Problem 1

Consider a company that markets and repairs small computers. To study the relationship between the length of a service call and the number of electronic components in the computer that must be repaired or replaced, a sample of records on service calls was taken. The data consist of the length of service calls in minutes (the response variable) and the number of components repaired (the predictor variable). The data are presented in the table below:

Minutes	23	29	49	64	74	87	96	97	109	119	149	145	154	166
Units	1	2	3	4	4	5	6	6	7	8	9	9	10	10

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.162      3.355    1.24    0.239
x             15.509      0.505   30.71 8.92e-13 ***
---

Residual standard error: 5.392 on 12 degrees of freedom
Multiple R-squared:  0.9874,    Adjusted R-squared:  0.9864
F-statistic: 943.2 on 1 and 12 DF,  p-value: 8.916e-13

Analysis of Variance Table

Response: y
          Df    Sum Sq     Mean Sq    F value    Pr(>F)
x          1    27419.5    27419.5    943.2      8.916e-13 ***
Residuals 12    348.8      29.1
```

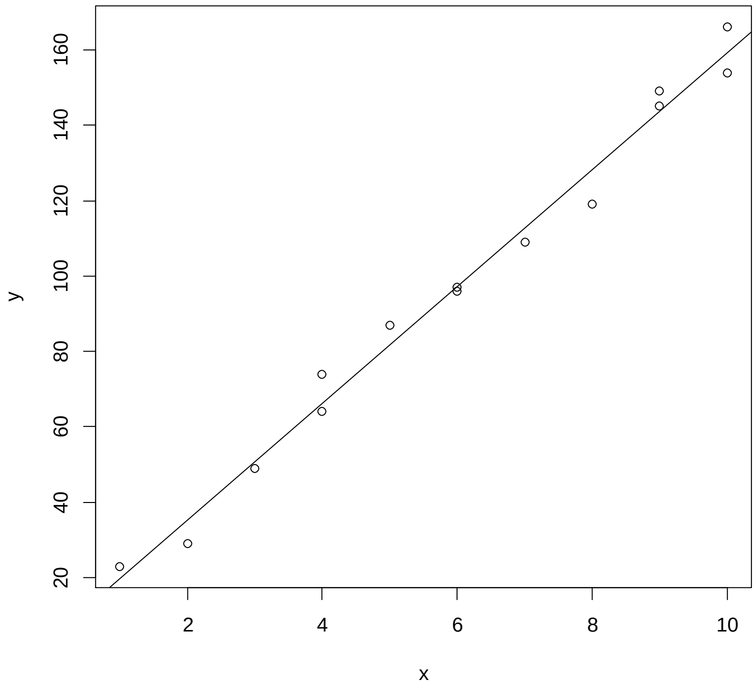
(a) Estimate the regression line and interpret the coefficients.

$$\hat{y} = b_o + b_1 x$$

$$\hat{y} = 4.16 + (15.51)x$$
$$(\widehat{service\ time}) = 4.16 + (15.51)(\#of\ units)$$

$b_1 =$ *The changes in service time when number of units increase by one.*
$b_o =$ *The intersection with y axis.*

```
> y=c(23,29,49,64,74,87,96,97,109,119,149,145,154,166)
>  x=c(1,2,3,4,4,5,6,6,7,8,9,9,10,10)
> model=lm(y~x)
> summary(model)
> with(plot(x,y),abline(model))
```



(b) Construct 90% confidence intervals for the model coefficients and explain the results.

