

# **Purpose**

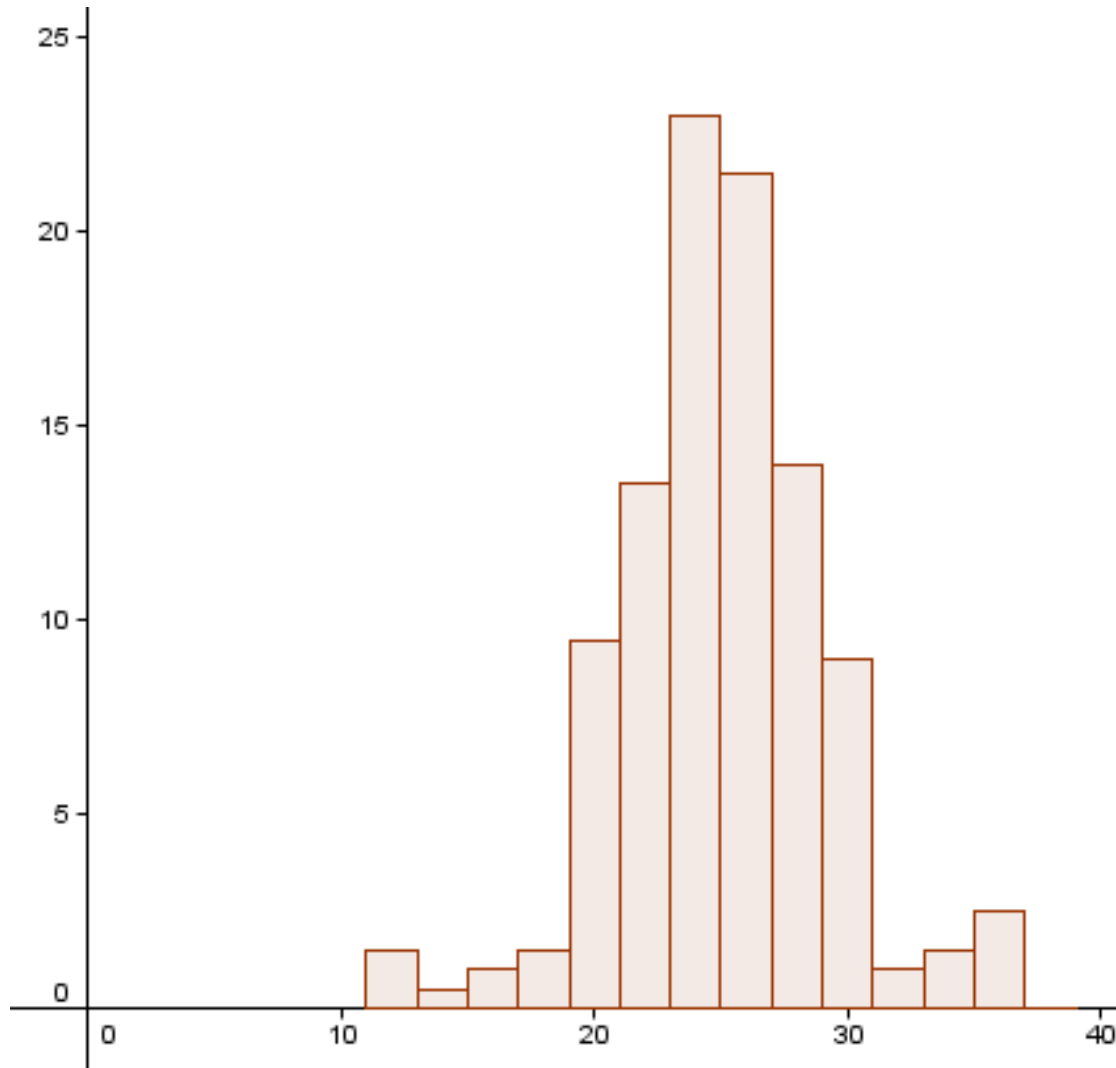
*The purpose of this section is to show you how you can get students to discover the Empirical Rule from their own data.*

# ***Right Foot Lengths***

*On your sheets you have data for right foot lengths.*

*Can you describe the distribution of the right foot lengths?*

# Right Foot Lengths



# Analysing the Distribution of Right Foot Lengths

1. Complete the table using the mean and the standard deviation.

Mean = 24.6cm

Standard Deviation = 4.06cm  $\approx$  4cm

Fill in the following table:

3 standard deviations below the mean	2 standard deviations below the mean	1 standard deviation below the mean	Mean	1 standard deviation above the mean	2 standard deviations above the mean	3 standard deviations above the mean
			24.6cm	28.6cm		

2. Fill in the left hand side of the larger table. (Hint: There are 20 numbers in each column)

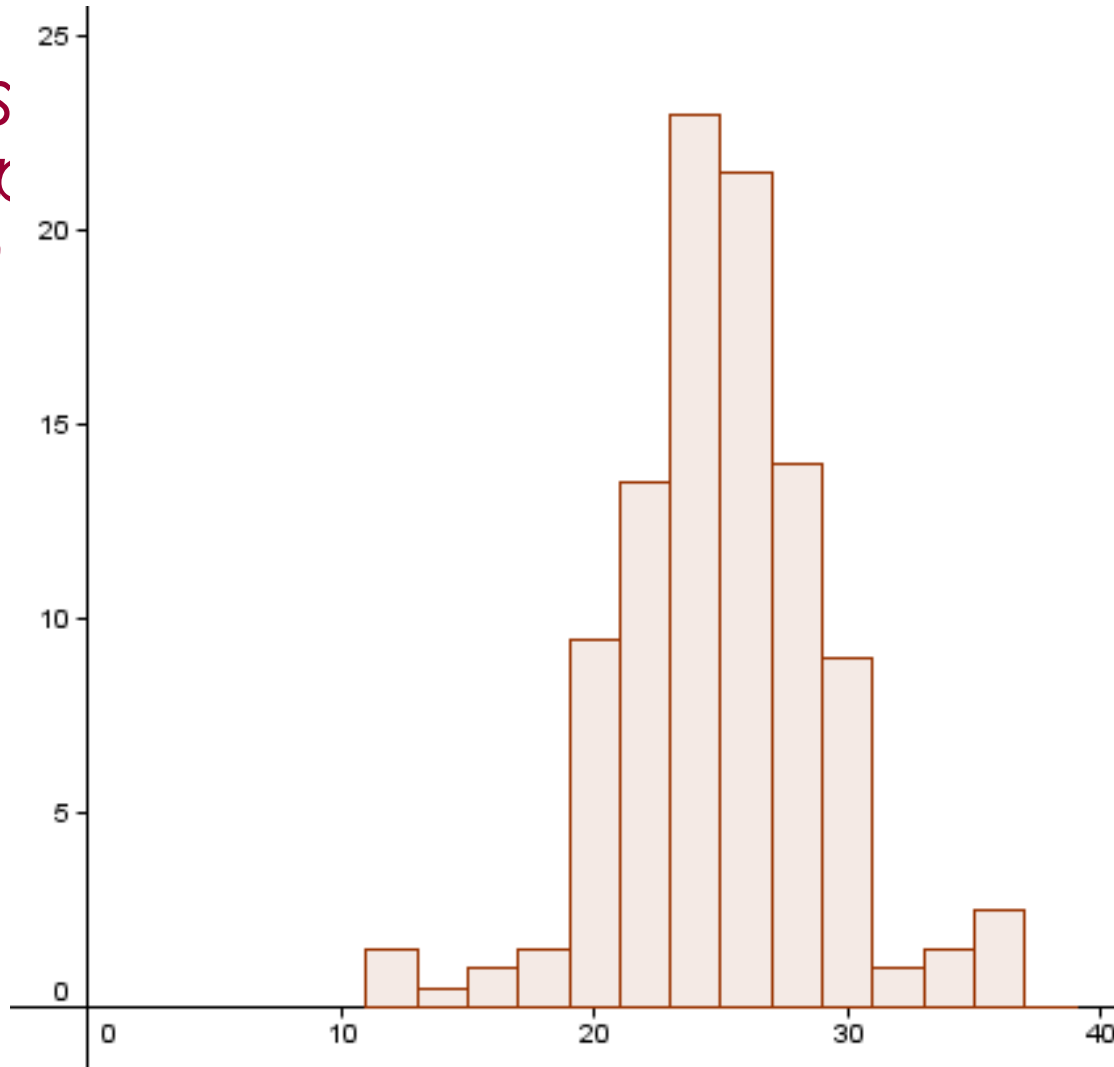
3. Fill in the right hand side of the larger table by working out the Percentage of the 200 students who are in each of the 3 categories.

From the table above, count how many numbers are between 1 standard deviation below the mean and 1 standard deviation above the mean? <input type="text"/>	What percentage of the 200 numbers is to be found within 1 standard deviation of the mean? <input type="text"/>
From the table above, count how many numbers are between 2 standard deviations below the mean and 2 standard deviations above the mean? <input type="text"/>	What percentage of the 200 numbers is to be found within 2 standard deviations of the mean? <input type="text"/>
From the table above, count how many numbers are between 3 standard deviations below the mean and 3 standard deviations above the mean? <input type="text"/>	What percentage of the 200 numbers is to be found within 3 standard deviations of the mean? <input type="text"/>

