

《論文》

立教大学生のコンペティション参加への取組みと課題  
—スポーツデータ解析コンペティションの参加を例に—

Efforts and Challenges of the Rikkyo University Students  
to Competitions Participation

大橋 洸太郎 Kotaro Ohashi  
大川内 隆朗 Takaaki Ohkawauchi  
小野寺 剛 Tsuyoshi Onodera  
丹野 清美 Kiyomi Tanno  
山口 和範 Kazunori Yamaguchi

In this report, we introduce initiatives of the Rikkyo University's Center for Statistics and Information, CSI initiatives, (CSI). For two years, CSI has supported students participating in competitions. Several problems have been identified. For solving these problems, we discuss the abilities required from the students, kind of support needed by students.

Key words : Statistics Education, Competitions, Sports Data Analysis

キーワード : 統計教育,コンペティション,スポーツデータ解析

I はじめに

近年、データ解析を主体としたコンペティションが多く開催されている。例えば経営学の分野では、経営科学系研究学部連合協議会が主催するデータ解析コンペティションがあり、毎年全国から多くの企業や大学からの参加者が集まり、それぞれの解析成果を競っている。このコンペティションでは学生の参加を受け入れており、ID-POS データやインターネットサイトの閲覧履歴といったデータが貸与され、それらを用いた分析が行なわれている。コンペティションの成果の例として大橋・豊田・久保(2012)では、株式会社マルイより与えられた3店舗1年分のID-POS データから来店回数に注目をし、まだ一度しか来店していない顧客の中から、複数回来店していたリピーターに近い購買行動を行っていたものを今後も来店する可能性の高い有望な顧客として抽出するモデルを考案している。

学生の参加を受け入れるコンペティションはこの他にも存在している。日本統計学会スポーツ統計分科会が事務局を務め、株式会社日本科学技術研修所の協賛の元、株式会社データスタジアムの持つ野球とサッカーに関するデータが提供されるスポーツデータ解析コンペティションも、今日多くの参加者を集めるコンペティションの1つに数えられる。スポーツデータ解析コンペティションでは、例えば2015年には、野球部門、野球トラッキング部門、サッカー部門、サッカートラッキング部門が用意され、それぞれの部門について全国から学生の参加があった。本コンペティションは2015年度で第5回目の開催となっている。野球のデータを用いた参加者の分析例としては、荒木・竹村(2014)のように、状態空間モデルを用いて打率がどのように推移していくのかを解析したものや、君島ら(2015)のように大谷翔平選手に分析対象を絞り、CDMCA という手法を用いて今後どのような選手と類似した成長を見せる可能性があるのかを分析したものがある。また永田ら(2015)のように、一般化加法モデルを用いストリートボールにコンタクトする確率を計算したものや、石原ら(2015)のように、あまり野球に詳しくない主婦の目線から野球中継を捉え、選手が活躍するチャンスが発生する確率を計算しテレビ画面に表示することで視聴率の向上に資することができないかを考えた発表など、様々な観点に注目した発表があった。サ

サッカーのデータを扱ったものにも、例えば徐ら（2014）のような、どのような攻撃プレーが得点に対してより効果的であるのかを行動の最適化計算をすることで表現したものがみられた。

これらのコンペティションの特徴として参加学生に数学や理学を初めとする自然科学系の所属だけでなく、社会学や経営学、文学といった社会科学系の所属の学生がいるという学際的な出自の存在が挙げられる。中には医学や薬学部の出身者もみられ、同じデータに対する多彩なアプローチの方法がみられる点も特徴である。データ解析コンペティションは、学生を受け入れる門戸が広いことが特徴である。

参加学生はデータ提供元から申請した区分のデータを貸与され、一定の期間の中で分析を行い、成果を報告する。スポーツデータ解析コンペティションの成果報告会は、2014年度、2015年度は共に12月下旬に開催され、2015年度では100名を越える出席者があった。これまでの出席者の中には参加チーム以外の大学教員や学生だけでなく、研究所研究員、スポーツデータアナリスト、スポーツをテーマにしたアプリケーションソフトウェアの開発会社や人材開発会社の社員、フリーライターといった業種の人々がいた。経営科学系研究学部連合協議会が主催するデータ解析コンペティションにおいても、成果報告会では研究所研究員や、データ分析や調査を主体とする企業の人々が参加し、これらのコンペティションは業種においても多彩な人々が注目している点が特徴的である。

