高知工科大学 経済・マネジメント学群

計量経済学

11. 回帰分析の応用

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

- 重回帰分析を理解する
  - 重回帰分析における「コントロール」の意味を理解する
  - 重回帰分析で生じ得るバイアスについて理解する

因果関係を知りたいとき

- 実験：因果関係を調べるための最善策
- しかし、どんな問題でも実験できるわけではない
- 調査・観察・観測データ（observational data）に頼るしかない

統計的因果推論

- 比較的大きな標本サイズのデータを使って理論を検証する
- 自然実験 (natural experiments)
- 準実験 (quasi-experiments)
  - 操作変数法 (instrumental variable method)
  - 回帰不連続デザイン (regression discontinuity design)
  - 差分の差分法 (difference-in-differences [DiD])
- 条件付け
  - **統制変数を伴う回帰分析：重回帰分析**
  - パネルデータ分析

重回帰分析

何を検証する？ (1)

- 検証したい理論：「X が Y を引き起こす」
- この関係は決定的 (deterministic) か？
- 例：「教育が政治参加を促す」
- 変数の操作化
    - 教育：大卒か否か
    - 政治参加：国政選挙での投票参加

何を検証する？ (2)

- 決定論的理論: 「大学の学位は国政選挙で投票するための必要十分条件である」
- 大卒だが投票していない人を「1人だけ」見つけたらどうする？
  - 決定論では、理論を否定する必要がある
  - 社会科学の理論として、それでいいの？

何を検証する？ (3)

- ほとんどの場合、私たちが検証したいのは、**確率的**理論：「大学に行くと、国政選挙での投票確率が上がる傾向にある」
- 理論的予測に合致しない人を少数見つけても、理論の否定にはならない
  - 十分大きな標本サイズで、少人数が理論に合致しなくても、大きな傾向に影響はない
- 大卒と大卒未満の2つのグループで、平均すると大卒の方が投票率が高いことを示せればよい

条件付け：「他の条件が等しければ」

- 2つ（以上）のグループを比較する
- 社会科学では、以下のような異質な個体を比較する
  - 人間
  - 国家
- 通常、調べている要因以外の「他の条件」は等しくない！
  - 大卒と大卒未満では、「親の年収」に違いがあるかもしれない
- 「他の条件が等しい(ceteris paribus)」状況で比較したい

重回帰分析

- 「他の条件が等しい」状況を作り出すため、重回帰分析を利用する
- 検証したい理論：「X が Y を上昇（減少）させる」
- 応答変数： Y
- 主な説明変数: X
- 統制（コントロール）変数: Z （複数あってよい）

コントロール・条件付け

- 変数 z を統制（コントロールする）：Z は統制（コントロール）変数と呼ばれることも
- Z は複数あってもよい: $Z_1, Z_2, \ldots$
- 私たちが比較したい個体が様々な面で異質なとき、複数の要因を統制する必要がある
- 複数の要因を統制するためには、大きな標本サイズが必要
  - $N = 2$ で一人は女性、もう一人は男性のとき、性別を統制できる？

重回帰モデル

- 理論的関心：X が Y に影響するかどうか
  - 問題：Z がセレクションバイアスを引き起こす
- 重回帰モデル

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

- 検証する仮説
  - 帰無仮説 $\beta = 0$ vs. 対立仮説 $\beta \neq 0$
  - γ は検証の対象ではない！

重回帰の結果の解釈

- 理論的関心：X が Y に影響するかどうか
- 重回帰モデル：$Y_i \sim \text{Normal}(\alpha + \beta X_i + \gamma Z_i, \sigma)$
- β の推定値：Z の影響を取り除いたとき、X 1単位の増加がY を何単位増加させるか
- γ の推定値：意味なし！
  - Z がコントロール変数なら、γ の意味を解釈しようとしてはいけない！

重回帰の結果の解釈（2）

- 理論的関心：Y に影響を与える変数は何か？
- 重回帰モデル：$Y_i \sim \text{Normal}(\alpha + \beta X_i + \gamma Z_i, \sigma)$
- β の推定値：Z の影響を取り除いたとき、X 1単位の増加がY を何単位増加させるか
- γ の推定値：X の影響を取り除いたとき、Z 1単位の増加がY を何単位増加させるか

因果推論のための
重回帰分析

どの変数を統制する？

- 重回帰で使う変数は何？
  - 応答変数（理論における結果）：絶対に必要
  - 主な説明変数（理論における原因）：絶対に必要
  - 統制変数（共変量）：必要かもしれない（ほとんどの場合、必要）
    - どの変数を統制する？
    - いくつの変数を統制する？