# The Empirical Rule

The ans  
the Em  
percen

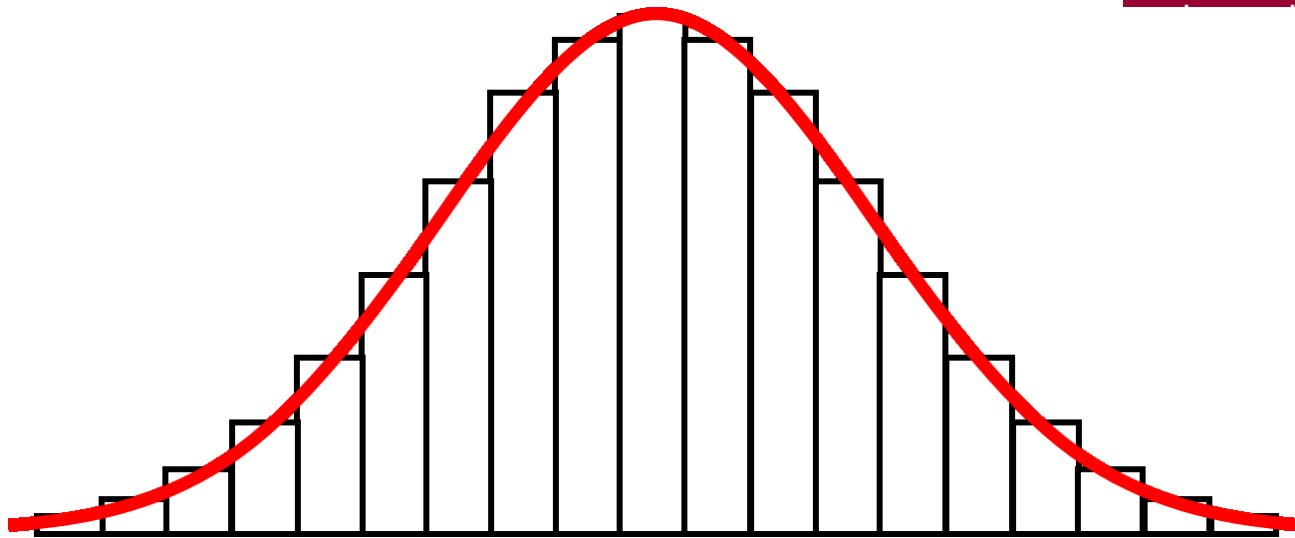
e to



# Normal Distribution

*A symmetric, unimodal, bell-shaped distribution of continuous random data is called a normal distribution.*

*It is the most common model in statistics, because so often, with a big enough sample, we see the data mostly collect around the mean.*

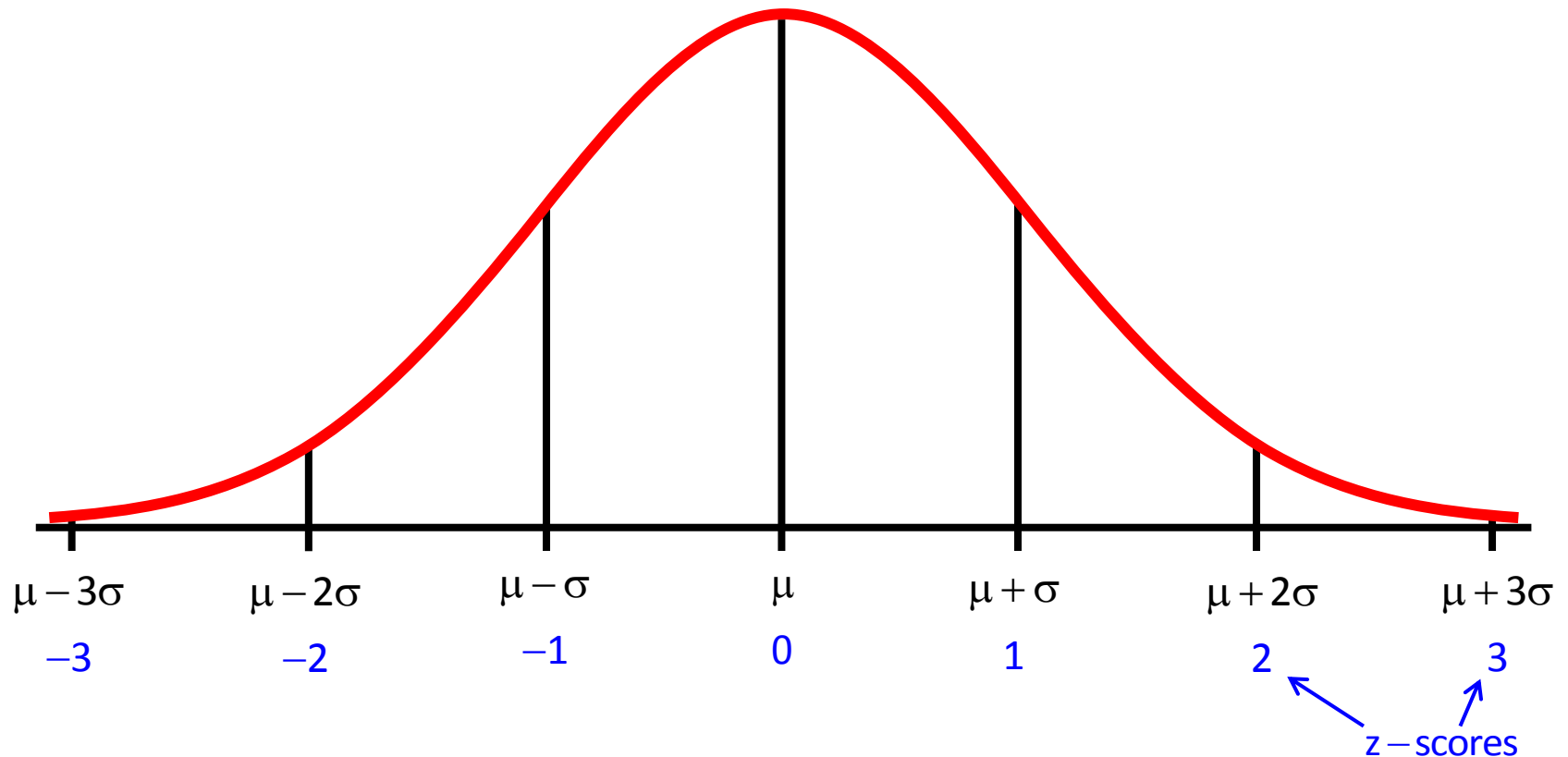


If we plot the right foot lengths or heights of all students we will start to see a normal distribution. The majority of students will be around an average height.

# Standard Normal Distribution

If  $\mu = 0$  and  $\sigma = 1$  we would plot  $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$

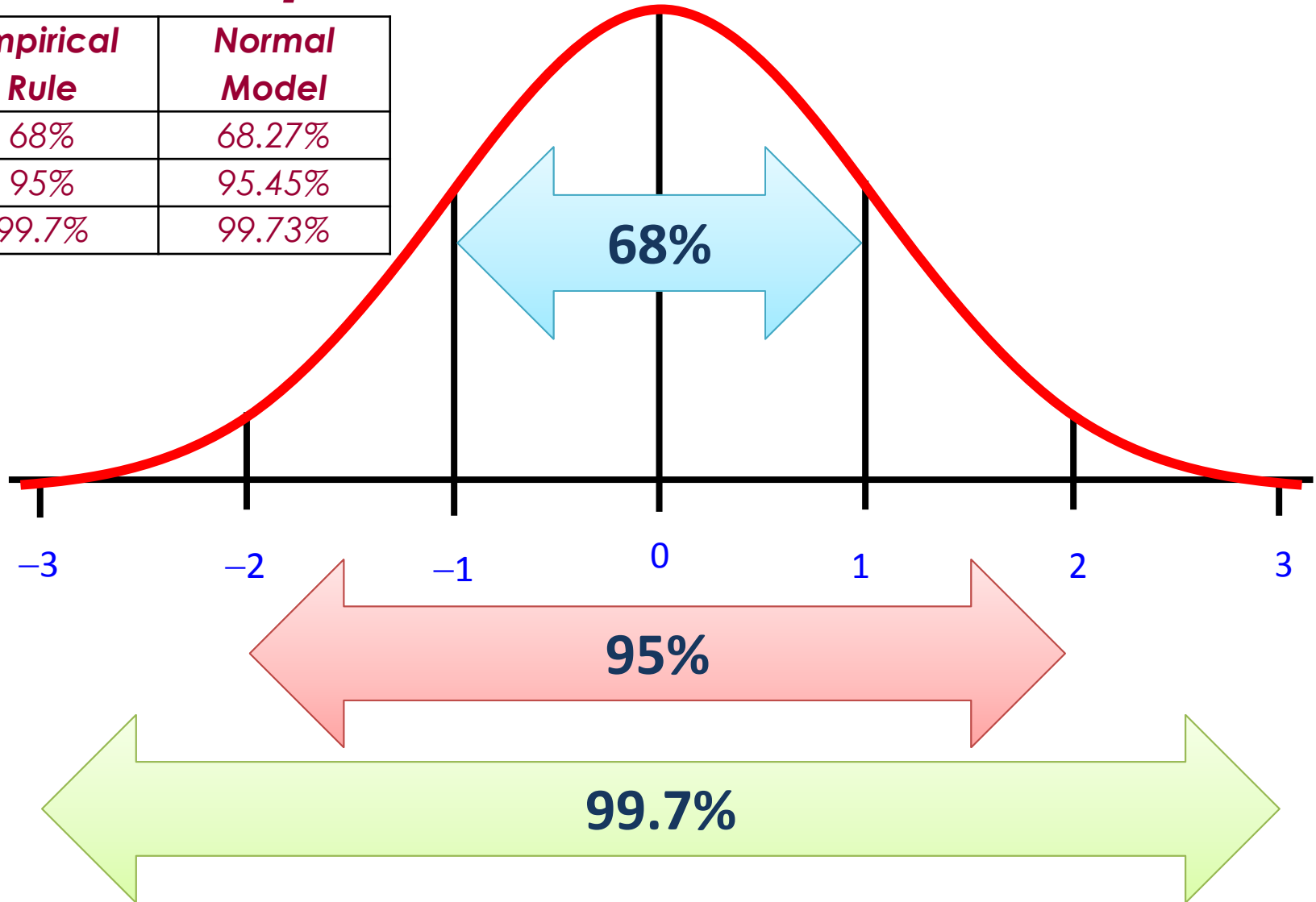
This graph gives the Standard Normal Graph with a standardised scale.



The area between the Standard Normal Curve and the z-axis between  $-\infty$  and  $+\infty$  is 1.

