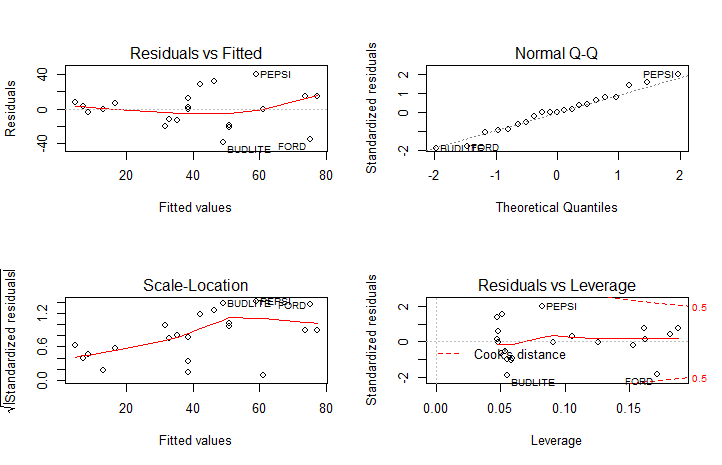
**Question 1**

(a)

**> model1<-lm(MILIMP~log(SPEND),data=advertspend)**

**> par(mfrow=c(2,2))**

**> plot(model1)**



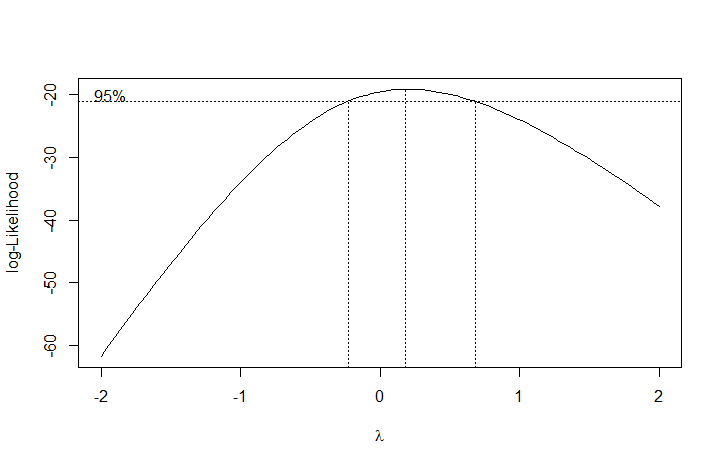
(b) The diagnostic plots indicate the problems of heteroscedasticity

(c)

**> library(MASS)**

**> par(mfrow=c(1,1))**

**> boxcox(model1)**



(d)

**> boxcox(model1,plotit = FALSE)**

$x

[1] -2.0 -1.9 -1.8 -1.7 -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5

[17] -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1

[33] 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0

$y

[1] -61.75260 -58.68416 -55.66156 -52.69010 -49.77590 -46.92597 -44.14834 -41.45215

[9] -38.84771 -36.34659 -33.96154 -31.70641 -29.59590 -27.64517 -25.86924 -24.28229

[17] -22.89676 -21.72252 -20.76603 -20.02979 -19.51213 -19.20733 -19.10613 -19.19654

[25] -19.46465 -19.89559 -20.47426 -21.18596 -22.01685 -22.95417 -23.98642 -25.10334

[33] -26.29590 -27.55620 -28.87737 -30.25346 -31.67929 -33.15040 -34.66290 -36.21342

