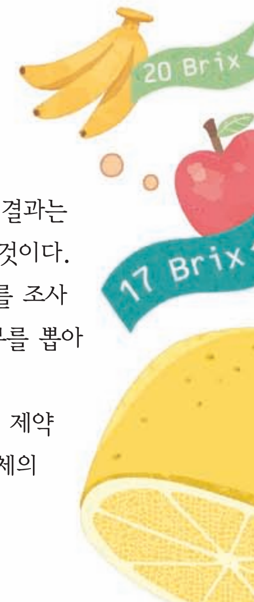


대통령 선거에서 개표가 시작되기 전에 발표되는 당선 예측 결과는 일부 유권자를 대상으로 한 출구 조사를 통하여 전체 투표 결과를 추측한 것이다.

또 신약의 안전성과 효과를 파악할 때, 농산물의 크기나 과일의 당도를 조사할 때, 제품의 품질이나 수명을 조사할 때에도 대상이 되는 자료의 일부를 뽑아 조사한다.

이와 같이 통계 조사에서 대상 전체를 조사하는 것은 시간이나 비용의 제약 때문에 현실적으로 어려운 경우가 많아 주로 일부를 조사하여 대상 전체의 특성을 추측한다.

이 단원에서는 모집단과 표본, 모평균의 추정을 알아본다.

2 통계적 추정

준비학습

1 확률변수 X의 확률분포가 오른쪽 표와 같을 때, X의 평균과 표준편차를 구하시오.

X	0	1	2	3	합계
$\mathrm{P}(X=x)$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	1

2 확률변수 X가 정규분포 $\mathrm{N}(100,\ 10^2)$을 따를 때, 다음을 구하시오.

(1) $\mathrm{P}(X\leq 110)$

(2) $\mathrm{P}(|X-100|\leq 19.6)$

1 모집단과 표본

- 모집단과 표본의 뜻을 알고 표본추출의 원리를 이해한다.
- 표본평균과 모평균의 관계를 이해하고 설명할 수 있다.

모집단과 표본이란 무엇일까

생각 톡

자동차 안전도 평가 제도는 다양한 성능 검사를 통하여 자동차의 안전도를 평가하고 그 결과를 소비자에게 공개하는 제도이다. 국토교통부에서 어느 신형 자동차의 안전도 평가를 위해 정면충돌 안전성 검사를 하려고 한다.

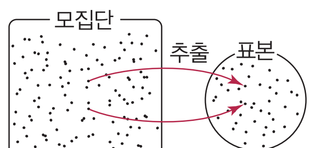
탐구 ① 검사의 대상을 말해 보자.

탐구 ② 검사의 대상을 모두 검사할 때 생기는 문제점을 말해 보자.

탐구 ③ 탐구 ②의 문제점을 해결하기 위한 방법을 추측해 보자.

통계 조사에서 조사하고자 하는 대상 전체를 **모집단**이라 하고, 모집단 전체를 조사하는 것을 **전수조사**라고 한다.

한편 조사 대상 중에서 일부를 뽑아 조사하기도 하는데, 이와 같이 모집단에서 뽑은 일부분을 **표본**이라 하고, 표본을 뽑아 조사하는 것을 **표본조사**라고 한다. 또 표본조사에서 뽑은 표본의 개수를 표본의 크기라고 한다.



모집단에서 표본을 뽑는 것을 추출이라고 한다.

전수조사는 모집단에 대한 정확한 정보를 얻을 수 있지만 비용이 많이 들고 시간이 오래 걸린다. 또 전구의 수명과 같이 전수조사가 불가능한 경우도 있다.

따라서 많은 통계 조사에서는 표본조사를 하고, 그 결과를 이용해 모집단의 특성을 추측한다.

문제 1 **다음은 전수조사와 표본조사 중에서 어느 것이 적합한지 말하시오.**

(1) A 고등학교 학생들의 흉부 엑스레이 검사
(2) B강의 수질 오염도 조사
(3) C 아파트 단지 각 가구의 8월 한 달 전력 소비량 조사
(4) D 회사에서 생산한 휴대 전화 액정의 강도 검사

표본조사의 목적은 모집단의 특성을 추측하는 데 있다. 따라서 표본은 모집단의 특성을 잘 나타낼 수 있도록 편중되지 않게 뽑아야 한다.

난수 주사위는 정이십면체의 각 면에 0부터 9까지의 숫자를 각각 두 번씩 적은 것이다.

표본을 추출하는 여러 가지 방법 중에서 모집단에 속하는 각 대상이 같은 확률로 추출되도록 하는 방법을 **임의추출**이라고 한다.

표본을 임의추출할 때에는 난수 주사위, 난수표, 컴퓨터의 난수 프로그램 등을 사용한다.

특별한 말이 없으면 임의추출은 복원추출로 간주한다.

한편 표본을 추출할 때 한 개의 자료를 뽑은 후 되돌려 놓고 다시 뽑는 것을 복원추출이라 하고, 되돌려 놓지 않고 뽑는 것을 비복원추출이라고 한다.

이때 모집단의 크기가 표본의 크기에 비해 매우 큰 경우에는 비복원추출도 복원추출로 볼 수 있다.

문제 2 **1부터 10까지의 자연수가 각각 하나씩 적힌 10개의 공이 들어 있는 주머니에서 3개의 공을 한 개씩 뽑아 표본으로 할 때, 다음을 구하시오.**

(1) 복원추출하는 경우의 수

(2) 비복원추출하는 경우의 수

분석하기

A 도시에서 시내버스 이용 만족도를 조사하기 위하여 다음과 같이 설문 조사를 실시하였다.

일시: 평일 10시부터 17시까지
방법: 자동 응답 조사
대상: A 도시 지역 번호를 사용하는 유선 전화 응답자

1 위와 같이 조사한 결과가 시내버스 이용 만족도를 잘 나타낸다고 할 수 있는지 말하고, 그 이유를 설명해 보자.

2 시내버스 이용 만족도를 잘 나타내도록 조사하기 위한 방법을 말해 보자.

모평균과 표본평균이란 무엇일까

생각톡

한국환경공단에서는 각 지역에 설치된 측정소에서 대기 오염 농도를 측정하여 웹 사이트를 통해 실시간으로 공개하고 있다.
오른쪽은 2016년 4월 한 달 동안 서울 중구의 일평균 미세 먼지 농도를 나타낸 것이다.

