Tutorial-06 MC3020

1. $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = \overline{1,10}$

y_i is the price of the kilo of flour.

x_i is the production of wheat.

β_0 is the intercept parameter.

β_1 is the slope parameter.

ε_i is the error term, $\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$

$\sum x_i = 286 \qquad \sum y_i = 354 \qquad \sum y_i^2 = 13268$

$\sum x_i^2 = 8468 \qquad \sum x_i y_i = 9734 \qquad n = 10$

$\bar{x} = 28.6 \qquad \bar{y} = 35.4$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{10} x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{9734 - 10 \times 28.6 \times 35.4}{8468 - 10 \times (28.6)^2}$$

$$= -1.353675$$

$$= -1.3537$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$= 35.4 - (-1.353675) \times 28.6$$

$$= 74.1151$$

$$\hat{y} = 74.1151 - 1.3537x$$

Scanned with CamScanner

b) $MSE = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}$.

Y_i	25	30	27	40	42	40
X_i	30	28	32	25	25	25
$\hat{Y}_i$	33.5041	36.21115	30.7967	40.2726	40.2726	40.2726
$(Y_i-\hat{Y}_i)^2$	72.3197	38.5827	14.4149	0.0743	2.9839	0.0743

Y_i	50	45	30	25
X_i	22	24	35	40
$\hat{Y}_i$	44.3337	41.6263	26.7356	19.9671
$(Y_i-\hat{Y}_i)^2$	32.107	11.3819	10.6563	25.3301

$$MSE = \frac{207.9251}{8}$$

$$= 25.9906.$$

c) 95% C.I for the slope (β_1)

$$\hat{\beta}_1 \pm t_{n-2,\alpha/2} * \frac{S_e}{\sqrt{\sum x_i^2 - n\bar{x}^2}} \qquad S_e = \sqrt{MSE}$$

$$= -1.3537 \pm t_{8,0.025} \times \sqrt{\frac{25.9906}{8468 - 10\times 28.6^2}}$$

$$= -1.3537 \pm 2.306 \times \sqrt{\frac{25.9906}{288.4}}$$

$$= (-2.046, -0.6614)$$

Scanned with CamScanner

d) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

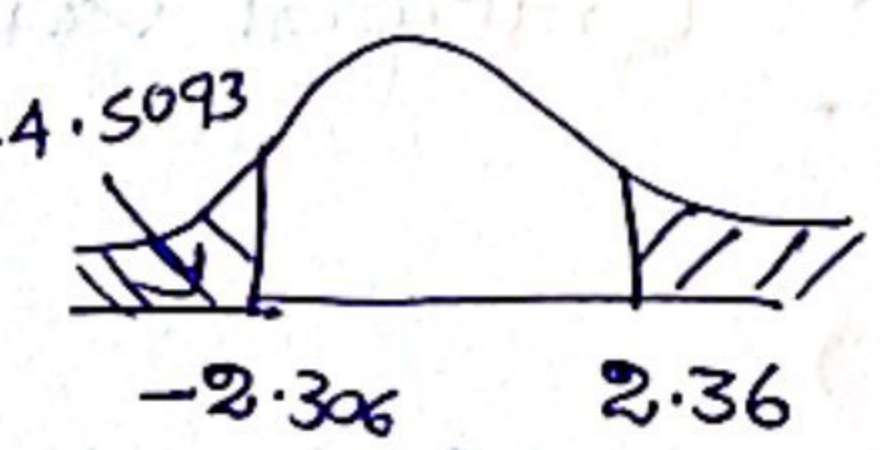
Test statistic $= \dfrac{\hat{\beta}_1 - 0}{S_e / \sqrt{\Sigma x_i^2 - n\bar{x}^2}}$

$= \dfrac{-1.3537}{\sqrt{\dfrac{25.9906}{288.4}}}$

$= -4.5093$

Critical value $t_{n-2, \alpha/2} = t_{8, 0.025} = 2.306$

Test value falls in rejection region. reject H_0.



