

# アメリカ市場の車の価格予測

学籍番号 192010

氏名 喜田圭伍

授業名 計量経済学

提出日 10月20日

## 結論

調整済み散布図、重回帰係数推定値の有意性、符合条件、Cp基準、正規分位点・分位点プロットによる残差の正規性診断、残差プロットによる分散不均一性診断を通じて、高級車30車種、レクサス 1 車種、メルセデス 2 車種、ジープ 1 車種、フォード 2 車種、ホンダ 1 車種、日産 3 車種の合計 10 車種のモデルを除いたデータにモデル 1 を適用した以下のモデルが最適モデルとなった。価格に関しては分布を正規分布に近づけるため自然対数をとったものである。

最適モデル ( $R^2=0.794$ )

$$\begin{aligned} \text{価格} = & 9.631(281.849) + 0.003 \cdot \text{最高出力}_i(22.447) + 0.301 \cdot \text{高級ブランド} \\ & i(11.897) + 0.0756 \cdot \text{全輪駆動者}_i(3.0168) + \text{残差}_i \end{aligned}$$

※かっこの中の値はt値

## データの説明

本レポートで使用したデータは乗用車314車種から8万ドル以上の30車種、まだ販売されていない-LEXUS UX、商業車 4 車種、水素車のToyota Miraiを取り除いた268車種のデータである。目的変数として価格（ドル）、説明変数として最高出力（Hp）、全輪駆動車（AWD）、スポーツ用多目的車（SUV）、ラグジュアリーカー（Luxuary）、ハイブリッドカー（Hybrid(option)）、データ内のメーカーの平均価格が約4万4千ドル（約500万円）以上の高級ブランド（exp\_brand)の6つを選択した。今回は乗用車の価格を予測することが目的なため、商業車については除外した。高級ブランドというダミー変数は元データになかったものだが、データ内のブランドの平均価格が4万4千ドル以上のものは1、それ以外は0として説明変数に加えた。

## 散布図行列

図 1 は上述の 6 つの変数に価格を加えたデータ同士の相関関係を見るための相関係数行列である。横軸と縦軸にそれぞれの変数が並んでおり、これを見ると最高出力と価格には線形のような関係が見える。また他の変数にも価格との関係が見られる。変数同士の相関関係はこのグラフでは見られないが、このグラフだけからは確信が持てないため調整済み散布図を見てみる。複数の調整済み散布図を書いたが必要な最適モデルと調整済み散布図を見てみる。

