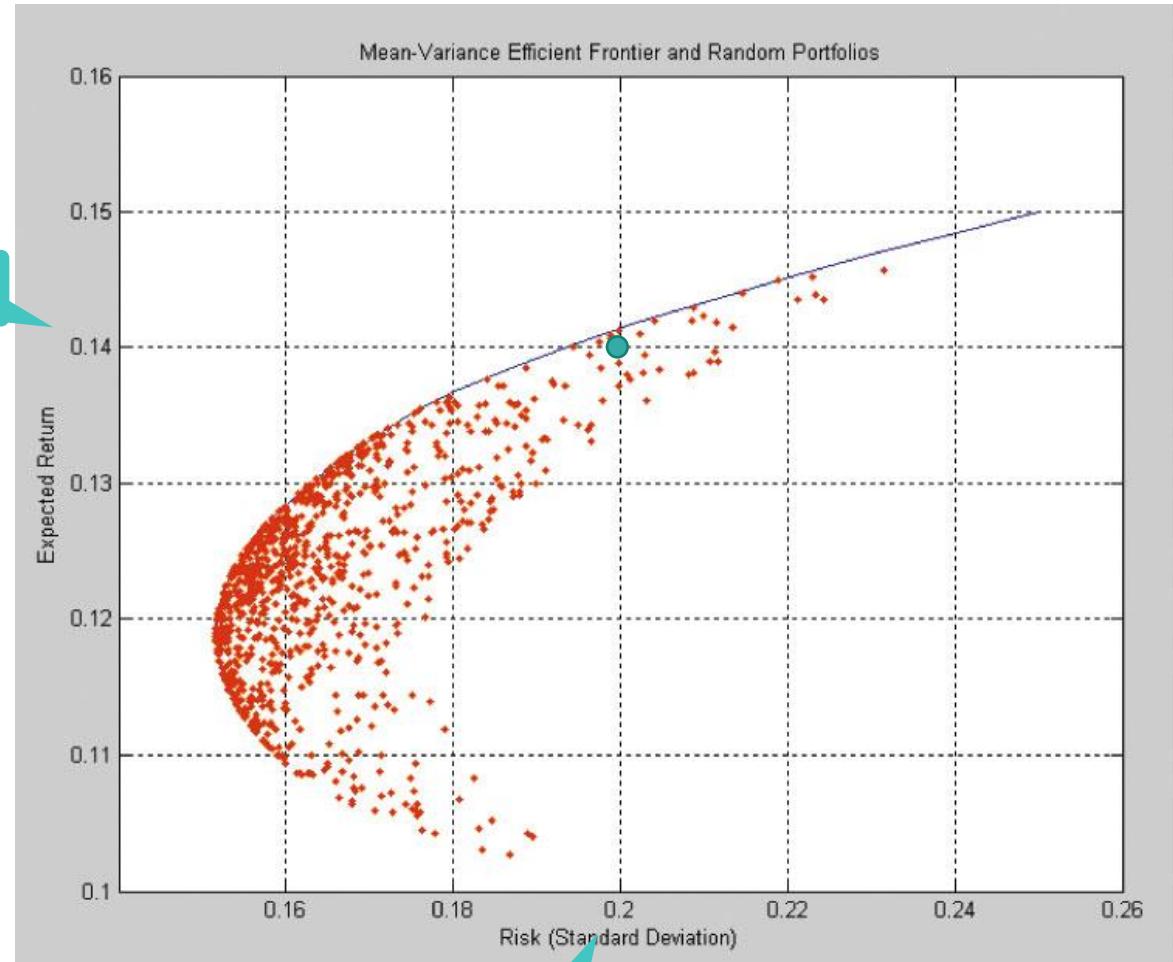
$r_i$  = expected rate of return of security  $i$

$k$  = minimum rate of return of the portfolio

$I$  and  $J$  are sets of positive integers

The resulting Lagrange function is therefore

$$L = v + \lambda \left( k - \sum r_i x_i \right) + \gamma \left( 1 - \sum x_i \right) \quad (48.5)$$



위험 20%

# 점이 아닌, 분포를 볼 수 있어야...

틀리기 쉬운 것에 집중한 것!

MPT의 본질은 위험분산 효과!

위험 분산은 자산간의 관계( $\rho$ )에 의해 나타나는데,  
이것을 이용하여 위험을 최소화

$$\min_{x_i x_j} v = \sum_{i \in I} x_i^2 \sigma_{ii} + \sum_{i \in I} \sum_{j \in J} x_i x_j \sigma_{ij} \quad (48.1)$$

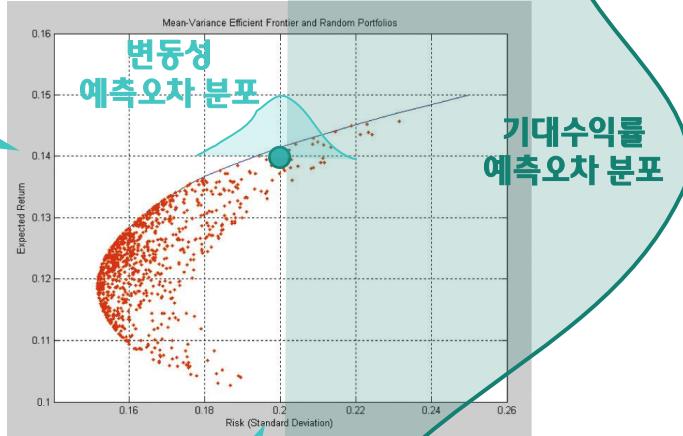
$$\text{subject to } \sum_{i \in I} r_i x_i \geq k \quad (48.2)$$

$$\sum_{i \in I} x_i = 1 \quad (48.3)$$

$$x_i \geq 0 \forall i \in I, \quad (48.4)$$

금융 이론이 확실하게 말할 수 있는 건, '위험분산 효과' 말고는 없다!!!

기대수익률 14%

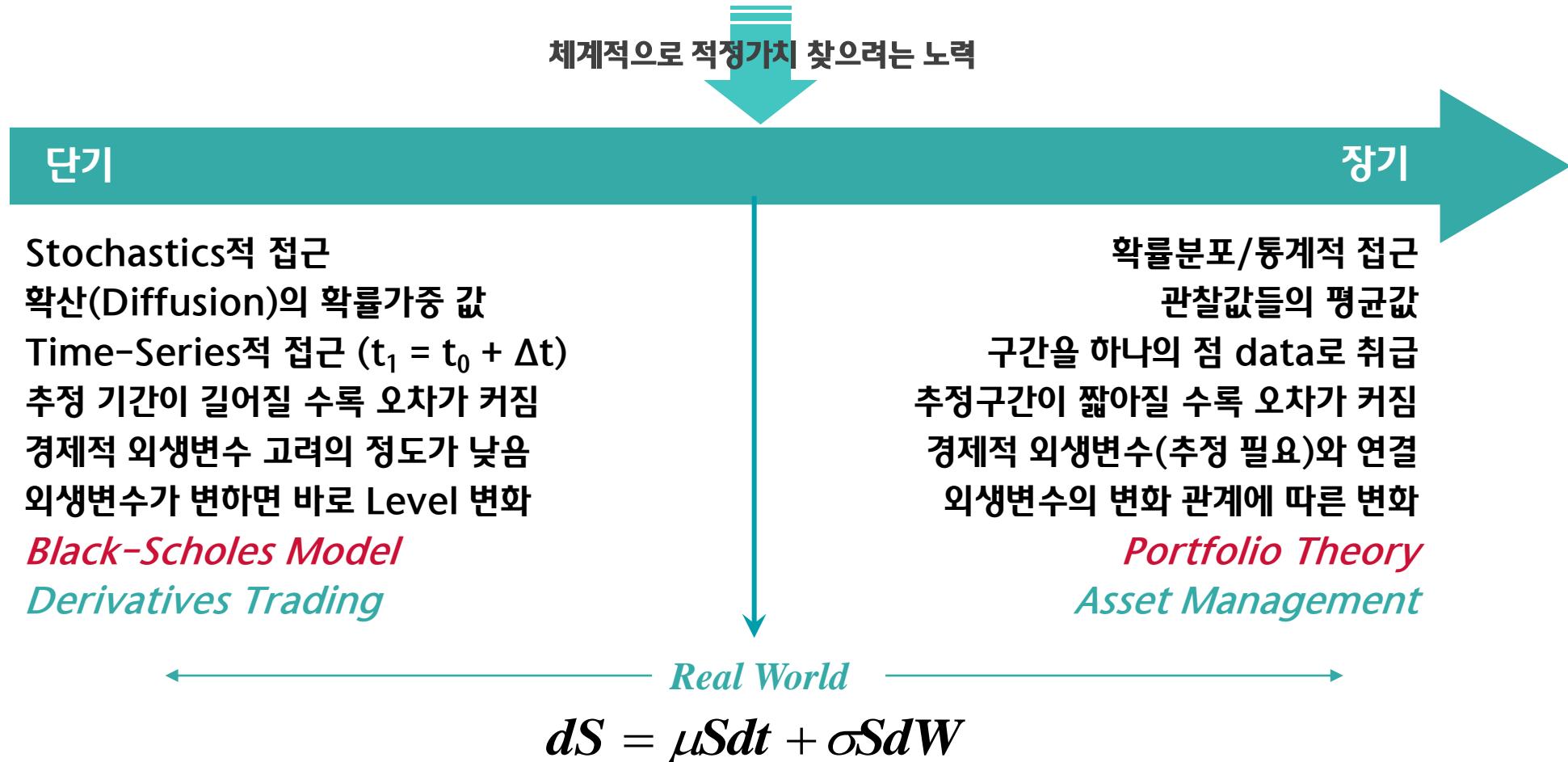


위험 20%

## 2. 재무 이론의 통계적 해석

# 복잡한 근본적 이유는 시간

## 수식에서 변수의 대부분의 것들은 미래의 값



## 작은 가능성에 대한 반복이 기적을 만든다



- 나쁜 아니라 참여자 모두가 확률 게임을 하고 있는 상황
- 영화 “머니볼”의 Oakland Athletics : 수 많은 게임이 반복되면서 평균에 수렴
- 프로 포커 게이머 : 멀티 플레이를 통해 평균에 수렴

## 핵심요약

- 수익률은 연속된 기간에 대해서는 로그수익률(기하평균) 데이터를 사용하고, 동기간의 포트폴리오에 대해서는 산술수익률(산술평균) 데이터를 사용
- 금리와 수익률은 시간, 시장 참여자 관점에서 다양한 정의와 계산이 존재
- 기대수익률은 미래 수익을 추정해 산출하는 값이고, 요구수익률(할인율)은 매수자가 원하는 투자 대상의 수익률임. 이 균형에 의해 가격이 움직이는 것.
- 재무/경제 이론은 확률론적 세계관에 입각한 ‘통계적 모델’
- 통계 모델은 ‘명시적’ 함수를 찾고, 기계학습 모델은 ‘내재적’으로 찾음
- 평균(기대값)은 대수의 법칙/장기적인 추세를, 위험(변동성)은 실현 불가능성/단기적 무작위성을 의미

# 5. 데이터 분석 기본기

---

# 학습목표

- ▶ 데이터 분석 프로세스 관점에서 분석툴 사용에 대한 개념 이해
- ▶ 데이터 저장/활용을 위한 테이블 구조의 관계형 데이터 모델의 의미를 이해
- ▶ 엑셀에서 데이터를 정의하고 참조하며, 처리하는 메커니즘 이해
- ▶ 데이터 분석에 자주 활용되는 함수들을 확인하고, 활용법 습득

## 1. 데이터 분석 프로세스와 관계형 데이터 모델

# 데이터 분석 프로세스와 필요 역량

## 데이터 획득



- 텍스트 파일 : CSV, TXT
- Open API : JSON, XML
- 내부 데이터베이스 : SQL
- 웹 (공공 데이터) : XLS, 텍스트
- 웹 페이지 : HTML(크롤링, 스크레이핑)

## 처리/분석 도구



- 데이터 입력/출력
- 데이터 가공/정제 : 셀, 시트, 함수
- 데이터 시각화 : 차트, 피벗 테이블
- 데이터 분석 : 함수, 데이터분석 도구, 해찾기 도구 등
- 보고서 출력 : XLS, PDF

## 분석 방법론



- 시각적 방법
- 통계적 방법 : 기술통계, 회귀분석, 판별분석 등
- 머신 러닝 : 클러스터링, 딥 러닝 등

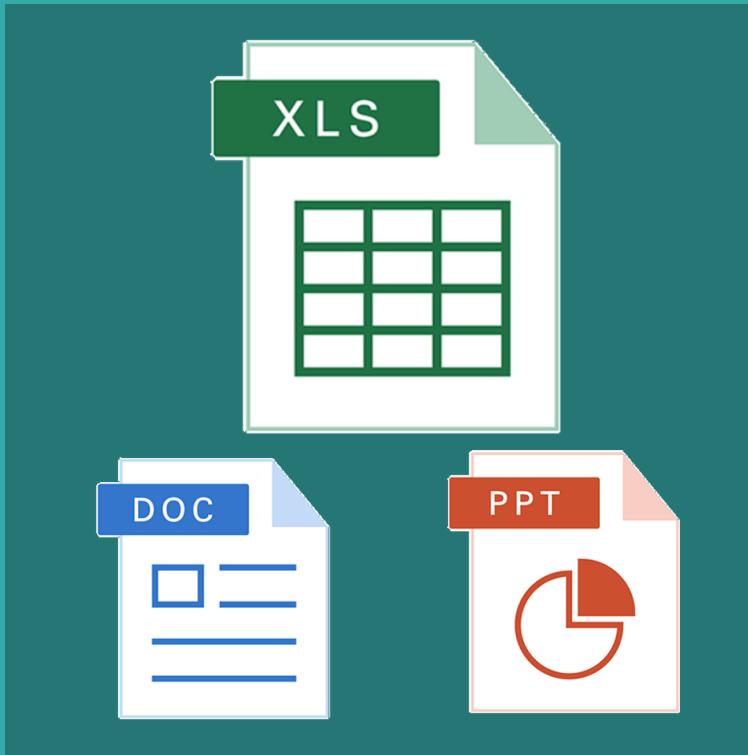
1. 데이터 분석 프로세스와 관계형 데이터 모델

## 실무에서 엑셀 데이터분석을 권하는 이유

친근하고 빠르게 접근!

머신러닝 이외 거의 다 됨!

VBA 활용시 자동화 가능!



## 1. 데이터 분석 프로세스와 관계형 데이터 모델

## (1) 파일/DB/웹으로부터 원 데이터 입력 후

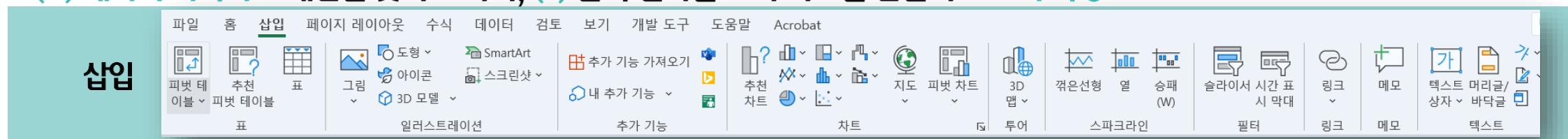
# 엑셀 메뉴 구조



(2) 수식/함수를 이용하여 데이터를 정제/가공하고, (3) 함수/데이터 분석 툴을 이용하여 가공한 데이터 통계분석 수행



(2) 데이터 시각화로 패턴을 찾기도 하여. (4) 분석 결과를 표나 차트를 만들어 보고서 작성



(5) 반복 작업, 정기 리포트가 있다면, VBA(매크로) 활용시 보고서 자동화를 통해 업무 효율 극대화



## 1. 데이터 분석 프로세스와 관계형 데이터 모델

# 확장성(라이브러리)과 자동화(프로그래밍)를 고려하면 파이썬!

## (1) 파일/DB/웹으로부터 원 데이터 입력 후

- |           |                                                                                                                                                                                                                                                                                                                                                                  |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 데이터<br>수집 | <ul style="list-style-type: none"> <li>▪ 텍스트 파일 (<a href="#">pandas</a>)</li> <li>▪ 웹데이터: OpenAPI (<a href="#">json</a>, <a href="#">xml</a>), 크롤링 (<a href="#">beautifulsoup</a>), 웹스크레이핑 (<a href="#">selenium</a>)</li> <li>▪ 데이터베이스: MySQL (<a href="#">pymysql</a>), Oracle (<a href="#">cx_Oracle</a>), MS-SQL Server (<a href="#">pymssql</a>)</li> </ul> |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## (2) 함수와 간단한 통계를 통해 데이터를 정제/가공하고

- |            |                                                                                                                                                                                                |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 데이터<br>전처리 | <ul style="list-style-type: none"> <li>▪ Raw data 전처리 (<a href="#">pandas</a>): 필터링, 검증, 문자열 처리, 조건 선별, 정렬, 그룹핑, 데이터 합치기/추가/삭제/수정</li> <li>▪ 수학적, 통계적 기초 계산 (<a href="#">numpy</a>)</li> </ul> |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## (3) 통계분석, 머신러닝, 딥러닝을 통해 데이터를 분석하여

- |           |                                                                                                                                                                                                                                                           |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 데이터<br>분석 | <ul style="list-style-type: none"> <li>▪ 통계분석 (<a href="#">numpy</a> / <a href="#">scipy</a> / <a href="#">statistics</a>)</li> <li>▪ 머신러닝 (<a href="#">statsmodels</a> / <a href="#">sklearn</a>)</li> <li>▪ 딥러닝 (<a href="#">tensorflow</a>)</li> </ul> |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## (4) 다양한 표준 형태로 데이터 저장하거나, 시각화/엑셀 문서를 통해 보고서 작성

- |             |                                                                                                                                                                                                                                                                                                                                                                                                          |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 저장 및<br>보고서 | <ul style="list-style-type: none"> <li>▪ 데이터 저장 (<a href="#">pandas</a>, <a href="#">json</a>, <a href="#">xml</a>, <a href="#">pymysql</a>, <a href="#">cx_Oracle</a>, <a href="#">pymssql</a>)</li> <li>▪ 차트 시각화 (<a href="#">pandas</a>, <a href="#">matplotlib</a>, <a href="#">seaborn</a>), 엑셀 파일/보고서 (<a href="#">openpyxl</a>, <a href="#">python-docx</a>, <a href="#">win32com</a>)</li> </ul> |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

1. 데이터 분석 프로세스와 관계형 데이터 모델

# 시각화 툴 도입 이유는 공유!

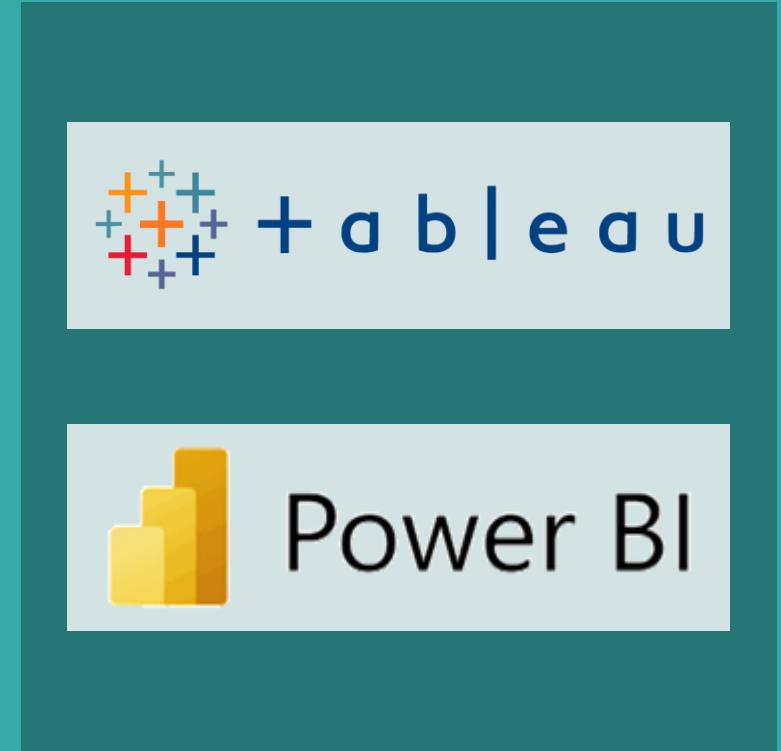
시각화에 초점!



조직내 공유 플랫폼



대표 시각화 패키지

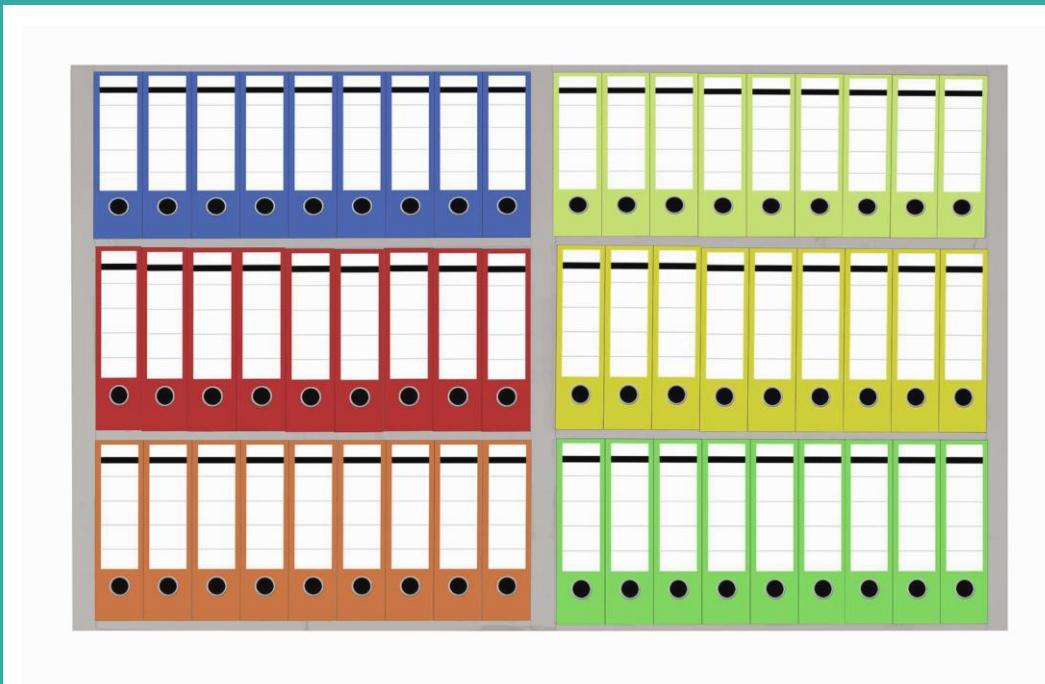


1. 데이터 분석 프로세스와 관계형 데이터 모델

## 데이터를 저장하는 두 가지 방법

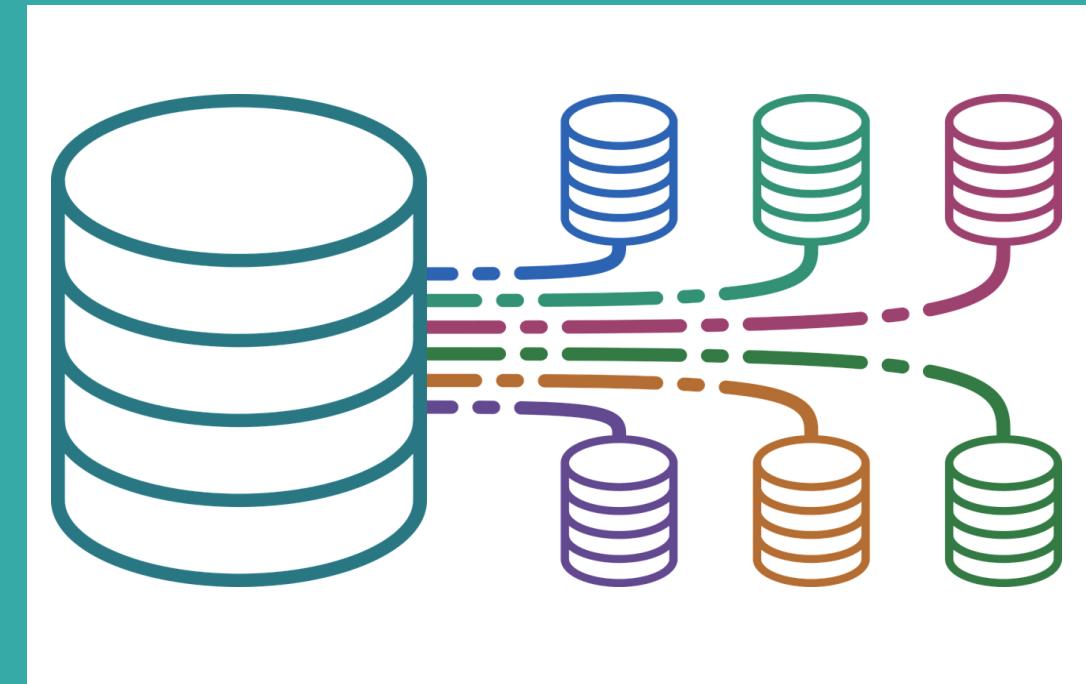
### 파일 처리

자료 저장의 기본 방법 (txt, xls, doc, pdf)



### 데이터베이스

구조화된 자료 저장과 핸들링 (DBMS)



## 5. 데이터 분석 기본기

### 1. 데이터 분석 프로세스와 관계형 데이터 모델

# 왜 데이터는 행과 열의 테이블 형태로 표현되고 있을까?

The image displays three distinct software interfaces side-by-side, illustrating different ways data is presented and managed:

- Microsoft Excel:** On the left, a screenshot of the Microsoft Excel application shows a simple table with columns labeled "이름" (Name), "학년" (Grade), and "이메일" (Email). The data includes rows for Hong Gil Dong, Jung Yeong Young, Shin Sa Im Dang, You Wan Sun, and You Wan Sun.
- Database Management System (DBMS) Interface:** In the center, a screenshot of a DBMS interface shows a table with columns "id" and "date". The data consists of 99 rows, each containing a unique identifier (e.g., 968, 969, 970, ..., 989) and a timestamp (e.g., 2023-01-24 00:00:00). The interface includes a toolbar at the top and a detailed view of the table structure on the right.
- Web Shopping Platform:** On the right, a screenshot of a web-based shopping platform (Naver Shopping) for coffee. It features a search bar, category filters for "커피/차류" (Coffee/Coffee), and a product listing for "Nestle Nescafe Gold Blend Instant Coffee". The page includes various filters for price, brand, and delivery type, along with a summary table of statistics like total products and average price.

## 1. 데이터 분석 프로세스와 관계형 데이터 모델

# 왜 데이터는 행과 열의 테이블 형태로 표현되고 있을까?

	A	B	C	D	E	F
1	이름	학년	이메일	수강과목명	학점	
2	홍길동	2	hgd@snowballproject.co.kr	수학	4	
3	정약용	1	jyy@snowballproject.co.kr	영어	3	
4	신사임당	3	saimdang@snowballproject.co.kr	영어	3	
5	유관순	2	yks@snowballproject.co.kr	수학	4	
6	유관순	2	yks@snowballproject.co.kr	영어	3	

<학생> (<학년>, <이메일>)이 <과목> (<학점>)을 <수강>한다.

관계(relation) : 두 개 이상의  
<속성>을 구성하는 원소들의 순서쌍

- 홍길동은 (2학년, 이메일 주소는 hgd@snowballproject.co.kr) 수학(4학점)을 수강한다.
- 정약용은 (1학년, 이메일 주소는 jyy@snowballproject.co.kr) 영어(3학점)를 수강한다.
- 신사임당은 (3학년, 이메일 주소는 saimdang@snowballproject.co.kr) 영어(3학점)를 수강한다.
- 유관순은 (2학년, 이메일 주소는 yks@snowballproject.co.kr) 수학(4학점)을 수강한다.
- 유관순은 (2학년, 이메일 주소는 yks@snowballproject.co.kr) 영어(3학점)를 수강한다.

관계들의  
나열(집합)

## 1. 데이터 분석 프로세스와 관계형 데이터 모델

# 관계형 데이터 모델

테이블 형식을 이용하여 체계적으로 데이터를 정의 (데이터 정제)

학생 수강 현황

필드

관계(relation) : 두개 이상의 '속성'을  
구성하는 원소들의 순서쌍



이름	학년	이메일	수강과목명	학점
홍길동	2	hgd@snowballproject.co.kr	수학	4
정약용	1	jyy@snowballproject.co.kr	영어	3
신사임당	3	saimdang@snowballproject.co.kr	영어	3
유관순	2	yks@snowballproject.co.kr	수학	4
유관순	2	yks@snowballproject.co.kr	영어	3

## 1. 데이터 분석 프로세스와 관계형 데이터 모델

# E-R (Entity–Relationship) 다이어그램

## 관계형 데이터베이스의 체계화를 위한 데이터 구조 추상화

필드

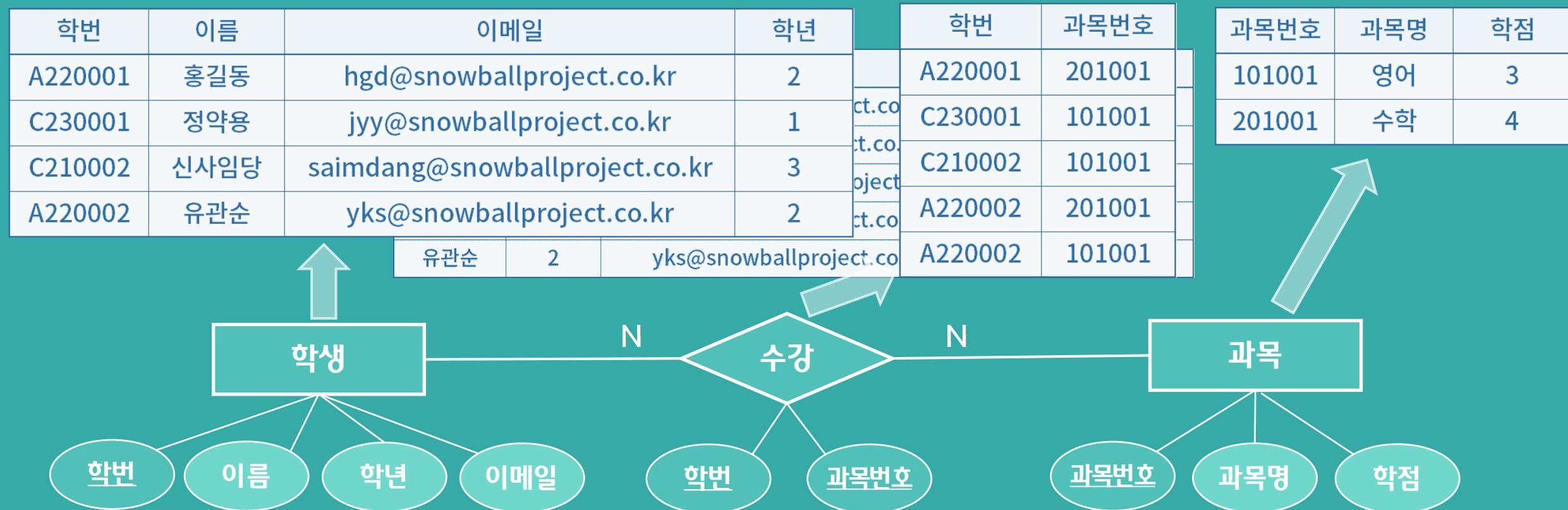
이름	학년	이메일	수강과목명	학점
홍길동	2	hgd@snowballproject.co.kr	수학	4
정약용	1	jyy@snowballproject.co.kr	영어	3
신사임당	3	saimdang@snowballproject.co.kr	영어	3
유관순	2	yks@snowballproject.co.kr	수학	4
유관순	2	yks@snowballproject.co.kr	영어	3



## 1. 데이터 분석 프로세스와 관계형 데이터 모델

# E-R (Entity–Relationship) 다이어그램

## 관계형 데이터베이스의 체계화를 위한 데이터 구조 추상화



## 1. 데이터 분석 프로세스와 관계형 데이터 모델

# Structured Query Language

## 관계형 데이터베이스에서 데이터를 조회, 삽입, 수정하는 언어

학생 (b)

학번	이름	이메일	학년
A220001	홍길동	hgd@snowballproject.co.kr	2
C230001	정약용	jyy@snowballproject.co.kr	1
C210002	신사임당	saimdang@snowballproject.co.kr	3
A220002	유관순	yks@snowballproject.co.kr	2

과목 (c)

과목번호	과목명	학점
101001	영어	3
201001	수학	4

수강 (a)

학번	과목번호
A220001	201001
C230001	101001
C210002	101001
A220002	201001
A220002	101001

SELECT [조회할 필드들]

FROM [조회 필드가 존재하는 테이블들]

WHERE [조회 조건들]

SELECT b.이름, b.학년, b.이메일, c.과목명, c.학점

FROM 수강 a, 학생 b, 과목 c

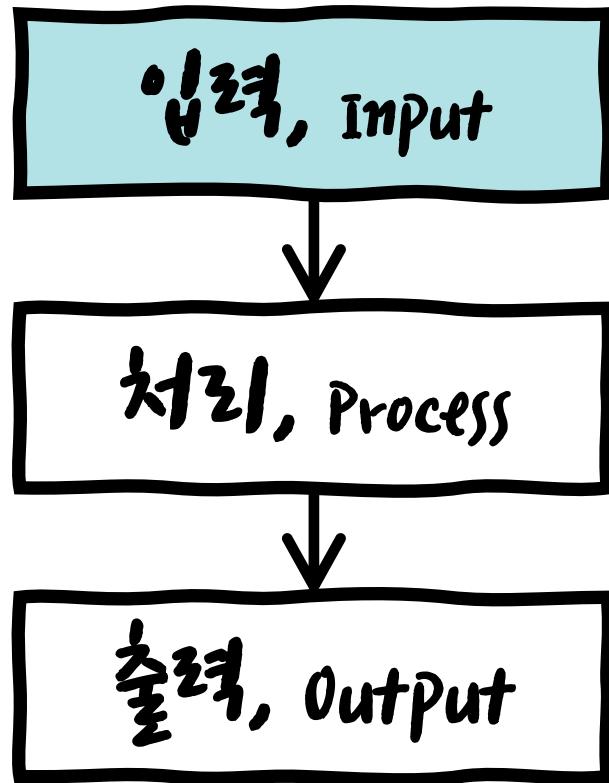
WHERE a.학번 = b.학번

AND a.과목번호 = c.과목번호

Key를 이용한 연결

이름	학년	이메일	수강과목명	학점
홍길동	2	hgd@snowballproject.co.kr	수학	4
정약용	1	jyy@snowballproject.co.kr	영어	3
신사임당	3	saimdang@snowballproject.co.kr	영어	3
유관순	2	yks@snowballproject.co.kr	수학	4
유관순	2	yks@snowballproject.co.kr	영어	3

## 컴퓨터에 입력하는 일반적인 메커니즘



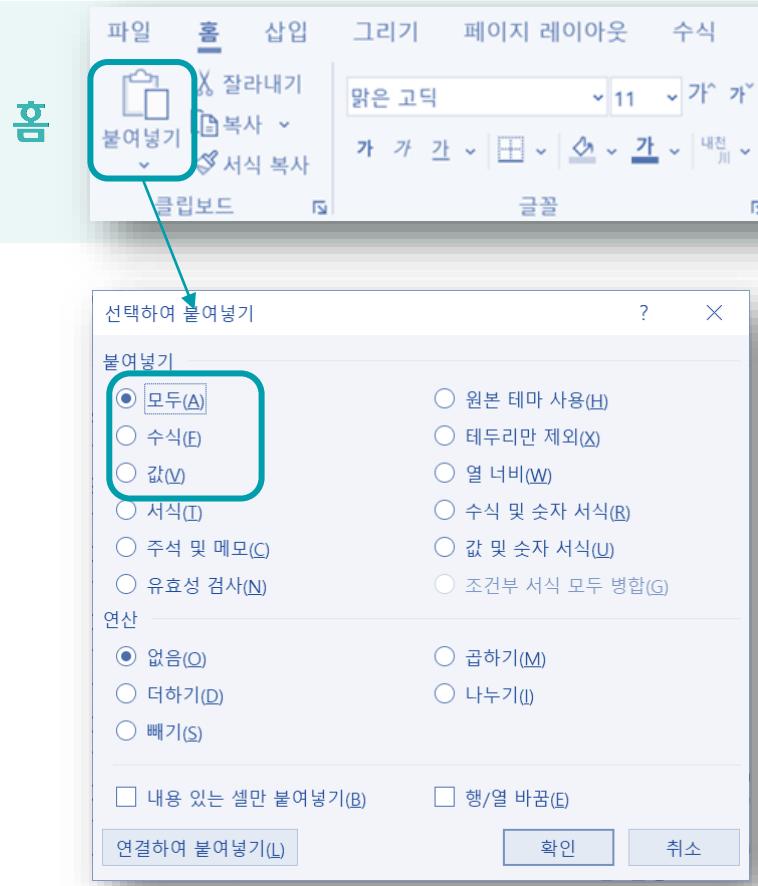
- ① 값의 종류 (형; Type) : 숫자, 문자, 날짜
- ② 값의 위치 배정 : 상수, 변수 / 배열
- ③ 값의 위치 참조 : 대입
- ① 기초 연산 : 사칙연산, 비교연산, 반복연산
- ② 함수 : 기초연산들의 집합으로서 목적이 있는 단위 연산

= 엑셀을 배우는 원리  
= 프로그래밍을 배우는 원리

- 값과 명령에 의해 도출된 값
- 명령에서 지정한 양식의 값, 그림, 소리, 영상 등

# 메뉴에 없지만 반드시 알아야 할 것

## 메뉴에는 꾸미는 것(서식)만 존재



The screenshot shows the Microsoft Excel ribbon with the Home tab selected. In the clipboard group, the 'Format Painter' icon is highlighted with a red box and a callout arrow pointing to the 'Paste Special' dialog box. The dialog box has two main sections: 'Paste' and 'Format'. Under 'Paste', the 'All (A)' radio button is selected. Under 'Format', the 'Text (T)' radio button is selected. Other options like 'Values (V)', 'Formulas (F)', 'Comments (C)', and 'Hyperlinks (H)' are also available.

