Chapter 7

Using Transformation in Regression Models

Overview

- Complex models may be changed into linear models thru appropriate transformations
- Use of power transformation
- Transformation of non-linear models
- Box and Cox Transformation
- Importance of checking residuals
- Variance stabilizing transformation
- Some commonly used transformations

7.1 Introduction

Reasons for a transformation of data

- Transformation of data can sometimes reduce complex models to linear ones.

- Transforming a non-linear model into a linear model

- Stabilizing the variance

Introduction (Continued)

Some examples of reducing complex models into linear models.

(i) reciprocal transformation

$$y = \beta_0 + \frac{\beta_1}{x_1} + \frac{\beta_2}{x_2} + \epsilon$$

By letting $w_1 = 1/x_1$ and $w_2 = 1/x_2$, then we have

$$y = \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \varepsilon$$

Introduction (Continued)

(ii) logarithm transformation

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \varepsilon$$

It can be written as

$$y = \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \varepsilon$$

where $w_1 = \log(x_1)$ and $w_2 = \log(x_2)$.

(Note: it is assumed here that x_1 and x_2 take only positive values.)

Introduction (Continued)

(iii) square root transformation

$$y = \beta_0 + \beta_1\sqrt{x_1} + \beta_2\sqrt{x_2} + \epsilon$$

(x_1 and x_2 take only positive values.)

It can be written as

$$y = \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \varepsilon$$

where $w_1 = \sqrt{x_1}$ and $w_2 = \sqrt{x_2}$

Introduction (Continued)

- There are many other transformations such as higher powers or lower powers of x_i's and y (x^r is said to be a higher power of x if $r > 1$ and lower power if $r < 1$.)

Further examples

(iv) $y = \beta_0 + \beta_1 x_1^2 + \beta_2 \log x_2 + \epsilon$

(v) $\sqrt{y} = \beta_0 + \beta_1 x^{-1/3} + \beta_2 x_2^2 + \epsilon$

7.2 Which Transformation to be Used

- The choice of which transformation to be used (if necessary) is often difficult to decide since we don't know the true model.

- However for the cases that involve only x and y, then a scatter plot may give some hints on which transformation to be used.

Which Transformation (Continued)

