Math 207
Introduction to Statistics

Regression and the Normal Curve

Suppose we have a list of points (x_1, y_1), (x_2, y_2), …, (x_n, y_n) and we draw a line (such as the regression line or perhaps some other line) to try to predict y from x. For each of the x-values x_i in our data list, we can use the line to get a predicted value $y_{\text{predicted}}$ of y. We also have the true value y_i of y. The difference

$$y_{\text{predicted}} - y_i$$

is the *residual error* at x_i. The RMS (root-mean-squared) size of these errors is:

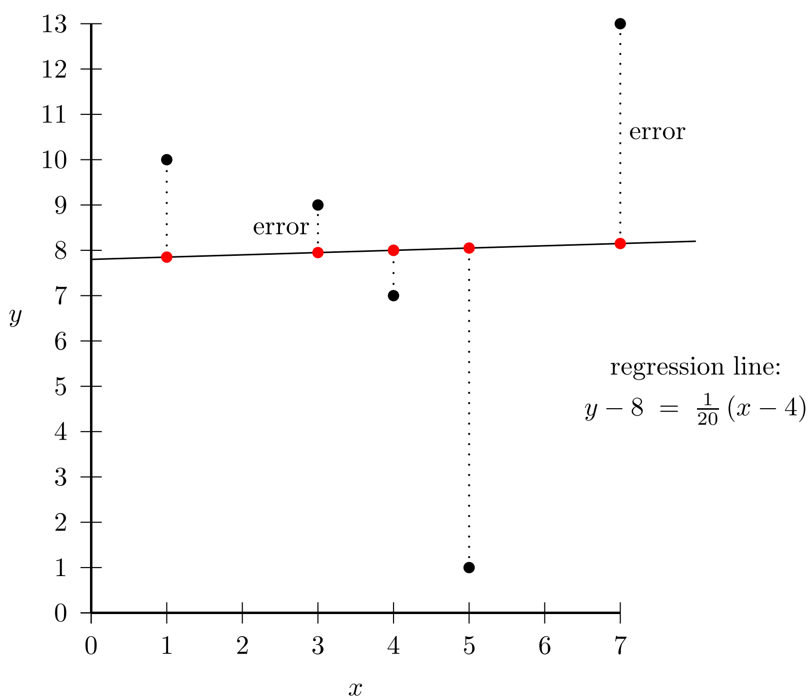
$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{residual error}_i)^2}$$

It is a measure of the total error of the line that we are using to fit the data. The regression line is the line that minimizes this error. It is the *best fit* line. Its RMS can be computed using:

$$\text{RMS}_{\text{reg}} = \text{sd}_y\sqrt{1-r^2}$$

where r is the correlation and sd_y is the standard deviation of the y-values in our data set. Notice that:

- For a fixed value of r, RMS increases with sd_y.
- If $r = 1$ or $r = -1$, the RMS is zero (since the points fall exactly on a line).
- If $r = 0$, then RMS $= \text{sd}_y$.



1. Compute the regression line RMS given sd_y and r

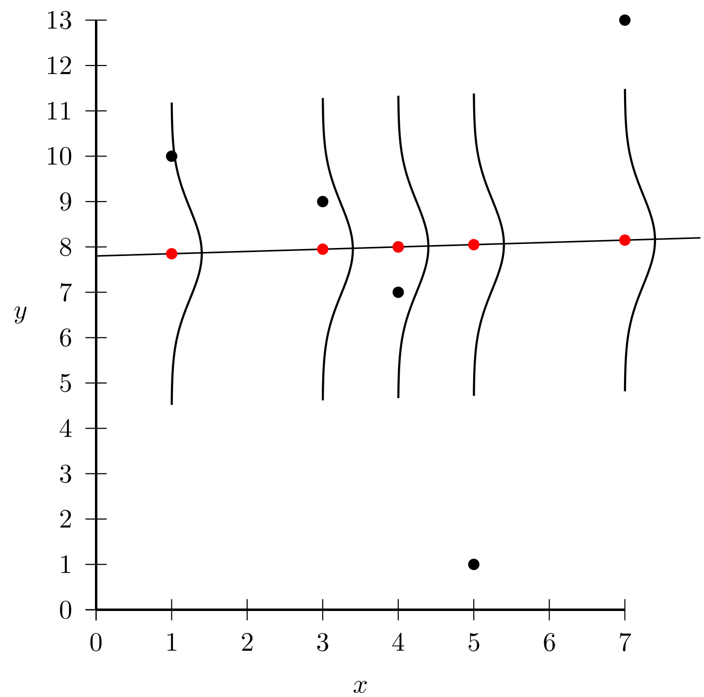
   a) $\text{sd}_y = 8$ and $r = \frac{\sqrt{3}}{2}$

   b) $\text{sd}_y = 1$ and $r = \frac{\sqrt{3}}{2}$

   c) $\text{sd}_y = 1$ and $r = -\frac{\sqrt{3}}{2}$

   d) $\text{sd}_y = 5$ and $r = 0.1$

One of the assumptions of the regression model is that for any given x, the histogram of the possible y values is a normal curve. Its mean is the predicted y value (the point on the line). Its sd is the RMS. The figure below shows a normal curve moving along with the regression line. The y-value of each red dot is the mean of the corresponding normal curve.



2. Use the information given to find the mean and sd of the moving normal curve. If $x = 3$, find a range for 68% of the y-values.

a) $y - 8 \;=\; 2\,(x-4)$, $\mathrm{sd}_y = 8$, $r = \frac{\sqrt{3}}{2}$.

b) $y - 2 \;=\; 5\,(x-1)$, $\mathrm{sd}_y = 1$, $r = \frac{\sqrt{3}}{2}$.

c) $y - 8 \;=\; -2\,(x-2)$, $\mathrm{sd}_y = 1$, $r = -\frac{\sqrt{3}}{2}$.