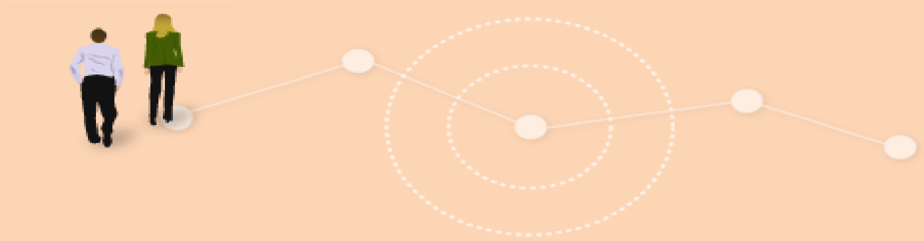
Chapter 07

가설검정 Test of Hypothesis

목 차

01 가설검정과 유의수준

:: Keywords 가설검정 | 가설검정의 절차 | 유의수준
가설검정의 오류 | 양측검정 | 단측검정



가설검정과 유의수준

스포츠 이온 음료의 용량이 과연 250ml일까?

[그림 7-1] 이온 음료

소비자는 제품을 살 때 제조사에서 표시하는 용량이나 성분을 신뢰한다. 음료수를 살 때도 용기에 표기된 용량을 별 의심 없이 받아들인다. 또한 강도 높은 운동을 한 후에 마시는 스포츠 이온 음료가 수분 섭취를 빠르게 해서 건강에 도움이 된다고 믿는 사람들이 많다.

그런데 실제로 운동을 한 후에 250ml로 표기된 이온 음료를 마셨을 때, 너무 땀을 많이 흘린 탓인지 표기된 것보다 양이 적다는 의문이 생긴다. 이런 경우, 이온 음료의 용량이 250ml가 맞는지 어떻게 확인할 수 있을까?

가장 정확한 방법은 대상이 되는 모든 음료수의 용량을 측정하는 것이다. 하지만 이 방법을 사용하면 상품성을 훼손하게 되므로 일정한 표본을 선택해서 조사하여 확인한다.

가설검정

■ 가설(hypothesis)

주어진 사실 혹은 조사하고자 하는 사실이 어떠하다는 주장이나 추측

→ 모수를 추정할 때, 모수가 어떠하다(혹은 어떠할 것이다)는 조사자의 주장이나 추측을 일컬음

가설검정

■ 귀무가설(歸無假說, null hypothesis)

일반적으로 믿어온 사실을 가설로 설정한 것.

영가설(零假設)이라고도 하며, 영(零)이라는 의미로 H_0로 표기

→ 귀무가설을 증명하기 위한 조 사는 하지 않는다.

연구 목적이 되는 가설에 상반되는 의미를 가지는 가설이므로

귀무가설 자체를 연구해도 어떤 의미를 찾아내긴 어렵기 때문이다.

H_0 : 스포츠 이온 음료의 용량은 제품에 표기된 250ml다.

가설검정

■ 대립 가설(代立假設, antihypothesis)

공공연하게 사실로 받아들여진 현상에 대립되는 가설이며, 연구목적이 된다.
→ 대립가설을 조사하여 대립가설이 지지를 받으면 큰 의미가 있다.

대립가설은 연구를 통한 증명이 목적이며,
표본에 근거한 강력한 증거로 입증하고자 하는 가설

연구를 진행하는 의미가 있으므로 연구가설(硏究假說, research hypothesis) 이라고도 하며,
영(零, 0)에 반대가 된다는 의미로 H_1으로 표기.

- 양측검정[3]

H_1 : 스포츠 이온 음료의 용량은 제품에 표기된 250ml가 아니다.

- 단측검정

H_1 : 스포츠 이온 음료의 용량은 제품에 표기된 250ml보다 적다.

혹은

H_1 : 스포츠 이온 음료의 용량은 제품에 표기된 250ml보다 많다.

가설검정의 오류

통계적 판단 = 가설을 검정한다는 의미
가설검정은 통계량으로 모수를 추정하는 과정
∴ 아무리 정확히 추정하더라도 가설과 모수 사이에는 항상 오차가 발생

따라서 모수를 추정할 때는 항상 오류 가능성(확률)을 제시해야 한다.

통계적 의사결정에서 틀리지 않는 것이 목적이라면
→ 추정된 모수의 틀릴 확률을 낮게 잡으면 된다.
→ 오차를 최소한으로 줄여 0에 가깝게 만들어야 의사결정의 실패를 줄일 수 있다.

반면, 의사결정의 오류와 상 관없이 특정한 수치를 도출하는 것이 목적이라면
표본으로부터 계산한 결과를 정확하게 제시하면 된다.

가설검정의 오류

- **제1종 오류**
  - **귀무가설을 채택해야 함에도 귀무가설을 기각하는 경우**를 말한다.
  - H_0를 기각할 확률을 α라고 한다면 채택하게 될 확률은 $1-\alpha$로 표시할 수 있다.
  - 가설검정을 위해 **유의수준을 나타내는** α**값과 관련**이 있다.

- **제2종 오류**
  - **귀무가설을 기각해야 함에도 귀무가설을 채택하는 경우**를 말한다.
  - 귀무가설을 기각해야 함에도 귀무가설을 채택하는 경우의 확률을 β로 표시한다면, 귀무가설을 기각하는 것이 옳을 확률은 $1-\beta$가 된다.

가설검정의 오류

■ 가설 검정력

귀무가설을 채택해야 하지만 귀무가설을 기각하는 경우의 확률은 유의수준 α로 표시

귀무가설을 기각해야 함에도 귀무가설을 채택하는 경우의 확률은 β로 표시

→ 기각해야 할 귀무가설을 기각하는 확률은 1- β

→ 대립가설(H_1)이 사실인 상황에서 귀무가설(H_0)을 기각할 확률

2종 오류가 발생하지 않을 확률 : 가설의 검정력(power of hypothesis testing)

→ 즉, 귀무가설이 거짓일 때 이를 기각할 확률을 의미.

양측검정과 단측검정

■ 기각의 판단 기준은 양측검정과 단측검정으로 구분

→조사 결과가 유의수준 α에 포함되면 기각, 포함되지 않으면 채택

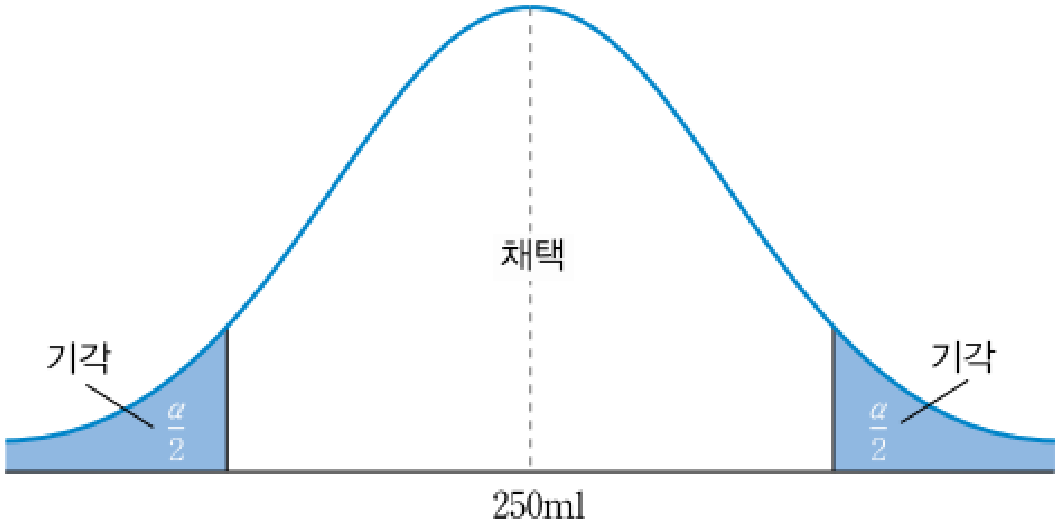
→채택/기각의 기준은 항상 귀무가설

양측검정과 단측검정

■ 양측검정(two-sided test)

조사하고자 하는 대립가설, 즉 ‘사실이 아니다’라는 것을 검정하여 귀무가설을 기각하고 대립가설을 채택하고자 하는 것

$$H_0: \mu = 250ml$$
$$H_1: \mu \neq 250ml$$

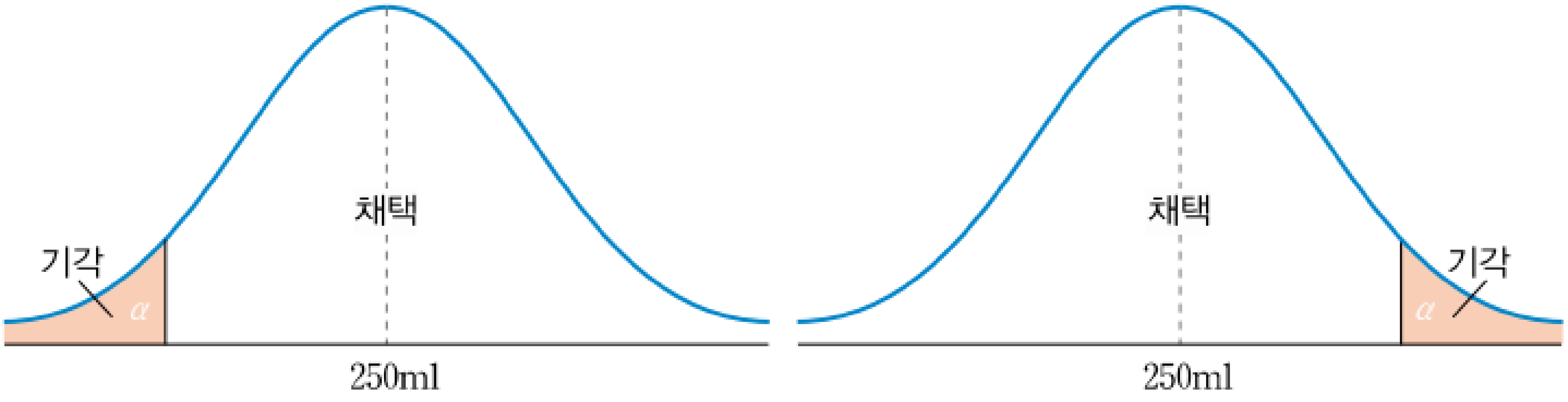


양측검정과 단측검정

■ 단측검정(one-sided test)

조사의 목적에 따라 대립가설을 스포츠 이온음료의 용량이 300ml보다 적다고 수립하거나, 혹은 300ml보다 많다고 수립하여 한 쪽만 살펴보는 것

$$H_0: \mu \geq 250ml \qquad H_0: \mu \leq 250ml$$
$$H_1: \mu < 250ml \quad \text{혹은} \quad H_1: \mu > 250ml$$



(a) 좌측검정　　(b) 우측검정

가설검정의 절차

가설 수립 → 유의수준 결정 → 기각역 설정 → 통계량의 계산 → 의사결정

① 가설수립

귀무가설 H_0와 대립가설 H_1을 수립한다.

② 유의수준 결정

수립된 귀무가설과 대립가설 중 어떤 가설을 채택할 것인지 판단하는 유의수준 a를 결정한다.

③ 기각역 설정

조사의 성격에 따라 양측검정을 할 것인지 단측검정을 할 것인지를 정해 기각역을 설정한다.

가설검정의 절차

④ 통계량의 계산

수집된 표본을 대상으로 조사에 필요한 통계량을 계산한 후 기각역과 비교한다.

⑤ 의사결정

가각역과 비교한 후에 귀무가설과 대립가설 중 어떤 가설을 채택할 것인지를 결정한다.

가설검정의 절차

Note 가설검정의 진행 및 해석 시 알아두어야 할 용어

- **가설**

통계적 가설검정의 대상이 되는 가설에는 기존에 믿어온 사실이 맞다는 귀무가설(H_0)과 기존에 믿어온 사실이 실제와 다르다는 대립가설(H_1)이 있다.

- **검정통계량**

표본통계량을 계산했을 때 영가설(H_0)이 가지는 값이 모수로부터 어느 정도 떨어져 있는지 나타내는 지표를 말한다.

- **임계치**

임계치는 영가설을 기각하거나 채택하기 위한 한계값을 의미한다. 검정통계량을 계산하여 임계치를 기준으로 좌/우 어느 쪽에 있는가에 따라 채택과 기각을 판단한다.

가설검정의 절차

• 유의수준과 p값

검정통계량을 해석하기 위해 영가설(H_0)이 발생할 가능성을 확률로 표시한 것이다. α는 유의수준(significance level)으로 전적으로 연구자(조사자)가 결정하게 되는데, 이는 연구자가 연구에 맞는 적절한 유의수준을 기준으로 가설을 검정할 것이기 때문이다. 유의수준은 확률(probability)로 표시되므로 약자를 사용하여 p값(p-value)으로 표시한다.

일반적으로 통계학에서는 95%, 99%, 99.9%를 기준으로 유의수준을 설정한다. 아래에 제시한 표기 방법은 절대적인 것이 아니며 일반적으로 연구자들이 보고서나 논문에 표기하는 방식이다. 따라서 연구자에 따라서는 다른 표기 방법을 써서 표시할 수 있다. 이때 반드시 주석으로 어떤 의미인지를 보고해야 한다.

예 95%유의수준 : $\alpha = 0.05$ 혹은 $p \leq 0.05$ ⇒ *으로 표시

99%유의수준 : $\alpha = 0.01$ 혹은 $p \leq 0.05$ ⇒ **으로 표시

99.9%유의수준 : $\alpha = 0.001$ 혹은 $p \leq 0.05$ ⇒ ***으로 표시

• 결과의 해석

검정통계량을 계산한 값이 어느 영역에 포함되는지를 확인하여 가설의 채택/기각 여부를 판단한다. 일반적으로 연구자는 유의수준을 보면서 결과를 해석한다. p값을 기준으로 값이 작게 나온다면 귀무가설을 기각하고 대립가설을 채택한다.

가설검정의 기본 요소

■ **무작위 추출**

→ 표본을 모집단으로부터 추출할 때 무작위로 추출해야 한다.

■ **통계적 독립성**

→ 표본으로부터 관측치를 수집할 때
어떤 하나의 관측치가 다른 관측치에 영향을 미치지 않아야 한다.

