확률 및 통계학

- 6장 정규 분포 -

문 기 운(kiwoon@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

목 차

균일분포

- 균등분포

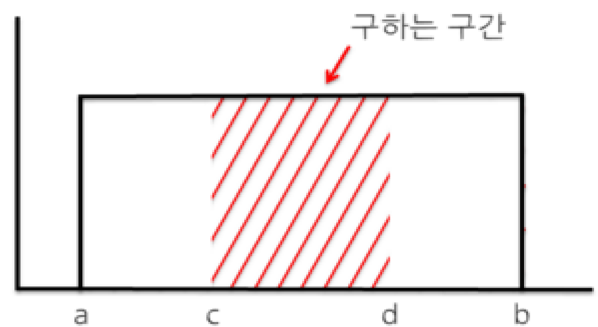
  - 연속확률분포들은 구간의 면적으로 확률을 구함
  - 균등분포의 확률을 구하기 위하여 총 구간 b-a와 구하고자 하는 구간 d-c를 파악

    - 균등분포의 확률값을 구하는 방법

구하는 구간

$$F(x) = \frac{d-c}{b-a}$$

총구간



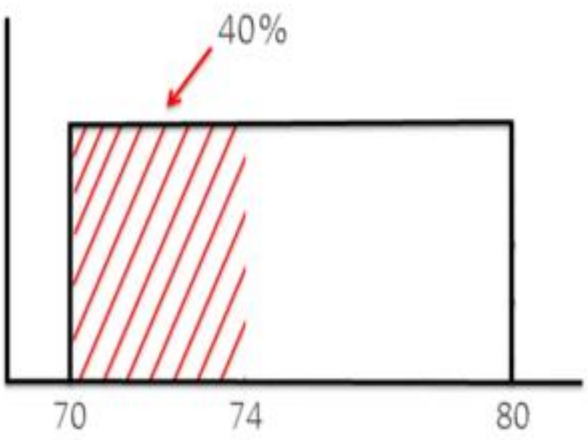
균일분포

[예제 1]
서울에서 춘천까지 고속버스를 타고 가는데, 70분에서 80분이 걸린다고 한다. 이때 74분 안에 춘천까지 도착할 확률을 구하라.

74분 안에 도착할 경우이므로, 74분 이하의 확률을 구하면
총 구간 b-a = 80-70, 구하고자 하는 구간은 d-c=74-70

$$\frac{74-70}{80-70} = 0.4$$

74분 안에 서울에서 춘천까지 도착할 확률은 0.4, 즉 40%



연속성 균일분포

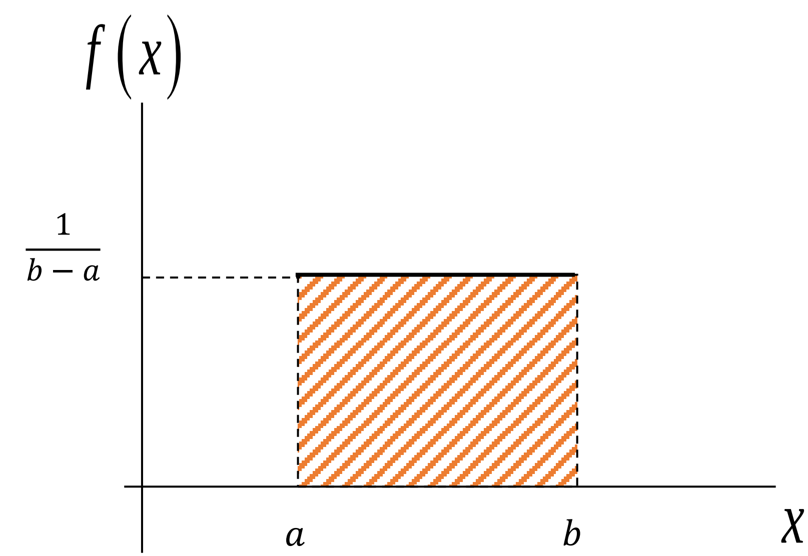
- 연속성 균일분포(continuous suniform distribution)

  - 균일분포는 밀도함수가 존재하는 구간, 예를 들어 구간 $[a,b]$에서 확률밀도가 균일한 분포를 나타냄
  - 임의의 구간에서 확률이 동일
  - 전 구간에서 동일

구간 $[a,b]$에서 정의되는 연속형 균일확률변수 X의 밀도함수

$$f(x;a,b) = f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{다른 곳에서} \end{cases}$$

