傾向性の検定における秘匿疫学調査プロトコル

佐藤智貴† 菊池浩明† 佐久間淳‡

†東海大学 ‡筑波大学

背景:傾向性の検定とプライバシー

投与した薬剤の用量に対しての反応に傾向があるか

ある新薬が血圧を下げる効果があるか

組織A	A群	B群	C群	D群
	10ppm	100ppm	1000ppm	10000ppm
組織B	8.06	7.97	7.66	8.00
	8.27	7.66	7.71	7.89
	8.45	8.30	7.88	7.40
平均	8.2	8.0	7.8	7.8

背景:傾向性の検定とプライバシー

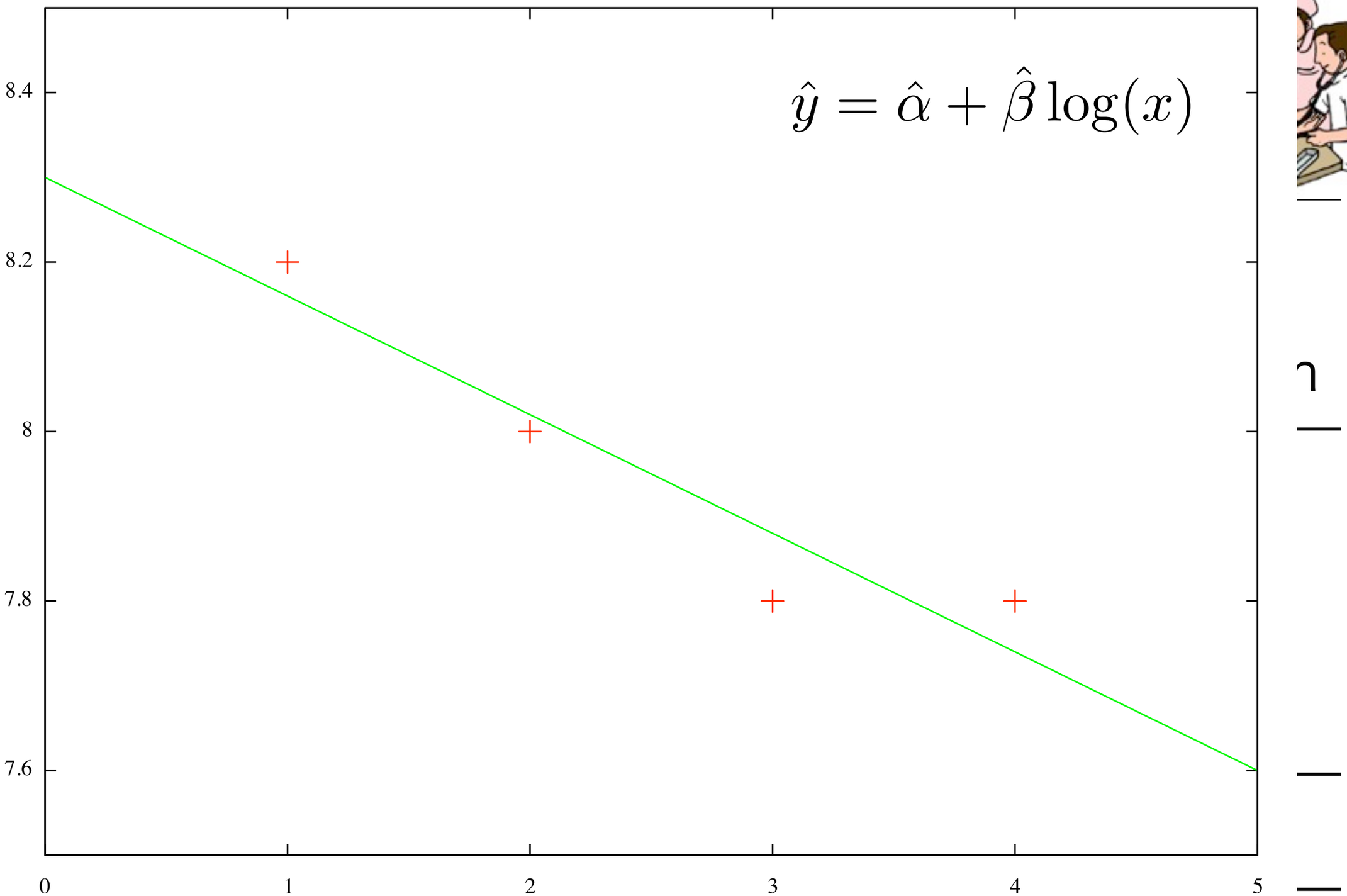
投与

あ

組織

組織

平



背景:傾向性の検定とプライバシー

投与した薬剤の用量に対しての反応に傾向があるか

ある新薬が血圧を下げる効果があるか

組織A	A群	B群	C群	D群
	10ppm	100ppm	1000ppm	10000ppm
組織B	8.06	7.97	7.66	8.00
	8.27	7.66	7.71	7.89
	8.45	8.30	7.88	7.40
平均	8.2	8.0	7.8	7.8

