Name:Shiva Kumar Peddapuram

UnivID: 811235874 (speddapu@kent.edu)
EXAM 1 – Math 40015/50015
Fall 2022

SHOW ALL YOUR WORK and write complete and coherent answers. No partial credit will be given if no work is shown. Please write as clearly and neatly as possible. If I cannot read your answers, I cannot give you any credit. Feel free to ask for more paper if you need more space. GOOD LUCK!!!

Questions-

Based on the data set "water" in package alr4, you are asked to predict stream runoff volume labelled BSAAM using precipitation measurements labelled by OPSLAKE.

1. Identify the response variable and predictor variable.

Response: - BSAAM

Predictor: - OPSLAKE

2. Draw the scatterplot of the response versus the predictor variable. Summarize the information in the graph.

Code: - I have initially installed all necessary libraries and then done with plotting and then done summary function.

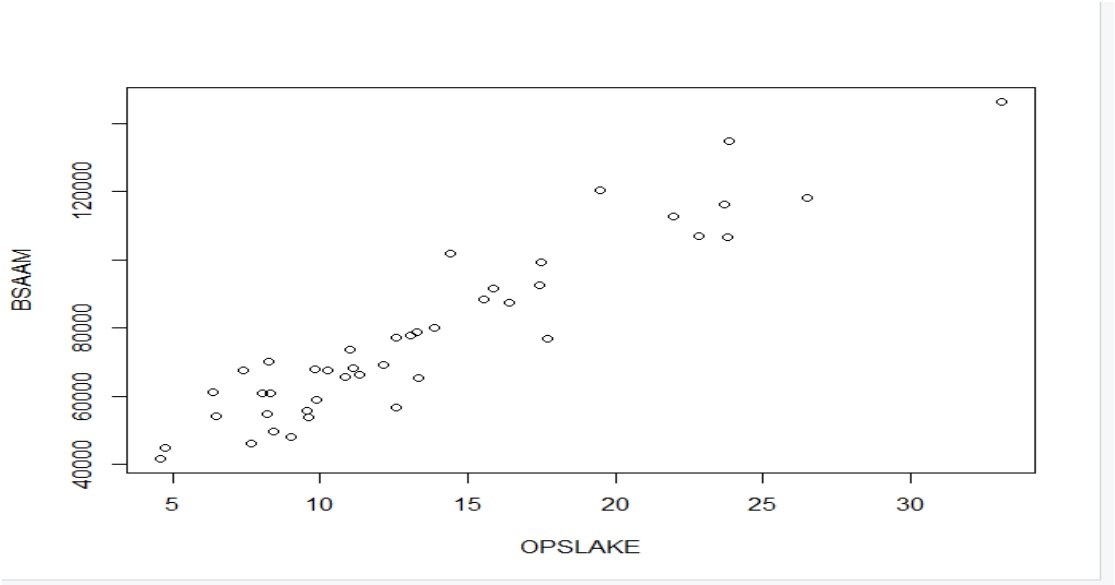
```
> install.packages("alr4")
Error in install.packages : Updating loaded packages
> install.packages("alr4")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools befor
e proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Warning in install.packages :
  package 'alr4' is in use and will not be installed
> library(alr4)
> a <- water #assigning the data package
> plot(y=water$BSAAM,x=water$OPSLAKE)
> lm(BSAAM ~ OPSLAKE, water)

Call:
lm(formula = BSAAM ~ OPSLAKE, data = water)

Coefficients:
(Intercept)      OPSLAKE
      27015         3752

> summary(water)
      Year          APMAM           APSAB          APSLAKE          OPBPC            OPRC          OPSLAKE          BSAAM
 Min.   :1948   Min.   : 2.700   Min.   : 1.450   Min.   : 1.77   Min.   : 4.050   Min.   : 4.350   Min.   : 4.600   Min.   : 41785
 1st Qu.:1958   1st Qu.: 4.975   1st Qu.: 3.390   1st Qu.: 3.36   1st Qu.: 7.975   1st Qu.: 7.875   1st Qu.: 8.705   1st Qu.: 59857
 Median :1969   Median : 7.080   Median : 4.460   Median : 4.62   Median : 9.550   Median :11.110   Median :12.140   Median : 69177
 Mean   :1969   Mean   : 7.323   Mean   : 4.652   Mean   : 4.93   Mean   :12.836   Mean   :12.002   Mean   :13.522   Mean   : 77756
 3rd Qu.:1980   3rd Qu.: 9.115   3rd Qu.: 5.685   3rd Qu.: 5.83   3rd Qu.:16.545   3rd Qu.:14.975   3rd Qu.:16.920   3rd Qu.: 92206
 Max.   :1990   Max.   :18.080   Max.   :11.960   Max.   :13.02   Max.   :43.370   Max.   :24.850   Max.   :33.070   Max.   :146345
> |
```



