Review Week Seven Answers

1. Chapter 7, question 1.
   (a)

1. It was mentioned in the chapter that a cubic regression spline with one knot at ξ can be obtained using a basis of the form x, x^2, x^3, $(x-\xi)^3_+$, where $(x-\xi)^3_+ = (x-\xi)^3$ if $x > \xi$ and equals 0 otherwise. We will now show that a function of the form

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x-\xi)^3_+$$

is indeed a cubic regression spline, regardless of the values of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$.

(a) Find a cubic polynomial

$$f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3$$

such that $f(x) = f_1(x)$ for all $x \leq \xi$. Express a_1, b_1, c_1, d_1 in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$.

Since the β_4 term is zero, $\beta_0 = a_1, \beta_1 = b_1, \beta_2 = c_1\ and \beta_3 = d_1$.

(b) Find a cubic polynomial

$$f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3$$

such that $f(x) = f_2(x)$ for all $x > \xi$. Express a_2, b_2, c_2, d_2 in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$. We have now established that $f(x)$ is a piecewise polynomial.

Ans: $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4(x^3 - 3x^2\xi + 3x\xi^2 - \xi^3)$

$a_2 = [\beta_0 - \beta_4\xi^3]; b_2 = [\beta_1 + 3\beta_4\xi^2];\ c_2 = [\beta_2 - 3\beta_4\xi];\ d_2 = [\beta_3 + \beta_4]$

(c) Show that $f_1(\xi) = f_2(\xi)$. That is, $f(x)$ is continuous at ξ.

$f_1(\xi) = \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3$

$f_2(\xi) = [\beta_0 - \beta_4\xi^3] + [\beta_1 + 3\beta_4\xi^2]\xi + [\beta_2 - 3\beta_4\xi]\xi^2 + [\beta_3 + \beta_4]\xi^3$

$f_2(\xi) = \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 - \beta_4\xi^3 + 3\beta_4\ \xi^3 - 3\beta_4\xi^3 + \beta_4\xi^3 = f_1(\xi)$

(d) Show that $f_1'(\xi) = f_2'(\xi)$. That is, $f'(x)$ is continuous at ξ.

$f_1'(x) = \beta_1 + 2\beta_2 x + 3\beta_3 x^2$

$f_1'(\xi) = \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2$

$f_2'(x) = [\beta_1 + 3\beta_4\xi^2] + 2x[\beta_2 - 3\beta_4\xi] + 3x^2[\beta_3 + \beta_4]$

$f_2'(\xi) = \ \beta_1 + \xi[2\beta_2] + \xi^2[3\beta_4 - 6\beta_4 + 3\beta_3 + 3\beta_4] = f_1'(\xi)$

(e) Show that $f_1''(\xi) = f_2''(\xi)$. That is, $f''(x)$ is continuous at ξ.

$f_1''(x) = 2\beta_2 + 6\beta_3 x$

$f_1''(\xi) = 2\beta_2 + 6\beta_3\xi$
$f_2''(x) = 2[\beta_2 - 3\beta_4\xi] + 6x[\beta_3 + \beta_4]$

