

論文

心理学科における初等統計学講義の試み
—受講生データをもとにして—An Essay on Elementary Statistics for Psychology Courses
-- An Analysis of Data on Students --持元 江津子
YOSHIMOTO Etsuko

要 約

筆者の経験より、大学及び短期大学で開講される初等統計学関連科目の講義において、受講生データの活用は有用である。とりわけ心理学科等で初等段階の因子分析について講義用の例題を作成する場合、受講生対象のアンケートから収集したデータを用いることで、学生の関心と学習意欲を喚起させやすくなる。この3年間の講義経験を再検討することにより、例題作成に適切なデータを収集しやすいアンケート内容を設定するコツが導き出される。それは、テーマを絞り、一般的で身近な話題を取り上げることである。

Key Word : 心理学科, 初等統計学, 因子分析, 受講生データ, 受講生対象アンケート, テーマ

1. はじめに

4年制大学及び短期大学において、筆者の統計学関連科目の講義経験は4教育機関13年に及ぶ。社会科学系の商業及び経営の他、ここ数年は、学際的な都市環境学や人文科学系の心理学専攻の学生を対象とする初等統計学関連

科目の講義にチャレンジして来た。これまで筆者が一貫して心掛けて来たことは、いかにして受講生の関心を引き続け、学習意欲を保持させ、統計学を身近なものとする納得させ、彼らの将来につなげるかである。併せて、統計学の持つ自然科学領域における側面と社会科学及び人文科学領域における側面の両方についても、受講生に理解させるよう腐心して来たつもりである。

そこで、筆者がどのような講義手段を実際に用いてきたかであるが、最大の特徴は受講生から生のデータを集め、それを講義内容にできる限り反映させてきていることである。特に心理学科では因子分析に関する知識が必須と考えられるため、この3年間は、その例題作成のためのデータを受講生から直接集めてきた。尤も、因子分析は元来多変量解析に分類されており、通常初等統計学講義で扱われることは滅多にない。しかし、心理学専攻の学生が学んで損することはない重要な解析技術であるため、筆者はこれを講義構成に組み入れている。

次に、筆者の受講生データの集め方であるが、具体的には、初回講義において、受講生対象のデータ集めを目的とする簡単なアンケートを実施している。アンケートによって集めたデータは、例え匿名であっても、対象人数が少なく個人特定もある程度可能である限り個人情報に当たることから、アンケートに答えることが受講生に単位取得のための義務ではないと断りを入れている。とはいえ、これまでのところ、記入済みアンケート用紙の提出を拒んだ受講生は1人もいない。アンケートの内容は、大まかにみて3種類の質問群から成り、「①身長と体重、性別」、「②高校での確率・統計を含む数学の履修状況や計算または図形に対する好悪感情」、「③その他の某かの質問」で、②と③については「はい／いいえ」のような二者択一の質問形式をとっている。①の身長や体重等のデータは、統計学の持つ自然科学的な側面を受講生に伝えるのに最適の材料となっている。②の高校での確率・統計などの履修状況等は、予め計画していた講義構成及び運営方法を微調整するための資料となる。③のその他の質問は、因子分析を含む諸例題作成のための資料ないし材料として活用される。

以下、聖泉大学人間学部人間心理学科における平成16～18年度の統計学講義での試みを事例として、本稿タイトルで示した通り、大学で心理学を学ぶ大学生を対象とする初等統計学講義において、効果的なデータの取り方及び扱い方について、自らの講義経験を踏まえて考察する。次章では例題作成のためのアンケート内容について筆者の経験に基づいた考えを述べ、第3章では受講生データを用いた因子分析を含む例題作成の成功事例について詳述し、第4章にて例題作成の成功要因について、不成功例と対比しながら検討し、最終章でまとめを述べる。

2. アンケート内容の概略

大学入学以前の学生の統計学履修状況にはばらつきがある。また、履修状況が良好であっても、学生が学んだことを一旦忘れてしまっていることも多々ある。できるだけ初歩の初歩から講義を進めることが不可欠となるが、同時に、既にある程度基礎のある受講生の関心をつなぎ止める工夫も必要となる。その工夫の鍵は、統計の基礎理論や基礎技術のレベルではなく、それらを伝えるために材料として用いるデータにこそある。