■ **정규성**

→ 표본을 구성하는 관측치는 충분히 많아서 분포가 정규분포를 구성해야 한다.

■ **등분산성**

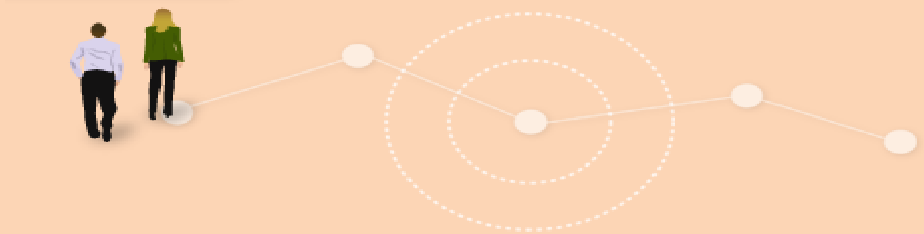
→ 둘 이상의 표본을 비교하는 가설검정을 하는 경우 이 두 집단은 서로 등분산을 구성

■ **표본의 크기**

→ 표본의 크기를 늘리면 표준오차가 줄어들어
모수에 접근하게 되므로 통계적 검정력은 높아진다

02 모집단 평균의 가설검정

:: Keywords 모분산을 아는 경우의 가설검정
모분산을 모르는 경우의 가설검정
p값을 이용한 가설검정



모분산을 아는 경우의 가설검정

■ 가설을 수립

귀무가설을 모평균이 계산된 특정 값과 동일하므로

$$H_0 : \mu = \mu_0$$

대립가설은 양측검정의 경우 모평균이 특정 값과 동일하지 않으므로

$$H_1 : \mu \neq \mu_0$$

단측검정의 경우에는 좌측검정: $H_1 : \mu < \mu_0$

우측검정: $H_1 : \mu > \mu_0$

모분산을 아는 경우의 가설검정

■ 유의수준의 결정

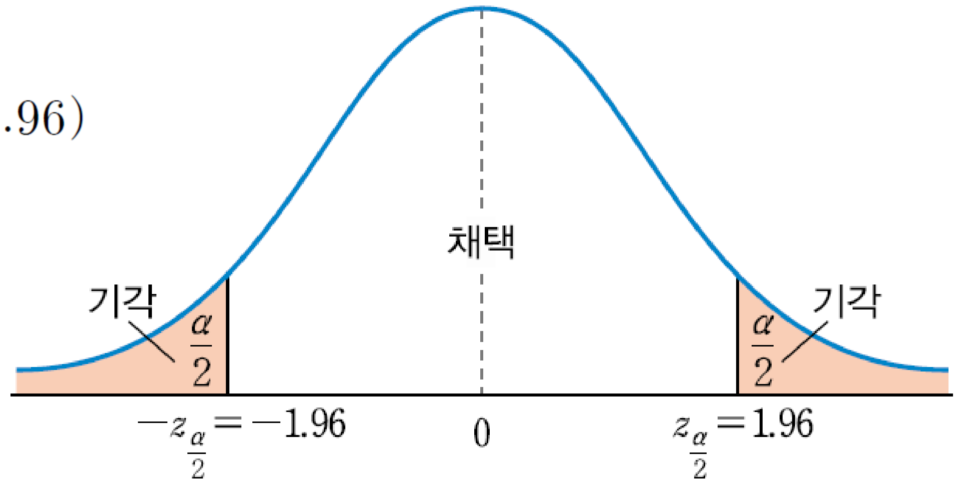
귀무가설과 대립가설에 대해 $\alpha = 0.05$에서의 가설검정

양측검정이면 $\pm z_{\frac{\alpha}{2}} = z_{0.025} = \pm 1.96$,

검정통계량은 $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

Z의 절대값이 1.96보다 크면 $(|z| > 1.96)$

귀무가설 H_0를 기각



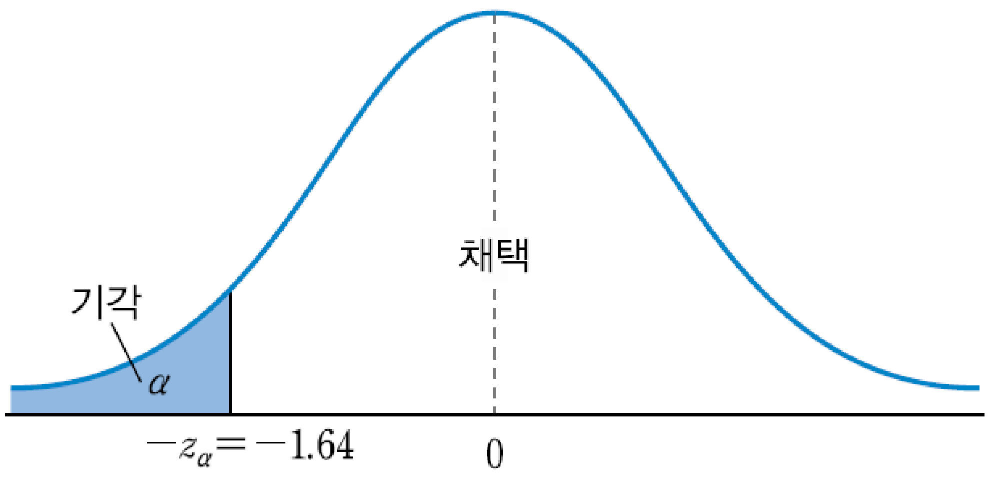
모분산을 아는 경우의 가설검정

■ 유의수준의 결정

귀무가설과 대립가설에 대해 $\alpha = 0.05$에서의 가설검정

좌측검정이면 $-z_{\alpha} = -z_{0.05} = -1.64$

검정통계량 z값이 -1.64보다 작으면 $(z < -1.64)$ 귀무가설 H_0를 기각



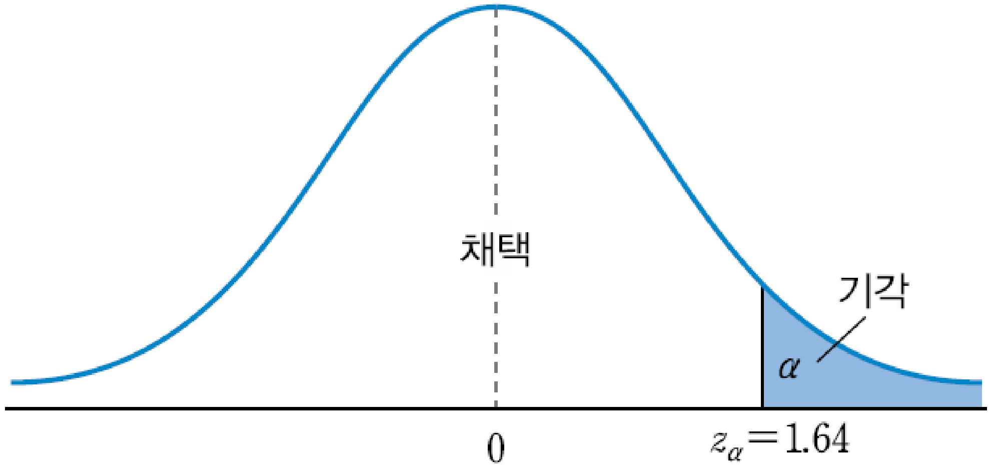
모분산을 아는 경우의 가설검정

■ 유의수준의 결정

귀무가설과 대립가설에 대해 $\boldsymbol{\alpha = 0.05}$에서의 가설검정

우측검정이면 $z_\alpha = z_{0.05} = 1.64$

검정통계량 z값이 -1.64보다 크면 $(z < 1.64)$ 귀무가설 H_0를 기각



모분산을 아는 경우의 가설검정

예제 7-1 모분산을 아는 경우의 가설검정 준비파일 | 7장_가설검정(모분산을 아는 경우).xlsx

스포츠 이온 음료의 용량이 제품에 표시된 250ml보다 적은 것 같아서 직접 조사를 해보고자 한다. 표본 300개를 대상으로 용량을 측정한 결과, 평균 244.65로 확인되었다. 표준편차가 20일 때, 가설을 수립하고 유의수준 0.05에서의 좌측검정을 실시하라.

모분산을 아는 경우의 가설검정 (예제 풀이)

■ 유의수준의 결정

귀무가설과 H_0:$\mu = 250$, 대립가설은 H_1: $\mu < 250$

유의수준 0.05에서의 좌측검정

n =300, $\bar{x} = 244.65$, $\sigma = 20$, $\sigma = 0.05$이므로

-z_α = -$z_{0.05} = -1.64$이며, $z < -1.64$이면 귀무가설 H_0를 기각

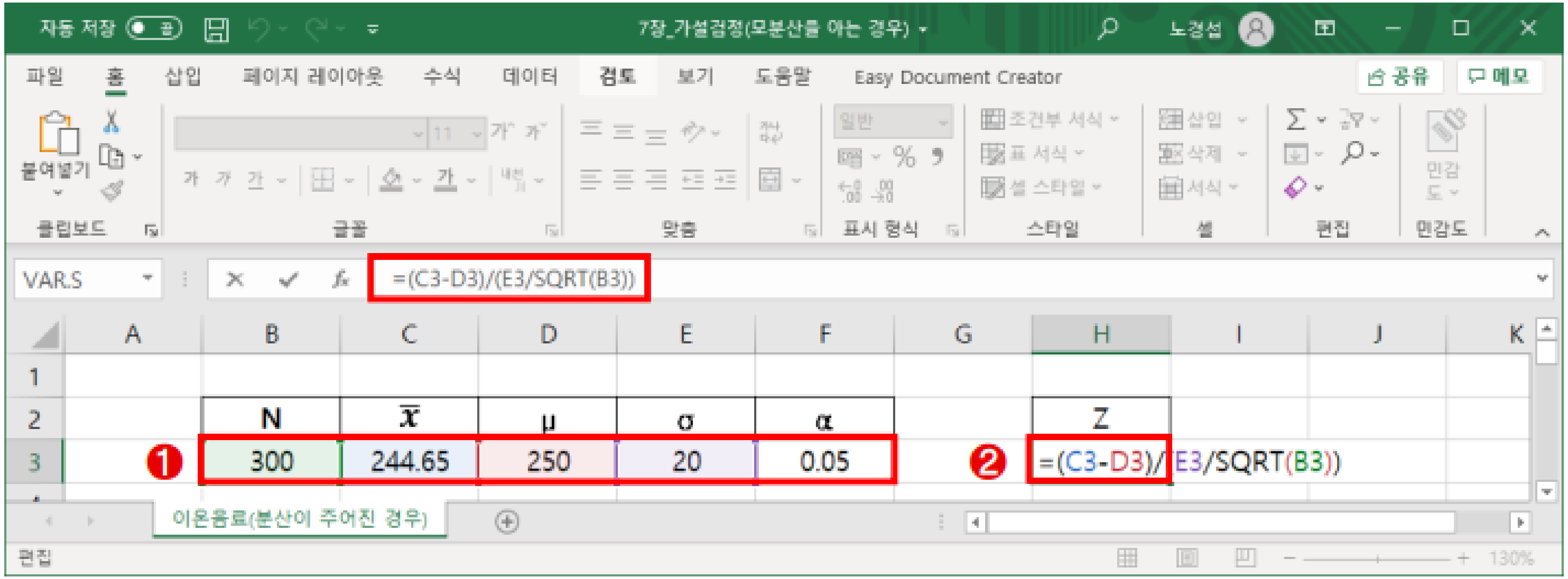
검정통계량을 구하면 $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{244.65 - 250}{20 / \sqrt{300}} = -4.633\cdots$

$z < -1.64$를 만족

따라서 스포츠 이온음료의 용량이 250ml라는 귀무가설 H_0를 기각하므로

대립가설 H_1을 채택

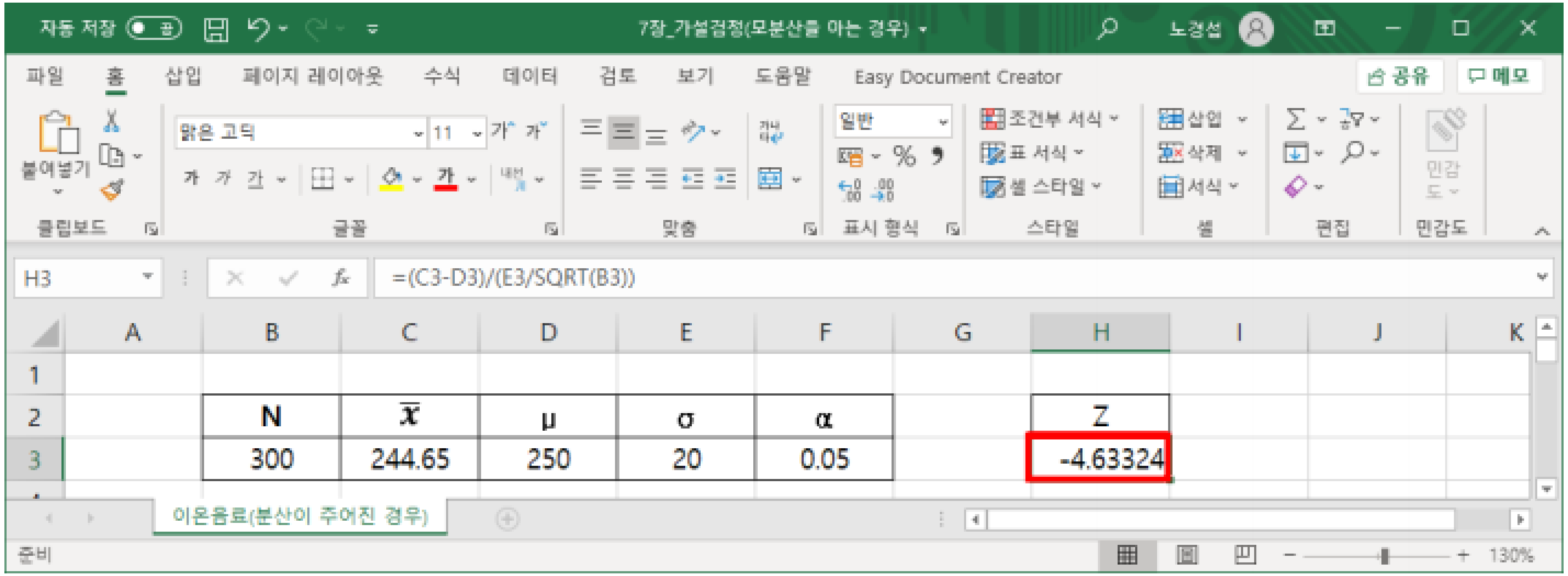
모분산을 아는 경우의 가설검정 (예제 Excel 풀이)



❶ 'N'은 300, '$\overline{X}$'는 244.65, 'μ'는 300, 'σ'는 20, 'α'는 0.05를 입력한다.

