ロジスティック回帰分析結果の解釈・利用のための新手法
─ 信用リスク・スコアリングモデルを例に ─

高田 直樹

New Techniques for Interpreting and Applying the Results of Logistic Regression Analysis
─ An Example of the Credit Risk Scoring Model ─

Naoki Takada

医学の分野で生まれたデータ分析手法であるロジスティック回帰分析は, 金融分野でも信用リスク・スコアリングモデルとして利用されている. 本論文はその分析結果をわかりやすく解釈, 利用するための新たな3手法を提案する. これらは, (1)各説明変数の増減がブラック率に与える影響を直接的に示す方法, (2)モデル構築用標本と推計対象標本のデフォルト率が異なる場合にロジスティック回帰式を調整する方法, (3)個別案件の各説明変数の値をブラック率への寄与の程度から評価する方法, である. これにより分析結果のより深い洞察も可能となる.

Logistic regression analysis was devised in the field of medicine, but is now also used as a credit risk scoring model in the field of finance. This article proposes three new techniques for interpreting and applying the results of this analysis in a way that is easy to understand. The three techniques are: (1) the method of directly showing the effect each explanatory variable has on the black ratio, (2) the method for adjusting the logistic regression formula in cases where the probabilities of default vary between the sample used for building the model and the sample used for estimates, and (3) the method for evaluating the value of each of the explanatory variables in a single item from the extent of the contribution to the black ratio. Using these methods enables an even deeper insight into the results of the analysis.

Key Words & Phrases： ロジスティック回帰分析, 信用リスク, スコアリングモデル, オッズ比, デフォルト率
logistic regression analysis, credit risk, scoring model, odds ratio, probability of default

1. はじめに

ロジスティック回帰モデル／分析は1960年代に生活習慣病の原因調査結果を分析する手法として生まれた[1]. この手法は医学や薬学, 生物学などの自然科学だけでなく社会科学の分野でも利用されるようになった. 金融分野では, 1970年代の終わりに, 企業や個人が支払い不能になる確率であるデフォルト率またはブラック率を推定する信用リスク・スコアリングモデルとして用いられ始めた[2][3]. この分野でロジスティック回帰分析は, ツリー分析[1]と並んで現在最も用いられている手法である[4]. なお本論文では便宜上デフォルトとなる確率を次のように使い分ける. モデルによって推計した個別案件の確率は「ブラック率」とし, 「ホワイト率(デフォルトしない確率)＝1−ブラック率」とする. また, 標本に含まれるデフォルト(ブラック)となった案件の割合を「デフォルト率」とする. 例えば1,000件の標本のうち20件が支払い不能だったとすれば, 2%がこの標本の「デフォルト率」である.

ロジスティック回帰分析は優れた手法ではあるが, その結果の解釈や利用にはある程度数学的な知見が要求されるため, 一般の人にとってはわかりづらい面がある. 筆者が本手法を個人向けローン商品の初期または途上の与信モデルや企業のブラック率推計

提出日：2006年08月31日　　再提出日：2007年3月5日

1 ロジスティック回帰分析のような明示的な関数を用いずに，一連の手順に沿ってデータを分けていくことにより，例えばブラック／ホワイトの判別やブラック率の予測を行う手法．

モデルを構築するプロジェクトを実施する中でも，構築したモデルをわかりやすく説明することが難しいと感じる場面があった．具体的には以下の3つの問題である．

(1) 各説明変数の増減がブラック率に与える影響の大きさが，直接的にはわからない．

(2) モデルを構築するために用いた標本（モデル構築用標本）とブラック率を推計する案件が属する標本（推計対象標本）が異なり，両標本のデフォルト率が異なる場合に，どのように調整すればよいかわからない．

(3) 構築したモデルを用いて各案件のブラック率は推計できるが，各説明変数の値がブラック率にどの程度寄与しているのか，わからない．

ロジスティック回帰分析は前述のように歴史も古くまた現在では広く使われつつある．しかし普及しつつあるということは，同時に統計分析を専門としない人も利用する機会が増えていることも意味する．本論文の出発点は，一般の人にもわかりやすく分析結果を解釈することについて，そのニーズはむしろ高まっているのではないか，また，その技術的な余地は残されているのではないかという筆者の問題意識である．本論文では，2章でロジスティック回帰分析について簡単に説明した後，3章〜5章において上記(1)〜(3)について検討した結果得られた知見を提案として説明する．3章〜5章での説明はそれぞれ，1. 論点，2. 提案と数学的根拠，3.数値例による確認，の順とした．

