



UNIVERSIDADE FEDERAL DE LAVRAS


**VARIÂNCIAS DO PONTO CRÍTICO DE  
EQUAÇÕES DE REGRESSÃO QUADRÁTICA**

**CEILE CRISTINA FERREIRA NUNES**

**2002**

CEILE CRISTINA FERREIRA NUNES

**VARIÂNCIAS DO PONTO CRÍTICO DE EQUAÇÕES DE  
REGRESSÃO QUADRÁTICA**



Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

Orientador

Prof. Augusto Ramalho de Moraes

LAVRAS  
MINAS GERAIS - BRASIL  
2002

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Nunes, Ceile Cristina Ferreira

Variâncias do ponto crítico de equações de regressão quadrática / Ceile Cristina  
Ferreira Nunes. -- Lavras : UFLA, 2002.

73 p. : il.

Orientador: Augusto Ramalho de Moraes.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Estatística. 2. Análise de variância. 3. Ponto crítico. 4. Regressão quadrática.  
5. Experimentação agropecuária. I. Universidade Federal de Lavras. II. Título.

CDD-519.537

**CEILE CRISTINA FERREIRA NUNES**

**VARIÂNCIAS DO PONTO CRÍTICO DE EQUAÇÕES DE  
REGRESSÃO QUADRÁTICA**

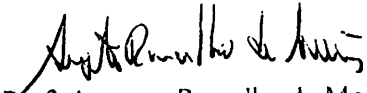
Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

APROVADA em 22 de fevereiro de 2002.

Profa. Thelma Sáfyadi UFLA

Prof. Joel Augusto Muniz UFLA

Pesq. Paulo Tácito Gontijo Guimarães EPAMIG

  
Prof. Augusto Ramalho de Moraes  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS - BRASIL

*À Nossa Senhora Aparecida,*

**OFEREÇO.**

*Aos meus pais, José Nunes e Luzia Ferreira,  
pelos exemplos de vida e fé;  
Aos meus irmãos, Celma Nunes e José Vicente,  
pelo carinho e amizade;  
Ao meu querido Alex,  
pelo companheirismo, equilíbrio e força em todos os momentos,*

**DEDICO.**

## AGRADECIMENTOS

A DEUS, que tudo nos proporciona.

À Universidade Federal de Lavras pela oportunidade de realização do curso.

À CAPES pela bolsa concedida.

À equipe de orientadores deste trabalho, pela orientação acadêmica.

Às duas grandes professoras da minha vida: Eliza Soares Pereira, que ensinou-me os primeiros passos rumo a uma vida estudantil de sucesso e Tânia Valquíria Menegon, que ensinou-me o gosto pela matemática.

Às amigas, Milena Spegiorim Moreno Gomes e Elizete Antunes Teixeira, que muito me incentivaram nos momentos difíceis.

À amiga e quase “mãe” D. Maria Antônia, que com sua simplicidade, força e carinho mostrou-me que há mais o que vencer na vida do que apenas distância e saudade.

Aos meus avós, que sempre torceram e rezaram por mim.

Ao João, meu querido afilhado, que com ternura de criança, trouxe-me paz quando mais precisei.

Aos que foram companheiros no curso de pós-graduação, que, de colegas, se transformaram em amigos.

A todos que, com consideração e carinho, estiveram presentes na defesa.

A todos que, mesmo sem o saberem, auxiliaram a levar a termo o projeto de pós-graduação.

## LISTA DE TABELAS

TABELA 1	Esquema de análise de variância para regressão.....	17
TABELA 2	Esquema de análise de variância para um modelo de regressão polinomial quadrático.....	43
TABELA 3	Valores médios dos coeficientes de $\hat{b}$ e $\hat{c}$ e respectivas variâncias, em função das diferentes variâncias reais utilizadas.....	54
TABELA 4	Valores médios do ponto crítico ( $\bar{\hat{x}}$ ) e variâncias do ponto crítico, calculadas pelos métodos: usual $Var_1(\hat{x})$ , fórmula deduzida por D'Aulísio (1976) $Var_2(\hat{x})$ , expressão dada por Mood, Graybill e Boes (1974) $Var_3(\hat{x})$ .....	55
TABELA 5	Valores médios dos coeficientes de assimetria, curtose e respectivas estatísticas do teste "t", em função das diferentes variâncias.....	56
TABELA 6	Valores médios do ponto crítico e intervalos de confiança para o ponto crítico, considerando as três variâncias, em função das variâncias teóricas.....	66

## LISTA DE FIGURAS

- FIGURA 1 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 0,1$  e valor médio de  $\bar{x} = 0,4560$  ..... 58
- FIGURA 2 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 0,5$  e valor médio de  $\bar{x} = 0,4564$  ..... 59
- FIGURA 3 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 1,0$  e valor médio de  $\bar{x} = 0,4569$  ..... 60
- FIGURA 4 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 5,0$  e valor médio de  $\bar{x} = 0,4616$  ..... 61
- FIGURA 5 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 10,0$  e valor médio de  $\bar{x} = 0,4680$  ..... 62
- FIGURA 6 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 15,0$  e valor médio de  $\bar{x} = 0,4751$  ..... 63
- FIGURA 7 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 20,0$  e valor médio de  $\bar{x} = 0,4835$  ..... 64
- FIGURA 8 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 50,0$  e valor médio de  $\bar{x} = 0,5586$  ..... 65



## RESUMO

NUNES, Ceile Cristina Ferreira. **Variâncias do ponto crítico de equações de regressão quadrática**. LAVRAS: UFLA, 2002. 73p. (Dissertação – Mestre)\*

O presente trabalho teve por objetivo a determinação de variâncias para o estudo do ponto crítico de uma equação de regressão de segundo grau, em situações experimentais com diferentes variâncias, por meio de simulação Monte Carlo. Em muitos estudos, teóricos ou aplicados, o pesquisador depara-se com o problema envolvendo quociente entre variáveis aleatórias e, principalmente, entre variáveis normais. Como exemplo, aquelas que surgem em pesquisas de dose econômica de nutrientes em experimentos de adubação, de compactação de solos e em outros problemas em que há interesse na variável aleatória  $\hat{x} = \hat{b}/(-2\hat{c})$ , estimador do ponto crítico na regressão  $\hat{y} = \hat{a} + \hat{b}x + \hat{c}x^2$ . Para estudar a distribuição do ponto crítico de uma equação de regressão quadrática, foram utilizados dados de produção de algodão de 536 ensaios, ajustando-se um modelo quadrático. A estimação dos parâmetros foi feita pelo método dos quadrados mínimos ordinários. A partir dessas estimativas implementou-se, por meio do software MATLAB®, uma rotina para simulação de duas séries, com cinco mil erros aleatórios de distribuição normal de média zero, relativos a cada uma das variâncias consideradas teóricas:  $\sigma^2=0,1; 0,5; 1; 5; 10; 15; 20$  e 50. As estimativas da variância do ponto crítico foram obtidas por meio de três métodos: (a) fórmula comum do cálculo de variâncias; (b) fórmula obtida através da diferenciação do estimador do ponto crítico e (c) fórmula demonstrada para o cálculo da variância de uma razão, considerando-se a covariância entre  $\hat{b}$  e  $\hat{c}$ . Os resultados obtidos para as estatísticas médias dos coeficientes de regressão  $\hat{b}$  e  $\hat{c}$ , bem como suas respectivas variâncias em função das diversas variâncias teóricas ( $\sigma^2$ ) adotadas, mostraram que esses valores teóricos estão próximos aos reais. Ainda, ocorre uma tendência de que, com o aumento da variância teórica, esses valores aumentem. Pôde-se concluir que a variância do ponto crítico calculada, usando-se a expressão que leva em consideração a covariância entre  $\hat{b}$  e  $\hat{c}$ , apresenta resultados mais satisfatórios e que não segue uma distribuição normal, pois apresenta uma distribuição de frequência com assimetria positiva e formato leptocúrtico.

---

\* Comitê Orientador: Prof. Augusto Ramalho de Moraes-UFLA (Orientador), Prof. Thelma Sáfydi-UFLA e Prof. Joel Augusto Muniz-UFLA.

## ABSTRACT

NUNES, Ceile Cristina Ferreira. **Variance of the critical point of quadratic regression equations.** LAVRAS: UFLA, 2002. 73p. (Dissertation – Master in Statistics and Agricultural Experimentation)\*

The present work was intended to determine variances for the study of the critical point of a second-degree regression equation under experimental situations with different variances by means of Monte Carlo simulation. In a number of studies, whether theoretical or applied, the researcher faces the problem, involving quotient among random variables and mainly among normal variables. As an example, those ones which appear in research of economic dose of nutrients in fertilization experiments, in soil compaction and in other problems in which there are interests in the random variable  $\hat{x} = \hat{b}/(-2\hat{c})$ , estimator of the critic point in the regression  $\hat{y} = \hat{a} + \hat{b}x + \hat{c}x^2$ . To study the distribution of the critical point of a quadratic regression equation, data of five hundred and thirty-six trials in cotton yield by adjusting a quadratic model were utilized. From these estimates, a routine for the simulating of two sets with five thousand random errors of normal distribution of zero mean relative to each of the variances considered theoretical:  $\sigma^2=0,1; 0,5; 1; 5; 10; 15; 20$  and 50 was implemented by means of the MATLAB® software. The estimates of the variance of the critical point were obtained through three methods: (a) common formula of the variance calculation; (b) formula obtained through the differentiation of the critical point estimator and (c) formula demonstrated for the variance calculation of a ratio, by taking into consideration the covariance between  $\hat{b}$  e  $\hat{c}$ . The results obtained for the average statistics of the regression between  $\hat{b}$  e  $\hat{c}$ , as well as its respective variances in terms of the several theoretical residual variances ( $\sigma^2$ ) adopted, show that those theoretical values are close to the real ones. Still, a trend occurs that with the increase of the theoretical variance those values increase. It can conclude that the critical point variance calculated by using the expression, which takes into consideration the covariance between  $\hat{b}$  e  $\hat{c}$  presents more satisfactory results and that does not follow a normal distribution, for it presents a frequency distribution with positive asymmetry and leptokurtic shape.

---

\* Guidance Committee: Professor Augusto Ramalho de Morais-UFLA (Major Professor), Professor Thelma Sáfaci-UFLA and Professor Joel Augusto Muniz-UFLA.

## SUMÁRIO

	<b>Página</b>
RESUMO.....	i
ABSTRACT.....	ii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 Estudos da distribuição de um quociente.....	3
2.2 Distribuição normal.....	8
2.3 Modelos de regressão.....	10
2.4 Regressão linear múltipla.....	12
2.4.1 Análise de variância da regressão linear múltipla.....	16
2.4.2 Matriz de variâncias e covariâncias.....	18
2.5 Regressão polinomial.....	20
2.6 Estimando o máximo ou o mínimo de uma equação de regressão quadrática.....	21
2.7 Estimativa da variância do ponto crítico.....	21
2.8 Coeficientes de assimetria e curtose.....	23
2.9 Intervalos de confiança.....	24
2.10 Desenvolvimentos assintóticos – série de Taylor.....	28
2.10.1 Série de Taylor para funções de uma variável.....	29
2.10.2 Série de Taylor para funções de duas ou mais variáveis.....	30
3 MATERIAL E MÉTODOS.....	35

3.1 Material.....	35
3.2 Métodos.....	35
<b>4 RESULTADOS E DISCUSSÃO.....</b>	<b>38</b>
4.1 Modelo de regressão polinomial quadrático.....	38
4.2 Análise de variância.....	42
4.3 Obtenção do ponto crítico.....	45
4.4 Esperança e variância para o quociente de variáveis aleatórias.....	46
4.4.1 Esperança do quociente de duas variáveis aleatórias.....	46
4.4.2 Variância do quociente de duas variáveis aleatórias.....	49
4.5 Análise dos resultados simulados.....	53
4.6 Distribuição de frequência dos valores do ponto crítico.....	57
4.7 Intervalos de confiança para o ponto crítico.....	66
<b>5 CONCLUSÕES.....</b>	<b>67</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>68</b>
<b>ANEXO.....</b>	<b>71</b>

# 1 INTRODUÇÃO

A necessidade de comparar e testar tratamentos é uma prática geral na experimentação agropecuária. Inicialmente, a hipótese de igualdade de médias de tratamentos em uma análise de variância é avaliada com o uso do teste “F”. Quando ocorre a rejeição dessa hipótese, a significância desse teste, em relação a um valor fixado de probabilidade, permite a inferência de que existe pelo menos uma diferença entre os tratamentos, não informando claramente onde estão essas diferenças.

Dependendo da natureza dos tratamentos, vários procedimentos de comparações múltiplas podem ser aplicados, o que dará uma continuidade nas análises dos dados. Quando os tratamentos são qualitativos, recomenda-se a aplicação de algum tipo de procedimento de comparações múltiplas. No entanto, nos experimentos em que os tratamentos são quantitativos, como por exemplo: doses crescentes de adubos, ou de inseticidas, recomenda-se o ajuste de equações de regressão. Isso porque muitas vezes se justifica a existência de uma correspondência funcional, que relaciona os valores dos tratamentos aos dados observados, denominada equação de regressão.

Entre os vários modelos de regressão usados nas pesquisas, merece destaque o modelo de regressão quadrático. Esse destaque deve-se ao seu amplo uso, facilidade de cálculos e enorme adaptação na explicação e interpretação de fenômenos biológicos. Entre esses, destaca-se a aplicação de doses crescentes de um nutriente, em uma determinada cultura.

Após a obtenção da equação de regressão do segundo grau que descreve e/ou interpreta os dados observados em função dos níveis quantitativos de um determinado fator, existe ainda o interesse do pesquisador na determinação do ponto crítico e na sua interpretação. Normalmente, apenas calcula-se esse ponto crítico pontualmente, dizendo se é máximo ou mínimo e nenhuma inferência por

intervalo é feita. Isto por existir pouca informação sobre a natureza do ponto crítico, com referência especial à sua variância e distribuição. Este fato envolve o quociente de variáveis aleatórias.

Assim, em muitos estudos, teóricos ou aplicados, o pesquisador depara-se com o problema envolvendo quociente entre variáveis aleatórias e, principalmente, entre variáveis aleatórias normais. Como exemplo, aquelas que surgem em pesquisas de dose econômica de nutrientes em experimentos de adubação, em experimentos de compactação de solos e em outros problemas em que há interesse na variável aleatória  $\hat{x} = \frac{\hat{b}}{-2\hat{c}}$ , estimador do ponto crítico na regressão quadrática  $\hat{y}_i = \hat{a} + \hat{b}x_i + \hat{c}x_i^2$ .

O presente trabalho teve por objetivo a determinação de variâncias para o estimador do ponto crítico de uma equação de regressão de segundo grau, em situações experimentais com diferentes variâncias, por meio de simulação Monte Carlo.

## 2 REFERENCIAL TEÓRICO

Um importante e amplo campo de estudos sobre curvas e superfícies de resposta abriu-se logo após Mitscherlich ter apresentado, em 1908, a lei que leva seu nome. Desde então, vários pesquisadores têm proposto inúmeros meios e métodos para o estudo e aplicação das curvas de resposta que representam determinados fenômenos. Em particular, e principalmente, essas pesquisas têm voltado atenção ao estudo das respostas das diversas culturas aos adubos e nutrientes aplicados ao solo.

É inegável a existência de inúmeros estudos na pesquisa agropecuária, nos quais as variáveis observadas estão relacionadas com o uso de doses de fertilizantes, herbicidas, irrigação, entre tantos outros. Esses fatores são aplicados, em geral, em doses crescentes o que pode sugerir a existência de dependência entre os efeitos dos tratamentos. Assim, há interesse em relacionar as variáveis respostas com as doses (níveis) ou concentrações através do estudo dos modelos de regressão, estabelecendo uma correspondência funcional, denominada de equação de regressão, que relaciona os valores dos tratamentos ( $x$ ) aos dados observados ( $y$ ).

Quando se ajusta uma equação de regressão do segundo grau a um conjunto de dados, tem-se grande interesse na determinação do ponto crítico que por sua vez é um quociente de variáveis aleatórias.

### 2.1 Estudos da distribuição de um quociente

Estudos sobre a distribuição de um quociente de duas variáveis aleatórias são bastante escassos na literatura. Um dos primeiros trabalhos a esse respeito foi o de Merrill (1928) que estudou a distribuição do quociente:

$$I = \frac{\bar{y} + y}{\bar{x} + x},$$

sendo que ambos os membros possuem distribuições normais e referem-se à medidas de crânios humanos, para o qual:

$y$  denota a variável de um conjunto de medidas de média  $\bar{y}$ ,  
 $x$  um segundo conjunto de medidas de média  $\bar{x}$ .

O autor verificou que, no caso da existência de correlação entre  $x$  e  $y$  ser pequena e seus respectivos coeficientes de variação baixos, os quocientes obtidos são aproximadamente normais. Nos demais casos, esses quocientes fogem consideravelmente da normalidade.

