データ解析 (第 12 回)

安藤 和敏

静岡大学システム工学科

[復習] y_i がしたがう分布

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2) \quad (i = 1, \cdots, n)$$

のとき,

y_i は期待値 $\beta_0 + \beta_1 x_i$, 分散 σ^2 の正規分布にしたがう:

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \quad (i = 1, \cdots, n).$$

ここで, $\beta_0, \beta_1, \sigma^2$ は未知のパラメータである.

[復習] $\hat{\beta}_1$ がしたがう分布

$\hat{\beta}_1$ は正規分布にしたがう確率変数 y_i の一次結合:

$$\hat{\beta}_1 = \frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x}) y_i .$$

であるから, $\hat{\beta}_1$ も正規分布にしたがう.
その正規分布の期待値は β_1, 分散は $\frac{\sigma^2}{S_{xx}}$ である.

$\hat{\beta}_1$ は期待値 β_1, 分散 $\frac{\sigma^2}{S_{xx}}$ の正規分布にしたがう:

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right). \tag{4.25}$$

ここで, β_1 と σ^2 は未知のパラメータである.

標準化

$\hat{\beta}_1$ を標準化 (式 (2.34)) すると,

$$u = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/S_{xx}}} \sim N(0, 1^2). \qquad (4.26)$$

ここで, β_1 と σ^2 は未知のパラメータである.

t-分布

$$u = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}} \sim N(0, 1^2). \tag{4.26}$$

の σ^2 にその不偏推定量

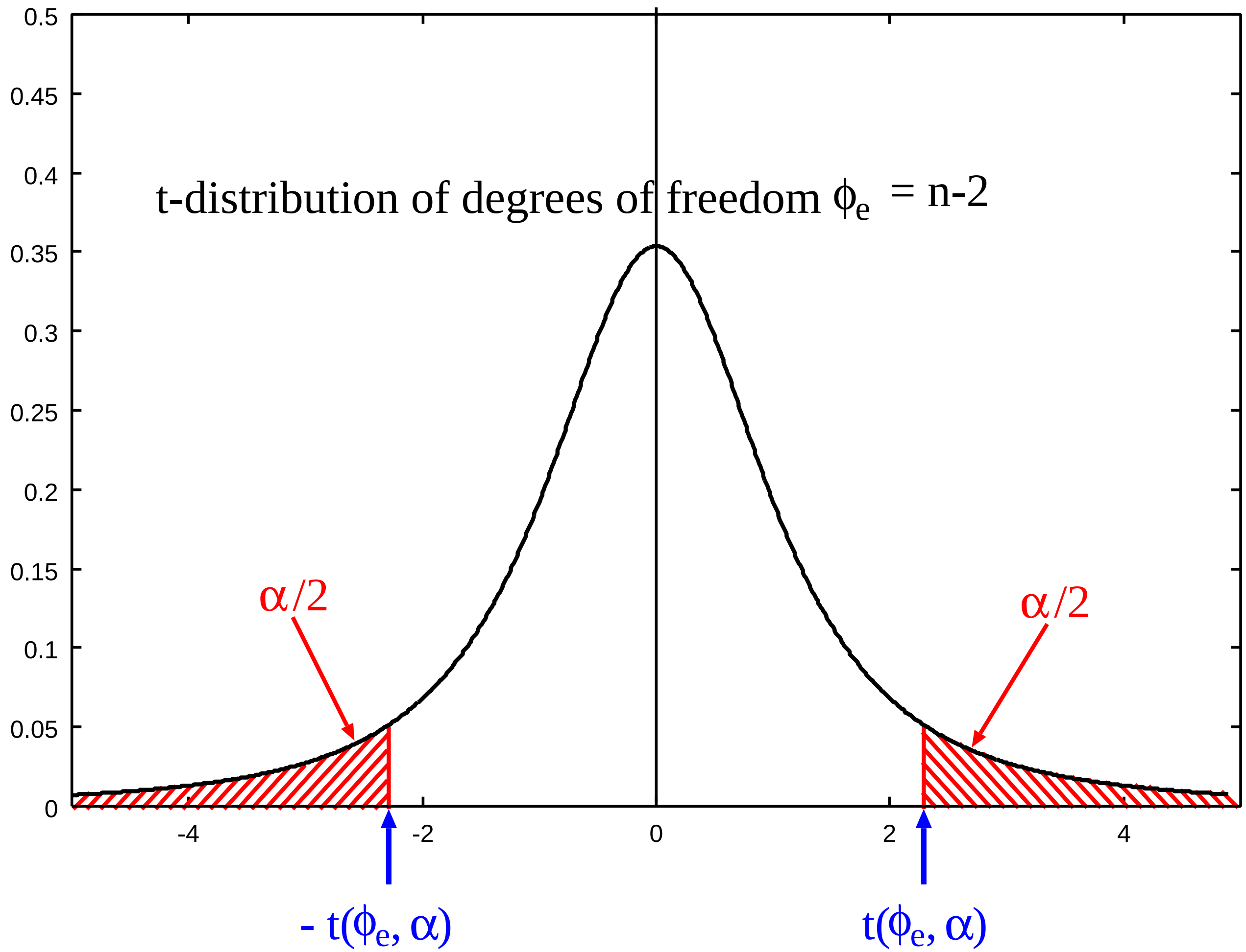
$$\hat{\sigma}^2 = V_e = \frac{S_e}{\phi_e} = \frac{S_e}{n-2} \tag{4.18}$$

を代入すると,

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V_e / S_{xx}}} \sim t(\phi_e). \tag{4.27}$$

ここで, $\phi_e = n - 2$.

