
重回帰分析の学習を可能にする 授業教材の提案

岸本有生

大阪電気通信大学高等学校

背景

- 小学校から高等学校の授業において、データサイエンスが重要になっており注目されている
- 高等学校「情報Ⅱ」では、重回帰分析を取り扱う

【授業に必要な教材】

- 分析ツール Excel, Python, R
- サンプルデータ 教科書
- 授業用スライド 教科書

実際に重回帰分析に必要な教材を全て用意して授業で使ってみました

授業提案のポイント

- 「情報I」で学習した散布図や単回帰分析を応用して重回帰分析が理解できる
- 難しい専門用語をあまり使用しない
- 説明変数を複数にすることで予測モデルの精度が単回帰分析よりも高くなると理解することが目的

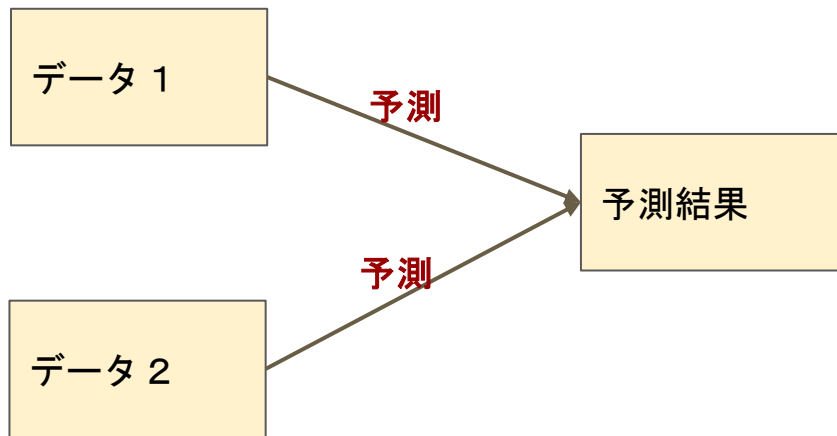
単回帰分析と重回帰分析の違い

単回帰分析イメージ図



1つのパラメータから予測をする

重回帰分析イメージ図



複数のパラメータから予測をする

授業例) マンションの家賃を予測する

正しく家賃を予測するには、どの説明変数が強く効いているかを考えさせる。

目的変数：家賃

説明変数：駅までの時間

築年数

面積

階

賃貸マンション

	家賃	駅からの時間	築年数	面積	階
1	69000	8	1	30.64	1
2	67000	5	34	61.56	3
3	62000	3	6	28.26	1
4	63000	5	16	36.92	7

データ分析ツールの使用

Connect DB

<https://cdb.eplang.jp>

- データ分析ツールConnect DBに機能を追加

(1) 散布図行列

- 全ての散布図を表示し、それぞれの関係性を調べることができる

(2) 重回帰分析

- 重回帰分析から導き出された予測結果と実際の値を比較できる

実演します

Connect DBで散布図行列

方法:

散布図行列

散布図行列を選択

表示

データ削除

入れ替え

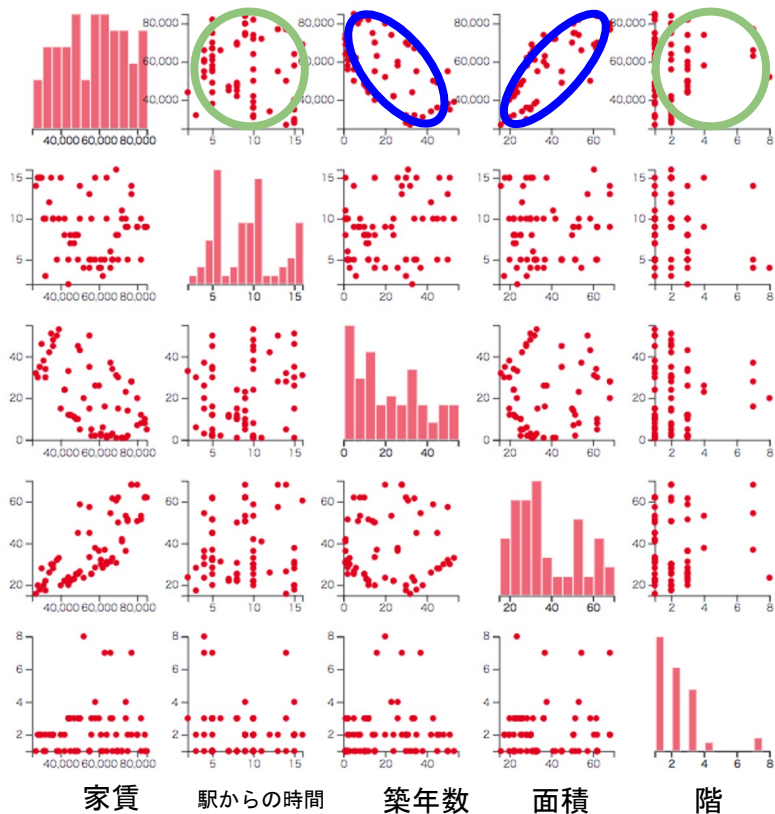
型

スクロールしてテーブルを見てね

<input type="checkbox"/>	<input checked="" type="checkbox"/> 家賃[円]	<input checked="" type="checkbox"/> 駅からの時間[分]	<input checked="" type="checkbox"/> 築年数[年]	<input checked="" type="checkbox"/> 面積[m2]	<input checked="" type="checkbox"/> 階
<input type="checkbox"/>	27000	14	32	15.84	1
<input type="checkbox"/>	28000	15	30	19.87	2
<input type="checkbox"/>	29000	15	35	17.79	2
<input type="checkbox"/>	30000	15	46	28	2
<input type="checkbox"/>	31000	10	38	19.87	1
<input type="checkbox"/>	32000	3	30	17.39	2

家賃、駅からの時間、築年数、面積、階を選択

散布図行列



散布図行列では、全てのデータにおける

- ・ ヒストグラム
- ・ 散布図

を同時に表示できる

「家賃」と「築年数」→右肩下がり
「家賃」と「面積」→右肩上がり

「家賃」と「駅からの時間」
「家賃」と「階」
相関がありそうなことを気付かせる

Connect DBで重回帰分析

方法:

回帰分析

回帰分析を選択

表示

予測したいデータ

予測に使用するデータを選択

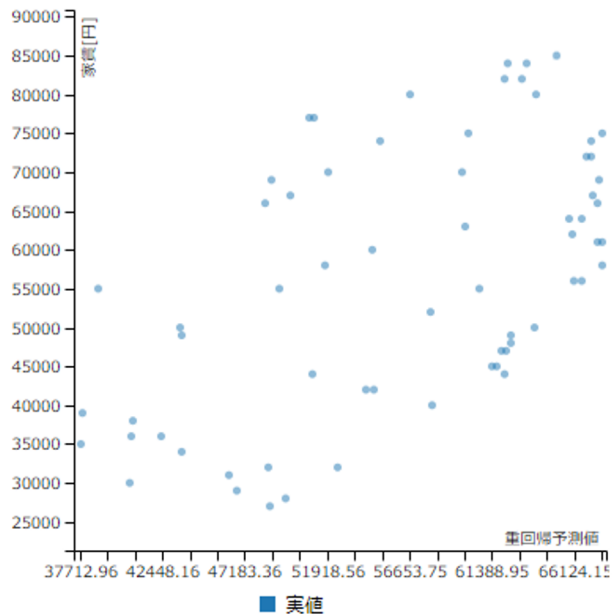
スクロールしてテーブルを見てね

<input type="checkbox"/>	<input checked="" type="checkbox"/> <u>家賃[円]</u>	<input checked="" type="checkbox"/> 駅からの時間[分]	<input checked="" type="checkbox"/> 築年数[年]	<input checked="" type="checkbox"/> 面積[m2]	<input checked="" type="checkbox"/> 階
<input type="checkbox"/>	27000	14	32	15.84	1
<input type="checkbox"/>	28000	15	30	19.87	2
<input type="checkbox"/>	29000	15	35	17.79	2

(1) 駅からの時間, 築年数で家賃を予測

重回帰分析の式: 家賃[円] = 70061.6602 - (247.1877 × 駅からの時間[分])
- (561.5859 × 築年数[年])

決定係数 R2 : 0.2855

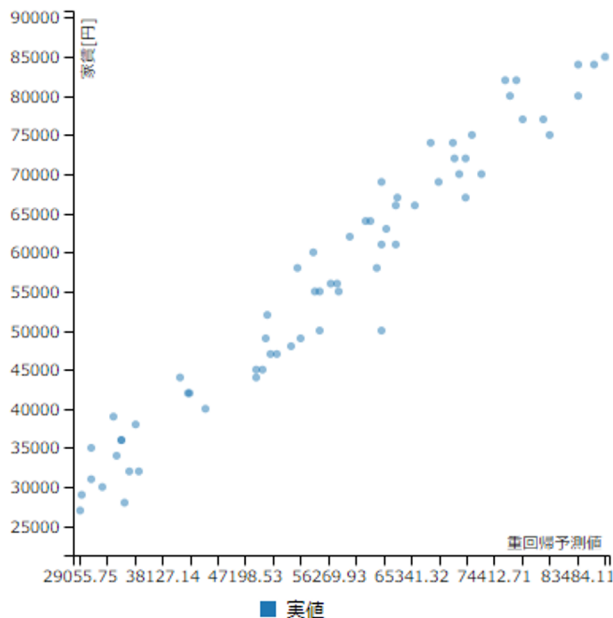


- ・ 説明変数が複数でも、散布図のばらつきが大きいため予測の精度があまり高くない

(2) 駅からの時間, 築年数, 面積, 階で家賃を予測

重回帰分析の式: 家賃[円] = 38029.5014 - (309.8782 × 駅からの時間[分])
- (589.6999 × 築年数[年]) + (866.1367 × 面積[m2]) + (515.3323 × 階)

