# Statistical Modeling I
# Practical in R – Output

## Practical in R – Output

In this practical, we will work with the Fire dataset. We will look at the predictive values and the creation of confidence and predictive intervals.

A fire insurance company wants to relate the amount of fire damage in major residential fires to the distance between the residence and the nearest fire station. The study was conducted in a large suburb of a major city; a sample of fifteen recent fires in the suburb was selected. The amount of damage, $y$ (£000), and the distance, $x$ (km), between the fire and the nearest fire station are given in the following table.

| $x$ | 3.4 | 1.8 | 4.6 | 2.3 | 3.1 | 5.5 | 0.7 | 3.0 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 26.2 | 17.8 | 31.3 | 23.1 | 27.5 | 36.0 | 14.1 | 22.3 |
| $x$ | 2.6 | 4.3 | 2.1 | 1.1 | 6.1 | 4.8 | 3.8 | |
| $y$ | 19.6 | 31.3 | 24.0 | 17.3 | 43.2 | 36.4 | 26.1 | |

We define the values of x and y

```
> x<- c(3.4,1.8,4.6,2.3,3.1,5.5,0.7,3.0,2.6,4.3,2.1,1.1,6.1,4.8,
                      3.8)
> y<- c(26.2,17.8,31.3,23.1,27.5,36.0,14.1,22.3,19.6,31.3,24.0,
                   17.3,43.2,36.4,26.1)
```

Once we import the two variables in R, we can proceed with the Practical

1. We plot the data and then we fit the linear regression model between $x$ and $y$.

```
> plot(x,y, main="Plot of Y versus X")
> fire<-lm(y~x)
> abline(fire)
> summary(fire)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4682 -1.4705 -0.1311  1.7915  3.3915
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2779     1.4203   7.237 6.59e-06 ***
x             4.9193     0.3927  12.525 1.25e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.316 on 13 degrees of freedom
Multiple R-squared:  0.9235,Adjusted R-squared:  0.9176
F-statistic: 156.9 on 1 and 13 DF,  p-value: 1.248e-08
```

In Figure 1.1, we have the fitted line plot, where the parameters of the intercept and of the slope are positive and highly significant. The $R^2$ is around $93\%$ and there is a positive linear relation between $x$ and $y$.

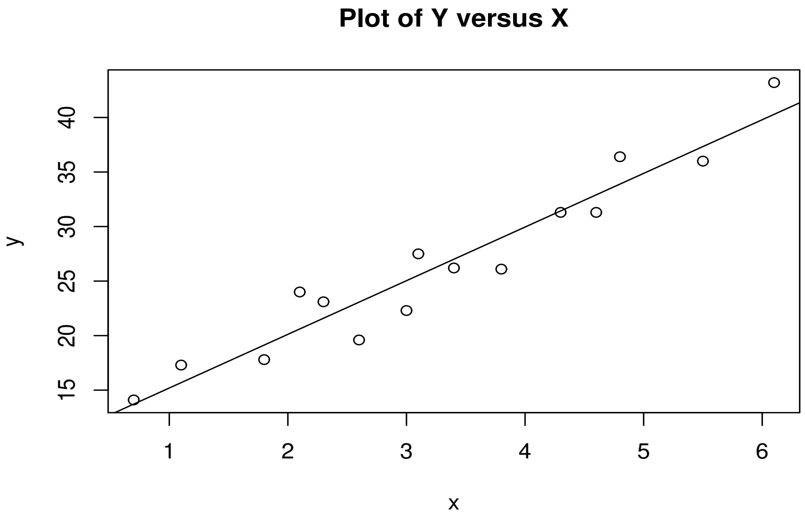

Figure 1.1: Plot of the fitted regression line.

