

LUCIANE DA SILVA OLIVEIRA

**SELEÇÃO DE COVARIÁVEIS PARA AJUSTE DE REGRESSÃO LOGÍSTICA  
NA ANÁLISE DA ABUNDÂNCIA DE INVERTEBRADOS EDÁFICOS EM  
DIFERENTES AGROECOSSISTEMAS**

Dissertação apresentada à  
Universidade Federal de Viçosa,  
como parte das exigências do  
Programa de Pós-Graduação em  
Estatística Aplicada e Biometria,  
para obtenção do título de  
*Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2011

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

O48s  
2011

Oliveira, Luciana da Silva, 1978-

Seleção de covariáveis para ajuste de regressão logística na análise de abundância de invertebrados edáficos em diferentes agroecossistemas / Luciana da Silva Oliveira.

– Viçosa, MG, 2011.

xi, 63f. : il. ; 29cm.

Inclui apêndice.

Orientador: Paulo Roberto Cecon.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 41-43

1. Análise de regressão. 2. Regressão. 3. Solos - Manejo.  
I. Universidade Federal de Viçosa. II. Título.

CDD 22. ed. 519.536

LUCIANE DA SILVA OLIVEIRA

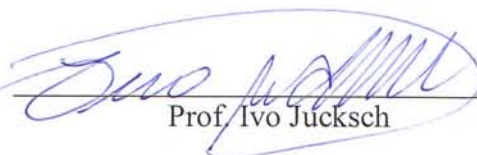
**SELEÇÃO DE COVARIÁVEIS PARA AJUSTE DE REGRESSÃO LOGÍSTICA  
NA ANÁLISE DA ABUNDÂNCIA DE INVERTEBRADOS EDÁFICOS EM  
DIFERENTES AGROECOSSISTEMAS**

Dissertação apresentada à  
Universidade Federal de Viçosa,  
como parte das exigências do  
Programa de Pós-Graduação em  
Estatística Aplicada e Biometria,  
para obtenção do título de  
*Magister Scientiae*.

APROVADA: 25 de fevereiro de 2011



Prof. Antonio Policarpo Souza Carneiro



Prof. Ivo Jucksch



Prof. Sebastião Martins Filho  
(Coorientador)



Prof. Paulo Roberto Cecon  
(Orientador)

*À minha família, pelo  
conforto nas horas de angústia,  
pelo carinho, amor e confiança.*

*“O correr da vida embrulha tudo, a vida é assim:  
esquenta e esfria, aperta e daí afrouxa,  
sossega e depois desinquieta.  
O que ela quer da gente é coragem.”*  
Guimarães Rosa

## AGRADECIMENTOS

Agradeço a Deus, por estar sempre presente em minha vida e permitir alcançar mais esse objetivo. Por colocar sempre em meu caminho as pessoas certas, nas horas certas.

À minha mãe, Luzia, pela força e compreensão e pelo exemplo na minha vida, sempre participando e me estimulando a ir mais além.

Ao meu pai, Adilson, que mesmo distante sempre me incentivou nos estudos e na profissão.

Às minhas irmãs Viviane e Emiliane pelos direcionamentos, por ter me escutado nas horas de angústia e indicado o melhor caminho a seguir nos momentos de dúvidas. Pelo apoio e incentivo sempre.

Aos amigos Alessandro, Erika e Andréia que dividiram as angústias e alegrias dessa fase da vida e sempre me incentivaram a vencer os desafios e obstáculos do caminho.

Às amigas Elisângela e Diana pelo estímulo, apoio e companheirismo durante todo curso.

Aos amigos e colegas do mestrado André, Rogério, Fernanda, Liliane, Gemma, Luis, Gilson, Iury, Rodrigo, Marconi e Suelen pelos momentos inesquecíveis de convivência e amizade.

Ao Prof. Dr. Paulo Roberto Cecon, pela orientação na realização deste trabalho, pelo carinho, atenção, simplicidade e generosidade como ser humano.

Aos meus Coorientadores Prof. Dr. Sebastião Martins Filho e Prof. Dr. Fabyano Fonseca e Silva pelo conhecimento, paciência, disponibilidade e ajuda necessária para a realização desta dissertação.

Aos membros da banca Prof. Dr. Antonio Policarpo Souza Carneiro, por todo o carinho e contribuições nesta conquista e Prof. Dr. Ivo Jucksch, pela amizade de sempre, carinho, incentivo e contribuições na análise dos dados.

À Universidade Federal de Viçosa, por intermédio do Programa de Pós Graduação de Estatística Aplicada e Biometria, pela oportunidade.

A todos os professores do Departamento de Estatística da Universidade Federal de Viçosa que contribuíram para minha formação, crescimento pessoal e profissional.

Ao coordenador do curso de mestrado em Estatística Aplicada e Biometria da UFV, Prof. Dr. Luiz Alexandre Peternelli, pelo exemplo de profissionalismo.

Ao antigo secretário Altino e o atual, Joel, pela eficiência sempre demonstrada e apoio técnico.

À secretária do DET-UFV Anita, pela presteza e simpatia.

À Dra. Joedna Silva por me fornecer o banco de dados para realização deste trabalho.

À Secretaria Estadual de Educação de Minas Gerais e Prefeitura Municipal de Espera Feliz/MG por autorizar meu afastamento das atividades para realização desse curso e pelo apoio financeiro.

À Faculdade do Vale do Carangola – UEMG/FAVALE pelo incentivo e apoio na formação profissional.

A todos que, de alguma forma, ajudaram e contribuíram para que este trabalho pudesse ser realizado. Agradeço às orações, as críticas e as sugestões que só fizeram-me crescer.

## **BIOGRAFIA**

LUCIANE DA SILVA OLIVEIRA, filha de Luzia da Silva Oliveira e Adilson Pinto de Oliveira, nasceu em 25 de fevereiro de 1978, em Carangola/MG.

Em 2000 graduou-se em licenciatura plena em Matemática pela Faculdade de Filosofia, Ciências e Letras de Carangola, hoje Faculdades Vale do Carangola – FAVALE/UEMG, em Carangola/MG.

Antes mesmo de sua formação acadêmica, já trabalhava como professora do Ensino Básico nas escolas de Espera Feliz e Caiana/MG.

Em 2005 concluiu o curso de especialização em Métodos Estatísticos Computacionais, pela Universidade Federal de Juiz de Fora - UFJF com a monografia intitulada: Aplicação de métodos estatísticos para análise da adesão ao exame de Papanicolau.

Em março de 2009, iniciou o curso de Mestrado em Estatística Aplicada e Biometria, na Universidade Federal de Viçosa – UFV, submetendo-se à defesa da dissertação em fevereiro de 2011.

Atualmente, é professora efetiva de Matemática da Secretaria Estadual de Educação de Minas Gerais, no município de Caparaó/MG e da Prefeitura Municipal de Espera Feliz/MG, além de trabalhar como professora e coordenadora do curso de Matemática das Faculdades Vale do Carangola – FAVALE/UEMG, em Carangola/MG.

## SUMÁRIO

RESUMO.....	viii
ABSTRACT.....	ix
1 INTRODUÇÃO .....	1
2 REFERENCIAL TEÓRICO .....	3
2.1 Importância dos invertebrados edáficos .....	3
2.2 Modelo de regressão logística simples .....	4
2.1.1 Estimação dos Parâmetros.....	6
2.1.2 Interpretação dos Coeficientes .....	8
2.3 Modelo de regressão logística múltipla.....	9
2.2.1 Estimação dos Parâmetros.....	10
2.3 Variáveis dummy no modelo .....	12
2.4 Teste de significância dos coeficientes .....	13
2.4.1 Teste da Razão de Verossimilhança.....	13
2.4.2 Teste de Wald.....	15
3 MATERIAL E MÉTODOS .....	16
3.1 Área de estudo/Procedência dos dados .....	16
3.2 Construção do modelo.....	17
3.3 Seleção de Covariáveis .....	19
3.4 Medidas de qualidade do ajuste .....	20
3.4.1 Teste de Hosmer e Lemeshow .....	20
3.4.2 Área sob a Curva ROC.....	22
4 RESULTADOS E DISCUSSÃO .....	26
4.1 Conjunto de dados da serapilheira e do solo no período seco – atributos físicos e químicos .....	26
4.2 Conjunto de dados da Serapilheira e do solo no período chuvoso – atributos físicos e químicos.....	33
5 CONCLUSÕES E RECOMENDAÇÕES .....	39
5.1 Conclusões .....	39
5.2 Recomendações.....	40
REFERÊNCIAS BIBLIOGRÁFICAS.....	41
APÊNDICE .....	44
APÊNDICE A – .....	46
APÊNDICE B – .....	55

## RESUMO

OLIVEIRA, Luciane da Silva, M.Sc., Universidade Federal de Viçosa, fevereiro de 2010. **Seleção de covariáveis para ajuste de Regressão Logística na análise da abundância de invertebrados edáficos em diferentes agroecossistemas.** Orientador: Paulo Roberto Cecon. Coorientadores: Sebastião Martins Filho e Fabyano Fonseca e Silva.

A regressão logística é o método estatístico usual de análise utilizado com a finalidade de verificar a relação entre uma variável resposta dicotômica e variáveis explicativas de interesse. Este trabalho teve como objetivo realizar um estudo sobre os fatores que influenciam a abundância de invertebrados no solo sob diferentes formas de manejo utilizando a Regressão Logística. Tal objetivo reside no fato destes invertebrados serem considerados excelentes indicadores do tipo de uso e qualidade do solo, atuando em vários processos fundamentais para a manutenção da fertilidade e qualidade dos solos de agroecossistemas e ecossistemas naturais de acordo com Brown et al. (1998) e Hendrix et al. (2006), citado Souza (2010). Para seleção de covariáveis foi utilizada a proposta de Collett (1994) e foram apresentados estimadores dos parâmetros envolvidos em cada modelo e suas interpretações, propriedades estatísticas e critérios para se julgar a adequabilidade dos modelos selecionados. A metodologia apresentada neste trabalho foi aplicada a dois conjuntos de dados reais (período seco e chuvoso). No modelo final ajustado para o conjunto de dados analisado no período seco verificou-se que as covariáveis *Tipo de Sistema*, *Cálcio em serapilheira*, *Matéria orgânica do solo*, *Potássio em serapilheira* e a interação entre *Cálcio e Potássio em serapilheira* foram importantes para explicar a presença de mais de 9 indivíduos, em média, no solo. Já no modelo final ajustado para o conjunto de dados analisado no período chuvoso, as covariáveis significativas para explicar a presença de 101 indivíduos, em média, no solo foram *Magnésio em serapilheira*, *Carbono orgânico total na serapilheira*, *Matéria orgânica da serapilheira* e *Temperatura ambiente*. Para os dois modelos citados houve bom desempenho discriminatório e excelentes áreas sob a curva ROC, confirmando assim a validade da utilização de técnicas de regressão logística na construção dos modelos para descrever os dados analisados.

## ABSTRACT

OLIVEIRA, Luciane da Silva, M.Sc., Universidade Federal de Viçosa, February, 2010. **Covariates selection for Logistic Regression adjustment in analysis of edaphic invertebrates abundance in different agroecosystems.** Advisor: Paulo Roberto Cecon. Co-Advisors: Sebastião Martins Filho and Fabyano Fonseca e Silva.

Logistic regression is the analysis' usual statistical method used to verify the relationship between a dichotomous variable response and the interest explanatory variables. This work aimed to carry out a study about the factors influencing the invertebrates' abundance on the soil under different management forms, using the logistic regression. This objective is that these invertebrates are considered excellent indicators of the use type and soil quality, working in several fundamental processes for maintaining the soil fertility and quality in agroecosystems and natural ecosystems, according to Brown et al. (1998), Hendrix et al. (2006), and Souza (2010). For covariates selection, the Collett (1994) proposal was used and the involved parameters estimators in each model, their interpretations, statistical properties, and some criteria for judging the suitability of the selected models were presented. The methodology presented by this work was applied to two real datasets (dry and rainy season). In the final adjusted model for the analyzed dataset in the dry season, it was verified that the covariates *System Type*, *Calcium in litter*, *Soil organic matter*, *Potassium in litter*, and the interaction between *Calcium* and *Potassium in litter* were important to explain the presence of more than nine individuals on the soil. In the final adjusted model for the analyzed dataset in the rainy season, the significant covariates to explain the presence of one hundred and one individuals on average on the soil were *Magnesium in litter*, *Total organic carbon in the litter*, *Litter organic matter*, and *Ambient temperature*. For two mentioned models, there were a good discriminatory performance and excellent areas under the ROC (Receiver Operating Characteristic) curve, thus confirming the validity of using logistic regression techniques for the models construction to describe the analyzed data.

## 1 INTRODUÇÃO

A regressão logística é o método estatístico usual de análise utilizado com a finalidade de verificar a relação entre uma variável resposta dicotômica e variáveis explicativas de interesse. A análise logística controla grande número de variáveis simultaneamente, permitindo que os dados sejam utilizados mais eficientemente.

Na regressão logística a variável resposta ( $Y$ ), geralmente binária ou dicotômica, apresenta duas possibilidades de resposta (sucesso ou fracasso). Ao “sucesso”, resultado mais importante ou aquele que se relaciona o acontecimento de interesse, geralmente atribui-se o valor 1 ( $y = 1$ ), e ao resultado complementar “fracasso” o valor 0 ( $y = 0$ ).

Um modelo de regressão logística prevê a probabilidade direta de um evento ocorrer e têm sido amplamente aplicado em importantes áreas como Agronomia, Biologia, Engenharia, Economia, Mineração, Transportes, Farmacologia, Medicina e nas Ciências Sociais.

Os modelos de regressão logística fazem parte da classe dos modelos lineares generalizados, ou seja, daqueles que se tornam lineares por meio da aplicação de algum tipo de transformação.

Nesse estudo foi verificado se a presença de invertebrados no solo é mais ou menos abundante considerando alguns fatores como o tipo de manejo agrícola em cultivos de café, a associação entre grupos da fauna edáfica e os atributos físicos, químicos e abióticos do solo e da serapilheira.

A diversidade e a abundância da fauna invertebrada do solo, assim como a presença de determinados grupos de organismos em um sistema, podem ser usadas como indicadores eficientes da qualidade dos solos (PAOLETTI (1999); BARROS et al., 2003) mas, podem ser afetadas por vários fatores edáficos (tipo de solo, minerais predominantes, temperatura, pH, matéria orgânica, umidade, textura e estrutura), eventos históricos (antropogênico e geológico), topográficos e climáticos (MELO *et al.* 2009).

Os invertebrados edáficos atuam em vários processos fundamentais para a manutenção da fertilidade e qualidade dos solos de agroecossistemas e ecossistemas naturais, exercem papel central na decomposição da matéria orgânica do solo e resíduos vegetais, influenciando a disponibilidade de nutrientes (BROWN et al. 1998, HENDRIX et al., 2006 *apud* SOUZA, 2010). São capazes de melhorar a estrutura do

solo pelo estabelecimento de relações com os microorganismos ou de forma direta, pela digestão, transporte e incorporação de partículas orgânicas (SILVA, 2010).

Esse trabalho teve como objetivo realizar um estudo sobre os fatores que influenciam a abundância de invertebrados no solo sob diferentes formas de manejo utilizando a Regressão Logística. Além disso, apresentar uma revisão bibliográfica sobre a teoria e os conceitos básicas dos modelos de Regressão Logística simples e múltipla, tais como suposições envolvidas, aspectos de inferência e metodologia para seleção de covariáveis.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Importância dos invertebrados edáficos**

O solo é um dos recursos naturais que juntamente com seus organismos contribui de modo decisivo para a manutenção da vida e para o equilíbrio da biosfera. Seu estudo tem sido intensificado nas últimas décadas, sobretudo nos aspectos relacionados aos seus organismos por exercerem significativo papel para a funcionalidade e manutenção dos ecossistemas naturais manejados e principalmente degradados. (SIQUEIRA; MOREIRA, 1996).

Muitas pesquisas envolvendo esses organismos têm sido desenvolvidas com o objetivo de aumentar a capacidade produtiva do solo, reduzir a utilização de insumos químicos e agroquímicos e contribuir para alcançar um padrão de agricultura mais sustentável e menos dependente de insumos.

É necessário aumentar a biodiversidade do solo, vegetação e fauna benéfica, a reciclagem de nutrientes, a produção de biomassa e o incremento de matéria orgânica. Os processos que contribuem para o funcionamento do agroecossistema devem ser favorecidos, tais como a redução dos níveis de resíduos de agrotóxicos; a otimização da decomposição da matéria orgânica e da ciclagem de nutrientes; a melhoria da conservação e da regeneração do solo, da água e da biodiversidade; e o aumento e sustentabilidade da produtividade a longo prazo.

A fauna edáfica tem papel fundamental em processos de decomposição da material orgânica, de produção de húmus, ciclagem de nutrientes e energia, produção de complexos que causam agregação do solo, entre outros, e podem ser alteradas pela simples modificações práticas de manejo do solo. Segundo Hu et al., 1997; Baretta et al., 2003 e Silva et al. (2006) a relação entre a diversidade de animais e os processos que ocorrem no solo permite identificar grupos funcionais da fauna edáfica mais sensíveis ao sistema de manejo.

Muitos invertebrados são bioindicadores da qualidade e da degradação ambiental, devido às várias funções que desempenham na natureza, estreita relação com a heterogeneidade dos ecossistemas e processos ecológicos, bem como seu alto grau de sensibilidade às mudanças ambientais (SILVA, 2010). Assim, sua diversidade pode revelar o nível de qualidade ambiental, a partir do qual podem ser determinadas

intervenções a fim de manter, recuperar ou restaurar a sanidade ambiental, atingindo a sustentabilidade ecológica dos ecossistemas (WINK et al., 2005).

Os invertebrados edáficos se distinguem quanto as suas características e sua funcionalidade. A meso e a macrofauna do solo incluem uma variedade enorme de formas biológicas distintas, de diferentes ordens.

A mesofauna abrange os organismos entre 0,2 a 2,0 mm, que inclui, por exemplo, ácaros, colêmbolos e outros insetos. As atividades tróficas desses animais incluem tanto o consumo de microorganismos e da microfauna, como a fragmentação de material vegetal em decomposição (CORREIA & ANDRADE, 1999).

A macrofauna, que inclui organismos visíveis a olho nu (>2,0 mm), é representada por mais de 20 grupos taxonômicos. Entre eles, cupins, formigas, minhocas, besouros, tatuzinhos, aranhas, centopéias, piolhos-de-cobra, baratas, tesourinhas, grilos, caracóis, escorpiões, percevejos, cigarras, larvas de mosca e de mariposas. (MELO et al., 2009).

A abundância e a diversidade da meso e macrofauna do solo dos ecossistemas naturais e dos agrossistemas podem ser afetadas por vários fatores edáficos (tipo de solo, minerais predominantes, temperatura, pH, matéria orgânica, umidade, textura e estrutura), vegetais (tipo de vegetação e cobertura), históricos (especialmente humano, mas também geológico), topográficos (posição fisiográfica, inclinação) e climáticos (precipitação fluvial, temperatura, vento, umidade relativa do ar). Assim, intervenções antrópicas tanto em sistemas naturais quanto em agrícolas podem potencialmente afetar a dinâmica dessas comunidades animais e, por consequência, as funções ecológicas nas quais estão envolvidas. (MELO et al., 2009).

