

(再掲) 重回帰分析をしてみよう～結果

• 重回帰分析の結果

	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)	0.11			

– ソフトウェアを利用することで、このような結果が出力される

• これ以降の本授業の目標

⇒ この表の数値の意味を正しく解釈できる

1

2

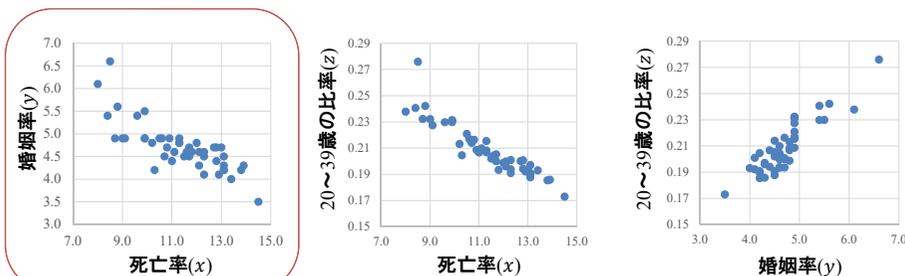
経済と社会 補助資料 ～重回帰分析の結果の解釈2～

2022年度1学期: 火曜2限
担当教員: 石垣 司

偏相関係数～人口動態の相関係数

• 各都道府県の人口動態データ(2015年)

- 1000人当たりの死亡率(x), 婚姻率(y)
⇒ 相関係数 $r_{xy} = -0.80$ (その解釈は?)
- 20～39歳の比率(z)
⇒ 相関係数 $r_{xz} = -0.92, r_{yz} = 0.88$



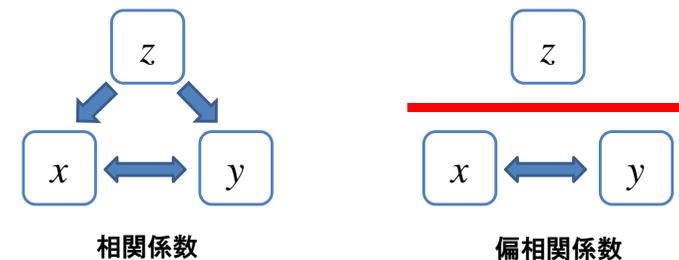
3

偏相関係数

• 3つの変数 (x, y, z) から1つの変数の影響を取り除いた, 残り2つの変数の相関係数

– z が x と y の共通の説明要因であると, x と y の間に疑似相関が発生

• 回帰により変数 z の影響を x と y からそれぞれ除去し, その残差間の相関

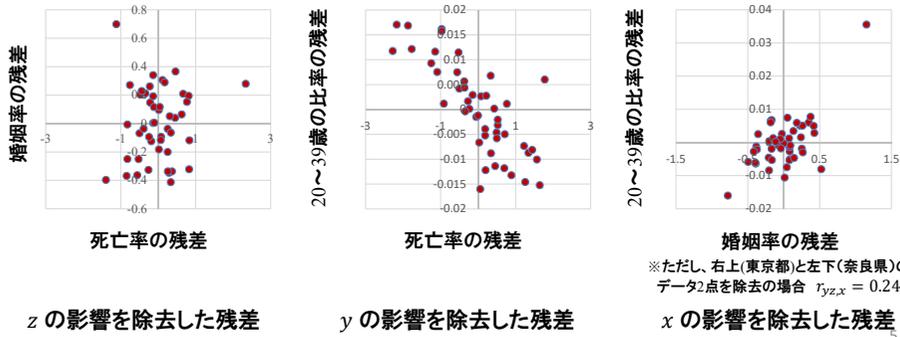


4

偏相関係数～人口動態の偏相関係数

偏相関係数

- 偏相関係数 $r_{xy,z} = 0.09$
 - 20～39歳の比率(z)が同じときの死亡率(x), 婚姻率(y)の相関係数
- 偏相関係数 $r_{yz,x} = 0.62, r_{xz,y} = -0.77$



重回帰係数の推定量 \hat{b}_p の解釈

- 重回帰モデルの最小2乗推定量 $\{\hat{b}_p\}_{p=1,\dots,P}$ は偏回帰係数と呼ばれる
- 偏相関係数と同様に, p 番目の説明変数以外の影響を取り除いた場合の, 目的変数 y と説明変数 x_p の単回帰係数に等しい

※その理由の説明は, 本講義の内容に合致しないため説明は省略。詳しく知りたい人は計量経済学の教科書を参照

重回帰分析の結果の解釈

	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)	0.11			

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」

重回帰分析の結果の解釈(総合)

	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)	0.11			

- 購買金額には年齢, 家族の人数, 家からの所要時間が影響を与えている

関数の当てはまりには改善の余地がある

- 当然, 購買は今回利用した顧客属性以外の要因に基づいて変化しうる。購買金額を予測するためには, よりよい説明変数の追加などのモデルの改善が必要である
- 一方, 予測が目的ではない場合, 決定係数は重要視する必要はない