**값 / 수식 / 서식의 명확한 구분!**

- 셀 안에는 세 가지 형태의 정보
- 값 : 형(type)이 지정된 데이터 또는 수식의 결과
- 수식 : 함수식 (참조, 수식, 함수식)
- 서식 : 값 또는 셀을 보여주는 양식 (글꼴, 맞춤, 표시형식, 스타일)

## 모든 데이터는 형(type)이 지정되어 있다

**1 ≠ “1”**

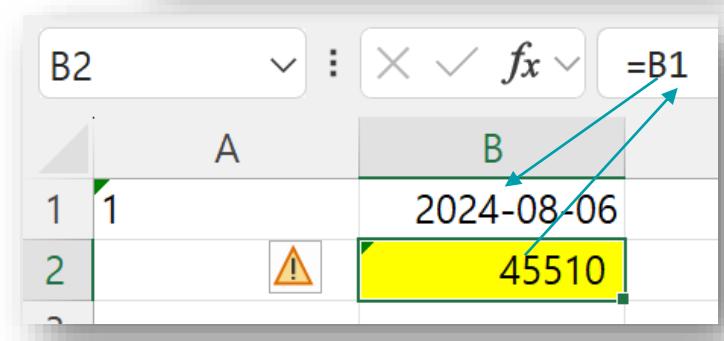
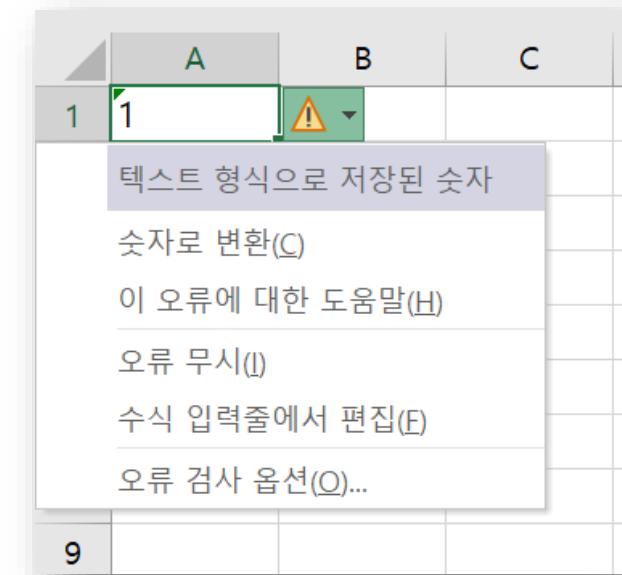
- 1은 숫자일 수도, 텍스트(문자열)일 수도 있다
- 숫자 1과 텍스트 1은 다르다
- ‘1’을 입력하거나, =“1”을 입력하면 텍스트 1로 자동 인식

**2024-08-06 = 45510**

- 날짜, 시간은 실제로 실수값
- 1900-01-01을 1로 하여, 1일(24시간)마다 1씩 증가
- 기간은 두 날짜를 빼면 바로 구해짐

### 형 (Type)

- 컴퓨터에서 모든 데이터는 형(Type)이 지정되어야 한다
- 컴퓨터는 궁극적으로 모든 Type을 숫자(이진수)로 변환 처리
- 문자열과 날짜의 처리 수준이 데이터 처리 능력의 바로미터!



# 서식은 데이터의 표현법

## 지름길 → 마우스 우측 버튼

- [셀 서식] 메뉴 이용
- 그 위치에서 할 수 있는 대부분의 것들이 메뉴에 나열됨
- 모르면 우측 버튼을 눌러라!

## 셀 서식 특징

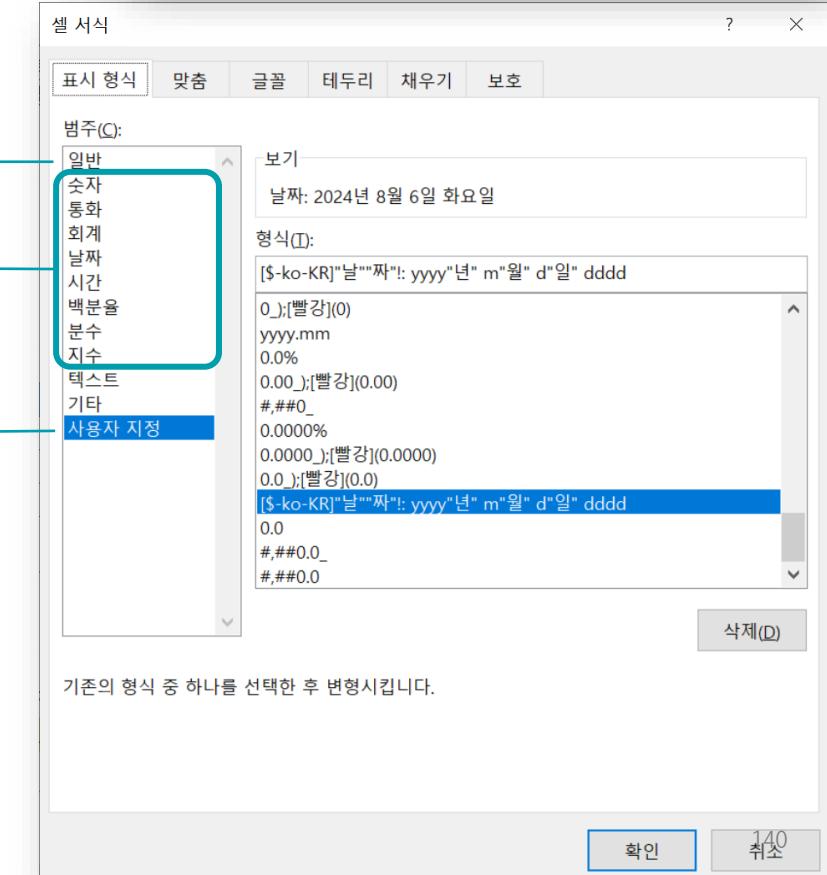
- [일반] 경우에 따라 표시형식이 역으로 Type을 결정 : 텍스트, 날짜, 백분율 등
- 서식 변경해도 Data는 변동없음
- 보여주는 형식만 바뀐 것
- 사용자가 직접 표시 형식을 정의할 수 있음

엑셀이 알아서 결정

모두 숫자

무조건 문자열화

C1	A	B	C
1	1	2024-08-06	날짜: 2024년 8월 6일 화요일
2		45510	



## 셀과 참조 - 데이터의 위치

### 좌표(행렬)를 통한 셀의 인식; B7 = R7C2

- A1식 표현: 행은 숫자, 열은 영문자; 열 → 행 순
- R1C1식 표현: 행, 열 모두 숫자; 행 → 열 순
- 수작업시는 A1식 표현이 간단, 자동화 작업(매크로, 인덱스 참조) 시에는 R1C1식 표현이 용이

### 절대-상대 위치

- 상대 위치 (Default)
  - ✓ A1식: 아무 표시 없이
  - ✓ R1C1식: 표현식 셀 기준으로 대괄호('[]')로 상대 위치 표현
- 절대 위치 (F4 토글 가능)
  - ✓ A1식: '\$'를 앞에 표시
  - ✓ R1C1식: 절대값 그대로 표현
- 반복되는 수식 복사할 때 매우 중요!

**A1식**

절대 =SUM(\$A\$4:\$B\$6)

	A	B	C
4	1	2	
5	3	4	
6	5	6	
7	SUM	21	
8			

**R1C1식**

절대 =SUM(R4C1:R6C2)

	1	2	3
4	1	2	
5	3	4	
6	5	6	
7	SUM	21	
8			

## 셀과 참조 – 데이터의 연속 사용

### 범위 (Range)

- 형식 : [좌측상단]:[우측하단]
- A1식 : A4:B6, \$A\$4:\$B\$6, \$A4:\$B6, A\$4:B\$6 등
- R1C1식 : R[-3]C[-1]:R[-1]C, R4C1:R6C2 등

### 데이터 연속 사용 예제

- 분석 예
  - 분석 : 키를 미터(m) 단위로 제곱한 수로 몸무게(kg)를 나누면 신체 질량지수
  - 수식 : =체중(kg) / [키(m)]<sup>2</sup>
- 계산 중간 과정도 의미가 있을 수 있어, 과정 데이터를 두는 것도 좋은 방법

F12

=C12/POWER(B12/100, 2)

	A	B	C	D	E	F
10	신체질량지수 (BMI; 카우프지수)					
11	이름	키(cm)	몸무게(kg)	키(m)제곱	비만도1	비만도2
12	홍길동	183	70	3.3	20.9	20.9
13	이도	175	85	3.1	27.8	27.8
14	정약용	170	71	2.9	24.6	24.6
15						
16	저체중(20 미만), 정상 (20~24), 과체중 (25~29), 비만(30이상)					
				=POWER(B12/100, 2)	=C12/D12	

## 셀과 참조 - 범위에 이름 부여하기

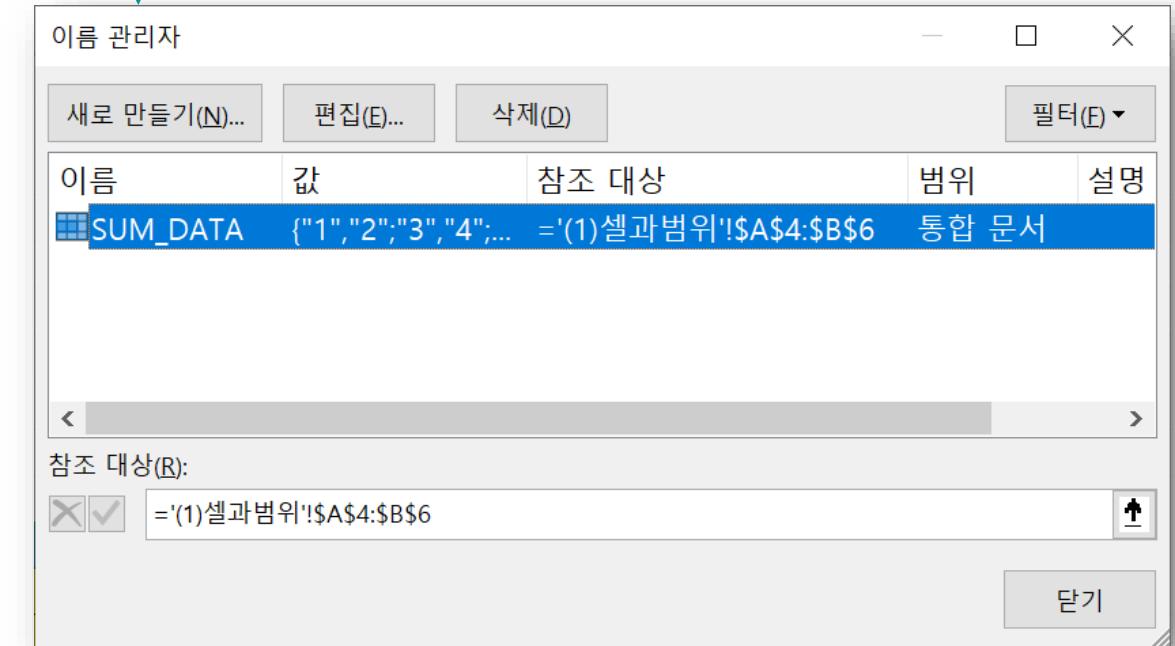
수식



(1) 최초 부여 : 셀/범위를 선택 후 이름상자에 이름 입력

SUM_DATA		fx	1
A	B		
4	1	2	
5	3	4	
6	5	6	
7	SUM	21	=(1)

(2) 이름 관리 : [수식]-[이름 관리자]에서 수정, 삭제



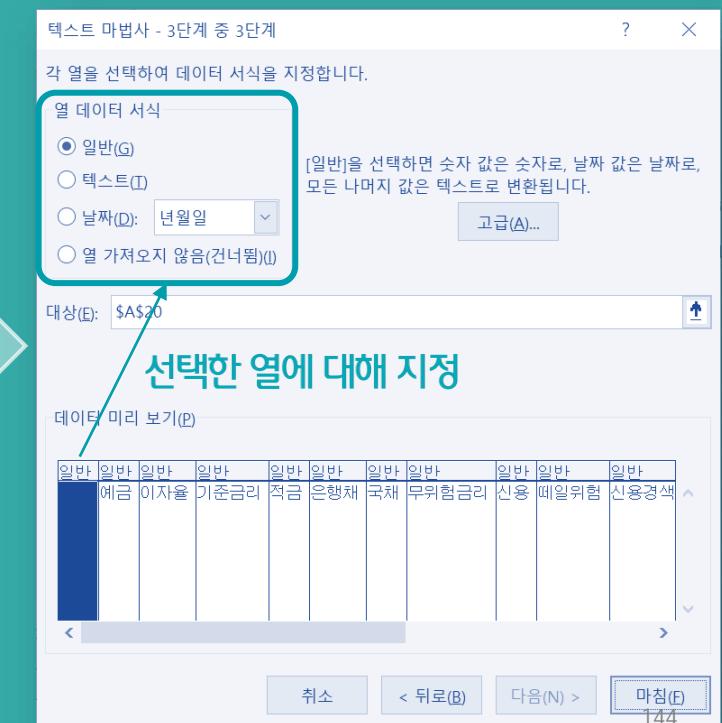
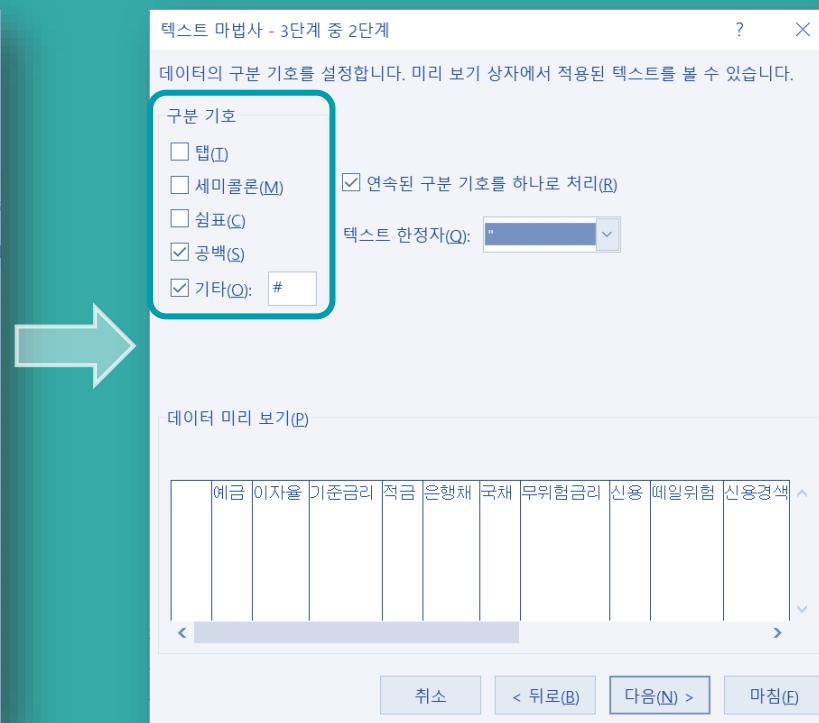
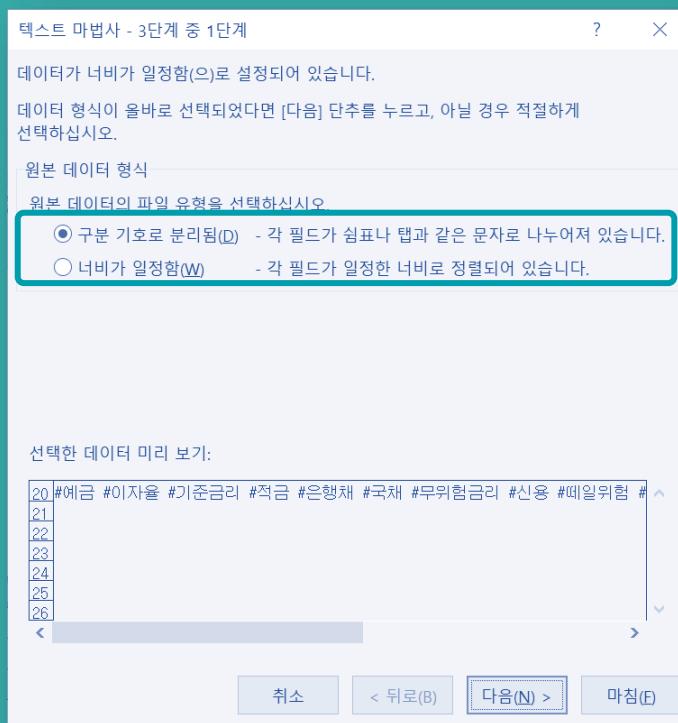
(3) 수식 관리 : 수식 변경 없이 이름 범위를 수정만 하면 끝!

## 텍스트 데이터 나누기

### 텍스트(파일)의 중요성

- 서로 다른 플랫폼 사이 데이터 교환의 일반 형태
- 문자열 내 데이터들의 구분 : 탭(Tab), 쉼표(콤마)를 사용하거나, 문자열의 인덱스(index)를 통해 각 필드를 구분
- 문자열 내의 문자열 숫자형(Type)을 구분 불가 → 지정

**해시태크(#)가 있는 이유!**  
→ 키워드 분리자



## 2. 엑셀 데이터 처리 기본기

# 선택적 정보 복사

## 수식(F)

- 참조, 수식, 함수 등 '='로 표현된 모든 것
- 상대참조인지, 절대참조인지 유의

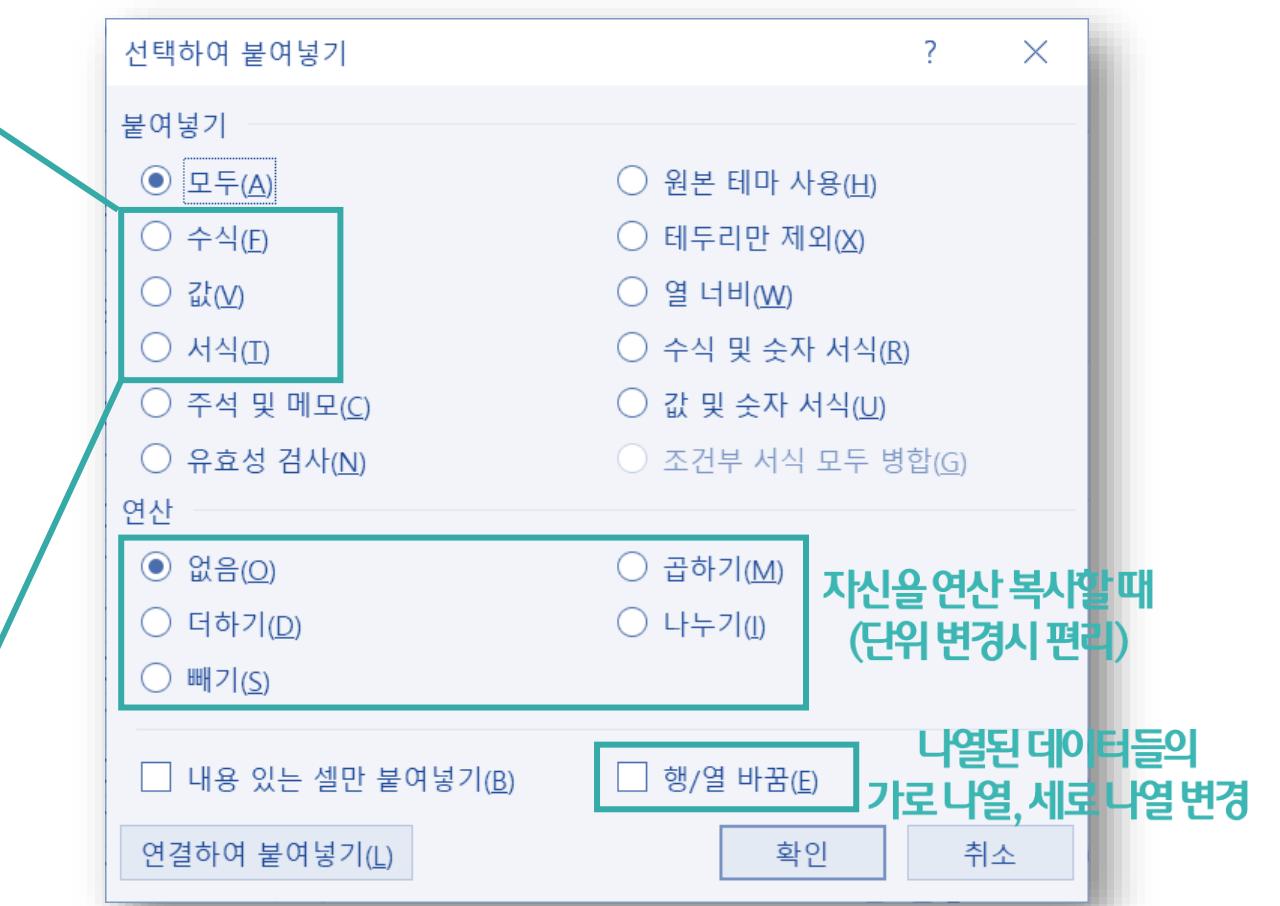
## 값(V)

- 입력하였거나 참조, 계산된 결과(result)
- 결과(result)가 빈 경우도 있음

## 서식(T)

- 글꼴, 표시형식, 테두리, 무늬 등

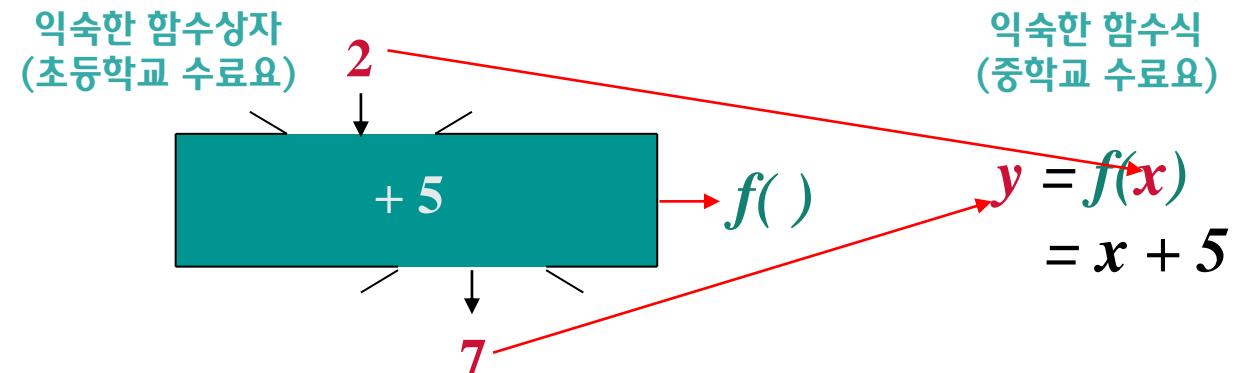
**지금까지 학습한 것들 모두 포함**



## 3. 데이터 분석을 위한 주요 함수

# 함수란 무엇인가?

함수(function)란 **‘정해진 규칙’**이다



함수는 정해진 규칙에 의해 입력값( $x$ )들을 받아 그 결과( $y$ )를 돌려주는 것이다

입력값( $x$ )

- 값이 없는 경우
- 단일값
- 배열, 범위 값
- 조건

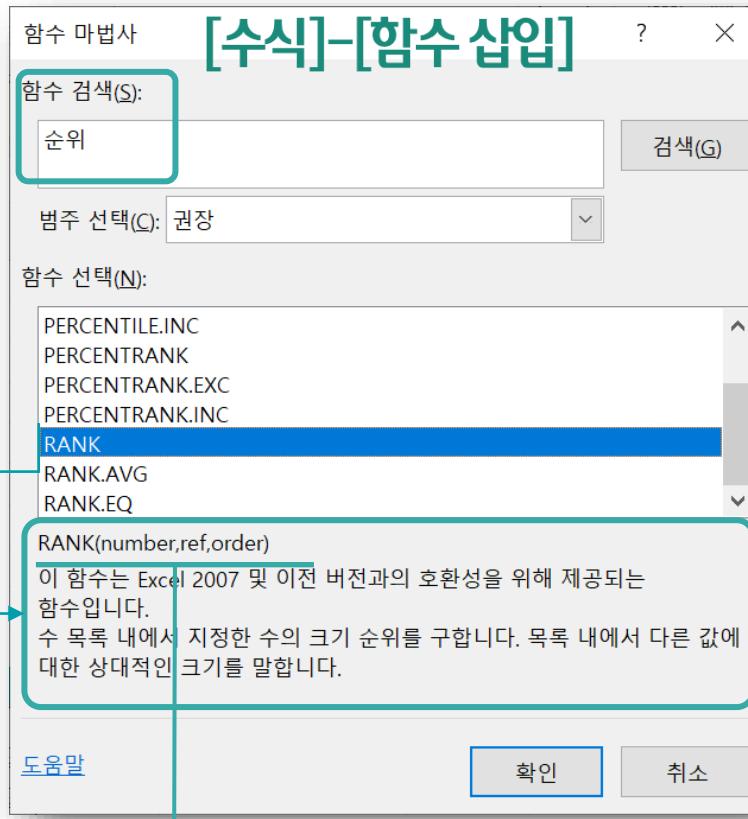
함수

출력값( $y$ )

- 단일값
- 배열값
- 오류값

## 3. 데이터 분석을 위한 주요 함수

# 함수마법사 활용법



**RANK (number, ref, order)**

값      참조범위      조건

A	B	C	D	E	F	G	
10	신체질량지수 (BMI; 카우프지수)						
11	이름	키(cm)	몸무게(kg)	키(m)제곱	비만도1	비만도2	순위
12	홍길동		183	70	3.3	20.9	20.9
13	이도		175	85	3.1	27.8	1
14	정약용		170	71	2.9	24.6	2
15							
16	저체중(20						

함수 인수

RANK f()

Number F12  
Ref \$F\$12:\$F\$14  
Order 0

이 함수는 Excel 2007 및 이전 버전과의 호환성을 위해 제공되는 함수입니다.  
수 목록 내에서 지정한 수의 크기 순위를 구합니다. 목록 내에서 다른 값에 대한 상대적인 크기를 말합니다.

Order 은(는) 순위를 정할 방법을 지정하는 수입니다. 0이나 생략하면 내림차순으로, 0이 아닌 값을 지정하면 오름차순으로 순위가 정해집니다.

수식 결과= 3 y 값

도움말(H)      확인      취소

→도움말 보기 습관을 기르자!

## 3. 데이터 분석을 위한 주요 함수

## 입력 인수 표현

### 단일값

- `number` 또는 `value : type`이 숫자인 값
- `text : type`이 문자열(텍스트)인 값
- `serial_number` : `type`이 날짜(숫자)인 값
- 기타 : 해당 인수의 성격을 나타내는 표현 사용 (`power`, `divisor`, `criteria`, `order`, `year`, `start_num`, `value_if_true`, `value_if_na`, `lookup_value`...)

### 배열

여러 개의 나열  
또는 범위

- `number1, number2,...` 또는 `value1, value2,...` : 숫자들의 나열 (위치가 연속될 필요 없음)
- `text1, text2,...` : 텍스트들의 나열
- `range` 또는 `ref` 또는 `reference` : 데이터의 범위 (행열이  $m \times n$  상관 없으며, 연속될 필요 없으며, 범위를 여러 개 나열도 가능)
- `array` : 데이터의 범위 (일반적으로, 행 또는 열 값들의 위치가 이어지는 값들)
- `table_array` : 데이터의 범위 (단, 행열이  $m \times n$ 의 테이블 형태)

### 기타

- `format...` 또는 `criteria` : 엑셀에서 약속된 표현법 (규칙 확인하여 사용)
- `logical_test` : 단일 논리식 (비교식)
- `logical1, logical2,...` : 논리식의 나열

## 3. 데이터 분석을 위한 주요 함수

## 기본 수학/삼각 함수

기본적으로 관계가 멀지만,  
몇개 기본 수학함수는 필수적

함수명	입력 인수	함수 설명	예제
<b>ABS</b>	(number)	절대값	ABS(-3) = 3
<b>LN</b>	(number)	자연로그값	LN(2) = 0.6931
<b>POWER</b>	(number,power)	number를 power로 거듭제곱	POWER(2, 3) = 8
<b>MOD</b>	(number,divisor)	number/divisor의 나머지 값	MOD(12, 5) = 2
<b>PRODUCT</b>	(number1,number2,...)	인수들의 곱 (number가 거듭되면, 범위표현 가능)	PRODUCT(2, 3, 4) = 24 PRODUCT(A2:A6)
<b>RAND</b>	()	0에서 1 범위 일양분포의 난수 생성	RAND() = 0.2159
<b>ROUND</b>	(number,num_digits)	number를 지정한 자리수만큼 반올림	ROUND(1.235, 2) = 1.24 ROUND(1234,-2) = 1200
<b>ROUNDDOWN</b>	(number,num_digits)	0에 가까워지도록 수를 내림	ROUNDDOWN(2.7, 0) = 2
<b>ROUNDUP</b>	(number,num_digits)	0에서 멀어지도록 수를 올림	ROUNDUP(2.3, 0) = 3
<b>SQRT</b>	(number)	number의 양의 제곱근	SQRT(9) = 3
<b>SUM</b>	(number1,number2,...)	인수들의 합	SUM(2, 3, 4) = 9
<b>SUMIF</b>	(range,criteria,sum_range)	range에서 criteria를 만족하는 대응되는 동일행 sum_range의 합	SUMIF(A:A, "남", D:D) A열에서 "남"값을 가진 행들의 D열 값을 합계
<b>SUMPRODUCT</b>	(array1,array2,array3,...)	배열에 대응되는 값끼리 곱한 값들의 합	SUMPRODUCT((1, 2), (3, 4)) = 1*3 + 2*4 = 11

## 3. 데이터 분석을 위한 주요 함수

## 기본 통계 함수

이용할 함수가 많음  
통계는 다수의 인수를 다루므로 배열형태가 기본

함수명	입력 인수	함수 설명	예제
<b>AVERAGE</b>	(number1,number2,..)	숫자들의 평균	AVERAGE(1,1,3,5,10) = 4
<b>CORREL</b>	(array1,array2)	두 데이터 집단사이의 상관계수	CORREL(B12:B14, C12:C14) = -0.19
<b>COUNT</b>	(number1,number2,...)	숫자가 들어있는 셀의 개수	COUNT(A1:A4) = 4
<b>COUNTA</b>	(value1,value2,...)	비어 있지 않은 셀의 개수	COUNTA(A1:A4) = 3
<b>COUNTIF</b>	(range,criteria)	지정한 range에서 criteria에 맞는 셀의 개수	COUNTIF(F12:F14, ">25") = 1
<b>COVAR</b>	(array1,array2)	두 데이터 집단사이의 공분산	COVAR(B12:B14, C12:C14) = -7
<b>MAX</b>	(number1,number2,..)	숫자들 중 제일 큰 값	MAX(1,2,3) = 3
<b>MEDIAN</b>	(number1,number2,...)	숫자들의 중앙값	MEDIAN(1,1,3,5,10) = 3
<b>MIN</b>	(number1,number2,..)	숫자들 중 제일 작은 값	MIN(1,2,3) = 1
<b>MODE</b>	(number1,number2,...)	숫자들의 최빈값	MODE(1,1,3,5,10) = 1
<b>NORM.S.DIST</b>	(z,cumulative)	표준정규분포 z값의 누적분포값(cumulative =True) 또는 확률밀도함수값(False)	NORM.S.DIST(1.96, TRUE) = 0.975
<b>NORM.S.INV</b>	(probability)	표준정규분포 누적분포값의 z값	NORM.S.INV(0.975) = 1.96
<b>RANK</b>	(number,ref,order)	number의 ref 범위에서 order(1 오름차순, 0 내림차순)로의 순위	RANK(F12, \$F\$12:\$F\$14, 0) = 3
<b>STDEV</b>	(number1,number2,..)	숫자들의 표준편차	STDEV(1, 3, 5) = 2
<b>VAR</b>	(number1,number2,..)	숫자들의 분산	VAR(1, 3, 5) = 4

## 3. 데이터 분석을 위한 주요 함수

**기본 텍스트, 날짜/시간 함수**

텍스트와 날짜/시간은 특수한 형태의 값으로 분류  
변환이 주된 관심사

함수명	입력 인수	함수 설명	예제
<b>CONCATENATE</b>	(text1,text2,..)	여러개의 문자열을 한개의 문자열로 통합	CONCATENATE("A","B","S") = "ABS"
<b>DATE</b>	(year,month,day)	정수 year, month, day 각각을 조합하여 날짜값으로 변환	DATE(2024, 8, 6) = "2024-08-06" = 45510
<b>DAY</b>	(serial_number)	날짜(연월일)에서 일값만 반환	DAY(45510) = 6
<b>FIND</b>	(find_text,within_text,start_num)	find_text를 within_text내 start_num부터 찾아, 해당 문자의 시작위치를 반환	FIND("is", "Time is money.", 1) = 6
<b>LEN</b>	(text)	문자열 내의 문자 개수 반환	LEN("Money") = 5
<b>MID</b>	(text,start_num,num_chars)	text의 start_num 위치에서부터 num_chars 만큼까지 문자열	MID("KOREA", 2, 2) = "OR"
<b>MONTH</b>	(serial_number)	날짜에서 월값만 반환	MONTH(45510) = 8
<b>NOW</b>	()	현재 날짜와 시각	NOW() = "2024-08-06 09:00:00"
<b>TEXT</b>	(value,format_text)	값(숫자)를 format_text 형식 문자열로 반환	TEXT(843, "00000") = "00843"
<b>TODAY</b>	()	현재 날짜	TODAY() = "2024-08-06"
<b>TRIM</b>	(text)	문자열의 양끝 공백을 없앰	TRIM(" 이진수 ") = "이진수"
<b>VALUE</b>	(text)	문자열인 숫자를 숫자로 변환	VALUE("00843") = 843
<b>WEEKDAY</b>	(serial_number,return_type)	날짜의 요일을 반환 (type0이 1일때, 1은 일요일, 7은 토요일)	WEEKDAY(45510) = 3
<b>YEAR</b>	(serial_number)	날짜에서 연값만 반환	YEAR(45510) = 2024

# 시계열에서 날짜 핸들링은 핵심 기술!

## 날짜(실수)로 or 문자열로?

- 날짜의 원래 값은 실수 : 2024-08-06 → 45510
- 다양한 날짜 data : 2024-08-06, 2024.08.06, 2024.08, 2024-08, ...
- 표준형(2024-08-06)외에 불러 오면 텍스트

## 월말, 주말 기준일은?

- 소스 데이터의 혼합 : 일간데이터, 주간데이터, 월간데이터
- 월말일 vs 영업일 기준 월말일
- 요일 기준 주말일 vs 영업일 기준 주말일

## 자주 활용되는 함수

함수명	입력 인수	함수 설명	예제
MID	(text,start_num,num_chars)	text의 start_num 위치에서부터 num_chars 만큼까지 문자열	MID("2024-08-06", 6, 2) = "08"
VALUE	(text)	문자열인 숫자를 숫자로 변환	VALUE("08") = 8
DATE	(year,month,day)	정수 year, month, day 각각을 조합하여 날짜값으로 변환	DATE(2024, 8, 6) = "2024-08-06" = 45510
DAY	(serial_number)	날짜(연월일)에서 일값만 반환	DAY(45510) = 6
MONTH	(serial_number)	날짜에서 월값만 반환	MONTH(45510) = 8
YEAR	(serial_number)	날짜에서 연값만 반환	YEAR(45510) = 2024

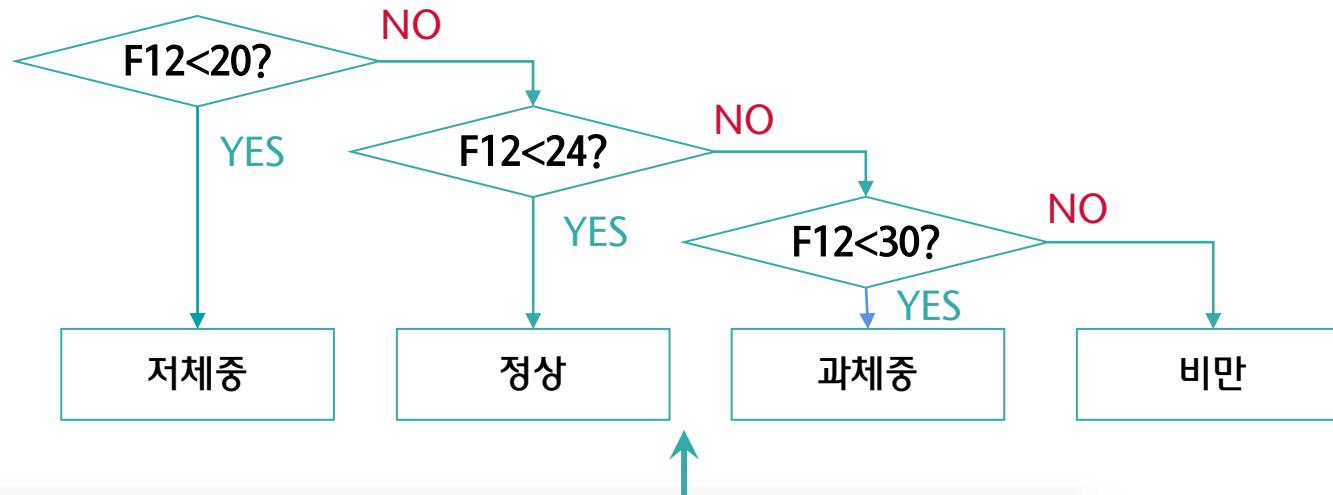
## 3. 데이터 분석을 위한 주요 함수

## 기본 논리 함수

조건의 참, 거짓에 의해 반환값을 달리 할 때

함수명	입력 인수	함수 설명	예제
<b>AND</b>	(logical1,logical2,...)	인수가 모두 참(TRUE)일 때, TRUE 반환	AND(2>1, 2<3) = TRUE AND(2>1, 2>3)= FALSE
<b>IF</b>	(logical_test,value_if_true,value_if_false)	logical_test의 결과가 참이면 value_if_true를 반환하고, 거짓이면 value_if_false를 반환	if(2>1, 1, 0) = 1 If(ISNA(VLOOKUP(A1, B:D, 3, 0)), “없음”, “참조있음”)
<b>IFERROR</b>	(value,value_if_error)	value가 오류인 경우 value_if_error를 반환, 그렇지 않으면 value 반환	IFERROR(2/0, 0) = 0
<b>IFNA</b>	(value,value_if_na)	Value가 #N/A면 value_if_na를 반환, 그렇지 않으면 value 반환	IFNA(#N/A, 1) = 1
<b>IFS</b>	(logical_test,value_if_true,...)	logical_test1이 참이면 value_if_true1 반환, 그렇지 않으면 logical_test2가 참이면 value_if_true2 반환, ...	ifs(1>2, 1, 3>2, 2) = 2
<b>OR</b>	(logical1,logical2,...)	인수중 하나라도 TRUE이면, TRUE 반환	OR(2>1, 2<3) = TRUE OR(2>1, 2>3) = TRUE

