

3. Random Variable and Probability Distribution

3.1 Random Variable

3.2 Probability Distribution of Discrete Random Variable

3.3 Probability Distribution of Continuous Random Variable

3.4 Binomial Distribution

3.5 Normal Distribution

Probability calculation for certain types of events is widely used in our daily lives. Let's discuss what is a random variable and how to calculate the probability of binomial distribution and normal distribution which are particularly commonly used.

3.1 Random Variable

☞ Think	The sample space that can appear by tossing a coin twice is as follows: $S = \{TT, TH, HT, HH\}$
Exploration	Which probabilities of various events in the sample space is used a lot in real problem?

- If there are only two possible outcomes in a trial, it is often observed around us to count the number of one outcome when the trial is repeated several times as follows:
 - Select 10 products produced in a factory and inspect whether they are ‘good’ or ‘defective’ and count the number of defectives out of 10 products.
 - Survey 10 voters whether they like or dislike a particular candidate in an election and count the number of voters who like the particular candidate.
- Probabilities can be obtained for various events in the sample space, but what we are more interested in is the events for the number of heads and their probabilities such as tossing a coin two times as follows/

Event of no head	{TT}	$P(\{TT\}) = \frac{1}{4}$
Event of one head	{TH, HT}	$P(\{TH, HT\}) = \frac{2}{4}$
Event of two heads	{HH}	$P(\{HH\}) = \frac{1}{4}$

- If X is the number of heads of the coin, the possible number of X is 0, 1, and 2. In other words, X is one-to-one correspondence from each element in the sample space S to one number as follows:

Sample space	X = Number of heads when tossing a coin two times
{TT}	0
{TH}	1
{HT}	1
{HH}	2

- The one-to-one relation in which a single real number is matched to each element in the sample space is called a **random variable**. Random variables are usually denoted with uppercase letters X, Y, Z, \dots and values of the random variables are denoted with lowercase letters x, y, z, \dots . When a random variable has finite values or countable numbers as a natural number, the random variable is called a **discrete random variable**, and when it has an arbitrary real value, the random variable is called a **continuous random variable**. For example, the number of heads that appear by tossing two coins is a discrete random variable, and the delivery time to home when we order a pizza is a continuous random variable because it can have any positive real value.

Example 3.1	<p>In the next experiments, distinguish whether the random variable X is a discrete random variable or a continuous random variable.</p> <ol style="list-style-type: none"> 1) The number of car accidents X that occur per day at an intersection. 2) The height of our classmates X.
Solution	<ol style="list-style-type: none"> 1) The number of car accidents occurring at an intersection can be 0, 1, 2, 3, ... Since it is a natural number, the number of car accidents occurring at an intersection X is a discrete random variable. 2) The height of our class students usually has an arbitrary real value within a certain range, so the height X is a continuous random variable.

Exercise 3.1	<p>In the next experiments, distinguish whether the random variable X is a discrete random variable or a continuous random variable.</p> <ol style="list-style-type: none"> 1) Number of typhoons X passing through Korea in a year 2) Annual precipitation X in Seoul 3) The number of goals X in a soccer game 4) Height X of the first-year male high school students
--------------	---

3.2 Probability Distribution of Discrete Random Variable

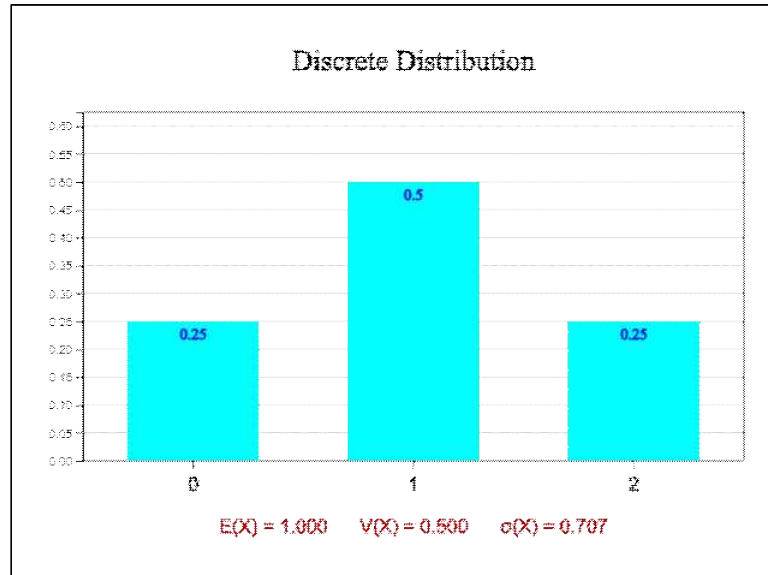
👁 Think	When a coin is tossed twice, if X is the number of heads, the possible values of X is 0, 1, and 2.
Exploration	Let's arrange all the probabilities for the possible numbers 0, 1, and 2 of X in a table?

- If the random variable X is the number of heads when a coin is tossed twice, the probability of event which has no head, $P(\{TT\}) = \frac{1}{4}$, can be written as $P(X=0) = \frac{1}{4}$. Similarly, the probability of events which has one head, $P(\{TH, HT\}) = \frac{2}{4}$, as $P(X=1) = \frac{2}{4}$, and the probability of event which has two heads, $P(\{HH\}) = \frac{1}{4}$, as $P(X=2) = \frac{1}{4}$. These probabilities can be summarized in a table as follows:

Sample space	$X = \text{Number of heads}$	$P(X=x)$
{TT}	$x = 0$	$\frac{1}{4}$
{TH} {HT}	$x = 1$	$\frac{2}{4}$
{HH}	$x = 2$	$\frac{1}{4}$

- The **probability distribution** of a discrete random variable X is a summary

of probabilities that X has a certain value x_i $P(X=x_i)$, as shown in this table. The above probability distribution can be expressed graphically as in <Figure 3.1>.



<Figure 3.1> Probability distribution of a discrete random variable

- In general, if a discrete random variable X can have values x_1, x_2, \dots, x_n and their corresponding probabilities are p_1, p_2, \dots, p_n respectively, it can be summarized as one-to-one relation as follows and it is called a **probability distribution** of X

$P(X=x_1)$	p_1
$P(X=x_2)$	p_2
...	...
$P(X=x_n)$	p_n

This probability distribution can also be represented in a row as follows:

X	x_1	x_2	...	x_n	total
$P(X=x_i)$	p_1	p_2	...	p_n	1

It may also be represented as a mathematical function and it is called a **probability density function** of X

$$P(X=x_i) = p_i, \quad i=1, 2, \dots, n$$

- Among discrete random variables, the most widely used in reality is the binomial distribution, which will be studied in detail in Section 3.4.
- According to the basic properties of probability, the probability distribution function has the following properties.

☞ Properties of probability distribution function

The probability distribution function $P(X=x_i) = p_i$ ($i=1,2,\dots,n$) of a discrete random variable X has the following properties.

- 1) $0 \leq p_i \leq 1$
- 2) $p_1 + p_2 + \dots + p_n = 1$

- The probability that a random variable X has a value between a and b is represented as $P(a \leq X \leq b)$.

Example 3.2	<p>There are five products in one box, two of which are defective. We select two products out of this box and inspect them. Let X be the number of defective items selected.</p> <p>1) Find the probability distribution of X and draw a graph using 『eStatH』 .</p> <p>2) Find $P(1 \leq X \leq 2)$.</p>										
Solution	<p>1) X can have values 0, 1, 2 and their probabilities are as follows:</p> $P(X=0) = \frac{{}_3C_2 \times {}_2C_0}{{}_5C_2} = \frac{3}{10}.$ $P(X=1) = \frac{{}_3C_1 \times {}_2C_1}{{}_5C_2} = \frac{6}{10}.$ $P(X=2) = \frac{{}_3C_0 \times {}_2C_2}{{}_5C_2} = \frac{1}{10}.$ <p>Therefore, the probability distribution of X is represented in a table as follows:</p> <table border="1" style="margin: 10px auto; border-collapse: collapse; text-align: center;"> <tr> <td style="padding: 5px;">X</td> <td style="padding: 5px;">0</td> <td style="padding: 5px;">1</td> <td style="padding: 5px;">2</td> <td style="padding: 5px;">Total</td> </tr> <tr> <td style="padding: 5px;">$P(X=x)$</td> <td style="padding: 5px;">$\frac{3}{10}$</td> <td style="padding: 5px;">$\frac{6}{10}$</td> <td style="padding: 5px;">$\frac{1}{10}$</td> <td style="padding: 5px;">1</td> </tr> </table> <p>If you select 'Discrete Distribution' from the 『eStatH』 menu, data input window such as in <Figure 3.2> appears. Enter the data here as shown in the figure and click [Execute] button to display the probability distribution graph as shown in <Figure 3.3>.</p>	X	0	1	2	Total	$P(X=x)$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	1
X	0	1	2	Total							
$P(X=x)$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	1							

Example 3.2
Solution 계속

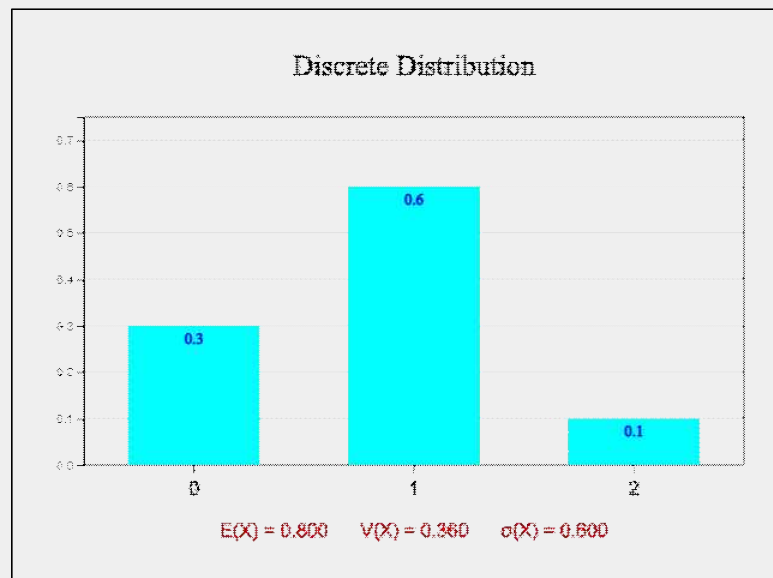


Discrete Distribution Menu

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	Total
X	0	1	2							
$P(X = x_i)$	0.3	0.6	0.1							1.000

Execute Mean $E(X) = 0.800$ Variance $V(X) = 0.360$ Std Dev $\sigma(X) = 0.600$

<Figure 3.2> Data input for a discrete distribution



<Figure 3.3> Graph of a discrete distribution using 『eStatH』

$$\begin{aligned}
 2) \quad P(1 \leq X \leq 2) &= P(X=1) + P(X=2) \\
 &= \frac{6}{10} + \frac{1}{10} = \frac{7}{10}
 \end{aligned}$$

Exercise 3.2



Let a random variable X be the number of heads when a coin is tossed three times. Find the probability distribution of X and draw a graph.

A. Mean and standard deviation of a discrete random variable

👁 Think	If $X =$ 'the number of heads' when a coin tossed twice, X can have values 0, 1, and 2 and their probabilities are $\frac{1}{4}$, $\frac{2}{4}$, and $\frac{1}{4}$ respectively.
Exploration	1) What would be a good representative value of X ? 2) What would be a good measure of dispersion for possible values of X ?

- The probability distribution of the random variable $X =$ 'the number of heads' when a coin tossed twice is as follows:

X	0	1	2	Total
$P(X=x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

- This probability distribution is similar to the frequency table of three possible values 0, 1, and 2 with the relative frequency.

values	frequency	relative frequency
0	1	$\frac{1}{4}$
1	2	$\frac{2}{4}$
2	1	$\frac{1}{4}$
Total	4	1

The mean μ of these numbers 0, 1, 1, 2 and the variance σ^2 are as follows:

$$\text{Mean } \mu = \frac{0+1+1+2}{4} = 0 \times \frac{1}{4} + 1 \times \frac{2}{4} + 2 \times \frac{1}{4} = 1$$

$$\begin{aligned} \text{Variance } \sigma^2 &= \frac{(0-1)^2 + (1-1)^2 + (1-1)^2 + (2-1)^2}{4} \\ &= (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{2}{4} + (2-1)^2 \times \frac{1}{4} = \frac{2}{4} = 0.5 \end{aligned}$$

Therefore, the standard deviation σ is $\sqrt{0.5} = 0.707$.

- In general, if a discrete random variable X can have values as x_1, x_2, \dots, x_n

and their probabilities are p_1, p_2, \dots, p_n respectively, then the probability distribution is as follows:

X	x_1	x_2	\dots	x_n	Total
$P(X=x)$	p_1	p_2	\dots	p_n	1

The mean or expectation of the random variable X denoted as $E\{X}$ or μ , the variance denoted as $V\{X}$ or σ^2 and the standard deviation $\alpha\{X}$ or σ are calculated as follows

$$E\{X\} = \mu = \sum_{i=1}^n x_i p_i$$

$$V\{X\} = \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 p_i$$

$$\alpha\{X\} = \sigma = \sqrt{V\{X\}}$$

- The mean or expectation of the random variable X , $E\{X\}$, which can have possible values x_1, x_2, \dots, x_n with their probabilities p_1, p_2, \dots, p_n can be considered as a weighted average of values using the weight of probabilities. The variance of X , $V\{X\}$, is a weighted average of all squared distances from the mean. The standard deviation of X , $\alpha\{X\}$, is an average distance from the mean.

