팀 SPC

미천한 수학자

촌철살인 확률과 통계 1, 2
네이버에 <촌철살인 확률과 통계 1>을 검색해보세요.!!

[정규분포 칼럼] 미천한 수학자

〈지엽적인 문제〉의 〈기준〉은 무엇인가?

- 다음의 기출문제를 풀어보자.

2008. 11. 나형(28%). 29번. 4점
다음은 어떤 모집단의 확률분포표이다.

X	10	20	30	계
$\mathrm{P}(\mathrm{X}=\mathrm{x})$	$\frac{1}{2}$	a	$\frac{1}{2}-a$	1

이 모집단에서 크기가 2인 표본을 복원 추출하여 구한 표본평균을 $\overline{X}$라 하자.
$\overline{X}$의 평균이 18일 때, $P(\overline{X}=20)$의 값은?

① $\frac{2}{5}$ ② $\frac{19}{50}$ ③ $\frac{9}{25}$ ④ $\frac{17}{50}$ ⑤ $\frac{8}{25}$

- 이 문제는 2008년 인문계열 수능문제이다.
이 문제가 객관식임에도 불구하고 정답률이 28%라는 점은 의미하는 바가 참 크다.
(그 당시에는 굉장히 풀기 어려웠다는 뜻이다.)
만약 비슷한 문제가 지금 나온다면 최소 60% ~ 80%의 정도의 정답률이 나올 것이라고 생각한다.

이 당시에 〈표본정규분포가 기본적으로는 이산확률분포〉를 따른다는 것과 〈복원추출〉의 개념은 이전에는 물어본 적이 없는 매우 지엽적인 문제였다. 물론 〈지엽적〉이라는 말의 기준이 〈이전의 기출문제, 사설 문제집에 나오는 빈도수〉라면 말이다.
즉, 문제가 객관적으로 어려웠던 것이 아니라 생각해 본 적이 없었던 것이다. 정답은 ④

[정규분포 칼럼] 미천한 수학자

- 이번에는 아주 간단한 <이항분포와 정규분포의 관계>를 풀어보자.

다음은 이항분포 $\mathrm{B}\left(100,\ \frac{1}{10}\right)$을 따르는 확률변수 X에 대하여 $\mathrm{P}(10 \leq \mathrm{X} \leq 11)$의 값을 다음 표준정규분포표를 이용하여 구하여라. (단, $\frac{1}{3}$은 0.33으로 계산한다.)

z	$\mathrm{P}(0 \leq \mathrm{Z} \leq z)$
0.33	0.1293
0.66	0.2454
0.99	0.3315

[정규분포 칼럼] 미친한 수학자

- 혹시 정답을 0.1293이라고 했을지도 모른다. 정답은 0.2454이다.
  왜 그럴까? 그럼 다음 문제를 풀어보시길 바란다. (다음 장에 해설도 참고하라.)

다음은 이항분포 $\mathrm{B}\left(100,\ \frac{1}{10}\right)$을 따르는 X의 확률질량함수

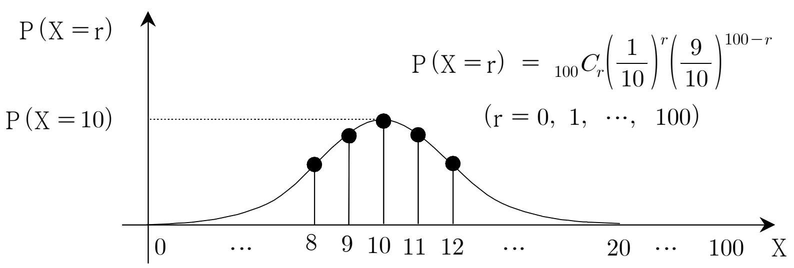
$$\mathrm{P}(\mathrm{X}=\mathrm{r}) = {}_{100}C_r\left(\frac{1}{10}\right)^r\left(\frac{9}{10}\right)^{100-r} \quad (\mathrm{r}=0,\ 1,\ \cdots,\ 100)$$

에 대하여 이 확률질량함수의 그래프와 정규분포를 이용하여 $\mathrm{P}(10 \le \mathrm{X} \le 11)$의 값을 구하는 과정이다.