自由度ϕ_eのt分布の両側100α%点: $t(\phi_e, \alpha)$



$$\alpha = \Pr(|t| \geq t(\phi_e, \alpha)) \tag{2.62}$$

$\beta_1 = 0$ の検定

x という変数が, 説明変数として予測に役立つかどうか？を検定してみよう. 有意水準 (⇒ 教科書 p. 25) を α とする.
役立たない, つまり, $\beta_1 = 0$ と仮定すると,

$$t_0 = \frac{\hat{\beta}_1 - 0}{\sqrt{V_e / S_{xx}}} \tag{4.29}$$

は, 自由度 $n - 2$ の t 分布にしたがう.

$\beta_1 = 0$ の検定 (続き)

したがって,

$$|t_0| = \frac{|\hat{\beta}_1|}{\sqrt{V_e/S_{xx}}} \geq t(\phi_e, \alpha) \qquad (*)$$

となる確率は, α (=小さい数) である. これが成り立つのは起りそうもないこと (100α%の確率) であるから, 式 (*) が成り立てば, この「x は予測に役立たない」という仮定は正しそうもないと判断し, 成り立たなければ, 正しそうだと判断する.

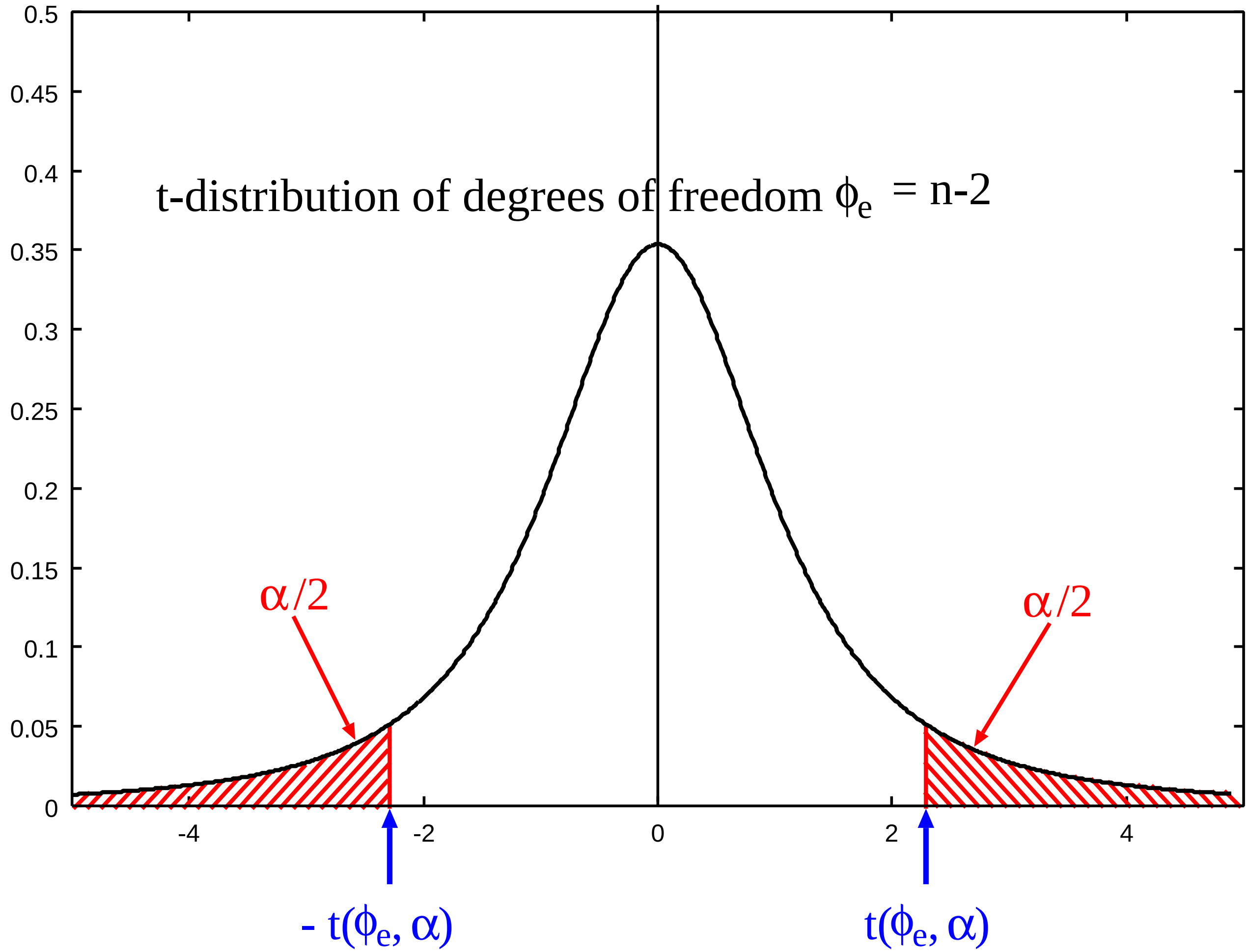
$\beta_1 = 0$ の検定 (続き)

これを統計学の言葉でしゃべると,

式 (*) が成り立てば, 危険率 α で仮説: $\beta_1 = 0$ を棄却し, 成り立たなければ仮説: $\beta_1 = 0$ を採択する.

