MA 4710 Homework 10

Benjamin Hendrick

April 13, 2016

Problem 8.4

Load the data and rename the variables.

```
Muscle <- read.table("~/GitHub/MA-4710/Homework 10/Muscle.txt", quote="\"", comment.char="")
names(Muscle) <- c("y","x")
```
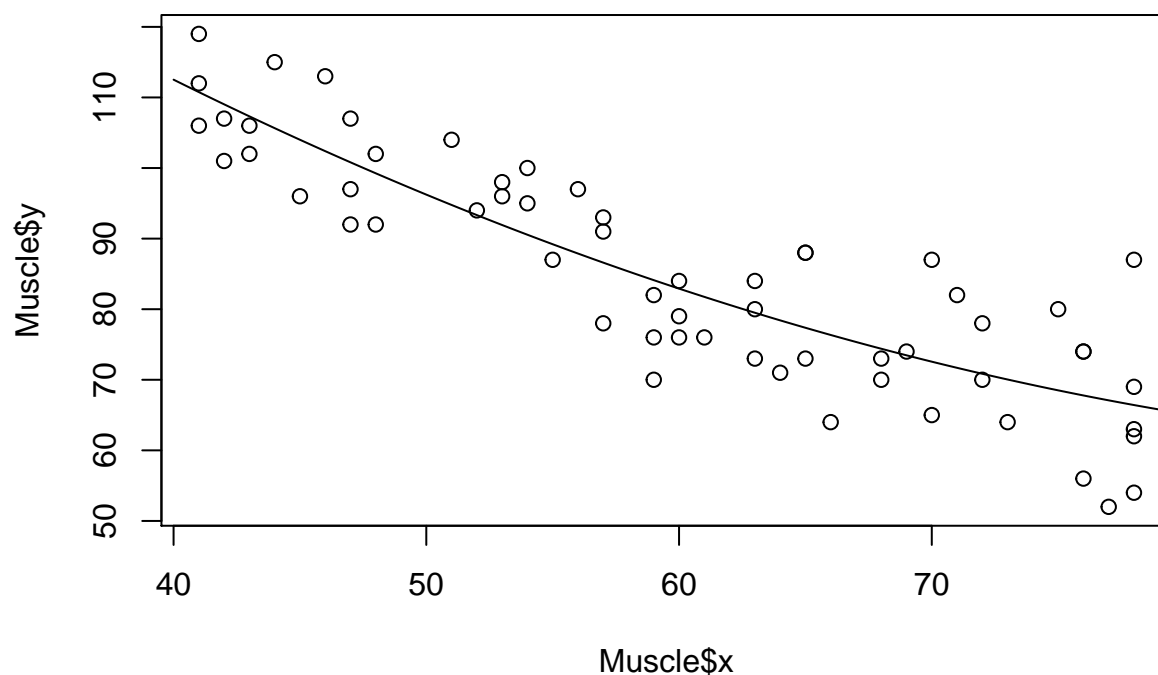
Part A

Fit the regression model.

```
muscle.fit <- lm(y~x+I(x^2), data = Muscle)
```

The regression model is $Y_i = 207.34961 - 2.96432x_i + 0.01484x_i^2$.

Plot the fitted data with the regression line.

```
plot(Muscle$x, Muscle$y)
x <- seq(40,80, by=0.01)
y <- muscle.fit$coefficients[1] + muscle.fit$coefficients[2]*x + muscle.fit$coefficients[3]*x^2
lines(x,y)
```


Find the R^2 value with the `summary` function.

```
summary(muscle.fit)
```

```
##
## Call:
## lm(formula = y ~ x + I(x^2), data = Muscle)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.086  -6.154  -1.088   6.220  20.578
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 207.349608  29.225118   7.095 2.21e-09 ***
## x            -2.964323   1.003031  -2.955  0.00453 **
## I(x^2)        0.014840   0.008357   1.776  0.08109 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## 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
```

$R^2 = 0.7631707$

Part B

Test for regression relation is done with the **summary** function.

```
summary(muscle.fit)
```

```
##
## Call:
## lm(formula = y ~ x + I(x^2), data = Muscle)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.086  -6.154  -1.088   6.220  20.578
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 207.349608  29.225118   7.095 2.21e-09 ***
## x            -2.964323   1.003031  -2.955  0.00453 **
## I(x^2)        0.014840   0.008357   1.776  0.08109 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## 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
```

Becasue the p-values for the intercept and x are less than 0.05, there is no relation. However, because the p-value for x^2 is greater than 0.05, there is regression relation for this variable.