Example 3.3	<p>Find the mean, variance and standard deviation of the probability distribution in [Example 3.2]. Check the result using 『eStatH』 .</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>X</td> <td>0</td> <td>1</td> <td>2</td> <td>Total</td> </tr> <tr> <td>$P(X=x)$</td> <td>$\frac{3}{10}$</td> <td>$\frac{6}{10}$</td> <td>$\frac{1}{10}$</td> <td>1</td> </tr> </table>	X	0	1	2	Total	$P(X=x)$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	1
X	0	1	2	Total							
$P(X=x)$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	1							
Solution	<p>The mean, variance and standard deviation of the random variable X are as follows:</p> $E\{X\} = 0 \times \frac{3}{10} + 1 \times \frac{6}{10} + 2 \times \frac{1}{10} = \frac{8}{10} = 0.8$ $V\{X\} = (0-0.8)^2 \times \frac{3}{10} + (1-0.8)^2 \times \frac{6}{10} + (2-0.8)^2 \times \frac{1}{10} = 0.36$ $\alpha\{X\} = \sqrt{0.36} = 0.60 \text{ 이다}$ <p>The mean, variance, and standard deviation are consistent with the results of 『eStatH』 in <Figure 3.2>.</p>										

Exercise 3.3

When a coin is tossed three times, find the probability distribution by setting the random variable as the number of heads. Find the mean, variance, and standard deviation of this probability distribution.

- If we expand the variance of a discrete random variable, $V(X) = \sum_{i=1}^n (x_i - \mu)^2 p_i$, it becomes as follows:

$$\begin{aligned}
 V(X) &= (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \cdots + (x_n - \mu)^2 p_n \\
 &= (x_1^2 p_1 - 2\mu x_1 p_1 + \mu^2 p_1) + (x_2^2 p_2 - 2\mu x_2 p_2 + \mu^2 p_2) + \cdots + (x_n^2 p_n - 2\mu x_n p_n + \mu^2 p_n) \\
 &= (x_1^2 p_1 + x_2^2 p_2 + \cdots + x_n^2 p_n) - 2\mu(x_1 p_1 + x_2 p_2 + \cdots + x_n p_n) + \mu^2(p_1 + p_2 + \cdots + p_n) \\
 &= E(X^2) - 2\mu^2 + \mu^2 \\
 &= E(X^2) - \mu^2 \\
 &= E(X^2) - E(X)^2
 \end{aligned}$$

Therefore, the variance of a random variable X can be calculated using a short cut formula as follows:

$$V(X) = E(X^2) - E(X)^2$$



- Using this short cut formula, the variance of [Example 3.3] can be calculated as follows:

$$V(X) = 0^2 \times \frac{3}{10} + 1^2 \times \frac{6}{10} + 2^2 \times \frac{1}{10} - (0.8)^2 = 0.36$$

Short cut formula of the variance of a discrete random variable X

If a discrete random variable X can have values x_1, x_2, \dots, x_n , their probabilities are p_1, p_2, \dots, p_n respectively and $E(X) = \mu$, then the variance of X can be calculated as follows.

$$\begin{aligned}
 V(X) &= E(X^2) - E(X)^2 \\
 &= (x_1^2 p_1 + x_2^2 p_2 + \cdots + x_n^2 p_n) - \mu^2
 \end{aligned}$$

<p>Example 3.4</p>	<p>When a dice is rolled, assume that the number of dots appeared is a random variable X. Find the mean, variance, and standard deviation of X. Draw a graph of the probability distribution using 『eStatH』 and check these values.</p>																																																	
<p>Solution</p>	<p>The possible values of the random variable X are 1, 2, 3, 4, 5, 6 and the probability distribution of X is as follows:</p> <table border="1" data-bbox="614 510 1324 611"> <thead> <tr> <th>X</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>$P(X=x)$</td> <td>$\frac{1}{6}$</td> <td>$\frac{1}{6}$</td> <td>$\frac{1}{6}$</td> <td>$\frac{1}{6}$</td> <td>$\frac{1}{6}$</td> <td>$\frac{1}{6}$</td> <td>1</td> </tr> </tbody> </table> <p>The mean, variance using a short cut formula and standard deviation are as follows:</p> $E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$ $E(X^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = \frac{91}{6}$ $V(X) = E(X^2) - E(X)^2 = \frac{91}{6} - (3.5)^2 = 2.916$ $\sigma(X) = \sqrt{2.916} = 1.708$ <p>If you select 'Discrete Distribution' from 『eStatH』 menu and enter data as shown in <Figure 3.4>, the graph as shown in <Figure 3.5> and the mean, variance and standard deviation appear. A slightly different value originated from the computing error for the number of decimal places.</p> <div data-bbox="343 1232 518 1411" style="text-align: center;">  </div> <div data-bbox="606 1191 1332 1393" style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;">Discrete Distribution Menu</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th>x_1</th> <th>x_2</th> <th>x_3</th> <th>x_4</th> <th>x_5</th> <th>x_6</th> <th>x_7</th> <th>x_8</th> <th>x_9</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>$P(X=x_i)$</td> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> <td>0.1667</td> <td></td> <td></td> <td></td> <td>1.000</td> </tr> </tbody> </table> <p style="text-align: center;"> <input type="button" value="Execute"/> Mean $E(X)$ = 3.501 Variance $V(X)$ = 2.915 Std Dev $\sigma(X)$ = 1.707 </p> </div> <p style="text-align: center;"><Figure 3.4> Data input for a discrete distribution</p> <div data-bbox="614 1444 1324 1971" style="border: 1px solid black; padding: 10px;"> <p style="text-align: center;">Discrete Distribution</p>  <p style="text-align: center;"> $E(X) = 3.501$ $V(X) = 2.915$ $\sigma(X) = 1.707$ </p> </div> <p style="text-align: center;"><Figure 3.5> Discrete distribution with mean, variance and standard deviation</p>	X	1	2	3	4	5	6	Total	$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	Total	X	1	2	3	4	5	6					$P(X=x_i)$	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667				1.000
X	1	2	3	4	5	6	Total																																											
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1																																											
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	Total																																								
X	1	2	3	4	5	6																																												
$P(X=x_i)$	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667				1.000																																								

Exercise 3.4



When a coin is tossed three times, find the probability distribution by setting the random variable X as the number of heads. Use 「eStatH」 to find the mean, variance, standard deviation and the graph of this probability distribution.

B. Mean and standard deviation of $aX+b$

☞ Think	When a coin tossed twice, the random variable $X =$ ‘number of heads’ can have values 0, 1, 2 and their probabilities are $\frac{1}{4}$, $\frac{2}{4}$, $\frac{1}{4}$ respectively.
Exploration	1) If 100 dollars is paid for each head, how much value can be expected in this game and how is the variance of the prize money? 2) If 50 dollars is basically paid and 100 dollars is paid for each head, how much value can be expected in this game and how is the variance of the prize money?

- If $X =$ ‘the number of heads’ when a coin is tossed twice, the above game is the same as the problem to examine the probability distribution of 1) $Y = 100X$ and 2) $Z = 50 + 100X$. The probability distribution of the random variable X is as follows and $E(X) = 1$, $V(X) = 0.5$, $\sigma(X) = 0.707$.

X	0	1	2	Total
$P(X=x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

- 1) The probability distribution of $Y = 100X$ is as follows:

$Y = 100X$	0	100	200	Total
$P(Y=y)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

Therefore, the expectation $E(Y)$, variance $V(Y)$ and standard deviation $\sigma(Y)$ are as follows:

$$\begin{aligned}
E(Y) &= 0 \times \frac{1}{4} + 100 \times \frac{2}{4} + 200 \times \frac{1}{4} = 100 \\
V(Y) &= (0-100)^2 \times \frac{1}{4} + (100-100)^2 \times \frac{2}{4} + (200-100)^2 \times \frac{1}{4} \\
&= \frac{20000}{4} = 5000 \\
\sigma(Y) &= \sqrt{5000} = 70.7
\end{aligned}$$

That is, $E(Y) = 100E(X)$, $V(Y) = 100^2 V(X)$, $\sigma(Y) = 100\sigma(X)$.

2) The probability distribution of $Z = 50 + 100X$ is as follows:

$Z = 50 + 100X$	50	150	250	Total
$P(Z=z)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	1

Therefore, the expectation $E(Z)$, variance $V(Z)$ and standard deviation $\sigma(Z)$ are as follows:

$$\begin{aligned}
E(Z) &= 50 \times \frac{1}{4} + 150 \times \frac{2}{4} + 250 \times \frac{1}{4} = 150 \\
V(Z) &= (50-150)^2 \times \frac{1}{4} + (150-150)^2 \times \frac{2}{4} + (250-150)^2 \times \frac{1}{4} = \frac{20000}{4} \\
&= 5000 \\
\sigma(Z) &= \sqrt{5000} = 70.7
\end{aligned}$$

That is, $E(Z) = 50 + 100E(X)$, $V(Z) = 100^2 V(X)$, $\sigma(Z) = 100\sigma(X)$.

- We noticed that if 50 is added to X the variance of Z is not affected, but if X is multiplied by 100, the variance of Z is increased by 100^2 .
- In general, assume that the probability distribution of a discrete random variable X is as follows and its mean, variance and standard deviation are $E(X)=\mu$, $V(X)=\sigma^2$, $\sigma(X)=\sigma$ respectively.

X	x_1	x_2	...	x_n	Total
$P(X=x)$	p_1	p_2	...	p_n	1

The probability distribution of $Y = aX + b$ is as follows:

$Y = aX + b$	$y_1 = ax_1 + b$	$y_2 = ax_2 + b$...	$y_n = ax_n + b$	Total
$P(Y=y)$	p_1	p_2	...	p_n	1

Therefore, the expectation $E(Y)$, variance $V(Y)$ and standard deviation $\sigma(Y)$ are as follows:

$$\begin{aligned}
 EY &= (ax_1+b)p_1+(ax_2+b)p_2+\dots+(ax_n+b)p_n \\
 &= a(x_1p_1+x_2p_2+\dots+x_np_n)+b(p_1+p_2+\dots+p_n) \\
 &= aEX+b = a\mu+b \\
 VY &= [(ax_1+b)-(a\mu+b)]^2p_1+[(ax_2+b)-(a\mu+b)]^2p_2+\dots+[(ax_n+b)-(a\mu+b)]^2p_n \\
 &= a^2(x_1-\mu)^2p_1+a^2(x_2-\mu)^2p_2+\dots+a^2(x_n-\mu)^2p_n \\
 &= a^2VX = a^2\sigma^2 \\
 \alpha Y &= \sqrt{VY} = \sqrt{a^2\sigma^2} = |a|\sigma
 \end{aligned}$$

- If the mean, variance, standard deviation of a random variable X are $EX=\mu$, $VX=\sigma^2$, $\alpha X=\sigma$ respectively, then the mean, variance and standard deviation of $Y= aX+b$ are as follows:

$$\begin{aligned}
 EY &= aEX+b = a\mu+b \\
 VY &= a^2 VX = a^2\sigma^2 \\
 \alpha Y &= |a|\alpha X = |a|\sigma
 \end{aligned}$$

This is true even if X is a continuous random variable.

☞ Mean, variance and standard deviation of $Y= aX+b$

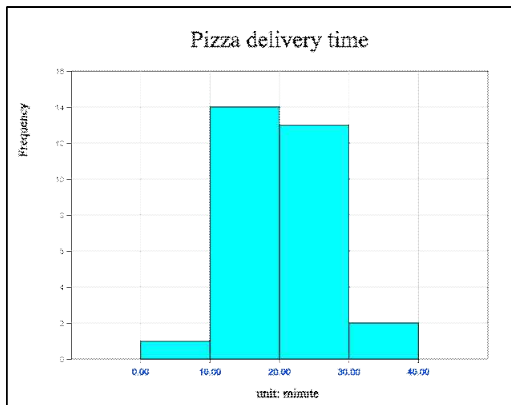
If X is a random variable and a, b are constants, the expectation variance and standard deviation of $Y=aX+b$ are as follows:

$$\begin{aligned}
 EY &= aEX+b = a\mu+b \\
 VY &= a^2 VX = a^2\sigma^2 \\
 \alpha Y &= |a|\alpha X = |a|\sigma
 \end{aligned}$$

3.3 Probability Distribution of Continuous Random Variable

☞ Think	<p>The delivery time to the house by ordering a pizza at a shop were examined 30 times is as follows:</p> <p>(Data 3.1) Pizza delivery time to the house (unit: minute)</p> <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;"> <p>15 27 37 9 16 26 17 22 30 23 13 19 22 14 16 18 12 22 16 19 26 11 24 20 21 19 22 11 25 27</p> </div>
Exploration	<ol style="list-style-type: none"> 1) What is the probability that the pizza will be delivered between 10 and 20 minutes? 2) What is the probability that the pizza will be delivered between 15 and 20 minutes? 3) How to obtain the probability that pizza is delivered to a more detailed time interval?

- Using 『Histogram - Frequency Table』 of 『eStatH』, the histogram of pizza delivery time is as <Figure 3.6> if the class interval starts at 0 and the interval size is 10 minutes. It can be seen that the probability that the pizza is delivered between 10 and 20 minutes is 0.47.

