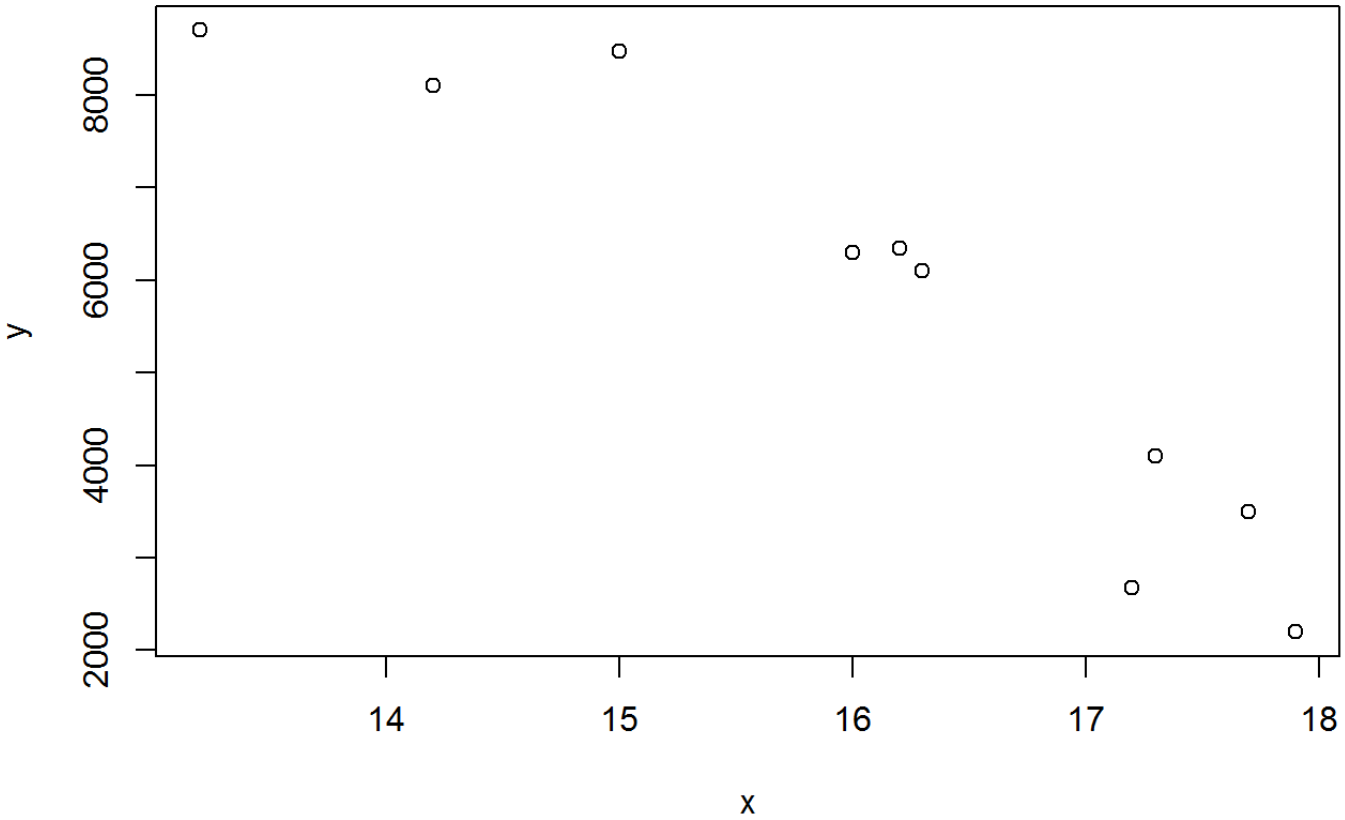
Business Analytics - Assignment 3

Prakkas Thulasithas

29 septembre 2015

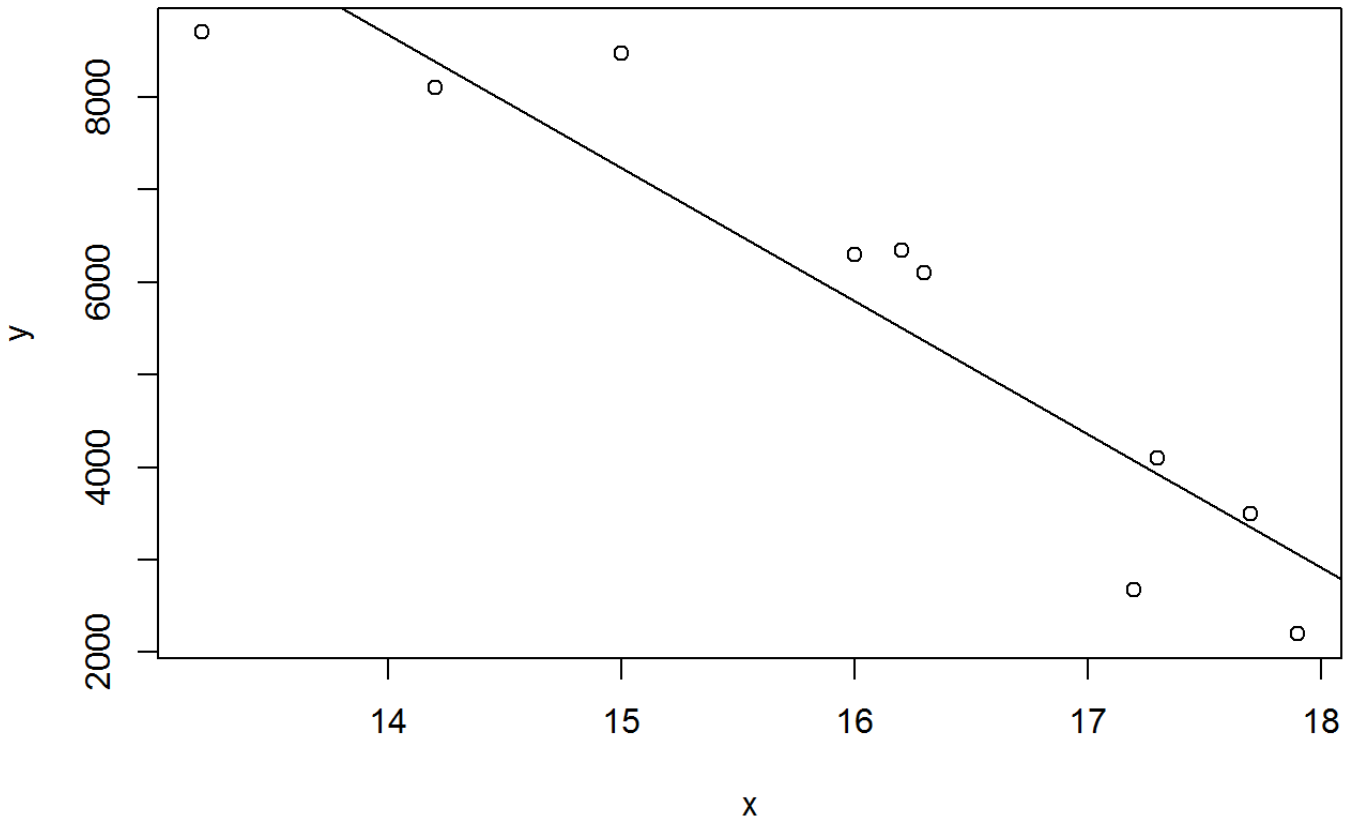
Question1:

```
x <- c(17.9, 16.2, 15.0, 16.0, 17.3, 13.2, 16.3, 17.2, 17.7, 14.2)
y <- c(2200, 6350, 8470, 6300, 4100, 8700, 6100, 2680, 3500, 8100)
```



The heavier is the bicycle, the less expensive it is.

