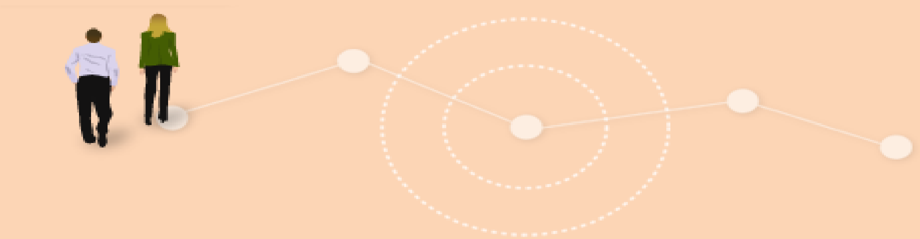
Chapter 05

확률분포 Probability Distribution

목 차

01 확률분포

:: Keywords 확률분포 | 균등분포 | 정규분포 | 표준정규분포



확률분포

타석에 들어선 타자의 운명은?

[그림 5-1] 타석에 선 타자

이번에 타석에 들어선 타자가 안타를 칠지, 아웃이 될지 미리 알 수 있을까? 누구나 자신이 응원하는 팀의 선수가 타석에 들어서면 안타를 치길 원할 것이다. 일반 관중의 입장에서 타자의 운명을 예측해보면 어떨까?

보통 한 시합에서 한 선수가 9회 말까지 타석에 들어서는 횟수는 4~5회다. 첫 번째 타자의 타율이 2할 5푼이라 가정하고 첫 번째 타석의 운명을 예측해보자.

이때 확률로 예측하면 거의 맞는데, 타율은 공격 기회가 왔을 때 성공 가능성을 확률로 나타낸 것이다. 만약 타율이 2할 5푼이면 안타가 성공할 가능성이 25%라는 의미다. 공격 횟수가 거듭될수록 타율은 단순한 확률이 아니라 타자가 안타를 칠지, 아웃이 될지에 대한 예측 근거가 된다.

첫 번째 타석에서는 75% 확률로 아웃이 되므로 아웃될 가능성이 높다. 하지만 횟수를 더해가면서 그 확률은 점점 더 정확해지게 되는데, 3번째 타석까지 한 번도 안타를 치지 못했다면 4번째 타석에서는 거의 100% 안타를 치는 것을 확인할 수 있다. 당연히 시즌 초기보다는 시즌 후반으로 갈수록 정확도는 더 올라간다.[1] 이러한 예측이 믿어지지 않는다면, 앞으로 야구 경기를 관람할 때 타율로 확률을 계산해서 안타 가능성을 예측해보기를 권한다. 경기 후반으로 갈수록 타자의 타율이 출루와 아웃의 기준이 된다는 것을 알 수 있을 것이다.

확률분포

■ 확률분포의 정의

확률분포(probability distribution)는 미래에 발생할 사건에 대해 확률을 나열한 것

→ 확률분포는 그래프나 표로 나타낸다.

여러 가지 대안 가운데 하나를 선택해야 하는 경우에는 정보의 양이 많을수록 의사결정을 하는 데 유리하다.

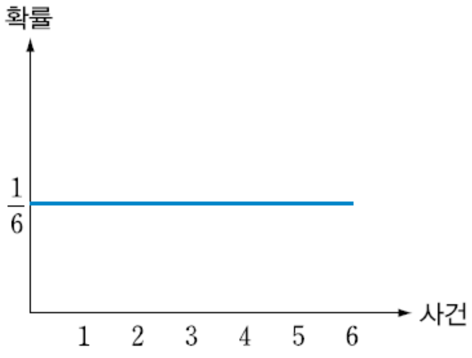
→ 4장에서 살펴본 월드컵과 치킨집의 사례에서 치킨집 점주가 위험 기피자(risk-averter)라면 135마리보다 더 적게 준비하여 낭비를 피할 것이며, 반대로 위험 선호자(risk-lover)라면 135마리보다 좀 더 준비하여 적극적으로 판촉 활동을 벌일 수도 있다.

확률분포

Ex. 주사위를 던지기

→ 주사위를 던지면 1, 2, 3, 4, 5, 6의 6개 중 하나의 결과가 발생

이에 대한 확률을 다음과 같이 나타낼 수 있다.



(a) 확률분포도

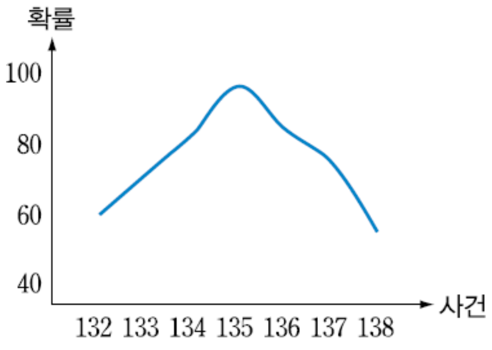
사건	1	2	3	4	5	6
확률	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(b) 확률분포표

cf. 치킨집의

판매량 예측에 대한

확률분포를 나타내면



(a) 확률분포도

사건	132	133	134	135	136	137	138
확률	60.1%	71.3%	82.2%	95.8%	83.7%	74.5%	55.2%

(b) 확률분포표

균등분포

■ 균등분포 (uniform distribution)

이러한 예와 같이 과거의 경험이 미래를 예측하는 데 어떤 영향도 미치지 않으며, 나타날 가능성이 모두 동일한 분포

→ 변수의 특성에 따라 이산 균등분포(discrete uniform distribution)와 연속 균등분포(continuous uniform distribution)로 구분

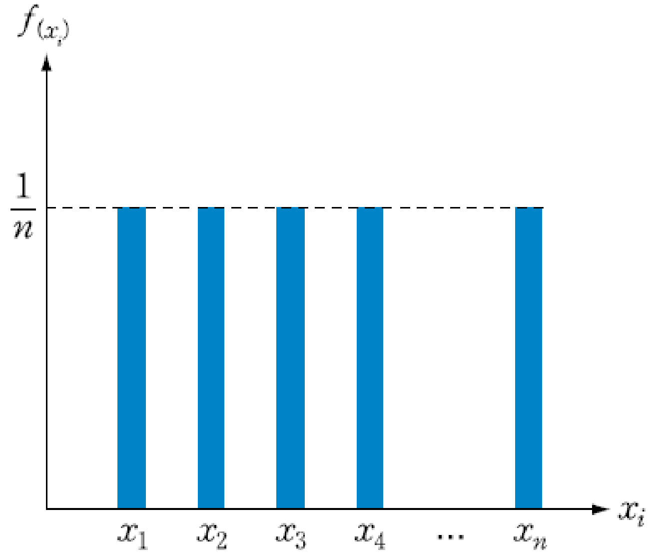
균등분포

■ 이산균등분포

정의된 구간에서 확률분포 함수의 모든 확률이 동일한 분포

확률변수가 $x_1, x_2, x_3, \cdots, x_n$으로 n개 일때 x_i의 확률은$\frac{1}{n}$

→ 확률함수 X의 확률함수는 $f_x(\chi) = \frac{1}{n}$



균등분포

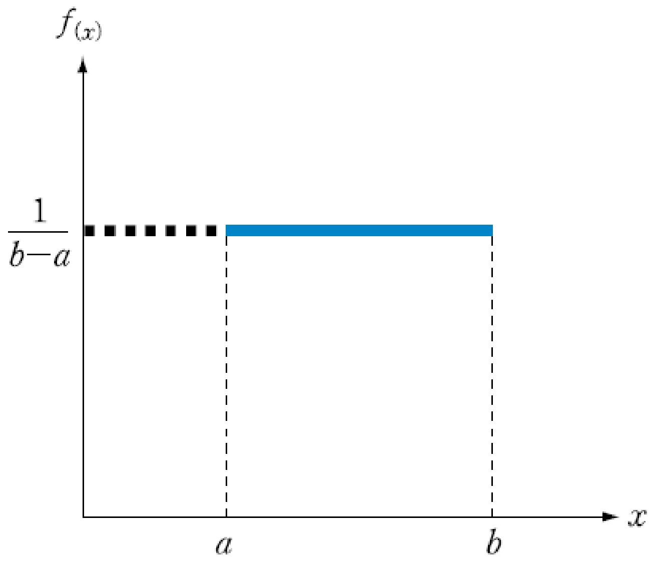
■ 연속균등분포

특정 범위 내에서 모든 확률함수가 동일한 분포

확률변수가 X가 a와 b구간에서 균등분포를 가진다면

→ 확률변수가 X의 확률함수는 $\frac{1}{b-a}(a \le x \le b)$이 되고

확률변수 X의 확률함수는 $f_X(x) = \ X \sim U(a, b)$ 로 나타낸다.



균등분포

예제 5-1 균등분포의 확률 계산하기

스톱워치를 작동시킨 후 편안히 잠을 자고 일어나서 시계를 확인했을 때, 분침과 상관없이 초침이 45 ~ 55 사이에 있을 확률을 구하라.

[그림 5-6] 스톱워치

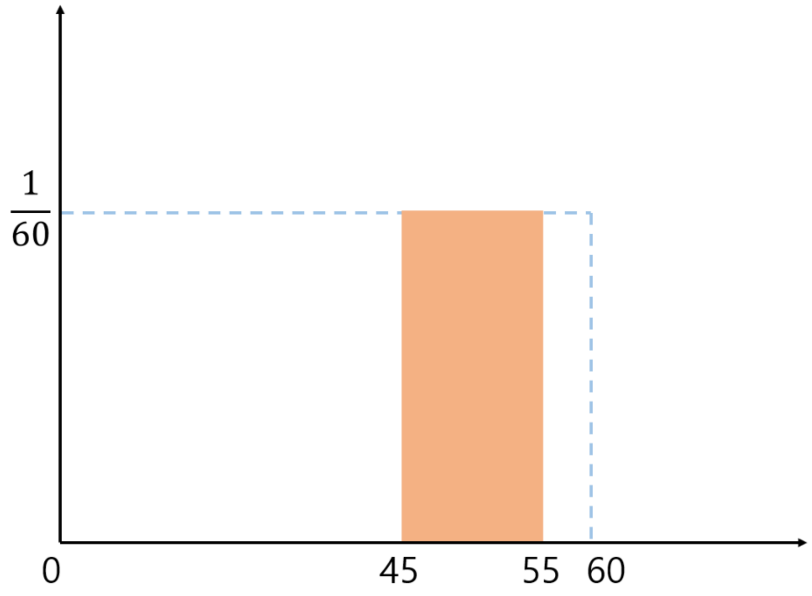
균등분포 (예제 풀이)

전체의 눈금 → 60개

$\therefore$ 눈금이 머무를 확률 = $\frac{1}{60}$

초침이 45~55 사이에 있어야 하므로, 간격은 55-45=10

$$\therefore P(45 \le X \le 55) = (55-45) \times \frac{1}{60} = \frac{1}{6}$$



정규분포와 표준정규분포

■ 정규분포 (normal distribution)

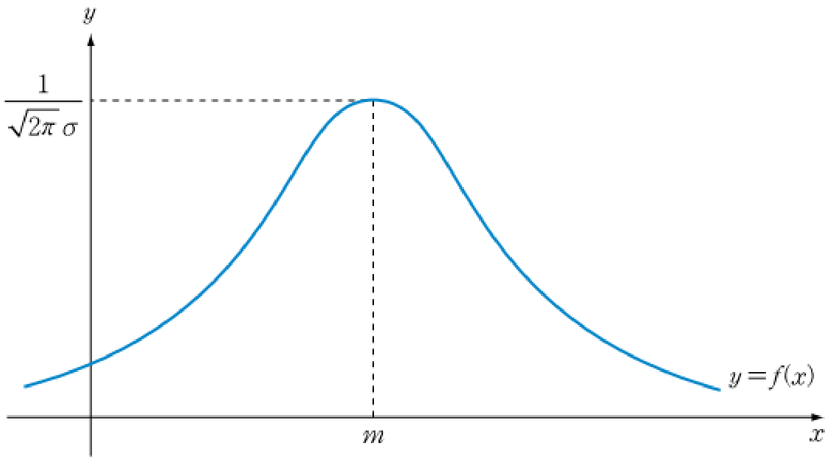
축적된 데이터를 기준으로 미래를 예측할 수 있는 분포

→ 통계학에서 가장 많이 사용되는 분포이며,

평균과 분산만으로 그 특성을 모두 설명할 수 있어 아주 편리

평균이 m, 분산이 σ^2인 정규분포의 확률함수는 $e = 2.71828\cdots$ 인 무리수를 사용하여

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (-\infty < x < \infty)$$



[그림 5-8] 정규분포

정규분포와 표준정규분포

Note **정규분포의 성질**

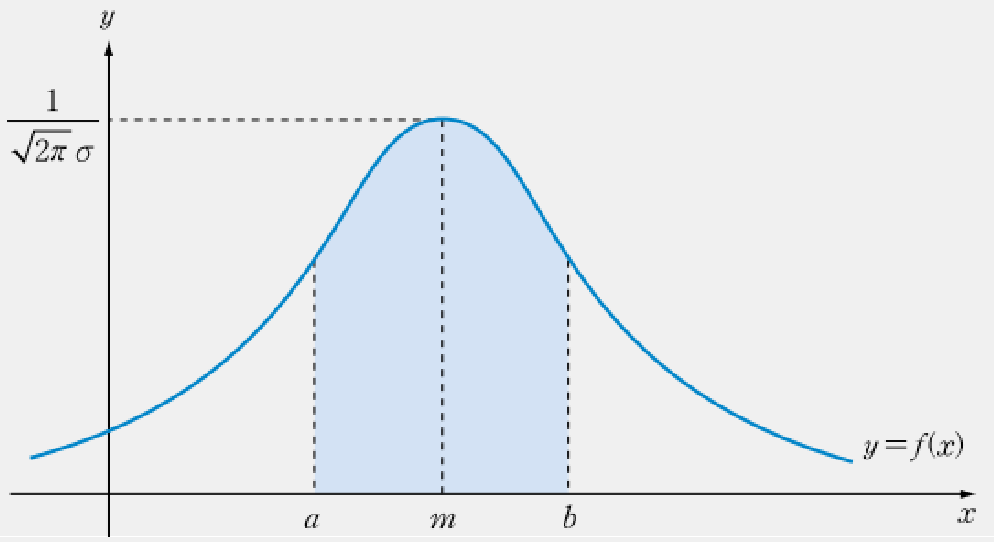
확률변수 X의 확률밀도함수가 $f_X(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}$ $(-\infty < x < \infty)$일 때,

❶ 어떤 실수 x에 대해 $f(x) > 0$

❷ 분포곡선은 평균 m을 기준으로 좌우대칭

❸ 곡선과 x축 사이의 넓이는 1

❹ 곡선 내 임의의 a, b가 $a < b$일 때 a와 b에 속할 확률 $P(a \le X \le b)$는 a, b와 곡선 사이의 넓이와 같다.



