두 개의 확률변수에 관한 확률분포

- 예제

  Xavier 와 Yvette 은 어느 지역에 있는 두 명의 부동산중개업자이다. Xavier가 한달 동안 판매한 집의 개수를 X, Yvette 이 한달 동안 판매한 집의 개수를 Y라고 할 때, 과거 경험에 의하면 X와 Y는 다음과 같은 결합확률을 가진다고 하자.

		X			
		0	1	2	
Y	0	.12	.42	.06	0.6
	1	.21	.06	.03	0.3
	2	.07	.02	.01	0.1
		0.4	0.5	0.1	

X	$f_X(x)$
0	0.4
1	0.5
2	0.1

◆ Xavier가 한달 동안 판매한 집의 개수에 관한 주변확률분포를 표로 정리하여라.

X

◆ Yvette이 한달 동안 판매한 집의 개수에 관한 주변확률분포를 표로 정리하여라.

Y

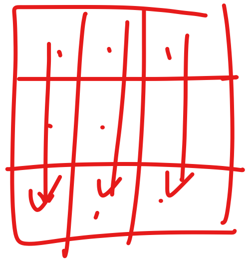
y	$f_Y(y)$
0	0.6
1	0.3
2	0.1

두 개의 확률변수에 관한 확률분포

Joint

```
> Pxy <- matrix( c( 0.12, 0.21, 0.07, 0.42, 0.06, 0.02, 0.06, 0.03, 0.01),
+                 nrow=3, ncol=3 )
> rownames( Pxy ) <- c(0, 1, 2)
> colnames( Pxy ) <- c(0, 1, 2)
> Pxy
     0    1    2
0 0.12 0.42 0.06
1 0.21 0.06 0.03
2 0.07 0.02 0.01
> Px <- apply( Pxy, 2, sum )
> Px
  0   1   2
0.4 0.5 0.1
> Py <- apply( Pxy, 1, sum )
> Py
  0   1   2
0.6 0.3 0.1
```

Marginal



apply(M , 1 , sum)

2

두 개의 확률변수에 관한 특성값

- **두 개의 확률변수에 관한 특성값**
  - **기대값, 분산, 표준편차**

◆ 확률변수 X와 Y의 기대값
- $\mu_X (= E[X])$: X의 기대값
- $\mu_Y (= E[Y])$: Y의 기대값

μ_X

◆ 확률변수 X와 Y의 분산
- σ_X^2 $(= V[X])$: X의 분산
- $\sigma_Y^2 (= V[Y])$: Y의 분산

σ_Y^2

◆ 확률변수 X와 Y의 표준편차
- $\sigma_X (= S[X])$: X의 표준편차
- $\sigma_Y (= S[Y])$: Y의 표준편차

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

$f(x)$

$X \rightarrow$ μ_X σ_X^2 σ_X

$Y \rightarrow$ μ_Y σ_Y^2 σ_Y

$f(y)$

X, Y f(x, y) → (x_i, y_i)

두 개의 확률변수에 관한 특성값

공분산 → 상관계수

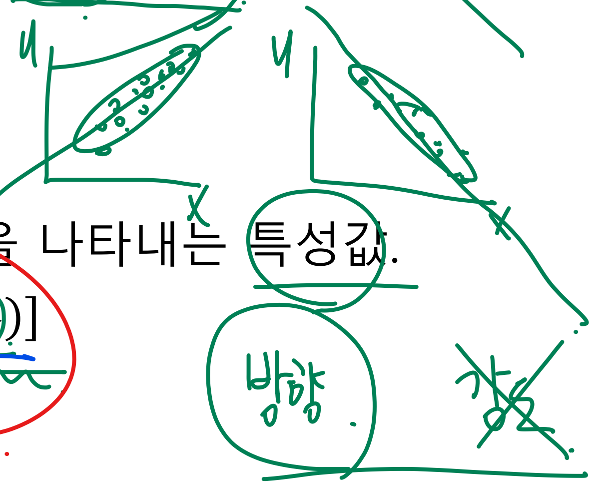
- 두 개의 확률변수에 관한 특성값
  - 공분산(covariance)

    두 개의 확률변수 X, Y에 관한 선형적 연관성을 나타내는 특성값.

