

Homework 11

Statistics 411/511: Spring 2018

Due: In class Wednesday, April 18

Do Exercise 23 on page 342-3 of your text where you will analyze study data to assess *Does pollution kill people?*. The predictors that quantify pollution are due to nitrogen NOX and sulphur dioxide S02. Possible confounding variables that we want to “account for” in the analysis are precipitation Precip, education level Educ and the percent of the population that is “non-white” Nonwhite. You must report your answers to this problem as a *Statistical Report*. The report, not including the Appendix that contains your R-code and R-output and figures and any tables, should not exceed two pages. Your grade will be determined by how well you answer the questions and by the organization and clarity of your write-up.

1. Load in the data and perform initial summaries including matrix and Trellis plots of all the quantitative variables. Pay special attention to Trellis plots of Mort, NOX and S02.
2. Based on your initial summaries, consider transforms of the quantitative variables in this data set.
3. Fit a simple model of (possibly transformed) Mort as a function of the (possibly transformed) other variables with no interactions. Using this model, test whether the coefficient of the S02 term is larger than the coefficient of the NOX term. *Hint: this is a linear combination of the coefficients.*
4. Consider interactions. Formulate a more complicated model of mortality as a function of the pollution variables that also accounts for the climate and socioeconomic variables. Consider 2-way interaction between NOX and S02; and also 2-way interaction between each of NOX and S02 and the possible confounding variables in the data set. Assess fit via normal probability and residual plots.
5. Are any unusual data evident in the residual plots? Which city or cities do these datum come from?
6. Consider R^2 and extra sum of squares F -tests when deciding on a “final” model. In the *State procedures* section of your report, summarize what models you investigated and what process led to the final model that you present. You do not need to include all of the plots that you generated in the exploratory phase of your analysis!
7. Indicate the sampling plan and the study design in the introduction to your report. In the *Scope of inference* part of your report you will indicate how the sampling plan and the study design affects the applicability (or lack of applicability) of your conclusion regarding extinction of birds.
8. Write out the final regression model that includes all parameters β_0, β_1, \dots in the *Stat procedures* section of your report.
9. Write out the fitted regression equation in the *Summary of statistical findings* section of your report.
10. State all relevant assumptions and include any plots or other output that indicate that the final model you fit satisfies these assumptions.
11. Answer the question posed by the book. To substantiate your answer, state the parameters relevant to this question, and report and interpret relevant CIs.