



ALLAN ALVES FERNANDES

**SENSIBILIDADE DAS OBSERVAÇÕES EM
COMPONENTES PRINCIPAIS: UM ESTUDO
DOS DADOS COMPORTAMENTAIS EM UM
SISTEMA DE TRABALHO NA CAFEICULTURA**

LAVRAS - MG

2014

ALLAN ALVES FERNANDES

**SENSIBILIDADE DAS OBSERVAÇÕES EM COMPONENTES
PRINCIPAIS: UM ESTUDO DOS DADOS COMPORTAMENTAIS EM UM
SISTEMA DE TRABALHO NA CAFEICULTURA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador

Dr. Renato Ribeiro de Lima

Coorientador

Dr. Marcelo Ângelo Cirillo

LAVRAS - MG

2014

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Fernandes, Allan Alves.

Sensibilidade das observações em componentes principais : um estudo dos dados comportamentais em um sistema de trabalho na cafeicultura / Allan Alves Fernandes. – Lavras : UFLA, 2014.

83 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2014.

Orientador: Renato Ribeiro de Lima.

Bibliografia.

1. Análise multivariada. 2. Análise de componentes principais. 3. Sensibilidade. 4. Cafeicultura. 5. Outliers. I. Universidade Federal de Lavras. II. Título.

CDD – 519.535

ALLAN ALVES FERNANDES

**SENSIBILIDADE DAS OBSERVAÇÕES EM COMPONENTES
PRINCIPAIS: UM ESTUDO DOS DADOS COMPORTAMENTAIS EM UM
SISTEMA DE TRABALHO NA CAFEICULTURA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 18 de fevereiro de 2014.

Dr. Marcelo Ângelo Cirillo UFLA

Dr. Marco Antônio Gomes Barbosa UFLA

Dra. Thelma Sáfadi UFLA

Dr. Renato Ribeiro de Lima
Orientador

LAVRAS - MG

2014

A Deus,
à minha família,
pelo amor incondicional.
DEDICO.

AGRADECIMENTOS

A Deus, por sua constante presença em minha vida, me fazendo alcançar lugares onde sozinho jamais chegaria.

Aos meus pais, Pedro Alves Fernandes e Marlene Maria Félix Fernandes, pelos ensinamentos, carinho, apoio e força para concretização de meus sonhos.

A meu irmão, Alysson Alves Fernandes, pela amizade e companheirismo.

Aos meus avós, Francisco Alves do Couto, in memoriam, e Maria Azarias Fernandes, pelo incentivo e força.

Aos professores Dr. Renato Ribeiro de Lima e Dr. Marcelo Ângelo Cirillo, pela orientação, paciência e compreensão para a concretização deste trabalho.

Ao Dr. Marco Antônio Gomes Barbosa e a Dra. Thelma Sáfadi, por terem aceitado o convite para participar da banca e contribuir de forma significativa com meu trabalho.

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária da UFLA, pelos conhecimentos transmitidos.

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária.

À Layla, por estar comigo em todos os momentos, sejam eles bons ou ruins.

Aos amigos do curso de mestrado pelo apoio e amizade, em especial à Laís, Érica, Tamara e Filipe (Carioca).

Aos funcionários do Departamento de Ciências Exatas da UFLA (DEX), em especial à Josi, secretária da Pós-Graduação.

Ao CNPq, pela concessão da bolsa de estudos.

Por fim, a todos que contribuíram de forma direta ou indireta para execução desta dissertação, muito obrigado!

“Nenhum mal te sucederá, nem praga alguma chegará à sua tenda.
Porque aos seus anjos dará ordem a teu respeito, para te guardarem em
todos os teus caminhos.”

