연립 방정식

예를 들어서

3x + 2y - 3 = 0(식1)과

x - y - 1 = 0(식2)

와 같이 2개 이상의 미지수를 포함하고 있는 방정식의 조를 연립 방정식이라고 한다.

그리고, x=1, y=0은 (식1)과 (식2)를 동시에 만족하고 있는 데, 이때 x=1, y=0를 연립 방정식의 해(또는 근)이라고 하고, 이 근을 구하는 것을 "연립 방정식을 푼다"라고 한다.

문제 :물건의 가격과 수량에 관한 문제

어느 슈퍼 마켓의 월드콘 가격이 1,000원이고 우유 가격이 500원이다.

철수 엄마는 월드콘과 우유를 합쳐 7개를 구입하고 4,500원을 지불하였다.

구입한 월드콘과 우유의 개수를 각각 구하시오.

풀이 :

1. 변수(미지수) 정하기

구입한 월드콘 개수를 x, 우유의 개수를 y 라고 가정한다.

2. 연립방정식 만들기

식1. x+y = 7

식2. 1000*x + 500*y = 4500

3. 연립방정식 풀기

식1에서 x = 7 - y이므로 이를 식2의 x 대신 7-y를 대입하면 다음과 같다.

1000*(7-y) + 500y = 4500→7000 - 1000y + 500y = 4500→500y = 2500

∴ y = 5

y=5를 식 x=7-y에 대입하면 x = 2

∴ x=2, y=5

답은 월드콘 2개, 우유 5개이다.

문제 :다음 연립 방정식을 풀어라.(가감법)

가감법이란 두 방정식의 같은 변끼리 더하거나 빼어서 하나의 미지수를 소거하여 해를 구하는 방식을 말한다.

(식2)에 3배수를 곱하여 (식1)에서 뺄셈을 하여 풀어 보도록 한다.

항목	식1	식2
문제01	3x + 2y - 3 = 0	x - y - 1 = 0
풀이	(식1) - 3 * (식2) = (3x + 2y - 3) - 3 * (x - y -1) = 3x + 2y - 3 - 3x + 3y + 3 = 5y = 0 즉, y = 0이다. 이것을 식1에 대입하면 3x + 2y - 3 = 3x + 2*0 - 3 = 0 x = 1이 된다.	

	정답 : x = 1, y = 0

문제 :다음 연립 방정식을 풀어라.(연립 방정식의 불능과 부정)

항목	식1	식2	표현
문제01	2x + 4y + 1 = 0	4x + 8y + 3 = 0	불능(직선이 평행하다.)
문제02	2x + 4y + 1 = 0	4x + 8y + 2 = 0	부정(해가 너무 많다.)

문제 :다음 연립 방정식을 풀어라.

항목	식1	식2	식3	정답
문제01	3x + 2y – z = 12	x + y + z = 6	x -2y – z = -2	x = 3, y = 2, z = 1
문제02	x + 2y = 0	2y + 3z = 4	x + 3z = 8	x = 2, y = -1, z = 2

문제 01 풀이 :

연립 방정식

3*x + 2*y - z = 12 (식 1)
x + y + z = 6 (식 2)
x - 2*y - z = -2 (식 3)

(식 1) + (식 2) -->(식 4)
4*x + 3*y = 18

(식 2) + (식 3) -->(식 5)
2*x - y = 4

(식 4) - 2*(식 5)
4*x + 3*y = 18 (식 6)
4*x - 2*y = 8 (식 7)

(식 6) - (식 7)
5y = 10 ∴ y = 2 이고, (식 7)에 대입하면
4*x - 2*2 = 8
4*x = 12 ∴ x = 3

구한 x, y를 식(1)에 대입
3*3 + 2*2 - z = 12
∴ z = 1

점과 좌표

수직선 위의 두 점 사이의 거리

수직선 상에 놓여 있는 두 점 사이의 거리는 다음과 같이 구하면 된다

수직선상의 두 점 사이의 거리
두 점 A(x1)과 B(x2) 사이의 거리 $\overline{AB}$는 x2 >= x1일때 $\overline{AB} = x2 - x1$이고, x2 < x1일때 $\overline{AB} = x1 - x2$이고, 이것을 하나의 식으로 표현하면 $\overline{AB} = \lvert x2 - x1 \rvert$이다.

문제 :두 점A(-3)과 B(6) 사이의 거리를 구하시오.

$\overline{AB} = \lvert x2 - x1 \rvert = \lvert 6 - (-3) \rvert = 9$이다.

좌표 평면 위의 두 점 사이의 거리

두 점 사이의 거리 공식은 피타고라스 정리에 의하여 다음과 같은 공식이 성립한다.

좌표 평면 위의 두 점 사이의 거리
두 점 A(x1, y1)과 B(x2, y2) 사이의 거리 $\overline{AB}$는 다음과 같이 구할 수 있다. $\overline{AB} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$이다.

문제 : 두 점A(2, 7)과 B(-1, 3) 사이의 거리를 구하시오.

$\overline{AB} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} = \sqrt{(2 - (-1))^2 + (7 - 3)^2} = 5$이다.

문제 : 두 점A(1, 2)과 B(4, 5)에서 같은 거리에 있는 직선 3x + y = 2위의 점의 좌표를 구하여라.

풀이 :

구하려는 점의 좌표를 P(a, b)라고 가정하면 $AP = BP \Rightarrow AP^2 = BP^2$이다.

$(a - 1)^2 + (b - 2)^2 = (a - 4)^2 + (b - 5)^2$이므로 이것을 정리하면 a + b = 6이다.

또한, 점 P(a, b)는 직선 3x + y = 2위에 있으므로 3a + b = 2이다.

풀이
a + b = 6…식1 3a + b = 2… 식2 에서 3*식1 – 식2를 이용하면 2b=16이므로 b=8이 된다. b=8을 식1에 대입하면 a=-2가 된다.

∴P(a, b) = P(-2, 8)

유제 : 두 점A(1, 2)과 B(3, 4)로 부터 동일한 거리에 있는 x축위의 점의 좌표를 구하시오.

풀이 :P(5, 0)

구하려는 점의 좌표를 P(a, 0)라고 가정하면 $\mathrm{AP} = \mathrm{BP} \ \Rightarrow \mathrm{AP}^2 = \mathrm{BP}^2$이다.
$(a-1)^2 + (0-2)^2 = (a-3)^2 + (0-4)^2$이므로 이것을 정리하면 4a = 20이다.
따라서 a = 5이므로, P(a, 0) = P(5, 0)이다.

유제 : 두 점 A(1, 2)과 B(3, 4)로 부터 동일한 거리에 있는 y축위의 점의 좌표를 구하시오.
풀이 : 없음

직선의 방정식

어떠한 조건이 주어졌을 때 직선의 방정식을 구하는 문제를 살펴 보도록 하자.

기본적인 직선의 방정식

기본적인 직선의 방정식
(1) x 절편 a, y축에 평행한 직선 ⇔ x = a y 축의 방정식 ⇔ x = 0 (2) y 절편 b, x 축에 평행한 직선⇔ y = b x 축의 방정식 ⇔ y = 0 (3) 기울기 a, y 절편 b인 직선 ⇔ y = ax + b (4) 기울기가 m이고 점(x1, y1)을 지나는 직선의 방정식⇔ y - y1 = m(x - x1) (5) 두 점 (x1, y1), (x2, y2)를 지나는 직선의 방정식 ⇔ y - y1 = $\frac{y2-y1}{x2-x1}$(x - x1)

문제 : 다음 조건을 만족하는 직선의 방정식을 구하여라.

항목	문제	정답
문제01	기울기가 2이고, 점(1, 3)을 지난다.	y = 2x + 1
문제02	기울기가 -3이고, 점(1, 2)을 지난다.	y = -3x + 5

문제01 풀이 :기본적인 직선의 방정식 (4번)에 의하면 다음과 같이 풀어낼 수 있다.
y - 3 = 2(x - 1)
y = 2x - 2 + 3 = 2x + 1

문제 : 다음 두 점 을 지나는 직선의 방정식을 구하여라

항목	문제	정답
문제01	점(2, 1), 점(4, 5)	y = 2x - 3
문제02	점(2, 1), 점(1, 7)	y = -6x + 13

도형의 이동

평행 이동

평행 이동

변환이란 점을 다른 점으로 이동시키거나 도형을 다른 도형으로 옮길 때, 또는 식을 형식의 식으로 바꾸는 것을 의미한다.
특히, 좌표 평면 위에서 도형의 모양과 크기를 바꾸지 않고 그 위치만 바꾸는 변환을 합동 변환이라고 한다.
합동 변환 중에서 평행 이동과 대칭 이동을 다뤄 보기로 한다.

평행 이동
점(x, y)를 x 축을 따라 a 만큼 평행이동하고, y 축을 따라 b 만큼 평행 이동하면 점(x+a, y+b)로 된다. 즉, (x, y) --> (x+a, y+b)

문제

점(1, 2)는 평행 이동 g:(x, y) --> (x+2, y+3)에 의하여 어떤 점으로 옮겨 지는가?

정답

점(1, 2)를 g:(x, y) --> (x+2, y+3)에 대입하면 g:(1, 2) --> (1+2, 2+3) --> (3, 5)
따라서, 점(1, 2)는 평행 이동 g에 의하여 점(3, 5)으로 옮겨 진다.

문제

점(2, 3)을 점(1, 5)에 대응 시키는 평행 이동을 나타내는 변환 g를 구하여라.

정답

g:(2, 3) --> (2+a, 3+b) = (1, 5) 이므로, 연립 방정식을 풀면 ∴ a = -1, b = 2이다.
따라서, g:(x, y) --> (x-1, y+2) 이다.

도형의 평행 이동
방정식 f(x, y)=0 으로 주어진 도형을 x 축 방향으로 p, y 축 방향으로 q 만큼 평행 이동한 도형의 방정식은 다음과 같다. f(x - p, y - q)=0

문제

직선 2x+3y+1=0에 대해서 다음 조건에 맞는 방정식을 구하여라.

(1) x축의 양의 방향으로 1만큼 평행 이동한 직선
(2) y축의 음의 방향으로 2만큼 평행 이동한 직선
(3) x축 방향으로 1, y축 방향으로 -2만큼 평행 이동한 직선

정답

(1) x대신 (x-1)을 대입하면 되므로 2(x-1)+3y+1=0
∴ 2x+3y-1=0

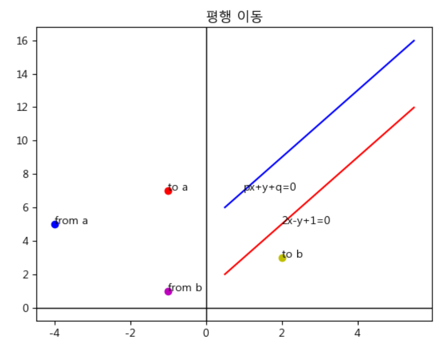
(2) y대신 (y+2)을 대입하면 되므로 2x+3(y+2)+1=0
∴ 2x+3y+7=0

(3) x대신 (x-1), y대신 (y+2)을 대입하면 되므로 2(x-1)+3(y+2)+1=0
∴ 2x+3y+5=0

문제

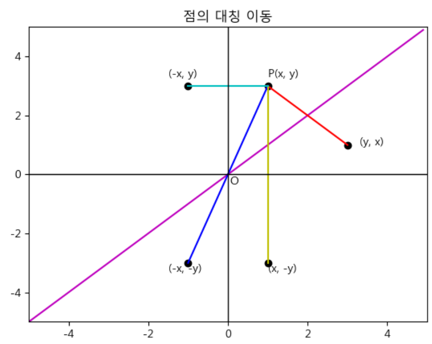
점(-1, 1)을 점(2, 3)으로 옮기는 평행 이동 f를 f:(x, y) --> (x+a, y+b)라고 할 때,
(1) 상수 a, b의 값을 구하여라.
(2) 평행 이동 f에 의하여 점(-4, 5)는 어떤 점으로 옮겨 지는가?
(3) 직선 px+y+q=0이 평행 이동 f에 의하여 2x-y+1=0으로 이동할 때, p, q의 값을 구하여라.

정답

(1) a = 3, b = 2 (그림에서 보라점, 노랑점) (2) f:(x, y) --> (x+3, y+2)에서 f:(-4, 5) --> (-4+3, 5+2) --> (-1, 7) (그림에서 파랑점, 빨강점) (3) 직선 px+y+q=0에 x대신 (x-3), y대신 (y-2)를 대입하면 직선 p(x-3) + (y-2) +q=0 ∴px + y - 3p + q - 2=0 이것이 2x-y+1=0와 동일해야 함으로 ∴ p = -2, q = -5 직선 px+y+q=0 : 그림 파랑 직선 직선 2x-y+1=0 : 그림 빨강 직선	 parallel_move_01.py

대칭 이동

점의 대칭 이동

항목	설명
그림에서 알 수 있는 바와 같이 좌표 평면 위의 임의의 점 P(x, y)에서의 x축, y축, 원점, y=x 등에 대하여 대칭 이동 시킨 점은 각각 (x, -y), (-x, y), (-x, -y), (y, x)이다. 이들에 대한 대칭 이동을 정리하면 다음과 같다. x축에 대한 대칭 이동 g:(x, y) --> (x, -y) y축에 대한 대칭 이동 g:(x, y) --> (-x, y) 원점에 대한 대칭 이동 g:(x, y) --> (-x, -y) y = x에 대한 대칭 이동 g:(x, y) --> (y, x)	 symmetry_move_01.py

문제

점(2, 5)는 x축, y축, 원점, y=x에 대한 대칭 이동에 의하여 어떤 점으로 옮겨 지는가?

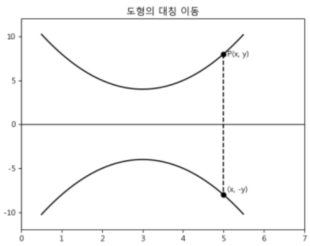
정답

x축 : (2, 5) --> (2, -5)
y축 : (2, 5) --> (-2, 5)
원점 : (2, 5) --> (-2, -5)
y=x : (2, 5) --> (5, 2)

도형의 대칭 이동

도형의 대칭 이동을 생각해 보자.

<table>
<tr><th>항목</th><th>설명</th></tr>
<tr><td>방정식 f(x, y)=0이 나타내는 도형을 다음과 같이 대칭 이동한 도형의 방정식은 다음과 같다.

x축에 대한 대칭 이동 : f(x, -y)=0
y축에 대한 대칭 이동 : f(-x, y)=0
원점에 대한 대칭 이동 : f(-x, -y)=0
y = x에 대한 대칭 이동 : f(y, x)=0</td><td>

symmetry_move_02.py</td></tr>
</table>

문제

다음 도형을 x축, y축, 원점, y = x에 대하여 각각 대칭 이동한 도형의 방정식을 구하여라.

(1) 2x+3y-1=0
(2) $y = x^2 - 2x$

힌트 : x축에 대한 대칭 이동은 y대신 -y를 입력하여 풀면 된다.

항목	x축	y축	원점	y = x
문제01	2x-3y-1=0	2x-3y+1=0	2x+3y+1=0	3x+2y-1=0
문제02	$y = -x^2 + 2x$	$y = x^2 + 2x$	$y = -x^2 - 2x$	$y = 1 \pm \sqrt{x+1}$

문제

직선 y=x-3 은 변환 f:(x, y) --> (y-2, -x+1)에 의하여 어떤 도형으로 옮겨 지는 가?

문제 풀이

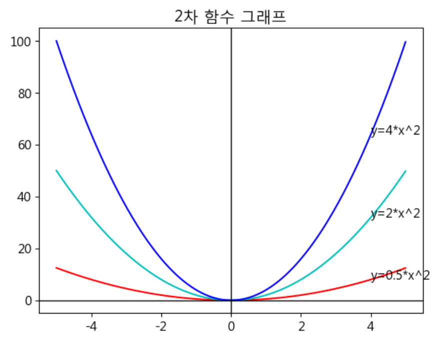
X= y-2, Y = -x+1 으로 놓고, x 와 y 에 대하여 풀면
x = -Y + 1, y = X + 2 이므로 이것을 y=x-3 에 대입하면
X + 2 = -Y + 1 - 3 ∴ Y = -X - 4

따라서, 구하는 도형은 y = -x -4 이다.

2차 함수의 그래프

2차 함수는 다음과 같은 그래프의 성질을 가지고 있다.

$y = ax^2$ 그래프의 성질
(1) 원점을 꼭지점으로 하고, y 축을 축으로 하는 포물선이다. (2) a > 0 이면 아래로 볼록하고, a < 0 이면 위로 볼록하다. (3) a 의 절대 값이 클 수록 y 축 쪽으로 오므라진다.

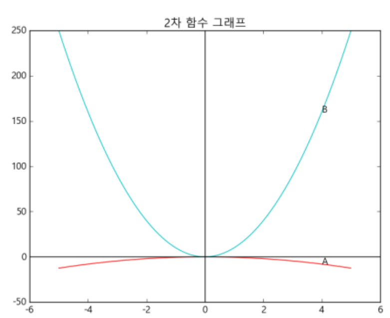


two_curve_01_image_01.png

문제 :

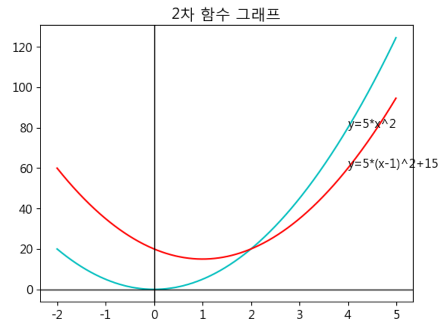
오른쪽 그림에서 포물선 $y = ax^2$그래프가 A와 같이 그려질 때 포물선 B를 나타내는 방정식은 ?정답 :②

$$①\ y = 2ax^2\ \ ②\ y = -2ax^2\ ③\ y = 3ax^2 ④\ y = -ax^2 ⑤\ y = -\frac{1}{2}ax^2$$



two_curve_02_image_01.png

$y = a(x - m)^2 + n$ 그래프의 성질
$y = ax^2$ 의 그래프를 x 축에 따라 m 만큼, y 축에 따라 n 만큼 평행 이동한 그래프이다. $y = ax^2$ 의 그래프와 합동인 포물선이고, $y = a(x - m)^2 + n$ $\Rightarrow$ 꼭지점(m, n), 축 : x = m 이다.



two_curve_03_image_01.png

문제 :
다음 각 함수의 그래프를 그려라

항목	문제	정답
$y = 2x^2 + 1$	$y = 2x^2 - 1$	$y = 2(x - 3)^2$
$y = 2(x + 3)^2$	$y = 2(x - 3)^2 + 1$	$y = 2(x + 3)^2 - 1$

$y = ax^2 + bx + c$ 그래프
이 함수는 다음과 같은 꼴로 변형할 수 있다 $y = a(x - m)^2 + n$ $\Rightarrow$ 꼭지점(m, n), 축 : x = m 이다.

문제 :함수 $y = -x^2 + 4x - 3$의 그래프를 그려라

정답
$y = -x^2 + 4x - 3 = -(x^2 - 4x) - 3 = -(x^2 - 4x + 4 - 4) - 3 = -(x^2 - 4x + 4) + 4 - 3 = -(x - 2)^2 + 1$

포물선의 방정식
(1) 꼭지점이 주어지는 경우 $\Rightarrow y = a(x - m)^2 + n$을 이용한다. (2) 세 점이 주어지는 경우 $\Rightarrow y = ax^2 + bx + c$ 을 이용한다. (3) x 축과의 교점이 주어지는 경우 $\Rightarrow y = a\left(x - \alpha\right)(x - \beta)$ 을 이용한다.

문제 :

꼭지점이 점(-2, 3)이고, 점(1, -6)을 지나는 포물선의 방정식을 구하여라.
단, 축은 y축에 평행하다.

정답
$y = -x^2 - 4x - 1$

문제 :

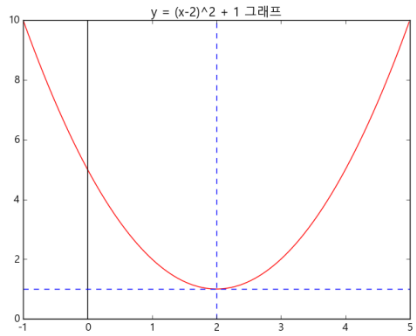
다음 조건을 만족시키는 포물선의 방정식을 구하여라
단, 축은 y축에 평행하다.

항목	설명	정답
문제01	세 점 (1, 1), (-1, 9), (2, 3)을 지난다.	$y = 2x^2 - 4x + 3$
문제02	세 점 (1, 0), (4, 0), (2, 2)을 지난다.	$y = -x^2 + 5x - 4$
문제03	두 점 점 (2, 0), (4, 0)을 지나고, 꼭지점의 y 좌표가 -2이다	$y = 2x^2 - 12x + 16$

문제 02
a+b+c=0 (식 1) 16a+4b+c=0 (식 2) 4a+2b+c=2 (식 3)   8a+4b+2c=4 (식 5) 2a+2b+2c=0 (식 4) (식 3)-(식 4) 2a -c = 2(식 6) (식 2)-(식 5) 8a-c=-4 (식 7) (식 6)과 (식 7)을 연립하여 풀면 a=-1 c=-4 b=5

이차 함수의 최대 최소

이차 함수 $y = x^2 - 4x + 5$에 있어서 최소 값은 다음과 같이 구할 수 있다.
$y = x^2 - 4x + 5 = (x-2)^2 + 1$으로 변형이 되므로 x=2에서 최소 값 1을 가진다.



two_curve_minmax.png

일반적으로 다음과 같이 정리할 수 있다.

함수 $y = ax^2 + bx + c$ 의 최대 최소
함수 $y = ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2-4ac}{4a}$ $(a \neq 0)$에서 a > 0 일때 x = $-\frac{b}{2a}$ 에서 최소 값 $-\frac{b^2-4ac}{4a}$ 을 가진다. a < 0 일때 x = $-\frac{b}{2a}$ 에서 최대 값 $-\frac{b^2-4ac}{4a}$ 을 가진다.

문제 :다음 이차 함수의 최대 값또는 최소 값을 구하시오.

항목	설명	정답
문제01	$y = 2x^2 + 8x + 5$	x = -2에서 최소 값 -3을 가진다.
문제02	$y = -2(x+1)(x-3)$	x = 1에서 최대 값 8을 가진다.

삼각 함수

삼각 함수의 정의

반지름의 길이가 r인 원호 위에 길이가 r인 호AB를 잡을 때, 이 호에 대한 중심각 AOB는 반지름의 길이 r에 상관 없이 항상 일정하게 된다.
이 중심각 AOB의 크기를 1호도(radian)라고 하고, 이것을 단위로 하는 각의 측정법을 호도법이라고 한다.

호도법과 60 분법
$\pi\mathrm{rad} = 180\,° = \begin{cases} 1\mathrm{rad} = \frac{180\,°}{\pi} \\ 1\,° = \frac{\pi}{180}\mathrm{rad} \end{cases}$

일반적으로 단위의 길이 rad는 생략하고 쓰지 않는 것이 보통이다.
즉, π = 180도이다.

비례식 다뤄 보기

60분법과 호도법을 다루고자 할 때 비례식을 사용하면 쉽게 접근할 수 있다.

비례식은 "내항의 곱은 외항의 곱과 같다."이다.
예를 들어서 1 : 3 = 2 : x 라는 비례식에서 x의 값을 구하려면
"내항의 곱 = 외항의 곱"이므로 1*x = 2*3 와 같은 식이 성립한다.
이것을 이용하여 x에 대하여 풀어 보면 x = 6이 된다.

문제 : 비례식

45 °는 pi/4인가요?
pi : 180 = x : 45
180 * x = 45 * pi
x = 45 * pi / 180 = pi / 4

문제

다음 60분법은 호도법으로 호도법은 60분법으로 변경해보세요.

항목				
문제	36 °	120 °	$\frac{\pi}{3}$rad	2rad
정답	$\frac{\pi}{5}$rad	$\frac{2\pi}{3}$rad	60 °	$\frac{360}{\pi}$ °

각도와 호도법 환산

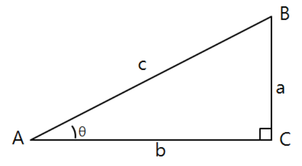
동일한 요령으로 다음과 같은 환산표를 얻을 수 있다.

60분법	0 °	30 °	45 °	60 °	90 °
호도법	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$

삼각 함수의 정의

삼각비의 정의

각 ∠C = 90 °(직각)이고, 각 ∠A = θ인 직각 삼각형 ABC에서 세변의 a, b, c 사이의 비의 값을 다음과 같이 정의한다.

$\sin\theta = \frac{a}{c}$	$\cos\theta = \frac{b}{c}$	$\tan\theta = \frac{a}{b}$	
$\csc\theta = \frac{c}{a}$	$\sec\theta = \frac{c}{b}$	$\cot\theta = \frac{b}{a}$	

특수각의 삼각비의 값

일반적으로 많이 사용되는 특수 각에 대한 삼각비는 다음과 같은 것들이 있다.
삼각비의 정의에 의하면 sin30은 cos(90-30)=cos(60)과 동일한 값을 가진다.

