

Rによる学修時間調査の分析

Analysis of the Questionnaire about Study Time using “R”

三好 善彦

MIYOSHI Yoshihiko

Now, we conduct many questionnaires which are the freshman questionnaire, the graduate satisfaction questionnaire, the class questionnaire and so on in Saitama W. J. College. So we summarize and report their questionnaires. In July 2013 we conduct the questionnaire about study time for the first time. In this paper, I analyzed this questionnaire using statistical analysis software “R”, which is freeware.

1. はじめに

現在、いろいろな場面において、多くのアンケート調査が行われている。これらのアンケートは集計することにより、その結果から傾向を簡単に分析することができる。そのとき、結果を整理し分類するのみの分析では、これらのアンケートの結果を表面的に捉えているに過ぎないといえる。そこで、表面には現れない傾向やデータ間の関係などを知るには、統計解析処理を行う必要が生じてくる。本ノートでは、アンケートの分析の簡単な手法の一つである独立性の検定と数量化Ⅲ類によるデータの分類を統計解析ソフトを利用して行った。

2. 統計解析ソフト R¹について

本ノートで使用する統計解析ソフト R とは、統計解析用のフリーソフトウェアであり、本ノート執筆時の最新バージョンは2013年9月25日リリースの3.0.2である。統計解析ソフトウェア²

には SAS、SPSS、Stata、JMP などさまざまなものがあるが、どれも市販されている製品であるため誰でも自由に使うことが出来ない。それに対し、R はフリーソフトウェアであるため誰でも自由に使うことが出来る。本ノートでは、誰でも簡単に統計処理ができることを目的として、統計解析ソフトウェア R の Windows 64ビット版を利用して統計解析を行った。R には Windows 版以外にも Linux 版や Max (OS X) 版も用意されている。



R を起動すると上記のようにコンソール画面が開く。ここに様々なコマンドを入力すると、その結果が出力される。R には統計解析以外にもグラフ表示などのさまざまな機能があるが、本ノートではよく使われる統計解析の一つである独立性の検定と数量化Ⅲ類をこの R を用いて行った。以下のすべての統計解析結果は R のコンソール画面に出力された内容をそのままコピーしたものである。

3. 学修時間調査について

本学では、2013年度春学期の学生の学修実態を調査する目的で春学期の最終基礎ゼミにおいて「学修時間調査」を行った。実施対象は全学生で、学科と学年のみの記名式で回答し、実施内容は以下のとおりである。

- 基礎ゼミに対する取り組み
 - 基礎ゼミ小テスト1回分の勉強時間
 - ✓ していない・30分未満・1時間未満・2時間未満・2時間以上
 - 勉強量
 - ✓ 少ない・ちょうど良い・多い
 - 勉強場所（複数回答可）
 - ✓ 自宅・電車やバス・図書館・学食・教室・その他
 - 勉強をしていないを回答した場合の理由
 - ✓ 忙しい・必要ない・やる気がない・その他
- 授業全体に対する取り組み
 - 授業に対する勉強は1日の勉強時間
 - ✓ していない・30分未満・1時間未満・2時間未満・2時間以上
 - 勉強量
 - ✓ 少ない・ちょうど良い・多い
 - 勉強内容（複数回答可）
 - ✓ 予習・復習・宿題・レポート・小テスト・資格・その他
 - 勉強場所（複数回答可）
 - ✓ 自宅・電車やバス・図書館・学食・教室・その他
 - 勉強をしていないを回答した場合の理由
 - ✓ 忙しい・必要ない・やる気がない・その他

基礎ゼミに対する取り組みの調査が他の授業とは別に行われているのは次のような理由からである。本学ではオリジナルの基礎ゼミ問題集を作成して毎回の授業でこの問題集から小テストを実施している。また、期末の定期試験においてもこの問題集からの出題となっている。このため、基礎ゼミにおける学修は全学生が必ず取り組まなければならないものであるからである。さらに言えば、講義や演習などに関しては授業以外での学修が必要になるが、実習や実技に関しては必要にならない場合もあるため、これら授業と明確に区別するため基礎ゼミに対してのみ別に調査することになった。