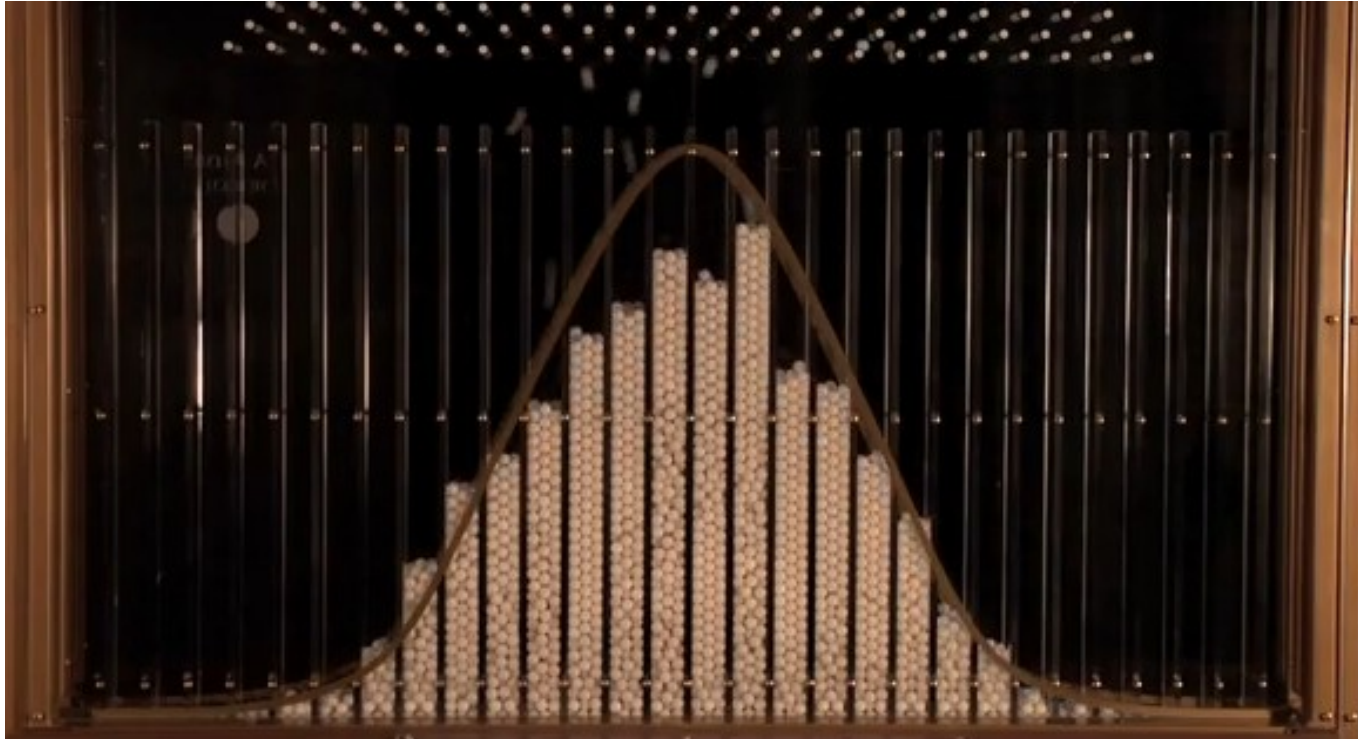
# The Normal Model also shows the Empirical Rule is accurate

<b>Empirical Rule</b>	<b>Normal Model</b>
68%	68.27%
95%	95.45%
99.7%	99.73%





# *The Normal Distribution Machine*



# Short Questions

*When data is normally distributed describe the chances of a piece of data being near the middle.*

*When data is normally distributed what are the chances of being within 1 standard deviation of the mean?*

*If we assume data is normally distributed what are the chances of being within 2 standard deviations of the mean?*

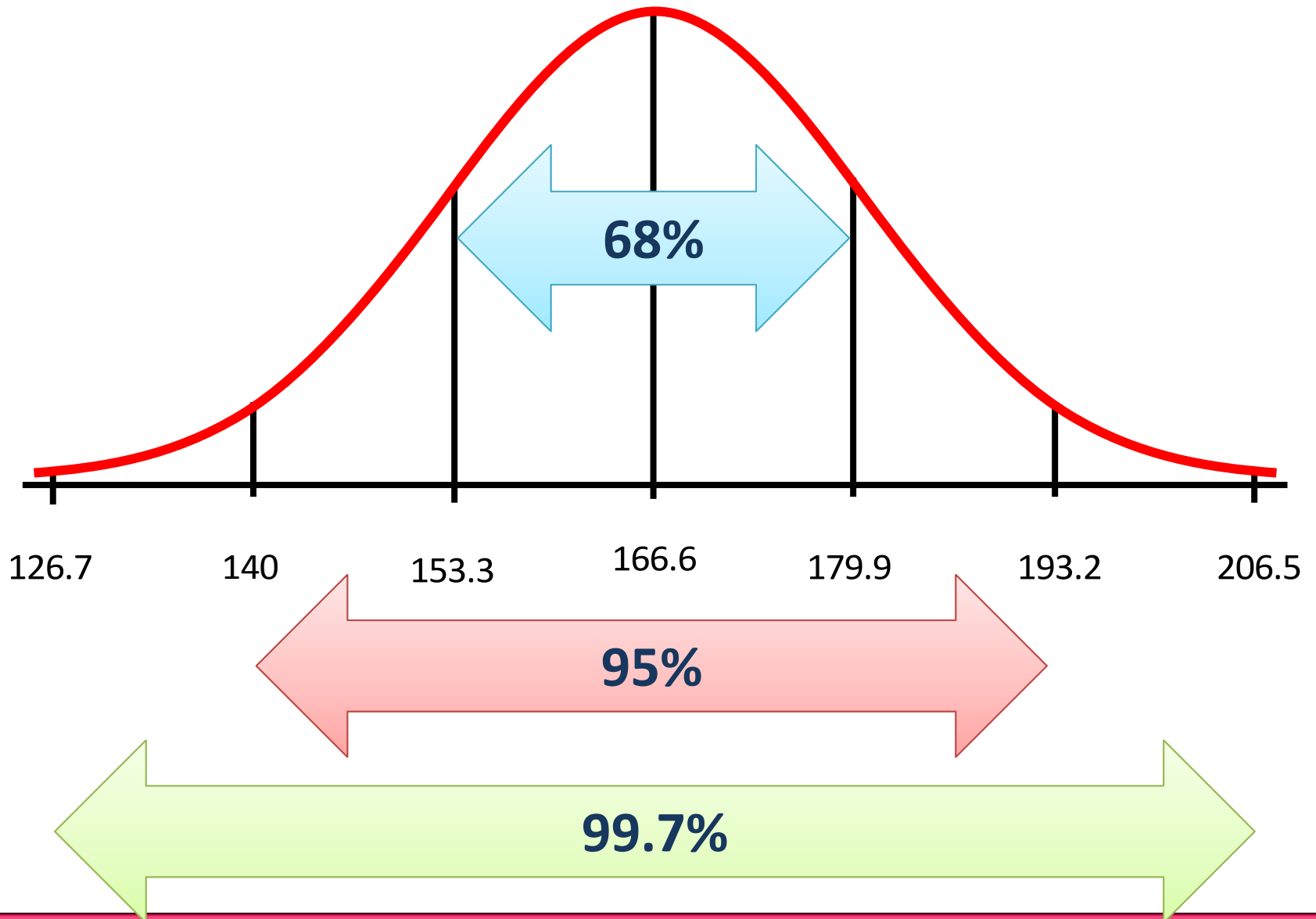
# Short Questions

The mean height is 166.6cm and the standard deviation is 13.3cm.

Based on an assumption that the distribution of heights is approximately normal, use the empirical rule for the following questions:

- (i) 68% of this school's students have heights between \_\_\_ and \_\_\_ cm.
  
- (ii) What percentage of students have heights between 140cm and 193.2cm?
  
- (iii) A school tour is being organised. All students can apply to go on it. There is a rollercoaster at one location on the tour. You have to be over 140cm to be allowed on the rollercoaster. What percentage of students wouldn't be tall enough?

