

# 最小絶対値回帰分析を利用した 中古車落札金額予測モデルの構築

高野 祐一, 水野 圭, 山菅 和人, 佐藤 齊行, 小川 直哉

## 1. はじめに

本研究では、「平成19年度データ解析コンペティション」において提供された中古車オークションデータを用いて、落札金額の回帰分析を行う。本研究の目的は、オークション主催者がオークション出品者に対して落札金額予測サービスを提供するための、中古車落札金額予測モデルを構築することとする。提供されたデータは、2005年6月から2007年6月までの約24カ月間の車両出品データ（サンプル数125,880）であり、オークションは17会場で860回開催され、入札は一度きりで他者の入札額は知らされず最高額の入札者が落札する第一価格秘密入札方式がとられている。

出品者に対して落札金額予測サービスを提供する、オークション主催者にとっての動機、利点は複数存在する。例えば、出品車両の落札金額を前もって見積もることで出品者の不安を取り除き、より多くの出品を促すことや、精度の高い予測サービスを提供しオークションの落札金額の安定性、妥当性を示すといったことが考えられる。一方で、落札金額予測サービスを提供するからにはユーザーに使ってもらわなくては意味が無く、ユーザーである出品者からの視点も必要となる。予測精度が良くなければならないのは当然としても、例えば、一度でも予測が大きく外れるようなことがあればユーザーは予測サービスの信頼性を疑うこと

となるだろう。また、出品した際の落札金額が予測額より下がるといったことが続くと、毎回高く売れると期待していた出品者の落胆は大きく、予測サービスを使おうという意欲を損なう可能性がある。さらに、このオークションでは相場より安い値段が付けられてしまっていると考え、今後は他のオークションに出品しようと出品者は考えるかもしれない。そして、このようなサービスを提供する上で最も重要であるのが、ユーザーの入力項目が少ないということである。

現在では、インターネットによる通信販売やPOSシステムの充実化によって、大規模データを容易に手に入れることができる。一方で、その規模から分析が難航することが少なくなく、その理由としてデータ処理や計算に時間を要したり、モデルが複雑になりその解釈が難しくなってしまうことが考えられる。特に回帰モデル構築の際には、なるべく説明変数を減らしてシンプルかつ入力項目の少ないモデルを作成したいと考えられるが、大規模データの下では使用する説明変数の選択は困難な問題となる。

そこで本研究ではKonno-Yamamoto[5]を参考に、説明変数の選択を組み込んだ最小絶対値回帰分析を0-1混合整数線形計画問題として定式化し、使用する説明変数の選択を回帰分析と同時に進行。また、様々な目的や条件を加味できる最小絶対値法の利点を活用し、Conditional Value-at-Riskを利用して予測を大きく外すリスクを回避する、非対称な偏差計算式の導入により様々な要求を反映して予測値を計算する、といった工夫を提案する。

本論文の構成は次の通りである。第2節でデータの概要を説明する。第3節では最適変数選択・最小絶対値回帰分析について説明し、第4節では2種類の工夫を提案する。第5節ではサンプルを距離と排気量でセグメント化し、セグメントごとに回帰式を構築することで回帰式の精度を向上させ、その精度を評価する。

たかの ゆういち, さとう なおゆき, おがわ なおや  
筑波大学 システム情報工学研究科  
〒305-8573 つくば市天王台1-1-1  
みずの けい  
日本電気(株) 第一キャリアソリューション事業本部  
〒108-8001 港区芝5-7-1  
やますげ かずと  
東日本電信電話(株) 東京支店総務部  
〒108-8019 港区港南1-9-1  
受付08.7.11 採択08.12.5

第6節では落札金額予測サービスを利用するユーザーの入力項目を削減するために説明変数の選択を行い、第7節では結論と今後の課題について述べる。

## 2. データ概要

今回分析に用いたデータについて説明する。回帰分析に使用した説明変数は次の通りである：

**質的変数 (25種)：**{エアコン, パワーステアリング, パワーウィンドウ, サンルーフ, 革シート, カーナビ, 左ハンドル, Wタイヤ, 冷凍冷蔵装置 (機械式), 冷凍冷蔵装置 (畜令式), 抹消区分, 保証書, 整備手帳, 記録, 修復歴, ヤブレ, 穴, 亀裂}の有無, 会場コード, 形状記号, 燃料, 駆動方式, ドア数, 走行メータ, 車歴.

**量的変数 (7種)：**排気量, 看板面数, 距離, 全長, 全幅, クレーン数, 新車価格.

次は提供されたデータを利用して本研究で独自に作成した説明変数である：

**作成した変数 (9種)：**

- ・ 累計月数 (出品日は2005年6月から何カ月目か, 量的)
- ・ 会場ごと累計月数 (例えば「会場1累計月数」は会場1で出品された車両には累計月数を入力し, それ以外には0を入力した変数, 量的)
- ・ 出品月 (出品日が何月か, 質的)
- ・ 登録月数 (初年度登録と出品日から計算, 量的)
- ・ 車検残存 (出品日における車検残存の有無, 質的)
- ・ 車検残存月数 (出品日と車検期限年月から計算し, 残っていない場合は0, 量的)
- ・ メーカー名 (サンプル数上位9メーカーで9値, その他国内メーカーで1値, 海外メーカーで1値, 計11値, 質的)
- ・ ミッション記号 (段数とミッションでグループ化, 質的)
- ・ 損傷ポイント (車両パネル損傷データから計算, 量的)

質的変数については適宜ダミー変数化を行い, 説明変数の数は133個となる。

本研究では, 車検期限年月, 損傷ポイント, 新車価格のいずれかに欠損のあるサンプルを除去した82,510サンプルを使用した。ただし, パラメータ設定のための実験 (図4) と第6節では計算の都合上, ランダムに抽出した8,251サンプルを使用した。

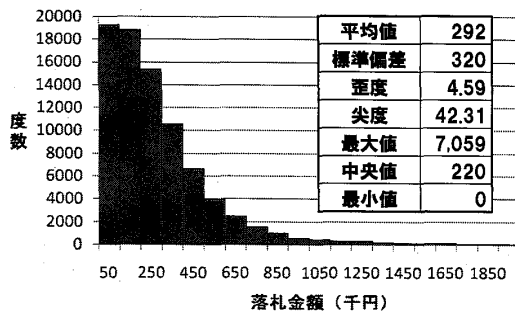


図1 落札金額の度数分布と基本統計量

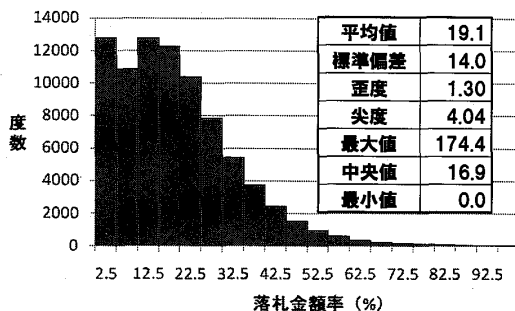


図2 落札金額率の度数分布と基本統計量

被説明変数は, 新車価格の何%の金額で落札されたかを示す落札金額率 (単位: %) を用いた:

$$(\text{落札金額率}) := 100 \times (\text{落札金額}) / (\text{新車価格}).$$

予備的な実験において, 落札金額率の予測値に (新車価格)/100 をかけて落札金額に変換すると, 落札金額を直接予測するよりも精度が良かったため, 本研究では落札金額率を被説明変数とした。なお, 入札の無かった車両 (サンプル数960) は落札金額を0円とした。また, 回帰分析の結果, 予測値が負になってしまう車両には価値が認められず入札がされないと考え, 後処理として予測値を0に変換した。

