Lab-03

Linear Regression

Part: 1

First we need to input the dataset. Here we create two vectors with both sets of observations

```
distance <-c(3.4, 1.8, 4.6, 2.3, 3.1, 5.5, .7, 3.0, 2.6, 4.3, 2.1, 1.1, 6.1, 4.8, 3.8)

damage <- c(26.1, 17.8, 31.3, 23.1, 27.5, 36.0, 14.1, 22.3, 19.6, 31.3, 24.0, 17.3, 43.2, 36.4, 26.1)
```

Then we put both into a dataframe to show the relationship

```
fd <- data.frame(distance, damage)
```

Now we can build the model - and display details (gradient & intercept)

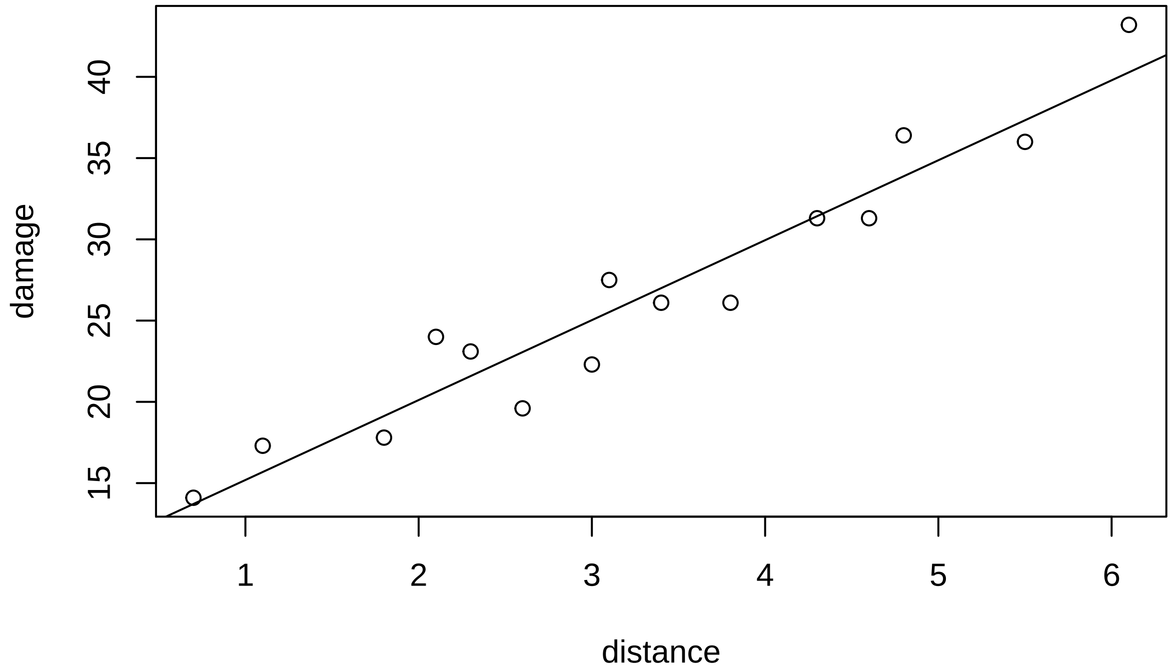
```
fit <- lm(damage ~ distance)

fit
```

```
##
## Call:
## lm(formula = damage ~ distance)
##
## Coefficients:
## (Intercept)     distance
##      10.272        4.919
```

Part: 2

We can now plot the two vectors and use the regression line from the model to show their relationship

```
plot(distance, damage)
abline(fit)
```



Part: 3

Part: 4

To check how well the model has fitted to the data, we can use the summary command

```
summary(fit)
```

```
##
## Call:
## lm(formula = damage ~ distance)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.462 -1.463 -0.124  1.798  3.398
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.2724     1.4220   7.224 6.71e-06 ***
## distance      4.9190     0.3932  12.509 1.27e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.319 on 13 degrees of freedom
## Multiple R-squared:  0.9233, Adjusted R-squared:  0.9174
## F-statistic: 156.5 on 1 and 13 DF,  p-value: 1.267e-08
```

This shows that the RSE is 2.316. This indicates that prediction for fire damage would on average be 2.316 off when using the least squares line.

The R-squared is 0.9176. This indictates that 91.76% of the sample varaiation can be explained by distance.

Part: 5

From the model summary we can see that the t-value is 12.525, and the p-value is 1.248e-08. As the t-value is large and the p-value is small, we can conclude that there is a relationship between fire damage and distance to the nearest fire station.