O conhecimento da comunidade da fauna edáfica pode contribuir para a avaliação do grau de sustentabilidade de uma prática, seja de recuperação de uma área degradada ou até mesmo no caso de um sistema natural interferido. (LINDEN et al., 1994).

## **2.2 Modelo de regressão logística simples**

Os métodos de regressão têm como objetivo descrever as relações entre a variável resposta ( $Y$ ), geralmente binária ou dicotômica e a variável explicativa ( $X$ ). Na regressão logística, a probabilidade de ocorrência de um evento pode ser estimada diretamente e a variável dependente  $Y$  assume apenas dois possíveis valores 1 ou 0,

sendo  $\pi_i = P(Y = 1 | X = x_i)$  a probabilidade de “sucesso” e  $1 - \pi_i = P(Y = 0 | X = x_i)$  a probabilidade de “fracasso”.

Numa série de eventos binários, em que  $(Y_1, Y_2, \dots, Y_n)$  são variáveis aleatórias independentes com distribuição Bernoulli, com probabilidade de sucesso  $(\pi_i)$ , isto é,  $Y_i \sim Ber(\pi_i)$  e denota-se  $x'_i = (1, x_i)$  a  $i$ -ésima linha da matriz  $(X)$  em que  $i = 1, 2, \dots, n$ , pode-se definir a probabilidade de sucesso no modelo de regressão logística simples como:

$$\pi_i = \pi(x_i) = P(Y = 1 | X = x_i) = \frac{\exp(\beta_0 + \beta_1 x_1)}{1 + \exp(\beta_0 + \beta_1 x_1)} \quad (1)$$

e a probabilidade de fracasso:

$$1 - \pi_i = 1 - \pi(x_i) = P(Y = 0 | X = x_i) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1)} \quad (2)$$

em que  $\beta_0$  e  $\beta_1$  são os parâmetros desconhecidos.

No modelo de regressão linear assume-se que a variável resposta é contínua, enquanto que no modelo de regressão logística a variável resposta é discreta: binária ou dicotômica. Em qualquer problema de regressão, a quantidade a ser modelada é o valor médio da variável resposta dado os valores das variáveis independentes. Assim sendo, essa quantidade é chamada de média condicional  $E(Y | X = x_i)$ , onde  $Y$  é a variável resposta e  $x_i$ , os valores das variáveis independentes.

Na regressão linear a equação  $E(Y | X = x_i) = \beta_0 + \beta_1 x_1$  representa a equação de uma reta, onde as constantes  $\beta_0$  e  $\beta_1$  são os parâmetros, sendo que o primeiro representa o valor de  $y$  quando  $x = 0$ , ou seja, o ponto em que a reta corta o eixo das ordenadas (eixo  $y$ ) no gráfico e o segundo representa quanto aumenta/diminui a variável resposta para variação de cada unidade de  $x$ . É importante notar que para qualquer valor de  $x$  dentro do intervalo de  $-\infty$  a  $+\infty$ , sempre existirá um valor esperado de  $Y$ , assim tem-se  $-\infty < E(Y | X = x_i) < +\infty$ .

Na regressão logística, devido a natureza da variável resposta, sua média condicional deve ser maior ou igual a zero e menor ou igual a um, isto é,  $0 \leq E(Y | X = x_i) \leq 1$ , e usando a definição de variável aleatória discreta, tem-se:

$$E(Y | X = x_i) = 1P(Y_i = 1 | X = x_i) + 0P(Y_i = 0 | X = x_i) = \pi_i$$

Quanto a distribuição condicional da variável resposta, no modelo de regressão logístico, o valor da variável resposta dado  $x_i$ , pode ser expressa por  $Y_i = \pi_i + \varepsilon_i$ , em que  $\varepsilon_i$  é chamado de erro, com distribuição Binomial  $(1, \pi_i)$ , com média zero e variância dada por  $\pi_i(1 - \pi_i)$  e pode assumir somente um de dois possíveis valores, isto é,  $\varepsilon_i = 1 - \pi_i$  para  $Y_i = 1$  ou  $\varepsilon_i = -\pi_i$  para  $Y_i = 0$  (HOSMER; LEMESHOW, 1989).

O modelo de regressão logístico binário é um caso particular dos modelos lineares generalizados, mas especificamente dos modelos *logit*, nos quais a variável dependente é associada a uma variável aleatória Bernoulli.

Assim, baseada no modelo *logit*, a forma do modelo de regressão logística é dada como:

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_1)}{1 + \exp(\beta_0 + \beta_1 x_1)} \quad (3)$$

Sendo o modelo acima não linear, aplica-se uma transformação denominada  $g(x)$  para tornar o modelo *logit* linear em seus parâmetros contínuos e fazer com que assumam valores entre  $-\infty$  e  $+\infty$ , dependendo do limite de  $x$ :

$$\begin{aligned} g(x_i) &= \ln \frac{P(\text{ocorrência do evento} | x_i)}{P(\text{não ocorrência do evento} | x_i)} \\ &= \ln \left[ \frac{\pi_i}{1 - \pi_i} \right] = \ln [\exp(\beta_0 + \beta_1 x_1)] = \beta_0 + \beta_1 x_1 \end{aligned} \quad (4)$$

No contexto de modelos lineares generalizados, uma função monótona e derivável que relaciona a média ao preditor linear é denominada de função de ligação,

assim  $g(x_i) = \ln \left( \frac{\pi_i}{1 - \pi_i} \right)$  é a função de ligação canônica para o modelo binomial.

### 2.1.1 Estimação dos Parâmetros

Para a estimação dos parâmetros desconhecidos  $\beta_0$  e  $\beta_1$  no modelo de regressão logística simples usa-se o princípio que orienta uma análise em regressão linear. Na regressão linear o método mais usado para estimação desses parâmetros é o dos mínimos quadrados. Neste método, a escolha de  $\beta_0$  e  $\beta_1$  é dada pelos valores que

minimizam a soma de quadrados dos desvios para os valores observados ( $y_i$ ) em relação ao valor predito ( $\hat{y}_i$ ) baseado no modelo (HOSMER; LEMESHOW, 1989).

No entanto, quando o método dos mínimos quadrados é utilizado em modelo com resultado dicotômico, os estimadores não apresentam as pressuposições básicas do modelo de regressão linear. Para solucionar o problema é utilizado o método da máxima verossimilhança, que produz valores para os parâmetros desconhecidos que maximizam a probabilidade de obtenção dos conjuntos de dados observados.

Segundo Meyer (1978), o método de máxima verossimilhança conduz a estimativas razoáveis para os dados dicotômicos. A função de distribuição de probabilidade  $Y_i$  para o modelo de regressão logístico simples com  $Y_i \sim Ber(\pi_i)$  é dada por:

$$f(y_i, \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

Como as observações são independentes, a função de verossimilhança é obtida pelo produto dos termos dados na equação anterior, ou seja, a função de distribuição de probabilidade conjunta de  $y_1, y_2, \dots, y_n$  será:

$$\prod_{i=1}^n f(y_i, \pi_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, y_i \in [0,1]$$

Então, a função de verossimilhança é dada por:

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, \beta \in \mathbf{R} \quad (5)$$

Aplicando-se logaritmo em ambos os lados da equação, tem-se a expressão:

$$\begin{aligned} l(\beta) &= \ln[L(\beta)] = \ln \left[ \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \right] \\ &= \sum_{i=1}^n \left[ y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right] \end{aligned} \quad (6)$$

Substituindo pelas equações (2) e (4), temos:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n \left[ y_i (\beta_0 + \beta_1 x_i) + \ln \left( \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \right) \right] \\ &= \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_i) - \ln(1 + \exp(\beta_0 + \beta_1 x_i))] \end{aligned} \quad (7)$$

Para determinar os valores de  $\beta$  que maximizam  $l(\beta)$ , deriva-se a função (7) em relação aos elementos do vetor  $\beta$ , que por sua vez são funções dos logaritmos presentes na equação, e assim obtendo duas equações:

$$\frac{\partial l(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left[ y_i - \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \exp(\beta_0 + \beta_1 x_i) \right]$$

$$\frac{\partial l(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left[ y_i x_i - \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \exp(\beta_0 + \beta_1 x_i) x_i \right]$$

e fazendo-se os resultados das equações acima iguais a zero, tem-se o seguinte sistema de equações:

$$\sum_{i=1}^n (y_i - \pi_i) = 0 \quad (8)$$

$$\sum_{i=1}^n x_i (y_i - \pi_i) = 0 \quad (9)$$

Como as equações acima são não lineares em  $\beta_0$  e  $\beta_1$ , para resolvê-las é necessário o uso de um procedimento iterativo, por exemplo, o método de Newton-Raphson, ver seção 3 do Capítulo XII de Casella e Berger (2002), ou algum outro método iterativo disponível em vários programas computacionais.

### 2.1.2 Interpretação dos Coeficientes

Para o estudo em que a variável resposta é dicotômica, a interpretação dos coeficientes se dá pela razão de chance (*odds ratio*), que é a razão das proporções para os dois resultados possíveis, isto é, a razão entre sucesso ( $\pi_i$ ) e fracasso ( $1 - \pi_i$ ).

Considerando a variável independente também dicotômica, em que a variável  $x$  é codificada em 0 e 1, a chance da resposta quando  $x = 1$  é definida como  $\pi(1)/1 - \pi(1)$  e quando  $x = 0$  como  $\pi(0)/1 - \pi(0)$ . O logaritmo da razão de chance é dado por

$$g(1) = \ln \left[ \frac{\pi(1)}{1 - \pi(1)} \right] \quad \text{e} \quad g(0) = \ln \left[ \frac{\pi(0)}{1 - \pi(0)} \right]$$

**Tabela 2.1 – Valores do Modelo de Regressão Logística quando a variável independente é dicotômica.**

Variável resposta Y	Variável independente X	
	x = 1	x = 0
y = 1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
y = 0	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$
<b>Total</b>	<b>1,0</b>	<b>1,0</b>

Fonte: HOSMER & LEMESHOW (1989)

Denotando a razão de chance (*odds ratio*) por  $\Psi$ , temos que:

$$\Psi = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} \quad (10)$$

O logaritmo da razão de chance é:

$$\ln(\Psi) = \ln \left[ \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} \right] = g(1) - g(0)$$

Usando as expressões apresentadas na tabela 1, definimos a razão de chance como:

$$\Psi = \frac{\left[ \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \right] / \left[ \frac{1}{1 + \exp(\beta_0 + \beta_1)} \right]}{\left[ \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \right] / \left[ \frac{1}{1 + \exp(\beta_0)} \right]} = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)$$

e o logaritmo da diferença, ou o logaritmo da razão de chance é:

$$\ln(\Psi) = \ln[\exp(\beta_1)] = \beta_1 \quad (11)$$

Devido a fácil interpretação, a razão de chance é uma medida de associação muito utilizada e possui propriedades estatísticas que a tornam fundamental em muitos tipos de estudos.

### 2.3 Modelo de regressão logística múltipla

Seja um conjunto com  $p$  variáveis independentes, denotadas por  $x'_i = (x_{i0}, x_{i1}, x_{i2}, \dots, x_{ip})$ , o vetor da  $i$ -ésima linha da matriz ( $X$ ) das variáveis

explicativas, em que cada elemento da matriz corresponde ao  $ij$ -ésimo componente  $(x_{ij})$ , em que  $i = 1, 2, \dots, n$  e  $j = 0, 1, \dots, p$ , com  $x_{i0} = 1$ . Denota-se por  $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$ , o vetor de parâmetros desconhecidos e  $\beta_j$  é o  $j$ -ésimo parâmetro associado a variável explicativa  $x_j$ .

No modelo de regressão logística múltipla a probabilidade de sucesso é dada por:

$$\begin{aligned} \pi_i = \pi(x_i) = P(Y = 1 | X = x_i) &= \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})} \\ &= \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} \end{aligned} \quad (12)$$

e a probabilidade de fracasso por:

$$\begin{aligned} 1 - \pi_i = 1 - \pi(x_i) = P(Y = 0 | X = x_i) &= \frac{1}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})} \\ &= \frac{1}{1 + \exp(x_i' \beta)} \end{aligned} \quad (13)$$

Assume-se que  $Y_i$  tem uma distribuição de Bernoulli com parâmetro de sucesso  $\pi_i$  e que o “logit” para o modelo de regressão logística múltipla é dada pela equação:

$$g(x_i) = \ln \left[ \frac{\pi_i}{1 - \pi_i} \right] = x_i' \beta = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (14)$$

### 2.2.1 Estimação dos Parâmetros

Para estimar os parâmetros deve-se utilizar o método da máxima verossilhança, similar ao caso da regressão logística simples.

A função de verossimilhança, neste caso, é idêntica a expressão (5), considerando que  $\pi(\cdot)$  é dada pela expressão (12). Especificamente, para uma amostra de tamanho  $n$ , tem-se que

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1 - y_i}, \text{ com } y_i \in \{0, 1\} \quad (15)$$

Seja o vetor  $\beta$  de parâmetros relacionados com a probabilidade condicional  $P(Y_i = 1 | x_i) = \pi(x_i)$  para  $i \in \{1, \dots, n\}$ . Então, o estimador de  $\beta$ , pelo método da máxima verossimilhança, denotado por  $\hat{\beta}$ , é a solução das equações de verossimilhança

$$\begin{aligned} \sum_{i=1}^n (y_i - \pi_i) &= 0 \\ \sum_{i=1}^n x_{ij} (y_i - \pi_i) &= 0 \quad \text{para } j \in \{1, \dots, p\} \end{aligned} \quad (16)$$

No entanto, agora tem-se  $p + 1$  equações de verossimilhança que são obtidas ao se diferenciar a função logaritmo de verossimilhança dada por

$$L(\beta) = \sum_{i=1}^n [y_i \ln \pi_i + (1 - y_i) \ln(1 - \pi_i)] \quad (17)$$

com respeito a cada um dos  $p + 1$  coeficientes. A expressão (17) é obtida a partir do logaritmo de função (15) e do uso das propriedades de somatório e de logaritmos.

As expressões das equações normais são apresentadas abaixo,

$$\begin{aligned} \frac{\partial L}{\partial \beta_0} &= \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\pi}_i = 0 \\ \frac{\partial L}{\partial \beta_j} &= \sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n x_{ij} \hat{\pi}_i = 0, \quad \text{para } j \in \{1, \dots, p\} \end{aligned} \quad (18)$$

onde  $\hat{\pi}_i$  indica o estimador pelo método da máxima verossimilhança de  $\pi_i$ .

Assim, pode-se representar todas as  $p + 1$  equações de verossimilhança, em notação matricial, como

$$\frac{\partial L(\beta)}{\partial \beta} X'(Y - \Pi) = 0 \quad (19)$$

em que,

$$\begin{aligned} Y &= (y_1, \dots, y_n)'_{1 \times n} \\ \Pi &= (\pi_1, \dots, \pi_n)'_{1 \times n} \\ \beta' &= (\beta_0, \dots, \beta_p)'_{1 \times (p+1)} \\ X &= \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}_{n \times (p+1)} \end{aligned}$$

$$\Sigma = \begin{bmatrix} \pi_1(1-\pi_1) & 0 & \cdots & 0 \\ 0 & \pi_2(1-\pi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_n(1-\pi_n) \end{bmatrix}_{n \times n}$$

onde  $Y$  e  $\Pi$  são matrizes  $n \times 1$ ,  $\beta$  é um vetor  $1 \times (p+1)$ ,  $X$  é uma matriz  $n \times (p+1)$  e  $\Sigma$  é uma matriz de variâncias e covariâncias  $n \times n$ .

Como no modelo univariado, as equações obtidas com a derivação da função de verossimilhança não são lineares, aqui também são necessários métodos iterativos para a resolução do sistema de equações resultante. Segundo Walker e Duncan (1967) citado por Martins (1988), a exata determinação dos parâmetros não é possível, em função da conhecida complexidade do problema resultante. Entretanto, o método iterativo de Newton-Raphson, apresenta a vantagem de convergir rapidamente para a solução.

Nesse método, o primeiro passo requer o uso de uma solução inicial para os valores que maximizam a função de verossimilhança. A função é aproximada, em uma vizinhança da solução inicial por um polinômio de segundo grau. A segunda solução alcançada, no processo iterativo, é o ponto de máximo valor do polinômio, e assim por diante. Dessa forma, o método gera uma sequência de soluções que convergem para o ponto de máximo da função de verossimilhança (FIGUEIRA, 2006).

O método de Newton-Raphson é apresentado de forma detalhada em Casella e Berger (2002).

### 2.3 Variáveis *dummy* no modelo

Quando as variáveis independentes categóricas possuem mais de duas categorias é necessário recodificá-las antes de incluí-las no modelo.

Admitindo que seja  $p$  o número de variáveis independentes e se a  $j$ -ésima variável independente da equação  $x_j$  assume  $k_j$  níveis, são feitas  $k_j - 1$  variáveis indicadoras (*dummy*) para representá-la. Será denotado como  $D_{ju}$  as variáveis *dummy* e os coeficientes dessas variáveis como  $\beta_{ju}$ , onde  $u = 1, 2, \dots, k_j - 1$ . Assim, a equação da transformação logarítmica assume a seguinte forma:

$$g(x_i) = \beta_0 + \beta_1 x_1 + \dots + \sum_{u=1}^{k_j-1} \beta_{ju} D_{ju} + \beta_p x_{p_1}$$

É importante ressaltar que sempre que um nível de uma variável independente categórica for incluído no modelo, todos os outros níveis dessa variável devem ser incluídos, o mesmo raciocínio vale também para a exclusão de algum nível, se um for excluído, todos os outros devem ser excluídos. Se recodificarmos uma variável em  $k$  níveis de uma variável categórica, a contribuição para os graus de liberdade será  $k - 1$  para o teste da razão de verossimilhança, na análise de inclusão ou não desta variável.

## 2.4 Teste de significância dos coeficientes

Após o ajuste do modelo (estimação dos parâmetros  $\beta_i$ 's) deve-se testar a significância das variáveis decorrentes no modelo. Nesse processo está envolvido o teste de hipóteses estatísticas, o qual determina se as variáveis independentes no modelo estão “significativamente” relacionadas com a variável resposta.

Os testes mais utilizados para testar a qualidade do modelo ajustado e a significância individual dos parâmetros ou de um conjunto de parâmetros do modelo são o teste da razão de verossimilhança, o teste de Wald e o teste de Escore. Neste estudo, porém, serão abordadas apenas as estatísticas para o teste da razão de verossimilhança e o teste de Wald.

### 2.4.1 Teste da Razão de Verossimilhança

Para testar a significância do coeficiente de uma variável no modelo comparam-se os valores observados da variável resposta com os valores preditos, por cada um dos dois modelos; o primeiro com a variável presente e o segundo sem essa variável. Para a comparação entre os valores preditos e observados, usando a função de verossimilhança, usa-se “menos duas vezes o logaritmo do quociente desses máximos”, pois sua distribuição equivale ao *Qui-Quadrado* com  $n - p$  graus de liberdade, e é baseada na seguinte expressão:

$$D = -2 \ln \left[ \frac{\text{verossimilhança do modelo atual}}{\text{verossimilhança do modelo saturado}} \right]$$

Essa estatística é chamada de *deviance* (desvio) e avalia o valor ajustado na regressão logística, desempenhando o mesmo papel que a soma de quadrados residuais tem na regressão linear (NELDER; WEDDERBURN, 1972).