So the price of flour depends linearly on the production of wheat. the test is carried out with 0.05 significance level.

e) 95% confidence interval for intercept (β_0)

$\hat{\beta}_0 \pm t_{n-2, \alpha/2} * S_e \sqrt{1/n + \dfrac{\bar{x}^2}{\Sigma x_i^2 - n\bar{x}^2}}$

$= 74.1151 \pm t_{8, 0.025} * \sqrt{25.9906 \left(1/10 + \dfrac{28.6^2}{288.4} \right)}$

$= 74.1151 \pm 2.306 . \sqrt{\dfrac{25.9906 \times 8468}{10 \times 288.4}}$

$= (53.9704 , 94.2578)$

Scanned with CamScanner

f) $H_0 : \beta_0 = 0$
$H_1 : \beta_0 \neq 0$.

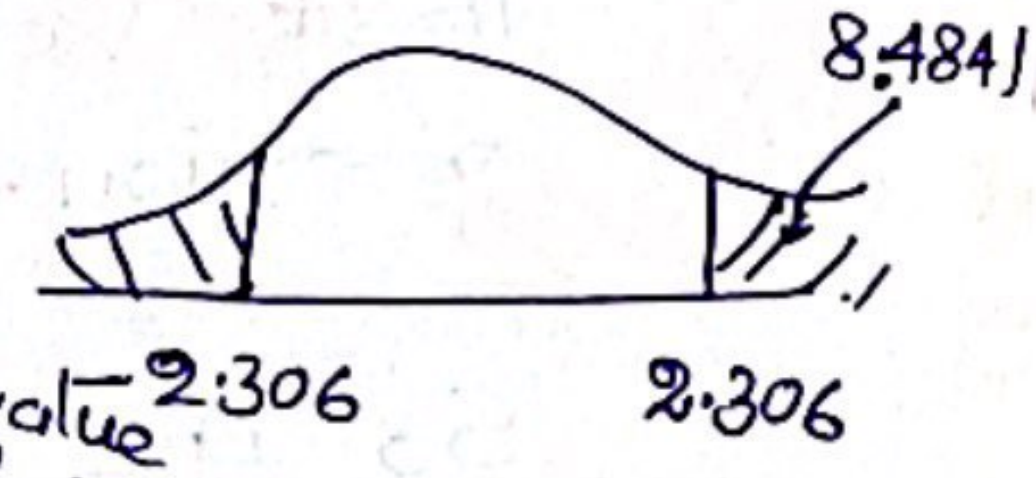
$$\text{Test statistic} = \frac{\hat{\beta}_0 - 0}{S_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2 - n\bar{x}^2}}}$$

$$= \frac{74.1151}{\sqrt{25.9906 \times \left(\frac{1}{10} + \frac{28.6^2}{288.4}\right)}}.$$

$$= \frac{74.1151}{\sqrt{\frac{25.9906 \times 8468}{10 \times 288.4}}}$$

$$= 8.4841.$$

Critical value $= t_{8,0.025} = 2.306$.



Test value $8.4841 > 2.306$ Critical value
Test value falls in rejection region.
So reject H_0.

$\therefore$ Regression line doesnot passes through the origin. the test is carried out at 0.05 significance level.

g) $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$
$= 74.1151 - 1.3537 \times 30$
$= 33.5041$.

Scanned with CamScanner

2)

a) Insurance premium depend on the driving experience.

X - Driving experience.

Y - Insurance premium.

b) $Y_i = B_0 + B_1 X_i + \varepsilon_i \quad i = \overline{1,8}$

Y_i is the Auto insurance premium.

x_i is the driving experience.

B_0 is the intercept parameter.