밀도함수는 밑변이 $\mathrm{b}-a$ 이고, 높이는 $\frac{1}{b-a}$ 로 일정한 직사각형



연속성 균일분포

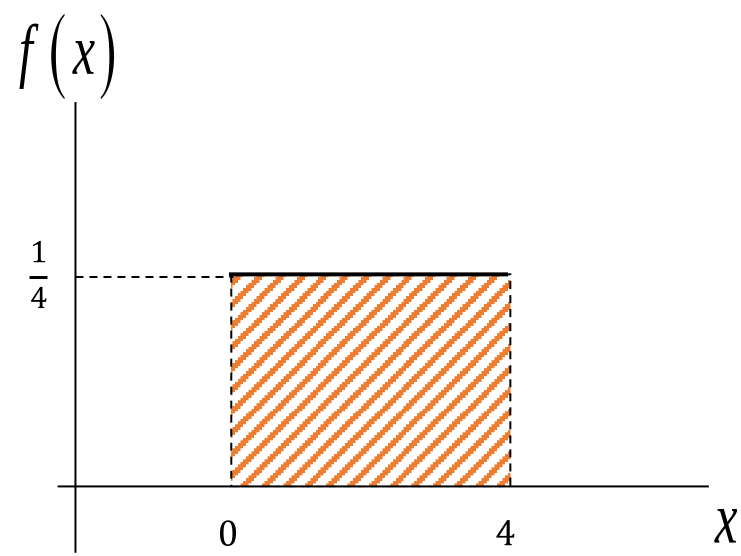
[예제 6.1]
어느 회사에서 대형회의실은 4시간을 초과하여 사용할 수 없다. 그 회의실에서는 긴 회의와 짧은 회의가 자주 열리며, 회의시간 X는 구간 [0,4]에서 정의되는 균일분포로 가정할 수 있다.

(a) 밀도 함수를 구하라.
(b) 어떤 회의가 최소한 3시간 이상 계속될 확률은 얼마인가?

(a) 구간 [0,4]에서 정의되는 연속형 균일확률변수 X의 밀도함수

$$f(x;0,4)=f(x)=\begin{cases}\frac{1}{4}, & 0\le x\le 4\\ 0, & \text{다른 곳에서}\end{cases}$$

밀도함수는 밑변이 4 이고, 높이는 $\frac{1}{4}$ 로 일정한 직사각형



(b) 어떤 회의가 최소한 3시간 이상 계속될 확률을 구하면

$$P[X\ge 3]=\int_3^4 \frac{1}{4}\,dx=\frac{1}{4}$$

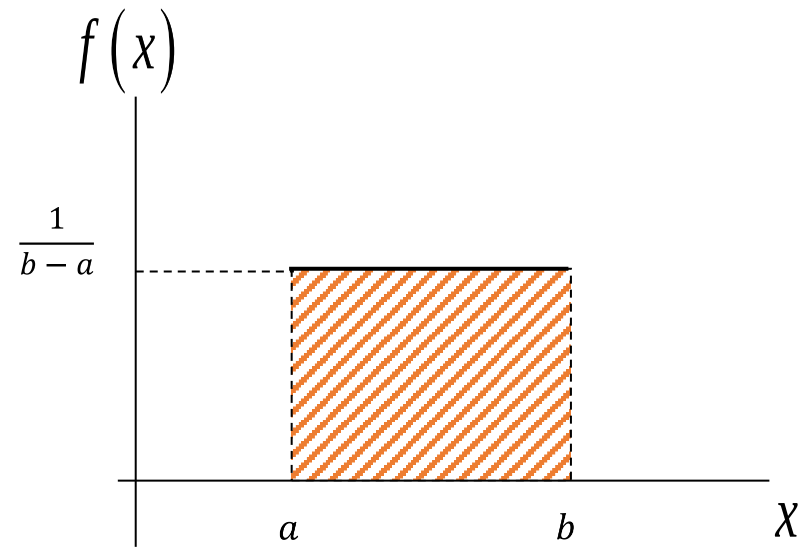
어떤 회의가 최소한 3시간 이상 계속될 확률은 $\frac{1}{4}$ 과 같음