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##       28818        -1439
```



Thus, the regression equation is: price = -1439*(weight) + 28818

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1387.1  -715.9   164.6   679.9  1237.1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28818.0     3267.3   8.820 2.15e-05 ***
## x            -1439.0      202.1  -7.121 9.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 942.3 on 8 degrees of freedom
## Multiple R-squared:  0.8637, Adjusted R-squared:  0.8467
## F-statistic:  50.7 on 1 and 8 DF,  p-value: 9.994e-05
```

hypothesis test for intercept parameter B0 Null hypothesis Ho: Bo =0 Alternative hypothesis HA: B0!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then BO is different than 0.

hypothesis test for parameter B1 Null hypothesis Ho: B1=0 Alternative hypothesis HA: B1!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then B1 is different than 0.

Our p values being less than 0.05 for both coefficients, we can assert that there is a strong relationship between the price and the weight of the bicycle.

The model tells us that for every additional pound, the price decreases by $1439.

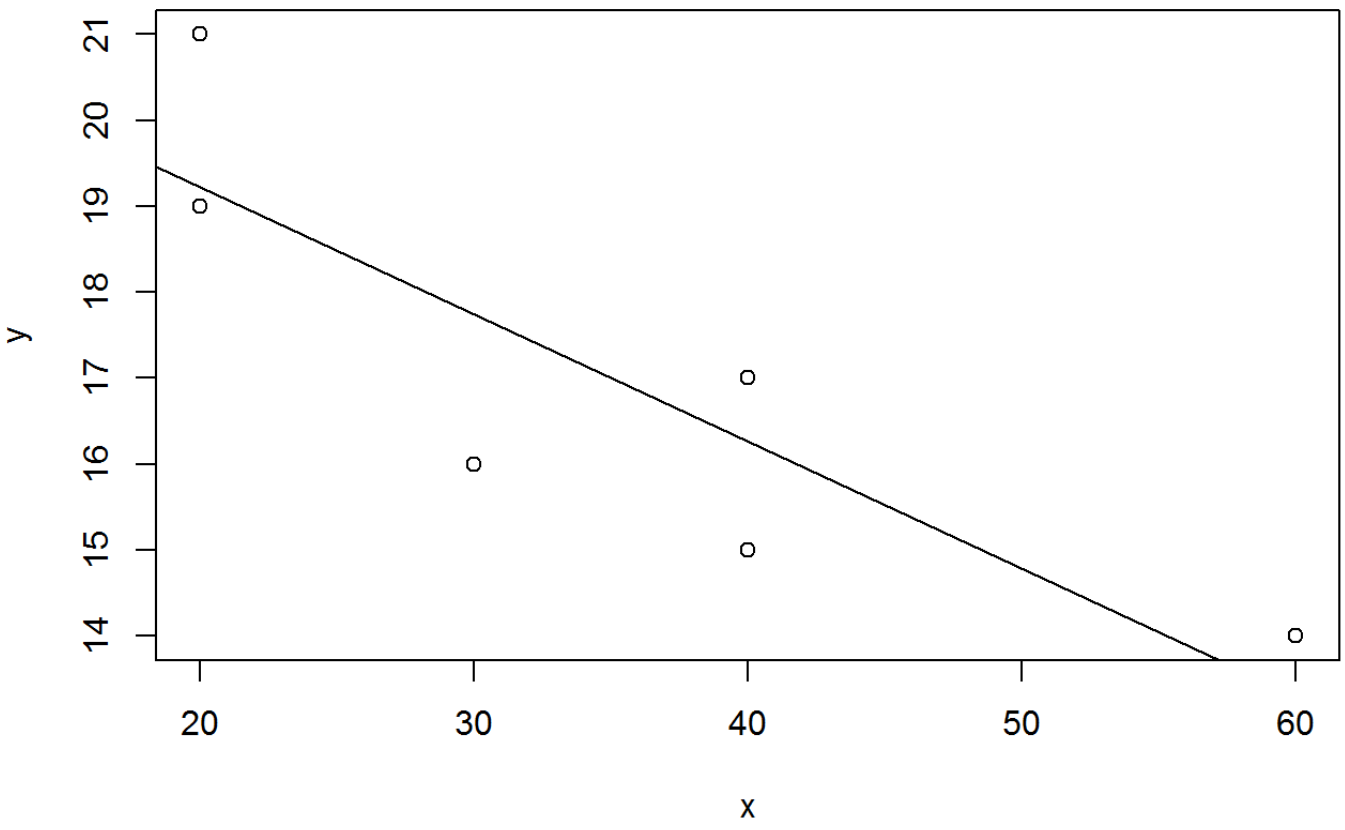
```
predicted_price=-1439*15+28818
predicted_price
```

```
## [1] 7233
```

The predicted price is $7233

Question 2:

```
x<-c(20,20,40,30,60,40)
y<-c(21,19,15,16,14,17)
```




```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     22.1739      -0.1478
```

The faster is the line, the less defects there are in the line.

According to our regression, the model's equation is: defects=-0.1478*(line_speed)+22.1739

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##       1       2       3       4       5       6
##  1.7826 -0.2174 -1.2609 -1.7391  0.6957  0.7391
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.17391    1.65275  13.416 0.000179 ***
## x           -0.14783    0.04391  -3.367 0.028135 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.489 on 4 degrees of freedom
## Multiple R-squared:  0.7391, Adjusted R-squared:  0.6739
## F-statistic: 11.33 on 1 and 4 DF,  p-value: 0.02813
```

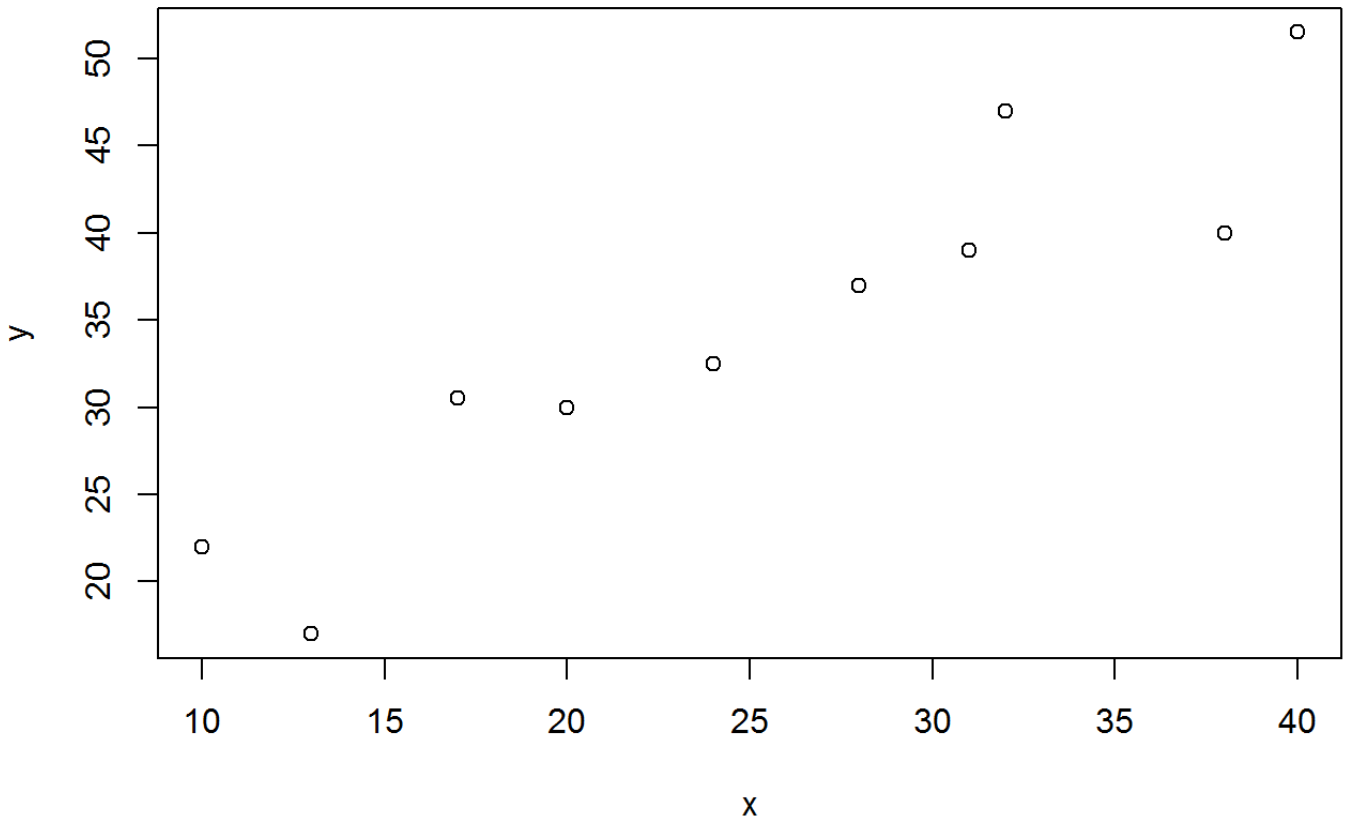
hypothesis test for intercept parameter B0 Null hypothesis Ho: Bo =0 Alternative hypothesis HA: B0!=0 since p<0.01, we can reject the null hypothesis at a 0.01 level of significance then BO is different than 0.

hypothesis test for parameter B1 Null hypothesis Ho: B1=0 Alternative hypothesis HA: B1!=0 since p>0.01, we cannot reject the null hypothesis at a 0.01 level of significance then there is significant evidence that B1=0.

Therefore, it means the line speed does not have any effect on the number of defective parts.

Question 3:

```
x<-c(13,10,20,28,32,17,24,31,40,38)
y<-c(17.0,22,30,37,47,30.5,32.5,39,51.5,40)
plot(x,y)
```



Annual maintenance costs increase with weekly usage.

The linear regression gives us the following equation: annual maintenance expense= 0.9534* (weekly_usage)+10.528

```
summary(model3)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7587 -1.0411  0.0895  2.6102  5.9619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.5280     3.7449   2.811 0.022797 *
## x             0.9534     0.1382   6.901 0.000124 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.25 on 8 degrees of freedom
## Multiple R-squared:  0.8562, Adjusted R-squared:  0.8382
## F-statistic: 47.62 on 1 and 8 DF,  p-value: 0.0001244
```

hypothesis test for intercept parameter B0 Null hypothesis Ho: Bo =0 Alternative hypothesis HA: B0!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then BO is different than 0.

hypothesis test for parameter B1 Null hypothesis Ho: B1=0 Alternative hypothesis HA: B1!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then B1 is different than 0.

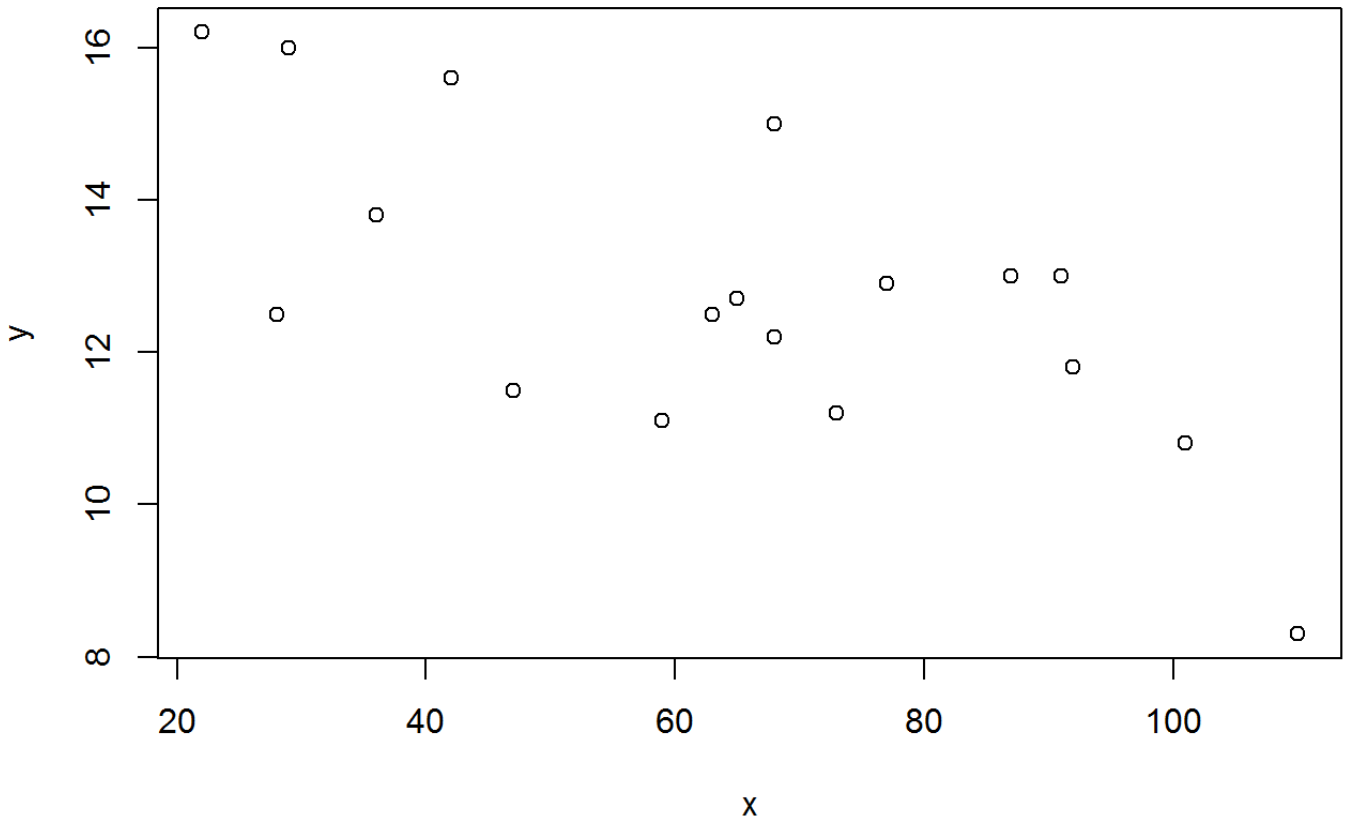
Our p values being less than 0.05 for both coefficients, we can assert that there is a strong relationship between the weekly usage and the annual maintenance expense.

The model tells us that for every additional hour, annual maintenance expenses increase by $953.4

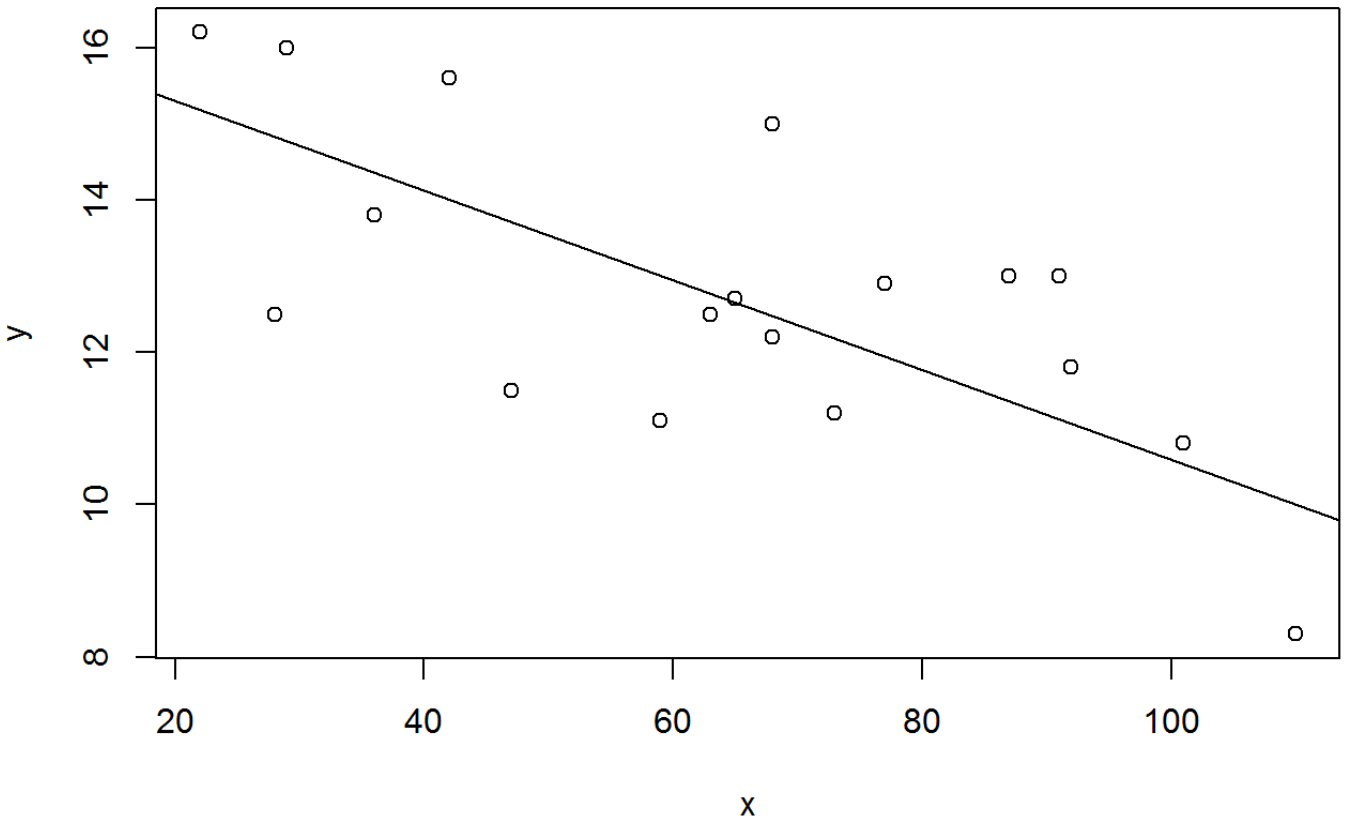
Since we saw that there is a strong relationship between weekly usage and annual expense, I would not recommend buying any maintenance contract which doesn't take into account the weekly usage.

Question 8:

```
x<-c(22,29,36,47,63,77,73,87,92,101,110,28,59,68,68,91,42,65,110)
y<-c(16.2,16,13.8,11.5,12.5,12.9,11.2,13,11.8,10.8,8.3,12.5,11.1,15,12.2,13,15.6,12.7,8.3)
plot(x,y)
```



The scatterplot indicates that the price decreases when miles increase.




```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32408 -1.34194  0.05055  1.12898  2.52687
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.46976    0.94876  17.359 2.99e-12 ***
## x           -0.05877    0.01319  -4.455 0.000348 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.541 on 17 degrees of freedom
## Multiple R-squared:  0.5387, Adjusted R-squared:  0.5115
## F-statistic: 19.85 on 1 and 17 DF,  p-value: 0.0003475
```

The linear regression gives us the following equation: price = -0.05877*(miles)+16.46976

hypothesis test for intercept parameter B0 Null hypothesis Ho: Bo =0 Alternative hypothesis HA: B0!=0 since p<0.01, we can reject the null hypothesis at a 0.01 level of significance then BO is different than 0.

hypothesis test for parameter B1 Null hypothesis Ho: B1=0 Alternative hypothesis HA: B1!=0 since p<0.01, we can reject the null hypothesis at a 0.01 level of significance then B1 is different than 0.

Our p values being less than 0.01 for both coefficients, we can assert that there is a strong relationship between the weekly usage and the annual maintenance expense.

For every addtional 1000 miles, the price decreases by $58.77

```
predicted_price8 <- function(x){
  out<--0.05877*x+16.46976
  return(out)
}
predicted_price8(x)
```

```
##  [1] 15.17682 14.76543 14.35404 13.70757 12.76725 11.94447 12.17955
##  [8] 11.35677 11.06292 10.53399 10.00506 14.82420 13.00233 12.47340
## [15] 12.47340 11.12169 14.00142 12.64971 10.00506
```

```
residual <- predicted_price8(x)-y
residual
```

```
##  [1] -1.02318 -1.23457  0.55404  2.20757  0.26725 -0.95553  0.97955
##  [8] -1.64323 -0.73708 -0.26601  1.70506  2.32420  1.90233 -2.52660
## [15]  0.27340 -1.87831 -1.59858 -0.05029  1.70506
```

```
sort(residual,TRUE)[1:2]
```

```
## [1] 2.32420 2.20757
```

Therefore the auromobiles with the following characteristcs are a good bargain: 47000miles/$11500 and 28000miles/$12500

```
predicted_price8(60)
```

```
## [1] 12.94356
```

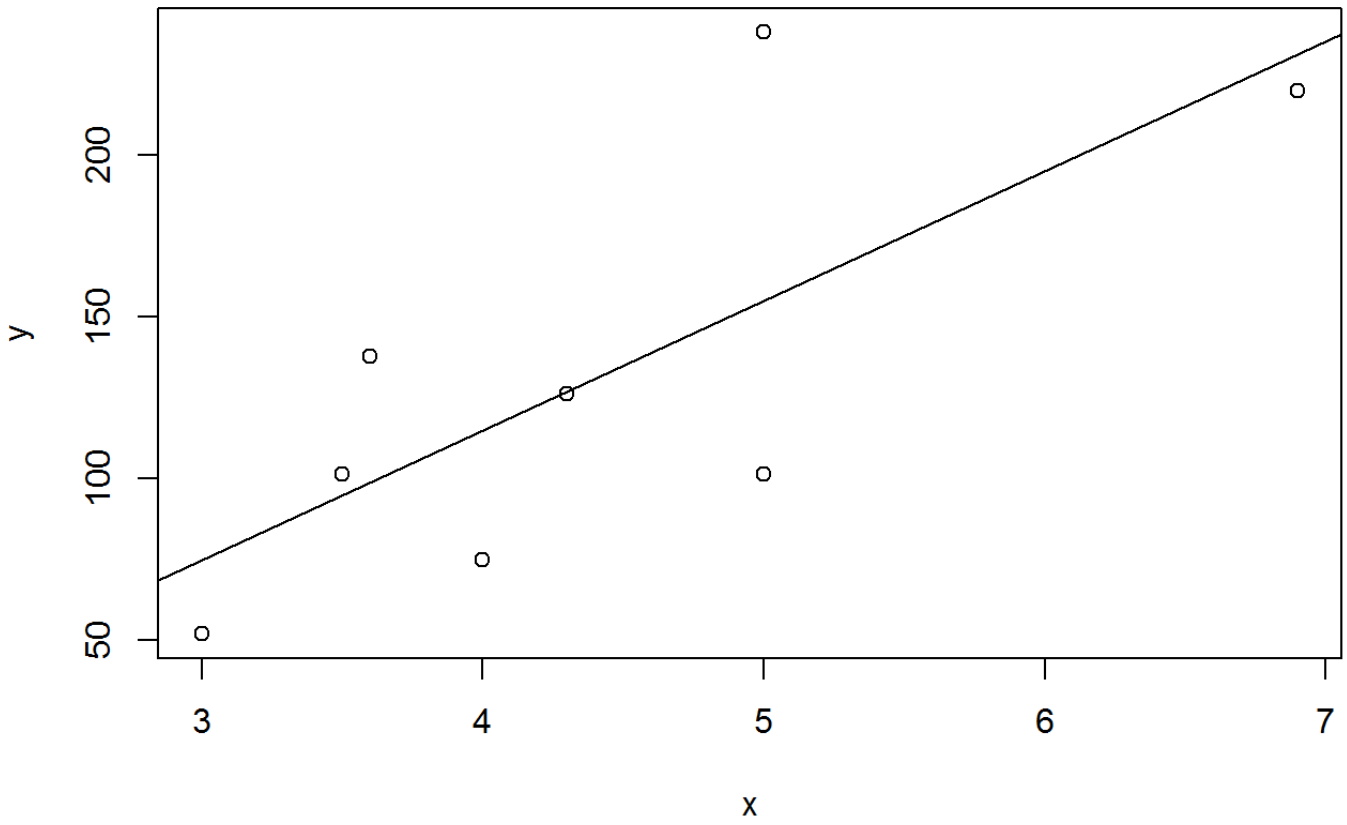
The predicted price by the model is $12943. However, I will always offer less to the seller for the sake of neogtaiating.

Question 9: a.

