レポート（テーマ２）の実例

テーマ２（統計的分析）に特有な事項（書式の３、４）についての参考情報を以下に記し、その後実例を示します。

３．方法

１）使用したデータについて

公開されているデータの例としては、例えば、以下のようなものがありますので、参考にして下さい。

●総務省統計局 e-Stat（政府統計のポータルサイト、
http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do）
●気象庁　各種データ・資料（http://www.jma.go.jp/jma/menu/menureport.html）
●スポーツ関連の各種データ、例えば、相撲レファレンス
（http://sumodb.sumogames.de/Default.aspx?l=j）等
●ソーシャルデータ（ツイッター等）

２）解析の方法について

データ分析については、以下の教材に、データ事例を含めて高校生向けの方法解説や分析の事例が多く載っていますので、参考にしてください。

●科学の道具箱
（日本統計学会統計教育委員会、https://rika-net.com/contents/cp0530/contents/）
●なるほど統計学園高等部
（総務省統計局、http://www.stat.go.jp/koukou/index.htm）

４．結果と考察

統計的な分析の事例として、最近実施されたデータ解析関連コンテストからの事例（高校生入賞作品）を紹介します。

●第1回データビジネス創造コンテスト（2014年、ツイッターの分析活用と新たなサービスのアイデア、
http://dmc-lab.sfc.keio.ac.jp/prize/）
○最優秀賞「花粉症患者のつぶやきを社会の意志決定に活かしたい」

次ページ以下に、統計的分析のレポートの実例として、大相撲の決まり手の数と力士の体重の相関について分析したレポートを示しますので、参考にして下さい。

作品名　大相撲の決まり手の数と力士の体重の相関についての分析

氏名　西川　哲夫

１．研究の要約

大相撲において、各力士の決まり手の数に力士の体重が影響を与えているかどうかを知るために、幕内 42 力士について体重と決まり手の数との相関を調べた。その結果、決まり手の数は、現在の体重との間には中程度の負の相関（-0.55）があり、各力士の初土俵から約２年の時期の体重との間には強い負の相関（-0.75）があることがわかった。また決まり手の数が多い力士は体重が軽い初土俵の頃に決まり手の数が多く、体重が増加するにつれて決まり手の数が減っていく傾向があることがわかった。

２．研究の動機と目的

大相撲の魅力は、その大きな体が生み出す迫力と共に、その多彩な技にあると思われる。大相撲には 82 手の決まり手があるという。最近、その技の多彩さが失われ、それは力士の体重の増加と関係があるということが新聞の記事で指摘されていた[1]。この記事で、つり出しや掛け技が減少し押し出しとはたきこみが増加していることがこの 40 年間で見られ、平均体重の増加と関係しているのではないかと指摘されていた。しかし、この記事では、幕内全体での平均体重や決まり手の頻度の変化をみているので、個々の力士について体重の大きい力士が決まり手の数が多いのかどうかはわからないと思われる。そこでここでは、力士の体重と決まり手の数とは本当に相関があるのかどうかを調べることを本研究の目的とした。

３．方法

（1）使用したデータについて

1）「日本相撲協会」サイトからの情報取得

日本相撲協会の番付表[2]から、平成 26 年 7 月場所の各幕内力士の力士プロフィールにアクセスし、平成 26 年 7 月場所時点での幕内力士名と身長及び体重情報を取得した。

2）「相撲レファレンス」サイトからの情報取得

過去の取組のデータは、「相撲レファレンス」[3]のサイトから取得した。

①　過去の体重情報

幕内力士について、「相撲レファレンス」の力士検索[4]で各力士の情報にアクセスした。各力士の過去の体重情報については、各力士の情報のページ下部の過去の場所の星取表にほぼ 1 年毎に記載された体重情報を取得した。

②　過去の取組の決まり手の情報

各力士の過去の取組の決まり手の情報は、各力士の情報のページにある「すべての決まり手を見る」のリンクのページから取得した。このページには、左側に各力士の過去の決まり手の統計情報（全決まり手数及び各決まり手の頻度情報）が記載してある。また、このページの右側には、各力士の過去の全ての取組について、勝ちの場合と負けの場合に分けて、決まり手ごとに取組情報を記載してある。取組情報としては、場所（年月）、日（何日目）、取組み時の地位、決まり手名称、相手の地位と四股名が、取組み毎に 1 行で記載してある。この取組情報のうち勝ちの場合の取組情報全ての行を取得してエクセルファイル上に記録した。