Para o caso particular em que a variável no denominador apresenta baixo coeficiente de variação, Geary (1930) afirmou que a variável aleatória:

$$G = \frac{m_2 Q - m_1}{(\sigma_2^2 Q^2 - 2\rho\sigma_1\sigma_2 Q + \sigma_1^2)^{\frac{1}{2}}},$$

tem distribuição aproximadamente normal.

Note-se que  $G$  é função de cinco parâmetros:  $m_1$ ,  $m_2$ ,  $\sigma_1$ ,  $\sigma_2$  e  $\rho$ . Pode-se verificar que com  $m_1$  e  $m_2$  não nulos, a expressão de  $G$  é dada por:

$$G = \frac{Q - \frac{\sigma_1}{\sigma_2} CV_2}{CV_2 \left[ Q^2 - 2\rho \frac{\sigma_1}{\sigma_2} Q + \left( \frac{\sigma_1}{\sigma_2} \right)^2 \right]^{\frac{1}{2}}},$$

de sorte que, agora,  $G$  é função, apenas, de  $\rho$ ,  $\frac{\sigma_1}{\sigma_2}$ ,  $CV_1$  e  $CV_2$ .

Fieller (1932) estudou a distribuição do quociente  $v = \frac{y}{x}$ , no qual  $x$  e  $y$  são variáveis aleatórias normalmente distribuídas e representam,



respectivamente, medições de ossos parietais e temporais, observadas em 787 crânios humanos. O autor verificou que os quocientes obtidos se aproximaram da normalidade.

Kendall e Stuart (1977) estudaram a distribuição do quociente:

$$z = \frac{\delta_1}{\delta_2}$$

no qual:

$\delta_1$  e  $\delta_2$  são variáveis aleatórias, com  $\delta_1 \sim N(m_1, \delta_1^2)$  e  $\delta_2 \sim N(m_2, \delta_2^2)$  respectivamente. No caso de  $m_1 = m_2 = 0$  e  $\delta_1$  e  $\delta_2$  serem independentes entre si, demonstraram que  $z$  tem distribuição de Cauchy.

Uma aplicação importante em medicina, em que há interesse na determinação do intervalo de confiança para o parâmetro  $\gamma = -\frac{\alpha}{\beta}$ , sendo  $\alpha$  e  $\beta$  parâmetros da regressão linear  $E(y_i) = \alpha + \beta x_i$ , foi citada por Marsaglia (1965).

Este autor estudou a variável aleatória  $W = \frac{a+x}{b+y}$ , em que  $a$  e  $b$  são constantes não negativas e  $x$  e  $y$  são variáveis aleatórias independentes normais reduzidas. Propôs aproximação para o caso em que  $b+y > 0$ . Apresentou gráficos para determinados valores de  $a$  e  $b$ . Embora o autor tenha afirmado que os valores de  $a$  e  $b$  conduziram às ilustrações das formas possíveis da função de densidade de probabilidade da variável quociente, os casos cobertos são tais que correspondem a  $CV_1 \geq 0,5$  e  $CV_2 \geq 1$ , todos eles de nenhum interesse em aplicações práticas.

D'Aulísio (1976) estudou a distribuição do quociente:

$$x = \frac{-\hat{b}}{2\hat{c}},$$

que se refere ao ponto de máximo ou de mínimo da função de produção estimada:

$$\hat{y} = \hat{a} + \hat{b}P_1(x) + \hat{c}P_2(x),$$

no qual

$P_1(x)$  e  $P_2(x)$  são polinômios ortogonais de primeiro e segundo graus, respectivamente;

$\hat{a}$ ,  $\hat{b}$  e  $\hat{c}$  são os estimadores dos parâmetros;

$\hat{y}$  são os valores preditos (esperados).

Para tanto, foram gerados 16.000 dados de distribuição normal. Após ajustado, foram obtidos 8.000 valores para  $x$ . As variâncias utilizadas foram: 0,015625; 0,0625; 0,25; 1,00; 2,00; 4,00; 6,25 e 9,00. O autor verificou que a distribuição de  $x$  foge completamente da normalidade, exceto talvez para o valor mais baixo de  $\sigma^2$  estudado. Além disso, à medida que crescem os valores das variâncias teóricas estudadas, a distribuição tende a ser leptocúrtica. Isso mostra que, no caso das variâncias dos membros de um quociente serem pequenas, a distribuição desse quociente é aproximadamente normal.

Freitas (1978) estudou a distribuição dos erros experimentais associados ao ponto que determina a produção máxima dado por:

$$\sqrt{x} = \frac{-\hat{b}}{2\hat{c}},$$

para a função de produção no modelo com raiz quadrada:

$$y_i = a + bx_i^{\frac{1}{2}} + cx_i + e_i \quad i = 1, 2, \dots, n,$$

no qual:

$y_i$  representa a produção obtida na  $i$ -ésima dose, em quilogramas por hectare,

$x_i$  representa a  $i$ -ésima quantidade de nutriente (N, P<sub>2</sub>O<sub>5</sub> ou K<sub>2</sub>O) em kg/ha.

Consideraram-se os erros como aleatórios e independentes com distribuição normal de média zero e variância  $\sigma^2$ . Verificou-se que a distribuição do quociente  $\frac{-\hat{b}}{2\hat{c}}$  foi aproximadamente normal para as variâncias  $\sigma^2 = 5$  e  $\sigma^2 = 10$ . Já para as variâncias  $\sigma^2 = 15, 20, 25, 30, 35, 40, 45$  e  $50$ , a distribuição caracterizou-se por ser leptocúrtica e com assimetria positiva.

Chiarini (1983) estudou o caso geral da distribuição de probabilidades do quociente de variáveis aleatórias normais, abordando casos em que elas são correlacionadas e, também não correlacionadas. Obteve a função de densidade de probabilidade para o quociente:  $Q = \frac{x_1}{x_2}$ , em que  $x_1$  e  $x_2$  têm distribuição normal de médias e variâncias  $m_1, \sigma_1^2$  e  $m_2, \sigma_2^2$ , respectivamente. Onde foi conveniente, foram utilizados os coeficientes de variação de  $x_1$  e  $x_2$ ,  $CV_1 = \frac{\sigma_1}{m_1}$  e  $CV_2 = \frac{\sigma_2}{m_2}$ , com  $m_1$  e  $m_2$  não nulos. Após verificar propriedades que simplificavam a obtenção e apresentação da tabela para a distribuição do quociente, o autor concluiu que era suficiente a obtenção da tabela para variáveis independentes, de variâncias unitárias e coeficientes de variação (ou médias) positivos.

## 2.2 Distribuição normal

A distribuição normal é a mais importante distribuição de variável aleatória contínua. Ela apresenta grande aplicação nas pesquisas científicas e tecnológicas, tanto nos aspectos práticos quanto teóricos. A maior parte das inferências são realizadas considerando que os resíduos dos modelos adotados para as variáveis consideradas possuem distribuição normal. Essa é uma das pressuposições básicas da análise de variância. Nas várias pesquisas realizadas no campo agropecuário, as variáveis como produção de grãos, altura de plantas, peso de animais, entre inúmeras outras, são consideradas como obedecendo a uma distribuição normal.

De acordo com Mood, Graybill e Boes (1974), uma variável aleatória  $y$  apresenta distribuição normal com parâmetros  $\mu$  e  $\sigma^2$ , se sua função de densidade de probabilidade for dada por:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}$$

na qual, os parâmetros  $\mu$  e  $\sigma^2$  são a média e a variância.

O gráfico desta distribuição tem um formato campanular e, é simétrico em relação à média  $\mu$ . As medidas de posição, mediana e moda coincidem com a média.

Para a descrição de uma variável aleatória são utilizados os momentos relativos a cada distribuição, sendo mais importantes e usados os momentos até a quarta ordem (Hoel, Port e Stone, 1978; Roussas, 1973). O momento ordinário  $\mu_1$  é muito informativo, pois corresponde ao valor médio da variável  $y$ . Já o momento  $\mu_2$ , que corresponde à variância da variável  $y$ , fornece uma idéia com relação à maior ou menor dispersão dos valores de  $y$  em torno da média. O

momento  $\mu_3$  é usado para verificar se a distribuição de  $y$  é ou não simétrica em torno de seu valor médio. Este fato é realizado por meio do “coeficiente de assimetria”, que é representado por  $\alpha_3$  ou por  $\sqrt{\beta_1}$ , sendo este último o mais comum e é definido por:

$$\alpha_3 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{(\mu_2)^3}}$$

em que  $\mu_2$  e  $\mu_3$  são os momentos de ordem 2 e 3, respectivamente, centrados na média, podendo ser obtidos pelas integrais

$$\mu_2 = \int_{-\infty}^{+\infty} (y - \mu)^2 f(y) dy \quad \text{e} \quad \mu_3 = \int_{-\infty}^{+\infty} (y - \mu)^3 f(y) dy .$$

Para que uma distribuição seja perfeitamente simétrica, deve-se ter  $\sqrt{\beta_1} = 0$ ; se a distribuição é assimétrica à direita, então  $\sqrt{\beta_1} > 0$  e, se é assimétrica à esquerda, tem-se  $\sqrt{\beta_1} < 0$ .

Uma outra medida que pode ser usada para verificar a natureza de uma distribuição é o “coeficiente de curtose”. Ele é representado por  $\alpha_4$  ou  $\sqrt{\beta_2}$ , com o qual pode-se medir o grau de achatamento de uma distribuição, tendo a distribuição normal como referência. O coeficiente de curtose é definido por

$$\alpha_4 = \sqrt{\beta_2} = \frac{\mu_4}{(\mu_2)^2}$$

em que  $\mu_4$  é o momento de ordem 4 centrado para a média, o qual pode ser obtido pela resolução da integral

$$\mu_4 = \int_{-\infty}^{+\infty} (y - \mu)^4 f(y) dy .$$

Considera-se que valores de  $\beta_2$  maiores que 3 ( $\beta_2 > 3$ ) indicam uma distribuição de probabilidade “afilada” em relação à normal, que é chamada de

leptocúrtica. Valores de  $\beta_2$  menores que 3 ( $\beta_2 < 3$ ), indicam um formato “achatado” na função de densidade, chamada de platicúrtica. Se  $\beta_2$  é igual a 3 ( $\beta_2 = 3$ ), tem-se um formato com grau de achatamento padrão em relação à distribuição normal e, nesse caso, a distribuição é chamada de mesocúrtica. Em muitas aplicações práticas é comum a subtração da constante 3, para facilidade de referência, que, neste caso, é zero.

### 2.3 Modelos de regressão

Historicamente, segundo Neter, Wasserman e Kutner (1990), a análise de regressão foi primeiramente desenvolvida por Francis Galton na segunda metade do século XIX. Galton tinha estudado a relação entre alturas de pais e filhos e notou que as alturas de filhos de pais altos e de baixos apareceram “invertidas” ou “regressivas” para a média do grupo. Ele considerou esta tendência como sendo uma regressão “mediocre”. Galton desenvolveu uma descrição matemática para esta tendência, o que constituiu o modelo precursor dos modelos de regressão de hoje.

O termo “regressão”, oriundo daquela época, persiste até hoje para descrever relações estatísticas entre variáveis.

Draper e Smith (1998) classificaram os modelos de regressão em relação aos seus parâmetros, em lineares, linearizáveis e não-lineares, a saber:

a) modelos lineares: aqueles que são lineares em relação aos parâmetros, isto é:

$$\frac{\partial}{\partial \theta_j} f_i(x, \theta) = g(x)$$

para  $i=1,2,\dots,n$  e  $j=1,2,\dots,p$ ;

sendo que  $n$  é o número total de observações e  $p$  é o número de parâmetros do modelo;

b) modelos linearizáveis: aqueles que podem ser transformados em lineares, por meio de alguma transformação. Tem-se o modelo:

$$y = a^x \cdot e$$

no qual o erro é dito multiplicativo. Aplicando-se logaritmo à igualdade, tem-se:

$$\ln y = \ln(a^x \cdot e)$$

$$\ln y = \ln a^x + \ln e$$

$$\ln y = x \ln a + \ln e$$

considerando  $z = \ln y$ ;  $b = \ln a$ ,  $e^* = \ln e$  o modelo fica:

$$z = bx + e^*$$

que é linear, pois

$$\frac{\partial f}{\partial b} = x = g(x)$$

logo, o modelo  $y = a^x \cdot e$  é dito linearizável;

c) modelos não lineares: aqueles que são baseados em considerações teóricas inerentes ao fenômeno que se tem interesse em estudar. Não se enquadram nos casos a) e b). Considera-se o modelo:

$$y = a^x + e$$

em que o erro é dito aditivo e não existe transformação capaz de tornar o modelo linear. Verifica-se que:

$$\frac{\partial f}{\partial a} = xa^{x-1} = g(x, a),$$

portanto, o modelo é dito não linear, pois a derivada em relação ao parâmetro é dependente do parâmetro.

Dentre os modelos de regressão, merece destaque o modelo de regressão múltipla. Ele possibilita a explicação de uma variável dependente, como a produção, em função de várias outras variáveis independentes, por exemplo doses de nitrogênio e fósforo.

## 2.4 Regressão linear múltipla

Um modelo de regressão linear múltipla, conforme Draper e Smith (1998) e Hoffmann e Vieira (1998), pode ser expresso como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i,$$

no qual

$y_i$  representa os valores observados da variável dependente (ou variável resposta);

$\beta_0, \beta_1, \dots, \beta_{p-1}$  são os parâmetros a serem estimados;

$x_{i1}, x_{i2}, \dots, x_{i,p-1}$  representam os valores das variáveis independentes;

$\varepsilon_i$  são os erros experimentais relacionados com os valores observados  $y_i$  que, em geral, são considerados independentes e normalmente distribuídos com média zero e variância constante.

Em termos matriciais, pode-se escrevê-lo como:

$$Y = X\beta + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & & & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & \cdots & x_{n,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

sendo:

$Y$  um vetor de respostas, de dimensões  $n \times 1$ ;

$X$  uma matriz de constantes conhecidas, denominada matriz do planejamento, de dimensões  $n \times p$ ;

$\beta$  um vetor de parâmetros, de dimensões  $p \times 1$ ;

$\varepsilon$  um vetor de erros aleatórios independentes e normalmente distribuídos, de dimensões  $n \times 1$ .



Existem vários métodos que podem ser usados para a estimação do vetor de parâmetros  $\beta$ . Entre esses, os mais comumente empregados são o método dos quadrados mínimos e o da máxima verossimilhança. Esses métodos conduzem aos mesmos estimadores em presença de normalidade, segundo Hoffmann e Vieira (1998).

De acordo com as pressuposições assumidas quanto aos erros, podem existir variações do método de estimação dos quadrados mínimos para o modelo de regressão linear. Assim obtêm-se estimadores não-tendenciosos de variância mínima.

Draper e Smith (1998) distinguiram a caracterização da regressão em função das suposições do vetor de erros  $\varepsilon$  da seguinte maneira:

a) modelos ordinários: aqueles cuja estrutura dos erros não viola nenhuma das pressuposições. Pode ser escrito de forma mais eficiente como  $\varepsilon \sim N(\mu, I\sigma^2)$ ;

b) modelos ponderados: são aqueles cuja estrutura dos erros viola a pressuposição de homogeneidade de variâncias. Nesse caso, diz-se que os erros são heterocedásticos. Escreve-se  $\varepsilon \sim N(\mu, D\sigma^2)$ , em que  $D$  é uma matriz diagonal, positiva definida, que pondera a variância  $\sigma^2$ ;

c) modelos generalizados: são aqueles cuja estrutura dos erros viola a pressuposição de independência dos erros e possivelmente a de homogeneidade de variâncias. Diz-se que os erros são correlacionados (e possivelmente heterocedásticos). Escreve-se  $\varepsilon \sim N(\mu, W\sigma^2)$ , sendo  $W$  uma matriz simétrica, positiva definida, que representa as variâncias e covariâncias dos erros.

Para o ajuste de um modelo pelo método dos quadrados mínimos ordinários, pressupõe-se que a média dos erros é nula, ou seja  $E(\varepsilon_i) = 0$ ; que

os erros são homocedásticos, isto é,  $E(\varepsilon_i^2) = \sigma^2$  e que os erros são não-correlacionados entre si,  $E(\varepsilon_i \varepsilon_j) = 0$  para  $i \neq j$ .

A teoria geral dos modelos de regressão, que pode ser encontrada em Steel, Torrie e Dickley (1998), mostra que o vetor  $\beta$  pode ser estimado por meio do método dos quadrados mínimos, considerando que a soma de quadrados dos erros seja mínima. Assim, a soma de quadrados dos desvios ou soma de quadrados de resíduo, para o modelo  $Y = X\beta + \varepsilon$ , é dado pela função:

$$Z = \varepsilon' \varepsilon = (Y' - \beta' X')(Y - X\beta).$$

A função  $Z$  apresenta ponto de mínimo para os valores de  $\beta$  que tomem a derivada parcial em relação a  $\beta$  identicamente nula, isto é:

$$(d\beta')(X'X\beta - X'Y) \equiv 0.$$

Assim, a derivada parcial em relação a  $\beta$  será identicamente nula para

$$X'X\hat{\beta} = X'Y,$$

que é o sistema de equações normais, citado por Neter, Wasserman e Kutner (1990).