(단, $\frac{1}{3}$은 0.33으로 계산한다.)

이항분포 $\mathrm{B}\left(100,\ \frac{1}{10}\right)$의 확률질량함수 X는 근사적으로 정규분포 $\mathrm{N}(10,\ 3^2)$을 따른다.

이것은 X의 확률질량함수 $\mathrm{P}(\mathrm{X}=\mathrm{r})$ $(\mathrm{r}=0,\ 1,\ \cdots,\ 100)$의 그래프를 좌표평면에 표시하여
각 함숫값을 연결했을 때 아래그림과 같이 근사적으로 정규분포와 비슷한 모양을 가진다는 뜻이다.



⋮

(중략)

이항분포 $\mathrm{B}\left(100,\ \frac{1}{10}\right)$의 확률질량함수 X에 대하여

$$\mathrm{P}(10 \le \mathrm{X} \le 11) = \mathrm{P}(10) + \mathrm{P}(\boxed{\text{(가)}})$$

이고, 이 값은 위의 그래프를 참조하면 정규분포 $\mathrm{N}(10,\ 3^2)$
에서 $\mathrm{P}(10 \le \mathrm{X} \le \boxed{\text{(나)}})$의 값으로 근사된다.

z	$\mathrm{P}(0 \le \mathrm{Z} \le \mathrm{z})$
0.33	0.1293
0.66	0.2454
0.99	0.3315

즉, 위의 표준정규분포 표를 참조하면

$$\mathrm{P}(10 \le \mathrm{X} \le \boxed{\text{(나)}}) = \mathrm{P}\left(0 \le \mathrm{Z} \le \frac{\boxed{\text{(나)}} - 10}{3}\right) = \boxed{\text{(다)}}$$이다.

위의 증명에서 (가)에 들어갈 수를 a, (나)에 들어갈 수를 b, (다)에 들어갈 숫자를 c라고 할 때, $a+b+c$의 값은?

① 22.1293 ② 23.1293 ③ 22.2454
④ 23.2454 ⑤ 24.3315

[정규분포 칼럼] 미천한 수학자

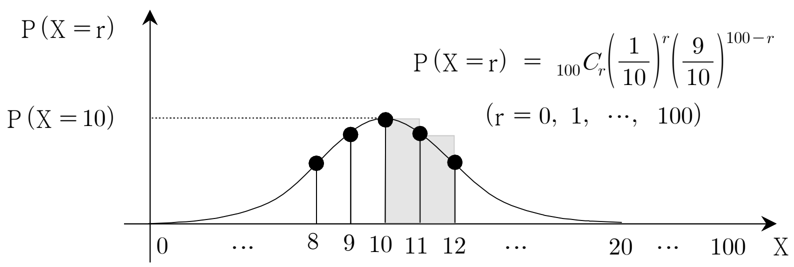
- 해설

X	$\cdots$	10	11	$\cdots$	총합
$P(X=r)$	$\cdots$			$\cdots$	1

이항분포 $B\left(100,\ \frac{1}{10}\right)$의 확률질량함수 X를 간단히 표로 나타내면 위와 같고, 이항분포는 역시 이산확률분포이므로

$$P(10 \le X \le 11) \ = \ P(10) + P(\boxed{11})$$이다.

즉, $\boxed{(가) \ = \ 11}$



$P(10)+P(\boxed{11})$은 주어진 그래프를 참조하면 위와 같이 밑변의 길이가 1이고 높이가 $P(10)$과 $P(11)$인 두 직사각형에 대응된다.

