

## A correlation example

The table below gives data on diameter (in millimeters) and mass (in grams) for 15 peanut butter M&M's (taken from the larger data set we've been using as an example). For the diameter data distribution, the mean is  $\bar{x} = 15.68$  mm and the standard deviation is  $s_x = 0.58$  mm. For the mass data distribution, the mean is  $\bar{y} = 1.79$  gm and the standard deviation is  $s_y = 0.26$  gm. The three columns on the right give the intermediate results needed in calculation the *correlation* for these distributions.

$X=\text{diameter}$	$Y=\text{mass}$	$\frac{x_i - \bar{x}}{s_x}$	$\frac{y_i - \bar{y}}{s_y}$	$\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$
15.44	1.67	-0.40	-0.44	0.18
15.27	2.01	-0.70	0.85	-0.59
14.73	1.37	-1.62	-1.58	2.56
16.20	2.06	0.89	1.04	0.93
16.04	1.87	0.62	0.32	0.20
15.13	1.46	-0.93	-1.24	1.16
15.77	1.67	0.16	-0.44	-0.07
15.40	1.60	-0.47	-0.71	0.33
16.19	2.09	0.88	1.15	1.01
16.83	2.32	1.97	2.02	3.99
16.06	1.81	0.66	0.09	0.06
15.66	1.69	-0.03	-0.37	0.01
14.76	1.52	-1.57	-1.01	1.58
16.19	1.97	0.88	0.69	0.61
15.48	1.69			

1. On the accompanying scatterplot, draw the coordinate axes for the “natural” coordinate system.
2. Locate the point on the scatterplot for the data in the last row. Determine the quadrant this point is in with respect to the “natural” coordinate system for the scatterplot. Based on this, determine whether this point's contribution to the correlation is positive or negative.
3. Fill in the missing values in the last row to get this point's contribution to the correlation.
4. Use the given values and the value you get in the previous step to calculate the correlation:

$$r = \frac{\sum \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)}{n - 1} = \frac{0.18 - 0.59 + \cdots + 0.61 + \cdots}{15 - 1} =$$

**Scatterplot of mass vs diameter**

