Modelo de Gauss-Markov de Regressão: Adequação de Normalidade e Inferência na Escala Original, após Transformação.

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sra. María Carola Alfaro Vives e aprovada pela Comissão Julgadora .

Campinas, 8 de maio de 1995

Clarice Azevedo de Luna Freire

Profa. Dra. Clarice Azevedo de Luna Freire .

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

UNIVERSIDADE ESTADUAL DE CAMPINAS
DEPARTAMENTO DE ESTATÍSTICA - IMECC

Modelo de Gauss-Markov de Regressão: Adequação de Normalidade e Inferência na Escala Original, após Transformação.

María Carola Alfaro Vives

Orientação
Profa. Dra. Clarice Azevedo de Luna Freire

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência de Computação da Universidade Estadual de Campinas para obtenção de título de Mestre em Estatística
Campinas - S.P.
1994

Agradecimentos

Gostaria de expressar meus agradecimentos a:

- CAPES pelo apoio financiero para realizar este trabalho de investigação,
- Clarice pela constante disponibilidade para orientar no desenvolvimento da pesquisa e pela amizade, fazendo deste trabalho uma experiência agradável e motivadora,
- Johnattan pelo generoso apoio na parte computacional e importantes sugestões,
- Meus professores, funcionários da universidade e amigos (em especial das minhas repúblicas e do predinho) pela amizade que me brindaram,
- Minha querida família que sempre esteve presente apesar da distância.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Frequentemente, é necessário modelar a relação entre uma variável resposta, y, e uma variável regressora, x (ou várias variáveis regressoras). O modelo linear de regressão de Gauss-Markov, normal, é de grande atrativo pela sua simplicidade, e disponibilidade de resultados teóricos e programas computacionais. Entre as vantagens deste modelo, poderia ser citada a inferência exata, mesmo para pequenas amostras. No entanto, nem sempre o ajuste do modelo linear parece ser adequado: relação entre y e x não linear ou o erro não parece ter distribuição normal. Uma alternativa, usualmente adotada, é a utilização do modelo linear de regressão da transformação da variável resposta, $g(y)$, em x. Após este ajuste de modelo em escala transformada, como fazer inferência na escala original?

O enfoque principal deste trabalho é a estimação pontual da esperança de y, dado x, e a construção de intervalo de predição para y, dado x, após o ajuste de modelo linear de regressão de $g(y)$ em x.

Para a estimação pontual da esperança de y, dado x, há vários estimadores na literatura. O que eles tem em comum é a utilização da transformação inversa e de quantidades decorrentes do ajuste do modelo linear de regressão de $g(y)$ em x. Em alguns destes estimadores fica explícita a interferência da variância do erro do modelo, σ^2. Esta interferência tem várias intensidades de atuação, conforme a transformação usada. Na transformação logaritmo, os estimadores podem vir a superestimar bastante as esperanças, se o valor de σ^2 é grande, e o seu estimador, $\hat{\sigma}^2$, for observado acima do valor verdadeiro,

σ^2. Na transformação raiz quadrada, o valor de σ^2 não interfere muito, desde que seja pequeno, em relação à magnitude da variável y. Portanto, o ajuste do modelo linear na escala transformada deve ser examinado com atenção, levando em conta a transformação usada.

Para a predição da variável y, dado x, é importante verificar a suposição de normalidade do erro no modelo Gauss-Markov de regressão de $g(y)$ em x. Não havendo evidência contra esta suposição, um intervalo de predição para $g(y)$, dado x, pode ser construído da maneira usual, e, usando a inversa da transformação g nos limites deste intervalo, está definido um intervalo de predição para y. Scott e Wild (1993) apresentam um exemplo onde ajustam o modelo linear de regressão de Gauss-Markov normal para duas transformações da variável resposta, separadamente. Nos dois casos, a suposição de normalidade do erro parece estar sendo atendida. Intervalos de predição para a variável resposta, na escala original, são obtidos usando a transformação inversa nos limites do intervalo de predição para a variável transformada. Estes intervalos são praticamente idênticos, para as duas transformações usadas, apesar destas transformações serem bem diferentes e apesar de que o ajuste do modelo para uma transformação ter sido bom e para a outra muito pobre. O objetivo deste exemplo foi o de mostrar que não faz nenhum sentido comparação de ajustes de modelos em escalas diferentes, e que para o intervalo de predição é necessário examinar a distribuição do erro.

No contexto descrito acima, seriam de grande utilidade:

- um procedimento para verificar a suposição de normalidade dos erros, no modelo Gauss-Markov, através dos resíduos;
- uma definição de intervalo de predição para o modelo de regressão Gauss-Markov, sem a suposição de normalidade dos erros.

Desenvolvemos neste trabalho um procedimento gráfico, que baseado na função de distribuição empírica dos resíduos de um modelo de Gauss-Markov normal, produz uma região crítica para detecção de falta de normalidade. Tomamos como base o desenvolvimento de Lange e Ryan (1989) que analisaram a função de distribuição empírica, pon-

derada, de efeitos aleatórios, de modelo linear misto. O resultado é o conhecido gráfico 'qq-plot', porém, com distribuição ajustada à estimação de parâmetros, com região crítica delimitada.

Com relação a um intervalo de predição sem a exigência de normalidade no erro do modelo de Gauss-Markov de regressão , investigamos o método proposto por Schmoyer (1992). Se a distribuição do erro não tiver uma assimetria muito acentuada, a utilização deste intervalo de predição parece ser adequada.

A sequência dos capítulos é a seguinte:

• No Capítulo 2 são apresentados os resultados em que se baseam procedimentos para verificação de suposição de normalidade do erro no modelo de regressão de Gauss-Markov.

• No Capítulo 3 são apresentados estimadores para a esperança de y, dado x.

• No Capítulo 4 são investigados alguns aspectos dos estimadores do capítulo 3, inclusive uma simulação para comparação dos mesmos.

• No Capítulo 5, o intervalo de predição de y, sob o modelo Gauss-Markov, sem normalidade, é apresentado.

No apêndice são apresentados breves resultados teóricos. Os programas desenvolvidos em linguagem SAS para a obtenção de estimadores pontuais e intervalos de predição são apresentados em disquete.

Os gráficos exibidos foram elaborados com o uso de S-plus.

Capítulo 2

Procedimentos para Verificação de Suposição de Normalidade

2.1 Introdução

Frequentemente, uma modelagem estatística se baseia em alguma suposição de normalidade. Torna-se necessário a adoção de procedimentos que verifiquem a adequação de tal suposição.

Neste capítulo, inicialmente, é apresentado um resultado geral de Pierce (1982) que define distribuições assintóticas para certo tipo de estatística. Tipicamente, essas estatísticas são funções de estimadores de um ou mais parâmetros. A distribuição assintótica da estatística é obtida por intermédio da distribuição desta estatística, considerando-se que os parâmetros sejam conhecidos. Duas aplicações são desenvolvidas.

A primeira aplicação, trata-se de estatística de assimetria, descrita de maneira sumária em Pierce (1982), aqui, com mais detalhes. Sob normalidade, a estatística de assimetria é aproximadamente $N(0, \frac{6}{n})$, onde n é o tamanho da amostra. Valores maiores, em módulo, que $2\sqrt{\frac{6}{n}}$, para a estatística de assimetria forneceriam evidência contra a suposição de que a amostra se origina de modelo normal.

A segunda aplicação concerne a distribuição empírica, na presença de estimadores de parâmetros. Lange e Ryan (1989) analisaram a função de distribuição empírica, ponderada, de efeitos aleatórios, de modelo linear misto. Neste trabalho, lidamos com

uma situação mais simples, que, no que possa ser comparado aos resultados de Lange e Ryan, parece haver concordância. O objetivo aqui é produzir um gráfico de normalidade com região demarcada, que indicaria violação à suposição de normalidade da amostra.

As duas aplicações se referem ao contexto do modelo Gauss-Markov normal de regressão simples, abaixo descrito:

$$\underset{\sim}{y} = \mathbf{X}\ \underset{\sim}{\beta} + \underset{\sim}{\epsilon}, \qquad \underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2 \mathbf{I_n}),$$

onde $\underset{\sim}{y}^T = (y_1, y_2, ..., y_n)$ é o vetor de observações de uma variavel aleatória y; $\mathbf{X}$ é matriz, de posto 2, com colunas $\underset{\sim}{1}$ e $\underset{\sim}{x}$, onde $\underset{\sim}{x}^T = (x_1, x_2, ..., x_n)$ são valores da variável regressora, x; $\underset{\sim}{\beta}^T = (\beta_0, \beta_1)$ é o vetor de parâmetros; $\underset{\sim}{\epsilon}$ é o vetor de erros e $\mathbf{I}_n$ é a matriz de identidade de dimensão $n \times n$. A adequação da suposição de normalidade de $\underset{\sim}{\epsilon}$ seria investigada através do exame dos resíduos.

2.2 Distribuição Assintótica Ajustada à Estimação de Parâmetros

Pierce (1982) apresentou o seguinte resultado.

Seja $y_1, y_2, ...$ uma sequência de variáveis aleatórias com distribuição conjunta que depende de um vetor de parâmetros $\underset{\sim}{\lambda}$. Estas variáveis aleatórias não são necessariamente independentes ou identicamente distribuídas. Seja $\underset{\sim}{\hat{\lambda}}_n\ (y_1, y_2, ..., y_n)$ uma sequência de estimadores para $\underset{\sim}{\lambda}$, eficiente e assintoticamente normal, denotada por $\underset{\sim}{\hat{\lambda}}_n$. Considere a sequência $T_n = T_n(y_1, y_2, ..., y_n; \underset{\sim}{\lambda})$ com distribuição assintoticamente normal, cuja média assintótica é constante em $\underset{\sim}{\lambda}$.

O objetivo é achar a distribuição limite da estatística $\hat{T}_n = T_n(y_1, y_2, ..., y_n; \underset{\sim}{\hat{\lambda}}_n)$, onde $\underset{\sim}{\lambda}$ é estimado por $\underset{\sim}{\hat{\lambda}}_n$. Sob condições de regularidade e sem perda de generalidade, $\hat{T}_n$ tem distribuição assintoticamente normal com esperança zero. As condições de regularidade são:

1. $\begin{bmatrix} \sqrt{n}T_n \\ \sqrt{n}(\hat{\underset{\sim}{\lambda}}_n - \underset{\sim}{\lambda}) \end{bmatrix} \xrightarrow{D} \begin{bmatrix} L \\ \hat{\delta} \end{bmatrix} \sim N\left(\underset{\sim}{0}, \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}\right)$,

para todo $\underset{\sim}{\lambda}$, onde assume-se que V_{22} é não singular.

2. Existe uma matriz $\mathbf{B}$, que geralmente dependerá de $\underset{\sim}{\lambda}$, tal que $\sqrt{n}\hat{T}_n = \sqrt{n}T_n + \mathbf{B}\sqrt{n}(\hat{\underset{\sim}{\lambda}}_n - \underset{\sim}{\lambda}) + O(1)$. Se T_n é diferenciável em $\underset{\sim}{\lambda}$, este resultado vem da expansão em série de Taylor até primeira ordem onde

$$\mathbf{B} = \lim_{n\to\infty} E\left[\frac{\partial T_n}{\partial \underset{\sim}{\lambda}}\right].$$

3. $\hat{\underset{\sim}{\lambda}}_n$ é assintoticamente eficiente.

Atendidas estas três condições de regularidade, tem-se que

$$\sqrt{n}\hat{T}_n \xrightarrow{D} \hat{L} \sim N(0, V_{11} - \mathbf{B}V_{22}\mathbf{B}^T). \tag{2.1}$$

2.2.1 Distribuição aproximada de estatística de assimetria

Pierce (1982) apresentou a seguinte ilustração.

Considere-se variáveis aleatórias, independentes,

$$y_i \sim N(\beta_0 + \beta_1 x_{i1}, \sigma^2), \qquad i = 1, 2, ..., n.$$

A distribuicão de y_i depende do vetor de parâmetros $\underset{\sim}{\lambda} = (\beta_0, \beta_1, \sigma^2)$. Seja a estatística de assimetria

$$\hat{T}_n = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{(y_i - \hat{\beta}_{0n} - \hat{\beta}_{1n}x_{i1})}{\hat{\sigma}_n}\right\}^3,$$

onde $\hat{\underset{\sim}{\lambda}}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n}, \hat{\sigma}_n^2)$ é o vetor de estimadores de máxima verossimilhança de $\underset{\sim}{\lambda}$.

Resultado 2.1

Considerando os parâmetros, $\underset{\sim}{\lambda}= (\beta_0, \beta_1, \sigma^2)$, conhecidos, e

$$T_n = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{(y_i - \beta_0 - \beta_1 x_{i1})}{\sigma}\right\}^3,$$

então, $\sqrt{n}T_n \sim N(0, 15)$.

Prova:

$$\sqrt{n}T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{(y_i - \beta_0 - \beta_1 x_{i1})}{\sigma}\right\}^3,$$

assim

$$E[\sqrt{n}T_n] = 0$$

porque o terceiro momento de uma variável de distribuição $N(0,1)$ é zero.

Portanto,

$$Var[\sqrt{n}T_n] = \frac{1}{n}\sum_{i=1}^{n} Var\left\{\frac{(y_i - \beta_0 - \beta_1 x_{i1})}{\sigma}\right\}^3$$

$$= \frac{1}{n} n Var\left\{\frac{(y_1 - \beta_0 - \beta_1 x_{11})}{\sigma}\right\}^3,$$

pois, $y_i - \beta_0 - \beta_1 x_{i1}$, $i = 1, 2, ..., n$, são independentes e identicamente distribuídas,

$$= Var\left\{\frac{(y_1 - \beta_0 - \beta_1 x_{11})}{\sigma}\right\}^3$$

$$= E\left[\left(\frac{y_1 - \beta_0 - \beta_1 x_{11}}{\sigma}\right)^6\right] - E^2\left[\left(\frac{y_1 - \beta_0 - \beta_1 x_{11}}{\sigma}\right)^3\right]$$

$$= E\left[\left(\frac{y_1 - \beta_0 - \beta_1 x_{11}}{\sigma}\right)^6\right]$$

$$= \frac{6!}{8 \times 3 \times 2},$$

pela função geratriz de momentos,

$$= 15 \,. \tag{2.2}$$

Portanto, para todo n, temos que

$$\sqrt{n}T_n \sim N(0, 15). \quad \square$$

Resultado 2.2

Usando os estimadores de máxima verossimilhança dos parâmetros, $\hat{\underset{\sim}{\lambda}}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n}, \hat{\sigma}_n^2)$, e

$$\hat{T}_n = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{(y_i - \hat{\beta}_{0n} - \hat{\beta}_{1n}x_{i1})}{\hat{\sigma}_n}\right\}^3,$$

então, $\sqrt{n}\hat{T}_n \xrightarrow{D} N(0, 6)$.

Prova:

$$V_{11} = Var[\sqrt{n}T_n] = 15$$

conforme (2.2).

Precisamos definir

$$\mathbf{B} = \lim_{n\to\infty} E\left[\frac{\partial T_n}{\partial \underset{\sim}{\lambda}^*}\right]\Bigg|_{\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}},$$

onde $\underset{\sim}{\lambda}^* = (\beta_0^*, \beta_1^*, \sigma^{*^2})$ é uma variável matemática.

Diferenciando T_n com respeito a β_0^*, temos

$$\frac{\partial T_n}{\partial \beta_0^*} = \frac{\partial}{\partial \beta_0^*}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0^* - \beta_1^* x_{i1}}{\sigma^*}\right)^3\right\}$$

$$= -\frac{3}{n\sigma^*}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0^* - \beta_1^* x_{i1}}{\sigma^*}\right)^2,$$

e no ponto $\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}$, obtemos

$$\left.\frac{\partial T_n}{\partial \beta_0^*}\right|_{\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}} = -\frac{3}{n\sigma}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0 - \beta_1 x_{i1}}{\sigma}\right)^2,$$

cuja esperança é

$$-\frac{3}{\sigma}. \tag{2.3}$$

Diferenciando T_n com respeito a β_1^* temos

$$\frac{\partial T_n}{\partial \beta_1^*} = \frac{\partial}{\partial \beta_1^*}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0^* - \beta_1^* x_{i1}}{\sigma^*}\right)^3\right\}$$

$$= -\frac{3}{n\sigma^*}\sum_{i=1}^{n} x_{i1}\left(\frac{y_i - \beta_0^* - \beta_1^* x_{i1}}{\sigma^*}\right)^2,$$

e no ponto $\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}$, obtemos

$$\left.\frac{\partial T_n}{\partial \beta_1^*}\right|_{\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}} = -\frac{3}{n\sigma}\sum_{i=1}^{n} x_{i1}\left(\frac{y_i - \beta_0 - \beta_1 x_{i1}}{\sigma}\right)^2,$$

cuja esperança é

$$-\frac{3}{\sigma}\,\bar{x}, \tag{2.4}$$

onde $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_{i1}$.

Diferenciando T_n com respeito a σ^{*2}, temos

$$\frac{\partial T_n}{\partial \sigma^{*2}} = \frac{\partial}{\partial \sigma^{*2}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0^* - \beta_1^* x_{i1}}{\sigma^*}\right)^3\right\}$$

$$= -\frac{3}{2n\sigma^{*2}}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0^* - \beta_1^* x_{i1}}{\sigma^*}\right)^3,$$

e no ponto $\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}$, obtemos

$$\left.\frac{\partial T_n}{\partial \sigma^{*2}}\right|_{\underset{\sim}{\lambda}^*=\underset{\sim}{\lambda}} = -\frac{3}{2n\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0 - \beta_1 x_{i1}}{\sigma}\right)^3,$$

cuja esperança é

$$0. \tag{2.5}$$

Das equações (2.3), (2.4) e (2.5), temos

$$\mathbf{B} = -\frac{3}{\sigma}(1, \bar{x}, 0). \tag{2.6}$$

Agora,

$$\begin{aligned} V_{22} &= Var[\sqrt{n}(\hat{\underset{\sim}{\lambda}}_n - \underset{\sim}{\lambda})] \\ &= Var[\sqrt{n}\ \hat{\underset{\sim}{\lambda}}_n] \\ &= n\, Var[\hat{\underset{\sim}{\lambda}}_n]. \end{aligned} \tag{2.7}$$

Por resultados apresentados no apêndice, temos

$$Var[\hat{\underset{\sim}{\lambda}}_n] = \begin{bmatrix} \frac{\sigma^2}{ns^2}(s^2 + \bar{x}^2) & -\frac{\sigma^2}{ns^2}\bar{x} & 0 \\ -\frac{\sigma^2}{ns^2}\bar{x} & \frac{\sigma^2}{ns^2} & 0 \\ 0 & 0 & \frac{2(n-2)\sigma^4}{n^2} \end{bmatrix},$$

onde $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 - \bar{x}^2$. Substituindo esta matriz em (2.7),

$$\begin{aligned} V_{22} &= n\, Var[\hat{\underset{\sim}{\lambda}}_n] \\ &= \begin{bmatrix} \frac{\sigma^2}{s^2}(s^2 + \bar{x}^2) & -\frac{\sigma^2}{s^2}\bar{x} & 0 \\ -\frac{\sigma^2}{s^2}\bar{x} & \frac{\sigma^2}{s^2} & 0 \\ 0 & 0 & \frac{2(n-2)\sigma^4}{n} \end{bmatrix}. \end{aligned} \tag{2.8}$$

Usando as equações (2.2), (2.6) e (2.8), temos

$$V_{11} - \mathbf{B}V_{22}\mathbf{B}^T = 15 - \frac{3}{\sigma}(1, \bar{x}, 0) \begin{bmatrix} \frac{\sigma^2}{s^2}(s^2 + \bar{x}^2) & -\frac{\sigma^2}{s^2}\bar{x} & 0 \\ -\frac{\sigma^2}{s^2}\bar{x} & \frac{\sigma^2}{s^2} & 0 \\ 0 & 0 & \frac{2(n-2)\sigma^4}{n} \end{bmatrix} \frac{3}{\sigma} \begin{pmatrix} 1 \\ \bar{x} \\ 0 \end{pmatrix}$$

$$= 15 - \frac{9}{s^2}(s^2, 0, 0) \begin{pmatrix} 1 \\ \bar{x} \\ 0 \end{pmatrix}$$

$$= 6.$$

Portanto,

$$\sqrt{n}\hat{T}_n \xrightarrow{D} \hat{L} \sim N(0,6) \qquad ou \qquad \hat{T}_n \dot{\sim} N(0, \frac{6}{n}) \,. \quad \square$$

Nesta aplicação podemos dizer que $\hat{T}_n \dot{\sim} N(0, \frac{6}{n})$, para qualquer n, já que a estatística $\hat{T}_n$ e $\hat{\underset{\sim}{\lambda}}_n$ tem distribuição normal, para qualquer n.

2.3 Distribuição Aproximada de Função de Distribuição Empírica

Considere as variáveis aleatórias, independentes,

$$y_i \sim N(\beta_0 + \beta_1 x_{i1}, \sigma^2), \qquad i = 1, 2, ..., n,$$

e $\underset{\sim}{\lambda} = (\beta_0, \beta_1, \sigma^2)$ é o vetor de parâmetros.

Definimos as variáveis aleatórias

$$z_i = \frac{y_i - (\beta_0 + \beta_1 x_{i1})}{\sigma},$$

onde $z_i \overset{iid}{\sim} N(0,1)$, para $i = 1, 2, ..., n$, se $\underset{\sim}{\lambda}$ é conhecido.

A função de distribuição acumulada empírica de z_i é

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x - z_i), \tag{2.9}$$

onde

$$I(x - z_i) = \begin{cases} 1, & se\ x \geq z_i \\ 0, & c.c. \end{cases}.$$

Usando $\hat{\underset{\sim}{\lambda}}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n}, \hat{\sigma}_n^2)$, onde $\hat{\beta}_{0n}, \hat{\beta}_{1n}$ são estimadores de máxima verossimilhança de β_0 e β_1, respectivamente e [1]$\hat{\sigma}_n^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_{0n} - \hat{\beta}_{1n}x_{i1})^2$, definimos as variáveis aleatórias

$$\hat{z}_i = \frac{y_i - (\hat{\beta}_{0n} + \hat{\beta}_{1n}\, x_{i1})}{\hat{\sigma}_n},\ i = 1, 2, ..., n,$$

e

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x - \hat{z}_i), \tag{2.10}$$

onde

$$I(x - \hat{z}_i) = \begin{cases} 1, & se\ x \geq \hat{z}_i \\ 0, & c.c. \end{cases}.$$

Temos alguns resultados conhecidos sobre a função F.

Resultado 2.3

Supondo $\underset{\sim}{\lambda}$ conhecido,

$$E[F_n(x)] = \Phi(x),$$

onde $\Phi(x)$ é o valor da função de distribuição acumulada da normal padrão, no ponto x, e

$$Cov[F_n(x), F_n(y)] = \frac{1}{n}\Phi(x)[1 - \Phi(y)].$$

Prova:

[1]Por simplicidade, mantemos a notação $\hat{\sigma}_n^2$, apesar desta ter sido usada para o estimador de máxima verossimilhança na seção anterior.

$$E[F_n(x)] = \frac{1}{n}\sum_{i=1}^{n} E[I(x - z_i)]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Pr[z_i \leq x]$$

$$= \Pr[z_1 \leq x]$$

$$= \Phi(x).$$

Suponha $x \leq y$, então,

$$I(x - z_i) = 1 \Rightarrow I(y - z_i) = 1 \Rightarrow$$

$$E[I(x - z_i)I(y - z_i)] = \Phi(x),$$

e para $i \neq j$,

$$E[I(x - z_i)I(y - z_j)] = \Phi(x)\Phi(y).$$

Então,

$$E[F_n(x)F_n(y)] = E\left[\frac{1}{n^2}\sum_{i=1}^{n} I(x - z_i)\sum_{i=1}^{n} I(y - z_i)\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n} I(x - z_i)I(y - z_i) + \sum_{i=1}^{n}\sum_{j\neq i} I(x - z_i)I(y - z_j)\right]$$

$$= \frac{1}{n^2}\left[n\Phi(x) + n(n-1)\Phi(x)\Phi(y)\right].$$

Logo,

$$Cov[F_n(x), F_n(y)] = E[F_n(x)F_n(y)] - E[F_n(x)]E[F_n(y)]$$

$$= \frac{1}{n^2}\left[n\Phi(x) + n(n-1)\Phi(x)\Phi(y)\right] - \Phi(x)\Phi(y)$$

$$= \frac{1}{n}\Phi(x)[1 - \Phi(y)]. \quad \square$$

O objetivo é achar a distribuição aproximada de $\hat{F}_n(x)$, aplicando o resultado dado por Pierce. Definimos

$$T_n(x) = F_n(x) - \Phi(x), \tag{2.11}$$

$$\hat{T}_n(x) = \hat{F}_n(x) - \Phi(x),$$

e com os vetores de parâmetros $\underset{\sim}{\lambda} = (\beta_0, \beta_1, \sigma^2)$ e $\hat{\underset{\sim}{\lambda}}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n}, \hat{\sigma}_n^2)$, estas estatísticas satisfazem às condições de regularidade, dadas anteriormente, obtendo-se o resultado dado em (2.1).

Resultado 2.4

Considerando $\underset{\sim}{\lambda}$ conhecido e $T_n(x) = F_n(x) - \Phi(x)$, $\sqrt{n}T_n \xrightarrow{D} N(0, \Phi(x)[1 - \Phi(x)])$.

Prova:

Usando (2.11) e o resultado (2.3), calculamos

$$E[\sqrt{n}T_n] = \sqrt{n}E[F_n(x) - \Phi(x)]$$

$$= 0.$$

$$V_{11} = Var[\sqrt{n}T_n]$$

$$= Var[\sqrt{n}(F_n(x) - \Phi(x))]$$

$$= nVar[F_n(x)]$$

$$= n\frac{1}{n}\Phi(x)[1 - \Phi(x)],$$

pelo resultado 2.3,

$$= \Phi(x)[1 - \Phi(x)]. \tag{2.12}$$

O resultado segue usando o Teorema Central do Limite. □

Resultado 2.5

Usando os estimadores de máxima verossimilhança de $\underset{\sim}{\lambda}$, $\hat{\underset{\sim}{\lambda}}_n$, e $\hat{T}_n(x) = \hat{F}_n(x) - \Phi(x)$, então, $\sqrt{n}\hat{T}_n \xrightarrow{D} N\left(0, \Phi(x)[1 - \Phi(x)] - \phi^2(x)\left(1 + \frac{nx^2}{2(n-2)}\right)\right)$.

Prova:

$$V_{11} = Var[\sqrt{n}T_n] = \Phi(x)[1 - \Phi(x)]$$

conforme (2.12).

Precisamos

$$\mathbf{B} = \lim_{n\to\infty} E\left[\frac{\partial T_n}{\partial \underset{\sim}{\lambda^*}}\right]\Bigg|_{\underset{\sim}{\lambda^*}=\underset{\sim}{\lambda}}$$

$$= \lim_{n\to\infty} \left[\frac{\partial E[T_n]}{\partial \underset{\sim}{\lambda^*}}\right]\Bigg|_{\underset{\sim}{\lambda^*}=\underset{\sim}{\lambda}},$$

onde $\underset{\sim}{\lambda^*} = (\beta_0^*, \beta_1^*, \sigma^{*^2})$ é uma variável matemática.

