

# 商品開発で重視すべき 要因を明らかにする 商品評価分析システム マニュアル

エクセルでできる簡単な要因分析ツール

品質好選から消費者ニーズまで



## 刊行にあたって

我が国の農業を取り巻く環境は農業従事者の高齢化や市場のグローバル化等の進展で年々厳しさを増しています。その一方で、消費市場では農産物においてもニーズの多様化は加速しており、より消費者ニーズに即した商品（農産物・食品）開発が求められるようになって来ました。その要請に応えるためには、消費者ニーズを的確に把握しておくことが重要になります。

本マニュアルでは、このような課題に対応するために、より簡便で使いやすいリサーチ手法を提示するとともに、表計算ソフトで利用可能なソフトウェアを開発しその利用方法を紹介しています。本手法は、簡単なアンケート調査を基に消費者が重視する商品属性を明らかにするだけでなく、その重視度の大きさや、属性の構成の違いによる重視度の変化を数量的に把握することができます。このような情報は、商品開発を進める上で大いに役立てることができるでしょう。

本マニュアルが、農産物・食品開発のために活用され、国内農業の活性化の一助になれば幸いです。

独立行政法人

農業・食品産業技術総合研究機構

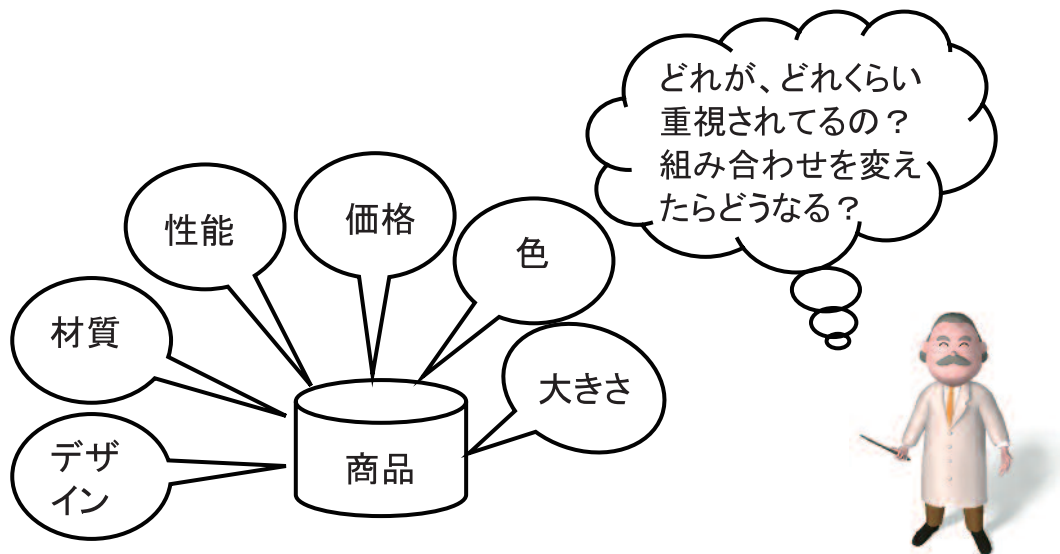
東北農業研究センター 所長 岡 三徳

商品開発で重視すべき要因を明らかにする  
商品評価分析システムマニュアル目次

1. 消費者は商品の何を重視するか	1
2. 理論編	2
1) ロジスティック回帰モデルとは	2
2) 計算方法	2
(1) ロジスティック回帰モデル式	2
(2) 具体的な計算方法	3
3) 目的変数（二値データ）、説明変数	3
(1) 目的変数	3
(2) 説明変数	3
4) オッズ・オッズ比	3
(1) オッズ (odds) とは？	3
(2) オッズ比とは	4
3. 実践編	5
1) 実際の調査・分析の手順	5
2) ロジスティック回帰分析の具体例	5
(1) 新作ジェラートの例	5
(2) 試食アンケートの方法	5
(3) 集計表の実際	6
(4) ロジスティック回帰分析ツールの使用方法	7
(a) 開始画面	7
(b) データの張り付け	8
(c) 基本設定（計算開始）	9
3) 分析結果出力表の見方	11
(1) モデル式の検定	11
(2) 回帰式	12
(3) カイ二乗検定（説明変数の検定）	12
(4) オッズ比の推定	13
(5) 偏回帰係数の意味とオッズ比の求め方	13
(6) シミュレーション	14
4) 交絡の有無の確認	16
5) 説明変数が量的データの時のオッズ比	17
6) 本ツールの使用に当たって	17
参考文献	18

