

重回帰分析

2008/05/07

M1 栗田季佳

回帰分析とは

- 1つ or 複数の独立変数から1つの従属変数を予測・説明するための分析法

単回帰分析・・・独立変数が1つ

重回帰分析・・・独立変数が複数

独立変数



従属変数

重回帰分析の予測式

$$\text{予測値} = b \times (\text{独立変数}) + b \times (\text{独立変数}) \\ + \dots + b_i \times (\text{独立変数}_i) + c$$

$b \dots$ 偏回帰係数

$c \dots$ 切片

予測値は従属変数ではない。

あくまで独立変数に基づいて予測した値

⇒ 誤差(残差)が生じる

重回帰モデル(1)

残差 = 従属変数 - 予測値

従属変数 = 予測値 + 残差

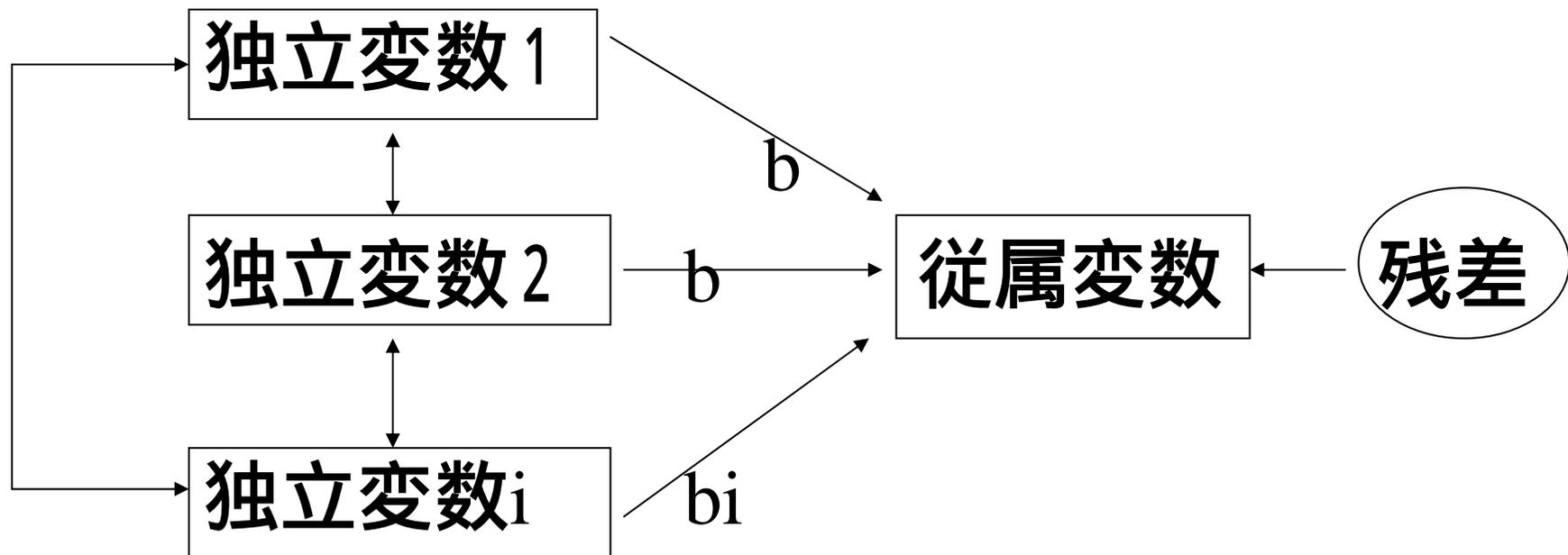
 予測式を代入

$$\begin{aligned} \text{従属変数} = & b \times (\text{独立変数}) + b \times (\text{独立変数}) \\ & + \dots + b_i \times (\text{独立変数}_i) + c + \text{残差} \end{aligned}$$

重回帰モデル

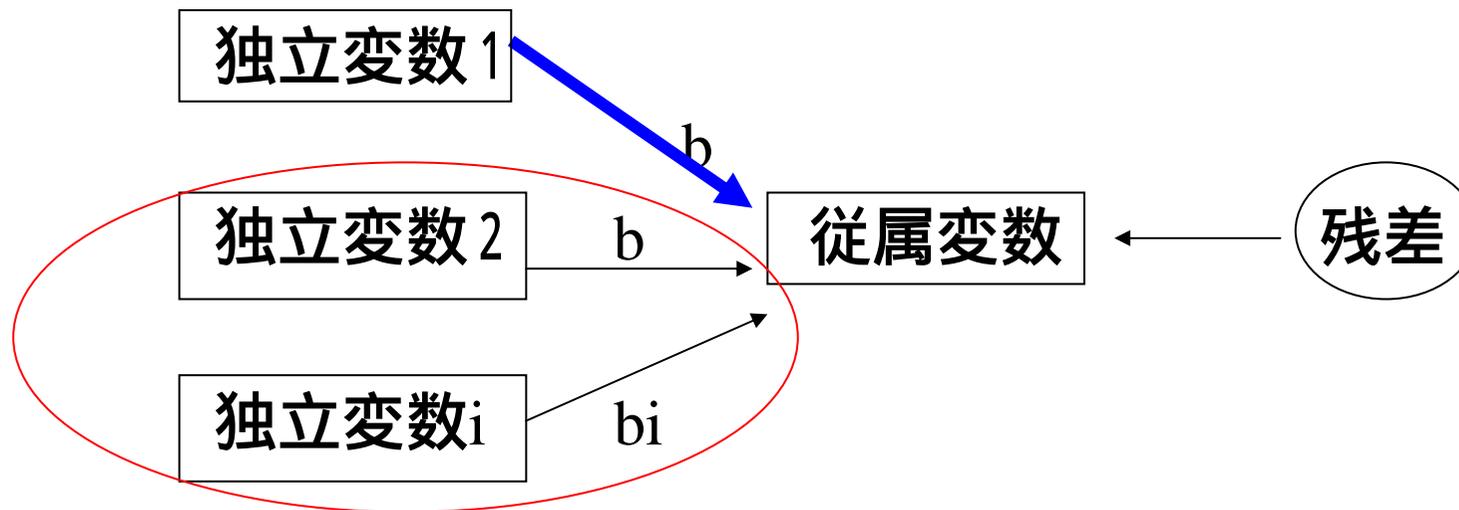
重回帰モデル(2)

