

# ロジスティック回帰による課題提出ログデータの解析

中村 永友<sup>1</sup>石川 千温<sup>2</sup>渡辺 慎哉<sup>3</sup>小池 英勝<sup>4</sup>

## 要 旨

札幌学院大学の情報教育は、コンピュータ・リテラシの習得と、ワードプロセッサや表計算ソフトウェアのスキルを身につけるために数多くの課題を課している。課題の成果は単位取得に直結している。このようなスタイルの授業を始めて以来問題となっているのは、課題の進捗状況を能動的に把握しない学生が散見されることである。そのような傾向の学生は途中で脱落したり、最後まで出席しているにもかかわらず残念にも単位未習得となるのである。そのようなにならないような策や、脱落防止指導のための基礎的資料を作るための研究を行った。課題提出日の集計を通して単位取得者と未取得者の違いを明らかにし、さらに課題合格状況のログデータを分析することで、その方向性を探った。

**キーワード：**ロジスティック回帰モデル、コンピュータ・リテラシ教育

## 1 はじめに

札幌学院大学におけるコンピュータ・リテラシ教育は1997年より実施している。1年生の前後期に、主としてメール、ワードプロセッサ、表計算、プレゼンテーションの各ソフトウェアの習得を目的としたカリキュラムを実施している（小池・中村・石川，2010）。2014年度時点で、前期の授業ではワードプロセッサのMicrosoft Office Word 2013を、後期の授業では表計算ソフトウェアのMicrosoft Office Excel 2013を中心に行っている。これらのソフトウェアをある程度使いこなし、スキルをつけさせるためには、類似の内容を含む多くの演習課題を行う必要があると考え、これらのソフトウェアの課題を前期に17、後期に16を課している。これら以外にMicrosoft PowerPoint 2013やメールの課題をあわせて10程度課している。さらに実力を確認するため、前期と後期にWordとExcelに絞って「単元まとめ試験」を各3回実施している。受講者を750名と仮定して、以上の全課題に対して全員が提出した場合、年間で単純計算して、のべ約3万5千

以上の課題を評価しなければいけない。この膨大な提出課題を教員と複数のTA（ティーチング・アシスタント）が公平な採点することは事実上困難である。この問題点を克服するために、WordとExcelの課題に対して、学生自らが課題の可否を確認するための課題採点ツール（以下、ツール）を2000年以来開発し、2005年度から本格的に運用し（石川・皆川・渡邊・中村・小池・梅田，2008；Koike, Akama, Ishikawa, Nakamura & Watanabe, 2007）、大規模な授業運営を実現している。このツールは採点した課題の完成度に関する履歴データをしかるべき場所に記録していて、不正行為防止の機能もある。

本論文は、2013年度の前後期の各課題の提出日、単位取得の有無、単元まとめ試験の受験の有無等の情報を分析することで、ツールによる教育効果の測定や、単位未取得とならないようにするための方策や指導法を確立するための基礎的な資料を作成するための方向性を探るものである。とくに、締め切りまでの課題の提出日の分布による比較、課題締め切り日における課題合格数による単位取得確率の推定、この2点について分析を進め、学生の行動特性と単位取得の関係を概観していく。

<sup>1</sup> 札幌学院大学経済学部；nagatomo@sgu.ac.jp.

<sup>2</sup> 札幌学院大学経営学部；chiharu@sgu.ac.jp.

<sup>3</sup> 札幌学院大学経営学部；wattan@sgu.ac.jp.

<sup>4</sup> 札幌学院大学社会情報学部；koike@sgu.ac.jp.