Expressando a esperança de T_n em termos da variável matemática $\underset{\sim}{\lambda^*}$,

$$E[T_n] = \frac{1}{n}\sum_{i=1}^{n} \Pr\left[\frac{y_i - (\beta_0^* + \beta_1^* x_{i1})}{\sigma^*} \leq x\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Pr\left[y_i \leq x\sigma^* + (\beta_0^* + \beta_1^* x_{i1})\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Pr\left[\frac{y_i - (\beta_0 + \beta_1 x_{i1})}{\sigma} \leq \frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Pr\left[z_i \leq \frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right).$$

Diferenciando $E[T_n]$ com respeito a β_0^*, temos

$$\frac{\partial E[T_n]}{\partial \beta_0^*} = \frac{\partial}{\partial \beta_0^*}\left\{\frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma}\phi\left(\frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right),$$

onde $\phi(x)$ é a função densidade da normal padrão, e no ponto $\underset{\sim}{\lambda^*} = \underset{\sim}{\lambda}$, obtemos

$$\frac{\partial E[T_n]}{\partial \beta_0^*}\Bigg|_{\underset{\sim}{\lambda^*}=\underset{\sim}{\lambda}} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma}\phi(x)$$

$$= \frac{1}{\sigma}\phi(x). \tag{2.13}$$

Diferenciando $E[T_n]$ com respeito a β_1^*, temos

$$\frac{\partial E[T_n]}{\partial \beta_1^*} = \frac{\partial}{\partial \beta_1^*}\left\{\frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{x_{i1}}{\sigma}\phi\left(\frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right),$$

e no ponto $\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}$, obtemos

$$\left.\frac{\partial E[T_n]}{\partial \beta_1^*}\right|_{\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_{i1}}{\sigma}\phi(x)$$

$$= \frac{\bar{x}}{\sigma}\phi(x). \tag{2.14}$$

Diferenciando $E[T_n]$ com respeito a σ^{*^2}, temos

$$\frac{\partial E[T_n]}{\partial \sigma^{*^2}} = \frac{\partial}{\partial \sigma^{*^2}}\left\{\frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right)\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{x}{2\sigma^*\sigma}\phi\left(\frac{x\sigma^* + (\beta_0^* + \beta_1^* x_{i1}) - (\beta_0 + \beta_1 x_{i1})}{\sigma}\right),$$

e no ponto $\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}$, obtemos

$$\left.\frac{\partial E[T_n]}{\partial \sigma^{*^2}}\right|_{\underset{\sim}{\lambda}^* = \underset{\sim}{\lambda}} = \frac{1}{2n\sigma^2}\sum_{i=1}^{n}x\phi(x)$$

$$= \frac{x}{2\sigma^2}\phi(x). \tag{2.15}$$

Das equações (2.13), (2.14) e (2.15),

$$\mathbf{B} = \frac{1}{\sigma}\phi(x)(1, \bar{x}, \frac{x}{2\sigma}). \tag{2.16}$$

Pelas equações (2.12), (2.16) e apêndice, temos

$$V_{11} - \mathbf{B}V_{22}\mathbf{B}^T = \Phi(x)[1-\Phi(x)] -$$

$$\frac{1}{\sigma}\phi(x)(1, \bar{x}, \frac{x}{2\sigma}) \begin{bmatrix} \frac{\sigma^2}{s^2}(s^2 + \bar{x}^2) & -\frac{\sigma^2}{s^2}\bar{x} & 0 \\ -\frac{\sigma^2}{s^2}\bar{x} & \frac{\sigma^2}{s^2} & 0 \\ 0 & 0 & \frac{2n\sigma^4}{n-2} \end{bmatrix} \frac{1}{\sigma}\phi(x) \begin{pmatrix} 1 \\ \bar{x} \\ \frac{x}{2\sigma} \end{pmatrix}$$

$$= \Phi(x)[1-\Phi(x)] - \phi^2(x)\left(1, 0, \frac{nx\sigma}{n-2}\right) \begin{pmatrix} 1 \\ \bar{x} \\ \frac{x}{2\sigma} \end{pmatrix}$$

$$= \Phi(x)[1-\Phi(x)] - \phi^2(x)\left(1 + \frac{nx^2}{2(n-2)}\right). \quad \square$$

Resultado 2.6

$$E[\Phi^{-1}(\hat{F}_n(x))] \doteq x$$

e

$$Var[\Phi^{-1}(\hat{F}_n(x))] \doteq \frac{1}{n\phi^2(x)}\left[\Phi(x)[1-\Phi(x)] - \phi^2(x)\left(1 + \frac{nx^2}{2(n-2)}\right)\right].$$

Prova:

Pela aproximação em série de Taylor, de primeira ordem, temos:

$$E[\Phi^{-1}(\hat{F}_n(x))] \doteq \Phi^{-1}(E[\hat{F}_n(x)])$$

$$= \Phi^{-1}(\Phi(x))$$

$$= x.$$

$$Var[\Phi^{-1}(\hat{F}_n(x))] \doteq \left\{\frac{\partial}{\partial u}\left[\Phi^{-1}(u)\right]\Big|_{u=\Phi(x)}\right\}^2 Var(\hat{F}_n(x)),$$

porque, se $\Phi(t) = u$, ou seja, $t = \Phi^{-1}(u)$, então,

$$\frac{\partial}{\partial t}\Phi(t) = \phi(t)$$

$$= \phi(\Phi^{-1}(u)).$$

Portanto,

$$\frac{\partial}{\partial u}[\Phi^{-1}(u)] = \frac{\partial t}{\partial u} = \frac{1}{\frac{\partial u}{\partial t}} = \frac{1}{\frac{\partial}{\partial t}\Phi(t)} = \frac{1}{\phi(t)} = \frac{1}{\phi(\Phi^{-1}(u))},$$

e, no ponto $u = \Phi(x)$,

$$\frac{\partial}{\partial u}\left[\Phi^{-1}(u)\right]\Big|_{u=\Phi(x)} = \frac{1}{\phi(x)}.$$

Assim,

$$Var[\Phi^{-1}(\hat{F}_n(x))] \simeq \frac{1}{n\phi^2(x)}\left[\Phi(x)[1-\Phi(x)] - \phi^2(x)\left(1 + \frac{nx^2}{2(n-2)}\right)\right]. \quad \square$$

Pelos resultados apresentados, um gráfico dos valores observados de z_i versus $\Phi^{-1}(\hat{F}_n(z_i))$, para $i = 1, 2, ..., n$, deveria ser aproximadamente linear, passando na origem, com inclinação aproximadamente igual a 1, sob suposição de normalidade. Além disso, para $k > 0$, a região dada por

$$z_i \pm k\left\{\frac{1}{n\phi^2(x)}\left[\Phi(x)[1-\Phi(x)] - \phi^2(x)\left(1 + \frac{nx^2}{2(n-2)}\right)\right]\right\}^{\frac{1}{2}},$$

poderia ser considerada como uma região indicadora de falta de evidência contra a suposição de normalidade. Ou seja, se os pares $(z_i, \Phi^{-1}(F_n(z_i)))$ se locarizarem dentro desta região, não estão comprometendo a suposição de normalidade.

Lange e Ryan (1989) definem uma região de confiança para verificar se os efeitos aleatórios de modelo linear misto são normais. Eles sugerem o uso de $k = 1$, ou seja, admitir uma flutuação dentre um desvio padrão.

Portanto, seguindo o mesmo rigor de Lange e Ryan, definimos a região crítica

$$z_i \pm \left\{ \frac{1}{n\phi^2(x)} \left[\Phi(x)[1-\Phi(x)] - \phi^2(x) \left(1 + \frac{nx^2}{2(n-2)} \right) \right] \right\}^{\frac{1}{2}}.$$

Observa-se que a razão entre os desvios padrões, de $\hat{F}_n(x)$ e $F_n(x)$,

$$\frac{(Var[\hat{F}_n(x)])^{\frac{1}{2}}}{(Var[F_n(x)])^{\frac{1}{2}}},$$

é aproximadamente 0,5, para valores de x próximos a zero (sendo 0,6 para x igual a zero), e esta razão decresce um pouco para valores de x afastados de zero.

2.4 Ilustrações

Os procedimentos apresentados nas seções anteriores são utilizados em oito amostras, para verificar a adequação da suposição de normalidade, no ajuste de uma reta. Para cada amostra temos os gráficos dos pares (y, x), aparecendo a reta de quadrados mínimos e o gráfico de normalidade, além do valor observado da estatística de assimetria.

As três primeiras amostras são conjuntos de dados de Carroll e Ruppert (1988). Estes conjuntos de dados claramente mostram situação de heterogeneidade de variância de y, dado x. Estão aqui expostos para vermos como se comportaria o gráfico de normalidade.

Na primeira amostra, os dados provém de um ensaio para conhecer o comportamento de uma enzima, esterase. Foi registrada a concentração observada de esterase num experimento de ligaduras onde o número de ligaduras foi contado. Vamos considerar o número de ligaduras como a variável resposta, y, e a concentração de esterase utilizada como a variável regressora, x. A figura (2-1) mostra evidência forte de violação da suposição de normalidade.

A segunda amostra, figura (2-2), se refere ao conjunto de dados observados no rio Skeena para modelar a relação entre a quantidade anual de ovas de peixes e a produção anual de peixes com tamanho suficiente para ser pescado, chamados 'recrutas'. Temos

o número de ovas de peixes como variável regressora, x, e o número de recrutas como variável resposta, y. A figura (2-2) não mostra evidência contra a suposição de normalidade, mas podemos observar que para x maior que 500, os resíduos padronizados correspondentes a estes valores de x, (com símbolo diferente), também têm os valores maiores, em módulo. Isto significa que a variância não é a mesma para todo x, sugerindo um provável aumento da variância com respeito a x.

A terceira amostra, figura (2-3), se refere ao conjunto de dados que foram recolhidos para comparar um método novo com um método velho ou de referência. O objetivo do estudo seria verificar a viabilidade de reproduzir as medidas do teste novo (variável resposta), usando as medidas do teste de referência (variável regresssora). A figura (2-3) mostra evidência forte de violação de suposição de normalidade.

As demais amostras (dos números quatro até oito) são dados simulados pelo modelo

$$y_i = 3 + x_{i1} + \epsilon_i, \quad i = 1, 2, ..., 30,$$

onde os valores entre 1 e 4,8 foram usados para x_{i1}. Para as amostras referentes às figuras (2-4), (2-5) e (2-6) foi utilizada a distribuição normal para o erro, $\epsilon_i \sim N(0, 0, 5)$, e para as amostras referentes às figuras (2-7) e (2-8), o erro tem distribuição exponencial, ajustada para esperança zero e variância 1. Apresentamos várias amostras de um mesmo modelo para ver a flutuação que existe entre elas.

Notemos que para a quarta, quinta e sexta amostras não temos evidência contra a suposição de normalidade, o que é adequado, já que os erros foram gerados segundo o modelo normal. Podemos notar que alguns resíduos se situam fora da faixa, porém próximos aos limites.

Para as amostras de número sete e oito, podemos observar que os pontos caem fora da região de confiança numa grande frequência. Nestes casos, os resíduos mais próximos de zero são justamente os que caem mais longe da região de confiança. Nota-se que os resíduos não assumem valores exclusivamente no intervalo $(-2, 2)$, o que deveria ocorrer sob normalidade.

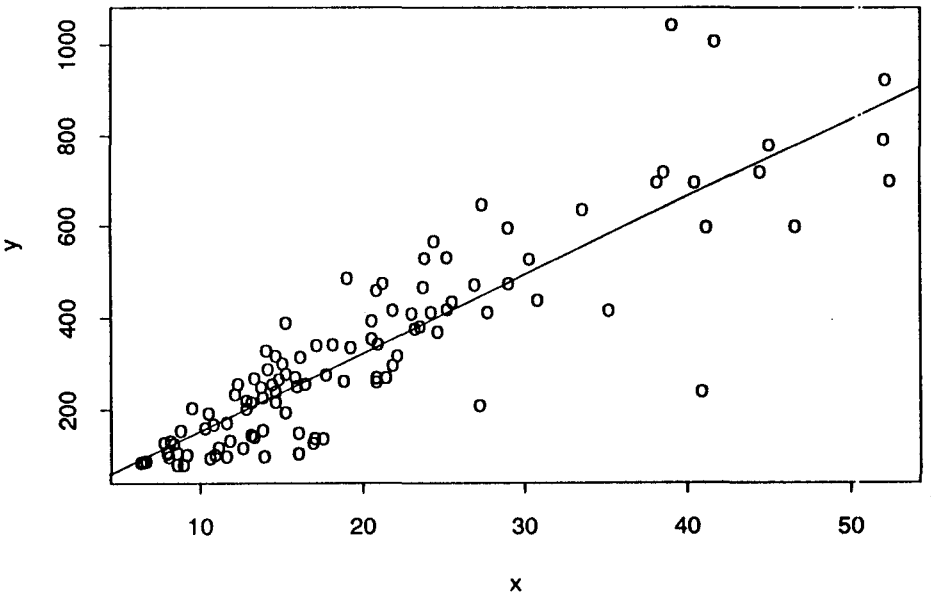
Quanto à estatística de assimetria foram observados os seguintes resultados:

- Nas amostras um, dois e três, os valores da estatística de assimetria são -0,09, 0,38 e 0,67, respectivamente, (desvios padrões: 0,23, 0,46 e 0,27, respectivamente), portanto, só para a primeira amostra o resultado foi diferente do que o resultado do gráfico de normalidade;

- Nas amostras quatro, cinco e seis (erros normais) foram observados valores próximos do desvio padrão, $\sqrt{\frac{6}{30}} = 0,45$; conclusão idêntica aos gráficos de normalidade;

- Nas amostras sete e oito (erros assimétricos), obtivemos um valor aproximadamente igual a três desvios padrões, indicando falta de normalidade, mas também obtivemos um valor um pouco acima do desvio padrão; uma conclusão diferente do gráfico de normalidade.

(a)



(b)

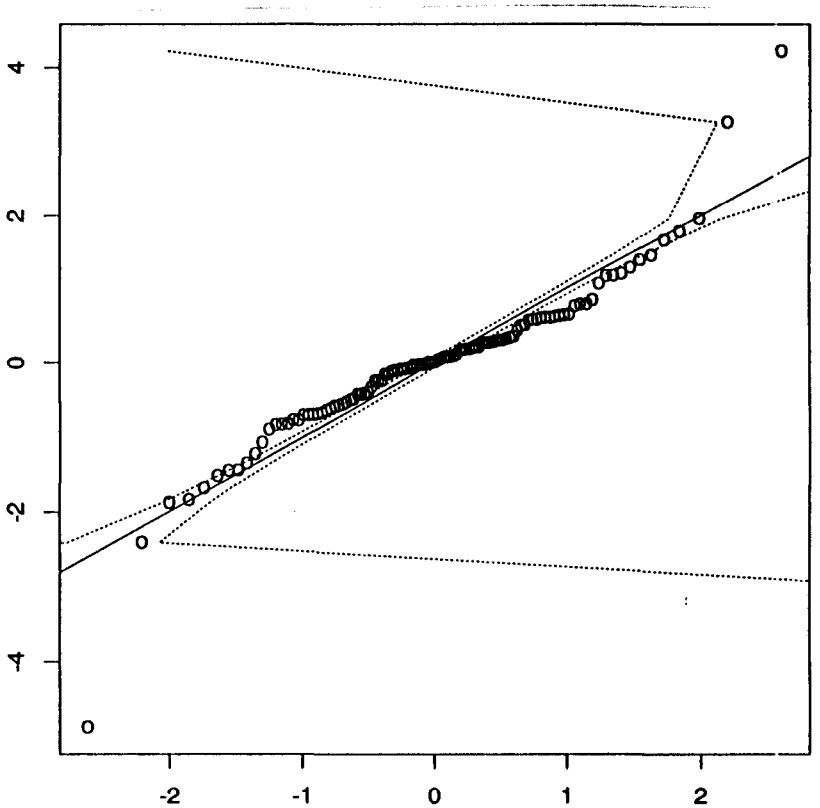
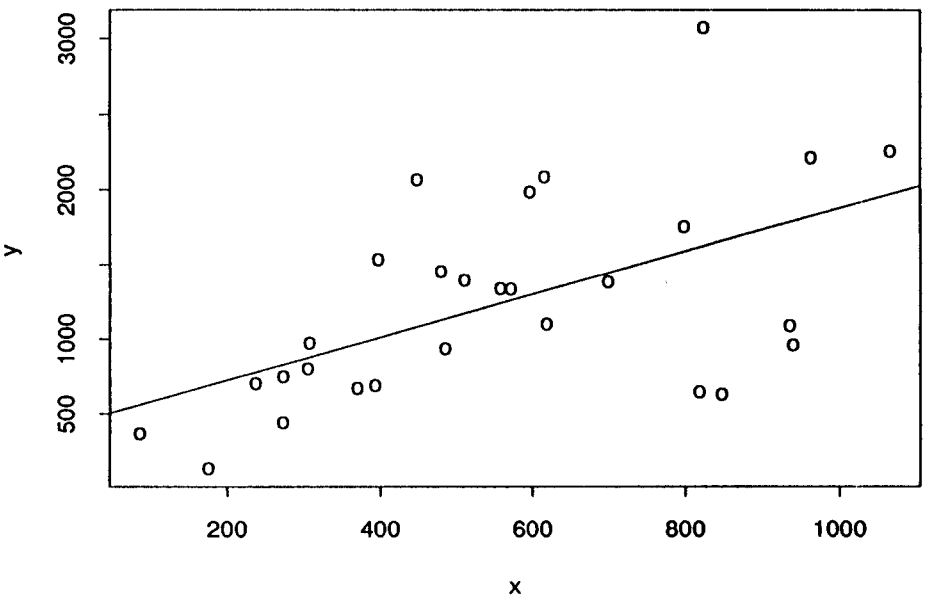


Figura 2-1: (a) Reta de quadrados mínimos de y =número de ligaduras e x = concentração de esterase; estatística de assimetria = -0,09; (b) Gráfico de normalidade ajustado

(a)



(b)

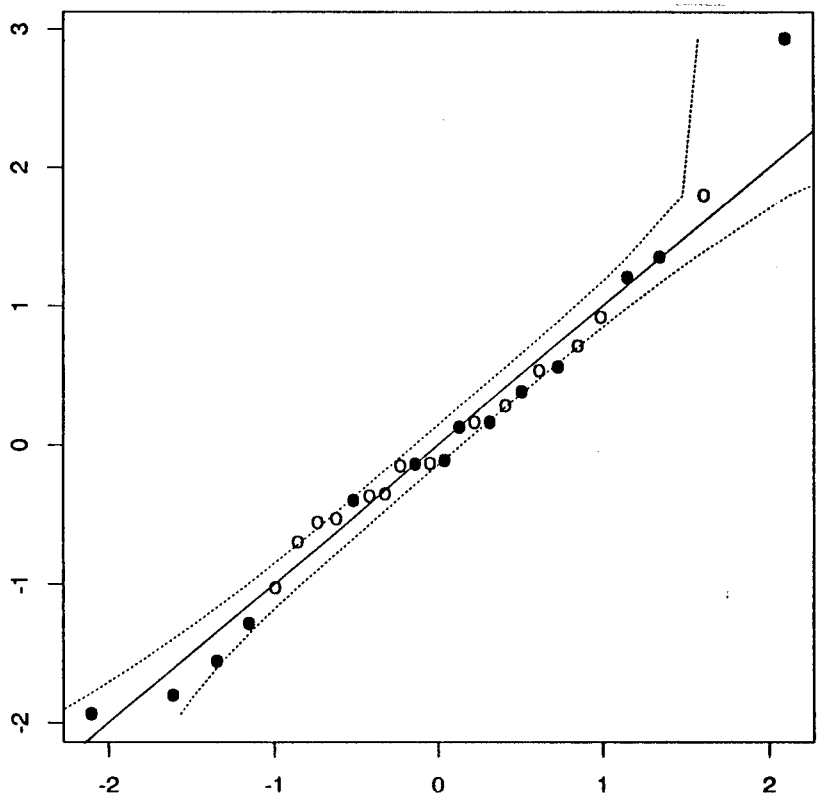
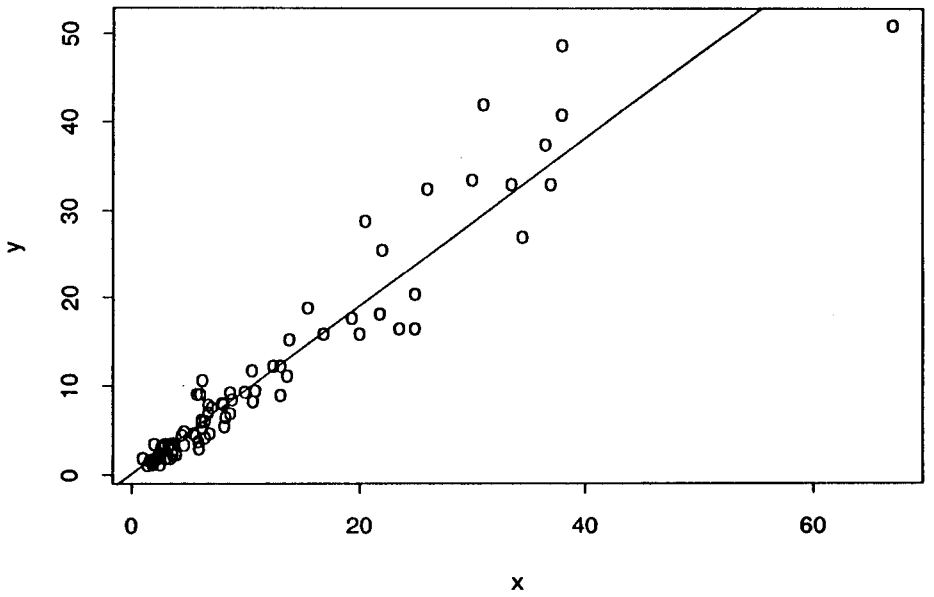


Figura 2-2: (a) Reta de quadrados mínimos de y =número de recrutas e x = número de ovas de peixes; estatística de assimetria = 0,38; (b) Gráfico de normalidade ajustado, onde o símbolo • indica os resíduos correspondentes a $x > 500$

(a)



(b)

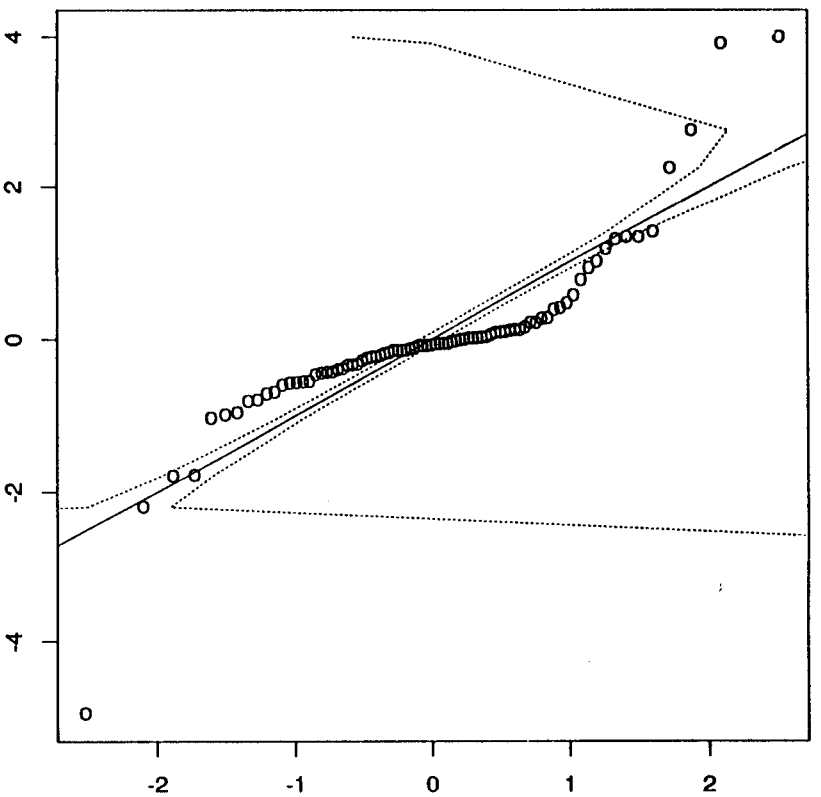
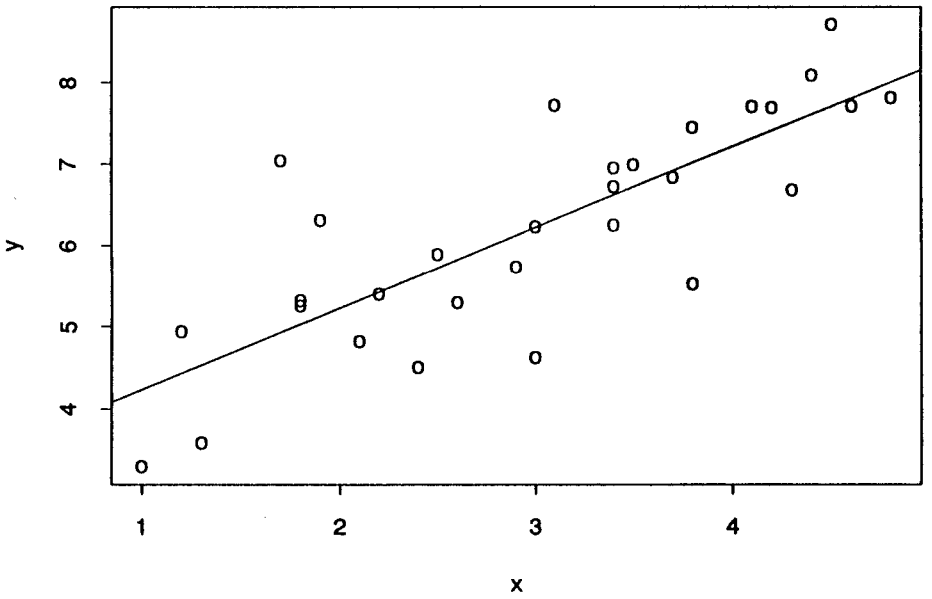


Figura 2-3: (a) Reta de quadrados mínimos de y =medidas do teste novo e x = medidas do teste de referência; estatística de assimetria = 0,67; (b) Gráfico de normalidade ajustado

(a)



(b)

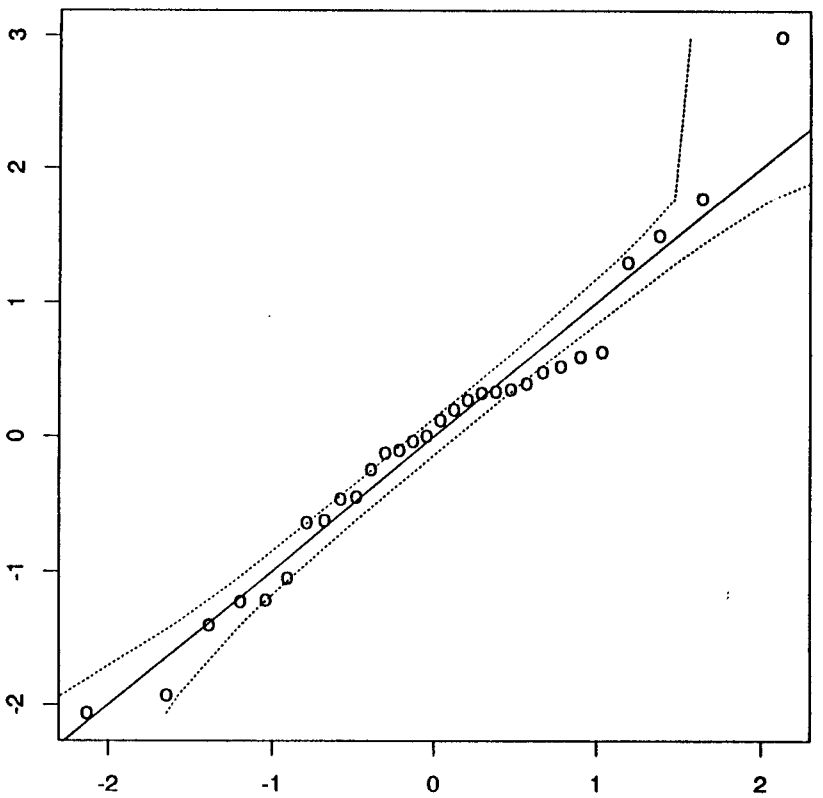
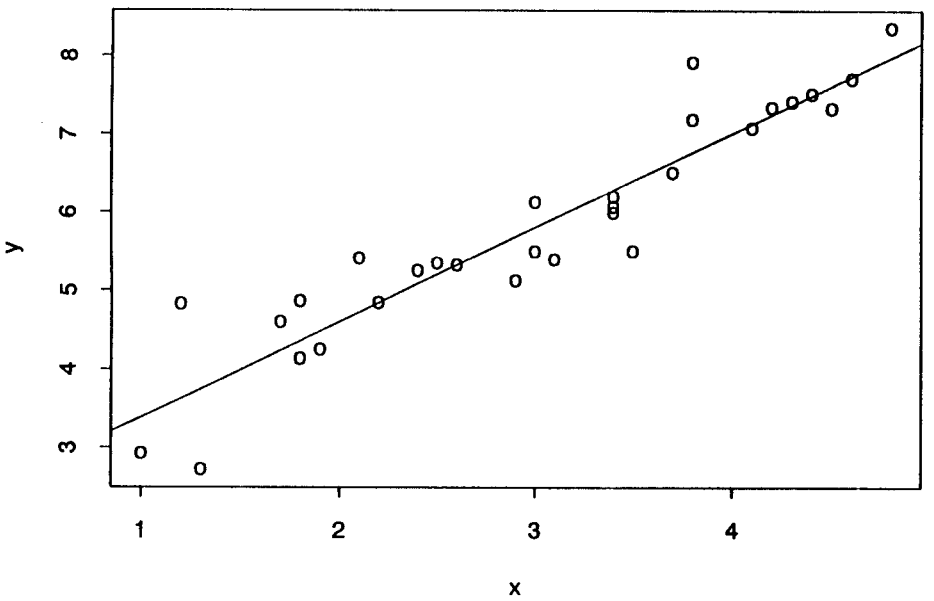


Figura 2-4: (a) Reta de quadrados mínimos de $y = 3 + x + \epsilon$, $\epsilon \sim N(0, 0, 5)$; estatística de assimetria = 0,25; (b) Gráfico de normalidade ajustado

(a)



(b)

