Correlation and Simple Linear Regression

Code

A. Covariance and correlation We can compute the covariance and correlation in R using the cov() and cor() functions. Ex. A pediatrician wants to study the relationship between a child's height and their head circumference (both measured in inches). She selects a SRS of 11 three-year old children and obtains the following data. (See lecture notes for data) Begin by reading in the data:

Hide

```
Height = c(27.75, 24.5, 25.5, 26, 25, 27.75, 26.5, 27, 26.75, 26.75, 27.5)
Circ = c(17.5, 17.1, 17.1,17.3, 16.9, 17.6, 17.3, 17.5, 17.3, 17.5, 17.5)
Dat = data.frame(Height,Circ)
attach(Dat)
```

```
The following objects are masked _by_ .GlobalEnv:

    Circ, Height
```

Hide

```
Dat
```

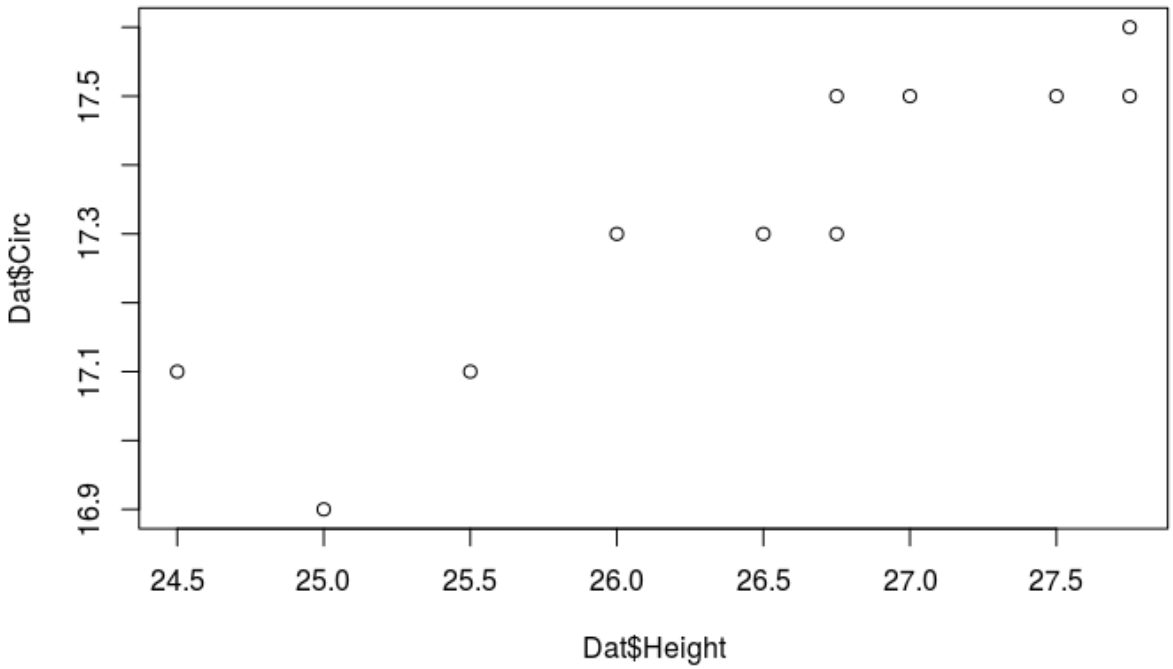
	Height <dbl>	Circ <dbl>
	27.75	17.5
	24.50	17.1
	25.50	17.1
	26.00	17.3
	25.00	16.9
	27.75	17.6
	26.50	17.3
	27.00	17.5
	26.75	17.3
	26.75	17.5

1-10 of 11 rows Previous **1** 2 Next

To make a scatter plot of circumference against height type: 17.6

Hide

```
plot(Dat$Height,Dat$Circ)
```



Studying the plot, there appears to be a linear relationship between the two variables. This relationship can be quantified by computing the covariance and correlation between variables.