Como a matriz  $X$  é de posto coluna completo, então  $X'X$  é uma matriz positiva definida e, assim,  $X'X$  é não singular. Portanto, existe a matriz inversa  $(X'X)^{-1}$  e a solução única para  $\beta$ , de acordo com Draper e Smith (1998), é:

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

Esta solução corresponde ao estimador linear não-tendencioso de variância mínima de  $\beta$ , o que é mostrado por Hoffmann e Vieira (1998), Neter, Wasserman e Kutner (1990), como se segue.

Substituindo  $Y = X\beta + \varepsilon$  em  $\hat{\beta} = (X'X)^{-1} X'Y$ , tem-se:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1} X'Y = (X'X)^{-1} X'(X\beta + \varepsilon) = \\ &= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon = \beta + (X'X)^{-1} X'\varepsilon\end{aligned}$$

assim, aplicando o operador esperança matemática, obteve-se:

$$\begin{aligned}E[\hat{\beta}] &= E[\beta + (X'X)^{-1} X'\varepsilon] \\ E[\hat{\beta}] &= \beta\end{aligned}$$

Portanto, o estimador de quadrados mínimos de  $\beta$ , dado por

$$\hat{\beta} = (X'X)^{-1} X'Y, \text{ é um estimador não viesado.}$$

Outra propriedade importante relacionada aos parâmetros de um modelo de regressão se refere à matriz de variâncias e covariâncias dos estimadores dos parâmetros. Sua obtenção, segundo Hoffmann e Vieira (1998), Neter, Wasserman e Kutner (1990) e Draper e Smith (1998), pode ser feita como descrito a seguir.

Da definição de covariância, como em Roussas (1973), e da expressão

$$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon, \text{ tem-se que:}$$

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1} X'\varepsilon\varepsilon'X(X'X)^{-1}],$$

dado que a matriz  $X'X$  é simétrica, positiva definida e considerando que os erros são homocedásticos e não-correlacionados entre si, tem-se:

$$E(\varepsilon\varepsilon') = I\sigma^2$$

Assim:

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = (X'X)^{-1} X'I\sigma^2 X(X'X)^{-1}$$

ou seja,

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \text{Var}[\hat{\beta}] = (X'X)^{-1} \sigma^2.$$

### 2.4.1 Análise de variância da regressão linear múltipla

Em Hoffmann e Vieira (1998) podem-se encontrar as expressões para as somas de quadrados que compõem a análise de variância. Os autores mostram que a soma de quadrados dos desvios, ou soma de quadrados residual, é dada por:

$$\varepsilon' \varepsilon = Y'Y - 2\hat{\beta}' X'Y + \hat{\beta}' X' X \hat{\beta},$$

mas, do sistema de equações normais tem-se que:

$$X' X \hat{\beta} = X'Y,$$

assim,

$$\text{SQRes} = \varepsilon' \varepsilon = Y'Y - 2\hat{\beta}' X'Y + \hat{\beta}' X' X \hat{\beta}$$

$$\text{SQRes} = \varepsilon' \varepsilon = Y'Y - \hat{\beta}' X'Y$$

somando e subtraindo uma constante apropriada  $C$ , dada por

$$C = \frac{\left( \sum_{i=1}^n y_i \right)^2}{n} = Y' J Y, \text{ referida como correção, a SQRes fica}$$

$$\text{SQRes} = (Y'Y - C) - (\hat{\beta}' X'Y - C)$$

na qual

$(Y'Y - C)$  é a soma de quadrados total,

$(\hat{\beta}' X'Y - C)$  é a soma de quadrados de regressão,

$C = Y' J Y$ , com  $J$  uma matriz de dimensões  $n \times n$  com todos os elementos iguais a  $\frac{1}{n}$ , é a correção.

Desenvolvendo-se as expressões, a soma de quadrados total, pode ser escrita como

$$SQ_{Tot} = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}$$

ou, ainda, como

$$SQ_{Tot} = \sum_{i=1}^n (y_i - \bar{y})^2$$

com  $\bar{y}$  sendo a média geral. Do mesmo modo a soma de quadrados de regressão pode ser escrita como:

$$SQ_{Reg} = \hat{\beta}' X' X \hat{\beta} - Y' J Y,$$

mas do sistema de equações normais, tem-se que

$$X' X \hat{\beta} = X' Y,$$

de forma que:

$$SQ_{Reg} = \hat{\beta}' X' Y - Y' J Y.$$

Analisando as expressões para as somas de quadrados pode-se observar que a seguinte relação é estabelecida:

$$SQ_{Res} = (SQ_{Tot}) - (SQ_{Reg}).$$

Conforme Draper e Smith (1998), esses resultados levam à construção da tabela de análise de variância, como está apresentado na Tabela 1.

TABELA 1 Esquema de análise de variância para regressão

F.V.	G.L.	S.Q.	Q.M.	F
Regressão	p-1	$\hat{\beta}' X' Y - C$	$\frac{SQ_{Reg}}{p-1}$	$\frac{QM_{Reg}}{QM_{Res}}$
Resíduo	n-p	$Y' Y - \hat{\beta}' X' Y$	$\frac{SQ_{Res}}{n-p}$	
Total	n-1	$Y' Y - C$		

O quadrado médio residual, dado pelo quociente  $\frac{SQRes}{n-p}$ , é, de acordo com Draper e Smith (1998) e Hoffmann e Vieira (1998), um estimador não-tendencioso da variância residual ( $\sigma^2$ ), pois:

$$E\left[\frac{SQRes}{n-p}\right] = E[QMRes] = E(s^2) = \sigma^2.$$

#### 2.4.2 Matriz de variâncias e covariâncias

A matriz de variâncias e covariâncias dos estimadores dos parâmetros é, por definição, segundo Neter, Wasserman e Kutner (1990) e Hoffmann e Vieira (1998), dada por:

$$E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = Var[\hat{\beta}] = (X'X)^{-1}\sigma^2,$$

e seu estimador é:

$$\hat{Var}[\hat{\beta}] = s^2 \{\hat{\beta}\} = QMRes(X'X)^{-1}$$

a qual pode ser representada por:

$$s^2 \{\hat{\beta}\} = \begin{bmatrix} s^2 \{\hat{\beta}_0\} & s \{\hat{\beta}_0, \hat{\beta}_1\} & \dots & \dots & s \{\hat{\beta}_0, \hat{\beta}_{p-1}\} \\ s \{\hat{\beta}_1, \hat{\beta}_0\} & s^2 \{\hat{\beta}_1\} & \dots & \dots & s \{\hat{\beta}_1, \hat{\beta}_{p-1}\} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ s \{\hat{\beta}_{p-1}, \hat{\beta}_0\} & s \{\hat{\beta}_{p-1}, \hat{\beta}_1\} & \dots & \dots & s^2 \{\hat{\beta}_{p-1}\} \end{bmatrix}$$

Nos termos da diagonal principal estão representadas as variâncias dos estimadores dos parâmetros e nos termos fora da diagonal, as covariâncias entre os estimadores dos parâmetros, ou seja,

$$\begin{aligned}
s^2 \{ \hat{\beta}_0 \} &= E \left[ (\hat{\beta}_0 - \beta_0)^2 \right] = \text{Vâr} [ \hat{\beta}_0 ] \\
s^2 \{ \hat{\beta}_1 \} &= E \left[ (\hat{\beta}_1 - \beta_1)^2 \right] = \text{Vâr} [ \hat{\beta}_1 ] \\
s^2 \{ \hat{\beta}_{p-1} \} &= E \left[ (\hat{\beta}_{p-1} - \beta_{p-1})^2 \right] = \text{Vâr} [ \hat{\beta}_{p-1} ] \\
s \{ \hat{\beta}_1, \hat{\beta}_0 \} &= E \left[ (\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) \right] = \text{Côv} [ \hat{\beta}_1, \hat{\beta}_0 ] \\
s \{ \hat{\beta}_0, \hat{\beta}_1 \} &= E \left[ (\hat{\beta}_1 - \beta_1)(\hat{\beta}_0 - \beta_0) \right] = \text{Côv} [ \hat{\beta}_0, \hat{\beta}_1 ] \\
s \{ \hat{\beta}_{p-1}, \hat{\beta}_0 \} &= E \left[ (\hat{\beta}_0 - \beta_0)(\hat{\beta}_{p-1} - \beta_{p-1}) \right] = \text{Côv} [ \hat{\beta}_{p-1}, \hat{\beta}_0 ] \\
s \{ \hat{\beta}_0, \hat{\beta}_{p-1} \} &= E \left[ (\hat{\beta}_{p-1} - \beta_{p-1})(\hat{\beta}_0 - \beta_0) \right] = \text{Côv} [ \hat{\beta}_0, \hat{\beta}_{p-1} ] \\
s \{ \hat{\beta}_{p-1}, \hat{\beta}_1 \} &= E \left[ (\hat{\beta}_1 - \beta_1)(\hat{\beta}_{p-1} - \beta_{p-1}) \right] = \text{Côv} [ \hat{\beta}_{p-1}, \hat{\beta}_1 ] \\
s \{ \hat{\beta}_1, \hat{\beta}_{p-1} \} &= E \left[ (\hat{\beta}_{p-1} - \beta_{p-1})(\hat{\beta}_1 - \beta_1) \right] = \text{Côv} [ \hat{\beta}_1, \hat{\beta}_{p-1} ]
\end{aligned}$$

Assim, pode-se reescrever a matriz de variância e covariância da seguinte forma:

$$\text{Vâr} [ \hat{\beta} ] = \begin{bmatrix} \text{Vâr} [ \hat{\beta}_0 ] & \text{Côv} [ \hat{\beta}_0, \hat{\beta}_1 ] & \dots & \dots & \text{Côv} [ \hat{\beta}_0, \hat{\beta}_{p-1} ] \\ \text{Côv} [ \hat{\beta}_1, \hat{\beta}_0 ] & \text{Vâr} [ \hat{\beta}_1 ] & \dots & \dots & \text{Côv} [ \hat{\beta}_1, \hat{\beta}_{p-1} ] \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \text{Côv} [ \hat{\beta}_{p-1}, \hat{\beta}_0 ] & \text{Côv} [ \hat{\beta}_{p-1}, \hat{\beta}_1 ] & \dots & \dots & \text{Vâr} [ \hat{\beta}_{p-1} ] \end{bmatrix}$$

Entre os modelos de regressão merecem um destaque especial os modelos de regressão polinomiais. Isto por apresentarem uma ampla aplicação nas pesquisas da experimentação agrônômica, principalmente nos casos onde os tratamentos são constituídos pelas doses de algum fertilizante.

## 2.5 Regressão polinomial

A regressão polinomial possui ampla aplicação nas pesquisas agropecuárias, principalmente com o uso de fertilizantes.

Em face dos elevados custos de produção e dos fertilizantes, torna-se necessário encontrar métodos e procedimentos que permitam indicar ao empreendedor agrícola as melhores doses de nutrientes que forneçam um máximo de produção.

Segundo Neter, Wasserman e Kutner (1990), o modelo de regressão quadrática ou de regressão polinomial quadrática, com uma variável independente, é:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

no qual

$y_i$  representa os valores observados da variável dependente;

$\beta_0, \beta_1, \beta_2$  são os parâmetros a serem estimados;

$x_i$  representam os valores das variáveis independentes;

$\varepsilon_i$  são os erros experimentais relacionados com os valores observados  $y_i$  que, em geral, são considerados independentes e normalmente distribuídos com média zero e variância constante.

Os autores estudaram este modelo por meio dos procedimentos de regressão linear múltipla. Para isso, fizeram as transformações  $x_{i1} = x_i$  e  $x_{i2} = x_i^2$  e, assim, o modelo passou a ser escrito da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,$$

Dessa forma, tem-se que o modelo de regressão polinomial é um caso particular do modelo de regressão linear múltipla.



## 2.6 Estimando o máximo ou o mínimo de uma equação de regressão quadrática

Seja o modelo de regressão quadrática ou de regressão polinomial quadrática, com uma variável independente, dado por Neter, Wasserman e Kutner (1990):

$$\hat{y}_i = \hat{a} + \hat{b}x_i + \hat{c}x_i^2$$

A obtenção de um ponto crítico de uma dada função é feita conforme Lang (1976) e Kaplan (1972); derivando-se  $\hat{y}_i$  em relação a  $x$  e fazendo sua derivada igual a zero, tem-se:

$$\frac{d\hat{y}_i}{dx} = \frac{d(\hat{a} + \hat{b}x_i + \hat{c}x_i^2)}{dx} = \hat{b} + 2\hat{c}x_i = 0,$$

logo, o estimador do ponto crítico da equação de regressão quadrática é:

$$\hat{x}_i = \frac{\hat{b}}{-2\hat{c}}$$

que será a abscissa de um ponto de máximo se  $\hat{c}$  for negativo e será de mínimo se  $\hat{c}$  for positivo (Neter, Wasserman e Kutner, 1990).

## 2.7 Estimativa da variância do ponto crítico

O conhecimento da variância de um ponto crítico pode ser de muita importância em algumas pesquisas. Ele possibilita que, em vez de se ter apenas uma estimativa pontual, como é bastante usual, construam-se intervalos de confiança para o verdadeiro valor do ponto crítico. Além disso, também possibilita testar este valor por meio da validade ou não de uma hipótese de interesse a seu respeito.

Para estimar a variância do ponto crítico em um estudo usando simulação, D'Aulísio (1976) utilizou a expressão usual:

$$Var(\hat{x}) = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1},$$

citada por Spiegel (1975), na qual  $x_i$  representa os valores simulados do ponto crítico;  $n$  é o número de observações amostrais ou de pontos críticos simulados.

Segundo Mood, Graybill e Boes (1974), a variância de um quociente de duas variáveis aleatórias  $x$  e  $y$  é dado, de forma aproximada, pela expressão:

$$Var\left[\frac{x}{y}\right] \approx \left(\frac{\mu_x}{\mu_y}\right)^2 \left(\frac{Var[x]}{\mu_x^2} + \frac{Var[y]}{\mu_y^2} - \frac{2Cov[x,y]}{\mu_x\mu_y}\right).$$

D'Aulísio (1976) estudando a variância dos pontos de máximo ou de mínimo de equações de regressão de segundo grau, demonstrou que, se  $x = \frac{-\hat{b}}{2\hat{c}}$  é um estimador consistente do ponto crítico, sua variância pode ser estimada aproximadamente pela primeira derivada do estimador do ponto crítico em relação a  $x$ . Portanto, o estimador da variância é dado pela expressão:

$$Var(\hat{x}) = \frac{1}{4} \left[ \frac{Var(\hat{b})}{\hat{c}^2} + \frac{\hat{b}^2 \cdot Var(\hat{c})}{\hat{c}^4} \right]$$

Deve-se ressaltar que D'Aulísio (1976) utilizou um modelo de regressão com a variável independente na forma de polinômios ortogonais de primeiro e segundo grau. Daí, verifica-se que, na expressão da variância, não aparece o termo de covariância entre os estimadores  $\hat{b}$  e  $\hat{c}$ , fato que ocorre somente neste caso.

## 2.8 Coeficientes de assimetria e curtose

Testes que avaliam assimetria e curtose são úteis e ajudam a verificar a significância da variação de certos momentos da amostra, em relação aos valores esperados, de uma população normal. Nas aplicações práticas, porém, pode-se ter valores próximos de zero e, principalmente neste caso, existe a necessidade de saber se o valor obtido é diferente ou não de zero.

O coeficiente de assimetria pode ser estimado, conforme definido por Fisher (1930), usando-se a expressão:

$$\hat{\gamma}_1 = \frac{\hat{\mu}_3}{\hat{\sigma}^3}.$$

Para verificar o grau de assimetria ou se os valores estimados de  $\hat{\gamma}_1$  diferem ou não de zero, Fisher (1930) sugeriu como estimador da variância do estimador do coeficiente de assimetria, a seguinte expressão:

$$\text{Var}(\hat{\gamma}_1) = \frac{6n(n-1)}{(n-2)(n+1)(n+3)},$$

sendo  $n$  o número de observações amostrais.

Nota-se que esta expressão depende apenas do número de observações, não envolvendo nenhum parâmetro.

