8-2 이론적 확률분포(연속형) 2

제주권역 대학 e-러닝 지원센터

정규분포의 표준화(표준정규분포)

· 확률변수 X가 $X \sim N(\mu, \sigma^2)$일 때, 다음과 같이 표준화한 Z는 평균이 0이고 분산이 1인 정규분포를 따르는데 이를 표준정규분포라고 한다.

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

·확률변수 Z가 평균이 0이고 분산이 1인 정규분포를 따를 때 Z는 표준정규분포를 따른다고 하며, $Z \sim N(0,1)$ 으로 표현한다.

Z의 확률밀도함수는 다음과 같다

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \ -\infty < z < \infty$$

표준정규분포

· 평균이 0이고 분산이 1인 정규확률변수는 Z로 나타내며 다음과 같이 표현한다.

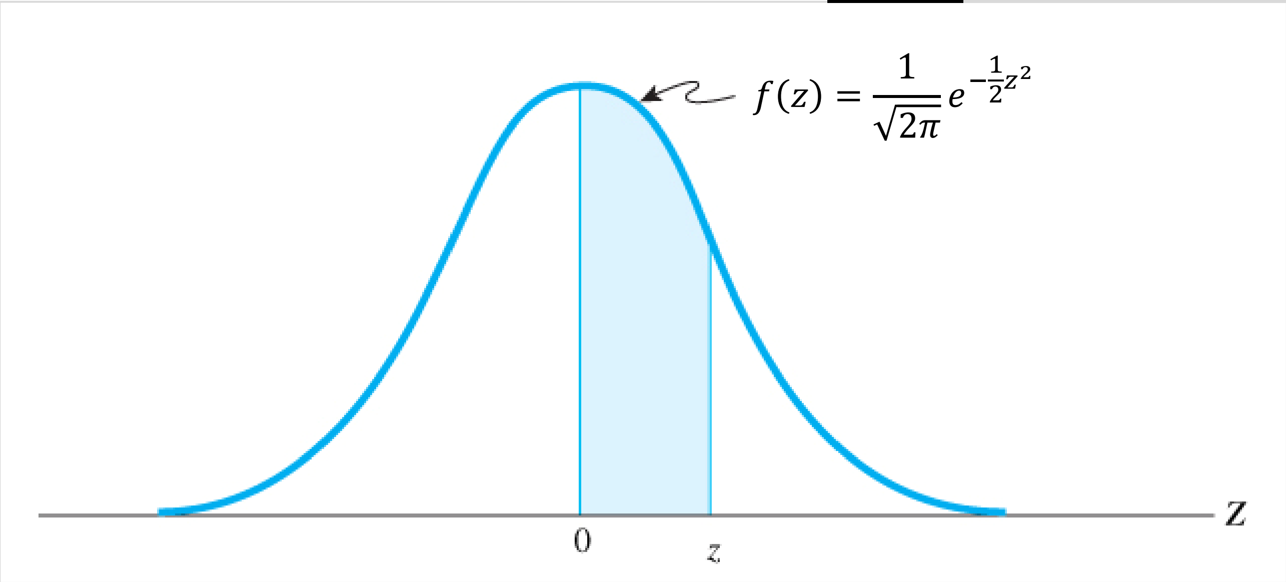
$$Z \sim N(0,\ 1)$$

· 이때 Z는 표준정규분포를 따른다고 한다.

