



西日本工業大学

Nishinippon institute of technology

# ソフトウェア技術を用いた バドミントンのデータ分析

井上 翼

西日本工業大学



# Abstract

## 本研究の概要

### 目的

- “誰でも”入手可能なデータによるデータ分析
- ソフトウェア技術を用いたデータ抽出の自動化

### ソフトウェア技術

- プログラミング言語を用いた技術

### データ分析

- 抽出データを用いた分析



# Background (1)

## スポーツとデータ分析

### ☆ 分析のタイミング

### ➤ リアルタイム性とICT

- ・ リアルタイムに分析が可能な種目

バレー : ICT機器をコートに持ち込める

スパイクやサーブのコースの分析・共有が可能

サッカー: 2018年ロシアW杯よりタブレット配布

リアルタイム映像,トラッキングデータ等の共有



## Background (2)

### スポーツとデータ分析

#### ☆ 分析のタイミング

#### ➤ リアルタイム性とICT

- ・ リアルタイムな分析が困難な種目

**バドミントン** : ICT機器の持ち込み不可

ラリーのスピードに記録が追い付かない

総て画像データが基本 = “誰でも”入手できるわけではない



# Background (3)

## スポーツとデータ分析

☆ 画像データ以外のデータ

➤ 「種目 データベース」で検索

①野球(約22,000,000件)



チーム・選手成績, セイバー系データ etc.

第97回 選抜高等学校野球大会 2019年3月23日(土) から12日間 (準々決勝日の休養日を含む 雨天順延) . 組み合わせ抽選会 3月16日(金) . 第101回 全国高等学校野球選手権大会 2019年8月6日(火) から16日間 (2日間の休養日を含む 雨天順延) .

HSBB高校野球データベース (@hsbb\_t) | Twitter  
[https://twitter.com/hsbb\\_t?lang=ja](https://twitter.com/hsbb_t?lang=ja)



# Background (4)

## スポーツとデータ分析

☆ 画像データ以外のデータ

➤ 「種目 データベース」で検索

②サッカー(約1,8500,000件)



J. League Data Site

<https://data.j-league.or.jp/>

Jリーグ公式サイト・日程・公式記録・順位など Jリーグのデータに関する情報を提供いたします

チーム・選手成績, 観客動員, 得点データ etc.

[www.football-lab.jp/](http://www.football-lab.jp/)

フットボールラボ(Football LAB)はサッカーをデータで分析し、新しいサッカーの観戦方法を伝えるサッカー情報サイトです。選手のプレーを評価するチャンスビルディングポイントやプレススタイル指標、チームの戦術を評価するチームスタイル指標といった独自の...

表監督、ユース代表試合結果などを掲載。

ワールドサッカーデータベース

[www.geocities.jp/k1a1w1a1i1/warudosakkadetabesu.htm](http://www.geocities.jp/k1a1w1a1i1/warudosakkadetabesu.htm)

# Background (5)

## スポーツとデータ分析

- ☆ 画像データ以外のデータ
- 「種目 データベース」で検索

### ③バドミントン(約1,540,000件)



動画データ, 練習データ, ノックデータ  
→ 明らかに他種目と質が異なる



# Background (6)

## スポーツとデータ分析

☆ 画像データ以外のデータ

➤ 「種目 データベース」で検索

野球 データベースに関連する	バドミントン データベースに関連する検索キーワード		データベースに関連する検索キーワード
プロ野球 データ csv	高校野球 バドミントン ライブ	バドミントン 代表	Jリーグ プレビュー
プロ野球 データ分析	高校野球 バドミントン スケジュール	ポリイバドミントン	関東 大学 サッカー データ
プロ野球 データ 2018	高校野球 バドミントン テレビ	バドミントン wtf 速報	Jリーグ 台風 情報
高校野球データベース千葉	野球：バドミントン 観戦	youtube バドミントン 2018	サッカーまとめ
プロ野球 データ 守備	野球：バドミントン 動画 桃田	マレーシア オープン バドミントン 2019	サッカー スタッツ どうやって

・ 分析に関連するキーワード(csv, excel, DB, スタッツ etc.)がない

“誰でも”入手可能かつ分析可能なデータはないのか？





## Background (7)

### スポーツとデータ分析

#### ☆ バドミントンにおける

“誰でも”入手可能かつ分析可能なデータ

#### ➤ tournamentsoftware.com

- ・ ランキング, スコア (時間経過, 最多連続得点), 試合時間 etc.
- ・ 膨大な量の過去の試合データ

自動化して抽出し, 分析利用できないか?

“誰でも”入手可能かつ分析可能なデータを

自動化して抽出, 分析する



# scraping (1)

## バドミントンデータの抽出

### ① アナログ的方法

- ・ とにかく時間がかかる, 自動化はできない

### ② ソフトウェア技術の利用

- ・ スクレイピング技術

スクレイピング(ウェブスクレイピング)

ウェブサイトから情報を抽出するソフトウェア技術.



# scraping (2)

## バドミントンデータの抽出

スクレイピング(ウェブスクレイピング)

ウェブサイトから情報を抽出するソフトウェア技術.

メリット: データ取得の自動化, 業務・生産性の効率向上

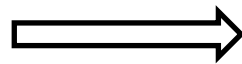
デメリット: サイトが変わる度にプログラムの修正が必要

Tournament days

WED 12 DEC	THU 13 DEC	FRI 14 DEC	SAT 15 DEC	SUN 16 DEC
------------------	------------------	------------------	------------------	------------------

Order of play of Sunday, December 16, 2018

Time	Draw	Score	Duration
Tianhe Gymnasium - 1			
1. Starting at 1:00 PM	WS PUSARLA V. Sindhu [ ] - Nozomi OKUHARA [ ]	21-19 21-17	1:02
2. Followed by	XD WANG Yilyu [ ] HUANG Dongping [ ] - ZHENG Siwei [1] HUANG Yaqiong [ ]	23-21 16-21 21-18	1:09
3. Followed by	MS SHI Yuqi [3/4] [ ] - Kento MOMOTA [2] [ ]	21-12 21-11	0:49
4. Followed by	WD Misaki MATSUTOMO [1] [ ] Ayaka TAKAHASHI [ ] - LEE So Hee [ ] SHIN Seung Chan [ ]	21-12 22-20	0:49
5. Followed by	MD LI Junhui [ ] LIU Yuchen [ ] - Hiroyuki ENDO [ ] Yuta WATANABE [ ]	21-15 21-11	0:45