2者間プロトコルの意義

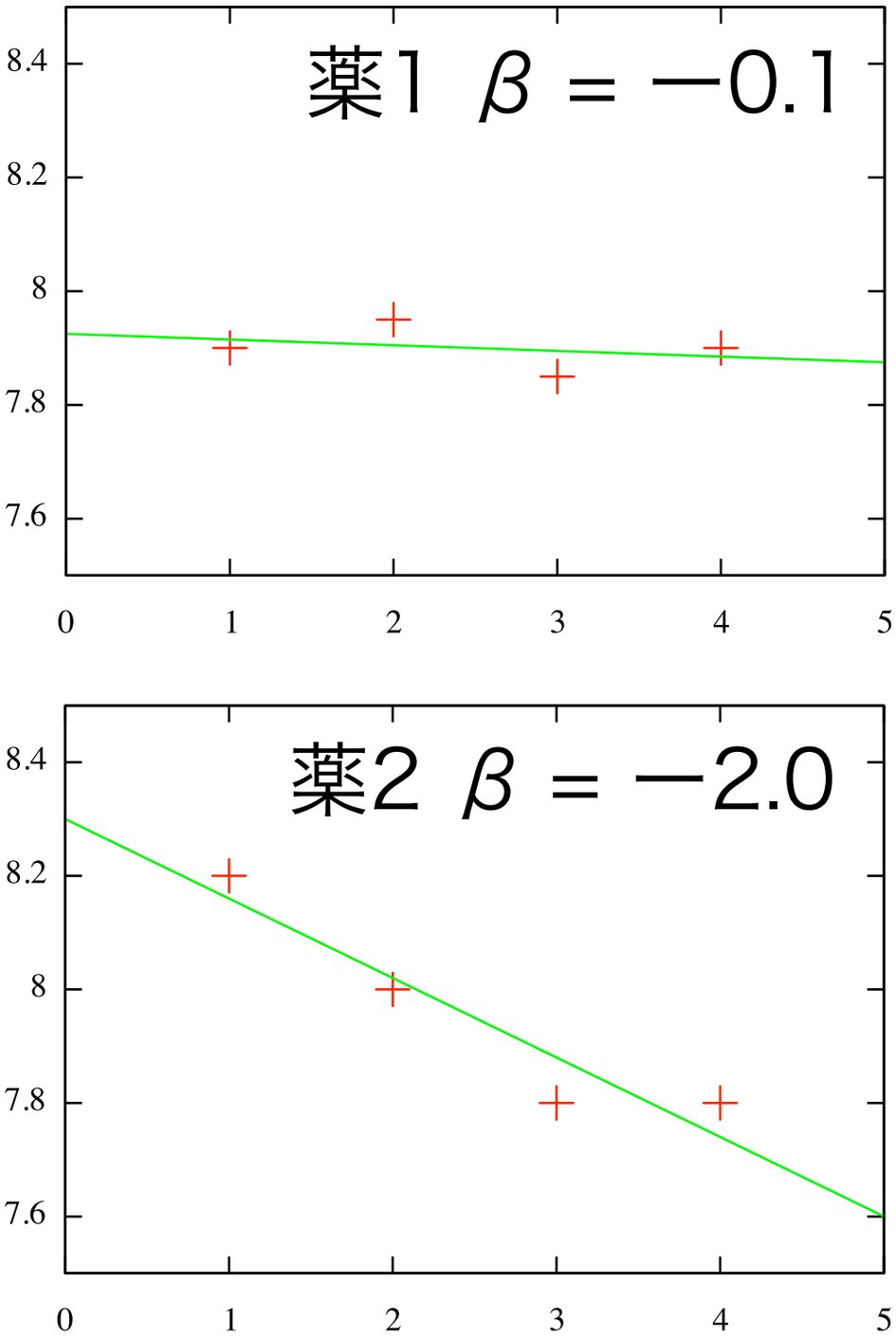
医療機関の情報公開と患者のプライバシー

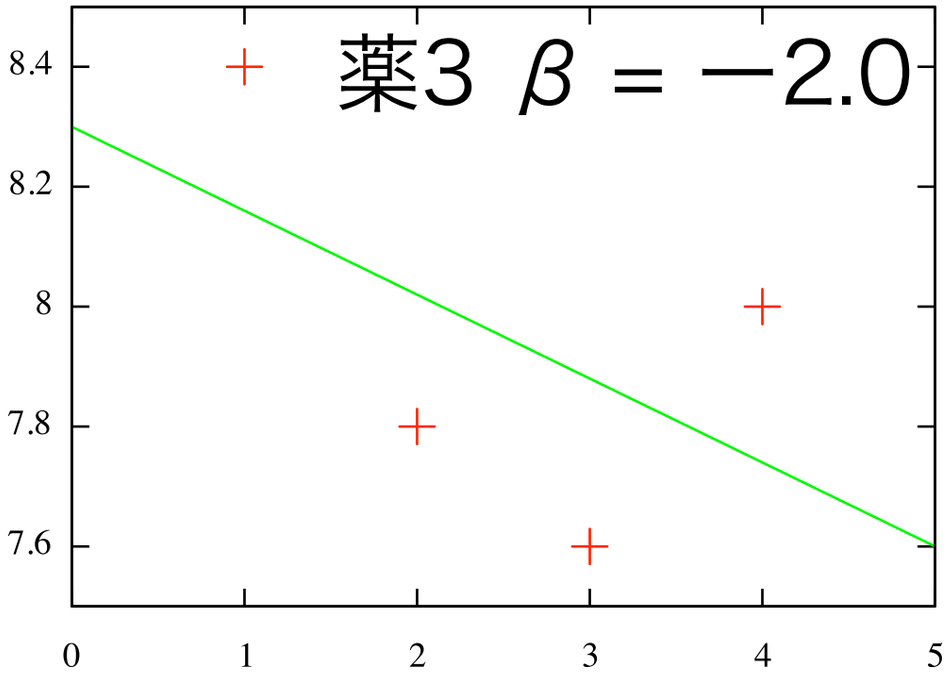
（2010年2月 7日 11:16） | トラックバック（0）

ツイート　チェック　+1 0　いいね！ 0