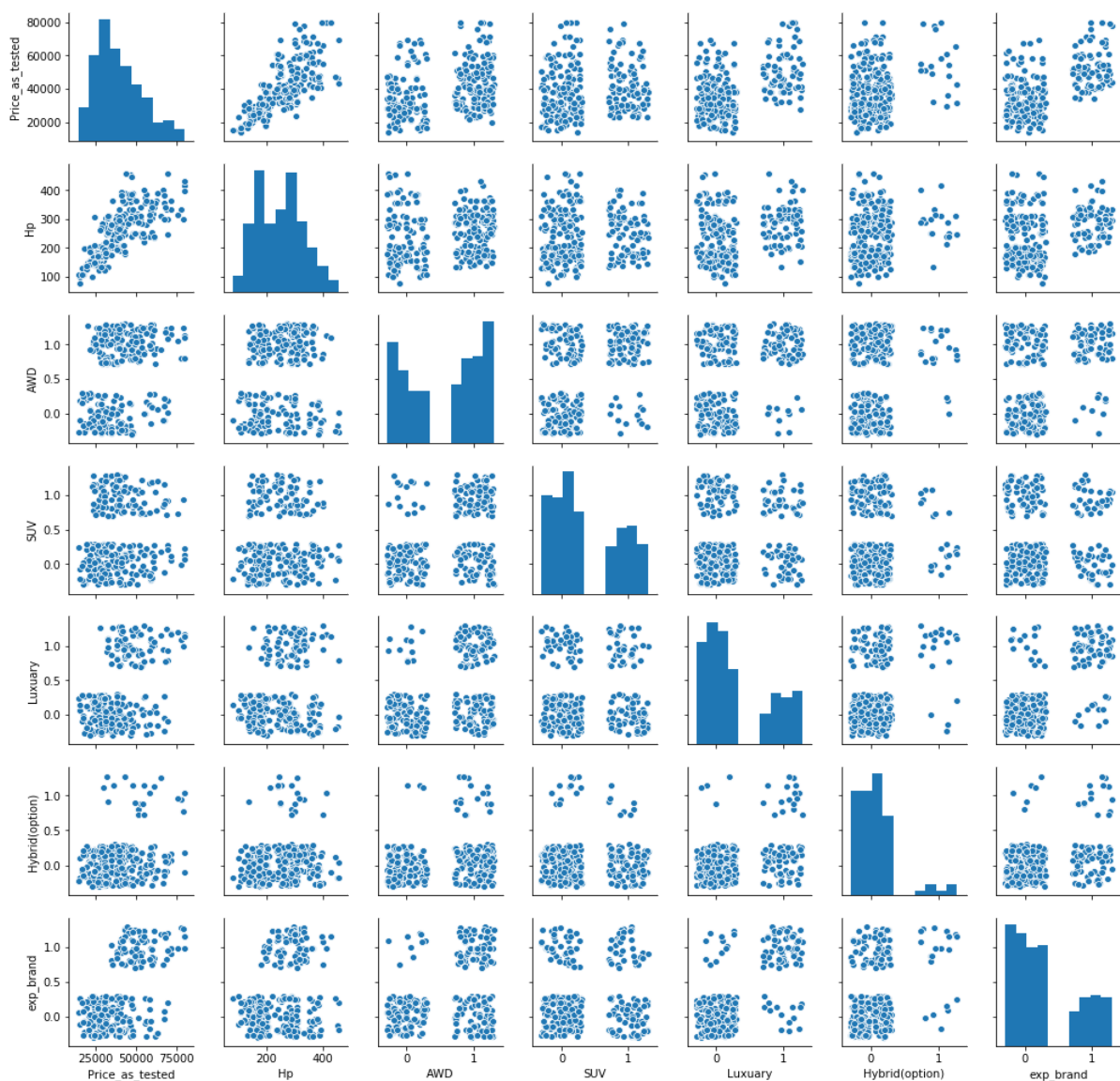
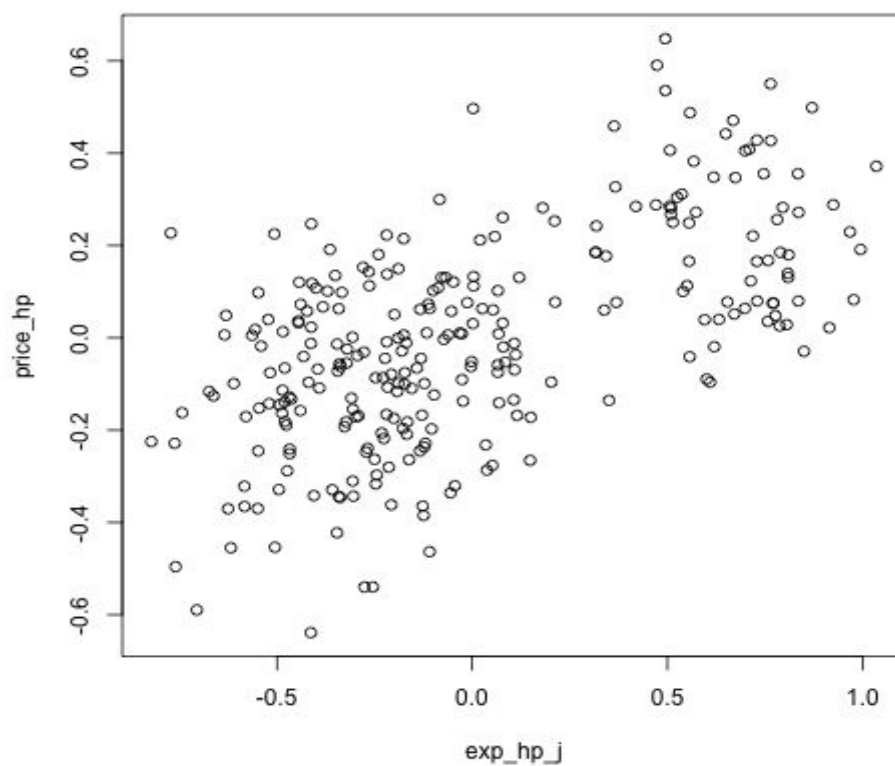


