

# ヘドニックモデルの理論と実証

## 住宅価格の推定

岩田真一郎（神奈川大学経済学部）\*

2025-01-26 作成，2025-01-26 改訂

第1節では，ローゼンが開発した住宅の属性を用いて住宅の市場価格を説明するヘドニックモデルについて紹介します\*<sup>1</sup>。第2節では，実際のデータを用いて，ヘドニックモデルに基づく回帰分析を通じて市場価格を推計します。

## 1 理論

### 1.1 市場価格

住宅価格に関するニュースは日々報じられており，その関心の高さがうかがえます。家計にとって，住宅は賃貸であれ持ち家であれ，大きな支出を占める（持ち家の場合はさらに大きな資産になる）ため重要です。また，住宅産業の規模や波及効果を考慮すると，企業や政府が住宅価格を注視するのは当然のことと言えるでしょう。

住宅価格と住宅政策では，完全競争的な住宅市場を考え，住宅価格は需要と供給が等しくなるように決定すると伝えました。このように，完全競争市場では，差別化されていない財が取引され，一つの財について一つの価格が設定される一物一価の状況を分析します。しかし，現実には一物一価の例を考えるには非常に細分化された市場を考える必要があります。例えば乾電池の1個の価格はいくらでしょうか？乾電池は，サイズ（単1，単2など），性能（アルカリ電池，リチウム電池など），使用推奨期限（5年，10年など）によって価格が異なるため，具体的な属性を指定しないと価格を示すことは困難です。むしろ，乾電池の価格はその属性によって異なると考えた方が自然であり，サイズや性能，使用推奨期限などに基づいて価格を分析するのが現実的です。住宅も同様に，属性によって価格が大きく異なる好例のサービスです。

実際に住宅の賃貸や購入を検討する際，消費者は不動産情報サービスを提供するポータルサイトやウェブサイトを参考にすることがあるでしょう。その際，自分の求める条件を選択し，物件を絞り込みます。具体的には，住宅の所有形態（賃貸，購入），立地属性（都道府県，市区町村，沿線など），建物属性（建て方，構造，間

---

\* 本稿は「費用便益分析の応用と実践」（2024年9月政策研究大学院大学）の一環として講義した内容に基づいています。このような貴重な機会を提供してくださった城所幸弘教授に感謝申し上げます。

<sup>\*1</sup> Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34-55.

取り、面積など）を選び、自分の希望に合った物件を見つけようとしています。そして、（最も知りたい情報である）住宅価格はこれらの属性に依存して決まっていると考えられます。

そこで以下では、住宅の市場価格（以下では、住宅価格は用いず、市場価格を用います）と住宅属性の関係について検討します。住宅の属性を  $h_1, h_2, h_3, \dots, h_n$  で表し、そのリストを属性ベクトル

$$\mathbf{h} = (h_1, h_2, h_3, \dots, h_n)$$

で表しましょう。すると、住宅 1 戸当たりの市場価格は次のように属性ベクトルの関数として表せます。

$$p = p(\mathbf{h}) \quad (1)$$

市場価格と住宅属性の情報を集め、(1) 式を特定化することによって、市場価格関数を推定することができます。この過程で、右辺に示されるそれぞれの住宅属性がどれだけ左辺の市場価格に影響を与えるのかという、住宅属性が持つ暗黙的な価格を見いだせます。この暗黙的な価格をヘドニック価格とよびます。

市場価格関数がどのような形状になるのかを調べるために、国土交通省の不動産情報ライブラリ、不動産価格（取引価格・成約価格）情報の検索・ダウンロードから価格情報をダウンロードして、散布図と近似曲線を描いてみましょう。なお、近似曲線とは散布図に対して最も当てはまりのよい形状を持つ曲線を指します。

ここでは、次の条件を設定して、データをダウンロードしました。

- 
- 地域：神奈川県横浜市神奈川区
    - － 価格情報区分：☒ 不動産取引価格情報 ☒ 成約価格情報
    - － 種類：中古マンション等
    - － 時期：2023 年第 3 四半期～2024 年第 2 四半期
- 

なお、不動産取引価格とは国土交通省が実施したアンケートに基づく市場価格であり、成約価格とはレインズマーケットインフォメーションに公表されている市場価格です。詳細はこちらを参照してください。データ上は両者を区別できますが、ここでは、不動産取引価格と成約価格をひとまとめにして市場価格とよびます。

それでは、ダウンロードした市場価格と住宅属性の関係をプロットしてみましょう。ここでは、住宅属性として面積（図 1A）と建築年（図 1B）をそれぞれ取り上げています。図 1 の各点は実際に観察される市場価格と住宅属性の組み合わせを表しています。市場価格と面積の関係をみると、近似曲線はおおむね右上がりになり、面積が拡大すると、市場価格が上昇する傾向にあることがわかります。建築年に関しては、面積ほどではないですが、近似曲線の形状から、建物が新しくなると、市場価格は上昇する傾向にあります。このように、住宅の面積や建築年で測った品質が良いほど、一般的に市場価格は高くなります。

図 1 の特徴を捉えるために、いま住宅属性の一つである  $h_1$  を取り上げ、 $h_1$  と市場価格の関係を描いたところ、図 2 のように右上がりの曲線  $p(h_1)$  になったとします。この曲線  $p(h_1)$  は市場価格曲線とよべれます。しかし、現実には、図 1 で見られるように、分析者は A 点や B 点のような、住宅属性と市場価格の組み合わせ

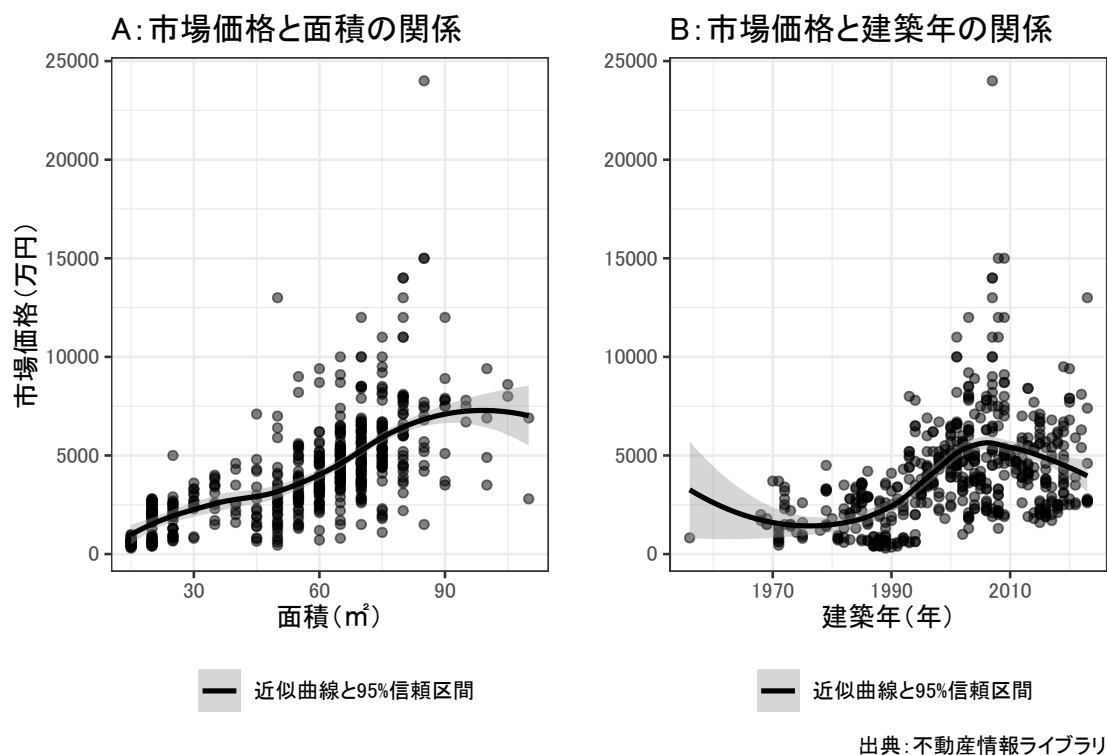


図1 中古マンション等の市場価格と属性の関係 <br> (横浜市神奈川区 2023/3Q~2024/2Q)

