

학통과 통계

- 정규분포 알아보기 -

참고서: Head First Statistics

(8장 정규분포 알아보기)

작성자 : 이예식

환경대학교, 컴퓨터공학과



f장. 정수형과 이야기

정수형과? (p. 380 ~ 381)

실생활에서 길이, 크기, 높이 등을 재는 것처럼

연속되어 있는 데이터를 다른 때 일반적으로

기재할 수 있는 값들의 집합.

정수형을 특징자음.

((지금까지 다른 데이터는
이산데이터임.)

* **이산데이터** : 술록머신에서 수학이 정해진 규칙들만

적하거나, 동전던지기에서 앞면이 나온 확수가 1, 2, 3 ...
등을 적하는 것처럼 특정 값들만을 대상으로 하는 데이터를
일컬음.

* **연속데이터** : 예를 들어, X가 한국남성들의

키를 가리킬 때, 키는 175, 176, 178 ... 등의

값 이외에 175.1, 175.12, 176.003 ... 등

이는 같은 사이의 실수를 모두 적할 수 있다.

이런 경우의 데이터를 연속 데이터라 부른다.

*주의: 교재 p.366 ~ p.377 내용은
다루지 않는다. 하지만
아래 내용을 이해한다면
쉽게 이해할 수 있는 내용이다.

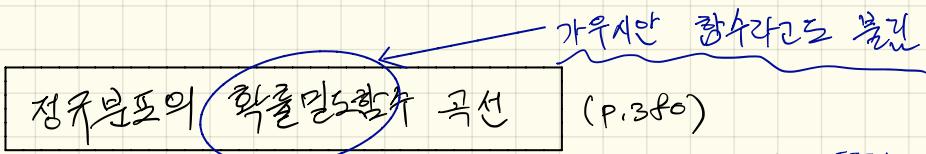
* "일반적으로 기대할 수 있는 값들의 분포"의 의미

한국 사람들의 키의 분포, 또는 미국 사람들의 풀무기의 분포,

환경대 학생들의 성적 평점의 분포, 환경대 컴퓨터공학과

학생들의 통학시간의 분포 등은 모두 정규분포를 따른다.

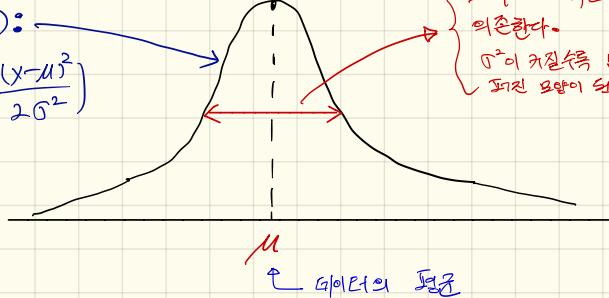
이유는 아래에 대한 이유는 존재하지 않는다. 다만 우리가 사는 우주에서는 정규분포가 일반적으로 성립한다.



가우시안 함수 $G(x)$:

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

↑
의우지 마세요!



위 그래프가 나타내는 정규분포를 $N(\mu, \sigma^2)$ 으로

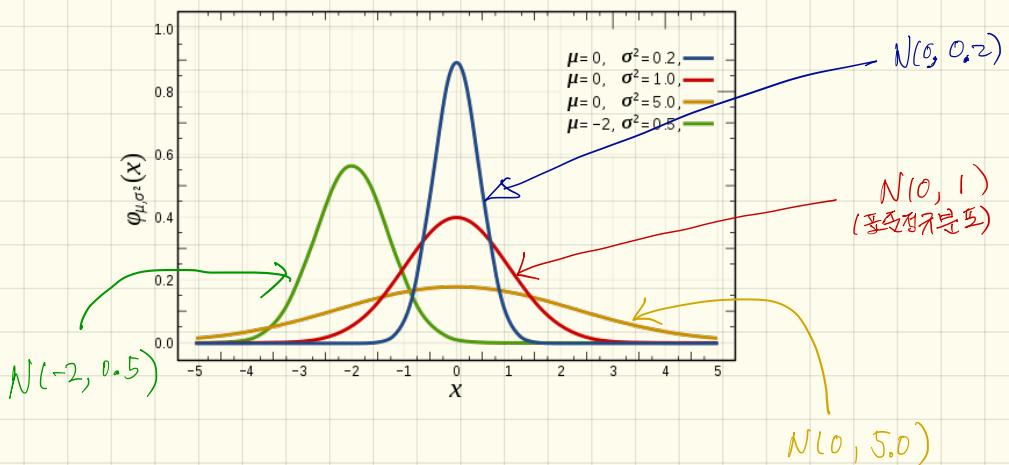
표기한다. 그리고 X 가 정규분포이면서 평균이 μ 이고

분산이 σ^2 일 때 아래와 같이 표기한다.

$$X \sim N(\mu, \sigma^2)$$

정규분포의 확률밀도함수 곡선의 특징 (P.380)

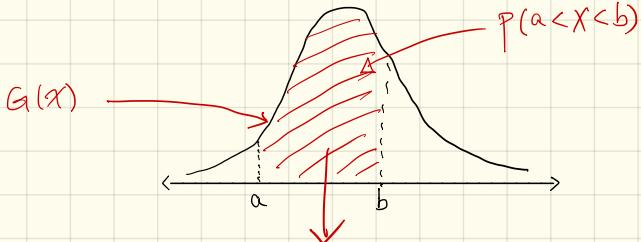
- ① σ^2 이 커지면 곡선은 넓고 평평해진다.
- ② σ^2 이 작아지면 곡선은 좁고 뾰족해진다.
- ③ 곡선은 x 축과 절대로 만나지 않는다.
하지만 평균에서 좌우로 멀어질 수록 점점 x 축에 가까워진다.
- ④ 곡선과 x 축 사이의 면적은 항상 1이다.



