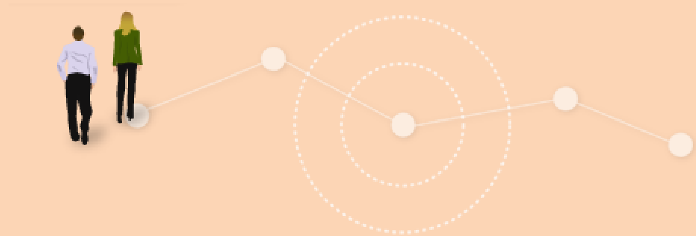
5 장

목 차

01 확률분포

:: Keywords 확률분포 | 균등분포 | 정규분포 | 표준정규분포



확률분포

■ 확률분포의 정의

확률분포(probability distribution)는 래에 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.

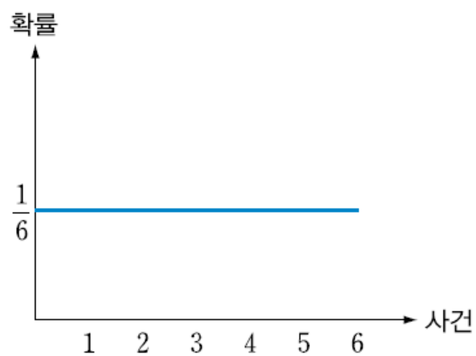
→ 4장에서 살펴본 월드컵과 치킨집의 사례에서 치킨집 점주가 위험 기피자(risk-averter)라면 135마리보다 더 적게 준비하여 낭비를 피할 것이며, 반대로 위험 선호자(risk-lover)라면 135마리보다 좀 더 준비하여 적극적으로 판촉 활동을 벌일 수도 있다.

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

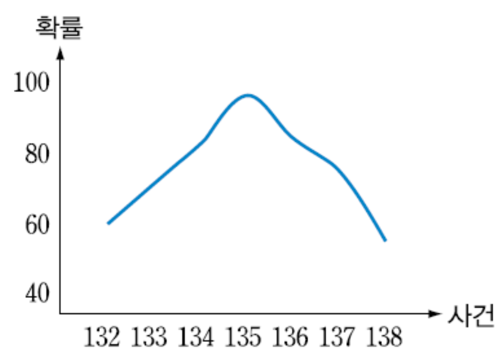
사건	1	2	3	4	5	6
확률	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(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

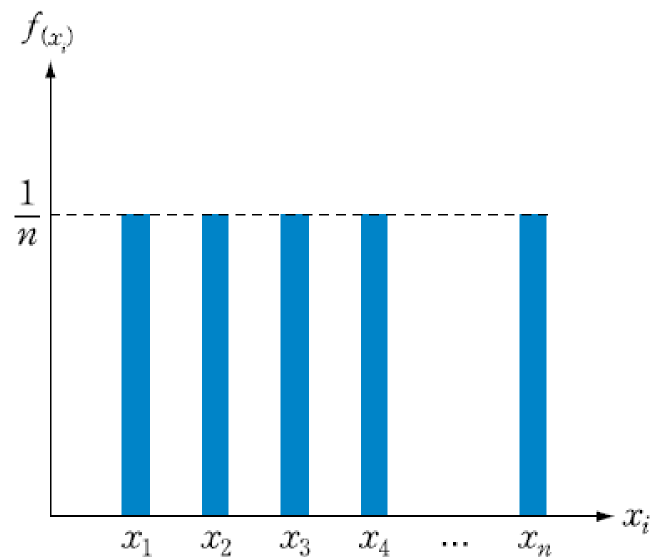
균등분포

■ 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

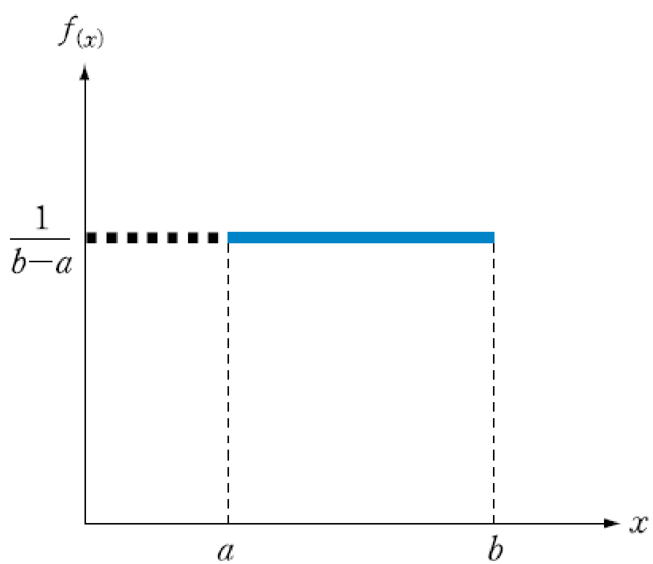
■ 연속균등분포

특정 범위 내에서 모든 확률함수가 동일한 분포

확률변수가 X가 a와 b구간에서 균등분포를 가진다면

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

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



정규분포와 표준정규분포

■ 정규분포 (normal distribution)

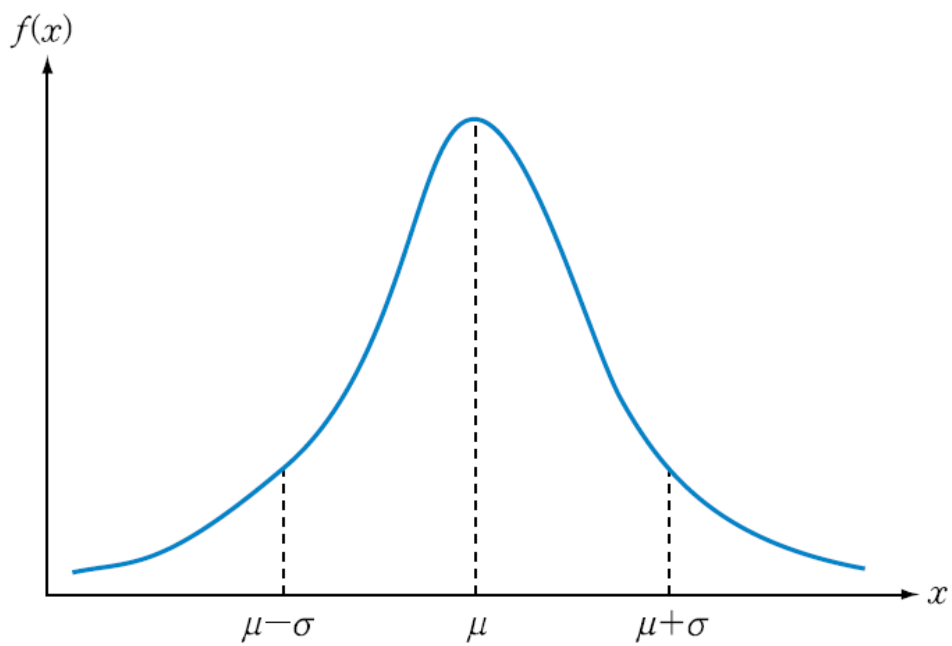
축적된 데이터를 기준으로 미래를 예측할 수 있는 분포

→ 통계학에서 가장 많이 사용되는 분포이며,

평균과 분산만으로 그 특성을 모두 설명할 수 있어 아주 편리

평균이 μ, 분산이 σ^2 인 정규분포의 확률함수는 $f(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$X \sim N(\mu,\ \sigma^2)$으로 나타낸다.



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

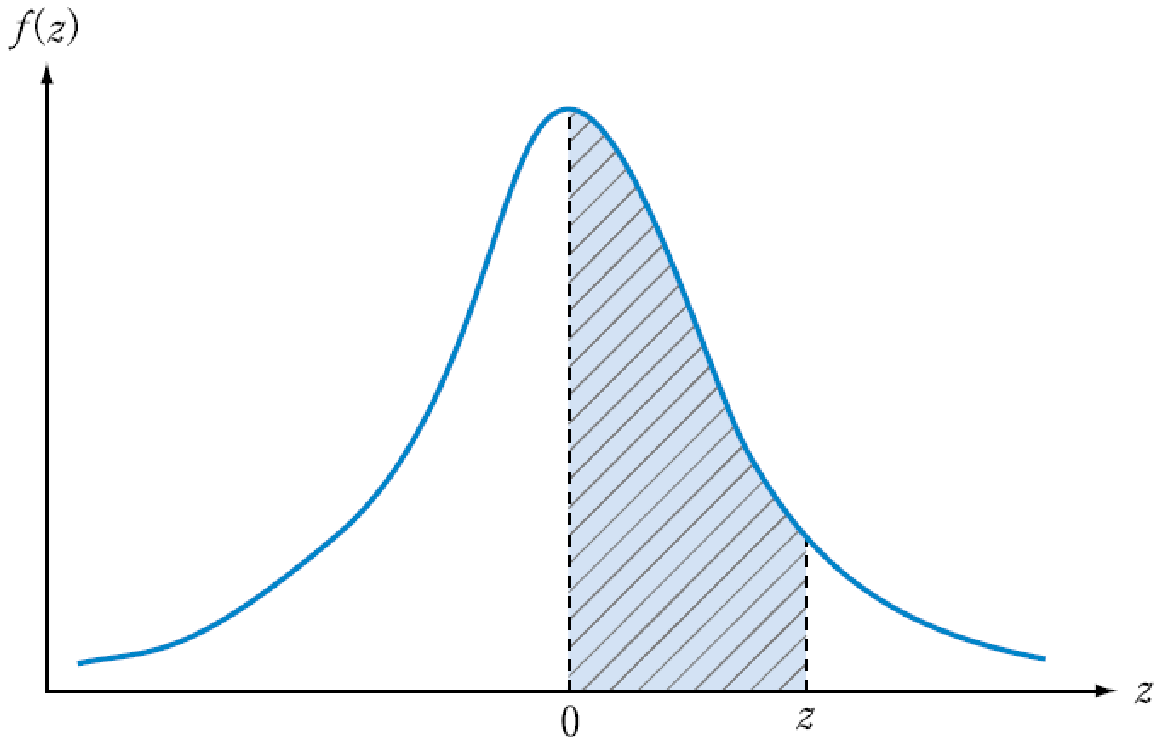
확률변수 X가 정규분포 N(μ, σ^2)을 따를 때,

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

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

정규분포와 표준정규분포

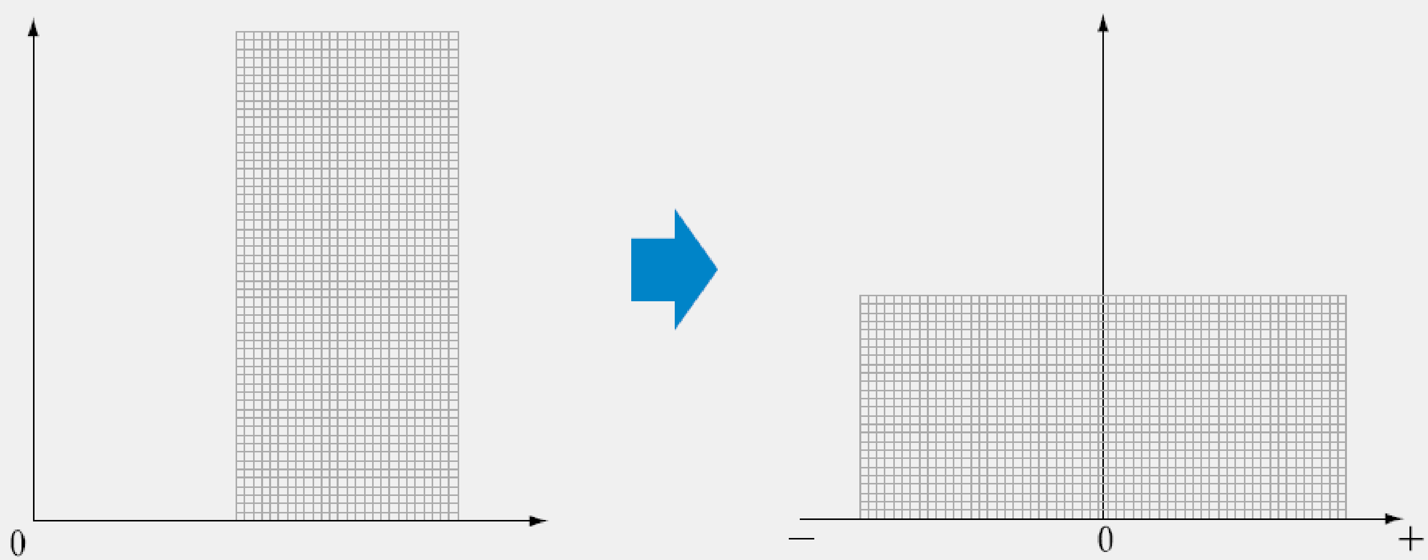
확률변수 Z의 값이 구간 $(0,\ z)$에 속할 확률 $P(0 \leq Z \leq z)$는



정규분포와 표준정규분포

Note **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.