（2）解析の方法について

力士毎に、決まり手数と体重の相関を調べるためには、決まり手数を数える期間をどうするかということと、いつの時期の体重を用いるかが問題であると考えられる。期間が短いと十分な決まり手数がカウントできないだろうし、期間が長いと決まり手の傾向がその期間の間に変化していくだろうと考えられるからである。また体重も時期によって変化していく（一般的には増加していく）と考えられる。

ここではまず、期間、時期については詳細に扱わず、平均的な情報として決まり手数と体重の相関を調べることとした。すなわち、幕内全力士（42 名）の各々について、全決まり手数と平成 26 年 7 月場所の体重（現在の体重と呼ぶ）との間の相関を調べた。比較のために、決まり手数と身長との相関についても調べた。その後、決まり手数の多い複数の力士について、決まり手数と体重の、それらを調べる期間の長さや時期に対する依存性を調べた。その結果を踏まえ、幕内全力士（42 名）の各々について、全決まり手数と初土俵から約 2 年間の平均体重との間の相関を調べた。

1）現在の体重と全決まり手数の相関分析

平成 26 年 7 月場所の各幕内力士 42 人について、現在の体重と全決まり手数について、度数分布を求めた。また、現在の体重と全決まり手数の間の相関、及び身長と全決まり手数の間の相関を調べた。散布図を作成し、相関係数を計算した。

2）一定期間における決まり手数と体重の時間的推移の分析

次に、一定期間における決まり手の数と体重の時間的推移の様子を知るために、決まり手数が多い 1 名の力士（白鵬、決まり手数 39）について、過去の全取組情報から、異なる複数の長さの一定の期間ごとに決まり手数を集計した。ここで、勝利の取組数（以下では取組み数と称する）を、期間を表す指標として用いた。また、初土俵からの取組数を、時期を表す指標として用いた。期間の長さとしては、50、100、150 取組を用いた。また、50 取組毎の決まり手数の、初土俵からの積算値を算出した。さらに、ほぼ 1 年毎に記録されている体重の記録の時期を、初土俵からの取組数に換算した。以上から、50、100、150 取組毎の決まり手数と初土俵後の 50 取組毎の決まり手数の積算値及び体重の、時期（初土俵後の取組み数）による推移が求められた。

上記で得られた結果より期間の大きさを一つに決定し、さらに決まり手数が多く（34 以上）、かつ現在と若い時期の体重差が大きい 4 名の力士（安美錦、北大樹、日馬富士、豊ノ島）について、決まり手の数と体重の推移の様子を調べた。総勢 5 名の力士について、各期間における決まり手の数と体重の相関係数を計算した。上記の推移グラフと相関により、体重情報として、初土俵から約 2 年間の体重の平均値（初期体重と呼ぶこととする）を算出して以降の分析に用いることとした。

3）初期体重と全決まり手数の相関分析

幕内全力士（42 力士）について、初期体重を求め、初期体重について度数分布を求めた。各力士の全決まり手数と初期体重の散布図を作成し、相関係数を計算した。

4．結果と考察

（1）現在の体重と全決まり手数の相関分析

図 1 と図 2 に、平成 26 年 7 月場所の各幕内力士 42 人の現在の体重と全決まり手数の度数分布を示す。図 1 から、現在の体重の平均値、最低値、最大値は、ぞれぞれ、158.4 kg、130.0 kg、199.0 kg である。

現在の体重の最頻区間は、150～160 kg であり、若干上方に伸びているが左右対称に近い正規分布に近い分布をしている。図２から、全決まり手数の平均値、最低値、最大値は、ぞれぞれ、26.0、15、42 である。最頻区間は、15～20 であり、下方にはほとんど伸びておらず、上方に長く伸びた分布をしている。全決まり手数の最小区間と最頻区間がほぼ一致しており、全決まり手数の下限（15～20）の存在が推察される。これが幕内力士としての下限なのかどうかは、より下位の力士について調べてみる必要があるが、ここでは行っていない。

図３に現在の体重と全決まり手数の散布図を示す。図３に示すように、現在の体重の増加に伴って全決まり手数は減少する傾向にある。この傾向を定量化するために相関係数を計算してみると、-0.55 であった。この値は、相関が中程度はあるといえるレベルである。図４に現在の身長と全決まり手数の散布図を示す。相関係数は、-0.15 であり、現在の身長と全決まり手数の間には相関がないことがわかった。

