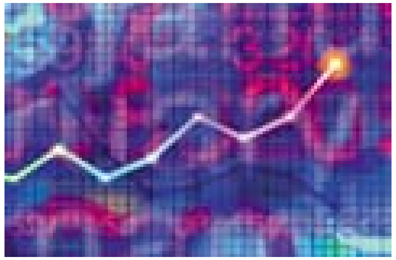
제6장 연속형 확률 변수



연속형 확률변수

❖ 연속형확률변수 **(Continuous Random Variable)**

: 확률변수가 취할 수 있는 값의 수가 셀 수 없이 많을 때

❖ 연속형확률변수의 확률밀도함수**(Probability Density Function)**

모든 실수 a,b 에 대하여 $P(a \le X \le b) = \int_a^b f(x)dx$ 를 만족시키는 $f(x)$

- 확률 밀도 함수가 되기 위한 성질 2가지
  ① 모든 실수 값 x 에 대하여 $f(x) \ge 0$ 이어야 한다
  ② $\int_{-\infty}^{\infty} f(x)\, dx = 1$을 만족 시킨다
- 특징
  임의의 상수에 대하여 $P(X = a) = \int_a^a f(x)\, dx = 0$
  $\Rightarrow P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$

연속형 확률변수

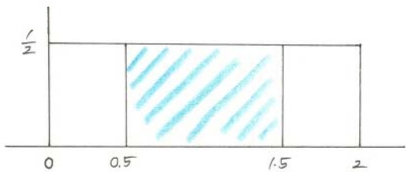
❖ **(예제 6.1)** 실수 구간 **[0,1]**에서 한 점을 임의로 선택하여 그 값을 **x**로 둔다. 확률변수 **X**는 **0**과 **1**사이의 값을 동일한 확률로 취하게 됨 따라서**,** 임의의 $0 \le a \le b \le 1$에 대하여 $P(a \le X \le b) = b - a$가 된다**.**

❖ **(예제 6.2)** 확률변수 **X**의 확률밀도함수가 아래와 같이

$$f(x) = \begin{cases} 0.5 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

주어졌을 때**,** 확률 $P(0.5 \le X \le 1.5)$ 값은**?**

$$P(0.5 \le X \le 1.5) = \int_{0.5}^{1.5} \frac{1}{2}\, dx = 0.5$$



연속형 확률변수

❖ 연속형 확률변수의 누적분포함수

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\, dx$$

- 누적 분포 함수 $F(x)$와 확률 밀도 함수 $f(x)$의 관계

$$\Rightarrow F'(x) = f(x)$$

❖ 연속형 확률변수의 평균과 분산

- $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$
- $V(X) = E(X - \mu)^2$
  $= E(X^2) - E(X)^2$
  $= \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - \left(\int_{-\infty}^{\infty} x \cdot f(x)\, dx \right)^2$

연속형 확률변수

❖ (예제 **6.3**) 예제 **6.2**에서 누적분포함수는 예제 **6.2**의 결과를 이용하여,

$$f(x)=\begin{cases}0.5 & 0\le x\le 2\\ 0 & \text{otherwise}\end{cases}$$

$$F(x)=\int_{-\infty}^{x} f(x)\,dx=\begin{cases}0 & \text{if } x<0\\ \dfrac{x}{2} & \text{if } 0\le x\le 2\\ 1 & \text{if } 2<x\end{cases}$$

임을 알 수 있다.

연속형 확률변수

❖ (예제 **6.4**)
확률변수 X의 확률밀도 함수가 $f(x)=3x^2, 0\le x\le 1$
로 주어졌을 때, $P(X\le 0.5)$, 평균, 그리고 분산을 구하라.

- $P(X\le 0.5)=\int_0^{0.5} f(x)\,dx=\int_0^{0.5} 3x^2\,dx=\frac{1}{8}$
- $E(X)=\int_0^1 x\cdot f(x)\,dx=\int_0^1 x\cdot 3x^2\,dx=\int_0^1 3x^3\,dx=\frac{3}{4}.$

  $E(X^2)=\int_0^1 x^2\cdot f(x)\,dx=\int_0^1 x^2\cdot 3x^2\,dx=\int_0^1 3x^4\,dx=\frac{3}{5}$
- $V(X)=E(X^2)-E(X)^2=\frac{3}{5}-\left(\frac{3}{4}\right)^2=\frac{3}{80}$

균일분포

❖ 균일분포 **(Uniform Distribution)**
: 확률변수 X가 어느 구간 (a, b)에서 정의되고, 확률밀도함수의 크기가 동일한 확률분포