Part C

```
x <- 48
y <- muscle.fit$coefficients[1] + muscle.fit$coefficients[2]*x + muscle.fit$coefficients[3]*x^2
newdata <- data.frame(y=y, x=x)
predict(muscle.fit, newdata, interval = "confidence")

##                   fit      lwr      upr
## (Intercept) 99.25461 96.28436 102.2249
```

The estimated mean muscle mass for women aged 48 years is 99.2546116. The 95 percent confidence interval is (96.2843639, 102.2248593). This means that we can be 95 confident that the mean muscle mass for women aged 48 is between these two bounds.

Part D

```
predict(muscle.fit, newdata, interval = "predict")

##                   fit     lwr      upr
## (Intercept) 99.25461 82.9116 115.5976
```

The predicted muscle mass for women aged 48 years is 99.2546116. The 95 percent confidence interval is (82.9116032, 115.5976201). This means that we can be 95 confident that the predicted muscle mass for women aged 48 is between these two bounds.

Part E

```
muscle.fit.simple <- lm(y~x, data = Muscle)
anova(muscle.fit.simple, muscle.fit)

## Analysis of Variance Table
##
## Model 1: y ~ x
## Model 2: y ~ x + I(x^2)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     58 3874.4
## 2     57 3671.3  1    203.13 3.1538 0.08109 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis is $H_0 : \beta_0 = \beta_1 = \beta_2$. The alternative hypothesis is $H_a : \beta_i \neq \beta j$ for some $i \neq j$. We reject H_0 if the p-value is less than $\alpha = 0.05$.

We fail to reject H_0 becasue the $0.08109 > 0.05$. This means that we can drop the quadratic term from the regression model.

Part F

Use the quadratic formula to find the regression model in terms of x.

$$x = \frac{2.96 \pm \sqrt{(-2.96)^2 - 4(207.35)(0.015) + 4(0.015)(y)}}{2(207.35)}$$

Problem 8.5

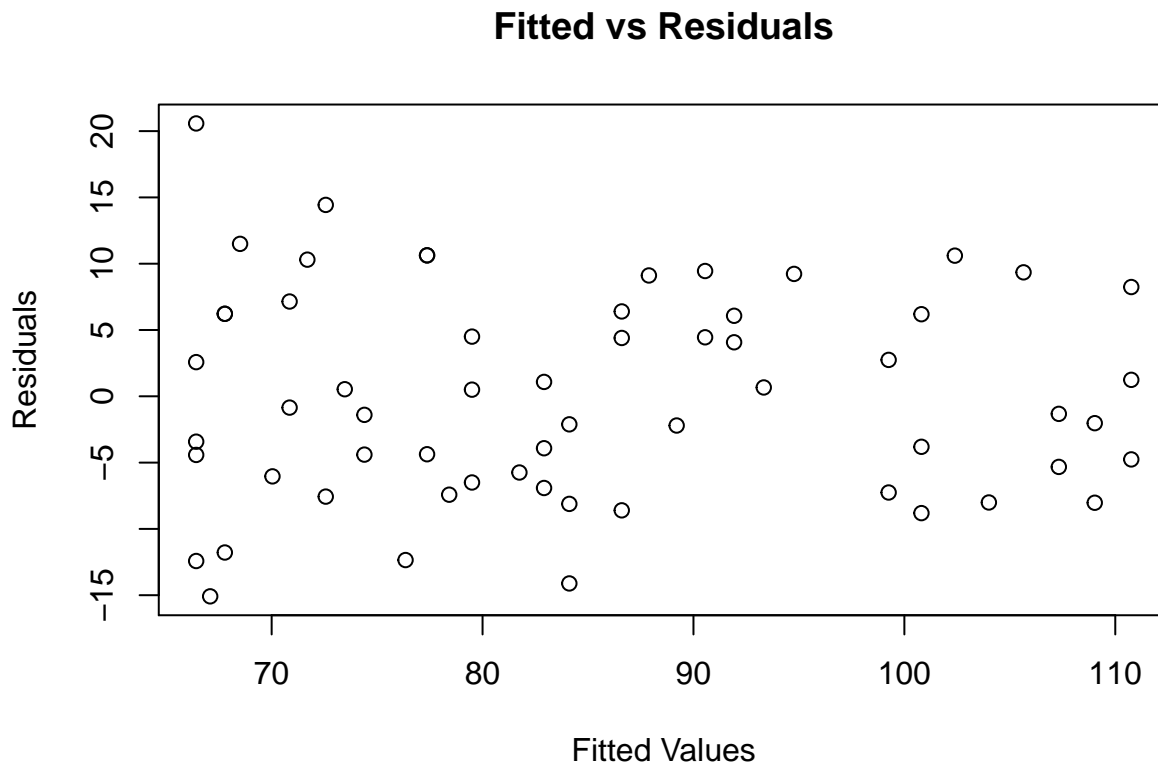Use the data and regression model from Problem 8.4.

Part A

Find the residuals.