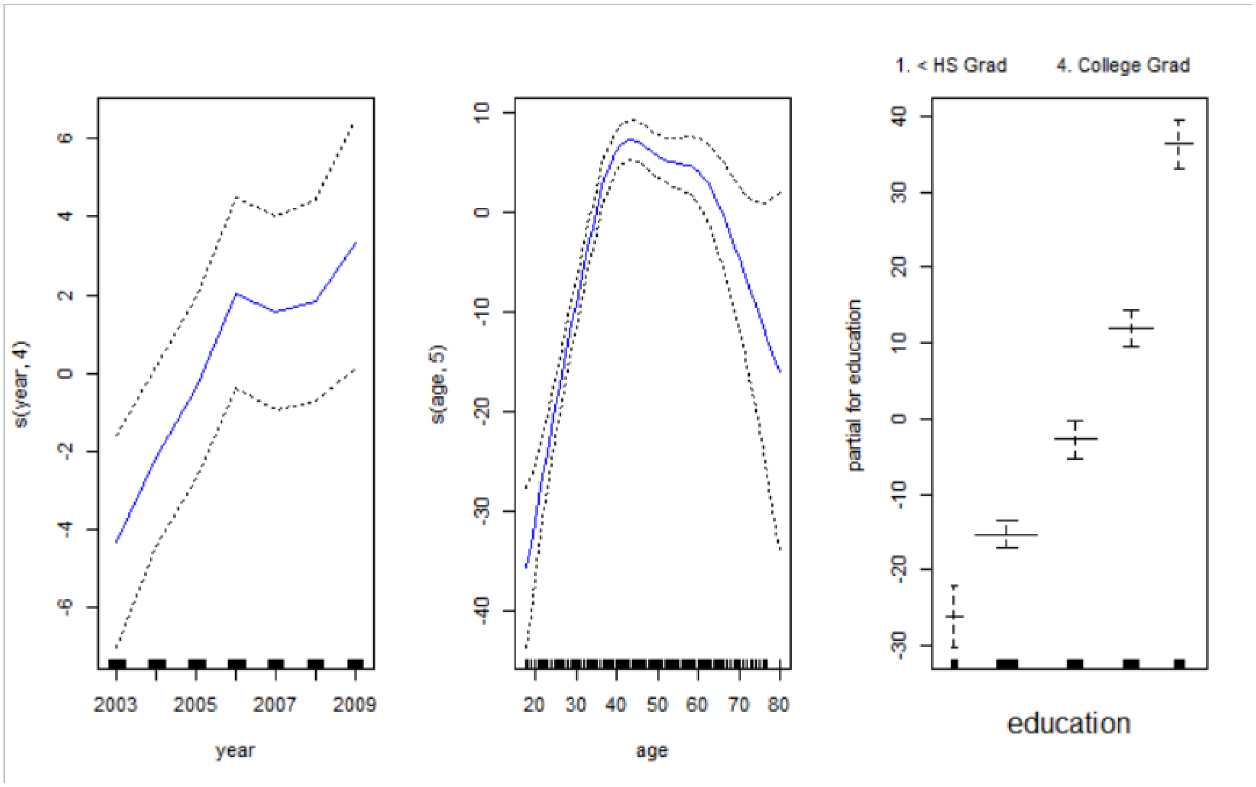
$f_2''(\xi) = 2[\beta_2 - 3\beta_4\xi] + 6\xi[\beta_3 + \beta_4] = 2\beta_2 + 6\beta_3\xi - 6\beta_4\xi + 6\beta_4\xi$
$f_2''(\xi) = f_1''(\xi)$