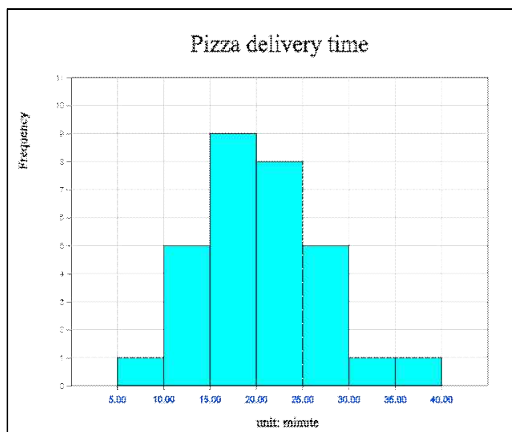


<Figure 3.6> Histogram of the pizza delivery times - interval size is 10 minutes

class	frequency	relative frequency
$0.00 \leq x < 10.00$	1	0.03
$10.00 \leq x < 20.00$	14	0.47
$20.00 \leq x < 30.00$	13	0.43
$30.00 \leq x < 40.00$	2	0.07
Total	30	1.00

[Table 3.1] Frequency table of the pizza delivery times - interval size is 10 minutes

- If the interval of the histogram starts at 5 and the interval size is 5 minutes, the histogram becomes as <Figure 3.7>. It can be seen that the probability in which the pizza will be delivered between 15 and 20 minutes is 0.30.

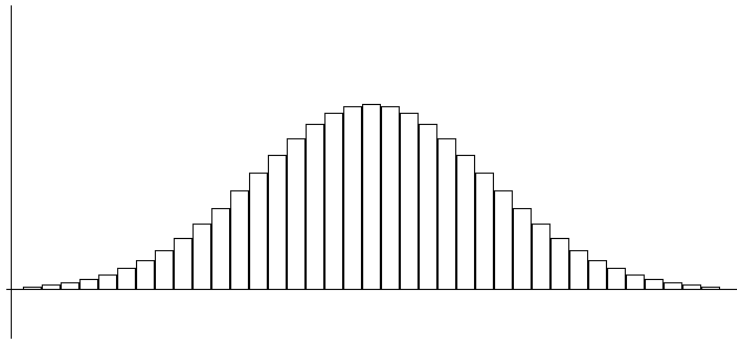


<Figure 3.7> Histogram of the pizza delivery times - interval size is 5 minutes

class	frequency	relative frequency
$5.00 \leq x < 10.00$	1	0.03
$10.00 \leq x < 15.00$	5	0.17
$15.00 \leq x < 20.00$	9	0.30
$20.00 \leq x < 25.00$	8	0.27
$25.00 \leq x < 30.00$	5	0.17
$30.00 \leq x < 35.00$	1	0.03
$35.00 \leq x < 40.00$	1	0.03
Total	30	1.00

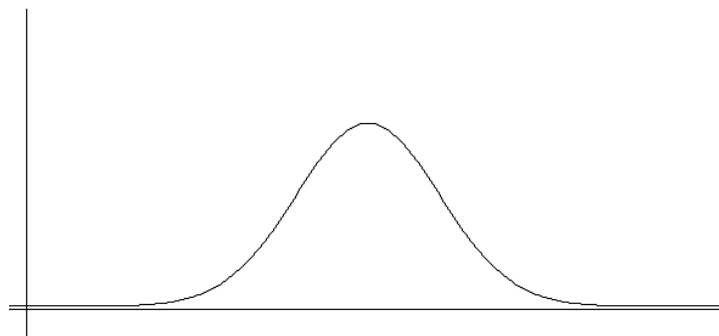
[Table 3.2] Frequency table of the pizza delivery times - interval size is 5 minutes

- In order to obtain the probability of a more detailed interval, a more detailed histogram such as <Figure 3.8> and its frequency table are required as shown below. However, for such a histogram, more data must be collected, and it is cumbersome to redraw the histogram every time to calculate the desired probability.



<Figure 3.8> Histogram with detailed interval

- There are many continuous data around us as the histogram above in which many data are located around the mean, data are symmetrical about the mean, and their shape looks like a bell. In order to easily find the probability for all of this type data, mathematicians searched for a function that can describe these kinds of histograms. This function has the same shape as in <Figure 3.9> and is called a **normal distribution function**. It is explained in detail in Section 3.5.



<Figure 3.9> Normal distribution function

- If the probability distribution function of a continuous random variable can be expressed as a mathematical function $f(x)$, it is possible to approximate the desired probability without finding a frequency table and histogram. In general, the probability distribution function $f(x)$ of a continuous random variable X has the following properties.

☞ **Properties of the probability distribution function $f(x)$ of a continuous random variable X**

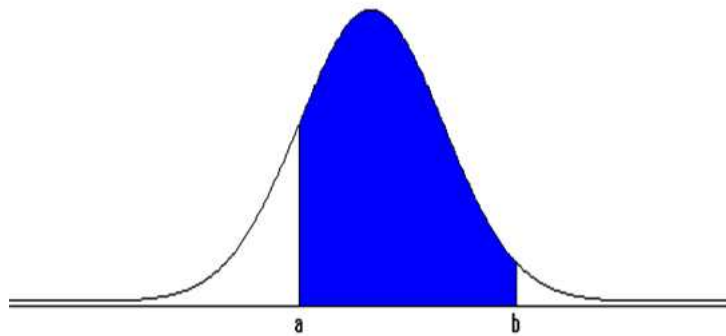
- 1) $f(x) \geq 0$
- 2) The sum of the total probabilities should equal to 1.

That is, if the domain of the function $f(x)$ is an interval $[\alpha, \beta]$, the area of the function $f(x)$ on the domain should be 1 as follows:

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x) dx = 1$$

- 3) The probability that a continuous random variable X is in the interval $[a, b]$, $P(a \leq X \leq b)$, can be found as the definite integral on the interval $[a, b]$ which is the area of $f(x)$. (<Figure 3.10>).

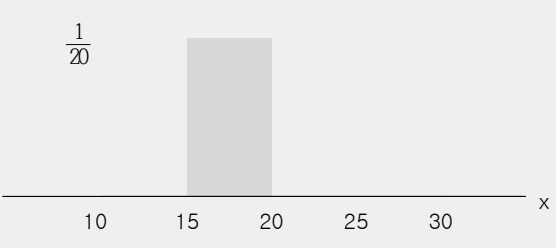
$$P(a \leq X \leq b) = \int_a^b f(x) dx$$



<Figure 3.10> The probability that a continuous random variable X is in the interval $[a, b]$

Example 3.5

Suppose the delivery time from ordering a pizza to its arrival home can be any time between 10 and 30 minutes. If X is the random variable of 'delivery time to home', find the probability distribution function of X and draw its graph. Find the probability that the delivery time is in between 15 and 20 minutes.

<p>Example 3.5 Solution</p>	<p>Since X can be any number between 10 and 30, the probability distribution function is as follows:</p> $f(x) = \begin{cases} \frac{1}{(30-10)}, & 10 \leq x \leq 30 \\ 0, & \text{기타} \end{cases}$ <p>It is called a uniform distribution and denoted as Uniform(10,30).</p> <p>The probability of the delivery time in between 15 and 20 minutes is plotted as <Figure 3.11>, so the calculation of the probability (area) is $(20-15) \times (1/20) = 0.25$.</p>  <p style="text-align: center;"><Figure 3.11> Calculation of $P(15 < X < 20)$ in Uniform(10,30)</p>
------------------------------------	--

<p>Exercise 3.5</p>	<p>If the probability distribution function of a continuous random variable X is $f(x) = 2x(0 \leq x \leq 1)$, find the followings.</p> <ol style="list-style-type: none"> 1) $P(0 \leq X \leq \frac{1}{4})$ 2) $P(\frac{1}{4} \leq X \leq 1)$
----------------------------	--

<p>Exercise 3.6</p>	<p>If the probability distribution function of a continuous random variable X is $f(x) = ax(0 \leq x \leq 2)$, find the followings. ($a$ is a constant)</p> <ol style="list-style-type: none"> 1) Constant a 2) $P(0 \leq X \leq \frac{1}{4})$ 3) $P(\frac{1}{4} \leq X \leq 1)$
----------------------------	---

3.4 Binomial Distribution

☞ Think	It is said that the defect rate of products produced in one factory is 0.1. Consider a random variable which is the number of defective products when three products are inspected.
Exploration	What is the probability distribution function of the random variable X ?

- When a product is inspected, if it is a good product, we denote it as G, and if it is a defective product, we denote it as B. When three products are inspected, possible events of the sample space, possible values of the random variable X which is the number of defective products and their probabilities are as follows:

Sample space	$X = \text{number of defectives}$	$P(X=x)$
GGG	0	$(0.9)^3$
GGB GBG BGG	1	$3(0.1)(0.9)^2$
GGB BGB BBG	2	$3(0.1)^2(0.9)$
BBB	3	$(0.1)^3$

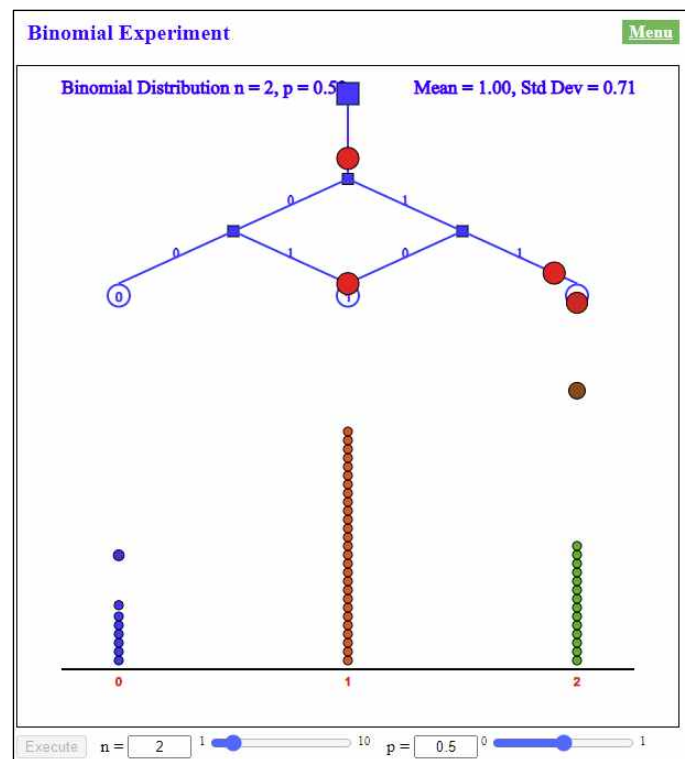
When the number of defectives is 0, the number of cases is one which is from ${}_3C_0$. When the number of defectives is 1, the number of cases is three which is from ${}_3C_1$. When the number of defectives is 2, the number of cases is three which is from ${}_3C_2$. When the number of defectives is 3, the number of cases is 1 which is from ${}_3C_3$. Therefore the probability distribution function of the random variable X can be expressed as follows:

$$P(X=x) = {}_3C_x(0.1)^x(1-0.1)^{3-x} \quad (x=0,1,2,3)$$

- Similar examples as the above problem, which examines the number of defective products by inspecting the product, are observed frequently around us.

- Toss a coin five times to count the number of heads.
- Count the number of voters in favor of a particular candidate in an election.

Another example can be found at the following instrument as <Figure 3.12> in science museums. This instrument drops a ball from the top and, if the ball hit a bar, it may fall to the left (0 point) or to the right (1 point) with a half chance. Then the ball drops to the next stage bar again and may fall to the left and right with a half chance. We examine the distribution of scores after the second stage when 100 balls are dropped.



<Figure 3.12> Simulation of binomial distribution

- What these examples have in common is that there are two possible outcomes of one trial such as {good, defective}, {head, tail}, {approval, opposition}, etc. and that this trial is repeated several times. However, the probability of outcomes in each trial is different. The trial with two possible outcomes ('binary') is called the Bernoulli trial, and the outcome of interest among the two is often called 'success' and the rest is called 'failure'. When the Bernoulli trial with probability of 'success' p is repeated n times independently, if X is the random variable of counting the number of 'success', then X can have values $0, 1, 2, \dots, n$ and the probability distribution function of X is as follows:

$$P(X=x) = {}_n C_x p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \dots, n)$$

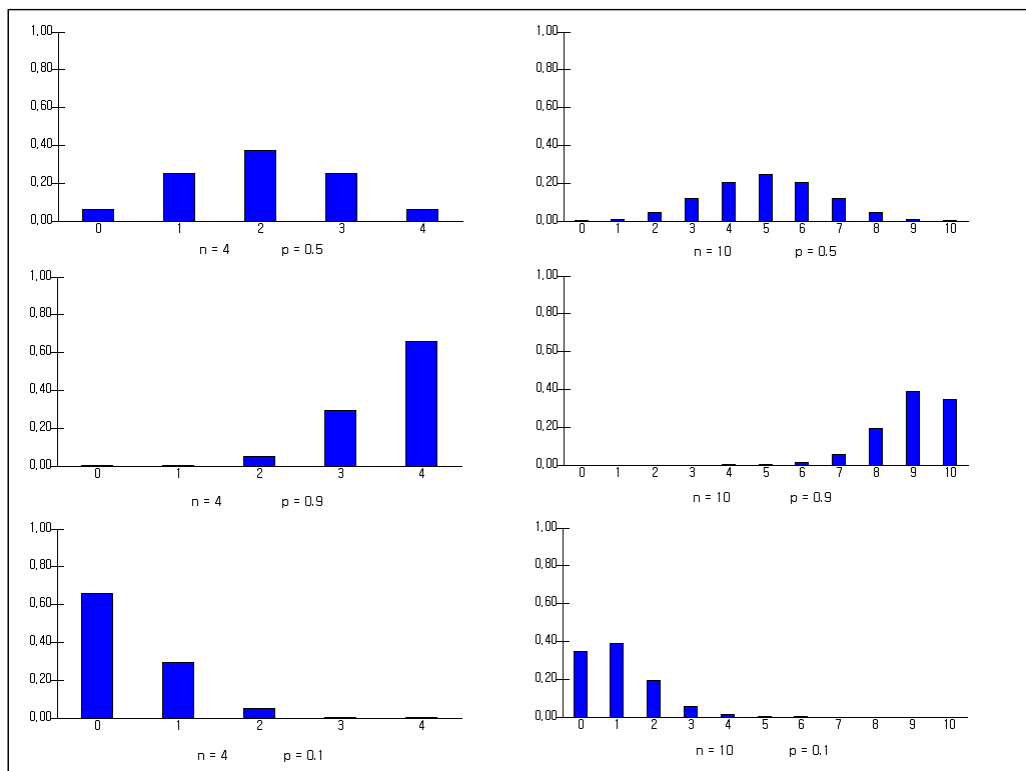
- This probability distribution is called a binomial distribution denoted as a symbol $B(n, p)$ and, if we denote $q=1-p$, the binomial distribution can be expressed as the following table.