Para avaliar a hipótese  $H_0 : \gamma_1 = 0$  contra  $H_a : \gamma_1 \neq 0$ , Fisher (1930) sugeriu a aplicação do teste de “t” de Student:

$$t = \frac{\hat{\gamma}_1 - \gamma_1}{s(\hat{\gamma}_1)}$$

em que o valor da estatística “t” é a razão entre a estimativa do parâmetro ( $\gamma_1$ ) e seu erro padrão, obtido a partir da raiz quadrada da variância  $\text{Var}(\hat{\gamma}_1)$ . Este teste é possível de ser realizado, pois, para grandes amostras,  $\hat{\gamma}_1$  pode ser aproximado por uma distribuição normal.

O coeficiente de curtose foi definido por Fisher (1930) como:

$$\hat{\gamma}_2 = \frac{\hat{\mu}_4}{\hat{\sigma}^4} - 3.$$

Para verificar se o grau de curtose ou de achatamento de uma distribuição ou se os valores estimados de  $\hat{\gamma}_2$  diferem ou não de zero, necessita-se da variância do estimador do coeficiente de curtose o qual apresenta a seguinte forma:

$$\text{Var}(\hat{\gamma}_2) = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)},$$

sendo  $n$  o número de observações amostrais.

Para avaliar a hipótese  $H_0 : \gamma_2 = 0$  contra  $H_a : \gamma_2 \neq 0$ , Fisher (1930) sugeriu o uso da estatística “t” de Student, empregando a seguinte expressão:

$$t = \frac{\hat{\gamma}_2 - \gamma_2}{s(\hat{\gamma}_2)}$$

na qual  $\hat{\gamma}_2$  representa a estimativa do parâmetro ( $\gamma_2$ ) e  $s(\hat{\gamma}_2)$  o seu erro padrão, obtido a partir da raiz quadrada da variância  $\text{Var}(\hat{\gamma}_2)$ .

## 2.9 Intervalos de confiança

Um método para se obter intervalo de confiança para um quociente é o baseado no teorema de E.C. Fieller, Finney (1964).

Sendo  $x_1$  e  $x_2$  duas variáveis aleatórias normalmente distribuídas com média zero e variância  $\sigma_1^2$  e  $\sigma_2^2$ , respectivamente, a nova variável  $u = x_1 - \mu x_2$ ,

função linear das observações  $x_1$  e  $x_2$  terá também distribuição normal de média zero e variância  $Var(u) = \sigma_1^2 + \mu^2 \sigma_2^2$ .

Sendo  $s_1^2$  e  $s_2^2$  estimadores não viesados de  $\sigma_1^2$  e de  $\sigma_2^2$ , respectivamente, o quociente:

$$\frac{u}{\sqrt{Var(\hat{u})}} = \frac{x_1 - \mu x_2}{\sqrt{s_1^2 + \mu^2 s_2^2}},$$

tem distribuição “t” de Student (Kendall e Stuart, 1977).

Assim sendo, a um certo nível de probabilidade  $\alpha$  apropriado, a inequação  $(x_2^2 + t^2 s_2^2) \mu^2 - 2x_1 x_2 \mu + x_1^2 - t^2 s_1^2 \leq 0$  dá os extremos do intervalo de confiança. Tomando-se  $p = x_2^2 + t^2 s_2^2$  e  $q = x_1^2 - t^2 s_1^2$ , as raízes da equação correspondente são dadas pela expressão:

$$\mu = \frac{x_1 x_2 \pm \sqrt{x_1^2 x_2^2 - pq}}{p}.$$

Para interpretação das raízes  $\mu$ , que irão fornecer o intervalo de confiança para  $\mu = \frac{x_1}{x_2}$  sob a hipótese de nulidade, D’Aulísio (1976) considerou os seguintes casos:

1.  $p > 0$  e  $q > 0$

Neste caso,  $\mu_1$  e  $\mu_2$  são raízes reais positivas. A solução da inequação será:

$$\mu_1 \leq \mu \leq \mu_2$$

2.  $p > 0$  e  $q = 0$

A solução da inequação será:

$$0 \leq \mu \leq \frac{2x_1x_2}{p}$$

3.  $p > 0$  e  $q < 0$

Neste caso, têm-se duas raízes reais e de sinais contrários,  $\mu_1 < 0$  e  $\mu_2 > 0$ .

A solução da inequação será:

$$\mu_1 \leq \mu \leq \mu_2$$

4.  $p = 0$

Neste caso, a inequação se torna:  $p\mu^2 - 2x_1x_2\mu + q \leq 0$

$$-2x_1x_2\mu + q \leq 0 \quad \Rightarrow \quad \mu \geq \frac{q}{2x_1x_2}$$

4.1.  $p = 0$  e  $q > 0$

A solução da inequação será:

$$\mu \geq \frac{q}{2x_1x_2} > 0 \quad \Rightarrow \quad \mu \geq 0$$

4.2.  $p = 0$  e  $q = 0$

A solução da inequação será:

$$\mu \geq \frac{q}{2x_1x_2} = 0 \quad \Rightarrow \quad \mu \geq 0$$

4.3.  $p = 0$  e  $q < 0$

A solução da inequação será:

$$\mu \geq \frac{q}{2x_1x_2}$$

5.  $p < 0$  e  $q > 0$

Neste caso, têm-se duas raízes reais e de sinais contrários,  $\mu_1 < 0$  e  $\mu_2 > 0$ .

A solução da inequação será:

$$\mu \leq \mu_1 < 0 \quad \text{ou} \quad \mu \geq \mu_2 > 0$$

6.  $p < 0$  e  $q = 0$

Neste caso, têm-se duas raízes reais, uma  $\mu_2$  é nula, e a outra

$\mu_1 = \frac{2x_1x_2}{p} < 0$ , de modo que a solução da inequação, será:

$$\mu \leq \mu_1 < 0 \quad \text{ou} \quad \mu \geq 0$$

7.  $p < 0$  e  $q < 0$

A natureza das raízes dependerá do valor do discriminante

$$x_1^2 x_2^2 - pq.$$

$$7.1. x_1^2 x_2^2 - pq > 0$$

Neste caso, as duas raízes são reais e negativas. A solução da inequação será:

$$\mu \leq \mu_1 < 0 \quad \text{ou} \quad \mu \geq \mu_2 < 0$$

$$7.2. x_1^2 x_2^2 - pq = 0$$

As raízes  $\mu_1$  e  $\mu_2$  são reais e iguais. A inequação estará satisfeita para qualquer valor real de  $\mu$ .

$$7.3. x_1^2 x_2^2 - pq < 0$$

As raízes são complexas. A inequação estará satisfeita para qualquer valor real de  $\mu$ .

Intervalos de confiança propriamente ditos só são obtidos quando  $p > 0$ .

Neste caso, interessam intervalos que incluam somente valores de  $x$  positivos que, no caso de experimentos de adubação correspondem às dosagens que proporcionam as produções máximas.

## 2.10 Desenvolvimentos assintóticos – série de Taylor

Kaplan (1972) dá a seguinte definição de série de Taylor: seja  $f(x)$  a soma de uma série de potências cujo intervalo de convergência é  $a - r < x < a + r$  ( $r > 0$ ):

$$f(x) = \sum_{n=0}^{\infty} c_n (x - a)^n, \quad a - r < x < a + r.$$

Esta série é chamada de série de Taylor de  $f(x)$  para  $x = a$  se os coeficientes  $c_n$  são dados pela da seguinte forma:

$$c_0 = f(a), \quad c_1 = \frac{f'(a)}{1!}, \quad c_2 = \frac{f''(a)}{2!}, \dots, \quad c_n = \frac{f^{(n)}(a)}{n!}, \dots,$$

tais que:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \dots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \dots$$

**Teorema:** Toda série de potências com raio de convergência não nulo é a série de Taylor de sua soma.

**Demonstração:** Seja  $f(x)$  dada por  $f(x) = \sum_{n=0}^{\infty} c_n (x - a)^n$ . Então, por repetidas diferenciações, encontra-se que:



$$\begin{aligned}
f(x) &= c_0 + c_1(x-a) + \dots + c_n(x-a)^n + \dots, \\
f'(x) &= c_1 + 2c_2(x-a) + \dots + n.c_n(x-a)^{n-1} + \dots, \\
f''(x) &= 2c_2 + 6c_3(x-a) + \dots + n(n-1).c_n(x-a)^{n-2} + \dots,
\end{aligned}$$

.....,

$$\begin{aligned}
f^{(n)}(x) &= n(n-1)(n-2)\dots 2.1.c_n + \\
&+ (n+1)n(n-1)\dots 2.c_{n+1}(x-a) + \dots,
\end{aligned}$$

Aqui todas as séries convergem em  $a - r^* < x < a + r^*$ . Fazendo agora

$x = a$ , segue que:

$$f(a) = c_0, \quad f'(a) = c_1, \quad f''(a) = 2c_2, \dots, \quad f^{(n)}(a) = n!c_n, \dots$$

Assim,  $c_0 = f(a)$  e

$$c_n = \frac{f^{(n)}(a)}{n!}, \quad n = 1, 2, \dots,$$

como queríamos demonstrar.

No caso de  $a = 0$ , a expressão para a série de Taylor de  $f(x)$  torna-se

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + \dots$$

Esta expressão é chamada série de Maclaurin de  $f(x)$  que, para muitos propósitos, é mais fácil de se usar. A substituição  $t = x - a$  reduz a série de Taylor geral à forma de Maclaurin, Courant (1963).

### 2.10.1 Série de Taylor para funções de uma variável

De acordo com Lang (1976), pela teoria das funções de uma variável, obtém-se a seguinte expressão para os valores de uma função  $f$  na vizinhança de um ponto  $a$ , por meio das derivadas:

$$f(a+h) = f(a) + f'(a)h + \frac{f^{(2)}(a)}{2!}h^2 + \dots + \frac{f^{(r-1)}(a)}{(r-1)!}h^{r-1} + R_r,$$

em que:

$$R_r = \frac{f^{(r)}(c)}{r!}h^r,$$

para algum ponto  $c$  entre  $a$  e  $a+h$ .

### 2.10.2 Série de Taylor para funções de duas ou mais variáveis

É possível, de acordo com Lang (1976), obter uma função de duas ou mais variáveis expandida em série de Taylor. Seja uma função  $f$  de  $n$  variáveis definida num conjunto aberto  $U$  e tendo derivadas parciais contínuas até a ordem  $r$ . Sejam  $P$  um ponto de  $U$  e  $H$  um vetor. Suponha que o segmento de reta:

$$P + tH, \quad 0 \leq t \leq 1,$$

esteja contido em  $U$ . Então, existe um número  $\tau$  entre 0 e 1 tal que

$$f(P+H) = f(P) + \frac{(H \cdot \nabla)f(P)}{1!} + \dots + \frac{(H \cdot \nabla)^{r-1}f(P)}{(r-1)!} + \frac{(H \cdot \nabla)^r f(P + \tau H)}{r!}.$$

Esta forma é obtida substituindo na Fórmula de Taylor para uma variável, as derivadas da função  $g(t) = f(P + tH)$  pelas suas expressões. Tem-se, assim, que:

$$g^{(s)}(0) = (H \cdot \nabla)^s f(P)$$

e

$$g^{(r)}(\tau) = (H \cdot \nabla)^r f(P + \tau H).$$

Reescrevendo, em termos das notações  $\frac{\partial}{\partial x}$  e  $\frac{\partial}{\partial y}$ , segue que:

$$f(a+h, b+k) = f(a, b) + \left( h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right) f(a, b) + \dots + \left( h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^{r-1} f(a, b) + \left( h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^r f(a + \tau h, b + \tau k).$$

As potências dos operadores diferenciais

$$\left( h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)$$

são calculadas pelo desenvolvimento do binômio. Por exemplo:

$$\left( h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^2 = h^2 \frac{\partial^2}{\partial x^2} + 2hk \frac{\partial^2}{\partial x \partial y} + k^2 \frac{\partial^2}{\partial y^2},$$

$$\left( h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^3 = h^3 \left( \frac{\partial}{\partial x} \right)^3 + 3h^2 k \left( \frac{\partial}{\partial x} \right)^2 \left( \frac{\partial}{\partial y} \right) + 3hk^2 \left( \frac{\partial}{\partial x} \right) \left( \frac{\partial}{\partial y} \right)^2 + k^3 \left( \frac{\partial}{\partial y} \right)^3$$

e assim por diante.

Segundo Kaplan (1972), há uma fórmula de Taylor com resíduo para funções de  $n$  variáveis. Para o caso de duas variáveis, tem-se:

$$F(x, y) = F(x_1, y_1) + dF(x_1, y_1; x - x_1, y - y_1) + \dots + \frac{1}{r!} d^r F(x_1, y_1; x - x_1, y - y_1) + \frac{1}{(r+1)!} d^{r+1} F(x^*, y^*; x - x_1, y - y_1);$$

com  $x^* = x_1 + t(x - x_1)$ ,  $y^* = y_1 + t(y - y_1)$ ,  $0 < t < 1$ .

O ponto  $(x^*, y^*)$  se encontra entre  $(x_1, y_1)$  e  $(x, y)$  na linha de junção desses pontos. Para  $r = 1$ , a fórmula se torna:

$$F(x, y) = F(x_1, y_1) + (x - x_1)F_x(x^*, y^*) + (y - y_1)F_y(x^*, y^*)$$

que é conhecida como a Lei da Média para funções de duas variáveis.

Para a demonstração deste resultado, pode-se escrever:

$$\phi(t) = F[x_1 + t(x - x_1), y_1 + t(y - y_1)], \quad 0 \leq t \leq 1$$

Tem-se que  $x$  e  $y$  são considerados fixos e  $\phi$  depende somente de  $t$ .  
Pela lei da média para  $\phi$ ,

$$\phi(1) = \phi(0) + \phi'(t^*), \quad 0 < t^* < 1$$

Mas  $\phi(1) = F(x, y)$ ,  $\phi(0) = F(x_1, y_1)$  e

$$\begin{aligned} \phi'(t) &= (x - x_1)F_x [x_1 + t(x - x_1), y_1 + t(y - y_1)] \\ &\quad + (y - y_1)F_y [x_1 + t(x - x_1), y_1 + t(y - y_1)] \end{aligned}$$

Substituindo  $t$  por  $t^*$ , obtém-se:

$$F(x, y) = F(x_1, y_1) + (x - x_1)F_x(x^*, y^*) + (y - y_1)F_y(x^*, y^*)$$

A fórmula de Taylor com resíduo para funções de  $n$  variáveis é demonstrada de forma análoga, ou seja, baseada na fórmula de Taylor para  $\phi$ :

$$\phi(1) = \phi(0) + \phi'(0) + \dots + \frac{\phi^{(r)}(0)}{r!} + \frac{\phi^{(r+1)}(t^*)}{(r+1)!},$$

em que  $0 < t^* < 1$ . Pode-se encontrar, por indução, que

$$\phi^{(r)}(t) = d^r F [x_1 + t(x - x_1), y_1 + t(y - y_1); x - x_1, y - y_1];$$

e a validade desta equação é assegurada se  $F(x, y)$  tem derivadas contínuas por meio da  $(r + 1)$ -ésima ordem, num domínio  $D$ , contendo o segmento de linha que liga  $(x, y)$  a  $(x_1, y_1)$ .

A série de Taylor ou Fórmula de Taylor pode ser usada para estudar a natureza de uma função perto de um ponto particular (Lima, 1993). Como foi dito anteriormente, os termos lineares resultam em  $dF$ , a melhor “aproximação linear” para  $F(x, y) - F(x_1, y_1)$ . Se  $dF = 0$ , os termos quadráticos  $d^2F/2!$  tornam-se de grande importância. Em particular, se a expressão quadrática

$$d^2F = A(x - x_1)^2 + 2B(x - x_1)(y - y_1) + C(y - y_1)^2$$

for positiva, exceto para  $x = x_1$ ,  $y = y_1$ . Então,  $F(x, y)$  tem um mínimo no ponto  $(x_1, y_1)$ .

Em estatística, segundo Leite e Singer (1990), uma das ferramentas mais utilizadas no estudo de métodos assintóticos é a aproximação de uma função real  $f$  de variável real por um polinômio. Embora a expansão de  $f$  em série de potências na vizinhança de um ponto  $x_0$  seja importante sob o ponto de vista teórico, o que realmente interessa nas aplicações é o estudo do resto proveniente da aproximação de  $f$  por um polinômio.

Para uma função expandida em série de Taylor, para todo  $x \in U$ ,  $R_n(x) = f(x) - P(x)$  é o resto ou o erro da aproximação de  $f(x)$  por  $P(x)$ . Logo, se  $f$  for derivável até a ordem  $n$  em um ponto  $x_0$ , tem-se:

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + R_n(x),$$

para todo  $x \in U$ . A expressão acima é denominada fórmula de Taylor, com resto  $R_n$ , de  $f$  em torno de  $x_0$ . Tem-se que, para  $n = 1$  e  $2$ ,  $R_n(x)$  tende a zero mais rapidamente do que  $(x - x_0)^n$  quando  $x$  tende a  $x_0$  (Lima, 1993).