❷ H3셀에 검정통계량을 구하는 계산식 '=(C3−D3)/(E3/SQRT(B3))'를 입력

모분산을 아는 경우의 가설검정 (예제 Excel 풀이)



❶ ‘N’은 300, ‘$\overline{X}$’는 244.65, ‘μ’는 300, ‘σ’는 20, ‘α’는 0.05를 입력한다.

❷ H3셀에 검정통계량을 구하는 계산식 ‘=(C3−D3)/(E3/SQRT(B3))’를 입력

모분산을 모르는 경우의 가설검정

■ 표본이 큰 경우의 가설검정

모수를 모르더라도 표본이 아주 큰 경우,

즉 $n \longrightarrow \infty$ 이면 모집단의 정규성과 상관없이 중심극한정리에 의해 s^2 은 σ^2 으로 수렴

그러므로 표본이 큰 경우의

검정통계량은 모분산이 주어진 것과 동일하게 $Z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ 로 계산

■ 표본이 작은 경우의 가설검정

표본이 충분하지 못한 경우는

표본통계량 $t_{n-1} = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ 는 자유도 $(n - 1)$ 에서 t분포를 따르게 된다.

모분산을 모르는 경우의 가설검정

Note ***t*검정과 *z*검정의 관계**

통계학을 학습할 때 검정 방법이 다양하여 혼란스러울 수 있다. t검정과 z검정의 관계만 보더라도 검정통계량을 구하는 공식은 비슷한데 도출 결과가 다르다. 이 두 가지 검정 방법은 중심극한정리에 의해 구분되는 것으로 판단하면 이해하기 쉽다.

즉 표본이 30보다 적으면 t분포를 이용하고, 그보다 많으면 z분포를 이용한다. 다만 t분포표를 보면 알 수 있듯이 t분포는 30보다 적은 표본의 분포를 나타내는 동시에 30보다 많은 표본의 분포도 포함한다. 다시 말해 z분포는 t분포에 포함된다고 할 수 있다.

[그림 7-9] ***t*검정과 *z*검정의 관계**

실제로 일반적으로 사용되는 통계프로그램 SPSS Statistics 에서는 z검정이라는 개념이 존재하지 않는다. 표본의 개수가 많아지면 t분포가 z분포로 수렴하고 표본의 개수가 ∞ 로 접근하면 t분포와 z분포는 같아지게 된다. 이는 t분포로 z분포를 설명할 수 있다고 할 수 있다.

모분산을 모르는 경우의 가설검정

예제 7-2 모분산을 모르는 경우의 가설검정

준비파일 | 7장_가설검정(모분산을 모르는 경우).xlsx

[예제 7-1]의 스포츠 이온 음료의 용량에 대한 조사를 다음과 같이 표본 15개를 대상으로 살펴보고자 한다. 직접 가설을 수립하고, 유의수준 0.05에서 좌측검정을 실시하라. 단, 모집단의 분포는 정규분포라 가정한다.

번호	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
용량(ml)	257	253	241	242	277	240	270	235	265	235	245	238	260	275	250

모분산을 모르는 경우의 가설검정 (예제 풀이)

귀무가설은 $H_0 : \mu = 250$

대립가설은 $H_1 : \mu < 250$ $\qquad n = 15,\ \alpha = 0.05$

$\bar{x}$와 s는 주어지지 않았으므로 구해야 한다.

$$\bar{x} = \frac{257 + 253 + 240 + \cdots + 250}{15} = 252.2$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{(257 - 252.2)^2 + (253 - 252.2)^2 + \cdots + (250 - 252.2)^2}{14} = 209.2$$

$$s = \sqrt{209.2} = 14.46$$

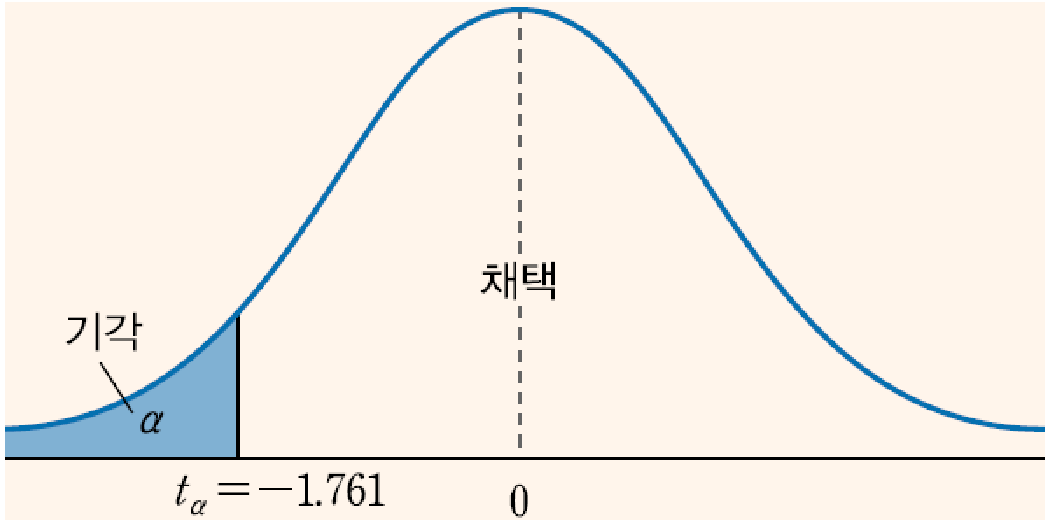
모분산을 모르는 경우의 가설검정 (예제 풀이)

검정통계량 t를 계산하기 위해 $\dfrac{\bar{x}-\mu}{s/\sqrt{n}}$에 대입하면 $t=\dfrac{252.2-250}{14.46/\sqrt{15}}=0.589$

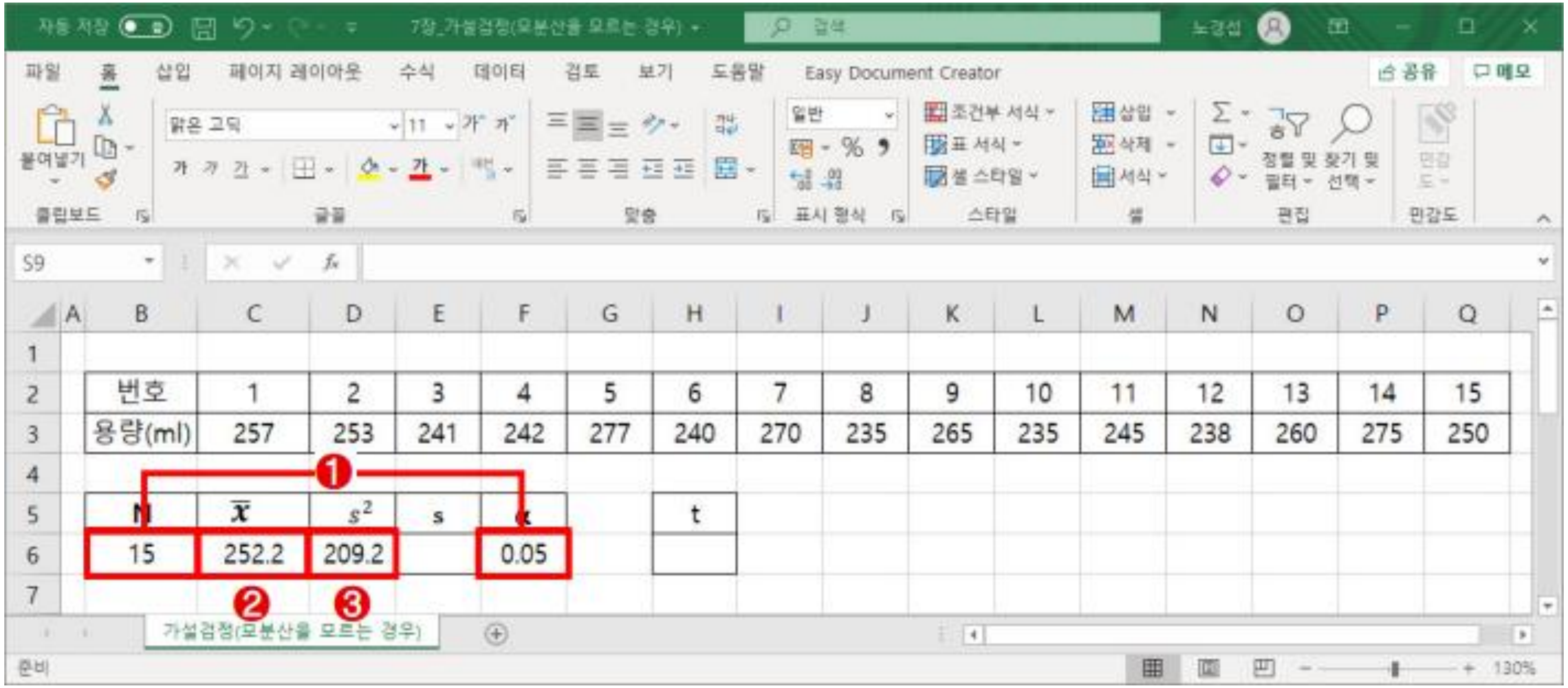
자유도는 n-1=14, 좌측검정 $\alpha=0.05$인 t값은 -1.761 이다.

좌측검정에 의해 t값의 검정 통계량이 0.589

→귀무가설 H_0를 기각할 수 없다.



모분산을 모르는 경우의 가설검정 (예제 풀이)



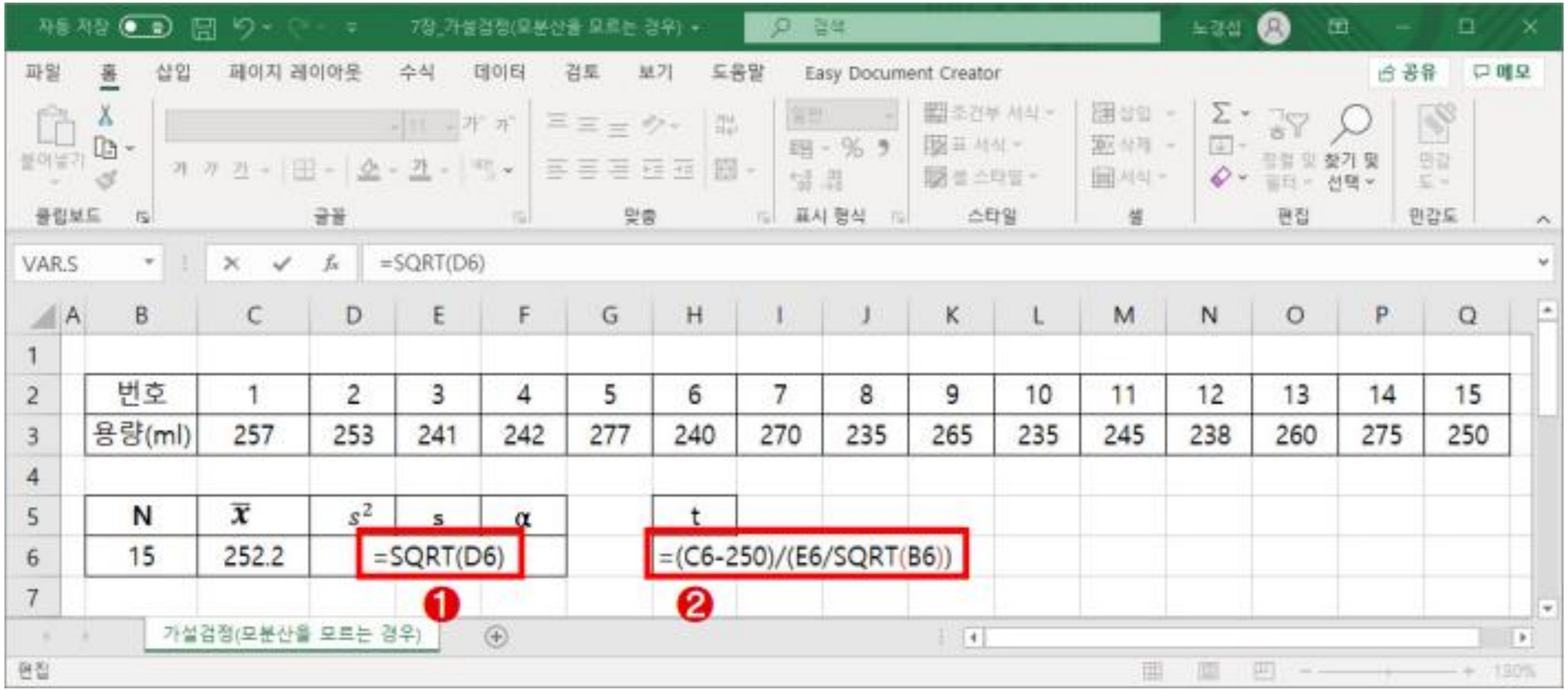
❶ ‘N’은 15, ‘α’는 0.05를 입력한다.