重回帰分析の共変量

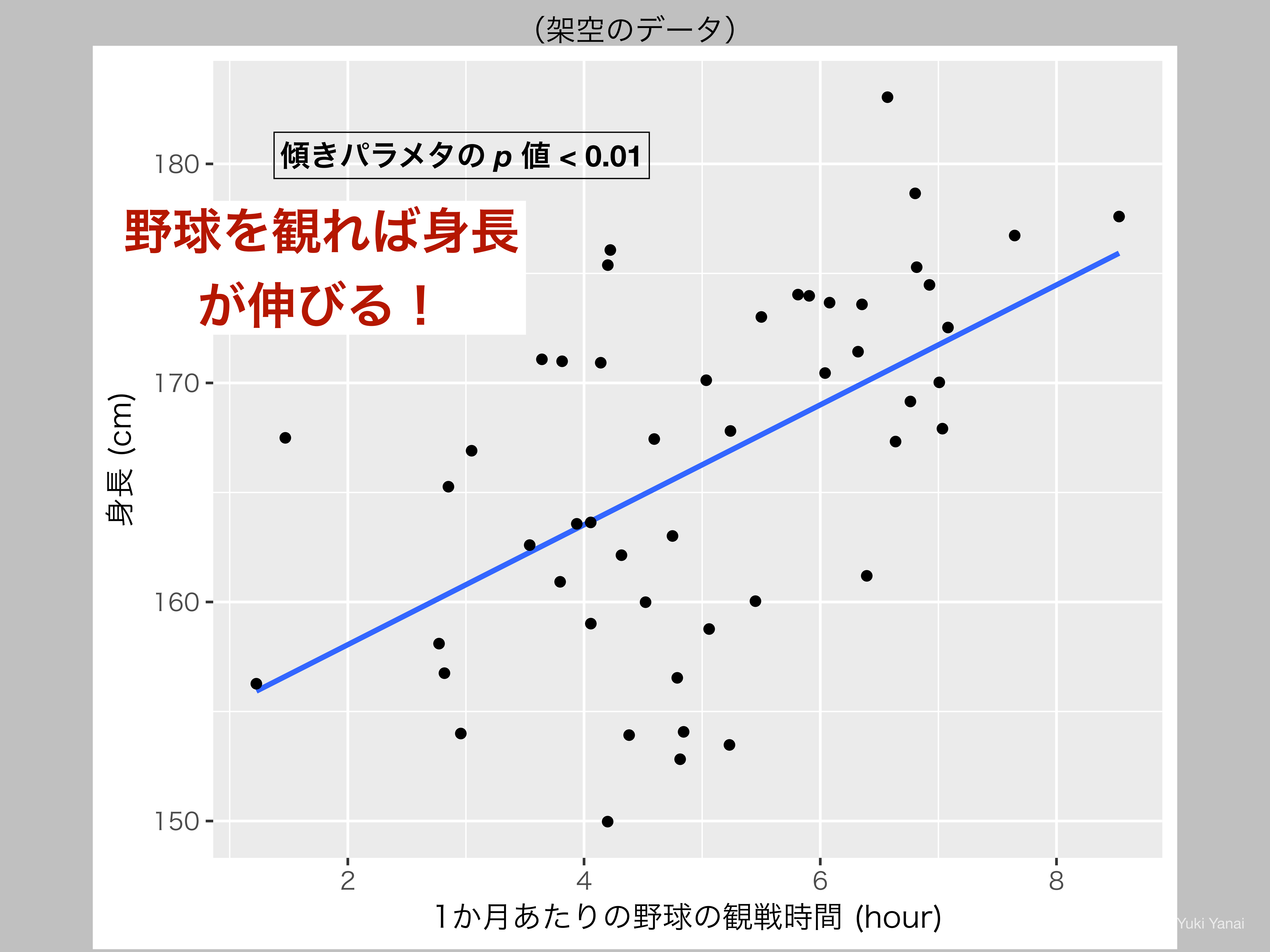
- 重回帰分析には、統制すべき変数と、統制すべきでない変数がある
  - 統制すべき変数
    - 交絡変数
  - 統制すべきでない変数
    - 処置後変数
      - 媒介変数
      - 合流点

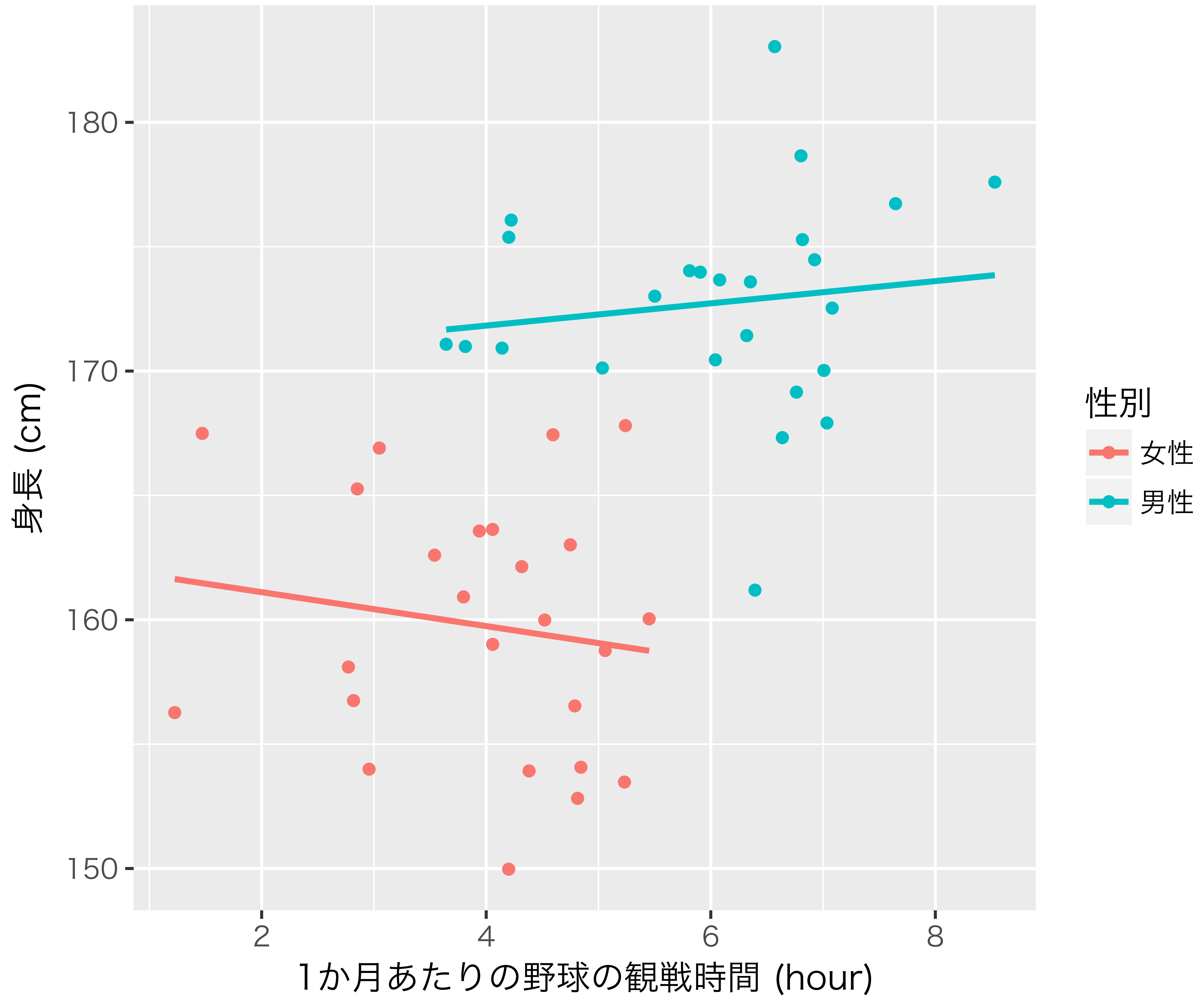
回帰分析

- 線形回帰モデル（最小二乗法で推定）を考える
  - Y： 応答変数
  - X：主な説明変数（原因と考えられる変数）
  - Z：統制変数（共変量）
- 私たちが知りたい（推定する）のは、X が Y に与える影響
  - X の Y に対する因果効果：X が1単位増加したとき、Y は何単位増加するか？
  - この効果を推定する：係数の点推定値と信頼区間