**Teorema:** Seja  $f$  uma função real de variável real derivável até a ordem  $n + 1$  em um intervalo aberto  $I$  e sejam  $x_0, x \in I$ . Então, existe pelo menos um ponto  $c$  entre  $x_0$  e  $x$  tal que:

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}.$$

Em outras palavras, o resto na fórmula de Taylor é dado por:

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}.$$

Observa-se que, quando  $n = 0$ , este teorema é precisamente o teorema do valor médio.

Em Leite e Singer (1990), ainda pode-se encontrar uma série de aplicações sobre o uso das expansões em Taylor e Maclaurin em estatística, como, por exemplo, a Lei dos Grandes Números e o Teorema Central do Limite.

## 3 MATERIAL E MÉTODOS

### 3.1 Material

Os dados utilizados neste trabalho foram os mesmos utilizados por Freitas (1978), e são provenientes de 536 ensaios conduzidos com a cultura de algodão, realizados nos estados do Paraná, Minas Gerais e Goiás, no período de 1970/1975.

### 3.2 Métodos

Este trabalho foi realizado por meio de simulação Monte Carlo, para avaliar distribuições de frequências do ponto crítico de uma equação de regressão do segundo grau.

Foram considerados como referência os valores médios obtidos por Freitas (1978) para os coeficientes de regressão cujas estimativas foram  $\hat{b} = 90,1713$  e  $\hat{c} = -98,8947$ , necessárias para o cálculo do ponto crítico. Considerou-se, ainda, que as variâncias eram proporcionais, sendo  $Var(\hat{c}) = 24,4532 Var(\hat{b})$ .

Para simular dados experimentais para  $\hat{b}$  e  $\hat{c}$ , com média e variância estipulados conforme a situação, foi desenvolvido um procedimento utilizando o software MATLAB® - Matrix Laboratory.

As seguintes dosagens foram consideradas com 4 repetições cada: 0.0; 0.2; 0.4; 0.6 e 0.8, caracterizadas pelas diferentes variâncias teóricas,  $\sigma^2 = 0,1$ ; 0,5; 1,0; 5,0; 10; 15; 20; 50.

A partir desses valores reais, geraram-se duas séries com 5.000 erros aleatórios cada, com distribuição normal de média zero para cada uma das

variâncias teóricas consideradas, denominando  $e_1$  a primeira série e  $e_2$  a segunda.

Assim:

$$e_1 \sim N(0, \sigma^2)$$

$$e_2 \sim N(0, \sigma^2)$$

Obtiveram-se então, 5.000 erros  $e_3$  após multiplicar  $e_2$  por  $\sqrt{24,4532}$ .

Assim as estimativas de  $\hat{b}$  e  $\hat{c}$  podem ser escritas como:

$$\hat{b} = 90,1713 + e_1$$

$$\hat{c} = -98,8947 + e_3.$$

Assim, construiu-se um vetor com 5.000 observações, cujos elementos são:

$$\hat{x} = \frac{\hat{b}}{-2\hat{c}}.$$

Determinaram-se então, os valores médios para as estimativas  $\hat{b}$ ,  $\hat{c}$  e  $\hat{x}$ , suas respectivas variâncias, com a finalidade de verificar a qualidade da simulação.

A variância do quociente ou do ponto crítico  $\hat{x} = \frac{\hat{b}}{-2\hat{c}}$  foi calculada utilizando-se três expressões  $Var_1(\hat{x})$ ,  $Var_2(\hat{x})$  e  $Var_3(\hat{x})$ , a saber:

$$Var(\hat{x}) = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}$$

$$Var_2(\hat{x}) = \frac{1}{4} \left[ \frac{Var(\hat{b})}{\hat{c}^2} + \frac{\hat{b}^2 \cdot Var(\hat{c})}{\hat{c}^4} \right]$$



$$Var_3(\hat{x}) = \frac{1}{4} Var \left[ \frac{\hat{b}}{\hat{c}} \right] = \frac{1}{4} \left[ \left( \frac{\mu_{\hat{b}}}{\mu_{\hat{c}}} \right)^2 \left( \frac{Var[\hat{b}]}{\mu_{\hat{b}}^2} + \frac{Var[\hat{c}]}{\mu_{\hat{c}}^2} - \frac{2Cov[\hat{b}, \hat{c}]}{\mu_{\hat{b}}\mu_{\hat{c}}} \right) \right]$$

Para cada situação do conjunto de 5.000 simulações foi obtida a média do ponto crítico e das três variâncias. Do mesmo modo, procedeu-se ao cálculo do coeficiente de assimetria e curtose, utilizando-se as expressões citadas em 2.8.1 e 2.8.2.

Para os coeficientes de assimetria e curtose, calculam-se  $\hat{\gamma}_1$ ,  $\hat{\gamma}_2$  cujas respectivas variâncias são constantes em todos os casos, visto que, todos os vetores têm 5.000 observações, assim:

$$Var(\hat{\gamma}_1) = 0,0012 \quad e \quad Var(\hat{\gamma}_2) = 0,0048$$

para o cálculo da estatística “t” de Student.

Conhecidas as estimativas das variâncias, determinaram-se os intervalos de confiança para  $\hat{x} = \frac{\hat{b}}{-2\hat{c}}$ , estimador do ponto crítico, a um coeficiente de confiança de 95%, estimando-se os seus extremos, da seguinte forma:

$$\bar{\hat{x}} \pm t \sqrt{Var_1(\hat{x})}$$

$$\bar{\hat{x}} \pm t \sqrt{Var_2(\hat{x})}$$

$$\bar{\hat{x}} \pm t \sqrt{Var_3(\hat{x})}$$

sendo t=2.

Procedeu-se também à construção dos histogramas de frequência para verificar graficamente se a distribuição dos pontos críticos aproxima-se de uma distribuição normal.

## 4 RESULTADOS E DISCUSSÃO

A seguir, apresenta-se o procedimento teórico para obtenção do ponto crítico de uma equação de regressão quadrática.

### 4.1 Modelo de regressão polinomial quadrático

Considerando-se o seguinte modelo de regressão polinomial quadrático:

$$y_i = a + bx_i + cx_i^2 + e_i \quad i = 1, \dots, n$$

em que:

$y_i$  representa o  $i$ -ésimo valor observado de uma variável dependente,

$x_i$  representa o  $i$ -ésimo valor fixo de uma variável dependente,

$a$ ,  $b$ ,  $c$  são os parâmetros da equação de regressão,

$e_i$  representa o  $i$ -ésimo erro aleatório, associado a observação  $y_i$ , considerados como independentes e normalmente distribuídos com média zero e variância constante  $\sigma^2$ .

As equações para as  $n$  observações são:

$$\begin{aligned} y_1 &= a + bx_1 + cx_1^2 + e_1 \\ y_2 &= a + bx_2 + cx_2^2 + e_2 \\ &\vdots \\ y_n &= a + bx_n + cx_n^2 + e_n \end{aligned}$$

Em termos matriciais, é possível escrevê-las como:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

ou, ainda:

$$Y = X\beta + \varepsilon,$$

no qual:

$Y$  é um vetor de realizações de variáveis aleatórias (observações da variável dependente), de dimensões  $n \times 1$ ;

$X$  é a matriz dos coeficientes associados aos parâmetros ou matriz do planejamento, de dimensões  $n \times 3$ ;

$\beta$  é o vetor de parâmetros, de dimensões  $3 \times 1$ ;

$\varepsilon$  é o vetor de variáveis aleatórias não observáveis (erros experimentais), com dimensões  $n \times 1$ , os quais são assumidos serem independentes e normalmente distribuídos com  $\varepsilon \sim N(\phi, I\sigma^2)$ .

O sistema de equações normais pode ser obtido como em Draper e Smith (1998), pela aplicação do método dos quadrados mínimos. Assim, o vetor de erros é:

$$\varepsilon = Y - X\beta$$

A função  $Z$ , que representa a soma de quadrados dos erros fica:

$$Z = \varepsilon' \varepsilon = (Y - X\beta)'(Y - X\beta) = Y'Y - 2\beta' X'Y + \beta' X' X \beta$$

O ponto crítico desta função pode ser obtido como sendo os valores de  $\beta$  que tornam nula a derivada de  $Z$  em relação a  $\beta$ . Assim,

$$dZ = -2d\beta' X'Y + d\beta' X' X \beta$$

Pode-se mostrar que os valores de  $\beta$  que minimizam a soma de quadrados dos resíduos é a solução do seguinte sistema:

$$X' X \hat{\beta} = X' Y$$

que é conhecido como sistema de equações normais, SEN (Searle, 1971). Para o caso do modelo de regressão quadrático, aqui tratado, as matrizes que constituem o SEN são as seguintes:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ x_1 & x_2 & \dots & \dots & x_n \\ x_1^2 & x_2^2 & \dots & \dots & x_n^2 \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

$$X'X = \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ x_1 & x_2 & \dots & \dots & x_n \\ x_1^2 & x_2^2 & \dots & \dots & x_n^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 y_i \end{bmatrix}$$

Assim, o sistema de equações normais  $X'X\hat{\beta} = X'Y$  tem a seguinte forma:

$$\begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 y_i \end{bmatrix}$$

Segundo Neter, Wasserman e Kutner (1990), o sistema de equações normais pode também ser escrito da seguinte forma:

$$\begin{aligned}\sum_{i=1}^n y_i &= \hat{a}n + \hat{b}\sum_{i=1}^n x_i + \hat{c}\sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i y_i &= \hat{a}\sum_{i=1}^n x_i + \hat{b}\sum_{i=1}^n x_i^2 + \hat{c}\sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 y_i &= \hat{a}\sum_{i=1}^n x_i^2 + \hat{b}\sum_{i=1}^n x_i^3 + \hat{c}\sum_{i=1}^n x_i^4\end{aligned}$$

Dado que a matriz  $X$  possui posto coluna completo, então  $X'X$  é uma matriz positiva definida e, portanto, invertível. Assim, como em Neter, Wasserman e Kutner (1990), a solução única é dada por:

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

Neste caso, o vetor  $\hat{\beta}$ , além de ser solução única, apresenta também status de estimador e todos os seus componentes são individualmente estimáveis (Searle, 1971).

A matriz inversa de  $X'X$  é obtida por:

$$(X'X)^{-1} = \begin{bmatrix} A & B & C \\ B & E & F \\ C & F & K \end{bmatrix}$$

As expressões necessárias para determinação da matriz inversa

$$X'X = \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{bmatrix}$$

são dadas, segundo Lipschutz (1972), por:

$$D = \det[X'X] = n \sum_{i=1}^n x_i^2 \sum_{i=1}^n x_i^4 + 2 \sum_{i=1}^n x_i \sum_{i=1}^n x_i^2 \sum_{i=1}^n x_i^3 - n \left( \sum_{i=1}^n x_i^3 \right)^2 - \left( \sum_{i=1}^n x_i \right)^2 \sum_{i=1}^n x_i^4 - \left( \sum_{i=1}^n x_i^2 \right)^3$$

Fazendo:

$$A = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n x_i^4 - \left( \sum_{i=1}^n x_i^3 \right)^2}{D}$$

$$B = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n x_i^4 - \sum_{i=1}^n x_i^2 \sum_{i=1}^n x_i^3}{D}$$

$$C = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n x_i^3 - \left( \sum_{i=1}^n x_i^2 \right)^2}{D}$$

$$E = \frac{n \sum_{i=1}^n x_i^4 - \left( \sum_{i=1}^n x_i^2 \right)^2}{D}$$

$$F = \frac{n \sum_{i=1}^n x_i^3 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i^2}{D}$$

$$K = \frac{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}{D}$$

## 4.2 Análise de variância

Obtida a solução para o sistema de equações normais  $X'X\hat{\beta} = X'Y$ , pode-se determinar as somas de quadrados do modelo quadrático adotado e que compõem a análise de variância.

De acordo com o modelo adotado, sabe-se que a soma de quadrados do resíduo é dada por:

$$SQRes = Y'Y - \hat{\beta}'X'Y,$$

no qual:

$Y'Y$  é definido como a soma de quadrados total e

$\hat{\beta}'X'Y$  é definida como a soma de quadrados de parâmetros ou a redução devida aos parâmetros.

Somando e subtraindo a constante  $C = \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = Y'JY$ , comumente conhecida como correção, tem-se:

$$SQRes = Y'Y - C - \hat{\beta}'X'Y + C = (Y'Y - C) - (\hat{\beta}'X'Y - C)$$

sendo que  $Y'Y - C = SQTot$  e  $\hat{\beta}'X'Y - C = SQPar$ .

O esquema de análise de variância para o modelo de regressão polinomial quadrático, mostrando o número de graus de liberdade, expressões para o cálculo das somas de quadrados, quadrados médios e teste "F" para regressão, está apresentado na Tabela 2.

TABELA 2 Esquema de análise de variância para um modelo de regressão polinomial quadrático

F.V.	G.L.	S.Q.	Q.M.	F
Regressão	2	$\hat{\beta}'X'Y - C$	$\frac{SQReg}{2}$	$\frac{QMReg}{QMRes}$
Resíduo	n-3	$Y'Y - \hat{\beta}'X'Y$	$\frac{SQres}{n-3}$	
Total	n-1	$Y'Y - C$		

Desenvolvendo as expressões anteriores, as somas de quadrados também podem ser obtidas da seguinte forma:

$$SQTot = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = Y'Y - Y'JY = Y'Y - C$$

que é a soma de quadrados total corrigida pela média, e a soma de quadrados de regressão é:

$$SQReg = \hat{\beta}' X'Y - C = \begin{bmatrix} \hat{a} & \hat{b} & \hat{c} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 y_i \end{bmatrix} - C$$

$$SQReg = \hat{a} \sum_{i=1}^n y_i + \hat{b} \sum_{i=1}^n x_i y_i + \hat{c} \sum_{i=1}^n x_i^2 y_i - C$$

logo, a soma de quadrados do resíduo é:

$$SQRes = SQTot - SQReg = \sum_{i=1}^n y_i^2 - \hat{a} \sum_{i=1}^n y_i - \hat{b} \sum_{i=1}^n x_i y_i - \hat{c} \sum_{i=1}^n x_i^2 y_i$$

Da definição de variância e covariância (Searle, 1971), obteve-se a matriz de variâncias e covariâncias

$$Var[\hat{\beta}] = Var[(X'X)^{-1} X'Y] = (X'X)^{-1} X' Var[Y] X (X'X)^{-1}$$

como  $Var[Y] = I\sigma^2$ , tem-se que:

$$Var[\hat{\beta}] = (X'X)^{-1} X' I\sigma^2 X (X'X)^{-1} = (X'X)^{-1} \sigma^2.$$

Assim, o estimador da matriz de variâncias e covariâncias é:

$$Var[\hat{\beta}] = QMRes (X'X)^{-1}$$



ou, ainda:

$$\text{Vâr}[\hat{\beta}] = \begin{bmatrix} \text{Vâr}[\hat{a}] & \text{Côv}[\hat{a}, \hat{b}] & \text{Côv}[\hat{a}, \hat{c}] \\ \text{Côv}[\hat{b}, \hat{a}] & \text{Vâr}[\hat{b}] & \text{Côv}[\hat{b}, \hat{c}] \\ \text{Côv}[\hat{c}, \hat{a}] & \text{Côv}[\hat{c}, \hat{b}] & \text{Vâr}[\hat{c}] \end{bmatrix}$$

### 4.3 Obtenção do ponto crítico

O ponto crítico da equação de regressão quadrática foi obtido derivando-se a função

$$\hat{y}_i = \hat{a} + \hat{b}x_i + \hat{c}x_i^2$$

em relação a  $x$  e fazendo sua derivada igual a zero,

$$\frac{d\hat{y}_i}{dx} = \frac{d(\hat{a} + \hat{b}x_i + \hat{c}x_i^2)}{dx} = \hat{b} + 2\hat{c}x_i = 0,$$

Assim, obtém-se a abscissa do estimador do ponto crítico da equação de regressão quadrática

$$\hat{x}_i = \frac{\hat{b}}{-2\hat{c}}$$

A ordenada do estimador do ponto crítico é obtida substituindo-se a expressão encontrada para a abscissa na equação de regressão:

$$\hat{y}_i = \hat{a} + \hat{b}x_i + \hat{c}x_i^2$$

$$\hat{y}_i = \hat{a} + \hat{b}\left(\frac{\hat{b}}{-2\hat{c}}\right) + \hat{c}\left(\frac{\hat{b}}{-2\hat{c}}\right)^2 = \hat{a} - \frac{\hat{b}^2}{2\hat{c}} + \frac{\hat{b}^2}{4\hat{c}} = \hat{a} - \frac{\hat{b}^2}{4\hat{c}}.$$

Assim, o estimador do ponto crítico em uma equação de regressão quadrática é dado por:

$$P = (\hat{x}_i, \hat{y}_i) = \left( \frac{\hat{b}}{-2\hat{c}}, \hat{a} - \frac{\hat{b}^2}{4\hat{c}} \right)$$

#### 4.4 Esperança e variância para o quociente de variáveis aleatórias

Segundo Mood, Graybill e Boes (1974), é possível obter uma expressão para a esperança e variância do quociente de duas variáveis aleatórias por meio da expansão da função  $f(x, y) = \frac{x}{y}$  em série de Taylor.