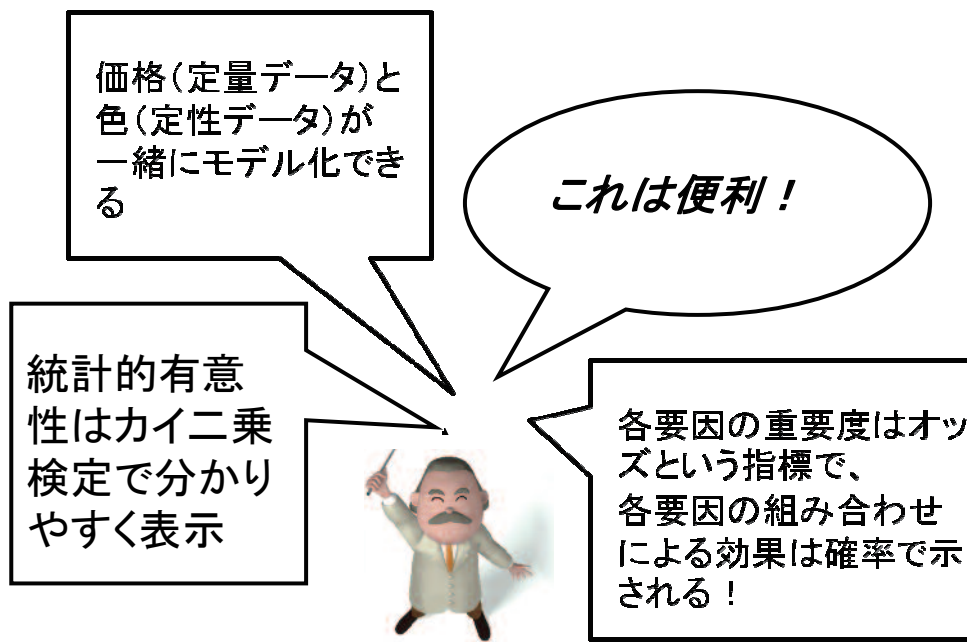
## 1. 消費者は商品の何を重視するか

「消費者は商品の何を重視するか」。これを見極めることは、これから商品開発を行おうとする際極めて重要です。また、単に重視する要因を突き止めるだけでなく、各要因が消費者の評価にどれだけの影響力を及ぼしているか、またその各要因の組み合わせによってどれだけの効果が期待できるかを数量的に把握することは商品開発の際に大いに役立ちます。



これまでも消費者ニーズ探索手法として多くの多変量解析やCS（顧客満足度）分析等が利用されてきました。しかし、定量データ（気温や価格等の数量的データ）と定性データ（男女や白黒等の質的データ）を混在させた分析が困難であったり、モデルの統計学的な妥当性や個々の変数の影響力が今ひとつ明確ではないという難点があります。さらに、それらの手法を用いるにしても専門的な統計学的知識が欠かせません。

マーケティングや社会科学を専門分野としない、より現場に近い研究者にとっては、データの取り方からモデルの構築、分析結果の解釈の仕方まで簡易でかつ統計学的にも裏付けのある分析手法が望まれていました。その1つの手法としてロジスティック回帰モデルが推奨できます。すでに、社会科学の分野でも利用されるようになって久しいのですが、まだ認知度は低く広汎に用いられるまでには至っていません。そこで、簡易に使えるエクセルのマクロを利用した「商品評価分析システム（ロジスティック回帰モデル分析ツール）」を作成しました。ロジスティック回帰モデルはこのような食味試験の分析の他にも、天候、気温、曜日等を説明変数にした直売所の売上予測や、借地料、面積、距離等を説明変数にした田畑の貸借予測等、多様な場面において適用可能です。本ツールはエクセルのアドインソフトです。本マニュアルはその理論から実際の使用方法までを分かりやすく解説してあります。



## 2. 理論編

### 1) ロジスティック回帰モデルとは

直線回帰分析や重回帰分析は体重から身長を予測するとか、店舗の面積と駅から店舗までの距離から売上高を予測する等という「数値」を予測するものです。一方、ロジスティック回帰分析は、ある模擬テストの結果(各教科の得点)から特定の大学への合格率を予測するとか、農薬の濃度の増加によって害虫の致死率がどのように変化するかといった「確率」を予測する手法です。したがって、その取り得る値は「確率」ですから、0～1の間になります。百分率にすると0%～100%ということです。これがロジスティック回帰分析の最大の特徴です。

### 2) 計算方法

#### (1) ロジスティック回帰モデル式

ロジスティック回帰のモデル式は以下のようなものです。

$$y = \frac{1}{1 + e^{-(a_1x_1 + a_2x_2 + \dots + a_px_p + b)}}$$

↑ 目的変数     
 ↑ 回帰係数     
 ↑ 説明変数     
 ↑ 切片

目的変数  $y$  は確率です。 $e$  は自然対数の底で  $e^{-a}$  は  $e$  の  $-a$  乗を意味します。

ヒント：自然対数のべき乗はエクセルの関数@ e x p (-a)  
で求めることができます。

## (2) 具体的な計算方法

ロジスティック回帰分析は多変量解析と呼ばれる手法の1つで対数を利用した「最尤法」という計算方法がとられています。多変量解析ですので、手計算では計算量が膨大すぎて不可能なためパソコンに頼らざるを得ません。具体的には行列式や逆行列式を用います。パソコンによる解の導出にはエクセルのソルバー機能が有効です。本ソフトでもソルバー機能を用いています。