# *Empirical Rule*

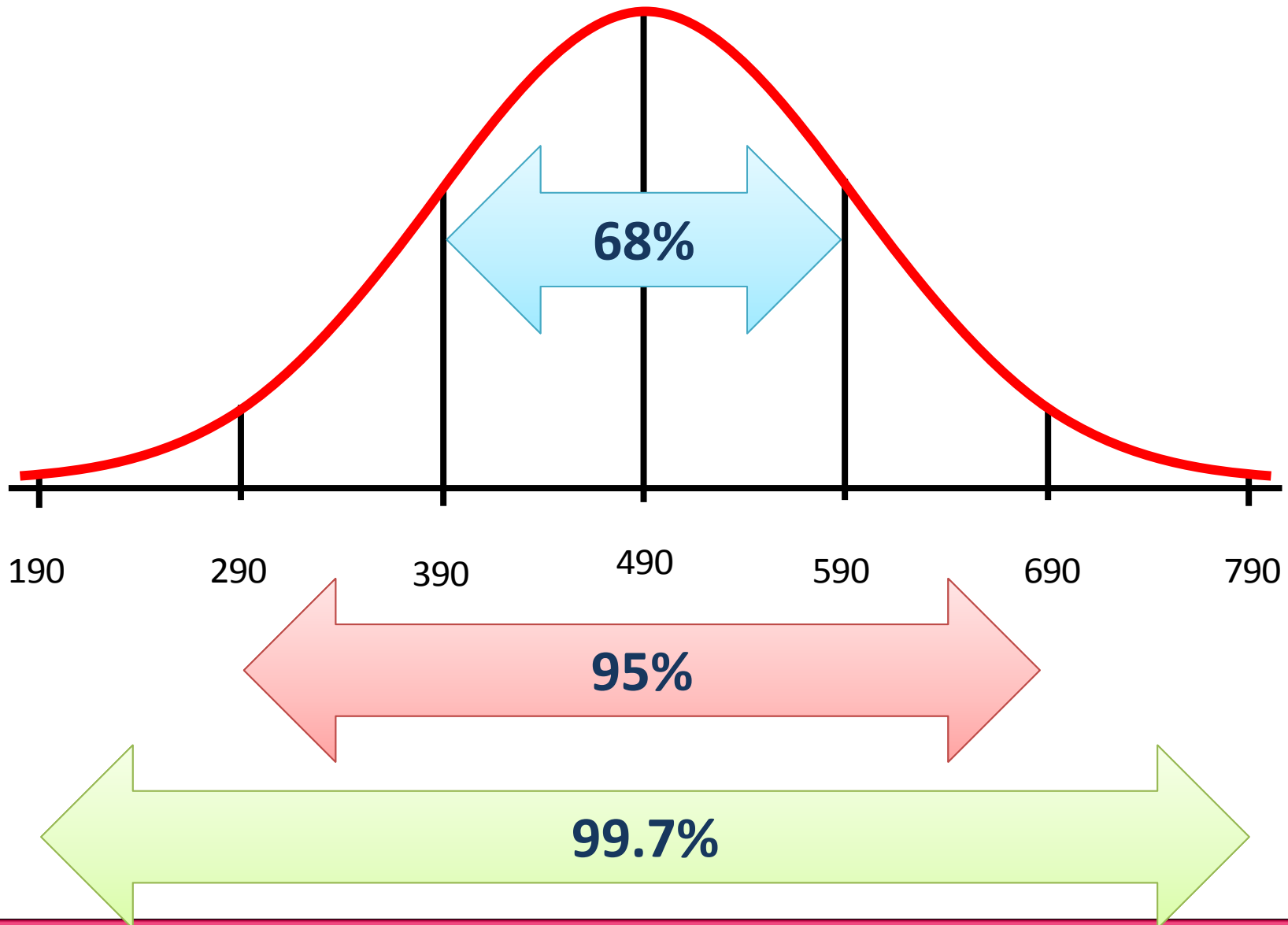


# NCCA Material

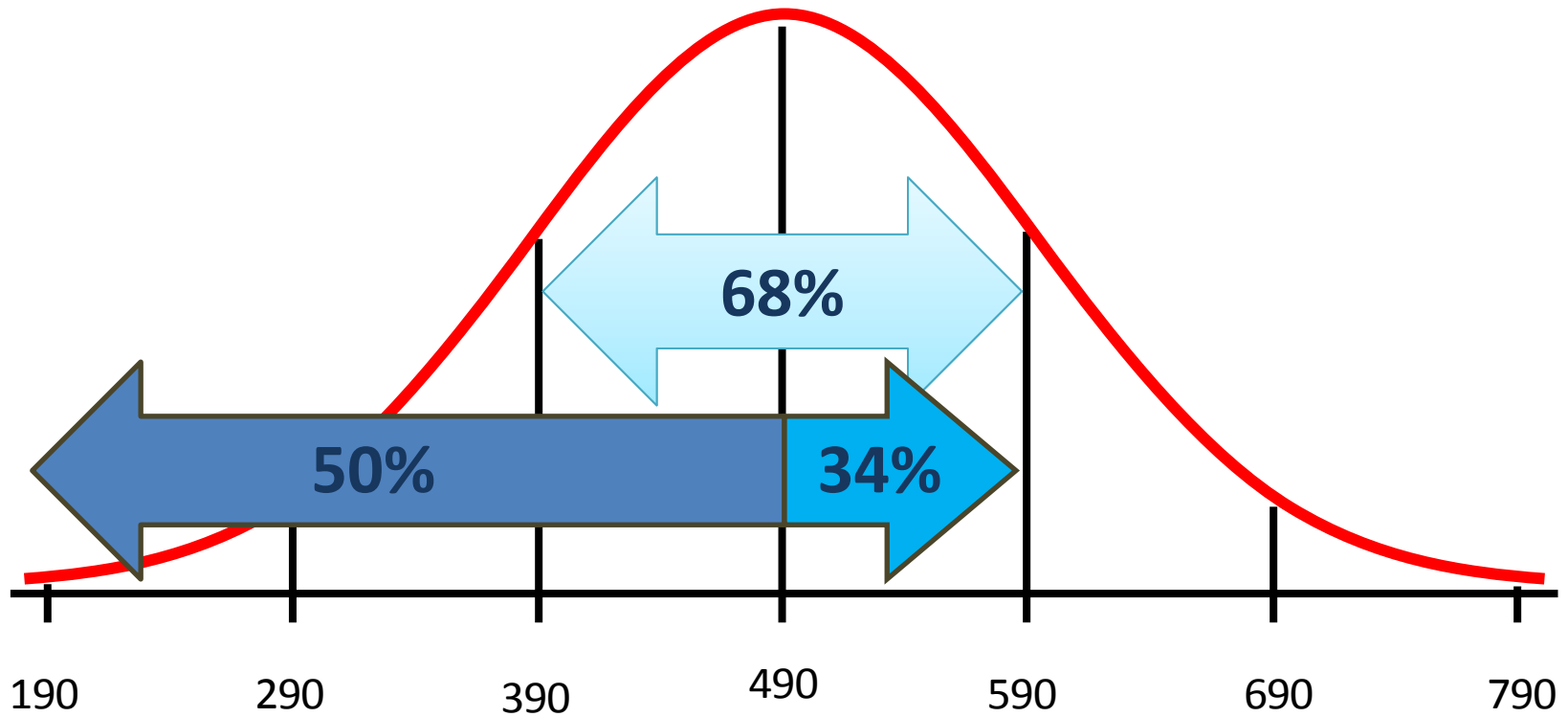
To enter a particular college course, candidates must complete an aptitude test. In 2010 the mean score was 490 with a standard deviation of 100. The distribution of the scores on the aptitude test is a normal distribution.

- a) What percentage of candidates scored between 390 and 590 on this aptitude test?
- b) One student scored 795 on this test. How does this student's score compare to the rest of the scores?
- c) The college admits only students who were among the highest 16% of the scores on this test. What score would a student need on this test to be qualified for admission to this college? Explain your answer.

# *Empirical Rule*



# Empirical Rule



# **Purpose**

*To identify all the concepts we must keep in mind when we use data from a sample to talk about the population.*

*We will also tease out how each concept influences the others.*

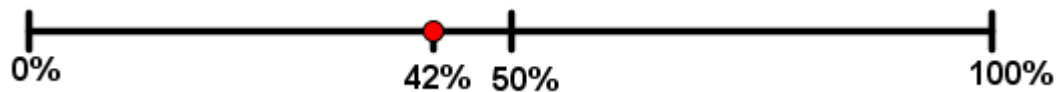


# Mobile Phones

*You wanted to work out which mobile phone company is the most popular among Irish second level students.*

*You used a sample of 100 students and found that Meteor topped the survey with 42%.*

*What percentage of all second level students in the country use Meteor?*



# Census at School Data Tool

Summary table Show Options

**Mobile Network**

Random sample	Frequency chart
Meteor	3
Vodafone	Meteor <b>50</b>
O2	O2
O2	Tesco Mobile 2
Meteor	Vodafone 25
Meteor	
Meteor	
Meteor	
Vodafone	

Number <small>not incl blanks</small>	Mean	Quartiles
100		Q1: <small>lower quartile</small>
Range	Standard deviation	Q2: <b>Median</b>
to		Q3: <small>upper quartile</small>

Summary table Show Options

**Mobile Network**

Random sample	Frequency chart
Meteor	3
Vodafone	Meteor <b>42</b>
O2	O2
Meteor	Vodafone 26
Meteor	
O2	
O2	
O2	

Number <small>not incl blanks</small>	Mean	Quartiles
100		Q1: <small>lower quartile</small>
Range	Standard deviation	Q2: <b>Median</b>
to		Q3: <small>upper quartile</small>

Summary table Show Options

**Mobile Network**

Random sample	Frequency chart
Vodafone	3
Vodafone	Meteor <b>39</b>
3	O2
O2	Tesco Mobile 3
O2	Vodafone 24
Meteor	
Vodafone	
O2	

Number <small>not incl blanks</small>	Mean	Quartiles
100		Q1: <small>lower quartile</small>
Range	Standard deviation	Q2: <b>Median</b>
to		Q3: <small>upper quartile</small>

Summary table Show Options

**Mobile Network**

Random sample	Frequency chart
Meteor	3
Meteor	Meteor <b>56</b>
Meteor	O2
O2	Vodafone 23
Meteor	
O2	
Vodafone	
Meteor	

Number <small>not incl blanks</small>	Mean	Quartiles
100		Q1: <small>lower quartile</small>
Range	Standard deviation	Q2: <b>Median</b>
to		Q3: <small>upper quartile</small>

Summary table Show Options

**Mobile Network**

Random sample	Frequency chart
O2	3
Vodafone	Meteor <b>41</b>
Vodafone	O2
Meteor	Tesco Mobile -
O2	Vodafone 28
Tesco Mobile	
Vodafone	
Vodafone	

Number <small>not incl blanks</small>	Mean	Quartiles
100		Q1: <small>lower quartile</small>
Range	Standard deviation	Q2: <b>Median</b>
to		Q3: <small>upper quartile</small>

Summary table Show Options

**Mobile Network**

Random sample	Frequency chart
Vodafone	3
O2	Meteor <b>45</b>
Vodafone	O2
Meteor	Vodafone 28
Vodafone	
Meteor	
Meteor	
Meteor	

Number <small>not incl blanks</small>	Mean	Quartiles
100		Q1: <small>lower quartile</small>
Range	Standard deviation	Q2: <b>Median</b>
to		Q3: <small>upper quartile</small>

# Confidence

*Make a comment about the following three statements:*

*A polling company said it was 100% certain that a candidate has between 0% and 100% of the vote.*

*A polling company said it was 50% certain that a candidate has between 30% and 35% of the vote.*

*A polling company said it was 99.9% certain that a candidate has between 4% and 94% of the vote.*

# ***Balancing Margin of Error, Level of Confidence and Sample Size***

# **Important Sample Sizes using the Formula for Margin of Error (Level of Confidence 95%)**

<b>Sample Size</b>	<b>Margin of Error</b>
25	$\pm 20\%$
64	$\pm 12.5\%$
100	$\pm 10\%$
256	$\pm 6.25\%$
400	$\pm 5\%$
625	$\pm 4\%$
1111	$\pm 3\%$
1600	$\pm 2.5\%$
2500	$\pm 2\%$
10000	$\pm 1\%$



# **Purpose**

*To use one open question to work out how we can use statistics to see what we can and can't say about claims that use percentages.*

*We will use the idea of a court room to test claims.*

# **A Claim**

*A drug company claims that their new drug relieves migraine 70% of the time.*

*A newspaper investigates this claim by getting migraine sufferers to try the new drug. They get 100 results that say it relieves migraine 62% of the time.*

*What could the newspaper say about this?*

*What could the newspaper headline be?*



# ***Let's Put the Claim on Trial***

- *What will we assume?*
- *What evidence do we have?*
- *What conclusion will you make?*
- *What action will you take?*

# Hypothesis Test Question Based On Our Data

A teacher claims that 30% of second level students in Ireland are 180cm or taller.

**(a)** If we treat the 200 results from our school as the results of a simple random sample of all second level students then what is the overall margin of error of the survey, at 95% confidence?

**(b)** Of the students in the sample above, 34 are 180cm or taller. Is this sufficient evidence to reject the teacher's claim, at the 5% level of significance?

