O USO DA REGRESSÃO DE CUMEEIRA EM EXPERIMENTOS AGRONÔMICOS

TIEMI MATSUO

Orientador: Prof. Dr. Décio Barbin

Dissertação apresentada a Escola Superior de Agricultura “Luiz de Queiroz”, da Universidade de São Paulo, para obtenção do título de Mestre em Agronomia, Área de concentração em “Estatística e Experimentação Agronômica”.

PIRACICABA
Estado de São Paulo – Brasil
Junho – 1986

Aos meus pais,

Aos meus irmãos,

As minhas cunhadas,

Aos meus sobrinhos

DEDICO

À família Ferreira Alves

OFEREÇO

AGRADECIMENTOS

Ao Prof. Dr. Décio Barbin, Prof. Titular do Departamento de Matemática e Estatística da ESALQ, pela valiosa orientação, apoio, amizade e estímulo constante.

Ao Prof. Dr. Humberto de Campos, Prof. Titular do Departamento de Matemática e Estatística da ESALQ, pelo apoio concedido durante o curso como Coordenador do Curso de Pós-Graduação em Estatística e Experimentação Agronômica e amizade.

À Família Ferreira Alves e Inês F. Ubukata Yada pelo carinho, amizade, apoio e incentivo.

Aos Professores do Departamento de Matemática e Estatística da ESALQ pelos ensinamentos e amizade.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela concessão da bolsa de estudo.

À Prof^ª Dra. Maria Aparecida Perre, Prof^ª Adjunto do Departamento de Matemática Aplicada da FUEL, pela colaboração e amizade.

Ao Prof. Dr. Octavio Nakano, Prof. Titular do Departamento de Entomologia da ESALQ, pela orientação do estágio, dados experimentais deste trabalho, amizade e incentivo.

Aos funcionários do Departamento de Matemática e Estatística da ESALQ, em especial à Srt^ª Maria Izalina Ferreira

Alves, Srt^a Rosa Maria Alves e Sr. Otávio Frasseto, pela amizade, apoio e incentivo.

Aos Estagiários do Departamento de Entomologia da ESALQ, Setor Entomologia Aplicada (1980 - 1982), pela amizade e incentivo.

À Fundação Universidade Estadual de Londrina através da Coordenadoria de Pesquisa e Pós-Graduação (CPG Projeto N^o 131064/84).

Aos Professores Samuel Fabre Sanches, Ubirajara Dorival Diniz e José Carlos Dalmas do Departamento de Matemática Aplicada.

Aos Professores do Departamento de Matemática Aplicada pelo apoio e amizade.

Ao Professor Jeferson Moriconi Cesário, Diretor do Centro de Ciências Exatas da FUEL pela colaboração e amizade.

À Thelma Herena Negri Borges, Édio Vizoni e Maria Cristina Neves de Oliveira pela amizade e incentivo.

Ao Engenheiro Florestal Valter João Diel pela amizade e orientação no uso do computador.

Aos Funcionários do Centro de Ciências Exatas, da Coordenadoria de Pesquisa e Pós-Graduação e da Editora da Fundação Universidade Estadual de Londrina pela colaboração e amizade.

A todos que de alguma forma colaboraram para a realização do curso de mestrado e deste trabalho.

ÍNDICE

	pag.
RESUMO	xiii
SUMMARY	xv
1. INTRODUÇÃO	01
2. REVISÃO DE LITERATURA	04
2.1. A multicolinearidade	04
2.2. Processos alternativos ao de mínimos quadrados quando ocorre a multicolinearidade	06
3. MATERIAL E MÉTODO	10
3.1. Material	10
3.1.1. Experimento I: "Controle de ácaro da falsa ferrugem (<u>Phyllocoptruta oleivora</u>) (Ashmead, 1879) (Acari, Eriophyidae) em citros (<u>Citrus</u> <u>sinensis</u> var. Natal)"	10
3.1.2. Experimento II: "Efeitos da intensidade da população de <u>Sitophilus oryzae</u> (Linné, 1763) (Coleoptera - Curculionidae) e do pe ríodo de armazenamento sobre diferentes va- riedades de milho"	13
3.2. Métodos	18
3.2.1. Regressão múltipla	18
3.2.2. Regressão de cumeeira	29
4. RESULTADOS E DISCUSSÃO	42
4.1. Dados de Avaliação do número de ácaros da falsa ferrugem (<u>Phyllocoptruta oleivora</u>) nas folhas de citros (<u>Citrus sinensis</u> var. Natal)	43

4.2. Dados de Avaliação do número de ácaros da falsa ferrugem (<u>Phyllocoptruta oleivora</u>) nos frutos de citros (<u>Citrus sinensis</u> var. Natal)	50
4.3. Dados de avaliação de danos em grãos de milho armazenados cultivar C-111, causados por gorgulhos (<u>Sitophilus oryzae</u>)	55
4.4. Dados de avaliação de danos em grãos de milho armazenados cultivar Flint, causados por gorgulhos (<u>Sitophilus oryzae</u>)	65
4.5. Dados de avaliação de danos em grãos de milho armazenados cultivar Piranão, causados por gorgulhos (<u>Sitophilus oryzae</u>)	74
5. CONCLUSÕES	84
6. BIBLIOGRAFIA	86

LISTA DE TABELAS

TABELA		pag.
1	Número médio de ácaros (y) da falsa ferrugem nas folhas por parcela, temperatura em $^{\circ}\text{C}$ (X_1), umidade em % (X_2); precipitação em mm (X_3) e insolação em horas por dia (X_4)	12
2	Número médio de ácaros (y) da falsa ferrugem nos frutos por parcela, temperatura em $^{\circ}\text{C}$ (X_1), umidade em % (X_2), precipitação em mm (X_3) e insolação em horas por dia (X_4)	12
3	Avaliações de danos em grãos de milho cultivar C-111	15
4	Avaliações de danos em grãos de milho cultivar Flint composto	16
5	Avaliação de danos em grãos de milho cultivar Piranão	17
6	Estimativa dos coeficientes de correlação, simples entre as variáveis independentes e dependentes ...	44
7	Estimativa dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,3$)	47
8	Estimativa do erro padrão dos coeficientes de re-	

TABELA

pag.

	gressão linear múltipla e de cumeeira ($k = 0,3$) ..	47
9	Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(6)} = 2,4469$	48
10	Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(6;0,05)} = 2,4469$	49
11	Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira	49
12	Estimativa dos coeficientes de correlação linear simples entre as variáveis independentes e dependentes ...	50
13	Estimativa dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,5$)	53
14	Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,5$) ..	53
15	Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeeira , com $t_{(4)} = 2,7764$	54

TABELA

pag.

16	Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(4;0,05)} = 2,7764$	54
17	Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira	55
18	Estimativa dos coeficientes de correlação linear simples entre as variáveis independentes e dependentes	57
19	Estimativa dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,8$)	60
20	Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,8$) ..	61
21	Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(27)} = 2,0518$	62
22	Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(27; 0,05)} = 2,0518$	63

TABELA

pag.

23	Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cuneeira	64
24	Estimativa dos coeficientes de correlação linear simples entre as variáveis independentes e dependentes	66
25	Estimativa dos coeficientes de regressão linear múltipla e de cuneeira ($k = 1,0$)	69
26	Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cuneeira ($k = 1,0$) ..	70
27	Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cuneeira , com $t_{(27)} = 2,0518$	71
28	Teste "t" de Student para os coeficientes de regressão linear múltipla e de cuneeira, com $t_{(27; 0,05)} = 2,0518$	72
29	Coeficiente de determinação (R^2), soma de quadrados de resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e	

TABELA

pag.

	coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira	73
30	Estimativa dos coeficientes de correlação linear simples entre as variáveis independentes de dependentes	75
31	Estimativa dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,6$)	78
32	Estimativa do erro padrão dos coeficientes de regressão linear múltipla de cumeeira ($k = 0,6$) ..	79
33	Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(27)} = 2,0518$	80
34	Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(27; 0,05)} = 2,0518$	81
35	Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira	82

LISTA DE FIGURAS

Figura		pag.
1	Valores dos coeficientes de regressão para os dados de avaliação do número de ácaros da falsa ferrugem nas folhas de citros, em função dos diversos níveis de k	46
2	Valores dos coeficientes de regressão para os dados de avaliação do número de ácaros da falsa ferrugem nos frutos de citros, em função dos diversos níveis de k	52
3	Valores dos coeficientes de regressão para os dados de avaliações de danos em grãos de milho cultivar C-111, causados por gorgulhos, em função dos diversos níveis de k	59
4	Valores dos coeficientes de regressão para os dados de avaliações de danos em grãos de milho cultivar Flint composto, causados por gorgulhos, em função dos diversos níveis de k	68
5	Valores dos coeficientes de regressão para os dados de avaliações de danos em grãos de milho cultivar Piranão, causados por gorgulhos, em função dos diversos níveis de k	77

O USO DA REGRESSÃO DE CUMEEIRA EM EXPERIMENTOS AGRONÔMICOS

Autora : Tiemi Matsuo

Orientador : Prof. Dr. Décio Barbin

R E S U M O

No presente trabalho foi desenvolvido o método da regressão de cumeeira proposto por HOERL e KENNARD (1970 a,b) como método alternativo ao método da regressão linear múltipla quando as variáveis independentes estão altamente correlacionadas.

Os métodos da regressão de cumeeira e o da regressão linear múltipla foram aplicados aos dados de alguns experimentos agronômicos com o objetivo de compará-los na prática.

Os dados experimentais utilizados referem-se aos parâmetros climáticos e a flutuação da população do ácaro da falsa ferrugem (Phyllocoptruta oleivora) em citros (Citrus sinensis var. Natal) e às avaliações de danos físicos e químicos causados por gorgulhos (Sitophilus oryzae) em três cultivares de grãos de milho armazenados.

Através desses métodos foram obtidos:

a) As estimativas dos coeficientes de regressão;

- b) As estimativas do erro padrão dos coeficientes de regressão;
- c) Os intervalos de confiança dos coeficientes de regressão;
- d) O coeficiente de determinação;
- e) A soma de quadrados dos desvios de regressão;
- f) O teste t ;
- g) O teste F .

Além da aplicação dos dois métodos foram construídos os gráficos de cumeeira com o objetivo de discutir o efeito das altas correlações entre as variáveis independentes, sobre as estimativas dos coeficientes de regressão.

A principal conclusão deste trabalho é que na prática, para os experimentos agrônômicos o método de regressão de cumeeira é uma das alternativas em relação ao método da regressão linear múltipla quando se evidencia um alto grau de multicolinearidade entre as variáveis independentes, principalmente por apresentar estimativas dos coeficientes de regressão com maior precisão.

THE USE OF THE RIDGE REGRESSION IN AGRONOMIC EXPERIMENTS

Candidate : Tiemi Matsuo

Adviser : Prof. Dr. Décio Barbin

SUMMARY

This research developed the ridge regression method, proposed by HOERL e KENNARD (1970 a,b) as an alternative method to the multiple linear regression method, when the predictor variables are highly correlated.

The ridge regression method and the multiple linear regression method were applied to the data obtained from some agronomic experiments, seeking to compare both in practice.

The experimental data used refer to the climatic parameters and fluctuation of population of rust mite (Phyllocoptruta oleivora) in citrus (Citrus sinensis var. Natal) and to the evaluation of physical and chemical caused by weevils (Sitophilus oryzae) in three crops of stored maize grains.

Trough these methods were obtained:

- a) The estimates of the regression coefficients;
- b) The estimates of the standart error of the regression coeffi

cientes;

- c) The confidence intervals of the regression coefficients;
- d) The determination coefficient;
- e) The sum of square of lack of fit;
- f) The "t" test to the regression coefficients;
- g) The "F" test to the regression.

Besides the applications of the two methods , ridge traces were built seeking to discuss the effect of high correlations among the predictor variables, over the regression coefficients estimates.

The main conclusion obtained from this research is that, in practice of agronomic experiments, the ridge regression method is one of the alternatives to the multiple linear regression method, when is evident a high degree of multicollinearity among the predictor variables, mainly for presenting estimates of regression coefficients with greater accuracy.

1. INTRODUÇÃO

A regressão linear múltipla tem sido um dos métodos estatísticos de maior utilização em diversas áreas tais como ciências e tecnologia bem como ciências sociais, economia, finanças e outras, para estabelecer uma função matemática que descreva a dependência entre as variáveis e que possibilite efetuar previsões.

A sua aplicação intensa também é devido à facilidade de compreensão da metodologia e de cálculos, e ainda, hoje existem pacotes de programas estatísticos em qualquer centro de computação que efetuam estas análises com rapidez.

Porém, a simples execução de um programa estatístico desta natureza poderá acarretar em sérios problemas devido às limitações das metodologias clássicas de regressão e que não devem ser esquecidas.

Uma das metodologias clássicas é a de mínimos

quadrados que fornece estimadores não tendenciosos desde que as variáveis sejam ortogonais e as demais pressuposições do modelo matemático sejam levadas em conta.