## 3) 目的変数（二値データ）、説明変数

### (1) 目的変数

実際に計算に用いる目的変数のデータは、大学の合格率を求める場合なら、不合格0、合格1、というように01の二値データになります。直売所で新作の漬け物が売れる確率を調べたい時は、売れない0、売れた1、という具合です。量的データであっても、01に変換して目的変数にすることもできます。例えば、来店者数が100人未満なら0、100人以上なら1として目的変数にすることができます。この場合、求められる確率は来店者数が100人を超える確率ということになります。

### (2) 説明変数

説明変数にはカテゴリデータ（例えば、男性なら0、女性なら1）や数量データ（金額や薬剤の量など）をそのまま用いることができます。また、カテゴリデータと数量データが混在していても問題ありません。ただし、説明変数間に強い相関がある場合、重回帰分析の多重共線性（マルチコ）に似た交絡（強い相関が実際の影響力以上に評価されてしまう等）を生ずることがあります。説明変数の選択には相関の強い変数同士を組み合わせることのないように注意しなければなりません。交絡の有無を検証する方法は後述します。

## 4) オッズ・オッズ比

オッズ・オッズ比の意味は確率とは違います。

オッズとオズ比について詳しく説明しましょう。

### (1) オッズ (odds) とは？・・・

ある事象の起こる確率を  $p$  として、 $p/(1-p)$  の値をいいます。

例 農薬Aを使用したら害虫 100 匹中 20 匹に効果があった  
効果有りの確率 0.2 効果なしの確率 0.8  
odds=0.2/0.8=0.25 (オッズは 0.25 となります)

農薬Bを使用したら 100 匹中 40 匹に効果があった  
効果有りの確率 0.4 効果なしの確率 0.6  
odds=0.4/0.6=0.667 (オッズは 0.667 となります)  
すなわち、オッズは「効いた割合」／「効かない割合」  
= 「効き目」を意味しています。

## (2) オッズ比とは

以上のようにオッズは「効いた割合」／「効かない割合」、つまり「効き目」を示すものです。農薬Aの効き目は 0.25、農薬Bの効き目は 0.667 でした。では、農薬Aから農薬Bに代えたとき「効き目」の変化はどのようになるでしょう。それは双方の「効き目」の割合で判断します。すなわち、農薬Aの「効き目(オッズ)」は 0.25、農薬Bの「効き目(オッズ)」は 0.667 ですからその比は  $A/B = 2.667$  です。これは、それぞれのオッズの比(オッズ比)ですから 1 ならば同じ、1 より大きければ効き目大、小さければ効き目小ということになります。この例では、農薬Aから農薬Bに代えたことで「効き目」が 2.667 倍になったことを意味しています。この 2.667 がオッズ比(害虫にとっては死んでしまう[リスク]の大きさ)ということになります。何故、確率ではなくオッズ比なのでしょう？

先ほどの例で説明しましょう。

農薬A薬を使用したら害虫 100 匹中 20 匹に効果があった  
効果有りの確率 0.2 効果なしの確率 0.8  
odds=0.2/0.8=0.25 (オッズは 0.25 となります)  
農薬B薬を使用したら害虫 100 匹中 40 匹に効果があった  
効果有りの確率 0.4 効果なしの確率 0.6  
odds=0.4/0.6=0.667 (オッズは 0.667 となります)

効果有りの確率を比較すると 0.2 から 0.4 に増加していますからB薬に代えることによって2倍効き目があったと考えるかも知れません。しかし、それは「効果有りの数」が2倍になったのであって、「効き目」が2倍になったのではありません。「効き目」とは先述のように、「効果有り」と「効果なし」との割合のことです。「効き目」(オッズ)をみると、農薬Aは 0.25、農薬Bは 0.667 です。「効き目」の比(オッズ比)は2倍ではなく、 $A/B = 2.667$  となります。