즉, $P(10)+P(\boxed{11})$는 이 이항분포가 근사되는 정규분포의 입장에서는 $P(10 \le X \le 11)$보다는 $P(10 \le X \le \boxed{12})$에 가깝다.
그리고 실제로 이 두 확률 값의 차이는 0.1161이나 날 정도로 절대 무시할 수 없는 차이가 생긴다. 즉, $\boxed{(나) = 12}$

결국 $P(10 \le X \le \boxed{12})$을 표준화시키면 $P(0 \le Z \le 0.66) \ = \ \boxed{0.2454}$이다. 즉, $\boxed{(다) = 0.2454}$
그러므로 정답은 ④

comment 1. 수많은 문제집들에 나오는 대표적인 오개념이 이항분포와 정규분포와의 관계이다.
이런 이유로 〈지금까지의 평가원과 교육청의 대부분의 문제〉는 이항분포의 확률을 정규분포로 구하는 과정에 대하여 $P(X \ge a)$만 물어봤다,

이렇게 물어보는 경우 〈아무 생각 없이 표준화〉한다고 하더라도 틀리지 않기 때문이다.
이것은 평가원의 아량이라고 생각한다.
2009 사관학교 문제처럼 $B\left(720,\ \frac{1}{6}\right)$에서 $P(110 \le X \le 145)$을 물어보는 경우가 있는데 $N(120,\ 10^2)$ 을 따르기 때문에 (이렇게 표준편차가 큰 경우는) $P(110 \le X \le 145)$으로 계산하든 $P(110 \le X \le 146)$으로 계산하든 유의미한 차이가 발생하지는 않는다.

comment 2. 2017수능에서 빈칸문제가 확률로 나와서 만들어 봤다.
빈칸문제는 〈배운 내용 + 추론할 수 있는 내용〉이 그 전에 대세를 이루었던 문항이었다.
(주로 점화식 문제가 나왔다.) 그래서 문제집들의 흔한 오 개념을 가지고 문제를 만들어 봤다.
심지어 $P(10 < X < 11) \ = \ 0$인데 아무 생각 없이 표준화 하는 경우도 많이 있다.

이 문제가 지엽적이라고 말하는 사람들은 다음 장을 참고해 보길 바란다.

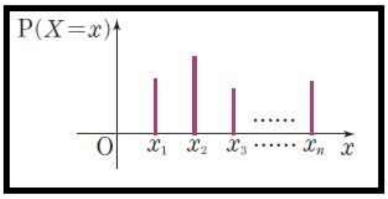
[정규분포 칼럼] 미천한 수학자

교학사 개정 교과서 발췌 1

- 교과서에서 이미 확률질량함수를 그래프로 그리는 시도를 여러 번 하고 있다.

확률은 0에서 1까지의 값을 가진다.

X	x_1	x_2	x_3	……	x_n	합계
$\mathrm{P}(X=x)$	p_1	p_2	p_3	……	p_n	1



확률질량함수를 그래프로 나타내는 교과서 내용

한편 확률변수 X가 a 이상 b 이하인 값을 가지는 확률을 $\mathrm{P}(a\le X\le b)$와 같이 나타낸다. 이때 X가 이산확률변수이면 $\mathrm{P}(a\le X\le b)=\sum_{x=a}^{b}\mathrm{P}(X=x)$이다.

일반적으로 확률질량함수 $\mathrm{P}(X=x_i)=p_i(i=1,\ 2,\ \cdots\cdots,\ n)$은 다음과 같은 성질을 갖는다.

❶ $0\le p_i\le 1$　　❷ $\sum_{i=1}^{n}p_i=1$

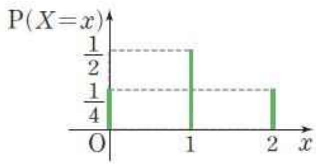
교학사 개정 교과서 발췌 2

- 이항분포도 정규분포이기 이전에 이산확률분포이다.

$\mathrm{P}(X=x)$
$={}_2\mathrm{C}_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{2-x}$
$={}_2\mathrm{C}_x\left(\frac{1}{2}\right)^2$
$(x=0,\ 1,\ 2)$

보기 **생각 열기**의 한 개의 동전을 2번 던지는 시행에서 앞면이 나오는 횟수를 확률변수 X라고 할 때, 확률변수 X의 확률분포를 표와 그래프로 나타내면 다음과 같다.