回帰分析の例

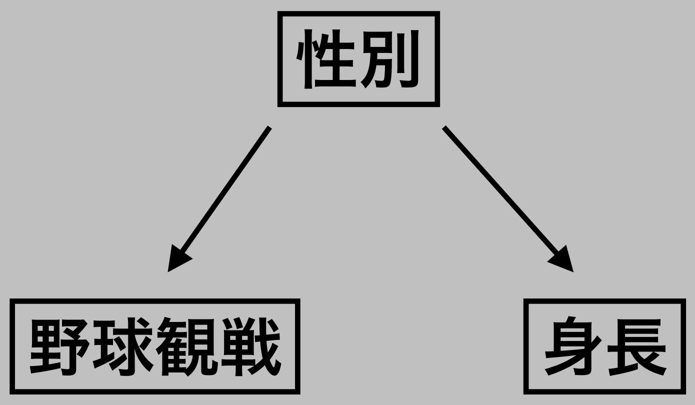
- 身長とプロ野球の観戦時間の関係は？
  - プロ野球の観戦時間は身長を伸ばす？
- 理論的に考えると、おそらく No!
- しかし、回帰分析をすると…
  - Yes ???


（架空のデータ）
傾きパラメタの p 値 < 0.01
野球を観れば身長
が伸びる！
身長 (cm)
180
170
160
150
2
4
6
8
1か月あたりの野球の観戦時間 (hour)
Yuki Yanai



身長 (cm)
180
170
160
150
2
4
6
8
1か月あたりの野球の観戦時間 (hour)
性別
女性
男性

何が問題か？



- 統制すべき「他の要因」が存在
- 女性と男性は同じではない
- 性別が野球の観戦時間 (X) と身長 (Y) の両者に影響を及ぼす
  - 男性の方が野球を観る
  - 男性の方が身長が高い

セレクションバイアス

- 統制変数を入れ忘れた回帰分析だと、なぜ間違えるのか？
  - 最小二乗推定量が、因果効果の推定を誤る：推定結果にバイアスが生じる
    - 欠落変数バイアス
    - 内生性 (endogeneity)
    - セレクションバイアス

セレクションバイアス（復習）

- Selection bias: $\mathbb{E}[Y(0) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0]$
  - $\mathbb{E}[Y(0) \mid D = 1]$: 処置を受けた群の個体が、処置を受けなかったときの潜在的結果の期待値
  - $\mathbb{E}[Y(0) \mid D = 0]$: 処置を受けなかった群の個体が、処置を受けなかったときの潜在的結果の期待値
- $\mathbb{E}[Y(0) \mid D = 1] = \mathbb{E}[Y(0) \mid D = 0]$　ならセレクションバイアスはない → その場合、ATT が推定できる (ATE ではないので注意）
- バイアスがある：処置の値と潜在的結果の値に相関がある
  - 処置を受けた群と受けていない群で、結果のベースラインに違いがある