図 1 : 乗用車データの 6 変数の散布図行列

## 調整済み散布図

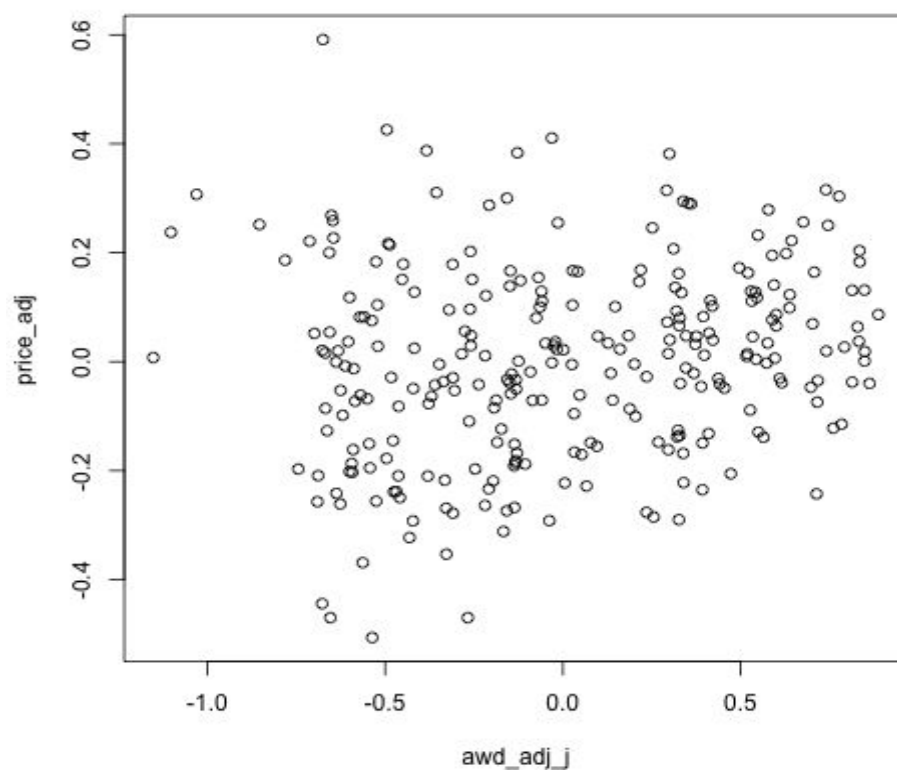
最終的に複数の調整済み散布図を書いたが本レポートでは最適モデルの調整済み散布図のみに触れる。図2は最高出力を説明変数とし、価格を目的変数とした単回帰モデルに高級ブランドを説明変数として加えられるかを見るための調整済み散布図である。図3は最高出力、高級ブランドを説明変数として価格を目的変数とする重回帰モデルに全輪駆動が変数として加えられるかを見るための調整済み散布図である。これらを見るとt値の絶対値が2以上で正になっているため符号条件も満たしており、この3変数を同時に説明変数として加えられることが確認できた。



Estimate	Std.Err	t-value	
Intercept	0.0000	0.0205	0.0000
X	1.2014	0.0895	13.4168

図2：価格を目的変数、最高出力を説明変数とする単回帰モデルに高級ブランドが説明変数として加えられるか確かめるための調整済み散布図

- ① 価格 =  $a_0 + a_1 \cdot \text{最高出力} + \text{価格を最高出力で調整した残差}_i$
- ② 高級ブランド =  $b_0 + b_1 \cdot \text{最高出力} + \text{高級ブランドを最高出力で調整した残差}_i$
- ③ ②の残差をx軸に、①の残差をy軸にプロット



	Estimate	Std.Err	t-value	Pr(> t )
Intercept	0.0000	0.0287	0.000	1.0000
X	0.4529	0.1619	2.797	0.0055

図3：価格を目的変数、最高出力と高級ブランドを説明変数とする重回帰モデルに全輪駆動が説明変数として加えられるか確かめるための調整済み散布図

① 価格 =  $d_0 + d_1 \cdot \text{高級ブランド}_i + d_2 \cdot \text{最高出力}_i + \text{価格を最高出力、高級ブランドで調整した残差}$

② 全輪駆動 =  $e_0 + e_1 \cdot \text{高級ブランド}_i + e_2 \cdot \text{最高出力}_i + \text{全輪駆動を最高出力、高級ブランドで調整した残差}$