21-19 21-17  
 23-21 16-21 21-18  
 21-12 21-11  
 21-12 22-20  
 21-15 21-11

スクレイピング元ページ

スクレイピング後



## scraping (3)

### バドミントンデータの抽出

#### ☆ スクレイピング方法

#### ➤ 汎用プログラミング言語Pythonの使用

- ・ コードがシンプル, 扱いやすい設計
- ・ 実行速度は他の言語より遅い
- ・ モジュール(関数群)を多用する

`import モジュール名`

- ・ 外部モジュール(世界中の開発者によるもの)による拡張が可能

`pip install モジュール名`



## scraping (4)

### バドミントンデータの抽出

#### ☆ スクレイピング方法

#### ➤ 汎用プログラミング言語Pythonの使用

- コードがシンプル, 扱いやすい設計
- 実行速度は他の言語より遅い
- モジュール(関数群)を多用する

```
import モジュール名
```

- 外部モジュール(世界中の開発者によるもの)による拡張が可能

```
pip install モジュール名
```



## scraping (5)

### バドミントンデータの抽出

#### ☆ スクレイピング方法

#### ➤ 用いた外部モジュール

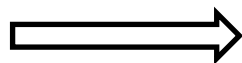
- Requests: webサイトにアクセスし, コードを取得する
- Beautiful soup: HTMLの構文を解析し, 必要部分を抽出する
- Selenium: ブラウザの自動操作を行う
- OpenPyXL: PythonでExcelを操作する
- matplotlib: Pythonでグラフを描画する

## バドミントンデータの抽出

### ☆ プログラミング例

Order of play of Sunday, December 16, 2018

Time	Draw	Score	Duration
1. Starting at 1:00 PM	WS PUSARLA V. Sindhu (IND) vs. Nozomi OKUHARA (JPN)	21-19 21-17	1:02
2. Followed by	XD WANG Yiliyu (CHN) vs. ZHENG Siwei [1] (CHN) HUANG Dongping (CHN) vs. HUANG Yaorong (CHN)	23-21 16-21 21-18	1:09
3. Followed by	MS SHI Yuqi [B4] (CHN) vs. Kento MOMOTA [2] (JPN)	21-12 21-11	0:49
4. Followed by	WD Misaki MATSUTOMO [1] (JPN) vs. LEE So Hee (KOR) Ayaka TAKAHASHI (JPN) vs. SHIN Seung Chan (KOR)	21-12 22-20	0:49
5. Followed by	MD LI Junhui (CHN) vs. Hiroyuki ENDO (JPN) LIU Yuchen (CHN) vs. Yuta WATANABE (JPN)	21-15 21-11	0:45



21-19 21-17  
23-21 16-21 21-18  
21-12 21-11  
21-12 22-20  
21-15 21-11

スクレイピング元ページ

スクレイピング後

```
import requests, bs4
url = 'https://bwf.tournamentsoftware.com/sport/matches.aspx?id=F0D25C8F-6A9A-49DE-97FC-E58E3DB74CF1&d=20181216'
res = requests.get(url)
soup = bs4.BeautifulSoup(res.text)
score_list = soup.select('table > tbody > tr > td > span')
for score in score_list:
    print(score.getText())
```



# analysis (1)

## バドミントンデータの分析例

### ☆ 分析対象

- tournamentsoftware.comに記載の国際大会
- 2019年2月12日時点の世界ランキング上位10名の出場試合
- MS(102), WS(111), MD(98), WD(86), XD(94) ※ ()はgame数

### ☆ 分析期間

- 2018年の1年間(合計491games)





## analysis (2)

### バドミントンデータの分析例

#### ☆ 分析テーマ

(1) 1st gameおよび2nd gameとfinal gameの勝率の関係

- 1stと2ndの内容がfinalの勝率に影響するのか？

- 二項検定により評価

(2) 2nd gameまでの得失点差とfinal gameの勝率関係

- 2<sup>nd</sup> gameまでの得失点差がfinalの勝率に影響するのか？

- 相関係数, 単回帰分析により評価



# analysis (3)

## バドミントンデータの分析例

### ☆ 分析結果(1)

Tab. 1 取得gameの違いによる勝率への影響

種目	Game数	1st, final	2nd, final	p値
MS	102	42	60	0.0918
WS	111	43	68	0.0223

WS: 2ndを取った方が勝率が高い  
XD: 1stを取った方が勝率が高い

XD	94	57	37	0.0495
----	----	----	----	--------

有意水準5% ( $p < .05$ )



# analysis (4)

## バドミントンデータの分析例

### ☆ 分析結果(2)

Tab. 2 2ndまでの得失点差の違いによる勝率への影響

種目	相関係数	相関関係	回帰直線
MS	0.023	ほぼ無相関	$y=0.076x+58.70$

WS	0.610	正の相関	$y=2.130x+54.08$
----	-------	------	------------------

WS, WD, XDで得失点差がfinalの勝率に影響

WD	0.517	正の相関	$y=2.507x+54.04$
----	-------	------	------------------

XD	0.453	正の相関	$y=1.720x+59.80$
----	-------	------	------------------



## analysis (5)

### バドミントンデータの分析例

#### ☆ 分析結果

(1) 1st gameおよび2nd gameとfinal gameの勝率の関係

- 1stと2ndの内容がfinalの勝率に影響するの？

→ **WSとXDにおいて勝率に影響**

(2) 2nd gameまでの得失点差とfinal gameの勝率関係

- 2<sup>nd</sup> gameまでの得失点差がfinalの勝率に影響するの？

→ **WS, WD, XDにおいて勝率に影響**



# conclusion

## ソフトウェア技術を用いたバドミントンのデータ分析

### ☆ 目的

“誰でも”入手可能かつ分析可能なデータを自動化して抽出・分析

### ☆ 抽出方法

プログラミング言語Pythonを用いたWebスクレイピング

### ☆ 分析方法

様々な統計的手法を用いた分析

### ☆ 結果

データからfinal gameの勝率について示唆



## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (MS)