- 예) 약속시간에 늦는 시간이 특별한 경향이 없이 0분과 60분 사이에 랜덤하게 골고루 퍼져 있을 때의 늦는 시간 X

❖ 균일분포의 확률밀도함수

$$f(x)=\frac{1}{b-a},\qquad a\le x\le b.$$

- $a\le c\le d\le b$를 만족하는 c 와 d 에 대하여

$$P(c\le X\le d)=\int_c^d f(x)\,dx=\frac{1}{b-a}\int_c^d 1\,dx=\frac{d-c}{b-a}.$$

균일분포

❖ 균일분포의 누적확률함수

$$F(x)=\int_{-\infty}^{x} f(x)\,dx=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 1 & b<x\end{cases}$$

❖ 균일분포의 평균과 분산

$$E(X)=\int_a^b x\cdot f(x)dx=\int_a^b x\cdot\frac{1}{b-a}\,dx=\frac{a+b}{2}$$

$$E(X^2)=\int_a^b x^2\cdot f(x)\,dx=\int_a^b x^2\cdot\frac{1}{b-a}\,dx=\frac{a^2+ab+b^2}{3}$$

$$\Rightarrow V(X)=E(X^2)-E(X)^2=\frac{(b-a)^2}{12}$$

균일분포

❖ 약속시간에 늦는 시간이 특별한 경향이 없이 0분과 60분 사이에 랜덤하게 골고루 퍼져 있을 때의 늦는 시간 X는 균일분포를 따름

❖ $a = 0$, $b = 60$

❖ 확률밀도함수 $f(x) = \frac{1}{b-a} = \frac{1}{60}, \quad 0 \le x \le 60.$

❖ 누적확률함수

$$F(x) == \begin{cases} 0 & x < 0 \\ \frac{x-a}{b-a} = \frac{x}{60} & 0 \le x \le 60 \\ 1 & 60 < x \end{cases}$$

❖ 평균과 분산

$$E(X) = \frac{a+b}{2} = \frac{0+60}{2} = 30 \qquad V(X) = \frac{(b-a)^2}{12} = \frac{(60-0)^2}{12} = 300$$

정규분포

❖ 정규분포 **(Normal Distribution)**

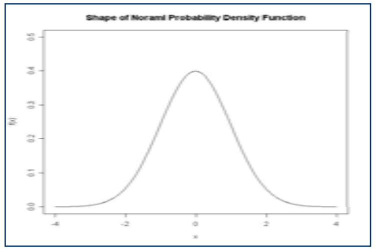
: 많은 자료들이 정규분포를 하며**,** 분포식이나 의미 해석에 있어 많은 좋은 특성이 있어**,** 여러 분야에서 중요히 널리 쓰임

- 확률밀도함수

평균이 μ 이고 분산이 σ^2 인 확률변수 X 의 정규분포

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

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



종형곡선

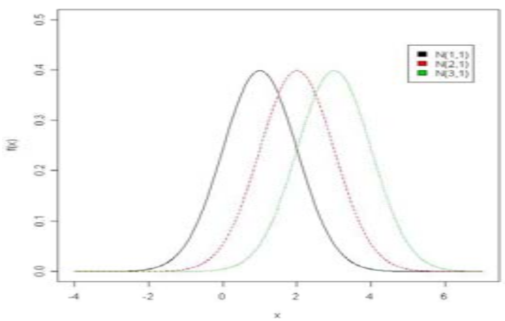
정규분포

❖ 정규분포의 특성

① 정규분포의 확률밀도함수는 평균 에서 가장 큰 값을 가짐

⇒ μ 는 평균이며 또한 최빈값

② 분산은 같고 평균이 서로 다른 정규분포의 정규 곡선들

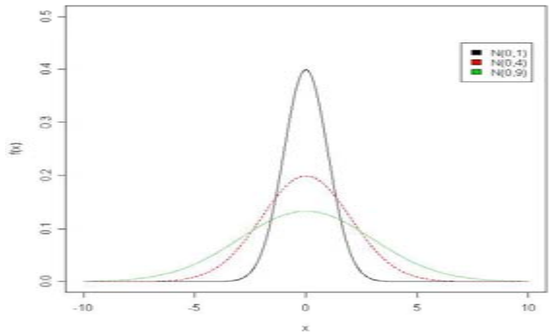


정규분포

❖ 정규분포의 특성

③ 정규곡선은 평균 μ에 대하여 대칭

④ 평균은 같고 분산이 서로 다른 정규분포의 정규 곡선들



정규분포

❖ 정규분포의 특성

⑤ $X \sim N(\mu, \sigma^2)$일 때, 확률변수 X가 구간 (a, b)의 값을 가질 확률은 정규곡선의 (a, b) 구간의 아래 부분의 영역의 넓이에 해당