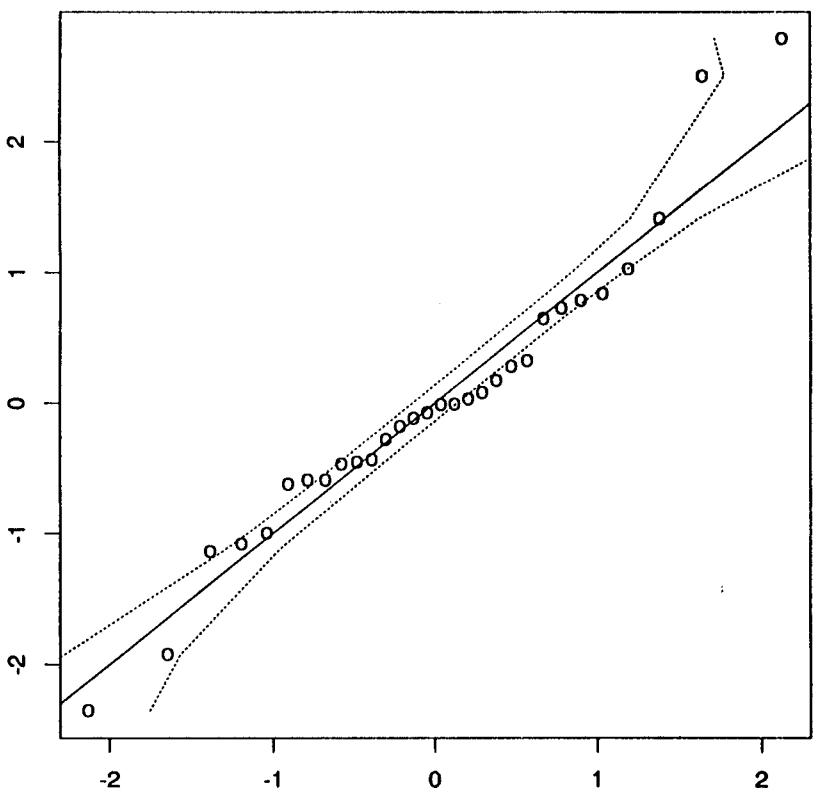
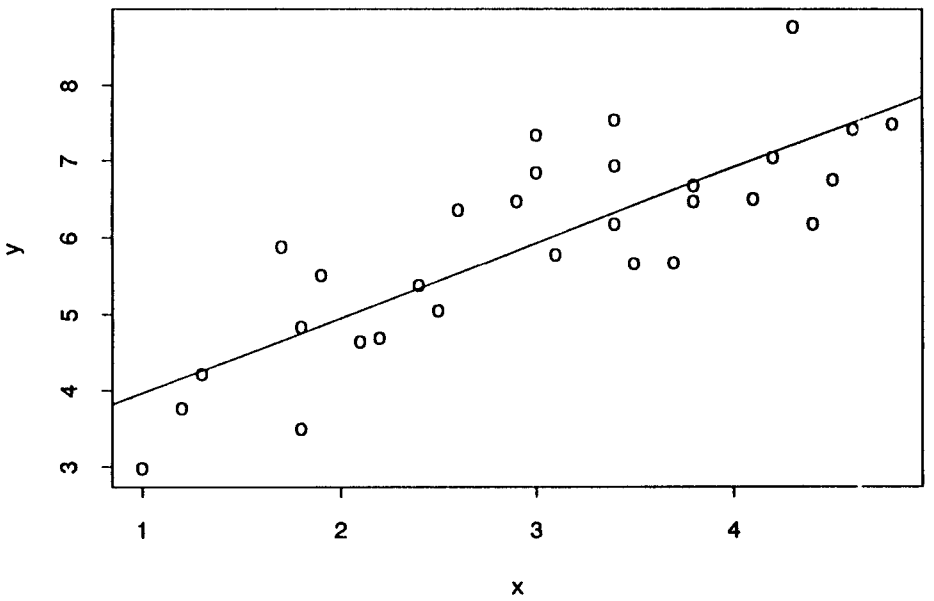


Figura 2-5: (a) Reta de quadrados mínimos de $y = 3 + x + \epsilon$, $\epsilon \sim N(0, 0, 5)$; estatística de assimetria = 0,40; (b) Gráfico de normalidade ajustado

(a)



(b)

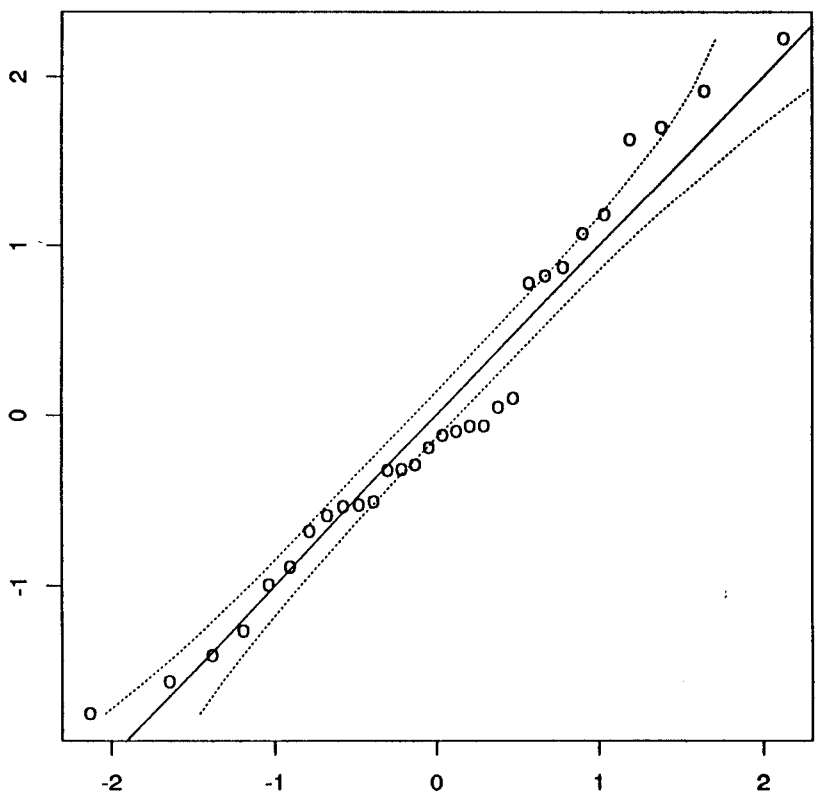
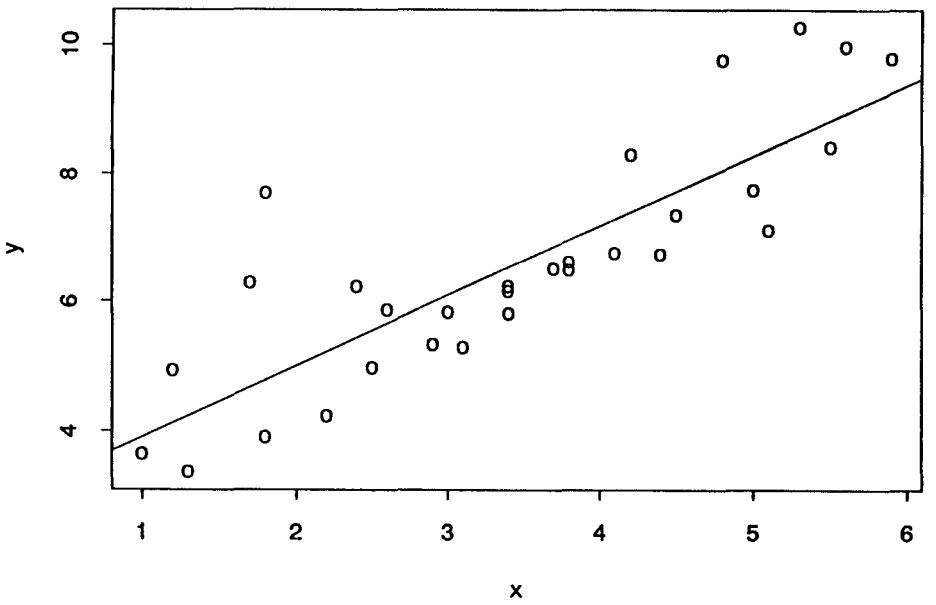


Figura 2-6: (a) Reta de quadrados mínimos de $y = 3 + x + \epsilon$, $\epsilon \sim N(0, 0, 5)$; estatística de assimetria = 0,46; (b) Gráfico de normalidade ajustado

(a)



(b)

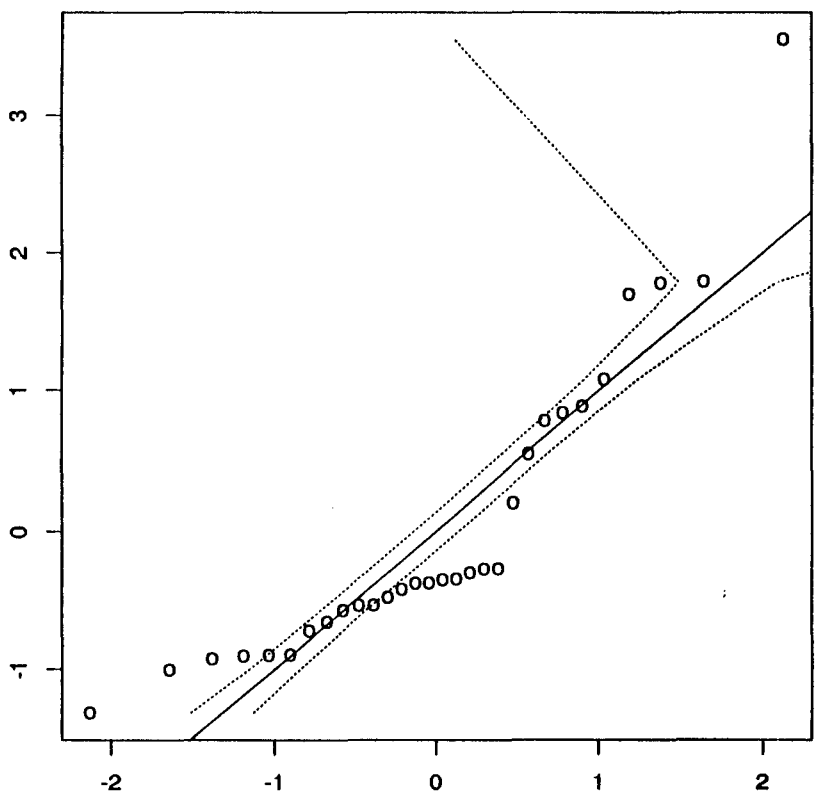
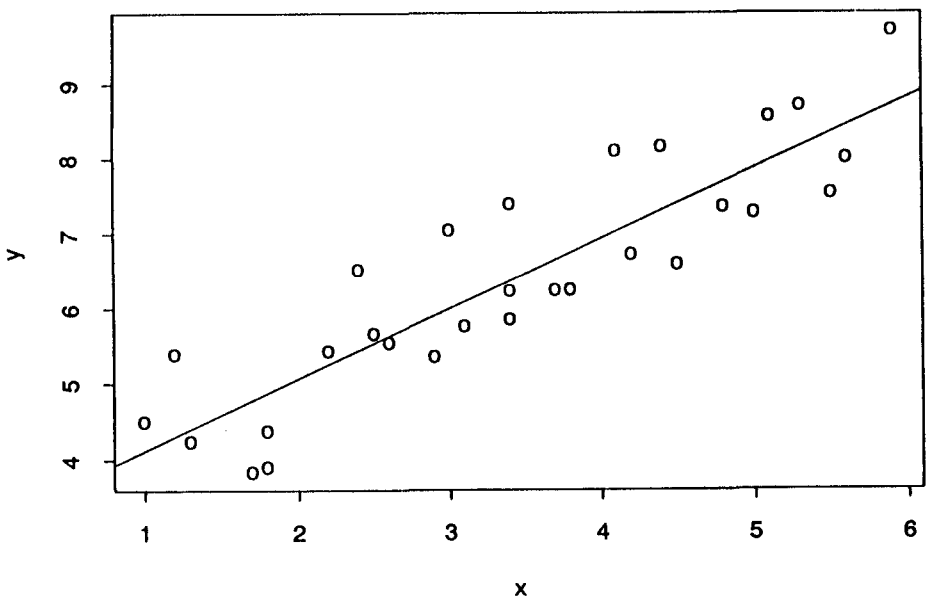


Figura 2-7: (a) Reta de quadrados mínimos de $y = 3+x+\epsilon$, ϵ tem distribuição exponencial ajustada com média zero e variância um; estatística de assimetria = 1,25; (b) Gráfico de normalidade ajustado

(a)



(b)

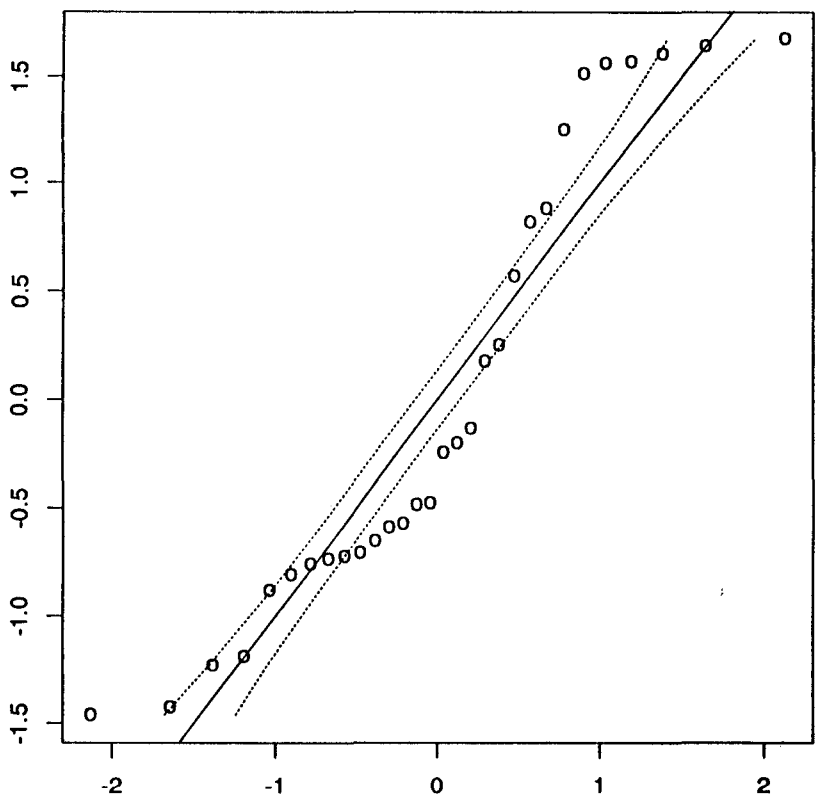


Figura 2-8: (a) Reta de quadrados mínimos de $y = 3+x+\epsilon$, ϵ tem distribuição exponencial ajustada com média zero e variância um; estatística de assimetria = 0,46; (b) Gráfico de normalidade ajustado

Capítulo 3

Estimadores para a Esperança, na Escala Original, após Transformação

3.1 Introdução

Inicialmente será descrito o modelo de Gauss-Markov, com destaque para as propriedades dos estimadores de quadrados mínimos. Uma metodologia que utilize, mesmo que indiretamente este modelo, poderia contar com as vantagens que o mesmo proporciona.

Definimos o modelo Gauss-Markov da seguinte forma

$$\underset{\sim}{y} = \mathbf{X}\,\underset{\sim}{\beta} + \underset{\sim}{\epsilon}, \tag{3.1}$$

onde $\underset{\sim}{y}^T = (y_1, y_2, ..., y_n)$ é o vetor de observações de uma variável aleatória y,
$\mathbf{X}$ é a matriz de modelo, de dimensão $n \times (p+1)$,
$\underset{\sim}{\beta}^T = (\beta_0, \beta_1, ..., \beta_p)$ é o vetor de parâmetros e
$\underset{\sim}{\epsilon}^T = (\epsilon_1, \epsilon_2, ..., \epsilon_n)$ é o vetor de variáveis aleatórias chamadas de erros,
com as suposições distribucionais

$$E[\epsilon_i] = 0,$$

$$Var[\epsilon_i] = \sigma^2,$$

$$E[\epsilon_i \epsilon_j] = 0, \quad i \neq j, \quad i, j = 1, 2, ..., n.$$

Se além das suposições mencionadas acima, assumimos também que os erros têm distribuição normal, ou seja,

$$\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2 \mathbf{I_n}),$$

onde $\mathbf{I_n}$ é a matriz identidade, de dimensão $n \times n$, então o modelo se chamará modelo Gauss-Markov normal.

O modelo de regressão linear é um caso particular do modelo (3.1) quando temos $\mathbf{X}^T = (\underset{\sim}{x_1}, \underset{\sim}{x_2}, ..., \underset{\sim}{x_n})$ como a matriz de modelo, identificando $\underset{\sim}{x_i}^T = (1, x_{i1}, ..., x_{ip})$ como o vetor de valores de p variáveis (regressoras), além do elemento 1, para a i-ésima observação, $i = 1, 2, ..., n$.

Neste trabalho trataremos, na maior parte do tempo, do modelo de regressão com suposição de normalidade. Seja o modelo considerado Gauss-Markov, com normalidade ou não, modelo de regressão ou não, vamos considerar a matriz $\mathbf{X}$ como sendo de posto coluna completo.

Uma forma de estimar o vetor de parâmetros $\underset{\sim}{\beta}$ para o modelo de regressão é usar o método de quadrados mínimos ordinários, que consiste em achar um estimador de $\underset{\sim}{\beta}$ que minimiza a soma dos quadrados das diferenças entre os valores observados e estimados, ou seja,

$$\min_{\underset{\sim}{b}} \{(\underset{\sim}{y} - \mathbf{X}\,\underset{\sim}{b})^T (\underset{\sim}{y} - \mathbf{X}\,\underset{\sim}{b})\}, \quad \underset{\sim}{b} \in R^{p+1}.$$

O estimador do vetor de parâmetros $\underset{\sim}{\beta}$ é

$$\underset{\sim}{\hat{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\,\underset{\sim}{y}, \tag{3.2}$$

onde a existência da inversa da matriz $\mathbf{X}^T\mathbf{X}$ está garantida por ser $\mathbf{X}$ de posto coluna completo.

O estimador $\hat{\underset{\sim}{\beta}}$ é o estimador de $\underset{\sim}{\beta}$, de menor variância dentre todos os estimadores de $\underset{\sim}{\beta}$, não viciados e lineares em $\underset{\sim}{y}$. Temos mais um parâmetro, σ^2, que é estimado por

$$\hat{\sigma}^2_{(1)} = \frac{(\underset{\sim}{y} - \mathbf{X}\,\hat{\underset{\sim}{\beta}})^T(\underset{\sim}{y} - \mathbf{X}\,\hat{\underset{\sim}{\beta}})}{n-p-1},$$

sendo $\hat{\sigma}^2_{(1)}$ é um estimador não viciado. Se temos um modelo Gauss-Markov normal, então, $\hat{\underset{\sim}{\beta}}$, dado em (3.2), é também o estimador de máxima verossimilhança, e o estimador de máxima verossimilhança de σ^2 é

$$\hat{\sigma}^2_{(2)} = \frac{(\underset{\sim}{y} - \mathbf{X}\,\hat{\underset{\sim}{\beta}})^T(\underset{\sim}{y} - \mathbf{X}\,\hat{\underset{\sim}{\beta}})}{n}.$$

Um estimador pontual da esperança de y, dado $\underset{\sim}{x}$, é

$$\hat{y} = \underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}$$

e uma propriedade importante deste estimador é que $\underset{\sim}{x}^T\hat{\underset{\sim}{\beta}}$ é o estimador de $\underset{\sim}{x}^T\underset{\sim}{\beta}$ de menor variância dentre os estimadores de $\underset{\sim}{x}^T\underset{\sim}{\beta}$, não viciados e lineares em $\underset{\sim}{y}$.

Na prática, quando tentamos o ajuste de um modelo da forma (3.1), estimamos o vetor de parâmetros pelo método de quadrados mínimos e usamos os resíduos observados,

$$\hat{\underset{\sim}{\epsilon}} = \underset{\sim}{y} - \hat{\underset{\sim}{y}},$$

para verificar as suposições feitas ao vetor de erros $\underset{\sim}{\epsilon}$.

Uma possível solução ao problema de não termos atendidas as suposições distribucionais do modelo Gauss-Markov normal é a transformação da variável resposta. Uma transformação pode afetar a forma da distribuição dos erros e o tipo da relação entre a tendência central da distribuição da variável resposta e os valores das variáveis regressoras.

Box e Cox (1964) propuzeram uma família de transformações de potência na variável

resposta, definindo o modelo:

$$y_i^{(\lambda)} = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \epsilon_i, \tag{3.3}$$

onde

$$\epsilon_i \overset{iid}{\sim} N(0, \sigma^2), \quad i = 1, 2, ..., n,$$

$$y_i^{(\lambda)} = \begin{cases} (y_i^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \ln(y_i), & \lambda = 0 \end{cases},$$

assumindo que para um dado λ, $y^{(\lambda)}$ é uma função monótona de y, sobre o intervalo admissível. A identidade e a transformação logarítmica são casos particulares desta família de transformações. Existem vários métodos para escolher o valor de λ. Usando o método de máxima verossimilhança, já que os erros têm distribuição normal, são definidos estimadores consistentes dos parâmetros $\underset{\sim}{\beta}$ e λ.

Existe o problema de interpretação dos parâmetros $\underset{\sim}{\beta}$ e λ, o que ocasionou uma discussão documentada em Bickel e Doksum (1981), Box e Cox (1982) e Hinkley e Runger (1984). Bickel e Doksum argumentaram que, considerando λ como um parâmetro extra, a ser estimado, a variância de $\underset{\sim}{\hat{\beta}}$ é muito maior, quando λ é estimado do mesmo conjunto de dados, do que quando λ é assumido conhecido. Hinkley e Runger alegam ser possível fazer inferência considerando o valor de λ como fixo. Salientaram também que o fato do valor de λ determinar a escala da variável resposta faz não comparáveis os modelos de diferentes λ 's. Não abordaremos a escolha do valor de λ nesta pesquisa pois é um tema extenso que por si mesmo merece um trabalho de investigação. Partiremos de um valor dado deste parâmetro.

Sabemos que o vetor $\underset{\sim}{\beta}$ representa os efeitos das variáveis regressoras na esperança de $y^{(\lambda)}$, a resposta transformada, dado um conjunto de valores $\underset{\sim}{x}$. Mas, os pesquisadores podem estar necessitando do conhecimento da relação entre as variáveis nas escalas originais. Tendo transformado a variável dependente, a idéia de interpretar o modelo ajustado na escala original nos leva automáticamente a pensar na inversa da transformação. Carroll e Ruppert (1984) sugeriram usar a inversa da transformação e fazer inferência na escala

original das variáveis. Pela proposição 1 do Apêndice, aplicar a função inversa à média de uma variável resposta transformada (por uma função estritamente monótona) nos dará a mediana e não a média da variável resposta na escala original de medida. Carroll e Ruppert estudaram as propriedades da mediana da variável resposta na escala original e demonstraram que há um custo ao estimar λ, mas este geralmente é pequeno e é muito menor que o custo obtido por Bickel e Doksum para os estimadores de $\underset{\sim}{\beta}$.

O procedimento de aplicação da função inversa da transformação, para voltar à escala original da variável resposta, é referido de várias maneiras na literatura. Por exemplo, Miller (1984), Seber e Wild (1989) usam o termo 'detransformation', Duan (1983), Carroll e Ruppert (1984), Taylor (1986), Sakia (1990) e Gianola e Hammond (1990) usam 'retransformation'.

Se quisermos predizer um valor futuro de y, dado $\underset{\sim}{x}$, então, a média (valor esperado) da variável resposta na escala original de medida será geralmente a quantidade de interesse (ver Miller, 1984). Mesmo que a escala transformada seja de muita utilidade, como acontece com o uso da escala logarítmica para medir magnitudes sísmicas, um objetivo importante em muitos casos é obter um estimador para a média na escala original (ver Shumway, Azari e Johnson, 1989). Vemos isto geralmente na área da economia, quando, por exemplo, a variável resposta é uma variável monetária e o valor da média é um resultado mais informativo do que o valor da mediana (ver Taylor, 1986). No caso de querermos estimar um valor total da variável resposta, usamos como estimador uma constante vezes o estimador da média (ver Miller, 1984). Box e Cox (1964) citam algumas situações em que se precisa da média para estimar o total: a produção total de um produto em um período longo de tempo, usando a média da produção por lote, e o número de componentes usadas em um período longo de tempo, usando o tempo médio de falha por componente. Rubin (1984) dá os exemplos de estimar produção total de petróleo por mês, quantidade total de poluição produzida por uma fonte, e gastos totais de um seguro de saúde.

Supondo que exista uma transformação na variável resposta com a qual conseguimos

o ajuste de um modelo de regressão linear, mostraremos a seguir alguns métodos para estimar a média ou esperança da variável resposta na escala original.

Na seção 3.2 descrevemos o trabalho desenvolvido por Miller (1984). Ele analisa algumas transformações mais comuns para transformar y, em geral, assumindo erros com distribuição normal. O objetivo é definir um estimador não viciado para a esperança de y. Este estimador está baseado nos estimadores de quadrados mínimos do modelo transformado. Miller obtém estimadores da média de y para modelos de regressão linear simples e neste trabalho o generalizamos para modelos de regressão linear múltipla, segundo o desenvolvimento feito para o caso anterior. Além disso, propomos outros estimadores para a média de y, com pequena modificação dos estimadores de Miller.

Na seção 3.3 mostramos o trabalho desenvolvido por Duan (1983). Ele considera quaisquer funções não lineares, que têm inversa, para transformar a resposta, (o caso linear não necessita desta metodologia, pois admite soluções muito mais simples) e propõe um estimador não paramétrico da esperança da variável resposta, na escala original. Não é assumida normalidade dos erros, apenas que os erros sejam independentes, identicamente distribuídos, com esperança zero.

Na seção 3.4 apresentamos um estimador de uma aproximação da esperança de y, após o uso da família de transformações de potência de Box-Cox, proposto por Taylor (1986). Taylor assume que os erros são independentes, identicamente distribuídos, com esperança zero e variância constante, e função de densidade simétrica.

Na seção 3.5 mostramos o estimador da esperança, dado por Taylor, aplicado a modelos lineares mistos balanceados, proposto por Sakia (1990). As transformações estudadas pertencem à família de transformações de potência de Box-Cox e é assumido que os erros são independentes com esperança zero e distribuição normal.

3.2 Estimador da Esperança de y após Transformação da Variável Resposta

Miller (1984) estudou a estimação da esperança da variável resposta na escala original para algumas transformações. As transformações estudadas foram as seguintes: logaritmo, raiz quadrada e inversa. Assumiu para todas um modelo de regressão linear simples com erros independentes e identicamente distribuídos com esperança zero, e assumiu normalidade para alguns modelos.

Miller apresentou o modelo (3.1), com uma variável regressora ($p = 1$), sendo $\underset{\sim}{\beta}^T = (\beta_0, \beta_1)$ o vetor de parâmetros, $\underset{\sim}{x}_i^T = (1, x_{i1})$, onde x_{i1} é a i-ésima observação da variável regressora.

Nós apresentaremos o desenvolvimento feito por Miller para definir um estimador da esperança de y no caso de regressão linear simples e o generalizamos ao caso de regressão linear múltipla, usando o modelo (3.1) com $p > 1$, ou seja, com mais de uma variável regressora.

Os estimadores de Miller podem ser vistos como a função inversa da transformação aplicada ao valor ajustado do modelo transformado, multiplicada ou somada a um termo de ajuste, que depende do estimador de quadrados mínimos de σ^2. Estes estimadores para a esperança de y são ligeiramente viciados pelo fato de ter usado o estimador de σ^2 no lugar do valor verdadeiro, σ^2.

Denotaremos por $\hat{\beta}_0, \hat{\beta}_1$ e $\hat{\sigma}^2$ os estimadores de quadrados mínimos de β_0, β_1 e σ^2 do modelo linear com a variável resposta transformada, e definiremos x_0 como um valor do intervalo $[x_{(1)}, x_{(n)}]$, onde $x_{(1)}$ e $x_{(n)}$ são os valores mínimo e máximo, respectivamente, da variável regressora.

3.2.1 Transformação logaritmo

Consideremos o modelo

$$\ln(y_i) = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \tag{3.4}$$

onde

$$y_i > 0, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2), \qquad i = 1, 2, ..., n.$$

Do modelo (3.4) temos que

$$\begin{aligned} y_i &= \exp(\ln(y_i)) \\ &= \exp(\beta_0 + \beta_1 x_{i1} + \epsilon_i) \\ &= \exp(\beta_0 + \beta_1 x_{i1}) \exp(\epsilon_i) \\ &= \exp(\beta_0 + \beta_1 x_{i1}) \epsilon_i^*, \qquad i = 1, 2, ..., n, \end{aligned}$$

onde $\epsilon_i^* = \exp(\epsilon_i)$. Logo, ϵ_i^* tem distribuição lognormal com parâmetros $(0, 1, \sigma)$, o que será denotado da seguinte maneira: $\epsilon_i^* \sim LN(0, 1, \sigma)$.