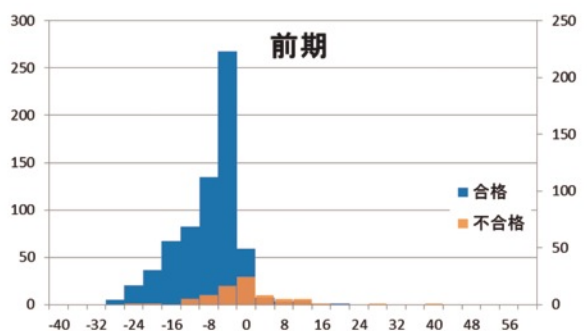


図1：締め切り日に対する合格課題数の度数分布

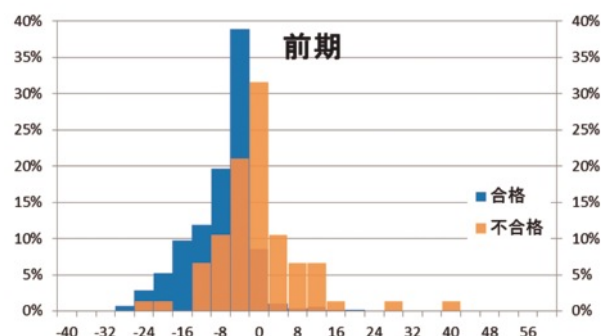
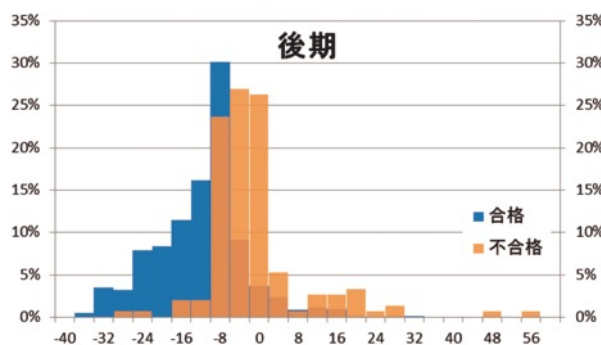
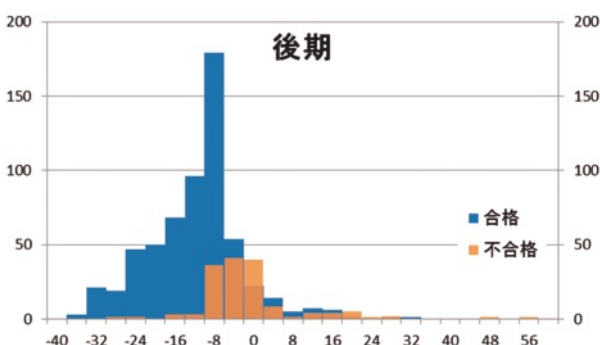


図2：締め切り日に対する合格課題数の相対度数分布



## 2 分析対象となる課題

### 2.1 課題の概要

Word の課題内容は、日本語のローマ字入力とその正確さ、見本通りのレイアウトで文章作成、タブとインデント、図表のあるレポート作成、長文の編集等である。Excel の課題内容は、基本的なデータの入力、数式・関数、参照概念、データの集計と分析、グラフの作成まで行う。これらのすべての課題はツールで学生各自が採点し、その履歴が記録される。

### 2.2 課題の合格

課題の合格までには以下の過程を経る。すべての課題には締め切りが設定されていて、毎回の授業日に1～2程度ある。授業中は教員やTAが指導している。締め切り時間は、その日の実習室の閉室時間である。学生は自らがツールを使って課題をチェックするため、授業時間以外にも課題を完成することは可能である。

ツールが採点対象としているのは、Word では用紙サイズ、余白、文字とその各種スタイル、ルーラー、図表等のかなり特種なことを除いた一般的な文章作成で使う機能で、これらの完成度が都度記録される。Excel ではセル値、数式、書式、条件付き書式、罫線、セル幅、グラフ、グラフ系列などの完成度（正解率）

である。対象項目の正解率が100%になると合格となる。

ツールによって合格サインが出た課題は、所定のフォルダに学生自ら提出する。それらのファイルは合格であることを再度確認してから、合格フォルダに管理サイドが移動して、初めて正規合格となる。締め切り日を過ぎて提出することもできるが、その場合は正規の課題点から一定の割合で減点される。

### 2.3 単元まとめ試験

単元まとめ試験は時間を決めて一斉に実施し（通常の定期試験や中間試験のイメージ）、当該ファイルの回収後に教員がツールで採点し、その完成度に応じて中間点を与えている。ちなみに、良い成績で単位取得するためには、前後期で各3回実施する単元まとめ試験を受験し、ある程度の点数を得られなければならない。一方、単元まとめ試験が未受験の場合でも単位取得が可能になる場合もある。