### 3. 実践編

#### 1) 実際の調査・分析の手順

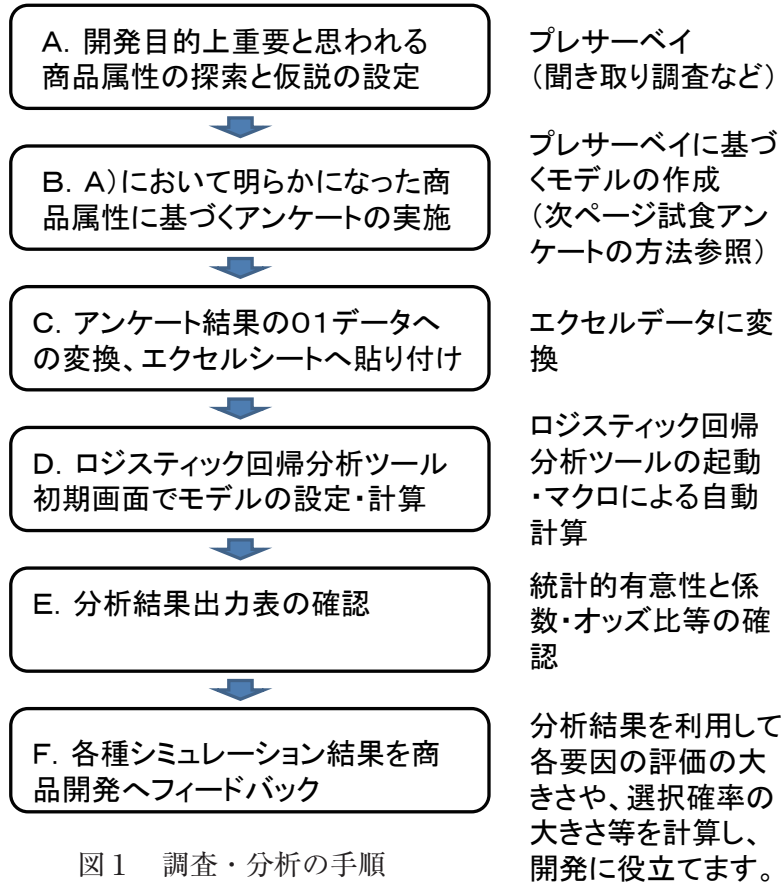


図1 調査・分析の手順

#### 2) ロジスティック回帰分析の具体例

##### (1) 新作ジェラートの例

村の直売所で特産いちごのジェラートを開発・販売することにしました。まずは、品質が問題になりますが、事前の試作・試食の結果、2つの検討課題が浮かび上がりました。一つはいちごの果肉をペースト状にするか果肉と分かる大きさに砕いて混ぜるか、もう一つは食紅を使用すべきか否かでした。

そこで、来店者を対象に試食アンケートを実施してその結果をロジスティック回帰分析で調べてみることにしました。この例ではサンプル数は20ですが、多い程結果の信頼度は増します。

##### (2) 試食アンケートの方法

まず、A：いちごペースト・食紅なし

B：いちご果肉・食紅あり

以上の2種類のジェラートを用意し、Aをコントロール（基準）にしてBについて評価し

てもらいます。AとBを食べ比べてもらい、Bの果肉入りが好きなら「果肉入り」の欄に○嫌いなら×、同じように食紅使用が好きなら○、嫌いなら×をつけてもらい、実際に購入するとしたらBを選ぶ場合は「選好」に○、Aを選ぶ場合は×をつけてもらいます。実際に分析に用いるのはこの○を1、×を0に変換した次ページのようなデータになります。

アンケート調査結果表

	評価項目		いちご	食紅
	Bのジェラート	選好	果肉入り	使用
回答者No	1	○	○	×
	2	○	○	○
	3	×	×	○
	⋮			
	⋮			
	20	○	○	×

(3) 集計表の実際

アンケート結果				
	評価項目		いちご	食紅
	Bのジェラート	選好	果肉入り	使用
回答者No	1	1	1	0
	2	1	1	1
	3	0	0	1
	4	1	1	0
	5	1	0	0
	6	1	1	0
	7	1	0	0
	8	1	1	0
	9	0	0	1
	10	0	0	0
	11	1	1	0
	12	0	0	1
	13	0	1	1
	14	1	0	1
	15	1	1	0
	16	1	1	1
	17	0	0	1
	18	1	1	0
	19	1	1	1
	20	1	1	0

このデータの「選好」がロジスティック回帰モデルのY、「いちご果肉入り」が $X_1$ 、「食紅使用」が $X_2$ になります。

(4) ロジスティック回帰分析ツールの使用方法

(a) 開始画面

CDの中に入っているファイル、ロジスティック回帰分析 2007)をクリックして立ち上げます。開始画面は以下のようになります。変数は16まで指定できますが、多すぎたりデータ数(サンプル)が少なかったりすると、ソルバーで解が求められなくなる可能性が大きくなります。相関の強い変数同士の組み合わせは避け、2~8変数程度に納めた方が良いモデルができるでしょう。

	A	B	C	D	E	F
1	変数	計算対象	変数名称			
2	Y				定数項	
3	X1	<input type="checkbox"/>			解(X1)	
4	X2	<input type="checkbox"/>			解(X2)	
5	X3	<input type="checkbox"/>			解(X3)	
6	X4	<input type="checkbox"/>			解(X4)	
7	X5	<input type="checkbox"/>			解(X5)	
8	X6	<input type="checkbox"/>			解(X6)	
9	X7	<input type="checkbox"/>			解(X7)	
10	X8	<input type="checkbox"/>			解(X8)	
11	X9	<input type="checkbox"/>			解(X9)	
12	X10	<input type="checkbox"/>			解(X10)	
13	X11	<input type="checkbox"/>			解(X11)	
14	X12	<input type="checkbox"/>			解(X12)	
15	X13	<input type="checkbox"/>			解(X13)	
16	X14	<input type="checkbox"/>			解(X14)	
17	X15	<input type="checkbox"/>			解(X15)	
18	X16	<input type="checkbox"/>			解(X16)	
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						

