

Chapter
01

통계학의 이해

Introduction of Statistics

목 차

01

통계학을 공부하는 이유

02

통계분석과 그 한계

03

Excel의 통계 기능과 데이터 입력

01 통계학을 공부하는 이유

:: **Keywords** 통계학을 공부하는 이유 | 통계학의 정의와 목적



인문/사회과학에서 통계학을 공부하는 이유

■ 인문/사회과학에서 통계학을 공부하는 이유

- 통계학은 이공 계열 학과에서는 주로 교양 과목
→ 사회과학을 포함해 대부분의 문과 계열 학과에서는 통계학을 필수과목
: 필수과목 → 전공 공부를 하는 데 반드시 필요하다는 의미
- 통계학이 어렵게 느껴지는 몇 가지 이유
→ 가장 큰 이유는 범용성의 문제
범용성은 '어떤 제약 조건도 없이 여러 분야에 두루 사용되는 것'
과연 통계는 우리 생활에서 범용성을 가지고 있을까?

인문/사회과학에서 통계학을 공부하는 이유

■ 통계학을 적용한 자료는 신뢰성을 가진다.

- 통계는 실생활에서 광범위하게 활용된다.
- 우리가 접하는 대부분의 정보는 통계학을 통해 가공되어 전달된다.

국내 최고의 미각소유자 vs. 신뢰할 만한 기관

→ 통계는 사회에서 발생할 수 있는 다양한 상황에서
신뢰할 수 있는 자료를 가공해 내고 이를 활용

인문/사회과학에서 통계학을 공부하는 이유

■ 통계는 의사결정에 필요한 과학적 근거 자료를 제시한다.

- 인문/사회과학의 연구(조사)나 실험결과는 다양한 결과를 도출
- 일상생활이나 현상 등을 수치화하기 위한 기준을 토대로
조사와 분석에서의 과학적 접근

획기적 신제품의 홍보방법은 어떻게 해야 할까?

→ 비용 대비 최대의 효과를 얻을 수 있는 의사결정 가능

인문/사회과학에서 통계학을 공부하는 이유

■ 통계는 현상을 분석하여 실증자료를 제시한다.

- 현상을 분석하여 문제의 해결을 위한 다양한 원인을 찾을 수 있도록 자료 제공

(이론) 인간의 소비행동은 자신의 불만족을 만족으로 바꾸려는 행동

Ex. 소비자의 스마트폰 선택 기준은? (디자인, 편의성, 유용성)

→ 3가지 원인에 대해 선택과 집중을 해야 한다면 어디에?

통계학의 정의와 목적

■ 통계학의 정의

통계학(statistics)은 수량적인 비교를 기초로 많은 사실을 관찰하고 처리하는 방법을 연구하는 학문

일반적으로 수집되는 데이터가 조사자, 시기, 방법, 목적 등에 따라 다르게 나타나는 불균형적인 데이터이지만, 통계학은 이 안에서 의미를 찾아내고, 실생활에서 적용 가능한 유용성을 찾아내 이를 수치로 표현할 수 있도록 한다.

- **기술통계(descriptive statistics) :**

표본에 대한 분석 결과의 각종 수치들을 활용하여 집단의 특성을 설명

- **추론통계(inference statistics) :** 표본을 활용하여 모집단의 특성을 나타내는 것

통계학의 정의와 목적

■ 통계학의 목적

- 의사결정

의사결정은 많은 정보를 지각하고 평가하여 하나를 선택하는 것

∴ 정보와 반응 사이의 다대일 대응으로 나타나므로

→ 여러 가지 대안 가운데 하나를 선택할 때, 기초자료를 제공

- 불확실성의 해소

의사결정을 하게 되면 그 결과가 정확한 것이라 할 수 있는가의 문제

∴ 빅데이터의 개념을 들여와 불확실성을 해소하려는 노력

→ 정보수집의 어려움, 시장의 변화와 대응의 어려움에 대한 극복

통계학의 정의와 목적

■ 통계학의 목적

- 요약

다양한 데이터를 신속히 이해할 수 있도록 다양한 형태로 표현

∴ 불확실성의 감소를 위해

→ 반복되어 생산되는 데이터를 정리된 보고서로 표현하여
불확실성이 낮은 상황의 의사결정이 가능하도록 함.

- 연관성 파악

요약된 보고서에서 주요한 항목들 간의 연관성을 파악한 경쟁우위의 확보

∴ 의사결정권자에게 항목 간 연관성을 제시하여 미래의 계획을 지원

→ 다양한 자료는 의사결정에 있어 세부적 판단에도 기여

통계학의 정의와 목적

■ 통계학의 목적

- 예측

인과관계 파악을 통해 패턴을 찾아내고 이러한 패턴을 통해 추세를 판단

∴ 다양한 변수의 대입과 삭제를 통해 예측 가능

→ 다양한 계량 기법과 여러 변수들을 활용하여

최소의 비용으로 최대의 수익을 얻을 수 있는 조합 확인

02 통계분석과 그 한계

:: **Keywords** 통계분석의 의미와 과정 | 통계분석의 한계



통계분석의 과정

■ 통계분석의 과정

통계분석은 표본을 통해 의미 있는 자료를 추출하고, 이를 기반으로 의사결정, 불확실성의 해소, 요약, 연관성 파악, 예측 등의 결과로 이어지도록 하는 일련의 과정

수집 → 정제 → 추정 → 검정의 과정을 따른다

통계분석의 과정

■ 통계분석의 과정

(자료의) 수집

1. 조사 과정에서 자연스럽게 수집
2. 조사자가 특정 목적에 맞는 자료를 얻기 위해 설계한 수집 도구를 이용

보통 자료가 많을수록 통계분석에서 더 좋은 결과를 얻을 것이라고 생각하지만, 조사의 목적에 맞는 적절한 자료를 수집해서 최적의 통계 방법으로 분석한 결과가 가장 정확도가 높음

