高知工科大学 経済・マネジメント学群

計量経済学

9. 回帰分析による統計的推測 II
仮説を検証する

やない ゆうき
矢内 勇生

https://yukiyanai.github.io

yanai.yuki@kochi-tech.ac.jp

今日の目標

- 回帰分析で「統計的検定」と「統計的推定」を行うための準備を整える
  - 母集団における回帰直線と標本の回帰直線を区別する
  - 回帰分析の帰無仮説と対立仮説を理解する

母集団の回帰直線と
標本の回帰直線

回帰分析による推定

- データから作った散布図への直線（平面）の当てはめは、標本データの要約
- 興味があるのは母集団の特徴

★ どのような方法で、標本から母集団を推定する？

統計モデルをつくる

- 自分が観察しているデータが生み出される過程をモデル化する
  - データ生成過程（data generating process; DGP）
  - モデル：目的に応じた現象の単純化
    - 本質的に「正しくない」
    - 「正しいかどうか」ではなく、「役に立つかどうか」で評価する

"All models are wrong,

but some are useful."

–George E. P. Box

Cf. Box, George. 1976. "Science and Statistics." *Journal of the American Statistical Association,* 71(356): 791-799.

単回帰

- 母集団における単回帰

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

  - α, β：パラメタ, 母数（推定の対象）
  - ϵ：誤差 (error)
    - 説明変数以外で応答変数に影響を与えるもの
    - 平均すると0

誤差をモデル化する

- 誤差 ϵ の分布を以下のように**仮定**する
  - $\epsilon_i \sim \text{Normal}(0, \sigma)$
    - 誤差の平均は 0
    - 誤差は、1つの正規分布から生み出される
      - 標準偏差は、i によらず一定

単回帰モデル

- 単回帰モデル：単回帰が想定するDGP
  - まず、X_i $(i = 1,2,\ldots)$の値が決まる
  - 次に、Y_i $(i = 1,2,\ldots)$の値が以下のように決まる

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta X_i$$

単回帰モデルの書き換え

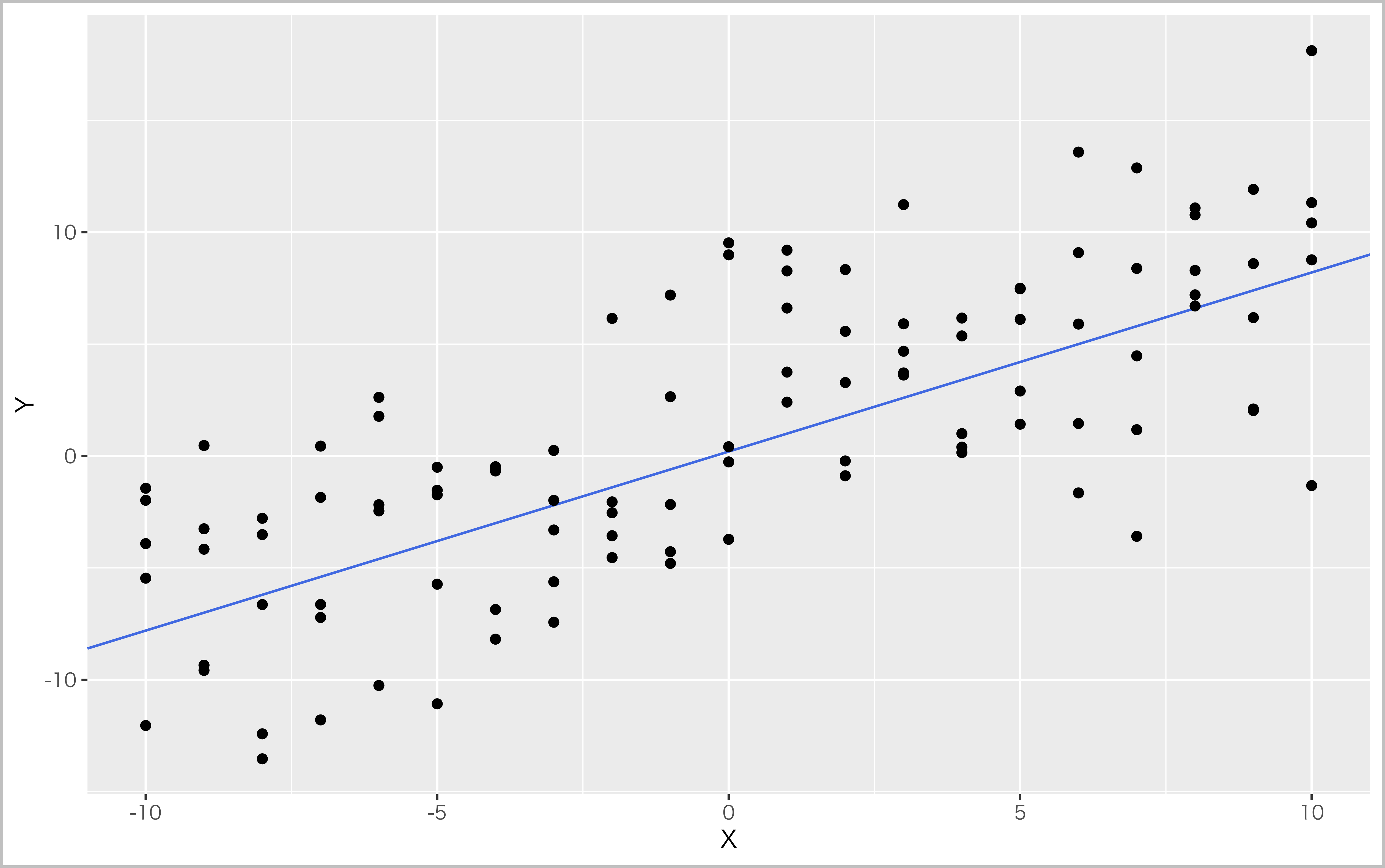
- 以下のような表記が使われることも多い（意味はどれも同じ）
  - 別表記（1）