B_1 is the slope parameter

ε_i is the error term $\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$

$\Sigma x_i = 90 \quad \Sigma x^2 = 1396 \quad \Sigma Y = 474 \quad \Sigma Y^2 = 29642.$

$\Sigma xy = 4739. \quad n = 8$

$\bar{x} = 11.25 \quad \bar{Y} = 59.25.$

$$\hat{B}_1 = \frac{\sum_{i=1}^{8} x_i Y_i - n\bar{x}\bar{Y}}{\Sigma x_i^2 - n\bar{x}^2}.$$

$$= \frac{4739 - 8 \times 11.25 \times 59.25}{1396 - 8 \times (11.25)^2}$$

$= -1.547588$

$= -1.5476$

$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{x}$

$= 59.25 - (-1.547588) \times 11.25$

$= 76.660365$

$= 76.6604.$

Scanned with CamScanner

$\hat{y} = 76.6604 - 1.5476x$

c) The value of $B_0 = 76.6604$ gives the value of $\hat{y}$ for $x=0$, that is, it gives the monthly auto insurance premium for a driver with no driving experience.
(We should not attach much important to this statement because the sample contains drivers with only 2 or more year of experience).

$B_1 = -1.5476$ indicates that on average for every extra year of driving experience, the monthly auto insurance premium decreases by 1.55.

d) $$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}}$$

$$= \frac{4739 - 8 \times 11.25 \times 59.25}{\sqrt{(1396 - 8 \times (11.25)^2)(29642 - 8 \times (59.25)^2)}}$$

$= -0.767934$
$= -0.7679.$

$r = -0.7679$ indicates that the driving experience and the monthly auto insurance premium are negatively related.

Scanned with CamScanner

$$r^2 = (-0.7679)^2$$

$$= 0.5897$$

58.97% of the total variation in insurance premium(Y) is explained by the linear relationship between years of driving experience (x) & insurance premium(Y)

e)

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

Test statistic : $T = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$

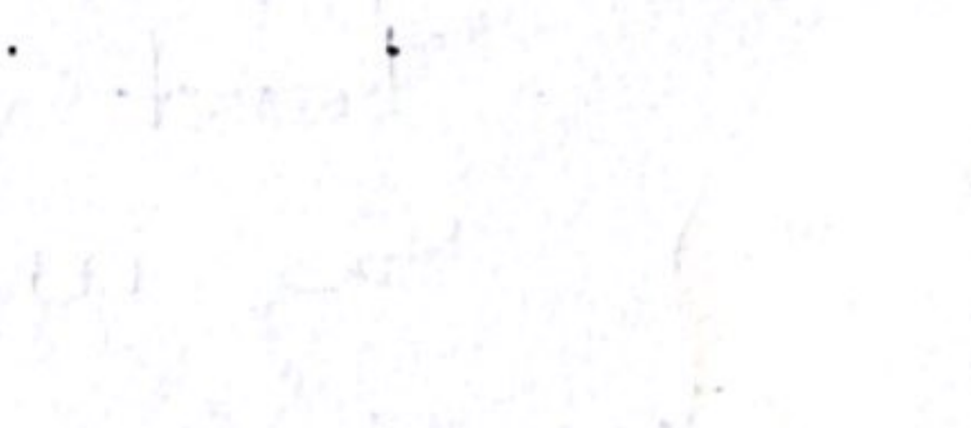
$$= \frac{-0.7679}{\sqrt{\dfrac{1-(-0.7679)^2}{6}}}$$

$$= -2.93649$$

$$= -2.9365$$

Critical value $= t_{n-2,\alpha/2} = t_{6,0.025} = -2.4469$

Test value falls in rejection region. So we can ~~to~~ reject H_0.



$\therefore$ We can conclude that $\rho \neq 0$ at 5% level of significance.

Scanned with CamScanner

f) Step 1.

$H_0 : B_1 \geq 0$

$H_1 : B_1 < 0$ (left tail test)

Step 2

$$\text{Test statistic} = \frac{\hat{B}_1 - 0}{S_e / \sqrt{\sum x_i^2 - n\bar{x}^2}}$$

$$= \frac{-1.5476}{\frac{10.3198}{\sqrt{383.5}}}$$

$$= -2.9368$$

Step-3

Critical value $= t_{n-2,\alpha} = t_{6,0.05} = -1.943$.

Step-4

(test value) $-2.9368 < -1.943$ (Critical value)

test value falls in rejection region.

So we can reject H_0.