⇒ 정규곡선 아래 전체의 넓이는 1

⇒ 대칭성에 의하여 X가 μ보다 크거나 작을 확률은 $\frac{1}{2}$

정규분포

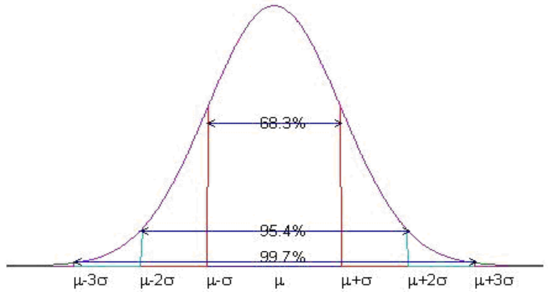
❖ 정규분포의 특성

⑥ 자주 사용되는 구간의 확률들

X가 구간 $(\mu-\sigma, \mu+\sigma)$ 에 포함되는 확률은 68.3%

$(\mu-2\sigma, \mu+2\sigma)$ 에 포함되는 확률은 95.4%

$(\mu-3\sigma, \mu+3\sigma)$ 에 포함되는 확률은 99.7%



표준정규분포

❖ 표준정규분포**(Standard Normal Distribution)**

: 평균이 **0** 분산이 **1**인 정규분포

- 표기

Z : 표준정규분포 따르는 변수

$\phi(z)$: 확률밀도함수

$\Phi(z)$: 누적확률밀도함수

❖ 표준 정규분포의 확률밀도함수

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

표준정규분포

❖ 정규분포의 확률 계산

⇒ 일반적인 공식이 존재하지 않아, 정규분포와 표준정규분포 사이의 관계를 정의하고, 컴퓨터로 계산된 표준 정규분포의 누적 확률분포 함수 값을 사용

❖ 표준정규분포의 확률

⇒ 확률변수 Z가 구간 (a, b)에 포함될 확률은 (a, b)에 해당되는 구간의 확률밀도 함수 곡선의 아래 영역

$$P(a \le Z \le b) = P(Z \le b) - P(Z \le a) = \Phi(b) - \Phi(a)$$

⇒ 표준정규분포의 누적확률함수로 계산

표준정규분포

❖ 표준정규 확률분포표

$P(0 \le Z \le z)$

TABLE 1 Normal Curve Areas

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.10	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.20	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.30	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.40	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.50	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.60	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.70	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.80	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.90	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.00	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.10	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.20	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.30	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.40	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.50	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.60	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545

표준정규분포

❖ 표준정규분포의 확률 계산

- 예) Z가 0과 1.21 사이에 포함될 확률

z	.00	.01	.02
⋮		⋮	
.9	.3159	.3186	.3212
1.0	**.3413**	.3438	.3461
1.1	.3643	.3665	.3686
1.2	**.3849**	.3869	.3888
⋮		⋮	

$$P(0 \le Z \le 1.21) = 0.3869$$

- 예) Z가 구간 -1.00과 0.00사이에 있을 확률

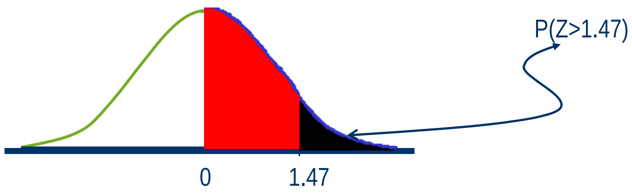
$$P(-1.00 \le Z \le 0) = P(0 \le Z \le 1.00) = 0.3413$$