# 함수는 얼마든지 중첩 사용할 수 있다



	A	B	C	D	E	F	G	H
10	신체질량지수 (BMI; 카우프지수)							
11	이름	키(cm)	몸무게(kg)	키(m)제곱	비만도1	비만도2	순위	결과
12	홍길동	183	70	3.3	20.9	20.9	3	정상
13	이도	175	85	3.1	27.8	27.8	1	과체중
14	정약용	170	71	2.9	24.6	24.6	2	과체중

$$y = f(x) = h(g(x))$$

IF 함수의 경우 조건이 여러 개로 분기될 때, result행에 다시 IF함수 사용

## 3. 데이터 분석을 위한 주요 함수

## 기본 찾기/참조, 정보 함수

찾기/참조함수는 데이터를 추적하여 연결,  
정보함수는 에러 처리를 위해 유용한 함수들

함수명	입력 인수	함수 설명	예제
<b>HLOOKUP</b>	(lookup_value,table_array, row_index_num,range_lookup)	table_array에서 lookup_value를 검색하여, 같은 열의 지정한 row_index_num의 값을 반환. range_lookup이 FALSE이면 정확한 값, TRUE이면 가장 유사한 값 찾음	HLOOKUP(A14, 1:5, 5, FALSE) = 25 * 찾는 값이 없으면 #N/A 반환
<b>INDEX</b>	(array,row_num,[column_num])	array 범위에서 row_num째 행과 column_num째 열이 만나는 셀의 값	INDEX({1,2;3,4},2,1) = 3
<b>ISBLANK</b>	(value)	해당셀이 비어 있으면 TRUE	ISBLANK(A10) = TRUE
<b>ISERROR</b>	(value)	값이 오류값일 때, TRUE 반환	ISERROR(#REF) = TRUE
<b>ISNA</b>	(value)	값이 #N/A일 때, TRUE 반환	ISNA(#N/A) = TRUE
<b>ISTEXT</b>	(value)	값이 텍스트일 때, TRUE 반환	ISTEXT("1234") = TRUE
<b>MATCH</b>	(lookup_value,lookup_array ,match_type)	lookup_array 범위에서 lookup_value가 있는 위치의 순서값을 반환	MATCH({4,3,2,1}, 3, 0) = 2
<b>VLOOKUP</b>	(lookup_value,table_array,col_index_num,range_lookup)	table_array에서 lookup_value를 검색하여, 같은 행의 지정한 col_index_num의 값을 반환. range_lookup이 FALSE이면 정확한 값, TRUE이면 가장 유사한 값 찾음	VLOOKUP(A14, A8:F10, 5, FALSE) = 20.90 * 찾는 값이 없으면 #N/A 반환

## 3. 데이터 분석을 위한 주요 함수

VLOOKUP(lookup\_value, table\_array, col\_index\_num, range\_lookup)

# VLOOKUP 함수를 이용한 데이터 연결

- 서로 다른 데이터셋 간의 연결을 위해 필수적 함수
- 관계형 데이터에서 테이블간의 조인(Join)과 동일한 역할

학생 (b)

학번	이름	이메일	학년
A220001	홍길동	hgd@snowballproject.co.kr	2
C230001	정약용	jyy@snowballproject.co.kr	1
C210002	신사임당	saimdang@snowballproject.co.kr	3
A220002	유관순	yks@snowballproject.co.kr	2

과목 (c)

과목번호	과목명	학점
101001	영어	3
201001	수학	4

수강 (a)

학번	과목번호
A220001	201001
C230001	101001
C210002	101001
A220002	201001
A220002	101001

SELECT [조회할 필드들]

FROM [조회 필드가 존재하는 테이블들]

WHERE [조회 조건들]

SELECT b.이름, b.학년, b.이메일, c.과목명, c.학점

FROM 수강 a, 학생 b, 과목 c

WHERE a.학번 = b.학번

AND a.과목번호 = c.과목번호

Key를 이용한 연결

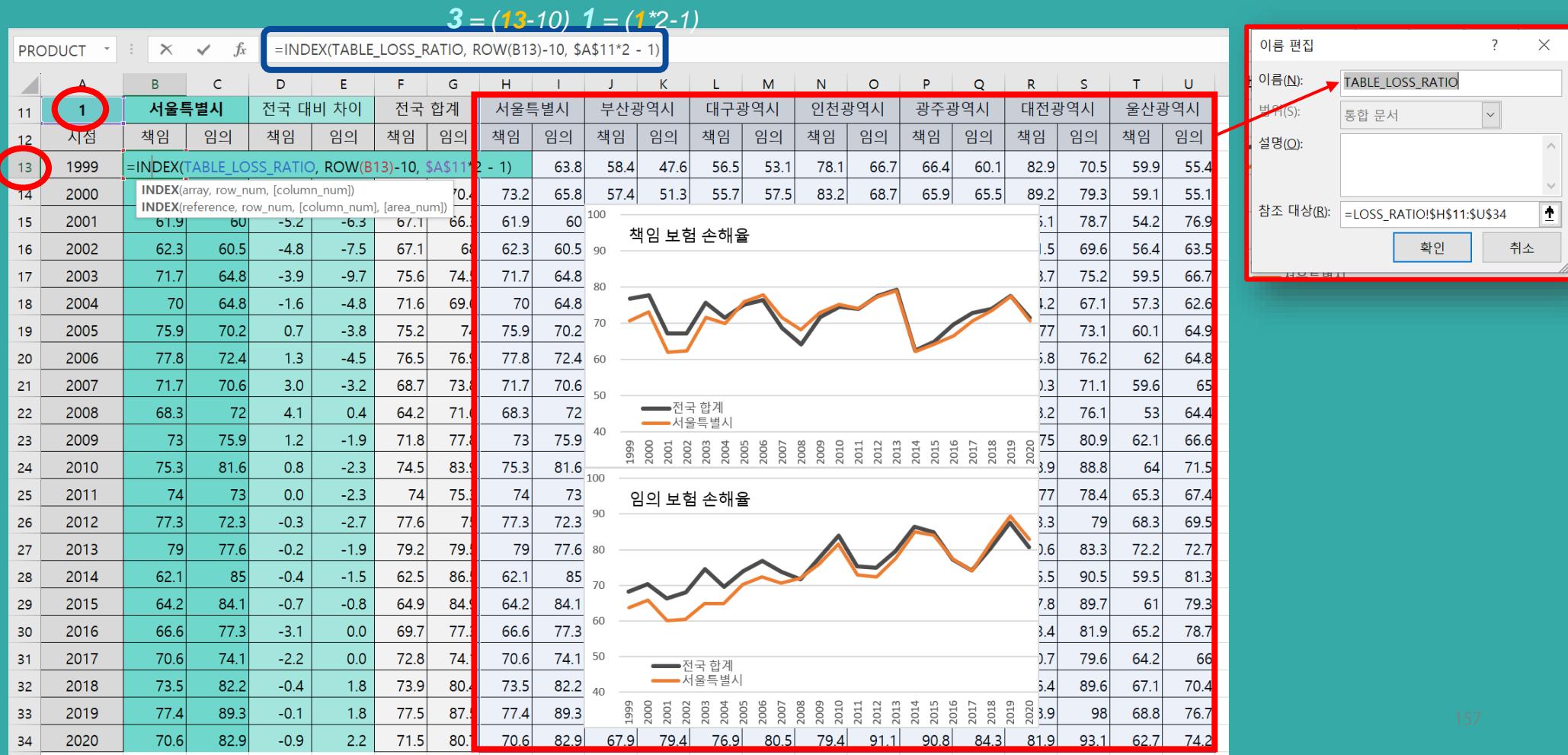
이름	학년	이메일	수강과목명	학점
홍길동	2	hgd@snowballproject.co.kr	수학	4
정약용	1	jyy@snowballproject.co.kr	영어	3
신사임당	3	saimdang@snowballproject.co.kr	영어	3
유관순	2	yks@snowballproject.co.kr	수학	4
유관순	2	yks@snowballproject.co.kr	영어	3

## 3. 데이터 분석을 위한 주요 함수

INDEX (array, row\_num, [column\_num])

## INDEX 함수를 활용한 데이터 전환

- 테이블 내의 위치 관계(행, 열)를 활용하여 다른 테이블의 데이터를 가져옴



## 핵심요약

- ‘데이터 획득’, ‘데이터 전처리’ 이후 ‘분석 방법론’을 적용하여 ‘통계적 분석’ 혹은 ‘머신러닝 분석’을 통해 결론 획득
- 실무 영역에서 ‘엑셀’은 강력한 도구이며, 보다 깊은 분석을 위해 ‘파이썬’을, 조직의 공유 플랫폼으로 ‘시각화 도구’ 활용
- 데이터베이스 뿐 아니라 엑셀을 비롯한 대부분의 데이터 처리 형태는 ‘관계형 데이터 모델’을 기반으로 설계되고 처리됨
- 컴퓨터는 입력–처리–출력의 구조로 움직이며, 컴퓨터를 도구로 사용한다는 것은 주로 입력하는 법을 익히는 것을 의미
- 함수는 입력값을 받아, 출력값을 돌려주는 가장 기본 단위 구조임.
- VLOOKUP은 서로 다른 테이블에 저장된 데이터를 연결해 주는 함수임

# 6. 데이터 획득과 데이터 전처리

---

# 학습목표

- ▶ 데이터를 획득할 수 있는 주요 소스들을 확인
- ▶ 다양한 종류의 데이터 형태를 읽는 방법을 엑셀 파워쿼리를 통해 습득
- ▶ 원 데이터를 분석에 적합한 데이터로 가공하는 데이터 전처리 방법론 이해

## 6. 데이터 획득과 데이터 전처리

### 1. 데이터 획득 방법과 데이터 형태

# 네비게이션은 어떻게 과속방지턱이 있는지 알까? 행정안전부 데이터 개방 - localdata.go.kr

The screenshot shows the LOCALDATA website interface. The top navigation bar includes tabs for LOCALDATA, 데이터받기, 데이터찾기, 생활편의정보 (which is highlighted with a teal border), and 데이터활용안내. Below the navigation is a search bar with dropdowns for '서울특별시' and '시군구(전체)', and input fields for '데이터기준일자 검색시작일' and '데이터기준일자 검색종료일'. A '검색' button is also present. The main content area is titled '생활밀착데이터 / 과속방지턱정보' and contains several categories: 공중화장실정보, 보호수정보, 자전거보관소정보, 생활쓰레기배출정보, 세차장정보, 과속방지턱정보 (highlighted with a teal border), 무료와이파이정보, CCTV정보, 안전비상벨위치정보, 낚시터정보, 비산먼지발생사업정보, and 민방위대피시설정보. A large teal arrow points from the 'Over-speed Prevention Barrier Information' button down to a data table. The table has columns for 번호, 과속방지턱형태구분, 주소, 과속방지턱재료, 관리기관명, and 데이터기준일자. The first two rows of data are shown: row 1 (번호 1, 원호형, 서울특별시 강서구 금낭화로24길 50, 아스팔트 콘크리트, 서울특별시 강서구청, 2024-03-08) and row 2 (번호 2, 원호형, 서울특별시 강서구 금낭화로24길 56, 아스팔트 콘크리트, 서울특별시 강서구청, 2024-03-08). A large teal box highlights the 'EXCEL' download button at the bottom right of the table. The table itself has a teal border around its data area.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
번호	과속방지 턱관리번 호	도로명	시도명	시군구명	소재지 도 로명 주소	소재지지 번주소	설치장소	과속방지 턱재료	과속방지 턱형태구 분	과속방지 턱높이	과속방지 턱폭	과속방지 턱연장	도로유형 구분	규격여부	WGS84위 도	WGS84경 도	보자분리 여부	연속형여 부	과속방지 턱설치연 도	관리기관 명	관리기관 전화번호	데이터기 준일자
2 1	500	탑골로	서울특별시금천구	서울특별시	도로	아스팔트 · 원호형	10	360	600	일반	Y	37.453349	126.91699	Y	N	2010	금천구청	02-2627-12020-08-31				
3 2	501	탑골로	서울특별시금천구	서울특별시	도로	아스팔트 · 원호형	10	360	500	일반	Y	37.452160	126.91594	Y	N	2010	금천구청	02-2627-12020-08-31				
4 3	502	탑골로8길	서울특별시금천구	서울특별시	도로	아스팔트 · 원호형	10	360	300	일반	Y	37.451082	126.91461	N	N	2010	금천구청	02-2627-12020-08-31				
5 4	596	시흥대로1	서울특별시금천구	서울특별시	도로	아스팔트 · 원호형	10	360	500	일반	Y	37.473679	126.89594	Y	N	2010	금천구청	02-2627-12020-08-31				
6 5	597	시흥대로1	서울특별시금천구	서울특별시	도로	아스팔트 · 원호형	10	360	600	일반	Y	37.475105	126.89996	Y	N	2010	금천구청	02-2627-12020-08-31				
7 6	598	시흥대로1	서울특별시금천구	서울특별시	도로	아스팔트 · 원호형	10	360	800	일반	Y	37.475341	126.90010	Y	N	2010	금천구청	02-2627-12020-08-31				

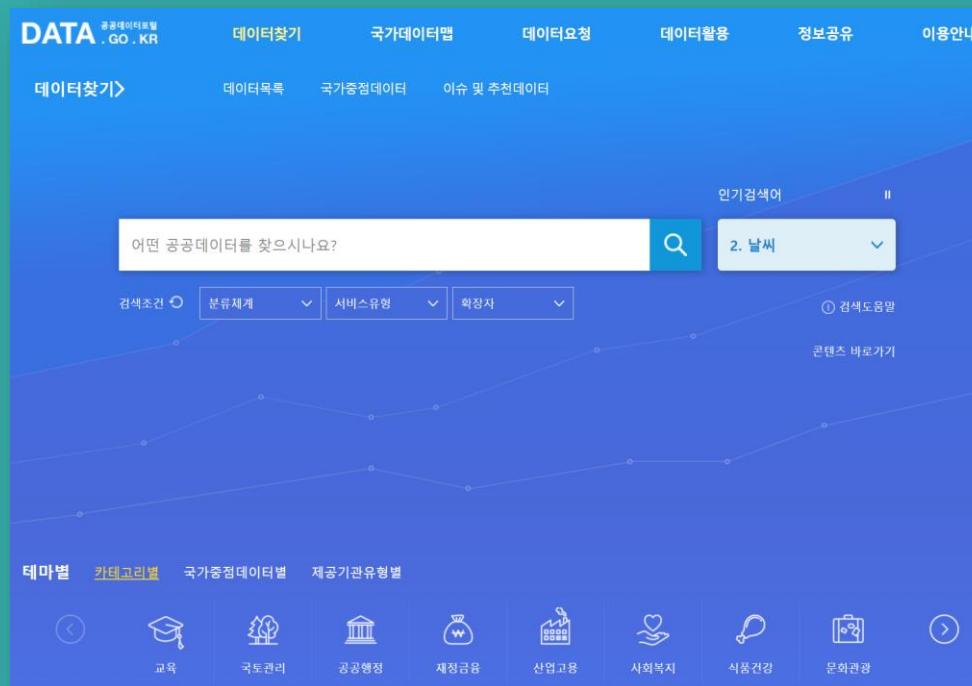
## 6. 데이터 획득과 데이터 전처리

### 1. 데이터 획득 방법과 데이터 형태

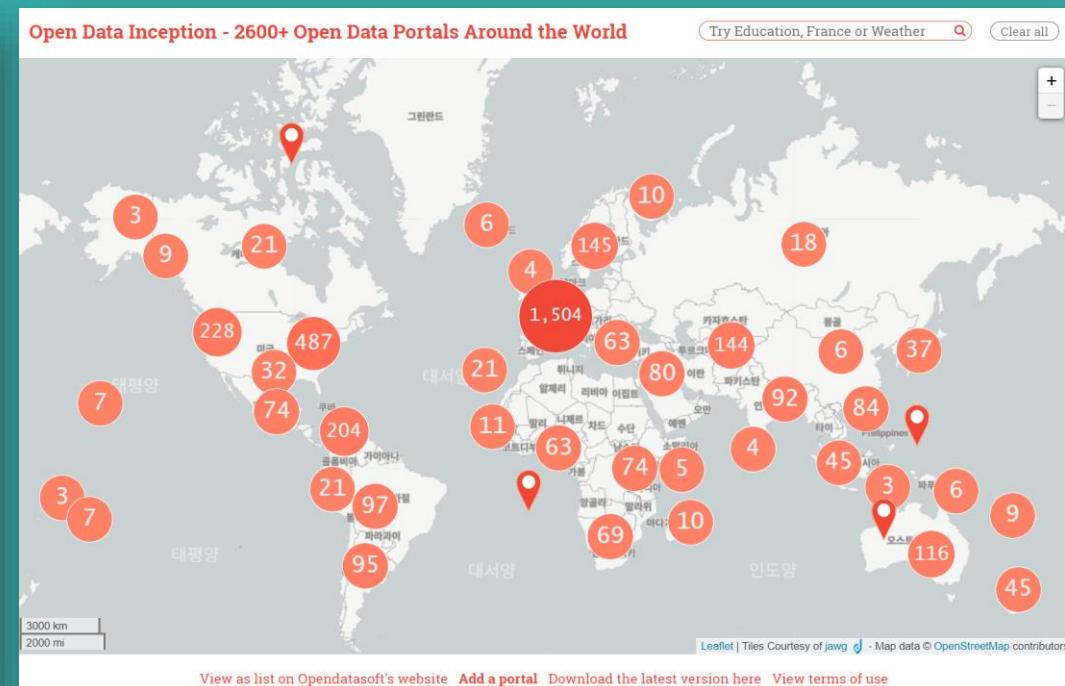
# 공공 데이터

공공기관 제공 데이터는 무료로 자유롭게 이용 가능

## 대한민국 공공데이터 포털 data.go.kr



## 세계 공공데이터 포털 opendatainception.io



## 6. 데이터 획득과 데이터 전처리

### 1. 데이터 획득 방법과 데이터 형태

# 통계 데이터

## 국가, 경제, 사회, 복지 등 신뢰성 있는 통계 자료의 획득

### 국가 통계 포털 kosis.kr

The screenshot shows the homepage of the KOSIS website. At the top, there is a banner with the text "이 누리집은 대한민국 공식 전자정부 누리집입니다." and "다시 찾았어요! 새로운 출발의 바다". Below the banner are links for "로그인", "회원가입", "English", and "어린이". The main navigation menu includes "국내통계", "국제·북한통계", "쉽게 보는 통계", "온라인간행물", "민원안내", "서비스 소개", and a search bar. A sidebar on the left lists categories such as "주제별 통계", "기관별 통계", "e-지방지표(통계표)", "e-지방지표(시각화)", "과거·중지통계", "실험적통계", "인구", "사회일반", "범죄·안전", "노동", and "스토리 서비스·교과서". A central content area displays a "통계목록" section with a search bar and several buttons for "주제전체보기", "관심주제설정", "목록받기", and "오름차순". A message at the bottom of this section says, "관심주제설정"에서 특정 주제를 선택한 경우 해당 주제의 목록만 보입니다. 전체 목록은 "주제 전체보기"를 클릭하세요.

### UN 통계 포털 data.un.org

The screenshot shows the homepage of the UNdata website. At the top, there is a logo with the text "un data A world of information". Below the logo, there is a navigation bar with links for "Datamarts", "Update Calendar", "Glossary", "API", and "More". A search bar is also present. A message below the navigation bar states "32 databases - 60 million records". A section titled "Other UNSD Databases:" lists "MBS", "SDG indicators", and "UNcomtrade". The main content area features a large orange banner titled "Popular statistical tables, country (area) and regional profiles". This banner contains a table of contents for "Population" with links to PDF and CSV files, updated on 18-Oct-2022. Other sections listed include "World", "Africa", "Northern Africa", "Sub-Saharan Africa", "Eastern Africa", "Middle Africa", "Southern Africa", "Western Africa", "Americas", and "Northern America".

## 1. 데이터 획득 방법과 데이터 형태

# 데이터 거래소 / 빅데이터 플랫폼

민간 데이터의 무료 유통 또는 유료 판매를 활성화하는 분야별 플랫폼

한국 데이터 거래소  
kdx.kr

금융 데이터 거래소  
findatamall.or.kr

The screenshot shows the main landing page of the KDX (한국데이터거래소) website. At the top, there is a navigation bar with links to '데이터 마켓', 'AI 비즈니스', '서비스', '커뮤니티', '기업회원 사이트', '데이터 바우처', a search icon, and a '로그인' button. Below the navigation bar, a large banner features the text '데이터와 사람이 만나는 공간 KDX'. The banner displays four key statistics: '10,445개 총 보유 데이터 수', '24,173명 총 회원 수', '3.09TB 보유 데이터 용량', and '43,308개 누적 거래 건수'. The background of the banner is dark blue with a network of white dots.

The screenshot shows the main landing page of the Financial Data Exchange (금융데이터거래소). At the top, there is a navigation bar with links to '데이터마켓', '데이터허브', '원격분석센터', 'AI테스트베드', and '맞춤데이터'. A search bar is located at the top right. Below the navigation bar, a prominent banner features the text '지금까지 이런 데이터 상품은 없었다!' (Until now, there were no such data products!). It lists several data products available for purchase, including '선택구매형 상품' (NH농협카드, KB캐피탈), '구독형 상품' (KCB, 신한카드), and '임대형 상품' (지역경제(가계) 통계데이터). To the right of the banner, there is a call-to-action section for '로그인' and '회원가입' (Login and Sign Up). Below this, a pink box highlights the '금융데이터거래소 멤버십' (Financial Data Exchange Membership) program, which offers various benefits to members.

## 6. 데이터 획득과 데이터 전처리

### 1. 데이터 획득 방법과 데이터 형태

# 통합 데이터 검색

## 공공/민간 데이터에 대한 종합 검색

통합 데이터 지도  
bigdata-map.kr

통합 데이터 지도

bigdata-map.kr

찾고 싶은 데이터를 검색하세요.

키워드 검색

인기검색어 | #특허

문화체육 교육과학 교통통신 경제금융 건설에너지 농수산식품 제조소비 건강의료복지 재!

인기 데이터

카테고리	데이터셋 명
교통통신	전국자동차정비업체표준데이터
교통통신	성남시 비정형도로 CCTV 이미지 데이터
문화체육	해외여행 TOP10 트렌드 데이터
교통통신	VDS기반 고속도로 지점별 교통 소통 통계 데이터(1시간 단위)

공지사항/이벤트

공간융합 빅데이터 플랫폼 홈페이지 신규가입 이벤트!  
참여기간: '23년 7월 1일 ~ 7월 31일(한 달간)  
공간융합 빅데이터 플랫폼

구글 데이터 서치  
datasetsearch.research.google.com

Google

검색어: 교통

데이터세트 100개 이상 발견됨

국토교통부\_국토교통부\_도로 교통량 데이터 AI 학습용 영상 데이터세트  
data.go.kr  
jpg  
Feb 16, 2021에 업데이트됨

도로교통공단\_교통사고 GIS정보  
data.go.kr  
xls  
Jun 20, 2023에 업데이트됨

교통시설(현황)  
data.nsdi.go.kr  
dataportal.asia  
Jan 7, 2018에 업데이트됨

국토교통부\_교통물류 분야 사전

다음 웹페이지에서 살펴보기: [data.go.kr](https://data.go.kr/ugs/selectPortalPolicyView.do)

## 6. 데이터 획득과 데이터 전처리

### 1. 데이터 획득 방법과 데이터 형태

# 공개 데이터 파일 제공

데이터셋을 파일 단위로 제공 (CSV, XLS, JSON, XML)

전체(481건) 파일데이터(417건) 오픈 API(64건) 표준데이터셋0개(0건)

정확도순 10개씩 정렬

공공데이터 포털에서 '관광지' 검색 사례

파일데이터 (417건)

미리보기

파일데이터 (417건)

미리보기

미리보기

민원서비스 게시판

## 6. 데이터 획득과 데이터 전처리

### 1. 데이터 획득 방법과 데이터 형태

# 오픈 API

※ API(Application Programming Interface)  
→ 프로그램 개발을 위해 미리 정해 놓은 약속 체계

## 데이터셋을 인터넷 주소로 실시간 제공하여 자동화 가능 (JSON, XML)

공공데이터 포털에서 '관광지' 검색 사례

The screenshot shows the Public Data Portal's search interface. At the top, there are four tabs: '전체(481건)', '파일데이터(417건)', '오픈 API(64건)' (which is highlighted with a blue border), and '표준데이터셋(0건)'. Below the tabs, there are filters for '정확도순' (sorted by relevance), '10개씩' (show 10 items at a time), and a '정렬' (sort) button. A large green speech bubble on the right says '공공데이터 포털에서 '관광지' 검색 사례'. The main content area displays three search results for tourism-related APIs:

- 인천광역시\_관광진흥/관광지** (JSON): Provides information about tourism and attractions in Incheon. It was published on September 2, 2022, with 30,441 views and 916 usage requests. The keywords are tourism and attractions.
- 경기도\_관광지\_현황** (XML): Provides information about tourism sites in Gyeonggi-do. It was published on October 8, 2021, with 5,467 views and 1,046 usage requests. The keywords are tourism sites and travel guides.
- 대구광역시\_관광지** (XML): Provides information about tourism in Daegu. It was published on October 26, 2021, with 12,641 views and 384 usage requests. The keywords are Daegu tourism and travel.

A blue circular icon in the bottom right corner says '민원서비스 게시판' (Citizen Service Bulletin Board).

## 6. 데이터 획득과 데이터 전처리

### 1. 데이터 획득 방법과 데이터 형태

# 웹 페이지 (HTML)

※ HTML(HyperText Markup Language)  
→ 웹페이지 문서를 작성하는 약속 체계

## 웹페이지 내 <TABLE> 태그로 표시한 데이터 셋

A screenshot of a web browser displaying a table from the Dart FSS website. The table has columns for date, change reason, type of transaction, change amount before, change amount after, and notes. The data shows various transactions like purchases and sales of stocks.

변경일	변경원인	주식의 종류	변경전주식수	증감주식수	변경후주식수	비고
2023-06-01	장내매수(+)	보통주식	9,919,790	6,263	9,926,053	-
2023-06-02	장내매도(-)	보통주식	9,926,053	-48,190	9,877,863	-
2023-06-02	장내매수(+)	보통주식	9,877,863	4,163	9,882,026	-
2023-06-05	장내매도(-)	보통주식	9,882,026	-2,823	9,879,203	-
2023-06-05	장내매수(+)	보통주식	9,879,203	6,804	9,886,007	-
2023-06-07	장내매도(-)	보통주식	9,886,007	-20,296	9,865,711	-
2023-06-07	장내매수(+)	보통주식	9,865,711	2,576	9,868,287	-
2023-06-08	장내매도(-)	보통주식	9,868,287	-8,768	9,859,519	-
2023-06-08	장내매수(+)	보통주식	9,859,519	2,886	9,862,405	-
2023-06-09	장내매도(-)	보통주식	9,862,405	-10,078	9,852,327	-
2023-06-09	장내매수(+)	보통주식	9,852,327	7,113	9,859,440	-
2023-06-12	장내매도(-)	보통주식	9,859,440	-114,876	9,744,564	-
2023-06-12	장내매수(+)	보통주식	9,744,564	10,685	9,755,249	-
2023-06-13	장내매도(-)	보통주식	9,755,249	-9,648	9,745,601	-
2023-06-13	장내매수(+)	보통주식	9,745,601	587	9,746,188	-

기업공시 자료 조회 사례 (dart.fss.or.kr)

## 웹 데이터 다운로드 (크롤링)

A screenshot of a spreadsheet application showing the downloaded data from the table above. The data is presented in a tabular format with columns labeled by the headers from the table. The rows show individual transaction entries with dates, reasons, types, and amounts.

변경일	변경원인	주식의 종류	변경전주식수	증감주식수	변경후주식수	비고
2023-06-01	장내매수(+)	보통주식	9919790	6263	9926053	-
2023-06-02	장내매도(-)	보통주식	9926053	-48190	9877863	-
2023-06-02	장내매수(+)	보통주식	9877863	4163	9882026	-
2023-06-05	장내매도(-)	보통주식	9882026	-2823	9879203	-
2023-06-05	장내매수(+)	보통주식	9879203	6804	9886007	-
2023-06-07	장내매도(-)	보통주식	9886007	-20296	9865711	-
2023-06-07	장내매수(+)	보통주식	9865711	2576	9868287	-
2023-06-08	장내매도(-)	보통주식	9868287	-8768	9859519	-
2023-06-08	장내매수(+)	보통주식	9859519	2886	9862405	-
2023-06-09	장내매도(-)	보통주식	9862405	-10078	9852327	-
2023-06-09	장내매수(+)	보통주식	9852327	7113	9859440	-
2023-06-12	장내매도(-)	보통주식	9859440	-114876	9744564	-
2023-06-12	장내매수(+)	보통주식	9744564	10685	9755249	-
2023-06-13	장내매도(-)	보통주식	9755249	-9648	9745601	-
2023-06-13	장내매수(+)	보통주식	9745601	587	9746188	-
2023-06-14	장내매도(-)	보통주식	9746188	-97835	9648353	-
2023-06-14	장내매수(+)	보통주식	9648353	964	9649317	-
2023-06-15	장내매도(-)	보통주식	9649317	-61249	9588068	-
2023-06-15	장내매수(+)	보통주식	9588068	12318	9600386	-

## 2. 파워쿼리를 활용한 데이터 입력

[예제] APT서울매매지수 불러오기

## 파일 데이터 엑셀로 가져오기

data\_hanaro/file\_data/

CSV

날짜, 매매지수
1998-12-31, 20.5
1999-01-31, 21.13
1999-02-28, 21.22
1999-03-31, 21.37
1999-04-30, 21.48
1999-05-31, 21.66
1999-06-30, 21.81
1999-07-31, 22.16
1999-08-31, 22.66
1999-09-30, 23.08
1999-10-31, 23.06
1999-11-30, 22.95
1999-12-31, 23.06
2000-01-31, 23.32
2000-02-29, 23.67
2000-03-31, 23.89
2000-04-30, 23.89
2000-05-31, 23.87
2000-06-30, 23.87
2000-07-31, 24.09
2000-08-31, 24.22
2000-09-30, 24.37
2000-10-31, 24.5
2000-11-30, 24.31
2000-12-31, 24.02
2001-01-31, 24.04
2001-02-28, 24.33
2001-03-31, 24.55

TXT(탭)

날짜 매매지수
1998-12-31 20.5
1999-01-31 21.13
1999-02-28 21.22
1999-03-31 21.37
1999-04-30 21.48
1999-05-31 21.66
1999-06-30 21.81
1999-07-31 22.16
1999-08-31 22.66
1999-09-30 23.08
1999-10-31 23.06
1999-11-30 22.95
1999-12-31 23.06
2000-01-31 23.32
2000-02-29 23.67
2000-03-31 23.89
2000-04-30 23.89
2000-05-31 23.87
2000-06-30 23.87
2000-07-31 24.09
2000-08-31 24.22
2000-09-30 24.37
2000-10-31 24.5
2000-11-30 24.31
2000-12-31 24.02
2001-01-31 24.04
2001-02-28 24.33
2001-03-31 24.55

JSON

```
[{"날짜": "1998-12-31", "매매지수": 20.5}, {"날짜": "1999-01-31", "매매지수": 21.13}, {"날짜": "1999-02-28", "매매지수": 21.22}, {"날짜": "1999-03-31", "매매지수": 21.37}, {"날짜": "1999-04-30", "매매지수": 21.48}, {"날짜": "1999-05-31", "매매지수": 21.66}, {"날짜": "1999-06-30", "매매지수": 21.81}, {"날짜": "1999-07-31", "매매지수": 22.16}, {"날짜": "1999-08-31", "매매지수": 22.66}, {"날짜": "1999-09-30", "매매지수": 23.08}, {"날짜": "1999-10-31", "매매지수": 23.06}, {"날짜": "1999-11-30", "매매지수": 22.95}, {"날짜": "1999-12-31", "매매지수": 23.06}, {"날짜": "2000-01-31", "매매지수": 23.32}, {"날짜": "2000-02-29", "매매지수": 23.67}, {"날짜": "2000-03-31", "매매지수": 23.89}, {"날짜": "2000-04-30", "매매지수": 23.89}, {"날짜": "2000-05-31", "매매지수": 23.87}, {"날짜": "2000-06-30", "매매지수": 23.87}, {"날짜": "2000-07-31", "매매지수": 24.09}, {"날짜": "2000-08-31", "매매지수": 24.22}, {"날짜": "2000-09-30", "매매지수": 24.37}, {"날짜": "2000-10-31", "매매지수": 24.5}, {"날짜": "2000-11-30", "매매지수": 24.31}, {"날짜": "2000-12-31", "매매지수": 24.02}, {"날짜": "2001-01-31", "매매지수": 24.04}, {"날짜": "2001-02-28", "매매지수": 24.33}, {"날짜": "2001-03-31", "매매지수": 24.55}]
```

