Homework 4

Name: Zixin Ouyang

1.

install.packages('alr4')

library(alr4)

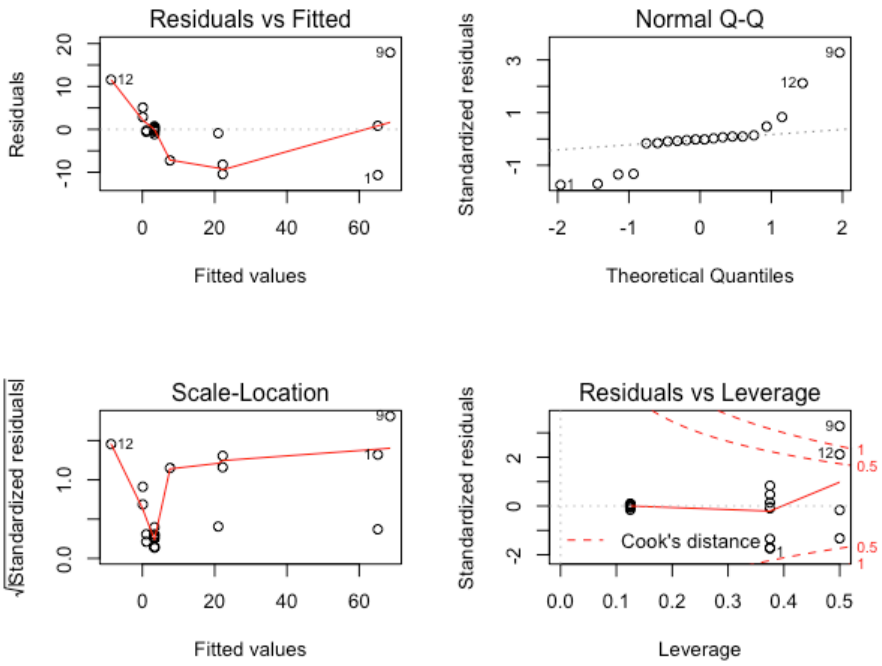
(a)

```
> lathe1mod<-lm(Life~Feed+Speed+I(Feed^2)+I(Speed^2)+Feed*Speed, data=lathe1)
> summary(lathe1mod)

Call:
lm(formula = Life ~ Feed + Speed + I(Feed^2) + I(Speed^2) + Feed *
    Speed, data = lathe1)

Residuals:
     Min       1Q   Median       3Q      Max
-10.6601  -0.9607  -0.1383   0.7062  17.9193

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.338      2.733   1.222 0.241998
Feed         -10.494      2.231  -4.703 0.000339 ***
Speed        -21.548      2.231  -9.657 1.44e-07 ***
I(Feed^2)      1.412      2.617   0.540 0.597837
I(Speed^2)    17.392      2.617   6.647 1.10e-05 ***
Feed:Speed    10.975      2.733   4.016 0.001274 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.729 on 14 degrees of freedom
Multiple R-squared:  0.9267,    Adjusted R-squared:  0.9005
F-statistic:  35.4 on 5 and 14 DF,  p-value: 1.831e-07
```

The p-value of interaction term is 0.001274, which is less than 0.05, so it is statistically significant.

(b)

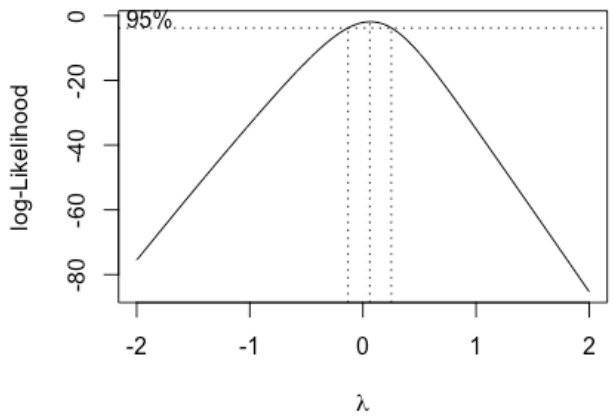
```
> par(mfrow=c(2,2))
> plot(lathe1mod)
```



(c)According to the Residuals vs Fitted plot, the model shows nonlinearity and heteroscedasticity. The line in the normal Q-Q plot is not straight, so it is not normal. The Residuals vs Leverage plot indicates that observation 9 is influential.

(d)

```
> library(MASS)
> boxcox(lathe1mod)
```



(e)

```
> lathe1mod2$x[which.max(lathe1mod2$y)]
[1] 0.1
```

(f)

The most "simple" λ value that is still within the confidence limits is 0. This corresponds to the log-transformation of the response variable.