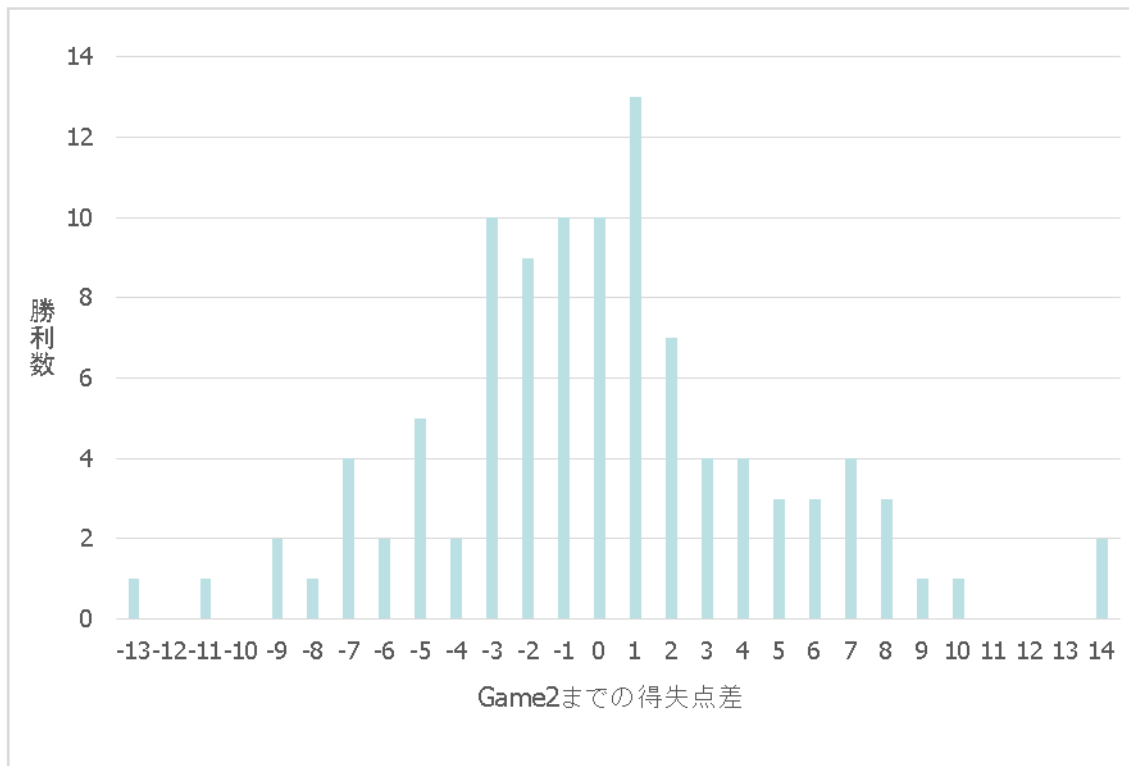


図1 Game2までの得失点差とそこから勝利した試合数

## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (MS)

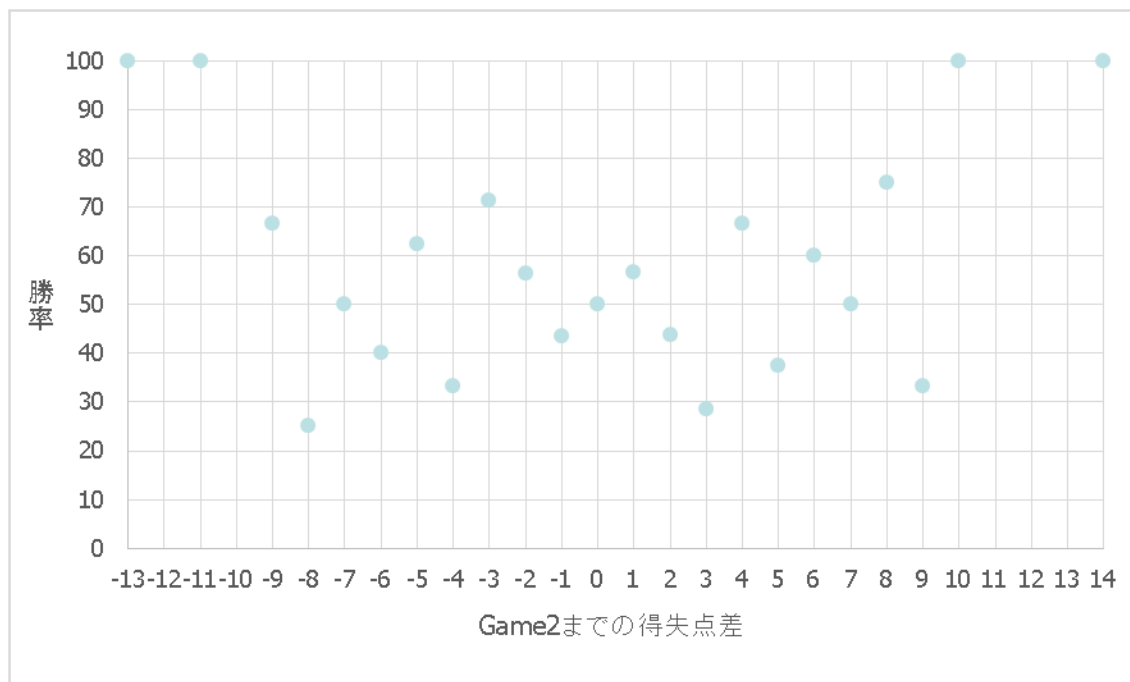


図2 Game2までの得失点差ごとのファイナルゲームの勝率



## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (MS)

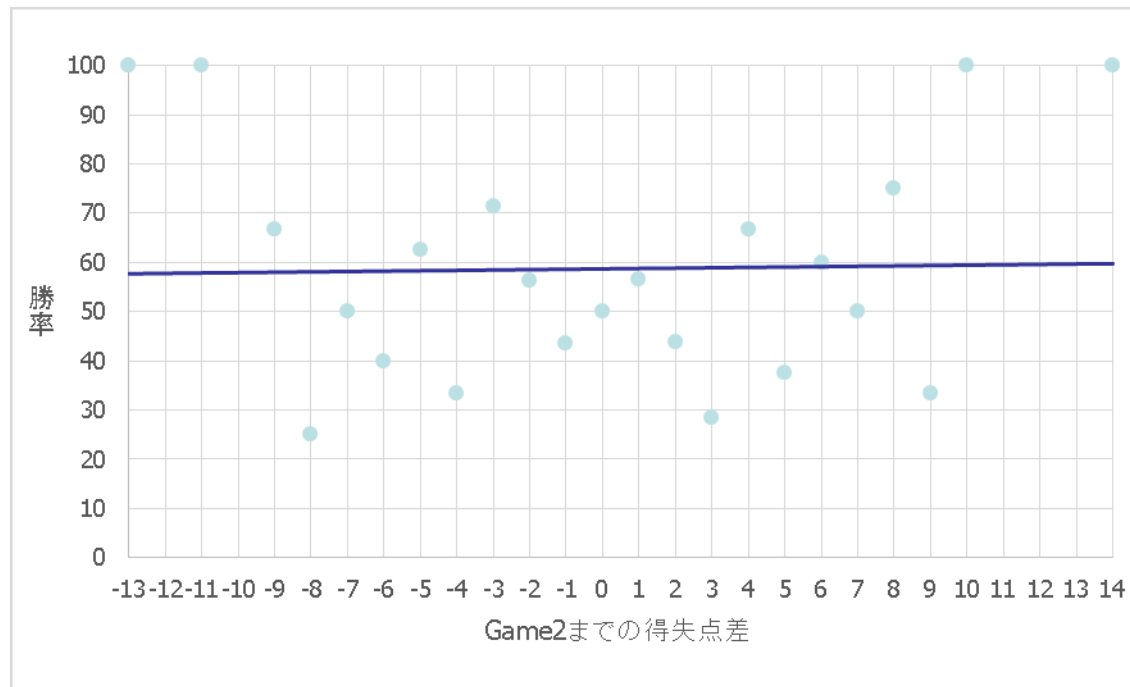


図3 単回帰分析の結果

## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (WS)

