

Chapter 10(Edition 8): 10.7, 10.8, 10.10

Jeremy Ling & Emmanuel Mejia

May 15, 2018

```
# loading libraries
library(car)
library(gplots)
#code function
coded=function(x) #a function to code variable x
{
  ifelse(x=="+", 1, -1)
}
```

10.7

The brake horsepower developed by an automobile engine on a dynamometer is thought to be a function of the engine speed in revolutions per minute (rpm), the road octane number of the fuel, and the engine compression. An experiment is run in the laboratory and the data that follow are collected:

```
# creating data table
Brake.HP = c(225,212,229,222,219,278,246,237,233,224,223,230)
RPM = c(2000,1800,2400,1900,1600,2500,3000,3200,2800,3400,1800,2500)
RON = c(90,94,88,91,86,96,94,90,88,86,90,89)
Compression = c(100,95,110,96,100,110,98,100,105,97,100,104)
automob = data.frame(Brake.HP,RPM,RON,Compression)
```

(a) Fit a multiple regression model to these data.

```
# linear regression
brake.lm <- lm(formula = Brake.HP ~ RPM + RON + Compression,
               data = automob)
```

(b) Test for significance of regression. What conclusions can you draw?

```
# summary output
summary(brake.lm)
```

```
##
## Call:
## lm(formula = Brake.HP ~ RPM + RON + Compression, data = automob)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.3282  -3.1701   0.8026   2.9400  11.5558
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.660e+02  9.267e+01  -2.871  0.02081 *
## RPM          1.071e-02  4.483e-03   2.390  0.04388 *
## RON          3.135e+00  8.444e-01   3.712  0.00594 **
## Compression  1.867e+00  5.345e-01   3.494  0.00816 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.812 on 8 degrees of freedom
## Multiple R-squared:  0.8065, Adjusted R-squared:  0.734
## F-statistic: 11.12 on 3 and 8 DF,  p-value: 0.00317
```

To check for regression significance, we test for the following:

$$H_0 : \beta_1 = \dots = \beta_p = 0$$

$$H_1 : \beta_i \neq 0, \text{ where } i = 1, \dots, p$$

From our summary output, we find that the p-value for our F test is .00317. Rejecting the null hypothesis at $\alpha = .05$, we find that at least one of the regressors is significant in our model.

(c) Based on t-tests, do you need all three regressor variables in the model? To determine whether or not our regressors are statistically significant to our model, we test for the following:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

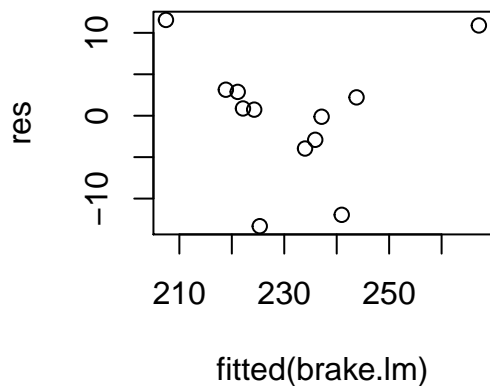
For each regressor, we reject the null hypothesis and conclude that we need all three regressor variables in the model.

10.8

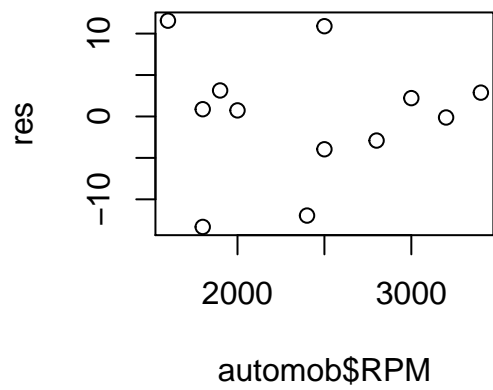
Analyze the residuals from the regression model in Problem 10.7. Comment on model adequacy.

```
# defining residuals
res <- automob$Brake.HP - fitted(brake.lm)

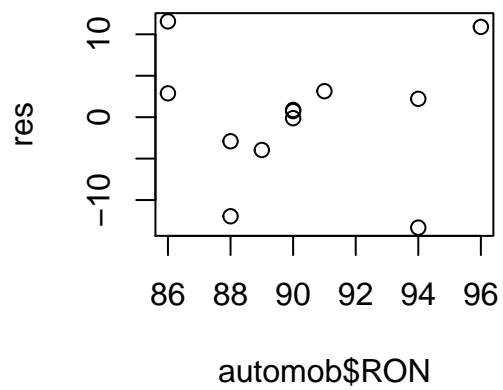
# plotting residuals
plot(fitted(brake.lm), res)
```



