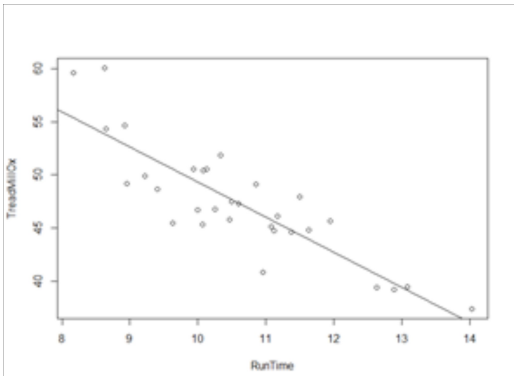
Homework 6

Stat 217- Due Monday 16th November

Researchers interested in studying methods for measuring fitness, collected data on 31 male volunteers. They measured various things like treadmill oxygen consumption (TreadMillOx, in ml per kg per minute), minutes to run 1.5 miles (RunTime), maximum pulse during 1.5 mile run (RunPulse, in beats per minute), resting pulse rate (RestPulse, beats per minute), body weight (BodyWeight, in kg), and age (Age, in years).

1. Researchers were initially interested in the relationship between a runners RunTime and their treadmill oxygen consumption. A simple linear regression model was fit using RunTime as the explanatory variable for TreadMillOx.

```
> tm1<-lm(TreadMillOx~RunTime,data=treadmill)
> summary(tm1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  82.4177     3.8544  21.383  < 2e-16
RunTime      -3.3102     0.3611  -9.167 4.57e-10

Residual standard error: 2.744 on 29 degrees of freedom
Multiple R-squared:  0.7434,  Adjusted R-squared:  0.7346
F-statistic: 84.03 on 1 and 29 DF,  p-value: 4.572e-10

> confint(tm1)
                2.5 %    97.5 %
(Intercept) 74.534484 90.300830
RunTime     -4.048784 -2.571671

> predict(tm1,newdata=data.frame(RunTime=11),interval="confidence")
       fit      lwr      upr
1 46.00515 44.95181 47.0585

> predict(tm1,newdata=data.frame(RunTime=11),interval="prediction")
       fit      lwr      upr
1 46.00515 40.29474 51.71557

> predict(tm1,newdata=data.frame(RunTime=14),interval="confidence")
       fit     lwr      upr
1 36.07447 33.3591 38.78985

> predict(tm1,newdata=data.frame(RunTime=14),interval="prediction")
       fit      lwr      upr
1 36.07447 29.83968 42.30926
```



   a. Report the estimated model.

   b. Interpret the RunTime slope coefficient from the estimated model.

   c. Predict the treadmill oxygen consumption for a new runner with a run time of 14 minutes

d. Use the output to provide and interpret an uncertainty interval for your answer to part c.

e. Predict the true mean treadmill oxygen value for a run time of 11 minutes.

f. Use the output to provide and interpret an uncertainty interval for your answer to part e.

2. Researchers began to question if using only run time to predict treadmill oxygen consumption was best. They fit a multiple linear regression model that also included running pulse (RunPulse), body weight (BodyWeight), and Age.

```
> mlr1<-lm(TreadMillOx~RunTime+RunPulse+BodyWeight+Age,data=treadmill)
> summary(mlr1)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 115.65011   11.22410  10.304 1.13e-10
RunTime      -2.77220    0.36487  -7.598 4.59e-08
RunPulse     -0.12929    0.05084  -2.543  0.01728
BodyWeight   -0.04940    0.05639  -0.876  0.38908
Age          -0.27622    0.09931  -2.781  0.00994

Residual standard error: 2.451 on 26 degrees of freedom
Multiple R-squared:  0.8165,            Adjusted R-squared:  0.7883
F-statistic: 28.92 on 4 and 26 DF,  p-value: 3.107e-09
```

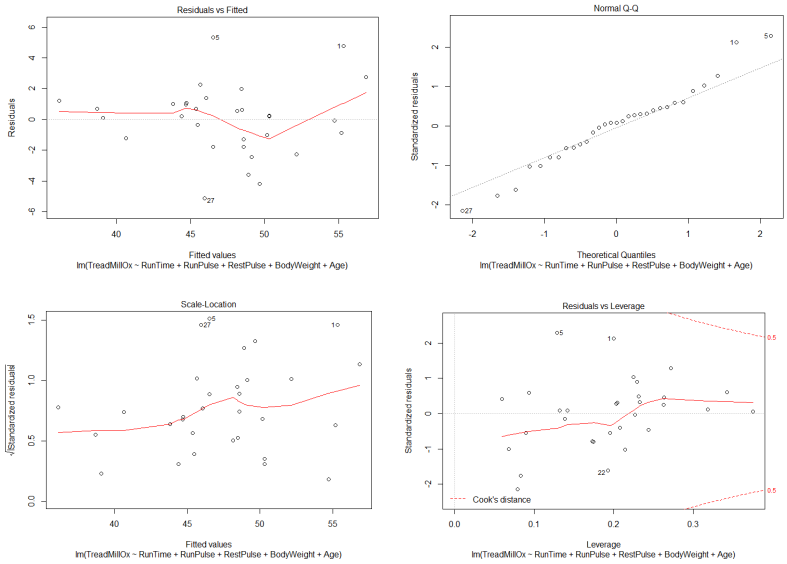
a. Report the estimated model.

b. Interpret the RunTime slope coefficient from this estimated model.

c. The MLR model with 4 predictors has an R^2 of 0.8165 and the SLR model with 1 predictor has an R^2 of 0.7434. Does this definitely mean the MLR model is better? Why or why not?

d. The MLR model with 4 predictors has an adjusted R^2 of 0.7883 and the SLR model with 1 predictor has an adjusted R^2 of 0.7346. Does this definitely mean the MLR model is better? Why or why not?

e. The diagnostic plots for this model are below. Use them to asses some of the assumptions. Be sure to justify.



Linear Relationship:

Constant Variance:

Normality of Residuals:

No Influential Points:

What are 3 other assumptions that need to be met?

What is multicollinearity?