데이터 분석의 이해_1

허경욱



Contents

01 데이터 분석 기초

- 06 데이터 요약 분석
- 02 데이터 도구 Excel
- 07 데이터 탐색 분석

03 데이터 준비

08 데이터 예측 분석

04 데이터 전처리

09 데이터시각화

05 데이터 모델링

10 다양한 데이터 도구

데이터 가상분석 및 요약분석

- 목표값 찾기
- 시나리오
- 에이터 표
- ◈ 해 찾기
- 에이터 요약 집계 분석을 위한 피벗 테이블 활용
- 관계 설정된 데이터 모델 기반 피벗테이블 요약 분석
- 🦩 피벗테이블 마법사의 활용

학습 목표:

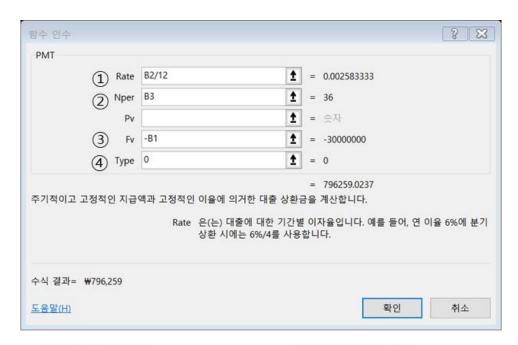
- □ 목표값을 활용해서 1개의 변수의 값을 예측할 수 있다.
- 시나리오를 활용해서 시나리오 요약 보고서를 만들 수 있다.
- □ 1~2개의 입력 변수를 사용하여 목표값의 변화를 데이터 표로 나타낼 수 있다.
- 피벗테이블 기능을 이해하고 기준에 따른 데이터 집계 요약 작업을 할 수 있다.

관계

■ 피벗테이블 기능을 확장하여 분리된 데이터 기능을 적용하여 요약 분석 작업을 할 수 있다.

- 목표값 찾기
 - ■목표로 설정한 값은 알고 있지만, 그 결과를 얻기 위한 입력값을 모르는 경우에 사용
 - ■수식이 있는 셀이 원하는 결과를 얻을 수 있도록 1개의 셀의 데이터를 변경하여 원하는 값을 찾는 경우에 사용
 - ■[목표값 찾기]는 목표값으로부터 특정 값을 거꾸로 찾아내는 방법

- 월 납입액과 목표값 찾기 실습
 - 3년 동안 3천만 원을 모으려고 한다. 연이율이 3.1%라면 매월 얼마를 납입해야 할까?



① 월이율구하기

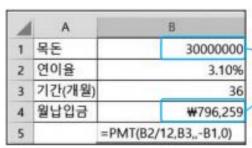
④ 1-주기초: 1~15 0-주기말:16~말

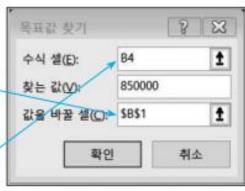
- ② 기간을 개월수로 계산
- ③ 목돈(대출이 아니기때문에 -를 붙임)

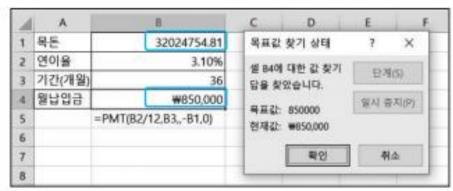
● 월 납입액과 목표값 찾기 실습

② 월 납입액과 목표값 구하기

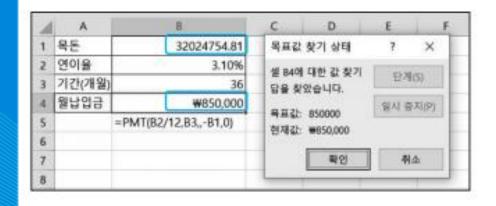
③ 결과 확인

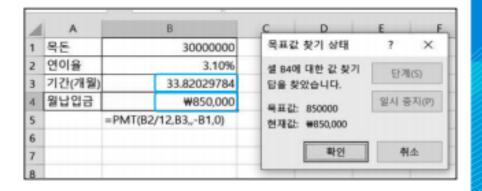






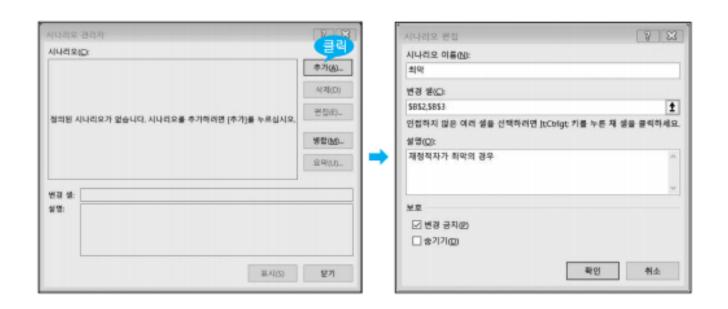
- 월 납입액과 기간 목표값 찾기 실습
 - 3년 동안 3천만 원을 모으려고 한다. 연이율이 3.1%라면 기간과 납입금액은?



