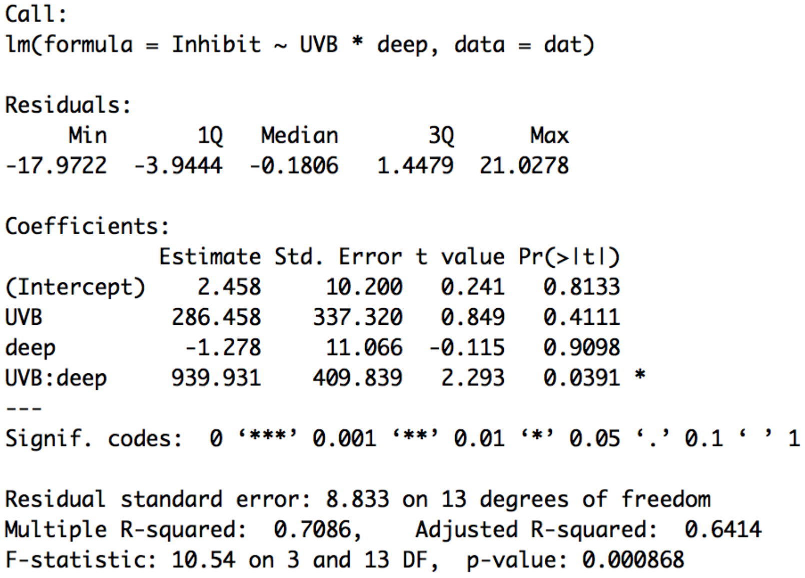
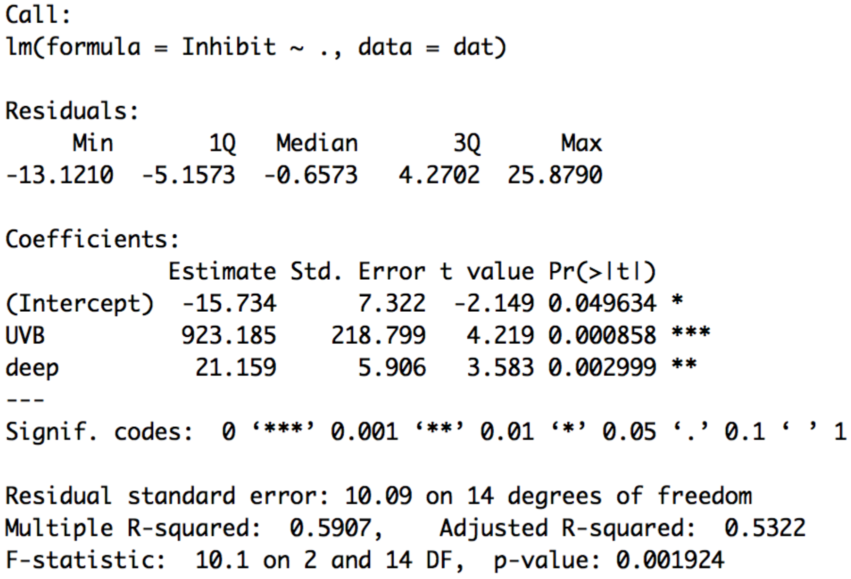
3.



```
Call:
lm(formula = Inhibit ~ UVB * deep, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-17.9722  -3.9444  -0.1806   1.4479  21.0278 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)    2.458     10.200   0.241   0.8133  
UVB          286.458    337.320   0.849   0.4111  
deep          -1.278     11.066  -0.115   0.9098  
UVB:deep     939.931    409.839   2.293   0.0391 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.833 on 13 degrees of freedom
Multiple R-squared:  0.7086,	Adjusted R-squared:  0.6414 
F-statistic: 10.54 on 3 and 13 DF,  p-value: 0.000868
```