```
x<-c(5,3,4,4.3,3.6,3.5,5,6.9)
y<-c(101.3,51.9,74.8,126.2,137.8,101.4,237.8,219.6)
plot(x,y)
model9<-lm(y~x)
model9
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##      -45.43        40.06
```

```
abline(model9)
```




```
summary(model9)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.588 -27.151  -6.026  14.707  82.912
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -45.43      66.75  -0.681   0.5215
## x              40.06      14.64   2.737   0.0339 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.55 on 6 degrees of freedom
## Multiple R-squared:  0.5552, Adjusted R-squared:  0.481
## F-statistic: 7.489 on 1 and 6 DF,  p-value: 0.03389
```

hypothesis test for intercept parameter B0 Null hypothesis Ho: Bo =0 Alternative hypothesis HA: B0!=0 since p>0.05, we cannot reject the null hypothesis at a 0.05 level of significance then there is significant evidence that BO=0.

hypothesis test for parameter B1 Null hypothesis Ho: B1=0 Alternative hypothesis HA: B1!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then B1 is different than 0.

Therefore the interpretation of this relationship is that there is not a strong relationship at a 0.05 level of significance between weekly gross revenues and television advertising.

b. For each additional $100 expected in revenue, the cost of TV advertising increase by $40.06

c.

```
z<-c(1.5,3,1.5,4.3,4,2.3,8.4,5.8)
model9c<-lm(y~x+z)
model9c
```

```
##
## Call:
## lm(formula = y ~ x + z)
##
## Coefficients:
## (Intercept)            x            z
##      -42.57        22.40        19.50
```

```
summary(model9c)
```

```
##
## Call:
## lm(formula = y ~ x + z)
##
## Residuals:
##       1       2       3       4       5       6       7       8
##   2.610 -31.233  -1.487 -11.404  21.727  20.715   4.570  -5.498
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -42.570     28.547  -1.491  0.19611
## x             22.402      7.099   3.156  0.02522 *
## z             19.499      3.697   5.274  0.00326 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.33 on 5 degrees of freedom
## Multiple R-squared:  0.9322, Adjusted R-squared:  0.9051
## F-statistic: 34.39 on 2 and 5 DF,  p-value: 0.001196
```

since p=0.001196<0.05, the overall regression is statiscally significant at a 0.05 level of significance.

hypothesis test for intercept parameter B0 Null hypothesis Ho: Bo =0 Alternative hypothesis HA: B0!=0 since p>0.05, we cannot reject the null hypothesis at a 0.05 level of significance then there is significant evidence that BO=0.

hypothesis test for parameter B1 Null hypothesis Ho: B1=0 Alternative hypothesis HA: B1!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then B1 is different than 0.

hypothesis test for parameter B2 Null hypothesis Ho: B2=0 Alternative hypothesis HA: B2!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then B2 is different than 0.

My interpretation is that both TV and newspaper advertising affect the weekly revenue which is obviously reasonable.

d. For every additional $100 in weekly revenue, we need to spend $22.402 in TV and $19.499 in newspaper advertising

e. My next step would be to try to find a linear model between the weekly revenue and newspaper advertising only and then. compare it to the previous ones.

f. The managerial implications is that they cannot cut expenses in the newspaper advertising.

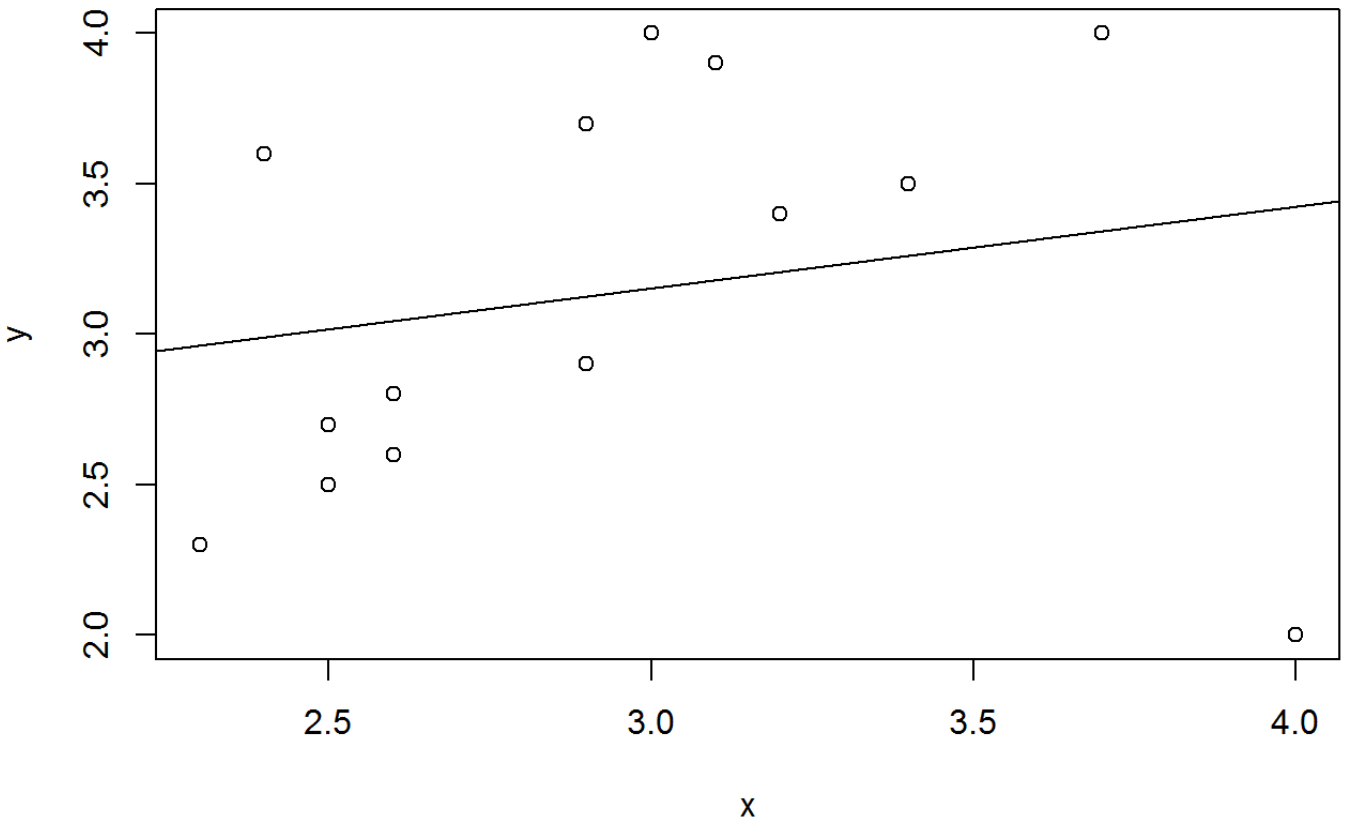
Question 11:

a. let's take x, the trade price, as the independent variable and y, the scores for overall satisfaction, as the dependent variable