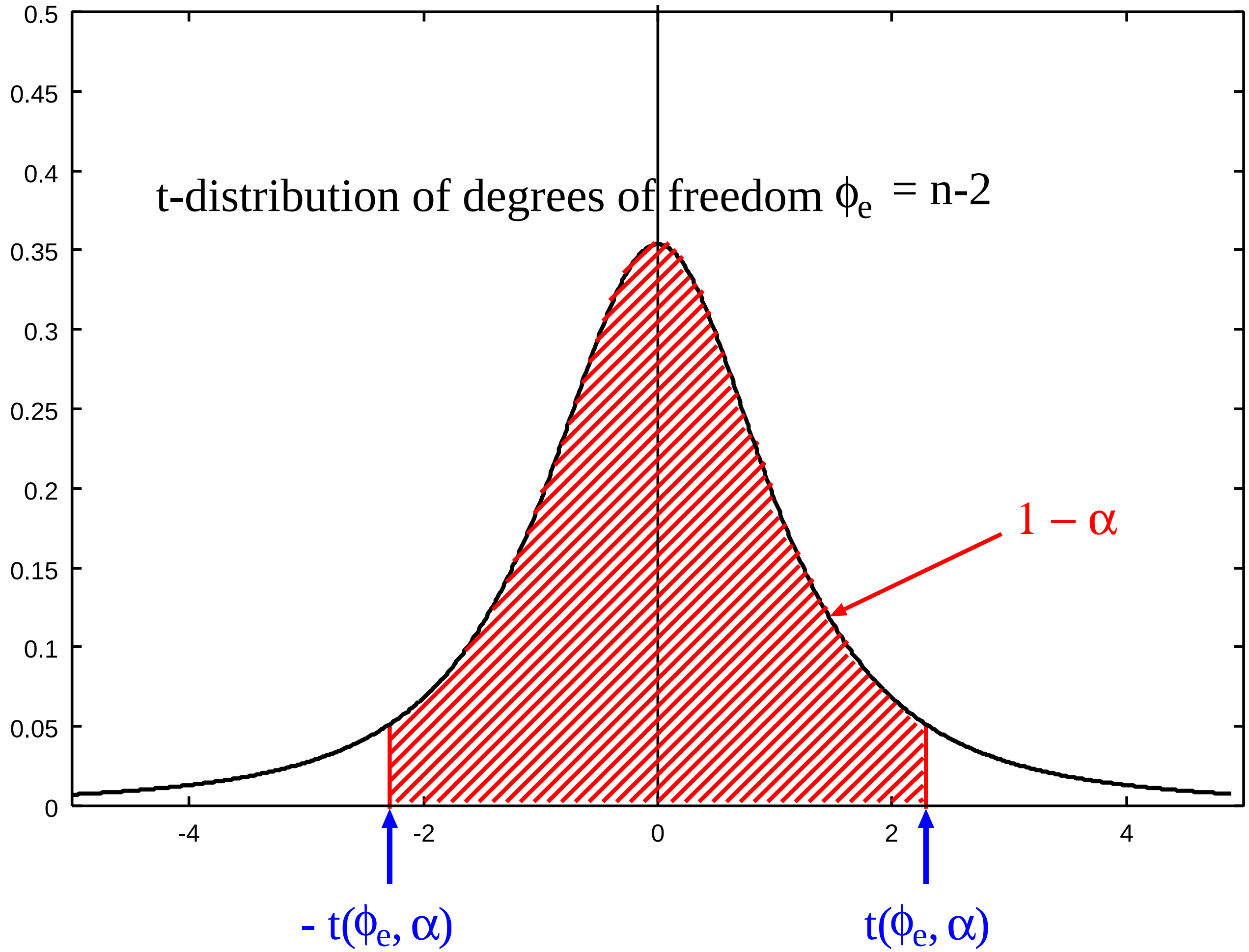
になる.

$\beta_1 = 0$の信頼区間



$$\alpha = \Pr(|t| \geq t(\phi_e, \alpha)) \tag{2.62}$$

$\beta_1 = 0$の信頼区間(続き)



$$1-\alpha = \Pr(|t| \leq t(\phi_e, \alpha))$$

$$1-\alpha = \Pr(|t| \leq t(\phi_e, \alpha))$$

で,

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V_e/S_{xx}}} \sim t(\phi_e). \tag{4.27}$$

なのだから,

$$\frac{|\hat{\beta}_1 - \beta_1|}{\sqrt{V_e/S_{xx}}} \leq t(\phi_e, \alpha)$$

が $(1-\alpha)100$%の確率で成り立つ.