Salmo 91:10-11 – Bíblia Sagrada

RESUMO

Em se tratando do estudo de dados comportamentais, em geral, a inferência estatística utilizada é usual, como por exemplo, análise de variância e testes de hipótese. Especificamente, a aplicação de técnicas multivariadas a esse tipo de dados, resume-se, basicamente, na utilização dos componentes principais, no entanto, o estudo da sensibilidade das observações na formação desses componentes é pouco explorado. No que tange à aplicabilidade no estudo comportamental de trabalhadores na cafeicultura, em lavouras de café do Sul de Minas Gerais, não é encontrado nenhum estudo do tipo na literatura. Diante disso, o presente trabalho foi conduzido com o objetivo de aplicar a análise de componentes principais, realizando um procedimento para identificar a influência que cada observação exerce em relação à formação desses componentes, de modo a priorizar uma interpretação contextualizada ao sistema de trabalho, associado às combinações posturais predefinidas durante atividades na lavoura. Concluiu-se, ao final deste estudo, que pela metodologia proposta permitiu-se identificar as observações discrepantes, cuja interpretação revelou as posturas que exerceram maior influência nas atividades realizadas por cada trabalhador, uma vez que os trabalhadores foram considerados como variáveis e as posturas como unidades amostrais na organização dos dados.

Palavras-chave: Análise multivariada. Cafeicultura. Componentes principais. Outliers. Sensibilidade.

ABSTRACT

Regarding the study of behavioral data, the statistical inference used is usual, such as the analysis of variance and the hypothesis tests. Specifically, the application of multivariate techniques to this type of data is basically summarized in the use of the principal components. However, the observation sensitivity study in the formation of these components is little explored. In regard to the applicability on the behavioral study of coffee culture workers, in coffee plantations in southern Minas Gerais, Brazil, no study of this type is found in literature. Thus, the present work was conducted with the objective of applying the principal component analysis, performing a procedure to identify the influence each observation exercises in relation to the formation of these components, in order to prioritize a contextualized interpretation to the working system, associated to the pre-defined postural combinations during the activities in the crop. At the end of this study, it was concluded that the proposed methodology allowed the identification of discrepant observations, of which the interpretation revealed the postures which exercise the most influence in the activities performed by each worker, since they were considered as variables and the postures as sample units in the data organization.

Keywords: Multivariate analysis. Coffee culture. Principal components. Outliers. Sensitivity.

LISTA DE TABELAS

Tabela 1	Descrição das 14 posturas selecionadas.....	42
Tabela 2	Correlações entre as variáveis*	47
Tabela 3	Autovalores associados a cada componente principal para o estudo da porcentagem de tempo para a média das 12 atividades	48
Tabela 4	Desvio Padrão, Proporção da variância e proporção da variância acumulada explicada por cada componente principal para o estudo da porcentagem de tempo para a média das 12 atividades	49
Tabela 5	Valores do índice de Gini para o estudo da porcentagem de tempo para a média das 12 atividades	53
Tabela 6	Comparação entre as contribuições e os valores da função de influência para os autovalores λ1 , λ2 , λ3 e λ4 (análise da porcentagem de tempo média para as 12 atividades)	54
Tabela 7	Comparação entre as contribuições e os valores da função de influência para os autovalores λ5 , λ6 , λ7 e λ8 (análise da porcentagem de tempo média para as 12 atividades)	55
Tabela 8	Comparação entre as contribuições e os valores da função de influência para os autovalores λ9 , λ10 , λ11 e λ12 (análise da porcentagem de tempo média para as 12 atividades).....	56
Tabela 9	Estimativa de Variância para cada atividade	58

Tabela 10	Autovalores associados a cada componente principal para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro	59
Tabela 11	Desvio padrão, proporção da variância e proporção da variância acumulada explicada por cada componente principal para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro	60
Tabela 12	Valores do índice de Gini para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro.....	64
Tabela 13	Comparação entre as contribuições e os valores da função de influência para os autovalores λ_{11} , λ_{12} , λ_{13} e λ_{14} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro)	65
Tabela 14	Comparação entre as contribuições e os valores da função de influência para os autovalores λ_{15} , λ_{16} , λ_{17} e λ_{18} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro)	66
Tabela 15	Comparação entre as contribuições e os valores da função de influência para os autovalores λ_{19} , λ_{110} , λ_{111} e λ_{112} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro)	67
Tabela 16	Autovalores associados a cada componente principal para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231	69
Tabela 17	Desvio padrão, proporção da variância e proporção da variância acumulada explicada por cada componente principal	