X	0	1	2	합계
$\mathrm{P}(X=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1



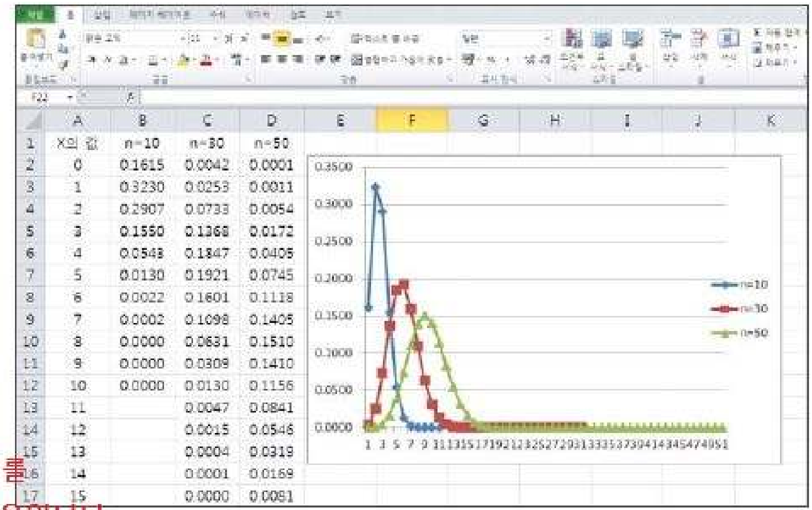
따라서 확률변수 X가 1 이상 2 이하인 값을 가질 확률은 다음과 같다.

이산확률분포는 확률 값을 오른쪽과 같이 <인지>하는 것이 가장 기본입니다.

$$\boxed{\begin{aligned}\mathrm{P}(1\le X\le 2)&=\mathrm{P}(X=1)+\mathrm{P}(X=2)\\&=\frac{1}{2}+\frac{1}{4}=\frac{3}{4}\end{aligned}}$$

[정규분포 칼럼] 미천한 수학자

교학사 개정 교과서 발췌 3



X의 값	n=10	n=30	n=50
0	0.1615	0.0042	0.0001
1	0.3230	0.0253	0.0011
2	0.2907	0.0733	0.0054
3	0.1550	0.1368	0.0172
4	0.0543	0.1847	0.0405
5	0.0130	0.1921	0.0745
6	0.0022	0.1601	0.1118
7	0.0002	0.1098	0.1405
8	0.0000	0.0631	0.1510
9	0.0000	0.0309	0.1410
10	0.0000	0.0130	0.1156
11		0.0047	0.0841
12		0.0015	0.0546
13		0.0004	0.0319
14		0.0001	0.0169
15		0.0000	0.0081

큰수의 법칙은 모든 교과서에
최소 2페이지 이상의 지면을
차지하고 있으며 <이항분포>를
이항분포 자체로 계산하는 내용입니다.
(물론 계산된 값을 이용하는 것이지요.)

[그림 1] 이항분포 $B\left(n,\ \frac{1}{6}\right)$의 표와 그래프

$\left|\frac{X}{n}-\frac{1}{6}\right|<0.1$

$\Leftrightarrow \frac{n}{15}<X<\frac{4n}{15}$

(i) $n=10$일 때,

$\frac{2}{3}<X<\frac{8}{3}$이므로

$X=1, 2$

(ii) $n=30$일 때,

$2<X<8$이므로

$X=3, 4, 5, 6, 7$

(iii) $n=50$일 때,

$\frac{10}{3}<X<\frac{40}{3}$이므로

$X=4, 5, \cdots\cdots, 13$

위의 표를 이용하여 $n=10, 30, 50$일 때, $\frac{X}{n}$와 $\frac{1}{6}$의 차가 0.1보다 작을 확률 $P\left(\left|\frac{X}{n}-\frac{1}{6}\right|<0.1\right)$을 각각 구하여 보자.