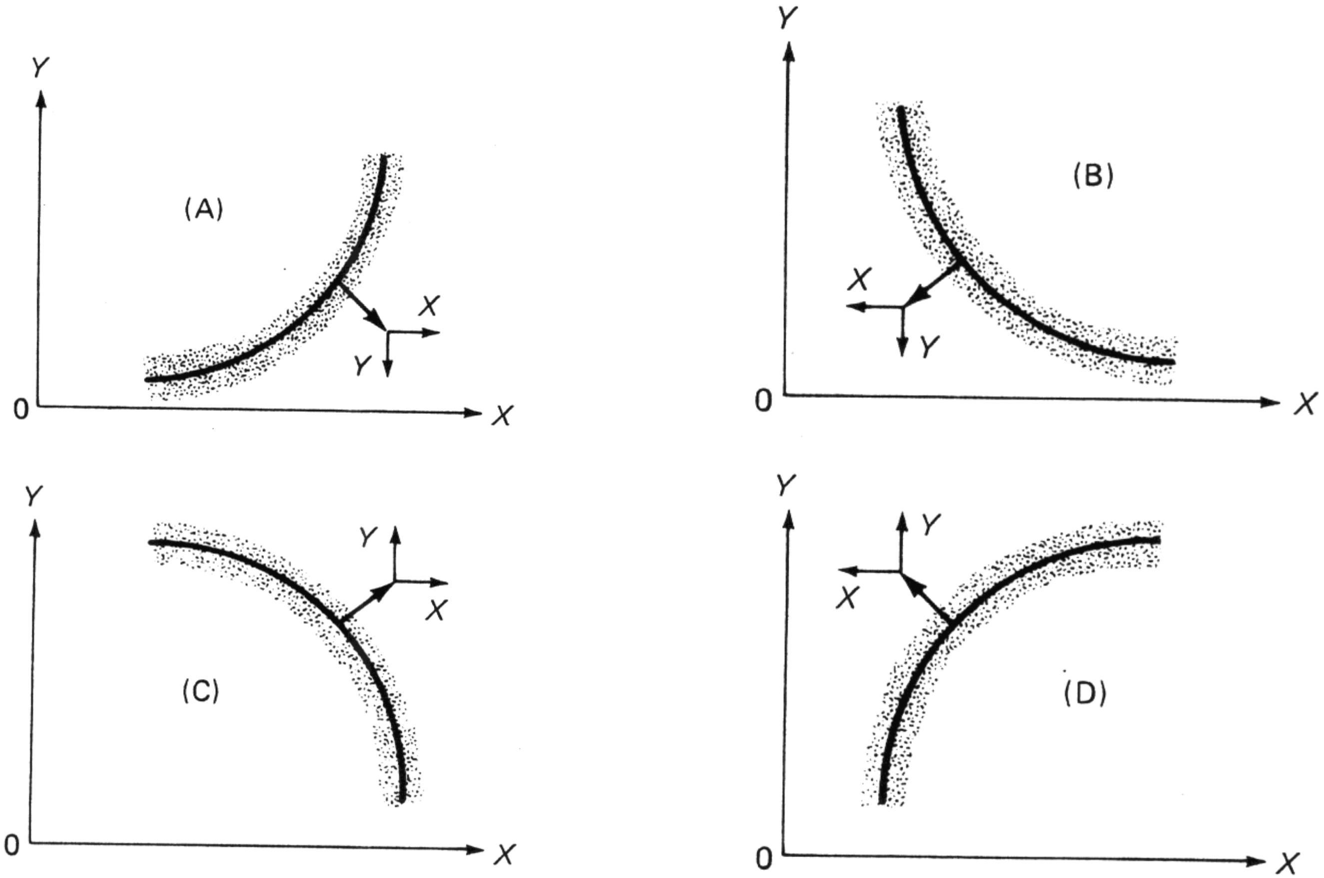


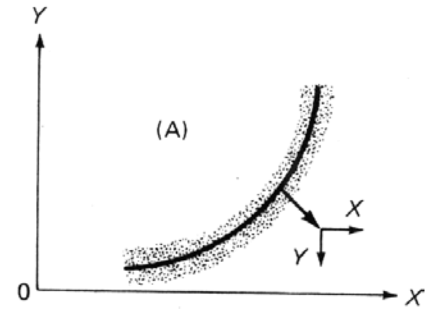
Figure 7-1 Patterns in curvilinear relationships indicating the direction of (power) re-expression for x and y.

Which Transformation (Continued)

Suggested transformations are

Case	A	B	C	D
x y	H* (and/or) L	L (and/or) L	H (and/or) H	L (and/or) H

For example for case A:



- Try $y = \beta_0 + \beta_1 w + \varepsilon$ with $w = x^2$ (i.e. a higher power of x)
- Or $z = \beta_0 + \beta_1 x + \varepsilon$ with $z = y^{1/2}$ or $z = \log(y)$ (i.e. a lower power of y)

7.3 Transformation of Nonlinear Models

- Some nonlinear models are intrinsically linear and by a suitable transformation, such models can be expressed as linear models.

For examples

(i) multiplicative model

$$y = \alpha x_1^{\beta} x_2^{\gamma} x_3^{\delta} \epsilon,$$

where α, β, γ and δ are unknown parameters and ε is a multiplicative random error.

Transformation of Nonlinear Models (Continued)

- Taking logarithm on both sides, we have
$$\log(y) = \log(\alpha) + \beta \log(x_1) + \gamma \log(x_2) + \delta \log(x_3) + \log(\epsilon),$$
which is the familiar multiple regression model for $\log(y)$ on $\log(x_1)$, $\log(x_2)$ and $\log(x_3)$.

- Note: $y = \alpha {x_1}^{\beta} {x_2}^{\gamma} {x_3}^{\delta} + \varepsilon$ cannot be transformed into a linear model since taking logarithm on both sides leads to $\log(y) = \log\left(\alpha {x_1}^{\beta} {x_2}^{\gamma} {x_3}^{\delta} + \varepsilon\right)$.

Transformation of Nonlinear Models (Continued)

(ii) exponential model

$$y = e^{\beta_0+\beta_1 x_1+\beta_2 x_2}\epsilon$$

Taking logarithm on both sides, we have

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \log(\epsilon)$$

(iii) reciprocal model

$$y = \frac{1}{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon}$$

Taking reciprocal on both sides, we have

$$\frac{1}{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

7.4 Examples

Example 1

The following set of data represents the number of days training (x) and performance score (y) for 10 sales trainees in a battery of simulated sales situations in an experiment.

x	0.5	0.5	1.0	1.0	1.5	1.5	2.0	2.5	3.0	3.5
y	43	40	71	74	107	109	158	209	270	341

Example 1 (Continued)