(단위: $\mu g/m^3$)

4월

일	월	화	수	목	금	토
					1 60	2 67
3 40	4 26	5 52	6 66	7 34	8 70	9 109
10 125	11 49	12 59	13 57	14 59	15 55	16 54
17 48	18 67	19 46	20 65	21 26	22 81	23 210
24 153	25 78	26 81	27 54	28 60	29 64	30 76

(출처: 에어코리아, 2016)

탐구 ① 2016년 4월 한 달 동안 서울 중구의 일평균 미세 먼지 농도의 평균을 구해 보자.

탐구 ② 크기가 5인 표본을 임의추출하여 평균을 구하고, **탐구 ①**의 평균과 비교해 보자.

탐구 ③ **탐구 ②**에서 짝이 구한 평균과 자신이 구한 평균을 비교해 보자.

모집단에서 조사하고자 하는 특성을 나타내는 확률변수를 X라고 할 때, X의 평균, 분산, 표준편차를 각각 **모평균**, **모분산**, **모표준편차**라 하고, 이것을 기호로 각각

$$m,\ \sigma^2,\ \sigma$$

와 같이 나타낸다.

한편 모집단에서 임의추출한 크기가 n인 표본을 $X_1,\ X_2,\ \cdots,\ X_n$이라고 할 때, **표본평균**, **표본분산**, **표본표준편차**를 각각 기호로

$$\overline{X},\ S^2,\ S$$

와 같이 나타내고, 다음과 같이 정의한다.

$$\overline{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n)$$

$$S^2=\frac{1}{n-1}\{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\cdots+(X_n-\overline{X})^2\}$$

$$S=\sqrt{S^2}$$

표본평균을 정의할 때와 달리 표본분산의 정의에서 $n-1$로 나누는 것은 표본분산과 모분산의 차이를 줄이기 위함이다.

이때 모평균 m은 고정된 상수이지만 표본평균 $\overline{X}$는 추출된 표본에 따라 여러 가지 값을 가질 수 있으므로 확률변수이다.

표본평균 $\overline{X}$의 확률분포와 평균, 분산 및 표준편차를 구해 보자.

예를 들어 1부터 3까지의 자연수가 각각 하나씩 적힌 3개의 공이 들어 있는 상자에서 한 개의 공을 임의추출할 때, 공에 적힌 수를 확률변수 X라고 하자.

이때 X의 확률분포, 즉 모집단의 확률분포를 표로 나타내면 오른쪽과 같고, 모평균, 모분산, 모표준편차는 다음과 같다.

X	1	2	3	합계
$\mathrm{P}(X=x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

$$m=2,\ \sigma^2=\frac{2}{3},\ \sigma=\frac{\sqrt{6}}{3}$$

한편 이 모집단에서 크기가 2인 표본을 임의추출하여 공에 적힌 수를 각각 X_1, X_2라고 할 때, 표본평균 $\overline{X}=\frac{X_1+X_2}{2}$는 다음과 같다.

(X_1, X_2)	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)
$\overline{X}$	1	1.5	2	1.5	2	2.5	2	2.5	3

이때 각 경우의 확률이 $\frac{1}{9}$이므로 표본평균 $\overline{X}$의 확률분포를 표로 나타내고 표본평균 $\overline{X}$의 평균, 분산, 표준편차를 구하면 다음과 같다.

$\overline{X}$	1	1.5	2	2.5	3	합계
$\mathrm{P}(\overline{X}=\overline{x})$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$	1

$$\mathrm{E}(\overline{X})=2,\ \mathrm{V}(\overline{X})=\frac{1}{3},\ \sigma(\overline{X})=\frac{\sqrt{3}}{3}$$

따라서 표본의 크기가 $n=2$일 때, 표본평균 $\overline{X}$의 평균, 분산, 표준편차를 각각 모평균, 모분산, 모표준편차와 비교하면 다음이 성립함을 알 수 있다.

$$\mathrm{E}(\overline{X})=m,\ \mathrm{V}(\overline{X})=\frac{\sigma^2}{n},\ \sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}$$

일반적으로 다음이 성립한다.

표본평균의 평균, 분산, 표준편차

모평균이 m이고 모표준편차가 σ인 모집단에서 임의추출한 크기가 n인 표본의 표본평균 $\overline{X}$에 대하여

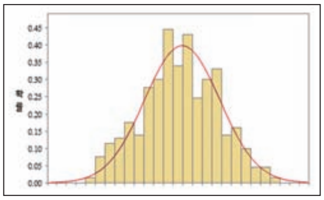
$$\mathrm{E}(\overline{X})=m,\quad \mathrm{V}(\overline{X})=\frac{\sigma^2}{n},\quad \sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}$$

문제 3 **모평균이 50이고 모표준편차가 4인 모집단에서 임의추출한 크기가 16인 표본의 표본평균 $\overline{X}$에 대하여 $\overline{X}$의 평균과 표준편차를 구하시오.**

표본평균의 분포는 어떻게 될까

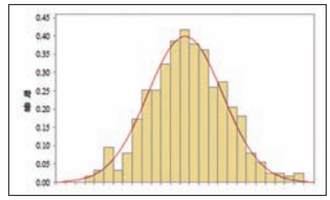
표본평균의 분포를 알아보자.

다음은 확률변수가 각 값을 가질 확률이 모두 같은 분포를 따르는 모집단에서 크기가 각각 10, 25, 100인 표본을 임의추출하였을 때, 표본평균 $\overline{X}$의 확률분포를 컴퓨터 프로그램을 이용하여 그래프로 나타낸 것이다.



크기가 10인 경우

크기가 25인 경우



크기가 100인 경우

위의 그래프에서 표본의 크기가 커지면 표본평균 $\overline{X}$의 분포는 정규분포에 가까워짐을 알 수 있다.

일반적으로 표본평균의 분포는 다음과 같다.

표본평균의 분포

모평균이 m, 모표준편차가 σ인 모집단에서 임의추출한 크기가 n인 표본의 표본평균 $\overline{X}$에 대하여 다음이 성립한다.

❶ 모집단이 정규분포를 따르면 n의 크기에 관계없이 표본평균 $\overline{X}$는 정규분포 $\mathrm{N}\left(m,\ \dfrac{\sigma^2}{n}\right)$을 따른다.