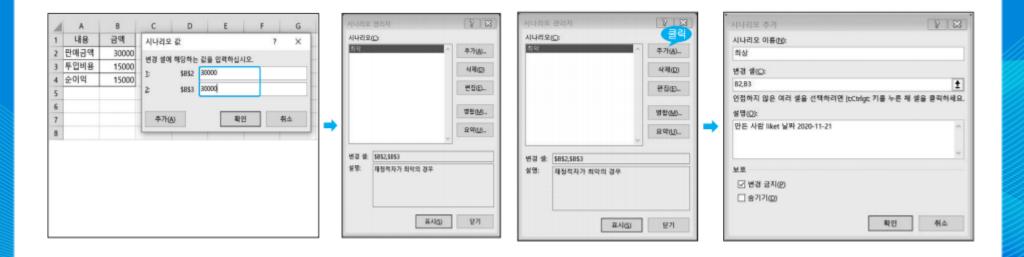


- 시나리오
 - 여러 가지 가상 상황(시나리오)을 고려한 결과값의 비교·분석
 - 각 가상 상황에 따른 변수값을 시나리오로 저장 → 다양한 결과
 - ■시나리오는 최대 32개의 변수값을 적용
 - ■목표로 설정한 값은 알고 있지만, 그 결과를 얻기 위한 입력값을 모르는 경우에 사용

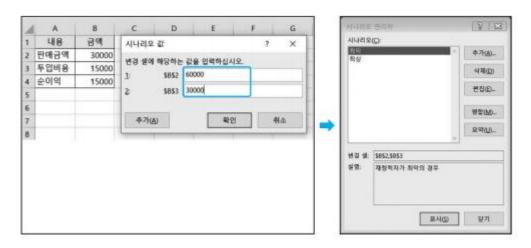
- 시나리오
 - ■[데이터] 탭 → [가상 분석] → 시나리오 관리자
 - ■[시나리오 이름] 지정 → [변경 셀] 지정

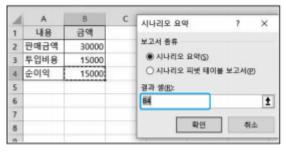


- 시나리오
 - ■[시나리오 값] 입력 → [확인] → [시나리오 관리자] → [추가]



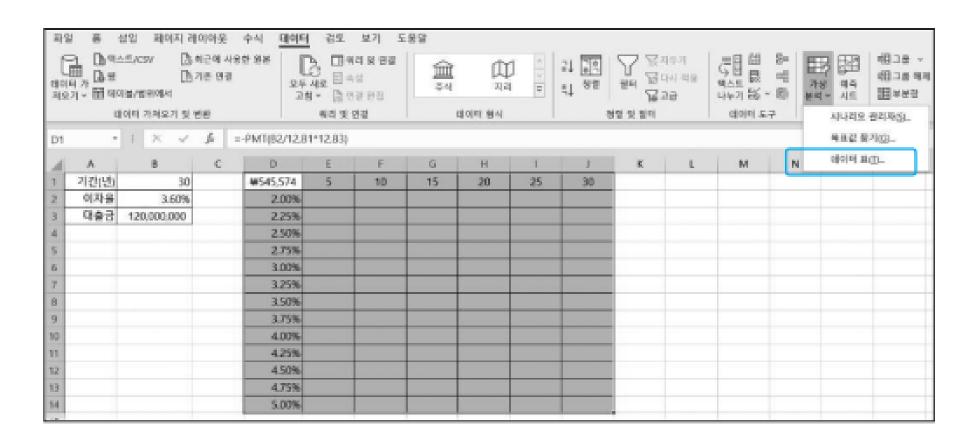
- 시나리오
 - ■[시나리오 값] 입력 → [확인] → [시나리오 관리자] → [요약]
 - ■[시나리오 요약] → [결과 셀] → [확인]



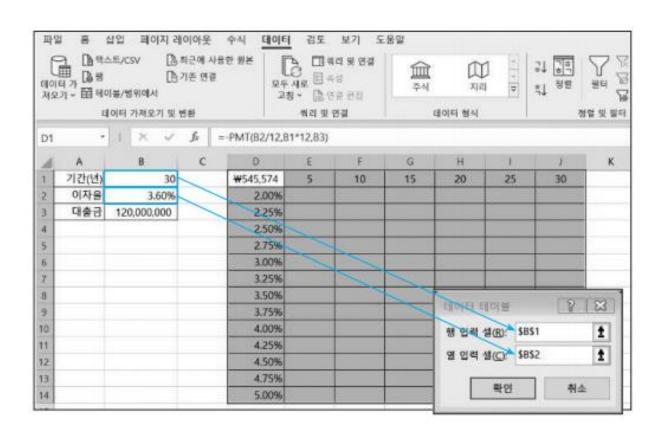


- 데이터 표
 - ■데이터 표는 일부 셀의 값을 변경하고 다양한 결과를 얻을 수 있는 셀 범위
 - ■데이터 표를 사용하면 여러 결과를 한 번의 계산
 - ■테스트해야 하는 변수 및 수식 개수에 따라 변수가 하나 또는 두 개인 데이터 표를 만들어 사용
 - 변수의 개수가 3개 이상인 경우 데이터 표로 분석할 수 없음