	(atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231).....	70
Tabela 18	Valores do índice de Gini para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231	75
Tabela 19	Comparação entre as contribuições e os valores da função de influência para os autovalores λ_{11} , λ_{12} , λ_{13} e λ_{14} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231).....	76
Tabela 20	Comparação entre as contribuições e os valores da função de influência para os autovalores λ_{15} , λ_{16} , λ_{17} e λ_{18} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231).....	77
Tabela 21	Comparação entre as contribuições e os valores da função de influência para os autovalores λ_{19} , λ_{110} , λ_{111} e λ_{112} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231).....	78

LISTA DE FIGURAS

Figura 1	Gráfico de dispersão para medidas positivamente correlacionadas em um sistema de coordenadas rotacionado	21
Figura 2	Diferentes tipos de <i>outliers</i> considerando um conjunto de dados tridimensional projetado sobre um espaço bidimensional ...	30
Figura 3	Curva de Lorenz para o j -ésimo autovalor λ_j	32
Figura 4	Registro das posturas básicas de tronco, pernas e braços	40
Figura 5	Combinações posturais	41
Figura 6	Curvas de Lorenz para os componentes de 1 a 6 (análise da porcentagem de tempo média para as 12 atividades)	50
Figura 7	Curvas de Lorenz para os componentes de 7 a 12 (análise da porcentagem de tempo média para as 12 atividades)	51
Figura 8	Curvas de Lorenz para os componentes de 1 a 6 (análise da porcentagem de tempo para atividade aplicação de herbicida no morro)	62
Figura 9	Curvas de Lorenz para os componentes de 7 a 12 (análise da porcentagem de tempo para atividade aplicação de herbicida no morro)	63
Figura 10	Curvas de Lorenz para os componentes de 1 a 6 (atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)	71
Figura 11	Curvas de Lorenz para os componentes de 7 a 12 (atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)	72

SUMÁRIO

1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO	17
2.1	Análise exploratória multivariada	17
2.2	Análise de componentes principais	18
2.2.1	Componentes principais populacionais	19
2.2.2	Componentes principais amostrais	22
2.2.3	Variáveis padronizadas	24
2.3	Funções de influência	25
2.4	Robustez em PCA	27
2.4.1	Outliers na análise de componentes principais	29
2.5	Curva de Lorenz e índice de concentração Gini	30
2.6	Café e agricultura familiar	33
2.7	Agricultura familiar e ergonomia	34
2.8	Trabalho agrícola e ergonomia	35
3	MATERIAIS E MÉTODOS	37
3.1	Material	37
3.1.1	Registro e processamento dos dados	38
3.2	Métodos	39
3.2.1	Posturas	39
3.2.2	Procedimento das análises	41
4	RESULTADOS E DISCUSSÃO	46
4.1	Estudo da sensibilidade para a porcentagem média de tempo das 12 atividades	46
4.2	Estudo da sensibilidade para a atividade aplicação de herbicida no morro	57
4.3	Estudo da sensibilidade para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231	68
5	CONCLUSÃO	80
	REFERÊNCIAS	81

1 INTRODUÇÃO

A estatística multivariada consiste em um conjunto de técnicas utilizadas quando o pesquisador deseja estudar duas ou mais variáveis simultaneamente em determinado experimento. Uma das técnicas multivariadas que podem ser utilizadas neste caso é a análise de componentes principais. Com esta técnica busca-se explicar a estrutura de variância e covariância de um vetor de variáveis aleatórias através de combinações lineares das mesmas, denominadas de componentes principais. Além disso, busca-se uma redução no número de variáveis a serem avaliadas.