(i) $n=10$일 때, $P\left(\left|\frac{X}{10}-\frac{1}{6}\right|<0.1\right)=P(X=1)+P(X=2)$

$=0.3230+0.2907=0.6137$

(ii) $n=30$일 때, $P\left(\left|\frac{X}{30}-\frac{1}{6}\right|<0.1\right)=P(X=3)+P(X=4)+\cdots\cdots+P(X=7)$

$=0.1368+0.1847+\cdots\cdots+0.1098=0.7835$

(iii) $n=50$일 때, $P\left(\left|\frac{X}{50}-\frac{1}{6}\right|<0.1\right)=P(X=4)+P(X=5)+\cdots\cdots+P(X=13)$

$=0.0405+0.0745+\cdots\cdots+0.0319=0.9455$

[정규분포 칼럼] 미천한 수학자

교학사 개정 교과서 발췌 4

- 연속확률분포를 사고하는 방식은 이산확률분포의 확률 값을 직사각형의 면적으로 인지하고 이를 통해 연속확률분포를 끌어내는 방식이다. 그리고 그 과정에서 <상대도수>라는 개념을 이용한다.
  (과거에도 개정 전 교과서에서도 같은 방식으로 서술하였다.)

생각열기

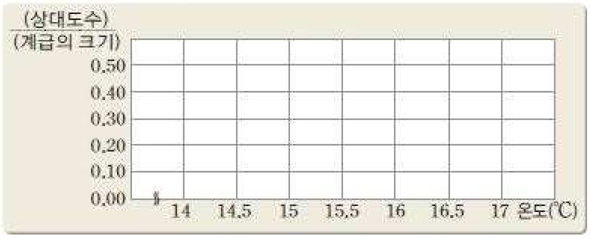
어느 날 A 도시의 하루 평균 온도를 알아보기 위하여 A 도시의 80개의 지역에서 하루 평균 온도를 측정하였다. 그 결과가 아래의 표와 같을 때, 다음 물음에 답하여 보자.

온도(℃)	도수(곳)	상대도수	$\frac{(상대도수)}{(계급의\ 크기)}$
14.0℃이상 ~ 14.5℃미만	8	0.1	
14.5℃ ~ 15.0℃	14		
15.0℃ ~ 15.5℃	16		
15.5℃ ~ 16.0℃	20	0.25	
16.0℃ ~ 16.5℃	16	0.2	
16.5℃ ~ 17.0℃	6		
합계	80	1	

교과서는 상대도수를 통해서 확률을 설명합니다.

1 각 계급의 상대도수를 구하여 위의 표를 완성하여 보자.

2 각 계급의 상대도수를 계급의 크기로 나눈 값을 히스토그램으로 나타내어 보자.



연속확률변수

σ 히스토그램에서 (직사각형의 넓이) $= \frac{(상대도수)}{(계급의\ 크기)} \times (계급의크기) = (상대도수)$

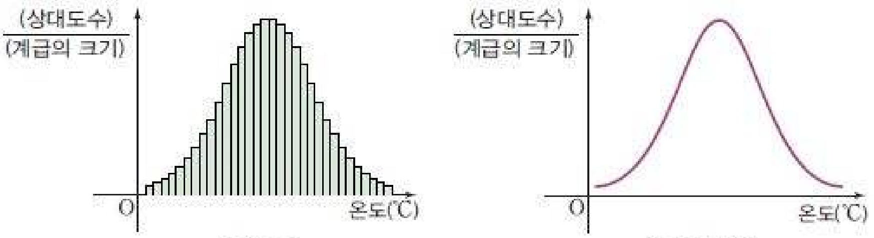
생각 열기에서 확률변수 X의 $\frac{(상대도수)}{(계급의\ 크기)}$를 히스토그램으로 나타내면 히스토그램에서 각 구간에 세워진 직사각형의 넓이는 각 구간의 상대도수를 나타내고, 상대도수의 합이 1이기 때문에 직사각형들의 넓이의 합은 항상 1이다.

이때 총 도수를 늘리고 계급의 크기를 더욱 작게 하여 히스토그램을 그리면 [그림 1]을 얻을 수 있다.

[정규분포 칼럼] 미천한 수학자

교학사 개정 교과서 발췌 5

또 조사 대상의 수를 한없이 늘리고, 계급의 크기를 0에 한없이 가깝게 하면 히스토그램은 [그림 2]와 같은 매끄러운 곡선이 된다.



[그림 1] [그림 2]

σ 시간, 길이, 무게, 온도 등과 같이 어떤 구간에 속하는 모든 실숫값을 가지는 확률변수 X를 연속확률변수라고 한다.