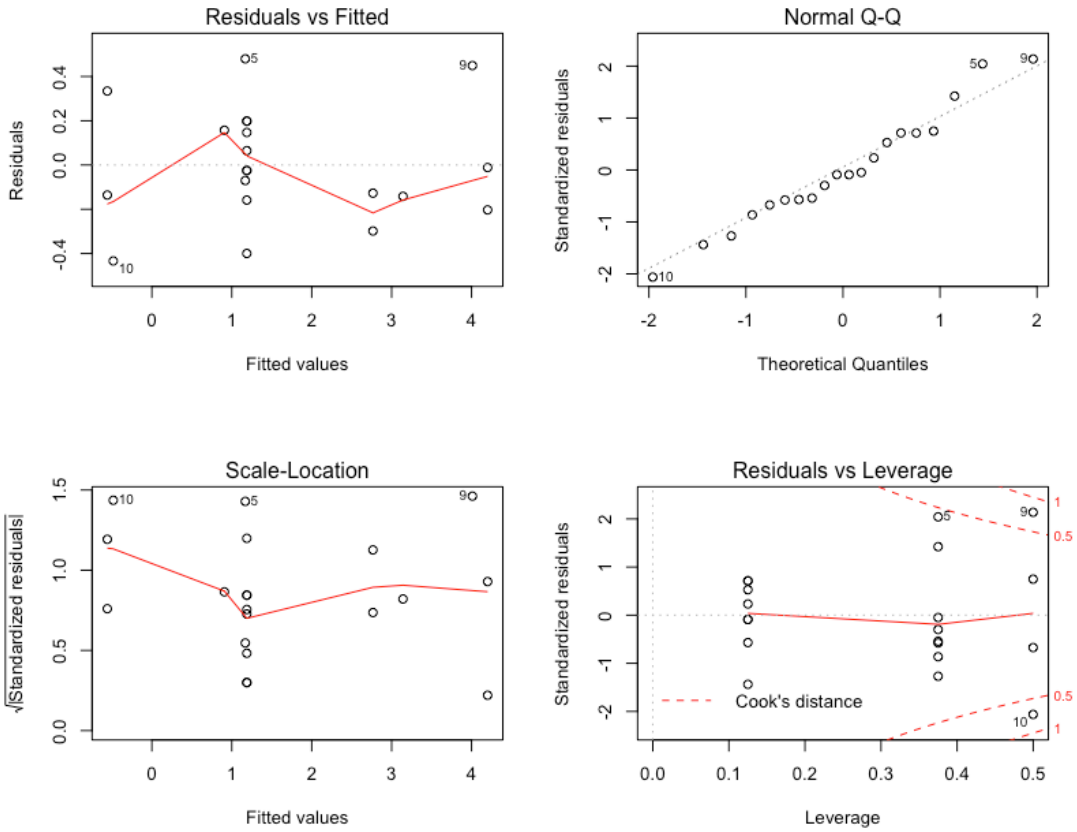
(g)

```
> lathe1mod3<-lm(log(Life)~Feed+Speed+I(Feed^2)+I(Speed^2)+Feed*Speed, data=lathe1)
> summary(lathe1mod3)

Call:
lm(formula = log(Life) ~ Feed + Speed + I(Feed^2) + I(Speed^2) +
    Feed * Speed, data = lathe1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.43349 -0.14576 -0.02494  0.16748  0.47992

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.18809    0.10508  11.307 2.00e-08 ***
Feed        -0.79023    0.08580  -9.210 2.56e-07 ***
Speed       -1.58902    0.08580 -18.520 3.04e-11 ***
I(Feed^2)    0.41851    0.10063   4.159 0.000964 ***
I(Speed^2)   0.28808    0.10063   2.863 0.012529 *
Feed:Speed  -0.07286    0.10508  -0.693 0.499426
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2972 on 14 degrees of freedom
Multiple R-squared:  0.9702,    Adjusted R-squared:  0.9596
F-statistic: 91.24 on 5 and 14 DF,  p-value: 3.551e-10
```

The p-value of the interaction term is greater than 0.05, so it is not statistically significant.

(h)

```
> plot(lathe1mod3)
```



The plots have improved because there are no nonlinearity or heteroscedasticity, and it is normal. The Residuals vs Leverage plot shows that observation 9 and observation 10 is influential.

2.