2. ロジスティック回帰分析概説

2.1　ロジスティック回帰分析とは

ロジスティック回帰モデルとは，ある現象が発生する確率Pを目的変数とし，その現象の出現を説明する変数群 $\vec{x} = (x_1, x_2,..., x_n)$ との間の関係が次のロジスティック関数によって表されるとするモデルである［1］［5］．

$$P(x_1, x_2,..., x_n) = \frac{1}{1 + \exp\{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)\}}$$

$$P(\vec{x}) = \frac{1}{1 + \exp\{-(\alpha + \vec{\beta} \cdot \vec{x})\}}, \quad P(z) = \frac{1}{1 + \exp\{-z\}}$$

ただし β_iは係数であり．$\vec{\beta} = (\beta_1, \beta_2,..., \beta_n)$．また，$z = (\alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n) = (\alpha + \vec{\beta} \cdot \vec{x})$である．ロジスティック関数を表す曲線は**図1**の実線のとおりであり，$P(z)$ は0と1の間の値（0%と100%の間の値）をとる．

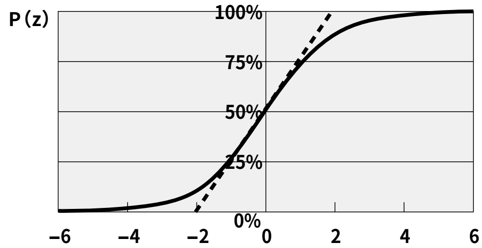

図1.　ロジスティック曲線

ロジスティック回帰分析とは，現象の有無と説明変数の値 x_i が明らかになっているデータを分析してこの式，すなわちαと各β_iを求めることである．説明変数の選択には，変数増減法[2]を利用し［6］，それらの係数α，β_iの決定には最尤法（さいゆうほう）[3]を利用する［1］［3］．この分析の利点は以下のとおりである．

(1) 種類が異なる複数の説明変数を考慮できる．（①複数の要因を説明変数として取り込むことができる．②連続数値項目だけでなく，カテゴリ項目も取り扱うことができる．）

(2) モデルの性質がよい．（①要因とオッズ（ブラック率とホワイト率の比，3.1で詳説）でみたリスクとの関係を把握しやすい．②構築したモデルが安定している．③統計ツールを用いれば短時間で分析できる．）

(3) 確率の予測に適している．（①“0”から“1”までの連続した値をとる$P(z)$を，ブラック率とみなすことができる．）

2.2　分析結果の数値例

分析結果の数値例を個人向け金融商品の信用リスク・スコアリングモデルでみてみよう．個人向けローン商品利用者は，最後まで完済する人が大多数だが，中には途中で返済不能となる人もいる．こうしたデフォルトの発生確率は，その人の個人属性や取引振りなどのプロファイルによってある程度予測できるだろう．これらを説明変数 $\vec{x}$ に，ブラック率 $P(\vec{x})$ を目的変数にすると，ロジスティック回帰モデルで表すことができる．

表1は1,000件弱のモデル構築用標本（デフォルト率＝2%）を用いたロジスティック回帰分析の結果例である．個別案件のブラック率 $P(\vec{x}) = P(z)$ は(a)列の説明変数 x_i（通常10個程度）で表される．各変数の係数 β_i は(b)列に示した．β_iが正の変数のときにはx_iの増加にともないブラック率が上昇し，β_iが負の変数のときには逆にブラック率は低下する．

---

2 回帰式に説明変数として採用する独立変数を探索的に選ぶ方法の一つ．回帰式に未採用の独立変数から最も予測に有効なものを順次回帰式に取り入れていくとともに，都度，既に取り入れた説明変数の中で除去すべきものがないかチェックすることにより，説明変数を決定する．

3 Maximum likelihood estimation，略してMLEともいう．与えられた標本からそれが従う確率分布の母数について推測するために よく用いられる方法．

表1. ロジスティック回帰分析の結果例

	(a) 説明変数	(b) 係数 β_i	(c) 係数 $\beta_i/4$	(d) オッズ比
0	定数 $\beta_0=\alpha$	x.xxx	−	−
1	勤務先規模_指数	-0.380	-0.095	0.684
2	自己資金比率(%)	-0.013	-0.003	0.987
3	配偶者(有=1, 無=2)	0.464	0.116	1.590
4	勤続開始年齢(歳)	0.022	0.006	1.022
5	ほかの借り入れ件数	0.284	0.071	1.329
⋮	⋮	⋮	⋮	⋮

