Please Click https://canvas.ust.hk and SFQ on the left panel To Fill out End-of-Term Course Survey.

**Thanks for your attention!**

# Chapter 6. Transformation of Variables

## *Outline*

# 6.1. Transformation to Achieve Linearity

# 6.1 Transformation to Achieve Linearity

## Introduction

Data do not always come in a form that is immediately suitable for analysis. We often have to **transform** the variables before carrying out the analysis. Transformations are applied to accomplish certain objectives such as to **ensure linearity**, to achieve **normality**, or to **stabilize the variance**. It often becomes necessary to fit a linear regression model to the transformed rather than the original variables. This is common practice. In this chapter, we discuss the situations where it is necessary to transform the data, the possible choices of transformation, and the analysis of transformed data.

We illustrate transformation **mainly using simple regression**. In multiple regression where there are several predictors, **some** may require transformation and **others may not**. Although the same technique can be applied to multiple regression, transformation in multiple regression requires more effort and care.

# 6.1 Transformation to Achieve Linearity

## Introduction

The necessity for transforming the data arises because the original variables, or the model in terms of the original variables, violates one or more of the standard regression assumptions. Two of the most commonly violated assumptions are the **linearity** of the model and the **constancy** of the error variance. As mentioned in Chapters 2 and 3, a regression model is linear when the parameters present in the model occur linearly even if the predictor variables occur nonlinearly. For example, each of the four following models is

$$
\begin{aligned}
Y &= \beta_0 + \beta_1 X + \varepsilon, \\
Y &= \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon, \\
Y &= \beta_0 + \beta_1 \log X + \varepsilon, \\
Y &= \beta_0 + \beta_1 \sqrt{X} + \varepsilon,
\end{aligned}
$$

because the model parameters $\beta_0, \beta_1, \beta_2$ enter linearly.

On the other hand, $Y = \beta_0 + e^{\beta_1 X} + \varepsilon$ is a **nonlinear** model because the parameter $\beta_1$ does not enter the model linearly. To satisfy the assumptions of the standard regression model, instead of working with the original variables, we sometimes work with transformed varaibles.

## 6.1 Transformation to Achieve Linearity

### Class Discussion

1. Which of the following statements is true. (3pts)

   (a) $e^Y = e^{\beta_0} X^{\beta_1} e^{\varepsilon}$ can be viewed as a linear model

   (b) $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$ is a non-linear model

## 6.1 Transformation to Achieve Linearity

### Reasons of Transformations

Transformations may be necessary for several reasons

1. Theoretical considerations may specify that the relationship between two variables is nonlinear. An appropriate transformation of the variables can make the relationship between the transformed variables linear. Consider an example from learning theory (experimental psychology). A learning model that is widely used states that the time taken to perform a task on the $i$th occasion $(T_i)$ is

$$T_i = \alpha\beta^i, \quad \alpha > 0, \quad 0 < \beta < 1. \tag{6.1}$$

The relationship between $(T_i)$ and $i$ as given in (6.1) is nonlinear, and we cannot directly apply techniques of linear regression. On the other hand, if we take logarithms of both sides, we get

$$\log T_i = \log\alpha + i\log\beta, \tag{6.2}$$

showing that $\log T_i$ and $i$ are linearly related. The transformation enables us to use standard regression methods. Although the relationship between the original variables was nonlinear, the relationship between transformed variables is linear. A transformation is used to achieve the linearity of the fitted model.

# 6.1 Transformation to Achieve Linearity

## Reasons of Transformations

Under the standard assumptions of multiple linear regression, we have $\mathrm{Var}(Y) = \mathrm{Var}(\varepsilon) = \sigma^2$.

2. The response variable $Y$, which is analyzed, may have a probability distribution whose variance is related to the mean. If the mean is related to the value of the predictor variable $X$, then the variance of $Y$ will change with $X$, and will **not be constant**. The distribution of $Y$ will usually also be non-normal under these conditions. **Non-normality** invalidates the standard tests of significance (although not in a major way with large samples) since they are based on the normality assumption. The unequal variance of the error terms will produce estimates that are unbiased, but are no longer best in the sense of having the **smallest variance**. In these situations we often transform the data so as to ensure normality and constancy of error variance. In practice, the transformations are chosen to ensure the constancy of variance (variance-stabilizing transformations). It is a fortunate coincidence that the variance-stabilizing transformations are also good normalizing transforms.

3. There are neither prior theoretical nor probabilistic reasons to suspect that a transformation is required. The evidence comes from examining the residuals from the fit of a linear regression model in which the original variables are used.

## 6.1 Transformation to Achieve Linearity

One of the standard assumptions made in regression analysis is that the model which describes the data is **linear**. From theoretical considerations, or from an examination of scatter plot of $Y$ against each predictor $X_j$, the relationship between $Y$ and $X_j$ may appear to be nonlinear. There are, however, several **simple nonlinear regression models** which by appropriate transformations can be **made linear**. We list some of these linearizable curves in Table 6.1. The corresponding graphs are given in Figures 6.1-6.4.

**Table 6.1** Linearizable Simple Regression Functions with Corresponding Transformations

| Function | Transformation | Linear Form | Graph |
|---|---|---|---|
| $Y = \alpha X^{\beta}$ | $Y' = \log Y, X' = \log X$ | $Y' = \log \alpha + \beta X'$ | Figure 6.1 |
| $Y = \alpha e^{\beta X}$ | $Y' = \ln Y$ | $Y' = \ln \alpha + \beta X$ | Figure 6.2 |
| $Y = \alpha + \beta \log X$ | $X' = \log X$ | $Y = \alpha + \beta X'$ | Figure 6.3 |
| $Y = \frac{X}{\alpha X - \beta}$ | $Y' = \frac{1}{Y}, X' = \frac{1}{X}$ | $Y' = \alpha - \beta X'$ | Figure 6.4(a) |
| $Y = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$ | $Y' = \ln \frac{Y}{1-Y}$ | $Y' = \alpha + \beta X$ | Figure 6.4(b) |

Here for simplicity, we do not include the random error $\varepsilon$.

Figure 6.1 - 6.4

# 6.1 Transformation to Achieve Linearity

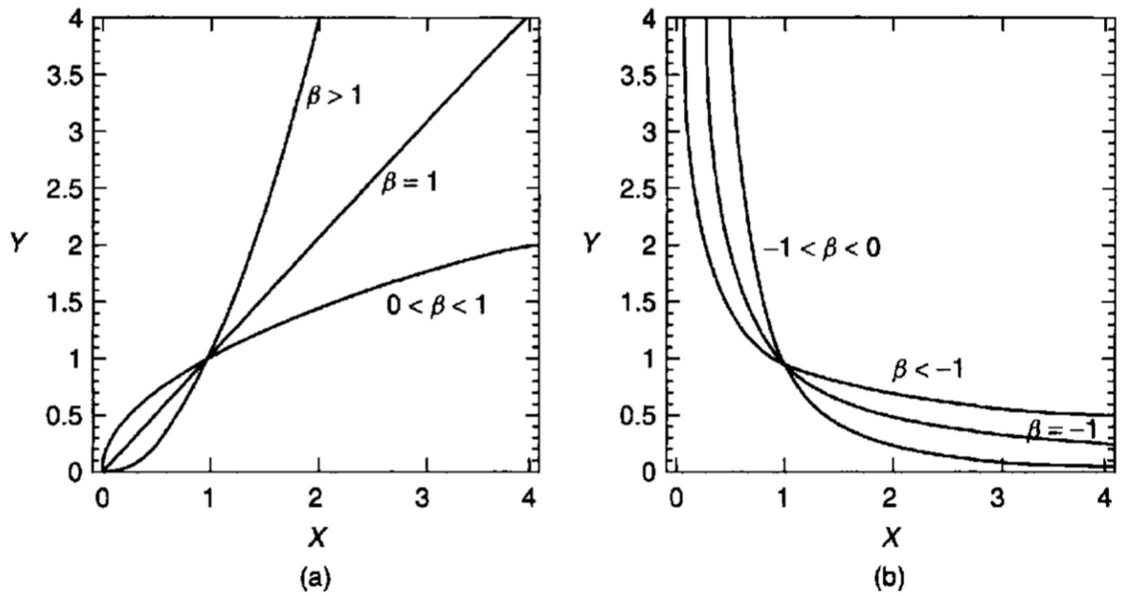



**Figure 6.1** Graphs of the linearizable function $Y = \alpha\, X^{\beta}$.

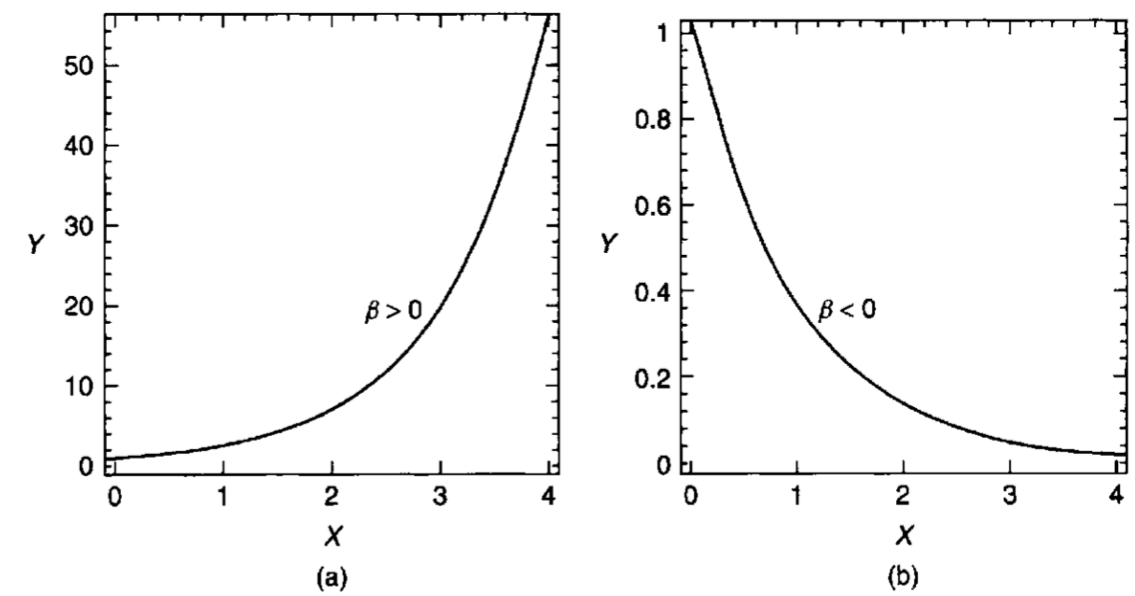



**Figure 6.2** Graphs of the linearizable function $Y = \alpha\, e^{\beta X}$.

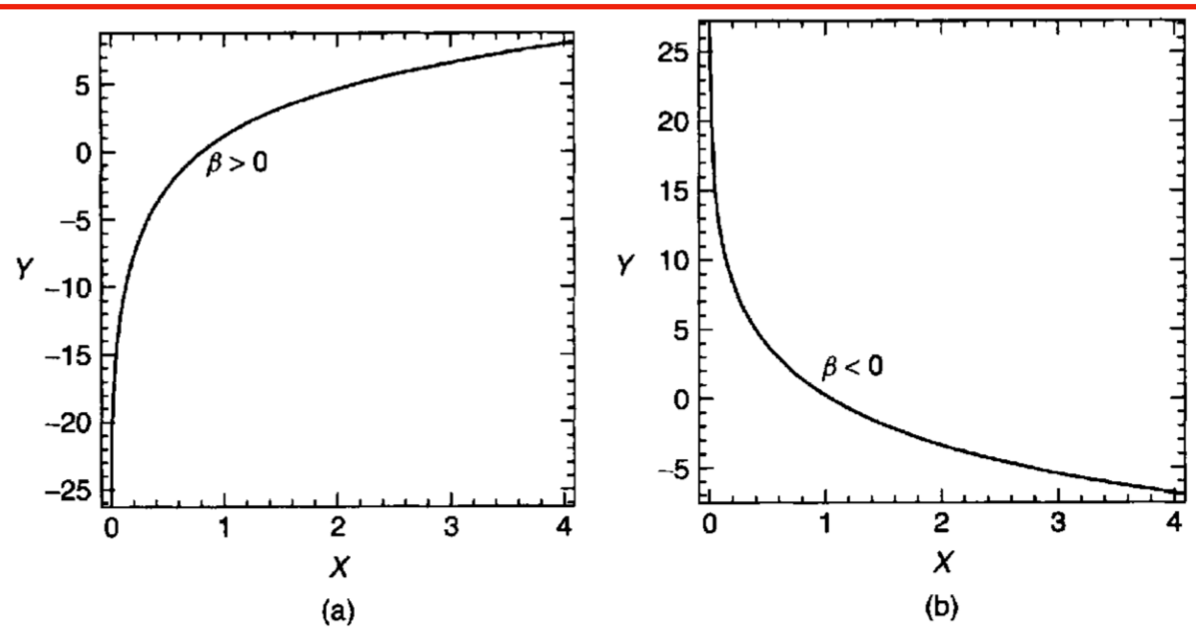



**Figure 6.3** Graphs of the linearizable function $Y = \alpha + \beta\ \log X$.

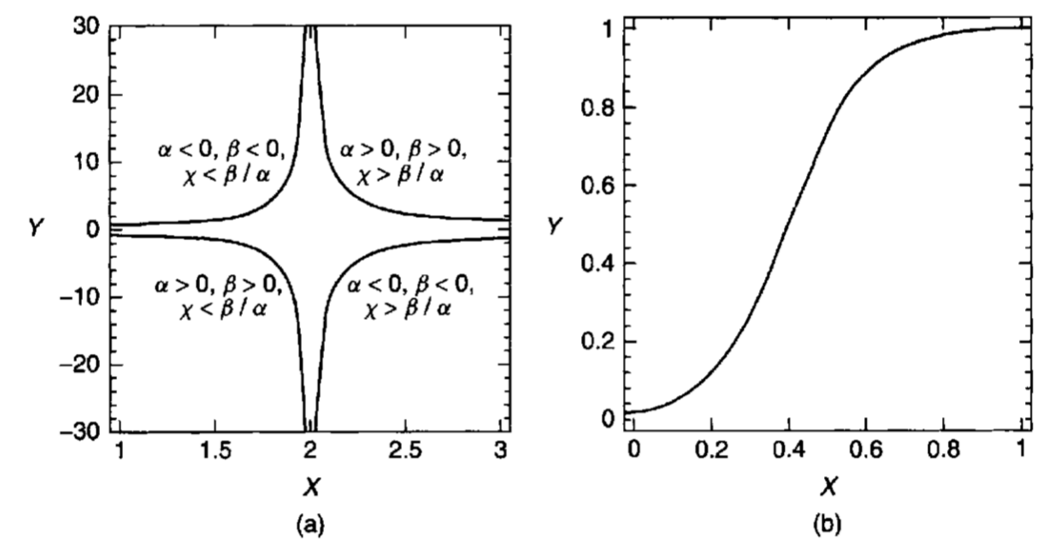



**Figure 6.4** Graphs of the linearizable functions: (a) $Y = X/(\alpha\ X - \beta)$ and (b) $Y = (e^{\alpha+\beta\ X})/(1 + e^{\alpha+\beta\ X})$.

## 6.1 Transformation to Achieve Linearity

When curvature is observed in the scatter plot of $Y$ against $X$, a linearizable curve from one of those given in Figures 6.1-6.4 may be chosen to represent the data. There are, however, many simple nonlinear models that **cannot be linearized**. Consider for example, $Y = \alpha + \beta\delta^X$ , a modified exponential curve, or

$$Y = \alpha_1 e^{\theta_1 X} + \alpha_2 e^{\theta_2 X},$$

which is the sum of two exponential functions. The **strictly nonlinear models** (i.e., those not linearizable by variable transformation) require very different methods. We do not describe them in this course.

In the following section, theoretical considerations lead to a model that is nonlinear. The model is, however, linearizable and we indicate the appropriate analysis.