計算開始

シートクリア

基本設定 / 元データ / 計算シート / 定数項のみ / 単位行列 / M1行列 / M2行列

(b) データの張り付け

データを入力するために開始画面の下部バーにある「元データ」タブを選択します。

	A	B	C	D	E	F
1	変数	計算対象	変数名称			
2	Y				定数項	
3	× 1	<input type="checkbox"/>			解(X1)	
4	× 2	<input type="checkbox"/>			解(X2)	
5	× 3	<input type="checkbox"/>			解(X3)	
6	× 4	<input type="checkbox"/>			解(X4)	
7	× 5	<input type="checkbox"/>			解(X5)	
8	× 6	<input type="checkbox"/>			解(X6)	
9	× 7	<input type="checkbox"/>			解(X7)	
10	× 8	<input type="checkbox"/>			解(X8)	
11	× 9	<input type="checkbox"/>			解(X9)	
12	× 10	<input type="checkbox"/>			解(X10)	
13	× 11	<input type="checkbox"/>			解(X11)	
14	× 12	<input type="checkbox"/>			解(X12)	
15	× 13	<input type="checkbox"/>			解(X13)	
16	× 14	<input type="checkbox"/>			解(X14)	
17	× 15	<input type="checkbox"/>			解(X15)	
18	× 16	<input type="checkbox"/>			解(X16)	
19						
20			計算開始			
21						
22						
23						
24			シートクリア			
25						
26						
27						
28						
29						

ここをクリック

「元データ」シートに、エクセルのデータシート（別途作成しておいたアンケート結果の集計表）からデータ部分をコピーして貼り付けます。変数Xは16まで追加できますが、変数が多くなると有意な結果が得られなくなる場合があります。説明変数は相互に相関の低い少数に絞り込んだ方が良いでしょう。

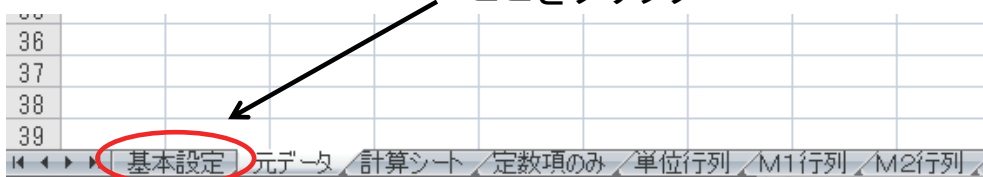
	A	B	C	D	E	F	G	H	I
1	Y	X1	X2	X3	X4	X5	X6	X7	X8
2	1	1	0						
3	1	1	1						
4	0	0	1						
5	1	1	0						
6	1	0	0						
7	1	1	0						
8	1	0	0						
9	1	1	0						
10	0	0	1						
11	0	0	0						
12	1	1	0						
13	0	0	1						
14	0	1	1						
15	1	0	1						
16	1	1	0						
17	1	1	1						
18	0	0	1						
19	1	1	0						
20	1	1	1						
21	1	1	0						
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									
39									

← データをコピーしてこの部分に貼り付け

(c) 基本設定 (計算開始)

データの貼り付けが終わったら、下部バーの「基本設定」をクリックします。

ここをクリック



チェックマーク  
説明変数としてモデルに組み込む場合チェックマークをつけます。変数名も書き込んでおくとい  
いでしょう。

「シート]クリア」  
は新しいデータ分析を行うときにク  
リックして、シート内  
をきれいにする時  
に使用します。

	A	B	C	D
1	変数	計算対象	変数名称	
2	Y		1 = 売れる、0 = 売れない	
3	X 1	<input checked="" type="checkbox"/>	果肉入り	
4	X 2	<input checked="" type="checkbox"/>	食紅	
5	X 3	<input type="checkbox"/>		
6	X 4	<input type="checkbox"/>		
7	X 5	<input type="checkbox"/>		
8	X 6	<input type="checkbox"/>		
9	X 7	<input type="checkbox"/>		
10	X 8	<input type="checkbox"/>		
11	X 9	<input type="checkbox"/>		
12	X 10	<input type="checkbox"/>		
13	X 11	<input type="checkbox"/>		
14	X 12	<input type="checkbox"/>		
15	X 13	<input type="checkbox"/>		
16	X 14	<input type="checkbox"/>		
17	X 15	<input type="checkbox"/>		
18	X 16	<input type="checkbox"/>		
19				
20			計算開始	
21				
22				
23				
24			シートクリア	
25				
26				
27				
28				

