工学ためのデータサイエンス入門
ーフリーな統計環境 R を用いたデータ解析ー

第 4 版用正誤表 (2007.03.21 現在)：初版、第 2 版、第 3 版用正誤表は、修正しておりませんので、各版の正誤表をお使いの方は、本正誤表を追加してください。

十分注意したつもりでしたが、それでも既に多くの誤植、ミスが見付かっております。以下に現在判明している箇所をメモしておきますので、ご参考にして頂ければ幸いです。読者の皆様に著者一同お詫びを申し上げますともに、もし更に間違いと思われる点に気づかれた場合は、ご面倒でも `mase@is.titech.ac.jp` 宛にお知らせくだされば幸いです。なお、今回の修正は主として石岡恒憲様による御指摘を元にしております。お礼を申し上げます。

12 頁下から 11 行	『間欠泉の噴出時間間隔』は『間欠泉の噴出継続時間 (分単位)』です。
30 頁 18 行	『$1-F(-x)+F(x-0)=P\{\|X\|\leq x\}$』は『$1-F(x-0)+F(-x)=P\{\|X\|\leq x\}$』です。
76 頁 3 行	『パラメータ n』は『パラメータ m』です
106 頁	線形単回帰モデルの正規方程式において 2 つの式において $\sum_{i=1}^{n}$ の前に係数 2 が抜けています。
114 頁 3 行	『上側 $100(p/2)\%$ 点』は『上側 $100((1-p)/2)\%$ 点』です。
119 頁 3 行と 4 行	『区間 [0.3,0.6]』は『区間 [0.2,0.8]』です。『$\alpha+\beta x_0$』は『$\alpha x_0+\beta$』です。『$\hat{\alpha}+\hat{\beta}x_0$』は『$\hat{\alpha}x_0+\hat{\beta}$』です。
130 頁 9 行	『 $AIC=2p+\frac{n}{2}+\frac{n}{2}\log(2\pi\widehat{\sigma}^2)=2p+\frac{n}{2}\log(\widehat{\sigma}^2)+$ 定数項』は『$AIC=2(p+1)+n\log\widehat{\sigma}^2+n+n\log(2\pi)$』です。
130 頁下から 2 行	『パラメータ数に σ^2 の分が加算されていないが，AIC 統計量は相対的な大小のみが意味を持つため，誤差の構造が同一である限り，無視してもよい.』は『パラメータ数が p ではなく $p+1$ となっているのは，誤差分散パラメータ　σ^2 を加えているからである.』です。
137 頁下から 3 行	『 $\widehat{\boldsymbol{\epsilon}}=(I-H)\boldsymbol{\epsilon}$』は『$\widehat{\boldsymbol{\epsilon}}=(I-H)y$』です。
139 頁下から 9 行	『 分散が小さの』は『分散が小さい』です。
145 頁下から 11 行	『各データの分散の自乗を重みとした 』は『各データの分散の自乗の逆数を重みとした』です。

157 頁下から 8 行	『 は自由度 $(s-1)(t-1)$ のカイ自乗分布に従うことが知られており， $$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{(s-1)(t-1)}\sum_i\sum_j \widehat{\epsilon}_{ij}^2 \\ &= \frac{\sigma^2}{(s-1)(t-1)}\chi_e^2\end{aligned}$$ が σ^2 の不偏推定量となる.』は『は自由度 $(s-1)(t-1)$ のカイ自乗分布に従うことが知られており，カイ自乗分布の期待値はその自由度であるから，χ_e^2 に $\sigma^2/\{(s-1)(t-1)\}$ をかけると σ^2 の不偏推定量 $$\hat{\sigma}^2 = \frac{\sigma^2}{(s-1)(t-1)}\chi_e^2 = \frac{1}{(s-1)(t-1)}\sum_i\sum_j \hat{\epsilon}_{ij}^2$$ を得る.』です。
184 頁 1 行	『$\log_{10}(V) = \alpha - t\beta\exp(-\gamma/T) + \epsilon$』は『$\log_{10}(V) = \alpha - t\beta\exp(-\gamma/(T + 273.16)) + \epsilon$』です。
184 頁 10 行	『R なら (42):最適解への収束状況の出力』を以下に変更する。

```
> fm <- nls(log10(y) ~ a - b*x1*exp(-c/(x2+273.16)), data=Nelson,
            start=c(a=1.13,b=6.375e+11,c=17065),trace=T)
0.7269386 : 1.1300e+00   6.3750e+11   1.7065e+04   # 初期値
0.722269  : 1.125919e+00 2.878341e+11 1.664299e+04 # 以下途中解
0.7220003 : 1.125479e+00 2.637889e+11 1.659572e+04
0.7213283 : 1.125151e+00 2.473455e+11 1.655993e+04
0.7206689 : 1.124662e+00 2.242644e+11 1.650580e+04
0.7194283 : 1.124181e+00 2.036685e+11 1.644994e+04
0.719196  : 1.124190e+00 2.037838e+11 1.644781e+04
0.719196  : 1.124190e+00 2.037473e+11 1.644772e+04  # 収束値
```

