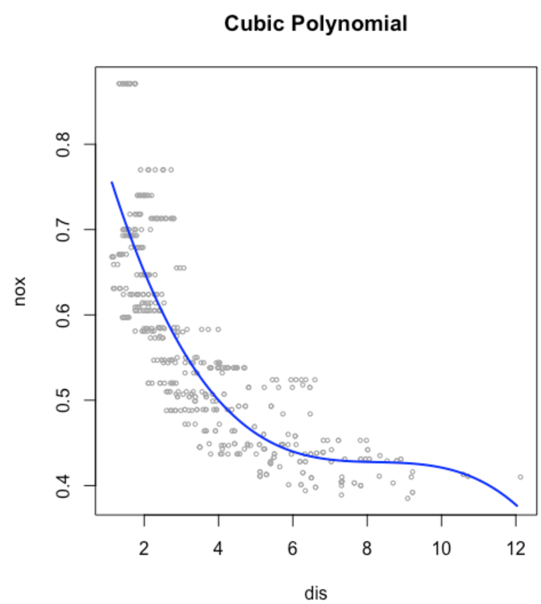
Hw 4
Marc Mailloux
ma147661

Exercise 9

#Chapter 7 Exercise 9

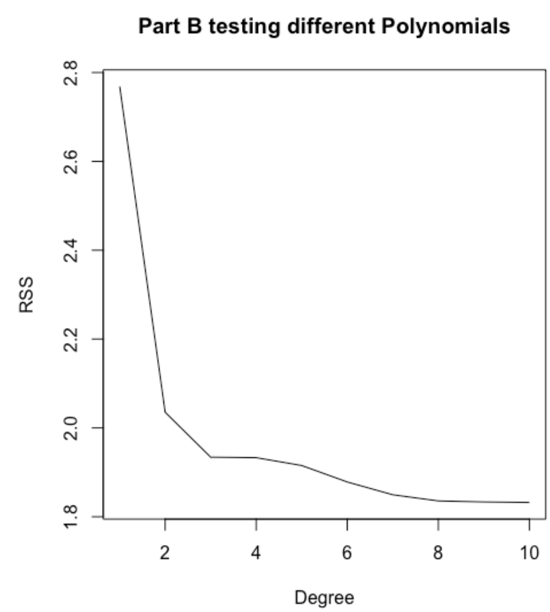
```
library( ISLR)
library(boot)
library(MASS)
set.seed(333)
fit <- lm(nox ~ poly(dis,3), data = Boston)
summary(fit)
attach(Boston)
#finding the smallest and largest values
dislims = range(Boston$dis)

dis.grid = seq(from = dislims[1],to = dislims[2], by = .1)
pred.fit <- predict(fit, newdata = list(dis = dis.grid))

plot(nox~ dis, data = Boston ,xlim = dislims, cex = .5 , col =
'darkgrey')
title("Cubic Polynomial")
lines(dis.grid,pred.fit,lwd = 2, col = 'blue')
```



#from the summary all of the polynomial degrees are significant

#b
#testing for all 10 variables relative to RSS'

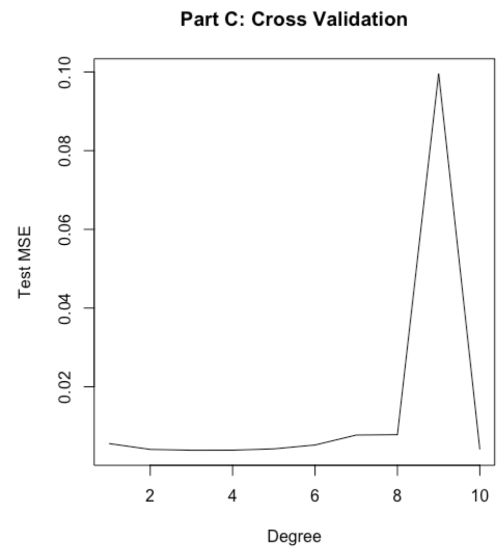
```
rss <- rep(NA, 10) #empty list

for ( i in 1:10){
  fit <- lm(nox ~ poly(dis,i), data = Boston)
  rss[i] <- sum(fit$residuals^2)
}

plot(1:10, rss , xlab = "Degree", ylab = "RSS", type = "l")
title('Part B testing different Polynomials')
```



