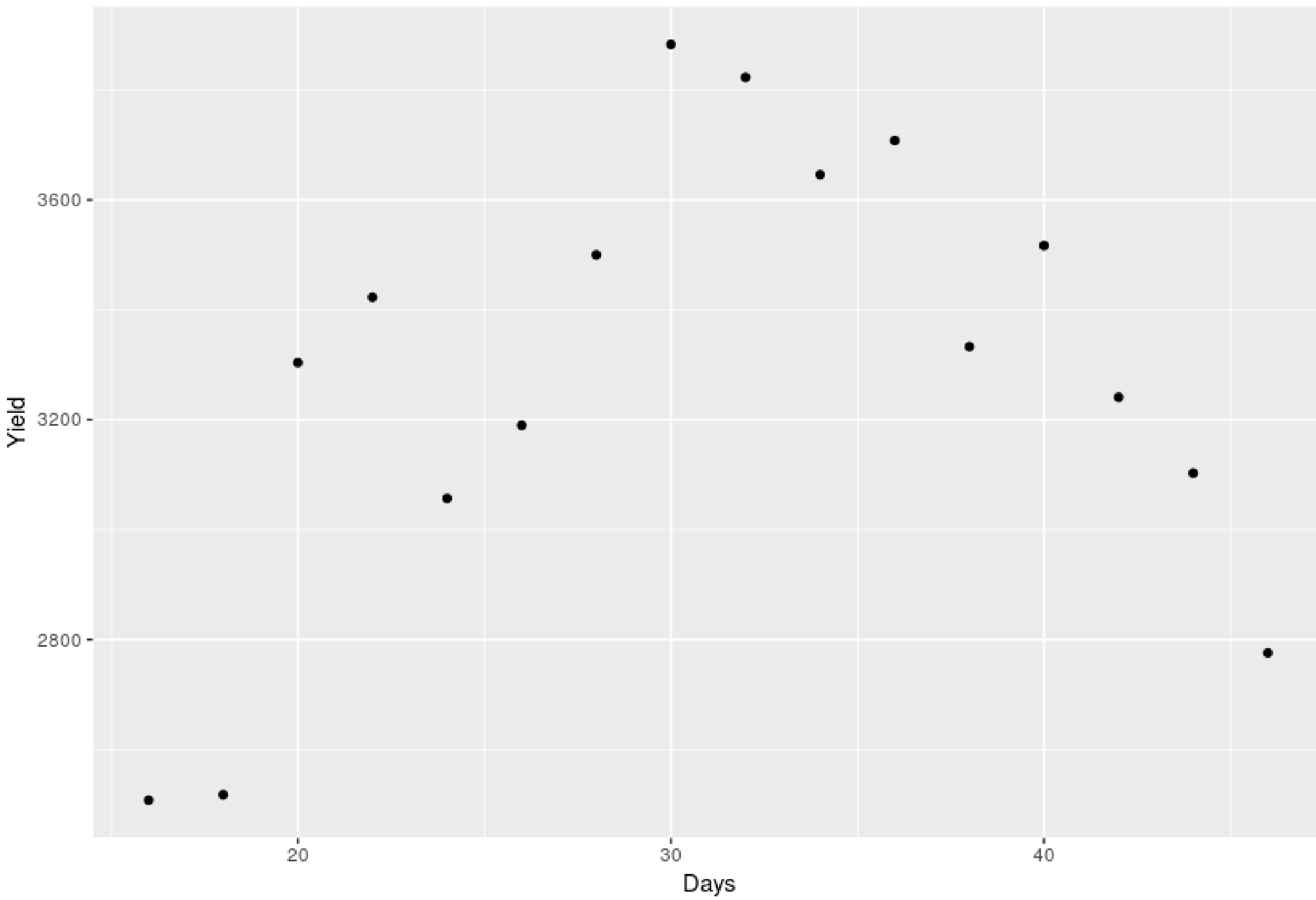
STAT 512 – Assignment 3

Vignesh M. Pagadala
Vignesh.Pagadala@ColoState.Edu

Feb 25, 2019

1. Create a scatterplot of Yield vs Days. Include this plot in your assignment.

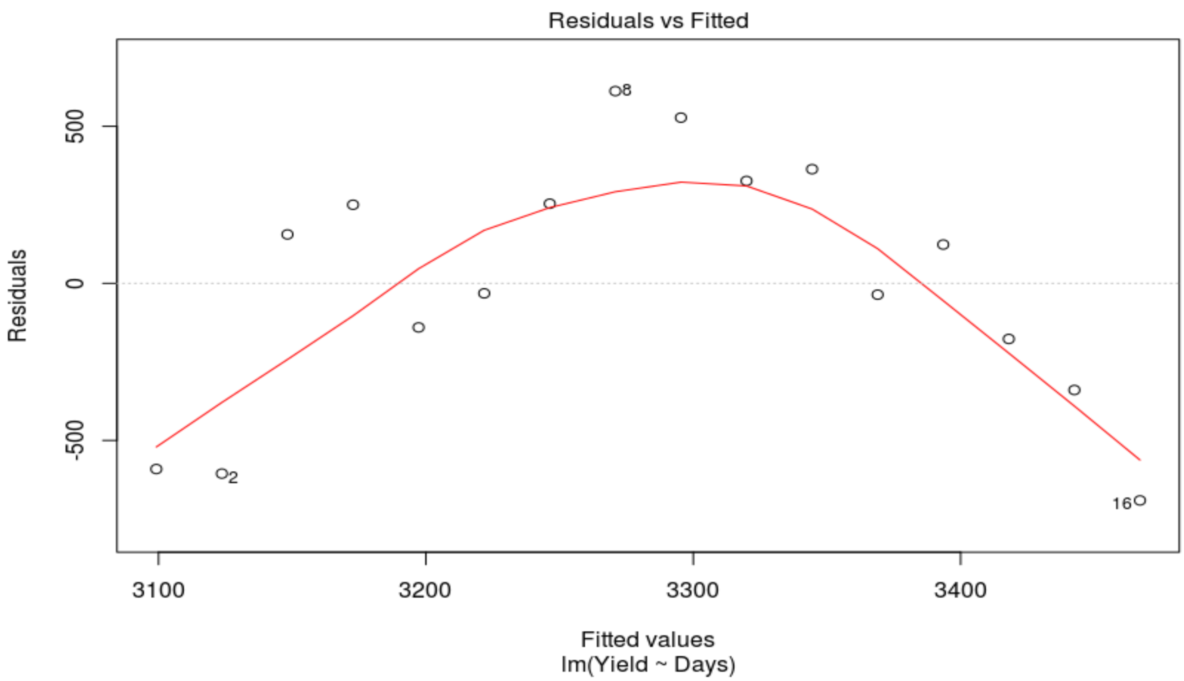
ANSWER:



2. Fit a linear regression model of Yield on Days. Include the parameter estimate information ("Coefficients" table) in your assignment. Examine a plot of the residuals versus predicted values. What does the residual plot suggest? (4 pts)