- **1차 자료** : 조사자가 직접 수집한 자료, 조사 목적에 맞게 측정 도구를 설계하고 개발할 수 있기에, 목적에 가장 부합되는 자료들을 수집할 수 있으며, 설문지가 가장 많이 활용됨(우편, 전화, 전자 질의, 직접 인터뷰 등의 방법으로 진행)
- **2차 자료** : 조사 목적에 맞아 활용할 수 있는 자료, 도서관이나 학술정보원 등에서 수집할 수 있으며 정부간행물, 연구보고서, 사내자료 등도 사용됨
(자료의 원래 목적이 조사자의 목적과 다를 수 있으므로 자료의 활용 가능성, 적합성, 신뢰성 등에 대한 사전 평가가 매우 중요)

통계분석의 과정

■ 통계분석의 과정

(자료의) 정제

1. 분석에 적합한 자료를 선별
2. 적합하지 않은 자료는 삭제

정확한 자료만을 수집했다면 정제 과정이 필요 없겠으나,
대체로 이는 불가능하므로 사후적으로 세심하게 주의가 요구됨

통계분석의 과정

■ 통계분석의 과정

추정

통계분석의 진행하는 것 → 결국 모수를 추정하는 것

표본을 대상으로 분석하는 것이므로, 표본의 특성을 설명하는 통계량을 통해 모집단의 모수를 추정하게 되므로 좋은 추정 결과를 위해서는 양질의 자료가 필요

참고 모집단, 모수, 표본, 통계량

모집단, 모수, 표본, 통계량에 대해서는 2장에서 자세하게 다루겠지만, 우선 다음 예를 통해 간단히 살펴보자. 선거 전에 선거구민의 투표 성향을 파악하고자 한다. 원래는 선거구민 전체(모집단)에 대한 투표율(모수)을 알아낸 후 선거구민의 투표 성향을 파악해야 한다. 그러나 선거구민 전체에 대한 정보를 알기 힘들기 때문에, 모집단에서 일부 추출한 표본(조사 대상 n 명)으로부터 구한 표본의 투표율(통계량)을 통해 선거구민의 투표 성향을 추정한다. 이는 투표자의 성향을 파악하기 위한 일반적인 조사 방법으로 모집단, 모수, 표본, 통계량에 대한 의미를 쉽게 이해할 수 있다.

통계분석의 과정

■ 통계분석의 과정

검정

통계조사의 목적 → 주장이나 믿어지는 사실이 실제로 옳은지 아닌지를 확인

→ 수립된 가설이 유의미하고 타당성을 가지는지를 통계적으로 확인하는 과정

확인을 통해 가설의 기각과 채택 여부를 판단

Note 추정과 예측

추정(statistical estimation)은 앞서 설명한 바와 같이 표본으로부터 얻은 통계량으로 모수를 추정하는 것이다. 즉, 모집단 전체를 분석할 수 없으므로 모집단으로부터 표본을 구성하고, 그 표본을 조사하여 얻은 수치로 모수를 추정하는 것이다. 추정과 혼동하기 쉬운 용어가 예측이다.

예측(statistical prediction)은 추정을 반복하여 얻은 결과들을 기준으로 일정한 패턴을 찾아낸 후, 미래에 활용할 수 있는 특정한 모수를 가늠하는 것이다. 일기예보를 보면 더욱 정확하게 이해할 수 있다. [그림 1-1]에 표시된 강수 확률은 과거 자료를 기준으로 현재의 날씨를 분석하고, 일정한 패턴을 매년 측정하여 내일(미래) 날씨를 예측한 것이다.

일시 도시	오늘오전 (8.04.)	오늘오후 (8.04.)	내일오전 (8.05.)	내일오후 (8.05.)
강화	 구름많음 기온 24℃ 강수확률 20%	 구름많음 기온 29℃ 강수확률 20%	 흐리고 가끔 비 기온 24℃ 강수확률 60%	 구름많음 기온 29℃ 강수확률 20%
동두천	 구름많음 기온 23℃ 강수확률 20%	 구름많음 기온 31℃ 강수확률 20%	 흐리고 가끔 비 기온 24℃ 강수확률 60%	 구름많고 한때 비 기온 30℃ 강수확률 60%
서울	 구름많음 기온 25℃ 강수확률 20%	 구름많음 기온 32℃ 강수확률 20%	 흐리고 가끔 비 기온 25℃ 강수확률 60%	 구름많음 기온 31℃ 강수확률 20%
수원	 구름많음 기온 24℃ 강수확률 20%	 구름많음 기온 32℃ 강수확률 20%	 구름많음 기온 24℃ 강수확률 20%	 구름많음 기온 31℃ 강수확률 20%
양평	 구름많음 기온 24℃ 강수확률 20%	 구름많음 기온 33℃ 강수확률 20%	 구름많음 기온 24℃ 강수확률 20%	 구름많음 기온 33℃ 강수확률 20%

[그림 1-1] 일기 예보의 강수 확률 예측

출처 : 네이버 날씨

통계분석의 한계

■ 확률이 없으면 의미가 없다

통계분석의 결과는 항상 확률과 연관되어 표현된다.

“미국 유권자 50%, 앞으로도 트럼프 지지 안 해”

올 11월 미국 대선이 3개월여 앞으로 다가왔지만 아직도 유권자의 13%가 지지 후보를 결정하지 못한 것으로 조사됐다.

월스트리트저널(WSJ)은 25일(현지시간) 등록 유권자 900명을 대상으로 시행한 여론조사(조사기간 7월 9~12일) 오차범위 $\pm 3.27\%$ 포인트에서 이 같은 결과가 나왔다고 보도했다.

부동층으로 분류되는 13%는 도널드 트럼프 대통령이나 조 바이든 전 부통령 중 누구도 지지하지 않거나, 지지하는 후보가 있지만 다른 후보로 바꿀 수 있다고 답했다.

선호 후보에 대한 질문에 대해 부동층 중 17%는 트럼프 대통령을, 14%는 바이든 전 부통령을 꼽았다. 53%는 선호하는 후보도 없다고 답했다.

부동층은 스스로를 민주당이나 무소속보다는 공화당 지지층에 가깝다고 규정했다.