Mas, em algumas situações, pode-se dizer, é impossível atender à todas as pressuposições devido à própria natureza dos dados, tornando inadequado o método dos mínimos quadrados que irá gerar estimadores tendenciosos.

Desta forma, é preferível em certos casos, adotar um método alternativo, mesmo que forneça estimadores tendenciosos mas, que gerem melhores resultados na prática.

Visando ao problema da não ortogonalidade ou da multicolinearidade, este trabalho irá descrever uma metodologia alternativa proposta por HOERL e KENNARD (1970 a, b): a regressão de cumeeira ("ridge regression") que equivale a uma estimação de regressão com restrição, apresentando em alguns casos práticos melhores estimativas do que o método dos mínimos quadrados.

Este método dá origem a um estimador tendencioso dos coeficientes de regressão mas, produzirá uma estimativa com menor esperança da soma de quadrados dos desvios do que aquela produzida pelo método dos mínimos quadrados além da variância dos coeficientes de regressão também ser menor.

Este trabalho, além de apresentar o método da regressão de cumeeira tem por objetivo:

- a) Divulgar uma metodologia alternativa ao de mínimos quadrados quando existe a multicolinearidade.

- b) Discutir o efeito da multicolinearidade nas análises para dados de um experimento com ácaros da falsa ferrugem (Phyllocoptruta oleivora) em citros (Citrus sinensis var. Natal) e de um com gorgulhos (Sitophilus oryzae), em milho.
- c) Comparar a aplicabilidade do método da regressão múltipla com o método da regressão de cumeeira para os dados experimentais utilizados para verificar o objetivo b.

2. REVISÃO DE LITERATURA

2.1. A Multicolinearidade

A multicolinearidade ocorre quando as variáveis independentes do modelo de regressão linear múltipla estão altamente correlacionadas.

Quando a correlação não é acentuada a precisão das estimativas dos coeficientes de regressão não chega a ser afetada significativamente segundo afirmam CHATTERJEE e PRICE (1977). Porém, quando existe um alto grau de multicolinearidade, as estimativas dos coeficientes de regressão serão de baixa precisão gerando um grande problema no processo de estimação pois, havendo dois fatores positivamente correlacionados, existirá tendência dos respectivos coeficientes de regressão estimados serem viesados.

Quando surgem problemas de multicolinearidade é

comum ocorrer a eliminação de variáveis ou observações por ser um processo elementar que satisfaz a condição da independência entre as variáveis independentes do modelo. Porém, não é muito correto na prática pois este procedimento pode dar origem a estimativas viesadas dos coeficientes no sentido de excluir uma variável relevante para a especificação do modelo conforme conclusão de JOHNSTON (1972). O mesmo autor salienta que as inferências com base nestes coeficientes não são confiáveis, já que a estimativa da variância residual é viesada. Por isso, é preferível incluir variáveis no modelo a excluí-las mas, o ideal consiste em incluir variáveis relevantes e excluir as variáveis irrelevantes do modelo de regressão múltipla.

CHATTERJEE e PRICE (1977) introduzem alguns indicativos para detectar a presença de multicolinearidade quando é adotado o processo de inclusão de variáveis ou observações para obter a equação de regressão adequada. Estes são:

- a) mudanças bruscas nos coeficientes de regressão quando uma variável é incluída ou excluída da equação;
- b) mudanças bruscas nos coeficientes de regressão quando uma observação é eliminada, substituída ou incluída;
- c) os sinais algébricos dos coeficientes de regressão não estão em concordância com os esperados "a priori";
- d) os erros padrões dos coeficientes de regressão são grandes.

Além destes indicativos os autores têm, como me-

dida para presença de multicolinearidade, a soma das recíprocas dos autovalores da matriz de correlação simples é dada por:

$$\psi = \sum_{i=1}^p \frac{1}{\lambda_i}$$

onde λ_i é o i -ésimo auto valor e p é o número de variáveis independentes.

As variáveis apresentam alto grau de multicolinearidade se $\psi > 5p$. Logo, para evitar erros grosseiros nas inferências, baseadas na equação de regressão estimada, convém verificar se a multicolinearidade está presente para evitar uma queda na precisão das estimativas dos coeficientes de regressão.

2.2. Processos alternativos ao de mínimos quadrados quando ocorre a multicolinearidade

Constatada a existência da multicolinearidade, recomenda-se a utilização de processos alternativos ao de mínimos quadrados que permitam desenvolver uma série de coeficientes que darão um bom desempenho nas previsões através da equação de regressão estimada.

CHATTERJEE e PRICE (1977) e WEISBERG (1980) recomendam a aplicação da análise de componentes principais. HUBER (1981) recomenda a utilização de métodos robustos de estimação,

definindo-os como aqueles em que os parâmetros são menos sensíveis às variações dos dados.

SUNG (1981) sugere a técnica que consiste na imposição do que chama "ótimas restrições" na estimação do quadrado médio residual. Sob estas restrições o autor afirma que os estimadores dos coeficientes de regressão de mínimos quadrados são praticamente idênticos aos estimadores de componentes principais.

HOERL e KENNARD (1970 a, b) propõem o uso da regressão de cumeeira que tem como estimativas dos coeficientes de regressão a solução do sistema

$$(X'X + kI_p)\hat{\beta}^* = g \quad \text{isto é,}$$

$$\hat{\beta}^* = (X'X + kI_p)^{-1}g \quad \text{onde } g = X'Y \quad \text{e para } k > 0,$$

normalmente um valor do intervalo real $[0; 1]$. Em geral, existe um valor ótimo de k para um problema mas, é preferível examinar a solução de cumeeira para uma série de valores de k , isto é, aquele que produza menor quadrado médio do desvio nos parâmetros do que a solução de mínimos quadrados.

Esses autores dão as seguintes propriedades aplicáveis à regressão de cumeeira:

a) A soma de quadrados do resíduo é uma função crescente de k ;

b) A soma de quadrados dos desvios de $\hat{\beta}^*$ é uma

função decrescente de k ;

c) $\hat{\beta}^*$ é uma transformação linear de $\hat{\beta}$:

$$\hat{\beta}^* = (X'X + kI_p)^{-1} (X'X) \hat{\beta}$$

logo, $\hat{\beta}^*$ é uma estimativa tendenciosa de β isto é,

$$E(\hat{\beta}^*) = (X'X + kI_p)^{-1} (X'X) \beta$$

$$d) V(\hat{\beta}^*) = (X'X + kI_p)^{-1} X'X(X'X + kI_p)^{-1} \sigma^2$$

$$\begin{aligned} e) E[SQD(\hat{\beta}^*)] &= k^2 \beta' (X'X + kI_p)^{-1} (X'X + kI_p)^{-1} \beta + \\ &\quad + \text{traço}[V(\hat{\beta}^*)] \\ &= \text{viés ao quadrado} + \text{variância}. \end{aligned}$$

Esta propriedade tem dois corolários importantes:

- i) a variância dos coeficientes é uma função decrescente;
- ii) o viés é uma função crescente de k .

g) Se $\beta'\beta$ é limitado, existe um valor $k > 0$ tal que

$$E[SQD(\hat{\beta}^*)] < E[SQD(\hat{\beta})] .$$

Esta última propriedade é também enfatizada por THEOBALD (1974).

MARQUARDT e SNEE (1975) trazem como uma das maiores vantagens da regressão de cumeeira o gráfico que poderá ser construído, e denominado gráfico de cumeeira e que auxilia o es

tatístico a ver quais coeficientes são sensíveis aos dados, ficando a análise de sensibilidade um aspecto importante da regressão.

MARQUARDT e SNEE (1975), HOCKING (1976), BESLEY et alii (1980) também sugerem a utilização do método de regressão de cumeeira pois, os resultados obtidos nos estudos foram similares aos obtidos pelos componentes principais, que produzem melhores estimativas dos coeficientes do que o método dos mínimos quadrados, quando existe alto grau de multicolinearidade.

NEWELL e LEE (1981) verificaram que a regressão de cumeeira tem se mostrado uma alternativa conveniente, não somente superando o problema da inflação e instabilidade das estimativas dos coeficientes de regressão como também possibilitando ao pesquisador estudar mais as interrelações entre as variáveis independentes através do traço de cumeeira.

3. MATERIAL E MÉTODO

3.1. Material

Para a aplicação dos métodos deste trabalho foram utilizados os dados dos seguintes experimentos:

3.1.1. Experimento I: "Controle do ácaro da falsa ferrugem (Phyllocoptruta oleivora) (Ashmead, 1879) (Acari - Eriophyidae) em citros (Citrus sinensis var. Natal).

O experimento foi conduzido na chácara do Sr. Luiz Campassi, localizada no município de Piracicaba, no estado de São Paulo, durante o período de 23/10/83 a 02/04/84, sob a orientação do Prof. Titular Dr. Octávio Nakano, do Departamento de Entomologia da ESALQ/USP.

Deste experimento foram utilizados os dados referentes às parcelas que não receberam produto químico para o controle da referida praga. Foram coletadas de 4 parcelas, sendo

que cada uma constou de 3 plantas, com aproximadamente 55 m^2 de área foliar e os frutos apresentavam, em média 2cm de diâmetro no momento da 1ª avaliação.

a) Para avaliação do número de ácaros da falsa ferrugem nas folhas foram coletadas 30 folhas por parcela, que correspondiam ao 3º ou 4º par de folhas localizadas na altura do peito; posteriormente, foram levadas ao laboratório e submetidas ao escovamento mecânico.

b) Para a avaliação do número de ácaros da falsa ferrugem nos frutos foram coletados 30 frutos por parcela, nos quais era feita a avaliação dos possíveis focos na região inferior, com o auxílio de uma lupa de bolso.

Além do número médio de ácaros da falsa ferrugem (y) foram coletados os dados referentes as seguintes variáveis:

X_1 = temperatura ($^{\circ}\text{C}$)

X_2 = Umidade relativa do ar (%)

X_3 = Precipitação (mm)

X_4 = Insolação (horas/dia)

De todos os dados obtidos somente as variáveis X_2 e y sofreram transformações visando a adequação às exigências do modelo matemático.

Os dados constam nas tabelas 1 e 2.

Tabela 1 - Número médio de ácaros (y) da falsa ferrugem nas folhas por parcela, temperatura em $^{\circ}\text{C}(X_1)$, umidade em % (X_2); precipitação em mm(X_3) e insolação em horas por dia (X_4). Piracicaba (SP)

y	X_1	X_2	X_3	X_4
63,00	23,9	90,2	9,1	0,6
81,00	24,2	81,0	5,7	0,4
48,00	23,0	50,0	0,0	10,2
15,00	24,1	52,8	0,0	11,9
15,00	24,5	65,3	7,0	7,3
0,00	24,9	67,6	0,0	6,7
12,00	24,7	77,6	10,5	4,3
36,00	27,1	56,8	0,0	12,2
66,00	23,2	94,3	25,8	1,1
90,00	27,2	72,1	1,5	8,7
87,00	28,8	67,3	0,0	8,1

Tabela 2 - Número médio de ácaros (y) da falsa ferrugem nos frutos por parcela, temperatura em $^{\circ}\text{C}(X_1)$, umidade em % (X_2); precipitação em mm(X_3) e insolação em horas por dia (X_4). Piracicaba (SP)

y	X_1	X_2	X_3	X_4
86,75	23,9	90,2	9,1	0,6
63,25	23,1	58,0	0,0	7,3
98,50	24,4	49,0	0,0	11,4
10,25	23,6	59,8	0,0	10,2

(continua)

Tabela 2 - continuação

y	X ₁	X ₂	X ₃	X ₄
56,25	23,8	79,3	1,0	4,0
12,00	25,5	66,1	0,0	11,6
54,25	22,6	80,3	0,0	7,3
112,50	27,1	70,7	0,0	9,1
263,25	27,9	75,7	0,0	4,6

3.1.2. Experimento II: "Efeitos da intensidade da população de Sitophilus oryzae (Linnê, 1763) (Coleoptera - Curculionidae) e do período de armazenamento sobre diferentes variedades de milho".

Este experimento foi conduzido pelo Eng^o Agr^o José Claret Mاتيoli, na câmara de condições controladas do Dept^o de Zoologia da Universidade Federal do Paraná, na cidade de Curitiba, estado do Paraná.

Utilizou-se um delineamento experimental inteiramente casualizado, em esquema fatorial com 3 repetições, sendo cada parcela constituída por um frasco contendo quinhentos grãos de milho. Determinou-se o comportamento das cultivares Flint Composto, Piranão e híbrido C-111 sobre populações iniciais de 0,5, 10 e 20 casais de gorgulhos (Sitophilus oryzae), mantidos em armazenamento durante 60, 105 e 150 dias. No final de cada período formavam-se, ao acaso, para cada cultivar 12 frascos (4 níveis x 3 repetições), determinando-se as seguintes

variáveis:

y = perda de peso (%)

X_1 = número de insetos

X_2 = produto do tempo de armazenamento por:

0 se a parcela não foi infestada e

1 se a parcela foi infestada;

X_3 = Teor de carboidratos (%)

X_4 = Teor de umidade (%)

X_5 = Teor de matéria seca (%)

X_6 = Teor de proteínas (%)

X_7 = Teor de ácidos graxos (%)

X_8 = Índice de acidez dos ácidos graxos (%)

Dos dados obtidos, somente as variáveis X_1 e y sofreram transformações visando a adequação às exigências do modelo matemático.