```
x<-c(3.4,3.2,3.1,2.9,2.9,2.5,2.6,2.4,2.6,2.3,3.7,2.5,3.0,4.0)
y<-c(3.5,3.4,3.9,3.7,2.9,2.7,2.8,3.6,2.6,2.3,4.0,2.5,4.0,2.0)
plot(x,y)
data11<-lm(y~x)
data11
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##      2.3409       0.2707
```

```
abline(data11)
```

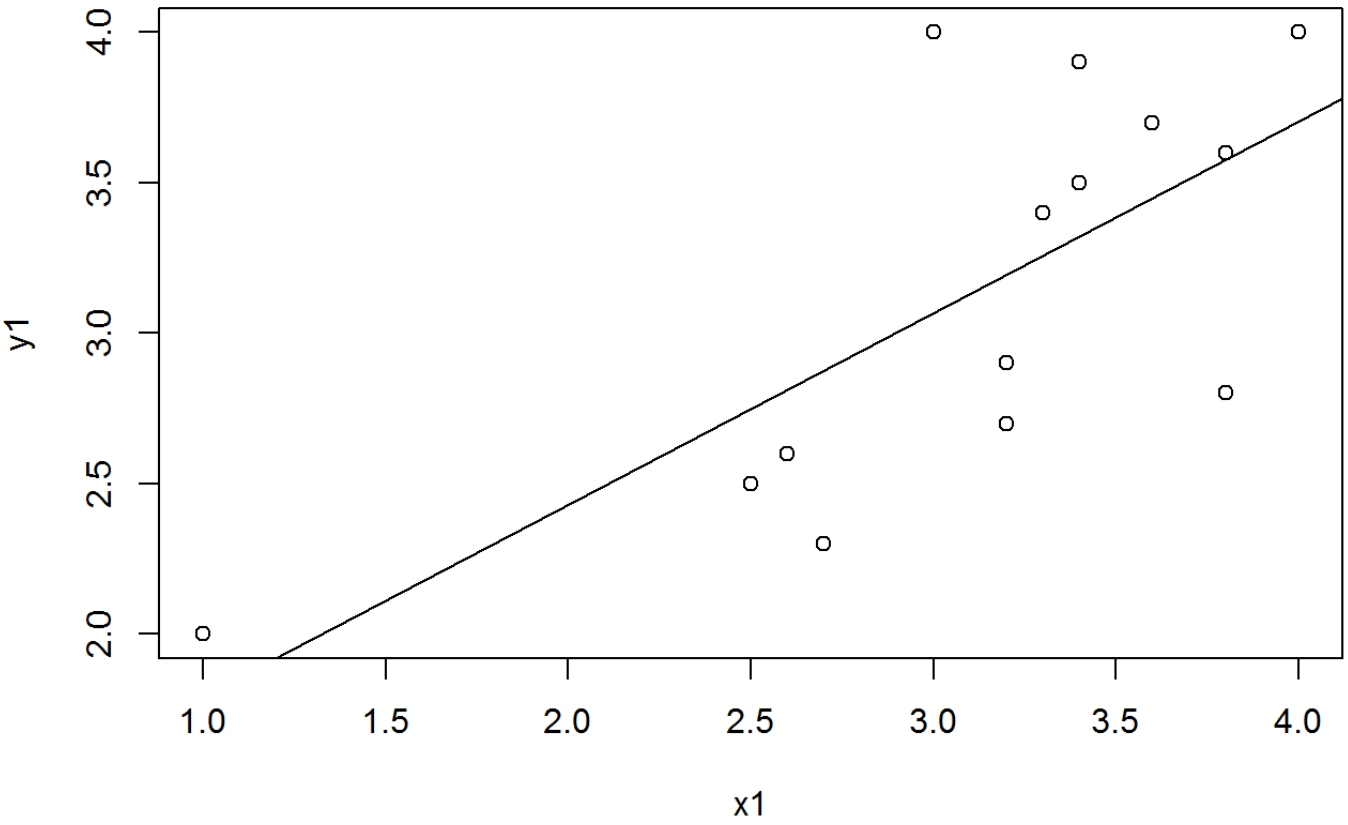


let's take x1, our independent variable as the speed of execution, and y1, the dependent variable as the scores for overall satisfaction.

```
x1<-c(3.4,3.3,3.4,3.6,3.2,3.2,3.8,3.8,2.6,2.7,4.0,2.5,3.0,1.0)
y1<-c(3.5,3.4,3.9,3.7,2.9,2.7,2.8,3.6,2.6,2.3,4.0,2.5,4.0,2.0)
plot(x1,y1)
data11b<-lm(y1~x1)
data11b
```

```
##
## Call:
## lm(formula = y1 ~ x1)
##
## Coefficients:
## (Intercept)           x1
##      1.1571       0.6368
```

