R Commander & 平滑化

- 基礎篇
  - セットアップ
  - 起動と終了
- 実演
  - 統計量
  - グラフ
- 非線形モデルの局所重みつき回帰モデルによる平滑化
  - lowess
  - loess

R Commanderとは

- McMaster大学(カナダ)のJohn Fox 教授が開発した画期的なパッケージ
- マウスでメニュー選択の形でRを操作する
- 2005年より日本語化
- 現在バージョン：1.3-4（2007/7/4）

セットアップ＆起動

- パッケージRcmdrをダウンロード＆インストール
  - Rcmdr.HHというパッケージもある
- library(Rcmdr)
  - 再起動するときは　Commander()
- Rを立ち上がると同時に起動する方法
  - ../R-2.5.1/etc/Rprofile.siteというファイルに
    setHook(packageEvent("grDevices","onLoad"),function(...)
    grDevices::ps.options(family="Japan1GothicBBB"))
    library(Rcmdr)
  - を追加する

実演 1

- データの読み込み
  - Excel（RExcel）との連携
  - クリップボードからの読み込み
- 統計量

実演　２

- グラフ
  - インデックスプロット：一つの変数について，点または線でデータを順番に表示する
  - 幹葉表示：
    - 幹葉図：ヒストグラムに似ているが，データの情報が分かる
    - 例：列車やバスの時刻表など

平滑化について

- R Commanderで散布図を描かせると同時に，「平滑線」というものも描かせる
- scatterplot関数でsmooth=TUREであれば，**局所重みつき回帰モデル（lowess）**を当てはめた結果が加えられる

局所回帰モデル（lowess）

- （Cleveland 1979）
- 曲線の「平滑さ」は引数「f」（0≤f≤1）
- 最小二乗法による直線（回帰直線）の安易な当てはめのデメリットを緩和する効果がある
- Rでは
  - line(lowess(data,f=0.5))
  - によって，回帰モデルの結果を当てはめる

局所回帰モデル(loess)

- モデルの説明
  - 目的とする平滑曲線が局所的に多項式で近似できると想定した平滑化を行う
  - （Cleveland and Devlin 1988）lowessの発展版である

局所回帰平滑化

- データ(t_i, y_i) i = 1, ... , n　に対して，その背後に滑らかな関数 f(t) が存在と仮定，f(t) を p 次多項式によって局所的に近似する。
- 重み関数 w(･)によって定められる近傍に重みをつけた最小二乗法によって求める。

$$\sum_{j=1}^{n} w\left(\frac{|t_j - t|}{d_\delta(t)}\right)\{y_j - f_t(t_j)\}^2 \underset{f_t}{\rightarrow} \min$$

$$f_t(s) = \sum_{k=0}^{p} \beta_k(t)(s-t)^k$$

$d_\delta(t) = \max_{j;t_j \in U_\delta(t)} |t_j - t|$

$U_\delta(t)$ は t の $[n\delta]$ 最近隣近傍

$[n\delta]$ は $n\delta$ を越えない最大整数

δ は平滑化パラメータ（span）

局所回帰モデル(loess)

$$w(x) = \begin{cases} (1-x^3)^3, & 0 \le x < 1 \\ 0\,, & \text{その他} \end{cases}$$

関数`loess`の引数`degree`によって多項式の次数pを指定

- $p = 0$

$$\hat{\beta}_0(t) = \sum_j \frac{1}{s_0(t)} \, w\left(\frac{|t_j - t|}{d_\delta(t)}\right) y_j$$

- $p = 1$

$$\hat{\beta}_0(t) = \sum_j \frac{s_2(t) - (t_j - t)\, s_1(t)}{s_0(t)\, s_2(t) - s_1(t)^2} \, w\left(\frac{|t_j - t|}{d_\delta(t)}\right) y_j$$

- $p = 2$

$$\begin{aligned} \hat{\beta}_0(t) \;=\; & \frac{1}{c(t)} \sum_j \Big[\left\{s_2(t)\, s_4(t) - s_3(t)^2\right\} - (t_j - t)\left\{s_1(t)\, s_4(t) - s_2(t)\, s_3(t)\right\} \\ & + (t_j - t)^2 \left\{s_1(t)\, s_3(t) - s_2(t)^2\right\}\Big] w\left(\frac{|t_j - t|}{d_\delta(t)}\right) y_j \end{aligned}$$