

# データ解析 (第2回)

静岡大学システム工学科

安藤 和敏

# 第2章 統計的方法の基礎知識

## 2.1 データのまとめ方

## (2) 2つの量的変数の場合

データ No.	変数 (変量)	
	$x$	$y$
1	$x_1$	$y_1$
2	$x_2$	$y_2$
⋮	⋮	⋮
$i$	$x_i$	$y_i$
⋮	⋮	⋮
$n$	$x_n$	$y_n$

データ No.	変数 (変量)	
	$x$	$y$
1	$x_1$	$y_1$
⋮	⋮	⋮
$n$	$x_n$	$y_n$
平均	$\bar{x}$	$\bar{y}$
平方和	$S_{xx}$	$S_{yy}$
分散	$V_x$	$V_y$
標準偏差	$S_x$	$S_y$

# 偏差積和

## 偏差積和

$$\begin{aligned} S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \sum_{i=1}^n x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} \end{aligned}$$

# 共分散, 相関係数

共分散

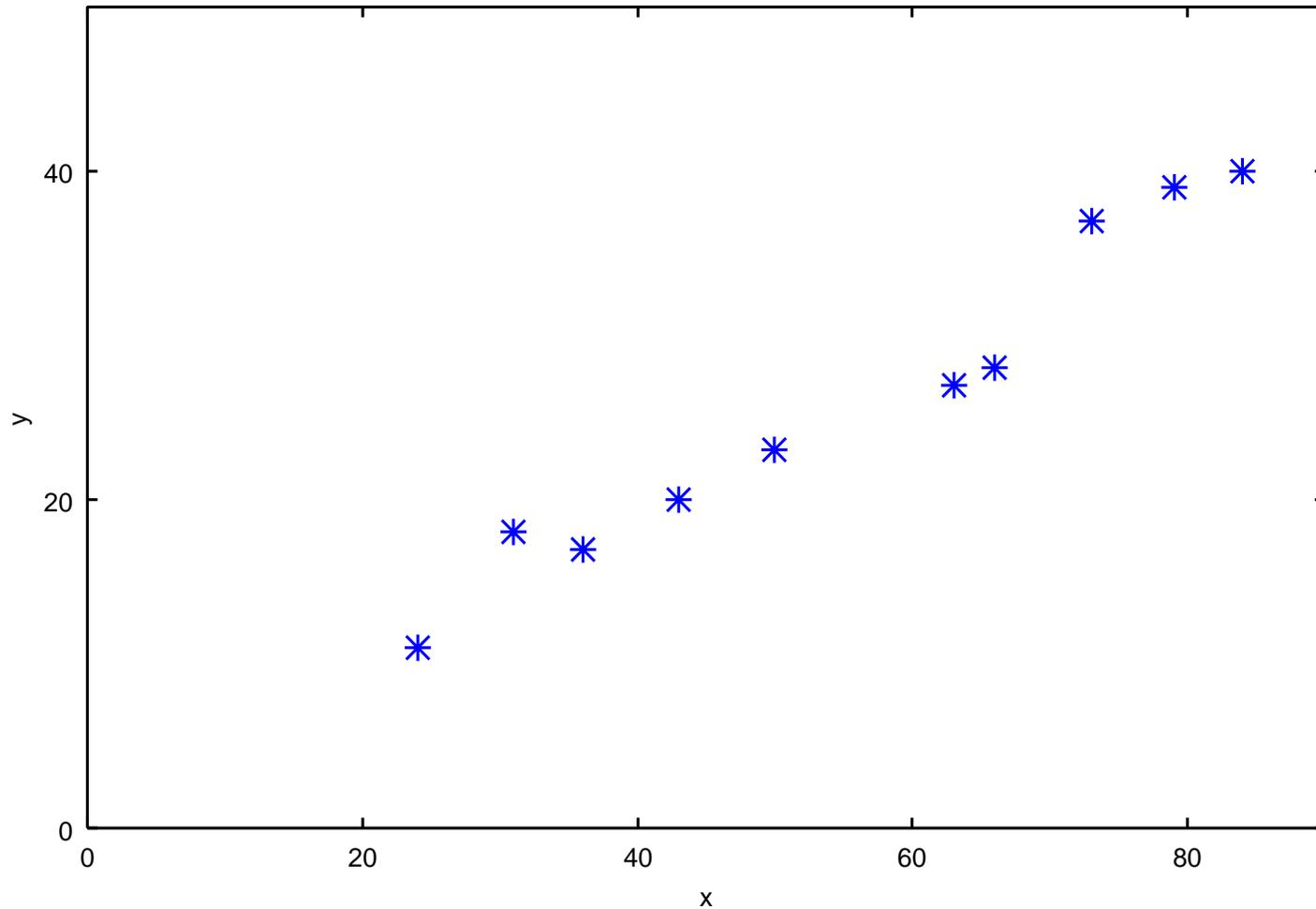
$$C_{xy} = \frac{S_{xy}}{n - 1} \quad (2.9)$$