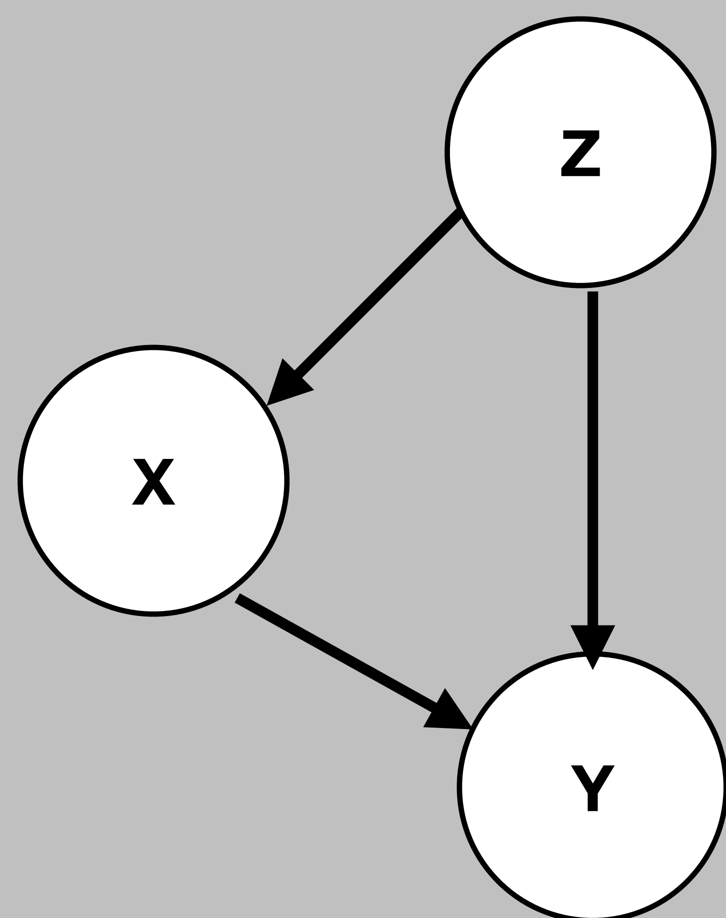
変数 X, Y, Z の関係

- Yは結果、Xは原因とする
- 3つの可能性

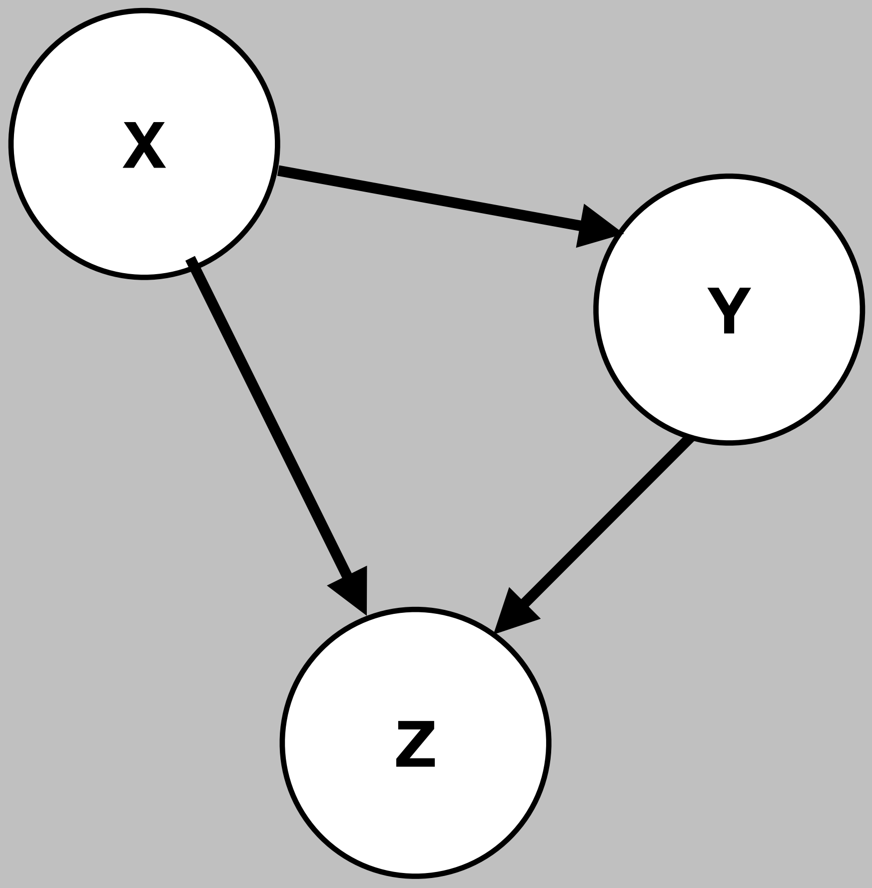
1. Z は X と Y の交絡変数 (confounder) である
2. Z は X と Y の合流点 (collider) である
3. Z は X と Y の媒介変数（mediator, 中間因子）である