```
muscle.fit.resid <- residuals(muscle.fit)
```

Plot the residuals against the fitted values.

```
plot(fitted(muscle.fit), muscle.fit.resid, main = "Fitted vs Residuals",
     xlab = "Fitted Values", ylab = "Residuals")
```
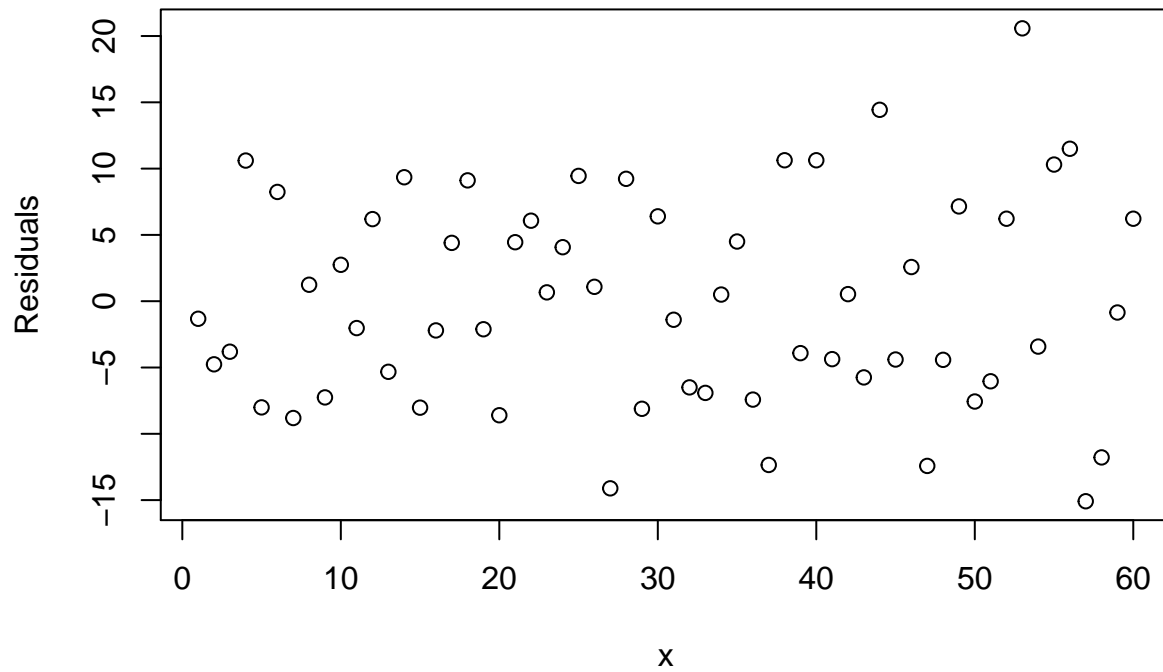
Fitted vs Residuals

There is nothing out of the ordinary with this plot.

Plot the residuals against the x values.