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

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

정규분포와 표준정규분포

예제 5-2 정규분포의 확률 계산

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

$$z = \frac{x-\mu}{\sigma} = \frac{185-170}{10} = 1.5$$

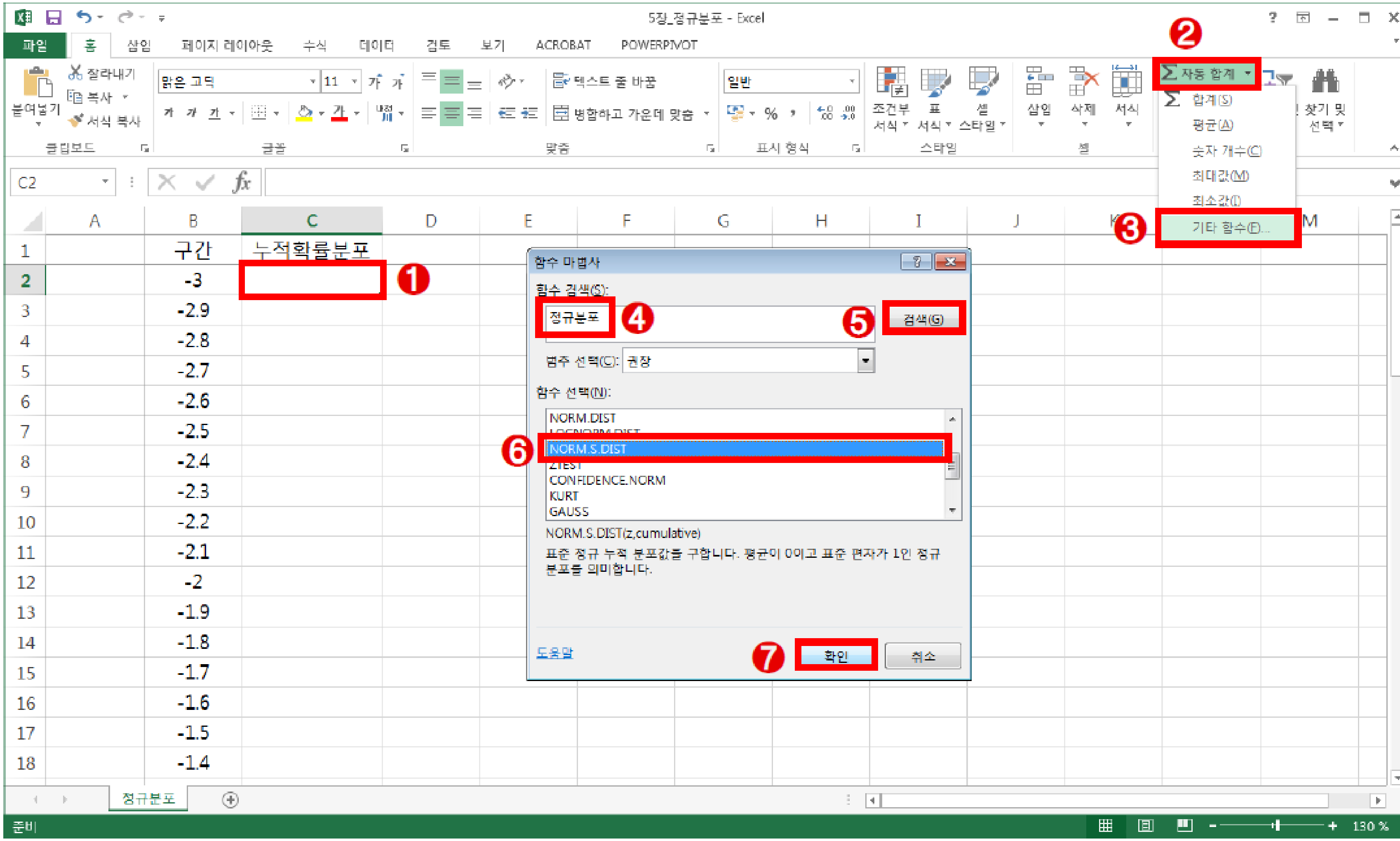
$P(z > 1.5)$

z가 1.5보다 큰 부분의 확률을 구해야하므로,

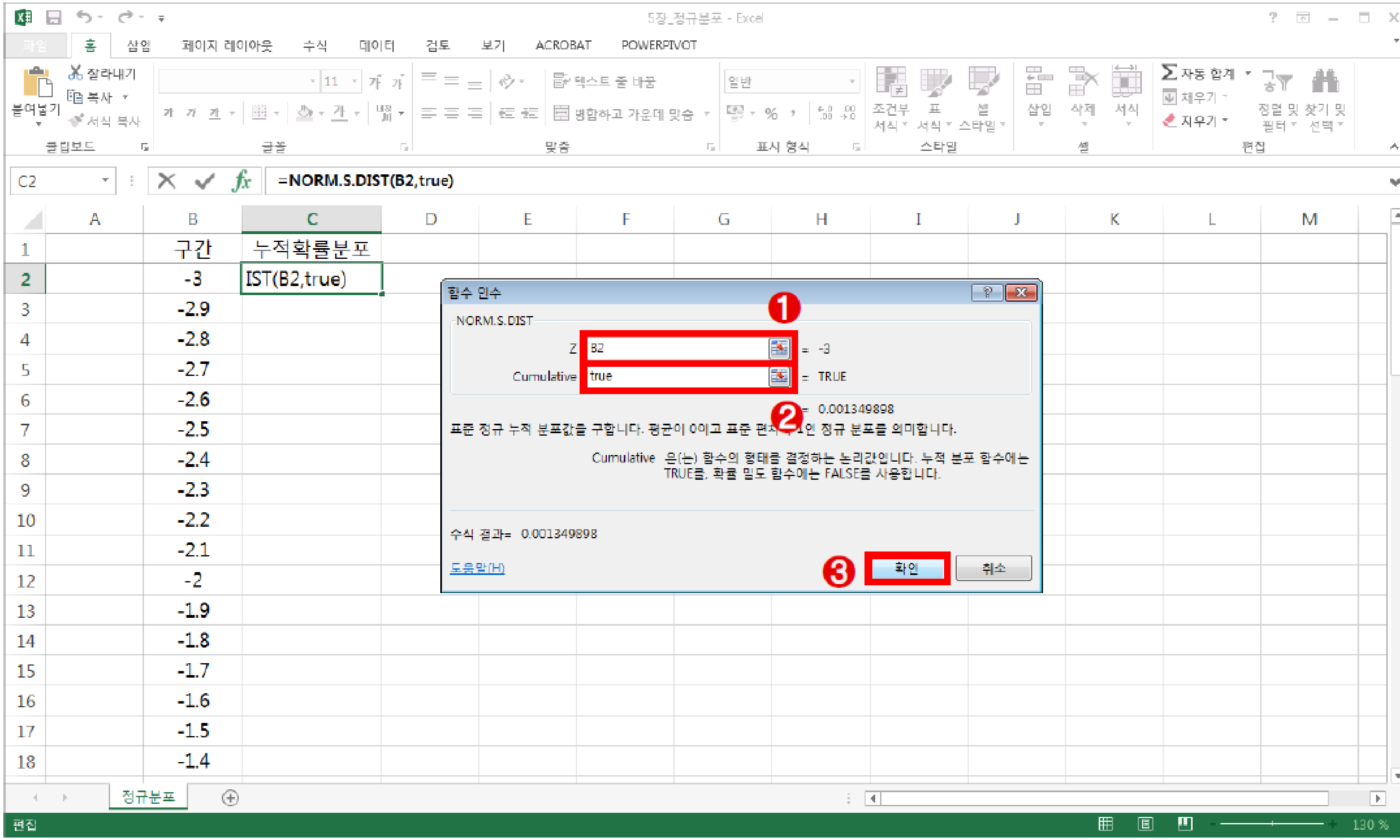
$z = 1.5$일 때까지의 누적확률을 빼야 한다.

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

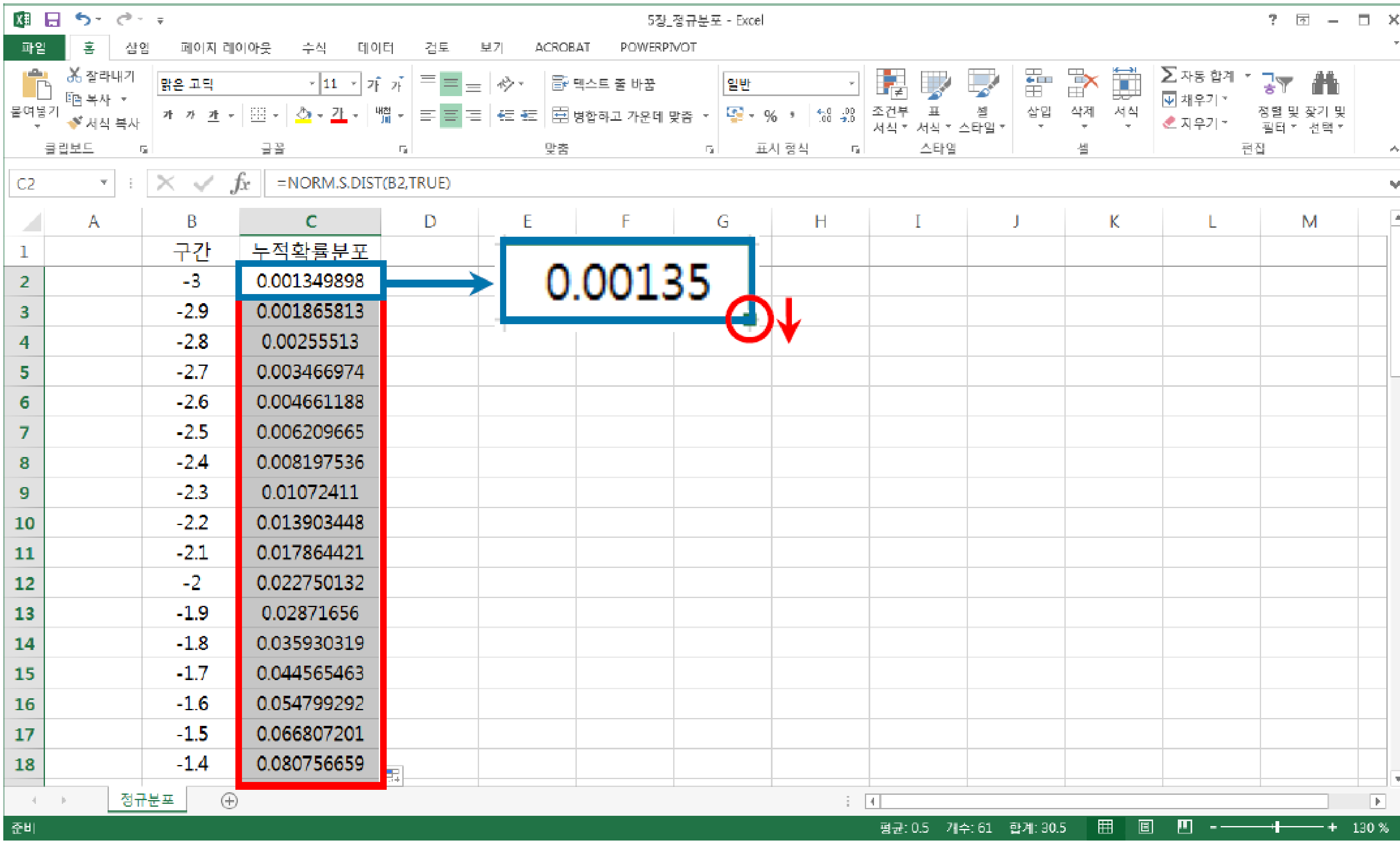
정규분포와 표준정규분포 (예제 Excel 풀이)