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

$$Var(z) = E[z-E(z)]^2 = E(\frac{X-\mu}{\sigma})^2 = \frac{1}{\sigma^2}E(X-\mu)^2 = 1$$

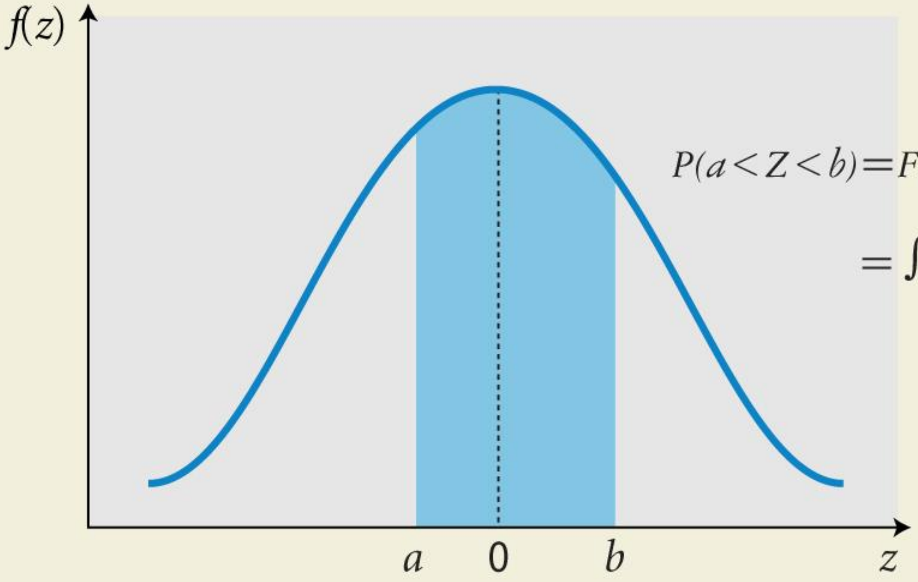
그림 표준정규분포



표준정규분포의 형태는 그림과 같으며, 중심 0에서부터 양의 값 Z까지의 확률은 색칠한 부분의 넓이와 같다.

