

The Wald Line

The Wald line is a linear regression line that can serve as a compromise between the Least-Squares model and the Median-Median model. The Wald line is named after Abraham Wald. As always, if the data shows a strong association, go with the Least-Squares model.

Consider an even number of points $(x_1, y_1), (x_2, y_2), \dots, (x_{2n}, y_{2n})$ whose scatter plot shows “reasonable” association. Sort the points on their x -coordinates. Now divide the points into the “lower” x points and the “higher” x points. For each of these groups, find the mean point:

$$(\overline{x}_L, \overline{y}_L) \quad \text{and} \quad (\overline{x}_H, \overline{y}_H).$$

Now determine the line through these two points. That is the Wald line.

The Wald line has the property that it goes through the mean point $(\overline{x}, \overline{y})$ for the entire set of data. This is the same point through which the Least-Squares line passes.

If you have an odd number of data points then you only need add $(\overline{x}, \overline{y})$ as a point. This now makes the number of points even. Since the Wald line goes through this point, adding it to the data set does not change the slope or intercept.