そして、本ノートの目的は、これらの調査結果をもとに学生の基礎ゼミと授業全般に対する学

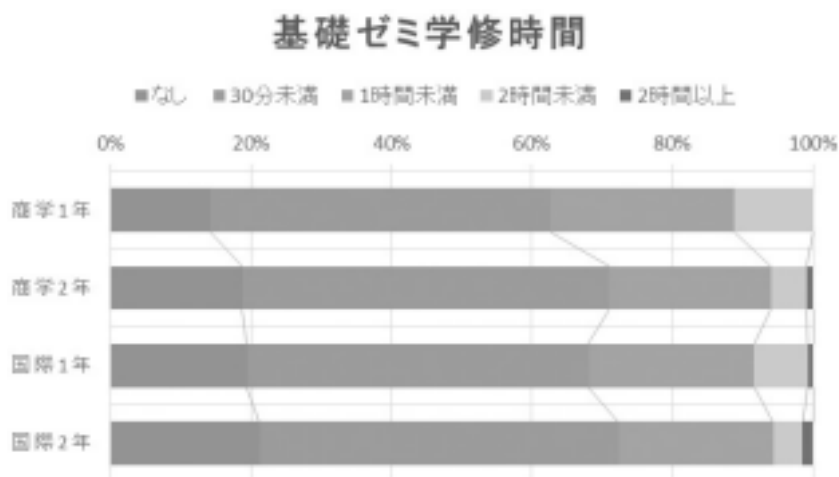
修への取り組みに関して分析することである。以降、統計解析ソフト R を用いて独立性の検定と数量化Ⅲ類による分析を行う。

4. 独立性の検定³による学科学年の分析

まず、この調査結果を集計して各学年学科における学修時間の違いを集計した。そこで、学年や学科における学修時間の割合の違いがあるかどうかを独立性の検定により分析することとした。

基礎ゼミに対する学修時間の集計は、以下の表およびグラフのとおりとなった。

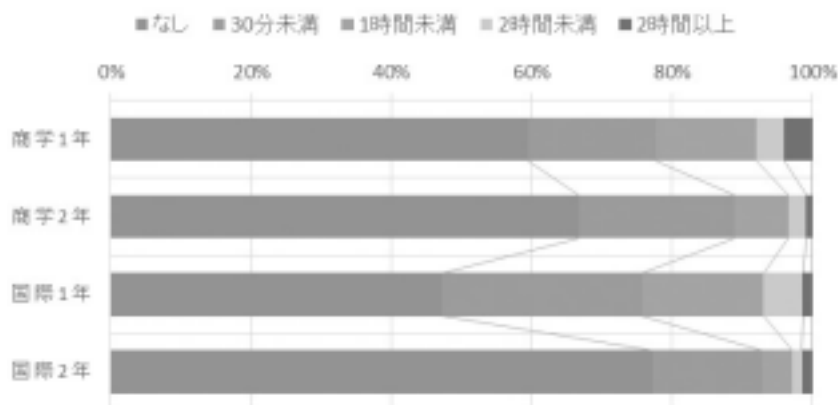
		学修時間					総計
		なし	30分未満	1時間未満	2時間未満	2時間以上	
学年・学科	商学	40	122	60	20	1	243
	1年	18	61	33	14	0	126
	2年	22	61	27	6	1	117
	国際	58	142	65	17	3	285
	1年	28	70	34	11	1	144
	2年	30	72	31	6	2	141
	総計	98	264	125	37	4	528



また、授業全般に対する学修時間の集計は、以下の表およびグラフのとおりとなった。

		学修時間					
		なし	30分未満	1時間未満	2時間未満	2時間以上	総計
学年・学科	商学	153	49	27	8	6	243
	1年	75	23	18	5	5	126
	2年	78	26	9	3	1	117
	国際	177	63	31	10	4	285
	1年	68	41	25	8	2	144
	2年	109	22	6	2	2	141
	総計	330	112	58	18	10	528

授業全般学修時間



これらの結果から考察すると、基礎ゼミにおいては学科学年における学修時間の割合の差はほとんど無いが、授業全般においては大きく差があるように判断できそうである。そこで、実際にこれらの結果から独立性の検定を行うことで理論的な結論を導き出すこととした。その結果、学修時間における割合の差について以下の様な分析をすることができた。

まず、基礎ゼミに対する学科・学年における学修時間の割合に関して独立性の検定を行う。

検定内容：

基礎ゼミの学修時間の集計結果をもとに、有意水準5%で次の仮設 H_0 を検定する。

H_0 : 学科学年において、学修時間の割合に違いはない。

検定結果：(R コンソール)

```
> ctbl <- c( # 基礎ゼミ
+ 18, 61, 33, 14, 0, # 商学科1年
+ 22, 61, 27, 6, 1, # 商学科2年
+ 28, 70, 34, 11, 1, # 国際科1年
+ 30, 72, 31, 6, 2 # 国際科2年
+ )
> chisq.test( matrix( ctbl, 4, 5, byrow=T), correct=F) # 4行5列の表でカイ2乗検定
```