例題を作成する際には、受講生対象のアンケート結果や行政機関等によって公表される信頼できるデータ、あるいは説明を容易にしたり受講生の直感に訴えたりするために、実在するデータをアレンジした架空データを駆使する。とりわけ、受講生に関心を持たせ続けるためには、受講生データはやはり欠かせない。

そこで、本章以降は、この3年の間に、特に因子分析のごくごく初歩的な扱いのあらましを伝えるための例題に用いた受講生データに注目して、筆者の経験を軸に議論を展開していく。本稿では、受講生データの中でも、前章で述べた「③その他の某かの質問」に対する回答から得られたデータが中核的な役割を担う。この③の部分について、筆者は年ごとに異なるテーマを設けて質問を作成し、アンケートを実施して来た。知りたいことは、受講生全体を1つの集団と捉え、「統計学の受講を選択した受講生が、講義開始時点

において、筆者の提示するテーマに関してどんな意識を持ち、どんな傾向を持つ人々の集まりであったのか」である。

まず1年目は、「スポーツや音楽、文学などに対する意識や好み」や、実際に絵画を見せた上での「絵画の好み」、「行動や気分の傾向」などについて、計18個の質問を行なった。

2年目は、「体格に関する意識」と「花粉症」の2テーマに絞り、計25個の質問を行なった。

3年目は、「萌え」という語を巡って、性差に関する社会通念についての問いを含め、さまざまな質問を計12個行なった。

上記のすべては、先述した通り、二者択一で答えを求める質問形式をとっている。そうして得られたデータは、各年の因子分析に関する例題を作成するための材料となっている。アンケート内容の概略は以上の通りである。

3. 例題としての因子分析事例について

本章では、前章で述べたアンケートから得られたデータを用いて、実際にどのような例題としての因子分析を行なうことが出来たのかについて、述べてゆく。先に結果を述べると、1年目の試みは、因子分析のごくごく初歩的な扱いのあらましを伝える例題として成功から最も遠く、2年目が最も成功に近かった。3年目は2年目ほどのインパクトは得られなかったが、1年目に比べればずっと成功に近い。以下、最も成功に近かった2年目における試みについて述べてゆく。

3.1 アンケートの具体的な内容

さて、2年目の試みにおけるテーマは「体格と花粉症」であった。受講生対象のアンケートで行なった質問の内、③のその他の範疇である程度の回答数を得て、因子分析を扱う例題作成に貢献した質問は8個あった。以下、その具体的な質問内容と結果を示す。便宜的に(1)～(8)という番号を付けておく。

- (1) あなたは自分自身を太っていると思いますか
 思う 22 (14) 思わない 23 (5)
- (2) あなたは自分よりも痩せている人を見て
 うらやましい 18 (14) うらやましくない 27 (5)
- (3) あなたは人から「細いね」と言われたら
 嬉しい 19 (12) 嬉しくない 26 (7)
- (4) あなたは、2ヶ月以内（現在を含む）に花粉症に罹りましたか？（医師に花粉症と診断された人も診断は受けていないが自覚はある人も「はい」と答えて下さい）
 はい 12 (8) いいえ 33 (11)
- (5) あなたは、過去に花粉症を自覚したり、花粉症であると診断されたりしたことはありますか？
 はい 18 (11) いいえ 27 (8)
- (6) あなたと血縁関係のある両親または兄弟姉妹に花粉症の人はいますか？
 はい 23 (10) いいえ 22 (9)
- (7) あなたは、花粉症は日本の国民病だと思いますか？
 はい 29 (13) いいえ 16 (6)
- (8) あなたは、日本政府は税金を使ってでも杉の木を減らすなど花粉症対策を行うべきだと思いますか
 はい 15 (8) いいえ 30 (0)

因みに、このアンケート実施日時は2005年4月12日第4限の統計学第1回講義開講時であり、アンケート実施場所が聖泉大学353教室、アンケート対象者は勿論、聖泉大学人間学部平成17年度統計学受講生で、アンケート回収数が45（内女子受講生によるものは19）、上記（ ）内の数字は女子の回答数である。