## 1. コンペティションに参加する意義

前述のように様々な分野の学生を受け入れ、多彩な業種の社会人が注目するコンペティションに学生が参加する意義は大きい。その理由としては第一に、扱うデータの質が非常に高いことが挙げられる。コンペティションでは、学生一人の手では入手が困難な現実の店舗のID-POSデータ、現実の野球チームの年間の全試合、全投球、全バッティングの詳細なデータ、秒単位で位置を記録したサッカーチームのメンバーすべての座標データといったものを扱うことができる。そしてこれは就職を期に実社会でこのような現実的なデータに直面する可能性のある学生にとって貴重な体験となる。また第二の理由としては、データ解析コンペティションでは数十万ケースを越えるビッグデータを扱うことが多いため、データの中の何に注目し、何を切り取り、自身の目的に沿う形にデータを成形していくかという経験ができることが挙げられる。今後は益々扱わなければならないデータの量が増えていくことが社会的に予想されるため、早いうちにこのようなデータと向き合う機会ができることも貴重である。

コンペティションにおける評価の基準は、統計的な分析手法の適切さや新規性、実質的な成果の有用性、着眼点の面白さに加えて、発表者のプレゼンテーションの質や発表資料がどれだけ整理されているか、提案手法の実質的な応用可能性といった点も加えられていることが多い。評価の基準はデータの提供元に対して、どれだけ有益な情報をフィードバックできるのかという側面が含まれているためである。このような要請に応じていくことで、ただ与えられたデータを分析するだけでなく、ニーズに応えた成果を提供していくという学生のマネジメント能力の向上にもコンペティションは資する。以上より、コンペティションの参加は学生にとって非常に意義深いと考える。

## 2. スポーツデータ解析コンペティションの流れ

図表1はスポーツデータ解析コンペティションの2014年から2015年までの2年間のおおよその流れを示したものである。本コンペティションは2014年度、2015年度共に夏季（2014年度：7月27日、2015年度：7月19日）に発会式と説明会が開催されてきた。参加に興味のある人々がこの会に参加し、その後、事務局へ参加の応募を行うこととなっている。応募の形式は電子メールでの投稿であり、その際に各年度に用意された部門の中か

ら参加希望をするものを選択する。そしてデータ管理の責任者として指導教員を 1 名以上付けることとなっている。また応募の際、構造方程式モデリング (SEM) を用いた分析を行い、SEM 因果分析特別賞の審査候補として参加するかどうかを選択することとなっている。SEM による分析を行う際には、株式会社日本科学技術研修所が貸与する分析ツール JUSE-StatWorks/V5 を用いることが可能である。応募時の制限としては 2015 年度の現在までは、学生であっても一般企業に勤めている者は応募できないことである。応募が受理された後、送付先の住所にデータの入ったディスクが送られ、その時点から分析が開始される。

成果報告会は冬季 (2014 年度 : 12 月 26 日, 2015 年度 : 12 月 26 日) に行われ、この会における発表を元に、各部門の優秀賞, 最優秀賞, データスタジアム特別賞, SEM 因果分析特別賞の授与チームが決定される。受賞チームは翌年の春季 (2014 年度 : 3 月 12 日, 2015 年度 : 執筆時には未定) に受賞者講演会を行い、この時に賞状等が授与される。受賞者講演会までに、成果報告会の分析を更に進めることも許されている。

またこの他、2014 年度、2015 年度には日本統計学会春季大会での参加者によるポスター発表の場が設けられていることや、受賞者講演会の後に論文集の特集号への投稿の場が設けられている。

時期	流れの概要
7 月下旬頃	コンペティション発会式, 説明会
8 月~12 月	応募, 分析期間
12 月下旬頃	成果報告会
翌年 3 月上旬頃	受賞者講演会

図表 1 : 2 年分のスポーツデータ解析コンペティションのおおよその流れ

## II CSI におけるコンペティションへの取組と課題

### 1. CSI における取組み

立教大学社会情報教育研究センター (Rikkyo University Center for Statistics and Information, CSI) では、2014 年度より立教大学の全学部生を対象に、スポーツデータ解析コンペティションの参加者を募り、CSI 統計教育部会教員が指導教員となってチームによる参加を促す事業を行っている。この事業は今年で 2 年目となり、1 年目で 1 名、2 年目では 5 名の応募者があり、実際にコンペティションに参加し、データ分析に関わり、成果報告会で発表までを行った。発表テーマとタイトルは以下の図表 2 の通りである。