❷ C6셀에 표본평균을 구하는 함수식 ‘=AVERAGE(C3:Q3)’를 입력하고,

❸ D6셀에 표본분산을 구하는 다음 계산식

=((C3−C6)^2+(D3−C6)^2+(E3−C6)^2+(F3−C6)^2+(G3−C6)^2+(H3−C6)^2+(I3−C6)^2+(J3−C6)^2+(K3−C6)^2+(L3−C6)^2+(M3−C6)^2+(N3−C6)^2+(O3−C6)^2+(P3−C6)^2+(Q3−C6)^2)/(B6−1) 을 입력

모분산을 모르는 경우의 가설검정 (예제 풀이)

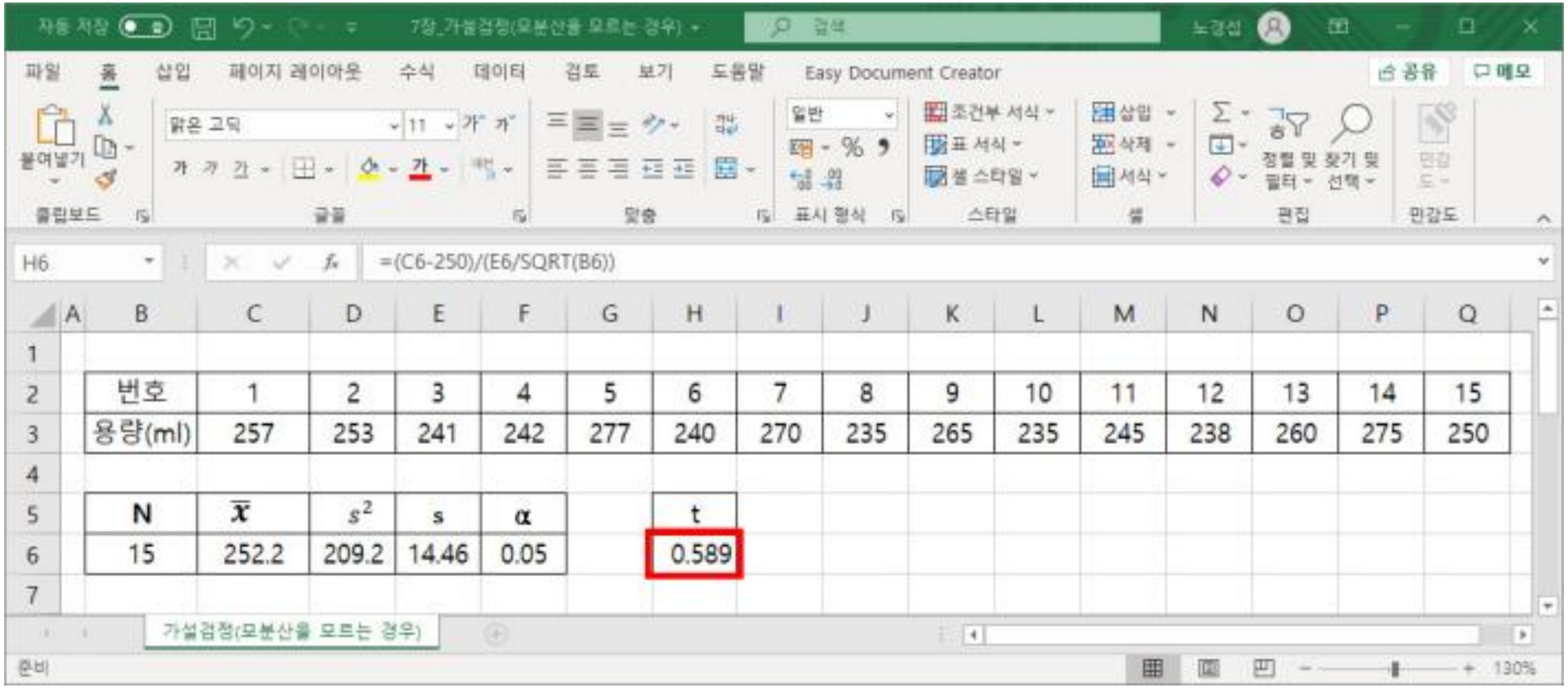


	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2	번호	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3	용량(ml)	257	253	241	242	277	240	270	235	265	235	245	238	260	275	250
5	N	$\bar{x}$	s^2	s	α		t									
6	15	252.2		=SQRT(D6)			=(C6-250)/(E6/SQRT(B6))									

❶ E6셀에 표본의 표준편차를 구하는 함수식 '=SQRT(D6)'를 입력하고

❷ H6셀에 검정통계량을 구하는 계산식 '=(C6−300)/(E6/SQRT(B6))'를 입력

모분산을 모르는 경우의 가설검정 (예제 풀이)



번호	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
용량(ml)	257	253	241	242	277	240	270	235	265	235	245	238	260	275	250

N	$\bar{x}$	s^2	s	α	t
15	252.2	209.2	14.46	0.05	0.589

t가 0.589로 계산됨

즉 $t > -1.761$을 만족

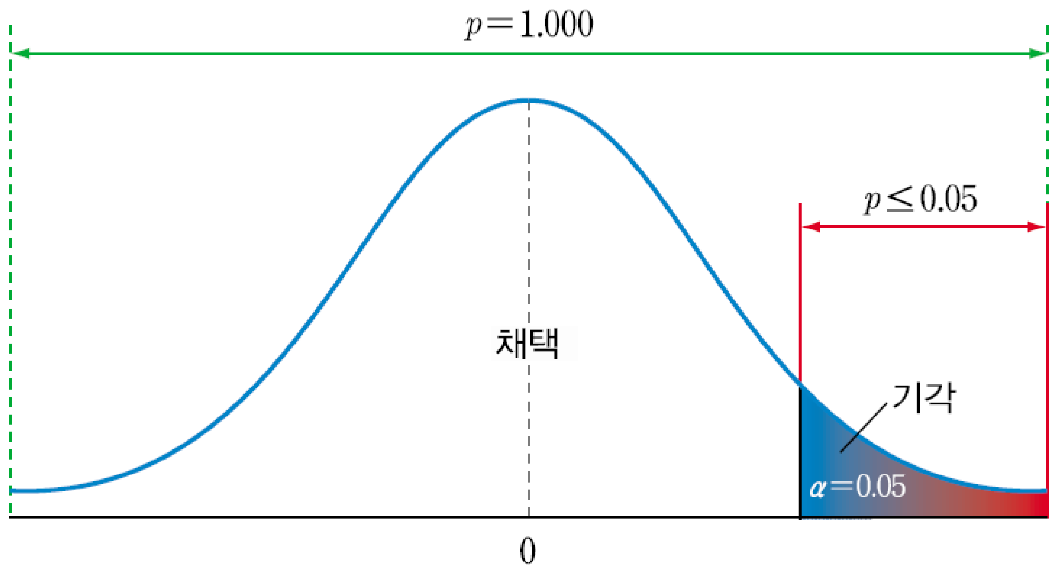
■ p값을 이용한 가설검정

유의수준에 따라 채택/기각을 결정한 지금까지의 방법은 신뢰범위에 포함되는지 그렇지 않는지만을 제시하므로 채택/기각에 대한 강도를 표현하기에는 충분하지 않음

→ 이러한 단점을 보완하기 위해 채택/기각에 대한 기준을 확률 p로 나타내려는 방법

■ P값을 이용한 가설검정

p값은 귀무가설을 기각하기 위한 최대한의 한계점을 나타내는데, 유의수준 α를 기준으로 보면 α로부터 멀리 떨어져 있는 확률을 나타낸다.



p값이 0.05보다 작거나 같다면 귀무가설을 기각

p값을 이용한 가설검정

Note **p값의 다른 표현들**

p값을 표현하는 방법이 다양하므로, 그 방법들을 알고 있어야 어떤 조사 결과를 접하더라도 이해하기가 수월하다.

p값의 다른 표현 : p-value, significance level(혹은 sig.), 유의수준, 유의확률

통계분석 방법에 따라 p값과 달리 표현되는 값들도 있는데, 실제로는 각각의 경우에 따라 p값과 1:1로 대응해서 나오는 수치들이다. Excel은 물론이고 통계 프로그램을 이용하여 도출된 결과를 바탕으로 수치들을 비교해보면, 분석 방법에 따라 다른 수치가 표현되지만 p값은 공통으로 표시된다. 그 이유는 p값이 확률이므로 0~1까지의 값을 가지게 되어 다른 수치들을 표준화한 효과가 나타나기 때문이다. 이러한 이유로 거의 모든 분석 결과에서 p값을 사용한다. [표 7-1]을 보면 통계분석 방법에 따라 유의수준은 각각 1:1로 대응하게 되는데, 공통적으로 p값이 모두 포함되어 있음을 확인할 수 있다.

[표 7-1] 통계분석 방법별 유의수준의 종류

t검정	분산분석	회귀분석	교차분석
t, p	F, p	F, t, p	χ^2, p

p값을 이용한 가설검정

예제 7-3 **p값을 이용한 가설검정** 준비파일 | 7장_가설검정(p값).xlsx

[예제 7-1]의 스포츠 이온 음료의 용량에 대한 조사에서 표본 300개를 대상으로 살펴보고자 한다. 용량을 직접 측정한 결과 평균이 244.65ml로 확인되었다. 표준편차가 45일 때, 직접 가설을 수립하고 유의수준 0.05에서 귀무가설의 기각 여부를 판단하라.

p값을 이용한 가설검정 (예제 풀이)

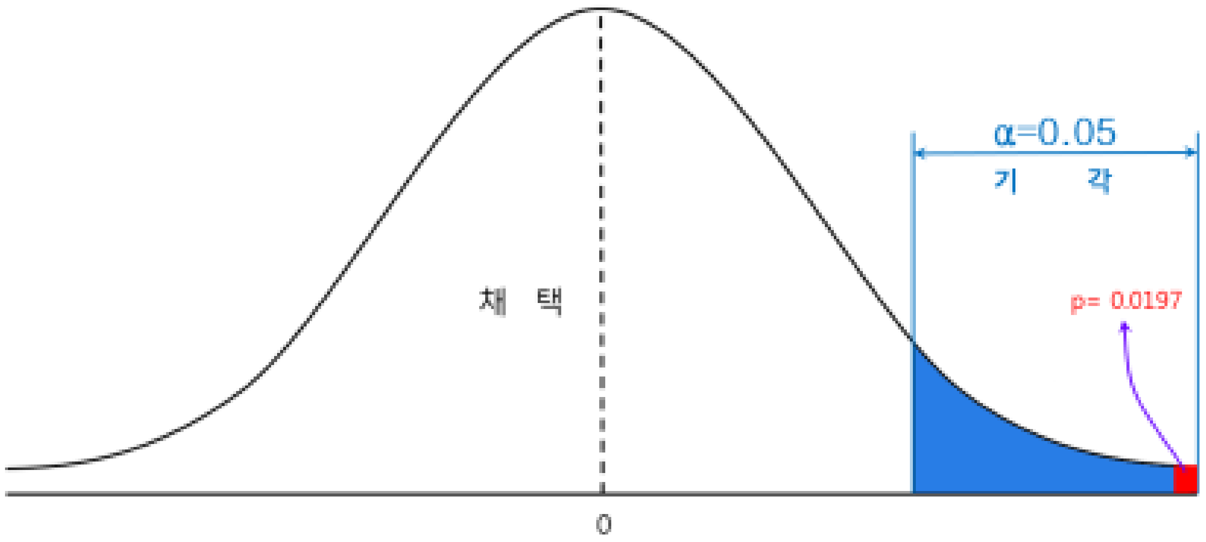
귀무가설은 $H_0 : \mu = 250$, 대립가설은 $H_1 : \mu < 250$

$n = 300$, $\bar{x} = 244.65$, $s = 45$, $\alpha = 0.05$

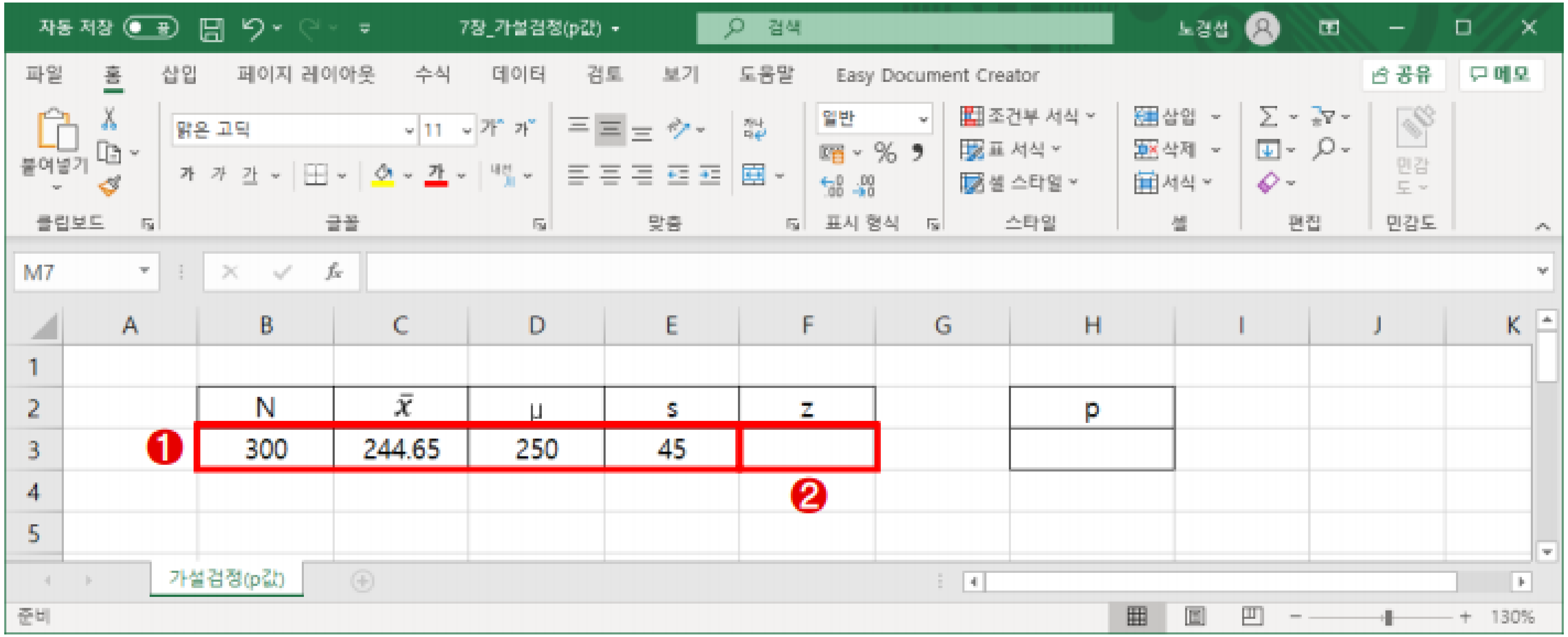
$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{244.65 - 250}{45/\sqrt{300}} = -2.059$$

$$\begin{aligned} p &= P(z < -2.059) \\ &= 0.5 - P(0 \le z \le 2.059) \\ &\fallingdotseq 0.5 - P(0 \le z \le 2.06) \\ &= 0.5 - 0.480 = 0.020 \end{aligned}$$