Os dados constam nas tabelas 3, 4 e 5.

Tabela 3 - Avaliações de danos em grãos de milho cultivar C-111

y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
4,360	0,707	0,0	72,086	12,391	0,987	10,005	4,530	24,336
5,437	0,707	0,0	72,865	12,607	0,951	8,936	4,641	26,212
5,425	0,707	0,0	72,568	12,529	0,951	9,629	4,322	22,630
9,418	6,819	60,0	70,977	12,949	1,031	10,469	4,572	22,113
8,148	6,519	60,0	72,167	12,581	0,999	9,777	4,474	28,147
7,649	4,743	60,0	72,195	12,615	1,105	9,495	4,588	27,263
15,370	12,786	60,0	71,073	13,142	1,041	9,909	4,735	44,133
12,485	9,823	60,0	72,216	12,979	1,001	9,035	4,767	36,757
14,940	13,473	60,0	72,634	13,039	1,107	8,576	4,644	40,565
17,276	13,209	60,0	71,390	13,005	1,040	9,673	4,891	42,461
14,586	11,379	60,0	71,801	13,280	1,030	9,027	4,861	40,492
16,118	12,980	60,0	71,956	13,048	1,035	9,341	4,620	41,236
6,812	0,707	0,0	72,680	12,951	1,155	8,792	4,420	31,440
6,632	0,707	0,0	72,867	12,835	1,224	8,520	4,552	33,835
7,927	0,707	0,0	72,400	13,024	1,235	8,897	4,444	29,245
13,644	8,031	105,0	70,700	13,371	1,287	9,897	4,744	42,660
17,484	11,022	105,0	71,325	13,498	1,281	9,323	4,571	43,550
9,842	6,123	105,0	71,334	13,206	1,209	9,660	4,590	36,945
22,991	14,983	105,0	69,840	13,556	1,393	10,273	4,938	67,534
25,814	17,762	105,0	69,274	13,659	1,340	10,856	4,870	97,399
25,718	18,207	105,0	69,728	13,457	1,411	10,221	5,182	93,434
22,640	14,747	105,0	70,049	13,118	1,326	10,346	5,161	56,625
31,502	22,726	105,0	70,028	13,412	1,412	10,473	4,674	148,556
30,361	21,737	105,0	68,504	13,951	1,445	11,883	4,215	172,521
4,009	0,707	0,0	72,932	12,470	1,210	9,147	4,240	15,842
4,887	0,707	0,0	72,225	12,777	1,146	9,456	4,395	15,647
5,239	0,707	0,0	72,872	12,842	1,207	8,680	4,397	15,934
23,085	13,693	150,0	70,312	12,898	1,226	9,734	4,628	37,867
24,806	15,475	150,0	69,680	13,428	1,356	10,740	4,795	36,988
22,747	14,300	150,0	71,289	13,503	1,228	9,199	4,780	34,643
22,317	13,583	150,0	69,964	13,953	2,418	9,892	3,773	41,520
24,088	14,439	150,0	69,531	13,587	2,495	9,588	4,798	35,405
30,147	17,930	150,0	70,322	13,241	1,939	9,849	4,648	43,233
31,096	19,634	150,0	69,936	13,345	2,307	10,195	4,217	71,769
24,646	16,688	150,0	71,369	13,000	1,961	9,023	4,646	38,676
30,432	19,912	150,0	70,026	12,934	2,472	10,377	4,189	66,272

Tabela 4 - Avaliações de danos em grãos de milho cultivar Flint Composto.

y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
4,314	0,707	0,0	68,457	11,953	0,958	13,810	4,819	52,665
5,367	0,707	0,0	68,382	12,183	1,007	13,689	4,736	46,486
4,592	0,707	0,0	68,400	12,120	0,910	13,740	4,819	43,469
6,398	4,301	60,0	67,695	12,294	1,263	14,010	4,735	49,628
10,861	7,516	60,0	67,152	12,645	1,309	14,119	4,773	45,535
9,390	7,382	60,0	66,999	12,810	1,283	14,138	4,767	49,102
11,683	11,811	60,0	66,246	12,924	1,353	14,393	5,082	73,487
10,216	8,689	60,0	67,148	12,495	1,333	14,012	5,010	67,999
10,504	9,772	60,0	67,616	12,974	1,218	13,660	4,529	72,560
14,998	13,509	60,0	66,833	13,258	1,026	13,645	5,235	76,588
12,828	11,853	60,0	67,052	12,703	1,104	14,082	5,057	51,152
13,156	10,653	60,0	67,379	12,692	1,327	13,651	4,949	59,561
8,177	0,707	0,0	67,103	13,029	1,808	13,435	4,622	31,989
7,321	0,707	0,0	67,544	12,759	1,592	13,470	4,633	32,437
7,636	0,707	0,0	67,285	12,690	1,602	13,633	4,787	31,176
16,674	12,469	105,0	66,420	13,138	1,768	13,734	4,936	49,933
15,150	9,354	105,0	65,472	13,516	1,637	14,442	4,931	39,769
10,488	5,244	105,0	66,463	13,245	1,326	14,059	4,905	47,242
16,882	13,729	105,0	66,817	13,061	1,668	13,675	4,777	78,162
11,922	8,514	105,0	65,875	13,042	1,914	14,350	4,817	81,616
16,342	11,510	105,0	66,094	13,540	2,254	13,418	4,732	86,797
17,532	13,838	105,0	65,098	13,521	2,048	14,530	4,801	57,121
15,808	12,629	105,0	66,961	13,142	1,349	13,767	4,779	53,729
16,647	14,017	105,0	66,134	13,219	1,569	14,451	4,624	55,329
6,447	0,707	0,0	68,095	12,979	1,497	13,071	4,356	26,710
7,156	0,707	0,0	66,437	13,066	1,460	14,253	4,782	22,299
7,606	0,707	0,0	67,919	13,203	1,462	12,968	4,445	21,305
22,904	14,815	150,0	63,292	14,523	2,387	14,951	4,844	42,634
16,041	9,513	150,0	64,680	13,978	1,592	14,762	4,729	31,535
21,393	14,089	150,0	64,384	14,040	1,734	14,894	4,946	35,211
23,779	13,656	150,0	65,650	14,069	1,435	14,085	4,759	47,302
19,382	13,133	150,0	66,222	13,689	1,186	14,039	4,861	40,428
26,115	16,896	150,0	64,783	14,049	1,421	15,028	4,716	49,292
24,675	16,263	150,0	63,391	14,422	1,139	16,703	4,343	64,506
22,855	15,732	150,0	64,398	13,752	1,217	15,662	4,969	43,935
32,691	21,459	150,0	62,353	15,646	1,434	18,235	2,330	115,699

Tabela 5 - Avaliação de danos em grãos de milho cultivar Piramã.

y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
4,761	0,707	0,0	69,498	12,096	1,036	12,849	4,520	20,684
4,131	0,707	0,0	69,014	12,124	1,000	13,068	4,791	18,622
4,360	0,707	0,0	68,621	12,112	0,850	13,584	4,830	13,974
10,167	7,176	60,0	67,591	12,509	1,028	14,067	4,801	13,759
12,762	11,423	60,0	66,703	13,057	1,076	14,049	5,112	17,155
10,120	8,573	60,0	67,670	12,649	1,157	13,895	4,628	15,751
11,394	10,793	60,0	67,106	12,831	1,271	13,910	4,879	17,238
11,762	10,024	60,0	67,667	12,835	1,201	13,553	4,743	16,273
8,730	7,314	60,0	67,891	12,126	1,194	14,133	4,653	15,453
11,155	9,974	60,0	69,535	12,981	1,182	12,807	4,493	13,039
11,269	10,464	60,0	67,635	12,489	1,254	14,087	4,533	15,607
11,955	9,924	60,0	66,860	12,969	1,301	14,195	4,672	20,872
6,971	0,707	0,0	68,479	12,716	1,306	13,175	4,321	10,775
6,802	0,707	0,0	68,663	12,721	1,408	13,008	4,296	9,189
7,158	0,707	0,0	68,937	12,648	1,362	12,939	4,112	8,624
14,485	9,513	105,0	68,593	13,031	1,311	12,531	4,541	20,206
15,435	13,583	105,0	68,719	13,003	1,379	12,420	4,477	20,411
17,299	11,379	105,0	68,932	12,725	1,337	12,489	4,514	20,352
20,418	16,598	105,0	67,298	13,071	1,344	13,597	4,687	31,702
17,248	14,781	105,0	67,670	12,875	1,351	13,553	4,548	21,293
12,644	10,977	105,0	68,197	12,951	1,369	13,237	4,244	15,885
19,985	16,263	105,0	66,351	13,090	1,427	14,610	4,552	33,712
17,427	12,780	105,0	67,155	13,820	1,386	13,674	3,962	47,854
21,564	17,902	105,0	67,300	12,915	1,309	14,082	4,392	21,446
5,802	0,707	0,0	68,850	12,780	1,558	12,680	4,129	10,712
6,083	0,707	0,0	66,584	12,713	1,949	14,394	4,359	10,222
5,724	0,707	0,0	68,251	12,478	1,812	12,952	4,505	9,307
18,006	11,247	105,0	67,515	13,258	1,450	13,282	4,491	20,511
23,489	14,815	150,0	63,129	14,215	1,716	16,712	4,227	49,043
17,133	11,510	150,0	64,194	13,373	1,582	16,271	4,578	32,980
21,939	13,209	150,0	66,350	13,594	1,467	14,478	4,109	34,350
23,001	16,263	150,0	65,977	13,807	2,092	13,439	4,682	21,882
17,022	10,885	150,0	65,310	13,398	3,064	13,863	4,363	26,330
32,163	24,259	150,0	63,145	14,825	3,453	15,762	2,812	125,530
29,271	17,449	150,0	61,989	14,708	2,039	17,031	4,230	66,520
26,660	21,644	150,0	65,631	14,611	3,000	14,329	2,427	86,761

3.2. Métodos

3.2.1. Regressão Múltipla

3.2.1.1. Modelo Matemático

O modelo matemático da regressão múltipla adotado será:

$$y = X\beta + \epsilon$$

onde, y = vetor das observações, de dimensão $n \times 1$;

X = matriz dos coeficientes dos parâmetros, de dimensão $n \times p$;

β = vetor dos coeficientes de regressão, de dimensão $p \times 1$;

ϵ = vetor dos erros, de dimensão $n \times 1$.

Admite-se que $\epsilon \sim N(\phi_n; I_n \sigma^2)$ e que os erros das observações são independentes.

3.2.1.2. Estimativa dos Coeficientes de Regressão Múltipla

Considere-se o modelo matemático

$$y = X\beta + \epsilon$$

$$\begin{array}{c} \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} \right] \\ nx1 \end{array} = \begin{array}{c} \left[\begin{array}{cccc} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{array} \right] \\ nxp \end{array} \cdot \begin{array}{c} \left[\begin{array}{c} b_1 \\ b_2 \\ \vdots \\ b_p \end{array} \right] \\ px1 \end{array} + \begin{array}{c} \left[\begin{array}{c} e_1 \\ e_2 \\ \vdots \\ e_n \end{array} \right] \\ nx1 \end{array}$$

sendo, n = número de observações e p = número de coeficientes de regressão, com $n > p$.

O erro aleatório ou casual será a diferença entre o valor observado e o valor esperado, isto é,

$$\varepsilon = y - E(y) \quad \text{onde} \quad E(y) = X\beta.$$

Utilizando o método de estimadores de mínimos quadrados, que consiste em adotar os estimadores que minimizam a soma de quadrados dos desvios entre os valores estimados e os valores observados da amostra, tem-se o seguinte sistema de equações normais:

$$X'X\hat{\beta} = X'y.$$

Como a matriz $X'X$ é não singular, pois $n > p$, então existirá $(X'X)^{-1}$ logo,

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

que é o vetor das estimativas dos coeficientes de regressão.

3.2.1.3. Propriedade dos estimadores

a) $\hat{\beta}$ é um estimador não viesado de β pois,

$$E(\hat{\beta}) = \beta$$

b) Matriz das variâncias e covariâncias de $\hat{\beta}$.