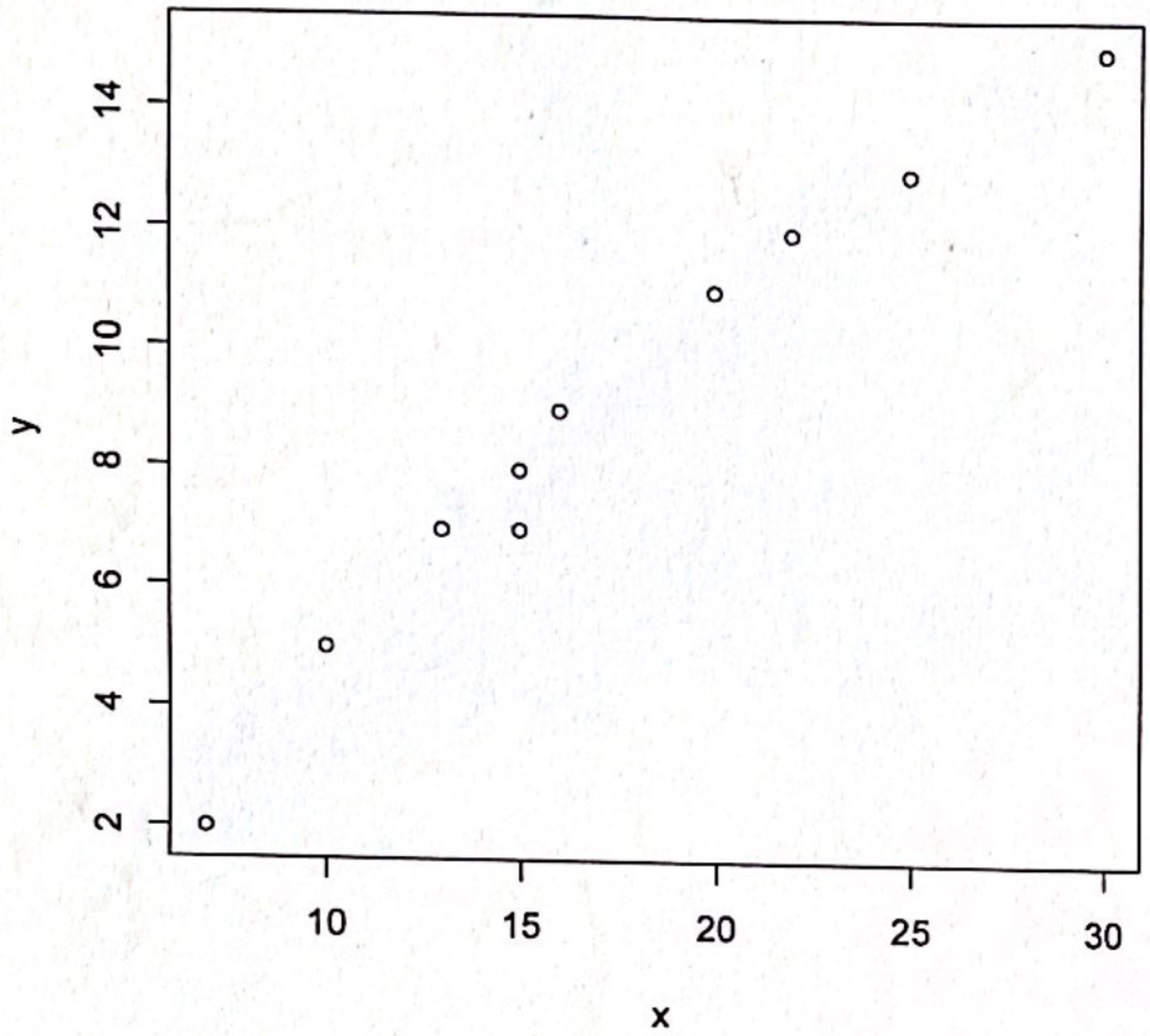
Step-5.

∴ The monthly auto insurance premium(Y) decreases with an increase in years of driving experience(X). and the test is carried out with 5% level of significance.