Considerando o modelo com as proporções estimadas  $\hat{\pi}_i$ , a *deviance* pode ser escrita como:

$$D = \sum_{i=1}^n \left[ y_i \ln \left( \frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left( \frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (20)$$

A *deviance* sempre é positiva e quanto menor, melhor é o ajuste do modelo.

Para estimar a significância de uma variável independente, comparam-se o valor de  $D$  com e sem a variável independente na equação. A alteração no valor de  $D$  esperada pela inclusão da variável independente no modelo é obtida através de:

$$G = D(\text{para o modelo sem a variável}) - D(\text{para o modelo com a variável})$$

Também é comum expressar a estatística  $G$  por:

$$G = -2 \ln \left[ \frac{(\text{verossimilhança sem a variável})}{(\text{verossimilhança com a variável})} \right]$$

No caso de uma única variável independente, verifica-se facilmente que se esta variável não está no modelo, o estimador de máxima verossimilhança de  $\beta_0$  é

$$\hat{\beta}_0 = \ln \left( \frac{n_1}{n_0} \right) \text{ onde } n_1 = \sum_{i=1}^n y_i \text{ e } n_0 = \sum_{i=1}^n (1 - y_i) \text{ e que o valor predito é constante, } \frac{n_1}{n}.$$

Assim, o valor de  $G$  pode ser escrito como:

$$G = -2 \ln \left[ \frac{\left( \frac{n_1}{n} \right)^{n_1} \left( \frac{n_0}{n} \right)^{n_0}}{\prod \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \right] \quad (21)$$

Para o caso em que analisamos se pelo menos uma das variáveis explicativas são significativas no modelo em estudo, temos as seguintes hipóteses:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_1 : \text{pelo menos um } \beta_i \neq 0 \end{cases}$$

Sob a hipótese de que pelo menos um  $\beta$  é igual a zero, a estatística  $G$  terá uma distribuição assintótica qui-quadrado ( $\chi^2$ ) com grau de liberdade igual à diferença do número de parâmetros dos modelos que estão sendo comparados. Compara-se a estatística  $G$  com o valor de ( $\chi^2$ ), a um nível de significância pré-estabelecido e conclui se as variáveis podem, ou não, ser retiradas do modelo.

## 2.4.2 Teste de Wald

Esta estatística é baseada na distribuição assintótica de  $\hat{\beta}$  e é uma generalização do teste  $t$  de Student (Wald, 1943) citado por Colosimo e Giolo (2006). É, geralmente, mais usada no caso de hipóteses relativas a um único parâmetro  $\beta_j$ . Assim, as hipóteses são as seguintes:

$$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

E a estatística do teste é dada pela seguinte expressão:

$$w = \hat{\beta}' I(\hat{\beta}) \hat{\beta} \quad (22)$$

em que  $I(\hat{\beta})$  é a matriz de informação de Fischer avaliada em  $\hat{\beta}$ , em que sob  $H_0$ , a estatística  $w$  apresenta uma distribuição qui-quadrado com número de graus de liberdade igual ao número de parâmetros.

No caso em que  $\beta$  é um escalar, a expressão (22) se reduz a:

$$w = \frac{\hat{\beta}^2}{\hat{SE}(\hat{\beta})^2}$$

Porém, o teste de Wald em alguns casos, costuma não rejeitar a hipótese nula quando esta deveria ser rejeitada. Hauck e Donner (1977) e Jennings (1986) examinando a eficiência do Teste de Wald recomendam que o teste da razão de verossimilhanças seja usado quando há dúvidas de que o teste de Wald tenha falhado.

### 3 MATERIAL E MÉTODOS

#### 3.1 Área de estudo/Procedência dos dados

Os dados utilizados nesse trabalho foram gentilmente cedidos por Joedna Silva, doutora em Agronomia/Fitotecnia pela Universidade Federal do Ceará e são provenientes de um estudo realizado no Município de Araponga, Zona da Mata de Minas Gerais, dentro da microrregião de Viçosa.

A coleta de dados foi realizada em quatro propriedades rurais de agricultores familiares e comerciais desse município, onde foram coletadas amostras de solo em diferentes agroecossistemas e sistemas naturais, sendo assim descritos:

- Sistemas convencionais (SC) de café (*Coffea arabica* L.) – cultivo de café solteiro a pleno sol com uso de fertilizantes e agrotóxicos.
- Sistemas de manejo agroecológico (AGRO) - cultivo de café solteiro a pleno sol com o surgimento e manutenção de vegetação espontânea, sem uso de agrotóxicos.
- Sistemas agroflorestais (SAF) - cultivo de café consorciado com árvores frutíferas ou não, com o surgimento e manutenção de vegetação espontânea, sem uso de agrotóxicos.
- Mata nativa próximas as propriedades rurais.

A coleta do solo foi efetuada na camada de 0 – 10 cm, realizada em dois períodos do ano: – seco (Junho – Setembro) e chuvoso (Dezembro – Março), com 60 amostras em cada período (15 amostras de cada sistema).

As amostras foram selecionadas em áreas demarcadas aleatoriamente em pontos distanciados entre dez e quinze metros um do outro. Para a avaliação da macro e mesofauna edáfica, foram coletados blocos de solo de 20 x 20 x 10 cm, sendo coletada, primeiramente, a serapilheira da superfície, considerando todo resíduo vegetal sobre a superfície do solo.

Foram extraídos os invertebrados edáficos para mensuração do número total de indivíduos por amostra e foram realizadas as análises físicas e químicas do solo e da serapilheira.

### 3.2 Construção do modelo

Buscou-se construir um modelo para determinar a quantidade média de indivíduos (invertebrados edáficos) por  $m^2$  que se poderia encontrar no solo sob diferentes condições (época seca e época chuvosa) e, além disso, analisar os fatores que influenciariam esse resultado. Para determinar os fatores ambientais responsáveis pela ocorrência de mais ou menos indivíduos por  $m^2$  no solo, foi utilizada a análise de regressão logística.

A variável resposta (variável dependente) analisada foi denominada como o “*número médio de indivíduos por  $m^2$  encontrados no solo*”. Foi usada a seguinte codificação para tornar a variável resposta dicotômica:

Em época seca:

- 0 para representar a presença de 9 ou menos indivíduos por  $m^2$  no solo
- 1 para representar a presença de mais de 9 indivíduos por  $m^2$  no solo

Em época chuvosa:

- 0 para representar a presença de 101 ou menos indivíduos por  $m^2$  no solo
- 1 para representar a presença de mais de 101 indivíduos por  $m^2$  no solo

Os valores 9 e 101 referem-se ao valor médio de indivíduos encontrados por  $m^2$ , nas amostras coletadas, em épocas diferentes e, serviram como valor de referência para categorização da variável resposta.

As covariáveis utilizadas na análise são apresentadas na Tabela 3.2 (ver matrizes de dados completas no apêndice A e B). Dentre elas, a covariável “*Tipo de Sistema*” foi recodificada antes de ser incluída no modelo por ser uma covariável categórica. Assim, a Tabela 3.1 ilustra a codificação através de covariáveis *dummy*.

**Tabela 3.1 - Codificação de covariáveis *dummy* utilizadas na análise realizada no solo e em serapilheira em período seco e chuvoso**

Tipo de Sistema	Variáveis <i>Dummy</i>		
	Vd1	Vd2	Vd3
Convencional	0	0	0
Mata	1	0	0
Agroecológico	0	1	0
Agro florestal	0	0	1

**Tabela 3.2 – Código, descrição e tipo de variáveis utilizadas na análise realizada no solo e em serapilheira em período seco e chuvoso**

<b>Código</b>	<b>Descrição</b>	<b>Tipo</b>
Vd1	<i>Dummy</i> sistema 1 (0: Convencional; 1: Mata)	Categórica
Vd2	<i>Dummy</i> sistema 2 (0: Convencional; 1: Agroecológico)	Categórica
Vd3	<i>Dummy</i> sistema 3 (0: Convencional; 1: Agroflorestal)	Categórica
V4	Nitrogênio total em serapilheira	Contínua
V5	Fósforo em serapilheira	Contínua
V6	Potássio em serapilheira	Contínua
V7	Cálcio em serapilheira	Contínua
V8	Magnésio em serapilheira	Contínua
V9	Manganês em serapilheira	Contínua
V10	Zinco em serapilheira	Contínua
V11	Ferro em serapilheira	Contínua
V12	Carbono orgânico total na serapilheira	Contínua
V13	Relação carbono/nitrogênio na serapilheira	Contínua
V14	Matéria orgânica da serapilheira	Contínua
V15	Porcentagem de umidade do solo	Contínua
V16	Temperatura ambiente	Contínua
V17	Temperatura do solo	Contínua
V18	Peso seco da serapilheira em t/há	Contínua
V19	Peso seco da serapilheira em g/kg	Contínua
V20	Microporosidade do solo	Contínua
V21	Macroporosidade do solo	Contínua
V22	Porosidade total do solo	Contínua
V23	Densidade do solo	Contínua
V24	Ph do solo	Contínua
V25	Fósforo no solo	Contínua
V26	Potássio no solo	Contínua
V27	Cálcio no solo	Contínua
V28	Magnésio no solo	Contínua
V29	Alumínio no solo	Contínua
V30	Soma de base	Contínua
V31	CTC (capacidade de troca de cátions do solo) efetiva	Contínua
V32	CTC (capacidade de troca de cátions do solo) total	Contínua
V33	Saturação de bases do solo	Contínua
V34	Saturação por alumínio	Contínua
V35	Matéria orgânica do solo	Contínua
V36	Fósforo remanescente do solo	Contínua

Para a seleção das covariáveis foi utilizado o método derivado da proposta de Collett executado com o auxílio do pacote estatístico R (*R Development Core Team*), versão 2.11.1. Para testar a qualidade do ajuste do modelo foram utilizados o teste Hosmer e Lemeshow e o cálculo da área sob a Curva ROC, através do programa estatístico SPSS (*Statistical Package for the Social Sciences*), versão 15.0 (2006) para Windows.

### 3.3 Seleção de Covariáveis

A aplicação da análise de regressão logística envolve selecionar o conjunto de covariáveis a ser usado no modelo e, é necessário que esse modelo forneça resultados satisfatórios, com boa estimativa de previsão e com o menor número de covariáveis possíveis que descreva o fenômeno estudado (segundo o Princípio da Parcimônia). O sucesso para a modelagem de um conjunto de dados está relacionado à área específica, aos métodos estatísticos e à experiência e bom senso do pesquisador (HOSMER e LEMESHOW, 1989).

Para obter um modelo com o menor número de covariáveis possível, mas que mantenha a eficiência nos resultados de previsão é necessário que se tenha um plano de escolha das covariáveis iniciais que serão testadas no modelo e um método que auxilie na seleção e adequação dessas covariáveis (HOSMER e LEMESHOW, 1989).

Alguns métodos como *forward*, *backward* ou *stepwise* são muito utilizados para seleção de covariáveis, cujos algoritmos foram implementados em programas computacionais. No entanto, apresentam algumas desvantagens, pois tendem a identificar um particular conjunto de covariáveis, em vez de possíveis conjuntos igualmente bons para explicar a resposta, impossibilitando o pesquisador escolher o mais relevante em sua área de aplicação (COLOSIMO; GIOLO, 2006).

Diante disso, nesse estudo optou-se por utilizar uma estratégia de seleção de covariáveis derivada da proposta de Collett (1994), citado por Colosimo e Giolo (2006), em que as informações do pesquisador podem ser incluídas no processo de decisão, o que envolve uma participação mais ativa do estatístico e pesquisador em cada passo do processo de seleção, podendo, por exemplo, incluir covariáveis relevantes no estudo independente de significância estatística.

Os passos utilizados no processo de seleção são descritos como se segue:

1. Primeiramente ajustar todos os modelos contendo uma única covariável. Em seguida, incluir todas as covariáveis significativas ao nível de 0,10. Nesse passo, utilizar o teste da razão de verossimilhanças.
2. Ajustar conjuntamente as covariáveis significativas no passo 1. Em seguida ajustar modelos reduzidos, excluindo uma única covariável de cada vez, pois na presença de certas covariáveis, outras podem deixar de ser significativas. Verificar quais as covariáveis que provocaram um aumento significativo na

estatística da razão de verossimilhanças. Somente aquelas que atingiram a significância devem permanecer no modelo.

3. Com as covariáveis que ficaram retidas no passo 2, ajustar um novo modelo e as covariáveis que foram excluídas no passo 2 retornaram ao modelo para confirmar se não são estatisticamente significativas.
4. Incluir ao modelo as eventuais covariáveis significativas no passo 3 juntamente com aquelas do passo 2. Neste passo retornam-se com as covariáveis excluídas no passo 1 para confirmar se elas não são estatisticamente significativas.
5. Ajustar um modelo incluindo as covariáveis significativas no passo 4 e testar se alguma delas pode ser retirada do modelo.
6. Com as covariáveis que “sobreviveram” ao passo 5, ajusta-se então o modelo final para os efeitos principais. Deve-se verificar a possibilidade de inclusão de termos de interação dupla entre as covariáveis incluídas no modelo. O modelo final será composto pelos efeitos principais identificados no passo 5 e os possíveis termos de interação significativos nesse passo.

Segundo Colosimo e Giolo (2005), ao ser utilizado este procedimento de seleção, deve-se evitar ser muito rigoroso ao testar cada nível individual de significância. O nível de significância não deve ser muito baixo para decidir se um termo deve ser incluído no modelo, sendo recomendado um valor próximo de 0,10.

### **3.4 Medidas de qualidade do ajuste**

Para saber se o modelo selecionado explica razoavelmente bem o comportamento da variável resposta deve-se testar a qualidade do ajuste do modelo, que é dada por estatísticas que medem a proporção da variação na variável resposta que é explicada pelo modelo.

#### **3.4.1 Teste de Hosmer e Lemeshow**

O teste Hosmer e Lemeshow corresponde a um teste Qui-quadrado que consiste em dividir o número de observações em dez classes e, em seguida, comparar as frequências preditas com as observadas. A finalidade desse teste é verificar se existem diferenças significativas entre as classificações realizadas pelo modelo e a realidade observada.

Assim, a hipótese a testar é:

$$\begin{cases} H_0 : o_j = \mathbf{e}_j, & \forall_j = 1, \dots, g \\ H_1 : \exists_j | o_j \neq \mathbf{e}_j, & j = 1, \dots, g \end{cases}$$

A estatística de teste sob a hipótese nula é a seguinte:

$$\chi_{HL}^2 = \sum_{j=1}^g \frac{(o_j - \mathbf{e}_j)^2}{\mathbf{e}_j \left(1 - \frac{\mathbf{e}_j}{n_j}\right)} = \sum_{j=1}^g \frac{(o_j - \bar{p}_j)^2}{n_j \bar{p}_j (1 - \bar{p}_j)} \sim \chi_{g-2}^2$$

em que

$n_j$  é o número de observações pertencentes ao grupo  $j$ , verificando-se  $n = \sum_{j=1}^g n_j$

$o_j$  é a frequência observada de sucesso no grupo  $j$ , onde  $o_j = \sum_{i=1}^{n_j} y_{ij}$  e  $y_{ij}$  é a  $i$ -ésima observação do grupo  $j$ .

$e_j$  é a frequência esperada de sucesso no grupo  $j$ , onde  $e_j = n_j \bar{p}_j$  e  $\bar{p}_j = \frac{\sum_{i=1}^{n_j} \hat{p}_{ji}}{n_j}$

$\hat{p}_j$  é a probabilidade predita correspondente à  $i$ -ésima observação do grupo  $j$ .

Antes do cálculo da estatística teste, é necessário estimar a probabilidade de sucesso para cada observação e ordenar as probabilidades preditas por ordem crescente, em seguida agrupar os dados de acordo com os decis das probabilidades preditas. Em cada decil, dividir os valores observados e os valores preditos para o sucesso e não sucesso.

A um nível de significância estabelecido, busca-se não rejeitar a hipótese de que não existem diferenças entre os valores preditos e observados. O critério de avaliação se distingue um pouco do convencional, pois geralmente o que se pretende é rejeitar a hipótese nula. Nesse caso, se houver diferenças significativas entre as classificações preditas pelo modelo e as observadas, então o modelo não representa a realidade de forma satisfatória. Em tal situação, o modelo não seria capaz de produzir estimativas e classificações muito confiáveis (HOSMER e LEMESHOW, 1989).

A estatística do teste de Hosmer e Lemeshow tem distribuição qui-quadrado com  $g - 2$  graus de liberdade, em que  $g = 10$  grupos.

### 3.4.2 Área sob a Curva ROC

No modelo logístico, como a variável resposta é binária é necessário escolher uma regra de predição. A probabilidade da resposta predita pode conseqüentemente formar a base para se classificar um indivíduo de acordo com um dos dois grupos ( $\hat{Y} = 0$  ou  $\hat{Y} = 1$ ), já que  $\hat{\pi}$  está entre 0 e 1. É preciso estabelecer um ponto de corte ( $\pi_c$ ) em que valores acima dele o indivíduo é classificado como evento ( $\hat{Y}_i = 1$ ) e valores abaixo dele o indivíduo é classificado como não evento ( $\hat{Y}_i = 0$ ). Geralmente  $\pi_c = 0,5$  é um valor razoável, entretanto, se os dois grupos não podem ser classificados como simétricos, um valor diferente de 0,5 deve ser considerado. Uma maneira de se determinar o ponto de corte é através da curva ROC (*Receiver Operating Characteristic*), a qual permite avaliar a capacidade preditiva de um modelo usando o ponto de corte escolhido (ABREU, 2004).

A avaliação da capacidade preditiva de um modelo, depois que foi feita a classificação das observações em um dos dois grupos ( $\hat{Y} = 0$  ou  $\hat{Y} = 1$ ) é baseada nos conceitos de sensibilidade e especificidade estatística, obtidas a partir da construção de matrizes de confusão, gerada pelo modelo. A sensibilidade é definida como a probabilidade de um indivíduo ser classificado como zero, dado que realmente é zero e a especificidade é a probabilidade de um indivíduo ser classificado como um, dado que realmente é um.

A matriz de confusão é uma tabela 2 x 2 para o ponto de corte escolhido, como apresentado na Tabela 3.3. Neste trabalho, a sensibilidade e a especificidade são utilizadas com uma nomenclatura diferente da usual, onde a sensibilidade será denominada como capacidade de acerto dos zeros e a especificidade como capacidade de acerto dos uns.