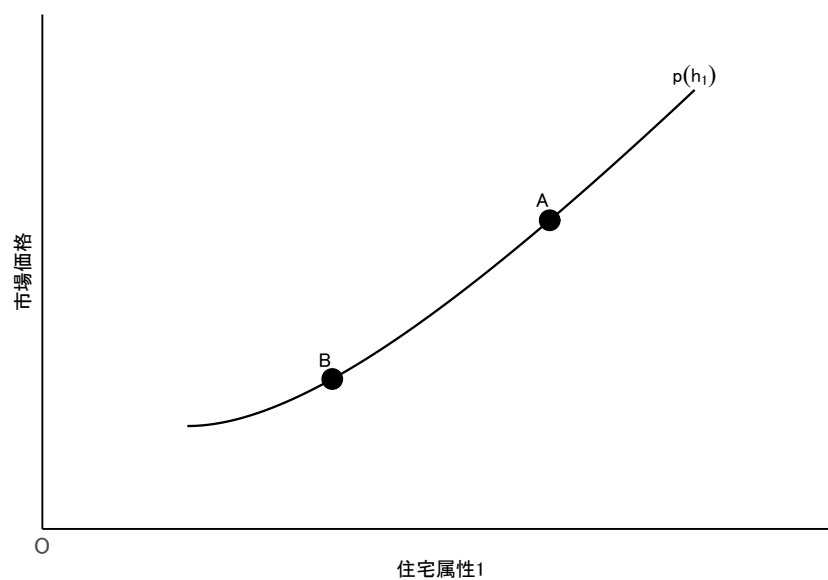


図2 市場価格曲線

しか観察できません．このため，図 1 で見たように，近似曲線を描いたり，第 2 節で見るように，住宅属性に回帰したりすることで市場価格と住宅属性の関係を明らかにしようと試みます．

それでは，なぜ  $A$  点や  $B$  点の様な組み合わせが観測されるのでしょうか．それは，これらの点において実際の契約（・取引）が交わされているからに他なりません．このことは消費者と生産者がこの点を選択していることを意味します．この理論的な基礎づけをしたのがローゼンであり，住宅価格がヘドニック価格として表現できることから，その理論モデルをヘドニックモデルとよびます．そこで，以下ではヘドニックモデルで登場する消費者行動，生産者行動を紹介し，消費者と生産者が  $A$  点や  $B$  点において契約を交わすことを示しましょう．

## 1.2 消費者行動

所得  $y_A$  をもつ消費者  $A$  は，住宅（ $h$ ）とその他の消費財の集合（合成財） $z$ ，の消費量を決定するとしましょう．ただし，以下では，簡単化のため，住宅は属性  $h_1$  のみで表現され，合成財は消費量ではなく消費額で表現します．

図 3 には，縦軸に金額で計測される合成財の消費量が，横軸に住宅（住宅属性 1）の消費量が示されています．また，縦軸には，消費者  $A$  の所得  $y_A$  も示されていますが，これは住宅消費量がゼロのときに購入できる合成財の最大額に相当します．ここで，消費者  $A$  が  $h_1^0$  の住宅を選択したとしましょう．消費者  $A$  の効用は  $h_1^0$  と合成財消費額に依存します．例えば，合成財消費額が  $z^0$  の場合，消費者  $A$  は  $U_A^0$  の効用を得ることになります（ $U_A^0$  は無差別曲線を示し，消費者  $A$  の効用水準が  $U_A^0$  であることを表しています）．このとき，消費者  $A$  は  $h_1^0$  の住宅に対して，所得から合成財消費額を差し引いた  $y_A - z^0$ ，すなわち図 3 の矢印で示される  $\theta_A(h_1^0)$  を支払っています．

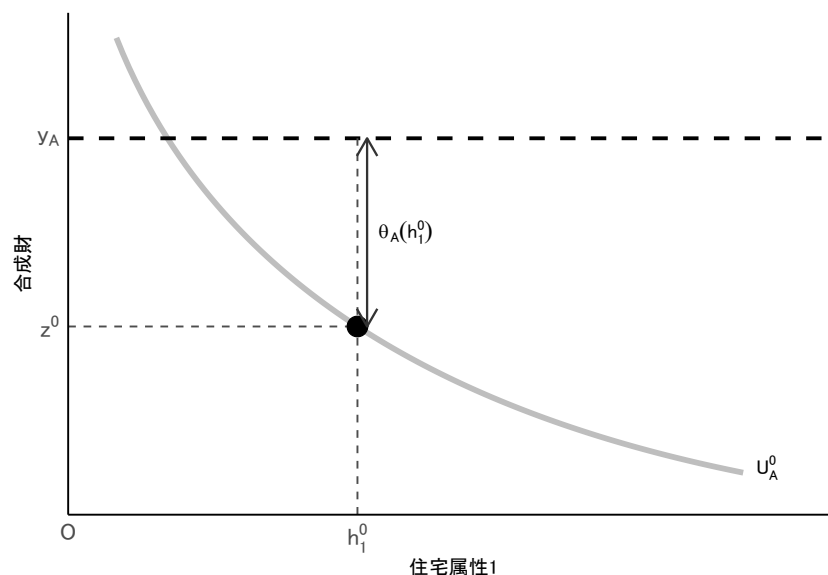


図3 無差別曲線と付け値 1

この矢印の長さは，消費者  $A$  が  $U_A^0$  を達成するために， $h_1^0$  の住宅に対して支払える最大額になります．例

例えば、図4の破線の矢印のように、より短い額を支払う場合は、消費者Aは合成財消費額を $z^0$ よりも増やせるため、 $U_A^0$ よりも高い効用( $U_A^1$ )を達成する無差別曲線に移動できます。このことは、消費者Aは $U_A^0$ を達成すればよいのであれば、より高い額を提示してもよいことを意味します。逆に、図には示していませんが、実線の矢印よりも高い額を支払ってしまう場合は、合成財消費額が $z^0$ よりも減らさなくてはならないため、効用は $U_A^0$ よりも低下します。以上から、消費者Aが $U_A^0$ を達成するために $h_1^0$ の住宅に対して支払える最大額は図3の $\theta_A(h_1^0)$ (図4の実線の矢印)に等しくなることがわかります。ヘドニックモデルでは、この $\theta_A(h_1^0)$ を付け値とよびます。