[그림 5-9] 정규분포곡선 내의 a~b

정규분포와 표준정규분포

■ 표준정규분포(standard normal distribution)

서로 다른 정규분포를 비교할 수 있도록 여러 개의 분포를 어떤 하나의 기준으로 (평균=0, 표준편차1)으로 재배치해서 각 분포를 비교할 수 있도록 표준화된 분포

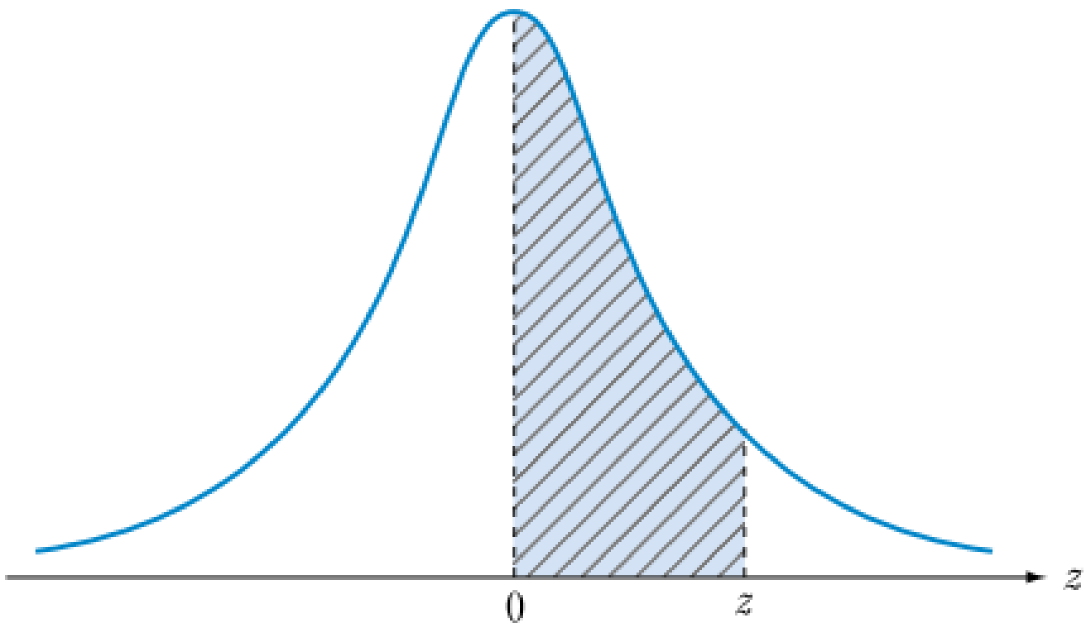
확률변수 X가 정규분포 $N(\mu, \sigma^2)$을 따를 때,

새로운 확률변수 $\frac{x-\mu}{\sigma} = z$로 변환하면

표준정규분포의 확률함수는 $\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

정규분포와 표준정규분포

확률변수 Z의 값이 구간 $(0,\ z)$에 속할 확률 $P(0 \leq Z \leq z)$는



정규분포와 표준정규분포

Note 정규분포의 표준화 과정

정규분포 $N(\mu, \sigma^2)$인 확률변수 X를 표준정규분포인 $N(0,1)$이 되는 $z = \frac{X-\mu}{\sigma}$로 변환하여 표준화하면 서로 다른 대상을 쉽게 비교할 수 있다.

[참고]

표준화하여 (평균)$=0$, (분산)$=1$이 되는 과정을 수식으로 나타내면 다음과 같다.[4]

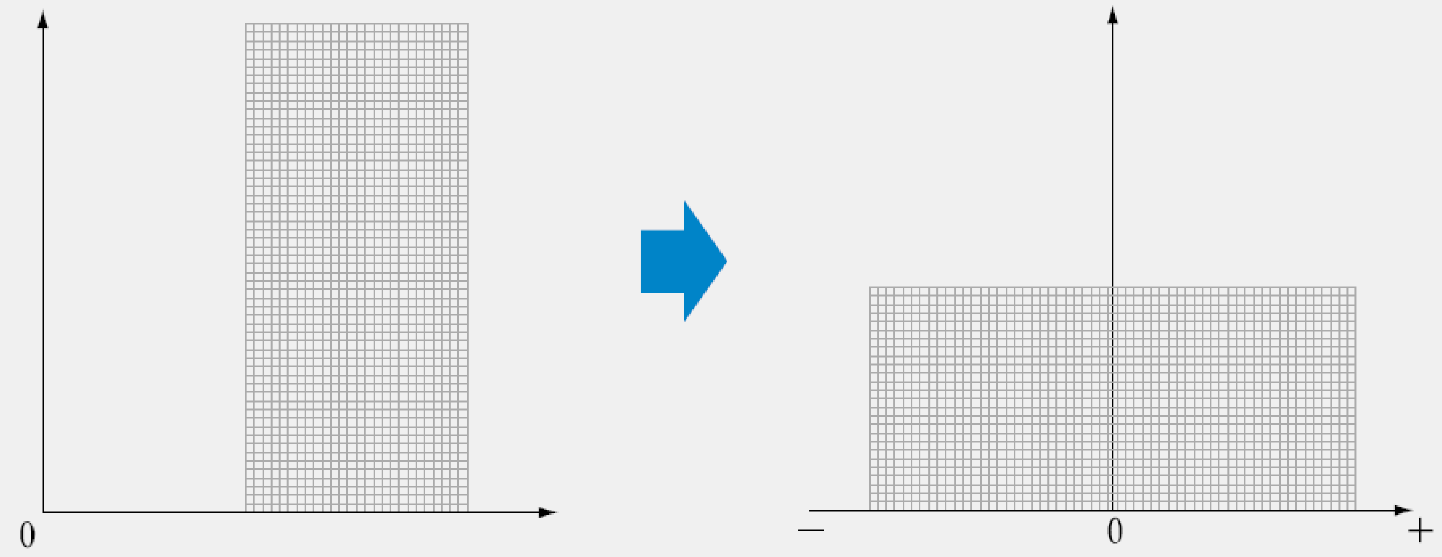
$$\begin{aligned} E(z) &= E\left(\frac{\overline{X}-m}{\sigma}\right) \\ &= \frac{1}{\sigma}E(\overline{X}) - \frac{m}{\sigma} \\ &= \frac{m}{\sigma} - \frac{m}{\sigma} = 0 \end{aligned}$$

$$\begin{aligned} V(z) &= V\left(\frac{\overline{X}-m}{\sigma}\right) \\ &= \frac{1}{\sigma^2}V(\overline{X}) \\ &= \frac{1}{\sigma^2}\cdot\sigma^2 = 1 \end{aligned}$$

정규분포와 표준정규분포

Note **평균 = 0, 분산 = 1로 표준화를 하는 이유는?**

일반적으로 좌우대칭인 정규분포를 많이 사용한다. 그런데 표준화 과정은 왜 필요할까? 단순히 비교 대상 없이 하나만을 기준으로 통계량을 계산한다면 굳이 표준화를 하지 않아도 상관없다. 하지만 서로 비교해야 하는 대상이 있는 경우에는 비교 기준이 필요하다. 예를 들어, 수학 60점과 영어 80점의 성적을 비교할 때 단순히 영어 80점을 더 좋은 성적이라 할 수 있을까? 서로 다른 대상을 비교할 때 기준을 0으로 설정하면 분포가 +와 − 중 어느 쪽으로 치우쳐져 있는지 바로 확인할 수 있다.



[그림 5-9] 표준화에 대한 설명

정규분포와 표준정규분포

[그림 5-9]는 이해를 돕기 위해 정규분포를 직사각형으로 표시한 것이다. 왼쪽의 세로 직사각형은 직접 조사를 수행해서 측정값을 표시한 것이고, 오른쪽의 가로 직사각형은 (평균)=0, (분산)=1로 측정값을 재구성한 것이다. 가로와 세로의 모양만 달라졌을 뿐이며 실제 내부의 면적은 변하지 않았다. 만약 다른 데이터 측정값을 표준화한다면 [그림 5-9]와는 크기와 모양이 다르면서 (평균)=0, (표준편차)=1이 되는 직사각형이 될 것이다. 즉 정규분포 $N(\mu, \sigma^2)$인 확률변수 X를 표준정규분포인 $N(0,1)$이 되는 $z = \dfrac{\overline{X}-\mu}{\sigma}$로 변환하여 표준화하면 서로 다른 대상을 쉽게 비교할 수 있다.

표준화하여 (평균)$=0$, (분산)$=1$이 되는 과정을 수식으로 나타내면 다음과 같다.

$$E(z) = E\left(\frac{\overline{X}-\mu}{\sigma}\right) = \frac{1}{\sigma}E\left(\overline{X}\right) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$V(z) = V\left(\frac{\overline{X}-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V\left(\overline{X}\right) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

정규분포와 표준정규분포

예제 5-2 정규분포의 확률 계산하기 준비파일 | 5장_정규분포.xlsx

A는 서울에서 부산까지 KTX의 운행 시간을 100회 측정하여 평균을 170분, 표준편차를 10분으로 측정했다. 무작위로 KTX를 이용했을 때, 185분보다 더 오래 걸릴 확률을 구하라.

정규분포와 표준정규분포 (예제 풀이)

$P(x > 185)$를 표준정규분포로 변경하면

$$z = \frac{x - m}{\sigma} \Rightarrow z = \frac{185 - 170}{10}, \ z = 1.5$$

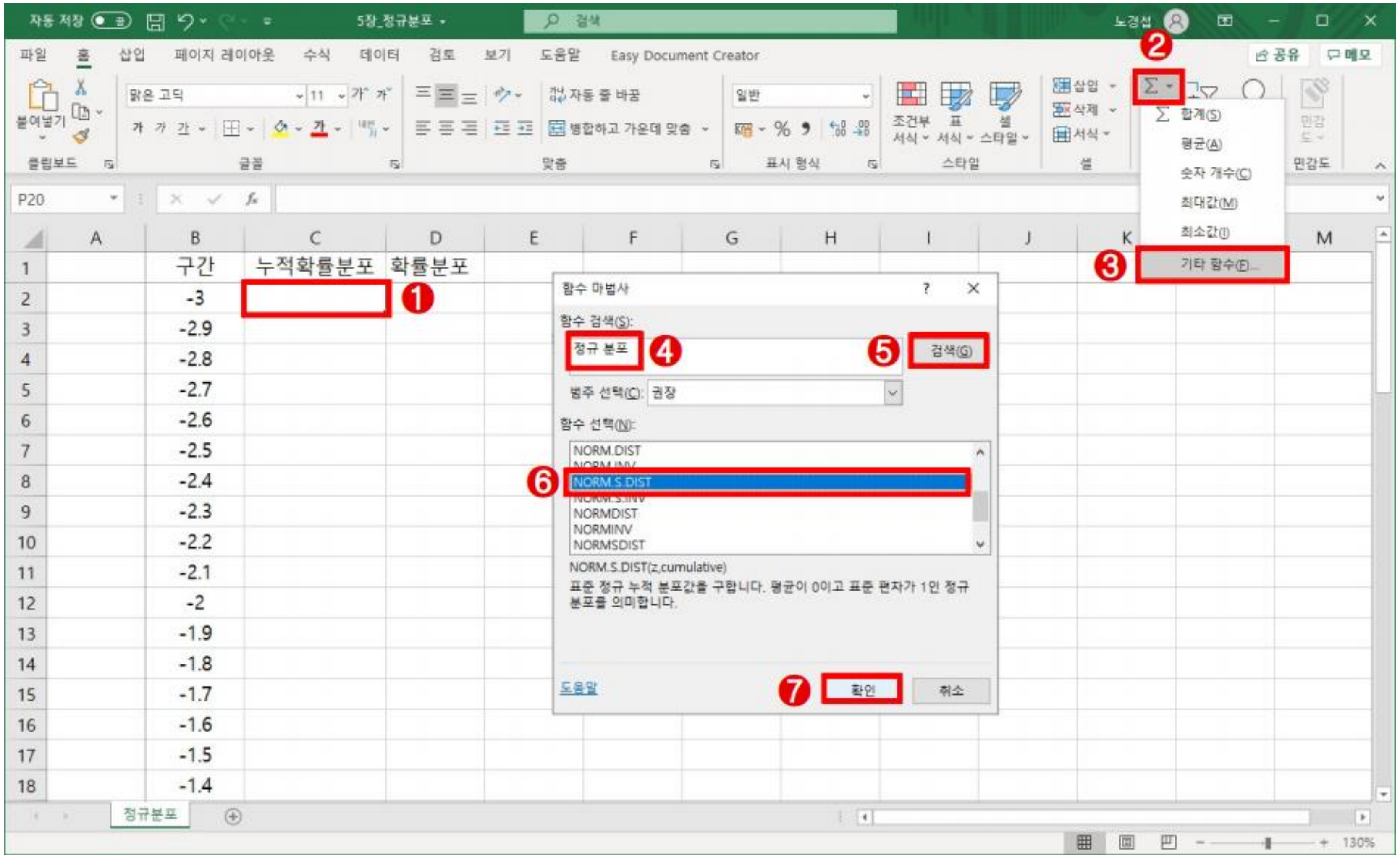
$P(z > 1.5)$

z가 1.5보다 큰 부분의 확률을 구해야하므로,

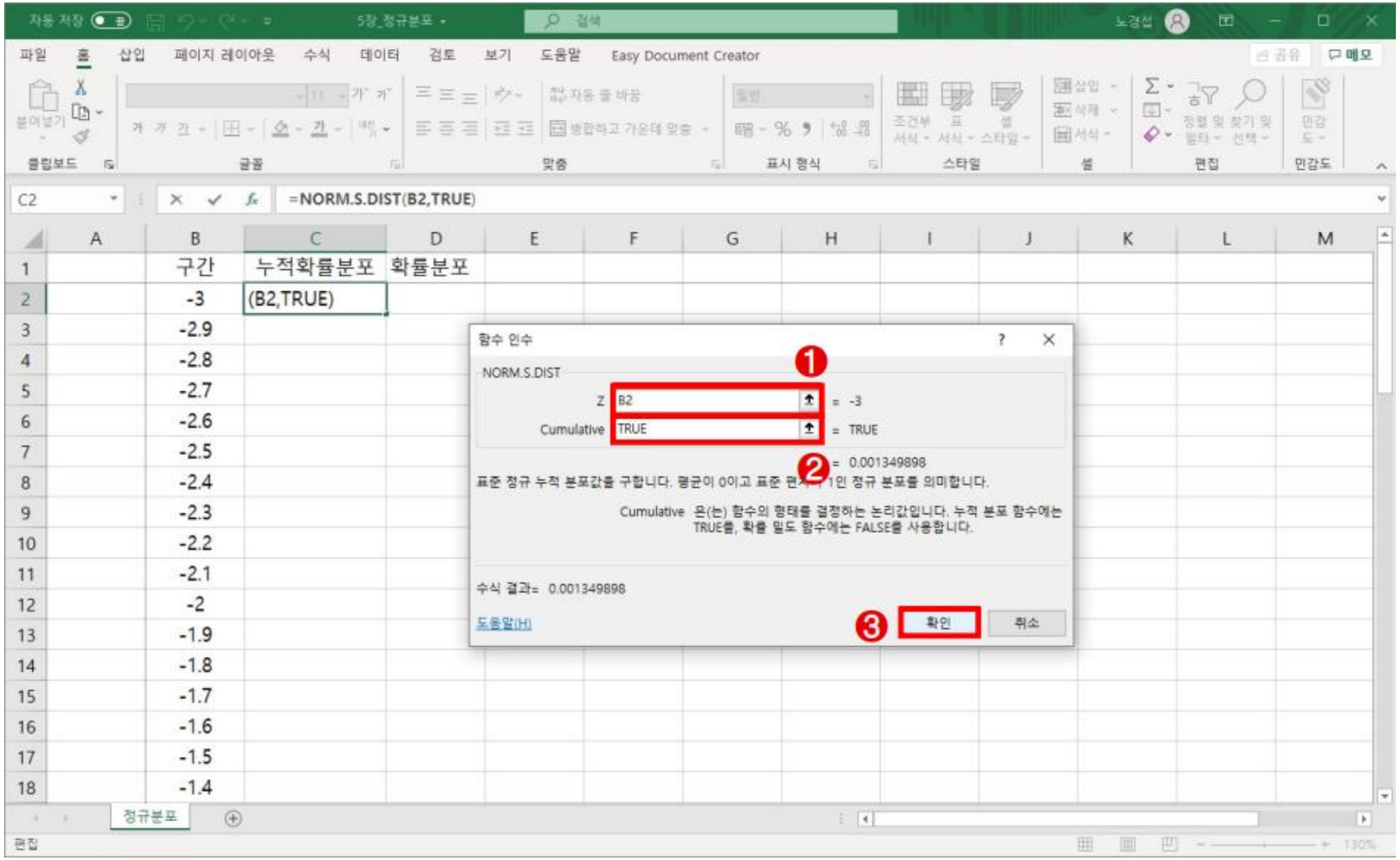
$z = 1.5$일 때까지의 누적확률을 빼야 한다.