⇒ 표준정규확률밀도함수는 0을 중심으로 대칭

표준정규분포

❖ 표준정규분포의 확률 계산

- 예) Z가 1.47 이상일 확률

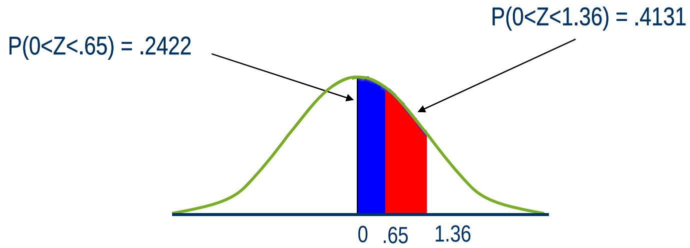


P(Z>1.47) = 0.5 - 0.4292 = 0.0708

표준정규분포

❖ 표준정규분포의 확률 계산

- 예) Z가 0.65와 1.36 사이에 있을 확률



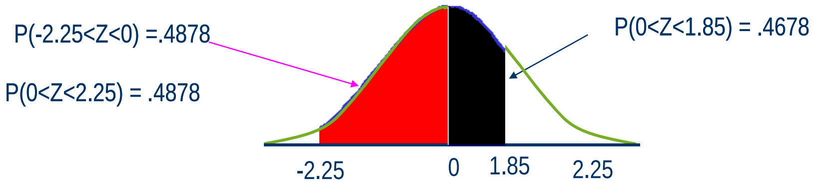
P(.65<Z<1.36) = .4131 - .2422 = .1709

표준정규분포

❖ 표준정규분포의 확률 계산

- 예) Z가 구간 **-2.25**과 **1.85** 사이에 있을 확률

$$P(-2.25 \le Z \le 1.85) = P(-2.25 \le Z \le 0.00) + P(0.00 \le Z \le 1.85)$$
$$= 0.4878 + 0.4678 = 0.9556$$



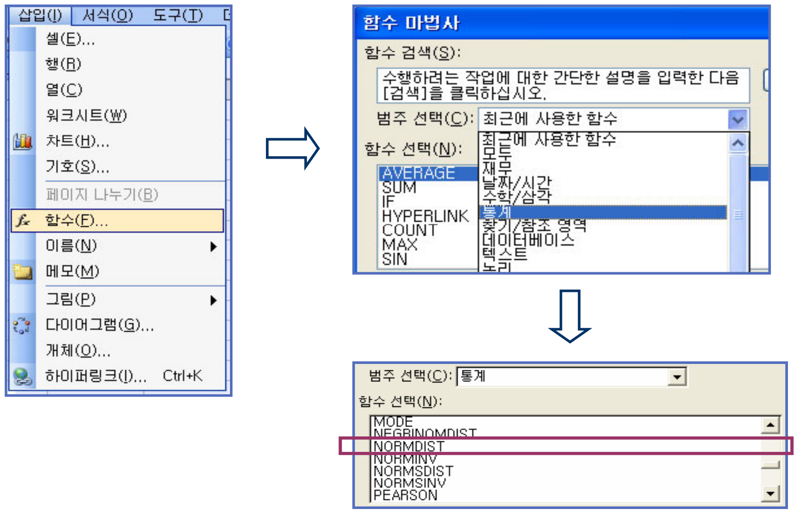
표준정규분포

- 예) Z가 구간 **-1.21**보다 작을 확률

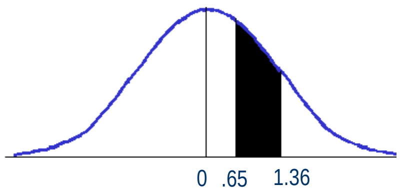
  ⇒ 표준정규확률밀도함수는 0을 중심으로 대칭

$$P(Z \ge 1.21) = 1 - P(Z \le 1.21)$$
$$= 1 - [P(Z \le 0.00) + P(0.00 \le Z \le 1.21)]$$
$$= 1 - 0.5 - 0.3869 = 1 - 0.8869 = 0.1131.$$

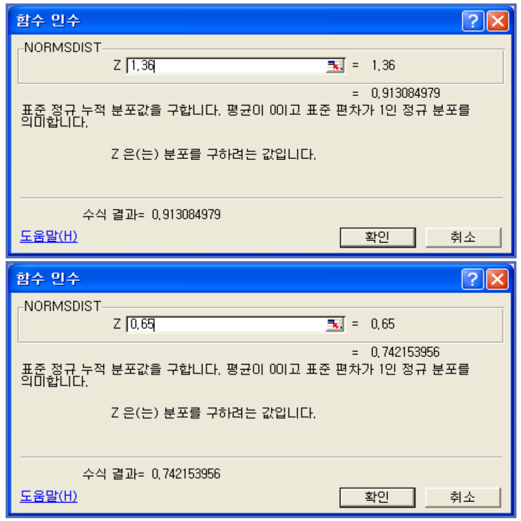
표준정규분포 Excel 연습



표준정규분포 Excel 연습



P(.65<Z<1.36)
= .9131 - .7422 = .1709



정규분포의 확률 계산

❖ 정규분포의 확률 계산

- 정규분포를 따르는 확률변수 $X \sim N(\mu, \sigma^2)$ 의 선형변환인 $Y = a + bX$ 도 정규분포를 따르고 평균은 $a + b\mu$ 이고, 분산은 $b^2\sigma^2$가 됨
- 확률변수 $X \sim N(\mu, \sigma^2)$라면, $Z = \dfrac{X - \mu}{\sigma}$ 은 표준정규분포

따라서 $\Rightarrow$ $P(a \le X \le b) = P\left(\dfrac{a-\mu}{\sigma} \le \dfrac{X-\mu}{\sigma} \le \dfrac{b-\mu}{\sigma}\right)$

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

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

정규분포의 확률 계산

❖ (예제 6.8) **1kg**인 추를 반복적으로 잴 때, 측정된 무게는 평균이 **1kg**이고, 표준편차가 **20**그램인 정규분포를 따른다.측정된 값이 **1kg**으로부터 **10**그램 이내에 있는 비율?