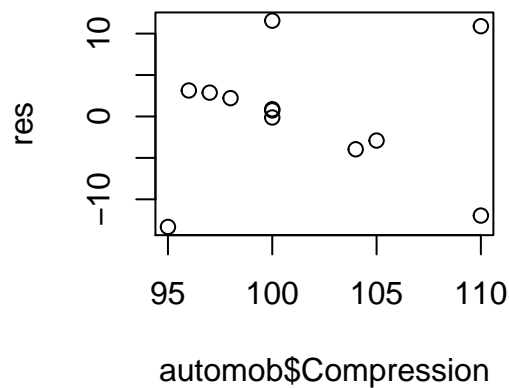
```
plot(automob$RPM, res)
```



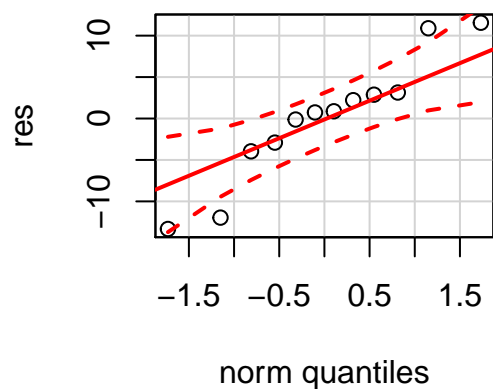
```
plot(automob$RON,res)
```



```
plot(automob$Compression,res)
```



```
# checking normality
qqPlot(res)
```



After analyzing our residuals. We see that residuals for our model is good because there is no pattern and that we have random residuals displayed. We check the residuals for our predictors and all residuals are random and patternless. We quickly check normality and we seem some residuals are slightly off, normality may be off but because of the residuals our model is good.

10.10

Consider the 2^4 factorial experiment in Example 6.2. Suppose that the last observation is missing. Reanalyze the data and draw conclusions. How do these conclusions compare with those from the original example?

```
A <- rep(x = c("-", "+"), times = 8)
B <- rep(x = c("-", "+"), each = 2, times = 4)
```

```

C <- rep(x = c("-", "+"), each = 4, times = 2)
D <- rep(x = c("-", "+"), each = 8)
FilRate <- c(45,71,48,65,68,60,80,65,43,100,45,104,75,86,70,96)
chemical <- data.frame(A,B,C,D,FilRate)

```

```

qqnorm(aov(FilRate~A*B*C*D, chemical), label = TRUE)#AC AD

```

#regular Linear Regression

```

chem.lm <- lm(FilRate ~ A*C + A*D, chemical); summary(chem.lm)

```

```

##
## Call:
## lm(formula = FilRate ~ A * C + A * D, data = chemical)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.3750 -1.5000  0.0625  2.9062  5.7500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    46.250      2.705   17.098 9.88e-09 ***
## A+              23.125      3.825    6.045 0.000124 ***
## C+              28.000      3.123    8.964 4.29e-06 ***
## D+             -2.000      3.123   -0.640 0.536368
## A+:C+          -36.250      4.417   -8.206 9.41e-06 ***
## A+:D+           33.250      4.417    7.527 2.00e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.417 on 10 degrees of freedom
## Multiple R-squared:  0.966, Adjusted R-squared:  0.9489
## F-statistic: 56.74 on 5 and 10 DF, p-value: 5.14e-07

```

```

chemicalMiss0 <- chemical[-c(16),]
qqnorm(aov(FilRate ~ A*B*C*D, chemicalMiss0), label = TRUE)#AC AD