図1と図2に落札金額と落札金額率の度数分布と基本統計量を示す (度数分布はそれぞれ, 200万円以上, 100%以上の部分については省略した)。落札金額の統計量 (図1) を見ると全サンプルの半数は22万円以下で落札もしくは入札無しとなっている一方で, 700万円もの高額で落札される車両が混在していることが分かる。同様の傾向は落札金額率 (図2) にも見られ, 新車価格の1.7倍で落札された車両が存在する。

## 3. 最適変数選択・最小絶対値回帰分析

回帰分析は, 被説明変数を回帰式と呼ばれる説明変数の式で表すことによって, その間の関係をとらえる統計手法である。説明変数の線形式による回帰式で予

測値を計算し、被説明変数の観測値と予測値との差の2乗和が最小となるように回帰式を構築する最小2乗法が代表的な推計手法である。

しかし、大量のサンプルや説明変数を用いて分析を行う場合には、計算時間や説明変数の組合せ数が増大し、回帰式に使用する説明変数の選択は困難な問題となる。その解決策として、説明変数の選択を組み込んだ回帰分析を0-1混合整数計画問題として定式化し、最適な説明変数集合を回帰式の構築と同時に求める方法がある。また、最小2乗法を用いて定式化すると目的関数は2次関数となるが、被説明変数の観測値と予測値との差の絶対値（絶対偏差）の和を最小化する最小絶対値法を用いると、線形計画問題に帰着でき、効率的に求解できる。本研究では、被説明変数の  $T$  個の観測値  $y_t, t \in \mathcal{T} = \{1, \dots, T\}$  と、 $J$  個の説明変数の  $T$  個の観測値  $x_t^j, j \in \mathcal{G} = \{1, \dots, J\}, t \in \mathcal{T} = \{1, \dots, T\}$  が与えられたとし、 $J$  個の説明変数から  $s$  個を選択する問題を文献[5]を参考に次のように定式化し、最適化ソルバー Xpress-MP (ver. 2007 B) を利用して計算を行った：

$$\begin{aligned} & \text{minimize}_{a, u, v, z} \sum_{t \in \mathcal{T}} \frac{1}{T} (u_t + v_t) \\ & \text{subject to} \quad u_t - v_t = y_t - (a_0 + \sum_{j \in \mathcal{G}} a_j x_t^j), t \in \mathcal{T} \\ & \quad u_t \geq 0, v_t \geq 0, t \in \mathcal{T} \\ & \quad \sum_{j \in \mathcal{G}} z_j = s \\ & \quad 0 \leq a_j \leq M_j z_j, z_j \in \{0, 1\}, j \in \mathcal{G}, \end{aligned} \quad (1)$$

ただし、 $M_j, j \in \mathcal{G}$  は十分大きな正の定数とする。ここで、前処理として各説明変数の被説明変数との相関係数を計算し、相関係数が負の場合には説明変数の符号を反転させておく。被説明変数との相関係数が非負だからといって偏回帰係数が必ず非負になるとはいえないが、整数計画問題は制約式を追加することで問題が解きやすくなる場合が多く、計算時間の短縮のために定式化(1)では偏回帰係数を表す決定変数  $a_j, j \in \mathcal{G}$  を非負変数としている。

定式化の仕組みについて、簡単に説明する。被説明変数の観測値から予測値を引いた残差は、非負決定変数の差  $u_t - v_t$  で表されている。目的関数の最小性により、残差が正（負）の場合は  $v_t = 0$  ( $u_t = 0$ ) となり、 $u_t$  ( $v_t$ ) に残差の絶対値が現れ、目的関数値は平均絶対偏差となる。また、0-1決定変数  $z_j$  が0になると、制約式により偏回帰係数  $a_j$  が0となり、対応する説明変数  $x^j$  は回帰式から削除される。なお、第6節以

外では回帰式の精度と計算効率を優先し、説明変数の選択回数に関する制約式  $\sum_{j \in \mathcal{G}} z_j = s$  は消去しておく。

なお、本研究では説明変数  $x^i, x^j (i > j)$  間の相関係数の絶対値がしきい値パラメータ  $c \in [0, 1]$  より大きい場合に制約式  $z_i + z_j \leq 1$  を定式化(1)に追加し、関連の強い2変数の少なくともどちらか一方を削除することで多重共線性を排除した。この方法では最小絶対値基準の下で、分析者の主観や恣意性が入ることなく多重共線性を排除することができる。しきい値パラメータ  $c$  の値を変えて実験を行い、その結果から、本研究では回帰式の精度を悪化させないよう余裕を持って  $c = 0.5$  と設定した。

#### 4. 回帰式の構築と工夫

最小絶対値法は、正規方程式を解いて計算を行う最小2乗法とは対照的に、線形計画問題として定式化することで、様々な目的や条件を加味できる利点がある[10]。最小2乗法も2次計画問題として定式化することは可能であるが、大規模な問題や整数変数を導入した問題を解くことを考えると線形計画問題として定式化可能な最小絶対値法に明らかな優位性がある。本節ではこの利点を活用し2種類の工夫を提案する。

##### 4.1 CVaR を用いたリスク回避

多重共線性を排除する制約式を追加した定式化(1)で落札金額率の最小絶対値回帰分析を行った結果、921サンプル（全サンプル中の約1.1%）で絶対偏差が25%以上となった。例えば新車価格200万円の車両が70万円で落札される場合に、120万円、または20万円と予測すると絶対偏差が25%となる。この例からも分かるように、25%以上の絶対偏差というのは非常に大きく、予測サービスの信頼を失うような事態も起こりかねない。こうしたリスクを避けるために、本研究ではConditional Value-at-Risk[7]（以下、CVaR）を利用する。