- We first use the simple regression model

$$y = \beta_0 + \beta_1 x + \epsilon$$

- By performing the routine calculations, or using SAS or R, we have the following results.

$$\hat{y} = -23.3906 + 47.4063x$$

Example 1 (Continued)

ANOVA Table

Source	SS	df	MS	F	p-value
Regression	91084.58	1	91084.58	398.4	< 0.0001
Error	1829.02	8	228.627		
Total	92913.6	9			

$R^2 = 0.9803.$

- Since $F_{obs} = 398.4 > F_{0.05}(1, 8) = 5.32$ (or p-value < 0.05), therefore we reject the null hypothesis that there is no significant model at the 5% level of significance and conclude that there is a significant relationship between y and x.

Example 1 (Continued)

- Since independent repeat observations are available, we would like to perform the lack of fit test.

x	$\sum_{k=1}^{n_j}(y_{jk}-\bar{y}_j)^2 = \sum_{k=1}^{n_j} y_{jk}^2 - n\bar{y}_j^2$	df
0.5	$43^2 + 40^2 - 83^2/2 = 4.5$	1
1.0	$71^2 + 74^2 - 145^2/2 = 4.5$	1
1.5	$107^2 + 109^2 - 216^2/2 = 2$	1
	SSPE = 11	3

Example 1 (Continued)

- $SSLF = SSE - SSPE$

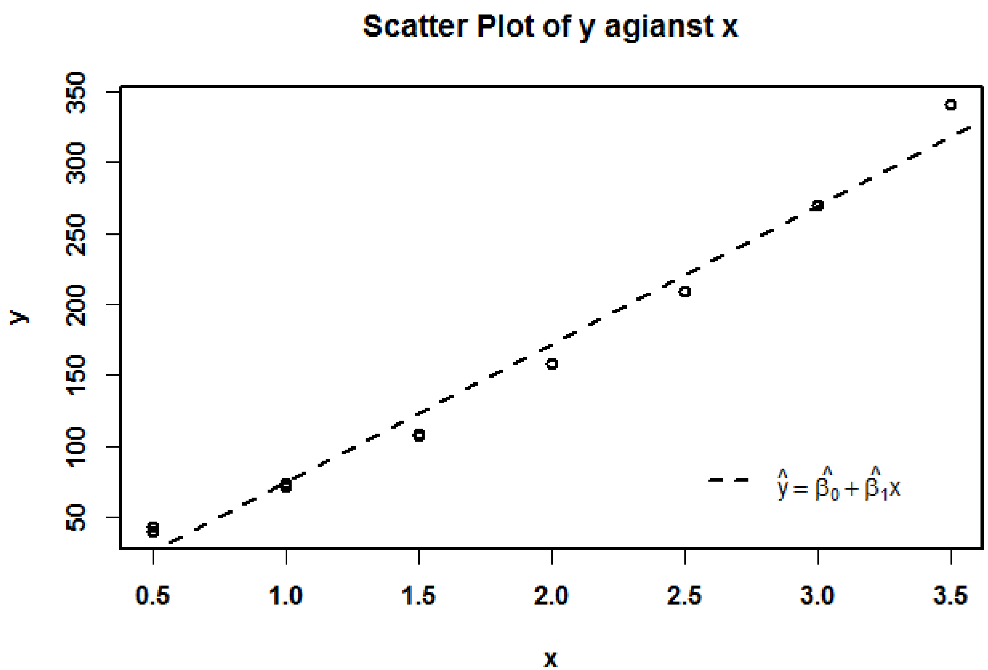
  $= 1829.016 - 11 = 1818.016$ with $8 - 3 = 5$ d.f.

- $F_L = MSLF \,/\, MSPE = 99.16.$

- Since $F_L = 99.16 > F_{0.05}(5, 3) = 9.01$ (or p-value $= 0.001575 < 0.05$) therefore we reject the hypothesis that there is no lack of fit and conclude that the simple linear regression model does not provide an adequate fit.

Example 1 (Continued)

- A scatter plot of y against x shows that the relation between y and x appears to be curvilinear. In fact the plot resembles that shown in Case A.



Example 1 (Continued)

- Hence a lower power transformation on y is suggested and the following model is used:

$$\sqrt{y} = \beta_0 + \beta_1 x + \epsilon$$

- By using SAS or R, we obtain the following results.