Scanned with CamScanner

3)

(a)



Scanned with CamScanner

3)

a) Plot

b) $\therefore \Sigma x^2 = 3433$ $\Sigma y^2 = 931$ $\Sigma xy = 1783$

$\Sigma x = 173$ $\Sigma y = 89$ $n = 10$

$\bar{x} = 17.3$ $\bar{y} = 8.9$

x - Number of Missing rivets.
y - Alignment Errors.

$$\hat{\beta}_1 = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}$$

$$= \frac{1783 - 10 \times 17.3 \times 8.9}{3433 - 10 \times 17.3^2}$$

$$= 0.5528289$$

$$= 0.5528$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$= 8.9 - (0.5528289)(17.3)$$

$$= -0.663940$$

$$= -0.6639$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$= -0.6639 + 0.5528x$$

Scanned with CamScanner

c) Step 1.

$H_0 : \beta_1 = 1$

$H_1 : \beta_1 \neq 1$

Step 2.

$$\text{Test statistic} = \frac{\hat{\beta}_1 - \beta_1}{S_e / \sqrt{\Sigma x_i^2 - n\bar{x}^2}}$$

$$\hat{Y}_i = 0.6639 + 0.5528 X_i$$

$$= \frac{0.5528 - 1}{0.7413 / 20.9786}$$

$$\sqrt{\Sigma x_i^2 - n\bar{x}^2}$$

$$= \sqrt{3433 - 10 \times (17.3)^2}$$

$$= \sqrt{440.1}$$

$$= 20.97856$$

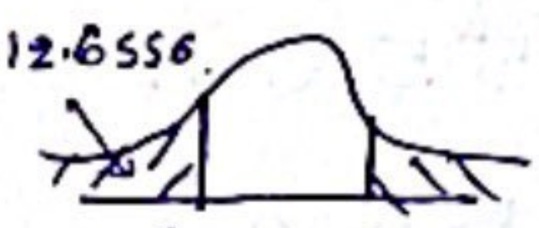
$$= 20.9786$$

$$= -12.6556$$

Step 3

Critical value $= t_{n-2, \alpha/2} = t_{8, 0.025} = -2.306$.

x_i	y_i	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$
13	7	6.5525	0.22800
15	7	7.6281	0.39450
10	5	4.8641	0.018468
22	12	11.4977	0.25230
30	15	15.9201	0.84658
7	2	3.2057	1.453712
25	13	13.1561	0.02436
16	9	8.1809	0.670924
20	11	10.3921	0.36954
15	8	7.6281	0.138309
			4.39673



Step 4

Test value falls in the rejection region. So we can reject H_0.

Step 5

So we can conclude that the Slope(β_1) is not equal to 1. the test is carried out with 5% significance level.

$$S_e = \sqrt{\frac{\Sigma (Y_i - \hat{Y}_i)^2}{n-2}}$$

$$= \sqrt{\frac{4.396731}{8}}$$

$$= \sqrt{0.549491}$$

$$= 0.7413$$

Scanned with CamScanner

d) $\hat{Y} = -0.6639 + 0.5528X$

when $X = 24$

$\hat{Y} = -0.6639 + 0.5528 \times 24$

$= 12.6033$.

$(11.98437, 13.22354)$

e) $x_0 = 24$.

90% C.I for mean response at $x_0 = 24$,

$y(x_0) = \hat{B}_0 + \hat{B}_1 x_0 = 12.6033$

$$\hat{B}_0 + \hat{B}_1 x_0 \pm t_{n-2,\alpha/2} * S_e \sqrt{1/n + \frac{(x_0 - \bar{X})^2}{\sum x_i^2 - n\bar{X}^2}}.$$

$$= 12.6033 \pm t_{8,0.05} * 0.7413 \sqrt{1/10 + \frac{(24-17.3)^2}{3433 - 10 \times 17.3^2}}$$

$$= 12.6033 \pm 1.8595 * 0.7413 \sqrt{1/10 + \frac{(24-17.3)^2}{3433 - 10 \times 17.3^2}}$$