表2. ブラック率の推計結果例

	(a) 説明変数	(b) 係数 β_i		(e) x_i		(f) = (b)×(e)
0	定数 $\beta_0=\alpha$	x.xxx	×	−	=	x.xxx
1	勤務先規模_指数	x.xxx	×	xx	=	x.xxx
2	自己資金比率(%)	-0.013	×	30	=	-0.390
3	配偶者(有=1, 無=2)	0.464	×	2	=	0.927
4	勤続開始年齢(歳)	0.022	×	36	=	0.794
5	ほかの借り入れ件数	0.284	×	0	=	0.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
z:合計						-1.684
$P(\vec{x})=P(z)$:ブラック率						15.70%

このロジスティック回帰モデルを用いた推計例を示したのが**表2**である. (e)列のような値(プロファイル)を持つ案件のブラック率 $P(\vec{x}) = P(z)$ は15.7%となる. これは標本のデフォルト率2%を大幅に上回っているから, デフォルトする確率は高いと考えられる. 通常, 標本のデフォルト率を上回るブラック率の案件は「ブラック案件である」と判断する.

3. 各説明変数の増減がブラック率に与える影響の大きさを直接的に示す方法

3.1 論点

ロジスティック回帰モデルに採用された説明変数がブラック率にどのように影響するかは, オッズ比によって説明される.

オッズ, そしてオッズ比について概説しておこう. ブラック率とホワイト率の比をオッズといいリスクの大きさを表す. 例えばブラック率＝0.5の場合, ホワイト率＝1ー0.5であるから, オッズ＝0.5 / 0.5 = 1 となる. 同様にブラック率＝0.75 の場合のオッズは0.75 / 0.25 = 3 となる. オッズ比は2つのオッズの比で定義され, 両者のリスクの差を意味する. ここでの例では, 前者と後者のオッズ比は, 3/1 =3 となり,「前者と比べて後者のリスクは3倍である」ことを意味する[5].

ロジスティック回帰モデルの結果として表示される各説明変数のオッズ比は, ある案件とほかの説明変数の値は同じで当該説明変数の値が "1単位 "だけ異なる案件のオッズ比である. 前述表1でみると(d)列に表示された数値がこれにあたる. 例えば表1では自己資金比率が1%増加するとオッズ比でみたリスクは0.987倍になる, すなわち増加によりリスクは減少する(ブラック率は減少する). 一方, 勤続開始年齢が1歳高くなると, リスクは1.022倍になる, すなわち増加によりリスクは増す(ブラック率は増加する). なおオッズ比(リスクの倍率)は, 説明変数の大きさには依存しない. 30歳と31歳でも, 40歳と41歳でもオッズ比は一定である. オッズがこのような意味を持つのはロジスティック回帰モデルの特徴である(というよりこうなることを前提としてモデルを構築している, といったほうが正確である). このように説明変数のブラック率への影響の方向(ブラック率の増減のいずれに寄与するか)とその大きさは, オッズ比を用いて解釈できる[1][5].

オッズ比による解釈は理路整然としている. しかし, 実務での利用者を含む一般の人にとってのわかりやすさという観点からすると課題がある. 第一に, 解釈する側はオッズ比という数学的な概念の理解を要求される. 第二に, オッズ比は説明変数の増減がブラック率の増加と減少どちらに作用するのかは教えてくれるが, ブラック率を増減させる大きさは教えてくれない.

3.2 提案と数学的根拠

筆者は, ①説明変数の増減がデフォルト率に与える影響を直接的に説明できる, ②理解のためにオッズ比のような数学的知見が不要である, という2条件を満たすロジスティック回帰モデルの解釈方法として次の方法を提唱したい.