표준정규분포의 확률계산

그림 표준정규분포의 a와 b구간확률



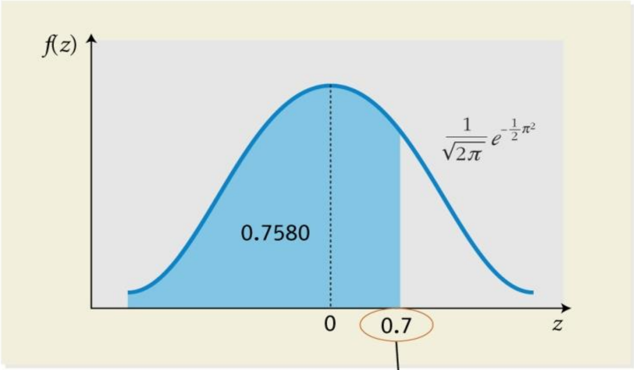
$$P(a<Z<b)=Fz(b)-Fz(a)$$

$$=\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

그림 z가 0.70보다 작을 확률

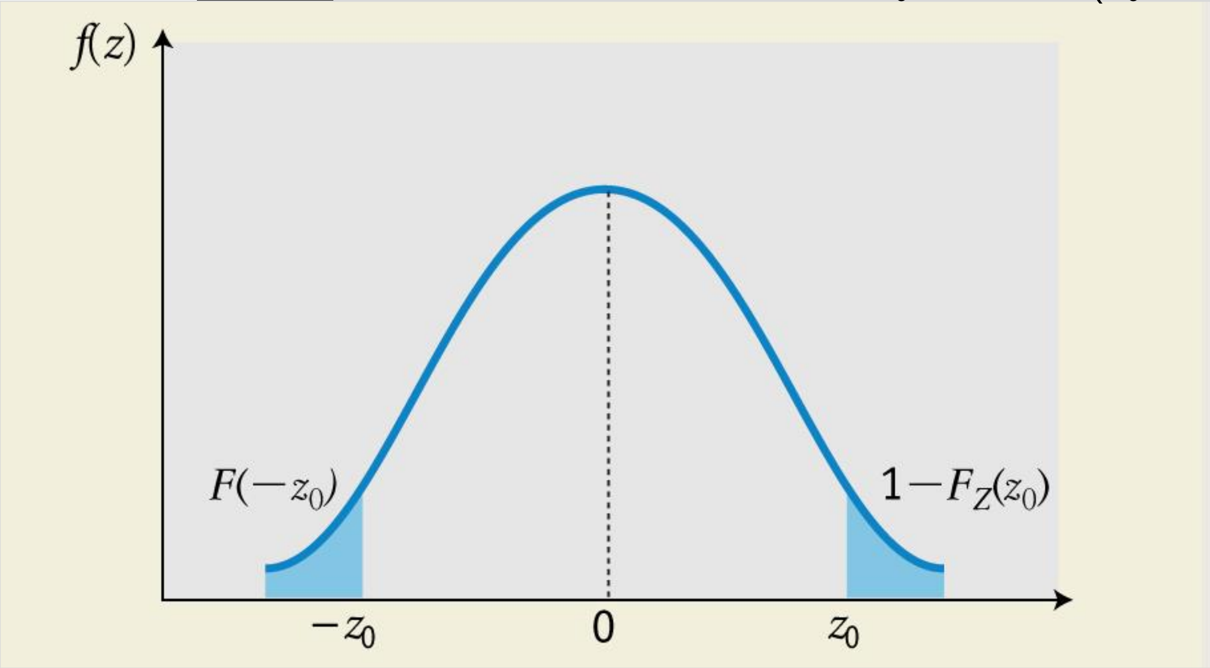
```r
zc<-rep(NA)

for(i in 1:101) {
  zc[i]<-pnorm((i-1)/100, 0, 1)
}
zc
```

```
> zc
  [1] 0.5000000 0.5039894 0.5079783 0.5119665 0.5159534 0.5199388 0.5239222 0.5279032 0.5318814
 [10] 0.5358564 0.5398278 0.5437953 0.5477584 0.5517168 0.5556700 0.5596177 0.5635595 0.5674949
 [19] 0.5714237 0.5753454 0.5792597 0.5831662 0.5870644 0.5909541 0.5948349 0.5987063 0.6025681
 [28] 0.6064199 0.6102612 0.6140919 0.6179114 0.6217195 0.6255158 0.6293000 0.6330717 0.6368307
 [37] 0.6405764 0.6443088 0.6480273 0.6517317 0.6554217 0.6590970 0.6627573 0.6664022 0.6700314
 [46] 0.6736448 0.6772419 0.6808225 0.6843863 0.6879331 0.6914625 0.6949743 0.6984682 0.7019440
 [55] 0.7054015 0.7088403 0.7122603 0.7156612 0.7190427 0.7224047 0.7257469 0.7290691 0.7323711
 [64] 0.7356527 0.7389137 0.7421539 0.7453731 0.7485711 0.7517478 0.7549029 0.7580363 0.7611479
 [73] 0.7642375 0.7673049 0.7703500 0.7733726 0.7763727 0.7793501 0.7823046 0.7852361 0.7881446
 [82] 0.7910299 0.7938919 0.7967306 0.7995458 0.8023375 0.8051055 0.8078498 0.8105703 0.8132671
 [91] 0.8159399 0.8185887 0.8212136 0.8238145 0.8263912 0.8289439 0.8314724 0.8339768 0.8364569
[100] 0.8389129 0.8413447
```



z	$F(z)$	z	$F(z)$	z	$F(z)$	z	$F(z)$
.00	.5000						
.01	.5040	.31	.6217	.61	.7291	.91	.8186
.02	.5080	.32	.6255	.62	.7324	.92	.8212
.03	.5120	.33	.6293	.63	.7357	.93	.8238
.04	.5160	.34	.6331	.64	.7389	.94	.8264
.05	.5199	.35	.6368	.65	.7422	.95	.8289
.06	.5239	.36	.6406	.66	.7454	.96	.8315
.07	.5279	.37	.6443	.67	.7486	.97	.8340
.08	.5319	.38	.6480	.68	.7517	.98	.8365
.09	.5359	.39	.6517	.69	.7549	.99	.8389
.10	.5398	.40	.6554	.70	.7580	1.00	.8413