Pearson's Chi-squared test

```
data: matrix(ctbl, 4, 5, byrow = T)
X-squared = 9.6608, df = 12, p-value = 0.6457
```

警告メッセージ:

```
In chisq.test(matrix(ctbl, 4, 5, byrow = T), correct = F) :
  カイ自乗近似は不正確かもしれません
```

上記結果から、 χ^2 値は9.6608で確率64.57%となり、有意水準5%より大きいため、仮説 H_0 は採択され、学科学年において学修時間の割合に違いはないといえる。

同様に、授業全般に対する学科・学年における学修時間の割合に関して独立性の検定を行う。

検定内容：

授業全般の学修時間の集計結果をもとに、有意水準5%で次の仮説 H_0 を検定する。

H_0 ：学科学年において、学修時間の割合に違いはない。

検定結果：(R コンソール)

```
> ctbl <- c( # 授業全般
+ 75, 23, 18, 5, 5, # 商学科1年
+ 78, 26, 9, 3, 1, # 商学科2年
+ 68, 41, 25, 8, 2, # 国際科1年
+ 109, 22, 6, 2, 2 # 国際科2年
+ )
> chisq.test( matrix( ctbl, 4, 5, byrow=T), correct=F) # 4行5列の表でカイ2乗検定
```

Pearson's Chi-squared test

```
data: matrix(ctbl, 4, 5, byrow = T)
X-squared = 38.4567, df = 12, p-value = 0.0001293
```

警告メッセージ:

```
In chisq.test(matrix(ctbl, 4, 5, byrow = T), correct = F) :
カイ自乗近似は不正確かもしれません
```

上記結果から、 χ^2 値は38.4567で確率0.01293%となり、有意水準5%より小さいため、仮説 H_0 は棄却され、学科学年において学修時間の割合に違いがあるといえる。

ここで、Rの結果において「カイ自乗近似は不正確かもしれません」と出力される原因⁴として、以下の様な場合が考えられる。

- 期待値が1未満の項目が1つでもある。
- 期待値が5未満の項目が全体の20%以上ある。

このような場合の対処方法として、

- カテゴリーを併合する。
- カテゴリーが併合できないような場合には、他の検定手法の適用の可能性を検討する。

といった方法があるので、今回は、学修時間が「1時間以上2時間未満」と「2時間以上」のカテゴリーを併合して、再度同様の検定を行った。

基礎ゼミの学修時間の検定結果：(R コンソール)

```
> ctbl <- c( # 基礎ゼミ
+ 18, 61, 33, 14, # 商学科1年
+ 22, 61, 27, 7, # 商学科2年
+ 28, 70, 34, 12, # 国際科1年
+ 30, 72, 31, 8 # 国際科2年
+ )
>
> chisq.test( matrix( ctbl, 4, 4, byrow=T), correct=F) # 4行4列の表でカイ2乗検定
```

Pearson's Chi-squared test

```
data: matrix(ctbl, 4, 4, byrow = T)
X-squared = 5.797, df = 9, p-value = 0.76
```

授業全般の学修時間の検定結果：(R コンソール)

```
> ctbl <- c( # 授業全般
+ 75, 23, 18, 10, # 商学科1年
+ 78, 26, 9, 4, # 商学科2年
+ 68, 41, 25, 10, # 国際科1年
+ 109, 22, 6, 4 # 国際科2年
+ )
>
> chisq.test( matrix( ctbl, 4, 4, byrow=T), correct=F) # 4行4列の表でカイ2乗検定

Pearson's Chi-squared test

data: matrix(ctbl, 4, 4, byrow = T)
X-squared = 35.3988, df = 9, p-value = 5.066e-05
```

上記結果から、基礎ゼミにおいては、 χ^2 値は5.797で確率76%となり、有意水準5%より大きいため仮説 H_0 は採択され、学科学年において学修時間の割合に違いはないといえる。また、授業全般においては、 χ^2 値は35.3988で確率0.005066%となり、有意水準5%より小さいため、仮説 H_0 は棄却され学科学年において学修時間の割合に違いがあるといえる。