- The regression equation

$$\widehat{\sqrt{y}} = 4.4617 + 4.0002x$$

Example 1 (Continued)

- ANOVA Table

Source	SS	df	MS	F	p-value
Regression	153.6121	1	153.6121	16183.12	< 0.0001
Error	0.075937	8	0.009492		
Total	153.6880	9			

- $R^2 = 0.999506$.
- Since $F_{\text{obs}} = 16183.12 > F_{0.05}(1, 8) = 5.32$ (or p-value < 0.05), therefore we reject H_0 at the 5% significance level and we conclude that at the 5% level of significance there is a significant relationship between $\sqrt{y}$ and x.

Example 1 (Continued)

- Furthermore, we can carry out a lack of fit test:

x	z_{jk}	$\sum_{k=1}^{n_j}\left(z_{jk} - \bar{z}_j\right)^2$	df
0.5	6.5574, 6.3246	0.0270970	1
1.0	8.4261, 8.6023	0.0155232	1
1.5	10.3441, 10.4403	0.0046272	1
	SSPE =	0.0472484	3

where $z_{jk} = \sqrt{y_{jk}}$ and $\bar{z}_j = \frac{1}{n_j}\sum_{k=1}^{n_j}\sqrt{y_{jk}}$

Example 1 (Continued)

- $SSLF = SSE - SSPE = 0.0759370 - 0.0472484$
  $= 0.0286886$ with 5 d.f.

- $F_L = MSLF / MSPE$
  $= (0.0286886/5)/(0.0472484/3) = 0.36$

- Since $F_L = 0.36 < F_{0.05}(5, 3) = 9.01$ (or p-value = 0.8501 > 0.05) therefore we do not reject the hypothesis that there is no lack of fit and conclude that the model $\sqrt{y} = \beta_0 + \beta_1 x + \epsilon$ is an appropriate model.

Example 2

- The pressure P of a gas corresponding to various volumes V is recorded as follows:

V (cm^3)	50	60	70	90	100
P (kg/cm^2)	64.7	51.3	40.5	25.9	7.8

- The ideal gas law is given by the functional form $PV^\gamma = C$, where γ and C are constants.

- Estimate the constants C and γ.

Solution to Example 2

- Since there are random errors in recording the data, therefore an appropriate model should be

$$P_i V_i^{\gamma} = C\epsilon_i, \quad i = 1, 2, 3, 4, 5.$$

- We can take natural logarithm of both sides of the model and a linear model is obtained as follows.

$$\log(P_i) = \alpha + \beta \log(V_i) + \epsilon_i^*, \quad i = 1, 2, 3, 4, 5,$$

where $\alpha = \log(C)$, $\beta = -\gamma$ and $\epsilon^* = \log(\epsilon)$.

Solution to Example 2

- By performing the routine calculations or using SAS or R, we have the following results.

$$\widehat{\log P} = 14.759 + 2.6535 \log V$$

- Hence $\hat{C} = e^{14.759} = 2568930$ and $\hat{\gamma} = -2.6535$

Solution to Example 2 (Continued)

- ANOVA Table

Source	SS	df	MS	F	p-value
Regression	2.28546	1	2.28546	13.27	0.0357
Error	0.51685	3	0.51685		
Total	2.80231	4			

- $R^2 = 0.8156$.
- Since $F_{\text{obs}} = 13.27 > F_{0.05}(1, 3) = 10.13$ (or p-value < 0.05), we conclude that there is a significant relationship between P and V.

7.5 Box and Cox Transformation

- Suppose that the data $(y_1, y_2, \ldots., y_n)$ on a response variable y have the following properties.

  (1) y is always positive,

  (2)
$$\frac{y_{\max}}{y_{\min}} > 10$$

- Then we may consider the possibility of transforming y.

Box and Cox Transformation (Continued)

- Box and Cox considered the following power transformation

$$w = \begin{cases} (y^{\lambda} - 1)/\lambda, & \text{for } \lambda \neq 0 \\ \log y, & \text{for } \lambda = 0 \end{cases} \qquad (1)$$

Or a modified form

$$v = \begin{cases} (y^{\lambda} - 1)/(\lambda \tilde{y}^{\lambda-1}), & \text{for } \lambda \neq 0 \\ \tilde{y} \log y, & \text{for } \lambda = 0 \end{cases} \qquad (2)$$

where $\tilde{y} = (y_1 y_2 \cdots y_n)^{1/n}$

Box and Cox Transformation (Continued)

- Find the appropriate value of λ such that $\underline{w}$ or $\underline{v}$ satisfies

$$\underline{w} = X\underline{\beta} + \underline{\epsilon},$$

or $\qquad (3)$

$$\underline{v} = X\underline{\beta} + \underline{\epsilon},$$

where $\underline{\epsilon} \sim N(\underline{0}, \sigma^2 I)$

- How to find the appropriate value of λ?