「説明変数が1単位変化すると, ブラック率は『係数β_iを1/4にした分だけ』変化する. ただし変化分はスコア“ 0.5 ”の近傍で有効であり, スコアが“ 0 ”または“ 1 ”に近づくほど, その変化の大きさは小さくなる.」

これは数学的には, ロジスティック回帰式 $P(\vec{x})=P(z)$ を $z=0$ の近傍でテーラー展開し, その一次近似をとったものに相当し, 次の式で表される[4].

$$P(z) = P(\vec{x}) = P(x_1, x_2,\ldots, x_n)$$

$$\frac{1}{2}+\frac{1}{4}\cdot\alpha+\frac{1}{4}\cdot\beta_1\cdot x_1+\frac{1}{4}\cdot\beta_2\cdot x_2+\cdots+\frac{1}{4}\cdot\beta_n\cdot x_n$$

$$\triangle P(\vec{x})=\frac{1}{4}\cdot\beta_i\cdot\triangle x_i$$

ただし, $(x_1, x_2,\ldots, x_n) = (x_1, x_2,\ldots, x_n \mid z = 0)$

これにより,「ある説明変数が1単位増加すると係数

4 導出過程は次のとおりである.

$$P(z) \quad P(0)+\left.\frac{dP}{dz}\right|_{z=0}\cdot z$$

ここで, $P(0)=\frac{1}{2}$, $\left.\frac{dP}{dz}\right|_{z=0}=\frac{1}{4}$ **であるので本文中の式を得る.**

IBMプロフェッショナル論文❶ IBMプロフェッショナル論文❷ **IBMプロフェッショナル論文❸** IBM事例報告

β_iを1/4にした分だけ」ブラック率が変化することがわかる. これを図示したものが前出の図1の破線である. これは Z=0, すなわちデフォルト率＝50%におけるロジスティック回帰曲線の近似直線である.

3.3　数値例による確認

例として再び表1に戻って列（c）をみてみよう. 自己資本比率の $\beta_i/4$ は -0.003 であるから, ほかの条件が同じで自己資本比率が1%増加するとブラック率は0.3%低下する. また10%増加するとブラック率は3%低下する. 同様に勤続開始年齢が1歳高くなるとブラック率は0.6%増加, 10歳高くなると同様に6%増加する. ただし, これはブラック率が20〜80%程度の範囲内でのみ有効であることには留意が必要である.

このように近似的にではあるが, 本方法によって説明変数の増減のブラック率への影響を直接的に把握できることがわかる.

4. モデル構築用標本と推計対象標本のデフォルト率が異なる場合の調整方法

4.1　論点

ロジスティック回帰モデルを用いて個々の案件のブラック率を推計するには, 式に説明変数の値を代入すればよいことは前述した. ただしこの前提として, 推計対象標本のデフォルト率がモデル構築に用いた標本のデフォルト率と同じであることが必要である.

しかし, 推計対象標本のデフォルト率がモデル構築用標本のデフォルト率とは異なり, 上記の前提が成り立たないことがある. 具体的には以下のような場合である.

（1）説明変数に欠損値や異常値があるデータは, モデル構築用標本から除くことがある. このため, モデル構築用標本と推計対象標本のデフォルト率が同じにならないことがある.

（2）ツリー分析を併用する際, モデルの精度を高めるためブラックとホワイトの案件が同数になるよう抽出したものをモデル構築用標本とすることがある. この場合, モデル構築用標本のデフォルト率は50%であり, 推計対象標本のそれとは明らかに異なる.

前提が成立しない場合に問題となるのは, モデル構築用標本に基づいて作成したロジスティック回帰モデルを推計対象標本に適用できるか否か, できる場合はどのような調整が必要になるかである.

4.2　提案と数学的根拠

このヒントは, 丹後［1］や古川［7］による考察にある. 本論文ではこれを拡張し, モデル構築用標本から求めたロジスティック回帰式を利用して, 推計対象標本の個別データのブラック率推計式を導出する方法を提示する.

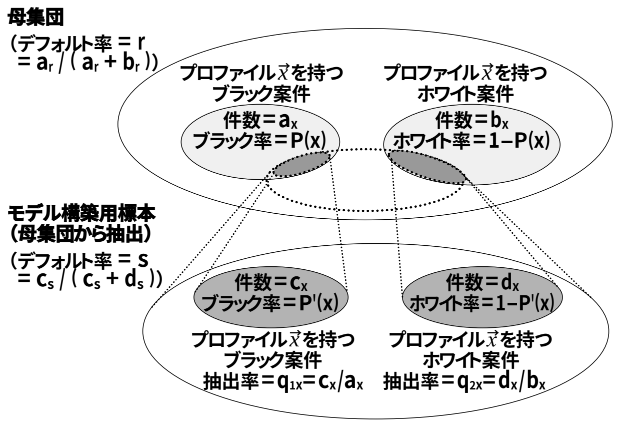

図2.　モデル構築用標本から求めたロジスティック回帰式から, 母集団のブラック率を算出する式を導出する検討の前提条件

まず, 丹後および古川の考察を確認しておこう. **図2**にはその前提の筆者なりの解釈を表している. 推計対象である母集団から抽出した標本を用いて, ブラック率 $P'(\vec{x})$を算出する次のようなロジスティック回帰モデルを作成する.

