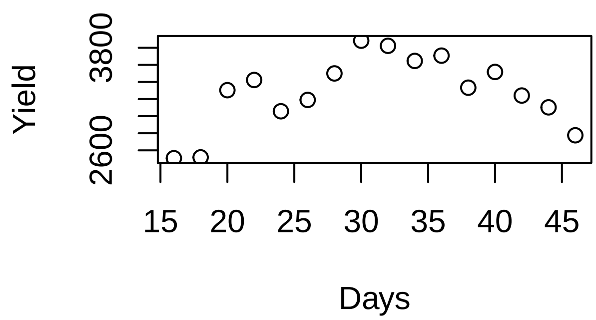
HW3 KEY

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

Wheat Yield

1. Scatterplot

```
Grain <- read.csv("C:/hess/STAT512/HW_2019/HW3/Grain.csv")
plot(Yield ~ Days, data = Grain)
```

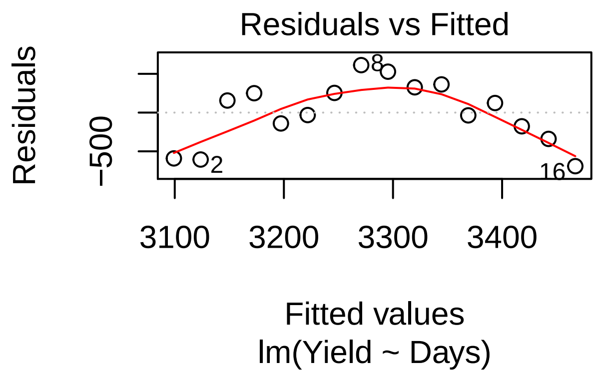


2. (**4 pts**) Linear Model
   There is a curve in the residual plot which indicates that a linear relationship may not be appropriate.

```
Model1 <- lm(Yield ~ Days, data = Grain)
summary(Model1)
```

```
##
## Call:
## lm(formula = Yield ~ Days, data = Grain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -691.07 -217.65   45.85  271.77  612.14
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2902.96     364.67   7.961 1.45e-06 ***
## Days           12.26      11.28   1.088    0.295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 415.8 on 14 degrees of freedom
## Multiple R-squared:  0.07791,    Adjusted R-squared:  0.01205
## F-statistic: 1.183 on 1 and 14 DF,  p-value: 0.2951
```

```
plot(Model1, which = 1)
```

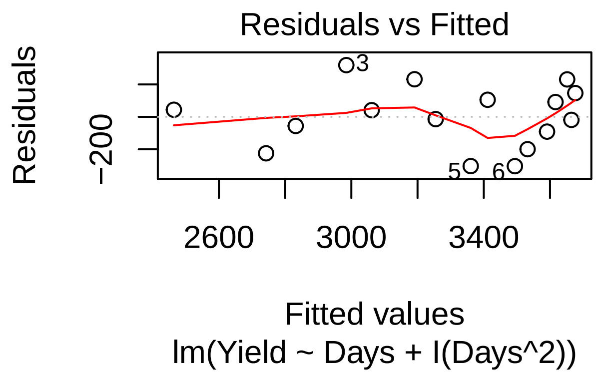


3. (**4pts**) Quadratic Regression
   The residuals are randomly scattered around zero indicating no apparent departures from the assumed model and the equal-variance assumption.

```
Model2 <- lm(Yield ~ Days + I(Days^2), data = Grain)
summary(Model2)
```

```
##
## Call:
## lm(formula = Yield ~ Days + I(Days^2), data = Grain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -303.96 -118.11   13.86  115.67  319.06
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1070.3977   617.2527  -1.734    0.107
## Days          293.4829    42.1776   6.958 9.94e-06 ***
## I(Days^2)      -4.5358     0.6744  -6.726 1.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 203.9 on 13 degrees of freedom
## Multiple R-squared:  0.7942, Adjusted R-squared:  0.7625
## F-statistic: 25.08 on 2 and 13 DF,  p-value: 3.452e-05
```

```
plot(Model2, which = 1)
```

