

東京学芸大学附属国際中等教育学校 2015 年度授業研究会

数学科 公開授業 I

数学を使い、生み出す統計授業

～バスケット分析を通して～

学 習 指 導 案

日時 平成 28 年 1 月 30 日（土曜日）9:00～9:50

対象 東京学芸大学附属国際中等教育学校

4 年 1・3 組 32 名

授業者 教諭 高橋広明

単元名 統計基礎

1. 扱う題材について

本授業では以下の課題の探究を行う。

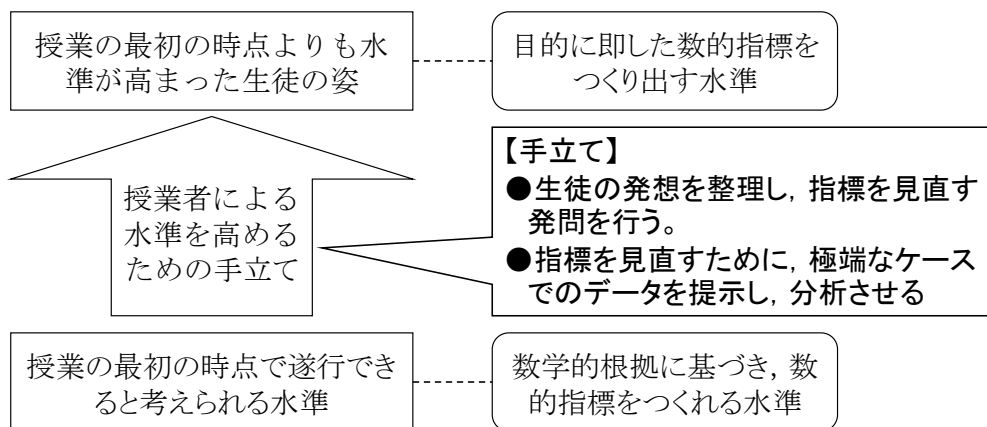
探究課題

私たちがスーパーマーケットなどで買い物をすると、レジでレシートが発行される。1枚のレシートを見ると、その客がどの商品とどの商品を同時に購入したのかが分かる。販売店ではこのような客の購買内容をデータとして蓄積し、そのデータを分析することによって販売促進に役立っている。次のデータは 20 枚のレシートについて購買の有無をそれぞれ 1 と 0 で表したものである。同時に購入する傾向の強い商品があれば、それらを近くに陳列すれば、商品の購入機会が増えることが期待でき、売上向上につながる可能性がある。これらのデータを分析し、同時に購入する傾向の強い商品を選んでみよう。

レシートID	商品1	商品2	商品3	商品4	商品5	商品6
1	1	0	0	0	0	1
2	1	1	1	1	0	1
3	1	1	1	1	0	1
4	1	1	0	1	1	1
5	1	0	0	0	0	1
6	0	0	0	0	0	1
7	1	0	1	1	0	0
8	1	0	1	1	0	1
9	0	0	0	0	0	1
10	1	0	0	1	0	1
11	1	0	1	1	0	1
12	1	0	0	1	1	1
13	0	1	1	0	0	0
14	1	0	1	1	0	1
15	0	0	0	0	1	1
16	0	0	1	0	1	1
17	1	0	0	0	0	1
18	0	1	1	1	0	1
19	0	1	0	0	0	1
20	1	0	1	0	0	1

2. 本題材で培いたい資質・能力

本題材は、販売促進・売上向上という目的のもと、同時に購入する傾向の強い商品を選び出すことにある。このためにはデータ分析を通し、数理的根拠に基づいてどれが対象となる商品なのかを判断する必要がある。数学を活用し、数学的論拠に基づき判断する力は学習指導要領の数学科の目標にも述べられている大切な力である。このとき、判断の基準となる数理的根拠は数値となる。すなわち、判断の基準となる数的指標を根拠に基づき自らつくり出すことが求められる。しかし何らかの根拠に基づいていても、それが目的に即していなければ使える指標とはならない。授業では、生徒がつくり出した指標を、販売促進・売上向上という目的に即した使える指標へと昇華させることをねらいとする。