##### 4.4.1 Esperança do quociente de duas variáveis aleatórias

Considerando-se a função  $f(x, y) = \frac{x}{y}$  tal que  $x$  e  $y$  são variáveis aleatórias, é possível escrever sua expansão em série de Taylor em torno de  $(\mu_x, \mu_y)$ . Assim, de acordo com Lima (1993), e Roussas (1973), a expansão de  $f(x, y)$  fica

$$\begin{aligned} f(x, y) = & f(\mu_x, \mu_y) + \frac{\partial f}{\partial x}(\mu_x, \mu_y)(x - \mu_x) + \frac{\partial f}{\partial y}(\mu_x, \mu_y)(y - \mu_y) + \\ & + \frac{1}{2!} \left[ \frac{\partial^2 f}{\partial x^2}(\mu_x, \mu_y)(x - \mu_x)^2 + \frac{\partial^2 f}{\partial y \partial x}(\mu_x, \mu_y)(x - \mu_x)(y - \mu_y) + \right. \\ & \left. + \frac{\partial^2 f}{\partial y^2}(\mu_x, \mu_y)(y - \mu_y)^2 + \frac{\partial^2 f}{\partial x \partial y}(\mu_x, \mu_y)(x - \mu_x)(y - \mu_y) \right] + \dots \end{aligned}$$

Fazendo as derivadas parciais da função e aplicando no ponto  $(\mu_x, \mu_y)$ ,

tem-se:

$$\frac{\partial f}{\partial x}(x, y) = \frac{1}{y} \quad \Rightarrow \quad \frac{\partial f}{\partial x}(\mu_x, \mu_y) = \frac{1}{\mu_y}$$

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = \frac{-1}{y^2} \quad \Rightarrow \quad \frac{\partial^2 f}{\partial y \partial x}(\mu_x, \mu_y) = \frac{-1}{\mu_y^2}$$

$$\frac{\partial^2 f}{\partial x \partial x}(x, y) = 0 \quad \Rightarrow \quad \frac{\partial^2 f}{\partial x \partial x}(\mu_x, \mu_y) = 0$$

$$\frac{\partial f}{\partial y}(x, y) = \frac{-x}{y^2} \quad \Rightarrow \quad \frac{\partial f}{\partial x}(\mu_x, \mu_y) = \frac{-\mu_x}{\mu_y^2}$$

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{-1}{y^2} \quad \Rightarrow \quad \frac{\partial^2 f}{\partial x \partial y}(\mu_x, \mu_y) = \frac{-1}{\mu_y^2}$$

$$\frac{\partial^2 f}{\partial y \partial y}(x, y) = \frac{2xy}{y^4} \quad \Rightarrow \quad \frac{\partial^2 f}{\partial y \partial x}(\mu_x, \mu_y) = \frac{2\mu_x \mu_y}{\mu_y^4}$$

Substituindo-se as derivadas parciais é possível reescrever a função expandida em série de Taylor, da seguinte forma:

$$f(x, y) = \frac{\mu_x}{\mu_y} + \frac{1}{\mu_y}(x - \mu_x) - \frac{\mu_x}{\mu_y^2}(y - \mu_y) - \\ - \frac{1}{\mu_y^2}(x - \mu_x)(y - \mu_y) + \frac{\mu_x}{\mu_y^3}(y - \mu_y)^2 + \dots$$

Fazendo-se uma aproximação da fórmula de Taylor para funções com duas variáveis, tem-se:

$$f(x, y) \approx \frac{\mu_x}{\mu_y} + \frac{1}{\mu_y}(x - \mu_x) - \frac{\mu_x}{\mu_y^2}(y - \mu_y) - \\ - \frac{1}{\mu_y^2}(x - \mu_x)(y - \mu_y) + \frac{\mu_x}{\mu_y^3}(y - \mu_y)^2$$

Aplicando o operador esperança em ambos os lados da igualdade na função aproximada, tem-se:

$$E[f(x,y)] \approx E\left[\frac{x}{y}\right] \approx E\left[\frac{\mu_x}{\mu_y} + \frac{1}{\mu_y}(x - \mu_x) - \frac{\mu_x}{\mu_y^2}(y - \mu_y) - \frac{1}{\mu_y^2}(x - \mu_x)(y - \mu_y) + \frac{\mu_x}{\mu_y^3}(y - \mu_y)^2\right]$$

Separando os termos, fica:

$$E\left[\frac{x}{y}\right] \approx E\left[\frac{\mu_x}{\mu_y}\right] + E\left[\frac{1}{\mu_y}(x - \mu_x)\right] - E\left[\frac{\mu_x}{\mu_y^2}(y - \mu_y)\right] - E\left[\frac{1}{\mu_y^2}(x - \mu_x)(y - \mu_y)\right] + E\left[\frac{\mu_x}{\mu_y^3}(y - \mu_y)^2\right]$$

Utilizando propriedades da esperança, obtém-se:

$$E\left[\frac{x}{y}\right] \approx \frac{\mu_x}{\mu_y} + \frac{1}{\mu_y}E[(x - \mu_x)] - \frac{\mu_x}{\mu_y^2}E[(y - \mu_y)] - \frac{1}{\mu_y^2}E[(x - \mu_x)(y - \mu_y)] + \frac{\mu_x}{\mu_y^3}E[(y - \mu_y)^2]$$

ou, ainda,

$$E\left[\frac{x}{y}\right] \approx \frac{\mu_x}{\mu_y} - \frac{1}{\mu_y^2}\text{Cov}[x,y] + \frac{\mu_x}{\mu_y^3}\text{Var}[y]$$

que é idêntica à expressão contida em Mood, Graybill e Boes (1974).

#### 4.4.2 Variância do quociente de duas variáveis aleatórias

A partir da definição de variância, Mood, Graybill e Boes (1974)

afirmam que

$$\text{Var}\left[\frac{x}{y}\right] = E\left[\left(\frac{x}{y}\right)^2\right] - \left(E\left[\frac{x}{y}\right]\right)^2$$

Sabe-se do resultado anterior que:

$$E\left[\frac{x}{y}\right] \approx \frac{\mu_x}{\mu_y} - \frac{1}{\mu_y^2} \text{Cov}[x, y] + \frac{\mu_x}{\mu_y^3} \text{Var}[y].$$

Logo, necessita-se determinar a esperança matemática de  $\left(\frac{x}{y}\right)^2$ . Com a

finalidade de obter uma forma aproximada para o valor esperado  $E\left[\left(\frac{x}{y}\right)^2\right]$ ,

considerou-se

$$f(x, y) = \frac{x^2}{y^2}.$$

Expandindo a função em série de Taylor em torno do ponto  $(\mu_x, \mu_y)$ ,

segue que  $f(x, y)$  tem a forma:

$$\begin{aligned} f(x, y) &= f(\mu_x, \mu_y) + \frac{\partial f}{\partial x}(\mu_x, \mu_y)(x - \mu_x) + \frac{\partial f}{\partial y}(\mu_x, \mu_y)(y - \mu_y) + \\ &+ \frac{1}{2!} \left[ \frac{\partial^2 f}{\partial x^2}(\mu_x, \mu_y)(x - \mu_x)^2 + \frac{\partial^2 f}{\partial y \partial x}(\mu_x, \mu_y)(x - \mu_x)(y - \mu_y) + \right. \\ &\left. + \frac{\partial^2 f}{\partial y^2}(\mu_x, \mu_y)(y - \mu_y)^2 + \frac{\partial^2 f}{\partial x \partial y}(\mu_x, \mu_y)(x - \mu_x)(y - \mu_y) \right] + \dots \end{aligned}$$

Obtendo-se as derivadas parciais de 1ª e 2ª ordem da função em relação a  $x$  e em relação a  $y$  e, ainda, as derivadas parciais mistas de 2ª ordem, e aplicando-as no ponto  $(\mu_x, \mu_y)$ , tem-se:

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \frac{2x}{y^2} & \Rightarrow & \frac{\partial f}{\partial x}(\mu_x, \mu_y) = 2 \frac{\mu_x}{\mu_y^2} \\ \frac{\partial^2 f}{\partial y \partial x}(x, y) &= \frac{-4x}{y^3} & \Rightarrow & \frac{\partial^2 f}{\partial y \partial x}(\mu_x, \mu_y) = \frac{-4\mu_x}{\mu_y^3} \\ \frac{\partial^2 f}{\partial x \partial x}(x, y) &= \frac{2}{y^2} & \Rightarrow & \frac{\partial^2 f}{\partial x \partial x}(\mu_x, \mu_y) = \frac{2}{\mu_y^2} \\ \frac{\partial f}{\partial y}(x, y) &= \frac{-2x^2}{y^3} & \Rightarrow & \frac{\partial f}{\partial y}(\mu_x, \mu_y) = \frac{-2\mu_x^2}{\mu_y^3} \\ \frac{\partial^2 f}{\partial x \partial y}(x, y) &= \frac{-4x}{y^3} & \Rightarrow & \frac{\partial^2 f}{\partial x \partial y}(\mu_x, \mu_y) = \frac{-4\mu_x}{\mu_y^3} \\ \frac{\partial^2 f}{\partial y \partial y}(x, y) &= \frac{6x^2}{y^4} & \Rightarrow & \frac{\partial^2 f}{\partial y \partial y}(\mu_x, \mu_y) = \frac{6\mu_x^2}{\mu_y^4} \end{aligned}$$

Substituindo-se as derivadas parciais de  $f(x, y)$  na expansão em série de Taylor, tem-se:

$$\begin{aligned} f(x, y) &= \left(\frac{x}{y}\right)^2 = \frac{\mu_x^2}{\mu_y^2} + \frac{2\mu_x}{\mu_y^2}(x - \mu_x) - \frac{2\mu_x^2}{\mu_y^3}(y - \mu_y) + \frac{1}{\mu_y^2}(x - \mu_x)^2 - \\ &\quad - \frac{4\mu_x}{\mu_y^3}(x - \mu_x)(y - \mu_y) + \frac{3\mu_x^2}{\mu_y^4}(y - \mu_y)^2 + \dots \end{aligned}$$

Fazendo-se uma aproximação da fórmula de Taylor para funções com duas variáveis, tem-se:

$$f(x, y) \approx \left(\frac{x}{y}\right)^2 \approx \frac{\mu_x^2}{\mu_y^2} + \frac{2\mu_x}{\mu_y^2}(x - \mu_x) - \frac{2\mu_x^2}{\mu_y^3}(y - \mu_y) + \frac{1}{\mu_y^2}(x - \mu_x)^2 - \\ - \frac{4\mu_x}{\mu_y^3}(x - \mu_x)(y - \mu_y) + \frac{3\mu_x^2}{\mu_y^4}(y - \mu_y)^2$$

Aplicando-se o operador esperança matemática e utilizando-se de suas propriedades, obtém-se:

$$E\left[\left(\frac{x}{y}\right)^2\right] \approx \frac{\mu_x^2}{\mu_y^2} + \frac{2\mu_x}{\mu_y^2}E[(x - \mu_x)] - \frac{2\mu_x^2}{\mu_y^3}E[(y - \mu_y)] + \frac{1}{\mu_y^2}E[(x - \mu_x)^2] - \\ - \frac{4\mu_x}{\mu_y^3}E[(x - \mu_x)(y - \mu_y)] + \frac{3\mu_x^2}{\mu_y^4}E[(y - \mu_y)^2]$$

Portanto,

$$E\left[\left(\frac{x}{y}\right)^2\right] \approx \frac{\mu_x^2}{\mu_y^2} + \frac{1}{\mu_y^2}\text{Var}[x] + \frac{3\mu_x^2}{\mu_y^4}\text{Var}[y] - \frac{4\mu_x}{\mu_y^3}\text{Cov}[x, y]$$

Da definição, de variância,

$$\text{Var}\left[\frac{x}{y}\right] = E\left[\left(\frac{x}{y}\right)^2\right] - \left(E\left[\frac{x}{y}\right]\right)^2$$

substituindo-se as expressões obtidas, tem-se

$$\text{Var}\left[\frac{x}{y}\right] \approx \frac{\mu_x^2}{\mu_y^2} + \frac{1}{\mu_y^2}\text{Var}[x] + \frac{3\mu_x^2}{\mu_y^4}\text{Var}[y] - \frac{4\mu_x}{\mu_y^3}\text{Cov}[x, y] - \\ - \left(\frac{\mu_x}{\mu_y} - \frac{1}{\mu_y^2}\text{Cov}[x, y] + \frac{\mu_x}{\mu_y^3}\text{Var}[y]\right)^2$$

Fazendo-se as simplificações necessárias, verificou-se que:

$$\text{Var} \begin{bmatrix} x \\ y \end{bmatrix} \approx \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}^2 \left( \frac{1}{\mu_x^2} \text{Var}[x] + \frac{1}{\mu_y^2} \text{Var}[y] - \frac{2}{\mu_x \mu_y} \text{Cov}[x, y] \right)$$

que é a expressão dada por Mood, Graybill e Boes (1974) para o cálculo da variância de um quociente de duas variáveis aleatórias.

Assim, a variância do estimador do ponto crítico de uma equação polinomial quadrática, que é um quociente de duas variáveis aleatórias, pode ser escrita como:

$$\text{Var} \begin{bmatrix} \hat{b} \\ -2\hat{c} \end{bmatrix} = \text{Var} \left[ \frac{-1}{2} \cdot \frac{\hat{b}}{\hat{c}} \right] = \left( \frac{-1}{2} \right)^2 \cdot \text{Var} \left[ \frac{\hat{b}}{\hat{c}} \right] = \frac{1}{4} \text{Var} \left[ \frac{\hat{b}}{\hat{c}} \right]$$

Logo,

$$\text{Var} \begin{bmatrix} \hat{b} \\ -2\hat{c} \end{bmatrix} = \frac{1}{4} \text{Var} \left[ \frac{\hat{b}}{\hat{c}} \right] = \frac{1}{4} \left[ \left( \frac{\mu_{\hat{b}}}{\mu_{\hat{c}}} \right)^2 \left( \frac{\text{Var}[\hat{b}]}{\mu_{\hat{b}}^2} + \frac{\text{Var}[\hat{c}]}{\mu_{\hat{c}}^2} - \frac{2\text{Cov}[\hat{b}, \hat{c}]}{\mu_{\hat{b}} \mu_{\hat{c}}} \right) \right]$$

Particularmente, para o caso em que a covariância entre  $\hat{b}$  e  $\hat{c}$  seja nula, a expressão anterior pode ser simplificada, do seguinte modo:

$$\begin{aligned} \text{Var} \begin{bmatrix} \hat{b} \\ -2\hat{c} \end{bmatrix} &= \frac{1}{4} \left[ \left( \frac{\mu_{\hat{b}}}{\mu_{\hat{c}}} \right)^2 \left( \frac{\text{Var}[\hat{b}]}{\mu_{\hat{b}}^2} + \frac{\text{Var}[\hat{c}]}{\mu_{\hat{c}}^2} \right) \right] = \\ &= \frac{1}{4} \left[ \frac{\mu_{\hat{b}}^2}{\mu_{\hat{c}}^2} \cdot \frac{\text{Var}[\hat{b}]}{\mu_{\hat{b}}^2} + \frac{\mu_{\hat{b}}^2}{\mu_{\hat{c}}^2} \cdot \frac{\text{Var}[\hat{c}]}{\mu_{\hat{c}}^2} \right] \end{aligned}$$



$$\therefore \text{Var} \begin{bmatrix} \hat{b} \\ -2\hat{c} \end{bmatrix} = \frac{1}{4} \left[ \frac{\text{Var}[\hat{b}]}{\mu_{\hat{c}}^2} + \frac{\mu_{\hat{b}}^2 \text{Var}[\hat{c}]}{\mu_{\hat{c}}^4} \right],$$

que é exatamente a expressão apresentada por D'Aulio (1976), obtida através da primeira derivada do estimador do ponto crítico em relação a  $x$ .

#### 4.5 Análise dos resultados simulados

Com a finalidade de verificar a aplicação das três expressões para o cálculo da variância do ponto crítico e de se ter uma idéia se os valores simulados do ponto crítico seguem uma distribuição normal, são apresentados a seguir os resultados oriundos de várias simulações realizadas.