相関係数

$$r_{xy} = \frac{C_{xy}}{\sqrt{V_x V_y}} \quad (2.10)$$

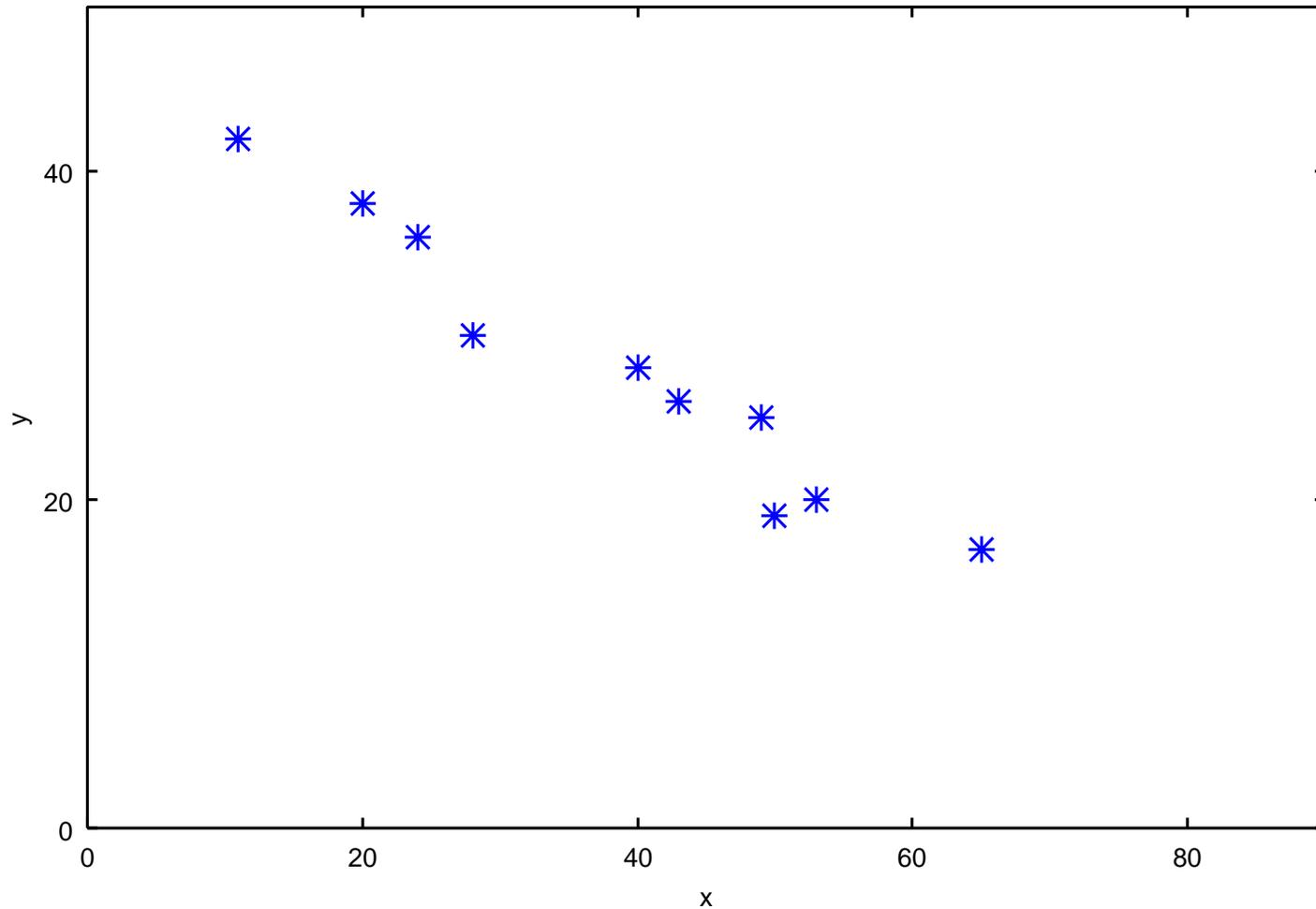
相関係数は  $-1 \leq r_{xy} \leq 1$  を満たす.