❷ 모집단이 정규분포를 따르지 않아도 n이 충분히 크면 표본평균 $\overline{X}$는 근사적으로 정규분포 $\mathrm{N}\left(m,\ \dfrac{\sigma^2}{n}\right)$을 따른다.

보통 $n \geq 30$이면 충분히 큰 표본으로 간주한다.

보기 모평균이 70이고 모표준편차가 12인 정규분포를 따르는 모집단에서 크기가 36인 표본을 임의추출할 때, 표본평균 $\overline{X}$의 평균과 표준편차는 각각

$$\mathrm{E}(\overline{X})=70,\ \sigma(\overline{X})=\frac{12}{\sqrt{36}}=2$$

이므로 표본평균 $\overline{X}$는 정규분포 $\mathrm{N}(70,\ 2^2)$을 따른다.

문제 4 모평균이 32이고 모표준편차가 5인 모집단에서 크기가 100인 표본을 임의추출할 때, 표본평균 $\overline{X}$는 근사적으로 어떤 분포를 따르는지 기호로 나타내시오.

예제 1 어느 병원에서 태어난 신생아의 체중은 평균이 3.2 kg, 표준편차가 0.6 kg인 정규분포를 따른다고 한다. 이 병원에서 태어난 신생아 중에서 9명을 임의추출하여 체중을 조사할 때, 체중의 평균이 2.6 kg 이상 3.8 kg 이하일 확률을 구하시오.

풀이 신생아의 체중이 정규분포 $\mathrm{N}(3.2,\ 0.6^2)$을 따르므로 9명의 평균 체중 $\overline{X}$ kg은 정규분포 $\mathrm{N}\left(3.2,\ \frac{0.6^2}{9}\right)$, 즉 $\mathrm{N}(3.2,\ 0.2^2)$을 따른다.

따라서 확률변수 $Z=\frac{\overline{X}-3.2}{0.2}$는 표준정규분포 $\mathrm{N}(0,\ 1)$을 따르므로 구하는 확률은

$$\begin{aligned}\mathrm{P}(2.6\le\overline{X}\le3.8)&=\mathrm{P}\left(\frac{2.6-3.2}{0.2}\le Z\le\frac{3.8-3.2}{0.2}\right)\\&=\mathrm{P}(-3\le Z\le3)\\&=2\mathrm{P}(0\le Z\le3)\\&=2\times0.4987=0.9974\end{aligned}$$

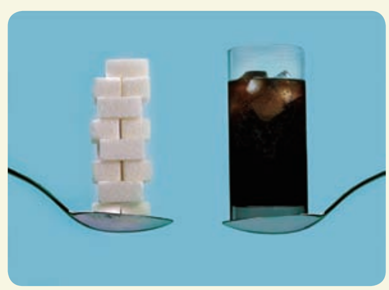
답 0.9974

문제 5 어느 도시에 거주하는 직장인의 6월 저축액은 평균이 67만 원, 표준편차가 20만 원인 정규분포를 따른다고 한다. 이 도시의 직장인 중에서 100명을 임의추출하여 조사한 6월 저축액의 평균이 65만 원 이하일 확률을 구하시오.

비교하기

세계보건기구는 2014년 성인 기준 당분의 하루 권장 섭취량을 25 g으로 지정하였다.

어느 회사에서 생산한 A 탄산음료 한 캔에 포함된 당분의 양은 평균이 24 g, 표준편차가 5 g인 정규분포를 따른다고 한다. 서윤, 시우, 도현이가 A 탄산음료를 각각 25캔, 100캔, 225캔을 임의추출하여 당분의 양을 조사할 때, 당분의 양의 평균이 25 g 이상일 확률을 각각 구하고 그 값을 표본의 크기와 관련하여 비교해 보자.



공학 도구 활용

임의추출하기

정보 처리 창의·융합

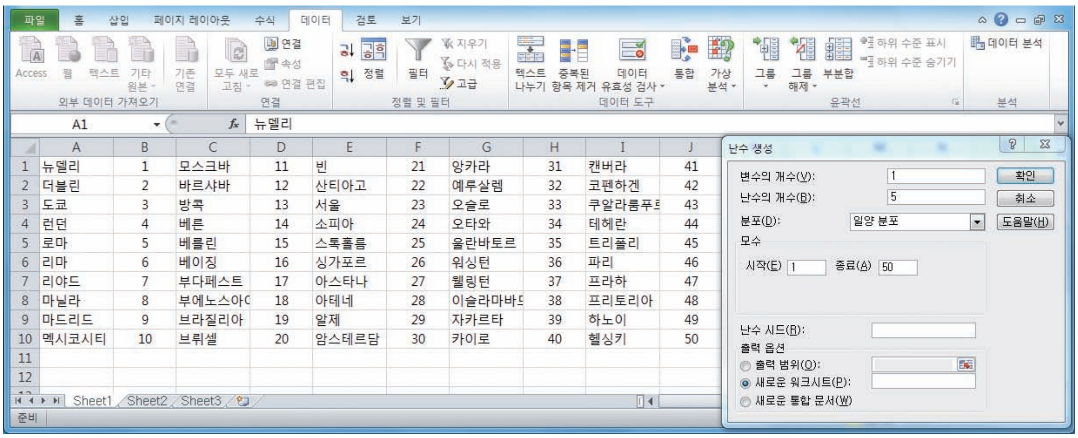
스프레드시트를 이용하여 임의추출하는 방법을 알아보자.

크기가 50인 모집단에서 다음과 같은 방법으로 크기가 5인 표본을 임의추출할 수 있다.

❶ 모집단에 속한 50개의 대상에 1부터 50까지의 번호를 붙인다.

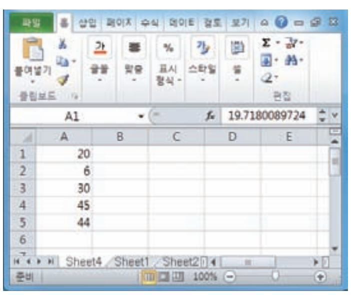
❷ **데이터−데이터 분석−난수 생성**을 선택한다.

❸ **난수 생성** 창의 변수의 개수에는 '1', 난수의 개수에는 '5'를 입력하고 분포에서 '일양 분포'를 선택한다. 또 모수의 시작과 종료에는 각각 '1', '50'을 입력한 후 확인 을 누른다.