XML

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Root xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<text>
<날짜>36160</날짜>
<매매지수>20.5</매매지수>
</text>
<text>
<날짜>36191</날짜>
<매매지수>21.13</매매지수>
</text>
<text>
<날짜>36219</날짜>
<매매지수>21.22</매매지수>
</text>
<text>
<날짜>36250</날짜>
<매매지수>21.37</매매지수>
</text>
<text>
<날짜>36280</날짜>
<매매지수>21.48</매매지수>
</text>
<text>
<날짜>36311</날짜>
<매매지수>21.66</매매지수>
</text>
<text>
<날짜>36341</날짜>
<매매지수>21.81</매매지수>
</text>
<text>
<날짜>36370</날짜>
<매매지수>22.16</매매지수>
</text>
```

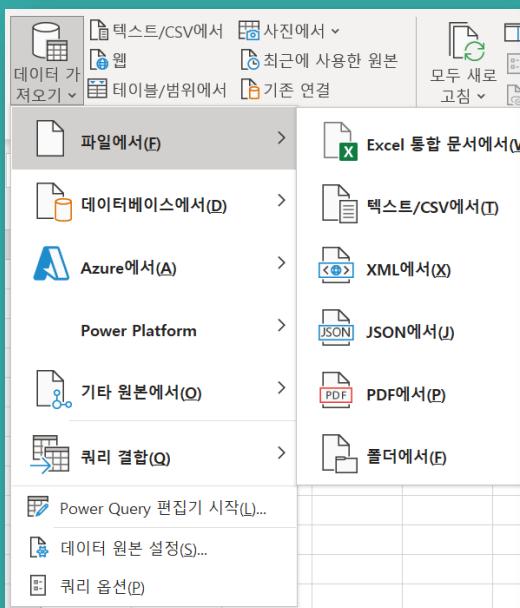
## 2. 파워쿼리를 활용한 데이터 입력

[예제] APT서울매매지수 불러오기

# 파워 쿼리 [데이터 가져오기]로 테이블 만들기

관계형 DB 구조의 테이블을 디자인하고, 실제 테이블처럼 사용 가능

## 메뉴



## Power Query 편집기

The screenshot shows the Power Query Editor window. The ribbon has 'File', 'Home', 'Transform', 'Add', 'View', and 'Help' tabs. The 'Transform' tab is selected. The main area shows a table with two columns: '날짜' (Date) and '매매지수' (Real Estate Price Index). Below the table, the formula bar shows '= Table.TransformColumnTypes(#"승격된", {{"날짜", type date}, {"매매지수", type number}})'. On the right, there's a 'Query Settings' pane with 'Name' set to 'APT전국매매지수' and 'Applied Changes' section showing 'Original Source' and 'Transformed Headers'.

## 데이터 로드 &amp; 관리

The screenshot shows the Power BI Data Load & Management pane. It lists a single query named 'APT전국매매지수' which contains 295 rows. The table shows data with columns '날짜' and '매매지수'.

날짜	매매지수
1998-12-31	20.5
1999-01-31	21.13
1999-02-28	21.22
1999-03-31	21.37
1999-04-30	21.48
1999-05-31	21.66
1999-06-30	21.81
1999-07-31	22.16
1999-08-31	22.66
1999-09-30	23.08
1999-10-31	23.06
1999-11-30	22.95
1999-12-31	23.06
2000-01-31	23.32
2000-02-29	23.67
2000-03-31	23.89
2000-04-30	23.89
2000-05-31	23.87
2000-06-30	23.87
2000-07-31	24.09
2000-08-31	24.22
2000-09-30	24.37
2000-10-31	24.5
2000-11-30	24.31

## 6. 데이터 획득과 데이터 전처리

### 2. 파워쿼리를 활용한 데이터 입력

[예제] 네이버 데이터 다운 받기

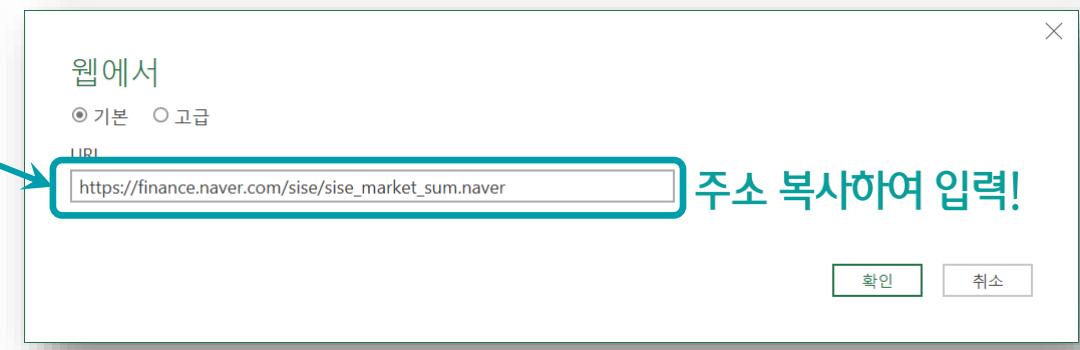
네이버 증권(시가총액)  
finance.naver.com



A screenshot of the finance.naver.com website. The address bar shows 'finance.naver.com/sise/sise\_market\_sum.naver'. The main navigation menu includes '증권 품', '국내증시' (selected), '해외증시', '시장지표', '리서치', '뉴스', and 'MY'. Below the menu, a message says '이 광고는 더 이상 보이지 않습니다'. The main content area is titled '시가총액' and displays various filtering options like '거래량', '시가', '고가', '저가', '매수호가', '매도호가', etc. A large table below lists top Korean stocks by market capitalization, including columns for name, current price, change, volume, market cap, P/E ratio, ROE, and turnover rate.

# 웹사이트 데이터 가져오기

[데이터 가져오기]-[기타 원본에서]-[웹]



[https://finance.naver.com/sise/sise\\_market\\_sum.naver](https://finance.naver.com/sise/sise_market_sum.naver)

## 2. 파워쿼리를 활용한 데이터 입력

[예제] 네이버 데이터 다운 받기

## 웹 페이지(HTML) 내에도 데이터는 테이블 형태!

The screenshot shows the Microsoft Power Query interface with the following details:

- Left Panel (Source):** Shows a list of sources. The first item, "https://finance.naver.com/sise/sise\_market\_su...", is highlighted with a red box. The second item, "코스피", is also highlighted with a red box and has a green arrow pointing to it from the text "선택시 미리보기" (Preview when selected).
- Right Panel (Preview):** A table titled "코스피" showing stock market data. The columns are: N, 종목명 (Stock Name), 현재가 (Current Price), 전일비 (Change from Previous Day), 등락률 (Change Rate), 액면가 (Face Value), 시가총액 (Market Capitalization), and 상장 (Listed). The data includes 14 rows of Korean stocks.
- Bottom Panel:** Buttons for "로드" (Load), "데이터 변환" (Data Transformation), and "취소" (Cancel).

N	종목명	현재가	전일비	등락률	액면가	시가총액	상장
1	삼성전자	71000	700	-0.98%	100	4238546	
2	LG에너지솔루션	561000	1000	0.18%	500	1312740	
3	SK하이닉스	115400	2100	-1.79%	5000	840115	
4	삼성바이오로직스	738000	3000	-0.40%	2500	525264	
5	LG화학	696000	9000	1.31%	5000	491323	
6	삼성전자우	59500	300	-0.50%	100	489618	
7	삼성SDI	687000	8000	-1.15%	5000	472412	
8	현대차	200500	1500	-0.74%	5000	424121	
9	POSCO홀딩스	498500	1000	-0.20%	5000	421588	
10	포스코퓨처엠	470500	9500	-1.98%	500	364464	
11	기아	84500	1500	-1.74%	5000	339727	
12	NAVER	205500	1000	0.49%	100	337121	
13	현대모비스	242000	2000	-0.82%	5000	228170	
14	카카오	50700	300	0.60%	100	225241	

## 2. 파워쿼리를 활용한 데이터 입력

[예제] 네이버 데이터 다운 받기

## 한 번 연결하면, 클릭 한번에 최신 데이터 업데이트

쿼리 및 연결

쿼리 | 연결

1개 쿼리

코스피  
50개의 행이 로드되었습니다.

‘새로고침’ 클릭하면,  
최신 데이터 업데이트

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	N	종목명	현재가	전일비	등락률	액면가	시가총액	상장주식수	외국인비율	거래량	PER	ROE	토론실
2	1	삼성전자	71100	600	-0.0084	100	4244515	5969783	52.99	5820747	10.73	17.07	
3	2	LG에너지솔루션	562000	2000	0.0036	500	1315080	234000	5.2	202037	126.29	5.75	
4	3	SK하이닉스	116200	1300	-0.0111	5000	845939	728002	51.92	1167981	-36.31	3.56	
5	4	삼성바이오로직스	738000	3000	-0.004	2500	525264	71174	10.44	19322	66.05	11.42	
6	5	LG화학	697000	10000	0.0146	5000	492029	70592	46.76	201944	33.45	6.95	
7	6	삼성전자우	59600	200	-0.0033	100	490440	822887	72.63	523193	9	N/A	
8	7	삼성SDI	688000	7000	-0.0101	5000	473100	68765	49.4	129325	23.8	12.52	
9	8	현대차	200500	1500	-0.0074	5000	424121	211532	32.73	285032	6.1	9.36	
10	9	POSCO홀딩스	500000	500	0.001	5000	422856	84571	35.76	921646	19.82	6.11	
11	10	포스코퓨처엠	471000	9000	-0.0187	500	364852	77463	7.89	1080422	301.15	4.87	
12	11	기아	84600	1400	-0.0163	5000	340129	402044	38.25	544057	5.28	14.57	
13	12	NAVER	205500	1000	0.0049	100	337121	164049	47.01	364934	52.34	3.29	
14	13	현대모비스	243000	1000	-0.0041	5000	229113	94285	37.42	88264	8.17	6.80	
15	14	카카오	50700	300	0.006	100	225241	444263	25.87	707067	177.27	13.54	
16	15	셀트리온	150300	400	0.0027	1000	220025	146391	20.08	232469	37.61	13.35	
17	16	LG전자	121700	1800	-0.0146	5000	199159	163648	29.71	311321	49.9	6.61	
18	17	KB금융	48400	250	0.0052	5000	195299	403511	72.21	403548	4.47	9.17	
19	18	삼성물산	103000	300	-0.0029	100	191159	185592	18.19	73990	8.99	7.16	
20	19	신한지주	33850	300	0.0089	5000	175460	518347	59.04	331801	3.87	9.70	
21	20	SK이노베이션	175300	5800	0.0342	5000	162092	92466	23.37	1442528	25.5	8.49	
22	21	LG	88100	1700	0.0197	5000	138582	157301	35.95	87982	12.84	8.48	
23	22	삼성생명	69200	100	0.0014	500	138400	200000	16.65	53323	6.85	5.73	
24	23	한국전력	20200	0	0	5000	129677	641964	14.77	275177	-0.55	-46.91	
25	24	HD현대중공업	142500	2300	-0.0159	5000	126502	88773	7.32	92494	-64.31	-6.47	
26	25	카카오뱅크	25550	50	0.002	5000	121814	476767	14.46	383416	40.81	4.68	
27	26	하나금융지주	39050	300	0.0077	5000	115550	295903	68.21	313933	3.08	10.06	
28	27	삼성전기	152800	400	-0.0026	5000	114132	74694	31.92	153362	15.27	13.76	
29	28	하이브	271000	7500	0.0285	500	112877	41652	16.13	157981	251.22	1.87	

## 2. 파워쿼리를 활용한 데이터 입력

# 데이터베이스 데이터 가져오기 - 시연

The screenshot shows the HeidiSQL interface version 11.3.0.6295. On the left, the 'DB 서버이름' (Database Server Name) sidebar lists several databases: information\_schema, mysql, performance\_schema, res, sn (selected), sys, and test. A green callout box highlights the 'sn' database with the text '선택된 데이터베이스' (Selected Database). On the right, the main pane displays a table of tables within the 'sn' database. The table has columns: 이름 (Name), 행 (Rows), 크기 (Size), 생성됨 (Created At), 엔진 (Engine), 코멘트 (Comment), and 유형 (Type). The table lists 15 tables related to YouTube data, such as yt\_clip, yt\_like, yt\_reply, etc. A green callout box highlights the entire table area with the text '선택된 데이터베이스를 구성하는 테이블들' (Tables that make up the selected database).

이름	행	크기	생성됨	엔진	코멘트	유형
yt_clip	183	144.0 Kib	2023-01-17 13:4...	InnoDB	유튜브 영상 정보	Table
yt_like	4	48.0 Kib	2023-01-17 13:4...	InnoDB	youtube 좋아요/싫...	Table
yt_reply	0	48.0 Kib	2023-01-17 13:4...	InnoDB	youtube 컨텐츠 댓글	Table
yt_reply_like	0	48.0 Kib	2023-01-17 13:4...	InnoDB	youtube 사용자 의...	Table
yt_reply_reply	0	48.0 Kib	2023-01-17 13:4...	InnoDB	youtube 컨텐츠 댓...	Table
yt_series	11	32.0 Kib	2023-01-17 13:4...	InnoDB	영상 시리즈	Table
search_request_keyword	70	32.0 Kib	2023-01-10 17:3...	InnoDB	사용자가 입력한 최...	Table
search_request_keyword_split	87	32.0 Kib	2023-01-10 17:3...	InnoDB	제출된 검색어를 단...	Table
search_result_youtube	746	80.0 Kib	2023-01-10 17:3...	InnoDB	유튜브 내에서의 검...	Table
yt_keyword_index	1,764	208.0 Kib	2023-01-10 17:2...	InnoDB	yt_clip에 등록된 #k...	Table
yt_channel	2	32.0 Kib	2023-01-05 14:5...	InnoDB	YouTube 채널	Table
yt_clip_stat	6,886	2.9 MiB	2023-01-03 18:0...	InnoDB	유튜브 영상 통계자...	Table
yt_group	25	48.0 Kib	2023-01-03 18:0...	InnoDB	YouTube 채널 재생...	Table
yt_clip_recommend	10	16.0 Kib	2023-01-03 18:0...	InnoDB	재키의 추천 영상 ...	Table

## 2. 파워쿼리를 활용한 데이터 입력

## 선택 테이블의 필드 현황

호스트: 127.0.0.1 | 데이터베이스: sn | 테이블: yt\_clip\_stat | 데이터 | 쿼리 |

기본 옵션 인덱스 (3) 외래 키 (0) 제약 조건 확인 (0) 분할 CREATE 코드 ALTER 코드

이름: yt\_clip\_stat 테이블 이름  
코멘트: 유튜브 영상 통계자료 및 성과

열: + 추가 × 제거 ▲ 위로 ▼ 아래로

선택된 테이블을 구성하는 필드들

#	이름	데이터 유형	길이/설정	부호 ...	NULL ...	0으...	기본값	코멘트
1	id	INT	11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	AUTO_INCREMENT	p.key
2	as_of_date	DATETIME		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	curdate()	통계기준일
3	yt_id	VARCHAR	12	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	"	영상 유튜브 ID(yt_clip.yt_id)
4	cur_title	VARCHAR	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL	영상의 현재 제목 (조회 당시)
5	total_score1	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0.5)	28일 종합 점수
6	total_score2	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0.5)	90일 종합 점수
7	before_score1	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0.5)	28일 사전 흥미 점수
8	before_score2	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0.5)	90일 사전 흥미 점수
9	after_score1	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0.5)	28일 시청 둘입도 점수
10	after_score2	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0.5)	90일 시청 둘입도 점수
11	exposure1	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0)	노출수 (28일)
12	exposure2	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0)	노출수 (90일)
13	click_rate1	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0)	노출클릭률 (28일)
14	click_rate2	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0)	노출클릭률 (90일)
15	view_count1	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0)	조회수 (28일)
16	view_count2	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0)	조회수 (90일)
17	view_hours1	FLOAT		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	abs(0)	시청시간 (28일)

## 2. 파워쿼리를 활용한 데이터 입력

## 여러 테이블에 있는 데이터를 원하는 형태로 가져오는 SQL 조회

SQL문 편집창