    선형적 연관성

$$COV[X,Y] = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$

Cov.



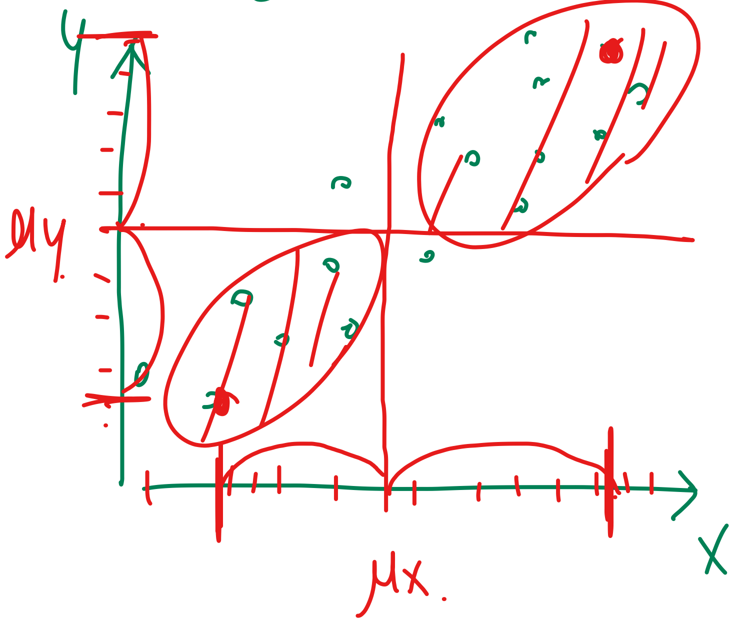
방향 ~~강도~~

- 공분산의 성질
  - $COV[X,Y] = \sigma_{XY} = E[XY] - \mu_X\mu_Y$
  - $COV[aX,bY] = ab \cdot COV[X,Y]$

$X, Y \sim f(x,y)$

(x_i, y_i) $i = 1 \ldots n.$

⊕의 선형.