## 6.2 Example: Bacteria Death Due to X-Ray Radiation

## 6.2 Example: Bacteria Death Due to X-Ray Radiation

### Introduction

The data given in Table 6.2 represent the number of surviving bacteria (in hundreds) as estimated by plate counts in an experiment with marine bacterium following exposure to 200-kilovolt X-rays for periods ranging from $t = 1$ to 15 intervals of six minutes. The response variable $n_t$ represents the number surviving after exposure time $t$. The experiment was carried out to test the single-hit hypothesis of X-ray action under constant field of radiation. According to this theory, there is a single vital center in each bacterium, and this must be hit by a ray before the bacteria is inactivated or killed. The particular bacterium studied does not form clumps or chains, so the number of bacterium can be estimated directly from plate counts.

**Table 6.2** Number of Surviving Bacteria (Units of 100)

| $t$ | $n_t$ | $t$ | $n_t$ | $t$ | $n_t$ |
|---|---|---|---|---|---|
| 1 | 355 | 6 | 106 | 11 | 36 |
| 2 | 211 | 7 | 104 | 12 | 32 |
| 3 | 197 | 8 | 60 | 13 | 21 |
| 4 | 166 | 9 | 56 | 14 | 19 |
| 5 | 142 | 10 | 38 | 15 | 15 |

## 6.2 Example: Bacteria Death Due to X-Ray Radiation

### Introduction

If the theory is applicable, then $n_t$ and $t$ should be related by

$$n_t = n_0 e^{\beta_1 t}, \quad t \geq 0, \tag{6.3}$$

where $n_0$ and $\beta_1$ are parameters. These parameters have simple physical interpretations; $n_0$ is the number of bacteria at the start of the experiment, and $\beta_1$ is the destruction (decay) rate. Taking logarithms of both sides of (6.3), we get

$$\ln n_t = \ln n_0 + \beta_1 t = \beta_0 + \beta_1 t, \tag{6.4}$$

where $\beta_0 = \ln n_0$ and we have $\ln n_t$ as a linear function of $t$. If we introduce $\varepsilon_t$ as the random error, our model becomes

$$\ln n_t = \beta_0 + \beta_1 t + \varepsilon_t \tag{6.5}$$

and we can now apply standard least squares methods.

## 6.2 Example: Bacteria Death Due to X-Ray Radiation

### Introduction

To get the error $\varepsilon_t$ in the transformed model (6.5) to be additive, the error must occur in the multiplicative form in the original model (6.3). The correct representation of the model should be

$$n_t = n_0 e^{\beta_1 t} \varepsilon_t', \tag{6.6}$$

where $\varepsilon_t'$ is the multiplicative random error. By comparing (6.5) and (6.6), it is seen that $\varepsilon_t = \ln \varepsilon_t'$. For standard least squares analysis $\varepsilon_t$ should be normally distributed, which in turn implies that $\varepsilon_t'$, has a log-normal distribution.[2] In practice, after fitting the transformed model we look at the residuals from the fitted model to see if the model assumptions hold. No attempt is usually made to investigate the random component, $\varepsilon_t'$, of the original model.

The random variable $Y$ is said to have a log-normal distribution if $\ln(Y)$ has a normal distribution.

# 6.2 Example: Bacteria Death Due to X-Ray Radiation

## Inadequacy of Linear Model

The first step in the analysis is to plot the raw data $n_t$ versus $t$. The plot, shown in Figure 6.5, suggests a **nonlinear** relationship between $n_t$ and $t$. However, we proceed by fitting the simple linear model and investigate the consequences of misspecification. The model is

$$n_t = \beta_0 + \beta_1 t + \varepsilon_t, \tag{6.7}$$

where $\beta_0$ and $\beta_1$ are constants; $\varepsilon_t$'s are the random errors, with zero means and equal variances, and are uncorrelated with each other. Estimates of $\beta_0$, $\beta_1$, their standard errors, and the square of the correlation coefficient are given in Table 6.3.

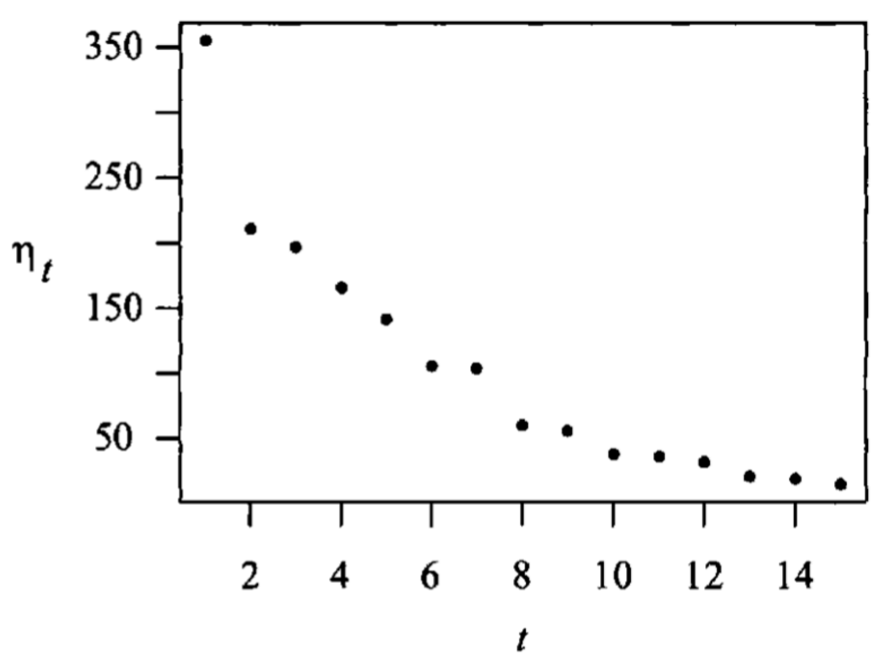



**Figure 6.5** Plot of $n_t$ against time $t$.

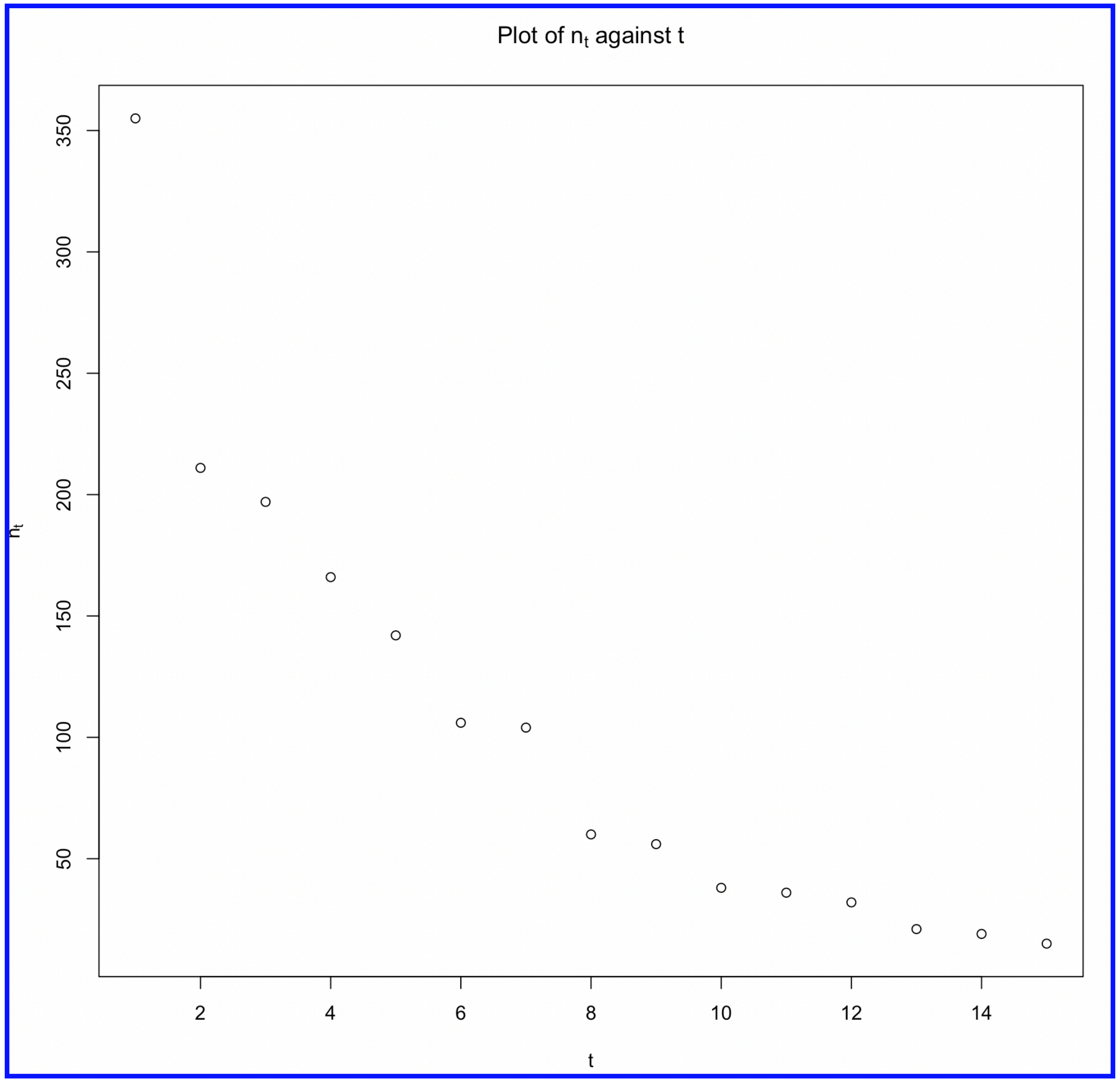
**Table 6.3** Estimated Regression Coefficients From Model (6.7)

| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | 259.58 | 22.73 | 11.42 | $< 0.0001$ |
| TIME ($t$) | $-19.46$ | 2.50 | $-7.79$ | $< 0.0001$ |
| | $n = 15$ | $R^2 = 0.823$ | $\hat{\sigma} = 41.83$ | df = 13 |

## 6.2 Example: Bacteria Death Due to X-Ray Radiation

### Use R for last example

```
> #######################################  Bacterial Examples
> library(latex2exp)
> bact_dat<-read.table('data/P168.txt',header=TRUE)   ## read the data
> plot(bact_dat$t,bact_dat$N_t,xlab="t", ylab=TeX(r'($n_t$)'), main=TeX(r'(Plot of $n_t$ against $t$)'))
```

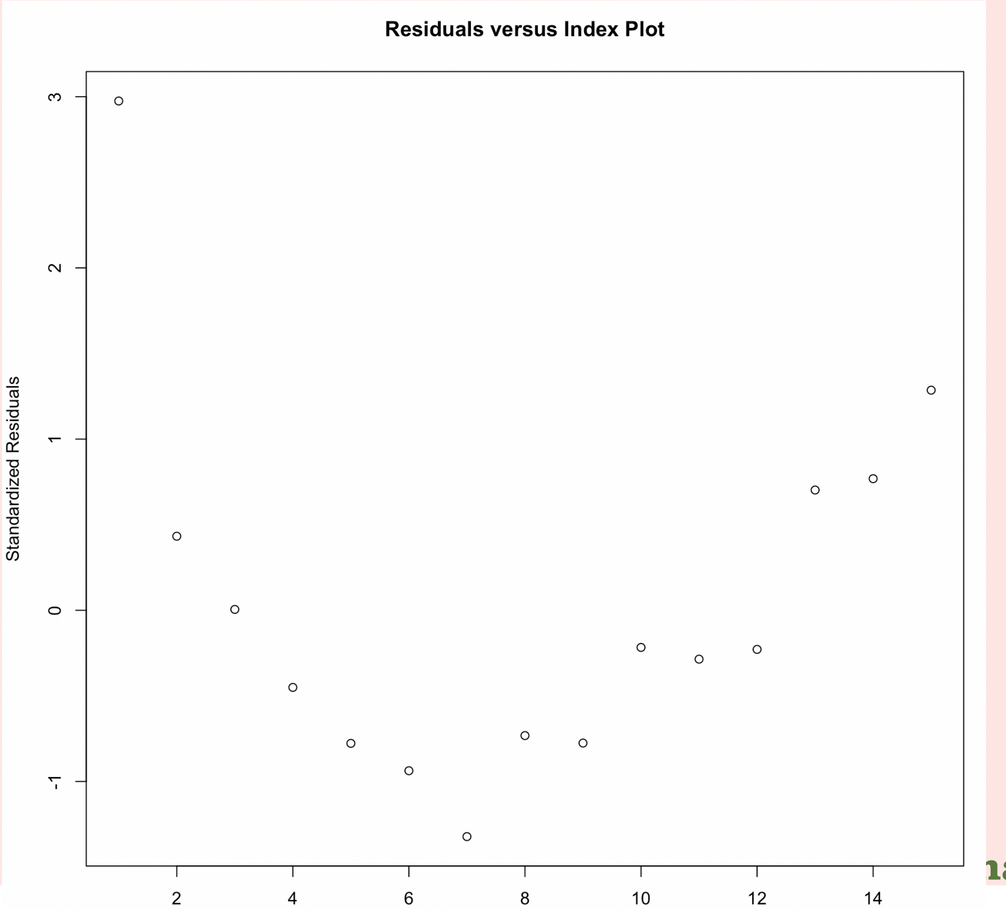


# 6.2 Example: Bacteria Death Due to X-Ray Radiation

## Use R for last example

```
> ################################### summary of linear model of nt = beta_0 + beta_1 t + eps
> mod1<-lm(N_t~.,data=bact_dat)
> summary(mod1)

Call:
lm(formula = N_t ~ ., data = bact_dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.867 -23.599  -9.652  10.223 114.883 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   259.58      22.73  11.420 3.78e-08 ***
t             -19.46       2.50  -7.786 3.01e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41.83 on 13 degrees of freedom
Multiple R-squared:  0.8234,	Adjusted R-squared:  0.8098 
F-statistic: 60.62 on 1 and 13 DF,  p-value: 3.006e-06
```

```
############# plot standardized residuals
pii=hatvalues(mod1)
plot(bact_dat$t,mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),ylab="Standardized Residuals",xlab="t",main="Residuals versus Index Plot")
```



## 6.2 Example: Bacteria Death Due to X-Ray Radiation

### Inadequacy of Linear Model

Despite the fact that the regression coefficient for the time variable is significant and we have a high value of $R^2$, the linear model is **not appropriate**. The plot of $n_t$ against $t$ shows departure from linearity for high values of $t$ (Figure 6.5). We see this even more clearly if we look at a plot of the standardized residuals against time (Figure 6.6). The distribution of residuals has a **distinct pattern**. The residuals for $t = 2$ through 11 are all negative, for $t = 12$ through 15 are all positive, whereas the residual for $t = 1$ appears to be an outlier. This systematic pattern of deviation confirms that the linear model in (6.7) does not fit the data.

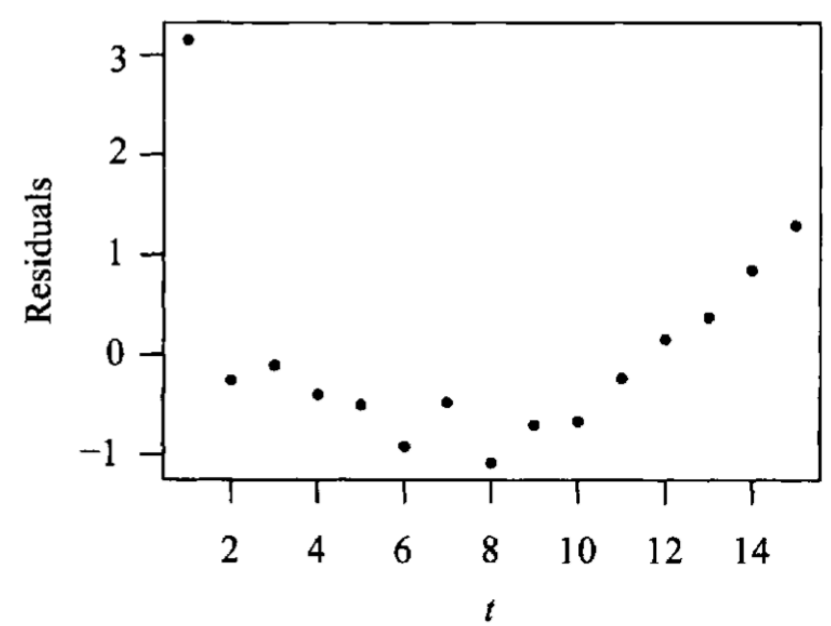



**Figure 6.6** Plot of the standardized residuals from (6.7) against time $t$.

## 6.2 Example: Bacteria Death Due to X-Ray Radiation

### Logarithmic Transformation for Achieving Linearity

The relation between $n_t$ and $t$ appears distinctly nonlinear and we will work with the transformed variable $\ln(n_t)$, which is suggested from theoretical considerations as well as by Figure 6.7. The plot of $\ln(n_t)$ against $t$ appears linear, indicating that the logarithmic transformation is appropriate. The results of fitting (6.5) appear in Table 6.4. The coefficients are highly significant, the standard errors are reasonable, and nearly 99% of the variation in the data is explained by the model. The standardized residuals are plotted against $t$ in Figure 6.8. There are no systematic patterns to the distribution of the residuals and the plot is satisfactory. The single-hit hypothesis of X-ray action, which postulates that $\ln(n_t)$ should be linearly related to $t$, is confirmed by the data.