X	0	1	2	...	x	...	n	total
$P(X=x)$	${}_n C_0 q^n$	${}_n C_1 p q^{n-1}$	${}_n C_2 p^2 q^{n-2}$...	${}_n C_x p^x q^{n-x}$...	${}_n C_n p^n$	1

Each probability of this binomial distribution is equal to each coefficient of the terms in the expansion of the equation $(q+p)^n$ using the binomial theorem.

$$(q+p)^n = {}_n C_0 q^n + {}_n C_1 p^1 q^{n-1} + {}_n C_2 p^2 q^{n-2} + \dots + {}_n C_x p^x q^{n-x} + \dots + {}_n C_n p^n$$

- The following is a graph of the binomial distribution for various values of n and p .




<Figure 3.13> Binomial distribution for various values of n and p .

Binomial distribution

When the Bernoulli trial with probability of 'success' p is repeated n times independently, if X is the random variable of counting the number of 'success', then X can have values $0, 1, 2, \dots, n$ and the probability distribution function of X is a binomial distribution $B(n, p)$ as follows:

$$f(x) = {}_n C_x p^x (1-p)^{n-x}, \quad (x=0,1,2,\dots,n)$$

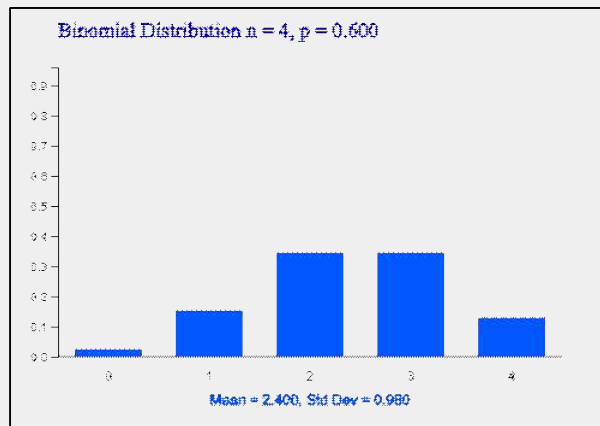
<p>Example 3.6</p>	<p>In this year baseball season, there are four more games for the team "Tiger" to play with the team "Bear". It is known that 'Tiger' team has a 60% chance of winning every game.</p> <p>1) Let the random variable X be the number of games in which 'Tiger' wins among 4 games. Suppose that there is no draw game. Find the probability distribution of X.</p> <p>2) Use 『eStat』 to find this probability distribution.</p>
<p>Solution</p> 	<p>Each game can be considered as a Bernoulli trial of 'Win' or 'Loss' and the trial is repeated four times. If 'Win' is denoted by O and 'Loss' is denoted by X, then there are $2^4 = 16$ possible elements in the sample space as follows:</p> $S = \{ 'XXXX', 'OXXX', 'XOXX', 'XXOX', 'XXXO', 'OOXX', 'OXOX', 'OXXO', 'XOOX', 'XOXO', 'XXOO', 'OOOX', 'OOXO', 'OXOO', 'XOOO', 'XXXX' \}$ <p>The probability that 'Tiger' will lose four times, $\{ 'XXXX' \}$, is $(0.4) \times (0.4) \times (0.4) \times (0.4) = (0.4)^4$.</p> <p>The probability that 'Tiger' will lose three times and win one time is $(0.6) \times (0.4) \times (0.4) \times (0.4)$. There are four cases of winning one time, $\{ 'OXXX', 'XOXX', 'XXOX', 'XXXO' \}$ which are the same number of cases in which O is seated in one when there are four seats. So, the probability that 'Tiger' will win once is ${}_4 C_1$. Therefore, the total probability of winning one game is as follows:</p> ${}_4 C_1 (0.6)(0.4)^3$ <p>The probability that 'Tiger' will lose two times and win also two times is $(0.6) \times (0.6) \times (0.4) \times (0.4)$. There are six cases of winning two times, $\{ 'OOXX', 'OXOX', 'OXXO', 'XOOX', 'XOXO', 'XXOO' \}$ which are ${}_4 C_2$. So the probability of winning two games is as follows:</p> ${}_4 C_2 (0.6)^2 (0.4)^2$ <p>Similarly, the probability of winning three games is as follows:</p> ${}_4 C_3 (0.6)^3 (0.4)$ <p>The probability of winning four games is as follows:</p> $(0.6) \times (0.6) \times (0.6) \times (0.6)$

**Example 3.6
Solution
(Continued)**

If we summarize, the probability distribution of the random variable X which is the number of games in which ‘Tiger’ wins is as follows:

X	0	1	2	3	4	Total
$P(X=x)$	${}_4C_0(0.4)$	${}_4C_1(0.6)(0.4)$	${}_4C_2(0.6)^2(0.4)$	${}_4C_3(0.6)^3(0.4)$	${}_4C_4(0.6)$	1

Select ‘Binomial Distribution’ from the menu of 『eStatH』 and enter $n = 4$, $p = 0.6$. If you click [Execute] button, a binomial distribution graph such as <Figure 3.14> appears.



<Figure 3.14> Binomial distribution, $n = 4$, $p = 0.6$

There are sliding bars of n , p and a probability calculation box under the graph, so put the desired value and press the key to calculate the value.

$n =$ $p =$

Show Probability Normal Approx

Probability $P($ $\leq X \leq$ $) =$

$P(X \leq$ $) =$

$P(X \geq$ $) =$

<Figure 3.15> Binomial probability calculation by providing intervals

To the right of the graph, a table of binomial distributions is shown. In addition to $P(X = x)$, this table shows the cumulative probabilities $P(X \leq x)$ and $P(X \geq x)$ to facilitate various probability calculations. If you select a new n and p and click [Execute] button, new binomial distribution table for this value is added below.

$n = 4$	$p = 0.600$		
x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
0	0.0256	0.0256	1.0000
1	0.1536	0.1792	0.9744
2	0.3456	0.5248	0.8208
3	0.3456	0.8704	0.4752
4	0.1296	1.0000	0.1296

Exercise 3.7



The defect rate of an electronic component produced in one factory is 5%. Use «eStatH» to find the following probabilities when there is a box containing 50 of these components.

- 1) What's the probability that there is no defective product?
- 2) What is the probability that there are 1, 2 or 3 defective products?
- 3) What is the probability of having more than three (\geq) defective products?

A. Mean and standard deviation of binomial distribution

- If a Bernoulli trial with the probability of 'success' p and of 'failure' $q=1-p$ is repeated three times, the random variable X which counts the number of 'success' is the following binomial distribution.

X	0	1	2	3	Total
$P(X=x)$	q^3	$3pq^2$	$3p^2q$	p^3	1

The random variable X has the mean $E(X)$, the variance $V(X)$ and the standard deviation $\sigma(X)$ as follows:

$$\begin{aligned}
 E(X) &= 0 \times q^3 + 1 \times 3pq^2 + 2 \times 3p^2q + 3 \times p^3 = 3p \\
 V(X) &= (0-3p)^2 \times q^3 + (1-3p)^2 \times 3pq^2 + (2-3p)^2 \times 3p^2q + (3-3p)^2 \times p^3 \\
 &= 3pq \\
 \sigma(X) &= \sqrt{3pq}
 \end{aligned}$$

- In general, if a random variable X follows the binomial distribution, $B(n,p)$, then its probability distribution is as follows:

X	0	1	2	...	x	...	n	Total
$P(X=x)$	${}_nC_0q^n$	${}_nC_1pq^{n-1}$	${}_nC_2p^2q^{n-2}$...	${}_nC_xp^xq^{n-x}$...	${}_nC_np^n$	1

The mean of X becomes $E(X) = np$, the variance becomes $V(X) = npq$, and the standard deviation becomes $\sigma(X) = \sqrt{npq}$.


☞ Mean, variance, standard deviation of binomial distribution

If a random variable X follows binomial $B(n, p)$,

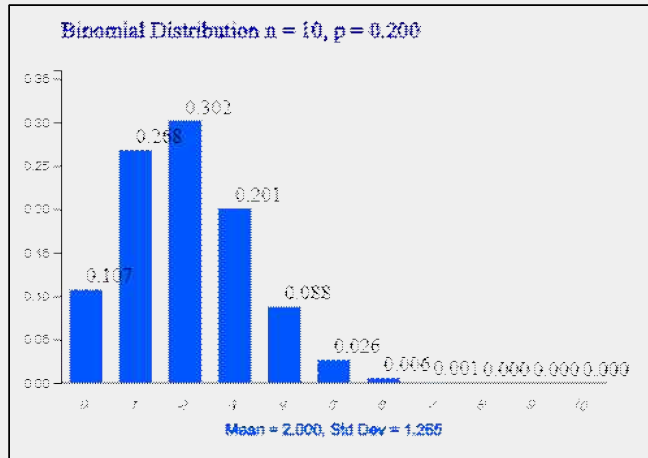
$$E(X) = np,$$

$$V(X) = npq,$$

$$\sigma(X) = \sqrt{npq} \quad (\text{단, } q=1-p)$$

<p>Example 3.7</p>	<p>The probability that a salesperson of an insurance company causes a customer to subscribe its insurance is 20% from past experience. A salesperson is scheduled to meet 10 customers this morning. Calculate the following probabilities and use 『eStatH』 to verify them.</p> <ol style="list-style-type: none"> 1) What is the probability that three customers will subscribe the insurance? 2) What is the probability that two or more people (\geq) will subscribe the insurance? 3) On average, how many people will subscribe the insurance? What is its standard deviation?
<p>Solution</p> 	<p>Since it is a binomial distribution with $n = 10$ and $p = 0.2$,</p> <ol style="list-style-type: none"> 1) Probability that three customers will subscribe the insurance is as follows: $P(X=3) = {}_{10}C_3(0.2)^3(1-0.2)^{10-3} = 0.2013$ 2) It is better to use the probability of complementary event to calculate the probability that two or more people (\geq) will subscribe the insurance as follows: $P(X \geq 2) = 1 - P(X=0) - P(X=1)$ $= 1 - {}_{10}C_0(0.2)^0(1-0.2)^{10} - {}_{10}C_1(0.2)^1(1-0.2)^{10-1}$ $= 1 - 0.1074 - 0.2684 = 0.6242$ 3) It is as follows: $q = 1-p$ $E(X) = np = 10 \times 0.2 = 2$ $V(X) = npq = 10 \times 0.2 \times 0.8 = 1.6$ $\sigma(X) = \sqrt{1.6} = 1.265$ <p>Select 'Binomial Distribution' in the menu of 『eStatH』 and select $n = 10$, $p = 0.2$. If you click [Execute] button, the graph as in <Figure 3.16> appears. You can check the probability of 1) if you check the 'probability' here.</p>

Example 3.7
Solution 계속



<Figure 3.16> Binomial distribution when $n = 10, p = 0.2$ using 『eStatH』

You can check the answer of 2) $P(X \geq 2) = 0.6242$ from the binomial table on the right hand side as follows:

n = 10	p = 0.200		
x	P(X = x)	P(X ≤ x)	P(X ≥ x)
0	0.1074	0.1074	1.0000
1	0.2684	0.3758	0.8926
2	0.3020	0.6778	0.6242
3	0.2013	0.8791	0.3222
4	0.0881	0.9672	0.1209
5	0.0264	0.9936	0.0328
6	0.0055	0.9991	0.0064
7	0.0008	0.9999	0.0009
8	0.0001	1.0000	0.0001
9	0.0000	1.0000	0.0000
10	0.0000	1.0000	0.0000

Exercise 3.8



In a presidential election, the approval rating of a particular candidate was 40%. In order to establish an election strategy, we will meet 20 voters this morning to see if they support the particular candidate. Calculate the following probabilities directly and check them using 『eStatH』. Assume that no person abstained.

- 1) What is the probability that 8 out of 20 will support the particular candidate?
- 2) What is the probability that 5 or more (\geq) and 12 or less (\leq) will support?
- 3) On average, how many people will support the candidate? What's the standard deviation?

