Flere variabler samtidig

Thea Bjørnland
Institutt for matematiske fag
NTNU

Motivasjon

Meteorologisk institutt   seklima.met.no

$\geq 17\,^{\circ}C$ Juni/Juli/August

Historiske data (1969 - 2019)

Hendelse A: 'Det er en varm sommerdag i Trondheim' $P(A) \approx 0.17$

Hendelse B: 'Det er en varm sommerdag i Oslo' $P(B) \approx 0.36$

? Er A og B uavhengige hendelser?

Er $P(A \cap B) = P(A)\,P(B)$?   <u>Nei</u>

$$P(A \cap B) \approx 0.12 \quad \neq \quad P(A)\,P(B) \approx 0.06$$

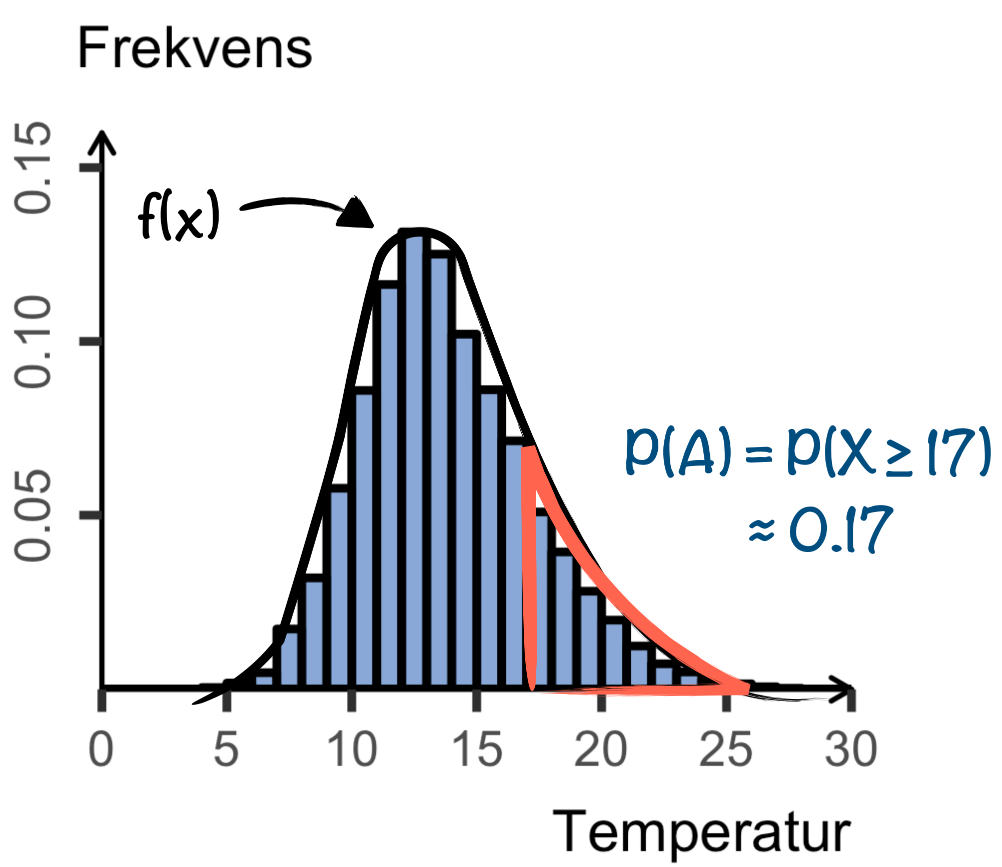
Motivasjon

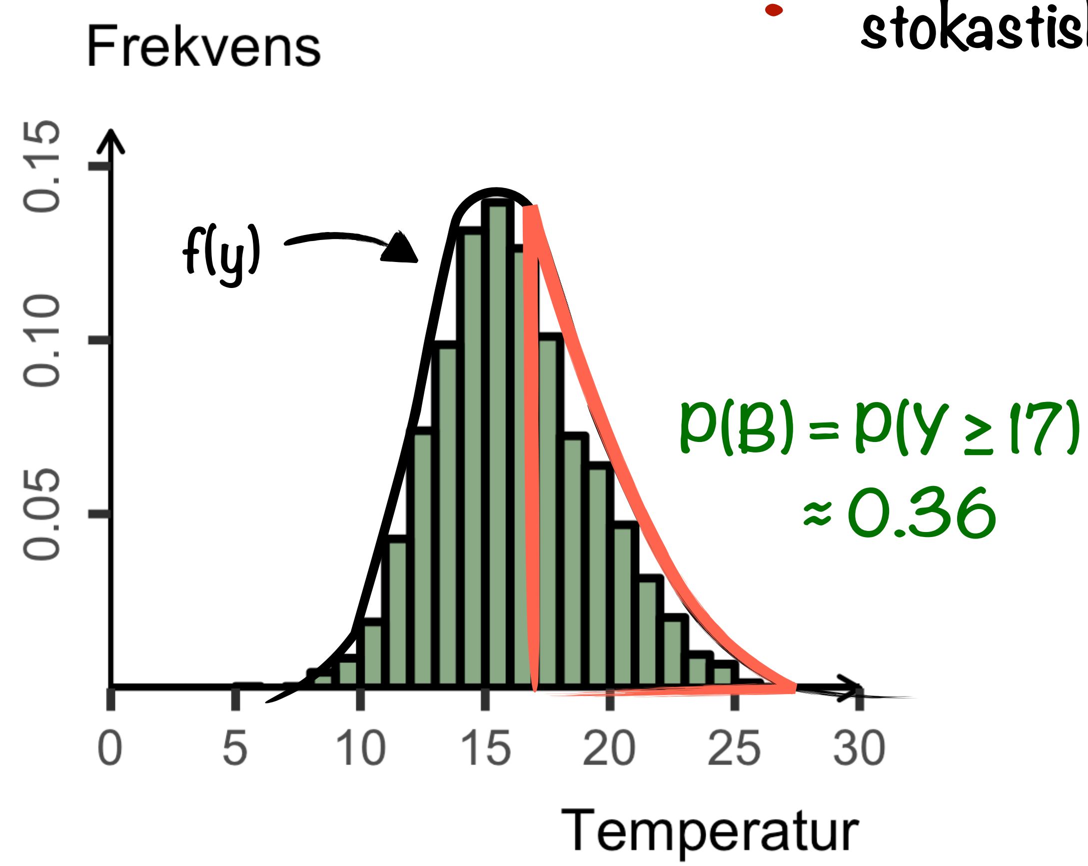
Stokastisk variabel

X: Temperatur i Trondheim på en tilfeldig valgt sommerdag

Y: Temperatur i Oslo på en tilfeldig valgt sommerdag

Er X og Y uavhengige stokastiske variabler?





Historiske data (1969 - 2019)