Em se tratando do estudo de dados comportamentais, em geral, as técnicas estatísticas empregadas são usuais, como por exemplo, análise de variância e testes de hipótese. Especificamente, a aplicação da análise multivariada a esse tipo de dados se resume, basicamente, na utilização dos componentes principais, no entanto, o estudo da sensibilidade das observações na formação desses componentes, bem como a aplicabilidade no estudo comportamental de trabalhadores na cafeicultura, não é encontrado na literatura. Neste contexto, destaca-se a motivação dada para realização deste trabalho.

No que tange ao estudo dos dados comportamentais de trabalhadores na lavoura cafeeira da região do Sul de Minas, a importância em identificar as observações sensíveis à formação dos componentes principais dar-se-á pelo fato de que a carga física de trabalho nesse seguimento é intensa e se distribui de forma desigual entre os sistemas de trabalho. Além do mais, pode haver diferença na intensidade da carga de trabalho quando o trabalho é executado em condições de relevo plano ou em aclive (BARBOSA, 2013).

Em virtude do que foi mencionado, este trabalho foi conduzido com o objetivo de aplicar a análise de componentes principais a dados comportamentais de trabalhadores em lavouras de café do Sul de Minas Gerais,

realizando um procedimento para identificar a influência que cada observação exerce em relação à formação dos componentes principais, de modo a priorizar uma interpretação contextualizada ao sistema de trabalho associado às combinações posturais predefinidas durante atividades na lavoura.

2 REFERENCIAL TEÓRICO

Nesta seção, é apresentada uma revisão de literatura, que visa abordar conceitos fundamentais relativos a componentes principais, funções de influência, robustez em PCA, curvas de Lorenz, cafeicultura, ergonomia e agricultura familiar.

2.1 Análise exploratória multivariada

A estatística multivariada consiste em um conjunto de métodos estatísticos que são utilizados em situações nas quais o pesquisador deseja estudar várias variáveis, simultaneamente, em um experimento, medidas em cada elemento amostral. Essa situação é normalmente vista na maioria dos experimentos, pois dificilmente o interesse é focado em apenas uma variável (JOHNSON; WICHERN, 2007). Geralmente o que se faz, é tratar uma situação legitimamente multivariada como univariada, ou seja, analisar as variáveis isoladamente. Dessa forma, são aplicadas técnicas estatísticas usuais, como a análise de variância, análise de regressão, entre outras, o que não é conveniente quando se deseja estudar essas variáveis em conjunto e, também, quando se tem um número elevado de variáveis.

É importante ressaltar que os métodos multivariados, em sua grande maioria, são exploratórios, o que de maneira nenhuma os torna menos importantes do que qualquer outro método de análise estatística. Eles são de extrema importância no tocante a reconhecimento de padrões, identificação de *outliers*, classificação de variáveis, redução de dimensionalidade e tratamento dos dados para a realização de uma inferência estatística. Além disso, é de fundamental importância a realização de uma análise descritiva/exploratória em

qualquer conjunto de dados a serem analisados, pois, além de tornar mais evidente a técnica que deve ser utilizada, evita, muitas vezes, a ocorrência de erros nas análises.

Os métodos de estatística multivariada são, geralmente, utilizados com o propósito de simplificar ou facilitar a interpretação do fenômeno em estudo. Isso é feito considerando-se índices ou novas variáveis alternativas que sintetizam a informação fornecida pelos dados originais. Por esses métodos procura-se construir novos grupos de elementos amostrais que apresentam similaridade entre si, proporcionando, assim, uma segmentação do conjunto de dados original. Além disso, tem como objetivo investigar as relações de dependência entre as variáveis associadas ao fenômeno em estudo.

A estatística multivariada se divide em dois grupos: um primeiro que consiste em técnicas exploratórias de sintetização ou simplificação da estrutura de variância e covariância dos dados, e o segundo que consiste em técnicas de inferência estatística (MINGOTI, 2005).