- Maximum Likelihood Method!!

Box and Cox Transformation (Continued)

Maximum Likelihood Method of Estimating λ

- Choose a value of λ from a selected range.
  - Usually we look at λ's in the range of $(-1, 1)$, or perhaps even $(-2, 2)$, at first, and modify the range later.

- For each chosen λ value, evaluate $\underline{v}$ using Equation (2).

- Next we fit the model in (3) and record $S(\lambda, \underline{v})$, the residual sum of squares for the regression.

Box and Cox Transformation (Continued)

Maximum Likelihood Method of Estimating λ

- Plot $S(\lambda, \underline{\nu})$ versus λ.

- Draw a smooth curve through the plotted points, and find out at what value of λ, the lowest point of the curve lies. That value, $\hat{\lambda}$, is the maximum likelihood estimate of λ.

Note:

- In regression models with normal random error, **maximizing the likelihood function** is equivalent to **minimizing the residual sum of squares**

Box and Cox Transformation (Continued)

Remarks

1. The fact that the "best λ" has been selected does not necessarily guarantee an equation useful in practice. The final equation must be evaluated in the usual ways on its own merits.

2. To allow for the fact that λ has been estimated, some statisticians remove one degree of freedom from SST and SSE for estimating $\hat{\lambda}$ in the ANOVA table in the subsequent analysis. However the reduction is optional.

Example 3

- The following table shows the Mooney Viscosity at 100°C as a function of filler level and oil level.

- A model $V = \beta_0 + \beta_1 f + \beta_2 p + \epsilon$ is fitted to the data, where f is the filler level and p is the oil level.

	Filter, f					
Oil, p	0	12	24	36	48	60
0	26	38	50	76	108	157
10	17	26	37	53	83	124
20	13	20	27	37	57	87
30	-	15	22	27	41	63

Example 3 (Continued)

- Note that the response data range from 13 to 157, a ratio of $157/13 = 12.1$.
- The geometric mean is $\tilde{y} = 41.5461$
- The following table shows the selected values of $S(\lambda, \underline{v})$ for various λ.

λ	−1.0	−0.8	−0.6	−0.4	−0.2	−0.15	−0.10
$S(\lambda, \underline{v})$	2456	1453	779.1	354.1	131.7	104.5	88.3

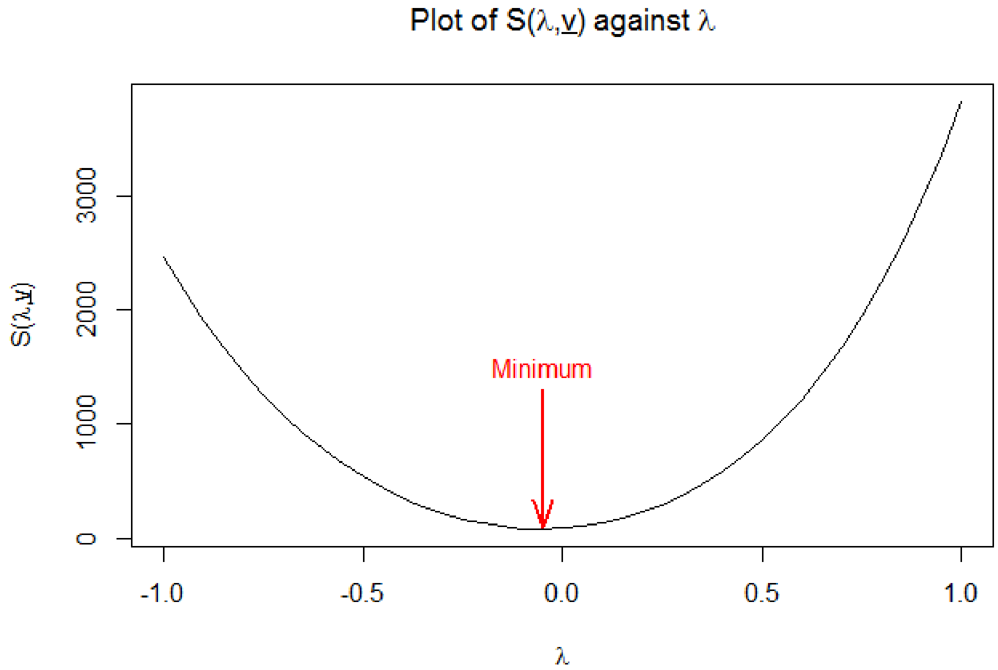
λ	−0.08	−0.06	−0.05	−0.04	−0.02	−0.00	0.05
$S(\lambda, \underline{v})$	84.9	83.3	83.2	83.5	85.5	89.3	106.7

λ	0.10	0.2	0.4	0.6	0.8	1.0
$S(\lambda, \underline{v})$	135.9	231.1	588.0	1222	2243	3821

Example 3 (Continued)

- A smooth curve through these points is plotted in the figure below.