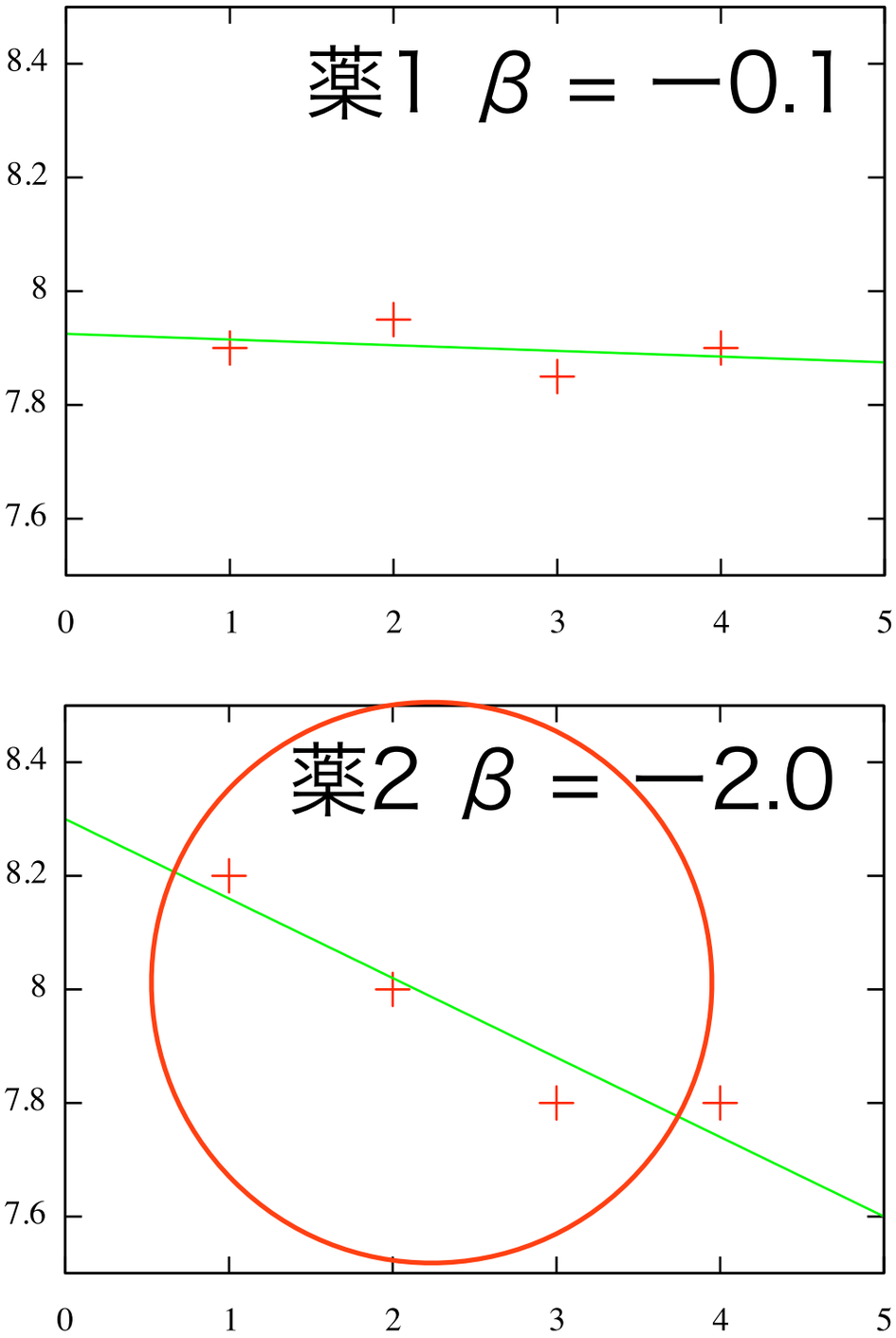
「散らかっている部屋に他人が入れば綺麗になる」―。医療事故や薬害を防止するため、医療機関が保有する情報をできる限り公開すべきだという考えがある。これに対して、患者のプライバシー保護の観点から、投薬や検査の内容、傷病名などの個人情報が他人に漏れる危険性を指摘する声もある。（新井裕充）