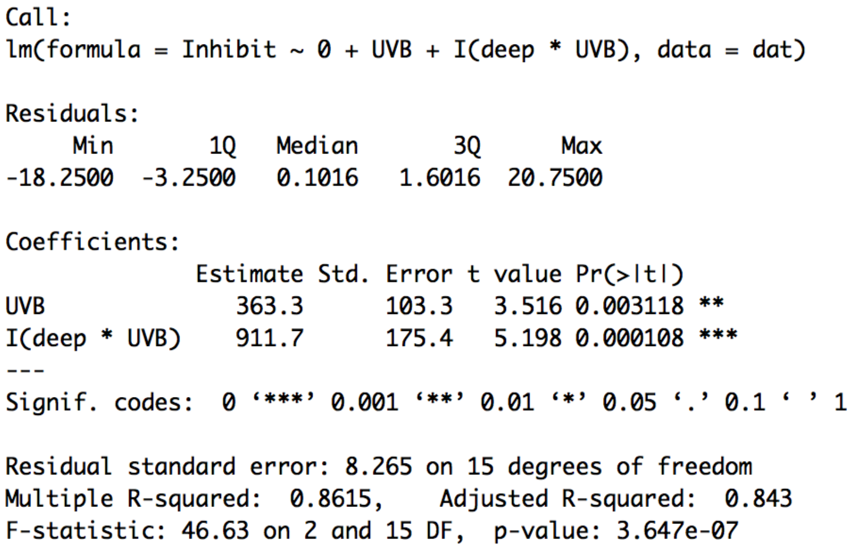
According to first (full) model (different slopes), UVB and deep seem to be insignificant variables. However, the indicator variable seems to be relevant.



```
Call:
lm(formula = Inhibit ~ ., data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-13.1210  -5.1573  -0.6573   4.2702  25.8790 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -15.734      7.322  -2.149 0.049634 *  
UVB          923.185    218.799   4.219 0.000858 ***
deep          21.159      5.906   3.583 0.002999 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.09 on 14 degrees of freedom
Multiple R-squared:  0.5907,	Adjusted R-squared:  0.5322 
F-statistic:  10.1 on 2 and 14 DF,  p-value: 0.001924
```

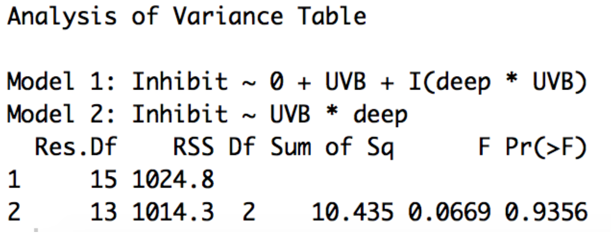
We analyze the model again (parallel slopes model) without the indicator variable. This time around, both variables seems relevant (p – value < 0.001).

In this case, we can see that it wouldn’t make sense to have inhibition without UVB exposure. Hence, we try another model to see the effect of UVB after accounting for depth.



```
Call:
lm(formula = Inhibit ~ 0 + UVB + I(deep * UVB), data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-18.2500  -3.2500   0.1016   1.6016  20.7500 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
UVB              363.3      103.3   3.516 0.003118 ** 
I(deep * UVB)    911.7      175.4   5.198 0.000108 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.265 on 15 degrees of freedom
Multiple R-squared:  0.8615,	Adjusted R-squared:  0.843 
F-statistic: 46.63 on 2 and 15 DF,  p-value: 3.647e-07
```

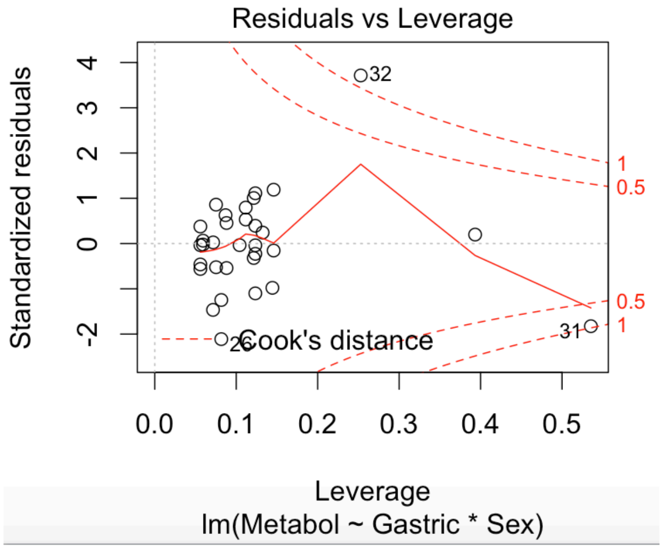
In this model, both variables seem relevant. The difference in slopes of deep vs. surface is 911.7. Extra sum of squares analysis can be found below:



```
Analysis of Variance Table

Model 1: Inhibit ~ 0 + UVB + I(deep * UVB)
Model 2: Inhibit ~ UVB * deep
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     15 1024.8                           
2     13 1014.3  2    10.435 0.0669 0.9356
```

