

津田塾大学総合政策学部総合政策学科 | データ政策科学

# データ分析演習

松本 崇斗(Takato Matsumoto)  
takato.matsumoto0114@gmail.com

# コンテンツ

1. 不動産データの分析
2. 不動産価格の予測

# コンテンツ

1. 不動産データの分析
2. 不動産価格の予測

# 不動産データの分析

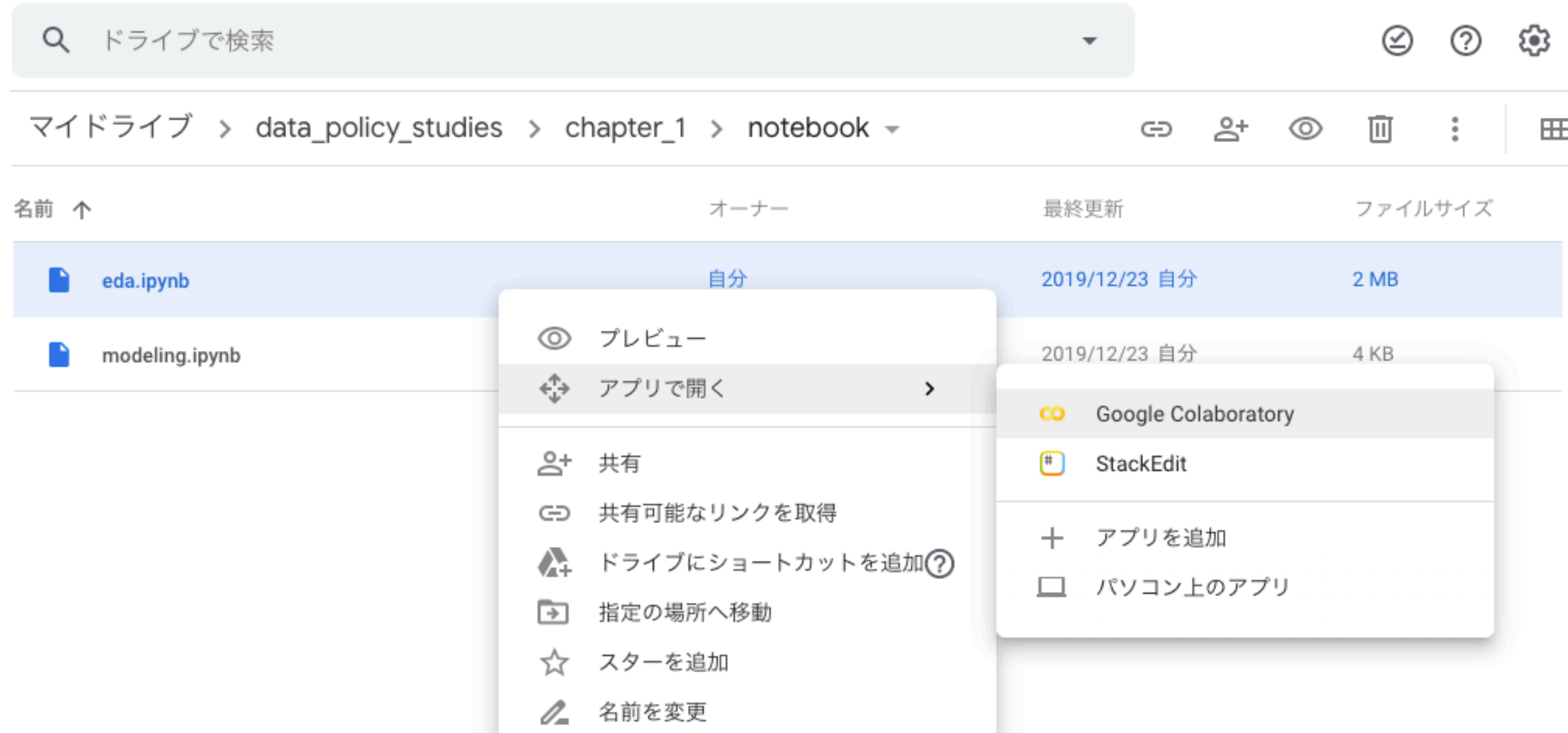
## ■不動産データの分析を行う

- データを分析して,データの特徴をつかむ
- どんな要素が物件価格に影響を与えているのかを分析する

# 不動産データの分析

## ■ Google Colaboratoryを使います

- data\_policy\_studies/chapter\_1/notebook/eda.ipynbを開いてください



# コンテンツ

1. 不動産データの分析
2. 不動産価格の予測

# 不動産価格の予測 | 線形回帰モデルの作成

- 不動産価格を予測するために線形回帰モデルを作成します
- 価格が分かっていない不動産の情報(立地や広さなど)をモデルに当てはめると、価格を予測することができます