● 데이터 표



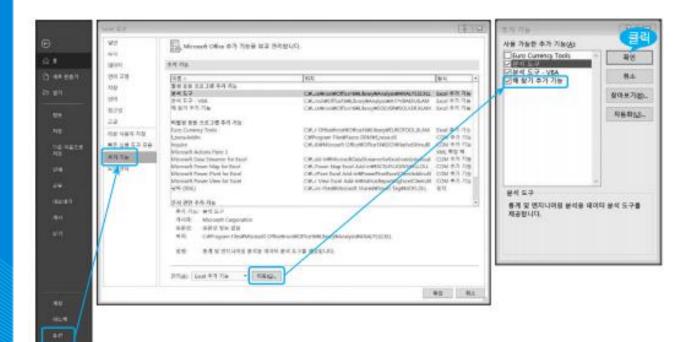
● 데이터 표 설정



● 데이터 표 결과

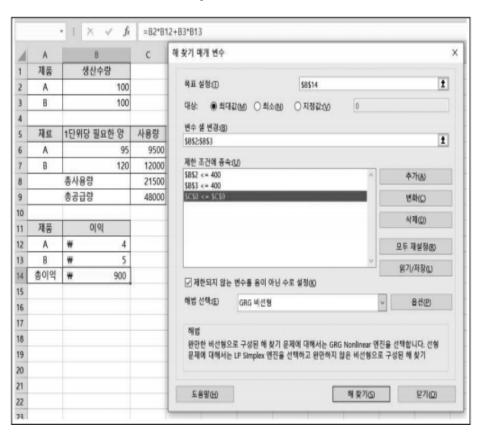
4	A	В	C	D	E	F	G	н	1	J
1	기간(년)	30		₩545,574	5	10	15	20	25	30
2	이자율	3.60%		2.00%	2103331	1104161	772210.4	607060	508625.2	443543.4
3	대출금	120,000,000		2.25%	2116481	1117648	786101.7	621369.9	523356.8	458695.3
4				2.50%	2129683	1131239	800147.1	635883.5	538340.1	474145.1
5				2.75%	2142937	1144932	814346	650599.6	553573	489889.4
6				3.00%	2156243	1158729	828698	665517.1	569053.6	505924.8
7				3.25%	2169600	1172628	843202.5	680634.9	584779.5	522247.6
8				3.50%	2183009	1186630	857859	695951.7	600748.3	538853.6
9				3.75%	2196470	1200735	872666.9	711466	616957.4	555738.7
10				4.00%	2209983	1214942	887625.5	727176.4	633404.2	572898.4
11				4.25%	2223547	1229250	902734.1	743081.4	650085.7	590327.9
12				4.50%	2237162	1243661	917991.9	759179.3	666999	608022.4
13				4.75%	2250829	1258173	933398.3	775468.4	684140.8	625976.8
14				5.00%	2264548	1272786	948952.4	791946.9	701508	644185.9

- 해 찾기
 - 최적화 문제를 해결
 - ■[파일]메뉴 → [옵션] 이동
 - ■[추가 기능] → [관리 상자] → [Excel 추가 기능] → [이동] 클릭
 - ■[사용 가능한 추가 기능] → [해 찾기 추가 기능]선택

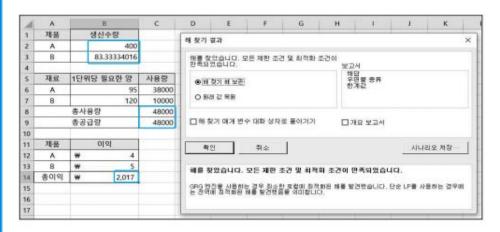


해 찾기 추가 기능을 로드 한 후에는 [데이터] 탭의 분석 그룹에서 [해 찾기] 기능을 사용할 수 있다.

총이익이 최대치가 되기 위해 각 제품의 생산량을 얼마나 해야 하는지 결정하고 자 한다. 재료의 총사용량은 총공급량보다 많으면 안 되고, 제품을 생산하는 시간 으로 인해 A제품, B제품은 400개 이하로 생산을 제한한다



- ① 목적은 최대화 또는 최소화할 '무엇'이고, 최적의 결과를 찾기 위해 목적 함수를 사용한다. 여기서 총이익은 A 제품의 총이익과 B 제품의 총이익의 합계이다.
- ② 목적을 달성하기 위해 변경하는 값인 결정 변수를 찾아 '값을 바꿀 셀 항목'에 입력한다. 여 기서 결정변수는 A 제품과 B 제품의 생산 수량이다.
- ③ 제한 조건을 추가한다. 총사용량은 총공급량보다 많으면 안 되고, 제품을 생산하는 시간으로 인해 A제품, B 제품은 각각 400개 이하로 생산하는 것을 추가한다.
- ④ [해 찾기]를 클릭한다.



해 찾기 결과 A 제품은 400개, B 제품은 약 83개를 생산하면 총이익 2,017 원을 기대할 수 있다는 사실을 알 수 있다.