```
plot(muscle.fit.resid, main = "x vs Residuals",
     xlab = "x", ylab = "Residuals")
```

x vs Residuals

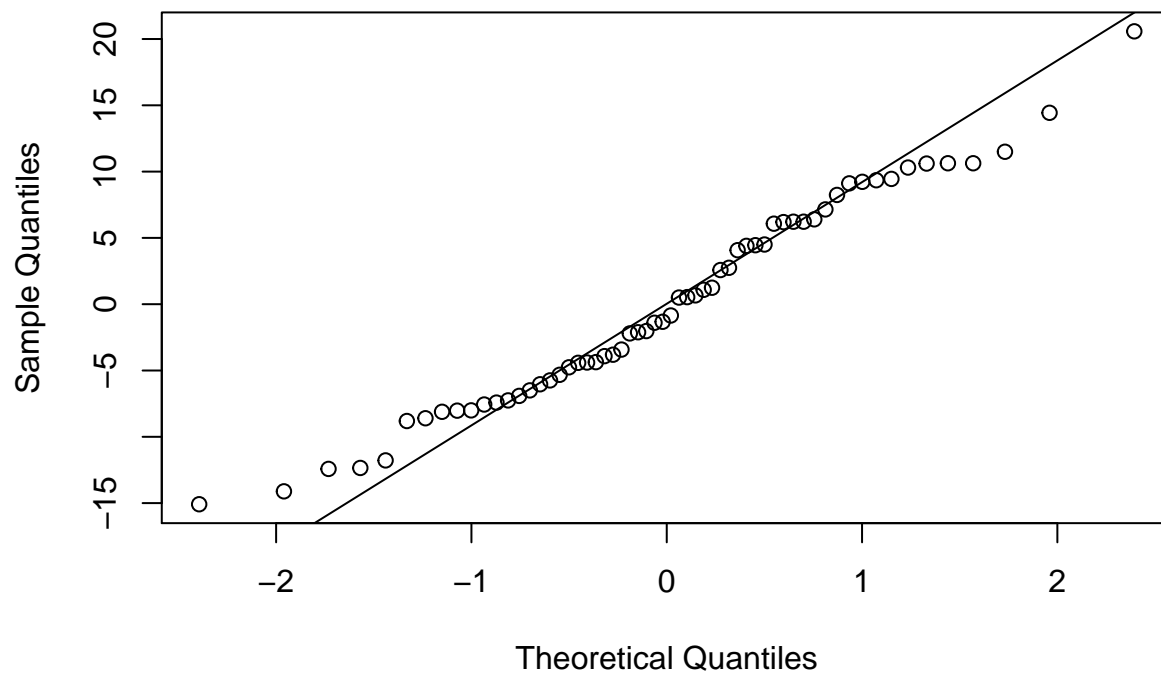There is nothing out of the ordinary with this plot.

Plot the Q-Q plot of the residuals.

```
qqnorm(muscle.fit.resid)
qqline(muscle.fit.resid)
```

Normal Q–Q Plot

The residuals appear to be normal, but with short tails.

Part B

```
anova(muscle.fit)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq  F value  Pr(>F)
## x          1 11627.5 11627.5 180.5258 < 2e-16 ***
## I(x^2)     1   203.1   203.1   3.1538 0.08109 .
## Residuals 57  3671.3    64.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis is $H_0 : \beta_0 = \beta_1 = \beta_2$. The alternative hypothesis is $H_a : \beta_i \neq \beta j$ for some $i \neq j$. We reject H_0 if the p-value is less than $\alpha = 0.05$.

The p-value is significantly less than 0.05, therefore we reject H_0

Problem 8.16

Load the data and rename the variables.

```
gpa2 <- read.delim("~/GitHub/MA-4710/Homework 10/gpa2.txt", header=FALSE)
names(gpa2) <- c("y", "x1", "x2")
```

Part A

The regression model is $Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2$

- β_0 is the intercept value of the GPA
- β_1 is the rate of change of ACT scores per unit
- β_2 is the rate of change of change of major concetration per unit.