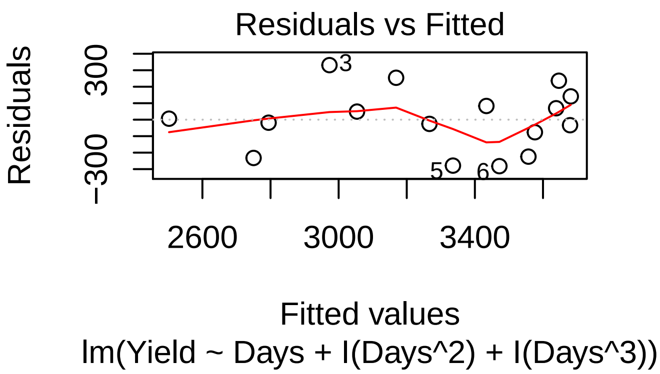


4. (**4pts**) Cubic Regression
   The residuals are randomly scattered around zero indicating no apparent departures from the assumed model and the equal-variance assumption.

```r
Model3 <- lm(Yield ~ Days + I(Days^2) + I(Days^3), data = Grain)
summary(Model3)
```

```
##
## Call:
## lm(formula = Yield ~ Days + I(Days^2) + I(Days^3), data = Grain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -281.97 -113.21   -6.11   97.75  330.92
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -203.60852 2285.13020  -0.089    0.930
## Days         199.07674  242.92513   0.819    0.428
## I(Days^2)     -1.32071    8.16843  -0.162    0.874
## I(Days^3)     -0.03457    0.08751  -0.395    0.700
##
## Residual standard error: 210.8 on 12 degrees of freedom
## Multiple R-squared:  0.7968, Adjusted R-squared:  0.746
## F-statistic: 15.68 on 3 and 12 DF,  p-value: 0.0001876
```

```r
plot(Model3, which = 1)
```



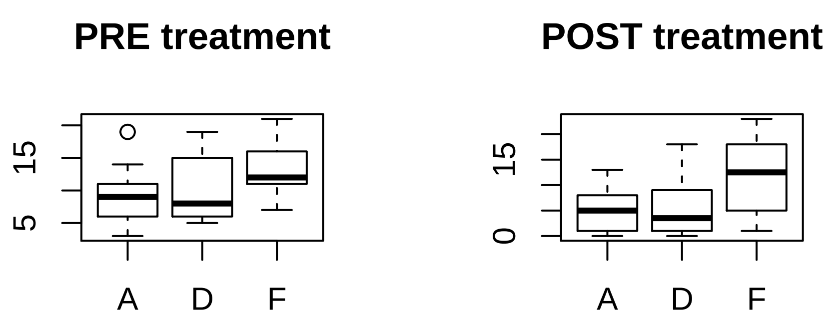

```r
rm(Model1, Model2, Model3)
```

5. From summary(Model3): F= 15.68, p-value=0.0002. Reject H0; conclude that the linear, quadratic and cubic regression coefficients are NOT all (simultaneously) zero.
6. I would choose the quadratic model since the residual plot looks OK and the quadratic coefficient is significantly different from 0.
   OR
   I would choose the quadratic model since it is the simplest model where assumptions are satisfied.

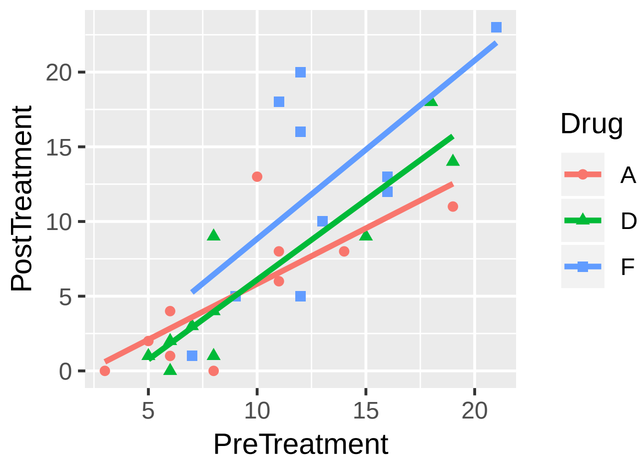
Drug Test

7. (**4pts**) Summary Plots

```r
DrugTest <- read.csv("C:/hess/STAT512/HW_2019/HW3/DrugTest.csv")
par(mfrow=c(1,2))
boxplot(PreTreatment ~ Drug, data = DrugTest, main = "PRE treatment")
boxplot(PostTreatment ~ Drug, data = DrugTest, main = "POST treatment")
```




```r
library(ggplot2)
qplot(PreTreatment, PostTreatment, shape = Drug, color = Drug, data = DrugTest) +
  geom_smooth(method = "lm", se = FALSE)
```



