

The background is a vibrant purple field filled with various colorful, rounded geometric shapes in shades of blue, orange, green, and yellow. A white L-shaped graphic element is positioned to the left of the text box.

統計的因果推論を用いた MLBルール改正の効果検証

鈴木研究室B4 大木基嗣

目次

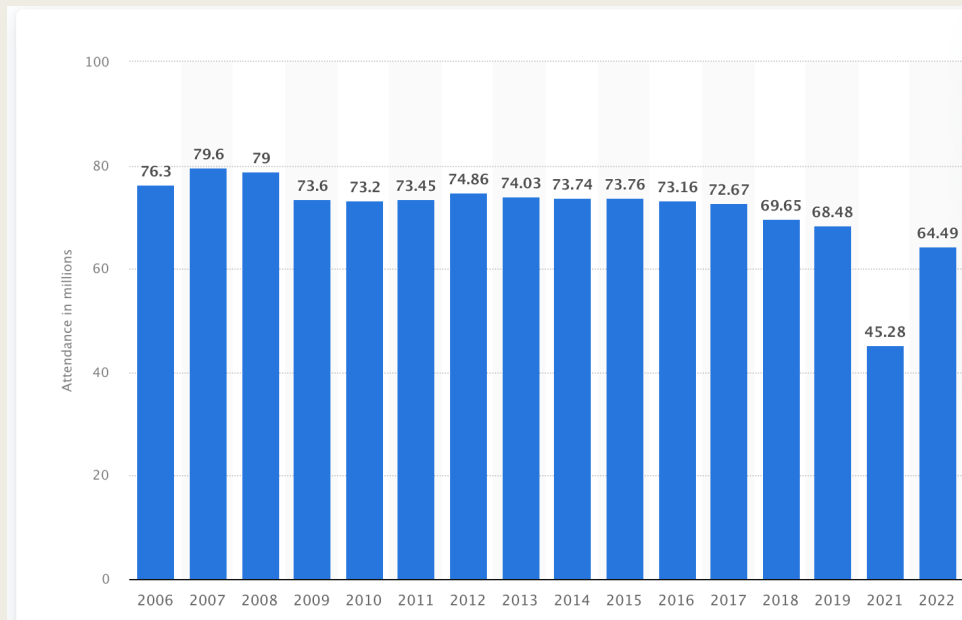
- 研究背景・目的
- 関連する既存研究・論文の紹介
- 活用する手法
- 活用する手法の疑問点
- 現状の報告&今後の予定
- データについて



研究背景・目的

- 背景：野球人気の低下がここ数年話題になっている。
 - その要因の一つは「試合時間が長いこと」だとされており、MLB (Major League Baseball) では2023年より「ピッチクロック」をルールとして新たに追加

MLBの年間観客動員数の変化



<https://www.statista.com/statistics/193421/regular-season-attendance-in-the-mlb-since-2006/>

