HW11 KEY

36 points total, 2 points per problem part unless otherwise noted.

Q1 Treadmill

```
library(car)
InData <- read.csv("C:/hess/STAT511_FA11/ASCII-comma/CH11/ex11-22.txt", quote = " ' ")
str(InData)
```

```
## 'data.frame':    20 obs. of  2 variables:
##  $ Treadmill: num  7.5 7.8 7.9 8.1 8.3 8.7 8.9 9.2 9.4 9.8 ...
##  $ X10.K    : num  43.5 45.2 44.9 41.1 43.8 44.4 42.7 43.1 41.8 43.7 ...
```

```
#1A
Fit <- lm(X10.K ~ Treadmill, data = InData)
summary(Fit)
```

```
##
## Call:
## lm(formula = X10.K ~ Treadmill, data = InData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9440 -1.5788  0.1860  0.7863  4.5603
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  59.9211     3.1166  19.226 1.90e-13 ***
## Treadmill    -1.9601     0.3164  -6.194 7.59e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.921 on 18 degrees of freedom
## Multiple R-squared:  0.6807, Adjusted R-squared:  0.6629
## F-statistic: 38.37 on 1 and 18 DF,  p-value: 7.589e-06
```

```
#1B
plot(X10.K ~ Treadmill, data = InData)
abline(coef(Fit))
```


```r
#1C
confint(Fit)
```

```
##                  2.5 %    97.5 %
## (Intercept) 53.373295 66.468942
## Treadmill   -2.624957 -1.295313
```

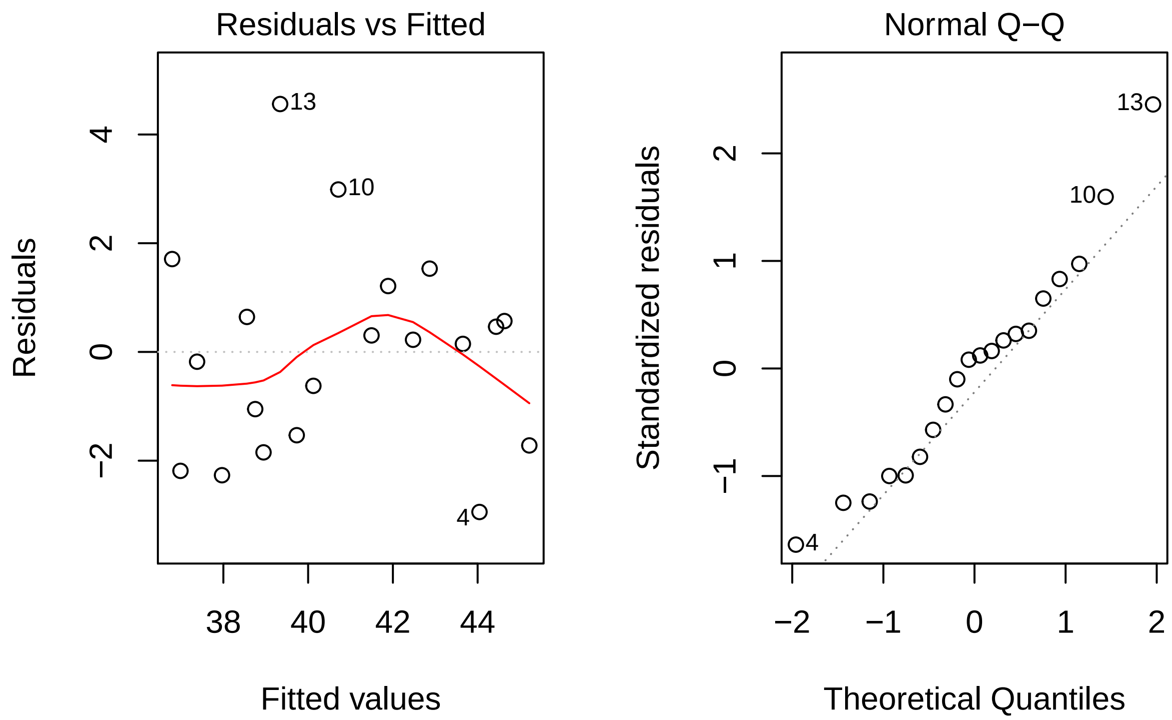
1C. (**4 pts**)
Estimated slope = -1.96, 95% CI = (-2.625, -1.295)
A 1 min increase in Treamill time is associated with a predicted **decrease** of 1.96 min in 10km race time.