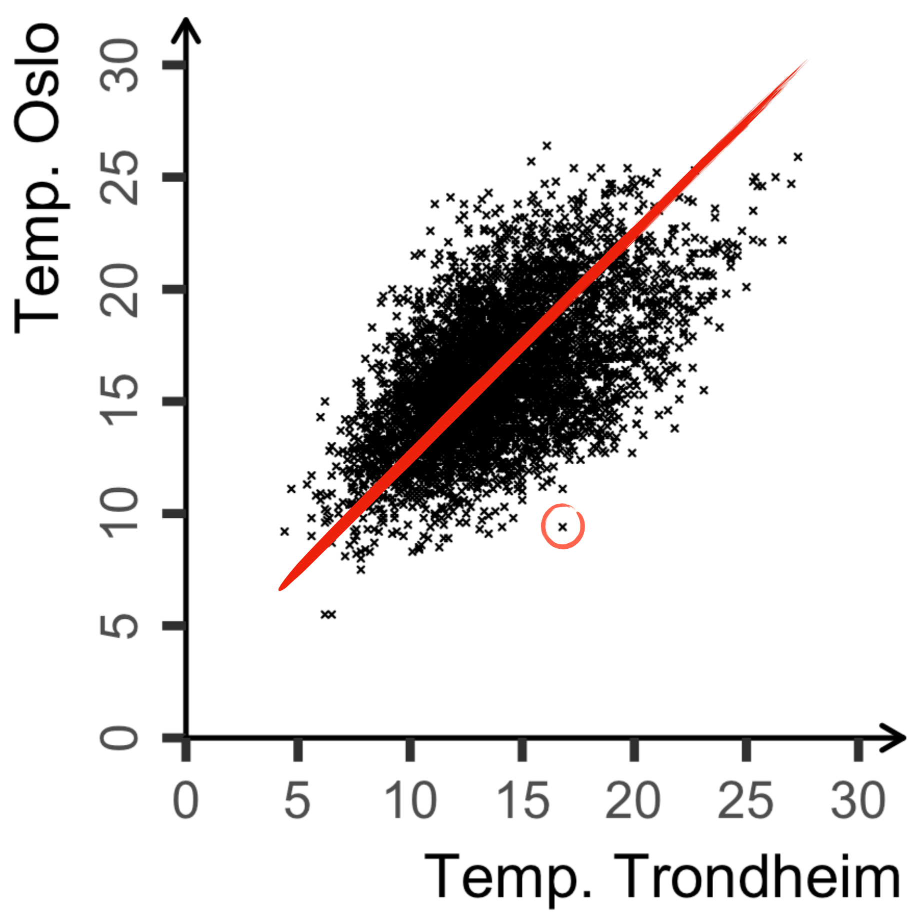
Meteorologisk institutt seklima.met.no

Motivasjon

Stokastisk variabel

X: Temperatur i Trondheim på en tilfeldig valgt sommerdag

Y: Temperatur i Oslo på en tilfeldig valgt sommerdag



? Er X og Y uavhengige stokastiske variabler?

Nei

Historiske data (1969 - 2019)

Diskret simultanfordeling

Simultanfordelingen til to diskrete stokastiske variabler X og Y beskrives av punktsannsynligheter $P(X = x \text{ og } Y = y)$

Tall i verdimengden V_x til X

Tall i verdimengden V_y til Y

“samtidig som”

Diskret simultanfordeling

Simultanfordelingen til to diskrete stokastiske variabler X og Y beskrives av punktsannsynligheter $P(X = x \text{ og } Y = y)$

X: Diskret stokastisk variabel med verdimengde $V_x = \{1, 2, 3\}$

Y: Diskret stokastisk variabel med verdimengde $V_y = \{1, 2\}$

	y	
x	1	2
1	0.1	0.1
2	0.5	0
3	0.2	0.1

$P(X = 1, Y = 2) = 0.1$

$P(X = 3, Y = 1) = 0.2$

Diskret simultanfordeling

Simultanfordelingen til to diskrete stokastiske variabler X og Y beskrives av punktsannsynligheter $P(X = x \text{ og } Y = y)$

X: Diskret stokastisk variabel med verdimengde $V_x = \{1, 2, 3\}$

Y: Diskret stokastisk variabel med verdimengde $V_y = \{1, 2\}$

	y	
x	1	2
1	0.1	0.1
2	0.5	0
3	0.2	0.1

$P(X = 1) = 0.1 + 0.1 = 0.2$

$P(Y = 1) = 0.1 + 0.5 + 0.2 = 0.8$

Diskret simultanfordeling

Simultanfordelingen til to diskrete stokastiske variabler X og Y beskrives av punktsannsynligheter $P(X = x \text{ og } Y = y)$

X: Diskret stokastisk variabel med verdimengde $V_x = \{1, 2, 3\}$

Y: Diskret stokastisk variabel med verdimengde $V_y = \{1, 2\}$

x	$y = 1$	$y = 2$	$P(X = x)$
1	0.1	0.1	0.2
2	0.5	0	0.5
3	0.2	0.1	0.3
$P(Y = y)$	0.8	0.2	

"Marginal sannsynlighetsfordeling for X"

"Marginal sannsynlighetsfordeling for Y"

Diskret simultanfordeling

Simultanfordelingen til to diskrete stokastiske variabler X og Y beskrives av punktsannsynligheter $P(X = x \text{ og } Y = y)$

X: Diskret stokastisk variabel med verdimengde $V_x = \{1, 2, 3\}$

Y: Diskret stokastisk variabel med verdimengde $V_y = \{1, 2\}$

x \ y	1	2	$P(X = x)$
1	0.1	0.1	0.2
2	0.5	0	0.5
3	0.2	0.1	0.3
$P(Y = y)$	0.8	0.2	

Betinget sannsynlighet:

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$P(X = 3 \mid Y = 1) = \frac{0.2}{0.8} = 0.25$$

Diskret simultanfordeling

Simultanfordelingen til to diskrete stokastiske variabler X og Y beskrives av punktsannsynligheter $P(X = x \text{ og } Y = y)$