- 重回帰モデルをパス図で表す



偏回帰係数 (B)

- 偏回帰係数は、他の独立変数の影響を除いたときに、ある独立変数の値が「1」だけ変わったとき、従属変数の予測値がどれだけ変化するかを表している



影響を取り除く

標準偏回帰係数 ()

- 偏回帰係数は、独立変数や従属変数の単位に依存する
- 単位に依存しない偏回帰係数を、**標準偏回帰係数**という
- **標準偏回帰係数** = 偏回帰係数 × (ある独立変数のSD / 従属変数のSD)

偏回帰係数の区間推定

- 区間推定・・・実際にはわからない母集団の真の値がどこからどこまでの範囲にあるかを推定すること
- 信頼区間・・・母集団が信頼できる確率でその範囲に入っている区間
(大抵95%か99%が使われる)

偏回帰係数は、区間推定を行って求めている

最小二乗法

- 残差の絶対値は小さいことが望ましい
⇒ できるだけ小さくする b, c の値を求める

最小二乗法・・・残差の二乗和を最小にする係数・切片の値を求める方法

予測の精度を測る その1(1)

- 最小二乗法によって重回帰分析の解が得られたら、その予測式がどのくらい正確に従属変数を予測できているかを検討する

従属変数の値と予測値が近ければ近いほど、予測の精度が高いといえる

⇒ 残差がどのくらい小さいかを見積もる

- 残差の分散を残差の大きさとして予測の精度を図る

予測の精度を測る その1(2)

□ 従属変数、予測値、残差のそれぞれの分散の関係

□ 従属変数の分散 = 予測値の分散 + 残差の分散

$$\frac{\text{予測値の分散}}{\text{従属変数の分散}} = 1 - \frac{\text{残差の分散}}{\text{従属変数の分散}}$$

決定係数

残差の大きさ

従属変数の分散の内、
何%を予測値の分散が
説明しているか

予測の精度を測る その1(3)

- 決定係数の大きさはどれくらいであればよいか？

$$0 < \text{決定係数} < 1$$

明確な基準はないが、0.5以上の値が目安となる

しかし、結局は研究の目的による

予測の精度を測る その2

- 従属変数と予測値との相関係数を指標とする
重相関係数 (R)
- 重相関係数が1に近づくほど、予測の精度は高いことになる

決定係数と重相関係数の関係

決定係数 = 重相関係数の2乗 (R²乗)

- 独立変数が1つのとき、つまり単回帰分析では、重相関係数は、独立変数と従属変数の相関係数に一致する

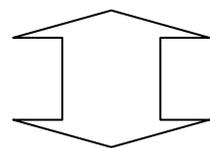
自由度調整済みR²乗

- 独立変数の数が増えると、一般に決定係数の値は大きくなる
- できるだけ少ない独立変数で、決定係数の高いモデルが理想的
 - ⇒ **自由度調整済みR²乗**・・・独立変数の数を考慮した決定係数
- 自由度調整済みR²乗は、予測に必要なのない独立変数が加えられると値が小さくなる
- 検定結果の、R²乗と自由度調整済みR²乗の値が大きく異なっている場合は、予測に必要なのない独立変数が含まれている可能性がある

重回帰分析を使用するときの注意点

- 独立変数、従属変数ともに量的データであること

量的データ・・・数量によって表されたデータ



(例: 孤独感得点)

質的データ・・・性質や内容を表すデータ

(例: 喫煙の有無)



重回帰分析を使用するときの注意点

- データの自由度数が、変数の数よりも十分に多いこと(目安としては変数の10倍以上)
- 重回帰分析の他に、パス解析や因子分析、構造方程式モデリングの分析対象となる場合も同様である

重回帰分析を使用するときの注意点

□ 多重共線性の問題

↳ 独立変数間の相関が強すぎる場合、
偏回帰係数 (標準偏回帰係数) の推
定値が過度に不安定になる

偏回帰係数の区間推定の値が
+ から - の値を含んでしまう

つまり、独立変数間の相関係数を求めなければならない！！

引用・参考文献

- 足立浩平(2007) 多変量データ解析法 心理・教育・社会系のための入門 ナカニシヤ出版
- アイスクリーム統計学によろこそ！ 相関から因子分析まで
<http://kogolab.jp/elearn/icecream/index.html>
- 石井秀宗(2005) 統計分析のここが知りたい 文光堂
- 渡部洋(1988) 心理・教育のための多変量解析入門 基礎編 福村出版
- 栗田季佳(2008) FUMIEテストを用いた障害者に対する態度の調査 平成19年度広島大学教育学部卒業論文