# **Purpose**

*To show parts of the Teaching and Learning Plan on The Correlation Coefficient*

*To introduce students to the idea of looking for and talking about relationships in bivariate data.*

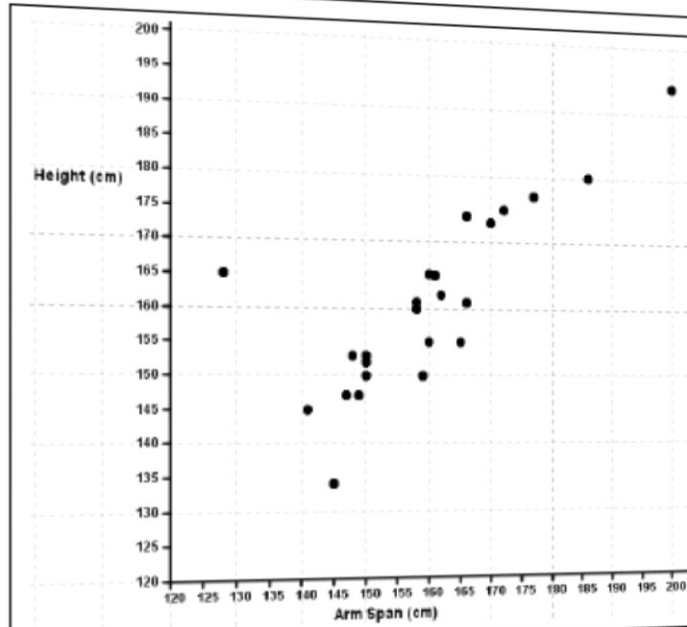
**Teaching & Learning Plans**  
**The Correlation Coefficient**  
Leaving Certificate Syllabus



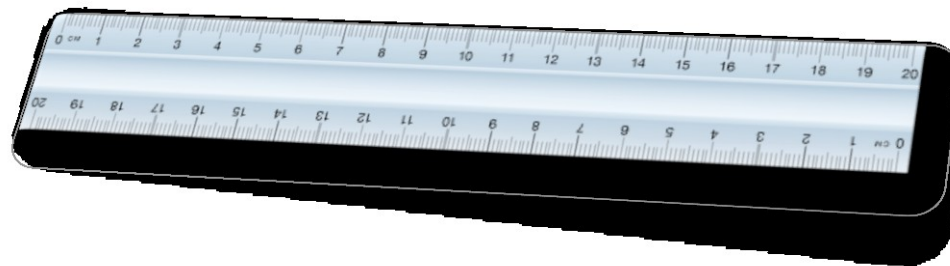
WS5.11

### Arm Span and Height

Arm Span	Height
128	165
141	145
145	134
147	147
148	153
149	147
150	152
150	150
150	153
158	161
158	160
159	150
160	165.1
160	155
161	165
162	162
165	155
166	161
166	174
170	173
172	175
177	177
186	180
200	194



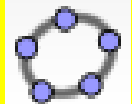
Page 14

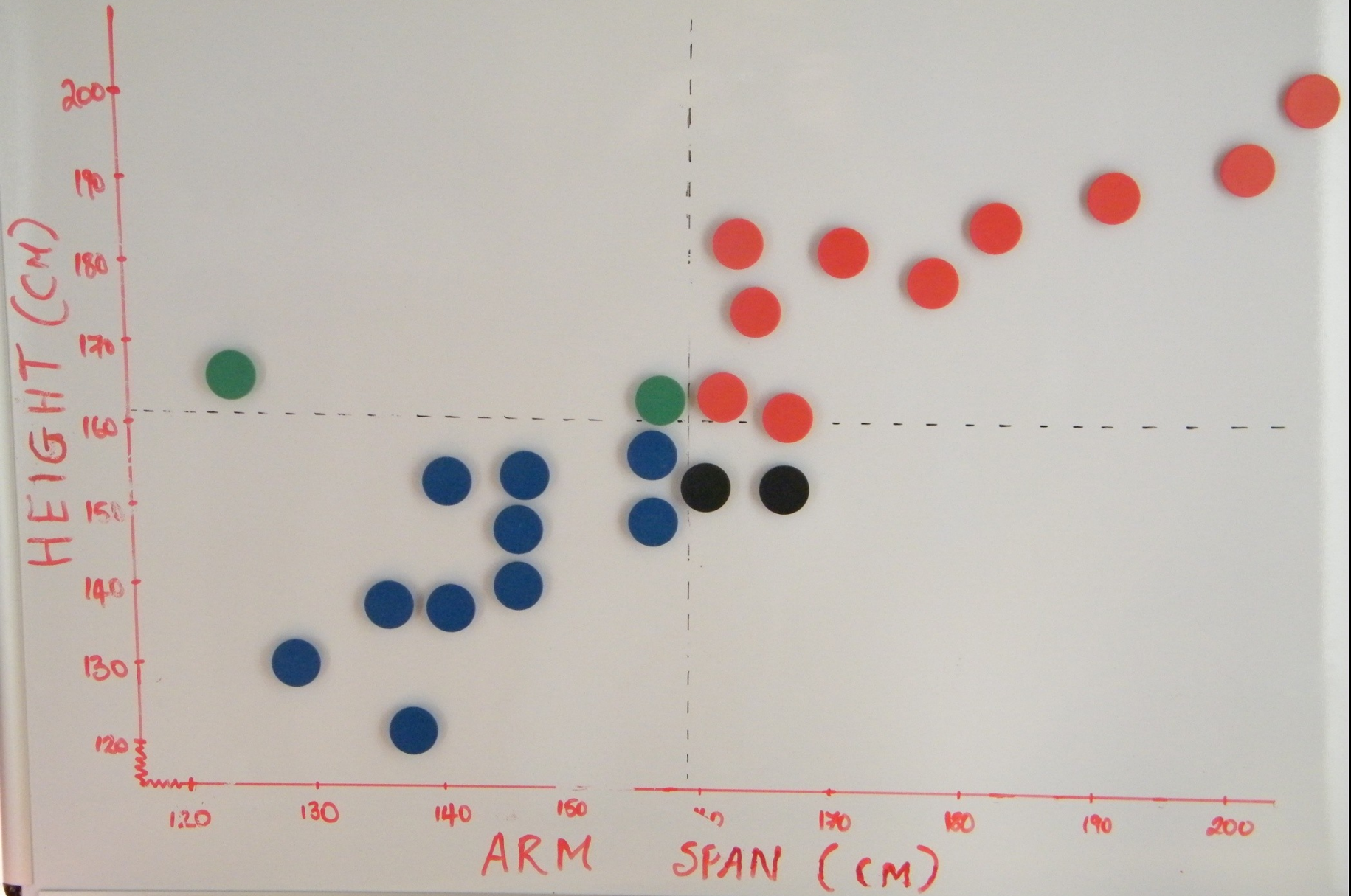


Mean

159.5

160.5





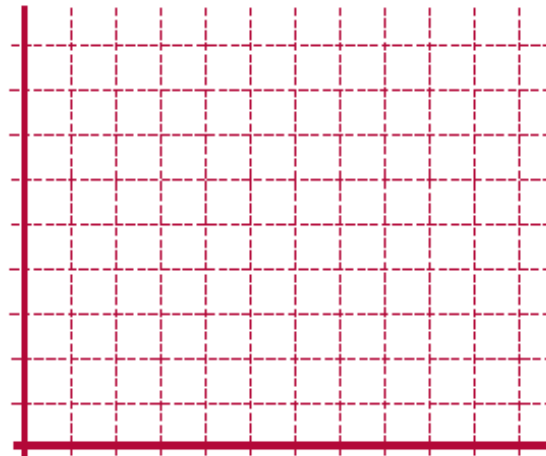
# Teaching & Learning Plan

## Line of Best Fit (HL only)

The following table shows the weekly rainfall ( $x$  cm) and the number of tourists ( $y$  thousand) visiting a certain beauty spot, for 9 successive weeks.

Rainfall ( $x$ cm)	4.5	3.0	5.2	5.0	2.1	0	0	1.2	3.2
No. of tourists ( $y$ thousand)	5.0	8.0	0.8	4.2	4.8	7.4	9.4	8.6	2.6

- i Draw a scatter plot for this data.

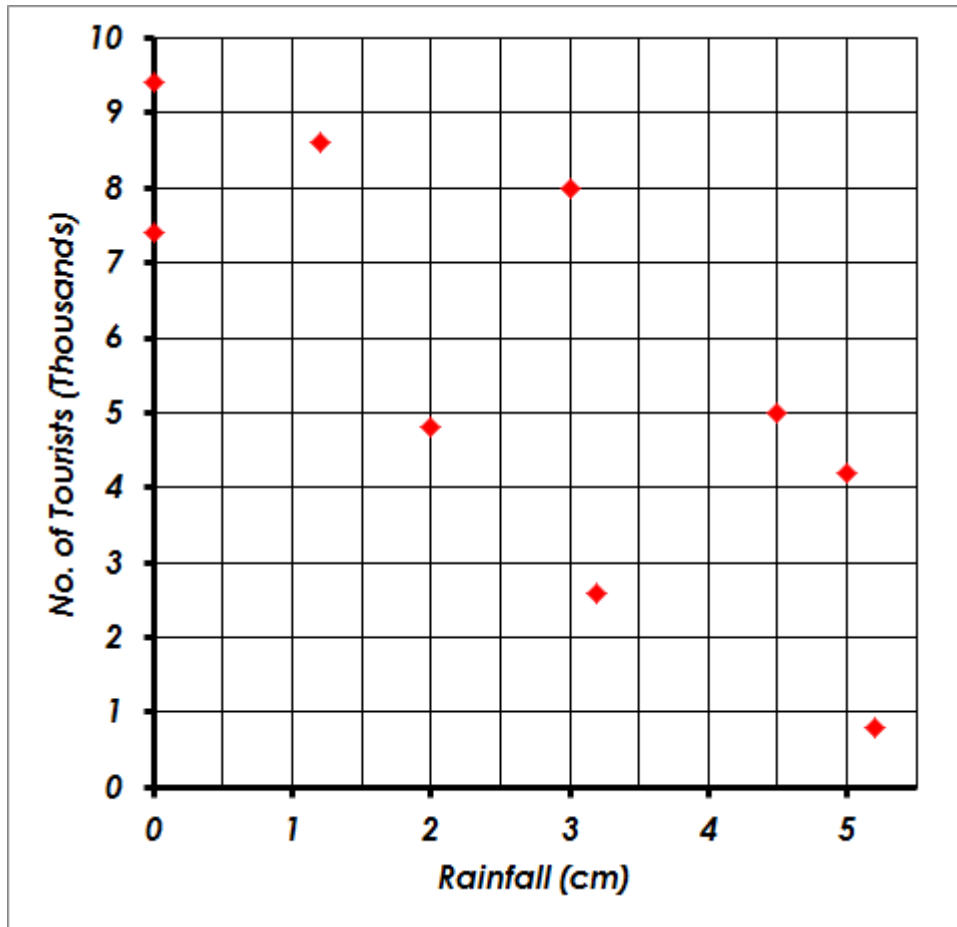


- ii Find the mean rainfall ( $\bar{x}$ ). \_\_\_\_\_
- iii Find the mean number of tourists ( $\bar{y}$ ). \_\_\_\_\_
- iv Plot the point ( $\bar{x}, \bar{y}$ ). Draw lines parallel to the  $x$ -axis and  $y$ -axis through this point.
- v This splits the scatter plot into 4 quadrants. In which quadrants do you find the most points? \_\_\_\_\_

