

統計学2及び演習

線形回帰分析の考え方



創域理工学部

Faculty of Science and Technology

東京理科大学
創域理工学部情報計算科学科
安藤宗司

2023年6月28日

Contents

- 回帰分析とは
- 相関関係と因果関係
- 相関係数の性質

多変量解析

- 多変量間の関係を解析するための統計手法
 - 重回帰分析, 判別分析, 主成分分析など
- 回帰分析
 - 多変量解析のなかで最も基本的で, 広く実用されている統計手法の一つ
 - 多変量間の因果関係や相関関係を解析

回帰分析の恩恵

□ 病気の診断

□ 予後の予測

□ 天気予報

□ 選挙予測

2変量間の関係を考える

- 身長と体重
- 血圧と年齢
- 所得と年齢
- （体積一定の気体の） 温度と圧力

「関係」の中身

- (体積一定の気体の) 温度と圧力
 - ボイル・シャルルの法則から、圧力と温度は比例する

- 身長と体重
 - 「身長が高い人は体重も重い」という「傾向」はある
 - 「血圧と年齢」, 「所得と年齢」も同様な関係

- 身長が高くなれば、体重も重くなる (傾向がある)
- ✕ 身長を高くすると、体重も重くなる

相関関係と因果関係

□ 相関関係

- 二つ以上の変量が
「規則正しさをもって、増減をともにする関係」

□ 因果関係

- 「一方の変量が他方の変量を決定する関係」

□ 2変量間に相関関係があるからといって因果関係があるというわけではない

見せかけの相関

- 「血圧と所得に正の相関がある」という命題
 - 統計的に真になる可能性が高い
 - 「年齢」を介して、「血圧と所得」に見せかけの正の相関
 - 「血圧と所得」には直接的な相関関係は認められない



- 二つの変量が相関している
 - 二つの変量間に何かしらの「関係」が存在していることを示唆しているにすぎない

統計学での相関関係

- 「線形な共変関係」という限定的な意味に使われることが多い
- 「関係はあっても相関はない」というケースもあることに注意する必要がある

相関係数

□ 2変量 X と Y に関する n 個の観測値

$$(x_i, y_i) \quad (i = 1, \dots, n)$$

□ 2変量間の「関係」を記述するための基本的な統計量

$$r = \frac{s_{xy}}{s_x s_y}$$

$$(-1 \leq r \leq 1)$$

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

具体例

| x | y_1 | y_2 | y_3 |
|-----|-------|-------|-------|
| 10 | 8.04 | 9.14 | 7.46 |
| 8 | 6.95 | 8.14 | 6.77 |
| 13 | 7.58 | 8.74 | 12.74 |
| 9 | 8.81 | 8.77 | 7.11 |
| 11 | 8.33 | 9.26 | 7.81 |
| 14 | 9.96 | 8.10 | 8.84 |
| 6 | 7.24 | 6.13 | 6.08 |
| 4 | 4.26 | 3.10 | 5.39 |
| 12 | 10.84 | 9.13 | 8.15 |
| 7 | 4.82 | 7.26 | 6.42 |
| 5 | 5.68 | 4.74 | 5.73 |

2変量間の関係を考えてみましょう！

- 横軸に x , 縦軸に y_1, y_2, y_3 とした散布図
- x の y_1, y_2, y_3 への回帰分析を実施

問

□ 相関係数

$$r = \frac{s_{xy}}{s_x s_y}$$

□ $-1 \leq r \leq 1$ となることを証明せよ

□ $|r| = 1$ となるとき, 2変量 X と Y にどのような関係があるかを述べよ