Exercise Series 7

1. The data-frame `parboot.dat` contains simulated data from the following model:

$$y = 8 \cdot x + 4 \cdot \cos(14 \cdot x) + \epsilon_i, \quad i \in 1, \ldots, 70,$$

where $x \in \{\frac{j}{70}, j = 1, \ldots, 70\}$ and $\epsilon_i \overset{\text{iid.}}{\sim} P$ for an unknown distribution P.

In this exercise we want to compare confidence-intervals for nonparametric-regression which are generated by 3 different techniques, that are:

a) hat-matrix approach (as in exercise 3)

b) "standard" (non-parametric) bootstrap (as in Series 6). It is recommended to use the bootstrap methods of the `boot` package; however, you may also use an adapted version of your own implementation that you did in Series 6.

c) parametric bootstrap with assumption $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

d) model-based bootstrap with no assumptions about the errors.

To do this, fit a smoothing-spline (automatic choice of degrees of freedom) to the `parboot`-data and compute confidence-intervals at selected locations. Those locations are a subset of all x-values:

```
x <- (1:70)/70
ind.pre <- seq(5, 62, by=3)
x.pre <- x[ind.pre]
```

Plot the data, the spline-fit, the original curve and and all confidence intervals at the selected locations into the same plot and comment on the results.

R-Hints: The data is located at `http://stat.ethz.ch/Teaching/Datasets/parboot.dat`. Use $B = 2000$ bootstrap-samples in each case. For the hat-matrix approach you need to compute the hat-matrix for smooth.spline for the given data. This can again be done by smoothing unit vectors as in exercise 3. Use the same degrees of freedom for fit and hat-matrix-generation. `smooth.spline` automatically calculates the degrees of freedom. For the parametric bootstrap approach you need an estimate for the error variance σ^2. You can use the same estimate as in hat-matrix-theory, that is

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{m}(x_i))^2}{n - df}.$$

As a hint for the interpretation you could check the Gaussian assumption that the parametrical bootstrap-technique makes by looking at the normal-plot (`qqnorm`) for the residuals.

Preliminary discussion: Friday, April 30, 2010.
Deadline: Friday, May 7, 2010.