決定係数 R2 : 0.9481

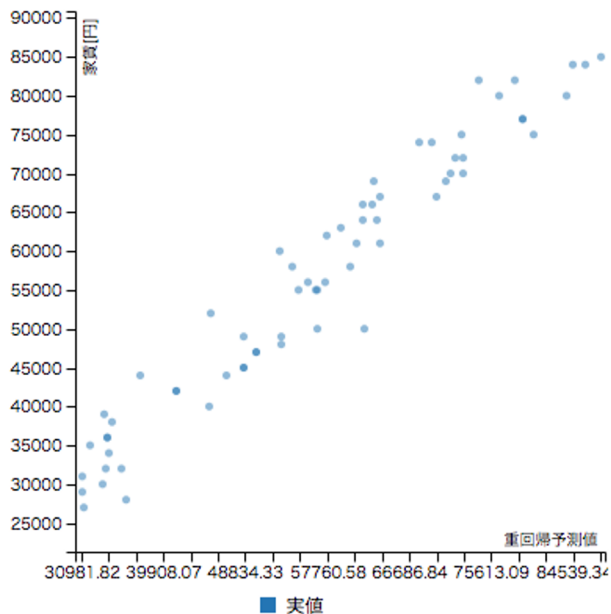


散布図が直線になってきていることから、目的変数を増やすと予測の精度が高くなってきている

(3) 築年数, 面積で家賃を予測

重回帰分析の式 : 家賃[円] = 36660.2314 - (603.7893 × 築年数[年])
+ (868.9271 × 面積[m2])

決定係数 R2 : 0.9418

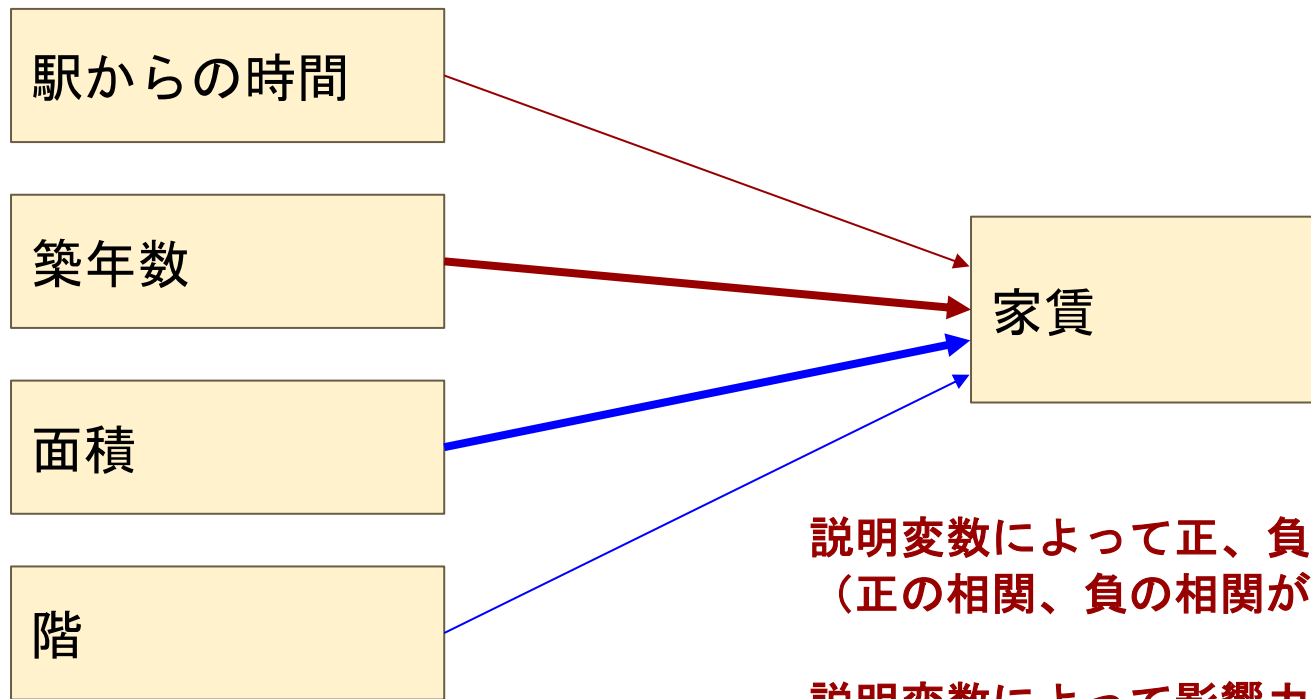


説明変数を減らしても精度が高い



影響力の低い説明変数が存在する
(駅からの時間や階など・・・)

