Hwk 06 Solutions L9
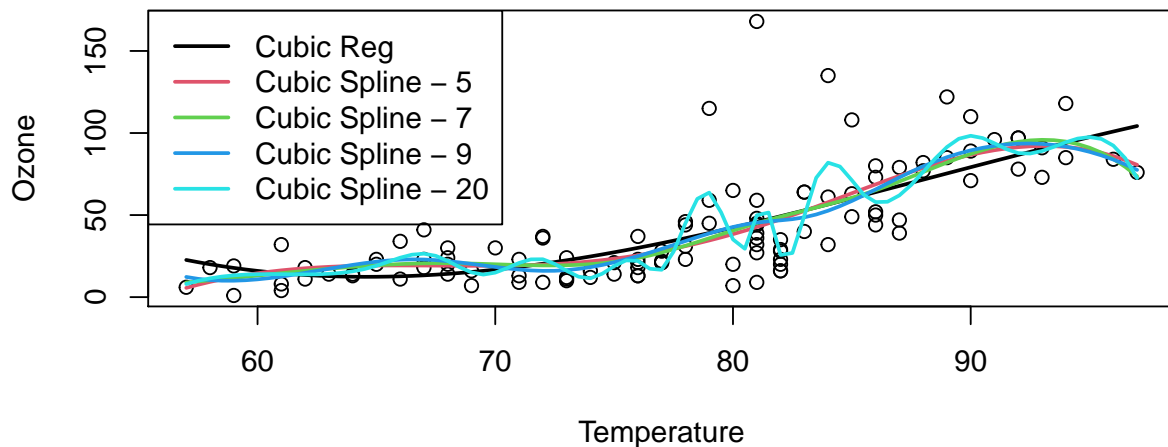
Due on Oct 23, 2020

Applications

Question 1

Figure 1: Cubic Regression and Cubic Splines

(a) See Figure 1 for a comparison of cubic splines.

(b) We know that if there is bias, then the least flexible model likely has the most bias. This is true here, as well, although it is hard to see. Look between 63–70, where it underestimates many consecutive points. Then between about 73–77 it underestimated a lot of points. The other models get closer to the centers of these points.

(c) The cubic spline with 20 degrees of freedom appears to be overfitting because it is very wiggly and seems to overreact to individual observations.

(d) If we had to choose a single model, we would use the something around 5–9 DF because they do a good job of following the general increasing, upward curved trend that we see in the dataset without following the errors too much.

Question 2

Based on the results from question 1, I will use 5, 7, and 9 degrees of freedom. I think that these cover a reasonable range of flexibility without being too extreme toward bias or variance.
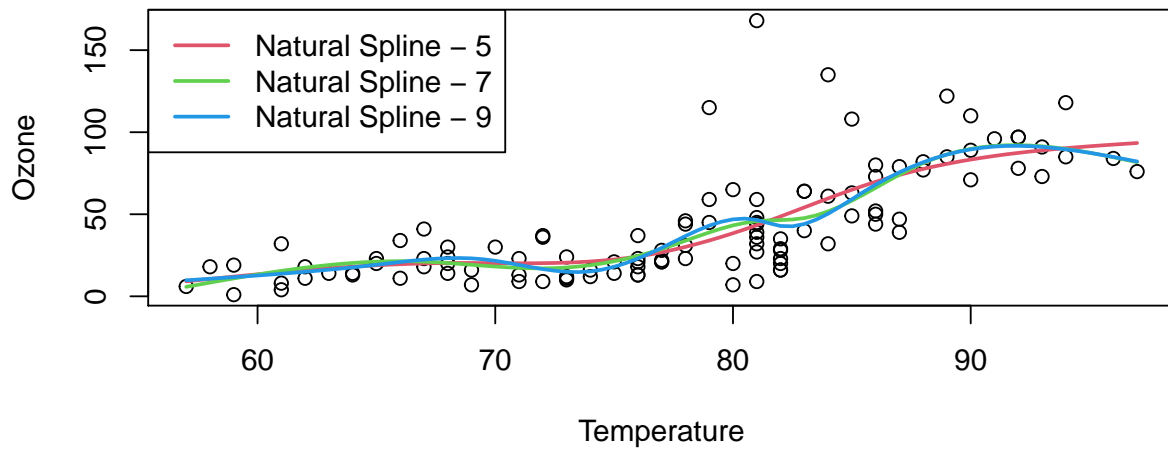

Figure 2: Natural Splines

(b) See Figure 2 for natural splines fit with my chosen degrees of freedom. (We now prefer the fit with 5 degrees of freedom, since the 7 DF fit curves down more than I would like around where temperature equals 82 degrees. Indeed, this 5df fit seems better than any of the cubic spline fits ast the upper end of temp, where I don't expect Ozone to suddenly start decreasing.)