**Tabela 3.3 – Matriz de Confusão para duas classes**

		Valores Observados		
		0	1	
Valores Preditos	0	$a_0$	$a_1$	$a$
	1	$b_0$	$b_1$	$b$
		A	B	$n$

Assim, estas medidas são definidas como:

$$\text{Capacidade de acerto dos zeros (CAZ)} = \frac{a_0}{A} \text{ (Sensibilidade)}$$

$$\text{Capacidade de acerto dos uns (CAU)} = \frac{b_1}{B} \text{ (Especificidade)}$$

$$\text{Capacidade de acerto total (CAT)} = \frac{a_0 + b_1}{n} \text{ (Acurácia)}$$

em que

$n$  é o número total de observações na amostra;

$a_0$  é o número de zeros que foram classificados como zero (acerto);

$b_1$  é o número de uns que foram classificados como um (acerto);

$a_1$  é o número de uns que foram classificados como zero (erro);

$b_0$  é o número de zeros que foram classificados como um (erro);

$A$  é o número de zeros na amostra;

$B$  é o número de uns na amostra;

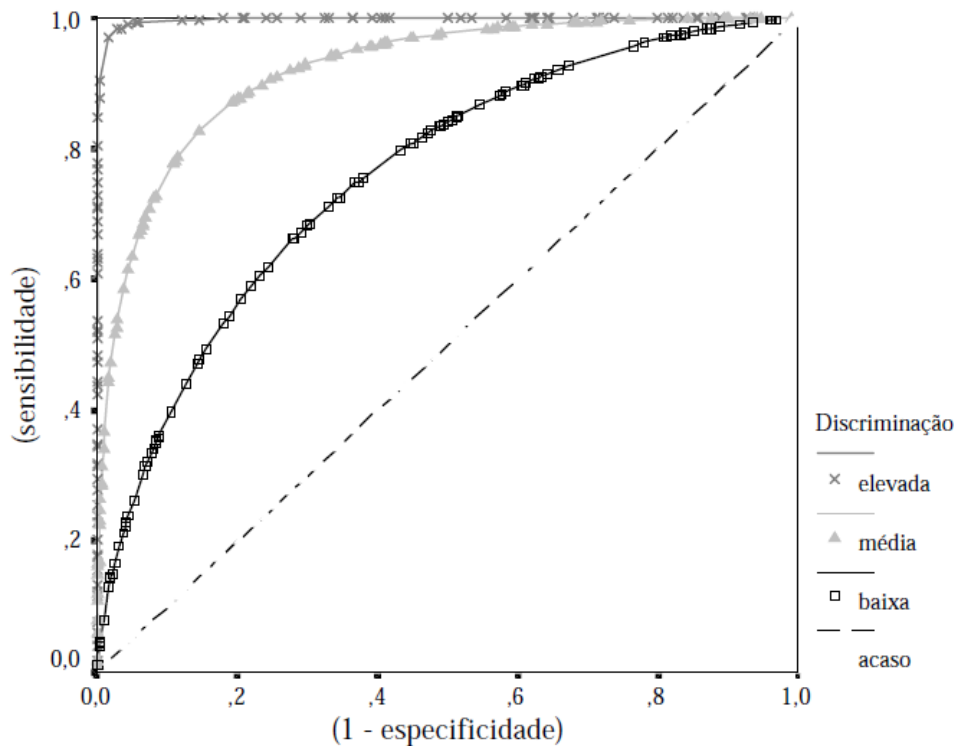
$a$  é o número total de observações classificadas como zero na amostra;

$b$  é o número total de observações classificadas como um na amostra.

Essa matriz (Tabela 2) descreve uma tabulação cruzada entre a classificação predita, através de um único ponto de corte e a condição real e conhecida de cada indivíduo, em que a diagonal principal corresponde às classificações corretas e os valores fora dessa diagonal representam os erros de classificação (BROCCO, 2006).

Para a construção da Curva ROC, são calculadas a sensibilidade e a especificidade para todas as observações da amostra, considerando diferentes pontos de corte do modelo. A curva é obtida registrando em um gráfico “sensibilidade” x “1 – especificidade” para os diversos pontos de corte.

Em geral, o melhor ponto de corte produz valores para a sensibilidade e a especificidade que se localizam no ponto mais à esquerda e superior possível, isto é, no “ombro” da curva, ou próximo dele.



**Figura 3.1 – Curvas ROC dos modelos ajustados para o conjunto de dados qualquer**

A linha diagonal indica uma classificação aleatória, ou seja, um modelo que aleatoriamente seleciona saídas como zero ou um. Uma curva perfeita corresponderia a uma linha horizontal no topo do gráfico, com elevada capacidade de discriminação, porém isto dificilmente será alcançado. As curvas que se apresentam entre a linha diagonal e a linha perfeita são consideradas de média e baixa capacidade de discriminação, assim quanto mais a curva estiver distante da diagonal principal, melhor o desempenho de modelo associado a ela. Esse fato sugere que quanto maior for a área entre a curva ROC produzida e a diagonal principal, melhor o desempenho global do modelo.

Alguns métodos, como o da estimação de máxima verossimilhança e o da aproximação da estatística  $U$  de Wilcoxon-Mann-Whitney, são usados para calcular a área abaixo de uma curva ROC, porém não serão demonstrados nesse trabalho. Para maiores informações, consultar Hanley e McNeil (1983).

A área delimitada pela curva mede a capacidade de discriminação do modelo. Hosmer e Lemeshow (2000) apresentaram uma regra geral para avaliação do resultado da área sob a Curva ROC:

- Se a área for igual a 0,5: não há discriminação;

- Área no intervalo entre 0,7 e 0,8: discriminação aceitável;
- Área no intervalo entre 0,8 e 0,9: excelente discriminação;
- Área acima de 0,9: excepcional discriminação.

Na comparação de modelos, o melhor modelo é aquele que apresentar a maior área sob a curva ROC.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Conjunto de dados da serapilheira e do solo no período seco – atributos físicos e químicos

No processo de seleção das covariáveis para a construção do modelo de regressão logística foi utilizada a estratégia de seleção de modelos derivada da proposta de Collett (1994), que envolve uma participação mais ativa do pesquisador em cada passo do processo de seleção. Os resultados são apresentados na Tabela 4.1.

No passo 1 foram ajustados todos os modelos contendo uma única covariável e pelo teste da razão de verossimilhanças foi observado que as covariáveis Vd1, V5, V7, V12, V14, V17, V22, V23, V29, V32, V34 e V35 foram significativas ao nível de 0,10, ou seja, mostraram ter alguma influência sobre a característica avaliada, que nesse caso, refere-se a ocorrência de mais de 9 indivíduos por m<sup>2</sup>, em média, no solo das áreas estudadas.

As covariáveis significativas anteriormente foram então ajustadas conjuntamente (passo 2), juntamente com as covariáveis Vd2 e Vd3 que foram incluídas no modelo mesmo não apresentando significância estatística, porque sempre que um nível de uma variável indicadora (*dummy*) for incluído no modelo, todos os outros níveis dessa variável devem ser incluídos. Neste passo ainda, foram também ajustados modelos reduzidos, excluindo uma única covariável de cada vez. Verificou-se que apenas as covariáveis Vd1, V5 e V7 provocam um aumento significativo na estatística da razão de verossimilhanças. Assim, no passo 3, ajustou-se um novo modelo com as covariáveis Vd1, Vd2, Vd3, V5 e V7 e outros modelos que incluíam também as covariáveis que foram excluídas no passo 2 para confirmar realmente se não eram estatisticamente significativas.

No passo 4, as covariáveis significativas no passo 3 (V22, V23 e V35) foram incluídas ao modelo juntamente com aquelas do passo 2. Neste passo retornou-se com as covariáveis excluídas no passo 1 para confirmar se elas não eram estatisticamente significativas. Ajustou-se então (passo 5), um modelo incluindo as covariáveis V6, V11, V33 e V36 significativas no passo 4 e foi testado se alguma delas poderia ser retirada do modelo. Foi verificado que as covariáveis V5, V11, V22, V23, V33 e V36 não apresentaram significância estatística nesse passo e foram retiradas do modelo.

No passo 6, com as covariáveis selecionadas ajustou-se então o modelo para as covariáveis que sobreviveram no passo 5. Para completar a modelagem foi verificada a possibilidade de inclusão de termos de interação dupla entre as covariáveis já incluídas no modelo. As interações Vd2\*V6 e V7\*V6 foram significativas ao nível de 0,10. Na etapa final chegou-se a três modelos que não apresentaram muita discrepância nos valores da estatística do Teste da Razão da Verossimilhança.

**Tabela 4.1 – Seleção de covariáveis usando o modelo de regressão logística para o conjunto de dados da serapilheira e do solo no período seco**

Passos	Modelo	-2 log L ( $\beta$ )	Estatística de teste TRV	Valor <i>p</i>
Passo 1	Nulo	78,859	-	-
	Vd1	71,288	7,571	0,006*
	Vd2	78,764	0,095	0,758
	Vd3	76,302	2,557	0,110
	V4	78,456	0,403	0,525
	V5	74,881	3,978	0,046*
	V6	78,140	0,719	0,397
	V7	73,462	5,397	0,020*
	V8	77,032	1,827	0,176
	V9	77,816	1,043	0,307
	V10	78,406	0,453	0,501
	V11	77,513	1,346	0,246
	V12	74,946	3,913	0,048*
	V13	78,189	0,670	0,413
	V14	74,661	4,198	0,040*
	V15	77,421	1,438	0,230
	V16	76,926	1,933	0,164
	V17	75,084	3,775	0,052*
	V18	77,268	1,591	0,207
	V19	77,260	1,599	0,206
	V20	78,461	0,398	0,528
	V21	76,887	1,972	0,160
	V22	75,633	3,226	0,072*
	V23	75,476	3,383	0,066*
	V24	77,020	1,839	0,175
	V25	78,832	0,027	0,870
	V26	77,594	1,265	0,261
	V27	78,508	0,351	0,553
	V28	78,465	0,394	0,530
	V29	72,611	6,248	0,012*
	V30	78,579	0,280	0,596
	V31	78,784	0,075	0,784
	V32	75,399	3,460	0,063*
	V33	77,648	1,211	0,271
	V34	72,151	6,708	0,010*
	V35	75,468	3,391	0,066*
	V36	78,703	0,156	0,693

...continuação

Passo 2	Vd1+Vd2+Vd3+V5+V7+V12+V14+V17+V22+V23+V29+V32+V34+V35	46,423	-	-
	Vd2+Vd3+V5+V7+V12+V14+V17+V22+V23+V29+V32+V34+V35	49,515	3,092	0,079*
	Vd1+Vd3+V5+V7+V12+V14+V17+V22+V23+V29+V32+V34+V35	48,126	1,703	0,192
	Vd1+Vd2+V5+V7+V12+V14+V17+V22+V23+V29+V32+V34+V35	46,725	0,302	0,583
	Vd1+Vd2+Vd3+V7+V12+V14+V17+V22+V23+V29+V32+V34+V35	49,873	3,450	0,063*
	Vd1+Vd2+Vd3+V5+V12+V14+V17+V22+V23+V29+V32+V34+V35	50,509	4,086	0,043*
	Vd1+Vd2+Vd3+V5+V7+V14+V17+V22+V23+V29+V32+V34+V35	47,978	1,555	0,212
	Vd1+Vd2+Vd3+V5+V7+V12+V17+V22+V23+V29+V32+V34+V35	48,176	1,753	0,186
	Vd1+Vd2+Vd3+V5+V7+V12+V14+V22+V23+V29+V32+V34+V35	46,763	0,340	0,560
	Vd1+Vd2+Vd3+V5+V7+V12+V14+V17+V23+V29+V32+V34+V35	47,130	0,707	0,400
	Vd1+Vd2+Vd3+V5+V7+V12+V14+V17+V22+V29+V32+V34+V35	46,495	0,072	0,788
	Vd1+Vd2+Vd3+V5+V7+V12+V14+V17+V22+V23+V32+V34+V35	46,468	0,045	0,832
	Vd1+Vd2+Vd3+V5+V7+V12+V14+V17+V22+V23+V29+V34+V35	46,610	0,187	0,665
	Vd1+Vd2+Vd3+V5+V7+V12+V14+V17+V22+V23+V29+V32+V35	46,630	0,207	0,649
	Vd1+Vd2+Vd3+V5+V7+V12+V14+V17+V22+V23+V29+V32+V34	47,542	1,119	0,290
	Passo 3	Vd1+Vd2+Vd3+V5+V7	55,479	-
Vd1+Vd2+Vd3+V5+V7+V12		54,790	0,689	0,407
Vd1+Vd2+Vd3+V5+V7+V14		54,713	0,766	0,381
Vd1+Vd2+Vd3+V5+V7+V17		53,048	2,431	0,119
Vd1+Vd2+Vd3+V5+V7+V22		51,682	3,797	0,051*
Vd1+Vd2+Vd3+V5+V7+V23		51,976	3,503	0,061*
Vd1+Vd2+Vd3+V5+V7+V29		55,478	0,001	0,975
Vd1+Vd2+Vd3+V5+V7+V32		53,475	2,004	0,157
Vd1+Vd2+Vd3+V5+V7+V34		55,476	0,003	0,956
Vd1+Vd2+Vd3+V5+V7+V35		52,392	3,087	0,079*
Passo 4	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35	50,567	-	-
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V4	50,567	0,000	1,000
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V6	43,143	7,424	0,006*
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V8	50,563	0,004	0,950
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V9	50,198	0,369	0,544
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V10	48,313	2,254	0,133
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V11	47,043	3,524	0,060*
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V13	49,827	0,740	0,390
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V15	48,159	2,408	0,121
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V16	49,886	0,681	0,409
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V18	48,988	1,579	0,209
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V19	48,963	1,604	0,205
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V20	50,338	0,229	0,632
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V21	50,413	0,154	0,695
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V24	48,011	2,556	0,110
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V25	48,826	1,741	0,187
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V26	49,518	1,049	0,306
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V27	49,652	0,915	0,339
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V28	50,564	0,003	0,956
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V30	49,439	1,128	0,288
Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V31	49,436	1,131	0,288	
Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V33	47,528	3,039	0,081*	
Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V36	47,255	3,312	0,069*	
Passo 5	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V6+V11+V33+V36	38,752	-	-
	Vd2+Vd3+V5+V7+V22+V23+V35+V6+V11+V33+V36	45,858	7,106	0,008*
	Vd1+Vd3+V5+V7+V22+V23+V35+V6+V11+V33+V36	39,997	1,245	0,265
	Vd1+Vd2+V5+V7+V22+V23+V35+V6+V11+V33+V36	38,885	0,133	0,715

...continuação

	Vd1+Vd2+Vd3+V7+V22+V23+V35+V6+V11+V33+V36	41,338	2,586	0,108
	Vd1+Vd2+Vd3+V5+V22+V23+V35+V6+V11+V33+V36	51,740	12,988	0,000*
	Vd1+Vd2+Vd3+V5+V7+V23+V35+V6+V11+V33+V36	38,830	0,078	0,780
	Vd1+Vd2+Vd3+V5+V7+V22+V35+V6+V11+V33+V36	38,754	0,002	0,964
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V6+V11+V33+V36	43,744	4,992	0,025*
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V11+V33+V36	44,098	5,346	0,021*
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V6+V33+V36	39,150	0,398	0,528
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V6+V11+V36	40,763	2,011	0,156
	Vd1+Vd2+Vd3+V5+V7+V22+V23+V35+V6+V11+V33	38,765	0,013	0,909
Passo 6	Vd1+Vd2+Vd3+V7+V35+V6	47,977	-	-
	Vd1+Vd2+Vd3+V7+V35+V6+Vd1*Vd2	47,977	0,000	1,000
	Vd1+Vd2+Vd3+V7+V35+V6+Vd1*Vd3	47,977	0,000	1,000
	Vd1+Vd2+Vd3+V7+V35+V6+Vd1*V7	47,187	0,790	0,374
	Vd1+Vd2+Vd3+V7+V35+V6+Vd1*V35	47,860	0,117	0,732
	Vd1+Vd2+Vd3+V7+V35+V6+Vd1*V6	47,975	0,002	0,964
	Vd1+Vd2+Vd3+V7+V35+V6+Vd2*Vd3	47,977	0,000	1,000
	Vd1+Vd2+Vd3+V7+V35+V6+Vd2*V7	47,805	0,172	0,678
	Vd1+Vd2+Vd3+V7+V35+V6+Vd2*V35	45,473	2,504	0,114
	Vd1+Vd2+Vd3+V7+V35+V6+Vd2*V6	45,155	2,822	0,093*
	Vd1+Vd2+Vd3+V7+V35+V6+Vd3*V7	47,879	0,098	0,754
	Vd1+Vd2+Vd3+V7+V35+V6+Vd3*V35	45,479	2,498	0,114
	Vd1+Vd2+Vd3+V7+V35+V6+Vd3*V6	47,959	0,018	0,893
	Vd1+Vd2+Vd3+V7+V35+V6+V7*V35	45,593	2,384	0,123
	Vd1+Vd2+Vd3+V7+V35+V6+V7*V6	44,536	3,441	0,064*
	Vd1+Vd2+Vd3+V7+V35+V6+V35*V6	47,711	0,266	0,606
Etapa Final	Vd1+Vd2+Vd3+V7+V35+V6+Vd2*V6+V7*V6	43,233		
	Vd1+Vd2+Vd3+V7+V35+V6+Vd2*V6	45,155		
	Vd1+Vd2+Vd3+V7+V35+V6+V7*V6	44,536		

\* Valor  $p \leq 0,10$

Desta forma, os possíveis modelos para a estimativa da probabilidade de ocorrência de mais de 9 indivíduos por  $m^2$  no solo, nas áreas estudadas foram:

- Modelo 1:

$$P_{(\text{Mais de 9 indivíduos})} = \frac{e^{(\beta_0 + \beta_1 Vd1 + \beta_2 Vd2 + \beta_3 Vd3 + \beta_4 V7 + \beta_5 V35 + \beta_6 V6 + \beta_7 Vd2*V6 + \beta_8 V7*V6)}}{1 + e^{(\beta_0 + \beta_1 Vd1 + \beta_2 Vd2 + \beta_3 Vd3 + \beta_4 V7 + \beta_5 V35 + \beta_6 V6 + \beta_7 Vd2*V6 + \beta_8 V7*V6)}}$$

- Modelo 2:

$$P_{(\text{Mais de 9 indivíduos})} = \frac{e^{(\beta_0 + \beta_1 Vd1 + \beta_2 Vd2 + \beta_3 Vd3 + \beta_4 V7 + \beta_5 V35 + \beta_6 V6 + \beta_7 Vd2*V6)}}{1 + e^{(\beta_0 + \beta_1 Vd1 + \beta_2 Vd2 + \beta_3 Vd3 + \beta_4 V7 + \beta_5 V35 + \beta_6 V6 + \beta_7 Vd2*V6)}}$$

- Modelo 3:

$$P_{(\text{Mais de 9 indivíduos})} = \frac{e^{(\beta_0 + \beta_1 Vd1 + \beta_2 Vd2 + \beta_3 Vd3 + \beta_4 V7 + \beta_5 V35 + \beta_6 V6 + \beta_7 V7*V6)}}{1 + e^{(\beta_0 + \beta_1 Vd1 + \beta_2 Vd2 + \beta_3 Vd3 + \beta_4 V7 + \beta_5 V35 + \beta_6 V6 + \beta_7 V7*V6)}}$$

Para avaliar se os modelos finais foram bem ajustados e então decidir qual deles deve ser usado, optou-se por utilizar o Teste Hosmer e Lemeshow e a área sob a Curva ROC, que testam a qualidade do ajuste, avaliando a capacidade preditiva dos modelos.

