# Response Plots for Experimental Design

David J. Olive*

Southern Illinois University

April 22, 2008

**Abstract**

A response plot of the fitted values versus the response simultaneously displays the fitted values, response and residuals. The plot is also used to visualize the model and to check whether the model is reasonable. The plot can be used to select a response transformation $Y = t(Z)$ since the plotted points will scatter about a line with unit slope and zero intercept if the transformation is reasonable.

**KEY WORDS: Goodness of Fit, Outliers, Response Transformations.**

*David J. Olive is Associate Professor, Department of Mathematics, Southern Illinois University, Mailcode 4408, Carbondale, IL 62901-4408, USA. E-mail address: dolive@math.siu.edu. This research was supported by NSF grant DMS 0600933.

# 1 INTRODUCTION

A model for an experimental design is $Y_i = E(Y_i) + e_i$ for $i = 1, ..., n$ where the error $e_i = Y_i - E(Y_i)$ and $E(Y_i) \equiv E(Y_i|\boldsymbol{x}_i)$ is the expected value of the response $Y_i$ for a given vector of predictors $\boldsymbol{x}_i$. Many models can be fit with least squares (OLS) and have the form

$$Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$$

for $i = 1, \ldots, n$. Often $x_{i,1} \equiv 1$ for all $i$. In matrix notation, these $n$ equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ design matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. If the fitted values are $\hat{Y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$, then $Y_i = \hat{Y}_i + r_i$ where the residuals $r_i = Y_i - \hat{Y}_i$.

A response plot is a plot of the fitted values on the horizontal axis versus the response on the vertical axis. Ignoring the residuals gives the line $Y = \hat{Y}$, so the plotted points scatter about the identity line with unit slope and zero intercept. Since the vertical deviations from the identity line are the residuals $r_i = Y_i - \hat{Y}_i$, the response plot simultaneously shows the response, fitted values and residuals. The response plot should be used as well as the residual plot of $\hat{Y}_i$ versus the residuals $r_i$.

If the residual degrees of freedom is not too small, these plots are useful for visualizing the model and for checking whether the model is reasonable. Section 2 gives some examples and Section 3 shows how to use response plots to select a response transformation.

# 2 Examples

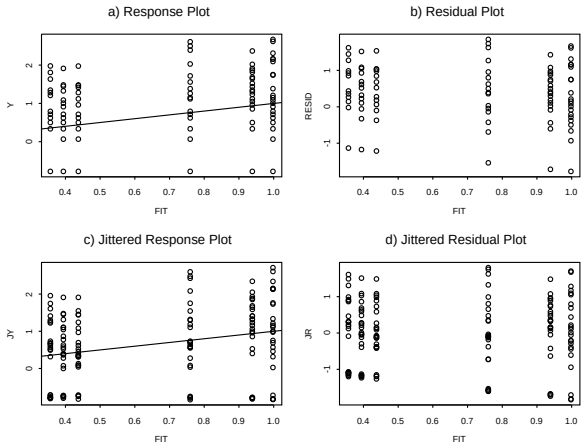


Figure 1: Plots for Crab Data

**Example 1.** Kuehl (1994, p. 128) gives data for counts of hermit crabs on 25 different transects in each of six different coastline habitats. Let $C$ be the count. Then the response variable $Y = \log_{10}(C + 1/6)$. Although the counts $C$ varied greatly, each habitat had several counts of 0 and often there were several counts of 1, 2 or 3. Hence $Y$ is not a continuous random variable. The one way Anova model $Y_{ij} = \mu_i + e_{ij} = \eta + \tau_i + e_{ij}$ was fit for $i = 1, ..., 6$ with $n_i = 25$, and $j = 1, ..., n_i$. Each of the six habitats was a level. Figure 1a and b shows the response plot and residual plot. There are 6 dot plots in each plot. Because several of the smallest values in each plot are identical, it does not always look like the identity line is passing through the six sample means $\overline{Y}_{i0}$ for $i = 1, ..., 6$. In particular, examine the dot plot for the smallest mean (look at the 25 dots furthest to the left that fall on the vertical line FIT $\approx 0.36$). Random noise (jitter) has been added to the response and residuals in Figure 1c and d. Now it is easier to compare the six dot plots. They seem to have roughly the same spread.

