확률과 통계

이변량 확률변수

1. 이변량 확률변수

이변량 확률변수의 정의

> *두 개의 확률변수를 동시에 고려할 때*
> *(X, Y)가 어떤 특성을 갖는지 알아보는 것*

- 확률변수 X와 Y가 모두

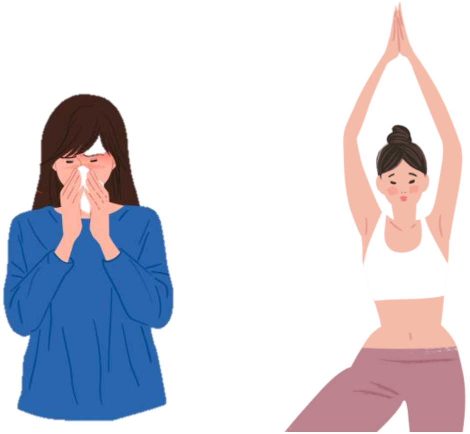
이산형 확률변수 **연속형 확률변수** 로 구분

1. 이변량 확률변수

이변량 확률변수의 구분

확률변수 X와 Y가 모두 **이산형** 확률변수인 경우

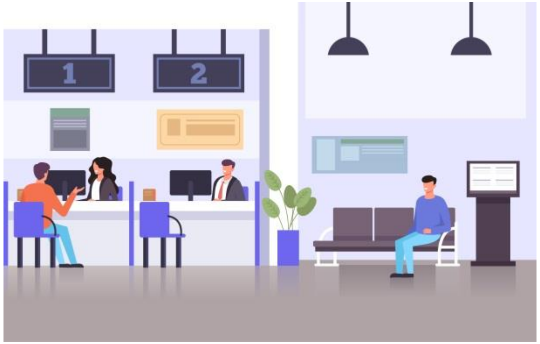
예



X : 감염자 수

Y : 완치된 환자수

예



X : 창구1에 온 고객 수

Y : 창구2에 온 고객 수

1. 이변량 확률변수

이변량 확률변수의 구분

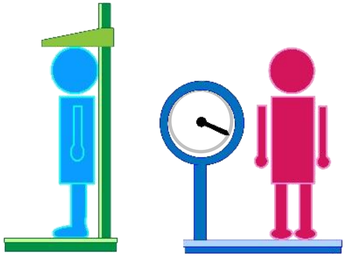
확률변수 X와 Y가 모두 **연속형** 확률변수인 경우

예

X : 광고비

Y : 매출액

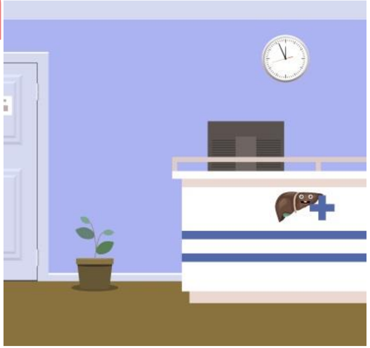
예



X : 키

Y : 몸무게

예



X : 진료 대기 시간

Y : 진료 시간

2. 결합확률함수

결합확률함수(Joint Probability Function)