항목	0 °	30 °	45 °	60 °	90 °
sinθ	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1
cosθ	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
tanθ	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	∞

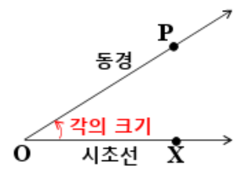
일반각의 삼각 함수

일반각의 삼각 함수의 정의

그림과 같은 ∠XOP가 있을 때, 이 각이 변 OP가 처음에는 OX의 위치에 있었던 것이 회전하여 현재의 위치에 와서 만들어진 것이라고 생각하면 편리할 때가 많다.

이때 OX를 ∠XOP의 시초선, OP를 ∠XOP의 동경이라고 한다.

그림과 같이 동경 OP가 x축의 양의 방향과 이루는 각을 θ라고 할때, 반시계 방향이면 양의 각, 시계 방향이면 음의 각이라고 한다.



일반각의 표시

동경 OP가 시초선 OX와 양의 방향과 이루는 최소의 양의 각을 α °라고 하면, 일반각 θ °는 다음과 같다.
θ ° = 360 °n + α °(단, n=0, ±1, ±2, …)

일반 각의 삼각 함수의 정의

일반적으로 일반각 θ의 삼각 함수를 다음과 같이 정의한다.

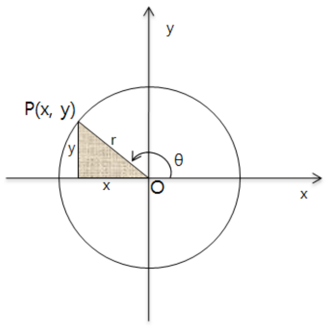
일반 각의 삼각 함수의 정의

그림과 같이 동경 OP가 x축의 양의 방향과 이루는 각은 θ라고 할 때,

$$\sin\theta = \frac{y}{r}, \cos\theta = \frac{x}{r}, \tan\theta = \frac{y}{x},$$

$$\csc\theta = \frac{r}{y}, \sec\theta = \frac{r}{x}, \cot\theta = \frac{x}{y},$$

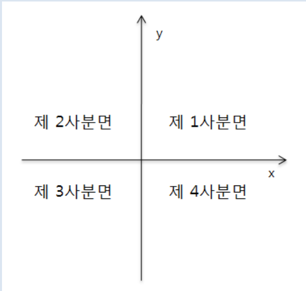
로 나타내고, 이들을 각각 θ의 사인 함수, 코사인, 탄젠트, 코시이컨트, 시이컨트, 코탄젠트 함수라고 하며, 이들을 통틀어서 삼각 함수라고 한다.



삼각 함수의 값의 부호

삼각 함수의 값은 어느 사분면인가에 따라서 다음과 같이 부호가 달라진다.

축	제 1사분면	제 2사분면	제 3사분면	제 4사분면
x	양수	음수	음수	양수
y	양수	양수	음수	음수
sin	양수	양수	음수	음수
cos	양수	음수	음수	양수
tan	양수	음수	양수	음수



다음 각 물음에 답하여라

(1) 원점과 점 P(-4, 3)을 맺는 선분을 동경으로 하는 각을 θ라고 할 때, $\sin\theta$, $\cos\theta$, $\tan\theta$의 값을 각각 구하여라.
(2) sin45 °, cos840 °, tan(-405 °)의 값을 구하여라.

정답	$\sin\theta$	$\cos\theta$	$\tan\theta$	
문제01	$\frac{3}{5}$	$-\frac{4}{5}$	$-\frac{3}{4}$	
문제02	$\frac{1}{\sqrt{2}}$	$-\frac{1}{2}$	-1	

삼각 함수의 기본 성질

삼각 함수의 기본 공식

(1) 역수 관계

$$\csc\theta = \frac{1}{\sin\theta}, \quad \sec\theta = \frac{1}{\cos\theta}, \quad \cot\theta = \frac{1}{\tan\theta}$$

(2) 상제 관계

$$\tan\theta = \frac{\sin\theta}{\cos\theta}, \quad \cot\theta = \frac{\cos\theta}{\sin\theta}$$

(3) 제곱 관계

$$\sin^2\theta + \cos^2\theta = 1, \quad 1 + \tan^2\theta = \sec^2\theta, \quad 1 + \cot^2\theta = \csc^2\theta$$

다음 각 물음에 답하여라

(1) θ가 예각이고, $\cos\theta = \frac{4}{5}$ 일 때, $\sin\theta$, $\tan\theta$의 값을 각각 구하여라.
(2) θ가 제 3사분면의 각이고, $\tan\theta = \frac{5}{12}$가 일 때, $\sin\theta$, $\cos\theta$의 값을 구하여라.

정답	$\sin\theta$	$\tan\theta$		
문제01	$\frac{3}{5}$	$\frac{3}{4}$		
문제02	$-\frac{5}{13}$	$-\frac{12}{13}$		

다음 각 물음에 답하여라

$\sin\theta + \cos\theta = \frac{1}{2}$일 때, 다음 각 값을 구하여라.(힌트 : 양변을 제곱해보세요.)

관련 힌트 공식

$(a+b)^2 = a^2 + 2ab + b^2$

$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

문제	$\sin\theta\cos\theta$	$\sin^3\theta + \cos^3\theta$	$\sin^4\theta + \cos^4\theta$	$\tan\theta + \cot\theta$	$\sin\theta - \cos\theta$
정답	$-\frac{3}{8}$	$\frac{11}{16}$	$\frac{23}{32}$	$-\frac{8}{3}$	$\pm\frac{\sqrt{7}}{2}$

문제 01 번 풀이

주어진 공식에 대하여 양변을 제곱하면

$(\sin\theta + \cos\theta)^2 = \sin^2\theta + 2\sin\theta\cos\theta + \cos^2\theta = \frac{1}{4}$

이다.

여기서 $\sin^2\theta + \cos^2\theta = 1$이므로

$$\sin^2\theta + 2\sin\theta\cos\theta + \cos^2\theta = 1 + 2\sin\theta\cos\theta = \frac{1}{4}$$

$$2\sin\theta\cos\theta = -\frac{3}{4}$$

따라서

$$\sin\theta\cos\theta = -\frac{3}{8}$$

이다.

문제 04 번 풀이

$$(\sin^2\theta + \cos^2\theta)^2 = \sin^4\theta + 2\sin^2\theta\cos^2\theta + \cos^4\theta$$

에서 구하고자 하는 항을 좌측으로 옮기면 다음과 같다.

이미 값을 알고 있는 항목들을 대입하여 연산을 수행하면

$$\sin^4\theta+\cos^4\theta = (\sin^2\theta+\cos^2\theta)^2-2\sin^2\theta\cos^2\theta = 1^2-2(\sin\theta\cos\theta)^2 = 1-2\left(-\frac{3}{8}\right)^2 = 1-\frac{18}{64}=\frac{23}{32}$$

와 같은 결과를 얻을 수 있다.

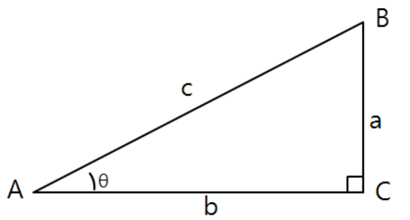
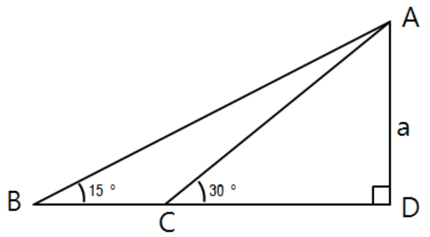
음각의 삼각 함수

각 θ의 동경과 -θ의 동경은 x축에 대하여 대칭이다.
따라서, 다음과 같은 공식이 성립한다.

-θ의 삼각 함수

(1) $\sin(-\theta) = -\sin\theta$
(2) $\cos(-\theta) = \cos\theta$
(3) $\tan(-\theta) = -\tan\theta$

문제 : 다음 물음에 대하여 각각 답하여라.

문제01	(1) ∠C = 90 °인, 직각 삼각형 ABC에서 두변 AB, BC의 길이 c, a 사이에 c=3a인 관계가 성립한다고 할 때, sinA, cosA, tanA의 값을 구하여라.	
문제02	(2) 우측 그림을 참조하여 sin15 °를 구하여라.	

풀이 :

(1) AB = c, BC = a이므로, CA = b라고 하면 문제의 조건식으로부터 c = 3a와 피타고라스 정리에 의하여 $a^2+b^2=c^2$이다.
여기서 c를 소거하면 $b=2\sqrt{a}$ 이다.
따라서, $\sin A = \frac{a}{c}=\frac{a}{3a}=\frac{1}{3}$
동일한 방식으로 각각 $\cos A = \frac{2\sqrt{2}}{3}, \tan A = \frac{\sqrt{2}}{4}$가 된다.

(2) AD = a으로 놓으면 ∠C = 30 °이므로 CD = $\sqrt{3a}$, AC=2a이다.
한편 삼각형 ABC는 이등변 삼각형이다.
따라서, BC = AC = 2a이므로 BD = BC + AD = 2a + $\sqrt{3a}$가 된다.

피타고라스 정리에 의하여 AB $= \sqrt{AD^2 + BD^2} = \sqrt{a^x + (2a + \sqrt{3}a)^2} = (\sqrt{6} + \sqrt{2})a$가 된다.

따라서 $\sin 15° = \frac{AD}{AB} = \frac{a}{(\sqrt{6}+\sqrt{2})a} = \frac{(\sqrt{6}-\sqrt{2})a}{4}$가 된다.

벡터(vector)

벡터는 공간적인 개념을 표현하는 데 매우 편리한 개념이며, 힘/속도/가속도 등의 물리적인 현상을 설명하는 기본적인 도구이다.
패턴 인식에서는 인식 대상이 되는 객체가 특징으로 표현이 되고, 특징은 차원을 가지는 벡터로 표현이 된다.

참조 사이트 : https://docs.sympy.org/latest/modules/physics/vector/vectors.html#vector

벡터(vector)의 기본 성질

벡터의 정의

크기만을 가지고 있는 스칼라(Scalar)와 구별하여 크기와 방향을 가진 물리량을 벡터(Vector)라고 한다.
힘, 모멘트, 위치, 속도, 가속도 등이 우리가 잘 알고 있는 벡터량들이다.
A점을 시점(始點)으로 하고 B점을 종점(終點)으로 하는 화살표로써 벡터 $\overrightarrow{AB}$를 나타내면, 화살표의 길이는 벡터의 크기, 화살표의 방향은 벡터의 방향과 대응된다.

Cartesian 좌표계를 이용한 벡터의 좌표 성분 표시법은 B점의 좌표 (B_1, B_2)에서 A점의 좌표 (A_1, A_2)를 감한 $\overrightarrow{AB} = (B_1 - A_1,\ B_2 - A_2)$ 으로 벡터 $\overrightarrow{AB}$를 나타낸다.

벡터의 크기와 기본 연산

벡터의 좌표 성분 표시에서 $a = (a_1, a_2, a_3)$이라고 할 때, 스칼라 $\sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2}$ 를 벡터 a의 크기(norm)이라고 하고, $|a| = \sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2}$ 로 쓴다.

용어	설명
단위 벡터	크기가 1인 벡터를 의미한다.
위치 벡터	좌표계의 원점을 시점으로 한 벡터를 위치 벡터라고 한다.
영 벡터	방향이 없고, 크기가 0인 벡터를 의미한다.

단위 벡터

벡터의 크기가 1인 벡터를 단위(unit) 벡터라고 한다.
영(0)이 아닌 벡터 v의 단위 벡터 u는 다음과 같이 구한다.
이렇게 주어진 벡터와 방향이 같은 단위 벡터를 구하는 과정을 벡터의 정규화(normalization)라고 한다.

$$u = \frac{v}{|v|}$$

벡터 문제

시작점 P와 끝점 Q가 다음과 같을 때 해당 벡터의 성분과 크기를 구하여라.

문제	정답

P(1, 0), Q(3, 2)	v(2, 2), $2\sqrt{2}$
P(1, 0, 0), Q(3, 2, 1)	v(2, 2, 1), 3
P(4, 0, 1), Q(1, 0, 3)	v(-3, 0, 2), $\sqrt{13}$

방향 여현(direction cosine)

크기가 0이 아닌 벡터를 자신의 크기로 나누어 얻은

$$\frac{\mathbf{a}}{|a|} = \left(\frac{a_1}{|a|}, \frac{a_2}{|a|}, \frac{a_3}{|a|}\right) = (l, m, n)$$

으로 되는 벡터의 각 성분을 "방향 여현"이라고 한다.

분모의 "방향 여현"은 위치 벡터 a가 x, y, z축와 이루는 사이의 각을 α, β, γ라고 할 때

$$l = \cos\alpha = \frac{a_1}{a}, \qquad m = \cos\beta = \frac{a_2}{a}, \qquad n = \cos\gamma \frac{a_3}{a}$$

으로 쓸 수 있다.

또한 $a^2 = {a_1}^2 + {a_2}^2 + {a_3}^2$이므로

$$l^2 + m^2 + n^2 = \cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$$

인 관계를 만족한다.

벡터 a의 방향 여현으로 이루어진 벡터 (l, m, n)은 단위 벡터이므로 e_a로 쓰기로 한다.

$$A = A(l, m, n) = Ae_A$$

문제 : 방향 여현

위치 벡터 r의 좌표가 $(2, 1, \sqrt{3})$으로 주어져 있다.
주어진 벡터의 방향 여현을 구하고, 또 x축과 이루는 각을 구하시오.

문제 풀이 : 방향 여현

$r^2 = 2^2 + 1^2 + \sqrt{3}^2 = 8$이므로, 벡터의 크기 r은 $r = 2\sqrt{2}$이다.
주어진 벡터의 방향 여현은 벡터 방향의 단위 벡터

$$e_r = \frac{\mathbf{r}}{|r|} = \frac{1}{2\sqrt{2}}(2, 1, \sqrt{3}) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{4}, \frac{\sqrt{6}}{4}\right) = (l, m, n)$$

에서 $l = \cos\alpha = \frac{\sqrt{2}}{2}$ 으로 부터, $\alpha = 45°$이다.

벡터(vector)의 내적(inner product)

내적의 정의

차원이 동일한 두 개의 벡터 a, b에 대하여 대응하는 성분별로 곱하는 것을 두 벡터의 내적이라고 한다.
따라서, 내적의 결과는 실수 스칼라(scalar)가 된다.

내적(內積)은 a・b=|a|・|b|・$\cos\theta$으로 정의하기도 한다.
스칼라 |a|와 |b|는 두 벡터의 크기이고, θ는 두 벡터가 이루는 사이 각을 의미한다.

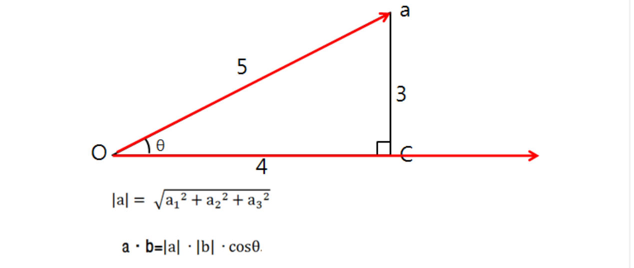
특히 자신과의 내적(θ=0)으로부터 $a \cdot a = a^2 = |\mathbf{a}|^2$ 임을 알 수 있다.

두 벡터의 내적이 0이 될 때, a · b = 0일 때, a와 b는 서로 수직(垂直, orthogonal)이라고 한다.
0이 아닌 두 벡터가 서로 수직이면 반드시 사이 각은 θ=90 °가 된다.

문제 : 벡터의 내적 구하기

벡터 a(4, 3)와 b(6, 0)의 내적을 구해보세요.

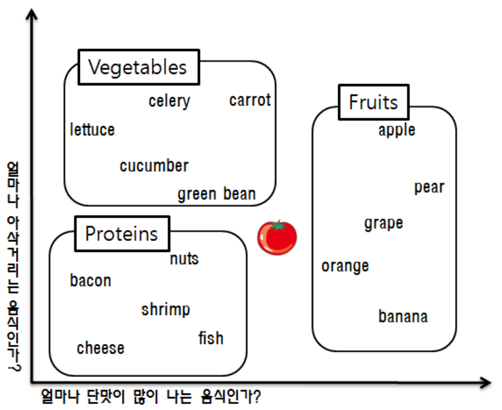
문제 풀이 :



공식에 의하여 $5 * 6 * \frac{4}{5} = 24$이다.

유사도

특정 데이터를 분류하기 위해서는 유사도란 개념을 사용한다.
예를 들어, 여러 가지 먹거리 중에서 "토마토는 어느 부류에 속할 것인가"를 구분하는 문제가 있다고 가정하자.



코사인 유사도

특성 값을 나타내는 벡터 a와 b가 있을 때, 두 벡터 간 각도의 코사인 값을 이용하여 벡터 간의 유사도를 측정하는 방법이다.

코사인 유사도

벡터 a와 b의 코사인 유사도는 다음과 같이 구할 수 있다.

$$\text{cosine similarity} = \cos(\theta) = \frac{a \cdot b}{|a||b|} = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n} (a_i)^2} \times \sqrt{\sum_{i=1}^{n} (b_i)^2}}$$

문제 : 코사인 유사도

다음 벡터 a(1, 0)와 b(1, $\sqrt{3}$)의 코사인 유사도를 구해보세요.

$$\cos(\theta) = \frac{a \cdot b}{|a||b|} = \frac{1 \times 1 + 0 \times \sqrt{3}}{\sqrt{1^2 + 0^2} \times \sqrt{1^2 + \sqrt{3}^2}} = 0.5$$

문제 : 코사인 유사도

다음 벡터 a(1, 2, -1), b(2, 0, 1) 의 코사인 유사도를 구해보세요.

$$\cos(\theta) = \frac{a \cdot b}{|a||b|} = \frac{1 \times 2 + 2 \times 0 - \ -1 \times 1}{\sqrt{1^2 + 2^2 + +(-1)^2} \times \sqrt{2^2 + 0^2 + 1^2}} = \frac{1}{\sqrt{30}}$$

행렬(matrix)

행렬이란 수를 나타내는 문자를 직사각형의 꼴로 배열하여 괄호 ()로 묶어 나타낸 것을 의미한다.
행렬은 가장 새롭고 중요한 수학의 한 대상이다.

행렬(matrix)

다음 표는 김철수와 홍길동의 국어, 영어, 수학 점수를 나타낸 것이다.

학생 이름	국어	영어	수학	
김철수	80	75	80	
홍길동	90	85	95	

위의 표에서 수의 배열을 () 안에 넣어서 나타내면 편리하다.

$$\begin{pmatrix} 80 & 75 & 80 \\ 90 & 85 & 95 \end{pmatrix}$$

이와 같이 나타낸 것을 행렬(matrix)이라고 한다.

행렬에서 가로의 줄을 행(row)이라고 하고, 위쪽에서부터 차례로 제 1행, 제 2행...이라고 한다.
세로의 줄은 열(column)이라고 하고, 왼쪽에서부터 차례로 제 1열, 제 2열...이라고 한다.
일반적으로 m행과 n열로 이루어진 행렬을 m행 n열의 행렬 또는 m X n 행렬이라고 한다.

행의 개수와 열의 개수가 동일한 행렬을 정방 행렬(정사각형 행렬)이라고 하고, 이것을 m차 정방 행렬이라고 한다.
한편, 행렬을 이루고 있는 낱낱의 숫자나 문자를 그 행렬의 성분 또는 원소라고 한다.

삼각 행렬과 대각 행렬

임의의 n X n 정사각형 행렬에서 주대각선 아래 쪽이 모두 0인 행렬을 상삼각 행렬(Upper triangular matrix), 그 반대를 하삼각 행렬(Lower triangular matrix)이라고 한다.

한편 대각선이 아닌 모든 원소가 0인 정사각 행렬은 대각 행렬(Diagonal martix)이라고 한다.

$$U = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}, \quad L = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}, \quad D = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

문제 : 다음 행렬의 꼴을 말하여라.

항목	설명
행렬	$A = (1 \quad 2)$, $B = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$, $C = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}$
풀어 보기	A : 1행 2열의 행렬(1 X 2 행렬) B : 2행 1열의 행렬(2 X 1 행렬) C : 2행 2열의 행렬(2 X 2 행렬)

행렬의 상등

두 행렬 A, B에서 그들의 행의 개수와 열의 개수가 각각 같을 때, 두 행렬 A, B는 같은 꼴 또는 동형이라고 한다.
한편 두 행렬 A, B 가 같은 꼴이고, 대응하는 위치의 성분들이 모두 동일한 값을 가지고 있을 때 두 행렬 A 와 B 는 같다 또는 상등하다라고 하며 A = B 로 표현한다.

행렬의 상등

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \Rightarrow \begin{Bmatrix} a_{11} = b_{11} & a_{12} = b_{12} \\ a_{21} = b_{21} & a_{22} = b_{22} \end{Bmatrix}$$

문제 : 다음 등식이 성립하도록 x, y, z의 값을 구하라.

$$\begin{pmatrix} x+y & -1 \\ -2 & y \end{pmatrix} = \begin{pmatrix} 3 & y-z \\ -y & 2 \end{pmatrix}$$

풀이

x + y = 3 (식1)
-1 = y - z (식2)
-2 = -y (식3)
y = 2 (식4)

(식4) y = 2
(식1) 에서 x = 3 - y = 3 - 2 = 1
(식2)에서 z = y + 1 = 2 + 1 = 3

정답 : x = 1, y = 2, z = 3

유제 : 다음 등식이 성립하도록 x, y, z, u의 값을 구하라.

$$\begin{pmatrix} 2x+y & y-z \\ u+1 & 5 \end{pmatrix} = \begin{pmatrix} 4 & z-u \\ y+3 & x+u \end{pmatrix}$$

풀이
2x + y = 4 (식1) y - z = z -u (식2) u + 1 = y + 3 (식3) 5 = x + u (식4) (식4) 에서 u = 5 - x (식5) (식3) 에서 u = y + 2 (식6) (식5)와 (식6)에서 5 - x = y + 2을 풀면 x + y = 3 (식7) (식1)- (식7)을 연립하여 풀면 2x + y = 4 x + y = 3 --------------- x = 1 따라서, y = 2가 된다. (식4)에서 u = 1이 된다. (식2)에서 z = 3이 된다.

정답 : x = 1, y = 2, z = 3, u = 4

행벡터 · 열벡터

(1, 2) 또는 (2, 3, -1) 등과 같이 순서를 정해서 가로로 나열한 수의 쌍을 행벡터라고 한다.

$\begin{pmatrix} 3 \\ 3 \end{pmatrix}$ 또는 $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$등과 같이 순서를 정해서 세로로 나열한 수의 쌍을 열벡터라고 한다.

여기서, 행벡터 또는 열벡터를 구성하고 있는 각각의 요소를 원소 또는 성분이라고 한다.

원소의 개수를 행벡터 또는 열벡터의 차원이라고 한다.

행렬 벡터 문제

다음 행/열벡터에 대하여 빈 칸을 채워 넣으시오.

행/열 벡터	설명	행/열 벡터	설명
(1, 2)	이차원 행벡터	(2, 3, -1)	?
$\begin{pmatrix} 3 \\ 3 \end{pmatrix}$	?	$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$	삼차원 열벡터

행렬의 곱셈에 대한 성질

단위 행렬

정방(정사각) 행렬 중에서 왼쪽 위에서 오른쪽 아래로의 대각선의 성분은 모두 1이고, 나머지 성분이 모두 0인 행렬을 단위 행렬이라고 하다.
단위 행렬은 일반적으로 I 또는 E 문자를 많이 사용한다.

예를 들면 $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$은 각각 2차 및 3차의 단위 행렬이다.

단위 행렬
A와 E가 각각 n차의 정방 행렬, n차의 단위 행렬일 때 AE = EA = A를 만족한다.

행렬의 곱셈

일반적으로 2 X 2 행렬 A, B에 대하여 두 행렬의 곱 AB는 다음과 같이 정의한다.

행렬의 곱셈

$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ 일 때

$$AB = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

행렬의 곱의 계산
두 행렬 A와 B의 곱은 (A의 열의 개수) = (B의 행의 개수) 일때만 정의할 수 있다.

행렬 문제

다음과 같은 행렬이 있다고 가정할 때 AX=A를 만족하는 X를 구하시오.

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, X = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

문제 풀이

$$AX = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + 2c & b + 2d \\ 3a + 4c & 3b + 4d \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

위의 연립 방정식을 풀면 a = 1, b = 0, c = 0, d = 1이다.

즉, $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

AB ≒ BA

행렬의 곱셈에서도 실수의 곱셈에서와 같이 결합 법칙, 분배 법칙이 성립한다.
그러나, 실수의 곱셈 과는 달리 교환 법칙은 성립하지 않는다.

AB ≒ BA
(1) 두 개의 행렬 A, B에 있어서 곱 AB가 정의되더라도 곱 BA가 정의되지 않는 경우가 있다. (2) 두 개의 행렬 A, B가 같은 꼴의 정방 행렬일 때 곱 AB와 BA가 정의되더라도 AB와 BA의 연산 결과가 같지 않는 경우가 있다.