3.2 「体格と花粉症」についての因子分析結果

2年目の試みでは因子分析に係るテーマに「体格」を含めたため、アンケートより得た質問内容①の範疇にある「身長と体重」のデータも活用した。体重について未記入の回答が女子のみ3ケースあったため、有効なデータセット数は42セットとなった。これは、社会科学や人文科学領域において因子分析可能とされる30セットを超えるデータセット数である。因みに、最低200セットを必要とする自然科学領域では、不十分と判定されるセット数である点に注意されたい。また、分析を容易にするために、「体格」についても、②や③と同様の二者択一的なデータに置き換える必要がある。ここでは、身長と体重のデータより算出したBMI指数をもとに、「太りすぎ又は太りすぎ」と「やせ又はふつう」の2つのグループに分類した。BMI指数は体格及び体重指数の1つであり、その求め方は次式の通りである。

$$\text{BMI指数} = \text{体重 (kg)} \div \text{身長 (m)} \div \text{身長 (m)}$$

データを分類するにあたっては、BMI指数20未満を「やせ」、20以上24未満「ふつう」、24以上26.5未満「太りすぎ」、26.5以上「太りすぎ（肥満）」という判定基準に基づいて行ない、男女とも同一基準に従った。¹⁾ その結果、「太りすぎ又は太りすぎ」が13（内女子4）名、「やせ又はふつう」29（12）名となった。

次に、因子分析を可能にするために、「女／男」や「はい／いいえ」といった二者択一的に得られたデータを、それぞれ「1／0」というダミー変数に

1) BMI指数による肥満度の分類方法は複数あるが、次のURLで紹介されているものを用いた。<http://www.diettown.net/bmical/> 参照。最終確認年月日は2006年12月17日。尚、この判定基準を選んだ理由は、日本肥満学会のものよりも、今時の若者である受講生の気持ちにフィットするかもしれないと考えたからである。

置き換えた。

その上で、「体格」と「性別（女／男）」, 前章で述べたアンケート質問（1）～（8）の計10項目について、基本統計量と相関係数行列を算出した。ここでは相関係数行列のみ提示しておく（表1）。表1より、「体格」と（1）, 「性別」と（2）, （1）と（2）, （4）と（5）について、相対的に大きな相関係数となっていることが分かる。

表1. 平成17年度統計学受講生データの相関係数行列

	体格	性別	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
体格	1.0000									
性別	-0.1010	1.0000								
(1)	0.5297	0.3706	1.0000							
(2)	0.0384	0.5409	0.6204	1.0000						
(3)	-0.2071	0.2933	0.0751	0.3362	1.0000					
(4)	0.0326	0.3721	0.2723	0.2986	0.1550	1.0000				
(5)	0.0446	0.4105	0.1795	0.2582	0.1132	0.7303	1.0000			
(6)	0.0912	0.1220	0.1533	-0.0214	-0.0751	0.2572	0.3038	1.0000		
(7)	-0.0728	0.1387	0.1353	0.2108	0.1387	0.2236	0.2041	0.1691	1.0000	
(8)	0.2534	0.2339	0.2211	0.2741	0.2339	0.2986	0.2582	-0.0214	0.3162	1.0000

そして、これら10項目について因子分析を行なった。使用ソフトは、受講生がインターネットを通じて自宅等で簡易に復習できるように、群馬大学青木繁伸教授が公開するオンラインソフト「Black-Box」を選択した。²⁾ また、「Black-Box」では因子分析の適切性を判断する基準として、「Kaiser-Meyer-Olkinのサンプリング基準」が算出されるので、この値をもとに、これら10項目について因子分析を行うことが妥当かどうかを判断した。尚、この適切性基準についての詳細は、青木教授のWebサイトを参照されたい。³⁾

まず、10項目全てについてのデータを丸ごと因子分析にかけたところ、因子分析を行うことが適切と言える最低水準が満たされた。しかしながら、因子負荷量の計算結果より、不適切な特性（項目）を取り除く必要のあることが分かった。

2) 群馬大学青木教授の統計学学習サイトについては次の URL を参照。http://aoki2.si.gunma-u.ac.jp/ 最終確認年月日は 2006 年 12 月 17 日。