▸セレクションバイアスが生じるのは、交絡変数 Z が存在するとき

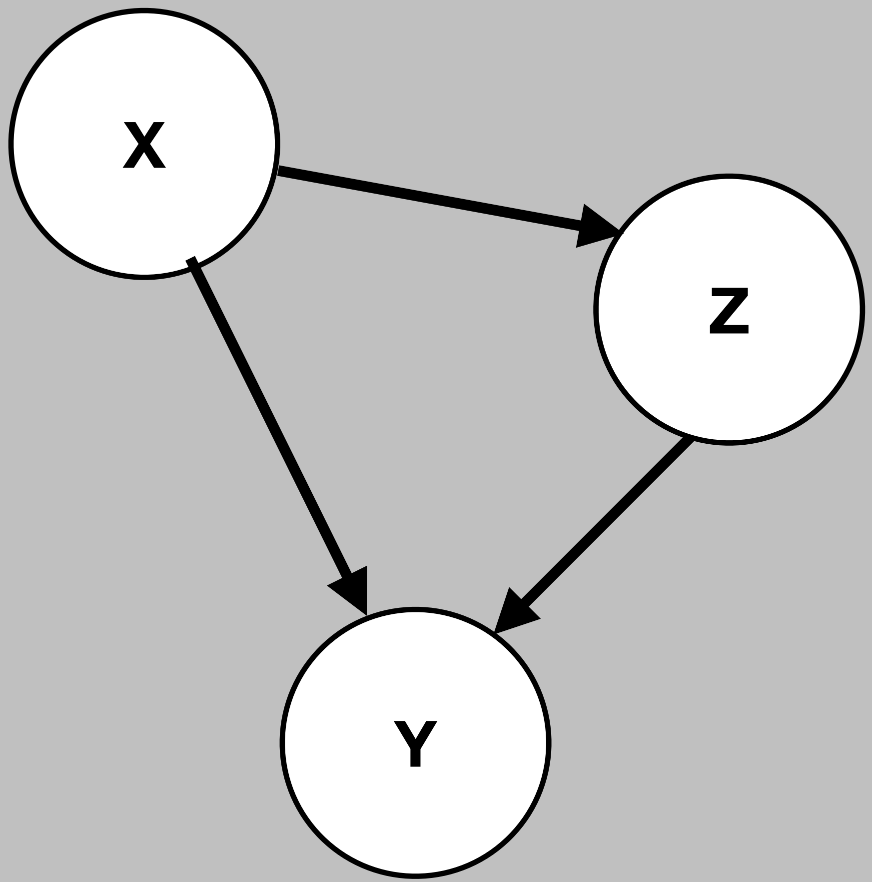
交絡変数 Z



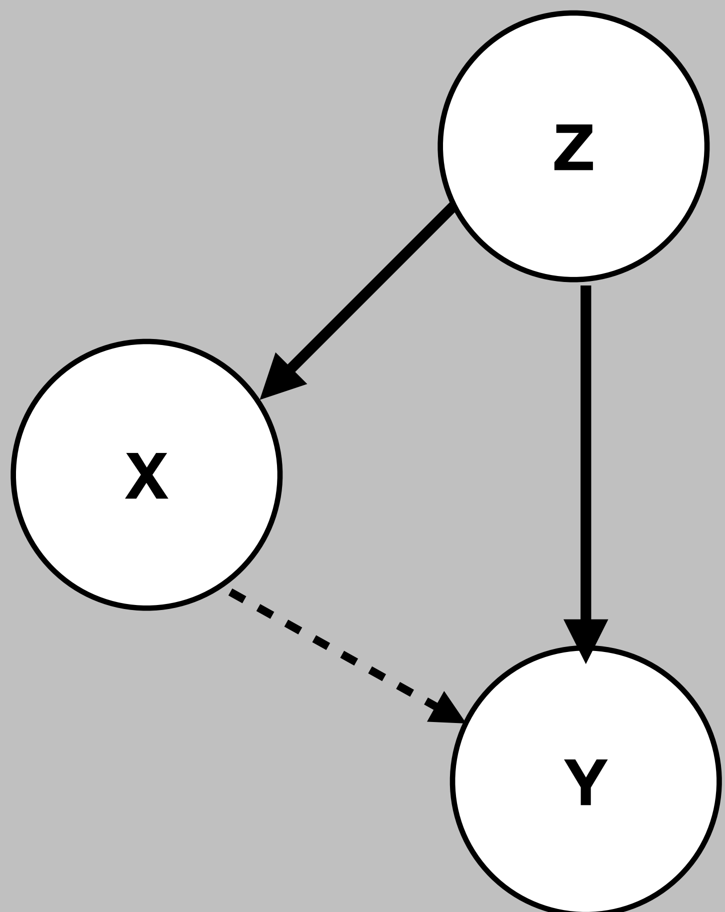
合流点 Z