확률변수 X와 Y가
모두 이산형 확률변수인 경우

$$f(x, y) = P(X = x, Y = y)$$

2. 결합확률함수

결합확률함수의 특징

$$0 \leq f(x, y) \leq 1$$

for all x, y

$$\sum_{all x} \sum_{all y} f(x, y) = 1$$

2. 결합확률함수

주변확률함수

때로는 X만의 움직임 또는 Y만의 움직임 특성을 살펴볼 필요가 있음

X의 주변확률함수

$$g(x) = \sum_{all y} f(x, y)$$

Y의 주변확률함수

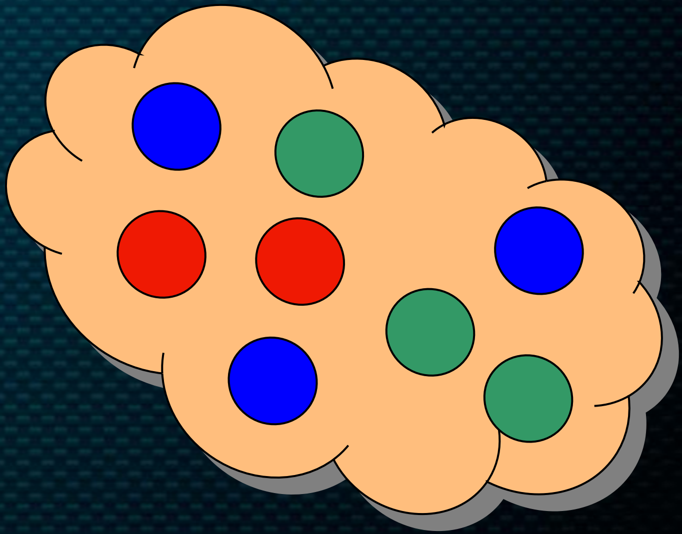
$$h(y) = \sum_{all x} f(x, y)$$

예제 풀어보기

Q1.

8개의 공(빨간색, 파란색, 초록색)이 들어있는 주머니에서
2개의 공을 무작위로 선택한다.

X = 파란색 공의 수
Y = 빨간색 공의 수



3. 주변확률함수

X의 주변확률함수

$$g(x) = \sum_{all\,y} f(x, y)$$

Y의 주변확률함수

$$h(y) = \sum_{all\,x} f(x, y)$$

예제 풀어보기

Q2. 확률변수 (X, Y)의 결합확률함수는 다음과 같다.

$$f(x, y) = \frac{x+y}{30}$$

$x = 0, 1, 2, 3$ $\quad$ $y = 0, 1, 2$

(1) X의 주변확률함수 $g(x)$를 구하여라.
(2) Y의 주변확률함수 $h(y)$를 구하여라.

4. 결합밀도함수

결합밀도함수의 특징

"확률변수 X와 Y가
모두 **연속형 확률변수**인 경우"

- $0 \leq f(x, y)$ for all x, y
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$
- $P(a < X < b, c < Y < d) = \int_{c}^{d}\int_{a}^{b} f(x, y)dxdy$

4. 결합밀도함수

주변밀도함수

X의 주변밀도함수

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

Y의 주변밀도함수

$$h(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

예제 풀어보기

Q3. 결합밀도함수는 다음과 같다.

$$f(x,y) = \begin{cases} \frac{2}{5} \cdot (2x+3y) & (0 \le x \le 1, 0 \le y \le 1) \\ 0 & \text{elsewhere} \end{cases}$$

다음의 확률을 구하라.

$$P(0 < x < \frac{1}{2}, \quad \frac{1}{4} < y < \frac{1}{2})$$

5. 주변밀도함수

X의 주변밀도함수

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

Y의 주변밀도함수

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

예제 풀어보기

Q4. 결합밀도함수는 다음과 같다.

$$f(x,y) = \begin{cases} 8xy & 0 \le x \le 1, \quad 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

X와 Y의 주변밀도함수를 구하라.

6. 공분산

공분산 Covariance

- 확률변수가 2개 이상일 때 존재

> “ *확률변수 X와 Y가 평균 E(X)와 E(Y)로 부터 어떻게 흩어져 있는지를 나타내는 척도* ”

6. 공분산

공분산 Covariance

X의 평균 : 상수   Y의 평균 : 상수

$$Cov(X,Y) = E(X - \boxed{\mu_X})(Y - \boxed{\mu_Y})$$
$$= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y) = EXY - \mu_X\mu_Y$$

예제 풀어보기

Q5. 결합밀도함수는 다음과 같다.

$$f(x,y) = \begin{cases} 8xy & 0 \leq x \leq 1, \quad 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$$

$Cov(X,Y)$를 구하라.

6. 공분산

공분산의 특징

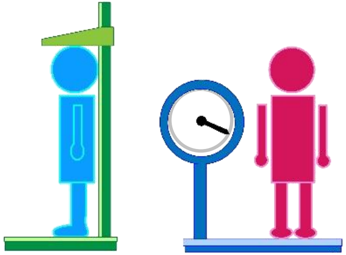
단위에 민감

- 예를 들어, a, b 는 상수라면

$$
\begin{aligned}
Cov(aX, bY) &= E(aX - a\mu_X)(bY - b\mu_y) \\
&= abE(X - \mu_X)(Y - \mu_Y) \\
&= ab\,Cov(X, Y)
\end{aligned}
$$