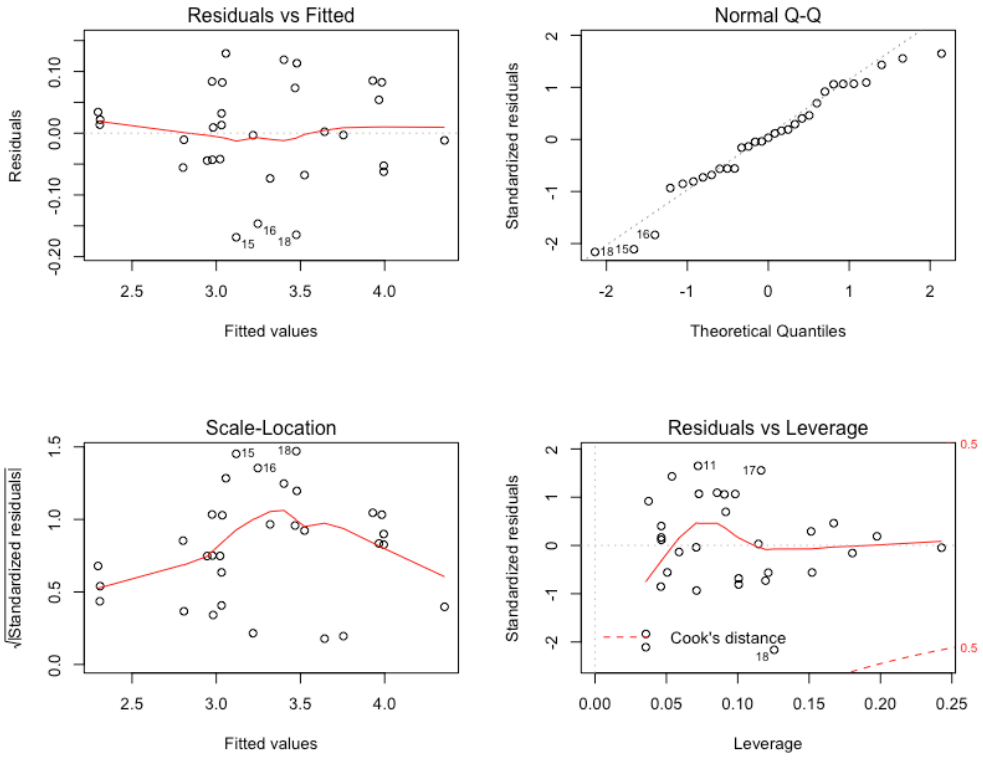
(a) Taking log: $\ln(\text{Volume}) = \ln(\gamma) + \beta 1 \ln(\text{Girth}) + \beta 2 \ln(\text{Height}) + \ln(e)$

(b)

```
> treesmod<-lm(log(Volume)~log(Girth)+log(Height), data=trees)
> summary(treesmod)

Call:
lm(formula = log(Volume) ~ log(Girth) + log(Height), data = trees)

Residuals:
      Min        1Q    Median        3Q       Max
-0.168561 -0.048488  0.002431  0.063637  0.129223

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.63162    0.79979  -8.292 5.06e-09 ***
log(Girth)   1.98265    0.07501  26.432  < 2e-16 ***
log(Height)  1.11712    0.20444   5.464 7.81e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08139 on 28 degrees of freedom
Multiple R-squared:  0.9777,    Adjusted R-squared:  0.9761
F-statistic: 613.2 on 2 and 28 DF,  p-value: < 2.2e-16
```

(c)

```
> plot(treesmod)
```



The variance is constant, the error is normal, and the problem of nonlinearity does not exist.

(d)

```
> confint(treesmod)
                2.5 %    97.5 %
(Intercept) -8.269912 -4.993322
log(Girth)   1.828998  2.136302
log(Height)  0.698353  1.535894
```

The 95% confidence intervals for β_1 is (1.828998, 2.136302), which contains the theoretical value 2. The 95% confidence intervals for β_2 is (0.698353, 1.535894), which contains the theoretical value 1.

(e)

```
> newtree<-data.frame(Girth=15.5, Height=83)
> loginterval<-predict(treesmod, newdata=newtree, interval="prediction")
> loginterval
       fit      lwr      upr
1 3.738899 3.566246 3.911552
```

(f)

```
> exp(loginterval)
       fit      lwr      upr
1 42.05167 35.38351 49.97646
```

3.

(a) the independent variables in the final model: SSF, Sex

```
> aismod1<-lm(Bfat~1, data=ais)
> indep.vars<-~Sex+Ht+Wt+LBM+BMI+SSF
```

```
> add1(aismod1,indep.vars,test='F' )
Single term additions

Model:
Bfat ~ 1
       Df Sum of Sq    RSS    AIC   F value    Pr(>F)
<none>              7701.1 737.45
Sex     1    3733.1 3968.0 605.51  188.1568 < 2.2e-16 ***
Ht      1     272.3 7428.9 732.18    7.3295  0.007370 **
Wt      1       0.0 7701.1 739.45    0.0000  0.998176
LBM     1    1008.3 6692.8 711.10   30.1326 1.214e-07 ***
BMI     1     270.9 7430.2 732.22    7.2921  0.007519 **
SSF     1    7142.0  559.1 209.64 2554.8760 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> aismod1<-update(aismod1, .~.+SSF)
```