행렬 문제

두 개의 행렬이 다음과 같다고 하자.
A : 2*2행렬, B : 2*3행렬이라고 하면 AB는 연산이 가능하지만, BA는 연산이 불가능하다.
즉, 2*3행렬 X 2*2행렬에 대한 연산은 존재하지 않는다.

행렬 문제

다음 행렬들에 대하여 행렬의 곱의 모두 구하여라.
힌트 : 두 행렬 A, B의 곱 AB는 (A의 열의 개수) = (B의 행의 개수)일 때만 정의 된다.

항목	설명
행렬	$A = (1 \quad 2)$, $B = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$, $C = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}$
풀어 보기	AB, AC, BA, CB

문제 풀이

항목	설명
접근 방식	A : 1*2, B : 2*1, C : 2*2이므로 AB = (1*2) · (2*1) = 1*-1+2*3 = 4 AC = (1*2) · (2*2) = (1 2) 와 같이 풀이할 수 있다.
AB	$AB = (1 \quad 2)\begin{pmatrix} -2 \\ 3 \end{pmatrix} = (1 \times (-2) + 2 \times 3) = (4) = 4$
AC	$AC = (1 \quad 2)\begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix} = (1 \times 3 + 2 \times (-1) \quad 1 \times 2 + 2 \times 0) = (1\ 2)$
BA	$BA = \begin{pmatrix} -2 \\ 3 \end{pmatrix}(1 \quad 2) = \begin{pmatrix} -2 \times 1 & -2 \times 2 \\ 3 \times 1 & 3 \times 2 \end{pmatrix} = \begin{pmatrix} -2 & -4 \\ 3 & 6 \end{pmatrix}$
CB	$CB = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \times -2 + 2 \times 3 \\ -1 \times -2 + 0 \times 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

행렬 문제

다음과 같은 행렬식이 있다고 가정할 때, a, b, c의 값을 구하여라.(단, a > 0)

$$\begin{pmatrix} a & b & c \\ -3 & -2 & a \end{pmatrix}\begin{pmatrix} 0 & 3 \\ -1 & -2 \\ b & a \end{pmatrix} = \begin{pmatrix} 4 & 2a \\ 2b & -4 \end{pmatrix}$$

문제 풀이

주어진 식의 좌변을 계산해 보면

좌변 $= \begin{pmatrix} a\times 0 + b\times -1 + c\times b & a\times 3 + b\times -2 + c\times a \\ -3\times 0 - 2\times -1 + a\times b & -3\times 3 - 2b\times -2 + a\times a \end{pmatrix} = \begin{pmatrix} -b+bc & 3a-2b+ac \\ 2+ab & -5+a^2 \end{pmatrix} = \begin{pmatrix} 4 & 2a \\ 2b & -4 \end{pmatrix}$

문제 풀이
A : 2*3 B : 3*2 AB = (2*3)·(3*2) = 2*2 a^2 = -4+5 = 1 a = ±1 a는 양수이므로 a = 1이다. 이것을 토대로 풀면 b = 2, c = 3이다.

역행렬(inverse matrix)

영행렬 O가 아닌 임의의 정방 행렬 A와 단위 행렬 E에 대하여 AX = XA = E를 만족하는 행렬 X가 존재한다고 할 때, X를 A의 역행렬이라고 하고, $X = A^{-1}$으로 표기한다.

역행렬의 관계
정사각형 A가 역행렬을 갖고,그 역행렬을 X라고 할 때 다음과 같은 공식이 성립한다. $X = A^{-1} \Leftrightarrow A = X^{-1}$

행렬 문제

다음 행렬의 역행렬을 구하여라.

$$A = \begin{pmatrix} -1 & 2 \\ -3 & 5 \end{pmatrix}$$

문제 풀이

구하는 역행렬을 $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$로 놓으면, AX=E에서

$AX = \begin{pmatrix} -1 & 2 \\ -3 & 5 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -a+2c & -b+2d \\ -3a+5c & -3b+5d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

연립하여 풀어 보면 a=5, b=-2, c=3, d=-1이 된다.

$$A^{-1} = \begin{pmatrix} 5 & -2 \\ 3 & -1 \end{pmatrix}$$

역행렬의 공식
행렬 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$에서 D = ad - bc라 할 때

(1) D 이면 $A^{-1} = \frac{1}{D}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$
(2) D = 0이면 A의 역행렬은 존재하지 않는다.

행렬 문제

다음 행렬이 역행렬이 존재하는 지를 확인하고, 존재하는 경우 그 행렬을 구하시오.

항목	설명	항목	설명
문제01	$A = \begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix}$	문제02	$A = \begin{pmatrix} 3 & -2 \\ -6 & 4 \end{pmatrix}$
정답	$A^{-1} = \frac{1}{2}\begin{pmatrix} 3 & -5 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 1.5 & -2.5 \\ -1 & 2 \end{pmatrix}$	정답	역행렬이 존재하지 않는다.

역행렬과 연립 1차 방정식

연립 1차 방정식

연립 1차 방정식은 행렬을 써서 나타낼 수 있고, 또 행렬을 사용하여 그 해를 구할 수 있다.
수학자 케일리가 행렬을 발견하게 된 것은 연립 1차 방정식을 푸는 방법을 연구한 것에 연유된다.

연립 1 차 방정식과 행렬

(1) $\begin{cases} ax + by = p \\ cx + dy = q \end{cases} \Leftrightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix}$

(2) $\begin{cases} a_1x + b_1y + c_1z \\ a_2x + b_2y + c_2z \\ a_3x + b_3y + c_3z \end{cases} \Leftrightarrow \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p \\ q \\ r \end{pmatrix}$

행렬 문제

다음 연립 1차 방정식을 행렬을 써서 나타내어라.

항목	내용	정답
문제01	2x - 3y = 5, x - y = 3	$\begin{pmatrix} 2 & -3 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$
문제02	3x = -6, x + 2y = 7	$\begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -6 \\ 7 \end{pmatrix}$

문제 01 풀이

연립 1 차 방정식과 행렬

위의 행렬식에서 역행렬을 구해보면

$$\frac{1}{D}\begin{pmatrix}-1 & 3\\ -1 & 2\end{pmatrix} = \begin{pmatrix}-1 & 3\\ -1 & 2\end{pmatrix}$$

이다.

이 역행렬을 양쪽에 추가하여, 연산을 수행하면

$$\begin{pmatrix}-1 & 3\\ -1 & 2\end{pmatrix}\begin{pmatrix}2 & -3\\ 1 & -1\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}-1 & 3\\ -1 & 2\end{pmatrix}\begin{pmatrix}5\\ 3\end{pmatrix} = \begin{pmatrix}-5+9\\ -5+6\end{pmatrix} = \begin{pmatrix}4\\ 1\end{pmatrix}$$

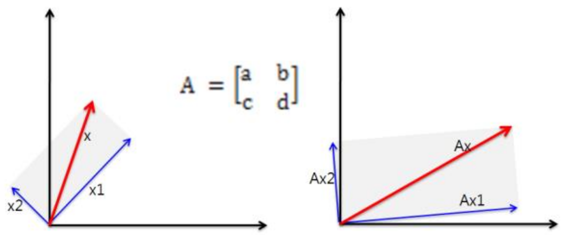
이다.

따라서, x= 4, y =1이 된다.

이것을 원래 방정식에 대입해보면 2x - 3y = 5 = 2*4 – 3*1이 성립한다.

고유 값과 고유 벡터

고유 값과 고유 벡터의 정의



고유 벡터의 기하학적 의미

그림에서 보듯이 대부분의 경우 행렬 A에 대응하는 선형 변환에 의하여 벡터 x의 사상(image) Ax는 방향이 바뀌게 된다.
그러나, 경우에 따라서 벡터 x와 벡터 Ax는 크기만 다를 뿐, 방향은 동일할 수 있다.(서로 평행하다.라고 한다.)
이와 같이 선형 변환에 의하여 크기는 바뀔 수 있지만 회전하지 않는 벡터가 존재한다.
이러한 벡터를 고유 벡터(eigenvector)라고 한다.

두 개의 "벡터가 서로 평행하다"라는 것은 한 벡터가 다른 벡터의 상수배(常數倍)가 되는 것을 의미한다.
또한, 상수배라고 했는데 이 상수(constants)를 고유 값 또는 고유치라고 부른다.

고유 값(eigen value)의 사용처 :
선형 연립 방정식 특이 값 분해 주성분 분석

따라서, 정방 행렬 A 에 대해 다음 식을 만족한다.
Ax = λx (A : 행렬, x : 벡터, λ : 숫치 정보)

위의 식은 다음과 같이 작성해도 된다.
Ax −λx = (A−λI)x = 0
단, I는 단위 행렬을 의미한다.

위 식을 만족하는 실수 λ 를 고유 값(eigenvalue)이라고 하고, 벡터 x를 고유 벡터(eigenvector)라고 한다.

항목	설명
고유 벡터(eigen vector)	특정 행렬에 의하여 선형 변환이 이루어졌을 때 방향이 바뀌지 않는 벡터를 의미한다.
고유 값 또는 고유 값(eigen value)	원래 벡터와 변형된 벡터와의 비율을 의미한다.

고유 값과 고유 벡터를 찾는 작업을 고유 분해(eigen-decomposition) 또는 고유 값 분해(eigenvalue decomposition)라고 한다.

특성 방정식(다항식)

간단한 2*2 행렬 A와 열벡터 x에 대해서 고유 값과 고유 벡터를 구해 보자.

$A=\begin{pmatrix} a & c \\ b & d \end{pmatrix}, x=\begin{pmatrix} x \\ y \end{pmatrix}$ 식을 Ax = λx에 대입하면 다음과 같다.

$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\lambda\begin{pmatrix} x \\ y \end{pmatrix}$ 가 되고, 정리하면 $\begin{matrix} (a-\lambda)x+cy=0 \\ bx+(d-\lambda)y=0 \end{matrix}$의 연립 방정식이 된다.
이 식을 다시 행렬식으로 명시하면 다음과 같다.

$\begin{pmatrix} a-\lambda & c \\ b & d-\lambda \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=0$ 으로 된다.

이 수식이 0이 아닌 해를 가질 필요 충분 조건은 행렬의 행렬식이 0이 되는 것이다.

$$\begin{vmatrix} a-\lambda & c \\ b & d-\lambda \end{vmatrix}=(a-\lambda)(d-\lambda)-bc=\lambda^2-(a+d)\lambda+ad-bc=0$$

의 수식이 성립하면 고유 벡터를 가지는 것을 알 수 있다.
λ 에 대한 이차식으로 이러한 다항식을 행렬 A의 특성 다항식이라고 한다.

특성 다항식에서 1차항의 계수는 행렬 A의 대각 원소의 합이고, 상수 항은 det A임을 주목한다.
정사각형 행렬에서 대각 원소의 합을

$$\text{trace } A = a_{11}+a_{22}+\cdots+a_{nn}=\sum_{k=1}^{n} a_{kk}$$

라고 정의하면 2*2 행렬의 특성 다항식은 다음과 같다.

$$\lambda^2-(\text{trace } A)\lambda+\det A=0$$

고유 값 문제

다음 행렬의 고유 값과 고유 벡터를 구하여라.

$$A = \begin{pmatrix} -2 & 4 \\ -1 & 3 \end{pmatrix}$$

특성 다항식으로부터 고유 값을 구할 수 있으므로

$$\lambda^2 - (\text{trace } A)\lambda + \det A = \lambda^2 - \lambda - 2 = 0$$

을 풀어 고유 값 $\lambda_1 = -1, \lambda_2 = 2$을 얻는다.

고유 값 λ_1에 대한 고유 벡터는 정의로 부터

$$Ax = \lambda_1 x = \begin{pmatrix} -2 & 4 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{matrix} -2x + 4y = -x \\ -x + 3y = -y \end{matrix}$$

여기서, x = 4, y = 1을 구할 수 있다.

$$x_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

위와 동일한 방식으로

$$x_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

을 얻을 수 있다.

고유 값 문제

다음 행렬의 고유 값과 고유 벡터를 구하여라.

$$A = \begin{pmatrix} -2 & 4 \\ -1 & 3 \end{pmatrix}$$

고유 값 구하기

$$\lambda^2 - (1+2)\lambda + 2 - 12 = \lambda^2 - 3\lambda - 10 = 0$$

λ = -2, 5가 된다.

λ = -2일 때,

$$Ax = \lambda_1 x = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{matrix} x + 3y = -2x \\ 4x + 2y = -2y \end{matrix}$$

을 연립하여 풀면 y=-x 이므로 $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -x \end{pmatrix}$ 에서

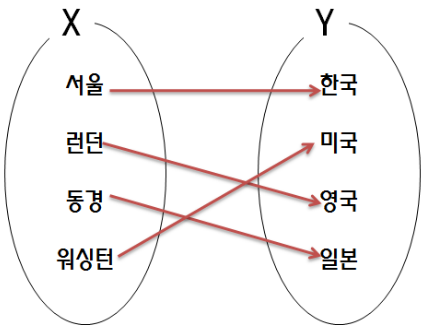
고유 벡터는 $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

함수

함수의 정의

두 개의 집합
X = {서울, 런던, 동경, 워싱턴},
Y = {한국, 미국, 영국, 일본} 이 있다고 가정하자.

X에 속하는 각 도시가 Y에 속하는 나라 중에서 어느 나라에 있는지를 조사하여, X의 각 도시에서 Y의 나라로의 대응을 『→』으로 나타내면 그림과 같음을 알 수 있다.



이 대응에서는 X의 원소 하나를 정하면 그 원소에 대하여 Y의 원소가 반드시, 그리고 오직 하나만 결정된다.
이와 같이 X의 각 원소에 대하여 Y의 원소 하나씩 대응될 때, 이 대응을 X에서 Y로의 함수라고 한다.
함수는 흔히, f, g, h 등의 문자로 나타내고, X에서 Y로의 함수 f를 기호

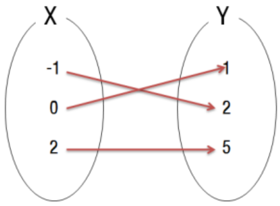
$$f:\ X \to Y$$

로 나타낸다.

이때, 집합 X를 함수 f의 정의역이라고 하고, 집합 Y를 함수 f의 공역이라고 한다.

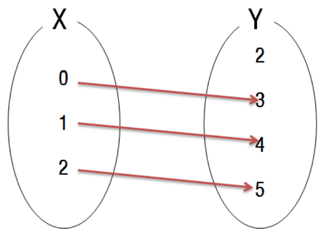
문제 : 두 집합 X = {-1, 0, 2}, Y = {1, 2, 5}가 있다.
X의 각 원소 x에 Y의 각 원소 y를 $y = x^2 + 1$인 관계에 의하여 대응 시키고, 그림으로 나타내어라.

문제 풀이:

항목	설명
x = -1일때, y = 2 x = 0일때, y = 1 x = 2일때, y = 5 따라서, X에서 Y로의 대응은 그림과 같다.	

함수 값

다음과 같은 대응 관계를 살펴 보도록 하자.



그림과 같이 나타내어 지는 함수

$$f:\ X \to Y$$

에서, X 의 한 원소를 x, 이에 대응하는 Y 의 한 원소를 y 라고 하면, x 와 y 사이에는 $y = x + 3$ 인 관계식이 성립한다.

이 x 와 y 와 같이, 어떤 집합에 속하는 여러 가지 값을 취하는 문자를 변수(variable)라고 하고, 일정한 값을 나타내는 숫자나 문자를 상수(constant)라고 한다.
또, 변수가 나타낼 수 있는 값의 집합을 그 변수의 변역이라고 한다.

이를 테면, $y = x + 3$ 에서 x 와 y 는 변수이고, 3 은 상수이다.
또, 변수 x 의 변역은 집합 X={0, 1, 2}이고, 변수 y 의 변역은 집합 Y={2, 3, 4, 5}이다.

한편 y = f(x)에서 정의역의 원소 x 의 값에 따라 공역의 원소 y 의 값이 정해지므로 x 를 독립 변수, y 를 종속 변수라고 한다.

일반적으로 집합 X 에서 Y 로의 함수

$$f:\ X \to Y$$

에서, X 의 각 원소 x 에 대하여 Y 의 단 한 개의 원소 y 가 대응한다는 것을 기호

$$y = f(x)$$

로 나타낸다.
이들에 대응하는 원소들은

$$f(0) = 0 + 3 = 3$$
$$f(1) = 1 + 3 = 4$$
$$f(2) = 2 + 3 = 5$$

와 같이 나타낼 수 있다.
이때, f(0), f(1), f(2)를 각각 함수 f(x)에서의 x=0, 1, 2 일 때의 함수 값이라고 하고, 함수 값 전체의 집합, 즉 {3, 4, 5}를 함수 f 의 치역이라고 한다.

함수 y = f(x)에서

항목	설명
정의역	x의 원소들을 의미한다.
공역	y의 원소들을 의미한다.
치역	정의역에 대응되는 값들의 집합을 의미한다. 치역은 공역의 부분 집합이다.

함수의 그래프

일차 함수 y = x + 2의 그래프를 그려 보자.
독립 변수 x가 임의의 값을 가지면서 변함에 따라 종속 변수 y도 따라서 변한다.
이것을 간단히 도표로 만들어 보면

x	...	-2	-1	0	1	2	3	...
y	...	0	1	2	3	4	5	...

와 같다.

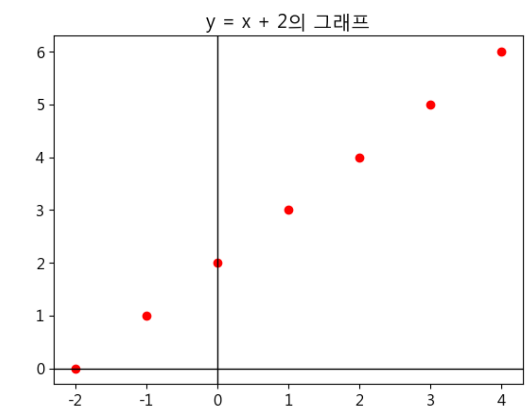
조건 제시법은 원소들의 논리적 관계를 기술하는 집합 표현 방식이다.
집합에 속하는 원소들이 가지는 공통된 성질을 조건으로 제시하여 집합을 나타낸다.
형식 : { 집합의 임의의 원소 형태 | 원소에 대한 조건 }

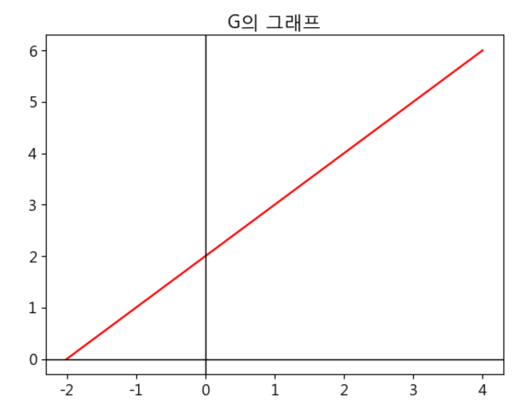
위의 일차 함수를 조건 제시법으로 표현하면 다음과 같다.
$\{ (x, y) \mid y = x + 2,\ x \in \text{정수} \}$

그런데, x는 실수 값을 가지면서 변하므로 점(x, y)의 집합은
$\{ (x, y) \mid y = x + 2,\ x \in R \}$로 표시한다.

위의 집합을 $G = \{ (x, y) \mid y = x + 2,\ x \in R \}$라고 할 때 G를 함수 y=x+2의 그래프라고 한다.

함수의 그래프
집합 X에서 집합 Y로의 함수 y=f(x)가 주어지면 이 함수에 대하여 집합
$$G = \{ (x, f(x)) \mid x \in X \}$$
가 정해진다.
이때 G를 함수 y=x+2의 그래프라고 한다.

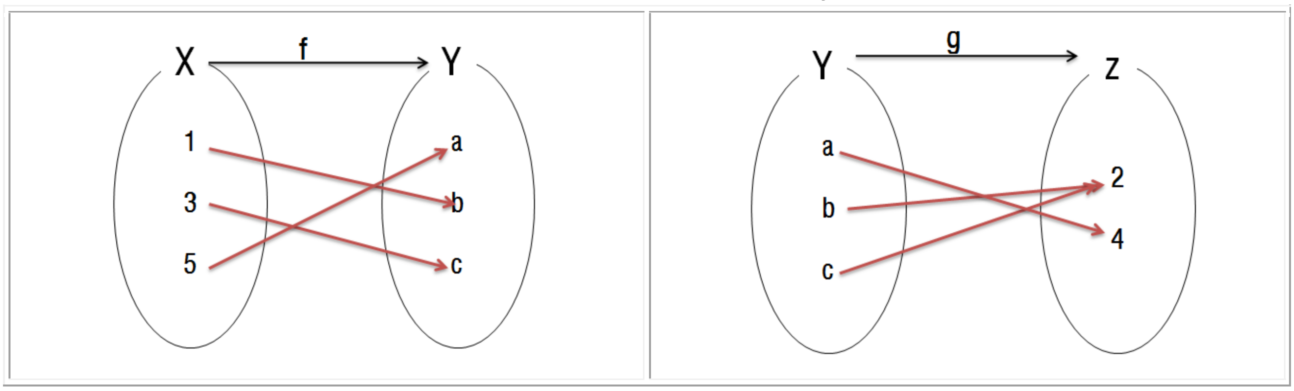




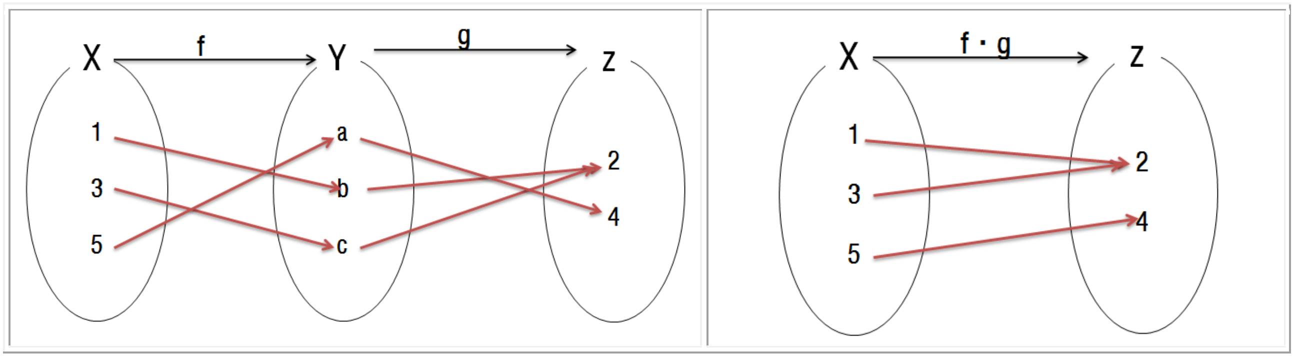
파일 이름 : func_test.py

합성 함수

세 집합 X = {1, 3, 5}, Y = {a, b, c}, Z = {2, 4}에 대하여 두 함수 f, g가 아래 그림과 같이 주어졌다고 하자.



여기서 두 함수 f와 g의 대응 관계를 그림으로 나타내면 다음과 같다.
또한, 중간 단계를 생략하고 X의 원소에 Z의 원소를 대응 시키는 것도 생각해 볼 수 있다.



두 함수 f와 g가 각각

$$f:\ X \to Y,\quad y=f(x)$$
$$g:\ Y \to Z,\quad z=g(y)$$

와 같이 주어 졌을 때, 함수 f에 의한 X의 임의의 원소 x의 상은 Y의 원소 f(x)이다.
또, 함수 g에 의한 f(x)의 상은 Z의 원소 g(f(x))이다.

이 때, X의 임의의 원소 x에 대한 Z의 원소 g(f(x))를 대응시킴으로써 X를 정의역, Z를 공역으로 하는 새로운 함수를 정의할 수 있다.

이 함수를 f와 g의 합성 함수라고 하고, 기호

$$g \circ f:\ X \to Z,\quad y=(g \circ f)(x)$$

로 나타낸다.

문제 :
두 함수 $f(x) = 2x - 1$, $g(x) = x^2$에 대하여 다음을 구하여라
$(f \circ g)(2)$, $(g \circ f)(2)$, $(f \circ g)(x)$, $(g \circ f)(x)$

문제 풀이

$(f^\circ g)(2) = f(g(2)) = f(4) = 7$

$(g^\circ f)(2)) = g(f(2)) = g(3) = 9$

$(f^\circ g)(x) = f(g(x)) = f(x^2) = 2x^2 - 1$

$(g^\circ f)(x) = g(f(x)) = g(2x - 1) = 4x^2 - 4x + 1$

합성 함수의 성질
(1) $f^\circ g \neq g^\circ f$
(2) $(h^\circ g)^\circ f = f^\circ (g^\circ f)$

문제 : 다음 물음에 답하여라.

(1) $f\left(\frac{x-1}{x+1}\right) = x^2 + x - 1$일 때 f(3)을 구하여라.

(2) $f(x) = \frac{x}{x-1}$ 일 때 $f(f(x)) = x^3$을 만족하는 x를 구하여라.

문제 풀이

(1) $\frac{x-1}{x+1} = 3$으로 놓으면 x-1=3x+3 ∴ x=-2이다.