Por definição a matriz das variâncias e covariâncias ou a matriz de dispersão de $\hat{\beta}$ é dada por

$$V(\hat{\beta}) = D(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\}, \text{ como } \hat{\beta} = (X'X)^{-1}X'y$$

$$D(\hat{\beta}) = (X'X)^{-1} \cdot \sigma^2$$

3.2.1.4. Erro padrão

A matriz de dispersão $D(\hat{\beta}) = (X'X)^{-1} \sigma^2$ tem em sua diagonal principal as variâncias dos coeficientes de regressão que serão denotadas por $V(\hat{\beta}_i)$ com $i = 1, 2, \dots, p$. A raiz quadrada destas variâncias dará o erro padrão dos coeficientes de regressão que será denotado por $S(\hat{\beta}_i)$, ou seja,

$$S(\hat{\beta}_i) = \sqrt{V(\hat{\beta}_i)}$$

e sua estimativa por

$$s(\hat{b}_i) = \sqrt{\hat{V}(\hat{b}_i)}.$$

3.2.1.5. Intervalo de confiança

Os intervalos de confiança para os coeficientes de regressão são dados por

$$\hat{b}_i \pm t \cdot s(\hat{b}_i)$$

sendo t um valor da tabela t de Student obtido com o nível de significância α e $\theta = n - p$ graus de liberdade.

3.2.1.6. Teste t

A fim de testar as hipóteses

$$H_0 : b_i = 0 \quad \text{vs} \quad H_1 : b_i \neq 0$$

poderá ser aplicado o teste t ao nível de significância α e com $\theta = n - p$ graus de liberdade e a estatística

$$t = \frac{\hat{b}_i}{s(\hat{b}_i)}$$

3.2.1.7. Análise de Variância

a) Soma de quadrados do resíduo

Uma vez conhecida a estimativa dos coeficientes de regressão, a soma de quadrados do resíduo (SQRes) será:

$$\text{SQRes} = \hat{\varepsilon}'\hat{\varepsilon} \quad \text{ou seja} \quad \text{SQRes} = X'y - \hat{\beta}'X'y$$

b) Decomposição da soma de quadrados

$$\text{SQRes} = X'y - \hat{\beta}'X'y \quad \text{ou} \quad \text{SQRes} = (X'y - C) - (\hat{\beta}'X'y - C)$$

$$\text{onde} \quad C = \frac{y'EE'y}{E'E} \quad \text{e} \quad E = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}$$

sendo, $y'y - C$ a soma de quadrados total, ou seja,

$$\text{SQTo} = y'y - C ;$$

$\hat{\beta}'X'y - C$ a soma de quadrados da regressão, isto é ,
 $\text{SQReg} = \hat{\beta}'X'y - C$ logo,

$$\text{SQRes} = \text{SQTo} - \text{SQReg} .$$

3.2.1.8. Esperança Matemática das Somas de Quadrados

Para facilitar a obtenção da esperança matemática será enunciado o teorema apresentado por SEARLE (1971).

Se $y \sim N(\mu; V)$ então

$$E(y'Ay) = t_R(AV) + \mu'A\mu.$$

Como $E(y) = \mu$ e $V(y) = V$ tem-se que

$$E(yy') = V + \mu'\mu$$

$$\text{portanto } E(y'Ay) = E[t_R(Ayy')] = t_R[AE(yy')]$$

$$= t_R(AV + A\mu'\mu)$$

$$= t_R(AV) + \mu'A\mu.$$

Desta forma, a esperança matemática da soma de quadrados do resíduo será:

$$E(\text{SQRes}) = E(\hat{\epsilon}'\hat{\epsilon}) = E[(y'y - C) - (\hat{\beta}'X'y - C)]$$

$$= E(y'y) - E(C) - E(\hat{\beta}'X'y) + E(C)$$

$$\text{onde, a) } E(y'y) = E[(X\beta + \epsilon)'(X\beta + \epsilon)] = E[(\beta'X' + \epsilon')(X\beta + \epsilon)]$$

$$\begin{aligned}
&= E(\beta'X'X\beta + \beta'X'\epsilon + \epsilon'X\beta + \epsilon'\epsilon) \\
&= \beta'X'X\beta + \beta'X'E(\epsilon) + E(\epsilon')X\beta + E(\epsilon'\epsilon)
\end{aligned}$$

como $\epsilon \sim N(\phi; I_n \sigma^2)$ então $E(\epsilon) = \phi$ e

$$E(\epsilon'\epsilon) = E(\epsilon'I_n\epsilon) = \text{tr}(I_n\sigma^2) + \phi'I_n\phi = n\sigma^2$$

portanto

$$E(y'y) = \beta'X'X\beta + n\sigma^2$$

b) $E(\hat{\beta}'X'y) = E\{[(X'X)^{-1}X'y]'X'y\} = E[y'X(X'X)^{-1}X'y]$, pois $(X'X)^{-1}$ é simétrica,

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

como $E(\epsilon) = \phi$,

$$E(\hat{\beta}'X'y) = \beta'X'X\beta + E[\epsilon'X(X'X)^{-1}X'\epsilon]$$

fazendo $A = X(X'X)^{-1}X'$,

$$E(\hat{\beta}'X'y) = \beta'X'X\beta + E(\epsilon'A\epsilon)$$

e, então, pelo mesmo teorema,

$$E(\hat{\beta}'X'y) = \beta'X'X\beta + t_R(A\sigma^2)$$

$$\text{como } t_R(A\sigma^2) = t_R[X(X'X)^{-1}X'\sigma^2] = t_R[X'X(X'X)^{-1}\sigma^2]$$

$$= t_R(I_p \sigma^2) = p \sigma^2 \quad \text{logo,}$$

$E(\hat{\beta}'X'y) = \beta'X'X\beta + p\sigma^2$ onde p é o número de parâmetros do modelo.

$$c) E(C) = E\left(\frac{y'EE'y}{E'E}\right)$$

Como $y \sim N(X\beta; I_n\sigma^2)$, pelo teorema tem-se que:

$$\begin{aligned} E(C) &= \frac{t_R(EE'\sigma^2)}{E'E} + \frac{\beta'X'EE'X\beta}{E'E} \\ &= \frac{n}{n} \sigma^2 + \frac{\beta'X'EE'X\beta}{E'E} \\ &= \sigma^2 + \frac{\beta'X'EE'X\beta}{E'E} \end{aligned}$$

Assim, tem-se que

$$E(\text{SQRes}) = (n - p)\sigma^2$$

onde $(n-p)$ representa o número de graus de liberdade do resíduo.

Um estimador não viefzado de σ^2 será $\hat{\sigma}^2 = \frac{\text{SQRes}}{n-p}$

que corresponde ao quadrado médio do resíduo.

O termo $y'y - C$ que compõe a soma de quadrados do resíduo é denominado soma de quadrados total (SQTo), cuja esperança matemática será:

$$\begin{aligned} E(\text{SQTo}) &= E(y'y) - E(C) = \beta'X'X\beta + n\sigma^2 - \left(\sigma^2 + \frac{\beta'X'EE'X\beta}{E'E}\right) \\ &= \beta'X'X\beta - \frac{\beta'X'EE'X\beta}{E'E} + (n-1)\sigma^2 \end{aligned}$$

denotando $\beta'X'X\beta - \frac{\beta'X'EE'X\beta}{E'E} = Z$, fica

$$E(\text{SQTo}) = Z + (n-1)\sigma^2.$$

O termo $\hat{\beta}'X'y - C$ que também compõe a soma de quadrados do resíduo e é denominada de soma de quadrados da regressão (SQReg) tem a seguinte esperança matemática:

$$\begin{aligned} E(\text{SQReg}) &= E(\hat{\beta}'X'y) - E(C) = \beta'X'X\beta + p\sigma^2 - \sigma^2 - \\ &\quad - \frac{\beta'X'EE'X\beta}{E'E} \end{aligned}$$

$$E(\text{SQReg}) = \beta'X'X\beta - \frac{\beta'X'EE'X\beta}{E'E} + (p-1)\sigma^2 \quad \text{ou ainda,}$$

$$E(\text{SQReg}) = Z + (p-1)\sigma^2.$$

3.2.1.9. Coeficiente de Determinação (R^2)

O coeficiente de determinação ou explicação é uma medida descritiva da proporção da variação de y que é explicada pela regressão e é obtida por

$$R^2 = \frac{\text{SQReg}}{\text{SQTo}}.$$

sendo seu campo de variação o intervalo real $0 < R^2 < 1$.

Esta estatística é útil para medir a validade do ajustamento entre os pontos observados e o plano de regressão. Assim sendo, quanto mais próximo da unidade estiver o valor R^2 , melhor "a qualidade" do ajuste do plano de regressão aos pontos observados e quanto mais próximo de zero pior será a "qualidade do ajuste".

3.2.1.10. Teste F

Na análise de regressão é utilizado o teste F a fim de testar as seguintes hipóteses:

$$H_0 : b_i = 0 \quad \text{vs} \quad H_1 : b_i \neq 0 \quad (i = 1, 2, \dots, p).$$

A distribuição F é definida pela razão de duas variáveis independentes, cada uma das quais com distribuição de qui-quadrado com θ_1, θ_2 graus de liberdade. Sendo que os graus de liberdade não são necessariamente iguais, isto é, se uma variável tem distribuição de qui-quadrado com m graus de liberdade e uma outra variável tem distribuição de qui-quadrado com p graus de liberdade então a razão entre estas duas variáveis terá distribuição F com m e p graus de liberdade, segundo GRAYBILL(1976), portanto

$$QMReg = \frac{\hat{\beta}'X'y - C}{p - 1} \sim \chi^2_{(p-1)} \quad e$$

$$QMRes = \frac{y'y - \hat{\beta}'X'y}{n - p} \sim \chi^2_{(n-p)} \quad \text{então}$$

$$F_o = \frac{QMReg}{QMRes} \sim F_{[(p-1); (n-p)]}$$

3.2.1.11. Quadro de análise de variância

C.V.	G.L.	S.Q.	Q.M.	E(Q.M.)
Regressão	p-1	$\hat{\beta}'X'y - C$	$\frac{\hat{\beta}'X'y - C}{p-1}$	$\frac{\sigma^2}{p-1} + \sigma^2$
Resíduo	n-p	$y'y - \hat{\beta}'X'y$	$\frac{y'y - \hat{\beta}'X'y}{n-p}$	σ^2
Total	n-1	$y'y - C$		

3.2.2. Regressão de Cumeeira

3.2.2.1. Modelo Matemático

Para este método utiliza-se o mesmo modelo matemático da regressão múltipla

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

com variáveis normalizadas, de forma que $X'X$ seja matriz de correlação simples e $X'y$ o vetor dos coeficientes de correlação da variável dependente com cada uma das variáveis independentes, ou seja,

$$x_{ij} = X_{ij} - \bar{X}_j \quad \text{onde } X_{ij} \text{ é a } i\text{-ésima observação da } j\text{-}$$

ésima variável independente; $\bar{X}_j$ é a média aritmética dos valores da j -ésima variável independente e

$$\frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}} \quad \text{os elementos da matriz } X, \text{ com } i = 1, 2, \dots, n$$

e $j = 1, 2, \dots, p$, e da mesma forma, $y_i = y_i - \bar{y}$ onde y_i é a i -ésima observação da variável dependente, $\bar{y}$ é a média aritmética dos valores da variável dependente e

$$\frac{y_i}{\sqrt{\sum_{i=1}^n y_i^2}} \quad \text{os elementos do vetor } y, \text{ com } i = 1, 2, \dots, n.$$

3.2.2.2. Estimativa dos Coeficientes de Regressão de Cumeeira

O estimador de cumeeira é obtido através do acréscimo de uma constante positiva aos elementos da diagonal principal da matriz de correlação simples $X'X$, em forma de matriz de correção, do vetor das estimativas dos coeficientes de regressão obtido pelo método dos mínimos quadrados, isto é,

$$\hat{\beta}^* = (X'X + kI_p)^{-1}X'y$$

sendo, k um valor pequeno e positivo, em geral, $k \in (0,1]$.

3.2.2.3. Propriedade dos Estimadores

a) $\hat{\beta}^*$ é um estimador viesado de β pois, $E(\hat{\beta}^*) \neq \beta$

$$\text{Demonstração: } E(\hat{\beta}^*) = E[(X'X + kI_p)^{-1}X'y]$$

$$\begin{aligned} E(\hat{\beta}^*) &= (X'X + kI_p)^{-1}X'E(y) = (X'X + kI_p)^{-1}X'E(X\beta + \epsilon) \\ &= (X'X + kI_p)^{-1}X'X\beta + (X'X + kI_p)^{-1}X'E(\epsilon) \end{aligned}$$

mas $E(\varepsilon) = \phi$ logo, $E(\hat{\beta}^*) = (X'X + kI_p)^{-1}X'X\beta$

b) Matriz das Variâncias e Covariâncias de $\hat{\beta}^*$.