マーケティング・リサーチと重回帰分析

• 重回帰分析は実証分析や予測の手段

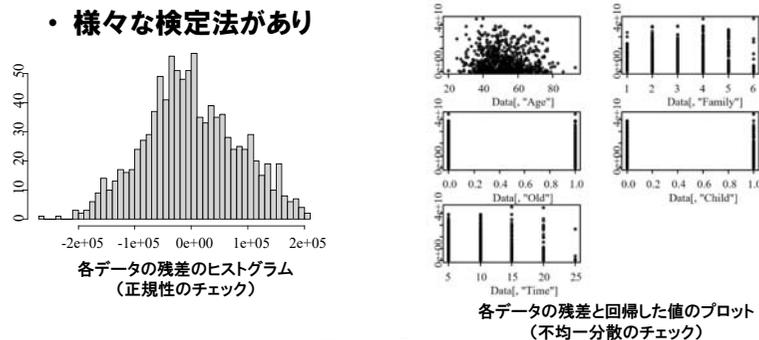
- 実証分析: 現状の定量的な理解
 - 「現状が分かる!」と「現状が分かるだけ...」の違いは大きい
- 予測: 定量的な将来の推計
 - 構造が変化しない場合という条件の理解は必須
- その結果に基づいて, 経営戦略の方針・マーケティング理論・法律や制度を参照し「何をすべきか?」を考える
 - 年齢: 高齢層の維持? 非高齢層の購買意欲喚起?
 - 家族人数: 家族人数と購買金額の比例関係は当たり前? それとも他の要因?
 - 家からの時間: コンビニ化? 無料宅配サービス?

回帰係数の検定を行うときの注意

- 次の3つの仮定を満たすとき, $\hat{b}_1 \sim N\left(b_1, \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2}\right)$
 1. 説明変数 x_1 は非確率的
 2. 誤差項は正規分布: $e_i \sim N(0, \sigma^2)$
 - 平均はゼロで不均一分散
 3. 誤差項は独立: $\text{Cov}[e_i, e_j] = 0$
- これらの仮定を満たさない場合, 線形重回帰分析の回帰係数の検定の結果は成り立たない

本重回帰分析での検証

1. 説明変数 x_1 は非確率的
 - 本講義の内容を超えるので, 詳細は省略
2. 誤差項は正規分布: $e_i \sim N(0, \sigma^2)$
 - 様々な検定法があり



3. 誤差項は独立: $\text{Cov}[e_i, e_j] = 0$
 - 顧客 i の購買が顧客 j の購買に影響を与えるとは考えにくい

内生性 (線形重回帰分析の回帰係数の検定を単純に適用できない例)

- 軍隊経験とその後の賃金は関係あるのか? (Angrist 1990)
 - 軍隊経験者と非経験者の賃金を比べる \Rightarrow 不十分
 - そもそも, 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?

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

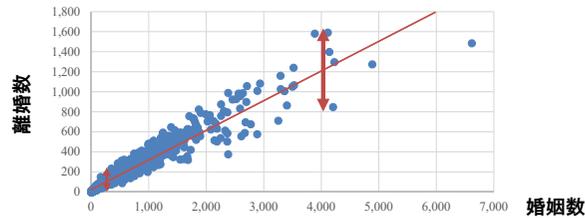
? (疑問符) with arrows pointing to the equation

- 内生性(同時方程式)バイアスの結果
 - 最小2乗推定量は不偏性も一貫性も持たない
 - 操作変数法などを学習
 - ※ 内生性の影響を除いた推定結果: ベトナム戦争(米国)では, 軍隊経験者は平均賃金が15%低い

不均一分散 (線形重回帰分析の回帰係数の検定を単純に適用できない例)

- 「所得と消費額」、「失業者数と犯罪発生率」に関係はあるか？

– 所得や都市の規模が大きいほど分散が大の傾向



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

– 不均一分散の結果

- 最小2乗推定量は不偏性と一致性を持つ
- 分散を過小評価の傾向 ⇒ 係数の検定は誤った結果を出す
- 加重最小2乗法や一般化最小2乗法などを学習

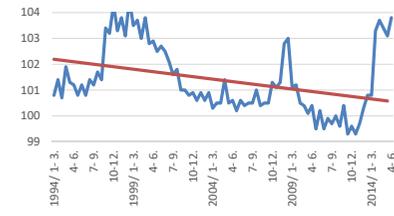
13

系列相関 (線形重回帰分析の回帰係数の検定を単純に適用できない例)

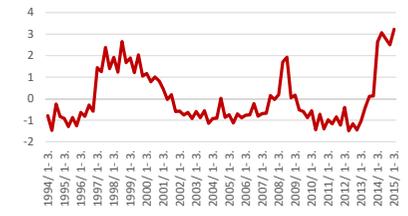
- 時系列データのほとんどは誤差項は非独立

– 過去の値が現在の値に影響を与える

- 例 今期の政策の効果は次期以降に徐々に現れる



消費者物価指数(CPI)の時系列データと回帰直線



残差系列

– 系列相関の結果

- 最小2乗推定量は不偏性と一致性を持つ
- 分散を過小評価の傾向 ⇒ 係数の検定は誤った結果を出す
- 自己相関を分析対象とする時系列分析を学習

14