以上の独立性の検定結果から、基礎ゼミと授業全般に関して以下の様な考察が可能である。

- 基礎ゼミにおいては、各学科・各学年とも基礎ゼミ問題集を利用した形式で授業、小テスト、定期試験を行っているので学修実態に関して割合に差はないといえる。
- 授業全般においては、学科や学年、さらには選択コースにより学修内容が異なっている。その結果、授業内でのプレゼンテーション準備や資格試験などの学修に費やす時間が大きく違ってきているのではないかと考えられる。

5. 数量化Ⅲ類⁵によるカテゴリーの分析

次に、各調査項目の分類をおこなうため、学科や学年などの調査項目をカテゴリーデータとして扱い数量化Ⅲ類により分析を行う。ここではパッケージ MASS の関数 `corresp` を用いることによって数量化Ⅲ類による分析を行った。

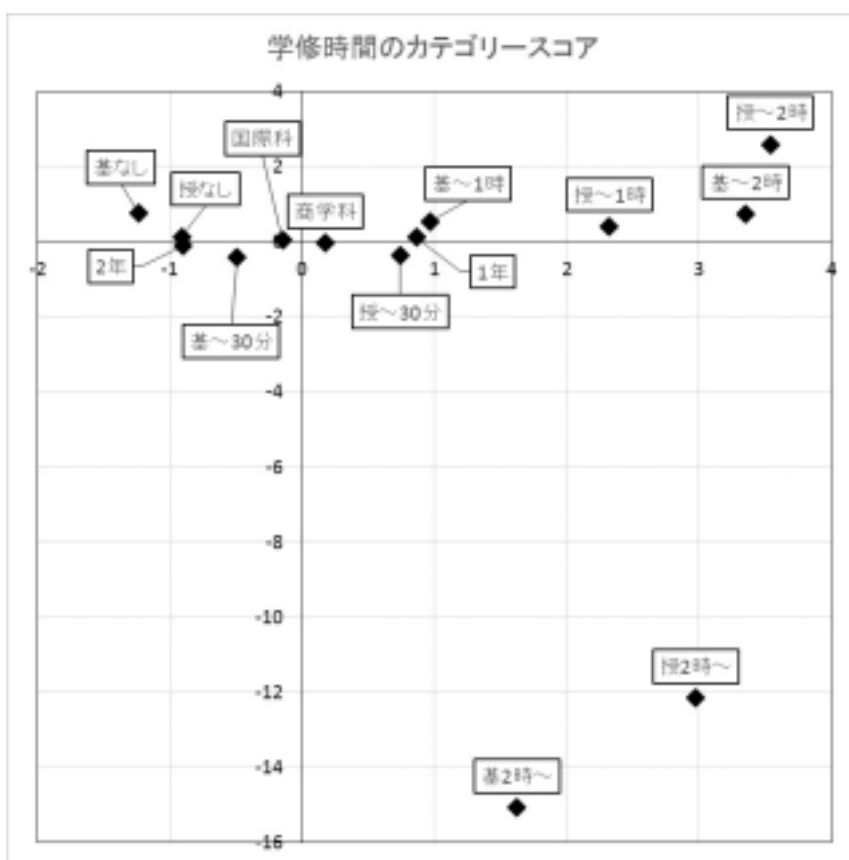
5.1. 学修時間の分析

まず、学修時間と学科や学年の分類を行った。これらに関しては、先ほどの独立性の検定から授業全般においては学科や学年での学修時間の違いを見出すことができています。ここでは、これら学修時間の分類をおこなうため、学科（商学科、国際科）、学年（1年、2年）、基礎ゼミ学修時間（基なし、基～30分、基～1時、基～2時、基2時～）、授業全般学修時間（授なし、授～30分、授～1時、授～2時、授2時～）の14カテゴリーで数量化Ⅲ類を行った。その結果、相関係数 $r_1=0.546719$, $r_2=0.4960908$ （上位2つまで、以下同様）、固有値 $\lambda_1=0.2989017$, $\lambda_2=0.246106$ 、固有ベクトル（カテゴリースコア）は以下のとおりとなった。

