We shall perform regression on the "Crest" data based on the model form,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

1. From R, we compute the least squares estimates as,

$$\beta_0 = 30625.907$$
$$\beta_1 = 3.893$$
$$\beta_2 = -29607.315$$
$$\beta_3 = 86.519$$

2. SSE, s^2 and s

   (a) We compute SSE, the sum of squared error, as

   $$SSE = 894568667.411638$$

   (b) In order to compute both s and s^2, we must compute k. k is equal to the subscript of the highest β_i. As such our model has $k = 3$. Then, noting that $n = 13$, we compute,

   $$s^2 = \frac{SSE}{n-(k+1)} = \frac{894568667.411638}{13-(3+1)} = \frac{894568667.411638}{9} \approx 99396519$$

   (c) We may then compute,

   $$s = \sqrt{s^2} \approx 9969.78$$

3. Variation

   (a) We compute, $TotalVariation = \sum(y_i - \bar{y})^2$. Then,

   $$T.V. \approx 21050596826.9231$$

   (b) Next, we compute $ExplainedVariation = \sum(\hat{y}_i - \bar{y})^2$. Then,

   $$E.V. \approx 20156028159.5114$$

   (c) Finally, we compute $UnexplainedVariation = \sum(y_i - \hat{y}_i)^2$. Then,

   $$U.V. \approx 894568667.4116$$

4. Using R, we compute $F(model) = 67.59$ with a $p-val = 1.709e-6$. Thus, with $\alpha = .05$, Thus, we reject the null hypothesis that all of the $\beta_i = 0$. Then, our alternative that $\beta_i \neq 0$ becomes plausible.

5. Using the generated t-statistics and associated p-values, we shall test each of the computed β_i using the null hypotheses $H_0: \ \beta_i = 0$ versus the alternative $H_A: \ \beta_i \neq 0$ with $\alpha = .05$. The results follow,

Variable	Beta Value	t-value	p-value	Significance
Intercept	β_0	1.546	0.15647	Not Significant at $\alpha = .05$
Budget	β_1	1.871	.09420	Not Significant at $\alpha = .05$
Ratio	β_2	-1.243	.24533	Not Significant at $\alpha = .05$
Income	β_3	4.628	.00124	Significant at $\alpha = .05$

   Similarly, we make the same comparison with $\alpha = .01$, and see that, again the income variable is the only one sigificant at this new level.

6. We shall now compute 90% confidence intervals for each β_i. The equation for our interval is,

$$\left[b_j \pm t_{\alpha/2}^{n-(k+1)} s_{b_j} \right]$$

   Using our previously computed k and n, we find

$$t_{.05}^{9} = 1.833$$

   Then, we compute the intervals as follows:

β_j	b_j	s_{b_j}	Interval
β_0	30625.907	19808.009	[-5682.173, 66933.988]
β_1	3.893	2.081	[.078527, 7.707473]
β_2	-29607.315	23822.087	[-73273.20047, 14058.57047]
β_3	86.519	18.693	[52.254731, 120.783269]

7. Next, we note that our equation of,

$$R^2 = \frac{Explained\ Variation}{Total\ Variation}$$

Thus,

$$R^2 \approx .9575039$$

To compute the adjusted R^2 value, we use the equation,

$$\bar{R}^2 = \left(R^2 - \frac{k}{n-1}\right)\left(\frac{n-1}{n-(k+1)}\right)$$

So,

$$\bar{R}^2 = \left(0.9575 - \frac{3}{13-1}\right)\left(\frac{13-1}{13-(3+1)}\right) \approx 0.9433385$$

8. We now consider the computation of a new point prediction and interval based on the new data of $Budget = 25,000$, $Ratio = 1.55$ and $Income = 1821.70$. Using our computed linear model and R, we compute,