$p < 0.05$를 만족



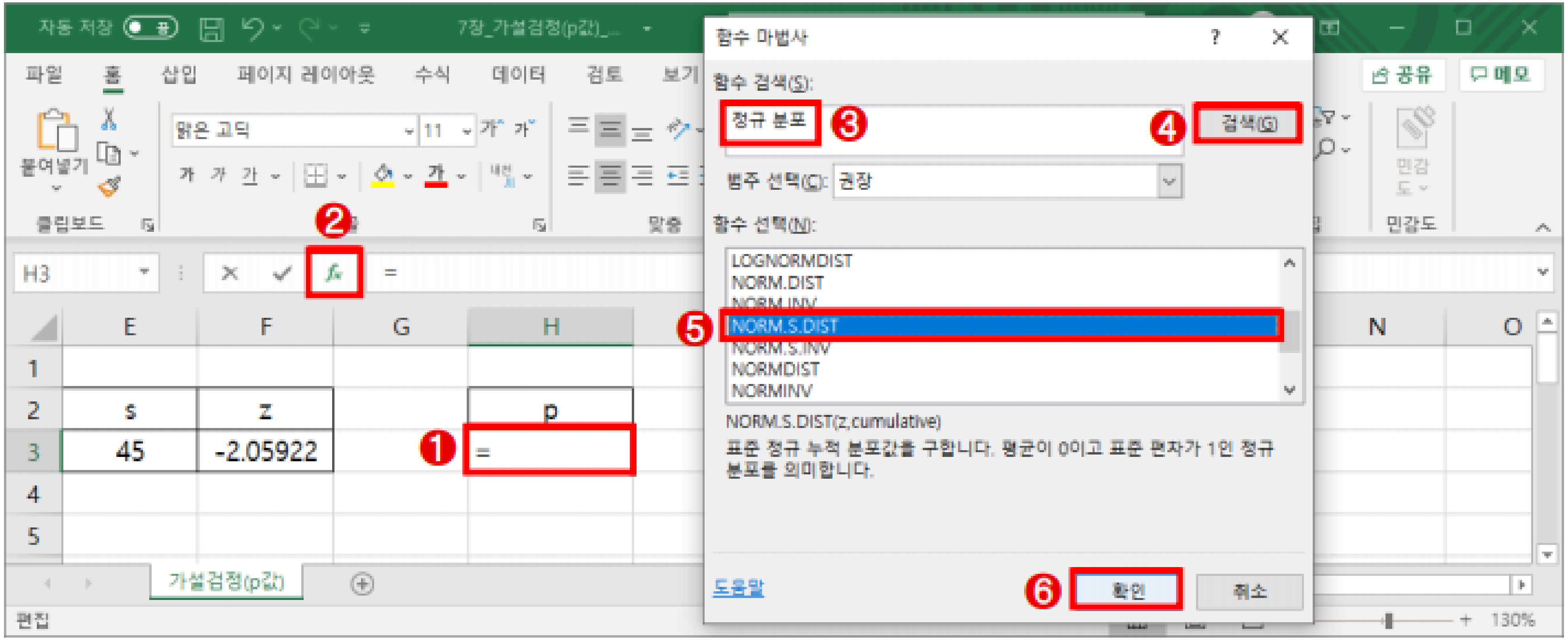
p값을 이용한 가설검정 (예제 Excel 풀이)



❶ 'N'은 300, 'μ'는 294.65, 'σ'는 45를 입력

❷ E3셀에 검정통계량을 구하는 계산식 '=(C3−B3)/(D3/SQRT(B3))'를 입력

p값을 이용한 가설검정 (예제 Excel 풀이)



❶ 셀 포인터를 G3셀에 위치하고 ❷ f_x 를 클릭

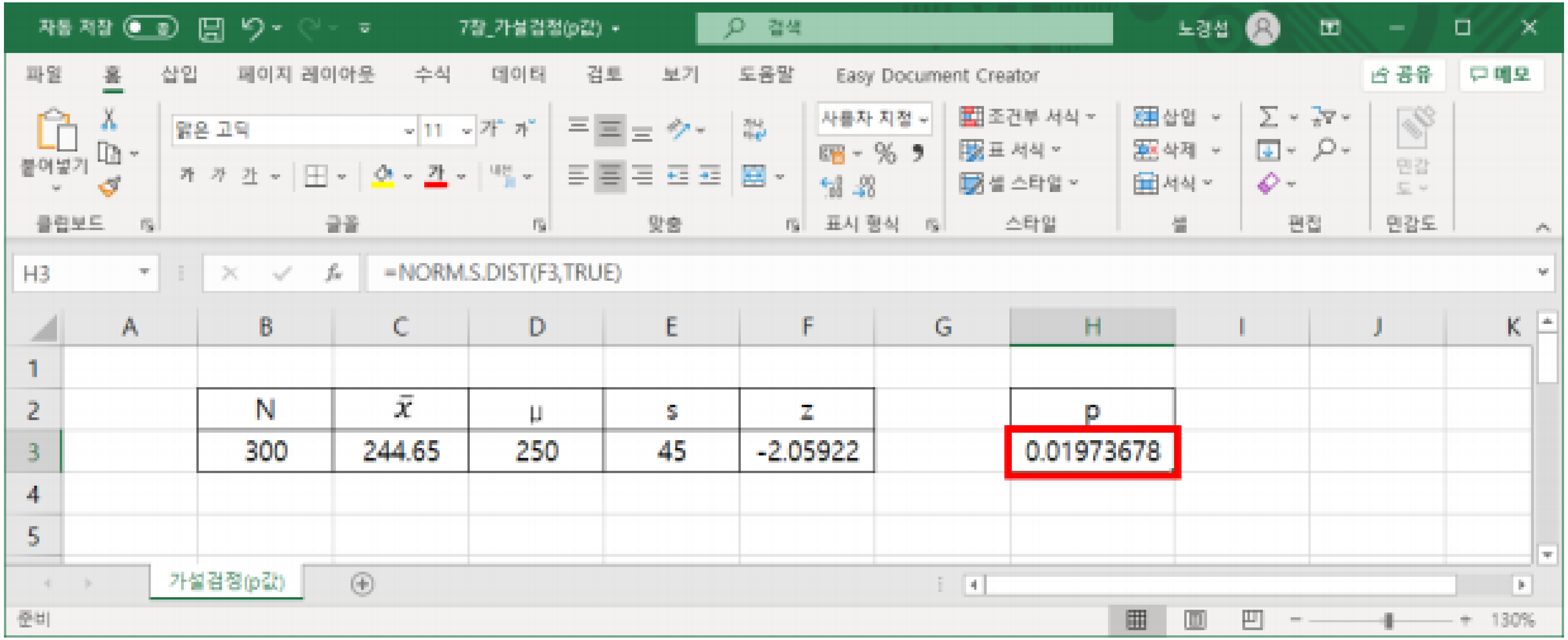
❸ '함수 검색(S):'에 '정규분포'를 입력한 후 ❹ [검색]을 클릭

❺ '함수 선택(N):'에서 'NORM.S.DIST' 함수를 선택한 후 ❻ [확인]을 클릭

p값을 이용한 가설검정 (예제 Excel 풀이)

❶ 'Z'에 E3셀을 입력 ❷ 'Cumulative'에는 누적확률분포인 TRUE를 입력한 후
❸ [확인]을 클릭

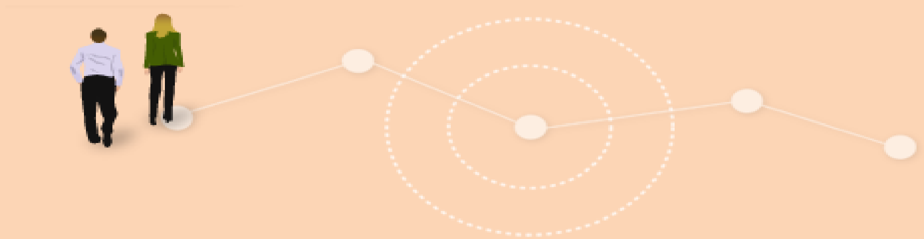
p값을 이용한 가설검정 (예제 Excel 풀이_완료)



p의 값이 0.01973678로 확인. $p < 0.05$를 만족

03 모집단 비율 및 분산의 가설검정

:: Keywords 모집단 비율의 가설검정
모집단 분산의 가설검정



모집단 비율의 가설검정

- **모집단 비율에 대한 가설검정**

  모집단에 대한 특성을 비율로 가늠하여 검정하는 것

→ 조사 목적에서 목표로 하는 비율을 기준으로 제시하고, 조사 결과를 기준으로 제시한 비율과 비교하여 높거나 혹은 낮으면 채택 또는 기각

ex. 4번 타자가 되기 위해서는 3할(30%)을 넘어야 한다.

혹은 3할 3푼(33%)을 넘어야 한다는 등의 기준을 제시

모집단 비율의 가설검정

- **모집단 비율에 대한 가설검정**

ex. 전자제품의 AS 센터에서 100명의 고객이 서비스를 받았다고 할 때,

과연 몇 %의 고객이 만족감을 느꼈을지 확인하는 조사

표본에서 구한 비율을 $\hat{p}$, 모집단의 비율을 p라 하면,

$$H_0 : \hat{p} = p$$

$$H_1 : \hat{p} \neq p \quad \text{(양측검정)}$$

$$H_1 : \hat{p} < p \quad \text{혹은} \quad \hat{p} > p \quad \text{(단측검정)}$$

모비율의 검정통계량은 $Z = \dfrac{p - p}{\sqrt{\dfrac{p(1-p)}{n}}}$

귀무가설을 기각하려면 양측검정에서는 $Z > z_{\frac{\alpha}{2}}, Z < z_{\frac{\alpha}{2}}$,

단측검정에서는 $H_1 : \hat{p} < p$이면 $Z < z_\alpha$이고,

$H_1 : \hat{p} > p$이면 $Z > z_\alpha$이다.

모집단 비율의 가설검정

예제 7-4 모집단 비율의 가설검정

준비파일 | 7장_모집단 비율의 가설검정.xlsx

S전자에 AS 센터 서비스에 대한 불만이 민원으로 접수되어, AS 센터를 방문한 고객을 대상으로 서비스 만족도를 조사하고자 한다. 방문한 고객의 80% 이상이 서비스에 만족한다면 서비스에 관한 재교육을 실시하지 않고, 그렇지 않다면 재교육을 실시할 예정이다. 무작위로 선택한 100명의 고객을 대상으로 만족도를 조사한 결과, 78명의 고객만 서비스에 만족한다고 했다. 이에 대해 유의수준 0.05에서 검정하라.

모집단 비율의 가설검정 (예제 풀이)

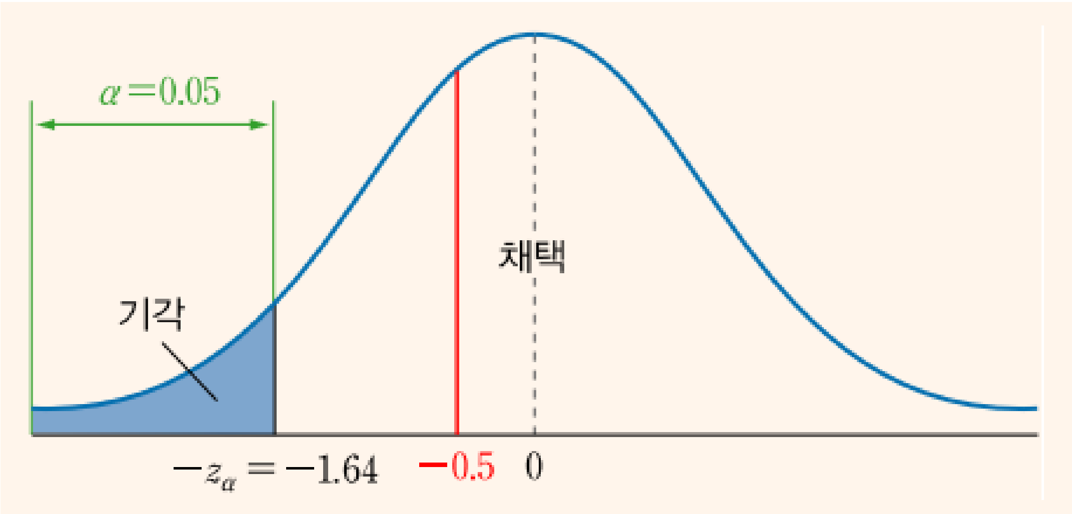
대립가설은 $p < 0.8$로 설정

80% 이상의 고객이 만족하다는 답변을 한다면 굳이 이러한 조사를 하지 않을 것이기 때문.
따라서 80%에 미치지 않는다는 것을 확인하는 단측검정

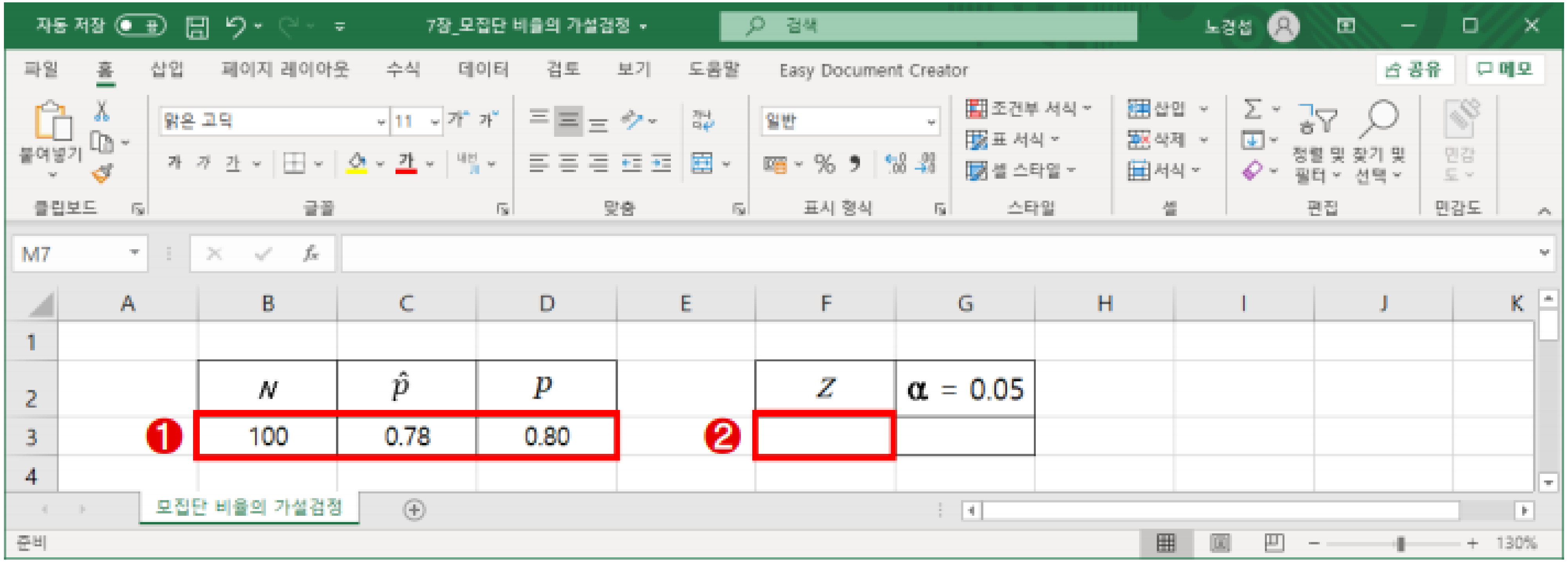
귀무가설은 $H_0 : p = 0.8$, 대립가설은 $H_1 : p < 0.8$

$$N = 100,\ p = 0.8,\ \hat{p} = 0.78,\ \alpha = 0.05$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.78 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{100}}} = -0.5$$