$cov[X,Y] > 0$

$= E[(X-\mu_X)(Y-\mu_Y)]$

⊕ ⊕

⊖ ⊖

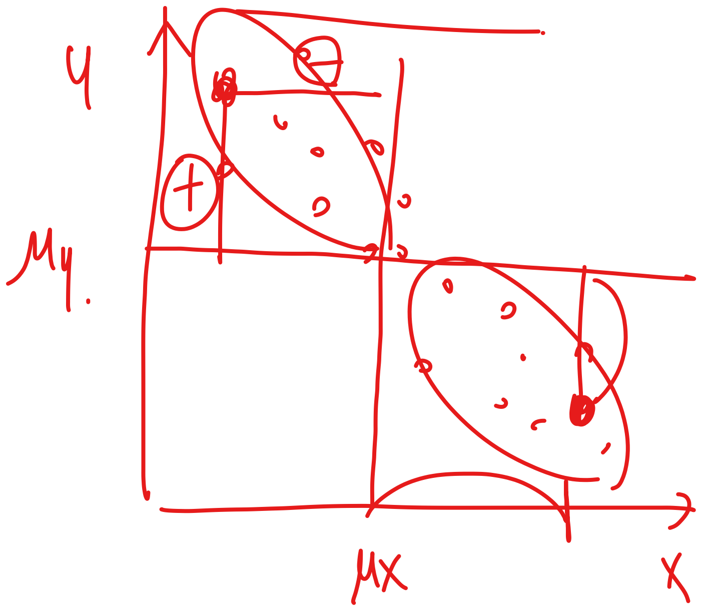
⊕

⊕

$cov > 0$

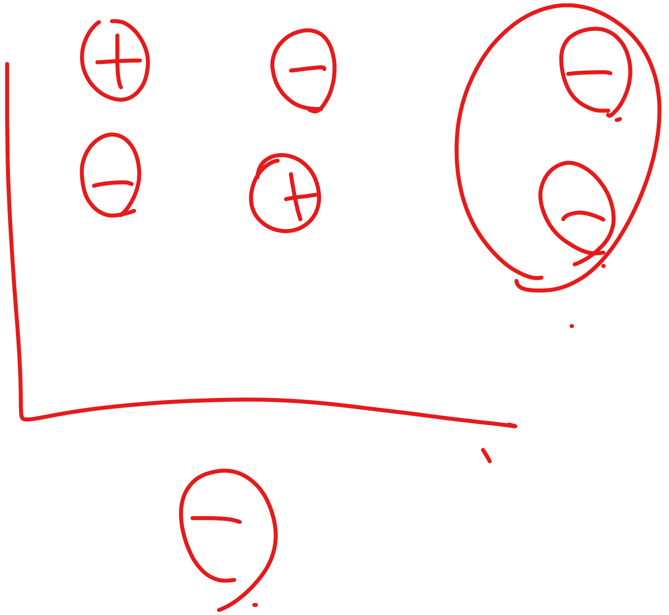
⊕

⊖ 의 선형



$Cov[X,Y] < 0.$

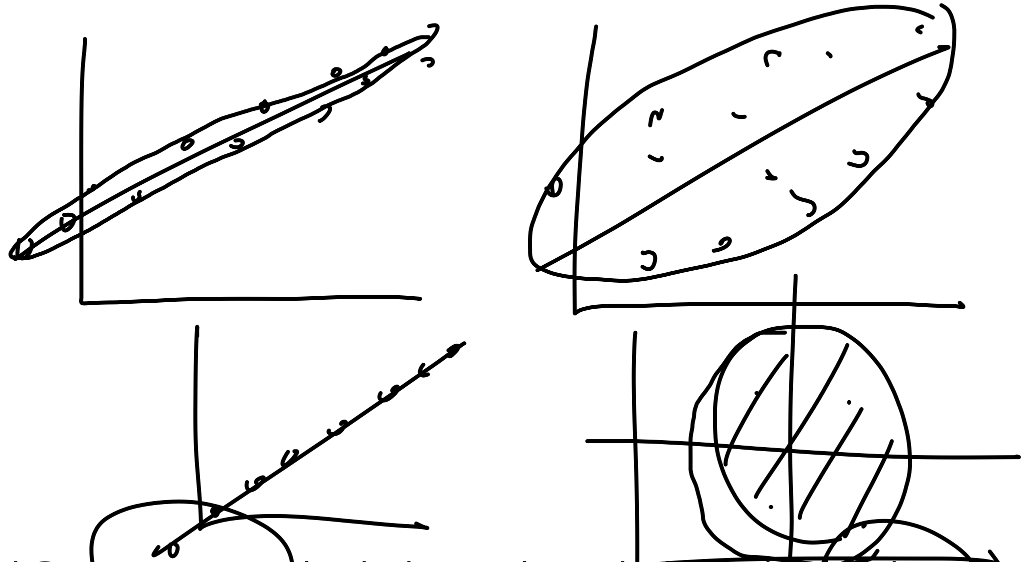
$E[(X-\mu_x)(Y-\mu_y)]$



두 개의 확률변수에 관한 특성값

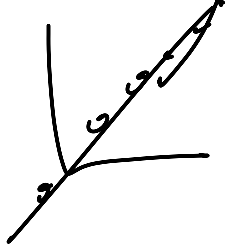
- **공분산의 성질**

  1) 선형관계의 **방향**을 파악할 수 있음.
     - $\sigma_{XY} > 0$: X, Y 는 양의 선형 관계를 가짐
     - $\sigma_{XY} < 0$: X, Y 는 음의 선형 관계를 가짐



  2) 선형관계의 **강도**는 공분산 만으로 파악할 수 없음. $COV[X,Y]$가 가질 수 있는 값의 범위는 X와 Y의 스케일에 의존하기 때문.
     - $-\sigma_X\sigma_Y \le \sigma_{XY} \le \sigma_X\sigma_Y$, by Cauchy−Schwarz inequality

$\sigma_{xy} = 30$?