피벗테이블

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

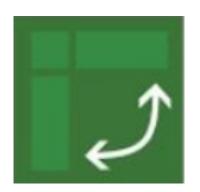




기준열과 계산열

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



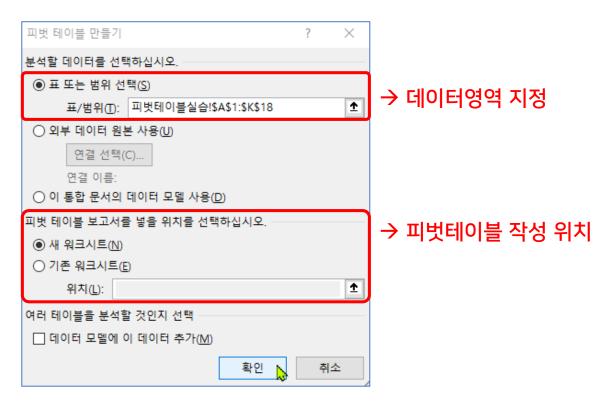


● 피벗테이블 작성을 위한 데이터 확인

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

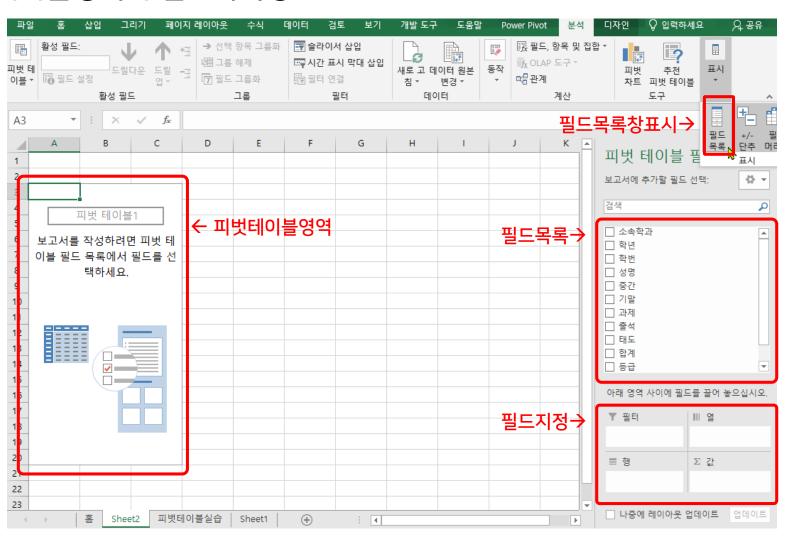
- 범주형 필드 → 기준 필드로 지정 가능
- ■개별 데이터 필드 → 계산 필드로 지정 가능