X: Diskret stokastisk variabel med verdimengde $V_x = \{1, 2, 3\}$

Y: Diskret stokastisk variabel med verdimengde $V_y = \{1, 2\}$

x	$y = 1$	$y = 2$	$P(X = x)$
1	0.1	0.1	0.2
2	0.5	0	0.5
3	0.2	0.1	0.3
$P(Y = y)$	0.8	0.2	

Betinget sannsynlighet:

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$P(X \leq 2 \mid Y = 1) = \frac{0.1 + 0.5}{0.8} = 0.75$$

Diskret simultanfordeling

Simultanfordelingen til to diskrete stokastiske variabler X og Y beskrives av punktsannsynligheter $P(X = x \text{ og } Y = y)$

X: Diskret stokastisk variabel med verdimengde $V_x = \{1, 2, 3\}$

Y: Diskret stokastisk variabel med verdimengde $V_y = \{1, 2\}$

	y		
x	1	2	$P(X = x)$
1	0.1	0.1	0.2
2	0.5	0	0.5
3	0.2	0.1	0.3
$P(Y = y)$	0.8	0.2	

$P(X = 1, Y = 2) = 0.1$

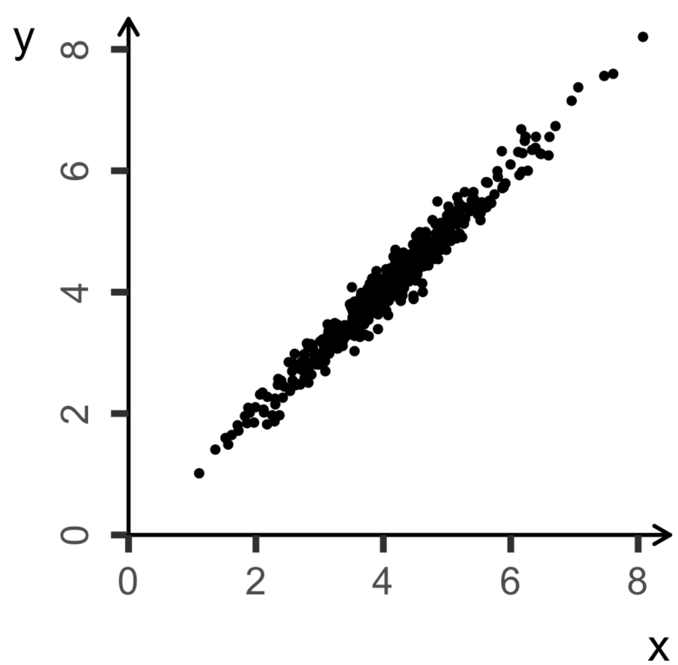
$P(X = 1)P(Y = 2)$

Er X og Y uavhengige? Er $P(X = x, Y = y) = P(X = x)P(Y = y)$ for alle x og y?