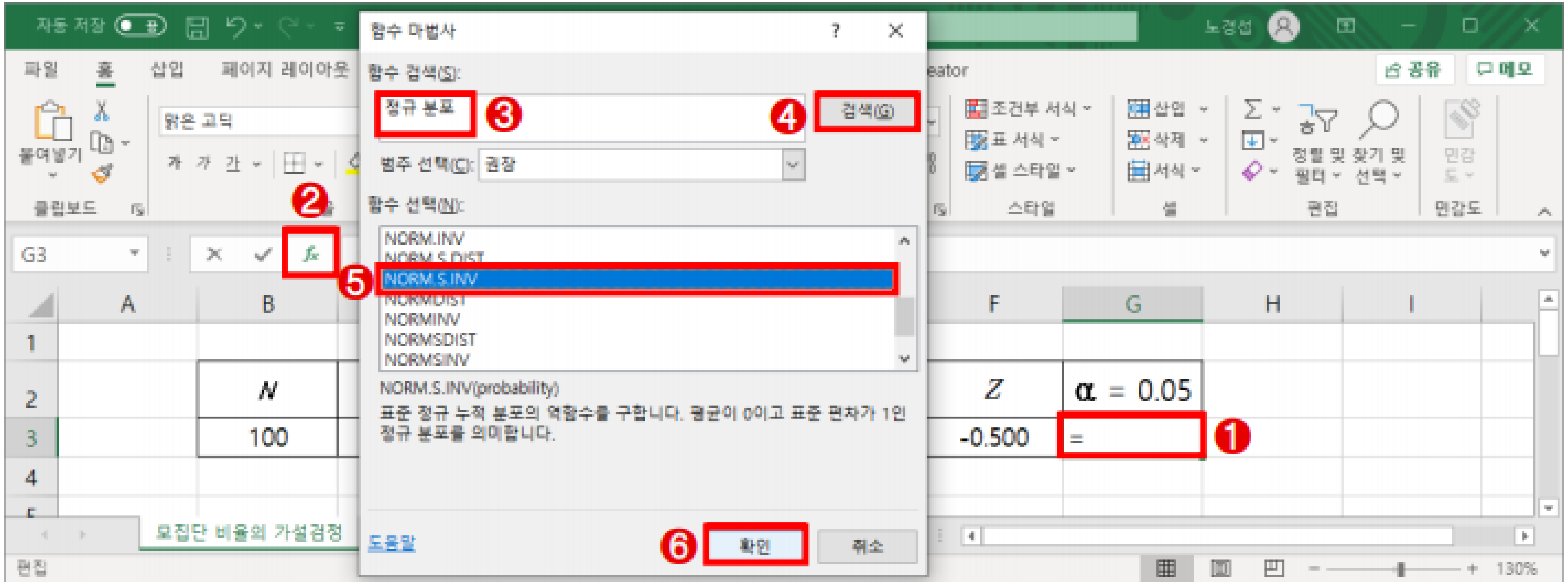


모집단 비율의 가설검정 (예제 Excel 풀이)



① ‘N’은 ‘’100’, ‘$\hat{p}$’은 ‘0.78’, ‘p’는 ‘0.80’을 입력한다.
② F3셀에 검정통계량을 구하는 계산식 ‘=((C3-D3)/SQRT((D3*(1-D3))/B3))’을 입력한다.

모집단 비율의 가설검정 (예제 Excel 풀이)

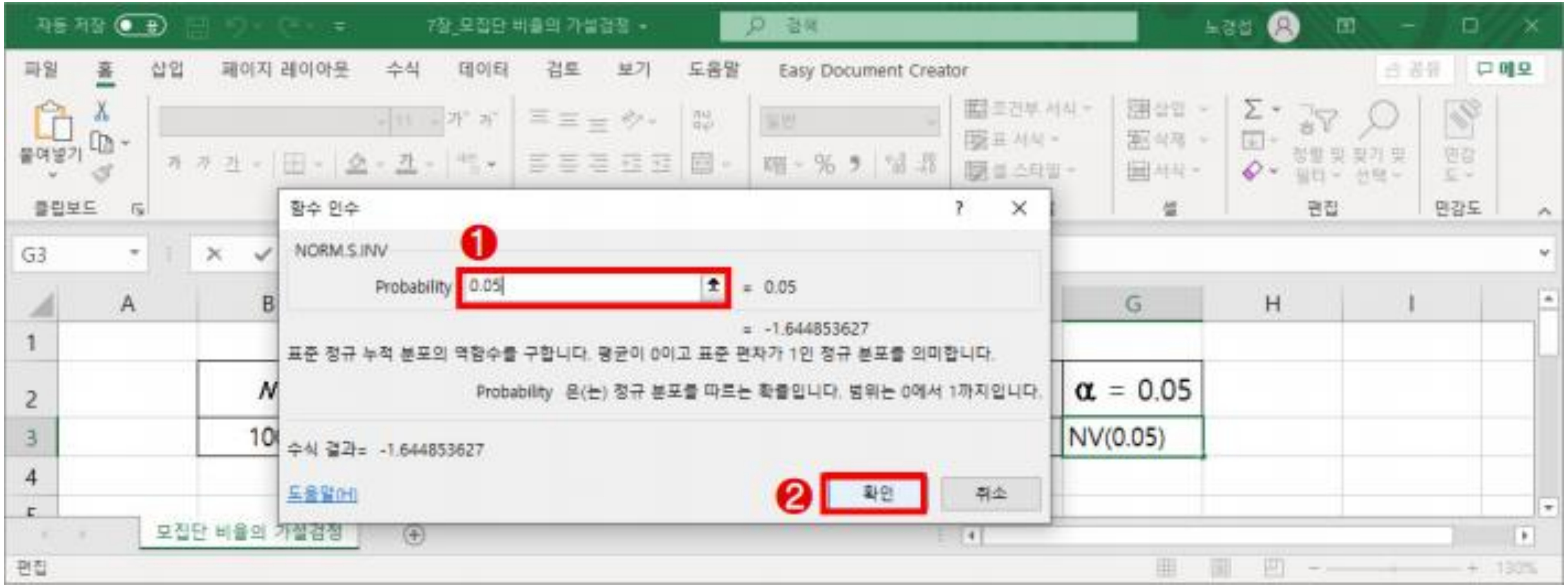


❶ 셀 포인터를 G3셀에 위치하고 ❷ f_x 를 클릭

❸ '함수 검색(S):'에 '정규분포'를 입력한 후 ❹ [검색]을 클릭

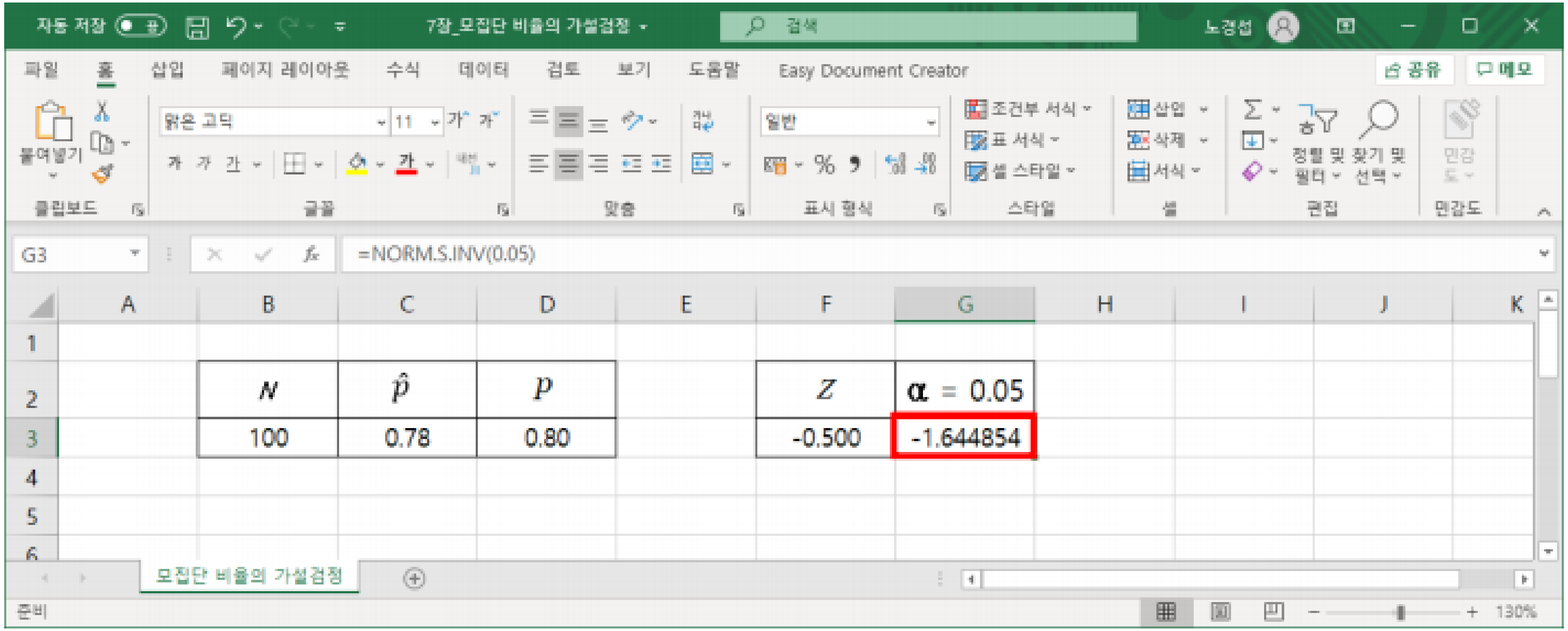
❺ '함수 선택(N):'에서 'NORM.S.INV' 함수를 선택한 후 ❻ [확인]을 클릭

모집단 비율의 가설검정 (예제 Excel 풀이)



❶ 'Probability'에 0.05를 입력하고 ❷ [확인]을 클릭

모집단 비율의 가설검정 (예제 Excel 풀이_완료)



N	$\hat{p}$	p		Z	α = 0.05
100	0.78	0.80		-0.500	-1.644854

모집단 분산의 가설검정

- **모집단 분산에 대한 가설검정**

  평균에 대해 어느 정도의 산포가 나타나는지를 살펴보는 가설검정

→분산에 대한 가설검정도 모집단 평균의 가설검정과 함께 많이 사용

분산이 σ^2인 모집단으로부터 n개의 표본을 추출해서 표본분산 s^2을 계산하면

이 표본은 자유도가 $(n-1)$인 χ^2분포를 따르며,

검정통계량은

$$\chi^2_{(n-1)} = \frac{(n-1)s^2}{\sigma^2}$$

귀무가설과 대립가설은 $H_0: s^2 = \sigma^2$

$H_1: s^2 \neq \sigma^2$ (양측검정)

$H_1: s^2 < \sigma^2$ 혹은 $s^2 > \sigma^2$ (단측검정)

모집단 분산의 가설검정

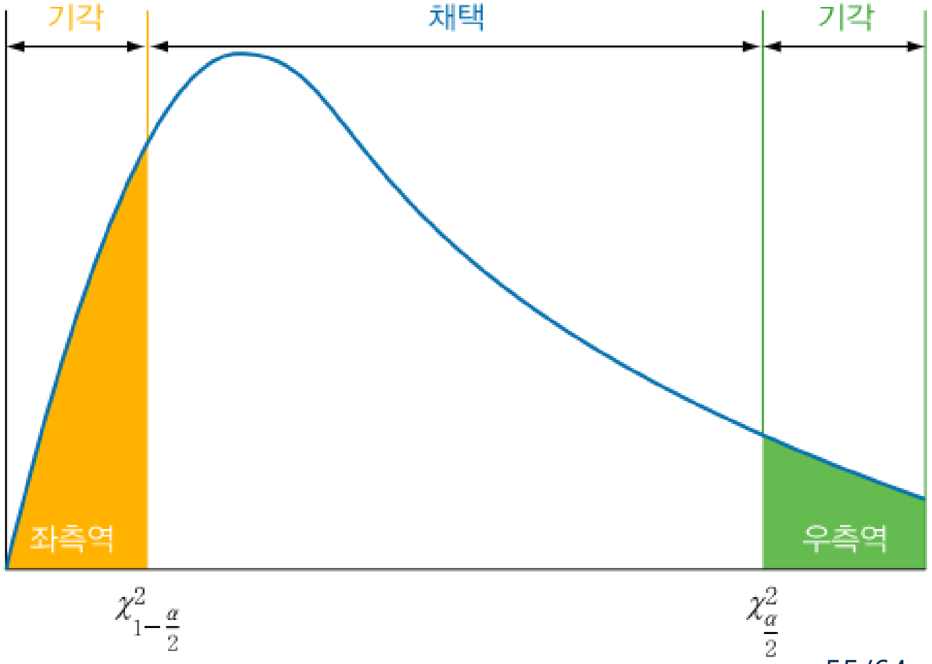
- **모집단 분산에 대한 가설검정**

귀무가설을 기각하려면,

양측검정에서는 $\chi^2 > \chi^2_{\left(n-1,\frac{\alpha}{2}\right)}$, $\chi^2 < \chi^2_{\left(n-1,1-\frac{\alpha}{2}\right)}$

단측검정에서는 $H_1 : \sigma^2 < \sigma_0^2$ 이면 $\chi^2 < \chi^2_{(n-1,1-\alpha)}$,

$H_1 : \sigma^2 > \sigma_0^2$ 이면 $\chi^2 > \chi^2_{(n-1,\alpha)}$



모집단 분산의 가설검정

예제 7-5 모집단 분산의 가설검정 준비파일 | 7장_모집단 분산의 가설검정.xlsx

자동차 제조사 H사에서 개발한 친환경 하이브리드 자동차는 연비가 25km/l, 분산은 6.25로 H사는 이 자동차를 홍보하면서 세계적인 자동차 회사로 거듭나고자 한다. 소비자 단체에서는 H사의 홍보 자료가 사실인지 확인하기 위해 표본으로 50대의 자동차 연비를 직접 측정하였다. 그 결과 연비는 28.4km/l이며 분산이 3.84로 조사되었다. H사의 주장대로 연비의 분산이 6.25가 맞는지 유의수준 0.05에서 검정하라.