それぞれのデータの関係



説明変数によって正、負の比例関係がある
(正の相関、負の相関がある)

説明変数によって影響力が違う

サンプルデータの設計

(1) 高校生が身近に感じる題材

- 初めから、知らないことを考えたり予測したりすることは難しい

(2) 複数の説明変数から必要なものだけを取り出す課題

- 説明変数を全て選択すれば良いといった課題では、考える必要がなくなるため学習として成り立たない

授業実践

情報Ⅱに相当する重回帰分析の授業を工学部の大学2年生149人に対して行った（90分×2コマ）

(1)気温とアイスクリームの売り上げ数を使った単回帰分析の説明

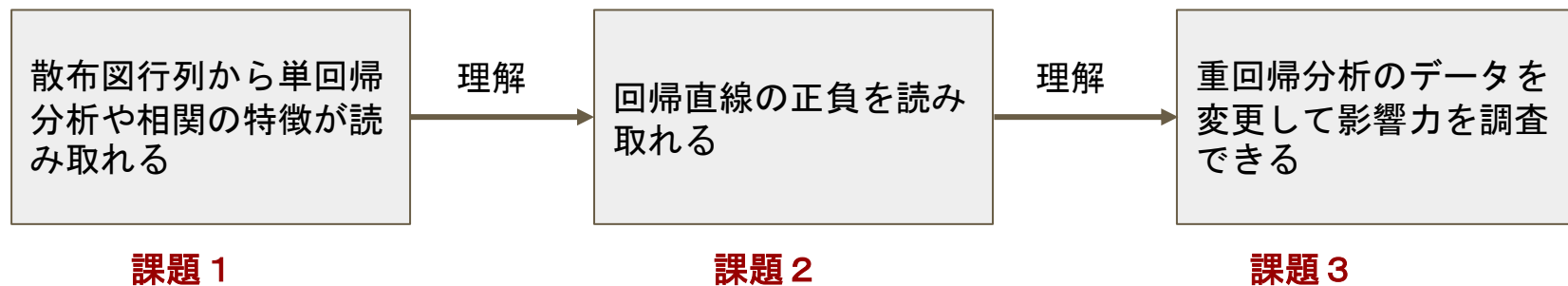
(2)家賃を散布図行列や重回帰分析を利用して予測する

課題の理解度チェック

課題1： 散布図行列からデータの特徴を答えなさい

課題2： 家賃が安くなる傾向がある説明変数を2つ答えなさい

課題3： 重回帰分析に利用しても影響力の低い説明変数を2つ答えなさい



課題1

課題1：散布図行列からデータの特徴を答えなさい

正解率77.9%

正解例では、単回帰分析を読み取れている解答があった

それとは逆に、不正解例はデータの偏りなどの解答が多かった

正解例	駅からの近さは、 そこまで明確な 差はない	築年数は少ない ほうが家賃が高 い	面積は大きいほ うが家賃が高い	階が高い物件は そもそも少ない	5～7階がない
不正解例	広いほうが良い	駅から近さの重 要性	階の問題はあま りない	築年数は0のほ うに集中してい る	当たり前ですが、 家賃を気にして いる人が多い

課題2,3

課題2：家賃が安くなる傾向がある説明変数を2つ答えなさい

正解：駅からの時間、築年数 **正解率：79.9%**

課題3：重回帰分析に利用しても影響力の低い説明変数を2つ答えなさい

正解：駅からの時間、階 **正解率：85.2%**

課題2, 3についても正答率が高い

理解できたことを自由記述

データの要因は複数あることを知った。	複数のデータを組み合わせて分析する方法を学べた。
複数のデータから、さらに正確なデータを予想できると知った。	重回帰分析では影響力の小さい値は省いても問題なかった。
散布図行列でいくつものグラフを一気に見比べることができ、データの中でどれが関係しているか比べる時に使える。	思いもよらない項目がデータに影響を与えていることを知った。
実際にデータ分析すると、自分の予想を超えた結果が得られる。	データ同士の関係を考えることが大切だと知った。

複数のデータを組み合わせることで予測精度が高くなる
影響力の低いデータは、省いても問題ない

おわりに

- 専門的な知識を必要とせず重回帰分析を学習する教材を開発した
- 実際に授業を行ったところ、約8割の学生が単回帰分析の応用で重回帰分析ができることに気付いた