	α_1	α_2
商学科	0.1759701	-0.04536823
国際科	-0.147984	0.04024758
1年	0.8637155	0.11012184
2年	-0.9016197	-0.11351479
基なし	-1.2359687	0.7625837
基～30分	-0.4916297	-0.41065669
基～1時	0.9682733	0.53715151
基～2時	3.3505937	0.73582309
基2時～	1.6235871	-15.06078687
授なし	-0.9113973	0.13010861
授～30分	0.7393747	-0.36784056
授～1時	2.3214224	0.40152832
授～2時	3.5387481	2.5841411
授2時～	2.9758712	-12.14652082

このカテゴリースコアをグラフに表すと以下の通りとなる。これらから判断できることは以下のとおりである。

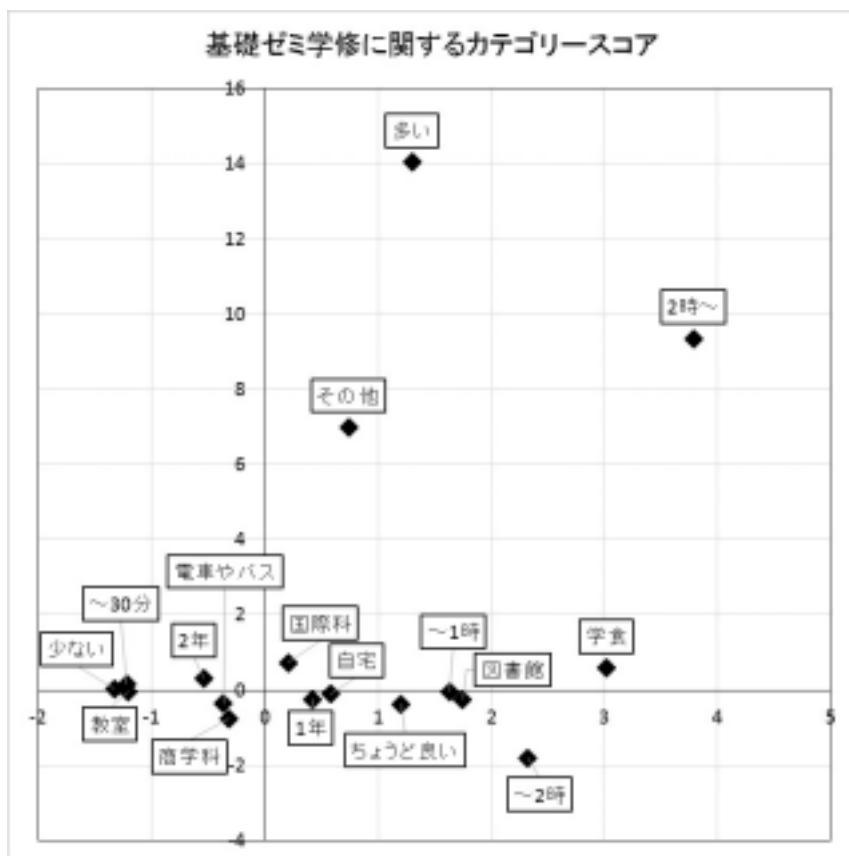
- 横軸（x 軸）が学修時間を表している。
- 2年生の学修時間はないかそれに等しいぐらい少ない。
- 1年生の学修時間は基礎ゼミ1時間未満、授業全般では30分未満に等しい。
- 商学科と国際コミュニケーション学科はほぼ同じぐらいの学修時間だが、商学科のほうが多少多い。



5.2. 学修内容と場所の分析

次に、基礎ゼミにおける学修時間と学科や学年、学修量、学修場所の分類を行った。ここでは、基礎ゼミにおいて実際に学修を行っている学生を対象に学修量や学修場所の分類を目的として行うため、学修していない学生は除いている。そのうえで、これらの分類をおこなうため、学科(商学科、国際科)、学年(1年、2年)、学修時間(～30分、～1時、～2時、2時～)、学修量(少ない、ちょうど良い、多い)、学修場所(自宅、電車やバス、図書館、学食、教室、その他)の17カテゴリで数量化Ⅲ類を行った。その結果、相関係数 $r_1=0.546719$, $r_2=0.4960908$ 、固有値 $\lambda_1=0.2989017$, $\lambda_2=0.246106$ 、固有ベクトル(カテゴリースコア)は以下のとおりとなった。