따라서 $1 - 0.933193 = 0.066807$이므로 확률은 6.68%

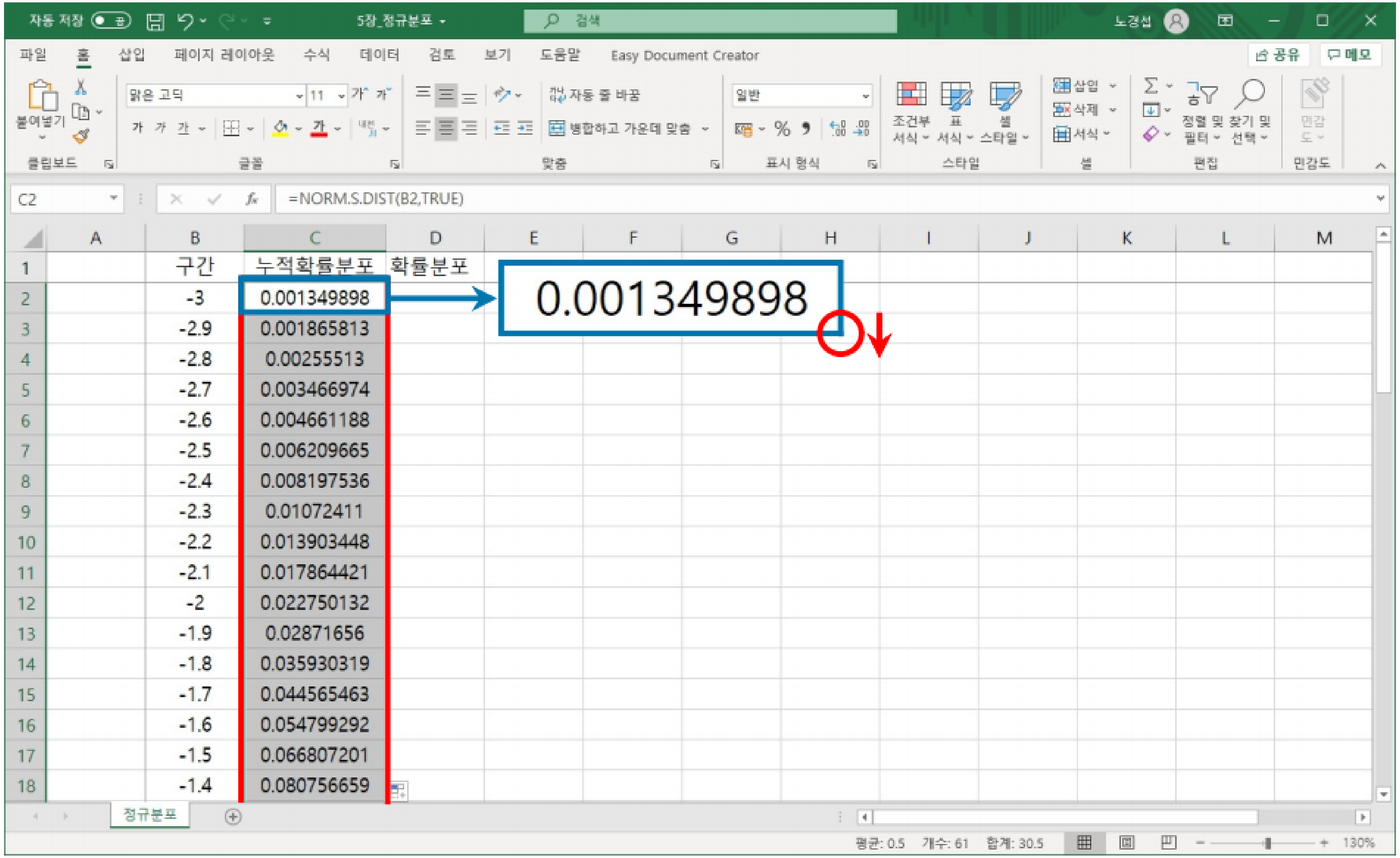
정규분포와 표준정규분포 (예제 Excel 풀이)



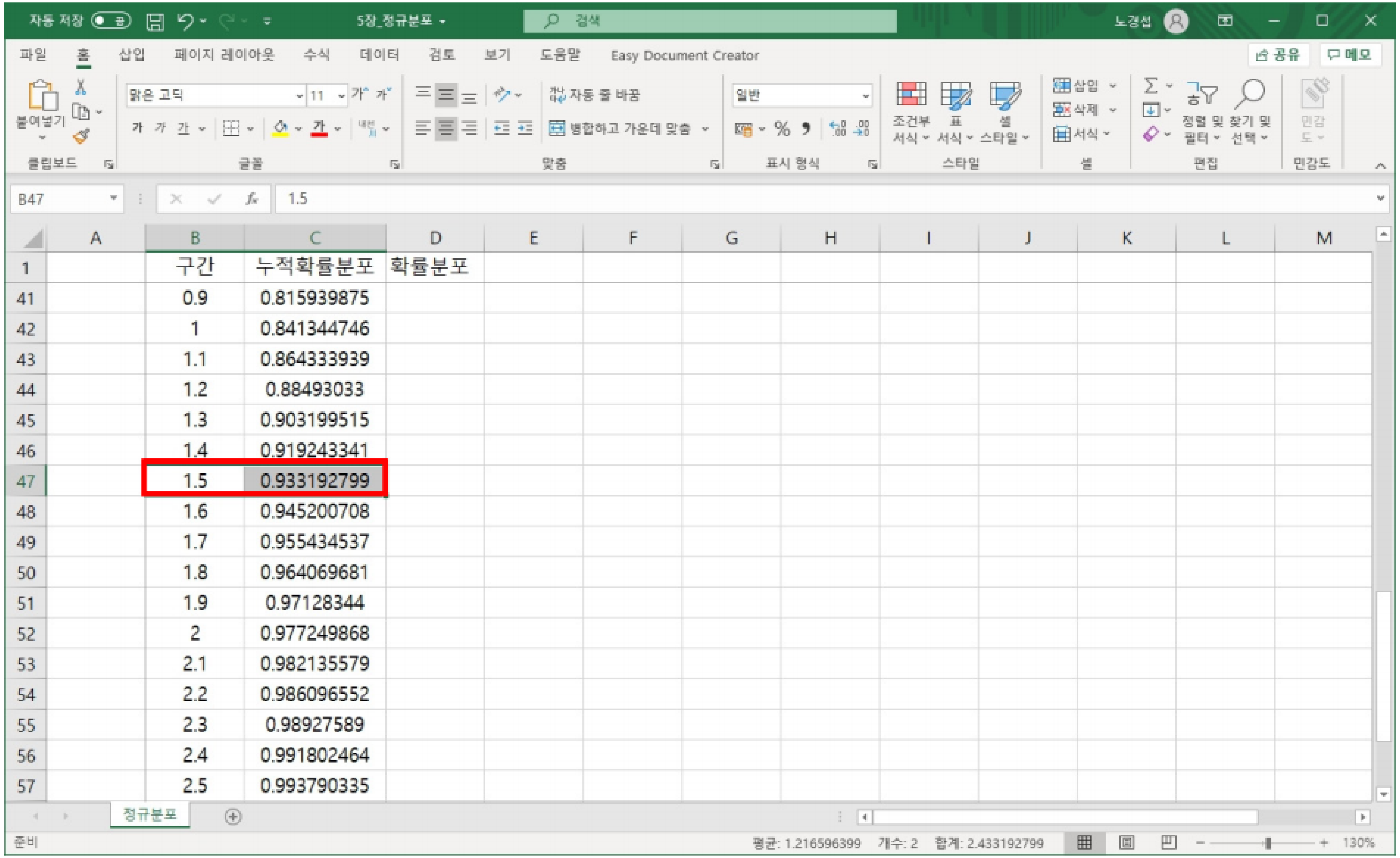
정규분포와 표준정규분포 (예제 Excel 풀이)



정규분포와 표준정규분포 (예제 Excel 풀이)

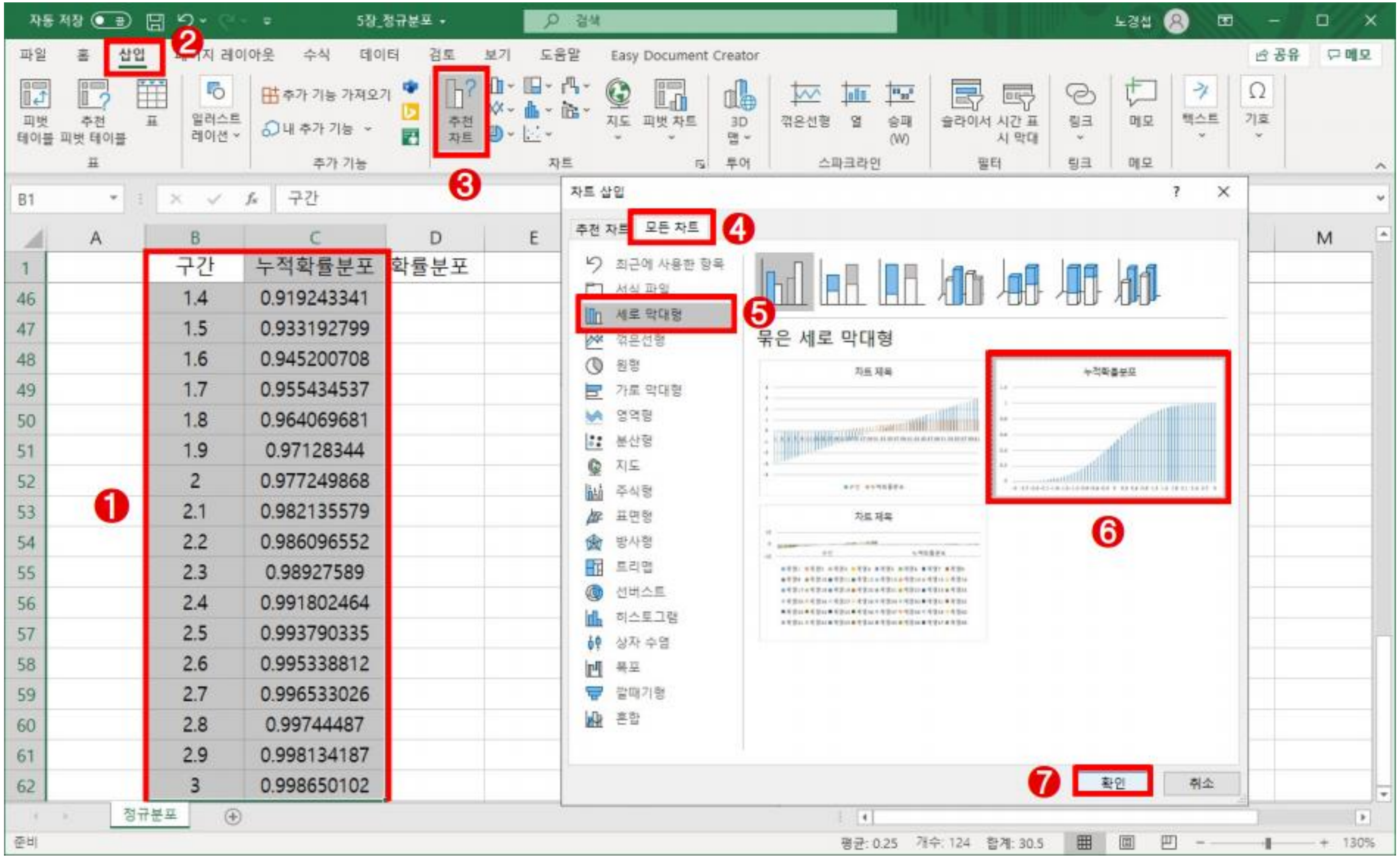


정규분포와 표준정규분포 (예제 Excel 풀이)

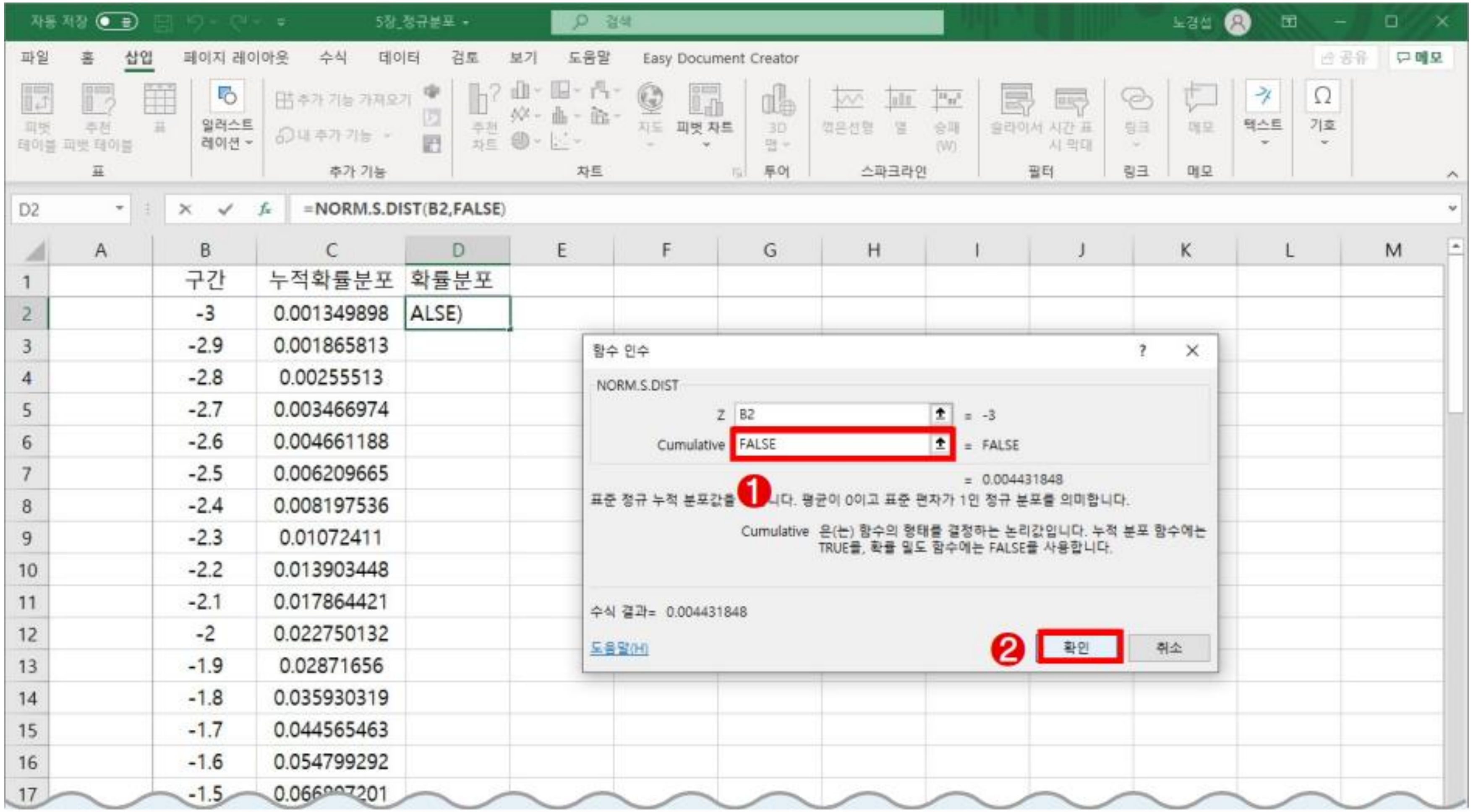


	B	C	D
1	구간	누적확률분포	확률분포
41	0.9	0.815939875	
42	1	0.841344746	
43	1.1	0.864333939	
44	1.2	0.88493033	
45	1.3	0.903199515	
46	1.4	0.919243341	
47	1.5	0.933192799	
48	1.6	0.945200708	
49	1.7	0.955434537	
50	1.8	0.964069681	
51	1.9	0.97128344	
52	2	0.977249868	
53	2.1	0.982135579	
54	2.2	0.986096552	
55	2.3	0.98927589	
56	2.4	0.991802464	
57	2.5	0.993790335	