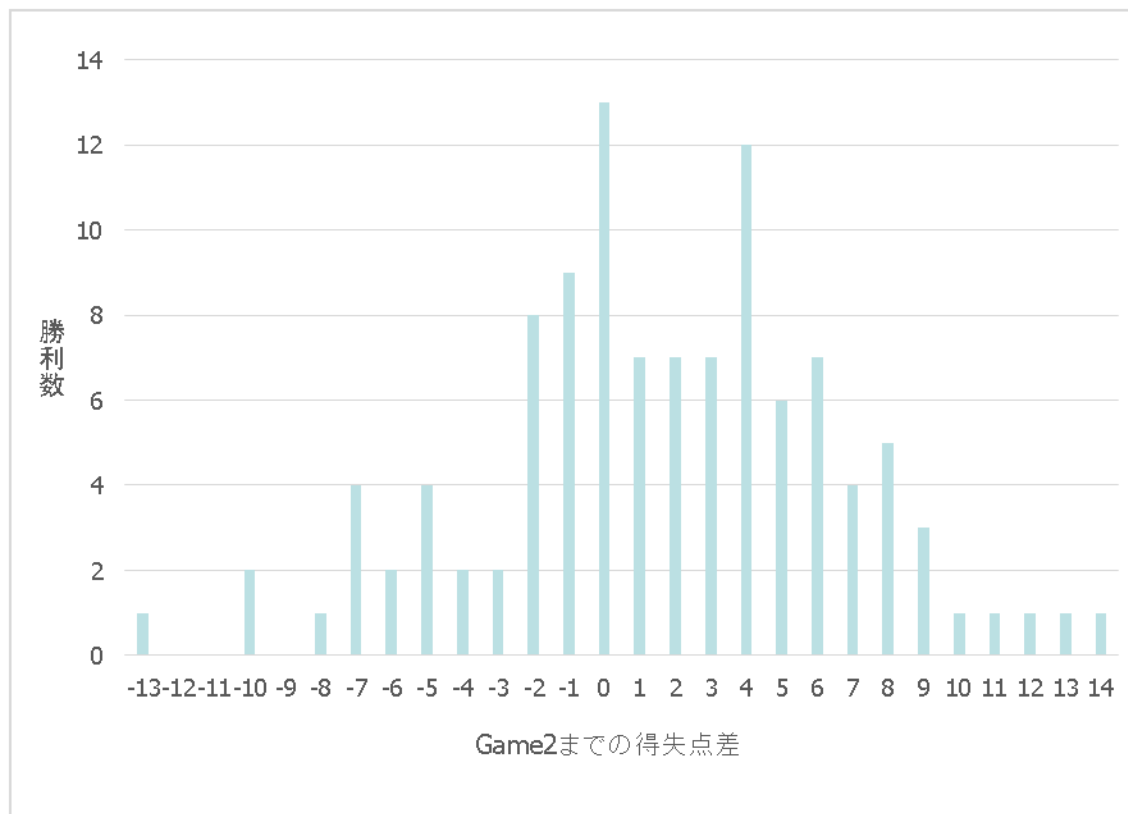


図4 Game2までの得失点差とそこから勝利した試合数

② Game2までの得失点差はファイナルゲームの勝率に影響するか (WS)

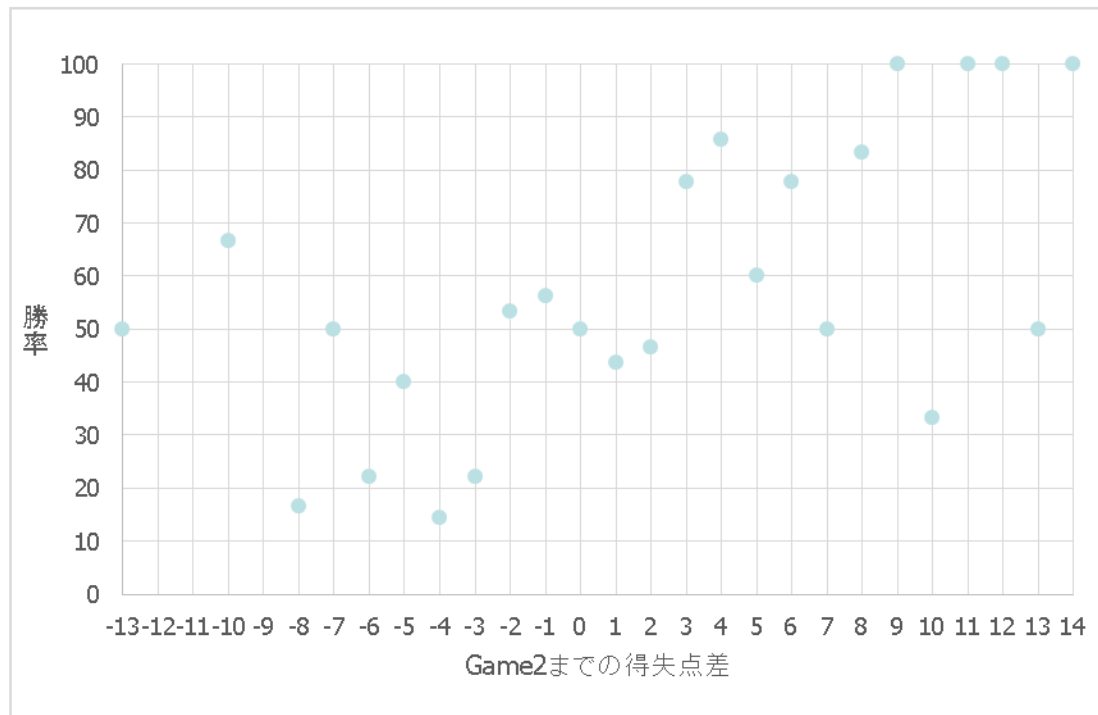


図5 Game2までの得失点差ごとのファイナルゲームの勝率

## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (WS)