따라서, 주어진 식의 양변에 x=-2를 대입하면

$$f\left(\frac{-2-1}{-2+1}\right) = (-2)^2 + (-2) - 1 = 1 \qquad \therefore f(3) = 1$$

(2) $f(x) = \frac{x}{x-1}$이므로 $f(f(x)) = \frac{f(x)}{f(x)-1} = \frac{\frac{x}{x-1}}{\frac{x}{x-1}-1} = \frac{x}{x-x+1} = x$ 이다 .

$\therefore x^3 = x,\ x^3 - x = 0,\ x(x+1)(x-1) = 0$ 이므로

$$\therefore x = 0, -1, 1$$

그런데, x=1은 분모를 0으로 하므로 버린다.

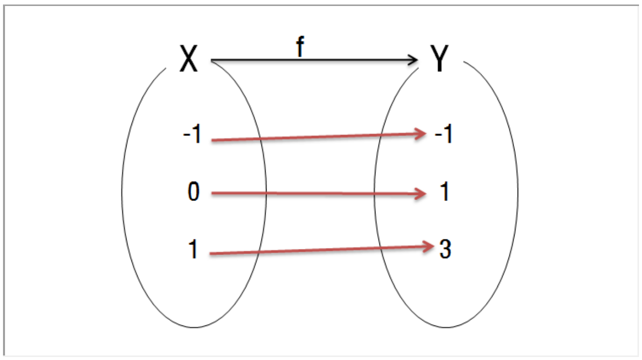
역함수

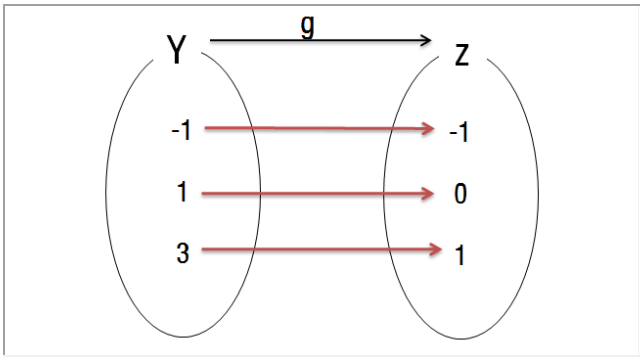
역함수

두 집합 X={-1, 0, 1}, Y={-1, 1, 3}에 대하여

함수 f: X→Y, y=2x+1은

대응 관계에서 알 수 있듯이 일대일 대응이다.





이제, 일대일 대응 f: X→Y에 대하여 우측 그림처럼 Y에서 X로의 대응을 생각하면 Y의 임의의 원소 y에 대하여 $x=\frac{1}{2}(y-1)$를 만족하는 X의 원소 x가 오직 하나 존재하므로 이 대응 관계 역시 하나의 함수이다.

함수 g: Y→X, $x=\frac{1}{2}(y-1)$

도 우측의 그림의 대응 관계에서 알 수 있듯이 ´일대일 대응´이다.
여기서, 함수 g를 함수 f의 역함수라고 하고, f^{-1}로 나타낸다. 즉,

f^{-1}: Y→X, $x=\frac{1}{2}(y-1)$

역 함수

함수 f: X→Y, y=f(x)가 일대일 대응일 때 함수 f(x)의 역함수 f^{-1}는 다음과 같이 나타낼 수 있다.

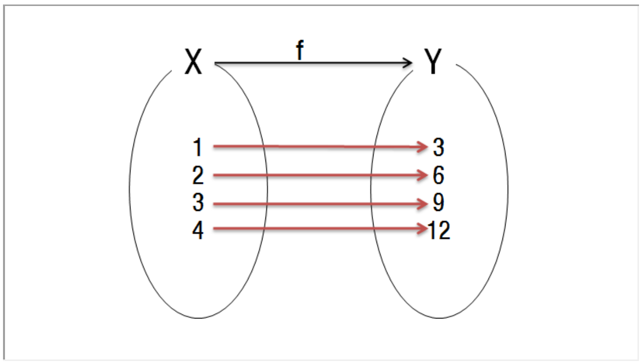
f^{-1}: Y→X, $y=f^{-1}(x)$

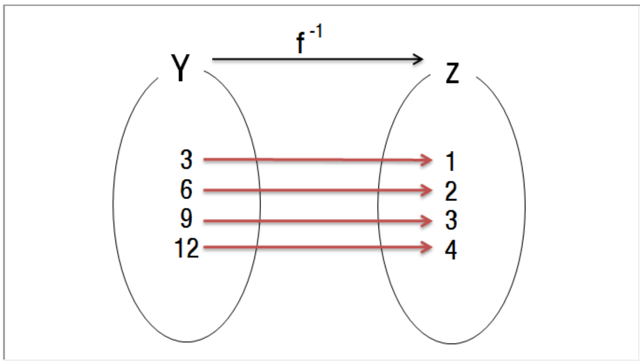
역함수를 구하는 방법

함수 f:x-->3x(=y), 즉 f:x-->y의 역함수 f^{-1}를 구한다는 것은 역의 대응 규칙

$$f^{-1}: y \longrightarrow \frac{y}{3}(=x), \text{즉}, f^{-1}: y \longrightarrow x$$

를 구하는 것과 같다.





위의 대응 관계에서 알 수 있는 바와 같이 함수 y=3x의 역함수는

y=3x를 x에 관하여 풀어서 $x=\frac{y}{3}$

이다.

그런데, 함수를 나타낼 때, 보통의 경우 독립 변수를 x, 종속 변수를 y로 표현하므로

$$x = \frac{y}{3} \text{에서 } x\text{와 } y\text{를 바꾸어 } y = \frac{x}{3}$$

로 나타낸다.

역 함수를 구하는 순서
첫째 : 주어진 함수 y=f(x)가 일대일 대응인지를 검토한다. 둘째 : y=f(x)를 x에 관하여 풀어서 x=g(y)의 꼴로 고친다. 셋째 : x=g(y) 꼴에서 x와 y를 바꾸어 y=g(x)로 고친다.

문제 : 함수 $f(x) = \frac{1}{2}x - 1$의 역함수를 구하여라.

문제 풀이

이 함수는 일차 함수이고, 일대일 대응이다.
따라서 f(x)의 역함수 $f^{-1}(x)$가 존재한다.
이 식을 x에 대하여 풀면
$y = \frac{1}{2}x - 1$이라고 놓고, x에 대하여 풀면 $\therefore x = 2(y+1)$
x와 y를 바꾸어 쓰면 $y = 2(x+1)$

정답 : $\therefore f^{-1}(x) = 2(x+1)$

역함수의 그래프

이제, 함수 y=f(x)의 그래프와 그 역 함수 $f^{-1}(x)$의 그래프 사이의 관계를 살펴 보자.
함수 y=f(x) 그래프 위의 점(a, b)를 택하면

$$b = f(a)\text{이므로 } f^{-1}(b) = a$$

따라서, 점(a, b)가 y=f(x)의 그래프 위에 있으면 점(b, a)는 f(x)의 역함수 $f^{-1}(x)$의 그래프 위에 있다.

그런데, 점(a, b)와 점(b, a)는 직선 y=x에 대하여 대칭이므로 다음이 성립한다.

역함수의 그래프
함수 y=f(x)의 그래프와 그 역함수 $y = f^{-1}(x)$의 그래프는 직선 y=x에 대하여 대칭이다.

지수

거듭 제곱

이를 테면

$$a \times a = a^2$$
$$a \times a \times a = a^3$$
$$\cdots$$
$$a \times a \times a \ldots \times a = a^n$$

에서 a^2을 a의 제곱, a^3을 a의 세제곱, a^n을 a의 n제곱이라고 하고, 이를 통틀어서 a의 거듭 제곱이라고 한다.

a의 거듭 제곱 개념을 이용하면 다음과 같은 법칙이 성립한다.

지수 법칙(지수의 확장)

임의의 실수 a, b와 양의 정수 m, n에 대하여 다음과 같은 법칙이 성립한다.

m, n이 양의 정수일 때		
$a^m \times a^n = a^{m+n}$	$a^m \div a^n$	a^{m-n} $(m > n)$
$(a^m)^n = a^{mn}$		$1 (m = n)(a \neq 0)$
$(ab)^n = a^n b^n$		$\frac{1}{a^{n-m}} (m < n)(a \neq 0)$

0의 지수, 음의 정수의 지수 규약	
$a^0 = 1$	$a^{-n} = \frac{1}{a^n}$(단, $a \neq 0$, n은정수)

거듭 제곱근의 계산

a의 n 제곱근과 $\sqrt[n]{a}$

3제곱을 해서 8이 되는 수를 8의 3제곱근이라고 한다.
답을 구해보면 $x^3 = 8$에서 $x^3 - 8 = x^3 - x2^3 = (x-2)(x^2+2x+4) = 0$이므로 이것을 풀어 보면

$$\therefore x = 2, -1 \pm \sqrt{3}i$$

이다.
이 중에서 실수인 2를 $\sqrt[3]{8}$으로 나타내기로 정의하고, $\sqrt[3]{8}$을 3제곱근 8이라고 읽는다.

a 의 n 제곱근과 $\sqrt[n]{a}$
(1) a의 n 제곱근의 정의 n 제곱해서 a가 되는 수 곧 $x^n = a$를 만족하는 x를 a의 n제곱근이라고 한다. (2) a의 n 제곱근과 $\sqrt[n]{a}$(n 제곱근 a)와의 관계 (가) n이 홀수인 경우 a를 임의의 실수라 할 때, a의 n제곱근이 되는 실수는 오직 한 개 있으며, 이것을 $\sqrt[n]{a}$로 나타낸다. (나) n이 짝수인 경우 a > 0 일때 a의 n제곱근이 되는 실수는 양, 음의 2개가 있으며, 양인 것은 $\sqrt[n]{a}$,음인 것을 $-\sqrt[n]{a}$로 나타낸다. a <= 0 일때 a의 n제곱근이 되는 실수는 없다.

문제 :다음에서 옳은 것은? 정답 (2)

(1) 125의 3제곱근은 $\sqrt[3]{125}$이다. (2)$\sqrt[3]{125}$는 125의 3제곱근 중의 하나이다. (3) 81의 4제곱근은 $\pm\sqrt[4]{81}$이다. (4) $\sqrt[4]{-81}$은 -81의 4제곱근 중의 하나이다.

거듭 제곱근의 계산 법칙

거듭 제곱근에 관하여 a > 0, b > 0이고, m, n은 2이상의 정수일 때다음과 같은 계산 법칙이 성립한다.

거듭 제곱근의 계산 법칙		
$\sqrt[n]{a}\sqrt[n]{b} = \sqrt[n]{ab}$	$\dfrac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\dfrac{a}{b}}$	$(\sqrt[n]{a})^m = \sqrt[n]{a^m}$
$\sqrt[m]{\sqrt[n]{a}} = \sqrt[mn]{a} = \sqrt[n]{\sqrt[m]{a}}$	$\sqrt[np]{a^{mp}} = \sqrt[n]{a^m}$ (단, p는 양의 정수)	

문제 :다음의 각 값을 구하여라.

항목	설명			
$\sqrt[4]{81}$	$\sqrt[5]{-32}$	$\sqrt[4]{2}\sqrt[4]{8}$	$\sqrt[3]{0.0001}\sqrt[3]{10}$	$\dfrac{\sqrt[3]{2}}{\sqrt[3]{16}}$
$\dfrac{\sqrt[4]{64}}{\sqrt[4]{4}}$	$\sqrt[4]{16^3}$	$\sqrt[3]{(\frac{8}{27})^2}$	$\sqrt[4]{\sqrt[3]{16}} \times \sqrt{\sqrt[3]{16}}$	$\sqrt[12]{3^4} \times \sqrt[9]{3^6}$

정답(문제 풀이)

항목	설명			
3	-2	2	0.1	0.5
2	8	4/9	2	3

분수 지수의 규약		
a > 0, m은 정수, n은 2 이상의 정수일 때,		
$a^{\frac{1}{n}} = \sqrt[n]{a}$	$a^{\frac{m}{n}} = \sqrt[n]{a^m}$	$a^{-\frac{m}{n}} = \dfrac{1}{a^{\frac{m}{n}}}$

확장된 지수 법칙		
a > 0, b > 0이고, m, n이 유리수일때		
$a^m \times a^n = a^{m+n}$	$a^m \div a^n = a^{m-n}$	$(a^m)^n = a^{mn}$

$(ab)^n = a^n b^n$	$\left(\frac{b}{a}\right)^n = \frac{b^n}{a^n}$	

다음의 식들을 간단히 하여라.

문제	$a^{-3} \times a^4 \div a^{-2}$	$(-a^{-5})^{-4}$	$\{(a^2b^{-4})^{-3}\}^{-1}$
정답	a^3	a^{20}	a^6b^{-12}

다음의 식들을 간단히 하여라.

문제	$27^{\frac{1}{3}}$	$\left\{\left(\frac{9}{16}\right)^{-\frac{3}{4}}\right\}^{\frac{2}{3}}$	$\left(-\frac{1}{3}\right)^{-3}$	$\{(-3)^4\}^{\frac{1}{4}}$
정답	3	$\frac{4}{3}$	-27	3

$e^{2x} = 3$일때, 다음의 각 값을 구하여라.

문제	$\left(\frac{1}{e^3}\right)^{-4x}$	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$	$\frac{e^{3x} - e^{-3x}}{e^x + e^{-x}}$
정답	729	$\frac{1}{2}$	$\frac{7}{3}$

로그

로그의 정의

로그의 정의
a > 0, a ≠1일 때, 임의의 양수 b에 대하여 $a^x = b$를 만족시키는 실수는 오직 하나 존재한다. 이 실수 x를 a를 밑으로 하는 b의 로그라고하고, $x = \log_a b$로 나타낸다. 이 때 b를 $\log_a b$의 진수라고 한다. a > 0, a ≠1, b > 0일때 $a^x = b$<===>$x = \log_a b$

로그의 성질

로그의 진수가 두수의 곱, 비, 거듭제곱 등의 형태일 때 사용하는 공식을 말한다.
진수가 곱이면 두 로그의 합으로 바꿀 수 있으며, 비면 차, 거듭제곱이면 곱으로 바꿀 수 있다.
또한 로그의 밑을 로그 거듭제곱하면 진수가 된다.(참조 : 네이버 지식백과)

로그			
a ≑ 0, a >0이고, x >0, y >0일때			
$\log_a a = 1$	$\log_a 1 = 0$	$\log_a xy = \log_a x + \log_a y$	$\log_a \frac{x}{y} = \log_a x - \log_a y$
$\log_a x^r = r \log_a x$	$a^{\log_a x} = x$		

다음 물음에 답하여라.

문제		정답
문제 01	$\log_3 6 = a$일때, $\log_3 288$의 값을 구하여라.	5a - 3
문제 02	log0.5=a,log9=b일때, log72의값을 구하여라.	-3a + b
문제 03	log6 = 2a,log1.5 = 2b일때, log24의값을 구하여라.	4a - 2b

문제 01 풀이
log3_6 = log3_3 + log3_2 = a log3_2 = a - 1 288=32*9=2^5*3^2 log3_288 = log3_2^5*3^2 = log3_2^5 + log3_3^2 = 5*log3_2 + 2*log3_3 = 5*(a - 1) + 2*1 = 5a - 5 + 2 = 5a - 3
문제 02 풀이
log_0.5 = log_1/2 = -log_2 = a log_2 = -a log_9 = log_3^2 = b log_3 = b/2 72 = 8*9 = 2^3*3^2 log_72 = log_2^3*3^2 = 3*log_2 + 2*log_3 = 3*(-a) + 2*b/2 = -3a + b
문제 03 풀이
log_0.5 = log_1/2 = -log_2 = a

```
log_6 = log_2*3 = log_2 + log_3 = 2a
log_1.5 = log_3/2 = log_3 - log_2 = 2b

2log_3 = 2a+2b
log_3 = a + b
log_2 = a - b

24 = 8*3 = 2^3*3
log_24 = log_2^3*3 = 3*log_2 + log_3 = 3*(a-b) + a+ b
        = 4a - 2b
```

다음 각 식을 만족하는 x의 값을 구하여라

문제	$\log_8 0.25 = x$	$\log_{2\sqrt{2}} 64 = x$	$\log_x 81 = 2$
	$\log_x 9 = \frac{2}{3}$	$\log_{25} x = 1.5$	$\log_6(\log_{64} x) = -1$
정답	$-\frac{2}{3}$	4	9
	27	125	2

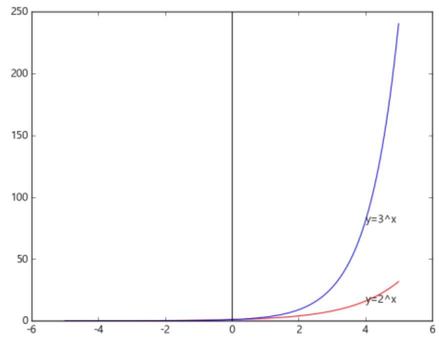
$\log_8 0.25 = x$ 풀이
log8_0.25 = log2^3_1/2^2 = -2/3*log2_2 = -2/3

지수 로그 함수

$y = 2^x, y = 3^x$등과 같이 $y = a^x (a \neq 1, a > 0)$으로 나타내어질 때, y는 a를 밑으로 하는 x의 지수 함수라고 한다.
또, $y = \log_2 x, y = \log_3 x$ 등과 같이 $y = \log_a x\,(a \neq 1, a > 0)$으로 나타내어질 때, y는 a를 밑으로 하는 x의 로그 함수라고 한다.
두 개의 함수는 직선 y=x에 대하여 대칭이므로 서로 역함수 관계이다.

지수 함수 $y = a^x$의 성질
(1) 정의역은 실수 전체의 집합 R이고, 치역은 {yly>0}이다. (2) 그래프는 점(0, 1)을 지난다. (3) 그래프는 직선 y=0(x축)을 점근선으로 한다. (4) a >= 1일때, 단조 증가하고, 0 <a < 1일때, 단조 감소 한다.

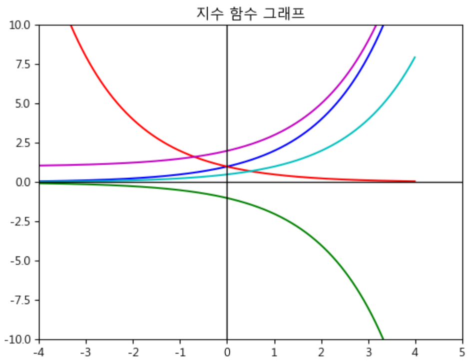
$y = 2^x, y = 3^x$ 그래프 그려 보기



exponential_02.py

$y = 2^x$ 그래프를 이용하여 다음 각 함수의 그래프를 그려라.

문제	$y = \left(\frac{1}{2}\right)^x$	$y = -2^x$	$y = 2^x + 1$	$y = 2^{x-1}$

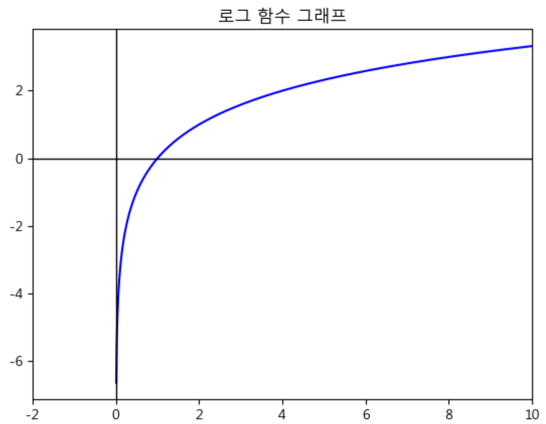


exponential_01.py, exponential_03.py,exponential_image_01.png

로그 함수 $y = \log_a x$의 성질
(1) 정의역은 {x|x>0}이고, 치역은 실수 전체의 집합 R이다. (2) 그래프는 점(1, 0)을 지난다. (3) 그래프는 직선 x=0(y축)을 점근선으로 한다. (4) a >= 1일때, 단조 증가하고, 0 <a < 1일때, 단조 감소 한다.

$y = \log_2{}^x$ 그래프를 이용하여 다음 각 함수의 그래프를 그려라.

문제	$y = \log_2(-x)$	$y = \log_2 \frac{1}{x}$	$y = \log_2(x-1)$	$y = \log_2 2x$



logarithm_01.py, log2.png

풀이

$y = \log_2(-x)$는 y 축에 관하여 대칭이다.
$y = \log_2 \frac{1}{x}$ 는 x 축에 관하여 대칭이다.
$y = \log_2(x-1)$는 x 축 에 1 만큼 평행 이동한 그래프이다.
$y = \log_2 2x$는 y 축 에 1 만큼 평행 이동한 그래프이다.

미분

변화율과 도함수

평균 변화율

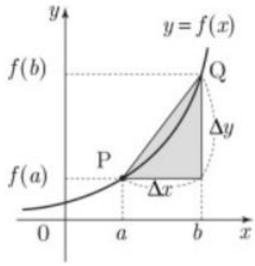
머신 러닝에는 경사 하강법이라는개념이 있다.
이것은 곡선에서 미분 계수를 이용하여 접선의 기울기를 구하는 것인데, 우선 평균 변화율의 개념부터 살펴 보기로 한다.

평균변화율은 독립변수의 변화량에 따른 종속변수의 변화량의 비를 말한다.
즉,변수 x의 증가량△x에 대한 y 값의 증가량△y를 말하며, 기하학적으로는 직선의 기울기를 의미한다.

함수 y = f(x)를 이용하여 평균변화율에 대하여 설명하도록 하겠다.

평균 변화율의 개념

$$\frac{\Delta y}{\Delta x} = \frac{f(b) - f(a)}{b - a} = \frac{f(a + \Delta x) - f(a)}{\Delta x}$$



문제 :평균 변화율 구해 보기

함수 $y = f(x) = x^2$ 이란 함수에 대하여 x가 1부터 3까지 변한다고 가정할 때 평균 변화율을 구하시오.
x의 값이 1부터 3까지 변하면 y는 1부터 9까지 변하게 된다.
△x = 3 - 1 = 2이므로 다음과 같은 공식으로 구할 수 있다.

$$\frac{\Delta y}{\Delta x} = \frac{f(a + \Delta x) - f(a)}{\Delta x} = \frac{f(3) - f(1)}{2} = \frac{9 - 1}{2} = 4$$

구간 [1, 3]에서의 y의 평균 변화율은 두 점 P(1, 1), Q(3, 9)를 잇는 직선의 기울기와 같다.

풀어 봅시다

함수 $y = f(x) = x^2 - 2x$ 이란 함수에서 주어진 구간 [3, 5]에서의 평균 변화율을 구하시오.
정답 : 6

실습 파일 :avg_rate_01.py

순간 변화율

평균 변화율에 극한이라는 개념을 적용하여 다음과 같은 식을 만들 수 있다.
이때Δx의값을아주작게 만들면 극한이라는 개념이 적용된다.

$$f'(\mathrm{x}) = \lim_{\Delta X \to 0}\left(\frac{f(x + \Delta x) - f(x)}{\Delta x}\right)$$

이때 $f'(\mathrm{x})$를 y = f(x)의 도함수 또는 미분이라고 한다.
표기할 때는 $f'(\mathrm{x})$라고 사용할 때도 있고, 다음과 같이 표기하기도 한다.
$f'(\mathrm{x})$는 prime(프라임) x로 읽는다.

$$\frac{dy}{dx} 또는 \frac{d}{dx}f(x)$$

극한을적용했을 때의 변화율을 순간 변화율(instantaneous rate of change) 또는 미분 계수(differential coefficient)라고 부른다.

즉, 미분 계수는 △x의 값이 0에 가까워질 때의 평균 변화율을 의미하고, 이 변화율은 접점 a에서의 접선의 기울기를 의미한다.

미분 계수의 기하학적 의미

미분 계수의 기하학적 의미
함수 y = f(x)의 점(x_1, y_1)에서의 미분 계수 $f'(x_1)$은 곡선 y = f(x) 위의 점 $(x_1,\ f(x_1))$에서의 접선의 기울기와 같다. 접선의 방정식은 $y - y_1 = f'(x_1)(x - x_1)$이다.

문제 :접선의 기울기 구하기

다음 함수 위의 점(1, 3)에서의 접선의 기울기를 구하시오.
$y = 4x - x^2$ 위의 점(1, 3)에서의 접선의 기울기를 구하시오.문제 :diff01.py
$f'(1) = \lim_{\Delta x \to 0} \frac{f(1+\Delta x) - f(1)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\{4(1+\Delta x) - (1+\Delta x)^2\} - 3}{\Delta x} = \lim_{\Delta x \to 0} (2 - \Delta x) = 2$

문제 :접선의 방정식 구하기

곡선 $y = x^3$ 위의 점(2, 8)에서의 접선의 방정식을 구하시오.
$y = f(x) = x^3$으로 놓으면, $f'(x) = 3x^2$이므로 x = 2에서의 접선의 기울기는 $f'(2) = 3 \cdot 2^2 = 12$이다.
따라서, 구하고자 하는 접선의 방정식은 $y - 8 = 12(x - 2)$에서 y = 12x - 16이다.