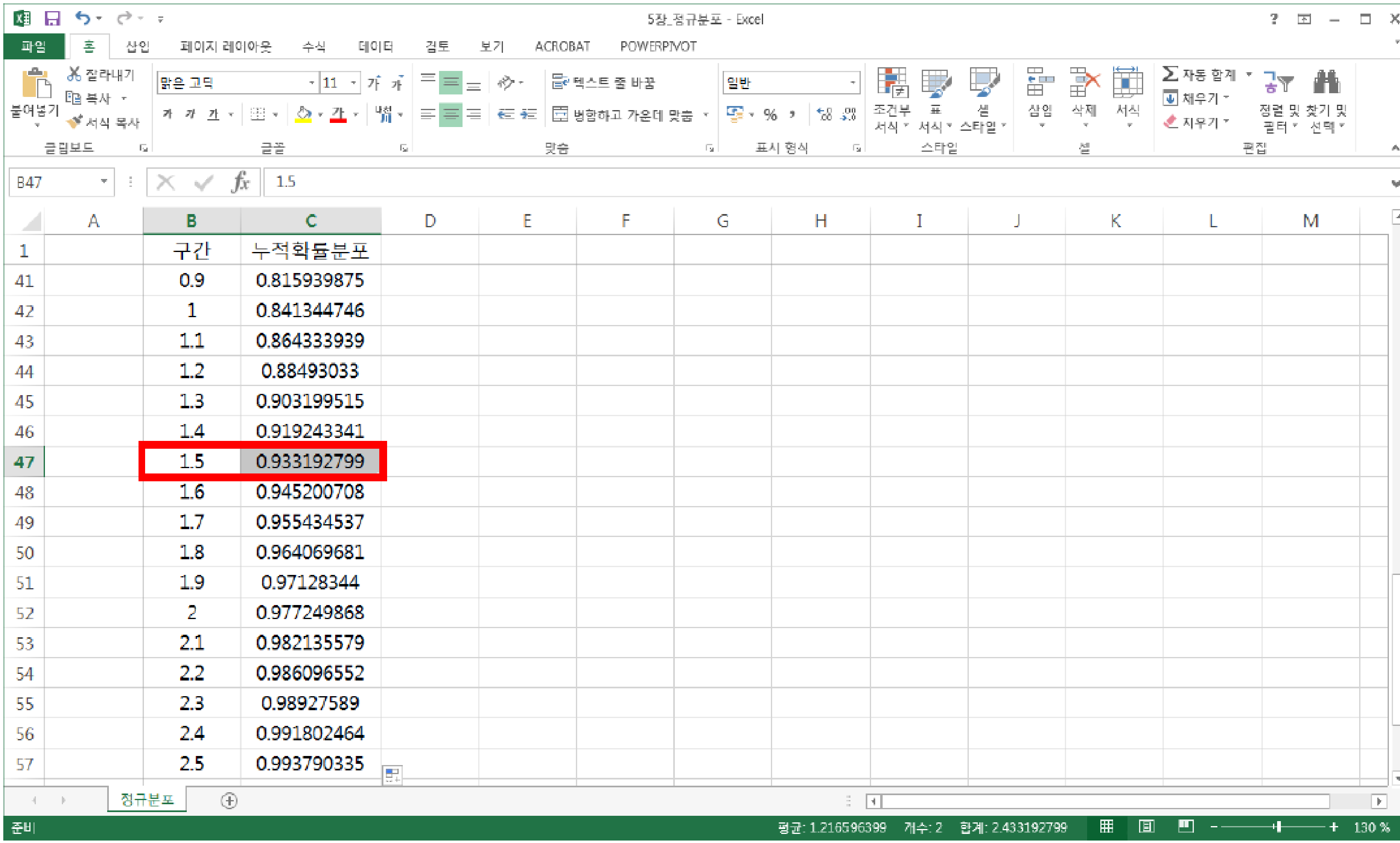
정규분포와 표준정규분포 (예제 Excel 풀이)



정규분포와 표준정규분포 (예제 Excel 풀이)

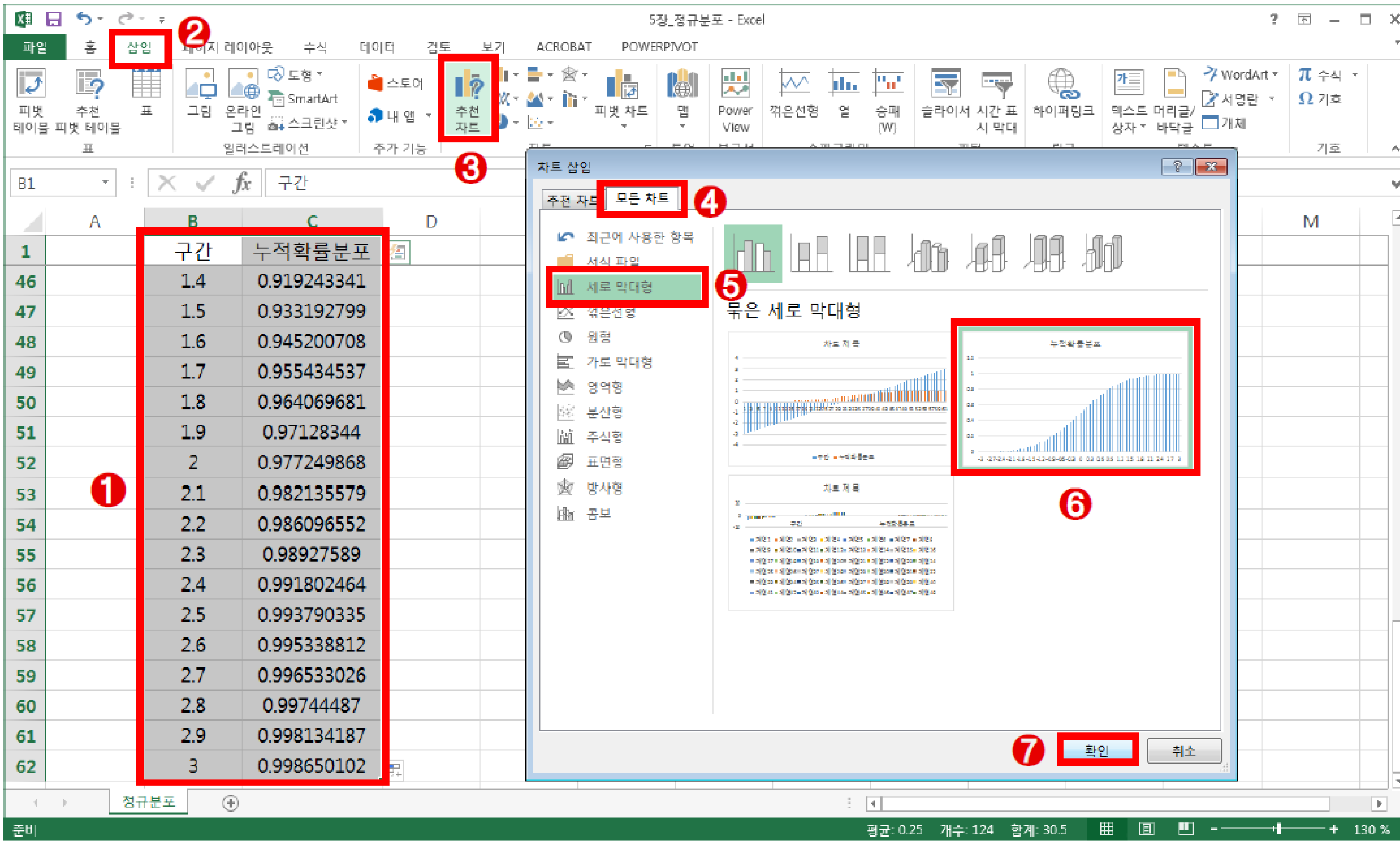


정규분포와 표준정규분포 (예제 Excel 풀이)

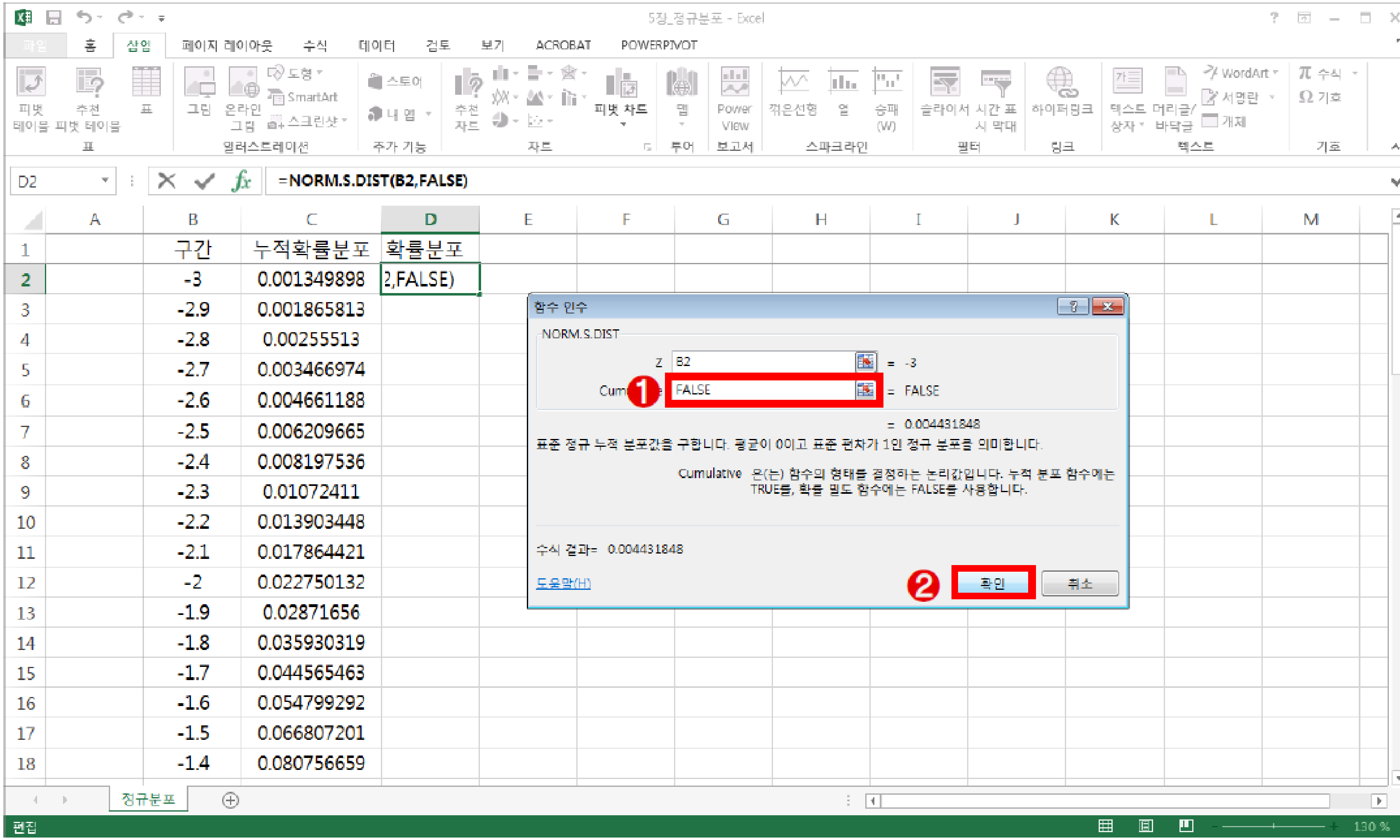


	A	B	C
1		구간	누적확률분포
41		0.9	0.815939875
42		1	0.841344746
43		1.1	0.864333939
44		1.2	0.88493033
45		1.3	0.903199515
46		1.4	0.919243341
47		1.5	0.933192799
48		1.6	0.945200708
49		1.7	0.955434537
50		1.8	0.964069681
51		1.9	0.97128344
52		2	0.977249868
53		2.1	0.982135579
54		2.2	0.986096552
55		2.3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