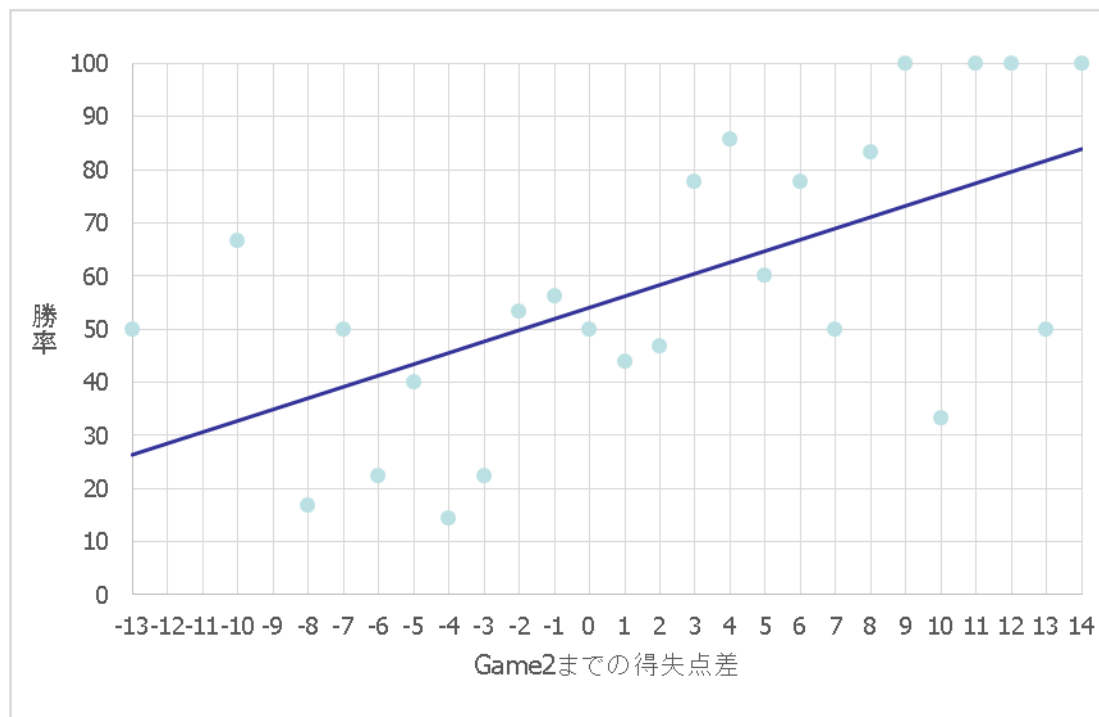


図6 単回帰分析の結果

② Game2までの得失点差はファイナルゲームの勝率に影響するか (MD)

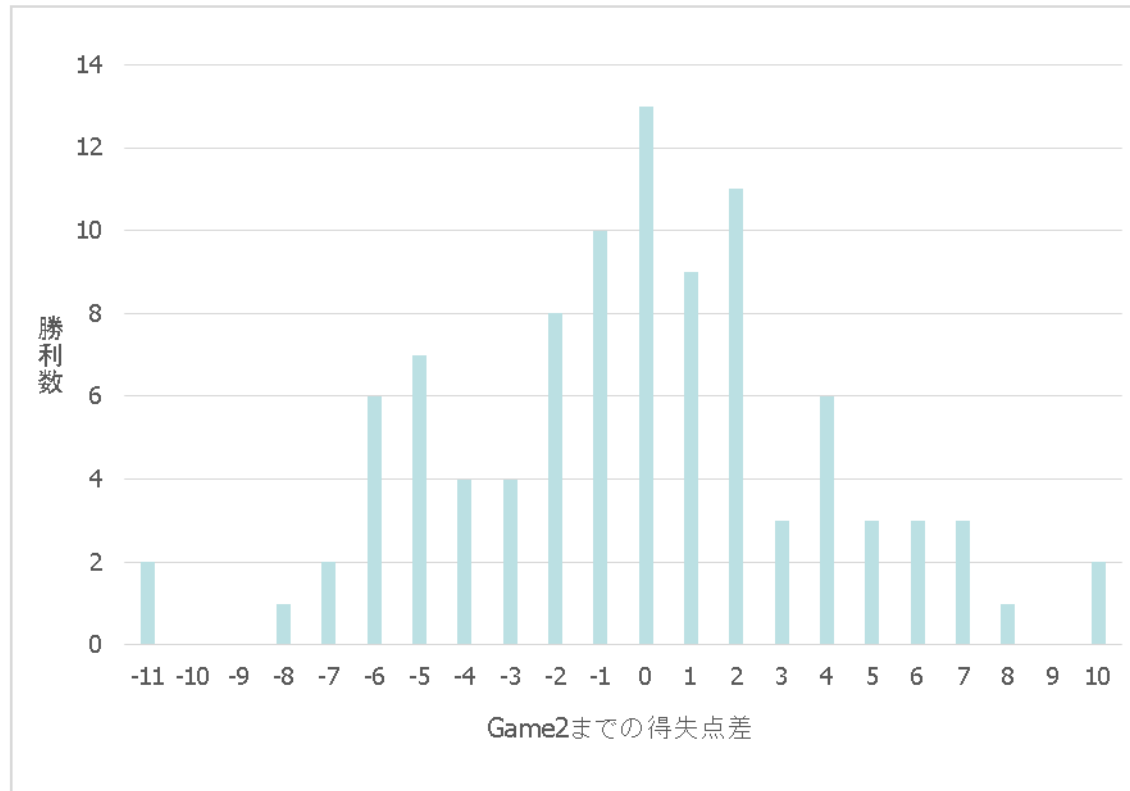


図7 Game2までの得失点差とそこから勝利した試合数

② Game2までの得失点差はファイナルゲームの勝率に影響するか (MD)

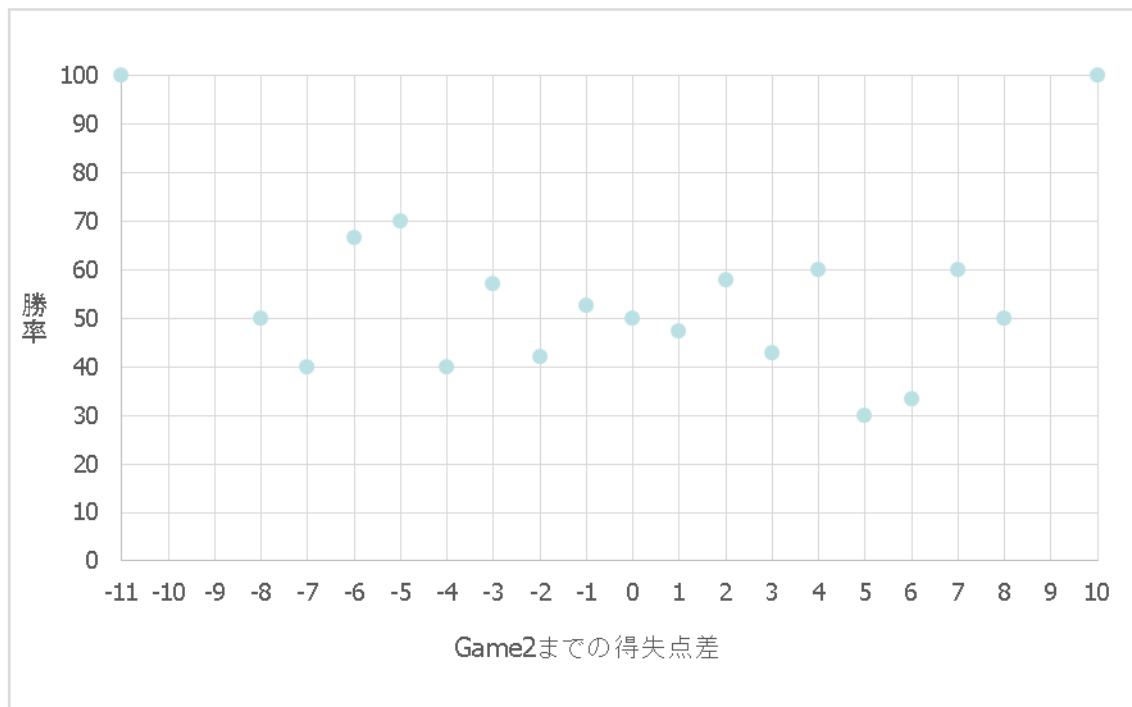


図8 Game2までの得失点差ごとのファイナルゲームの勝率

② Game2までの得失点差はファイナルゲームの勝率に影響するか (MD)

