# MATH 3307
# Lesson 18

# Correlation Coefficient

Abbreviation: r

The **correlation coefficient** measures the strength and direction of the linear relationship between two quantitative variables. The formula to find *r* is:

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

The point $(\bar{x}, \bar{y})$ is: (the mean of x-values, the mean of y-values)

The values of $s_x$ and $s_y$ are the individual standard deviations of *x* and *y* respectively.

*n* represents the number of data pieces.

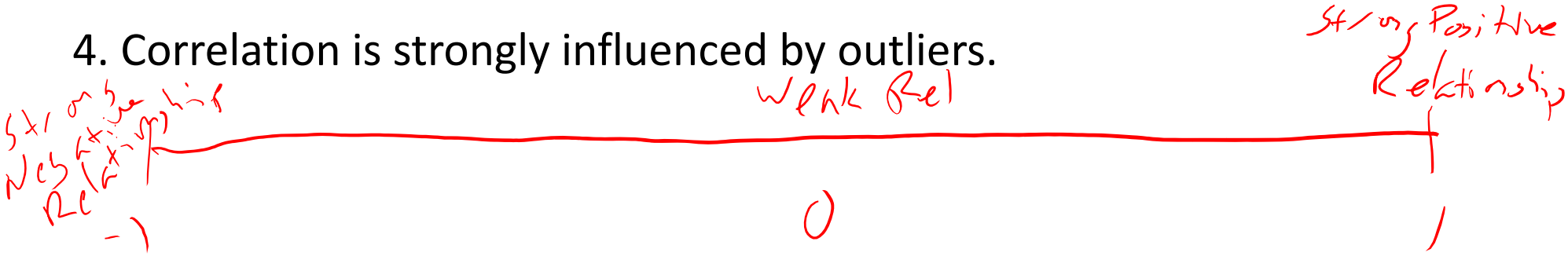
# Facts about Correlation

1. Positive *r* indicates positive association and negative *r* indicates negative association between variables.

2. *r* is always between –1 and 1.

3. The closer |*r*| is to 1, the *stronger* the association. A *weak* association will have an *r* value close to 0.

4. Correlation is strongly influenced by outliers.



# Example of a Correlation Coefficient

Calculating in R-Studio:

cor(a,b)

Using the monopoly example from Section 5.1:

assign("spaces",c(1,3,5,6,8,9,11,12,13,14,15,16,18,19,21,23,24,25,26,27,28,29,31,32,34,35,37,39))

assign("cost",c(60,60,200,100,100,120,140,150,140,160,200,180,180,200,220,220,240,200,260,260,150,280,300,300,320,200,350,400))

Determine the Correlation Coefficient.

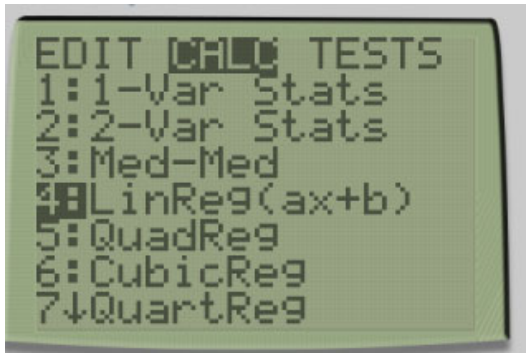
```
cor(spaces,cost)
1] 0.8779736
```

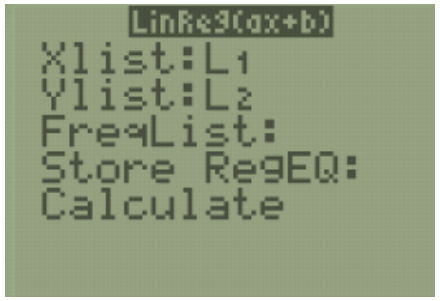
What does this mean? Positive association between these variables. Moderately strong.

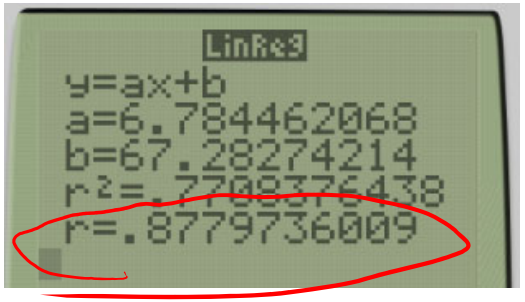
Do this once (not for every question)



Do this for each question:



# Popper 13

Create a scatter plot from the data.