- If the value of n is large, it is not easy to calculate the probability of the binomial distribution even using a calculator or computer. In the 『eStatH』, you can find the binomial probability when $n \leq 100$. If n is more than 100, the probability can be approximated using a normal distribution with mean np and variance $np(1-p)$. See details in Section 3.5.

B. Law of large number

👁 Think	Consider an experiment which rolls a dice n times in which the mathematical probability of each side is $\frac{1}{6}$. If X is a random variable which counts the number of single dot on the upper side, X follows the binomial distribution $B(n, \frac{1}{6})$.
Exploration	When you roll a dice n times, let's compare the relative frequency $\frac{X}{n}$ with the mathematical probability $\frac{1}{6}$.

- Using the binomial distribution, it can be seen that the mathematical probability of $\frac{1}{6}$ is a reasonable model. When you roll a dice n times, the probability that the difference between the relative frequency, $\frac{X}{n}$, and the mathematical probability, $\frac{1}{6}$, is less than a small value ε can be written as follows:

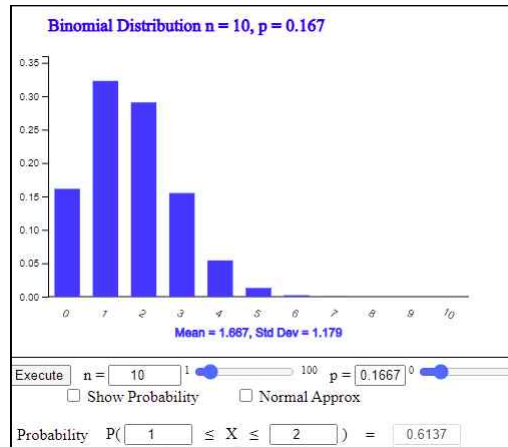
$$\begin{aligned} P\left(\left|\frac{X}{n} - \frac{1}{6}\right| < \varepsilon\right) &= P\left(-\varepsilon < \frac{X}{n} - \frac{1}{6} < \varepsilon\right) \\ &= P\left(\frac{1}{6} - \varepsilon < \frac{X}{n} < \frac{1}{6} + \varepsilon\right) \\ &= P\left(n\left(\frac{1}{6} - \varepsilon\right) < X < n\left(\frac{1}{6} + \varepsilon\right)\right) \end{aligned}$$

If $\varepsilon = 0.1$, the above formula becomes as follows:

$$P\left(\left|\frac{X}{n} - \frac{1}{6}\right| < 0.1\right) = P\left(n\left(\frac{1}{6} - 0.1\right) < X < n\left(\frac{1}{6} + 0.1\right)\right)$$

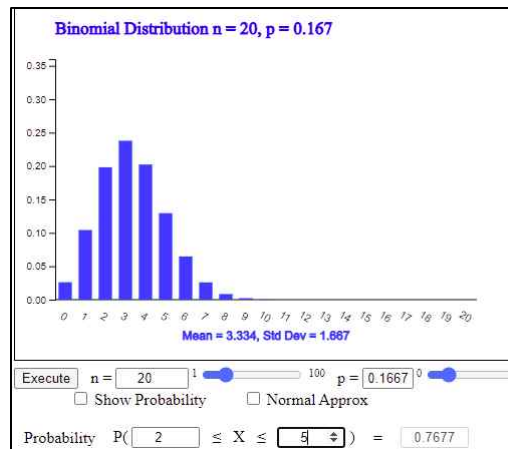
If the number of rolls n is 10, 20, 50, 100, then the above formula becomes as follows and its probability can be calculated using $B(n, \frac{1}{6})$ in 『eStatH』.

$$\begin{aligned} n = 10 \quad P\left(10\left(\frac{1}{6} - 0.1\right) < X < 10\left(\frac{1}{6} + 0.1\right)\right) &= P(0.6667 < X < 2.6667) \\ &= P(X=1) + P(X=2) = 0.6137 \end{aligned}$$



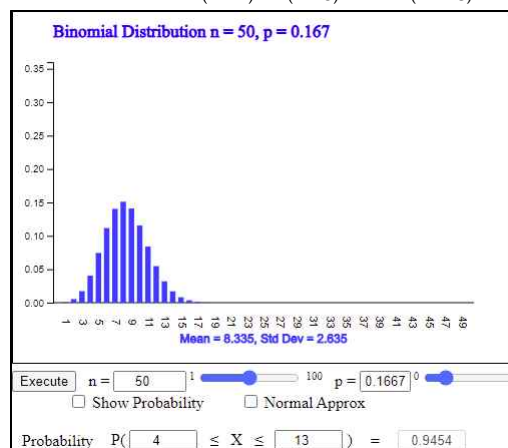
<Figure 3.17> $B(10, 0.1667)$ using 『eStatH』

$$\begin{aligned}
 n = 20 \quad P\left(20\left(\frac{1}{6} - 0.1\right) < X < 20\left(\frac{1}{6} + 0.1\right)\right) &= P(1.3333 < X < 5.3333) \\
 &= P(X=2) + P(X=3) + P(X=4) + P(X=5) = 0.7677
 \end{aligned}$$



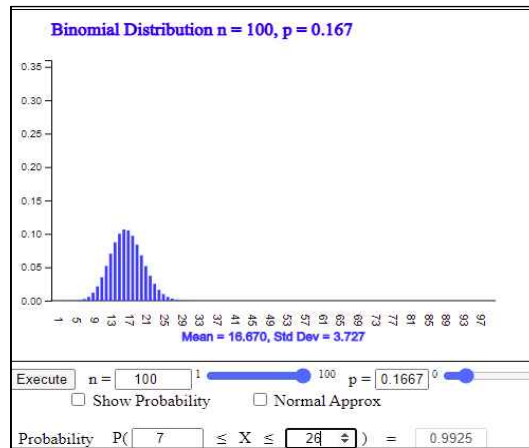
<Figure 3.18> $B(20, 0.1667)$ using 『eStatH』

$$\begin{aligned}
 n = 50 \quad P\left(50\left(\frac{1}{6} - 0.1\right) < X < 50\left(\frac{1}{6} + 0.1\right)\right) &= P(3.3333 < X < 13.3333) \\
 &= P(X=4) + P(X=5) + \dots + P(X=13) = 0.9454
 \end{aligned}$$



<Figure 3.19> $B(50, 0.1667)$ using 『eStatH』

$$\begin{aligned}
 n = 100 \quad P\left(100\left(\frac{1}{6} - 0.1\right) < X < 100\left(\frac{1}{6} + 0.1\right)\right) &= P(6.6667 < X < 26.6667) \\
 &= P(X=7) + P(X=8) + \dots + P(X=26) = 0.9925
 \end{aligned}$$



<Figure 3.20> $B(100, 0.1667)$ using 『eStatH』

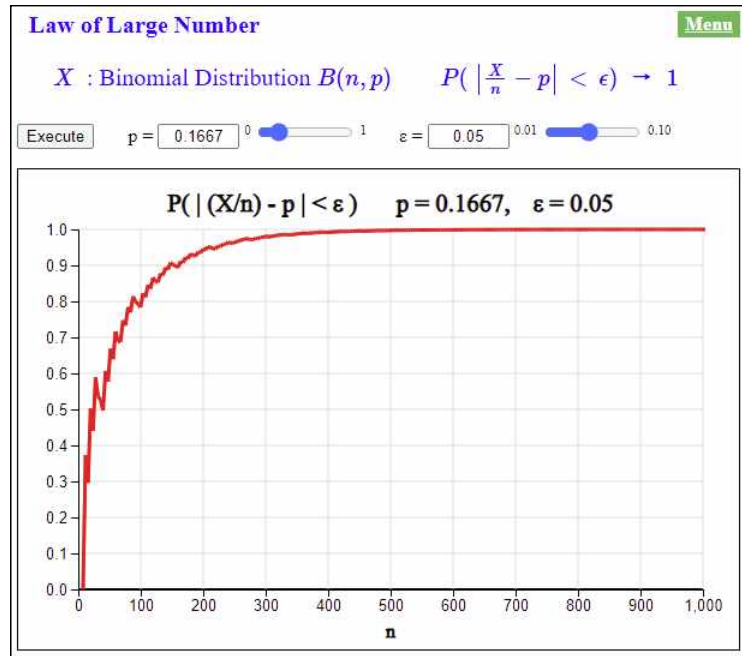
As you noticed here, the probability $P\left(\left|\frac{X}{n} - \frac{1}{6}\right| < 0.1\right)$ is close to 1 as n increases.

- This result is true when ϵ is smaller than 0.1 such as 0.01, 0.001, ... which means that the relative frequency $\frac{X}{n}$ is close to the mathematical probability $\frac{1}{6}$ if n is sufficiently large. This is called the **law of large number**.

☞ Law of large number

Consider a Bernoulli trial with ‘success’ probability of p and let a random variable X be the number of ‘success’ when the trial is repeated n times. The relative frequency $\frac{X}{n}$ is close to p if n is sufficiently large.

- <Figure 3.21> shows that $P\left(\left|\frac{X}{n} - \frac{1}{6}\right| < \epsilon\right)$ is close to 1 when $\epsilon = 0.05$ if n is increasing up to 1000 by using 『eStatH』. You can change p and ϵ using the slide bars.



<Figure 3.21> Simulation experiment of law of large number using 『eStatH』

3.5 Normal Distribution

Think	<p>The histogram of delivery times to the house investigated randomly when pizza is ordered is as follows:</p> <p><Figure 3.22> Histogram using many data of pizza delivery time</p>
Exploration	<ol style="list-style-type: none"> 1) What are the characteristics of this histogram? 2) What is the mathematical function that can describe this histogram?

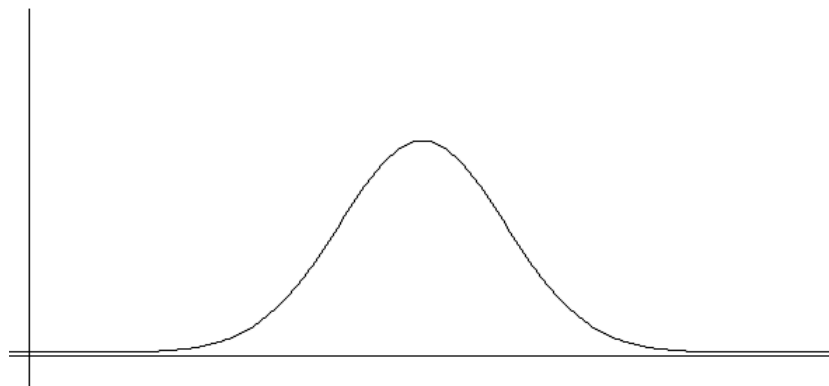
- Among the data of continuous random variables around us, there are many data which have similar shape such as the histogram above. More data are concentrated near the mean, smaller data are located far from the mean, and

data are symmetrical with respect to the mean. In order to find easily the probability of any interval on this shape of data, mathematicians found a function that can describe this shape of data. This mathematical function allows you to approximate the desired probability without finding a frequency table or histogram.

- This function was first discovered by Abraham de Moivre (1667–1754), and then widely applied to physics and astronomy by German mathematician Karl Friedrich Gauss (1777–1855). This function is called the normal distribution function or the Gaussian distribution function, and the equation and graph are as follows:

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

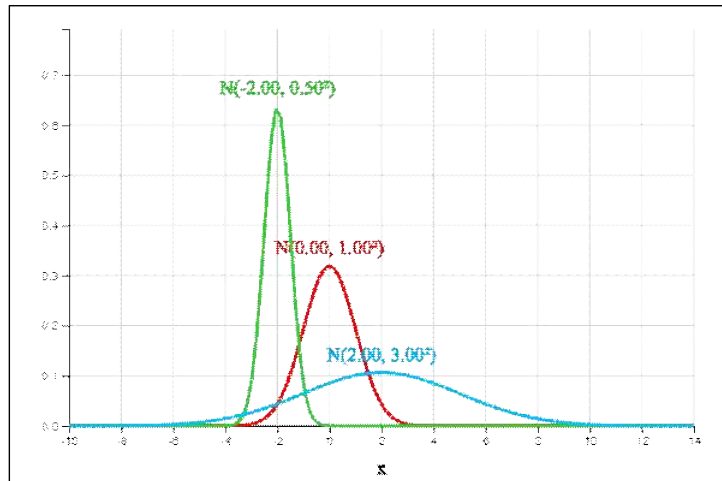
where μ is a constant, σ is a positive constant, e is the irrational number as 2.71828...



<Figure 3.23> probability distribution function of the normal distribution

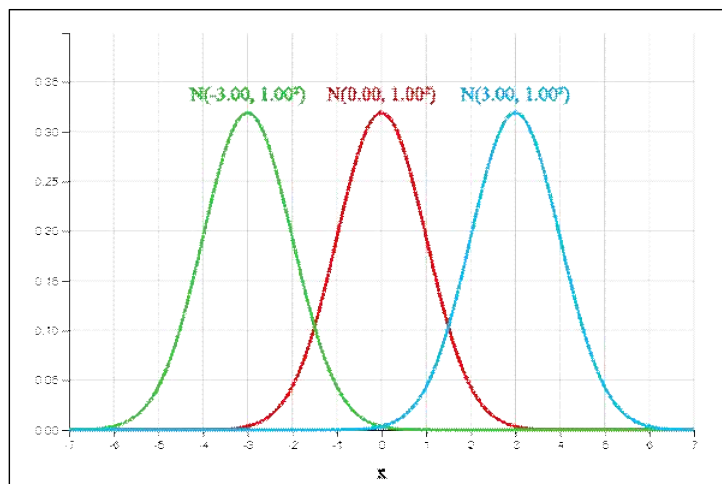
The mean and standard deviation of the normal distribution are μ and σ respectively. If a random variable X follows the normal distribution with the mean μ and standard deviation σ , that is the variance of σ^2 , it is denoted as $X \sim N(\mu, \sigma^2)$.