법선의 방정식

곡선 위의 접점을 지나며 접선에 수직인 직선을 법선이라고 한다.
접선과 법선으로 서로 수직이므로, 접선의 기울기가놓으면, $f'(x)$일 때, 법선의 기울기는 $-\frac{1}{f'(x)}$이다.

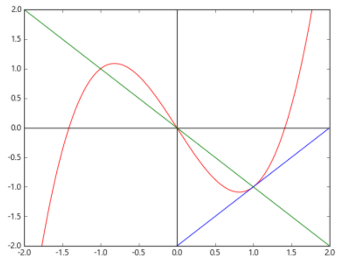
법선의 방정식
곡선 y = f(x) 위의 한 점 P(x_1, y_1)에서의 법선의 방정식은 다음과 같다. $y - y_1 = -\frac{1}{f'(x)}(x - x_1)$이다.

문제 :접선과 법선의 방정식 구하기

곡선 $y = x^3 - 2x$ 위의 점(1, -1)에서의 접선의 방정식과법선의 방정식을 구하시오.

정답	접선의 방정식	법선의 방정식
	y = x - 2	y = -x

실습파일 :make_graph_01.py (접선과 법선의 방정식.png)



미분 공식

다음 공식은 도함수를 구할 때 일일이 도함수의 정의에 의하지 않고도 간단히 셈을 할 수 있는 데 도움을 주는 중요한 미분 공식이다.

항목	설명
미분의 기본/성질 공식	$(c)' = 0$ (c 는 상수이다.) $(x^n)' = nx^{n-1}$ (n 은 자연수이다.) $(cf(x))' = c(f'(x))$ $\left(f(x) + g(x)\right)' = f'(x) + g'(x)$ $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$ $\left(\frac{\mathrm{f(x)}}{\mathrm{g(x)}}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$
유리 함수	$\frac{\mathrm{d}}{\mathrm{dx}}x^r = rx^{r-1}$ $\frac{\mathrm{d}}{\mathrm{dx}}\sqrt{x} = \frac{1}{2\sqrt{x}}$
로그 함수	$\frac{\mathrm{d}}{\mathrm{dx}}logx = \frac{1}{x}$ $\frac{\mathrm{d}}{\mathrm{dx}}log_a x = \frac{1}{xloga}$
지수 함수	$\frac{\mathrm{d}}{\mathrm{dx}}e^{ax} = ae^{ax}$ $\frac{\mathrm{d}}{\mathrm{dx}}a^x = a^x loga$

문제 :도함수 구하기

다음 함수의 도함수를 구하시오.

문제	$\mathrm{y = 100}$	$\mathrm{y = x^3}$	$\mathrm{y = x^{10}}$	$\mathrm{y = 6x^5}$

정답	$y' = 0$	$y' = 3x^2$	$y' = 10x^9$	$y' = 30x^4$

문제 :도함수 구하기

다음 함수의 도함수를 구하시오.

문제	$y = x^2(1-x^3)$	$y = (2x+1)(3x^2+x-1)$	$y = (2x+3)(x^2+x)(1-x)$
정답	$y' = 2x - 5x^4$	$y' = 18x^2 + 10x - 1$	$y' = -8x^3 - 9x^2 + 4x + 3$

풀이 :$y = (2x+1)(3x^2+x-1)$

2(3x^2+x-1) + (2x+1)(6x+1)
=6x^2 + 2x - 2 + 12x^2 + 2x +6x + 1
=18x^2 + 10x - 1

문제 :도함수 구하기

다음 함수의 도함수를 구하시오.

문제	$y = 2x^3 + 6x^2 - 5x + 1$	$y = 5x^{10} + 7x^6 - 3x^3 + x - 2$
정답	$y' = 6x^2 + 12x - 5$	$y' = 50x^9 + 42x^5 - 9x^2 + 1$

문제 :도함수 구하기

다음 함수를 미분하여라.

문제	$y = (x^2-3x+1)(x^2+1)$	$y = (x-2)(x+1)(2x+1)$
정답	$y' = 4x^3 - 9x^2 + 4x - 3$	$y' = 6x^2 - 2x - 5$

합성 함수

함수 $y = (2x+1)^2$의도함수를 구해 보도록 하자.
먼저 우변을 전개하면 $y = (2x+1)^2 = 4x^2 + 4x + 1$이다.
여기에 도함수를 구해 보면 $\frac{dy}{dx} = y' = 8x + 4$이다.

주어진 식을 전개하지 않고, 도함수를 구하는 방법을 알아 보자.
함수 $y = (2x+1)^2$ 은 $y = t^2$, $t = 2x + 1$ 두 함수의 합성 함수로 이해할 수 있다.
여기에 도함수를 구해 보면 $\frac{dy}{dt} = (t^2)' = 2t$이고, $\frac{dt}{dx} = (2x+1)' = 2$이다.
$\frac{dy}{dx} = \frac{dy}{dt} \cdot \frac{dt}{dx} = 2t \cdot 2 = 4t = 4(2x+1) = 8x + 4$가 된다.

이와 같은 미분법을 합성 함수의 미분법이라고 한다.

합성 함수의 미분법
함수 y = f(t)와 함수 t = g(x)가 미분이 가능하면 합성 함수 y = f(g(x))도 미분이 가능하고, 이때 다음과 같은 식이 성립한다. y = f(t), t = g(x)이면 $\Rightarrow \frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx} = f'(g(x)) \cdot g'(x)$

문제 :도함수 구하기

다음 함수를 합성 함수의 미분법을 이용하여 미분하여라.

문제	$y = (x^2 + x)^3$	$y = (1 - 2x)^4$
정답	$y' = 3(x^2 + x)^2(2x + 1)$	$y' = -8(1 - 2x)^3$

합성 함수의 미분법
y = t^3 t = x^2 + x dy/dt = 3*t^2 dt/dx = 2x + 1 dy/dt * dt/dx = 3*(t)^2(2x + 1)= 3*(x^2 + x)^2(2x + 1)

편미분

이전까지 봤던 함수는 변수 x라는 1개의 변수를 가진 함수이었다.
그러나, 다음과 같이 2개 이상의 변수를 가지는 다변수 함수도 존재한다.

다변수 함수 예시
$f(x_1, \quad x_2, \quad x_3, \quad x_4, \dots x_n,) = x_1 + {x_2}^2 + \dots {x_n}^n$

머신 러닝에 나오는 최적화 문제는 매개 변수의 개수 만큼의 변수가 존재하므로 비용 함수가 다변수 함수 형태로 존재한다.
다변수 함수를 미분할 때는 미분할 변수에만 주목하고 나머지 다른 변수는 모두 상수(constant) 형태로 취급하는 데, 이러한 미분 방식을 편미분이라고 한다.

편미분은 일반 미분과 구분하기 위하여 ∂ 기호를 사용한다.

편도 함수

$$\frac{\partial z}{\partial x} = \lim_{\Delta x \to 0}\left(\frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}\right)$$

$$\frac{\partial z}{\partial y} = \lim_{\Delta y \to 0}\left(\frac{f(x, y+\Delta y) - f(x,y)}{\Delta y}\right)$$

함수z=f(x,y)=x^2+3y+1에 대하여 편미분을 수행해 보도록 한다.

$$\frac{\partial z}{\partial x} = \lim_{\Delta x \to 0}\left(\frac{(x+\Delta x)^2 + 3y + 1 - (x^2 + 3y + 1)}{\Delta x}\right) = \lim_{\Delta x \to 0}\left(\frac{2x\Delta x + \Delta x^2}{\Delta x}\right) = 2x$$

$$\frac{\partial z}{\partial y} = \lim_{\Delta x \to 0}\left(\frac{x^2 + 3(y\ +\ \Delta y) + 1 - (x^2 + 3y + 1)}{\Delta y}\right) = \lim_{\Delta x \to 0}\left(\frac{3\Delta y}{\Delta y}\right) = 3$$

편미분 정의식의 일반화

$$\frac{\partial z}{\partial x_i} = \lim_{\Delta x_i \to 0}\left(\frac{f(x_1 + x_2 + \cdots \Delta x_i \ldots x_n) - f(x_1 + x_2 + \cdots + x_n)}{\Delta x_i}\right)$$

편미분 예시

2개의 변수 x, y로 구성된 함수 z가 있다고 가정하자.
편미분 계수에 대하여 살펴 보기 위하여 이 함수를 살펴 보기로 하자.

이변수 함수 예시
$\mathrm{z} = \mathrm{f}(\mathrm{x}, \quad \mathrm{y}) = \mathrm{x}^2 + \mathrm{y}^3$

x에 대하여 편미분을 수행하게 되면 y를 임의의 값으로 고정하게 되므로 $\mathrm{z} = \mathrm{g}(\mathrm{x}) = x^2 + b^2$라고 봐도 무방하다.
즉, y는 상수(constant)값으로 이해하면 된다.

변수x, y에 대하여 각각편미분을 수행하면 다음과 같다.

$$\frac{\partial}{\partial x} z = 2x \frac{\partial}{\partial y} z = 3y^2$$

미분의 기본 공식

일반적인 미분 공식과 유사하게 편미분에도 기본 적인 산술 연산자를 이용한 공식들이 있다.
변수 x, y로 구성이 되어 있는 함수 f(x, y), g(x, y)에 대하여 다음과 같은 공식이 성립한다.

연산 방식	설명
덧셈,뺄셈	$\frac{\partial}{\partial x}\big(f(x,y)\ \pm\ g(x,y)\big) = \frac{\partial}{\partial x}\big(f(x,y)\big)\ \pm\ \frac{\partial}{\partial x}(g(x,y))$
곱셈	

	$\frac{\partial}{\partial x}(f(x,y)g(x,y)) = \left(\frac{\partial}{\partial x}f(x,y)\right)g(x,y) + f(x,y)\left(\frac{\partial}{\partial x}g(x,y)\right)$
나눗셈	$\frac{\partial}{\partial x}\left(\frac{f(x,y)}{g(x,y)}\right) = \frac{\left(\frac{\partial}{\partial x}f(x,y)\right)g(x,y) + f(x,y)\left(\frac{\partial}{\partial x}g(x,y)\right)}{(g(x,y))^2}$
정수배	정수 c에 대한 배수 연산 $\frac{\partial}{\partial x}cf(x,y) = c\frac{\partial}{\partial x}f(x,y)$

적분

부정적분

예를 들어서 $F(x)=x^2$일 때 $F'(x) = 2x$이다.
이것은 "미분한다" 또는 "도함수를 구한다"라고 학습한 바 있다.

이제는 역으로 도함수를 알 때, 이것을 써서 미분하기 전의 원래의 함수를 구하는 방법을 살펴 보도록 한다.
원래의 함수를 원시 함수 또는 부정적분이라고 한다.

이 때 부정적분을 구하는 것은 적분한다라고 하고, 그 계산법을 적분법이라고 한다.

부정적분의 뜻
부정적분(indefinite integral)은 정확하게 미분과 반대되는 개념, 즉 반-미분(anti-derivative)이다. 함수f(x)가 어떤 함수를 미분하여 나온 결과인 도함수라고가정하자. 이 도함수 f(x)에 대한 미분되기 전의 원래의 함수를 찾는 과정(integration)을 의미한다.

여기에서 도함수가 f(x)이므로 미분하기 전의 함수를F(x)으로 표기한다.

$$\frac{dF(x)}{dx} = f(x) \leftrightarrow F(x) = \int f(x)dx + C$$

부정적분의 표시
함수 f(x)의 부정적분의 하나를 F(x)라고 할 때, f(x)의 임의의 부정적분은 F(x) + C(C는 임의의 상수)가 되며, 이것을 다음과 같이 나타낸다. $\int f(x)dx = F(x) + C$

여기서, C 는 상수항(constant)을 뜻한다.
상수항은어차피 미분하면 0이 되므로 부정적분은 무한 개의 해가 있을 수 있다.
일반적으로 C는 생략하고 쓰는 경우도 있다.

문제 :다음에 대한 부정적분을 구하여라.

문제	func = 2 * x + 5	func2 = 2 * x + 3 * y (y에 대하여 부정적분)
정답	x**2 + 5*x	2*x*y + 3*y**2/2

실습 파일 : integrate01.py

x^n의부정적분

n이 양의 정수일 때

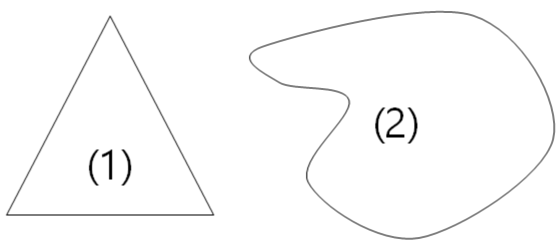
$$\int \mathbf{x^n} dx = \frac{\mathbf{1}}{\mathbf{n+1}} \mathbf{x^{n+1}} + C$$

문제 :다음에 대한 부정적분을 구하여라.

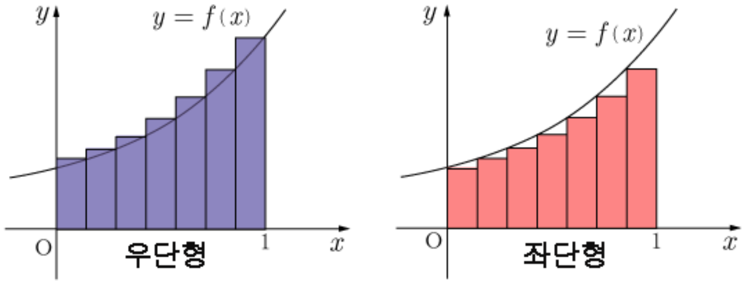
문제	$\int 3x^2dx$	$\int 6x^5dx$	$\int 10dx$	$\int 4xdx$	$\int 9x^2dx$	$\int 12x^3dx$
정답	$x^3 + C$	$x^6 + C$	$10x + C$	$2x^2 + C$	$3x^3 + C$	$3x^4 + C$

구분 구적법

삼각형 또는 사각형의 넓이는 길이 정보를 알면 쉽게 구할 수 있다.(그림(1))



그러나, 그림(2)와 같이 만들어진 도형은 너비나 높이의 개념으로 면적을 구할 수 없다.
이러한 경우에는 구분 구적법이라는 개념을 도입하여 면적을 구할 수 있다.



참조 사이트 :http://ictmath.honam.ac.kr/category/sekibun/kubun-kyuuseki-hou.html

그림에서 원점과 점(1, 0)과 함수 f(x)의 좌우 끝점을 연결하는 도형의 면적을 S라고 하자.
우단형은 S보다 조금 넓은 데 이 면적을 S_r이라고 하고, 좌단형은 S보다 면적이 작다.
이 면적을 S_l이라고 한다면 다음과 같은 공식이 성립한다.

$$S_l < S < S_r$$

또한, 분할하는 개수 n을 무한히 크게 하면 다음과 같은 공식이 성립한다.

$$S = \lim_{n\to\infty} S_l = \lim_{n\to\infty} S_r$$

이와 같이, 도형의 넓이나 부피를 구할 때 주어진 도형을 잘게 쪼개어 그 도형의 넓이나 부피의 근사 값을 구한 다음, 그 근사 값의 극한 값으로 도형의 넓이나 부피를 구하는 방법을 구분 구적법이라고 한다.

정적분

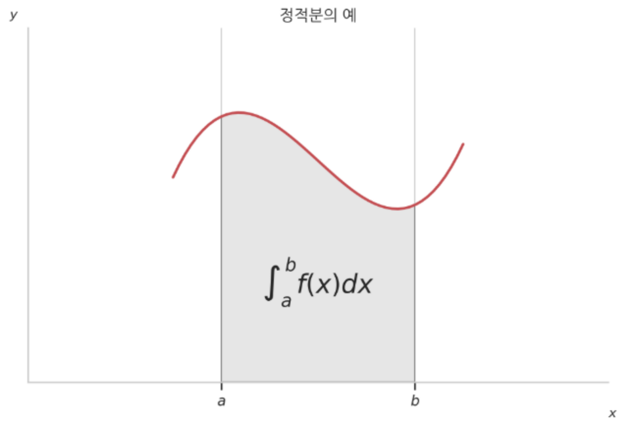
정적분은 기본적으로 구분 구적법을 기본 토대로 하고 있다.
정적분(definite integral)은 독립변수 x가 어떤 구간[a,b]사이일 때 그 구간에서 함수f(x) 의 값과 수평선(x 축)이 이루는 면적을 구하는 행위(integration) 혹은 그 값(integral)을 말한다.

어떤 함수가 f(x)≥0이라고 할 때, 구간 [a, b]사이의 면적은 다음과 같이 구할 수 있다.

정적분의 기본 정리

연속 함수 f(x)의 부정적분 중의 하나를 F(x)라고 한다면 다음과 같은 공식이 성립한다.

$$\int_a^b f(x)dx = [F(x)] = F(b) - F(a)$$



그림참조 :https://datascienceschool.net/view-notebook/c970d15a74d24656988c8f020e96ce45/

문제 :다음 정적분의 값을 구하여라.

문제	$\int_0^2 (3x^2+2x-1)dx$	$\int_{-1}^2 4(x+1)(x^2-1)dx$
정답	10	9

문제 풀이: 실습 파일 :Integral01.py, Integral02.py

정답 풀이	$\int_0^2 (3x^2+2x-1)dx = [x^3+x^2-x] = [8+4-2]-[0+0-0] = 10$
	$\int_{-1}^2 4(x+1)(x^2-1)dx$ $= \int_{-1}^2 (4x^3-4x+4x^2-4)dx = \left[x^4-2x^2+\frac{4}{3}x^3-4x\right]$ $= \left[2^4-2*2^2+\frac{4}{3}2^3-4*2\right]-\left[(-1)^4-2*(-1)^2+\frac{4}{3}(-1)^3-4*(-1)\right] = 9$

정적분의 기본 정리

연속 함수 f(x)의 부정적분 중의 하나를 F(x)라고 한다면 다음과 같은 공식이 성립한다.

$$\int_a^b (ax+b)^n dx = \frac{1}{a(n+1)}(ax+b)^{n+1} + C$$

문제 :다음 정적분의 값을 계산하여라.

문제	$\int_1^2 (x-1)^3 dx$	$\int_0^2 \left(1-\frac{x}{2}\right)^5 dx$	$\int_{-2}^1 (x-1)(x+2)^2 dx$
정답	1/4	1/3	-27/4

실습파일 :Integral04.py

순열(Permutation)

숫자가 적혀 있는 4개의 카드 1, 2, 3, 4가 있다고 가정하자.
서로 다른 두 개의 카드로 만들 수 있는 2자리의 정수의 개수를 알아 보자.
1 다음에 2가 나오고, 2다음에 3이 나오는 소위 사전식 배열법을 사용해보면 다음과 같다.

순열

앞수	뒷수	앞수	뒷수	앞수	뒷수	앞수	뒷수
1	2 3 4	2	1 3 4	3	1 2 4	4	1 2 3

위의 표와 같이 두 자리 정수는 4*3=12(개)이다.
이와 같이 서로 다른 4개에서 2개를 택하여 일렬로 나열하는 것을 4개에서 2개를 택하는 순열이라고 한다.

순열의 공식 :

서로 다른 n 개의 원소에서 r 개를 택하여 순서 있게 늘어 놓은 것을 n 에서 r 개를 택한 순열이라고 한다.
공식은 다음과 같다.

$$ {}_nP_r = \frac{n!}{(n-r)!} $$

실습 파일 : Permutation01.py, permutations_01.R

n 의 계승의 뜻(factorial)

1 에서 n 까지의 자연수의 곱을 n 의 계승이라고 하고, n!으로 나타낸다.

n! = n X (n-1) X … X 3 X 2 X 1

문제 : 순열 문제

10 개의 역이 있는 지하철 회사에서 출발역과 도착역을 명기한 차표를 만들려고 한다.
몇 가지를 만들어야 하는가?
단, 왕복표와 1, 2, 3 등의 구별은 없다.

문제 풀기 :

10개 중에서 2개를 순서를 따지는 것이므로 $_{10}P_2$ = 10! / (10-2)! = 10 * 9 = 90(가지)

실습 파일 : permutations_02.R

문제 : 순열 추가 문제

야구 팀에서 9 명의 타순으로 정하는 방법은 몇 가지인가?

정답 : factorial(9) = 362880 가지

서로 다른 마을에 있는 다섯 친구의 집을 방문하는 경우의 수는 몇 가지인가?
정답 : 120 가지

60 명의 학급에서 반장, 부반장, 학습 부장을 각각 1 명씩 선출하는 방법은 몇 가지인가?
반장, 부반장 및 학습 부장의 구분이 있어야 하므로 이 문제는 순열이다.
정답 : 205320 가지
실습 파일 : permutations_03.R

조합(Combination)

세 사람 A, B, C 중에서 대표를 뽑는다고 가정하자.

순열 문제 : 반장, 부반장 각 1명씩을 뽑는 경우

예를 들어, A(반장), B(부반장)과 A(부반장), B(반장)은 서로 다른 것이다.
이처럼 순서를 따지는 것이므로 이것은 순열이다.

문제 풀기 :

3 명 중에서 2 명를 순서를 따지는 것이므로 $_3P_2$ = 3! / (3-2)! = 6(가지)

조합 문제 : 대표 2 명을 뽑는 경우

순서를 따지지 않는 것이므로 이것은 조합이다.

문제 풀기 :

3 명 중에서 2 명를 순서를 따지는 것이므로 $_3C_2$ = 3! / 2! / (3-2)! = 3(가지)

조합의 공식 :

서로 다른 n 개의 원소에서 순서를 고려 하지 않고, r 개를 택하고자 하는 경우를 조합(Combination)이라고 한다.
공식은 다음과 같다.

$$_nC_r = \frac{n!}{r!(n-r)!}$$

실습 파일 : Combination01.py, combinations_01.R

문제 : 추가 문제

남자 6명, 여자 4명이 있다.
이 중에서 남자 3 명, 여자 2 명을 뽑는 경우의 수는 얼마인가?
정답 : $_6C_3$ * $_4C_2$ = 120 가지

실습 파일 : combinations_02.R

문제 :풀어 봅시다.

홍길동을 포함하여 10 명의 사람이 있다.
이중에서 3 명의 위원을 뽑는다고 가정하자.

1) ´홍길동´이가 반드시 뽑힐 경우의 수는 36 가지이다.
이유를 설명하시오.

정답 : ${}_{9}C_{2}$
홍길동이는 반드시 뽑혀야 하므로, 나머지 9 명에서 2 명을 뽑으면 된다.

2) ´홍길동´이는 절대로 뽑히지 않을 경우의 수는 84 가지이다.
이유를 설명하시오.

정답 : ${}_{9}C_{3}$
홍길동이는 뽑히지 말아야 하므로, 나머지 9 명에서 3 명을 뽑으면 된다.

문제 :풀어 봅시다.
평면 위에 있는 10개의 점중에서 4개의 점이 일직선 상에 있다고 가정하자.
이들 점을 꼭지점으로 하는 삼각형의 개수를 구하여라.
힌트 : 4개의 점은 일직선 상에 있으므로 삼각형을 만들 수 없다.
정답 : 116개

파일 이름 : combinations_03.R

통계의 첫 걸음

변량, 도수 분포
학교에 보관되어 있는 성적 일람표라든지, 키/몸무게 등의 신체 검사표와 같은 통계 자료를 수량으로 나타낼 때 그 수량을 변량이라고 한다.

다음 예시를 이용하여 통계에서 자주 사용하는 용어의 뜻을 살펴 보자.

다음은 3 학년 2 반 학생들의 수학 성적을 명시하고 있는 도수 분포표이다.

점수	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100
학생수	1	3	8	12	18	10	6	2

용어 정리

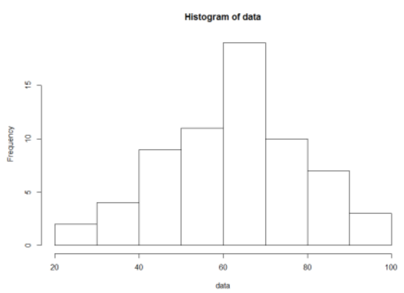
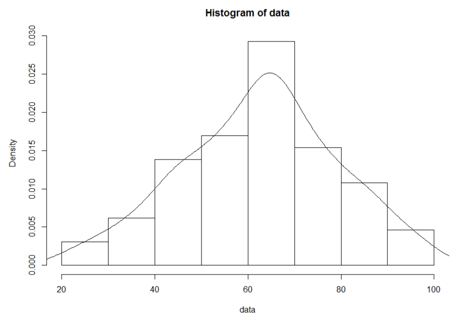
통계와 관련된 용어를 다음과 같이 정리해본다.

변량	대상을 수량으로 나타낸 것(20~30, 30~40 등)을 말한다.
계급	20~30, 30~40 처럼 적당한 간격으로 나누는 것을 말한다.
계급의 크기	구간의 크기 10 을 말한다.
계급값	계급의 가운데 있는 값을 말한다.(25, 35, 45 등등)
도수	변량이 가지고 있는 개수를 의미한다.
분포	자료가 흩어져 있는 모양을 말한다.
도수 분포표	변량과 도수와의 관계를 표로 만든 것이다.

히스토 그램(histogram)

다음 도수 분포표를 그래프로 나타내어 보자.

