

項目反応理論によるテスト問題の分析

2001MM105 吉田 茉莉

指導教員 木村 美善

1 はじめに

塾にはたいてい入塾テストというものがある。私は入塾テストにより測りたいものを、より短時間で、より効果的に、正確に測ることができないかと考えていた。そして項目反応理論という、これを調べる最適な理論を知った。そこで、塾のテストの代わりに、ある大学のテストを代用に項目反応理論を使ったテスト問題の分析をしていくことにした。

2 データの詳細

ある大学の講義「統計学」の試験の結果をデータとしている。被験者は文系の学生で、P 学部 103 名、K 学部 86 名、その他 5 名である。男女の内訳は男子 106 名、女子 88 名である。試験時間は学部により異なり、P 学部が 80 分、K 学部が 50 分である。問題数も学部により異なり、試験時間の差の分だけ P 学部の方が 3 題多い。問題は、統計学の基本的な問題で大きく 6 題、全部で 25 問の構成になっている。25 問は両学部共通である。試験時には教科書・ノート類、および電卓の持ち込みが可である。得点のデータはすべて「0」「2」「4」で表されていて、その評価基準はどの問題も同じである。解き方も答えも間違っている場合は「0」、解き方はあっているが答えが間違っている場合は「2」、解き方も答えもあっている場合は「4」としている。

3 分析手法

3.1 コルモゴルフ = スミルノフ検定

2 つの分布関数が一致するとみなして良いかを検定するものである。得られたデータの特性値をグループ分けし、経験分布関数によって出力されたグラフの差に注目し、特徴のあるグループ分けであれば分析を進めていく。

3.2 項目反応理論

テストは普通、多くの項目によって構成される。項目反応理論では、個々の項目の正解率をただ求めるのではなく、測りたい特性（困難度や識別度など）の関数として表すことにより、個々の項目の特性を表すことができる。そして、被験者がどの程度の困難度の項目にどのように正誤しているかを基に、被験者の特性を推定する方法である ([2] 参照)。

3.3 項目反応モデル

個々の項目の特徴は項目特性関数によって表される。いくつかのパラメタをもつ関数を仮定して、これを項目反応モデルと言い、その中の 1 つにロジスティックモデルがある。主に使われるのが、識別力と困難度の 2 つのパラメタをもつロジスティックモデルであり、2

パラメタ・ロジスティックモデルと言う。今回、このモデルを使用するにあたり断っておきたいことがある。普通、項目反応理論を用いるときの条件で必要なのは、項目一つ一つが独立性を持っているということである。このテストの場合、問題自体は項目一つ一つに関連性があるのだが、項目反応理論を適用できるように、評価の仕方に工夫を入れた。データを「0」「1」に変換するとき、データの「0」を「0」で表し、データの「2」「4」を「1」で表した。

4 分析結果

4.1 最尤推定法

コルモゴルフ = スミルノフ検定により、男女間、学部間での関連性をみた。差は殆ど無いという結果が出たが、男性の経験分布関数がおかしな形をしていたことから、被験者の中に 1 次元性を持たない者がいるかを確かめるため、対数尤度関数を用いて最尤推定法により、モデルから逸脱している被験者を見てみた。予測に反し、逸脱している被験者に男女間の差は無かった。図 1 より、中間学力の被験者はレベルに応じた問題に正解している者が多く、低・高学力の被験者はレベルに応じた問題に答えられていない者がいることが分かった ([1] 参照)。

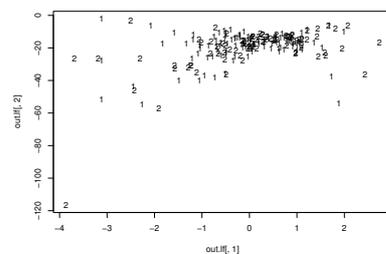


図 1 最大・最小尤度の被験者 (男女別)

4.2 項目反応理論

項目反応理論により出力されたテスト情報曲線から、中心より少し上の学力 (0.50 付近) の被験者に効果的なテストであることが分かった。識別度も非常に高く、学力を測る上で大変良いテストといえる。実際の合計得点と特性値を比較しても、56.8 点前後の被験者に効果的であることが分かり、大学のテストとして良好と判断して良いだろう。

図 2 は、テスト項目 1 問 1 問の情報曲線である。ここから識別度の高い項目は、13, 14, 15, 16, 17 の項目であることが分かった。5 つの項目がテスト全体の識