ANSWER:

```
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
```

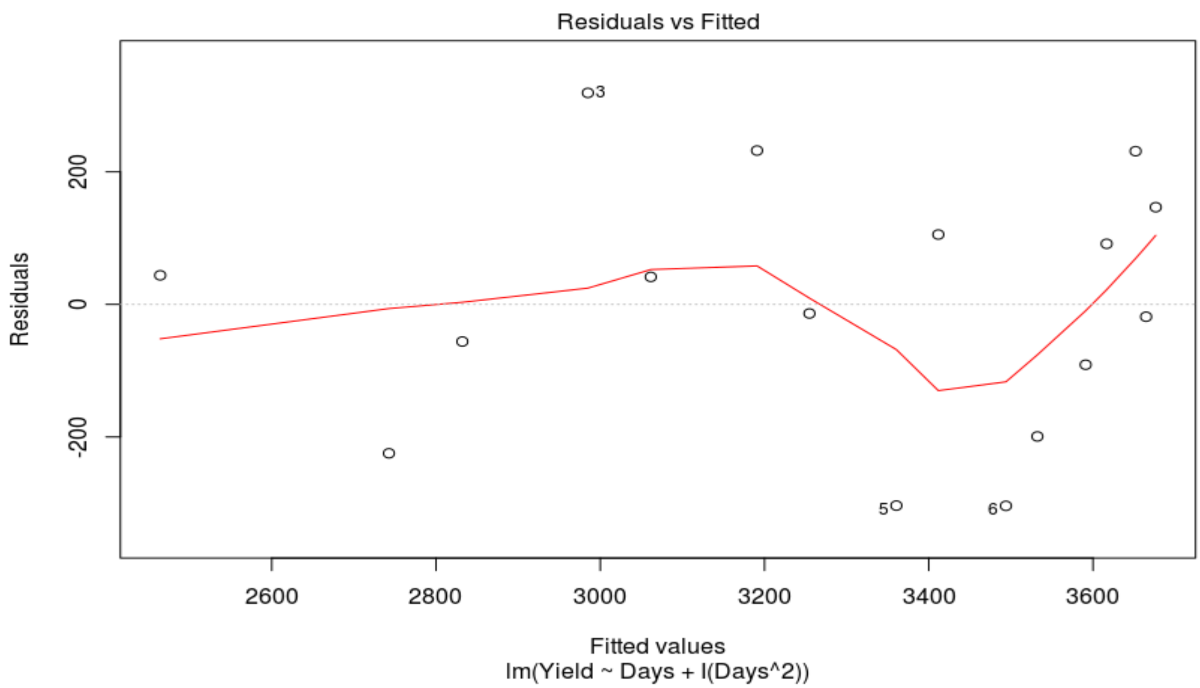


The residual plot shows a clear trend (upside-down U shape). Therefore, the linearity assumption does not hold, and this is not a good model.

3. Fit a quadratic regression model (including both linear and quadratic terms). Include the parameter estimate information ("Coefficients" table) in your assignment. Examine a plot of the residuals versus predicted values and comment. (4 pts)