X : 측정된 추의 무게

$P(990 \le X \le 1010)$

$$= P\left(\frac{990-1000}{20} \le \frac{X-\mu}{\sigma} \le \frac{1010-1000}{20}\right)$$

$$= P(-0.5 \le Z \le 0.5) = 2 \times 0.1915 = 38.29\%$$

정규분포의 백분율

❖ 정규분포의 백분율

확률 변수 **X**가 평균이 μ , 분산이 σ^2 인 정규분포를 따를 때 이의 100(1- α)% 백분율 η_α는 $P(X \le \eta_\alpha) = 1 - \alpha$ 의 해이다.

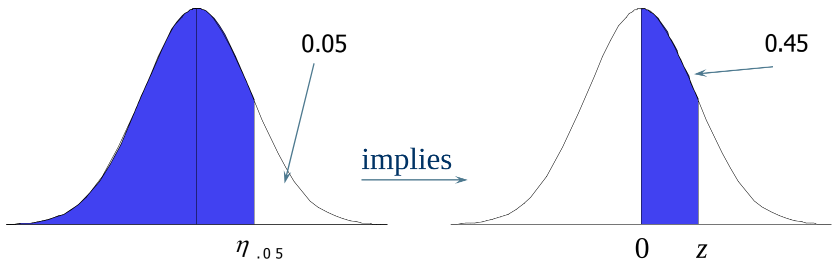
$$P(X \le \eta_\alpha) = P\left(\frac{X-\mu}{\sigma} \le \frac{\eta_\alpha - \mu}{\sigma}\right) = P\left(Z \le \frac{\eta_\alpha - \mu}{\sigma}\right) = 1 - \alpha$$

따라서, 표준정규분포의 100(1- α)% 백분율을 Z_α 로 표시하면, $\dfrac{\eta_\alpha - \mu}{\sigma} = Z_\alpha$ 이므로,

$$\eta_\alpha = \mu + \sigma \times Z_\alpha$$

표준정규분포의 백분율은 표준 정규분포표를 역으로 읽음으로써 구할 수 있다.

표준정규분포의 백분율

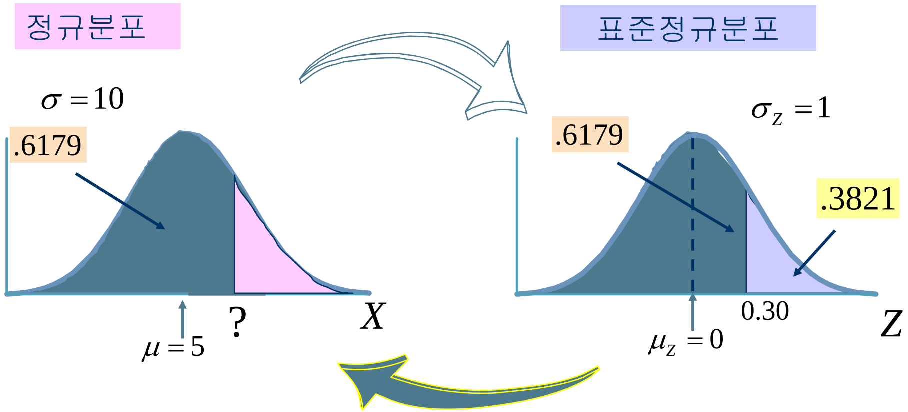


z	0.00	0.01	0.02	0.03	0.04		0.05
1.6					0.4495	0.4500	0.4505

표준정규분포의 95% 백분율 = 1.645

정규분포의 백분율

표준정규분포의 61.79% 백분율



$$\eta_{.3821} = \mu + z\sigma = 5 + (.30)(10) = 8$$

표준정규분포의 백분율

❖ 표준 정규분포의 백분율 구하는 연습

90%의 백분율을 계산하여 보면,

$P(Z \le 1.28) = 0.8997$ 이고 $P(Z \le 1.29) = 0.9015$ 이므로

90%의 백분율은 1.28과 1.29사이의 값이다.