	α_1	α_2
商学科	-0.3124073	-0.75377435
国際科	0.209117	0.71938137
1年	0.4241442	-0.2477244
2年	-0.5386283	0.31928953
～30分	-1.2161518	0.16803663
～1時	1.6320511	-0.03571767
～2時	2.3234234	-1.81171405
2時～	3.7854766	9.35495883
少ない	-1.3249801	0.02947848
ちょうど良い	1.2052822	-0.36875619
多い	1.3023759	14.03898893
自宅	0.5807146	-0.10686451
電車やバス	-0.3640052	-0.35694564
図書館	1.7481853	-0.24623566
学食	3.0163181	0.58428902
教室	-1.2051597	-0.04825581
その他	0.7415431	6.99010896



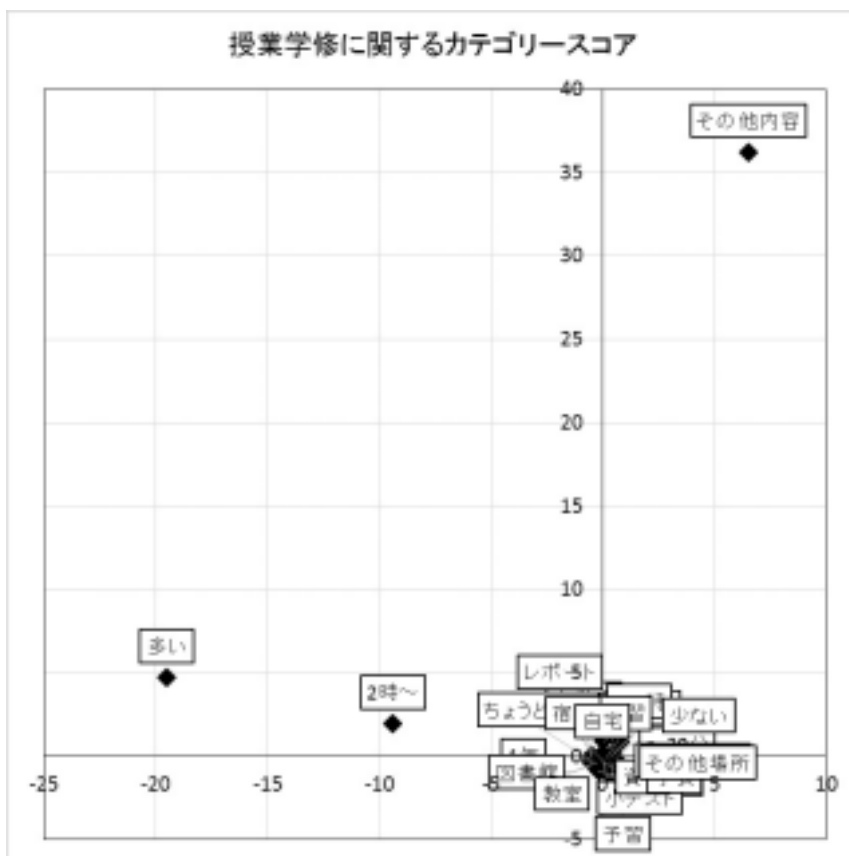
このカテゴリースコアをグラフに表すと上記の通りとなる。これらから判断できることは以下のとおりである。

- 横軸（x 軸）が学修時間を表している。
- 2年生よりは1年生の学修時間は多い。
- 商学科よりは国際コミュニケーション学科の学修時間は多い。
- 学修時間が30分未満の学生は、学修量が少ないと感じている。また、これらの学生は学修場所として教室か電車やバスを利用している。
- 学修時間が30分を超え2時間未満の学生は、学修量はちょうど良いと感じている。また、これらの学生は学修場所として自宅か図書館か学食を利用している。