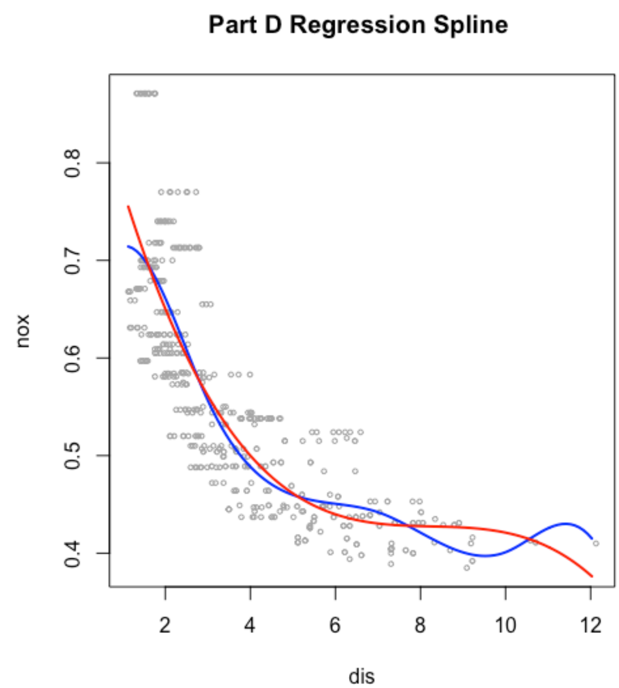
#10 has the lowest RSS

#part C perform CV to select the optimal polynomial

```
cv <- rep(NA,10)

for (i in 1:10){
  fit <- glm(nox ~poly(dis,i), data = Boston)
  cv[i] <- cv.glm(Boston, fit, K=10)$delta[1]
}

plot(1:10, cv , xlab = "Degree", ylab = "Test MSE", type = "l")
title('Part C: Cross Validation')
summary(cv)
```

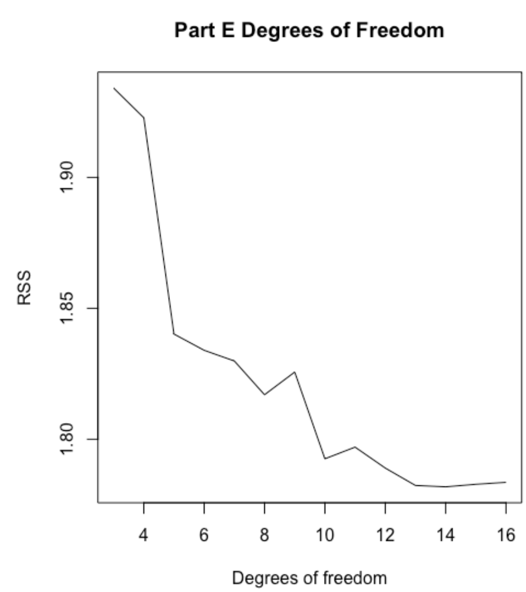



```
#part d

#fitting a regression spline

library(splines)

fit = lm(nox~ bs(dis, knots = c(3,7,10)))
pred = predict(fit,list(dis = dis.grid))

plot(nox~ dis, data = Boston ,xlim = dislims, cex = .5 , col =
'darkgrey')
title("Part D Regression Spline ")
lines(dis.grid,pred,lwd = 2, col = 'blue')
lines(dis.grid,pred.fit,lwd = 2, col = ‘red')
```



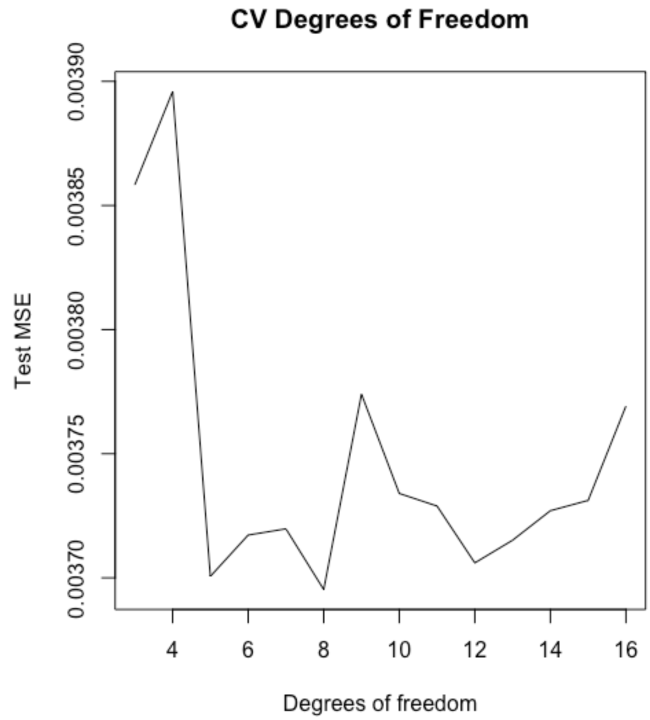
#all of the polynomials are significant except for the first order.

```
#part e

rss <- rep(NA, 16)
for (i in 3:16) {
  fit <- lm(nox ~ bs(dis, df = i), data = Boston)
  rss[i] <- sum(fit$residuals^2)
}
plot(3:16, rss[-c(1, 2)], xlab = "Degrees of freedom", ylab =
"RSS", type = "l")
title('Part E Degrees of Freedom')
```