**CVaRの定義と回帰分析への応用** CVaRは、損失を

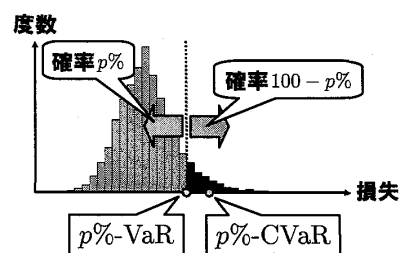


図3 VaRとCVaRのイメージ

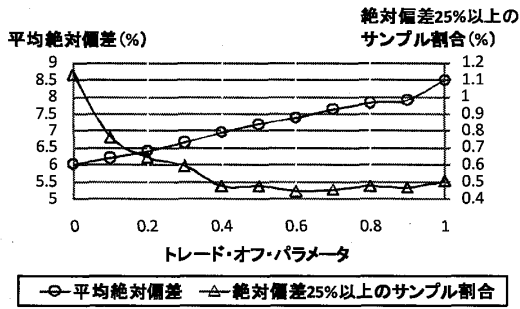


図4 トレード・オフ・パラメータλと回帰の結果

表す確率変数  $\tilde{X}$  を入力とするリスク尺度であり、Value-at-Risk (以下, VaR) 以上という条件付きの損失の期待値として説明できる。なお本研究では、残差の正負に関係なく予測値が観測値から大きく外れることが予測サービスの信頼性を損なうリスクであると考え、回帰式の絶対偏差を用いて損失を定義している。

まず VaR を定義しよう。信頼水準を表すパラメータ  $p \in [0, 100]$  に対して (本研究では,  $p=99$  とした),  $p\%$ -VaR は次の式で定義される:

$$\min_{\eta \in \mathbb{R}} \{\eta | \Psi_{\tilde{X}}(\eta) \geq p/100\}, \quad (2)$$

ただし,  $\Psi_{\tilde{X}}$  は損失  $\tilde{X}$  の分布関数, すなわち  $\Psi_{\tilde{X}}(\eta) = \text{Prob}(\tilde{X} \leq \eta)$  である。確率変数  $\tilde{X}$  が連続分布に従う場合には,  $p\%$ -VaR は  $\tilde{X}$  の  $p\%$ 分位点と一致する。この  $p\%$ -VaR を超えるような損失の期待値が  $p\%$ -CVaR であり, 確率変数が有限個の実現値を持つ離散分布に従う場合,  $p\%$ -CVaR は  $p\%$ -VaR を表す決定変数  $\eta \in \mathbb{R}$  と,  $(\max\{X_t - \eta, 0\})$  を表す決定変数  $w \in \mathbb{R}^T$  を用いた次の定式化で計算できる [7]:

$$\begin{cases} \text{minimize} & \eta + \frac{1}{1-p/100} \sum_{t \in \mathcal{J}} q_t w_t \\ \text{subject to} & w_t \geq X_t - \eta, w_t \geq 0, t \in \mathcal{J}, \end{cases} \quad (3)$$

ただし,  $X_t, t \in \mathcal{J} = \{1, \dots, T\}$  は確率変数  $\tilde{X}$  の実現値で, 各生起確率は  $q_t, t \in \mathcal{J}$  であるとする。本研究では, 定式化(1)において,  $X_t := u_t + v_t, q_t := 1/T, t \in \mathcal{J}$  とすることで, 損失を絶対偏差で定義した CVaR を計算している。CVaR は計算面, 理論面の両面から好ましいリスク尺度であり [6], また, 回帰分析に CVaR を導入した研究として [8] などがあるが, そこでは変数選択は考慮されていない。

本研究では, 絶対偏差の平均値と 99%-CVaR のトレード・オフ・パラメータ  $\lambda \in [0, 1]$  を導入した次の式(4)を目的関数とし, 最小化することで両方を同時に考慮する, 絶対偏差の平均-CVaR モデルを用いる:

$$(1-\lambda) \times (\text{平均絶対偏差}) + \lambda \times (99\text{-CVaR}). \quad (4)$$

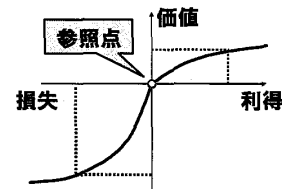


図5 プロスペクト理論における価値関数

本研究では, トレード・オフ・パラメータ  $\lambda=0.2$  と設定した。平均絶対偏差の悪化は小さく, 絶対偏差が 25%以上となるサンプルの割合は大きく減少していることが図4 (ただしサンプル数 8,251 で計算) から分かる。

#### 4.2 非対称な偏差計算式の導入

本節では, 予測値より観測値が大きい場合の偏差  $u_t$  と, 予測値より観測値が小さい場合の偏差  $v_t$  の重みに差を付けることが可能な, 次の非対称な偏差計算式を導入し, 式(4)の平均絶対偏差として使用する:

$$\sum_{t \in \mathcal{J}} \frac{1}{T} (\theta u_t + (2-\theta)v_t), \quad (5)$$

ただし,  $\theta \in (0, 2)$  は偏差の相対的重要度を表すパラメータである。  $\theta=1$  に設定すると式(5)は元の平均絶対偏差の計算式になる。また, 式(5)の偏差計算による最小絶対値回帰は, Koenker-Bassett [4] によって提案された分位点回帰と見なすこともでき, 落札金額の  $100 \times \theta/2\%$ 分位点回帰に対応する。

例えば  $\theta$  を 1 より大きく設定すれば予測落札金額が上がり, 予測額を見た出品者の出品意欲を高められる可能性がある。また, 出品者は落札金額に満足できなければ成約せずに流札を選択することも可能だが, オークション主催者が流札を減らしたいと考えている場合には  $\theta$  を 1 より小さく設定すれば, 出品者に示す予測落札金額が下がることで出品者の期待値が下がり, 流札を減らす効果が期待できる。また,  $\theta$  を 1 より小さく設定する根拠の一つとして, Kahneman-Tversky [3] の提唱したプロスペクト理論がある。

プロスペクト理論では, 人々の評価は図5のような価値関数によって説明され, その特徴として, 「価値評価はある期待 (参照点) からの乖離によって行われる」, 「参照点からの乖離が同程度なら, 利得に対する満足感より損失に対する拒否感の方が大きい」の2点が挙げられる [11]。落札金額の予測を行い, それを出品者に示すことによって, 出品者は落札金額の予測値, すなわち参照点を持つようになる。プロスペクト理論を考慮すると, 予測額より実際の落札金額が下がった

場合の拒否感は、実際の落札金額が上がった場合の満足感よりかなり大きいと考えられる。よって、「実際に出品すると落札金額が予測額より下がった」といった出品者からの苦情にオークション主催者が頭を悩ませている場合などは、 $\theta$  を1より小さく設定し、参照点に対して非対称な偏差計算式を採用することが図5の価値関数の形からも自然な方法であると考えられる。ただし、このような非対称な偏差計算式を導入することで予測精度を大きく悪化させる可能性があり、注意する必要がある。本研究では、 $\theta=2/3$ と設定し、予測値より観測値が小さい場合の偏差には相対的に2倍の重みをかけることにした。