これらの試験を受験・未受験、合格・不合格という情報も同時に確認できるので、これらと時間分布、それまでの通常課題の合格状況などから、総合的な分析が可能となる。

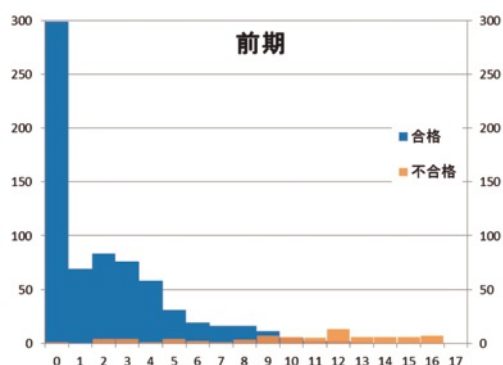


図3：未提出課題数の度数分布

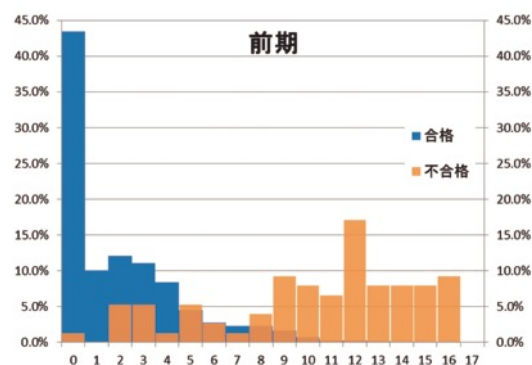
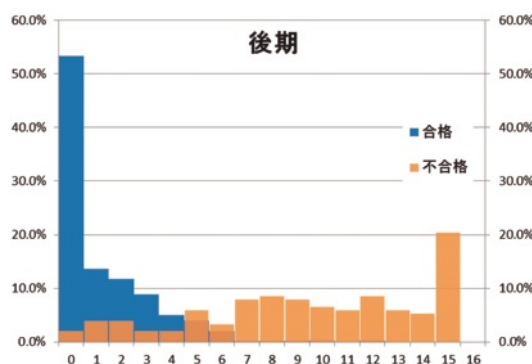
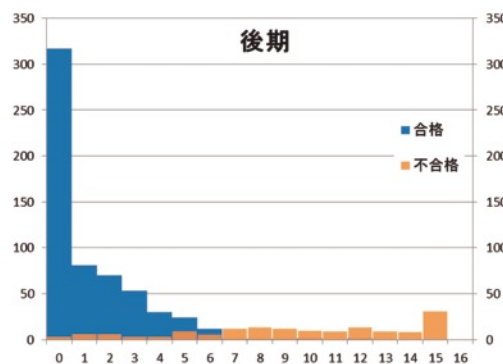


図4：未提出課題数の相対度数分布



## 2.4 分析対象のデータ

分析対象となるデータは、ツールによって合格となって提出された課題ファイルの、締め切り日に対する相対的日数、任意の日付に対する課題合格数（課題不合格数）である。

## 3 課題提出日データの分析

設定された各課題の締め切り日に対して、その日に課題が提出されれば0、それ以前に提出していればマイナスの値、期限以降の場合はプラスの値として、各個人の提出までの平均日数を計算し、単位取得者（合格）と非取得者（不合格）について集計したグラフが図1と図2である。前者は度数で、後者は相対度数である。

一方、課題の未提出数と合格の関係を見るために、未提出数の回数の分布を、合格ごとに集計したグラフが図3と図4である。

図1と図2では、単位の取得と未取得（合格）あわせて当日以前に前期89%、後期90%の学生が課題を提出している。不合格者は平均的に締め切り日付近に多く分布している様子がうかがえる。前後期とも最頻値は共にずれている。

また、ここでは集計数字を示さないが、分析対象と

なった前期16課題のうち、締め切り以前に提出された課題の割合は52%、当日が28%、それ以降が4%、未提出が16%であった。後期17課題では、締め切り以前の提出割合は56%、当日が18%、以降が5%、未提出が22%であった。

図3と図4から合格者と不合格者の分布に差があることが読み取れる。まとめ試験なしに単位取得はかなり困難であるが、その前に未提出課題を1つでもなくしていくことが必須であることがわかる。

成績評価後のデータ集計によって合格の傾向を観察したが、毎週の課題の提出状況を見ながら、落ちこぼれないようにする方策を次節では考える。単純には未提出課題を1つでも減らすことであるが、締め切り日に課題提出をしないような指導が必要であろう。

## 4 課題合否によるロジスティック回帰モデルによる分析

通常課題の合否履歴を通して、単位を落とさないような学生指導の材料を検討するために、任意の課題締め切り日における課題合格数と、最終的な合否の関係をロジスティック回帰モデル（中村，2009）で分析する。ロジスティック回帰モデルは、