A design matrix for this model consists of indicator variables for each treatment, and values of $\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$ other than the observed six values $\overline{Y}_{i0}$ may not have much meaning. The plots still contain a great deal of information. The response plot can be used to explain the model, check that the sample from each population (treatment) has roughly the same shape and spread, and to see which populations have similar means. Since the response plot closely resembles the residual plot in Figure 1, there may not be much difference in the six populations. Linearity seems reasonable since the samples scatter about the identity line. The residual plot makes the comparison of "similar shape" and "spread" easier.

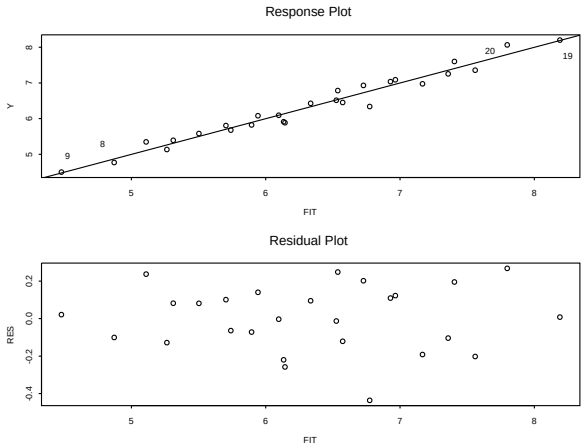


Figure 2: Plots for Textile Data

In industry, the levels are often fixed values of a continuous variable such as temperature. Then values of $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{\beta}}$ are interesting even for unobserved values of $\boldsymbol{x}$ if $\boldsymbol{x}$ is in the factor space. That is, interpolation is informative although extrapolation is dangerous. If $h_{\boldsymbol{x}} = \boldsymbol{x}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}$, then a rule of thumb is that $\boldsymbol{x}$ is in the factor space and interpolation is being done if $h_{\boldsymbol{x}} \leq \max(h_{\boldsymbol{x}_1}, ..., h_{\boldsymbol{x}_n})$.

For testing to be informative and for the response and residual plots to be informative, the residual degrees of freedom should not be too small. Large interactions can be omitted from the OLS design matrix, perhaps after making a normal plot of the effects. See Box, Hunter and Hunter (2005, p. 203).

**Example 2.** In their pioneering paper on response transformations, Box and Cox (1964) analyze data from a $3^3$ experiment on the behavior of worsted yarn under cycles of repeated loadings. The response $Y = \log(Z)$ where $Z$ is the number of cycles to failure and the three predictors are the length, amplitude and load. To make Figure 2, a constant was used in the design matrix, but no interactions. For this data set, there is one value of the response for each of the 27 treatment level combinations.

Figure 2 shows that linearity with constant variance is reasonable, and that the signal to noise ratio is high. To use the response plot to visualize the conditional distribution of $Y|\boldsymbol{x}^T\boldsymbol{\beta}$, use the fact that the fitted values $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{\beta}}$. For example, suppose that log(cycles to failure) given fit = 6 is of interest. Mentally examine the plot about a narrow vertical strip about fit = 6, perhaps from 5.75 to 6.25. The cases in the narrow strip have a mean close to 6 since they fall close to the identity line. Similarly, when the fit = $w$ for $w$ between 4.5 and 8.5, the cases have log(cycles to failure) near $w$, on average. Notice that cases 19 and 20 had the largest time until failure. These cases correspond to wool specimens with long length, short amplitude of loading cycle and low load. Cases 8 and 9 had the shortest times with low length, high amplitude and high load.

# 3 Response Transformations

The applicability of an experimental design model can be expanded by allowing response transformations. An important class of *response transformation models* adds an additional unknown transformation parameter $\lambda_o$, such that

$$Y_i = t_{\lambda_o}(Z_i) \equiv Z_i^{(\lambda_o)} = E(Y_i) + e_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i.$$

If $\lambda_o$ was known, then $Y_i = t_{\lambda_o}(Z_i)$ would follow the experimental design model.