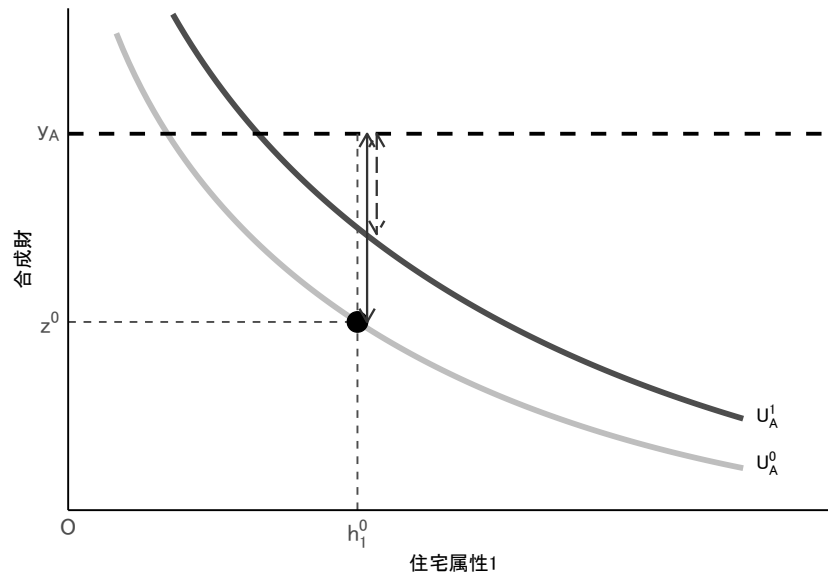


図4 無差別曲線と付け値 2

住宅属性1の消費量を増やし、より品質の高い住宅を消費する場合、 $U_A^0$ を達成するための付け値はどのように変化するでしょうか？例えば、図5の $h_1^1$ に対して消費者Aの付け値は $\theta(h_1^1)$ に変化します。すなわち、消費者Aの付け値は、品質の高い住宅を選択することに伴い上昇することがわかります。

ここまでの話から、 $U_A^0$ を達成するための付け値は消費者Aの $y_A$ と $U_A^0$ の垂直距離に等しくなります。そこで、 $y_A$ を通る水平の破線を軸に図5の上下を反転させてみましょう。すると、図6を得ます。図6は、住宅属性1の消費量が増えてゆくと、消費者Aの付け値が上昇してゆくことが確認できます。ヘドニックモデルでは、図6の曲線 $\theta_A^0$ を付け値曲線とよびます。

消費者は市場価格曲線 $p(h_1)$ を所与として、効用が最大になる住宅を選択するはずです。そこで、図2に消費者Aの付け値曲線を加え、消費者Aの効用が最大になる住宅(とその市場価格の組み合わせ)を考えましょう。図7には、消費者Aの付け値曲線が3本描かれています。図7の付け値曲線 $\theta_A^0$ に対応する効用水準は図4の $U_A^0$ になり、付け値曲線 $\theta_A^1$ に対応する効用水準は $U_A^1$ になります。消費者にとって、住宅水準が一定の場合に、支払い価格が低いほど嬉しいはずですので、付け値曲線の位置が低いほど消費者Aの効用は高くなります。したがって、消費者Aは市場価格曲線上でできるだけ下方に位置する付け値曲線上の住宅、すなわち市場価格曲線と付け値曲線の接点であるA点を満たす $h_1^0$ 、を選択することで効用を最大にします。

消費者が同質的な場合、すべての消費者が消費者Aと同様に図7のA点を選択するようになります。この

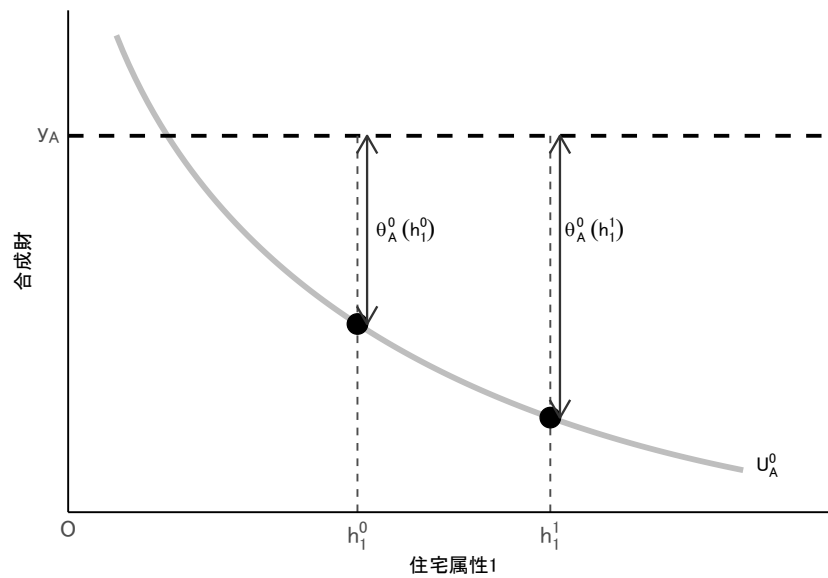


図5 住宅属性と付け値の関係

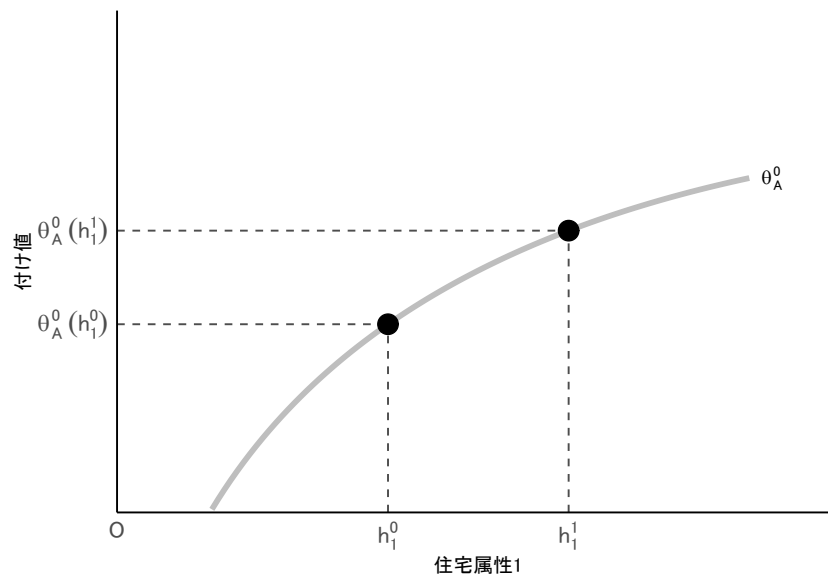


図6 付け値曲線

場合、図 1 の散布図のような様々な住宅属性と市場価格の組み合わせは観察できないはずです。そこで、消費者の異質性を考えてみましょう。消費者が異質的な場合、付け値曲線は消費者 A のそれとは異なった形状になります。図 8 には、消費者 A だけではなく、消費者 B の付け値曲線も描かれています。消費者 B の付け値曲線が消費者 A の付け値曲線の形状と異なるのは、消費者 B の選好（無差別曲線の形状）や所得が消費者 A のそれらと異なることを反映しています。ここで、消費者 B も消費者 A 同様に効用を最大にするような住宅を選択すると考えると、消費者 B が選択する住宅属性と市場価格の組み合わせは B 点になります。このように、消費者の異質性を導入することにより、様々な住宅属性と市場価格の組み合わせを表現できるようになりま