書き直すと

$$\hat{\beta}_1 - t(\phi_e, \alpha)\sqrt{V_e/S_{xx}} \leq \beta_1 \leq \hat{\beta}_1 + t(\phi_e, \alpha)\sqrt{V_e/S_{xx}}$$

が $(1-\alpha)100$%の確率で成り立つ.
ゆえに,

β_1 の $(1-\alpha)100$%信頼区間は,

$$\hat{\beta}_1 \pm t(\phi_e, \alpha)\sqrt{V_e/S_{xx}}$$

である.

$\hat{\beta}_0$ がしたがう分布

$\hat{\beta}_0$ は期待値 β_0, 分散 $\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2$ の正規分布にしたがう:

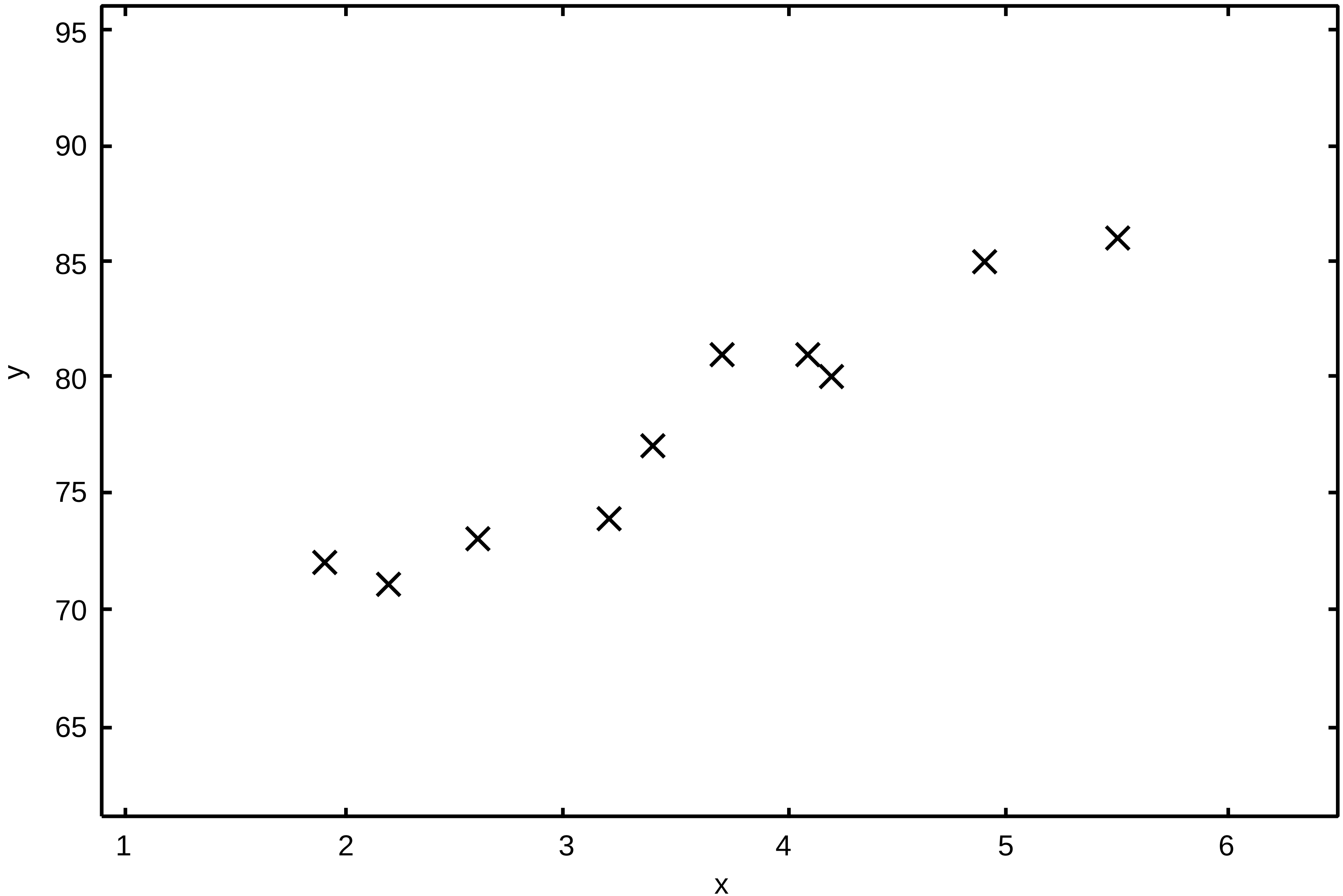
$$\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2\right).$$

ここで, β_0 と σ^2 は未知のパラメータである.