공분산의 특징

예

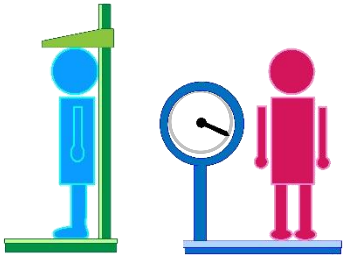


	기존 단위	변환
X : 키	cm	m
Y : 몸무게	kg	g
	공분산 Cov(X, Y)	공분산 Cov(aX, bY)

6. 공분산

공분산의 특징

예



공분산 Cov(aX, bY)

- a는 100분의 1(cm → m로 변환)
- b는 1000 (kg → g로 변환)

10배

단위가 바뀌면 공분산의 값이 바뀜

공분산만으로 두 확률변수의 상대적 흩어짐을 나타내기에 부적절

6. 공분산

공분산의 특징

$$Cov(aX, aX) = a^2 Cov(X, X) = a^2 VarX$$

$$E(X - \mu_X)(X - \mu_X) = VarX$$

6. 공분산

공분산의 특징

$$
\begin{aligned}
&Var(X+Y) \\
&= E((X+Y)-(\mu_X+\mu_Y))^2 \\
&= E\left((X+Y)^2 - 2(X+Y)(\mu_X+\mu_Y) + (\mu_X+\mu_Y)^2\right) \\
&= E((X^2+Y^2+2XY) - 2(\mu_X+\mu_Y)E(X+Y) + ({\mu_X}^2 + 2\mu_X\mu_Y + {\mu_Y}^2) \\
&= EX^2 + EY^2 + 2EXY - 2({\mu_X}^2 + 2\mu_X\mu_Y + {\mu_Y}^2) + ({\mu_X}^2 + 2\mu_X\mu_Y + {\mu_Y}^2) \\
&= EX^2 - {\mu_X}^2 + EY^2 - {\mu_Y}^2 + 2(EXY - \mu_X\mu_Y) \\
&= VarX + VarY + 2Cov(X,Y)
\end{aligned}
$$

$$Var(X+Y) = Var\,X + Var\,Y + 2Cov(X,Y)$$
$$Var(X-Y) = Var\,X + Var\,Y - 2Cov(X,Y)$$

7. 확률변수의 독립성

확률변수 X, Y의 독립

- 사상 A, B가 독립이면,

$$P(AB) = P(A)P(B)$$

조건부확률 계산이 필요 없음

7. 확률변수의 독립성

확률변수 X, Y의 독립

- 확률변수 X와 Y가 독립이면,

$$f(x, y) = g(x)h(y)$$

- 확률변수 X와 Y가 독립인지 확인하는 방법

결합확률함수 또는 결합밀도함수가 각각 주변확률함수 또는 주변밀도함수의 곱인지 아닌지 확인

7. 확률변수의 독립성

확률변수들의 상호독립

- $X_1, X_2, \dots, X_n$가 상호독립이면,

$$f(x_1, x_2, \dots, x_n) = \underbrace{f(x_1)}_{}f(x_2) \dots f(x_n)$$

$f(x_1)$는 확률변수 X_1의 확률함수 또는 밀도함수

7. 확률변수의 독립성

확률변수 X, Y 독립의 특징

- 확률변수 X와 Y가 독립이면, $\boldsymbol{f(x,y) = g(x)h(y)}$
- EXY = $\mu_X\mu_Y$

$$EXY = \int\int xyg(x)h(y)dxdy = \int xg(x)dx \int yh(y)dy$$
$$= EXEY = \mu_X\mu_Y$$

- 확률변수 X와 Y가 독립이면,

$$\boldsymbol{Cov(X,Y) = 0}$$

7. 확률변수의 독립성

확률변수 X, Y 독립의 특징

Q. 과연 $Cov(X, Y)$가 0이라면, X와 Y가 독립일까요?

일방향

- 확률변수 X와 Y가 독립이면, 반드시 $Cov(X, Y) = 0$
- 그러나 $Cov(X, Y) = 0$ 는 X와 Y의 독립성을 보장하지 않음

7. 확률변수의 독립성

확률변수 X, Y 독립의 특징

Q. $Cov(X,Y)$가 0일 때, X와 Y의 독립 판별 방법은?

$f(x,y) = g(x)h(y)$ *인지 확인*

모든 x, y에 대해서 확인해야함

7. 확률변수의 독립성

확률변수 X, Y 독립의 특징

- 확률변수 X와 Y가 독립이면,

$$Var(X \pm Y) = Var\,X + Var\,Y$$

Why?