```
> confint(model,level=0.90)
                   5 %       95 %
(Intercept)    -1.81810   10.14141
     x         14.60875   16.40879
```

$\beta_0 \in (-1.82\ , 10.14\ )$ $\beta_1 \in (14.61\ , 16.41\ )$

We are 90% sure that, when the number of units increase by one the service time increase somewhere between (14.41 , 16.61) *(c) Test the linearity by using two different approaches.*

1. *Using T-test :*

$$H_0: \beta_1 = 0 \quad vs \quad H_a: \beta_1 \neq 0$$

Since p-value = 0.000 <0.05 then the decision:
We reject H_0 (there is a linear association).

2. *Using F-test :*

$$H_0: \beta_1 = 0 \quad vs \quad H_a: \beta_1 \neq 0$$

Since p-value = 0.000 <0.05 then the decision:
We reject H_0 (there is a linear association).

(d) Calculate the residual at Units=4 and Minutes=64

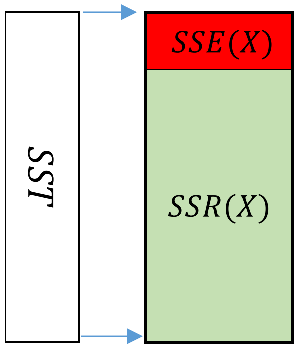
```
> summary(model)$res
        1          2          3          4          5          6          7
3.3295739 -6.1791980 -1.6879699 -2.1967419  7.8032581  5.2944862 -1.2142857
        8          9         10         11         12         13         14
-0.2142857 -3.7230576 -9.2318296  5.2593985  1.2593985 -5.2493734  6.7506266
```

(e) Estimate the standard deviation of the residuals.

$$S = \sqrt{MSE} = \sqrt{29.1} = 5.39$$

ANOVA:

	df	*SS*	*MS*	*F*
Regression (R)	1	SSR	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
Error (E)	$n - 2$	SSE	$MSE = \frac{SSE}{n-2}$	
Total (T)	$n - 1$	SST		



Problem 2:

A linear regression was run on a set of data. You are given only the following partial information:

Predictor	Coef	SE Coef	T
Constant	293.89	5.62	$\frac{293.89}{5.62} = 52.29$
X	$0.13 \times -13.13 = -1.7069$	0.13	-13.13

Analysis of Variance

Source	DF	SS	MS	F
Regression	1	7621.667	$172.3969 \times 44.21 = 7621.667$	$(-13.13)^2 = 172.3969$
Residual Error	5	$5 \times 44.21 = 221.05$	44.21	
Total	6	$7621.667 + 221.21 = 7842.717$		

(a) Compute the 90% Confidence intervals for β_0 and β_1

$$\left(b_0 \pm t_{(1-0.1/2, 7-2)}\ S(b_0)\right)$$

$$(293.89 \pm 2.015 \times\ 5.62)$$

$$(282.5657\ , 305.2143)$$

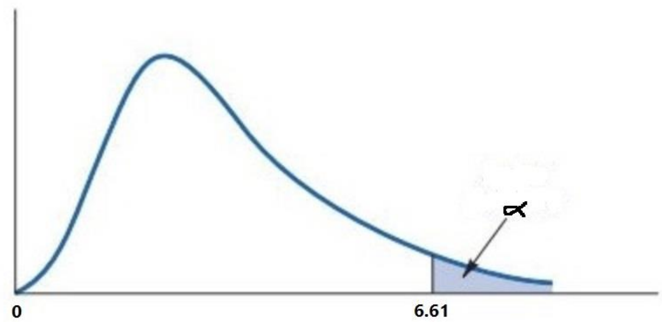
$$\left(b_1 \pm t_{(1-0.1/2, 7-2)}\ S(b_1)\right)$$

$$(-1.7069 \pm 2.015 \times\ 0.13)$$

$$(-1.96885\ , -1.44495)$$

(b) Give the F-statistic and test $H_0: \beta_1 = 0\ \ vs\ \ H_1: \beta_1 \neq 0$

$$F - statistics = 172.3969$$



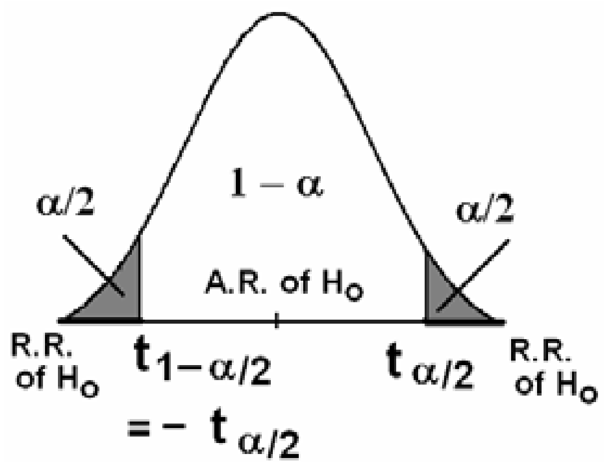
$$F_{(1-\alpha, 1, n-2)} = F_{(0.95, 1, 5)} = 6.60789 < 172.3969$$

$$We\ reject\ H_0$$

(c) Test $H_0: \beta_0 = 0 \ \ vs \ \ H_1: \beta_0 \neq 0$

$$T = \frac{b_0}{S(b_0)} = 52.2936$$

$$t_{1-\frac{\alpha}{2}, n-2} = t_{0.975\,,5} = 2.57058$$