	A	B	C	D	E	F	G	H	I	J
1	뉴델리	1	모스크바	11	빈	21	앙카라	31	캔버라	41
2	더블린	2	바르샤바	12	산티아고	22	예루살렘	32	코펜하겐	42
3	도쿄	3	방콕	13	서울	23	오슬로	33	쿠알라룸푸르	43
4	런던	4	베른	14	소피아	24	오타와	34	테헤란	44
5	로마	5	베를린	15	스톡홀름	25	울란바토르	35	트리폴리	45
6	리마	6	베이징	16	싱가포르	26	워싱턴	36	파리	46
7	리야드	7	부다페스트	17	아스타나	27	웰링턴	37	프라하	47
8	마닐라	8	부에노스아이	18	아테네	28	이슬라마바드	38	프리토리아	48
9	마드리드	9	브라질리아	19	알제	29	자카르타	39	하노이	49
10	멕시코시티	10	브뤼셀	20	암스테르담	30	카이로	40	헬싱키	50

❹ 새로운 시트에 나타난 5개의 수를 **홈 −표시 형식**에서 **자릿수 줄임**을 눌러 자연수로 만든 후 그 수에 해당하는 대상을 추출한다.



활동 우리 학급의 학생들을 모집단으로 할 때, 스프레드시트를 이용하여 크기가 4인 표본을 임의추출해 보자.

2 모평균의 추정

● 모평균을 추정하고, 그 결과를 해석할 수 있다.

모평균은 어떻게 추정할까

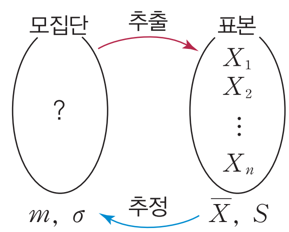
생각 톡

어느 밭에서 수확한 옥수수 1개의 낟알의 개수의 평균을 조사하기 위해 옥수수 50개를 임의추출하여 조사하였더니 평균이 400, 표준편차가 50이었다.

탐구 ① 이 밭에서 수확한 옥수수 1개의 낟알의 개수의 모평균을 추측해 보자.

탐구 ② 탐구 ①에서 추측한 값이 모평균과 같다고 할 수 있는지 말해 보자.

표본을 조사해 얻은 정보를 이용하여 모평균, 모표준편차와 같이 모집단의 특성을 나타내는 값을 추측하는 것을 **추정**이라고 한다.



표본평균을 이용하여 모평균을 추정하는 방법을 알아보자.

정규분포 $\mathrm{N}(m, \sigma^2)$을 따르는 모집단에서 크기가 n인 표본을 임의추출할 때, 표본평균 $\overline{X}$는 정규분포 $\mathrm{N}\left(m, \dfrac{\sigma^2}{n}\right)$을 따르므로 확률변수

$$Z=\frac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$$

은 표준정규분포 $\mathrm{N}(0, 1)$을 따른다.

한편 표준정규분포표로부터

$$\mathrm{P}(-1.96 \le Z \le 1.96)=0.95$$

이므로

$$\mathrm{P}\left(-1.96 \le \frac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}} \le 1.96\right)=0.95$$

이다.

이를 정리하면 다음과 같다.

$$\mathrm{P}\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}}\le m\le \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)=0.95$$

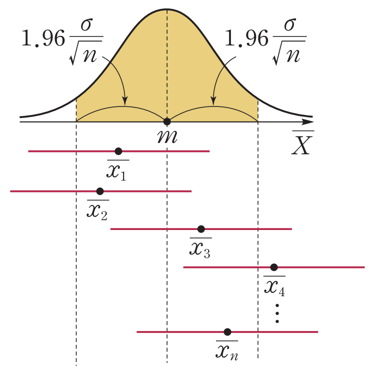
여기서 표본평균 $\overline{X}$의 값이 $\overline{x}$일 때,

$$\overline{x}-1.96\frac{\sigma}{\sqrt{n}}\le m\le \overline{x}+1.96\frac{\sigma}{\sqrt{n}}$$

를 모평균 m에 대한 **신뢰도** 95 $\%$의 **신뢰구간**이라고 한다.

표본평균 $\overline{X}$는 확률변수이므로 추출하는 표본에 따라 $\overline{x}$가 달라지고 신뢰구간도 달라진다. 이 신뢰구간 중에는 오른쪽 그림과 같이 모평균 m을 포함하는 것과 포함하지 않는 것이 있을 수 있다.

이때 모평균 m에 대한 신뢰도 95 $\%$의 신뢰구간이란 크기가 n인 표본을 여러 번 추출하여 신뢰구간을 만들 때, 이들 중에서 95 $\%$ 정도는 모평균 m을 포함할 것으로 기대된다는 뜻이다.



한편 $\mathrm{P}(-2.58\le Z\le 2.58)=0.99$임을 이용하여 모평균 m에 대한 신뢰도 99 $\%$의 신뢰구간을 구하면 다음과 같다.

$$\overline{x}-2.58\frac{\sigma}{\sqrt{n}}\le m\le \overline{x}+2.58\frac{\sigma}{\sqrt{n}}$$

이상을 정리하면 다음과 같다.

모평균에 대한 신뢰구간

정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는 모집단에서 임의추출한 크기가 n인 표본의 표본평균 $\overline{X}$의 값이 $\overline{x}$일 때, 모평균 m에 대한 신뢰구간은

❶ 신뢰도 95 $\%$의 신뢰구간

$$\overline{x}-1.96\frac{\sigma}{\sqrt{n}}\le m\le \overline{x}+1.96\frac{\sigma}{\sqrt{n}}$$

❷ 신뢰도 99 $\%$의 신뢰구간

$$\overline{x}-2.58\frac{\sigma}{\sqrt{n}}\le m\le \overline{x}+2.58\frac{\sigma}{\sqrt{n}}$$

표본의 크기 n이 충분히 크면 표본표준편차 S의 값 s는 모표준편차 σ와 큰 차이가 없음이 알려져 있다. 따라서 정규분포를 따르는 모집단에서 모표준편차 σ를 모르는 경우에 σ 대신 s를 이용하여 모평균에 대한 신뢰구간을 구할 수도 있다.

예제 1

어느 호수에 서식하는 붕어의 길이는 평균이 m cm, 표준편차가 2 cm인 정규분포를 따른다고 한다. 이 호수에 서식하는 붕어 16마리를 임의추출하여 길이를 측정하였더니 평균이 15 cm이었다. 이 호수에 서식하는 붕어의 길이의 모평균 m에 대한 신뢰도 95 %의 신뢰구간을 구하시오.