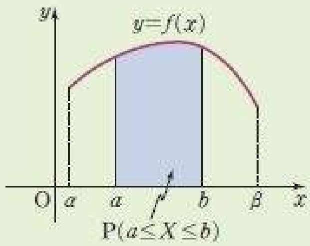
이와 같이 확률변수 X가 어떤 범위의 모든 실숫값을 가질 때, X를 **연속확률변수**라고 한다. 또 $\alpha \leq X \leq \beta$의 모든 실숫값을 가지는 연속확률변수 X에 대하여 함수 $f(x)$가 다음과 같은 성질을 가질 때, 함수 $f(x)$를 연속확률변수 X의 **확률밀도함수**라 하고, X는 확률밀도함수가 $f(x)$인 확률분포를 따른다고 한다.

위의 그림과 같이 결국
(상대도수/계급의 크기)X계급의 크기
를 통해서 <확률값 = 상대도수>를 직사각형으로 인지하고
이것이 정규분포로 근사됨을 설명한다.
이항분포는 여기에서 <계급의 크기를 1>이라고 볼 수 있다.

α,

및

와



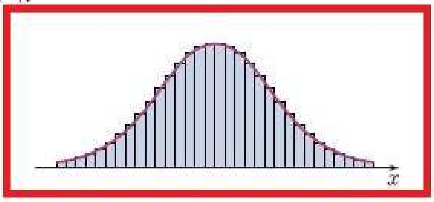
[정규분포 칼럼] 미천한 수학자

교학사 개정 교과서 발췌 6

정규분포

가우스
(Gauss, K. F. ; 1777~1855)
정규분포를 발견하였다.

생각 열기에서와 같이 고등학생 전체의 키, 몸무게, 지능 지수, 어떤 현상 또는 대상에 대한 여러 번의 측정값 등 자연 현상이나 사회 현상에서 얻은 자료를 히스토그램으로 그리면 자료의 개수가 커짐에 따라 다음 그림과 같이 좌우 대칭인 종 모양의 곡선에 가까워지는 경우가 많다.



확률변수 X가 모든 실숫값을 가지고, 두 상수 m, $\sigma(\sigma>0)$에 대하여 그 확률밀도 함수 $f(x)$가

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}\quad(-\infty<x<\infty)$$

일 때, X의 확률분포를 **정규분포**라고 한다. 이때 e는 2.718281……인 무리수이고, 연속확률변수 X의 평균과 표준편차는 각각 m, σ이다.

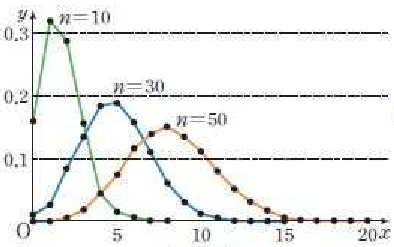
이산확률분포에서 확률은 직사각형의 넓이로 표현되고 이 값은 정규분포곡선의 면적으로 근사된다.

교학사 개정 교과서 발췌 7

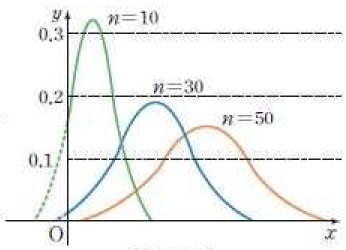
이항분포와 정규분포의 관계

한 개의 주사위를 n번 던질 때, 1의 눈이 나오는 횟수를 확률변수 X라고 하면 X는 이항분포 $\mathrm{B}(n, p)$를 따른다.

여기서 $n=10$, 30, 50일 때 이항분포 $\mathrm{B}\left(n, \frac{1}{6}\right)$의 그래프는 [그림 1]과 같고, n이 커지면 [그림 2]와 같이 좌우 대칭인 정규분포곡선에 가까워진다.



[그림 1]



[그림 2]

교학사 개정 교과서 발췌 8

- 모든 교과서에 공통적으로 수록되어있는 충분히 크다는 말의 의미.
  이 말의 의미를 설명할 수 있다면 위의 문제도 자연스럽게 이해할 수 있다고 생각한다.

일반적으로 n이 충분히 크다는 것은 $np\geq5$, $n(1-p)\geq5$일 때를 뜻한다.