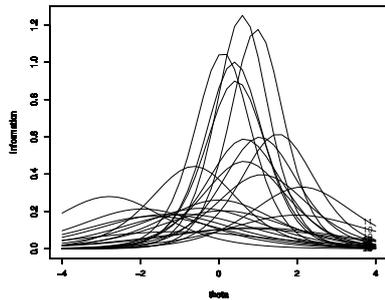


図2 項目情報曲線

別に、特に影響しているといえる。項目の識別度、困難度、正解率などから項目の特性が分かってきたが、特におもしろい項目について取り上げる。項目9は表からグラフを描く問題である。図を描くだけの簡単な問題であるにも関わらず、正解率が低い。これは教科書に例として図は載っているが、その図には間違いがあり、授業中に注意深く訂正されている。しかし、教科書通り描いた人が多かったと考えられる。よって、授業を聞いていなかった人が多かったということにつながる。もう一つ考えられるのが、累積度数分布表を勘違いして描いた人が多かったともいえる。他の項目についても詳しく分析できた ([3] 参照)。

5 テスト作成

5.1 作成方法

識別度の高い5つの項目をもとに、より精度の高いテスト作成をする。まず初めに、ここから60%で合否を判定するテストとして使用することを仮定する。精度の高いテストとはできる限り少ない項目数で同一の結果が表れるテストと考える。項目をZ問選出することを繰り返し、最良な結果を出す。

5.2 分析結果

Z問	24	10	23
True True	69	68	69
True False	8	16	10
False True	0	1	0
False False	117	109	115

表1 項目数ごとのテスト結果

識別度の高い5つの項目は、どれも学力が中間の被験者を識別するのに適した項目である。5つの項目がテストの中心的な項目となっている理由として、どれも教科書の公式から単純に解く計算問題であるため簡単な問題と判断されがちだが、項目13, 14, 15は度数分布表から数値を読み取る力が別に必要であるため、勉強をしていない人にとっては、テスト中だけで導き出すのは難しいと考えられる。そのため、識別度の高い項目となった。

項目16, 17は実際は単純な計算問題だが、回帰直線という関数の問題となっているため、数学に対して苦手意識をもっている者にとっては、ただそれだけで解けなくなってしまうのではないかと考えた。項目17は、特に導き出した関数に x を代入するだけで解答を出すことができるが、識別度が高い項目となっているのは、数学の苦手意識にあると考えた。

ある10項目を選出したとき、本テストの結果にほぼ近い値が出た。その項目は5つの項目と6, 19, 22, 23, 24である。項目22, 23, 24は確率変数の問題で、確率の知識も必要であるし、どの公式を使ったらよいかも問題をしっかり把握していないと分からない。計算も複雑であり、特性値の低い(学力の低い)被験者はこの項目に手をつけなかった者が多いと考えられる。そのため、しっかり勉強してきたものだけが正解を出すことができる問題となった。困難度が高いところで、識別度が高い項目である。

項目19は決定変数 r^2 を求める問題である。公式は単純であるにも関わらず、正解率が一番悪かった。これは1つに、 $r^2 = 1 - Se / \sum (y_i - \bar{y})^2$ の公式から解答を導き出そうとし、複雑な公式、分かりにくい解き方により、解くことを諦めた被験者が多かったのではないかと考える。 $r = \frac{S_{xy}}{S_x S_y}$ の2乗が決定変数である、という本文から $r^2 = \frac{S_{xy}^2}{S_x^2 S_y^2}$ を導くことができたか、できなかったかの違いによる結果ではないだろうか。しっかりと勉強していれば、気づく点である。

精度が高く、本テストと変わらない結果を出すテストを完璧に作ろうとすると、23項目必要であった。ほぼ全部の項目が必要になった理由として、困難度が低いところで識別度の高い項目(学力の低い被験者を測る効果的な項目)がなかったことが原因といえる。よって、本テストには簡単な項目で識別度の高い項目をいくつか用意すれば、約15項目のテスト作成ができるだろうと予測できる。項目反応理論を繰り返すことにより、困難度別に識別度の高い項目を見つけ出すことができる。そこから、いくつかの項目を選出し、項目の分析もしつつテストを作成すると、精度の良いテストができる。([1] 参照)

6 おわりに

本研究を通して、おもしろいテスト問題の分析ができた。今回使用したデータは非常に良かったので、まだ他に色々な研究をする価値があると思った。是非、後輩にはこのデータを使用し、おもしろい研究をして頂きたい。

参考文献

- [1] 安藤雅和・清水俊紀・木村美善：項目反応理論による多肢選択問題の分析，南山経営研究(1995)。
- [2] 芝 祐順：項目反応理論 - 基礎と応用 - ，東京大学出版会(1991)。
- [3] 白旗慎吾：統計解析入門，共立出版(1992)。