「分析の方法」で述べたように、時期による決まり手数や体重の変化を考慮にいれた分析を次に行うために、以下では、異なる時期で一定期間での決まり手の数と体重を求め、その時間的推移について調べてみることとした。

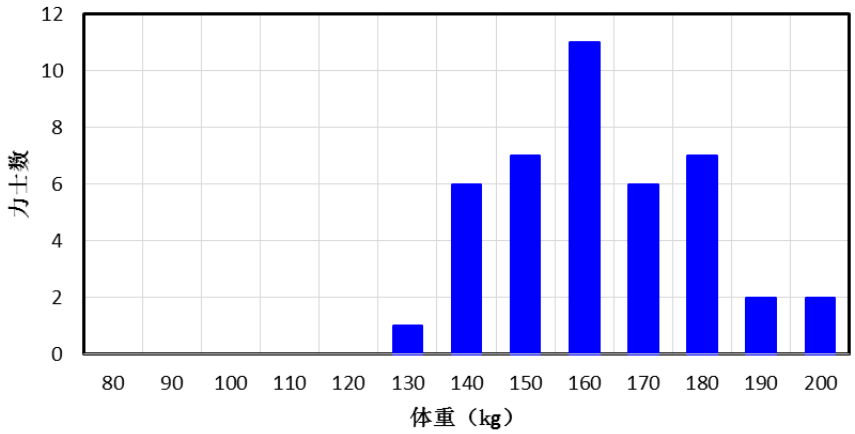

図１　幕内力士の現在の体重の度数分布

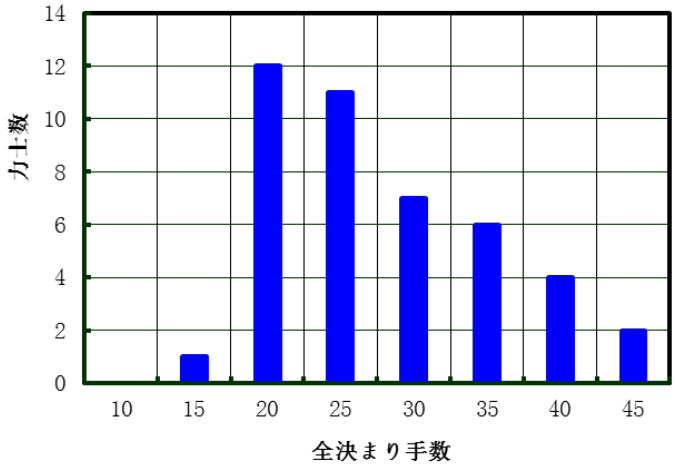

図２　幕内力士の全決まり手数の度数分布

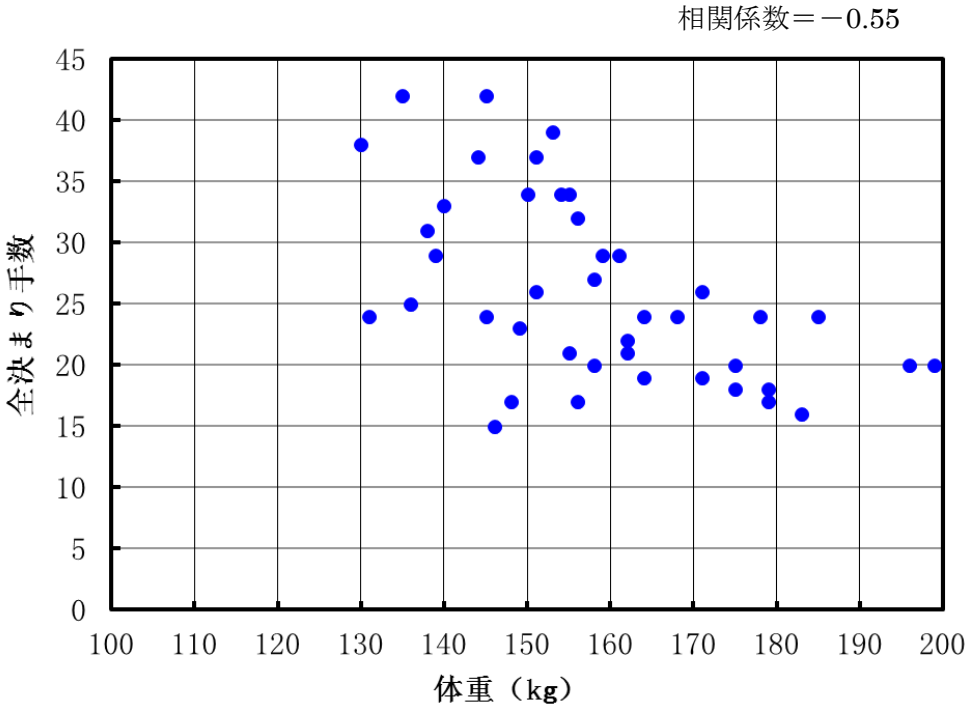

図３　幕内力士の現在の体重と全決まり手数の散布図

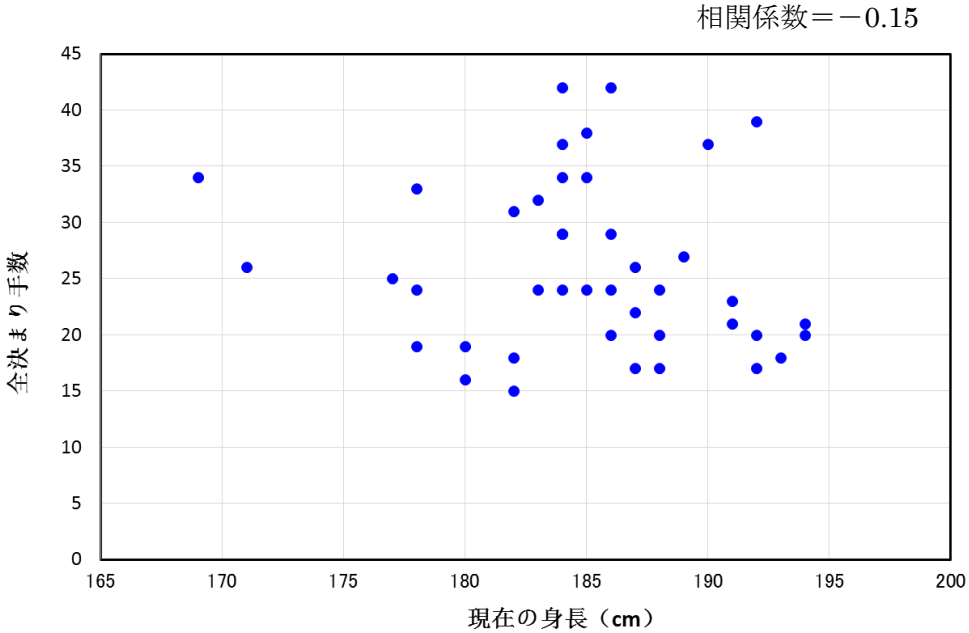

図４ 幕内力士の現在の身長と全決まり手数の散布図

（２）一定期間における決まり手数と体重の時間的推移の分析

白鵬の 50、100、150 取組毎の決まり手数と初土俵後の 50 取組毎の決まり手数の積算値及び体重の、時期（初土俵後の取組み数）による推移を、図５に示す。図５より、決まり手数は、取組み数の増加と共に減少した後増加し、その後再び減少することがわかる。集計の期間が 50 取組の場合は細かな変動が多いので、期間をより大きくしたほうが望ましいと考えられる。集計の期間が 100 と 150 取組の場合は、細かな変動もなくほぼ同じ傾向を示すので、ここではよりプロット数がとれる 100 取組を以降の分析に用いることとする。