[41] -37.79903

**> index = which.max(boxcox(model1,plotit = FALSE)$y)**

**> lamda = boxcox(model1,plotit = FALSE)$x[index]**

**> lamda**

[1] 0.2

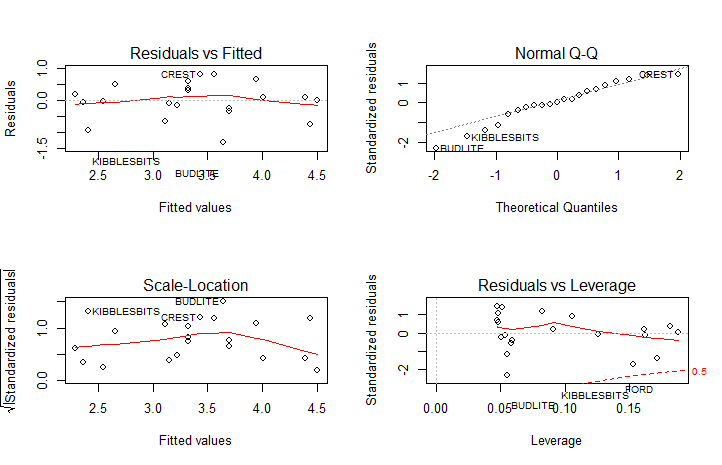
(e) When lamda = 0, the transformation of y is log(y).

(f)

**> model2<-lm(log(MILIMP)~log(SPEND),data=advertspend)**

**> par(mfrow=c(2,2))**

**> plot(model2)**



The problem of heteroscedasticity has been improved

**Question 2**

(a)

(b)

**> fit1<-lm(log(Displacement)~log(Length)+log(Beam)+log(Draft),data=destroyers)**

**> summary(fit1)**

Call:

lm(formula = log(Displacement) ~ log(Length) + log(Beam) + log(Draft),

data = destroyers)

Residuals:

Min 1Q Median 3Q Max

-0.181718 -0.038040 0.000452 0.051843 0.150074

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.30889 0.78417 -6.770 1.97e-07 \*\*\*

log(Length) 0.93189 0.22313 4.176 0.000247 \*\*\*

log(Beam) 1.98125 0.19008 10.423 2.56e-11 \*\*\*

log(Draft) 0.05670 0.07854 0.722 0.476124

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

Residual standard error: 0.07122 on 29 degrees of freedom

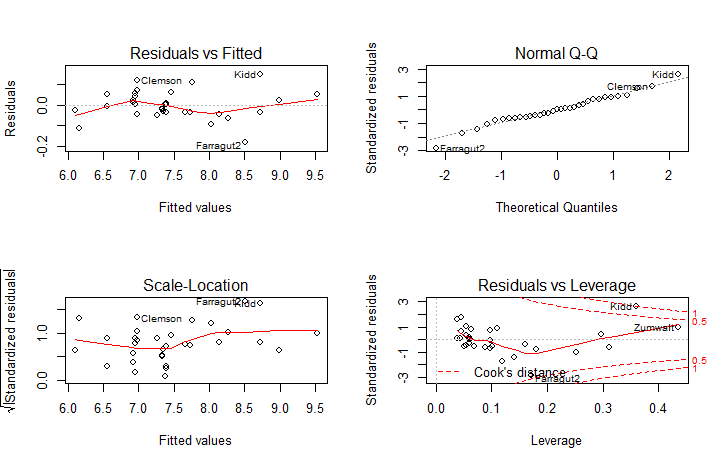
Multiple R-squared: 0.9927, Adjusted R-squared: 0.9919

F-statistic: 1309 on 3 and 29 DF, p-value: < 2.2e-16

(c)

**> par(mfrow=c(2,2))**

**> plot(fit1)**



The model appears to have linearity and homoscedasticity, but the residuals vs. leverage plot and scale-location plot indicate Kidd and Farragut2 may be outliers

(d)

**> x = data.frame(Length=500, Beam=69.5, Draft=24)**

**> predict(fit1, x, interval='prediction',level=0.95)**

fit lwr upr

1 9.065737 8.88916 9.242315

(e)

**> lower\_bound = exp(1)^(8.88916)**

**> lower\_bound**

[1] 7252.924

**> higher\_bound = exp(1)^(9.242315)**

**> higher\_bound**

[1] 10324.91 Yes. The prediction interval includes 8400.

