

**Stat 401, Section F Homework 11**

**Due Date:** Wednesday, November 28.

Read problem 17, Chapter 12 of the *Statistical Sleuth* (the data is on the course web-site: `Mortality.txt`). Please answer the following questions instead of the ones in the book.

1. Use stepwise regression to select a regression model involving weather and socioeconomic values. Set the  $p$ -to-enter as 0.25 and the  $p$ -to-stay (or leave) as 0.25. What variables are included in this model?
2. Use stepwise regression (if you are using JMP, choose "Mixed") to select a regression model involving weather and socioeconomic values (i.e. all variables except city name, hc, nox, and so2). Set the  $p$ -to-enter as 0.15 and the  $p$ -to-stay (or leave) as 0.15. What variables are included in this model?
3. Use the  $C_p$  criterion to select the best fitting regression model involving weather and socioeconomic values (i.e. all variables except city name, hc, nox, and so2). What variables are included in this model?
4. Do the three approaches select the same variables?
5. Calculate (or get the software to compute) the AIC for the models from (1), (2) and (3). Which of the models is selected by the AIC statistic?
6. Using the model selected in part (3), estimate the coefficient for education, compute a 95% confidence interval for that coefficient, and test whether it is zero.
7. A friend sees the estimate from part (6) and is very impressed. He decides to return to school, 'because you proved that more education reduces mortality'! Is this interpretation appropriate? Why or why not?
8. Calculate Cook's distance for the model from part (3). Are any of the points unusually influential?
9. The point of the study was to assess the contribution of pollution variables after adjusting for the influence of socioeconomic variables. Let's look at just one pollution variable: SO2. Add SO2 to the model selected in 3). Is the SO2 level in the city associated with the mortality, after adjusting for socioeconomic variables? How big is that effect? You do not need to compute a confidence interval.