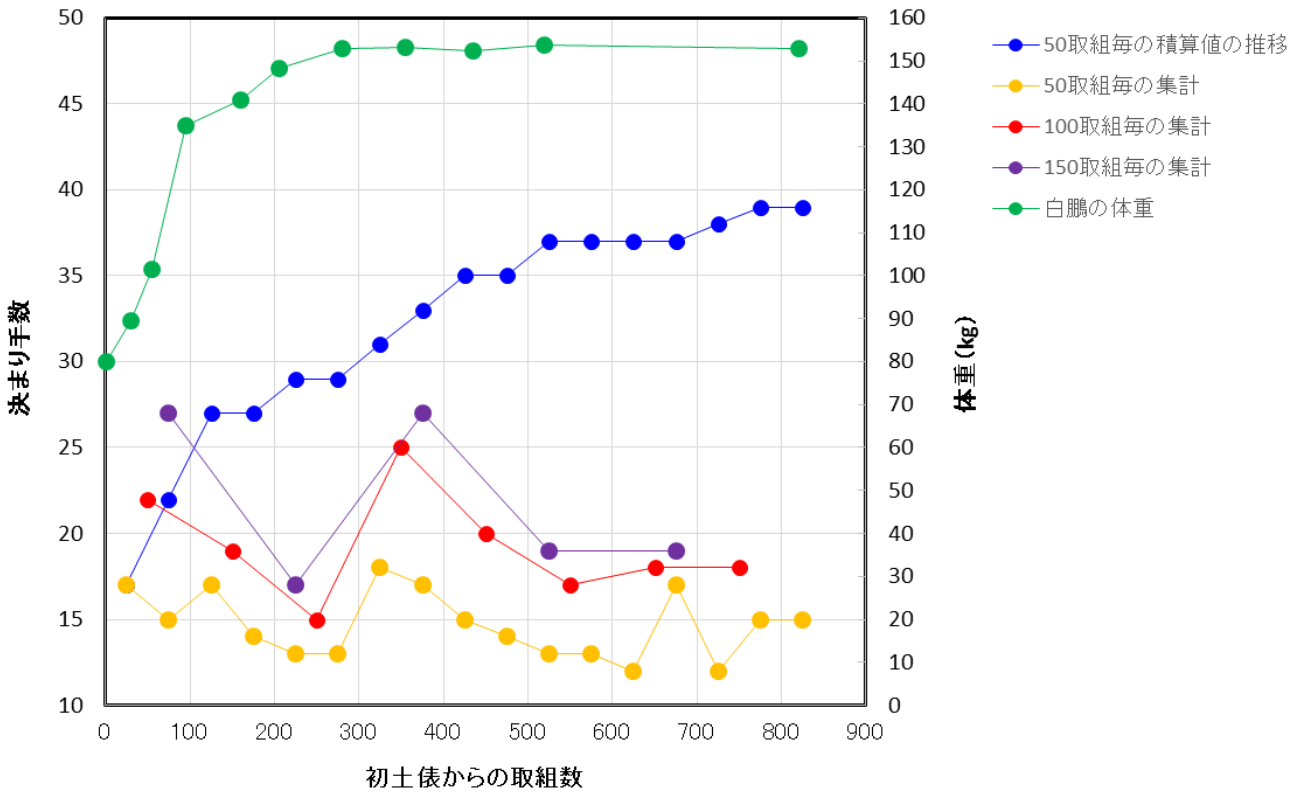

図５　白鵬の初土俵後からの決まり手数の積算値の推移と
50、100、150取組毎の決まり手集計数の推移、及び体重の推移

体重の推移と決まり手数の推移を比較すると、若い時期には体重は軽く、決まり手数が多い傾向にあることがわかる。最初の決まり手数の減少とその後の決まり手数の増加の時期は、それぞれ白鵬が横綱になった頃と優勝 10 回目の頃であり、多くの決まり手が必要ないほど充実した横綱昇進の頃と優勝を重ねるために決まり手数を増やしていた頃に対応しているのではないかと推定される。決まり手数の積算値は、2 回目の決まり手数のピークの頃に増加しており、最初のピークのときとは異なる決まり手を増加させていたことがうかがえる。

次に、ここで得られた「体重の軽い若い時期に決まり手数が多い傾向」にどの程度一般性があるのかを調べるために、決まり手数が多く（34 以上）、かつ現在と若い時期の体重差（約 40 kg 以上）が大きい 5 名の力士（白鵬、安美錦、北大樹、日馬富士、豊ノ島）について、同様な分析を行った。図６に上記 5 名の力士について、100 勝利取組毎の決まり手数と体重の推移を示した。力士によっては単調な減少ではないが、取組み数が増加するにつれて、体重の増加と共に、決まり手数は減少していく傾向にあることがわかる。このことを定量的に確かめるために、図７に各力士の各期間における体重と決まり手数の散布図を示した。相関係数は、−0.52 であったので、中程度の相関であることがわかる。

また図６で、各力士において、体重の軽い初土俵後の 100 取組ぐらいの期間における決まり手数が、最も多いことがわかる。このように、決まり手数の多い力士の決まり手数には、その力士の初土俵後の 100 取組の期間における決まり手数が最も寄与していることがわかった。そこで、決まり手数と体重の相関分析に用いる体重として、初土俵後の 100 勝利取組の期間における体重を用いることとする。100 取組は 2〜3 年に相当するので、ここでは初土俵後 2 年間における体重の平均値を用いることとした。