## 5. セグメント化と回帰式の精度評価

本節では距離と排気量を用いてサンプルをセグメント化し、セグメントごとに回帰式を作ることで回帰式の精度を向上させる。距離はおおよそ4分位点となる値で分類し、排気量は税金の基準を参考に6種類に分類した。最小絶対値法は、多重共線性を排除する制約式を加えた定式化(1)による「モデル ABS」と、モデル ABS の目的関数を式(4)とし、式(5)の非対称偏差計算( $\theta=2/3$ )を導入した定式化による「モデル CVaR」の2種類のモデルを考える。MATLAB 7.6.0 (R 2008 a) の Statistics Toolbox を利用した最小2乗法と合わせて、 $4 \times 6 = 24$ 種類のセグメントごとに計算結果を表1に示す。なお、最小2乗法では相関係数の絶対値が0.5より大きい変数の組に対し、被説明変数との相関が弱い方の説明変数を削除し多重共線性を排除した。

表1では回帰分析の評価基準として、予測誤差の大きさを示す「(b)平均絶対偏差」、モデルの当てはめの良さを示す「(c)決定係数」、(d)予測値より観測値が小さくなるサンプル割合、そして、予測を大きく外したサンプルの割合を示す「(e)絶対偏差25%以上のサンプル割合」、の4種類を考える。プロスペクト理論を考慮すると、予測額より実際の落札金額が下がると予測サービスを利用した出品者が大きな不満を持つ可能性がある。よって、「(d)予測値より観測値が小さくなるサンプル割合」は、なるべく小さくしたいという要求が考えられ、ここでは評価基準として用いている。落札金額率の予測(表1上段) 第3節で説明したように説明変数と被説明変数の相関係数を見て偏回帰係数の符号を限定しているために、最小絶対値法は最小2乗法に比べて少し不利になっている。それでも表1

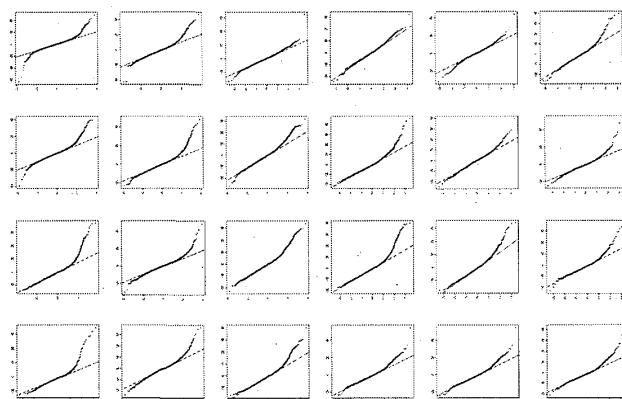


図6 セグメントごとの残差の正規 Q-Qプロット (セグメントの並び順は表1と同様)

上段で「(b)平均絶対偏差」を比較するとモデル ABS は最小2乗法を上回る結果を示している。なお、全セグメントのサンプルにおける絶対偏差の平均値(単位:%)は、モデル ABS が4.54、モデル CVaR が4.87、最小2乗法が4.55となっている。しかし、「(c)決定係数」を比較すると最小2乗法が最も良い結果を得ており、予測精度においてはモデル ABS と最小2乗法は優劣つけ難い。一方でモデル CVaR は工夫を加えることで回帰式の精度を犠牲にしている部分があり、精度では他のモデルに及ばない。しかし、「(d)予測値より観測値が小さくなるサンプル割合」を見ると非対称な偏差計算の効果によってすべてのセグメントで最も小さい値となり、最小2乗法と比較すれば予測値より観測値が小さくなるサンプルを10%以上減少させている。また、「(e)絶対偏差25%以上のサンプル割合」でも最も良い結果を得るセグメントが多く、全セグメントのサンプルにおける絶対偏差が25%以上となるサンプル割合(単位:%)は、モデル ABS が0.69、モデル CVaR が0.38、最小2乗法が0.45となり、CVaR 導入の効果が確認できる。さらに、モデル CVaR はトレード・オフ・パラメータ $\lambda$ を変更することで平均絶対偏差と CVaR の優先度を調整することが可能であり、回帰式の精度を良くしたいと考えればパラメータ $\lambda$ をより小さく、絶対偏差が大きくなるサンプルを少なくしたいと考えれば、パラメータ $\lambda$ をより大きく設定することで目的を達成できる。

ここで、セグメントごとの「(b)平均絶対偏差」に着目すると、走行距離が5万 km 以下の車両は比較的回帰式の精度が悪い、排気量660 cc以下の車両すなわち軽自動車の回帰式は比較的精度が良く、対照的に排気量が2,001 ccから3,000 ccの車両すなわち大型乗

表1 セグメントごとの回帰分析の結果

(a) サンプル数 (b) 平均絶対偏差 (c) 決定係数 (d) 予測値より観測値が小さくなるサンプル割合 (%) (e) 絶対偏差 25%以上のサンプル割合 (%) ただし、5分割交差検証法においては、5回の検証の平均値を示す。

