Group 8 - Lab 4

David Wiley / Duy Truong

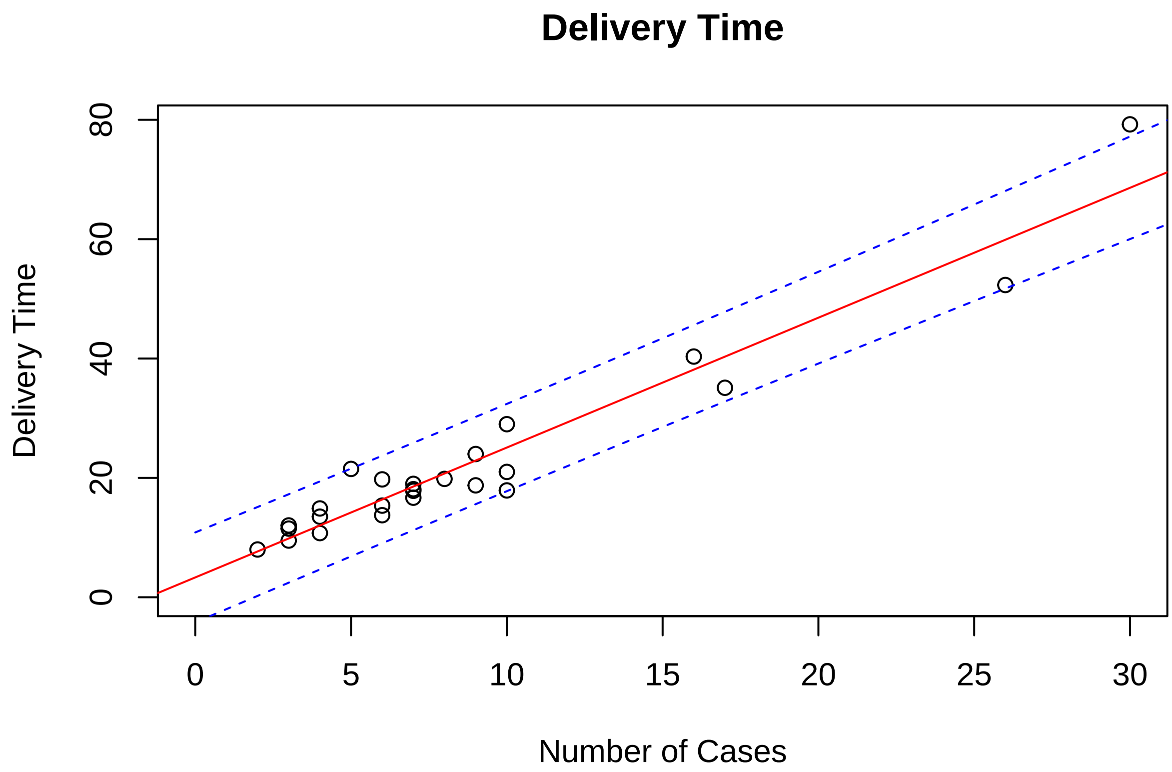
February 4, 2019

```
# reading the data in

dat <- read.csv("C:\\Users\\Nick\\Documents\\0_Spring 2019\\Applied Regression\\Labs_HW\\Data_Sets\\Chap
x = dat[,2]
y = dat[,1]
n = length(x)
```

Part 1:

For the soft drink delivery time data in problem 2.9, plot the 90% prediction interval and interpret.

```
# plotting the information

newx =  data.frame(x = seq(from = 0, to = max(y), length.out = max(y)))
fit = lm(y~x)

# calculating a prediction interval of %90
pred_int = as.data.frame(predict(fit, newx, level = 0.90, interval = "pred"))

# plotting the data
plot(x,y,
     xlim = c(0, max(x)),
     ylim = c(0, max(y) ),
     main = "Delivery Time",
     xlab = "Number of Cases",
     ylab = "Delivery Time")

abline(fit, col="red")

# plotting the prediction intervals
lines(cbind(newx,pred_int$lwr), col = "blue", lty = "dashed")
lines(cbind(newx,pred_int$upr), col = "blue", lty = "dashed")
```



Interpretation:

The %90 prediction interval on the plot is used to estimate where a future observation will fall. The observation has a probability of .9 that the observation will fall between these two intervals.

Part 2:

Calculate R^2 and interpret.

To calculate R^2, we can use two formulas:

$$R^2 = SS_R/SS_T$$

or:

$$R^2 = 1 - \frac{SS_{RES}}{SS_T}$$

We will use both methods to show they are the same.

```
# calculating variables to be used to calculate R^2

Sxx = sum(x^2) - sum(x)^2 / n
Sxy = sum(x * y) - sum(x) * sum(y) / n

B1H = Sxy / Sxx
B0H = mean(y) - mean(x)*B1H

yH = B0H + sort(x)*B1H
```

```r
res = y - (B0H + x*B1H)
SSres = sum(res^2)

SSr = sum( (yH - mean(y))^2 )
SSt = SSr + SSres

# Calculating R^2 in two ways

R2_1 = SSr / SSt

# or

R2_2 = 1-SSres / SSt

R2_1
```

```
## [1] 0.9304813
```

```r
R2_2
```

```
## [1] 0.9304813
```

$R^2 =$ 0.9305

Interpretation:

R^2 is the proportion of the variance in the dependent variable that is predictable from the independent variable. It is used to measure the accuracy of the model and how well it predicts future outcomes. This means that the model can explain $\%$93.05 of variability of the response data around its mean.