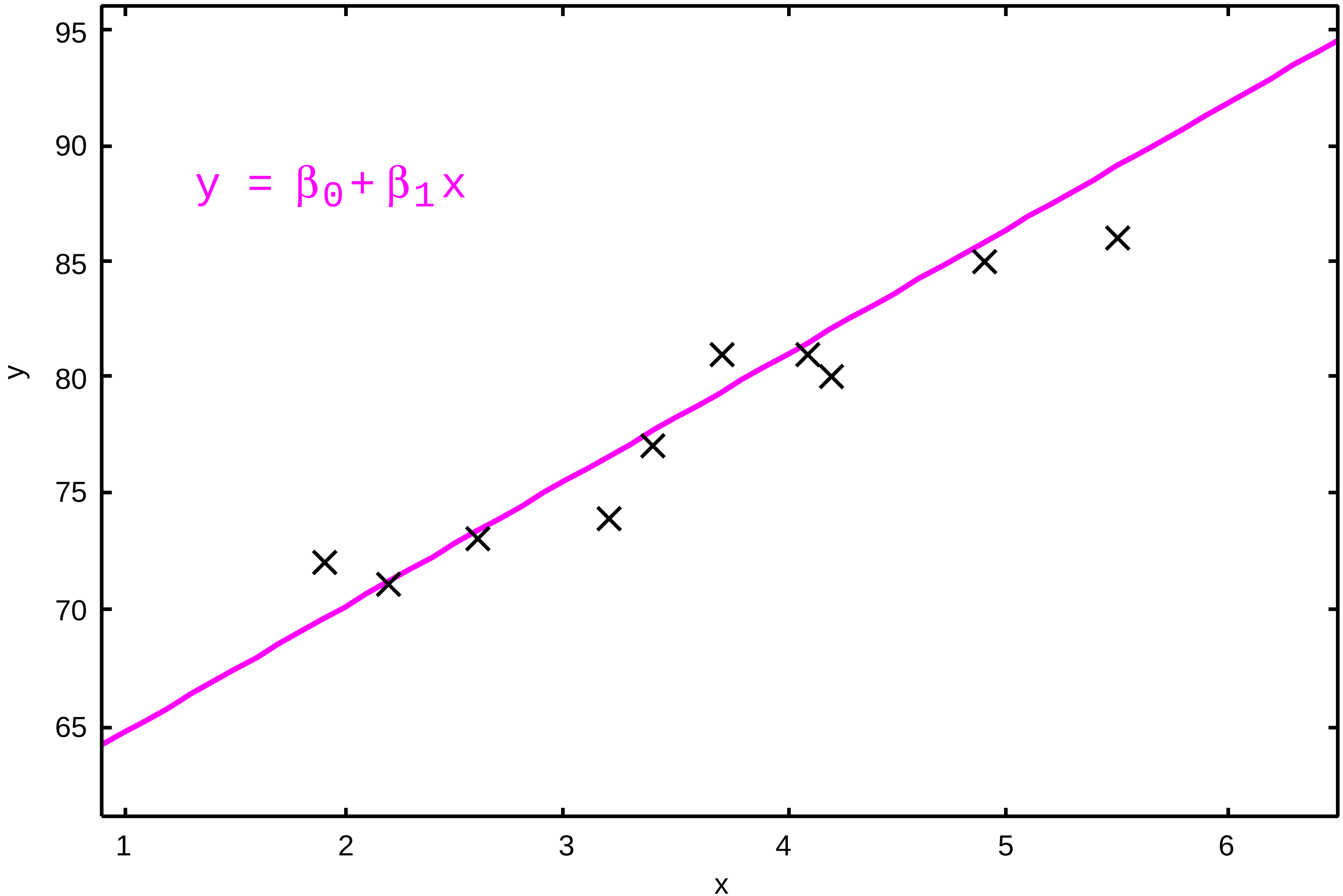
例題 2 をやってみよう.

(5) 得られた回帰式の利用 (p. 54–55)