모집단 분산의 가설검정 (예제 풀이)

평균연비의 분산이 6.25인지를 확인하는 것이므로,

귀무가설은 $H_0 : \sigma^2 = 6.25$, 대립가설은 $H_1 : \sigma^2 \neq 6.25$

$n = 50,\ \bar{x} = 28.4,\ s^2 = 3.84,\ \alpha = 0.05$

검정통계량을 구하면

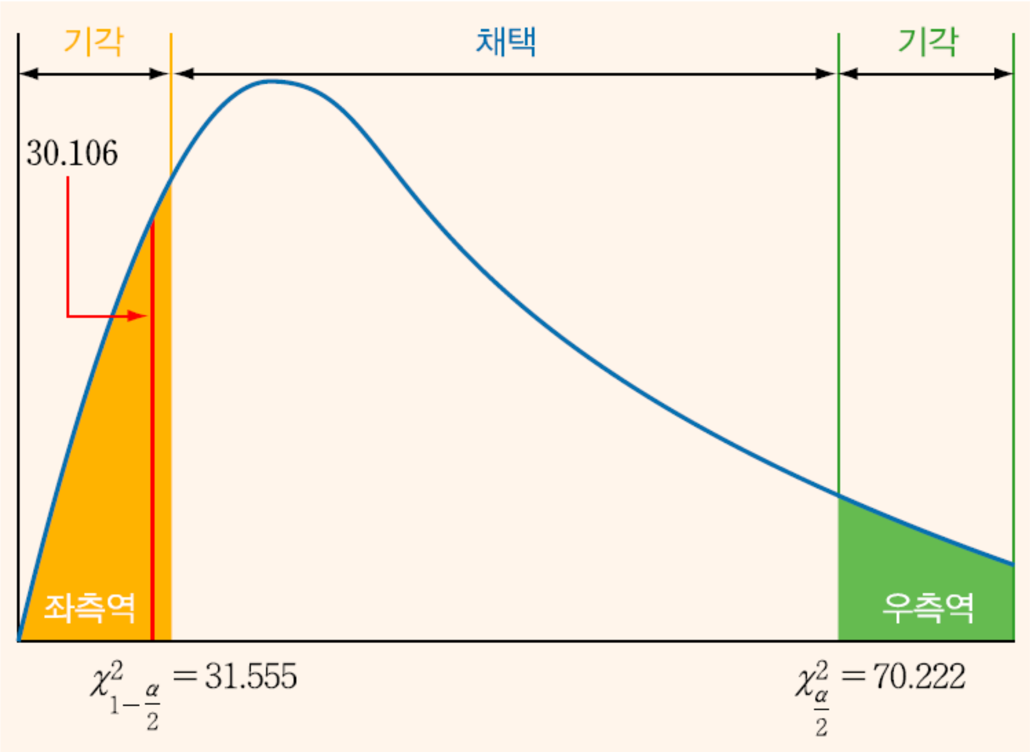
$$\chi^2_{(n-1)} = \frac{(n-1)s^2}{\sigma^2} = \frac{49 \times 3.84}{6.25} = 30.106$$

$$\chi^2_{\left(n-1,1-\frac{\alpha}{2}\right)} = 31.555,\quad \chi^2_{\left(n-1,\frac{\alpha}{2}\right)} = 70.222$$

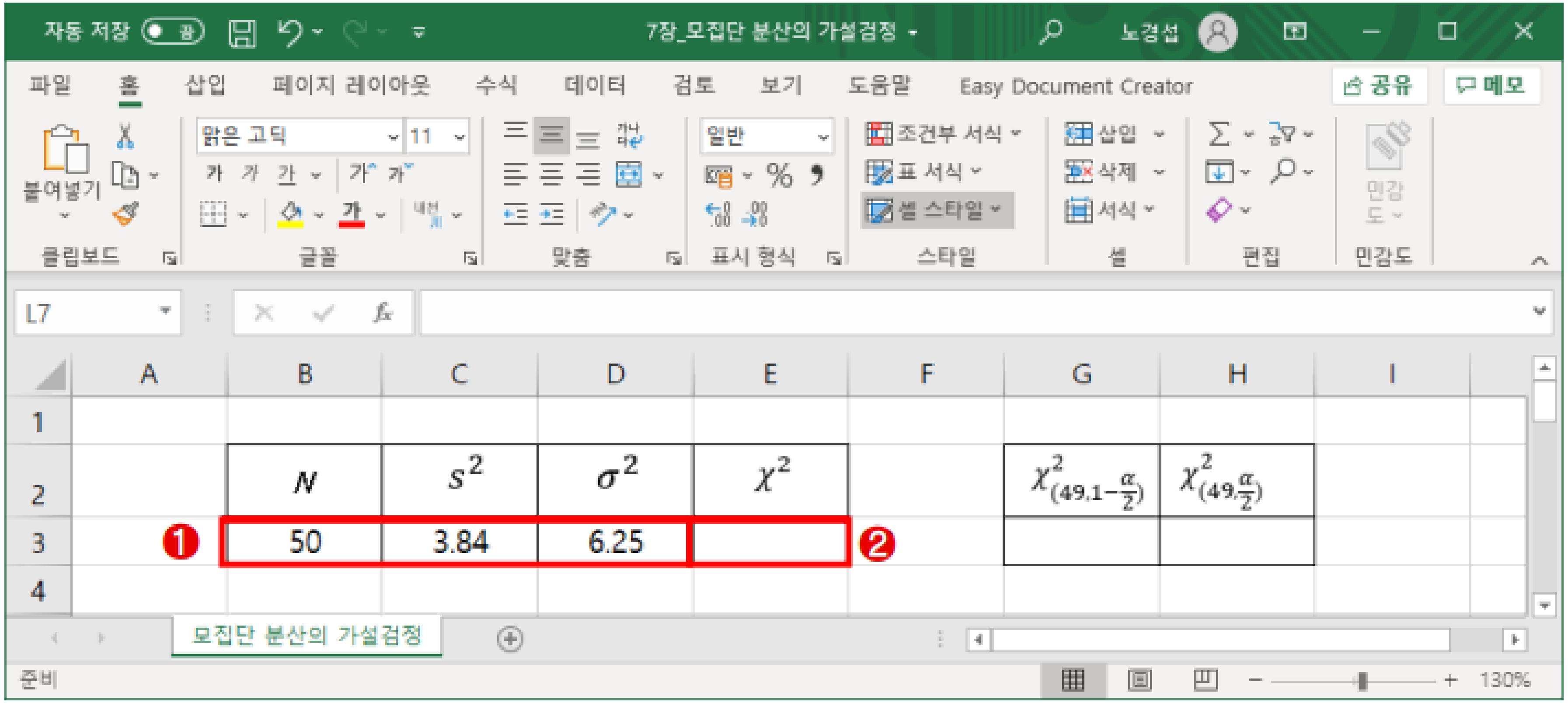
모집단 분산의 가설검정 (예제 풀이)

$\chi^2_{(n-1)}$이 31.555 ~ 70.222의 범위에 있지 않으므로 귀무가설을 기각

제조사는 분산이 6.25라고 주장하고 있지만 실제로는 그렇지 않다.



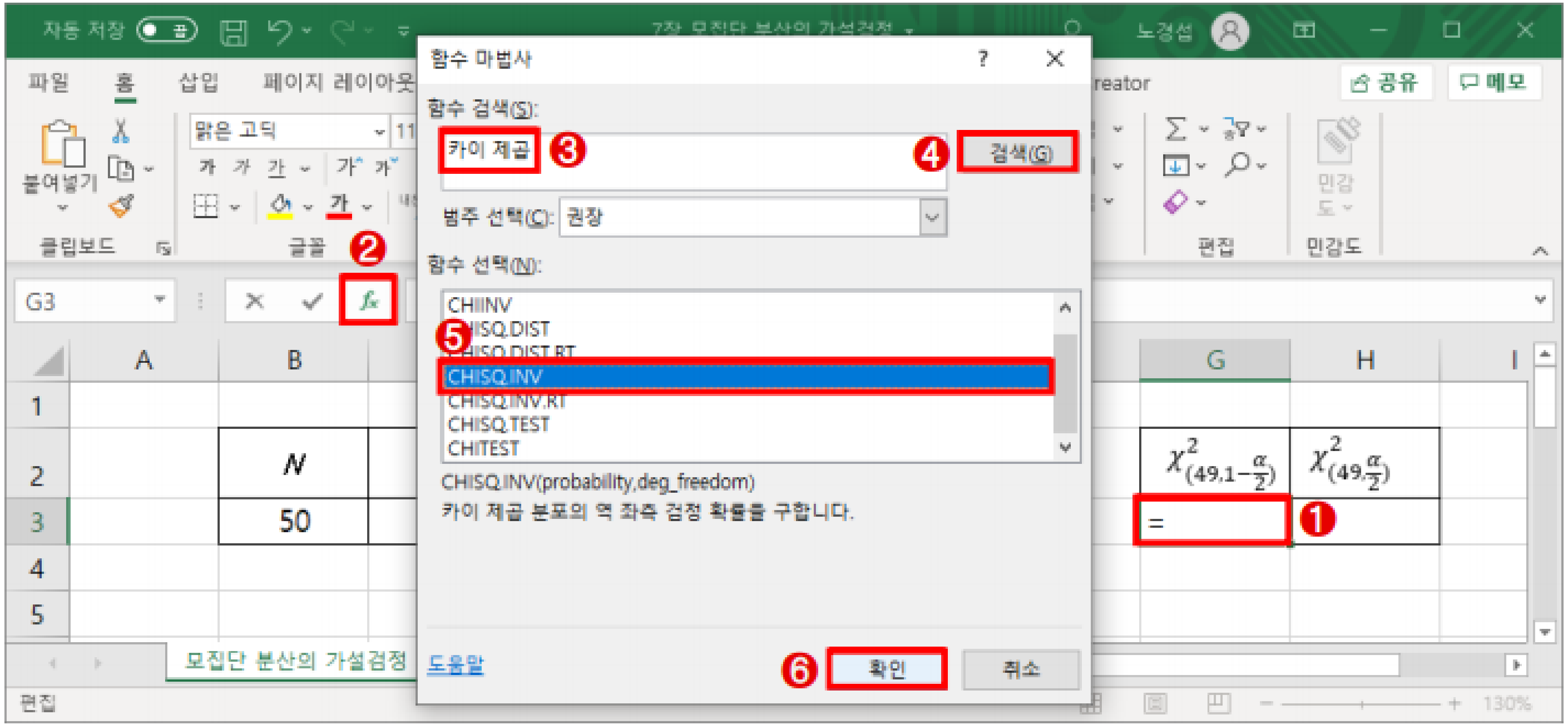
모집단 분산의 가설검정 (예제 Excel 풀이)



❶ 'N'은 50, 's^2'은 3.84, 'σ^2'은 6.25를 입력

❷ E3셀에 검정통계량을 구하는 계산식 '=((B3−1)*C3)/D3'를 입력

모집단 분산의 가설검정 (예제 Excel 풀이)