연속성 균일분포

• 균등분포의 평균

구간 $[a,b]$에서 정의되는 연속형 균일확률변수 X의 밀도함수

$$f(x;a,b) = f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{다른 곳에서} \end{cases}$$



$$\mu = E(X) = \int_a^b x\, f(x)dx = \int_a^b \frac{x}{b-a}dx$$

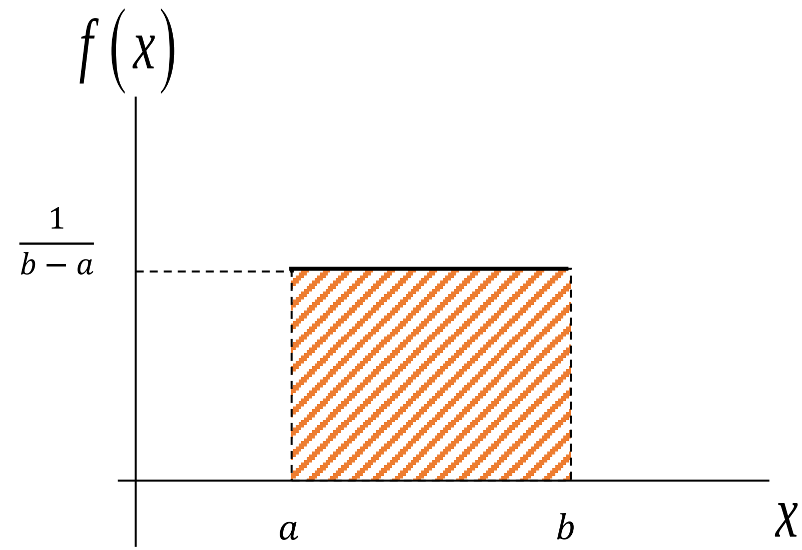
$$= \frac{1}{b-a}\left(\frac{x^2}{2}\right)\Bigg|_a^b = \frac{a+b}{2}$$

연속성 균일분포

- ## 균등분포의 분산

구간 $[a,b]$에서 정의되는 연속형 균일확률변수 X의 밀도함수

$$f(x;a,b) = f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{다른 곳에서} \end{cases}$$



$$E(X^2) \quad \int_a^b x^2 f(x)dx = \int_a^b \frac{x^2}{b-a}dx$$

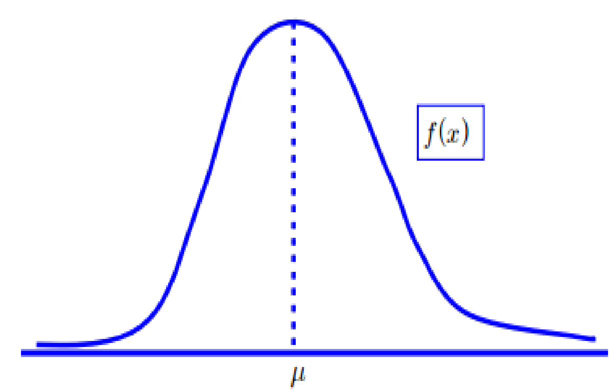
$$= \frac{1}{b-a}\left(\frac{x^3}{3}\right)\Bigg|_a^b = \frac{a^2+ab+b^2}{3}$$

$$\sigma^2 = Var(X) = E(X^2) - E(X)^2$$

$$= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \mathbf{\frac{(b-a)^2}{12}}$$

정규분포

- 정규분포(normal distribution)
  - 1733년에 드무아브르가 정규곡선의 모형을 제안
  - 1820년 라플라스에 의해 분포의 함수식이 도출
  - 동일한 양의 반복측정에서의 오차에 관한 연구에서 정규곡선의 방정식을 유도해낸 가우스를 기리기 위해 가우스 분포라고도 함
  - 자연과학, 기업, 각종 연구분야에서 발생하는 여러 현상들은 대부분 정규분포를 따름



정규분포

- 정규분포(normal distribution)

확률변수 X가 음의 무한대에서 양의 무한대사이의 어떠한 값도 취할 수 있는 연속확률분포로서 μ(기대값)와 σ(표준편차) 2개의 모수에 의해 특징지어지는 분포

기대값 μ와 분산 σ^2 을 가지는 정규확률변수 X의 확률분포는

$$\mathrm{n}(\mathrm{x};\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad , -\infty < x < \infty$$

$$\pi = 3.14159\cdots, e = 2.71828\cdots$$

정규분포

- 정규분포의 특징

  - 정규곡선은 종모양을 나타냄
  - 정규분포는 평균을 중심으로 좌우대칭을 이룸
    - 평균=중앙치=최빈값=최대값
  - 정규분포의 형태와 위치는 평균과 표준편차가 결정
  - 정규곡선은 x축에 닿지 않으므로 확률변수 X의 범위는 $-\infty < x < \infty$
  - 정규곡선 밑의 면적은 1
  - 정규곡선 밑의 두 점 사이의 면적은 정규확률변수가 이들 두 점 사이를 취할 확률