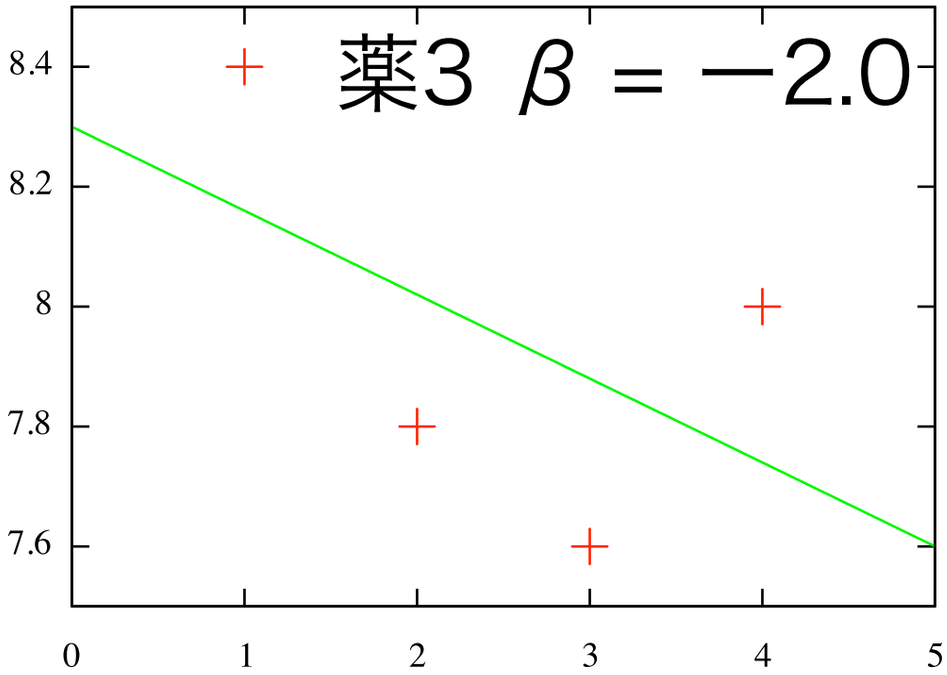
問題定義:秘匿回帰と傾向性の検定



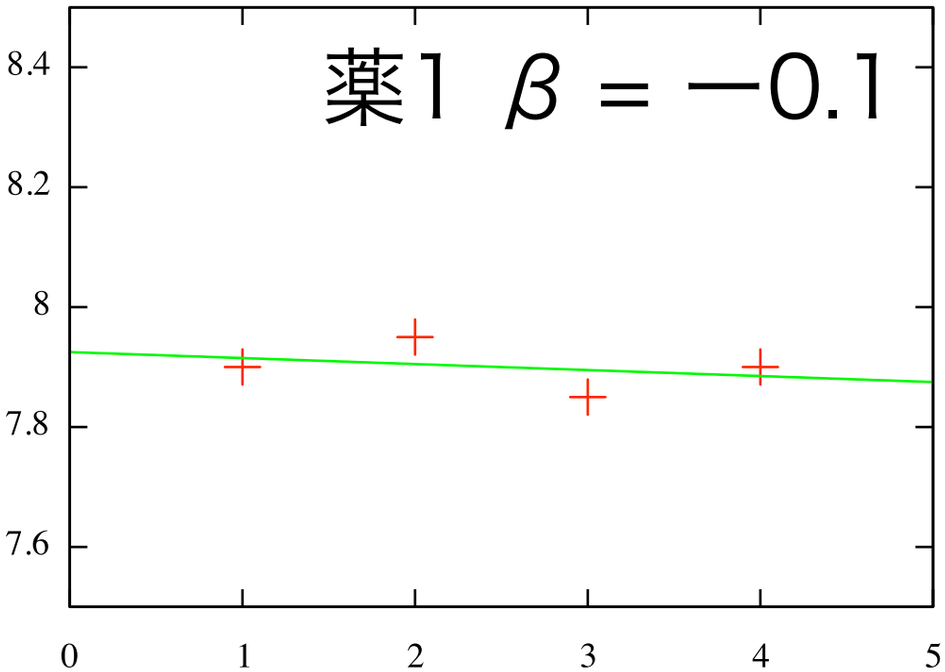


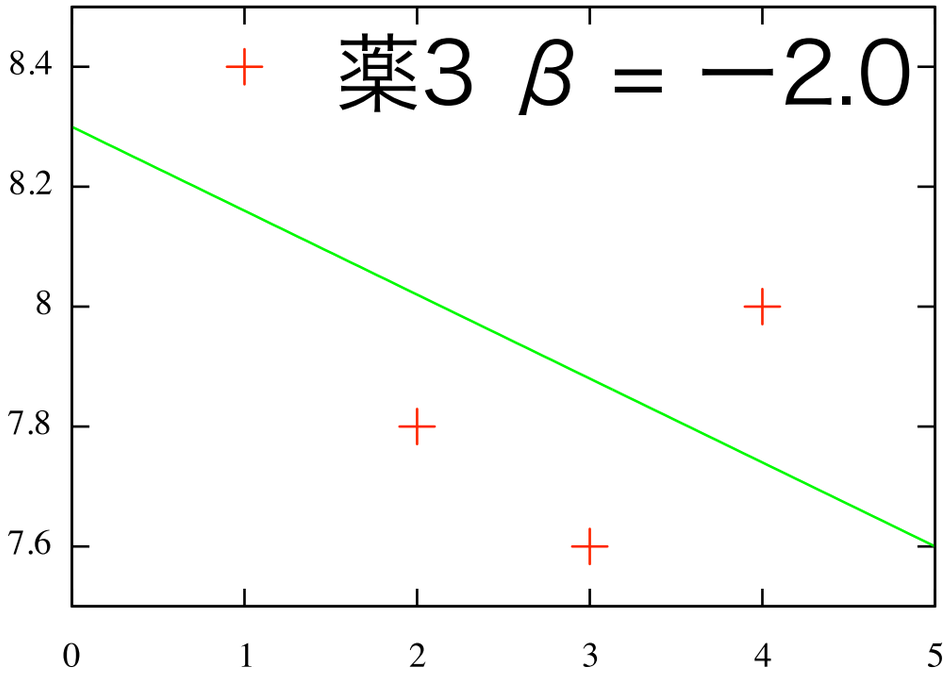
問題定義:秘匿回帰と傾向性の検定