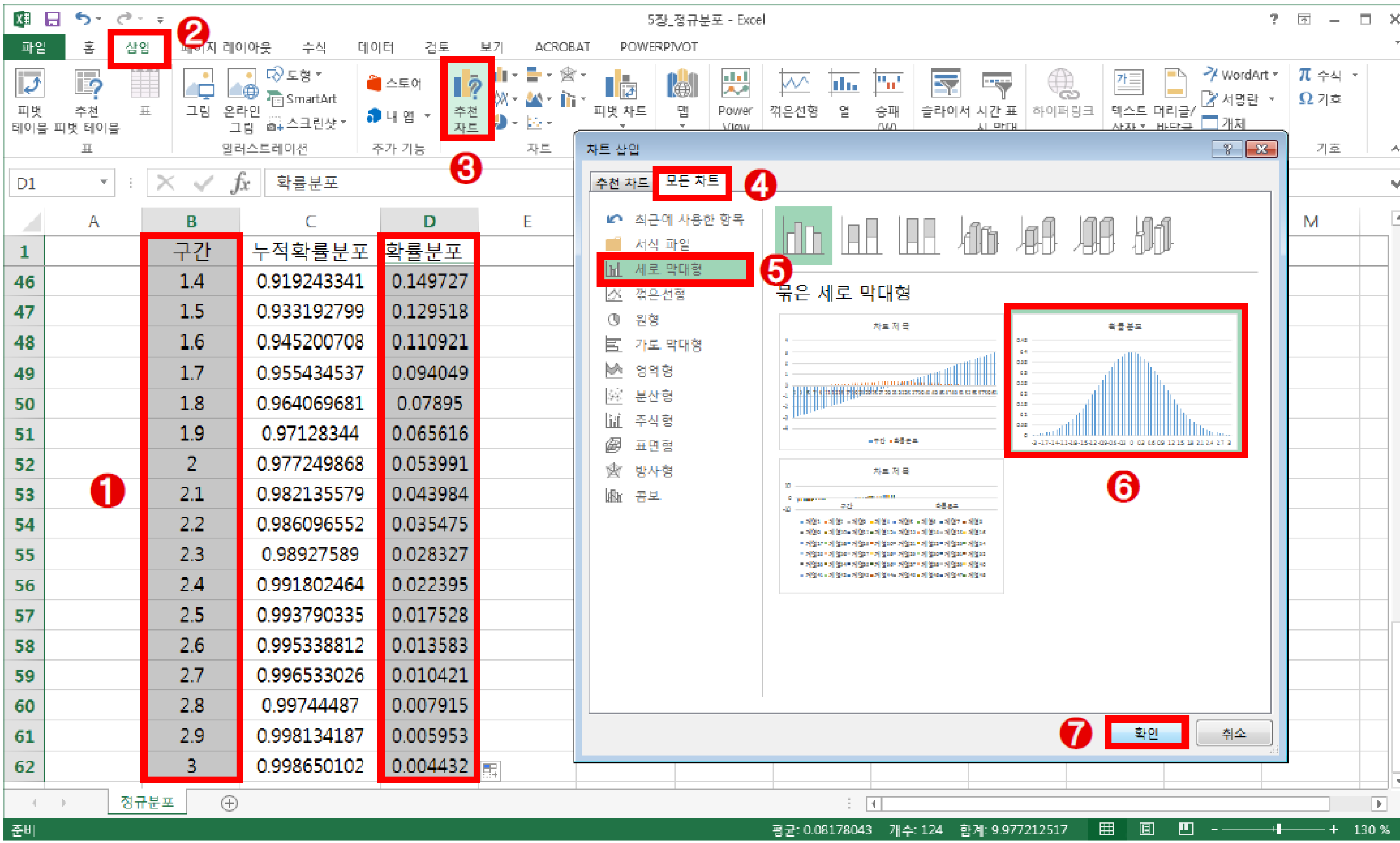
정규분포와 표준정규분포 (예제 Excel 풀이)



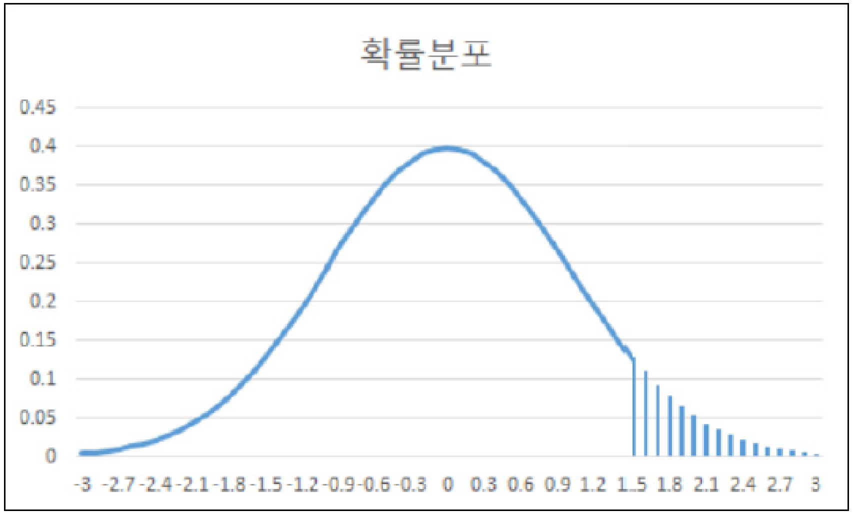
정규분포와 표준정규분포 (예제 Excel 풀이)

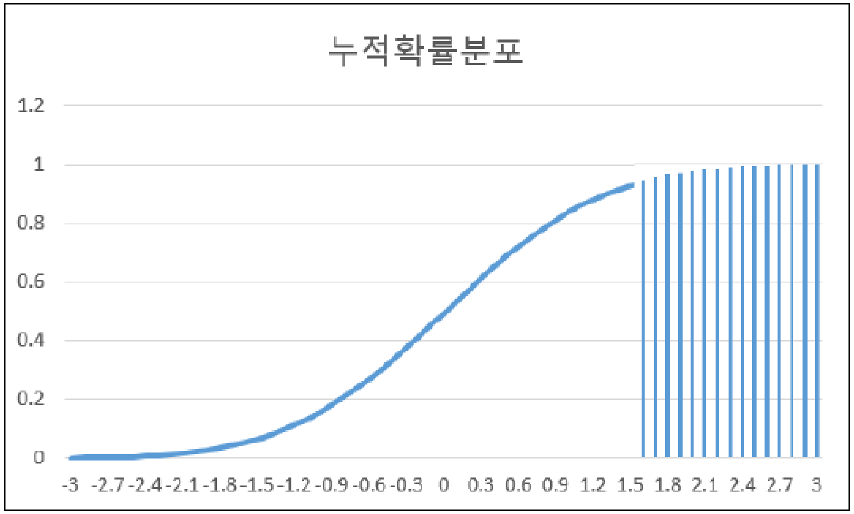


정규분포와 표준정규분포 (예제 Excel 풀이)



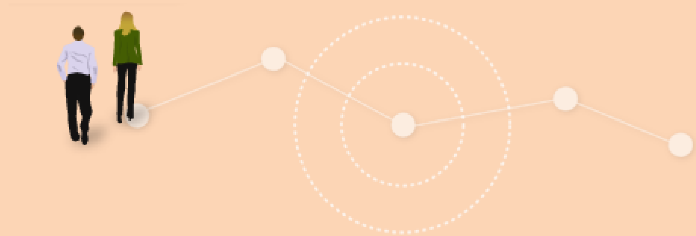
정규분포와 표준정규분포 (예제 Excel 풀이 완성)





02 이항분포

:: Keywords 베르누이 분포 | 이항분포 | 이항분포의 확률 계산



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

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

이항분포

예제 5-3 이항분포의 확률 계산

준비파일 | 5장_이항분포의 확률계산_함수.xlsx

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.

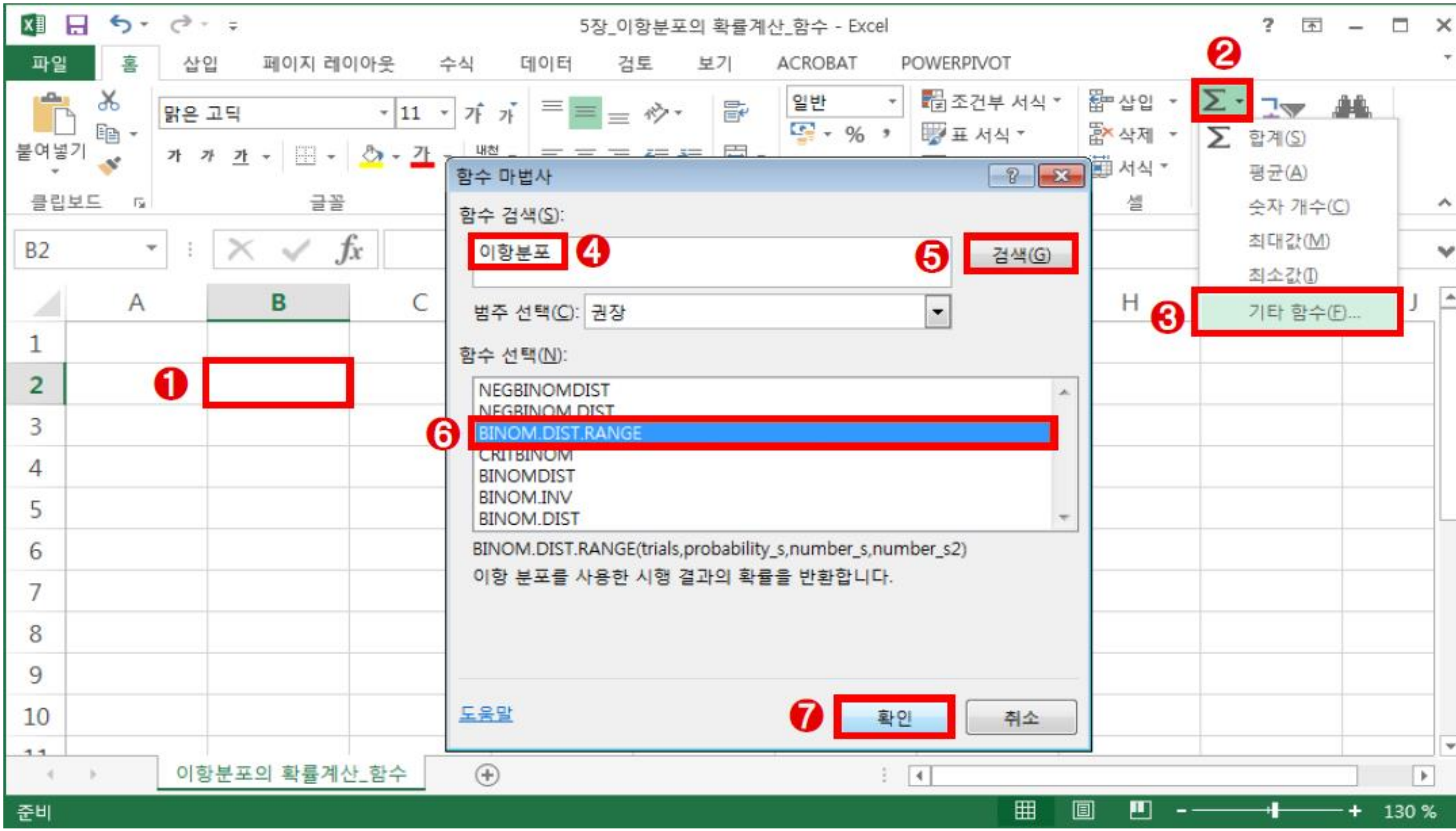
이항분포 (예제 풀이)