8. (**4pts**) One-way ANOVA
   Based on the one-way ANOVA F-test (p = 0.03) we can conclude that there is a difference between the Drugs. Based on Tukey adjusted pairwise comparisons, the mean for post treatment is significantly higher for Drug F as compared to Drug A (p = 0.04).

```
library(car)
library(emmeans)
Model1 <- lm(PostTreatment ~ Drug, data = DrugTest)
Anova(Model1, type =3)
```

```
## Anova Table (Type III tests)
##
## Response: PostTreatment
##             Sum Sq Df F value  Pr(>F)
## (Intercept)  280.9  1  7.6216 0.01024 *
## Drug         293.6  2  3.9831 0.03049 *
## Residuals    995.1 27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
em1 <- emmeans(Model1, ~Drug)
pairs(em1)
```

```
##  contrast estimate       SE df t.ratio p.value
##  A - D        -0.8 2.714979 27  -0.295  0.9533
##  A - F        -7.0 2.714979 27  -2.578  0.0403
##  D - F        -6.2 2.714979 27  -2.284  0.0754
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

9. (**4pts**) ANCOVA NO interaction
   Based on the one-way ANOVA F-test (p = 0.13) or Tukey adjusted pairise comparisons (all p-value > 0.05), we cannot conclude that there is a difference between the Drugs.

```
Model2 <- lm(PostTreatment ~ PreTreatment + Drug, data = DrugTest)
Anova(Model2, type =3)
```

```
## Anova Table (Type III tests)
##
## Response: PostTreatment
##               Sum Sq Df F value    Pr(>F)
```

```
## (Intercept)   61.26  1  3.8177   0.06155 .
## PreTreatment 577.90  1 36.0145 2.454e-06 ***
## Drug          68.55  2  2.1361   0.13838
## Residuals    417.20 26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
em2 <- emmeans(Model2, ~Drug)
pairs(em2)
```

```
##  contrast  estimate       SE df t.ratio p.value
##  A - D    -0.108971 1.795135 26  -0.061  0.9980
##  A - F    -3.446138 1.886781 26  -1.826  0.1809
##  D - F    -3.337167 1.853866 26  -1.800  0.1893
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

10. Looking at the boxplots from question 7, there appear to be differences between the Drugs at pre-treatment. After accounting for pre-treatment (by including it as a covariate), we no longer find significant differences between the Drugs as post-treatment.

11. (**4pts**) One-way ANOVA with Differences
Based on the one-way ANOVA F-test ($p = 0.11$) or Tukey adjusted pairise comparisons (all p-value $>$ 0.05), we cannot conclude that there is a difference between the Drugs. (This conclusion is the same as from the ANCOVA model.)

```
DrugTest$Diff <- DrugTest$PostTreatment - DrugTest$PreTreatment
Model3 <- lm(Diff ~ Drug, data = DrugTest)
Anova(Model3, type =3)
```

```
## Anova Table (Type III tests)
##
## Response: Diff
##             Sum Sq Df F value   Pr(>F)
## (Intercept) 160.00  1  10.352 0.003349 **
## Drug         74.87  2   2.422 0.107780
## Residuals   417.30 27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
em3 <- emmeans(Model3, ~Drug)
pairs(em3)
```

```
##  contrast estimate       SE df t.ratio p.value
##  A - D        -0.1 1.758156 27  -0.057  0.9982
##  A - F        -3.4 1.758156 27  -1.934  0.1486
##  D - F        -3.3 1.758156 27  -1.877  0.1647
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```