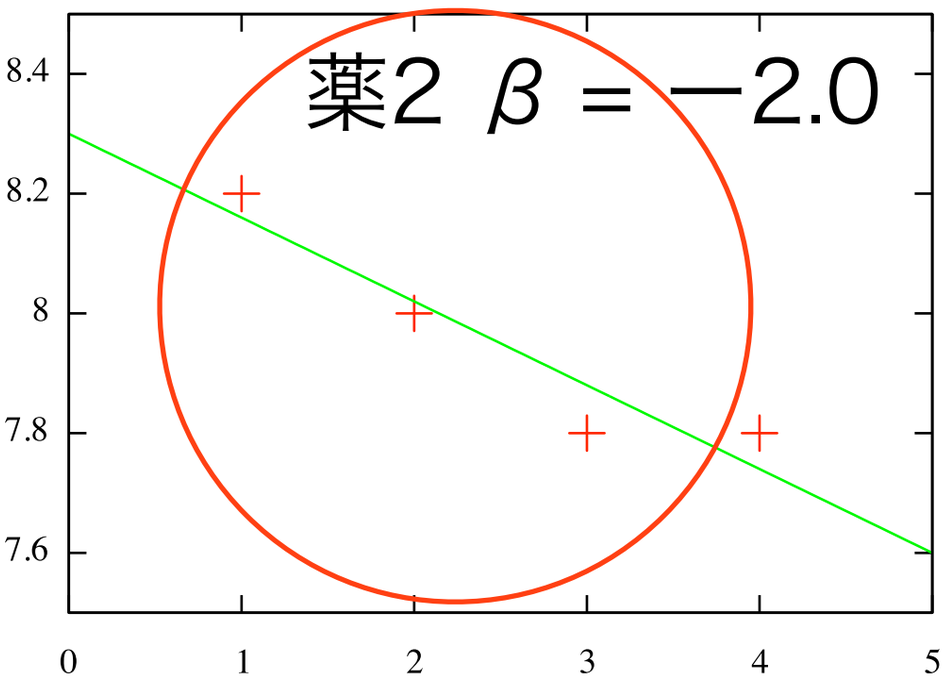




問題定義:秘匿回帰と傾向性の検定



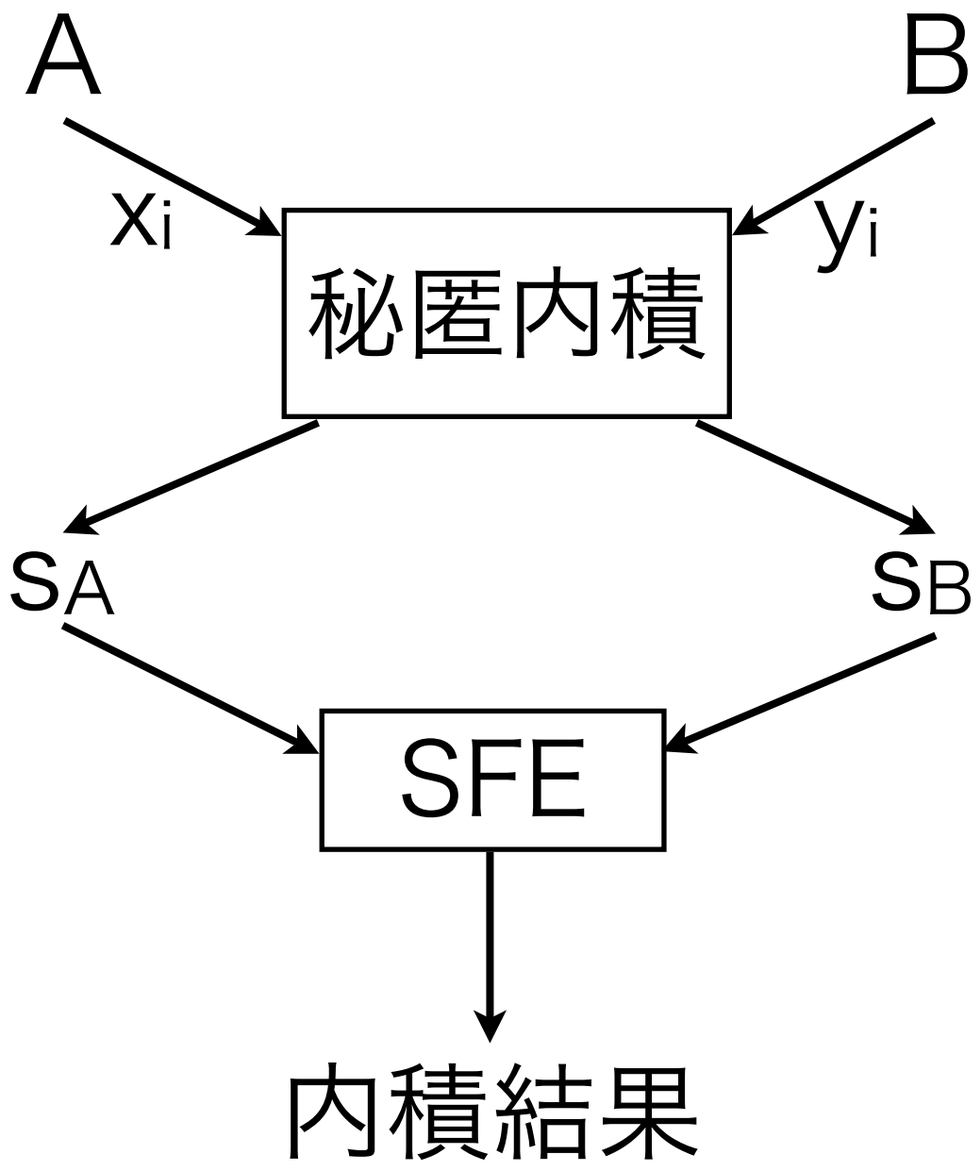




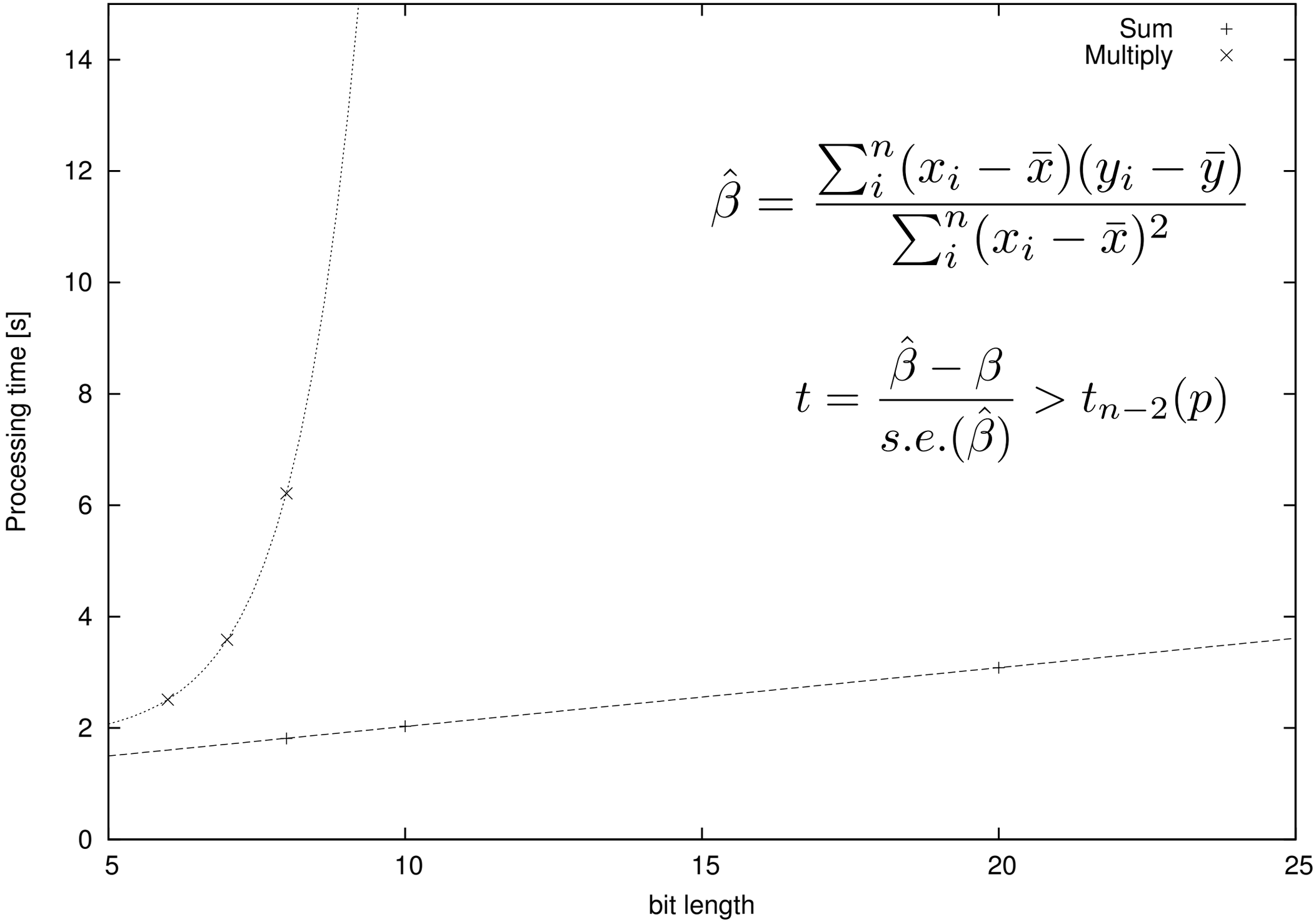
目的：プライバシーを保護して
傾向性を検定する

従来研究 [4], [5]