- <Figure 3.24> is the graph of three normal distributions with different mean and variance such as $N(-2, 0.5^2)$, $N(0, 1)$ and $N(2, 2^2)$. It can be observed that all of them are symmetrical around the mean. If the variance increases, the normal distribution becomes flat and, if the variance decreases, the normal distribution becomes sharp. This is because the total area under the normal distribution function must be 1.



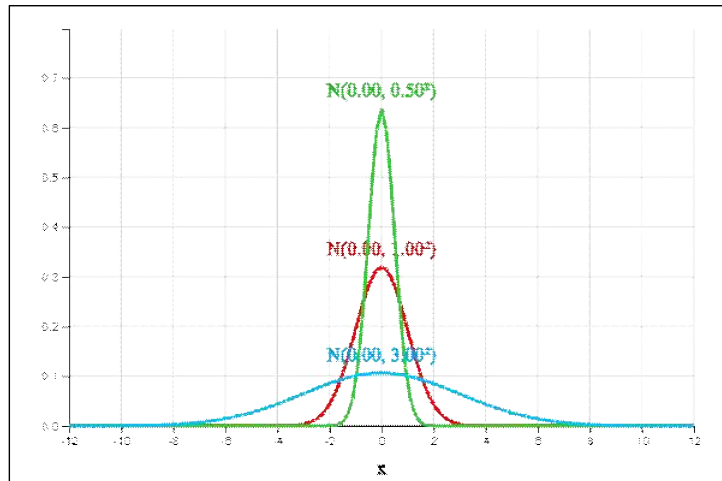
<Figure 3.24> Comparison of three graphs of normal distribution, $N(-2, 0.5^2)$, $N(0, 1)$ and $N(2, 2^2)$

- <Figure 3.25> is the graph of three-normal distributions $N(-3, 1)$, $N(0, 1)$ and $N(3, 1)$ in which the mean is different from each other and the variance is all 1. If the mean is different, the graph of the same shape is moved horizontally.



<Figure 3.25> Comparison of three graphs of normal distribution, $N(-3, 1)$, $N(0, 1)$, $N(3, 1)$ 의 그림

- <Figure 3.26> is the graph of three-normal distributions $N(0, 0.5^2)$, $N(0, 1)$ and $N(0, 3^2)$ in which all means are zero and variances are different. It can be observed that all of them are symmetrical around the average 0, and that the normal distribution becomes flat as the variance increases, and the normal distribution becomes sharp as the variance decreases.

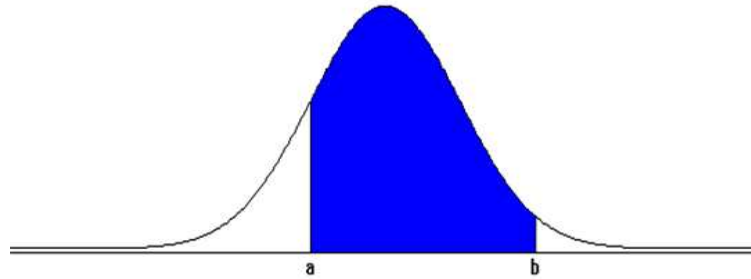


<Figure 3.26> Comparison of three graphs of normal distribution, $N(0, 0.5^2)$, $N(0, 1)$, $N(0, 3^2)$

- The characteristics of the normal distribution are summarized as follows:
 - 1) It is a bell-shaped function symmetrical with respect to a straight line $x = \mu$ and the asymptote is the x axis. Therefore, the probabilities of the left and right sides of the mean are 0.5 in each side.
 - 2) When $x = \mu$, the maximum value is $\frac{1}{\sqrt{2\pi\sigma}}$.
 - 3) When the value of σ is constant, if the value of μ is changed, only the position of the symmetry axis changes and the shape of curve is the same.
 - 4) When the value of μ is constant, as the value of σ increases, the curve becomes low and spreads widely. As the value of σ decreases, the curve becomes higher and sharper.
 - 5) The area between the graph $y=f(x)$ and the x axis is 1.

A. Probability calculation of Normal distribution

- The normal distribution is the most widely used distribution in reality, and it requires a lot of calculation of the probability of the interval $[a, b]$ of the random variable X . As explained earlier, the probability of the interval $[a, b]$, $P(a \leq X \leq b)$, when X follows the normal distribution $N(\mu, \sigma^2)$ is the area of the function enclosed between the x -axis and interval $[a, b]$ as in the colored part in <Figure 3.27>.



<Figure 3.27> Probability $P(a \leq X \leq b)$ of the normal distribution

Mathematically, this area should be obtained as the following definite integral, but it is only possible to use a computer because the function $f(x)$ is impossible to integrate.

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

- The probability of the interval $[\mu - \sigma, \mu + \sigma]$ calculated using the computer is 0.68, the probability of the interval $[\mu - 2\sigma, \mu + 2\sigma]$ is 0.95, and the probability of the interval $[\mu - 3\sigma, \mu + 3\sigma]$ is 0.997. That is, the normal distribution has most values around the mean, and there are few values that are more than three times the standard deviation to the left and right from the mean.
- If a random variable X follows $N(\mu, \sigma^2)$, the transformed variable $Z = \frac{X-\mu}{\sigma}$ follows a normal distribution with mean 0 and standard deviation 1, $N(0,1)$. This fact implies that if we can find all probabilities of $N(0,1)$, we can find the probabilities of any normal distribution. Therefore, $N(0,1)$ is specifically called the **standard normal distribution** or **Z distribution**. The transformation $Z = \frac{X-\mu}{\sigma}$ that turns a random variable X into Z is called a **standardized transformation**.

👁 Standardized transformation

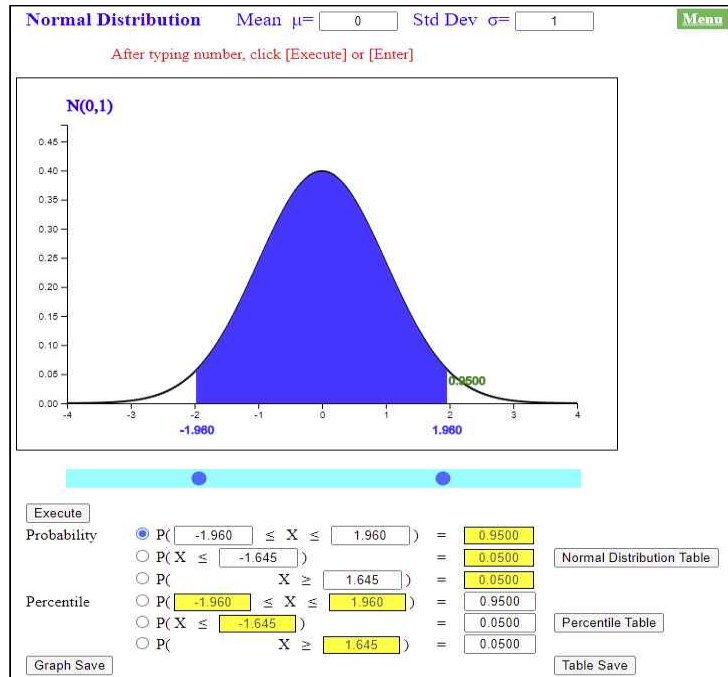
If a random variable X follows a normal distribution with mean μ and variance σ^2 , $N(\mu, \sigma^2)$, the transformed variable $Z = \frac{X - \mu}{\sigma}$ follows a normal distribution with mean 0 and standard deviation 1, $N(0, 1)$. The transformation is called the standardized transformation.

- For the standard normal distribution $N(0, 1)$, a table is created by finding the probabilities from the left end to z for various real values z , $P(Z \leq z)$, which is called the **standard normal distribution table**. The following table is a part of the standard normal distribution table obtained using 『eStatH』.

[Table 3.3] Cumulated probability of standard normal distribution in 『eStatH』

Normal Distribution	$\mu = 0$		$\sigma = 1.000$													
	x	$P(X \leq x)$	x	$P(X \leq x)$	x	$P(X \leq x)$	x	$P(X \leq x)$	x	$P(X \leq x)$	x	$P(X \leq x)$	x	$P(X \leq x)$	x	$P(X \leq x)$
-3.99	0.0000	-2.99	0.0014	-1.99	0.0233	-0.99	0.1611	0.01	0.5040	1.01	0.8438	2.01	0.9778	3.01	0.9987	
-3.98	0.0000	-2.98	0.0014	-1.98	0.0239	-0.98	0.1635	0.02	0.5080	1.02	0.8461	2.02	0.9783	3.02	0.9987	
-3.97	0.0000	-2.97	0.0015	-1.97	0.0244	-0.97	0.1660	0.03	0.5120	1.03	0.8485	2.03	0.9788	3.03	0.9988	
-3.96	0.0000	-2.96	0.0015	-1.96	0.0250	-0.96	0.1685	0.04	0.5160	1.04	0.8508	2.04	0.9793	3.04	0.9988	
-3.95	0.0000	-2.95	0.0016	-1.95	0.0256	-0.95	0.1711	0.05	0.5199	1.05	0.8531	2.05	0.9798	3.05	0.9989	
-3.94	0.0000	-2.94	0.0016	-1.94	0.0262	-0.94	0.1736	0.06	0.5239	1.06	0.8554	2.06	0.9803	3.06	0.9989	
-3.93	0.0000	-2.93	0.0017	-1.93	0.0268	-0.93	0.1762	0.07	0.5279	1.07	0.8577	2.07	0.9808	3.07	0.9989	
-3.92	0.0000	-2.92	0.0018	-1.92	0.0274	-0.92	0.1788	0.08	0.5319	1.08	0.8599	2.08	0.9812	3.08	0.9990	
-3.91	0.0000	-2.91	0.0018	-1.91	0.0281	-0.91	0.1814	0.09	0.5359	1.09	0.8621	2.09	0.9817	3.09	0.9990	
-3.90	0.0000	-2.90	0.0019	-1.90	0.0287	-0.90	0.1841	0.10	0.5398	1.10	0.8643	2.10	0.9821	3.10	0.9990	
-3.89	0.0001	-2.89	0.0019	-1.89	0.0294	-0.89	0.1867	0.11	0.5438	1.11	0.8665	2.11	0.9826	3.11	0.9991	
-3.88	0.0001	-2.88	0.0020	-1.88	0.0301	-0.88	0.1894	0.12	0.5478	1.12	0.8686	2.12	0.9830	3.12	0.9991	
-3.87	0.0001	-2.87	0.0021	-1.87	0.0307	-0.87	0.1922	0.13	0.5517	1.13	0.8708	2.13	0.9834	3.13	0.9991	
-3.86	0.0001	-2.86	0.0021	-1.86	0.0314	-0.86	0.1949	0.14	0.5557	1.14	0.8729	2.14	0.9838	3.14	0.9992	
-3.85	0.0001	-2.85	0.0022	-1.85	0.0322	-0.85	0.1977	0.15	0.5596	1.15	0.8749	2.15	0.9842	3.15	0.9992	
-3.84	0.0001	-2.84	0.0023	-1.84	0.0329	-0.84	0.2005	0.16	0.5636	1.16	0.8770	2.16	0.9846	3.16	0.9992	
-3.83	0.0001	-2.83	0.0023	-1.83	0.0336	-0.83	0.2033	0.17	0.5675	1.17	0.8790	2.17	0.9850	3.17	0.9992	
-3.82	0.0001	-2.82	0.0024	-1.82	0.0344	-0.82	0.2061	0.18	0.5714	1.18	0.8810	2.18	0.9854	3.18	0.9993	
-3.81	0.0001	-2.81	0.0025	-1.81	0.0351	-0.81	0.2090	0.19	0.5753	1.19	0.8830	2.19	0.9857	3.19	0.9993	
-3.80	0.0001	-2.80	0.0026	-1.80	0.0359	-0.80	0.2119	0.20	0.5793	1.20	0.8849	2.20	0.9861	3.20	0.9993	

- In 『eStatH』, it is easy to calculate the probability for the interval $[a, b]$ of any normal distribution, $P(a \leq X \leq b)$, and the percentile x for a given cumulated probability p , that is $P(X \leq x) = p$, as shown in <Figure 3.28>.





<Figure 3.28> Normal probability calculation using 『eStatH』

- In 『eStatH』, the probability of an interval can be calculated from $\mu - 4\sigma$ to $\mu + 4\sigma$. If the probability which is less than $\mu - 4\sigma$ or greater than $\mu + 4\sigma$, the probability is 0.0000. [Table 3.4] is the percentile table of the standard normal distribution of 『eStatH』.