③ ②の残差をx軸に、①の残差をy軸にプロット

### 3つのモデル

複数の調整済み散布図を描いた結果、比較的、決定係数が高かった以下の3つのモデルが導き出された。最高出力の予測への貢献が大きかった。

モデル1 ( $R^2=0.783$ )

価格 =  $9.631(271.351) + 0.003(21.9176) \cdot \text{最高出力}_i + 0.301(11.1694) \cdot \text{高級ブランド}_i + 0.0756(3.1750) \cdot \text{全輪駆動者}_i + \text{残差}_i$

※データにあったブランドの平均価格が約4万4千ドル以上の車を高級ブランドとし、ダミー変数exp\_brandを加えた。33ブランドのうち16車種が該当した。

モデル2 ( $R^2=0.735$ )

価格 =  $9.628(249.7556) + 0.00312(20.5967) \cdot \text{最高出力}_i + 0.230(8.1001) \cdot \text{ラグジュアリーカー}_i + 0.146(3.0728) \cdot \text{ハイブリッド}_i + 0.0639(2.5683) \cdot \text{スポーツ用多目的車}_i + \text{残差}_i$

モデル3 ( $R^2=0.736$ )

価格 =  $9.606(246.3482) + 0.00312(20.6928) \cdot \text{最高出力}_i + 0.218(7.4894) \cdot \text{ラグジュアリーカー}_i + 0.108(4.1288) \cdot \text{全輪駆動者}_i + \text{残差}_i$

### Cp基準によるモデル選択

多数の説明変数の候補の中から説明力のある部分集合を選ぶためにCpプロットを書く。図4は調整済み散布図を書いたモデルに対してCp値を計算し、上記の3つのモデルが含まれる部分のCp値をプロットしたものである。各モデルがどこに位置するかをモデル1、モデル2、モデル3のように追記してある。Cp基準によるとモデル1が最適なモデルであることが分かる。