1D. R2 = 0.68. Hence 68% of variability in 10km race time is explained by the linear regresison on treadmill time.

```r
#1E
newdata = data.frame(Treadmill = 11)
predict(Fit, newdata = list(Treadmill = 11), interval = "prediction")
```

```
##        fit      lwr      upr
## 1 38.35963 34.14223 42.57704
```

```r
#1F
par(mfrow = c(1,2))
plot(Fit, which =c(1,2))
```




```
#1G
outlierTest(Fit, cutoff = 1)
```

```
##    rstudent unadjusted p-value Bonferonni p
## 13 2.925728          0.0094335      0.18867
```

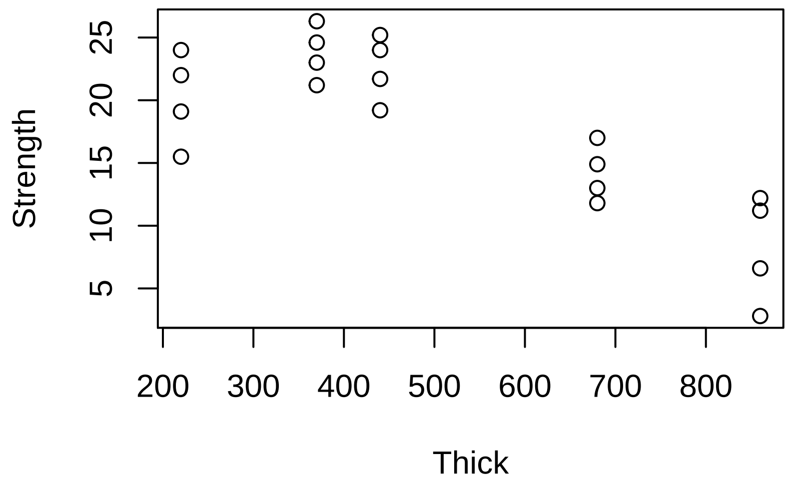
1G. Bonferonni p-value = 0.18867, Fail to Reject H0.
Hence we cannot conlude that subject 13 is an outlier.

Q2 Steel Quadratic

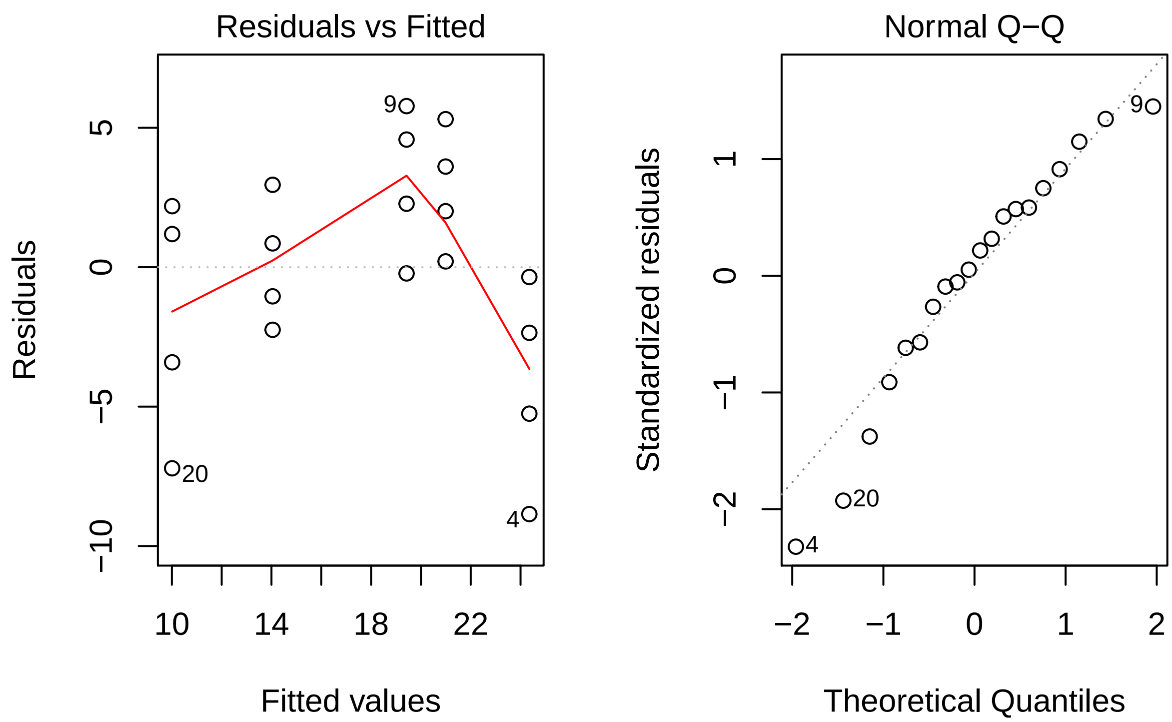
```
Steel<-read.csv("C:/hess/STAT511_FA11/HW-2018/HW11/Steel.csv")
str(Steel)
```

```
## 'data.frame':    20 obs. of  2 variables:
##  $ Thick   : int  220 220 220 220 370 370 370 370 440 440 ...
##  $ Strength: num  24 22 19.1 15.5 26.3 24.6 23 21.2 25.2 24 ...
```

```
#2A
Fit1 <- lm(Strength ~ Thick, data = Steel)
plot(Strength ~ Thick, data = Steel)
```




```r
par(mfrow=c(1,2))
plot(Fit1, which =c(1,2))
```