```

1 SELECT a.keyword, COUNT(b.yt_id) 영상수, SUM(b.view_count1) 조회수, SUM(b.exposure1) 노출수,
2      SUM(b.view_count1)/SUM(b.exposure1) 클릭률
3 FROM yt_keyword_index a, yt_clip_stat b → 2개의 테이블에서 조합하여 데이터 획득
4 WHERE b.as_of_date=(SELECT MAX(as_of_date) FROM yt_clip_stat)
5     AND a.yt_id=b.yt_id
6 GROUP BY a.keyword
7 ORDER BY SUM(b.view_count1) DESC

```

yt\_keyword\_index (1,089r × 5c)

keyword	영상수	조회수	노출수	클릭률
신용카드	4	1,508	13,423	0.1123
신용대출	7	1,495	13,371	0.1118
담보대출	6	1,492	13,385	0.1115
카드론	2	1,484	13,227	0.1122
저축은행	3	1,426	12,197	0.1169
소득	11	1,314	11,349	0.1158
이자	2	1,190	9,980	0.1192
재산	2	1,189	9,914	0.1199
마이너스통장	3	1,189	9,948	0.1195
연체	1	1,188	9,891	0.1201

SQL문에 따라 조회된 결과

## 2. 파워쿼리를 활용한 데이터 입력

## 엑셀에 테이블 데이터 가져오기

**마이크로소프트 DB**

**윈도우에서 미리 셋업한 DB 선택**

**선택한 DB의 테이블 리스트**

**선택한 테이블의 데이터 미리보기**

**이후 다른 파워 쿼리 사용법과 동일**

The screenshot shows the Microsoft Power Query ribbon with several options highlighted in blue boxes:

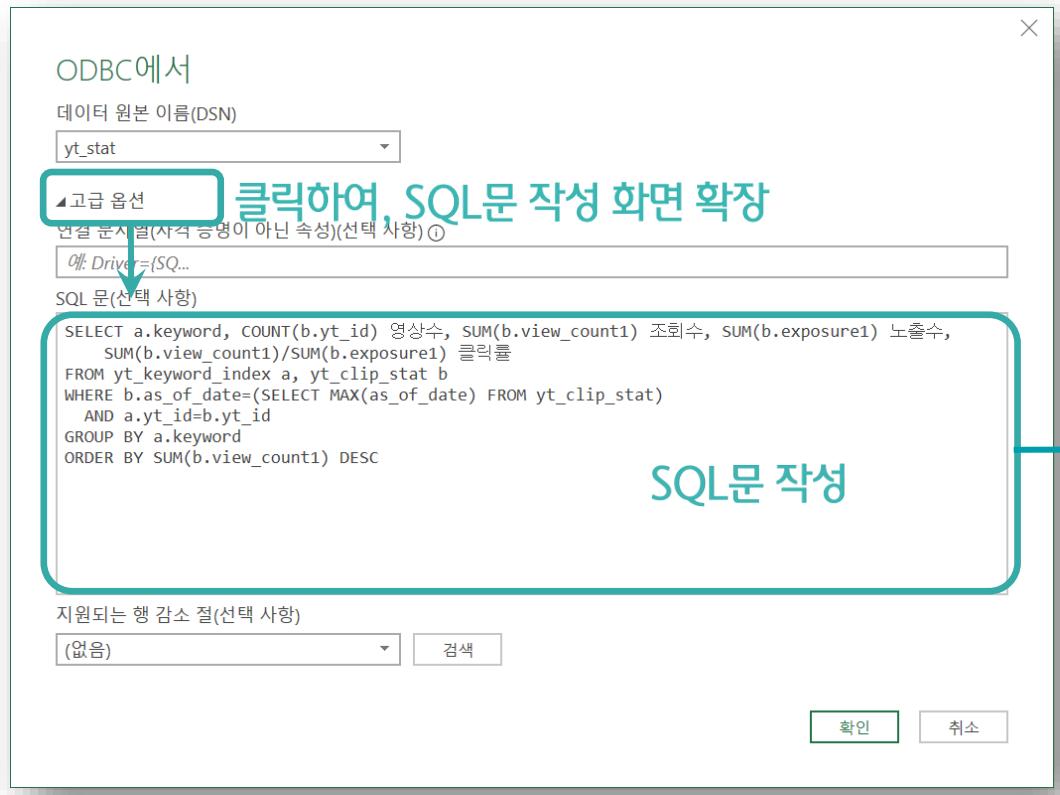
- 데이터 가져오기 > 데이터베이스에서(D) (highlighted)
- 파일에서(F)
- Azure에서(A)
- Power Platform 기타 데이터베이스
- 기타 원본에서(O) (highlighted)
- 쿼리 결합(Q)
- Power Query 편집기 시작(L)...
- 데이터 원본 설정(S)...
- 쿼리 옵션(P)
- ODBC에서(D) (highlighted)
- OLEDB에서(B)
- 사진에서(F)
- 빈 쿼리(Q)

The "ODBC에서(D)" option is selected, opening a dialog box titled "ODBC에서" (ODBC from). It shows a dropdown menu for "데이터 원본 이름(DSN)" with "yt\_stat" selected. Below it is a "선택한 DB의 테이블 리스트" (List of tables in selected DB) window showing a table named "yt\_clip\_stat" with 16 rows of data.

ID	as_of_date	yt_id	cur_title
1	2022-12-28 오전 12:00:00	KHPwHd4G-E	은퇴 준비, 30대부터 하라 1편 - 연령대별 자산, 소득, 지출 추세
2	2022-12-28 오전 12:00:00	efbKmSR12-E	은퇴 준비, 30대부터 하라 2편 - 소득의 종류와 연령대별 소비 추세
3	2022-12-28 오전 12:00:00	lbeDX25i4zU	가계자산의 중심, 부동산 1편 - 거시적 국가 자산/부채 구성
4	2022-12-28 오전 12:00:00	Ddeig6Ti0D8	가계자산의 중심, 부동산 2편 - 연령대별 자산/부채 구성 현황
5	2022-12-28 오전 12:00:00	O1Hjw57KPUM	불확실성 확대의 결과, 위험한 것들의 유혹 1편 - 성과급과 불확실성 확대의 결과, 위험한 것들의 유혹 2편 - 투자 실패 예방
6	2022-12-28 오전 12:00:00	6WpMc8C66EU	"백전불태(百戰不殆)"의 포트폴리오 이론 - 투자를 해야 하는 이유
7	2022-12-28 오전 12:00:00	yxCs2xhfBsU	"백전불태(百戰不殆)"의 포트폴리오 이론 - 뛰어 중단 2편
8	2022-12-28 오전 12:00:00	xD2WagJSU	"백전불태(百戰不殆)"의 포트폴리오 이론 - 뛰어 중단 2편
9	2022-12-28 오전 12:00:00	H8-GjeOVqjQ	"나의 재무 상황은?" - 재무 상황별 재테크 제안   재테크 잘 나에게 필요한 은퇴 재산은? - 재테크의 시작점   은퇴시기
10	2022-12-28 오전 12:00:00	5a4BxjZXVyo	"은퇴자금 얼마나 마련할 수 있을까?" - 장기 재테크 목표 수립
11	2022-12-28 오전 12:00:00	-YbUgnBvz4	"은퇴자금 얼마나 마련할 수 있을까?" - 장기 재테크 목표 수립
12	2022-12-28 오전 12:00:00	24RJpapJ	"복지자금 만들기, 언제쯤? 그리고 성공 가능성은?" - 복지자금 만들기, 언제쯤? 그리고 성공 가능성은?
13	2022-12-28 오전 12:00:00	VtkuLrm4Q	"복지자금 만들기, 언제쯤? 그리고 성공 가능성은?" - 복지자금 만들기, 언제쯤? 그리고 성공 가능성은?
14	2022-12-28 오전 12:00:00	J8EKBDAu83Y	"얼마쯤 모아야 내 집 마련해도 괜찮을까?" - 주택 구매 분석
15	2022-12-28 오전 12:00:00	LVC3vf0cq2w	"얼마쯤 모아야 내 집 마련해도 괜찮을까?" - 주택 구매 분석
16	2022-12-28 오전 12:00:00	Gre_5vPz2zc	"앞으로 기회가 없다고요? 1편 - 주식, 부동산 장기 평균 수익률"

## 2. 파워쿼리를 활용한 데이터 입력

## 엑셀에 SQL문으로 데이터 가져오기



클릭하여, SQL문 작성 화면 확장

SQL문 작성

ODBC(dsn=yt\_stat SELECT a.keyword, COUNT(b.yt\_id) 영상수, S...

keyword	영상수	조회수	노출수	클릭률
저축은행	3	1716	12169	0.141
신용대출	7	1678	12320	0.1362
신용카드	3	1677	12063	0.139
담보대출	5	1666	12111	0.1376
카드론	2	1664	11972	0.139
소득	10	1599	16014	0.0999
마이너스통장	3	1405	9684	0.1451
이자	2	1403	9868	0.1422
재산	2	1399	9573	0.1461
비담보대출	1	1397	9528	0.1466
신용점수	1	1397	9528	0.1466
신용관리	1	1397	9528	0.1466
계좌	1	1397	9528	0.1466
통장	1	1397	9528	0.1466
마통	1	1397	9528	0.1466
연체	1	1397	9528	0.1466
통신요금	1	1397	9528	0.1466
원리금	1	1397	9528	0.1466
현금서비스	1	1397	9528	0.1466
리볼빙	1	1397	9528	0.1466

SQL문에 의해 조회되는 데이터 미리보기

① 크기 제한으로 인해 미리 보기의 데이터가 잘렸습니다.

이후 다른 파워 쿼리 사용법과 동일

로드 데이터 변환 취소

## 6. 데이터 획득과 데이터 전처리

### 2. 파워쿼리를 활용한 데이터 입력

# 템플릿 만들어 놓고, '새로고침' 한번으로 보고서 업데이트

테이블 째 불려온 원본 테이블

번호	제목	조회수	노출수	클릭률	조회율
1	관심지역 전세가율을 머리 속에 저장하자	209	6,998	2.1%	22%
2	부동산 가격 보는 방법 (ft. KB부동산 시세)	173	4,569	2.4%	41%
3	부동산 가격, 어떻게 봐야 하는거야? (ft. 호가, 실거래가, 감정가, 공시지가)	102	873	9.3%	35%
4	집 가격이 비싼지 알아보는 최우선 지표 PIR (소득대비주택가격)	84	2,784	2.1%	33%
5	전국적으로 집 가격이 비싼지 알 수 있는 주택구입부담지수	61	2,343	2.1%	23%
6	첫 집 장만은 수도권, 지방 관계 없이 이렇게 접근하세요 (엑셀 템플릿 제공)	47	921	3.3%	21%
7	금리가 주식이나 부동산에 영향을 주는 원리 (ft. 할인율)	11	307	2.0%	39%
8	왜 부동산은 평당 가격을 따지는거야?	9	454	1.3%	83%
9	경제 흐름 잘 읽는 방법은? (ft. 자기개발, 투자)	8	141	4.3%	32%
10	인구도 감소하는데, 내 집 마련 해도 되나?	5	219	1.8%	73%
11	투자 수익을 위해 위험을 어떻게 떼어야 하는거야?	5	518	0.4%	26%
12	가계자산의 중심, 부동산 - 연령대별 자산/부채 구성 현황	4	58	6.9%	18%
13	자본주의의 은행 시스템 어떻게 이해해야 하는거야? (ft. 실리콘밸리 은행 파산)	2	111	0.0%	1%
14	너무 비싼 집값, 나에게도 살 기회가 있는 걸까? (ft. 소득 대비 주택가격, PII)	2	130	0.0%	2%
15	앞으로 기회가 없다고요? - 글로벌 부동산 가격 추이 리뷰	2	29	6.9%	51%
16	이자율과 수익률 - 돈의 이해 3	1	19	0.0%	0%
17	처음에 투자해야 할 것은? (투자의 순서)	1	78	1.3%	87%
18	앞으로 기회가 없다고요? - 주식, 부동산 장기 평균 수익률 리뷰	1	26	3.8%	42%
19	도대체 종부세가 뭐야? (종합부동산세)	0	30	0.0%	0%
20	가계자산의 중심, 부동산 - 거시적 국가 자산/부채 구성 현황	0	25	0.0%	0%
21	인플레이션을 대비해야 하나, 금리인상을 대비해야 하나? - 합리적 대응 방	0	47	0.0%	0%
22	도대체 공시지가가 뭐야?	0	25	0.0%	0%
23	재테크 보충수업 Q&A 2부	0	22	0.0%	0%

SQL로 불려온 원본 테이블

번호	제목	조회수	노출수	클릭률	조회율
1	관심지역 전세가율을 머리 속에 저장하자	209	6,998	2.1%	22%
2	부동산 가격 보는 방법 (ft. KB부동산 시세)	173	4,569	2.4%	41%
3	부동산 가격, 어떻게 봐야 하는거야? (ft. 호가, 실거래가, 감정가, 공시지가)	102	873	9.3%	35%
4	집 가격이 비싼지 알아보는 최우선 지표 PIR (소득대비주택가격)	84	2,784	2.1%	33%
5	전국적으로 집 가격이 비싼지 알 수 있는 주택구입부담지수	61	2,343	2.1%	23%
6	첫 집 장만은 수도권, 지방 관계 없이 이렇게 접근하세요 (엑셀 템플릿 제공)	47	921	3.3%	21%
7	금리가 주식이나 부동산에 영향을 주는 원리 (ft. 할인율)	11	307	2.0%	39%
8	왜 부동산은 평당 가격을 따지는거야?	9	454	1.3%	83%
9	경제 흐름 잘 읽는 방법은? (ft. 자기개발, 투자)	8	141	4.3%	32%
10	인구도 감소하는데, 내 집 마련 해도 되나?	5	219	1.8%	73%
11	투자 수익을 위해 위험을 어떻게 떼어야 하는거야?	5	518	0.4%	26%
12	가계자산의 중심, 부동산 - 연령대별 자산/부채 구성 현황	4	58	6.9%	18%
13	자본주의의 은행 시스템 어떻게 이해해야 하는거야? (ft. 실리콘밸리 은행 파산)	2	111	0.0%	1%
14	너무 비싼 집값, 나에게도 살 기회가 있는 걸까? (ft. 소득 대비 주택가격, PII)	2	130	0.0%	2%
15	앞으로 기회가 없다고요? - 글로벌 부동산 가격 추이 리뷰	2	29	6.9%	51%
16	이자율과 수익률 - 돈의 이해 3	1	19	0.0%	0%
17	처음에 투자해야 할 것은? (투자의 순서)	1	78	1.3%	87%
18	앞으로 기회가 없다고요? - 주식, 부동산 장기 평균 수익률 리뷰	1	26	3.8%	42%
19	도대체 종부세가 뭐야? (종합부동산세)	0	30	0.0%	0%
20	가계자산의 중심, 부동산 - 거시적 국가 자산/부채 구성 현황	0	25	0.0%	0%
21	인플레이션을 대비해야 하나, 금리인상을 대비해야 하나? - 합리적 대응 방	0	47	0.0%	0%
22	도대체 공시지가가 뭐야?	0	25	0.0%	0%
23	재테크 보충수업 Q&A 2부	0	22	0.0%	0%

## 3. 금융 데이터 전처리 방법론

# 이 데이터는 어떻게 처리해야 할까?



### 3. 금융 데이터 전처리 방법론

## 데이터 전처리는 데이터 탐색과 손질의 과정

인사이트 정도에 따라 데이터 정제의 질이 달라진다

### 원본 데이터

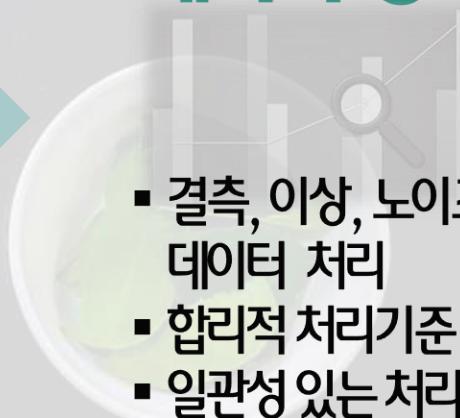


### 데이터 변환



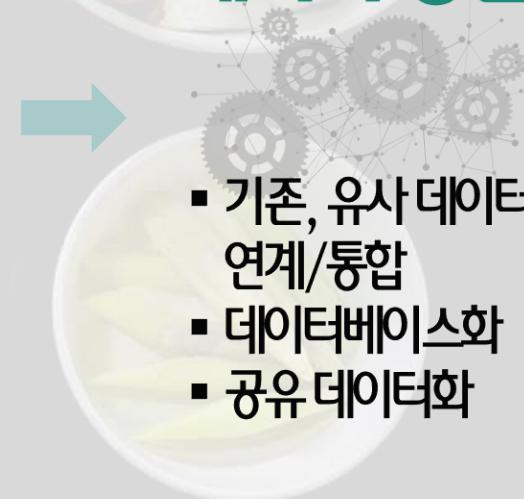
- 통일된 형식으로
- 비정형 데이터를 구조적 데이터로
- 기존 데이터와 통합

### 데이터 정제



- 결측, 이상, 노이즈 데이터 처리
- 합리적 처리기준 수립
- 일관성 있는 처리

### 데이터 통합



- 기존, 유사 데이터 연계/통합
- 데이터베이스화
- 공유 데이터화

시각화와 기술통계량을 통한 발견



### 3. 금융 데이터 전처리 방법론

## 결측값을 채우거나 이상값을 수정하여 데이터 신뢰도를 높이는 작업

**결측값  
(Missing Value)**

**노이즈  
(Noise)**

**이상값  
(Outlier)**

### 레코드(행)에서 특정 필드(열) 값이 없는 경우

- 판단 : 해당 레코드를 제거할 것인가, 빈 값을 어떤 규칙으로 수정 입력할 것인가?
- 중심 경향값으로 넣기 : 평균(Mean), 중앙값(Median), 최빈값(Mode)
- 분포 기반 값으로 넣기 : 랜덤에서 자주 나타나는 값 (일종의 최빈값)

### 값은 있으나, 발생했다고 보기 어려운 값이 있는 경우

- 판단 : 데이터가 실제와 다르게 잘못 입력되는 경우가 발생한 것인가?
- 이동평균 값으로 대체 : 이어지는 수에서 튀는 경우 (시계열)
- 일정 범위 중간값으로 대체

### 데이터 값이 일반적인 값보다 편차가 큰 값

- 판단 : 특별히 범위가 상당히 벗어나는 값이 분포의 상단/하단에 비해 의미가 있는가?
- 하한값 / 상한값으로 대체

### 3. 금융 데이터 전처리 방법론

## 원 값의 척도를 의미 있는 척도로 변환하거나 분포를 변환하는 작업

### 변화율 적용

#### 원 값의 구간별 변화율 비교가 유용할 때

- 시계열 데이터 분석에서 일반적인 방법, 계절 효과를 제거해야 할 때
- 새로운 데이터 = 기말값 / 기초값 - 1 (YoY, QoQ, MoM, DoD)
- 예 : 물가상승률(인플레이션), 주가 수익률, 매출액 증가율

### 로그 변환

#### 분포가 지수함수, 멱급수 관계에 있을 때

- 세상의 많은 분포는 파레토 법칙을 따른다!