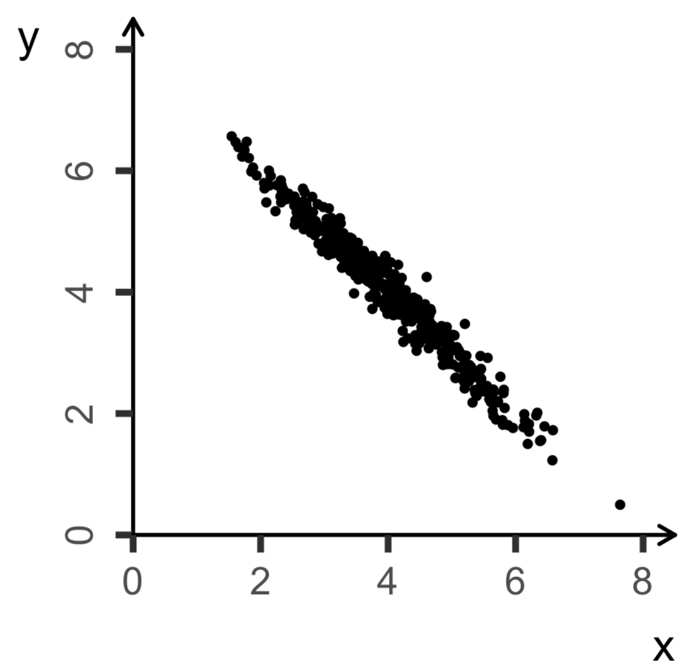
Nei

Kovarians og korrelasjon

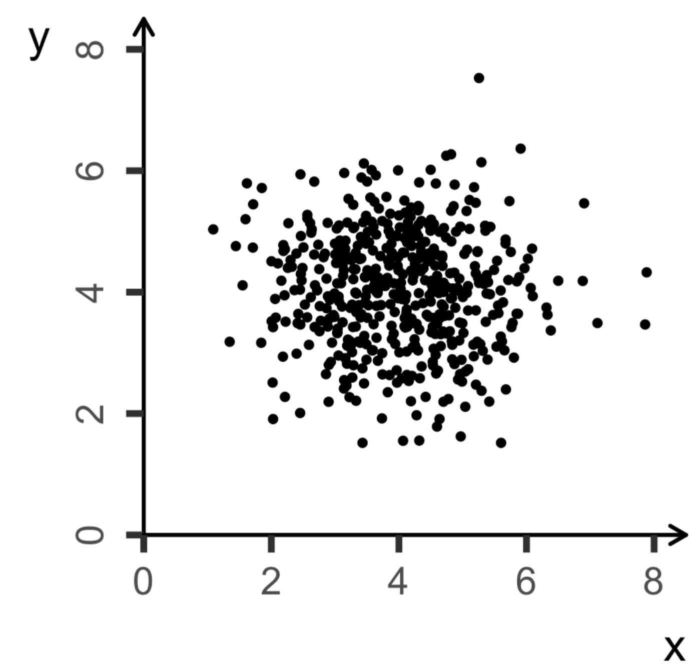
- mål på lineær samvariasjon mellom to variabler