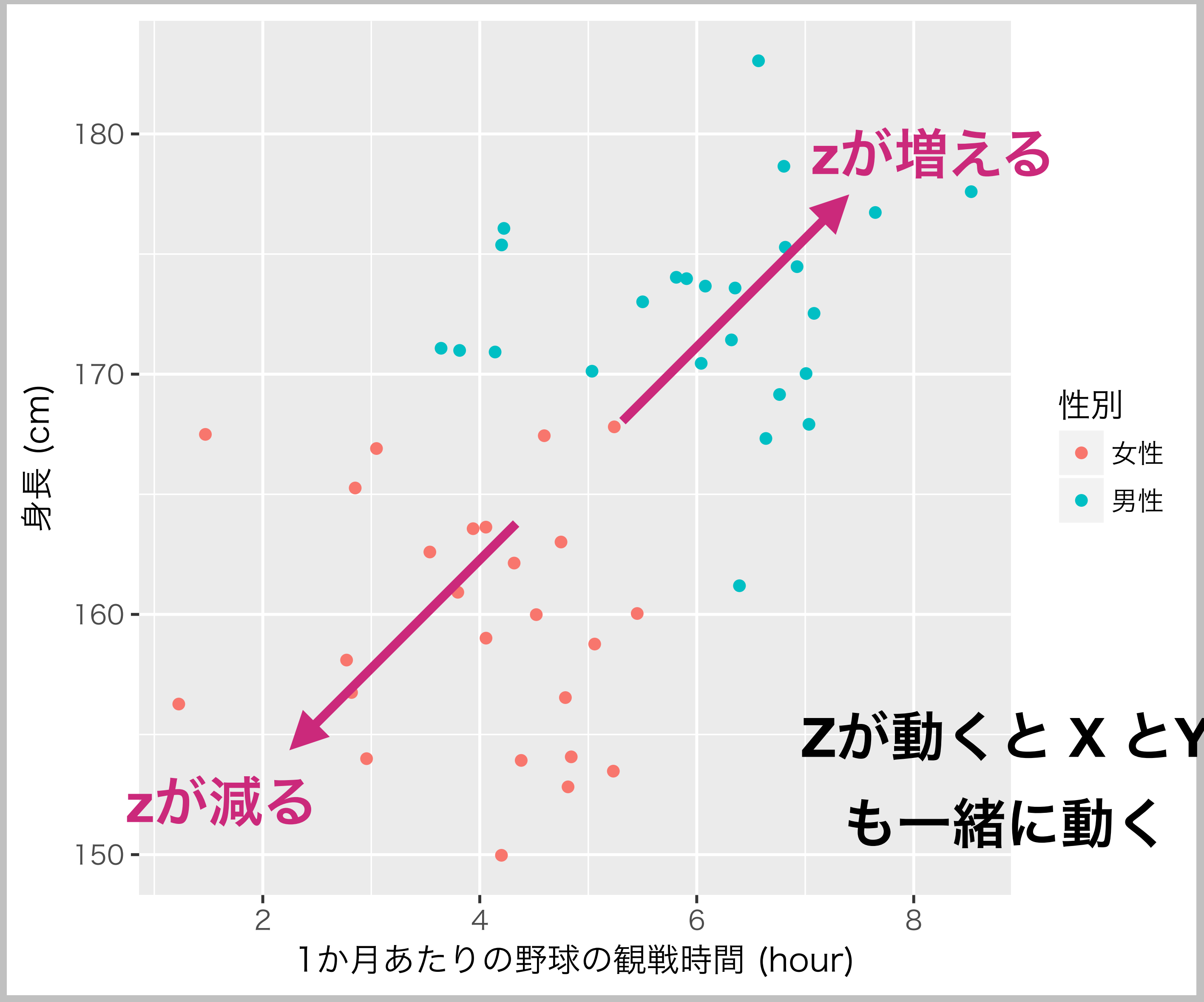


媒介変数 Z



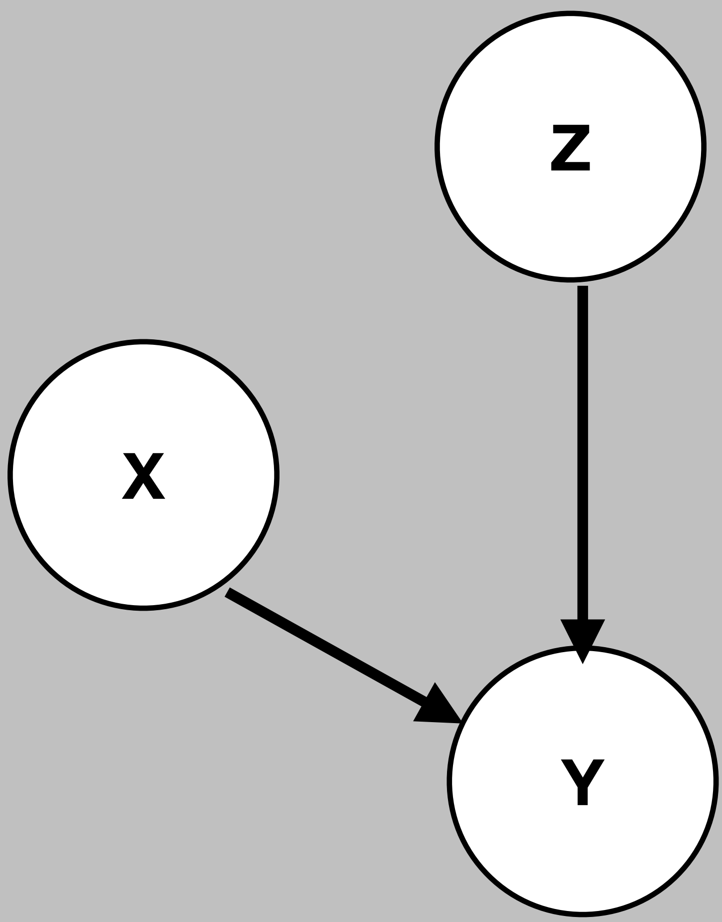
Zが交絡変数のとき

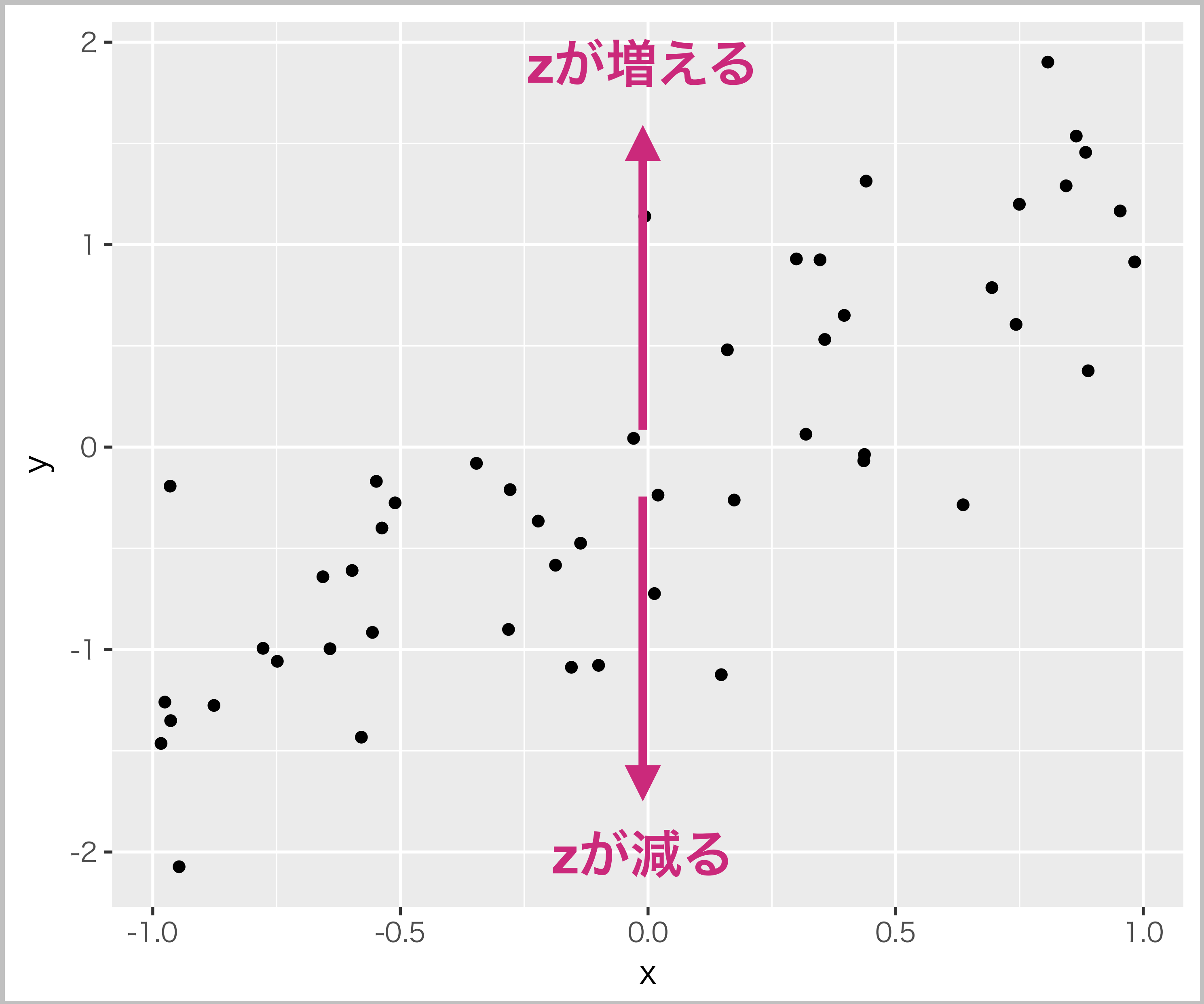




