1a) Find the least squares estimator of $\hat{\boldsymbol{\beta}}$

```
          [,1]
[1,] -2.381630
[2,]  2.401792
[3,]  1.443504
```

1b) Compute the estimate s^2 of σ^2

```
         [,1]
[1,] 125.5816
```

1c) Find the variance covariance matrix of $\hat{\boldsymbol{\beta}}$

```
           [,1]      [,2]      [,3]
[1,] 102.086172  4.931655 -12.72358
[2,]   4.931655  4.229413  -6.61544
[3,] -12.723579 -6.615440  10.65219
```

1d) Find the variance of the fitted values $\hat{Y}$

```
          [,1]      [,2]        [,3]       [,4]       [,5]       [,6]       [,7]      [,8]      [,9]
[1,]  40.578424  2.649368   4.0777925  9.0653077 -27.866245  20.899684  36.157490 23.325611 16.694211
[2,]   2.649368 19.016398  22.8388298 15.1381062  29.964992   8.921667   3.447371 11.203954 12.400958
[3,]   4.077792 22.838830 101.1855691 -0.0980403  -1.593725 -25.715059 -13.498781 32.374959  6.010098
[4,]   9.065308 15.138106  -0.0980403 17.8627146  27.527664  19.719780  13.525001  8.075785 14.765325
[5,] -27.866245 29.964992  -1.5937249 27.5276641  91.183179  16.828731 -13.780362 -8.905707 12.223117
[6,]  20.899684  8.921667 -25.7150586 19.7197802  16.828731  32.229038  28.275506  6.679280 17.743014
[7,]  36.157490  3.447371 -13.4987812 13.5250013 -13.780362  28.275506  36.889373 16.734111 17.831934
[8,]  23.325611 11.203954  32.3749589  8.0757847  -8.905707   6.679280  16.734111 22.990450 13.103200
[9,]  16.694211 12.400958   6.0100984 14.7653251  12.223117  17.743014  17.831934 13.103200 14.809785
```

1e) Find the variance covariance matrix of the residuals

```
          [,1]       [,2]        [,3]        [,4]       [,5]       [,6]       [,7]       [,8]       [,9]
[1,]  85.003220  -2.649368  -4.0777925  -9.0653077  27.866245 -20.899684 -36.157490 -23.325611 -16.694211
[2,]  -2.649368 106.565246 -22.8388298 -15.1381062 -29.964992  -8.921667  -3.447371 -11.203954 -12.400958
[3,]  -4.077792 -22.838830  24.3960745   0.0980403   1.593725  25.715059  13.498781 -32.374959  -6.010098
[4,]  -9.065308 -15.138106   0.0980403 107.7189291 -27.527664 -19.719780 -13.525001  -8.075785 -14.765325
[5,]  27.866245 -29.964992   1.5937249 -27.5276641  34.398465 -16.828731  13.780362   8.905707 -12.223117
[6,] -20.899684  -8.921667  25.7150586 -19.7197802 -16.828731  93.352605 -28.275506  -6.679280 -17.743014
[7,] -36.157490  -3.447371  13.4987812 -13.5250013  13.780362 -28.275506  88.692271 -16.734111 -17.831934
[8,] -23.325611 -11.203954 -32.3749589  -8.0757847   8.905707  -6.679280 -16.734111 102.591193 -13.103200
[9,] -16.694211 -12.400958  -6.0100984 -14.7653251 -12.223117 -17.743014 -17.831934 -13.103200 110.771858
```

2a) Fit the quadratic model to the data, then conduct a test to determine if the incidence rate is curvilinearly related to the estimate rate.

Call:
lm(formula = Rate ~ Estimation + I(Estimation^2))

Residuals:

Min	1Q	Median	3Q	Max
-101356	-801	-61	207	104498

Coefficients:

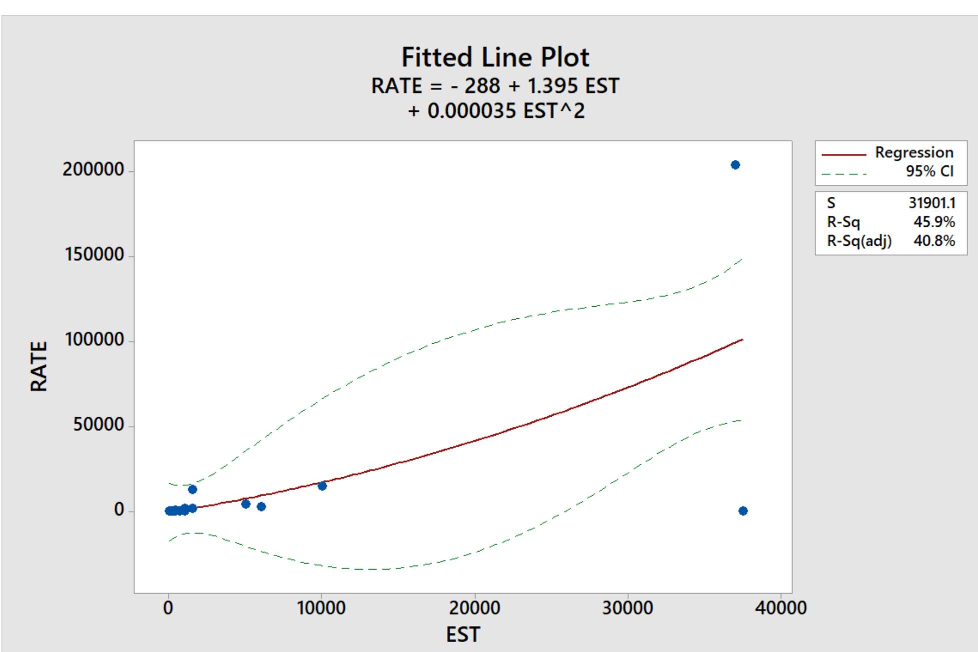
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.884e+02	8.049e+03	-0.036	0.972
Estimation	1.395e+00	3.651e+00	0.382	0.706
I(Estimation^2)	3.509e-05	9.724e-05	0.361	0.722

Residual standard error: 31900 on 21 degrees of freedom
Multiple R-squared: 0.4591, Adjusted R-squared: 0.4076
F-statistic: 8.912 on 2 and 21 DF, p-value: 0.001577

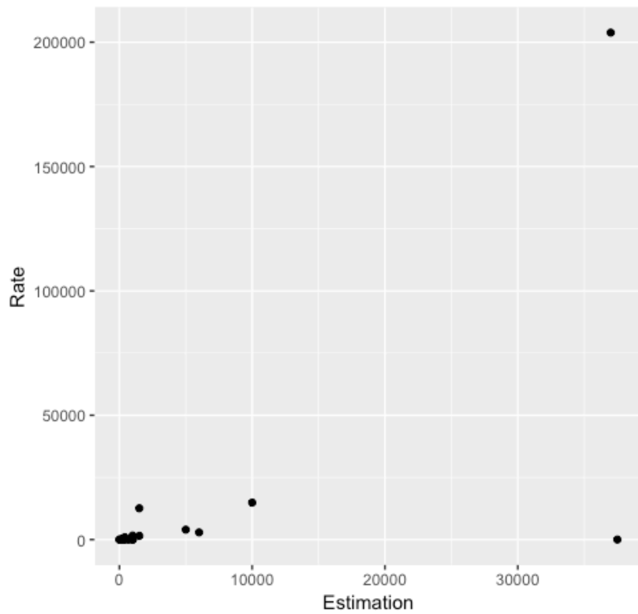
The regression equation is
RATE = - 288 + 1.395 EST + 0.000035 EST^2

We test at $\propto = .05$ $H_o: \beta_2 = 0$ vs $H_A: \beta_2 \neq 0$

Our p value is $\frac{0.722}{2} = 0.361 > .05$. We fail to reject our null hypothesis and conclude that the data is not curvilinearly related.



2b) Construct a scatterplot for the data. Locate the data point for Botulism on the graph. What do you observe?



The point for Botulism is located in the right bottom corner of our scatter plot. The data appears to be curvilinear expect for that particular point. This point might be an outlier and might be affecting the R-Squared and Adj.-R Squared