O teste de Hosmer e Lemeshow avalia as diferenças entre as classificações previstas pelo modelo e as observadas (HOSMER e LEMESHOW, 1989). Se as diferenças forem significativas, o grau de acurácia do modelo não é bom. A Tabela 4.2 exibe o resultado do teste para os três modelos. Ao nível de significância de 5%, não foi possível rejeitar a hipótese nula de que não houve diferenças significativas entre os valores preditos e observados para os modelos 1 e 3, o que indicou que esses modelos foram capazes de produzir classificações confiáveis.

**Tabela 4.2 – Teste de Hosmer e Lemeshow para o conjunto de dados da serapilheira e do solo no período seco**

Modelo	Qui-quadrado	g.l.	Valor <i>p</i>
1	66,001	8	0,580
2	169,000	8	0,031
3	64,149	8	0,601

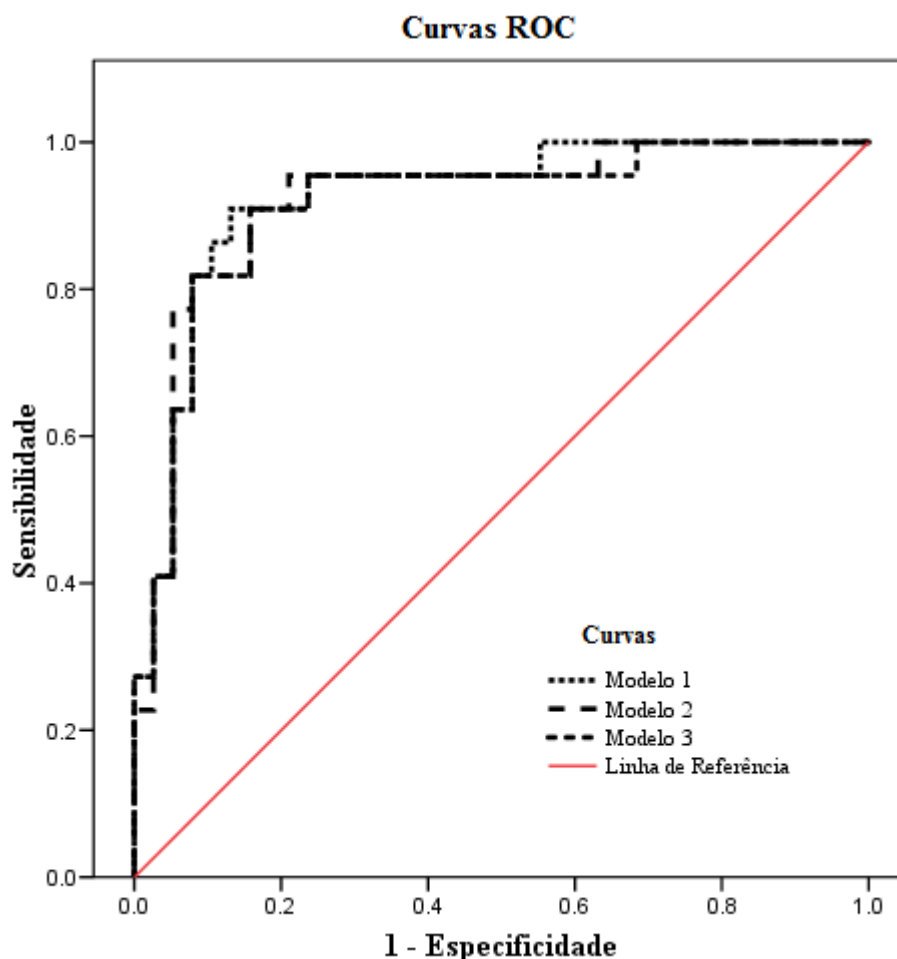
Da mesma maneira, a área sob a curva ROC que varia de zero a um, pode ser empregada como um índice da exatidão do modelo, sendo tanto melhor quanto maior for esta área.

Como a probabilidade da variável dependente estimada pode assumir um valor entre zero e um, neste estudo, foi utilizado um ponto de corte de 0,5 para o modelo, de forma que as amostras com resultados inferiores a 0,5 são classificadas como “presença de 9 ou menos indivíduos por m<sup>2</sup>, em média, no solo” e as amostras com resultados superiores a esse valor são classificadas como “presença de mais de 9 indivíduos por m<sup>2</sup>, em média, no solo.

Na Tabela 4.3 foram apresentadas as medidas de capacidade preditiva (CAT, CAU e CAZ) e a área sob a curva ROC (ASC) para os três modelos e na figura 4.1, suas respectivas curvas ROC.

**Tabela 4.3 - Medidas de capacidade preditiva no modelo ajustado para o conjunto de dados da serapilheira e do solo no período seco**

Modelo	CAT	CAU	CAZ	ASC
1	88,3%	81,8%	92,1%	0,923
2	86,7%	77,3%	92,1%	0,920
3	86,7%	77,3%	92,1%	0,914



**Figura 4.1 – Curvas ROC dos modelos ajustados para o conjunto de dados da serapilheira e do solo no período seco**

Pela análise das capacidades preditivas, notou-se que não houve diferenças entre os modelos 2 e 3. A capacidade preditiva total (CAT) e capacidade preditiva dos uns (CAU) foi maior no modelo 1 que nos modelos 2 e 3, mesmo assim, essa variação foi pequena. A capacidade preditiva dos zeros (CAZ) foi igual nos três modelos. A área sob a curva (ASC) foi maior que 0,90 nos três modelos analisados, representando uma excepcional capacidade de discriminação.

Como foi visto, pela curva ROC os três modelos discriminaram bem os dados, porém, pelo teste de Hosmer e Lemeshow apenas os modelo 1 e 3 foram bem ajustados. Assim, o modelo mais adequado para a análise da quantidade média de indivíduos no solo em época seca, foi o modelo 3, uma vez que ele é bem ajustado e é mais parcimonioso. O modelo final ficou composto pelas covariáveis: *Dummy* sistema 1 - Sistema Convencional/Mata (Vd1), *Dummy* sistema 2 - Sistema Convencional/

Agroecológico (Vd2), *Dummy* sistema 3 - Sistema Convencional/ Agroflorestal (Vd3), Cálcio em serapilheira (V7), Matéria orgânica do solo (V35), Potássio em serapilheira (V6) e a interação entre Cálcio e Potássio em serapilheira (V7\*V6).

Além de obter um modelo, testar a significância de seus parâmetros e verificar a acurácia e eficiência desse modelo encontrado, outra análise interessante de ser feita é a da razão das chances, calculada por  $\exp(\hat{\beta})$ . A Tabela 4.4 mostra os valores dessas razões para o modelo final.

**Tabela 4.4 – Razão de chance do modelo final ajustado para o conjunto de dados da serapilheira e do solo no período seco**

Variáveis	$\hat{\beta}$	Erro padrão	Valor <i>p</i>	Razão de Chance $\exp(\hat{\beta})$
Constante	0,852	4,270	0,842	
Vd1	3,800	1,501	0,011	44,688
Vd2	0,190	1,265	0,880	1,210
Vd3	-0,022	1,258	0,986	0,979
V7	-0,730	0,323	0,024	0,482
V35	0,431	0,169	0,011	1,538
V6	-0,903	0,858	0,293	0,405
V7*V6	0,118	0,070	0,092	1,125

Pode-se observar que dentre os fatores que influenciam a presença de mais ou menos indivíduos por m<sup>2</sup> no solo nas áreas estudadas, o *cálcio* e o *potássio em serapilheira*, atuaram de forma negativa, isto é, quando a quantidade desses elementos for alta na serapilheira, as chances de aumentar o número de indivíduos no solo diminuem. Assim, o aumento de uma unidade (em gkg-1) de cálcio e de potássio, separadamente, diminui em aproximadamente 52% e 60% respectivamente, as chances de ocorrência de mais de 9 indivíduos por m<sup>2</sup> no solo. Porém, a interação entre esses dois elementos no solo atuou de forma positiva. Verificou-se que com o aumento de uma unidade da interação entre cálcio e potássio, existe a possibilidade de se aumentar em aproximadamente 13% as chances de ocorrência de mais de 9 indivíduos por m<sup>2</sup> no solo.

Com o aumento de uma unidade de *matéria orgânica no solo*, as chances de ocorrência de mais de 9 indivíduos por m<sup>2</sup> no solo aumentam em 54% aproximadamente. Em relação ao Sistema Convencional, categoria de referência utilizada na codificação das variáveis *dummies*, o Sistema “Mata” aumenta aproximadamente 45 vezes a chance de se encontrar mais de 9 indivíduos por m<sup>2</sup> no

solo das áreas estudadas, em época seca. Os Sistemas Agroflorestal e Agroecológico não apresentaram significância estatística em relação ao Sistema Convencional.

#### **4.2 Conjunto de dados da Serapilheira e do solo no período chuvoso – atributos físicos e químicos**

Foi utilizada a estratégia de seleção de modelos derivada da proposta de Collett (1994) para a seleção das covariáveis na construção do modelo de regressão logística. Os resultados podem ser vistos na Tabela 4.5.

Foram ajustados todos os modelos contendo uma única covariável (passo 1) e pelo teste da razão de verossimilhanças foi observado que as covariáveis V5, V6, V8, V12, V14, V25, V26, V27, V30, V31, V32, V35 foram significativas ao nível de 0,10, ou seja, têm alguma influência sobre a ocorrência de mais de 101 indivíduos por m<sup>2</sup>, em média, no solo das áreas estudadas.

No passo 2, as covariáveis significativas anteriormente foram então ajustadas conjuntamente. Neste passo ainda, foram também ajustados modelos reduzidos, excluindo uma única covariável de cada vez. Verificou-se que apenas as covariáveis V8, V12 e V14 provocam um aumento significativo na estatística da razão de verossimilhanças. Assim, no passo 3, ajustou-se um novo modelo com essas covariáveis e foram ajustados também outros modelos que incluíam as covariáveis que foram excluídas no passo 2 para verificar se realmente não eram estatisticamente significativas. Foi possível observar que todas as covariáveis excluídas no passo 2 foram novamente significativa, devendo, dessa forma retornar ao modelo.

No passo 4, os modelos ajustados voltaram a conter as doze variáveis iniciais do passo 2. Neste passo retornou-se com as covariáveis excluídas no passo 1, uma a uma, para confirmar se elas não eram estatisticamente significativas. Ajustou-se então (passo 5), um modelo incluindo as covariáveis V11, V16 e V29 significativas no passo 4 e foi testado se alguma delas poderia ser retirada do modelo. As únicas covariáveis que sobreviveram nesse passo foram V8, V12, V14 e V16, pois apresentaram significância estatística.

No passo 6, com as quatro covariáveis selecionadas ajustou-se o modelo final. Para completar a modelagem foi verificada a possibilidade de inclusão de termos de interação dupla entre as covariáveis já incluídas no modelo. Nenhuma interação foi significativa ao nível de 0,10.

**Tabela 4.5 – Seleção de covariáveis usando o modelo de regressão logística para o conjunto de dados da serapilheira e do solo no período chuvoso**

Passos	Modelo	-2 log L ( $\beta$ )	TRV	Valor p
Passo 1	Nulo	65,193	-	-
	Vd1	63,973	1,220	0,269
	Vd2	64,132	1,061	0,303
	Vd3	65,071	0,122	0,727
	V4	65,059	0,134	0,714
	V5	51,011	14,182	0,000*
	V6	56,962	8,231	0,004*
	V7	63,215	1,978	0,160
	V8	61,860	3,333	0,068*
	V9	64,855	0,338	0,561
	V10	65,011	0,182	0,670
	V11	65,191	0,002	0,964
	V12	61,764	3,429	0,064*
	V13	65,145	0,048	0,827
	V14	61,549	3,644	0,056*
	V15	62,800	2,393	0,122
	V16	63,051	2,142	0,143
	V17	63,432	1,761	0,184
	V18	64,652	0,541	0,462
	V19	64,655	0,538	0,463
	V20	64,803	0,390	0,532
	V21	65,092	0,101	0,751
	V22	64,721	0,472	0,492
	V23	64,791	0,402	0,526
	V24	64,941	0,252	0,616
	V25	59,723	5,470	0,019*
	V26	59,862	5,331	0,021*
	V27	59,760	5,433	0,020*
	V28	63,483	1,710	0,191
	V29	63,737	1,456	0,228
	V30	59,602	5,591	0,018*
	V31	59,579	5,614	0,018*
	V32	56,411	8,782	0,003*
	V33	64,621	0,572	0,449
	V34	62,928	2,265	0,132
	V35	54,531	10,662	0,001*
V36	64,633	0,560	0,454	
Passo 2	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35	32,083	-	-
	V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35	33,123	1,040	0,308
	V5+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35	32,348	0,265	0,607
	V5+V6+V12+V14+V25+V26+V27+V30+V31+V32+V35	35,961	3,878	0,049*
	V5+V6+V8+V14+V25+V26+V27+V30+V31+V32+V35	34,800	2,717	0,099*
	V5+V6+V8+V12+V25+V26+V27+V30+V31+V32+V35	35,020	2,937	0,087*
	V5+V6+V8+V12+V14+V26+V27+V30+V31+V32+V35	32,110	0,027	0,869
	V5+V6+V8+V12+V14+V25+V27+V30+V31+V32+V35	32,739	0,656	0,418
	V5+V6+V8+V12+V14+V25+V26+V30+V31+V32+V35	32,686	0,603	0,437
	V5+V6+V8+V12+V14+V25+V26+V27+V31+V32+V35	32,392	0,309	0,578
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V32+V35	32,083	0,000	1,000
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V35	33,750	1,667	0,197
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32	32,285	0,202	0,653
	Passo 3	V8+V12+V14	55,751	-
V8+V12+V14+V5		47,072	8,679	0,003*
V8+V12+V14+V6		51,972	3,779	0,052*
V8+V12+V14+V25		50,385	5,366	0,021*

...continuação

	V8+V12+V14+V26	50,189	5,562	0,018*
	V8+V12+V14+V27	52,126	3,625	0,057*
	V8+V12+V14+V30	51,865	3,886	0,049*
	V8+V12+V14+V31	51,671	4,080	0,043*
	V8+V12+V14+V32	39,884	15,867	0,000*
	V8+V12+V14+V35	41,760	13,991	0,000*
Passo 4	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35	32,083	-	-
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+Vd1	30,794	1,289	0,256
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+Vd2	30,698	1,385	0,239
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+Vd3	31,592	0,491	0,483
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V4	31,912	0,171	0,679
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V7	30,755	1,328	0,249
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V9	31,835	0,248	0,618
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V10	30,935	1,148	0,284
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V11	27,159	4,924	0,026*
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V13	32,059	0,024	0,877
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V15	32,046	0,037	0,847
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V16	16,772	15,311	0,000*
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V17	31,508	0,575	0,448
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V18	30,500	1,583	0,208
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V19	30,497	1,586	0,208
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V20	31,819	0,264	0,607
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V21	31,776	0,307	0,580
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V22	31,463	0,620	0,431
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V23	31,455	0,628	0,428
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V24	32,081	0,002	0,964
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V28	32,060	0,023	0,879
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V29	28,761	3,322	0,068*
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V33	32,030	0,053	0,818
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V34	31,224	0,859	0,354
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V36	32,083	0,000	1,000
Passo 5	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V11+V16+V29	16,335	-	-
	V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V11+V16+V29	17,371	1,036	0,309
	V5+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V11+V16+V29	16,488	0,153	0,696
	V5+V6+V12+V14+V25+V26+V27+V30+V31+V32+V35+V11+V16+V29	22,608	6,273	0,012*
	V5+V6+V8+V14+V25+V26+V27+V30+V31+V32+V35+V11+V16+V29	20,988	4,653	0,031*
	V5+V6+V8+V12+V25+V26+V27+V30+V31+V32+V35+V11+V16+V29	21,437	5,102	0,024*
	V5+V6+V8+V12+V14+V26+V27+V30+V31+V32+V35+V11+V16+V29	16,474	0,139	0,709
	V5+V6+V8+V12+V14+V25+V27+V30+V31+V32+V35+V11+V16+V29	17,034	0,699	0,403
	V5+V6+V8+V12+V14+V25+V26+V30+V31+V32+V35+V11+V16+V29	16,396	0,061	0,805
	V5+V6+V8+V12+V14+V25+V26+V27+V31+V32+V35+V11+V16+V29	16,630	0,295	0,587
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V32+V35+V11+V16+V29	16,657	0,322	0,570
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V35+V11+V16+V29	16,340	0,005	0,944
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V11+V16+V29	16,660	0,325	0,569
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V16+V29	16,352	0,017	0,896
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V11+V29	24,994	8,659	0,003*
	V5+V6+V8+V12+V14+V25+V26+V27+V30+V31+V32+V35+V11+V16	16,753	0,418	0,518
Passo 6	V8+V12+V14+V16	39,251	-	-
	V8+V12+V14+V16+V8*V12	39,208	0,043	0,836
	V8+V12+V14+V16+V8*V14	39,217	0,034	0,854
	V8+V12+V14+V16+V8*V16	37,344	1,907	0,167
	V8+V12+V14+V16+V12*V14	39,057	0,194	0,660
	V8+V12+V14+V16+V12*V16	39,206	0,045	0,832
	V8+V12+V14+V16+V14*V16	39,186	0,065	0,799
Modelo Final	V8+V12+V14+V16			

\* Valor  $p \leq 0,10$

Desta forma, o modelo final para a estimativa da probabilidade de ocorrência de mais de 101 indivíduos por m<sup>2</sup> no solo, nas áreas estudadas foram:

$$P_{\text{(Mais de 101 indivíduos)}} = \frac{e^{(\beta_0 + \beta_1 V_8 + \beta_2 V_{12} + \beta_3 V_{14} + \beta_4 V_{16})}}{1 + e^{(\beta_0 + \beta_1 V_8 + \beta_2 V_{12} + \beta_3 V_{14} + \beta_4 V_{16})}}$$

Para avaliar a qualidade do ajuste para o modelo final utilizou-se o Teste Hosmer e Lemeshow e a área sob a Curva ROC.

Pelo teste de Hosmer e Lemeshow, ao nível de significância de 5%, não foi possível rejeitar a hipótese nula de que não houve diferenças significativas entre os valores preditos e observados, indicando que o modelo foi capaz de produzir classificações confiáveis, isto é, o grau de acurácia do modelo é bom (Tabela 4.6).