育成すべき資質・能力の三要素との関連

i) 「何を知っているか、何ができるか」

…既習の内容や今までの経験をもとに何らかの指標をつくり出すこと

ii) 「知っていること・できることをどう使うか」

…すでにある指標を目的に照らしてその適切性を考え、よりよい指標を生み出すこと

iii) 「どのように社会・世界と関わり、よりよい人生を送るか」

…数学が現実に使われている場面での問題解決をし、協働的によりよい指標を考えること

3. 探究課題設定の理由

(1) 社会的背景からの理由

近年、統計教育の中で PPDAC サイクルが強調されてきた。これは統計的問題解決過程における一つの枠組みで、解決すべき問題が事前に明確に存在する際に、それを解決する目的のもとに、「その問題を明確化し」、「データ収集に向けての計画を立て」、「実際にデータを収集、クリーニング」、「そのデータを分析し」、「結論を得る」というサイクルである。どのようなデータを収集するかはその問題に依存する。そのため、どのようなデータをどのように収集するかという計画が大切になってくる。一方、世は“大量性(Volume)”，“多様性(Variety)”，“高速性(Velocity；リアルタイムなど高頻度更新)”の3Vによって特徴づけられる¹ビッグデータの時代である。特許庁ではビッグデータの分析・活用に関する期待と課題を、次

¹ IBM ではデータの矛盾や曖昧さを排除したものが大切であるという考え方のもと、“正確さ(Veracity)”を合わせた「4V」を提唱している [網野知博, 2013]。

の4つの大きな枠組みのもとにとらえている [特許庁, 2014]。

1. データドリブン社会の到来
2. 分析対象データの多様化
3. 付加価値の源泉は分析技術
4. データサイエンティストの確保が必須

ビッグデータ社会においては次から次へと大量かつ多様のデータが蓄積されていくため、どのようなデータを集めるのか、というPPDACサイクルにおけるPlan（計画）の相はもはやあまり意味をなさなくなる。したがって、このようなビッグデータ社会においてはすでに存在するデータからいかに有効な付加価値を見いだしていくのが重要である。いわば、データマイニングに基づくデータ分析の重要性である。本課題のレシート分析もPOSデータ分析の一つである。データから、意味のある有益な情報をマイニングしていく経験をもとに、データの多様な見方や分析の視点を共有していきたい。その一つとして、データ分析のための指標をつくり出す活動を行う。

(2) 学習指導からの理由

本校では統計領域について、中学1学年で1変量のデータ分析として、データの散らばりの視覚化としてのヒストグラムや箱ひげ図、データの特徴の数値化としての代表値や四分位数を学習している。中学2年生では2変量のデータを視覚化する方法として散布図で表すことを学習し、概念としての相関（正の相関・負の相関、相関の強弱）についても学習している。そのような学習歴を経て、4学年では1変量のばらつきの数値化としての分散、そして2変量の関係の強さの数値化としての共分散・相関係数を学習する。

	1 変量		2 変量	
	数値化	視覚化	数値化	視覚化
1 学年 (中 1)	代表値・四分位数	ヒストグラム・箱ひげ図		
2 学年 (中 2)				散布図
4 学年 (高 1)	分散		共分散・相関係数	

4年次にこれらの数値化について学習する際、いくつかの対象を比較し判断する場面を設定することで、数値化するという発想は生み出すことができよう。また、分散についてもその数値化に対する考え方も理解することはできるであろう。一方、共分散や相関係数についてはその考え方を理解させることは非常に困難であると考え。特に相関係数については、なぜその値の大小により線形性が判断できるのかを指導するのは難しい。いずれの教科書にもそれを合理的に説明しているものはない。各変量の平均の偏差の符号により、平均を原点とする座標系についてどの象限に点が属するかをもとに相関係数を説明しているに過ぎない。しかしそれでは、なぜ共分散を標準偏差の積で割った値により線形性が判断できるのかの説明にはなっていない。これに対し、大谷は相関係数の意味指導を回帰直線の最小2乗法に基づき試乗する流れを提案している [大谷晋, 2012]。だが、教師が手立てを細かく設定しないと、これを生徒自らが生み出させることは非常に困難であろう。これがピアソンの積率相関係数を高等学校で学ぶ限界と考えられる。しかし多変量データについてもその傾向を数値化する発想を生み出す経験はさせたい。その一つの方策として、探究課題のようなカテゴリカルデータの分析を提案したい。後述するように、この分析を通して相関係数までは行かずとも、その発想に近い考えは導き出せることも期待している。

4. 探究課題の社会的背景と数学的背景

本課題は POS データ分析の一つである。POS データ分析を販売促進に生かす取り組みは一般的であり、実際に成果をあげている。実際金子は「現代マーケティングでは当たり前になった「POS 明細データ」による、多品目間の購買実績の相関分析」を行うと、関連陳列すべき商品の組み合わせは意外なところから発見されることがある」として、POS データ分析から卵とインスタントカレーの間での購買の相関関係が非常に強いことに基づき、この 2 つを近くに陳列したことにより、以前を大幅に上回る販売実績を残したことを紹介している [金子哲雄, 2013]。