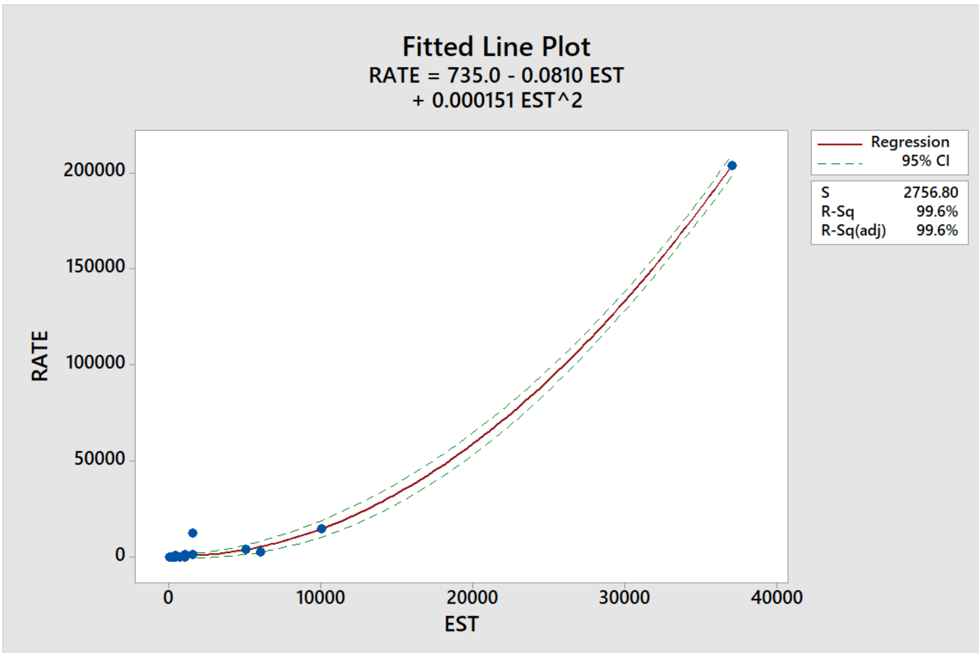
2c) Repeat part a, but omit the data point for Botulism from the analysis. Has the fit of the model improved?

```
Call:
lm(formula = Rate_Subset ~ Estimation_Subset + I(Estimation_Subset^2))

Residuals:
    Min      1Q  Median      3Q     Max
-2742.0  -726.9  -686.6   -81.4 11666.8

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             7.350e+02  6.959e+02   1.056    0.303
Estimation_Subset      -8.096e-02  3.167e-01  -0.256    0.801
I(Estimation_Subset^2)  1.505e-04  8.683e-06  17.336 1.62e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2757 on 20 degrees of freedom
Multiple R-squared:  0.9961,	Adjusted R-squared:  0.9958
F-statistic:  2582 on 2 and 20 DF,  p-value: < 2.2e-16
```

The regression equation is
RATE = 735.0 - 0.0810 EST + 0.000151 EST^2

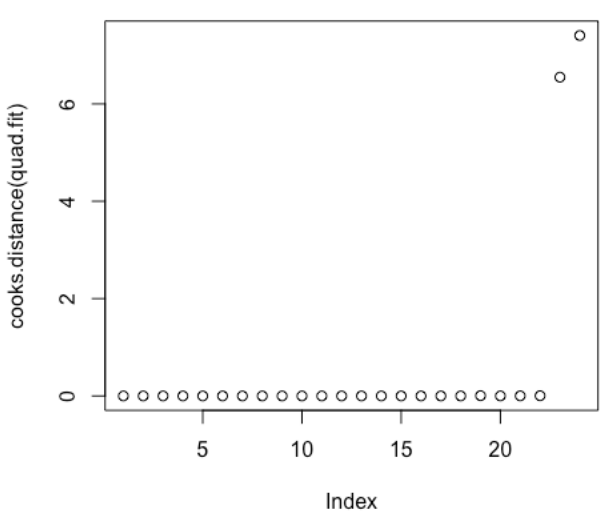
We test at $\propto = .05$ $H_o: \beta_2 = 0 \ vs \ H_A: \beta_2 \neq 0$

Our p value is $\frac{1.62(10^{-13})}{2} < .05$ We reject our null hypothesis and conclude that the data is curvilinearly related. The fit of the model has improved. Prior to removing the Botulism data point our Adj. R Squared was 0.4076. After removing the data point our Adj. R Squared increased to 0.9958



3) Identify any outliers in the INFECTION dataset. Are any of these outliers influential? If so, how should the researchers proceed?

	Rate	Estimation	Cook.Distance	F.50thpercentile
1	0.25	300.0	3.620496e-07	0.8148716
2	1.00	1000.0	2.172167e-05	0.8148716
3	1.75	691.0	8.459856e-06	0.8148716
4	2.00	200.0	2.163351e-09	0.8148716
5	3.00	17.5	1.679587e-06	0.8148716
6	5.00	0.8	2.031311e-06	0.8148716
7	9.00	1000.0	2.141802e-05	0.8148716
8	10.00	150.0	1.718607e-07	0.8148716
9	22.00	326.5	4.479051e-07	0.8148716
10	23.00	146.5	2.490307e-07	0.8148716
11	39.00	370.0	7.438326e-07	0.8148716
12	98.00	400.0	6.157626e-07	0.8148716
13	119.00	225.0	1.786763e-07	0.8148716
14	152.00	200.0	5.496425e-07	0.8148716
15	179.00	200.0	7.507440e-07	0.8148716
16	936.00	400.0	8.569887e-06	0.8148716
17	1514.00	1500.0	2.296657e-06	0.8148716
18	1627.00	1000.0	3.935263e-06	0.8148716
19	2926.00	6000.0	5.696970e-03	0.8148716
20	4019.00	5000.0	1.075258e-03	0.8148716
21	12619.00	1500.0	1.945861e-03	0.8148716
22	14889.00	10000.0	4.563122e-03	0.8148716
23	203864.00	37000.0	6.548829e+00	0.8148716
24	15.00	37500.0	7.405753e+00	0.8148716



From Cooks Distance it seems that data point 23 and 24 are influential outliers. The researchers should investigate why these outliers occurred.