3) http://aoki2.si.gunma-u.ac.jp/lecture/PFA/pfa6.html 参照。

その後、不適切な特性を取り除くための再計算を繰り返し、表2の最終計算結果に至った。

表2. 平成17年度統計学受講生データ因子分析最終計算結果

	因子1	因子2	因子3	共通性
体 格	-0.02339	0.05489	-0.78465	0.61923
性 別	-0.32797	-0.62148	0.09668	0.50315
(1)	-0.07899	-0.66891	-0.71967	0.97161
(2)	-0.14393	-0.80534	-0.09905	0.67910
(4)	-0.70119	-0.24767	-0.05317	0.55584
(5)	-0.98932	-0.13537	-0.02686	0.99781
寄 与 率 (%)	26.75831	26.08223	19.27185	
累積寄与率 (%)	26.75831	52.84054	72.11239	

表2より、各因子の内容を決定する特性が「体格（太りぎみ又は太りすぎ／やせ又はふつう）」、「性別（女／男）」、(1)、(2)、(4)、(5)の6つに絞られたことが分かる。振り返っておくと、後者4特性（項目）は、(1)あなたは自分自身を太っていると思いますか「思う／思わない」、(2)あなたは自分よりも痩せている人を見て「うらやましい／うらやましくない」、(4)あなたは、2ヶ月以内（現在を含む）に花粉症に罹りましたか？（医師に花粉症と診断された人も診断は受けていないが自覚はある人も「はい」と答えて下さい）「はい／いいえ」、(5)あなたは、過去に花粉症を自覚したり、花粉症であると診断されたりしたことはありますか？「はい／いいえ」である。

また、特性毎の因子間の共通性が全て0.5以上と高くなった。3因子の累積寄与率は71%であり、これは、上記3つの因子によって、分析の対象となっている集団の約7割を説明できることを意味している。

では、各因子の示す内容について考えてみよう。各因子負荷量の絶対値の大きい特性（項目）に注目する。勿論、受講生アンケート結果ダミー変数表（表3）なども適宜参照することになる。

表3. 平成17年度統計学受講生アンケート結果ダミー変数表

整理番号	体格	性別	(1)	(2)	(4)	(5)
	太り気味又は肥満 1 やせ又はふつう 0	女 1 男 0	思う 1 思わない 0	うらやましい 1 うらやましくない 0	はい 1 いいえ 0	はい 1 いいえ 0
1	0	0	0	0	0	0
2	0	1	1	1	0	0
3	0	0	0	1	0	0
4	0	1	0	0	1	1
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	1	0	0	0	1
8	0	1	1	1	1	1
9	0	0	0	0	0	0
10	0	1	1	1	1	1
11	0	0	0	0	0	0
12	0	0	0	0	0	1
13	0	1	0	1	0	0
14	0	1	0	0	0	1
15	0	1	1	1	0	0
16	0	0	0	0	0	0
17	0	1	1	1	1	1
18	0	0	0	0	0	0
19	0	0	1	1	0	0
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	0	0	0	0	1	1
23	0	0	0	0	0	0
24	0	0	0	0	1	1
25	0	1	1	1	0	0
26	0	1	1	1	1	1
27	0	0	0	0	0	1
28	0	0	0	0	0	0
29	0	1	0	0	1	1
30	1	1	1	1	1	1
31	1	0	1	1	1	1
32	1	0	1	0	0	0
33	1	1	1	1	1	1
34	1	1	1	0	0	0
35	1	0	1	1	0	1
36	1	0	1	0	1	1

37	1	1	1	1	0	1
38	1	0	0	0	0	0
39	1	0	1	0	0	0
40	1	0	1	0	0	0
41	1	0	1	0	0	0
42	1	0	0	0	0	0

因子1について、特性(4)と(5)の因子負荷量の絶対値が大きいので、この因子は、花粉症に罹った経験のあることを意味していると考えられる。

因子2について、「性別」と(1),(2)の因子負荷量の絶対値が大きいので、自分が太っていると感じ、やせている人をうらやましく思っている女子の姿が浮かび上がってくる。とはいえ、データに占める女子の比率も絶対数も4割弱16名と小さいので、この因子は、さしずめ体重を気にしている様子を意味するものと考えればよいだろう。