Two families of transformations are frequently used. Assume that **all** of the values of the "response" $Z_i$ are **positive**. A power transformation has the form $Y = t_\lambda(Z) = Z^\lambda$ for $\lambda \neq 0$ and $Y = t_0(Z) = \log(Z)$ for $\lambda = 0$ where $\lambda \in \Lambda_L = \{-1, -1/2, 0, 1/2, 1\}$. The *modified power transformation family*

$$Y_i = t_\lambda(Z_i) \equiv Z_i^{(\lambda)} = \frac{Z_i^\lambda - 1}{\lambda}$$

for $\lambda \neq 0$ and $t_0(Z_i) = \log(Z_i)$ for $\lambda = 0$ where $\lambda \in \Lambda_L$. See Box, Hunter and Hunter (2005, p. 321).

There are several reasons to use a coarse grid $\Lambda_L$ of powers. First, several of the powers correspond to simple transformations such as the log, square root, and reciprocal. These powers are easier to interpret than $\lambda = .28$, for example. Secondly, if the estimator $\hat{\lambda}_n$ can only take values in $\Lambda_L$, then sometimes $\hat{\lambda}_n$ will converge in probability to $\lambda^* \in \Lambda_L$. Thirdly, Tukey (1957) showed that neighboring modified power transformations are often very similar, so restricting the possible powers to a coarse grid is reasonable.

Box and Cox (1964) give a numerical method for selecting the response transformation for the modified power transformations. Although the method gives a point estimator

$\hat{\lambda}_o$, often an interval of "reasonable values" is generated (either graphically or using a profile likelihood to make a confidence interval), and $\hat{\lambda} \in \Lambda_L$ is used if it is also in the interval.

A graphical method for response transformations computes the fitted values $\hat{W}_i$ from the experimental design model using $W_i = t_\lambda(Z_i)$ as the "response" for each of the five values of $\lambda \in \Lambda_L$. The plotted points follow the identity line in a (roughly) evenly populated band if the model is reasonable for $(\hat{W}, W)$. If more than one value of $\lambda \in \Lambda_L$ gives a linear plot, consult subject matter experts and use the simplest or most reasonable transformation. Olive (2004) gives a similar method for linear models, and alternative methods are given in Cook and Olive (2002) and Box and Fung (1995).

After selecting the transformation, the usual checks should be made. A variant of the method would plot the residual plot or both the response and the residual plot for each of the five values of $\lambda$. Residual plots are also useful, but they no not distinguish between nonlinear monotone relationships and nonmonotone relationships. See Fox (1991, p. 55).