以上の作業を終えたら「計算開始」をクリックします。

「計算開始」をクリックすると自動的に計算が開始され、モニターに次のような「分析結果出力表」が表示されます。

### 3) 分析結果出力表の見方

- ケースの要約はデータ全体の要約です。
- 基本統計量はデータ全体の平均値や分散を示しています。

	A	B	C	D	E	F	G
1	<b>多重ロジスティック回帰分析出力表</b>						
2							
3	● ケースの要約		n			Y	
4	有効ケース		20			水準	n
5	目的変数のみ不明		0			0	6
6	説明変数不明		0			1	14
7							
8	● 基本統計量		平均		分散	標準偏差	
9	Y		0.7		0.21		
10	× 1		0.6		0.24	0.489897949	
11	× 2		0.45		0.2475	0.497493719	
12							
13	● モデル式の検定		対数尤度		-2対数尤度	尤度比	自由度
14	定数項のみ		-12.21728604		24.43457208		
15	定数項と説明変数		-6.8825265		13.765053	10.66951908	2
16							
17	● 回帰式					95%信頼限界	
18	説明変数		偏回帰係数		標準誤差	下限	上限
19	× 1		2.930861925		1.465722177	0.058046458	5.803677392
20	× 2		-2.550703325		1.487829109	-5.466848379	0.365441729
21	定数項		0.947877995		1.185030629	-1.374782038	3.270538028
22							
23	● カイ二乗検定(説明変数の検定)		Wald統計量		自由度	P 値	判定
24	× 1		3.998410692		1	0.045543196	*
25	× 2		2.939096149		1	0.086459115	
26	定数項		0.639802203		1	0.423782432	
27							
28	● オッズ比の推定					95%信頼限界	
29			オッズ比		下限	上限	
30	× 1		18.74377927		1.059764229	331.5164371	
31	× 2		0.078026769		0.004224525	1.441150466	

#### (1) モデル式の検定

モデルの検定は定数項のみの場合と、制約式の数  $q$  (ここでは  $x_1$  と  $x_2$  なので 2、「定数項と説明変数」) の対数尤度統計量に基づいて行われます。ここで求められた対数尤度比 (10.6695) は自由度  $q$  (ここでは 2) の  $\chi$  二乗分布に従うことが知られています。したがって P 値は 0.0048 となり、このモデルの帰無仮説は 1% 有意で棄却されます (意味のあるモデルであるといえます)。この尤度比はエクセルのソルバー機能を用いた計算で求められた解によって算出できるようになります。計算の実際は、計算シートの計算過程の入力関数を見ることによって理解できるようになると思います。

注：変数が多くなるとモデルの有意性が増す傾向があります。モデル自体の検定が有意であっても、変数の P 値がほとんど有意ではない場合は変数を減らしてモデルを再構築する必要があります。

●モデル式の検定	-2対数尤度	尤度比	自由度	P 値	* 5%有意
定数項のみ	24.4345				** 1%有意
定数項と説明変数	13.765	10.6695	2	0.00482	**

P値(000482)はエクセルの関数CHIDIST(X,自由度)、したがってここではCHIDIST(10.6695,2)で求めることができます。

## (2) 回帰式

●回帰式	説明変数	偏回帰係数	標準誤差	95%信頼限界	
				下限	上限
× 1		2.930861925	1.465722177	0.058046458	5.803677392
× 2		-2.550703325	1.487829109	-5.466848379	0.365441729
定数項		0.947877995	1.185030629	-1.374782038	3.270538028

$$y = \frac{1}{1 + e^{-(a_1x_1 + a_2x_2 + \dots + a_px_p + b)}}$$

分析結果出力表の「偏回帰係数」というのは上記モデル式の  $a_1 \dots a_p$  を示しています。

95%信頼区間は95%の確率で取り得る偏回帰係数の上限と下限を示しています。

## (3) ●カイ二乗検定(説明変数の検定)

説明変数の検定は「偏回帰係数 ( $a_1 \dots a_p$ )」が統計的に有意であるか否かを調べた結果です。

●カイ二乗検定(説明変数の検定)					
	wald統計量	自由度	P値	判定	* 5%有意
×1	3.9984	1	0.0455	*	** 1%有意
×2	2.939	1	0.0864		
定数項	0.6398	1	0.4237		

尤度比検定と同様に  $\chi^2$  二乗値をもとに検定されP値によって判定されます。このモデルでは  $a_1$  (X1) は5%有意で、 $a_2$  (X2) は5%有意に届きません。5%有意というのは当該係数



は95%の確率で意味がある。裏返すと5%の確率で意味がない危険性があることを意味しています。ですから、1%有意の場合はさらに信頼度が高まります。P値はエクセルでは統計関数のCHIDISTを用いて容易に求めることができます。(このカイ二乗統計量は行列式や逆行列式によって計算する必要があります。計算シートを見ると計算過程を学習することができます。)

#### (4) オッズ比の推定

分析結果出力表はそれぞれのオッズ比とその95%信頼区間の下限から上限が示されています。この結果ではX1(果肉入り)は18.7ですから、果肉入りのほうがペースト状よりも18.7倍の効果があり、X2(使用)は0.078ですから未使用よりはるかに－(負)の効果があることが分かります。

ですから、新たに売り出すいちごのジェラートは「果肉入り」で食紅は「未使用」のものが売れるだろうと予測できます。

●オッズ比の推定		95%信頼区間	
	オッズ比	下限	上限
×1	18.744	1.0597	331.5184
×2	0.078	0.0042	1.4411

#### (5) 偏回帰係数の意味とオッズ比の求め方

出力表にはオッズ比が示されますが、このオッズ比は簡単に求めることができます。それには、出力表の回帰式の各説明変数の偏回帰係数を利用します。

説明変数X1の偏回帰係数は2.931です。これを $e^{(a)}$ すなわち、自然対数の2.931乗したものがオッズ比(前ページ参照)となります。この場合、x1のデータ「1」は「果肉入り」、「0」は「ペースト状」なので、オッズ比は「果肉入り」と「ペースト状」オッズの比を表します。すなわち「果肉入り」の方が「ペースト状」の18.7倍購買確率に影響力を及ぼすこととなります。