Cov(X, Y) > 0



Cov(X, Y) < 0



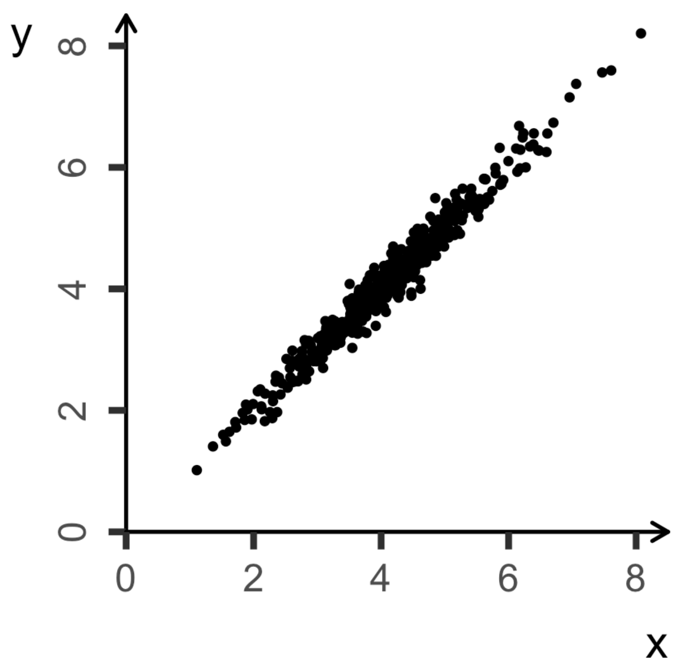
Cov(X, Y) = 0

$\Uparrow$

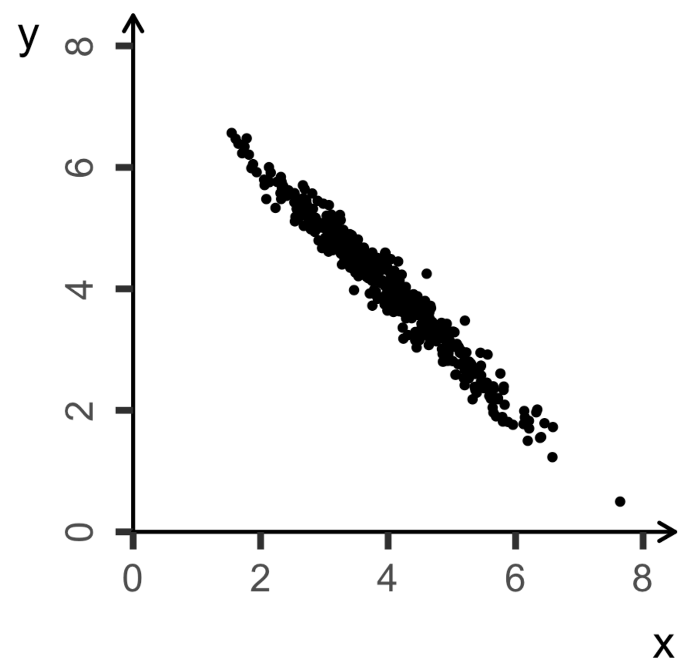
X og Y uavhengige

Kovarians og korrelasjon

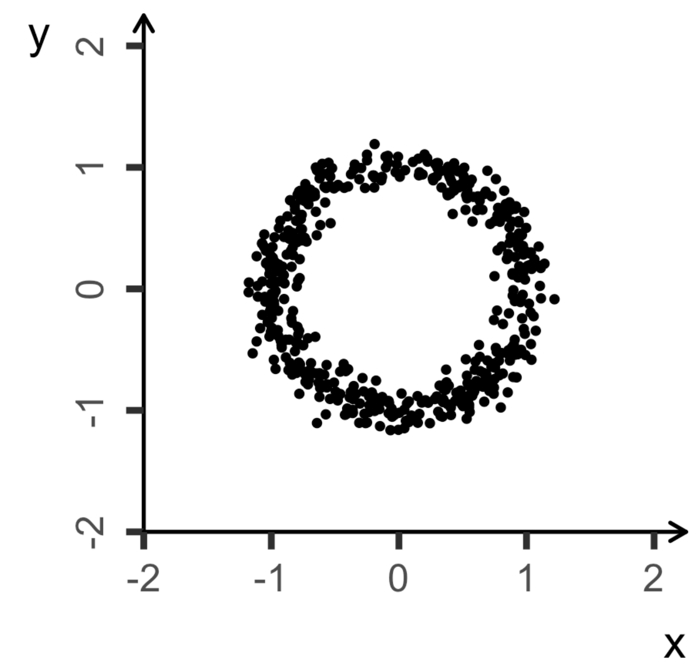