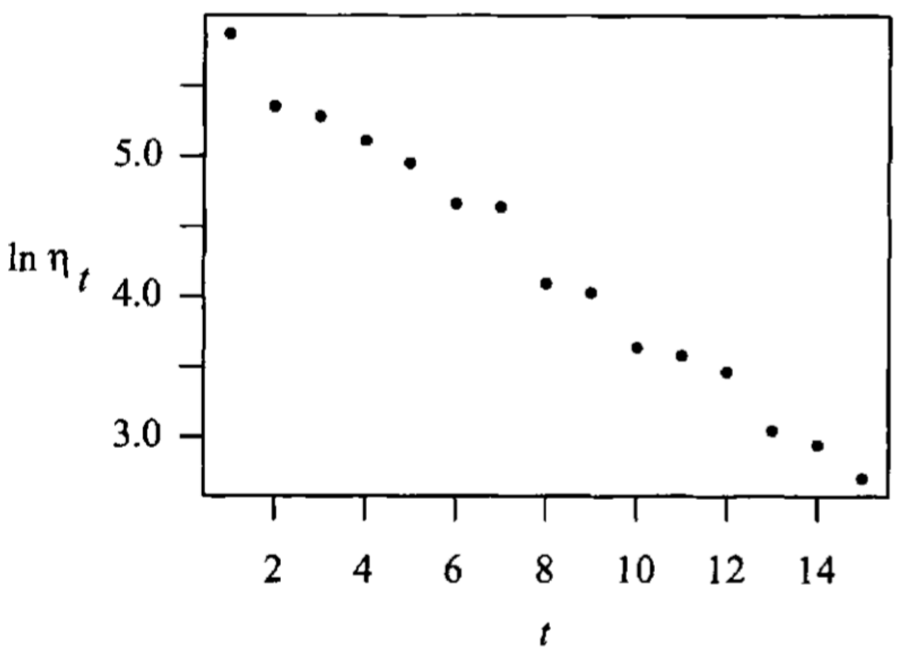



**Figure 6.7** Plot of $\ln n_t$ against time $t$.

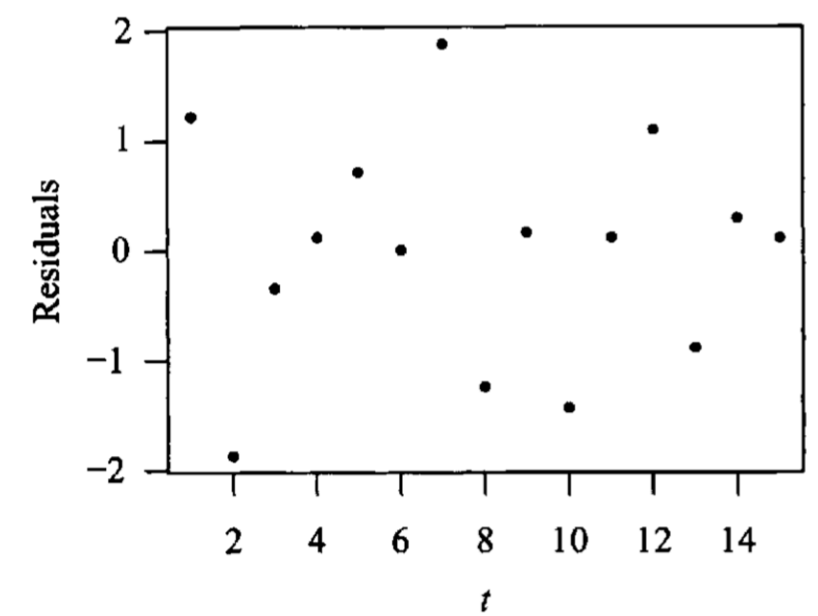



**Figure 6.8** Plot of the standardized residuals against time $t$ after transformation.

# 6.2 Example: Bacteria Death Due to X-Ray Radiation

## Logarithmic Transformation for Achieving Linearity

While working with transformed variables, careful attention must be paid to the estimates of the parameters of the model. In our example the point estimate of $\beta_1$ is $-0.218$ and the 95% confidence interval for the same parameter is $(-0.232, -0.204)$. The estimate of the constant term in the equation is the best linear unbiased estimate of $\ln(n_0)$. If $\hat{\beta}_0$ denotes the estimate, $e^{\hat{\beta}_0}$ may be used as an estimate of $n_0$. With $\hat{\beta}_0 = 5.973$, the estimate of $n_0$ is $e^{\hat{\beta}_0} = 392.68$. This estimate is not an unbiased estimate of $n_0$; that is, the true size of the bacteria population at the start of the experiment was probably somewhat smaller than 392.68. A correction can be made to reduce the bias in the estimate of $n_0$. The estimate $\exp[\hat{\beta}_0 - \mathrm{Var}(\hat{\beta}_0)/2]$ is nearly unbiased of $n_0$. In our present example, the modified estimate of $n_0$ is 391.98. Note that the bias in estimating $n_0$ has no effect on the test of the theory or the estimation of the decay rate.

$$\boxed{\ln(n_t) = \ln(n_0) + \beta_1 t + \varepsilon_t}$$

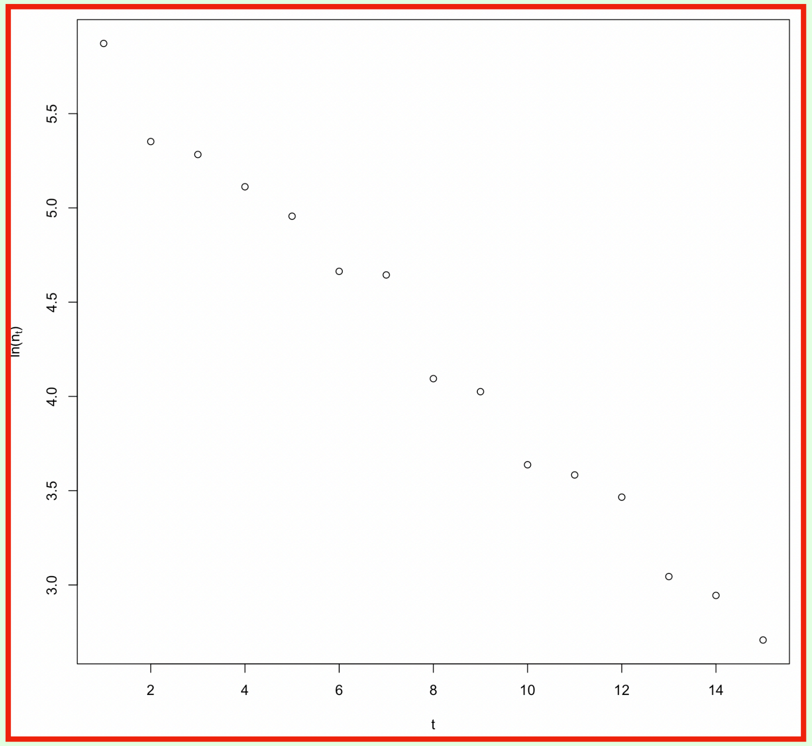
**Table 6.4** Estimated Regression Coefficients When $\ln n_t$ is Regressed on Time $t$

| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | 5.973 | 0.0598 | 99.9 | $< 0.0001$ |
| TIME ($t$) | $-0.218$ | 0.0066 | $-33.2$ | $< 0.0001$ |
| | $n = 15$ | $R^2 = 0.988$ | $\hat{\sigma} = 0.11$ | df = 13 |

# 6.2 Example: Bacteria Death Due to X-Ray Radiation

## Use R for last example

```
> plot(bact_dat$t,bact_dat$log_nt,ylab=TeX(r'($\ln(n_t)$)'),xlab="t")
```

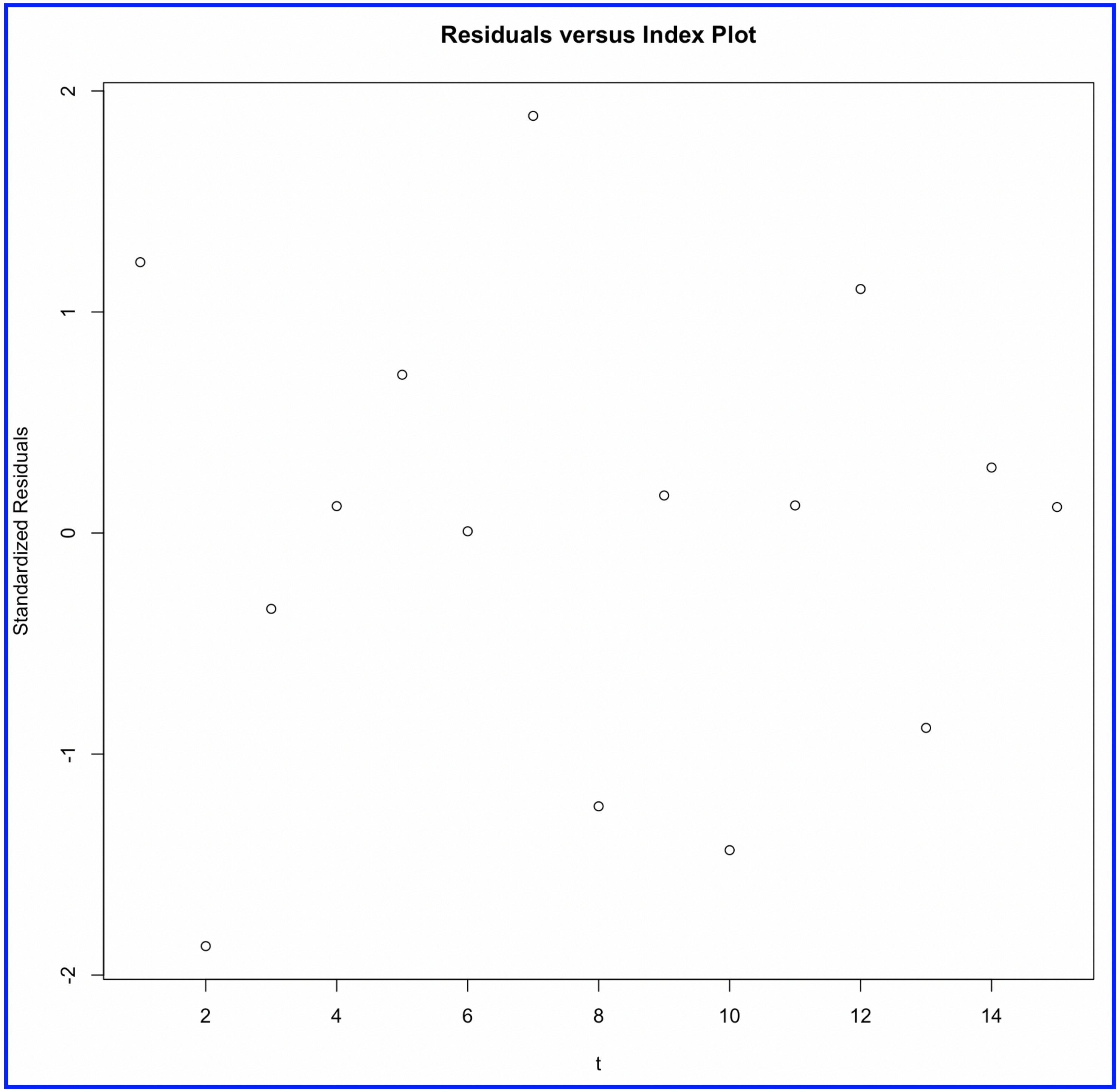



```
> ################################## linear model on transformed variable:  ln(nt) = beta_0 + beta_1 t + eps
> bact_dat$log_nt<-log(bact_dat$N_t)
> mod2<-lm(log_nt~t,data=bact_dat)
> summary(mod2)

Call:
lm(formula = log_nt ~ t, data = bact_dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.18445 -0.06189  0.01253  0.05201  0.20021 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.973160   0.059778   99.92  < 2e-16 ***
t           -0.218425   0.006575  -33.22 5.86e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.11 on 13 degrees of freedom
Multiple R-squared:  0.9884,	Adjusted R-squared:  0.9875 
F-statistic:  1104 on 1 and 13 DF,  p-value: 5.86e-14
```

# 6.2 Example: Bacteria Death Due to X-Ray Radiation

## Use R for last example

```
pii=hatvalues(mod2)
plot(bact_dat$t,mod2$residuals/(summary(mod2)$sigma * sqrt(1-pii)),ylab="Standardized Residuals",xlab="t",main="Residuals versus Index Plot")
```



# 6.3 Transformations to Stabilize Variance

## 6.3 Transformations to Stabilize Variance

We have discussed in the preceding section the use of transformations to **achieve linearity** of the regression function. Transformations are also used to **stabilize the error variance**, that is, to make the error variance constant for all the observations. The constancy of error variance is one of the standard assumptions of least squares theory. It is often referred to as the assumption of **homoscedasticity**. When the error variance is not constant over all the observations, the error is said to be **heteroscedastic**. **Heteroscedasticity** is usually detected by suitable graphs of the residuals such as the scatter plot of the standardized residuals against the fitted values or against each of the predictor variables. A plot with the characteristics of Figure 6.9 typifies the situation. The residuals tend to have a funnel-shaped distribution, either fanning out or closing in with the values of X.

If heteroscedasticity is present and no corrective action is taken, application of the ordinary least squares to the raw data will result in estimated coefficients which **lack precision** in a theoretical sense. The estimated standard errors of the regression coefficients are often understated, giving a false sense of accuracy.

Heteroscedasticity can be removed by means of a suitable transformation. We describe an approach for (a) detecting heteroscedasticity and its effects on the analysis and (b) removing heteroscedasticity from the data analyzed using transformations.

Figure 6.9

# 6.3 Transformations to Stabilize Variance

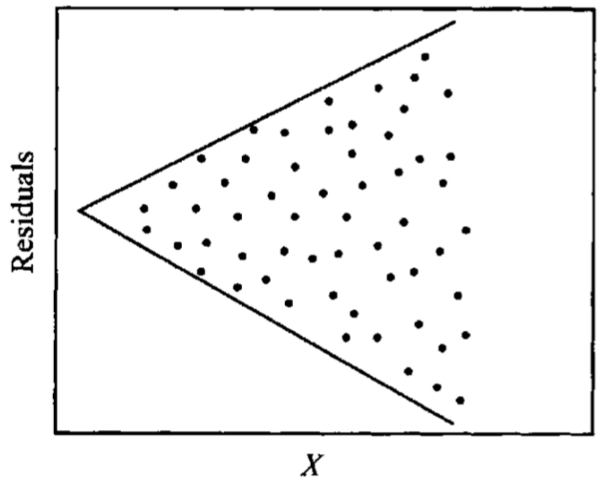



**Figure 6.9** An example of heteroscedastic residuals.

If $Y \sim \text{Poisson}(\mu)$, we have $E(Y) = \mu$ and $\text{Var}(Y) = \mu$.

If $Y \sim \text{Binomial}(n, \mu/n)$, we have $E(Y) = \mu$ and $\text{Var}(Y) = \mu(1 - \mu/n)$.