A análise de componentes principais é uma das técnicas multivariadas que se encaixa no primeiro grupo, buscando, assim, sintetizar a estrutura de variância e covariância dos dados.

2.2 Análise de componentes principais

A análise de componentes principais (PCA) se concentra em explicar a estrutura de variância e covariância de um vetor aleatório por meio de combinações lineares das variáveis originais. Em geral, o que se deseja é obter uma redução no número de variáveis e interpretação das combinações lineares obtidas (JOHNSON; WICHERN, 2007). Essas combinações lineares são chamadas de componentes principais e são mutuamente não correlacionadas.

Considerando-se inicialmente p variáveis originais, é possível obter-se p componentes principais. Embora p componentes principais sejam necessários, para representar a variabilidade total do sistema, tem-se que, frequentemente, grande parte desta variabilidade pode ser explicada por um número $k < p$, de componentes principais. Sendo assim, há tanta informação nos k componentes, como existe nas p variáveis originais. Dessa forma, os k componentes principais podem substituir as p variáveis, e o conjunto de dados original, que era constituído por n medidas em p variáveis, é reduzido para um conjunto constituído por n medidas em k componentes principais. Dessa forma, tem-se uma simplificação da estrutura de covariância (JOLLIFFE, 2002).

A qualidade dessa aproximação depende da quantidade de componentes mantidos no sistema e pode ser medida com base na proporção da variância total explicada pelos componentes remanescentes. Determinados os componentes, os seus valores numéricos, que são denominados de escores, podem ser calculados para cada um dos elementos amostrais (FERREIRA, 2011).

2.2.1 Componentes principais populacionais

Seja $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ um vetor aleatório, com matriz de covariâncias dada por $\Sigma_{p \times p}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ os autovalores associados a essa matriz, e sejam $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$, os correspondentes autovetores normalizados. Os autovetores \mathbf{e}_i satisfazem às seguintes condições:

- (i) $\mathbf{e}_i' \mathbf{e}_j = 0$, para todo $i \neq j$;
- (ii) $\mathbf{e}_i' \mathbf{e}_i = 1$, para todo $i = 1, 2, \dots, p$;

$$(iii) \sum_{p \times p} \square \mathbf{e}_i = \lambda_i \mathbf{e}_i, \text{ para todo } i = 1, 2, \dots, p.$$

Os componentes principais são definidos como:

$$\begin{aligned} Y_1 &= \mathbf{e}'_1 \mathbf{X} = e_{11}X_1 + e_{12}X_2 + \dots + e_{1p}X_p \\ Y_2 &= \mathbf{e}'_2 \mathbf{X} = e_{21}X_1 + e_{22}X_2 + \dots + e_{2p}X_p \\ &\vdots \\ Y_p &= \mathbf{e}'_p \mathbf{X} = e_{p1}X_1 + e_{p2}X_2 + \dots + e_{pp}X_p \end{aligned}$$

A variância do j -ésimo componente principal é dada pelo correspondente autovalor λ_j , ou seja:

$$Var[Y_j] = \mathbf{e}'_j \sum_{p \times p} \square \mathbf{e}_j = \lambda_j, \quad j = 1, 2, \dots, p$$

Além disso,

$$Cov(Y_j, Y_k) = \mathbf{e}'_j \sum_{p \times p} \square \mathbf{e}_k = 0, \text{ para } j \neq k.$$

Portanto, como já foi mencionado anteriormente, os componentes são não correlacionadas entre si.

Algebricamente, os componentes principais são combinações lineares de p variáveis aleatórias particulares X_1, X_2, \dots, X_p . Por outro lado, geometricamente, essas combinações lineares representam a seleção de um novo sistema de coordenadas, obtido pela rotação do sistema original que tem $X_1,$

X_1, \dots, X_p como as coordenadas dos eixos. Os novos eixos Y_1, Y_2, \dots, Y_p , que são os componentes principais, representam as direções de máxima variabilidade e proporcionam uma descrição mais simples e parcimoniosa da estrutura de covariância (JOHNSON; WICHERN, 2007).