점수	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100	계
학생수	2	4	9	11	19	10	7	3	65

항목	설명
	
파일 이름 : 히스트그램.R	

왼쪽 그림은 각 계급의 크기(너비)가 같고, 도수가 높이로 나타나 있으므로 직사각형의 넓이는 계급의 도수에 비례한다.
이와 같은 그래프를 히스토그램이라고 한다.

오른쪽 그림은 자료의 수가 많을 때, 계급의 크기를 작게 잡고, 구간 수를 많이 늘려서 얻은 도수 분포 다각형으로 거의 곡선에 가까워질 수 있다.
이것을 도수 분포 곡선이라고 한다.

대표 값, 평균

자료의 중심적인 경향이나 특징을 하나의 수로 나타내어야 하는 경우가 있다.
이러한 개념으로 사용되는 것이 대표 값이다.
대표 값에는 평균(mean), 중앙값(median), 최빈수 등이 있으나, 주로 평균이 많이 쓰인다.

평균

변량의 전체의 합을 변량의 수, 즉 도수의 합으로 나눈 값은 평균 또는 평균 값 또는 산술 평균이라고 한다.
기호는 M 또는 m 또는 $\bar{x}$ 등으로 나타낸다.

문제 : 평균 구하기

다음 학생의 수학 성적의 평균을 구하시오.

회수	1	2	3	4	5	6
김철수	65	80	45	75	85	70

평균을 m이라고 하면, $\mathrm{m} = \dfrac{65+80+45+75+85+70}{6} = 70$이다.
실습 파일 : mean_test_01.R

문제 : 연속 변량의 평균 구하기

씨름부 학생 10명의 몸무게에 대한 도수 분포표가 다음과 같을 때, 이들 학생의 몸무게의 평균을 구하여라.

몸무게	50~60	60~70	70~80	80~90	90~100	계
학생수	1	0	3	4	2	10

자료가 연속 변량으로 주어지는 경우 계급의 평균 값을 이용하여 풀면 된다.
평균을 m이라고 하면, $\mathrm{m} = \dfrac{55*1+65*0+75*3+85*4+95*2}{1+0+3+4+2} = 81$이다.

실습 파일 : mean_test_02.R

확률, 통계 분석

빅 데이터 분석은 기존의 사실에 대한 객관적인 근거를 제시하고, 다변화된 현대 사회를 정확하게 예측 및 대응하여 우리 사회의 전 분야에 거쳐서 가치 있는 정보를 제공하는 신기술의 학문이다.

이러한 빅 데이터 분석에서 확률과 통계는 중요한 영역을 차지한다.
분석을 위하여 필요한 확률과 통계의 개념에 대해서 살펴 본다.

확률 변수와 확률 분포

동전의 앞면에 1의 눈이, 뒷면에 2의 눈이 새겨진 있다고 가정하자
이와 같은 2개의 동전 A, B를 동시에 던질 때, 표본 공간 S는 다음과 같다.

S = {(1, 1), (1, 2), (2, 1), (2, 2)}

이 때 나오는 눈의 합을 X라고 하면, X가 취할 수 있는 값은 2, 3, 4을 가질 수 있다.

A \ B	1	2
1	2	3
2	3	4

이들 값을 취할 확률은 1/4, 2/4, 1/4이므로 이 대응 관계를 표로 나타내면 다음과 같다.

X	2	3	4
P(X)	1/4	2/4	1/4
확률 분포표			

이와 같이 변수 X가 취할 수 있는 각 값에 확률이 정해져 있을 때, 변수 X를 확률 변수라고 하고, 위와 같은 대응 관계를 확률 변수 X의 확률 분포라고 한다.

항목	설명
확률 변수	변수 X가 각 값에 취할 수 있는 확률이 정해져 있을 때, **변수 X를 확률 변수**라고 한다.
확률 분포	위와 같은 대응 관계를 **확률 분포 표**라고 한다.

문제 : 확률 구해보기

확률 변수 X의 분포가 다음과 같다고 가정하자.
이 때 다음의 각 값의 확률을 구하시오.

X	0	1	2	3	4	합
P(X)	0.1	0.25	0.3	0.25	0.1	1

문제 풀이

조건	P(X=2)	P(X=2 or X=4)	P(X<2)	P(0<=X<=2)	P(X>=2)
정답	0.3	0.4	0.35	0.65	0.65

확률 변수의 평균과 표준 편차

다음 표와 같이 상금이 붙은 10개의 제비가 있다고 가정하자.

등급	1등	2 등	3 등	4 등
제비수	1	2	3	4
상금(원)	1000	500	200	0

위의 제비에서 1개를 뽑을 때 상금의 액수를 X원이라고 한다면, X가 취할 수 있는 값과 취할 수 있는 확률은 다음과 같다.

X	1000	500	200	0
P(X)	1/10	2/10	3/10	4/10

제비 1개에 대한 상금은 평균 금액은 다음과 같다.

평균 금액
(1000*1+500*2+200*3+0*4) / 10 = 260 원

즉, 제비 1개를 가지고 있는 사람은 260원의 기대치가 있게 된다.
이런 의미에서 이 값을 확률 변수 X의 평균 또는 기대값(Expectation)이라고 한다.

확률 변수의 평균 공식

확률 변수 X의 확률 분포가 다음과 같다고 가정하자.

X	x_1	x_2	x_3	...	x_n	합
P(X)	p_1	p_2	p_3	...	p_n	1

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

를 X의 평균 또는 기대 값이라고 하고, m 또는 E(X)로 표시한다.

또한, 분산 공식은 다음과 같이 "확률 변수의 제곱"에 "평균의 제곱" 값을 뺄셈하여 구할 수 있다.

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

문제 : 확률 변수의 평균과 표준 편차 구하기

한 개의 주사위를 던질 때 , 나오는 눈을 확률 변수 X 라 한다.
X 의 평균, 분산 및 표준 편차를 구하여라.

X	1	2	3	4	5	6	합
P(X)	1/6	1/6	1/6	1/6	1/6	1/6	1

평균(m)	1*1/6+2*1/6+3*1/6+4*1/6+5*1/6+6*1/6 = 7/2
분산(V(X))	(1)^2*1/6+(2)^2*1/6+(3)^2*1/6+(4)^2*1/6+(5)^2*1/6+(6)^2*1/6 - (7/2)^2 = 35/12
표준 편차	분산에 루트를 씌운 값이다.

실습 파일 : probability01.R

문제 :
아래와 같은 상금이 걸린 20장의 복권이 있다.
이 중에서 복권 1 장을 뽑을 때의 기대값(평균)을 구하시오.

등급	1등	2 등	3 등	4 등
복권 개	1	3	6	10
상금(원)	1000	100	10	0

정답 : 68원

probability02.R

이항 정리

다음 공식인 일반적인 곱셈 공식이다.

$(a + b)^2 = a^2 + 2ab + b^2$ ... 식(1)
$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$ ... 식(2)

이제, 수식들을 우변의 각 항의 계수를 조합 기호로 $_nC_r$을 사용하여 표시해보도록 하자.
우선 식(1)을 살펴 보자.
제곱의 구하는 것이므로 n = 2이고, r은 0부터 2까지이다.
1 = $_2C_0$, 2 = $_2C_1$, 1 = $_2C_2$이므로, 부견의 계수를 조합 기호로 사용하여 나타내면 다음과 같다.
$(a + b)^2 = {}_2C_0a^2 + {}_2C_1ab + {}_2C_2b^2$

식(2)는 n = 3이고, r은 0부터 3까지이다.
$(a+b)^3 = {}_3C_0a^3 + {}_3C_1a^2b + {}_3C_2ab^2 + {}_3C_3b^3$ 으로 표현이 가능하다.

실습 파일 : binomial01.py

이항 정리의 일반식

n이 양의 정수일 때, $(a+b)^n$의 전개식은 다음과 같다.
$(a+b)^n = {}_nC_0a^n + {}_nC_1a^{n-1}b + \cdots + {}_nC_ra^{n-r}b^r + {}_nC_nb^n$
이 식을 이항 정리라고 한다.

문제 : 이항 정리를 이용하여 다음 식을 전개하여라.
문제 : $(a+2b)^6$
정답 : $a^6 + 12a^5b^1 + 60a^4b^2 + 160a^3b^3 + 240a^2b^4 + 192a^1b^5 + 64b^6$

실습 파일 : binomial02.py

이항 분포(Binomial Distribution)

일상 생활에서 2가지 중에서 하나를 선택해야 하는 경우가 있다.

이항 분포의 예시

표본 조사에서의 찬성 또는 반대의 결정
출생하는 아기의 성별 판단
어떤 제품이 합격품/불합격품 여부

이렇듯 2개 중에서 하나의 값을 취할 수 있는 확률 분포를 이항 분포라고 한다.
이항 분포는 시행 횟수가 많아질수록 정규 분포에 가까워진다.

다음의 예시를 살펴 보자.
한 개의 주사위를 3번 던질 때 1의 눈이 나오는 횟수를 X라 하면, X가 취할 수 있는 값과 확률은 다음과 같다.

한 번 던질 때1의 눈이 나올 확률은 1/6이므로 확률 분포표는 다음과 같이 표현할 수 있다.

X	0	1	2	3	합
P(X)	${}_3C_0(1/6)^0(5/6)^3$	${}_3C_1(1/6)^1(5/6)^2$	${}_3C_2(1/6)^2(5/6)^1$	${}_3C_3(1/6)^3(5/6)^0$	1

위와 같은 확률 분포를 이항 분포라고 하고, 기호 B(3, 1/6)으로 표현한다.
B는 Binomial의 줄인말이다.

B(3, 1/6)	3은 독립 시행의 횟수를 의미한다. 즉, 3번 시행한다는 것이다.
	1/6 은 한 번 일어날 확률이다. 주사위 눈금 1 이 일어날 확률이다.

문제 : 이항 분포 문제

확률 변수 X가 이항 분포 B(2, 1/3)을 따를 때, X의 확률 분포표를 만들어라.

X	0	1	2	합
P(X)	${}_2C_0(1/3)^0(2/3)^2$	${}_2C_1(1/3)^1(2/3)^1$	${}_2C_2(1/3)^2(2/3)^0$	1
P(X)	4/9	4/9	1/9	

이항 분포의 평균, 분산, 표준 편차

확률 변수 X는 특히 이항 분포를 이룰 때, X의 평균, 분산, 표준 편차는 다음 공식에 의하여 간단히 구할 수 있다.

공식 :

확률 변수 X 가 이항 분포 B(n, p)을 따를 때, X 의 평균, 분산은 다음과 같다.
단, p + q = 1 이다.

E(X) = np, V(X) = npq, $\sigma(X) = \sqrt{npq}$ (단, q = 1 - p)

문제 : 이항 분포의 평균과 분산 구하기

어떤 질병에 대한 치유율이 90%인 의약품으로 10명의 환자가 치료를 받고 있다.
치유될 환자의 수를 X라고 할때, X의 평균과 분산을 구하시오.

문제 풀기 :

평균 = np = 10*0.9 = 9

분산 = npq = 10*0.9*0.1 = 0.9

연속 확률 변수와 확률 분포

주사위의 눈금, 동전의 앞면 등은 연속적인 데이터가 아니다.
일반적으로 이산형 데이터 형식으로 되어 있다.
이러한 자료에서 다루는 확률 변수를 이산 확률 변수라고 한다.

반면, 시간, 온도, 사람의 키 등과 같이 연속적으로 어떤 구간 내에 있는 모든 값을 취하는 확률 변수에 대하여 살펴 보도록 하자.

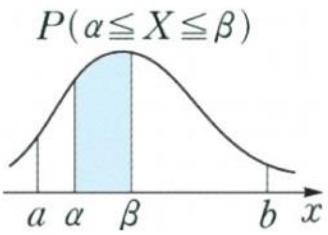
원의 중심 O를 중심으로 자유롭게 도는 바늘이 달린 숫자 판이 있다고 가정하자.
이 바늘을 힘껏 회전시켰을 때, 바늘이 저절로 멈추면서 가리키는 눈금을 X라고 한다면 X는 0에서 1까지의 모든 값을 취할 수 있는 확률 변수이다.
이 때 바늘 끝은 구간 [0, 1]의 어느 값이라도 모두 균등하게 가리킬 수 있으므로 X가 구간 [0, 1] 사이에 속할 확률은 1이다.

구간 [0.2, 0.5] 사이에 놓일 확률은 0.5 - 0.2 = 0.3이다.

이제 구간 [0, 1]에서 함수 f(x)를
f(x) = 1 (0 <= x <= 1)
이라고 정의 하면 곡선 y = f(x)의 넓이는 1이다.

연속 확률 변수, 확률 밀도 함수

변수 X가 어떤 구간 [a, b] 안의 모든 값을 취하고, f(x)가 구간 [a, b]를 정의역으로 하는 함수로써 다음 세 조건

(1) f(x) ≧ 0(단, 0 ≦ x ≦ b) # 모든 확률은 0 이상의 값을 가져야 한다.
(2) $\int_a^b f(x)dx = 1$ # 모든 확률의 합은 반드시 1이다.
(3)$P(\alpha \le x \le \beta) = \int_\alpha^\beta f(x)dx$ (단, $a \le \alpha \le \beta \le b$)


를 만족할 때, 변수 X를 연속 확률 변수, f(x)를 확률 밀도 함수라 한다.

우리가 많이 알고 있는 **정규 분포**도 연속 확률 밀도 함수의 한 종류이다.
연속 확률 밀도 함수에 대한 문제를 해결하려면 **적분에 대한 기본 이해**가 조금 필요하다.

연속 확률 변수의 평균, 분산, 표준 편차

설명 :

연속 확률 변수 X가 구간 [a, b]의 모든 값을 취하고, 이 구간에서 정의된 확률 밀도 함수가 f(x)로 주어질 때, X의 평균, 분산 및 표준 편차는 다음과 같이 정의한다.

평균 : $m = E(X) = \int_a^b x \cdot f(x)dx$
분산 : $V(X) = \int_a^b (x-m)^2 f(x)dx = \int_a^b x^2 f(x)dx - m^2$
표준 편차 : $\sigma(X) = \sqrt{\int_a^b (x-m)^2 f(x)dx} = \sqrt{\int_a^b x^2 f(x)dx - m^2}$

문제 : 연속 확률 변수

확률 변수 X의 확률 밀도 함수 f(x)가 $f(x) = kx^3$ (0 <= x <= 1)로 주어질 때, 다음 각 값을 구하시오.

1) 상수 k	2) P(X<=0.5)	3) P(X<=a) = 0.0625인 a
4	0.0625	0.5
4) 평균 E(X)	5) 분산 V(X)	6) 표준 편차
0.8	2/75	sqrt(6)/15

$$\int_0^1 kx^3 dx = [\frac{1}{4}kx^4]_0^1 = \frac{1}{4}k - 0 = \frac{k}{4} = 1$$

$$\int_0^{0.5} 4x^3 dx = [x^4]_0^{0.5} = 0.5^4 - 0 = 0.0625$$

$$a^4 = 0.0625 \quad \therefore a = 0.5$$

$$\int_0^1 xf(x)dx = \int_0^1 4x^4 dx = [\frac{4}{5}x^5]_0^1 = \frac{4}{5} = 0.8$$

$$\int_0^1 x^2 f(x)dx - Ex^2 = \int_0^1 4x^5 dx - (0.8)^2 = [\frac{4}{6}x^6]_0^1 - 0.64$$

$$= \frac{2}{75} = 0.02666667$$

$$\sqrt[2]{(\frac{2}{75})} = 0.163299317$$

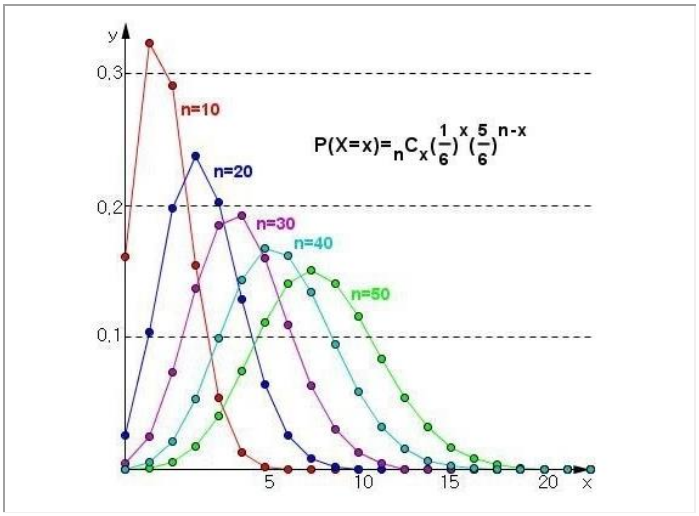
실습 파일 : probability_density_01.py

이항 분포와 정규 분포와의 관계

주사위를 n 회 던져서 1 의 눈이 r 번 나오는 확률을 Pr 이라고 할 때

Pr = nCr(1/6)^r(5/6)^(n-r) (n = 1, 2, 3 ...)

이 식에서 n 의 값을 충분히 크게 하면 다음 그림과 같다.



그래프에서와 같이 이항 분포 B(n, p)의 그래프가 n이 커질 때 정규 분포 곡선에 가까워진다.

따라서, n이 클 때에는 이항 분포 B(n, p)에서의 확률의 근사 값은 정규 분포 N(np, npq)를 이용하여 구한다.

정규 분포

연속 확률 변수는 연속된 구간 내의 모든 값을 취할 수 있다.
그러므로, 특정한 1개의 점을 취할 확률은 거의 0에 가깝다.
가령, 어느 집단에서 한 사람을 뽑았을 때 이 사람의 키가 167.23이 될 확률은 거의 0이라고 보면 된다.
그러나 실제로 키를 잴 때는 측정의 정확도에 한계가 있어 우리는 일반적으로 이산형과 같이 관측하게 된다.
한 사람의 키가 167이라면 이는 166.5 ~ 167.5 사이의 숫자라고 이해할 것이다.

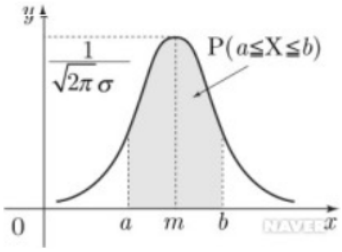
정규 분포는 도수 분포 곡선이 평균 값을 중심으로 좌우 대칭인 종(Bell-Shape) 모양을 이루고 있다.
K.F 가우스가 측정 오차의 부분에서 중요성을 강조하였기 때문에 **가우스 분포**라고도 한다.
모든 정규 분포의 확률을 구하기 위하여 평균(0), 표준 편차(1)의 표준화 시킨 정규 분포를 **표준 정규 분포**라고 한다.
평균과 표준 편차에 의하여 그 모양이 결정된다.

정규 분포 곡선의 성질 :

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

정규 확률 변수 X의 평균은 m, 분산은 σ^2이므로 이때 X는 평균 m, 분산 σ^2의 정규 분포를 따른다. 라고 말하고 N(m, σ^2)으로 표기한다.

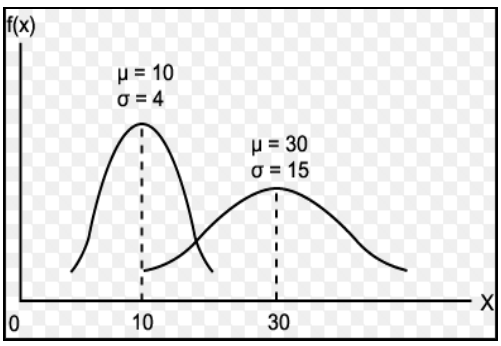
(1) $y = f(x)$의 그래프는 직선 $x = m$에 관하여 대칭이다.
(2) 곡선과 x축 사이의 넓이는 1이다.
(3) $x = m$일 때 최댓값 $\frac{1}{\sqrt{2\pi}\,\sigma}$을 가지며, x의 값이 m에서 멀어질수록 $f(x)$의 값은 0에 가까워진다. 즉 정규분포곡선은 x축을 점근선으로 한다.
(4) X가 구간[a, b] 에 속할 확률P($a \le X \le b$)는 구간[a, b]에서 곡선과 $x = a$, $x = b$, x축으로 둘러싸인 넓이와 같다.
(5) m의 값이 일정할 때, σ의 값이 커질수록 곡선의 모양은 낮아지며 좌우로 퍼진다.
(6) σ의 값이 일정하고 m의 값이 달라지면 곡선의 모양은 같고 대칭축만 바뀐다.

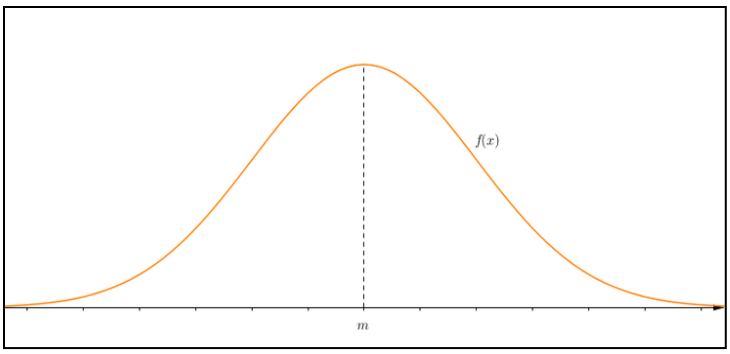


중심 극한 정리

일반적으로 표본의 크기가 커지면 근사적으로 모평균은 표본 평균과 같다.
또한, 모분산도 표본 분산과 같은 정규 분포를 이룬다.
최소 도수의 개수가 30이상이어야 한다.

구분	특징
변수	연속 변수
분포	평균을 중심으로 좌우 대칭 구조를 이룬다.
대표값	평균 = 중앙값 = 최빈값
왜도/첨도	왜도 = 0, 첨도 = 0 또는 3
모양	표준 편차에 의하여 모양이 달라진다.
위치	평균에 의하여 위치가 달라진다.
넓이	정규 분포의 전체 면적은 1 이다.





문제 : 정규 분포 확률 문제

확률 변수 X가 정규 분포 N(70, 3^2)를 따른다고 한다.
즉, 평균이 70이고 분산이 9라고 가정할 때 다음 표를 이용하여 각각의 확률을 구해 보세요.
여기서, m은 평균이다.

a	P(m <= X <= a)
m + σ	0.3413
m + 2σ	0.4772

문제 : 풀이

평균이 70이므로 m=70이고, 분산이 9이므로 표준 편차 σ는 3이다.
또한, 76 = 70 + 2 * 3 = m + 2 * σ이다.

문제 1) P(70 <= X <= 76)	P(70<=X<=76) = P(70<=X<=70+2*3) = P(m <= X <= m+2σ) = 0.4772
문제 2) P(67 <= X <= 73)	P(67<=X<=73) = P(70-3<=X<=70+3) = P(m-σ<=X<=m+σ) = 2 * P(m<=X<=m+σ) = 2*0.3413 = 0.6826

실습 파일 : normal_distribution_01.py, normal_distribution_01.R

균등 분포

균등 분포(uniform distribution)은 연속형 확률 분포 중에서 가장 간단한 형태의 분포이다.
어떠한 구간 [min = a, max = b]에서 값이 균등하게 퍼져 있는 집단, 즉 일어날 확률이 균등한 분포를 말한다.

확률변수 X가 구간 (a, b)에서의 확률밀도함수가

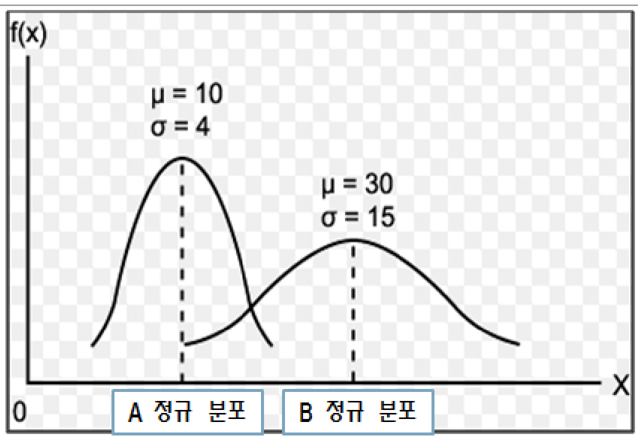
$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{기타} \end{cases}$$

표준 정규 분포

정규 분포를 대상으로 표준화한 표준 정규 분포의 과정을 파악해 보고, 이를 토대로 신뢰 수준과 신뢰 구간의 의미를 자세히 알아보도록 한다.