WSJ은 부동층 규모가 크지는 않지만, 앞선 지지율 여론조사에서 11%포인트 차이로 바이든 전 부통령에 밀리고 있는 트럼프 대통령으로서는 이들을 흡수하는 것이 지지율 격차를 좁히는 한 가지 방법이라고 분석했다.

다만 2016년 대선 당시 부동층의 절반에 가까운 48%가 투표를 하지 않았거나, 제3의 후보에게 투표했기 때문에 부동층 표심의 행방에는 여전히 의문이 따른다고 WSJ은 전했다.

한편 이번 여론조사에서 응답자의 50%는 도널드 트럼프 대통령을, 37%는 조 바이든 전 부통령을 앞으로도 지지할 의사가 전혀 없다고 답했다.

지지율을 보면, 부동층 중

트럼프 17%, 바이든 14%로 사실상 동률이라고 분석하고 있다.

이때 오차 범위를 $\pm 3.27\%$ 로 표시하여 100% 맞는 것이 아님을 언급하고 있는 것과 같이 결과는 항상 확률과 연관되어 표현된다.

통계분석의 한계

■ 항상 틀릴 가능성을 내포한다

신뢰수준 100% → 신뢰구간에 해당하는 값 $-\infty \sim +\infty$

이런 무의미한 결과를 배제하기 위해 통계 결과의 범위가 줄어듦.

범위가 줄면 줄수록 신뢰구간은 100%에서 점점 더 멀어지고,

→ 보기에 명쾌하고 정확한 결과가 도출될수록 그 결과가 틀릴 가능성은 점점 상승

통계분석의 한계

■ 결론은 항상 추론

모집단을 대상으로 조사하지 않기 때문에 통계는 항상 모집단에 대한 추론이 된다.
추론이기 때문에 확률을 동반하여 그 부족함을 채움

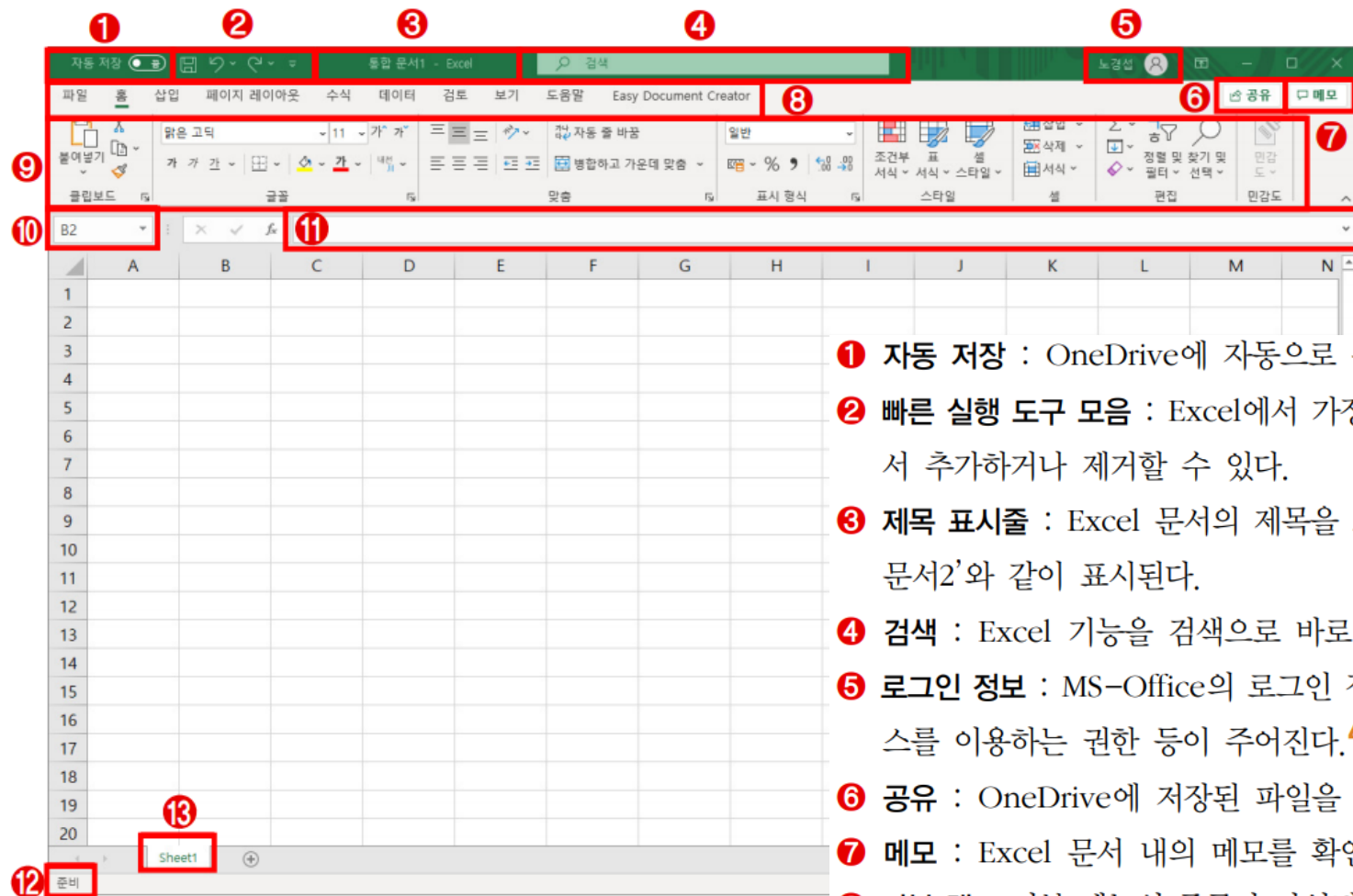
일부에서는 이러한 한계를 '부정확한 것의 일반화'라고 하지만
모집단에 대한 조사가 불가능하다고 해서 포기하는 것 보다는 적극적으로 한계를
극복하려는 시도와 오류를 배제하려는 노력이 더 많은 학문적 결과와 인문/사회과
학의 발전을 가져오게 될 것임