#the RSS decreases and levels off at 13 or 14 degrees of freedom

```
#part f CV for degrees of freedom
cv <- rep(NA, 16)
for (i in 3:16) {
  fit <- glm(nox ~ bs(dis, df = i), data = Boston)
  cv[i] <- cv.glm(Boston, fit, K = 10)$delta[1]
}
plot(3:16, cv[-c(1, 2)], xlab = "Degrees of freedom",
ylab = "Test MSE", type = "l")
title("CV Degrees of Freedom")
#the minimum look to be 11 for this instance
```



Hw 7

1) $f(x) = B_0 + B_1 x + B_2 x^2 + B_3 x^3 + B_4 (x - \xi)_+^3$

a) Cubic poly $\qquad (x - \xi)^3 = x^3 - 3x^2\xi + 3x\xi^2 - \xi^3$

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

$f_1(x) = f(x) \qquad x \le \xi$

So $x^0 \quad a_1 = B_0 \qquad x^3 \quad d_1 = B_3$

$x^1 \quad b_1 = B_1$

$x^2 \quad c_1 = B_2$

b) $f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3$ for all $x > \xi$

$f_2(x) = f(x)$

$x^0 \quad a_2 = B_0 - B_4 \xi^3 \qquad x^3 \quad d_2 = B_3 + B_4$

$x^1 \quad b_2 = B_1 + B_4 3\xi^2$

$x^2 \quad c_2 = B_2 - 3\xi B_4$

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

$f_1(\xi) = B_0 + B_1\xi + B_2\xi^2 + B_3\xi^3$

$f_2(\xi) = B_0 - \cancel{B_4\xi^3} + \xi(B_1 + \cancel{3B_4\xi^2}) + \xi^2(B_2 - \cancel{3B_4\xi}) + \xi^3(B_3 + \cancel{B_4})$

$= B_0 + B_1\xi + B_2\xi^2 + B_3\xi^3 = f_1(\xi) = f_2(\xi)$

D) Show $f_1'(z) = f_2'(z)$ and are cont for z

$f_1(z) = B_0 + B_1 z + B_2 z^2 + B_3 z^3$

$f_1'(z) = B_1 + 2B_2 z + 3B_3 z^2$

$f_2(z) = (B_0 - B_4 z^3) + (B_1 + 3z^2 B_4)z + (B_2 - 3B_4(z))z^2 + (B_3 + B_4)z^3$

$f_2'(z) = B_1 + 2B_2 z + 3B_3 z^2$ Since $f_1(z) = f_2(z)$ then $f_1'(z) = f_2'(z)$

e) $f_1''(z) = 2B_2 + 6B_3 z = f_2''(z)$

Since $f_1(z) = f_2(z)$ $f_2''(z) = f_1''(z)$

$f_2''(z) = 2B_2 + 6B_3 z$

2 $\hat{g} = \arg\min_g \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int [g^{(m)}(x)]^2 dx \right)$

a) $\lambda = \infty$ $m = 0$

So $\hat{g} = 0$ $g^{(0)}(x) \to 0$

b) $\lambda = \infty$ $m = 1$

$\hat{g} = c$ $g^{(1)}(x) \to 0$

c) $\lambda = \infty$ $m = 2$

$\hat{g} = cx + d$ $g^2(x) \to 0$

d) $\lambda = \infty$ $m = 3$

$\hat{g} = cx^2 + dx + e$ $g^3(x) \to 0$

e) $\lambda = 0$ $m = 3$

the penalty isn't on, so it's the spline