$= 12.6033 \pm 0.619534$

$= (11.9838, 13.2228)$

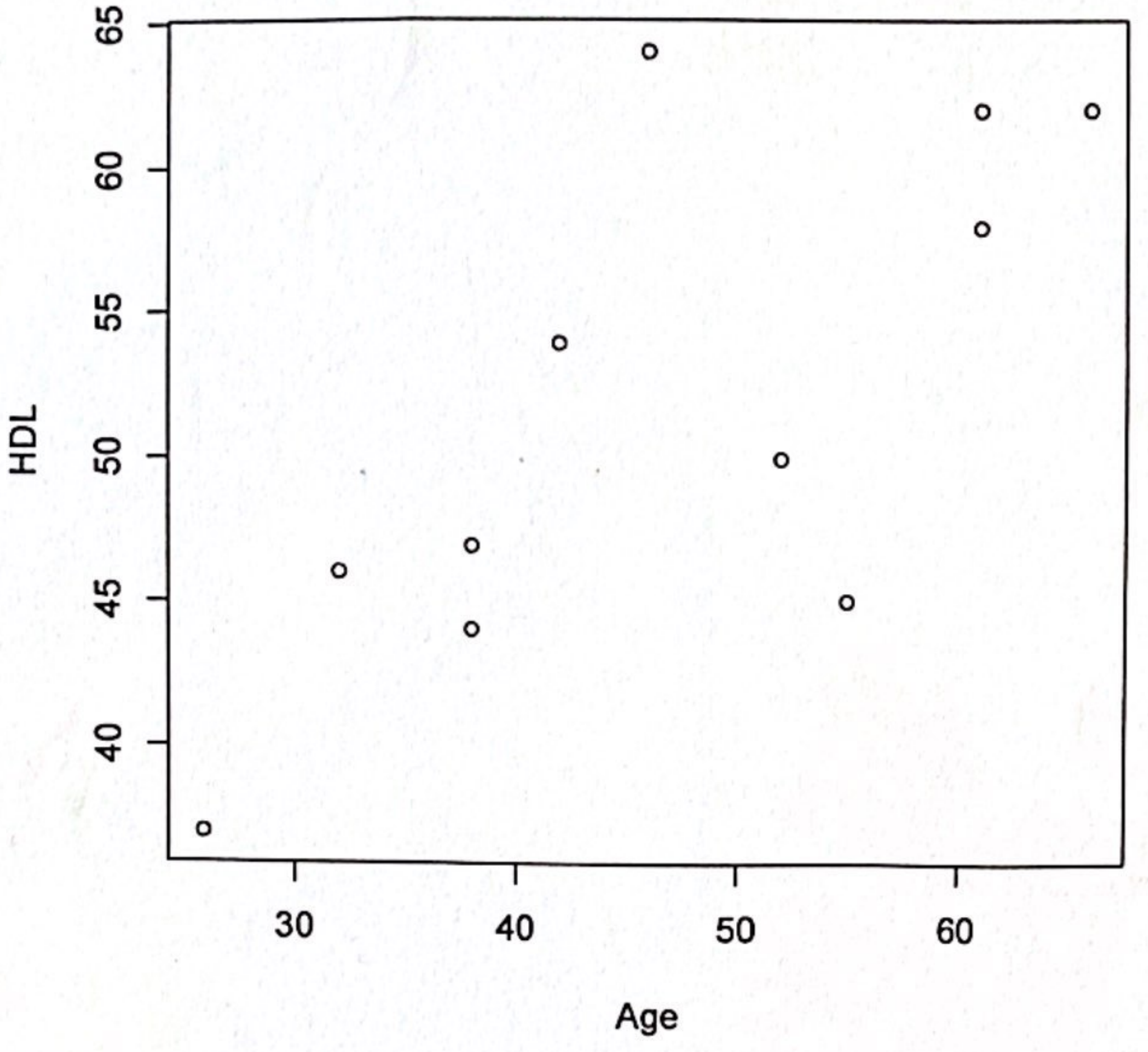
4)

```
> Age<-c(38,42,46,32,55,52,61,61,26,38,66)
> HDL<-c(47,54,64,46,45,50,62,58,37,44,62)
>
> plot(HDL~Age)
> cor(HDL,Age)
[1] 0.734381
> reg<-lm(HDL~Age)
> reg

Call:
lm(formula = HDL ~ Age)

Coefficients:
(Intercept)        Age
     28.23        0.50

> summary(reg)

Call:
lm(formula = HDL ~ Age)

Residuals:
     Min       1Q   Median       3Q      Max
-10.7273  -3.7273  -0.2273   2.5227  12.7727

Coefficients:
            Estimate Std. Error t value  Pr(>|t|)
(Intercept)   28.227      7.488   3.770   0.00442 **
Age            0.500      0.154   3.246   0.01006 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.344 on 9 degrees of freedom
Multiple R-squared:  0.5393,    Adjusted R-squared:  0.4881
F-statistic: 10.54 on 1 and 9 DF,   p-value: 0.01006
```