$$P'(\vec{x}) = \frac{1}{1+\exp\{-(\alpha+\vec{\beta}\cdot\vec{x})\}}$$

このうち, 任意のプロファイルを持つ案件（例えば前出の表2（e）列のような案件）に注目する. 母集団からモデル構築標本の抽出率がブラック案件とホワイト案件でそれぞれq_{1x}とq_{2x}だったとすると, モデル構築標本のブラック率$P'(\vec{x})$は母集団のブラック率$P(\vec{x})$からベイズの定理を用いて次のように表される.

$$P'(\vec{x}) = \frac{q_{1x}P(\vec{x})}{q_{1x}P(\vec{x})+q_{2x}(1-P(\vec{x}))}$$

抽出率q_{1x}とq_{2x}がプロファイル$\vec{x}$に関係なく無作為に適用されるとすれば, その抽出率をq_1, q_2として次の式を得る.

$$P'(\vec{x}) = \frac{q_1P(\vec{x})}{q_1P(\vec{x})+q_2(1-P(\vec{x}))}$$

以上の丹後および古川の考察を出発点とし, 以降で本論文の考察を進める. その主眼は, $P(\vec{x})$が$P'(\vec{x})$を用いてどのように表されるかにある.

第一にこの式を$P(\vec{x})$について変形し$P'(\vec{x})$のロジスティック回帰式を代入すると次式を得る.

$$\begin{aligned}P(\vec{x}) &= \frac{1}{1+\frac{q_1}{q_2}\exp\{-(\alpha+\vec{\beta}\vec{x})\}}\\ &= \frac{1}{1+\exp\{-(-\log\frac{q_1}{q_2}+\alpha+\vec{\beta}\cdot\vec{x})\}}\\ &= \frac{1}{1+\exp\{-(\alpha'+\vec{\beta}\cdot\vec{x})\}}\end{aligned}$$

この式を$P'(\vec{x})$と比較すると, q_1とq_2が入っている項$\log q_1/q_2$の分のみ異なっていることがわかる. これにより, モデル構築用標本からロジスティック回帰式 が作成されており抽出率q_1とq_2がわかっていれば, 母集団の推計式が得られる. この場合$P'(\vec{x})$のロジスティック回帰モデルに採用された説明変数とその係数$\vec{\beta}$は$P(\vec{x})$でもそのまま使うことができる. なお本論文と導出方法は異なるが,「定数項だけが変化し, それ以外の係数はまったく同じ解釈ができることがわかる」ことは, 丹後によっても述べられている[1].

ここまで$P'(\vec{x})$から$P(\vec{x})$が導出されることがわかったが, これに必要な抽出率, q_1/q_2はそのままではわからない. これは何か, が第二の考察である.

抽出率は案件のプロファイルに因らないので, 母集団のデフォルト率とブラック率が同じ案件($a_r/(a_r+b_r)=r$)についても適用してもよいだろう. これら案件がモデル構築標本として抽出されると, そのブラック率は標本全体のデフォルト率に等しくなる($c_s/(c_s+d_s)=s$)はずである. したがって, 抽出率の項q_1/q_2は以下のように表すことができる.

$$\frac{q_1}{q_2}=\frac{q_{1rs}}{q_{2rs}}=\frac{c_s}{a_r}\cdot\frac{b_r}{d_s}=\frac{a_r+b_r}{a_r}\cdot\frac{b_r}{a_r+b_r}\cdot\frac{c_r}{c_s+d_s}\cdot\frac{c_s+d_s}{d_s}$$
$$=\frac{1-r}{\mathrm{r}}\cdot\frac{s}{1-r}$$

これを$P(\vec{x})$に代入すると次式を得る.

$$P(\vec{x})=\frac{1}{1+\exp\{-(-\log\frac{1-r}{r}\cdot\frac{s}{1-s}+\alpha+\vec{\beta}\cdot\vec{x})\}}$$

すなわち,「モデル構築用標本からロジスティック回帰式 が作成されており, 母集団と抽出した標本それぞれのデフォルト率rとsがわかってさえいれば, 母集団の各案件のブラック率の推計式が得られる」ことになる.