[Table 3.4] Percentile of standard normal distribution in 『eStatH』

정규분포	$\mu = 0$	$\sigma = 1.000$							
p	$P(X \leq x) = p$	p	$P(X \leq x) = p$	p	$P(X \leq x) = p$	p	$P(X \leq x) = p$	p	$P(X \leq x) = p$
0.005	-2.576	0.205	-0.824	0.405	-0.240	0.605	0.266	0.805	0.860
0.010	-2.326	0.210	-0.806	0.410	-0.228	0.610	0.279	0.810	0.878
0.015	-2.170	0.215	-0.789	0.415	-0.215	0.615	0.292	0.815	0.896
0.020	-2.054	0.220	-0.772	0.420	-0.202	0.620	0.305	0.820	0.915
0.025	-1.960	0.225	-0.755	0.425	-0.189	0.625	0.319	0.825	0.935
0.030	-1.881	0.230	-0.739	0.430	-0.176	0.630	0.332	0.830	0.954
0.035	-1.812	0.235	-0.722	0.435	-0.164	0.635	0.345	0.835	0.974
0.040	-1.751	0.240	-0.706	0.440	-0.151	0.640	0.358	0.840	0.994
0.045	-1.695	0.245	-0.690	0.445	-0.138	0.645	0.372	0.845	1.015
0.050	-1.645	0.250	-0.674	0.450	-0.126	0.650	0.385	0.850	1.036
0.055	-1.598	0.255	-0.659	0.455	-0.113	0.655	0.399	0.855	1.058
0.060	-1.555	0.260	-0.643	0.460	-0.100	0.660	0.412	0.860	1.080
0.065	-1.514	0.265	-0.628	0.465	-0.088	0.665	0.426	0.865	1.103
0.070	-1.476	0.270	-0.613	0.470	-0.075	0.670	0.440	0.870	1.126
0.075	-1.440	0.275	-0.598	0.475	-0.063	0.675	0.454	0.875	1.150
0.080	-1.405	0.280	-0.583	0.480	-0.050	0.680	0.468	0.880	1.175
0.085	-1.372	0.285	-0.568	0.485	-0.038	0.685	0.482	0.885	1.200
0.090	-1.341	0.290	-0.553	0.490	-0.025	0.690	0.496	0.890	1.227
0.095	-1.311	0.295	-0.539	0.495	-0.013	0.695	0.510	0.895	1.254
0.100	-1.282	0.300	-0.524	0.500	0.000	0.700	0.524	0.900	1.282

<p>Example 3.9</p>	<p>If Z is the standard normal random variable, find the following probabilities using the standard normal distribution table and confirm it using 『eStatH』 .</p> <p>1) $P(Z \leq 1.96)$ 2) $P(-1.96 \leq Z \leq 1.96)$ 3) $P(Z \geq 1.96)$</p>
<p>Solution</p> 	<p>1) In the standard normal distribution table, you can look up $P(Z \leq 1.96)$ is 0.975.</p> <p>2) $P(-1.96 \leq Z \leq 1.96) = P(Z \leq 1.96) - P(Z \leq -1.96)$ $= 0.975 - 0.025 = 0.95$</p> <p>3) $P(Z \geq 1.96) = 1 - P(Z \leq 1.96) = 1 - 0.975 = 0.025$</p> <p>When using 『eStatH』 for 1), select the second option under the graph, enter 1.96, and click [Execute] button.</p> <p>In the same way, 2) is calculated by entering -1.96 and 1.96 in the first option under the graph.</p> <p>3) is calculated by selecting the third option under the graph and entering 1.96.</p>

<p>Example 3.10</p>	<p>If Z is the standard normal random variable, find x which satisfies the following formula and confirm it using 『eStatH』 .</p> <p>1) $P(Z \leq x) = 0.90$ 2) $P(-x \leq Z \leq x) = 0.99$ 3) $P(Z \geq x) = 0.05$</p>												
<p>Solution</p> 	<p>1) In the standard normal distribution table, you can look up the value of x as 1.28 approximately.</p> <p>2) The x is which both ends become 0.005 is 2.575.</p> <p>3) It is the same as looking up $P(Z \leq x) = 0.95$, x is 1.645 in the table.</p> <p>When using 『eStatH』 for 1), enter $p = 0.90$ in the right box in the fifth option under the graph and click [Execute] button. It can be seen that the correct 90% percentile is 1.282.</p> <div data-bbox="651 1585 1289 1765" style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">Probability</td> <td style="padding: 2px;"><input type="radio"/> $P(-1.960 \leq X \leq 1.960) = 0.9500$</td> </tr> <tr> <td style="padding: 2px;"></td> <td style="padding: 2px;"><input type="radio"/> $P(X \leq -1.645) = 0.0500$</td> </tr> <tr> <td style="padding: 2px;"></td> <td style="padding: 2px;"><input type="radio"/> $P(X \geq 1.645) = 0.0500$</td> </tr> <tr> <td style="padding: 2px;">Percentile</td> <td style="padding: 2px;"><input checked="" type="radio"/> $P(-2.576 \leq X \leq 2.576) = 0.9900$</td> </tr> <tr> <td style="padding: 2px;"></td> <td style="padding: 2px;"><input type="radio"/> $P(X \leq 1.282) = 0.9000$</td> </tr> <tr> <td style="padding: 2px;"></td> <td style="padding: 2px;"><input type="radio"/> $P(X \geq 1.645) = 0.0500$</td> </tr> </table> </div> <p>For 2), use the fourth option under the graph, enter $p = 0.99$ in the right box and click [Execute] button. It can be seen that the correct bilateral percentile are -2.5758 and 2.576.</p> <p>For 3), use the 6th option under the graph, enter $p = 0.05$ in the box on the right and click [Execute] button. It can be seen that the correct right 5% percentile is 1.645.</p>	Probability	<input type="radio"/> $P(-1.960 \leq X \leq 1.960) = 0.9500$		<input type="radio"/> $P(X \leq -1.645) = 0.0500$		<input type="radio"/> $P(X \geq 1.645) = 0.0500$	Percentile	<input checked="" type="radio"/> $P(-2.576 \leq X \leq 2.576) = 0.9900$		<input type="radio"/> $P(X \leq 1.282) = 0.9000$		<input type="radio"/> $P(X \geq 1.645) = 0.0500$
Probability	<input type="radio"/> $P(-1.960 \leq X \leq 1.960) = 0.9500$												
	<input type="radio"/> $P(X \leq -1.645) = 0.0500$												
	<input type="radio"/> $P(X \geq 1.645) = 0.0500$												
Percentile	<input checked="" type="radio"/> $P(-2.576 \leq X \leq 2.576) = 0.9900$												
	<input type="radio"/> $P(X \leq 1.282) = 0.9000$												
	<input type="radio"/> $P(X \geq 1.645) = 0.0500$												

Exercise 3.10



If Z is the standard normal random variable, find the following probabilities using the standard normal distribution table and confirm it using 「eStatH」.

- 1) $P(-1 \leq Z \leq 1)$
- 2) $P(-2 \leq Z \leq 2)$
- 3) $P(-3 \leq Z \leq 3)$

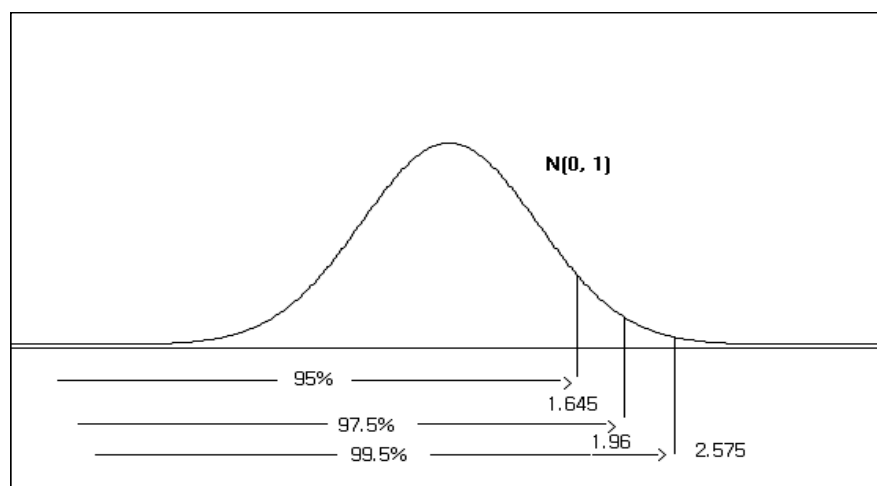
Exercise 3.11



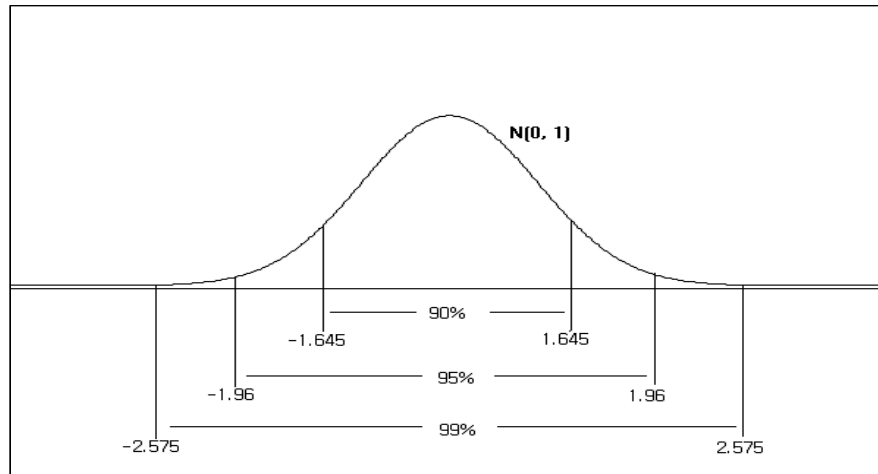
If Z is the standard normal random variable, find x which satisfies the following formula and confirm it using 「eStatH」.

- 1) $P(Z \leq x) = 0.99$
- 2) $P(-x \leq Z \leq x) = 0.95$
- 3) $P(Z \geq x) = 0.01$

- It is good to remember probabilities of some intervals in the standard normal distribution that are often used. <Figure 3.29> shows the percentiles of 95%, 97.5%, and 99.5% from the left of the standard normal distribution. <Figure 3.30> shows the 95% and 99% probabilities when both ends are equally excluded.



<Figure 3.29> 95%, 97.5%, 99.5% percentiles of standard normal distribution, that is, $P(Z \leq 1.645) = 0.95$, $P(Z \leq 1.96) = 0.975$, $P(Z \leq 2.575) = 0.995$



<Figure 3.30> Middle 95%, 99% percentiles of standard normal distribution, that is, $P(-1.96 \leq Z \leq 1.96) = 0.95$, $P(-2.575 \leq Z \leq 2.575) = 0.99$

- Using the standard normal distribution table, you can find the probability of a general normal distribution. If X is the normal distribution with mean μ and variance σ^2 , $\frac{X-\mu}{\sigma}$ follows the standard normal distribution. Therefore, the probability of the interval $[a, b]$ of X , $P(a \leq X \leq b)$, is obtained by finding the probability of the interval $[\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}]$ from the standard normal distribution.

☞ Calculation of probability on interval $[a, b]$ in normal distribution

If the random variable X is the normal distribution with mean μ and variance σ^2 , the probability of the interval $[a, b]$ of X , $P(a \leq X \leq b)$ is as follows:

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

Example 3.11

Find the following probability when the midterm score of the Statistics course, X , follows a normal distribution with a mean of 70 point and a standard deviation of 10 points. Check the calculated value using 「eStatH」.

- $P(X < 94.3)$
- $P(X > 57.7)$
- $P(57.7 < X < 94.3)$

Example 3.11
Solution

The probability calculation of each question is as follows:

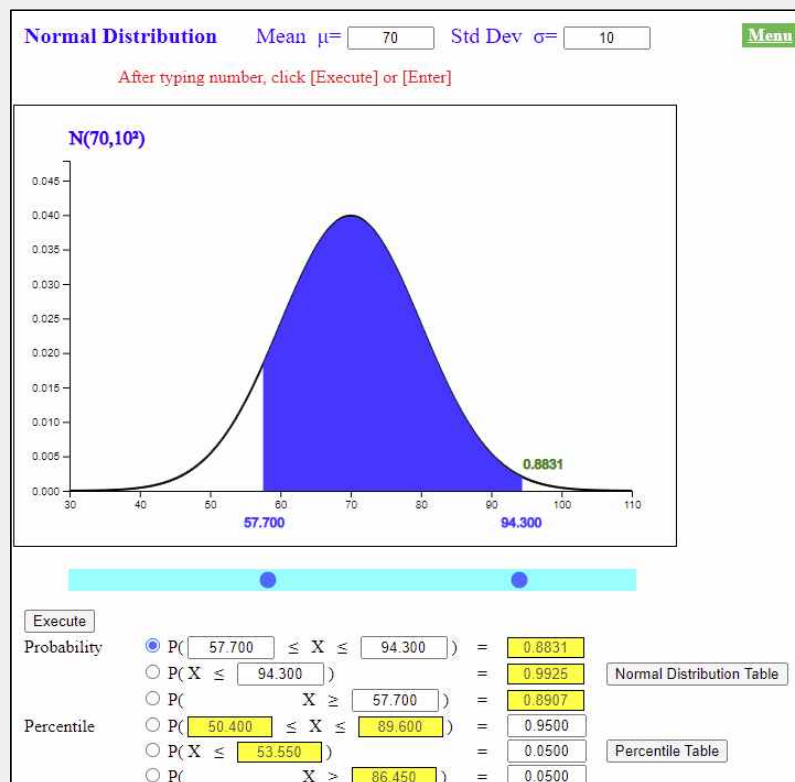
- 1) $P(X \leq 94.3) = P\left(\frac{X-70}{10} \leq \frac{94.3-70}{10}\right) = P(Z \leq 2.43) = 0.9925$
- 2) $P(X \geq 57.7) = P\left(\frac{X-70}{10} \geq \frac{57.7-70}{10}\right) = P(Z \geq -1.23) = 0.8907$
- 3) $P(57.7 \leq X \leq 94.3) = P\left(\frac{57.7-70}{10} \leq \frac{X-70}{10} \leq \frac{94.3-70}{10}\right)$
 $= P(-1.23 \leq Z \leq 2.43) = 0.8831$

To obtain the probability of a normal distribution using 『eStatH』, first enter the mean as 70 and the standard deviation as 10 on the screen of <Figure 3.31>.