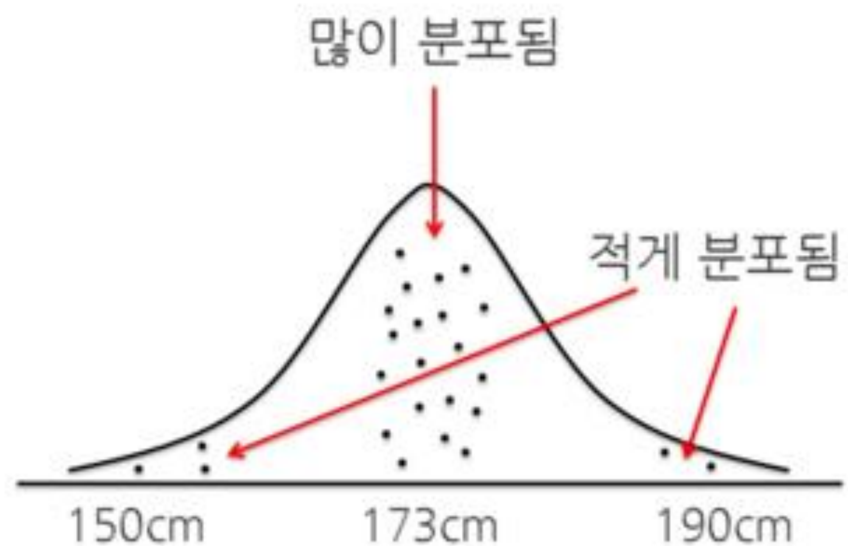
정규분포

- 정규분포의 특징

- 정규곡선은 종모양을 나타냄
- 정규분포는 평균을 중심으로 좌우대칭을 이룸
    - 평균=중앙치=최빈값=최대값

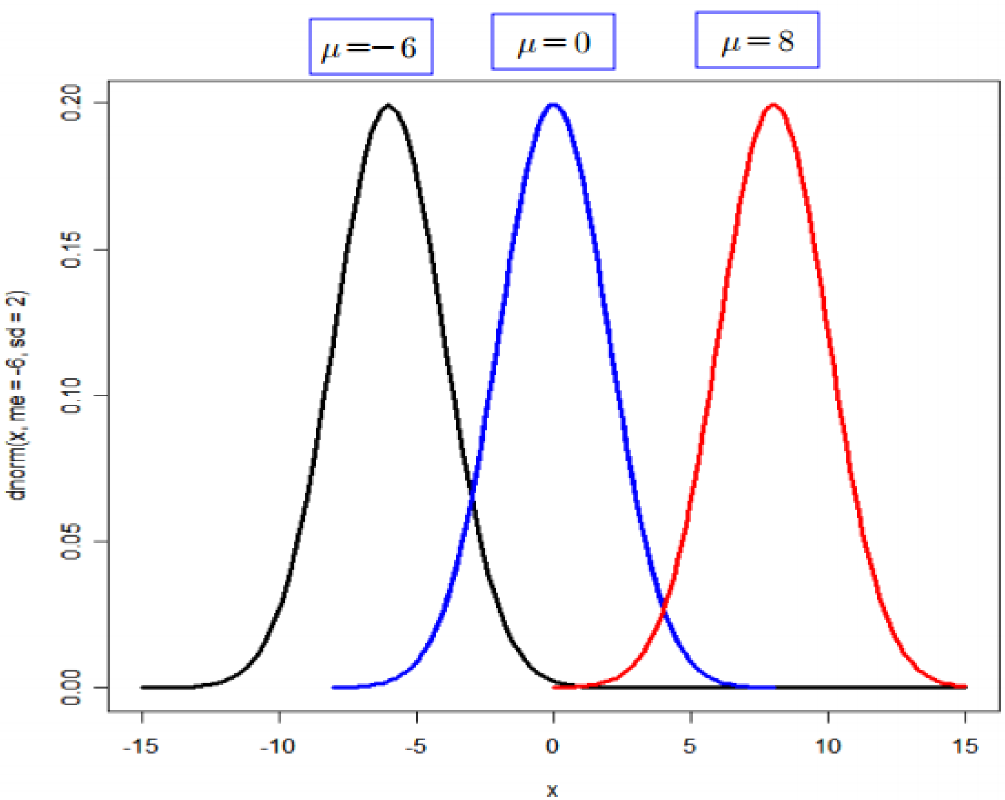
대한민국 성인 남성의 키는 정규분포를 나타냄