**Tabela 4.6 – Teste de Hosmer e Lemeshow no modelo ajustado para o conjunto de dados da serapilheira e do solo no período chuvoso**

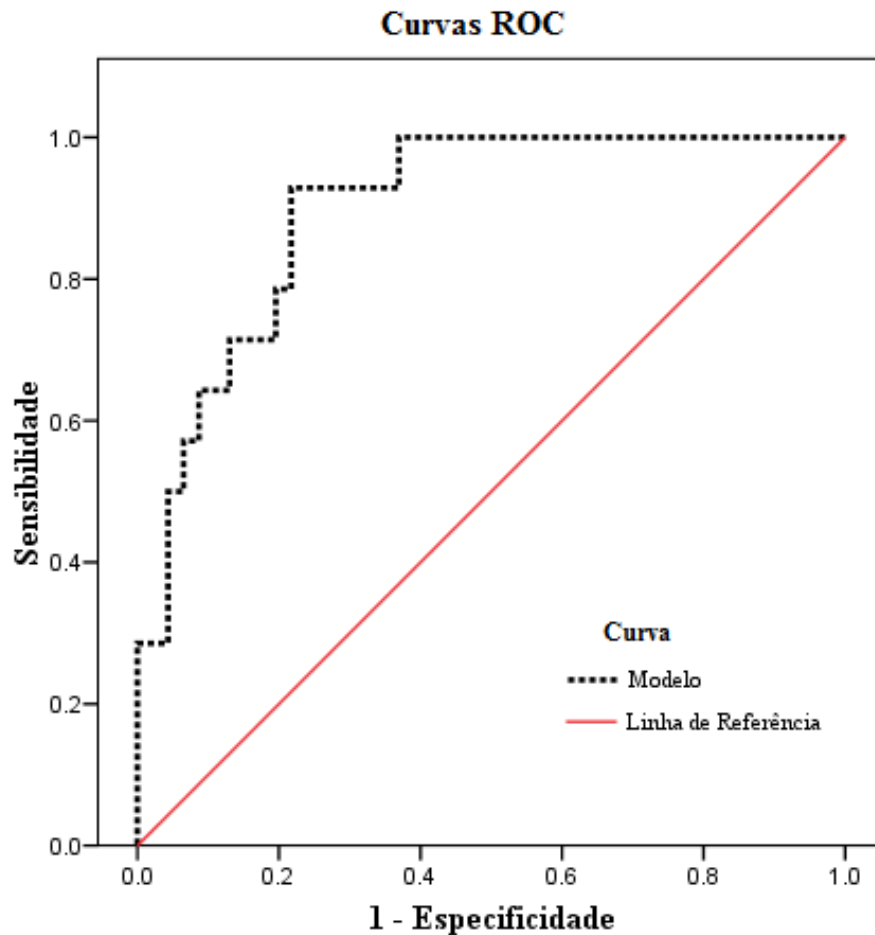
Qui-quadrado	g.l.	Valor <i>p</i>
5,7390	8	0,676

Da mesma maneira, a área sob a curva ROC que varia de zero a um, pode ser empregada como um índice da exatidão do modelo, sendo tanto melhor quanto maior for esta área.

Como a variável dependente estimada pode assumir um valor entre zero e um, neste estudo, foi utilizado um ponto de corte de 0,5 para o modelo, de forma que as amostras com resultados inferiores a 0,5 são classificadas como “presença de 101 ou menos indivíduos por m<sup>2</sup>, em média, no solo” e as amostras com resultados superiores a esse valor são classificadas como “presença de mais de 101 indivíduos por m<sup>2</sup>, em média, no solo”. Na Tabela 4.7 foram apresentadas as medidas de capacidade preditiva (CAT, CAU e CAZ) e a área sob a curva ROC (ASC) para o modelo final e na figura 4.2, suas respectivas curvas ROC.

**Tabela 4.7 - Medidas de capacidade preditiva no modelo ajustado para o conjunto de dados da serapilheira e do solo no período chuvoso**

CAT	CAU	CAZ	ASC
83,3%	57,1%	91,3%	0,899



**Figura 4.2 – Curvas ROC do modelo ajustado para o conjunto de dados da serapilheira e do solo no período chuvoso**

O modelo apresentou boa capacidade preditiva total (CAT) e boa capacidade preditiva dos zeros (CAZ), porém capacidade preditiva dos uns (CAU) foi menor que 60%. A área sob a curva (ASC) foi 0,899, representando uma excelente capacidade de discriminação.

Como foi verificado, o modelo final foi bem ajustado e ficou composto pelas covariáveis: Magnésio em serapilheira (V8), Carbono orgânico total na serapilheira (V12), Matéria orgânica da serapilheira (V14) e Temperatura ambiente (V16).

Além de obter um modelo, testar a significância de seus parâmetros e verificar a acurácia e eficiência desse modelo encontrado foi realizada a análise da razão das chances, calculada por  $\exp(\hat{\beta})$  (Tabela 4.8).

**Tabela 4.8 – Razão de chance do modelo final ajustado para o conjunto de dados da serapilheira e do solo no período chuvoso**

Variáveis	$\hat{\beta}$	Erro Padrão	Valor <i>p</i>	Razão de Chance $\exp(\hat{\beta})$
Constante	25,507	8,908	0,004	
V8	2,781	0,878	0,002	16,135
V12	4,588	2,449	0,061	98,268
V14	-2,736	1,379	0,047	0,065
V16	-0,834	0,271	0,002	0,434

Dentre os fatores que influenciam a presença de mais ou menos indivíduos por m<sup>2</sup> no solo nas áreas estudadas, o *magnésio* e o *carbono orgânico total na serapilheira*, atuaram de forma positiva, isto é, quando a quantidade desses elementos for alta na serapilheira, as chances de aumentar o número de indivíduos no solo também aumentam. Assim, o aumento de uma unidade de magnésio e de carbono orgânico total na serapilheira aumenta em aproximadamente 16 e 98 vezes, respectivamente, as chances de ocorrência de mais de 101 indivíduos por m<sup>2</sup> no solo.

Com o aumento de uma unidade de *matéria orgânica na serapilheira*, as chances de ocorrência de mais de 101 indivíduos por m<sup>2</sup> no solo diminuem em 94% aproximadamente, e se houver o aumento de um grau na *temperatura ambiente* em relação à temperatura ambiente registrada na época da pesquisa, pode ocorrer uma diminuição de aproximadamente 57% da chance de se encontrar mais de 101 indivíduos por m<sup>2</sup> no solo das áreas estudadas, em época chuvosa.

Sendo a matéria orgânica composta por carbono orgânico, observa-se que o carbono orgânico na serapilheira contribui de forma positiva para o aumento de indivíduos no solo, enquanto que a matéria orgânica na serapilheira contribui de forma negativa para esse aumento. Essa incoerência pode ter ocorrido devido o experimento ser realizado em ambiente não controlado e com pequeno número de repetição.

## 5 CONCLUSÕES E RECOMENDAÇÕES

### 5.1 Conclusões

Das trinta e seis covariáveis iniciais utilizadas para estudar o problema, somente quatro foram importantes no modelo final, tanto para o período seco quanto para o período chuvoso. No período seco as covariáveis significativas foram: *Dummy* sistema 1 - Sistema Convencional/Mata, Cálcio em serapilheira, Matéria orgânica do solo, Potássio em serapilheira, além disso, foram incluídos os outros níveis da covariável “*Tipo de sistema*” por se tratar de uma variável indicadora (*Dummy* sistema 2 - Sistema Convencional/Agroecológico e *Dummy* sistema 3 - Sistema Convencional/Agroflorestal) e uma interação significativa entre Cálcio e Potássio em serapilheira.

As covariáveis significativas no período chuvoso foram: Magnésio em serapilheira, Carbono orgânico total na serapilheira, Matéria orgânica da serapilheira e Temperatura ambiente.

O desempenho discriminatório do modelo final para cada período analisado (seco e chuvoso), avaliado pelo teste de Hosmer e Lemeshow e pela curva ROC, não apresentou diferença estatisticamente significativa entre os valores observados e preditos. Os resultados obtidos mostraram excelentes áreas sob a curva, confirmando assim a validade da utilização de técnicas de regressão logística na construção dos modelos para descrever os dados analisados.

A covariável “*tipo de sistema*” mostrou-se importante apenas no período seco, pois o solo da mata apresentou elevada razão de chance de apresentar maior quantidade de indivíduos (invertebrados edáficos) em relação ao solo onde se pratica o manejo do café de forma convencional, com o uso de fertilizantes e agrotóxicos. O solo onde se cultiva o café sob o sistema de manejo agroecológico ou agroflorestal não apresentou diferenças significativas em relação ao solo onde se cultiva o café sob o sistema convencional.

A covariável “*matéria orgânica*” esteve presente no modelo final para o período seco e chuvoso. No período seco, foi possível observar que quanto maior a quantidade de matéria orgânica no solo, maior sua contribuição para o aumento da quantidade de invertebrados edáficos, porém, no período chuvoso uma grande quantidade de matéria

orgânica na serapilheira contribui para a diminuição do número de invertebrados no solo.

Sendo a matéria orgânica composta por carbono orgânico, foi observada certa incoerência no período chuvoso – o carbono orgânico na serapilheira contribui de forma positiva para o aumento de indivíduos no solo, enquanto que a matéria orgânica na serapilheira contribui de forma negativa para esse aumento, isto é, contribui para a diminuição.

Somente o modelo final para o período seco apresentou interações significativas entre covariáveis. A interação entre Cálcio e Potássio contribui em aproximadamente 13% para o aumento do número de indivíduos por m<sup>2</sup> no solo em período seco.

A temperatura ambiente só foi relevante no modelo final para o período chuvoso onde é maior a concentração de indivíduos por m<sup>2</sup>, sendo que se seu valor for elevado, aumenta-se a chance de diminuir a quantidade e a variedade de invertebrados no solo.

## **5.2 Recomendações**

Desenvolver outras pesquisas com aplicação dessa mesma metodologia, com os ajustes necessários a cada caso, em outras localidades e com um maior número de repetições, a fim de ampliar os conhecimentos e melhor compreender a relação entre a variável resposta e as demais covariáveis do modelo, buscando resultados mais representativos.

Testar interações entre outras covariáveis e verificar a significância dessas interações no modelo.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, H. J. *Aplicação da Análise de Sobrevida em um problema de Credit Scoring e comparação com Regressão Logística*. 2004. 118 f. Dissertação (Mestrado) - Universidade Federal do São Carlos. 2004.

BARROS, E.; NEVES, A.; BLANCHART, E.; FERNANDES, E.C.; WANDELLI, E.; LAVELLE, P. Development of the soil macrofauna community under silvopastoral and agrosilvicultural systems in Amazonia. *Pedobiologia*, Jena, v.47, p.273-280, 2003.

BROCCO, J. B. *Ponderação de modelos com aplicação de Regressão Logística binária*. 2006. 78 f. Dissertação (Mestrado) - Universidade Federal do São Carlos. 2006.

CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2nd ed. Pacific Grove: Duxbury, 2002.

COLLETT, D. *Modelling Survival Data in Medical Research*. Chapman and Hall, London, 1994.

COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevida Aplicada*. ABE - Projeto Fisher. São Paulo: Edgar Blücher, 2006.

CORREIA, M. E. F.; ANDRADE, A. G. Formação da serapilheira e ciclagem de nutrientes. In: SANTOS, G. A.; CAMARGO, F. A. O. (Eds.). *Fundamentos da matéria orgânica do solo: ecossistemas tropicais e subtropicais*. Porto alegre: Gênese, 1999. p.197-225.

FIGUEIRA, C. V. *Modelos de regressão logística*. 2006. 138 f. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul. 2006.

HANLEY, J. A.; MCNEIL, B. J. A Method of Comparing the Areas under Receiver Operating Characteristic Curves. *Radiology*, 148: 839-843. 1983.

HAUCK, W. W.; DONNER, A. Wald's Tests Applied to Hypotheses in Logit Analysis. *Journal of the American Statistical Association*. v. 72. 851-853.1977.

HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. 2nd ed. Massachusetts: John Wiley & Sons, 2000.

HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. Massachusetts: John Wiley & Sons, 1989.

HU, F.; LI, H.X.; WU, S.M. Differentiation of soil fauna populations in conventional tillage and no-tillage red soil ecosystems. *Pedosphere*. v.7, p.339-348, 1997.

JENNINGS, D. E. Judging Inference Adequacy in Logistic Regression. *Journal of the American Statistical Association*. v. 81. 471-476. 1986a.

JENNINGS, D. E. Judging Inference Adequacy in Logistic Regression. *Journal of the American Statistical Association*. v. 81. 987-990. 1986b.

LINDEN, D. R., HENDRIX, P. F., COLEMAN, D. C. Faunal indicators of soil quality. In: *Defining Soil Quality for a Sustainable Environment*. Madison, SSSA, 1994. p. 91-106. (Special Publication, 35).

MARTINS, C. A. C. *Análise de regressão logística*. 1998. 53 f. Dissertação (Mestrado) - Universidade Federal de Pernambuco. 1988.

MELO, F. V. et al. A importância da meso e macrofauna do solo na fertilidade e como bioindicadores. *Boletim Informativo da Sociedade Brasileira de Ciência do Solo*, 34(01), 38-43. Jan/abr. 2009

MEYER, P. L. *Probabilidade: aplicações à estatística*. Rio de Janeiro: Livros Técnicos e Científicos, 1978.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, London, v.135, p370-384, 1972.

PAOLETTI, M.G. The role of earthworms for assessment of sustainability and as bioindicators. *Agriculture, Ecosystems and Environment*, Amsterdam, v.74, p.137-155, 1999.

R Development Core Team 2009. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org>>. Acesso em: Nov. 2010.

SILVA, J. *Invertebrados Edáficos em Sistemas de Produção com Café na Zona da Mata de Minas Gerais*. 2010. 218 f. Tese (Doutorado) – Universidade Federal do Ceará. 2010.

SIQUEIRA, J. O.; MOREIRA, F. M. S. *Microbiologia do solo e sustentabilidade agrícola: enfoque em fertilidade do solo e nutrição vegetal*. In: Reunião Brasileira em Fertilidade do Solo e Nutrição de Plantas, 22, 1996, Manaus. Resumos... Manaus: SBCS, 1996, p.1-42.

SOUZA, M. E. P. *Oligochaetas em solos sob sistemas de manejos a pleno sol e agroflorestal e vermicompostagem associada com pós de rochas*. 2010. 58 f. Dissertação (Mestrado) - Universidade Federal de Viçosa, 2010.

WAGNER, M. B.; MOTTA, V. T.; DORNELLES, C. *SPSS passo a passo: Statistical Package for the Social Sciences*. Caxias do Sul: Educs, 2004. 172p.

WINK C., GUEDES J. V. C., FAGUNDES C.K., ROVEDDER A. P. *Insetos edáficos como indicadores da qualidade ambiental* In: Revista de Ciências Agroveterinárias, Lages, v.4, n.1, p. 60-71, 2005.

## **APÉNDICE**

## **Análise dos dados da serapilheira e do solo no período de seca e no período chuvoso – atributos físicos e químicos**

### **LEGENDA**

**Vd1** = *Dummy* sistema 1 (*Sistema Convencional/Mata*); **Vd2** = *Dummy* sistema 2 (*Sistema Convencional/Sistema Agroecológico*); **Vd3** = *Dummy* sistema 3 (*Sistema Convencional/Sistema Agroflorestal*); **V4** = Nitrogênio total em serapilheira; **V5** = Fósforo em serapilheira; **V6** = Potássio em serapilheira; **V7** = Cálcio em serapilheira; **V8** = Magnésio em serapilheira; **V9** = Manganês em serapilheira; **V10** = Zinco em serapilheira; **V11** = Ferro em serapilheira; **V12** = Carbono orgânico total na serapilheira; **V13** = Relação carbomo/nitrogênio na serapilheira; **V14** = Matéria orgânica da serapilheira; **V15** = Porcentagem de umidade do solo; **V16** = Temperatura ambiente; **V17** = Temperatura do solo; **V18** = Peso seco da serapilheira em t/ha; **V19** = Peso seco da serapilheira em g/kg; **V20** = Microporosidade do solo; **V21** = Macroporosidade do solo; **V22** = Porosidade total do solo; **V23** = Densidade do solo; **V24** = Ph do solo; **V25** = Fósforo no solo; **V26** = Potássio no solo; **V27** = Cálcio no solo; **V28** = Magnésio no solo; **V29** = Alumínio no solo; **V30** = Soma de base; **V31** = CTC (capacidade de troca de cátions do solo) efetiva; **V32** = CTC (capacidade de troca de cátions do solo) total; **V33** = Saturação de bases do solo; **V34** = Saturação por alumínio; **V35** = Matéria orgânica do solo; **V36** = Fósforo remanescente do solo.

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais												
	Vd1	Vd2	Vd3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
1	0	0	0	14,95	0,96	5,7	10,75	5,21	65,68	11,46	42,25	36,94	25
2	0	0	0	13,54	0,79	6,4	9,91	4,73	48,70	6,23	20,89	34,83	26
3	0	0	0	13,89	0,98	5,1	10,96	5,12	86,72	20,25	64,14	34,12	25
4	0	0	0	11,78	2,08	4,0	10,03	4,36	86,22	21,68	72,09	34,63	29
5	0	0	0	14,25	1,84	10,4	12,31	5,56	64,13	19,11	28,34	34,08	24
6	0	0	0	13,19	0,75	6,4	7,62	2,62	141,2	18,56	9,09	35,51	27
7	0	0	0	15,30	0,30	11,6	8,81	2,81	143,16	17,89	11,92	37,02	24
8	0	0	0	13,54	0,73	4,9	9,57	3,35	167,09	16,04	23,45	36,69	27
9	0	0	0	15,65	0,68	8,3	9,75	2,99	238,44	12,99	31,53	34,80	22
10	0	0	0	14,60	0,29	7,8	8,54	3,20	173,91	14,94	27,9	37,84	26
11	0	0	0	13,89	0,70	3,4	12,43	3,12	93,96	304,24	79,82	30,54	22
12	0	0	0	13,54	0,79	4,9	12,50	3,26	66,75	157,36	42,85	33,96	25
13	0	0	0	16,00	0,75	4,0	13,42	3,42	116,13	247,06	106,41	29,57	18
14	0	0	0	13,89	0,47	3,9	11,78	3,22	88,34	199,95	90,43	33,11	24
15	0	0	0	14,60	0,77	3,1	15,88	3,57	67,71	309,95	38,45	34,63	24
16	1	0	0	13,19	0,51	1,8	4,05	1,16	204,73	440,78	127,05	26,54	20
17	1	0	0	12,14	0,36	1,8	1,74	0,66	211,77	215,62	91,32	26,61	22
18	1	0	0	9,32	0,38	1,3	2,14	0,79	194,94	53,50	138,07	25,82	28
19	1	0	0	8,97	0,44	2,1	2,29	0,89	200,08	14,59	136,39	25,22	28
20	1	0	0	9,67	0,38	1,6	2,12	0,92	215,09	34,41	64,72	28,12	29
21	1	0	0	10,02	0,66	2,5	13,81	1,69	204,64	475,94	81,72	38,45	38
22	1	0	0	12,49	1,58	3,6	13,34	2,06	275,15	958,23	69,46	39,03	31
23	1	0	0	12,49	1,22	3,4	16,68	1,58	319,55	376,44	54,02	37,12	30

