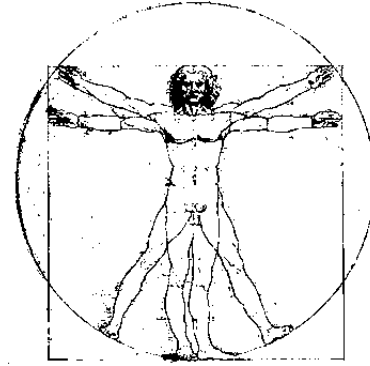**Task 1:** Check your understanding of correlation

This is a version of a famous drawing of the Vitruvian Man by
Leonardo di Vinci in 1487. In this drawing Leonardo is
representing the work of an ancient Roman architect named
Vitruvius who connected the proportions of the male figure to the
proportions used classical architecture.

(1) Vitruvius wrote that a man's arm span is equal to his
height. How does Leonardo's drawing communicate this
relationship between arm span and height?

(2) Graph the arm span and height measurements of 5
hypothetical men who fit the proportions of the
Vitruvian Man.
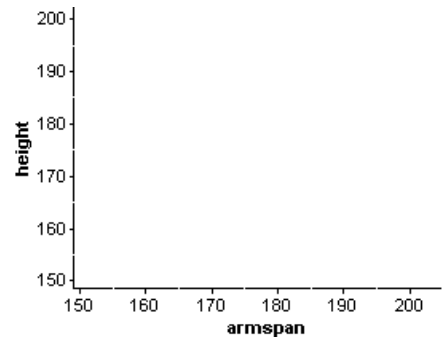
(3) Without doing any calculations, what is the correlation for your set of 5 men?
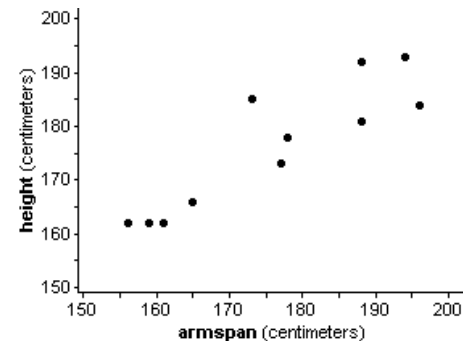
(4) Here is a scatterplot of real arm span and height measurements for a sample of 11 men. The
relationship between arm span and height is roughly linear, though these men are not perfectly
proportioned according to Vitruvius. Which of the following values for the correlation coefficient
describes the relationship in the scatterplot?

0.89   0.23   0.02   -0.45   -0.78

Task 2: Introduction to the correlation formula.

We will discuss the questions in this section together as a class. $r = \dfrac{\sum\left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right)}{n-1}$

(1) How is the formula for r related to things we have seen before?

(2) In the formula for r we see the following terms: $x,\ \bar{x},\ s_x,\ y,\ \bar{y},\ s_y,\ n-1$.

- If we are calculating the correlation for the relationship between $x = $ *arm span* and $y = $ *height* for a sample of 11 men, describe what each of these terms represents.

- For our sample of 11 men, which of these terms will be fixed in the formula (calculated once and then used as a constant) and which will have different values for the different men?

- How can you tell in the formula that r is an average of something?

Task 3: Digging into the correlation formula

Here we have arm span and height measurements for a sample of 11 men represented in a table, in dotplots and in a scatterplot. The summary statistics for both measurements are given below. The means are marked in each graph.

	armspan	height
1	161 cm	162 cm
2	196 cm	184 cm
3	177 cm	173 cm
4	188 cm	181 cm
5	159 cm	162 cm
6	178 cm	178 cm
7	194 cm	193 cm
8	188 cm	192 cm
9	173 cm	185 cm
10	165 cm	166 cm
11	156 cm	162 cm

	Mean	Standard deviation
Arm span	175.91	14.36
Height	176.18	11.90
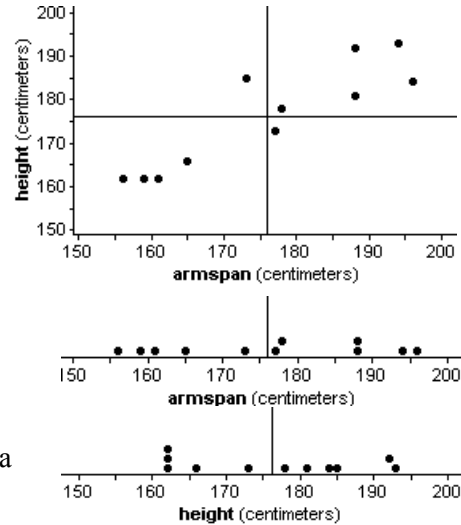
To get familiar with the formula for correlation, let's look at just a few men.

(1) Locate the male with an arm span of 188 cm and a height of 192 cm in the table and in each graph.

- This man has (*circle one*: above average, below average, average) arm span and (*circle one*: above average, below average, average) height.

- For this man $x - \bar{x}$ is (*circle one*: positive, negative, zero).

- For this man $y - \bar{y}$ is (*circle one*: positive, negative, zero).