03 Excel의 통계기능과 데이터 입력

:: **Keywords** Excel 화면과 워크시트의 구성 | Excel의 통계기능 | 데이터 입력

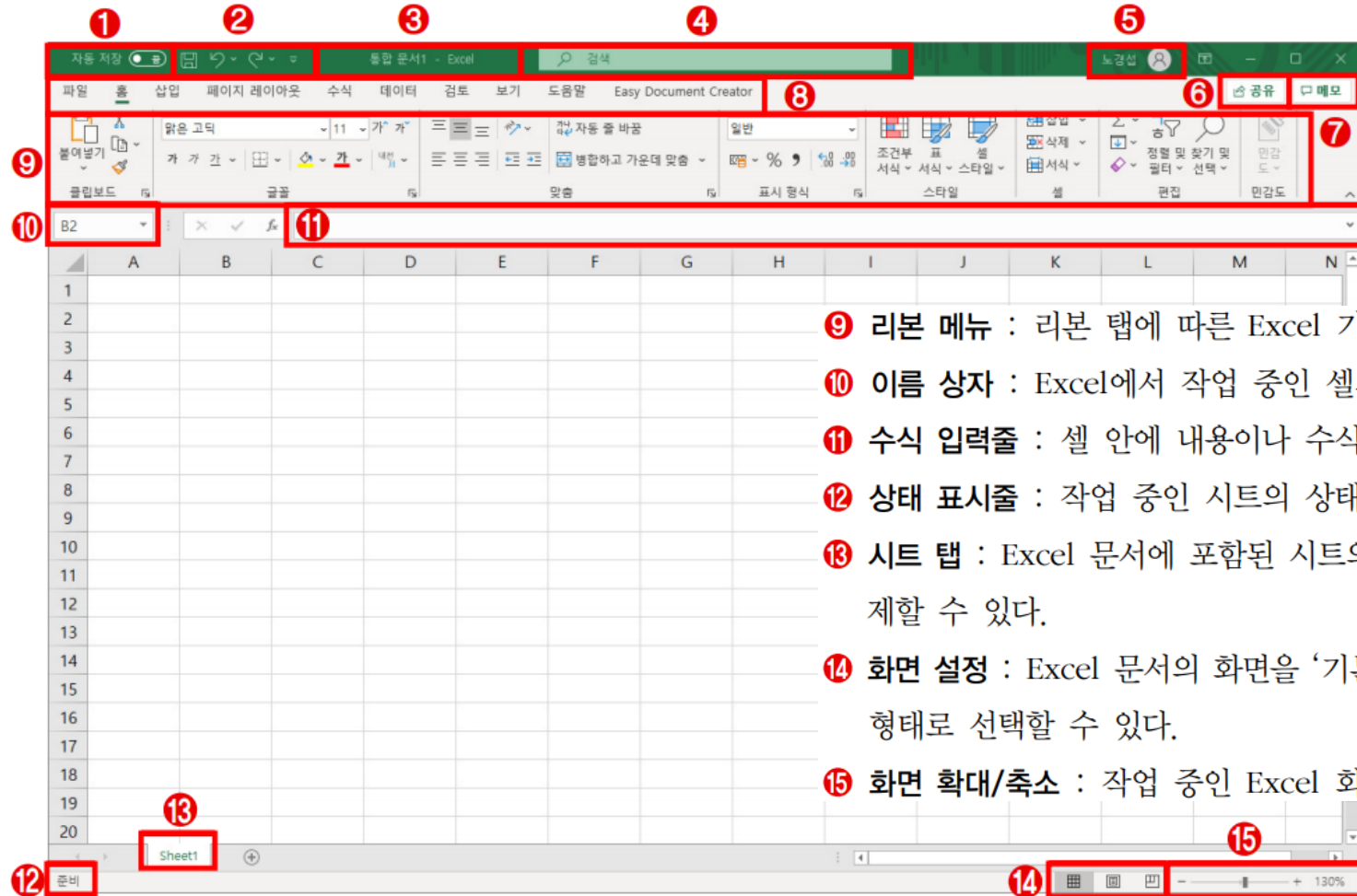


Excel 화면과 워크시트의 구성



- ① 자동 저장 : OneDrive에 자동으로 문서를 저장할 수 있는 옵션이다.
- ② 빠른 실행 도구 모음 : Excel에서 가장 빈번히 쓰이는 기능이 모여 있다. Excel 옵션에서 추가하거나 제거할 수 있다.
- ③ 제목 표시줄 : Excel 문서의 제목을 표시한다. 저장하지 않았다면 '통합 문서1', '통합 문서2'와 같이 표시된다.
- ④ 검색 : Excel 기능을 검색으로 바로 실행할 수 있다.
- ⑤ 로그인 정보 : MS-Office의 로그인 정보를 나타낸다. 로그인하면 MS의 클라우드 서비스를 이용하는 권한 등이 주어진다.⁴
- ⑥ 공유 : OneDrive에 저장된 파일을 공유할 수 있다.
- ⑦ 메모 : Excel 문서 내의 메모를 확인할 수 있다.
- ⑧ 리본 탭 : 리본 메뉴의 종류가 나열되어 있다. 탭 이름을 클릭하면 기능별로 분류된 리본 메뉴가 나타난다.

Excel 화면과 워크시트의 구성



⑨ 리본 메뉴 : 리본 탭에 따른 Excel 기능을 직접 실행할 수 있는 아이콘이 모여 있다.

⑩ 이름 상자 : Excel에서 작업 중인 셀의 이름이나 선택된 영역을 표시한다.

⑪ 수식 입력줄 : 셀 안에 내용이나 수식을 입력할 수 있다.

⑫ 상태 표시줄 : 작업 중인 시트의 상태나 지정된 범위의 통계를 표시한다.