The decision: we reject H_0

(d) Compute the coefficient of determination and hence the correlation coefficient.

$$R^2 = \frac{SSR}{SST} = \frac{7621.667}{7842.717} = 0.9718 \quad \Longrightarrow \quad r = -\sqrt{0.9718} = -0.9858$$

b_1 اشارة

Problem 3:

The computer repair data gives the length of time of service calls in minutes (y) and the number of components repaired in a computer (x). Some summary measures for this data are:

$$\boldsymbol{n = 14 \qquad \sum x_i = 84 \qquad \sum y_i = 1361}$$

$$\boldsymbol{S_{XX} = 114 \quad S_{YY} = 27768.36 \quad S_{XY} = 1768}$$

a. Find the point estimate of the intercept and slop to model the length of service call as a linear function of the number of units serviced.

$$\hat{y} = b_o + b_1 x$$

$$\bar{y} = \frac{\sum y}{n} = \frac{1361}{14} = 97.21 \qquad \bar{x} = \frac{\sum x}{n} = \frac{84}{14} = 6$$

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{1768}{114} = 15.51$$

$$b_0 = \bar{y} - b_1\bar{x} = 97.21 - 15.51 \times 6 = 4.16$$

$$\hat{y} = 4.16 + (15.51)x$$

$$(\widehat{time\ of service}) = 4.16 + (15.51)(\#of\ components)$$

b. Show that the error sum of square can be written as

$$SSE = S_{YY} - \hat{\beta}_1 S_{XY}$$

$$SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

$$= S_{yy} - \frac{S_{xy}S_{xy}}{S_{xx}}$$

$$= S_{yy} - b_1 S_{xy}$$

c. Give 95% confidence interval for the slop.

$$\left(b_1 \pm t_{(1-\alpha/2, n-2)}\ S(b_1)\right)$$

- $SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = 27768.36 - \frac{(1768)^2}{114} = 348.85$
- $MSE = \frac{SSE}{n-2} = \frac{348.85}{12} = 29.07$
- $Var(b_1) = \frac{MSE}{S_{xx}} = \frac{29.07}{114} = 0.255 \quad \Longrightarrow \quad S(b_1) = 0.505$
- $t_{(1-\alpha/2, n-2)} = t_{(1-0.05/2, 14-2)} = t_{(0.975, 12)} = 2.178813$

$$\left(b_1 \pm t_{(1-\alpha/2, n-2)}\ S(b_1)\right)$$

$$(15.51 \pm 2.178813 \times \ 0.505)$$

$$(\ 14.41\ ,16.61\ )$$

When the number of component increase by one the service time increase somewhere between (14.41 ,16.61)

Problem 4:

It is of interest to study the effect of population size in various cities in certain country on ozone concentration. The following data consists of the population in million and the amount of ozone present per hour in (parts per billion). The data is gives as follows.

i	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Ozone Y	*126*	*135*	*124*	*128*	*130*	*128*	*126*	*128*	*128*	*130*
Population X	*0.6*	*4.9*	*0.2*	*0.5*	*1.1*	*0.1*	*1.1*	*2.3*	*0.6*	*2.3*

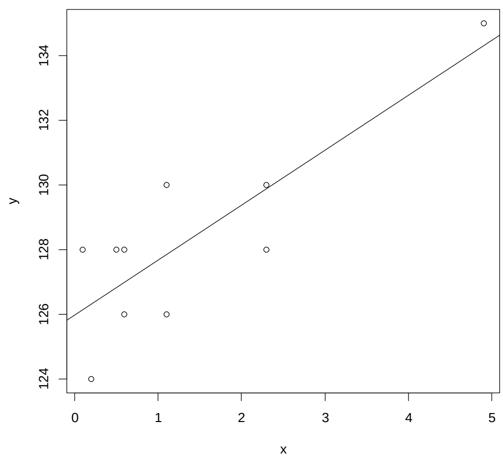
a. *Fit the linear regression model relating ozone concentration to the population size and explain the estimated model.*