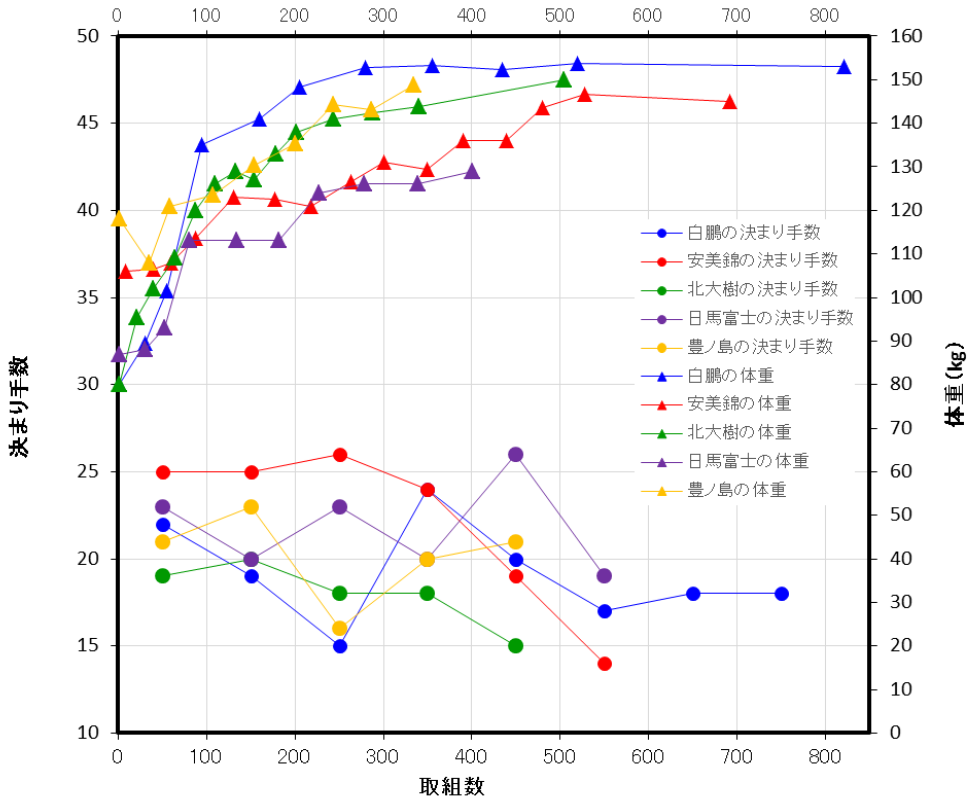

図６　５力士の100取組毎の決まり手数と体重の推移

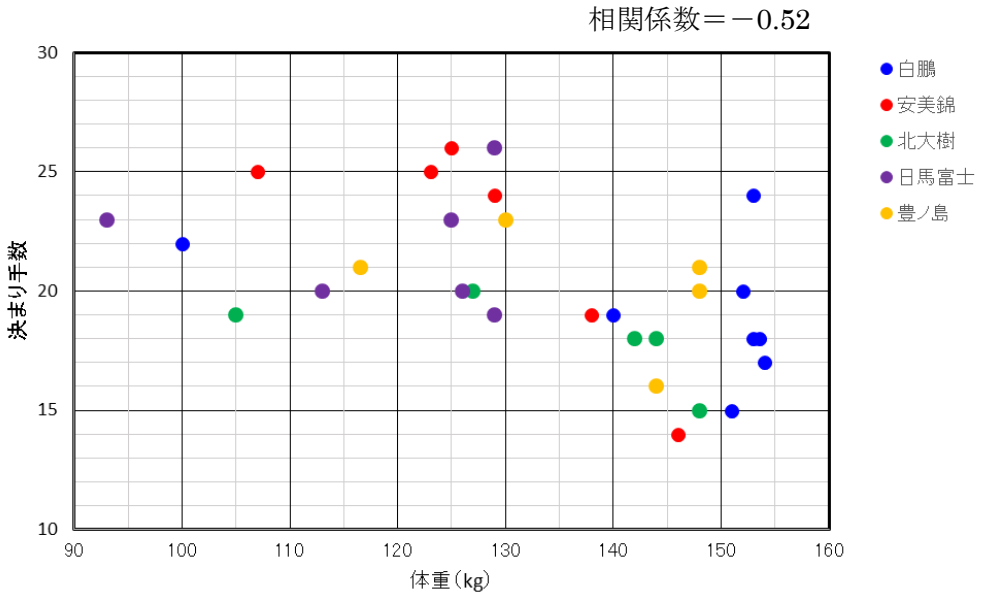

図７　5力士の100取組毎の体重と決まり手数の散布図

（３）初期体重と全決まり手数の相関分析

幕内全力士（42 力士）について、平成 26 年 7 月場所の体重（現在の体重）と初土俵から約 2 年間の平均体重（初期体重）の度数分布を図８に示す。初期体重については、平均値、最低値、最大値は、ぞれぞれ、128.3 kg、89.4 kg、164.0 kg であった。初期体重は、現在の体重と比べると、平均で 30.1 kg、最低値で 40.6 kg、最大値で 35.0 kg 小さいことがわかる。図９に、幕内全力士について、初期体重と全決まり手数の散布図を、現在の体重と全決まり手数の散布図と共に示す。図９に示すように、初期体重を用いた方が、全決まり手数のばらつきが減少しよりシャープな相関が観察される。相関係数は、現

在の体重の場合の-0.55 に対して、初期体重の場合、-0.75 となり強い相関が観察された。現在の体重との相関の場合、150 kg 付近での全決まり手数のばらつきは 50 kg 以上あるが、初期体重との相関では、全体重区間で全決まり手数のばらつきは 40 kg 以内に収まっている。このことによって相関係数の絶対値が大きくなったと考えられる。

ばらつきの減少の原因は、以下のように考えられる。現在の体重が 150 kg 付近の力士は、初期体重が大きい力士と小さい力士に分かれており、初期体重が小さいほうが全決まり手数が大きいため、初期体重で相関をみると、150 kg 付近のばらつきが解消されて、よりばらつきの小さい負の相関が観察されるのだと考えられる。以上示したように、初期体重と全決まり手数の間には強い相関があることがわかった。