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

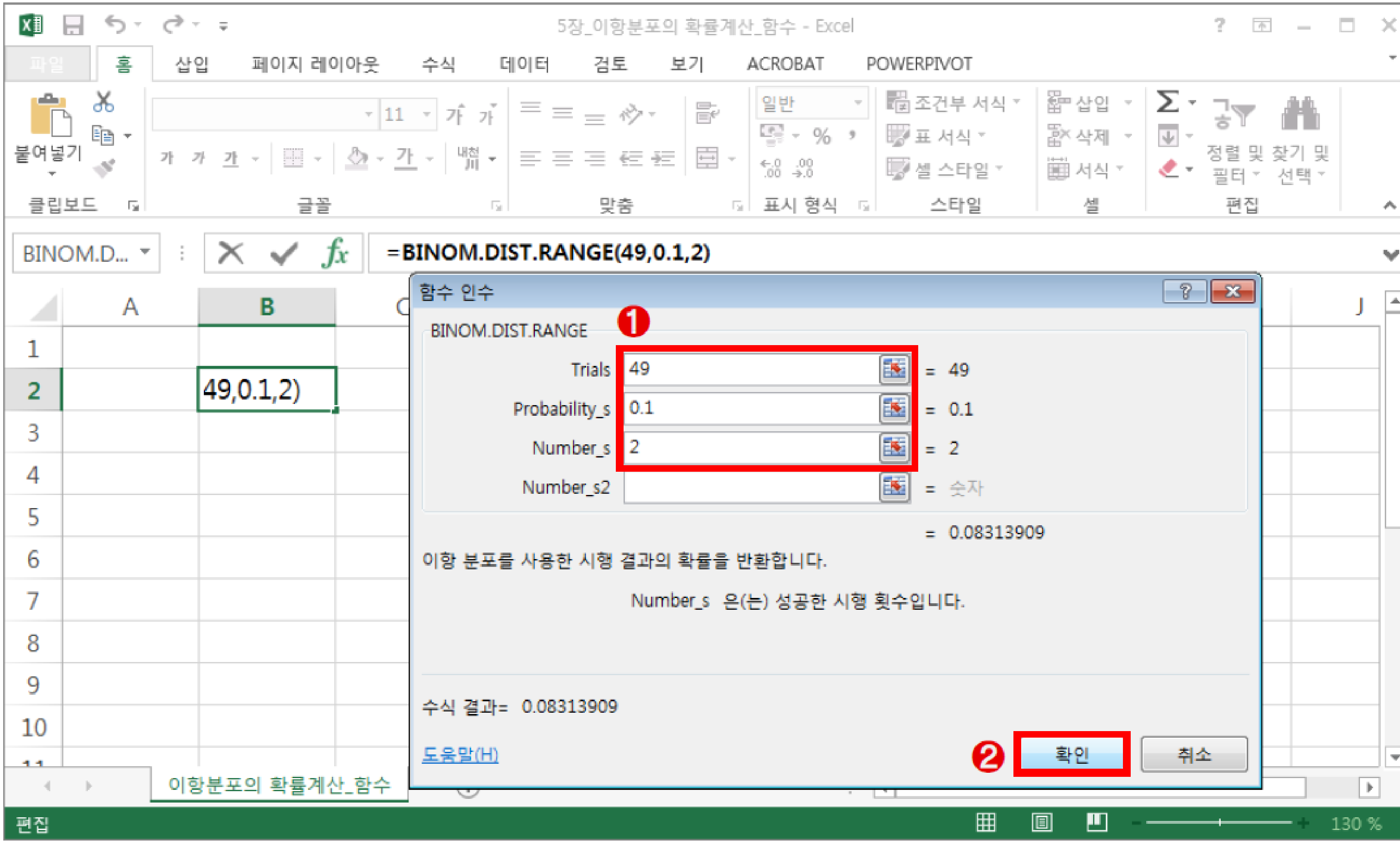
$$P(X=2) = \frac{49!}{2! \cdot (49-2)!} \cdot 0.1^2 \cdot (1-0.1)^{(49-2)}$$
$$= 0.08313909$$

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

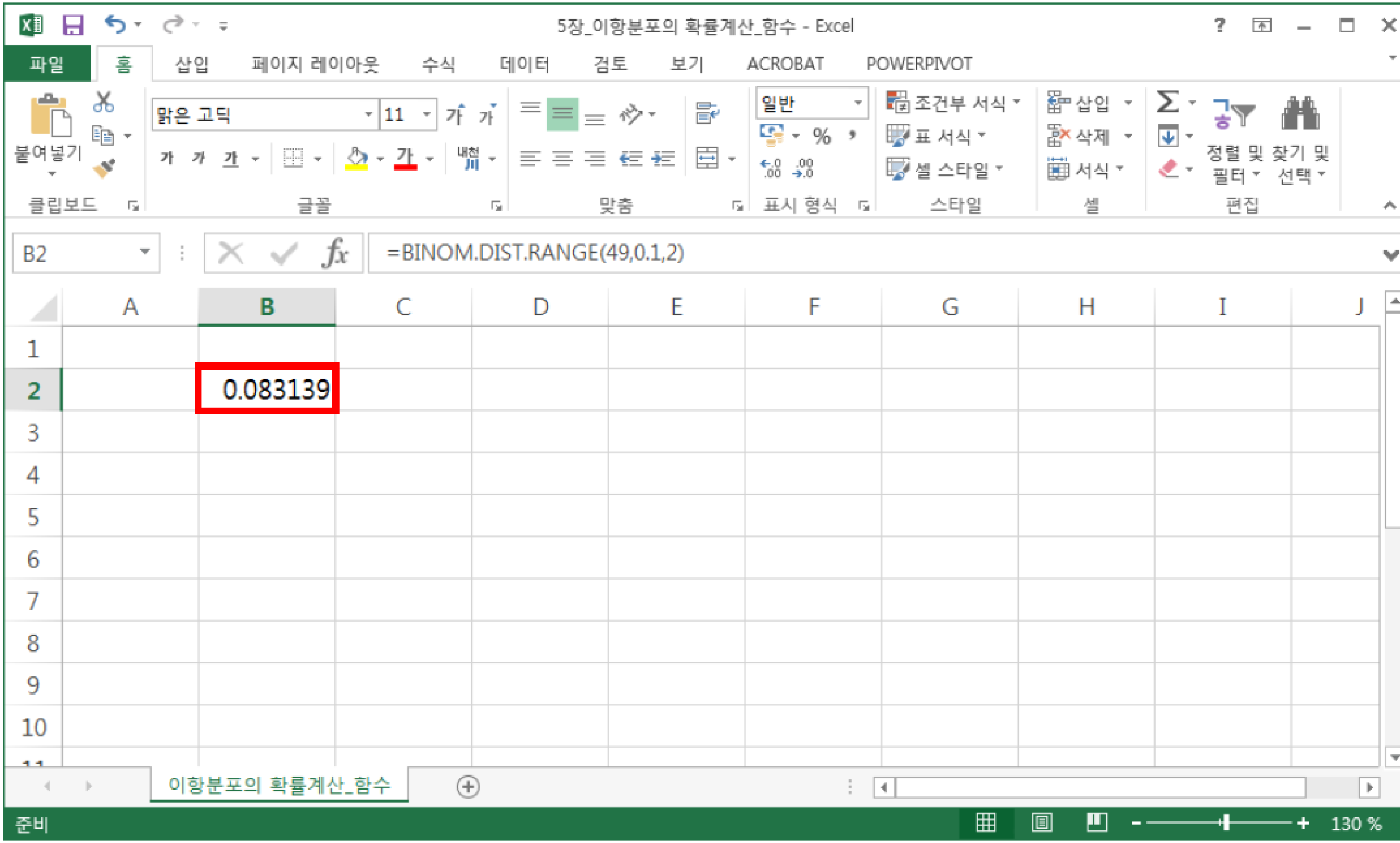
이항분포 (예제 Excel 풀이)



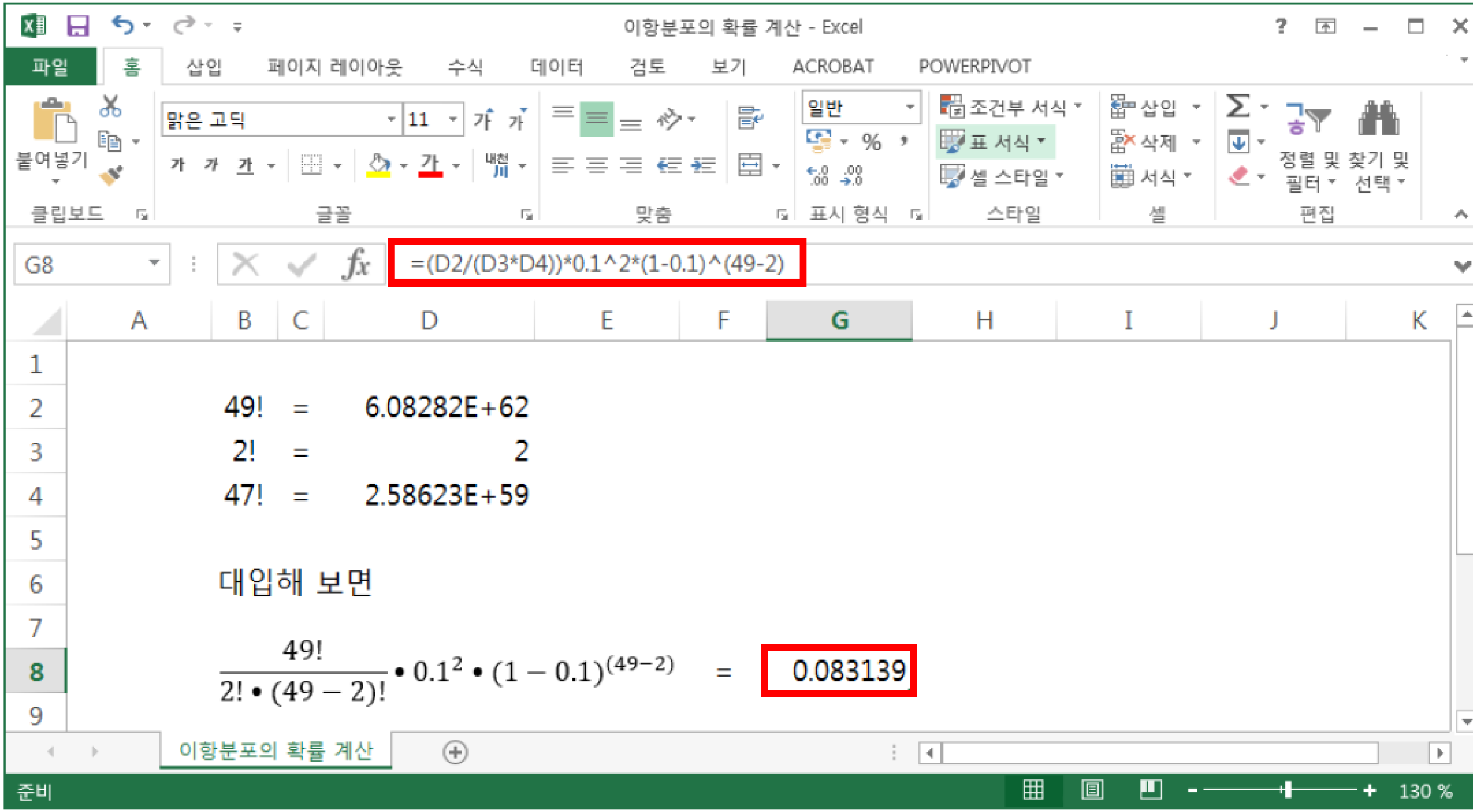
이항분포 (예제 Excel 풀이)