Pela distribuição de ϵ_i^* sabemos que $E[\epsilon_i^*] = \exp(\frac{\sigma^2}{2})$ e $Mediana[\epsilon_i^*] = 1$, então a esperança e a mediana da variável resposta, na escala original, são dadas por

$$E[y_i \mid x_{i1}] = \exp(\beta_0 + \beta_1 x_{i1}) \exp(\frac{\sigma^2}{2}) \tag{3.5}$$

e

$$Mediana[y_i \mid x_{i1}] = \exp(\beta_0 + \beta_1 x_{i1}). \tag{3.6}$$

Notamos que se o valor de σ^2 for muito pequeno, as medidas acima podem ser consideradas praticamente iguais.

Agora, trabalharemos com o modelo ajustado de (3.4),

$$\hat{E}[\ln(y) \mid x] = \hat{\beta}_0 + \hat{\beta}_1 x.$$

A variável $\hat{\beta}_0 + \hat{\beta}_1 x$ tem distribuição normal com

$$E[\hat{\beta}_0 + \hat{\beta}_1 x] = \beta_0 + \beta_1 x,$$

$$Var[\hat{\beta}_0 + \hat{\beta}_1 x] = \sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right],$$

pelo resultado 1 do apêndice, onde

$$\bar{x} = \frac{\sum_{i=1}^{n} x_{i1}}{n}.$$

Logo,

$$\exp(\hat{\beta}_0 + \hat{\beta}_1 x) \sim LN \left(0, \exp(\beta_0 + \beta_1 x), \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right]} \right)$$

e a esperança de $\exp(\hat{\beta}_0 + \hat{\beta}_1 x)$ é

$$E[\exp(\hat{\beta}_0 + \hat{\beta}_1 x)] = \exp(\beta_0 + \beta_1 x) \exp \left(\frac{\sigma^2}{2} \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right] \right). \tag{3.7}$$

Analisando (3.7) vemos que, quando o valor de n tende a infinito, a quantidade $\left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right]$ é aproximadamente zero e $\exp \left(\frac{\sigma^2}{2} \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right] \right)$ é próximo de um, logo, a esperança de $\exp(\hat{\beta}_0 + \hat{\beta}_1 x)$ atinge a mediana de y, dada em (3.6). Ou seja, quando n é grande, a esperança acima será um valor muito próximo da mediana da variável resposta, na escala original. Mas, o que nós desejamos é estimar o valor dado em (3.5), ou seja, queremos um estimador não viciado para a esperança de y, dado x.

Miller propõe estimar a esperança da variável resposta, na escala original, por:

$$\hat{E}[y \mid x] = \exp(\hat{\beta}_0 + \hat{\beta}_1 x) \exp(\frac{\hat{\sigma}^2}{2}). \tag{3.8}$$

Miller define este estimador para qualquer valor de n.

Agora, quando n não é muito pequeno,

$$\frac{1}{n} \leq \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right] \leq 1, \tag{3.9}$$

pelo resultado 2 do apêndice, o que leva a que

$$\exp\left(\frac{\sigma^2}{2}\left[\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]\right) \leq \exp(\frac{\sigma^2}{2}).$$

Assim, a esperança dada em (3.7) é, em geral, menor que a esperança da variável resposta na escala original, dada em (3.5), já que o fator $\exp\left(\frac{\sigma^2}{2}\left[\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]\right)$ aparece no lugar de $\exp(\frac{\sigma^2}{2})$.

Para não sub-estimar a esperança de y, dado x, inicialmente consideramos o estimador

$$\hat{E}[y \mid x] = \exp(\hat{\beta}_0+\hat{\beta}_1 x)\exp\left(\frac{\sigma^2}{2}\left[1-\frac{1}{n}-\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]\right).$$

Mas, como na prática os parâmetros não são conhecidos, propomos o estimador

$$\hat{E}[y \mid x] = \exp(\hat{\beta}_0+\hat{\beta}_1 x)\exp\left(\frac{\hat{\sigma}^2}{2}\left[1-\frac{1}{n}-\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]\right). \tag{3.10}$$

3.2.2 Transformação raiz quadrada

Consideremos o modelo

$$y_i^{\frac{1}{2}} = \beta_0+\beta_1 x_{i1}+\epsilon_i, \tag{3.11}$$

onde

$$y_i > 0, \quad \epsilon_i \overset{iid}{\sim} N(0,\sigma^2), \quad i = 1,2,...,n.$$

Portanto,

$$y_i = (\beta_0+\beta_1 x_{i1}+\epsilon_i)^2,$$

com esperança

$$E[y_i \mid x_{i1}] = E[(\beta_0+\beta_1 x_{i1}+\epsilon_i)^2 \mid x_{i1}]$$

$$= Var[\beta_0+\beta_1 x_{i1}+\epsilon_i \mid x_{i1}] + E^2[\beta_0+\beta_1 x_{i1}+\epsilon_i \mid x_{i1}]$$

$$= \sigma^2 + (\beta_0 + \beta_1 x_{i1})^2, \tag{3.12}$$

e, pela proposição 1 do apêndice, a mediana é

$$Mediana[y_i \mid x_{i1}] = (\beta_0 + \beta_1 x_{i1})^2. \tag{3.13}$$

É importante observar que, quando σ^2 é um valor pequeno em relação à magnitude da variável y, a esperança e a mediana podem ser consideradas aproximadamente iguais.

Agora, trabalharemos com o modelo ajustado de (3.11)

$$\hat{E}[y^{\frac{1}{2}} \mid x] = \hat{\beta}_0 + \hat{\beta}_1 x.$$

A esperança de $(\hat{\beta}_0 + \hat{\beta}_1 x)^2$ é

$$E[(\hat{\beta}_0 + \hat{\beta}_1 x)^2] = Var[\hat{\beta}_0 + \hat{\beta}_1 x] + E^2[\hat{\beta}_0 + \hat{\beta}_1 x]$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right] + (\beta_0 + \beta_1 x)^2. \tag{3.14}$$

Nesta equação, como no caso anterior, no qual usamos transformação logarítmica, podemos ver que, quando o valor de n é grande, a quantidade $\left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right]$ é aproximadamente zero, portanto, a esperança dada em (3.14) é aproximadamente o valor da mediana dada em (3.13), ou seja,

$$E[(\hat{\beta}_0 + \hat{\beta}_1 x)^2] \simeq (\beta_0 + \beta_1 x)^2.$$

Miller estima a esperança da variável resposta na escala original, para todo valor de n, da seguinte maneira:

$$\hat{E}[y \mid x] = \hat{\sigma}^2 + (\hat{\beta}_0 + \hat{\beta}_1 x)^2. \tag{3.15}$$

Agora, quando n não é muito pequeno, temos a condição (3.9), logo,

$$\sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right] \leq \sigma^2.$$

Portanto, a esperança dada em (3.14), em geral, vai ser menor que a esperança da variável resposta na escala original, dada em (3.12). Para este caso nós propomos o estimador, não viciado, para a esperança de y,

$$\hat{E}[y \mid x] = \hat{\sigma}^2 \left[1 - \frac{1}{n} - \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_{i1} - \bar{x})^2} \right] + (\hat{\beta}_0 + \hat{\beta}_1 x)^2. \tag{3.16}$$

3.2.3 Transformação inversa

Consideremos o modelo

$$\frac{1}{y_i} = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \tag{3.17}$$

onde

$$y_i > 0,$$

$$E[\epsilon_i] = 0,$$

$$Var[\epsilon_i] = \sigma^2,$$

$$E[\epsilon_i \epsilon_j] = 0, \quad i \neq j, \quad i, j = 1, 2, ..., n.$$

Resultado 3.1

Seja w variável aleatória com $E[w] \neq 0$ (finita). Então

$$\frac{1}{w} = \frac{1}{E[w]} \left(1 + \frac{w - E[w]}{E[w]} \right)^{-1}.$$

Prova:

$$\frac{1}{E[w]} \left(1 + \frac{w - E[w]}{E[w]} \right)^{-1} = \frac{1}{E[w]} \frac{1}{\left(1 + \frac{w - E[w]}{E[w]} \right)}$$

$$= \frac{1}{E[w]} \frac{E[w]}{(E[w] + w - E[w])}$$

$$= \frac{1}{E[w]} \frac{E[w]}{w}$$

$$= \frac{1}{w}. \quad \square$$

Usando a aproximação

$$\frac{1}{1-a} \simeq 1 + a + a^2, \quad a \in (-1, 1),$$

para $a = \frac{E[w]-w}{E[w]}$ temos

$$\frac{1}{\left(1 - \frac{E[w]-w}{E[w]}\right)} = \left(1 + \frac{w - E[w]}{E[w]}\right)^{-1} \simeq 1 + \frac{E[w] - w}{E[w]} + \left(\frac{E[w] - w}{E[w]}\right)^2$$

e, usando-a no termo direito do resultado (3.1), obtemos

$$\frac{1}{w} \simeq \frac{1}{E[w]} \left(1 + \frac{E[w] - w}{E[w]} + \left(\frac{E[w] - w}{E[w]}\right)^2\right). \tag{3.18}$$

Observa-se que se w tem distribuição normal, com esperança $E[w] \neq 0$ e variância $Var[w]$,

$$\frac{w - E[w]}{E[w]} \sim N\left(0, \frac{Var[w]}{(E[w])^2}\right).$$

Logo, se $\frac{Var[w]}{(E[w])^2}$ é um valor pequeno, por exemplo, $\sqrt{\frac{Var[w]}{(E[w])^2}} < \frac{1}{2}$, ou seja, se o desvio padrão de w é bem menor do que a metade de sua esperança, então

$$\Pr\left[\left|\frac{w - E[w]}{E[w]}\right| < 1\right] \simeq 1.$$

Agora, se $w_i = \frac{1}{y_i}$, no modelo (3.17), teremos

$$w_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \quad i = 1, 2, ..., n.$$

Usando a equação (3.18), substituindo w pela expressão de w_i, na linha acima, temos a seguinte aproximação

$$y_i = \frac{1}{\beta_0 + \beta_1 x_{i1} + \epsilon_i} \simeq \frac{1}{\beta_0 + \beta_1 x_{i1}} \left(1 + \frac{-\epsilon_i}{\beta_0 + \beta_1 x_{i1}} + \left(\frac{-\epsilon_i}{\beta_0 + \beta_1 x_{i1}}\right)^2\right).$$

(Note que essa aproximação requer $\sigma < \frac{\beta_0 + \beta_1 x_{i1}}{2}$.)

Calculando a esperança obtemos

$$E[y_i \mid x_{i1}] \simeq \frac{1}{\beta_0 + \beta_1 x_{i1}} \left(1 + \frac{\sigma^2}{(\beta_0 + \beta_1 x_{i1})^2}\right), \tag{3.19}$$

e, pela proposição 1 do apêndice, a mediana é

$$Mediana[y_i \mid x_{i1}] = \frac{1}{\beta_0 + \beta_1 x_{i1}}. \tag{3.20}$$

Devemos notar que, se o valor de σ^2, divido por $(\beta_0 + \beta_1 x_{i1})^3$, é pequeno em relação à magnitude da variável y, a esperança e a mediana são aproximadamente iguais. Conforme comentário acima, deveríamos ter $\frac{\sigma^2}{(\beta_0 + \beta_1 x_{i1})^2} < \frac{1}{4}$, para que a aproximação usada seja adequada.

Agora, trabalharemos com o modelo ajustado de (3.17),

$$\hat{E}\left[\frac{1}{y}\right] = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Substituindo w por $\hat{\beta}_0 + \hat{\beta}_1 x$ na equação (3.18), temos

$$\frac{1}{\hat{\beta}_0 + \hat{\beta}_1 x} \simeq \frac{1}{\beta_0 + \beta_1 x} \left(1 + \frac{\hat{\epsilon} - \epsilon}{\beta_0 + \beta_1 x} + \left(\frac{\hat{\epsilon} - \epsilon}{\beta_0 + \beta_1 x}\right)^2\right), \tag{3.21}$$

onde $\hat{\epsilon}$ é o resíduo do modelo (3.17), ou seja, $\hat{\epsilon} = (\beta_0 + \beta_1 x + \epsilon) - (\hat{\beta}_0 + \hat{\beta}_1 x)$. Calculamos

a esperança de $\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}$, aproximada,

$$E\left[\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\right] \simeq \frac{1}{\beta_0+\beta_1 x}\left(1+\sigma^2\frac{\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^n(x_{i1}-\bar{x})^2}\right)}{(\beta_0+\beta_1 x)^2}\right). \tag{3.22}$$

Note que para que a aproximação (3.21) seja razoável, é necessário que

$$\left(Var[\hat{\beta}_0+\hat{\beta}_1 x]\right)^{\frac{1}{2}} = \sigma\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^n(x_{i1}-\bar{x})^2}\right)^{\frac{1}{2}} < \frac{\beta_0+\beta_1 x}{2} = \frac{1}{2}E[\hat{\beta}_0+\hat{\beta}_1 x],$$

ou seja,

$$\sigma^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^n(x_{i1}-\bar{x})^2}\right) < \frac{(\beta_0+\beta_1 x)^2}{4}.$$

Assim, o lado direito da expressão (3.22) deve ser menor que $(1+\frac{1}{4})$. Além disso, em (3.22) notamos que, quando n é grande, a quantidade $\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^n(x_{i1}-\bar{x})^2}\right)$ é aproximadamente zero, sendo a aproximação da esperança dada em (3.22) e a mediana dada em (3.20) quase iguais. Logo, Miller estima a esperança de y por

$$\hat{E}[y \mid x] = \frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\left(1+\frac{\hat{\sigma}^2}{\left(\hat{\beta}_0+\hat{\beta}_1 x\right)^2}\right), \tag{3.23}$$

para qualquer valor de n.

Mas, se o valor de n não é muito grande, temos a condição (3.9), então

$$\sigma^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^n(x_{i1}-\bar{x})^2}\right) \leq \sigma^2.$$

Logo, a esperança dada em (3.22), em geral, é menor que o valor dado em (3.19) se $(\beta_0+\beta_1 x)^2 \geq 1$, e maior que o valor dado em (3.19) se $(\beta_0+\beta_1 x)^2 \leq 1$. Neste caso

propomos o estimador

$$\hat{E}[y \mid x] = \frac{1}{\hat{\beta}_0 + \hat{\beta}_1 x} \left\{ \frac{\left(1 + \frac{\hat{\sigma}^2}{\left(\hat{\beta}_0 + \hat{\beta}_1 x\right)^2}\right)}{\left(1 + \hat{\sigma}^2 \frac{\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right)}{\left(\hat{\beta}_0 + \hat{\beta}_1 x\right)^2}\right)} \right\}. \tag{3.24}$$

3.2.4 Estimadores para o caso de regressão linear múltipla

Seja $x_{\underset{\sim}{0}}^T = (1, x_{01}, ..., x_{0p})$ um vetor onde x_{0j} é um valor do intervalo $[x_{j(1)}, x_{j(n)}]$, com $x_{j(1)}$ e $x_{j(n)}$ os valores mínimo e máximo, respectivamente, da variável regressora x_j, para $j = 1, 2, ..., p$,.

Definimos

$$r_i = x_{\underset{\sim}{i}}^T (\mathbf{X}^T\mathbf{X})^{-1} x_{\underset{\sim}{i}}, \quad i = 0, 1, ..., n.$$

Podemos notar que, se $p = 1$, então,

$$r_0 = \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n}(x_{i1} - \bar{x})^2}\right],$$

onde $x = x_{01}$.

A tabela (3.1) mostra os estimadores para y no caso de modelo de regressão linear múltipla, cujo cálculo foi desenvolvido neste trabalho seguindo o mesmo procedimento usado por Miller no caso de regressão linear simples.

Tabela 3.1: Estimadores de regressão múltipla

Transformacões	*Estimadores Miller*	*Estimadores propostos*
Logaritmo	$\exp(\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}})\exp(\frac{\hat{\sigma}^2}{2})$	$\exp(\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}})\exp(\frac{\hat{\sigma}^2}{2}(1-r_0))$
Raiz quadrada	$\hat{\sigma}^2+(\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}})^2$	$\hat{\sigma}^2(1-r_0)+(\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}})^2$
Inversa	$\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\left(1+\frac{\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)$	$\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\left\{\frac{\left(1+\frac{\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)}{\left(1+\frac{r_0\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)}\right\}$

3.3 Estimador Não Paramétrico da Esperança de y

Duan (1983) propos um estimador não paramétrico para a esperança de y, a seguir apresentado.

Sejam g e h funções reais monótonas, continuamente diferenciáveis, onde h é a função inversa de g.

Sejam η_i as observações transformadas da variável resposta, y_i, tal que

$$\eta_i = g(y_i), \quad i = 1, 2, ..., n,$$

portanto,

$$y_i = h(\eta_i), \ (h = g^{-1}). \tag{3.25}$$

Consideremos o modelo (3.1), tomando as observações transformadas da variável resposta, ao invés das observações na escala original,

$$\eta_i = \underset{\sim}{x}_i^T\underset{\sim}{\beta} + \epsilon_i, \ i = 1, 2, ..., n, \tag{3.26}$$

onde

$$\epsilon_i \overset{iid}{\sim} F, \ E[\epsilon_i] = 0, \ Var[\epsilon_i] = \sigma^2,$$

notando que não assumimos que a função de distribuição F seja normal, nem simetria na função de densidade.

Estimando $\underset{\sim}{\beta}$, do modelo (3.26), por quadrados mínimos obtemos

$$\underset{\sim}{\hat{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\underset{\sim}{\eta},$$

onde $\mathbf{X}^T = (\underset{\sim}{x}_1, \underset{\sim}{x}_2, ..., \underset{\sim}{x}_n)$ é a matriz de modelo, de posto coluna completo e $\underset{\sim}{\eta}^T = (\eta_1, \eta_2, ..., \eta_n)$ é o vetor das observações transformadas da variável resposta.

Por (3.25) e (3.26) temos

$$E[y_0 \mid \underset{\sim}{x}_0] = E[h(\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \epsilon_0)] = \int h(\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \omega)\partial F(\omega),$$

onde a integral usada é a integral de Riemann-Stieltjes. Como a função de distribuição dos erros, F, não é conhecida, é estimada por

$$\hat{F}_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} I\{\hat{\epsilon}_i \leq \omega\}, \quad \omega \in R,$$

onde $\hat{\epsilon}_i = \eta_i - \underset{\sim}{x}_i^T\underset{\sim}{\hat{\beta}}$, $i = 1, 2, ..., n$, são os resíduos de quadrados mínimos ordinários do modelo (3.26) e $I\{.\}$ é a função indicadora do evento '$\{.\}$'. Logo, usando a função de distribuição estimada dos erros, $\hat{F}_n$, temos

$$E[y_0 \mid \underset{\sim}{x}_0] = \int h(\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \omega)\partial F(\omega) \simeq \int h(\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \omega)\partial \hat{F}_n(\omega)$$

$$= \frac{1}{n}\sum_{i=1}^{n} h(\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \hat{\epsilon}_i).$$

Duan denota por estimador 'smearing' a estatística

$$\hat{Y}_{0s} = \frac{1}{n}\sum_{i=1}^{n} h(\underset{\sim}{x}_0^T\underset{\sim}{\hat{\beta}} + \hat{\epsilon}_i), \tag{3.27}$$

considerando-o como um estimador da esperança da variável resposta, dado $\underset{\sim}{x}_0$, na escala

original,

$$\hat{E}[y_0 \mid \underset{\sim}{x}_0] = \hat{Y}_{0s}.$$

A seguir, vamos definir o estimador 'smearing' quando a transformação da variável resposta é da família de transformações de potência de Box-Cox.

Consideremos o modelo

$$y_i^{(\lambda)} = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \epsilon_i,$$

onde

$$\epsilon_i \overset{iid}{\sim} N(0, \sigma^2), \qquad i = 1, 2, ..., n,$$

para um valor de $\lambda \in R$, onde,

$$g(y_i) = y_i^{(\lambda)} = \begin{cases} (y_i^\lambda - 1)/\lambda, \lambda \neq 0 \\ \ln(y_i), \qquad \lambda = 0 \end{cases}.$$

Então, da equação (3.27)

$$h(\underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \hat{\epsilon}_i) = \begin{cases} (1 + \lambda \, \underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \lambda \hat{\epsilon}_i)^{\frac{1}{\lambda}}, & \lambda \neq 0 \\ \exp(\underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \hat{\epsilon}_i), & \lambda = 0 \end{cases}, \tag{3.28}$$

onde os resíduos $\hat{\epsilon}_i$, dados pela equação $\hat{\epsilon}_i = \eta_i - \hat{\eta}_i$, são

$$\hat{\epsilon}_i = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} - \underset{\sim}{x}_i^T \, \hat{\beta}, & \lambda \neq 0 \\ \ln(y_i) - \underset{\sim}{x}_i^T \, \hat{\beta}, & \lambda = 0 \end{cases},$$

e substituindo-os na equação (3.28) temos

$$h(\underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \hat{\epsilon}_i) = \begin{cases} (1 + \lambda \, \underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \lambda(\frac{y_i^\lambda - 1}{\lambda} - \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}))^{\frac{1}{\lambda}}, & \lambda \neq 0 \\ \exp(\underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \ln(y_i) - \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}), & \lambda = 0 \end{cases}$$

$$= \begin{cases} (y_i^\lambda + \lambda(x_0^T - x_i^T)\,\hat{\beta})^{\frac{1}{\lambda}}, & \lambda \neq 0 \\ y_i \exp\left\{(x_0^T - x_i^T)\,\hat{\beta}\right\}, & \lambda = 0 \end{cases}.$$

Logo, o estimador 'smearing' é

$$\hat{Y}_{0s} = \begin{cases} \frac{1}{n}\sum_{i=1}^n (y_i^\lambda + \lambda(x_0^T - x_i^T)\,\hat{\beta})^{\frac{1}{\lambda}}, & \lambda \neq 0 \\ \frac{1}{n}\sum_{i=1}^n y_i \exp\left\{(x_0^T - x_i^T)\,\hat{\beta}\right\}, & \lambda = 0 \end{cases}.$$

Duan observa que o estimador 'smearing' é consistente. No caso da distribuição dos erros normal, os estimadores baseados em normalidade também são consistentes, e além disso podem ser mais eficientes do que o estimador 'smearing'.

3.4 Estimador de Aproximação da Esperança de y

Taylor (1986) apresentou um estimador da esperança da variável resposta, na escala original, baseado num método de aproximação.

Para um elemento da família de transformações de Box-Cox dada em (3.3), temos o modelo

$$y_i^{(\lambda)} = x_i^T \beta + \epsilon_i, \quad i = 1, 2, ..., n, \tag{3.29}$$

onde

$$y_i^{(\lambda)} = \begin{cases} (y_i^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \ln(y_i), & \lambda = 0 \end{cases},$$

$$E[\epsilon_i] = 0, Var[\epsilon_i] = \sigma^2,$$

com função de densidade simétrica, não necessariamente normal.

Definindo

$$e_i = \frac{\epsilon_i}{\sigma}, \tag{3.30}$$

temos

$$E[e_i] = 0, Var[e_i] = 1,$$

e seja F o modelo de probabilidade de e_i. Substituindo (3.30) no modelo (3.29), temos

$$y_i^{(\lambda)} = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sigma e_i, \quad i = 1, 2, ..., n.$$

Inicialmente consideremos o caso $\lambda \neq 0$, onde

$$\frac{y_i^\lambda - 1}{\lambda} = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sigma e_i \Rightarrow y_i = (1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \lambda \sigma e_i)^{\frac{1}{\lambda}}.$$

A esperança de y_i, dado $\underset{\sim}{x}_i$, é

$$\begin{aligned} E[y_i \mid \underset{\sim}{x}_i] &= \int (1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \lambda \sigma e_i)^{\frac{1}{\lambda}} \partial F(e_i) \\ &= \int (1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta})^{\frac{1}{\lambda}} (1 + \frac{\lambda \sigma e_i}{1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta}})^{\frac{1}{\lambda}} \partial F(e_i) \\ &= \int (1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta})^{\frac{1}{\lambda}} (1 + \theta(\underset{\sim}{x}_i) e_i)^{\frac{1}{\lambda}} \partial F(e_i), \end{aligned}$$

onde

$$\theta(\underset{\sim}{x}_i) = \frac{\lambda \sigma}{1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta}}. \tag{3.31}$$

Assim,

$$\begin{aligned} E[y_i \mid \underset{\sim}{x}_i] &= (1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta})^{\frac{1}{\lambda}} \int (1 + \theta(\underset{\sim}{x}_i) e_i)^{\frac{1}{\lambda}} \partial F(e_i) \\ &= (1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\beta})^{\frac{1}{\lambda}} I_1, \end{aligned} \tag{3.32}$$

onde

$$I_1 = \int (1 + \theta(\underset{\sim}{x}_i) e_i)^{\frac{1}{\lambda}} \partial F(e_i). \tag{3.33}$$

Agora, vamos trabalhar só com o termo I_1. Pela expansão em série de Taylor da função $f(x) = (1 + x)^{\frac{1}{\lambda}}$ em torno de zero, temos que

$$(1+\theta(\underset{\sim}{x_i})e_i)^{\frac{1}{\lambda}} = 1+\sum_{k=1}^{\infty}\frac{(\theta(\underset{\sim}{x_i})e_i)^k}{k!}\Pi_{j=0}^{k-1}\left(\frac{1}{\lambda}-j\right),$$

para $\theta(\underset{\sim}{x_i})e_i$ próximo de zero. Substituindo em (3.33) temos

$$I_1 = \int\left\{1+\sum_{k=1}^{\infty}\frac{(\theta(\underset{\sim}{x_i})e_i)^k}{k!}\Pi_{j=0}^{k-1}\left(\frac{1}{\lambda}-j\right)\right\}\partial F(e_i).$$

Colocando a restrição [1] $\theta(\underset{\sim}{x_i}) \ll 1$, ou seja, $\theta(\underset{\sim}{x_i})$ bem menor do que um, consideramos desprezíveis os termos $(\theta(\underset{\sim}{x_i}))^j$, para j maior que 2, logo obtemos

$$I_1 \simeq \int\left\{1+\frac{1}{\lambda}\theta(\underset{\sim}{x_i})e_i+\frac{1}{2}(\theta(\underset{\sim}{x_i})e_i)^2\frac{1}{\lambda}\left(\frac{1}{\lambda}-1\right)\right\}\partial F(e_i)$$

$$= \int 1\partial F(e_i)+\int\frac{1}{\lambda}\theta(\underset{\sim}{x_i})e_i\partial F(e_i)$$

$$+\int\frac{1}{2}(\theta(\underset{\sim}{x_i})e_i)^2\frac{1}{\lambda}\left(\frac{1}{\lambda}-1\right)\partial F(e_i)$$

$$= 1+\frac{1}{\lambda}\theta(\underset{\sim}{x_i})E[e_i]+\frac{1}{2}(\theta(\underset{\sim}{x_i}))^2\frac{1}{\lambda}\left(\frac{1}{\lambda}-1\right)E[e_i^2]$$

$$= 1+\frac{1}{2}(\theta(\underset{\sim}{x_i}))^2\frac{1}{\lambda}\left(\frac{1}{\lambda}-1\right),$$

porque $E[e_i]=0$ e $E[e_i^2]=1$. Substituindo (3.31) na equação anterior,

$$I_1 \simeq 1+\frac{\sigma^2(1-\lambda)}{2(1+\lambda\,\underset{\sim}{x_i}^T\underset{\sim}{\beta})^2}.$$

Substituindo esta aproximação de I_1 na equação (3.32), tem-se um valor aproximado da esperança de y_i, dado $\underset{\sim}{x_i}$, o qual é

$$E[y_i\,|\underset{\sim}{x_i}] \simeq Y_a = (1+\lambda\,\underset{\sim}{x_i}^T\underset{\sim}{\beta})^{\frac{1}{\lambda}}\left\{1+\frac{\sigma^2(1-\lambda)}{2(1+\lambda\,\underset{\sim}{x_i}^T\underset{\sim}{\beta})^2}\right\}.$$

[1] Esta condição é denotada na literatura por θ-pequeno ('small-θ').

Logo, um estimador da esperança de y, dado $\underset{\sim}{x}$, é

$$\hat{E}[y\,|\underset{\sim}{x}] = \hat{Y}_a = (1+\lambda\,\underset{\sim}{x}^T\,\underset{\sim}{\hat{\beta}})^{\frac{1}{\lambda}}\left\{1+\frac{\hat{\sigma}^2(1-\lambda)}{2(1+\lambda\,\underset{\sim}{x}^T\,\underset{\sim}{\hat{\beta}})^2}\right\}, \tag{3.34}$$

onde $\underset{\sim}{\hat{\beta}}$ e $\hat{\sigma}^2$ são estimadores de máxima verossimilhança.