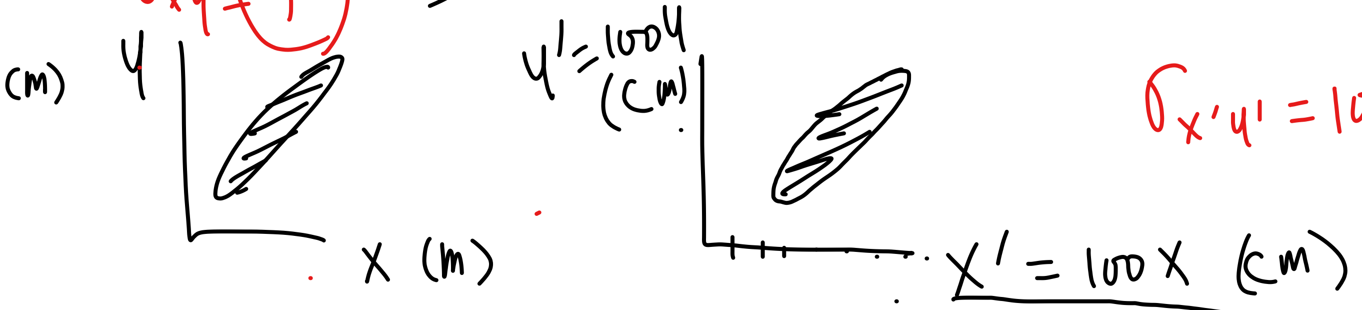


$\sigma_x\sigma_y = 30$

5 6

  3) 자료의 단위를 변경하면 공분산 값도 변함.
     - $X' = aX + b$ 이고, $Y' = cY + d$ 일 때, $\sigma_{X'Y'} = ac \cdot \sigma_{XY}$

$\sigma_{xy} = 5$

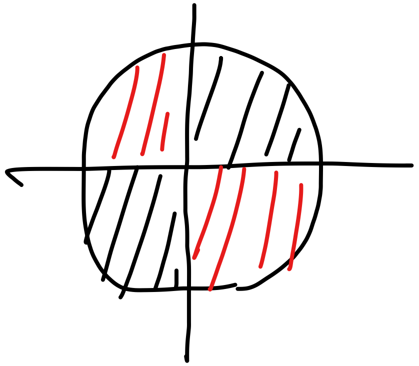


$\sigma_{x'y'} = 100 \times 100\ \sigma_{xy}$

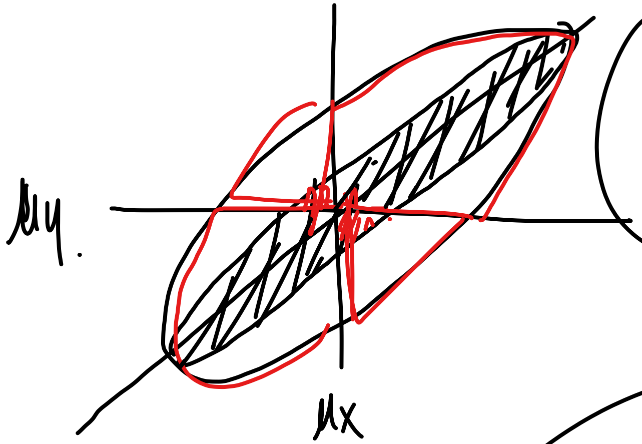
$= 50000.$

$\sigma_x\sigma_y = 300$

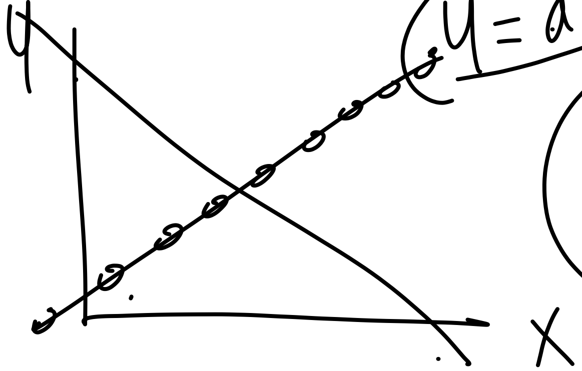
10 30



$|\sigma_{xy}| = 0$.