Os resultados obtidos para as estatísticas médias dos coeficientes de regressão  $\hat{b}$  e  $\hat{c}$ , bem como suas respectivas variâncias em função das diversas variâncias teóricas ( $\sigma^2$ ) adotadas, estão apresentadas na Tabela 3. Pode-se verificar que os valores médios dos coeficientes de regressão  $\bar{\hat{b}}$  e  $\bar{\hat{c}}$  estão próximos aos valores teóricos reais considerados. Os valores de  $\hat{b}$  variam de 90,1741 a 90,2330 e os valores de  $\hat{c}$  variam de -98,9050 a -99,1258. Notou-se uma tendência de que, com o aumento da variância teórica, esses valores tiveram um certo aumento.

Já para os valores das variâncias de  $\hat{b}$  e  $\hat{c}$ , à medida que a variância teórica ( $\sigma^2$ ) aumentou, seus valores também aumentaram; nota-se que as variâncias  $\text{Var}(\hat{b})$  estão próximas e com tendência na variância teórica ( $\sigma^2$ ); o mesmo ocorrendo para  $\text{Var}(\hat{c})$ , numa proporção próxima de 24,45 vezes, o que já era esperado.

Estes resultados sugerem que os valores de  $\hat{b}$  e  $\hat{c}$  foram simulados corretamente, considerando que essas estatísticas são variáveis aleatórias normais.

TABELA 3 Valores médios dos coeficientes de  $\hat{b}$  e  $\hat{c}$  e respectivas variâncias, em função das diferentes variâncias reais utilizadas.

$\sigma^2$	$\bar{\hat{b}}$	$\bar{\hat{c}}$	$Var(\hat{b})$	$Var(\hat{c})$
0,1	90,1741	-98,9050	0,1008	2,4370
0,5	90,1775	-98,9178	0,5040	12,1851
1	90,1800	-98,9274	1,0080	24,3701
5	90,1908	-98,9678	5,0402	121,8507
10	90,1989	-98,9981	10,0804	243,7013
15	90,2051	-99,0213	15,1206	365,5520
20	90,2103	-99,0409	20,1608	487,4026
50	90,2330	-99,1258	50,4020	1218,5000

Os valores das estimativas  $\bar{\hat{b}}$  e  $\bar{\hat{c}}$ , assim como  $Var(\hat{b})$  e  $Var(\hat{c})$  aumentaram à medida que as variâncias teóricas consideradas aumentaram.

TABELA 4 Valores médios do ponto crítico ( $\bar{\hat{x}}$ ) e variâncias do ponto crítico, calculadas pelos métodos: usual  $Var_1(\hat{x})$ , fórmula deduzida por D'Aulísio (1976)  $Var_2(\hat{x})$ , expressão dada por Mood, Graybill e Boes (1974)  $Var_3(\hat{x})$ .

$\sigma^2$	Ponto crítico		Variâncias		
	$\bar{\hat{x}}$	$Var_1(\hat{x})$	$Var_2(\hat{x})$	$Var_3(\hat{x})$	
0,1	0,4560	0,000054	0,000054	0,000001	
0,5	0,4564	0,000274	0,000271	0,000009	
1	0,4569	0,000555	0,000543	0,000019	
5	0,4616	0,0030	0,0027	0,000097	
10	0,4680	0,0069	0,0054	0,000195	
15	0,4751	0,0123	0,0081	0,000293	
20	0,4835	0,0211	0,0108	0,000390	
50	0,5586	2,7808	0,0270	0,000977	

A Tabela 4 apresenta os valores médios do ponto crítico e as variâncias obtidas pelos três métodos considerados.

Os resultados mostram que houve um aumento nos valores do ponto crítico, à medida que a variância teórica aumentou. Os valores médios do ponto crítico variaram de 0,4560 na variância 0,1 até 0,5586, na maior variância considerada de 50.

Observando-se os valores obtidos para as três variâncias, verificou-se que a variância  $Var_3(\hat{x})$  apresentou valores menores do que os calculados por  $Var_1(\hat{x})$  ou  $Var_2(\hat{x})$ . Pode-se inferir por este, que a expressão  $Var_3(\hat{x})$ , sugerida por Mood, Graybill e Boes (1974), deve ser preferida em lugar das outras duas.

Estes resultados mostram que, para as variâncias consideradas, os valores do ponto crítico não seguem uma distribuição normal Isto porque

diferiram de zero a 1% de probabilidade pelo teste "t" de Student. Para a menor variância estudada, os coeficientes de curtose  $\gamma_2$  não

positiva para todas as variâncias e uma tendência leptocúrtica. incremento das variâncias teóricas consideradas, evidenciando uma assimetria de curtose. Os valores de  $\gamma_1$  e  $\gamma_2$  foram todos positivos e aumentaram com o aumento da variância ( $\sigma^2$ ), houve um aumento do coeficiente de assimetria e respectivos testes "t" estão apresentados na Tabela 5. Verifica-se que, com o Os valores médios obtidos para os coeficientes de assimetria e curtose e

$\sigma^2$	$\gamma_1$	$\gamma_2$	$t_{\gamma_1}$	$t_{\gamma_2}$
0,1	0,0962	0,1311	2,7775(**)	1,8928(NS)
0,5	0,2197	0,2281	6,3447(**)	3,2947(**)
1	0,3157	0,3478	9,1163(**)	5,0232(**)
5	0,7906	1,5481	22,8301(**)	22,3559(**)
10	1,3329	4,3136	38,4902(**)	62,2927(**)
15	2,1221	11,2430	61,2783(**)	162,3599(**)
20	3,8834	37,6100	112,1372(**)	543,1243(**)
50	12,4604	1277,7000	359,8073(**)	18451,0000(**)

TABELA 5 Valores médios dos coeficientes de assimetria, curtose e respectivas estatísticas do teste "t", em função das diferentes variâncias.

Coefficiente de assimetria

Coefficiente de curtose

Valores do teste "t"

apresentam assimetria positiva e curtose do tipo leptocúrtica ou “afilada”, que não são características de uma distribuição normal.

#### **4.6 Distribuição de frequência dos valores do ponto crítico**

A seguir são apresentados, na forma de histogramas, Figuras 1, 2, 3, 4, 5, 6, 7 e 8, os valores simulados do ponto crítico para as diferentes variâncias adotadas.

Pode-se observar, como na Figura 1, que a distribuição de frequência dos valores do ponto crítico não estão muito distantes dos de uma curva normal. Este foi o caso em que o coeficiente de curtose se aproximou do de uma normal, não diferindo de 3 ou de zero quando subtrai-se esta constante.

Nas Figuras 2, 3 e 4, e de acordo com a Tabela 5, pode-se observar uma leve tendência de assimetria à direita. Valores dos coeficientes de assimetria são bem próximos e significativos a 1% de probabilidade pelo teste “t” de Student e coeficiente de curtose também significativo a 1% de probabilidade pelo teste “t”, caracterizando uma distribuição leptocúrtica.

Nas Figuras 5, 6, 7 e 8 observa-se uma forte tendência de assimetria à direita e ao formato leptocúrtico (curtose maior que 3), que já havia sido detectado anteriormente, nas análises realizadas na Tabela 5.

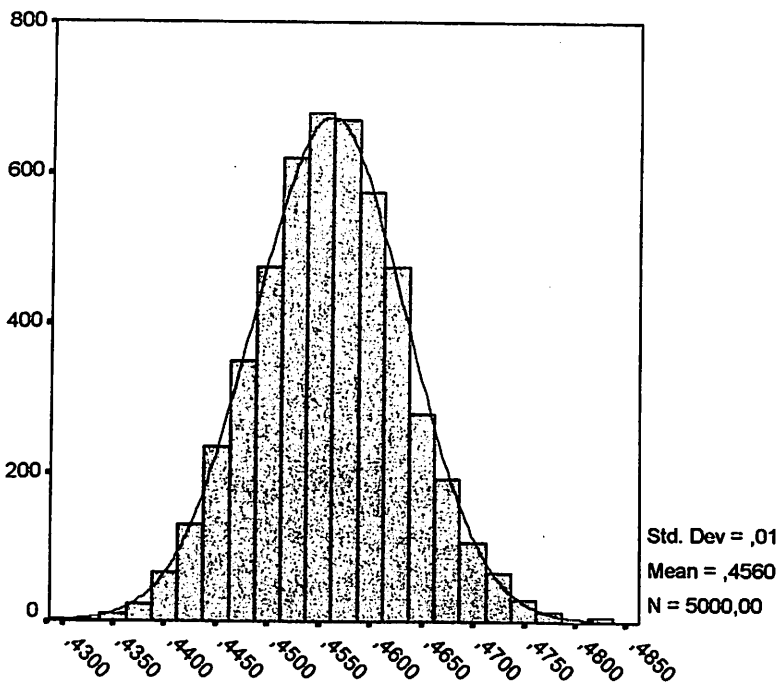


FIGURA 1 Representação gráfica da distribuição de frequência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 0,1$  e valor médio de  $\bar{x} = 0,4560$ .

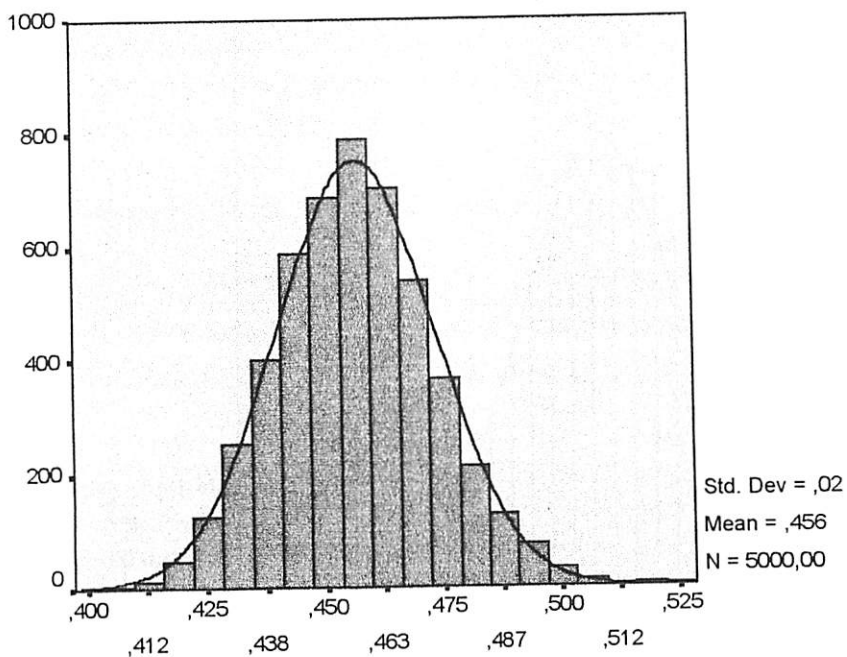


FIGURA 2 Representação gráfica da distribuição de freqüência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 0,5$  e valor médio de  $\bar{x} = 0,4564$ .

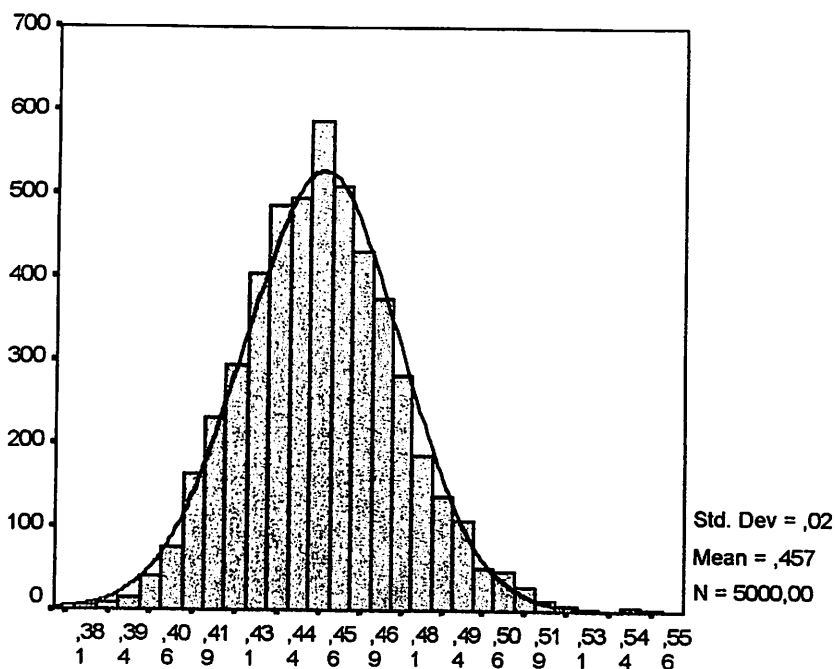


FIGURA 3 Representação gráfica da distribuição de freqüência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 1,0$  e valor médio de  $\bar{\hat{x}} = 0,4569$ .



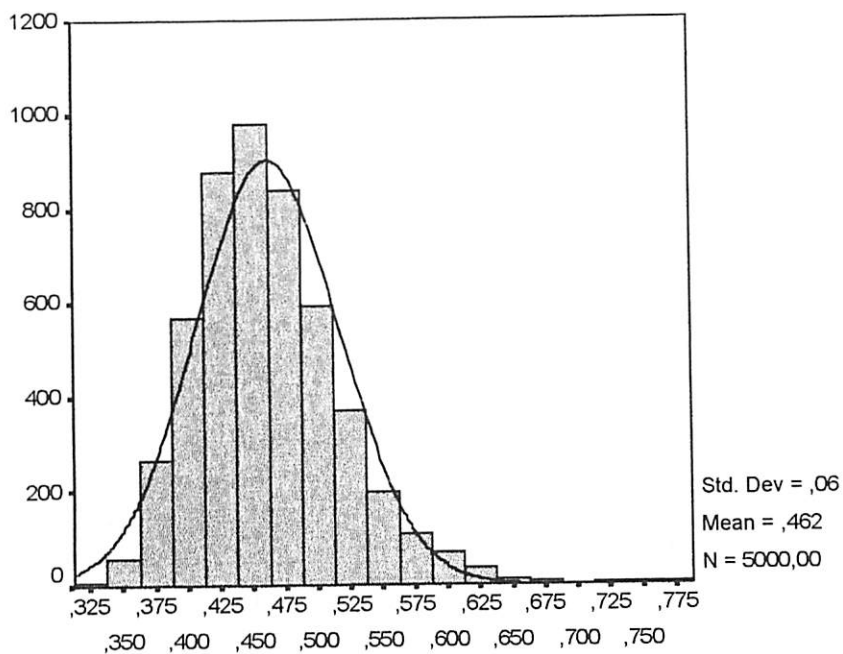


FIGURA 4 Representação gráfica da distribuição de freqüência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 5,0$  e valor médio de  $\bar{x} = 0,4616$ .

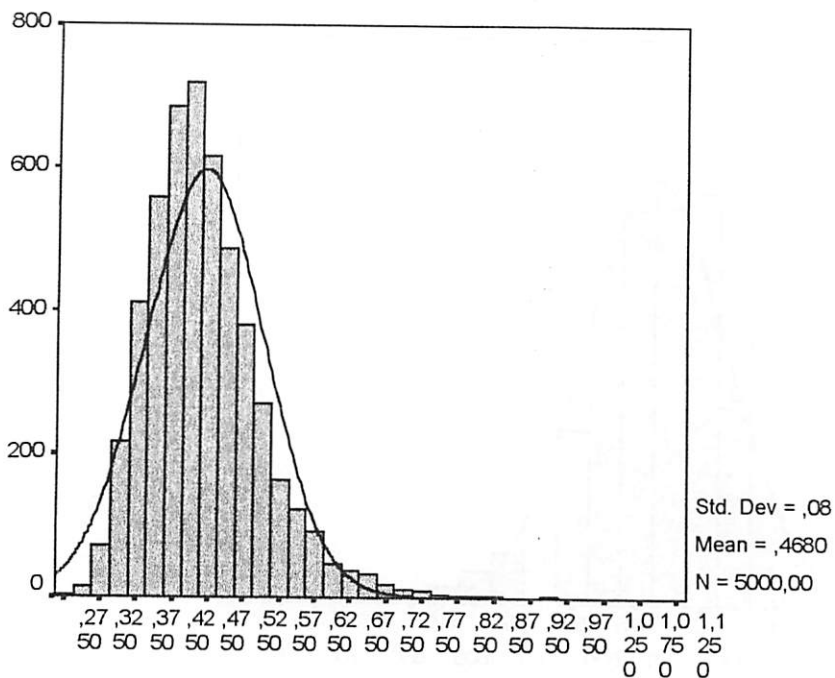


FIGURA 5 Representação gráfica da distribuição de freqüência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 10,0$  e valor médio de  $\bar{x} = 0,4680$ .