In linear regression, $Y = \beta_0 + \beta_1 X + \varepsilon$, given $X = x$, we have $E(Y) = \beta_0 + \beta_1 x$ and $\text{Var}(Y) = \text{Var}(\varepsilon) = \sigma^2$, **if the assumption of homogeneous variance holds.**

**Table 6.5** Transformations to Stabilize Variance

| Probability Distribution of $Y$ | $\text{Var}(Y)$ in Terms of Its Mean $\mu$ | Transformation | Resulting Variance |
|---|---|---|---|
| Poisson[a] | $\mu$ | $\sqrt{Y}$ or $(\sqrt{Y} + \sqrt{Y+1})$ | 0.25 |
| Binomial[b] | $\mu(1 - \mu/n)$ | $\sin^{-1}\sqrt{Y}$ (degrees) | $821/n$ |
| | | $\sin^{-1}\sqrt{Y}$ (radians) | $0.25/n$ |
| Negative Binomial | $\mu + \lambda^2\mu^2$ | $\lambda^{-1}\sinh^{-1}(\lambda\sqrt{Y})$ or | |
| | | $\lambda^{-1}\sinh^{-1}(\lambda\sqrt{Y} + 0.5)$ | 0.25 |

a. For small values of $Y$, $\sqrt{Y + 0.5}$ is sometimes recommended.
b. The sample size is denoted by $n$;

# 6.3 Transformations to Stabilize Variance

The response variable Y, in a regression problem, may follow a probability distribution whose variance is a function of the mean of that distribution. One property of the normal distribution, that many other probability distributions do not have, is that its mean and variance are independent in the sense that one is not a function of the other. The Binomial and Poisson are but two examples of common probability distributions that have this characteristic. We know, for example, that a variable that is distributed Binomially with parameters $n$ and $\pi$ has mean $n\pi$ and variance $n\pi(1-\pi)$. It is also known that the mean and variance of a Poisson random variable are equal. When the relationship between the mean and variance of a random variable is known, it is possible to find a simple transformation of the variable, which makes the variance approximately constant (stabilizes the variance). We list in Table 6.5, for convenience and easy reference, transformations that stabilize the variance for some random variables with commonly occurring probability distributions whose variances are functions of their means. The transformations listed in Table 6.5 not only stabilize the variance, but also have the effect of making the distribution of the transformed variable closer to the normal distribution. Consequently, these transformations serve the **dual purpose** of normalizing the variable as well as making the variance functionally independent of the mean.

Table 6.5

## 6.3 Transformations to Stabilize Variance

As an illustration, consider the following situation: Let $Y$ be the number of accidents and $X$ the speed of operating a lathe in a machine shop. We want to study the relationship between the number of accidents $Y$ and the speed of lathe operation $X$. Suppose that a linear relationship is postulated between $Y$ and $X$ and is given by

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where $\varepsilon$ is the random error. The mean of $Y$ is seen to increase with $X$. It is known from empirical observation that rare events (events with small probabilities of occurrence) often have a Poisson distribution.

For a given X, the mean of $Y$ is $E(Y) = \beta_0 + \beta_1 X$ is a function of $X$.

Let us assume that $Y$ has a Poisson distribution. Since the mean and variance of $Y$ are the same, it follows that the variance of $Y$ is a function of $X$, and consequently the assumption of homoscedasticity will not hold. From Table 6.5 we see that the square root of a Poisson variable ($\sqrt{Y}$) has a variance independent of the mean and is approximately equal to 0.25. To ensure homoscedasticity we, therefore, regress $\sqrt{Y}$ on $X$. Here the transformation is chosen to stabilize the variance, the specific form being suggested by the assumed probability distribution of the response variable. An analysis of data employing transformations suggested by probabilistic considerations is demonstrated in the following example.

Example

## 6.3 Transformations to Stabilize Variance

### Example: Injury Incidents in Airlines

The number of injury incidents and the proportion of total flights from New York for nine ($n = 9$) major U.S.A. airlines for a single year is given in Table 6.6 and plotted in Figure 6.10. Let $f_i$ and $y_i$ denote the total flights and the number of injury incidents for the $i$th airline that year. Then the proportion of total flights $n_i$ made by the $i$th airline is

$$n_i = \frac{f_i}{\sum_i f_i}$$

**Table 6.6** Number of Injury Incidents $Y$ and Proportion of Total Flights $N$

| Row | $Y$ | $N$ | Row | $Y$ | $N$ | Row | $Y$ | $N$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 11 | 0.0950 | 4 | 19 | 0.2078 | 7 | 3 | 0.1292 |
| 2 | 7 | 0.1920 | 5 | 9 | 0.1382 | 8 | 1 | 0.0503 |
| 3 | 7 | 0.0750 | 6 | 4 | 0.0540 | 9 | 3 | 0.0629 |

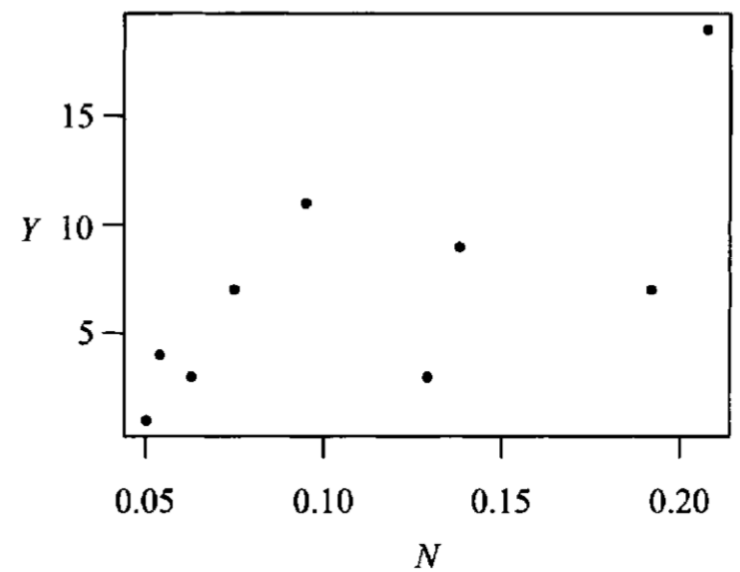



**Figure 6.10** Plot of $Y$ against $N$.

# 6.3 Transformations to Stabilize Variance

## Example: Injury Incidents in Airlines

If all the airlines are equally safe, the injury incidents can be explained by the model

$$y_i = \beta_0 + \beta_1 n_i + \varepsilon_i$$

where $\beta_0$ and $\beta_1$ are constants and $\varepsilon_i$ is the random error.

The results of fitting the model are given in Table 6.7. The plot of residuals against $n_i$ is given in Figure 6.11. The residuals are seen to increase with $n_i$ in Figure 6.11 and, consequently, the assumption of homoscedasticity seems to be violated. This is not surprising, since the injury incidents may behave as a Poisson variable which has a variance proportional to its mean. To ensure the assumption of homoscedasticity, we make the square root transformation. Instead of working with $Y$ we work with $\sqrt{Y}$ , a variable which has an approximate variance of 0.25, and is more normally distributed than the original variable.

**Table 6.7** Estimated Regression Coefficients (When $Y$ is Regressed on $N$)

| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | −0.14 | 3.14 | −0.045 | 0.9657 |
| $N$ | 64.98 | 25.20 | 2.580 | 0.0365 |
| | $n = 9$ | $R^2 = 0.487$ | $\hat{\sigma} = 4.201$ | df = 7 |

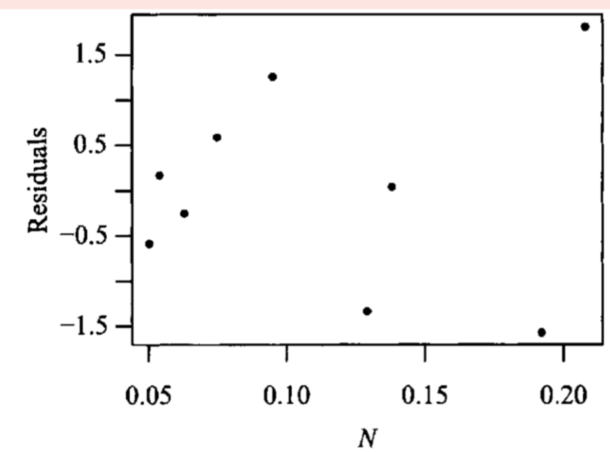
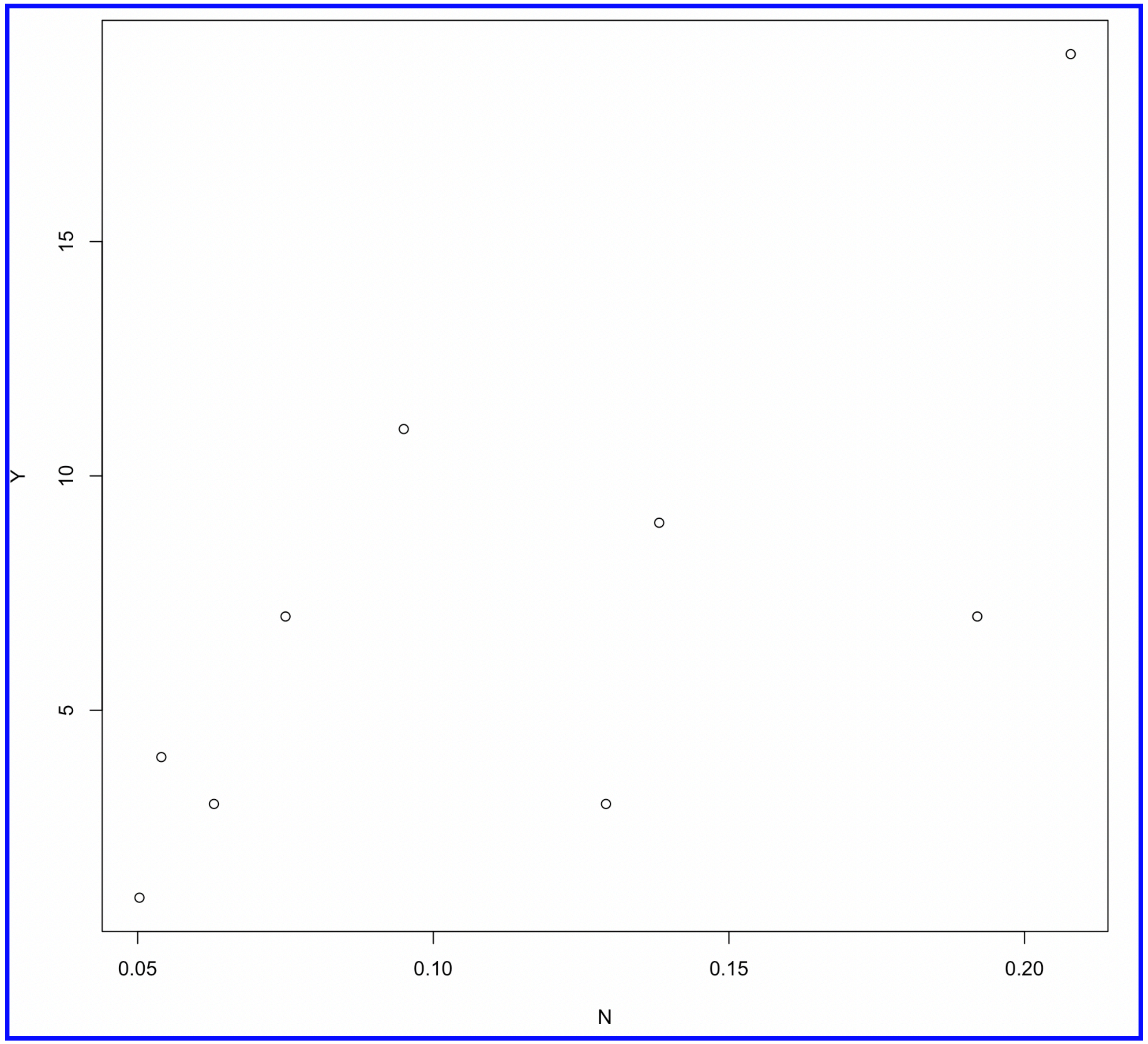



**Figure 6.11** Plot of the standardized residuals versus $N$.

# 6.3 Transformations to Stabilize Variance

## Use R for results on last slide

```
> #####################################  Example of Injury Accidents of Airlines
> inj_dat<-read.table('data/P174.txt',header=TRUE)   ## read the data
> plot(inj_dat$N, inj_dat$Y,xlab="N",ylab="Y")
```

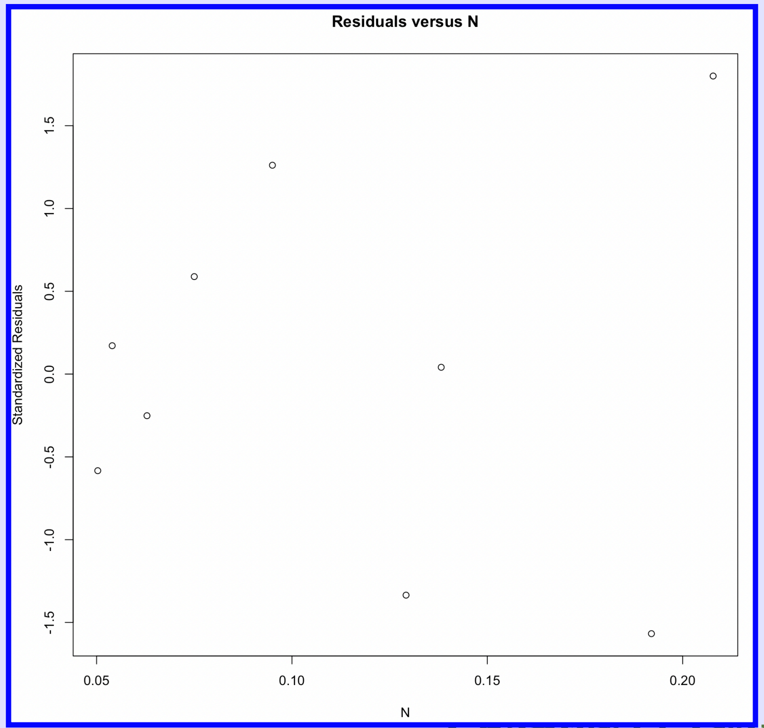


# 6.3 Transformations to Stabilize Variance

## Use R for results on last slide

```
> ######################################  Linear model on the original data
> mod1<-lm(Y~N,data=inj_dat)
> summary(mod1)

Call:
lm(formula = Y ~ N, data = inj_dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.3351 -2.1281  0.1605  2.2670  5.6382 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -0.1402     3.1412  -0.045   0.9657  
N            64.9755    25.1959   2.579   0.0365 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.201 on 7 degrees of freedom
Multiple R-squared:  0.4872,    Adjusted R-squared:  0.4139 
F-statistic:  6.65 on 1 and 7 DF,  p-value: 0.03654
```

```
pii<-hatvalues(mod1)
plot(inj_dat$N,mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="N",ylab="Standardized Residuals",main="Residuals versus N")
```



# 6.3 Transformations to Stabilize Variance

## Example: Injury Incidents in Airlines

Consequently, the model we fit is

$$\sqrt{y_i} = \beta_0' + \beta_1' n_i + \varepsilon_i. \tag{6.8}$$

The result of fitting (6.8) is given in Table 6.8. The residuals from (6.8) when plotted against $n_i$ are shown in Figure 6.12. The residuals for the transformed model do not seem to increase with $n_i$. This suggests that for the transformed model the homoscedastic assumption is not violated. The analysis of the model in terms of $\sqrt{y_i}$ and $n_i$ can now proceed using standard techniques. The regression is significant here (as judged by the t statistic) but is not very strong. Only 48% of the total variability of the injury incidents of the airlines is explained by the variation in their number of flights. It appears that for a better explanation of injury incidents other factors have to be considered.