풀이 $n=16$, $\overline{x}=15$, $\sigma=2$이므로 모평균 m에 대한 신뢰도 95 %의 신뢰구간은

$$15-1.96\times\frac{2}{\sqrt{16}}\le m\le 15+1.96\times\frac{2}{\sqrt{16}}$$

$$14.02\le m\le 15.98$$

답 $14.02\le m\le 15.98$

문제 1

어느 공장에서 생산하는 건전지의 수명은 평균이 m분, 표준편차가 20분인 정규분포를 따른다고 한다. 이 공장에서 생산한 건전지 중에서 100개를 임의추출하여 수명을 조사하였더니 평균이 300분이었을 때, 이 공장에서 생산하는 건전지의 수명의 모평균 m에 대한 신뢰도 99 %의 신뢰구간을 구하시오.

문제 2

어느 음식점에서 판매하는 오징어튀김 1개의 열량은 평균이 m kcal인 정규분포를 따른다고 한다. 이 음식점에서 판매하는 오징어튀김 256개를 임의추출하여 열량을 조사하였더니 평균이 70 kcal, 표준편차가 4 kcal이었다. 이 음식점에서 판매하는 오징어튀김의 열량의 모평균 m에 대한 신뢰도 95 %의 신뢰구간을 구하시오.

판별하기

어느 지역의 고등학생의 점심 식사 시간은 평균이 m분, 표준편차가 5분인 정규분포를 따른다고 한다. 이 지역에서 고등학생 n명을 임의추출하여 점심 식사 시간을 조사한 후 추정한 모평균 m에 대한 신뢰도 95 %의 신뢰구간이 $a\le m\le b$라고 한다. $b-a$의 값을 '신뢰구간의 길이'라고 할 때, 다음 세 사람 중 옳게 말한 사람을 찾아보자.

중단원 마무리

❶ 모집단과 표본

(1) **전수조사와 표본조사**

① 전수조사: 모집단 전체를 조사하는 것

② $\boxed{\qquad}$: 모집단에서 표본을 뽑아 조사하는 것

(2) **표본평균의 분포**

모평균이 m이고 모표준편차가 σ인 모집단에서 임의추출한 크기가 n인 표본의 표본평균 $\overline{X}$에 대하여

① $\mathrm{E}(\overline{X})=\boxed{\quad}$, $\mathrm{V}(\overline{X})=\boxed{\qquad}$, $\sigma(\overline{X})=\boxed{\qquad}$

② 모집단이 정규분포를 따르면 n의 크기에 관계없이 표본평균 $\overline{X}$는 정규분포 $\mathrm{N}\left(m,\ \dfrac{\sigma^2}{n}\right)$을 따른다.

③ 모집단이 정규분포를 따르지 않아도 n이 충분히 크면 표본평균 $\overline{X}$는 근사적으로 정규분포 $\mathrm{N}\left(m,\ \dfrac{\sigma^2}{n}\right)$을 따른다.

❷ 모평균의 추정

정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는 모집단에서 임의추출한 크기가 n인 표본의 표본평균 $\overline{X}$의 값이 $\overline{x}$일 때, 모평균 m에 대한 신뢰구간은 다음과 같다.

① 신뢰도 95 %의 신뢰구간

$\overline{x}-\boxed{\qquad}\le m\le \overline{x}+\boxed{\qquad}$

② 신뢰도 99 %의 신뢰구간

$\overline{x}-\boxed{\qquad}\le m\le \overline{x}+\boxed{\qquad}$

기본 문제

1 다음은 전수조사와 표본조사 중에서 어느 것이 적합한지 말하시오.

(1) A 고등학교 2학년 학생들의 수학여행 참여 여부 조사

(2) B 농장에서 재배한 수박의 당도 조사

2 모평균이 20, 모분산이 2인 모집단에서 임의추출한 크기가 9인 표본의 표본평균을 $\overline{X}$라고 할 때, 다음을 구하시오.

(1) $\mathrm{E}(\overline{X})$

(2) $\mathrm{V}(\overline{X})$

(3) $\sigma(\overline{X})$

3 어느 체육관 회원의 운동 시간은 평균이 60분, 표준편차가 16분인 정규분포를 따른다고 한다. 이 체육관 회원 중에서 16명을 임의추출하여 조사한 운동 시간의 평균을 $\overline{X}$분이라고 할 때, 다음을 구하시오.

(1) $\mathrm{P}(56\le\overline{X}\le66)$

(2) $\mathrm{P}(\overline{X}\le48)$

4 정규분포 $\mathrm{N}(m,\ 30^2)$을 따르는 모집단에서 크기가 n인 표본을 임의추출하여 구한 표본평균이 324일 때, 모평균 m에 대하여 다음을 구하시오.

(1) $n=100$일 때, 신뢰도 95 %의 신뢰구간

(2) $n=25$일 때, 신뢰도 99 %의 신뢰구간

표준 문제

5 평균이 25이고 표준편차가 σ인 모집단에서 크기가 n인 표본을 임의추출할 때, 표본평균을 $\overline{X}$라고 하자. $\mathrm{E}(\overline{X})=\frac{n}{4}$, $\mathrm{V}(\overline{X})=\frac{1}{20}$일 때, σ의 값을 구하시오.

6 확률변수 X의 확률분포를 표로 나타내면 다음과 같다. 이 모집단에서 크기가 4인 표본을 임의추출할 때, 표본평균 $\overline{X}$에 대하여 $\mathrm{V}(4\overline{X})$를 구하시오.

X	0	1	2	3	합계
$\mathrm{P}(X=x)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	1

추론 7 정규분포 $\mathrm{N}(50,\ 8^2)$을 따르는 모집단에서 크기가 16인 표본을 임의추출하여 그 합을 확률변수 X라고 할 때, $\mathrm{P}(768\le X\le 832)$를 구하시오.

서술형 8 정규분포 $\mathrm{N}(m,\ 20^2)$을 따르는 모집단에서 크기가 n인 표본을 임의추출할 때, 표본평균과 모평균의 차가 1 이하일 확률은 0.96이다. 이때 자연수 n의 값을 구하는 풀이 과정과 답을 쓰시오. (단, Z가 표준정규분포를 따르는 확률변수일 때, $\mathrm{P}(0\le Z\le 2)=0.48$로 계산한다.)