- 秘匿内積プロトコル[4]
  - 2者間で互いのデータを秘匿したまま内積が可能
    - $s_A + s_B = \mathbf{x} \cdot \mathbf{y}$
- 秘密関数計算(SFE)[5]
  - 互いのデータを秘匿したまま任意の計算が可能



SFE：Fairplayの制約



$$\hat{\beta} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

$$t = \frac{\hat{\beta} - \beta}{s.e.(\hat{\beta})} > t_{n-2}(p)$$

我々のアイデア

目的関数を内積と分散和の形式に変形する

従来	本提案
$\hat{\beta} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$	$\hat{\beta} = \sum_i^n x_i \cdot y_i = \beta_1 + \beta_2$ 秘匿内積とSFEで計算可

本研究の新規性

1. 効率的な秘匿回帰プロトコルの提案

2. 効率的な秘匿回帰検定プロトコルの提案

3. パフォーマンスの評価

1.秘匿回帰プロトコル

$$\hat{\beta} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

$$\hat{\beta} = (x'_1, \cdots, x'_{n+1}) \cdot (y_1, \cdots, y_{n+1}) = \beta_1 + \beta_2$$

SFEで加算するだけ

1.秘匿回帰プロトコル

$$\hat{\beta} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

$$\hat{\beta} = (x'_1, \cdots, x'_{n+1}) \cdot (y_1, \cdots, y_{n+1}) = \beta_1 + \beta_2$$

SFEで加算するだけ

$$\hat{\beta} = \frac{\sum_i^n x_i y_i - (\sum_i^n x_i)(\sum_i^n y_i)/n}{\sum_i^n x_i^2 - (\sum_i^n x_i)^2/n}$$

1.秘匿回帰プロトコル

$$\hat{\beta} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

$$\hat{\beta} = (x'_1, \cdots, x'_{n+1}) \cdot (y_1, \cdots, y_{n+1}) = \beta_1 + \beta_2$$

SFEで加算するだけ

$$\hat{\beta} = \frac{\sum_i^n x_i y_i - (\sum_i^n x_i)(\sum_i^n y_i)/n}{\sum_i^n x_i^2 - (\sum_i^n x_i)^2/n}$$

分子

$$\sum_i^n x_i y_i - (\sum_i^n x_i)(\sum_i^n y_i)/n$$

$$x_{n+1} = -(\sum_i^n x_i) \quad y_{n+1} = (\sum_i^n y_i)/n$$

1.秘匿回帰プロトコル

$$\hat{\beta} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

$$\hat{\beta} = (x'_1, \cdots, x'_{n+1}) \cdot (y_1, \cdots, y_{n+1}) = \beta_1 + \beta_2$$

SFEで加算するだけ

$$\hat{\beta} = \frac{\sum_i^n x_i y_i - (\sum_i^n x_i)(\sum_i^n y_i)/n}{\sum_i^n x_i^2 - (\sum_i^n x_i)^2/n}$$

分子

$$\sum_i^n x_i y_i - (\sum_i^n x_i)(\sum_i^n y_i)/n$$

$$x_{n+1} = -(\sum_i^n x_i) \quad y_{n+1} = (\sum_i^n y_i)/n$$

分母

Aのみで計算可

$$x'_i = \frac{x_i}{\sum_j^n x_j^2 - (\sum_j^n x_j)^2/n}$$

2.秘匿回帰検定プロトコル

$$t = \frac{\hat{\beta} - \beta}{s.e.(\hat{\beta})} > t_{n-2}(p)$$

$$t_2(0.05) = 4.303$$

$$T \equiv t^2_{n-2}(p)$$

$$C \equiv 1 + \frac{t^2_{n-2}(p)}{n-2}$$

$$\frac{T \cdot SS_Y}{n-2} < (\beta_1^2 SS_X + 2SS_X \beta_1 \cdot \beta_2 + SS_X \cdot \beta_2^2) \cdot C$$

2.秘匿回帰検定プロトコル

$$t = \frac{\hat{\beta} - \beta}{s.e.(\hat{\beta})} > t_{n-2}(p)$$

$$t_2(0.05) = 4.303$$

$$T \equiv t^2_{n-2}(p)$$

$$C \equiv 1 + \frac{t^2_{n-2}(p)}{n-2}$$

$$\frac{T \cdot SS_Y}{n-2} < (\beta_1^2 SS_X + 2SS_X \beta_1 \cdot \beta_2 + SS_X \cdot \beta_2^2) \cdot C$$

2.秘匿回帰検定プロトコル

$$t = \frac{\hat{\beta} - \beta}{s.e.(\hat{\beta})} > t_{n-2}(p) \qquad t_2(0.05) = 4.303 \qquad T \equiv t^2_{n-2}(p)$$

$$C \equiv 1 + \frac{t^2_{n-2}(p)}{n-2}$$

$$\frac{T \cdot SS_Y}{n-2} < (\beta_1^2 SS_X + 2SS_X\beta_1 \cdot \beta_2 + SS_X \cdot \beta_2^2) \cdot C$$

$(2SS_X\beta_1, SS_X) \cdot (\beta_2, \beta_2^2) = \gamma_1 + \gamma_2$ を秘匿内積でとく

$$\frac{T \cdot SS_Y}{n-2} - \gamma_2 \cdot C < (\beta_1^2 SS_X + \gamma_1) \cdot C$$

2.秘匿回帰検定プロトコル

$$t = \frac{\hat{\beta} - \beta}{s.e.(\hat{\beta})} > t_{n-2}(p) \qquad t_2(0.05) = 4.303 \qquad T \equiv t^2_{n-2}(p)$$

$$C \equiv 1 + \frac{t^2_{n-2}(p)}{n-2}$$

$$\frac{T \cdot SS_Y}{n-2} < (\beta_1^2 SS_X + 2SS_X\beta_1 \cdot \beta_2 + SS_X \cdot \beta_2^2) \cdot C$$