Sabe-se que $\hat{\beta}^* = (X'X + kI_p)^{-1}X'y$ ou seja,

$$\hat{\beta}^* = (X'X + kI_p)^{-1}X'X\beta + (X'X + kI_p)^{-1}X'\varepsilon$$

$$\hat{\beta}^* = E(\hat{\beta}^*) + (X'X + kI_p)^{-1}X'\varepsilon.$$

Por definição, a matriz das variâncias e covariâncias, ou matriz de dispersão de $\hat{\beta}^*$ é,

$$V(\hat{\beta}^*) = D(\hat{\beta}^*)' = E\{[\hat{\beta}^* - E(\hat{\beta}^*)][\hat{\beta}^* - E(\hat{\beta}^*)]'\} \quad \text{logo, se}$$

$$\hat{\beta}^* - E(\hat{\beta}^*) = (X'X + kI_p)^{-1}X'\varepsilon \quad \text{fica,}$$

$$D(\hat{\beta}^*) = E\{[(X'X + kI_p)^{-1}X'\varepsilon][(X'X + kI_p)^{-1}X'\varepsilon]'\}$$

$$= E\{[(X'X + kI_p)^{-1}X'\varepsilon][\varepsilon'X(X'X + kI_p)^{-1}]\}$$

$$D(\hat{\beta}^*) = (X'X + kI_p)^{-1}X'E(\varepsilon\varepsilon')X(X'X + kI_p)^{-1}$$

$$\text{como, } E(\varepsilon\varepsilon') = I_n\sigma^2 \quad \text{fica,}$$

$$D(\hat{\beta}^*) = (X'X + kI_p)^{-1}X'I_n\sigma^2X(X'X + kI_p)^{-1}$$

$$= (X'X + kI_p)^{-1}X'X(X'X + kI_p)^{-1}\sigma^2$$

3.2.2.4. Análise de Variância

Para facilitar as demonstrações seguintes, será considerado $(X'X + kI_p)^{-1} = S$ e ainda, $S^{-1} = X'X + kI_p$ ou $X'X = S^{-1} - kI_p$.

Para um determinado valor de k , com $k \in (0,1]$, a soma de quadrados do resíduo da equação estimada pelo método da regressão de cumeeira será:

$$SQRes^* = \hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{\beta}^*)'(y - X\hat{\beta}^*) \quad \text{efetuando o produto,}$$

$$\hat{\varepsilon}'\hat{\varepsilon} = y'y - y'X\hat{\beta}^* - \hat{\beta}^{*'}X'y + \hat{\beta}^{*'}X'X\hat{\beta}^*, \text{ como } X'X = S^{-1} - kI_p,$$

$$\hat{\varepsilon}'\hat{\varepsilon} = y'y - y'X\hat{\beta}^* - \hat{\beta}^{*'}X'y + \hat{\beta}^{*'}(S^{-1} - kI_p)\hat{\beta}^*$$

$$= y'y - y'X\hat{\beta}^* - \hat{\beta}^{*'}X'y + \hat{\beta}^{*'}S^{-1}\hat{\beta}^* - \hat{\beta}^{*'}kI_p\hat{\beta}^*$$

$$= y'y - y'X\hat{\beta}^* - \hat{\beta}^{*'}X'y + \hat{\beta}^{*'}S^{-1}\hat{\beta}^* - k\hat{\beta}^{*'}\hat{\beta}^*$$

$$\text{como } \hat{\beta}^* = SX'y \quad \text{e} \quad \hat{\beta}^{*'} = y'XS, \text{ pois } S = S'$$

$$SQRes^* = y'y - y'X\hat{\beta}^* - \hat{\beta}^{*'}X'y + y'XSS^{-1}\hat{\beta}^* - k\hat{\beta}^{*'}\hat{\beta}^*$$

$$= y'y - y'X\hat{\beta}^* - \hat{\beta}^{*'}X'y + y'X\hat{\beta}^* - k\hat{\beta}^{*'}\hat{\beta}^*$$

portanto,

$$SQRes^* = y'y - \hat{\beta}^{*'}X'y - k\hat{\beta}^{*'}\hat{\beta}^* .$$

b) Decomposição das Somas de Quadrados

Somando e subtraindo o termo

$$C = \frac{y'EE'y}{E'E}, \text{ com } E = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}, \text{ na expressão da soma}$$

de quadrados do resíduo tem-se:

$$SQRes^* = (y'y - C) - (\hat{\beta}^{*'}X'y + k\hat{\beta}^{*'}\hat{\beta}^* - C) \quad \text{sendo, } y'y - C$$

denominada soma de quadrados total (SQTo) e $\hat{\beta}^{*'}X'y + k\hat{\beta}^{*'}\hat{\beta}^* - C$ a soma de quadrados da regressão de cumeeira (SQReg*).

3.2.2.5. Coeficiente de Determinação

O coeficiente de determinação R^2 sendo uma medida descritiva, terá validade no método da regressão de cumeeira e será obtida por:

$$R^2 = \frac{SQReg^*}{SQTo} .$$

3.2.2.6. Testes de Significância

Supondo que os erros têm distribuição normal com $E(\epsilon) = 0$ e $E(\epsilon'\epsilon) = I_n \sigma^2$ então $\hat{\beta}$ tem distribuição multinormal, e a aplicação dos testes de significância usuais é válida nas análises de regressão mas, no caso da análise de regressão de cumeieira estes testes não são válidos pois, conforme demonstração anterior, as estimativas de cumeieira são tendenciosas.

3.2.2.7. Intervalos de Confiança

Devido à tendenciosidade das estimativas dos coeficientes de regressão de cumeieira, os intervalos de confiança para essas estimativas devem ser interpretados como medidas descritivas.

3.2.2.8. Soma de quadrados dos desvios de regressão

a) de $\hat{\beta}$ em relação a β

Seja $\hat{\beta} = (X'X)^{-1}X'y$ a estimativa dos coeficientes de regressão pelo método dos mínimos quadrados, isto é,
 $\hat{\beta} = (X'X)^{-1}X'(X\beta + \epsilon)$

$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon \quad \text{ou ainda,} \quad \hat{\beta} - \beta = (X'X)^{-1}X'\epsilon \quad \text{assim,}$$

$$E[\text{SQD}(\hat{\beta})] = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] = E[\epsilon'X(X'X)^{-1}(X'X)^{-1}X'\epsilon].$$

Pelo teorema enunciado em 3.2.1.5 e sabendo que $t_R(AB) = t_R(BA)$, temos

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

portanto, $E[SQD(\hat{\beta})] = \sigma^2 t_R(X'X)^{-1}$

b) Soma de quadrados dos desvios de $\hat{\beta}^*$ em relação a β

Seja, $\hat{\beta}^* = (X'X + kI_p)^{-1}X'y$ o vetor das estimativas dos coeficientes de regressão de cumeeira. Somando e subtraindo o vetor β ao vetor $\hat{\beta}^*$ tem-se

$$\begin{aligned}\hat{\beta}^* - \beta &= (X'X + kI_p)^{-1}X'y - \beta \\ &= (X'X + kI_p)^{-1}X'X\beta + (X'X + kI_p)^{-1}X'\epsilon - \beta\end{aligned}$$

fazendo $(X'X + kI_p)^{-1} = S$ fica

$$\hat{\beta}^* - \beta = SX'X\beta + SX'\epsilon - \beta \quad \text{logo, a esperança matemática}$$

da soma de quadrados dos desvios de $\hat{\beta}^*$ será:

$$\begin{aligned}E[SQD(\hat{\beta}^*)] &= E[(\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta)] \\ &= E[(SX'X\beta + SX'\epsilon - \beta)'(SX'X\beta + SX'\epsilon - \beta)]\end{aligned}$$

como S é uma matriz simétrica, $S' = S$ logo,

$$\begin{aligned} E[SQD(\hat{\beta}^*)] &= E[(\beta'X'XS + \epsilon'XS - \beta')(SX'X\beta + SX'\epsilon - \beta)] \\ &= E \left[\begin{array}{l} \beta'X'XSSX'X\beta + \beta'X'XSSX'\epsilon - \beta'X'XS\beta + \\ + \epsilon'XSSX'X\beta + \epsilon'XSSX'\epsilon - \epsilon'X'S\beta - \\ - \beta'SX'X\beta - \beta'SX'\epsilon + \beta'\beta \end{array} \right] \end{aligned}$$

$$\begin{aligned} E[SQD(\hat{\beta}^*)] &= \beta'X'XSSX'X\beta + \beta'X'XSSX'E(\epsilon) - \beta'X'XS\beta + \\ &+ E(\epsilon')XSSX'X\beta + E(\epsilon'XSSX'\epsilon) - E(\epsilon')XS\beta - \\ &- \beta'SX'X\beta - \beta'SX'E(\epsilon) + \beta'\beta \end{aligned}$$

como $E(\epsilon) = E(\epsilon') = \phi$ fica

$$\begin{aligned} E[SQD(\hat{\beta}^*)] &= \beta'X'XSSX'X\beta - \beta'X'XS\beta + E(\epsilon'XSSX'\epsilon) - \\ &- \beta'SX'X\beta + \beta'\beta \end{aligned}$$

aplicando o teorema enunciado em 3.2.1.5.,

$$\begin{aligned} E[SQD(\hat{\beta}^*)] &= \beta'X'XSSX'X\beta - \beta'X'XS\beta + \sigma^2 t_R(X'XSS) - \\ &- \beta'SX'X\beta + \beta'\beta \end{aligned}$$

mas, $X'X = S^{-1} - kI_p$

$$\begin{aligned}
E[\text{SQD}(\hat{\beta}^*)] &= \beta'(S^{-1} - kI_p)SS(S^{-1} - kI_p)\beta - \beta'(S^{-1} - kI_p)S\beta + \\
&+ \sigma^2 t_R(X'XSS) - \beta'S(S^{-1} - kI_p)\beta + \beta'\beta \\
&= \beta'S^{-1}SSS^{-1}\beta - \beta'S^{-1}SSkI_p\beta - \beta'kI_pSSS^{-1}\beta + \\
&+ \beta'kI_pSSkI_p\beta - \beta'S^{-1}S\beta + \beta'kI_pS\beta + \\
&+ \sigma^2 t_R(X'XSS) - \beta'SS^{-1}\beta + \beta'SkI_p\beta + \beta'\beta \\
&= \beta'\beta - k\beta'S\beta - k\beta'S\beta + k^2\beta'SS\beta - \beta'\beta + \\
&+ k\beta'S\beta + \sigma^2 t_R(SX'XS) - \beta'\beta + k\beta'S\beta + \beta'\beta
\end{aligned}$$

portanto, $E[\text{SQD}(\hat{\beta}^*)] = k^2\beta'SS\beta + \sigma^2 t_R(SX'XS)$ como

$\sigma^2 t_R(SX'XS) = t_R D(\hat{\beta}^*)$ pois $SX'XS = D(\hat{\beta}^*)$ onde

$S = (X'X + kI_p)^{-1}$ fica

$$E[\text{SQD}(\hat{\beta}^*)] = k^2\beta'SS\beta + t_R D(\hat{\beta}^*) .$$

Para a expressão acima existirá um valor de k para o qual:

$$E[\text{SQD}(\hat{\beta}^*)] < E[\text{SQD}(\hat{\beta})] .$$

Esta propriedade poderá ser adotada como um critério para a utilização do método da regressão de cumeeira.

Comparando os resultados obtidos na esperança ma

temática das somas de quadrados dos desvios, percebe-se que quando $k = 0$, o estimador de cumeieira será igual ao estimador de mínimos quadrados referente a regressão múltipla, ou seja, para $k = 0$ a

$$\begin{aligned} E[\text{SQD}(\hat{\beta}^*)] &= k^2 \beta' S S \beta + \sigma^2 t_R(SX'XS) \\ &= 0 + \sigma^2 t_R[(X'X)^{-1} X'X (X'X)^{-1}] \\ &= \sigma^2 t_R(X'X)^{-1} = E[\text{SQD}(\hat{\beta})] \end{aligned}$$

pois $S = (X'X + kI_p)^{-1} = (X'X)^{-1}$, e ainda

$$\hat{\beta}^* = SX'y = (X'X)^{-1} X'y = \hat{\beta} \quad \text{portanto,}$$

$$E(\hat{\beta}^*) = \beta .$$

3.2.2.9. Quadro da análise de variância

C.V.	G.L.	S.Q.	Q.M.
Regressão*	p-1	$\hat{\beta}^{*'} X'y + k \hat{\beta}^{*'} \hat{\beta}^* - C$	$\frac{\text{SQReg}^*}{p-1}$
Resíduo*	n-p	$y'y - (\hat{\beta}^{*'} X'y + k \hat{\beta}^{*'} \hat{\beta}^*)$	$\frac{\text{SQRes}^*}{n-p}$
Total	n-1	$y'y - C$	

3.2.2.10. Traço de Cumeeira

Diversos autores consideram como uma das maiores vantagens da regressão de cumeeira o traço de cumeeira, que através da análise de sensibilidade, mostra quais coeficientes de regressão são sensíveis aos dados.

O traço de cumeeira é um gráfico bidimensional do valor de cada coeficiente versus k , a constante que será acrescentada aos elementos da diagonal principal da matriz $X'X$ ou à matriz de correlações simples.