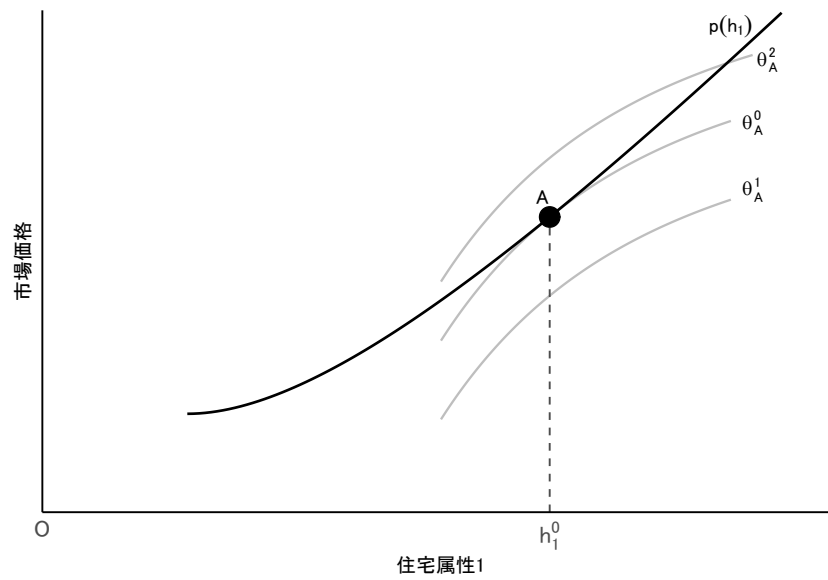


図7 消費者 A の効用最大化

す。

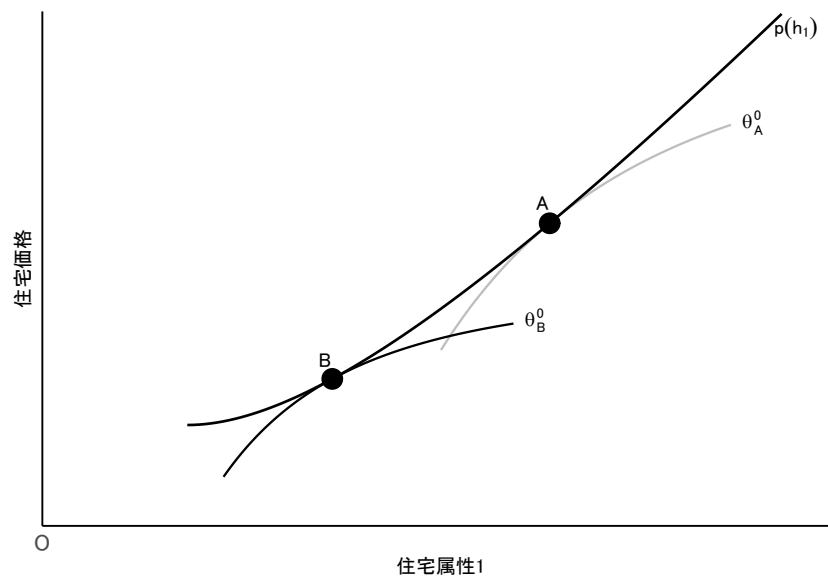


図8 消費者の異質性

### 1.3 生産者行動

しかし、消費者が選択したい住宅を、消費者が支払える範囲の額で生産者が供給しなくては、契約は成立しません。そこで、消費者 A が選択したい住宅を生産者 A が供給するかを考えてみましょう。

生産者 A が住宅属性 1 の住宅を供給する費用を、 $h_1$  の関数として、 $c_A(h_1)$  で表しましょう。ただし、 $h_1$  の供給量が増えると、費用は上昇すると仮定します。ここで、生産者 A は  $\pi_A^0$  以上の利潤を獲得したいと考えているとします。このとき、この利潤の大きさに加えて、費用  $c_A(h_1)$  を賄えなる価格以上で消費者と契約できなければ、 $\pi_A$  以上の利潤を獲得できません。したがって、 $\pi_A$  を獲得するためには、最小限、提示価格を  $\pi_A + c(h_1)$  に設定する必要があります。この、提供する住宅属性に対して、設定された利潤を達成するために必要な最小額はオファー価格とよべます。

例えば、生産者 A は  $h_1^0$  に対しては、 $\pi_A + c_A(h_1^0)$  に等しいオファー価格で、 $\pi_A$  を得られます。そこで、このオファー価格を  $\phi_A^0(h_1^0)$  で表しましょう。図には住宅属性  $h_1^0$  とオファー価格  $\phi_A^0(h_1^0)$  の組み合わせが示されています。生産者 A は、住宅属性 1 の供給量を  $h_1^0$  から  $h_1^1$  に増やす（品質を上昇させる）と、費用の増加に伴いオファー価格を  $\phi_A^0(h_1^0)$  から  $\phi_A^0(h_1^1)$  に上昇させます。このように住宅属性 1 の供給量とオファー価格の組み合わせを求め、それを繋げると、図のオファー価格曲線  $\phi_A^0$  を得られます。ここで、オファー価格曲線の縦軸の切片は  $\pi_A^0$  になり、縦軸の切片から始まる曲線は  $c_A(h_1)$  に等しくなります。なお、オファー価格曲線  $\phi_A^0$  上は生産者 A の利潤は  $\pi_A^0$  で一定になります。生産者 A は  $\phi_A^0$  より高い利潤を得たい場合は、オファー価格を高くすれば良いので、オファー価格曲線は上方にシフトします。

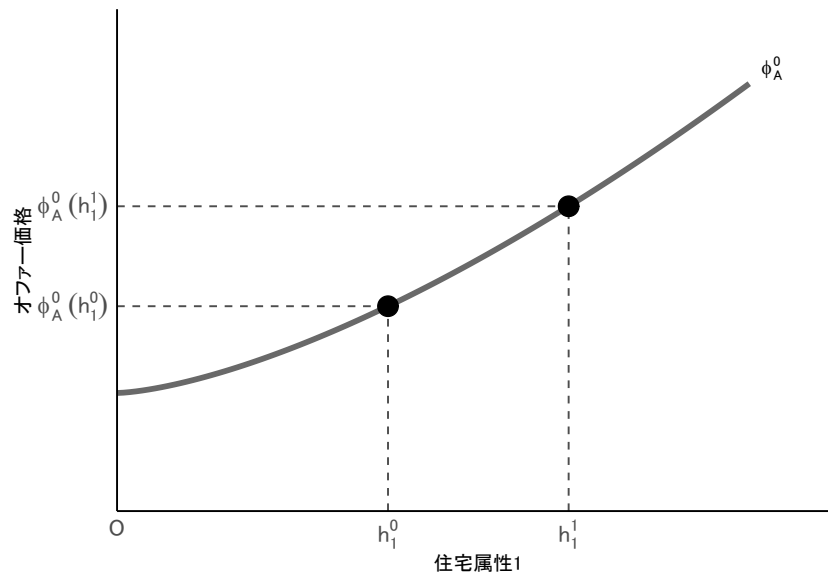


図9 オファー価格曲線

生産者は市場価格曲線  $p(h_1)$  を所与として、利潤が最大になる住宅を選択します。そこで、図 2 に生産者 A のオファー価格曲線を加え、生産者 A の利潤が最大になる住宅（とその市場価格の組み合わせ）を考えましょう。図 10 には、生産者 A のオファー価格曲線が 3 本描かれています。図 10 のオファー価格曲線  $\phi_A^0$  に対応する利潤水準は  $\pi_A^0$  になります。生産者にとって、住宅水準が一定の場合に、受け取り価格が高いほど嬉しいはずですので、オファー価格曲線の位置が高いほど生産者 A の利潤は高くなります。したがって、生産者 A は市場価格曲線上でできるだけ上方に位置するオファー価格曲線上の住宅、すなわち市場価格曲線とオファー価格曲線の接点である A 点を満たす  $h_1^0$ ，を選択することで利潤を最大にできます。

