

# 経済と社会 補助資料

## マーケティングと実証分析3

2024年度2学期: 火曜2限  
担当教員: 石垣 司

1

## 再掲: これ以降のお話の設定

背景: マーケティングコンサルタントとして、スーパーマーケットチェーンの販売促進に関する戦略立案を担うことになった



### 利用できるデータ

- 1年分の ID-POS データ: レジ通過時にポイントカードを提示した顧客の購買履歴データ。「誰が、いつ、何を、何個、いくらで」購入したのかが記録されているデータ
- 顧客属性データ: 登録顧客の年齢, 家族人数, 世帯内の高齢者の有無, 子供の有無, 自宅から店舗までの所要時間

目標: 戦略立案のためにまずは顧客属性と購買金額の関係を定量的に把握現状を知りたい

2

## 実証分析とは?

仮説や理論をデータを用いて統計的に検証すること

- 例: 広告費を上げると商品の売り上げは増えるか?
- 例: リフレッシュ休暇をとると生産性は上がるか?

この授業の目標: 線形回帰モデルを利用した分析結果の数値の意味を正しく解釈できる

学習済み

	Estimate (推定値)	Std.Error (標準誤差)	t value (t値)	Pr(> t ) (p値)
切片 ( $b_0$ )	106146	24196	4.39	0.000***
年齢 ( $b_1$ )	841	382	2.21	0.028*
家族人数 ( $b_2$ )	23170	2602	8.91	0.000***
高齢者の有無 ( $b_3$ )	-1063	8202	-0.13	0.897
子供の有無 ( $b_4$ )	7941	7633	1.04	0.299
家からの時間 ( $b_5$ )	-3208	598	-5.37	0.000**
Adjusted R-squared (自由度調整済み決定係数)				0.106

3

## 重回帰分析の結果の解釈 #1

	Estimate (推定値)	Std.Error (標準誤差)	t value (t値)	Pr(> t ) (p値)
切片 ( $b_0$ )	106146	24196	4.39	0.000***
年齢 ( $b_1$ )	841	382	2.21	0.028*
家族人数 ( $b_2$ )	23170	2602	8.91	0.000***
高齢者の有無 ( $b_3$ )	-1063	8202	-0.13	0.897
子供の有無 ( $b_4$ )	7941	7633	1.04	0.299
家からの時間 ( $b_5$ )	-3208	598	-5.37	0.000**
Adjusted R-squared (自由度調整済み決定係数)				0.106

### 赤字部分の解釈

「年齢 ( $b_1$ ), 家族人数 ( $b_2$ ), 家からの距離 ( $b_5$ ) は、有意水準5%で統計的に有意に購買金額に影響を与えている」

- 高齢者の有無 ( $b_3$ ), 子供の有無 ( $b_4$ ) は購買金額に影響を与えているかどうかは分からない

4

## 重回帰分析の結果の解釈 #2

	Estimate (推定値)	Std.Error (標準誤差)	t value (t値)	Pr(> t ) (p値)
切片 ( $b_0$ )	106146	24196	4.39	0.000***
年齢 ( $b_1$ )	841	382	2.21	0.028*
家族人数 ( $b_2$ )	23170	2602	8.91	0.000***
高齢者の有無 ( $b_3$ )	-1063	8202	-0.13	0.897
子供の有無 ( $b_4$ )	7941	7633	1.04	0.299
家からの時間 ( $b_5$ )	-3208	598	-5.37	0.000**
Adjusted R-squared (自由度調整済み決定係数)	0.106			

### 誤った赤字部分の解釈

- 高齢者の有無 ( $b_3$ ), 子供の有無 ( $b_4$ ) は購買金額に影響を与えていない
- 家族人数 ( $b_2$ ) のP値が一番低いので、家族人数が最も強く購買金額に影響を与えている
- 有意水準を1%に設定すると、年齢 ( $b_1$ ) が有意ではないので有意水準5%の方が良い分析結果である

5

## 重回帰分析の結果の解釈 #3

	Estimate (推定値)	Std.Error (標準誤差)	t value (t値)	Pr(> t ) (p値)
切片 ( $b_0$ )	106146	24196	4.39	0.000***
年齢 ( $b_1$ )	841	382	2.21	0.028*
家族人数 ( $b_2$ )	23170	2602	8.91	0.000***
高齢者の有無 ( $b_3$ )	-1063	8202	-0.13	0.897
子供の有無 ( $b_4$ )	7941	7633	1.04	0.299
家からの時間 ( $b_5$ )	-3208	598	-5.37	0.000**
Adjusted R-squared (自由度調整済み決定係数)	0.106			

### 赤字部分の年齢の解釈

- 「 $b_2, \dots, b_5$  の影響を取り除いた場合、年齢 ( $b_1$ ) が1歳上がることに購買金額が841円大きくなる」
- 「 $b_1, b_3, b_4, b_5$  の影響を取り除いた場合、家族の人数が1人増えることに購買金額が23,170円大きくなる」
- 「 $b_1, \dots, b_4$  の影響を取り除いた場合、家からの所要時間 ( $b_5$ ) が1分増えることに購買金額が3,208円小さくなる」