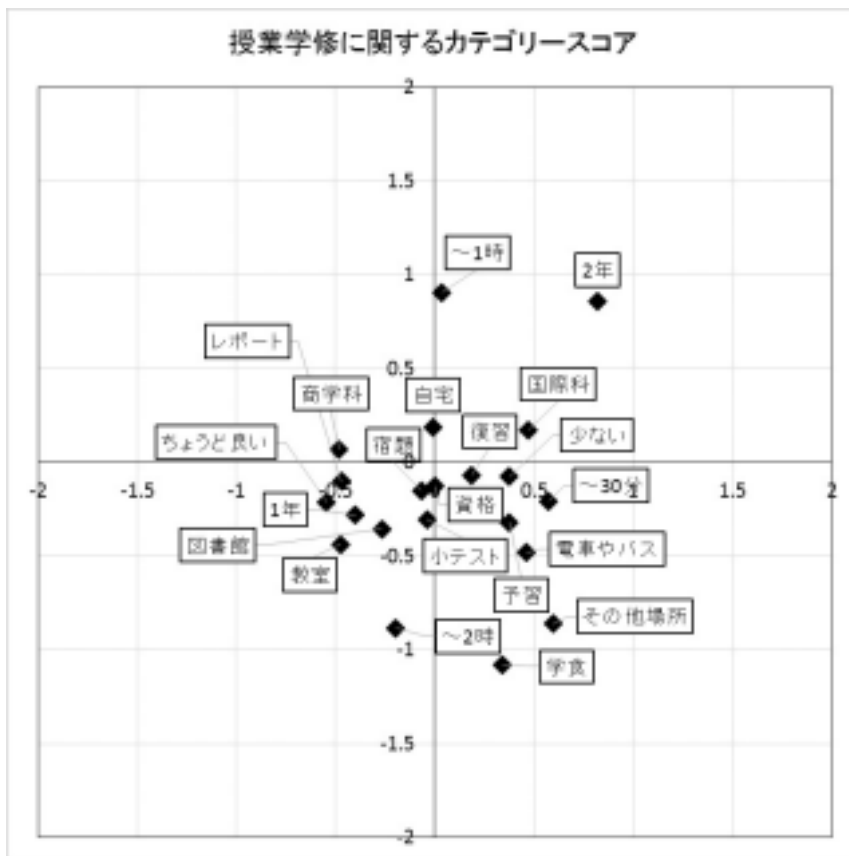
最後に、授業全般における学修時間と学科や学年、学修量、学修内容、学修場所の分類を行っ

た。ここでは、先ほどの基礎ゼミの場合と同様に、学修していない学生は除いている。そのうえで、これらの分類をおこなうため、学科（商学科、国際科）、学年（1年、2年）、学修時間（～30分、～1時、～2時、2時～）、学修量（少ない、ちょうど良い、多い）、学修内容（予習、復習、宿題、レポート、小テスト、資格、その他内容）、学修場所（自宅、電車やバス、図書館、学食、教室、その他場所）の24カテゴリーで数量化Ⅲ類を行った。その結果、相関係数 $r_1=0.4901462$, $r_2=0.4601403$ 、固有値 $\lambda_1=0.24024326$, $\lambda_2=0.21172913$ 、固有ベクトル（カテゴリースコア）は以下のとおりとなった。

	a_1	a_2
商学科	-0.483727022	0.06938142
国際科	0.47057057	0.17298438
1年	-0.400707608	-0.28235245
2年	0.819381068	0.85613242
～30分	0.57433186	-0.20715521
～1時	0.03408405	0.9057954
～2時	-0.197321046	-0.88351267
2時～	-9.378715555	1.92530282
少ない	0.374532577	-0.07728691
ちょうど良い	-0.546620813	-0.21127936
多い	-19.50611778	4.70171145
予習	0.375546132	-0.32027029
復習	0.185268926	-0.0681227
宿題	-0.068302349	-0.14978303
レポート	-0.471108957	-0.10113599
小テスト	-0.038840591	-0.30485665
資格	0.000259284	-0.12715235
その他内容	6.539181133	36.2094624
自宅	-0.008245999	0.18861569
電車やバス	0.463008861	-0.48010301
図書館	-0.268145322	-0.35556543
学食	0.339588484	-1.07736207
教室	-0.476995189	-0.43867502
その他場所	0.595769872	-0.86096123



このカテゴリースコアをグラフに表すと上記の通りとなる。ここで、いくつかのカテゴリースコア値が大きい（小さい）く、それ以外のカテゴリースコアが集中し判別不能となってしまうので、グラフを以下のように拡大した。



これらから判断できることは以下のとおりである。

- 横軸（x 軸）が学修時間を表している。ここで、これまでとは異なり、カテゴリースコアが小さい（左側）と学修時間が多いことになる。
- 基礎ゼミと同様に2年生よりは1年生の学修時間は多い。ここで、2年生は学修時間が少ないと感じており、1年生は学修時間がちょうど良いと感じている。
- 基礎ゼミとは逆に国際コミュニケーション学科よりは商学科の学修時間は多い。ここで、国際コミュニケーション学科の学生は学修時間が少ないと感じており、商学科の学生は学修時間がちょうど良いと感じている。
- 基礎ゼミと同様に学修時間が30分未満の学生は、学修量が少ないと感じている。また、これらの学生は電車やバスか学食において授業の予習か復習を行っている。
- 基礎ゼミと同様に学修時間が30分を超え2時間未満の学生は、学修量はちょうど良いと感