# 散布図1 (正の相関)



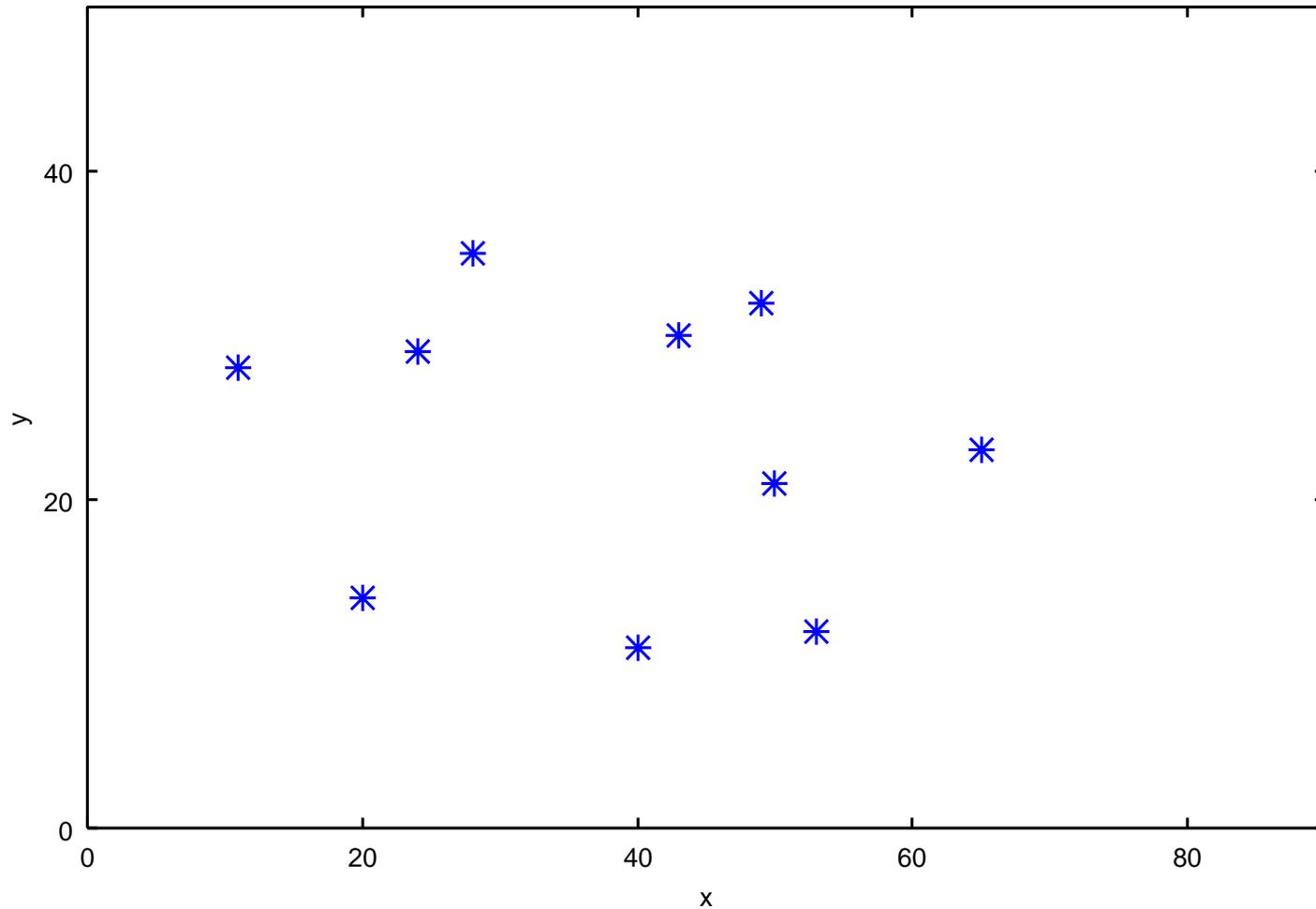
$$r_{xy} = 1$$

# 散布図2 (負の相関)



$$r_{xy} = -1$$

# 散布図3 (無相関)



$$r_{xy} \doteq 0$$

## 2.2 確率分布

多変量解析では、変数  $x$  はなんらかの確率分布にしたがっていること、つまり、変数  $x$  は**確率変数**であること、を仮定することが多い。確率変数の分布はその確率密度関数によって定義される。  
⇒ **母集団分布**