		落札金額率予測(単位:%)																	
		排気量(cc)																	
距離 (km)		-660			661-1,500			1,501-2,000			2,001-2,500			2,501-3,000			3,001-		
		最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小		
		ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗
	(a)	6,646			3,882			2,228			621			453			571		
-5万	(b)	4.70	5.00	4.59	4.97	5.48	5.01	6.07	6.64	6.08	5.46	5.82	5.59	6.19	7.52	6.23	6.40	7.82	6.70
	(c)	0.88	0.87	0.88	0.82	0.81	0.83	0.80	0.79	0.81	0.83	0.84	0.85	0.73	0.70	0.76	0.73	0.68	0.76
	(d)	49.64	38.32	51.29	49.25	43.07	53.97	48.61	41.38	51.30	44.93	37.68	52.33	43.93	33.77	49.89	45.36	42.21	53.24
	(e)	0.60	0.50	0.63	1.06	0.49	0.77	0.72	0.40	0.54	0.64	0.16	0.16	1.77	0.00	1.10	3.33	5.25	2.28
	(a)	8,711			10,822			5,626			1,721			1,156			1,189		
5万-10万	(b)	4.26	4.53	4.36	3.99	4.27	3.98	4.72	5.01	4.81	5.60	5.94	5.41	5.17	5.65	5.28	4.98	5.84	5.14
	(c)	0.73	0.72	0.74	0.78	0.76	0.78	0.71	0.69	0.72	0.63	0.62	0.68	0.74	0.72	0.75	0.64	0.61	0.67
	(d)	48.94	42.66	53.52	49.69	43.47	53.63	49.40	44.28	54.35	48.34	42.94	56.77	47.49	39.79	51.12	47.69	43.15	55.09
	(e)	0.63	0.36	0.44	0.40	0.32	0.34	0.46	0.25	0.21	1.34	0.58	0.76	0.69	0.17	0.35	1.43	5.00	0.84
	(a)	3,197			9,921			5,813			2,126			1,268			1,105		
10万-15万	(b)	3.60	3.99	3.69	4.12	4.33	4.01	4.56	4.76	4.60	5.44	5.75	5.23	5.25	5.99	5.37	4.50	5.00	4.71
	(c)	0.56	0.53	0.58	0.76	0.75	0.78	0.69	0.68	0.70	0.53	0.51	0.60	0.65	0.61	0.68	0.64	0.60	0.65
	(d)	48.95	43.45	54.77	49.72	41.87	51.48	49.44	45.19	54.02	48.49	44.73	55.50	47.71	43.45	54.18	46.79	41.00	55.66
	(e)	0.81	0.19	0.59	0.16	0.17	0.15	0.48	0.38	0.26	1.93	1.79	1.22	1.03	0.16	0.71	1.09	0.27	0.54
	(a)	776			5,030			4,040			2,496			1,392			1,720		
15万-	(b)	3.55	4.15	3.82	4.04	4.37	4.00	4.52	4.76	4.54	5.53	5.77	5.61	5.67	6.02	5.79	4.21	4.49	4.37
	(c)	0.38	0.43	0.49	0.75	0.72	0.76	0.63	0.61	0.66	0.53	0.53	0.56	0.57	0.58	0.61	0.63	0.64	0.65
	(d)	49.41	42.53	58.63	49.30	42.90	51.47	49.11	45.40	53.81	48.84	45.23	55.45	47.99	41.59	55.10	48.08	41.74	56.86
	(e)	2.06	0.00	0.90	0.24	0.10	0.18	0.89	0.30	0.22	1.76	0.56	1.24	1.51	0.14	0.50	0.70	0.17	0.23
			落札金額率の予測値を落札金額に変換することによる落札金額予測(単位:千円)																
		排気量(cc)																	
距離 (km)		-660			661-1,500			1,501-2,000			2,001-2,500			2,501-3,000			3,001-		
		最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小		
		ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗
	(b)	39.0	41.5	38.1	65.4	71.2	65.8	114.0	124.8	115.9	138.1	144.6	140.3	178.0	208.9	176.4	271.6	345.0	275.1
-5万	(c)	0.90	0.90	0.91	0.84	0.84	0.85	0.86	0.85	0.86	0.88	0.88	0.89	0.83	0.83	0.86	0.87	0.82	0.89
	(d)	35.9	38.4	37.0	50.9	54.2	51.1	83.5	88.7	85.5	111.3	118.3	108.8	135.6	148.1	137.1	180.5	211.0	184.3
	(e)	0.76	0.75	0.76	0.79	0.78	0.80	0.74	0.73	0.75	0.76	0.75	0.78	0.77	0.74	0.78	0.84	0.82	0.86
	(a)	30.3			33.8			31.1			52.9			55.1			51.7		
	(b)	30.3	33.8	31.1	52.9	55.1	51.7	79.1	82.3	79.7	93.8	99.8	91.9	117.2	135.5	120.4	148.2	164.0	154.6
5万-10万	(c)	0.60	0.55	0.61	0.77	0.77	0.79	0.72	0.71	0.74	0.62	0.59	0.66	0.69	0.65	0.71	0.81	0.78	0.82
	(d)	28.8	34.0	30.9	52.5	56.3	51.8	77.6	82.3	78.3	86.5	90.2	88.0	109.6	118.6	113.2	127.7	134.5	132.1
	(e)	0.38	0.37	0.46	0.77	0.76	0.78	0.67	0.65	0.69	0.55	0.55	0.58	0.54	0.53	0.57	0.78	0.79	0.79
	(a)	30.3			33.8			31.1			52.9			55.1			51.7		
	(b)	30.3	33.8	31.1	52.9	55.1	51.7	79.1	82.3	79.7	93.8	99.8	91.9	117.2	135.5	120.4	148.2	164.0	154.6
		落札金額率予測の5分割交差検証法(単位:%)																	
		排気量(cc)																	
距離 (km)		-660			661-1,500			1,501-2,000			2,001-2,500			2,501-3,000			3,001-		
		最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小	最小絶対値	最小		
		ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗	ABS	CVaR	2乗
	(b)	4.78	5.09	4.64	5.10	5.60	5.12	6.40	6.90	6.42	6.84	7.05	6.79	7.53	8.86	7.59	7.59	9.32	7.98
-5万	(c)	0.87	0.87	0.88	0.81	0.79	0.82	0.79	0.78	0.80	0.82	0.81	0.83	0.72	0.68	0.74	0.71	0.66	0.74
	(d)	50.64	39.54	51.78	50.89	44.66	53.85	50.06	42.91	51.76	51.37	40.45	52.68	49.51	40.12	48.51	49.42	46.50	54.63
	(e)	0.65	0.63	0.65	0.85	0.65	0.80	0.76	0.94	0.67	0.94	1.88	1.42	1.93	3.98	2.38	3.99	6.73	3.88
	(a)	4.33			4.61			4.41			4.03			4.31			4.02		
	(b)	4.33	4.61	4.41	4.03	4.31	4.02	4.83	5.10	4.87	5.92	6.31	5.73	5.72	6.20	5.76	5.39	6.47	5.55
5万-10万	(c)	0.71	0.70	0.72	0.75	0.74	0.76	0.68	0.66	0.69	0.61	0.58	0.66	0.72	0.69	0.73	0.64	0.58	0.66
	(d)	52.06	44.57	54.93	51.11	44.32	54.28	51.29	45.38	54.68	51.89	46.08	56.68	49.39	41.16	51.43	51.20	43.83	55.50
	(e)	0.61	0.46	0.51	0.37	0.42	0.36	0.44	0.45	0.23	1.26	1.22	0.99	0.61	0.70	0.72	1.50	0.98	0.91
	(a)	3.72			4.10			3.80			4.16			4.36			4.05		
	(b)	3.72	4.10	3.80	4.16	4.36	4.05	4.70	4.90	4.71	5.80	6.05	5.52	5.66	6.32	5.82	5.03	5.11	5.15
10万-15万	(c)	0.53	0.50	0.54	0.74	0.73	0.76	0.65	0.64	0.67	0.51	0.49	0.57	0.63	0.60	0.65	0.62	0.58	0.64
	(d)	53.27	47.75	57.25	50.85	42.66	52.19	50.97	46.40	54.79	51.87	46.86	55.71	50.35	44.69	53.70	51.31	42.90	56.81
	(e)	0.72	0.22	0.66	0.17	0.17	0.15	0.43	0.58	0.38	1.80	1.84	1.60	0.99	1.15	1.00	0.99	1.16	0.81
	(a)	3.88			4.83			4.30			4.16			4.49			4.06		
	(b)	3.88	4.83	4.30	4.16	4.49	4.06	4.64	4.84	4.66	5.73	6.17	5.80	6.11	6.43	6.16	4.52	4.87	4.62
15万-	(c)	0.38	0.38	0.43	0.72	0.70	0.73	0.59	0.58	0.62	0.49	0.47	0.53	0.55	0.56	0.60	0.62	0.61	0.64
	(d)	53.18	49.75	61.27	51.25	43.41	52.32	51.15	46.99	55.33	51.62	47.79	56.26	51.24	44.26	54.93	52.24	44.89	56.88
	(e)	1.75	0.77	1.36	0.26	0.16	0.22	0.68	0.49	0.40	1.75	0.84	1.23	1.43	1.22	0.93	0.53	0.35	0.29
	(a)	3.88			4.83			4.30			4.16			4.49			4.06		
	(b)	3.88	4.83	4.30	4.16	4.49	4.06	4.64	4.84	4.66	5.73	6.17	5.80	6.11	6.43	6.16	4.52	4.87	4.62