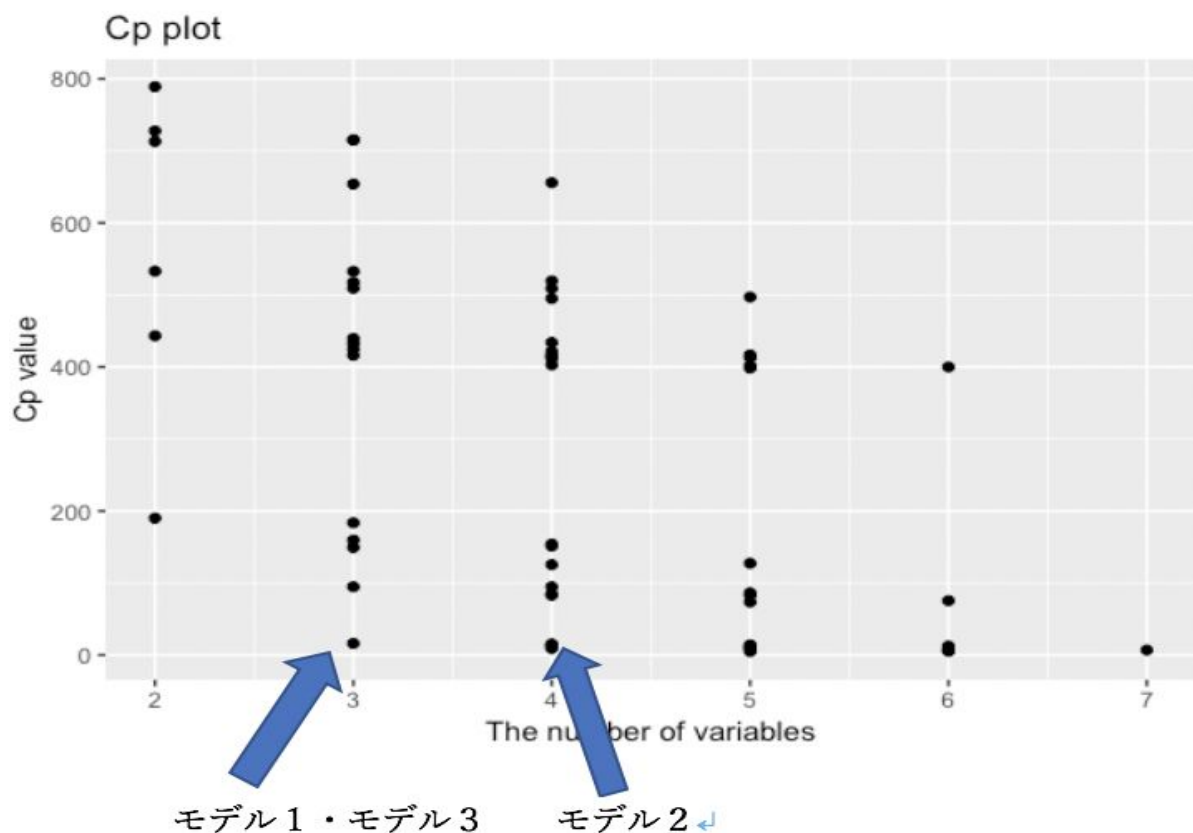


図4: X軸に（Y切片を含めた）説明変数の数をY軸にCpをとったCpプロット

Cp基準によるとモデル1が最適なモデルであることが分かる

## 正規分位点・分位点プロット

選び出した重回帰モデルがどれほど正確に価格を説明しているのかを確認するため正規分位点・分位点プロットを書いてみる。モデルに使用した説明変数が価格をうまく説明していれば価格の残差が正規分布をしているはずである。そのため正規分位点・分位点プロットは直線になるはずである。直線から逸脱している点があればそれらを異常値として取り除く。図5はモデル1に基づいた正規分位点・分位点プロットである。異常値があることが確認出来る。残差の最も大きかった車種はジープグランドチェローキー限定版であったため価格が通常よりも高くなっていると判断し、異常値として除外した。図6は異常値を取り除いたあと再度プロットを描いた物である。当てはまりの良さが確認できる。図7はモデル2に基づいて描かれた正規分位点・分位点プロットである。継続的な当てはまりの悪さが見える。