年度	参加部門	発表タイトル
2014	野球 (SEM 因果分析特別賞応募)	観客数が増えたら野球選手への心理的影響 ~弱くても勝てるか~
2015	サッカートラッキング	選手交代が与える影響について ~流れを変えることはできるのか~

図表 2 : 本学学生のスポーツデータ解析コンペティションへの参加状況

現在の CSI の取組では、7月の発会式を受けて参加学生の募集を開始している。参加学生が集まり、データや参加部門の内容の説明が学内で行われ、この時にチームの編成等が行われている。参加学生の学年や学部は問わず、興味を持って分析に従事できる者を募集し、指導にあたった。その結果、2014、2015年度共に成果報告会で発表することができ、2014年度では、日本統計学会春季大会でのポスター発表を行うことができた。

## 2. 本学学生が抱える課題

図表2の成果はCSI教員の指導の下、双方共に学生の手によるものであった。ただし、始終順調に分析が進んだわけではなく、年度や時期によって惹起した様々な問題を解決して漕ぎ着けた結果であった。このため、本誌では、この2年間における分析の進捗状況を記載し、今後のためにこれらの結果から考えられる本学学生のコンペティション参加に係る課題とその対策について考察していきたい。また、チームの学生代表者に成果報告会後に行った各問100字以内の自由記述型の質問結果についても報告する。質問文は図表3の3問であった。

質問	質問内容
1	コンペティションに参加した動機についてお書きください。
2	コンペティションに参加して得られたことがあればお書きください。
3	コンペティションにおいて大変だったことがあればお書きください。

図表3：自由記述質問の内容

これらの質問のうち、課題点を挙げる上での示唆となるものは質問番号3である。この質問に対する回答は2014年度、2015年度のチーム代表者の意見として以下の2つがあった。

年度	回答内容
2014	分析に関してエラーが何度も出たこと。一見、関連のあるデータでも分析を続けると全く無関係である場合が多々あった。発表に間に合わないのではないかと不安になり、何度もコンペティションを辞退しようと考えていた時期もあった。しかし、先生方から様々なアドバイスを頂いたおかげで、満足行く発表へと繋がられた。
2015	知識を補うことが大変でした。統計素人の私にとって今回の取り組みはわからないことの連続。調べたり人に聞いたりすることも多く、分析を進めるのに時間がかかってしまいました。未知のものを1から学んでいく、そんな大変さを感じながらの1ヶ月だったと思います。

図表4：コンペティションにおいて大変だったこと（質問3）

図表4の内容からまず挙げられるものは、限られた期間の中で分析結果を出さなければならぬという状況の中で、統計的な素養の不足を不安視する声が大きいのということである。意欲があれば本学の学生が学部や学年を問わずに参加できるという受け入れ方針は、学生に質の高いデータに触れさせるという意義からは非常に良い点ではある。しかしなが

らこれまでに統計学を専攻してきていないことがネックとなり、限られた期間の中でどのようにデータを加工し、どのように分析をして成果の着地点を決定していくかという具体的な行動を起こしづらいという現状が見て取れる。また統計を専門とする大学院生ではなく、他の専攻を持った学部生という立場上、本コンペティションのみに時間を割くことはできず、授業期間の合間を見て慣れない作業をしなければならないという点も留意すべきであると考えられる。

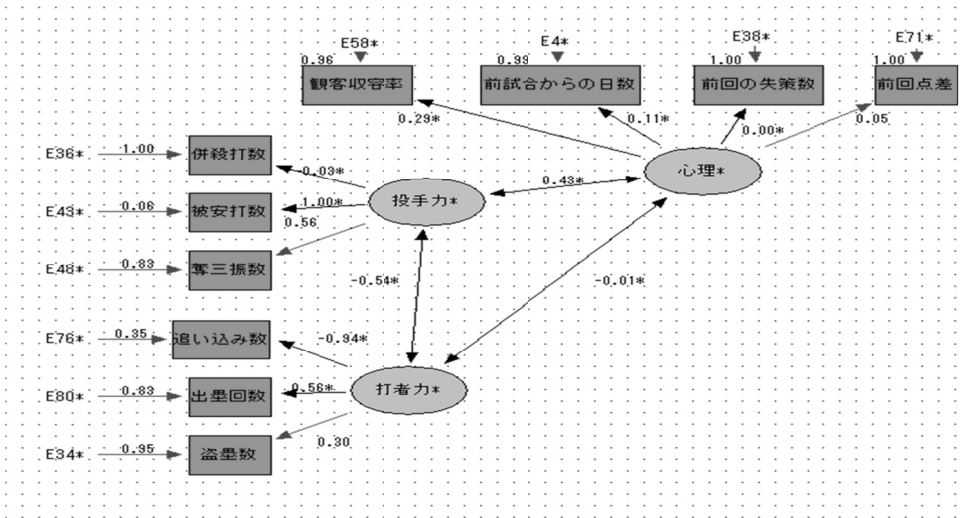
年度	使用した分析手法	分析ツール
2014	構造方程式モデリング	Microsoft Excel 2010, JUSE-StatWorks/V5
2015	散布図, カイ二乗検定	Microsoft Excel 2010