Analisando o caso $\lambda = 0$, temos

$$\ln(y_i) = \underset{\sim}{x}_i^T\,\underset{\sim}{\beta} + \sigma e_i \Rightarrow y_i = \exp(\underset{\sim}{x}_i^T\,\underset{\sim}{\beta} + \sigma e_i).$$

A esperança de y_i é

$$\begin{aligned} E[y_i\,|\underset{\sim}{x}_i] &= \int \exp(\underset{\sim}{x}_i^T\,\underset{\sim}{\beta} + \sigma e_i)\partial F(e_i) \\ &= \exp(\underset{\sim}{x}_i^T\,\underset{\sim}{\beta})\int \exp(\sigma e_i)\partial F(e_i) \\ &= \exp(\underset{\sim}{x}_i^T\,\underset{\sim}{\beta})I_2, \end{aligned} \tag{3.35}$$

onde

$$I_2 = \int \exp(\sigma e_i)\partial F(e_i). \tag{3.36}$$

Neste caso, podemos obter o cálculo exato e o cálculo aproximado da esperança condicional de y_i, se supormos que a função de distribuição F é normal. Vimos, na seção 3.2, que, se e_i tem distribuição $N(0,1)$, então $\exp(\sigma e_i) \sim LN(0,1,\sigma)$, logo

$$E[y_i\,|\underset{\sim}{x}_i] = \exp\left(\underset{\sim}{x}_i^T\,\underset{\sim}{\beta} + \frac{\sigma^2}{2}\right). \tag{3.37}$$

Fazendo o cálculo aproximado do termo I_2, usando a expansão de Taylor até segundo grau para a função exponencial, em torno de zero, obtemos

$$\exp(\sigma e_i) \simeq 1 + \sigma e_i + \frac{\sigma^2}{2}e_i^2, \tag{3.38}$$

para σe_i próximo de zero. Substituindo em (3.36),

$$I_2 \simeq \int \left\{1 + \sigma e_i + \frac{\sigma^2}{2} e_i^2\right\} \partial F(e_i)$$

$$= \int 1 \partial F(e_i) + \int \sigma e_i \partial F(e_i) + \int \frac{\sigma^2}{2} e_i^2 \partial F(e_i)$$

$$= 1 + \frac{\sigma^2}{2}.$$

Então, um valor aproximado da esperança de y_i, dado $\underset{\sim}{x}_i$, é dado pela substituição do valor aproximado de I_2 (acima) na expressão (3.35),

$$E[y_i \,|\underset{\sim}{x}_i] \simeq Y_a = \exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta}) \left(1 + \frac{\sigma^2}{2}\right).$$

Logo, um estimador da esperança condicional de y, dado $\underset{\sim}{x}$, é

$$\hat{E}[y \,|\underset{\sim}{x}] = \hat{Y}_a = \exp(\underset{\sim}{x}^T \underset{\sim}{\hat{\beta}}) \left(1 + \frac{\hat{\sigma}^2}{2}\right), \tag{3.39}$$

onde $\underset{\sim}{\hat{\beta}}$ e $\hat{\sigma}^2$ são estimadores de máxima verossimilhança.

Comparando (3.37) e (3.39), observa-se que, se σ^2 não é pequeno, o estimador da esperança de y sob normalidade é muito diferente do estimador da aproximação da esperança. Assim, Taylor considera que a aproximação da esperança de y não é boa quando σ^2 é um valor grande e λ é próximo de zero. Observamos que a aproximação em (3.38) já requer valor pequeno para σ.

Voltando ao caso $\lambda \neq 0$, Taylor sugere interpretar Y_a como a mediana de y, $(1 + \lambda \underset{\sim}{x}^T \underset{\sim}{\beta})^{\frac{1}{\lambda}}$, multiplicada por um fator de correção, $\left\{1 + \frac{\sigma^2(1-\lambda)}{2(1+\lambda \underset{\sim}{x}^T \underset{\sim}{\beta})^2}\right\}$. Se $\lambda < 1$, então o fator de correção é maior que um. Ele também destaca que, se os parâmetros $(\underset{\sim}{\beta}, \sigma)$ são estimados consistentemente, então, $\hat{Y}_a$ converge em probabilidade a Y_a, quando n tende a ∞.

Resultado 3.2

O estimador do valor aproximado da esperança, $\hat{Y}_a$, e o estimador 'smearing', da seção anterior, $\hat{Y}_s$, são aproximadamente iguais, exceto quando λ está próximo de zero.

Prova:

Para o caso $\lambda \neq 0$

$$\hat{Y}_s = \frac{1}{n}\sum_{i=1}^{n}(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}} + \lambda\hat{e}_i)^{\frac{1}{\lambda}}.$$

Substituindo $\hat{\epsilon}_i = \hat{\sigma}\hat{e}_i$ para $i = 1, 2, ..., n$, temos

$$\hat{Y}_s = \frac{1}{n}\sum_{i=1}^{n}(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}} + \lambda\hat{\sigma}\hat{e}_i)^{\frac{1}{\lambda}}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}})^{\frac{1}{\lambda}} \left(1 + \frac{\lambda\hat{\sigma}\hat{e}_i}{(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}})}\right)^{\frac{1}{\lambda}}$$

$$= \frac{1}{n}(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}})^{\frac{1}{\lambda}} \sum_{i=1}^{n} \left(1 + \frac{\lambda\hat{\sigma}\hat{e}_i}{(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}})}\right)^{\frac{1}{\lambda}}$$

$$= \frac{1}{n}(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}})^{\frac{1}{\lambda}} \sum_{i=1}^{n}(1 + \hat{\theta}(\underset{\sim}{x})\hat{e}_i)^{\frac{1}{\lambda}},$$

onde $\hat{\theta}(\underset{\sim}{x}) = \frac{\lambda\hat{\sigma}}{1+\lambda \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}}}$, que assumimos como um valor pequeno. Logo,

$$\hat{Y}_s = \frac{1}{n}(1 + \lambda\, \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}})^{\frac{1}{\lambda}} \sum_{i=1}^{n} \left\{1 + \frac{\hat{\theta}(\underset{\sim}{x})\hat{e}_i}{\lambda} + \frac{(\hat{\theta}(\underset{\sim}{x})\hat{e}_i)^2}{2}\frac{1}{\lambda}\left(\frac{1}{\lambda} - 1\right)\right.$$

$$\left. + O((\hat{\theta}(\underset{\sim}{x}))^3)\right\}.$$

Como a primeira coluna da matriz de modelo é uma coluna de uns, temos que $\sum_{i=1}^{n} \hat{e}_i = 0$

e $\sum_{i=1}^{n} \hat{e}_i^2 = n$. Portanto,

$$\hat{Y}_s \simeq (1 + \lambda \underset{\sim}{x}^T \underset{\sim}{\hat{\beta}})^{\frac{1}{\lambda}} \left(1 + \frac{(\hat{\theta}(\underset{\sim}{x}))^2}{2} \frac{1}{\lambda}\left(\frac{1}{\lambda} - 1\right)\right) = \hat{Y}_a. \quad \square$$

3.5 Estimador de Aproximação da Esperança de y Aplicado a Modelo Linear Misto Balanceado

Sakia (1990) generalizou o procedimento proposto por Taylor (1986), para estimar a esperança de y, ao modelo linear misto e estudou as diferenças entre os estimadores propostos da média na escala original e os estimadores que resultam da aplicação direta da função inversa da transformação à esperança estimada no modelo transformado.

Um modelo linear misto balanceado, usando uma transformação de Box-Cox na variável resposta, é

$$y_i^{(\lambda)} = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sum_{l=1}^{q} \zeta_l, \quad i = 1, 2, ..., n, \tag{3.40}$$

onde

$$y_i^{(\lambda)} = \begin{cases} (y_i^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ \ln(y_i), & \lambda = 0 \end{cases},$$

$$\zeta_l \sim N(0, \sigma_l^2), \quad l = 1, 2, ..., q,$$

onde os ζ_l's são variáveis independentes, notando-se que a diferença deste modelo para o modelo (3.3) está no termo de erro. Fazendo $e_l = \frac{\zeta_l}{\sigma_l}$, obtemos

$$e_l \sim N(0, 1), \quad l = 1, 2, ..., q,$$

e o modelo dado em (3.40) pode ser reescrito como

$$y_i^{(\lambda)} = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sum_{l=1}^{q} \sigma_l e_l, \quad i = 1, 2, ..., n; \quad l = 1, 2, ..., q. \tag{3.41}$$

A matriz **X** deste modelo não é composta por valores de variáveis regressoras, mas por

variáveis indicadoras.

Para o caso $\lambda \neq 0$,

$$\frac{y_i^{\lambda}-1}{\lambda} = \underset{\sim}{x_i^T}\underset{\sim}{\beta} + \sum_{l=1}^{q}\sigma_l e_l \Rightarrow y_i = (1+\lambda\,\underset{\sim}{x_i^T}\underset{\sim}{\beta} + \lambda\sum_{l=1}^{q}\sigma_l e_l)^{\frac{1}{\lambda}},$$

portanto, a esperança de y_i, dado $\underset{\sim}{x_i}$, é

$$E[y_i \mid \underset{\sim}{x_i}] = \int (1+\lambda\,\underset{\sim}{x_i^T}\underset{\sim}{\beta} + \lambda\sum_{l=1}^{q}\sigma_l e_l)^{\frac{1}{\lambda}}\partial F(e_1, e_2, ..., e_q)$$

$$= \int (1+\lambda\,\underset{\sim}{x_i^T}\underset{\sim}{\beta})^{\frac{1}{\lambda}}(1+\sum_{l=1}^{q}c_l e_l)^{\frac{1}{\lambda}}\partial F(e_1, e_2, ..., e_q),$$

onde F é a função de distribuição acumulada da normal padrão q-dimensional e $c_l = \frac{\lambda\sigma_l}{1+\lambda\sum_{i=1}^{n}\underset{\sim}{x_i^T}\underset{\sim}{\beta}}$.
Esta esperança pode ser expressa como

$$E[y_i \mid \underset{\sim}{x_i}] = (1+\lambda\,\underset{\sim}{x_i^T}\underset{\sim}{\beta})^{\frac{1}{\lambda}} I_1, \tag{3.42}$$

onde $I_1 = \int(1+\sum_{l=1}^{q}c_l e_l)^{\frac{1}{\lambda}}\partial F(e_1, e_2, ..., e_q)$. Supondo $\mid c_l \mid$ um valor bem pequeno, $\mid c_l \mid \ll 1$, podemos aproximar o integrando de I_1

$$(1+\sum_{l=1}^{q}c_l e_l)^{\frac{1}{\lambda}} \simeq 1 + \frac{\sum_{l=1}^{q}c_l e_l}{\lambda} + (1-\lambda)\frac{(\sum_{l=1}^{q}c_l e_l)^2}{2\lambda^2}$$

e, substituindo em I_1, temos

$$I_1 \simeq \int \left\{1 + \frac{\sum_{l=1}^{q}c_l e_l}{\lambda} + (1-\lambda)\frac{(\sum_{l=1}^{q}c_l e_l)^2}{2\lambda^2}\right\}\partial F(e_1, e_2, ..., e_q).$$

Como os erros são independentes,

$$I_1 \simeq \int \cdots \int \left\{1 + \frac{\sum_{l=1}^{q}c_l e_l}{\lambda} + (1-\lambda)\frac{(\sum_{l=1}^{q}c_l e_l)^2}{2\lambda^2}\right\}\partial F(e_1)...\partial F(e_q)$$

$$= \int \cdots \int \partial F(e_1)...\partial F(e_q) + \int \cdots \int \frac{\sum_{l=1}^{q} c_l e_l}{\lambda} \partial F(e_1)...\partial F(e_q)$$

$$+ \int \cdots \int (1-\lambda) \frac{(\sum_{l=1}^{q} c_l e_l)^2}{2\lambda^2} \partial F(e_1)...\partial F(e_q)$$

$$= 1 + E\left(\frac{\sum_{l=1}^{q} c_l e_l}{\lambda}\right) + \frac{(1-\lambda)}{2\lambda^2} Var[(\sum_{l=1}^{q} c_l e_l)^2]$$

$$= 1 + \frac{(1-\lambda)}{2\lambda^2} \sum_{l=1}^{q} c_l^2 Var[e_l]$$

pois $E[\sum_{l=1}^{q} c_l e_l] = 0$,

$$= 1 + \frac{(1-\lambda)}{2\lambda^2} \sum_{l=1}^{q} c_l^2.$$

Substituindo $c_l = \frac{\lambda \sigma_l}{1+\lambda \sum_{i=1}^{n} x_i^T \beta}$ na equação acima, temos

$$I_1 \simeq 1 + \frac{(1-\lambda)}{2\lambda^2} \sum_{l=1}^{q} \left(\frac{\lambda \sigma_l}{1 + \lambda \sum_{i=1}^{n} x_i^T \beta} \right)^2$$

$$= 1 + \frac{(1-\lambda)}{2} \sum_{l=1}^{q} \frac{\sigma_l^2}{(1 + \lambda \sum_{i=1}^{n} x_i^T \beta)^2},$$

então, um valor aproximado da esperança de y_i, dado x_i, é

$$E[y_i \mid x_i] \simeq Y_a = (1 + \lambda\, x_i^T \beta)^{\frac{1}{\lambda}} \left\{ 1 + \frac{(1-\lambda)}{2} \sum_{l=1}^{q} \frac{\sigma_l^2}{(1 + \lambda \sum_{i=1}^{n} x_i^T \beta)^2} \right\},$$

substituindo a aproximação de I_1 na equação (3.42). Logo, um estimador da esperança de y, dado x, é

$$\hat{E}[y \mid x] = \hat{Y}_a = (1 + \lambda\, x^T \hat{\beta})^{\frac{1}{\lambda}} \left\{ 1 + \frac{(1-\lambda)}{2} \sum_{l=1}^{q} \frac{\hat{\sigma}_l^2}{(1 + \lambda \sum_{i=1}^{n} x^T \hat{\beta})^2} \right\}, \qquad (3.43)$$

onde $\hat{\sigma}_l^2$ é estimador de máxima verossimilhança de σ_l^2.

Para o caso $\lambda = 0$,

$$\ln(y_i) = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sum_{l=1}^{q} \sigma_l e_l \Rightarrow y_i = \exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sum_{l=1}^{q} \sigma_l e_l)$$

e a esperança de y_i, dado $\underset{\sim}{x}_i$, é

$$E[y_i \mid \underset{\sim}{x}_i] = E[\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sum_{l=1}^{q} \sigma_l e_l)]$$

$$= \exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta}) E[\exp(\sum_{l=1}^{q} \sigma_l e_l)].$$

Como a função de distribuição dos erros F é normal, então

$$\exp(\sigma_l e_l) \sim LN(0, \sigma_l^2), \quad l = 1, 2, ..., q,$$

logo,

$$E[y_i \mid \underset{\sim}{x}_i] = \exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta}) \exp(\frac{1}{2} \sum_{l=1}^{q} \sigma_l^2).$$

Portanto, um estimador da esperança de y, dado $\underset{\sim}{x}$, é

$$\hat{E}[y \mid \underset{\sim}{x}] = \exp(\underset{\sim}{x}^T \underset{\sim}{\hat{\beta}}) \exp(\frac{1}{2} \sum_{l=1}^{q} \hat{\sigma}_l^2). \tag{3.44}$$

Sakia analisou a função inversa da transformação à esperança estimada no modelo transformado. Isto porque esta prática é comum. Ele destaca a diferença entre o seu estimador proposto e esta simples aplicação da inversa da transformação. Os passos da análise são apresentados a seguir.

Para o caso $\lambda \neq 0$,

$$\frac{y_i^\lambda - 1}{\lambda} = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \sum_{l=1}^{q} \sigma_l e_l, \quad i = 1, 2, ..., n.$$

Portanto,

$$E\left[\frac{y_i^{\lambda}-1}{\lambda}\,|\underset{\sim}{x}_i\right] = E\left[\underset{\sim}{x}_i^T\underset{\sim}{\beta} + \sum_{l=1}^{q}\sigma_l e_l\right]$$

$$= \underset{\sim}{x}_i^T\underset{\sim}{\beta},$$

pois $e_l \sim N(0,1)$. Logo,

$$E[y_i^{\lambda}\,|\underset{\sim}{x}_i] = 1 + \lambda\,\underset{\sim}{x}_i^T\underset{\sim}{\beta},$$

estimada por $1+\lambda\,\underset{\sim}{x}_i^T\underset{\sim}{\hat{\beta}}$, onde $\underset{\sim}{\hat{\beta}}$ é o estimador de máxima verossimilhança de $\underset{\sim}{\beta}$. Aplicando a função inversa da transformação da variável resposta à esperança estimada de y_i^{λ}, dado $\underset{\sim}{x}_i$, e denotando este valor por $\tilde{E}[y_i\,|\underset{\sim}{x}_i]$, temos

$$\tilde{E}[y_i\,|\underset{\sim}{x}_i] = \left(1 + \lambda\,\underset{\sim}{x}_i^T\underset{\sim}{\hat{\beta}}\right)^{\frac{1}{\lambda}}. \tag{3.45}$$

De forma semelhante, no caso $\lambda = 0$,

$$\ln(y_i) = \underset{\sim}{x}_i^T\underset{\sim}{\beta} + \sum_{l=1}^{q}\sigma_l e_l, \quad i = 1, 2, ..., n,$$

implica em

$$E\left[\ln(y_i)\,|\underset{\sim}{x}_i\right] = E\left[\underset{\sim}{x}_i^T\underset{\sim}{\beta} + \sum_{l=1}^{q}\sigma_l e_l\right]$$

$$= \underset{\sim}{x}_i^T\underset{\sim}{\beta},$$

estimada por $\underset{\sim}{x}_i^T\underset{\sim}{\hat{\beta}}$. Aplicando a função inversa da transformação da variável resposta à esperança estimada de $\ln(y_i)$, dado $\underset{\sim}{x}_i$, e denotando este valor por $\tilde{E}[y_i\,|\underset{\sim}{x}_i]$, temos

$$\tilde{E}[y_i\,|\underset{\sim}{x}_i] = \exp(\underset{\sim}{x}_i^T\underset{\sim}{\hat{\beta}}). \tag{3.46}$$

Definimos como fator de ajuste a diferença entre o estimador da esperança de y_i, dado

$\underset{\sim}{x}_i$, e $\tilde{E}[y_i \mid \underset{\sim}{x}_i]$, ao qual denotamos por $B_i(\lambda)$, ou seja,

$$B_i(\lambda) = \hat{E}[y_i \mid \underset{\sim}{x}_i] - \tilde{E}[y_i \mid \underset{\sim}{x}_i], \quad i = 1, 2, ..., n.$$

Para o caso $\lambda \neq 0$, pelas equações (3.43) e (3.45), temos

$$B_i(\lambda) = (1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}})^{\frac{1}{\lambda}} \left\{ 1 + \frac{(1-\lambda)}{2} \sum_{l=1}^{q} \frac{\hat{\sigma}_l^2}{(1 + \lambda \sum_{i=1}^{n} \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}})^2} \right\} - \left(1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}\right)^{\frac{1}{\lambda}}$$

$$= \frac{1}{2}(1 + \lambda \, \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}})^{\frac{1}{\lambda}-2}(1-\lambda) \sum_{l=1}^{n} \hat{\sigma}_l^2.$$

Para o caso $\lambda = 0$, pelas equações (3.44) e (3.46), temos

$$B_i(\lambda) = \exp(\underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}) \exp(\frac{1}{2} \sum_{l=1}^{q} \hat{\sigma}_l^2) - \exp(\underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}})$$

$$= \exp(\underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}) \left\{ \exp(\frac{1}{2} \sum_{l=1}^{q} \hat{\sigma}_l^2) - 1 \right\}.$$

Definem-se as quantidades $B(\lambda)$ e $\tilde{y}$ por

$$B(\lambda) = \sum_{i=1}^{n} \frac{B_i(\lambda)}{n}$$

e

$$\tilde{y} = \sum_{i=1}^{n} \frac{\tilde{E}[y_i \mid \underset{\sim}{x}_i]}{n}.$$

Logo, um estimador em percentagem de vício relativo é definido como

$$VR = \frac{B(\lambda)}{\tilde{y}}.$$

Capítulo 4

Efeitos da Estimação de Esperança sob Escala Transformada na Estimação sob Escala Original

4.1 Introdução

No capítulo anterior foram apresentados vários estimadores para a esperança de y, dado x, após a regressão de $g(y)$ em x, onde g é uma transformação monótona. Agora, primeiramente, vamos analisar o efeito da estimação do modelo
$g(y_i) = \beta_0 + \beta_1 x_{i1} + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$, na estimação de $E[y_i \mid x_{i1}]$, na escala original. Na primeira seção, usando um modelo de regressão linear simples, consideraremos a variável resposta como uma variável resposta transformada, tal que possamos aplicar os estimadores propostos para a esperança na escala original. Na segunda seção, apresentaremos resultados de simulações e mostraremos gráficos para ilustrar a teoría descrita anteriormente.

4.2 Aspectos dos estimadores da média na escala original

Estamos considerando uma amostra do modelo de regressão linear simples

$$g(y_i) = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \tag{4.1}$$

onde

$$y_i > 0, \quad x_{i1} > 0,$$

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, ..., n,$$

onde g é uma transformação estritamente monótona e utilizamos os estimadores de quadrados mínimos $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\sigma}^2$ dos parâmetros β_0, β_1 e σ^2, com base nesta amostra. Vamos examinar as transformações: (a) logaritmo, (b) raiz quadrada, (c) inversa e as transformações da família Box-Cox (seção 3.1) usando (d) $\lambda = -1$ e (e) $\lambda = \frac{1}{2}$. Notemos que logaritmo é transformação de Box-Cox usando
$\lambda = 0$. Os estimadores propostos por Duan (seção 3.3) poderão ser aplicados a todas estas transformações, mas, os estimadores propostos por Miller (seção 3.2) só serão aplicados às transformações (a), (b) e (c), e os estimadores propostos por Taylor (seção 3.4) serão aplicados aos casos (a), (d) e (e).

A seguir, veremos se este comportamento permanece o mesmo para os estimadores para a esperança de y na escala original, tratando cada transformação específica.

4.2.1 Transformação logaritmo

Seja o modelo (4.1), usando $g(y) = \ln(y)$,

$$\ln(y_i) = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \quad i = 1, 2, ..., 30,$$

$$\epsilon_i \sim N(0, \sigma^2).$$

Logo, para x, o valor esperado do modelo transformado é $\beta_0 + \beta_1 x$, e voltando à escala original, o valor esperado na escala original é

$$\exp(\beta_0 + \beta_1 x)\exp(\frac{\sigma^2}{2}).$$

O estimador proposto por Miller está dado na equação (3.8), $\exp(\hat{\beta}_0 + \hat{\beta}_1 x)\exp(\frac{\hat{\sigma}^2}{2})$. Este estimador superestima o valor esperado se

$$\exp(\hat{\beta}_0 + \hat{\beta}_1 x)\exp(\frac{\hat{\sigma}^2}{2}) > \exp(\beta_0 + \beta_1 x)\exp(\frac{\sigma^2}{2}).$$

A superestimação dos estimadores, $\hat{\beta}_0, \hat{\beta}_1$ e $\hat{\sigma}^2$, com respeito aos parâmetros β_0, β_1 e σ^2, afeta muito a magnitude do estimador da média. Mesmo que $\hat{\beta}_0$ e $\hat{\beta}_1$ estejam próximos de sua esperança, se σ^2 é grande, uma pequena superestimação deste parâmetro causa uma grande superestimação do estimador da média na escala original.

O estimador proposto neste trabalho está dado na equação (3.10), $\exp(\hat{\beta}_0 + \hat{\beta}_1 x)\exp\left(\frac{\hat{\sigma}^2}{2}\left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]\right)$. Este estimador superestima o valor esperado se

$$\exp(\hat{\beta}_0 + \hat{\beta}_1 x)\exp\left(\frac{\hat{\sigma}^2}{2}\left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]\right) > \exp(\beta_0 + \beta_1 x)\exp(\frac{\sigma^2}{2}).$$

O nosso estimador não difere muito do estimador de Miller mas tem a vantagem que o efeito da magnitude de $\hat{\sigma}^2$ é um pouco reduzido por ser $\hat{\sigma}^2$ multiplicado por $\left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]$, onde $\left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]$ é um valor entre zero e um, pela condição (3.9).

O estimador proposto por Taylor está dado na equação (3.39), $\exp(\hat{\beta}_0 + \hat{\beta}_1 x)\left(1 + \frac{\hat{\sigma}^2}{2}\right)$. Este estimador superestima o valor esperado se

$$\exp(\hat{\beta}_0 + \hat{\beta}_1 x)\left(1 + \frac{\hat{\sigma}^2}{2}\right) > \exp(\beta_0 + \beta_1 x)\exp(\frac{\sigma^2}{2}).$$

Notemos que se $\hat{\sigma}^2$ é um valor pequeno, $\exp(\frac{\hat{\sigma}^2}{2}) \simeq 1 + \frac{\hat{\sigma}^2}{2} + O(\hat{\sigma}^2)$, o estimador proposto

por Taylor é aproximadamente igual ao estimador proposto por Miller, e se $\hat{\sigma}^2$ não é pequeno, o estimador proposto por Taylor não é adequado, pois sua definição é baseada em aproximação de $E[y \mid x]$, que exige σ^2 pequeno.

O estimador proposto por Duan está dado na equação (3.27), $\frac{1}{n}\sum_{i=1}^{n}\exp(\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\epsilon}_i)$, e este estimador superestima o valor esperado se

$$\frac{1}{n}\sum_{i=1}^{n}\exp(\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\epsilon}_i) > \exp(\beta_0 + \beta_1 x)\exp(\frac{\sigma^2}{2}).$$

4.2.2 Transformação raiz quadrada

Seja o modelo (4.1), usando $g(y) = \sqrt{y}$,

$$\sqrt{y_i} = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \quad i = 1, 2, ..., 30,$$

$$\epsilon_i \sim N(0, \sigma^2).$$

Logo, para x, o valor esperado do modelo transformado é $\beta_0 + \beta_1 x$, e voltando à escala original, o valor esperado na escala original é

$$(\beta_0 + \beta_1 x)^2 + \sigma^2.$$

O estimador proposto por Miller está dado na equação (3.15), $(\hat{\beta}_0 + \hat{\beta}_1 x)^2 + \hat{\sigma}^2$. Este estimador superestima o valor esperado se

$$(\hat{\beta}_0 + \hat{\beta}_1 x)^2 + \hat{\sigma}^2 > (\beta_0 + \beta_1 x)^2 + \sigma^2.$$

Podemos observar que se $\hat{\beta}_0$ e $\hat{\beta}_1$ estão próximos de sua esperança e se a magnitude de σ^2 com relação a $(\beta_0 + \beta_1 x)^2$ é pequena, teremos bons resultados, então a superestimação dos parâmetros afetará ao estimador da média na escala original, mas não fortemente. Para o caso de ter um valor grande de σ^2, vemos que seu efeito é aditivo, diminuindo dependência do estimador da média ao valor de σ^2.

O estimador proposto neste trabalho está dado na equação (3.16), $(\hat{\beta}_0 + \hat{\beta}_1 x)^2 + \hat{\sigma}^2 \left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]$. Este estimador superestima o valor esperado se

$$(\hat{\beta}_0 + \hat{\beta}_1 x)^2 + \hat{\sigma}^2 \left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right] > (\beta_0 + \beta_1 x)^2 + \sigma^2.$$

Como acontece com a transformação logaritmo, o nosso estimador não difere muito do estimador de Miller mas o efeito da magnitude de $\hat{\sigma}^2$ é um pouco reduzido por ser $\hat{\sigma}^2$ multiplicado por $\left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]$, onde $\left[1 - \frac{1}{n} - \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_{i1}-\bar{x})^2}\right]$ é um valor entre zero e um.

O estimador proposto por Duan está dado na equação (3.27), $\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\epsilon}_i)^2$. Este estimador superestima o valor esperado se

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\epsilon}_i)^2 > (\beta_0 + \beta_1 x)^2 + \sigma^2.$$

4.2.3 Transformação inversa

Seja o modelo (4.1), usando $g(y) = \frac{1}{y}$,

$$\frac{1}{y_i} = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \quad i = 1, 2, ..., 30,$$

$$\epsilon_i \sim N(0, \sigma^2).$$

Logo, para x, o valor esperado do modelo transformado é $\beta_0 + \beta_1 x$, e voltando à escala original, o valor esperado na escala original é aproximadamente

$$\frac{1}{\beta_0 + \beta_1 x}\left(1 + \frac{\sigma^2}{(\beta_0 + \beta_1 x)^2}\right).$$

O estimador proposto por Miller está dado na equação (3.23),

$\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\left(1+\frac{\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)$. Este estimador superestima o valor esperado se

$$\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\left(1+\frac{\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right) > \frac{1}{\beta_0+\beta_1 x}\left(1+\frac{\sigma^2}{(\beta_0+\beta_1 x)^2}\right) \Leftrightarrow$$

$$\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}+\frac{\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^3} > \frac{1}{\beta_0+\beta_1 x}+\frac{\sigma^2}{(\beta_0+\beta_1 x)^3}.$$

Vemos que $\frac{\sigma^2}{(\beta_0+\beta_1 x)^3}$ é decrescente em $\beta_0+\beta_1 x$. Se o valor esperado na escala transformada é subestimado, e se $\hat{\sigma}^2$ é superestimado, o valor esperado na escala original é superestimado.