このようなデータからの知識発見(knowledge discovery and data mining)はデータマイニングの一つの産物であるが、近年はコンピュータの目覚ましい発達により、それをコンピュータに行わせること、すなわち機械学習が注目されている。教師なし機械学習の一つの応用例がレコメンデーションシステムである。これは、Amazon®での「この商品を買った人はこんな商品も買っています」に代表されるシステムである。その中でよく用いられるのが協調フィルタリングで、購買関連性の高い 2 つの商品 A, B に対して、商品 A を購入したユーザに商品 B を推薦するアイテムベースの協調フィルタリングと、購買類似性の高いユーザ A, B に対して、A が購入していて B が購入していない商品を B に推薦するようなユーザベースフィルタリングがある [D.Jannach 他, 2012]。Amazon のようなケースはアイテムベースの協調フィルタリングである。いずれのその指標としてピアソンの積率相関係数を用いることが多い。すなわち、アイテムベース協調フィルタリングでは、2 つの商品間の相関係数を、ユーザベース協調フィルタリングでは 2 人のユーザ間の相関係数をそれぞれ算出することとなる。本課題では購買の有無を 1 と 0 で表現したダミー変数を用いているが、ユーザが購買商品について例えば 5 段階で評価するなどのようなデータも考えられる。その場合も同じような評価のされ方、あるいは仕方について、アイテムベースおよびユーザベースでそれぞれ相関係数を用いて判断することもできる。本課題でのデータをアイテムベース及びユーザベースで相関係数による協調フィルタリングを行った結果が次の通りである。

	商品1	商品2	商品3	商品4	商品5	商品6
商品1						
商品2	-0.206					
商品3	0.105	0.218				
商品4	0.524	0.218	0.400			
商品5	-0.157	-0.055	-0.250	0.000		
商品6	0.105	-0.145	-0.333	0.000	0.167	

	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6	ID 7	ID 8	ID 9	ID 10	ID 11	ID 12	ID 13	ID 14	ID 15	ID 16	ID 17	ID 18	ID 19	ID 20
ID 1																				
ID 2	0.32																			
ID 3	0.32	1.00																		
ID 4	0.32	-0.20	-0.20																	
ID 5	1.00	0.32	0.32	0.32																
ID 6	0.63	0.20	0.20	0.20	0.63															
ID 7	0.00	0.45	0.45	-0.45	0.00	-0.45														
ID 8	0.50	0.63	0.63	-0.32	0.50	0.32	0.71													
ID 9	0.63	0.20	0.20	0.20	0.63	1.00	-0.45	0.32												
ID 10	0.71	0.45	0.45	0.45	0.71	0.45	0.33	0.71	0.45											
ID 11	0.50	0.63	0.63	-0.32	0.50	0.32	0.71	-0.25	0.63	0.00	-0.25	-0.25	0.50	-0.50	-0.25					
ID 12	0.50	-0.32	-0.32	0.63	0.50	0.32	0.00	0.25	0.32	0.71	0.25									
ID 13	-0.50	0.32	0.32	-0.63	-0.50	-0.32	0.00	-0.25	-0.32	-0.71	-0.25	-1.00								
ID 14	0.50	0.63	0.63	-0.32	0.50	0.32	0.71	1.00	0.32	0.71	1.00	0.25	-0.25							
ID 15	0.25	-0.63	-0.63	0.32	0.25	0.63	-0.71	-0.25	0.63	0.00	-0.25	0.50	-0.50	-0.25						
ID 16	0.00	-0.45	-0.45	-0.45	0.00	0.45	-0.33	0.00	0.45	-0.33	0.00	0.00	0.00	0.00	0.71					
ID 17	1.00	0.32	0.32	0.32	1.00	0.63	0.00	0.50	0.63	0.71	0.50	0.50	-0.50	0.50	0.25	0.00				
ID 18	-0.25	0.63	0.63	-0.32	-0.25	0.32	0.00	0.25	0.32	0.00	0.25	-0.50	0.50	0.25	-0.25	0.00	-0.25			
ID 19	0.25	0.32	0.32	0.32	0.25	0.63	-0.71	-0.25	0.63	0.00	-0.25	-0.25	0.25	-0.25	0.25	0.00	0.25	0.50		
ID 20	0.71	0.45	0.45	-0.45	0.71	0.45	0.33	0.71	0.45	0.33	0.71	0.00	0.00	0.71	0.00	0.33	0.71	0.00	0.00	

今回は商品の陳列を考察の対象にしているので、アイテムベースの協調フィルタリングが妥当である。その結果、商品 1 と商品 4 の相関が強いため、この 2 つの商品を近くに配置して陳列することがよいという結論が得られる。