以上から、消費者 A が効用を最大にする住宅属性と住宅価格の組み合わせを生産者 A が提供することがわ



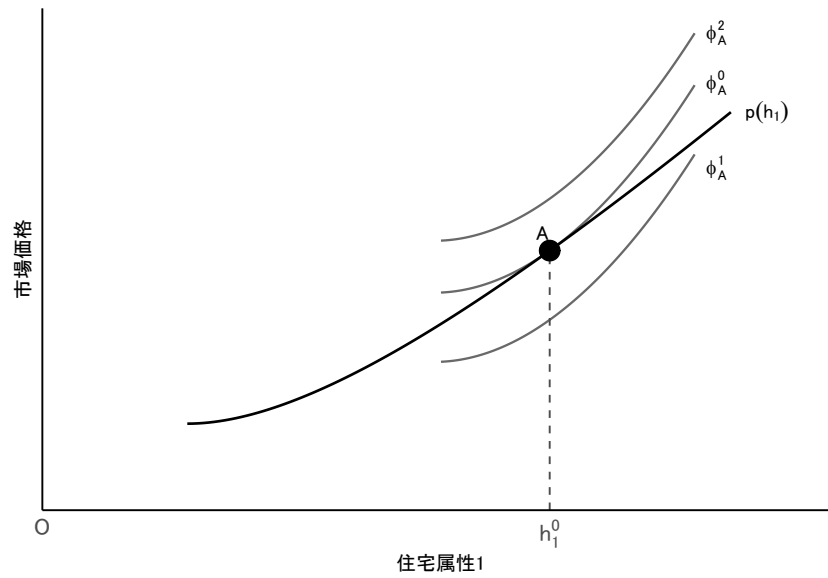


図10 生産者 A の利潤最大化

かりました．これを生産者側から見ると，生産者 A が利潤を最大にする住宅属性と住宅価格の組み合わせを消費者 A が望むということです．この結果，図 11 で示されるように，消費者 A と生産者 A は A 点において契約が成立し，分析者は A 点を観察できるようになります．

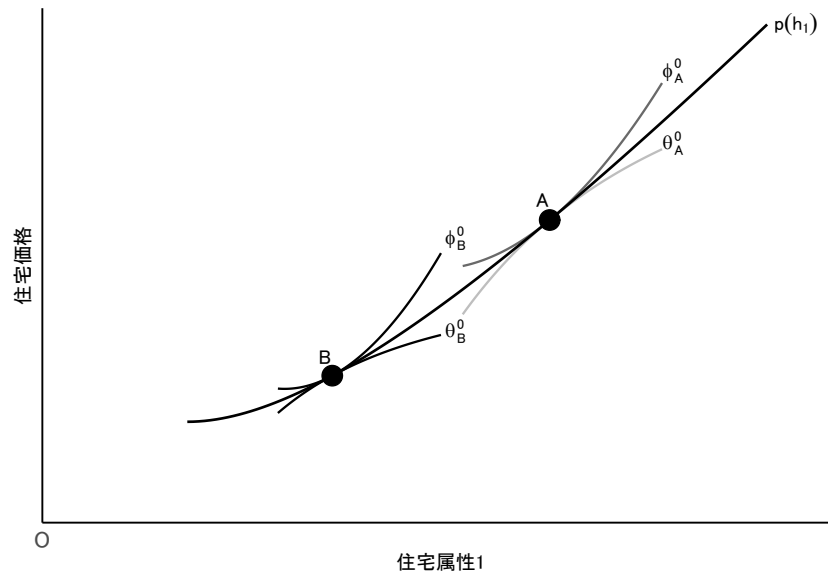


図11 生産者の異質性

生産者が同質的な場合，すべての生産者が生産者 A と同様に A 点の組み合わせを選択するようになります．この場合，消費者 B が望む B 点の組み合わせは選択できません．そこで，生産者の異質性を考えてみましょう．生産者が異質的な場合，オファー価格曲線は生産者 A のそれとは異なった形状になります．図 11 には，生産者 A とは異質な生産者 B のオファー価格曲線 ( $\phi_B^0$ ) も描かれています．生産者 B のオファー価格曲線

が消費者 A のオファー価格曲線の形状と異なるのは、生産者 B の費用曲線の形状が生産者 A のそれと異なることを反映しています。ここで、生産者 B も生産者 A 同様に利潤を最大にするような住宅を選択すると考えると、生産者 B が選択する住宅属性と住宅価格の組み合わせは  $B$  点になります。この結果、消費者 B と生産者 B は  $B$  点において契約が成立し、分析者は  $B$  点も観察できるようになります。

以上から、 $A$  点や  $B$  点の様な組み合わせが観測される理由が解明されました。

## 2 実証

### 2.1 準備

ここからは、実際のデータを用いて市場価格曲線を推定しましょう。ここでは、第 1.1 節市場価格で紹介したデータを、R Markdown 上で R コードを使用して、分析します。R Markdown を使用するためには、フリーソフトである R と RStudio をダウンロードサイトからダウンロードして、インストールする必要があります。

R に慣れていない場合は、分析に入る前に作業フォルダ（ディレクトリ）を作成しましょう。まず、ダウンロードした ZIP ファイルを作業フォルダに格納します。次に ZIP ファイルを右クリックして、「すべて展開」、「展開先の選択とファイルの選択」、「展開を実行」します。この展開後のデータ（拡張子.csv）を使用しますので、最初にダウンロードした ZIP ファイルは削除してかまいません。

データを読み込むために、RStudio を起動し、R Markdown を選択します。この作業ファイル（拡張子.Rmd）も作業フォルダに格納します。

以下の分析で利用するパッケージは `tidyverse`, `modelsummary`, `fastDummies` になります。

```
# データクリーニング・変数作成・図作成
```

```
library(tidyverse)
```

```
# 表の作成
```

```
library(modelsummary)
```

```
# ダミー変数の作成
```

```
library(fastDummies)
```

それでは、オブジェクト名を `hp` として、データを読み込みます。

```
# ダウンロードしたデータの読込
```

```
hp <- read.csv(
```

```
  "Kanagawa Prefecture_Kanagawa Ward_20233_20242/Kanagawa Prefecture_Kanagawa Ward_20233_20242.csv",  
  fileEncoding="CP932")
```

ここで、`fileEncoding` で文字化けを回避していますが、完全ではありません。上手く変換できない文字は「.」として読み込まれます。下記にデータ内容の一部を示していますが、1 行目の変数名に「.」が散見されている

のが分かります。実際に何が書かれているかは作業フォルダにある csv ファイルをメモ帳や Excel などのアプリで開けば、確認できます。

# データ内容

```
hp %>%
  select(市区町村コード, 最寄駅.距離.分., 取引価格.総額.,
         面積..., 建築年, 建物の構造) %>%
  head()
```

```
##   市区町村コード 最寄駅.距離.分. 取引価格.総額. 面積... 建築年 建物の構造
## 1           14102             5      62000000      75 1997年      R C
## 2           14102             9      42000000      80 2001年      R C
## 3           14102            18      18000000      75 1980年      R C
## 4           14102             5      35000000      35 2019年      R C
## 5           14102             4      53000000      70 2014年      R C
## 6           14102             3       4000000      15 1987年      鉄骨造
```