그림 **z의 확률밀도 함수 : $F_Z(-z_0)$과 $1-F_Z(z_0)$**



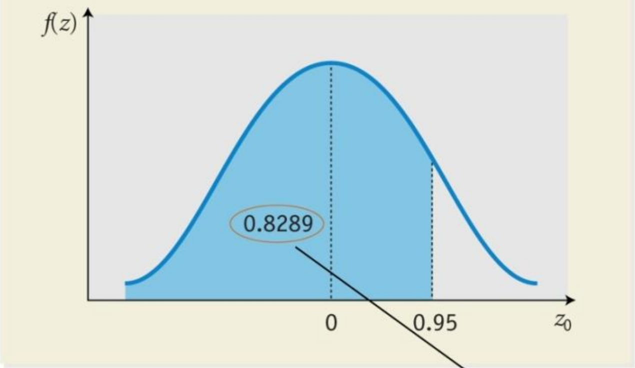
예 Z가 표준정규확률변수라면 Z값이 -1과 0.5사이에 놓일 확률을 구하라.

풀이

이 확률은 표준정규분포표를 이용하여 다음과 같이 계산할 수 있다.

$$
\begin{aligned}
P(-1<Z<0.5) &= F_Z(0.5)-F_Z(-1) \\
&= F_Z(0.5)-[1-F_Z(1)] \\
&= 0.6915-[1-0.8413] \\
&= 0.5328
\end{aligned}
$$

그림 $F_Z(z_0)=0.8289$가 되는 z_0 값



z	F(z)	z	F(z)	z	F(z)	z	F(z)
.00	.5000						
.01	.5040	.31	.6217	.61	.7291	.91	.8186
.02	.5080	.32	.6255	.62	.7324	.92	.8212
.03	.5120	.33	.6293	.63	.7357	.93	.8238
.04	.5160	.34	.6331	.64	.7389	.94	.8264
.05	.5199	.35	.6368	.65	.7422	.95	.8289
.06	.5239	.36	.6406	.66	.7454	.96	.8315
.07	.5279	.37	.6443	.67	.7486	.97	.8340
.08	.5319	.38	.6480	.68	.7517	.98	.8365
.09	.5359	.39	.6517	.69	.7549	.99	.8389
.10	.5398	.40	.6554	.70	.7580	1.00	.8413

예

표준정규확률변수 Z가 어떤 구간에 속할 확률이 0.90이고
이 구간이 0을 기준으로 대칭을 이룰 때 이 구간을 구하라.

풀이

0을 기준으로 대칭을 이루며 Z가 놓일 확률이 0.9인 구간은

$$P(-z_0 < Z < z_0) = 0.9$$

가 되며, 이를 만족하는 z_0 을 구하면 된다.
2• $F_Z(-z_0) = 1-0.9$와 $F_Z(z_0) = 0.9 + F_Z(-z_0)$를 이용하면

$F_Z(z_0) = 0.9 + 0.05 = 0.95$가 되므로 누적분포의 값이 0.95가 되는 z_0를
표준정규분포표로부터 구하면 그 값은 1.645이다.
따라서 구하고자 하는 구간은 -1.645부터 1.645까지가 된다.

```r
z<-rep(NA)

for(i in 1:110) {
  z[i]<-pnorm((i-1)/100, 0, 1)-0.5
}
z