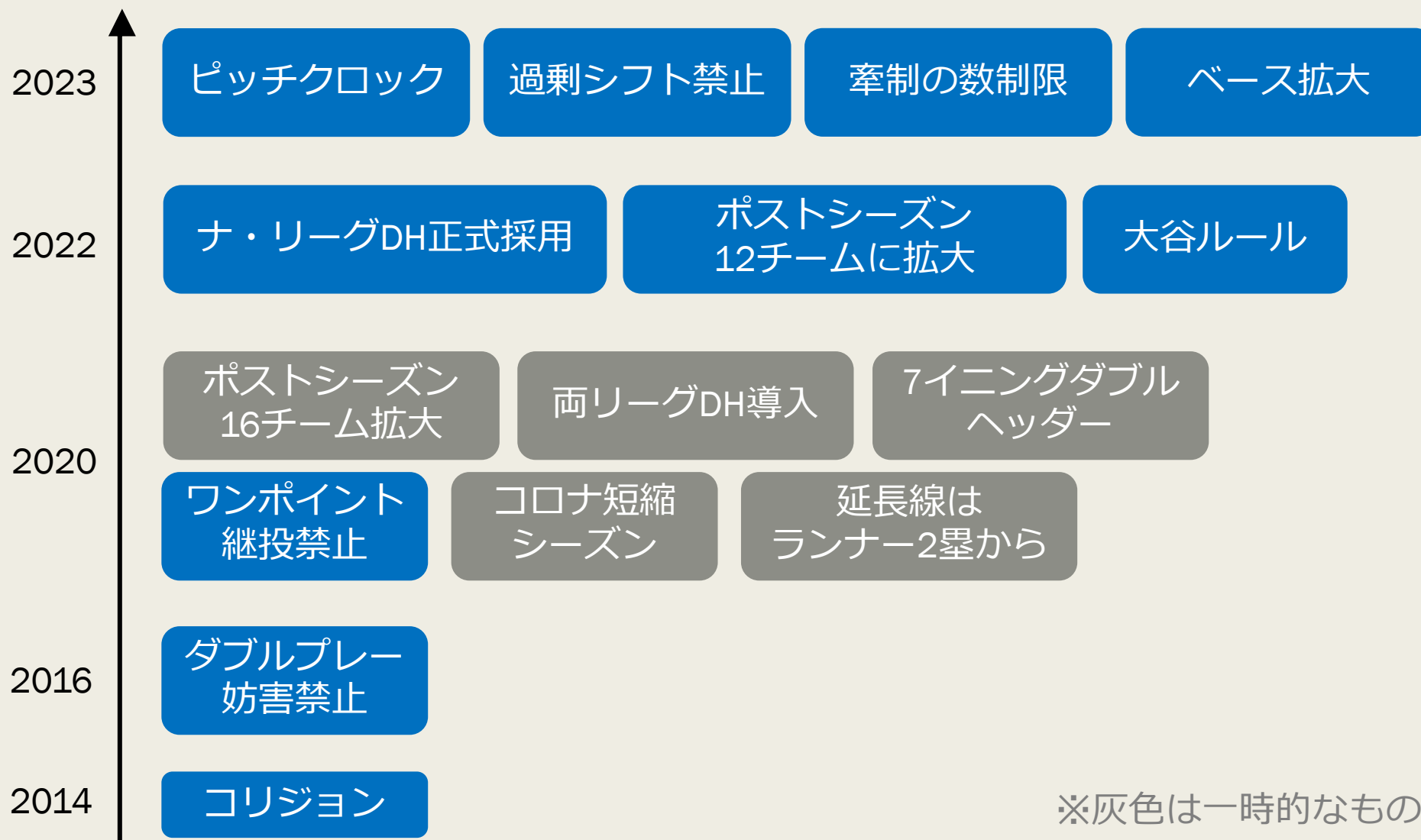
観客者数

- ・ 2012年から9シーズン連続減少
- ・ 2022年はコロナ前（2019）と比較して5.9%↓

視聴者数

2198万人(1992) → 751万人(2022)
ワールドシリーズでさえも、
1180万人(2022)で、2016年から50%減

研究背景・目的



※灰色は一時的なもの

研究背景・目的

- 目的：
 - 「ピッチクロック」が試合時間減少に効果があるのか
 - 「ピッチクロック」が観客動員数の減少に歯止めをかけているか

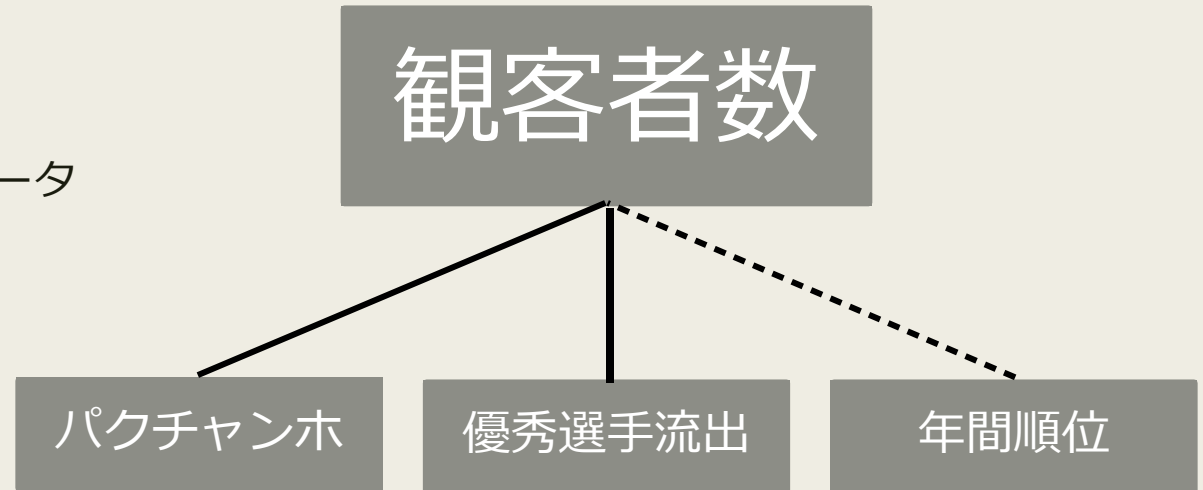
- ピッチクロック：
 - 投手：
 - ランナーなし：ボールを受け取ってから15秒以内に投球しなければボール
 - ランナーあり：ボールを受け取ってから20秒以内に投球しなければボール
 - 打者と打者の間は30秒
 - 打者
 - 制限時間が残り8秒になるまでにピッチャーに注意を向けなければストライク