ANSWER:

```
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 ***
```

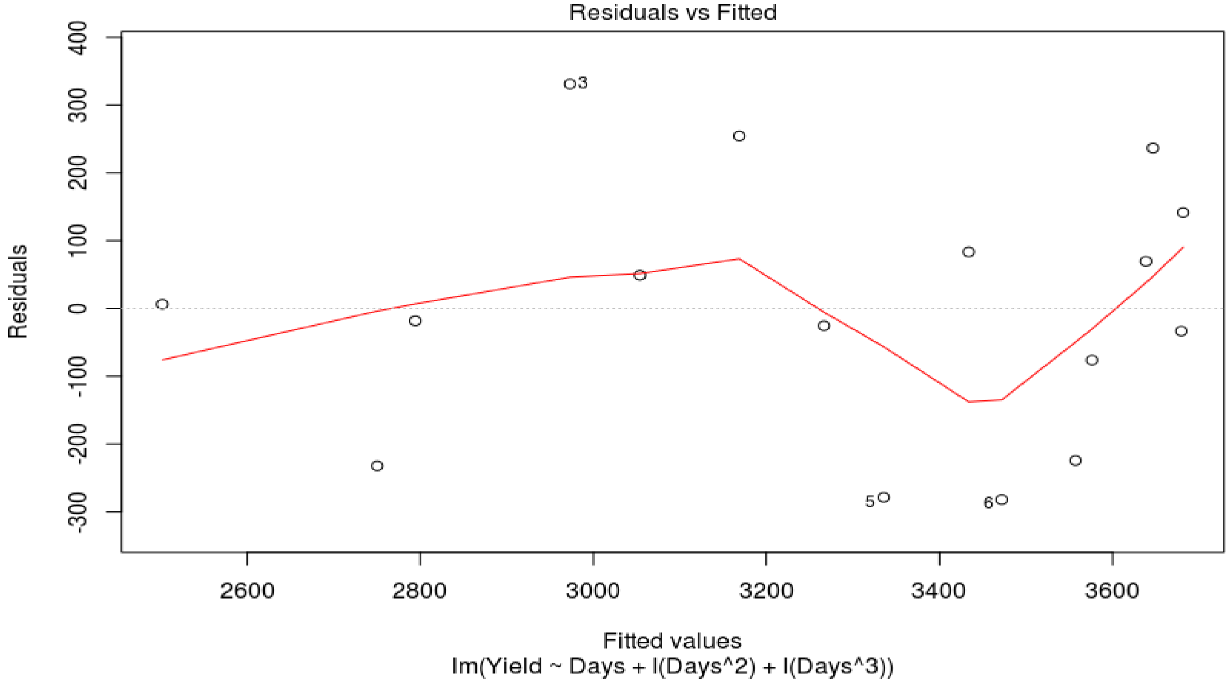


There is no clear trend which we can observe from the plot. Also, the points seem scattered more or less, randomly. This indicates that this model is actually pretty good, and that the assumptions do hold.

4. Fit a cubic regression model (including linear, quadratic, and cubic terms). Include the parameter estimate information (“Coefficients” table) in your assignment. Again examine a plot of the residuals versus predicted values and comment. (4 pts)

ANSWER:

```
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
```



This plot does not look different from the same plot for the quadratic model. It clearly does not seem to show a trend, and seems scattered more or less equally. However, the p values here are very bad.

5. In the cubic model (#4), test the hypothesis that the linear, quadratic and cubic regression
coefficients are all simultaneously zero. Give the F-statistic and p-value and make a conclusion about the test.

ANSWER:

```
Linear hypothesis test

Hypothesis:
Days = 0
I(Days^2) = 0
I(Days^3) = 0

Model 1: restricted model
Model 2: Yield ~ Days + I(Days^2) + I(Days^3)

  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     15 2625168
2     12  533451  3   2091717 15.684 0.0001876 ***
```

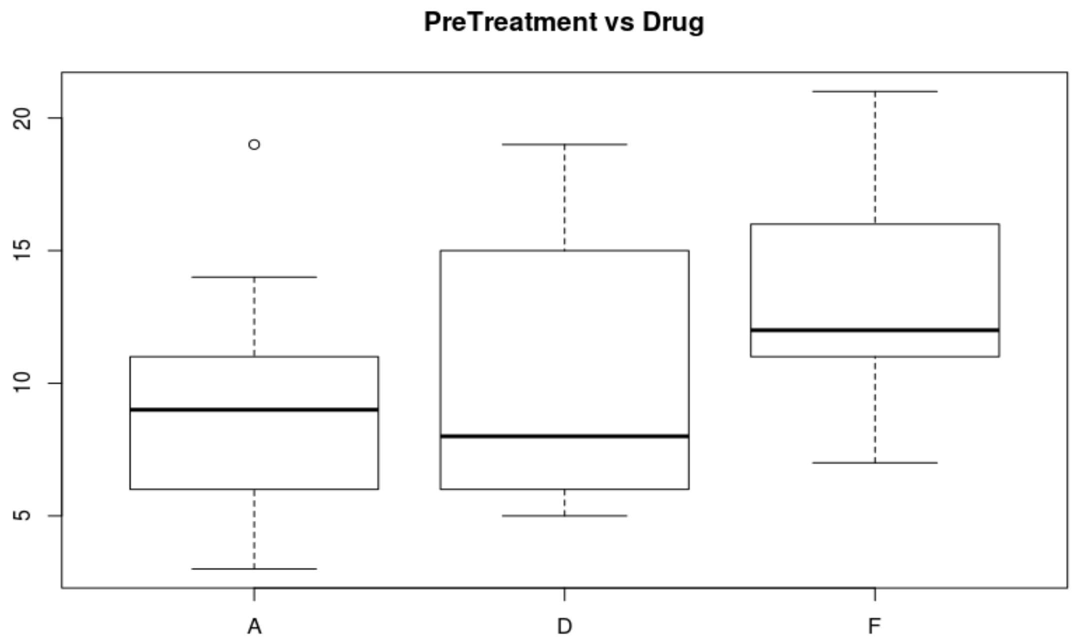
The p-value is low (<0.05). Therefore, we can reject the null hypothesis for this model. At least one predictor contributes positively to the model.