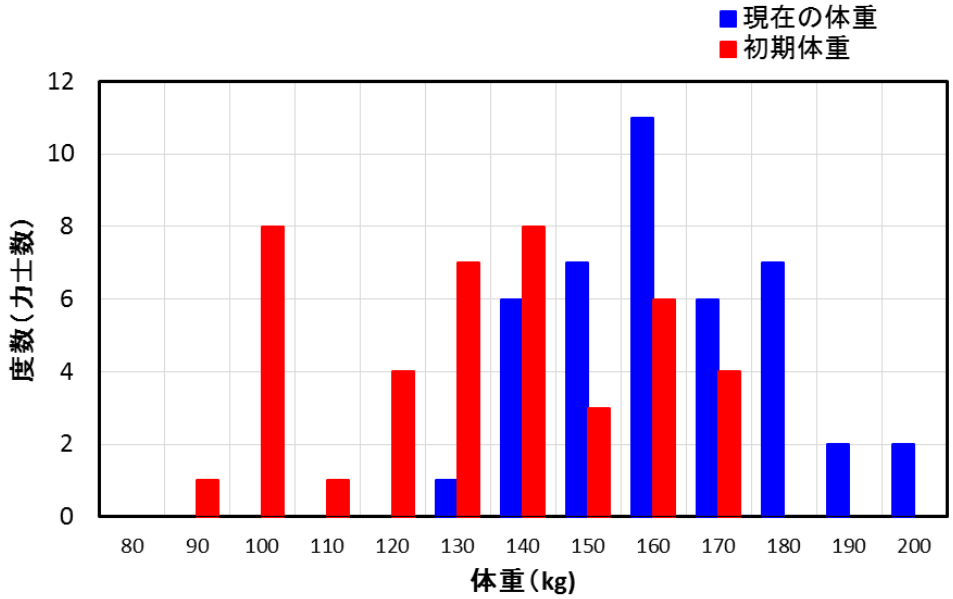

図 8 幕内力士の体重分布（現在の体重と初期体重）

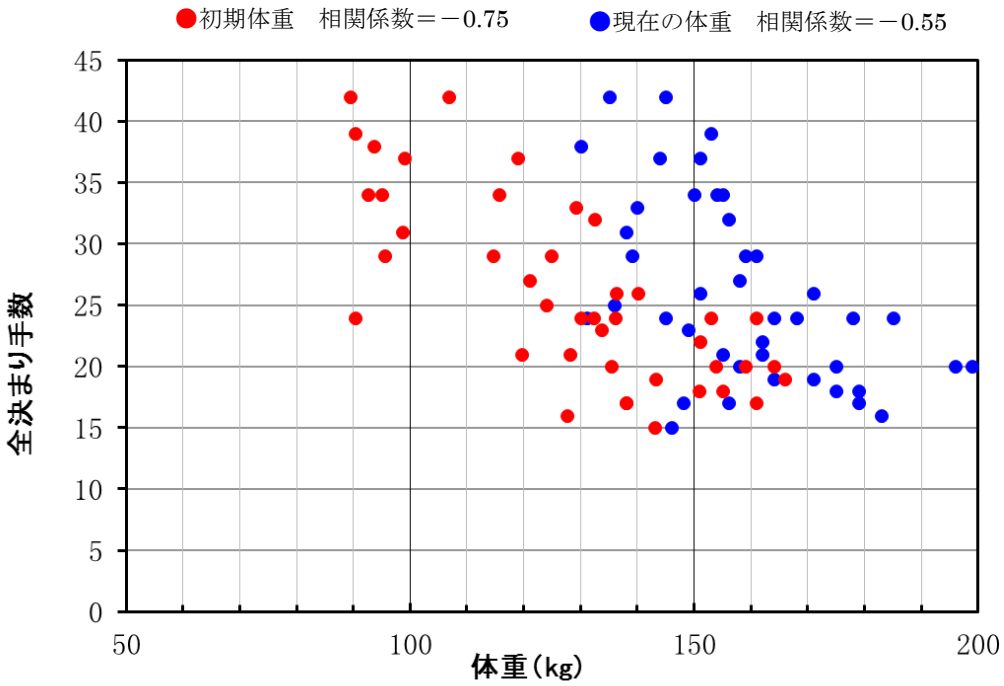

図 9 幕内力士の体重と全決まり手数の散布図
（現在の体重と初期体重との比較）

５．結論と今後の課題及び感想

１）結論

大相撲において、各力士の決まり手の数に力士の体重が影響を与えているかどうかを知るために、幕内 42 力士について体重と決まり手の数との相関を調べた結果、以下のことが得られた。

①幕内力士の全決まり手数の平均値、最大値、最低値はそれぞれ 26.0、42、15 であった。全決まり手数の最小区間と最頻区間がほぼ一致し、全決まり手数の下限（15～20）の存在が推察された。

②現在の体重と全決まり手数には中程度の相関（相関係数 = -0.55）があった。

③初期体重と全決まり手数には強い相関（相関係数 = -0.75）があった。

④全決まり手数の多い力士は初土俵から約 2 年間における決まり手数が最も多く、体重の増加と共に次第に減少していく傾向にあった。これは、軽量の力士が幕内で勝つためには、決まり手の多さが必要であり、体重が増加するにつれて少ない決まり手で勝てるようになったものと思われる。

結論として、力士の体重と決まり手の数との間には実際に相関があり、体重として初期体重を考えたときに強い相関があることがわかった。

２）今後の課題

今後検討すべき課題として、以下のような課題が考えられる。

①今回の分析では、決まり手の具体的な中身については触れていないので、具体的な決まり手に着目した分析を行ってみたい。

②決まり手の種類は、力士本人の体重だけではなく取組相手の体重も関係すると思われるので、取組相手との体重差を考慮した分析を行ってみたい。