(f)

**> confint(fit1,level=0.95)**

2.5 % 97.5 %

(Intercept) -6.9126990 -3.7050857

log(Length) 0.4755299 1.3882442

log(Beam) 1.5924891 2.3700103

log(Draft) -0.1039294 0.2173255

If Displacement is proportional to the product of Length, Beam and Draft, then all estimates of betas should be one. But only the 95% confidence interval of beta of log(Length) contains one.

(g)

Suggested model: beta1 = 1, beta2 = 2, beta3 = 0

So Displacement is proportional to the product of Length and (Beam)^2.

**Question 3**

(a)

**> library(faraway)**

**> foo1<-lm(hipcenter~1,data=seatpos)**

**> predictors<- ~ Age + Weight + Ht + Leg**

**> add1(foo1,predictors,test='F')**

Single term additions

Model:

hipcenter ~ 1

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 131639 311.71

Age 1 5541 126098 312.07 1.582 0.2166

Weight 1 53975 77664 293.66 25.020 1.493e-05 \*\*\*

Ht 1 84023 47616 275.07 63.526 1.831e-09 \*\*\*

Leg 1 81568 50071 276.98 58.646 4.587e-09 \*\*\*

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

**> foo1<-update(foo1, .~.+Ht) #Ht has the greatest F-value**

**> add1(foo1,predictors,test='F')**

Single term additions

Model:

hipcenter ~ Ht

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 47616 275.07

Age 1 2353.51 45262 275.14 1.8199 0.1860

Weight 1 195.86 47420 276.91 0.1446 0.7061

Leg 1 2781.10 44835 274.78 2.1711 0.1496

**> foo1**

Call:

lm(formula = hipcenter ~ Ht, data = seatpos)

Coefficients:

(Intercept) Ht

556.255 -4.265

Since all F-values now are smaller than 3, we stop adding new variables into the model. The final model includes Ht.

(b)

**> foo2<-lm(hipcenter~Age + Weight + Ht + Leg,data=seatpos)**

**> drop1(foo2,test='F')**

Single term deletions

Model:

hipcenter ~ Age + Weight + Ht + Leg

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 41891 276.20

Age 1 2498.3 44390 276.40 1.9681 0.16999

Weight 1 46.8 41938 274.24 0.0369 0.84887

Ht 1 3642.9 45534 277.37 2.8697 0.09968 .

Leg 1 3370.8 45262 277.14 2.6553 0.11271

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

> foo2<-update(foo2, .~.-Weight)

> drop1(foo2,test='F')

Single term deletions

Model:

hipcenter ~ Age + Ht + Leg

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 41938 274.24

Age 1 2896.6 44835 274.78 2.3483 0.13467

Ht 1 4238.3 46176 275.90 3.4360 0.07248 .

Leg 1 3324.2 45262 275.14 2.6950 0.10988

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

**> foo2<-update(foo2, .~.-Age)**

**> drop1(foo2,test='F')**

Single term deletions

Model:

hipcenter ~ Ht + Leg

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 44835 274.78

Ht 1 5236.3 50071 276.98 4.0877 0.05089 .

Leg 1 2781.1 47616 275.07 2.1711 0.14957

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

**> foo2<-update(foo2, .~.-Leg)**

**> drop1(foo2,test='F')**

Single term deletions

Model:

hipcenter ~ Ht

Df Sum of Sq RSS AIC F value Pr(>F)

<none> 47616 275.07

Ht 1 84023 131639 311.71 63.526 1.831e-09 \*\*\*

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

Since the only remain F-value is greater than 3, we stop dropping any variables. The final model includes Ht.