- We see that the minimum of $S(\lambda, \underline{v})$ occurs at about $\lambda = -0.05$.



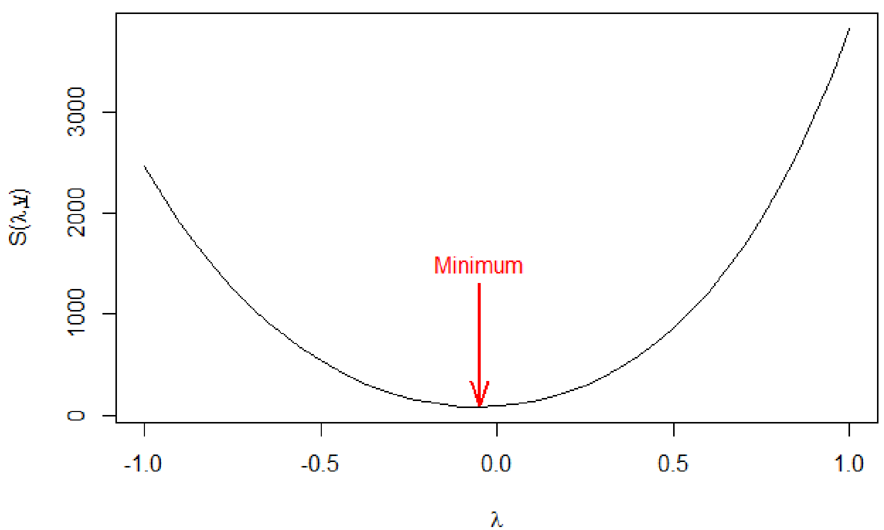
Example 3 (Continued)

- We see that the minimum of $S(\lambda, \underline{v})$ occurs at about $\lambda = -0.05$. This is close to zero, suggesting that the transformation

$$v = \tilde{y} \log y$$

or more simply $\log(y)$, might be suitable for this set of data.

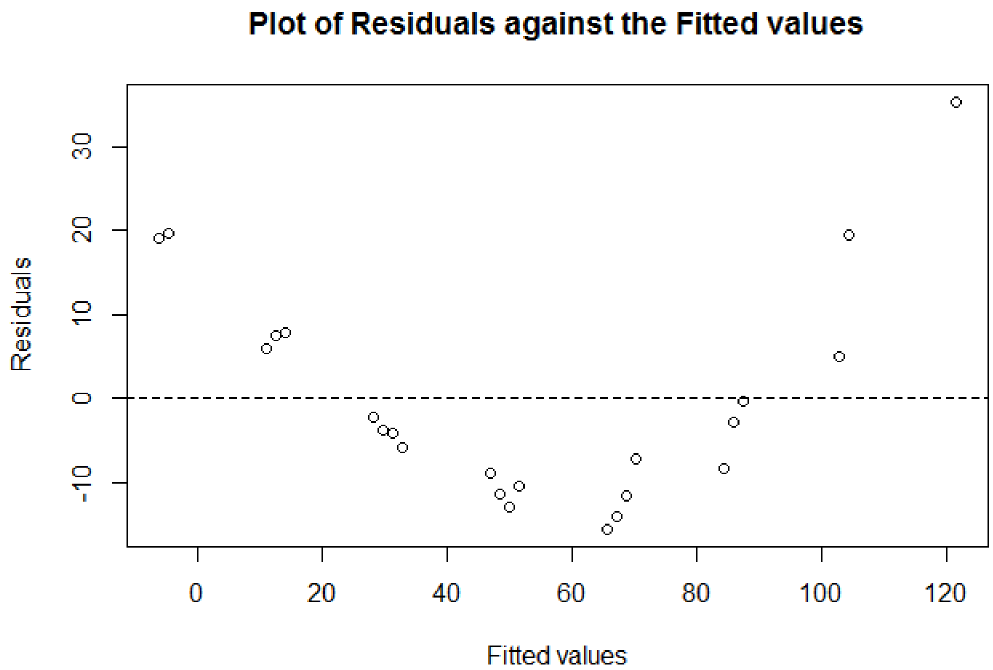
Plot of S(λ,v) against λ



Example 3 (Continued)