Conclusion: - I have concluded above is the graph and code

3. Obtain the estimated intercept, slope, and variance of the simple linear regression.

Estimated intercepts: -

```
> ## estimated intercepts
> p <- lm(BSAAM ~ OPSLAKE, water)
> coef(p)
(Intercept)     OPSLAKE
  27014.587    3752.486
> |
```

Slope: -

```
> slope(p)
Error in slope(p) : could not find function "slope"
> slope <- p
> slope

Call:
lm(formula = BSAAM ~ OPSLAKE, data = water)

Coefficients:
(Intercept)      OPSLAKE
      27015         3752

> |
```

Variance: -

```
> var(water)
                Year          APMAM          APSAB       APSLAKE         OPBPC          OPRC       OPSLAKE         BSAAM
Year     1.576667e+02 -2.952381e-02     1.3354762     4.824762 1.144905e+01 1.418571e+00 1.106119e+01 5.446117e+04
APMAM   -2.952381e-02  9.595279e+00     5.2616214     5.708760 2.914581e+00 2.405093e+00 2.125960e+00 1.885845e+04
APSAB    1.335476e+00  5.261621e+00     4.2116278     4.172506 6.238808e-01 1.090094e+00 3.878260e-01 9.599259e+03
APSLAKE  4.824762e+00  5.708760e+00     4.1725060     5.099938 1.622438e+00 1.208006e+00 1.449675e+00 1.436938e+04
OPBPC    1.144905e+01  2.914581e+00     0.6238808     1.622438 5.910628e+01 3.342705e+01 4.628457e+01 1.737757e+05
OPRC     1.418571e+00  2.405093e+00     1.0900937     1.208006 3.342705e+01 2.528275e+01 2.949468e+01 1.180011e+05
OPSLAKE  1.106119e+01  2.125960e+00     0.3878260     1.449675 4.628457e+01 2.949468e+01 4.072823e+01 1.528321e+05
BSAAM    5.446117e+04  1.885845e+04 9599.2586988 14369.383549 1.737757e+05 1.180011e+05 1.528321e+05 6.512147e+08
```

4. Obtain the standard errors of the estimated intercept and slope, respectively.

Answer: after plotting the graph I have got that standard errors of the estimated Intercept is 3218.9 and slope is 215.7

5. Add the fitted regression line to the scatterplot.

Code: -

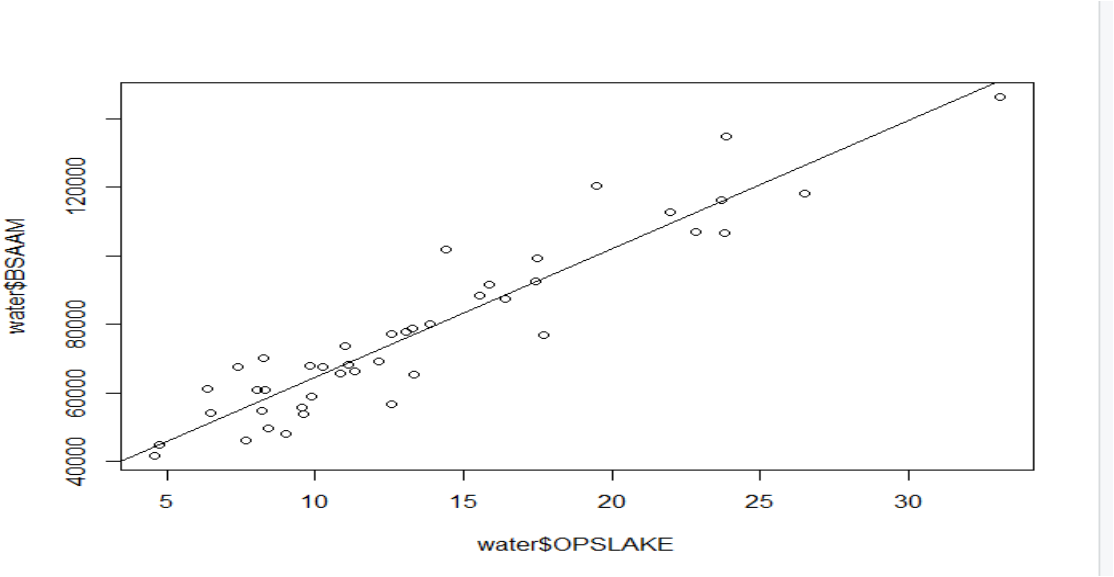
```
> ###fitted regression line
> lm(BSAAM ~ OPSLAKE, water)

Call:
lm(formula = BSAAM ~ OPSLAKE, data = water)

Coefficients:
(Intercept)      OPSLAKE
      27015         3752

> abline(lm(BSAAM ~ OPSLAKE, water))
>
```

Plot: -



Conclusion: - from the above graph by using abline function we can fit the regression line to the scatterplot.