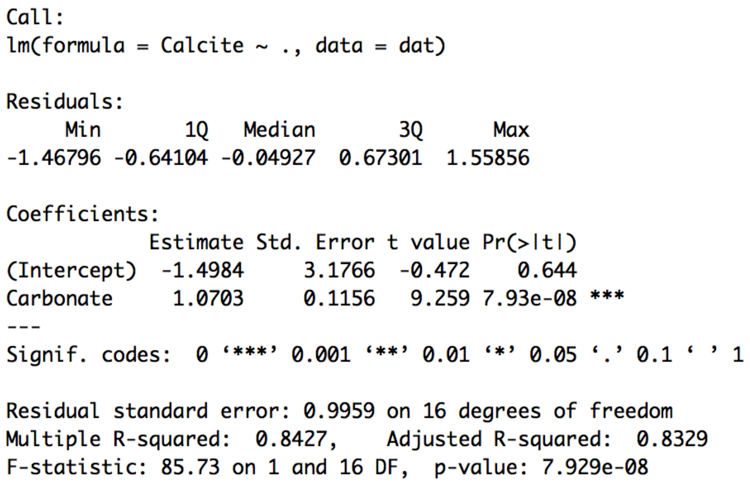
This confirms the relevance of the last model.

5.

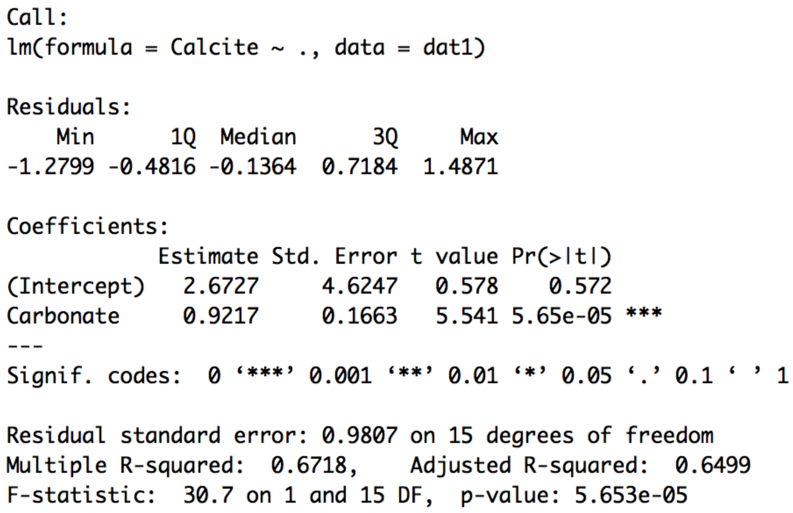


Leverage = 0.2528
Studentized Residual = 5.12
Cook’s Distance = 1.167

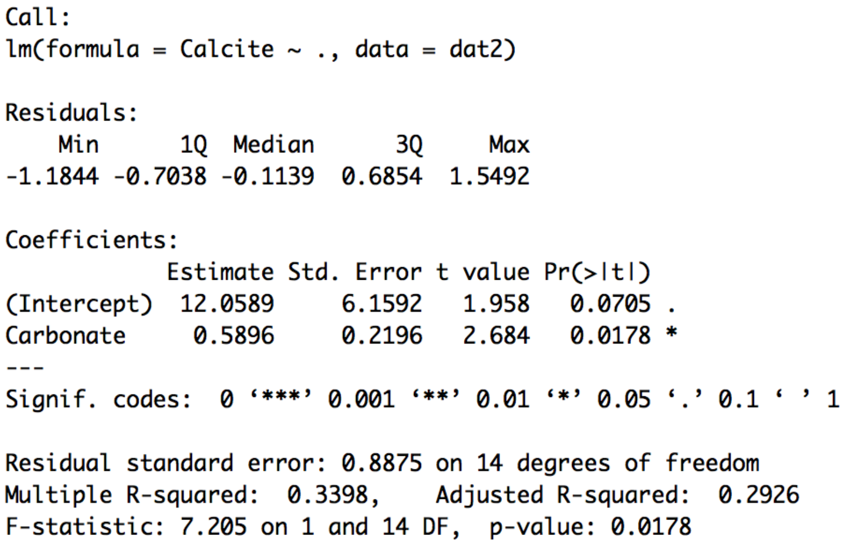
6. a.



```
Call:
lm(formula = Calcite ~ ., data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.46796 -0.64104 -0.04927  0.67301  1.55856 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.4984     3.1766  -0.472    0.644    
Carbonate     1.0703     0.1156   9.259 7.93e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9959 on 16 degrees of freedom
Multiple R-squared:  0.8427,	Adjusted R-squared:  0.8329 
F-statistic: 85.73 on 1 and 16 DF,  p-value: 7.929e-08
```



```
Call:
lm(formula = Calcite ~ ., data = dat1)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2799 -0.4816 -0.1364  0.7184  1.4871 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.6727     4.6247   0.578    0.572    
Carbonate     0.9217     0.1663   5.541 5.65e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9807 on 15 degrees of freedom
Multiple R-squared:  0.6718,	Adjusted R-squared:  0.6499 
F-statistic:  30.7 on 1 and 15 DF,  p-value: 5.653e-05
```



```
Call:
lm(formula = Calcite ~ ., data = dat2)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.1844 -0.7038 -0.1139  0.6854  1.5492 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  12.0589     6.1592   1.958   0.0705 .
Carbonate     0.5896     0.2196   2.684   0.0178 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8875 on 14 degrees of freedom
Multiple R-squared:  0.3398,	Adjusted R-squared:  0.2926 
F-statistic: 7.205 on 1 and 14 DF,  p-value: 0.0178
```