이항분포 (예제 Excel 풀이 완성)

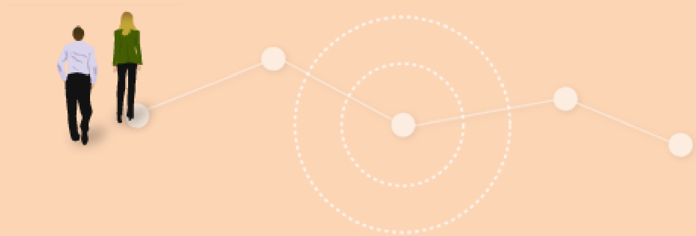


이항분포 (예제 Excel 풀이 완성-참고)



03 포아송분포

:: Keywords 포아송분포 | 포아송분포의 확률 계산
포아송분포와 정규분포와의 관계



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

→ 서로 반대되는 사건이란

반드시 두 개만 존재하며 절대 동시에 발생하지 않는 배타적인 사건

Ex. 동전 던지기와 주사위 던지기

포아송분포

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

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

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 $\dfrac{\lambda}{n}$

${}_nC_x \cdot p^x \cdot (1-p)^{(n-x)}$ 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.

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

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

포아송분포의 확률 계산

예제 5-4 포아송분포의 확률 계산

준비파일 | 5장_포아송분포의 확률 계산.xlsx

A는 건망증이 심해서 출근하거나 퇴근할 때 충전하던 휴대전화를 집이나 사무실에 종종 놓고 간다. 1주일 단위로 살펴봤을 때 휴대전화를 놓고 간 평균 횟수가 3회였다. 일주일에 단 1회 이하로 휴대전화를 놓고 갈 확률과 4 ~ 5회 휴대전화를 놓고 갈 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 $P(4 \leq X \leq 5)$