```

```
> z
  [1] 0.000000000 0.003989356 0.007978314 0.011966473 0.015953437 0.019938806 0.023922183 0.027903170 0.031881372 0.035856393
 [11] 0.039827837 0.043795313 0.047758426 0.051716787 0.055670005 0.059617692 0.063559463 0.067494932 0.071423716 0.075345435
 [21] 0.079259709 0.083166163 0.087064423 0.090954115 0.094834872 0.098706326 0.102568113 0.106419873 0.110261248 0.114091881
 [31] 0.117911422 0.121719522 0.125515835 0.129300019 0.133071736 0.136830651 0.140576433 0.144308755 0.148027292 0.151731727
 [41] 0.155421742 0.159097026 0.162757273 0.166402179 0.170031446 0.173644780 0.177241890 0.180822491 0.184386303 0.187933051
 [51] 0.191462461 0.194974269 0.198468212 0.201944035 0.205401484 0.208840313 0.212260281 0.215661151 0.219042691 0.222404675
 [61] 0.225746882 0.229069096 0.232371107 0.235652708 0.238913700 0.242153889 0.245373085 0.248571105 0.251747770 0.254902906
 [71] 0.258036348 0.261147932 0.264237502 0.267304908 0.270350003 0.273372648 0.276372708 0.279350054 0.282304562 0.285236116
 [81] 0.288144601 0.291029912 0.293891946 0.296730608 0.299545807 0.302337457 0.305105479 0.307849798 0.310570345 0.313267057
 [91] 0.315939875 0.318588745 0.321213620 0.323814458 0.326391220 0.328943874 0.331472393 0.333976754 0.336456941 0.338912940
[101] 0.341344746 0.343752355 0.346135770 0.348494997 0.350830050 0.353140944 0.355427700 0.357690346 0.359928910 0.362143428
```

정규분포의 표준화와 확률계산

정규분포의 표준화

X를 평균 μ와 분산 σ^2을 갖는 정규확률변수라고 하자. 이때 X를

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

로 선형변환하면 Z는 표준정규분포한다. 즉 $Z \sim N(0, 1)$이다. X를 Z로 선형변환시키는 것을 **표준화**(standardized)한다고 한다. 물론 역으로 Z를 X로 다음과 같이 선형변환시킬 수도 있다.

$$X = \mu + Z\sigma$$

정규분포의 확률계산

$X \sim N(\mu, \sigma^2)$이고 $a < b$인 두 실수 a와 b가 존재한다고 하자. 이때

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

$$= F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)$$

확률변수 X가 $X \sim N(50,\ 10^2)$일 때 X가 60보다 클 확률을 표준화 공식을 이용하여 계산해 보자.