$|\sigma_{xy}| \uparrow$



$y = a + bx$

$|\sigma_{xy}| = \sigma_x \sigma_y$

$-\sigma_x\sigma_y \leq \sigma_{xy} \leq \sigma_x \sigma_y$

by C-S.

$y = a + bx$.

두 개의 확률변수에 관한 특성값

- **상관계수(correlation coefficient)** :
  두 개의 확률변수 X, Y의 선형적 연관성의 방향과 강도를 동시에 파악할 뿐 아니라 자료의 단위에도 의존하지 않도록 공분산을 보정한 지표.

$$CORR[X,Y] = \rho_{XY} = \frac{COV[X,Y]}{S[X]S[Y]} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

◆ **상관계수의 성질**

- $-1 \leq \rho_{XY} \leq 1$

$$-\sigma_x\sigma_y \leq \sigma_{xy} \leq \sigma_x\sigma_y$$

$$-1 \leq \frac{\sigma_{xy}}{\sigma_x\sigma_y} \leq 1$$

$$\frac{\sigma_{xy}}{\sigma_x\sigma_y} = \rho_{xy}.$$

두 개의 확률변수에 관한 특성값

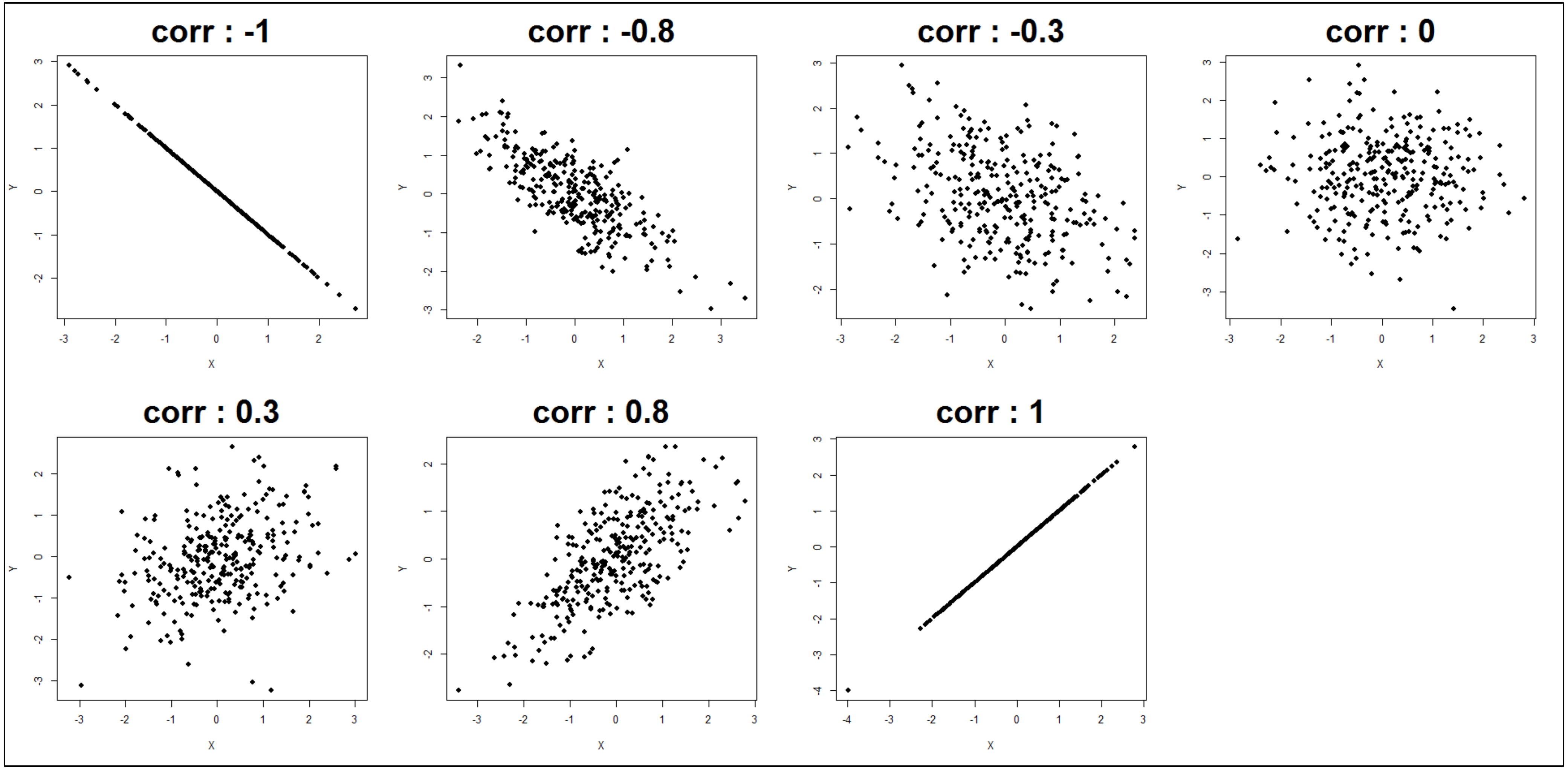
- **상관계수의 성질**

1) 선형관계의 **방향**을 파악할 수 있음.
   - $\rho_{XY} > 0$: X, Y 는 양의 선형 관계를 가짐
   - $\rho_{XY} < 0$: X, Y 는 음의 선형 관계를 가짐

2) 선형관계의 **강도**를 파악할 수 있음.
   - $|\rho_{XY}| \approx 0$: 강도가 약함.
   - $|\rho_{XY}| \approx 1$: 강도가 강함.

3) 자료의 단위를 변경해도 상관계수 값은 변하지 않음.
   - $X' = aX + b$ 이고 , $Y' = cY + d$ 이고, $ac > 0$ 일 때, $\rho_{X'Y'} = \rho_{XY}$

두 개의 확률변수에 관한 특성값



두 개의 확률변수에 관한 특성값

- 두 확률변수 X 와 Y의 합 또는 차에 관한 특성값
  - $E[X \pm Y] = E[X] \pm E[Y]$
  - $V[X \pm Y] = V[X] + V[Y] \pm 2COV[X, Y]$