$$Y_i \sim \text{Normal}(\alpha + \beta X_i,\ \sigma)$$

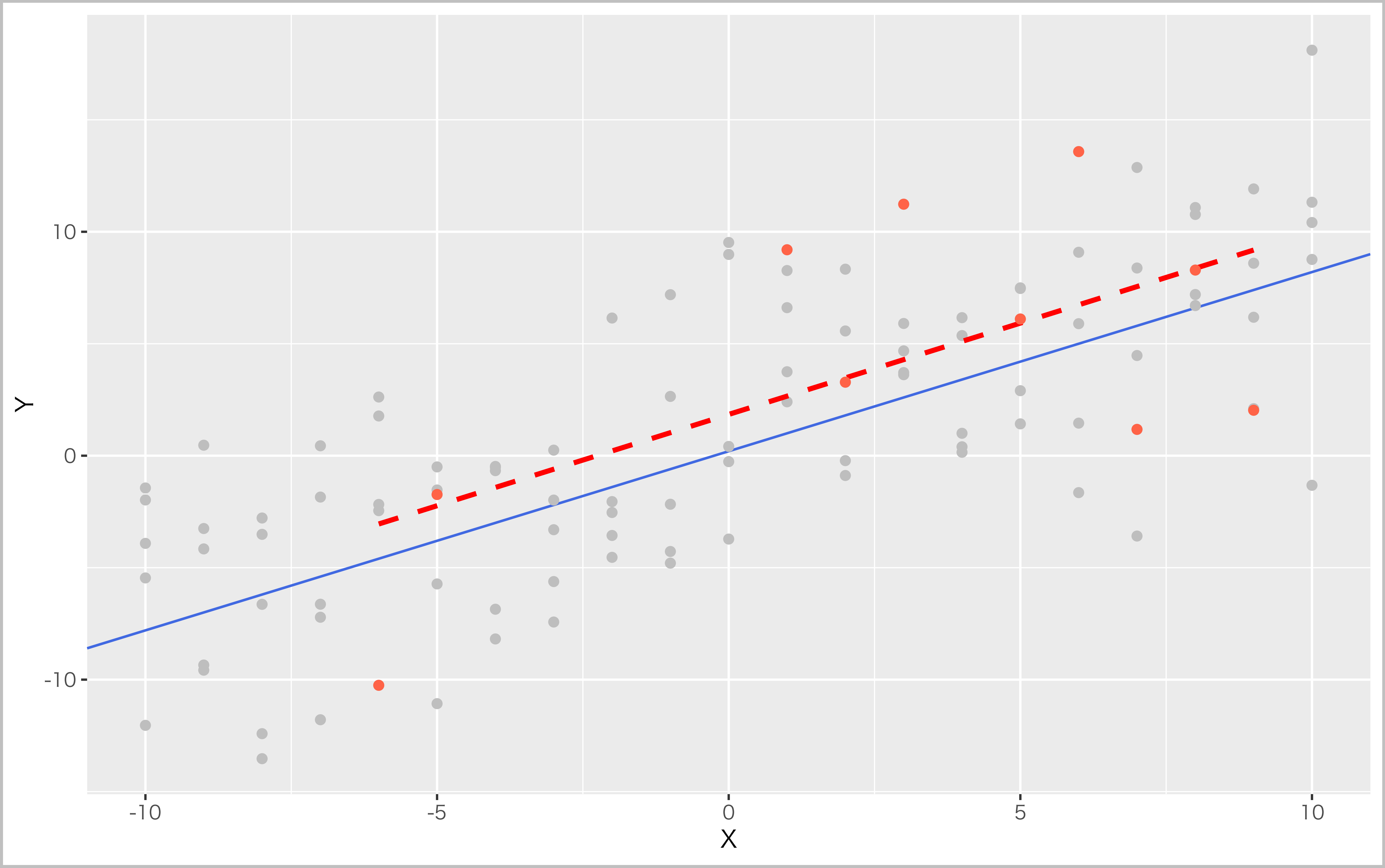
  - 別表記 (2)

