

Simple Linear Regression Model Assessment & Refinement Examples

In the following examples, you will use residual analysis to assess the adequacy of various regression models.

1: The data for this example consists of lengths of a spring subjected to loads of varying weight. According to Hook's Law there should be a linear relationship between length and weight. Taking into account errors in measurement this implies the simple linear regression model

$$\text{length}_i = \beta_0 + \beta_1 \text{weight}_i + \epsilon_i$$

Note that β_1 is the spring constant. Open the spring length vs weight data set at the bottom of the course website and analyze it as follows:

i. Create a scatterplot of the data using **Graph -> Plot**. Does the relationship look, at least approximately, linear?

ii. Compute the least squares estimates and the residuals by

1. Accessing the regression menu via

Stat -> Regression -> Regression -> Fit Regression Model ...

2. Selecting **length** for the response (y) variable and **weight** for the predictor (x) variable

3. Clicking the *Storage* option then checking *Residuals*.

You should see a new column labeled RESI1.

iii. What are the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 ?

iv. Manually compute the residual e_1 for the first observation where

$$\begin{aligned} e_1 &= y_i - \hat{y}_i \\ &= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \end{aligned}$$

v. Plot the residuals versus hydrocarbon level (x) using **Graph -> Plot**. What do you conclude with respect to residual properties I and II?

vi. Can you use the residuals to check the assumption that the random errors are independent?

vii. Can you assess the normality of the random errors using the residuals? If so, apply **Stat -> Basic Statistics -> Normality Test** to the residuals. What do you conclude at $\alpha = 0.05$.

(*A Phat application*) The most accurate method for measuring intraabdominal fat requires a CT which is expensive and exposes the patient to x-ray radiation. Despres, et al, investigated the possibility of estimating intraabdominal fat area using a person's waist size, clearly a cheaper, safer measurement. They collected the waistsize (x) and the CT-measured intraabdominal fat area (y) of 109 randomly selected men between the ages of 18 and 42 years. They then investigated predicting intraabdominal fat using the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Construct a model for this data as follows:

i. Open the data set for this experiment:

<http://www.rose-hulman.edu/~inlow/waistdeepfat.MTW>

- ii. Fit the linear model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ and compute the residuals using Minitab. To the extent that you can, check and comment on the four random error assumptions using appropriate plots and procedures. Comment on the fifth assumption that the predictor variable is measured with no or negligible error.
- iii. Does your model satisfy the regression assumptions? If not, modify your model by using transformations, adding higher order terms, etc., until you devise a model which meets the assumptions, if possible.