6

## 線形回帰モデルで実証分析を行うときの注意

### 偏回帰係数の仮説検定が成り立つための条件

- 仮定1: 説明変数は確率変数ではなく定数である
- 仮定2: 説明変数間に多重共線性はない
- 仮定3: 誤差項  $e$  は平均0, 分散  $\sigma^2$  に従う確率変数であり,  $\{e_1, \dots, e_N\}$  は互いに独立である ( $V[e] = \sigma^2 I$ )
- 仮定4: 誤差項  $e$  は正規分布  $N(0, \sigma^2)$  に従う
- 仮定N1: (仮定1を満たさない場合) 誤差項  $e$  はすべての説明変数と互いに独立

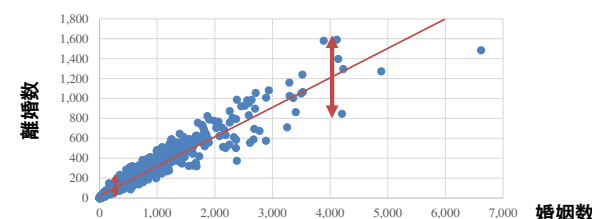
どんなデータに対しても、前述のような偏回帰係数の仮説検定ができるとは限らない

7

## 不均一分散 (仮定3の分散 $\sigma^2$ の均一性を満たさない例)

### サンプル $i$ 毎に誤差項 $e_i$ の分散が異なる事象

- 例: 所得と消費額, 失業者数と犯罪発生率などの関係では、所得や都市の規模が大きいほど分散が大きくなる傾向



総務省統計局: 統計でみる市区町村のすがた2015より1877市町村の婚姻と離婚数

$x \leq 300$  付近と  $x \geq 2000$  付近では分散が異なる

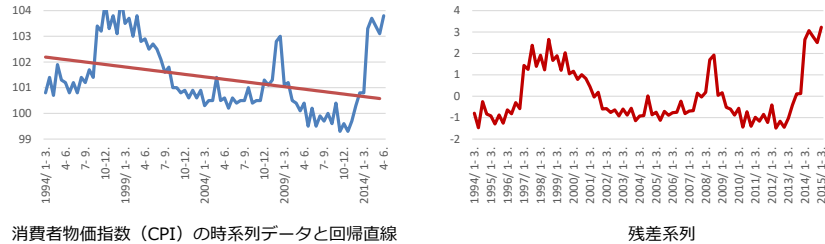
- 対応方法: 加重最小2乗推定や一般化最小2乗推定を学習

8

# 系列相関 (仮定3の誤差項間の独立性を満たさない例)

## 異なる点の誤差項 $e_i, e_j$ に相関がある事象

- 例: 時系列など過去の値が現在の値に影響を与えるデータ  
 今期の政策の効果は次期以降に徐々に現れる



$e_i$  が大きければ  $e_{i+1}$  も大きい傾向

- 対応方法: 時系列分析を学習

# 内生性 (仮定N1を満たさない例) #1

## 説明変数と誤差項の間に相関がある事象

### 内生性の例1: 観測できない説明変数

- $v_i$ : サンプル  $i$  に対して観測できない属性変数
- 例: 学歴  $x_i$  から給与  $y_i$  を説明する回帰モデル。 $v_i$  は個人  $i$  の観測できない能力で学歴と給与の両方と相関がある

より正確なモデル 我々が作成できるモデル

$$y_i = b_0 + b_1x_i + b_2v_i + e_i \qquad y_i = b_0 + b_1x_i + v_i e_i$$

- $x_i$  が大きいと  $v_i e_i$  も大きい傾向 これを誤差項  $e_i$  として観測してしまう

# 内生性 (仮定N1を満たさない例) #2

## 内生性の例2: 目的変数が説明変数に影響を与えている

例: 従軍経験とその後の賃金は関係あるのか? (Angrist 1990)  
 満足できる職に就けなかった人、つまり、元々賃金が低くなる傾向にある人たちが軍隊に入る傾向があるのでは?

例: 警察予算を増額すると犯罪件数は減るか? (Levitt 1997)  
 前年度の犯罪発生数に基づき予算を決める

$$\text{賃金} = a + b_1 \text{軍隊経験} + \dots + b_p x_p + e$$

- 内生性(同時方程式)バイアス
- 対策方法: 操作変数法などを学習



J. D. Angrist, Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records, The American Economic Review, Vol. 80, No. 3, pp. 313-336, (1990)  
 S.D. Levitt, Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime, Ameri. Econ. Review, 87(3), 270-90 (1997)

# マーケティング・リサーチと回帰分析

## 線形回帰分析は実証分析や予測の手段

- 実証分析: 現状の定量的な理解  
 「現状が分かる!」と「現状が分かるだけ...」の違いは大きい
- 予測: 定量的な将来の推計

線形回帰分析の結果に基づいて、経営戦略の方針・マーケティング理論・法律や制度を参照し「何をすべきか?」を考える

- 年齢: 高齢層の維持? 非高齢層の購買意欲喚起?
- 家族人数: 家族人数と購買金額の比例関係は当たり前? それとも他の要因?
- 家からの時間: コンビニ化? 無料宅配サービス?