第三の考察としてこれを次のように拡張することができる. 同じ母集団から得られた任意の2標本について, それぞれのデフォルト率sとtがわかっており, 一方の案件のブラック率を表すロジスティック回帰モデル$P_s(\vec{x})$がわかっていれば, 本調整によって他方のブラック率$P_t(\vec{x})$の推計式を得ることができる(**図3**).

4.3　数値例による確認と本手法の応用

簡単な数値例を使って, この調整前後の回帰式の関係を確認しておこう. **図4**の実線の回帰曲線は, デフォルト率r=1%の標本からブラック案件とホワイト案件を同数抽出し標本(デフォルト率s=50%)となし, これに最尤法を適用して算出したロジスティック回帰曲線である. 破線は, この式に前節の調整を加えて作成した, デフォルト率1%の標本の個別案件のブラック率を算出するロジスティック回帰曲線である. 後者は前

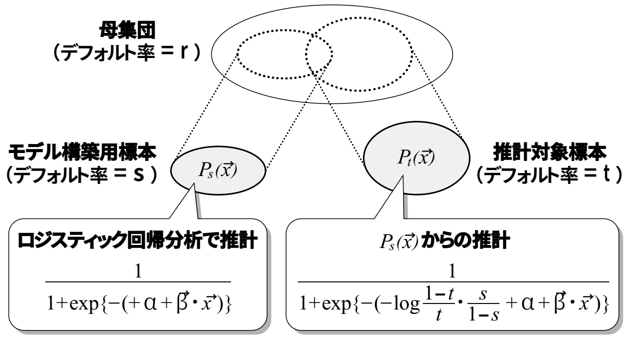

図3.　母集団が同じ標本のロジスティック回帰式からのブラック率の推計

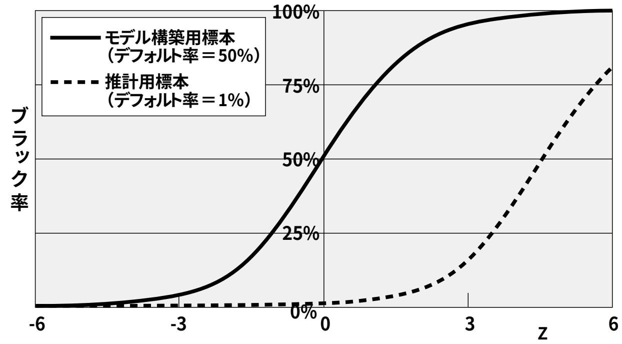

図4.　モデル構築用標本から推計したロジスティック曲線(実線)とこれを調整し推計用標本に適用するロジスティック曲線(破線)

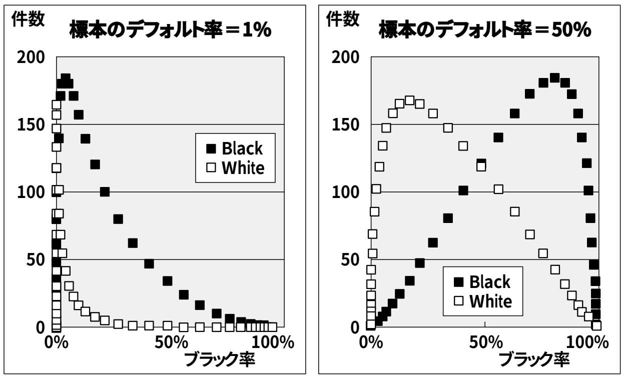

図5.　ブラック率1%ごとの案件の分布

者をZ軸方向に沿ってlog{(1-r)s/(1-s)r}だけ平行移動したものであることが確認できる. ちょうどとなる案件のブラック率は, 実線の回帰モデル(デフォルト率s=50%)では50%, 破線のモデル(同r=1%)では1%となる.