Por definição, o primeiro componente principal é a combinação linear com maior variabilidade e esse valor decresce até o p -ésimo componente, que é o que possui menor variância (GIRI, 2004).

Esse fato se deve à própria forma de construção, pois como pode ser observado na Figura 1, cada ponto no sistema de coordenadas original X_1 e X_2 passa a ser projetado ortogonalmente pela rotação por meio de um ângulo θ , em um novo sistema de coordenadas Y_1 e Y_2 , que nada mais são que os dois primeiros componentes principais, CP₁ e CP₂ (JOHNSON; WICHERN, 2007).

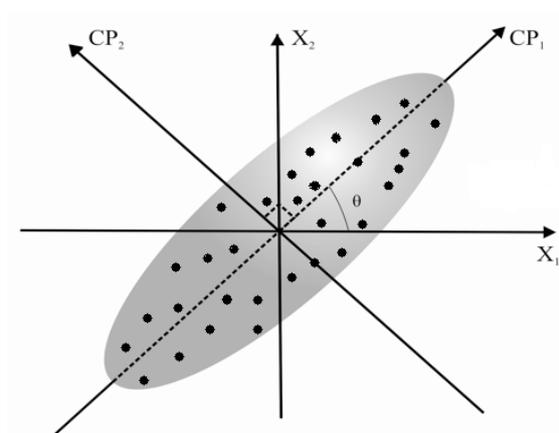


Figura 1 Gráfico de dispersão para medidas positivamente correlacionadas em um sistema de coordenadas rotacionado

A proporção da variância total explicada pelo j -ésimo componente principal é dada por:

$$\frac{Var[Y_j]}{Var\ total\ de\ X} = \frac{\lambda_j}{Tr(\Sigma_{p \times p})} = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_p}. \quad (1)$$

Considerando-se a definição apresentada em (1), tem-se que a proporção da variância total explicada pelos k primeiros componentes é dada por:

$$\begin{aligned} \frac{Var[Y_1] + Var[Y_2] + \dots + Var[Y_k]}{Var\ total\ de\ X} &= \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{Tr(\Sigma_{p \times p})} = \\ &= \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}. \end{aligned}$$

Se os k primeiros componentes principais explicarem uma grande parte da variabilidade total do vetor X , pode-se restringir o estudo apenas a eles. Dessa

forma, a matriz de covariâncias $\Sigma_{p \times p}$ é aproximada por:

$$\Sigma_{p \times p} \approx \sum_{j=1}^k \lambda_j e_j e_j'$$

2.2.2 Componentes principais amostrais

Na seção anterior, as modelagens foram apresentadas considerando-se a

matriz de covariância populacional, $\Sigma_{p \times p}$. Na prática, a matriz $\Sigma_{p \times p}$ é desconhecida e deve ser estimada pelo conjunto de dados amostrais coletados.

Dessa forma, a matriz de covariâncias populacional $\Sigma_{p \times p}$ é estimada pela matriz

de covariâncias amostral $\mathbf{S}_{p \times p}$. É importante ressaltar que todos os procedimentos descritos anteriormente são válidos, também, para o caso amostral (JOLLIFFE, 2002).