Hide

```
cov(Dat) # Covariance matrix
```

```
          Height       Circ
Height 1.1977273 0.21886364
Circ   0.2188636 0.04818182
```

From the output we see that the variance of Height and Circ is 1.198 and 0.048, respectively. The covariance between the two variables is 0.219 indicating a positive relationship.

Hide

```
cor(Dat) # Correlation matrix
```

```
          Height      Circ
Height 1.0000000 0.9110727
Circ   0.9110727 1.0000000
```

Simple Linear Regression

If there exists a strong linear relationship between two variables it is often of interest to model the relationship using a regression line. The main function for performing regression in R is lm(). It has many options that we will explore throughout the semester. To perform simple linear regression we can use the command: lm(response ~ explanatory) Here the terms response and explanatory in the function should be replaced by the names of the response and explanatory variables, respectively, used in the analysis. Ex. Fit a regression line that describes the relationship between Height and Circumference.

Hide

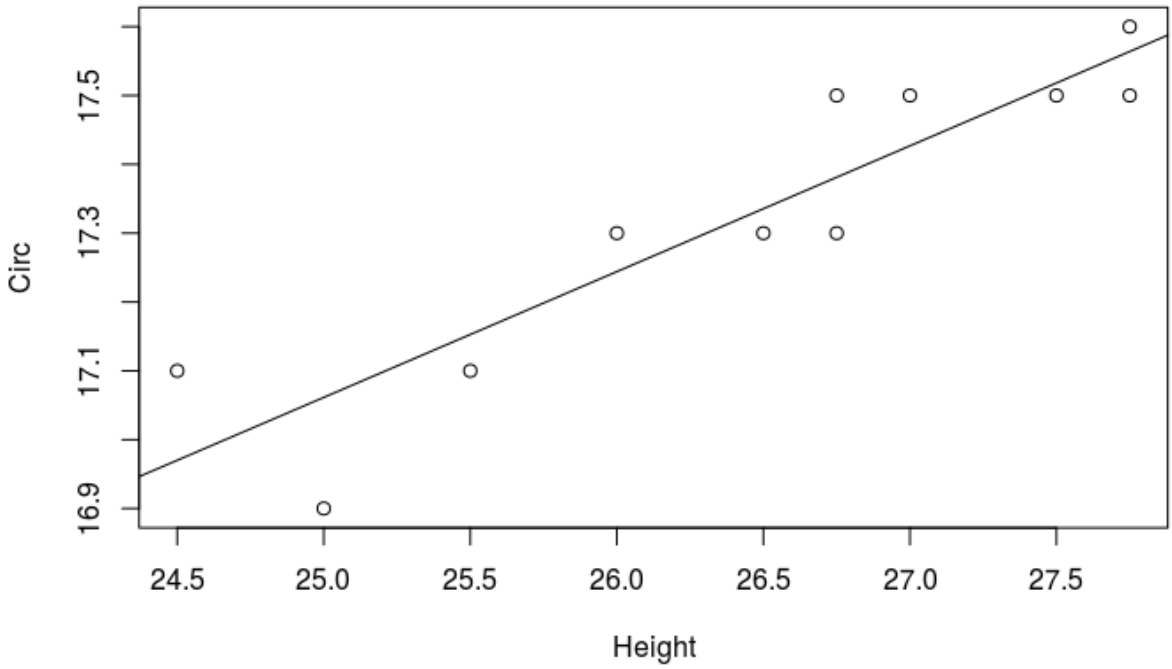
```
results = lm(Circ ~ Height)
results
```

```

Call:
lm(formula = Circ ~ Height)

Coefficients:
(Intercept)       Height
    12.4932       0.1827
```

The results indicate that the least squares regression line takes the form: yhat = 12.493 + 0.183x. Hence the model states that a one inch increase in height would lead to a 0.183 inch increase in head circumference. To superimpose the regression line over the data first make a scatter plot of Circ against Height, and therefore overlay the regression line using the command abline(results). Here results contains all relevant information about the regression line.

Hide

```
plot(Height,Circ)
abline(results)
```


Circ
16.9
17.1
17.3
17.5
24.5
25.0
25.5
26.0
26.5
27.0
27.5
Height