$$Y_i = \alpha + \beta X_i + \epsilon_i$$
$$\epsilon_i \sim \text{Normal}(0, \sigma)$$

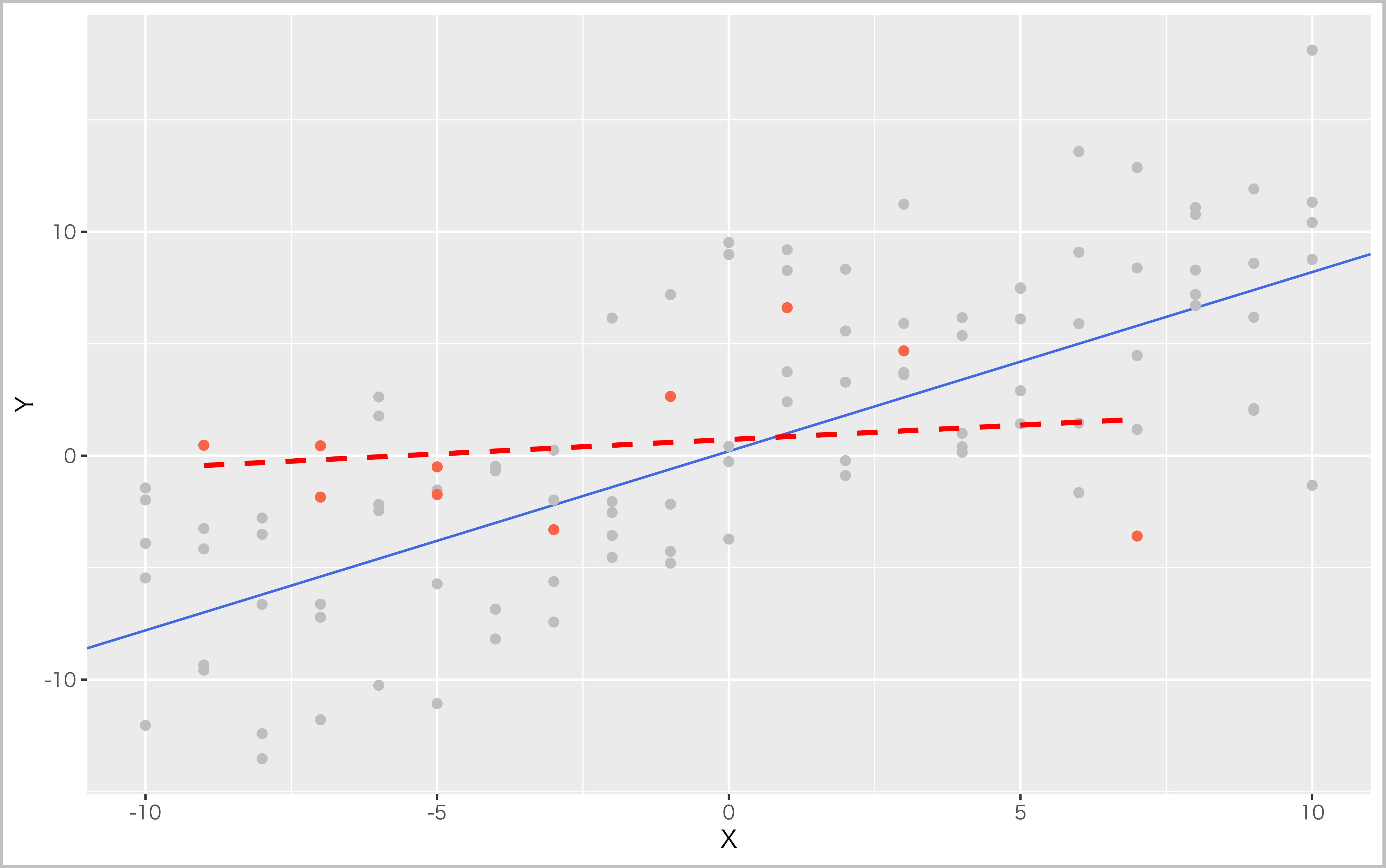
母集団の回帰直線



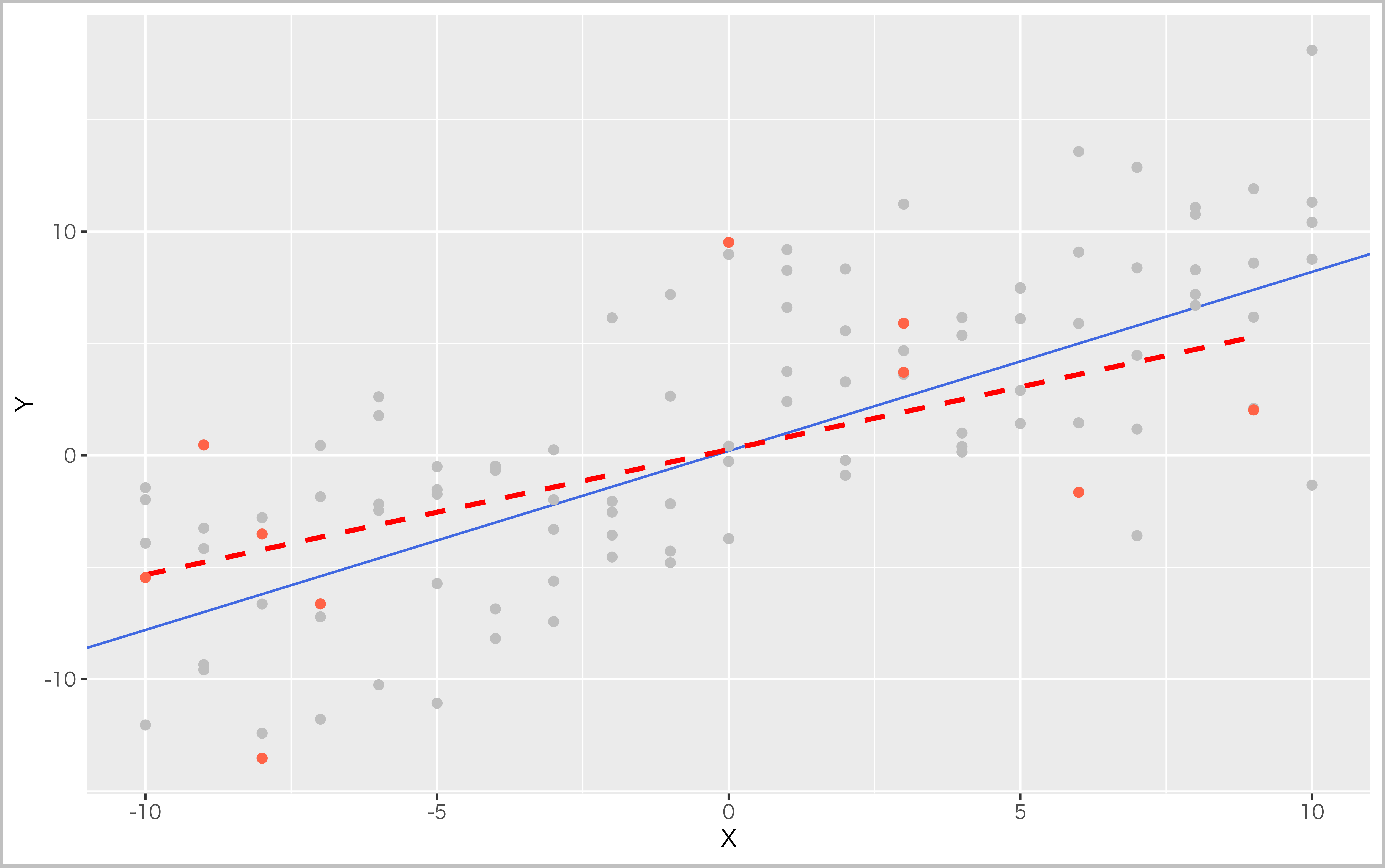
標本の回帰直線 (1)