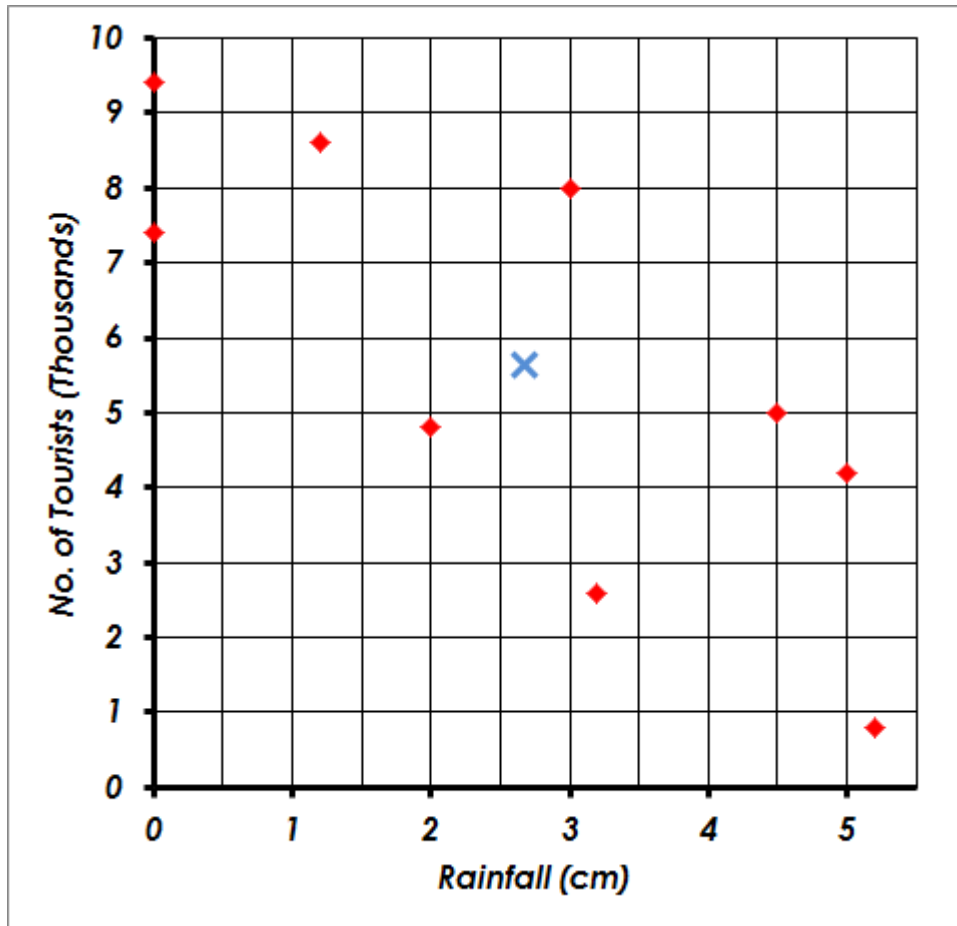




# Teaching & Learning Plan – Correlation Coefficient



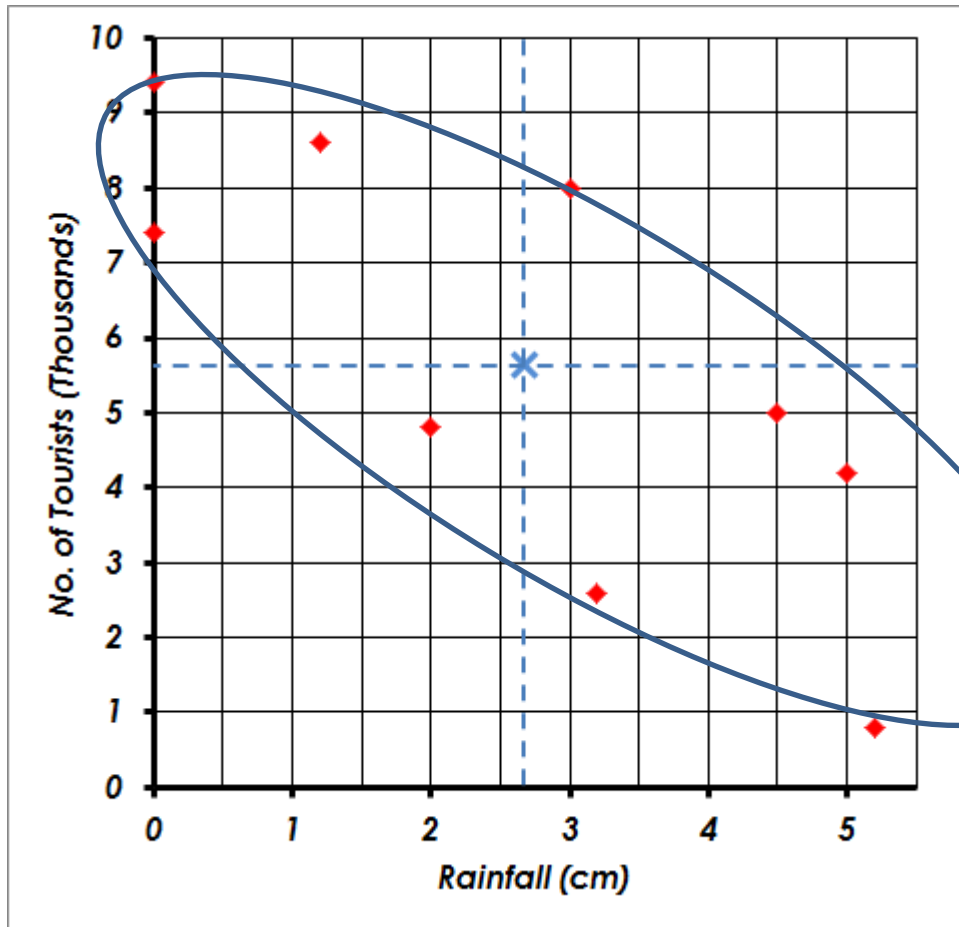
# Teaching & Learning Plan – Correlation Coefficient



