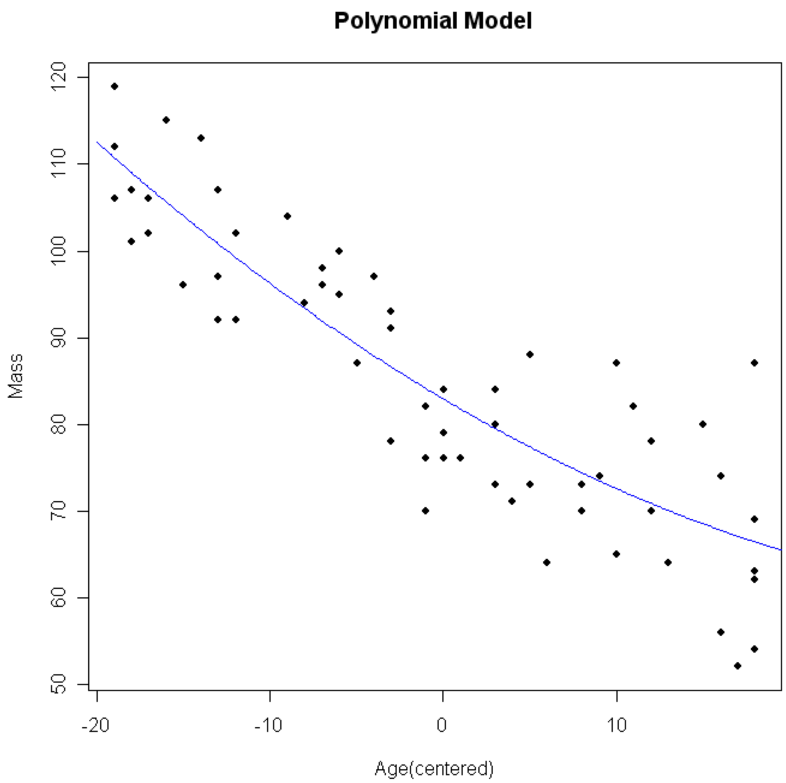
Homework 7 Solutions: Chapter 8: #4abeg, 17, 34

Assigned Wed, Nov 13:

8.4 **a.** Define $x_i = X_i - \bar{X}$. Fit regression model 8.2: $\hat{Y} = 82.936 - 1.184x + 0.0148x^2$. Plot the fitted regression function and the data: See plot below. The quadratic function does appear to be a good fit. R^2 = .7623 or 76.32%.



There are various ways you could to this in R. Here is the R code for one way of doing it, assuming the data frame is called "Data" and the variables are called Mass and Age:

```
#Center the Age variable
x1 = Data$Age - mean(Data$Age)
x1sq = x1^2
Data = cbind( Data, x1, x1sq)

#fit a quadratic model (8.2)
Quad = lm( Mass ~ x1 + x1sq, data=Data )

#plot the points, then the fitted quadratic model
plot( Data$x1, Data$Mass, main="Polynomial Model",
xlab="Age(centered)", ylab="Mass", pch=19 )
#create (x,y) points for quadratic model; plot it using a blue line
x = seq(-20, 20, by=.1)
y = 82.935749 - 1.183958*x + 0.014840*x^2
lines( x, y , col="blue")
```

R Code and some results for Parts (b) and (e)

```
summary(Quad)
Call:
lm(formula = Mass ~ x1 + x1sq, data = Data)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 82.935749   1.543146  53.745   <2e-16 ***
x1          -1.183958   0.088633 -13.358   <2e-16 ***
x1sq         0.014840   0.008357   1.776   0.0811 .
Residual standard error: 8.026 on 57 degrees of freedom
Multiple R-Squared: 0.7632,   Adjusted R-squared: 0.7549
F-statistic: 91.84 on 2 and 57 DF,  p-value: < 2.2e-16
```

b. Test whether or not there is a regression relation; use $\alpha = 0.05$:
H_0: $\beta_1 = \beta_{11} = 0$; Ha: β_1 and β_{11} and not both 0. (You could use notation β_2 instead of β_{11}.) You can use either the decision rule: Reject H_0 if $F^* > F(0.95, 2, 57) = 3.15884$, or you can use the p-value and reject H_0 if the p-value $< \alpha=0.05$. The relevant part of the R output is: `F-statistic: 91.84 on 2 and 57 DF,  p-value: < 2.2e-16.`
F-statistic = F^* = 91.84 and the p-value is essentially 0, thus, we reject H_0, and conclude that there is sufficient evidence to indicate that a regression relation exists between muscle mass and the centered age variable and its squared value.

e. The test of whether the quadratic term can be dropped from the model is a test of H_0: $\beta_{11} = 0$ versus Ha: $\beta_{11} \neq 0$. From the R output the appropriate test statistic is $t^* = 1.776$, p-value = 0.0811. Because 0.0811 > 0.05, do not reject H_0. Conclude that the coefficient for the quadratic term is not statistically significantly different from 0, so the quadratic term can be dropped from the model without too much loss. Using the decision rule approach, Reject H_0 if $|t^*| \geq 2.00247$. Because $t^* < 2.00247$, do not reject H_0.

g. Correlation between X and X^2 = .996. Correlation between x and x^2 = –0.0384. Yes, it was helpful to use the centered data because of the high correlation between X and X^2.

8.17 No, the conclusion that $\beta_2 = 0$ would not have the same meaning for the two equations. In equation 8.33, β_2 represents how much greater (or less) the response function is for stock companies compared to mutual companies, which remains constant for all sizes of firms. When $\beta_2 = 0$ it indicates that the response function is identical for the two types of firms. In equation 8.49 there is an interaction, so the interpretation of β_2 isn't as straight-forward. When $\beta_2 = 0$ the intercepts for the two types of firms are the same, but the slopes are different (unless $\beta_3 = 0$ as well).

8.34 **a.** $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$

b. Commercial: $Y_i = (\beta_0 + \beta_2) + \beta_1 X_{i1} + \varepsilon_i$

Mutual savings: $Y_i = (\beta_0 + \beta_3) + \beta_1 X_{i1} + \varepsilon_i$

Savings and loan: $Y_i = (\beta_0 - \beta_2 - \beta_3) + \beta_1 X_{i1} + \varepsilon_i$

c. Interpretations all use the fact that β_0 is the average of the 3 intercepts. Then:

- β_2 is the additional intercept for the relationship between Y and X_1 for Commercial banks;
- β_3 is the additional intercept for the relationship between Y and X_1 for Mutual savings;
- $-\beta_2$ $-\beta_3$ is the additional intercept for the relationship between Y and X_1 for Savings and loans.