- For this man calculate each of the following:

$$x - \bar{x} = \underline{\hspace{2cm}} \qquad \frac{x - \bar{x}}{s_x} = \underline{\hspace{2cm}} \qquad \left. \begin{array}{c} \\ \\ \end{array} \right\} \left(\frac{x - \bar{x}}{s_x} \right)\left(\frac{y - \bar{y}}{s_y} \right) = \underline{\hspace{2cm}}$$

$$y - \bar{y} = \underline{\hspace{2cm}} \qquad \frac{y - \bar{y}}{s_y} = \underline{\hspace{2cm}}$$

(2) Find a man who has <u>below average</u> arm span and <u>below average</u> height. Locate this man in all three graphs.

- For this man $\dfrac{x-\bar{x}}{s_x}$ is (*circle one*: positive, negative, zero). (Determine this without making any calculations if you can.)

- For this man $\dfrac{y-\bar{y}}{s_y}$ is (*circle one*: positive, negative, zero). (Determine this without making any calculations if you can.)

- For this man calculate each of the following:

$$x-\bar{x} = \underline{\hspace{2cm}}$$

$$y-\bar{y} = \underline{\hspace{2cm}}$$

$$\left.\begin{array}{c}\dfrac{x-\bar{x}}{s_x} = \underline{\hspace{2cm}} \\[2mm] \dfrac{y-\bar{y}}{s_y} = \underline{\hspace{2cm}}\end{array}\right\} \left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right) = \underline{\hspace{2cm}}$$

- Locate <u>all of</u> the men in the scatterplot who have a <u>below average</u> arm span and <u>below average</u> height.
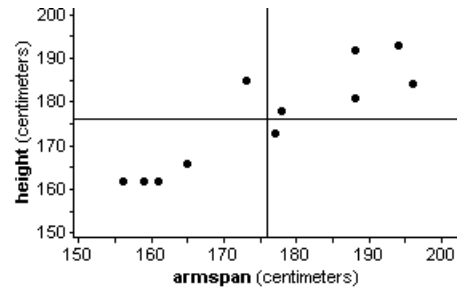
(3) Locate the men in the scatterplot who have both a <u>below average</u> arm span and an <u>above average</u> height. How many men in the sample fit this description?

For each of these men is the product $\left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right)$ positive, negative, or zero? (Determine this without making any calculations if you can.)

(4) In the scatterplot, circle all of the men for whom the product $\left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right)$ is positive.

(5) In the table, write the appropriate expression above the appropriate column in the table.

$$x, \ y, \ \frac{x-\bar{x}}{s_x}, \ \frac{y-\bar{y}}{s_y}, \ \left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$$ Note: $\frac{x-\bar{x}}{s_x}$ is called a *z-score* for x. Here x is armspan.

Complete the blanks in the table.

	armspan	zscore_armspan	height	zscore_height	product_zscores
1	161 cm	-1.04	162 cm	-1.19	1.24
2	196 cm		184 cm	0.657	
3	177 cm	0.076	173 cm	-0.267	-0.0203
4	188 cm	0.842	181 cm	0.405	0.341
5	159 cm	-1.18	162 cm	-1.19	1.4
6	178 cm	0.146	178 cm	0.153	0.0223
7	194 cm	1.26	193 cm	1.41	1.78
8	188 cm	0.842	192 cm	1.33	1.12
9	173 cm	-0.203	185 cm	0.741	-0.15
10	165 cm	-0.76	166 cm	-0.856	0.65
11	156 cm	-1.39	162 cm		

(6) Calculate the correlation by taking the sum of product of the z-scores and dividing by n-1. Double check that your answer matches your estimate in Task 1.

Task 4: Homework

(1) Is the correlation in this scatterplot positive, negative, or close to zero?

(2) In the scatterplot find a data point that fits each description:
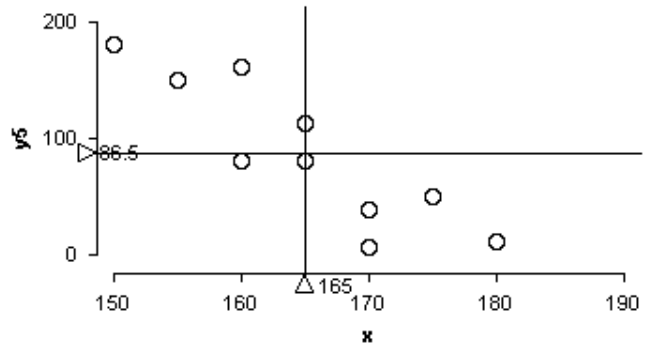
- Label a point A for which $\dfrac{x-\bar{x}}{s_x}$ is negative but $\dfrac{y-\bar{y}}{s_y}$ is positive. (How many points fit this description?)

- Label a point as B for which $\dfrac{x-\bar{x}}{s_x}$ is equal to 0 (How many points fit this description?)

- Label a point as C for which the x-value is above the mean of x and the y-value is below the mean of y (How many points fit this description?)

(3) For how many of the 10 points shown in the scatterplot is $\left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right)$ negative? For how many of the 10 points is $\left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right)$ positive? For how many of the 10 points is $\left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right)$ equal to zero?

(4) Estimate the correlation for this scatterplot.

(5) Here is the data set for the scatterplot shown above. Use technology to calculate the correlation coefficient. Check that the answer is consistent with your answer in (1), your observations about $\left(\dfrac{x-\bar{x}}{s_x}\right)\left(\dfrac{y-\bar{y}}{s_y}\right)$ in (3), and your estimate in (4).

x	y5
150	180
155	150
160	160
165	80
170	38
175	50
160	80
180	10
165	112
170	5