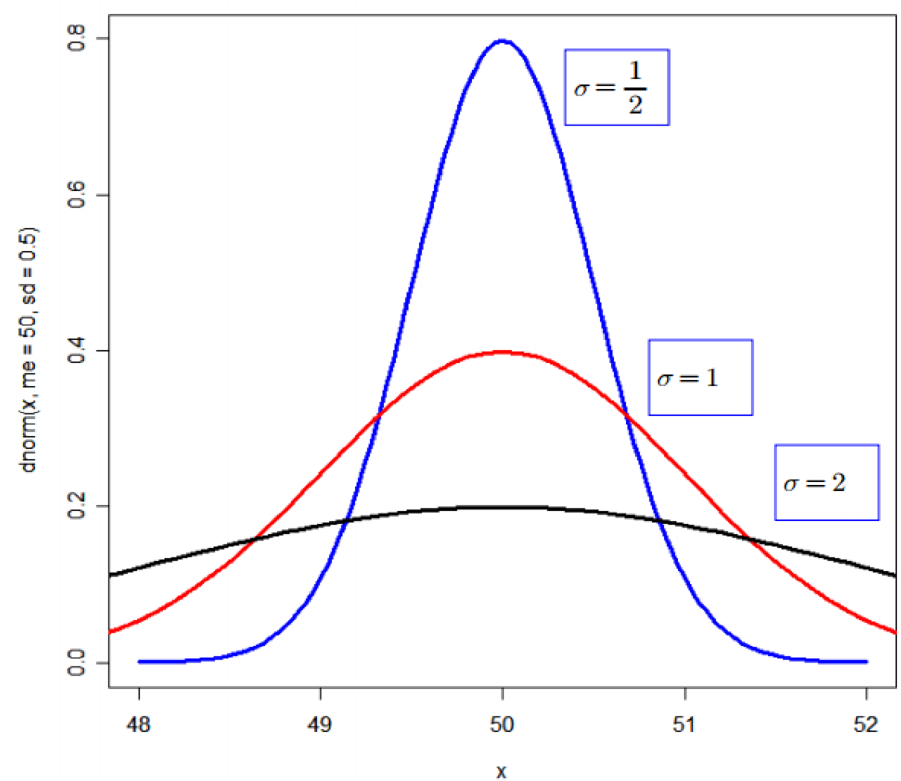
평균에 가까울 수록 데이터의 분포가 많고
평균에서 멀어질 수록 데이터의 분포가 감소



정규분포

- 정규분포의 특징
  - 정규분포의 형태와 위치는 평균과 표준편차가 결정
  - 정규곡선은 x축에 닿지 않으므로 확률변수 X의 범위는 $-\infty < x < \infty$