じている。また、これらの学生は学修場所として自宅か図書館か教室において資格の勉強か宿題かレポートを行っている。

6. まとめ

以上、統計解析処理として独立性の検定による学科や学年による違いの分析と数量化Ⅲ類による学修時間や場所などカテゴリーの分類を行ってきた。今回の学修時間調査は春学期に行われたものなので、秋学期も同様の統計解析処理を行うことでより詳しく分析することが可能となるであろう。また、本学ではさまざまなアンケート調査を行っているので、今回の処理をそれらに应用することで今まで以上に詳しく分析することができると考えられる。

また、今回は統計解析ソフトウェアとして R を使用したが、前にも述べたように R 以外にも多くのソフトウェアがある。しかし、それらのソフトウェアは導入コストがかかったり維持コストがかかったりするものが多い。それに対し R はフリーソフトウェアであるため導入コストも維持コストもかからない。さらに、サポート面に関しても多くの書籍が出版されていたりホームページで紹介されていたり他ソフトウェアと比較しても遜色ない。というより、充実しているとさえいえる。そのため、今後新たに統計解析処理を行ってみたいと考えている人には大いにお勧めできるソフトウェアである。

最後に、本ノートの統計解析処理はすべて R で行っているが、グラフ表示に関しては表計算ソフトの Excel で行っている。今回は準備不足や調査不足のため Excel を使ったが、R には多彩なグラフィック機能が用意されているので、今後はそれらの機能を使って最終的なグラフ表示まで R で行うようにしたい。

注

1. The R Project for Statistical Computing (<http://www.r-project.org/>) では、“R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.” と紹介され

ている。R プロジェクト日本語トップページ (http://sourceforge.jp/projects/sfnet_rproject_mirror/) では、「R は、統計コンピューティングおよびグラフィック向けの言語環境です。これは GNU プロジェクトの一つであり、ジョン・チェンバースおよび同僚によって Bell 研究所（旧 AT&T、現在はルーセントテクノロジー）にて開発された S 言語環境と似ています。R は S の別の実装として見なすことができます。幾つかの重要な違いもありますが、S のために書かれる多くのコードは R の下で不変で利用できます。」と訳されている。

2. 統計解析ソフトウェアの比較としては、「統計インフラのつくり方」の「1. 統計ソフトの比較」(<https://sites.google.com/site/statinfra/comparison>) に詳しく紹介されている。
3. 独立性の検定については、「統計処理ポケットリファレンス～Excel&R 対応」182ページ「79. $m \times n$ 分割表の検定（独立性の検定）」により行った。
4. 「統計処理ポケットリファレンス～Excel&R 対応」では、「適用基準として、期待度数が5未満のセルが全体の20%未満であることが望まれる。」と注意されている。また、「おしゃべりな部屋（プラネタリウム、星、植物、熱帯魚、統計学）」(<http://aoki2.si.gunma-u.ac.jp/>) の「統計学自習ノート、検定・推量、独立性の検定、いくつかの注意点」(<http://aoki2.si.gunma-u.ac.jp/lecture/Cross/warning.html>) では、注意や対処法も紹介されており、本ノートではこれらを参考にして統計解析処理を行った。
5. 数量化Ⅲ類については、「R によるデータサイエンス」87ページ「第3章 対応分析」により行った。また、独立性の検定同様「おしゃべりな部屋（プラネタリウム、星、植物、熱帯魚、統計学）」の「統計学自習ノート、検定・推量、多変量解析、数量化Ⅲ類」(<http://aoki2.si.gunma-u.ac.jp/lecture/Qt/qt3.html>) にも詳しく示されており、R のプログラムも紹介されている。

参考文献

- 有馬哲、石村貞夫（1987） 多変量解析のはなし．東京図書：東京．
- 涌井良幸、涌井貞美（2013） 統計処理ポケットリファレンス～Excel&R 対応．技術評論社：東京．
- 金明哲（2007） R によるデータサイエンス．森北出版：東京．
- 舟尾暢男（2009） The R Tips 第2版—データ解析環境 R の基本技・グラフィックス活用集—．オーム社：東京．
- 菅民郎（2007） らくらく図解 アンケート分析教室．オーム社：東京．
- 菅民郎（2011） 実例でよくわかる アンケート調査と統計解析．ナツメ社：東京．