Introduction to Statistics and Data Science using *eStat*

Chapter 5 Probability Distribution

5.4 Continuous Random Variable - Normal Distribution -

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

5.4.1 Normal Distribution

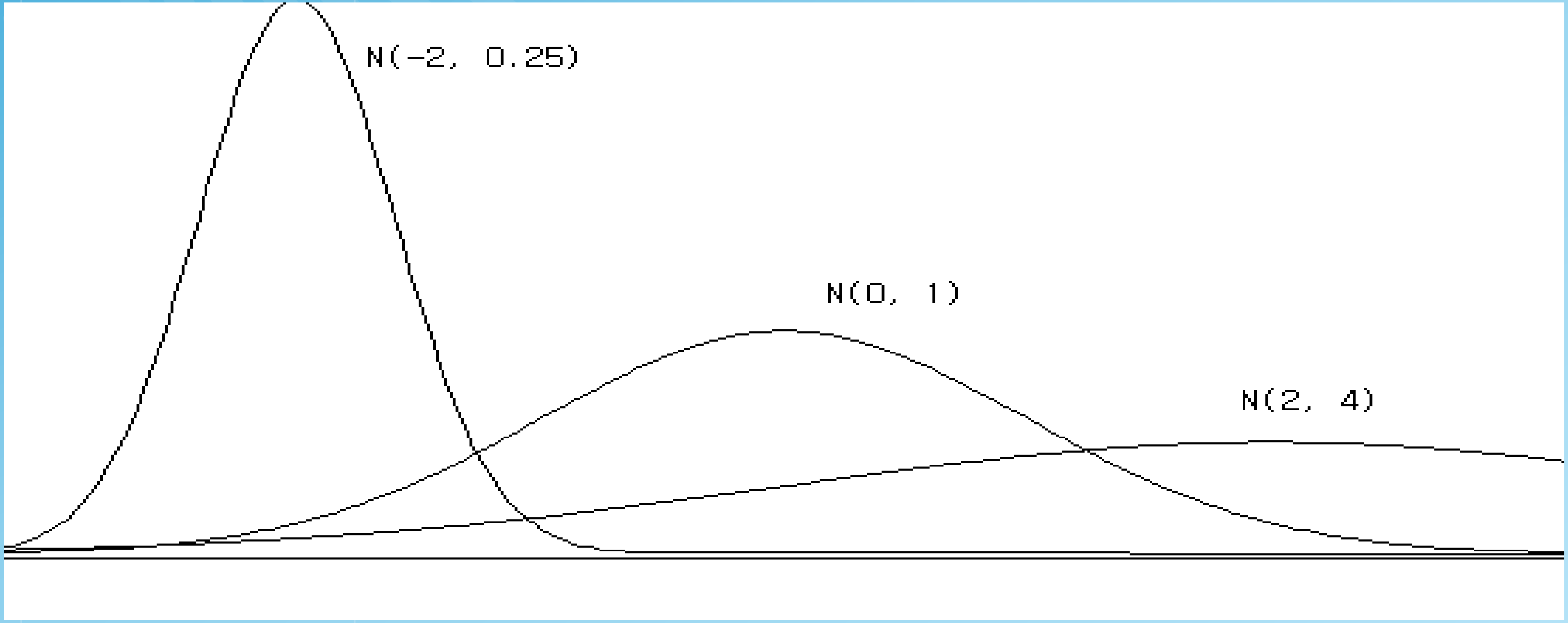
5.4.1 Normal Distribution

- Continuous data that appears more often in the form of a bell-shape
  ⇨ large collection of data near mean,
  fewer data as it moves away from the mean,
  symmetrical around the mean.
- This type of data is called a normal distribution.
  ⇨ height, weight, length of bolt
- Mathematicians tried to find a function to describe this distribution type.
  Abraham de Moivre (1667-1754) was first discovered the function
  Carl Friedrich Gauss (1777-1855) applied to physics and astronomy.
  ⇨ normal distribution function or a Gaussian function