用車と呼ばれるような車両に関しては、比較的精度が悪いといった傾向が見られる。これらは、「長い距離を走っている車両の落札金額は安いに違はなく予測は簡単だろう」、「大型乗用車は値段が高い物から安い物まで多くの種類があるために予測は難しいだろう」といった直感に近い結果といえる。

残差の正規性検定(図6) 最小絶対値法は、残差がLaplace分布に従う場合に偏回帰係数が最尤推定値になる[1]など、残差が正規分布に従わない(特に外れ値を含む)場合には最小2乗法より信頼度の高い偏回帰係数推定値を与えるという利点がある[10]。そこで、

セグメントごとの正規Q-Qプロットを行い、最小2乗法による落札金額率予測の残差の正規性を調べた(図6)。残差が正規分布に従う場合、プロットは直線状になることが期待されるが、図6を見ると、プロットの右端で対照直線からの系統的なずれが見られるセグメントが多い。よって、今回のデータでは残差が正規分布に従うとはみなせず、偏回帰係数の信頼度の点では最小絶対値法が有効であると考えられる。また、この系統的なずれは第2節で観察したような高額で落札されている車両の混在と関連する。

落札金額率の予測における説明変数(表3) 表3は、

表2 各モデルの結果比較 (表1の結果まとめ)

落札金額率予測				5分割交差検証法			
最小絶対値		最小2乗		最小絶対値		最小2乗	
ABS	CVaR			ABS	CVaR		
(b)	18	0	6	(b)	17	0	7
(c)	2	0	24	(c)	0	0	24
(d)	0	24	0	(d)	0	24	0
(e)	0	18	7	(e)	5	7	12

落札金額率予測			
最小絶対値		最小2乗	
ABS	CVaR		
(b)	18	0	6
(c)	2	0	24

表1の評価基準において、各モデルが最良値を得るセグメント数をカウントした。

落札金額率予測において、24セグメント中8以上のセグメントで有意(最小2乗法による回帰分析でP値が0.05以下)となった説明変数41個について、有意となったセグメント数、各モデルで偏回帰係数が正もしくは負になったセグメント数を示している。例えば、累計月数は最小2乗法による回帰分析で24セグメント中15セグメントでP値が0.05以下となり、かつモデルABSでは17のセグメントで偏回帰係数が正、一つのセグメントで偏回帰係数が負となり、残りの6セグメントでは多重共線性の排除に関連して削除されたかもしくは偏回帰係数が0となったことが読み取れる。「落札金額率は3月と9月に上がり、1月と12月に下がる傾向がある」、「ワゴン、バス、トラックといった大型車は落札金額率が高くなる傾向がある」といった全モデル共通の興味深い傾向が見られるが、一方で顕著に異なった傾向を見せる変数が、エアコンとパワーステアリングである。両変数とも落札金額率に対して負の影響を与えると直感的には考え難い。しかし、最小絶対値法とは対照的に最小2乗法では偏回帰係数が負になるセグメントが多い。変数の有意性を考慮しても、最小2乗法では、エアコンは4セグメント、パワーステアリングは7セグメントで有意となりながら偏回帰係数が負になっている。

**落札金額の予測 (表1中段)** 落札金額率の予測値に(新車価格)/100をかけて落札金額の予測とみなした結果を表1中段で示す。落札金額率の予測と同様に、各モデルの「(b)平均絶対偏差」を比較するとモデルABSが最も良い結果となるセグメントが多く、「(c)決定係数」を比較すると最小2乗法が最も良い結果を得ている。また、セグメントごとの「(b)平均絶対偏差」を比較すると排気量が高くなるほど単調に値が大きくなっていくことが分かる。これは、落札金額率の予測値に新車価格をかけて落札金額の予測値としているために、落札金額率における絶対偏差は同程度でも、排

表3 説明変数の有意性と偏回帰係数の符号

説明変数名	有意	最小絶対値		最小2乗			
		CVaR		ABS			
		正	負	正	負		
累計月数	15	17	1	17	1	23	1
出品:1月	8	3	16	5	15	6	18
出品:2月	9	16	2	14	3	17	5
出品:3月	10	19	1	16	1	15	7
出品:9月	10	11	5	12	5	14	9
出品:12月	9	3	17	3	16	3	19
登録月数	24	0	23	0	24	0	24
車検残存月数	8	8	1	7	0	11	0
メーカー:トヨタ	19	15	0	16	0	20	0
メーカー:ニッサン	8	2	10	2	9	4	7
メーカー:ミツビシ	12	1	16	2	14	7	14
メーカー:マツダ	12	3	16	2	14	5	17
メーカー:ホンダ	13	11	8	9	6	11	8
メーカー:ダイハツ	8	9	3	9	4	13	5
メーカー:スズキ	9	9	2	9	2	14	2
メーカー:スバル	11	2	9	2	11	7	10
メーカー:海外	10	1	15	1	14	3	14
ミッション:4速オート	9	1	5	1	5	5	6
ミッション:5速オート	9	5	2	4	3	9	7
ミッション:5速マニュアル	12	3	1	4	0	14	0
駆動方式:2WD	15	5	11	5	9	9	13
ドア数:3	11	5	13	5	12	11	11
看板面数	9	5	10	5	7	13	11
走行メータ:通常	8	13	0	12	0	16	1
距離	22	0	23	0	21	0	24
全長	9	6	0	8	1	7	3
全幅	11	11	0	8	0	12	0
エアコン	8	10	1	10	2	10	13
パワーステアリング	8	7	0	7	0	5	11
パワーウィンドウ	12	11	2	6	4	11	7
カーナビ	11	15	2	12	2	14	3
冷凍冷蔵装置(機械式)	15	17	1	18	1	20	2
修復歴	15	0	21	0	22	1	23
損傷ポイント	18	0	22	0	21	0	24
形状:ボンネットバン	17	4	13	4	10	5	15
形状:キャブバン	14	4	9	5	8	7	12
形状:ボンネットワゴン	14	19	0	18	0	14	6
形状:キャブワゴン	14	12	3	12	7	18	6
形状:ハードトップ	10	3	19	3	17	4	19
形状:バス	12	13	0	12	0	12	1
形状:ボンネットトラック	15	14	0	14	0	15	0