レコメンデーションシステムにはこの他にも、商品 A を購入している人は商品 B も購入する傾向がみられるので、商品 A を購入した人に商品 B を推奨する考え方もある。これがアソシエーション分析である。レシートから、商品 A を購入した人が商品 B を購入する確率は $P_A(B)$ で算出することができる。この値が大きい方がよいわけだが、商品 A とは関係なく商品 B は購入されるのかもしれない。そこで、何の条件もなく商品 B を購入する確率 $P(B)$ と商品 A を購入した人が商品 B を購入する確率は $P_A(B)$ の比、すなわち、 $P_A(B)/P(B)$ が大きければ、商品 A と B は同時に購入される傾向が強いので、商品 A を購入した人に商品 B を推奨したり、商品 A と B を近くに配置して陳列し、売上向上を期待したりすることも考えられる。これに基づいて計算しても、商品 1 と商品 4 との組み合わせが最も数値が大きくなる。

5. 予想される生徒の反応とその分析

[考え方①]

商品 A と B において、全レシート（20 枚）に対する 2 つの商品を同時に購入したレシート数の割合で比較

この値は、「Russell-Rao 係数」と呼ばれている。

	商品1	商品2	商品3	商品4	商品5	商品6
商品1		0.15	0.35	0.45	0.1	0.6
商品2			0.2	0.2	0.05	0.25
商品3				0.35	0.05	0.4
商品4					0.1	0.45
商品5						0.2
商品6						

[考え方②]

商品 A と B において少なくとも一方を購入したものに対する同時に購入したものの割合で比較

この値は、「Jaccard 係数」と呼ばれている。

	商品1	商品2	商品3	商品4	商品5	商品6
商品1		0.1875	0.53846	0.64286	0.13333	0.63158
商品2			0.33333	0.33333	0.11111	0.26316
商品3				0.53846	0.07692	0.4
商品4					0.16667	0.47368
商品5						0.22222
商品6						

$$p = \frac{P(A \cap B)}{P(A \cup B)}$$

[考え方③]

商品 A を購入したもののうち、商品 B を購入したものの割合（条件付き確率）で比較
この値は、「 $A \Rightarrow B$ の信頼度」と呼ばれている。

	商品1	商品2	商品3	商品4	商品5	商品6
商品1		0.23077	0.53846	0.69231	0.15385	0.92308
商品2	0.5		0.66667	0.66667	0.16667	0.83333
商品3	0.7	0.4		0.7	0.1	0.8
商品4	0.9	0.4	0.7		0.2	0.9
商品5	0.5	0.25	0.25	0.5		1
商品6	0.66667	0.27778	0.44444	0.5	0.22222	

※下の表で、値 p は商品 A を購入したもののうち商品 B も購入したものの割合を表す。

	商品B
商品A	p

$$p = P_A(B) = \frac{P(A \cap B)}{P(A)}$$

6. 水準を昇華させるための手立て

[手立て①] データから読み取れる商品の特徴を問い、そこから推察されることを考える。

『特徴的な商品はないだろうか？その商品はどのような商品と考えられるだろうか？』

→商品 6 はほとんどの人が購入している。すなわちこれは通常よく購入される商品であることが考えられるので、この商品と他の商品 X を近くに陳列しても商品 X が購入されやすくなるとは限らない。

[手立て②-1] 生徒の発想を整理し、そこから目的に即した指標へと導く発問をする。

『この値の意味は何だろう？何が言えそう？』

→後述の考え方⑥を引き出す際、条件付き確率 $P_A(B)$ だけでなく、事前確率 $P(B)$ にも触れている反応を用いて、それらの意味を場面に即して解釈する。すなわち、条件付き確率 $P_A(B)$ は商品 A を購入した人が商品 B も購入する確率（割合）であるのに対し、事前確率 $P(B)$ は何の条件も考えずに商品 B を購入する確率（割合）であることに注目し、それらを比較することによって商品 A の効果が判断できることに気づかせる。

[手立て②-2] 極端なデータを示し、目的に即した指標を考えさせる。

『次のようなデータだった場合、商品 A と B を近くに陳列する場合と商品 C と D を近くに陳列する場合とではどちらの方が売り上げが向上することが期待できるか？』

→商品 D は、あるいは商品 C も商品 6 と同じように通常よく購入される商品である。したがって、商品 A と B の方が近くに陳列すると効果的であると考えられる。そうなるような指標

商品A	商品B	商品C	商品D
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
0	1	1	1
1	0	1	1
0	0	0	1
0	0	0	1
0	0	1	1
0	0	1	0
0	0	1	1
0	0	0	1

を考える。

	[考え方①]	[考え方②]	[考え方③]	
商品 A と B	4/12=0.333	4/6=0.667	$P_A(B)=4/5=0.8$	$P_B(A)=4/5=0.8$
商品 C と D	6/12=0.5	8/12=0.667	$P_C(D)=8/9=0.889$	$P_D(C)=8/11=0.727$

7. 引き出したい反応

[考え方④]