Por exemplo, Baumgartner (1979), citado por NEWELL e LEE (1981), investigou os efeitos do pH e concentração de sal sobre a capacidade de emulsão de alimentos sistêmicos. O modelo quadrático da forma

$$y = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2 + b_{22}X_2^2 + b_{12}X_1X_2$$

foi ajustado aos dados, onde y é a capacidade de emulsão, X_1 é o pH e X_2 é a % de NaCl. A regressão linear múltipla e a regressão cumeeira foram aplicadas aos dados.

A figura a seguir mostra o traço de cumeeira para vários valores de k .

Além da análise de sensibilidade o traço de cumeira tem como objetivo encontrar um valor de k que gere uma série de coeficientes com menor soma de quadrados dos desvios do que a solução de mínimos quadrados. Certamente, como k aumenta, a soma de quadrados do resíduo também aumentará, mas isto não é da maior importância porque o objetivo não é obter a melhor estimativa mas, desenvolver uma série de coeficientes de regressão estáveis que darão um bom desempenho nas previsões futuras. A estabilidade ocorre quando os coeficientes de regressão não são sensíveis às pequenas mudanças na estimação dos dados. Pois, se as variáveis independentes são altamente correlacionadas, os coeficientes mudarão de valor rapidamente para pequenos valores de k , se estabilizando para maiores valores de k e assim, o valor de k no qual os coeficientes tenham estabilizado dão a série desejada dos coeficientes.

Se as variáveis são ortogonais, os coeficientes

se alterarão muito pouco, o que mostra que os coeficientes já estão estáveis, indicando que a solução de mínimos quadrados fornece uma boa série de coeficientes,

4. RESULTADOS E DISCUSSÃO

Com o auxílio do programa "Análise de regressão pelo método de cumeeira (Ridge regression)", desenvolvido por KALIL (1978) obtiveram-se os resultados a seguir, que estão acompanhados da discussão.

A presença de multicolinearidade é evidente nos dados dos diversos experimentos devida as altas correlações entre algumas das variáveis independentes e que podem ser vistos nas tabelas 6, 12, 18, 24 e 30.

Aplicando os métodos da regressão de cumeeira e da regressão linear múltipla foram obtidas as estimativas dos coeficientes de regressão para os diversos níveis de k , para os quais foram traçados os gráficos de cumeeira.

Com o auxílio desses gráficos determinou-se o valor ideal de k para se obterem as estimativas de cumeeira. Este valor de k , segundo proposto por HOERL e KENNARD (1970 a),

é determinado pelo ponto em que as estimativas dos coeficientes de regressão começam a se estabilizar.

Conforme esse critério, a estabilização dos coeficientes de regressão ocorre nos seguintes valores de k :

- a) $k = 0,3$ para os dados de avaliação do número de ácaros da falsa ferrugem nas folhas de citros (figura 1);
- b) $k = 0,5$ para os dados de avaliação do número de ácaros da falsa ferrugem nos frutos de citros (figura 2);
- c) $k = 0,8$ para os dados de avaliações de danos em grãos de milho armazenado cultivar C-111, causados por gorgulhos (figura 3);
- d) $k = 0,1$ para os dados de avaliações de danos em grãos de milho armazenado cultivar Flint, causados por gorgulhos (figura 4);
- e) $k = 0,6$ para os dados de avaliações de danos em grãos de milho armazenado cultivar Piranão, causados por gorgulhos (figura 5).

4.1. Dados de Avaliação do número de ácaros da falsa ferrugem (Phyllocoptruta oleivora) nas folhas de citros (Citrus sinensis Var. Natal)

Para os dados desse experimento, onde se evidencia um alto grau de multicolinearidade ($\psi = 44,97$) observou-se que o coeficiente de regressão b_2 , relativo à variável X_2 ,

umidade relativa do ar, foi o mais afetado, ocorrendo uma mudança no sinal quando são comparados os métodos de regressão de cumeeira e linear múltipla e que pode-se notar na tabela 7. Este resultado está coerente com aqueles para as correlações lineares simples (Tabela 6) onde se observam os maiores valores para essa variável com as demais, exceção feita com ela e X_1 .

Os testes t para os coeficientes de regressão e F para as regressões (Tabelas 10 e 11) relativas a esses dados não apresentam significância em ambos os métodos, ao nível de 5% de probabilidade. Isto indica que, embora o método de cumeeira tenha sido mais eficiente, esses modelos de regressão, não se ajustaram a esse tipo de dados.

Tabela 6 - Estimativa dos coeficientes de correlação linear, simples entre as variáveis X_1 , X_2 , X_3 , X_4 e y .

Variáveis	Coeficientes de Correlação Linear Simples				
	X_1	X_2	X_3	X_4	y
X_1	1,0000	-0,1959	-0,4661	0,3989	0,2696
X_2	-0,1959	1,0000	0,7888	-0,9233	0,3386
X_3	-0,4661	0,7888	1,0000	-0,7063	0,1697
X_4	0,3989	-0,9233	-0,7063	1,0000	-0,2637
y	0,2696	0,3386	0,1697	-0,2637	1,0000

onde X_1 = temperatura em $^{\circ}\text{C}$

X_2 = umidade relativa do ar (%)

X_3 = precipitação em mm

X_4 = insolação em horas/dia

y = número médio de ácaros da falsa ferrugem

Figura 1 - Valores dos coeficientes de regressão para os dados de avaliação do número de ácaros da falsa ferrugem nas folhas de citros, em função dos diversos níveis de k .

Tabela 7 - Estimativa dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,3$)

i	Coeficientes Linear múltipla ($\hat{\beta}_i$)	de Regressão Cumeeira ($\hat{\beta}_i^*$)
1	0,8098	0,4644
2	-0,0286	0,0340
3	0,0723	0,0209
4	-0,3082	-0,0951
termo constante	-10,3864	-7,2446

Tabela 8 - Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,3$)

i	Erro Linear múltipla $s(\hat{\beta}_i)$	Padrão Cumeeira $s(\hat{\beta}_i^*)$
1	0,9747	0,4307
2	0,3315	0,0407
3	0,2998	0,1050
4	0,9115	0,1613

Verifica-se, pelos resultados desta tabela, que para as estimativas de cumeeira os erros padrões são menores, o que confirma as indicações bibliográficas.

Além disso, a diferença maior ocorre para $\hat{b}_2$, relativo a variável X_2 , umidade relativa do ar, que é, como vimos, a variável com maiores valores de correlações simples.

Tabela 9 - Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeieira, com $t_{(6)} = 2,4469$

i	Intervalo de confiança	
	Linear múltipla	Cumeieira
	$b_i \pm ts(\hat{b}_i)$	$b_i^* \pm ts(\hat{b}_i^*)$
1	(-1,5752 ; 3,1948)	(-0,5895 ; 1,5183)
2	(-0,8474 ; 0,7902)	(-0,0658 ; 0,1338)
3	(-0,6613 ; 0,8059)	(-0,2360 ; 0,2778)
4	(-2,5386 ; 1,9222)	(-0,4898 ; 0,2996)

Como era esperado, em função dos valores de $s(b^*)$, a menor amplitude dos intervalos de confiança ocorreu para a variável X_2 .

Tabela 10 - Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com

$$|t_{(6;0,05)}| = 2,4469$$

i	Teste t	
	Linear múltipla	Cumeeira
	(t_i)	(t_i^*)
1	0,8308	1,0783
2	-0,0862	0,8337
3	0,2412	0,1986
4	-0,3381	-0,5897

Tabela 11 - Coeficiente de determinação (R^2), soma de quadrados de resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira

i	Regressão	
	Linear múltipla	Cumeeira
R^2	0,2469	0,1848
SQR	63,19	64,87
$F^{(1)}$	0,492	0,358
SQD	0,5418	0,1406
CV(%)	52,31	52,31

$$(1) F_{(4;6;0,05)} = 4,53$$

4.2. Dados de Avaliação do número de ácaros da falsa ferrugem (Phyllocoptruta oleivora) nos frutos de citros (Citrus sinensis var. Natal)

Os dados desse experimento apresentam um baixo grau de multicolinearidade ($\psi = 10,24$). Este fator não provocou grandes alterações nos valores das estimativas dos coeficientes de regressão quando comparados os métodos da regressão linear múltipla e de cumeeira, conforme mostra a tabela 13.

Os testes t e F não apresentaram significância ao nível de 5% de probabilidade no método da regressão linear múltipla e o mesmo ocorrendo no método da regressão de cumeeira.

Tabela 12 - Estimativa dos coeficientes de correlação linear simples entre as variáveis X_1 , X_2 , X_3 , X_4 e y .

Variáveis	Coeficientes de correlação linear simples				
	X_1	X_2	X_3	X_4	y
X_1	1,0000	0,0389	-0,1769	0,0566	0,6082
X_2	0,0389	1,0000	0,6234	-0,8148	0,2297
X_3	-0,1769	0,6234	1,0000	-0,7248	0,0874
X_4	0,0566	-0,8148	-0,7248	1,0000	0,4384
X_5	0,6082	0,2297	0,0874	0,4384	1,0000

onde X_1 = temperatura em $^{\circ}\text{C}$

X_2 = umidade relativa do ar (%)

X_3 = precipitação em mm

X_4 = insolação em horas/dia

y = número médio de ácaros da falsa ferrugem

Figura 2 - Valores dos coeficientes de regressão para os dados de avaliação do número de ácaros da falsa ferrugem nos frutos de citros, em função dos diversos níveis de k .

Tabela 13 - Estimativa dos coeficientes de regressão linear múltipla e de cumeieira ($k = 0,5$)

i	Coeficientes de Regressão	
	Linear múltipla ($\hat{b}_i$)	Cumeieira ($\hat{b}_i^*$)
1	1,3823	0,8885
2	-0,1557	-0,0074
3	-0,3441	-0,0642
4	-1,1446	-0,3631
termo constante	-5,9571	-10,2083

Tabela 14 - Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cumeieira ($k = 0,5$)

i	Erro Padrão	
	Linear múltipla s($\hat{b}_i$)	Cumeieira s($\hat{b}_i^*$)
1	0,5827	0,4848
2	0,1385	0,0629
3	0,5053	0,2837
4	0,5408	0,2097

Pelas indicações bibliográficas e pelo resultado apresentado nesta tabela, confirma-se que para as estimativas de cumeieira os erros padrões são menores.

Tabela 15 - Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(4)} = 2,7764$

i	Intervalo de Confiança	
	Linear múltipla $b_i \pm t.s(\hat{b}_i)$	Cumeeira $b_i \pm t.s(\hat{b}_i^*)$
1	(-0,2356 ; 3,0002)	(-0,4576 ; 2,2346)
2	(-0,5403 ; 0,2289)	(-0,1822 ; 0,1674)
3	(-1,7470 ; 1,0588)	(-0,8518 ; 0,7234)
4	(-2,6462 ; 0,3570)	(-0,9454 ; 0,2192)

Em função dos valores de $s(b^*)$ serem menores, a amplitude dos intervalos de confiança no método de cumeeira também foram menores.

Tabela 16 - Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com $|t_{(4;0,05)}| = 2,7764$

i	Teste t	
	Linear múltipla (t_i)	Cumeeira (t_i^*)
1	2,3721	1,8327
2	-1,1244	-0,1172
3	-0,6809	-0,2264
4	-2,1164	-1,7313

Tabela 17 - Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeieira

i	Regressão	
	Linear múltipla	Cumeieira
R^2	0,7263	0,3928
SQR	33,32	56,08
$F^{(1)}$	2,65	0,852
SQD	.1,9344	0,2930
C.V. (%)	34,17	44,34

$$^{(1)}F_{(4;4;0,05)} = 6,39$$

A não significância de teste F nos dois métodos, indica que esses modelos de regressão não se ajustaram aos dados desse experimento.

4.3. Dados de avaliação de danos em grãos de milho armazenados cultivar C-111, causados por gorgulhos (Sitophilus oryzae)

Estes dados apresentam um altíssimo grau de multicolinearidade ($\psi = 2534,78$) que afeta sensivelmente as estima

tivas dos coeficientes de regressão linear múltipla. Nota-se que as estimativas dos coeficientes de regressão $\hat{b}_4$, $\hat{b}_5$, $\hat{b}_6$ e $\hat{b}_7$ tiveram os seus sinais alterados quando comparados com as estimativas de cumeeira, havendo inclusive uma mudança sensível nos seus valores.

Os coeficientes de regressão apresentam estimativas do erro padrão menores no método da regressão de cumeeira e também menor amplitude do intervalo de confiança.

O teste t (5%) apresentou significância para os coeficientes de regressão b_1^* , b_2^* , b_3^* , b_4^* , e b_5^* no método da regressão de cumeeira enquanto que no método da regressão linear múltipla somente o coeficiente de regressão b_1 apresentou significância.