- 새로운 데이터 =  $\ln(\text{기존 데이터})$
- 예 : 상장주식 시가총액, 제품별 매출액, 가계 순자산, 시계열 수익률 분석

### 정규화 (Normalization)

#### 서로 다른 척도 값을 동일 척도로 하여 비교 또는 결합할 때

- Z-score : 순서를 세우고, 변환한 값들의 분포가 정규분포에 근사할 때
- Percentile Rank : 분포의 특성이 관찰되지 않을 때
- 예: 주가이익률(PER)의 역수 (Earnings Yield)

## 핵심요약

- 데이터는 다양한 원천에서 구할 수 있는데, 파일, 오픈API, 웹페이지, 데이터베이스 등 대부분의 원천 데이터를 엑셀에 입력할 수 있음
- 데이터의 원천에 관계 없이, 최종 입력 형태는 테이블 형태로 동일하며, 엑셀이 제공하는 이 데이터 처리 구조를 ‘파워쿼리’라 함
- 전처리에는 데이터 변환, 데이터 정제, 데이터 통합 등이 있으며, 데이터 분석 전체 과정의 절반 이상의 노력과 시간이 요구됨
- 결측값, 노이즈, 이상값에 대한 처리와, 데이터의 척도 변환은 데이터가 기록된 실무를 이해하는 사람이 제대로 처리 가능

## 7. 시각화와 기술통계량을 활용한 사전 데이터분석

---

# 학습목표

- ▶ 엑셀 차트를 활용하여, 데이터 셋의 패턴을 빠르게 찾아내는 방법 이해
- ▶ 기술 통계량을 활용하여, 데이터 셋의 분포 패턴을 빠르게 읽어내는 방법 이해

## 1. 시각화를 활용한 사전 데이터 분석

# 데이터 형태와 목적에 따른 시각화 종류

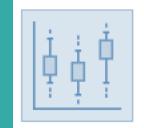
## 데이터의 분포(기술통계)

### 단일 필드 특성값 분포



- 데이터 순서 관계 없이 계급간 도수 생성
- 기술통계량의 시각적 확인

## 상자수염



- 기술통계량을 단일 차트에 표현
- 복수의 특성값 계열 분포 동시에 확인

## 연속된 필드 값의 경향

### (연속계열, 특성값) 추이

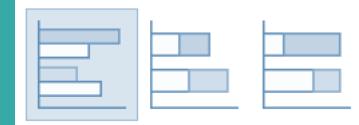
## 꺾은선형



## 세로 막대형



## 가로 막대형



- 연속계열(x축)에 따른 특성값(y축) 관찰
- 특성값 계열이 많을 경우 꺾은선형

## 영역형



- 꺾은선형의 아래 영역이 채워진 것
- 계열별 값을 누적할 경우 유용

## 데이터(쌍)의 관계

### (특성값1, 특성값2) 관계

## 분산형



- 데이터 순서에 관계 없이 관계를 볼 때
- 상관관계 분석(회귀분석)을 시각화

## 혼합



- 다른 계열을 명확히 구분하고 싶을 때
- 연속적이면서, 관계를 동시에 볼 때

## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)원형차트

# 차트에 데이터를 맵핑(입력)하는 법

## (1) 그리려 하는 데이터 범위 선택

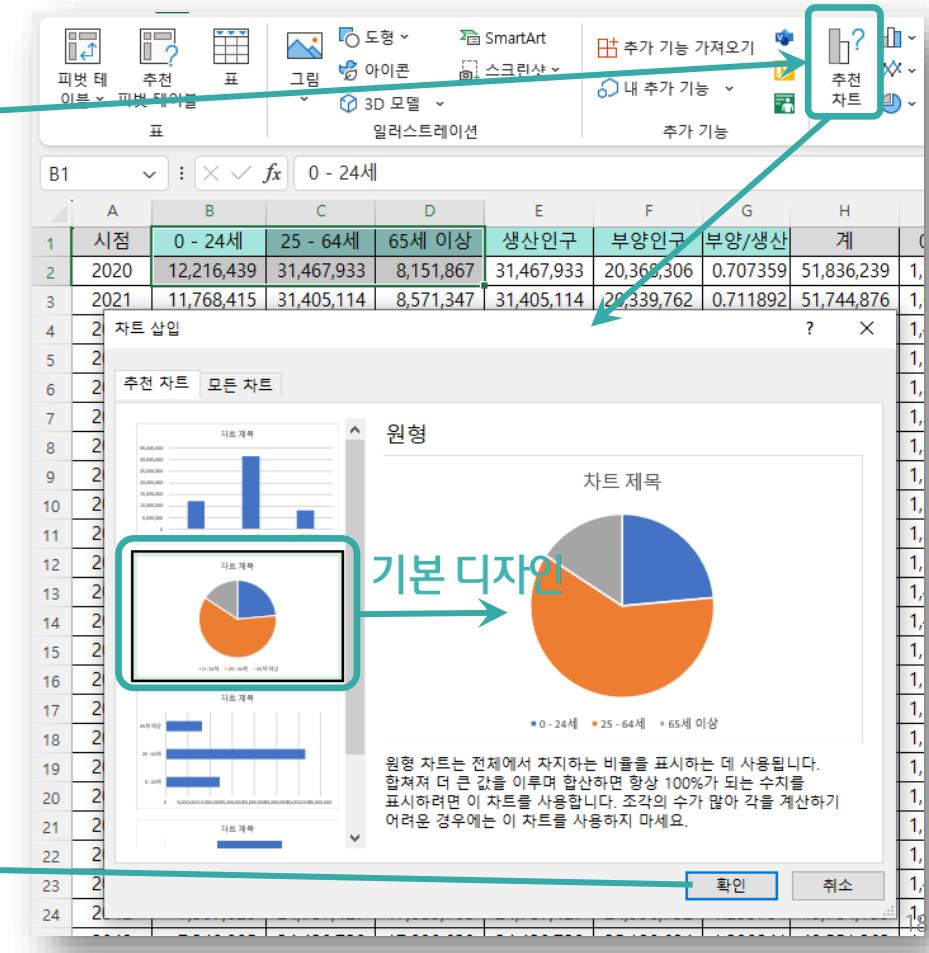
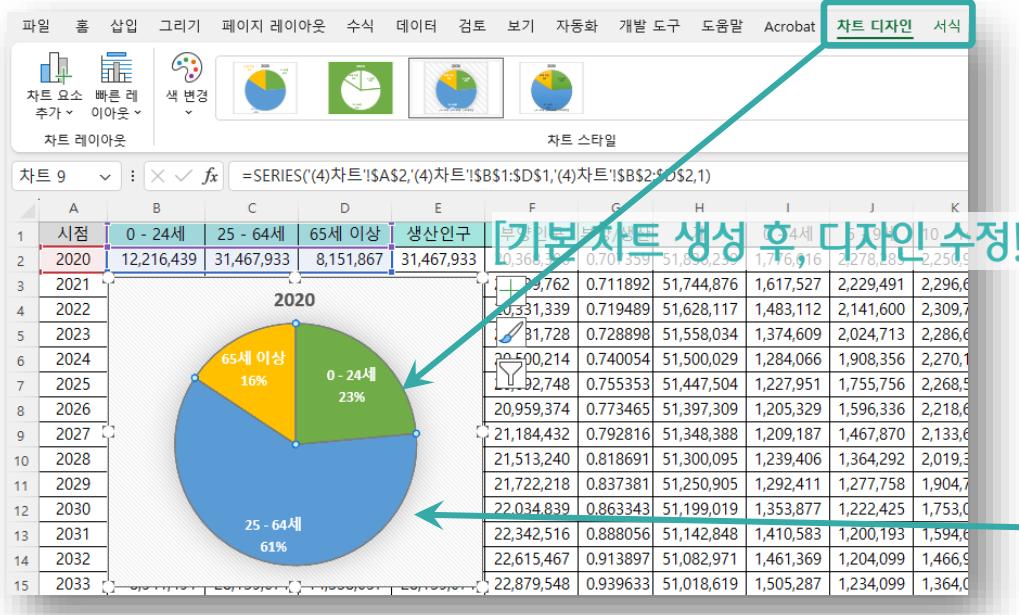
최초 선택요소는 '데이터 계열'과 '차트 종류'

	B1	C1	D1	E1	F1	G1	H1	I1	J1	K1
시점	0 - 24세	25 - 64세	65세 이상	생산인구	부양인구	부양/생산	계	0 - 24세	25 - 64세	65세 이상
2020	12,216,439	31,467,933	8,151,867	31,467,933	20,360,306	0.707359	51,836,239	12,216,439	31,467,933	8,151,867
2021	11,768,415	31,405,114	8,571,347	31,405,114	20,339,762	0.711892	51,744,876	11,768,415	31,405,114	8,571,347
2022	11,212,227	31,326,770	9,012,412	31,326,770	20,309,135	0.716335	51,641,092	11,212,227	31,326,770	9,012,412

선택 범위 : \$B\$1:\$D\$2

선택 후 클릭!

## (3) [차트디자인]과 [서식]에서 디자인 설정



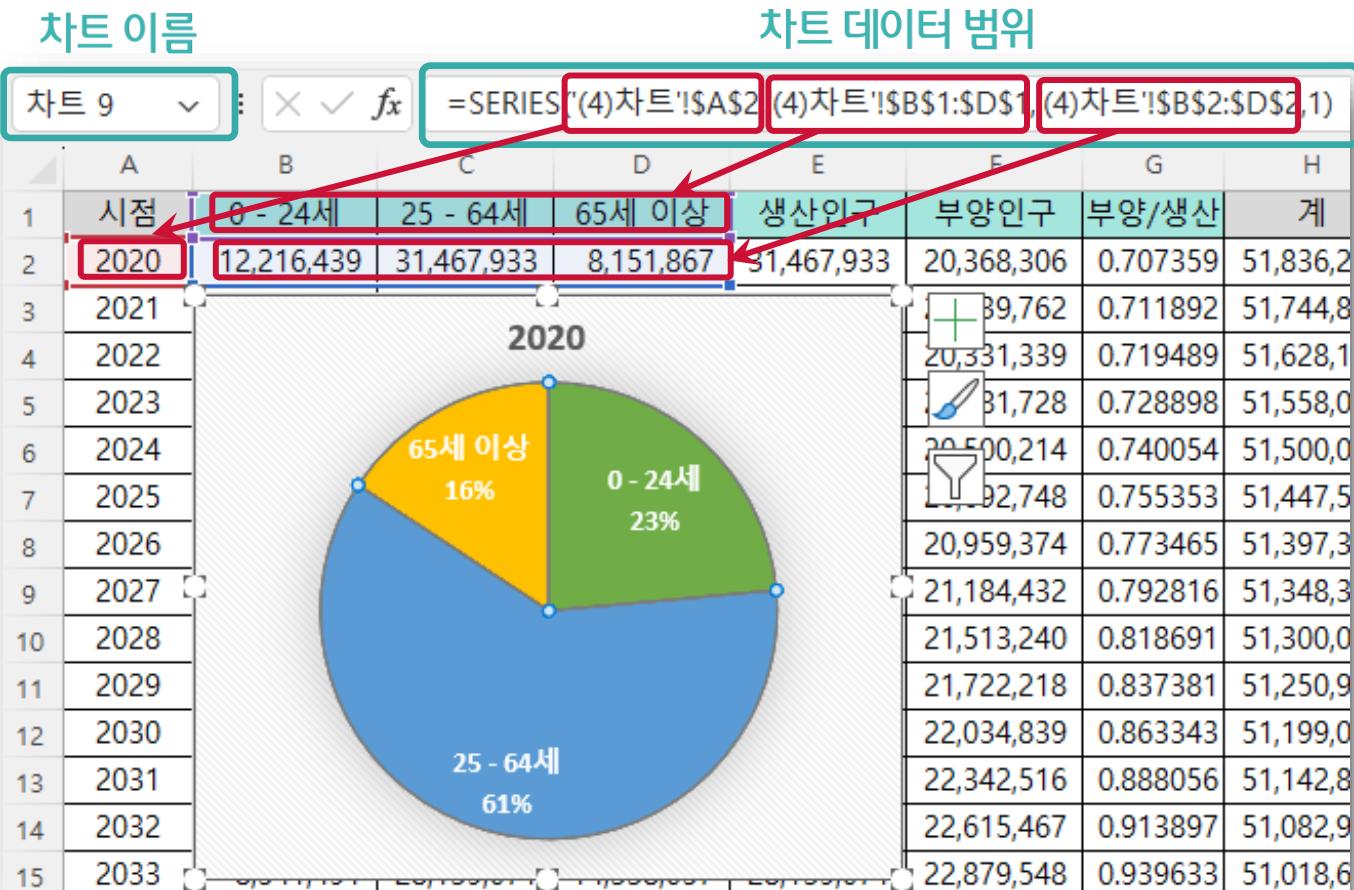
## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)원형차트

# 차트 선택을 통한 범위의 확인과 수정

함수식 형태!

=SERIES(범례 위치, 값제목 or 기준값(x) 데이터 범위, 값(y) 데이터 범위, 값(y) 데이터 순서)



## 수정 방법 3가지

- 선택된 범위를 마우스로 끌어서 변경하기
- [범위] 식을 직접 수정하기
- [데이터 원본 선택] 대화 상자 활용하기

## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)원형차트

## 데이터 원본 다루기

=SERIES(범례 위치, 값제목 or 기준값(x) 데이터 범위, 값(y) 데이터 범위, 값(y) 데이터 순서)

[데이터 선택]

차트 9 :  $=\text{SERIES}((4)\text{차트}'!$A$2,(4)\text{차트}'!$B$1:$D$1,(4)\text{차트}'!$B$2:$D$2)$

A	B	C	D	E	F	G	H
시점	0 - 24세	25 - 64세	65세 이상	생산인구	부양인구	부양/생산	계
2020	12,216,439	31,467,933	8,151,867	31,467,933	20,368,306	0.707359	51,836,2
2021	12,216,439	31,467,933	8,151,867	31,467,933	20,368,306	0.711002	51,774,1

Acrobat | 차트 디자인

행/열 전환 데이터 선택 차트 종류 변경 데이터 종류

축 레이블

축 레이블 범위(A):  $=\text{(4)차트}'!$B$1:$D$1$

확인 취소

데이터 원본 선택

차트 데이터 범위(D):  $=\text{(4)차트}'!$A$1:$D$2$

행/열 전환(W)

범례 항목(계열)(S)

추가(A) 편집(E) 제거(R)

2020

가로(항목) 축 레이블(C)

편집(I)

0 - 24세  
25 - 64세  
65세 이상

y 값 계열에 대한 편집

대응 x값 계열에 대한 편집

계열 편집

계열 이름(N):  $=\text{(4)차트}'!$A$2$

계열 값(V):  $=\text{(4)차트}'!$B$2:$D$2$

확인 취소

숨겨진 셀/빈 셀(H)

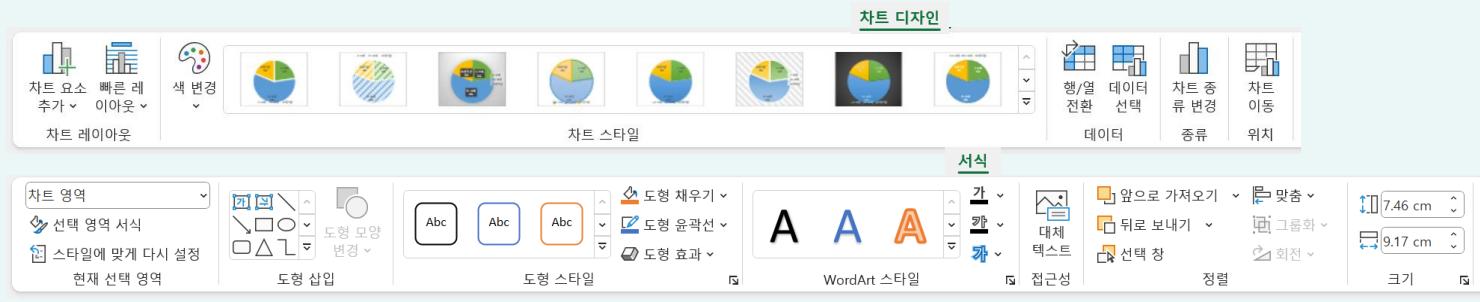
확인 취소

## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)원형차트

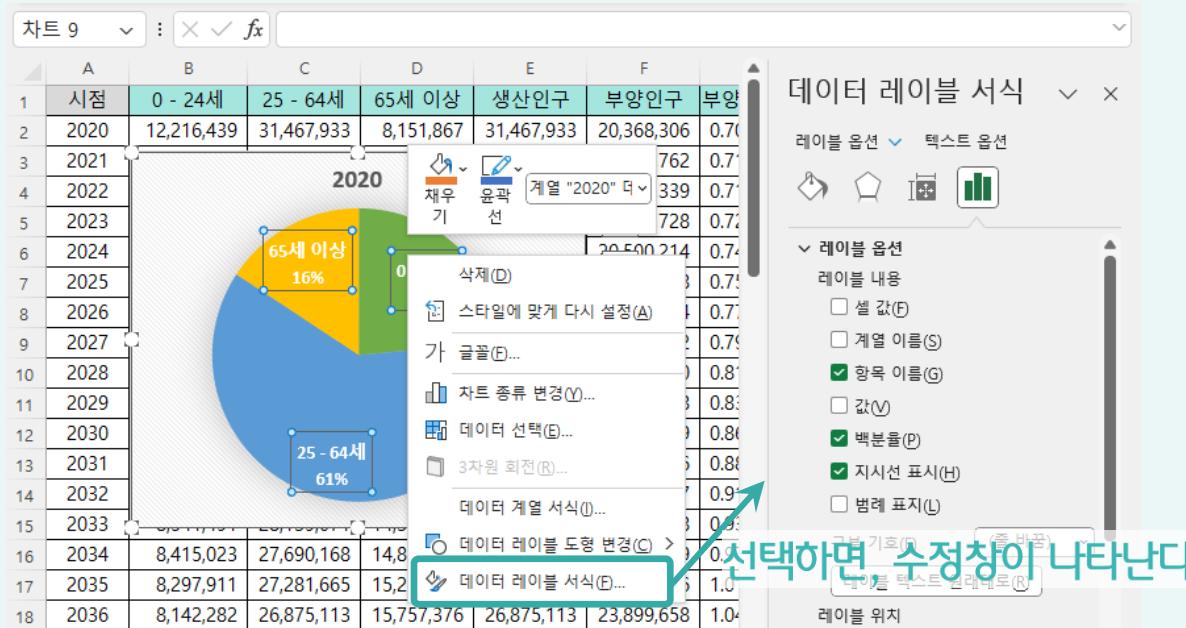
# 차트 디자인 꿀팁

리본 메뉴  
사용을 통한 설정



팝업 메뉴  
사용을 통한 설정

수정하고 싶은 부분 선택 후,  
오른쪽 마우스 클릭!

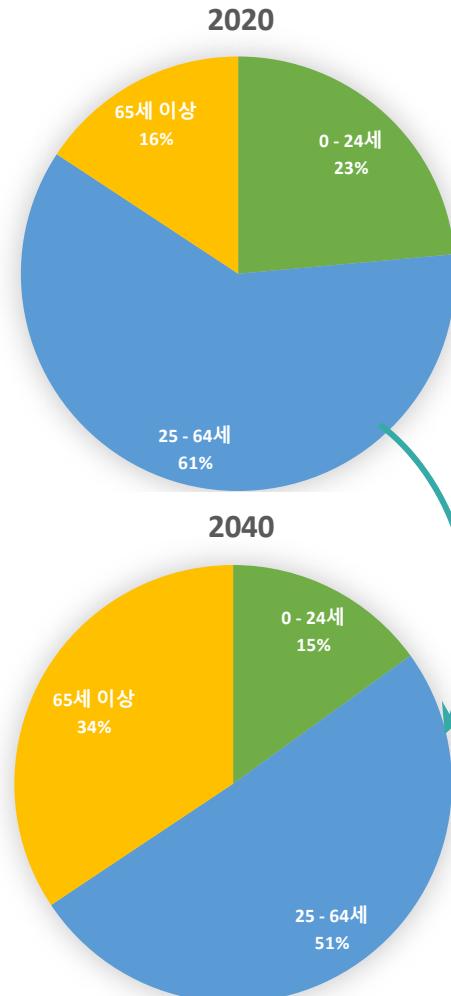
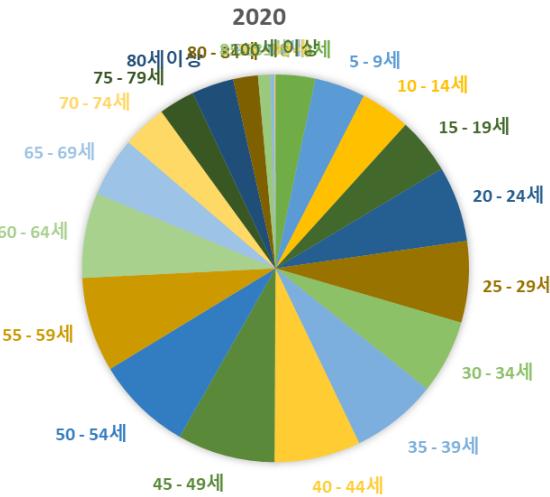


## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)인구추계

# 원형(파이) 차트가 인포그래픽용인 이유

무엇을 알 수 있는가?



생산 인구 비중 축소  
= 부양 인구 비중 확대

값들이 많아, 값들의 비교가 되지 않는다!

## 원형 차트 활용

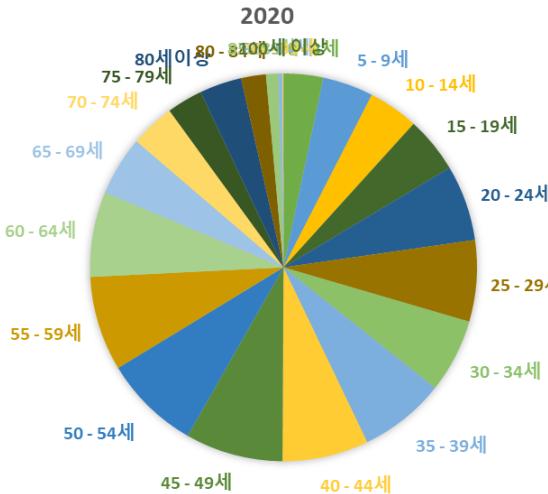
- 계열 값의 개수가 많지 않을 때
- 계열 값 중, 압도적으로 큰 값이 있을 때

## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)인구추계

# 막대 차트를 활용한 데이터 시각화 분석

무엇을 알 수 있는가?



값들이 많아, 값들의 비교가 되지 않는다!

## 막대 차트 활용

- 계열 값의 개수가 20개 이내
- 계열 내의 값들의 비교 또는 추세 확인

두 차트의 데이터  
매핑은 동일!

## 세로 막대형 차트



## 가로 막대형 차트

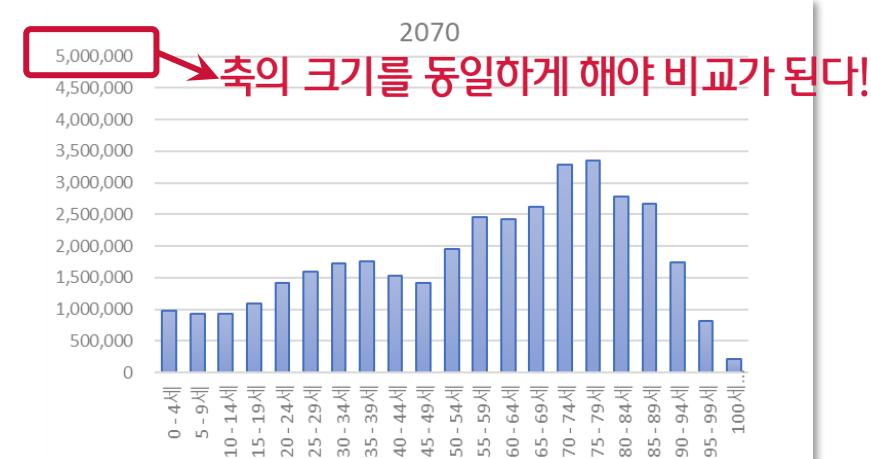
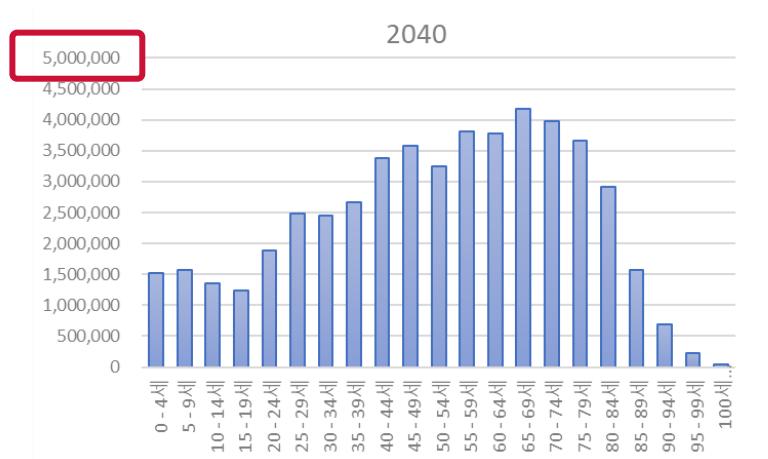
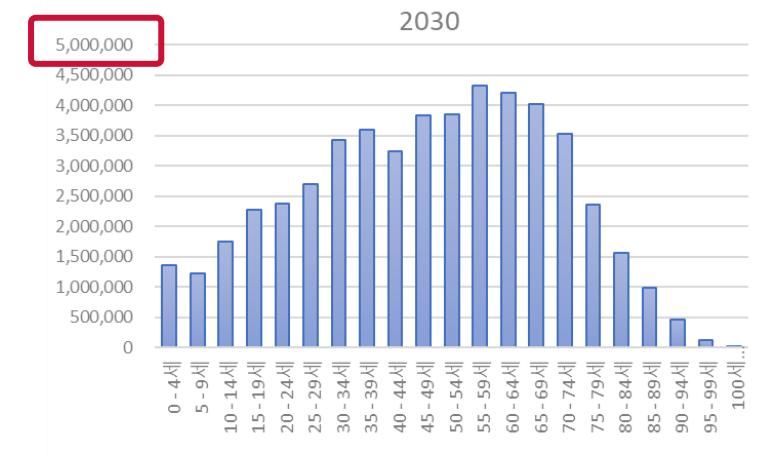
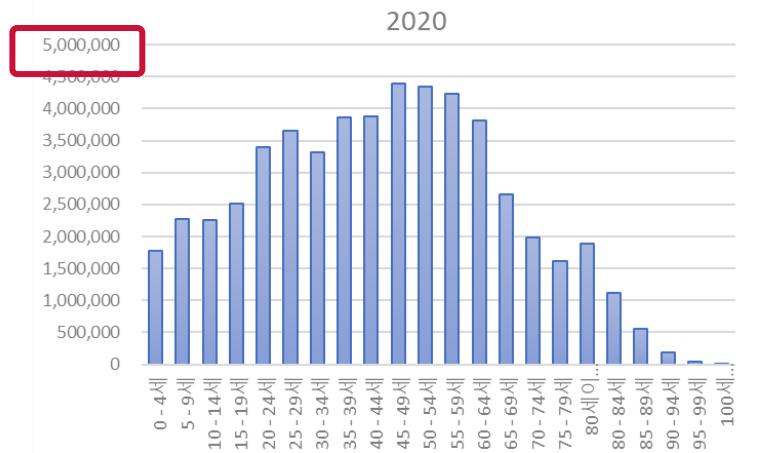


## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)인구추계

# 서로 다른 계열의 데이터 비교 시각화

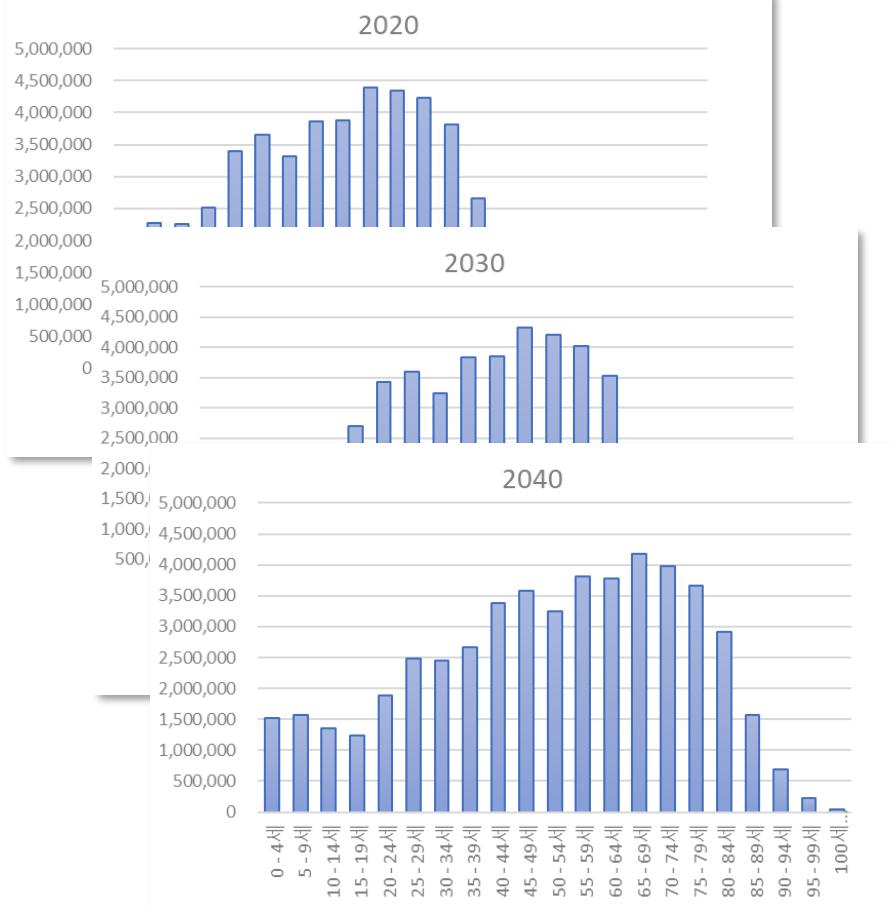
## 매 시기마다의 막대 차트를 그릴 것인가?



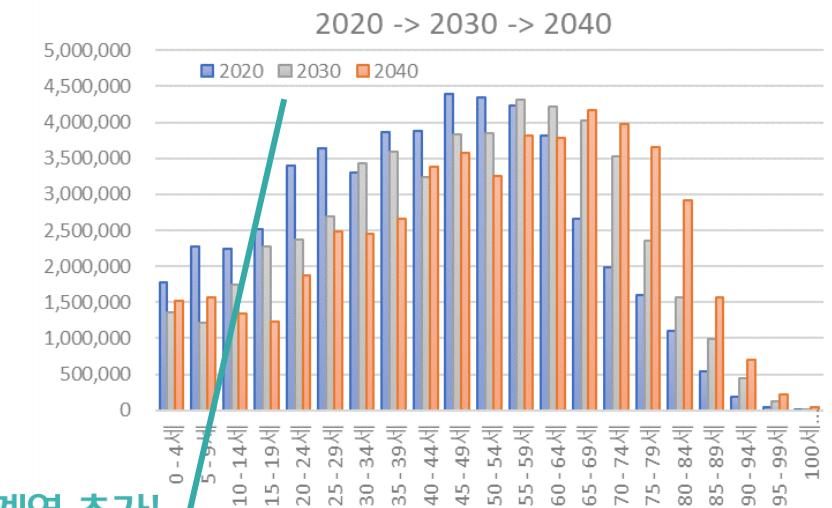
## 1. 시각화를 활용한 사전 데이터 분석

## [예제] (6)인구추계

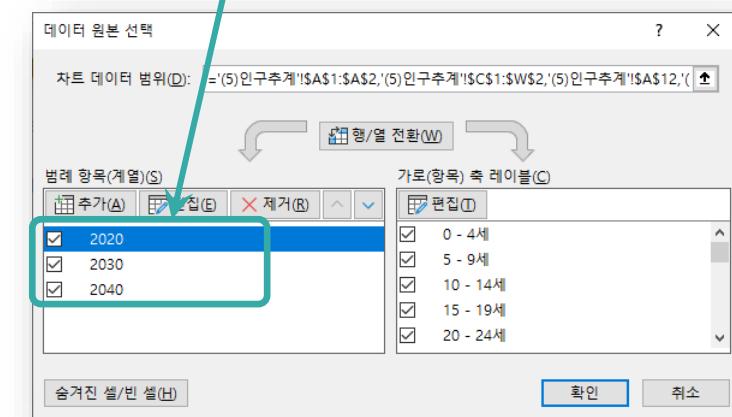
# 매 시기마다의 막대 차트를 그릴 것인가?



## 하나의 차트에 같이 그리기



계열 추가!



## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)인구추계

## 계열의 추세를 비교할 때는 선 차트!

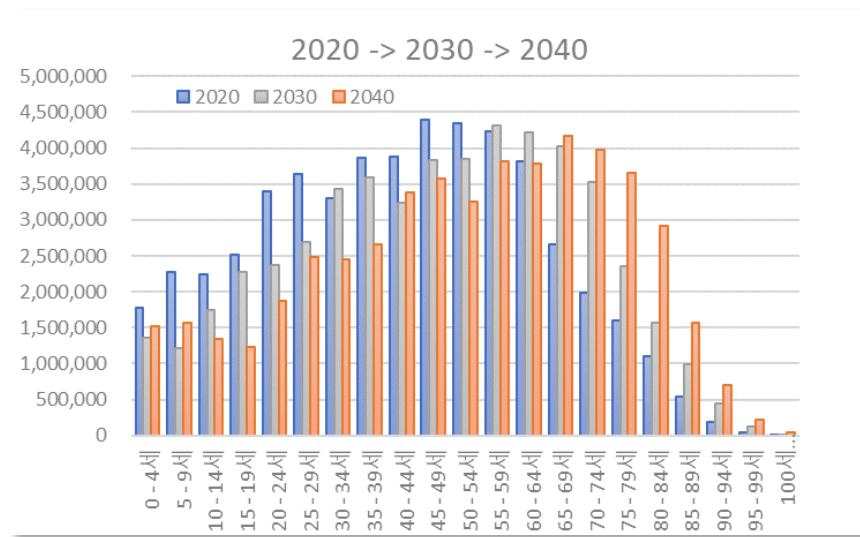
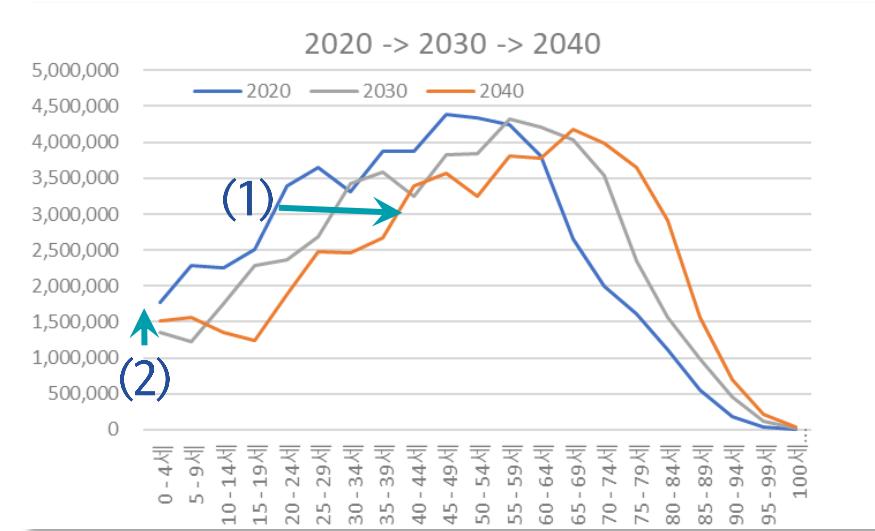


차트 종류만  
변경!  
→



추세의 변화를 포착하기에는 혼잡성이 큼!

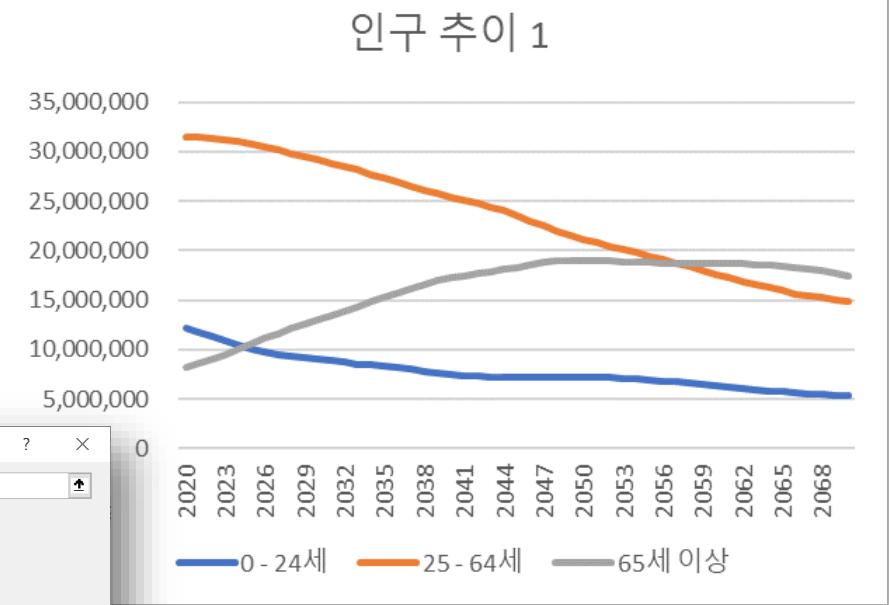
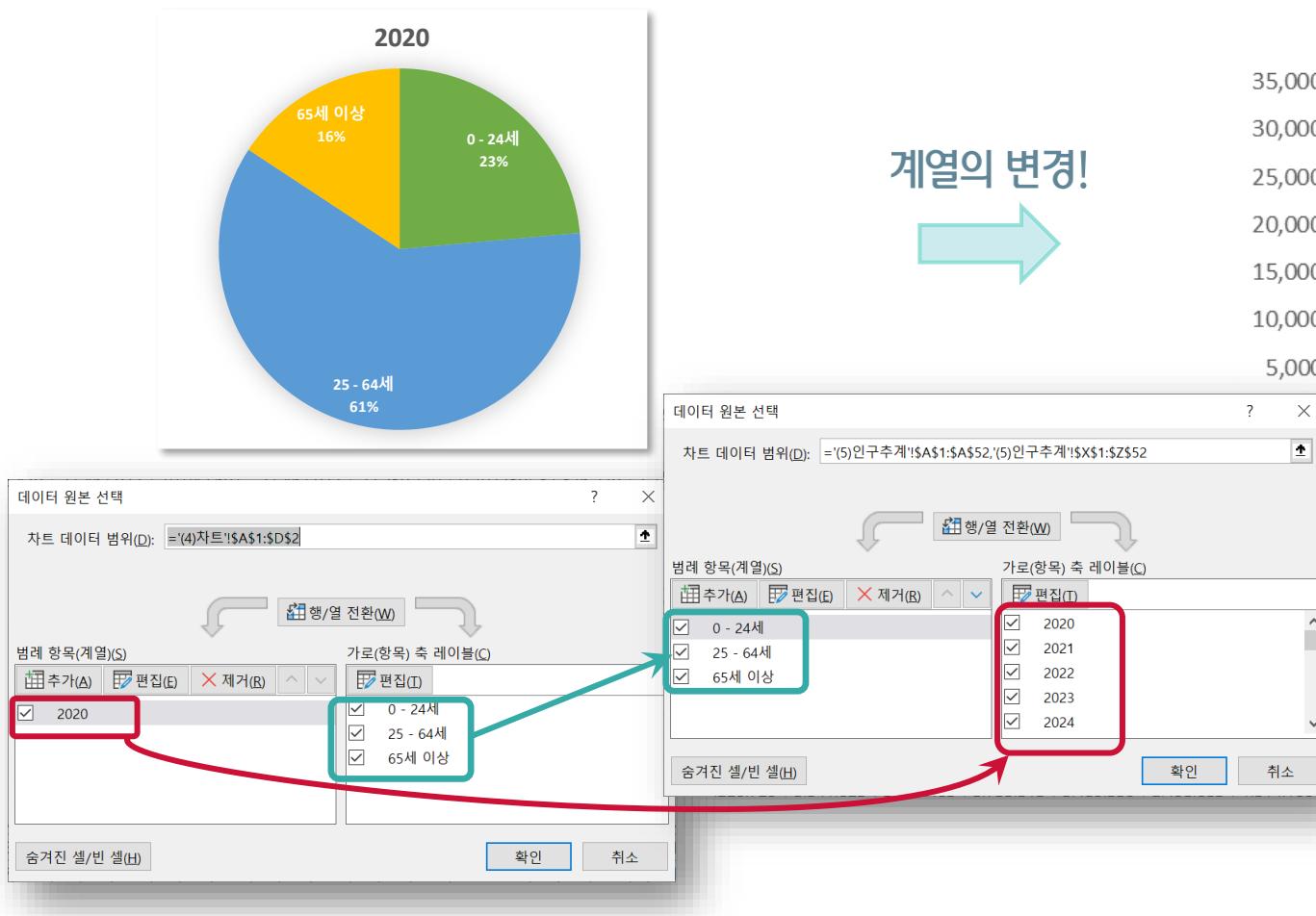
- 1) 현재 인구의 우측 평행 이동 경향을 알 수 있고,
- 2) 20년 후에는 신생아 출산이 증가하는 것으로 예측했음이 관찰됨

## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)인구추계

# 시계열에 따른 추이 비교에는 선 차트

연도별로 생산 인구 추이를 보려면?



- 패턴 변화의 기준이 되는 값을 가로(레이블; x) 축에,
- 보고자 하는 값을 세로(범례; y) 축에 위치

## 1. 시각화를 활용한 사전 데이터 분석

[예제] (6)인구추계

# 합계를 같이 파악할 수 있는 영역 차트

전체 인구 추이를 같이 보려면?

인구 추이 1

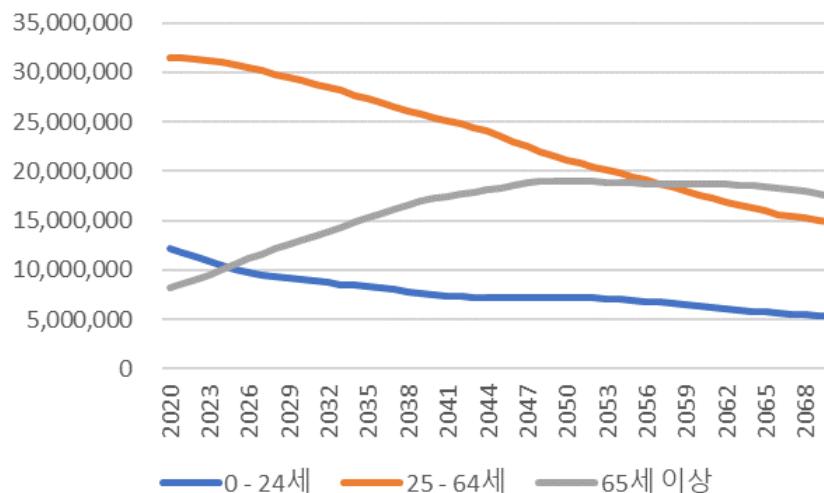
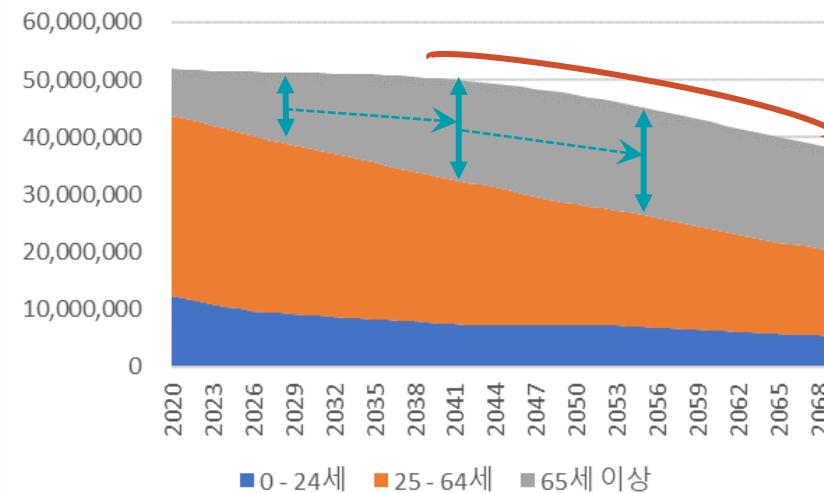


차트 종류만  
변경!

누적 영역 차트

인구 추이 2



계열별 인구는 알아도, 전체 인구는 알 수 없다!

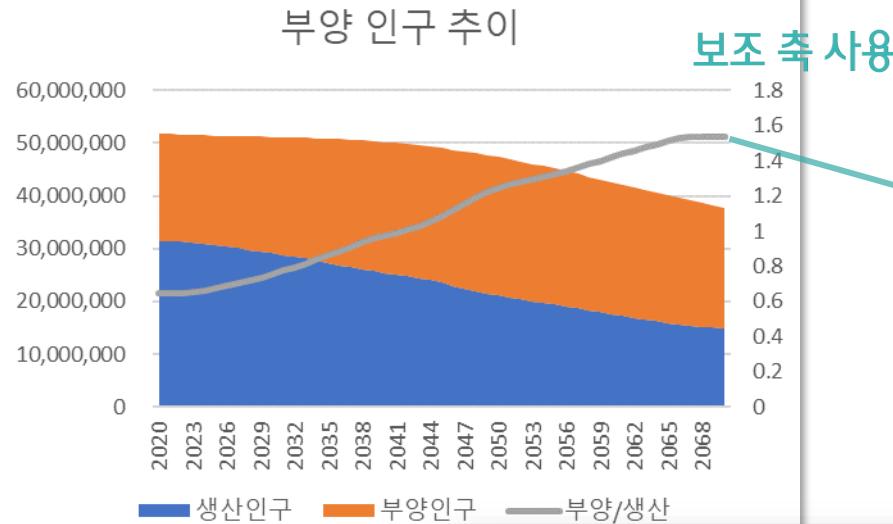
- 계열별로는 두께의 변화
- 가장 상단의 경계선은 전체 인구 합계
- 총인구의 감소 속도는 2040년대 이후 빨라지는 모습

## 1. 시각화를 활용한 사전 데이터 분석

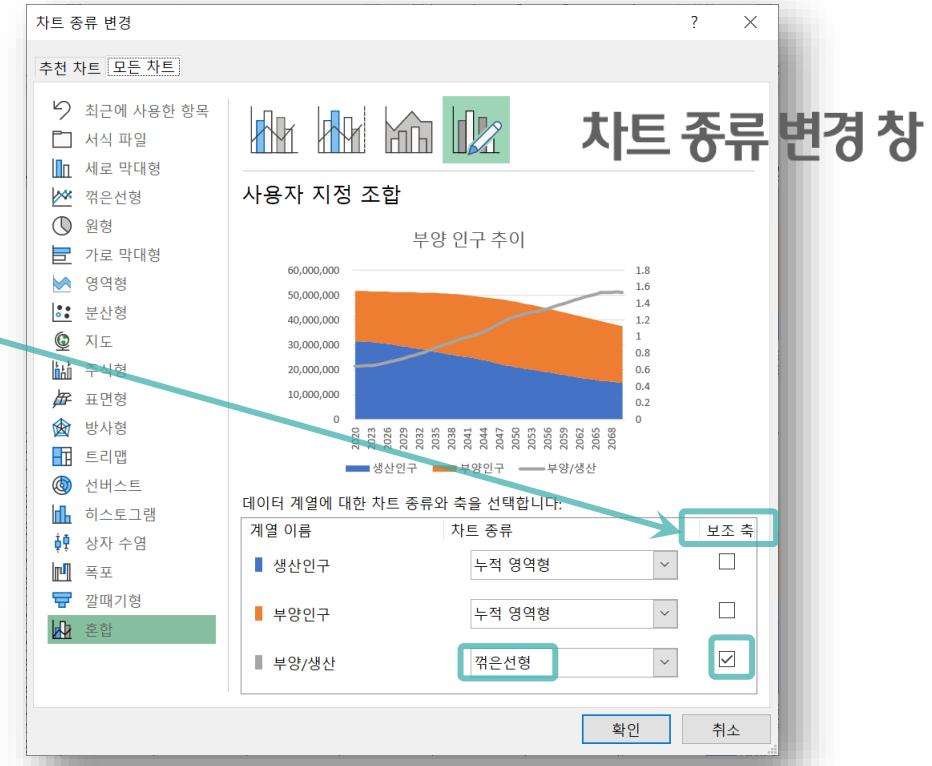
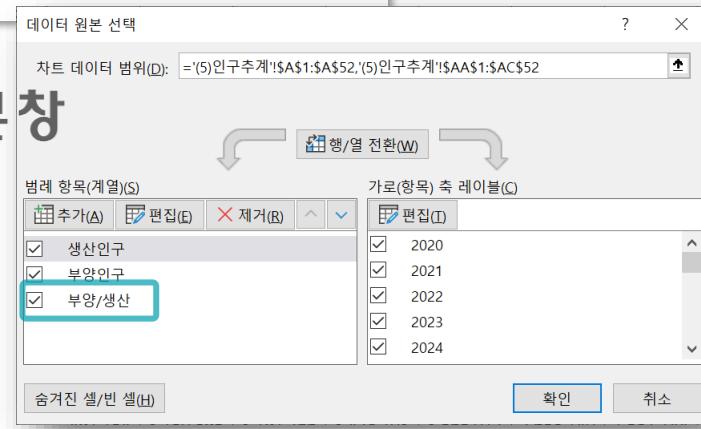
[예제] (6)인구추계

# 성격(값 범위)이 다른 계열을 함께 보는 혼합 차트

인구 뿐 아니라 비율 추이를 같이 보려면?



데이터 원본 창



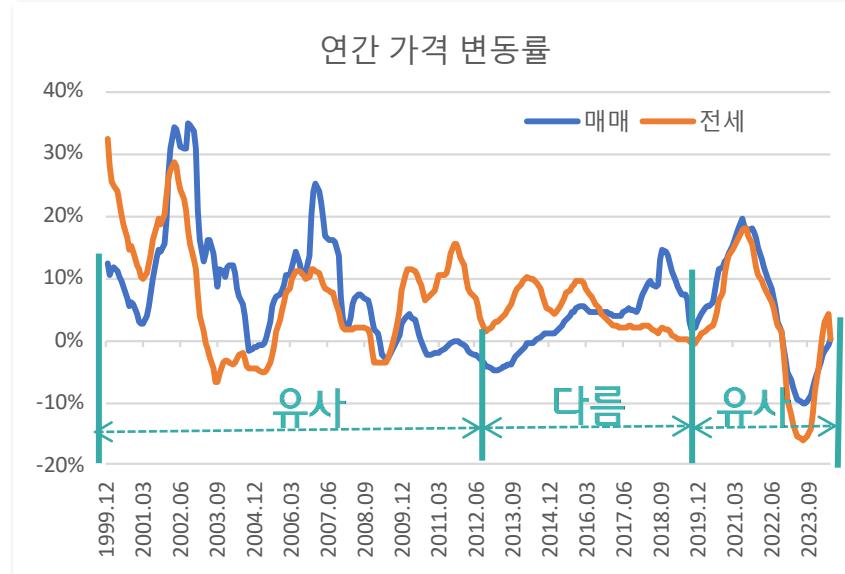
- 계열의 X값은 같이 사용
- 서로 다른 값(범위)의 특성을 갖는 계열
- 해석: 생산인구 대비 부양인구는 꾸준하게 상승

## 1. 시각화를 활용한 사전 데이터 분석

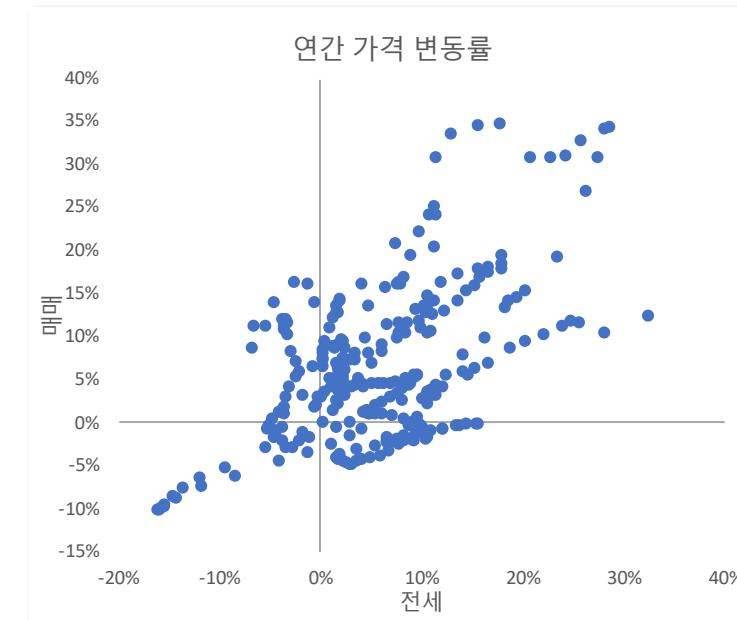
[예제] (5)아파트가격2

# 순서쌍의 관계를 파악하는 분산 차트

**매매가와 전세가는 같은 방향성을 갖는가?**



날짜가 없어짐  
→



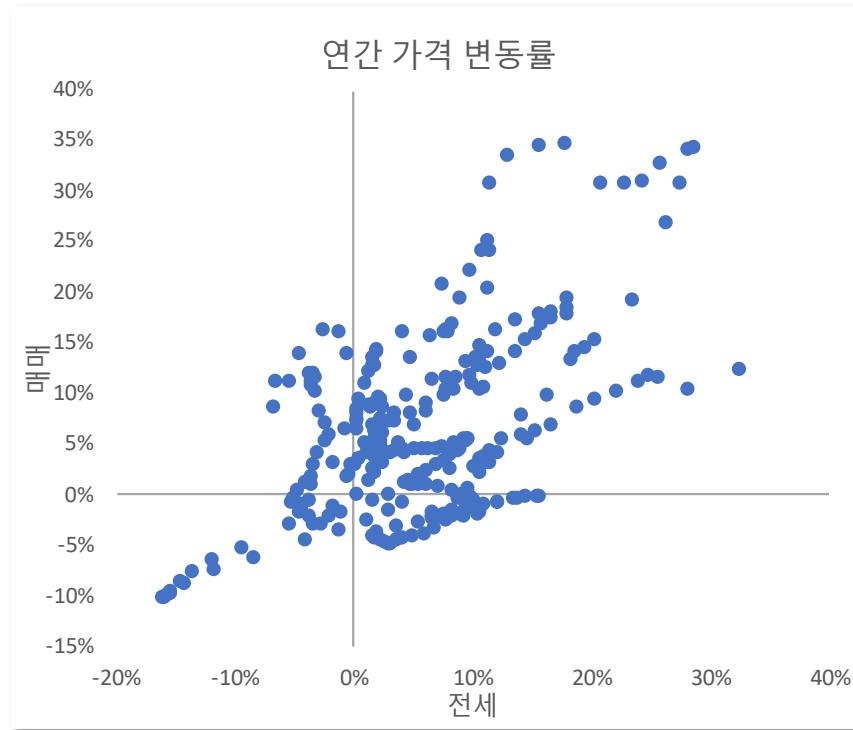
- 그런 것 같긴 한데, 어느 정도인지 말하기 어려움
- 눈에 띄게 다른 구간이 있다면, 해당 기간에 영향을 준 다른 관계를 찾아야 함

- 시간에 관계없이 두 값의 관계성 파악
- 흩어져 있지만, 우상향하는 관계성 확인
- 관계성에 대한 추가 통계 분석 가능

## 1. 시각화를 활용한 사전 데이터 분석

[예제] (5)아파트가격2

## 전세가의 변화는 매매가 변화를 설명할까?



- X값 계열 : 독립 변수 (영향을 준다고 가설을 세운 값들)
- Y값 계열 : 종속 변수 (영향을 받는다고 가설을 세운 값들)

**데이터 원본 창**

데이터 원본 선택

차트 데이터 범위(D):

데이터 범위가 너무 복잡해서 표시할 수 없습니다. 새 범위를 선택하면 계열 패널의 모든 계열이 바뀝니다.

범례 항목(계열)(S)

가로(항목) 측 레이블(C)

추가(A) 편집(E) 제거(R) 편집(I)

서울특별시

필요 없어 비활성

숨겨진 셀/빈 셀(H)

확인 취소

**범례 계열에서 (x, y) 모두 입력**

계열 편집

계열 이름(N): `='(5)아파트가격2'!$B$12:$C$12` = 서울특별시

계열 X 값(X): `='(5)아파트가격2'!$C$14:$C$30` = 32.5%, 28.1%, ...

계열 Y 값(Y): `='(5)아파트가격2'!$B$14:$B$30` = 12.5%, 10.4%, ...

확인 취소

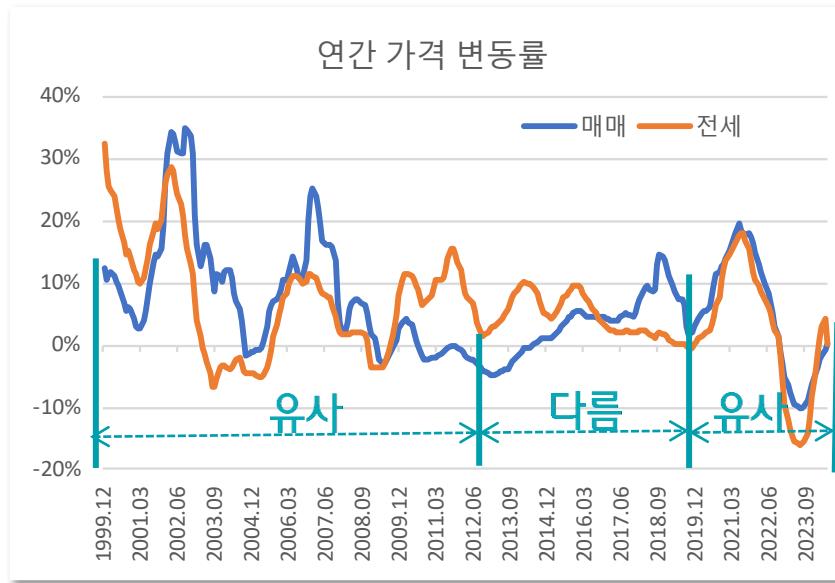
**전세가 변동**

**매매가 변동**

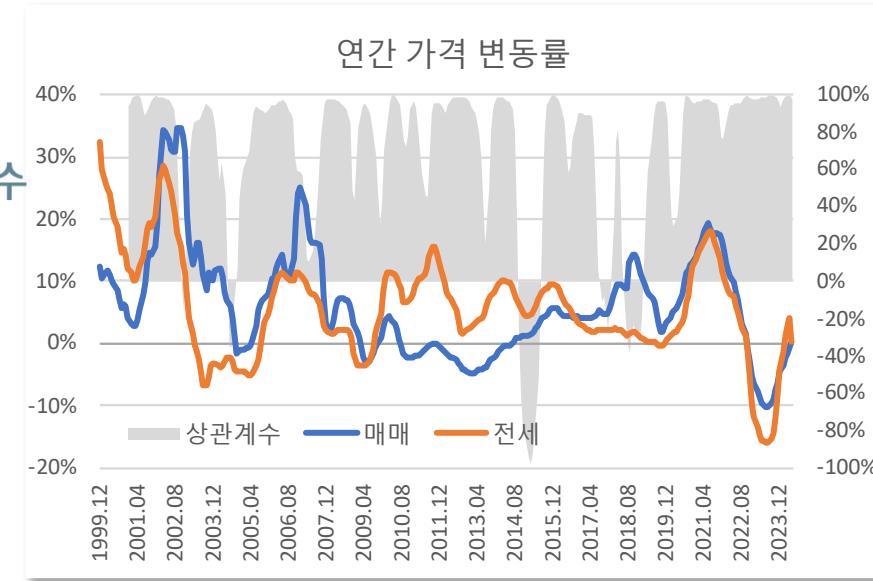
## 1. 시각화를 활용한 사전 데이터 분석

[예제] (5)아파트가격2

# 상관계수 함수를 이용한 2차 데이터 가공 (구간 관찰)



이동 상관계수



- 눈에 띄게 다른 구간이 있다면, 해당 기간에 영향을 준 다른 관계를 찾아야 함

- 12개월 상관계수를 이동하며 산출
- 혼합 차트 이용하여 같이 시각화
- 발견: 2014년 말 ~ 2015년 초의 특수성!  
(매매가 상승 전환의 신호???)

# 입력된 데이터 분포 특성 파악하기

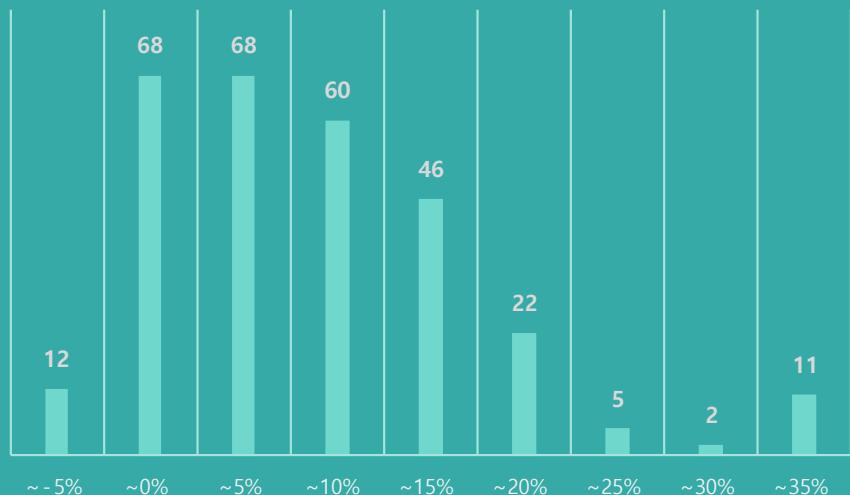
표 / 그래프를 이용

기술통계량 이용

도수분포표

범위	~-5%	~0%	~5%	~10%	~15%	~20%	~25%	~30%	~35%
빈도수	12	68	69	60	46	22	5	2	11

히스토그램

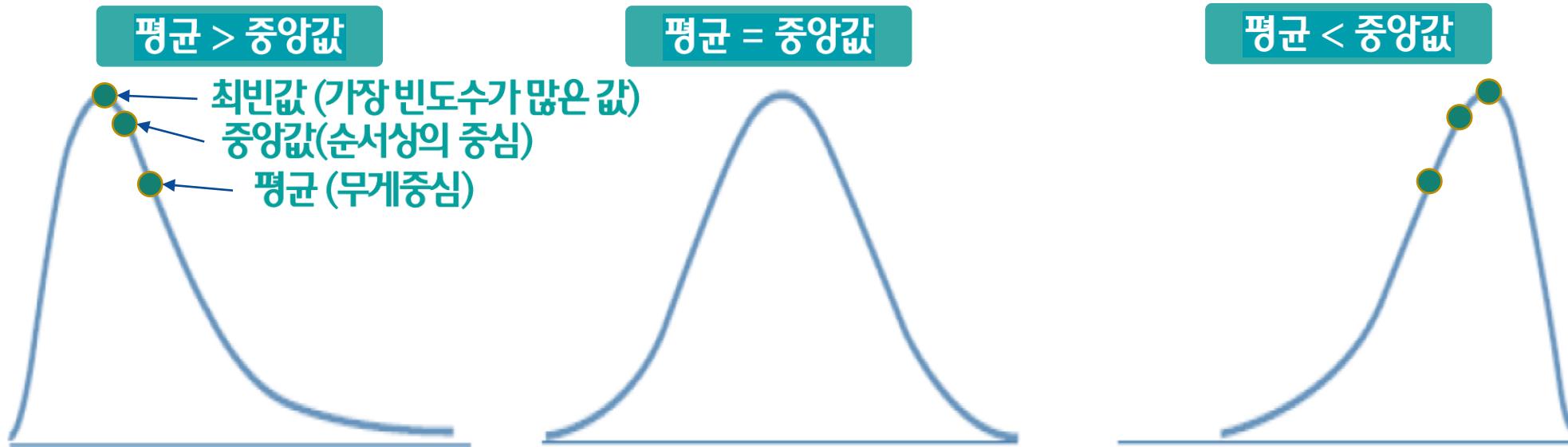


기술통계량

종류	통계량	값
대표치	평균 (mean)	6.3%
	중앙값 (median)	4.7%
	최빈수 (mode)	#N/A