O estimador proposto neste trabalho está dado na equação (3.24),

$$\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\left\{\frac{\left(1+\frac{\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)}{\left(1+\hat{\sigma}^2\frac{\left(1-\frac{1}{n}-\frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_{i1}-\bar{x})^2}\right)}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)}\right\}$$

. Este estimador superestima o valor esperado se

$$\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x}\left\{\frac{\left(1+\frac{\hat{\sigma}^2}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)}{\left(1+\hat{\sigma}^2\frac{\left(1-\frac{1}{n}-\frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_{i1}-\bar{x})^2}\right)}{(\hat{\beta}_0+\hat{\beta}_1 x)^2}\right)}\right\} > \frac{1}{\beta_0+\beta_1 x}\left(1+\frac{\sigma^2}{(\beta_0+\beta_1 x)^2}\right).$$

Temos de analisar a razão entre os termos

$$1+\frac{\sigma^2}{(\beta_0+\beta_1 x)^2} \quad \text{e} \quad 1+\sigma^2\frac{\left(1-\frac{1}{n}-\frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_{i1}-\bar{x})^2}\right)}{(\beta_0+\beta_1 x)^2}.$$

Vemos que a razão entre eles dá sempre um valor maior que um, aplicando a condição

(3.9). A magnitude dessa razão depende fortemente da magnitude de σ^2 em relação a $(\beta_0 + \beta_1 x)^2$. Se $\frac{\sigma^2}{(\beta_0+\beta_1 x)^2}$ é pequeno, então a razão é aproximadamente um, dando um bom estimador do valor esperado porque $\frac{\sigma^2}{(\beta_0+\beta_1 x)^2}$ também será pequeno.

O estimador proposto por Duan está dado na equação (3.27), $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\hat{\beta}_0+\hat{\beta}_1 x+\hat{\epsilon}_i}\right)$. Este estimador superestima o valor esperado se

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\epsilon}_i}\right) > \frac{1}{\beta_0 + \beta_1 x}\left(1 + \frac{\sigma^2}{(\beta_0 + \beta_1 x)^2}\right).$$

4.2.4 Transformação de Box-Cox usando $\lambda = -1$

Seja o modelo (4.1), usando $g(y) = 1 - \frac{1}{y}$,

$$1 - \frac{1}{y_i} = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \quad i = 1, 2, ..., 30,$$

$$\epsilon_i \sim N(0, \sigma^2).$$

Logo, para x, o valor esperado do modelo transformado é $\beta_0 + \beta_1 x$, e voltando à escala original, o valor esperado na escala original é aproximadamente

$$\frac{1}{1 - \beta_0 - \beta_1 x}\left(1 + \frac{\sigma^2}{(1 - \beta_0 - \beta_1 x)^2}\right).$$

O estimador proposto por Taylor está dado na equação (3.34), $\frac{1}{1-\beta_0-\beta_1 x}\left(1 + \frac{\sigma^2}{(1-\beta_0-\beta_1 x)^2}\right)$. Este estimador superestima o valor esperado se

$$\frac{1}{1 - \hat{\beta}_0 - \hat{\beta}_1 x}\left(1 + \frac{\hat{\sigma}^2}{(1 - \hat{\beta}_0 - \hat{\beta}_1 x)^2}\right) > \frac{1}{1 - \beta_0 - \beta_1 x}\left(1 + \frac{\sigma^2}{(1 - \beta_0 - \beta_1 x)^2}\right).$$

Temos de analisar o termo

$$\frac{\sigma^2}{(1 - \beta_0 - \beta_1 x)^3}.$$

Vemos que é crescente em $\beta_0 + \beta_1 x$. Se o valor esperado na escala transformada é superestimado, o valor esperado na escala original é superestimado.

O estimador proposto por Duan está dado na equação (3.27), $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1-\hat{\beta}_0-\hat{\beta}_1 x-\hat{\epsilon}_i}\right)$. Este estimador superestima o valor esperado se

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1-\hat{\beta}_0-\hat{\beta}_1 x-\hat{\epsilon}_i}\right) > \frac{1}{1-\beta_0-\beta_1 x}\left(1+\frac{\sigma^2}{(1-\beta_0-\beta_1 x)^2}\right).$$

Notemos que o estimador da média, proposto por Duan, $\left(\frac{1}{n}\sum_{i=1}^{n} h(\hat{\beta}_0+\hat{\beta}_1 x+\hat{\epsilon}_i)\right)$, é relativamente afetado pela magnitude de $\hat{\sigma}^2$, dependendo da transformação utilizada.

4.3 Simulações

Para realizar a simulação, usando o modelo dado em (4.1), os parâmetros $\beta_0, \beta_1, \sigma^2$ e o vetor $\underset{\sim}{x_1}$ foram definidos, com valores para x_1 entre 1 e 5,9, $\beta_0 = 3$, $\beta_1 = 1$. O parâmetro σ^2 tomou os valores: 0,5, 1 e 1,5. O tamanho de amostra foi $n = 30$. Apartir desses valores, construímos os valores de $\beta_0 + \beta_1 x_{i1}$. Geramos erros com distribuição normal de média zero e variância σ^2 e exponencial com parâmetro um, padronizado de média zero e variância um. Em seguida, construímos os valores de $g(y_i)$, somando o erro a $\beta_0 + \beta_1 x_{i1}$. Obtivemos os valores de y_i, escala original, usando a transformação inversa. Das várias amostras geradas e várias transformações investigadas, calculamos as estimativas das esperanças na escala original.

Destacamos três dessas amostras, denotadas **A**, **B** e **C**, para ilustrar as comparações entre os vários estimadores:

- A amostra **A** foi gerada com erros normais, $\sigma^2 = 0,5$, observou-se $\hat{\sigma}^2 = 0,58$, $\hat{\beta}_0 > \beta_0$ e $\hat{\beta}_1 < \beta_1$;
- A amostra **B** foi gerada com erros normais, $\sigma^2 = 1$, observou-se $\hat{\sigma}^2 = 1,34$, $\hat{\beta}_0 < \beta_0$ e $\hat{\beta}_1 > \beta_1$;
- A amostra **C** foi gerada com erros exponenciais, $\sigma^2 = 1$, observou-se $\hat{\sigma}^2 = 1,35$, $\hat{\beta}_0 > \beta_0$ e $\hat{\beta}_1 < \beta_1$.

Cada figura, números (4-1) a (4-12), ilustra a amostra na escala original de y e as

estimativas das esperanças na escala original para os estimadores indicados, comparadas com os valores verdadeiros de $E[y \mid x]$, para x da faixa de 1 a 5,9. Consta da legenda da figura a transformação utilizada.

Notamos de maneira geral comportamento semelhante para os três estimadores. Com respeito à transformação logaritmo, só na amostra **A** os resultados foram razoáveis. Nas amostras **B** e **C** valores de $\hat{\sigma}^2$ pouco maiores do que σ^2 causaram uma superestimação, com a transformação logaritmo. Numa outra amostra, não ilustrada, com erros com distribuição normal de variância $\sigma^2 = 10$, observamos $\hat{\sigma}^2 = 12,1$, o que causou uma superestimação muito pronunciada (transformação logaritmo). As demais transformações decorreram bem, σ^2 era relativamente pequeno, com respeito à transformação empregada.

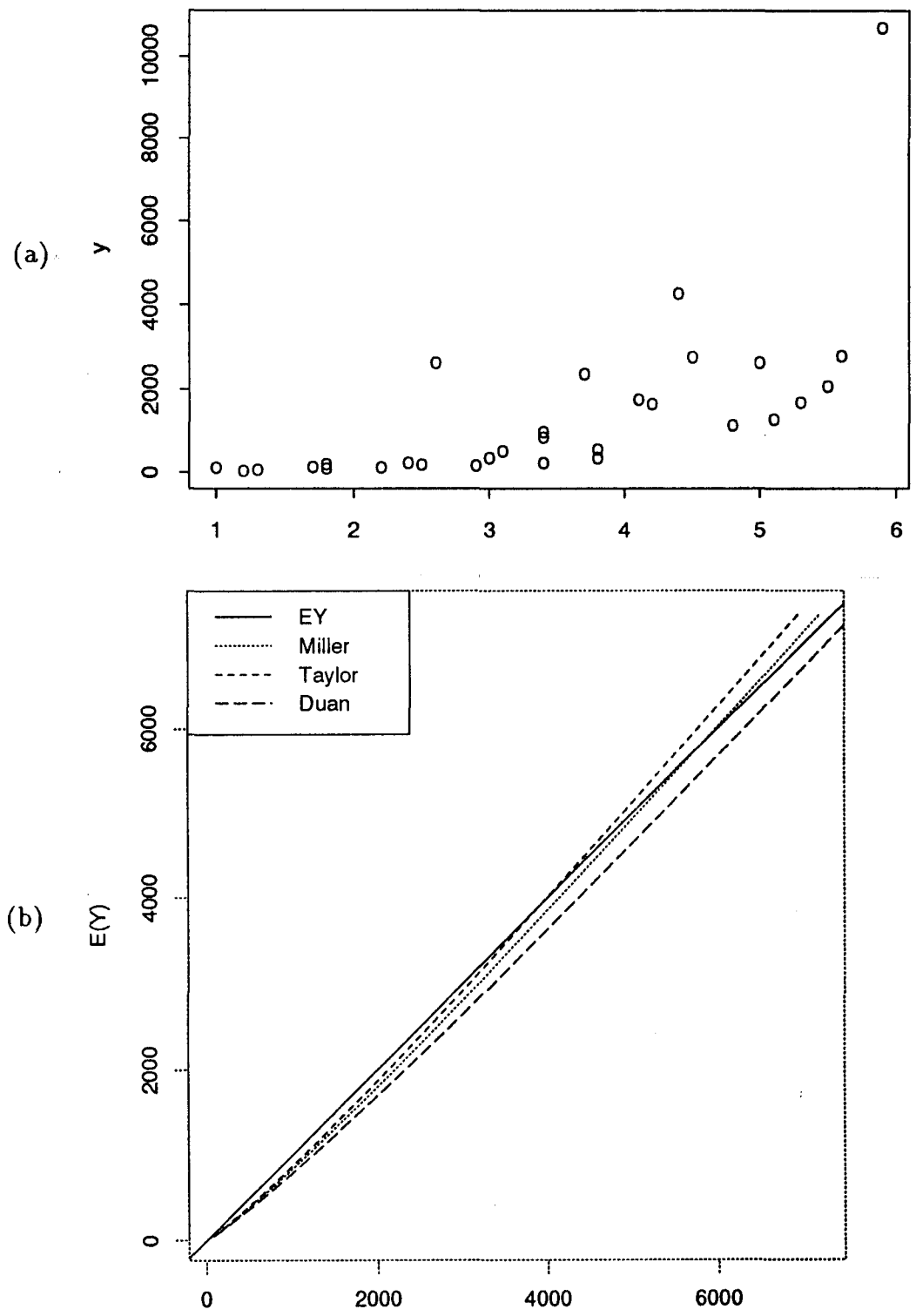


Figura 4-1: (a) Gráfico de y versus x da amostra A - o modelo linear é ajustado a logaritmo de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

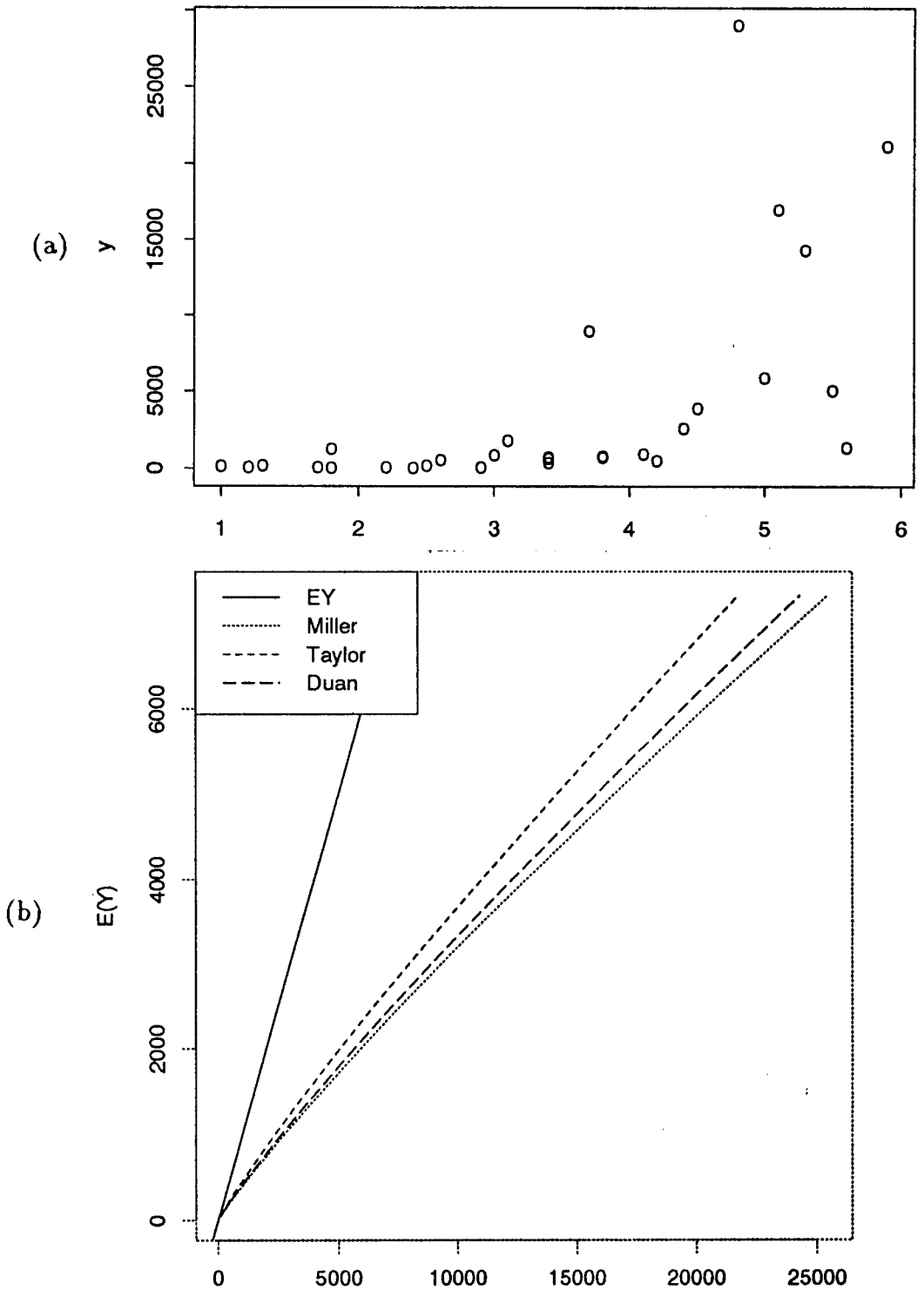


Figura 4-2: (a) Gráfico de y versus x da amostra B - o modelo linear é ajustado a logaritmo de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

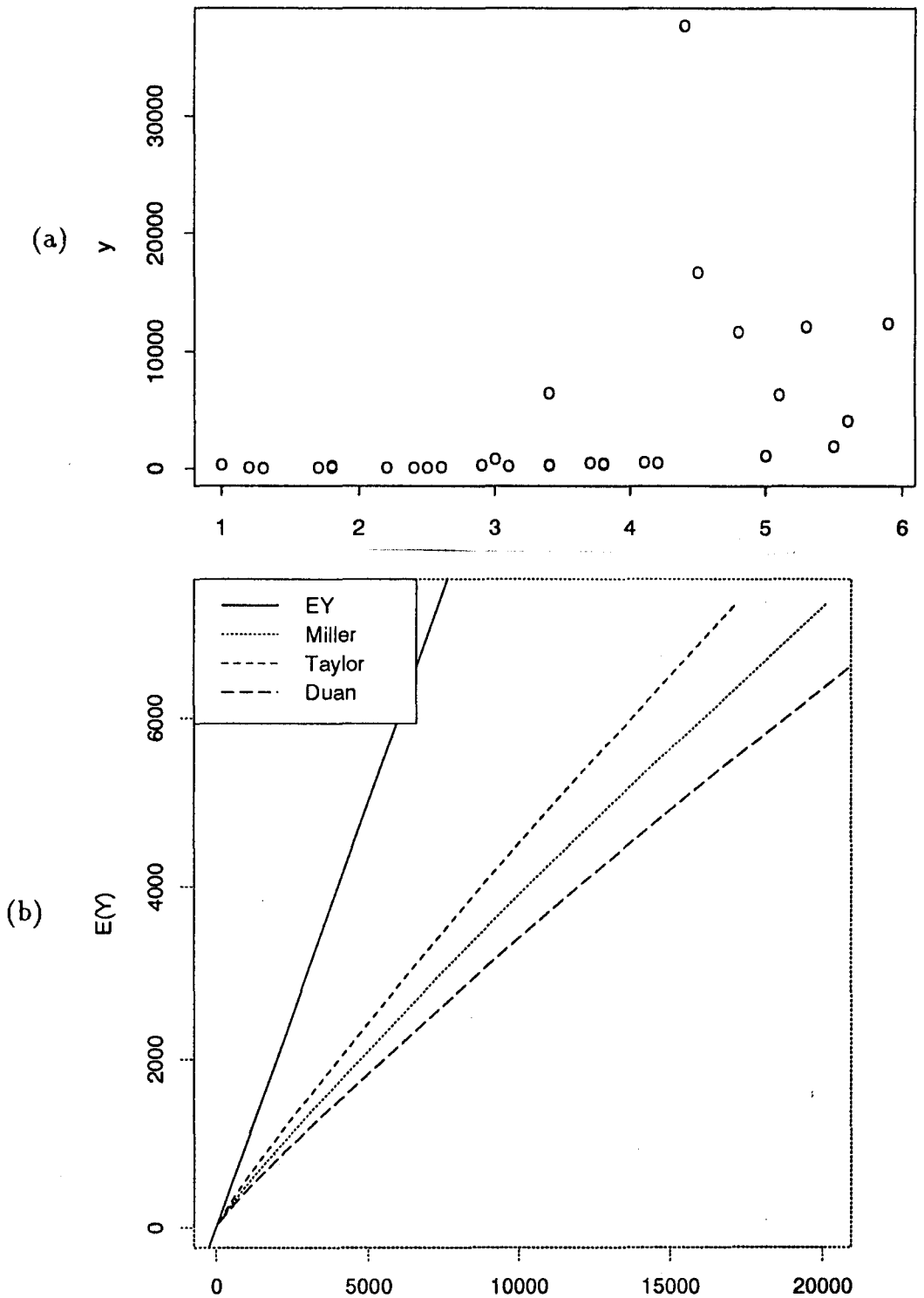


Figura 4-3: (a) Gráfico de y versus x da amostra C - o modelo linear é ajustado a logaritmo de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

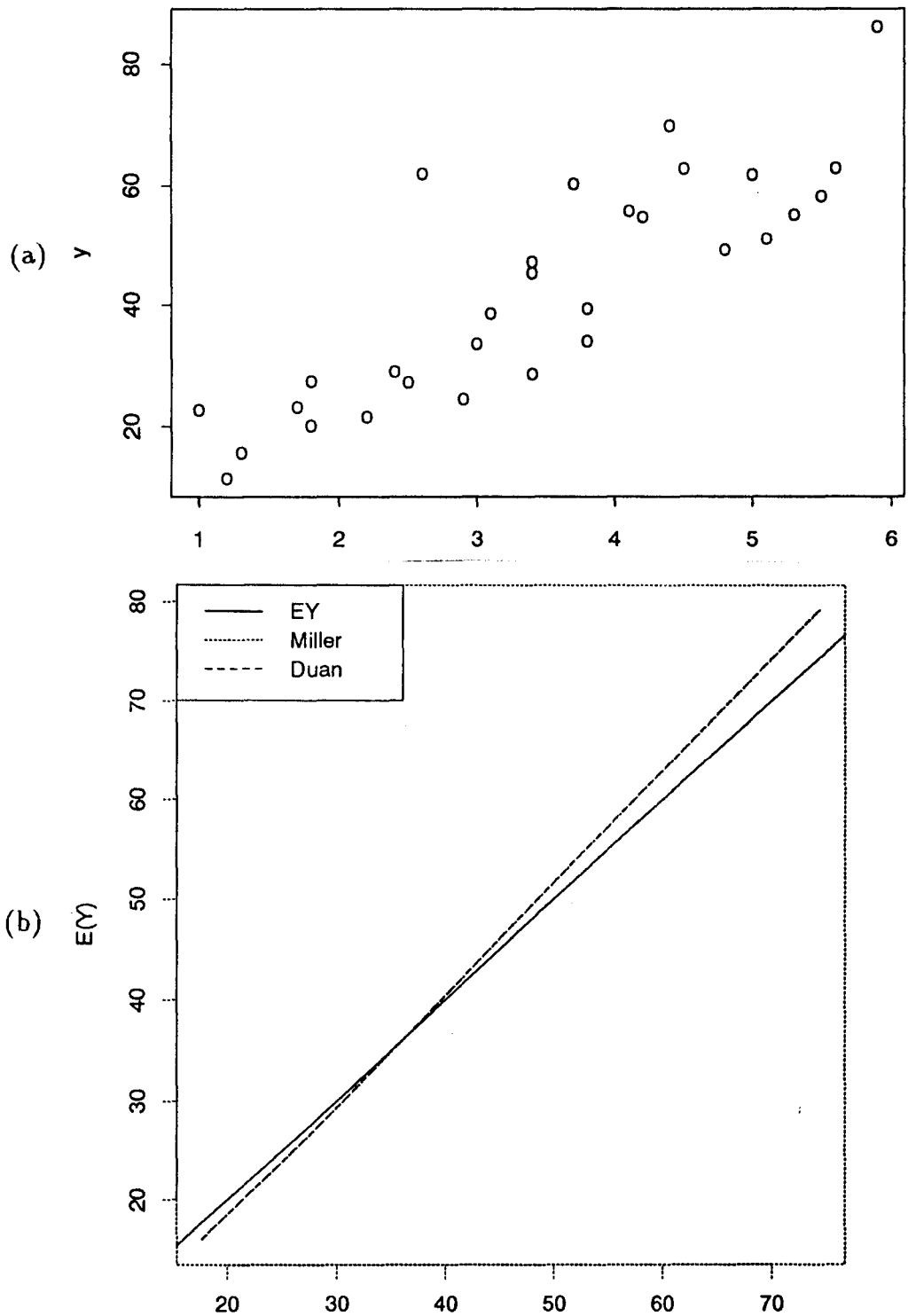


Figura 4-4: (a) Gráfico de y versus x da amostra A - o modelo linear é ajustado à raiz quadrada de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

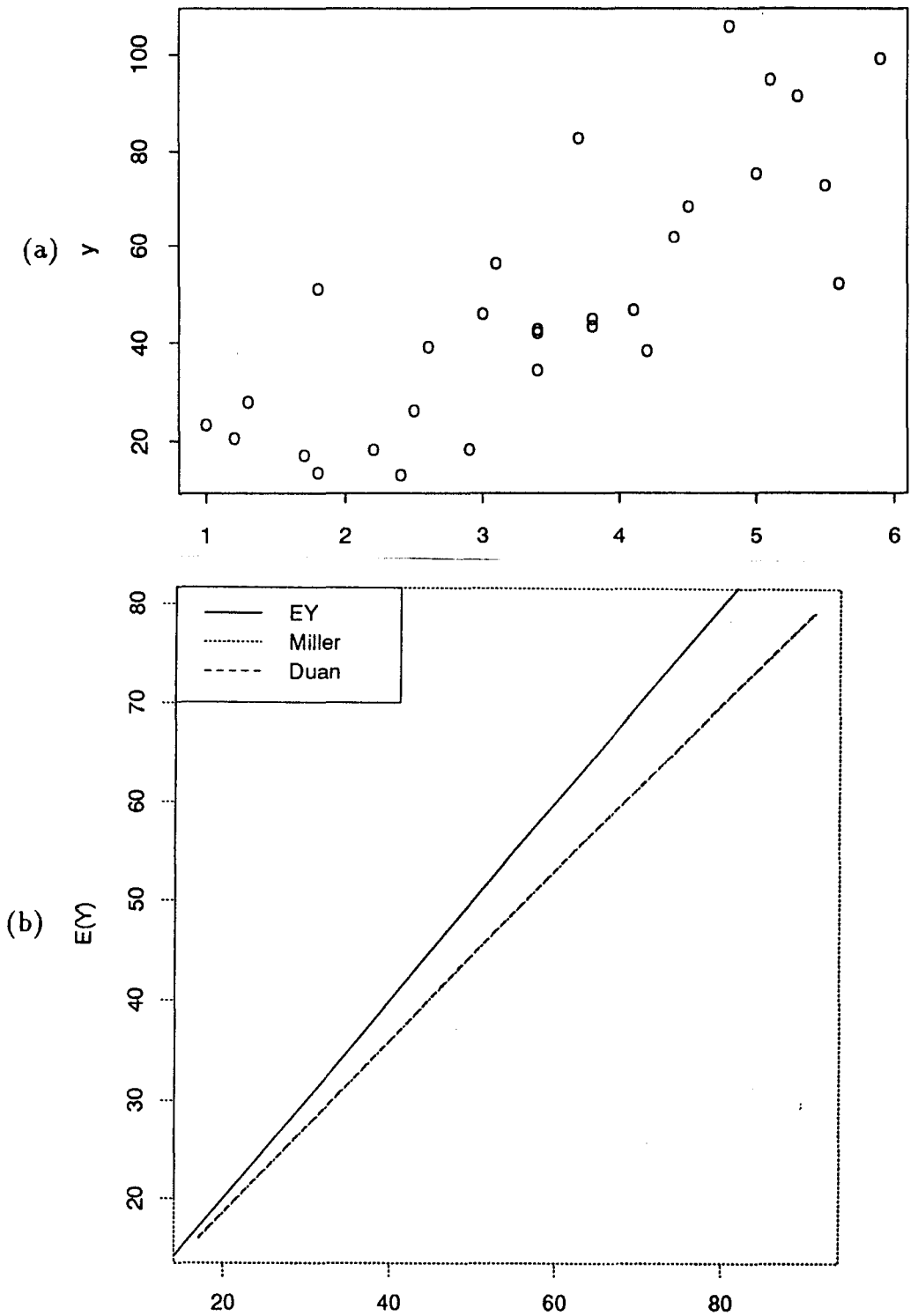


Figura 4-5: (a) Gráfico de y versus x da amostra B - o modelo linear é ajustado á raiz quadrada de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

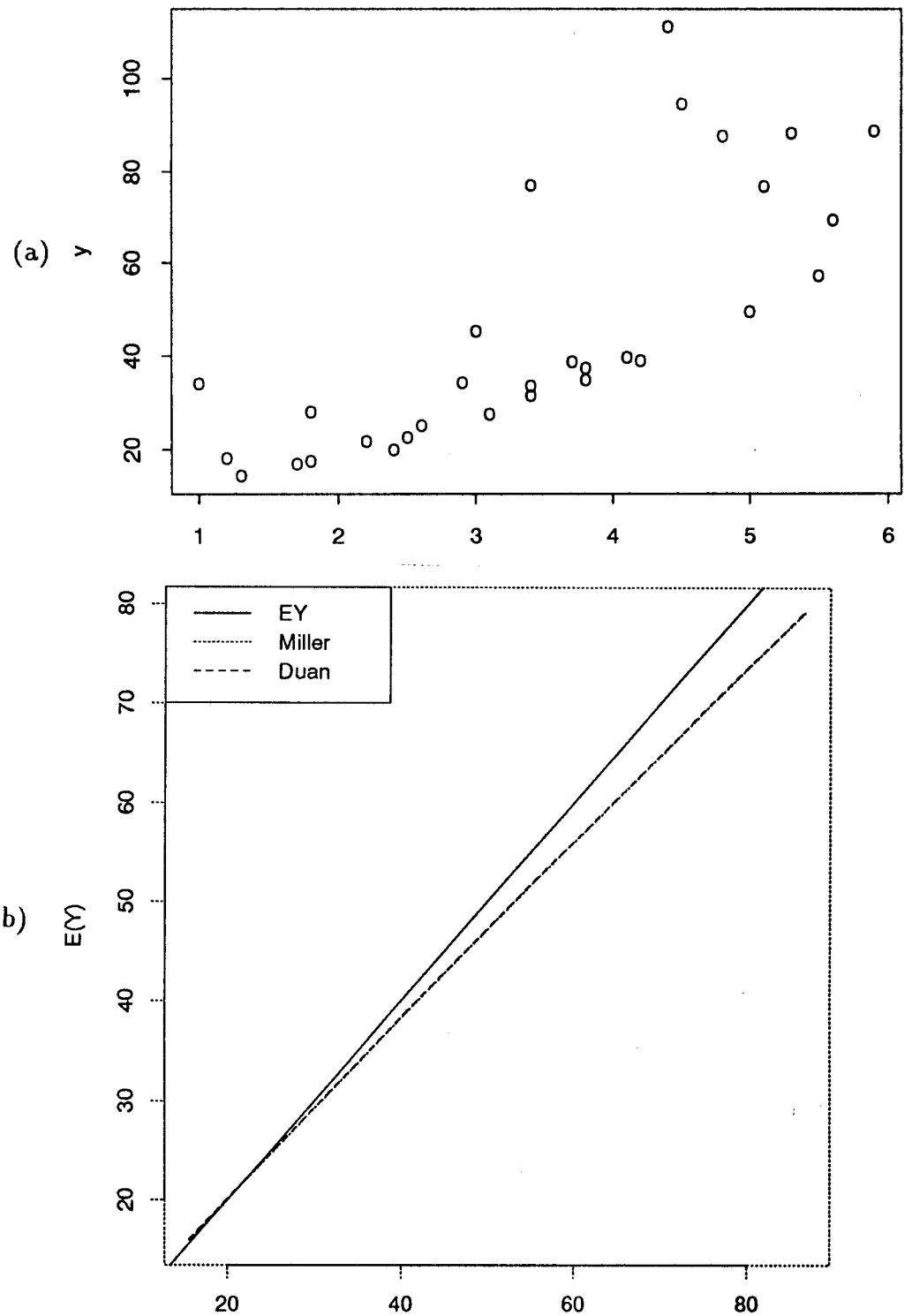


Figura 4-6: (a) Gráfico de y versus x da amostra C - o modelo linear é ajustado á raiz quadrada de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