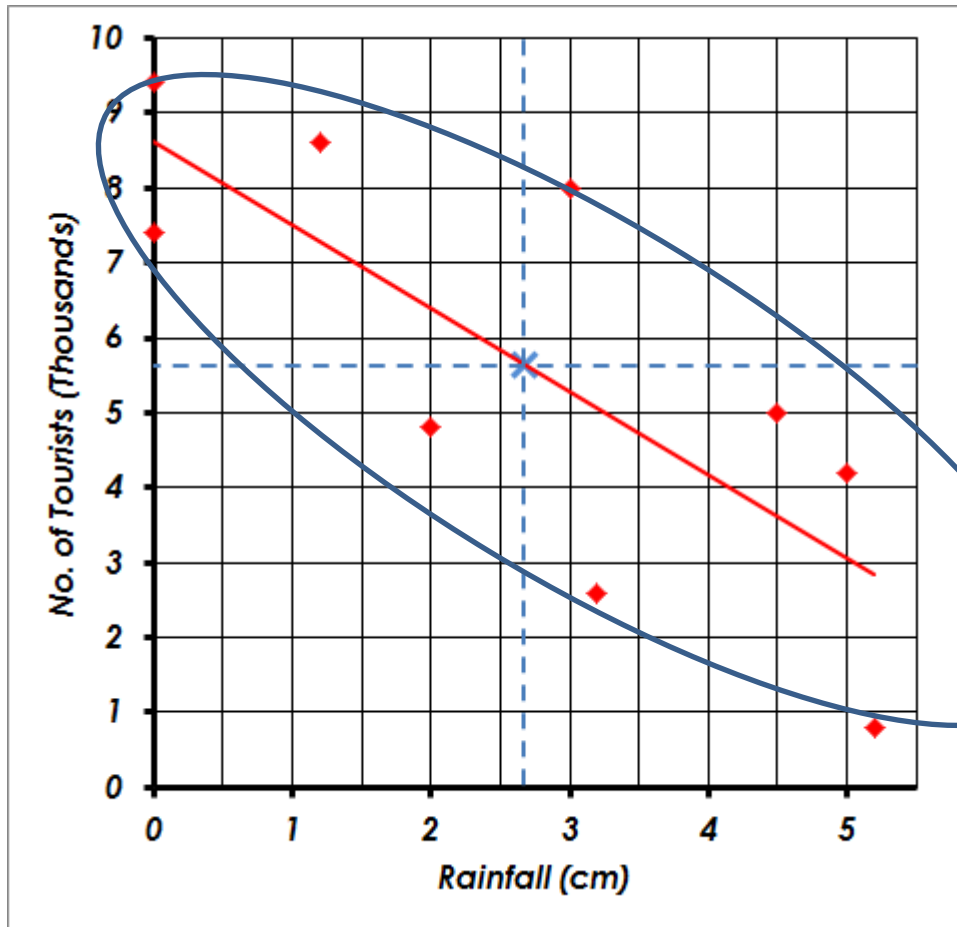
$$\bar{x} = 2.69$$

$$\bar{y} = 5.64$$

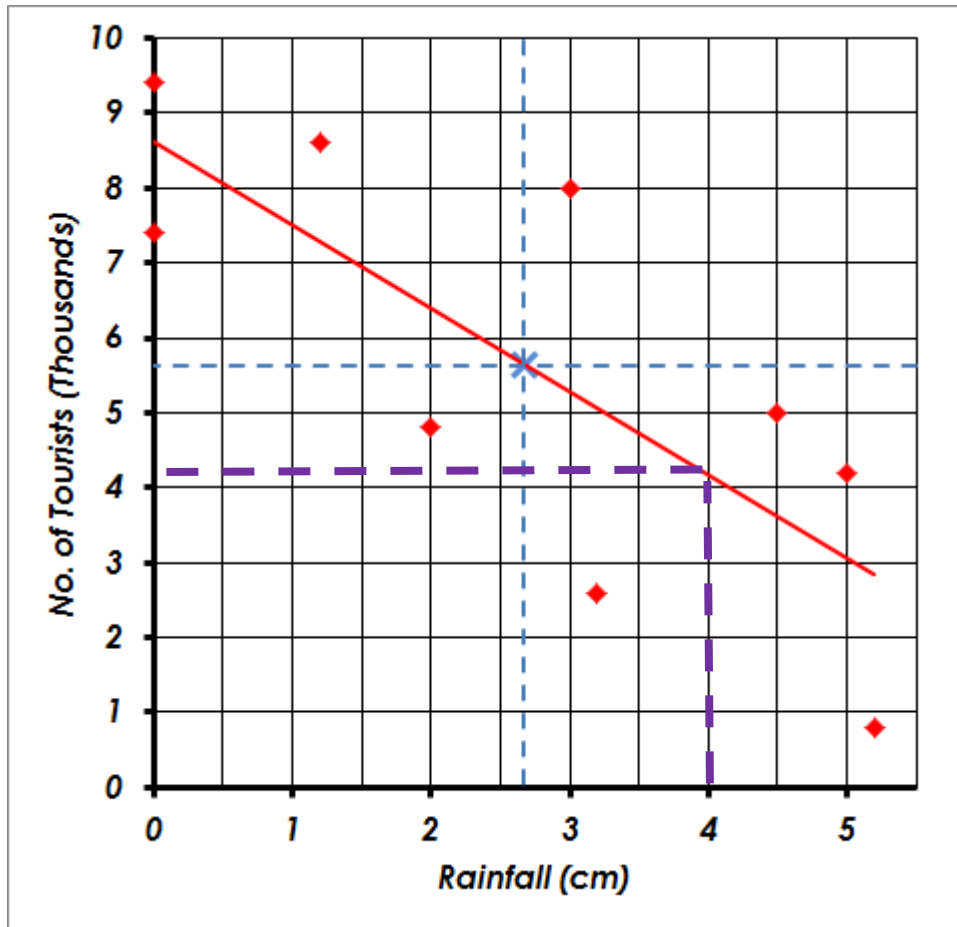
# Teaching & Learning Plan – Correlation Coefficient



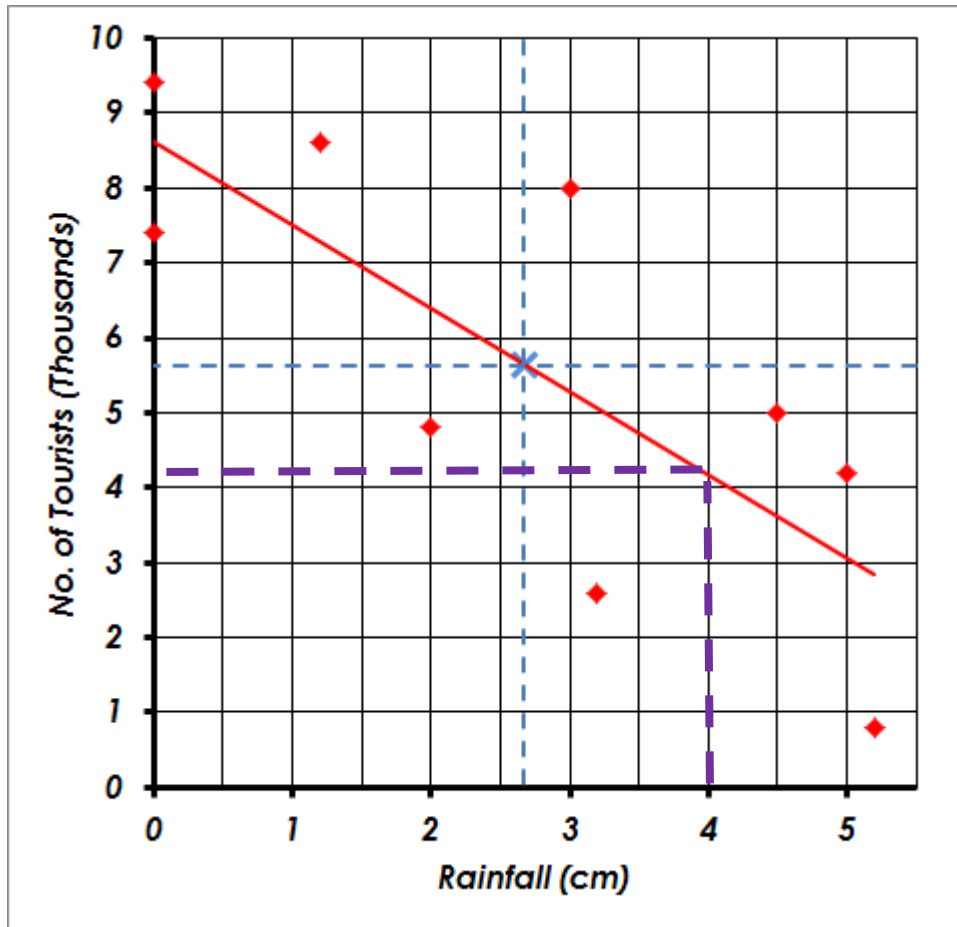
# Teaching & Learning Plan – Correlation Coefficient



# Teaching & Learning Plan – Correlation Coefficient

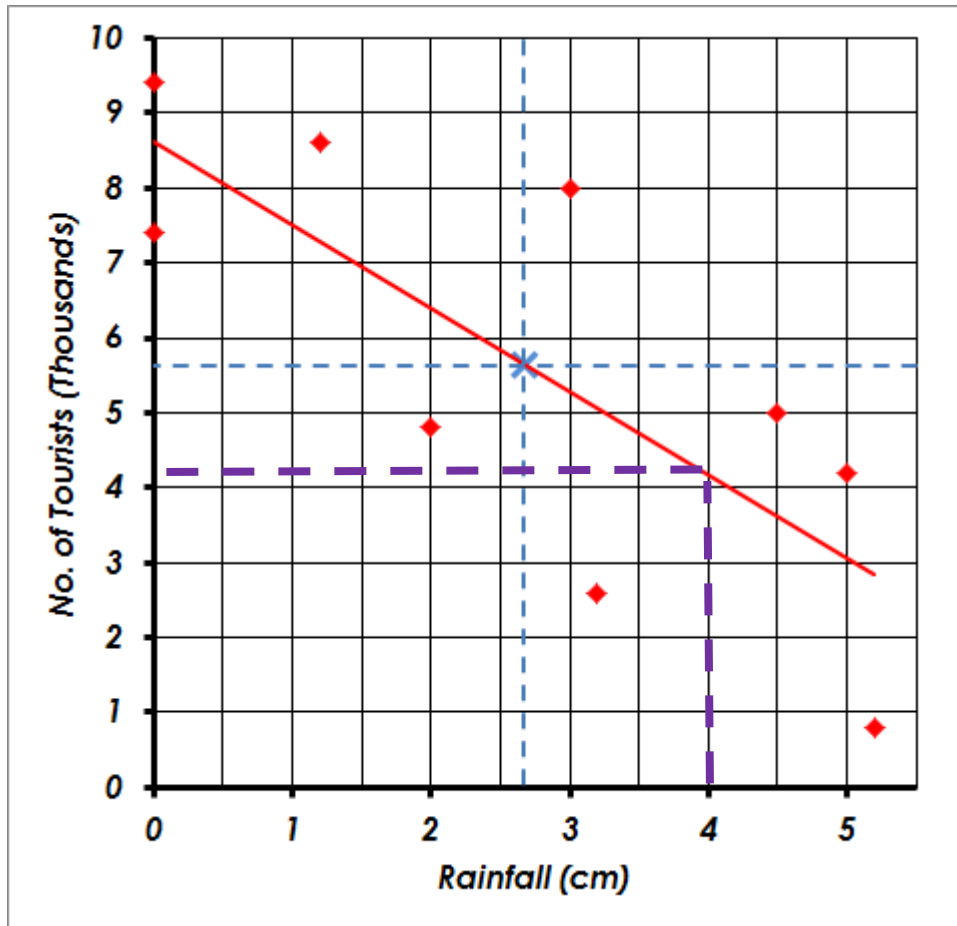


# Teaching & Learning Plan – Correlation Coefficient



(viii) 
$$m = \frac{y_2 - y_1}{x_2 - x_1}$$
$$m = \frac{3 - 8}{5 - 0.5} = -1.11$$

# Teaching & Learning Plan – Correlation Coefficient



$$\begin{aligned} \text{(x)} \quad y - y_1 &= m(x - x_1) \\ y - 8 &= -1.11(x - 0.5) \\ y - 8 &= -1.11x + 0.555 \\ y &= -1.11x + 8.56 \end{aligned}$$

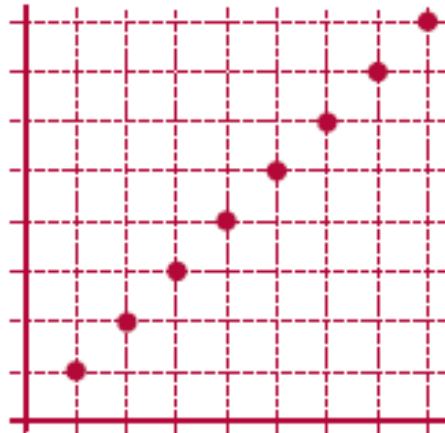
$$\begin{aligned} y &= -1.11(4) + 8.56 \\ &4.12 \text{ thousand} \end{aligned}$$

# Matching Correlations Coefficients to scatter plots

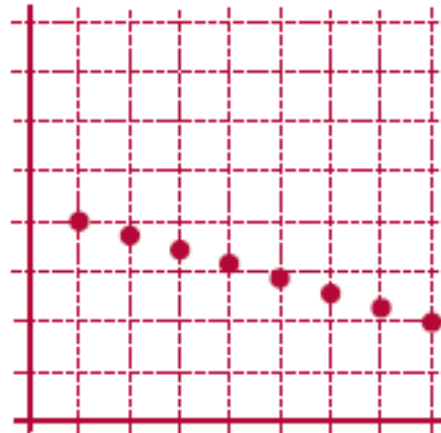
The table shows the correlations for the four graphs below. Match each graph to the correlation coefficient.