- Zの変化がXとYの変化を同時に引き起こす
- Y を Xだけに回帰すると、バイアスが生じる

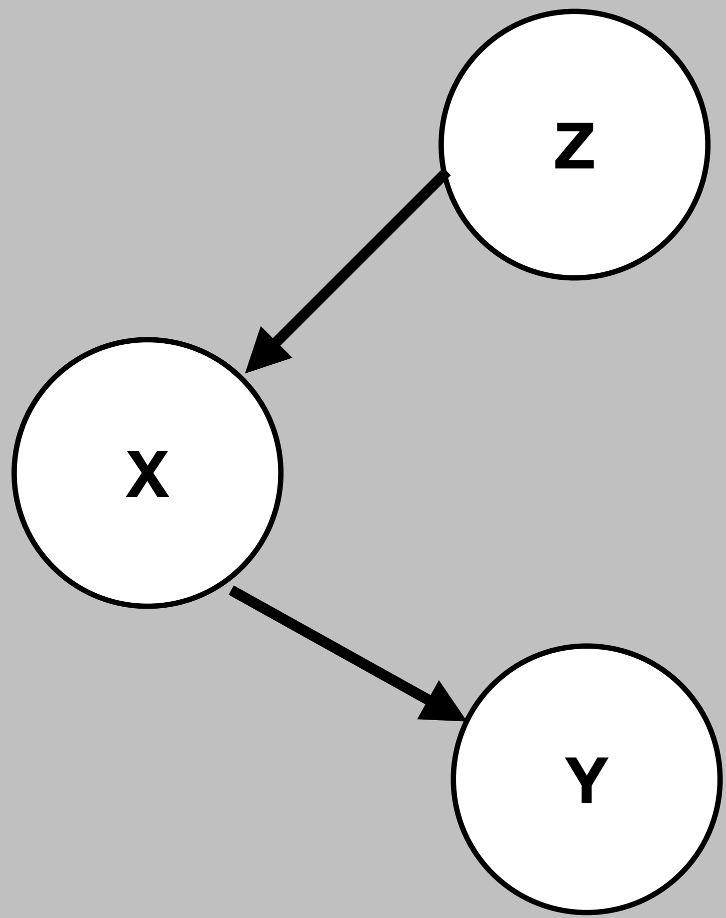
Z が交絡ではない場合 (1)