Scanned with CamScanner

4)

a)



According to the scatter diagram, we can say that there is a moderate positive correlationship between HDL, Age.

Scanned with CamScanner

4) a) plot.

b) $r = 0.734381$
$= 0.7344$

There is a positive correlation between age and HDL.

c) $\hat{y} = 28.23 + 0.5X$

d) the value of $\beta_0 = 28.23$ gives the value of $\hat{y}$ for $x = 0$, However in general if $x = 0$ is not within the range of the given x values the interpretation of β_0 is meaningless as is the case here.

$\hat{\beta}_1 = 0.5$ implies that, when age is increase by one year, the HDL will increase by 0.5 unit.

e) $HDL = 28.23 + (0.5 \times 43)$
$= 49.73.$

f) $R^2 = 53.98\%$

which means that 53.98% of the total variation in HDL (Y) can be explained by the linear relationship between Age (X) & HDL (Y).

Scanned with CamScanner

5)

```
> cor.test(Price,Consumption,alternative ="two.sided",conf.level = 0.95)

        Pearson's product-moment correlation

data:  Price and Consumption
t = 1.0152, df = 8, p-value = 0.3397
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3706380  0.7977637
sample estimates:
      cor
0.3378241

> reg<-lm(Consumption~Price)
> reg

Call:
lm(formula = Consumption ~ Price)

Coefficients:
(Intercept)       Price
    -0.4878      0.6110

> summary(reg)

Call:
lm(formula = Consumption ~ Price)

Residuals:
    Min       1Q    Median       3Q      Max
-0.10546 -0.03927 0.01587 0.04462 0.06287

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.4878     0.8223  -0.593    0.569
Price         0.6110     0.6018   1.015    0.340

Residual standard error: 0.0597 on 8 degrees of freedom
Multiple R-squared:  0.1141,	Adjusted R-squared:  0.003391
F-statistic: 1.031 on 1 and 8 DF,  p-value: 0.3397
```

Scanned with CamScanner

```
> anova(reg)
Analysis of Variance Table

Response: Consumption
          Df   Sum Sq    Mean Sq   F value Pr(>F)
Price      1  0.0036732 0.0036732  1.0306 0.3397
Residuals  8  0.0285124 0.0035641
>
> reg1<-lm(Consumption~Price+Income+Temperature)
> reg1

Call:
lm(formula = Consumption ~ Price + Income + Temperature)

Coefficients:
(Intercept)      Price      Income    Temperature
  -0.052594   0.746743   -0.002199     0.003033

> summary(reg1)

Call:
lm(formula = Consumption ~ Price + Income + Temperature)

Residuals:
     Min       1Q   Median       3Q      Max
-0.02451 -0.01865 -0.01377  0.01142  0.07797

Coefficients:
              Estimate  Std. Error t value Pr(>|t|)
(Intercept) -0.0525944  0.5969299  -0.088   0.9327
Price        0.7467433  0.4694484   1.591   0.1628
Income      -0.0021990  0.0018121  -1.214   0.2705
Temperature  0.0030330  0.0008346   3.634   0.0109 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03835 on 6 degrees of freedom
Multiple R-squared:  0.7259,    Adjusted R-squared:  0.5888
F-statistic: 5.296 on 3 and 6 DF,  p-value: 0.04014
```