● [삽입]메뉴탭 → [피벗테이블]메뉴

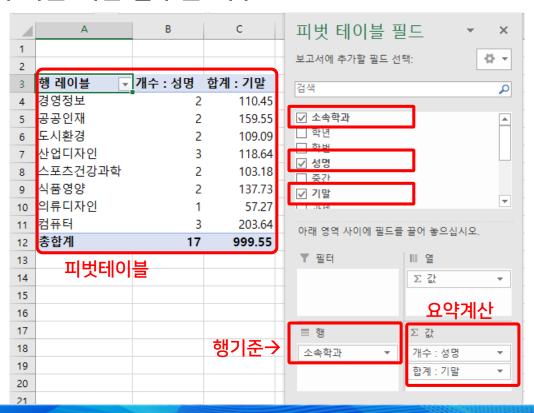


■데이터 시트 왼쪽에 새 워크시트가 생성되고 피벗테이블 영역이 설정됨

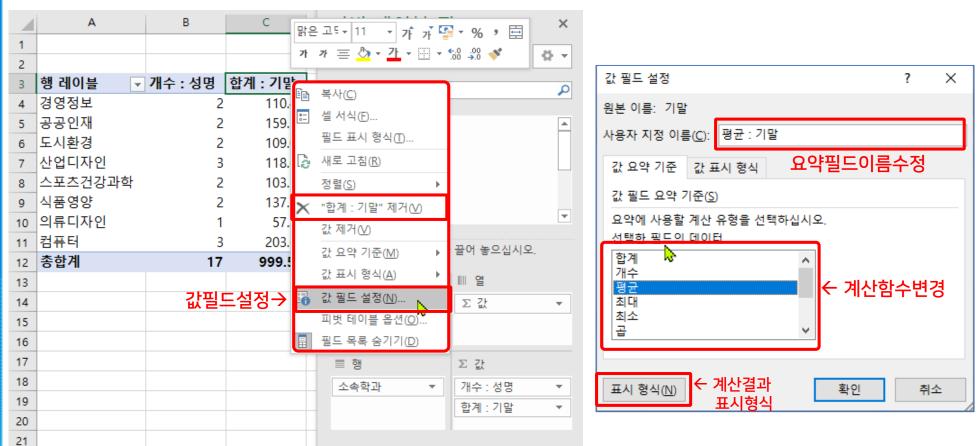
■ 피벗테이블영역과 필드목록창



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

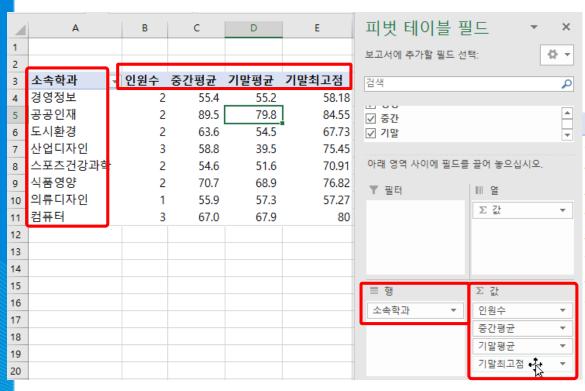


● 계산 필드 편집 → 필드 선택 후 [값 필드 설정]메뉴



■계산 필드를 제거: 단축메뉴 또는 필드이름 마우스드래그드랍으로 필드 삭제

● 소속학과별 인원수와 기말성적평균 요약집계 결과



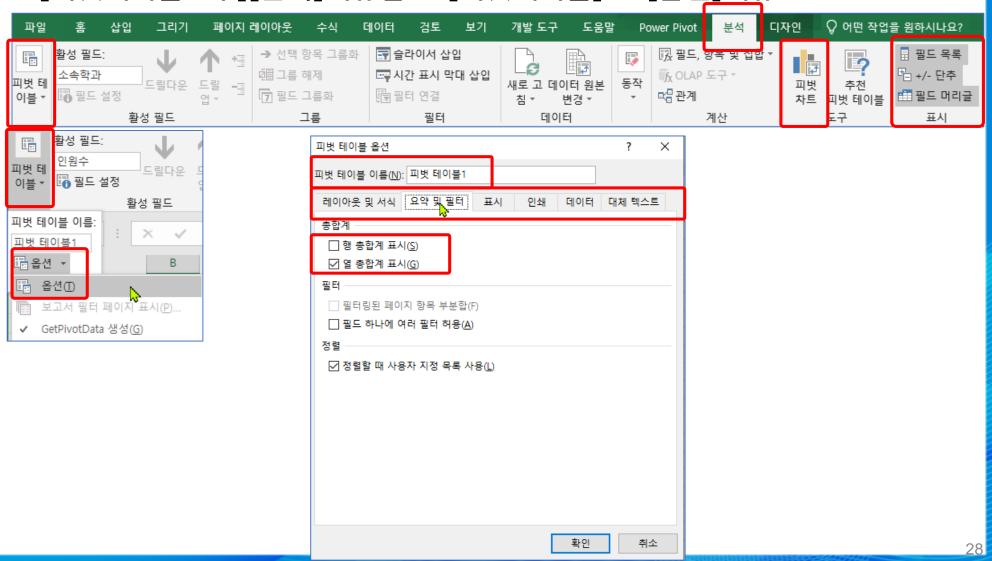
구성된 데이터 중 소속학과별 인원수, 중간/기말 평균과 기말 최대값 데이터 파악

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

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

피벗테이블 편집

● [피벗테이블도구][분석]메뉴탭 → [피벗테이블] → [옵션]메뉴



피벗테이블 편집

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

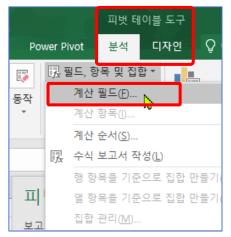


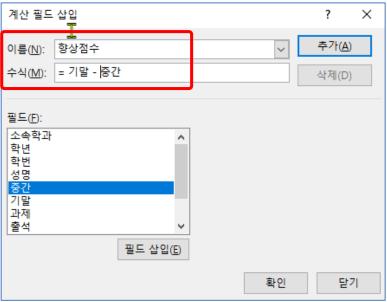
계산필드 추가하기

● [피벗테이블도구][분석]메뉴탭 → [계산]그룹 → [필드,항목및집합] →

[계산필드]

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





행 레이블	▼ 인원수	중간평균	기말평균	향상점수평균←
경영정보	2	55.4	55.2	-0.3
공공인재	2	89.5	79.8	-19.4
도시환경	2	63.6	54.5	-18.1
산업디자인	3	58.8	39.5	₩57.6
스포츠건강과학	<u>}</u> 2	54.6	51.6	-6.0
식품영양	2	70.7	68.9	-3.7
의류디자인	1	55.9	57.3	1.3
컴퓨터	3	67.0	67.9	2.6