※ 自然対数の2.931乗はエクセルの関数@exp(2.931)で求めることができます。  
同様にX2は@exp(-2.55)=0.078と求めることができます。

(6) ●シミュレーション

分析結果出力表の「●シミュレーション」で確率とイベントの発生の有無（この例の場合「売れる」「売れない」）のシミュレートが可能です。確率が0.5より大きければ「売れる」1、0.5以下なら「売れない」0と判定します。「果肉入り」なら1、「ペースト」なら0、同様に食紅「使用」1、「未使用」0を□の中に入力します。

果肉入り、食紅使用						
●シミュレーション				理論値		
x1	果肉入り		1	ロジット	1.328	
x2	食紅		1	確率	0.7905	
				判定	1	売れる
果肉入り、食紅未使用						
				理論値		
x1	果肉入り		1	ロジット	3.878	
x2	食紅		0	確率	0.9797	
				判定	1	売れる
ペースト、食紅未使用						
●シミュレーション				理論値		
x1	果肉入り		0	ロジット	0.9478	
x2	食紅		0	確率	0.7206	
				判定	1	売れる
ペースト、食紅使用						
●シミュレーション				理論値		
x1	果肉入り		0	ロジット	-1.602	
x2	食紅		1	確率	0.1675	
				判定	0	売れない

以上から、ペーストで食紅使用以外の場合、「売れる」ことが分かります。

●シミュレーション・・・応用編 I

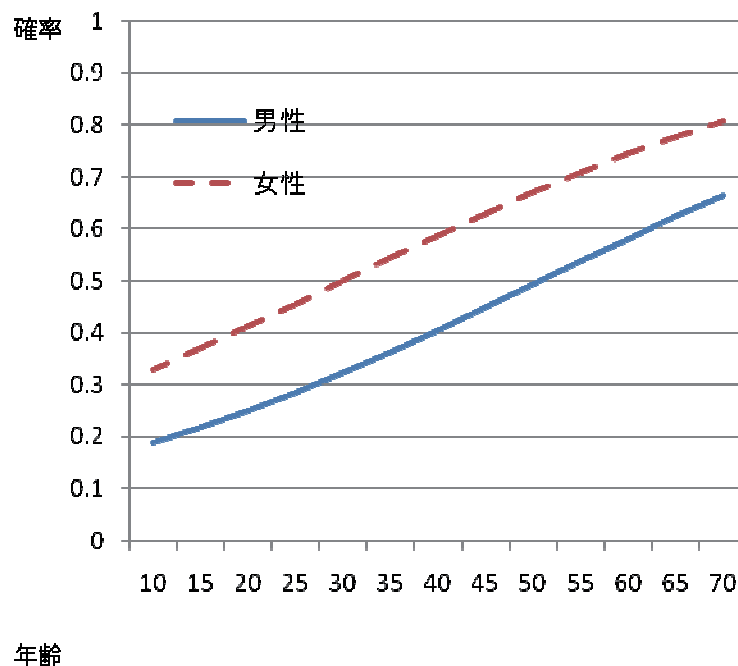
説明変数に価格や気温、年齢等の数量データがある場合、数量の変化によるシミュレートをするともとても分かりやすくなります。例えばある商品を選ぶ時に説明変数 x 1 に性別

(女性 = 1, 男性 = 0)、年齢 x 2 として計算した結果、 $X1 = 0.7421$ 、 $X2 = 0.0356$ 、定数項 = -1.8111 が得られた場合、以下のようにロジット（確率計算のための数値）を求めます。

男性のロジット  $0.7421 \times 0 + 0.0356 \times \text{年齢} - 1.8111$

女性のロジット  $0.7421 \times 1 + 0.0356 \times \text{年齢} - 1.8111$

以上で求められたロジットをエクセルの関数  $= 1 / (1 + \text{EXP}(-\text{ロジット}))$  で逆変換して確率を求めます。このように年齢ごとの確率を計算しグラフ化すると以下のようなグラフを描くことができます。



年齢や価格、気温などによってどのような確率変化をするのか明瞭になり、マーケティングの重要な情報として利用できます。

#### ● 購買反応曲線・・・応用編 II

ロジスティック回帰分析を応用した手法に「購買反応曲線」があります。これは「安すぎて不安を感じ始める価格」と「高すぎて買うのをためらい始める価格」の2つの回答から「購買反応曲線」（安過ぎるとも高過ぎるとも感じない価格）を求める方法です。本システムを利用することもできますが、より簡便な LOGNORMDIST というエクセルの関数を用いる方法をご紹介します。CD の「購買反応曲線」の添付シートを参照して下さい。

データ	安い価格	対数	高い価格	対数
1	20	2.995732	80	4.382027
2	45	3.806662	120	4.787492
3	50	3.912023	140	4.941642
4	55	4.007333	102	4.624973
5	48	3.871201	125	4.828314
6	65	4.174387	75	4.317488
7	80	4.382027	90	4.49981
8	25	3.218876	110	4.70048
9	48	3.871201	70	4.248495
10	25	3.218876	120	4.787492
	平均	3.745832	平均	4.611821
	標準偏差	0.451855	標準偏差	0.237906