(c)

**> library(leaps)**

**> foo\_x<-model.matrix(hipcenter ~ Age + Weight + Ht + Leg - 1, data=seatpos)**

**> foo\_y<-seatpos$hipcenter**

**> best<-leaps(foo\_x, foo\_y, nbest=1)**

**> best**

$which

1 2 3 4

1 FALSE FALSE TRUE FALSE

2 FALSE FALSE TRUE TRUE

3 TRUE FALSE TRUE TRUE

4 TRUE TRUE TRUE TRUE

$label

[1] "(Intercept)" "1" "2" "3" "4"

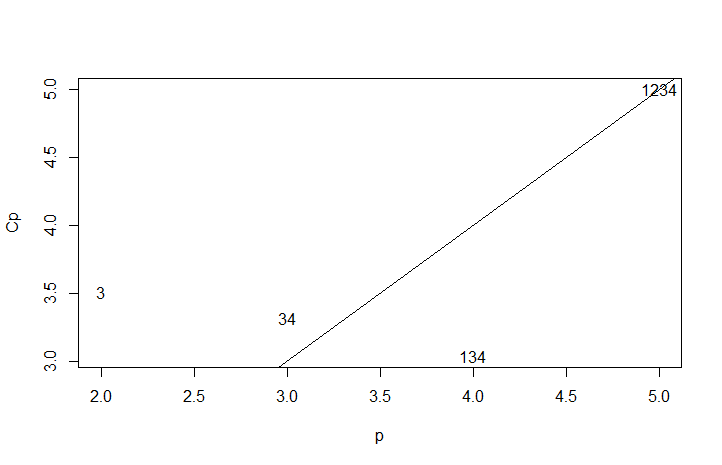
$size

[1] 2 3 4 5

$Cp

[1] 3.509521 3.318700 3.036888 5.000000

**> Cpplot(best)**



**> colnames(foo\_x)[c(1,3,4)]**

[1] "Age" "Ht" "Leg"

According to the minimum Cp, the best model includes Age, Ht and Leg.

(d)

**> stepwise<-step(object = lm(hipcenter ~ 1, data=seatpos),**

**+ scope = ~ Age + Weight + Ht + Leg,**

**+ direction= "both")**

Start: AIC=311.71

hipcenter ~ 1

Df Sum of Sq RSS AIC

+ Ht 1 84023 47616 275.07

+ Leg 1 81568 50071 276.98

+ Weight 1 53975 77664 293.66

<none> 131639 311.71

+ Age 1 5541 126098 312.07

Step: AIC=275.07

hipcenter ~ Ht

Df Sum of Sq RSS AIC

+ Leg 1 2781 44835 274.78

<none> 47616 275.07

+ Age 1 2354 45262 275.14

+ Weight 1 196 47420 276.91

- Ht 1 84023 131639 311.71

Step: AIC=274.78

hipcenter ~ Ht + Leg

Df Sum of Sq RSS AIC

+ Age 1 2896.6 41938 274.24

<none> 44835 274.78

- Leg 1 2781.1 47616 275.07

+ Weight 1 445.1 44390 276.40

- Ht 1 5236.3 50071 276.98

Step: AIC=274.24

hipcenter ~ Ht + Leg + Age

Df Sum of Sq RSS AIC

<none> 41938 274.24

- Age 1 2896.6 44835 274.78

- Leg 1 3324.2 45262 275.14

- Ht 1 4238.3 46176 275.90

+ Weight 1 46.8 41891 276.20

**> summary(stepwise)**

Call:

lm(formula = hipcenter ~ Ht + Leg + Age, data = seatpos)

Residuals:

Min 1Q Median 3Q Max

-79.715 -22.758 -4.102 21.394 60.576

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 452.1976 100.9482 4.480 8.04e-05 \*\*\*

Ht -2.3254 1.2545 -1.854 0.0725 .

Leg -6.7390 4.1050 -1.642 0.1099

Age 0.5807 0.3790 1.532 0.1347

---

Signif. codes: 0 ¡®\*\*\*¡¯ 0.001 ¡®\*\*¡¯ 0.01 ¡®\*¡¯ 0.05 ¡®.¡¯ 0.1 ¡® ¡¯ 1

Residual standard error: 35.12 on 34 degrees of freedom

Multiple R-squared: 0.6814, Adjusted R-squared: 0.6533

F-statistic: 24.24 on 3 and 34 DF, p-value: 1.426e-08