**Table 6.8** Estimated Regression Coefficients When $\sqrt{Y}$ is Regressed on $N$

| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | 1.169 | 0.578 | 2.02 | 0.0829 |
| $N$ | 11.856 | 4.638 | 2.56 | 0.0378 |
| | $n = 9$ | $R^2 = 0.483$ | $\hat{\sigma} = 0.773$ | df = 7 |

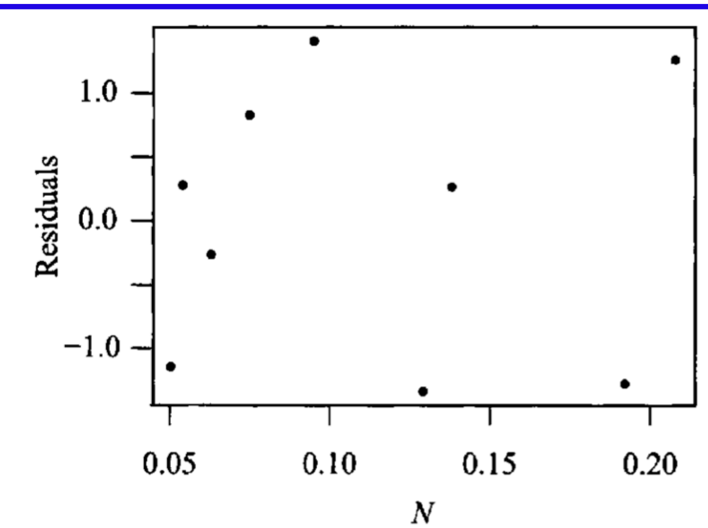


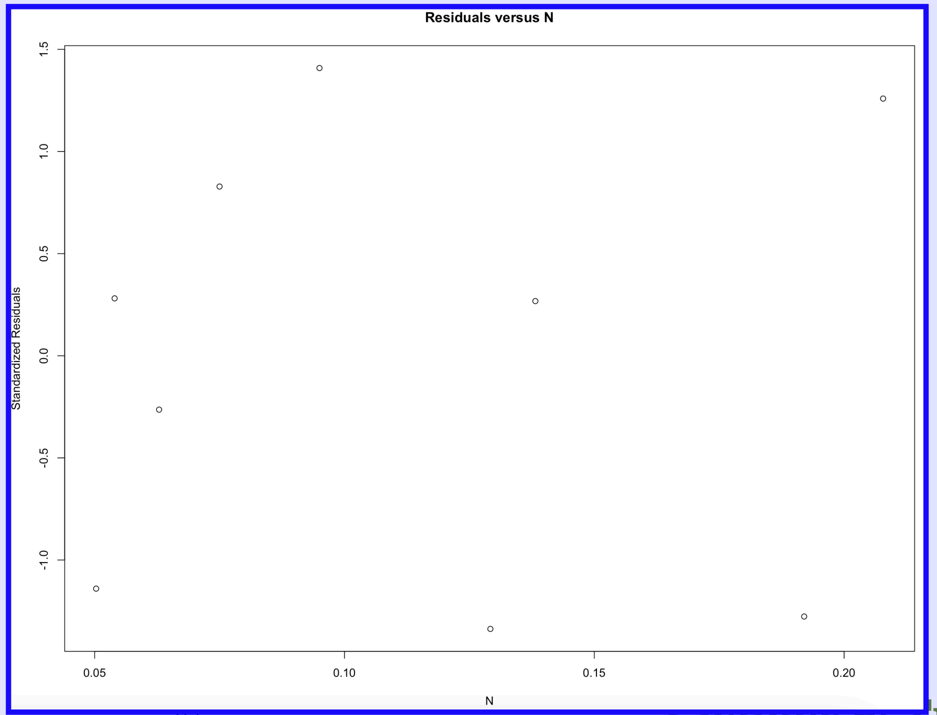

**Figure 6.12** Plot of the standardized residuals from the regression of $\sqrt{y_i}$ on $n_i$.

# 6.3 Transformations to Stabilize Variance

## Use R for results on last slide

```
> ####################################  Linear model using sqrt(Y)
> inj_dat$sqrtY<-sqrt(inj_dat$Y)
> mod2<-lm(sqrtY~N, data=inj_dat)
> summary(mod2)

Call:
lm(formula = sqrtY ~ N, data = inj_dat)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9690 -0.7655  0.1906  0.5874  1.0211

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.1692     0.5783   2.022   0.0829 .
N            11.8564     4.6382   2.556   0.0378 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7733 on 7 degrees of freedom
Multiple R-squared:  0.4828,	Adjusted R-squared:  0.4089
F-statistic: 6.535 on 1 and 7 DF,  p-value: 0.03776
```

```
pii<-hatvalues(mod2)
plot(inj_dat$N,mod2$residuals/(summary(mod2)$sigma * sqrt(1-pii)),xlab="N",ylab="Standardized Residuals",main="Residuals versus N")
```



## 6.3 Transformations to Stabilize Variance

In the preceding example the nature of the response variable (injury incidents) suggested that the error variance was not constant about the fitted line. The square root transformation was considered based on the well-established empirical fact that the occurrence of accidents tend to follow the Poisson probability distribution. For Poisson variables, the square root is the appropriate transformation (Table 6.5). There are situations, however, when the error variance is not constant and there is no a priori reason to suspect that this would be the case. Empirical analysis will reveal the problem, and by making an appropriate transformation this effect can be eliminated. If the unequal error variance is not detected and eliminated, the resulting estimates will have large standard errors, but will be unbiased. This will have the effect of producing wide confidence intervals for the parameters and tests with low sensitivity. We illustrate the method of analysis for a model with this type of heteroscedasticity in the next section.

## 6.4 Detection and Removal of Heteroscedastic Errors

# 6.4 Detection and Removal of Heteroscedastic Errors

## Detection

In a study of 27 industrial establishments of varying size, the number of supervised workers ($X$) and the number of supervisors ($Y$) were recorded (Table 6.9). It was decided to study the relationship between the two variables, and as a start a linear model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (6.9)$$

was postulated. A plot of $Y$ versus $X$ suggests a simple linear model as a starting point (Figure 6.13). The results of fitting the linear model are given in Table 6.10.

Table 6.10

**Table 6.9** Number of Supervised Workers and Supervisors in 27 Industrial Establishments

| Row | $X$ | $Y$ | Row | $X$ | $Y$ | Row | $X$ | $Y$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 294 | 30 | 10 | 697 | 78 | 19 | 700 | 106 |
| 2 | 247 | 32 | 11 | 688 | 80 | 20 | 850 | 128 |
| 3 | 267 | 37 | 12 | 630 | 84 | 21 | 980 | 130 |
| 4 | 358 | 44 | 13 | 709 | 88 | 22 | 1025 | 160 |
| 5 | 423 | 47 | 14 | 627 | 97 | 23 | 1021 | 97 |
| 6 | 311 | 49 | 15 | 615 | 100 | 24 | 1200 | 180 |
| 7 | 450 | 56 | 16 | 999 | 109 | 25 | 1250 | 112 |
| 8 | 534 | 62 | 17 | 1022 | 114 | 26 | 1500 | 210 |
| 9 | 438 | 68 | 18 | 1015 | 117 | 27 | 1650 | 135 |

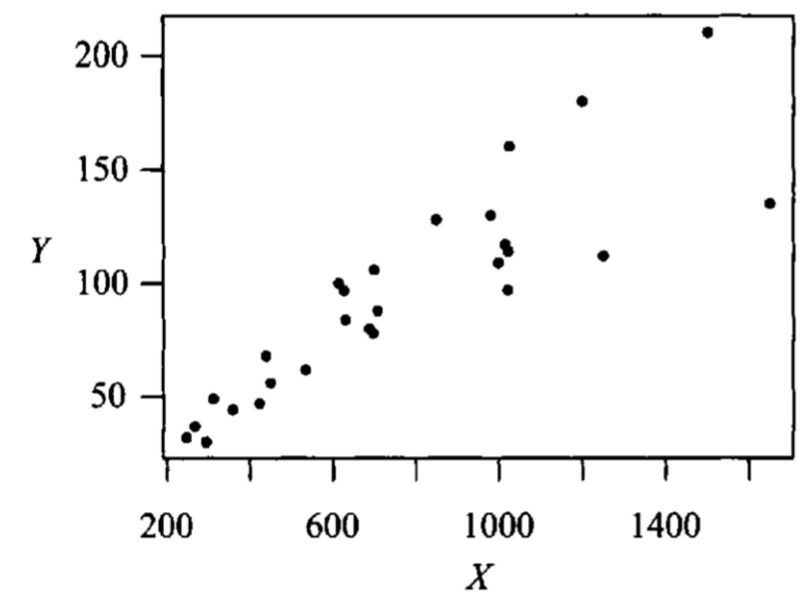



**Figure 6.13** Number of supervisors ($Y$) versus number supervised ($X$).

# 6.4 Detection and Removal of Heteroscedastic Errors

## Detection

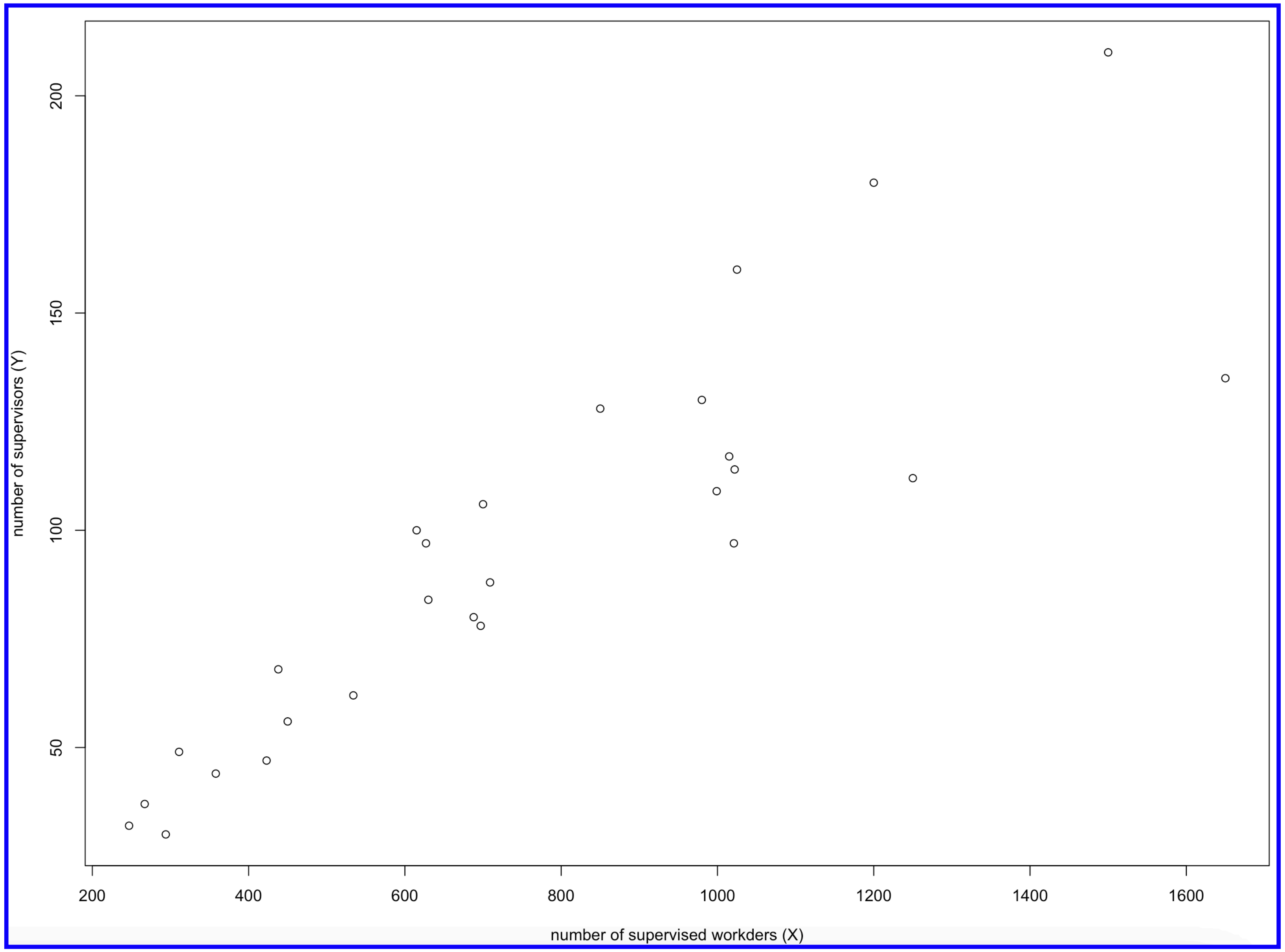
**Table 6.10** Estimated Regression Coefficients When Number of Supervisors ($Y$) is Regressed on the Number Supervised ($X$)

| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | 14.448 | 9.562 | 1.51 | 0.1433 |
| $X$ | 0.105 | 0.011 | 9.30 | $< 0.0001$ |
| | $n = 27$ | $R^2 = 0.776$ | $\hat{\sigma} = 21.73$ | df = 25 |

# 6.4 Detection and Removal of Heteroscedastic Errors

## Use R for last example

```
####################################  Example of Supervisor and worker dataset
supv_dat<-read.table('data/P176.txt',header=TRUE)    ## read the data
plot(supv_dat$X,supv_dat$Y,xlab="number of supervised workders (X)",ylab="number of supervisors (Y)")
```

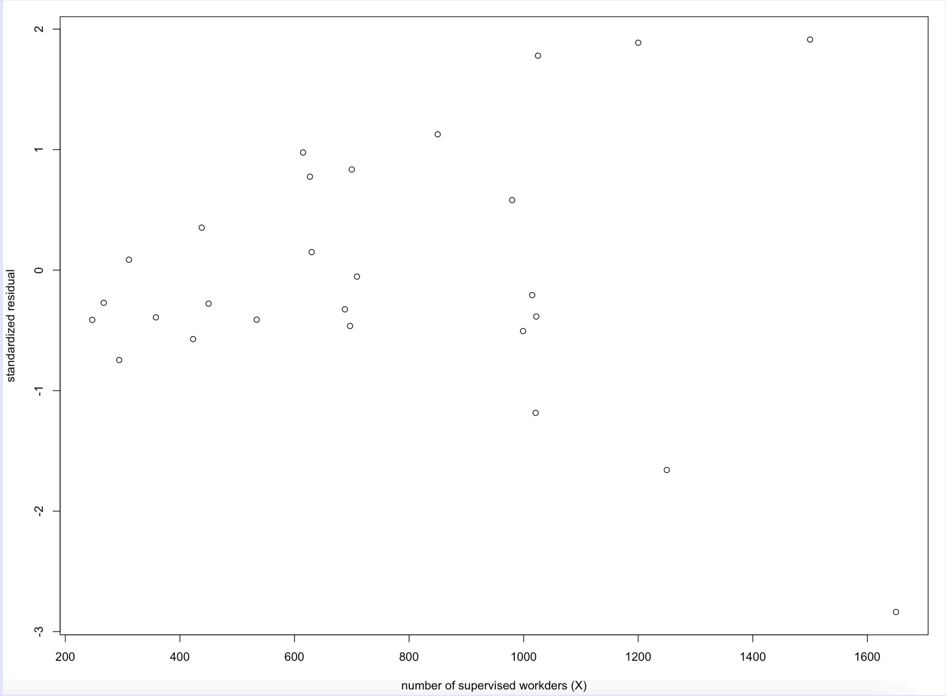


# 6.4 Detection and Removal of Heteroscedastic Errors

## Use R for last example

```
> ############# linear model on Y = beta_0 + beta_1 X +epsilon
> mod1<-lm(Y~X,data=supv_dat)
> summary(mod1)

Call:
lm(formula = Y ~ X, data = supv_dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-53.294  -9.298  -5.579  14.394  39.119 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14.44806    9.56201   1.511    0.143    
X            0.10536    0.01133   9.303 1.35e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.73 on 25 degrees of freedom
Multiple R-squared:  0.7759,    Adjusted R-squared:  0.7669 
F-statistic: 86.54 on 1 and 25 DF,  p-value: 1.35e-09
```

```
pii=hatvalues(mod1)
plot(supv_dat$X,mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="number of supervised workders (X)",ylab="standardized residual")
```



# 6.4 Detection and Removal of Heteroscedastic Errors

## Detection

The plot of the standardized residuals versus $X$ (Figure 6.14) shows that the residual variance tends to increase with $X$. The residuals tend to lie in a band that diverges as one moves along the $X$ axis. In general, if the band within which the residuals lie diverges (i.e., becomes wider) as $X$ increases, the error variance is also increasing with $X$. On the other hand, if the band converges (i.e., becomes narrower), the error variance decreases with $X$. If the band that contains the residual plots consists of two lines parallel to the $X$ axis, there is no evidence of heteroscedasticity. A plot of the standardized residuals against the predictor variable points up the presence of heteroscedastic errors. As can be seen in Figure 6.14, in our present example the residuals tend to increase with $X$.