분산도	표준편차 (std. dev)	8.8%
	사분위수 (percentile)	(-0.3%, 4.7%, 11.2%)
	최소값 (minimum)	-10.2%
분포도	최대값 (maximum)	34.8%
	왜도 (skewness)	0.99
	첨도 (kurtosis)	1.35

## 중앙화 경향 (Central tendency)

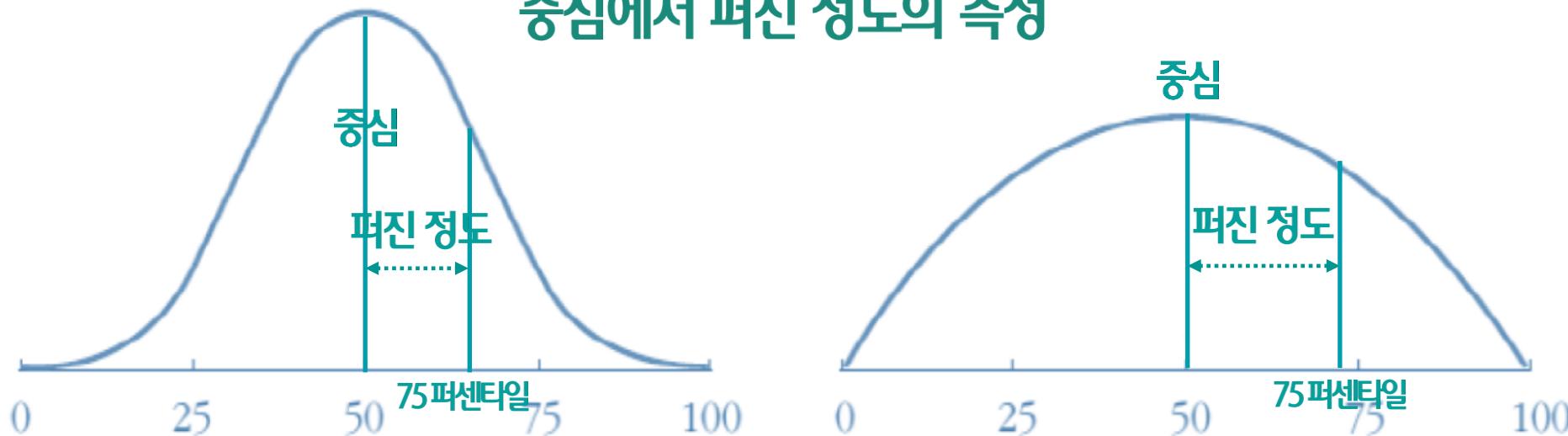
대표값으로 평균을 사용할지 중앙값을 사용할지



함수명	입력 인수	함수 설명	통계량
AVERAGE	(number1, number2, ...)	평균 산출	6.3%
MEDIAN	(number1, number2, ...)	중앙값 산출	4.7%
MODE	(number1, number2, ...)	최빈값 산출	#N/A

## 분산도 (Variation)

### 중심에서 퍼진 정도의 측정

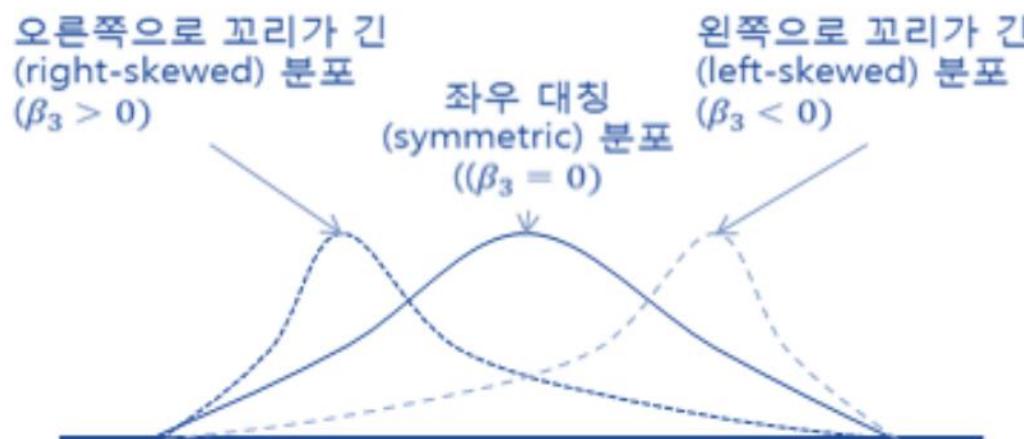


함수명	입력 인수	함수 설명	통계량
VAR	(number1, number2, ...)	분산 산출	0.8%
STDEV	(number1, number2, ...)	표준편차 산출	8.8%
PERCENTILE	(array, k)	K 퍼센타일 산출	(-0.3%, 4.7%, 11.2%)
MIN	(number1, number2, ...)	최소값	-10.2%
MAX	(number1, number2, ...)	최대값	34.8%

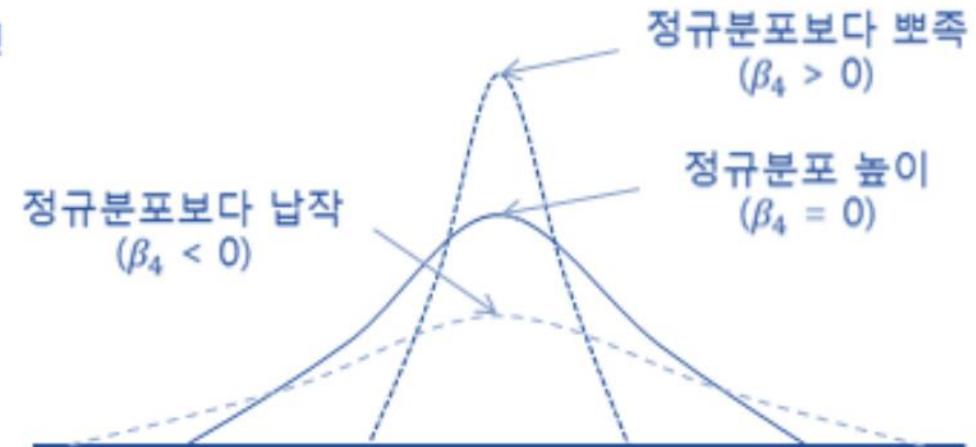
# 분포 (Distribution)

정규분포와 비교해서 치우침이 있는지, 많이 퍼져 있는지

## 왜도 (Skewness)



## 첨도(Kurtosis)

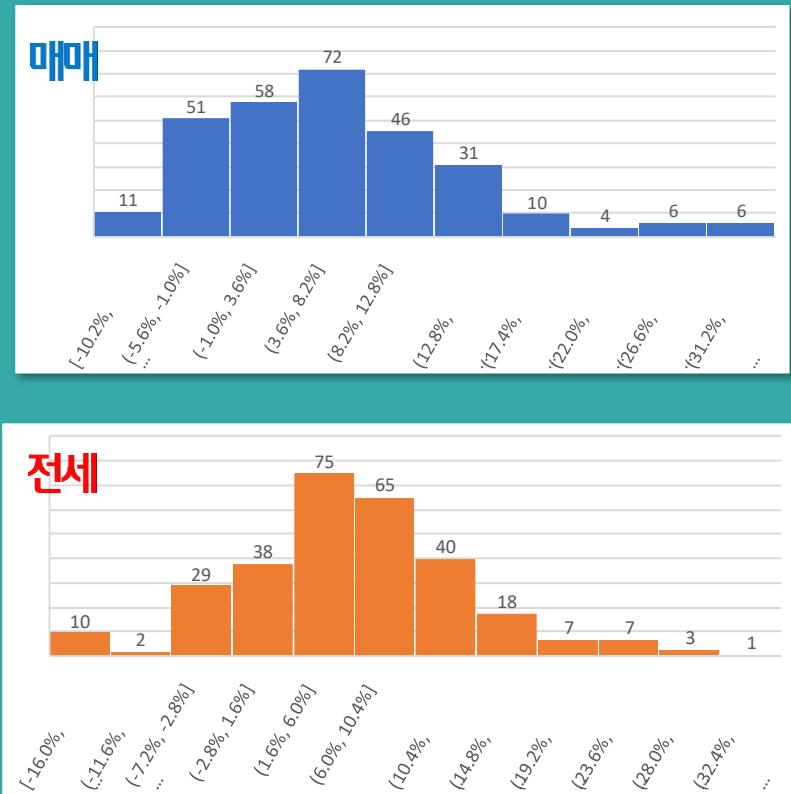


함수명	입력 인수	함수 설명	통계량
SKEW	(number1, number2, ...)	왜도 산출 (정규분포 = 0)	0.99
KURT	(number1, number2, ...)	첨도 산출 (정규분포 = 0)	1.35

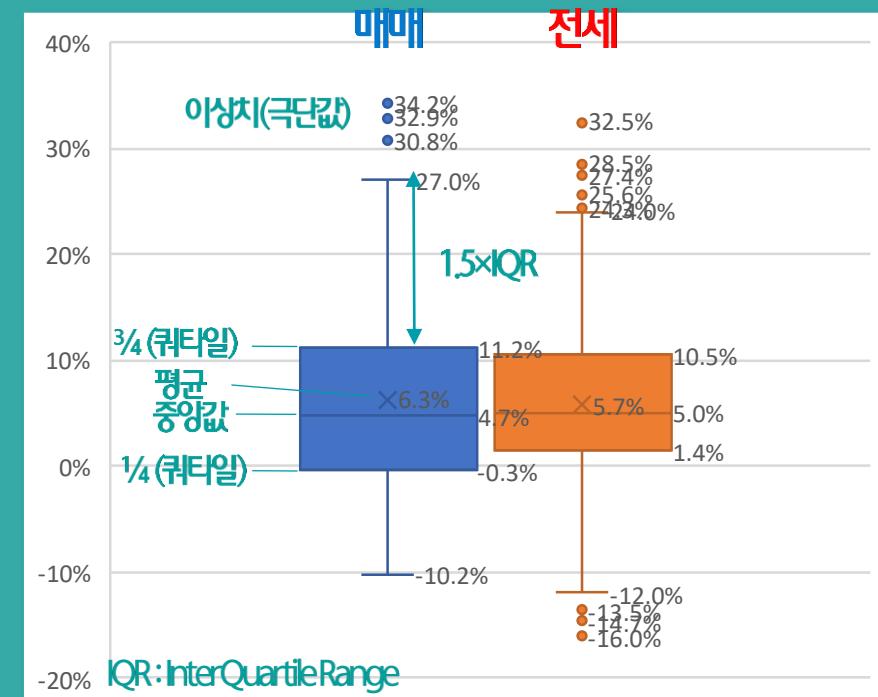
## 2. 기술 통계량을 활용한 사전 데이터 분석

[예제] (5)아파트가격2

## 변동률(기술통계량)에 대한 시각화 매매, 전세 각각의 분포 히스토그램



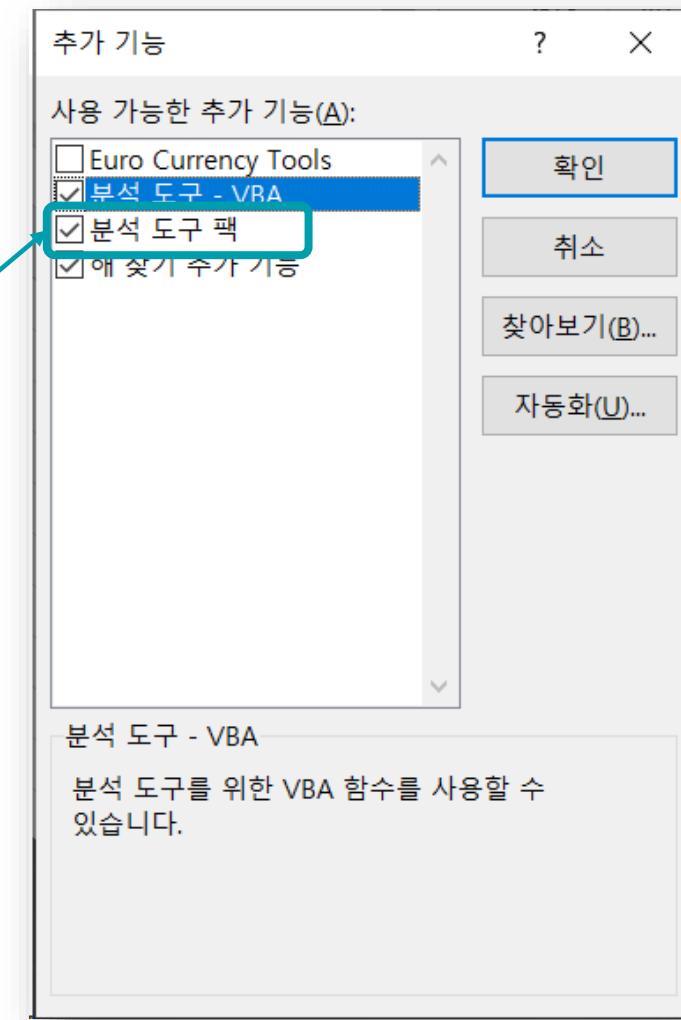
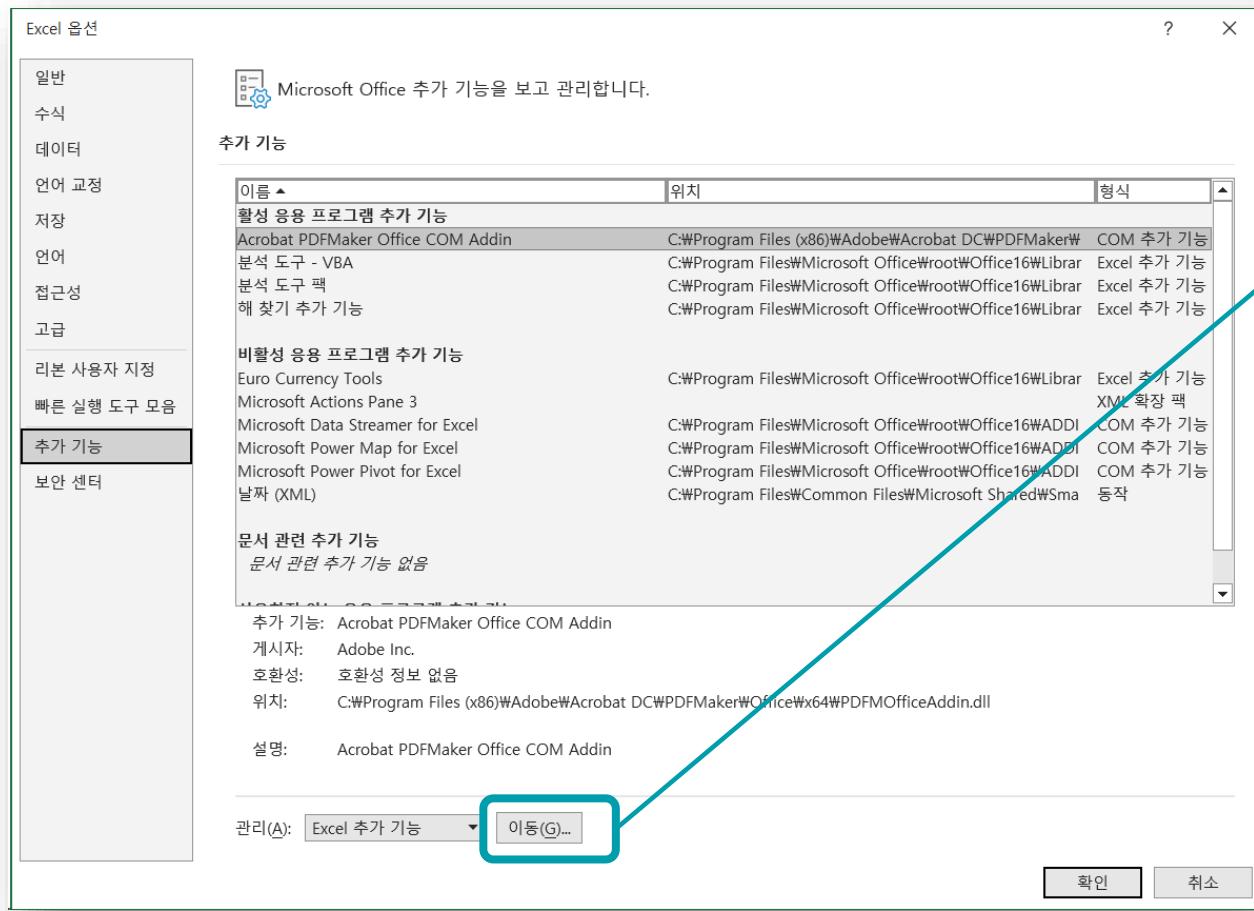
## 매매, 전세 분포를 하나의 차트에서 상자수염 차트



## 2. 기술 통계량을 활용한 사전 데이터 분석

# 데이터 분석 도구 설치

## [파일]-[옵션]-[추가기능]



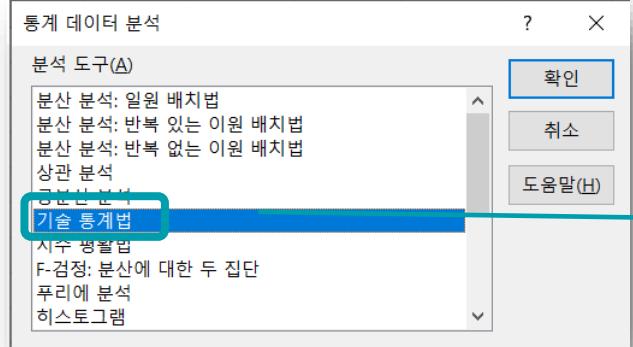
## 2. 기술 통계량을 활용한 사전 데이터 분석

[예제] (5)아파트가격2

# 컴퓨터 사용법보다 이론적 뒷받침이 중요한 이유!

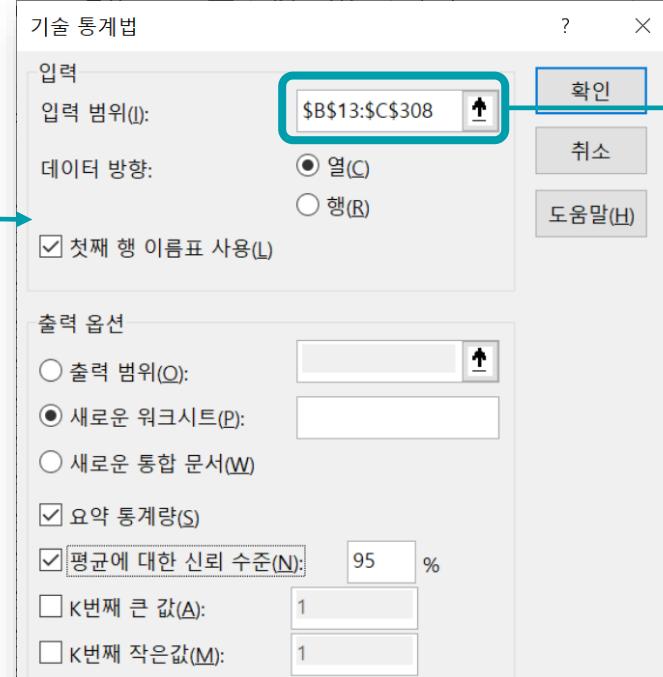
## 가격 변동 기술 통계량

(1) 분석하려는 방법을 선택하고,



[데이터]-[데이터 분석]

(2) 분석 대상 값 범위만 입력하면



(3) 끝!!!

	A	B	C	D
1	매매		전세	
2				
3	평균	6.3%	평균	5.7%
4	표준 오차	0.5%	표준 오차	0.5%
5	중앙값	4.7%	중앙값	5.0%
6	최빈값	#N/A	최빈값	#N/A
7	표준 편차	8.8%	표준 편차	8.3%
8	분산	0.8%	분산	0.7%
9	첨도	1.35	첨도	0.80
10	왜도	0.99	왜도	0.24
11	범위	45.0%	범위	48.5%
12	최소값	-10.2%	최소값	-16.0%
13	최대값	34.8%	최대값	32.5%
14	합	1854.1%	합	1695.6%
15	관측수	295	관측수	295
16	신뢰 수준(95%)	1.0%	신뢰 수준(95%)	1.0%

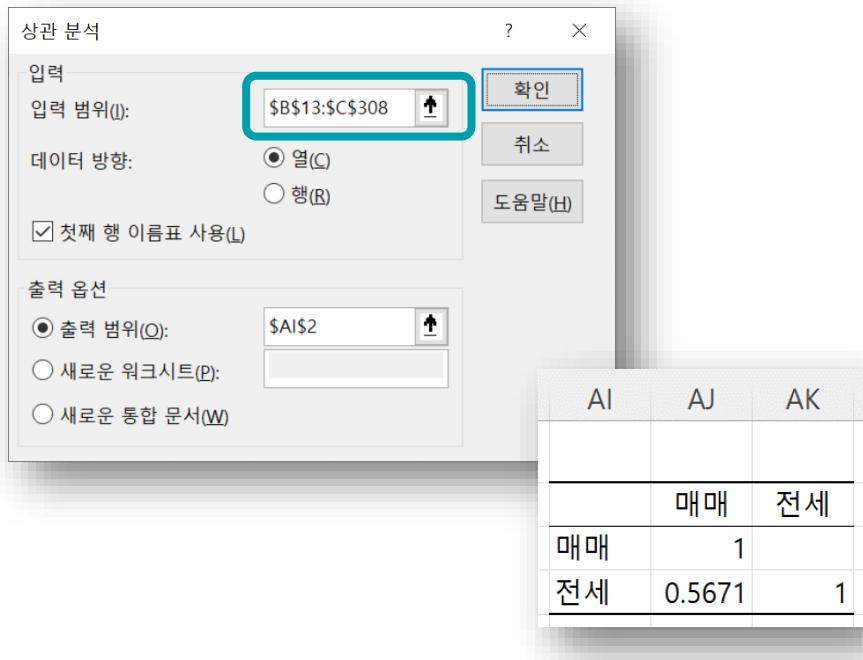
## 2. 기술 통계량을 활용한 사전 데이터 분석

[예제] (5)아파트가격2

# 데이터 분석 도구는 여러 계열을 동시에 분석할 경우 유용!

## 2 계열 vs 16계열 상관분석

### 서울 특별시 매매 vs 전세



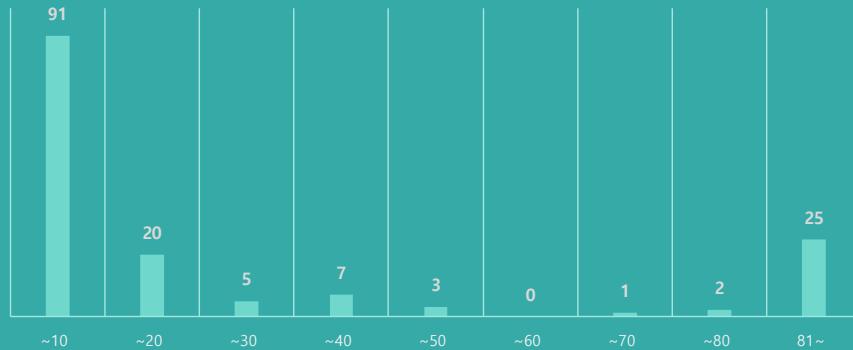
### 모든 지역 매매 vs 전세

		전국 평균		서울특별시		부산광역시		대구광역시		인천광역시		광주광역시		대전광역시		울산광역시	
		매매	전세	매매	전세												
전국 평균	매매	1															
	전세	69%	1														
서울특별시	매매	88%	51%	1													
	전세	66%	96%	57%	1												
부산광역시	매매	73%	73%	41%	62%	1											
	전세	58%	89%	36%	80%	85%	1										
대구광역시	매매	70%	72%	49%	70%	61%	52%	1									
	전세	59%	89%	40%	84%	60%	73%	86%	1								
인천광역시	매매	88%	57%	84%	57%	54%	45%	49%	41%	1							
	전세	67%	91%	60%	93%	57%	75%	65%	78%	71%	1						
광주광역시	매매	49%	36%	13%	22%	62%	38%	54%	44%	21%	11%	1					
	전세	49%	60%	17%	43%	66%	62%	58%	64%	20%	32%	83%	1				
대전광역시	매매	73%	42%	49%	32%	79%	49%	59%	40%	52%	29%	58%	55%	1			
	전세	65%	82%	50%	77%	78%	88%	53%	68%	55%	73%	22%	43%	60%	1		
울산광역시	매매	80%	62%	53%	53%	69%	54%	66%	60%	55%	45%	64%	69%	62%	49%	1	
	전세	59%	82%	36%	73%	64%	75%	52%	74%	43%	66%	44%	65%	33%	60%	78%	1

## 2. 기술 통계량을 활용한 사전 데이터 분석

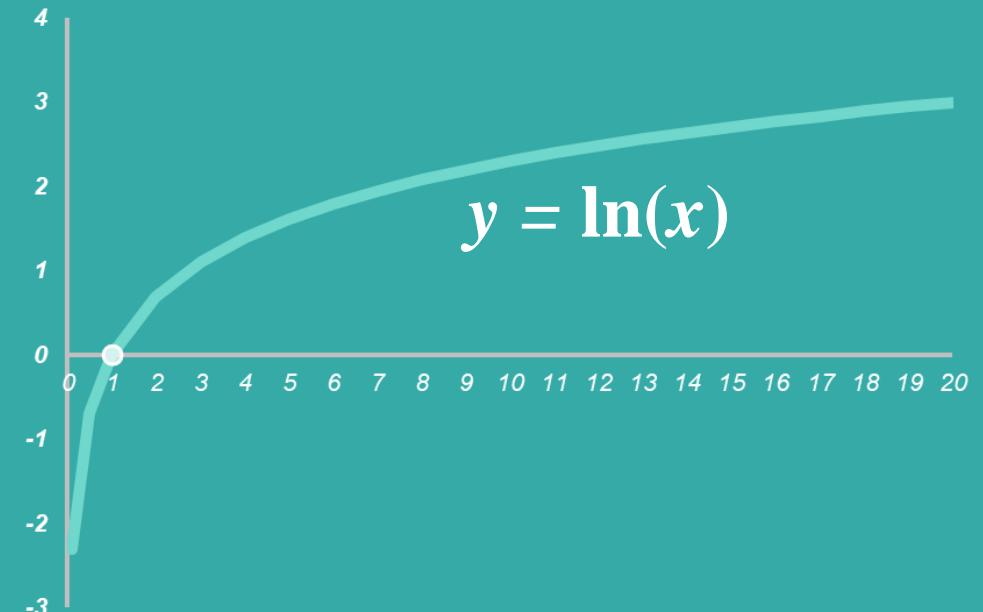
# 멱급수 분포(지수 함수)를 보이는 경우 ‘로그 변환’이 일반적

변경!!! 조회수의 로그 변환



‘반토막’을 회복하려면, ‘더블’!

로그 함수

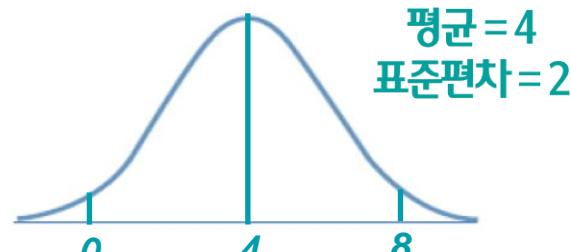


## 2. 기술 통계량을 활용한 사전 데이터 분석

# Z-score를 통한 표준화로 서로 다른 데이터 결합

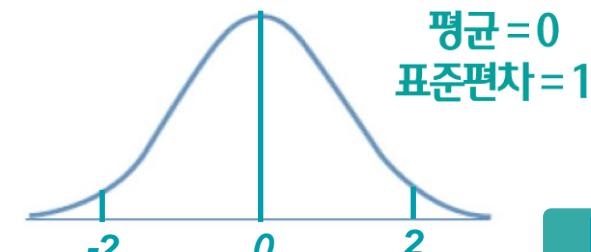
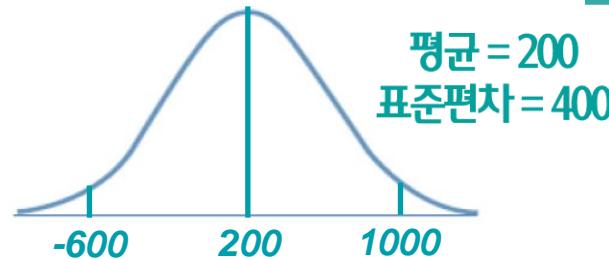
‘정규화’는 값의 범위 차이를 왜곡하지 않고 동일한 척도로 변경하는 것

다른 범위값 분포

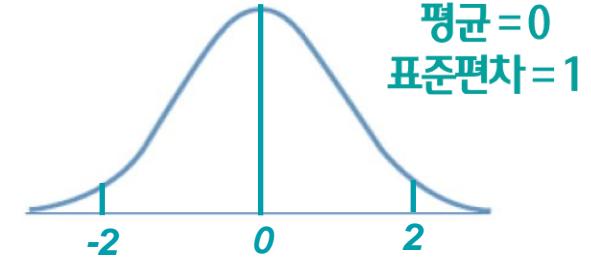


$$z = (x - \mu) / \sigma$$

- (1) 중심을 이동하고  
(2) 범위를 동일하게



표준 정규분포



함수명	입력 인수	함수 설명	예제
NORM.S.DIST	(z, cumulative)	표준정규분포 z값의 누적분포값(True) 또는 확률밀도함수(False) 반환	NORM.S.DIST(1.96, TRUE) = 0.975
NORM.S.INV	(probability)	표준정규분포 누적분포값의 z값 반환	NORM.S.INV(0.975) = 1.96

## 핵심요약

- 차트는 데이터 계열을 매핑 시켜 주는 것으로, 데이터 원본 대화상자 뿐 아니라, 직접 범위를 수정하거나, 마우스로 범위를 끌어 수정 가능
- 값이 적을 때는 막대차트도 유용하나, 대체적으로 데이터 분석에는 선 차트가 가장 유용하며, 필요에 따라 영역 차트는 이중 축 혼합차트를 통해 활용
- 두 계열의 상관관계 또는 인과관계 여부를 분석할 때는, 분산형 차트를 활용
- 기술 통계량은 시각화 없이 원 데이터의 특성을 파악하는 방법으로, 대표성, 분산 도, 분포도를 수치화 한 것. 이를 시각화 한 것이 히스토그램과 상자수염 차트
- 데이터 분석 도구를 활용하면 통계 분석이 수월하나, 도구 사용법 보다는 통계 이론 지식이 중요!
- 세상에는 지수함수 또는 역급수 분포가 혼하여, 이 경우 Log 변환을 사용. 서로 다른 데이터 척도의 결합 분석시 Z-score를 통한 정규화 활용 가능

q i i i  
i i a i a i i a  
i a i i a i a i  
a a i a i a a  
q i i i a i i i  
a i a

# 8. 통계적 가설검정 방법론과 데이터 분석

-	-	-	-	-	-	-
a	-	a	-	a	-	a
a	-	a	-	a	-	a
a	-	a	-	a	-	a
a	-	a	-	a	-	a
a	-	a	-	a	-	a
-	-	-	-	-	-	-
a	-	a	-	a	-	a
a	-	a	-	a	-	a
a	-	a	-	a	-	a
a	-	a	-	a	-	a
-	-	-	-	-	-	-

# 학습목표

- ▶ 통계적 가설검정법을 이해하여, 통계적 분석의 적용을 위한 인사이트 함양
- ▶ 엑셀의 [데이터 분석 도구]를 활용하여 빠르게 주요 통계 분석 수행법 습득
- ▶ 회귀분석법을 활용하여 상관관계를 찾고, 해석하며, 적용하는 방법 이해
- ▶ 통계적 분석 방법론을 통해 주요 재무이론의 의미와 적용법을 정확히 이해

1. 통계적 가설 검정 방법론

# 가설검정은 아는 정도와 모르는 정도를 확인하는 것

## 공자의 얕에 대한 철학

由(유)야!

誨女知之乎(회여지지호)

知之爲知之(지지위지지)

不知爲不知(부지위부지)

是知也(시지야)

자로야!

내가 너한테 안다는 것에 대하여  
가르쳐 주겠노라.

아는 것은 안다고 하고,

모르는 것은 모른다고 하는 것,

이것이 진정 아는 것이다.



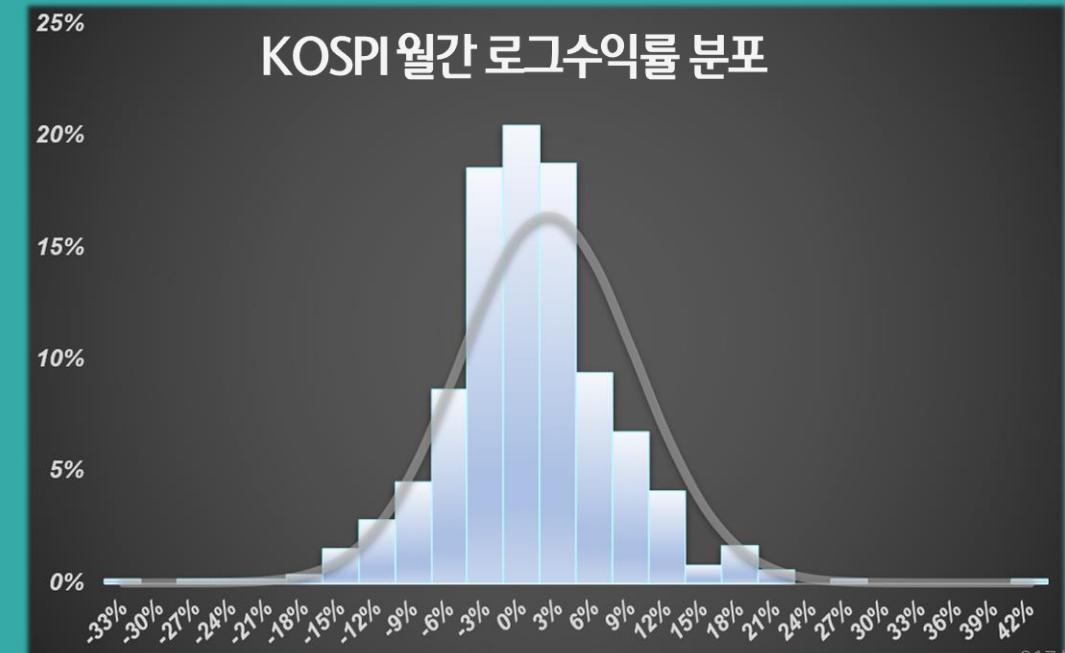
## 1. 통계적 가설 검정 방법론

# 수익률 분포는 모집단 분포일까, 표본 분포일까?

1980년 1월 4일 100부터 모두 수집한 데이터

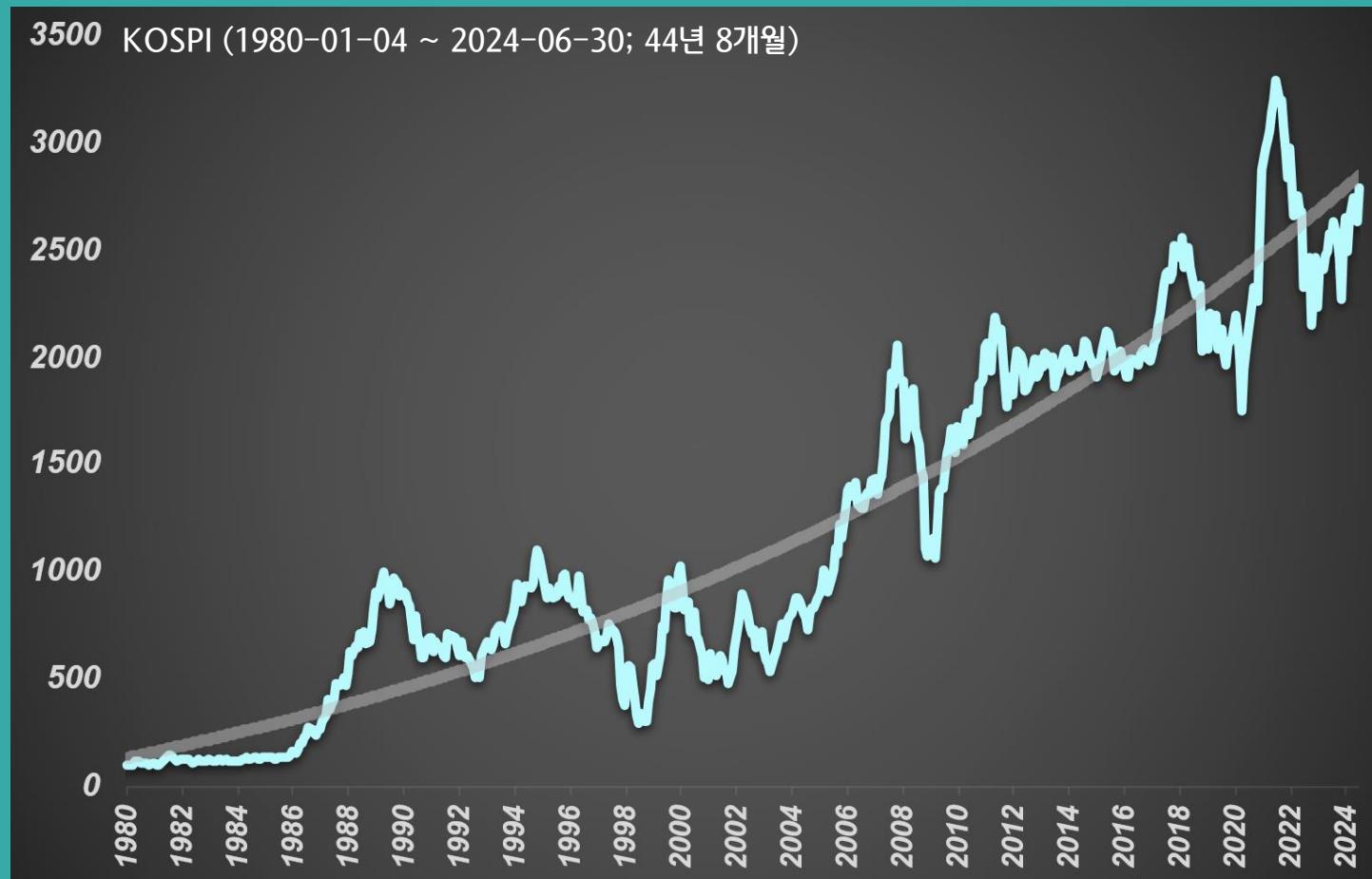


미래의 모든 데이터가 포함되어야 “모집단”



## 1. 통계적 가설 검정 방법론

# 미래에도 연 7.8%의 수익률을 낼 것인가?



평균 수익률  $R$

과거 관찰값을 기술한 것!

기대 수익률  $E(R)$

미래 예상값을 표현한 것!

모평균  $\mu$

모집단의 평균값!

기대값  $\bar{x}$

(모집단 추정을 위한)  
표본집단의 평균값!

## 1. 통계적 가설 검정 방법론

# 통계적 가설 검정법

## 귀류법

### 대립가설에 대한 거짓 증명

귀무가설  
명제  $p$

대립가설  
명제  $p^c$

“그래, 네 말( $p^c$ )이 맞다고 치자.”

“그런데, 이러이러하니까 말이 안되네.”

“네 말이 틀렸으니까, 반대가( $p$ ) 맞아.”

[예시]  $\sqrt{2}$  가 무리수임을 증명하라!

- 무리수의 정의 : 유리수가 아닌 실수
- 대립가설 :  $\sqrt{2}$  가 유리수
- 유리수 : 서로 소인 두 정수  $p/q$  로 표현 가능
- $p/q$  로 표현이 불가능함을 증명

## 통계적 가설 검정(추론)

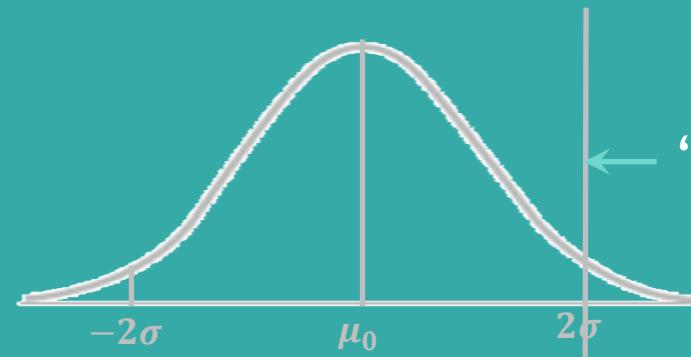
### 대립가설에 대한 증명

귀무가설  
 $\mu = \mu_0$

대립가설  
 $\mu \neq \mu_0$

경계 모호

“사전에 경계 정하기!”



- 경계를 넘어가지 못하면 ‘귀무가설’ 채택!
- 경계 넘어가면 ‘대립가설’ 채택!

## 1. 통계적 가설 검정 방법론

# 가설 검정 방법

## 임계값은 귀무가설 채택 오류 가능성의 허용치 (제1종 오류; $\alpha$ 오류)

모집단



샘플링

표본집단



관찰(측정)

추정분포

관찰값  
초과가능성  
 $p$ 

가설검정

임계값  
초과가능성  
 $\alpha$ 귀무가설  
( $H_0$ )  
채택  
 $\mu = \mu_0$ 대립가설  
( $H_1$ )  
채택  
 $\mu > \mu_0$ 

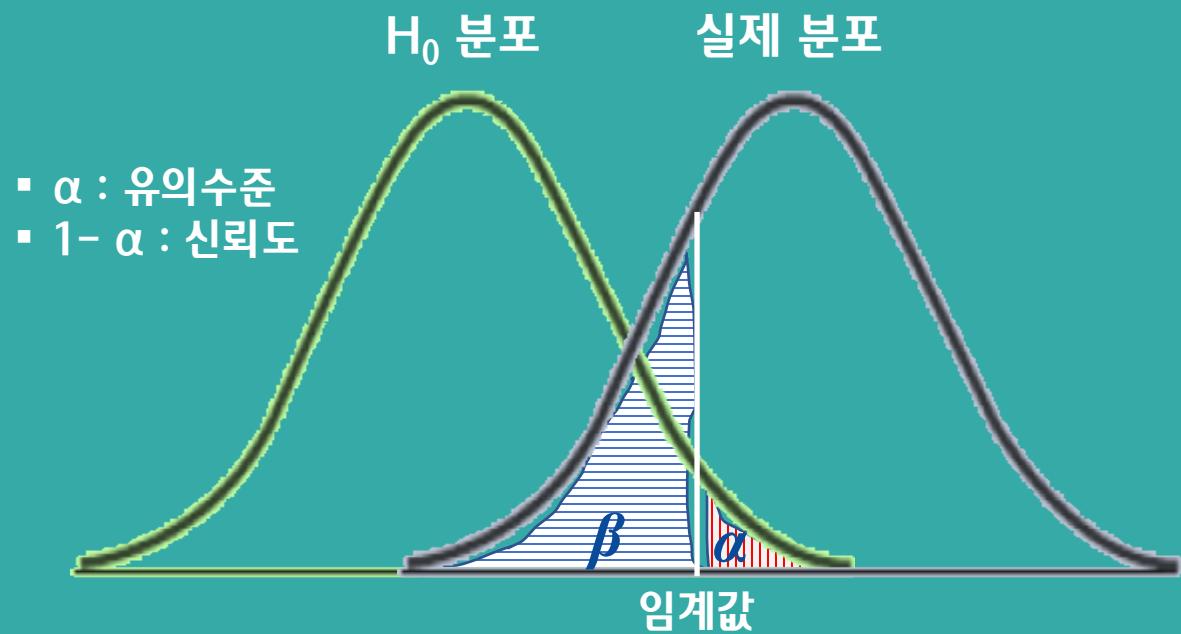
- ✓ “분포에 관계 없는” 확률값
- ✓ “흑백논리 결론”에 대한 틀릴 가능성 내포

## 1. 통계적 가설 검정 방법론

# 1종 오류와 2종 오류는 상충 관계 (Trade-off)

일반적인 경우에 1종 오류 중심 임계값을 정하고 검정  
암 진단 경우에는 2종 오류가 더 치명적 (대신 1종 오류에 관대 경향)

결정	
채택	기각
일치 $(1 - \alpha)$	1종 오류 $(\alpha)$
2종 오류 $(\beta)$	일치 $(1 - \beta)$



## 중심극한정리 (Central Limit Theorem)

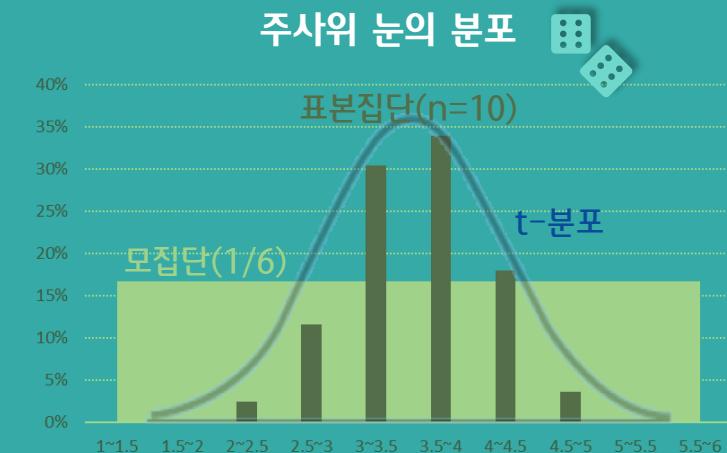
“무작위 추출시” 모수의 분포와 관계 없이 “표본평균의 분포”는 정규분포에 접근한다

**모수** 모집단의 통계적 척도 → 전수조사 하지 않는 한, 알 수 없는 값 ( $\mu, \sigma^2$ )

**통계량** 모집단에서 추출한 표본집단의 통계적 척도 → 모집단 추정 근거 ( $\bar{x}, s^2$ )

### 표본평균의 분포는 정규분포 ( $\mu, \sigma^2/n$ )에 접근

- 자책금지 : 통계 수업 들어도 이해 안 된 상태로 넘어가는 것이 태반!
- 한번 표본을 뽑을 때  $n$  개를 뽑아서 그 표본의 평균을  $x_1$  이라고 하고, 또  $n$  개를 뽑은 평균을  $x_2$ 라고 하고, 이렇게  $m$  번 반복하면,  $(x_1, x_2, \dots, x_m)$ 의 분포가 정규분포
- 쉽게 확인 1 : 표본( $n$ )을 1개 뽑으면 표본평균 분포는  $(\mu, \sigma^2)$
- 쉽게 확인 2 : 표본( $n$ )을 무한대로 뽑으면 표본평균 분포는  $(\mu, 0)$ . 모평균 그 자체
- t-분포 :  $n$ 이 25보다 작을 경우 → 통계 Tool 이용 시 t-test 하는 이유



## 1. 통계적 가설 검정 방법론

# 모든 정규분포는 표준정규분포로 해석할 수 있다!

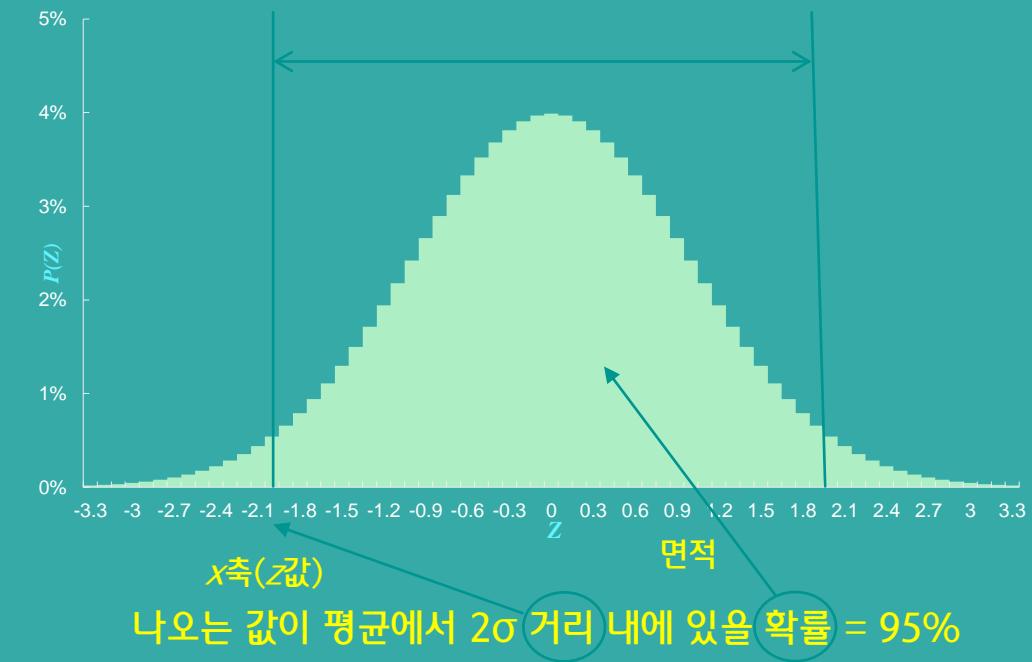
**무작위(Random) 분포는 중심극한정리에 의해 표준정규분포의 수식으로 표현됨**

## 왜 정규분포를 사용할까?

- 현실답: 이용하기 쉬워서 → 평균( $\mu$ )과 표준편차( $\sigma$ )만으로 계산
- 근사답: 세상 확률분포가 정규분포에 근사한 것이 많아서
- 무작위 발생 분포: 중심극한정리 (가설검정법)

## 이용하기 편한 이유

- 표현 용이: “평균보다  $2\sigma$  이상 나올 확률” → 2.5%