  - $V[X \pm Y] = V[X] + V[Y]$, 단, X 와 Y는 서로 독립.

두 개의 확률변수에 관한 특성값

- n개의 확률변수 $X_1, X_2, \dots, X_n$의 합에 관한 특성값
  - $E[\sum_{i=1}^{n} X_i] = E[\sum_{i=1}^{n} E[X_i]]$
  - $V[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} V[X_i] \pm \sum_{i \neq j} COV[X_i, X_j]$
  - $V[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} V[X_i]$, 단, $X_1, X_2, \dots, X_n$은 서로 독립.

두 개의 확률변수에 관한 특성값

- Xavier 와 Yvette에 관한 예제 (계속)

		X		
		0	1	2
Y	0	.12	.42	.06
	1	.21	.06	.03
	2	.07	.02	.01

- ◆ Xavier가 한달 동안 판매한 집의 개수에 대한 기댓값과 표준편차를 구하여라.

X	f(x)
0	0.4
1	0.5
2	0.1

$E[X] = \sum x f(x)$

$= 0 \times 0.4 + 1 \times 0.5 + 2 \times 0.1$

$= 0.7$

$V[X] = E[X^2] - E[X]^2$

$= \sum x^2 f(x) - 0.7^2$

$= 0^2 \times 0.4 + 1^2 \times 0.5 + 2^2 \times 0.1 - 0.7^2$

$= 0.41$

$S[X] = \sqrt{V[X]} = 0.6403$

- ◆ Yvette이 한달 동안 판매한 집의 개수에 대한 기댓값과 표준편차를 구하여라.

y	f(y)

$\Rightarrow$ $E[Y] = 0.5$

$V[Y] = 0.45$

$S[Y] = 0.6708$

두 개의 확률변수에 관한 특성값

$E[XY] = \sum_{\text{all } x} \sum_{\text{all } y} xy\, f(x,y)$

$P[X = x \; \& \; Y = y]$

- Xavier 와 Yvette에 관한 예제 (계속)

		X		
		0	1	2
Y	0	.12	.42	.06
	1	.21	.06	.03
	2	.07	.02	.01

- ◆ Xavier와 Yvette의 한달 동안 판매한 집의 개수에 대한 공분산과 상관계수를 구하여라.