9 어느 공장에서 생산하는 축구공의 무게는 평균이 m g, 표준편차가 3 g인 정규분포를 따른다고 한다. 이 공장에서 생산한 축구공 중에서 900개를 임의추출하여 무게를 조사하였더니 평균이 430 g이었다. 이 공장에서 생산하는 축구공의 무게의 모평균 m에 대한 신뢰도 95 %의 신뢰구간을 구하시오.

10 어느 고등학교 학생의 일주일 동안의 자기 주도적 학습 시간은 평균이 m시간, 표준편차가 5시간인 정규분포를 따른다고 한다. 이 학교 학생 중에서 임의로 36명을 뽑아 일주일 동안의 자기 주도적 학습 시간을 조사한 표본평균에 대하여 모평균 m을 신뢰도 99 %로 추정할 때, 모평균과 표본평균의 차의 최댓값을 구하시오.

11 어느 회사에서 생산하는 휴대 전화 배터리를 한 번 충전하여 사용하는 시간은 평균이 m분, 표준편차가 40분인 정규분포를 따른다고 한다. 이 회사에서 생산하는 휴대 전화 배터리 중에서 n개를 임의추출한 후 한 번 충전하여 사용하는 시간을 조사하였더니 평균이 $\overline{x}$분이었다. 이 회사에서 생산하는 휴대 전화 배터리의 모평균 m에 대한 신뢰도 95 %의 신뢰구간이 $590.2 \le m \le 609.8$일 때, $\overline{x}$의 값과 자연수 n의 값을 구하시오.

발전 문제

12 1이 적힌 공 1개와 3이 적힌 공 n개가 들어 있는 주머니에서 임의로 1개의 공을 꺼내 공에 적힌 수를 확인한 후 다시 주머니에 넣는 시행을 2번 반복할 때, 꺼낸 공에 적힌 수의 평균을 $\overline{X}$라고 하자. $\mathrm{P}(\overline{X}=2)=\frac{3}{8}$일 때, 자연수 n의 값을 구하시오.

서술형

13 모집단의 확률변수 X가 정규분포 $\mathrm{N}(100,\ 5^2)$을 따를 때, 이 모집단에서 크기가 4인 표본을 임의추출하여 구한 평균을 $\overline{X}$라고 하자. $\mathrm{P}(X<110)=\mathrm{P}(\overline{X}\ge a)$를 만족시키는 상수 a의 값을 구하는 풀이 과정과 답을 쓰시오.

추론

14 정규분포를 따르는 어느 모집단에서 크기가 50인 표본을 임의추출하여 추정한 모평균 m에 대한 신뢰도 95 %의 신뢰구간은 $a \le m \le b$이고, 신뢰도를 바꾸지 않고 크기가 n인 표본을 임의추출하여 추정한 모평균 m에 대한 신뢰구간은 $c \le m \le d$이다. $d-c=\frac{1}{2}(b-a)$일 때, 자연수 n의 값을 구하시오.

대단원 마무리

• 정답 및 풀이 144쪽

01 다음 중 이산확률변수인 것은?

① A 반 학생들의 충치의 개수
② B 공장에서 생산한 전구의 수명
③ C 정류장에서 버스를 기다리는 시간
④ D 목장에서 하루 동안 생산된 우유의 양
⑤ E 지역 주민들이 소유한 자동차의 주행 거리

02 1, 2, 3, 4, 5가 각각 하나씩 적힌 5장의 카드 중에서 임의로 2장의 카드를 동시에 뽑을 때, 카드에 적힌 두 수의 차를 확률변수 X라고 하자. 이때 $\mathrm{P}(X \geq 2)$를 구하시오.

03 이산확률변수 X의 확률질량함수가

$$\mathrm{P}(X=x)=k\left(\frac{1}{x}-\frac{1}{x+1}\right) \; (x=1,\ 2,\ \cdots,\ 10)$$

일 때, 상수 k의 값을 구하시오.

04 한 개의 주사위를 던져서 나온 눈의 수가 6의 약수이면 두 개의 동전을 동시에 던지고, 6의 약수가 아니면 세 개의 동전을 동시에 던질 때, 앞면이 나오는 동전의 개수를 확률변수 X라고 하자. 이때 $\mathrm{E}(X)$를 구하시오.

05 확률변수 X의 확률분포를 표로 나타내면 다음과 같을 때, $\mathrm{V}\left(\frac{X}{a}\right)$를 구하시오. (단, a는 상수이다.)

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

06 이산확률변수 X의 확률질량함수가

$$\mathrm{P}(X=x)={}_{20}\mathrm{C}_x\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{20-x}$$
$$(x=0,\ 1,\ 2,\ \cdots,\ 20)$$

일 때, $\mathrm{E}(X)+\mathrm{V}(2X)$의 값을 구하시오.

07 확률변수 X가 이항분포 $\mathrm{B}(n,\ p)$를 따르고 $\mathrm{V}(X)=\frac{15}{16}$, $15\mathrm{P}(X=n)=\mathrm{P}(X=n-1)$일 때, $\mathrm{E}(X)$를 구하시오.

08 정규분포 $\mathrm{N}(7, 2^2)$을 따르는 확률변수 X에 대하여 $\mathrm{P}(a-1 \le X \le a+3)$이 최대가 되도록 하는 상수 a의 값을 구하시오.

09 어느 영화관에서 판매하는 팝콘 한 봉지에 들어 있는 팝콘의 개수는 평균이 68, 표준편차가 6인 정규분포를 따른다고 한다. 이때 팝콘의 개수가 56 이상 74 이하인 봉지는 전체의 몇 %인지 구하시오.

10 한 개의 동전을 던져서 앞면이 나오면 5점을 얻고 뒷면이 나오면 2점을 잃는다고 하자. 0점에서 시작하여 한 개의 동전을 100번 던진 후의 점수가 220점 이상일 확률을 구하시오.

11 모평균이 m, 모표준편차가 σ인 정규분포를 따르는 모집단에서 크기가 n_1인 표본을 임의추출하여 구한 표본평균을 $\overline{X}$, 크기가 n_2인 표본을 임의추출하여 구한 표본평균을 $\overline{Y}$라고 할 때, 옳은 것만을 보기에서 있는 대로 고르시오.