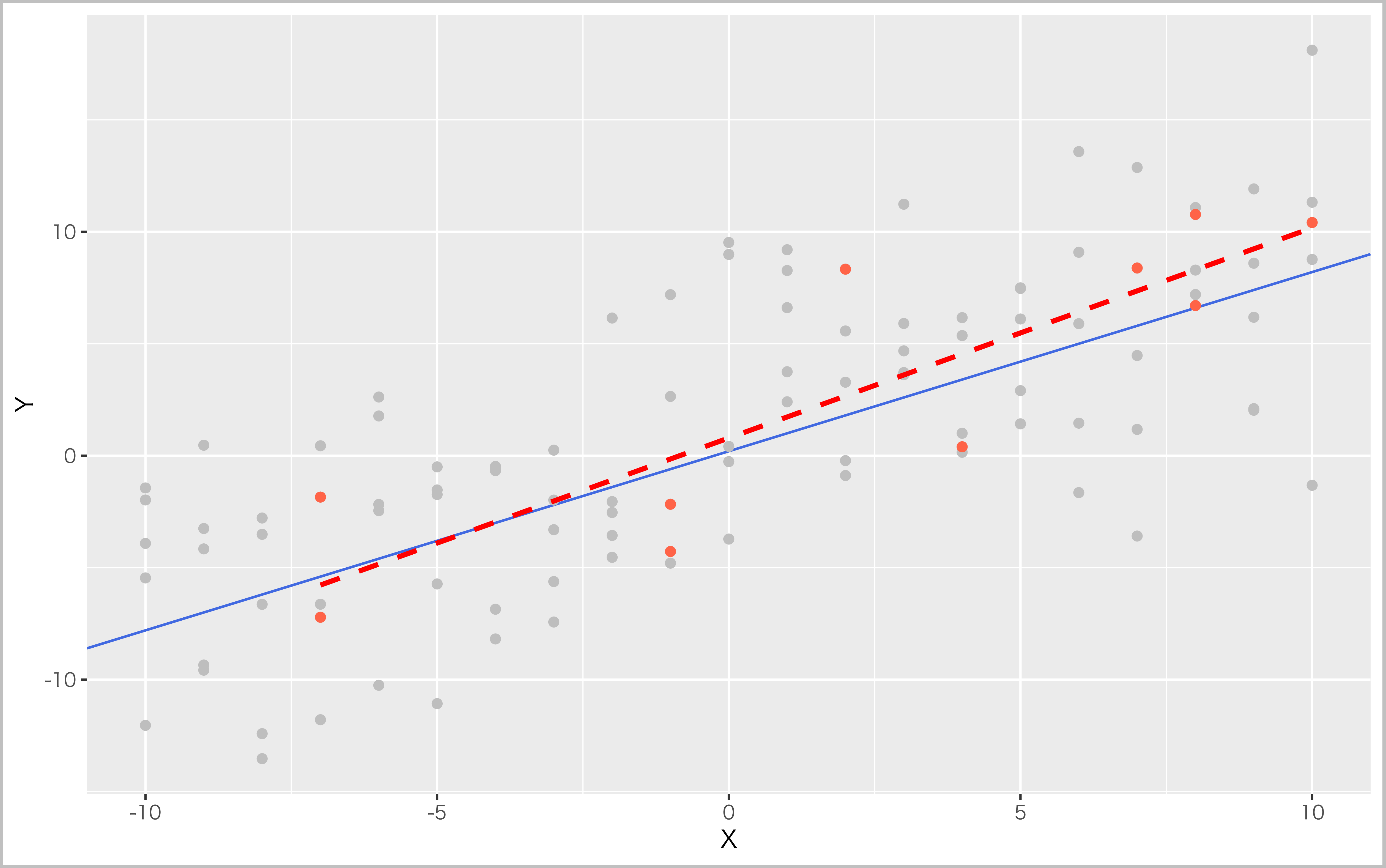
標本の回帰直線 (2)



