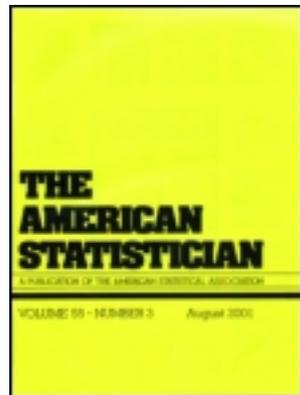


This article was downloaded by: [Moskow State Univ Bibliote]

On: 30 August 2013, At: 11:03

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41  
Mortimer Street, London W1T 3JH, UK



## The American Statistician

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/utas20>

### Parallel Lines in Residual Plots

Shayle R. Searle <sup>a</sup>

<sup>a</sup> Biometrics Unit, New York State College of Agriculture and Life Sciences, Cornell University,  
Ithaca, New York, 14850, USA

Published online: 27 Feb 2012.

To cite this article: Shayle R. Searle (1988) Parallel Lines in Residual Plots, The American Statistician, 42:3, 211-211

To link to this article: <http://dx.doi.org/10.1080/00031305.1988.10475569>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# Parallel Lines in Residual Plots

SHAYLE R. SEARLE\*

Attention is drawn to a little-known feature of residuals plotted against predicted values: When observed values of  $y$  are repeated in the data, then the plot exhibits a series of parallel lines having slope  $-1$ .

It is common practice after fitting a model to data to make a plot of residuals, of  $y - \hat{y}$ , say, against predicted values  $\hat{y}$ . The utility of doing so is promoted and described in many textbooks (e.g., Draper and Smith 1981, p. 147). Figure 1 shows an example of such a plot. It is for data that consist of 80 records of an observer's visual assessment of the percentage of a plot of potato plants infected with a leaf disease. The 80 plots consist of 4 plots on each of 20 varieties of potatoes.

At first sight the startling feature of Figure 1 is the occurrence of parallel lines. But, in fact, this is not startling. Pathologists eyeballing the percentage of a plot of potato plants that has leaf disease record their assessment as a code, A, B, C, . . . , and later convert that code to a preassigned percentage of disease, in these data to just the six percentages 10%, 25%, 50%, 75%, 90%, and 97%. Furthermore, these data had 6, 9, 8, 14, 18, and 25 plots recorded with those percentages, respectively. And the six parallel lines of Figure 1 correspond precisely to these six groups of data. (Points in the figure labeled x, 2, and 3 represent 1, 2, and 3 observations, respectively, and those with a thick-looking symbol represent two or more sets of closely adjacent points.)

The reason for this characteristic is very simple. The plot is of  $y - \hat{y}$  against  $\hat{y}$ . Consider that plot just for  $y$  of some particular value,  $c$  say. Then it is a plot of conditional variables, of  $[(y - \hat{y})|y = c]$  against  $(\hat{y}|y = c)$ , that is, of  $c - (\hat{y}|y = c)$  against  $(\hat{y}|y = c)$ . Clearly, this is a straight line, one for each  $c$ , every such line having a slope of  $-1$ . Although this result is so simple, it seems to have largely evaded the literature, both books and journals; for example, Draper and Smith (1981) do not mention it. One place where it *has* been seen is in McCullagh and Nelder (1983, p. 216), as kindly pointed out by H. V. Henderson.

Five general properties of this feature of plots of residuals against predicted values are worthy of note:

1. When analyzing data that include a new observation  $y = c$ , that new observation will give rise to a point on the line already established by the  $y = c$  data without the additional observation.
2. Even with data having many different observed values, these parallel lines do exist, although they may not be readily apparent. A vivid illustration of this would be with truncated data, as suggested by M. P. Meredith: If several observations were truncated to the same value and most other ob-

served values each occurred only once, then the straight line corresponding to the truncated value would be patently apparent in the residual-predicted plot. In general, though, it is the presence of only a few and clearly different values that makes the lines very apparent; for example, Figure 1 has only six different values. A special case of this would be for  $y$  being a dichotomous variable [e.g., the data of Neter, Wasserman, and Kutner (1984, p. 358)]: There would be just two distinct lines in the residual-predicted plot.

3. Even when a data value occurs only once, it implicitly gives rise to a line that would be evident were there more than one datum with that value.

4. Parallel lines of this nature occur no matter what model is fitted to  $y$ , and correspondingly no matter how  $\hat{y}$  is calculated, be it based on linear or nonlinear estimation. So long as  $y - \hat{y}$  is plotted against  $\hat{y}$ , the parallel lines will exist either explicitly, for repeated  $y$ -values, or implicitly, for single  $y$ -values.

5. The slope of all such lines is always  $-1$ , and the line corresponding to  $y = c$  crosses the line  $y - \hat{y} = 0$  at  $\hat{y} = c$ .

This is paper BU-515 in the Biometrics Unit, Cornell University.

[Received October 1987.]

## REFERENCES

- Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis* (2nd ed.), New York; John Wiley.
- McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman & Hall.
- Neter, J., Wasserman, W., and Kutner, M. (1984), *Applied Linear Regression Models*, Homewood, IL: Richard D. Irwin.

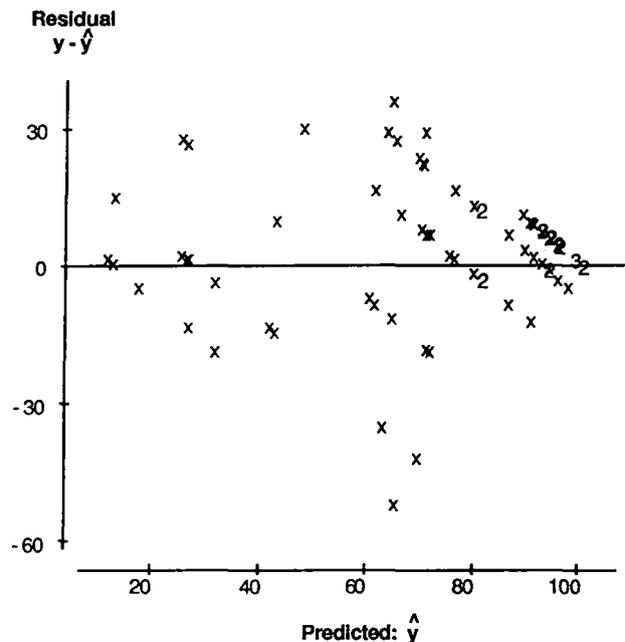


Figure 1. Plot of  $y - \hat{y}$  Against  $\hat{y}$ .

\*Shayle R. Searle is Professor of Biological Statistics, Biometrics Unit, New York State College of Agriculture and Life Sciences, Cornell University, Ithaca, New York 14850.