本調整の応用例を示しておこう. 第一に回帰式の視覚的な検証での利用である. **図5**左グラフはデフォルト率1%の推計対象標本を用いて推計したモデルから各案件のブラック率$P_1(\vec{x})$を算出, ブラックおよびホワイトの案件ごとにその度数をブラック率1%刻みで示したものである. この図は推計対象標本の実態を正確に表している. しかし, ブラック案件とホワイト案件の分布はブラック率が低いほうに偏っており,

推計したモデルが両案件を見極めているか確認することは難しい. 一方**図5**右グラフは先に示した調整によって標本のデフォルト率が50%になるように調整したモデルを作成して$P_2(\vec{x})$を算出, 同様にブラックおよびホワイト案件の分布を示したものである. これによりブラック案件とホワイト案件が左右に分かれて分布しており, モデルの見極めの程度を視覚的に確認できる[5].

第二の応用例は, 推計対象標本とは異なるデフォルト率を持つ標本のブラック率を「スコア」として用いることである. 例えば, デフォルト率1%の標本の各案件について, ブラック率とともにスコアの情報も提供することを考えよう. このスコアとして例えば標本のデフォルト率が50%とした場合のブラック率$P_2(\vec{x})$を採用することが考えられる. 前出の図5をみると明らかなように, このスコアを用いるとブラックとホワイトの境目をより拡大してみることができるからである. $P_1(\vec{x})$から$P_2(\vec{x})$への調整を前節で示した方法で行い, 算出したブラック率(標本のデフォルト率＝1%)とスコア(標本のデフォルト率＝50%)を縦横に示したのが**図6**である. 例えばスコア50の案件はブラック率1%, スコア80の案件はブラック率約3.5%となる.

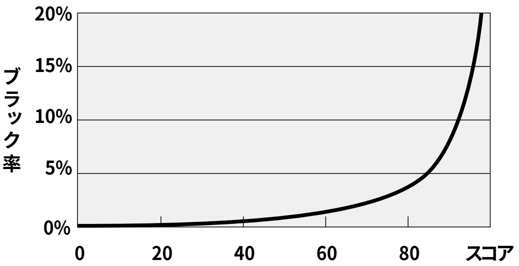

図6.　スコアごとのブラック率

このように本章で示した調整方法を用いて, 標本のデフォルト率を変えてみせることにより, 分析結果を使いやすくできる.

5. 個別案件の各説明変数の値をブラック率への寄与の程度から評価する方法

5.1　論点

構築されたロジスティック回帰モデルを用いれば, 各案件のブラック率を推計できることは前述した. しかし, ある案件の各説明変数の値が, ブラック率からみてどの程度であるかについてはわからない. 例えばブラック率が同じ程度高い案件であっても, その要因が「ある説明変数が非常に悪いこと」にあるのか, 「複数の説明変数が少しずつ悪くその合わせ技」であるのかは不明である. 当該案件の説明変数の値x_iをみても, こうした情報を得るのは難しい.

5.2　提案と数学的根拠

本章では, 各案件の説明変数の値がブラック傾向に効いたのかホワイト傾向に効いたのか評価するとともに, これを一覧することができる方法を提示する. これは3章で提示した式を変形した以下の式に基づく.

$$P(\vec{x}) \quad \frac{1}{2}+\frac{1}{4}\cdot\beta_1\cdot x'_1+\frac{1}{4}\cdot\beta_2\cdot x'_2+\cdots+\frac{1}{4}\cdot\beta_n\cdot x'_n$$

ここで, $x'_i = x_i - x_{i0} \quad x_i - \bar{x}_i$

$(x_{10}, x_{20},\ldots, x_{n0}) = (x_1, x_2,\ldots, x_n \mid z = 0)$

$\bar{x}_i$：モデル構築用標本の平均値

この式をみると, 各説明変数について, 標本の平均値からの差 $x'_i = x_i - \bar{x}_i$ に係数β_i/4を乗じた値により, その説明変数がブラック率にどのように効いているかを評価できる.

5.3　数値例による確認

実際の数値例で確認してみよう. まず, ロジスティック回帰式によって算出したブラック率が高かった案件のうちから, 特徴があるものを選び, 各変数について$\beta_i \cdot x'_i$/4を計算した. 次にこの値をみやすいよう, レーダーチャート上の中心から離れる方向にプロットし線で結んだのが**図7**である. 中心から離れるほどブラック率に効いている. また, 図上で“0”であることは, その変数の値x_iがこのモデル構築に利用した標本の平均値に等しいことを意味する.