0.9는 1.28의 누적확률인 0.8997에 보다 가까우므로

1.28을 취하든지, 선형근사 보간법에 의하여

1.28+0.01* (0.9-0.8997)/(0.9015-0.8997)=1.2817 로 계산.

z	.07	.08	.09
⋮		⋮	
1.1	.3790	.3810	.3830
1.2	.3980	**.3997**	**.4015**
1.3	.4147	.4162	.4177
1.4	.4292	.4306	.4319
⋮		⋮	

정규분포의 백분율

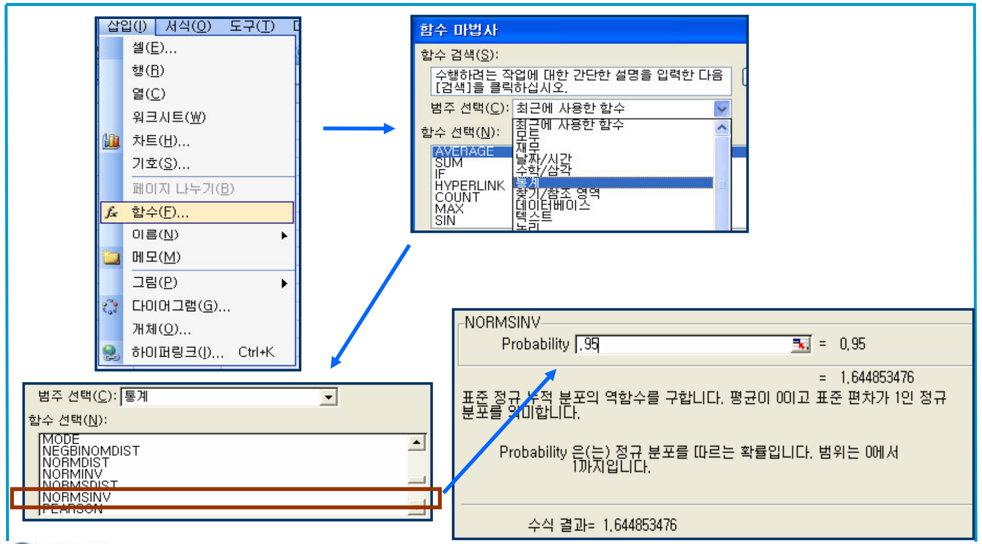
❖ **(예제 6.9)** 우체국에서 소포 무게의 상한선을 설정하고자 하여 기존 고객이 부치는 짐의 무거운 **5%** 정도를 제한하고자 한다. 만약 기존 고객의 소포 무게의 분포가 평균 **5kg,** 표준편차 **1kg**인 정규분포를 따른다면 상한선은 얼마로 해야 할까?

⇒ 무게의 상한선은 평균5kg, 표준편차 1kg인 정규분포의 95% 백분율이 됨

⇒ 5+1*(표준정규분포의 95% 백분율)

= 5+1*1.645=6.645

표준정규분포의 백분율 Excel 연습



이항분포의 정규근사

❖ 이항분포

$X \sim B(n, p)$ 을 따르는 확률변수에서

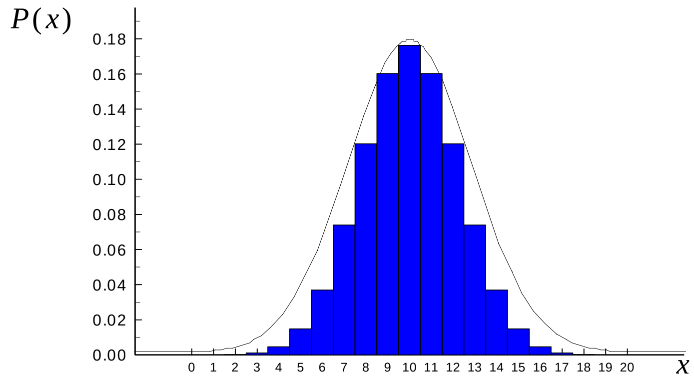
$P(a \le X \le b) = \sum_{k=a}^{k=b} \binom{n}{k} p^k (1-p)^{n-k}$ 을 계산하고자 할 때,

- n 이 작은 경우
  : 이항분포표를 이용, 혹은 직접 계산

- n 이 매우 크거나, 확률의 정확한 값을 알 필요가 없을 때
  : 정규분포를 이용한 근사 계산
  : np≥5와 n(1-p)≥5 의 조건이 필요
  예) p=0.5 이면, n≥10
  p=0.01 또는 p=0.99 이면, n≥500