b. Without points 1 and 2, we can notice that the remaining points are only a cluster. But points 1 and 2, make the data set more linear. Hence, R-squared values increase with these points

c. Max Leverage = 0.5577 (Point 1)
Max studentized residual = 1.704 (Point 10)

Max Cook’s Distance = 0.917 (Point 1)

d. Max Leverage = 0.5578 (Point 2)
Min studentized residual = -2.077 (Point 2)
Max Cook’s Distance = 2.229 (Point 2)

e. The maximum leverage in both cases is the same. However, the extreme studentized residual increases when point 1 is removed. Maximum Cook’s distance also increases in the second case. Both studentized residual and Cook’s distance indicate that point 2 could be a possible outlier. This is not the case when point 1 is included, possibly because point 2 reduces its effect.

The cases influential statistics are calculated removing one point at a time. This makes it difficult to calculate pairs.

f. The p-value of F-statistic when both points are excluded is 0.0178 >> 0.001 when these points are included. This indicates that these points are very influential.

CODES

```
#Question 3
library(Sleuth3)
attach(ex1026)
dat <- ex1026
dat$deep <- ifelse(Surface == "Deep", 1,0)
dat <- dat[-3]
model1 <- lm(Inhibit ~ UVB*deep, data = dat)
summary(model1)
model2 <- lm(Inhibit~., data = dat)
summary(model2)
model3 <- lm(Inhibit ~ 0 + UVB + I(deep*UVB), data = dat)
summary(model3)
anova(model3,model1)

#Question 5
library(MASS)
library(Sleuth3)
attach(case1101)
dat <- case1101
model <- lm(Metabol ~ Gastric*Sex)
summary(model)
hatvalues(model)[32] #Leverage
studres(model)[32] #Studentized Residual
cooks.distance(model)[32] #Cook's distance

#Question 6
library(MASS)
library(Sleuth3)
attach(ex1120)
dat <- ex1120
dat1 <- dat[-1,]
dat2 <- dat1[-1,]
#Part a
model <- lm(Calcite ~., data = dat)
model1 <- lm(Calcite ~., data = dat1)
model2 <- lm(Calcite ~., data = dat2)

#Part c
max(hatvalues(model)) #Leverage
which(hatvalues(model) == max(hatvalues(model))) #Index = 1
max(studres(model)) #Studentized Residual
which(studres(model) == max(studres(model))) #Index = 10
max(cooks.distance(model)) #Cook's distance
which(cooks.distance(model) == max(cooks.distance(model))) #Index = 1

#Part d
max(hatvalues(model1)) #Leverage
which(hatvalues(model1) == max(hatvalues(model1))) #Index = 2
min(studres(model1)) #Studentized Residual
which(studres(model1) == min(studres(model1))) #Index = 2
max(cooks.distance(model1)) #Cook's distance
which(cooks.distance(model1) == max(cooks.distance(model1))) #Index = 2
```