Scanned with CamScanner

Q5)

a) $H_0: \rho = 0$

$H_1: \rho \neq 0$

$\alpha - 0.05$

P-value $= 0.3397$

P-value $> \alpha$

$\therefore$ So Do not reject H_0

So There is no linear correlation between Consumption & price at 5% significance Level.

b) $\hat{y} = -0.4878 + 0.6110X$.

a) $R^2 = 11.41\%$

Which means that 11.41% of the total variation in Consumption (Y) can be explained by the model

e) $\hat{y} = -0.526 + 0.7467\,(\text{price}) + 0.0022\,(\text{Income}) + 0.0030\,(\text{Temperature})$

d) Temperature.

g) $R^2 = 0.5888$

58.88% of total variation in Constumption (Y) can be explained by the model.

d) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.05$

P-value $= 0.340$

P-value > 0.05

$\therefore$ So Do not reject H_0,

Predicting consumption by using price is not significant at 5% level of consumption.

Scanned with CamScanner

h)

H_0 : $\beta_1 = \beta_2 = \beta_3 = 0$

H_1 : at least one of the coefficient ($\beta_1, \beta_2, \beta_3$) not equal to zero

$\alpha = 0.05$

P-value = 0.04014

P-value < α (0.05)

$\therefore$ Reject H_0

The model in predicting consumption using price, income and temperature as a predicting variable is significant.

Scanned with CamScanner

6)

a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i=\overline{1,5}$

y_i is the compressive strength.

x_i is the curing time.

β_0 is the intercept parameter.

β_1 is the slope parameter.

ε_i is the error term. $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

$\sum x_i = 30 \quad \sum y_i = 120 \quad \sum x_i y_i = 996$

$\sum x_i^2 = 220 \quad \sum y_i^2 = 5358 \quad n=5 \quad \bar{x}=6 \quad \bar{y}=24$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{996 - 5\times6\times24}{220 - 5\times36}$$

$$= 6.9$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$= 24 - 6.9\times6$$

$$= -17.4$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$= -17.4 + 6.9x$$

b) $\beta_1 = 6.9$

When curing time is increased by one day compressive strength is increased by 6.9 Unit (MPa)

Scanned with CamScanner

c) When $x_0 = 12$.

$$\hat{y} = -17.4 + 6.9x$$
$$= -17.4 + 6.9 \times 12$$
$$= 65.4$$

d) $$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}}$$

$$= \frac{996 - 5 \times 6 \times 24}{\sqrt{(220 - 5 \times 6^2)(5358 - 5 \times 24^2)}}$$

$$= 0.8767$$

$$r^2 = 0.7685$$

76.85% of the total variation in compressive strength(y) is explained by the model.

e) $x_0 = 12$

$\hat{y} = 65.4$, $\quad t_{n-2, \alpha/2} = t_{3, 0.025} = 3.182$.

$$S_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

$$= 13.8275$$

Scanned with CamScanner

∴ 95% C.I for the avg compressive strength of concrete at 12 days of curing time

$$= \hat{\beta}_0 + \hat{\beta}_1 x_0 + t_{n-2,\alpha/2} * s_e \sqrt{1/n + \frac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}}$$

$$= 65.4 \pm 3.182 * 13.8275 \sqrt{1/5 + \frac{(12-6)^2}{40}}$$

$$= (19.2469,\ 111.5532)$$

f) 95% Prediction interval for an individual compressive strength of concrete at 12 days

$$\hat{B}_0 + \hat{B}_1 x_0 \pm t_{n-2,\alpha/2} * s_e \sqrt{1 + 1/n + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}}$$

$$= 65.4 \pm 3.182 * 13.8275 \sqrt{1 + 1/5 + \frac{(12-6)^2}{40}}$$

$$= (1.6303,\ 129.1697)$$

Scanned with CamScanner