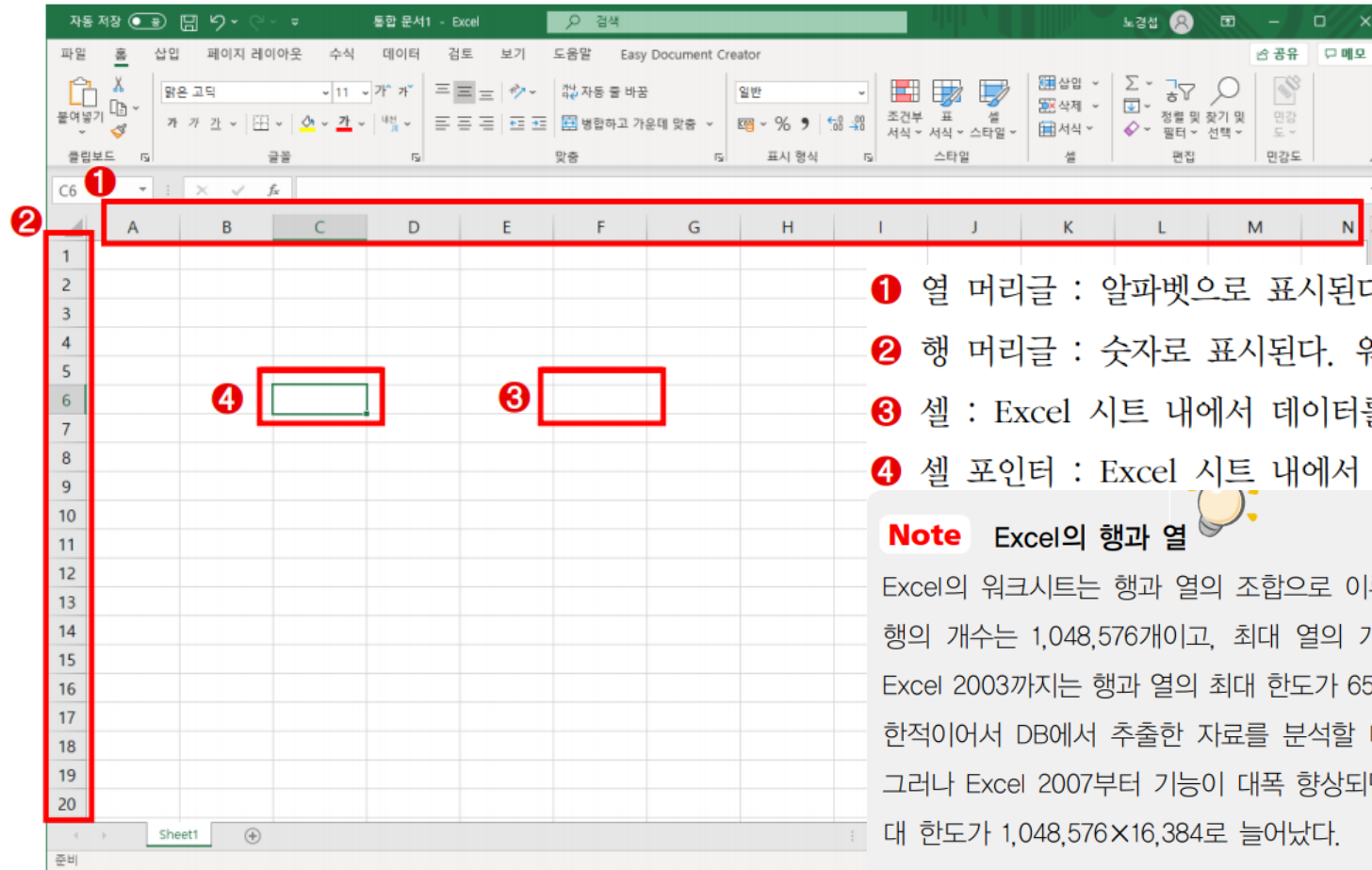
⑬ 시트 탭 : Excel 문서에 포함된 시트의 이름이 표시된다. 시트 이름을 변경, 추가, 삭제할 수 있다.

⑭ 화면 설정 : Excel 문서의 화면을 '기본, 페이지 레이아웃, 페이지 나누기 미리보기'의 형태로 선택할 수 있다.

⑮ 화면 확대/축소 : 작업 중인 Excel 화면의 내용을 키우거나 줄일 수 있다.

Excel 화면과 워크시트의 구성

워크시트의 구성



- ① 열 머리글 : 알파벳으로 표시된다. 워크시트 제일 위에 있는 열의 이름을 말한다.
- ② 행 머리글 : 숫자로 표시된다. 워크시트 제일 왼쪽에 있는 행의 이름을 말한다.
- ③ 셀 : Excel 시트 내에서 데이터를 입력하는 기본 단위다.
- ④ 셀 포인터 : Excel 시트 내에서 데이터가 입력되는 셀이다.

Note Excel의 행과 열

Excel의 워크시트는 행과 열의 조합으로 이루어져 있다. 최대 행의 개수는 1,048,576개이고, 최대 열의 개수는 16,384개다. Excel 2003까지는 행과 열의 최대 한도가 65,536×256으로 제한적이어서 DB에서 추출한 자료를 분석할 때 제약이 있었다. 그러나 Excel 2007부터 기능이 대폭 향상되면서 행과 열의 최대 한도가 1,048,576×16,384로 늘어났다.