Correlation	-0.72	1	-1	0
Graph				

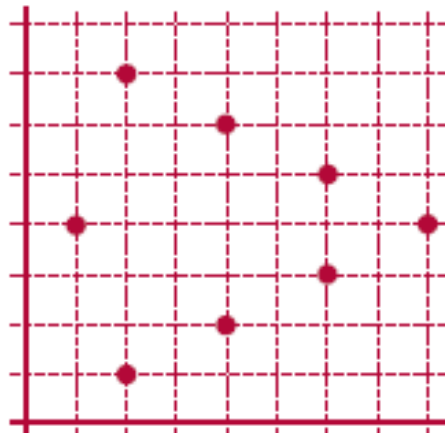
Graph A



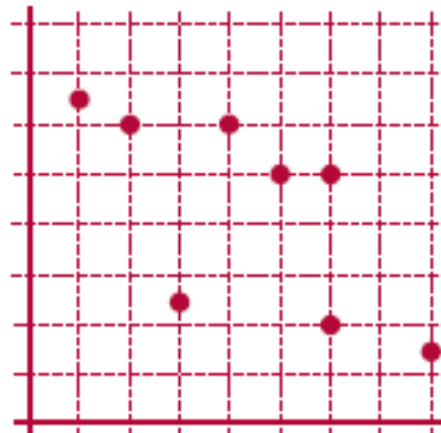
Graph B



Graph C



Graph D



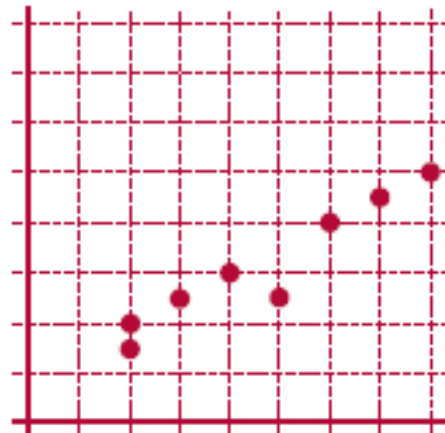


# Matching Correlation Coefficients to scatter plots

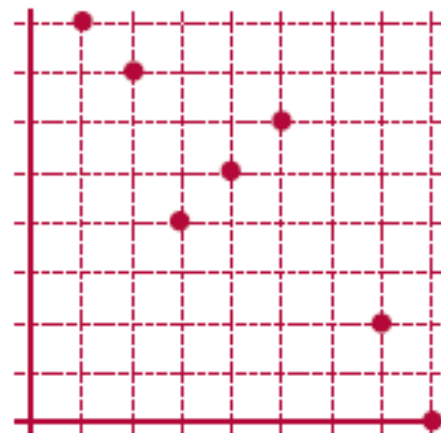
The table shows the correlations for the four graphs below. Match each graph to its correlation coefficient.

Correlation	0.72	-0.90	0.96	-0.42
Graph				

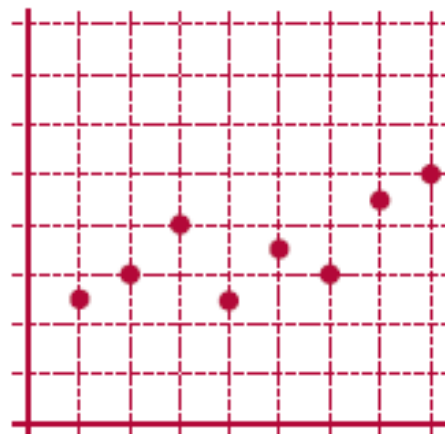
Graph A



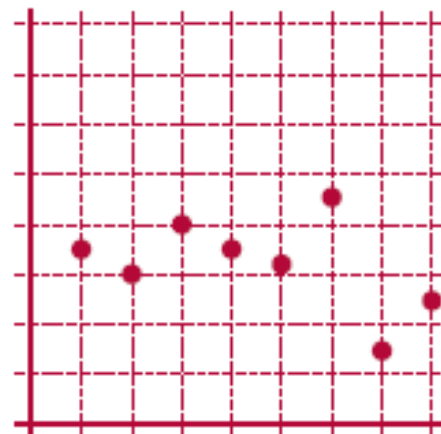
Graph B



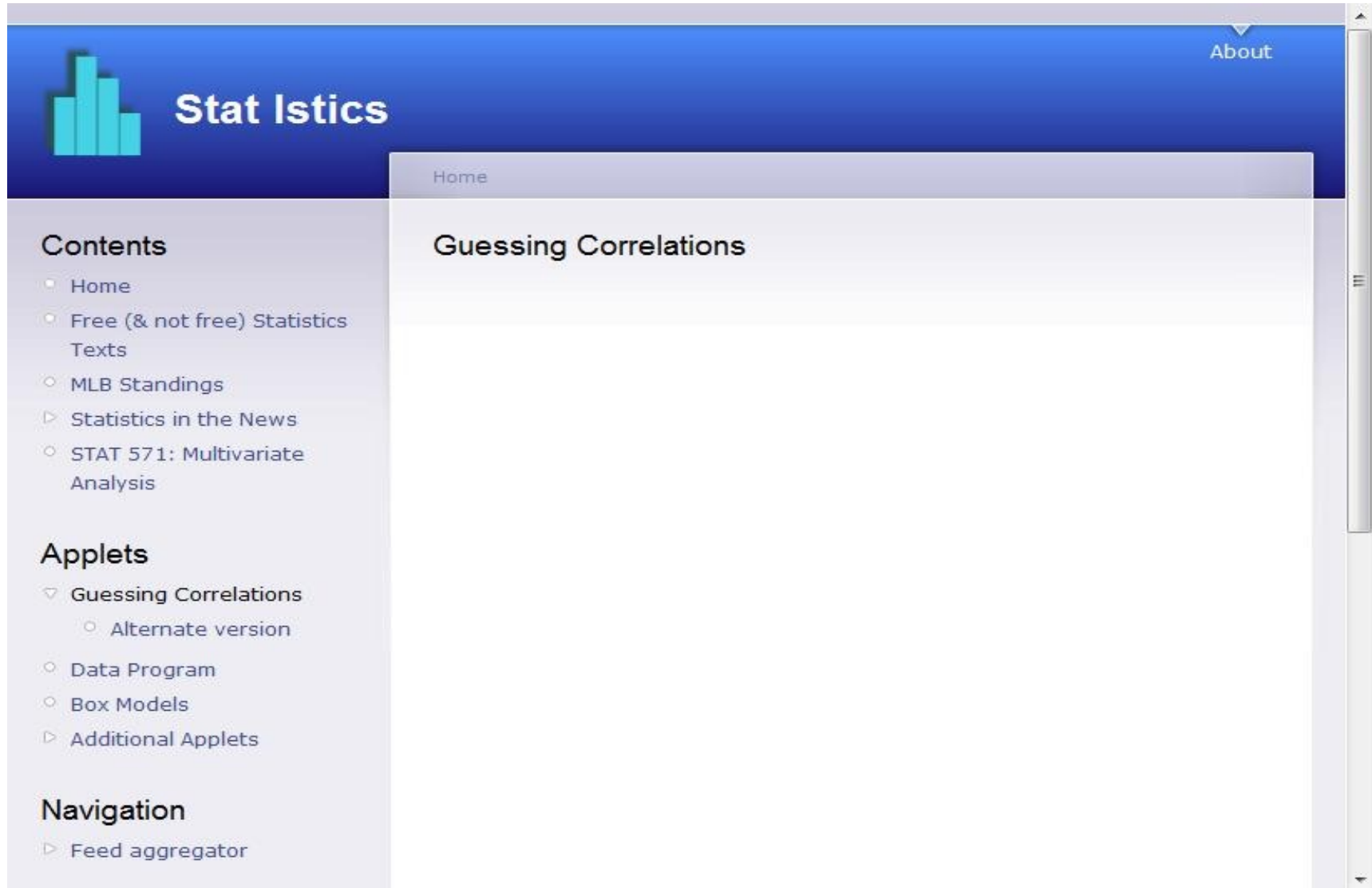
Graph C



Graph D



<http://www.istics.net/stat/Correlations>



The screenshot shows a web browser window with a blue header. On the left, there is a logo consisting of four vertical bars of increasing height, followed by the text "Stat Istics". In the top right corner of the header, there is a link labeled "About". Below the header, the main content area is titled "Home" and "Guessing Correlations". On the left side of the page, there is a sidebar with three sections: "Contents", "Applets", and "Navigation".

**Contents**

- Home
- Free (& not free) Statistics Texts
- MLB Standings
- ▷ Statistics in the News
- STAT 571: Multivariate Analysis

**Applets**

- ▽ Guessing Correlations
  - Alternate version
- Data Program
- Box Models
- ▷ Additional Applets

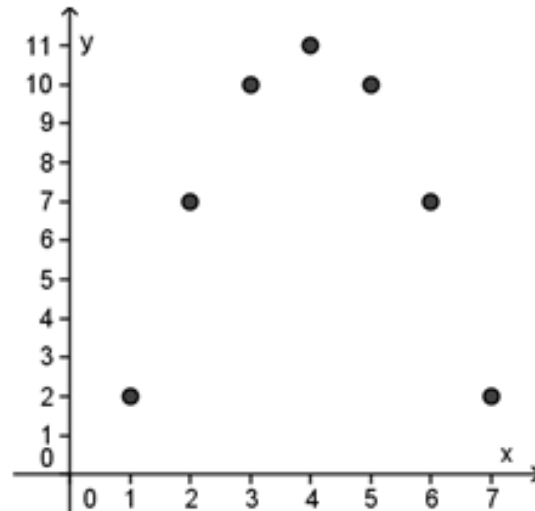
**Navigation**

- ▷ Feed aggregator

# A Question

The data given in the table below and represented in the scatter diagram are pairs of observations of the variables  $x$  and  $y$ .

$x$	1	2	3	4	5	6	7
$y$	2	7	10	11	10	7	2



- (i) From looking at the diagram would it be appropriate to work out the correlation coefficient of the data? Explain your reasoning.
- (ii) From looking at the diagram would it be appropriate to draw in the line of best fit of the data? Explain your reasoning.
- (iii) What kind of relationship, if any, do the observed data suggest exists between  $x$  and  $y$ ?