# 不動産価格の予測 | データの用意

■ 学習用のデータを使って回帰モデルを作ることによって  
物件価格の予測ができます

■ 以下の例を考えてみます

- Training data: 物件の価格と部屋の広さの関係のデータ
- Test data: 物件の部屋の広さのみのデータ

Training data  
モデル学習用のデータ

price	area
240000	100
300000	200
240000	300
300000	400

Test data  
(priceが分からないがareaが入っているデータ)

price	area
?	150
?	200
?	350
?	400

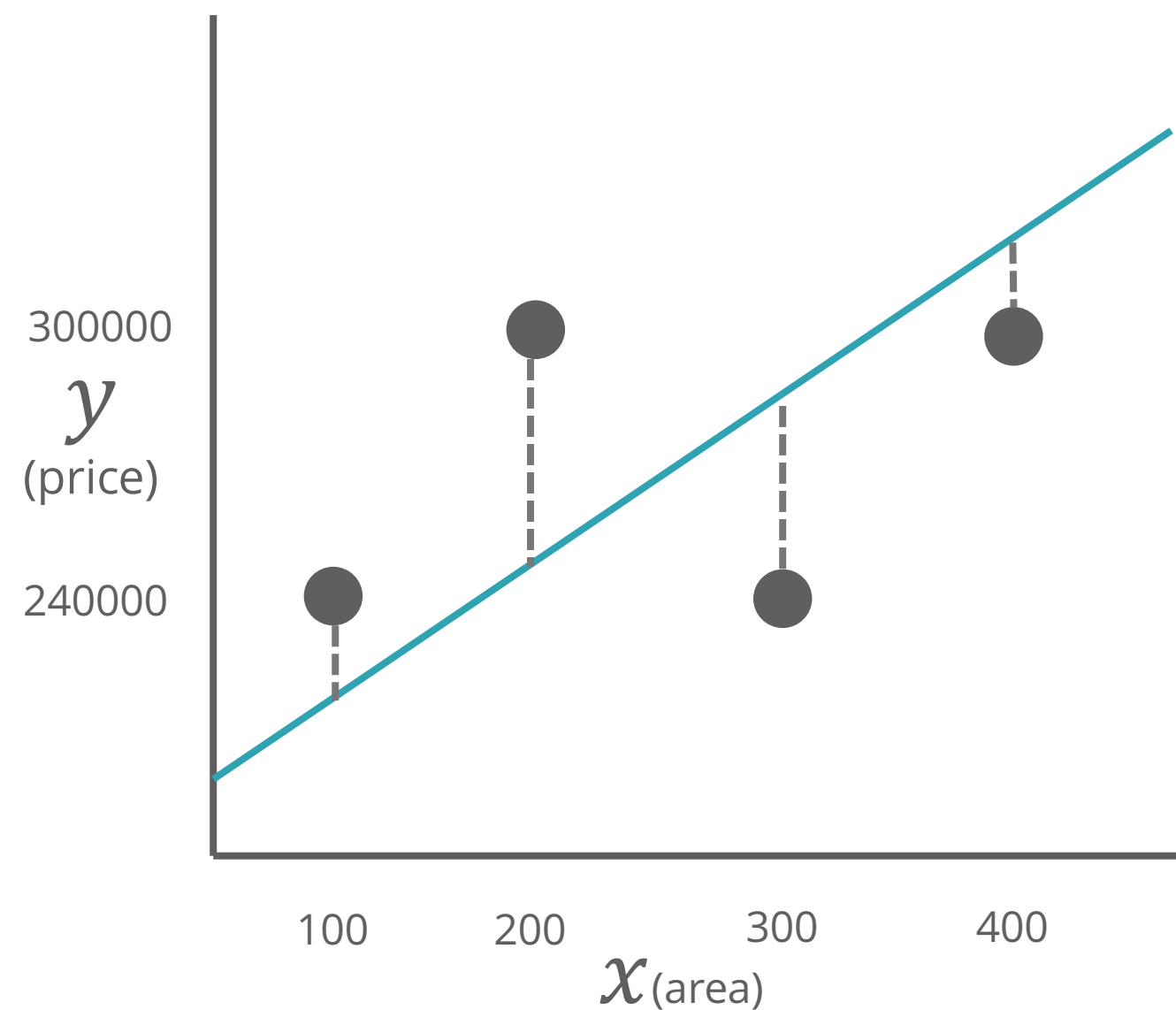


# 不動産価格の予測 | 線形回帰モデルの作成

## ■ Training dataのプロットに対して

点線の長さの合計が最も短くなるよう直線を引くイメージです

- 最小二乗法という手法を使っています (点線の長さの2乗の合計を最小化する)
- 求まった直線が線形回帰モデルです

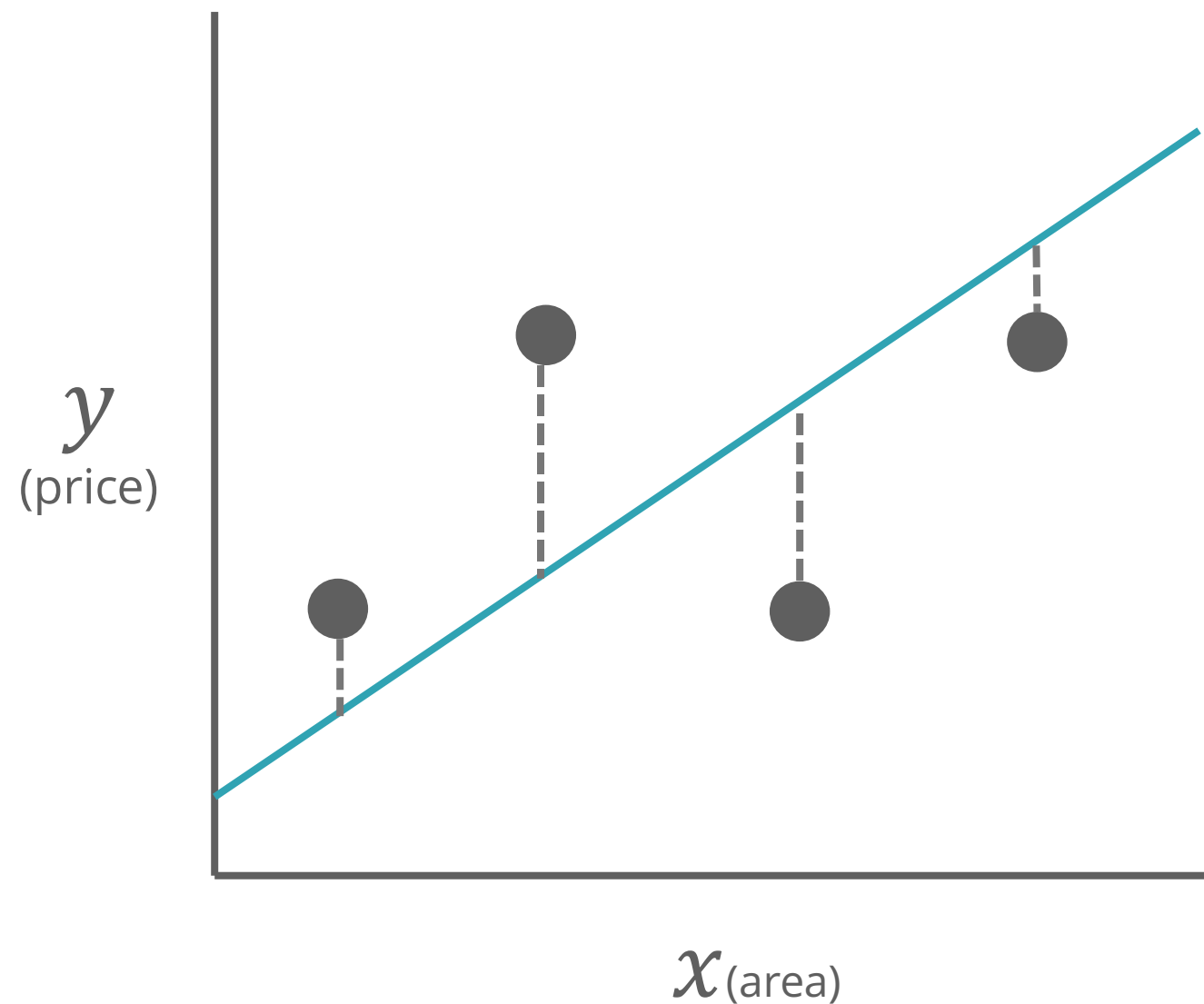


Training data