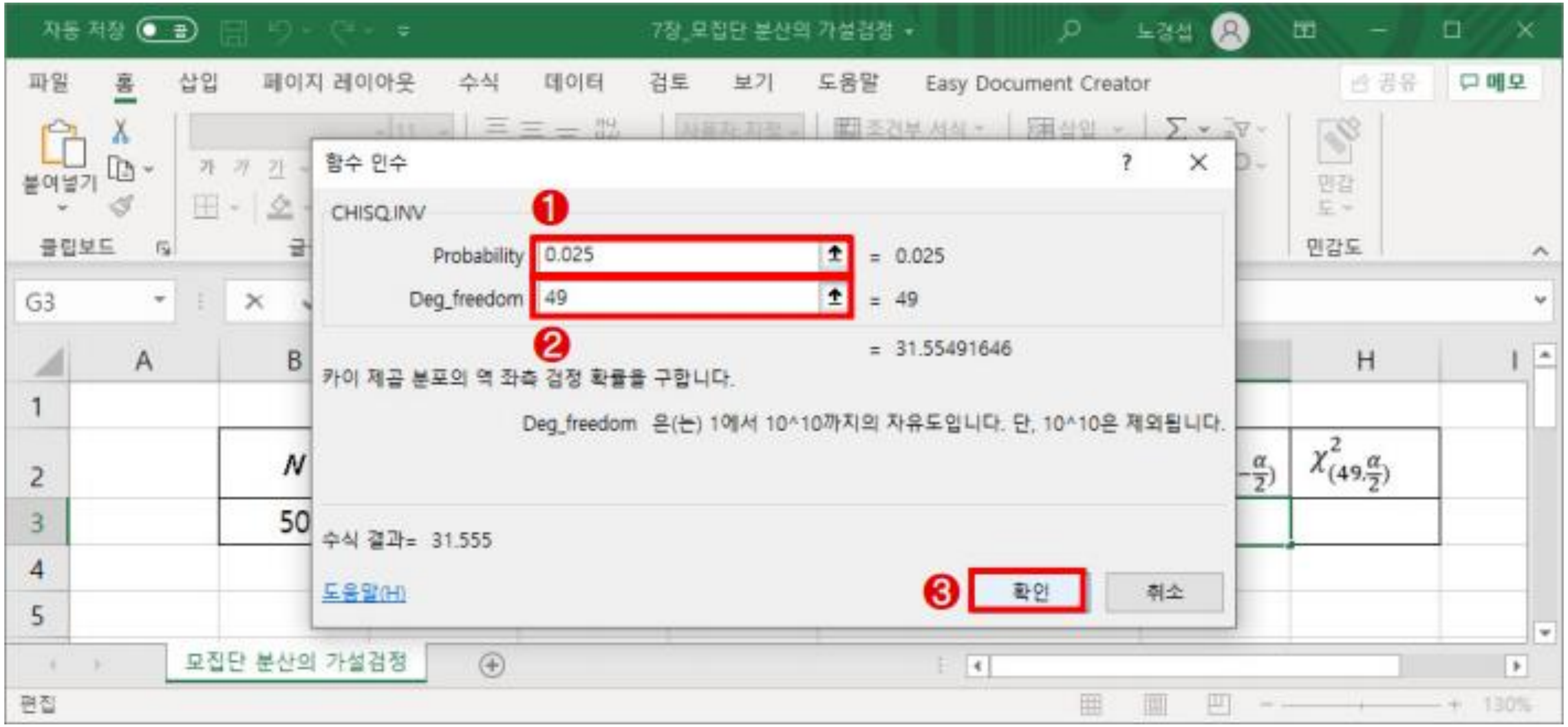
좌측역

❶ 셀 포인터를 G3셀에 위치하고 ❷ f_x 를 클릭

❸ '함수 검색(S):'에 '카이제곱'을 입력한 후 ❹ [검색]을 클릭

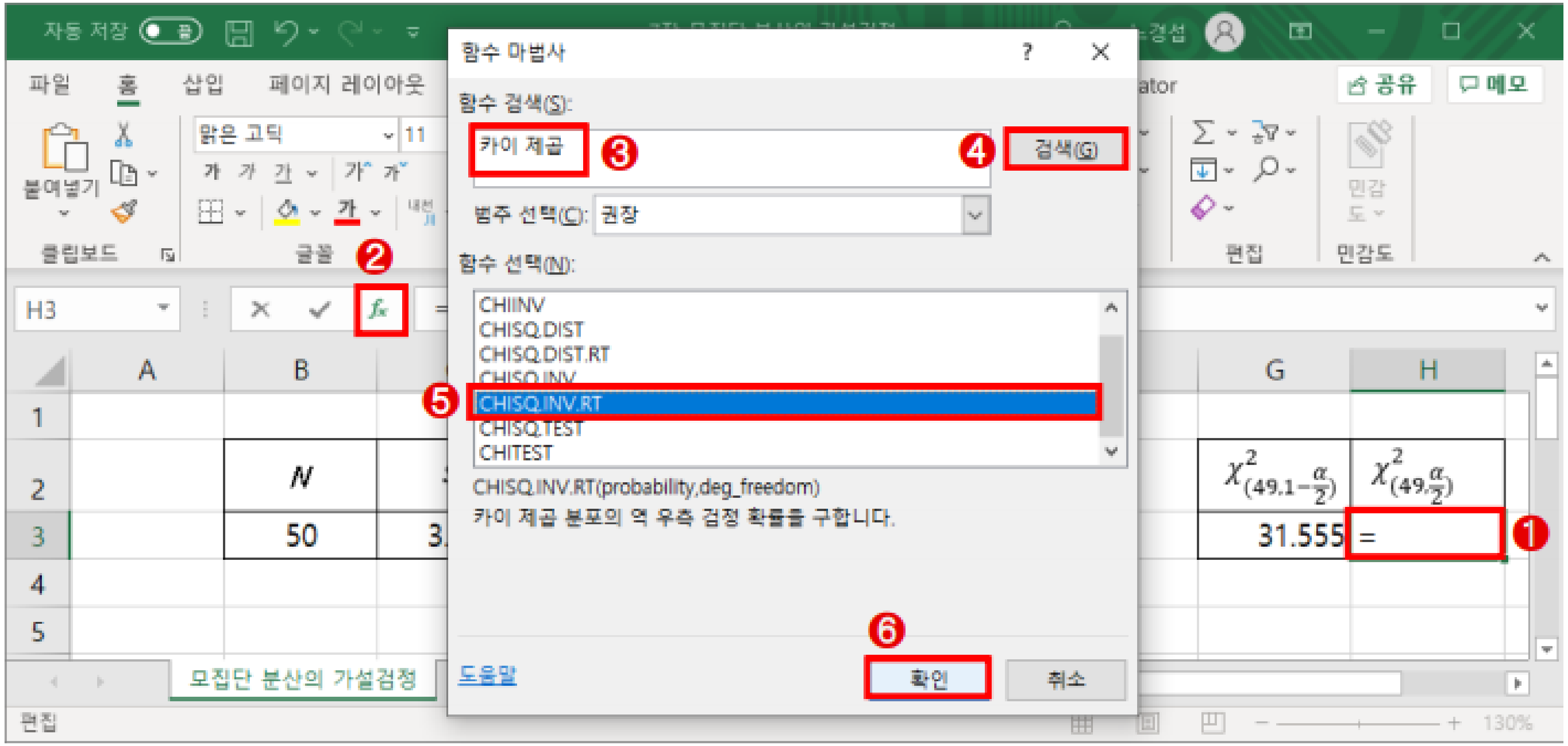
❺ '함수 선택(N):'에서 'CHISQ.INV' 함수를 선택한 후 ❻ [확인]을 클릭

모집단 분산의 가설검정 (예제 Excel 풀이)



❶ 'Probability'에 0.025를 입력 ❷ 자유도를 의미하는 'Deg_freedom'에 49를 입력한 후 ❸ [확인]을 클릭

모집단 분산의 가설검정 (예제 Excel 풀이)



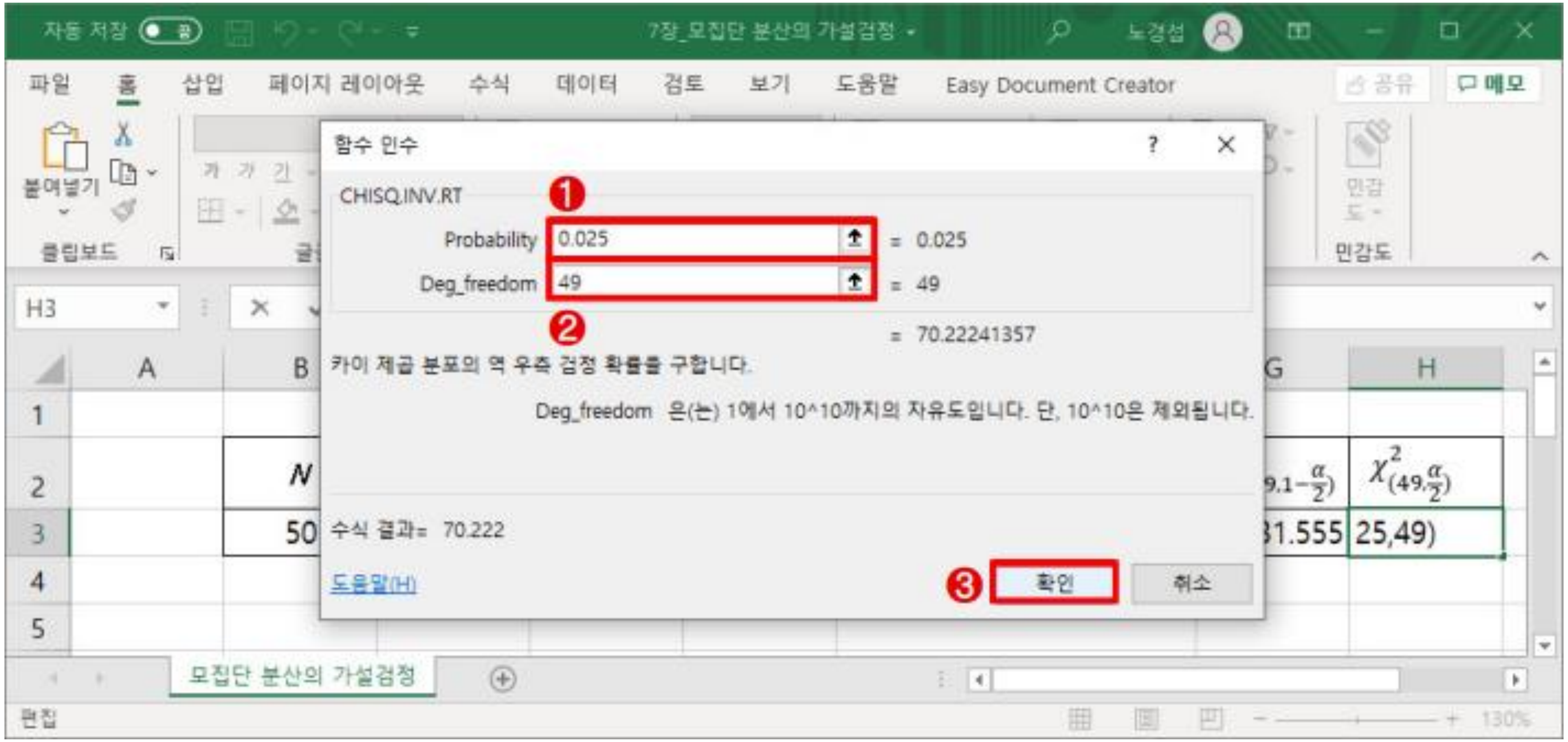
우측역

❶ 셀 포인터를 H3셀에 위치하고 ❷ f_x 를 클릭

❸ '함수 검색(S):'에 '카이제곱'을 입력한 후 ❹ [검색]을 클릭

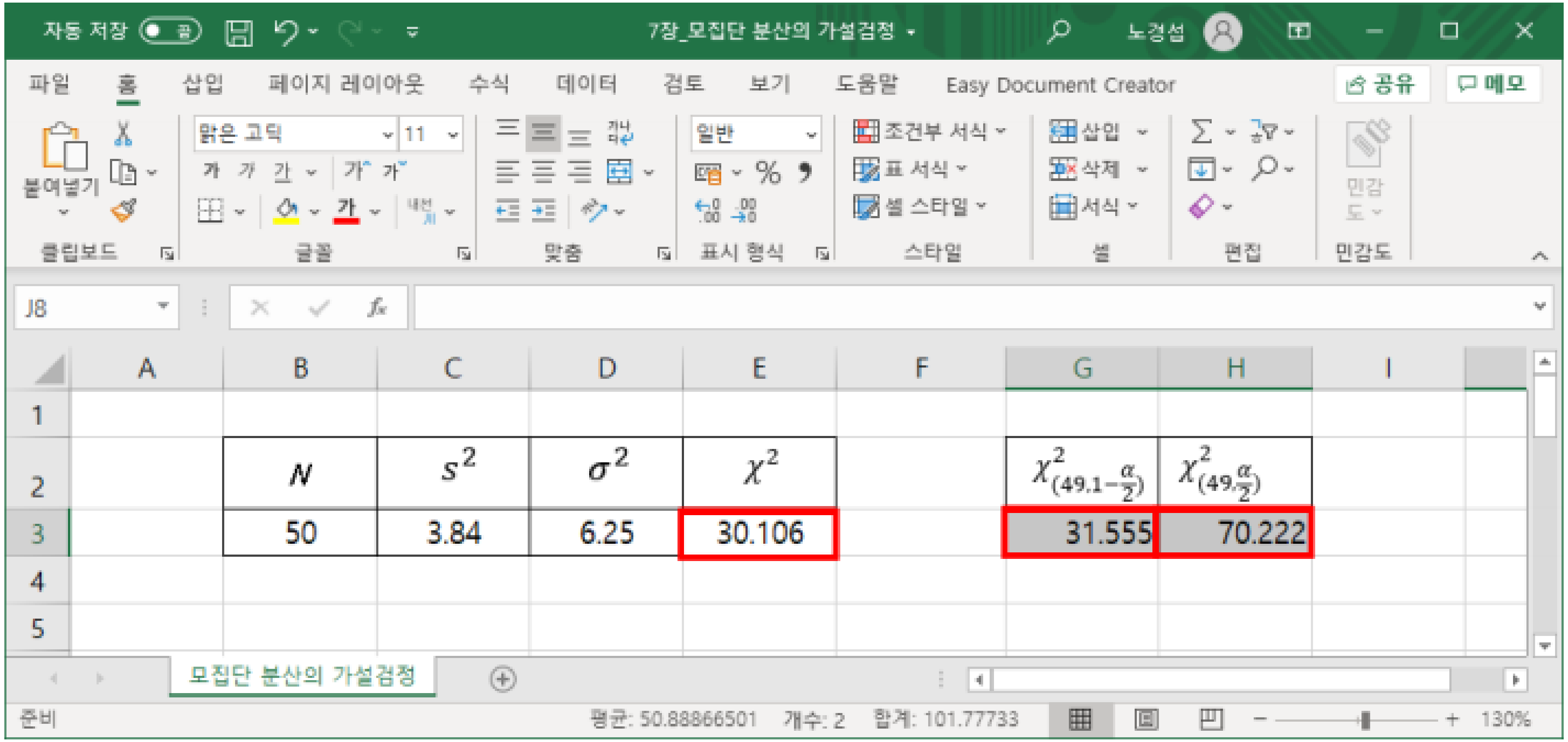
❺ '함수 선택(N):'에서 'CHISQ.INV.RT' 함수를 선택한 후 ❻ [확인]을 클릭

모집단 분산의 가설검정 (예제 Excel 풀이)



❶ 'Probability'에 0.025를 입력 ❷ 자유도를 의미하는 'Deg_freedom'에 49를 입력한 후 ❸ [확인]을 클릭

모집단 분산의 가설검정 (예제 Excel 풀이_완료)



N	s^2	σ^2	χ^2		$\chi^2_{(49,1-\frac{\alpha}{2})}$	$\chi^2_{(49,\frac{\alpha}{2})}$
50	3.84	6.25	30.106		31.555	70.222

$\chi^2_{(n-1)} = 30.106$과 비교해보면,

귀무가설의 좌측 기각역에 속하므로 대립가설을 채택

Q&A

통계학, 제대로 시작하자!

통계의 쓰임을 이해하고,
실제로 활용할 수 있어야 한다.