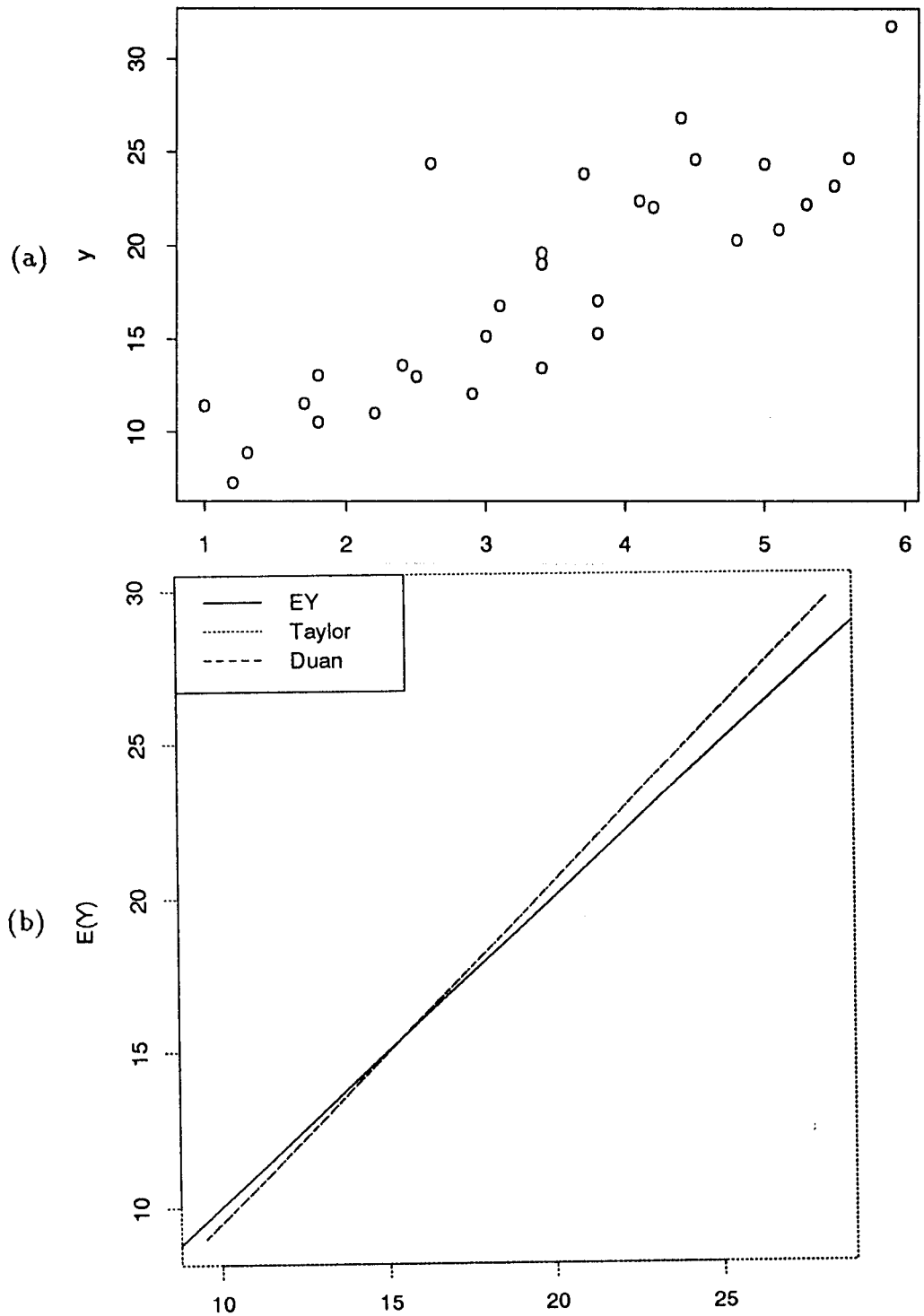


Figura 4-7: (a) Gráfico de y versus x da amostra A - o modelo linear é ajustado á transformação Box-Cox, $\lambda = 1/2$ de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

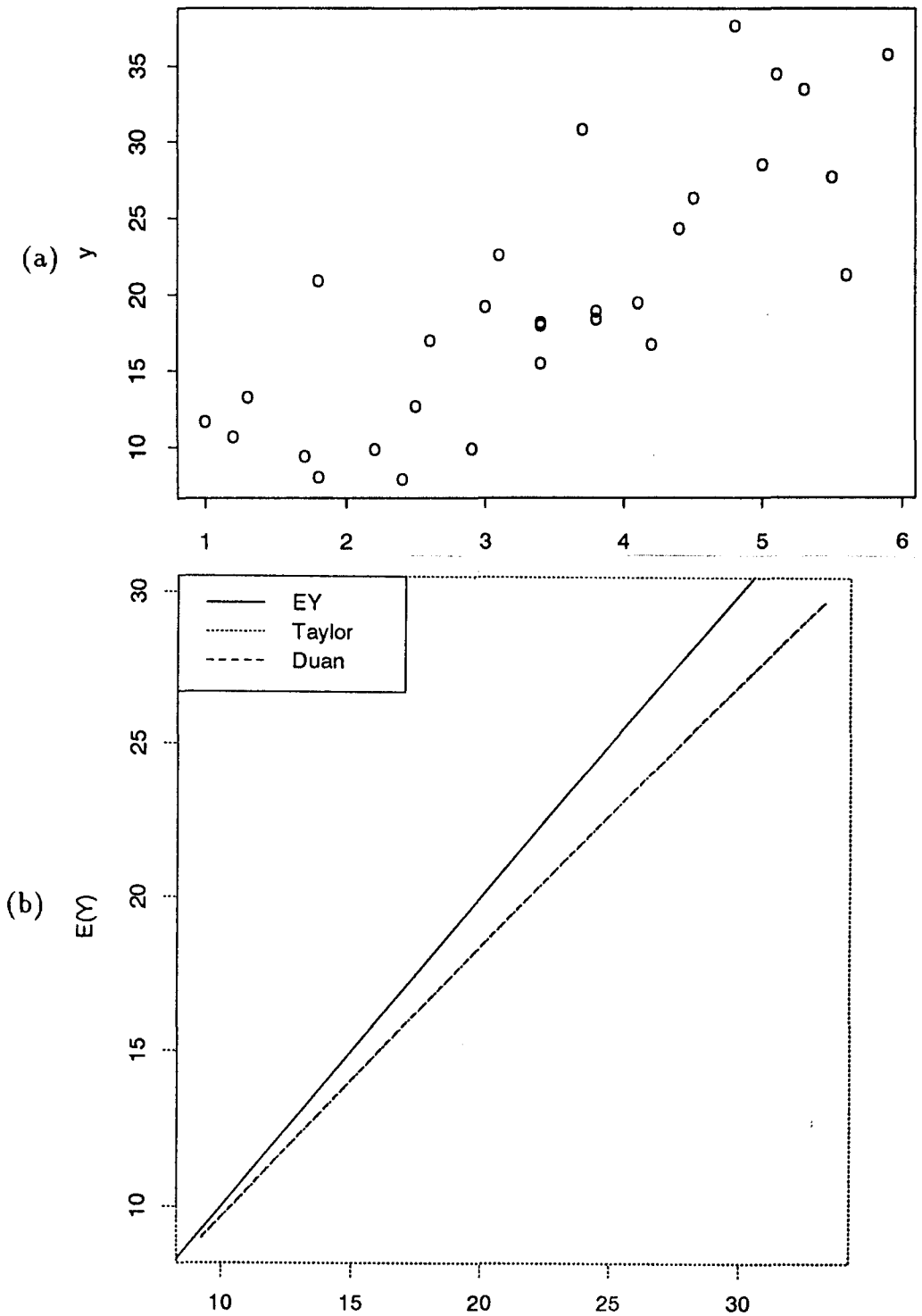


Figura 4-8: (a) Gráfico de y versus x da amostra B - o modelo linear é ajustado á transformação Box-Cox, $\lambda = 1/2$ de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

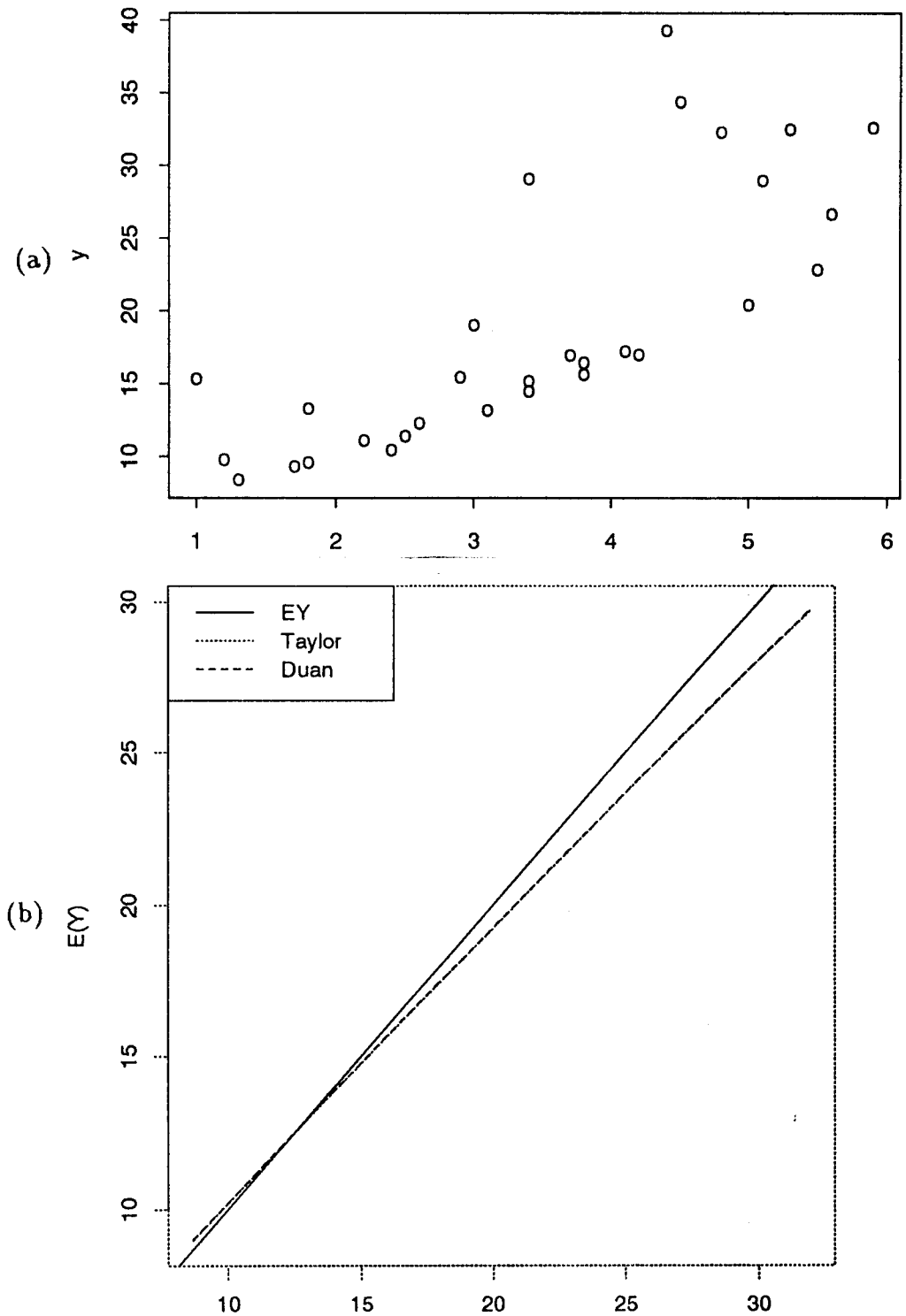


Figura 4-9: (a) Gráfico de y versus x da amostra C - o modelo linear é ajustado á transformação Box-Cox, $\lambda = 1/2$ de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

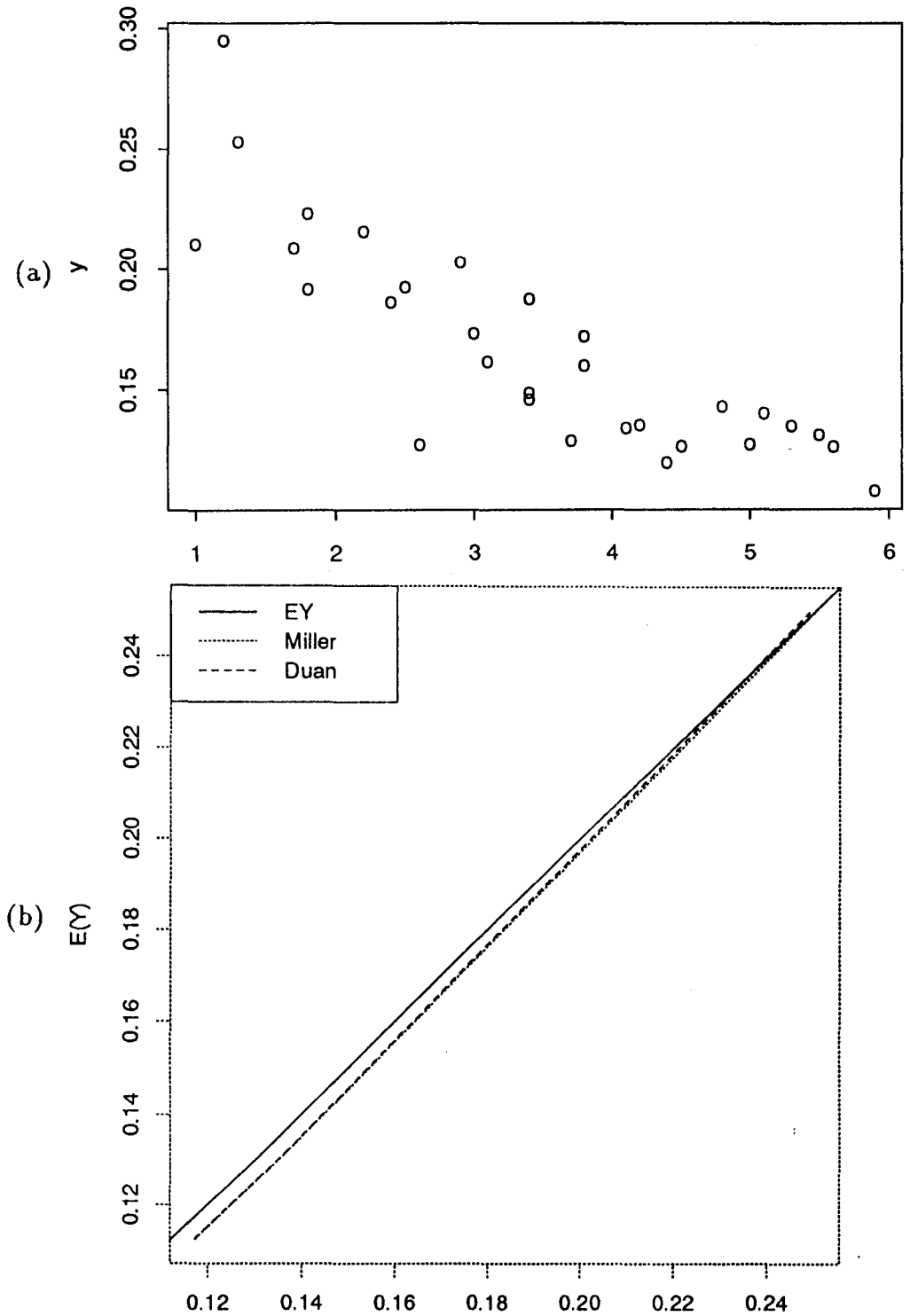


Figura 4-10: (a) Gráfico de y versus x da amostra A - o modelo linear é ajustado á inversa de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

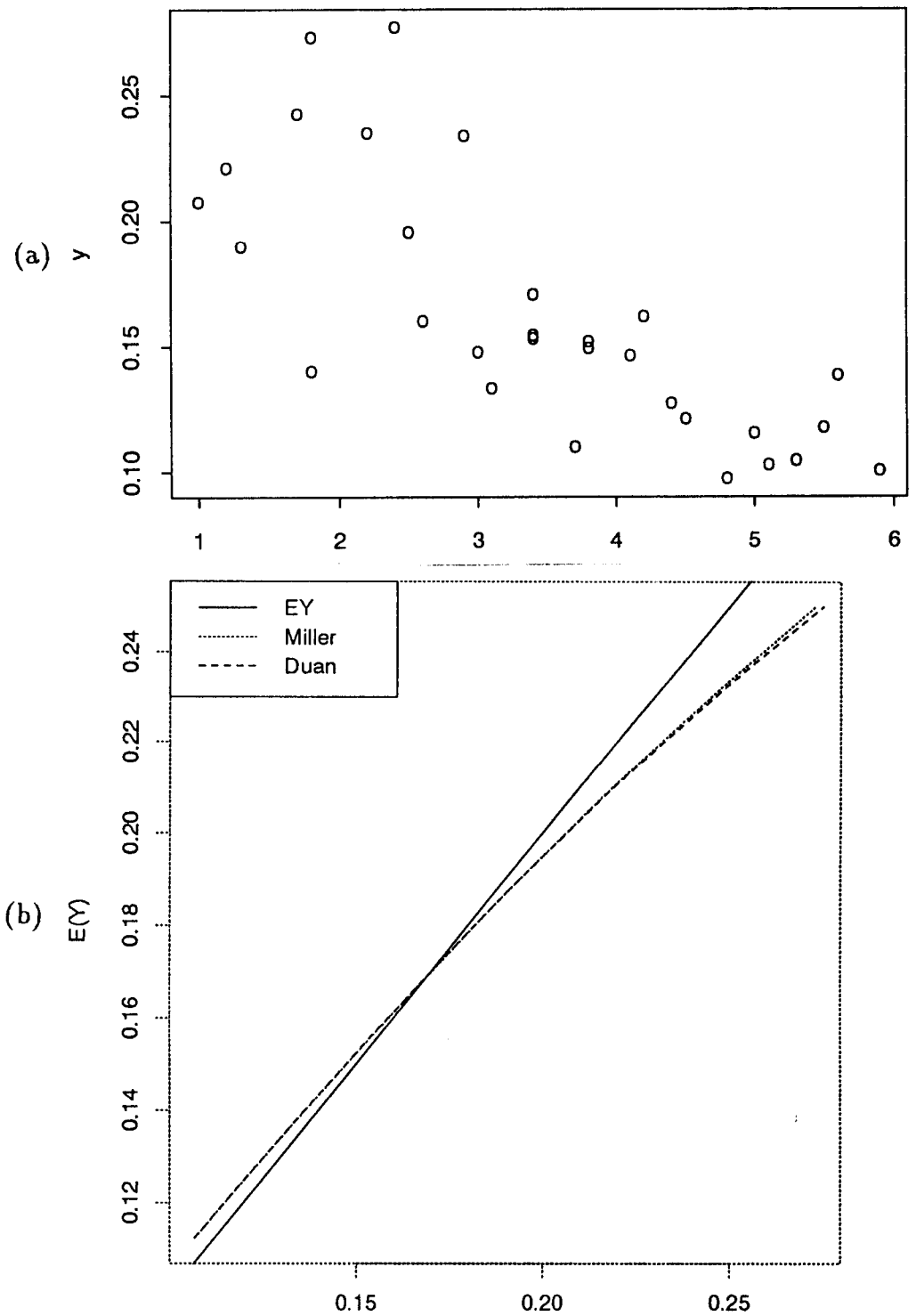


Figura 4-11: (a) Gráfico de y versus x da amostra B - o modelo linear é ajustado á inversa de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

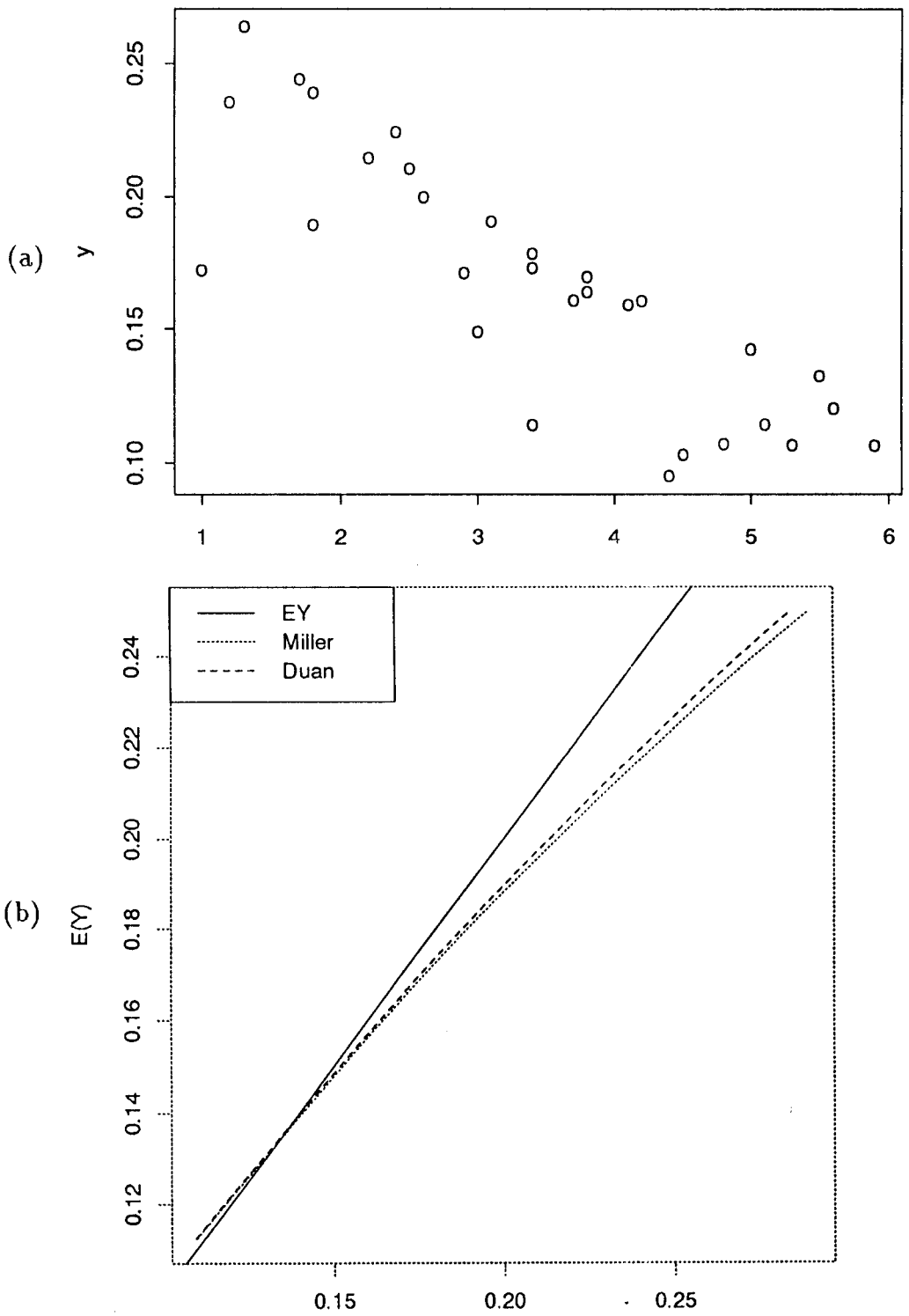


Figura 4-12: (a) Gráfico de y versus x da amostra C - o modelo linear é ajustado á inversa de y; (b) Gráfico de valores verdadeiros da esperança versus estimadores

Capítulo 5

Intervalos de Predição

5.1 Introdução

Um dos objetivos principais de usar a análise de regressão é fazer predição da variável resposta para valores específicos das variáveis independentes. Scott e Wild (1991) citam o seguinte exemplo: no National Women's Hospital em Auckland, Nova Zelândia, foram obtidas medidas ultrassônicas em fetos normais, desde o ano 1983 até 1988, quando a variável resposta era o comprimento do fígado do feto e a variável regressora era o tempo de gestação. O objetivo do estudo era ajustar um modelo para obter intervalos de predição, para o comprimento do fígado, dado o tempo de gestação.

Os limites de predição para uma nova observação dependem dos estimadores dos parâmetros do modelo (frequentemente, pelo menos assintoticamente normais) e da distribuição do erro do modelo (que pode ser normal ou não). Vamos apresentar neste capítulo a construção de um intervalo de predição sob as condições do modelo Gauss-Markov normal e também a construção de intervalo de predição assumindo um modelo Gauss-Markov, sem suposição de normalidade, definido em Schmoyer (1992). Este último é assintoticamente válido para uma variedade de tipos de distribuições para o erro. Schmoyer realizou simulações, concluindo que seu método é comparável a métodos de reamostragem, sendo, no entanto, de computação menos intensa. Quando a distribuição do erro não é normal, o intervalo proposto tem um desempenho razoável, a não ser que a assimetria da distribuição seja muito acentuada.

Com relação à inferência na escala original, se o modelo Gauss-Markov normal parece ser adequado à modelagem da variável resposta transformada, o intervalo de predição usual (Gauss-Markov normal) seria utilizado e a transformação inversa dos limites do intervalo forneceria um intervalo de predição na escala original. Porém, frequentemente, após a transformação da variável resposta, os resíduos parecem indicar alguma evidência contra normalidade, como, por exemplo, alguma assimetria na distribuição dos erros. Neste caso, o intervalo de predição, proposto por Schmoyer, apresenta-se como uma alternativa mais razoável, e, novamente, a transformação inversa dos limites do intervalo proporcionaria um intervalo de predição na escala original. No exemplo mencionado no primeiro parágrafo, a variável resposta é transformada. A suposição de normalidade, conferida através do exame dos resíduos não detectou nenhuma evidência contrária, assim, o intervalo de predição usual foi utilizado.

Apresentamos na seção 5.2 o desenvolvimento para a definição de intervalo de predição sob o modelo Gauss-Markov normal e na seção 5.3 alguns resultados expostos por Schmoyer para a definição de intervalo de predição. Na seção 5.4 expomos resultados de simulações, realizados por Schmoyer, além de pequena simulação por nós realizada.

O programa utilizado para o cálculo do intervalo de predição proposto por Schmoyer, foi desenvolvido neste trabalho, usando a linguagem SAS.

5.2 Intervalo de Predição Sob o Modelo Gauss-Markov Normal

Seja o modelo Gauss-Markov normal, apresentado no capítulo 3, (3.1), isto é,

$$y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \epsilon_i,$$

onde $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, 2, ..., n$ e $\mathbf{X}$ é matriz de posto coluna completo, com a i-ésima linha, $\underset{\sim}{x}_i^T = (1, x_{i1}, ..., x_{ip})$. Sejam $\underset{\sim}{\hat{\beta}}$ e $\hat{\sigma}^2$ os estimadores de quadrados mínimos de $\underset{\sim}{\beta}$ e σ^2,

baseados nesta amostra de tamanho n.