$(2SS_X\beta_1, SS_X) \cdot (\beta_2, \beta_2^2) = \gamma_1 + \gamma_2$ を秘匿内積でとく

$$\frac{T \cdot SS_Y}{n-2} - \gamma_2 \cdot C < (\beta_1^2 SS_X + \gamma_1) \cdot C$$

2.秘匿回帰検定プロトコル

$$t = \frac{\hat{\beta} - \beta}{s.e.(\hat{\beta})} > t_{n-2}(p) \qquad t_2(0.05) = 4.303 \qquad T \equiv t^2_{n-2}(p)$$

$$C \equiv 1 + \frac{t^2_{n-2}(p)}{n-2}$$

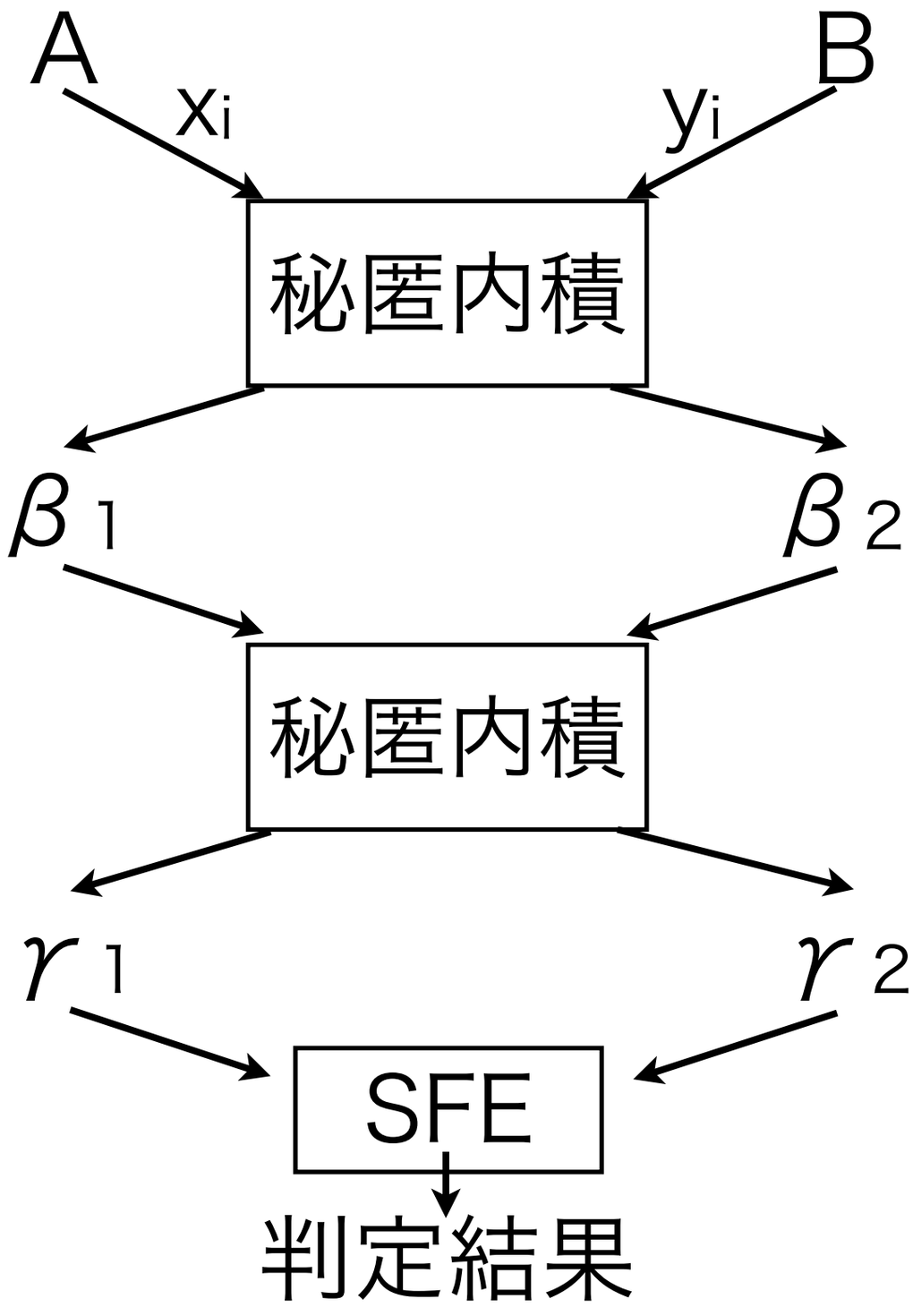
$$\frac{T \cdot SS_Y}{n-2} < (\beta_1^2 SS_X + 2SS_X\beta_1 \cdot \beta_2 + SS_X \cdot \beta_2^2) \cdot C$$

$(2SS_X\beta_1, SS_X) \cdot (\beta_2, \beta_2^2) = \gamma_1 + \gamma_2$　を秘匿内積でとく

$$\frac{T \cdot SS_Y}{n-2} - \gamma_2 \cdot C < (\beta_1^2 SS_X + \gamma_1) \cdot C$$

Bの入力＜Aの入力　の判定をするだけ

処理の流れ



アルゴリズム

A

B

$$E(x'_1), \cdots, E(x'_{n+1}) \longrightarrow \{E(x'_1)^{y_1}+, \cdots, +E(x'_{n+1})^{y_{n+1}}\}/E(\beta_2) = g$$

$$\beta_1 = D(g) \longleftarrow$$

アルゴリズム

A

B

$$E(x'_1), \cdots, E(x'_{n+1}) \longrightarrow \{E(x'_1)^{y_1}+, \cdots, +E(x'_{n+1})^{y_{n+1}}\}/E(\beta_2) = g$$

$$\beta_1 = D(g)$$

$$E(2SS_X\beta_1), E(SS_X) \longrightarrow \{E(2SS_X\beta_1)^{\beta_2} + E(SS_X)^{\beta_2^2}\}/E(\gamma_2) = f$$

$$\gamma_1 = D(f)$$

アルゴリズム

A

B

$E(x'_1), \cdots, E(x'_{n+1})$ ⟶ $\{E(x'_1)^{y_1}+, \cdots, +E(x'_{n+1})^{y_{n+1}}\}/E(\beta_2) = g$