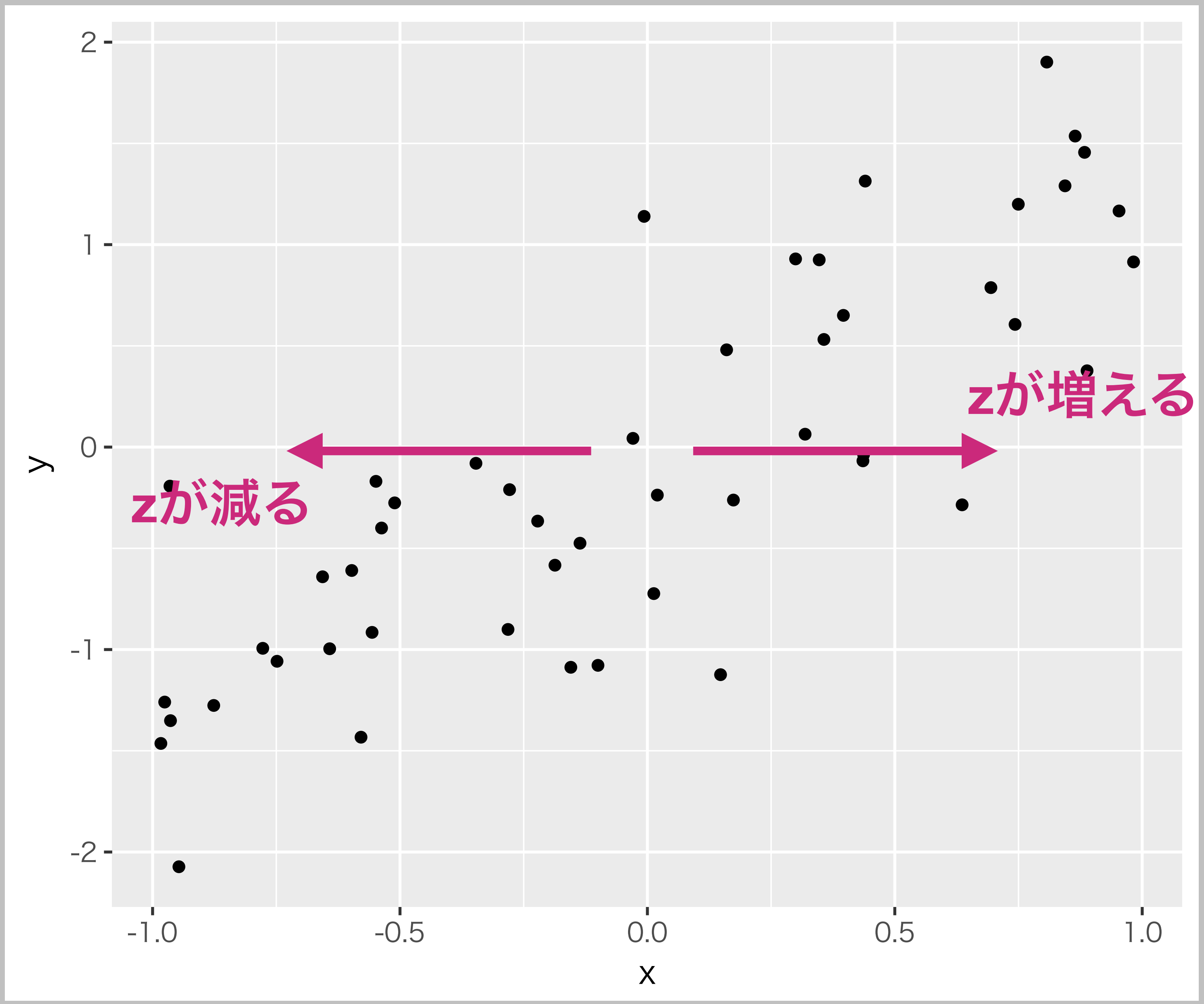




- Zの変化は、Xの変化には影響しない

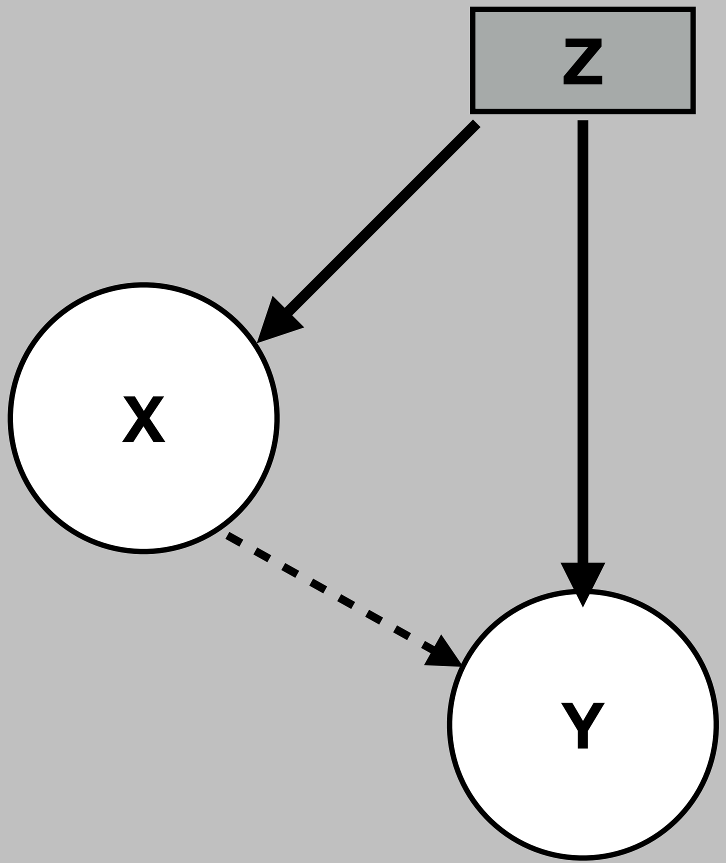
Z が交絡ではない場合 (2)

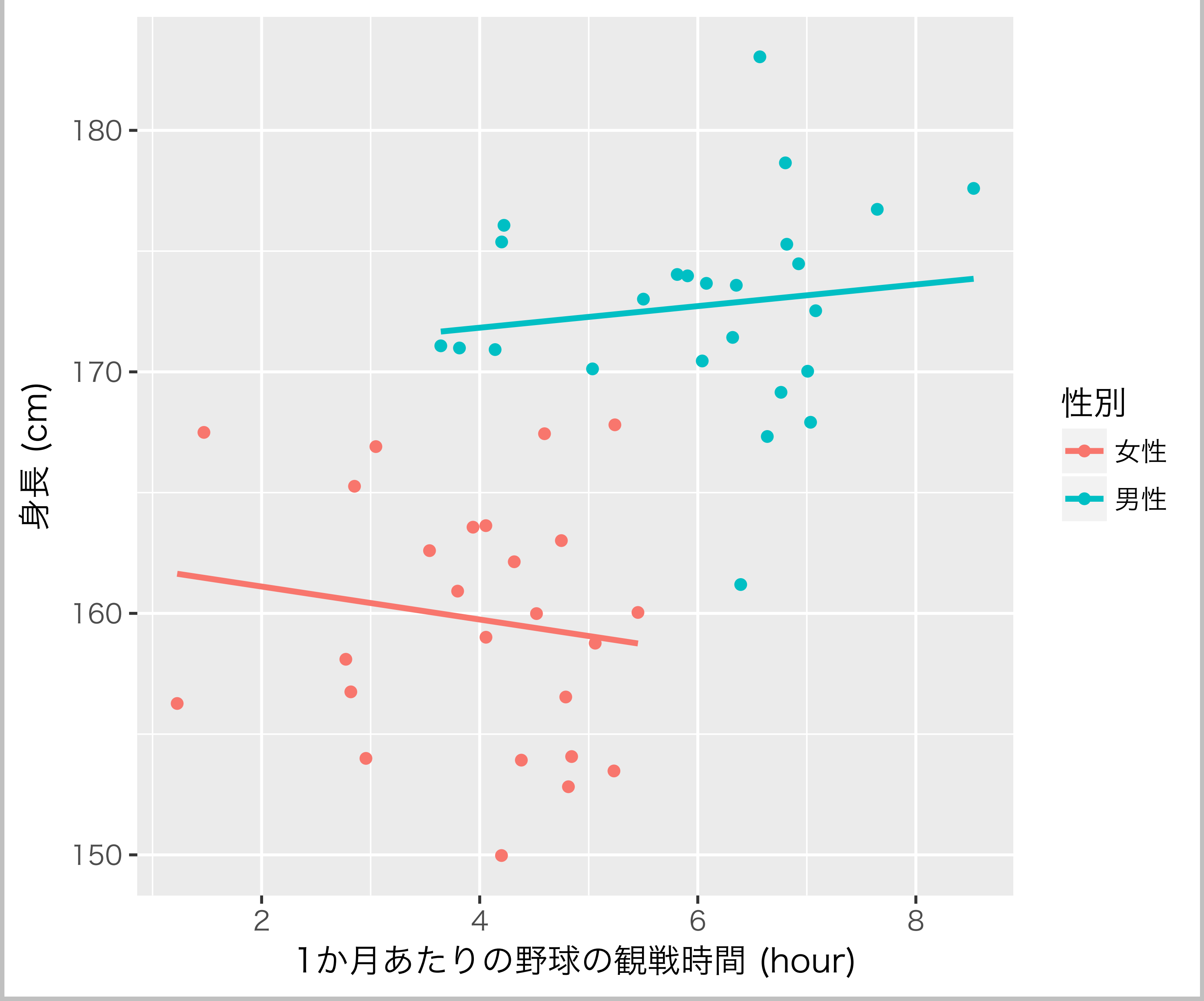




- Zの変化は、Yの変化には影響しない

バイアスを取り除くには？





- Zの値を「固定」すればよい
  - Zを統制した重回帰分析

回帰分析における交絡変数の扱い方

- 交絡は統制（コントロール）せよ！
  - 交絡を、重回帰分析の説明変数に加えればよい！
- 交絡を統制すれば、バイアスを防げる
- 交絡を統制し損ねると、**欠落変数バイアス** (omitted variable bias; OVB, 経済学では**セレクションバイアス** [selection bias] とも呼べれる) が生じる

回帰分析における処置後変数の扱い方

- 因果推論では、原因となる変数のことを「処置変数」と呼ぶ
  - 処置から延びる矢印が向かう先の変数を処置後変数と呼ぶ
- 2つの処置後変数：合流点と媒介変数
  - 理論的に考えて**処置後変数**だと思われる変数は、**回帰分析から外す**
  - 回帰分析に含めてしまうと、「処置後変数バイアス」と呼ばれる問題が生じる

次回

回帰分析の応用（続）