2A. (**4pts**) The relationship does NOT appear to be linear (based on scatter plot and also plot of resids vs fitted values). Regression assumptions NOT satisfied.

```r
#2B
Fit2<-lm(Strength ~ as.factor(Thick), data = Steel)
anova(Fit2, Fit1)
```

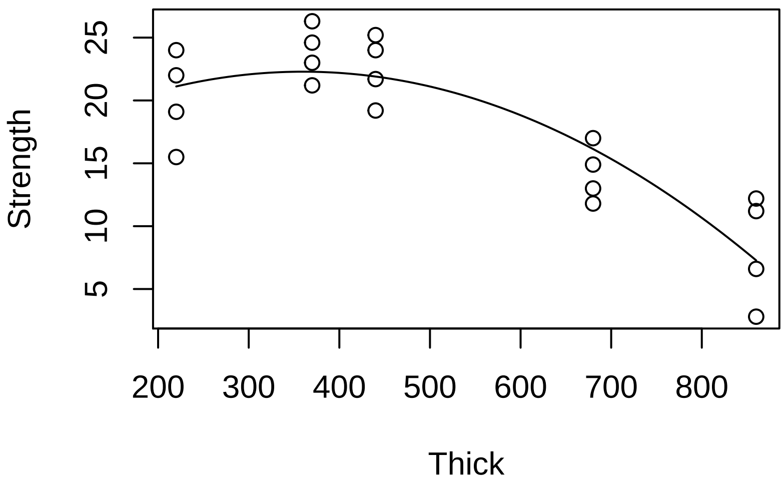
```
## Analysis of Variance Table
##
## Model 1: Strength ~ as.factor(Thick)
## Model 2: Strength ~ Thick
##   Res.Df    RSS Df Sum of Sq    F  Pr(>F)
## 1     15 148.57
## 2     18 301.90 -3   -153.33 5.16 0.01195 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2B. (**4pts**)
Lack of Fit test p-value = 0.01195, Reject H0.
Conclude the linear regression model does NOT fit.

```r
#2C (4 pts)
Fit3<-lm(Strength ~ Thick + I(Thick^2), data = Steel)
summary(Fit3)
```

```
##
## Call:
## lm(formula = Strength ~ Thick + I(Thick^2), data = Steel)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6222 -2.1960  0.2443  2.4491  4.8763
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.452e+01  4.752e+00   3.057  0.00713 **
## Thick        4.318e-02  1.980e-02   2.181  0.04354 *
## I(Thick^2)  -5.994e-05  1.786e-05  -3.357  0.00374 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.268 on 17 degrees of freedom
## Multiple R-squared:  0.7796, Adjusted R-squared:  0.7537
## F-statistic: 30.07 on 2 and 17 DF,  p-value: 2.609e-06
```