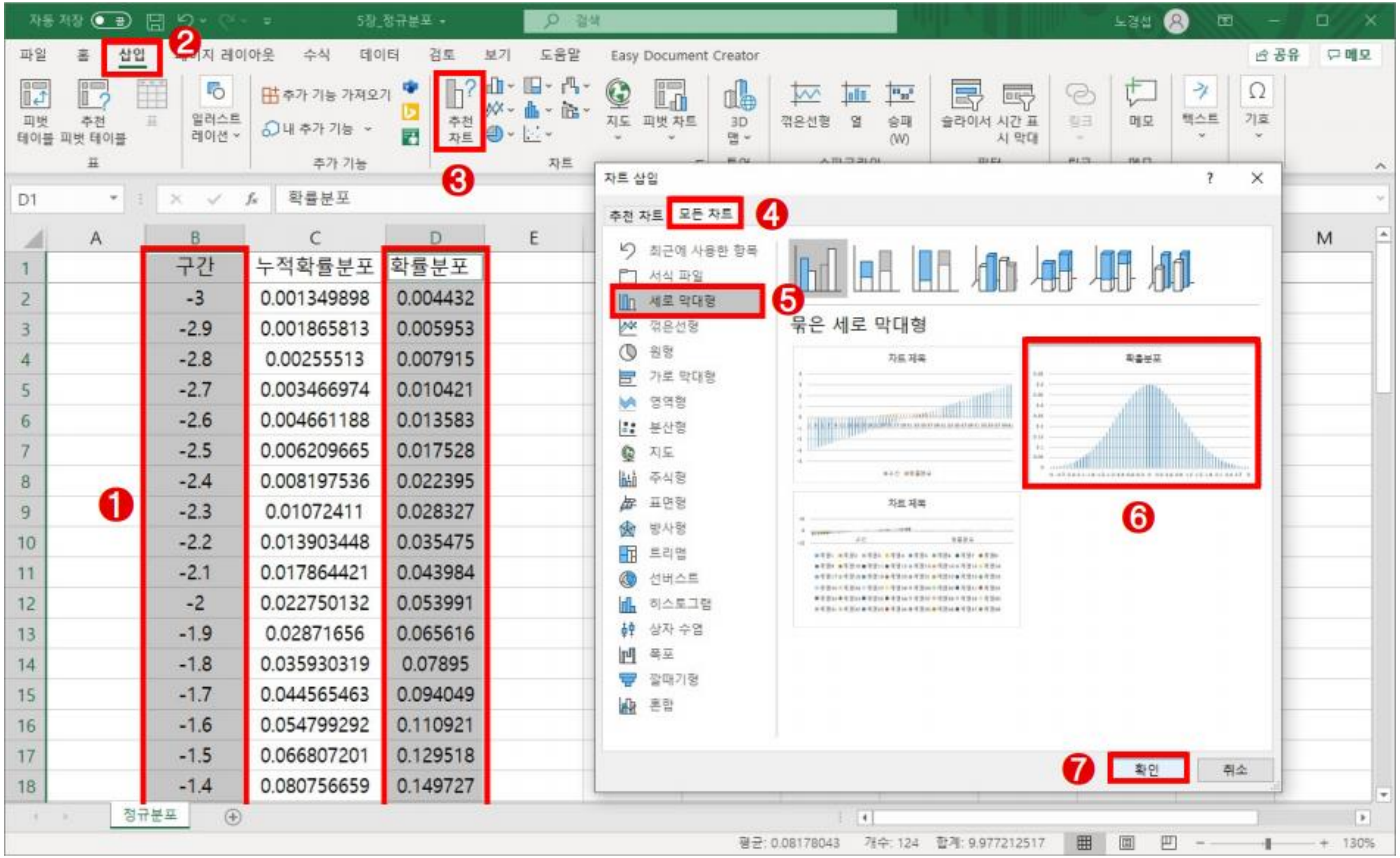
정규분포와 표준정규분포 (예제 Excel 풀이)



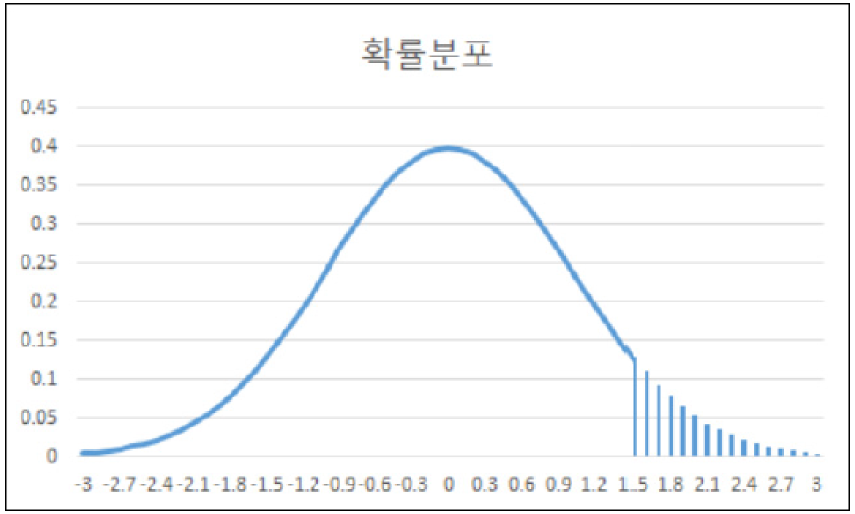
정규분포와 표준정규분포 (예제 Excel 풀이)

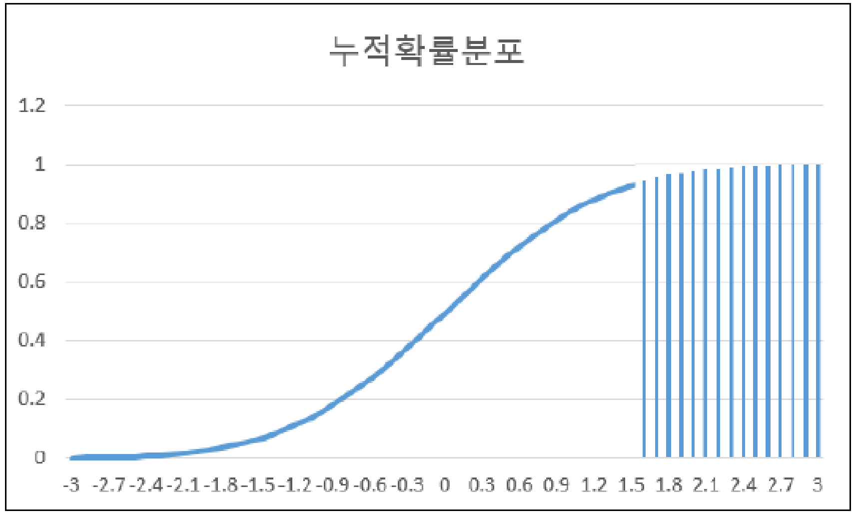


정규분포와 표준정규분포 (예제 Excel 풀이)



정규분포와 표준정규분포 (예제 Excel 풀이 완성)

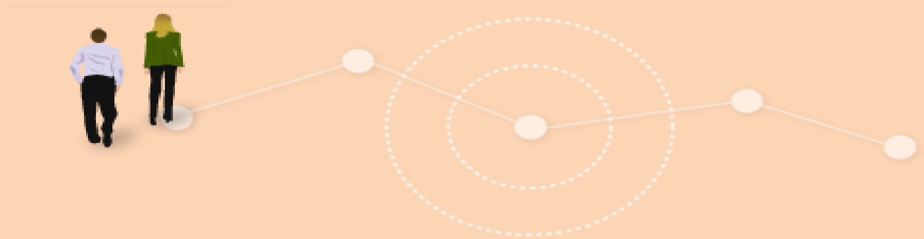




02 이항분포

:: Keywords 베르누이 분포 | 이항분포 | 이항분포의 확률 계산

정규근사를 통한 이항분포의 연속성 수정



이항분포

태어날 우리 아기는 아들일까, 딸일까?

[그림 5-19] 이항분포의 예

결혼해서 아이가 생기는 것은 부부에게 큰 축복이다. 아이가 딸이든 아들이든 부모 된 마음으로야 다 사랑스럽겠지만, 그래도 딸인지 아들인지 궁금하기 마련이다. 태아의 성별은 임신이 되는 순간 결정되지만, 의학적으로 확인이 가능하려면 12주가 지나야 한다. 원래 성별 확인은 불법이었으나 2008년에 관련 법이 개정되어 32주가 넘으면 병원에서 성별을 확인할 수 있다.

그렇다면 통계학적으로 확률분포를 이용해서 태아가 남자인지 여자인지 미리 알 수 있을까? 한 가정의 출산 수는 표본 수가 아주 적거나 1인 경우가 많기 때문에, 기계적으로 대입할 수 없지만, 전체 출산에 대한 표본을 구성한다면 통계적 적용이 가능하다. 이항분포는 아이의 성별과 같이 결과가 반드시 두 가지 중 한 가지로만 나타나며, 사건이 독립적으로 발생한다.

베르누이분포

이항분포에서 베르누이 시행과 베르누이 분포를 먼저 알아야 하는 이유는 베르누이 시행의 결과를 바탕으로 이항분포를 설명하기 때문

■ 베르누이 시행 (Bernoulli's trials)

서로 반대되는 사건이 일어나는 실험을 반복적으로 실행하는 것

→ 서로 반대되는 사건이란
반드시 두 개만 존재하며 절대 동시에 발생하지 않는 배타적인 사건

Ex. 동전 던지기와 주사위 던지기

베르누이분포

■ 베르누이 분포(bernoulli's distribution)

베르누이 시행을 확률분포로 나타낸 것

→ 성공 확률을 $p(x=1$인 경우$)$라 할 때, 실패 확률은 $1-p(x=0$인 경우$)$라고 가정

Ex. 주사위를 던져서 1과 2가 나오면 성공, 3, 4, 5, 6이 나오면 실패

→ 성공 확률(p)은 $\frac{2}{6}=\frac{1}{3}$, 실패 확률$(1-p)$은 $\frac{4}{6}=\frac{2}{3}$

베르누이 분포에서 평균과 분산은

$$\mu = E(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\begin{aligned}\sigma^2 = Var(X) &= E(X^2) - [E(X)]^2 \\ &= 0^2 \cdot p^0 (1-p)^{1-0} + 1^2 \cdot p^1 (1-p)^{1-1} - p^2 \\ &= p - p^2 = p(1-p)\end{aligned}$$

이항분포

■ 이항분포 (binomial distribution)

연속적인 베르누이 시행을 거쳐 나타나는 확률분포

→ 서로 독립된 베르누이 시행을 n회 반복할 때 성공한 횟수를 X라 하면,

성공한 X의 확률분포를 이항분포라 한다.

이항분포의 평균(μ)과 분산(σ^2)은

$$\mu = np,$$

$$\sigma^2 = np(1-p)$$

→ 성공확률$= p$, 베르누이시행을 n회한 이항분포를

$X \sim B(n, p)$ 으로 표현

이항분포

- 이항분포의 확률은 n번의 시행에서 성공확률 (p)이 r번 나타날 확률이며 n번의 시행에서 r번 관찰되는 것은 조합 ${}_nC_r$ 로 표현할 수 있음

r번 성공할 확률과 $(n-r)$번의 실패할 확률을 곱하면,

이항분포의 확률함수는

$$P(X=r)=\frac{n!}{r!\cdot(n-r)!}\cdot p^r\cdot(1-p)^{(n-r)}$$

예제 5-3 이항분포의 확률 계산하기

준비파일 | 5장_이항분포의 확률계산_함수.xlsx

대학생 A는 이제까지 많은 아르바이트를 경험했다. 다양한 아르바이트 경험으로 볼 때 10%의 확률로 A의 적성과 맞는 업무가 주어졌다. 졸업 시즌을 맞이하여 학교에서 주최하는 졸업자를 위한 취업 캠프에 약 50개의 업체가 참여한다고 하는데, A는 이 취업 캠프에 참가하려고 한다. 취업 캠프에 최종적으로 49개의 업체가 참가했을 때, A가 취업하고 싶은 회사가 2개의 업체가 될 확률을 구하라.

이항분포 (예제 풀이)

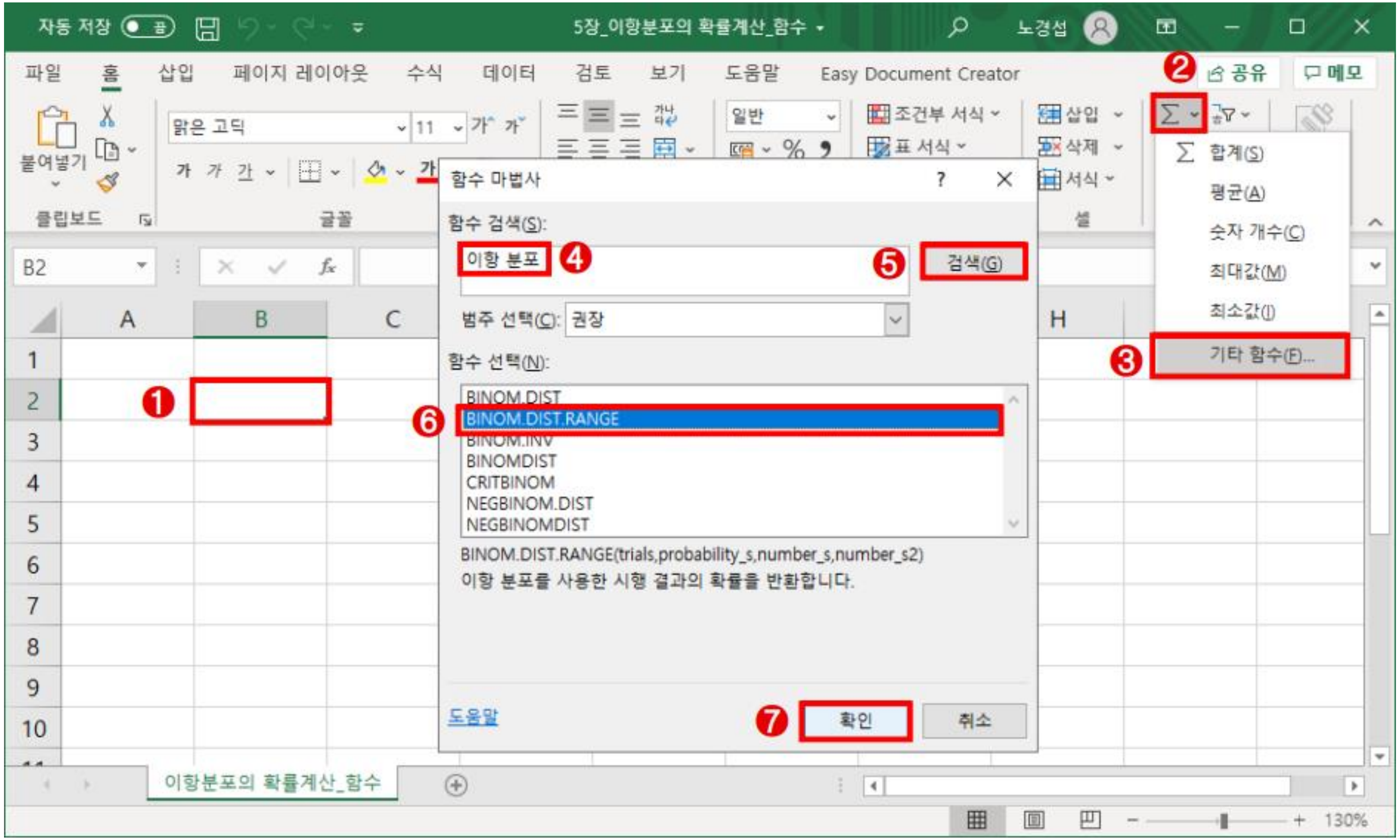
$n = 49, r = 2, p = 0.1$ 이므로 확률함수를 구하며 다음과 같다.