アンケートで「安すぎて不安を感じ始める価格」と「高すぎて買うのをためらい始める価格」の2つの回答を得ます。それをそれぞれ対数に変換し、平均値と標準偏差を求めます。それを、CDの「購買反応曲線」のファイルのように

LOGNORMDIST (平均値、標準偏差) によって価格ごとの確率を求め、表の様に  $L(X)$ 、 $L(H)$ 、 $B(X)$  を求めます。添付シートのようにグラフ化するとより分かりやすくなります。

#### 4) 交絡の有無の確認

説明変数同士に強い相関がある場合、良好なモデルが作れなくなることがあります。交絡の有無は次のようにして確認することができます。

説明変数が5つあったとします。

その5つの説明変数をバラバラにして、説明変数が1つだけのモデルを5つ計算します。これが単変量モデルで、求められたオッズは「調整しないオッズ比」といいます。

今度は、5つの説明変数すべてを組み込んだモデルを計算します。これがフルモデルで、求められたオッズ比は「調整したオッズ比」といいます。

**表1) オッズ比による交絡の見つけ方**

	調整しない オッズ比	調整した オッズ比
X1	1.021	1.045
X2	1.568	0.873
X3	2.554	2.419
X4	1.003	0.995
X5	3.965	6.599

求められたオッズ比を見ると、X1、X3、X4は「調整しないオッズ比」と「調整したオッズ」はほとんど同じ値ですが、X2、X5は大きく異なっています。このような場合X2、X5は「交絡因子」である可能性があります。この対処法としては「交絡因子」を削除したモデルを作ることです。ただし、多変量解析の場合、必ずしもそれで良好なモデルが作れるとは限りません。試行錯誤を繰り返し、よりよいモデルを完成させるように工夫することが大切です。

時には、説明変数が5%有意にならなくても（あまりにもかけ離れていれば別ですが）モデルとして採用することもあります。

### 5) 説明変数が量的データの時のオッズ比

ジェラートの例では「果肉入り」1、「ペースト」0、食紅「使用」1、「未使用」0でした。「果肉入り」のオッズ比（0に対して1の時）は18.7、「使用」のオッズ比（0に対して1の時）は0.078でした。では、量的データ（薬剤量、気温、重量など）の場合、オッズ比はどのように理解したらよいのでしょうか。例えば、気温のデータが18度～29度あったとして、オッズ比が3.45だとしたら、このオッズ比は何を意味しているのでしょうか。

このような量的データの場合は1単位の変化に対応したオッズ比になります。すなわち、気温が1度上がった時のオッズ比です。従って、気温が1度上昇すると3.45倍イベントの発生に寄与することを意味しています。くれぐれも、18度から29度に上昇した時のオッズ比ではないことに注意してください。

### 6) 本ツールの使用に当たって

ロジスティックモデルは汎用性が高く、かつ統計学的な裏付けを持った優れた手法です。しかし、クリック一発で謎が解ける魔法の玉手箱ではありません。もっとも重要なことは、どのような説明変数をモデルに組み込むかということ、すなわち「仮説」の構築です。

そのためには、聞き取り調査や各種のプレサーベイを実施する必要があります。また、多変量解析の場合、微妙な数値のぶれや重複がモデルの適合性を悪化させてしまうことが珍しくありません。変数の追加や削除などを繰り返し、また多くのシミュレートを繰り返す中で分析者が学習していくことも大切です。手法はマーケットの学習ツールと位置づけるべきでしょう。

#### 参考文献

- [1] W.O.Spitzer,et al.:The use of  $\beta$  agonists and the risk of death and near death from a sthm.1992.
- [2] D.W.,Hosmer,& S.,Lemeshow:Applied Logistic Regression,John Wiley & Sons.1989.
- [3] 鷺田豊明・栗山浩一・竹内憲司（編）『環境評価ワークショップ 評価手法の現状』第2章、1999。
- [4] 丹後俊郎・山岡和枝・高木晴良『ロジスティック回帰分析－ SAS を利用した統計解析の実際－』朝倉書店、1996。
- [5] 高橋善弥太『－医者のための－ロジスティック・COX 回帰入門』日本医学館、第2章、1995。
- [6] 浜田知久馬『学会・論文発表のための統計学』真興交易医書出版部、1999。
- [7] 東京大学教養学部統計学教室編『自然科学の統計学』東京大学出版会、1992。
- [8] 対馬栄輝『SPSS で学ぶ医療系データ解析』東京図書、2007。
- [9] 下山禎「ロジスティック回帰モデルによる食味評価の分析」フードシステム研究、1-13、2009。

---

## 商品評価分析システムマニュアル

発行年月日 2011年2月

発行 独立行政法人 農業・食品産業技術総合研究機構

東北農業研究センター

〒020-0198 岩手県盛岡市下厨川字赤平4

編集 東北地域活性化研究チーム

Tel・Fax 019-643-3491

印刷会社 河北印刷株式会社

---

**リサイクル適性 (B)**

この印刷物は、板紙へ  
リサイクルできます。