보기

ㄱ. $\mathrm{E}(\overline{X})=\mathrm{E}(\overline{Y})$

ㄴ. $n_1<n_2$이면 $\mathrm{V}(\overline{X})>\mathrm{V}(\overline{Y})$이다.

ㄷ. $n_1=2n_2$이면 $2\sigma(\overline{X})=\sigma(\overline{Y})$이다.

12 정규분포 $\mathrm{N}(24, 6^2)$을 따르는 모집단에서 크기가 4인 표본을 임의추출할 때, 표본평균 $\overline{X}$가 18 이상 21 이하일 확률을 구하시오.

13 어느 농장에서 생산하는 블루베리 1개의 무게는 평균이 1.5 g, 표준편차가 0.2 g인 정규분포를 따른다고 한다. 이 농장에서 생산한 블루베리 100개를 한 상자에 포장하는데, 한 상자 안에 들어 있는 블루베리의 전체 무게가 144 g 미만이면 불량품으로 판정한다고 한다. 이 농장에서 생산한 100만 개의 블루베리를 상자에 포장하였을 때, 불량품으로 판정될 상자의 개수의 평균을 구하시오.

14 향신료로 사용되는 사프란은 사프란 알뿌리를 몇 년 동안 배양하느냐에 따라 추출되는 양이 달라진다고 한다. 어느 농장에서는 5년 배양한 사프란 알뿌리를 재배하여 사프란을 추출하는데 1개의 알뿌리에서 추출되는 사프란의 양은 평균이 m mg인 정규분포를 따른다고 한다. 이 농장에서 재배한 사프란 알뿌리 36개를 임의추출하여 사프란을 추출하였더니 그 양의 평균이 7.5 mg, 표준편차가 0.6 mg이었을 때, 모평균 m에 대한 신뢰도 95 %의 신뢰구간을 구하시오.

서술형

• 정답 및 풀이 145쪽

15 열쇠 5개와 이 중 하나에 맞는 잠긴 자물쇠 1개가 있다. 열쇠 5개 중에서 임의로 1개를 택하여 잠긴 자물쇠를 열 때, 자물쇠를 열 때까지 확인한 열쇠의 개수를 X라고 하자. 확률변수 X의 평균을 구하는 풀이 과정과 답을 쓰시오.

풀이

16 어느 회사에서 서류 전형에 합격한 입사 지원자 2000명을 대상으로 필기시험을 치러 120명을 뽑았다. 필기시험 점수는 평균이 220점, 표준편차가 40점인 정규분포를 따른다고 할 때, 필기시험 합격자의 최저 점수를 구하는 풀이 과정과 답을 쓰시오. (단, Z가 표준정규분포를 따르는 확률변수일 때, $\mathrm{P}(0\le Z\le 1.55)=0.44$로 계산한다.)

풀이

17 1이 적힌 카드 5장, 2가 적힌 카드 4장, 3이 적힌 카드 3장, 4가 적힌 카드 2장, 5가 적힌 카드 1장이 들어 있는 상자가 있다. 이 상자에서 2장의 카드를 임의추출할 때, 카드에 적힌 수의 평균을 $\overline{X}$라고 하자. 표본평균 $\overline{X}$의 표준편차를 구하는 풀이 과정과 답을 쓰시오.

풀이

18 평균이 m, 표준편차가 σ인 정규분포를 따르는 모집단에서 크기가 n인 표본을 임의추출하여 추정한 모평균 m에 대한 신뢰도 99 %의 신뢰구간은 $a\le m\le b$이다. $b-a\le 0.86\sigma$를 만족시키는 자연수 n의 최솟값을 구하는 풀이 과정과 답을 쓰시오.

풀이

자기 평가

	만족	보통	미흡
❶ 확률변수와 확률분포의 뜻을 알고, 이산확률변수의 평균과 표준편차를 구할 수 있다.			
❷ 이항분포, 정규분포의 뜻을 알고, 그 성질을 이해하고 있다.			
❸ 모집단과 표본의 뜻을 알고 표본추출의 원리를 이해하고 있다.			
❹ 표본평균과 모평균의 관계를 이해하고, 모평균을 추정해 그 결과를 해석할 수 있다.			

보충 계획

부족한 부분을 어떻게 채울지 계획을 세워 보세요.

창의·융합 프로젝트

실생활에서 통계적 추정하기

다음과 같이 실생활에서 자료를 수집하여 표본평균과 표본표준편차를 구하고, 이를 이용하여 모평균을 추정해 보자.

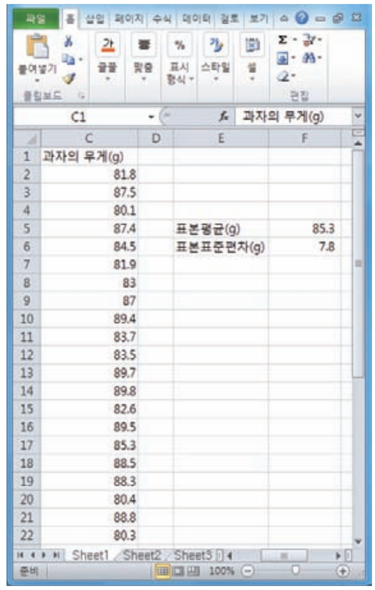
1단계 조사 주제와 조사 대상 선정하기

은영이네 반 학생 36명은 A 과자 한 개의 무게를 조사하기로 정하고, 각자 A 과자를 임의로 한 개씩 구매하였다.

2단계 표본평균과 표본표준편차 구하기

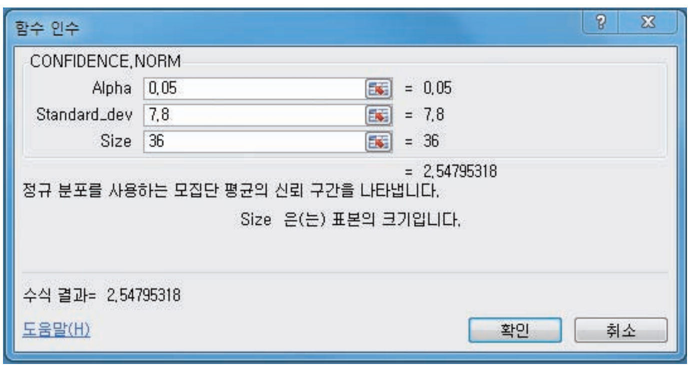
36명이 구매한 A 과자의 무게를 각각 측정하여 스프레드시트에 입력하고, 표본평균과 표본표준편차를 구한다.