$Cov(X, Y) = 0$이므로,

$\pm 2Cov(X, Y)$가 사라지기 때문!

8. 확률변수의 선형결합

- 확률변수 X와 Y의 선형결합 $U = aX + bY$

$$E(U) = aE(X) + bE(Y)$$
$$Var(U) = a^2Var(X) + b^2Var(Y) + 2ab\,Cov(X,Y)$$

확률변수 X와 Y가 독립이면,
$Cov(X,Y) = 0$

$$Var(U) = a^2Var(X) + b^2Var(Y)$$

8. 확률변수의 선형결합

- 만약 $X = X_1 + X_2 + \cdots + X_n$이며, $X_1, X_2, \ldots, X_n$ 독립이라면

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

$$Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$

$$Cov(X_i, X_j) = 0$$

9. 상관계수

상관계수(correlation coefficient)

확률변수 X의 증감에 따른 확률변수 Y의 증감 정도를 나타내는 척도

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var\ X}\sqrt{Var\ Y}}$$

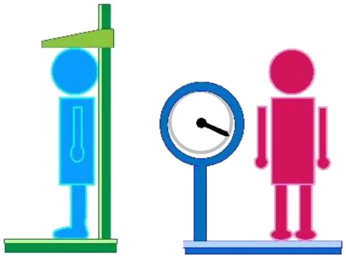
- $-1 \le \rho_{X,Y} \le 1$이며, $\rho_{X,Y}$는 단위가 없는 값

9. 상관계수

상관계수(correlation coefficient)

예

키와 몸무게 간 상관계수를 구한다고 가정하면



X : 키 (cm)　Y : 몸무게(kg)

$Cov(X,Y)$ 단위가 살아있음

$Var\,X,\ Var\,Y$ 단위가 살아있음

→단위들이 서로 상쇄됨

$-1 < \rho_{X,Y} < 1$, 단위는 없음

! 단위에 민감한 Cov에 대한 해결책 : 상관계수($\rho_{X,Y}$)

9. 상관계수

상관계수(correlation coefficient)

Q. 왜 $-1 < \rho_{X,Y} < 1$의 값을 가질까?

[증명]

확률변수 X의 분산은 σ_x^2	확률변수 Y의 분산은 σ_y^2

1

$$0 \le Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \frac{1}{\sigma_X^2} Var\ X + \frac{1}{\sigma_Y^2} Var\ Y + 2\frac{1}{\sigma_X}\frac{1}{\sigma_Y} Cov(X,Y)$$

$$\therefore \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \ge -1$$

9. 상관계수

상관계수(correlation coefficient)

[증명]

확률변수 X의 분산은 σ_x^2

확률변수 Y의 분산은 σ_y^2

2

$$0 \le Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = \frac{1}{\sigma_X^2} Var\,X + \frac{1}{\sigma_Y^2} Var\,Y - 2\frac{1}{\sigma_X}\frac{1}{\sigma_Y} Cov(X,Y)$$

$$\therefore \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \le 1$$

[결론]