2. We set the values at which the confidence and prediction intervals are to be calculated from the smallest x to the largest. We call them newx and one hundred values should be enough.

```
> newx <- seq(min(x), max(x), length.out=100)
> print(newx)
  [1] 0.7000000 0.7545455 0.8090909 0.8636364 0.9181818 0.9727273 1.0272727 1.0818182 1.1363636 1.1909
 [11] 1.2454545 1.3000000 1.3545455 1.4090909 1.4636364 1.5181818 1.5727273 1.6272727 1.6818182 1.7363
 [21] 1.7909091 1.8454545 1.9000000 1.9545455 2.0090909 2.0636364 2.1181818 2.1727273 2.2272727 2.2818
 [31] 2.3363636 2.3909091 2.4454545 2.5000000 2.5545455 2.6090909 2.6636364 2.7181818 2.7727273 2.8272
 [41] 2.8818182 2.9363636 2.9909091 3.0454545 3.1000000 3.1545455 3.2090909 3.2636364 3.3181818 3.3727
 [51] 3.4272727 3.4818182 3.5363636 3.5909091 3.6454545 3.7000000 3.7545455 3.8090909 3.8636364 3.9181
 [61] 3.9727273 4.0272727 4.0818182 4.1363636 4.1909091 4.2454545 4.3000000 4.3545455 4.4090909 4.4636
 [71] 4.5181818 4.5727273 4.6272727 4.6818182 4.7363636 4.7909091 4.8454545 4.9000000 4.9545455 5.0090
 [81] 5.0636364 5.1181818 5.1727273 5.2272727 5.2818182 5.3363636 5.3909091 5.4454545 5.5000000 5.5545
 [91] 5.6090909 5.6636364 5.7181818 5.7727273 5.8272727 5.8818182 5.9363636 5.9909091 6.0454545 6.1000
```

3. We compute the new predictive values of $y$ based on the linear regression model and the relative confidence intervals at all the values of newx by using the command predict

```
 > preds <- predict(fire, newdata = data.frame(x=newx),
+                  interval = ’confidence’)
```

Then we show the initial elements of preds

```
 > head(preds)
       fit      lwr      upr
1 13.72146 11.17951 16.26341
2 13.98979 11.48758 16.49199
3 14.25811 11.79543 16.72080
4 14.52644 12.10303 16.94985
5 14.79477 12.41039 17.17915
6 15.06310 12.71748 17.40872
```

As one can see, the first column is the predict value of $y$ for the new values of $x$, while the second column refers to the lower bound confidence interval and the third the upper bound of the confidence interval. If nothing is add, the confidence interval is by default fixed at $95\%$, otherwise one can decide the confidence interval, by using the command level as stated below for the $90\%$ confidence interval:

```
> preds90 <- predict(fire, newdata = data.frame(x=newx),
+                  interval = ’confidence’,level =0.90)
> head(preds90)
       fit      lwr      upr
1 13.72146 11.63773 15.80519
2 13.98979 11.93864 16.04094
3 14.25811 12.23936 16.27687
4 14.52644 12.53988 16.51300
5 14.79477 12.84020 16.74933
6 15.06310 13.14031 16.98589
```

As one can see the predicted values are not changing since we are using the same regression model, but the lower and upper bound of the confidence intervals are changing.

4. In this part, similarly we calculate the new predictive values and the relative prediction interval by using the following command

```
> preds1 <- predict(fire, newdata = data.frame(x=newx),
+                  interval = ’prediction’)
> head(preds1)
       fit      lwr      upr
1 13.72146 8.108694 19.33423
2 13.98979 8.394909 19.58466
```

```
3 14.25811 8.680798 19.83543
4 14.52644 8.966357 20.08653
5 14.79477 9.251582 20.33795
6 15.06310 9.536472 20.58972
```

As stated above, the first column is the predicted value (which is not changing since we are using the same x and the same regression model), while the second and third column refer to lower and upper bound of the predictive interval at $95\%$. Obviously as stated above, we can decide to have a different level of predictive interval, for example at $90\%$ and the results will change for the second and third column

```
> preds190 <- predict(fire, newdata = data.frame(x=newx),
+                     interval = 'prediction',level =0.90)
> head(preds190)
       fit       lwr      upr
1 13.72146  9.120470 18.32245
2 13.98979  9.403461 18.57611
3 14.25811  9.686184 18.83005
4 14.52644  9.968636 19.08425
5 14.79477 10.250816 19.33872
6 15.06310 10.532720 19.59347
```

5. We are refering to preds[,2] and preds[,3] for the lower and upper bound of the confidence interval, on the other hand, if we are working with the predictive interval, we should use preds1[,2] and preds1[,3] for the lower and upper bound of the $95\%$ predictive interval.

6. In conclusion we add both the $95\%$ intervals to the plot by using the following commands:

```
> plot(x,y, main="Plot of Y versus X")
> abline(fire)
> lines(newx, preds[ ,3], lty = 'dashed', col = 'blue')
> lines(newx, preds[ ,2], lty = 'dashed', col = 'blue')
> lines(newx, preds1[ ,3], lty = 'dashed', col = 'red')
> lines(newx, preds1[ ,2], lty = 'dashed', col = 'red')
```

Figure 1.2 shows the data, in black the fitted regression line. Moreover, we include in blue the $95\%$ confidence intervals, while in red, we have included the $95\%$ predictive intervals.

