Computing the Median-Median Line for a Data Set

One way to estimate a line of best fit for data is to construct a median-median line.

Procedure:

1. Make a scatterplot of the points and continue if the association is approximately linear.
2. Separate the data points into three groups of approximately equal size based on the x-values of the points.
3. Find the *summary point* for each group based on the median x-value and median y-value.
4. Find the equation of the line (Line $L$) through the summary points of the outer groups.
5. Slide $L$ one-third of the way to the middle summary point by doing the following:
   a. Find the y-coordinate of the point on $L$ with the same x-coordinate as the middle summary point.
   b. Find the vertical distance between the middle summary point and the line by subtracting y-values.
   c. Find the coordinates of the point $P$ one-third of the way from the line $L$ to the middle summary point.
6. Find the equation of the line through the point $P$ that is parallel to line $L$.

Example:

Find the median-median line for the following data about the “agreeableness” and “behavioral problems” of 10 participants in a psychological study of 104 adolescents (all data points are at integer coordinates).
Line of Best Fit Homework

Respond to the questions on separate paper, writing complete sentences to explain your solutions. Credit is awarded for correctness of mathematics (10 points), correctness and completeness of explanations (15 points) and neatness and grammar (5 points). You may use any resource but are expected to turn in your own individual work.


<table>
<thead>
<tr>
<th># of Year</th>
<th>0</th>
<th>6</th>
<th>11</th>
<th>13</th>
<th>17</th>
<th>18</th>
<th>24</th>
<th>25</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly Wage</td>
<td>0.75</td>
<td>1.00</td>
<td>1.15</td>
<td>1.25</td>
<td>1.40</td>
<td>1.60</td>
<td>2.00</td>
<td>2.10</td>
<td>2.30</td>
</tr>
</tbody>
</table>

Tasks:

1. Create a scatter plot of this set of data. Be sure to pick a good scale and label your axes.

2. Are the variables positively or negatively associated? Is the linear association strong or weak?

3. Find the median-median line for the data by hand. Show your work.

4. Calculate the least-squares line for the data using a calculator or computer.

5. Carefully draw both the median-median line and the least-squares line onto your data set. It may help to generate a couple points on each line and use a ruler.

6. Find the residual using the least-square line when the # of years is 24. Explain the meaning of this value.

7. Explain why the equations of the two lines (median-median and least-square line) are different. Which line do you believe would be more successful in predicting hourly wages between 1950 and 1980? Justify your answer.

8. Use each line to estimate the hourly wage in 2011. Compare your answers to the actual federal minimum wage and explain the meaning of any residuals.