気量が高くなるほど新車価格は上がり、落札金額における絶対偏差の値が大きくなるためと考えられる。また、全セグメントのサンプルにおける絶対偏差の平均値(単位:千円)は、モデルABSが69.9、モデルCVaRが75.4、最小2乗法が70.4となる。数百万円で落札される車両があることを考慮すれば、平均絶対偏差が7万円を切るモデルABSの結果はモデルの有効性を示すものと認められる。さらに、全セグメントのサンプルにおいて、観測値と予測値の相関係数を2乗し決定係数に対応する値を計算すると、モデルABSが0.86、モデルCVaRが0.84、最小2乗法が0.87となり、非常に説明力の高いモデルが作れていることも分かる。

**落札金額率予測の5分割交差検証法 (表1下段)** ここまでは全サンプルを利用して回帰式を構築し、全サンプルを利用して回帰式の精度評価を行ってきたが、ここでは予測モデルとしての性能を評価するために5分割交差検証法を行う。まず、サンプルを5グループに分割し、その中の4グループを学習用サンプルとしてセグメントごとに回帰式を構築し、回帰式の構築に利用しなかった1グループを検証用サンプルとして回

帰式の精度評価を行うということを5通りの組合せで行った(表1下段)。全セグメントのサンプルの平均絶対偏差の、5回の検証の平均値(単位:%)は、モデルABSが4.71、モデルCVaRが5.07、最小2乗法が4.70であり、交差検証を行うことでサンプル数が少ないセグメントでは回帰式の精度が悪化する場合もあるが、全体として平均絶対偏差の悪化の度合いは小さく、どのモデルも学習用サンプルに当てはまり過ぎて検証用サンプルでは精度が悪くなるといった過学習を生じていないことが確認できる。この結果から予測モデルとしての有効性についても実証できたといえる。さらに、「(d)予測値より観測値が小さくなるサンプル割合の、5回の検証の平均値」を見ると、モデルCVaRは5分割交差検証法においても全セグメントにおいて最も小さい値を得ていることが確認できる。一方で、「(e)絶対偏差25%以上のサンプル割合の、5回の検証の平均値」を見ると、モデルCVaRが最も悪い結果となるセグメントが多い。これは、(絶対偏差が大きい)1%のサンプルの絶対偏差の期待値である99%-CVaRの性質に起因するもので、サンプル数が少ないセグメントではCVaRが考慮する学習用サンプル数が極端に少なくなってしまう、検証用サンプルでの性能が悪くなっていると考えられる。サンプル数が十分に多いセグメントでは効果が得られていることを考慮すると、十分なサンプル数があれば99%-CVaRを利用した予測モデルがその性能を発揮できると考えられる。また、どうしてもサンプル数が少なくなってしまう場合には信頼水準 $\alpha$ をより小さい値(0.9,0.7など)に設定して考慮するサンプル数を増やす、サンプル数が少なくなってしまうセグメントではCVaRを考慮しないモデルを利用する(モデルABSに非対称偏差計算を導入する)などの対応策が必要となる。

## 6. 変数選択による入力項目の削減

落札金額予測サービスを提供する上での問題点として、例えどれだけ精度の良い予測でも、入力項目が多過ぎるとユーザーが入力を面倒に感じて予測サービスの利用をためらうといったことが考えられる。ここまでの分析では定式化(1)の、説明変数の選択個数に関する制約式 $\sum_{j \in S} z_j = s$ を消去していたが、本節ではこの制約式を利用して入力項目を絞り込む。

予測サービスでは説明変数の値が入力となるが、多値の質的変数については入力を選択式にすることで簡

表4 変数選択を行った落札金額率予測の結果  
質的変数4種以外の説明変数84個から26, 23, 15, 10個を選択。

	選択された説明変数の数(個)			
	26	23	15	10
最適変数選択・最小絶対値法(ABS)				
平均絶対偏差(%)	6.01	6.01	6.01	6.05
決定係数	0.66	0.66	0.66	0.66
ステップワイズ変数選択・最小2乗法				
投入基準P値	0.05	5.0E-03	5.0E-04	5.0E-09
除去基準P値	0.10	1.0E-02	1.0E-03	1.0E-08
平均絶対偏差(%)	6.02	6.02	6.05	6.11
決定係数	0.68	0.68	0.67	0.67

単に入力できる。また、複数個のダミー変数として回帰式に使用されているため、質的変数1個の入力で複数個のダミー変数の値を確定でき効率が良い。よって、表3の結果などを参考に、質的変数である出品月、メーカー名、ミッション記号、形状記号は多重共線性に関連して削除される以外はすべて回帰式に使用することとし、その他の説明変数(84個)から使用する変数を選択することとする。また、最小2乗法においては説明変数のP値を基準に変数の投入と除去を繰り返すステップワイズ法を利用して変数選択を行った。

ステップワイズ変数選択を用いた最小2乗法を行い、そこで選択された説明変数の数に合わせて最適変数選択を用いた最小絶対値法(ABS)を行った結果を表4(ただし、サンプル数8,251で計算)に示す。なお、ここではセグメント化は行わない。選択される説明変数の数が少なくなるにつれて、最小絶対値法と最小2乗法の平均絶対偏差の差は大きくなる傾向があり、説明変数の選択方法として定式化(1)による最適変数選択が有効であることが分かる。最小2乗法は2次計画問題として定式化されるために、変数選択を組み込んだ定式化を解くことは難しい。よって、この結果は最小絶対値法の優位性を示すものと見ることもでき、質的変数4個と選択された説明変数10個の計14個の入力項目で、平均絶対偏差6.05%の落札金額(率)予測サービスを行うことが可能である。

最後に、選択された説明変数とその偏回帰係数の符号を表5で示す。両手法で選択された説明変数を比較すると、変数の数が少なくなるにつれてほぼ同様の変数が選択されていくことが分かる。両手法で常に選択される説明変数としては例えば累計月数があり、偏回帰係数の値は常に正になる。このことから本オークションの落札金額率は上昇傾向にあることが読み取れる。一方で、顕著な違いを表す例としては、走行メータが挙げられる。ほぼすべてのサンプルで走行メータは通常、タコグラフ一体、交換のいずれかとなっているが、



最適変数選択では走行メータが交換されているかどうかだけをモデルに取り入れている。また、最小2乗法ではパワーステアリングの偏回帰係数が負になる場合があるが、パワーステアリングが装備されていることで落札金額が下がるということは直感的に考え難いため、これは望ましい結果とはいえない。