## 2.2 推定式

(1) 式を推定するために、その形状を特定化しましょう。ここでは、理解のしやすい (2) 式のような線形の形状を考えます。

$$p = \gamma_0 + \gamma_1 h_1 + \gamma_2 h_2 + \gamma_3 h_3 \cdots + \gamma_n h_n + \epsilon \quad (2)$$

(2) 式は線形回帰モデルとよばれます。 $\gamma_0, \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_n$  は未知のパラメーターであり、これらの回帰係数はヘドニックモデルにおける各属性の暗黙価格、すなわちヘドニック価格に該当します。実際の市場価格は 1 戸あたりの総額として記録されていますが、属性の量と市場価格の関係は分かりません。しかし、線形回帰モデルを実行すると、回帰係数の値が出力されます。この回帰係数を使って、ヘドニック価格が分かれば、各属性の変化が市場価格にどれだけ影響を与えるかを把握できます。例えば、住宅属性  $h_1$  が 1 単位変化すると、市場価格は  $\gamma_1$  だけ変化することになります。

なお、 $\epsilon$  は誤差項であり、与えられた住宅属性で市場価格を説明できない部分に該当します。

住宅属性に何を含めるかは分析者が決めます。ここでは、データ `hp` に含まれる面積、建築年、最寄駅から距離（最寄駅距離）を取り上げてみましょう。すると、(2) 式は (3) 式のように書き換えることができます。

$$\text{市場価格} = \gamma_0 + \gamma_1 \text{面積} + \gamma_2 \text{建築年} + \gamma_3 \text{最寄駅距離} + \epsilon \quad (3)$$

## 2.3 データクリーニング

回帰分析を実行する前に、ほとんどの場合、データを整備（きれいに）する必要があります。この作業はデータクリーニングとよばれ、非常に手間がかかります。そのため、分析者の多くがこの作業に苦勞することになります。ただし、ここではデータは `hp1` つのみ、使用する変数も 4 つのため、それほどの作業ではありません。

最初に、分析しやすいように 8, 9, 11 列目の変数名を短くします。次に、市場価格は単位を円から万円に変更します。さらに、建築年については末尾の年を削除し、文字列を整数列に変更します。

```
# 変数名の変更
```

```
hp <- hp %>%  
  rename(最寄駅距離 = 8, 市場価格 = 9, 面積 = 11)
```

```
# 市場価格（万円）に変更
```

```
hp <- hp %>%  
  mutate(市場価格 = 市場価格/10000)
```

```
# 建築年を整数列に変更
```

```
hp <- hp %>%  
  mutate(建築年 = as.integer(gsub(" 年", "", 建築年)))
```

この時点で、使用する変数に問題がないか調べてみましょう。ここでは、使用する変数をデータ `hp` から抽出し、オブジェクト名を `hp_test` とします。そして、`datasummary_skim()` を使用して、変数の特徴を見ます。

```
# 使用する変数の抽出
```

```
hp_test <- hp %>%  
  select(市場価格, 面積, 建築年, 最寄駅距離)
```

```
# 変数の特徴
```

```
hp_test %>%  
  datasummary_skim()
```

この表の `Unique` 列は重複しない値の数を、`Missing Pct.` 列は欠損値 (NA) の割合を、`Histogram` 列は変数の度数分布を示しています。表から、サンプルの平均、最小値、最大値などには問題がないように見えます。しかし、`建築年` や `最寄駅距離` には欠損値が含まれているようです。これは、`建築年` や `最寄駅距離` の情報のない物件が存在し、それが欠損値としてデータに記載されることが要因です。`建築年` については、データクリーニング前に空欄箇所だった要素がデータクリーニング過程で NA に変更されます。それに対して、`最寄駅距離` については、データクリーニング前から NA があることがわかります。これから行う回帰分析では、使用する変数に NA を含む物件が存在しても、その物件は自動的にサンプルから排除して推定が行われるため、問題は





	Unique	Missing Pct.	Mean	SD	Min	Median	Max	Histogram
市場価格	121	0	4035.9	2673.1	300.0	3600.0	24000.0	
面積	20	0	54.6	22.7	15.0	60.0	110.0	
建築年	56	2	2000.3	13.3	1956.0	2002.0	2023.0	
最寄駅距離	26	2	7.9	4.4	0.0	7.0	29.0	

表 1: 記述統計量

	平均	標準偏差	最小値	最大値	観察数
市場価格（万円）	4022.28	2662.99	300.00	24 000.00	552
面積（㎡）	54.77	22.65	15.00	110.00	552
建築年（年）	2000.31	13.26	1956.00	2023.00	552
最寄駅距離（分）	7.87	4.42	0.00	29.00	552

ありません。ただし、論文などで発表する際は、回帰分析で使ったサンプルの記述統計を示すことが一般的なため、このままだと都合が悪いことになります。

そこで、次のコードを用いて欠損値 NA を含む物件をサンプルから排除しましょう。

# 欠損値のあるサンプルを取り除く

```
hp <- hp %>%
  drop_na(建築年, 最寄駅距離)
```

記述統計（表 1）を示すと、観察数がきれいに 552 で揃っていることが確認できます。

# 記述統計

```
datasummary(Heading(" 市場価格（万円）")* 市場価格 +Heading(" 面積（㎡）")* 面積 +
  Heading(" 建築年（年）")* 建築年 +Heading(" 最寄駅距離（分）")* 最寄駅距離 ~
  Heading(" 平均")*Mean+Heading(" 標準偏差")*SD+
  Heading(" 最小値")*Min+Heading(" 最大値")*Max+
  Heading(" 観察数")*N, data = hp,
  title=" 記述統計量")
```

## 2.4 推定結果

線形回帰モデルの推定結果は次のコードによって出力できます。なお、市場価格 ~ 面積 + 建築年 + 最寄駅距離の部分は formula オブジェクトとよべれます。

表 2: 市場価格関数の推定結果 1

推定式 (3)	
定数項	−151 376.776*** (11 058.751)
面積	77.916*** (3.232)
建築年	75.770*** (5.522)
最寄駅距離	−54.767** (16.897)
Num.Obs.	552
R2	0.608
+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001 括弧内は標準誤差	

# 推定式 1

```
model1 <- lm(市場価格 ~ 面積 + 建築年 + 最寄駅距離, data=hp)
```

# 推定結果の出力

```
modelsummary(list(" 推定式 (3)"=model1), stars=TRUE,
  coef_rename=c("(Intercept)"=" 定数項"),
  gof_map=c("nobs", "r.squared"),
  title=" 市場価格関数の推定結果 1",
  notes=list(" 括弧内は標準誤差"))
```

表 2 から、市場価格は、面積が 1 m<sup>2</sup>増加すると約 78 万円上昇し、建築年が 1 年新しくなると約 76 万上昇し、最寄駅から距離が 1 分長くなると約 55 万円低下することが確認できます\*2。

## 2.5 説明変数の追加

せっかくですので、他の変数も加えて分析してみましょう。市場価格に影響を与える変数として取引の時点と地点があげられますが、(3) 式はそれらが考慮されていません。それらを考慮した推定式としては、次の (4) 式が考えられます。