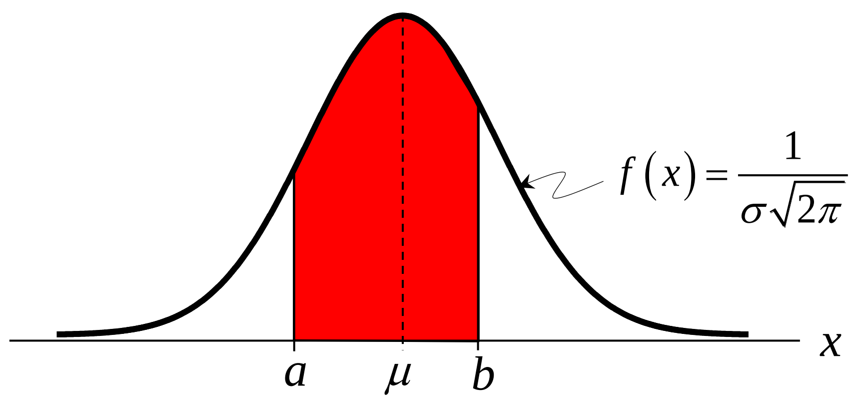




정규분포

X가 모수 m, s 의 정규분포를 갖는 확률변수일 때

X가 가질 수 있는 값 x가 범위 (a, b) 에 속할 확률 $P(a \le X \le b)$ 은



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

영역의 면적이다. 이 값은 적분계산 $P(a \le X \le b) = \int_a^b f(x)dx$ 으로 구함

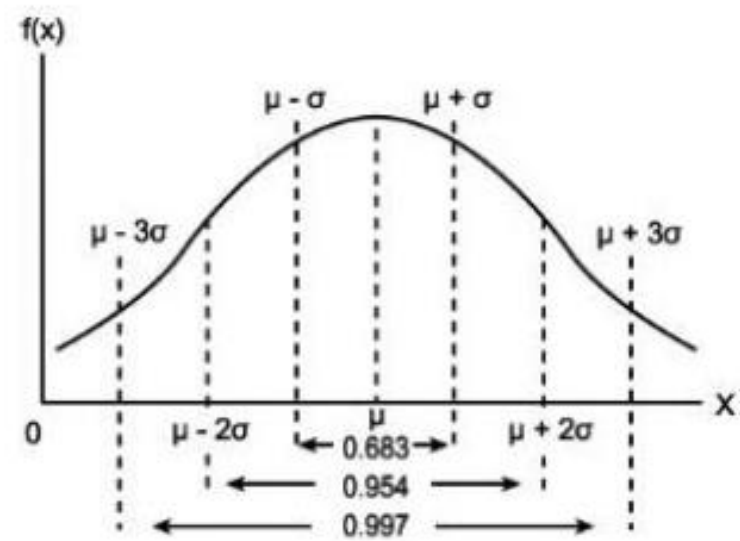
- 확률변수가 갖게 될 값의 범위에서의 확률을 구하기 위하여 정규밀도함수의 적분을 그때 마다하는 것은 어려움

$$P(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

정규분포

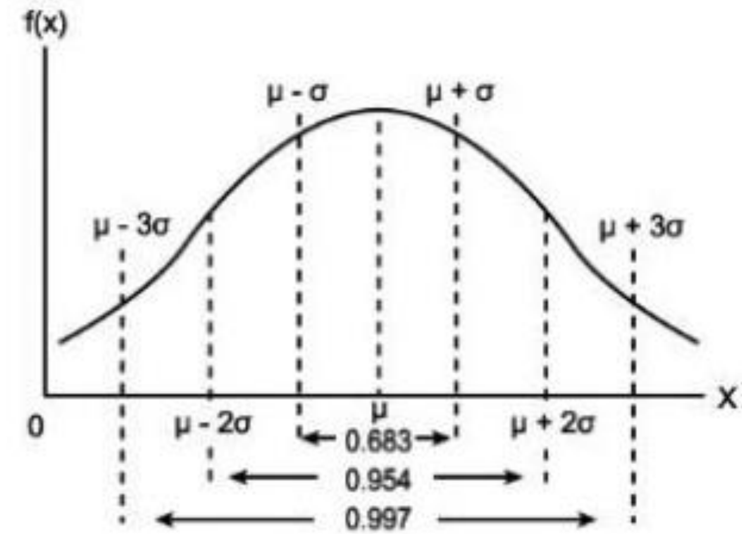
- **정규분포를 가지는 확률밀도함수와 평균, 표준편차의 관계**
  - 정규 곡선의 위치는 기대값 μ에 의하여 결정되고, 폭은 표준편차 σ 크기에 의해서 결정
  - 자료 집단에 따라 기대값과 표준편차의 크기는 다르나, 평균으로 부터 표준편차의 k배 이내의 범위에서 확률변수 X값을 갖게 될 확률은 같음
  - 즉, 확률변수 X가 기대값 μ 와 분산이 σ^2 인 정규분포를 취한다면, 기대값 μ 로부터 $\pm\sigma$, $\pm2\sigma$, $\pm3\sigma$ 범위내의 확률변수 값이 포함된 확률은 각각 68.3%, 95.4%, 99.7% 임

$$P(\mu - 1\sigma < X < \mu + 1\sigma) = 0.6826$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9974$$



정규분포

$$P(\mu - 1\sigma < X < \mu + 1\sigma) = 0.6826$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9974$$



"어느 지역의 4인 가구의 생활비를 평균 170만원 표준편차 20만원 정규분포로 가정했다."

"이 지역에 있는 가구의 약 68%가 생활비를 170-(1*20)= 150만원, 170+(1*20)= 190안원 사이에서 쓴다."

"이 지역의 170만원보다 2배의 표준편차를 더 쓰는 가구는, 170+(2*20)=210만원을 쓰는 사람들은 가구전체의 2.28% ($\frac{1-0.9544}{2}$) 이다."