```
> y=c(126,135,124,128,130,128,126,128,128,130)
> x=c(0.6,4.9,0.2,0.5,1.1,0.1,1.1,2.3,0.6,2.3)
> model=lm(y~x)
> summary(model)
Coefficients:
             Estimate   Std. Error   t value     Pr(>|t|)
(Intercept)  125.9677     0.7741     162.732     2.27e-15 ***
x            1.7024       0.3969     4.289       0.00266 **
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.742 on 8 degrees of freedom
Multiple R-squared:  0.6969,    Adjusted R-squared:  0.659
F-statistic: 18.39 on 1 and 8 DF,  p-value: 0.002656

> with(plot(x,y),abline(model))
```

$$\hat{y} = b_o + b_1 x$$

$$\hat{y} = 125.97 + (1.7)x$$

$$(\widehat{Ozone}) = 125.97 + (1.7)(Pop)$$



b_1 =*the changes in Ozone concentration when the population increase by one million.*

b_o =*the Ozone concentration when the population =0, and its the intersection with y axis.*

b. *Test the hypothesis* $H_0: \beta_1 = 0$.

Since p-value=0.00266 < 0.05

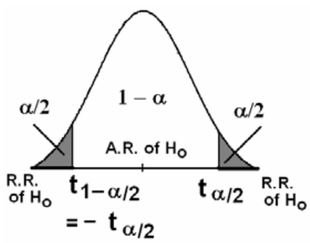
The decision: reject H_0 *(there is positive linear association between population size and Ozone concentration).*

c. *Test the hypothesis* $H_0: \beta_0 = 0.6$.

$$T = \frac{b_0 - \beta_0^{(0)}}{S(b_0)} = \frac{125.9677 - 0.6}{0.7741} = 161.95$$

$161.95 \notin (-2.306\,, 2.306)$

The decision: reject H_0



$$t_{1-\frac{\alpha}{2}, n-2} = t_{0.975\,,8} = 2.306$$

d. *Construct 90% confidence interval for the coefficients.*

```
> confint(model,Level=0.90)
                 2.5 %       97.5 %
(Intercept)  124.1826711  127.752742
x              0.7870611    2.617747
```

$\beta_0 \in (124.18\,, 127.75)$

$\beta_1 \in (0.787\quad, 2.618\ )$

e. *Find the coefficient of determination and the correlation and interpret the result.*

$$R^2 = 0.6969$$

The model explain 69.69% of variation in Ozone concentration(Y) by using population size(X)

f. *Construct 95% confidence interval for the mean Y when X=11*

```
> newx=data.frame(x=11)
> predict(model,newx,level=0.95,int="confidence")
       fit          lwr         upr
1 144.6941    135.7883      153.6
```

$Y \in (135.79\,, 153.6\ )$

Problem 5:

Suppose a sample of size 12 is used to estimate a simple linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$ *and obtain a 95% level confidence interval for the slope coefficient of (-0.045,-0.021). Based on the given information, complete the following statements (keep three decimal digits during the calculations):*

$$\beta_1 \in (-0.045, -0.021) \quad , \quad \alpha = 0.05 \quad , \quad n = 12$$

***(a)** The point estimate for the slope is:*

$$b_1 = \frac{(-0.045) + (-0.021)}{2} = -0.033$$

***(b)** The standard error for the slope is:*

$$b_1 - t_{\left(1-\frac{\alpha}{2}, n-2\right)} \; S(b_1) = -0.045$$

$$b_1 - t_{\left(1-\frac{0.05}{2}, 12-2\right)} \; S(b_1) = -0.045$$

$$-0.033 - (2.22814) \; S(b_1) = -0.045$$

$$-(2.22814) \; S(b_1) = -0.045 + 0.033$$

$$-(2.22814) \; S(b_1) = -0.012$$

$$\boxed{S(b_1) = 0.00539}$$

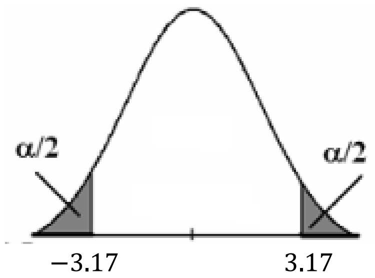
***(c)** The value of the test statistic for testing the slope is equal to 0 is:*

$$T = \frac{b_1}{S(b_1)} = \frac{-0.033}{0.00539} = -6.122$$

***(d)** The decision of the test in (c) at 1% level of significance is:*

$$H_0: \beta_1 = 0 \quad vs \quad H_1: \beta_1 \neq 0$$

$$\alpha = 0.01 \quad \Longrightarrow \quad t_{(1-\alpha/2, n-2)} = t_{(0.995,10)} = 3.16927$$



$$T = -6.122 \notin (-3.17, 3.17) \quad \textit{Then we reject } H_0$$

***(e)** The probability that the true (population) slope is between -0.045 and -0.021 is:*

0.95