Part B

```
gpa2.fit <- lm(y~x1+x2, data = gpa2)
```

The model is $Y_i = 2.198 + 0.038 x_1^2 - 0.094 x_2^2$

Part C

```
gpa2.fit.simple <- lm(y~x1, data = gpa2)
anova(gpa2.fit.simple, gpa2.fit)
```
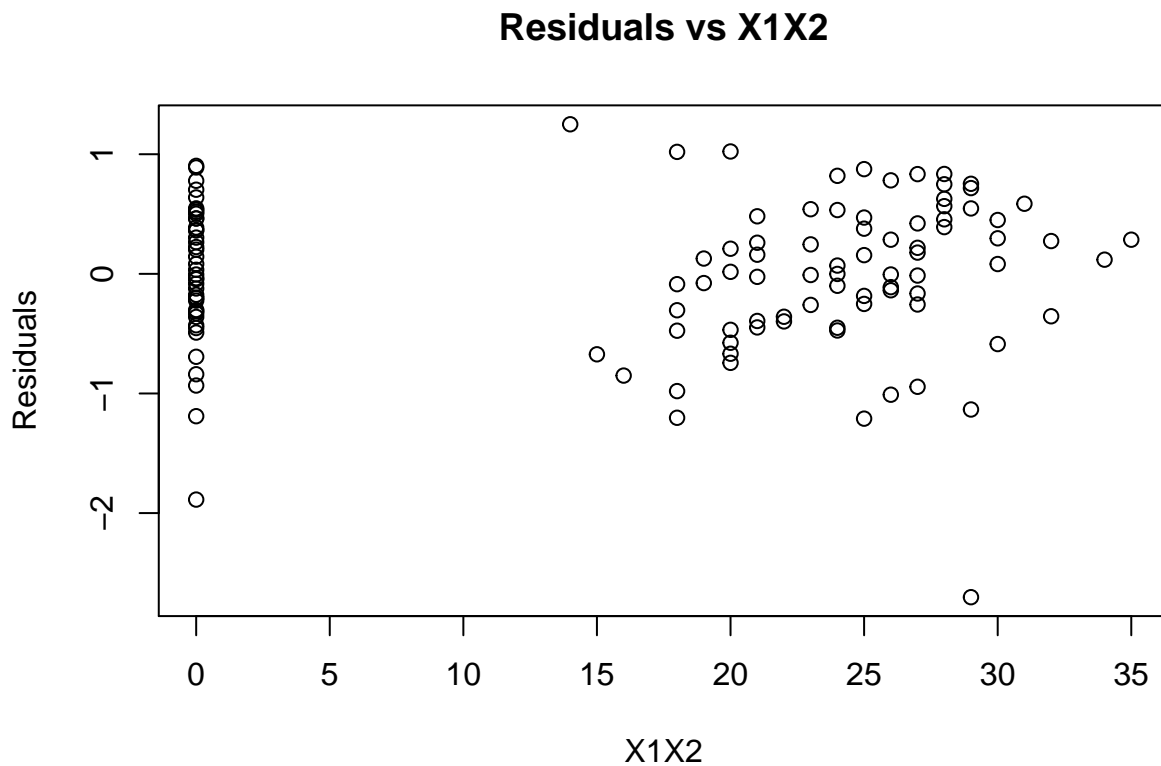
```
## Analysis of Variance Table
##
## Model 1: y ~ x1
## Model 2: y ~ x1 + x2
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    118 45.818
## 2    117 45.577  1   0.24071 0.6179 0.4334
```

The null hypothesis is $H_0 : \beta_0 = \beta_1 = \beta_2$. The alternative hypothesis is $H_a : \beta_i \neq \beta j$ for some $i \neq j$. We reject H_0 if the p-value is less than $\alpha = 0.05$.

We fail to reject H_0 becasue the $0.4334 > 0.05$. This means that we can drop the quadratic term from the regression model.

Part D

```
gpa2.fit.resid <- resid(gpa2.fit)
plot(gpa2$x1*gpa2$x2, gpa2.fit.resid, main = "Residuals vs X1X2",
     xlab = "X1X2", ylab = "Residuals")
```

Residuals vs X1X2

The residuals are split into two clusters. Therefore, there is an interaction term.

Problem 8.20

Use the data from Problem 8.20.

Part A

```
gpa2.fit2 <- lm(y~x1+x2+x1*x2, data = gpa2)
```

The fitted regressional model is $Y_i = 3.23 - 0.003x_1 - 1.650x_2 + 0.062x_1x_2$.

Part B

```
anova(gpa2.fit, gpa2.fit2)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x1 + x2
## Model 2: y ~ x1 + x2 + x1 * x2
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    117 45.577
## 2    116 43.506  1    2.0713 5.5226 0.02046 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis is $H_0 : \beta_0 = \beta_1 = \beta_2$. The alternative hypothesis is $H_a : \beta_i \neq \beta j$ for some $i \neq j$. We reject H_0 if the p-value is less than $\alpha = 0.05$.

We reject H_0 becasue the $0.02046 < 0.05$. This means that we cannot drop the quadratic term from the regression model.