·계산필드 추가 후 함수를 평균으로 수정

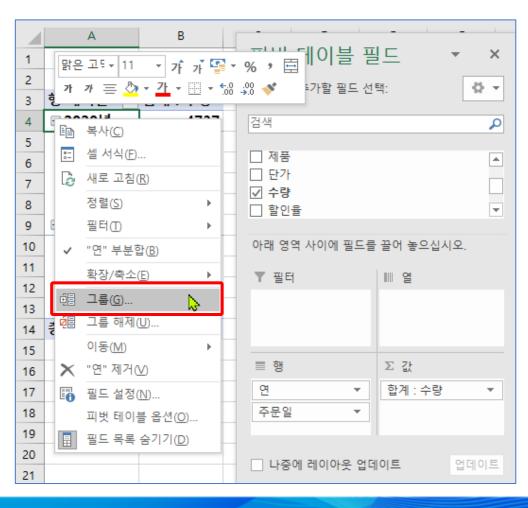
※계산 필드 완전 삭제 : [계산필드]메뉴 → 목록버튼 → [삭제]버튼

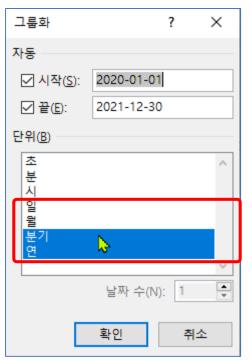


● 기준 필드 데이터가 수치데이터이면 다양한 요약 집계를 위하여 그룹 설정 가능

4	Α	В	С	D	Е	F	G	Н	I	J
1 7	ㅐ래번호	고객	담당	주문일	분류	제품	단가	수량	할인율	판매액
2	N-0705	S&C무역 ㈜	오서윤	R020-01-01	복사기	컬러레이저복사기 XI-3200	1,176,000	3	15%	2,998,800
3	N-0705	S&C무역 ㈜	오서윤	2020-01-01	바코드스캐너	바코드 Z-350	48,300	3	0%	144,900
4	N-0705	S&C무역 ㈜	오서윤	2020-01-01	팩스	잉크젯팩시밀리 FX-1050	47,400	3	0%	142,200
5	N-0706	드림씨푸드 ㈜	박현우	2020-01-04	복사용지	프리미엄복사지A4 2500매	17,800	9	0%	160,200
6	N-0706	드림씨푸드 ㈜	박현우	2020-01-04	바코드스캐너	바코드 BCD-100 Plus	86,500	7	0%	605,500
7	N-0707	자이언트무역 ㈜	정시우	2020-01-05	복사용지	고급복사지A4 500매	3,500	2	0%	7,000
8	N-0707	자이언트무역 ㈜	정시우	2020-01-05	바코드스캐너	바코드 Z-350	46,300	7	0%	324,100
9	N-0707	자이언트무역 ㈜	정시우	2020-01-05	바코드스캐너	바코드 BCD-100 Plus	104,500	8	0%	836,000
10	N-0708	진왕통상 ㈜	오서윤	2020-01-05	복합기	잉크젯복합기 AP-3300	79,800	1	0%	79,800
44	N 0700	지와토사 🖂	оио	2020 01 05	보하기	이그제보하기 AD 2200	00 200	0	00/	714 400
2681	N-1749	동화상사 ㈜	오서윤	2021-12-29	복사기	흑백레이저복사기 TLE-9000	896,500	2	5%	1,703,350
2682	N-1750	한성트레이딩 ㈜	이서연	2021-12-29	복합기	레이저복합기 L950	504,400	4	10%	1,815,840
2683	N-1750	한성트레이딩 ㈜	이서연	2021-12-29	팩스	잉크젯팩시밀리 FX-1050	51,000	8	5%	387,600
2684	N-1751	신화식품 ㈜	박현우	2021-12-29	출퇴근기록기	지문인식 FPIN-1000+	130,000	8	5%	988,000
2685	N-1751	신화식품 ㈜	박현우	2021-12-29	복합기	레이저복합기 L500	286,200	5	5%	1,359,450
2686	N-1752	용도교역 ㈜	박현우	2021-12-29	복합기	무한잉크젯복합기 AP-3300W	101,000	4	0%	404,000
2687	N-1752	용도교역 ㈜	박현우	2021-12-29	복사용지	프리미엄복사지A4 5000매	29,700	8	0%	237,600
2688	N-1753	노을백화점 ㈜	유준혁	2021-12-29	문서세단기	오피스 Z-05C	120,500	10	5%	1,144,750
2689	N-1753	노을백화점 ㈜	유준혁	2021-12-29	문서세단기	문서세단기 SCUT-1500B	686,800	1	5%	652,460
2690	N-1753	노을백화점 ㈜	유준혁	2021-12-29	출퇴근기록기	지문인식 FPIN-2000F	172,800	8	5%	1,313,280

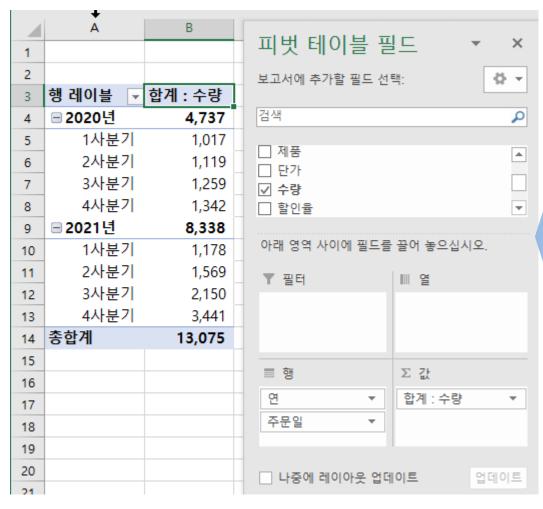
- 날짜 데이터의 그룹 설정
- 피벗테이블 작성 후 기준 열 데이터를 선택한 후 단축메뉴 [그룹]메뉴 실행





