

## 第10講 一般線型モデル

田中重人 (東北大学文学部准教授)

[テーマ] 固定因子と共変量

### 1 前回課題について

年齢と Q39G は間隔尺度と考えてよいが、学歴（3段階）は順序尺度。

- 年齢 × Q39G → 相関係数 (Pearson / Spearman / Kendall)
- 学歴 × Q39G → クロス表、平均値の比較、または順位相関係数 (Pearson の積率相関係数は使えない)
- 学歴 × 年齢 → クロス表 (年齢を適当にカテゴリにわけると)、平均値の比較、または順位相関係数 (Pearson の積率相関係数は使えない)

### 2 モデルとパラメータ

前回の一般線形モデルの推定結果では、Q39g の値が次の式で近似されていることになる：

$$Q39g = \text{切片} + B_1X_1 + B_2X_2 + B_3X_3 \quad (1)$$

ただし、

- $X_1$  は年齢
- $X_2$  は初等教育のものについて1、それ以外は0とする
- $X_3$  は中教育のものについて1、それ以外は0とする

推定された係数 (切片と  $B$ ) それぞれについて、区間推定と統計的検定がおこなわれる

### 3 固定因子と共変量

**固定因子:** 名義尺度の変数。自動的にカテゴリに分割され、そのうちひとつが「基準」になる。推定される係数は、カテゴリ数 - 1。

**共変量:** 間隔尺度の変数。そのままの値が投入される。推定される係数はひとつだけ。

### 3.1 固定因子ひとつだけのモデル

カテゴリ別平均から係数が計算される

初等:  $3.591 - 0.700 = 2.891$

中等:  $3.591 + 0.011 = 3.602$

高等:  $3.591 + 0.000 = 3.591 \leftarrow$  基準

「分散分析表」に表示されるものは、平均値の比較の際に使われるものと同等であるが、用語が少し違う：

- 決定係数  $R^2 =$  相関比  $\eta$  の 2 乗 = edu3 / 修正総和
- edu3 + 誤差 = 修正総和

おなじ変数について平均値の比較をおこない、結果を照らし合わせてみよう。

### 3.2 共変量ひとつのモデル

最小 2 乗法 (least square method) で係数を求める。これは、適当な直線  $A + BX$  によって  $Y$  の値を近似する方法であり、 $Y$  と  $A+BX$  とのずれの大きさを評価するために差の 2 乗和をとる。この 2 乗和  $\sum(Y - A - BX)^2$  が最小になるように  $A$  と  $B$  の組み合わせを求める。

回帰係数  $B$  の意味:  $X$  が 1 単位増えたとき  $Y$  がどれだけ増えるか

### 3.3 独立変数が複数の場合

- 独立変数がひとつの場合と何が異なるか?
- 「コントロール」することの意味
  - 媒介効果
  - 疑似相関
- 分散分析表から独立変数の影響力の大きさを読む

## 4 期末レポート

期限: 2/5 (金)

提出先: ISTU

内容: 相関係数、対応のある分析、多変量解析について、それぞれ適当な分析をして結果を解釈する。すべての分析について、推定または検定結果をつける。データは何を使ってもよいが、SSM データ以外のものを使うときはデータについての説明をつけること。

備考: レポート提出後に、SSM データのコピーをすべて消去すること。レポートは、採点後に返却する。

## 文献

大野高裕 (1998) 『多変量解析入門』同友館。

吉川徹・轟亮 (1996) 「学校教育と戦後日本の社会意識の民主化」『教育社会学研究』58:87-101.

三土修平 (1997) 『初歩からの多変量統計』日本評論社。