```

#missing observation Linear Regression

```

chemmo.lm <- lm(FilRate ~ A*C + A*D, chemicalMiss0); summary(chemmo.lm)

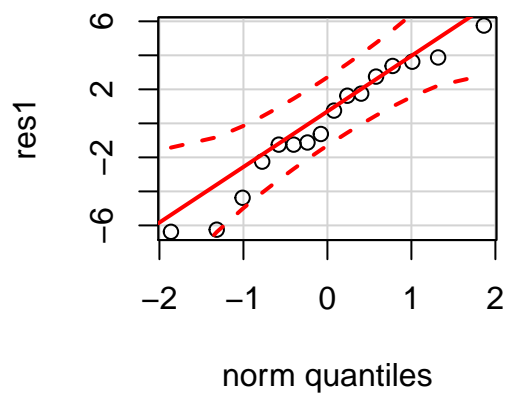
```

```

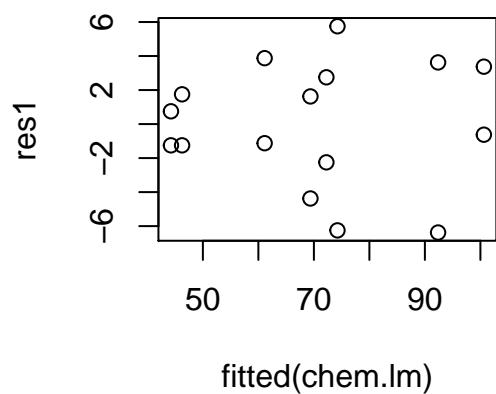
##
## Call:
## lm(formula = FilRate ~ A * C + A * D, data = chemicalMiss0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -6.25  -1.75   0.10   2.25   5.75
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    46.250      2.693   17.172 3.47e-08 ***
## A+              23.850      3.872    6.160 0.000167 ***
## C+              28.000      3.110    9.003 8.51e-06 ***
## D+             -2.000      3.110   -0.643 0.536209
## A+:C+          -37.700      4.613   -8.173 1.87e-05 ***
## A+:D+           31.800      4.613    6.894 7.12e-05 ***

```

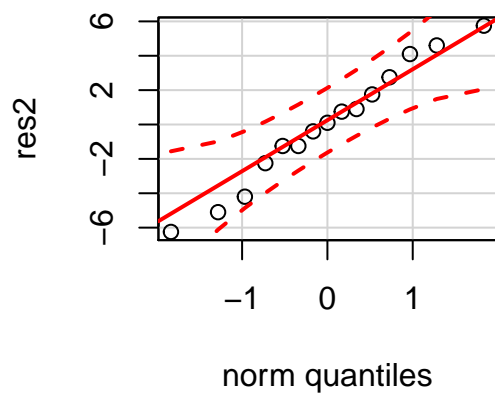
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.398 on 9 degrees of freedom
## Multiple R-squared:  0.9653, Adjusted R-squared:  0.946
## F-statistic: 50.03 on 5 and 9 DF,  p-value: 2.686e-06
#residual analysis for regular linear model
res1 = chemical$FilRate - fitted(chem.lm)
qqPlot(res1)
```



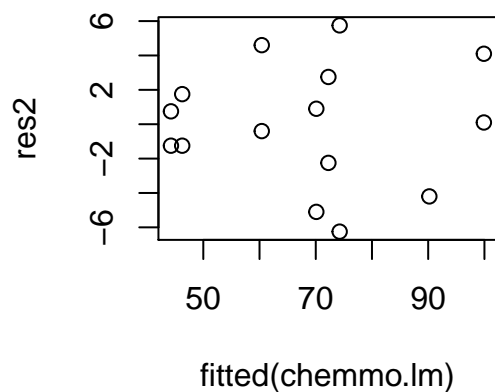
```
plot(fitted(chem.lm), res1)
```



```
#residual analysis for missing observation linear model
res2 = chemicalMiss0$FilRate - fitted(chemmo.lm)
qqPlot(res2)
```



```
plot(fitted(chemmo.lm), res2)
```



After dropping the last observation, we still are dealing with four factors and must pick the most significant factors to the response variable. We choose to use half normal probability plot to pick out those important factors. We still get interaction effects AC and AD from the plot. We check the linear model's p-values and check and compare the residual plots Both regression types have excellent residuals so we can state both regression models are good. As we mentioned earlier AC and AD are important and we see that those effects are importantly the same in the regular linear regression without the missing observation.