

Checking Assumptions in Simple Linear Regression

Recall: All observations come in pairs (X, Y) e.g. $(X=\text{student weight}, Y=\text{student height})$

For any given value X of the explanatory variable, there are many possible values of Y that could potentially be observed & some Y values are more likely to occur or to be observed than others

(e.g., even if all students in a group have the exact same weight X , they may vary in their heights (Y -values possible) and some heights (Y -values) are more typical than others for students with a given weight X)

That is, potential Y -values for a given X have a distribution which satisfies

1. All potential Y values are all independent (though some paired observations may have same X)
2. Given an X , the mean/typical value of response variable Y : $\mu(Y | X) = \beta_0 + \beta_1 X$
3. Given an X , distribution of all possible Y -values is normal with center $\mu(Y | X)$
4. Given an X , spread/standard deviation of Y -values is $\sigma(Y | X) = \sigma$

An **alternative way** to view the same assumptions is that:

If we sample n pairs of data $(X_1, Y_1), \dots, (X_n, Y_n)$ then for each X_i

the corresponding $Y_i = \mu(Y | X_i) + e_i$

$$= \beta_0 + \beta_1 X_i + e_i$$

where e_1, \dots, e_n are independent normal errors with mean 0 and standard deviation

Mean/typical value of Y
in pairs (X, Y) with $X = X_i$

+ random error/deviation term

Estimate $\mu(Y | X_i) = \beta_0 + \beta_1 X_i$ with $\hat{Y}_i = \hat{\mu}(Y | X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i$

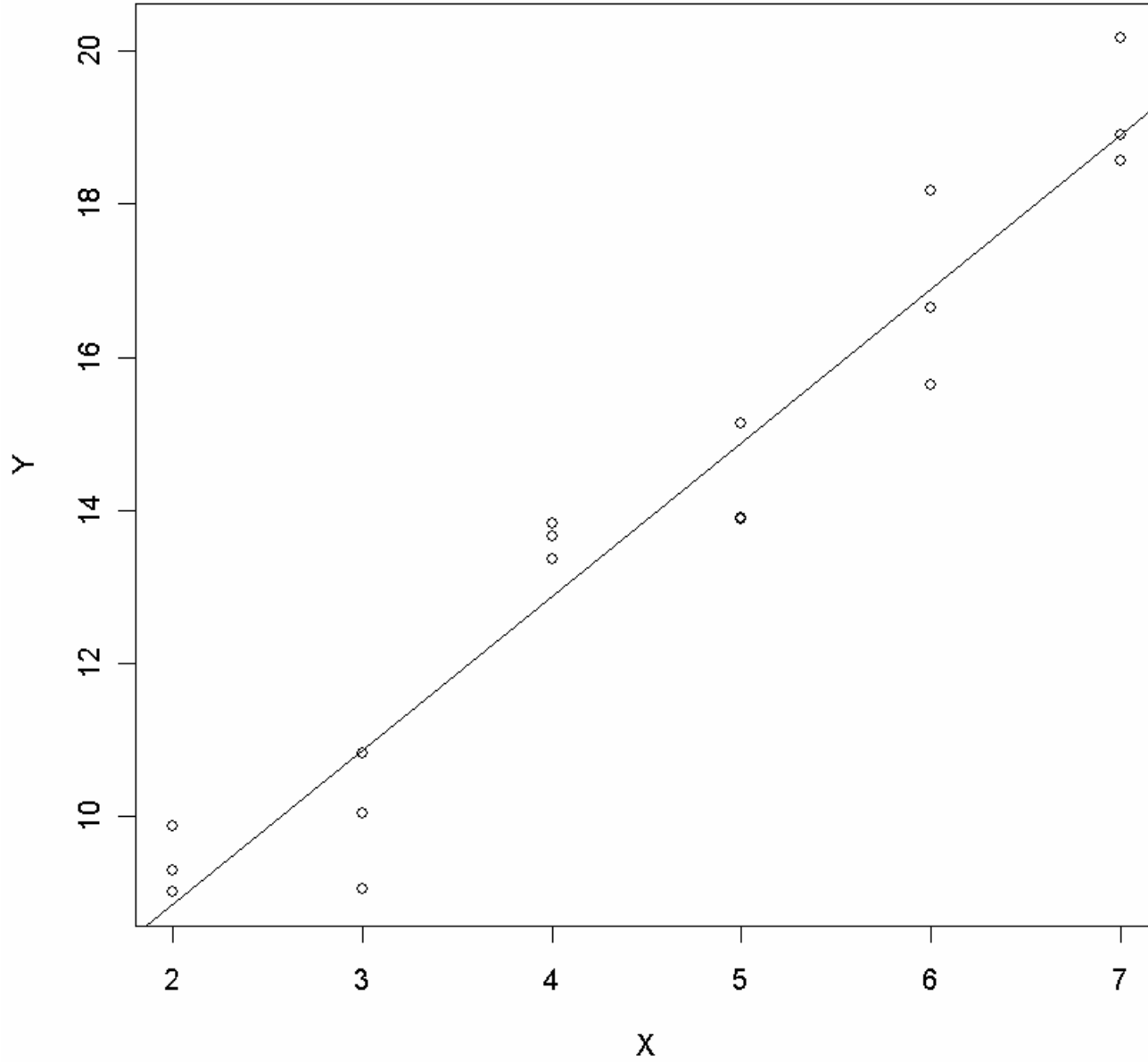
& estimate error e_i with residual $\hat{e}_i = Y_i - \hat{Y}_i$

To verify the assumptions of simple linear regression,
it's possible to use

1. scatterplot of (X_i, Y_i) to check for linearity & constant standard deviation (Assumptions 2,4)
2. residual plot of (\hat{Y}_i, \hat{e}_i) to check for linearity & constant standard deviation (Assumptions 2,4)
3. histogram and normal probability plot of residuals to check normality (Assumption 3)

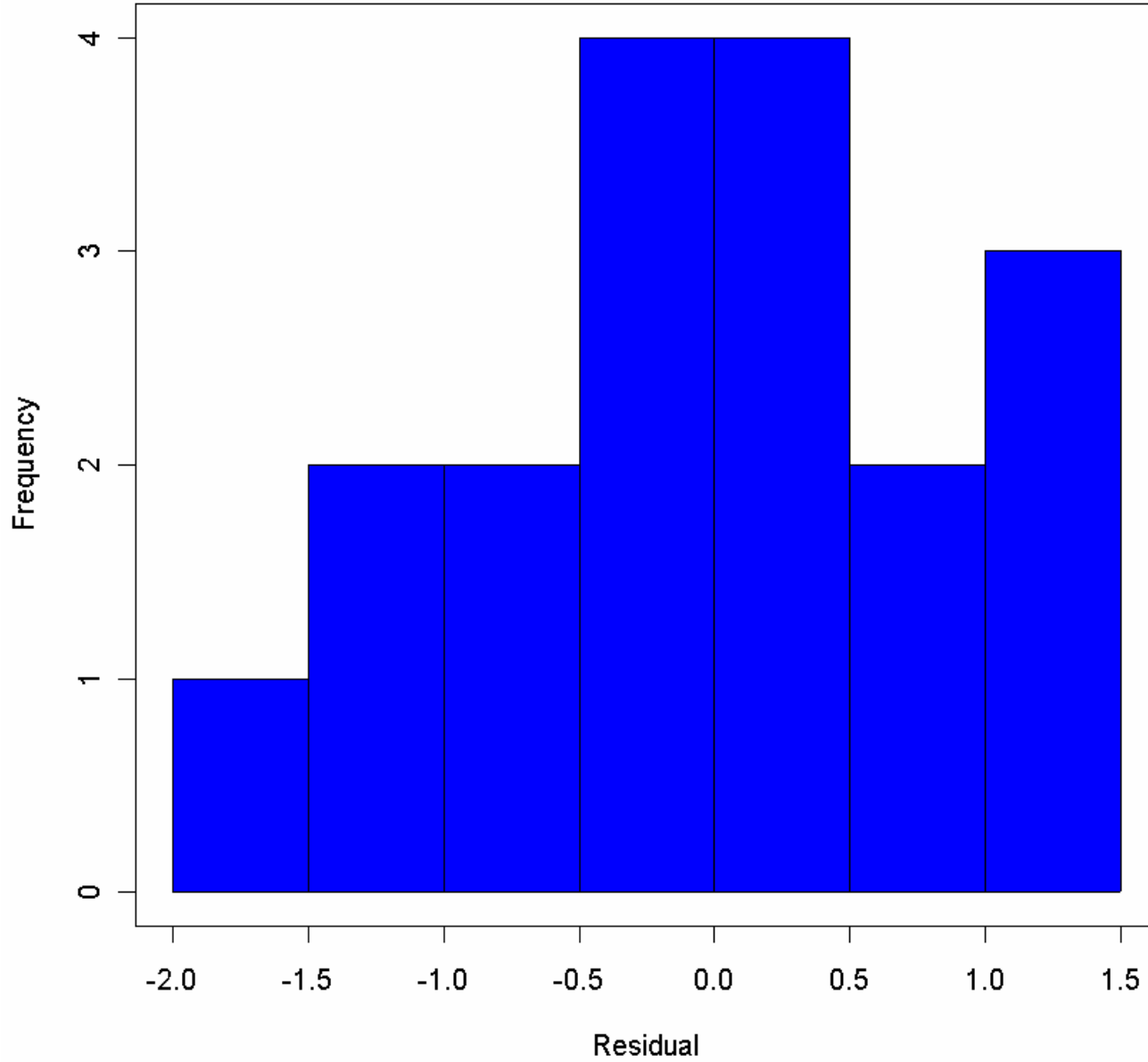
Example A:

Scatter Plot of Y vs. X



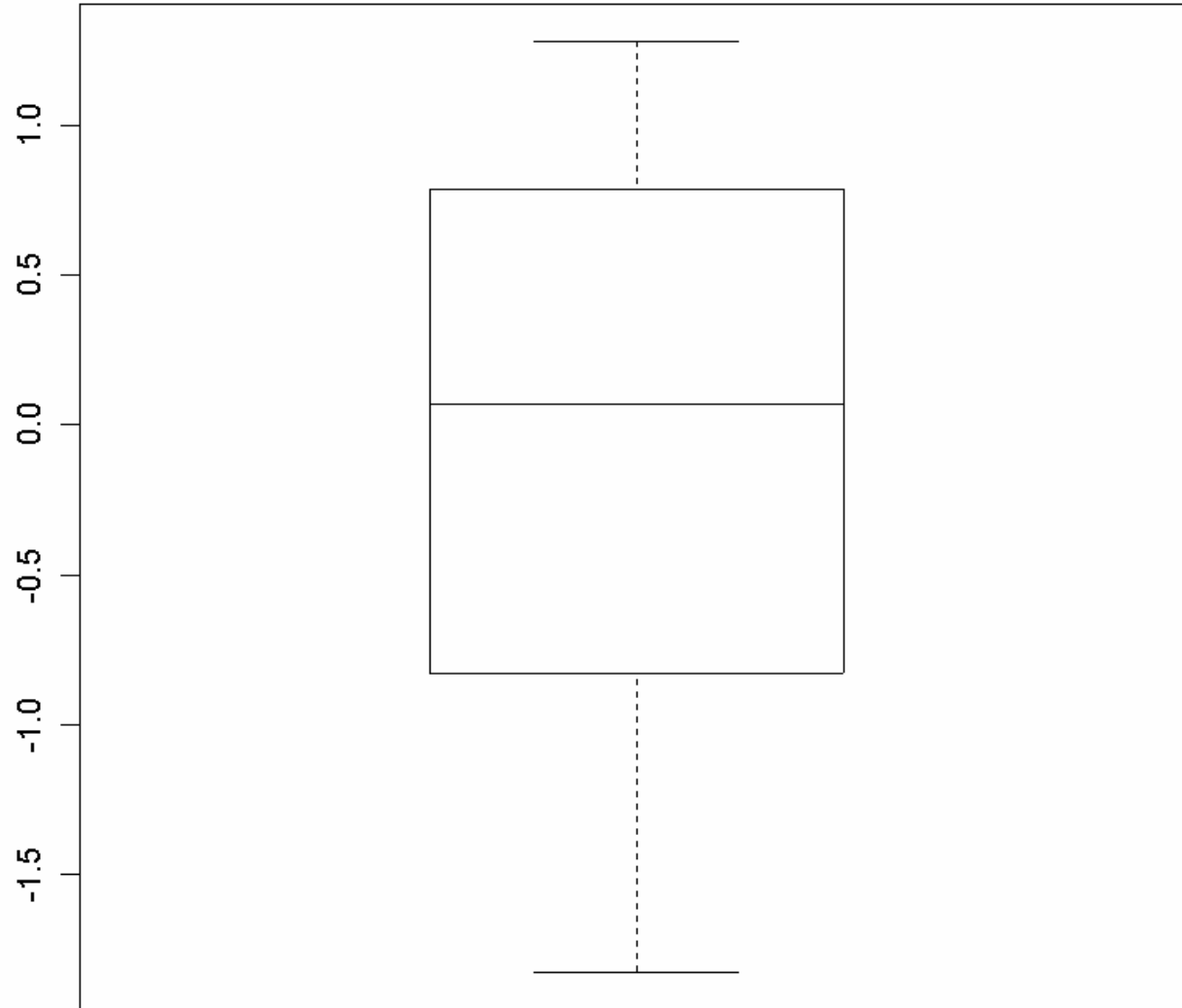
Example A:

Histogram of Residuals



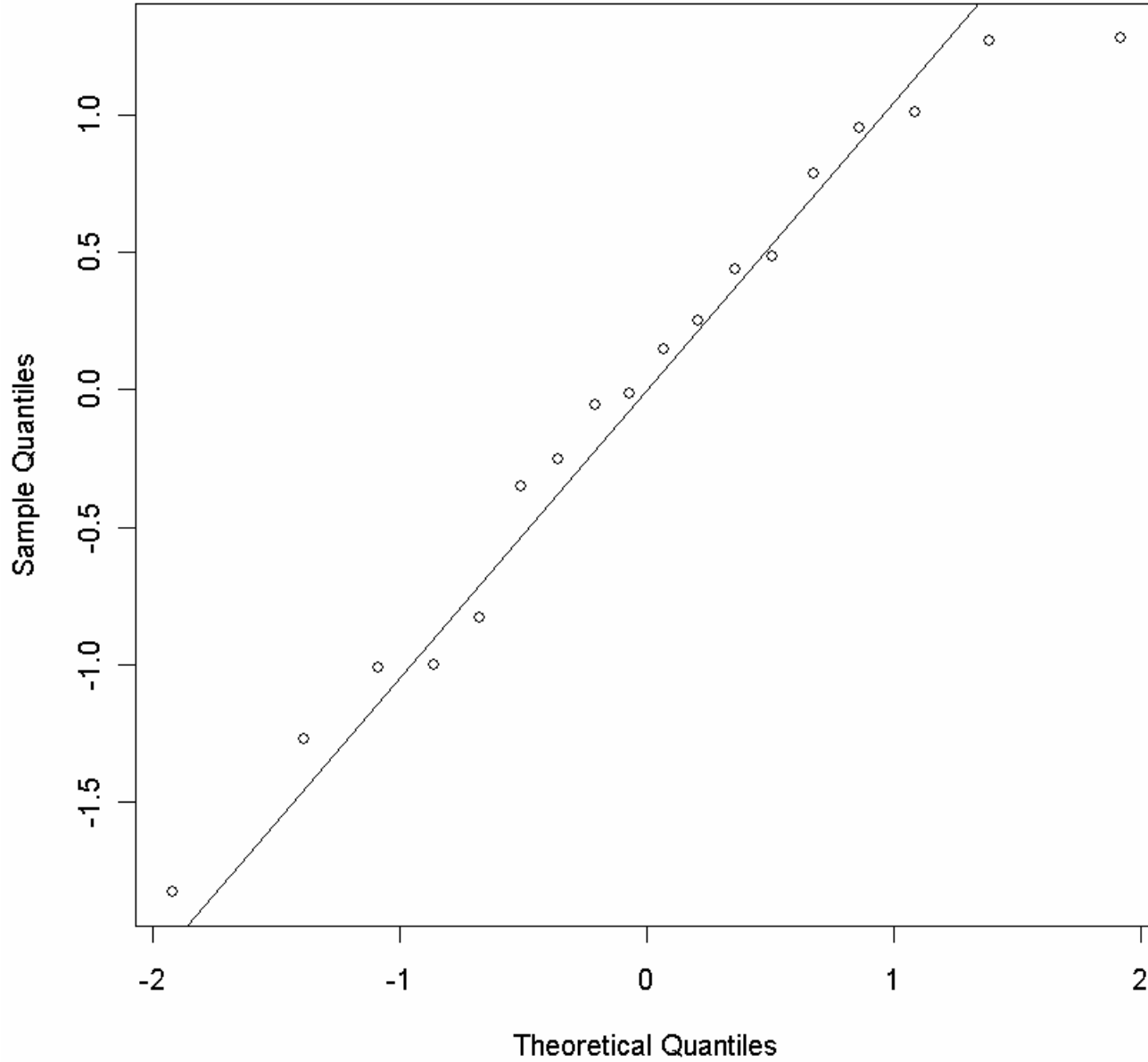
Example A:

Boxplot of Residuals



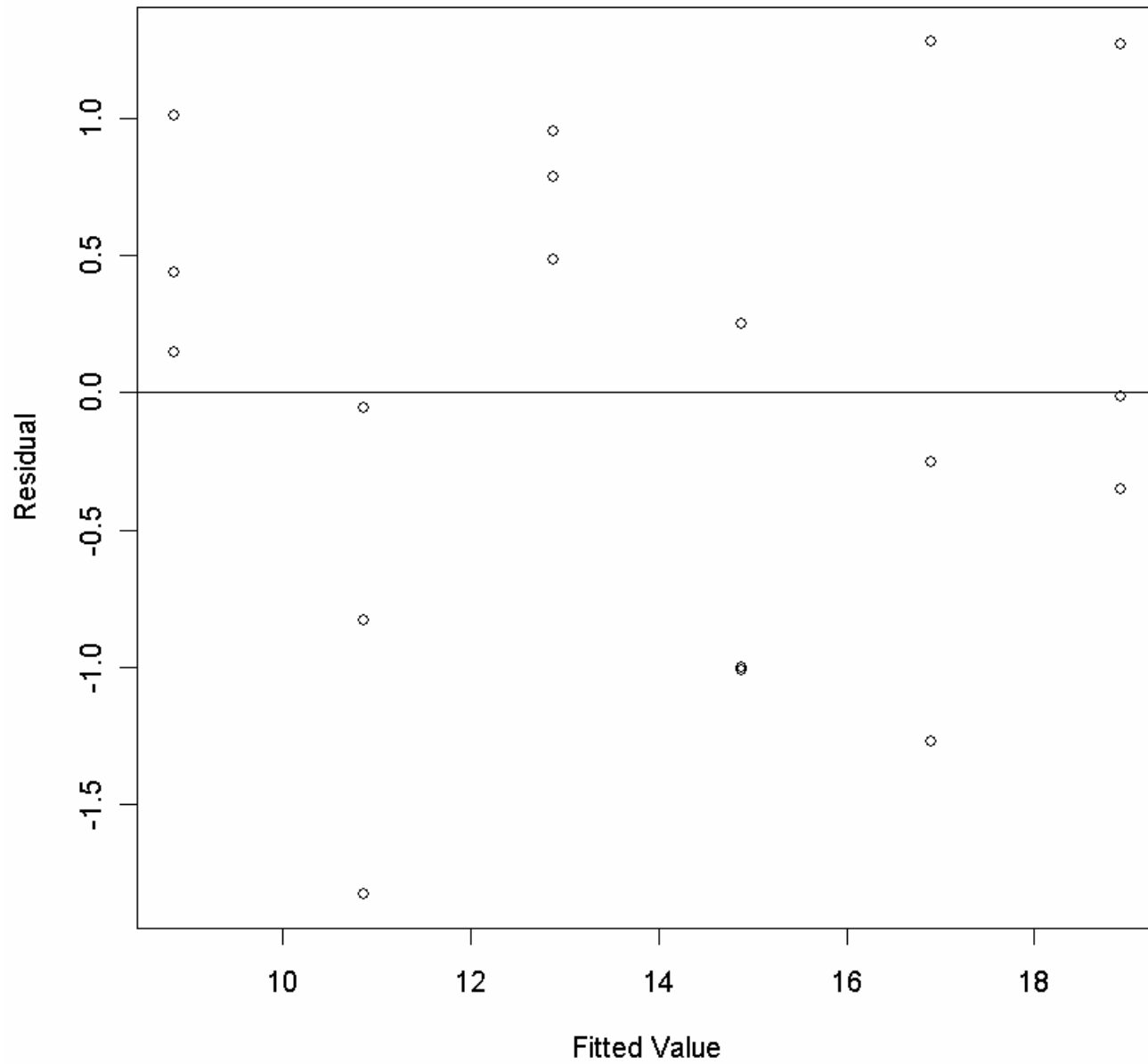
Example A:

Normal Probability Plot



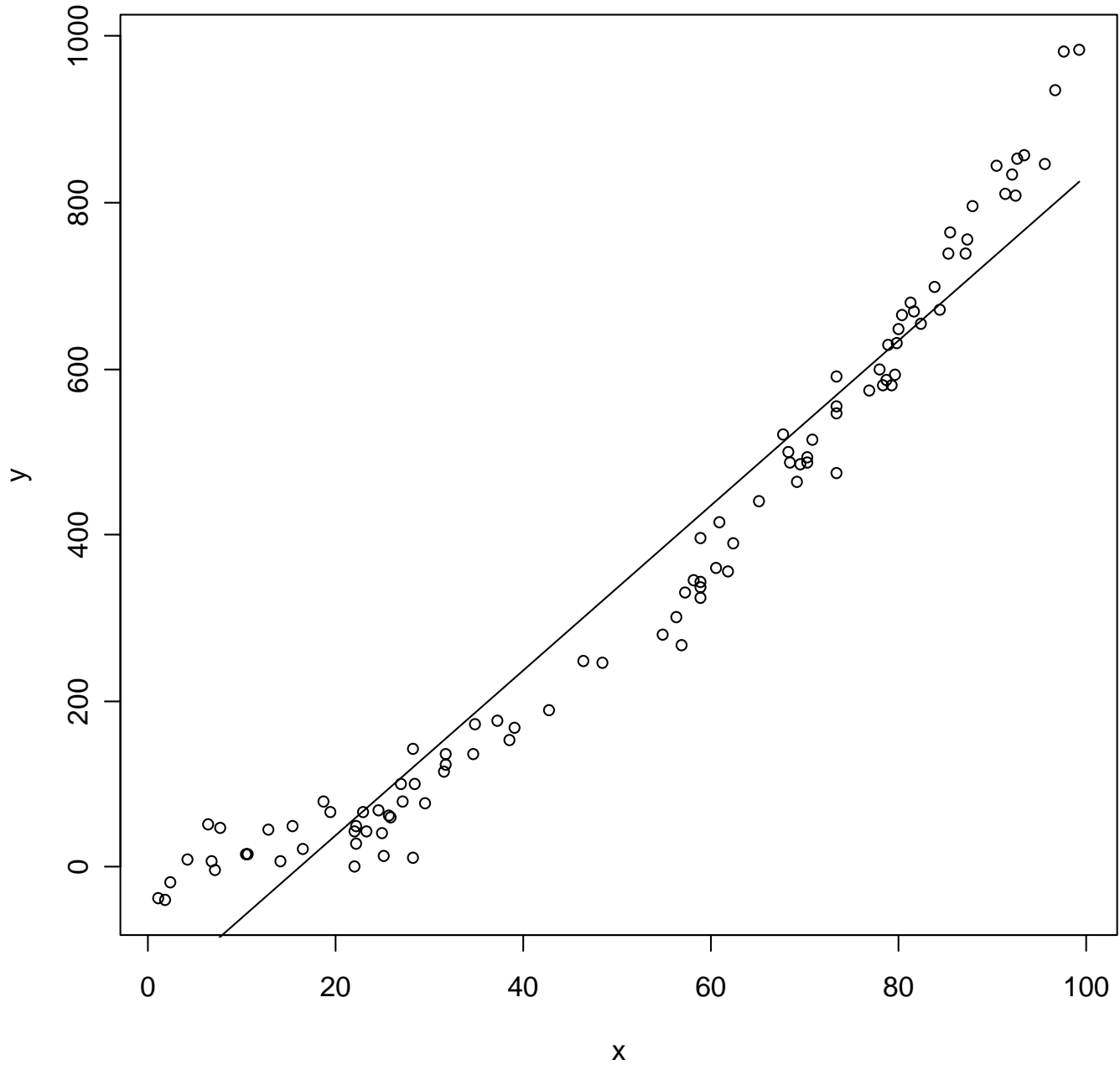
Example A:

Residual Plot



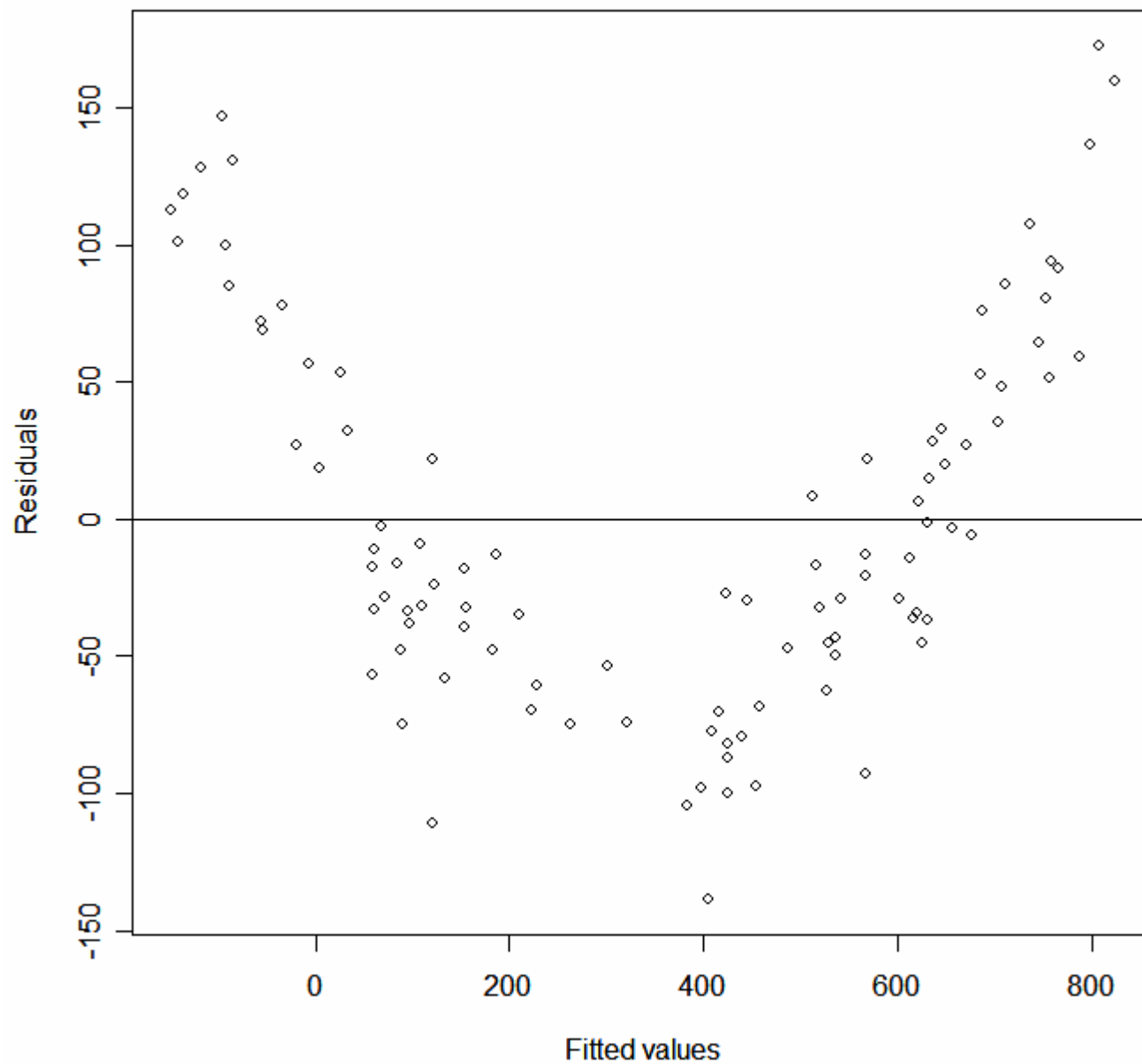
Example B:

Scatterplot of Y vs X



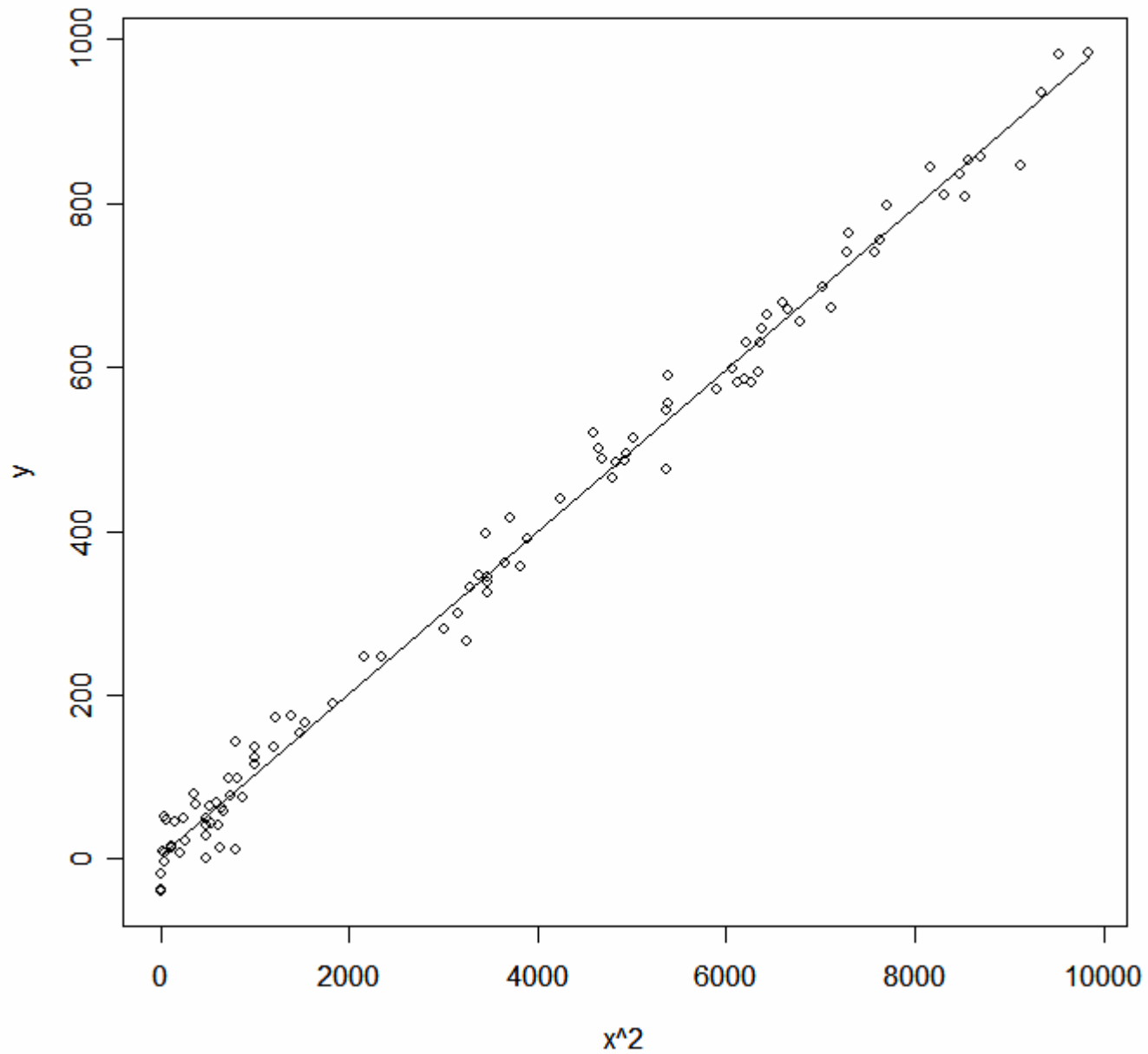
Example B:

Residual Plot



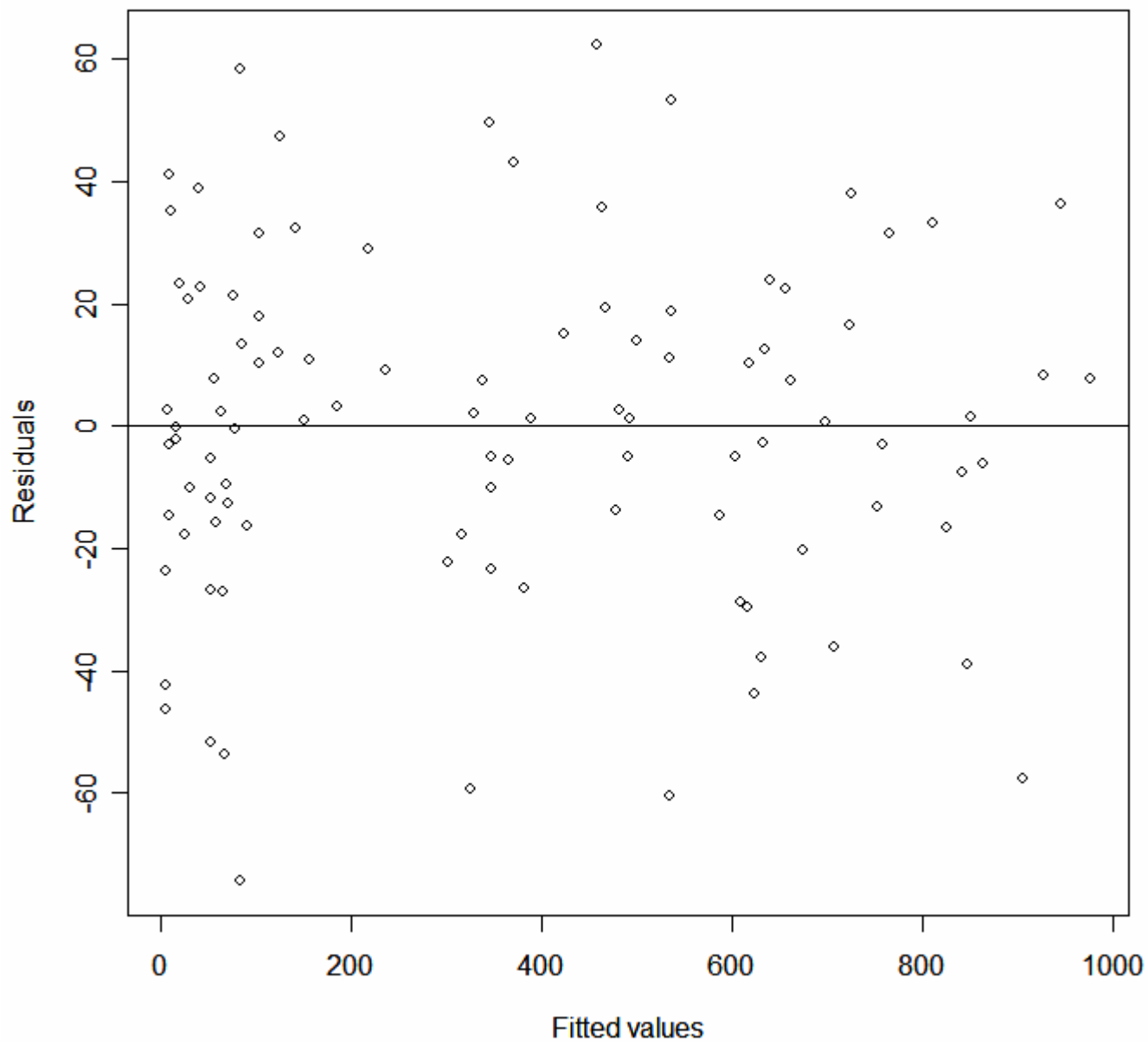
Example B:

Scatterplot of Y vs X²



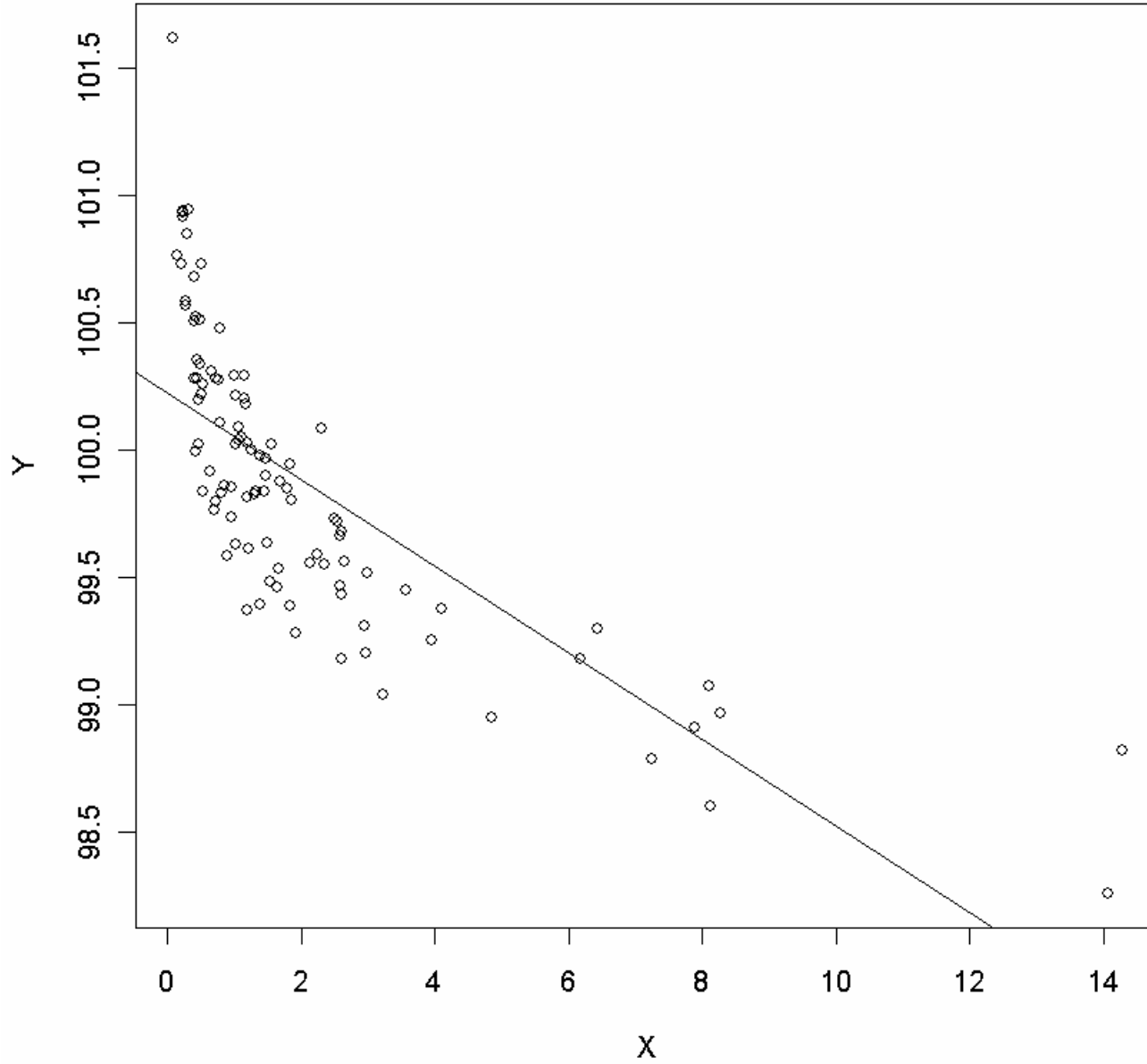
Example B:

Residual Plot



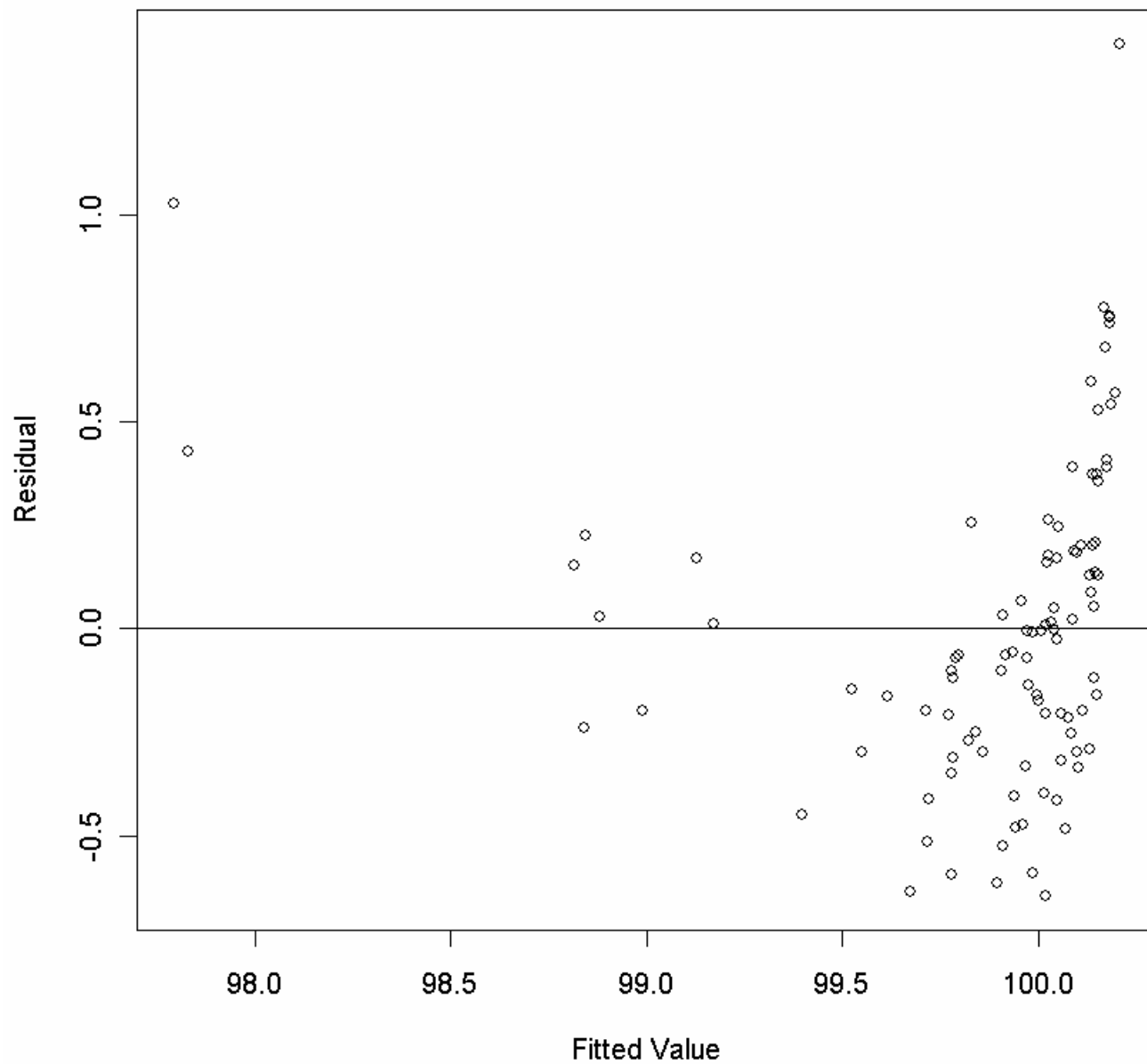
Example C:

Scatter Plot of Y vs. X



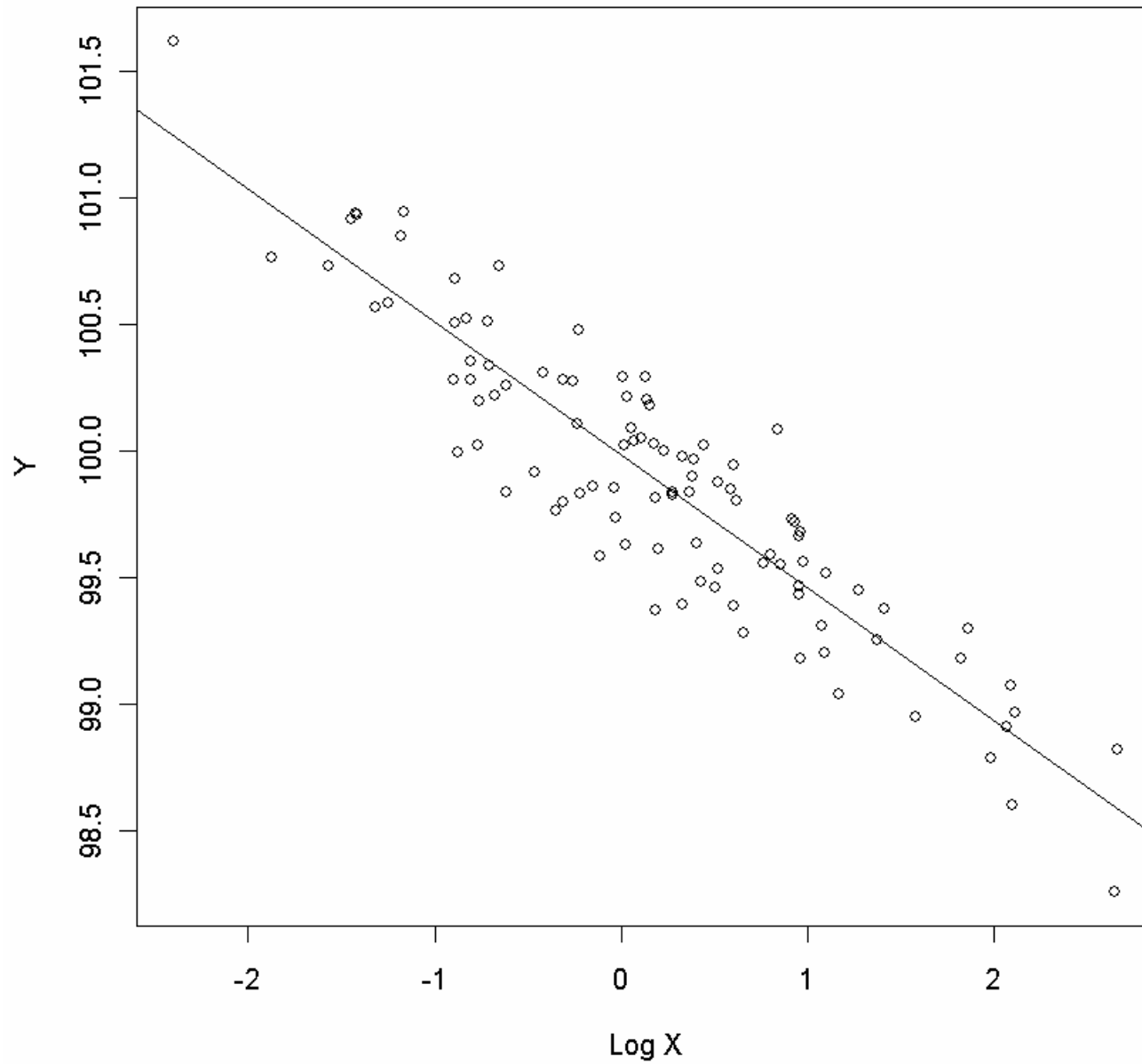
Example C:

Residual Plot



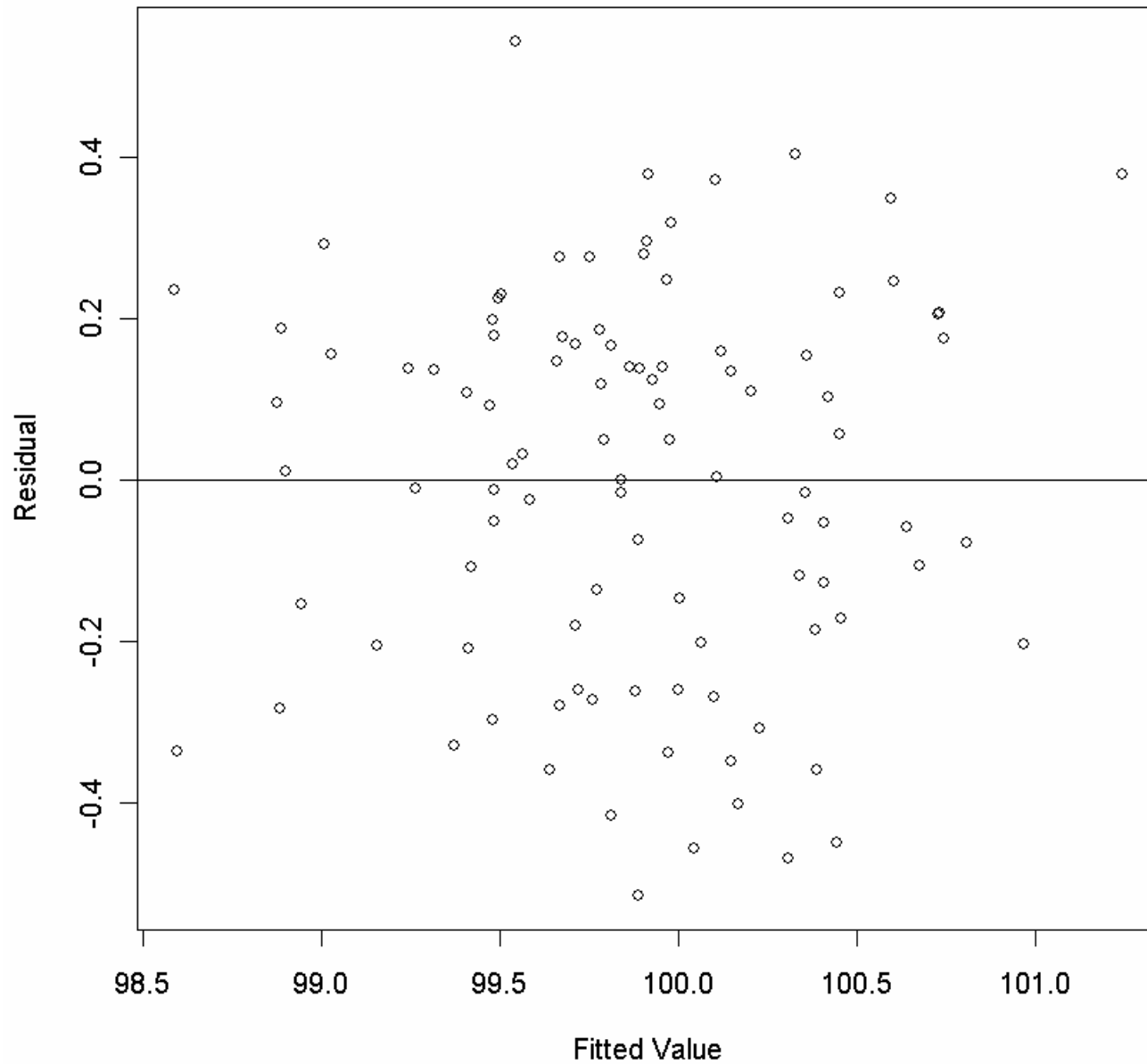
Example C:

Scatter Plot of Y vs. Log X



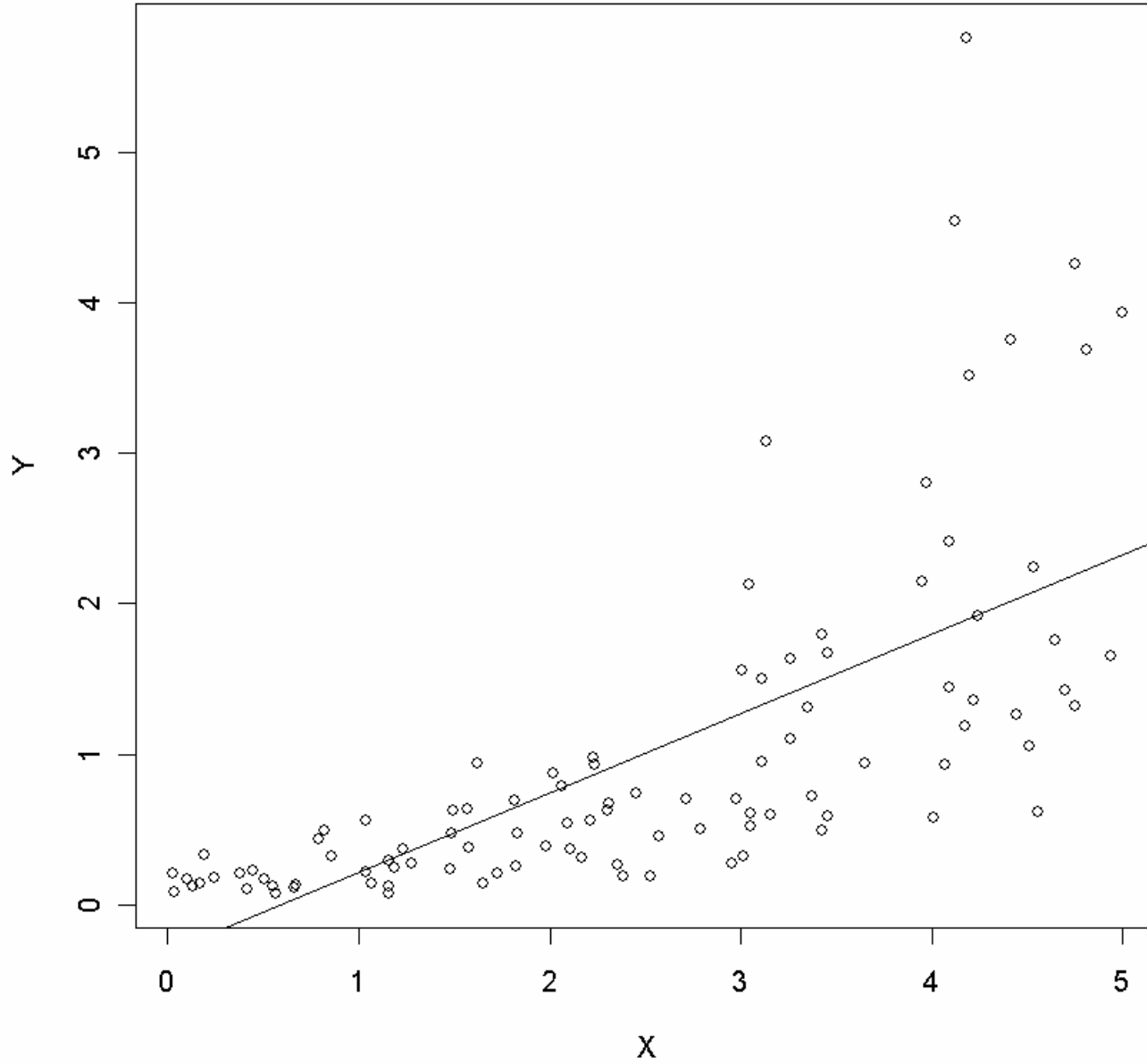
Example C:

Residual Plot



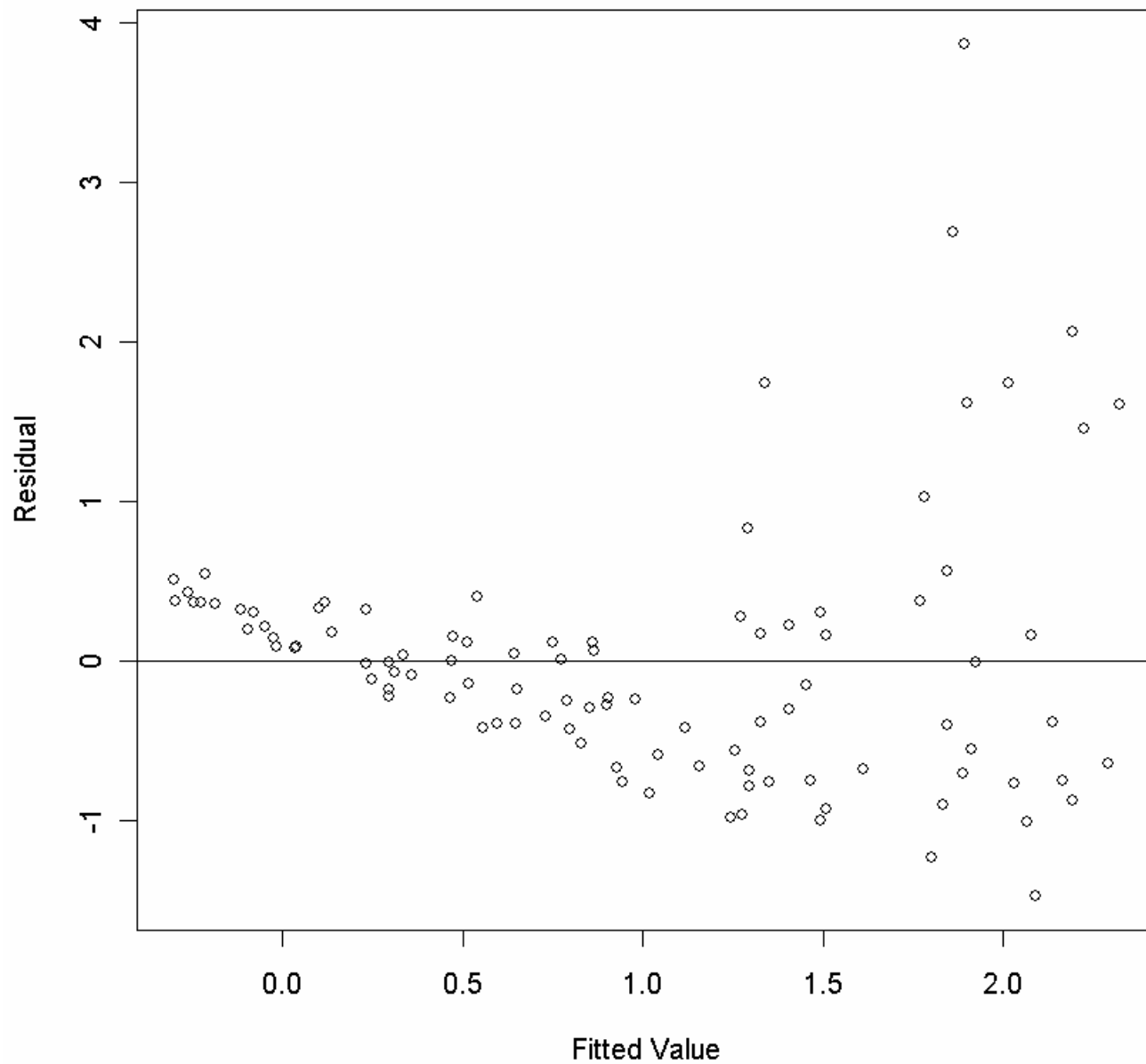
Example D:

Scatter Plot of Y vs. X



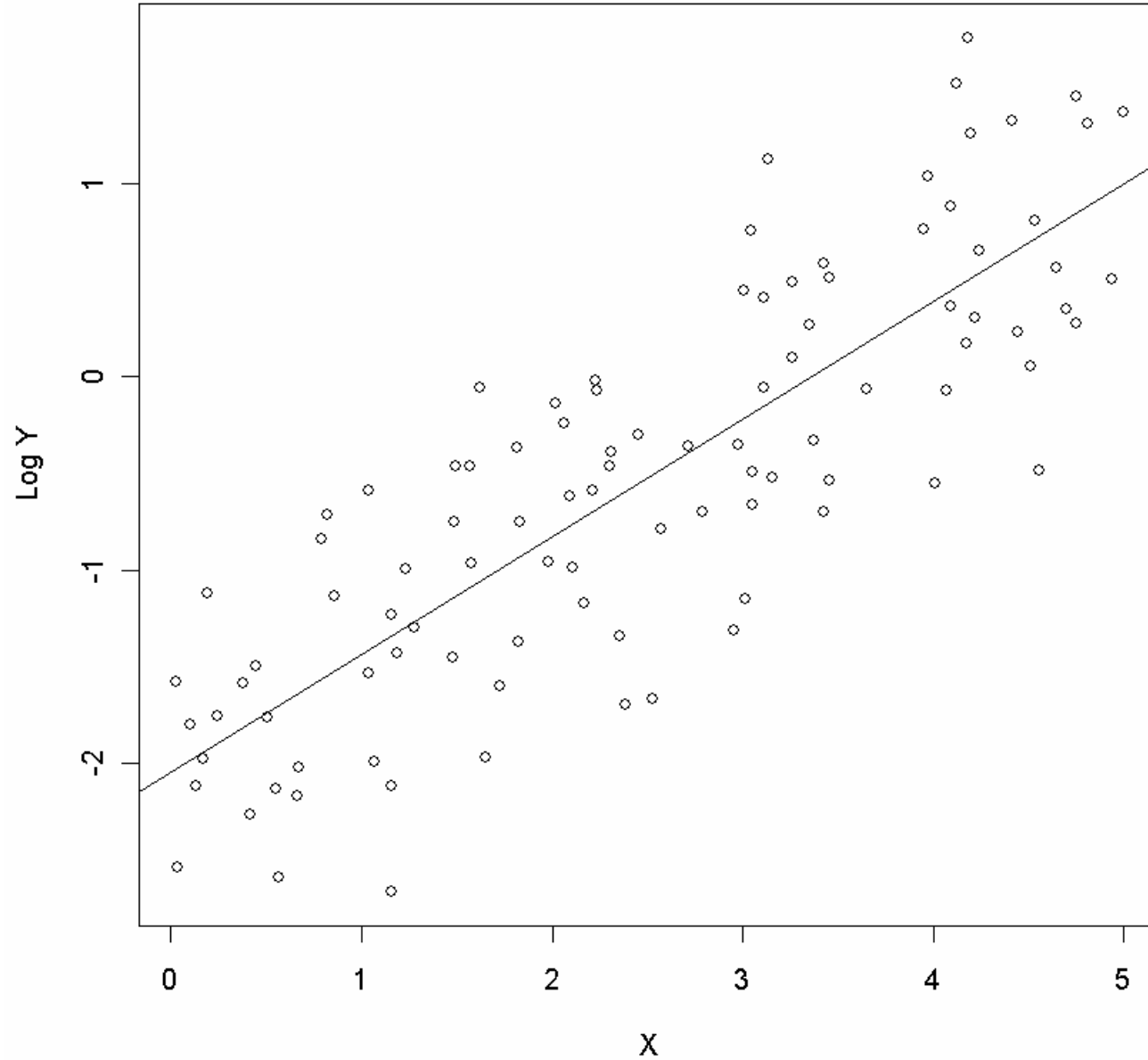
Example D:

Residual Plot



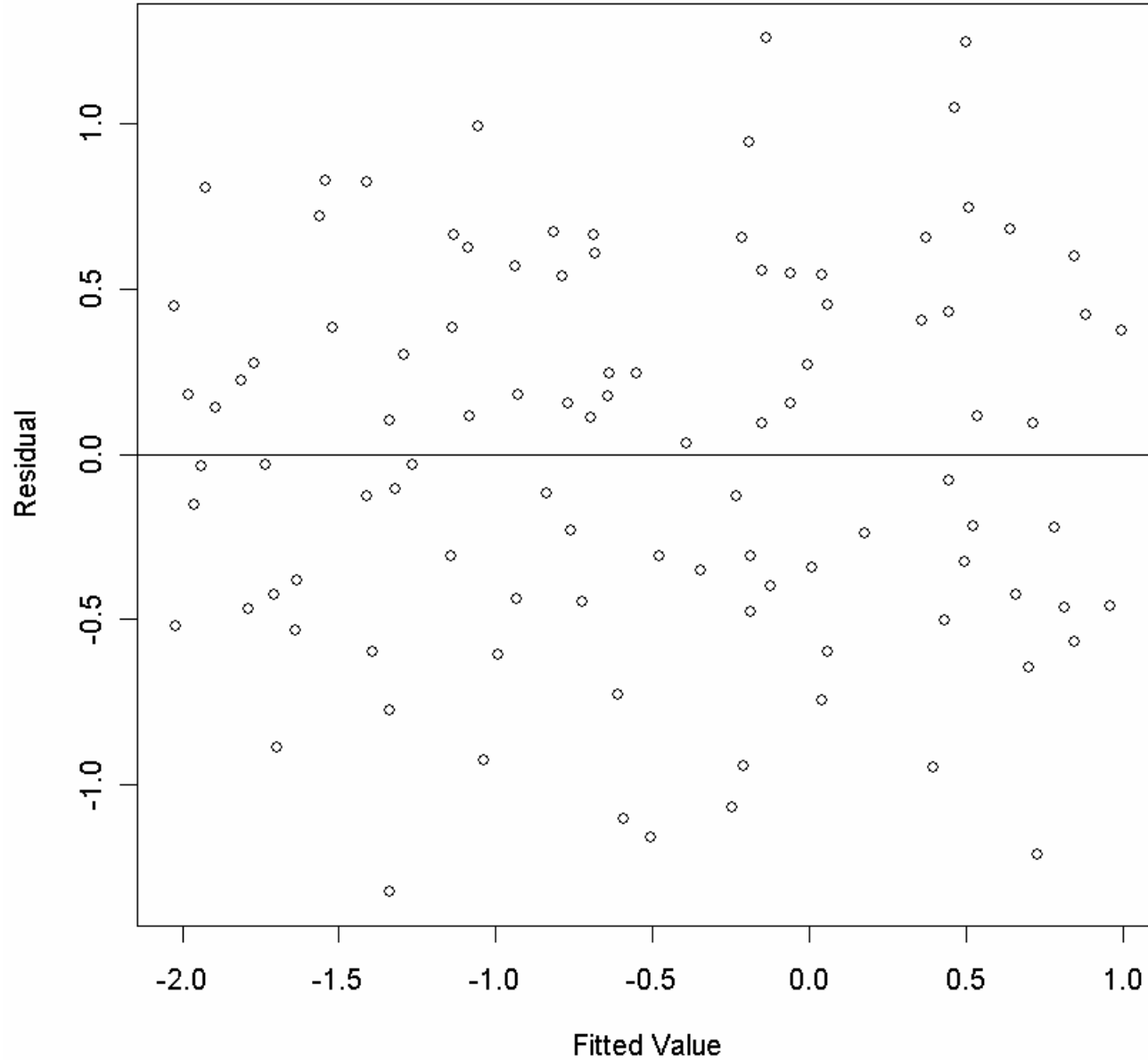
Example D:

Scatter Plot of Log Y vs. X



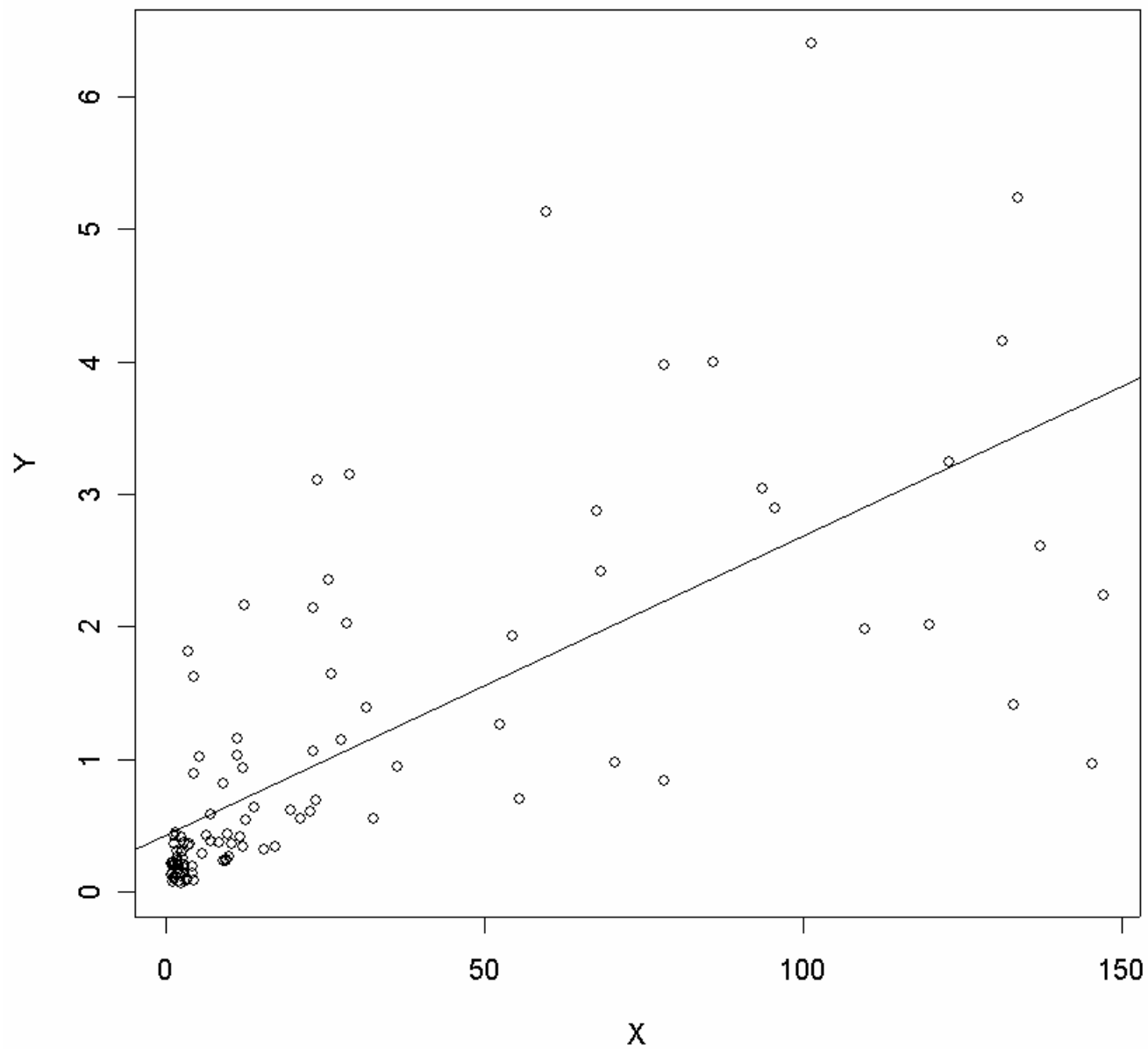
Example D:

Residual Plot



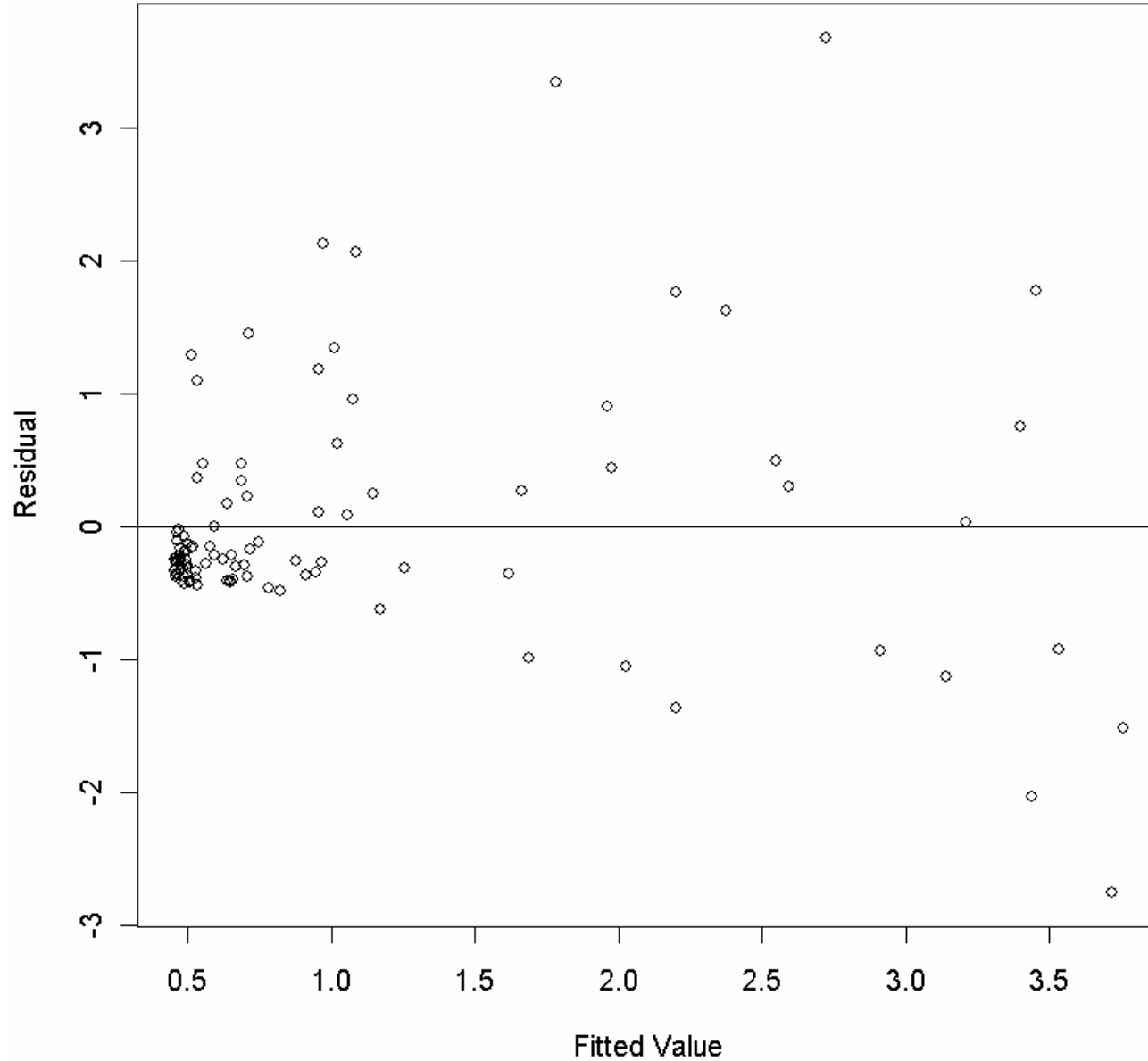
Example E:

Scatter Plot of Y vs. X



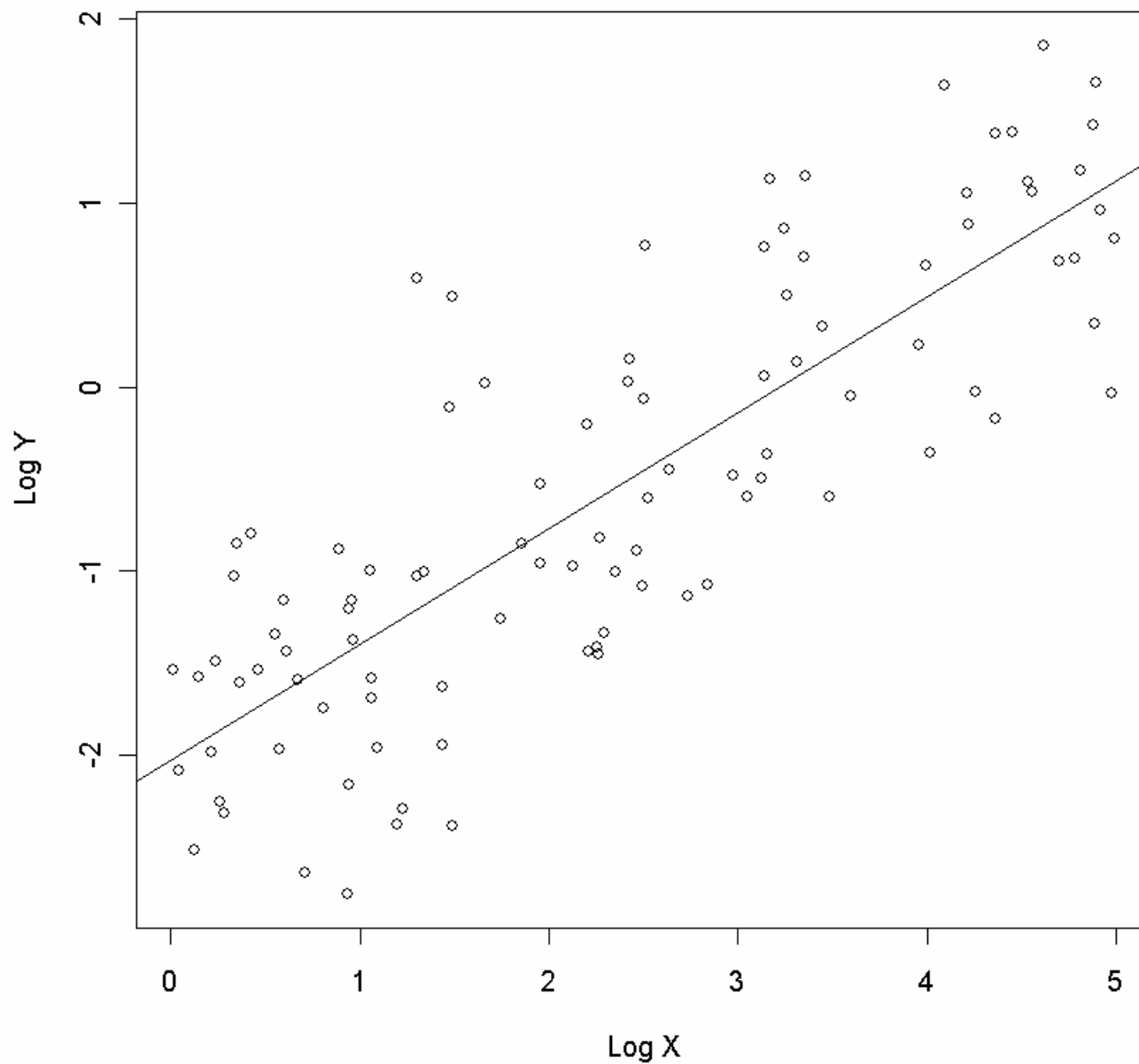
Example E:

Residual Plot



Example E:

Scatter Plot of Log Y vs. Log X



Example E:

Residual Plot