```r
par(mfrow=c(1,1))
plot(Strength ~ Thick, data = Steel)
curve(14.52 + 0.04318*x - 0.00006*x^2, add = TRUE)
```

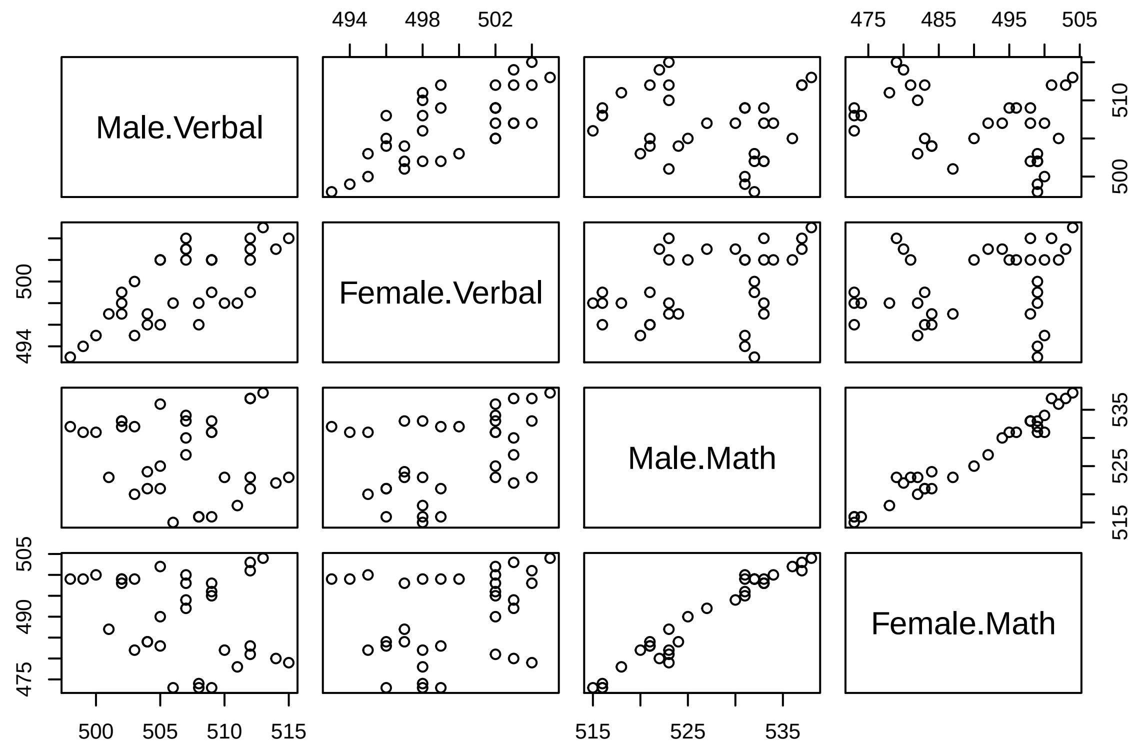


Q3 SAT Scores

```
SATData <- read.csv("C:/hess/STAT511_FA11/ASCII-comma/CH11/ex11-50.txt",
quote = " ' ", row.names = 1)
str(SATData)
```

```
## 'data.frame':    34 obs. of  4 variables:
##  $ Male.Verbal  : int  506 508 509 508 511 514 515 512 512 510 ...
##  $ Female.Verbal: int  498 496 499 498 498 503 504 502 499 498 ...
##  $ Male.Math    : int  515 516 516 516 518 522 523 523 521 523 ...
##  $ Female.Math  : int  473 473 473 474 478 480 479 481 483 482 ...
```

```
#3A
pairs(SATData)
```




```
#3B
cor(SATData)
```

```
##               Male.Verbal Female.Verbal  Male.Math Female.Math
## Male.Verbal     1.0000000     0.7081389 -0.1329501  -0.2884984
## Female.Verbal   0.7081389     1.0000000  0.3915856   0.2637590
## Male.Math      -0.1329501     0.3915856  1.0000000   0.9773392
## Female.Math    -0.2884984     0.2637590  0.9773392   1.0000000
```

3B. Male.Math and Female.Math have the strongest correlation.

```
#3C
cor.test(SATData$Female.Verbal, SATData$Female.Math)
```

```
##
##  Pearson's product-moment correlation
##
## data:  SATData$Female.Verbal and SATData$Female.Math
## t = 1.5468, df = 32, p-value = 0.1317
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.08169324  0.55263303
## sample estimates:
##      cor
## 0.263759
```

3C. Correlation p-value = 0.1317, Fail to Reject H0.
We cannot conclude that there is a linear association.