

第8章 回帰分析 後半(p.167～p.180)

■ 8-3 重回帰分析の流れ

➤ 8-3-1 強制投入法

例：「高校1年生120名がこれまで受けた試験の得点から、年度末模試の得点をどの程度予測できるか」

- 従属変数・・・年度末模試の得点
- 独立変数・・・プレイスメントテスト、前期模試、中間模試、後期試験

<操作手順>テキスト参照

➤ 8-3-2 出力結果の見方（強制投入法）

(1) [記述統計]

- ・データが正しく分析されたかの確認
- ・[相関] から多重共線性の疑いがある高い相関(.80以上または.90以上)がないかの確認
- ・独立変数間に高い相関が見られない場合→次の結果へ

(2) [モデル要約]

- ・[R2乗]：決定係数(R^2)
- ・[調整済みR2乗] (Adjusted R^2)
 - ：独立変数の影響力の大きさに関わらず、独立変数の数が増えるほど R^2 は大きくなる傾向。
 - 調整済み R^2 は、この欠点を自由度で調整し、母集団を推定、若干 R^2 より小さい値になる
- ・[Durbin-Watson]：値が2に近いので、各独立変数における残差の独立性に問題ない

(3) [係数]：最も重要な表

- ・[B]、[ベータ]：偏回帰係数および標準化偏回帰件数のこと
 - 各独立変数が従属変数の予測にどの程度寄与しているか
- ※有意な変数だけで重回帰式を作成したい場合
 - ・[プレイスメントテスト]を外して再度、重回帰分析
 - ・ステップワイズ法(8-3-3)
- ・[相関]：[ゼロ次] →ピアソンの相関係数、[偏] →偏相関係数、[部分] →部分相関係数
- ・[共線性の統計量]：多重共線性を診断する2つの指標が示される

(4) [共線性の診断]

- ・[次元]：定数項と独立変数の数
- ・[条件指数]：15以上であれば強い多重共線性、30以上であれば重大な多重共線性
- ・[分散プロパティ] (分散の比率)
 - ：その次元におけるそれぞれの独立変数の関係の強さを示す指標、どの独立変数間で共線性が表れているか診断できる

(5)外れ値の診断1

- ・ [ケースごとの診断] の [標準化残差] で指定した±2以上のケース（データ）が表示される
- ・ 分析から外すかどうかの判断は、他の指標を含めて検討

※問題のあるケースの診断だけを表にまとめるとき

メニュー→ [分析(A)] → [報告書(P)] → [ケースの要約(M)] で調べたい指標を選ぶ

(6)外れ値の診断2

- ・ [残差の統計量] : 外れ値指標
 - ・ [データ(D)] : 指定した指標のケースごとの値が保存されている
- ① [マハラノビス距離] : 説明基準（サンプルサイズが100で独立変数が3つの場合は、カットポイントが15以上）に従って判断
 - ② [Cookの距離] : 基準の1を超えているかどうか
 - ③ [中心化したてこ比の値]
 - : 4つの独立変数、120のサンプルで、平均てこ比= $(k+1)/n=(4+1)/120=0.042$ となり、問題ない→この値の2倍、3倍以上になると問題

(7)残差の正規性 : [標準化された残差の回帰の正規 P-P プロット]

- ・ 残差が正規直線上に並んでいるか、あるいは標準化された残差の [ヒストグラム] が正規分布していれば正規性が成り立つ

(8)残差の等分散性と線形性

- ・ [偏残差プロット] : 各独立変数と従属変数の偏残差分布を示す
- ・ [回帰の標準化された残差] と [回帰の標準化された予測値] の散布図 :
ゼロを中心に水平に分散しており、残差の等分散性および線形性が満たされているか

➤ 8-3-3 ステップワイズ法

「強制投入法」による分析→プレレイズメントテストは年度末模試の予測に適さない

→有効な独立変数のみで、再度、強制投入法で分析

→ステップワイズ法 : 有効な独立変数のみの重回帰式を一度に得ることができる

<操作手順>テキスト参照

➤ 8-3-4 出力結果の見方 (ステップワイズ法)

(1) [投入済み変数または除去された変数]

- ・ 独立変数がモデルに投入された順と、どの時点で除去されたか

(2) [モデル要約]

- ・ 従属変数への予測力が大きい独立変数順に重回帰式に投入されたときの決定係数の推移

(3) [係数]

- ・ 投入された独立変数のステップごとの偏回帰係数 [B]、[標準化係数 (ベータ)]
- ・ i 番目の生徒の予測得点を、以下の重回帰式から得ることができる

(式 8.11) i 番目の年度末試験得点=

$$-4.148+(0.714 \times \text{中間模試 } i)+(0.212 \times \text{後期試験 } i)+(0.179 \times \text{前期試験 } i)$$

➤ 8-3-5 論文への記載

- ・各変数の記述統計および変数間の相関関係を記載
- ・表 8.1 参照
決定係数(R^2)、調整済み決定係数 (調整済み R^2)、
各独立変数の標準化されていない偏回帰係数(B)、標準偏差誤差(SE,B)、標準化偏回帰係数(β)、
t 値(t)および有意確率(p)
- ・記載例 テキスト参照(p.174)

■ 8-4 ダミー変数を使った回帰分析

- 回帰分析では、間隔尺度の量的データだけでなく、性別などの名義尺度も、0 と 1 の 2 値データに変換して使用することができる
- 3 つ以上のカテゴリに分かれている名義尺度も、他の量的データとは比較できないが、ダミー変数(dummy variables)という 0 と 1 から成るデータに変換して分析できる

➤ 8-4-1 2 値の名義尺度

- 操作手順 テキスト参照
- 出力結果の見方
 - (1) [モデル要約]
 - ・決定係数 [R2 乗] : 4 つの独立変数で従属変数を説明
 - (2) [係数]
 - ・[有意確率] : [前期試験] [中間模試] [クラス] が有意に [年度末模試] の予測に寄与している
 - ・名義尺度 [クラス] は、偏回帰係数 [B] が、[普通コース] を基準として [進学コース] になると、どの程度まで得点が変わるか

➤ 8-4-2 ダミー変数の作成

- ・独立変数が 3 つ以上のカテゴリの場合、カテゴリ数より 1 つ少ないダミー変数を作成する
- 操作手順 テキスト参照

➤ 8-4-3 ダミー変数を含んだ階層的回帰分析

- ・作成した 2 つのダミー変数は、1 つの独立変数として重回帰分析に同時 (同じブロック) に投入
- ・ステップワイズ法などで削除するときも 2 つの変数を同時に扱う
- ・作成したダミー変数にどの程度説明率があるのかを調べる
- 操作手順 テキスト参照
- 出力結果の見方
 - (1) [モデル要約]
 - ・ブロックごとの [R2 乗]、[調整済み R2 乗] : [有意確率 F 変化量] と照らし合わせる

(2) [係数]

- ・ 偏回帰係数 [B] から、以下の式で、年度末模試の得点が求められる

$$(式 8.12) \quad \text{年度末試験の得点} = -4.177 + (7.811 \times \text{理系進学コース} + 3.707 \times \text{文系進学コース}) \\ + 0.672 \times \text{中間模試} + 0.171 \times \text{前期模試} + 0.190 \times \text{後期試験}$$

- ・ ダミー変数を作成することで、基準となる変数との比較が可能

➤ **8-4-4 論文への記載**

- ・ ステップワイズ法による重回帰分析を行った場合と同様、ブロック（ステップ）ごとに報告
- ・ 記載例 テキスト参照(p.180)