$$\hat{y} = 239675.4$$

$$P.I. = [209507.3, 269843.5]$$

R-Code

```
> setwd("~/Desktop/PSM/Fall 2018/Regression/HW4")
> crest=read.table('crest.txt',header = TRUE)
> mod<-lm(Sales~Budget+Ratio+Income, data=crest)
> summary(mod)

Call:
lm(formula = Sales ~ Budget + Ratio + Income, data = crest)

Residuals:
   Min     1Q Median     3Q    Max
-24447  -2561   1195   3885   9527

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  30625.907  19808.009   1.546  0.15647
Budget           3.893      2.081   1.871  0.09420 .
Ratio       -29607.315  23822.087  -1.243  0.24533
Income          86.519     18.693   4.628  0.00124 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9970 on 9 degrees of freedom
Multiple R-squared:  0.9575,Adjusted R-squared:  0.9433
F-statistic: 67.59 on 3 and 9 DF,  p-value: 1.709e-06

> test=data.frame(Budget=16300, Ratio=1.25, Income=547.9)
> predict(mod,test)
       1
104479.3
> 105000-104479.3
[1] 520.7
> mod$residuals
          1           2           3           4           5
   520.6815   1195.3334   9527.2175  -2708.3077  -6270.0669
          6           7           8           9          10
  7144.6314   3885.3353 -24446.9075    868.6300   2758.4026
         11          12          13
  1446.1279   8639.8221  -2560.8997
> SSE=sum(mod$residuals^2)
> s2=SSE/(13-(3+1))
> s=sqrt(s2)
> s2
[1] 99396519
> s
[1] 9969.78
> print(s2)
[1] 99396519
> totvar=sum(crest$Sales-mean(crest$Sales))
> mean(crest$Sales)
[1] 148723.1
> diff=crest$Sales-mean(crest$Sales)
> diff
 [1] -43723.077 -43723.077 -27123.077 -34973.077 -34973.077
 [6] -19798.077  -6223.077 -22723.077  13276.923  42901.923
[11]  40276.923  61276.923  75526.923
> diff=diff^2
```

```
> sum(diff)
[1] 21050596827
> totvar=sum((crest$Sales-mean(crest$Sales))^2)
> totvar
[1] 21050596827
> X=data.frame(crest[3:5])
> test
  Budget Ratio Income
1  16300  1.25  547.9
> X
   Budget Ratio Income
1   16300  1.25  547.9
2   15800  1.34  593.4
3   16000  1.22  638.9
4   14200  1.00  695.3
5   15000  1.15  751.8
6   14000  1.13  810.3
7   15400  1.05  914.5
8   18250  1.27  998.3
9   17300  1.07 1096.1
10  23000  1.17 1194.4
11  19300  1.07 1311.5
12  23056  1.54 1462.9
13  26000  1.59 1641.7
> yhat=predict(mod,X)
> yhat
       1        2        3        4        5        6        7
104479.3 103804.7 112072.8 116458.3 120020.1 121780.4 138614.7
       8        9       10       11       12       13
150446.9 161131.4 188866.6 187553.9 201360.2 226810.9
> expvar=sum((yhat-mean(crest$Sales))^2)
> unexvar=sum((crest$Sales-yhat)^2)
> totvar-expvar-unexvar
[1] -4.529953e-06
> expvar
[1] 20156028160
> unexvar
[1] 894568667
> anova(mod)
Analysis of Variance Table

Response: Sales
          Df     Sum Sq    Mean Sq F value    Pr(>F)
Budget     1 1.7008e+10 1.7008e+10 171.117 3.681e-07 ***
Ratio      1 1.0184e+09 1.0184e+09  10.246   0.01081 *
Income     1 2.1292e+09 2.1292e+09  21.421   0.00124 **
Residuals  9 8.9457e+08 9.9397e+07
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
> fmodel=(expvar/3)/(unexvar/(13-(3+1)))
> summary(mod)

Call:
lm(formula = Sales ~ Budget + Ratio + Income, data = crest)

Residuals:
   Min     1Q Median     3Q    Max
-24447  -2561   1195   3885   9527
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  30625.907  19808.009   1.546  0.15647
Budget           3.893      2.081   1.871  0.09420 .
Ratio       -29607.315  23822.087  -1.243  0.24533
Income          86.519     18.693   4.628  0.00124 **
---
Signif. codes:
0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9970 on 9 degrees of freedom
Multiple R-squared:  0.9575,Adjusted R-squared:  0.9433
F-statistic: 67.59 on 3 and 9 DF,  p-value: 1.709e-06

> expvar/totvar
[1] 0.9575039
> (expvar/totvar-3/12)*(12/9)
[1] 0.9433385
> topred=data.frame(Budget=25000, Ratio=1.55, Income=1821.70)
> predict(mod, topred)
       1
239675.4
> predict.lm(mod, topred)
       1
239675.4
> predict(mod, topred, interval = "prediction")
       fit      lwr      upr
1 239675.4 209507.3 269843.5
```