5.4 Continuous Random Variable

5.4.1 Normal Distribution

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



5.4 Continuous Random Variable

5.4.1 Normal Distribution

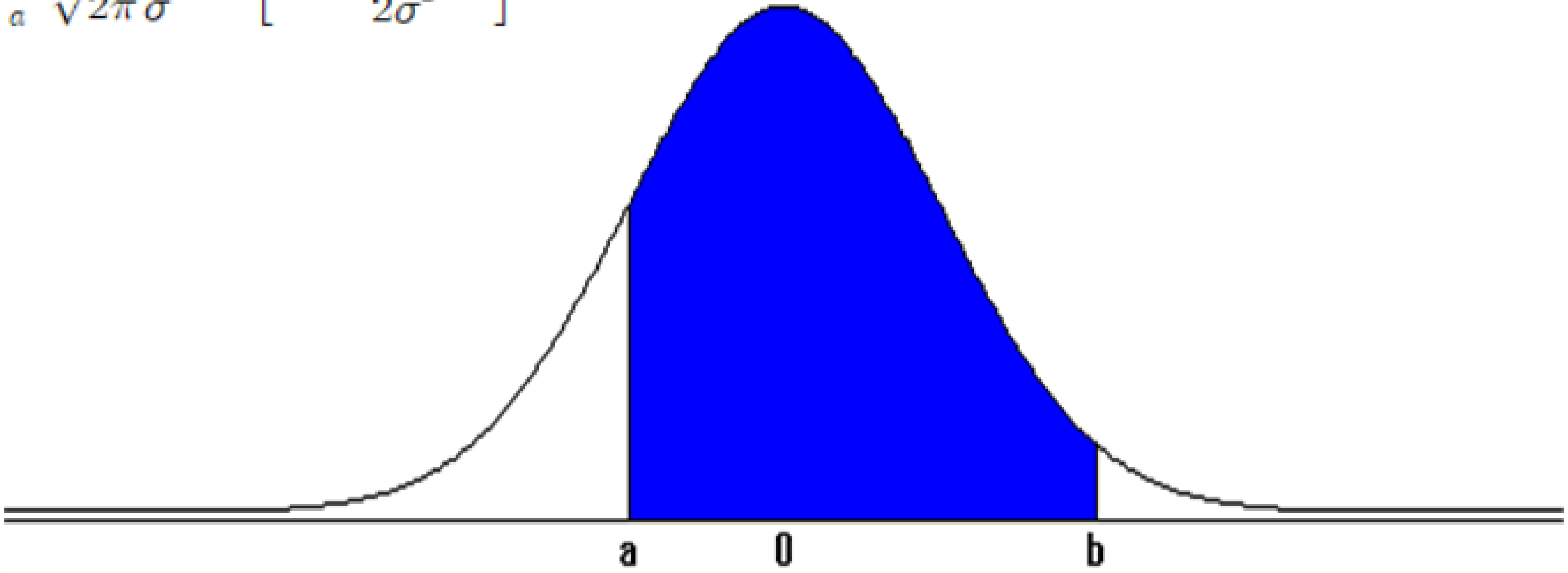
- Characteristics of the normal distribution.

1) continuous function in the shape of a bell.
2) symmetrical with respect to the mean μ. So the probability of the left and right sides of the mean is 0.5 each.
3) There are an infinite number of normal distributions according to μ and σ.
4) Probability of interval $[\mu - \sigma, \mu + \sigma]$ is 0.68,
   Probability of interval $[\mu - 2\sigma, \mu + 2\sigma]$ is 0.95,
   Probability of interval $[\mu - 3\sigma, \mu + 3\sigma]$ is 0.997.
   ⇨ most of the values around the interval of $\mu \pm 3\sigma$
   few values outside of this interval

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Probability Calculation

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



<Figure 5.4.8> probability of X on the interval $[a,b]$, $P(a \le X \le b)$

5.4 Continuous Random Variable

5.4.1 Normal Distribution

❖ Standardized Normal Random Variable

- If X is a normal random variable with the mean μ and variance σ^2, i.e. $X \sim N(\mu, \sigma^2)$,