6. Which model would you choose: linear, quadratic or cubic? Justify your choice.

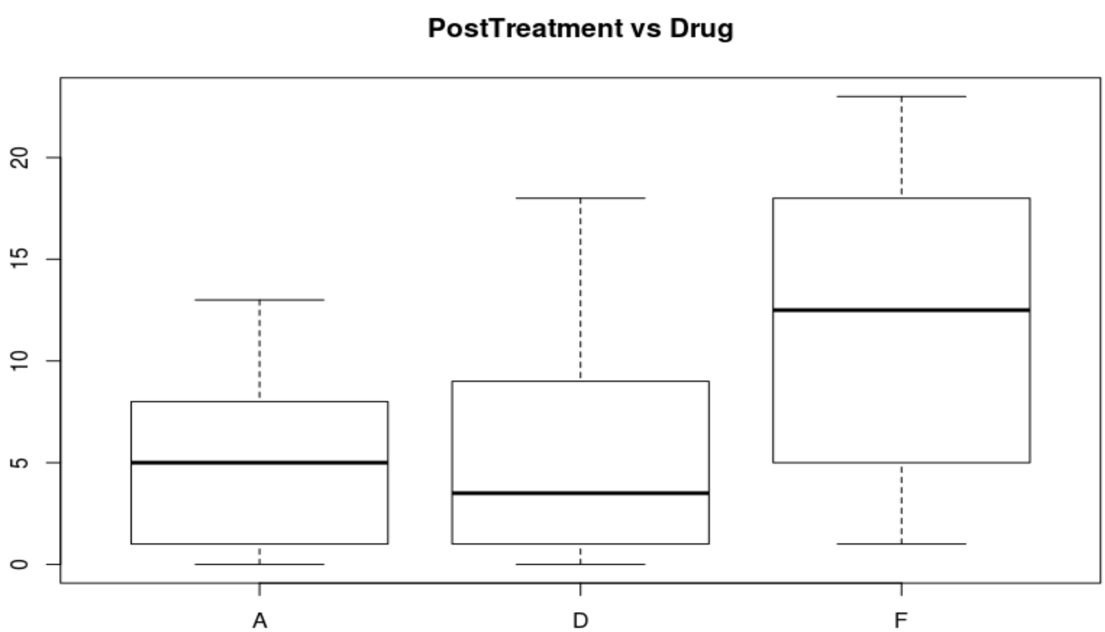
ANSWER:
I would choose the quadratic model. In this model, the quadratic term is clearly associated with the lowest (best) p-values. By both forward and backward elimination, the quadratic model is the best fit.

7. Construct side-by-side boxplots of (1) PreTreatment vs Drug and (2) PostTreatment vs Drug. Also construct (3) a scatterplot of PostTreatment vs PreTreatment for all Drugs on the same plot. Overlay a fitted regression line for each Drug. Include these plots in your assignment. (4 pts)

