Lab3 Answer

Anyi Guo

16/10/2018

Lab 3 Distance vs. Fire Damage

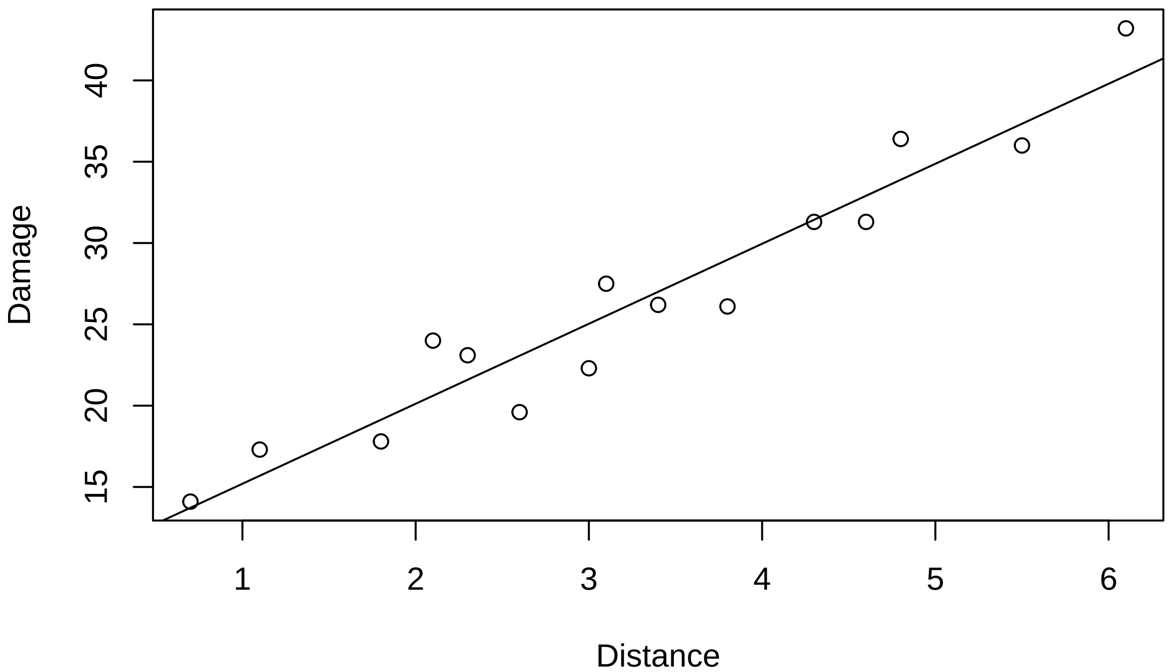
1) Use R to estimate the unknown parameters of the hypothesised model. Find the least squares estimates of the slope `b1` and intercept `b0` on the printout. Write down the least squares equation for this model.

```
library(ISLR)
library(MASS)
Distance<-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)
Damage<-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)
FireData<-data.frame(Distance,Damage)
fire_fit<-lm(Damage~Distance,data=FireData)
```

- Least squares estimates of the slope `b1` is:`4.919`
- Intercept `b0` is: `10.278`
- Least squares equation is: `Damage = 10.278 + 4.919 * Distance`

2) Plot the dataset and the least squares line on the same figure.

```
plot(Distance,Damage)+abline(fire_fit)
```




```
## integer(0)
```

3) How to interpret the slope `b1` and y-intercept `b0` of the least squares line?

- `b1`: For every increase of 1 mile, the fire damage would cost $4919.3 more.
- `b0`: Even if the fire takes place exactly where the fire station is (e.g. Distance=0), it would still cost $10,277.9 to put off the fire.

4) Measuring the extent to which the model fits the data.

```r
summary(fire_fit)
```

```
##
## Call:
## lm(formula = Damage ~ Distance, data = FireData)
##
## 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 ***
## Distance      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
```

a. What is the RSE of this model? How to interpret it?

- `RSE` is 2.316. It means that even if this is correct, any prediction on cost of damage on the basis of distance would still be off by $2316 on average

b. What is the R-squared value? How to interpret it?

- `R-squared` is 0.9176 (Using Adjusted R-Squared). It means that 91.76% of the variation of Damage Cost can be explained by changes in Distance.

5) Test the null hypothesis that `b1` is 0. Find the t-value and p-value. What conclusions can you draw from the two values?

- `t-value` is 12.525
- `p-value` is 1.248e-08.
- The low p-value($<$ 5%) indicates that there is a relationship between Distance and Damage Cost, so we can reject the null hypothesis.