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais												
	Vd1	Vd2	Vd3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
24	1	0	0	11,78	1,37	3,2	17,29	2,01	312,61	643,90	50,34	36,67	31
25	1	0	0	11,08	0,57	2,4	20,49	1,57	335,78	402,83	15,31	36,32	33
26	1	0	0	13,19	1,65	2,8	12,50	2,62	604,47	44,84	39,50	34,09	26
27	1	0	0	12,49	1,41	3,3	9,82	2,44	403,98	50,42	40,45	36,71	29
28	1	0	0	12,14	1,26	2,9	9,15	2,38	667,87	75,17	55,82	36,47	30
29	1	0	0	11,43	1,17	2,9	15,01	2,26	410,09	31,99	36,37	34,47	30
30	1	0	0	12,14	1,67	4,0	10,39	2,64	640,96	73,81	79,34	32,63	27
31	0	1	0	11,43	1,65	3,3	14,81	6,21	47,23	5,00	38,28	35,03	31
32	0	1	0	12,84	0,75	11,4	11,48	2,95	43,77	5,11	26,06	34,98	27
33	0	1	0	11,08	0,81	5,2	10,52	3,05	39,30	6,35	97,38	33,80	31
34	0	1	0	11,08	0,85	6,2	11,14	2,98	46,93	9,10	106,16	31,57	28
35	0	1	0	11,78	0,55	3,7	8,97	2,08	50,59	10,32	47,71	40,80	35
36	0	1	0	12,84	2,02	11,4	13,92	3,25	197,80	143,49	57,93	32,15	25
37	0	1	0	13,89	1,22	10,0	13,39	3,66	202,05	251,77	80,59	32,02	23
38	0	1	0	14,25	0,92	2,1	15,62	3,27	157,26	214,46	46,28	38,33	27
39	0	1	0	16,71	0,98	3,0	17,84	3,33	121,56	126,53	49,14	35,98	22
40	0	1	0	14,60	1,48	6,0	15,12	3,20	269,33	254,06	70,41	34,00	23
41	0	1	0	16,36	1,35	7,1	11,31	2,85	451,44	60,61	72,15	25,17	15
42	0	1	0	16,36	1,48	11,0	12,86	2,99	280,14	25,29	26,73	33,55	21
43	0	1	0	15,65	1,33	6,7	13,91	2,95	500,68	23,58	58,88	25,02	16
44	0	1	0	15,30	1,27	7,8	10,54	2,62	445,50	18,74	26,10	22,67	15
45	0	1	0	15,30	1,42	12,0	13,05	3,35	410,03	102,49	15,31	29,32	19
46	0	0	1	14,60	0,92	2,4	11,41	1,51	274,74	149,71	73,60	31,53	22

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais												
	Vd1	Vd2	Vd3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
<b>47</b>	0	0	1	13,54	0,85	6,8	10,36	2,03	186,97	16,07	45,22	35,35	26
<b>48</b>	0	0	1	15,30	0,81	4,5	15,53	2,06	222,52	20,26	71,31	32,57	21
<b>49</b>	0	0	1	16,36	1,03	2,4	11,09	1,44	445,75	265,66	51,31	35,46	22
<b>50</b>	0	0	1	15,30	0,98	5,8	12,40	1,80	281,36	222,12	50,23	24,51	16
<b>51</b>	0	0	1	8,27	0,47	3,9	15,97	3,56	108,89	18,67	25,94	33,74	41
<b>52</b>	0	0	1	13,89	0,21	5,1	13,75	3,72	201,82	29,38	25,32	33,63	24
<b>53</b>	0	0	1	8,62	0,60	7,0	12,25	2,84	76,14	10,02	74,69	36,66	43
<b>54</b>	0	0	1	8,62	0,38	8,9	13,59	4,16	147,26	10,50	14,25	30,96	36
<b>55</b>	0	0	1	8,62	1,20	1,7	14,72	2,78	312,63	293,77	112,66	31,05	36
<b>56</b>	0	0	1	13,89	2,30	4,0	14,02	2,96	102,77	647,85	23,11	35,18	25
<b>57</b>	0	0	1	14,60	5,14	8,3	12,86	4,19	459,28	338,31	96,06	26,13	18
<b>58</b>	0	0	1	19,17	2,88	5,2	8,85	2,77	94,55	540,50	76,04	28,99	15
<b>59</b>	0	0	1	18,12	2,34	2,5	12,68	2,32	169,35	389,45	76,03	29,29	16
<b>60</b>	0	0	1	19,17	2,08	3,5	18,36	2,36	251,97	168,17	49,79	26,40	14

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais												
	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26
<b>1</b>	66	26	23,6	18,5	19,32	4,8	0,36	0,33	0,69	0,79	7,49	11,5	63
<b>2</b>	63	27	21,2	18,3	22,38	5,6	0,35	0,21	0,56	0,99	7,13	3,1	97
<b>3</b>	61	28	20,3	18,3	26,88	6,7	0,34	0,26	0,60	1,00	7,28	5,0	79
<b>4</b>	62	26	18,9	17,5	12,10	3,0	0,37	0,18	0,55	1,09	7,44	3,5	116
<b>5</b>	61	30	18,9	19,2	34,78	8,7	0,41	0,24	0,65	0,89	6,88	0,9	238
<b>6</b>	64	15	30,0	20,3	11,24	2,8	0,32	0,31	0,63	0,83	6,22	2,5	143
<b>7</b>	67	14	27,4	19,4	30,39	7,6	0,34	0,29	0,63	0,88	6,50	7,2	219
<b>8</b>	66	14	27,5	19,7	6,21	1,6	0,30	0,40	0,70	0,71	6,21	1,7	195
<b>9</b>	63	13	31,2	20,2	21,60	5,4	0,35	0,30	0,65	0,87	5,18	3,1	238
<b>10</b>	68	14	30,1	20,1	20,76	5,2	0,26	0,36	0,62	0,91	5,01	1,4	176
<b>11</b>	55	34	15,0	16,7	16,45	4,1	0,31	0,19	0,50	1,16	8,04	10,1	115
<b>12</b>	61	30	16,7	16,6	20,33	5,1	0,33	0,13	0,46	1,16	7,73	5,8	104
<b>13</b>	53	35	16,7	17,1	21,12	5,3	0,27	0,27	0,54	1,04	7,57	2,5	94
<b>14</b>	60	32	16,2	17,0	32,58	8,1	0,28	0,24	0,52	1,02	7,58	3,0	80
<b>15</b>	62	34	17,3	17,1	27,34	6,8	0,29	0,12	0,41	1,27	7,65	7,9	94
<b>16</b>	48	33	22,9	17,0	42,67	10,7	0,37	0,17	0,54	1,11	4,17	0,0	49
<b>17</b>	48	33	21,7	16,9	24,20	6,1	0,36	0,11	0,47	1,23	4,37	0,3	26
<b>18</b>	46	33	22,5	17,1	27,83	7,0	0,33	0,22	0,55	1,08	4,58	1,6	50
<b>19</b>	45	30	23,9	17,4	22,55	5,6	0,34	0,26	0,60	0,94	5,13	3,6	73
<b>20</b>	51	31	22,4	16,9	14,53	3,6	0,34	0,22	0,56	1,02	4,81	1,7	37
<b>21</b>	69	34	14,3	15,5	22,64	5,7	0,23	0,09	0,32	1,43	5,48	0,4	176
<b>22</b>	70	33	14,9	15,5	23,39	5,8	0,37	0,04	0,40	1,20	5,47	0,2	86
<b>23</b>	67	33	14,6	15,7	16,38	4,1	0,35	0,06	0,42	1,36	5,51	0,1	103

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais												
	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26
24	66	37	14,6	15,8	24,41	6,1	0,39	0,02	0,41	1,34	6,21	0,0	186
25	65	49	14,1	15,7	34,91	8,7	0,31	0,25	0,56	1,00	6,20	0,0	161
26	61	18	19,3	16,5	15,25	3,8	0,34	0,26	0,59	0,98	4,64	0,6	52
27	66	18	21,7	16,5	20,83	5,2	0,28	0,31	0,59	0,95	5,56	1,7	119
28	66	16	21,6	16,8	26,60	6,7	0,28	0,23	0,51	1,07	4,97	1,3	75
29	62	21	22,2	16,8	28,60	7,2	0,28	0,24	0,52	1,11	5,72	1,5	81
30	59	20	20,6	16,8	23,44	5,9	0,29	0,29	0,57	1,03	5,02	1,7	80
31	63	24	30,8	19,5	26,82	6,7	0,41	0,29	0,71	0,72	6,06	6,3	49
32	63	19	33,4	18,1	13,95	3,5	0,48	0,18	0,66	0,81	6,50	1,8	152
33	61	22	30,5	19,4	18,81	4,7	0,43	0,26	0,68	0,74	6,02	0,5	69
34	57	19	29,1	20,4	22,66	5,7	0,39	0,31	0,70	0,77	5,86	2,3	58
35	73	19	30,0	20,2	21,48	5,4	0,41	0,25	0,66	0,77	5,26	2,0	62
36	58	56	17,0	16,0	44,45	11,1	0,30	0,18	0,48	1,22	6,89	32,0	338
37	58	62	15,1	16,4	18,91	4,7	0,33	0,20	0,52	1,19	6,26	11,5	121
38	69	56	13,8	15,8	9,83	2,5	0,34	0,11	0,45	1,25	7,08	64,5	94
39	65	57	14,7	16,2	14,39	3,6	0,33	0,19	0,52	1,09	6,65	23,2	98
40	61	49	16,5	16,5	20,60	5,2	0,32	0,24	0,56	1,01	6,22	9,7	59
41	45	34	23,1	16,5	19,03	4,8	0,25	0,46	0,71	0,78	7,10	4,9	160
42	60	25	25,6	17,6	23,00	5,8	0,27	0,28	0,54	1,02	7,15	2,8	207
43	45	24	27,4	17,6	23,42	5,9	0,29	0,33	0,62	0,96	7,39	6,9	123
44	41	21	25,1	17,8	120,91	30,2	0,32	0,28	0,60	0,95	7,47	11,5	116
45	53	24	24,5	17,7	16,46	4,1	0,31	0,36	0,66	0,87	7,31	5,3	106
46	57	27	24,9	16,9	56,68	14,2	0,31	0,31	0,63	0,96	6,25	0,9	59

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais												
	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26
<b>47</b>	64	22	26,0	16,8	31,83	8,0	0,26	0,32	0,57	1,04	7,05	1,5	120
<b>48</b>	59	24	27,4	18,3	47,28	11,8	0,28	0,28	0,56	1,03	6,52	7,5	119
<b>49</b>	64	28	24,6	17,0	63,45	15,9	0,30	0,22	0,52	1,10	6,47	11,3	187
<b>50</b>	44	29	25,7	17,7	49,11	12,3	0,26	0,31	0,56	1,05	5,73	5,3	105
<b>51</b>	61	17	21,7	16,9	13,82	3,5	0,36	0,28	0,65	0,81	6,63	2,3	88
<b>52</b>	61	15	23,2	18,6	8,25	2,1	0,39	0,24	0,63	0,93	6,09	0,7	175
<b>53</b>	66	15	23,8	19,4	9,22	2,3	0,41	0,21	0,62	0,96	7,55	2,7	131
<b>54</b>	56	20	23,7	17,8	10,55	2,6	0,42	0,21	0,63	0,92	6,27	0,5	68
<b>55</b>	56	20	23,6	16,6	5,86	1,5	0,45	0,2	0,65	0,83	6,37	1,4	105
<b>56</b>	63	47	14,6	15,6	17,50	4,4	0,29	0,2	0,49	0,96	6,72	6,7	203
<b>57</b>	47	48	15,2	17,2	44,91	11,2	0,29	0,35	0,64	0,86	7,53	10,6	982
<b>58</b>	52	41	14,2	16,1	20,52	5,1	0,38	0,13	0,51	1,12	6,86	16,4	210
<b>59</b>	53	43	16,5	16,0	22,36	5,6	0,33	0,34	0,67	0,79	7,2	5,3	298
<b>60</b>	48	52	15,6	15,1	37,92	9,5	0,35	0,19	0,54	1,07	6,87	3,8	244

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais									
	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36
1	8,10	2,63	0,00	10,89	10,89	13,39	81,3	0,0	10,56	23,7
2	5,19	4,91	0,00	10,35	10,35	14,05	73,7	0,0	9,92	20,7
3	7,60	4,02	0,00	11,82	11,82	14,72	80,3	0,0	10,88	24,2
4	6,83	5,06	0,00	12,19	12,19	15,09	80,8	0,0	9,60	21,5
5	4,74	3,59	0,00	8,94	8,94	13,84	64,6	0,0	9,60	21,1
6	3,06	2,17	0,00	5,60	5,60	11,20	50,0	0,0	5,25	24,0
7	3,52	2,25	0,00	6,33	6,33	10,93	57,9	0,0	6,14	27,9
8	2,91	1,75	0,00	5,16	5,16	11,06	46,7	0,0	5,37	28,2
9	2,38	1,11	0,10	4,10	4,20	11,90	34,5	2,4	5,12	30,1
10	1,09	0,60	0,96	2,14	3,10	12,84	16,7	31,0	5,50	25,5
11	9,77	1,15	0,00	11,21	11,21	11,21	100,0	0,0	6,78	16,7
12	8,21	2,07	0,00	10,55	10,55	11,35	93,0	0,0	6,40	24,7
13	7,04	2,94	0,00	10,22	10,22	11,32	90,3	0,0	6,40	25,9
14	4,95	1,96	0,00	7,11	7,11	7,61	93,4	0,0	4,86	32,9
15	8,44	2,13	0,00	10,81	10,81	12,11	89,3	0,0	5,89	28,6
16	0,33	0,11	1,83	0,57	2,40	18,07	3,2	76,3	9,98	14,1
17	0,11	0,02	1,35	0,20	1,55	14,04	1,4	87,1	8,70	12,4
18	0,14	0,09	1,64	0,36	2,00	18,96	1,9	82,0	11,39	14,1
19	0,14	0,07	1,45	0,40	1,85	14,70	2,7	78,4	8,70	17,9
20	0,18	0,05	1,45	0,32	1,77	12,92	2,5	81,9	6,65	22,2
21	5,40	1,42	0,00	7,27	7,27	16,77	43,4	0,0	8,96	22,1
22	3,40	0,74	0,10	4,36	4,36	15,06	29,0	2,2	9,28	17,9
23	5,40	1,14	0,00	6,80	6,80	16,30	41,7	0,0	9,92	22,3

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais									
	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36
24	5,36	1,47	0,00	7,31	7,31	12,11	60,4	0,0	8,32	32,2
25	11,92	1,95	0,00	14,28	14,28	19,68	72,6	0,0	12,15	35,4
26	1,43	0,33	0,67	0,89	1,56	9,59	9,3	42,9	3,97	22,4
27	1,14	1,05	0,19	2,49	2,68	9,54	26,2	7,1	5,37	29,5
28	0,94	0,56	0,39	1,69	2,08	10,39	16,3	18,8	4,73	25,1
29	2,52	1,59	0,00	4,32	4,32	13,02	33,2	0,0	7,55	32,3
30	1,37	0,72	0,48	2,29	2,77	12,99	17,6	17,3	5,37	20,8
31	3,47	1,35	0,10	4,95	5,05	14,35	34,5	2,0	7,68	17,6
32	4,53	1,51	0,00	6,43	6,43	13,13	49,0	0,0	7,16	20,1
33	2,46	0,87	0,00	3,51	3,51	11,91	29,5	0,0	5,37	18,3
34	2,54	0,79	0,00	3,48	3,48	13,78	25,3	0,0	7,04	17,5
35	0,97	0,31	0,67	1,44	2,11	13,84	10,4	31,8	6,53	15,6
36	8,57	2,19	0,00	11,62	11,62	15,72	73,9	0,0	11,83	26,9
37	7,10	1,51	0,00	8,92	8,92	16,52	54,0	0,0	12,15	20,7
38	12,62	3,74	0,00	16,6	16,6	19,90	83,4	0,0	14,39	24,0
39	10,46	2,63	0,00	13,34	13,34	14,94	89,3	0,0	14,07	24,2
40	11,96	2,02	0,00	14,13	14,13	20,83	67,8	0,0	14,71	24,2
41	7,17	1,52	0,00	9,10	9,10	10,80	84,3	0,0	6,01	29,8
42	5,37	1,56	0,00	7,46	7,46	7,46	100,0	0,0	5,50	35,4
43	6,89	0,95	0,00	8,15	8,15	9,85	82,7	0,0	4,61	31,9
44	7,43	1,56	0,00	9,29	9,29	11,09	83,8	0,0	6,40	38,9
45	6,35	1,12	0,00	7,74	7,74	9,14	84,7	0,0	4,22	36,0
46	5,27	1,79	0,00	7,21	7,21	12,01	60,0	0,0	5,76	27,6

**APÊNDICE A – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período seco.**

Amostra	Covariáveis ambientais									
	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36
47	5,33	0,93	0,00	6,57	6,57	10,27	64,0	0,0	4,73	20,1
48	7,09	1,82	0,00	10,02	10,02	15,22	65,8	0,0	7,04	23,7
49	5,37	1,33	0,00	7,18	7,18	12,28	58,5	0,0	6,53	30,8
50	7,09	1,65	0,00	9,01	9,01	16,21	55,6	0,0	8,57	28,6
51	4,51	1,61	0,00	6,35	6,35	11,25	56,4	0,00	5,50	23,5
52	2,39	1,09	0,00	3,93	3,93	10,33	38,0	0,00	4,09	18,7
53	6,93	1,82	0,00	9,09	9,09	11,29	80,5	0,00	5,12	23,8
54	4,18	1,34	0,00	5,69	5,69	11,89	47,9	0,00	5,76	23,0
55	3,98	1,83	0,00	6,08	6,08	13,58	44,8	0,00	6,40	19,7
56	9,21	3,17	0,00	12,90	12,90	18,60	69,4	0,00	14,39	20,2
57	10,03	3,90	0,00	16,44	16,44	19,44	84,6	0,00	11,51	18,6
58	8,02	2,36	0,00	10,92	10,92	17,62	62,0	0,00	11,51	19,6
59	8,26	3,11	0,00	12,13	12,13	17,03	71,2	0,00	12,79	21,9
60	6,71	2,47	0,00	9,80	9,80	17,60	55,7	0,00	13,43	17,7

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais												
	Vd1	Vd2	Vd3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
1	0	0	0	16,78	0,83	1,4	4,23	2,19	37,95	1,33	20,03	34,15	20
2	0	0	0	23,11	0,70	1,4	3,91	2,17	24,00	5,46	17,39	37,38	16
3	0	0	0	30,14	0,85	2,2	4,19	2,20	25,51	90,79	18,08	39,54	13
4	0	0	0	20,30	0,89	2,3	3,51	1,76	15,71	1,47	13,49	36,05	18
5	0	0	0	21,70	0,68	1,4	1,32	0,41	81,45	363,70	45,57	40,56	19
6	0	0	0	15,02	1,61	6,1	8,28	3,19	229,06	6,42	12,22	26,01	17
7	0	0	0	19,24	1,63	5,4	6,07	2,41	237,14	25,96	220,43	22,53	12
8	0	0	0	15,02	1,70	8,0	7,43	2,99	208,00	13,00	80,76	33,97	23
9	0	0	0	15,02	1,78	14,4	7,83	3,16	213,51	8,05	50,51	34,35	23
10	0	0	0	15,37	1,65	6,1	7,52	3,19	153,46	10,54	39,88	34,17	22
11	0	0	0	17,83	1,91	3,7	13,70	3,52	100,74	91,86	42,75	39,35	22
12	0	0	0	18,54	2,10	2,2	14,68	3,40	109,00	108,00	80,41	37,27	20
13	0	0	0	15,72	1,97	1,8	11,28	2,33	114,03	91,83	186,21	33,22	21
14	0	0	0	17,83	2,23	1,8	16,15	3,65	112,53	137,13	40,27	42,21	24
15	0	0	0	20,30	2,02	1,6	17,58	4,11	108,96	109,04	52,39	37,09	18
16	1	0	0	15,37	0,33	0,4	0,98	0,29	44,49	246,71	54,18	39,47	26
17	1	0	0	17,13	0,40	0,3	1,34	0,37	94,99	289,71	34,78	31,76	19
18	1	0	0	19,59	0,42	0,4	1,15	0,35	74,37	534,86	27,00	37,34	19
19	1	0	0	18,89	0,36	0,3	1,11	0,24	85,89	153,09	36,12	38,63	20
20	1	0	0	14,32	0,42	0,2	0,05	0,02	15,63	0,00	0,12	39,84	28
21	1	0	0	7,98	1,07	1,5	14,84	1,80	192,73	562,44	47,18	39,65	50
22	1	0	0	13,96	1,30	1,4	19,47	1,81	214,34	352,93	28,77	36,88	26
23	1	0	0	13,26	1,22	1,1	19,08	1,85	252,17	517,86	19,29	34,57	26