For 1), enter the value of 94.3 in the second option under the graph and click [Execute] button.

In a similar way, for 2), enter 57.7 in the third option to calculate.

For 3), enter the interval as [57.7, 94.3] in the first option and click [Execute] button.



<Figure 3.31> Calculation of probability in normal distribution

Example 3.12

Find the following percentiles in [Example 3.11] and confirm the percentiles using 『eStatH』.

- 1) What is the 95% percentile of the midterm test scores?
- 2) What is the 95% percentile of excluding the both end sides of the midterm test scores?

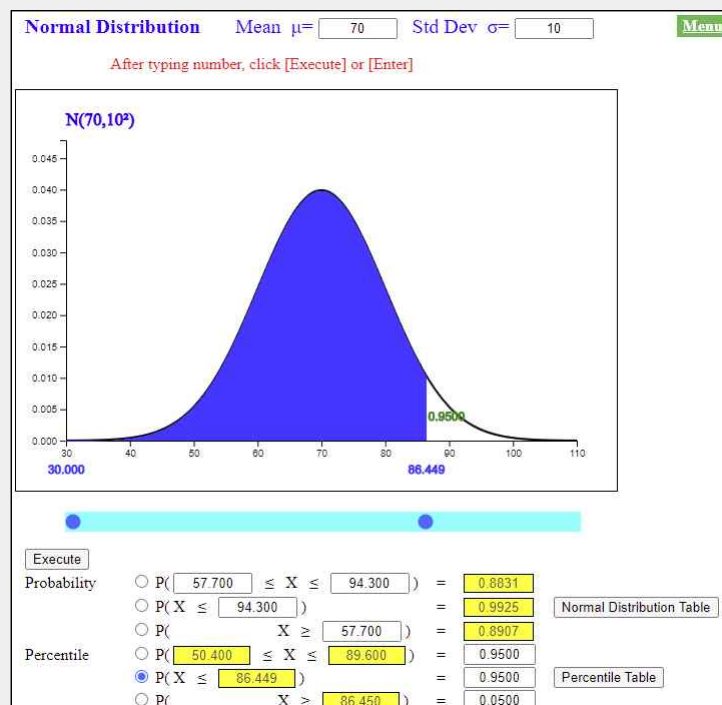
Solution

- 1) The percentile $P(Z \leq z) = 0.95$ of the standard normal distribution is 1.645, so the percentile of the normal distribution is $70 + 1.645 \times 10 = 86.45$.
- 2) Since the sum of the probabilities at both ends of the 95% percentile is 5%, you need to first find the 97.5% percentile from the left end. Since the percentile z , $P(z \leq Z \leq z) = 0.95$ of the standard normal distribution is 1.960, the 95% percentile of excluding the both end sides of the midterm test scores is $[70 - 1.96 \times 10, 70 + 1.96 \times 10]$, that is, $[50.4, 89.6]$.

To obtain the percentile of a normal distribution using 『eStatH』, first enter the mean as 70 and the standard deviation as 10 on the screen of <Figure 3.32>.

For 1), enter 0.95 in the right box of the fifth option under the graph and click [Execute] button, the 95% percentile 86.449 appears.

For 2), enter 0.95 in the right box of the fourth option under the graph screen and click [Execute] button to display both 95% percentiles $[50.400, 89.600]$.



<Figure 3.32> Calculation of percentile in normal distribution

Exercise 3.12

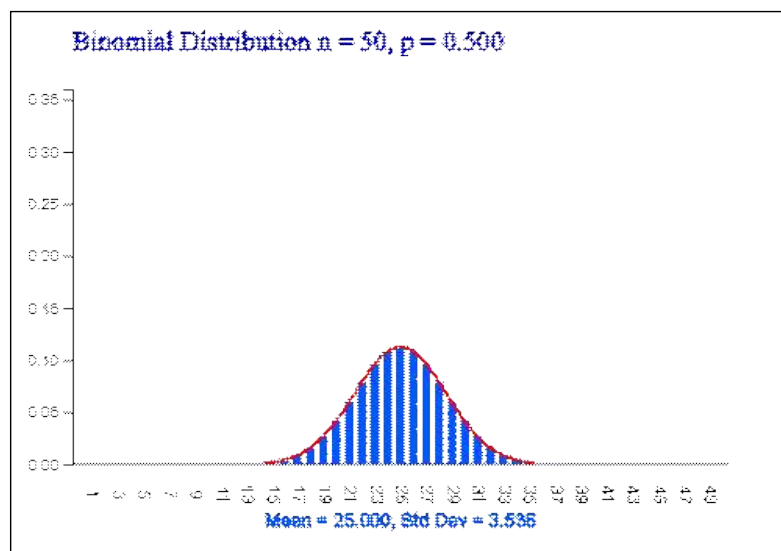


It is said that the weight of melons follows a normal distribution with a mean of 250g and a standard deviation of 12g. Find the following probabilities and check the calculation using 『eStatH』 .

- 1) Find the probability that the weight of the melon is less than 260g.
- 2) Find the probability that the weight of the melon is greater than 240g.
- 3) Find the probability that the weight of the melon is greater than 240g and less than 260g.
- 4) Find the weight corresponding to the top 10% of melons.

B. Approximation of Binomial distribution using Normal distribution

- If n is large (about 50 or more) in case of the binomial distribution, $B(n, p)$, the probability calculation cannot be obtained even by using 『eStatH』 . In this case, the binomial distribution can be approximated using a normal distribution with the mean np , variance $np(1-p)$. <Figure 3.33> is a graph obtained by approximating the binomial distribution with $n = 50$, $p = 0.5$ using the normal distribution with the mean $np = 50 \times 0.5 = 25$ and variance $np(1-p) = 50 \times 0.5 \times 0.5 = 12.5$.



<Figure 3.33> Normal approximation of binomial when $n = 50$, $p = 0.5$

<p>Example 3.13</p>	<p>It is said that the defective rate of products produced in one factory is 5%. When 100 products were sampled on a day,</p> <ol style="list-style-type: none"> 1) What is the probability that there are less than 2 defective products? 2) What is the probability of 3 to 7 defective products?
<p>Solution</p>	<p>If X is the number of defective products, then X is a binomial distribution with $n = 100$ and $p = 0.05$. In this case, n is large and the probability can be approximated using the normal distribution. Since the mean of this binomial distribution is $np = 100 \times 0.05 = 5$, and the variance is $np(1-p) = 100 \times 0.05 \times (1-0.05) = 4.75$, the probability calculation using the normal distribution $N(5, 4.75)$ is as follows:</p> <ol style="list-style-type: none"> 1) $P(X \leq 2) = P\left(Z \leq \frac{2-5}{\sqrt{4.75}}\right) = P(Z \leq 1.376) = 0.0845$ 2) $P(3 \leq X \leq 7) = P\left(\frac{3-5}{\sqrt{4.75}} \leq Z \leq \frac{7-5}{\sqrt{4.75}}\right)$ $= P(-0.918 \leq Z \leq 0.918) = 0.642$

<p>Exercise 3.13</p>	<p>In one presidential election, a particular candidate's approval rating was found to be 40%. To develop an election strategy, we will meet with 100 voters today to see if they support the particular candidate. Find the following probability and check it using 「eStatH」. Assume that there is no abstention.</p> <ol style="list-style-type: none"> 1) What is the probability that 40 or more (\geq) and less than (\leq) 50 out of 100 voters will support it? 2) What is the probability that more than 20 people (\geq) will support it?
-----------------------------	--

Exercise

3.1 Which one is an appropriate function as a probability distribution function? (Answer ②)

- ① $f(x) = \frac{x}{5}$ ($x = -1, 0, 1, 2, 3$) ② $f(x) = \frac{(5-x)}{14}$ ($x = 0, 1, 2, 3$)
 ③ $f(x) = x^2 - 1$ ($x = 0, 1, 2, 3$) ④ $f(x) = x^3 - x^2 + 5$ ($x = 0, 1, 2, 3$)

3.2 Which one is not the property of a probability distribution function? (Answer ④)

- ① $f(x) \geq 0$ ② $\int_{-\infty}^{\infty} f(x) dx = 1$
 ③ $P(a < X < b) = \int_a^b f(x) dx$ ④ $\int_{-\infty}^{\infty} xf(x) dx = 1$

3.3 If a random variable X has values 0, 1, 2, 3 and its probability distribution function is $f(x) = \frac{x}{6}$, what is the mean of the random variable X? (Answer ③)

- ① $\frac{5}{3}$ ② $\frac{6}{3}$ ③ $\frac{7}{3}$ ④ $\frac{8}{3}$

3.4 If a continuous random variable X has the probability distribution function $f(x)$ and it is defined on all real number, which one of the followings is not correct? (Answer ④)

- ① $f(x) \geq 0$ ② $\int_{-\infty}^{\infty} f(x) dx = 1$
 ③ $P(a \leq x \leq b) = \int_a^b f(x) dx$ ④ $\int_a^b f(x) dx = \frac{1}{2}$

3.5 If a probability distribution function of a random variable X is as follows, what are the mean $E(X)$ and standard deviation $\sigma(X)$? (Answer ①)

X	0	1	2	합 계
P(X=x)	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	1

- ① $E(X)=0.8, \sigma(X)=0.6$ ② $E(X)=0.6, \sigma(X)=0.8$
 ③ $E(X)=0.7, \sigma(X)=0.1$ ④ $E(X)=0.1, \sigma(X)=0.7$

3.6 Let the random variable X be the number of dots, 1, 2, ..., 6, which are appeared on the top when you roll a dice. Find the mean of the random variable X. (Answer ④)

- ① 2 ② 2.5

③ 3

④ 3.5

3.7 Let the random variable X be the number of dots when you roll a dice. What is the expected value of the squared value of each X ?

(Answer ④)

① $\frac{23}{6}$ ② $\frac{41}{6}$ ③ $\frac{53}{6}$ ④ $\frac{91}{6}$

3.8 If a random variable X has $E(X)=5$ and $E(X^2)=25$, what is the variance of X , $V(X)$?

(Answer ②)

① 5

② 0

③ 5

④ 25

3.9 If a random variable X is multiplied by 5 for each value of X , the standard deviation becomes what?

(Answer ①)

① 5 times.

② $1/5$ times.

③ 10 times.

④ $1/10$ times.

3.10 Midterm scores of a math class have the mean 24 and standard deviation 3. If each midterm score is multiplied by 2 and add 10 points, what are the mean and standard deviation of the new midterm scores?

(Answer ④)

① 24, 3

② 48, 3

③ 48, 6

④ 58, 6

3.11 If a random variable X has the mean of 20, what is the mean of $Y=2X+3$?

(Answer ④)

① 20

② 40

③ 23

④ 43

3.12 If a random variable X has the variance of 2, what is the variance of $Y=2X+3$?

(Answer ③)

① 2

② 3

③ 8

④ 18

3.13 Which one the followings is wrong?

(Answer ④)

① $E(aX+b)=aE(X)+b$ ② $E(X+Y)=E(X)+E(Y)$ ③ $V(aX+b)=a^2V(X)$ ④ $V(aX+b)=a^2V(X)+b$

3.14 What is the mean and variance of binomial distribution $B\left(100, \frac{1}{5}\right)$?

(Answer ④)

- ① 20, 4 ② 100, 16
 ③ 100, 4 ④ 20, 16

3.15 What is the variance of the binomial distribution which has the distribution function as $P(X=x) = {}_4C_x(0.2)^x(1-0.2)^{4-x}$?

(Answer ④)

- ① 0.16 ② 0.8
 ③ 0.32 ④ 0.64

3.16 There are 10 multiple choice questions where there is one correct answer among four possible answers. If a person chooses one answer randomly in each question, what is the probability that he would choose four correct answers? (Answer ③)

- ① 0.0162 ② 0.0487
 ③ 0.1460 ④ 0.2050

3.17 If a normal distribution has the mean μ and standard deviation σ , what is the probability of the interval $\mu \pm 3\sigma$?

(Answer ①)

- ① 99.73% approximately ② 97.73% approximately
 ③ 95.73% approximately ④ 68.73% approximately

3.18 The entrance examination scores of a company follows a normal distribution with the mean 400 and standard deviation 50. What is the probability that a score belongs to the interval between 450 and 500. ?

(Answer ④)

- ① 약 80% ② 약 50%
 ③ 약 30% ④ 약 14%

Answer

3.1 ②, 3.2 ④, 3.3 ③, 3.4 ④, 3.5 ①, 3.6 ④, 3.7 ④, 3.8 ②, 3.9 ①, 3.10 ④,
 3.11 ④, 3.12 ③, 3.13 ④, 3.14 ④, 3.15 ④, 3.16 ③, 3.17 ①, 3.18 ④,