- →選手にとっても観客にとっても大きな変化



関連する既存研究・論文

- An Assessment of Global Market Penetration of Major League Baseball: Case of the Korean Baseball League
 - 目的
 - 韓国プロ野球リーグ（KPB）で二つの出来事（パクチャンホシンドローム&優秀選手海外流出）による観客者数の増減への影響
 - データ
 - 1991/4～2001/9の平均入場者数の月次データ
 - 手法
 - 中断時系列分析
 - スピアマンの順位相関係数
 - 結論
 - 2つの出来事ともに5%有意
 - 年間順位と観客数の関係性は有意にならず（チーム順位は観客者数にプラスの影響を与えるらしい）
→2つの出来事が観客者数へ影響している



活用する手法

・ 中断時系列分析 (Interrupted Time Series Analysis)

→ある時点の介入によって、それまでの時系列データのトレンドが中断されたかどうかでその介入の影響の有無を判断する手法

線形モデルの例

$$y_t = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} + b_4X_{4t} + b_5X_{5t} + e_t$$

y_t : 1991/4-2001/9の月次観客数

b_0 : パクチャンホシンドローム以前の水準

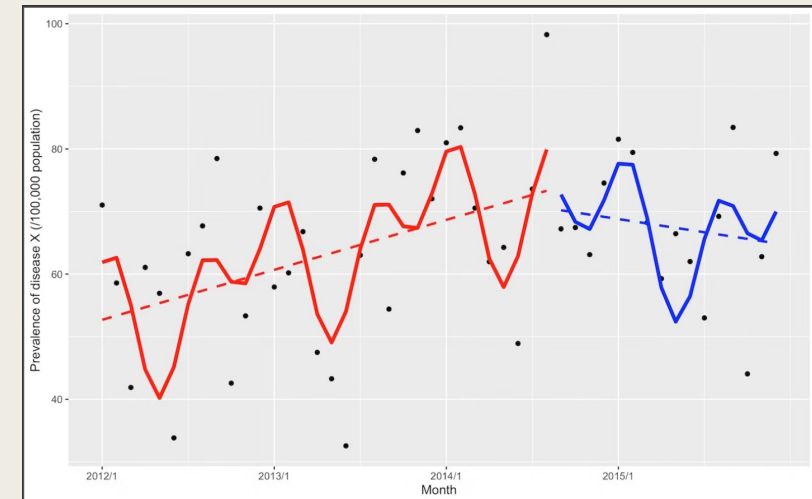
b_1 : パクチャンホシンドローム以前の傾き、 X_{1t} : $N=1\sim66$ の観測値の数