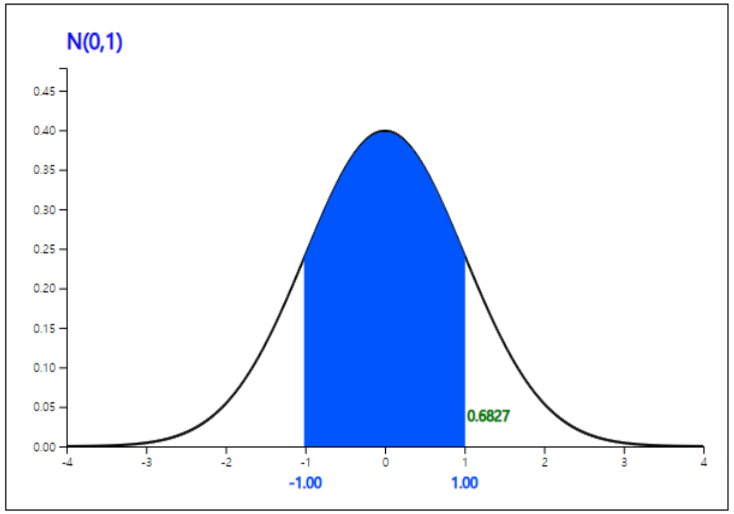
  then $Z = \dfrac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

5.4 Continuous Random Variable

5.4.1 Normal Distribution – 『eStatU』Probability Calculation

Normal Distribution

Mean μ= 0   Std Dev σ= 1   (If number is typed) Execute



P(-1 < Z < 1) = 0.6827

P(Z < 1.6449) = 0.95

P(-1.9600 < Z < 1.9600) = 0.95

Graph Save   Reference Site : Wikipedia Wolfram StatTrek KhanAcademy

Normal Distribution Table   Percentile Table   Table Save

정규분포	μ = 0	σ = 1.000													
x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ 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

5.4 Continuous Random Variable

[Ex 5.4.2] When Z is a standard normal random variable, obtain the following probability using the standard normality distribution table. Then use 『eStatU』.
1) $P(Z < 1.96)$ 2) $P(-1.96 < Z < 1.96)$ 3) $P(Z > 1.96)$

<Answer>
1) By using the standard normal distribution table, $P(Z < 1.96) = 0.975$.
2) $P(-1.96 < Z < 1.96) = P(Z < 1.96) - P(Z < -1.96) = 0.975 - 0.025 = 0.95$
3) $P(Z > 1.96) = 1 - P(Z < 1.96) = 1 - 0.975 = 0.025$

- By using the normal distribution module of 『eStatU』,

1) enters the interval –4, 1.96 in the first of the options below the graph, then clicks the [Execute] button.

P(-4.00 < Z < 1.96) = 0.9750

2) The answer is calculated by entering interval of 1.96 and 1.96,
3) is calculated by entering interval of 1.96 and 4

5.4 Continuous Random Variable

[Example 5.4.3] When Z is a standard random variable, obtain x that satisfies the following formula. Then use 『eStatU』 to find this value x.

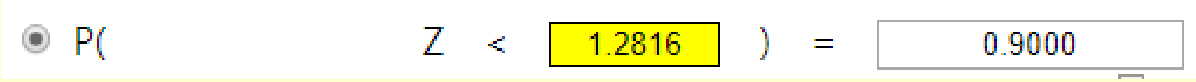
1) $P(Z < x) = 0.90$ 2) $P(-x < Z < x) = 0.99$ 3) $P(Z > x) = 0.05$

<Answer>

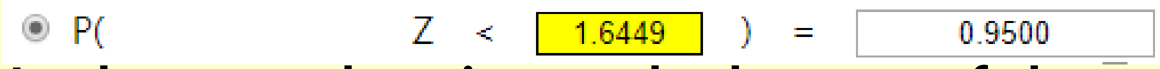
1) By using the standard normal distribution, the value of x is 1.2826
2) By using the standard normal distribution, the percentile of 0.995 is 2.575. .
3) By using the standard normal distribution, the value of x is 1.645.

- by using 『eStatU』,

1) Enter p = 0.90 in the right box in the second option at the bottom of the graph screen, then clicks the [Execute] button.