図表 5 : 本学学生の使用した分析手法と分析ツール

図表 5 は、本学学生の成果報告会までに使用した分析手法とデータの加工と、分析に用いたツールを示している。2014年度はSEM 因果分析特別賞に応募し、JUSE-StatWorks/V5を用いて構造方程式モデリングを行った。野球部門のデータをもとに、チームの投手力と打者力に、前試合での失策数や前試合での点数の開きに加え、会場にどれだけの観客がいるかといった心理的要因がどれだけ影響しているのかを把握するための相関モデルを作成した。以上の詳細は安池ら (2015) にまとめられている。

2015年度では、散布図を用いてサッカーコートにおける前半と後半の選手の位置を把握した後に、前半と後半でディフェンスの人数に対するオフENSEの数的有利な状況がどれ程の頻度で起きていたのかをクロス集計表でまとめ上げ、このクロス集計表についてカイ二乗検定を行った (山口ら, 2015)。

貸与された JUSE-StatWorks/V5 を除くと、学生達は Microsoft Excel を用いてデータハンドリングで行うことが精一杯であり、一括で大規模に処理するようなデータの加工に困難さを覚える中で作業を行っている様子が窺えた。分析手法については SEM 因果分析特別賞に応募した 2014年度は構造方程式モデリングに限定されていたため比較的分かりやすかった。2015年度では、数ある統計手法の多くを知らない状態でデータの加工に取り組み、指導教員との面接を経て最終的にクロス集計表とカイ二乗検定に落ち着いた。2014年度に CSI チームが作成した構造方程式モデリングのモデル図を図表 6 に、2015年度のチームが作成したクロス集計表を参考例として図表 7 に記載する。



図表 6 : 2014 年度 CSI チーム (SEM)

湘南の戦況	説明	前半	後半
攻勢	攻めと守りの人数が同等以上でシュートの隙が生まれやすい状況	1287 (0.51%)	1242(0.49%)
中間	若干攻め手にかけるが攻勢にも転じやすい状況	21190(0.53%)	18789(0.47%)
劣勢	攻め手が圧倒的に少なくシュートを撃ちにくい状況	9362(0.57%)	6959(0.43%)
攻め手 0	湘南が攻め入っていない状況	37333(0.43%)	50216(0.57%)

図表 7 : 2015 年度 CSI チーム (クロス集計表)

### III 対策とまとめ

#### 1. 本学学生の抱える問題への対策

以上を受け、本学の参加学生の特徴を考慮した上での今後の対策としては、第一にデータハンドリングについてのサポートが必要ということである。統計教育部会では、毎年統計教育に係るセミナーを開催しており、単純集計や SPSS や Amos といったソフトウェアの使い方についてはオンデマンド形式で PC から視聴できる体制が整っている。このコンテンツの一環として、表計算ソフトウェアである Excel や、簡単なプログラミング言語であり、統計解析環境でもある R といったソフトウェアをデータハンドリングのツールとして扱う講座があると良いと考えられる。このような講座はコンペティションに参加する学生達に資するものとなるだけでなく、データの加工に関して同様な悩みを持つユーザーに役立つコンテンツとなることが予想される。

第二の対策としては、データマイニングをテーマとした分析手法のオムニバスのなサポ

ートが必要であることが考えられる。様々なニーズに応える形で提案されてきたデータ解析手法の中で、大きなデータを扱う場合によく用いられる手法にどのようなものがあるのかを俯瞰的に知る機会があれば、自身のテーマに沿った分析手法を見つけ出す助けになる。こちらについてもセミナー形式で展開し、参加学生が自由に閲覧できる状態になると、より効果的であると考えられる。

## 2. まとめ

ここでは最後に、図表 3 における質問 1、質問 2 の回答について記載する。

年度	回答内容
2014	<b>セイバーメトリクスに興味があり参加した。</b>
2015	<b>実際にプロの現場で使われるようなデータを扱えることと、そのテーマが私の好きなスポーツだったため興味を持ちました。統計に関しては素人で不安もありましたが、勉強をする良い機会だと思い参加を決めました。</b>