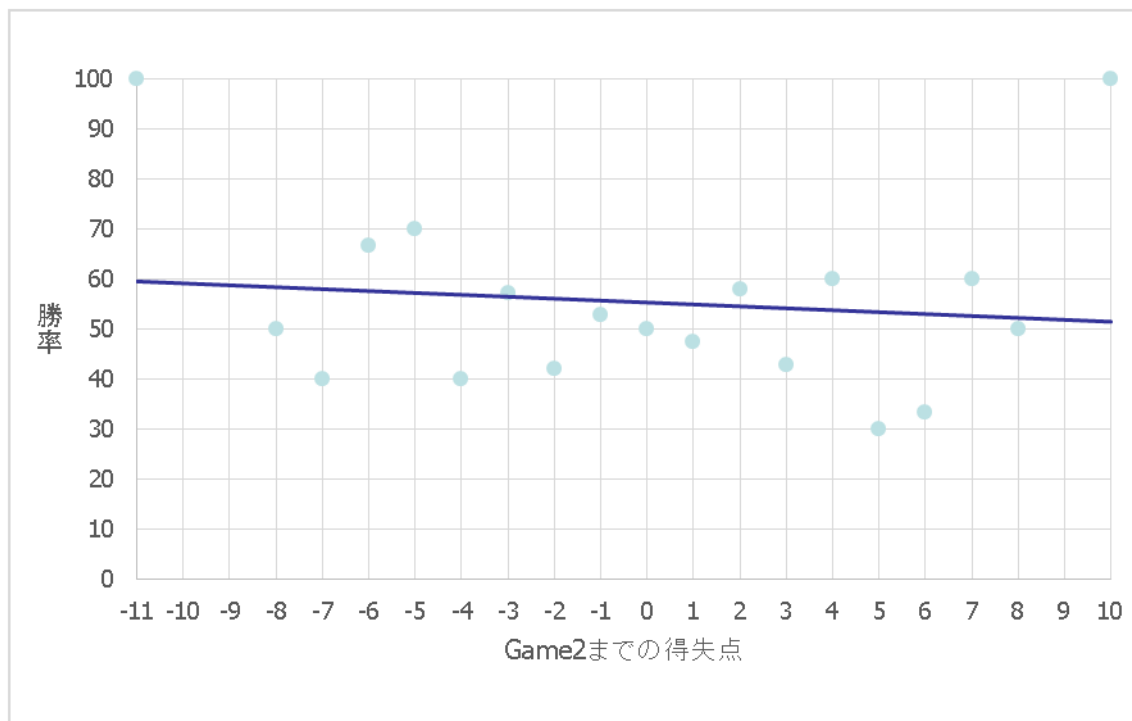


図9 単回帰分析の結果

② Game2までの得失点差はファイナルゲームの勝率に影響するか (WD)

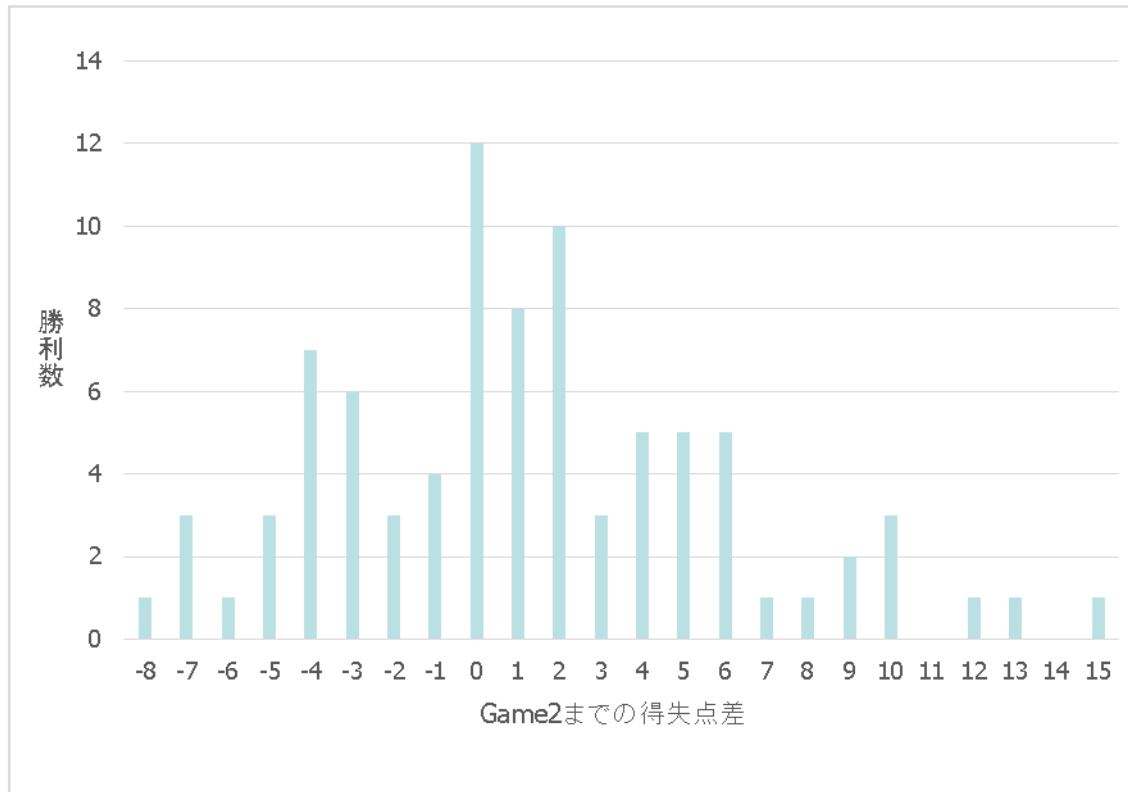


図10 Game2までの得失点差とそこから勝利した試合数



## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (WD)