$$P(X=2) = \frac{49!}{2! \cdot (49-2)!} \cdot 0.1^2 \cdot (1-0.1)^{(49-2)}$$

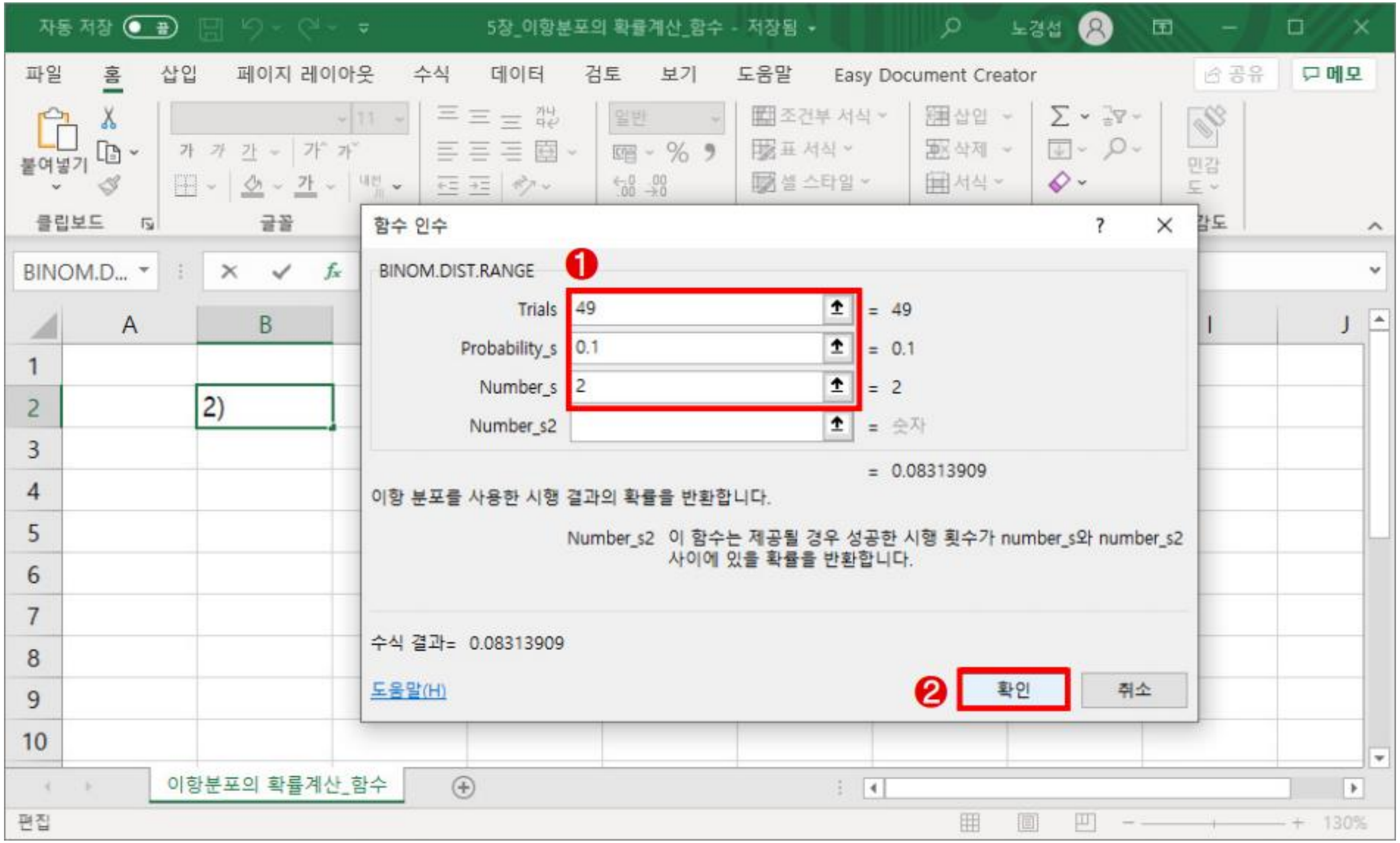
$$= 0.08313909$$

→ 8.31%의 확률로 A의 적성에 맞는 2개의 업체가 참가하게 될 것

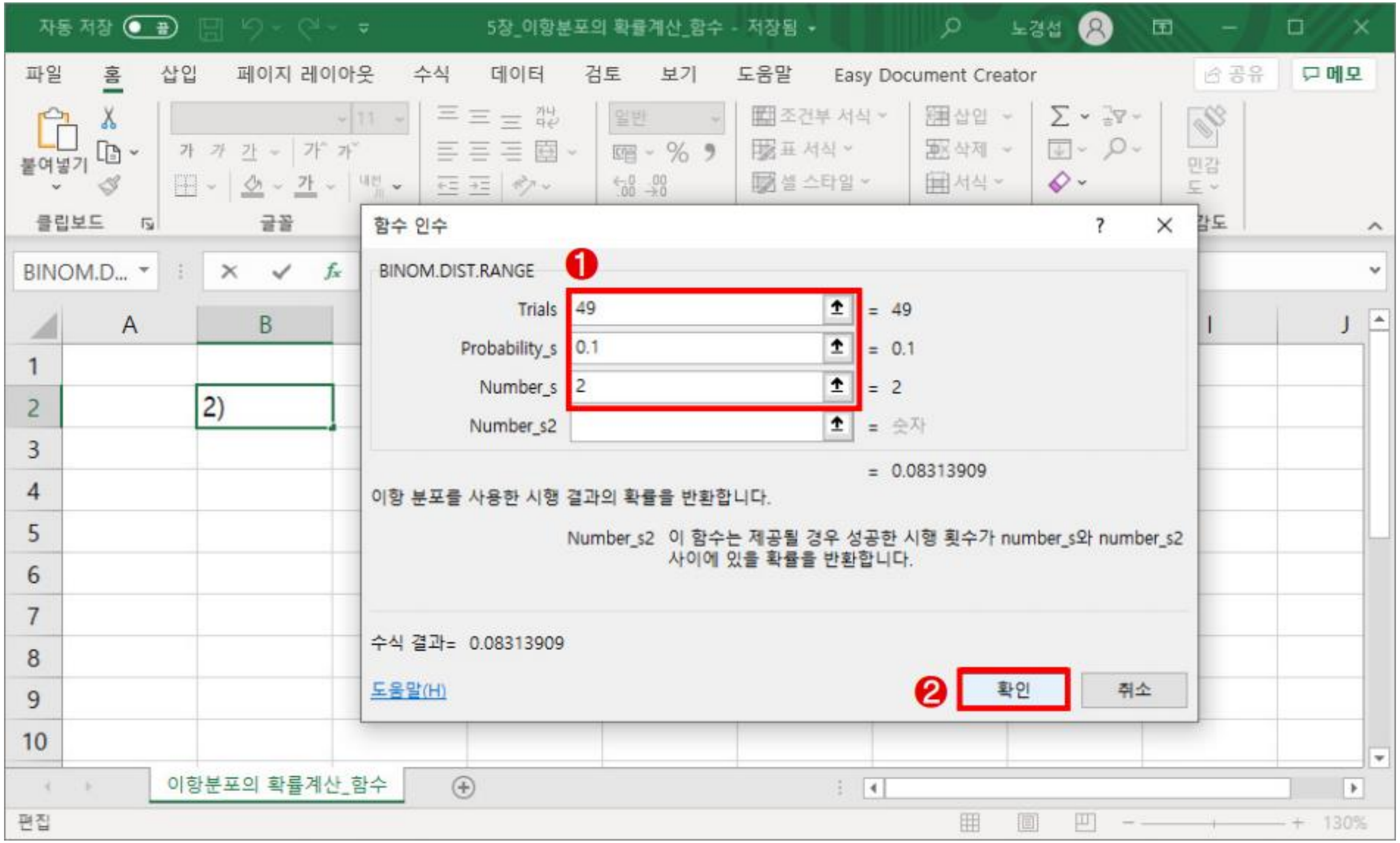
이항분포 (예제 Excel 풀이)



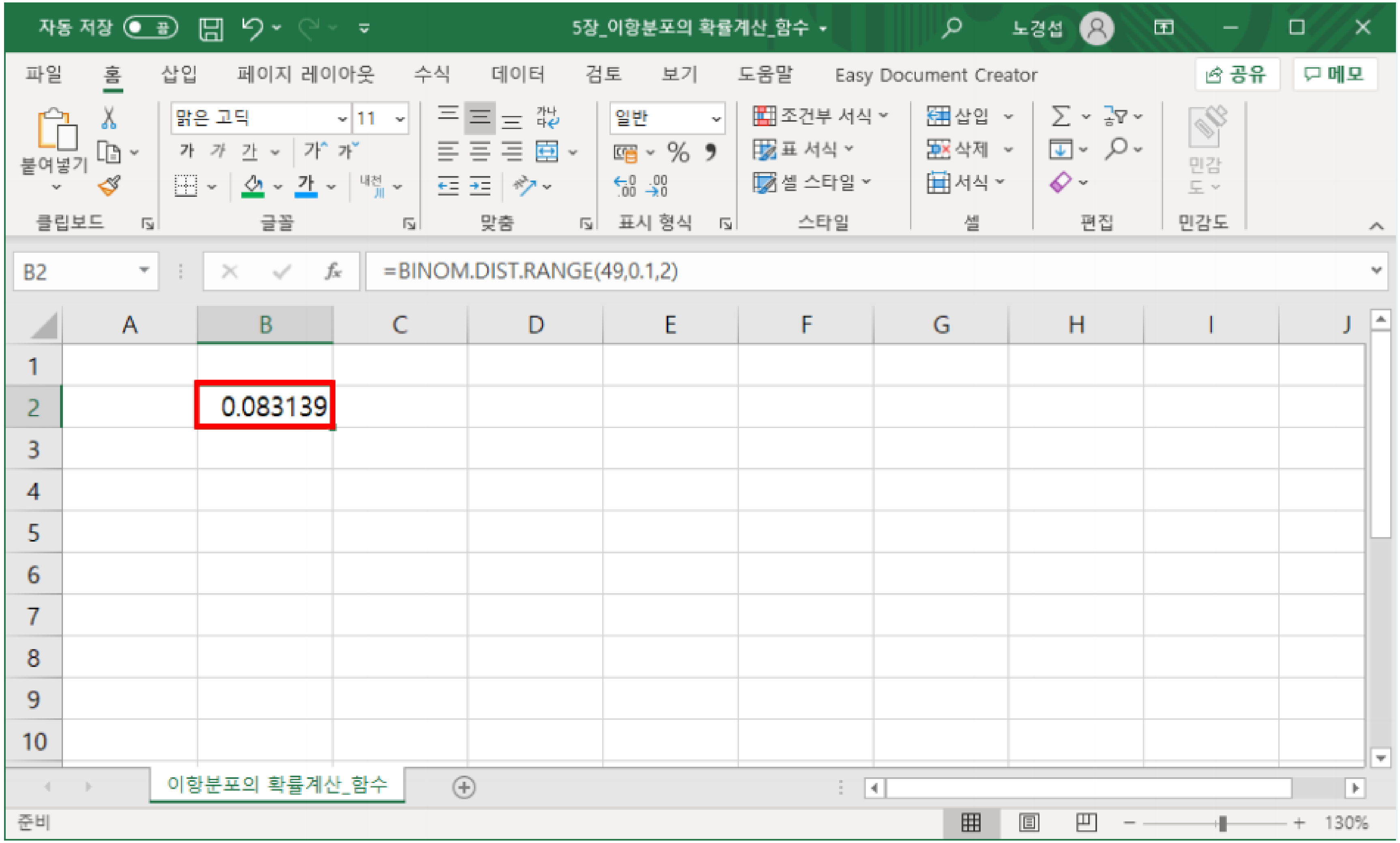
이항분포 (예제 Excel 풀이)



이항분포 (예제 Excel 풀이 완성)



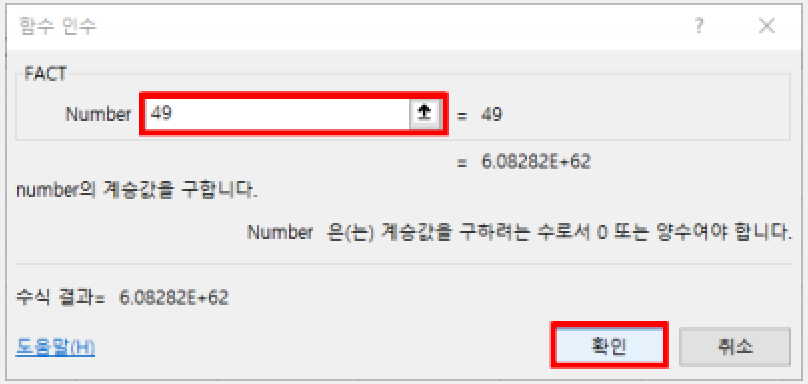
이항분포 (예제 Excel 풀이 완성-참고)



이항분포 (예제 Excel 풀이 완성-참고)

Note Excel에서 계승값(factorial) 계산하기

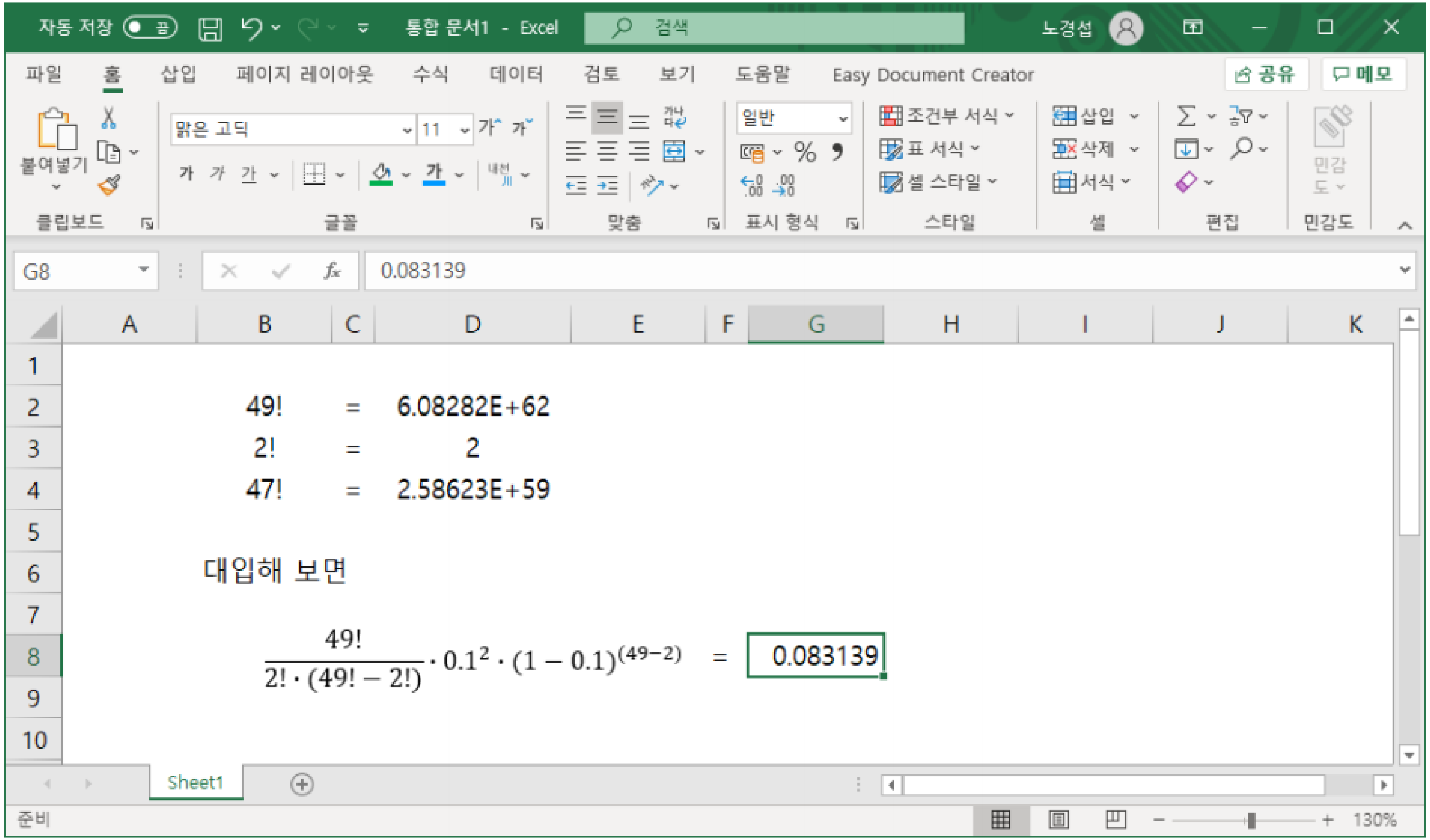
Excel에서 함수 'FACT'를 이용하면 계승값(factorial)을 구할 수 있다. 셀 포인터를 계산할 셀에 두고 [그림 5-23]과 같이 FACT 함수의 인수 창에서 'Number'에 49를 입력한다. [확인]을 클릭하면 49!을 계산할 수 있다.