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais												
	Vd1	Vd2	Vd3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
24	1	0	0	7,63	0,92	1,4	10,10	1,48	279,24	602,20	120,88	33,62	44
25	1	0	0	9,74	1,00	1,0	19,03	1,56	226,61	317,01	12,95	37,13	38
26	1	0	0	13,26	1,26	1,4	10,74	2,39	649,97	755,77	39,76	38,64	29
27	1	0	0	13,61	1,17	3,3	9,82	2,21	374,04	90,04	42,33	44,52	33
28	1	0	0	12,21	1,20	1,3	6,87	2,12	660,12	973,14	67,84	32,26	26
29	1	0	0	19,24	1,71	1,0	10,09	2,11	645,53	534,10	113,13	38,25	20
30	1	0	0	19,24	1,48	1,5	12,13	2,47	549,34	447,94	21,97	34,46	18
31	0	1	0	16,43	1,37	0,4	11,77	2,20	68,22	29,88	97,90	34,31	21
32	0	1	0	14,32	1,09	1,4	10,71	1,96	51,31	3,75	21,37	41,31	29
33	0	1	0	16,78	1,56	0,9	15,05	2,76	51,67	13,32	59,48	37,19	22
34	0	1	0	16,07	1,71	1,0	13,51	2,43	39,97	7,12	68,35	37,50	23
35	0	1	0	13,96	1,46	2,0	10,35	2,13	38,13	6,48	116,87	31,50	23
36	0	1	0	11,50	3,24	7,1	9,41	2,41	209,10	118,82	100,32	34,29	30
37	0	1	0	11,50	2,60	6,1	13,00	2,96	138,02	173,92	32,82	39,76	35
38	0	1	0	12,56	2,38	4,9	7,63	1,92	177,33	295,58	87,91	33,13	26
39	0	1	0	12,91	2,73	9,4	6,89	2,12	187,16	347,51	87,05	34,14	26
40	0	1	0	9,74	3,63	7,8	9,14	2,30	347,47	328,46	179,49	26,94	28
41	0	1	0	17,48	2,55	4,4	11,23	2,83	539,73	28,42	40,46	36,31	21
42	0	1	0	22,05	2,71	8,1	9,96	2,58	395,45	15,03	13,77	36,46	17
43	0	1	0	15,37	2,92	6,2	11,68	2,73	375,54	41,63	39,80	37,25	24
44	0	1	0	13,61	2,53	6,2	10,24	3,09	281,29	33,68	54,11	39,79	29
45	0	1	0	16,78	2,81	9,8	10,60	3,34	248,74	20,13	64,31	34,19	20
46	0	0	1	18,89	1,50	2,1	13,13	2,49	138,00	218,00	71,11	39,27	21

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais												
	Vd1	Vd2	Vd3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
47	0	0	1	19,24	1,33	0,8	12,07	2,00	133,12	221,59	113,42	35,29	18
48	0	0	1	12,21	1,20	1,6	12,18	2,30	90,48	483,31	55,48	40,30	33
49	0	0	1	17,13	1,39	1,1	12,91	2,35	142,18	99,97	62,35	40,05	23
50	0	0	1	17,83	1,48	2,0	15,38	3,30	185,09	66,78	53,20	37,92	21
51	0	0	1	24,17	1,78	1,4	10,87	1,52	272,93	122,96	75,19	32,52	13
52	0	0	1	18,19	1,07	2,5	9,83	1,18	147,17	89,42	46,73	36,21	20
53	0	0	1	21,35	1,69	2,3	10,48	1,55	204,80	55,67	85,96	32,17	15
54	0	0	1	20,65	1,24	2,7	10,26	1,44	138,73	59,10	13,74	37,36	18
55	0	0	1	22,76	1,33	4,1	11,78	1,90	224,21	22,18	50,29	37,46	16
56	0	0	1	27,68	2,32	4,9	13,67	3,03	146,79	188,09	21,18	40,40	15
57	0	0	1	27,68	2,40	3,1	12,82	2,30	315,63	274,14	58,43	38,31	14
58	0	0	1	24,17	2,02	5,4	12,86	1,93	325,38	148,51	24,97	36,56	15
59	0	0	1	23,11	2,45	1,7	14,67	2,97	278,17	292,84	87,05	32,43	14
60	0	0	1	23,11	3,91	7,1	11,97	2,50	112,04	104,61	53,64	34,64	15

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais												
	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26
<b>1</b>	61	33	27,4	19,2	15,38	3,8	0,36	0,33	0,69	0,79	7,49	11,5	63
<b>2</b>	67	36	28,4	19,0	22,24	5,6	0,35	0,21	0,56	0,99	7,13	3,1	97
<b>3</b>	71	47	29,8	19,0	25,72	6,4	0,34	0,26	0,60	1,00	7,28	5,0	79
<b>4</b>	65	33	28,5	18,5	44,43	11,1	0,37	0,18	0,55	1,09	7,44	3,5	116
<b>5</b>	73	38	28,9	19,9	26,49	6,6	0,41	0,24	0,65	0,89	6,88	0,9	238
<b>6</b>	47	23	27,0	22,8	12,21	3,1	0,32	0,31	0,63	0,83	6,22	2,5	143
<b>7</b>	41	24	29,5	22,3	18,57	4,6	0,34	0,29	0,63	0,88	6,50	7,2	219
<b>8</b>	61	25	26,5	21,8	12,26	3,1	0,30	0,40	0,70	0,71	6,21	1,7	195
<b>9</b>	62	24	27,6	22,8	13,69	3,4	0,35	0,30	0,65	0,87	5,18	3,1	238
<b>10</b>	62	24	27,5	23,3	37,72	9,4	0,26	0,36	0,62	0,91	5,01	1,4	176
<b>11</b>	71	29	32,1	22,6	14,93	3,7	0,31	0,19	0,50	1,16	8,04	10,1	115
<b>12</b>	67	34	32,4	22,5	27,34	6,8	0,33	0,13	0,46	1,16	7,73	5,8	104
<b>13</b>	60	30	31,9	22,5	16,26	4,1	0,27	0,27	0,54	1,04	7,57	2,5	94
<b>14</b>	76	30	28,5	21,6	18,27	4,6	0,28	0,24	0,52	1,02	7,58	3,0	80
<b>15</b>	67	28	29,5	22,6	20,96	5,2	0,29	0,12	0,41	1,27	7,65	7,9	94
<b>16</b>	71	49	22,0	18,0	54,72	13,7	0,37	0,17	0,54	1,11	4,17	0,0	49
<b>17</b>	57	55	22,2	17,9	40,85	10,2	0,36	0,11	0,47	1,23	4,37	0,3	26
<b>18</b>	67	28	21,3	18,2	32,83	8,2	0,33	0,22	0,55	1,08	4,58	1,6	50
<b>19</b>	70	46	22,0	18,6	77,01	19,3	0,34	0,26	0,60	0,94	5,13	3,6	73
<b>20</b>	72	49	20,4	18,3	25,86	6,5	0,34	0,22	0,56	1,02	4,81	1,7	37
<b>21</b>	71	49	22,7	18,3	38,28	9,6	0,23	0,09	0,32	1,43	5,48	0,4	176
<b>22</b>	66	38	21,3	18,8	29,65	7,4	0,37	0,04	0,40	1,20	5,47	0,2	86
<b>23</b>	62	42	21,5	18,9	28,52	7,1	0,35	0,06	0,42	1,36	5,51	0,1	103

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais												
	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26
24	61	40	22,2	18,7	51,45	12,9	0,39	0,02	0,41	1,34	6,21	0,0	186
25	67	48	20,6	18,6	62,25	15,6	0,31	0,25	0,56	1,00	6,20	0,0	161
26	70	23	25,1	21,6	61,43	15,4	0,34	0,26	0,59	0,98	4,64	0,6	52
27	80	25	24,1	20,6	74,48	18,6	0,28	0,31	0,59	0,95	5,56	1,7	119
28	58	31	25,5	21,2	81,61	20,4	0,28	0,23	0,51	1,07	4,97	1,3	75
29	69	34	25,8	21,1	67,11	16,8	0,28	0,24	0,52	1,11	5,72	1,5	81
30	62	23	23,8	21,2	39,51	9,9	0,29	0,29	0,57	1,03	5,02	1,7	80
31	62	36	32,8	22,3	45,72	11,4	0,41	0,29	0,71	0,72	6,06	6,3	49
32	74	30	27,1	20,3	24,73	6,2	0,48	0,18	0,66	0,81	6,50	1,8	152
33	67	30	28,3	21,5	57,88	14,5	0,43	0,26	0,68	0,74	6,02	0,5	69
34	67	30	28,4	21,8	29,79	7,4	0,39	0,31	0,70	0,77	5,86	2,3	58
35	57	24	27,9	21,9	22,97	5,7	0,41	0,25	0,66	0,77	5,26	2,0	62
36	62	52	31,2	19,1	15,89	4,0	0,30	0,18	0,48	1,22	6,89	32,0	338
37	72	40	22,1	18,6	29,05	7,3	0,33	0,20	0,52	1,19	6,26	11,5	121
38	60	48	23,3	19,0	51,24	12,8	0,34	0,11	0,45	1,25	7,08	64,5	94
39	61	37	28,0	19,3	46,87	11,7	0,33	0,19	0,52	1,09	6,65	23,2	98
40	48	47	25,6	19,3	55,45	13,9	0,32	0,24	0,56	1,01	6,22	9,7	59
41	65	28	26,3	24,3	16,66	4,2	0,25	0,46	0,71	0,78	7,10	4,9	160
42	66	29	25,6	21,9	31,49	7,9	0,27	0,28	0,54	1,02	7,15	2,8	207
43	67	30	26,3	23,0	42,71	10,7	0,29	0,33	0,62	0,96	7,39	6,9	123
44	72	31	25,8	21,2	23,47	5,9	0,32	0,28	0,60	0,95	7,47	11,5	116
45	62	36	26,0	21,2	26,28	6,6	0,31	0,36	0,66	0,87	7,31	5,3	106
46	71	28	25,0	21,0	48,81	12,2	0,36	0,28	0,65	0,81	6,63	2,3	88

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais												
	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26
47	64	29	24,6	20,5	64,35	16,1	0,39	0,24	0,63	0,93	6,09	0,7	175
48	73	31	25,5	21,2	62,18	15,5	0,41	0,21	0,62	0,96	7,55	2,7	131
49	72	32	25,1	20,4	28,59	7,1	0,42	0,21	0,63	0,92	6,27	0,5	68
50	68	28	25,9	20,6	28,70	7,2	0,45	0,20	0,65	0,83	6,37	1,4	105
51	59	29	28,3	20,5	67,75	16,9	0,31	0,31	0,63	0,96	6,25	0,9	59
52	65	29	24,3	21,7	34,68	8,7	0,26	0,32	0,57	1,04	7,05	1,5	120
53	58	34	24,8	21,2	28,72	7,2	0,28	0,28	0,56	1,03	6,52	7,5	119
54	67	33	24,9	21,4	48,14	12,0	0,30	0,22	0,52	1,10	6,47	11,3	187
55	67	33	25,2	20,9	15,47	3,9	0,26	0,31	0,56	1,05	5,73	5,3	105
56	73	52	23,5	18,4	20,78	5,2	0,29	0,35	0,64	0,86	7,53	10,6	982
57	69	48	24,0	18,8	32,94	8,2	0,38	0,13	0,51	1,12	6,86	16,4	210
58	66	51	24,9	18,6	41,49	10,4	0,29	0,20	0,49	0,96	6,72	6,7	203
59	58	40	27,5	19,2	44,82	11,2	0,33	0,34	0,67	0,79	7,20	5,3	298
60	62	43	25,8	18,2	28,65	7,2	0,35	0,19	0,54	1,07	6,87	3,8	244

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais									
	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36
<b>1</b>	8,10	2,63	0,00	10,89	10,89	13,39	81,3	0,0	10,56	23,7
<b>2</b>	5,19	4,91	0,00	10,35	10,35	14,05	73,7	0,0	9,92	20,7
<b>3</b>	7,60	4,02	0,00	11,82	11,82	14,72	80,3	0,0	10,88	24,2
<b>4</b>	6,83	5,06	0,00	12,19	12,19	15,09	80,8	0,0	9,60	21,5
<b>5</b>	4,74	3,59	0,00	8,94	8,94	13,84	64,6	0,0	9,60	21,1
<b>6</b>	3,06	2,17	0,00	5,60	5,60	11,20	50,0	0,0	5,25	24,0
<b>7</b>	3,52	2,25	0,00	6,33	6,33	10,93	57,9	0,0	6,14	27,9
<b>8</b>	2,91	1,75	0,00	5,16	5,16	11,06	46,7	0,0	5,37	28,2
<b>9</b>	2,38	1,11	0,10	4,10	4,20	11,90	34,5	2,4	5,12	30,1
<b>10</b>	1,09	0,60	0,96	2,14	3,10	12,84	16,7	31,0	5,50	25,5
<b>11</b>	9,77	1,15	0,00	11,21	11,21	11,21	100,0	0,0	6,78	16,7
<b>12</b>	8,21	2,07	0,00	10,55	10,55	11,35	93,0	0,0	6,40	24,7
<b>13</b>	7,04	2,94	0,00	10,22	10,22	11,32	90,3	0,0	6,40	25,9
<b>14</b>	4,95	1,96	0,00	7,11	7,11	7,61	93,4	0,0	4,86	32,9
<b>15</b>	8,44	2,13	0,00	10,81	10,81	12,11	89,3	0,0	5,89	28,6
<b>16</b>	0,33	0,11	1,83	0,57	2,40	18,07	3,2	76,3	9,98	14,1
<b>17</b>	0,11	0,02	1,35	0,20	1,55	14,04	1,4	87,1	8,70	12,4
<b>18</b>	0,14	0,09	1,64	0,36	2,00	18,96	1,9	82,0	11,39	14,1
<b>19</b>	0,14	0,07	1,45	0,40	1,85	14,70	2,7	78,4	8,70	17,9
<b>20</b>	0,18	0,05	1,45	0,32	1,77	12,92	2,5	81,9	6,65	22,2
<b>21</b>	5,40	1,42	0,00	7,27	7,27	16,77	43,4	0,0	8,96	22,1
<b>22</b>	3,40	0,74	0,10	4,36	4,36	15,06	29,0	2,2	9,28	17,9
<b>23</b>	5,40	1,14	0,00	6,80	6,80	16,30	41,7	0,0	9,92	22,3

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais									
	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36
24	5,36	1,47	0,00	7,31	7,31	12,11	60,4	0,0	8,32	32,2
25	11,92	1,95	0,00	14,28	14,28	19,68	72,6	0,0	12,15	35,4
26	1,43	0,33	0,67	0,89	1,56	9,59	9,3	42,9	3,97	22,4
27	1,14	1,05	0,19	2,49	2,68	9,54	26,2	7,1	5,37	29,5
28	0,94	0,56	0,39	1,69	2,08	10,39	16,3	18,8	4,73	25,1
29	2,52	1,59	0,00	4,32	4,32	13,02	33,2	0,0	7,55	32,3
30	1,37	0,72	0,48	2,29	2,77	12,99	17,6	17,3	5,37	20,8
31	3,47	1,35	0,10	4,95	5,05	14,35	34,5	2,0	7,68	17,6
32	4,53	1,51	0,00	6,43	6,43	13,13	49,0	0,0	7,16	20,1
33	2,46	0,87	0,00	3,51	3,51	11,91	29,5	0,0	5,37	18,3
34	2,54	0,79	0,00	3,48	3,48	13,78	25,3	0,0	7,04	17,5
35	0,97	0,31	0,67	1,44	2,11	13,84	10,4	31,8	6,53	15,6
36	8,57	2,19	0,00	11,62	11,62	15,72	73,9	0,0	11,83	26,9
37	7,10	1,51	0,00	8,92	8,92	16,52	54,0	0,0	12,15	20,7
38	12,62	3,74	0,00	16,60	16,60	19,90	83,4	0,0	14,39	24,0
39	10,46	2,63	0,00	13,34	13,34	14,94	89,3	0,0	14,07	24,2
40	11,96	2,02	0,00	14,13	14,13	20,83	67,8	0,0	14,71	24,2
41	7,17	1,52	0,00	9,10	9,10	10,80	84,3	0,0	6,01	29,8
42	5,37	1,56	0,00	7,46	7,46	7,46	100,0	0,0	5,50	35,4
43	6,89	0,95	0,00	8,15	8,15	9,85	82,7	0,0	4,61	31,9
44	7,43	1,56	0,00	9,29	9,29	11,09	83,8	0,0	6,40	38,9
45	6,35	1,12	0,00	7,74	7,74	9,14	84,7	0,0	4,22	36,0
46	4,51	1,61	0,00	6,35	6,35	11,25	56,4	0,0	5,50	23,5

**APÊNDICE B – Matriz de dados brutos tendo como covariáveis ambientais as características físicas e químicas da serapilheira e do solo, nas 60 amostras coletadas em período chuvoso.**

Amostra	Covariáveis ambientais									
	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36
47	2,39	1,09	0,00	3,93	3,93	10,33	38,0	0,0	4,09	18,7
48	6,93	1,82	0,00	9,09	9,09	11,29	80,5	0,0	5,12	23,8
49	4,18	1,34	0,00	5,69	5,69	11,89	47,9	0,0	5,76	23,0
50	3,98	1,83	0,00	6,08	6,08	13,58	44,8	0,0	6,40	19,7
51	5,27	1,79	0,00	7,21	7,21	12,01	60,0	0,0	5,76	27,6
52	5,33	0,93	0,00	6,57	6,57	10,27	64,0	0,0	4,73	20,1
53	7,09	1,82	0,00	10,02	10,02	15,22	65,8	0,0	7,04	23,7
54	5,37	1,33	0,00	7,18	7,18	12,28	58,5	0,0	6,53	30,8
55	7,09	1,65	0,00	9,01	9,01	16,21	55,6	0,0	8,57	28,6
56	10,03	3,90	0,00	16,44	16,44	19,44	84,6	0,0	11,51	18,6
57	8,02	2,36	0,00	10,92	10,92	17,62	62,0	0,0	11,51	19,6
58	9,21	3,17	0,00	12,90	12,90	18,60	69,4	0,0	14,39	20,2
59	8,26	3,11	0,00	12,13	12,13	17,03	71,2	0,0	12,79	21,9
60	6,71	2,47	0,00	9,80	9,80	17,60	55,7	0,0	13,43	17,7