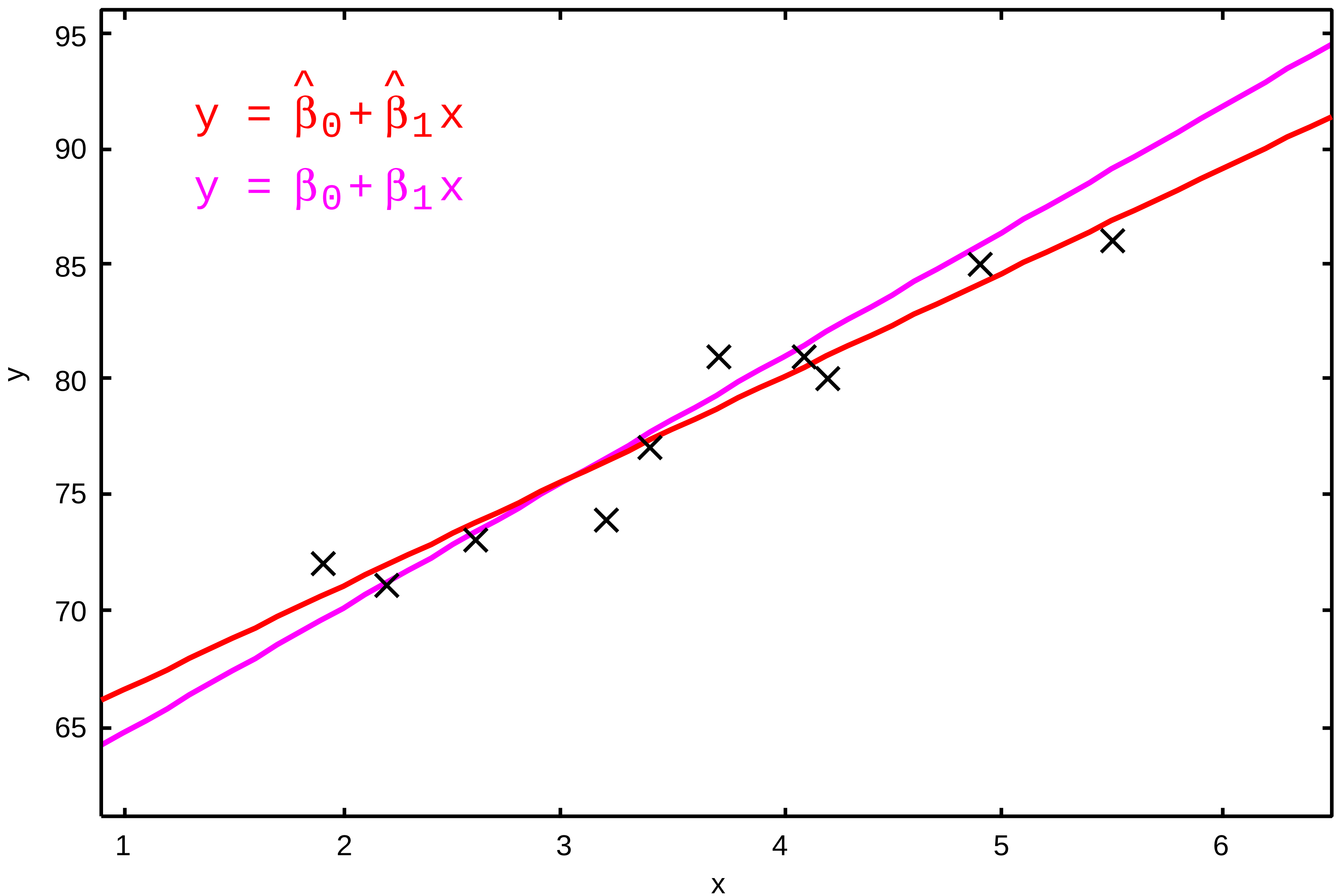
データ (x_i, y_i)