이항분포의 정규근사

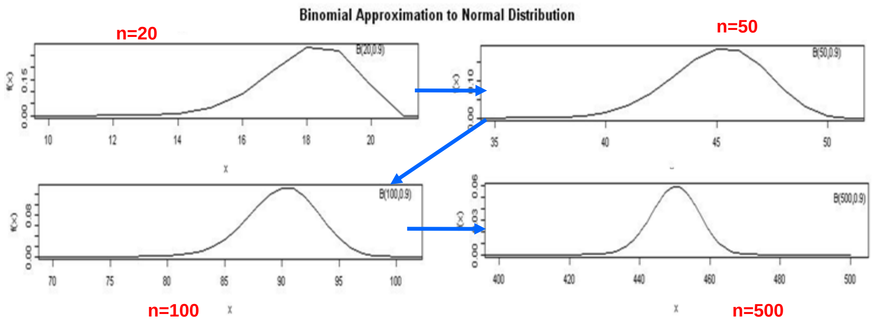
***n* = 20, *p* = 0.5** 인 이항분포와 정규근사



이항분포의 정규근사

❖ 이항분포의 정규근사

⇒ ***n*** 이 증가함에 따라 이항분포의 확률질량함수는 종모양의 정규분포에 가까워짐



이항분포의 정규근사

❖ 이항분포의 정규근사

$\mu = np, \ \sigma = \sqrt{np(1-p)}$ 일 때

$$P(a \le X \le b) \approx \Phi\left(\frac{b+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right)$$

⇒ 0.5는 연속성 정정계수: 이산형 분포를 연속형 분포로 이용하면서 생기는 오차를 보정함

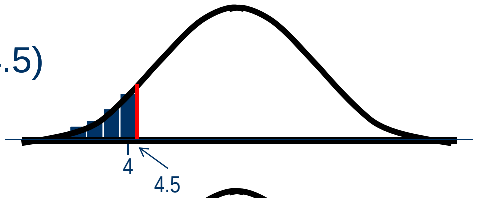
$$P(a \le X < b) \approx \Phi\left(\frac{b-0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right)$$

$$P(a < X \le b) \approx \Phi\left(\frac{b+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a+0.5-\mu}{\sigma}\right)$$

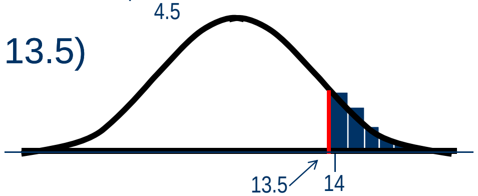
$$P(a < X < b) \approx \Phi\left(\frac{b-0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a+0.5-\mu}{\sigma}\right)$$

이항분포의 정규근사

$P(X \le 4) \cong P(Y < 4.5)$



$P(X \ge 14) \cong P(Y > 13.5)$



이항분포의 정규근사

❖ **(예제 6.10)**

한국 성인의 **50%**가 정기적으로 알코올 음료를 마시는 것으로 조사되었다. **1000**명의 임의 추출된 표본에서 알코올 성분의 음료를 정기적으로 마시는 사람의 수가 **480**명 보다 작을 확률은 얼마인가?

(풀이) X(1000명중 알코올 음료를 정기적으로 마시는 사람의 수)라고 할 때, X는 B(1000,0.5)를 따른다. $np \ge 5$와 $n(1-p) \ge 5$ 의 조건을 만족하므로 이항분포의 정규 근사를 이용하면,

$$P(X < 480) = P(X \le 479.5) = P\left(\frac{X-500}{\sqrt{250}} \le \frac{479.5-500}{\sqrt{250}}\right)$$

$$\approx P(Z \le -1.30) = 0.0968$$

지수분포

❖ **지수분포(Exponential Distribution)**

- 랜덤 시간의 분포를 설명하기 위한 분포 중 하나

  예) 수술 후 환자가 완전히 회복되는데 걸리는 시간

  어느 방사능 원소가 분해될 때까지 걸리는 시간

  공장에서 특별한 조립 과정에서 소요되는 시간

  전구의 수명
- 포아송분포(단위시간당 발생 건수)와 역의 관계

❖ 지수분포의 확률밀도함수

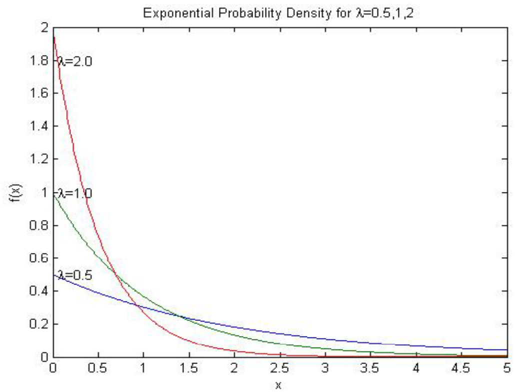
$$f(x) = \lambda \exp(-\lambda x), \qquad x \ge 0$$