Part: 6

We can use the predict function to predict value for new test data

```
newData <- c(0.5, 1.5, 2.5, 3.5)

predict(fit,data.frame(distance=newData))
```

```
##        1        2        3        4
## 12.73189 17.65087 22.56986 27.48884
```

Part: 7

We can use the confit function to check confidence level parameters

```
confint(fit, level=0.95)
```

```
##                2.5 %    97.5 %
## (Intercept) 7.200332 13.344455
## distance    4.069472  5.768499
```

Part: 8

We can use the predict function, with the additional parametr interval to get predictions with confidence values

```
predict(fit, data.frame(distance=c(0.5, 3, 5)), interval = "confidence")
```

```
##        fit      lwr      upr
## 1 12.73189 10.03914 15.42463
## 2 25.02935 23.71402 26.34468
## 3 34.86732 32.91578 36.81886
```

Part: 9

We can use the same process as in part 8 for prediction intervals

```
predict(fit,data.frame(distance=(c(0.5, 3, 5))), interval="prediction")
```

```
##        fit       lwr      upr
## 1 12.73189  7.043865 18.41991
## 2 25.02935 19.849314 30.20939
## 3 34.86732 29.490409 40.24424
```

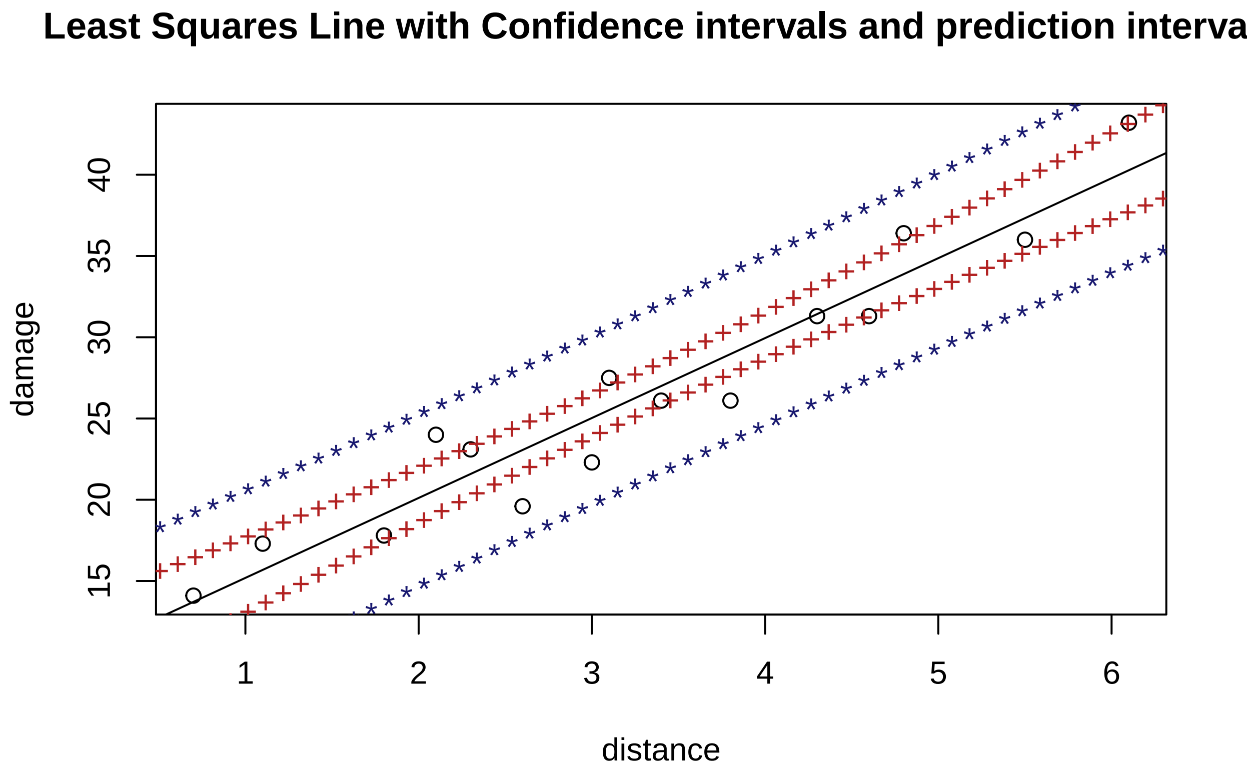
Part: 10

From the results in parts 8 & 9, we can see that prediction intervals are always larger than confidence intervals

Part: 11

To plot bothconfidence intervals & prediction intervals on the same graph, we can use the following commands

```
plot(distance, damage, xlab="distance", ylab = "damage", main = "Least Squares Line with Confidence inte
abline(fit)
plotLine <- data.frame(distance=seq(0,6.5,length=65)) # dummy plotline to show intervals
predict_c = predict(fit,plotLine,interval="confidence")
predict_p = predict(fit,plotLine,interval="prediction")
lines(plotLine$distance,predict_c[,"lwr"],col="firebrick", type="b",pch="+")
lines(plotLine$distance,predict_c[,"upr"],col="firebrick", type="b",pch="+")
lines(plotLine$distance,predict_p[,"upr"],col="midnightblue", type="b",pch="*")
lines(plotLine$distance,predict_p[,"lwr"],col="midnightblue",type="b",pch="*")
```


Least Squares Line with Confidence intervals and prediction interva
damage
distance