휴대전화를 4회 놓고 가는 경우의 확률함수

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

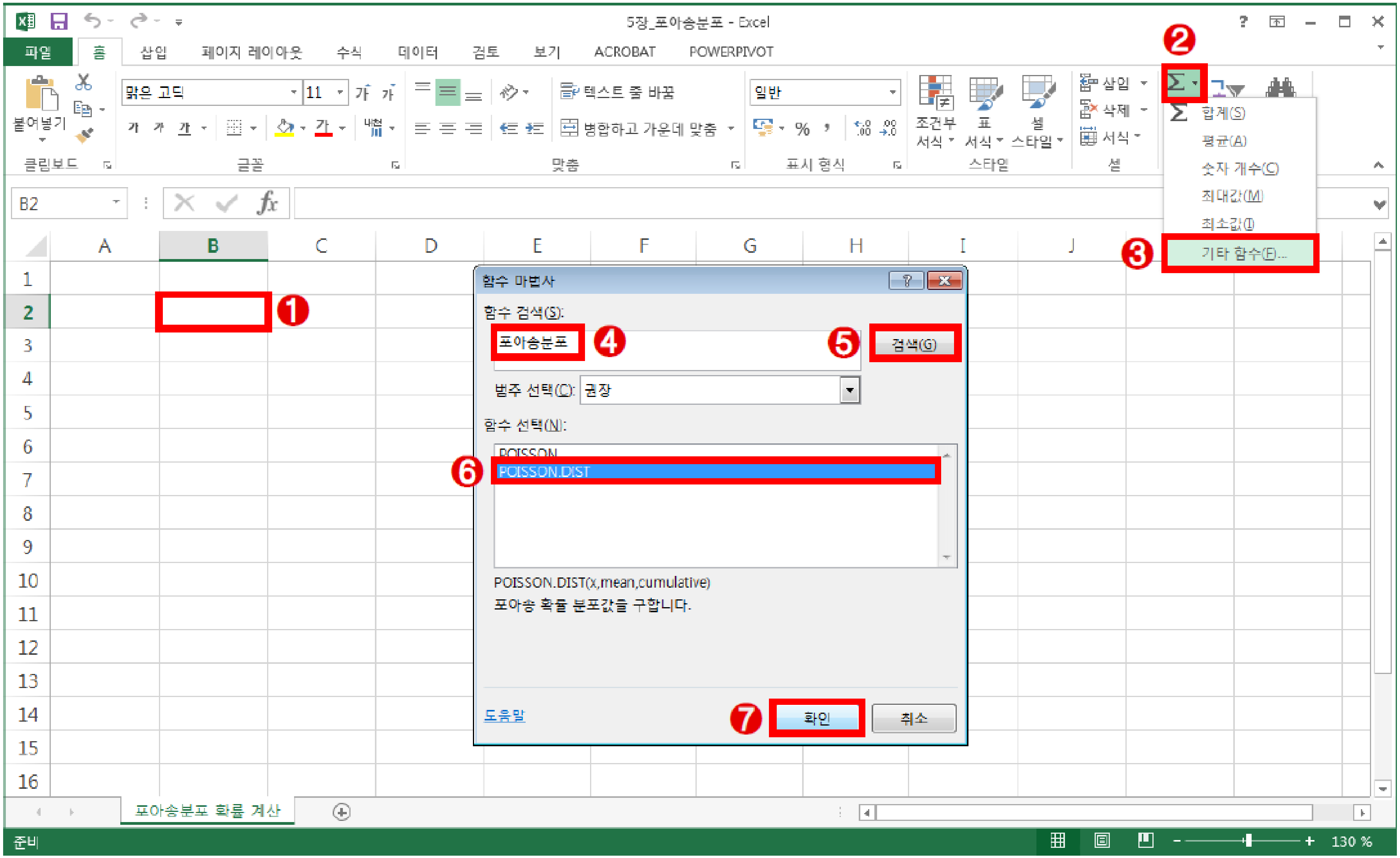
휴대전화를 5회 놓고 가는 경우의 확률함수

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

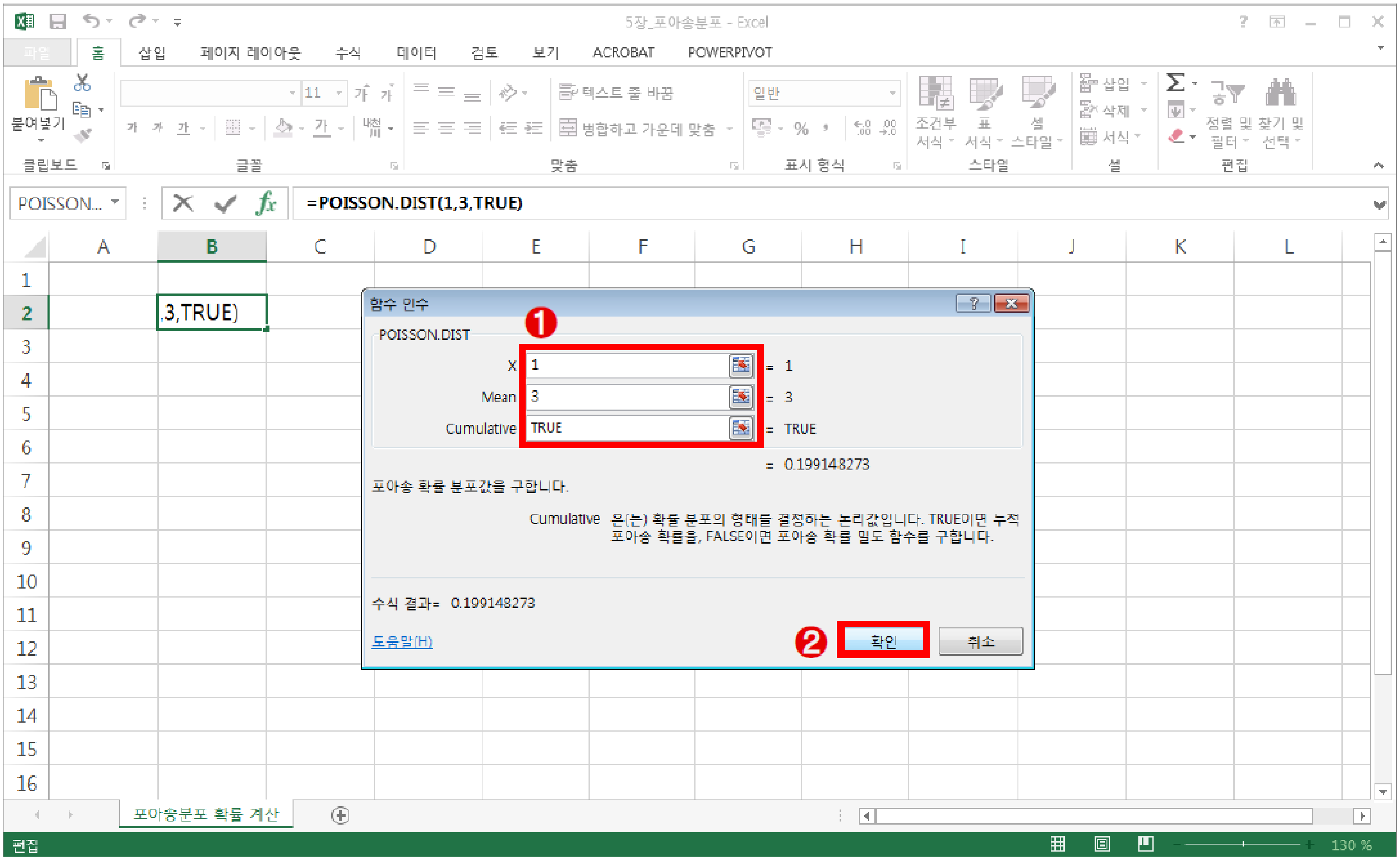
두 값을 더하면 0.26885

$\therefore$ 4~5회 휴대전화를 놓고 갈 확률은 26.89%

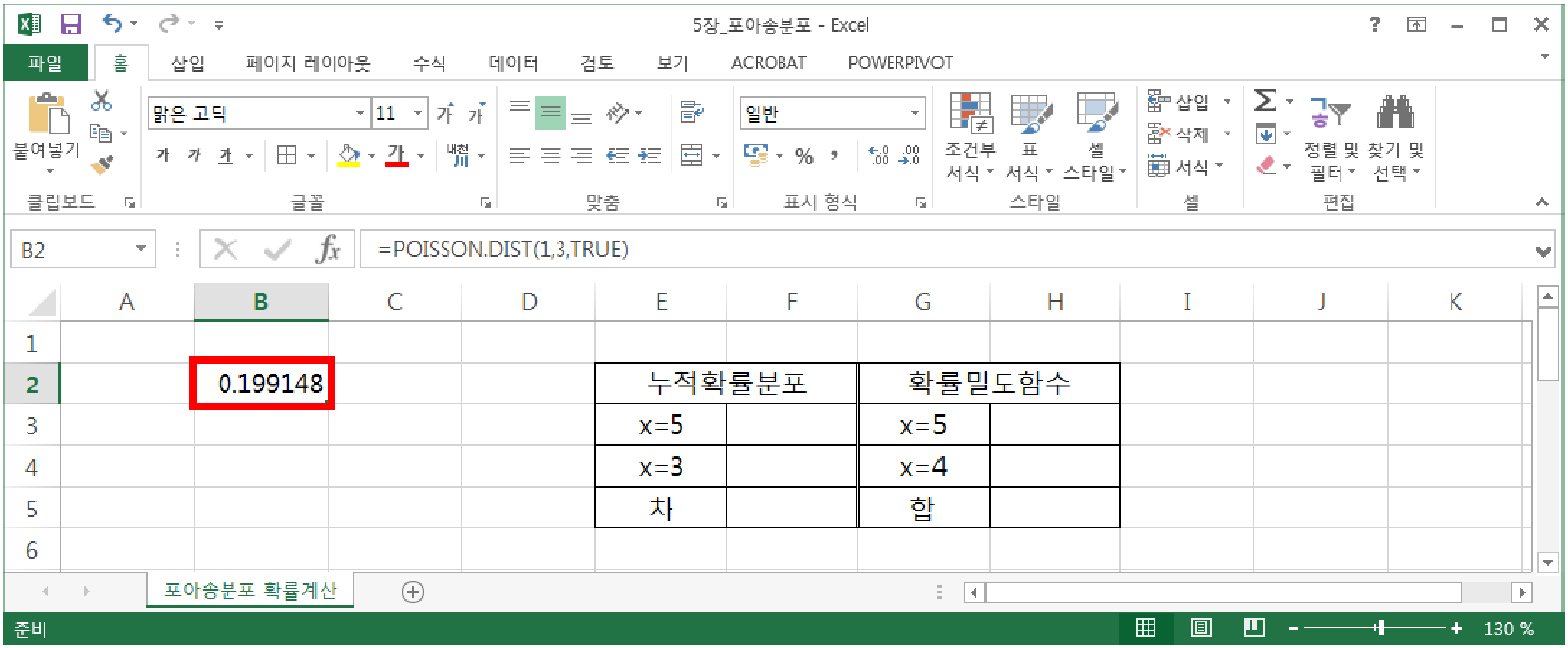
포아송분포의 확률 계산 (예제 Excel 풀이)



포아송분포의 확률 계산 (예제 Excel 풀이)

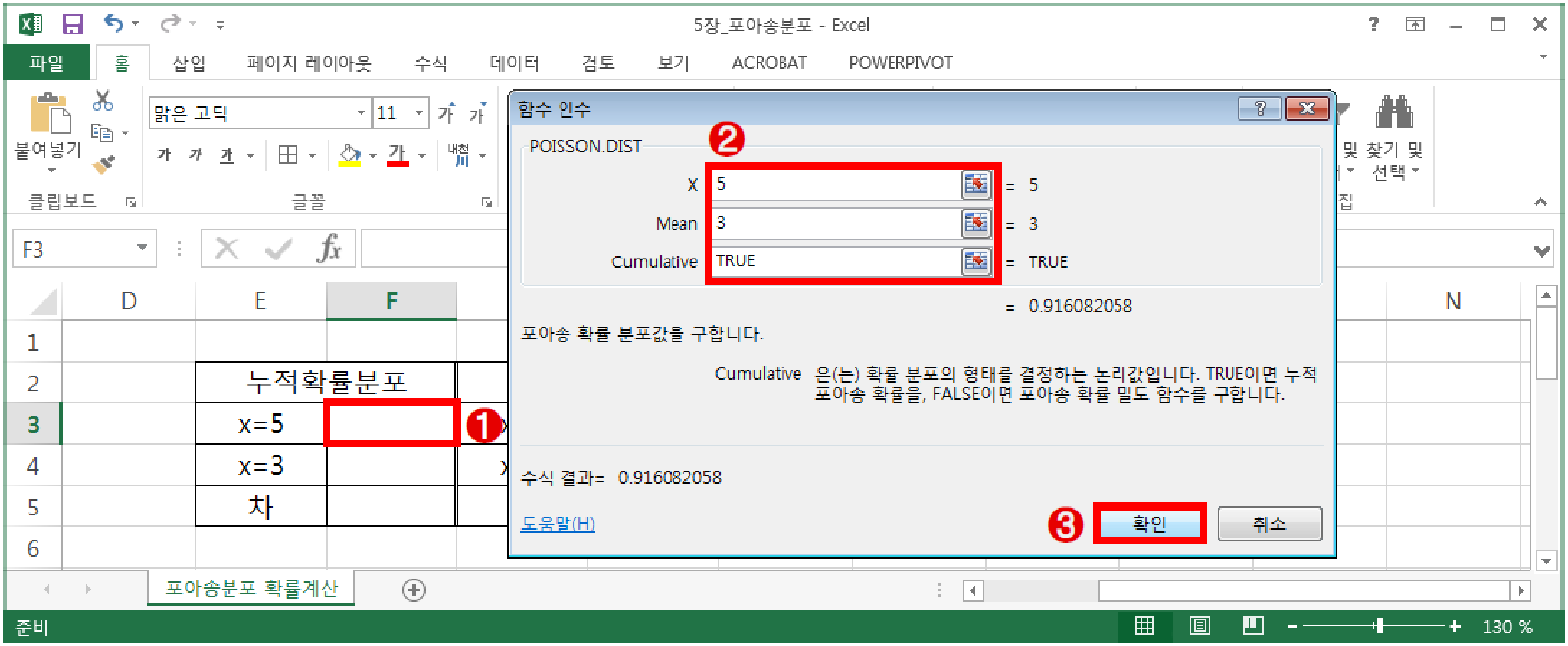


포아송분포의 확률 계산 (예제 Excel 풀이)

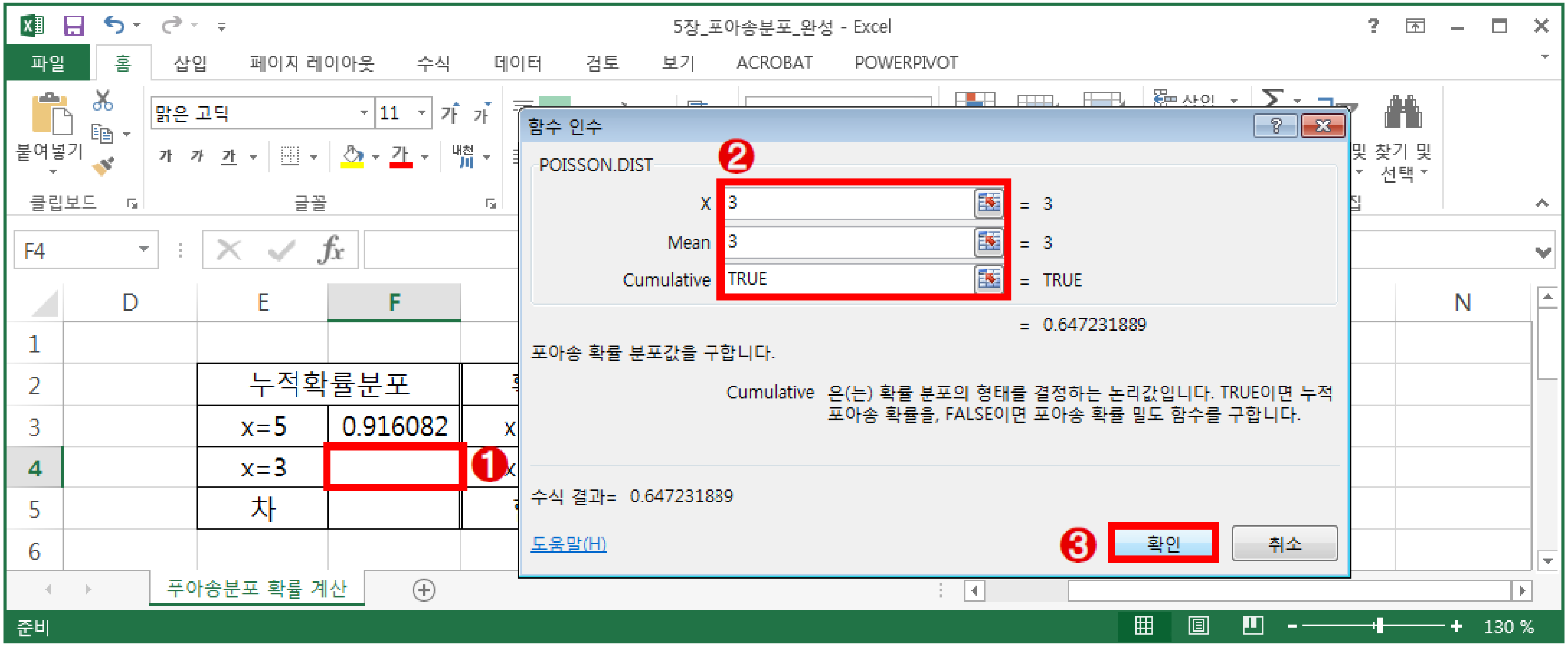


포아송분포의 확률 계산 (예제 Excel 풀이)

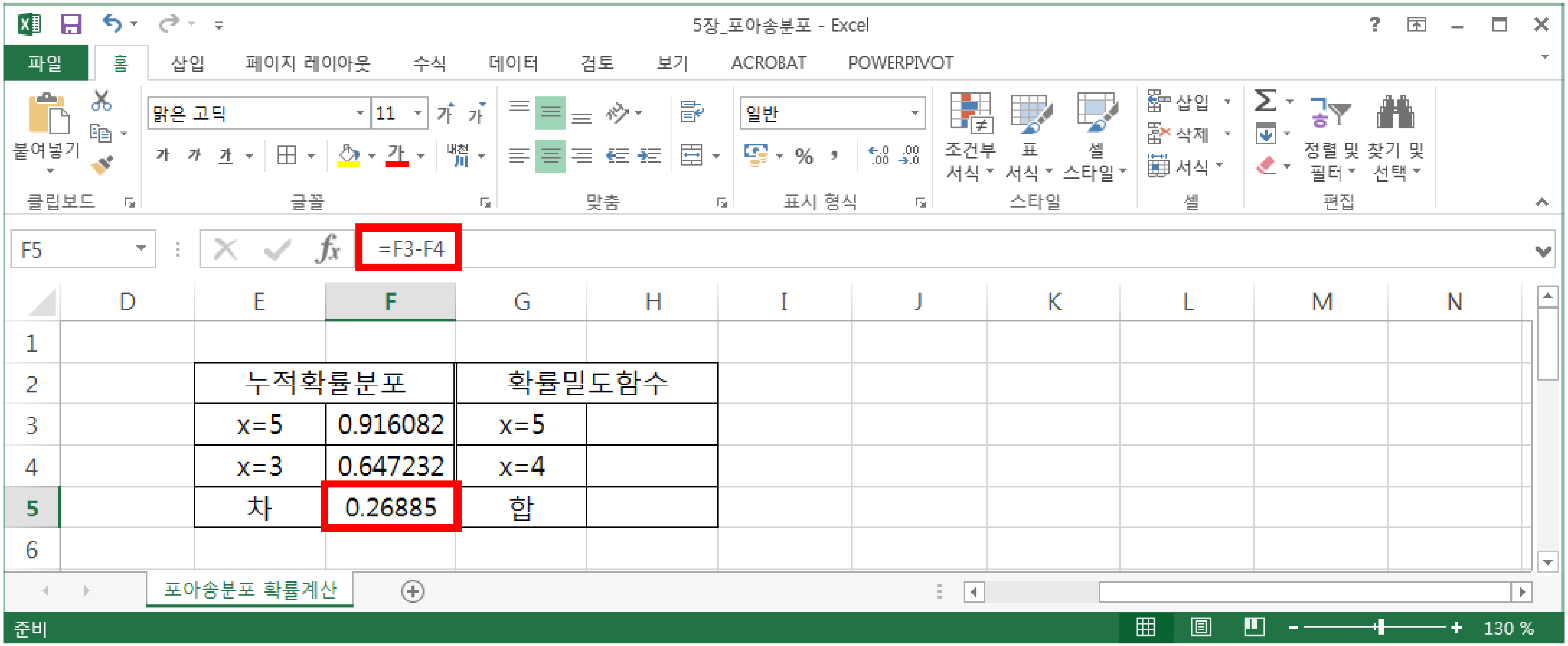
■ 4~5회 휴대전화를 놓고 갈 확률



포아송분포의 확률 계산 (예제 Excel 풀이)