- mål på lineær samvariasjon mellom to variabler



Cov(X, Y) > 0



Cov(X, Y) < 0



Cov(X, Y) = 0

Definisjon: Kovarians

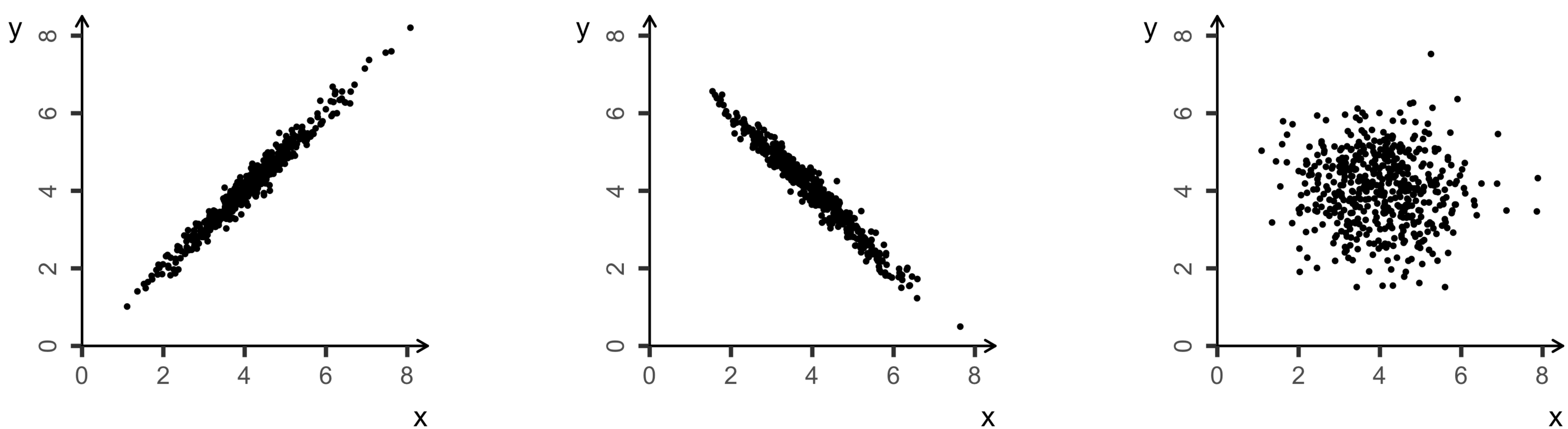
- mål på lineær samvariasjon mellom to variabler