単回帰モデル $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

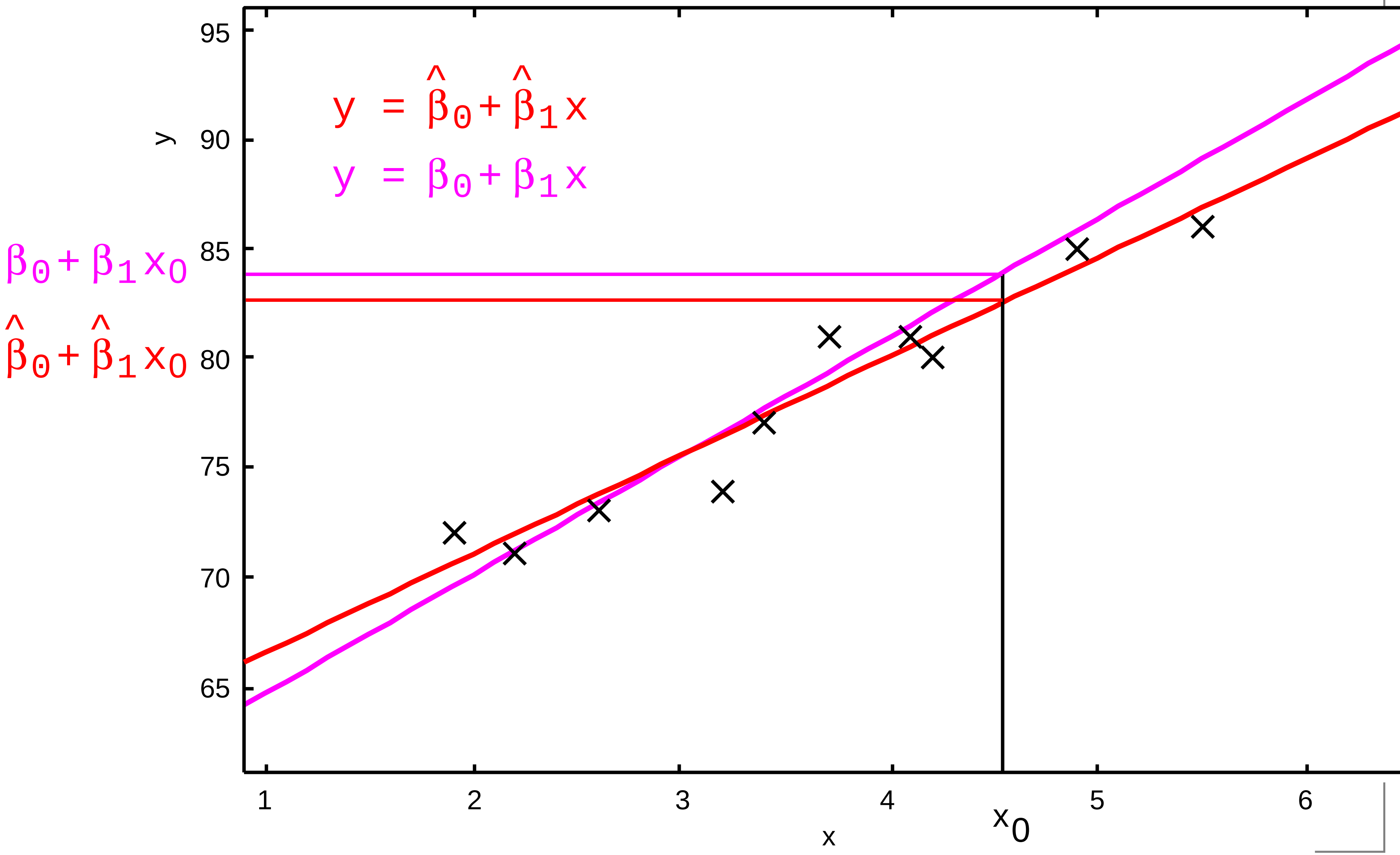


直線の推定



$y = \hat{\beta}_0 + \hat{\beta}_1 x$
$y = \beta_0 + \beta_1 x$
95
90
85
80
75
70
65
y
1
2
3
4
5
6
x

x_0 に対する母回帰 $\beta_0 + \beta_1 x_0$ の推定



$$
\begin{aligned}
E(\hat{\beta}_0 + \hat{\beta}_1 x_0) &= E(\hat{\beta}_0) + x_0 E(\hat{\beta}_1) \\
&= \beta_0 + \beta_1 x_0
\end{aligned}
$$

だから,

$\hat{\beta}_0 + \hat{\beta}_1 x_0$ は, $\beta_0 + \beta_1 x_0$ の不偏推定量である.

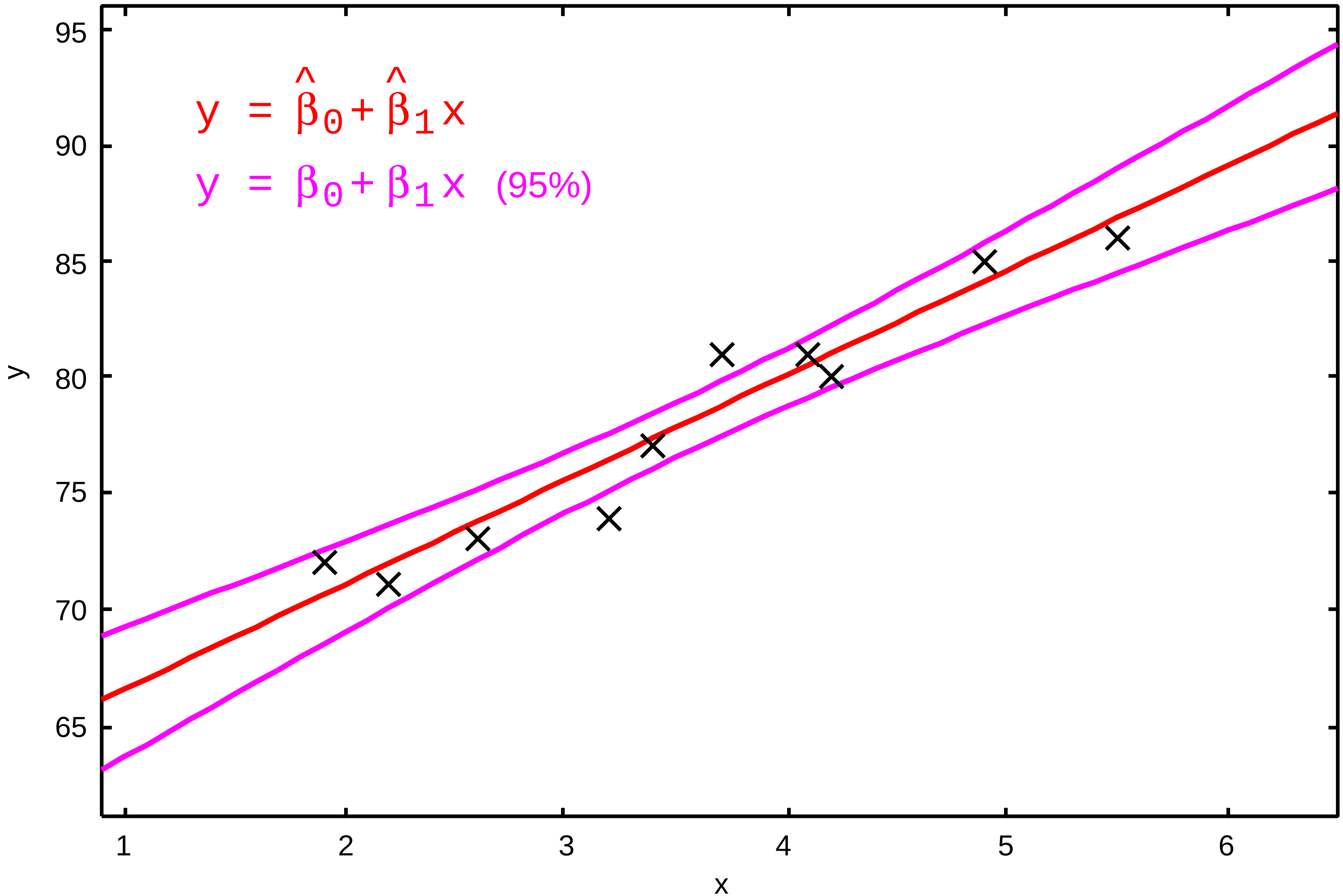
x_0 に対する母回帰 $\beta_0 + \beta_1 x_0$ の区間推定