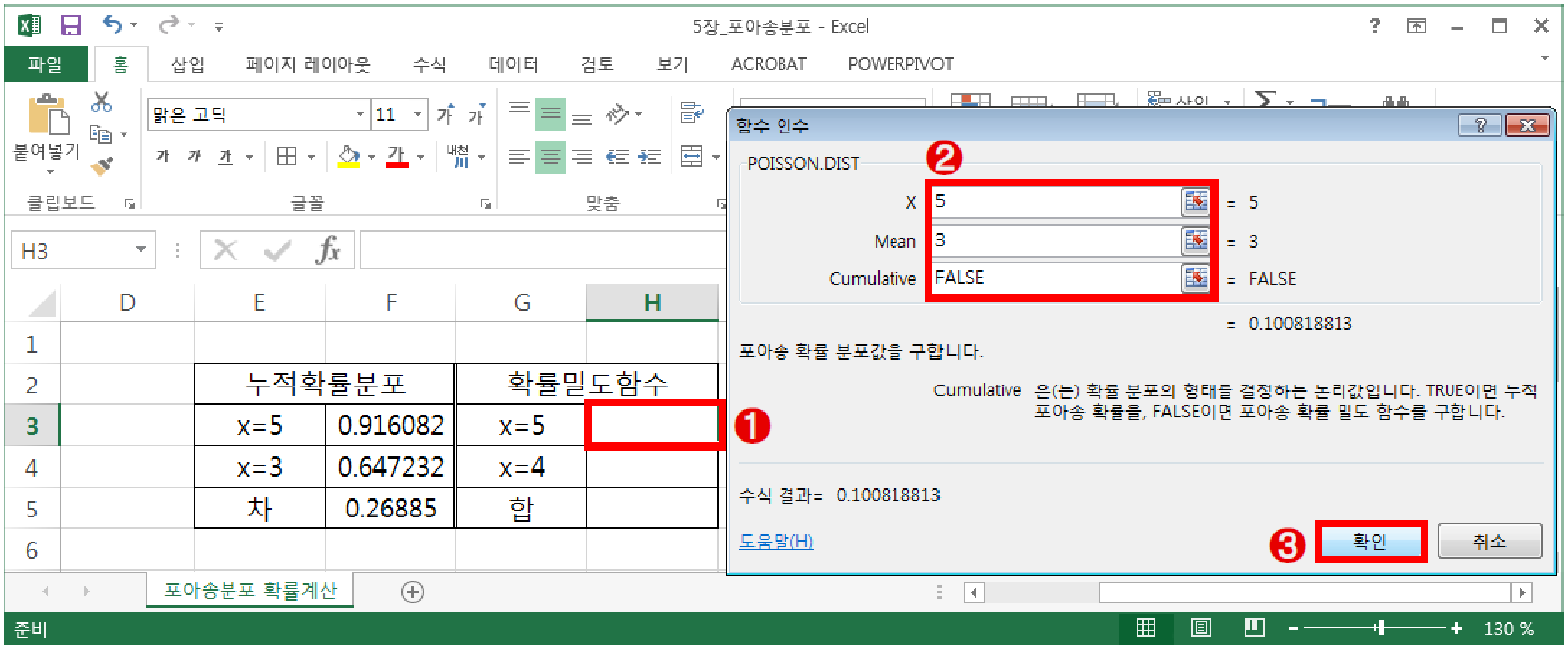


포아송분포의 확률 계산 (예제 Excel 풀이)

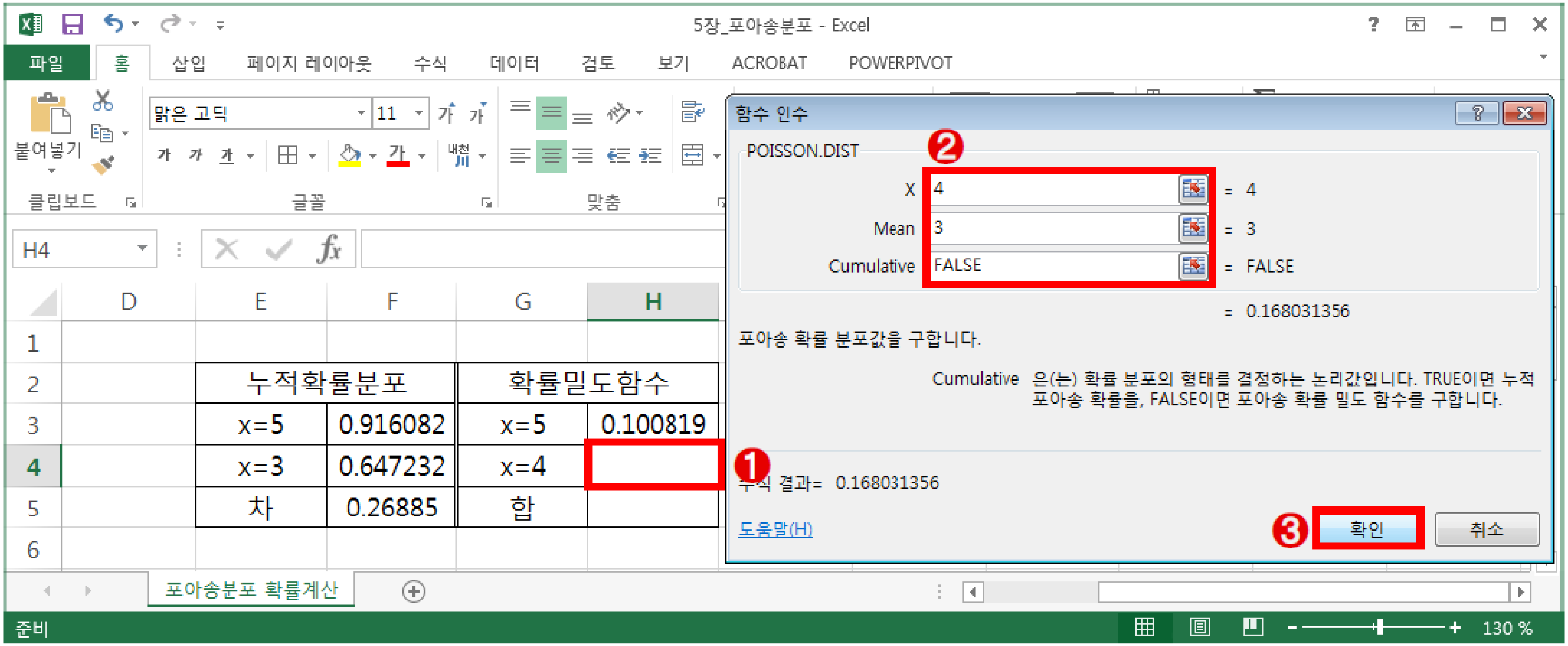


포아송분포의 확률 계산 (예제 Excel 풀이)

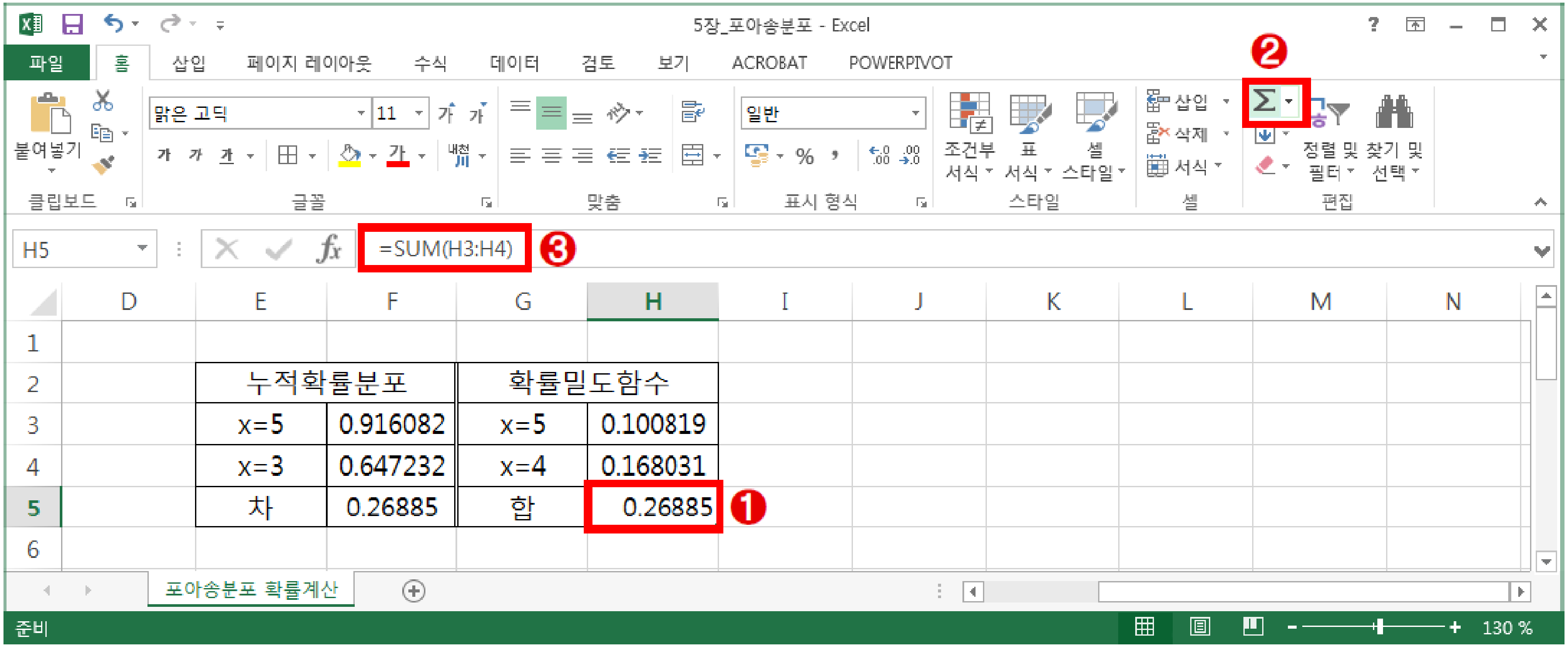
■ 확률밀도함수를 이용하는 경우



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
x	1	0.3678794	0.2706706	0.1493612	0.0732626	0.0336897	0.0148725	0.0063832
	2	0.1839397	0.2706706	0.2240418	0.1465251	0.0842243	0.0446175	0.0223411
	3	0.0613132	0.180447	0.2240418	0.1953668	0.1403739	0.0892351	0.0521293
	4	0.0153283	0.0902235	0.1680314	0.1953668	0.1754674	0.1338526	0.0912262
	5	0.0030657	0.0360894	0.1008188	0.1562935	0.1754674	0.1606232	0.1277167
	6	0.0005109	0.0120298	0.0504094	0.1041956	0.1462228	0.1606232	0.1490028
	7	7.299E-05	0.0034371	0.021604	0.0595404	0.1044449	0.137677	0.1490028
	8	9.124E-06	0.0008593	0.0081015	0.0297702	0.065278	0.1032577	0.1303774
	9	1.014E-06	0.0001909	0.0027005	0.0132312	0.0362656	0.0688385	0.1014047
	10	1.014E-07	3.819E-05	0.0008102	0.0052925	0.0181328	0.0413031	0.0709833