③決まり手の数には、力士の経歴（例えばモンゴル出身など）も影響すると思われるので、力士の経歴など体重以外の要因を考慮した分析を行ってみたい。

④今回幕内力士のみを分析の対象にしたので、下位の力士や過去の力士についても調べてみたい。

３）感想

新聞の記事をきっかけにテーマを考えてみたが、取り組みのデータをいかにして収集するかが課題であった。幸い「相撲レファレンス」という非常に有用なサイトが見つかり、分析を始めることができたが、データの収集というのは統計的な分析には一番の問題になるのものだと思った。データが初めからまとまった形で収集できたわけではないので、必要なデータを収集するのに苦労した。エクセルの様々な機能を用いて分析を行うなどデータ分析に苦労した。

目的としたことは、新聞で見つけた当たり前と思えるようなことをデータで実証することであったが、実際に分析を行うにつれて関連する様々な新しい疑問や課題が浮かんできて、非常に興味深く感じ、研究はやってみればおもしろいものだと思った。大相撲の力士自体にも興味が沸いてきた。今後身の回りの当たり前と思えるようなことにも、気をつけていきたいと思う。

６．参考文献

[1] 朝日新聞 2013 年 12 月 14 日朝刊、「力士大きく決まり手少なく」

[2] 日本相撲協会の番付表（https://www.sumo.or.jp/ResultBanzuke/table/）

[3] 「相撲レファレンス」(http://sumodb.sumogames.de/Default.aspx?l=j)

[4] 「相撲レファレンス」の力士検索（http://sumodb.sumogames.de/Rikishi.aspx?l=j）