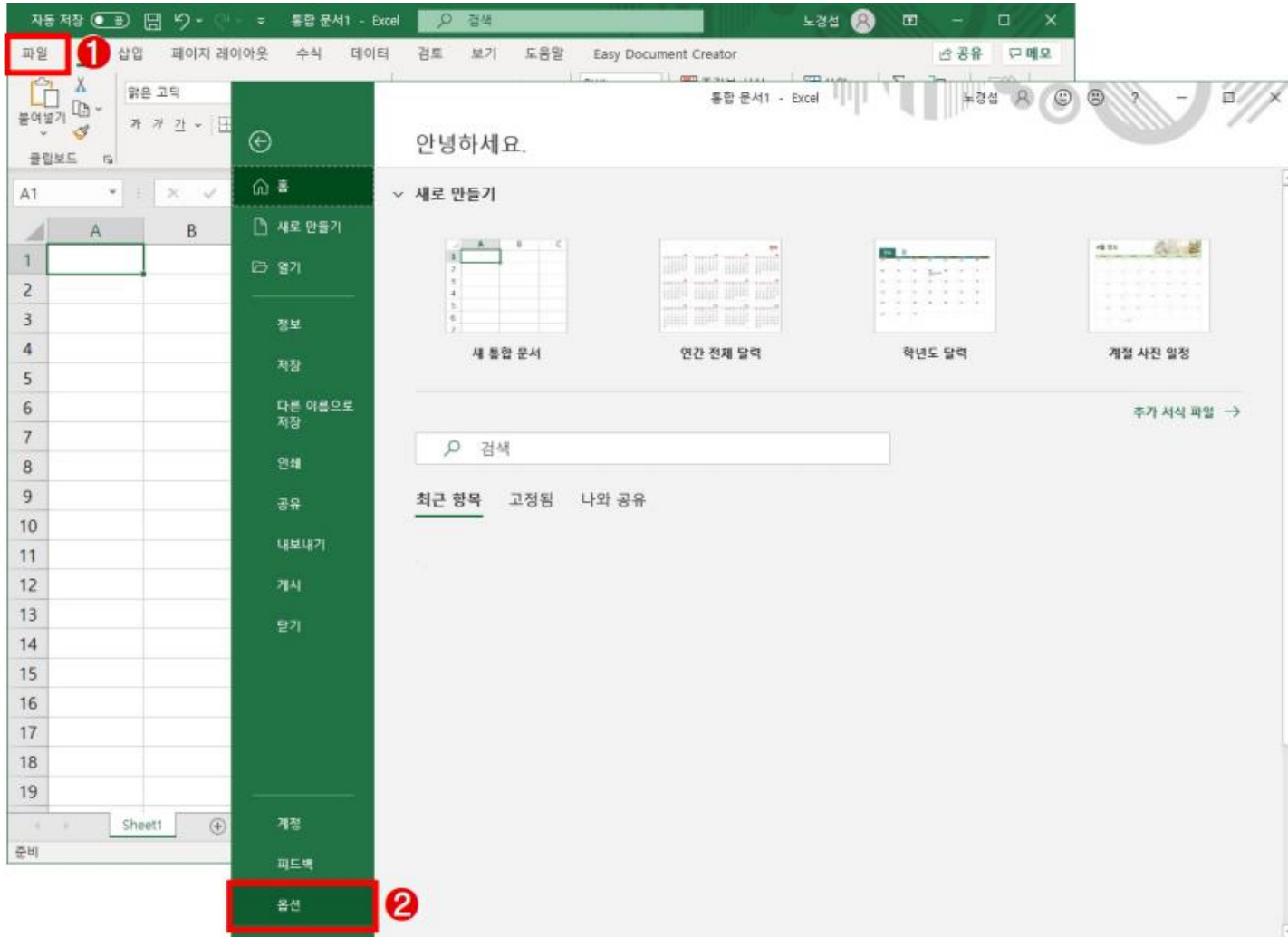
[그림 1-4]는 Microsoft의 Excel에서 제공하는 도움말이다. Excel에서 기본적으로 제공하는 워크시트에 대한 사양과 제한 사항을 보여준다.

도움말	
Excel 사양 및 제한	
워크시트와 통합 문서 사양 및 제한	
기능	8000 행
일 수 있는 통합 문서의 수	사용 가능한 메모리와 시스템 리소스에 따라 제한됨
워크시트의 총 행 수	1,048,576행(16,384열)
열 수	255개
행 높이	4096포인트
행 높이 늘리기	가용 메모리 1,024개
행 높이 늘릴 수 있는 총 용량	32,767개
각 행을 늘리는 데 필요한 용량	255
행 높이 늘릴 수 있는 총 용량	255
통합 문서의 시트 수	사용 가능한 메모리와 시스템 리소스에 따라 제한됨(기본값은 1개)
통합 문서의 개수	16,384개(16,384개 워크시트당 16,384개 시트 포함)
통합 문서에서 열 수 있는 총 용량	사용 가능한 메모리와 시스템 리소스에 따라 제한됨