However, the third derivatives are different

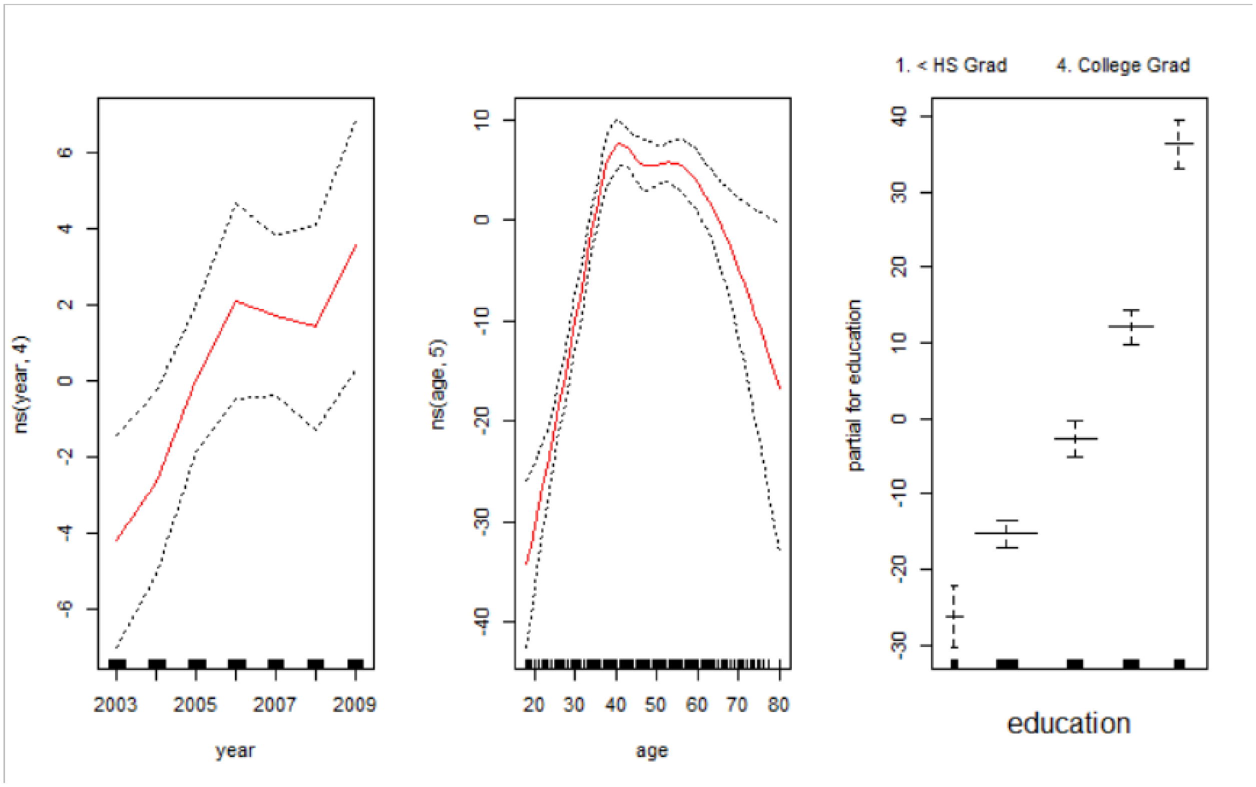
$$f_1'''(\xi) = 6\beta_3 \quad f_2'''(\xi) = 6(\beta_3 + \beta_4)$$

2. Chapter 7, , do the lab exercise in section 7.8.3

```
library(ISLR2)
attach(Wage)
library(gam)
gam1 <- lm(wage~ ns(year,4) + ns(age,5) + education, data =
Wage)
gam.m3 <- gam(wage~ s(year,4) + s(age , 5) + education,data =
Wage)
par(mfrow = c(1, 3))
plot(gam.m3, se = TRUE , col = "blue")
```




```
plot.Gam(gam1 , se = TRUE , col = "red") # Since gam1 was
created from "lm"
                                          # we use "plot.Gam"
```




```
# Book does hypotheses tests to compare a linear model of
year with the
# spline model. We will use a test set of data.
set.seed(3)
train<- sample(1:length(Wage$wage),0.8*length(Wage$wage))
gam.basic<- gam(wage~ s(age , 5) + education,data = Wage,
subset=train)
gam.linear<- gam(wage~ year+s(age , 5) + education,data =
Wage, subset=train)
gam.full<- gam(wage~ s(year,4)+s(age , 5) + education,data =
Wage, subset=train)
# Estimate MSE from the test set
basic.mse<- mean((Wage$wage[-train]-predict(gam.basic,
newdata = Wage[-train,]))^2)
linear.mse<- mean((Wage$wage[-train]-predict(gam.linear,
newdata = Wage[-train,]))^2)
full.mse<- mean((Wage$wage[-train]-predict(gam.full, newdata
= Wage[-train,]))^2)
# Display results, I manually reduced the number of digits
cat(rbind(c("Basic=", "Linear=",
"Full="),c(basic.mse,linear.mse,full.mse)))
Basic= 1369.4 Linear= 1360.6 Full= 1363.4
```