図表 8：コンペティションに参加した動機（質問 1）

年度	回答内容
2014	<b>順序立てて分析を行う姿勢。「仮説を立て統計手法に則して結果を導き出す」という一連の流れを理解することは、今後の課題解決へと活かしていけるのではと考えている。</b>
2015	<b>実践的なデータに触れることできたのは本当に良い経験でした。ファイルを開くのも大変なほど膨大なデータを編集し分析することは、知識だけでなく根気も養われたと思います。</b>

図表 9：コンペティションに参加して得たもの（質問 2）

図表 8 は質問 1 に対する回答である。コンペティションの参加の意義について述べた箇所にもあったように、プロの現場で用いられるようなデータに触れられる機会は貴重である。また、セイバーメトリクスやスポーツに興味のある学生達が参加を希望していることから、分析や統計に関わる部分のサポートを充実させ、このように興味を持った学生を広く受け入れられるような体制を取っていくことが今後も CSI の取組を続けていく上で重要であることが窺える。

図表 9 は質問 2 に対する回答である。成果報告会まで行うことができた 2 つのチームでは、今回の経験が今後活かしていけるという様子と達成感が共に窺える。2014 年の参加学生は、このコンペティションの経験を就職活動で話すことにより、スポーツ関連企業のインターンシップを得ることができたことを後に報告しており、学部生としてコンペティションに参加したことの具体的な成果となった。質問 1、質問 2 の回答からも、統計に関する膨大なデータの処理や統計解析に不慣れなことに対する不安、そして順序立てて分析を行っていく姿勢に対するコメントが見受けられる。この点からも、今後の CSI の対策が重要であるといえるだろう。

本誌では、これまでコンペティションに学生が参加することの意義や、スポーツデータコンペティションの流れ、そして CSI における学生のコンペティション参加への支援事業と現状について述べてきた。そして 2 年間に渡る支援事業の中で浮かび上がってきた課題とその対策について述べた。また併せて参加学生の声に掲載することで、コンペティシ

ン参加の意義や実際に参加した上での課題点について浮彫りにすることができたと考える。今後のより良い支援体制構築に資することができるよう、一層の努力をしていきたい。

## 参考文献

- [1] 荒木優・竹村彰通, 2014, 『状態空間モデルを用いたプロ野球における打率推移の解析』, スポーツデータ解析コンペティション第4回成果報告会.
- [2] 石原渚・圓城寺啓人・大屋拓磨・酒折文武, 2015, 『主婦が料理しても手を止める!? 魅力的な野球中継〜ロジスティック回帰を使った得点率の推測〜』, スポーツデータ解析コンペティション第4回成果報告会.
- [3] 徐広孝・大澤啓亮・見汐翔太・安藤梢・熊谷紗希・猶本光・横尾智治・西嶋尚彦, 2014, 『サッカーの攻撃プレー分析』, スポーツデータ解析コンペティション第4回成果報告会.
- [4] 君島亮・齋藤秀哉・竹森悠渡・中嶋雅彦・西塚真太郎・酒折文武, 2015, 『大谷翔平はどんなバッターになるのか?—最新手法 CDMCA を用いた若手選手の発掘—』, スポーツデータ解析コンペティション第5回成果報告会.
- [5] 永田大貴・大石惇喜・樫山文音・早瀬亮・南美穂子, 2015, 『一般化加法モデルを用いたストレートにおけるコンタクト確率の解析—ノビの正体とは?—』, スポーツデータ解析コンペティション第5回成果報告会.
- [6] 大橋洸太郎・豊田秀樹・久保沙織, 2012, 『有望な顧客の分類と特定—ランダムフォレストとゼロ過剰ポアソンモデルを利用した ID-POS データの分析—』, オペレーションズ・リサーチ 56 (2), 71~76.
- [7] 山口拓哉・奈良一毅・内藤優夏理・戒晴瑠・濱田薫・大橋洸太郎・小野寺剛・丹野清美, 2015, 『選手交代が与える影響について〜流れを変えることはできるのか〜』, スポーツデータ解析コンペティション第5回成果報告会.
- [8] 安池美紀・大橋洸太郎・丹野清美・大川内隆朗, 2015, 『観客数をもたらす野球選手への心理的影響—弱くても勝てるか!?—』, 統計数理研究所共同研究レポート 334, スポーツデータ解析における理論と事例に関する研究集会第2巻, 43~46.