\*2 定数項  $\gamma_0$  が負の値で推定されていますが、これは建築年の測り方に影響されています。建築年は最小値が 1956 (1956 年築) ですから、これに回帰係数 75.770 を掛け合わせると、146,700 になり、定数項の負の値をほぼ相殺します。

$$\text{市場価格} = \gamma_0 + \gamma_1 \text{面積} + \gamma_2 \text{建築年} + \gamma_3 \text{最寄駅距離} + \text{地域固定効果} + \text{時間固定効果} + \epsilon \quad (4)$$

ここで、地域固定効果は特定の地域に関連する変動しない固有の影響を、時間固定効果は特定の時期に関連する変動しない固有の影響を考慮します。

しばしば固定効果はダミー変数として調整されます。ダミー変数とは、ある条件を考え、その条件を満たす場合は 1 を、満たさない場合は 0 を取る二値変数です。

### 2.5.1 時間固定効果

最初に、時間固定効果を捉える変数をダミー変数として作成します。hp データには取引時期という変数があるので、この変数と `ifelse()` を用いて、次のように四半期ダミー変数を作成します。

# 時間固定効果ダミーの作成

```
hp <- hp %>%
  mutate("2023/3Q"=ifelse(取引時期=="2023 年第 3 四半期", 1, 0),
         "2023/4Q"=ifelse(取引時期=="2023 年第 4 四半期", 1, 0),
         "2024/1Q"=ifelse(取引時期=="2024 年第 1 四半期", 1, 0),
         "2024/2Q"=ifelse(取引時期=="2024 年第 2 四半期", 1, 0))
```

それでは、この時間固定効果を含んだ式を線形回帰モデルで推定しましょう。コードは以下の通りです。なお、数字から始まる変数は「` (バックティック)」で制御しないと、エラーが発生します。

# 時間固定効果の追加

```
quarter <- lm(市場価格 ~ 面積 + 建築年 + 最寄駅距離 +
              `2023/3Q` + `2023/4Q` + `2024/1Q` + `2024/2Q`, data=hp)
```

# 推定結果の出力

```
modelsummary(list(" 市場価格"=quarter), stars=TRUE,
              coef_rename=c("(Intercept)"=" 定数項"),
              gof_map=c("nobs", "r.squared"),
              title=" 時間固定効果を考慮したモデルの結果",
              notes=list(" 括弧内は標準誤差"))
```

表 3 を見ると、2024/2Q がありませんが、それは参照時点のためです。すなわち、推定結果は、2024 年第 2 四半期に比べて、2023 年第 3 四半期と 2024 年第 1 四半期は有意に市場価格が低いことを示しています。

### 2.5.2 地域固定効果

次に、地域固定効果を捉える変数をダミー変数として作成します。データ hp は地点を示す候補として、都道府県名、市区町村名、地区名、最寄駅名（最寄駅：名称）があります。今回は、神奈川県横浜市神奈川区の

表 3: 時間固定効果を考慮したモデルの結果

	市場価格
定数項	−150 742.737*** (11 106.708)
面積	77.630*** (3.234)
建築年	75.623*** (5.543)
最寄駅距離	−55.274** (16.875)
2023/3Q	−457.034* (207.997)
2023/4Q	−282.314 (229.271)
2024/1Q	−391.824+ (218.271)
Num.Obs.	552
R2	0.611
+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001 括弧内は標準誤差	

データを用いていますので、すべての物件が同じ都道府県、同じ市区町村に立地しています。このため、これらの変数では地点の違いを捉えられません。

そこで、地区名を考慮しましょう。先の四半期ダミー同様に、地区名にあわせて地区ダミー変数を一つ一つ作ることも考えられますが、地区の数を数えると、

```
# 重複しない地区名の数
hp %>% select(地区名) %>%
  distinct() %>% # 重複する行（地区名）を削除
  nrow() # 行（地区名）の数
```

```
## [1] 56
```



と、56 地区にのびります。そこで、`dummy_cols()` を用いて、次のように地区ダミー変数を一括して作成します。

```
# 地区ダミー変数の作成の一括作成
```

```
hp <- hp %>%  
  dummy_cols(select_columns=" 地区名")
```

この作業を通じて、地区ダミー変数が作成できましたので、最初の式に 56 地区のダミー変数（ただし、推定の際は参照地区は取り除かれます）を + で加えてゆけばよいのですが、これも大変面倒な作業です。これを回避する手段として、文字列で回帰式を表現して、それを `formula` オブジェクトに変更する手段を取ることが考えられます。コードは以下の通りです。

```
# 地域固定効果の追加
```

```
formula <-  
  as.formula(paste(" 市場価格 ~ 面積 + 建築年 + 最寄駅距離 +",  
    paste(" 地区名", collapse="+")))
```

```
# 推定モデル
```

```
district <- lm(formula, data=hp)
```

それでは、地域固定効果を含んだ式を線形回帰モデルで推定しましょう。ただし、紙幅の都合上、地区ダミーの一部は表 4 に掲載されていません。具体的には、`coef_omit` を用いて、25 から 59 行目の回帰係数を削除しています。

```
# 推定結果の出力
```

```
modelsummary(list(" 市場価格"=district), fmt=1,  
  coef_rename=c("(Intercept)"=" 定数項"),  
  coef_omit=25:59,  
  gof_map=c("nobs", "r.squared"),  
  title=" 地域固定効果を考慮したモデルの結果",  
  notes=list("+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001",  
    " 括弧内は標準誤差",  
    " 地区ダミーは紙幅の都合上一部省略"),  
  estimate="{estimate}{stars}({std.error})",  
  statistic=NULL)
```

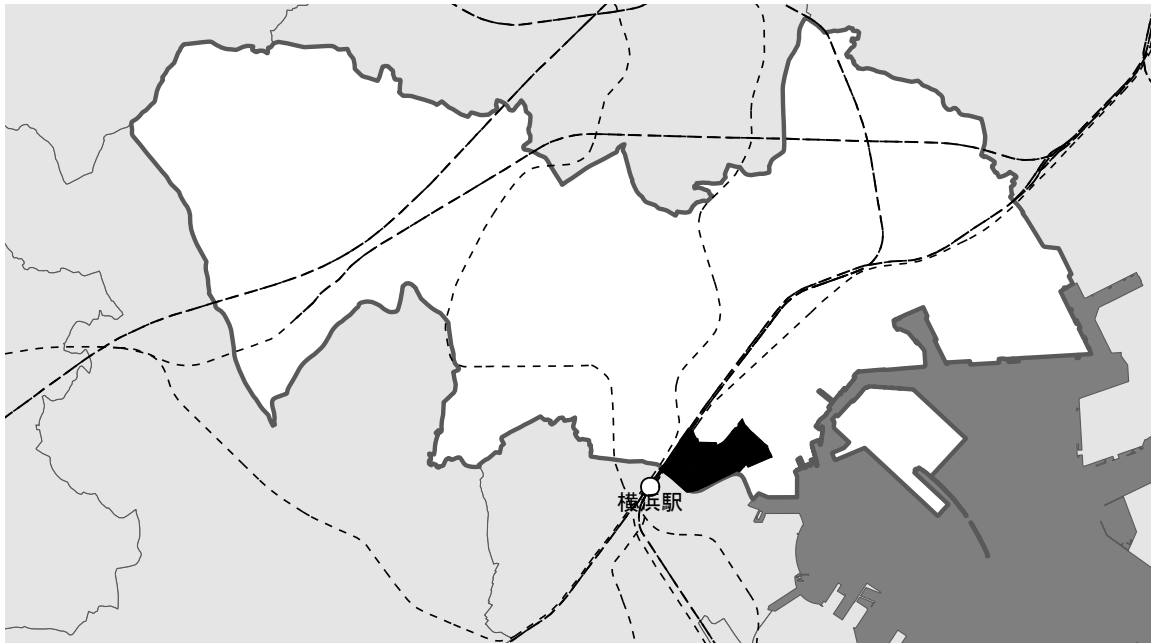
表 4 から、参照地区（旭ヶ丘）に比べて栄町、金港町が有意に市場価格が高いことがわかります。紙幅の都合上、表 4 には大野町が記載されていませんが、大野町も有意に市場価格が高くなります。図 12 ではこれらの地区が黒で色付けされています。栄町、金港町、大野町はヨコハマポートサイド地区の一面にあり、最寄駅が横浜駅であることから交通の利便性が高く、それを反映して市場価格が高くなったと考えられます。