## 7. おわりに

本研究では、最小絶対値回帰分析を基に落札金額率の回帰式を構築した。最小絶対値法には偏回帰係数の信頼度といった統計的性質の良さだけでなく、線形計画問題として定式化することで、様々な目的や条件を加味できる利点がある。本研究ではその利点を活用し、落札金額予測サービスを提供するためのモデルであることを考慮して、CVaRを用いたリスク回避、非対称な偏差計算式の導入といった工夫を提案した。さらにサンプルをセグメント化し、セグメントごとに回帰式を構築することで回帰式の精度を改善した。最後に、落札金額予測サービスにおいてはユーザーの入力項目を削減することが重要であると考え、最適変数選択を

表5 選択された説明変数の偏回帰係数の符号（+は偏回帰係数が正，-は負であることを表す）とP値とt値（ステップワイズ変数選択で変数26個の場合の値）

説明変数名	最適変数選択				ステップワイズ変数選択				P値	t値
	26	23	15	10	26	23	15	10		
会場:1	-	-	-	-	-	-	-	-		
会場:2	-	-	-	-	-	-	-	-		
会場:5	-	-	-	-	-	-	-	-	0.014	-2.45
会場:6	+	+	-	-	-	-	-	-		
会場:14	+	+	-	-	-	-	-	-		
累計月数	+	+	+	+	+	+	+	+	0.000	14.14
会場5 累計月数	-	-	-	-	-	-	-	-		
会場8 累計月数	-	-	-	-	-	-	-	-	0.002	-3.13
会場14 累計月数	-	-	-	-	+	-	-	-	0.033	2.14
会場39 累計月数	+	+	-	-	-	-	-	-		
登録月数	-	-	-	-	-	-	-	-	0.000	-73.90
車検残存月数	+	+	+	+	+	+	+	+	0.000	17.02
燃料:軽油	-	-	-	-	+	+	+	+	0.000	7.84
駆動方式:2WD	+	+	+	+	+	+	+	+	0.000	5.62
排気量	+	+	-	-	-	-	-	-		
ドア数:3	-	-	-	-	-	-	-	-	0.000	-14.39
ドア数:5	+	-	-	-	-	-	-	-		
看板面数	-	-	-	-	-	-	-	-		
走行メータ:通常	-	-	-	-	+	+	-	-	0.000	4.44
【走行メータ: タコグラフ一体】	-	-	-	-	+	+	-	-	0.002	3.03
走行メータ:交換	-	-	-	-	-	-	-	-		
距離	-	-	-	-	-	-	-	-	0.000	-34.70
パワーステアリング	-	-	-	-	-	-	-	-	0.004	-2.86
パワーウィンドウ	+	+	+	+	+	+	+	+	0.000	7.53
サンルーフ	+	-	-	-	+	+	+	+	0.004	2.85
革シート	-	-	-	-	+	+	+	+	0.000	3.53
カーナビ	+	+	+	+	+	+	+	+	0.000	5.66
Wタイヤ	-	-	-	-	-	-	-	-	0.000	-5.09
【冷凍冷蔵装置 (機械式)】	+	+	+	+	+	+	+	+	0.000	11.97
クレーン数	+	+	+	+	+	+	+	+	0.000	5.74
車歴:一般・自家用	+	+	+	+	-	-	-	-		
車歴:リース・自家用	-	-	-	-	-	-	-	-	0.037	-2.09
記録	-	-	-	-	-	-	-	-	0.000	-6.06
修復歴	-	-	-	-	-	-	-	-	0.000	-7.21
ヤブレ	-	-	-	-	-	-	-	-	0.000	-3.67
損傷ポイント	-	-	-	-	-	-	-	-	0.000	-10.34
新車価格	-	-	-	-	-	-	-	-	0.000	-4.20

利用することで入力項目を削減し、ヒューリスティックな選択方法と比較して有効な場合があることを示した。

本研究で計算したような大規模な混合整数計画問題が最適化ソルバーを利用して解けるようになったことは最適化手法の急速な発展によるところが大きい。また、今回は中古車オークションデータを利用して分析を行ったが、本研究で用いた手法は汎用性の高い方法ばかりであり、その用途は今回のデータに限ったものではない。

今回のように何らかの割合を予測する場合にはロジスティック曲線を当てはめた非線形回帰がよく応用されるが（例えば文献[9]など）、今後の課題として、こうした非線形回帰に本研究で提案したような工夫を組み込んで分析を行うことが考えられる。また、本研究では実験結果が良好であった距離×排気量のセグメント化を採用したが、近年では整数計画法を利用してサンプルの分類と回帰分析を行う手法が提案されており[2]、こうした手法を応用することも今後の課題としたい。

謝辞 2名の匿名の査読者からのコメントにより論文が著しく改善されました。心より感謝申し上げます。また、研究内容に関して有意義なアドバイスをいただいた住田潮先生（筑波大学）、後藤順哉先生（中央大学）、生田目崇先生（専修大学）、原稿に助言をいただいた山本芳嗣先生（筑波大学）に深く感謝を申し上げます。データをご提供いただき、研究発表の際には多くの貴重なコメントをいただいたデータ解析コンペティション関係者の皆様方に心からの謝意を表します。最後になりましたが、問題を解くためにXpress-MPを使わせていただいたDash社に御礼申し上げます。

## 参考文献

- [1] Arthanari, T. S. and Dodge, Y.: *Mathematical Programming in Statistics*, John Wiley & Sons, New York (1981).
- [2] Bertsimas, D. and Shioda, R.: "Classification and Regression via Integer Optimization," *Operations Research*, Vol. 55, No. 2, pp. 252-271 (2007).
- [3] Kahneman, D. and Tversky, A.: "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, Vol. 47, No. 2, pp. 263-291 (1979).
- [4] Koenker, R. and Bassett, G.: "Regression

- Quantiles," *Econometrica*, Vol. 46, No. 1, pp. 33-50 (1978).
- [5] Konno, H. and Yamamoto, R.: "Choosing the Best Set of Variables in Regression Analysis Using Integer Programming," to appear in *Journal of Global Optimization*.
- [6] Pflug, G. Ch.: "Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk," in *Probabilistic Constrained Optimization: Methodology and Applications*, S. Uryasev, eds., Kluwer Academic Publishers, Dordrecht (2000).
- [7] Rockafellar, R. T. and Uryasev, S.: "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking & Finance*, Vol. 26, No. 7, pp. 1443-1471 (2002).
- [8] Takeda, A.: "Support Vector Machine Based on Conditional Value-at-Risk Minimization," Department of Mathematical and Computing Sciences, Research Report, Tokyo Institute of Technology (2007).
- [9] 岡太彬訓, 木島正明, 守口剛編集: 「マーケティングの数理モデル」, 朝倉書店 (2001).
- [10] 末吉俊幸: "最小絶対値法による回帰分析," *Journal of the Operations Research Society of Japan*, Vol. 40, No. 2, pp. 261-275 (1997).
- [11] 古川一郎, 守口剛, 阿部誠: 「マーケティング・サイエンス入門」, 有斐閣 (2003).