정규확률 구하기 (P. 381 ~ 382)

$X \sim N(\mu, \sigma^2)$ 일 때, X 가 a 와 b 사이의
값을 취할 확률 $P(a < X < b)$ 는?

정답: X 의 확률밀도 함수 곡선과 구간 a 에서 b 사이의
면적이다.



질문: 그런데 이 면적을 어떻게
구하나?

* 정규곡선 아래에 있는 면적을 직접 계산하는
일은 어렵다. (점분 활용).

하지만 확률 테이블을 이용하면 쉽게
구할 수 있다.

그리고 확률테이블을 이용하여면
표준화 라성을 거치면 된다.

정규확률 구하기 세 단계

① 분포 결정: μ 와 σ^2 확인

② 표준화 하기: 표준점수 구하기

③ 확률 찾기: 확률 테이블 활용

예제

주의: 책에서 미국 남자들의 키를 대푯값으로 사용함.

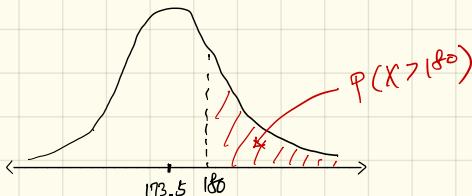
하지만 여기서는 한국 남자들(20~24세)의 키를 대푯값으로 사용함 하지만 설명은 동일함.

문제 ①: 분포 구하기 (P. 3 f3)

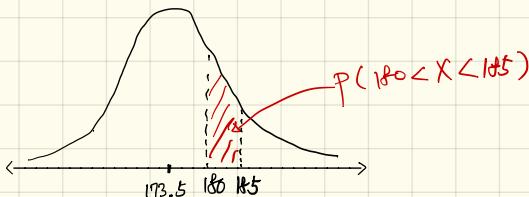
20~24세 사이의 한국 남성들의 키를 X 라 할 때,
 X 의 평균은 173.5 cm이고 표준편차는 5.17 cm이다.

즉, $X \sim N(173.5, 5.17^2)$

문제 ①: 키 180 이상의 남성을 소개팅에서 만날 확률 $P(X > 180)$ 은?



문제 ②: 키 180 이상, 185 이하의 남성을 만날 확률 $P(180 < X < 185)$ 은?



단계 ② : $N(0, 1)$ 를 표준화 하기 (P. 384)

X 의 표준점수 절수는 다음과 같이 구한다:

$$Z = \frac{X - \mu}{\sigma}$$

$\nwarrow X \text{의 평균 } \mu$
 $\nearrow X \text{의 표준편차 } \sigma$

주의: 이렇게 하면 Z 는 표준정규분포를 따른다:
 $Z \sim N(0, 1)$

예제 180과 185의 표준점수를 구하자.

$$180 \text{의 표준점수} = \frac{180 - 173.5}{5.17} = 1.26$$

$$185 \text{의 표준점수} = \frac{185 - 173.5}{5.17} = 2.22$$

따라서 아래 식이 성립한다.

$$P(X > 180) = P(Z > 1.26)$$

$$P(180 < X < 185) = P(1.26 < Z < 2.22)$$

단계 ③ 확률 테이블을 이용하기 (P. 389)

확률 테이블은 $P(Z < a)$ 를 구할 때 활용!

주의: 그레프와
 X 축 사이의 면적은
1이다.



표준정규분포 확률테이블

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511966	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621720	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.663757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802337	0.805105	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823814	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.879000	0.881000	0.882977
1.2	0.884930	0.886861	0.888768	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903200	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.931913	0.934478	0.93745	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958183	0.959070	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965620	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.983297	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.98679	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999534	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999651
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999822	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999915	0.999918	0.999922	0.999925	
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

확률테이블에 의해 아래 결과를 얻는다 =

$$\begin{cases} P(Z < 1.26) \approx 0.896 \\ P(Z < 2.22) \approx 0.987 \end{cases}$$

이 결과를 이용하여 원하는 확률을 구할 수 있다.
먼저 다음 사항에 주의한다.

확률 테이블 사용법 ($\uparrow .391$)

① 확률테이블은 $P(Z < a)$ 의 값을 보여준다.

$$\textcircled{2} \quad P(Z > a) = 1 - P(Z < a)$$

$$\textcircled{3} \quad P(a < Z < b) = P(Z < b) - P(Z < a)$$

문제 해결

소개팅에서 카카 180 이상 또는 네이버에서 185의
남성을 만날 확률 계산하기

문제 ①의 경우 :

$$\begin{aligned}
 P(X > 180) &= P(Z > 1.26) \\
 &= 1 - P(Z < 1.26) \\
 &= 1 - 0.896 \\
 &= 0.104 \quad (\text{즉, } 10.4\%)
 \end{aligned}$$

문제 ③의 경우

$$\begin{aligned}
 P(180 < X < 185) &= P(1.26 < Z < 2.22) \\
 &= P(Z < 2.22) - P(Z < 1.26) \\
 &= 0.987 - 0.896 \\
 &= 0.091 \quad (\text{즉, } 9.1\%)
 \end{aligned}$$

정규분포 활용하기 요점.

- ① 정규분포를 따른다는 사실 확인
- ② 평균과 분산 확인
- ③ 표준정규분포로 표준화
(표준점수 활용)
- ④ 확률테이블을 활용하여 원하는 확률 구하기

주의사항 :

연속데이터를 다룰 때 특정값을 가질 확률은
구하지 않는다.

예를 들어, 소개팅에서 카가 정확히 180cm인
남성을 만날 확률은 0이다.

즉, 연속데이터의 경우 일반적으로 다음 식이
성립한다.

$$P(X = a) = 0.$$