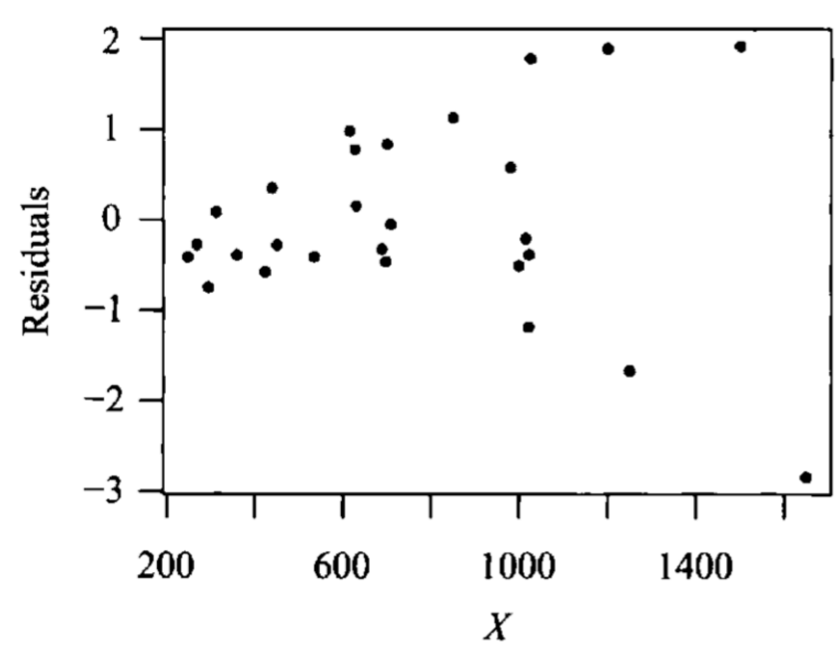



**Figure 6.14** Plot of the standardized residuals against $X$ when number of supervisors ($Y$) is regressed on the number supervised ($X$).

## 6.4 Detection and Removal of Heteroscedastic Errors

### Removal

In many industrial, economic, and biological applications, when unequal error variances are encountered, it is often found that the standard deviation of residuals tends to increase as the predictor variable increases. Based on this empirical observation, we will hypothesize in the present example that the standard deviation of the residuals is proportional to $X$ (some indication of this is available from the plot of the residuals in Figure 6.14):

$$\mathrm{Var}(\varepsilon_i) = k^2 x_i^2, \quad k > 0. \tag{6.10}$$

Dividing both sides of (6.9) by $x_i$, we obtain

$$\frac{y_i}{x_i} = \frac{\beta_0}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i}\,. \tag{6.11}$$

Now, define a new set of variables and coefficients,

$$Y' = \frac{Y}{X}, \quad X' = \frac{1}{X}, \quad \beta_0' = \beta_1, \quad \beta_1' = \beta_0, \quad \varepsilon' = \frac{\varepsilon}{X}.$$

In terms of the new variables (6.11) reduces to

$$y_i' = \beta_0' + \beta_1' x_i' + \varepsilon_i'. \tag{6.12}$$

# 6.4 Detection and Removal of Heteroscedastic Errors

## Removal

Note that for the transformed model, $\mathrm{Var}(\varepsilon_i')$ is constant and equals $k^2$. If our assumption about the error term as given in (6.10) holds, to fit the model properly we must work with the transformed variables: $Y/X$ and $1/X$ as **response** and **predictor** variables, respectively. If the fitted model for the transformed data is $\hat{\beta}_0' + \hat{\beta}_1'/X$, the fitted model in terms of the original variables is

$$\frac{\hat{Y}}{X} = \hat{\beta}_0' + \frac{\hat{\beta}_1'}{X} \Longrightarrow \qquad \hat{Y} = \hat{\beta}_1' + \hat{\beta}_0' X. \tag{6.13}$$

The constant in the transformed model is the regression coefficient of $X$ in the original model, and vice versa. This can be seen from comparing (6.11) and (6.12)

The residuals obtained after fitting the transformed model are plotted against the predictor variable in Figure 6.15. It is seen that the residuals are randomly distributed and lie roughly within a band parallel to the horizontal axis. There is no marked evidence of heteroscedasticity in the transformed model. The distribution of residuals shows no distinct pattern and we conclude that the transformed model is adequate. Our assumption about the error term appears to be correct; the transformed model has homoscedastic errors and the standard assumptions of least squares theory hold. The result of fitting $Y/X$ and $1/X$ leads to estimates of $\beta_0'$ and $\beta_1'$ which can be used for the original model.

Figure 6.15

## 6.4 Detection and Removal of Heteroscedastic Errors

### Removal

The equation for the transformed variables is $Y/X = 0.121 + 3.803/X$. In terms of the original variables, we have $Y = 3.803 + 0.121X$. The results are summarized in Table 6.11. By comparing Tables 6.10 and 6.11 we see the reduction in standard errors that is accomplished by working with transformed variables. The variance of the estimate of the slope is reduced by 33%.

**Table 6.11** Estimated Regression Coefficients of the Original Equation When Fitted by the Transformed Variables $Y/X$ and $1/X$

| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | 3.803 | 4.570 | 0.832 | 0.4131 |
| $X$ | 0.121 | 0.009 | 13.44 | $< 0.0001$ |
| | $n = 27$ | $R^2 = 0.758$ | $\hat{\sigma} = 21.577$ | df = 25 |

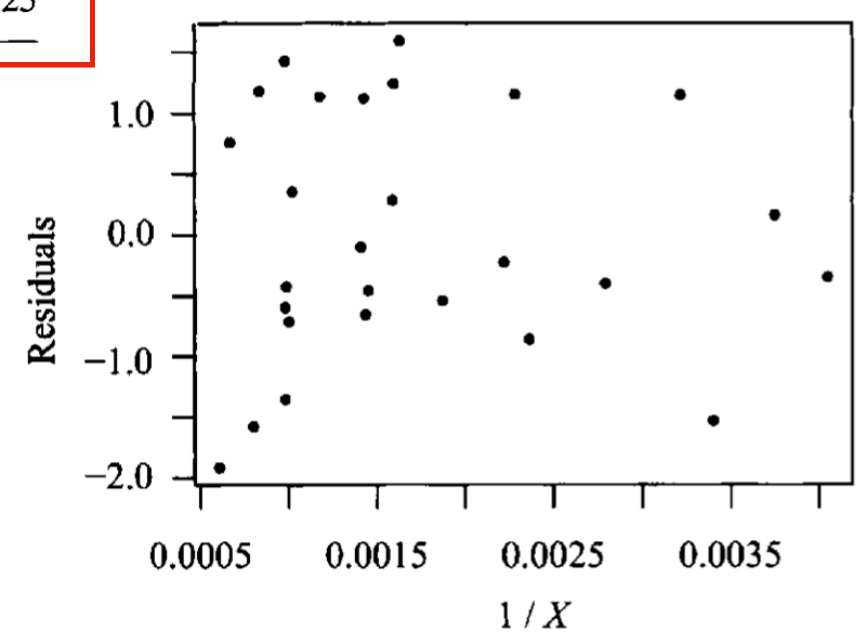


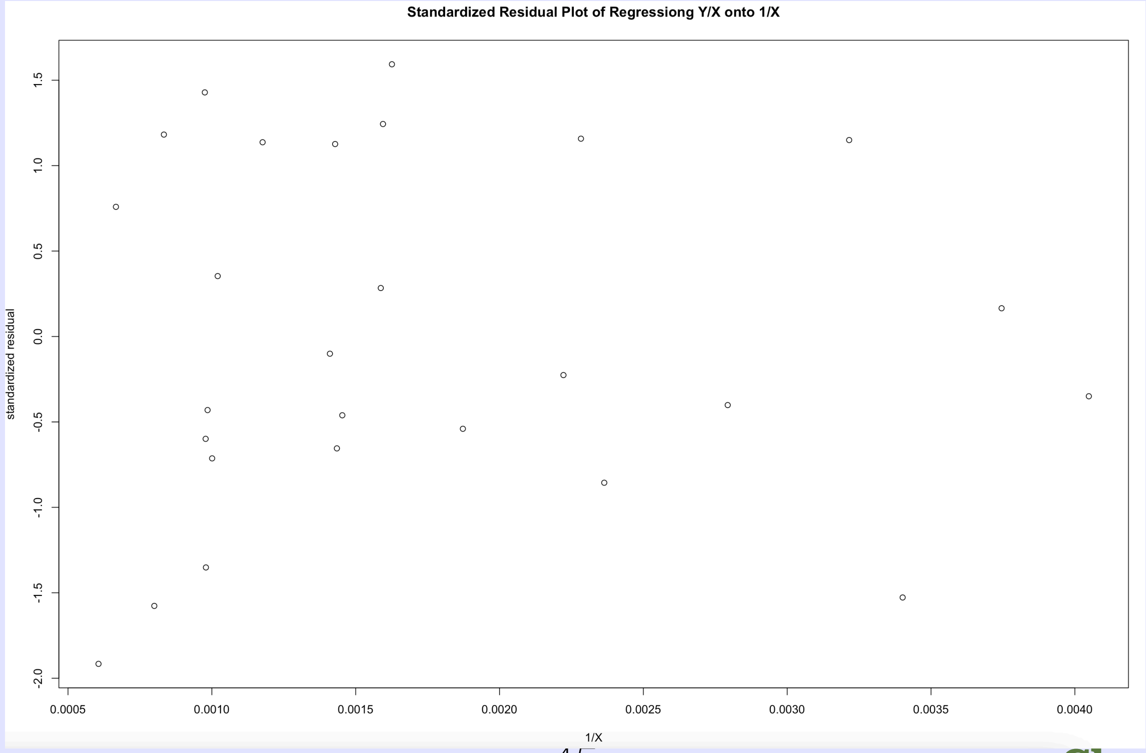

**Figure 6.15** Plot of the standardized residuals against $1/X$ when $Y/X$ is regressed on $1/X$.

# 6.4 Detection and Removal of Heteroscedastic Errors

## Use R for last example

```
> ##############  Transform the variable
> supv_dat$invX<-1/supv_dat$X
> supv_dat$newY<-supv_dat$Y / supv_dat$X
> mod2<-lm(newY~invX,data=supv_dat)
> summary(mod2)

Call:
lm(formula = newY ~ invX, data = supv_dat)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.041477 -0.013852 -0.004998  0.024671  0.035427 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.120990   0.008999  13.445 6.04e-13 ***
invX        3.803296   4.569745   0.832    0.413    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02266 on 25 degrees of freedom
Multiple R-squared:  0.02696,	Adjusted R-squared:  -0.01196 
F-statistic: 0.6927 on 1 and 25 DF,  p-value: 0.4131
```

```
pii=hatvalues(mod2)
plot(supv_dat$invX,mod2$residuals/(summary(mod2)$sigma * sqrt(1-pii)),xlab="1/X",ylab="standardized residual",main="Standardized Residual Plot of Regressiong Y/X onto 1/X")
```



## 6.5 Weighted Least Squares

## 6.5 Weighted Least Squares

Linear regression models with heteroscedastic errors can also be fitted by a method called the *weighted least squares* (WLS), where parameter estimates are obtained by minimizing a weighted sum of squares of residuals where the weights are inversely proportional to the variance of the errors. This is in contrast to ordinary least squares (OLS), where the parameter estimates are obtained by minimizing equally weighted sum of squares of residuals. In the preceding example, the WLS estimates are obtained by minimizing

$$\sum \frac{1}{x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2 \tag{6.14}$$

as opposed to minimizing

$$\sum (y_i - \beta_0 - \beta_1 x_i)^2. \tag{6.15}$$

It can be shown that WLS is equivalent to performing OLS on the transformed variables $Y/X$ and $1/X$.

## 6.5 Weighted Least Squares

In weighted least squares, our goal is to minimize

$$\sum_{i=1}^{n} \frac{1}{x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2$$

with respect to $\beta_0$ and $\beta_1$.

We can equivalently write

$$\sum_{i=1}^{n} \frac{1}{x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} \left( \frac{y_i}{x_i} - \beta_0 \cdot \frac{1}{x_i} - \beta_1 \right)^2$$

If we transform the data and denote $y_i' = \frac{y_i}{x_i}$ and $x_i' = x_i$, the goal is equivalent to minimizing

$$\sum_{i=1}^{n} (y_i' - \beta_0 x_i' - \beta_1)^2$$

which becomes a standard least square estimate.

## 6.6 Logarithmic Transformation and Power Transformation

# 6.6 Logarithmic Transformation and Power Transformation

## Logarithmic Transformation

The logarithmic transformation is one of the most widely used transformations in regression analysis. Instead of working directly with the data, the statistical analysis is carried out on the logarithms of the data. This transformation is particularly useful when the variable analyzed has a **large** standard deviation compared to its mean. Working with the data on a log scale often has the effect of **dampening variability** and **reducing asymmetry**. This transformation is also effective in **removing heteroscedasticity**. We illustrate this point by using the industrial data given in Table 6.9, where heteroscedasticity has already been detected. Besides illustrating the use of log (logarithmic) transformation to remove heteroscedasticity, we also show in this example that for a given body of data there may exist several adequate descriptions (models).

Instead of fitting the model given in (6.9), we now fit the model

$$\ln y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{6.16}$$

(i.e., instead of regressing $Y$ on $X$, we regress $\ln Y$ on $X$). The corresponding scatter plot is given in Figure 6.16. The results of fitting (6.16) are given in Table 6.12. The coefficients are significant, and the value of $R^2$ (0.77) is comparable to that obtained from fitting the model given in (6.9).