➡ 표본평균: 85.3 g, 표본표준편차: 7.8 g



3단계 모평균 추정하기

스프레드시트에서 수식−함수 삽입−CONFIDENCE.NORM을 선택하여 A 과자의 무게의 모평균을 신뢰도 95 %로 추정한다.



➡ 반올림하여 소수점 아래 첫째 자리까지 구한 신뢰구간은

$$85.3-2.5\le m\le 85.3+2.5,\ \text{즉}\ 82.8\le m\le 87.8$$

따라서 은영이네 반 학생 36명이 조사한 A 과자 한 개의 평균 무게를 신뢰도 95 %로 추정하면 82.8 g 이상 87.8 g 이하이다.

과제 모둠별로 주제를 선정하여 모평균을 추정하고, 그 과정을 정리하여 발표해 보자.

빅 데이터가 세상을 바꾼다

컴퓨터와 이동 통신 기기의 이용이 생활화되면서 우리는 수많은 정보와 데이터가 생산되는 ‘빅 데이터’ 시대에 살고 있다. 빅 데이터란 과거 아날로그 환경에서 생성되던 데이터에 비해 규모가 방대하고 생성 주기가 짧으며, 수치 데이터뿐 아니라 문자와 영상 데이터를 포함하는 대규모 데이터를 말한다.

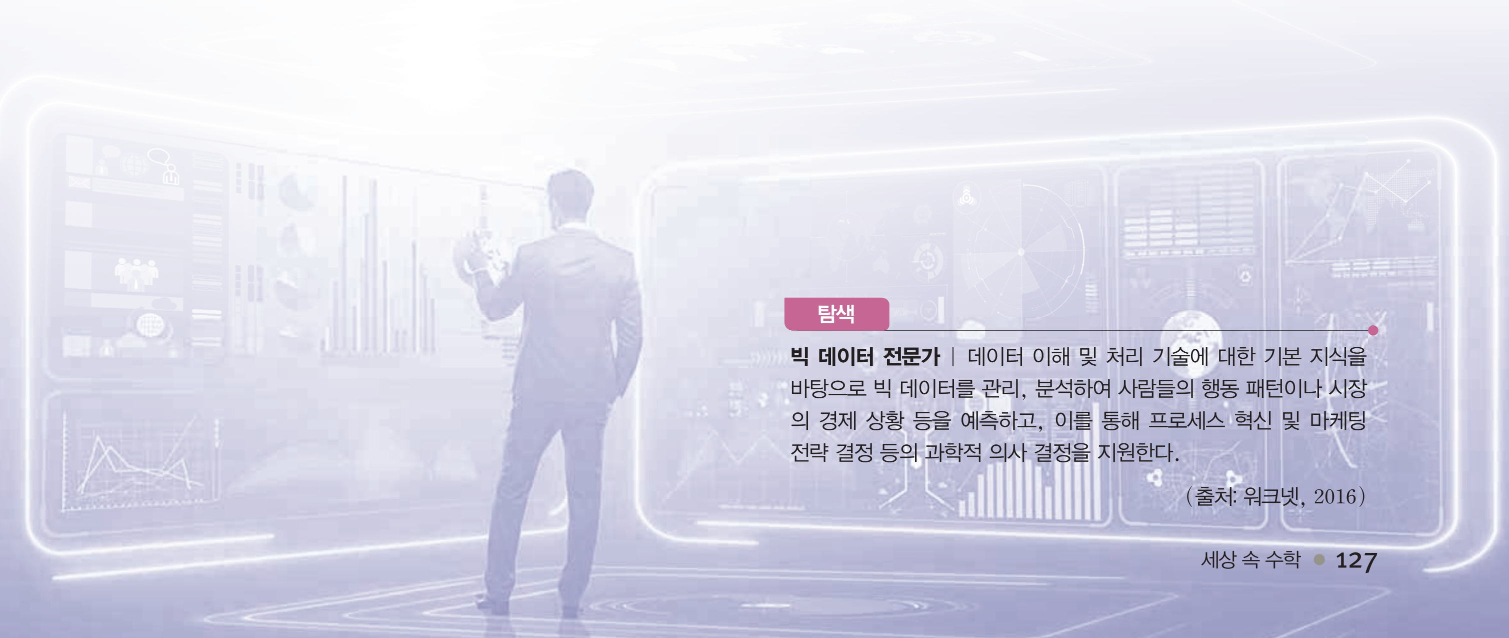
최근에는 빅 데이터에서 필요한 정보를 추출하고 그 정보를 바탕으로 미래를 예측하는 기술이 중요해지고 있다.

기업은 빅 데이터를 활용하여 고객의 행동을 예측하고 대처 방안을 마련하여 판매력을 강화할 수 있다. 예를 들어 고객이 온라인 쇼핑몰에서 책을 주문한 경우, 빅 데이터를 이용해 주문한 책과 유사한 분야의 책이나 비슷한 성향을 가진 고객이 구매한 책을 추천함으로써 또 다른 책의 구매를 유도할 수 있다.

또한 날씨와 관련된 빅 데이터를 분석해 폭우와 같은 기상 특보를 예보하고, 과거 교통 상황 자료를 활용해 교통량을 예측하여 교통 체증을 줄이기 위한 방안을 마련한다. 의료 분야에서는 과거 진료 기록과 약의 효능, 치료 비용 등의 데이터를 분석하여 더욱 효과적인 진료 방법을 선택할 수 있고, 전국의 의료 관련 데이터를 종합하여 전염병 발생과 같은 긴박한 순간에 빠른 의사 결정을 가능하게 한다.

우리에게 유용한 정보를 제공해 주는 빅 데이터를 활용할 수 있는 분야는 앞으로 더욱 확대될 전망이다.

(출처: 빅토르 마이어 쇤버거, 케네스 쿠키어, 『빅 데이터가 만드는 세상』)



탐색

빅 데이터 전문가 | 데이터 이해 및 처리 기술에 대한 기본 지식을 바탕으로 빅 데이터를 관리, 분석하여 사람들의 행동 패턴이나 시장의 경제 상황 등을 예측하고, 이를 통해 프로세스 혁신 및 마케팅 전략 결정 등의 과학적 의사 결정을 지원한다.

(출처: 워크넷, 2016)