O teste F (5%) foi significativo nos dois métodos. Porém, a equação de regressão deve ser aquela relativa ao método de cumeeira, cuja expressão é:

$$y = 0,3089X_1 + 0,0286X_2 - 0,9256X_3 + 2,1560X_4 + 3,0476X_5 + \\ + 0,6254X_6 + 1,5882X_7 + 0,0357X_8$$

Tabela 18 - Estimativa dos coeficientes de correlação linear simples entre as variáveis

$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$ e y

Varia- veis	X_1	X_2	X_3	Coeficientes de correlação X_4	X_5	linear X_6	simples X_7	X_8	y
X_1	1,0000	0,8416	-0,8332	0,7047	0,5362	0,6023	0,2358	0,7147	0,9964
X_2	0,8416	1,0000	-0,8035	0,7350	0,6743	0,4834	0,1376	0,3911	0,8728
X_3	-0,8332	-0,8035	1,0000	-0,7930	-0,5728	-0,8390	-0,1701	-0,6954	-0,8704
X_4	0,7047	0,7350	-0,7930	1,0000	0,4338	0,4672	0,1087	0,5660	0,7401
X_5	0,5362	0,6743	-0,5728	0,4338	1,0000	0,2510	-0,3221	0,2351	0,6553
X_6	0,6023	0,4834	-0,8390	0,4672	0,2510	1,0000	0,0403	0,6957	0,6001
X_7	0,2358	0,1376	-0,1701	0,1087	-0,3221	0,0403	1,0000	0,0824	0,1647
X_8	0,7147	0,3911	-0,6954	0,5660	0,2351	0,6957	0,0824	1,0000	0,6927
y	0,9964	0,8728	-0,8704	0,7401	0,6553	0,6001	0,1647	0,6927	1,0000

X_1 = número de insetos

X_2 = produto do tempo de armazenamento por:

0 se a parcela não foi infestada e

1 se a parcela foi infestada

X_3 = teor de carboidratos (%)

X_4 = teor de umidade (%)

X_5 = teor de matéria seca (%)

X_6 = teor de proteínas (%)

X_7 = teor de ácidos graxos (%)

X_8 = índice de acidez dos ácidos graxos (%)

y = perda de Peso (%)

Figura 3 - Valores dos coeficientes de regressão para os dados de avaliações de danos em grãos, de milho C-111, causados por gorgulhos, em função dos diversos níveis de k .

Tabela 19 - Estimativa dos coeficientes de regressão linear múltipla e de cumeieira ($k = 0,8$)

i	Coeficientes de Regressão	
	Linear múltipla ($\hat{\beta}_i$)	Cumeieira ($\hat{\beta}_i^*$)
1	0,7939	0,3089
2	0,0159	0,0286
3	-9,0441	-0,9256
4	-8,1911	2,1560
5	-4,2009	3,0476
6	-8,8429	0,6254
7	-6,8975	1,5882
8	0,0336	0,0357
termo constante	880,1689	29,5258

Tabela 20 - Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cumeieira ($k = 0,8$)

i	Erro	Padrão
	linear múltipla $s(\hat{\beta}_i)$	Cumeieira $s(\hat{\beta}_i^*)$
1	0,1374	0,0277
2	0,0168	0,0036
3	9,0142	0,1105
4	9,4980	0,5888
5	8,7092	0,5451
6	9,0572	0,3201
7	8,8716	0,9519
8	0,0212	0,0071

Pelos resultados desta tabela, os erros padrões das estimativas de cumeieira são menores, como nos casos anteriores, principalmente para as variáveis que foram afetadas pela multicolinearidade, tais como X_3 , X_4 , X_5 , X_6 e X_7 .

Tabela 21 - Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeeira, com

$$t_{(27)} = 2,0518$$

i	Intervalo de Confiança	
	Linear múltipla $\hat{b}_i \pm t.s(\hat{b}_i)$	Cumeeira $\hat{b}_i^* \pm t.s(\hat{b}_i^*)$
1	(0,5120 ; 1,0758)	(0,0568 ; 0,3657)
2	(-0,0186 ; 0,0504)	(0,0212 ; 0,0360)
3	(-27,5394 ; 9,4512)	(-1,1524 ; -0,6988)
4	(-27,6790 ; 11,2968)	(0,9478 ; 3,3642)
5	(-22,0704 ; 13,6686)	(1,9291 ; 4,1661)
6	(-27,4265 ; 9,7407)	(-0,0314 ; 1,2822)
7	(-25,1004 ; 11,3054)	(-0,3650 ; 3,5414)
8	(-0,0100 ; 0,0772)	(0,0210 ; 0,0504)

Com os valores de $s(b^*)$ menores, a amplitude dos intervalos de confiança também serão menores, ressaltando as variáveis X_3 , X_4 , X_5 , X_6 e X_7 .

Tabela 22 - Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com
 $|t_{(27;0,05)}| = 2,0518$

i	Teste t	
	Linear múltipla t_i	Cumeeira t_i^*
1	5,7763 (*)	11,1584 (*)
2	0,9494	7,9729 (*)
3	-1,0033	-8,3752 (*)
4	-0,8624	3,6614 (*)
5	-0,4823	5,5907 (*)
6	-0,9763	1,9536
7	-0,7775	1,6684
8	1,5821	4,9950 (*)

(*) significativo ao nível de 5% de probabilidade.

Com os resultados desta tabela, verifica-se que, no método de cumeeira, o teste t apresenta maior quantidade de valores significativos para as estimativas dos coeficientes de regressão.

Tabela 23 - Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira

	Regressão	
	Linear múltipla	Cumeeira
R^2	0,9709	0,7921
SQR	84,6238	239,3121
$F^{(1)}$	112,81 (*)	32,63 (*)
SQD	2,2657	0,1574
C.V. (%)	10,63	17,86

$^{(1)}F_{(8;27;0,05)} = 2,31$

(*) significativo ao nível de 5% de probabilidade

4.4. Dados de avaliações de danos em grãos de milho armazenados cultivar Flint composto, causados por gorgulhos (Sitophilus oryzae)

Para os dados desse experimento percebe-se que há um altíssimo grau de multicolinearidade ($\psi = 2812,26$), dificultando a determinação do valor de k pois nota-se que a estabilização das estimativas dos coeficientes de regressão ocorrerá para um valor de $k > 1,0$, conforme pode ser visto pelo gráfico 4.

Pelo fato dos dados terem sido analisados para somente até $k = 1,0$, ele será adotado para obter as estimativas dos coeficientes de regressão, embora não sejam as adequadas mas pode-se verificar pelos resultados (tabela 25) que há uma mudança sensível nos valores das estimativas dos coeficientes de regressão de cumeeira em relação às estimativas dos coeficientes de regressão linear múltipla, inclusive ocorrendo a troca de seus sinais, nos casos de $\hat{b}_3^*$, $\hat{b}_7^*$ e $\hat{b}_8^*$.

As estimativas dos erros padrões e dos intervalos de confiança dos coeficientes de regressão apresentaram menores valores e amplitudes, respectivamente, na regressão de cumeeira.

O teste F (5%) apresentou significância nos dois métodos. A semelhança do caso anterior, a equação de regressão deve ser a referente ao método de cumeeira, cuja expressão é:

$$y = 0,2417X_1 + 0,0220X_2 - 0,6759X_3 + 1,6572X_4 + \\ + 0,4571X_5 + 0,7029X_6 - 0,6281X_7 + 0,0129X_8$$

Tabela 24 - Estimativa dos coeficientes de correlação linear simples entre as variáveis

$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$ e γ

Variáveis	Coeficientes de correlação linear simples							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	γ
X_1	1,0000	0,8795	-0,7700	0,7278	0,1885	0,6497	-0,2100	0,9172
X_2	0,8795	1,0000	-0,8261	0,7719	0,2910	0,6199	-0,1272	0,8902
X_3	-0,7700	-0,8261	1,0000	-0,9221	-0,4061	-0,8492	0,3854	-0,8910
X_4	0,7278	0,7719	-0,9221	1,0000	0,3808	0,7525	-0,5404	0,9029
X_5	0,1884	0,2910	-0,4061	0,3808	1,0000	-0,0017	-0,0276	0,2668
X_6	0,6497	0,6199	-0,8492	0,7525	-0,0017	1,0000	-0,6228	0,7505
X_7	-0,2100	-0,1272	0,3854	-0,5404	-0,0276	-0,6228	1,0000	-0,3812
X_8	0,5794	-0,3425	-0,2988	0,2731	0,0248	0,4308	-0,4000	0,4099
γ	0,9172	0,8902	-0,8910	0,9029	0,2668	0,7505	-0,3812	1,0000

X_1 = número de insetos

X_2 = produto do tempo de armazenamento por:

0 se a parcela não foi infestada e

1 se a parcela foi infestada

X_3 = teor de carboidratos (%)

X_4 = teor de umidade (%)

X_5 = teor de matéria seca (%)

X_6 = teor de proteínas (%)

X_7 = teor de ácidos graxos (%)

X_8 = índice de acidez dos ácidos graxos (%)

y = perda de peso (%)

Figura 4 - Valores dos coeficientes de regressão para os dados de avaliações de danos em grãos de milho Flint Composto, causados por gorgulhos, em função dos diversos níveis de k .

Tabela 25 - Estimativa dos coeficientes de regressão linear múltipla e de cuneeira (k = 1,0)

i	Coeficiente de Regressão	
	Linear múltipla $\hat{\beta}_i$	Cuneeira $\hat{\beta}_i^*$
1	0,6383	0,2417
2	0,0182	0,0220
3	12,0743	-0,6759
4	15,5857	1,6572
5	11,5939	0,4571
6	12,2976	0,7029
7	11,0107	-0,6281
8	-0,0416	0,0129
termo constante	11,2913	24,6395

Tabela 26 - Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cuneeira

i	Erro Padrão	
	Linear Múltipla	Cuneeira
	$s(\hat{\beta}_i)$	$s(\hat{\beta}_i^*)$
1	0,1176	0,0242
2	0,0113	0,0026
3	5,6553	0,0732
4	5,7436	0,1758
5	5,6295	0,5870
6	5,6844	0,1483
7	5,7193	0,4089
8	0,0195	0,0097

Tabela 27 - Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeeira, com $t_{(27)} = 2,0518$

i	Intervalo de Confiança	
	Linear Múltipla	Cumeeira
	$\hat{\beta}_i \pm t.s(\hat{\beta}_i)$	$\hat{\beta}_i^* \pm t.s(\hat{\beta}_i^*)$
1	(0,3970 ; 0,8796)	(0,1920 ; 0,2914)
2	(-0,0049 ; 0,0413)	(0,0166 ; 0,0274)
3	(0,4708 ; 23,6778)	(-0,8262 ; -0,5256)
4	(3,8009 ; 27,3705)	(1,2965 ; 2,0179)
5	(0,0433 ; 23,1445)	(-0,7474 ; 1,6616)
6	(0,6343 ; 23,9609)	(0,3987 ; 1,0071)
7	(-0,7242 ; 22,7456)	(-1,4671 ; 0,2109)
8	(-0,0816 ; -0,0016)	(-0,0007 ; 0,0328)

Tabela 28 - Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com

$$|t_{(27;0,05)}| = 2,0518$$

i	Teste t	
	Linear múltipla t_i	Cumeeira t_i^*
1	5,4284 (*)	9,9723 (*)
2	1,6181	8,4372 (*)
3	2,1350 (*)	-9,2263 (*)
4	2,7136 (*)	9,4262 (*)
5	8,0595 (*)	0,7786
6	2,1634 (*)	4,7406 (*)
7	1,9252	-1,5362
8	-2,1350 (*)	1,3254

(*) significativo ao nível de 5% de probabilidade

Pelos resultados desta tabela e pelo gráfico da figura 4 percebe-se que há necessidade de aplicar o método de cumeeira para valores de k maiores do que 1,0 e assim, obter um melhor resultado do que o da regressão linear múltipla.

Tabela 29 - Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira

	Regressão	
	Linear múltipla	Cumeeira
R^2	0,9711	0,7570
SQR	47,3374	165,2283
$F^{(1)}$	113,64 (*)	25,63 (*)
SQD	14,1082	0,1433
C.V. (%)	9,47	17,60

$^{(1)}F_{(8;27;0,05)} = 2,31$

(*) significativo ao nível de 5% de probabilidade.

4.5. Dados de avaliação de danos em grãos de milho armazenado do cultivar Piranão, causados por gorgulhos (Sitophilus oryzae)

Para os dados desse experimento notou-se um alto grau de multicolinearidade ($\psi = 263,22$) que afeta sensivelmente algumas das estimativas dos coeficientes de regressão linear múltipla, tais como b_3 , b_5 , b_6 , b_7 e b_8 e podem ser vistos na figura 5. Este fator pode também ser justificado pelas altas correlações entre as variáveis independentes apresentadas na tabela 30.