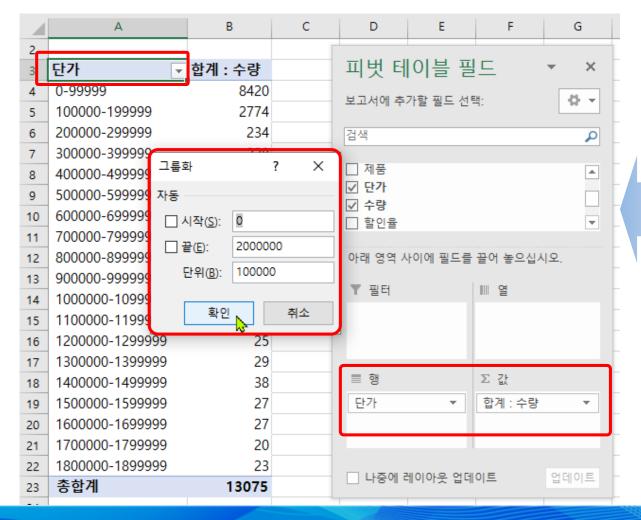
※ 날짜데이터의 연/분기/월/일 단위로 그룹 설정 가능 클릭하여 설정, 다시 클릭하면 설정 해제

● 날짜데이터 기준의 그룹 설정 결과



연도와 분기별 '수량'데이터의 합계를 요약 집계하여 파악함

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

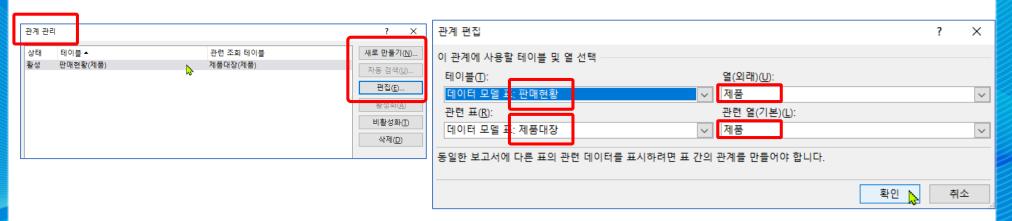


'단가'데이터를 십만 단위로 구분하여 해당 단가 그룹의 '수량' 데이터 합계 파악

관계 설정된 표들의 요약 집계

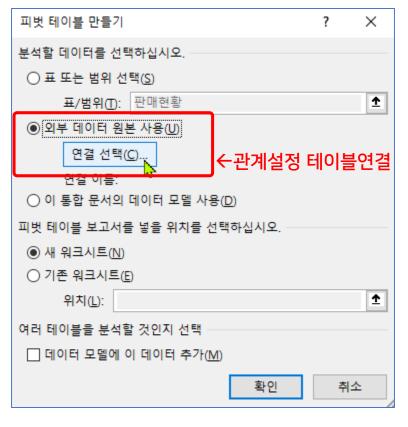
- 데이터 모델의 피벗테이블 작성
- 관계 설정된 두 테이블을 기반으로 요약 집계
 - ■분리된 2개 이상의 표
 - ■[데이터]메뉴탭 → [데이터도구]그룹 → [관계]메뉴

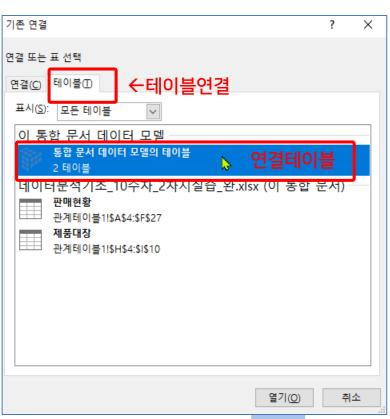




관계 설정된 표들의 요약 집계

● [삽입]메뉴탭 → [피벗테이블]메뉴

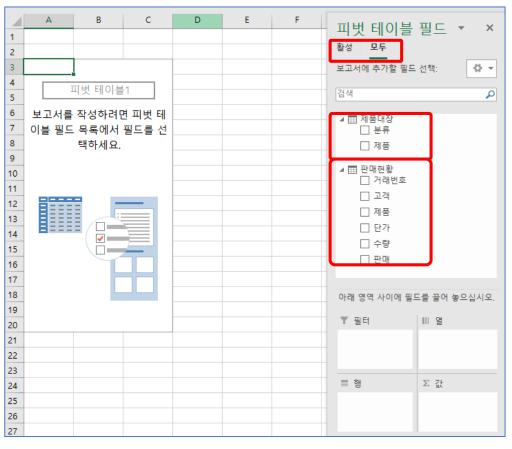


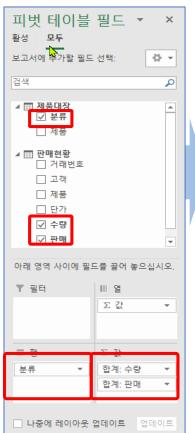


※ 연결이름: ThisWorkbookDataModel ● 외부 데이터 원본 사용(U) 면결 선택(C)... 연결 이름: ThisWorkbookDataModel