因子3について、「体格」と(1)の因子負荷量の絶対値が大きいので、実際に「太りすぎみ又は太りすぎ」で且つ自分が太っていると自覚している人物の様子が窺える。

以上より、このアンケート対象者集団は、簡潔に述べると「花粉症に罹りやすく体重を気にしやすい傾向のある人」から成っていると結論できる。

この分析結果は、誰にでも分かりやすく、受講生にとって共感の持てるものであった。よって、この2年目における受講生データを用いた因子分析の例題作成は、ほぼ成功したものと言えよう。成功した理由については続く次章において考察する。

4. 成功要因と不成功要因

本章では、各年の試みにおける成功要因と不成功要因について考察する。

まず1年目の試みについて、受講生データのみでは適当な例題を作成できなかったため、窮余の策として、受講生データの他に、別のグループに同様のアンケートを行なうことで、2グループの比較を可能にする形を整え、例題を作成するに至った。別のグループは、滋賀県彦根市内の某公民館に定期

的に集う、年輩の女性中心のあるサークル例会出席者から成る。この通り、受講生データのみから講義開始時における受講生の集団的特性を明らかにしたいという意図からやや外れてしまったため、この年の試みは不成功である。その主原因は、アンケート内容「③その他の某かの質問」について、成功の可能性を大きくしようと多岐に渡る質問を設定したものの、そのために、かえってテーマが曖昧になってしまった点に求められよう。ここでの不成功要因は、テーマの散漫さである。

次に、2年目の試みでは、体重に主眼をおく「体格」と、「花粉症」という一般的で馴染みのある分かりやすい話題に照準を当て、テーマを絞って設定している。特に4月中旬に行なうアンケートであるため、「花粉症」は季節柄身近な話題でもある。アンケート対象者である受講生も、興味深く回答できたという。よって、前章で事例として詳述した通り、例題作成に適した分析結果が得られたものと考えられる。

3年目の試みについては、受講生が若者中心であることと流行とを意識して、「萌え」という語と性差に関する社会通念を巡るテーマを設定したが、得られた結果はいまひとつぼんやりとしたものとなってしまった。具体的には、統計学講義開始時点での受講生集団について、「萌え」という語に対する思い入れはそれぞれであるが、性差に対する意識はごく標準的であるような人々から構成されていると結論された。つまり、「萌え」の現代用法に肯定的な評価を与えようと与えまいと、性差に対する社会通念的なものの見方を当然のことと受け止めている人々である。2年目の試みに比べて3年目が不成功に終わった原因の1つは、受講生の少なさ故に、23セットしかデータを集めることができなかったというデータセット不足に求められる。しかし、原因はこれのみではない。実際のところ、アンケートに対して脈絡のない回答を示す個票データが少なくなかった。これは、「萌え」という語が想定していたほど受講生の間で一般的ではなかったため、アンケート対象者の方で回答しづらかったことの現れであろう。この点も、不成功の原因として認めざるを得ない。従って、データ不足とテーマにおける一般性の欠如が3年目

の不成功要因である。

以上より、受講生対象のアンケート内容「③その他の某かの質問」について、一般的で身近な話題を扱い、分かりやすいテーマに絞ることが、主要な成功要因であると結論できる。また、3年目の試みの反省として、アンケートで扱う話題を選定する際に、若者を意識し過ぎる必要はないということも指摘できよう。

5. まとめ

以上の考察では、心理学科において度々取り上げられる因子分析を例に取り、受講生の興味と意欲を引き出すのに有効と考えられる受講生データの初等統計学講義への活用を巡って、望ましい受講生データを得るためのアンケート内容について論じて来た。

前章で述べたように、初等統計学に相当する範囲内で因子分析に関する例題を作成するためには、アンケート内容を設定する際に、テーマを絞り、且つ、一般的で身近な話題に関連するテーマを選ぶことが肝要であることが明らかになった。

とはいうものの、受講生データの活用にも大きな欠点がある。例えば、受講生データは恣意的なデータであり、ランダムサンプリングによるものではない。限りある講義時間の中で受講生データを大いに活用すればするほど、自然科学領域での観察調査や社会調査等において当然とされるランダムサンプリングの大切さを、肌で感じるように受講生に学ばせることは困難である。ここをどう乗り越えて行くかは、講義構成及び講義運営上、今後の1つの重要な課題となろう。

《参考文献》

室淳子・石村貞夫 著『Excelでやさしく学ぶ多変量解析』東京図書, 1999年。