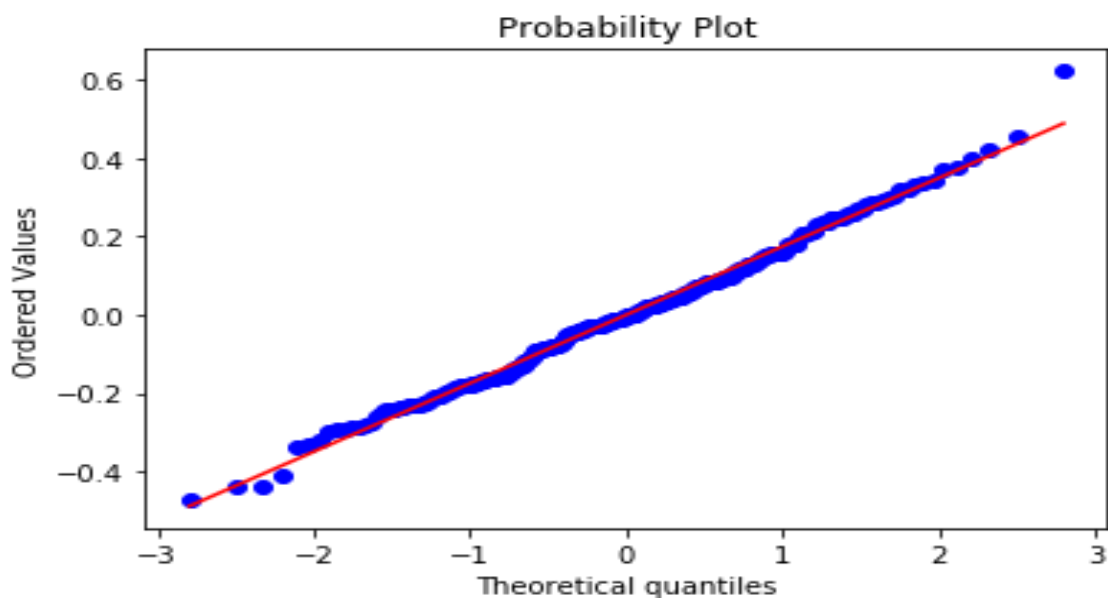


図5：モデル1に基づいた正規分位点・分位点プロット



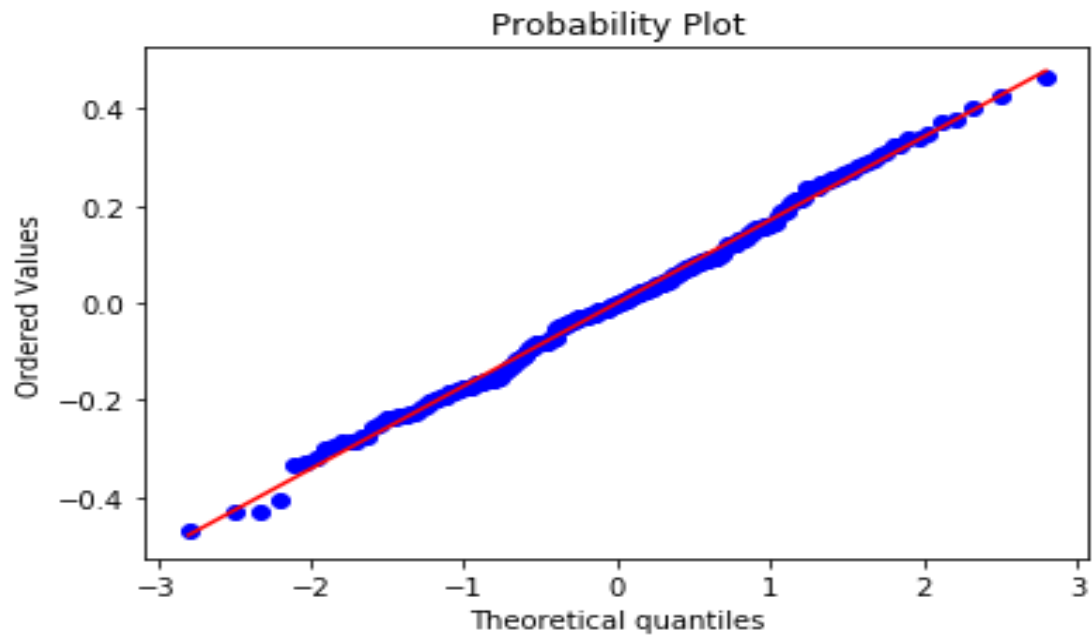


図6：モデル1に基づいた正規分位点・分位点プロット

ジープ限定版を異常値として取り除いた

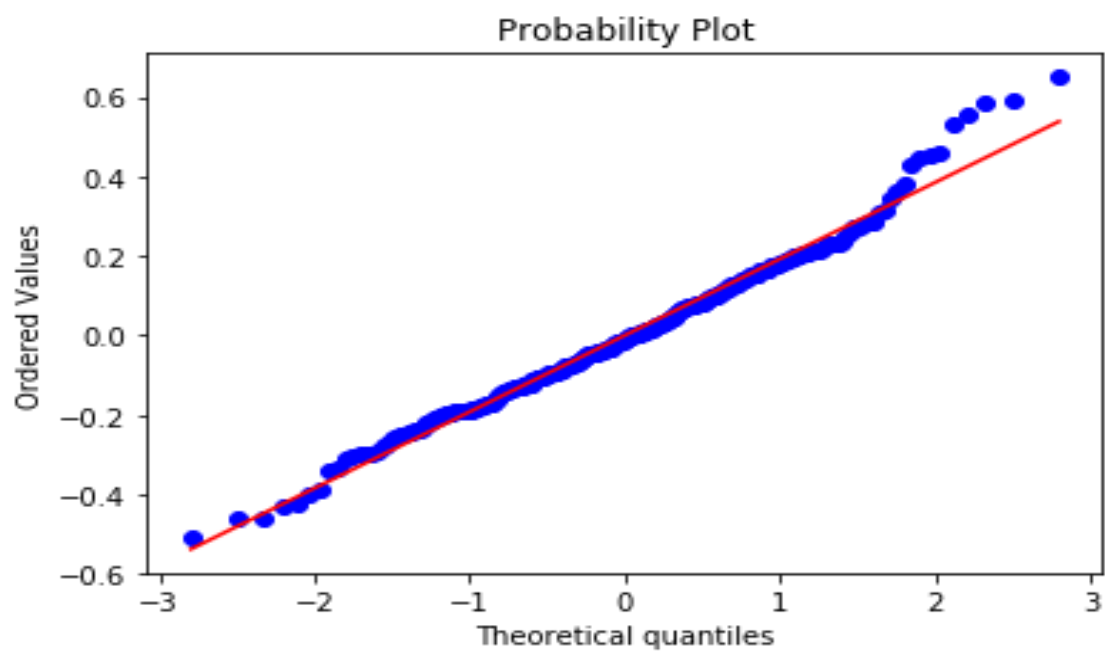


図7：モデル2に基づいた価格の正規分位点、分位点プロット

残差が大きく、当てはまりの悪さがある

## 残差プロット

図8はX軸に価格の予測値、Y軸に残差をプロットしたものである。残差が-0.4以下の4つの点（2つの点が重なっている）がある。フォードフィエスタSE2種、ホンダシビックセダン、日産ヴァルサSVセダンの4車種であり、価格帯がかなり低い物であり、どれもセダンであったため不人気で価格が落ちていると想定し、除外した。