同じ行動（購入した・購入しなかった）の割合で比較

この値は、「一致係数」あるいは「Rand 係数」と呼ばれている。

		商品 Y	
		0	1
商品 X	0	a	b
	1	c	d

$$\frac{a + d}{a + b + c + d}$$

の値で比較

商品 A と B	10/12=0.833
商品 C と D	8/12=0.667

[考え方⑤]

同じ行動をしたものから、同じ行動をしなかったものの差の割合で比較

この値は、「Hamann 係数」と呼ばれている。

		商品 Y	
		0	1
商品 X	0	a	b
	1	c	d

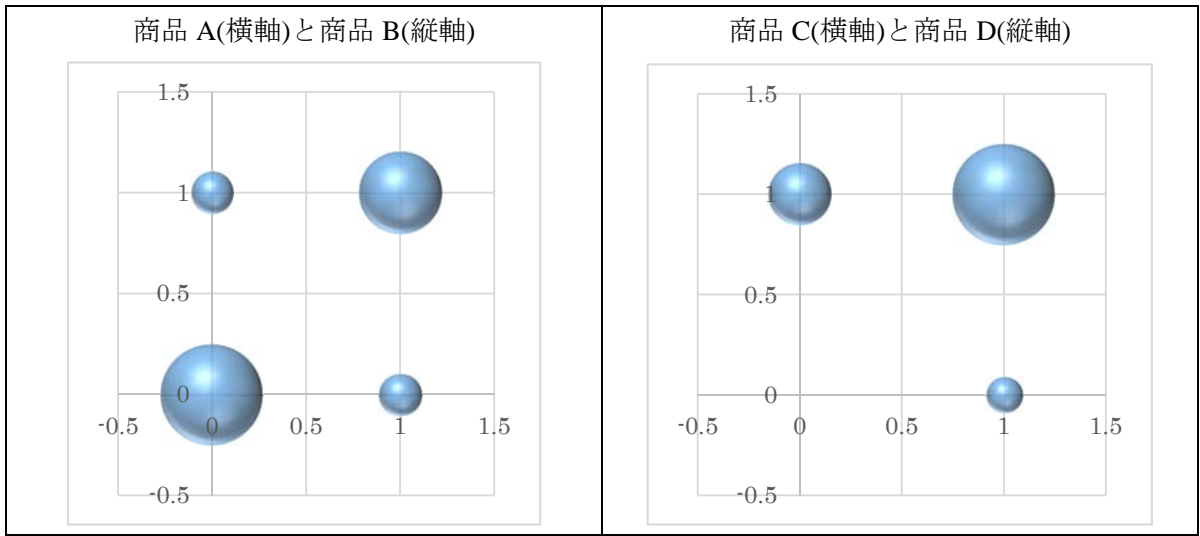
$$\frac{(a + d) - (b + c)}{a + b + c + d}$$

の値で比較

商品 A と B	$((6+4)-(1+1))/12=0.667$
商品 C と D	$((0+8)-(3+1))/12=0.333$

2組ずつの商品について、2値データをバブルチャートにて表したものが次のグラフである。このグラフから、右の方が直線傾向が強いことを見てとれることが期待できる。このグラフは2次元表より作成した。すなわち、グラフエリアを4領域に分けたときのそれぞれの領域に属する個数を表現したものとみることができる。今のデータはダミー変数である2値変数データであるが、これが量的変数である場合はどう考えられるだろうか。散布図としてデータをプロットしたとき、どのようにそのグラフエリアを4領域に分割することが自然であろうか。データがなるべく均等に4つの領域に属するように分割しようとすると、各データの平均値で区切るという発想が期待できる。相関係数につなげようとするとき、ここからの算出が天下り的になってしまう。その要因は、相関係数が回帰分析と密接に関連しており、回帰直

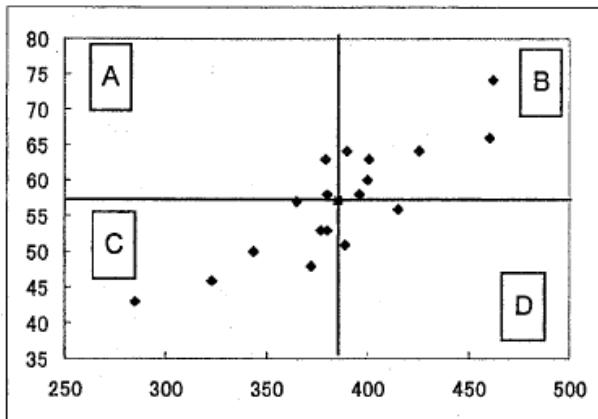
線を求める過程で相関係数が生み出されたが、その回帰直線を求めることが高等学校まででは難しいことにある。