Residual plot against the fitted values

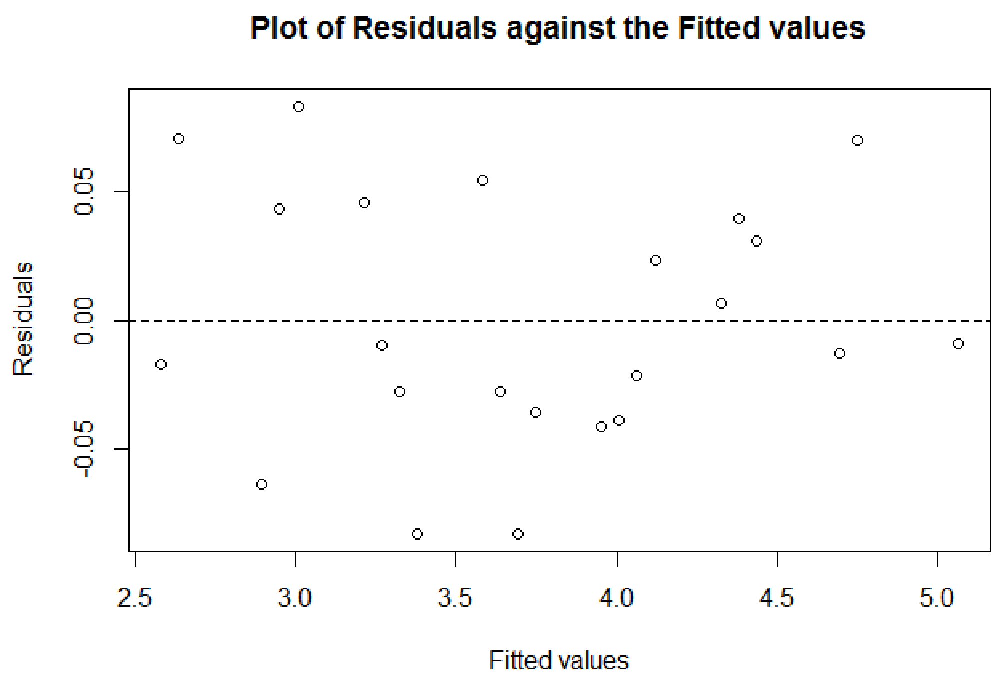
(a) Model: $y = \beta_0 + \beta_1 f + \beta_2 p + \epsilon$



Example 3 (Continued)

Residual plot against the fitted values

(b) Model: $\log(y) = \beta_0 + \beta_1 f + \beta_2 p + \epsilon$



7.6 Importance of Checking Residuals

- Transformation on the response variable affects the distribution errors.
- The assumptions that the errors are independent and follow a normal distribution with mean zero and variance σ^2 apply to the transformed data.
- Hence it is important to study the residuals from the model fitted, to see if those assumptions still hold.
- We shall discuss how to check residuals in Chapter 10.

7.7 Variance Stabilizing Transformation

- Another reason for transforming the data is to stabilize the variance when the assumption of constant variance is violated.
- Suppose that the variance of the un-transformed data y, $\sigma_y{}^2$ say, is a function $g(\eta)$ of the mean value, $\eta = E(y)$.
- We can then obtain an appropriate transformation by using the transformed variable $h(y)$ where

$$\frac{\partial h(y)}{\partial y} \propto \frac{1}{\sqrt{g(y)}}$$

Variance Stabilizing Transformation (Continued)

- Since

$$\frac{\partial h(y)}{\partial y} \propto \frac{1}{\sqrt{g(y)}}$$

therefore,

$$h(y) \propto \int \frac{1}{\sqrt{g(y)}} dy$$

Variance Stabilizing Transformation (Continued)

- For example,

  If $\sigma_y^2 \propto E(y)^2$, [i.e. $g(\eta) = \eta^2$]

  then

$$h(y) = \int \frac{1}{y}\, dy = \log(y)$$

Some Commonly Used Transformations

Nature of Dependence $\sigma_y^2 \propto g(\eta),\ \eta = E(y)$	**Variance Stabilizing Transformation**
$\sigma_y^2 \propto \eta$	$\sqrt{y}$
$\sigma_y^2 \propto \eta^2$	$\log(y)$
$\sigma_y^2 \propto \eta^3$	$y^{-1/2}$
$\sigma_y^2 \propto \eta^{-1}$	$y^{3/2}$
$\sigma_y^2 \propto \eta^k$	$y^{1-k/2}$
$\sigma_y^2 \propto \eta(1-\eta)$	$\sin^{-1}(\sqrt{y})$