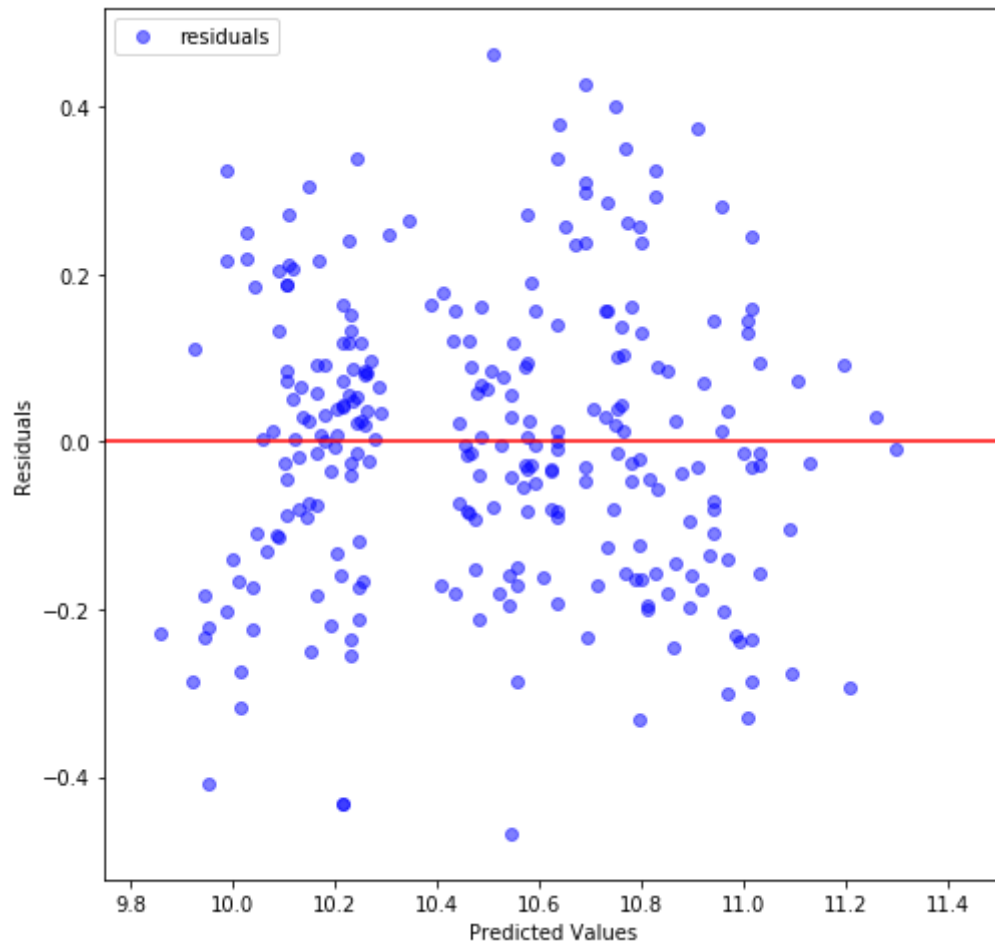


図8：モデル1に基づいた価格の残差対価格の予測値の散布図

## 最後に

以上の解析から、始めに述べたように最適モデル ( $R^2=0.794$ )

$$\begin{aligned} \text{価格} = & 9.631(281.849) + 0.003 \cdot \text{最高出力}_i(22.447) + 0.301 \cdot \text{高級ブランド} \\ & _i(11.897) + 0.0756 \cdot \text{全輪駆動者}_i(3.0168) + \text{残差}_i \end{aligned}$$

と結論付けた。