Considere y_0 uma nova observação, correspondente a $\underset{\sim}{x}_0$, independente das n observações anteriores, ou seja,

$$y_0 = \underset{\sim}{x}_0^T \underset{\sim}{\beta} + \epsilon_0, \tag{5.1}$$

onde

$$\epsilon_0 \sim N(0, \sigma^2),\ E[\epsilon_0 \epsilon_i] = 0,\ \ i = 1, 2, ..., n.$$

Logo, para $\hat{y}_0 = \underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}}$,

$$E[y_0] = \underset{\sim}{x}_0^T \underset{\sim}{\beta},$$

$$E[y_0 - \hat{y}_0] = 0,$$

$$Cov(y_0, \hat{y}_0) = 0$$

pois

$$\begin{aligned} Cov(y_0, \hat{y}_0) &= Cov(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \epsilon_0, \underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}}) \\ &= Cov(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \epsilon_0, \underset{\sim}{x}_0^T (X^T X)^{-1} X^T \underset{\sim}{y}) \\ &= Cov(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \epsilon_0, \underset{\sim}{x}_0^T (X^T X)^{-1} X^T (X \underset{\sim}{\beta} + \underset{\sim}{\epsilon})) \\ &= Cov(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \epsilon_0, \underset{\sim}{x}_0^T (X^T X)^{-1} X^T X \underset{\sim}{\beta} + \underset{\sim}{x}_0^T (X^T X)^{-1} X^T \underset{\sim}{\epsilon}) \\ &= 0 \end{aligned}$$

porque $E[\epsilon_0 \epsilon_i] = 0$, $i = 1, 2, ..., n$. Logo,

$$\begin{aligned} Var[y_0 - \hat{y}_0] &= Var[y_0] + Var[\hat{y}_0] \\ &= \sigma^2 + \sigma^2 \left(\underset{\sim}{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \underset{\sim}{x}_0 \right) \\ &= \sigma^2 \left(1 + \underset{\sim}{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \underset{\sim}{x}_0 \right), \end{aligned}$$

que será estimada por

$$s^2_{y_0} = \hat{\sigma}^2 \left(1 + \underset{\sim}{x}_0^T \, (\mathbf{X}^T\mathbf{X})^{-1} \, \underset{\sim}{x}_0\right).$$

Então, pelo lema 1 do Apêndice, $\frac{y_0 - \hat{y}_0}{s_{y_0}}$ tem distribuição $t - Student$ com $n - p - 1$ graus de liberdade, ou seja,

$$\frac{y_0 - \hat{y}_0}{s_{y_0}} \sim t_{n-p-1}.$$

Portanto,

$$P[\hat{y}_0 + s_{y_0} t_{(n-p-1,\frac{\alpha}{2})} \leq y_0 \leq \hat{y}_0 + s_{y_0} t_{(n-p-1,1-\frac{\alpha}{2})}] = 1 - \alpha, \tag{5.2}$$

onde $t_{(n-p-1,\frac{\alpha}{2})}$ é o quantil $\frac{\alpha}{2}$ da distribuição $t - Student$ com $n-p-1$ graus de liberdade.

5.3 Intervalo de Predição Assintótico Sob o Modelo Gauss-Markov

Schmoyer (1992) propõe um tipo de intervalo de predição para uma observação futura, sob o modelo Gauss-Markov. A distribuição do erro não é necessariamente normal, portanto, a sua utilização se direciona às situações onde a distribuição do erro possa ser, por exemplo, assimétrica.

Sejam y_i, $\underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}, \hat{\sigma}^2$ e $\mathbf{X}$, $i = 1, 2, ..., n$, conforme definidos no início da seção anterior. Considere a função $H_n(t)$, definida para $t \in R$,

$$H_n(t) = \frac{1}{n}\left[\sum_{i=1}^{n} \Psi_{n-p-1}\left(\frac{t - \hat{\epsilon}_i}{s_0}\right)\right],$$

onde $\hat{\epsilon}_i = y_i - \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}, s_0^2 = \hat{\sigma}^2 \left(\underset{\sim}{x}_0^T \, (\mathbf{X}^T\mathbf{X})^{-1} \, \underset{\sim}{x}_0\right)$, e Ψ_{n-p-1} é a função de distribuição acumulada da $t - Student$ com $n - p - 1$ graus de liberdade.

Um teorema relaciona a função $H_n(t)$ a $\Pr[y_0 \leq \hat{y}_0 + t]$, onde $\hat{y}_0 = \underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}}$.

Teorema 5.1

Se $\underset{\sim}{x}_0^T \, (\underset{\sim}{\hat{\beta}} - \underset{\sim}{\beta})$ é aproximadamente distribuído como $N(0, \sigma^2 \, \underset{\sim}{x}_0^T \, (\mathbf{X}^T\mathbf{X})^{-1} \, \underset{\sim}{x}_0)$, então, $\Pr[y_0 \leq \hat{y}_0 + t] \simeq H_n(t)$.

Prova:

$$\Pr[y_0 \leq \hat{y}_0 + t] = \Pr\left[\underset{\sim}{x_0^T}\,(\underset{\sim}{\beta} - \underset{\sim}{\hat{\beta}}) \leq t - \epsilon_0\right]$$

$$= \int \Pr\left[\underset{\sim}{x_0^T}\,(\underset{\sim}{\beta} - \underset{\sim}{\hat{\beta}}) \leq t - \epsilon_0 \mid \epsilon_0\right] \partial F(\epsilon_0),$$

onde F é a função de distribuição acumulada de ϵ_0,

$$= \int \Pr\left[\underset{\sim}{x_0^T}\,(\underset{\sim}{\beta} - \underset{\sim}{\hat{\beta}}) \leq t - \epsilon_0\right] \partial F(\epsilon_0)$$

porque os erros $\epsilon_1, ..., \epsilon_n$ são independentes de ϵ_0,

$$\simeq \int \Psi_{n-p-1}\left[\frac{t - \epsilon_0}{s_0}\right] \partial F(\epsilon_0).$$

Este termo pode ser estimado por

$$\int \Psi_{n-p-1}\left[\frac{t - \epsilon_0}{s_0}\right] \partial \hat{F}_n(\epsilon_0),$$

onde $\hat{F}_n$ é a função de distribuição empírica dos resíduos. Então,

$$\int \Psi_{n-p-1}\left[\frac{t - \epsilon_0}{s_0}\right] \partial \hat{F}_n(\epsilon_0) = \frac{1}{n}\sum_{i=1}^{n} \Psi_{n-p-1}\left[\frac{t - \hat{\epsilon}_i}{s_0}\right] = H_n(t). \quad \square$$

Portanto, um estimador de $\Pr[y_0 \leq \hat{y}_0 + t]$ é $H_n(t)$.

Schmoyer conclui que, para n pequeno, resultados de simulações sugerem que os quantis de H_n são limites de predição anticonservativos, então $\tilde{H}_n(t) = \frac{1}{n+1}\left[\frac{1}{2} + nH_n(t)\right]$ por ser mais dispersa, parece ser mais adequada para a definição de intervalo de predição.

Portanto, o intervalo proposto se baseia na seguinte função

$$\tilde{H}_n(t) = \frac{1}{n+1}\left[\frac{1}{2} + \sum_{i=1}^{n} \Psi_{n-p-1}\left(\frac{t - \hat{\epsilon}_i}{s_0}\right)\right].$$

Podemos observar que a função $\tilde{H}_n(t)$ é contínua e estritamente crescente, pois é diretamente proporcional à Ψ_{n-p-1}, que é estritamente crescente. Logo, podemos garantir a

existência da função inversa de $\tilde{H}_n(t)$.

Sejam

$$\gamma = \tilde{H}_n(T_\gamma),$$

portanto,

$$T_\gamma = \tilde{H}_n^{-1}(\gamma),$$

isto é, T_γ é o γ-ésimo quantil de $\tilde{H}_n$.

Resultado 5.1

T_γ está definido para os valores de γ que estão no intervalo $\left[\frac{1}{2(n+1)}, 1 - \frac{1}{2(n+1)}\right]$.

Prova:

$$\begin{aligned}
\lim_{t\to-\infty} \tilde{H}_n(t) &= \lim_{t\to-\infty} \frac{1}{n+1}\left[\frac{1}{2} + \sum_{i=1}^{n} \Psi_{n-p-1}\left(\frac{t-\hat{\epsilon}_i}{s_0}\right)\right] \\
&= \lim_{t\to-\infty} \frac{1}{n+1}\left[\frac{1}{2}\right] + \lim_{t\to-\infty} \frac{1}{n+1}\left[\sum_{i=1}^{n} \Psi_{n-p-1}\left(\frac{t-\hat{\epsilon}_i}{s_0}\right)\right] \\
&= \frac{1}{n+1}\left[\frac{1}{2}\right]
\end{aligned}$$

e

$$\begin{aligned}
\lim_{t\to\infty} \tilde{H}_n(t) &= \lim_{t\to\infty} \frac{1}{n+1}\left[\frac{1}{2} + \sum_{i=1}^{n} \Psi_{n-p-1}\left(\frac{t-\hat{\epsilon}_i}{s_0}\right)\right] \\
&= \lim_{t\to\infty} \frac{1}{n+1}\left[\frac{1}{2}\right] + \lim_{t\to\infty} \frac{1}{n+1}\left[\sum_{i=1}^{n} \Psi_{n-p-1}\left(\frac{t-\hat{\epsilon}_i}{s_0}\right)\right] \\
&= \frac{1}{n+1}\left[\frac{1}{2}\right] + \frac{1}{n+1}[n] \\
&= 1 - \frac{1}{2(n+1)}. \qquad \square
\end{aligned}$$

Baseados no teorema anterior, calculamos o intervalo de predição assintoticamente válido da seguinte forma:

$$\left[\hat{y}_0 + T_{\frac{\alpha}{2}}, \hat{y}_0 + T_{1-\frac{\alpha}{2}}\right], \tag{5.3}$$

tal que quando n aumenta, a probabilidade de y_0 estar contido neste intervalo converge a $1 - \alpha$.

Para achar os limites do intervalo (5.3), temos de resolver a equação $\tilde{H}_n(t) = \gamma$ para t, usando $\gamma = \frac{\alpha}{2}$ e $\gamma = 1 - \frac{\alpha}{2}$.

Definimos a probabilidade de exclusão (PE) de um intervalo de predição $[I, S]$ para y_0 como a probabilidade de que $y_0 < I$ ou $y_0 > S$, onde tanto y_0 como I e S são aleatórios. Para intervalos assintoticamente válidos, a probabilidade de exclusão calculada converge à probabilidade de exclusão nominal.

Schmoyer argumenta que a probabilidade de exclusão nominal de um intervalo de predição deveria ser limitada, conforme discutido a seguir. Se conhecêssemos o vetor de parâmetros β, o problema de fazer predições seria predizer ϵ_0 dados $\epsilon_1, \epsilon_2, ..., \epsilon_n$. Sem um maior conhecimento sobre a função de distribuição acumulada dos erros, o intervalo $(\epsilon_{(1)}, \epsilon_{(n)})$ seria o intervalo de predição mais largo para ϵ_0, podendo-se calcular a probabilidade de exclusão. Para o caso em que distribuição é contínua, temos que

$$\Pr\left[\epsilon_0 \leq \epsilon_{(1)}\right] = \Pr\left[\epsilon_0 \geq \epsilon_{(n)}\right] = \frac{1}{n+1}.$$

Então, quando não fazemos nenhuma suposição sobre a distribuição dos erros, a probabilidade de exclusão nominal unilateral deveria ser maior que $\frac{1}{n+1}$. Paralelamente, pelo resultado 5.1, deveríamos evitar probabilidade de exclusão nominal unilateral próxima a $\frac{1}{2(n+1)}$.

5.4 Simulações

Schmoyer realizou simulações de amostras de pares (x, y), onde $y = \beta_0 + \beta_1 x + \epsilon$, para várias distribuições do erro, ϵ, padronizadas para esperança zero e variância um. No que se refere a comparações entre o intervalo de predição usual (5.2) e o proposto (5.3), chegou a várias conclusões, que aqui expomos.

Quando a distribuição do erro é normal, as probabilidades de exclusão unilaterais são menores do que $\frac{\alpha}{2}$, para os dois tipos de intervalo de predição de probabilidade $1 - \alpha$. Ou seja, intervalos mais largos que o necessário, sendo o intervalo proposto menor

que o intervalo usual. Eventualmente, para amostras maiores, essas probabilidades são aproximadamente $\frac{\alpha}{2}$. Quanto menor o valor de α, maior deveria ser a amostra.

Foram examinadas as distribuições χ^2, com 3, 6 e 9 graus de liberdade. Observou-se que maior assimetria, $\chi^2_{(3)}$, implicava em pequena probabilidade de exclusão inferior (menor do que $\frac{\alpha}{2}$) e grande probabilidade de exclusão superior (pouco maior do que $\frac{\alpha}{2}$) para ambos intervalos de predição. A medida que a assimetria era suavizada, maiores graus de liberdade, essas duas probabilidades se aproximavam de $\frac{\alpha}{2}$. No entanto, para amostras pequenas, a soma das duas probabilidades de exclusão não atinge α. Portanto, intervalos maiores que o necessário são obtidos.

Para erros segundo distribuição $t - \textit{Student}$, com 3 graus de liberdade, o intervalo proposto apresentou probabilidades de exclusão próximas a $\frac{\alpha}{2}$. E para a distribuição lognormal, a probabilidade de exclusão inferior se apresentou muito pequena e a probabilidade de exclusão superior próxima a $\frac{\alpha}{2}$, um pouco menor.

Em todos estes casos, onde a distribuição do erro não é normal, o intervalo de predição proposto apresentou melhor desempenho do que o intervalo usual. Porém, para erros com uma distribuição com assimetria acentuada, como a lognormal, o intervalo de predição proposto não parece ser a solução.

Cabe observar que toda simulação elaborada por Schmoyer visava a obtenção de intervalos de predição para y, dado $x = 0$.

Neste trabalho foi simulada a seguinte situação: amostra de 30 pares (x, y), onde $y = 3 + x + \epsilon$, tomando x valores entre 1 e 5,9, e ϵ com distribuição exponencial, de parâmetro (1), padronizada para esperança zero e variância um. Foram feitas 30 repetições. Duas destas amostras são ilustradas no capítulo 2, figuras (2-7) e
(2-8), cujos gráficos de normalidade indicam assimetria. Nosso objetivo é verificar as probabilidades de exclusão dos intervalos de predição de probabilidade 0,9 para y, em três situações: $x = 1$, $x = 2,6$ e $x = 4,8$ (um valor central e dois valores extremos de x). Notamos que as probabilidades de exclusão, dos intervalos usual e proposto, quase não variam para os diferentes valores de x. Apresentaremos as probabilidades de

exclusão para $x = 2,6$. No referente ao intervalo usual, a probabilidade de exclusão do limite inferior foi observada no intervalo 0,014 $\pm 0,014$ e a probabilidade de exclusão do limite superior em 0,021$\pm$0,023. Quanto ao intervalo proposto, a probabilidade de exclusão do limite inferior foi observada no intervalo 0,052$\pm$0,019 e a probabilidade de exclusão do limite superior em 0,007$\pm$0,011. Podemos destacar que os intervalos usuais são mais largos que os propostos, com probabilidades de exclusão menores que $\frac{\alpha}{2}$. As probabilidades de exclusão, para o intervalo proposto, são aproximadamente $\frac{\alpha}{2}$ para o limite inferior e menores que $\frac{\alpha}{2}$ para o limite superior.

Apêndice

Proposição 1

Para toda transformação, função real estritamente monótona de uma variável aleatória positiva, com probabilidade um, a transformação da mediana da distribuição original é a mediana da distribuição transformada.

Prova:

Seja g uma função estritamente monótona. Seja y uma variável aleatória com espaço amostral $\chi_y \subset R^+$.

$$\{y \in \chi_y \,/\, y < y_{med}\} \Leftrightarrow \{y \in \chi_y \,/\, g(y) < g(y_{med})\},$$

$$\{y \in \chi_y \,/\, y \leq y_{med}\} \Leftrightarrow \{y \in \chi_y \,/\, g(y) \leq g(y_{med})\},$$

onde y_{med} é a mediana da variável y. Como

$$P[y < y_{med}] \leq \frac{1}{2} \leq P[y \leq y_{med}],$$

pela definição de mediana, com a igualdade dos eventos acima temos

$$P[g(y) < g(y_{med})] = P[y < y_{med}] \leq \frac{1}{2} \leq P[y \leq y_{med}] = P[g(y) \leq g(y_{med})].$$

Logo, $g(y_{med})$ é a mediana da variável $g(y)$.

De maneira semelhante, se g é decrescente, então os seguintes eventos são iguais:

$$\{y \in \chi_y \,/\, y < y_{med}\} \Leftrightarrow \{y \in \chi_y \,/\, g(y) > g(y_{med})\},$$

$$\{y \in \chi_y \,/\, y \leq y_{med}\} \Leftrightarrow \{y \in \chi_y \,/\, g(y) \geq g(y_{med})\}.$$

Assim,

$$P[g(y) > g(y_{med})] = P[y < y_{med}] \leq \frac{1}{2} \leq P[y \leq y_{med}] = P[g(y) \geq g(y_{med})].$$

Logo, $g(y_{med})$ é a mediana da variável $g(y)$. □

Resultado 1

Sejam $\underset{\sim}{y}$ do modelo Gauss-Markov, $\underset{\sim}{\hat{\beta}}$ estimador, não viciado, de quadrados mínimos de $\underset{\sim}{\beta}$, então

$$Var[\underset{\sim}{\hat{\beta}}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

Nota:

Se $p = 1$, ou seja, tendo só uma variável regressora, $\underset{\sim}{\beta}^T = (\beta_0, \beta_1)$, $\underset{\sim}{x_i}^T = (1, x_{i1})$. Logo,

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 \end{bmatrix}$$

e

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{n\sum_{i=1}^n x_{i1}^2 - (n\,\bar{x})^2} \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & -\sum_{i=1}^n x_{i1} \\ -\sum_{i=1}^n x_{i1} & n \end{bmatrix}$$

$$= \frac{1}{\sum_{i=1}^n x_{i1}^2 - n\,\bar{x}^2} \begin{bmatrix} \frac{\sum_{i=1}^n x_{i1}^2}{n} & -\frac{\sum_{i=1}^n x_{i1}}{n} \\ -\frac{\sum_{i=1}^n x_{i1}}{n} & 1 \end{bmatrix}$$

$$= \frac{1}{ns^2} \begin{bmatrix} s^2 + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

onde $s^2 = \frac{1}{n}\sum_{i=1}^n x_{i1}^2 - \bar{x}^2$,

$$= \begin{bmatrix} \frac{1}{ns^2}(s^2 + \bar{x}^2) & -\frac{1}{ns^2}\bar{x} \\ -\frac{1}{ns^2}\bar{x} & \frac{1}{ns^2} \end{bmatrix}.$$

Portanto,

$$Var[\underset{\sim}{\hat{\beta}}] = \begin{bmatrix} \frac{\sigma^2}{ns^2}(s^2 + \bar{x}^2) & -\frac{\sigma^2}{ns^2}\bar{x} \\ -\frac{\sigma^2}{ns^2}\bar{x} & \frac{\sigma^2}{ns^2} \end{bmatrix}.$$

Resultado 2

Sob o modelo dado para o resultado 1, seja $r_i = \underset{\sim}{x}_i^T (\mathbf{X}^T\mathbf{X})^{-1} \underset{\sim}{x}_i$ para $i = 1, 2, ..., n$, então

$$0 \leq r_i \leq 1.$$

Prova:

Para o modelo Gauss-Markov temos que

$$\underset{\sim}{y} = \mathbf{X}\,\underset{\sim}{\beta} + \underset{\sim}{\epsilon},$$

onde

$$E[\underset{\sim}{\epsilon}] = \underset{\sim}{0} \qquad e \qquad Var[\underset{\sim}{\epsilon}] = \sigma^2\mathbf{I_n}.$$

Logo, o vetor de resíduos é

$$\underset{\sim}{\hat{\epsilon}} = \underset{\sim}{y} - \underset{\sim}{\hat{y}},$$

e tem esperança $\underset{\sim}{0}$ com matriz de variância-covariância $\sigma^2(\mathbf{I_n} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)$. Para $i = 1, 2, ..., n$, $Var[\epsilon_i] = \sigma^2(1 - r_i)$. Consequentemente, $r_i \leq 1$, e sendo a matriz $\mathbf{X}$ de posto coluna completo, então $(\mathbf{X}^T\mathbf{X})^{-1}$ é definida positiva, logo $r_i \geq 0$. Portanto, $0 \leq r_i \leq 1$. □

Nota:

Para o modelo de regressão com uma variável regressora,

$$r_i = \frac{1}{n} + \frac{(x_{i1} - \bar{x})^2}{\sum_{j=1}^{n}(x_{j1} - \bar{x})^2}, \quad i = 1, 2, ..., n.$$

Vemos que se existe i, entre 1 e n, tal que $x_{i1} \neq x_{j1}$, com $j \neq i$, $j = 1, 2, ...n$, obtemos

$$0 \leq \frac{(x_{i1} - \bar{x})^2}{\sum_{j=1}^{n}(x_{j1} - \bar{x})^2},$$

então

$$\frac{1}{n} \leq \frac{1}{n} + \frac{(x_{i1} - \bar{x})^2}{\sum_{j=1}^{n}(x_{j1} - \bar{x})^2} = r_i. \tag{.1}$$

Temos também que

$$\frac{(x_{i1} - \bar{x})^2}{\sum_{j=1}^{n}(x_{j1} - \bar{x})^2} = \frac{[(n-1)x_{i1} - x_{11} - ... - x_{i-1,1} - x_{i+1,1-...} - x_{n1}]^2}{[(n-1)x_{11} - x_{21} - ... - x_{n1}]^2 + ... + [(n-1)x_{n1} - x_{11} - ... - x_{n-1,1}]^2}.$$

Para o caso em que $x_{11} \neq x_{21}$ e $x_{21} = x_{31} = ... = x_{n1}$, e para $i = 1$,

$$\frac{(x_{11} - \bar{x})^2}{\sum_{j=1}^{n}(x_{j1} - \bar{x})^2} = \frac{[(n-1)x_{11} - (n-1)x_{21}]^2}{[(n-1)x_{11} - (n-1)x_{21}]^2 + (n-1)[x_{21} - x_{11}]^2}$$

$$= \frac{(n-1)^2[x_{11} - x_{21}]^2}{(n-1)^2[x_{11} - x_{21}]^2 + (n-1)[x_{21} - x_{11}]^2}$$

$$= \frac{(n-1)^2}{(n-1)^2 + (n-1)}$$

$$= \frac{n-1}{n}, \tag{.2}$$

e para $i \neq 1$, por exemplo $i = 2$,

$$\frac{(x_{21} - \bar{x})^2}{\sum_{j=1}^{n}(x_{j1} - \bar{x})^2} = \frac{[(n-1)x_{21} - x_{11} - (n-2)x_{21}]^2}{[(n-1)x_{11} - (n-1)x_{21}]^2 + (n-1)[x_{21} - x_{11}]^2}$$

$$= \frac{[x_{11} - x_{21}]^2}{(n-1)^2[x_{21} - x_{11}]^2 + [x_{21} - x_{11}]^2}$$

$$= \frac{1}{(n-1)^2 + (n-1)}$$

$$= \frac{1}{n(n-1)}. \qquad (.3)$$

Portanto, das equações (.1), (.2) e (.3), para todo $i = 1, 2, ..., n$,

$$\frac{1}{n} \leq \frac{1}{n} + \frac{(x_{i1} - \bar{x})^2}{\sum_{j=1}^{n}(x_{j1} - \bar{x})^2} \leq 1.$$

Resultado 3

Sob o modelo Gauss-Markov normal, sendo $\hat{\sigma}^2_{(1)}$ estimador de quadrados mínimos de σ^2 e $\hat{\sigma}^2_{(2)}$ estimador de máxima verossimilhança de σ^2, então $\frac{(n-p-1)\hat{\sigma}^2_{(1)}}{\sigma^2}$ e $\frac{n\hat{\sigma}^2_{(2)}}{\sigma^2}$ tem distribuição $\chi^2_{(n-p-1)}$. Logo,

$$Var[\hat{\sigma}^2_{(1)}] = \frac{2\sigma^4}{n-p-1} \quad \text{e} \quad Var[\hat{\sigma}^2_{(2)}] = \frac{2(n-p-1)\sigma^4}{n^2}.$$

Lema 1

Sejam $\underset{\sim}{y}$ do modelo Gauss-Markov normal, $\hat{\underset{\sim}{\beta}}$ e $\hat{\sigma}^2$ estimadores de quadrados mínimos de $\underset{\sim}{\beta}$ e σ^2, respectivamente. Seja $\underset{\sim}{\lambda} \neq \underset{\sim}{0}$, com $\underset{\sim}{\lambda}^T \underset{\sim}{\beta}$ função estimável, então

$$\frac{\underset{\sim}{\lambda}^T \hat{\underset{\sim}{\beta}} - \underset{\sim}{\lambda}^T \underset{\sim}{\beta}}{\sqrt{\hat{\sigma}^2 \; \underset{\sim}{\lambda}^T \; (\mathbf{X}^T\mathbf{X})^{-1} \; \underset{\sim}{\lambda}}} \sim t_{n-p-1}.$$

Referências Bibliográficas

[1] Andrews, H. P., Snee, R. D. e Sarner, M. H. (1980). Graphical Display of means. The American Statistician, 34, 195-199.

[2] Arnold, Steven F. (1981). The Theory of Linear Models and Multivariate Analysis, John Wiley & Sons, New York.

[3] Bickel, Peter J. e Doksum, Kjell A. (1981). An Analysis of transformations Revisited. Journal of the American Statistical Association, 76, 296-311.

[4] Box, G. E. P. e Cox, D. R. (1964). An Analysis of Transformations. Journal of the Royal Statistical Society, Ser. B, 26, 211-252.

[5] Box, G. E. P. e Cox, D. R. (1982). An Analysis of Transformations Revisited, Rebuted. Journal of the American Statistical Association, 77, 209-210.

[6] Buxton, J. R. (1991). Some Comments on the Use of Response Variable Transformations in Empirical Modelling. Applied Statistics, 40, 391-400.

[7] Carroll, R. J. e Ruppert, David (1981). On Prediction and the Power Transformation Family. Biometrika, 68, 609-615.

[8] Carroll, R. J. e Ruppert, David (1984). Comentário em 'The Analysis of Transformed Data' por D. V. Hinkley e G. Runger. Journal of the American Statistical Association, 79, 312-313.

[9] Carroll, R. J. e Ruppert, David (1988). Transformation and Weighting in Regression, Chapman and Hall, New York.

[10] Cox, D. R. e Snell, E. J. (1981). Applied Statistics, Principles and Examples, Chapman and Hall, New York.

[11] Dempster, A. P. e Ryan, Louise M. (1985). Weighted Normal Plots. Journal of the American Statistical Association, 80, 845-850.

[12] Draper, N. R. e Smith, H. (1981). Applied Regression Analysis, Second Edition, John Wiley & Sons, New York.

[13] Duan, Naihua (1983). Smearing Estimate: A Nonparametric Retransformation Method. Journal of the American Statistical Association, 78, 605-610.

[14] Federer, Walter T. (1973). Statistics and Society, Marcel Dekker, New York.

[15] Gan, F.F. , Koehler, Kenneth J. e Thompson, John C. (1991). Probability Plots and Distribution Curves for Assessing the Fit of Probability Models. The American Statistician, 45, 14-21.

[16] Gianola, D. e Hammond, K. (1990). Advances in Statistical Methods for Genetic Improvement of Livestock, Springer-Verlag, Berlin.

[17] Hinkley, D. V. e Runger, G. (1984). The Analysis of Transformed Data. Journal of the American Statistical Association, 79, 302-320.

[18] Lange, Nicholas e Ryan, Louise (1989). Assessing Normality in Random Effects Models. The Annals of Statistics, 17, 624-642.

[19] Madansky, Albert (1988). Prescriptions for Working Statisticians, Springer-Verlag, New York.

[20] Miller, Don M. (1984). Reducing Transformation Bias in Curve Fitting. The American Statistician, 38, 124-126.

[21] Pierce, Donald A. (1982). The Asymptotic Effect of Substituting Estimators for Parameters in Certain Types of Statistics. The Annals of Statistics, 10, 475-478.

[22] Raktoe, B. L. e Hubert, J.J. (1979). Basic Applied Statistics, Marcel Dekker, New York.

[23] Royston, Patrick (1991). Constructing Time-especific Reference Ranges. Statistics in Medicine, 10, 675-690.

[24] Rubin, Donald B. (1984). Comentário em 'The Analysis of Transformed Data' por D. V. Hinkley e G. Runger. Journal of the American Statistical Association, 79, 309-312.

[25] Sakia, R. M. (1990). Retransformation Bias: A look at the Box-Cox Transformation to Linear Balanced Mixed ANOVA Models. Metrika, 37, 345-351.

[26] Sakia, R. M. (1992). The Box-Cox Transformation Technique: A Review. The Statistician, 41, 169-178.

[27] SAS Institute Inc., SAS Guide to Macro Processing, Version 6, Second Edition, Cary, NC: SAS Institute Inc., 1990.

[28] SAS Institute Inc., SAS/IML Software: Usage and Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc., 1989.

[29] SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc., 1989.

[30] Schmoyer, Richard L. (1992). Asymptotically Valid Prediction Intervals for Linear Models. Technometrics, 34, 399-408.

[31] Scott, Alastair e Wild, Chris (1991). Transformations and R^2. The American Statistician, 45, 127-129.

[32] Seber, G. A. F. (1977). Linear Regression Analysis, John Wiley & Sons, New York.

[33] Seber, G. A. e Wild, C. J. (1989). Nonlinear Regression, John Wiley & Sons, New York.

[34] Sen, Ashish e Srivastava, Muni (1990). Regression Analysis: Theory, Methods, and Applications, Springer-Verlag, New York.

[35] Shumway, R. H., Azari, A. S. e Johnson, P. (1989). Estimating Mean Concentrations under Transformation for Environmental Data with Detection Limits. Technometrics, 31, 347-356.

[36] Snee, Ronald D. (1986). An Alternative Approach to Fitting Models when Reespression of the Response is Useful. Journal of Quality Technology, 18, 211-225.

[37] Taylor, Jeremy M. G. (1986). The Retransformed Mean after a Fitted Power Transformation. Journal of the American Statistical Association, 81, 114-118.

[38] Wooldridge, Jeffrey M. (1992). Some Alternatives to the Box-Cox Regression Model. International Economic Review, 33, 935-955.