$\text{Var}(X) = \text{E}\left((X - \mu_x)^2\right)$

X: Stokastisk variabel med forventningsverdi $\text{E}(X) = \mu_x$ og varians $\text{Var}(X) = \sigma_x^2$

Y: Stokastisk variabel med forventningsverdi $\text{E}(Y) = \mu_y$ og varians $\text{Var}(Y) = \sigma_y^2$

$$\text{Cov}(X, Y) = \text{E}\left((X - \mu_x) \cdot (Y - \mu_y)\right)$$



Cov(X, Y) > 0 | Cov(X, Y) < 0 | Cov(X, Y) = 0

Definisjon: Korrelasjon

- mål på lineær samvariasjon mellom to variabler

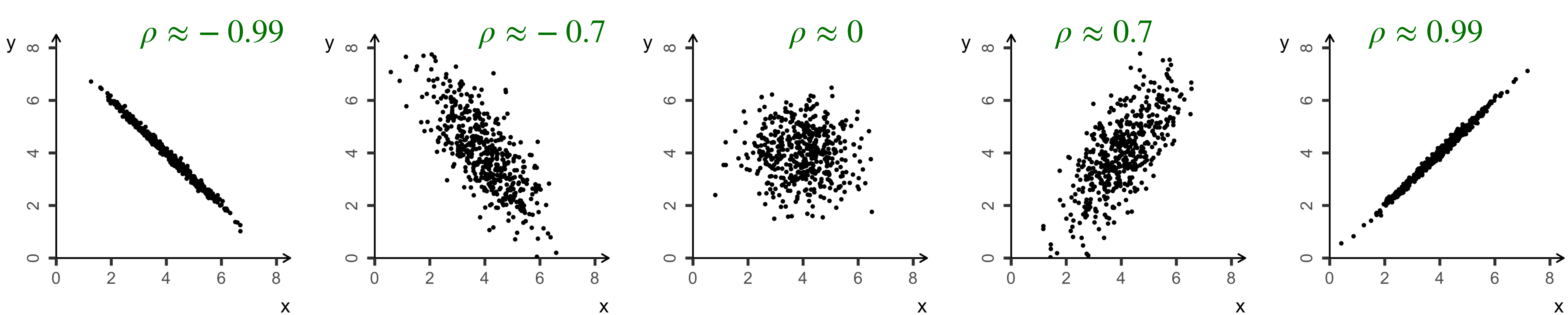
$\text{Var}(X) = \text{E}\left((X - \mu_x)^2\right)$

X: Stokastisk variabel med forventningsverdi $\text{E}(X) = \mu_x$ og varians $\text{Var}(X) = \sigma_x^2$

Y: Stokastisk variabel med forventningsverdi $\text{E}(Y) = \mu_y$ og varians $\text{Var}(Y) = \sigma_y^2$

$$\text{Cov}(X, Y) = \text{E}\left((X - \mu_x) \cdot (Y - \mu_y)\right)$$

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_x \cdot \sigma_y} \qquad -1 \leq \rho_{X,Y} \leq 1$$



Eksempler fra https://snl.no/korrelasjon

Regneregler for forventning og varians

Stokastiske variabler

Lineærkombinasjon: $X + Y$

$$E(X+Y) = E(X) + E(Y)$$

$$Var(X+Y) = Var(X) + Var(Y) + \cancel{2 \cdot Cov(X,Y)}$$

= 0 dersom f.eks X og Y er uavhengige

Regneregler for forventning og varians

Stokastiske variabler

Lineærkombinasjon: $aX + bY$

konstanter

$$E(aX + bY) = aE(X) + bE(Y)$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab \cdot Cov(X, Y)$$

?

Differanse:

$$X - Y$$

$$E(X - Y) = 1 \cdot E(X) + (-1) \cdot E(Y)$$
$$= E(X) - E(Y)$$

$$a = 1 \quad b = -1$$

Regneregler for forventning og varians

Stokastiske variabler

Lineærkombinasjon: $aX + bY$

konstanter

$$E(aX + bY) = aE(X) + bE(Y)$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab \cdot Cov(X, Y)$$

Differanse:

$$X - Y$$

$$a = 1 \quad b = -1$$

$$Var(X - Y) = 1^2 \cdot Var(X) + (-1)^2 \cdot Var(Y) + 2 \cdot 1 \cdot (-1) \cdot Cov(X, Y)$$

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