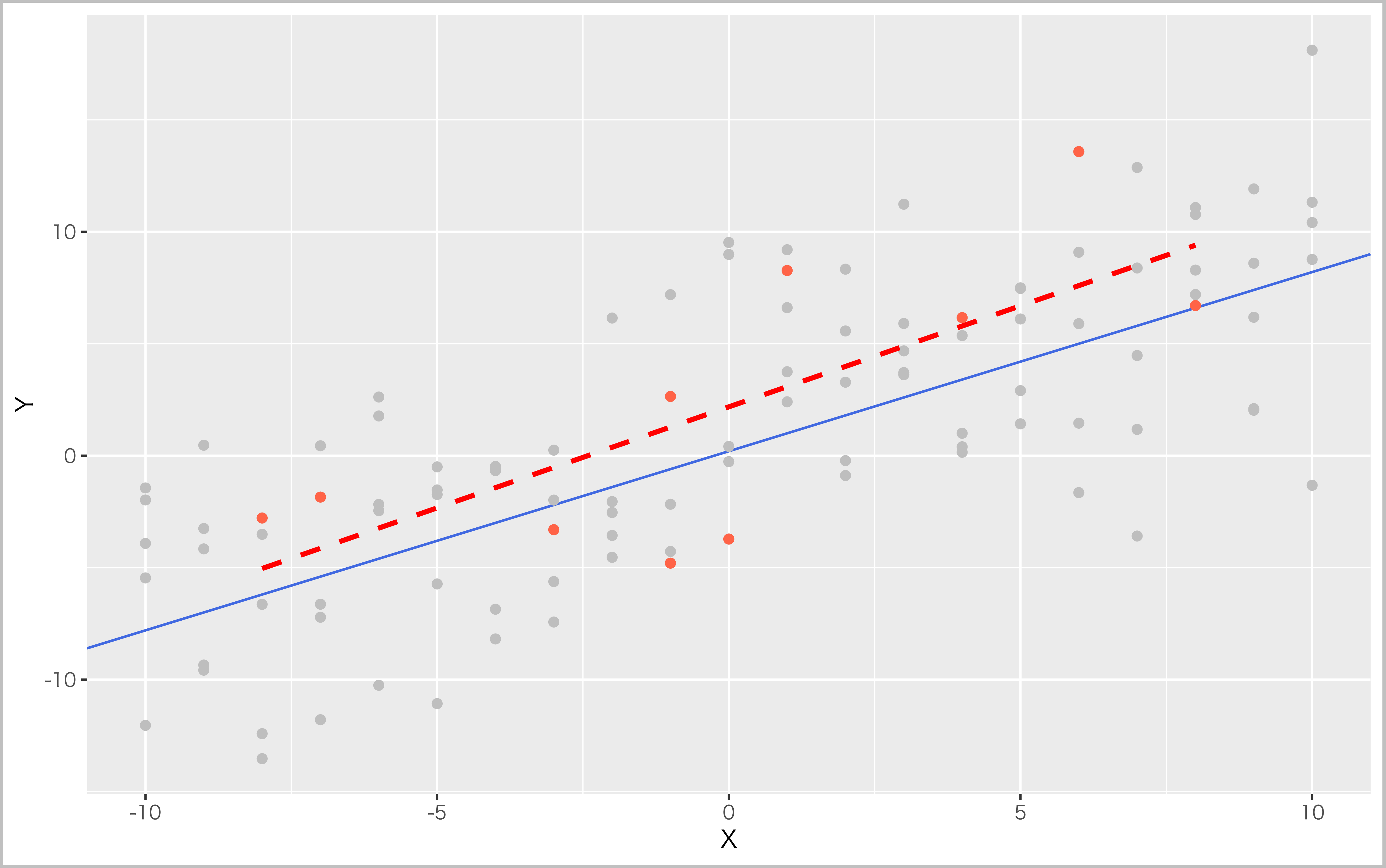
標本の回帰直線 (3)



標本の回帰直線 (4)