- 양쪽이 대칭이라 반쪽만으로도 사용 가능
- 모든 정규분포  $N(\mu, \sigma^2)$ 는 표준 정규분포  $Z(0, 1^2)$ 로 표현 가능
- 표준위치값  $z = (x - \mu) / \sigma$
- 원위치값  $x = \mu + z\sigma$



## 2. 데이터 분석 도구의 활용

[예제] (4)아파트가격

# 두 집단(매매가와 전세가 변동률의 평균)의 차이가 있는가?

## t-검정 (쌍체비교)

t-검정: 쌍체비교

입력

변수 1 입력 범위(1): \$B\$13:\$B\$308  
변수 2 입력 범위(2): \$C\$13:\$C\$308

가설 평균차(E):

이름표(L)  
유의 수준(A): 0.05

귀무가설: 차이 없다

출력 옵션

출력 범위(O):  
 새로운 워크시트(P):  
 새로운 통합 문서(W):

	A	B	C
1	t-검정: 쌍체 비교		
2			
3	매매	전세	
4	평균	6.28%	5.75%
5	분산	0.77%	0.70%
6	관측수	295	295
7	피어슨 상관 계수	0.57	
8	가설 평균차	0.00%	
9	자유도	294	
10	t 통계량	1.16	
11	P(T<=t) 단측 검정	12.4%	
12	t 기각치 단측 검정	1.65	
13	P(T<=t) 양측 검정	24.8%	
14	t 기각치 양측 검정	1.97	
15			

0.53% 차이가 유의한가?

비교적 높은 상관관계

귀무가설 채택,  
= 대립가설 기각  
→ “차이없다”

## 2. 데이터 분석 도구의 활용

[예제] (4)아파트가격

# 세 집단(서울, 부산, 대구 매매가 변동률의 평균)의 차이가 있는가?

## 분산 분석 (일원 배치법)



분산 분석: 일원 배치법

요약표

세 값의 차이가 유의한가?

	A	B	C	D	E	F	G
1	분산 분석: 일원 배치법						
2							
3	요약표						
4	인자의 수준	관측수	합	평균	분산		
5	서울	295	18.5	6.28%	0.77%		
6	부산	295	14.0	4.74%	0.58%		
7	대구	295	12.2	4.13%	0.52%		
8							
9							
10	분산 분석						
11	변동의 요인	제곱합	자유도	제곱 평균	F 비	P-값	F 기각치
12	처리	0.07	2	3.65%	5.84	0.30%	3.01
13	잔차	5.51	882	0.63%			
14							
15	계	5.59	884				
16							

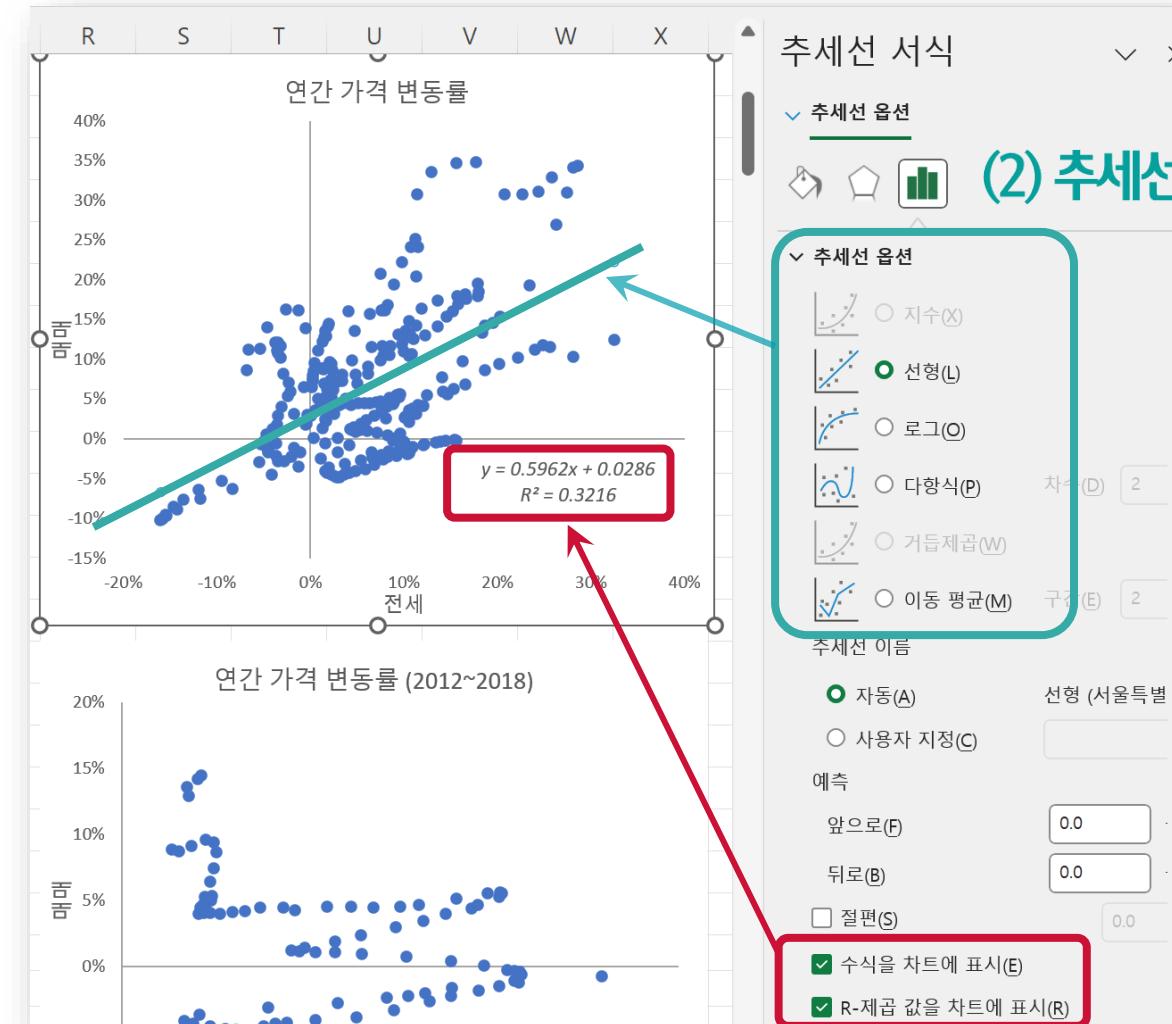
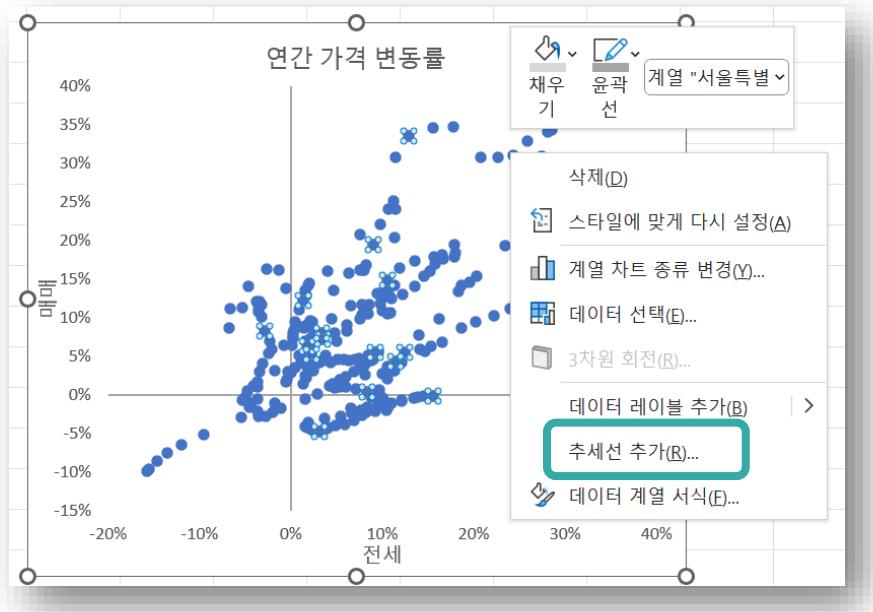
귀무가설 기각  
= 대립가설 채택  
→ “차이 있다”

## 2. 데이터 분석 도구의 활용

[예제] (5)아파트가격2

# 분산 차트에서 발견한 추세선(회귀식) 작성

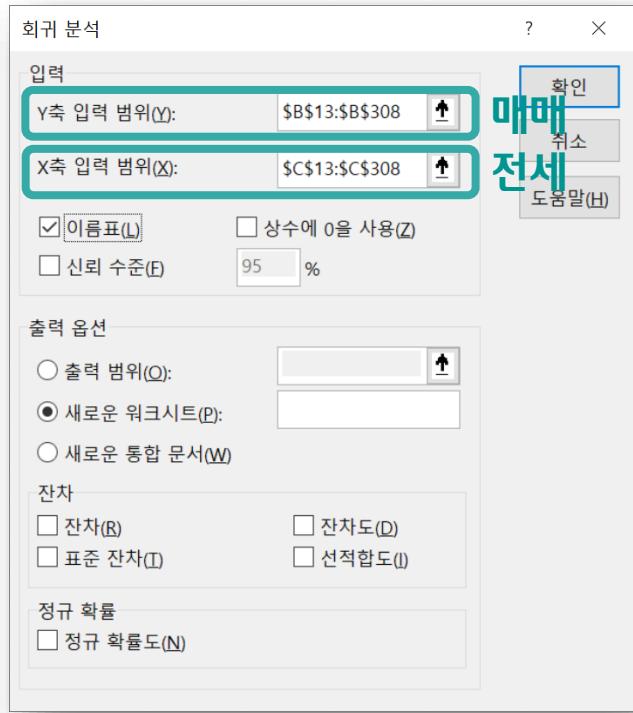
(1) 계열 선택 후, [팝업메뉴]-[추세선 추가] 클릭!



## 2. 데이터 분석 도구의 활용

[예제] (5)아파트가격2

# 툴 사용법 << 해석 능력



A	B	C	D	E	F	G
1	요약 출력					
2						
3	회귀분석 통계량					
4	다중 상관계수	0.5671				
5	결정계수	0.3216	$R^2$			
6	조정된 결정계수	0.3193				
7	표준 오차	7.24%	$\sigma$			
8	관측수	295				
9						
10	분산 분석					
11		자유도	제곱합	제곱 평균	F 비	유의한 F
12	회귀	1	0.7274	0.7274	138.88	0.0000
13	잔차	293	1.5345	0.0052		
14	계	294	2.2619			
15						
16		계수	표준 오차	t 통계량	P-값	하위 95% 상위 95%
17	Y 절편 $\alpha$	2.86%	0.51%	5.58	0.0000%	1.85% 3.87%
18	전세 $\beta$	0.60	0.05	11.78	0.0000%	0.50 0.70

$$y = 0.5962x + 0.0286$$

$R^2 = 0.3216$

P-값이 5%보다 작으면 계수 신뢰성 높은 것!

# 알고 있는 함수식으로 상관관계 분석

## 선형 회귀분석

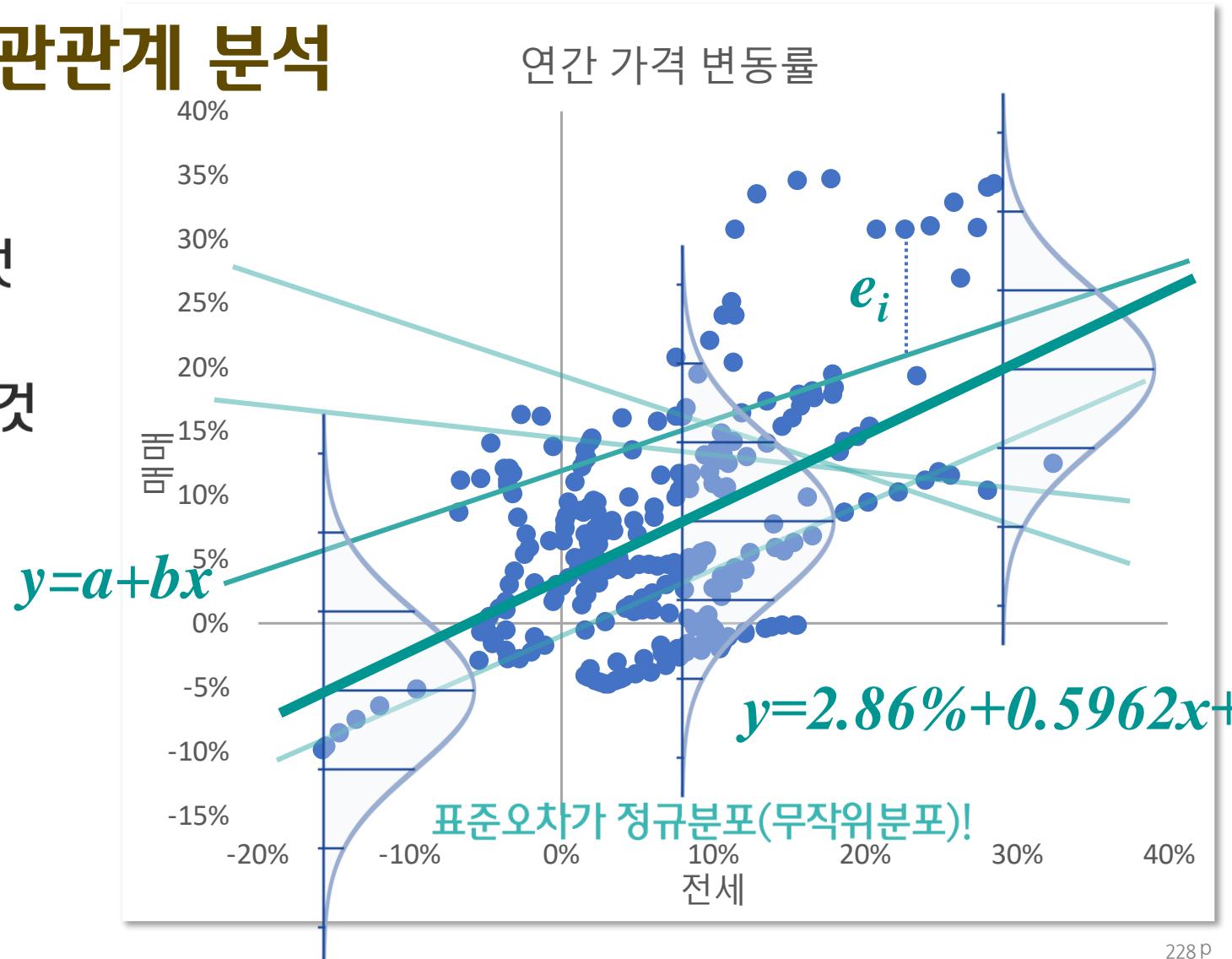
- 1차함수  $y = a + bx$ 에 근접시키는 것
- 최소자승법(Least Squared Method) :  
에러의 크기가 가장 작은  $a$ 와  $b$ 를 찾는 것

$a = 2.86\%$  절편

$\beta = 0.5962$  기울기

$\varepsilon = 7.24\%$  표준오차

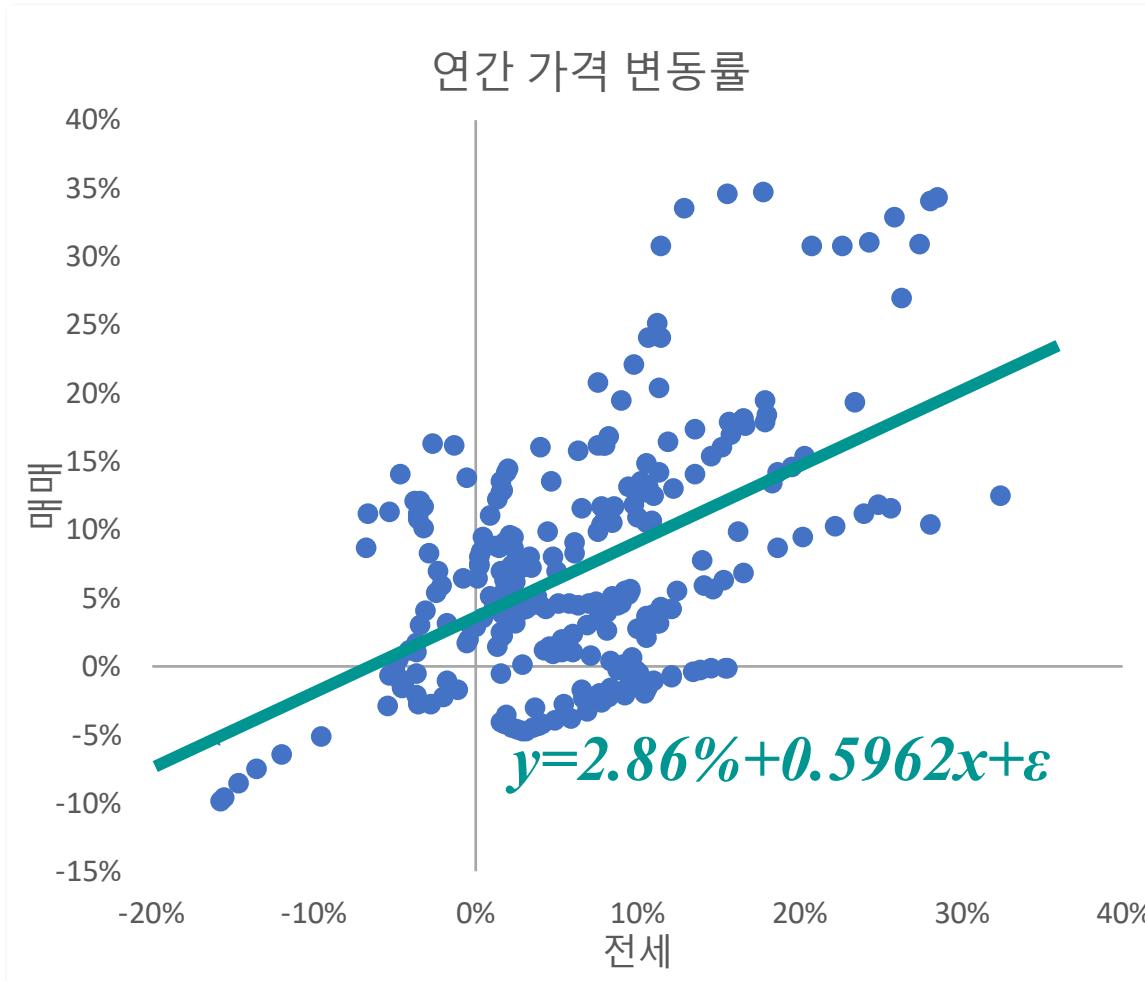
$R^2 = 0.3216$  설명력



## 3. 상관관계와 회귀분석법

[예제] (5)아파트가격2

## 자료 해석



$$\alpha = 2.86\% \quad \text{절편}$$

매매가는 기본적으로 연 평균 2.86% 상승  
→ (추정) 물가상승률 등 다른 원인에 의함

$$\beta = 0.5962 \quad \text{기울기}$$

전세가와 매매가는 동락의 방향이 대체적으로 같으며,  
전세가 1단위 당 매매가는 0.5962단위 변동

$$\varepsilon = 7.24\% \quad \text{표준오차}$$

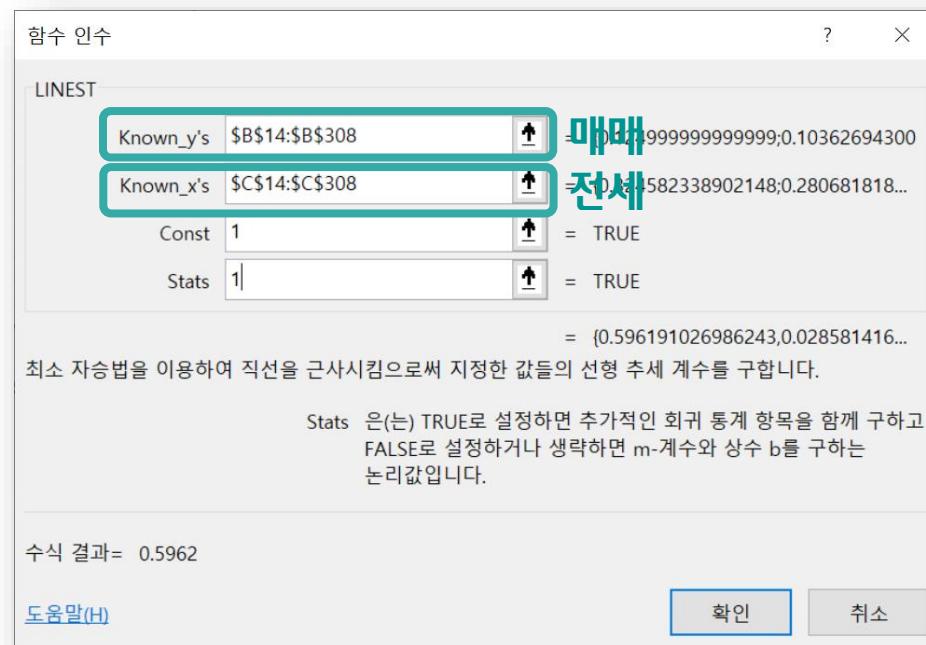
이 관계식에 의해 매매가를 추정할 때,  
- 67% 신뢰구간으로 7.24% 이내  
- 95% 신뢰구간으로 14.48% 이내의  
추정 오류 가능성이 있음

오차를 줄이는 법  
→ 다른 관계변수를 찾아 중회귀분석!

## 3. 상관관계와 회귀분석법

[예제] (5)아파트가격2

## LINEST 함수를 활용한 회귀분석



$$y = 2.86\% + 0.5962x + \varepsilon \quad \varepsilon = 7.24\%$$

5×2개 다중출력 (Ctrl-Shift-Enter)

선형 회귀분석		$\beta$	$\alpha$	t-값( $\beta$ )	t-값( $\alpha$ )
오차( $\beta$ )	오차( $\alpha$ )	0.0506	0.51%	11.78	5.58
$R^2$	오차( $y$ )	0.3216	0.0724	0.00%	0.00%
$F$	자유도	138.88	293		
SS <sub>reg</sub>	SS <sub>residual</sub>	0.7274	1.5345		

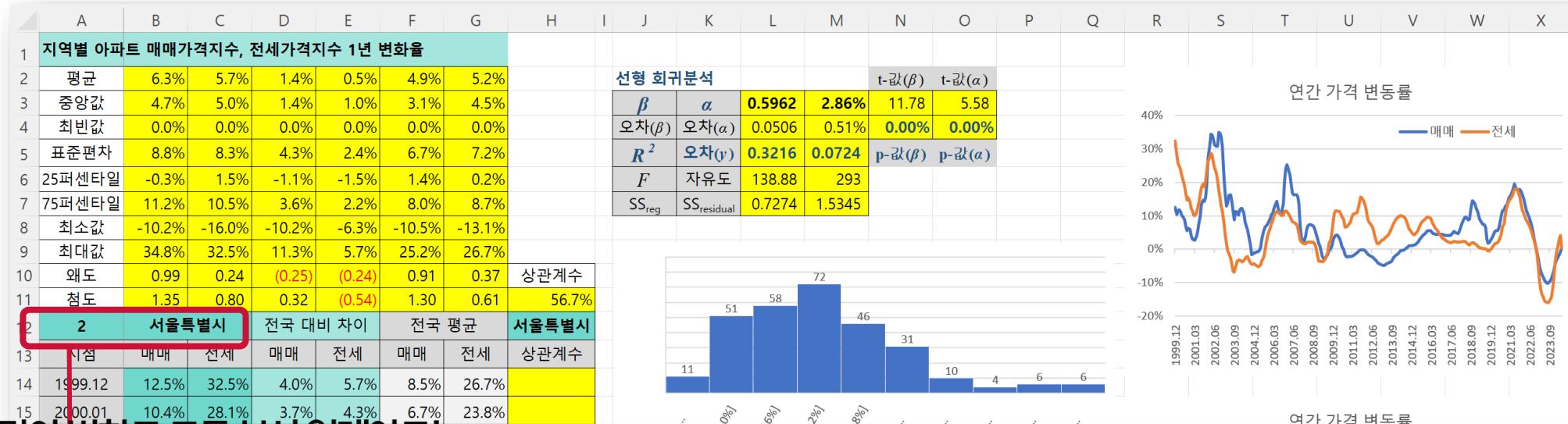
직접 계산

이름	수식	적용식	값
t-값 ( $\beta$ )	$\beta$ / 오차( $\beta$ )	=L3/L4	11.78
t-값 ( $\alpha$ )	$\alpha$ / 오차( $\alpha$ )	=M3/M4	5.58
p-값 ( $\beta$ )	1 - 정규분포 누적확률[t-값 ( $\beta$ )]	=1-NORM.S.DIST(N3, 1)	0.00%
p-값 ( $\alpha$ )	1 - 정규분포 누적확률[t-값 ( $\alpha$ )]	=1-NORM.S.DIST(O3, 1)	0.00%

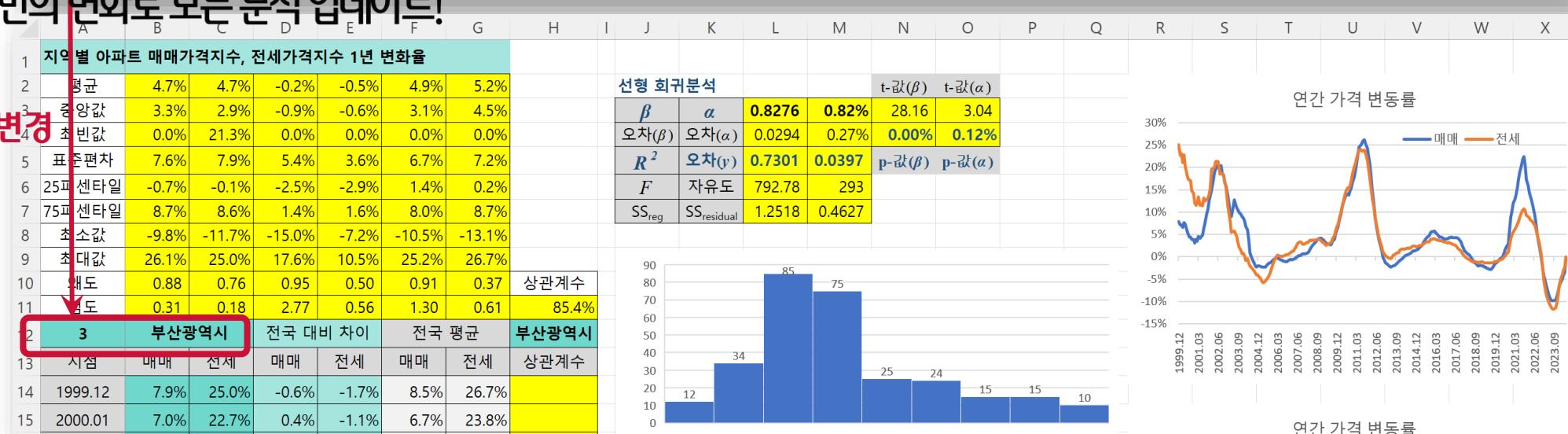
## 3. 상관관계와 회귀분석법

## 함수 사용의 장점 : 반자동화 (분석 템플릿)

[예제] (5)아파트가격2



분석 지역 변경



## 3. 상관관계와 회귀분석법

# 이제는 논문을 읽어도 두렵지 않다!

## Stocks versus Bonds: Explaining the Equity Risk Premium

Clifford S. Asness

$$D/P = \gamma_0 + \gamma_1 Y + \gamma_2 \sigma(\text{Stocks}) + \gamma_3 \sigma(\text{Bonds}) + \varepsilon_{D/P,t}. \quad (6)$$

$$D/P = 4.10\% - 0.03Y \quad (7)$$

(40.72) (-2.26)

(with an adjusted  $R^2$  of 0.7 percent) and

$$D/P = 2.02\% + 0.14\sigma(\text{Stocks}) - 0.07\sigma(\text{Bonds}) \quad (8)$$

(11.87) (18.96) (-5.24)

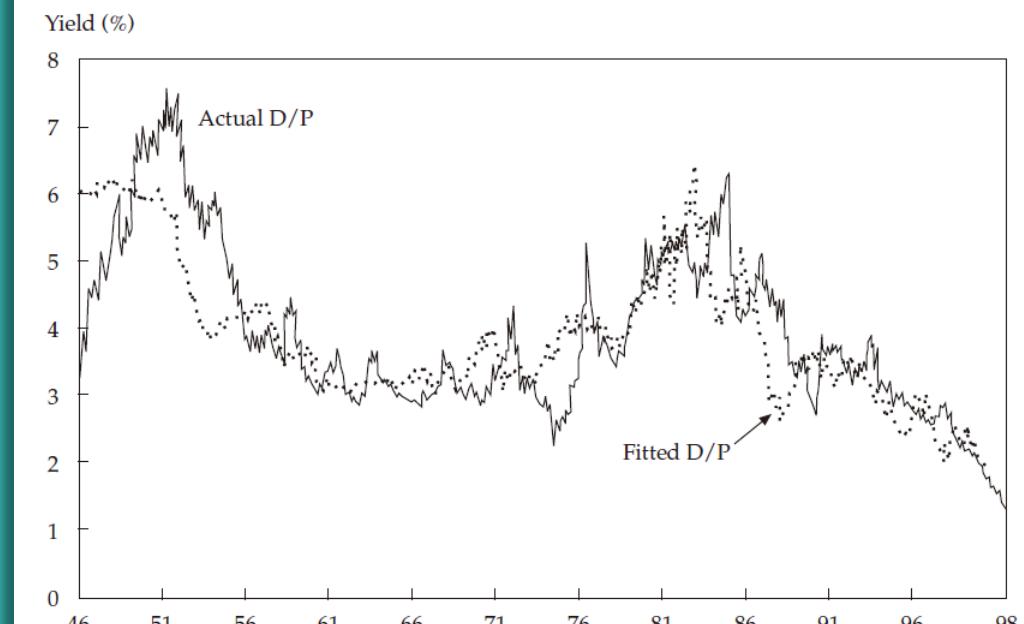
(with an adjusted  $R^2$  of 43.0 percent).<sup>12</sup>

$$D/P = 0.00\% + 0.35Y + 0.23\sigma(\text{Stocks}) - 0.31\sigma(\text{Bonds}) \quad (9)$$

(-0.05) (28.77) (39.51) (-25.69)

(with an adjusted  $R^2$  of 75.4 percent).

Figure 3. Actual S&P 500 Dividend Yield and In-Sample Dividend Yield, January 1946–May 1998



Note: In-sample D/P fitted from the regression in Equation 9.

Fitted D/P : 각 시기의 Input 값 ( $Y, \sigma(\text{Stocks}), \sigma(\text{Bonds})$ )을 회귀식에 대입하여 얻은 D/P 값의 그래프

## 4. 통계 해석을 통한 재무 이론 이해 심화

# 퀀트 매니저의 통계적 접근법

다수의 주식에 공통적으로  
적용되는 주가변동 요인

Key Factor



일관성

반복효과 및 상쇄효과를 통해  
Key Factor의 설명력을 높임

## 1개월 동안 30% 수익을 올린 주식

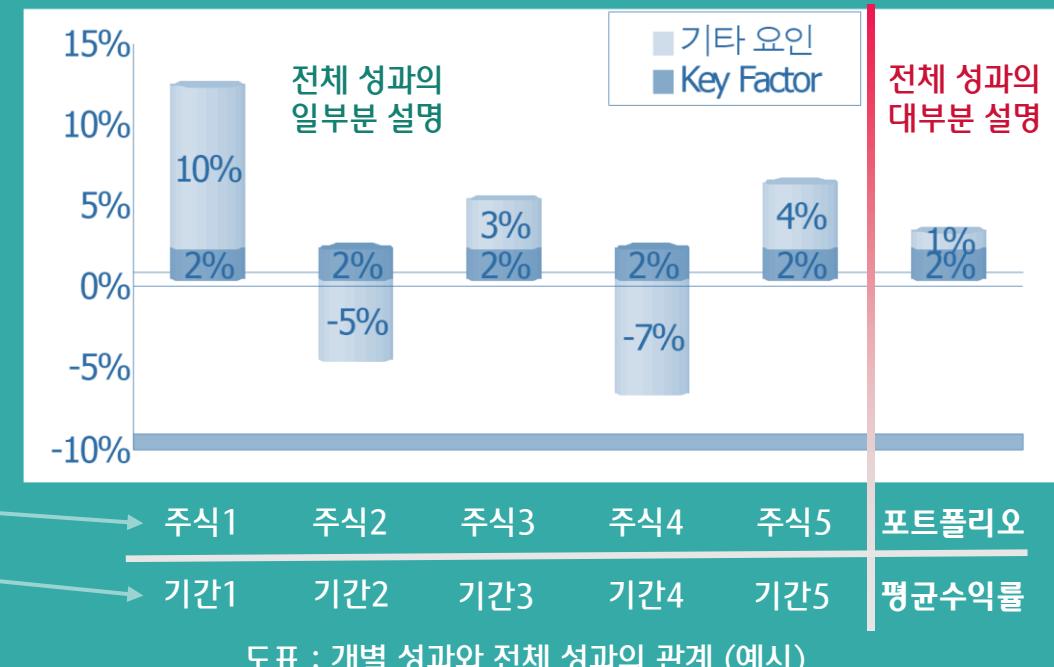
Key Factor의 수익률 기여도는?

- 가능성 1 : 매우 작을 것임. 특수한 원인에 의한 급상승
- 가능성 2 : 그 동안 반영되지 않다가, 어떤 계기로 반영됨

## Key Factor의 설명력 높이는 방법

- 방법 1 : 포트폴리오의 구축
- 방법 2 : 동일 전략을 장기간 적용

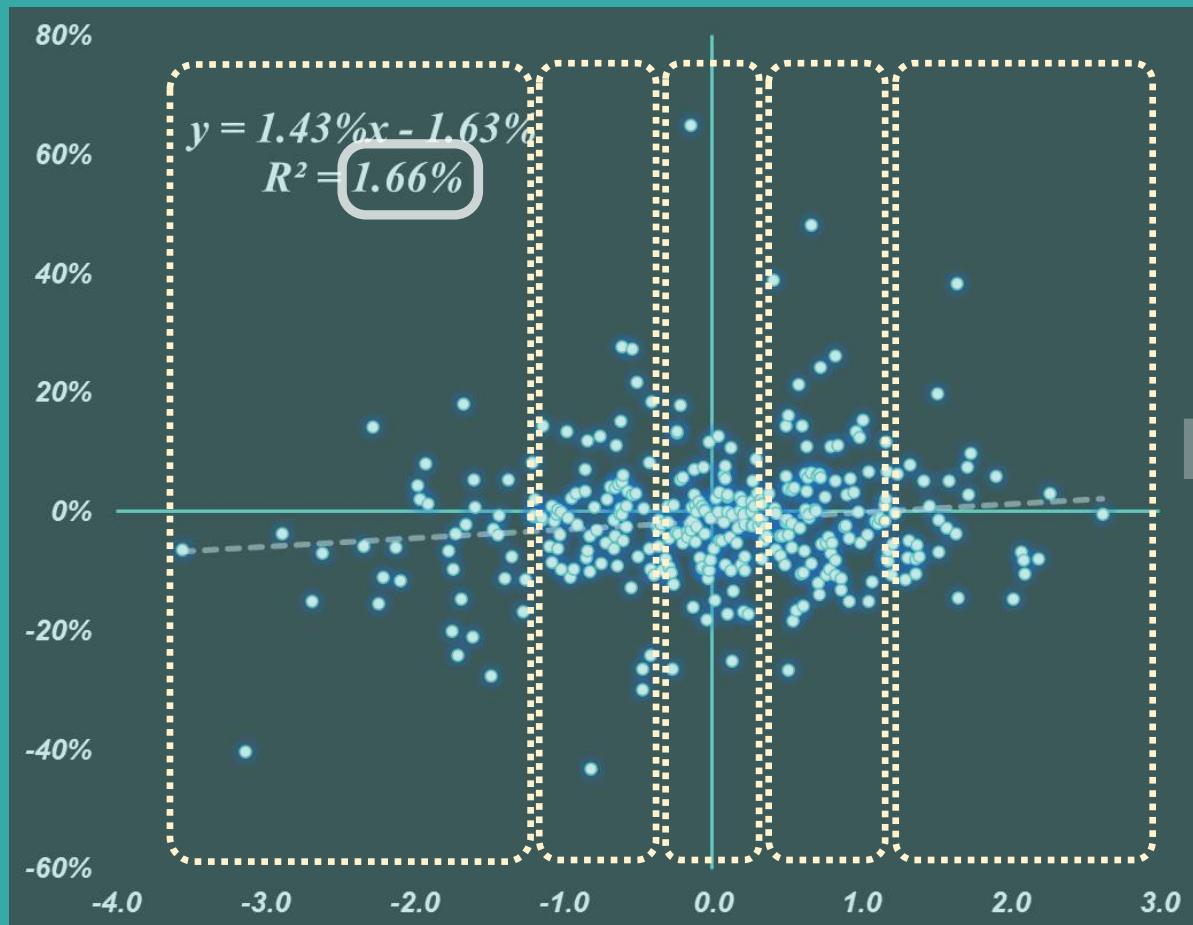
기타효과의 상쇄효과로 Key Factor의 기여도가 높아짐



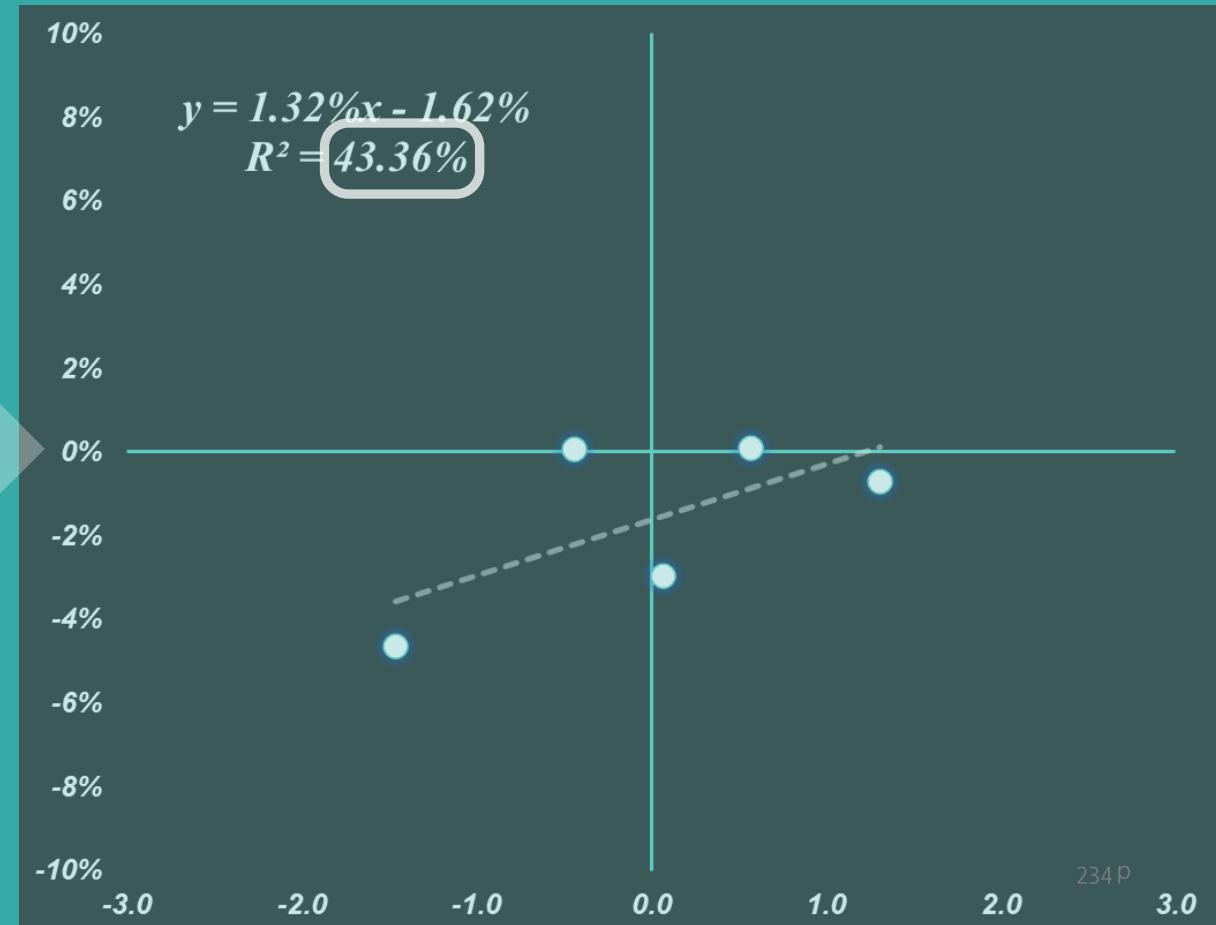
## 4. 통계 해석을 통한 재무 이론 이해 심화

# 퀀트 매니저 통계적 접근법

## 개별종목 단위 회귀분석



## 포트폴리오 단위 회귀분석



## 운칠기삼(運七技三)이지만, ‘기’에 집중해야 하는 이유

$$dS_t = \mu S_t dt + \sigma S_t dw_t$$

Drift

Diffusion

연기의 중심방향 (Drift)

기(技) : 바람 (편더멘탈)

연기의 무작위 퍼짐 (Diffusion)

운(運) : 무작위 움직임  
 기(技) : 확산 속도 (위험관리)

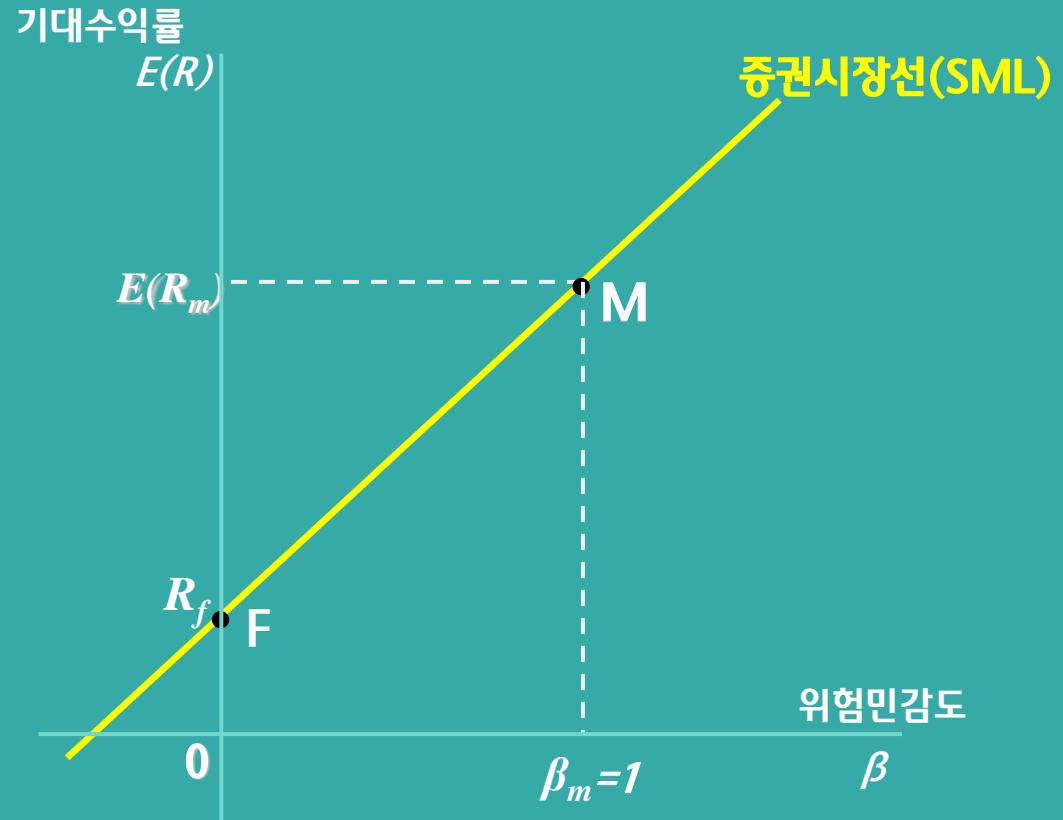
## 4. 통계 해석을 통한 재무 이론 이해 심화

# CAPM은 단일 요인 선형회귀 모형이었다!

## 자본자산가격결정모형 (Capital Asset Pricing Model)

$$\begin{aligned}
 E(R_i) &= R_f + [E(R_m) - R_f] \cdot Cov(R_i, R_m) / \sigma_m^2 \\
 &= R_f + [E(R_m) - R_f] \cdot \beta_i \\
 y &= \alpha + x \cdot \beta
 \end{aligned}$$

$$\begin{array}{l}
 E(R_i) - R_f = [E(R_m) - R_f] \cdot \beta_i \\
 \text{종목초과수익률} \quad \text{시장초과수익률}
 \end{array}$$



## 4. 통계 해석을 통한 재무 이론 이해 심화

# 상관관계는 인과관계가 아니다!

## 차익거래가격결정이론 (Arbitrage Pricing Theory)

$$E(R_p) = R_f + \frac{[E(R_{F1}) - R_f] \cdot \beta_{p1}}{\text{요인 1} \quad \text{민감도 1}} + \frac{[E(R_{F2}) - R_f] \cdot \beta_{p2}}{\text{요인 2} \quad \text{민감도 2}} + \dots + \frac{[E(R_{Fn}) - R_f] \cdot \beta_{pn}}{\text{요인 n} \quad \text{민감도 n}} + \varepsilon$$

요인(factor)으로 제시되는 것들

→ GDP 성장률, 인플레이션, 금리, 유가, 주식시장, .....

이론적으로 끌리지만... ‘상관관계’가 있어도 ‘인과관계(예측력)’은 없다!

## 4. 통계 해석을 통한 재무 이론 이해 심화

# 3-factor 모델은 예측력이 있다!

## Fama, French의 3-factor 모델

$$R_{it} - R_{ft} = \underbrace{\alpha_{it} + \beta_1(R_{Mt} - R_{ft})}_{\text{CAPM}} + \underbrace{\beta_2 SMB_t}_{\text{Size premium}} + \underbrace{\beta_3 HML_t}_{\text{Value premium}} + \varepsilon_{it}$$

- SMB (Small Minus Big) : 중소형주(저 시가총액) 프리미엄
- HML (High Minus Low) : 가치주 (고 B/P) 프리미엄

➡ 기초값을 적용하여... ‘예측력’ 있는 ‘인과관계’가 있다!

예측력 있는 모델 :  $y_{\text{다음기값}} = a + bx_{\text{이전기값}}$

4. 통계 해석을 통한 재무 이론 이해 심화

# 신이 나에게 내려준 예제주제

매주 로또 당첨 방송 30분 전에  
나에게는 당첨번호 6개가 보인다

- 당첨방송 : 토요일 20:45
- 판매마감 : 토요일 20:00



## 핵심요약

- 통계적 가설검정은 흑백 결정을 하지만, 아는 것과 모르는 것의 과학적 확인 과정으로, 목표에 적합한 의사결정이 요구됨
- 회귀분석은 상관관계를 분석하는 대표적 통계 분석 기법으로, 데이터 분석 도구를 활용하거나, LINEST 함수를 사용할 수 있음
- 금융 데이터를 통한 분석은 표본에 의해 모수 추정을 하는 일로서, 대부분의 재무/경제 모델은 통계적 회귀분석 모형이므로, 확률/통계적 인사이트가 필수
- 상관관계가 있다고 해서, 인과관계 또는 예측력이 있는 것은 아님. 시계열 상에서 선후 관계가 존재해야 인과관계 혹은 예측력이 있는 것임

**THANK YOU**