x_0 に対する母回帰 $\beta_0 + \beta_1 x_0$ の 95%信頼区間は,

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t(\phi_e, 0.05)\sqrt{\left\{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\} V_e} \tag{4.40}$$

である.

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t(\phi_e, 0.05)\sqrt{\left\{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\} V_e}$$



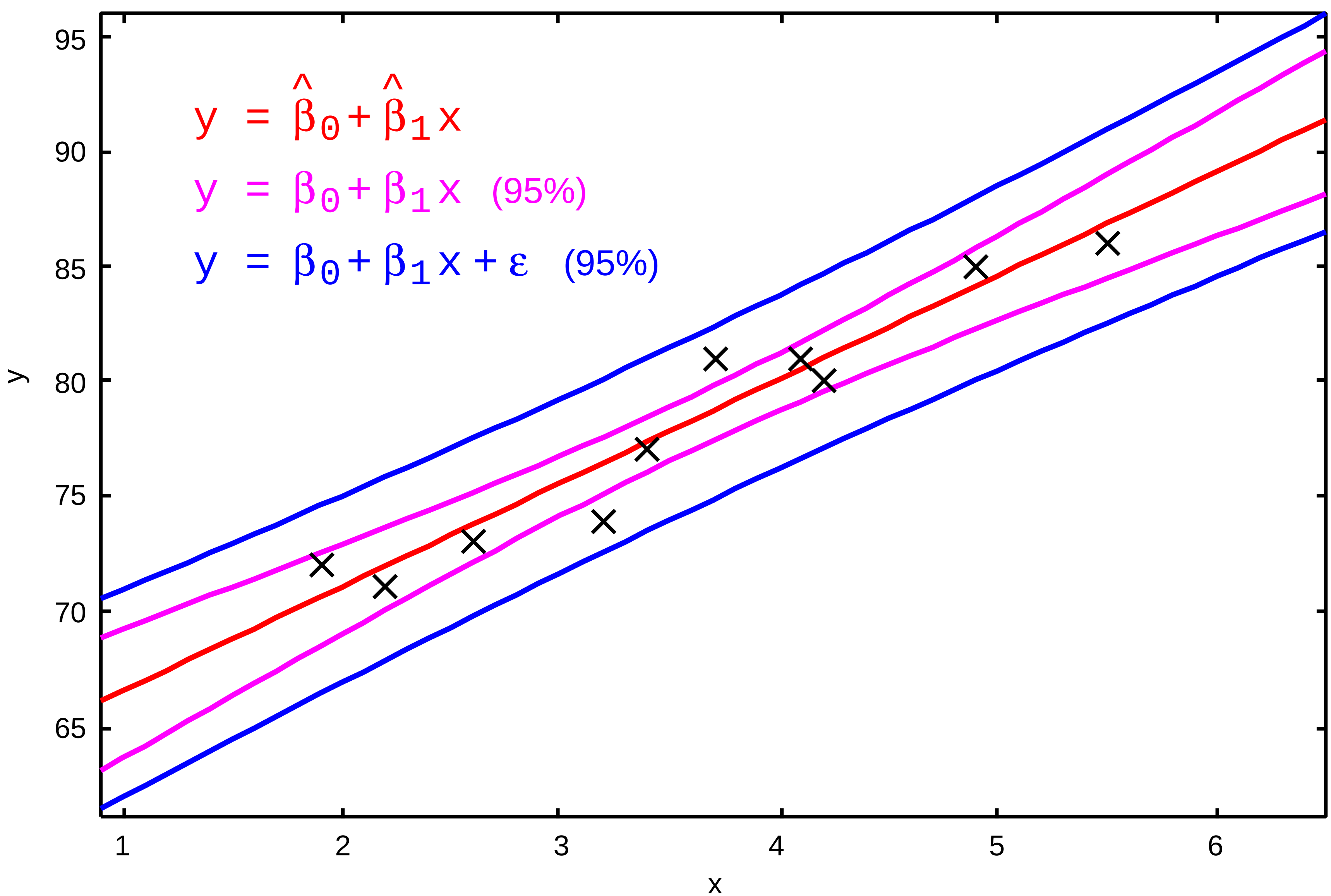
$\beta_0 + \beta_1 x_0 + \epsilon$ の区間推定

$\beta_0 + \beta_1 x_0 + \epsilon$ の 95%信頼区間は,

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t(\phi_e, 0.05)\sqrt{\left\{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\} V_e} \tag{4.41}$$

で与えられる.

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t(\phi_e, 0.05)\sqrt{\left\{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\} V_e}$$



例題4をやってみよう．