price	area
240000	100
300000	200
240000	300
300000	400

# 不動産価格の予測 | 線形回帰モデルの作成

- 求まった直線は以下の式で表されます



$$y = ax + b$$

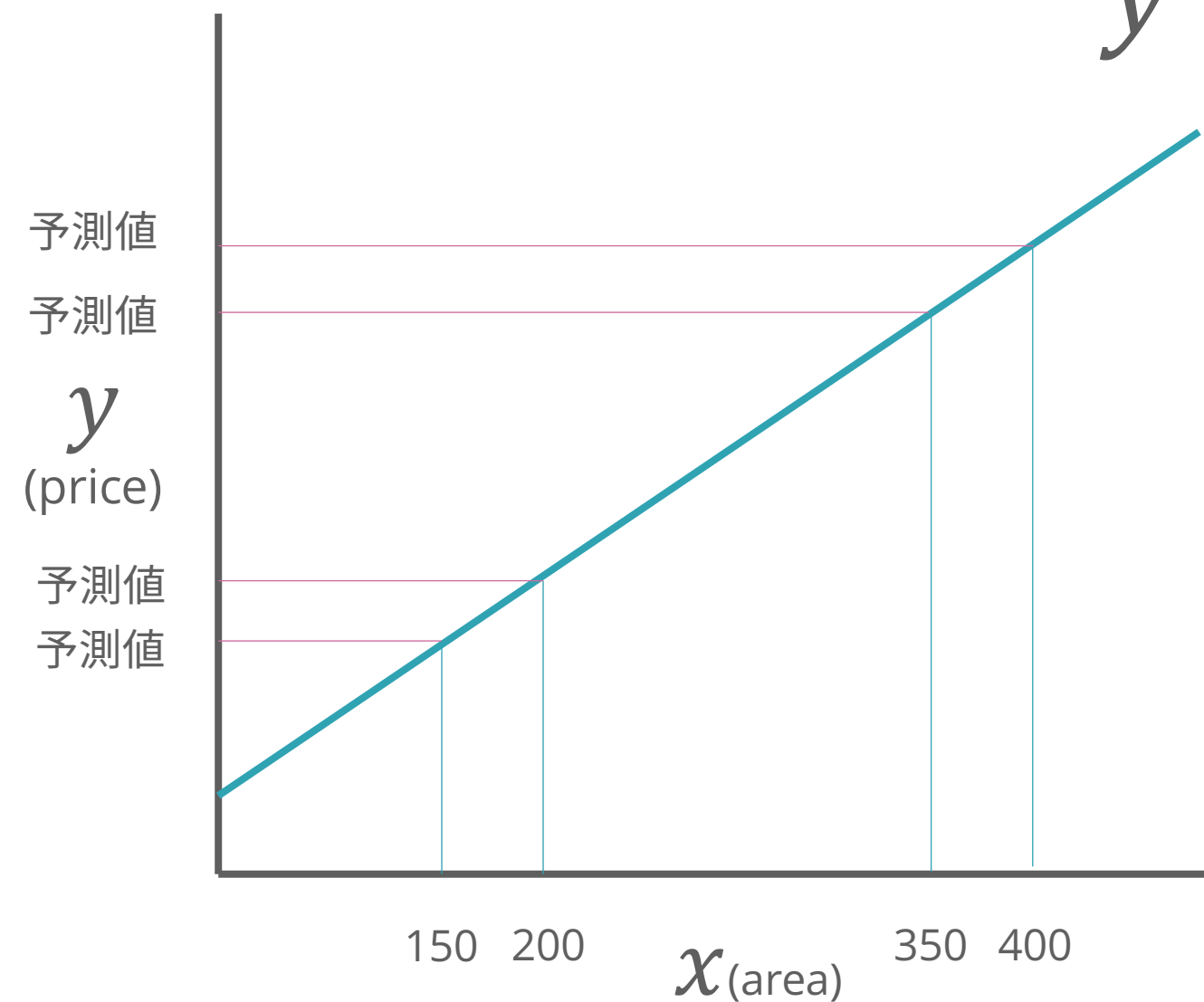
目的変数 = 係数 × 説明変数 + 切片

# 不動産価格の予測 | priceの予測

## ■ 線形回帰モデルに対して

Test dataのareaを代入してpriceを予測します

$$y = ax + b$$



Test data

price	area
?	150
?	200
?	350
?	400

# 不動産価格の予測 | モデルの評価

## ■ 予測モデルの評価

- 予測した価格と実際の価格の誤差の計算を行い, 評価する
  - 誤差が小さいほど良い予測ができており, 精度が高いと言える
- 評価はRMSLE(Root Mean Squared Logarithmic Error)という評価関数を用いる

$$RMSLE = \sqrt{\frac{1}{N} \sum_{i=0}^n (\log(y_i + 1) - \log(y'_i + 1))^2}$$

$y'_i$ : i番目の要素の予測値

$y_i$ : i番目の要素の正解値

Prediction data

id	price
3	468604
4	268328



True data

id	price
3	450000
4	290000

# 不動産価格の予測 | 実践

- `data_policy_studies/chapter_1/notebook/modeling.ipynb`  
を開いてください

# 不動産価格の予測 | 実践

## ■ 予測精度の向上

- 価格に影響を与えると分析した要素(カラム)を入力する
- どんな組み合わせのとき, 精度が高くなったでしょうか?

# 不動産価格の予測 | 実践

## ■ 価格に影響を与えるデータを使うと精度が良くなります

- 例えば、立地、部屋の広さは大きく価格に影響を与えそうなど
- 以下の組み合わせの時RMSLEは0.3179でした
  - sqft\_living
  - grade
  - lat
  - yr\_built
  - view

# 演習課題

- どんな条件が物件価格に影響を与えそうか仮説を立て、分析を行ってください
  - 分析結果のスクリーンショットを提出