Computational Statistics
Splines, generalized additive models

Predicting the box office success of movies is a favorite exercise for econometricians. The file `movie.data` contains data about 62 films released en 2009. The meaning of the variables is the following :

- `BOX` : receipt (in $) ;
- `MPRATING` : classification by the *Motion Picture Association of America* (a factor with four levels) ;
- `BUDGET` : movie budget ;
- `STARPOWR` : an index measuring the popularity of actors ;
- `BUZZ` : an index measuring the internet buzz (constructing by aggregating numbers of views, comments and votes on different web sites) ;
- `ACTION` : dummy variable, equals 1 for an action film.

1. Plot the response variable log(BOX) as a function of each of the predictors log(BUDGET), STARPOWR and BUZZ.
2. Try different smoothers on this data (polynomial regression, natural splines, smoothing splines). For each method, tune the degree of freedom by cross-validation.
3. Fit generalized additive models to these data. Compare their prediction errors using cross-validation.