We can decide to include in the plot also the $90\%$ confidence intervals that we have computed earlier

```
> plot(x,y, main="Plot of Y versus X")
> abline(fire)
> lines(newx, preds[ ,3], lty = 'dashed', col = 'blue')
```

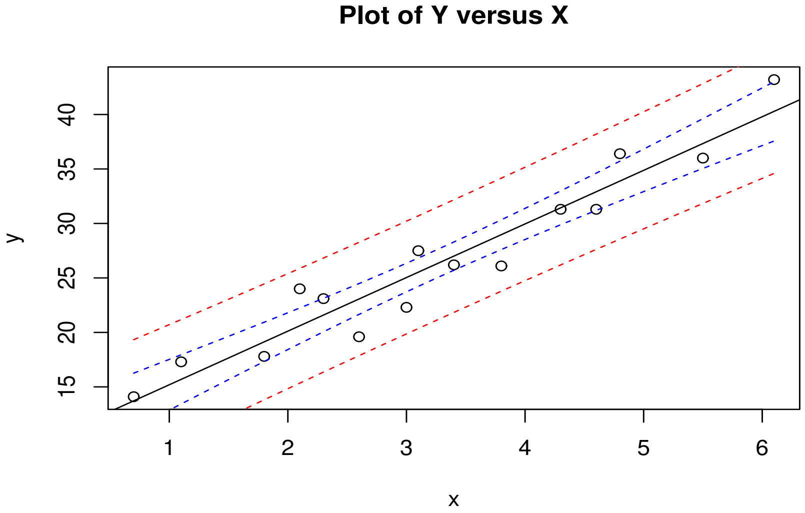

Figure 1.2: Fire data, plot of fitted data (black line) and confidence (blue dashed) and prediction intervals (red dashed).

```
> lines(newx, preds[ ,2], lty = 'dashed', col = 'blue')
> lines(newx, preds90[ ,3], lty = 'dashed', col = 'red')
> lines(newx, preds90[ ,2], lty = 'dashed', col = 'red')
```

The left panel of Figure 1.3 show the 95% confidence interval in blue and the 90% confidence interval in red. On the other hand, the right panel of Figure 1.3 shows the 95% prediction interval in blue and the 90% prediction interval in red.

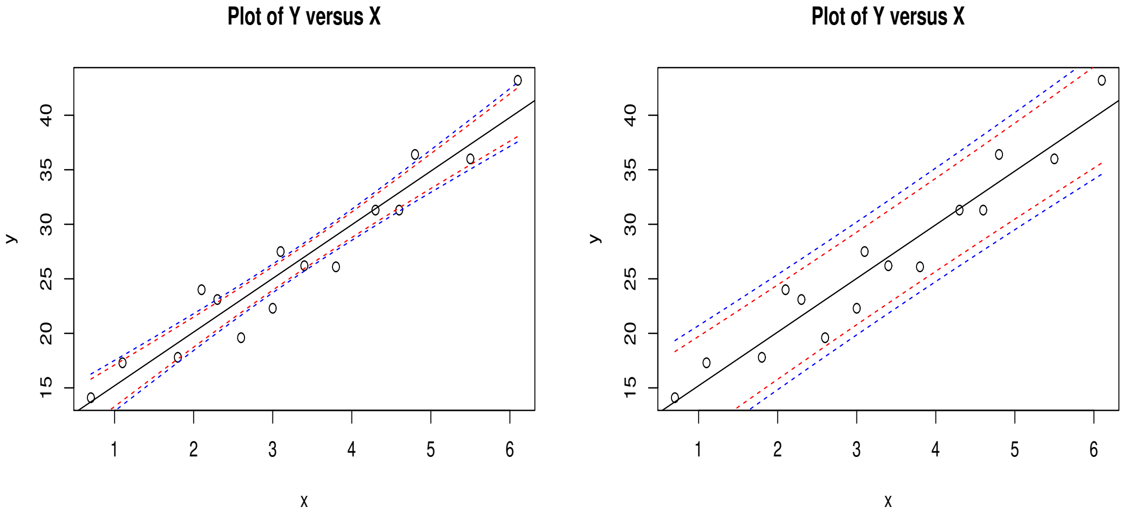

Figure 1.3: Left Panel: Fire data, plot of fitted data (black line) and 95% confidence (blue dashed) and 90% confidence intervals (red dashed). Right Panel: Fire data, plot of fitted data (black line) and 95% prediction (blue dashed) and 90% prediction intervals (red dashed).