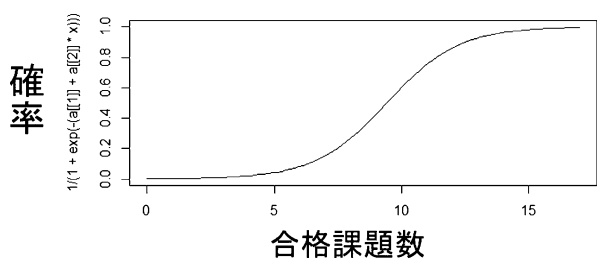


図5：ロジスティック回帰曲線

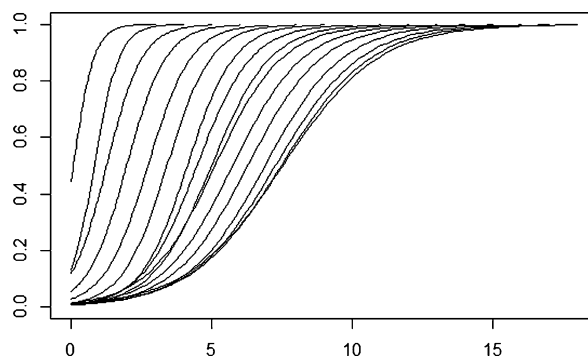
$$y = \frac{1}{1 + \exp(-\alpha - \beta x)}$$

で表され、図5のような曲線となる。αは曲線の立ち上がる場所を表し、βは曲線の勾配を表すパラメータである。

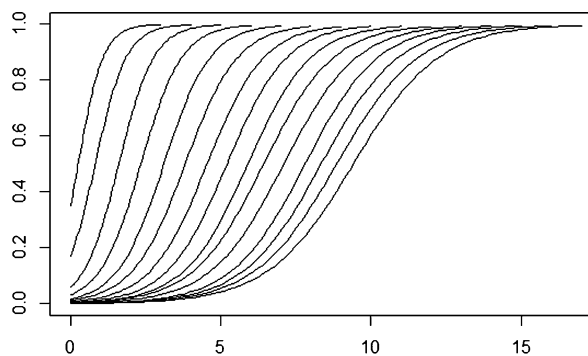
学生ごとに特定の課題の締め切り日までの課題合格数が、ロジスティック回帰モデルの説明変数である。目的変数は最終的な単位の取得(取得=1, 未取得=0)である。課題の締め切りごとに回帰曲線が得られる。

前期と後期のデータに対して、あてはめた結果が図6の(a)と(b)である。回帰曲線が多数あるが、一番右端の曲線は、最後の課題の締め切り時点におけるロジスティック回帰曲線である。例えば、(b)後期の右端の曲線は、後期の最終課題である16課題目の締め切り日における回帰曲線である。横軸が10の値のとき、たて軸が約0.6の値となっている。これは、最後の課題締め切り日に10課題合格していれば、60%の合格率ということを表している(実際、最後の課題の締め切り日においては、単位取得の有無はほぼわかっている。さらに、単位取得のためには、一定以上の出席率(総ログイン時間)や、まとめ試験の点数も関係している。この曲線による確率での説明が有効なのは、それ以前の締め切り日においてであろう)。

このようにして得られる合格確率を表として示したのが、前期が表1で、後期が表2である。前期の表頭のK11~K27は課題11~課題27を、後期の表頭のK31~K46は課題31~課題46の締め切りを表す。また、前期の表側の1~17、後期の1~16は、いくつの課題が合格しているかを表す。例えば、前期の課題20(表1のK20)の締め切り日において、課題合格数が4つの学生は、この段階での単位取得確率が28%であることから、この情報を基に指導を行うことができる、ということである。この表を学生に見せる必要はないと思うが、教員やTAが学生を指導する際に用いることで、学生に課題達成の目標を示すための資料となる。



(a) 前期



(b) 後期

図6：課題締め切り日における課題合格数に対する単位取得確率

## 5 おわりに

コンピュータ基礎の授業で課す数多くの課題の提出状況と単位の取得状況を分析することで、学生への指導資料として用いるための若干の知見が得られた。より効果的に指導を行うためには、学生個別の進捗状況を個別に把握し、同時に全学生の類型化が可能な最新の統計手法である関数データ解析法(Ramsay & Silverman, 2005)による分析を考えている。これは今後の研究課題としたい。