X_{2t} : パクチャンホシンドローム前は0、後は1

X_{3t} : パクチャンホシンドローム前は0、後は1, 2, 3, ...の月次カウンター

X_{4t} : 優秀選手流出前は0、後は1

X_{5t} : 優秀選手流出前は0、後は1, 2, 3, ...の月次カウンター

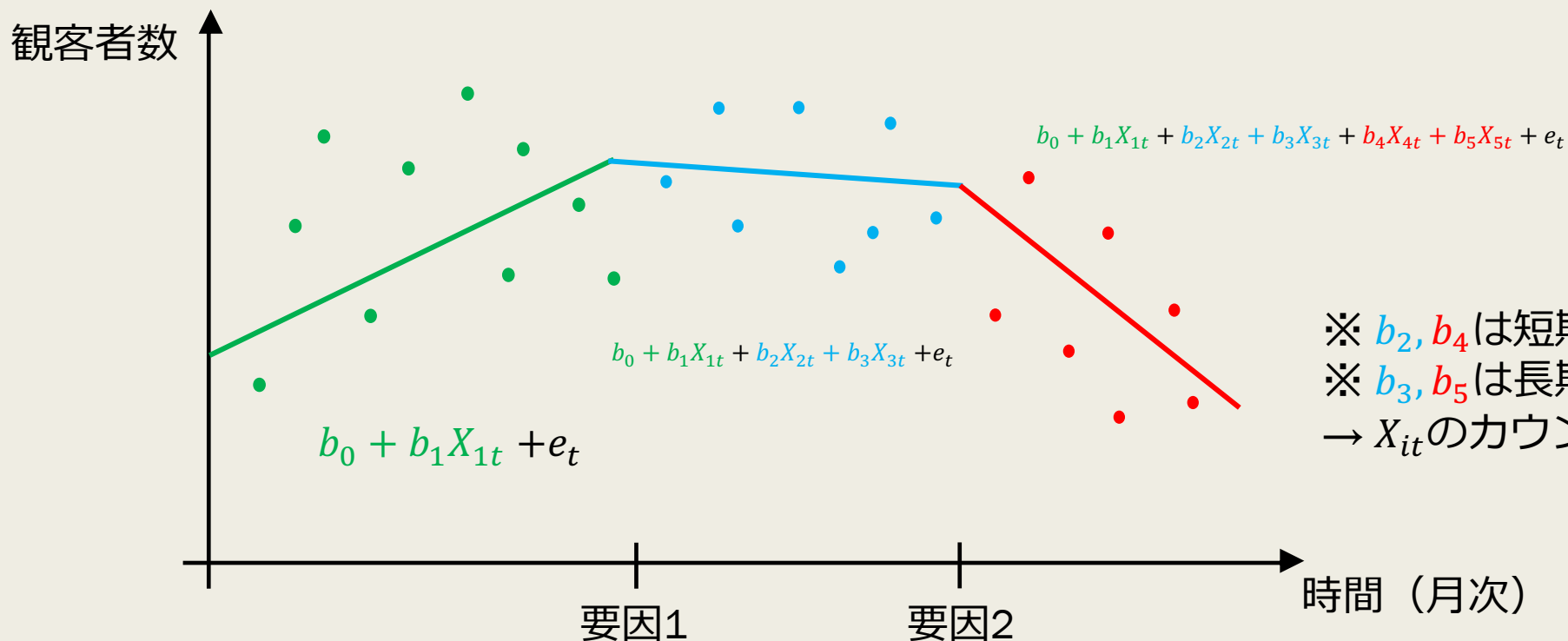


・ 中断時系列分析 (Interrupted Time Series Analysis)

→モデルの各係数が有意かの検定を行い、影響を調べる

線形モデルの例

$$y_t = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} + b_4X_{4t} + b_5X_{5t} + e_t$$



※ b_2, b_4 は短期的影響
※ b_3, b_5 は長期的影響
→ X_{it} のカウンターによる

係数たちが有意かどうかによって、トレンドが変化したかを見る
→水色ゾーンの説明に、もし $b_2, b_3 = 0$ なら、緑の変数のみで説明可能
→トレンドは変化していない、と考える

活用する手法の疑問点

・ 中断時系列分析 (Interrupted Time Series Analysis)

1. 「ある時点を境にトレンドが変化したから、その時に起こった因子が影響している」
→今年から観客者数が増えていれば、「2023年に起きたこと全てが影響した」は言えそうだけど、**「ピッチクロックが」はどうやって言う？**
→特に2023年はコロナ明け、WBC、大谷翔平の影響、、、
2. 論文では「観客者数への影響を及ぼす第3の因子としてチーム順位がある。」
としてパクチャンホシンドローム、優秀選手の海外流出、そしてチーム順位の三つの観点で分析していたが、**絶対に他の因子の影響もあるはず**
→経済、天候、他スポーツの盛り上がり、有名選手の移籍、、、
3. データが2015-2023年までのものしかない
→2023年に起きたことを調べるので**データが足りないのでは？**
→論文では二つの因子は1994と1997。データは1991から2001
4. **コロナシーズン (2021)**
→これは外れ値として外す？
5. そもそも中断時系列分析でいいの？

現状の報告・今後の予定

- 8月上旬
 - テーマの決定
- 8月中旬～下旬
 - 因果推論の勉強・データの手配
- 9月
 - 因果推論の勉強・既存研究調査（ITSのもの、その他因果推論手法）
- 10月
 - 既存研究調査・モデル構築
- 11月
 - モデル構築・考察
- 12月
 - 論文執筆



データについて

- 集めるのに予想以上に時間がかかった
 - *Baseball Reference*, *sportsnavi*, *sportstrader*,,,,
 - 様々なところに問い合わせたが、ほとんどダメだった。
- 野球関連：
 - *MLB-Stats API*というpythonライブラリ <https://github.com/toddrob99/MLB-StatsAPI>
 - *MLB Stats API*のpython ラッパー。非商用利用でOK。
 - 試合時間、観客者数、その他試合データなど
 - *Baseball Savant*
 - 利用規則はMLB.com同様。非商用利用ならOK。
 - 選手データ