[그림 5-23] FACT 함수의 인수 설정

그러면 $\dfrac{49!}{2!\cdot(49-2)!}\cdot 0.1^2\cdot(1-0.1)^{(49-2)}$을 Excel로 계산해보자. [그림 5-24]와 같이 2!, 47!을 각각 계산한 후, '=(D2/(D3*D4))*0.1^2*(1-0.1)^(49-2)'의 계산식으로 이항분포의 확률을 구할 수 있다.

이항분포 (예제 Excel 풀이 완성-참고)



정규근사를 통한 이항분포의 연속성 수정

이항분포에서도 정규분포를 이용할 수 있다

이항분포 : 이산형 분포

정규분포 : 연속적인 분포

이항분포의 확률을 알고 있더라도 충분히 많은 실험(예를 들어 1천 회)을 해서 그 확률이 맞다는 것을 확인하는 과정을 거쳐 정규근사(asymptotic normality)한다는 것을 확인하기 위함

(그러나 충분히 사건을 발생시키더라도 이항분포와 정규분포의 차이는 생긴다.)

주사위를 던져서 1과 2가 몇 번이나 나오는지 확인한다고 할 때,

1과 2는 이산된 분포지만 충분히 많은 실험을 한다 해도 확률이 정확하게 1/3이 아님.

정규근사를 통한 이항분포의 연속성 수정

그러므로 정규분포로 구성하는 과정에서 이 중간을 구성하는 값들이 존재하게 되고,
이런 중간의 값들을 좀 더 정확하게 확인하기 위하여 0.5로 보정

이때 구간이라면
하한에서는 -0.5
상한에서는 0.5를 추가하여 분포를 확인

→ 이것을 연속성 수정(continuity correction)이라 한다.

정규근사를 통한 이항분포의 연속성 수정

예제 5-4 이항분포의 연속성 수정하기

주사위 1개를 20번 던지는 실험을 통해 관찰된 확률변수 X가 있을 때, 정규근사를 이용하여 $P(X \geq 5)$의 근사값을 구하는 근사식을 연속성 수정을 통해 구하라.

정규근사를 통한 이항분포의 연속성 수정 (예제 풀이)

[풀이]

이항분포를 정규분포로 근사시키기 위해

$X \sim B(n, p)$를 $X \sim N(np.\,npq)$, (단, $q = 1 - p$, npq : 분산)로 구성

$P(X \geq 5)$를 정규근사 시킨 후

$P(Z \geq ?)$를 계산하는 것이므로, $Z = \dfrac{X - np}{\sqrt{np(1-p)}}$를 확인

정규근사를 통한 이항분포의 연속성 수정 (예제 풀이)

❶ 정규근사

$$n=20,\ p=\frac{1}{6},\ np=20\times\frac{1}{6}=\frac{10}{3},\ np(1-p)=\frac{10}{3}\times\frac{5}{6}=\frac{25}{9}$$

$$P(X\geq 5)=P\left(Z\geq\frac{5-\frac{10}{3}}{\sqrt{\frac{25}{9}}}\right)\qquad\qquad \therefore P(Z\geq 1)$$

정규근사를 통한 이항분포의 연속성 수정 (예제 풀이)

❷ 연속성 수정

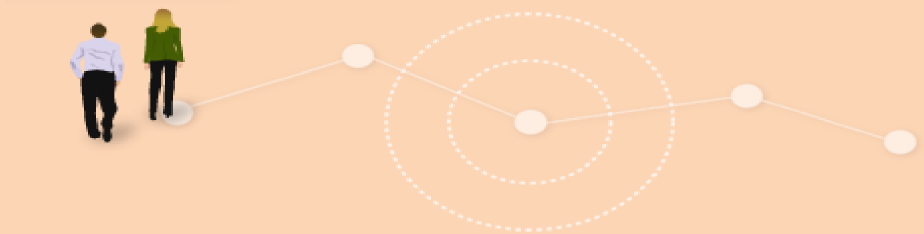
$P(X \geq 5)$ 구간의 하한이 5이므로, 하한을 −0.5로 수정하여 계산하면 다음과 같다.

$$P(X \geq 5) = P\left(Z \geq \frac{4.5 - \frac{10}{3}}{\sqrt{\frac{25}{9}}}\right) \qquad \therefore P(Z \geq 0.7)$$

그러므로 $P(X \geq 5) = P(Z \geq 0.7)$

03 포아송분포

:: Keywords 포아송분포 | 포아송분포의 확률 계산
포아송분포와 정규분포와의 관계



포아송분포

시메옹 드니 포아송(Siméon-Denis Poisson)

시메옹 드니 포아송(Siméon-Denis Poisson, 1781~1840)은 프랑스의 수학자이자 물리학자다. 그는 프랑스의 기마대에서 군의관으로 복무했다. 기마대에서 복무하는 병사들이 의무대를 찾는 경우는 일반 병원이나 타 의무대 상황과 별반 다르지 않았다. 전쟁 중이라면 낙마 관련 사고가 자주 발생하겠지만, 당시는 평화로운 시기여서 낙마 사고가 별로 없었다. 잊을 만하면 낙마 병사가 생겨 의무대로 실려 오곤 했다.

포아송은 여기서 착안하여 낙마를 하거나, 말과 관련한 사고로 다치거나 죽는 병사가 발생할 확률을 계산해냈다.

[그림 5-25] 기마대

포아송분포

■ 포아송분포 (poisson distribution)

특정한 사건이 발생할 가능성이 매우 드문 경우의 확률분포

ex. 기마대의 기병의 낙마사고 발생

손을 씻다 세면기에 휴대전화를 빠트릴 횟수

야구 관람 중에 홈런볼이나 파울볼을 받을 횟수

버스를 탈 때 1초도 기다리지 않고 정류장에 동시에 도착할 횟수 등

포아송분포

■ 말을 타는 횟수(n) 중 말에서 떨어지는 사고가 발생하라 횟수(x)

$$P(X=x) = {}_{n}C_{x} \cdot p^{x} \cdot (1-p)^{(n-x)}$$

포아송분포의 확률함수를 도출하기 위해서

포아송분포의 평균(μ)과 분산(σ^2)이 모두 λ로 같다$(\mu = \sigma^2 = \lambda)$는 가정이 필요

n개의 구간에서 발생할 확률은 $\dfrac{\lambda}{n}$

${}_{n}C_{x} \cdot p^{x} \cdot (1-p)^{(n-x)}$ 에 대입하면 $P(X=x) = \dfrac{\lambda^{x} e^{-\lambda}}{x!}$

포아송분포의 확률 계산

단위 시간당 평균 사건 발생 건수를 $\lambda(\mathrm{lambda})$

$X \sim Poisson(\lambda)$ 로 표현

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

포아송분포

참고 **포아송분포의 확률함수 유도 과정**

각 n개의 구간에서 발생할 확률 $\dfrac{\lambda}{n}$를 ${}_nC_x \cdot p^x \cdot (1-p)^{(n-x)}$에 대입하면 다음과 같다.

$$\begin{aligned}
{}_nC_x \cdot p^x \cdot (1-p)^{(n-x)} &= \frac{n!}{x! \cdot (n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{(n-x)} \\
&= \frac{n!}{x! \cdot (n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n} \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \lim_{n\to\infty} \frac{1}{x!}\,\frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x}\,\lambda^x \left(1-\frac{\lambda}{n}\right)^{n} \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \lim_{n\to\infty}\left[\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n} 1\right]
\end{aligned}$$

$\lim\limits_{n\to\infty}\left(1-\dfrac{\lambda}{n}\right)^n = e^{-\lambda}$이므로 포아송분포의 확률함수는 $P(X=x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$이 된다.

포아송분포의 확률 계산

예제 5-5 포아송분포의 확률 계산하기

준비파일 | 5장_포아송분포의 확률 계산.xlsx

A는 건망증이 심해서 출근하거나 퇴근할 때 충전하던 휴대폰을 집이나 사무실에 종종 놓고 나간다. 일주일 단위로 살펴봤을 때 휴대폰을 놓고 나간 평균 횟수가 3회였다. 일주일에 단 1회 이하로 휴대폰을 놓고 나갈 확률과 4~5회 휴대폰을 놓고 나갈 확률을 구하라.

포아송분포의 확률 계산 (예제 풀이)

■ 1회 이하로 휴대전화를 놓고 갈 확률 $P(X \leq 1)$

1주일동안 휴대전화를 놓고 갈 평균 횟수=3

$\lambda = 3$

$x = 1$을 대입하면

$$P(X=1) = \frac{3^1 \cdot e^{-3}}{1!} = \frac{3^1 \cdot 2.7182818^{-3}}{1!} = 0.149361$$

$P(X \leq 1)$은 1번 이하로 휴대전화를 놓고 가는 경우이므로,

한 번도 실수를 하지 않는 경우

$$P(X=0) = \frac{3^0 \cdot e^{-3}}{0!} = \frac{3^0 \cdot 2.7182818^{-3}}{0!} = 0.049787$$

$\lambda = 3,\ x = 1$ 의 확률과 $\lambda = 3,\ x = 0$ 의 확률을 더하면

$$P(X \leq 1) = 0.199148$$

∴ 1회 이하로 휴대전화를 놓고 갈 확률은 19.91%

포아송분포의 확률 계산 (예제 풀이)

■ 4~5회 휴대전화를 놓고 갈 확률 $P(4 \le X \le 5)$

휴대전화를 4회 놓고 가는 경우의 확률함수

$$P(X=4)=\frac{3^4\cdot e^{-3}}{4!}=\frac{3^4\cdot 2.7182818^{-3}}{4!}=0.168031$$

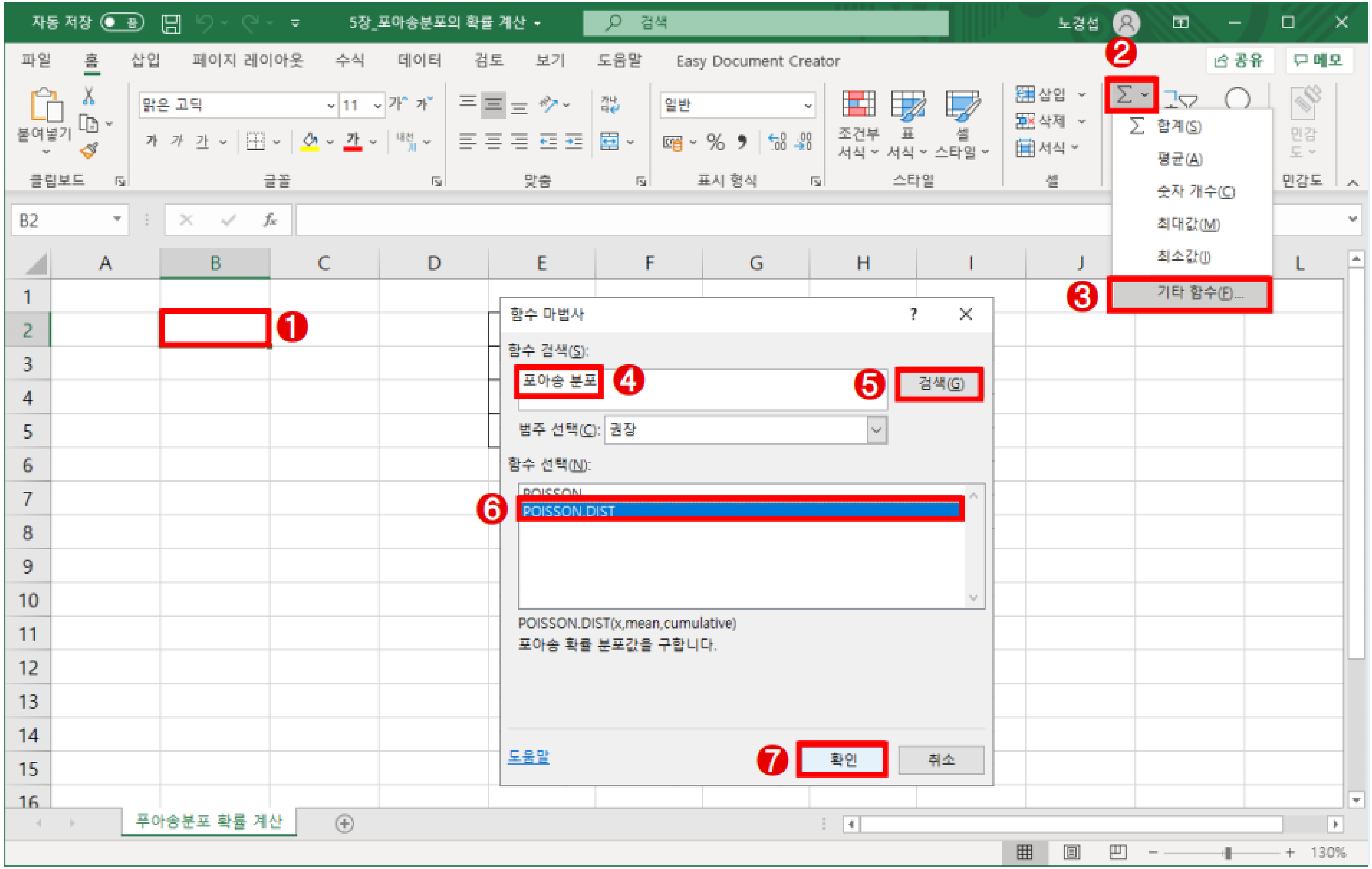
휴대전화를 5회 놓고 가는 경우의 확률함수

$$P(X=5)=\frac{3^5\cdot e^{-3}}{5!}=\frac{3^5\cdot 2.7182818^{-3}}{5!}=0.100819$$