標本の回帰直線 (5)



最小二乗法による母数の推定：単回帰の場合

- 標本データを使い、最小二乗法よって求めた回帰係数 a, b は、単回帰モデルに登場する α, β の点推定値

- 最小二乗推定量は以下の望ましい性質をもつ
  - 不偏性 (unbiasedness)：$\mathbb{E}[a] = \alpha, \quad \mathbb{E}[b] = \beta$
  - 一致性 (consistency)：標本サイズを無限大にすると、推定値は母数に一致する

重回帰

- 母集団における重回帰

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK} + \epsilon_i$$

- β_k ：パラメタ, 母数（推定の対象）、$k = 0, 1, 2, \ldots, K$
- ϵ ：誤差
  - $\epsilon_i \sim \text{Normal}(0, \sigma)$

重回帰モデル

- 重回帰モデル：重回帰が想定するDGP
  - まず、X_{ik} $(i = 1,2,\ldots; k = 0,2,\ldots,K)$の値が決まる
  - 次に、Y_i $(i = 1,2,\ldots)$の値が以下のように決まる

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_K X_{iK}$$

最小二乗法による母数の推定：重回帰の場合

- 標本データを使い、最小二乗法によって求めた回帰係数 $b_0, b_1, \dots, b_K$ は、$\beta_0, \beta_1, \dots, \beta_K$ の推定値である
    - 不偏性：$\mathbb{E}[b_k] = \beta_k$（$k = 0,1,2,\dots, K$）
    - 一致性

回帰分析の
帰無仮説と対立仮説

何のために回帰分析を行うのか

- 目的：理論（理論仮説）を検証したい
  - そのために作業仮説を用意する
  - 回帰分析で検証可能な作業仮説を用意する
    - 1つの応答変数
    - 1つ以上の説明変数
    - 説明変数が応答変数に与える影響についての仮説
      - 例：「X が Y を増加させる」

帰無仮説と対立仮説

- 帰無仮説：「説明変数は応答変数に影響を与えない」
- 対立仮説：「説明変数が応答変数に影響する」
  - 自分が「正しい」ことを示したい理論の作業仮説を対立仮説にする
- 統計的検定（方法は次回説明する）で帰無仮説が棄却されたとき、「作業仮説が統計的に正しい」と判断する
  - 作業仮説が正しいと考えられるので、操作化がうまくできていれば、理論仮説の蓋然性が高まる
    - 操作化（作業仮説と理論仮説の類似度）が重要

単回帰の場合

- モデル：$Y_i \sim \text{Normal}(\alpha + \beta X_i,\ \sigma)$

- 検証する仮説
  - 帰無仮説：$\beta = 0$
  - 対立仮説：$\beta \neq 0$

重回帰の場合 (1) 包括的検定

- モデル：$Y_i \sim \text{Normal}(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_K X_{iK}, \sigma)$
- 検証する仮説のパタン1
  - 帰無仮説：$\beta_1 = \beta_2 = \cdots = \beta_K = 0$
  - 対立仮説：「$\beta_1, \beta_2, \ldots, \beta_K$ のうち、少なくとも1つについて $\beta_k \neq 0$」

重回帰の場合 (2) 個別的検定

- モデル：$Y_i \sim \text{Normal}(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_K X_{iK}, \sigma)$
- 検証する仮説のパタン2

	β_1の仮説	β_2の仮説		β_Kの仮説
帰無仮説：	$\beta_1 = 0$	$\beta_2 = 0$	$\dots$	$\beta_K = 0$
対立仮説：	$\beta_1 \neq 0$	$\beta_2 \neq 0$		$\beta_K \neq 0$

  - 実際は、すべての k について仮説を立てて検証するわけではなく、理論における「原因」とみなされるものについてのみ個別に仮説を検証する

「影響がない」を検証する？？？

- 通常、「影響がない」は帰無仮説
  - 「影響がない」を対立仮説にすると、帰無仮説「影響がある」は棄却できない（検証する対象が無限にある）
  - 「影響がない」という帰無仮説を棄却できなくても、それは「影響がない」ことを意味しない
    - 「影響がある」という証拠が見つからないだけ
    - 「証拠の不在」は「不在の証拠」ではない！

★「影響がない」ことを主張する理論は、（これまで勉強してきた）統計的分析では検証不可能

次回

仮説を検証する (2)