これを改善する一つの方策として、アメリカの統計教育では、2変量のデータの関連性の強さを、四分度数比(Quadrant Count Ratio)と呼ばれる指標で考察することを提案している [藤井良宜, 2007]。この四分度数比は次の図のように、各データの平均値で A,B,C,D の4つの領域に分け、それぞれの領域に含まれるデータの個数 a,b,c,d に対して、

$$\frac{(b+c) - (a+d)}{a+b+c+d}$$

の値で与えられるものである。これは考え方④と類似しており、考え方⑤とはまったく同じ指標である。したがって、量的データの相関の強さを表す指標を自ら生み出す可能性が期待できる。



[考え方⑥]

条件付き確率を単体で購入する割合との割合で比較

この値は、「 $A \Rightarrow B$ のリフト値」と呼ばれている。

$P_A(B)$ は A を購入したときの B を購入する割合(確率)であるが、これと何の条件もなく B を購入した

ときの割合（確率）である $P(B)$ との割合を考える。すなわち、

$$\frac{P_A(B)}{P(B)}$$

が大きいほど、近くに陳列したほうが効果があると考えられる。

商品 A と B	$\frac{P_A(B)}{P(B)} = \frac{4}{\frac{5}{12}} = 1.92$	商品 C と D	$\frac{P_C(D)}{P(D)} = \frac{\frac{8}{9}}{\frac{11}{12}} = 0.97$
----------	---	----------	--

8. 指標についての整理

商品 X と商品 Y の状態が次のようになっているとする。

		商品 Y	
		0	1
商品 X	0	a	b
	1	c	d

今までの考え方の指標をまとめると、以下のようになる。

考え方	名称	指標
①	Russell-Rao 係数	$\frac{d}{a + b + c + d}$
②	Jaccard 係数	$\frac{d}{b + c + d}$
③	$X \Rightarrow Y$ の信頼度	$\frac{d}{c + d}$
④	一致係数	$\frac{a + d}{a + b + c + d}$
⑤	Hamann 係数	$\frac{(a + d) - (b + c)}{a + b + c + d}$
⑥	$X \Rightarrow Y$ のリフト値	$\frac{\frac{d}{c + d}}{\frac{b + d}{a + b + c + d}}$

これから分かるように、③や⑥の指標は、「ある性質 X をもつ対象が性質 Y を生じさせる傾向の強さ」を数値化したものであると見ることができる。すなわち、いわば因果関係を指標化したものである。したがって、例えば③の指標においては、「 $X \Rightarrow Y$ の信頼度」と「 $Y \Rightarrow X$ の信頼度」は一般には等しくならない。一方、③、⑥以外の指標は因果関係に基づいているわけではない。これらは 2 つの性質 X, Y の類似度を数値化したものであると見ることができる。類似度の高いものから順にグルーピングするなどして、事象を分類することなどに用いることができる。

9. 単元の指導計画と本時の位置づけ

(1) 単元の指導計画

「レシート分析」2時間（本時はその2時間目）

- レシートデータから売上向上を目的とし、同時に購入する傾向の強い商品を抽出する指標を作成し、その指標を吟味する。

「テストのでき具合は同じ？」1時間

- 平均点と同じ数学と英語のテストのデータから1変量データの散らばりを数値化する。
 - ・分散，標準偏差

「アイスクリームはどれだけ売れるだろうか？」2時間

- 複数の地域の気温とアイスクリームの売上額のデータから2変量の類似度を数値化し、どの地域のデータが売上高を予測するのに適しているかを判断する。
 - ・共分散，相関係数

「世論調査はどのように行われているの？」1時間

- 各新聞社が調査した内閣支持率の推移をもとに、全数調査と標本調査の違いを理解する。
 - ・標本調査，全数調査，母集団，無作為抽出

「標本調査は信頼できる？」1時間

- 母平均と標本平均との違いを標本の大きさを変えて実験することにより理解する。
 - ・標本平均

「オオクチバスの個体数を推定しよう」1時間

- 標識再捕法を例に、点推定により個体数を推定する方法を考える。

(2) 前時までの流れ

[第0時] 冬季休業中のレポート課題（別紙参照）としてレシートの分析を行った。

[第1時] レポートを返却し、グループでそれぞれの考え方を共有する。そして、どの考え方がもっとも適切と考えられるかを話し合い、グループとしての考え方をまとめる。

10. 本時案 別紙

参考文献

- D.Jannach 他. (2012). 情報推薦システム入門. 共立出版.
- 金子哲雄. (2013). 「衝動買い」が止まらない！ 商業界.
- 上田隆穂・他. (2014). リテールデータ分析入門. 中央経済社.
- 大谷晋. (2012). 「相関係数」学習指導案. 東京学芸大学附属高等学校情報教育公開研究会資料.
- 藤井良宜. (2007). 米国統計学会の統計教育ガイドライン. 日本数学教育学会誌第89巻第7号.
- 特許庁. (2014). 平成25年度特許出願技術動向調査報告書（概要）ビッグデータ分析.
- 豊田秀樹. (2006). 購買心理を読み解く統計学. 東京図書.
- 網野知博. (2013). 会社を強くするビッグデータ活用入門. 日本能率協会マネジメントセンター.