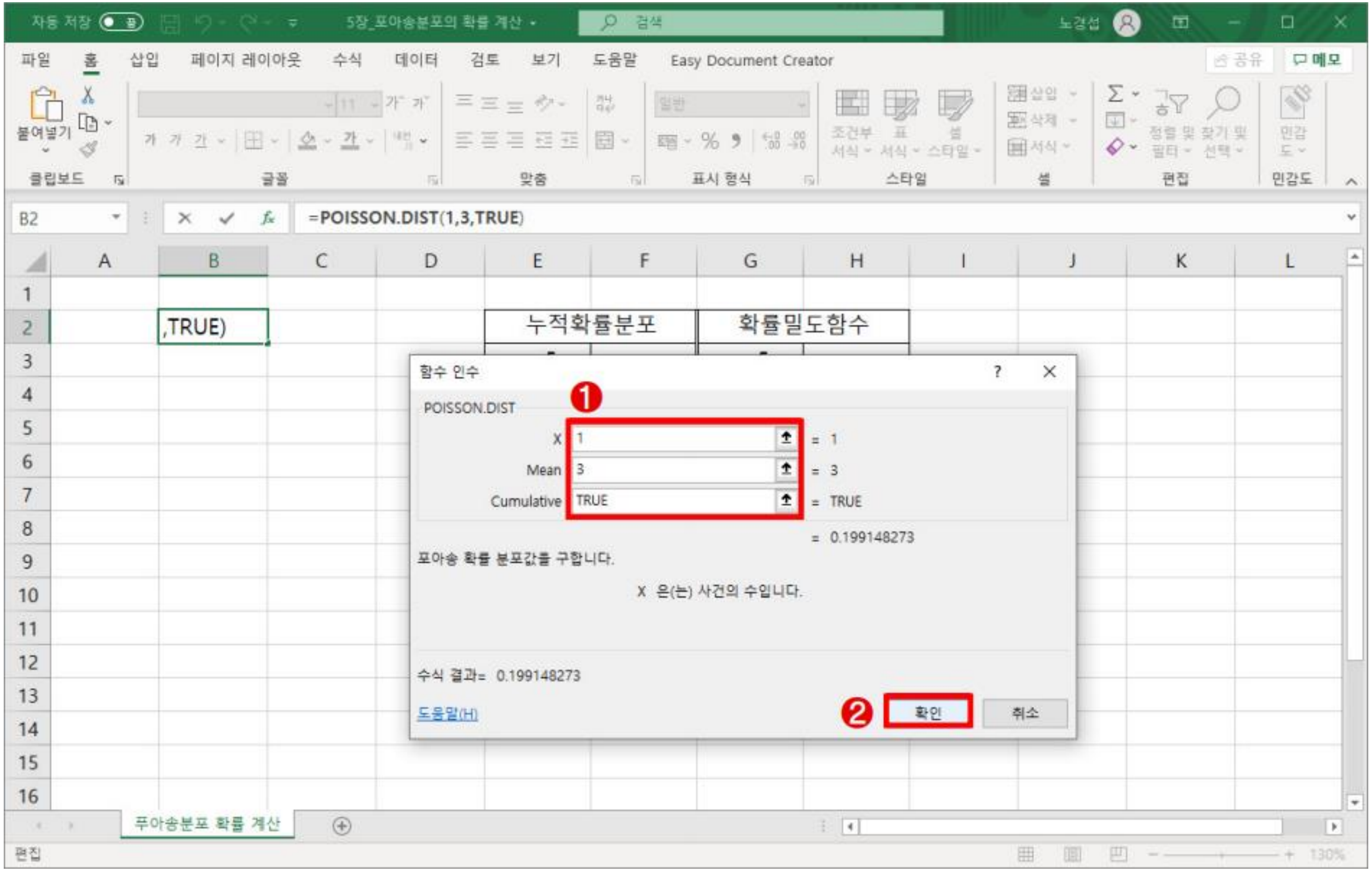
두 값을 더하면 0.26885

∴ 4~5회 휴대전화를 놓고 갈 확률은 26.89%

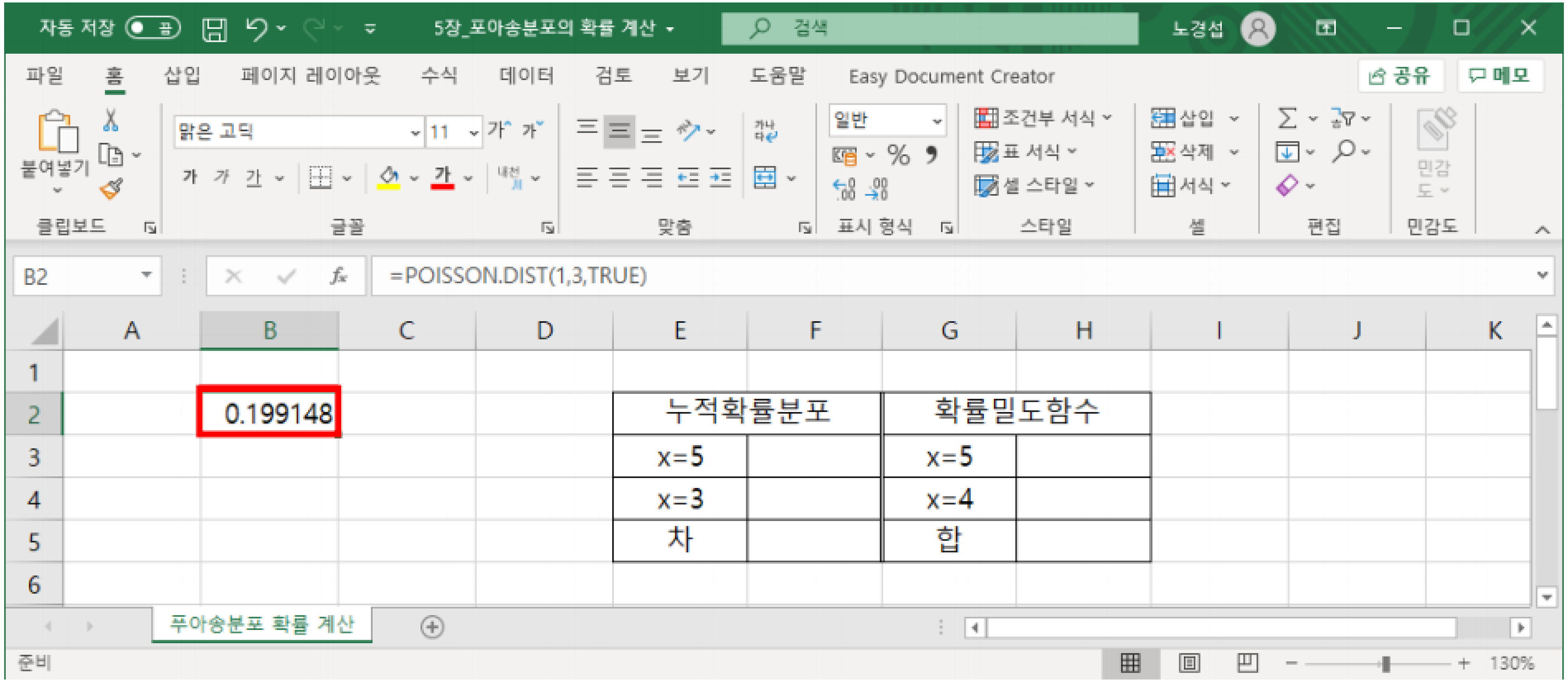
포아송분포의 확률 계산 (예제 Excel 풀이)



포아송분포의 확률 계산 (예제 Excel 풀이)

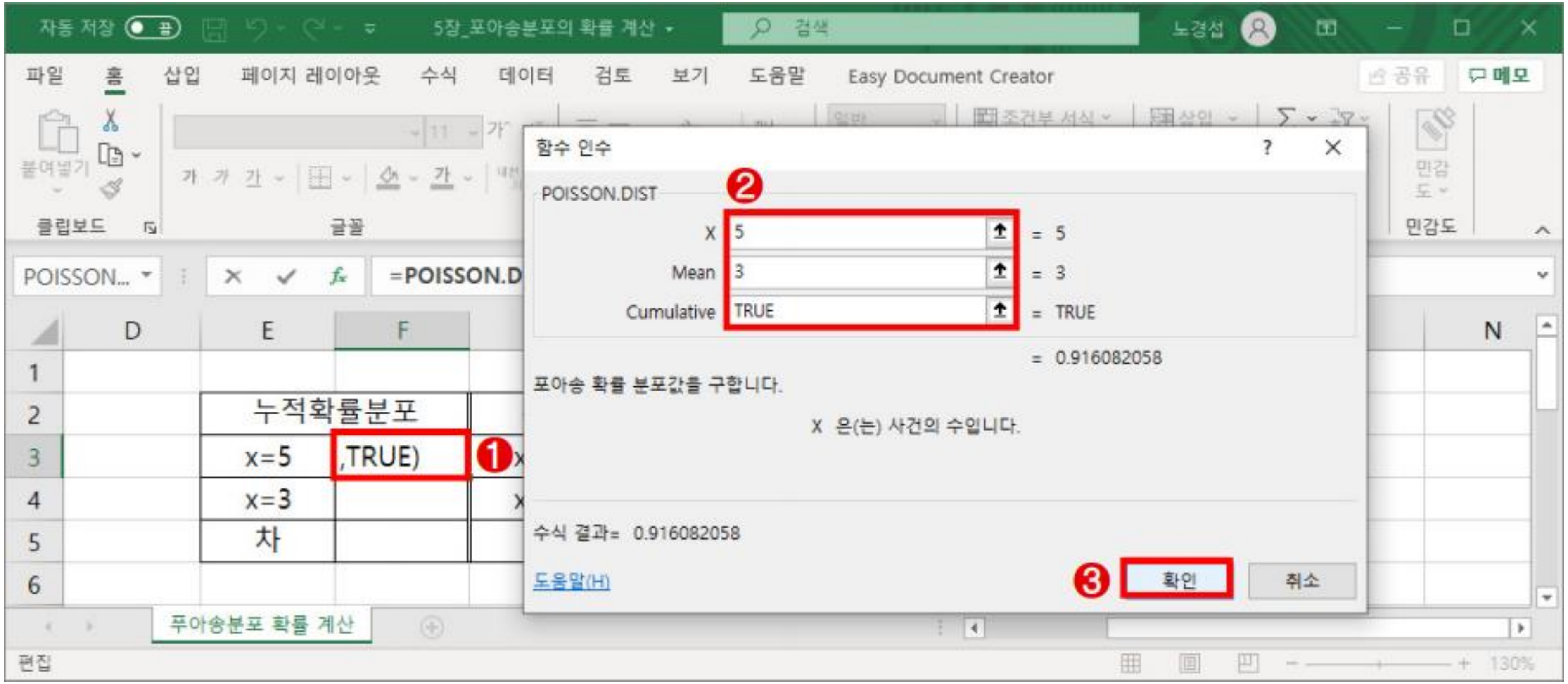


포아송분포의 확률 계산 (예제 Excel 풀이)

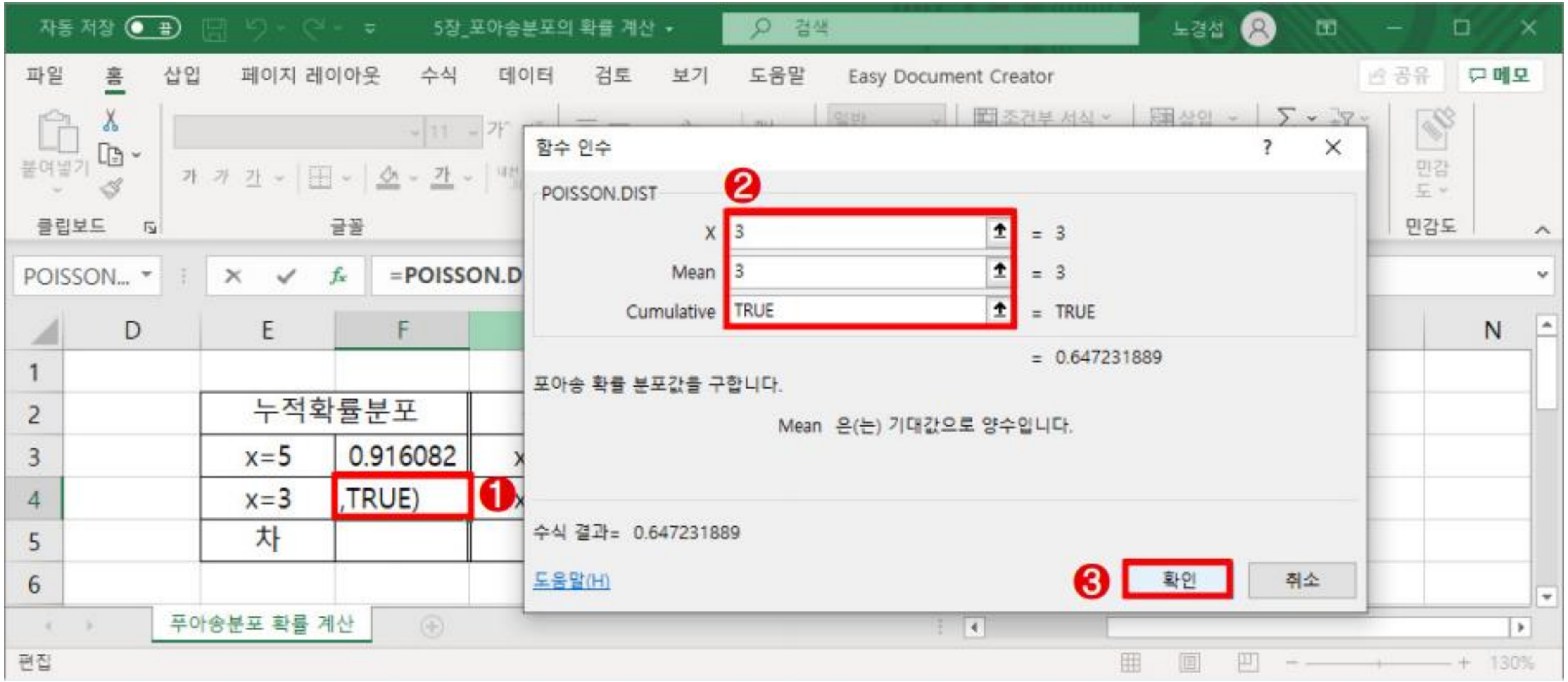


포아송분포의 확률 계산 (예제 Excel 풀이)

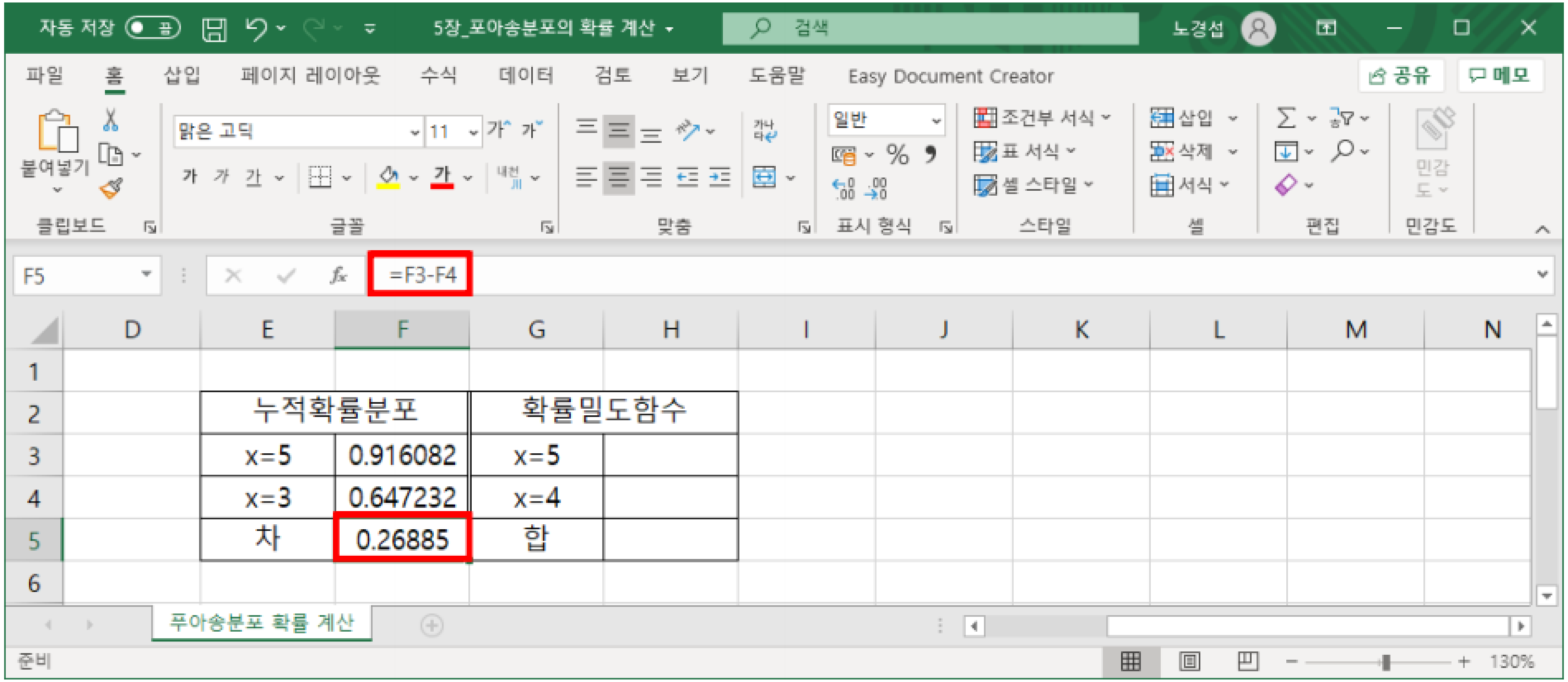
■ 4~5회 휴대전화를 놓고 갈 확률



포아송분포의 확률 계산 (예제 Excel 풀이)