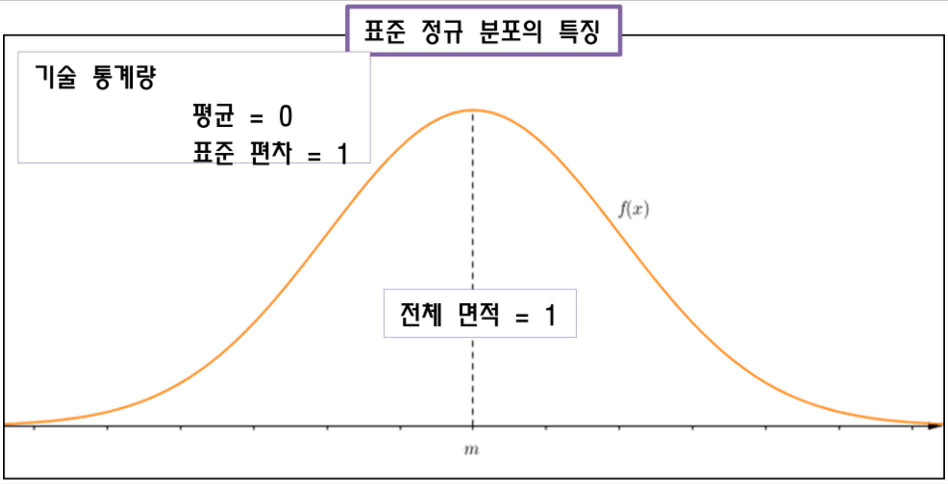
표준 정규 분포

평균과 분산이 다른 여러 개의 정규 분포를 표준화시킨 분포를 말한다.
평균과 표준 편차를 각각 0과 1로 고정 시키는 과정을 의미한다.



그림을 보면 A와 B 정규 분포가 있다.
각각의 평균과 표준 편차가 서로 달라서, 확률을 구하기 위하여 서로 접근하는 방법이 다르다.

이러한 분포를 평균 0, 분산 1로 표준화시킨 분포가 표준 정규 분포이다.



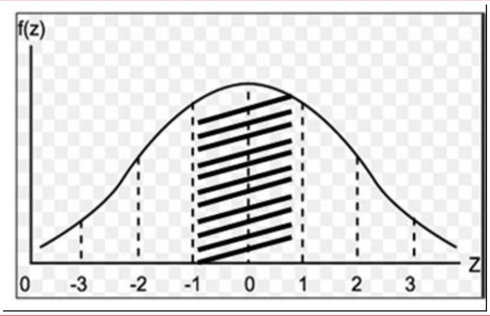
표준화 변수 Z

정규 분포의 확률 변수 X를 구하기 위하여 표준 정규 분포로 바꾸는 변수를 표준화 변수라고 한다.
공식은 다음과 같다.

정규 분포의 표준화 공식 :	
$\mathrm{X} \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow Z \sim N(0,1)$ $\mathrm{Z} = \dfrac{\mathrm{X}-\mu}{\sigma}$ X = σ * Z + μ	X : 확률 변수 μ 또는 m : 평균 σ : 표준 편차

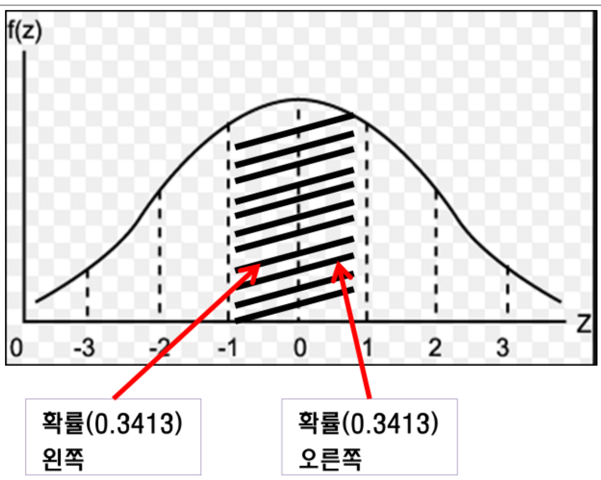
문제 : 표준 정규 분포 확률 문제

강남 고등학교 2학년 6반 학생들의 국어 점수가 평균 75점, 표준 편차가 5점인 정규 분포를 나타내고 있다.
이 경우 어느 학생의 점수가 70점 ~ 80점 사이일 확률은 얼마인가?
힌트 : 표준 정규 분포표를 이용하시오.

문제 풀이 :	
Z = (X - μ) / σ 에 의하면 70 점은 Z = (70 - 75) / 5 = -1 이고, 80 점은 Z = (80 - 75) / 5 = 1 이다. 즉, P(70<=X<=80) = P(-1<=Z<=1)이므로 평균 0 을 기준으로 ±σ의 표준 정규 분포로 나타낼 수 있다. 표준 정규 분포표를 보면 1.0에 해당하는 값이 0.3413이므로 구하고자 하는 확률은 0.6826이다.	

표준 정규 분포표

정규 분포는 확률을 구하기 위하여 해당 영역의 면적을 구해야 한다.
이것을 표준 정규화시켜서 만든 표를 표준 정규 분포표라고 한다.

 	앞의 예시에서 Z의 확률 구간은 P(-1 < Z < +1)이었다. 이 구간에 해당하는 확률을 표(하단의 표 참고)에서 찾아 보면 0.3413이다. 또한 좌우 대칭이므로 확률은 0.6826이 된다. 다시 말해서, 평균이 75점, 표준 편차가 5점일 때 어느 학생의 점수가 70~80점 사이일 확률은 약 68.26%가 된다는 의미이다.

표준 정규 분포표

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

문제 : 정규 분포 확률 문제

어느 고등학교 1학년 학생 1000명의 성적 분포가 평균 80점, 표준 편차 20점인 정규 분포로 나타났다.
이 경우에 60점이상 100점 이하의 점수를 얻은 학생은 대략 몇 명인가?
힌트) 표준화변수 z와 z값과 확률구간 참조
정답 : 약 682명

실습 파일 : normal_distribution_02.py

문제 : 풀어 봅시다.

확률 변수 X가 정규 분포 N(80, 25)를 따른다고 한다.
다음 확률을 구하시오.
P(75 <= X <= 90)

정답 : 0.8185
실습 파일 : normal_distribution_03.py

문제 : 표준화 변환 문제

1000 점 만점인 어느 대학교의 입학 시험에서, 지원자 2000 명의 득점 분포는 평균 450 점, 표준 편차 75 점의 정규 분포를 이룸을 알았다.
정규 분포표를 이용하여 다음 물음에 답하시오.
(1) 600 점 이상 득점한 수험생은 몇 명인가?
(2) 이 대학의 입학 정원이 320 명이라고 할 때, 합격의 최저 점수는 몇 점인가?

실습 파일 : standard_01.R

풀이 :

n = 2000
mu = 450
sigma = 75
Z = (X-mu)/sigma = (X-450)/75
X = 600일때 Z=2이고, 이때 확률은 0.4772
따라서 , n*(0.5-0.4772)=2000*(0.5-0.4772)=45.6명

문제 : 풀어 봅시다.

어떤 학년의 수학 성적은 평균 72점, 표준 편차 9점의 정규 분포를 이룬다고 가정하자.
수험생 중에서 상위 10%는 평점 ´수´가 주어진다고 할 때, 수험자가 평점 ´수´를 받기 위해서는 적어도 몇점을 받아야 하는가 ?

정답 : 84점

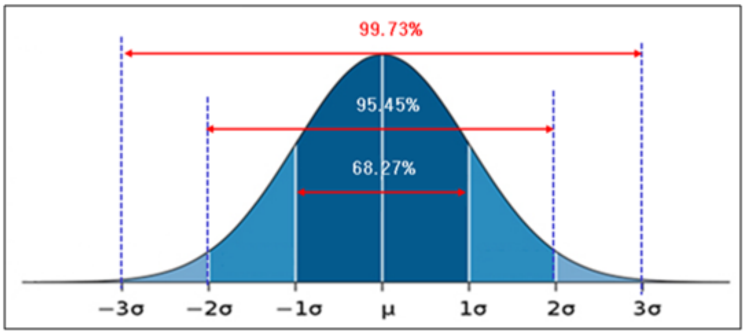
Z 값과 확률 구간

평균과 표준 편차에 의하여 산출된 Z의 값과 표준 정규 분포의 확률을 정리하면 다음과 같다.

예를 들어 2.0 을 표준 정규 분포포에서 찾아 보게 되면 0.4772 이다.
따라서, ±2σ는 2 * 0.4772 = 0.9544 즉 95.44%이다.

Z 값과 확률

Z 값	확률
평균 ±1σ	68.26%
평균 ±2σ	95.44%
평균 ±3σ	99.74%



평균 m을 중심으로 데이터가 ±σ내에 들어 있을 확률은 68.26%라는 의미이다.

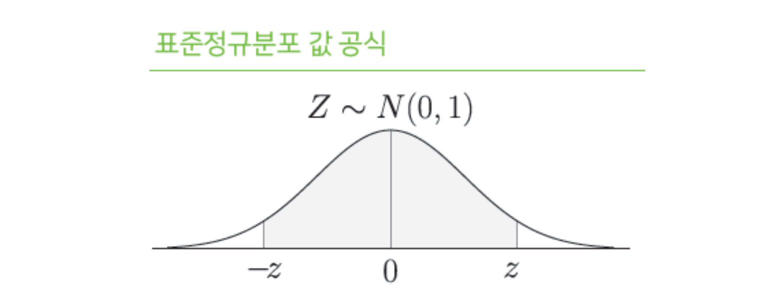
추정(estimation)과 신뢰도

전국 고등학교 3학년을 모집단으로 하여 임의의 표본 1,600명을 뽑아서 신장을 조사해보니 평균 165cm, 표준 편차 4를 얻었다고 가정하자.
전국 고교 3학년의 전체의 평균을 165라고 단정 지을 수 없겠지만, 165라는 값 근처의 어느 값이라고 판단은 가능하다.
예를 들어서, "**대략 160에서 170사이**"라고 판단할 수 있을 것이다.(**신뢰 구간**이라고 한다.)
그런데 이 말은 약 "**90% 정도 믿어도 좋다**"라고 적당히 대답할 것이다.(**신뢰도**라고 부른다.)

신뢰 수준과 신뢰 구간

신뢰 수준이란 계산된 구간이 모수를 포함할 확률을 말한다.
사회 과학 분야에서 일반적으로 사용하는 신뢰 수준은 90%, 95%, 99%을 사용한다.
의학 및 생명 과학 분야에서는 99%를 많이 사용한다.
신뢰 수준에 해당하는 Z 값을 구하면 다음과 같다.

신뢰 수준(확률)	Z 값
90%	1.65
95%	1.96
99%	2.58
표준 정규 분포 표에서 1.96을 찾아 보면 값이 0.475이다.(즉, 95%의 절반이다.)	



신뢰 구간이란 신뢰 수준하에서 모수를 포함하는 구간을 말한다.
유의 수준과 신뢰 수준에 의해서 결정된 표준화 변수 Z 값은 다음과 같다.

유의 수준(α)	신뢰 수준(%)	Z 값
0.10	90%	1.65
0.05	95%	1.96
0.01	99%	2.58

예를 들어 유의 수준이 0.05이라는 말은 "표본에 대한 신뢰도가 95%이상이다"라는 의미이고, 우리가 어떠한 "의사 결정을 할 때 실수할 확률이 5%정도 있다."라는 의미이다.

신뢰 구간 추정식

모집단이 정규 분포를 이루고 있다면 다음과 같은 ´모 평균의 신뢰 구간 추정식´에 의하여 신뢰 구간을 계산할 수 있다.

신뢰 수준(확률)	Z 값
90%	1.65
95%	1.96
99%	2.58

신뢰도가 95%일 때 모평균의 신뢰 구간

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

X: 평균, Z : 표준화 변수, μ : 모평균, σ : 표준 편차, n : 응답자수, α : 유의 수준

문제 : 신뢰 구간 구하기 문제

다음은 통계 분석 패키지에 의하여 나타난 결과이다.
이것을 이용하여 신뢰 구간을 구해보자.
전체 응답자는 290명이고, 평균은 45.11이다.
표준 편차는 13.752이고 α : 0.05(신뢰 수준 95%), Z : 표준화 변수는 1.96이다.
계산된 통계량을 신뢰 구간 추정식에 적용하면 다음과 같다.

신뢰 구간 추정식을 토대로 신뢰 구간을 나타내면 다음과 같다.

신뢰 구간 :
P(45.11 – 1.96 * 13.752 / 루트(290) <= μ <= 45.11 + 1.96 * 13.752 / 루트(290)) = **0.95**

따라서, 평균은 45.11은 신뢰 수준 95%에서 신뢰 구간 하한값 : 43.528, 상한값 : 46.692으로 계산된다.

신뢰 구간 :
[45.11 – 1.96 * 13.752 / 루트(290), 45.11 + 1.96 * 13.752 / 루트(290)] = [**43.528, 46.692**]

실습 파일 : confidence_level_01.R

문제 : 신뢰 구간 구하기 문제2

크기 n인 표본으로 신뢰 수준 95%를 갖도록 모평균을 추정하였더니 신뢰 구간의 길이가 10이었다.
동일한 조건하에서 표본의 크기만을 1/4로 줄이면 신뢰 구간의 길이는 어떻게 되는 가?
① 1/4로 줄어든다. ② 1/2로 줄어든다. ③ 2배로 늘어난다. ④ 4배로 늘어난다.

문제 : 평균 수명 추정하기

전구를 대량으로 생산하고 있는 공장에서 임의의 100개의 전구를 추출하여 수명을 조사하였다.
평균이 500 시간, 표준 편차가 40 시간이었다.

소스 코드 :
문제 1) 신뢰도 95%로 모집단에서의 평균 수명을 추정하여라. 문제 2) 신뢰도 99%로 모집단에서의 평균 수명을 추정하여라.
정답 1) 492.16 <= m <= 507.84

정답 2) 489.68 <= m <= 510.32

신뢰도 95%로 풀어 보는 신뢰 구간은 다음과 같다.
신뢰도 99%로 풀어 보는 경우에는 Z = 2.58 값을 사용하면 된다.

95%로 풀어 보는 신뢰 구간 :
[500 – 1.96 * 40 / 루트(100), 500 + 1.96 * 40 / 루트(100)] = [**492.16, 507.84**]

문제 : 드론의 평균 수명 추정하기

(주)대한 물산은 드론을 생산하는 회사이다.
이 회사의 드론 수명은 정규 분포를 따르고, 모 표준 편차는 200 시간으로 알려졌다.
무작위로 100개의 드론을 표본으로 추출하여 드론의 수명을 측정하였더니 평균 수명이 3,000시간으로 나타났다.
이때 (주)대한 물산에서 생산되고 있는 드론의 평균 수명에 대한 99% 신뢰 구간을 구하시오.
힌트) 모평균의 신뢰구간 추정 식 참조

정답 : [2948.4, 3051.6]

실습 파일 : confidence_level_02.R

표본 오차

표본에서 계산된 추측 값과 모집단의 실제 값과의 차이를 의미 한다.
전수조사를 하지 않는 이상 반드시 생길 수 밖에 없는 오차이다.
여론 조사 발표시 몇 % 신뢰 수준에 오차는 ±몇%라고 하는 것이 바로 표본 오차이다.

신뢰 수준(95%, 99% 등등)이 결정되면 다음과 같은 '허용 오차 계산식'에 의해서 계산될 수 있다.

허용 오차 계산식 :	
$\pm Z\sqrt{\frac{p(1-p)}{n}}$	Z : 표준화 변수 n : 표본 수 p : 확률

문제 : 표본 오차 구하기

20세 이상 유권자 1,500명을 대상으로 A 후보 대선 출마에 대한 찬성과 반대를 조사하는 설문 조사를 시행하였다.
설문 조사 95% 신뢰 수준에서 찬성 55%, 반대 45%가 나왔다.
이때 표본 오차는 얼마인가?

소스 코드 :

신뢰 수준 95%의 Z 값은 1.96, n = 1500, p = 0.55, (1 - p) = 0.45 이다.
이것을 공식에 대입하면 0.025177 이 출력된다.

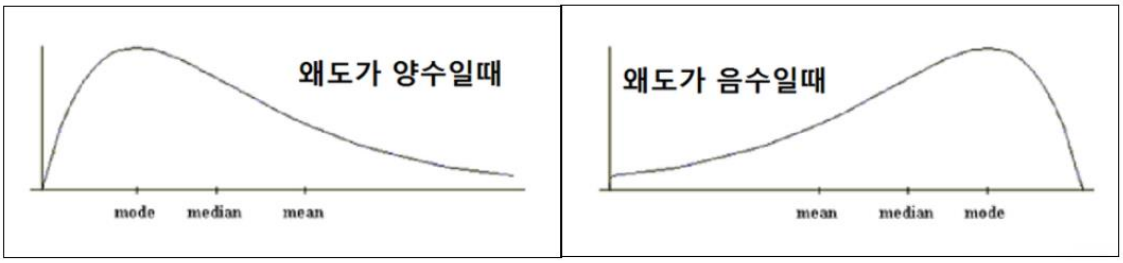
즉, ±2.5%의 표본 오차가 나온다.

왜도와 첨도

왜도는 평균을 중심으로 한 확률 분포의 비대칭 정도를 나타내는 지표이다.
즉, 분포의 기울어진 방향과 정도를 나타내는 양을 의미한다.

항목	설명
왜도	정규 분포와 다르게 얼마나 비대칭인가?
첨도	정규 분포에 비하여 얼마나 더/덜 뾰족한가?

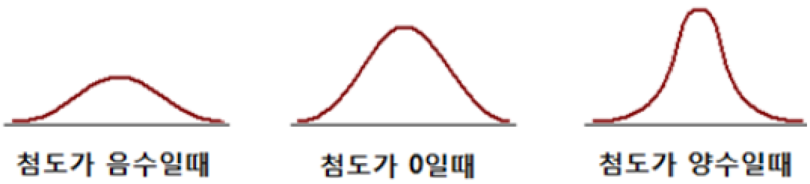
관련 그림



´평균(mean)´, 중위수(median), 최빈수(maximun)가 모두 동일한 경우에는 좌우 대칭인 분포가 되며, 이는 곧 정규 분포이다.

첨도는 표준 정규 분포와 비교하여 얼마나 뾰족한지를 측정하는 지표이다.

관련 그림



실습 파일 : 왜도와 첨도 확인하기.R

통계 관련 용어

추론 통계 분석을 위해서는 통계와 관련된 용어들에 대하여 정리할 필요가 있다.

통계학 분류

통계학(Statistics)이란 논리적 사고와 객관적인 사실에 의거하며, 일반적이고 확률적 결정론에 의해서 인과 관계를 규명한다. 특히 연구 목적에 의해 설정된 가설들에 대하여 분석 결과가 어떤 결과를 뒷받침하고 있는지를 통계적 방법으로 검정할 수 있다.

현재 통계학은 사회학, 경제학, 경영학, 정치학, 교육학, 공학/의학/생명 등 대부분의 모든 학문 분야에서 폭넓게 이용되고 있다.

구분	기술(Descriptive) 통계학	추론(Inferential) 통계학
기능	수집된 자료의 특성을 쉽게 파악하기 위해서 자료를 정리 및 요약	모집단에서 추출한 표본의 정보를 이용하여 모집단의 다양한 특성을 과학적으로 추론
방법	표, 그래프, 대푯값 등	회귀 분석, T-검정, 분산 분석 등

전수 조사와 표본 조사

우리 나라 전체의 인구 조사는 전체 인구를 정확히 파악하기 위하여 국민 전체를 대상으로 빠짐 없이 조사해야 한다.
이를 전수 조사라고 한다.
조사의 대상이 되는 집단 전체를 모집단이라고 부른다.

이에 대하여 전구의 수명을 조사할 때는 한번 검사를 받은 전구는 사용을 못하게 되므로, 전체의 일부를 검사하여 전체를 추측하는 것이 좋다.
이를 표본 조사라고 한다.
전체에서 조사 대상의 일부를 추출해야 하는 데 이것을 표본이라고 부른다.

유형	항목	설명
전수 조사	장점	전체를 대상으로 실시하기 때문에 모집단의 특성을 정확히 반영할 수 있다.
	단점	시간과 비용이 많이 소요된다.
	예시	인구 조사
표본 조사	장점	시간과 비용을 줄일 수 있다.
	단점	표본이 추출되지 못할 경우 수집된 자료가 무용지물이 될 수 있다.
	예시	선거 여론 조사, 임상 실험 등등

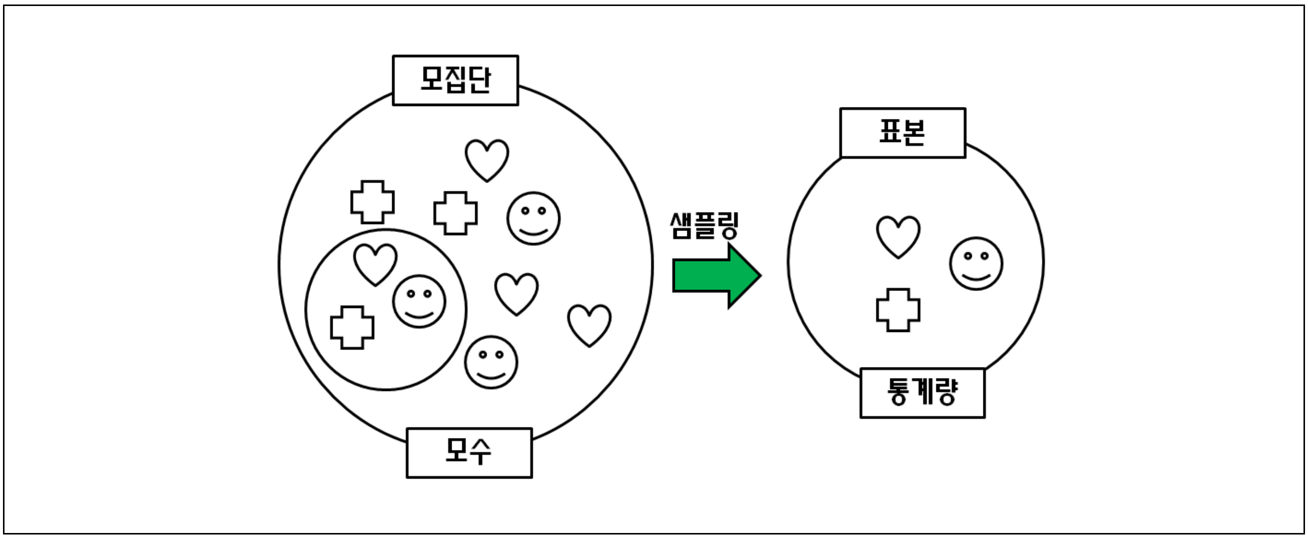
모집단과 표본(카페 1223)

통계와 관련된 용어들을 다음과 같이 정리해본다.

유형	설명	예시
모집단	조사 대상이 되는 자료 전체를 말한다.	전국 고등학교 3 학년 학생
모수	모집단의 특성을 나타내는 수치를 말한다. ↔ 통계량	모 평균/모 표준 편차 등
표본	모집단에서 뽑은 일부의 자료를 말한다.	
통계량	표본의 성격을 나타낸 수치를 말한다. ↔ 모수	표본 평균/ 표본 표준 편차 등
전수 조사	조사의 대상 전체를 빠짐 없이 조사하는 것을 말한다. 시간과 비용이 많이 소요 된다.	인구 조사
표본 조사	대상이 되는 자료 중에서 그 일부분을 빼내서 조사하고 전체를 추정하는 방법이다.	선거 여론 조사, 마케팅 조사 등등
샘플링(추출)	모집단으로부터 표본을 뽑는 행위를 말한다.	
표본의 크기	표본에 포함되는 자료의 개수를 말한다.	
복원 추출	추출된 내용을 다시 도로 넣으면서 추출하는 방법이다. ↔ 비복원 추출	
추정 (estimation)	표본(sample)을 이용하여 모집단(모평균, 모표준 편차 등 등)의 값들을 확률적으로 추측하는 방법이다.	
신뢰도	어떤 추정이 적중할 확률을 의미한다.	공 1 개를 추출할 때 흰 공이 확률은 95%이다.

모집단에서의 표본 추출 예시

다음 그림은 모집단과 표본 그리고 표본 추출 과정을 도식화한 그림이다.



모수와 통계량 표기 방법 (카페 1223)

모수란, 모집단의 특성을 나타내는 값을 모수(parameters)라고 한다.
선거에서 "김말똥" 후보에 대한 지지율을 알고자 하는 경우 모집단은 대한 민국 유권자 전체가 된다.
이때 "김말똥" 후보에 대한 지지율이 모수가 된다.

모수 예시:

"김말똥" 후보에 대한 전체 유권자의 지지율 : P

20대 성인 남자 키의 평균 μ 와 표준 편차 σ

통계량은 표본으로부터 관찰되는 표본의 특성을 통계량이라고 한다.
통계량의 종류에는 평균, 표준 편차, 중앙값 등이 있다.

집단의 특성을 나타내는 모수와 표본의 특성을 나타내는 통계량의 표기 방법은 다음과 같다.

구분	모수(모집단)	통계량(표본)
의미	모집단의 특성을 나타내는 수치	표본의 특성을 나타내는 수치
표기	그리스, 로마자	영문 알파벳
평균	(μ)모 평균	(x)표본의 평균
표준 편차	(σ)모 표준 편차	(S)표본의 표준 편차
분산	(σ^2)모 분산	(S^2)표본의 분산
대상 수	(N)사례수	(n)표본 수

표본 평균의 평균과 분산

모집단의 분포에서 확률 변수 X의 평균, 분산, 표준 편차를 각각 모평균, 모분산, 모표준편차라고 한다.
또, 어떤 모집단에서 크기 n인 표본을 추출하는 경우, 추출된 n개의 변량을 각각 $X_1, X_2, X_3, ..X_n$이라고 할 때 이들의 평균

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + X_3, + \cdots + X_n)$$

을 표본 평균이라고 한다.
이 때 $\overline{X}$는 확률 변수이다.