$\beta_1 = D(g)$

$E(2SS_X\beta_1), E(SS_X)$ ⟶ $\{E(2SS_X\beta_1)^{\beta_2} + E(SS_X)^{\beta_2^2}\}/E(\gamma_2) = f$

$\gamma_1 = D(f)$

$(\beta_1^2 SS_X + \gamma_1) \cdot C$ ⟷ (SFEで比較) $\frac{T \cdot SS_Y}{n-2} - \gamma_2 \cdot C$

アルゴリズム

A　　B

$E(x'_1), \cdots, E(x'_{n+1})$ ⟶ $\{E(x'_1)^{y_1}+, \cdots, +E(x'_{n+1})^{y_{n+1}}\}/E(\beta_2) = g$

$\beta_1 = D(g)$ ⟵

秘匿回帰

$E(2SS_X\beta_1), E(SS_X)$ ⟶ $\{E(2SS_X\beta_1)^{\beta_2} + E(SS_X)^{\beta_2^2}\}/E(\gamma_2) = f$

$\gamma_1 = D(f)$ ⟵

秘匿回帰検定

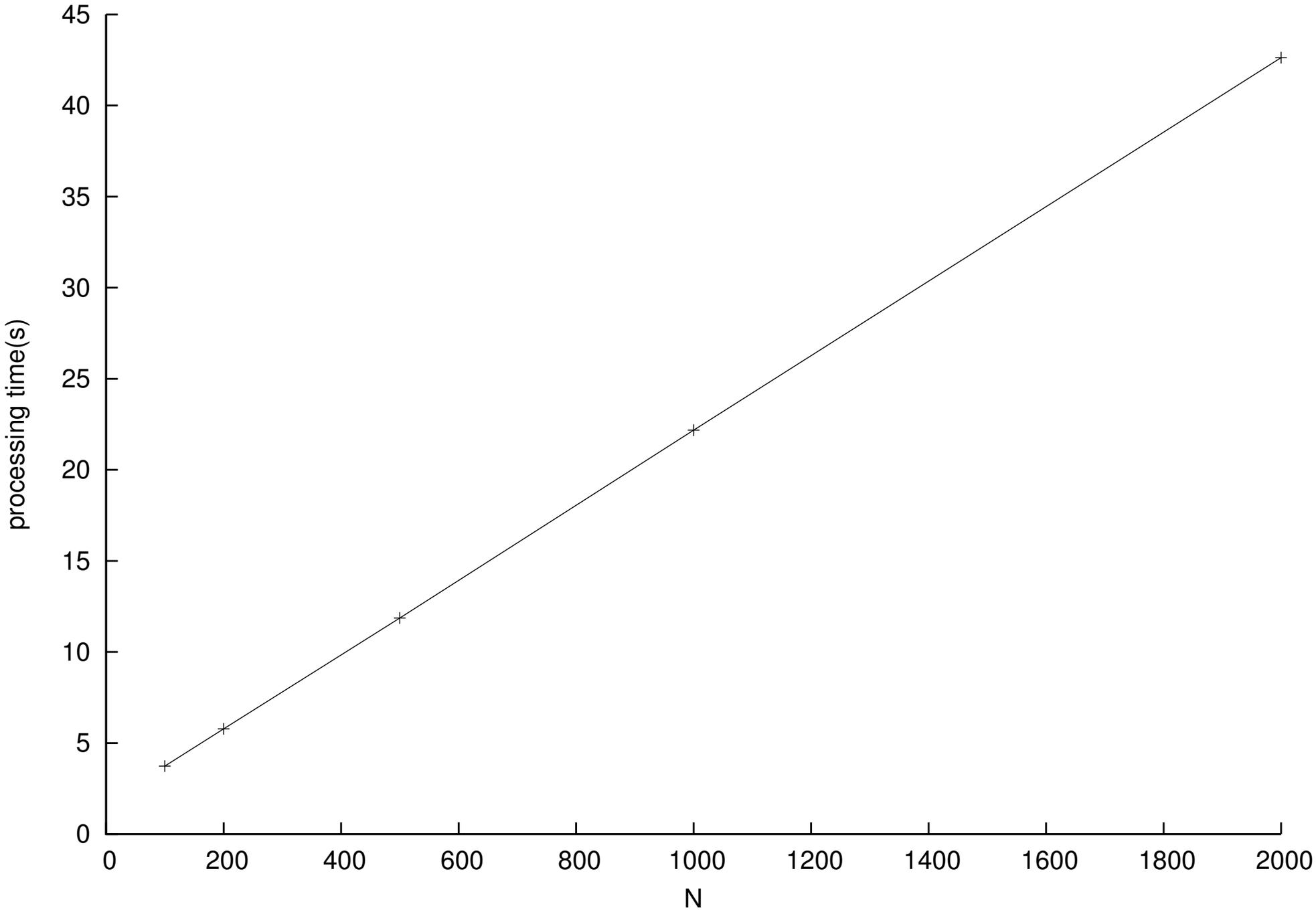
$(\beta_1^2 SS_X + \gamma_1) \cdot C$ ⟷ SFEで比較 ⟷ $\dfrac{T \cdot SS_Y}{n-2} - \gamma_2 \cdot C$

パフォーマンス

プロトコル	暗号化の回数
秘匿回帰	n+2回
秘匿回帰検定	3回

n = 14万件の時，2855.5[s]で
処理を行うことができる

パフォーマンス


processing time(s)
N
0
5
10
15
20
25
30
35
40
45
0
200
400
600
800
1000
1200
1400
1600
1800
2000

まとめ

- 秘匿内積プロトコルと秘密関数計算を用いた秘匿回帰プロトコル，及び秘匿回帰検定プロトコルを提案した
- 出力は回帰直線の傾き β^と，それが有意か否かの判定結果のみである
- データ数 n = 14万件の時，2855.5[s]で処理を行うことができる