

第二十一章 多項式回帰(R で)

重回帰の特殊な場合、 $Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \dots + e$ は多項式回帰と呼ばれる。多項式回帰の場合も R を使えば同様に解析できる。

例題 21.1 X は河口からの距離(km)で、 Y は河川の水の鉄濃度($\mu\text{g}/\ell$)を調べた値であるⁱ。講義用の Web ページからダウンロードしておくⁱⁱ。このデータに $Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \beta_4 X_1^4$ という多項式回帰モデル(4次なのでモデル4としておく)を当てはめてみる。

以下で、R を操作するコードは赤字(ファイル名や変数名などユーザーが変更する箇所は緑で網掛け)で、R の出力結果は青字で示す。

```
➤ d <- read.table("Ex21.1.txt", header=T) #読み込み
➤ model4 <- lm(Y ~ X + I(X^2) + I(X^3) + I(X^4), data = d)
➤ summary(model4)
```

```
Call:
lm(formula = Y ~ X + I(X^2) + I(X^3) + I(X^4), data = d)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.45159 -0.24518 -0.00307  0.14294  0.70094
```

```
Coefficients:
            Estimate Std. Error  t value Pr(>|t|)
(Intercept)  6.9265      7.2855   0.951   0.35787
X            55.8348     12.4946  4.469   0.00053 ***
I(X^2)      -31.4866     7.6054  -4.140   0.00100 **
I(X^3)       7.7625     1.9573   3.966   0.00141 **
I(X^4)      -0.6751     0.1808  -3.735   0.00222 **
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.3413 on 14 degrees of freedom Multiple R-squared:  0.9909,
Adjusted R-squared:  0.9883 F-statistic: 380.6 on 4 and 14 DF,  p-value: 4.141e-14
```

結果より、4 次多項式回帰モデルは

$$Y = 6.9265 + 55.8348X - 31.4866X^2 + 7.7625X^3 - 0.6751X^4 \quad (\text{式 1})$$

となった。下段にある Adjusted R-squared (自由度調整済み R² 値) : 0.9883 はこの重回帰モデルの当てはまりが非常に良いことを示している。多項式回帰も MuMIn パッケージⁱⁱⁱを使ったモデル選択法を行うことができ、

```
➤ library(MuMIn)
```

Ex21.1	
X(km)	Y($\mu\text{g}/\text{L}$)
1.22	40.9
1.34	41.8
1.51	42.4
1.66	43
1.72	43.4
1.93	43.9
2.14	44.3
2.39	44.7
2.51	45
2.78	45.1
2.97	45.4
3.17	46.2
3.32	47
3.5	48.6
3.53	49
3.85	49.7
3.95	50
4.11	50.8
4.18	51.1

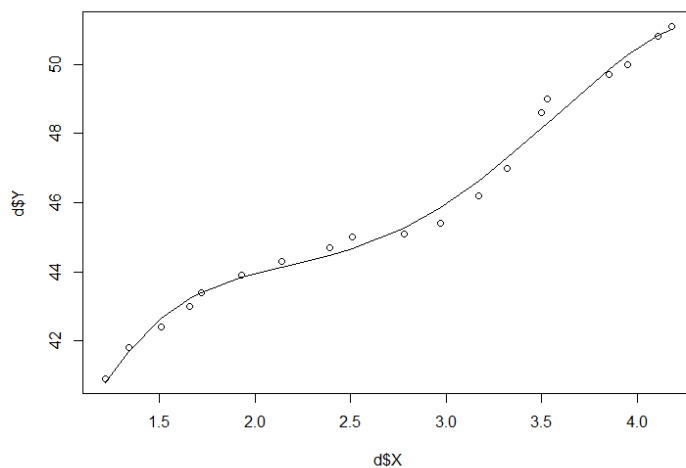
- `options(na.action = "na.fail")` #オプション設定
- `dredge(model4, rank="AIC")`

```
Fixed term is "(Intercept)"
Global model call: lm(formula = Y ~ X + I(X^2) + I(X^3) + I(X^4), data = d)
---
Model selection table
  (Inter) X      X^2      X^3      X^4 df logLik  AIC delta weight
16  6.927 55.8300 -31.4900  7.76200 -0.6751000  6  -3.631  19.3  0.00  0.989
  8 32.770 10.4100 -3.3870  0.47010          5 -10.198  30.4 11.13  0.004
12 34.890  6.9460 -1.4160          0.0400900  5 -10.785  31.6 12.31  0.002
14 36.610  4.3020          -0.31630  0.0652900  5 -11.226  32.5 13.19  0.001
10 39.150  2.0990          0.0107000  4 -12.460  32.9 13.66  0.001
  3 41.110          0.5745          0.05932  4 -13.929  33.9 14.59  0.001
  6 39.550  1.7300          0.05932  4 -12.935  33.9 14.61  0.001
15 39.340          2.3720 -0.86710  0.1113000  5 -12.052  34.1 14.84  0.001
  4 40.300  0.6666  0.4540          0.1113000  4 -13.618  35.2 15.97  0.000
  7 41.000          0.6262 -0.01168          0.1113000  4 -13.887  35.8 16.51  0.000
11 41.070          0.5873          -0.0006954  4 -13.920  35.8 16.58  0.000
13 41.700          0.27370 -0.0347900  4 -15.111  38.2 18.96  0.000
  2 37.390  3.1270          0.12780  3 -17.229  40.5 21.20  0.000
  5 42.360          0.12780  0.0297800  3 -18.611  43.2 23.96  0.000
  9 43.010          0.0297800  3 -24.008  54.0 34.75  0.000
  1 45.910          2 -48.263 100.5 81.26  0.000
Model ranked by AIC(w)
```

4次までの多項式モデルの中で式1が最もAICの低い回帰モデルであることが分かる。

推定された多項式回帰モデルの当てはまり具合を図示すると、下のようになる。

- `P <- predict(model4)`
- `plot(d$Y ~ d$X)`
- `reorder <- order(d$X)`
- `lines(d$X[reorder], P[reorder])`



高次の多項式回帰モデル式は1次線形(=直線)ではないが、残差は最小二乗法で線形モデルとして推定されている。

ⁱ Zar JH. 『Biostatistical Analysis.(5版)』 (2009) Prentice Hall の p459

ⁱⁱ <http://www.eonet.ne.jp/~vor-dem-gesetz/Ex21.1.txt>

ⁱⁱⁱ パッケージのインストール法は <http://www.eonet.ne.jp/~vor-dem-gesetz/Rintro.pdf> を参照