

## Regression and basis expansions

---

You continue in this exercise with the LA ozone data. You should also keep as in the previous exercise 2/3 of the dataset as a training dataset and 1/3 as a test dataset.

**Question 5.1.** Consider as in the book, page 155, the regression of ozone concentration on the daggot pressure gradient (dpg) alone. Use the `smooth.spline` function to reproduce a similar regression.

In the following couple of exercises you are asked to investigate if the suggested non-linearity from the one-dimensional regression remains if we do a multiple regression model of ozone concentration.

**Question 5.2.** Use `bs` combined with ordinary least squares regression to estimate a full model where the dpg variable is expanded in the B-spline basis and the other variables have a linear effect. Try different choices of degrees of freedom and plot the resulting estimated effects of `bs`.

**Question 5.3.** Do a formal ANOVA test where you test the above model against a model with only the linear effect of dpg.

**Question 5.4.** Compute the training and test error using the full model with dpg expanded in a B-spline basis for a range of choices of degrees of freedom.

**Question 5.5.** Use the basis for the natural cubic splines. Compute the  $\Omega_N$  matrix in this case and write an R-function that does this in practice. Show that we don't penalize the intercept and the linear regression parameter when we do a smoothing spline fit.