$Cov[X,Y] = E[(X-0.7)(Y-0.5)]$

$= \sum\sum (x-0.7)(y-0.5)\, f(x,y)$

$= (0-0.7)(0-0.5)\,0.12 +$

$\cdots$

$+ (2-0.7)(2-0.5) \times 0.01$

$= E[XY] - 0.7 \times 0.5$

$= \sum_{\text{all } x} \sum_{\text{all } y} xy\, f(x,y) - 0.35$

$= (0 \times 0 \times 0.12 + 0 \times 1 \times 0.21 + \cdots + 2 \times 2 \times 0.01) - 0.35$

$= -0.15$

두 개의 확률변수에 관한 특성값

- Xavier 와 Yvette에 관한 예제 (계속)

		X		
		0	1	2
Y	0	.12	.42	.06
	1	.21	.06	.03
	2	.07	.02	.01

- 두 중개업자에 의해 한달 동안 판매된 집의 총 개수에 대한 기대값과 분산을 구하여라.

두 개의 확률변수에 관한 특성값

```
> Ex <- sum( (0:2) * Px )
> Sx <- sqrt( sum( (0:2)^2 * Px ) - Ex^2 )
> Ex
[1] 0.7
> Sx
[1] 0.6403124
> Ey <- sum( (0:2) * Py )
> Sy <- sqrt( sum( (0:2)^2 * Py ) - Ey^2 )
> Ey
[1] 0.5
> Sy
[1] 0.6708204
```

두 개의 확률변수에 관한 특성값

```
> COVxy <- sum( outer( 0:2, 0:2 ) * Pxy ) - Ex*Ey
> CORxy <- COVxy / ( Sx * Sy )
> COVxy
[1] -0.15
> CORxy
[1] -0.3492151
> Ex + Ey
[1] 1.2
> Sx^2+Sy^2+2COVxy
[1] 0.56
```

두 개의 확률변수에 관한 특성값

- 예제
  - 주식 A의 수익률을 X라고 할 때 X의 기대값은 4, 분산은 100이다. 또한 주식 B의 수익률을 Y라고 할 때 Y의 기대값은 7, 분산이 144이다. 두 주식 A와 B의 수익률은 서로 독립이라고 하자. $Z = 0.5X + 0.5Y$로 정의되는 Z의 기대값과 분산은 얼마인가?

두 개의 확률변수에 관한 특성값

- 예제
  - X, Y, Z 3사의 주식으로 포트폴리오가 구성되어 있으며, 주식의 구성비율 및 기대수익률과 이들 3사 수익률 간의 분산과 공분산이 다음과 같다고 하자. 이 포트폴리오의 기대수익률과 분산을 구하시오.

기업	가중치(주식구성비)	기대수익률
X	50%	8%
Y	10%	12%
Z	40%	16%

$$V[X] = 0.20, V[Y] = 0.30, V[Z] = 0.25,$$

$$COV[X,Y] = 0.15, COV[X,Z] = 0.10, COV[Y,Z] = 0.12$$

두 개의 확률변수에 관한 특성값

```
> wt <- c(0.5, 0.1, 0.4)
> Evec <- c(0.08, 0.12, 0.16)
> Epf <- sum( wt * Evec )
> Epf
[1] 0.116
> Vvec <- c(0.2, 0.3, 0.25)
> Cmat <- diag( Vvec )
> Cmat[1, 2] <- Cmat[2, 1] <- 0.15
> Cmat[1, 3] <- Cmat[3, 1] <- 0.10
> Cmat[2, 3] <- Cmat[3, 2] <- 0.12
> Cmat
     [,1] [,2] [,3]
[1,] 0.20 0.15 0.10
[2,] 0.15 0.30 0.12
[3,] 0.10 0.12 0.25
> Vpf <- wt %*% Cmat %*% wt
> Vpf
       [,1]
[1,] 0.1576
```