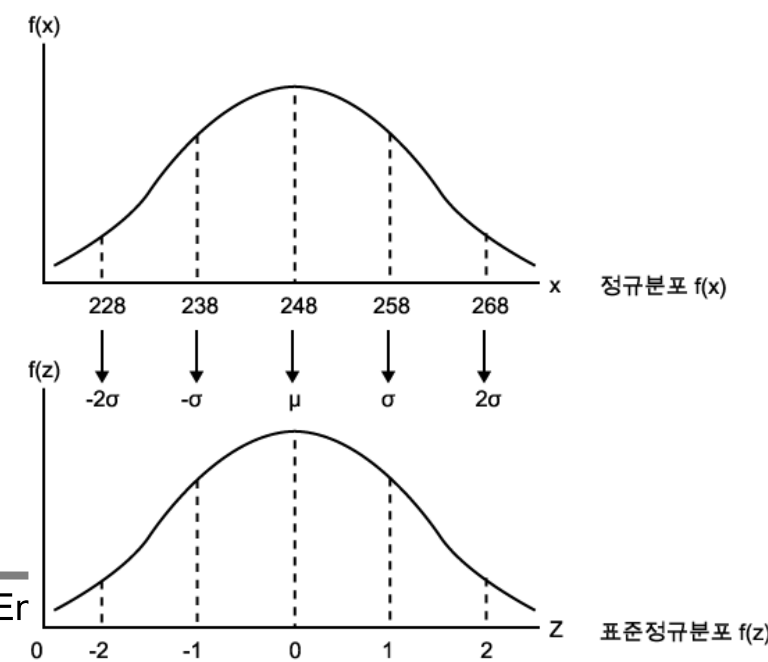
표준정규분포

- **표준정규분포**

  정규분포를 표준화한 특수한 형태의 정규분포를 말하며, 확률변수 X를 Z값으로 표준화한 분포로서 기대값(μ)이 0이고 표준편차(σ) 가 1인 정규분포를 따를 때 표준정규분포라고 함

- 정규분포의 표준화
  - 매 실험이나 조사 때 마다 그래프의 함수와 면적의 넓이를 계산하는 일은 매우 번거로움
  - 정규분포를 공식을 거쳐 표준화한 특수한 형태의 표준정규분포는 Z분포라고도 함

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

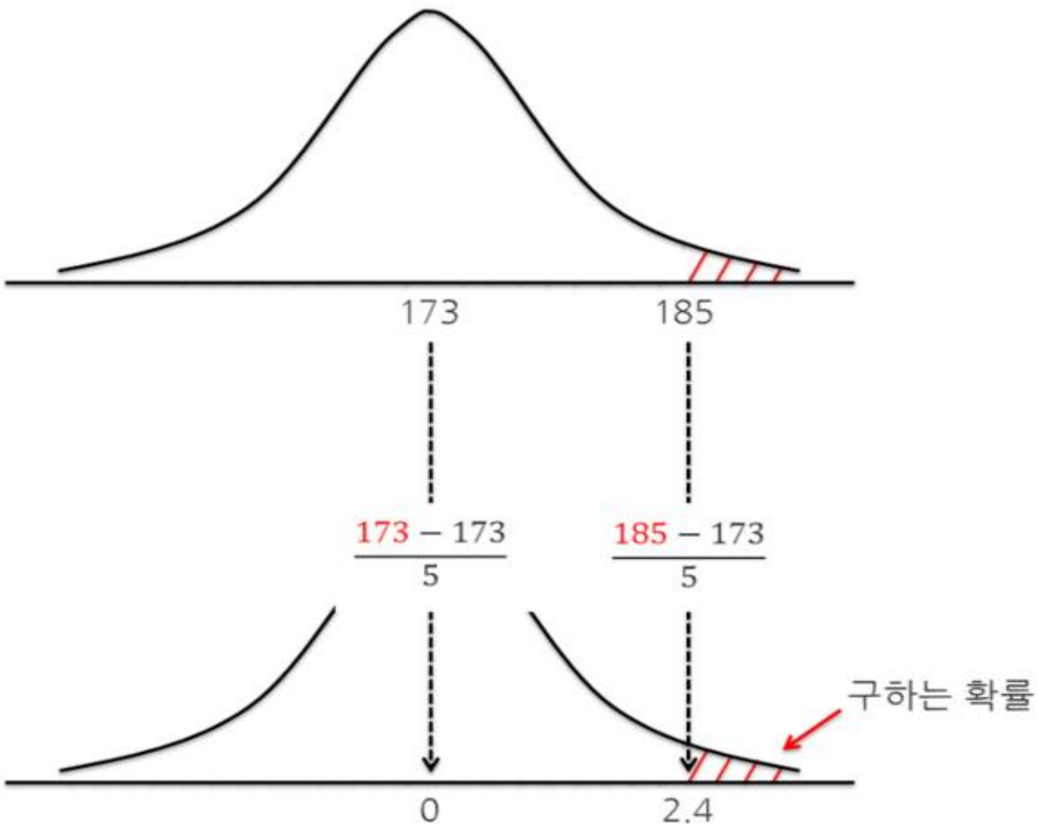


표준정규분포

- 정규분포의 표준화의 예

우리나라 성인 남성의 키는 평균 173cm, 표준편차 5인 정규분포를 따른다고 할 때, 185cm 이상인 남성의 비율을 구하기 위하여 정규분포를 표준화 하여라