관계 설정된 표들의 요약 집계

● 관계 데이터 모델의 피벗테이블 필드 목록





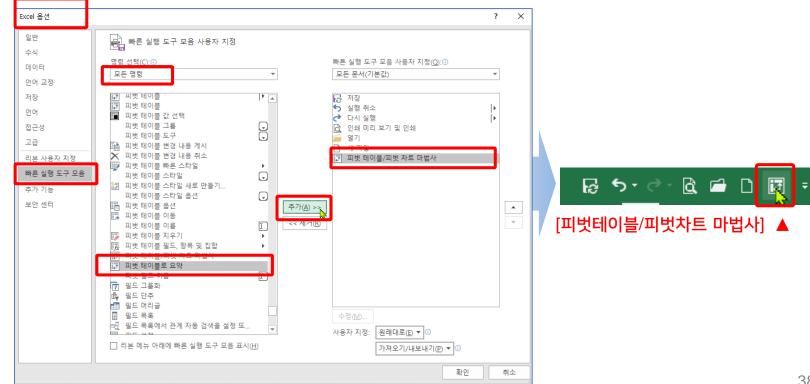


■분리된 테이블의 필드를 기준으로 원하는 필드 요약 집계 계산

- 집계표 형식의 분리된 표를 하나로 취합하기 위한 피벗테이블마법사 기능 제공
- 피벗테이블마법사 메뉴
 - 기본 메뉴 탭에 표시 없음 → 빠른실행도구모음에 등록하여 사용가능

■[파일] → [옵션] → [빠른실행도구모음] → [모든명령] → [피벗테이블/피벗차트

마법사]

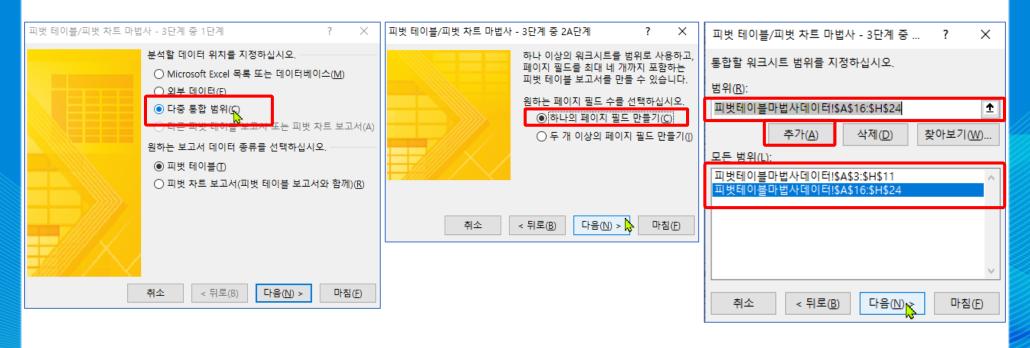


● 피벗테이블 / 피벗차트 마법사를 적용할 수 있는 표

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

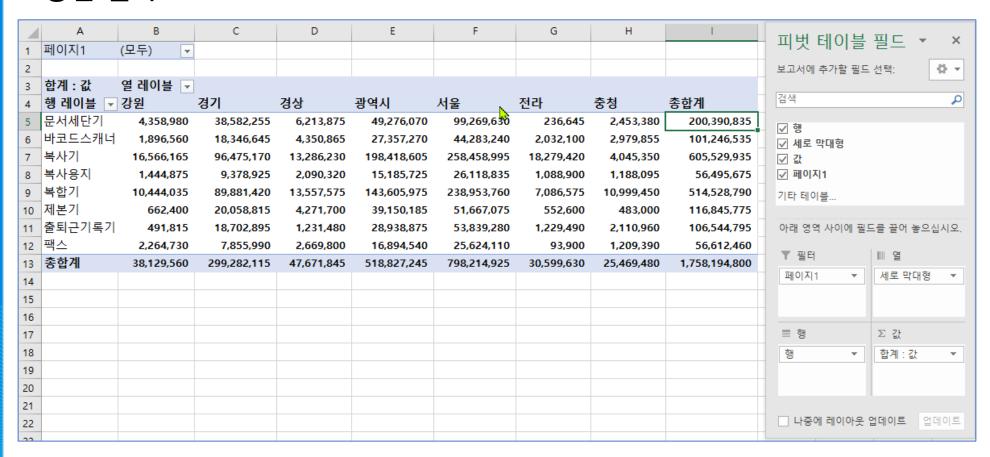
■행제목과 열제목의 순서가 일치하지 않더라도 같은 항목끼리 통합

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



■데이터 범위를 지정시 총합은 제외하고 지정

● 통합 결과



■행과 열 기준 항목이 같은 데이터의 합계 요약 계산되어 표시



□ 실습할 Excel 파일 열기 → [데이터분석의이해실습_1.xlsx]