```
abline(data11b)
```




```
summary(data11)$r.squared
```

```
## [1] 0.04173612
```

```
summary(data11b)$r.squared
```

```
## [1] 0.5175687
```

a. coefficient of determination: r^2= 0.04 Which is very less than 1 and r squared= 0.51 for data between speed of execution and overall satisfaction

b. Using the F test: We can say that the p-value is much greater than 0.05 when we compare the score of trade price with overall score. Thus we can say that there is no relationship between them but p-value is smaller than 0.05 when we compare the score of speed of execution with overall score. Thus we can say that there is a relationship between them Using T test: for significance of each independent variable. p value for scores for trade price is >0.05 p value for scores for speed of execution is <0.05. Thus we have enough evidence that there is relationship between score of speed of execution with overall score but no realtion between score of trade price with overall score

c. interperting regression parameters: According to me for case A, at 0.05 level of significance,B0 and B1 is equal to 0 because for we cannot reject the null hypothesis for case B, at 0.05 level of significance, B0 and B1 is not equal to 0 yes the realtions are what I would expect.

d. equation for score of trade price and overall electronic trades: y = 0.2707 (x) + 2.3409, if x=3 then y= 3.153. Thus overall satisfaction level = 3.15 y = 0.2707*3 + 2.3409 equation for score of speed of execution and overall electronic trades score: y = 0.6329 (x) + 1.1621, if x=3 then y= 3.06. Thus overall satisfaction level = 3.06 y = 0.6329*3 + 1.1621*

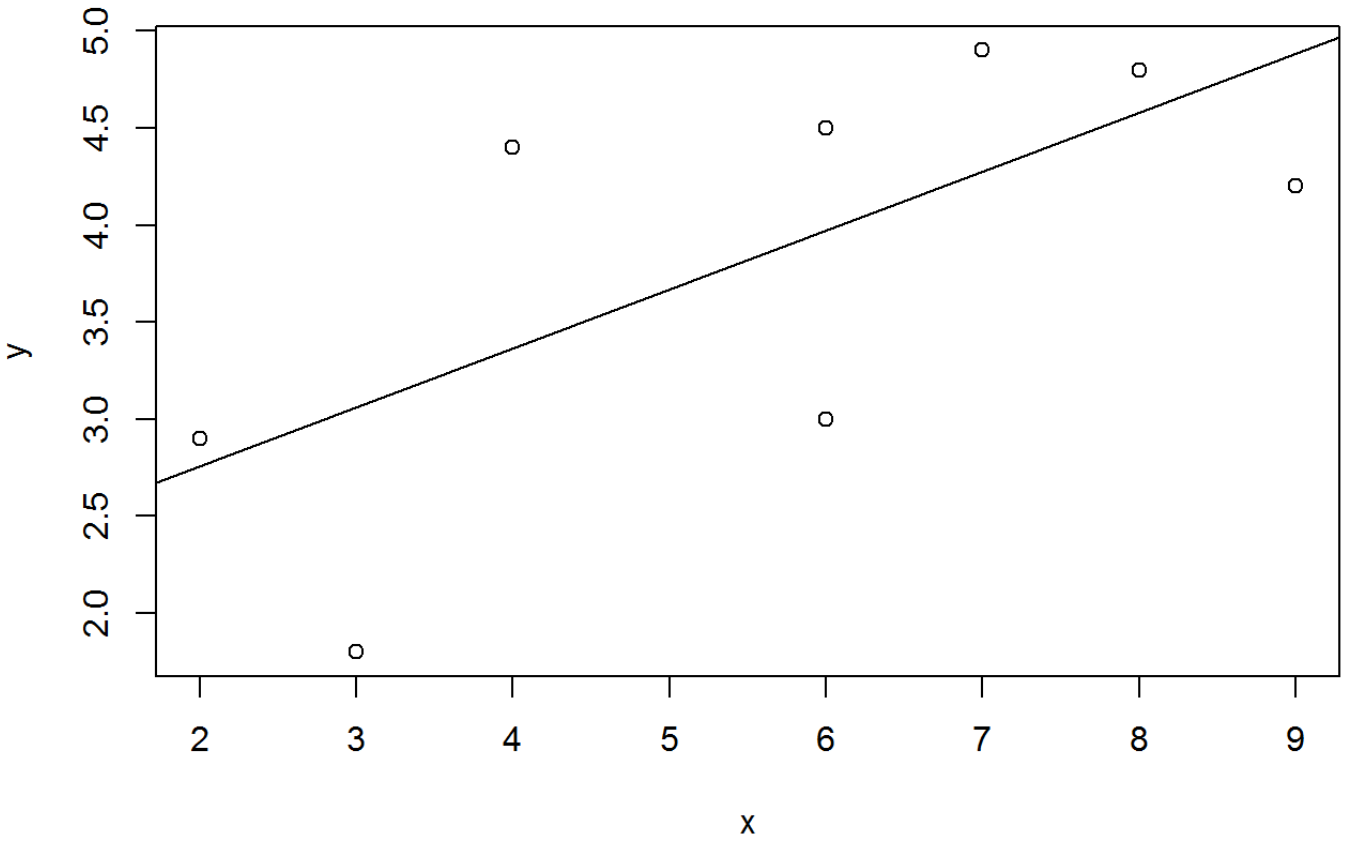
e. The possible response the respondents can select on the survey would be more scores for satisfaction with speed execution than for satisfaction with trade price.

Question 13 a.

```
x<-c(2,6,8,3,2,7,9,8,4,6)
y<-c(2.9,3.0,4.8,1.8,2.9,4.9,4.2,4.8,4.4,4.5)
plot(x,y)
model13<-lm(y ~ x)
summary(model13)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2597 -0.4772  0.1821  0.4509  1.0362
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1473     0.6050   3.549  0.00752 **
## x             0.3041     0.1004   3.029  0.01634 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.781 on 8 degrees of freedom
## Multiple R-squared:  0.5342, Adjusted R-squared:  0.4759
## F-statistic: 9.174 on 1 and 8 DF,  p-value: 0.01634
```

```
abline(model13)
```



Hypothesis = there is no relationship between repair time and the number of months since the last maintenance which means B0 and B1=0

hypothesis test for intercept parameter B0 Null hypothesis Ho: Bo =0 Alternative hypothesis HA: B0!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then BO is different than 0.

hypothesis test for parameter B1 Null hypothesis Ho: B1=0 Alternative hypothesis HA: B1!=0 since p<0.05, we can reject the null hypothesis at a 0.05 level of significance then B1 is different than 0.

Therefore we can reject our initial hypothesis.

```
summary(model13)$r.squared
```

```
## [1] 0.5341765
```

R^2 is equal to 0.534 so our model describes accurately 53% of the data.

b.

```
predicted_price13 <- function(x){
  out<-0.3041*x+2.1473
  return(out)
}
predicted_price13(x)
```

```
##  [1] 2.7555 3.9719 4.5801 3.0596 2.7555 4.2760 4.8842 4.5801 3.3637 3.9719
```

```
residual13 <- predicted_price13(x)-y
residual13
```

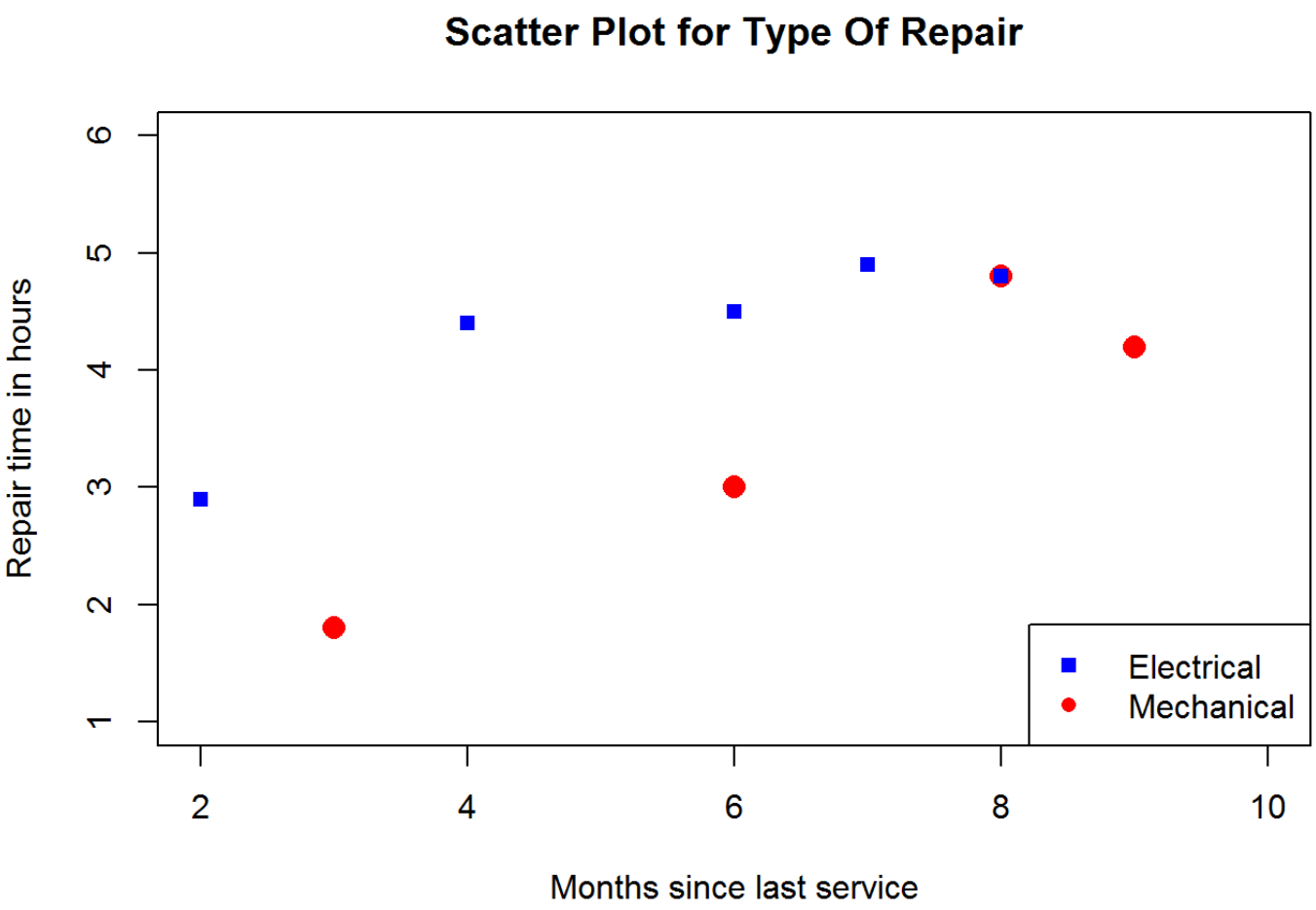
```
##  [1] -0.1445  0.9719 -0.2199  1.2596 -0.1445 -0.6240  0.6842 -0.2199
##  [9] -1.0363 -0.5281
```