## 参考文献

- [1] 石川千温・皆川雅章・渡邊慎哉・中村永友・小池英勝・梅田充(2008). 自動採点システムを用いたコンピュータリテラシー教育の質保証, 教育システム情報学会第33回全国大会予稿集, 熊本大学.
- [2] Hidekatsu Koike, Kiyoshi Akama, Chiharu Ishikawa, Nagatomo Nakamura & Shinya Watanabe (2007). Toward a Software Development Model for Automatic Marking Software, Proc. of ACM Special Interest Group on University and College Computing Services (SIGUCCS) Fall 2007 Confer-

表 1 : 締め切り課題時点における合格課題数から算出される単位取得格率 (前期; 単位: %)

合格 課題数	締め切りの課題																
	K11	K12	K13	K14	K15	K16	K17	K18	K19	K20	K21	K22	K23	K24	K25	K26	K27
1	2	92	60	40	20	9	4	2	3	2	3	2	2	1	2	2	2
2		99	93	76	51	28	14	7	7	5	7	5	4	3	3	3	4
3			99	94	81	59	37	20	18	13	14	10	8	6	6	6	6
4				99	95	84	68	46	38	28	28	20	15	11	10	10	11
5					99	95	88	74	63	50	47	35	27	20	18	17	18
6						99	96	91	83	72	68	55	45	34	30	28	29
7							99	97	93	87	83	73	63	52	46	43	42
8								99	97	95	92	86	79	69	62	58	57
9									99	98	96	93	89	82	76	72	71
10										99	98	97	95	90	86	83	81
11											99	99	97	95	92	90	89
12												99	99	98	96	95	93
13													99	99	98	97	96
14														99	99	98	98
15															99	99	99
16																100	99
17																	100

表 2 : 締め切り課題時点における合格課題数から算出される単位取得格率 (後期; 単位: %)

合格 課題数	締め切りの課題															
	K31	K32	K33	K34	K35	K36	K37	K38	K39	K40	K41	K42	K43	K44	K45	K46
1	2	86	59	26	12	6	4	2	1	1	1	1	0	0	0	0
2		99	91	66	38	19	11	6	4	2	2	1	1	1	1	1
3			99	92	73	47	28	16	10	6	5	3	2	2	1	1
4				98	92	76	55	36	23	14	11	7	4	3	3	2
5					98	92	80	62	45	30	23	15	9	7	5	4
6						98	92	82	69	52	41	29	18	14	11	8
7							97	93	86	74	62	48	34	26	20	15
8								98	94	88	79	68	54	43	35	27
9									98	95	90	83	73	62	53	43
10										98	96	92	86	78	70	60
11											98	96	93	89	83	75
12												98	97	95	91	86
13													99	97	95	93
14														99	98	96
15															99	98
16																99

ence, 190-193, Oct. 7-10, 2007, Orland, Florida, America.

- [3] 小池英勝・中村永友・石川千温 (2010). レポート・プレゼンが楽々! パソコン活用術, ムイスリ出版, 東京.

- [4] 中村永友 (2009). 多次元データ解析法, 共立出版, 東京.

- [5] James Ramsay & Silverman, B.W. (2005). Functional Data Analysis, Springer, New York.

## Analysis of Submission Logs on the Learning System via Logistic model

Nagatomo NAKAMURA<sup>1</sup>

Chiharu ISHIKAWA<sup>2</sup>

Shin-ya WATANABE<sup>3</sup>

Hidekatsu KOIKE<sup>4</sup>

### Abstract

The information processing education's main objectives at Sapporo Gakuin University are to offer all the students the ideal environment to acquire basic computer literacy and skills such as the functions of a word processor and a spread sheet software. Classes require a number of learning tasks and their results are directly linked to the credits. A recent problem is that some students do not fully grasp the progress of their tasks and stop learning at the middle of the course. Even so, they attend classes till the end of the course but are unable to get the required credits. The purpose of this study is to make a basic teaching plan for such students. We have to clarify the differences between the successful and unsuccessful candidates through the analysis of submission dates for learning tasks and the log data of passing the learning tasks.

**Keywords:** Logistic Regression Model, Computer Literacy Education.

---

<sup>1</sup>Department of Economics, Sapporo Gakuin University; nagatomo@sgu.ac.jp.

<sup>2</sup>Department of Bussines Administration, Sapporo Gakuin University; chiharu@sgu.ac.jp.

<sup>3</sup>Department of Bussines Administration, Sapporo Gakuin University; wattan@sgu.ac.jp.

<sup>4</sup>Department of Social Infomation, Sapporo Gakuin University; koike@sgu.ac.jp.