(SSF has biggest F)

```
> add1(aismod1,indep.vars,test='F' )
Single term additions

Model:
Bfat ~ SSF
       Df Sum of Sq    RSS     AIC F value    Pr(>F)
<none>              559.09 209.644
Sex     1    315.09 244.00  44.155 256.984 < 2.2e-16 ***
Ht      1    110.36 448.73 167.228  48.940 3.929e-11 ***
Wt      1    174.40 384.69 136.123  90.216 < 2.2e-16 ***
LBM     1    210.66 348.43 116.122 120.318 < 2.2e-16 ***
BMI     1    127.14 431.95 159.529  58.572 8.278e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> aismod1<-update(aismod1, .~.+Sex)
> add1(aismod1,indep.vars,test='F' )
Single term additions

Model:
Bfat ~ SSF + Sex
       Df Sum of Sq     RSS    AIC F value Pr(>F)
<none>               244.00 44.155
Ht      1   0.61479 243.38 45.646  0.5002 0.4803
Wt      1   0.26549 243.73 45.935  0.2157 0.6429
LBM     1   0.79043 243.21 45.500  0.6435 0.4234
BMI     1   0.04465 243.95 46.118  0.0362 0.8492
```

(b) the independent variables in the final model: Sex, Ht, Wt, LBM, SSF

```
> aismod2<-lm(Bfat~Sex+Ht+Wt+LBM+BMI+SSF, data=ais)
> drop1(aismod2, test='F')
Single term deletions

Model:
Bfat ~ Sex + Ht + Wt + LBM + BMI + SSF
       Df Sum of Sq    RSS      AIC  F value    Pr(>F)
<none>              105.03 -118.120
Sex     1    22.694 127.72  -80.603  42.1354 6.888e-10 ***
Ht      1     1.719 106.74 -116.842   3.1909    0.0756 .
Wt      1    59.652 164.68  -29.264 110.7556 < 2.2e-16 ***
LBM     1   136.496 241.52   48.096 253.4299 < 2.2e-16 ***
BMI     1     0.695 105.72 -118.788   1.2907    0.2573
SSF     1    24.310 129.34  -78.064  45.1351 1.969e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> aismod2<-update(aismod2, .~.-BMI)
```

(BMI has least F)

```
> drop1(aismod2, test='F')
Single term deletions

Model:
Bfat ~ Sex + Ht + Wt + LBM + SSF
       Df Sum of Sq    RSS      AIC  F value    Pr(>F)
<none>              105.72 -118.788
Sex     1    22.163 127.88  -82.343  41.0886 1.061e-09 ***
Ht      1     4.136 109.86 -113.035   7.6687  0.006158 **
Wt      1   134.239 239.96   44.786 248.8715 < 2.2e-16 ***
LBM     1   137.661 243.38   47.646 255.2150 < 2.2e-16 ***
SSF     1    24.709 130.43  -78.362  45.8081 1.474e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(c)

```
> library(leaps)
> bestmods<-regsubsets(Bfat~Sex+Ht+Wt+LBM+BMI+SSF,data=ais,nbest=1, nvmax=6)
```

```
> summary(bestmods)
Subset selection object
Call: regsubsets.formula(Bfat ~ Sex + Ht + Wt + LBM + BMI + SSF, data = ais,
    nbest = 1, nvmax = 6)
6 Variables  (and intercept)
    Forced in Forced out
Sex     FALSE      FALSE
Ht      FALSE      FALSE
Wt      FALSE      FALSE
LBM     FALSE      FALSE
BMI     FALSE      FALSE
SSF     FALSE      FALSE
1 subsets of each size up to 6
Selection Algorithm: exhaustive
         Sex Ht  Wt  LBM BMI SSF
1  ( 1 ) " " " " " " " " " " "*"
2  ( 1 ) " " " " "*" "*" " " " "
3  ( 1 ) " " " " "*" "*" " " "*"
4  ( 1 ) "*" " " "*" "*" " " "*"
5  ( 1 ) "*" "*" "*" "*" " " "*"
6  ( 1 ) "*" "*" "*" "*" "*" "*"
```

(d) the independent variables in the final model: SSF, SEX

```
> step(aismod1,indep.vars)
Start:  AIC=44.16
Bfat ~ SSF + Sex

       Df Sum of Sq    RSS    AIC
<none>               244.0  44.16
+ LBM   1       0.8  243.2  45.50
+ Ht    1       0.6  243.4  45.65
+ Wt    1       0.3  243.7  45.94
+ BMI   1       0.0  244.0  46.12
- Sex   1     315.1  559.1 209.64
- SSF   1    3724.0 3968.0 605.51

Call:
lm(formula = Bfat ~ SSF + Sex, data = ais)

Coefficients:
(Intercept)          SSF          Sex
     1.1307       0.1579       2.9844
```