Considere uma matriz de dados $\mathbf{X}_{n \times p}$ cujos elementos x_{ij} , $i = 1, \dots, n$, $j = 1, \dots, p$ são os valores de n observações em p variáveis, medidas de forma que a soma de x_{ij} para $i = 1, \dots, n$ seja igual a zero, ou seja, utilizando-se uma matriz de dados padronizada. Sejam $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p$ os autovalores da matriz $\mathbf{S}_{p \times p}$ e sejam $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p$ os autovetores associados à matriz. Dessa forma, o estimador do j -ésimo componente principal amostral é definido por:

$$\hat{Y}_j = \hat{e}_j' \mathbf{X} = \hat{e}_{j1} X_1 + \hat{e}_{j2} X_2 + \dots + \hat{e}_{jp} X_p \quad (2)$$

Assim, como no caso populacional, a variância estimada de \hat{Y}_j é igual a $\hat{\lambda}_j$ para $j = 1, 2, \dots, p$, e a covariância entre dois componentes quaisquer \hat{Y}_j e \hat{Y}_k é igual a zero, para todo $j \neq k$, o que implica em dizer que eles são não correlacionados.

Além disso, a proporção da variância total explicada pelo j -ésimo componente principal e a proporção da variância total explicada pelos k primeiros componentes é calculada de forma similar ao caso populacional. A única diferença é que será usada a matriz de covariâncias amostral $\mathbf{S}_{p \times p}$ em vez

de $\Sigma_{p \times p}$, e os valores dos componentes principais e dos autovalores serão valores estimados.

Como pode ser verificado em Bénasséni (2005), os escores, que são os valores estimados dos componentes principais para cada observação, podem, ainda, ser definidos de outra forma como:

$$z_{ij} = \sum_{k=1}^p e_{jk} x_{ik}, \text{ para } i = 1, \dots, n. \quad (3)$$

Além disso, baseando-se na propriedade de que a variância do j -ésimo componente principal é dada pelo j -ésimo autovalor λ_j , pode-se obter o j -ésimo autovalor em função dos escores como:

$$\lambda_j = \frac{1}{n} \sum_{i=1}^n z_{ij}^2, \text{ com } j = 1, 2, \dots, p.$$

2.2.3 Variáveis padronizadas

A obtenção dos componentes principais envolve a decomposição da matriz de covariâncias do vetor aleatório de variáveis em estudo, portanto, caso seja realizada alguma transformação nesse vetor, os componentes devem ser obtidos considerando a matriz de covariâncias relativa ao vetor transformado (MINGOTI, 2005).

Uma transformação, geralmente, utilizada é a padronização das variáveis, utilizando-se suas médias e desvios padrões, da seguinte maneira:

$$\begin{aligned} Z_1 &= \frac{(X_1 - \mu_1)}{\sqrt{\sigma_1^2}} \\ Z_2 &= \frac{(X_2 - \mu_2)}{\sqrt{\sigma_2^2}} \\ &\vdots \quad \quad \quad \vdots \end{aligned}$$

$$Z_p = \frac{(X_p - \mu_p)}{\sqrt{\sigma_p^2}},$$

em que $\mu_1, \mu_2, \dots, \mu_p$ são as médias de cada variável e $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$ suas respectivas variâncias. A obtenção dos componentes principais por dados padronizados é equivalente a obtê-los utilizando a matriz de correlação das variáveis originais (JOHNSON; WICHERN, 2007).

2.3 Funções de influência

As funções de influência são importantes ferramentas no estudo da sensibilidade em componentes principais. Sua teoria foi introduzida por Hampel (1974). Seja \mathbf{X} um vetor aleatório p -variado com função de distribuição acumulada (f.d.a.) dada por $F(x)$ e função de distribuição $f(x)$. Se θ é um vetor de parâmetros de f , então θ pode ser escrito como função de F , isto é, $\theta = t(F)$. Agora, seja $F(x)$ contaminada para se tornar

$$\tilde{F}(x) = (1 - \varepsilon)F(x) + \varepsilon\delta_y(x),$$

em que $0 < \varepsilon < 1$ e $\delta_y(x)$ é a função de distribuição acumulada da variável aleatória que certamente assume o valor x . Além disso, seja $\hat{\theta}$ o valor de θ quando F torna-se \tilde{F} . Dessa forma, a função de influência $IF(x; \theta)$ para θ avaliado em x é definida como:

$$IF(x, \theta) = \lim_{\varepsilon \rightarrow 0} \frac{t(\tilde{F}(x)) - t(F)}{\varepsilon} = \frac{\hat{\theta} - \theta}{\varepsilon}.$$