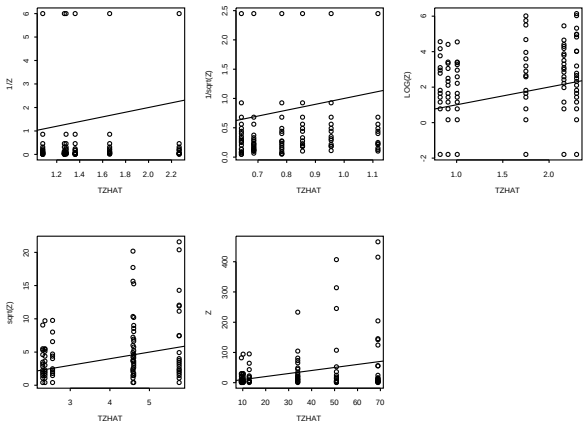


Figure 3: Transformation Plots for Crab Data

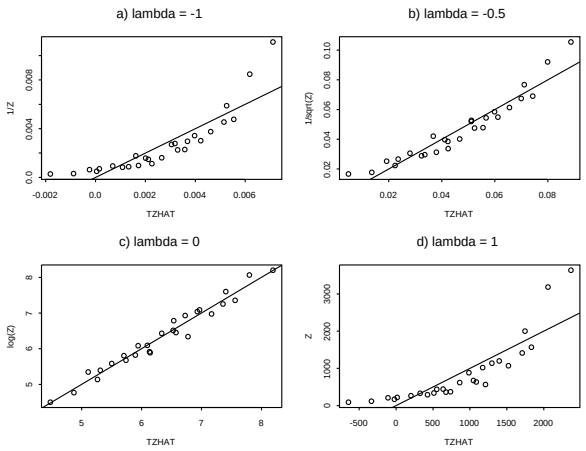


Figure 4: Transformation Plots for Textile Data

In the two following examples, the plots show $t_\lambda(Z)$ on the vertical axis. The label "TZHAT" of the horizontal axis are the fitted values that result from using $t_\lambda(Z)$ as the "response" in the software.

**Example 1 continued.** Following Kuehl (1994, p. 128), let $C$ be the count of crabs and let the "response" $Z = C + 1/6$. Figure 3 shows the five *transformation plots*. The transformation $\log(Z)$ results in dot plots that have roughly the same shape and spread. The transformations $1/Z$ and $1/\sqrt{Z}$ do not handle the 0 counts well, while the transformations $\sqrt{Z}$ and $Z$ have variance that increases with the mean.

**Example 2 continued.** For the textile data, $Z$ = number of cycles until failure. Figure 4 shows four of the five transformation plots. The plotted points curve away from the identity line in three of the four plots. The plotted points for the log transformation follow the identity line with roughly constant variance.

# 4 Summary

The analysis of the response, not that of the residuals, is of primary importance. The response plot can be used to analyze the response in the background of the fitted model. For linear models such as experimental designs, the estimated mean function is the identity line and should be added as a visual aid.

Assume that the residual degrees of freedom are large enough for testing. Then the response and residual plots contain much information. Linearity and constant variance may be reasonable if the plotted points scatter about the identity line in a (roughly) evenly populated band. Then the residuals should scatter about the $r = 0$ line in an evenly populated band. It is easier to check linearity with the response plot and constant variance with the residual plot. Curvature is often easier to see in a residual plot, but the response plot can be used to check whether the curvature is monotone or not. The response plot is more effective for determining whether the signal to noise ratio is strong or weak, and for detecting outliers, influential cases or a critical mix.

Transformation plots of $\hat{W}$ versus $W = t(Z)$ can be used to assess the success of a transformation or used to choose a transformation.

The response plots and transformation plots are simple to make, and useful for explaining the experimental design (linear model) to clients and students.

Experts in experimental design should be able to find many more applications of fitted values and the response. For example, suppose there are three response variables, and three fits using $(\boldsymbol{x}_i, Y_{ij})$ are found for $j = 1, 2, 3$ and $i = 1, ..., n$. Then a scatterplot matrix of the three variables and the three fits may be useful.

# 5 References


Box, G.E.P., and Cox, D.R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society, B,* 26, 211-246.

Box, G.E.P., and Fung, C. (1995), "The Importance of Data Transformation in Designed Experiments for Life Testing," *Quality Engineering*, 7, 625-638.

Box, G.E.P, Hunter, J.S., and Hunter, W.G. (2005), *Statistics for Experimenters*, 2nd ed., John Wiley and Sons, NY.

Cook, R.D., and Olive, D.J. (2001), "A Note on Visualizing Response Transformations in Regression," *Technometrics,* 43, 443-449.

Fox, J. (1991), *Regression Diagnostics*, Sage Publications, Newberry Park, CA.

Kuehl, R.O. (1994), *Statistical Principles of Research Design and Analysis*, Duxbury Press, Belmont, CA.

Olive, D.J. (2004), "Visualizing 1D Regression," in *Theory and Applications of Recent Robust Methods*, eds. Hubert, M., Pison, G., Struyf, A., and Van Aelst, S., Series: Statistics for Industry and Technology, Birkhäuser, Basel, Switzerland, 221-233.

Tukey, J.W. (1957), "Comparative Anatomy of Transformations," *Annals of Mathematical Statistics,* 28, 602-632.