案件ごとにブラック率を高める要因を考察してみよう. 濃い実線で表される案件はほかの借り入れ件数が突出して多いことが要因となっている. 破線の案件はほかの借り入れ件数が多いことと勤続開始年齢が高いことが要因である. 薄い線で表される案件は,

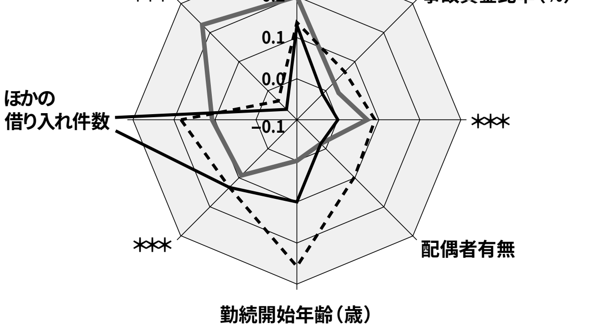

図7.　各説明変数のブラック率への寄与

5. ただし, 回帰式がブラック案件とホワイト案件とをよく見極めているかの正確な確認は, KS値やダイバージェンス, ARなどを用いて別途行う必要がある[8].

勤務先の規模が小さいことと本論文では示さなかったほかの変数が要因である．

これらの例によりブラック率が同程度でも，案件ごとに特徴があること，そして本手法によってその違いを一目瞭然（いちもくりょうぜん）に表せることがわかる．

6. おわりに

本論文では信用リスク・スコアリングモデルとして利用されているロジスティック回帰分析の結果を，より平易にかつ深く解釈，利用するための3手法を提示した．筆者はこれらによりこれまで十二分に解釈・活用されていなかった部分に新たに光を当てることができたと考える．

同時に，実際に本手法により分析結果を一般の人にも容易に説明できることも別途確認したことから，同手法がロジスティック回帰分析のさらなる理解促進と普及に貢献できると確信している．

謝辞

本論文執筆にあたっては，同じ組織の八子信行氏，ビュルガー・コンサルティングの野本恵美氏，山下諭史氏にデータ分析の面で多大な協力をいただきました．また，筆者の上司である簑輪圭樹氏と丸谷淳氏，船場正史氏には論文執筆を勧めていただくとともに，貴重なコメントをいただきました．ここにあらためて深謝いたします．

参考文献

[1] 丹後俊郎，山岡和枝，高木晴好：ロジスティック回帰分析，統計ライブラリー,朝倉書店，ISBN4-254-12656-5（1996）．

[2] A. D. Servigny, O. Renault : *Measuring and Managing Credit Risk*, McGraw-Hill ISBN0-07-41755-9（2004）．

[3] 安田隆二，大久保豊：信用リスク・マネジメント革命，きんざい，ISBN4-322-15871-4（1998）．

[4] 古川幸一，木村公昭，圷雅博："信用リスクに関する研究，" 三菱研究所所報，No.40，pp.34-47（2002）．

[5] 石村貞夫，石村園子：SPSSによるリスク解析のための統計処理，東京図書，ISBN4-489-00673-X（2004）．

[6] 久米均，飯塚悦功：回帰分析，シリーズ入門統計的方法2，岩波書店，ISBN4-00-007762-7（1987）．

[7] 古川俊之：寿命の数理，行動計量学シリーズ13，朝倉書店，ISBN4-254-12653-0（1996）．

[8] 山下智志，川口昇，敦賀智裕：信用リスクモデルの評価方法に関する考察と比較，金融研究研修センター・ディスカッション・ペーパー・シリーズ（2003） http://www.fsa.go.jp/frtc/seika/discussion/2003/20031031.pdf.

アイ・ビー・エム ビジネス コンサルティングサービス株式会社
金融事業本部
バンキングコンサルティング
シニア・コンサルタント

高田 直樹 Naoki Takada

［プロフィール］
銀行系，メーカー系シンクタンクや外資系コンサルティング会社を経て，2002年にPwCコンサルティングに入社．PwCコンサルティングとIBMのグローバル統合により，現職．銀行をお客様とする各種コンサルティング・プロジェクトに従事（信用リスク・スコアリングモデル構築や新BIS規制リスクパラメータ推計のほか，業績評価指標設計，業務プロセス設計，PMO支援，など）．
ntakada@jp.ibm.com