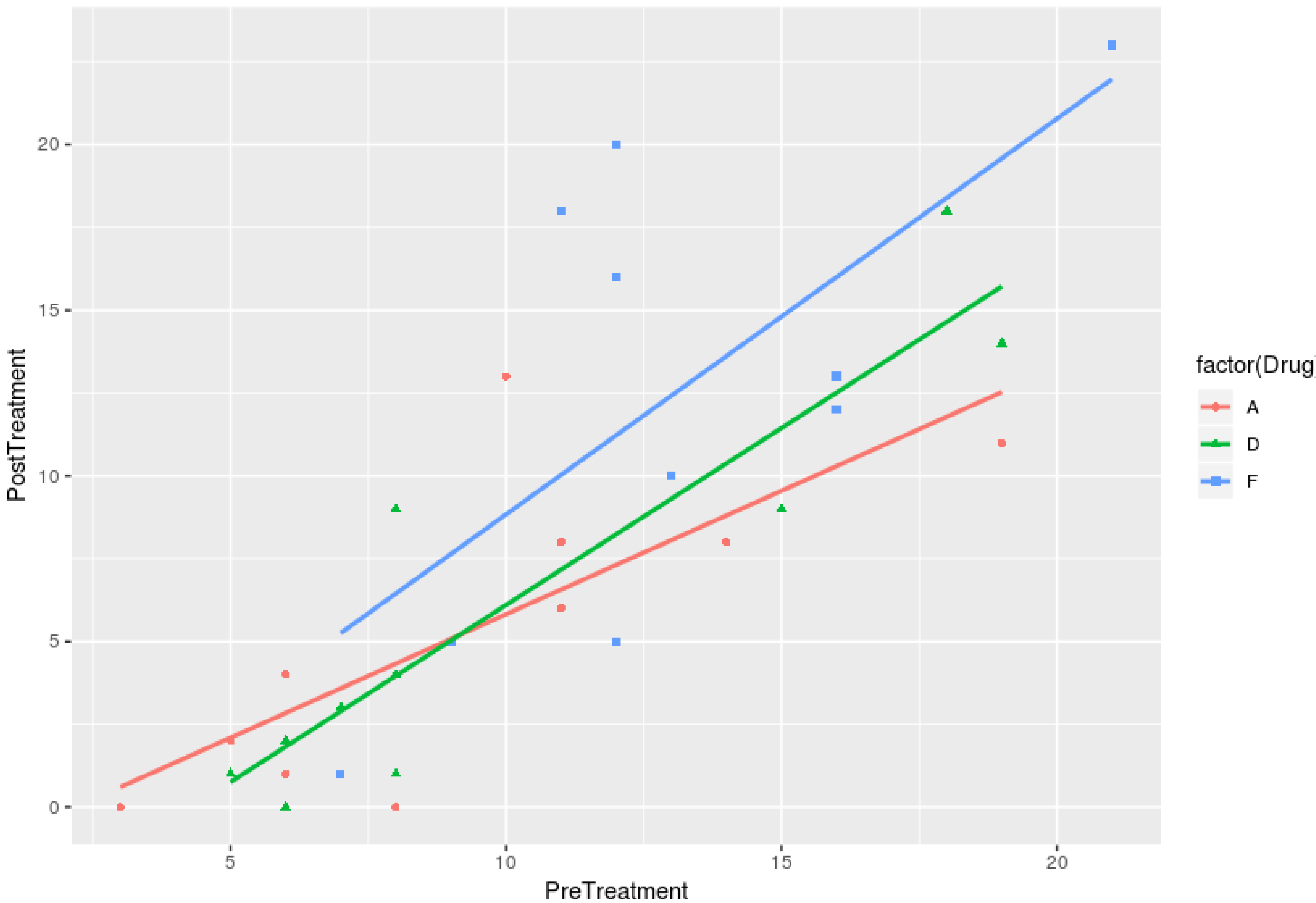
ANSWER:
(1)



(2)



(3)



8. Fit the one-way ANOVA model (using Drug as the only predictor). Include the ANOVA table and Tukey adjusted pairwise comparisons in your assignment. What can we conclude about differences between the Drugs? (4 pts)