평균 m, 모분산 σ^2인 모집단에서 크기 n인 표본을 복원 추출할 때, 표본 평균 $\overline{X}$에 대하여 다음의 성질이 알려져 있다.

표본 평균의 평균과 분산 :

$$E(\overline{X}) = E(X) = m \qquad V(\overline{X}) = \frac{\sigma^2}{n} \qquad \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

즉, 표본의 평균은 모평균과 동일하다.
표본의 분산은 (모분산 나누기 표본의 크기)이다.

문제 : 표본 평균의 평균과 분산 구하기

문제 풀기 :
모평균이 5, 모 표준 편차가 0.4 인 모집단에서 크기 100 의 임의의 표본을 복원 추출하는 경우, 표본 평균 X 의 평균, 분산, 표준 편차를 각각 구하여라. 평균 : 5, 분산 : 0.0016, 표준 편차 : 0.04 표본 평균의 공식은 **모 평균과 동일**하므로 5 이다. 표본 분산은 **(모표준편차)^2 / 표본수** = (0.4)^2 / 100 = 0.0016 이다. 표본 표준 편차는 **루트(표본 분산)**이므로 0.04이다.

표본 평균의 분포

평균 m, 모분산 σ^2인 모집단에서 크기 n인 표본을 임의로 추출할 때,
1) 모집단의 분포가 정규 분포이면 표본 평균 X는 정규 분포 N(m, σ^2/n)을 따른다.
2) 모집단의 분포가 정규 분포가 아닐 때에도 n이 충분히 크면, 표본 평균 X는 정규 분포 N(m, σ^2/n)을 따른다.

문제 : 표본 평균의 분포

문제 풀기 :
정규 분포 N(60, 10^2)을 따르는 모집단에서 크기 20 인 표본을 임의 추출할 때, 표본 평균 X 가 이루는 분포는? 1) B(60, 10)　2) B(20, 10)　3) N(60, 10^2/20)　4) N(60, 5^2) 정답) 3)번

통계적 추정(카페 1223)

표본을 이용하여 모수(모집단)들의 값을 추측해나가는 방법을 **통계적 추정**이라고 한다.
통계적 추정 방법은 다음과 같은 항목이 있다.

구분	점 추정	구간 추정
방식	모집단의 특성을 **하나의 값으로 추정**하는 방식	모집단의 특성을 적절한 **구간을 이용하여 추정**하는 방식
특징	모수와 동일할 가능성이 가장 높은 값을 선택한다. 가능성은 희박하다.	모수가 속하는 일정 구간(하한값, 상한값)으로 추정한다. **일반적으로 많이 사용**하는 기법이다.

구간 추정 주요 용어(카페 1223)

구간 추정에 사용되는 주요한 용어는 다음과 같다.

용어	설명
신뢰 수준	Confidence Level 계산된 구간이 모수를 포함할 확률을 의미한다. 통상적으로 90%, 95%, 99%등으로 표현한다.
신뢰 구간	Confidence Interval 신뢰 수준하에서 모수를 포함하는 구간을 말한다. (하한값, 상한값)의 형식으로 표현한다.
표본 오차	Sampling Error 모집단에서 추출한 표본이 모집단의 특성과 정확히 일치하지 않아서 발생되는 확률의 차이를 말한다.

일반적으로 표본의 오차와 표본의 크기는 반비례한다.
표본을 많이 수집할 수록 오차는 줄어든다.

구간 추정 예시 :

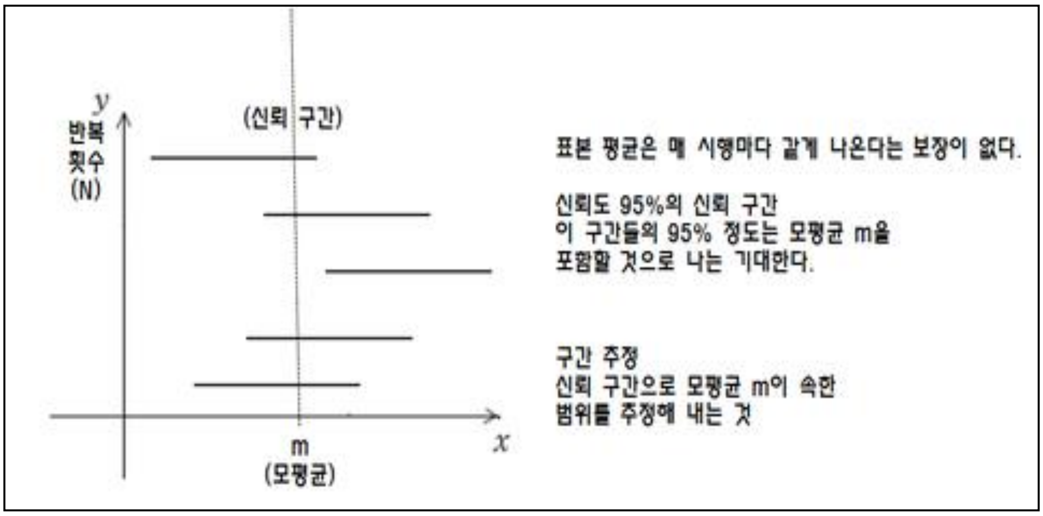
어떤 후보의 지지율 여론 조사에서 95% 신뢰 수준에서 표본 오차 ±3% 범위에서 32.5%로 조사되었다고 가정하자.
실제 지지율은 29.% ~ 35.4%(-3% ~ +3%) 사이에 나타날 수 있다는 의미이다.
이 얘기는 95% 정도의 신빙성이 있다는 의미이기도 하다.
→ 신뢰 수준 : 95%, 표본 오차 : ±3%, 신뢰 구간 : 29.% ~ 35.4%

구간 추정 주요 예시(카페 1223)

SS 고교 학생 25명의 수학 시험의 평균 점수가 76이고, 표준 편차가 10점이라고 하자.
누군가가 전체 학생의 수학 성적을 말할 때 다음과 같다고 가정하자.

-	평균 점수가	추정이 옳을 확률
①	73 점에서 79 점 사이	95%
②	70 점에서 82 점 사이	99%
③	0 점에서 100 점 사이	100%

이때 ①의 구간 [73, 79], ②의 구간 [70, 82]를 각각 95%, 99%의 신뢰 구간이라고 한다.
즉, 이 구간 내에 실제 모수(여기서는 모평균이 된다.)가 존재할 것으로 예측이 되는 구간을 말한다.



가설(hypothesis)

누군가에 의하여 내세우는 주장을 말한다.
실증적인 증명에 앞서 세우는 잠정적인 진술을 말한다.

가설 예시
한국 사람의 평균 수명은 82.4세이다. Caffe Bene의 커피 가격은 스타 벅스의 그것과 동일하다.

가설 검정(hypothesis testing) (카페 1124)

가설 검정이란 유의 수준과 표본의 검정 통계량을 이용하여 모수(모평균, 모분산 등등)에 대한 주장(통계적 가설)의 진위를 검정하는 과정을 말한다.
주장에 대한 검정은 반대되는 주장을 설정하여 어느 주장이 참인지 결정한다.

가설 검정의 특징
대립 가설과 귀무 가설은 상호 배타적이다. 귀무 가설을 참이라고 가정하고 시작한다.

가설의 유형

가설은 귀무 가설과 대립 가설로 나누어 진다.

유형	설명

귀무 가설 (H_0_영 가설)	**거짓이 규명이 될 때 까지 참인 것으로 인정이 되는 가설**이다. 타당성을 입증해야 하는 가설로 **부정적인 형태로 진술**한다. ~같다, ~다르지 않다, ~차이가 없다, ~효과가 없다, ~관련이 없다.
대립 가설 (H_1_연구 가설)	검정하고자 하는 현상에 대한 예측이다. ~다르다, ~차이가 있다, ~효과가 있다, ~관련이 있다. 3가지 형태 양측 검정(two sided test) : 같지 않다. 단측 검정(one sided test) : 크다, 작다

귀무 가설의 예시

두 변수가 서로 독립적이다.(서로 연관성이 없다.)
동전을 던질 때 앞면/뒷면이 나올 확률의 차이가 없다.
특정한 약이 질병에 효과가 없다.
교육 수준과 흡연율 간의 관련성은 없다.
두 가지 교육 방법에 따라 교육생의 만족률에 차이가 없다.
세 가지 교육 방법에 따른 집단 간 만족률에 차이가 없다.

문제 : 귀무 가설과 대립 가설

문제 예시 :

살인자는 재판의 유죄 판결이 있기 전까지는 무죄로 취급 받도록 법적인 보장을 받는다.

검사는 문서, 영상 자료 등 증거를 제시함으로써 유죄의 주장을 뒷받침해야 한다.
이러한 검사의 주장 또는 자료 등은 통계학에서 **대립 가설**이라고 한다.

이에 대하여 불충분한 증거 자료로 유죄를 충분히 입증하지 못하면 원래의 상태로 돌아간다.
즉, 무죄의 상태를 말한다.
이것을 통계학에서는 **귀무 가설**이라고 한다.

무죄의 혐의자가 유죄를 받는 오판이 발생할 수 있다.
귀무 가설이 참일때 이를 기각하는 과오이다.
통계학에서는 이것을 **제1종 오류**라고 한다.

반면 살인자가 무죄 판결을 받고 풀려 나는 경우를 **제 2종 오류**라고 한다.

기각역과 채택역(카페 1124)

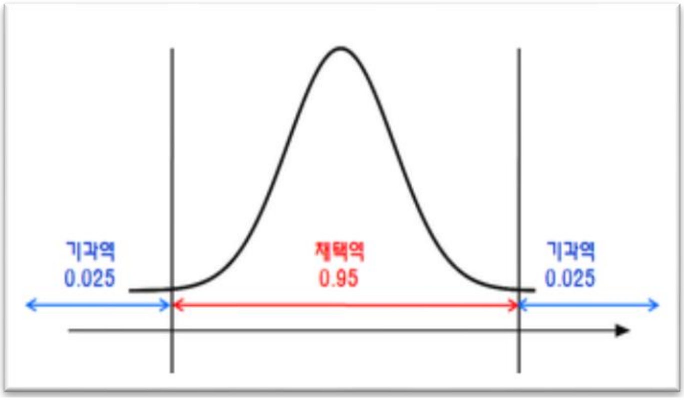
내가 세운 가설(귀무 가설)이 옳음에도 불구하고, **틀렸다고 판단할 수 있는 확률을 유의 수준**이라고 한다.
일반적으로 유의 수준은 5%를 많이 사용한다.

그럼 귀무 가설을 채택할 확률은 유의 수준을 뺀 95%가 된다.(확률이므로)
이 95%를 귀무 가설을 채택하는 영역이므로 채택역이라고 부른다.
이 때 유의 수준 5%의 영역을 기각역이라고 부른다.

기각 또는 채택을 하기 위한 판단 기준을 임계 값이라고 한다.
임계 값을 기준으로 각각 기각역, 채택역이라고 부른다.

항목	설명
채택역	귀무 가설 H_0을 채택하는 검정 통계량 영역을 말한다.
기각역	귀무 가설 H_0을 기각시키는 검정 통계량 영역을 말한다.

즉, 검정 통계량의 양이 기각역에 속하게 되면 귀무 가설을 기각한다는 의미이다.



가설을 검정하기 위해서는 귀무 가설과 연구 가설을 설정하여 통계 분석을 통하여 하나를 채택하게 된다.
이러한 검정 과정에서 발생하게 되는 오류는 제1종 오류와 제2종 오류가 있다.

예를 들어서 무죄라고 생각하는 혐의자가 유죄 판결을 받는 경우가 제 1종 오류에 해당한다.
실제 살인자가 무죄 판결을 받고 풀려 나는 경우가 제 2종 오류에 해당한다.

실제 상황 검정 결과	H_0가 참	거짓
H_0 채택	옳은 결정이다.	제 2 종 오류라고 한다. β (베타)로 표기한다.
H_0 기각	제 1 종 오류라고 한다. α (알파)로 표기한다.	옳은 결정이다.

귀무 가설의 채택역과 오류와의 상관 관계

채택역	기각역	1 종 오류	2 종 오류	
커짐	작아짐	작아짐	커짐	

작아짐	커짐	커짐	작아짐	

양측 검정과 단측 검정

검정 통계량의 분포에서 유의수준 α에 의하여 기각역의 크기가 결정이 된다.
기각역의 위치는 연구 가설(H1)의 형태에 의하여 결정이 된다.
즉 연구 가설의 형태에 따라서 양측 검정과 단측 검정으로 나누어진다.

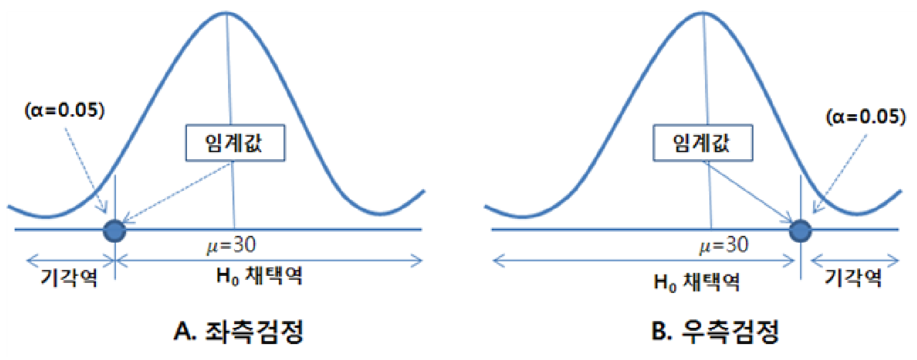
항목	설명
양측 검정	방향성(어느 한쪽이 많고 적음이 없음)이 없는 경우 적용하는 검정 방법이다.
단측 검정	단측 검정(1-sided))은 어느 한쪽이 많거나 적은 가설인 경우에 해당된다.

양측 검정 예시

양측 검정 예시	설명
귀무 가설	성별에 따라서 만족도에 차이가 없다.(만족도가 같다.)
연구 가설	성별에 따라서 만족도에 차이가 있다. (만족도가 같지 않다.)

단측 검정 예시

단측 검정 예시	설명
귀무 가설	1 일 생산되는 불량품의 갯수는 평균 30 개 이다.(μ =30)
연구 가설	1 일 생산되는 불량품의 갯수는 평균 30 개 이하이다.($\mu <$ 30) 1 일 생산되는 불량품의 갯수는 평균 30 개 이상이다.($\mu >=$ 30)



검정 통계량(카페 1124)

가설을 검정하기 위하여 수집된 데이터로부터 계산된 통계량을 말한다.

즉, 가설 검정에서 **기각역을 결정하는 기준이 되는 통계량**을 말한다.

검정 통계량은 분석 방법에 따라서 달라진다.

분석 방법	검정 통계량	분석 방법	검정 통계량
상관 분석	r 값	회귀 분석	t 값
T 검정	t 값	분산 분석	F 값
카이 제곱	X^2 값		

모수와 비모수 검정

모수 검정이란 특정 표본에서 평균, 표준 편차, 분산 등의 지표(parameter)를 뽑아서 특정 집단 사이의 특징을 비교하는 검정 방법이다.
비모수 검정 방법은 평균, 표준 편차 등의 개념이 존재하지 않는다.

검정 방법	설명
모수 검정	관측된 값이 어느 특정한 확률 분포(정규 분포, 이항 분포)를 따른다고 전제한 후 그 분포의 모수에 대한 검정을 실시하는 방법이다.
비모수 검정	관측된 값이 어느 특정한 확률 분포(정규 분포, 이항 분포)를 따른다고 전제를 할 수 없는 경우에 사용한다.

정규성 검정 방법 :
케이스의 수가 너무 적거나 정확한 정규 분포를 검정하기 위해서는 수행하는 검정 방법이다.
정규성 검정을 통하여 정규 분포는 모수 검정을, 그렇지 않으면 비모수 검정을 수행한다.

검정 방법	모수(정규 분포)	비모수(비정규 분포)
t 검정	독립 표본 t 검정	윌콕슨((Wilcoxon) 검정
	대응 표본 t 검정	맨-휘트니(Mann-Whitney) 검정
분산 분석	일원 배치 분산 분석	크루스칼-월리스(Kruskal-Wallis)검정
관계 분석	상관 분석	비모수적 상관 분석

분석 절차

빅 데이터 분석이란 다양한 분석 기법을 적용하여 새로운 통찰이나 새로운 가치를 발견하고 예측하는 일련의 과정을 말한다.

특히 논문 및 보고서 작성을 목적으로 실제 세계의 현상을 관찰하여 수량화하고, 이를 통해서 추론 통계를 바탕으로 예측하는 분석 절차는 데이터의 크기에 상관없이 다음과 같은 일반적인 절차에 의해서 분석을 수행한다.

일반적인 분석 절차 :

1. 가설 설정
2. 유의 수준 결정
3. 측정 도구 선정
4. 데이터 수집(설문지, 웹, SNS 등등)
5. 데이터 코딩/프로그래밍
6. 통계 분석 수행(R, SPSS, SAS)
7. 결과 분석(논문/보고서 작성)

유의 수준(카페 1124)

연구 가설의 채택 또는 기각은 유의 수준(Significant Level)을 기준으로 하여 가설 채택 여부가 결정이 된다.
유의 수준은 α(알파) 또는 P(probability : 확률)로 표시하며, 제 1종의 오류를 범하는 확률이라고 보면 된다.
일반적으로 사회 과학 분야에서는 0.05를 사용한다.
의학, 생명 분야에서는 0.01을 사용한다.

유의 수준 예시) 유의 수준이 0.05 이다.

"표본에 대한 신뢰도가 95%이상이다"라는 의미이다.
즉, 의사 결정을 할 때 우리가 실수할 확률이 5%정도라는 의미이다.

유의 확률과 유의 수준에 따른 기각역 채택 여부

유의 수준	설명
p-value > 0.05	귀무 가설 채택
p-value < 0.05	귀무 가설 기각

통계적으로 유의하다.

유의하다.	유의하지 않다.
우연이 아닌 실제로 의미가 있는 데이터이다.	이 사건은 우연히 발생한 사건이다.

척도(카페 1291)

통계 분석 방법을 결정할 때, 척도에 대한 이해가 필요하다.
척도는 연구 대상을 측정하기 위한 측정 도구로 사용된다.

숫자/기호 등으로 응답자에게 값을 선택할 수 있도록 측정하게 하는 단위이다.

척도의 예시

(설문 조사) 1) 매우 그렇다. 2) 그렇다. 3) 보통이다. 4) 그렇지 않다. 5) 매우 그렇지 않다.

척도는 크게 범주형 척도와 연속형 척도로 구분이 된다.
일반적으로 나누는 기준은 ′수량화 가능 여부′이다.

척도	예시	세부 설명
범주형	성별, 직업.	크기를 비교할 필요가 없다.
연속형	만족도, 키, 몸무게	만족도는 1 점에서 5 점까지의 점수로 평가가 가능하다.

정성적-질적 척도(범주형 변수)		정량적-양적 척도(연속형 변수)	
명목 척도	이름이나 범주를 표시하는 의미 없는 숫자	등간 척도	속성에 대한 각 수준 간의 간격이 동일한 경우
서열 척도	측정 대상의 높고 낮음	비율 척도	등간 척도 절대 원점이 존재한다. 비율 계산이 가능하다.(사칙 연산)

척도의 종류

종류		설명	예시
범주형	명목 척도	단지 구분을 하기 위한 목적으로 숫자를 부여하여 만든 척도이다. 숫자의 양적인 의미는 없다.	성별 : 숫자 1 은 남자, 2 는 여자이다. 직업 : 숫자 1(학생), 2(교수), 3(직원) 학력 : 숫자 1(초졸), 2(중졸), 3(고졸), 4(대졸) 선호하는 색상 1 : 검정, 2 : 흰색, 3 : 쥐색, 4 : 청색, 5 : 녹색)
	서열 척도	순서적 상하 관계를 나타낸다. 대상 간의 크고 작다. 양이 많고 적다. 선호도가 높다 낮다 숫자의 양적인 의미는 없다.	직급 : 1(사원), 2(대리), 3(과장), 4(부장), 5(이사)
연속형	등간 척도	균일한 간격을 두고 분할하여 측정하는 척도를 의미한다.	설문 조사 : 1) 매우 그렇다. 2) 그렇다. 3) 보통이다. 4) 그렇지 않다. 5) 매우 그렇지 않다.
	비율 척도	등간 척도 + 비율 개념이 첨가된 척도이다. 절대적 0 점을 출발점으로 하여 속성을	나이가 얼마입니까? 귀하의 몸무게는 얼마입니까?

		양적으로 표현하는 척도이다.	

척도의 예시 1

변수	resident	gender	job	age	position	price	survey
척도	명목	명목	명목	비율	서열	비율	등간
범위	1~5	1, 2	1~3	20~69	1~5	2.1~7.9	1~5
설명	거주지	성별	직업	나이	직위	구매 금액	만족도

척도의 예시 2

gender	age	group	interest
성별	나이	광고의 유형(종류)	관심도 유무
명목(M, F) 척도	비율 척도	명목(1, 2) 척도	명목(1, 0) 척도

척도에 따른 분석 방법

척도에 따른 분석 방법은 다음과 같은 항목들이 존재한다.
자신이 세운 가설에 해당 요인들이 어떠한 형태의 척도인지 파악하고 그것을 토대로 분석 방법이 결정된다고 이해하면 된다.

기술 통계량(카페 1152)

자료를 요약할 수 있는 기초 통계량을 말한다.
전체적인 데이터의 분포와 통계적 수치 정보들을 제공한다.

특징

- 모집단의 특성을 유추하는 데 사용할 수 있다.

기술 통계량의 유형에는 다음과 같은 항목들이 있다.

유형	설명/예시
대표값	자료 전체를 대표하는 값이다. 평균(mean), 합계(sum), 중위수(median), 최빈수(mode), 사분위수(quartile)
산포도	자료들이 대표 값으로부터 얼마나 흩어져 분포하고 있는 가를 보여 주는 값 분산, 표준 편차, 최소값, 최대값, 범위(range)
비대칭도	첨도(kurtosis) : 분포도가 얼마나 중심에 집중되어 있는가? 왜도(skweness) : 분포가 기울어진 방향과 정도

빈도 분석

- 범주형 데이터에 대한 비율을 알아 보고자 하는 경우에 사용한다.
- table() 함수를 이용한다.

빈도 분석 사용 예시

- 특정 후보의 지지율은 50%이다.
- 응답자 중에서 남자는 30%, 여자는 70%이다.
- 연령대 별로 차지하는 비율 구하기

통계 분석 수행(카페 1304)

전문 통계 분석 프로그램(R, SAS)등을 이용하여 분석을 수행하는 단계이다.
변수의 척도에 따라서 통계 분석 방법이 결정되기 때문에 변수의 척도 선정과 모델링이 무엇보다도 중요하다.

영향을 주는 변수	영향을 받는 변수	분석 방법
범주형 자료	범주형 자료	카이 제곱 검정
범주형 자료	연속형 자료	T 검정, 분산 분석(Anova)
연속형 자료	범주형 자료	로지스틱 회귀 분석
연속형 자료	연속형 자료	회귀 분석, 구조 방정식(고차 방정식)

통계 분석 방법과 변수 척도와의 관계

파일 이름	설명	척도
빈도 분석	가장 기초적인 분석 방법이다. 변수의 분포를 제공하며 인구 통계적 특성을 제시하는 데 용이하다.	모든 척도
교차 분석 (카이 제곱)	변수간의 분포와 백분율을 나타내주는 교차표를 작성한다. 두 변수의 독립성과 관련성(카이 제곱 검정)을 분석한다.	명목, 서열 척도
요인 분석	측정하고자 하는 변수들의 상관 관계가 높은 것끼리 묶어서 변수를 단순화 시키는 데 이용된다.(타당성 검정) 잘못 적재된 변수나 설명력이 부족한 변수를 제거한다.	등간, 비율 척도
신뢰도 분석	요인 분석으로 추출된 요인들이 동질적인 변수들로 구성이 되었는 가를 파악하는 분석 기법이다.	등간, 비율 척도
상관 관계 분석	설정한 가설을 검정하기 전에 모든 연구 가설에 사용되는 측정 변수들간의 관계 정도를 제시하며 변수들간의 관련성에 대한 윤곽을 제시한다.	피어슨 : 등간, 비율 척도 스피어만 : 서열 척도
회귀 분석	독립 변구가 종속 변수에 어떠한 영향을 미치는 지를 파악하기 위해 실시하는 분석 방법이다. 두 변수간의 인과 관계를 분석하는 방법이다.	등간, 비율 척도
t 검정	종속 변수에 대한 독립 변수의 집단간 평균의 차이를 검정한다. 독립 표본 t-test와 대응 표본 t-test 분류	독립 변수 : 명목 척도 종속 변수 : 등간, 비율 척도

분산 분석 (anova)	t 검정과 같이 집단간 평균의 차이를 구하는 분석 기법으로 다른 점은 3 집단 이상의 평균을 검정할 때 이용된다.	독립 변수 : 명목 척도 종속 변수 : 등간, 비율 척도