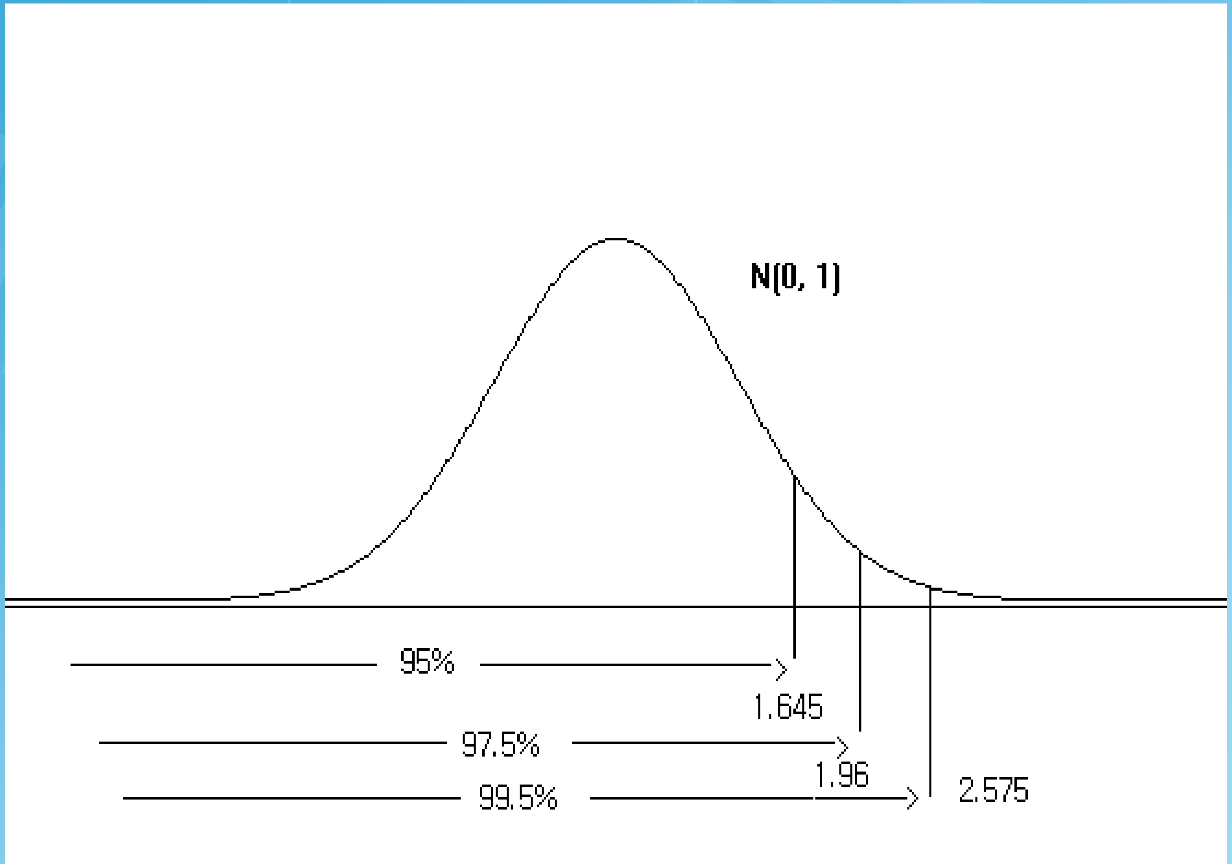
2) Enter p = 0.99 in the right box in the third option and click the [Execute] button.