レポート課題 No.3 「商品の効果的な陳列をデータから探ろう」

課題

私たちがスーパーマーケットなどで買い物をすると、レジでレシートが発行される。1枚のレシートを見ると、その客がどの商品とどの商品を同時に購入したのかが分かる。販売店ではこのような客の購買内容をデータとして蓄積し、そのデータを分析することによって販売促進に役立っている。次のデータは20枚のレシートについて購買の有無をそれぞれ1と0で表したものである。例えば、レシートID1の客は、商品1と商品6を同時に購入していることが分かる。

同時に購入する傾向の強い商品があれば、それらを近くに陳列すれば、商品の購入機会が増えることが期待でき、売上向上につながる可能性がある。実際にこのような方法で飛躍的に売り上げを伸ばした事例が報告されている。ではこれらのデータを分析し、同時に購入する傾向の強い商品を選んでみよう。

レシートID	商品1	商品2	商品3	商品4	商品5	商品6
1	1	0	0	0	0	1
2	1	1	1	1	0	1
3	1	1	1	1	0	1
4	1	1	0	1	1	1
5	1	0	0	0	0	1
6	0	0	0	0	0	1
7	1	0	1	1	0	0
8	1	0	1	1	0	1
9	0	0	0	0	0	1
10	1	0	0	1	0	1
11	1	0	1	1	0	1
12	1	0	0	1	1	1
13	0	1	1	0	0	0
14	1	0	1	1	0	1
15	0	0	0	0	1	1
16	0	0	1	0	1	1
17	1	0	0	0	0	1
18	0	1	1	1	0	1
19	0	1	0	0	0	1
20	1	0	1	0	0	1

Criterion D: 実生活への応用 (Applying mathematics in real-life contexts)

	Task Specific
0	いかのいずれのもの達していない。
1-2	どの商品とどの商品を近くに陳列すべきかについて、データ分析の方法を少なくとも1つを提案しているが、論理的根拠に乏しい。
3-4	どの商品とどの商品を近くに陳列すべきかについて、データ分析の複数の方法を提案しているが、そのいずれも論理的根拠に乏しい。
5-6	どの商品とどの商品を近くに陳列すべきかについて、データ分析の方法を少なくとも1つを提案しており、それが論理的根拠に基づいている。
7-8	どの商品とどの商品を近くに陳列すべきかについて、データ分析の複数の方法を提案しており、そのいずれもが論理的根拠に基づいている。

10. 本時案

(1) 本時の目標

規準C：レポートですでに創出した指標について、売上向上の目的に即して、それらの指標が適切に機能するかどうかを考え、伝え合うことができる。

規準B：売上向上の目的に即して、すでに考え出された指標を改善・改良することができる。

(2) 本時の展開案

時間	学習内容と主な発問、予想される生徒の反応	留意点と手立て
15	<p>T：前回の班での考え方をまとめてみました。考えてもらっていたのは、数的根拠に基づいて、並べ方などを工夫すると売り上げの向上が期待できるような商品の組み合わせを見つけてもらうことでした。その根拠となる数値を指標とிட்டが、数値を導く式があるはずでず。そこで、各班の考え方を式で確認してみます。</p> <p>・教師がそれぞれの班の考え方を全体で確認し、各班の考え方を理解させる。</p> <p>【板書例】</p> <p>商品Aと商品Bを比較</p> <p>① $\frac{AとBの購入数}{Aの購入数}$ …1,3,5,6班</p> <p>② $\left(\frac{AとBの購入数}{Aの購入数}\right) + \left(\frac{AとBの購入数}{Bの購入数}\right)$ …2班</p> <p>③ $\left(\frac{AとBの購入数}{AまたはBの購入数}\right)$ …4班</p> <p>④ $\left(\frac{AとBの購入数}{総数}\right)$ …6班</p>	<p>・板書では式でまとめる。</p> <p>3班と5班については記述からはよく分からないので、確認する。</p>
10	<p>T：では結局どの指標が適切だと考えられる？</p> <p>S1：Aを購入した人がBも購入する傾向が強ければそれを近くに配置すれば売上向上が期待できるので、①の指標。</p> <p>S2：Aを購入した人がBも購入する確率と、Bを購入した人がAを購入する確率の和が大きければ、AでもBでも売り上げが向上する可能性が高い。</p> <p>T：結果的に選り出された商品AとBは近くに配置すると売上向上が期待できるけど、これは商品Aを購入すれば商品Bも購入されやすい、という向きがあるの？それとも向きは関係ないの？</p> <p>S3：向きは関係あると思う。</p> <p>S4：②だと、例えば確率が0.9+0.1も0.5+0.5も同じになるが、結果的に2つは近くに並ぶので、それなら確率が0.9ある方が効果が高くなるから、2つを足す必要はない。</p>	<p>・最終的には考えた班の多い①の指標に焦点化する。</p> <p>・向きは関係ないという意見に傾くときは、コーヒー売りにチョコレートが置いてある写真を提示し、コーヒーを買う人はチョコレートをかう傾向があるからと考えられることを提示する。</p>