表 4: 地域固定効果を考慮したモデルの結果

	市場価格
定数項	−115 228.7*** (8552.1)
面積	72.1*** (2.5)
建築年	57.6*** (4.3)
最寄駅距離	−48.2** (17.6)
地区名羽沢町	−273.5 (1127.6)
地区名羽沢南	−570.8 (1122.1)
地区名浦島丘	−46.0 (1170.9)
地区名浦島町	23.9 (1107.1)
地区名栄町	1871.7+ (1077.1)
地区名橋本町	1427.0 (1126.9)
地区名金港町	6921.2*** (1114.6)
地区名広台太田町	732.3 (1231.4)
地区名高島台	1007.9 (1173.1)
地区名斎藤分町	333.4 (1506.0)
地区名三ツ沢下町	−283.5 (1143.9)
地区名三ツ沢上町	−523.4 (1122.9)
地区名三ツ沢西町	184.8 (1148.7)
地区名三ツ沢中町	470.2 (1243.3)
地区名三ツ沢南町	106.3 (1501.2)
地区名三枚町	−524.6 (1152.6)
地区名子安台	−1102.9 (1194.3)
地区名子安通	−154.0 (1078.2)
地区名七島町	165.8 (1127.6)
地区名松ヶ丘	−73.9 (1234.2)
Num.Obs.	552
R <sup>2</sup>	0.858

+  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

地区ダミーは紙幅の都合上一部省略



出典:統計地理情報システム, 境界データ, 国勢調査  
国土交通省国土数値情報, 行政区域

図12 神奈川区栄町, 金港町, 大野町

それに対して、その他の地区の市場価格は参照地区の市場価格と有意に変わらない結果となっています\*3。

### 2.5.3 地域固定効果+時間固定効果

最後に地域固定効果と時間固定効果を含めた (4) 式を推定しましょう。そして、これらの固定効果を考慮していない推定結果と並べて出力しましょう。ただし、地域固定効果と時間固定効果の回帰係数の推定結果は、`coef_omit` を使用して出力せず、コンパクトにまとめます。このようなまとめ方は、紙幅の都合上よく見かけるまとめ方です。

# 地域固定効果の追加

`formula2 <-`

```
as.formula(paste(" 市場価格 ~ 面積 + 建築年 + 最寄駅距離 +  
                `2023/3Q`+`2023/4Q`+`2024/1Q`+`2024/2Q`+",  
                paste(" 地区名", collapse="+")))
```

\*3 その理由の一つとして、各地区の物件観測数（サンプル数）の少なさが挙げられます。各地区のサンプル数が少ないと、ダミー変数の標準誤差が大きくなり、回帰係数の帰無仮説（回帰係数は 0 に等しい）を棄却できなくなる確率が高まります。これを回避するには、観測時点を増やすことが考えられます。

表 5: 市場価格関数の推定結果 2

	推定式 (3)	推定式 (4)
定数項	-151 376.776*** (11 058.751)	-115 724.106*** (8552.906)
面積	77.916*** (3.232)	72.173*** (2.535)
建築年	75.770*** (5.522)	57.835*** (4.263)
最寄駅距離	-54.767** (16.897)	-46.835** (17.604)
Num.Obs.	552	552
R2	0.608	0.859

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

推定式 (4) は地域固定効果と時間固定効果を考慮

# 推定モデル

```
model2 <- lm(formula2, data=hp)
```

# 推定結果の出力

```
modelsummary(list(" 推定式 (3)"=model1,
                  " 推定式 (4)"=model2), stars=TRUE,
              coef_rename=c("(Intercept)"=" 定数項"),
              gof_map=c("nobs", "r.squared"),
              coef_omit=c(-1, -2, -3, -4),
              title=" 市場価格関数の推定結果 2",
              notes=list(" 括弧内は標準誤差",
                          " 推定式 (4) は地域固定効果と時間固定効果を考慮"))
```

表 5 から、面積、建築年、最寄駅距離の回帰係数は地域固定効果と時間固定効果を考慮すると（絶対値が）小さくなります。したがって、これらの住宅属性が市場価格に及ぼす影響は小さくなりますが、それでもなお市場価格に対して有意な影響を与えていることが確認できます。

### 3 テキストガイド

付け値曲線、オファー価格曲線、市場価格曲線は周辺環境の属性（治安、災害危険度、公園）にも依存します。そのため、ヘドニックモデルは公共政策の費用や便益（治安の改善、災害危険度の改定、公園の開発）を

計測する際に役立ちます。政策分析の応用例や分析の際の注意点については、以下のテキストや論文が参考になります。

- 金本良嗣・中村良平・矢澤則彦（1989）「ヘドニック・アプローチによる環境の価値の測定」『環境科学会誌』2 巻 4 号 251-266
- 高橋孝明（2012）『都市経済学』有斐閣ブックス
- 金本良嗣・藤原徹（2016）『都市経済学』（第 2 版）東洋経済新報社

また、ヘドニックモデルを用いて因果推論を行うことも可能です。その方法や問題点については、以下の論文が参考になります。

- Christopher F. Parmeter and Jaren C. Pope (2013) Quasi-experiments and hedonic property value methods. *Handbook on Experimental Economics and the Environment*, 3–66, Edward Elgar Publishing.
  - SSRNから 2009 年バージョンがダウンロードできます。
- 牛島光一（2016）「ヘドニック・アプローチにおける因果識別」『都市住宅学』92 号 25-30

R による実証分析に興味がある方は、以下のテキストを推薦します。R コードは `tidyverse` 流で記述されています。

- 星野匡郎・田中久稔・北川梨津（2023）『R による実証分析（第 2 版）回帰分析から因果分析へ』，オーム社

## 4 R 環境

### セッション情報

- R version 4.4.2 (2022-10-31)
  - RStudio 2024.12.0+467
  - rmarkdown 2.28

### 使用したパッケージ

- `fastDummies`, `ggspatial`, `modelsummary`, `sf`, `tidyverse`

## 5 リンク

- 都市経済学講義ノート
- R による地理空間データの可視化

- Shinichiro Iwata

アイコン

- Font Awesome 6.3.0

生成型 **AI** の使用について

本稿の作成過程において，ChatGPT を使用して可読性の向上とコードの最適化に関する助言を得ました．サービスを利用した後，内容を確認し，必要に応じて編集を行いました