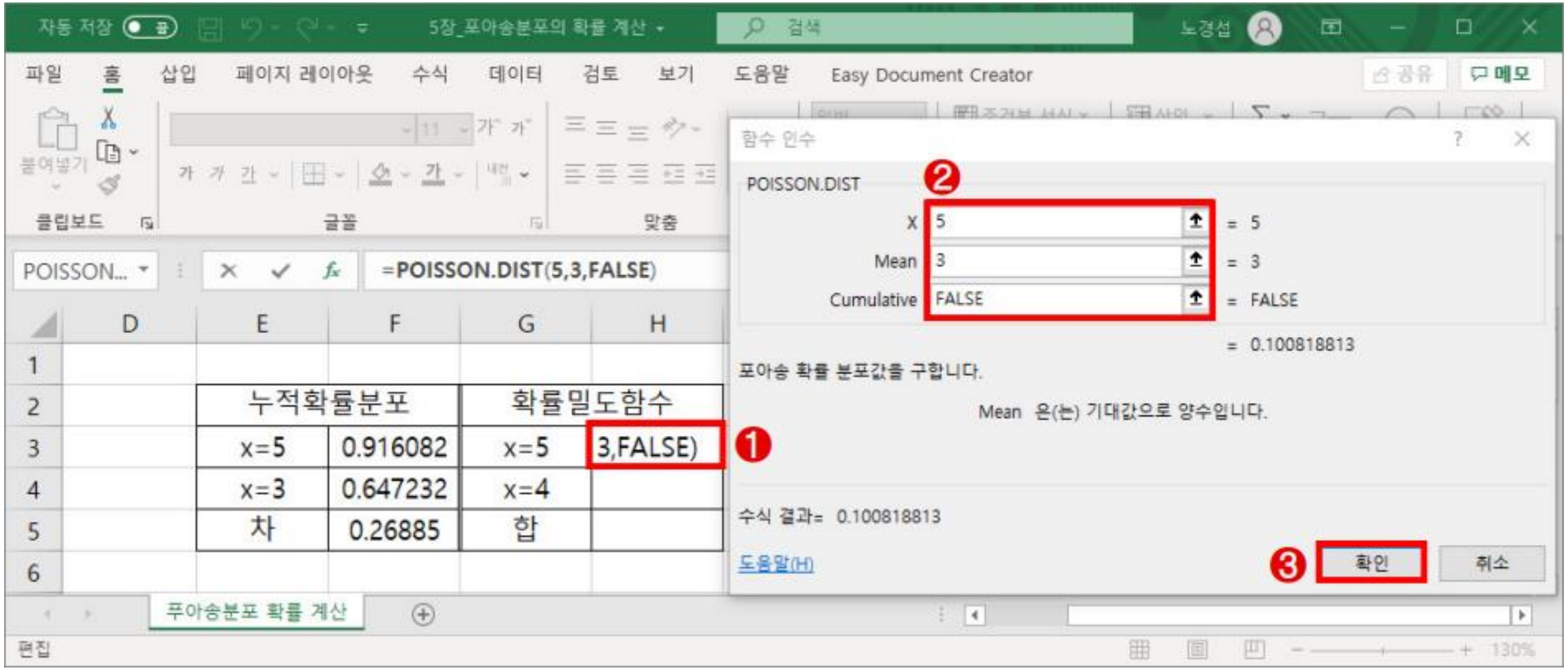


포아송분포의 확률 계산 (예제 Excel 풀이)

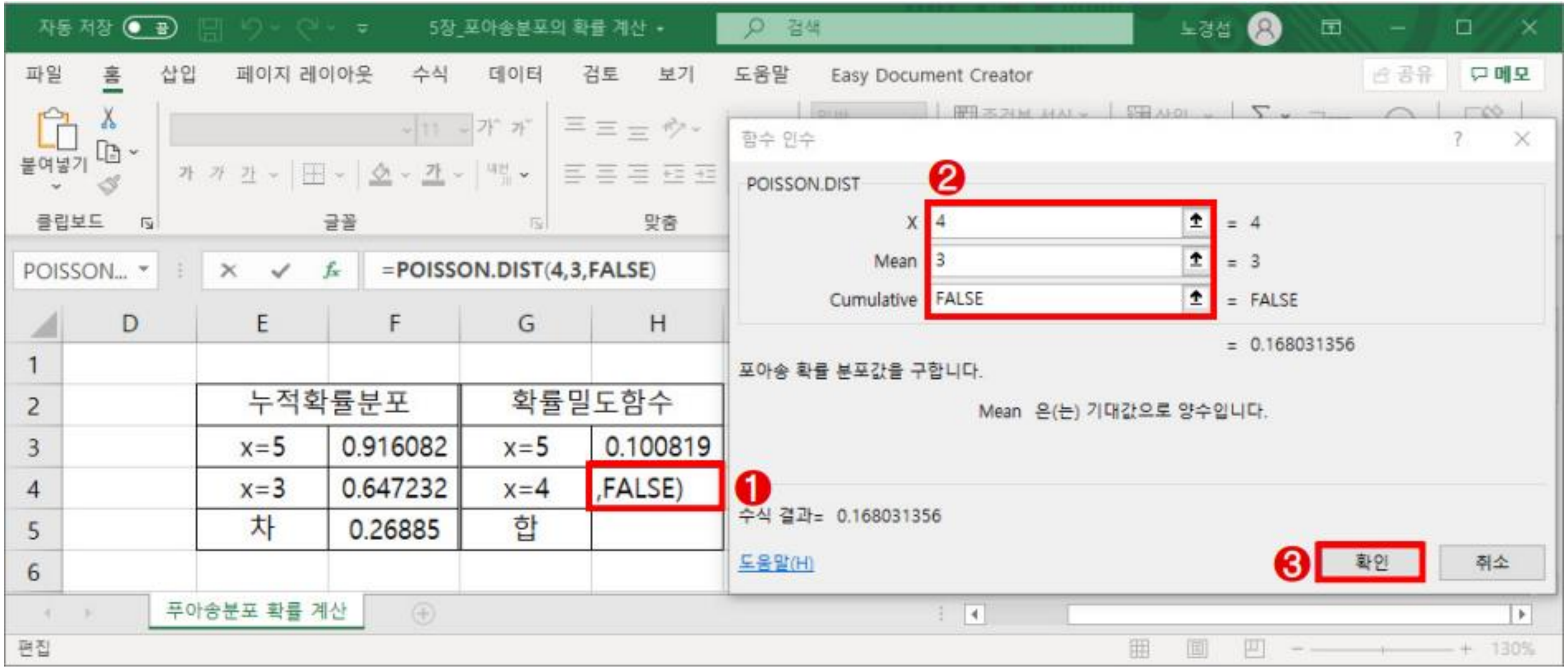


포아송분포의 확률 계산 (예제 Excel 풀이)

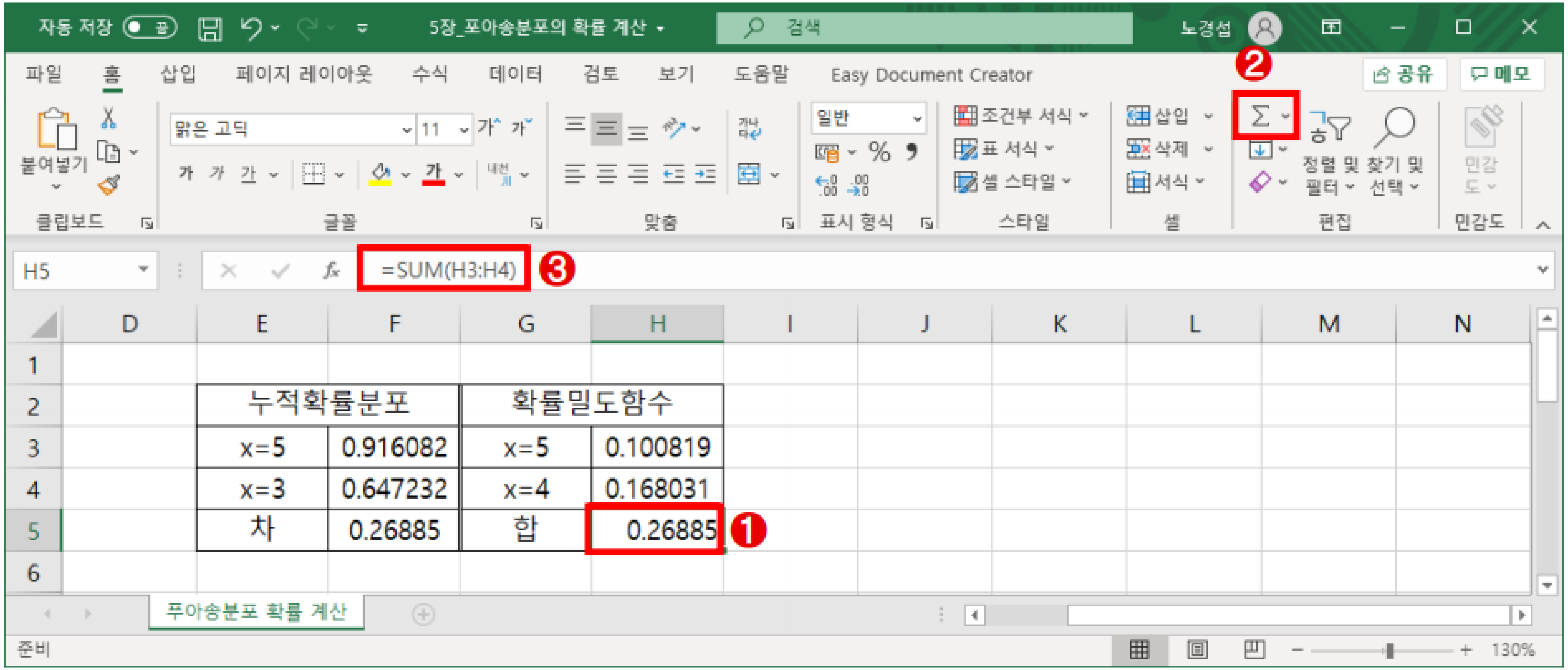
■ 확률밀도함수를 이용하는 경우



포아송분포의 확률 계산 (예제 Excel 풀이)

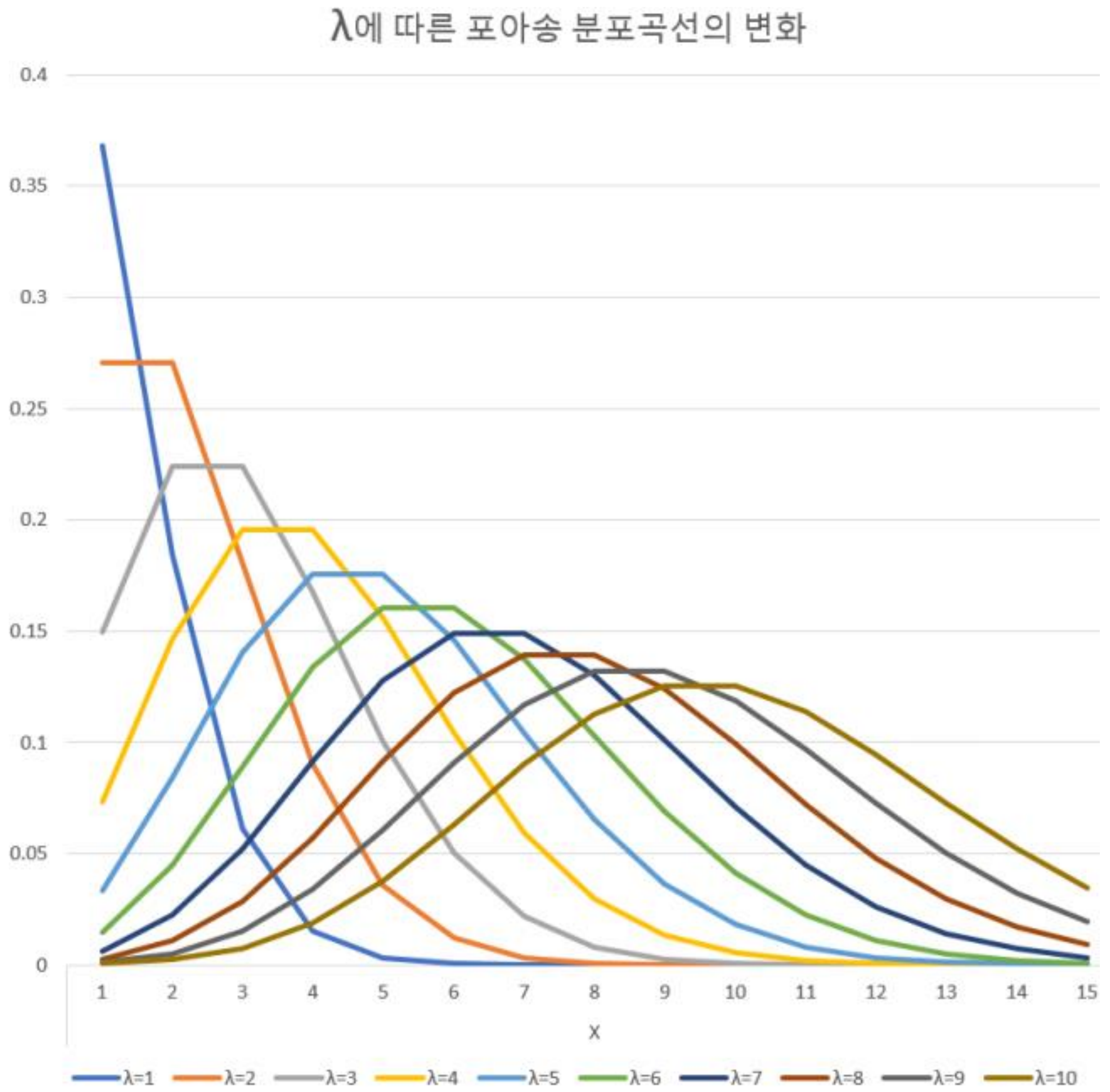


포아송분포의 확률 계산 (예제 Excel 풀이 완성)



포아송분포와 정규분포의 관계

구분		λ=1	λ=2	λ=3	λ=4	λ=5	λ=6	λ=7	λ=8	λ=9	λ=10
X	1	0.36788	0.27067	0.14936	0.07326	0.03369	0.01487	0.00638	0.00268	0.00111	0.00045
	2	0.18394	0.27067	0.22404	0.14653	0.08422	0.04462	0.02234	0.01073	0.005	0.00227
	3	0.06131	0.18045	0.22404	0.19537	0.14037	0.08924	0.05213	0.02863	0.01499	0.00757
	4	0.01533	0.09022	0.16803	0.19537	0.17547	0.13385	0.09123	0.05725	0.03374	0.01892
	5	0.00307	0.03609	0.10082	0.15629	0.17547	0.16062	0.12772	0.0916	0.06073	0.03783
	6	0.00051	0.01203	0.05041	0.1042	0.14622	0.16062	0.149	0.12214	0.09109	0.06306
	7	7.3E-05	0.00344	0.0216	0.05954	0.10444	0.13768	0.149	0.13959	0.11712	0.09008
	8	9.1E-06	0.00086	0.0081	0.02977	0.06528	0.10326	0.13038	0.13959	0.13176	0.1126
	9	1E-06	0.00019	0.0027	0.01323	0.03627	0.06884	0.1014	0.12408	0.13176	0.12511
	10	1E-07	3.8E-05	0.00081	0.00529	0.01813	0.0413	0.07098	0.09926	0.11858	0.12511
	11	9.2E-09	6.9E-06	0.00022	0.00192	0.00824	0.02253	0.04517	0.07219	0.09702	0.11374
	12	7.7E-10	1.2E-06	5.5E-05	0.00064	0.00343	0.01126	0.02635	0.04813	0.07277	0.09478
	13	5.9E-11	1.8E-07	1.3E-05	0.0002	0.00132	0.0052	0.01419	0.02962	0.05038	0.07291
	14	4.2E-12	2.5E-08	2.7E-06	5.6E-05	0.00047	0.00223	0.00709	0.01692	0.03238	0.05208
	15	2.8E-13	3.4E-09	5.5E-07	1.5E-05	0.00016	0.00089	0.00331	0.00903	0.01943	0.03472



Q&A

통계학, 제대로 시작하자!

통계의 쓰임을 이해하고,
실제로 활용할 수 있어야 한다.