6) Predict the value of damage (y) for a new set of distances (x = 0.5, 1.5, 2.5, 3.5). Can you obtain the above new y's simultaneously?

```r
predict(fire_fit,data.frame(Distance=c(0.5,1.5,2.5,3.5)))
```

```
##        1        2        3        4
## 12.73759 17.65692 22.57626 27.49559
```

7) Find the confidence interval for the prediction of damage (y) for a new set of distance (x = 0.5, 3, 5.5). Can you obtain the above confidence intervals simultaneously?

```r
predict(fire_fit,data.frame(Distance=c(0.5,3,5.5)),interval="confidence")
```

```
##        fit      lwr      upr
## 1 12.73759 10.04812 15.42707
## 2 25.03592 23.72219 26.34965
## 3 37.33425 35.05007 39.61843
```

8) Find the prediction interval for the prediction of damage (y) for a new set of distance (x = 0.5, 3, 5.5). Can you obtain the above confidence intervals simultaneously?

```r
predict(fire_fit,data.frame(Distance=c(0.5,3,5.5)),interval="prediction")
```

```
##        fit       lwr      upr
## 1 12.73759  7.056494 18.41869
## 2 25.03592 19.862187 30.20965
## 3 37.33425 31.833418 42.83508
```

9) Compare your results in 7) and 8) and comment on the confidence intervals and prediction intervals.

- Althought the fit figures (first column) is the same for both Prediction and Confidence intervals, Prediction intervals have wider lower and upper boundary than Confidence intervals.

10) Plot the confidence intervals and prediction intervals in the same figure.

```r
More_Distance<-data.frame(Distance=seq(0,7,length=51))
p_conf<-predict(fire_fit,interval="confidence",newdata=More_Distance)
p_pred<-predict(fire_fit,interval="prediction",newdata=More_Distance)
p<-plot(Distance,Damage,data=FireData)+ abline(fire_fit)+lines(More_Distance$Distance,p_conf[,"lwr"],col
```

```
## Warning in plot.window(...): "data" n'est pas un paramètre graphique
```

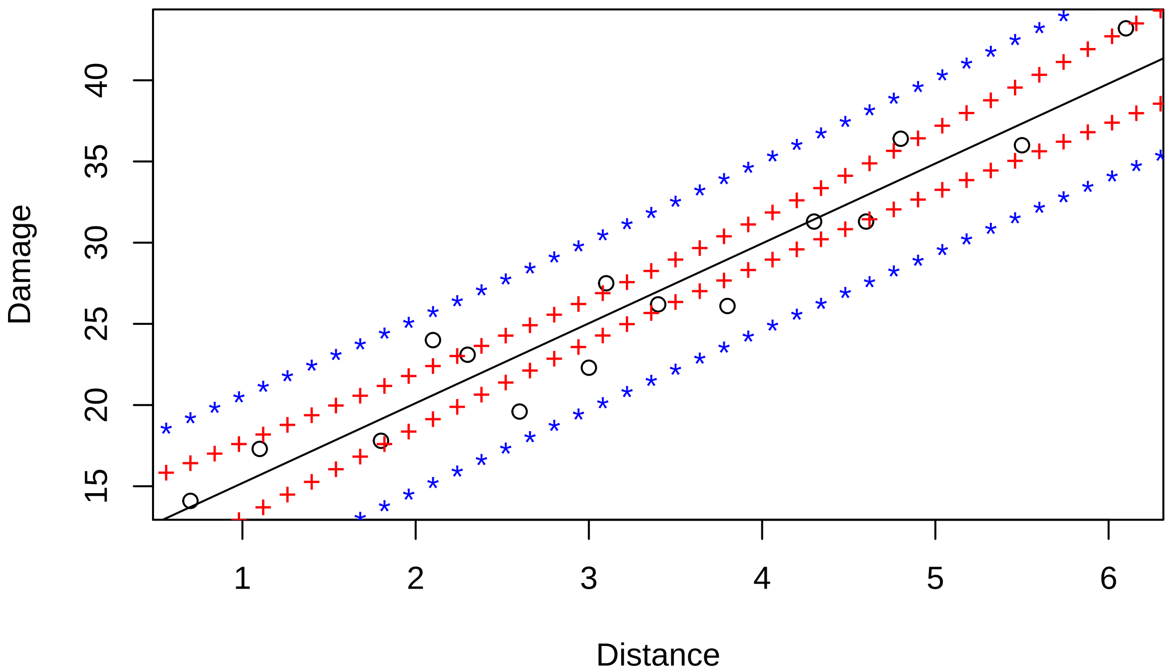
```
## Warning in plot.xy(xy, type, ...): "data" n'est pas un paramètre graphique
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "data" n'est
## pas un paramètre graphique
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "data" n'est
## pas un paramètre graphique
```

```
## Warning in box(...): "data" n'est pas un paramètre graphique
```

```
## Warning in title(...): "data" n'est pas un paramètre graphique
```


Damage
40
35
30
25
20
15
1
2
3
4
5
6
Distance