$$P(X>60)=P\left(\frac{X-\mu}{\sigma}>\frac{60-\mu}{\sigma}\right)$$

$$=P\left(Z>\frac{60-\mu}{\sigma}\right)$$

$$=P\left(Z>\frac{60-50}{10}\right)$$

$$=P(Z>1)$$

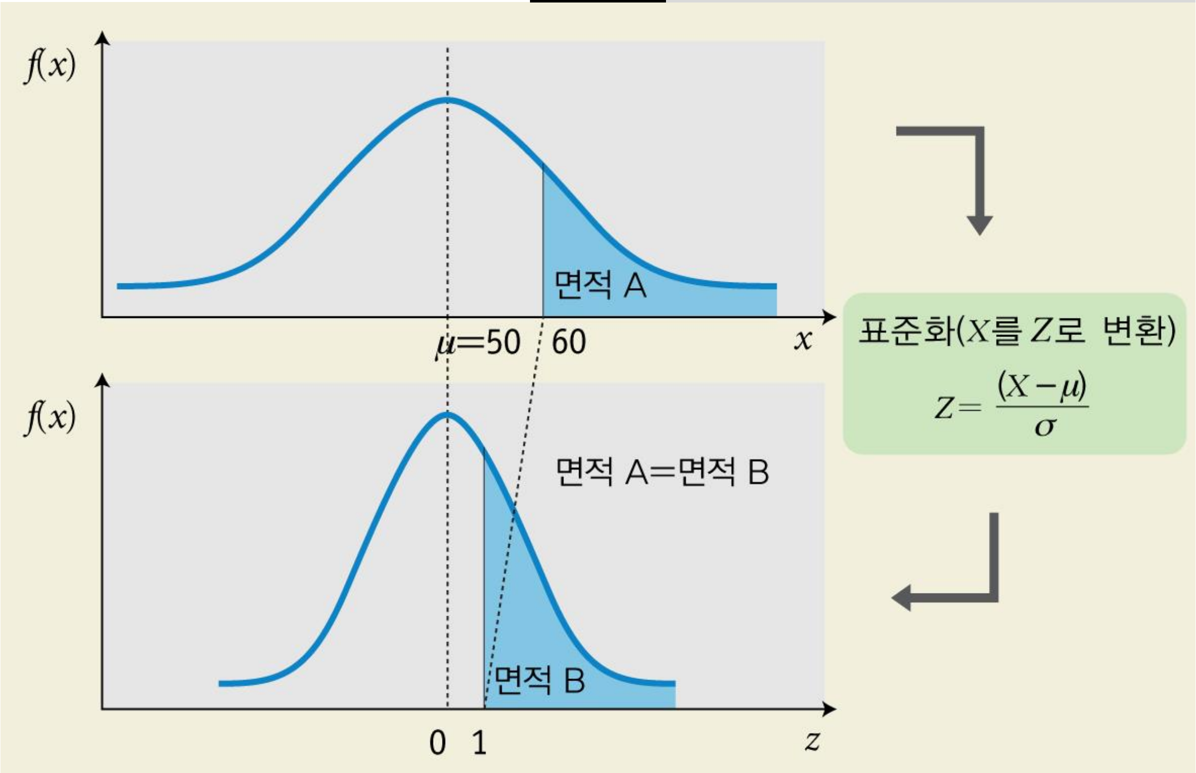
$$=1-F_Z(1)$$

$$=1-0.8413$$

$$=0.1587$$

z	$F(z)$	z	$F(z)$
.61	.7291	.91	.8186
.62	.7324	.92	.8212
.63	.7357	.93	.8238
.64	.7389	.94	.8264
.65	.7422	.95	.8289
.66	.7454	.96	.8315
.67	.7486	.97	.8340
.68	.7517	.98	.8365
.69	.7549	.99	.8389
.70	.7580	1.00	.8413

그림 X와 Z의 확률밀도함수



- 어느 투자가 A씨는 1천만 원을 증권에 투자하려고 한다. 증권회사에서 추천한 두 개의 증권에 대한 수익률 분포를 조사해보니 두 증권은 다음과 같은 정규분포를 한다. (단위 : %)

$$X_1 \sim N(10, 4^2)$$
$$X_2 \sim N(14, 6^2)$$

- A씨는 두 개의 증권 중에서 하나에만 투자하려고 한다. 증권 2는 높은 평균수익률을 갖는 대신에 수익률의 분산이 커서 투자에 대한 위험성은 증권 1보다 크다.

① A씨는 수익률이 0%보다 작게 되는 확률이 작은 증권을 택하려고 한다면 어느 증권을 택하게 되는가?

$$P(X_1 < 0) = P\left(z < \frac{0-10}{4}\right) = P(z < -2.5) = 0.0062$$

$$P(X_2 < 0) = P\left(z < \frac{0-14}{6}\right) = P(z < -2.333) = 0.0098$$

따라서, A씨는 증권 1을 택한다.

② A씨는 수익률이 10%보다 크게 되는 확률이 큰 증권을 택하려고 한다면 어느 증권을 택하게 되는가?

$$P(X_1 > 10) = P\left(z > \frac{10 - 10}{4}\right) = P(z > 0) = 0.5$$

$$P(X_2 > 10) = P\left(z > \frac{10-14}{6}\right) = P(z > -0.67) = 0.7486$$

따라서, A씨는 증권 2를 선택한다.