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



표준정규분포

- **정규분포와 표준정규분포와의 관계**

X가 모수 μ, σ 의 정규분포를 갖는 확률변수일 때

X가 가질 수 있는 값 x가 범위 (a, b) 에 속할 확률 $P\left(a \le X \le b\right)$ 은

표준화 Z의 값 z가 표준화된 범위 $(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma})$ 에 속할 확률 $P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$ 와 같음

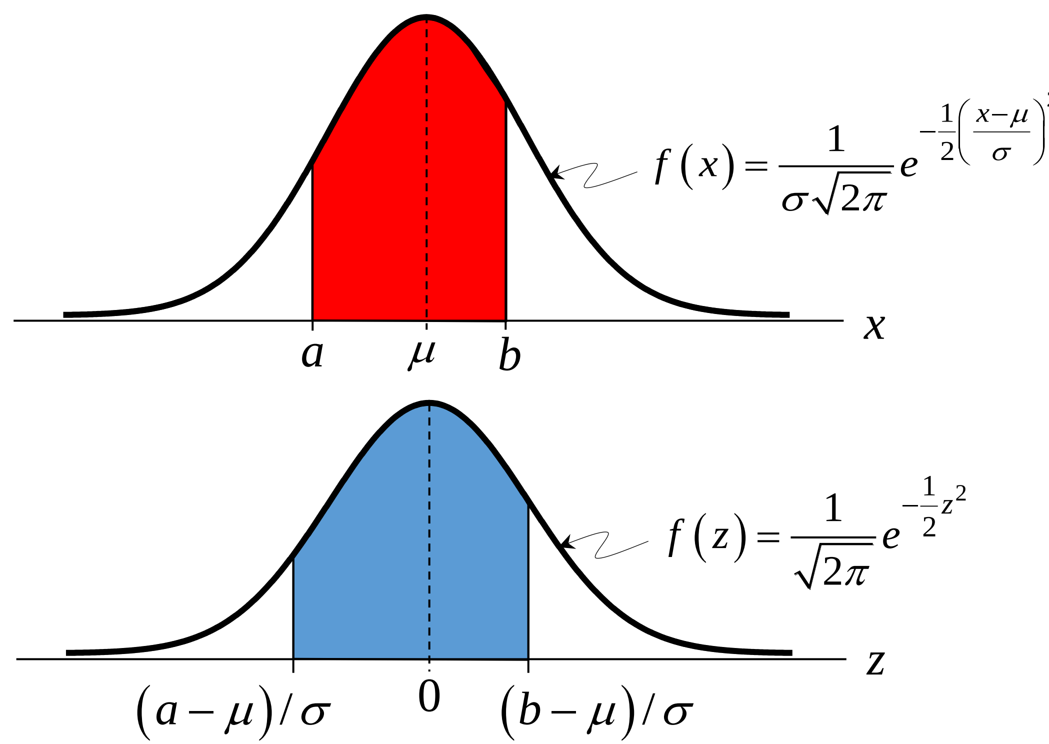
$P\left(a \le X \le b\right) =$ 의 면적

표준화

$P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) =$ 의 면적

의 면적과

의 면적은 같다



표준정규분포

- 표준정규분포표

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	0.662757	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	0.933193	0.934478	0.935745	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	0.958185	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	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	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	0.999912	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

표준정규분포

- **정규분포의 적용**

어느 축전지는 평균 수명이 3년이고 표준편차가 0.5년인 것으로 알려져 있다. 축전지의 수명이 정규분포를 따른다고 가정할 때, 임의로 주어진 전지의 수명이 2.3년보다 짧을 확률을 구하라.

$P(X < 2.3)$ 을 구해야 하므로 표준화하면 $Z = \dfrac{2.3 - 3}{0.5} = -1.4$

$$P(X < 2.3) = P(Z < -1.4) = 0.0808$$

표준정규분포

- **정규분포의 적용**

어느 전기회사에서는 평균수명이 800시간이고 표준편차가 40시간인 정규분포의 수명분포를 가지는 전구를 생산하고 있다. 임의로 선정된 전구의 수명이 778시간과 834시간 사이에 있을 확률을 구하라.

$P(778 < X < 834)$ 를 구해야 하므로 표준화하면

$$Z_1 = \frac{778 - 800}{40} = -0.55 \qquad Z_2 = \frac{834 - 800}{40} = 0.85$$

$$P(778 < X < 834) = P(-0.55 < Z < 0.85) = P(Z < 0.85) - P(Z < -0.55)$$

$$= 0.8023 - 0.2912 = 0.5111$$

감사합니다!