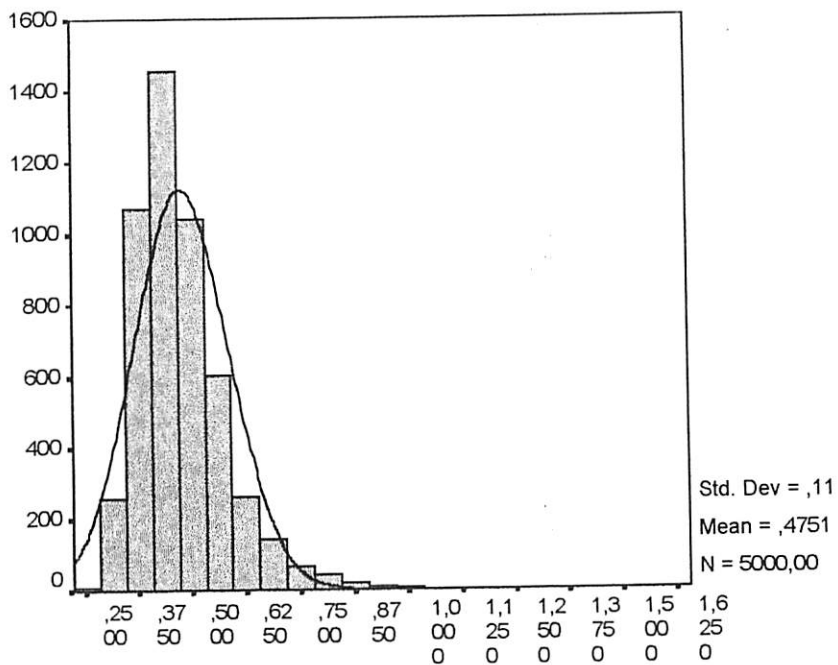


FIGURA 6 Representação gráfica da distribuição de freqüência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 15,0$  e valor médio de  $\bar{x} = 0,4751$ .

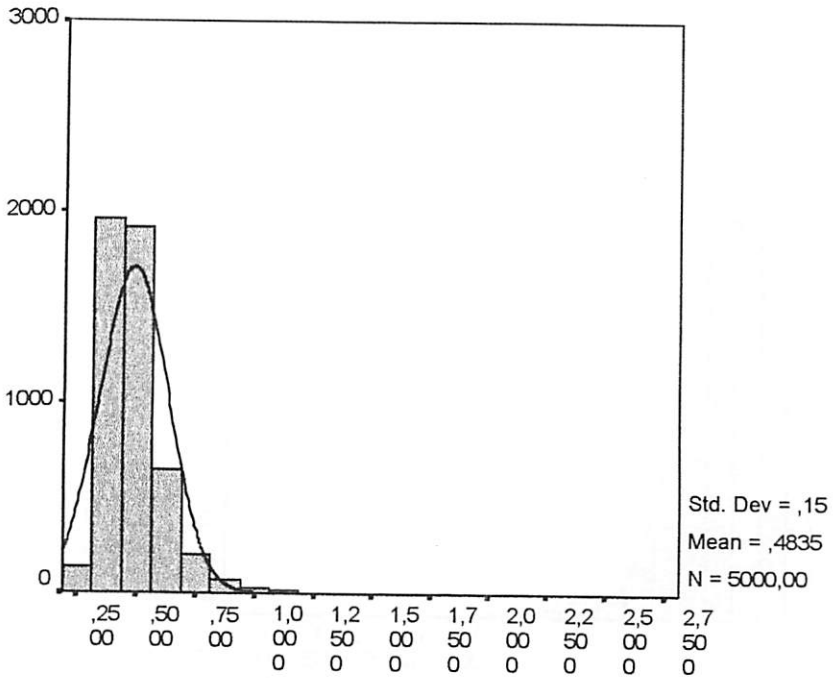


FIGURA 7 Representação gráfica da distribuição de freqüência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 20,0$  e valor médio de  $\bar{x} = 0,4835$ .

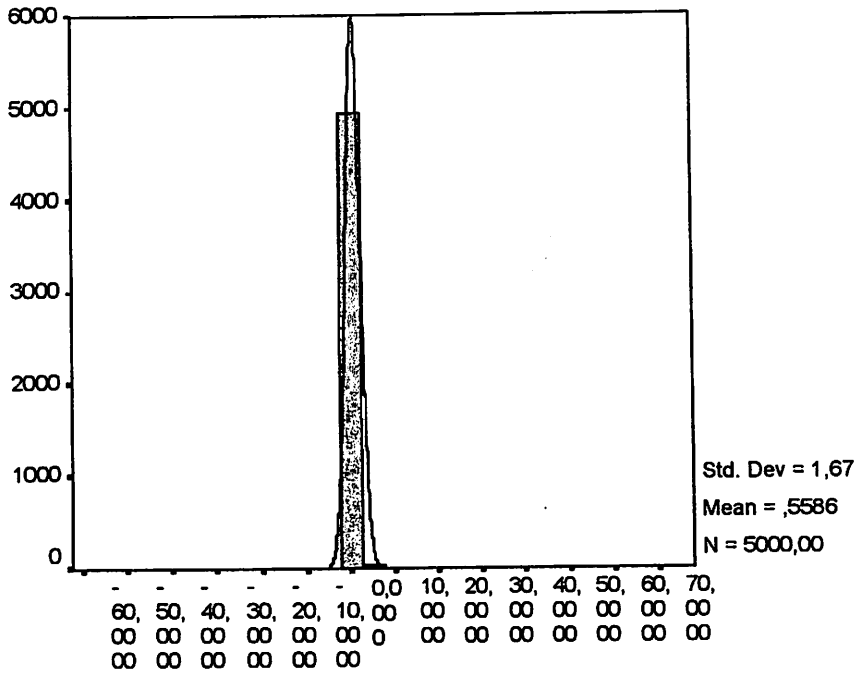


FIGURA 8 Representação gráfica da distribuição de freqüência dos valores simulados do ponto crítico, para a variância  $\sigma^2 = 50,0$  e valor médio de  $\bar{x} = 0,5586$ .

#### 4.7 Intervalos de confiança para o ponto crítico

TABELA 6- Valores médios do ponto crítico e intervalos de confiança para o ponto crítico, considerando as três variâncias, em função das variâncias teóricas.

Variância $\sigma^2$	Ponto crítico médio $\bar{x}$	Intervalos de confiança		
		$Var_1(\hat{x})$	$Var_2(\hat{x})$	$Var_3(\hat{x})$
0,1	0,4560	[0,4412; 0,4707]	[0,4412; 0,4707]	[0,4532; 0,4588]
0,5	0,4564	[0,4232; 0,4896]	[0,4234; 0,4894]	[0,4501; 0,4626]
1	0,4569	[0,4098; 0,5041]	[0,4103; 0,5035]	[0,4481; 0,4658]
5	0,4616	[0,3514; 0,5718]	[0,3574; 0,5657]	[0,4418; 0,4813]
10	0,4680	[0,3016; 0,6343]	[0,3208; 0,6152]	[0,4400; 0,4959]
15	0,4751	[0,2535; 0,6968]	[0,2949; 0,6554]	[0,4409; 0,5094]
20	0,4835	[0,1929; 0,7742]	[0,2755; 0,6916]	[0,4440; 0,5231]
50	0,5586	[-2,7765; 3,8937]	[0,2301; 0,8871]	[0,4961; 0,6211]

Analisando-se os intervalos de confiança para o ponto crítico calculados usando-se as três diferentes variâncias (Tabela 6), nota-se que esses intervalos são bastante satisfatórios, principalmente quando se utilizou a variância  $Var_3(\hat{x})$ .

Como já era esperado, todos os intervalos de confiança tornaram-se menos precisos com o aumento das variâncias teóricas. Desse modo, os intervalos de confiança menos satisfatórios foram aqueles obtidos com  $Var_1(\hat{x})$ , e os intervalos mais satisfatórios obtidos com  $Var_3(\hat{x})$ . Mesmo, neste caso, quando se usou  $Var_3(\hat{x})$ , para variâncias ( $\sigma^2$ ) maiores, existe uma tendência dos intervalos serem mais amplos, perdendo em precisão.

Em todos os casos, considerou-se  $t=2$  com 95% de confiança.

## 5 CONCLUSÕES

As variâncias dos coeficientes de regressão aumentaram proporcionalmente com o aumento das variâncias teóricas ( $\sigma^2$ ), consideradas, caracterizando uma simulação de boa qualidade.

A variância do ponto crítico calculada usando-se a expressão de Mood, Graybill e Boes (1974) apresenta-se como a mais satisfatória pois leva em consideração a covariância entre os coeficientes de regressão  $b$  e  $c$ .

A variância  $\sigma^2$  afeta a estimativa do ponto crítico; maiores variâncias estão relacionados com maiores valores médios.

A distribuição de frequência do ponto crítico não segue a de uma distribuição normal.

O ponto crítico apresenta uma distribuição de frequência com assimetria positiva e formato do tipo leptocúrtico.

Os intervalos de confiança para o ponto crítico tornam-se menos precisos com o aumento das variâncias teóricas ( $\sigma^2$ ) em todos os três casos analisados.

Os intervalos de confiança com menor amplitude foram os obtidos utilizando-se a variância calculada por  $Var_3(\hat{x})$ .

## REFERÊNCIAS BIBLIOGRÁFICAS

CHIARINI, C.A. **Distribuição de probabilidades do quociente de variáveis aleatórias normais: determinação, propriedades, tabela e algumas aplicações.** Piracicaba: ESALQ/USP, 1983. 42p. (Tese - Doutorado em Estatística e Experimentação Agronômica).

COURANT, R. **Cálculo diferencial e integral.** Porto Alegre: Editora Globo S.A., 1963. v.1, 616p.

COURANT, R. **Cálculo diferencial e integral.** Porto Alegre: Editora Globo S.A., 1963. v.2, 685p.

D'AULISIO, M.B.G. **A variância dos pontos de máximo ou de mínimo de equações de regressão de segundo grau.** Piracicaba: ESALQ/USP, 1976. 61p. (Dissertação - Mestrado em Estatística e Experimentação Agronômica).

DRAPER, N.R.; SMITH, H. **Applied regression analysis.** 3.ed. New York: John Wiley & Sons, 1998. 706p.

FIELLER, E.C. **The distribution of the index in a normal bivariate population.** *Biometrika*, London, v.24, p.428-440, 1932.

FINNEY, D.J. **Statistical method in biological assay.** London: Charles Griffin & Company, 1964, 668p.

FISHER, R.A. **The moments of the distribution for normal samples of measures of departure from normality.** *Royal Society, serie A*, v.130, p.17-28, 1930.



FREITAS, A.R. **A distribuição do ponto de máximo ou de mínimo de uma função usada em experimentos de adubação.** Piracicaba: ESALQ/USP, 1978. 81p. (Dissertação - Mestrado em Estatística e Experimentação Agronômica).

GEARY, R.C. The frequency distribution of the quotient of two normal variates. **Journal of the Royal Statistics Society, London, v.93, p.442-446, 1930.**

HOEL, P.G., PORT, S.C., STONE, C.J. **Introdução à teoria da probabilidade.** Rio de Janeiro: Interciência, 1978, 269p.

HOFFMANN, R., VIEIRA, S. **Análise de regressão: uma introdução à econometria.** 3.ed. São Paulo: HUCITEC, 1998. 379p.

KAPLAN, W. **Cálculo avançado.** São Paulo: Edgard Blucher, Ed. da Universidade de São Paulo, v.2, 1972, 750p.

KENDALL, M.G., STUART, A. **The advanced theory of statistics.** London: Charles Griffin & Company, v.1, 1977.

LANG, S. **Cálculo.** Rio de Janeiro: Ao Livro Técnico S.A., v.1, 1976, 388p.

LIMA, E.L. **Análise Real.** Vol.1.; 2.ed. Rio de Janeiro: IMPA, 1993, 189p.

LIPSCHUTZ, S. **Álgebra Linear.** Rio de Janeiro: McGraw-Hill, 1972, 413p.

LEITE, J.G.; SINGER, J.M. **Métodos assintóticos em estatística: fundamentos e aplicações.** In: **Simpósio Brasileiro de Probabilidade e Estatística, 9.IME – USP, São Paulo, 1990, 130p.**

MARSAGLIA, G., Ratios of normal variables and ratios of sums of uniform variables. **Journal of the American Statistical Association**, Washington, v.60, p.193-204, 1965.

MERRIL, A.S. Frequency distribution of an index when both the components follow the normal law. **Biometrika**, London, v.20, p.53-63, 1928.

MOOD, A.M., GRAYBILL, F.A., BOES, D.C. **Introduction to the theory of statistics**. 3 ed. Tokio: McGraw-Hill, 1974. 564p.

NETER, J., WASSERMAN, W., KUTNER, M.H. **Applied linear statistical models: regression, analysis of variance, and experimental designs**. 3.ed. Homewood: Richard D. Irwin, Inc., 1990. 1181p.

ROUSSAS, G.G. **A first course in mathematical statistics**. Reading Mass: Addison-Wesley Publishing Company, 1973, 506p.

SPIEGEL, M.R. **Estatística**. Rio de Janeiro: McGraw-Hill do Brasil Ltda, 1975. 580p.

SEARLE, S.R. **Linear Models**. New York: John Wiley, 1971, 532p.

STEEL, R.G.D.; TORRIE, J.H.; DICKLEY, J.E. **Principles and Procedures of Statistics**. 2 ed. London: McGraw-Hill, 1998, 666p.

## ANEXO

Rotina MATLAB®, programa utilizado para obtenção dos vetores de estimativas de  $\hat{b}$ ,  $\hat{c}$  e  $\hat{x}$ , bem como as variâncias  $Var_1(\hat{x})$ ,  $Var_2(\hat{x})$ ,  $Var_3(\hat{x})$  e os respectivos intervalos de confiança.

Clear

% Vetor de variâncias

variancias=[0.1 0.5 1 5 10 15 20 50]';

% Matriz de doses 0.0, 0.2, 0.4, 0.6 e 0.8 com 4 repetições

X=[1 0.0 (0.0)^2

1 0.0 (0.0)^2

1 0.0 (0.0)^2

1 0.0 (0.0)^2

1 0.2 (0.2)^2

1 0.2 (0.2)^2

1 0.2 (0.2)^2

1 0.2 (0.2)^2

1 0.4 (0.4)^2

1 0.4 (0.4)^2

1 0.4 (0.4)^2

1 0.4 (0.4)^2

1 0.6 (0.6)^2

1 0.6 (0.6)^2

1 0.6 (0.6)^2

1 0.6 (0.6)^2

1 0.8 (0.8)^2

```

1 0.8 (0.8)^2
1 0.8 (0.8)^2
1 0.8 (0.8)^2];
B=inv(X'*X);
xest=zeros(5.000,1);
% Gerar 5.000 números aleatórios normais  $(0, \sigma^2)$ 
A=1;
for ii=1:8
    varia(ii,1)=variâncias(ii,1);
    sigma=sqrt(varia(ii,1));
    randn('state',0);
    erro=normrnd(0,sigma,5000,2);
    n=5000;
% Variáveis aleatórias  $\hat{b}$  e  $\hat{c}$ 
    for i=1:5000
        e11(i,1)=erro(i,1);
        e31(i,1)=erro(i,2)*sqrt(24.4532);
        best1(i,1)=90.1713+e11(i,1);
        cest1(i,1)=-98.8947+e31(i,1);
        xest1(i,1)=best1(i,1)/(-2*cest1(i,1));
    end
% Matriz de valores  $\hat{x}$ 
    xest=[xest xest1];
% Valores médios e variâncias
    bbar1=mean(best1);
    cbar1=mean(cest1);
    xbar1=mean(xest1);

```

```

varb1=std(best1)^2;
varc1=std(cest1)^2;
% Variâncias dos estimadores do ponto crítico
C=B*varia(ii,1);
varx11=std(xest1)^2;
varx21=0.25*((varb1/cbarra1^2)+((bbarra1^2)/cbarra1^4)*varc1);
varx31=1/4*[(90.1713/-98.8947)^2*((C(2,2)/(90.1713)^2)+(C(3,3)/
(-98.8947)^2)-(2*C(2,3)/(90.1713*(-98.8947))))];
% Coeficientes de assimetria e curtose
gama11=skewness(xest1);
gama21=kurtosis(xest1)-3;
% Variância dos coeficientes de assimetria e curtose
vargama11=(6*n*(n-1))/((n-2)*(n+1)*(n+3));
vargama21=(24*n*(n-1)^2)/((n-3)*(n-2)*(n+3)*(n+5));
% Valores de t calculado
tgama11=gama11/sqrt(vargama11);
tgama21=gama21/sqrt(vargama21);
% Intervalos de confiança
Int1=[xbarra1-2*sqrt(varx11) xbarra1+2*sqrt(varx11)];
Int2=[xbarra1-2*sqrt(varx21) xbarra1+2*sqrt(varx21)];
Int3=[xbarra1-2*sqrt(varx31) xbarra1+2*sqrt(varx31)];
cont1=variancias(A,1);
A=A+1;
resultado=[cont1 bbarra1 cbarra1 xbarra1 varb1 varc1 varx11 varx21 varx31
gama11 gama21 vargama11 vargama21 tgama11 tgama21 Int1 Int2 Int3]
end
% Histogramas
hist(xest(:,2));

```