Example 4

- The average monthly income from food sales and the corresponding annual advertising expenses for 30 restaurants are shown below:

y	81464	72661	72344	90734	98588	96507	126574	114113
x	3000	3150	3085	5225	5350	6090	8925	9015

y	115814	123181	131434	10564	151352	146926	130963	144630
x	8885	8950	9000	11345	12275	12400	12525	12310

y	147041	179021	162000	180732	178187	185304	155931	172579
x	13700	15000	15175	14995	15050	15200	15150	16800

y	188851	192424	203112	192482	218715	214317
x	16500	17830	19500	19200	19000	19350

y: income, x: advertising expense

Example 4 (Continued)

- A first order model is fitted to the data and the following results were obtained:

- The regression equation is given by

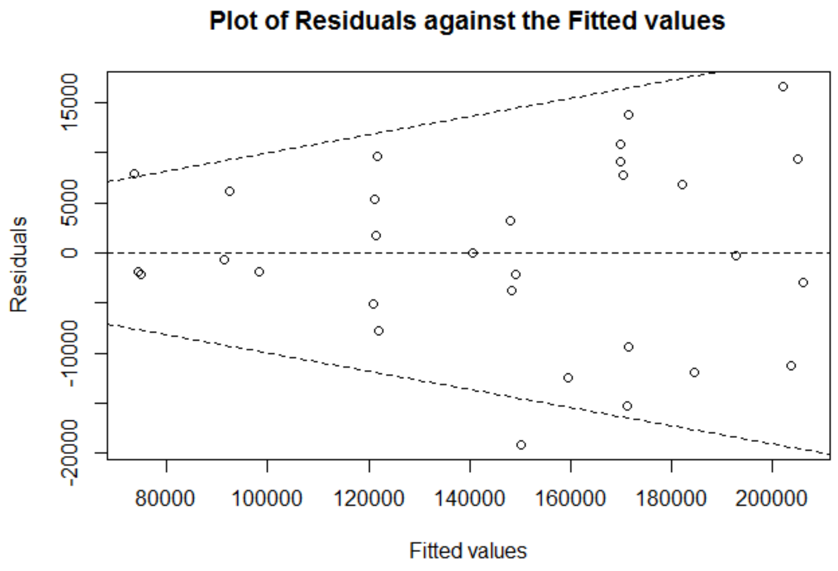
$$\hat{y} = 49507.4958 + 8.03162x$$

$F_{\text{obs}} = MSR/MSE = 588.89 > F_{0.05}(1, 28) = 4.20$ and

$R^2 = 0.954611$.

Example 4 (Continued)

- Although the model is significant at the 5% level ($F = 588.89 > F_{0.05}(1, 28) = 4.20$) and the value of the coefficient of determination is high ($R^2 = 0.954611$), a residual plot against $\hat{y}$ indicates that the residuals are more spread out when $\hat{y}$ increases, hence the assumption of constant variance is not valid.



Example 4 (Continued)

Note: $\hat{y}$ is an estimate of $\mu_{Y|\underline{x}}$

- Hence the residual plot indicates to some extent that σ_y^2 increases as $\mu_{Y|\underline{x}}$ increases
- That is, $\sigma_y^2 \propto \mu^k$, k : any number ≥ 1
- Let us assume $\sigma_y^2 \propto \mu$
- Hence the corresponding variance stabilizing transformation is $y^* = \sqrt{y}$
- A model is fitted to the transformed data and the following results are obtained.

$$\widehat{y^*} = 246.9852 + 0.01090x$$

Example 4 (Continued)

ANOVA Table

Source	SS	df	MS	F	p-value
Regression	90244.71	1	90244.71	673.34	< 0.0001
Error	3752.72	28	134.026		
Total	93997.43	29			

- $R^2 = 0.960076$.
- Since $F_{obs} = 673.34 > F_{0.05}(1, 28) = 4.20$ (or p-value < 0.05), we reject H_0 at the 5% significance level conclude that there is a significant relationship
- Hence, we conclude that there is a significant linear relationship between y^* and x

Example 4 (Continued)

- The residual plot with e against $\widehat{y^*}$ shows no serious violation of the assumption of constant variance. Therefore, the model $\sqrt{y} = \beta_0 + \beta_1 x + \epsilon$ is adequate.