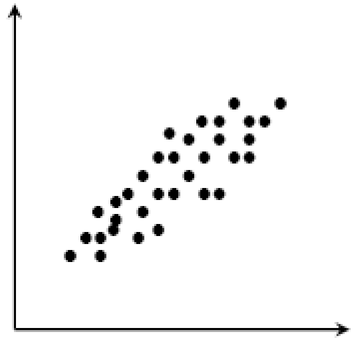
$$-1 \le \rho_{X,Y} \le 1$$

9. 상관계수

상관계수 $\rho_{X,Y}$ 값의 의미

정비례

$0 < \rho_{X,Y} < 1$

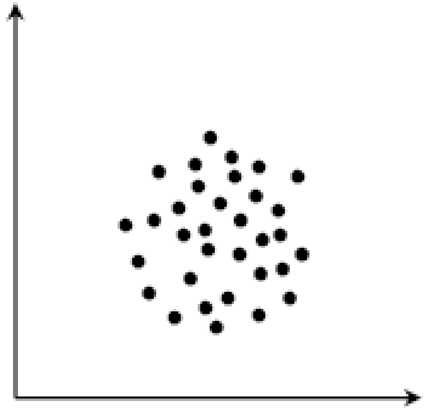


9. 상관계수

상관계수 $\rho_{X,Y}$ 값의 의미

무상관

$\rho_{X,Y} = 0$



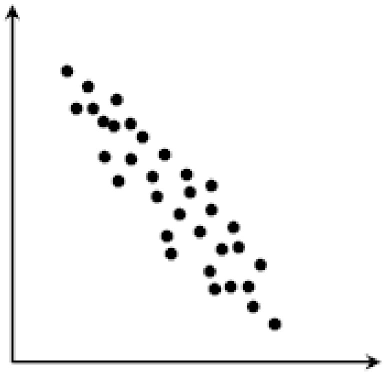
서로 정보에 대해서 아무런 공유가 없다는 의미

9. 상관계수

상관계수 $\rho_{X,Y}$ 값의 의미

반비례

$$-1 < \rho_{X,Y} < 0$$



9. 상관계수

상관계수 $\rho_{X,Y}$ 값의 의미

$\rho_{X,Y} = 1$

완벽한
정비례

$\rho_{X,Y} = -1$

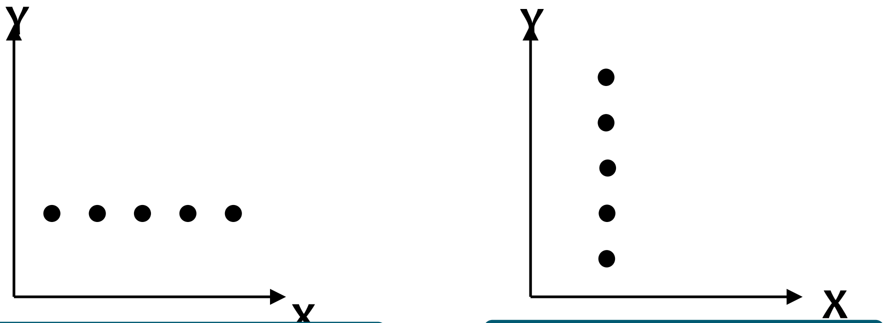
완벽한
반비례

X와 Y가 100% 공유

9. 상관계수

상관계수 $\rho_{X,Y}$ 값의 의미

undefined



$Var\,\mathbf{Y} = 0$

$Var\,\mathbf{X} = 0$

상관계수가 정의되지 않음

예제 풀어보기

Q6. 확률변수 (X, Y)의 결합밀도함수는 다음과 같다.

$$f(x, y) = 6x^2y \text{ for } 0 < x < 1, 0 < y < 1$$

(1) E(X)와 E(Y)을 각각 구하여라.
(2) 상관계수 $\rho_{X,Y}$를 구하여라.

예제 풀어보기

Q7.

$$f(x,y) = e^{-x-y} \qquad x > 0, y > 0$$

$$= 0 \qquad \text{otherwise}$$

확률변수 X와 Y는 서로 독립인가?

예제 풀어보기

Q8. $Cov(X, X+Y) = Var(X) + Cov(X, Y)$임을 증명하라.

예제 풀어보기

Q9. 확률변수 (X, Y)의 결합 확률함수는 다음과 같다.

y\x	1	2	3	$h(y)$
1	1/8	1/8	1/8	3/8
2	1/8	0	1/8	2/8
3	1/8	1/8	1/8	3/8
$g(x)$	3/8	2/8	3/8	1

(1) $Cov(X,Y)$를 구하라
(2) $Corr(X,Y)$를 구하라
(3) 확률변수 X와 Y는 서로 독립인가?

예제 풀어보기

Q9. 확률변수 (X, Y)의 결합 확률함수는 다음과 같다.

y\x	1	2	3	$h(y)$
1	1/8	1/8	1/8	3/8
2	1/8	0	1/8	2/8
3	1/8	1/8	1/8	3/8
$g(x)$	3/8	2/8	3/8	1

(1) $Cov(X,Y)$를 구하라
(2) $Corr(X,Y)$를 구하라
(3) 확률변수 X와 Y는 서로 독립인가?