FINAL EXAM: Problem 1

STAT 472, Spring 2020

The following will read in the appropriate data and make it easy to work with (the `attach(prob1.dat)` will allow you to use the variables `x`, `y`, and `m` (where `m` contains the true values of $E(Y|x)$ that were used to produce the values of `y` by adding random amounts to them)):

```
prob1.dat = read.table("http://mason.gmu.edu/~csutton/FEs20prob1.txt", header=TRUE)
dim(prob1.dat)
head(prob1.dat)
attach(prob1.dat)
plot(x,y)
```

The scatter plot suggests that it might be good to model $E(Y|x)$ with a linear spline model. Using the bottom portion of p. 7-8 of the class notes (which can be accessed from the *Week 9* section of the *Blackboard* site), and the values $\xi_1 = 20$ and $\xi_2 = 29$, create vectors `b2` and `b3` from the values of `x`, so that `b2` and `b3` contain the values of $b_2(x)$ and $b_3(x)$ described in the class notes. (Note that `x` contains the values of the basis function $b_1(x) = x$.) Then use the `lm()` function to fit the linear spline model. (Note that the basis function $b_0(x) = 1$ will be automatically included when `lm()` is used. (Using $b_0(x) = 1$ as a basis function is equivalent to including a constant/intercept term in a regression model.))

(a) (1 point) What is the value of R^2?

(b) (2 points) Rounded to the nearest thousandth, what is the slope of the estimate of $E(Y|x)$ for x between 20 and 29? (Provide justification for your answer.)

(c) (1 point) What p-value results from the t test about the coefficient of $b_3(x)$? (The large p-value suggests that the second knot at $x = 29$ is not needed.)