Based on the plot:

1. Is this a positive, negative or no relationship?

```
plot(age,height)
```

a. positive b. negative c. none

2. Is the relationship linear or not?

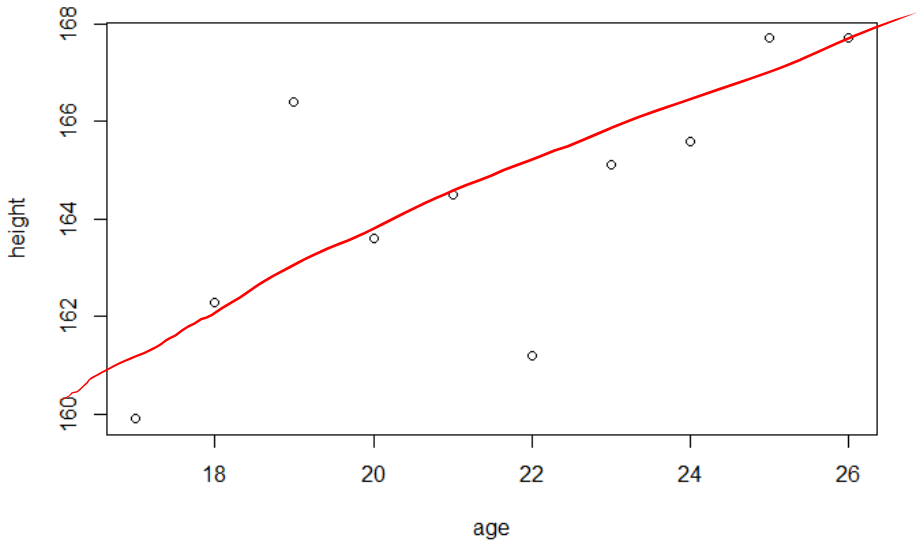
a. linear b. not linear

| Age | Height (cm) |
|---|---|
| 17 | 159.9 |
| 18 | 162.3 |
| 19 | 166.4 |
| 20 | 163.6 |
| 21 | 164.5 |
| 22 | 161.2 |
| 23 | 165.1 |
| 24 | 165.6 |
| 25 | 167.7 |
| 26 | 167.7 |

*International Journal of Morphology*

# To Copy Into RStudio

```
assign("age",c(17,18,19,20,21,22,23,24,25,26))
assign("height",c(159.9,162.3,166.4,163.6,164.5,161.2,165.1,165.6,167.7,167.7))
```



# Popper 13…continued

```
cor(age,height)
1] 0.7281053
```

3. Calculate the correlation coefficient.

a. 0.2265 b. 0.2393 c. 0.7281 d. 0.0794

4. Based on the correlation coefficient, determine the direction of the relationship?

a. positive b. negative c. neither

5. Based on the correlation coefficient, is this relationship strong ($|r| > 0.75$), moderate ($0.5 < |r| < 0.74$) or weak ($|r| < 0.5$)?

a. strong b. moderate c. weak

The RStudio data package Orange contains data of the "age" (in days) and "circumference" (in mm) of five different orange trees.

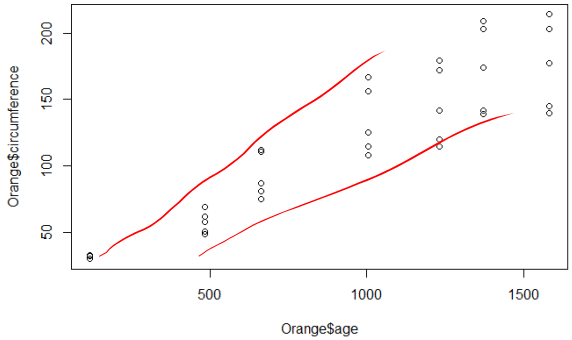
Explanatory: age; Response: circumference

```
plot(Orange$age,Orange$circumference)
```

Plot the scatterplot comparing tree age with tree circumference.

What do you notice about the scatterplot?

(direction, strength, shape)




Positive strong linear

Determine the correlation coefficient.

Does this agree with the predictions from the graph?

```
> cor(Orange$age,Orange$circumference)
[1] 0.9135189
```

positive, strong (linear: from r-definition)