184 頁下から 3 行	『初期値 (1.13,55589,4091)』は『初期値 (1.13,6.375e+11,17065)』です。
186 頁 14 行	『R なら (44):回帰分析要約 `summary(fm)` の出力』を以下に変更する。

```
Formula: log10(y) ~ a - b * x1 * exp(-c/(x2 + 273.16))

Parameters:
   Estimate Std. Error t value Pr(>|t|)
a 1.124e+00  8.244e-03 136.360   <2e-16 ***
b 2.037e+11  4.227e+11   0.482    0.631
c 1.645e+04  1.137e+03  14.469   <2e-16 ***
---
```

```
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Residual standard error: 0.07585 on 125 degrees of freedom
```

186 頁下から 3 行	『非線形回帰曲線 $\log_{10}(y) = \alpha - \beta x_1 \exp(-\gamma/x_2)$』は『非線形回帰曲線 $\log_{10}(y) = \alpha - \beta x_1 \exp(-\gamma/(x_2 + 273.16))$』です。
187 頁 2 行	推定値に関する表を以下のように変更する。

	最小自乗推定値	標準偏差推定値	対応 t 値	その p 値
α	1.124	8.244e-3	136.360	2e-16
β	2.037e+11	4.227e+11	0.482	0.631
γ	1.645e+4	1.137e+3	14.469	2e-16

187 頁 7 行	『(4)`Correlation of Parameter Estimates` は母数の推定値の相関を示している.』は R2.4.1 では出力されないので削除する。
193 頁 4,5 行	『$(a = 1.13, b = 55589, c = 4091)$』は『$(a = 1.13, b = 6.375e + 11, c = 17065)$』です。
198 頁下から 13 行	『`x <- numeric(10000)`』は『`x <- numeric(rep)`』です。
198 頁下から 9 行	『`x[i] <- wa`』は『 x[i] <- wa}』です。
201 頁下から 10 行	『# b[a] は i 番目の女が選んだ男が選んだ女の番号』は『# b[a] は i 番目の男が選んだ女が選んだ男の番号』です
204 頁下から 10 行	『また，各確率分布の密度・確率関数を求めるには』は『また，たとえば正規分布の密度・確率関数を求めるには』です。
221 頁 2 行	『範囲 [0,M-1) 』は『範囲 [0,M-1]』です。
245 頁下から 11 行	第 7 章の問題 2 の回答の部分を以下のように変更する。

２．色々なやり方が考えられるが Nelson の論文にある一例を示す.

(2) 図 7.3 において，4 種類のデータごとに目測で適当に回帰直線 (実際は曲線) を 4 本引く．この 4 本の直線が $t = 0$ で交わる点が V_0 の初期値の推定値 (仮に 13.5 とする) と考える．すると $\widehat{\alpha} = \log_{10} 13.5{=}1.13$ となる.

(2) モデル式 $\log_{10} V_i = \widehat{\alpha} - t\beta \exp(-\gamma/(T_i + 273.16))$ $(i = 1, 2)$ から γ と β の初期値推定式を作ると次のようになる.

$$\begin{aligned}\widehat{\gamma} &= \frac{(T_1 + 273.16)(T_2 + 273.16)}{T_1 - T_2} \log[\log_{10}(V_0/V_1)/\log_{10}(V_0/V_2)], \\ \widehat{\beta} &= (1/t)\exp(\gamma/(T_1 + 273.16))\log_{10}(V_0/V_1)\end{aligned}$$

例えば $t = 32, T_1 = 250, T_2 = 275$ での V_1, V_2 を目測で決定 (例えば $V_1 = 9.8, V_2 = 3.27$) する．これを上の式に代入して初期値推定値 $\widehat{\beta} = 6.375e + 11, \widehat{\gamma} = 17065$ を得る.

247 頁 20 行	正しい URL は『http://www.okada.jp.org/RWiki/]』です