```
sort(residual13,TRUE)[1:10]
```

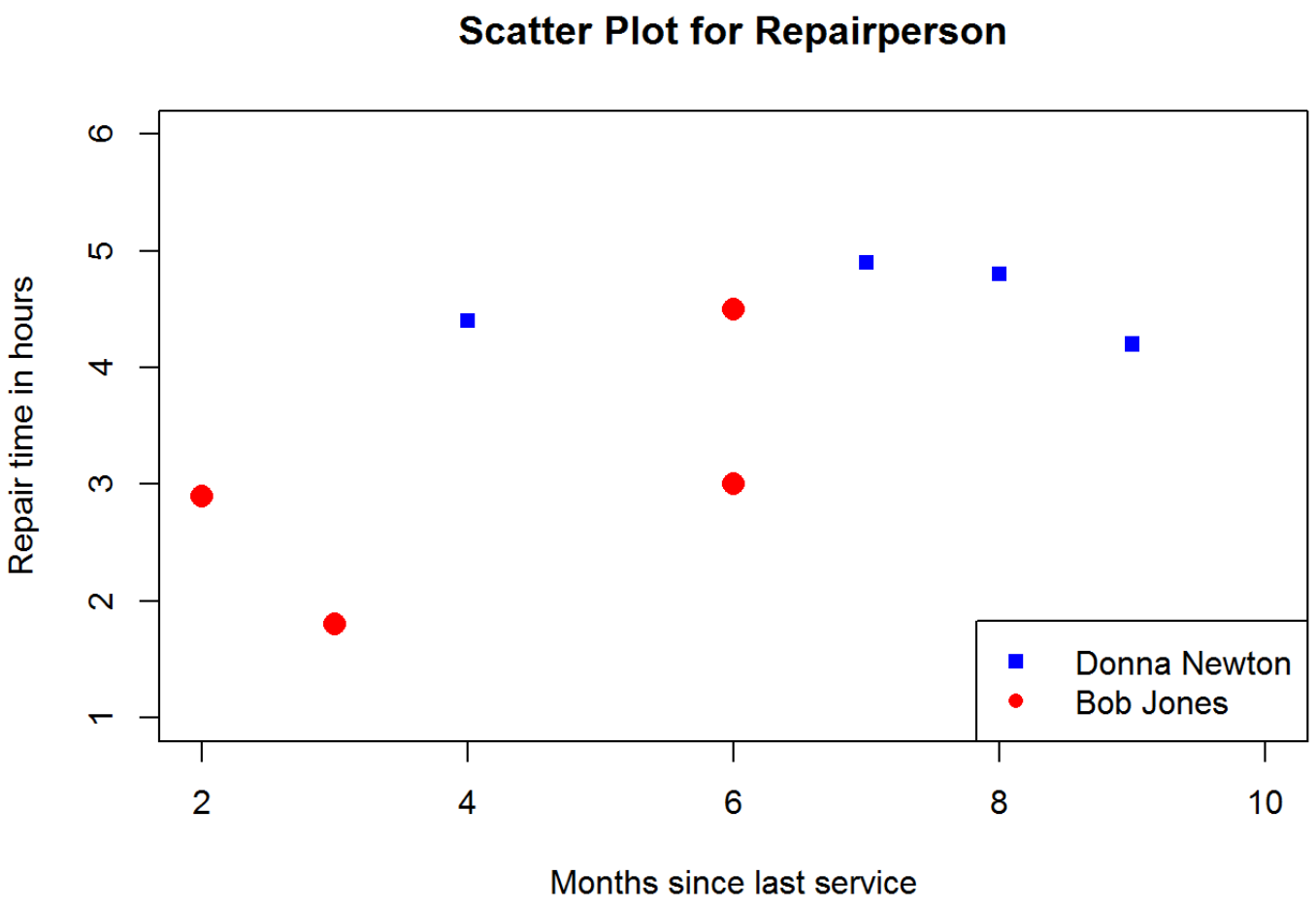
```
##  [1]  1.2596  0.9719  0.6842 -0.1445 -0.1445 -0.2199 -0.2199 -0.5281
##  [9] -0.6240 -1.0363
```

We can see that mechanical repairs are the less accurately predicted with our model

```
x.electrical<-c(2,8,7,4,6)
y.electrical<-c(2.9,4.8,4.9,4.4,4.5)
x.mechanical<-c(6,3,9,8)
y.mechanical<-c(3,1.8,4.2,4.8)
plot(x.mechanical,y.mechanical,cex=1.5,pch=16,col="red",ylim=c(1,6),xlim=c(2,10),xlab="Months si
nce last service",ylab="Repair time in hours")
title(main="Scatter Plot for Type Of Repair")
points(x.electrical,y.electrical,cex=1,pch=15,col="blue")
legend( x="bottomright", legend=c("Electrical", "Mechanical"), col=c("blue","red"),lwd="1", lt
y=c(0,0), pch=c(15,16) )
```



```
x.donna<-c(2,6,3,2,6)
y.donna<-c(2.9,3,1.8,2.9,4.5)
x.bob<-c(8,7,9,8,4)
y.bob<-c(4.8,4.9,4.2,4.8,4.4)
plot(x.donna,y.donna,cex=1.5,pch=16,col="red",ylim=c(1,6),xlim=c(2,10),xlab="Months since last s
ervice",ylab="Repair time in hours")
title(main="Scatter Plot for Repairperson")
points(x.bob,y.bob,cex=1,pch=15,col="blue")
legend( x="bottomright", legend=c("Donna Newton", "Bob Jones"), col=c("blue","red"),lwd="1", lt
y=c(0,0), pch=c(15,16) )
```



Those charts shows us different results than our previous linear regression. In the first chart, we can clearly see that their is a relationship between the type of repair and the repair time, which means we have to create 2 different linear model according to the type of repair. However, in the second chart, we can see that the repairs done by Bob have all almost the same repair time (4.5 hour in average) whereas the ones done by Donna have a more disparate repair time.