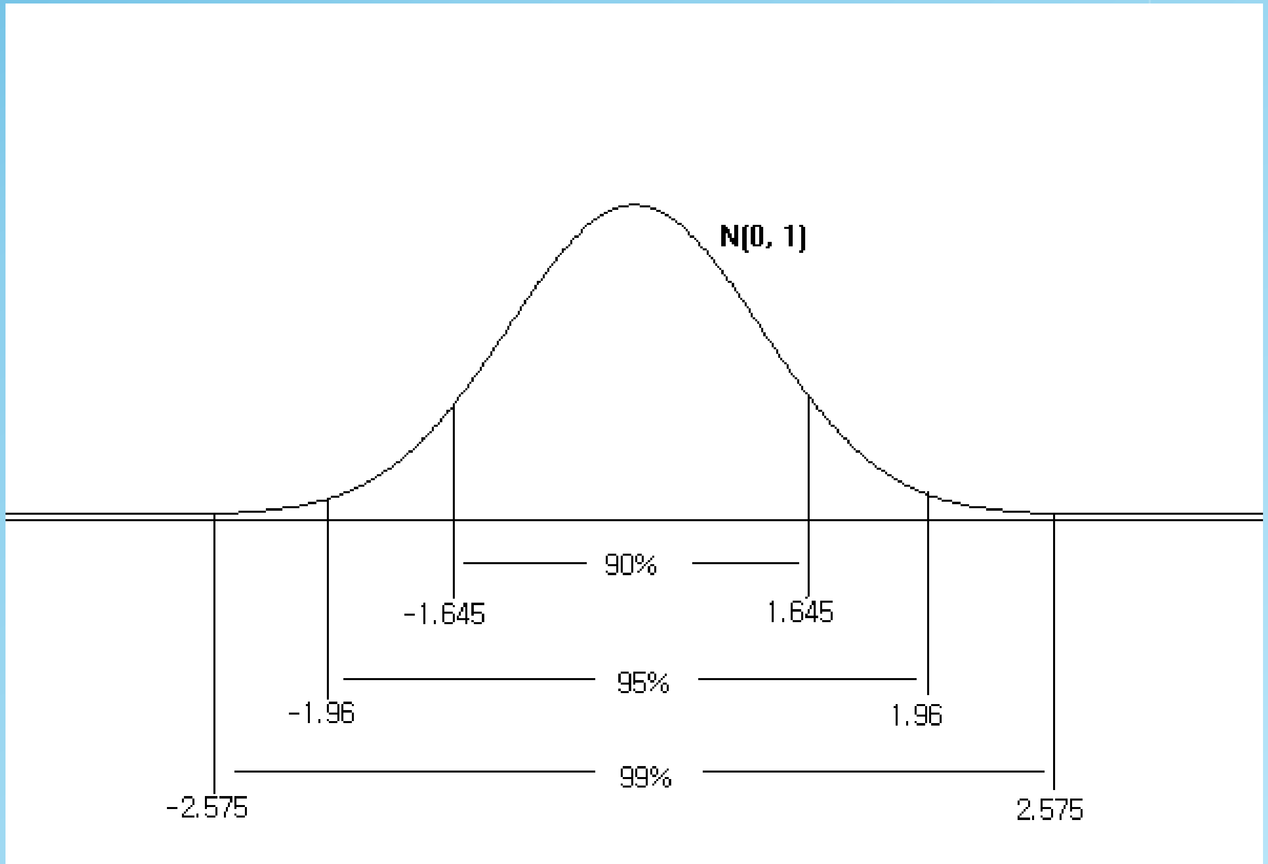


3) In the second option at the bottom of the graph screen, type p = 0.95 in the right box and click the [Execute] button.

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Frequently used percentile





5.4 Continuous Random Variable

5.4.1 Normal Distribution

❖ Probability Calculation of Normal Random Variable

- If X is a normal random variable with the mean μ and variance σ^2, i.e. $X \sim N(\mu, \sigma^2)$,

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

5.4 Continuous Random Variable

[Example 5.4.4] If the mid-term scores (X) of the Statistics course follows a normal distribution with an average of 70 points and a standard deviation of 10 test results X, calculate the following probabilities. Check the calculated value by using 『eStatU』.