10	<p>T: では①の指標が最も適切と考えていい?</p> <p>S5: 商品 A を購入する人が少ないときは①の値は大きくなる。例えば 20 人のうち商品 A を 1 人しか買っていなくても、その人が商品 B を買っていれば①の値は 1 となってしまう。</p> <p>T: このように極端な場合を例に考えようとする考え方は素晴らしい。ではこの場合は売り上げの向上は期待できないの?</p> <p>S6: 商品 A はほとんど売れない商品だから、これと商品 B を並べても、売り上げの向上にはあまり関係ない。</p> <p>T: では①の指標でも値が大きいからといって必ずしも売り上げの向上が期待できるというわけではないのかな。このレシートデータにはそういうパターンはある?</p> <p>S7: 6 班の表から、商品 E (商品 5) を購入した人が商品 F (商品 6) を購入する確率が 100%</p>	
15	<p>T: こうなる要因は何だろう? (特徴的な商品はない?)</p> <p>S8: 商品 5 があまり購入されないから。</p> <p>S9: 商品 6 が比較的良好に購入されるものだから。</p> <p>T: こういう場合どうする?</p> <p>S10: それらの商品は除いて考える。</p> <p>T: ではではどの程度よく買われたり、買われなかったりする商品なら除くという判断基準はあるの?</p> <p>S11: 買われるのは 9 割以上 S12: 8 割以上でいいかも</p> <p>T: 判断基準が主観的になってしまうね。では今のデータでは商品 5 を買う人は商品 6 も買うからこの 2 つを近くに配置しておけば売り上げが伸びることが期待できそう?</p> <p>【以下、指標を見直し改善する反応が出てこないときは発問する】</p> <p>T: 6 班は①の指標で計算した値を表にして書いてくれているけど、そこから何が分かる?</p> <p>T: 6 班は表のラベルのところにもパーセンテージが記されているけどこれは何?</p> <p>S13: A の 65%は商品 A (商品 1) を買った人が全体の 65%</p> <p>T: では商品 5 と商品 6 は結局どうなの?</p> <p>S14: 商品 5 を買った人が商品 6 を買う確率は 100%だけど、商品 6 はそもそも買われる確率が 90%あるから、商品 5 の効果はあったとしてもわずか。</p>	<p>手立て①</p> <p>手立て②- 1</p>
	<p>T: そのようなことを踏まえると、陳列を工夫すると売上向上が期待できる効果を測る指標はどう考えればよいただろうか。その指標に基づいて再度商品を洗い出してみよう。</p> <p>●最終的な結論は宿題とする。</p>	

【参考】 指標①～④を用いたレシート分析の値

①条件付き確率

	商品1	商品2	商品3	商品4	商品5	商品6
商品1		0.23	0.54	0.69	0.15	0.92
商品2	0.5		0.67	0.67	0.17	0.83
商品3	0.7	0.4		0.7	0.1	0.8
商品4	0.9	0.4	0.7		0.2	0.9
商品5	0.5	0.25	0.25	0.5		1
商品6	0.67	0.28	0.44	0.5	0.22	

②条件付き確率の和

	商品1	商品2	商品3	商品4	商品5	商品6
商品1		0.73	1.24	1.59	0.65	1.59
商品2			1.07	1.07	0.42	1.11
商品3				1.4	0.35	1.24
商品4					0.7	1.4
商品5						1.22
商品6						

③Jaccard 係数

	商品1	商品2	商品3	商品4	商品5	商品6
商品1		0.19	0.54	0.64	0.13	0.63
商品2			0.33	0.33	0.11	0.26
商品3				0.54	0.08	0.4
商品4					0.17	0.47
商品5						0.22
商品6						

④Russell-Rao 係数

	商品1	商品2	商品3	商品4	商品5	商品6
商品1		0.15	0.35	0.45	0.1	0.6
商品2			0.2	0.2	0.05	0.25
商品3				0.35	0.05	0.4
商品4					0.1	0.45
商品5						0.2
商品6						