Figure 6.16, Table 6.12

# 6.6 Logarithmic Transformation and Power Transformation

## Logarithmic Transformation

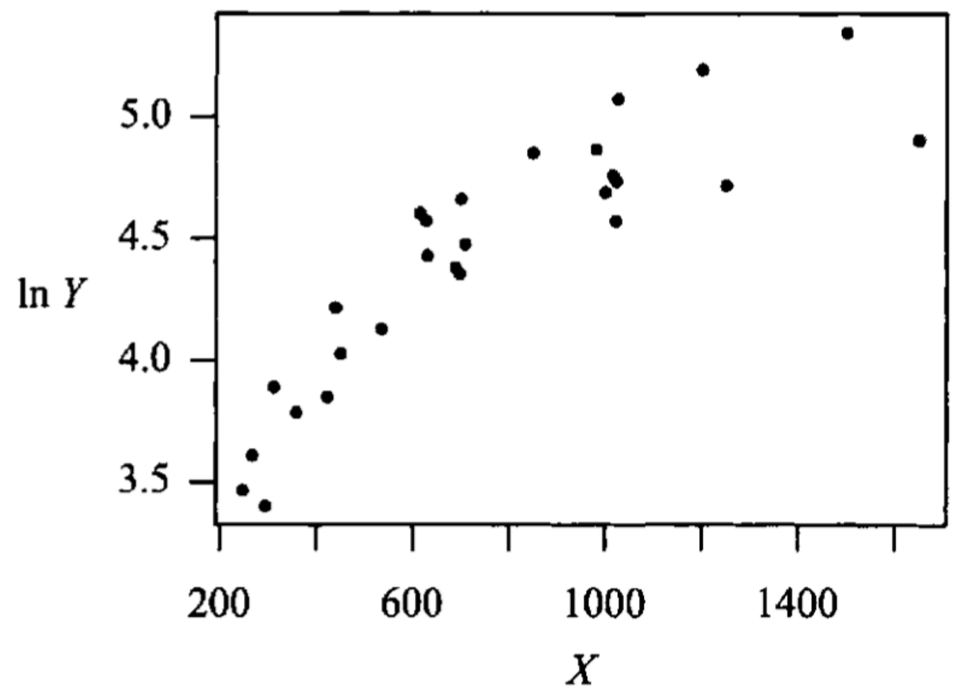



**Figure 6.16** Scatter plot of $\ln Y$ versus $X$.

**Table 6.12** Estimated Regression Coefficients When $\ln Y$ is Regressed on $X$

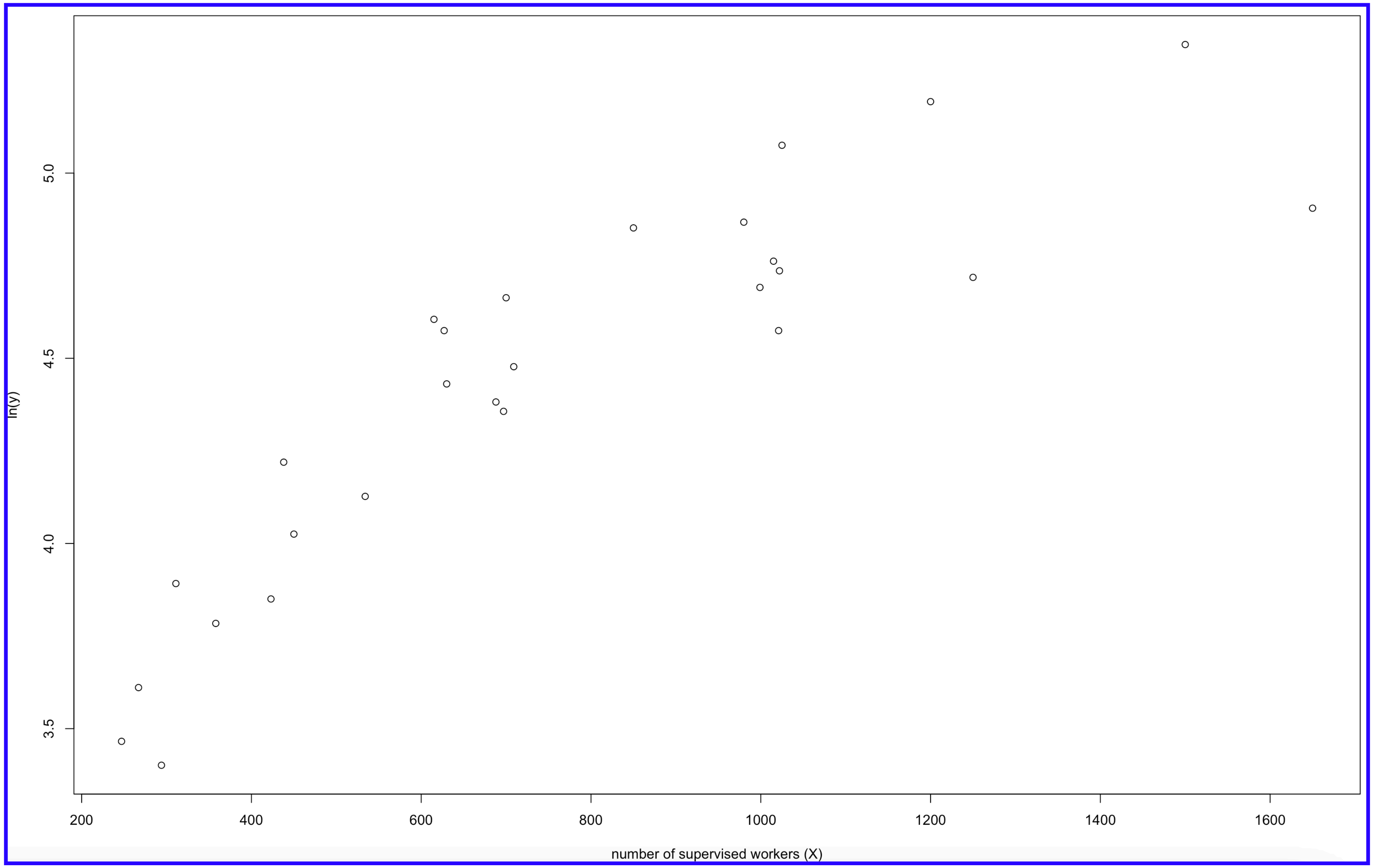
| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | 3.5150 | 0.1110 | 31.65 | $< 0.0001$ |
| $X$ | 0.0012 | 0.0001 | 9.15 | $< 0.0001$ |
| | $n = 27$ | $R^2 = 0.77$ | $\hat{\sigma} = 0.252$ | df = 25 |

# 6.6 Logarithmic Transformation and Power Transformation

## Use R for last example

on the model $\ln(Y) = \beta_0 + \beta_1 X + \varepsilon$

```
########################################### Log transformation on the Supervisors and supervisee dataset
supv_dat<-read.table('data/P176.txt',header=TRUE)    ## read the data
supv_dat$lny<-log(supv_dat$Y)
plot(supv_dat$X,supv_dat$lny,xlab="number of supervised workers (X)",ylab=TeX(r'($\ln(y)$)'))
```



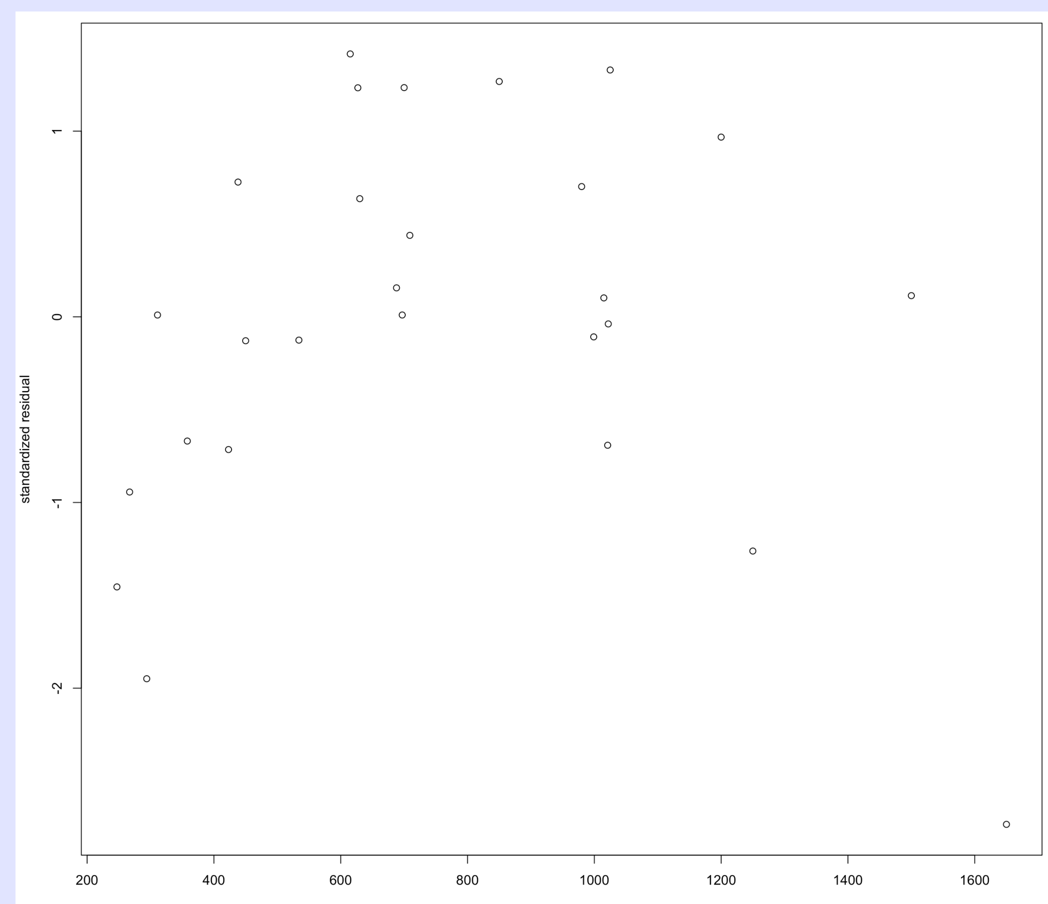
# 6.6 Logarithmic Transformation and Power Transformation

## Use R for last example

on the model $\ln(Y) = \beta_0 + \beta_1 X + \varepsilon$

```
> ######### model 1:  ln(Y) = beta_0 + beta_1 X + epsilon
> mod1<-lm(lny~X, data=supv_dat)
> summary(mod1)

Call:
lm(formula = lny ~ X, data = supv_dat)

Residuals:
     Min       1Q   Median       3Q      Max
-0.59648 -0.16578  0.00244  0.17481  0.34964

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.5150232  0.1110670  31.648  < 2e-16 ***
X           0.0012041  0.0001316   9.153 1.85e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2524 on 25 degrees of freedom
Multiple R-squared:  0.7702,    Adjusted R-squared:  0.761
F-statistic: 83.77 on 1 and 25 DF,  p-value: 1.855e-09
```

```
pii<-hatvalues(mod1)
plot(supv_dat$X,mod1$residuals/(summary(mod1)$sigma * sqrt(1-pii)),xlab="X",ylab="standardized residual")
```



# 6.6 Logarithmic Transformation and Power Transformation

## Logarithmic Transformation

The plot of the residuals against $X$ is shown in Figure 6.17. The plot is quite revealing. Heteroscedasticity has been removed, but the plot shows distinct nonlinearity. The residuals display a quadratic effect, suggesting that a more appropriate model for the data may be

$$\ln y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i. \tag{6.17}$$

Equation (6.17) is a multiple regression model because it has two predictor variables, $X$ and $X^2$. As discussed in Chapter 4, residual plots can also be used in the detection of model deficiencies in multiple regression. To show the effectiveness of residual plots in detecting model deficiencies and their ability to suggest possible corrections, we present the results of fitting model (6.17) in Table 6.13. Plots of the standardized residuals against the fitted values and against each of the predictor variables $X$ and $X^2$ are presented in Figures 6.18–6.20, respectively.[4]

Residuals from the model containing a quadratic term appear satisfactory. There is no appearance of heteroscedasticity or nonlinearity in the residuals. We now have two equally acceptable models for the same data. The model given in Table 6.13 may be slightly preferred because of the higher value of $R^2$. The model given in Table 6.11 is, however, easier to interpret since it is based on the original variables.

Figure 6.17, Table 6.13

# 6.6 Logarithmic Transformation and Power Transformation

## Logarithmic Transformation

**on the model** $\ln(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$

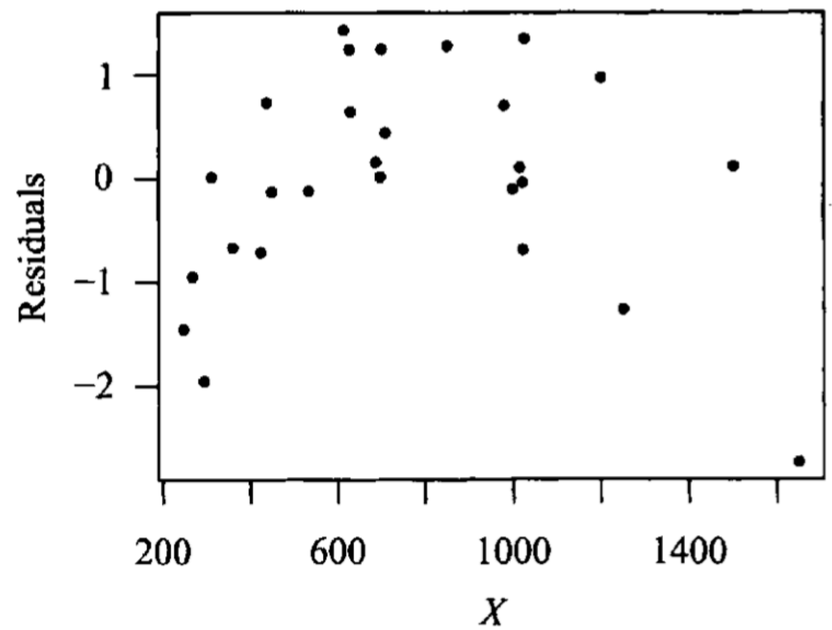



**Figure 6.17** Plot of the standardized residuals against $X$ when $\ln Y$ is regressed on $X$.

**Table 6.13** Estimated Regression Coefficients When $\ln Y$ is Regressed on $X$ and $X^2$

| Variable | Coefficient | s.e. | $t$-Test | $p$-value |
|---|---|---|---|---|
| Constant | 2.8516 | 0.1566 | 18.2 | $< 0.0001$ |
| $X$ | 3.11267E−3 | 0.0004 | 7.80 | $< 0.0001$ |
| $X^2$ | −1.10226E−6 | 0.220E−6 | −4.93 | $< 0.0001$ |
| | $n = 27$ | $R^2 = 0.886$ | $\hat{\sigma} = 0.1817$ | df = 24 |