1) $P(X < 94.3)$ 2) $P(X > 57.7)$ 3) $P(57.7 < X < 94.3)$

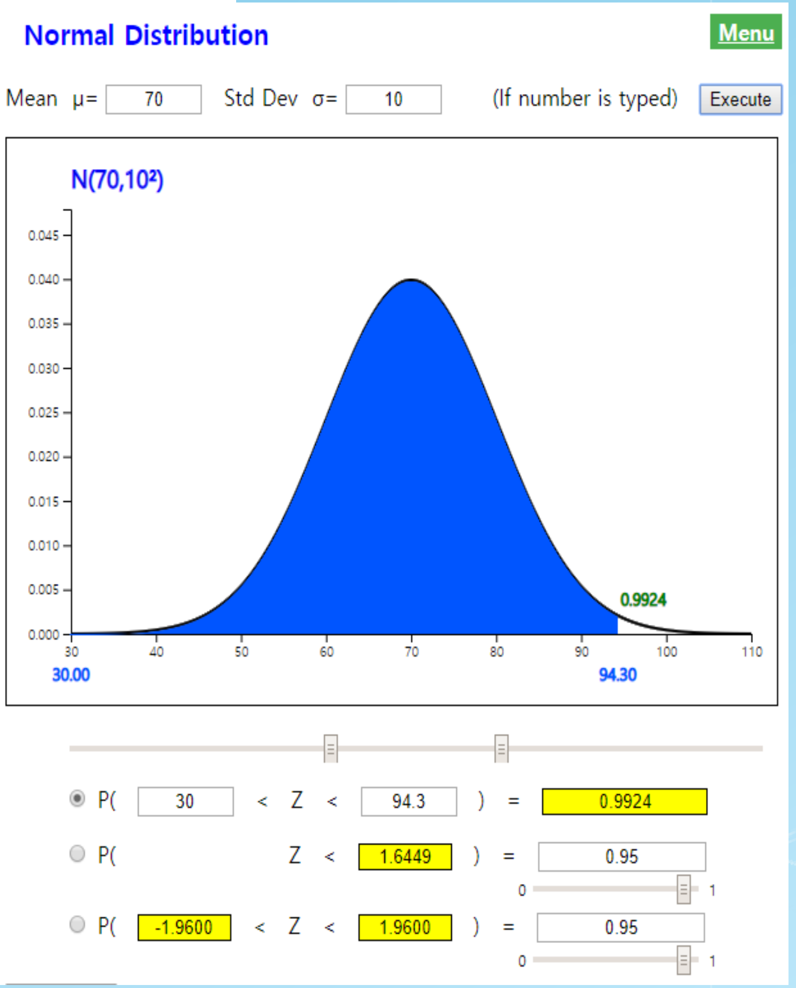
<Answer>

1) $P(X < 94.3) = P\left(\frac{X-70}{10} < \frac{94.3-70}{10}\right) = P(Z < 2.43) = 0.9925$

2) $P(X > 57.7) = P\left(\frac{X-70}{10} > \frac{57.7-70}{10}\right) = P(Z > -1.23) = 0.8907$

3) $P(57.7 < X < 94.3) = P\left(\frac{57.7-70}{10} < \frac{X-70}{10} < \frac{94.3-70}{10}\right)$

$= P(-1.23 < Z < 2.43) = 0.8832$



5.4 Continuous Random Variable

[Example 5.4.5] In [Example 5.4.4], obtain the following percentiles by using 「eStatU」.

1) What is the 95% percentile of the mid-term test scores?
2) What is the 95% percentile of two sided of the mid-term scores?

<Answer>

1) The 95 percentile $P(Z < ?) = 0.95$ in N(0,1) is 1.645, so the percentile in the $N(70,10^2)$ is 70 + 1.645 × 10 = 86.45.

2) The 95 percentile of two-sided type $P(? < Z < ?) = 0.95$ in N(0,1), which can be calculated from $P(Z < ?) = 0.975$, is 1.96. So the two-sided 95% percentile interval is [70 − 1.96 × 10, 70 + 1.96 × 10], I.e., [50.4, 89.6].

- To obtain a percentile for a general normal distribution using 「eStatU, enter the mean as 70 and the standard deviation as 10 on the screen in <Fig. 5.4.13>. 1) enters 0.95 in the second right box of options under the graph screen and press the [Execute] button to display the 95% percentile 86.4485.

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Binomial probability calculation approximately

- **In the case of large n in a binomial distribution, a direct probability calculation is not possible. In such cases, a normal distribution with an average of np and a variance of np (1-p) is used to calculate an approximated probability.**

5.4 Continuous Random Variable

5.4.1 Normal Distribution – Binomial probability calculation approximately

[Example 5.4.6] The defect rate of products produced in a factory is 5 per cent. One day, a sample of 100 products is collected
1) What is the probability that there are less than two defective products?
2) What is the probability that there are defectives between 3 and 7?

<Answer>

- X is a binomial distribution of n = 100, p = 0.05. The mean is np = 100 × 0.05 = 5, and the variance is np(1-p) = 100 × 0.05 x (1-0.05) = 4.75. The probability calculation using N(5, 4.75) is as follows.

1) $P(X \le 2) = P\left(Z \le \frac{(2-5)}{\sqrt{4.75}} \right) = P(Z \le -1.376) = 0.0845$

2) $P(3 \le X \le 7) = P\left(\frac{(3-5)}{\sqrt{4.75}} \le Z \le \frac{(7-5)}{\sqrt{4.75}} \right)$

$= P(-0.918 \le Z \le 0.918) = 0.642$

Thank you