[그림 1-4] Excel 도움말

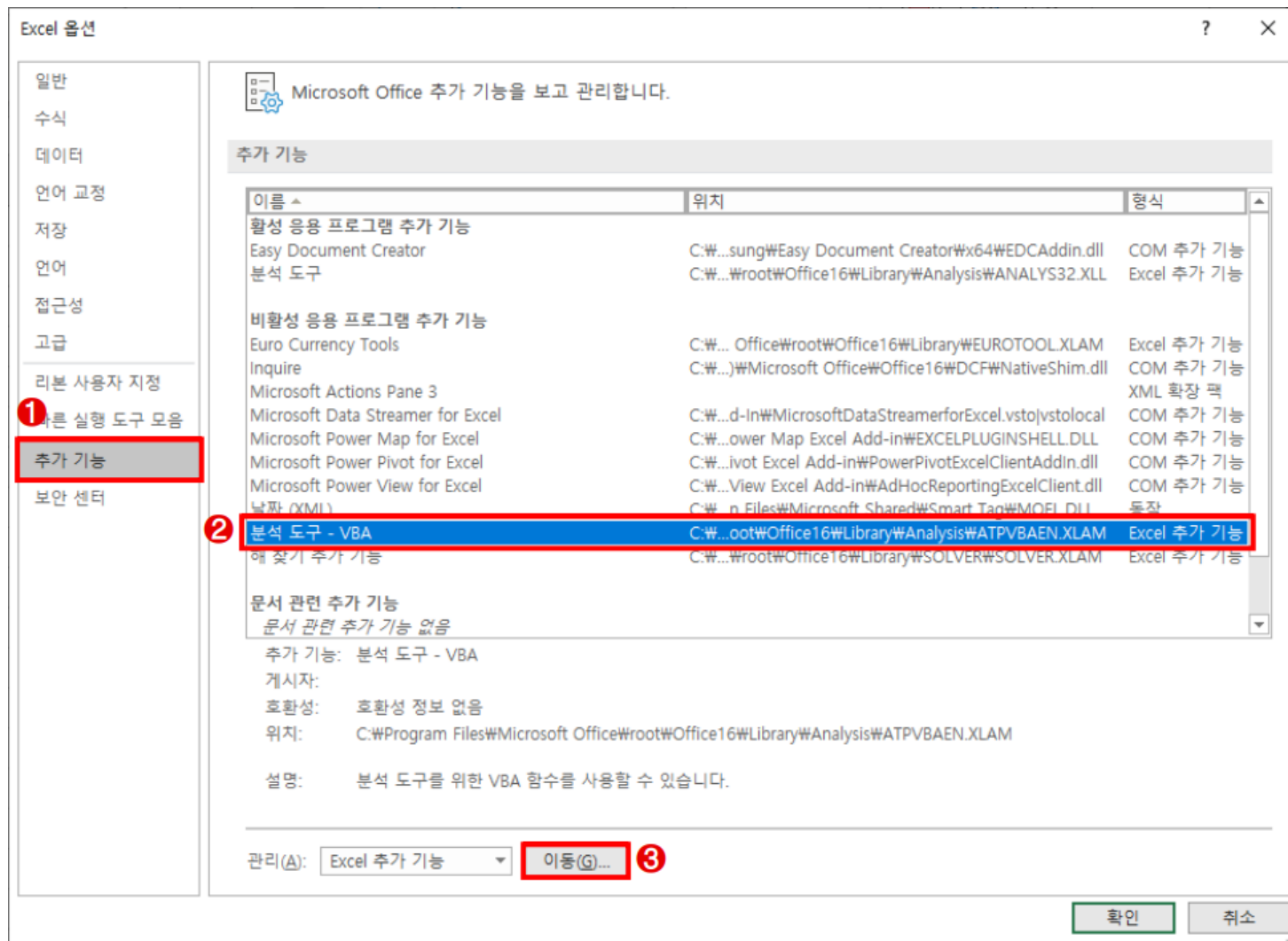
Excel의 통계 기능 사용하기

■ 통계 기능을 위한 '데이터 분석' 메뉴



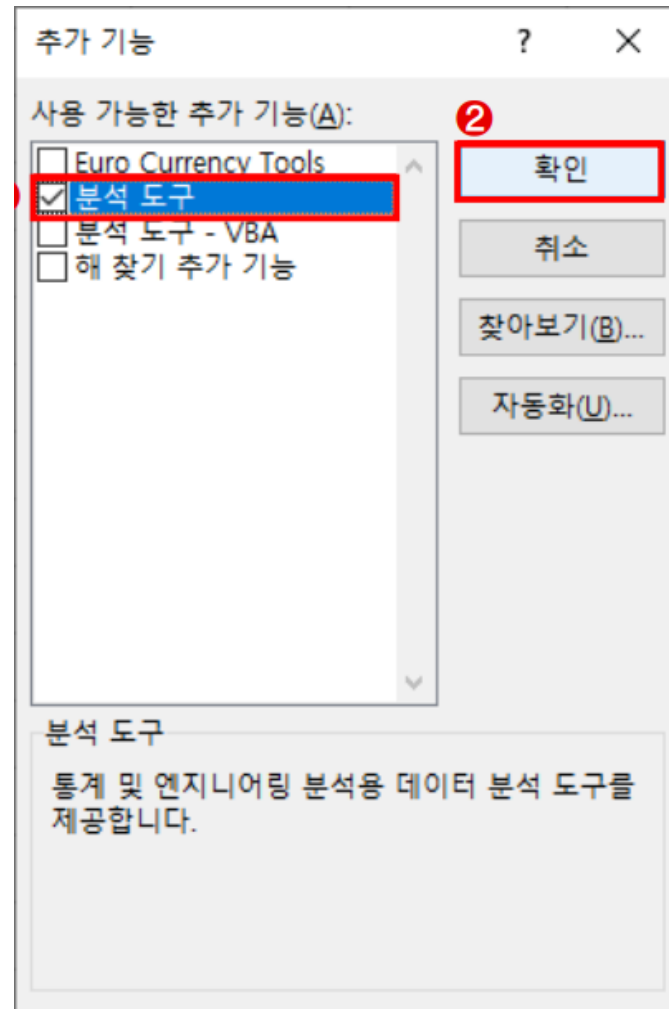
Excel의 통계 기능 사용하기

■ 통계 기능을 위한 '데이터 분석' 메뉴



Excel의 통계 기능 사용하기

■ 통계 기능을 위한 '데이터 분석' 메뉴



■ 통계 기능을 위한 '데이터 분석' 메뉴



Excel의 통계 기능 사용하기

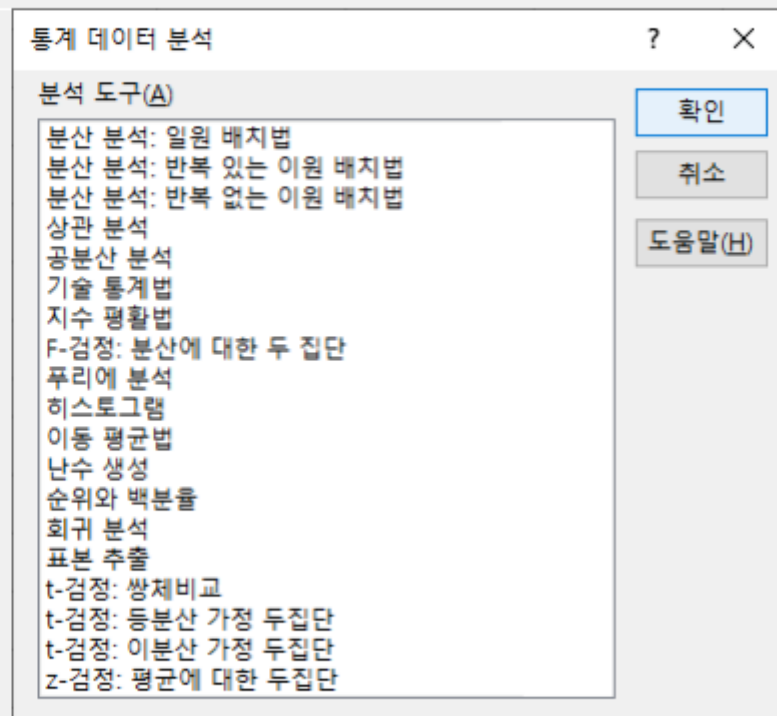
■ 통계 기능을 위한 '데이터 분석' 메뉴

Note Excel의 통계 데이터 분석



Excel 고급 사용자는 다양한 함수를 활용하여 통계분석을 진행할 수 있다. 함수를 사용해서 분석한다는 것은 직접 계산할 부분을 Excel 프로그램이 지원한다는 의미다.

Excel은 19가지 통계분석 메뉴를 지원하고 있다. 물론 [통계 데이터 분석] 창에서 지원하는 통계의 분석 결과가 SPSS와 같은 전문 통계 프로그램으로 계산한 결과와 약간 다르지만, 통계학을 학습하는 수준에서는 만족할 만하다.⁵



[그림 1-9] Excel의 통계 데이터 분석 도구

데이터의 입력, 수정, 삭제

■ 입력

한글, 영문, 특수기호 등의 문자를 입력 후 'Enter'를 눌러 입력을 완료