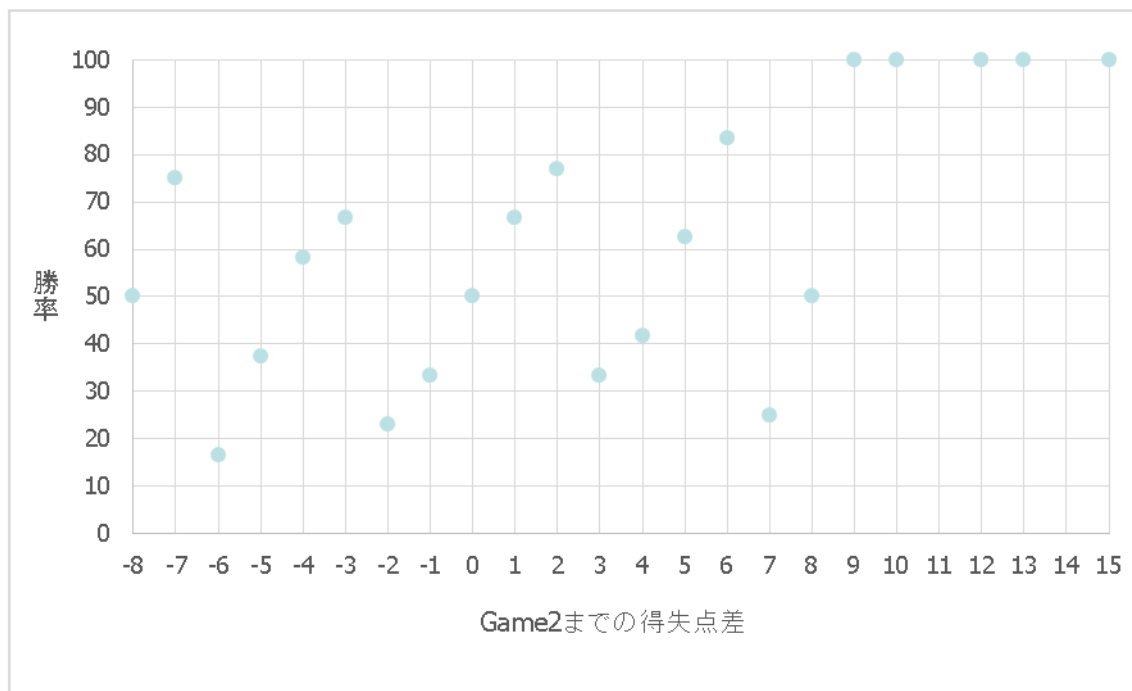


図11 Game2までの得失点差ごとのファイナルゲームの勝率

② Game2までの得失点差はファイナルゲームの勝率に影響するか (WD)

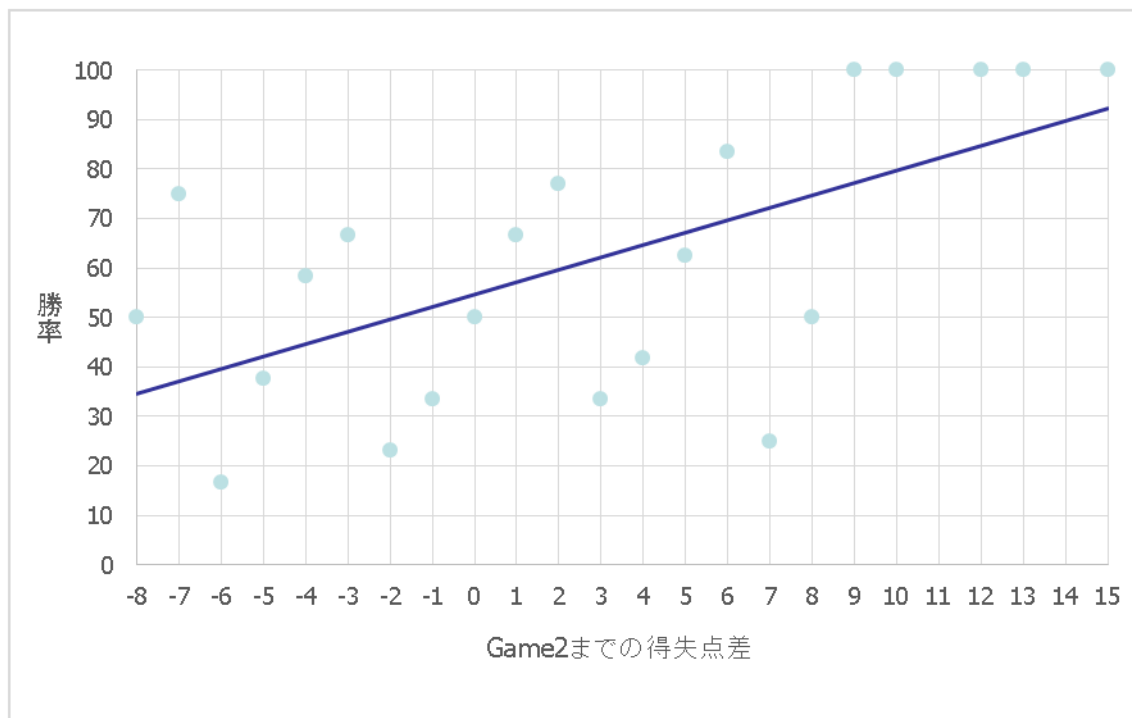


図12 単回帰分析の結果

## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (XD)

