

応用統計学

竹村彰通教員

2007/02

以下の4問にすべて答えること。なお問1については別紙の資料を参照のこと。

1. 別紙に示す中古マンション価格に関する回帰分析の計算結果について、どのように結果を解釈するか説明せよ。
2. $n \times p$ の素得点行列 X の各要素をすべて基準化した「標準化変数行列」を Z とする。すなわち Z の要素は、 X の各要素を標本平均を引き標準偏差で割った値でおきかえたものである。 Z の分散共分散行列 $Z'Z$ は X の相関行列 R に一致することを示せ。 R に基づく主成分分析に対応して、 R のスペクトル分解を $R = ADA'$ とおく。ただし $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, $\lambda_1 \geq \dots \geq \lambda_p > 0$ とおく。 $Y = ZA$ の要素は主成分得点を表す。ここまでの記述について説明せよ。

また X の第 i 変数 (列) と Y の第 j 変数 (列) の間の相関係数が $\sqrt{\lambda_j} a_{ij}$ と表されることを示せ。*1

3. クラスの41人の血液型を調べたところ、A:13人、O:14人、B:9人、AB:5人となった。昨年のクラスでは、24人の血液型を調べたところ、A:8人、O:7人、B:6人、AB:3人となった。昨年のクラスと今年のクラスは、同じ血液型の母集団からの標本と考えてよいか、の検定について考察せよ。その際並べ替え検定の考え方に重点をおいて説明せよ。また 2×4 の分割表の独立性の検定との関連を説明せよ。*2
4. 2元配置分散分析について説明し、平方和の分解：

$$W_T = W_A + W_B + W_{AB} + W_E$$

を証明せよ。また交互作用項 W_{AB} の解釈を説明せよ。ただし

$$W_T = \sum_{i,j,k} (x_{ijk} - \bar{x} \dots)^2, \quad W_A = br \sum_i (\bar{x}_{i..} - \bar{x} \dots)^2, \quad W_B = ar \sum_j (\bar{x}_{.j.} - \bar{x} \dots)^2,$$
$$W_{AB} = r \sum_{i,j} (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x} \dots), \quad W_E = \sum_{i,j,k} (x_{ijk} - \bar{x}_{ij.})^2$$

(a, b はそれぞれ要因 A, B の水準数, r は各水準の組での共通の繰り返し数。)

*1 ヒント: A の第 j 列を a_j とし、これが $R = Z'Z$ の固有ベクトルであることを用いて、 $Z'a_j$ の要素および $\|Z'a_j\|_H$ を考察せよ。

*2 一つの考え方: 壺の中に A と書いた玉を $13 + 8 = 21$ 個入れる。 O, B, AB についても同様で計 65 個入れる。そこから 41 人を無作為に抜き出す。

講義のホームページにある中古マンションデータについて、統計解析ソフト R を用いて重回帰分析を行った。目的変数を price, 説明変数を time, bus, walk, area, bal, kosuu, floor, tf, muki, year, build とした。ただし muki は S(南) を 1 とし, N(北) を -1 として 0.5 刻みで評価した。year は建築年を値とし build は (2001-year) の値を用いた。

まずは, time, bus, walk, area, bal, kosuu, floor, tf, muki, year の 10 変数で関数 lm() を用いて重回帰分析を行った。主要な結果を以下に示す。

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3266.8335   883.6338  -3.697 0.000296 ***
time         -62.7444     6.2181 -10.091 < 2e-16 ***
bus          -68.5791    21.0347  -3.260 0.001349 **
walk         -42.9814    19.6449  -2.188 0.030065 *
area          72.0401     3.3970  21.207 < 2e-16 ***
bal          -32.8128    14.1768  -2.315 0.021856 *
kosuu         0.3507     0.4420   0.793 0.428659
floor        31.5028    41.9712   0.751 0.453961
tf           -52.0170    35.0519  -1.484 0.139694
muki         -10.4029   212.3447  -0.049 0.960985
year          44.3902     9.3415   4.752 4.32e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1096 on 167 degrees of freedom
Multiple R-Squared:  0.8352,    Adjusted R-squared:  0.8254
F-statistic: 84.67 on 10 and 167 DF,  p-value: < 2.2e-16

```

図 1

重回帰分析の結果について plot したものが、次のグラフである。

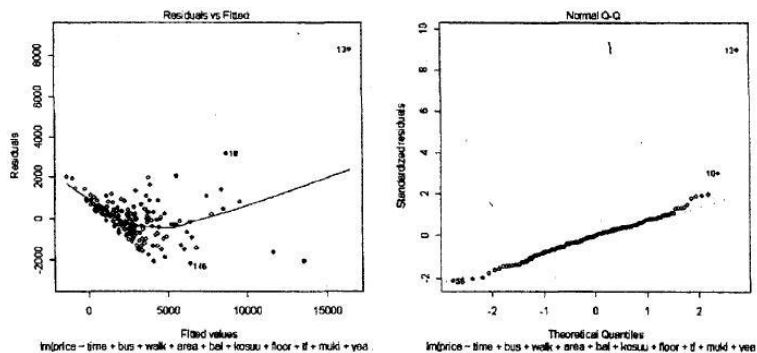


図 2 10 変数での分析 (全データ)