문자 데이터는 기본 값으로 왼쪽 정렬

숫자 : 입력하면 셀의 오른쪽으로 정렬

날짜 : 직접 입력할 수도 있으며 하이픈(-)이나 슬래쉬(/)로 구분하여 입력

시간 : 숫자를 콜론(:)으로 구분하여 '시, 분, 초'로 입력

수식 : 가감승제(加減乘除)의 경우, '등호(=)'를 먼저 입력한 후 수식을 입력하여 계산

함수 : '등호(=)'를 먼저 입력한 후 함수를 입력.

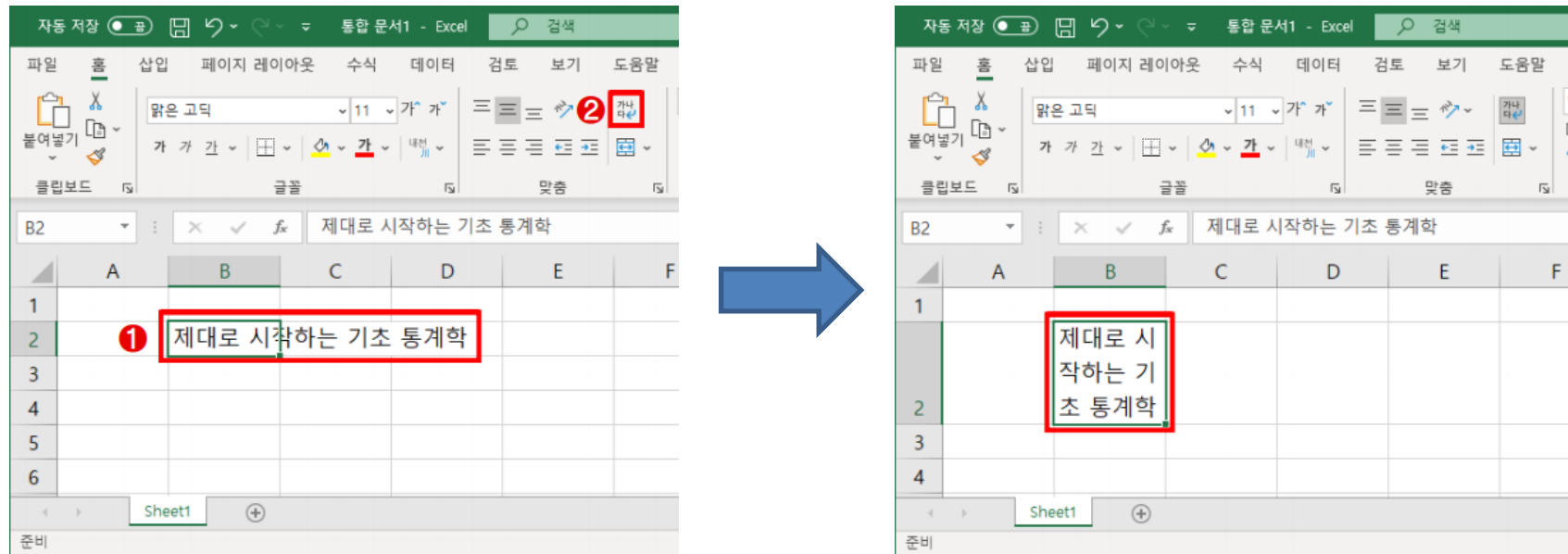
함수명을 아는 경우에는 직접 함수명을 입력

수식 입력줄 왼쪽의 함수 아이콘(f_x)을 클릭하고 함수를 선택하여 입력

데이터의 입력, 수정, 삭제

■ 수정

수정할 데이터가 있는 곳에 셀 포인터를 이동한 후 F2 키를 누르면, 입력된 데이터의 가장 마지막으로 커서가 이동하여 수정가능



셀 포인터를 벗어나는 데이터의 수정은 텍스트 줄 바꿈 아이콘(📄)을 클릭하여 셀의 폭에 맞도록 수정할 수 있다.

데이터의 입력, 수정, 삭제

■ 삭제

삭제하려는 셀로 셀 포인터를 이동-한 후 **Delete** 나 **Back Space** 를 누른다.

Delete 는 해당하는 셀 포인터 내의 전체 데이터를 삭제한다.

Back Space 는 셀 포인터 내의 데이터를 모두 삭제하고, 삭제 후에는 셀 포인터 내에 커서가 위치하여 입력을 준비한다.



Q&A

통계학, 제대로 시작하자!
통계의 쓰임을 이해하고,
실제로 활용할 수 있어야 한다.