Com a significância do teste $F(5\%)$ nos dois métodos, verifica-se que a equação de regressão referente ao método de cumeeira melhor se adapta aos dados e a equação de regressão é dada pela seguinte expressão:

$$y = 0,3198 x_1 + 0,0335 x_2 - 0,4380 x_3 + 0,0034 x_4 + \\ + 0,7832 x_5 + 0,2683 x_6 - 0,9774 x_7 + 0,0419 x_8 \cdot \cdot$$

Obteve-se a significância para teste $t(5\%)$ dos coeficientes b_1^* , b_2^* , b_3^* , b_4^* , b_5^* , b_7^* e b_8^* no método da regressão de cumeeira, enquanto que a regressão linear múltipla somente os coeficientes b_1 , b_2 , b_3 , b_6 e b_7 apresentam significância.

Tabela 30 - Estimativa dos coeficientes de correlação linear simples entre as variáveis $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$ e y

Variáveis	Coeficientes de correlação linear simples								
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	y
X ₁	1,0000	0,8807	-0,6348	0,2030	0,4911	0,4905	-0,4154	0,7098	0,9367
X ₂	0,8807	1,0000	-0,6751	0,2201	0,5018	0,5030	-0,3328	0,5945	0,9047
X ₃	-0,6348	-0,6751	1,0000	-0,5005	-0,6380	-0,9230	0,3700	-0,7163	-0,7617
X ₄	0,2030	0,2201	-0,5005	1,0000	0,1552	0,4909	-0,0592	0,2942	0,3533
X ₅	0,4911	0,5018	-0,6380	0,1552	1,0000	0,3897	-0,7473	0,7342	0,6125
X ₆	0,4905	0,5030	-0,9234	0,4909	0,3897	1,0000	-0,2496	0,6027	0,5909
X ₇	-0,4154	-0,3328	0,3700	-0,0592	-0,7473	-0,2496	1,0000	-0,7833	-0,5173
X ₈	0,7098	0,5945	-0,7163	0,2942	0,7342	0,6027	-0,7833	1,0000	0,7915
y	0,9367	0,9047	-0,7617	0,3533	0,6125	0,5909	-0,5173	0,7913	1,0000

X_1 = número de insetos

X_2 = produto do tempo de armazenamento por:

0 se a parcela não foi infestada

1 se a parcela foi infestada

X_3 = teor de carboidratos (%)

X_4 = teor de umidade (%)

X_5 = teor de matéria seca (%)

X_6 = teor de proteínas (%)

X_7 = teor de ácidos graxos (%)

X_8 = índice de acidez dos ácidos graxos (%)

y = perda de peso (%)

Figura 5 - Valores dos coeficientes de regressão para os dados de avaliações de danos em grãos de milho Piratã causados por gorgulhos, em função dos diversos níveis de k .

Tabela 31 - Estimativa dos coeficientes de regressão linear múltipla e de cumeieira ($k = 0,6$)

i	Coeficientes de Regressão	
	Linear múltipla $\hat{b}_i$	Cumeieira $\hat{b}_i^*$
1	0,5310	0,3198
2	0,0314	0,0335
3	-2,4436	-0,4380
4	0,0041	0,0034
5	-2,3774	0,7832
6	-2,5178	0,2683
7	-2,5793	-0,9774
8	0,0228	0,0419
termo constante	0,7443	36,1045

Tabela 32 - Estimativa do erro padrão dos coeficientes de regressão linear múltipla e de cumeeira ($k = 0,6$)

i	Erro	Padrão
	Linear múltipla	Cumeeira
	$s(\hat{\beta}_i)$	$s(\hat{\beta}_i^*)$
1	0,1028	0,0306
2	0,0116	0,0036
3	0,9725	0,0797
4	0,0021	0,0016
5	1,3961	0,3733
6	1,1150	0,1747
7	1,2789	0,4143
8	0,0311	0,0078

Tabela 33 - Intervalo de confiança ($\alpha = 0,05$) para os coeficientes de regressão linear múltipla e de cumeieira, com $t_{(27)} = 2,0518$

i	Intervalo de Confiança	
	Linear múltipla	Cumeieira
	$\hat{\beta}_i \pm t.s(\hat{\beta}_i)$	$\hat{\beta}_i^* \pm t.s(\hat{\beta}_i^*)$
1	(0,3200 ; 0,7420)	(0,2569 ; 0,3827)
2	(0,0075 ; 0,0553)	(0,0261 ; 0,0409)
3	(-4,4389 ; -0,4483)	(-0,6016 ; -0,2744)
4	(-0,0003 ; 0,0085)	(-0,0001 ; 0,0067)
5	(-5,2418 ; 0,4870)	(0,0173 ; 1,5491)
6	(-4,8056 ; -0,2300)	(-0,0902 ; 0,6268)
7	(-5,2035 ; 0,0449)	(-1,8276 ; -0,1272)
8	(-0,0411 ; 0,0867)	(0,0258 ; 0,0580)

Tabela 34 - Teste "t" de Student para os coeficientes de regressão linear múltipla e de cumeeira, com

$$|t_{(27;0,05)}| = 2,0518$$

i	Teste	t
	Linear múltipla	Cumeeira
	t_i	t_i^*
1	5,1626 (*)	10,4302 (*)
2	2,7037 (*)	9,2568 (*)
3	-2,5128 (*)	-5,4938 (*)
4	1,9176	2,1211 (*)
5	-1,7029	2,0979 (*)
6	-2,2581 (*)	1,5359
7	-2,1574 (*)	-2,3589 (*)
8	0,7323	5,3626 (*)

(*) significativo ao nível de 5% de probabilidade.

Nesta tabela observa-se que o método de cumeeira apresenta uma maior quantidade de valores significativos para os coeficientes de regressão assim, um melhor resultado do que o método da regressão linear múltipla.

Tabela 35 - Coeficiente de determinação (R^2), soma de quadrados do resíduo (SQR), teste F da regressão (F), soma de quadrados dos desvios de regressão (SQD) e coeficiente de variação (C.V.) em % para as regressões linear múltipla e de cumeeira

	Regressão	
	Linear múltipla	Cumeeira
R^2	0,9669	0,8093
SQR	61,8639	142,0921
$F^{(1)}$	98,75 (*)	36,01 (*)
SQD	0,8800	0,1913
C.V. (%)	10,56	15,99

$^{(1)}F_{(8;27;0,05)} = 2,31$

(*) significativo ao nível de 5% de probabilidade.

Pelas análises desses experimentos pode-se comprovar a propriedade da menor variância das estimativas dos coeficientes de regressão e menor soma de quadrados dos desvios de regressão pelo método da regressão de cumeira do que pelo método da regressão linear múltipla.

Com relação ao coeficiente de variação e ao coeficiente de determinação observou-se um aumento nos seus valores. No valor da estatística F , da análise de variância, houve uma diminuição, que é uma consequência esperada em função do aumento do valor da soma de quadrados do resíduo pelo método da cumeira, para um determinado valor de k .

Observou-se que quando o teste F da regressão linear múltipla é não significativo, a análise de regressão de cumeira apresentará também uma indicação do teste F não significativo.

Com a significância do teste F da regressão linear múltipla e da regressão de cumeira para um determinado valor de k , o teste t , neste último, apresentou uma maior quantidade de valores significativos para as estimativas dos coeficientes de regressão.

5. CONCLUSÕES

1- Pelos dados dos experimentos agronômicos utilizados neste trabalho confirmaram-se as seguintes conclusões principais dadas por HOERL e KENNARD (1970 a, b).

a) Os estimadores dos coeficientes de regressão pelo método da regressão de cumeeira são tendenciosos.

b) As estimativas dos coeficientes de regressão pelo método de cumeeira têm a vantagem de apresentar menores variâncias em relação ao método da regressão linear múltipla.

c) Levando-se em conta a soma de quadrados dos desvios de regressão como um critério de comparação dos dois métodos, o método da regressão de cumeeira é o que apresentou melhores resultados ou seja, uma menor soma de quadrados dos desvios de regressão.

2- Independente do alto grau de multicolinearidade não é possível encontrar um valor exato para k , pelo método proposto

por HOERL e KENNARD, pois existirão diversos pontos que podem ser considerados como o início da estabilização dos coeficientes de regressão, tornando k uma variável aleatória.

3- Através do gráfico de cumeeira a escolha do valor de k é bastante subjetiva principalmente quando o grau de multicolinearidade é altíssimo, pois alguns dos coeficientes não tendem à estabilização dentro do intervalo de k , comumente utilizado, sendo necessário então, uma análise para valores de k superiores a 1,0.

4- Quando o teste F da regressão linear múltipla é não significativo a análise de regressão de cumeeira apresentará também uma indicação do teste F não significativo, de forma que se aconselha a análise de regressão de cumeeira somente quando o teste F da análise de regressão linear múltipla é significativo. Fica como um dos critérios para escolha do valor de k , um valor tal que o teste F continue apresentando significância.

5- Na prática, para os experimentos agronômicos o método da regressão de cumeeira é uma das alternativas em relação ao método da regressão linear múltipla quando se evidencia um alto grau de multicolinearidade entre as variáveis independentes, principalmente por apresentar estimativas dos coeficientes de regressão com maior precisão.

6. BIBLIOGRAFIA

BESLEY, D. A.; E. KUH e R. E. WELSCH, 1980. Regression Diagnostics. New York, John Wiley e Sons, 292 pp.

CHATTERJEE, S. e B. PRICE, 1977. Regression Analysis by Example. New York, John Wiley e Sons, 228 pp.

GRAYBILL, F. A., 1976. Theory and Application of the Linear Model. California, Duxbury Press, 704 pp.

HOCKING, R. R., 1976. The Analysis and Selection of Variables in Linear Regression. Biometrics, 32 : 1 - 40.

HOERL, A. E., 1962. Application of Ridge Analysis to Regression Problems. Chemical Engineering Progress, New York , 58(3) : 53 - 60.

- HOERL, A. E. e R. W. KENNARD, 1970(a). Ridge Regression : Biased Estimation for Nonorthogonal Problems. Technometrics, Richmond, 12 : 55 - 67.
- HOERL, A. E. e R. W. KENNARD, 1970(b). Ridge Regression : Applications to Nonorthogonal Problems. Technometrics , Richmond, 12 : 69 - 82.
- HUBER, P. J., 1981. Robust Statistics. Nova York, John Wiley, 308 pp.
- JOHNSTON, J., 1972. Econometric Methods. 2ª edição. New York, McGraw-Hill Book Company , 437 pp.
- KALIL, M. N., 1977. Aplicação do Método de Regressão de Cumeei ra ("Ridge Regression") na Estimação de Funções de Demanda e de Produção. Piracicaba, ESALQ/USP, 136 pp. (Dissertação de Mestrado).
- KALIL, M. N., 1978. Programa "Análise de Regressão pelo Método de Cumeeira ("Ridge Regression"). Piracicaba, ESALQ/USP , 25 pp. (Série Manual nº 2).
- KMENTA, J., 1971. Elementos de Econometria. 1ª edição, São Paulo, Editora Atlas, 670 pp.
- MARQUADT, D. W., 1970. Generalized Inverse, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation. Technometrics, Richmond, 12 : 591 - 612.

- MARQUADT, D. W. e P. R. SNEE, 1975. Ridge Regression in Practice. American Statistician, Washington, 29 : 3 - 20.
- MATTIOLLI, C. H., 1983. Métodos de Seleção de Equações de Regressão Linear Múltipla. Piracicaba, ESALQ/USP, 104 pp. (Dissertação de Mestrado).
- MATTIOLLI, J. C., 1978. Efeitos da Intensidade da População de Sitophilus oryzae (Linné, 1763) (Coleoptera, curculionidae) e do Período de Armazenamento sobre Diferentes Variedades de Milho. Curitiba, Universidade Federal do Paraná, 192 pp. (Dissertação de Mestrado).
- MCDONALD, G. C. e D. I. GALARNEAU, 1975. A Monte Carlo Evaluation of Some Ridge - Type Estimators. Journal of the American Statistical Association, Washington, 70 (350) : 407-416.
- NAKANO, O., 1984. Ensaio Visando o Controle do Ácaro da Falsa Ferrugem (Phyllocoptruta oleivora), Ácaro Amarelo (Lorrya formosa) e Ácaro da Leprose (Brevipalpus phoenicis) em Citros (Citrus sinensis var. Natal). Piracicaba, ESALQ/USP, 26 pp. (Relatório do Experimento).
- NEWELL, G. J. e B. LEE, 1981. Ridge Regression: An Alternative to Multiple Linear Regression for Highly Correlated Data. Journal of Food Science, 46 : 968 - 969.

- SEARLE, S. R., 1971. Linear Models. New York, John Wiley e Sons, 531 pp.
- SUNG, H. P., 1981. Collinearity and Optimal Restrictions on Regression Parameter for Estimating Responses. Technometrics, Richmond, 23 : 289 - 296.
- THEOBALD, C. M., 1974. Generalizations of Mean Square Error Applied to Ridge Regression. J. Royal Statistical Society , Series B(36) : 103 - 106.
- WEISBERG, S., 1981. A Statistic for Allocating Cp to individual cases. Technometrics, Richmond, 23 : 27 - 32.