

Problems: Linear Models

1. The file at <http://socserv.mcmaster.ca/jfox/Books/Applied-Regression-2E/datasets/UnitedNations.txt> contains data on social indicators collected by the United Nations in 1998; a codebook for the data set is available at <http://socserv.mcmaster.ca/jfox/Books/Applied-Regression-2E/datasets/UnitedNations.pdf> >. The data can be read into R with the command

```
UN <- read.table(  
  "http://socserv.mcmaster.ca/jfox/Books/Applied-Regression-2E/datasets/UnitedNations.txt",  
  header=TRUE)
```

After exploring the data graphically, perform a linear least-squares regression of the total fertility rate (`tfr`) on GDP per capita (`GDPperCapita`), the female illiteracy rate (`illiteracyFemale`), and the rate of contraceptive use by married women (`contraception`).

Introduce the factor `region` into the regression, first fitting an additive model to the data and then considering the possibility that region interacts with the other explanatory variables.

2. Using the Canadian occupational prestige data (the `Prestige` data frame in the `car` package – see `?Prestige`), replicate the linear regression of `prestige` on education, income, and women (i.e., percent women in the occupation). Use “regression diagnostics” to check for problems in this regression and attempt to correct any problems that you discover.
3. Apply “regression diagnostics” to the models fit to the United Nations data in Problem 1.