X : 발생률이 λ 인 지수분포,

또는 평균이 $1/\lambda$인 지수분포

지수분포

❖ 지수분포의 확률밀도 함수



- x=0에서 최대값을 갖고 x값이 커짐에 따라 기하적으로 감소

지수분포

❖ 지수분포의 확률

X ~ 발생률이 λ인 지수분포를 따르는 확률변수

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

$$= \int_a^b \lambda \cdot \exp(-\lambda x)\,dx = \exp(-\lambda a) - \exp(-\lambda b)$$

$$P(X \ge a) = \exp(-\lambda a)$$
$$P(X \le a) = 1 - \exp(-\lambda a)$$

지수분포

❖ 지수분포의 평균과 분산

$$E(X) = \frac{1}{\lambda}, \qquad V(X) = \frac{1}{\lambda^2}$$

❖ 지수분포의 무기억성(**memoryless property**)

양의 실수 s 와 t 에 대하여,

$$\Rightarrow \quad P(X > t + s \mid X > s) = P(X > t)$$

지수분포

❖ **(예제 6.11)**

트랜지스터의 수명은 평균 **100**시간인 지수분포를 따른다. 라디오 트랜지스터를 새것으로 교환했을 때, 이것의 수명이 **50**시간도 되지 않을 확률은?

평균이 100이므로, 발생률은 $\lambda = 1/100 = 0.01$이므로

$$P(X < 50) = 1 - \exp\left(-50 \cdot \frac{1}{100}\right) = 1 - \exp(-0.5)$$

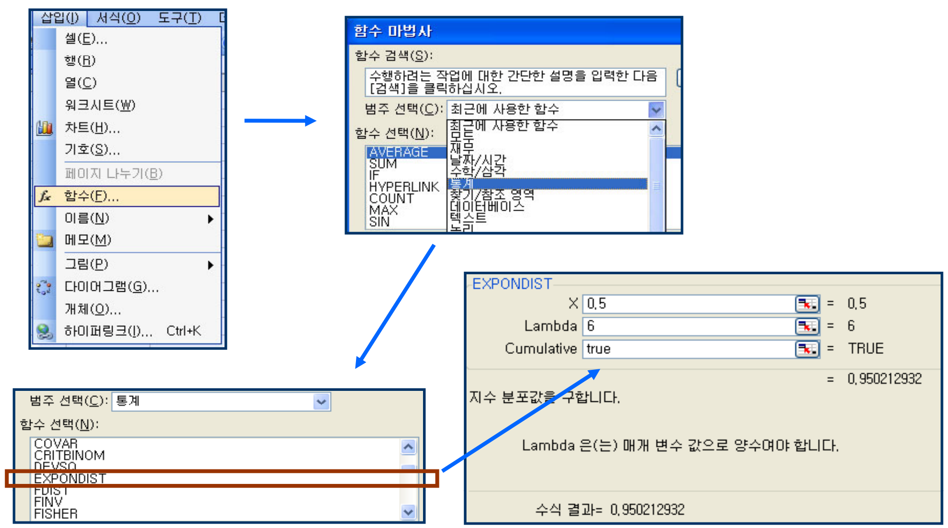
❖ **(예제 6.12)**

발생률이 λ인 지수분포의 $p \times 100th$ 백분율은?

$P(X \le \eta_p) = 1 - \exp(-\lambda \cdot \eta_p) = p$의 해 η_p는 $p \times 100\%$의 해

$$\Rightarrow \quad \eta_p = -(1/\lambda)\log(1 - p)$$

지수분포 Excel 연습



Excel 연습

❖ (예제 6.6)
Step 1. 다음의 값을 입력한다.

	A	B	C	D
1	=NORMSDIST(-0.155)+(1-NORMSDIST(1.6))			

❖ Step 2. 위의 값을 계산하면 다음의 결과를 얻는다.

	A
1	0.49321

Excel 연습

❖ (예제 6.7)
Step 1. 다음의 값을 입력한다.

	A	B
1	=NORMSINV(0.025)	

❖ Step 2. 위의 값을 계산하면 다음의 결과를 얻는다.

	A
1	-1.95996

Excel 연습

❖ (예제 6.8)
Step 1. 다음의 값을 입력한다.

	A	B	C	D	E	F
1	=NORMDIST(1010,1000,20,TRUE)-NORMDIST(990,1000,20,TRUE)					

❖ Step 2. 위의 값을 계산하면 다음의 결과를 얻는다.

	A
1	0.382925