6. Test the hypothesis that the slope is 0 versus the alternative that it is positive using the significance level $\alpha = 0.01$. You need to report the computed test statistic, p-value and summarize your decision and conclusion carefully.

7. Construct a 99% confidence interval for the slope. Report your interval with both lower and upper bounds and conclude what you find.

conclusion: - in the below graph I have constructed a 99% confidence interval for the slope and I have also found the lower upper bounds I conclude that the lower and upper bounds are varying.

```
> ###7
> predict(p,level=.99)
       1        2        3        4        5        6        7        8        9       10       11       12       13       14
51293.17 65515.09 69605.30 68779.75 112608.78 54820.51 76997.70 63788.95 92382.89 57972.59 92608.04 62888.35 55721.10 60786.96
      15       16       17       18       19       20       21       22       23       24       25       26       27       28
93358.53 85290.69 57784.97 76885.12 74145.81 115798.40 58085.17 151109.29 68291.93 67616.48 57259.62 86529.01 88630.40 74145.81
      29       30       31       32       33       34       35       36       37       38       39       40       41       42
44763.84 44276.02 109344.12 79099.09 116248.69 72569.76 100000.43 116548.89 81125.43 76022.05 126305.36 50880.40 64089.14 62963.40
      43
58572.99
>
```

By using interval=prediction

```
> ###7
> predict(p,interval="prediction",level=.99)
          fit       lwr       upr
1    51293.17  26569.48  76016.86
2    65515.09  41061.32  89968.86
3    69605.30  45192.69  94017.91
4    68779.75  44360.16  93199.34
5   112608.78  87635.47 137582.10
6    54820.51  30181.94  79459.07
7    76997.70  52617.64 101377.75
8    63788.95  39312.88  88265.01
9    92382.89  67897.53 116868.24
10   57972.59  33400.01  82545.18
11   92608.04  68119.41 117096.66
12   62888.35  38399.49  87377.21
13   55721.10  31102.36  80339.84
14   60786.96  36265.18  85308.73
15   93358.53  68858.66 117858.41
16   85290.69  60882.85 109698.52
17   57784.97  33208.73  82361.21
18   76885.12  52504.97 101265.27
19   74145.81  49759.59  98532.03
20 115798.40  90713.07 140883.72
21   58085.17  33514.77  82655.57
22 151109.29 124199.61 178018.96
23   68291.93  43867.90  92715.96
24   67616.48  43185.91  92047.05
25   57259.62  32672.95  81846.29
26   86529.01  62111.20 110946.82
27   88630.40  64192.21 113068.59
28   74145.81  49759.59  98532.03
29   44763.84  19851.55  69676.14
30   44276.02  19348.03  69204.01
31 109344.12  84475.75 134212.49
32   79099.09  54718.42 103479.75
33 116248.69  91146.81 141350.58
34   72569.76  48176.69  96962.84
35 100000.43  75377.15 124623.71
36 116548.89  91435.87 141661.92
37   81125.43  56740.04 105510.82
38   76022.05  51640.79 100403.31
39 126305.36 100786.46 151824.25
40   50880.40  26145.97  75614.82
41   64089.14  39617.17  88561.12
42   62963.40  38475.64  87451.16
43   58572.99  34011.89  83134.09
Warning message:
In predict.lm(p, interval = "prediction", level = 0.99) :
  predictions on current data refer to _future_ responses
```

8. For a precipitation measurement not in the data with OPSLAKE=22, obtain a 99% prediction interval for BASAAM.

```
> ##For a precipitation measurement not in the data with OPSLAKE=22, obtain a 99% prediction interval for BASAAM
> predict(p,data.frame(OPSLAKE=22),interval="prediction",level=.99)
       fit      lwr      upr
1 109569.3 84693.98 134444.6
>
```

Conclusion: - ii have found the precipitation measurement that is not in the data of OPSLAKE=22, and I have also obtained 99% prediction interval for Bassam.