ANSWER:

```
Analysis of Variance Table

Response: PostTreatment
          Df Sum Sq Mean Sq F value  Pr(>F)
Drug       2  293.6 146.800  3.9831 0.03049 *
Residuals 27  995.1  36.856

$emmeans
 Drug emmean   SE df lower.CL upper.CL
 A       5.3 1.92 27     1.36     9.24
 D       6.1 1.92 27     2.16    10.04
 F      12.3 1.92 27     8.36    16.24
```

```
Confidence level used: 0.95

$contrasts
 contrast estimate   SE df t.ratio p.value
 A - D        -0.8 2.71 27 -0.295  0.9533
 A - F        -7.0 2.71 27 -2.578  0.0403
 D - F        -6.2 2.71 27 -2.284  0.0754
```

From the anova table, we can see that the p-value is less than 0.05. The variation of PostTreatment across different drugs is larger than the variation within the types of drugs. Also, variation in Drug has a significant effect on the response.

9. Now fit the ANCOVA model with NO Interaction (using Drug and PreTreatment as the predictors). Include the ANOVA table and Tukey adjusted pairwise comparisons in your assignment. What can we conclude about differences between the Drugs? (4 pts)

ANSWER:

Anova Table (Type III tests)

```
Response: PostTreatment
               Sum Sq Df F value    Pr(>F)
(Intercept)   61.26  1  3.8177   0.06155 .
Drug          68.55  2  2.1361   0.13838
PreTreatment 577.90  1 36.0145 2.454e-06 ***
Residuals    417.20 26
```

$emmeans

```
 Drug PreTreatment emmean   SE df lower.CL upper.CL
 A            10.7   6.71 1.29 26     4.07     9.36
 D            10.7   6.82 1.27 26     4.21     9.44
 F            10.7  10.16 1.32 26     7.46    12.87

Confidence level used: 0.95

$contrasts
 contrast                                estimate   SE df t.ratio
p.value
 A,10.7333333333333 - D,10.7333333333333   -0.109 1.80 26 -0.061
0.9980
```

```
 A,10.7333333333333 - F,10.7333333333333   -3.446 1.89 26 -1.826
0.1809
 D,10.7333333333333 - F,10.7333333333333   -3.337 1.85 26 -1.800
0.1893
```

From the Anova table, we can see that PreTreatment has a very low p-value (<0.05). Hence, this predictor is significant in terms of its effect on the PostTreatment value. However, Drug is not significant due to low p-value.

10. Comparing your conclusions from #8 (one-way ANOVA) vs #9 (ANCOVA) you should have found different conclusions regarding significant differences between the Drug treatments. Give a brief explanation of why the conclusions change when we include PreTreatment as a covariate. Hint: Consider the boxplots from #7.

ANSWER:

From the figure in 7, we can see that Drug and PostTreatment are not very correlated (some correlation still exists). However, PreTreatment and Drug are more correlated, due to which they interfere with each other.

11. An alternative approach to the ANCOVA above is to calculate the difference (Diff = PostTreatment – PreTreatment) and use this Diff as the response in a one-way ANOVA model. Do this and include the ANOVA table and Tukey adjusted pairwise comparisons in your assignment. What can we conclude about differences between the Drugs? (4 pts)

ANSWER:

```
Analysis of Variance Table

Response: I(PostTreatment - PreTreatment)
           Df Sum Sq Mean Sq F value Pr(>F)
Drug        2  74.87  37.433   2.422 0.1078
Residuals 27 417.30  15.456

$emmeans
 Drug emmean   SE df lower.CL upper.CL
 A      -4.0 1.24 27    -6.55    -1.45
 D      -3.9 1.24 27    -6.45    -1.35
 F      -0.6 1.24 27    -3.15     1.95

Results are given on the I (not the response) scale.
Confidence level used: 0.95
```

```
$contrasts
 contrast estimate   SE df t.ratio p.value
 A - D        -0.1 1.76 27 -0.057  0.9982
 A - F        -3.4 1.76 27 -1.934  0.1486
 D - F        -3.3 1.76 27 -1.877  0.1647
```

This result indicates that variation seen in Drug does not manifest as variation in the response (which in this case, is the difference between PreTreatment and PostTreatment), due to the high p value.