이항분포와 정규분포의 관계

확률변수 X가 이항분포 $\mathrm{B}(n, p)$를 따를 때, n이 충분히 크면 X는 근사적으로 정규분포 $\mathrm{N}(np, npq)$를 따른다. (단, $q=1-p$)

[정규분포 칼럼] 미천한 수학자

교학사 개정 교과서 발췌 9

- 아래와 같이 프로그램을 이용하여 이항분포의 실제 확률 값을 계산하도록 유도하고 있다.

만약 이항분포 $\mathrm{B}\left(100,\ \frac{1}{10}\right)$을 <정규분포 $\mathrm{N}\left(10,\ 3^2\right)$라고만 생각>한다면 $\mathrm{P}\left(\mathrm{X}=10\right)$일 때의 확률 값은 0이다. 0은 아니더라도 대략 0에 가깝다고 생각할 수도 있다. 하지만 실제로 계산해보면

$$\mathrm{P}\left(\mathrm{X}=10\right) = {}_{100}C_{10}\left(\frac{1}{10}\right)^{10}\left(\frac{9}{10}\right)^{90} = 0.1293$$

으로 굉장히 유의미한 값이 나온다.

컴퓨터 프로그램을 이용하면 이항분포, 정규분포 등의 다양한 확률분포의 확률을 구할 수 있다. 확률변수 X가 이항분포 $\mathrm{B}(10,\ 0.2)$를 따를 때, 확률 $\mathrm{P}(X=4)$를 구하여 보자.

❶ [그림 1]과 같이 '수식' ⟶ '함수 추가' ⟶ '통계(S)' ⟶ 'BINOM.DIST'를 클릭한다.

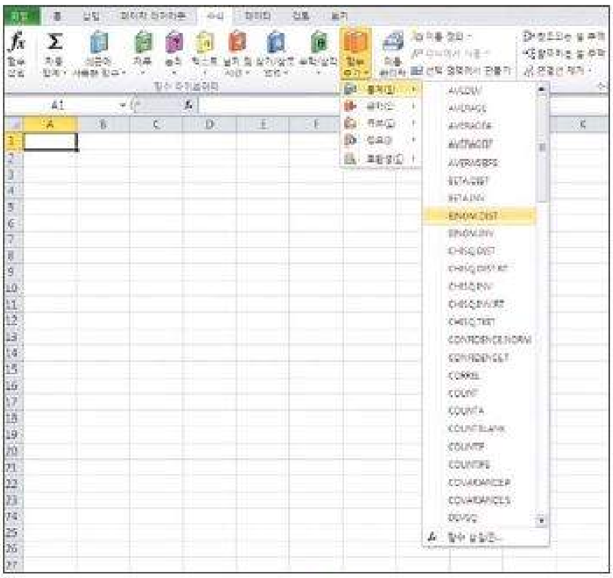
❷ [그림 2]와 같이 Numbers_s에 구하는 확률 $\mathrm{P}(X=k)$에서 k의 값 4를 입력한다.

❸ Trials에 $\mathrm{B}(n,\ p)$에서 n의 값 10을 입력한다.

❹ Probability_s에 $\mathrm{B}(n,\ p)$에서 p의 값 0.2를 입력한다.

❺ Cumulative에 false를 입력하면 0.088080384가 나온다.

따라서 확률 $\mathrm{P}(X=4)$의 값은 0.088080384임을 알 수 있다.



[그림 1]

[그림 2]

과제 1 확률변수 X가 이항분포 $\mathrm{B}(40,\ 0.4)$를 따를 때, 컴퓨터 프로그램을 이용하여 확률 $\mathrm{P}(X=18)$의 값을 구하여라.

즉, 앞서 출제한 문제는 얼마든지 후에 수능에서 나올 수 있다고 생각한다.

수능은 기존의 경향성을 80%정도 유지하면서 <교과서 내에서> 새로운 느낌의 문제를 20%정도 출제하기 때문이다.
교과서를 근거로 명확하게 설명할 수 있고, 학생들이나 선생님들도 대부분 착각하는 문제라면
후에 얼마든지 나올 수 있다고 본다.