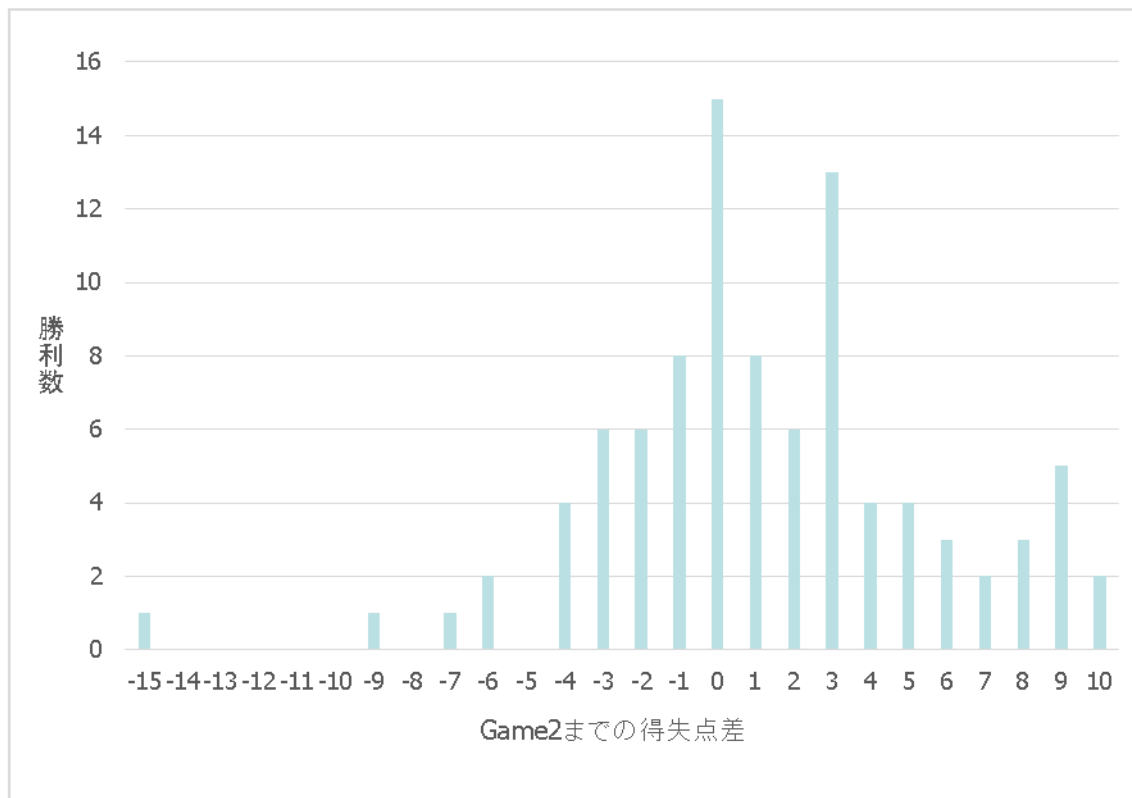


図13 Game2までの得失点差とそこから勝利した試合数

## ② Game2までの得失点差はファイナルゲームの勝率に影響するか (XD)

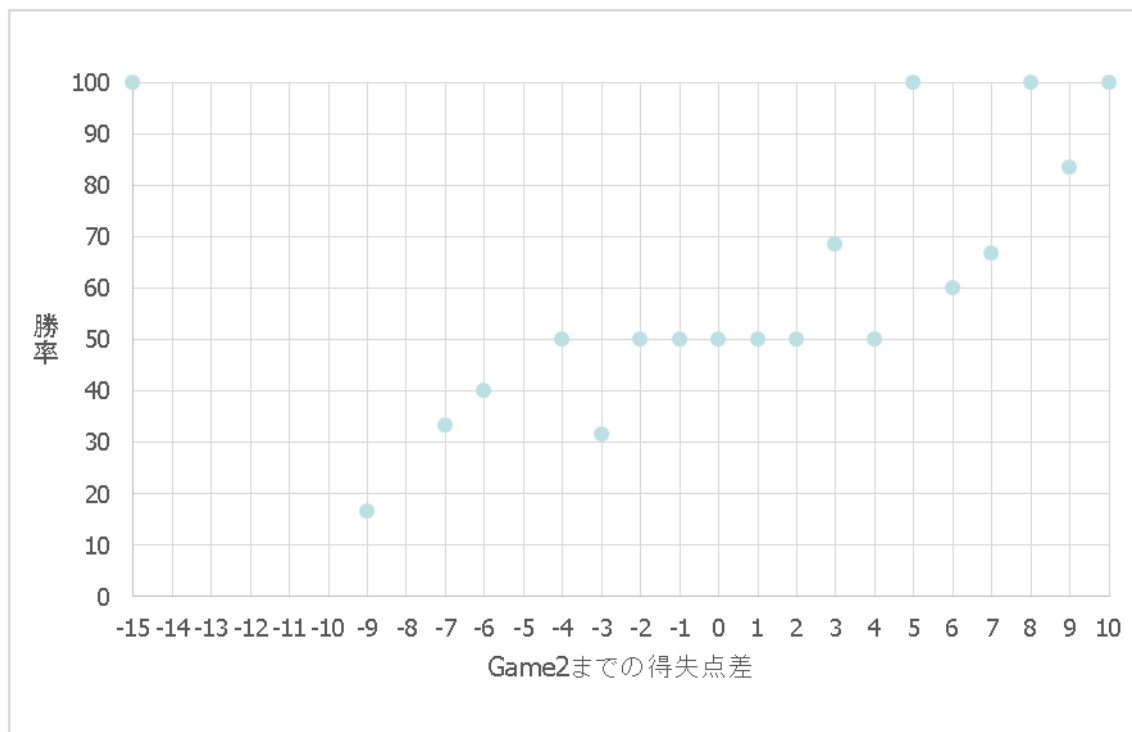


図14 Game2までの得失点差ごとのファイナルゲームの勝率

② Game2までの得失点差はファイナルゲームの勝率に影響するか (WD)

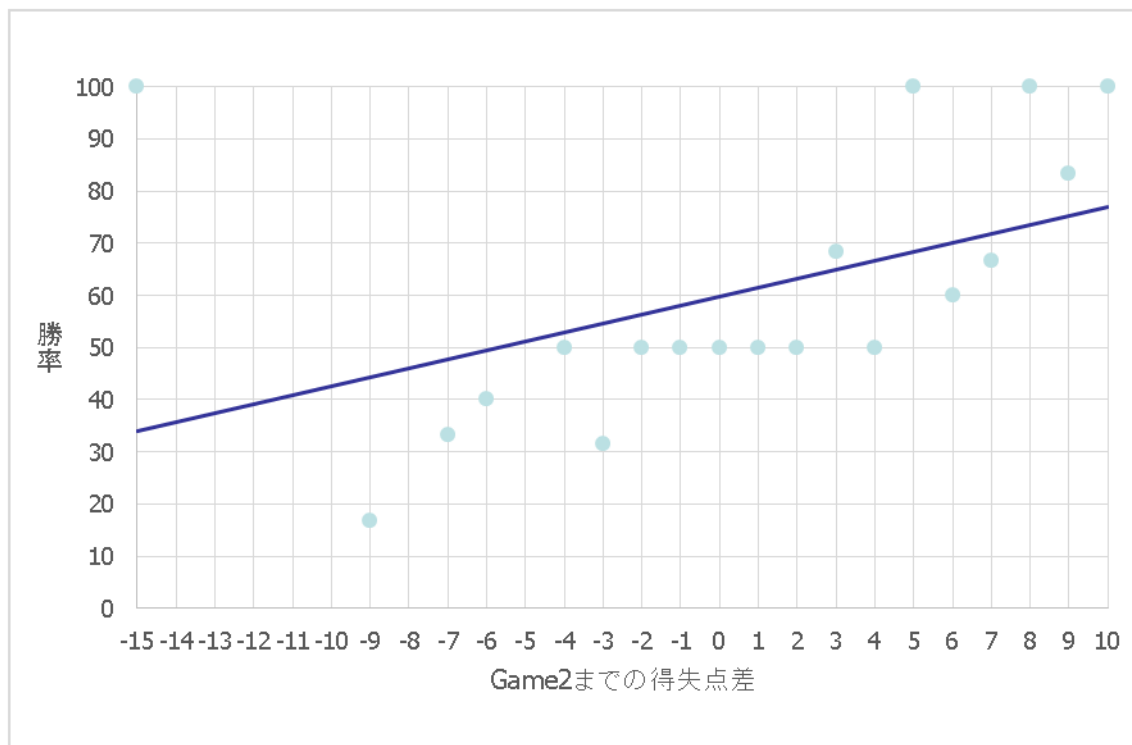


図15 単回帰分析の結果