**Figure 6.18 - 6.20**

# 6.6 Logarithmic Transformation and Power Transformation

## Logarithmic Transformation

on the model $\ln(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$

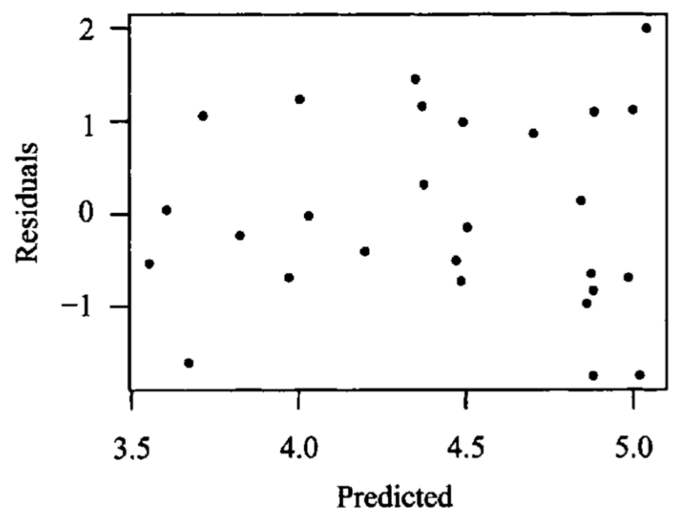



**Figure 6.18** Plot of standardized residuals against the fitted values when $\ln Y$ is regressed on $X$ and $X^2$.

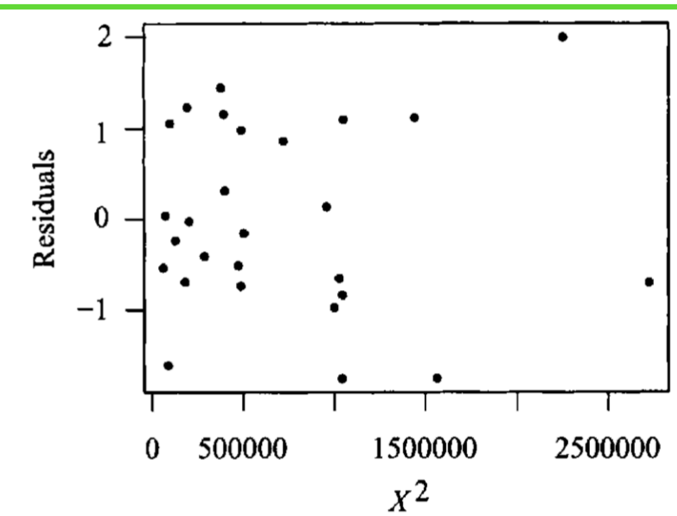



**Figure 6.20** Plot of standardized residuals against $X^2$ when $\ln Y$ is regressed on $X$ and $X^2$.

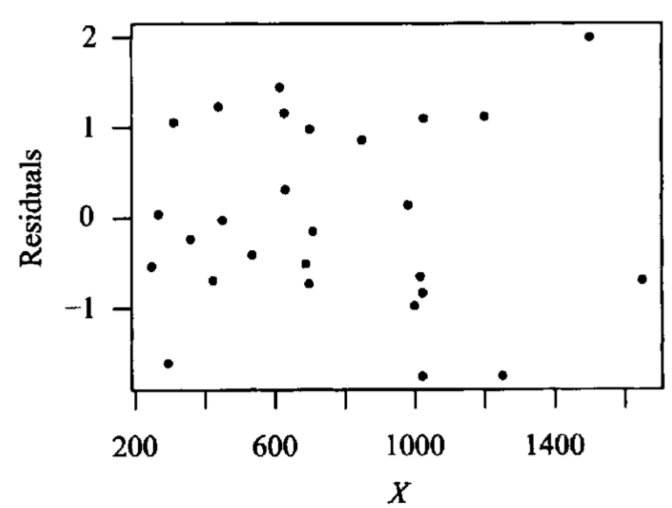


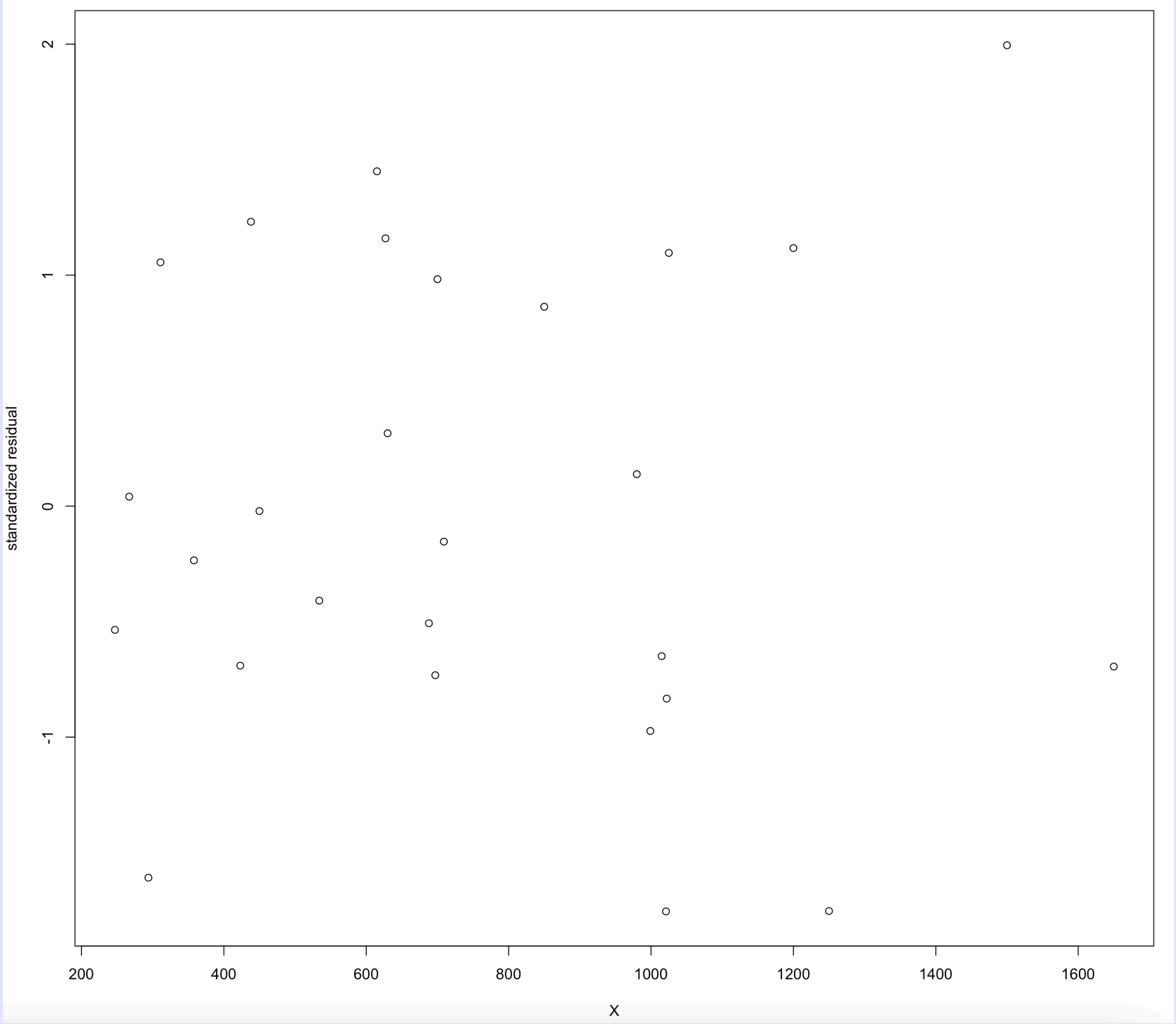

**Figure 6.19** Plot of standardized residuals against $X$ when $\ln Y$ is regressed on $X$ and $X^2$.

# 6.6 Logarithmic Transformation and Power Transformation

## Use R for last example

```
> ######### model 1:  ln(Y) = beta_0 + beta_1 X + beta_2 X^2 + epsilon
> supv_dat$X2<-(supv_dat$X)^2
> mod2<-lm(lny~X+X2, data=supv_dat)
> summary(mod2)

Call:
lm(formula = lny ~ X + X2, data = supv_dat)

Residuals:
     Min       1Q   Median       3Q      Max
-0.30589 -0.11705 -0.02707  0.17593  0.30657

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.852e+00  1.566e-01  18.205 1.50e-15 ***
X            3.113e-03  3.989e-04   7.803 4.90e-08 ***
X2          -1.102e-06  2.238e-07  -4.925 5.03e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1817 on 24 degrees of freedom
Multiple R-squared:  0.8857,    Adjusted R-squared:  0.8762
F-statistic: 92.98 on 2 and 24 DF,  p-value: 4.976e-12
```

```
pii<-hatvalues(mod2)
plot(supv_dat$X,mod2$residuals/(summary(mod2)$sigma * sqrt(1-pii)),xlab="X",ylab="standardized residual")
```



# 6.6 Logarithmic Transformation and Power Transformation

## Power Transformation

In the previous section we used several types of transformations (such as the reciprocal transformation, $1/Y$ , the square root transformation, $\sqrt{Y}$ , and the logarithmic transformation, $\ln(Y)$). These transformation have been chosen based on theoretical or empirical evidence to obtain linearity of the model, to achieve normality, and/or to stabilize the error variance. These transformation can be thought of as a general case of power transformation. In power transformation, we raise the response variable $Y$ and/or some of the predictor variables to a power. For example, instead of using $Y$ we use $Y^\lambda$, where $\lambda$ is an exponent to be chosen by the data analyst based on either theoretical or empirical evidence. When $\lambda = -1$ we obtain the reciprocal transfonnation, $\lambda = 0.5$ gives the square root transformation, and when $\lambda = 0$ we obtain the logarithmic transformation. Values of $\lambda = 1$ implies no transformation is needed.

If $\lambda$ cannot be determined by theoretical considerations, the data can be used to determine the appropriate value of $\lambda$. This can be done using numerical methods. In practice, several values of $\lambda$ are tried and the best value is chosen. Values of $\lambda$ commonly tried are: $2, 1.5, 1.0, 0.5, 0, -0.5, -1, -1.5, -2$. These values of $\lambda$ are chosen because they are easy to interpret. They are known as a **ladder of transformation**. This is illustrated in the following example.

# 6.6 Logarithmic Transformation and Power Transformation

## Choosing Power Transformation

Oftentimes, given the dataset $(x_1, y_1), \cdots, (x_n, y_n)$, we have to determine an appropriate power transformation $X^{\lambda_1}$ and $Y^{\lambda_2}$ **emprically** for choices of $\lambda_1$ and $\lambda_2$.

A common practice is to choose a list of possible values for $\lambda_1$ and $\lambda_2$, e.g., $2, 1.5, 1.0, 0.5, 0, -0.5, -1, -1.5, -2$ and calculate the sample correlation coefficient between $X^{\lambda_1}$ and $Y^{\lambda_2}$. Since the most important purpose of transformation is to achieve linearity (e.g., **strong correlation**), we choose a value $\lambda_1$ and $\lambda_2$ so that the sample correlation coefficient between $X^{\lambda_1}$ and $Y^{\lambda_2}$ is the largest.

**Example**

| | $X^{-2}$ | $X^{-1.5}$ | $X^{-1}$ | $X^{-0.5}$ | $X^{0}$ | $X^{0.5}$ | $X^{1}$ | $X^{1.5}$ | $X^{2}$ |
|---|---|---|---|---|---|---|---|---|---|
| $Y^{2}$ | 0.45 | 0.56 | 0.61 | 0.74 | 0.63 | 0.84 | 0.89 | 0.27 | 0.31 |
| $Y^{1.5}$ | | | | ...... | | | | | |
| $Y^{1}$ | | | | | | | | | |
| $Y^{0.5}$ | | | | | | | | | |
| $Y^{0}$ | | | | ...... | | | | | |
| $Y^{-0.5}$ | | | | | | | | | |
| $Y^{-1}$ | | | | | | | | | |
| $Y^{-1.5}$ | | | | | | | | | |
| $Y^{-2}$ | | | | ...... | | | | | |

# 6.6 Logarithmic Transformation and Power Transformation

## Power Transformation

### Example: the Brain Data

The data set shown in Table 6.14 represent a sample taken from a larger data set. The original sources of the data is Jerison (1973). It has also been analyzed by Rousseeuw and Leroy (1987). The average brain weight (in grams), $Y$, and the average body weight (in kilograms), $X$, are measured for 28 animals. One purpose of the data is to determine whether a larger brain is required to govern a heavier body. Another purpose is to see whether the ratio of the brain weight to the body weight can be used as a measure of of intelligence.

**Table 6.14** The Brain Data: Brain Weight (Grams) and Body Weight (Kilograms)

| Name | Brain Weight | Body Weight | Name | Brain Weight | Body Weight |
|---|---|---|---|---|---|
| Mountain beaver | 8.1 | 1.35 | African elephant | 5712.0 | 6654.00 |
| Cow | 423.0 | 465.00 | Triceratops | 70.0 | 9400.00 |
| Gray wolf | 119.5 | 36.33 | Rhesus monkey | 179.0 | 6.80 |
| Goat | 115.0 | 27.66 | Kangaroo | 56.0 | 35.00 |
| Guinea pig | 5.5 | 1.04 | Hamster | 1.0 | 0.12 |
| Diplodocus | 50.0 | 11700.00 | Mouse | 0.4 | 0.02 |
| Asian elephant | 4603.0 | 2547.00 | Rabbit | 12.1 | 2.50 |
| Donkey | 419.0 | 187.10 | Sheep | 175.0 | 55.50 |
| Horse | 655.0 | 521.00 | Jaguar | 157.0 | 100.00 |
| Potar monkey | 115.0 | 10.00 | Chimpanzee | 440.0 | 52.16 |
| Cat | 25.6 | 3.30 | Brachiosaurus | 154.5 | 87000.00 |
| Giraffe | 680.0 | 529.00 | Rat | 1.9.0 | 0.28 |
| Gorilla | 406.0 | 207.00 | Mole | 3.0 | 0.12 |
| Human | 1320.0 | 62.00 | Pig | 180.0 | 192.00 |

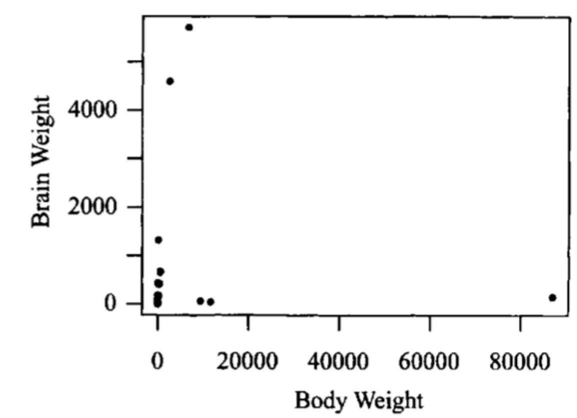



**Figure 6.21** Brain data: Scatter plots of Brain Weight versus Body Weight.

# 6.6 Logarithmic Transformation and Power Transformation

## Power Transformation

### Example: the Brain Data

The scatter plot of the data (Figure 6.21) does not show an obvious relationship. This is mainly due to the presence of very large animals (e.g., two elephants and three dinosaurs). Let us apply the power transformation to both $Y$ and $X$. The scatter plots of $Y^\lambda$ versus $X^\lambda$ for several values of $\lambda$ in the **ladder of transformation** are given in Figure 6.22. It can be seen that the values of $\lambda = 0$ (corresponding to the log transformation) is the most appropriate value. For $\lambda = 0$, the graph looks linear but the three dinosaurs do not conform to the linear pattern suggested by the other points. The graph suggests that either the brain weight of the dinosaurs is underestimated and/or their body weight is overestimated.

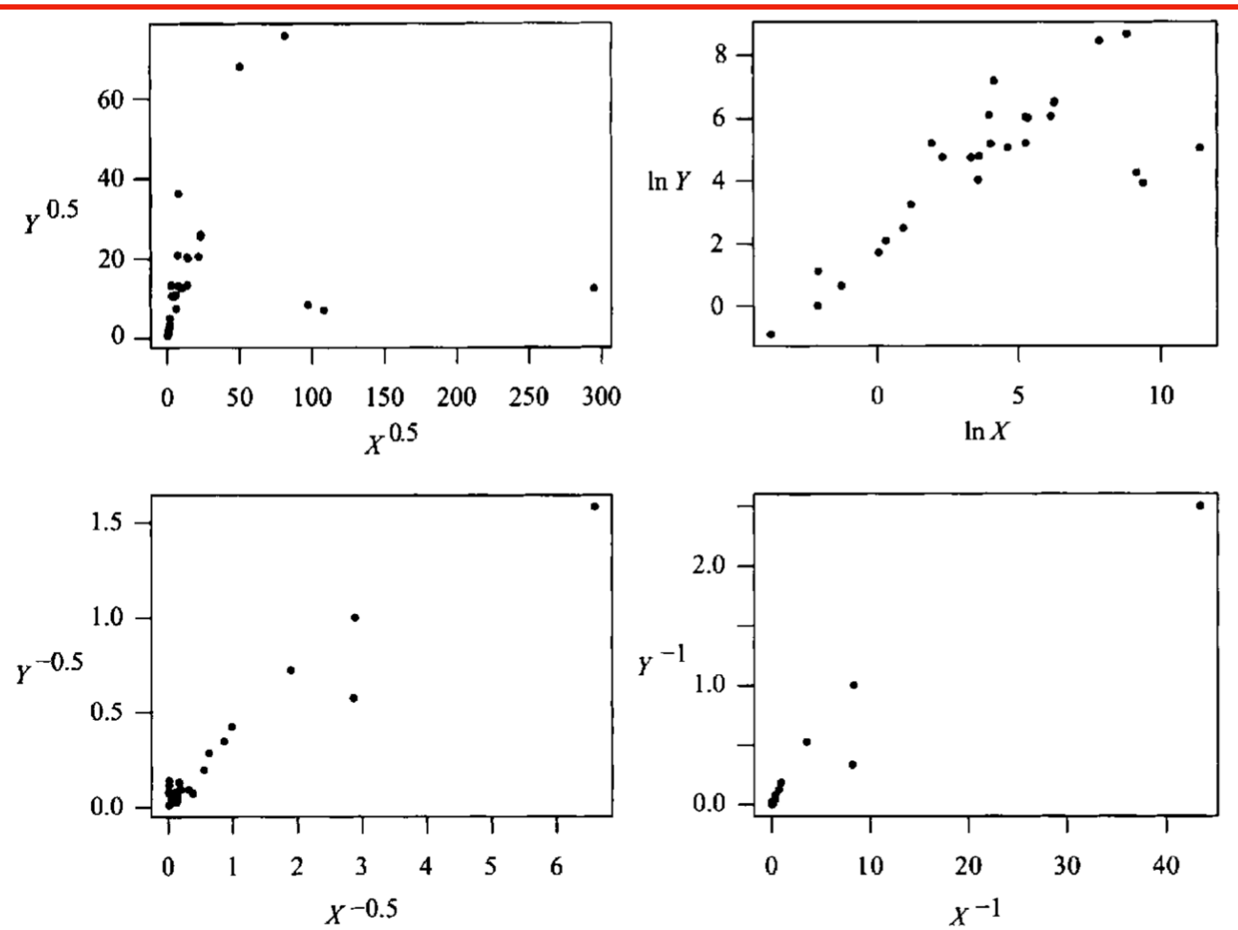



**Figure 6.22** Scatter plots of $Y^\lambda$ versus $X^\lambda$ for various values of $\lambda$.

*Note that in this example we transformed both the response and the predictor variables and that we used the same value of the power for both variables. In other applications, it may be more appropriate to raise each value to a different power and/or to transform only one variable.*

# 6.6 Logarithmic Transformation and Power Transformation

## Use R for last example

```
################################################################## Power Transformation on the Brain dataset
brain_dat<-read.table('data/P184.txt',header=TRUE)   ## read the data
brain_dat<-data.frame(Name=brain_dat$Name,X=brain_dat$BodyWeight,Y=brain_dat$BrainWeight)
par(mfrow=c(2,2))
plot(sqrt(brain_dat$X),sqrt(brain_dat$Y),xlab=TeX(r'($X^{0.5}$)'),ylab=TeX(r'($Y^{0.5}$)'))
plot(log(brain_dat$X),log(brain_dat$Y),xlab=TeX(r'($\ln X$)'),ylab=TeX(r'($\ln Y$)'))
plot(1/sqrt(brain_dat$X),1/sqrt(brain_dat$Y),xlab=TeX(r'($X^{-0.5}$)'),ylab=TeX(r'($Y^{-0.5}$)'))
plot(1/brain_dat$X,1/brain_dat$Y,xlab=TeX(r'($X^{-1}$)'),ylab=TeX(r'($Y^{-1}$)'))
```