Em muitas situações, pode ser obtida uma expressão explícita para a função de influência. A função de influência para λ_j e e_j , o j -ésimo autovalor e o j -ésimo autovetor da matriz de covariância populacional, respectivamente, são dados por Critchley (1985) e Radhakrishnan e Kshirsagar (1981). Para a matriz de covariância populacional, a função de influência para λ_j pode ser escrita como:

$$IF(x, \lambda_j) = z_{ij}^2 - \lambda_j, \quad (4)$$

em que z_{ij} é o escore correspondente ao valor dado de x . Portanto, a influência de uma observação sobre λ_j depende somente do escore para o j -ésimo componente principal; uma observação pode ser extrema em qualquer ou em todos os componentes principais sem afetar λ_j . Isso ilustra a afirmação de que *outliers* não são, necessariamente, influentes para cada uma das partes de uma análise. Para a matriz de covariância populacional, a função de influência teórica pode ser definida para e_j e escrita como:

$$IF(x, e_j) = z_{ij} \sum_{h \neq j}^p \frac{z_h e_h}{\lambda_j - \lambda_h}. \quad (5)$$

Comparando (4) e (5), observa-se que uma observação pode ser influente para λ_j , e não para e_j , e vice versa. Observa-se, também, que $IF(x, e_j)$ depende de todos os escores, ao contrário de $IF(x, \lambda_j)$ que depende apenas de z_{ij} .

Na prática, a função de influência possui termos desconhecidos, como é o caso da função $F(x)$ que é substituída pela função de distribuição empírica baseada na amostra. Dessa forma, a correspondente função de influência empírica para $\hat{\lambda}_j$ é obtida por:

$$I_i(j) = \widehat{IF}(x, \lambda_j) = z_{ij}^2 - \hat{\lambda}_j. \quad (6)$$

A função de influência empírica para e_j é dada por:

$$\widehat{IF}(x, e_j) = z_{ij} \sum_{h \neq j} \frac{z_h \hat{e}_h}{\hat{\lambda}_j \hat{\lambda}_h},$$

em que $\hat{\lambda}_j$ é o j -ésimo autovalor, \hat{e}_k é o k -ésimo autovetor da matriz de covariância amostral, e z_k é o valor do escore para cada observação individual.

Como os critérios utilizados neste trabalho são baseados em Bénasséni (2005), o cálculo dos valores da função de influência empírica será realizado como visto na expressão (6).

2.4 Robustez em PCA

As análises estatísticas estão sempre apoiadas sobre pressuposições básicas, que podem ser explícitas ou implícitas, a fim de garantir validade dos resultados obtidos. A palavra “robusto” pode ter vários significados, muitas vezes inconsistentes. Na Estatística, o termo robustez, geralmente, é utilizado, em um sentido relativamente estreito, como sendo a insensibilidade para pequenos desvios das pressuposições (HUBER, 1981).

Uma observação é chamada influente quando sua remoção acarreta um grande efeito sobre os resultados da análise. A maioria das observações pode ser influente em algum aspecto, mas nem toda observação discrepante é, potencialmente, influente (WEDLAKE, 2008).

A análise de componentes principais mostrou-se sensível a observações discrepantes por muitos autores, como Devlin, Gnanadesikan e Kettenring (1981) e Wedlake (2008). Diante disso, podem ser tomadas várias ações. Uma possível estratégia é a utilização de estimadores robustos para a matriz de covariância ou correlação, obtendo-se as funções de influência e as variâncias assintóticas correspondentes para estes estimadores robustos, como pode ser observado em Croux e Haesbroeck (2000).

Para Ibazizen e Dauxois (2003), o estudo das observações influentes em PCA deve estar em concordância com a essência da análise exploratória do comportamento das unidades experimentais. Por essa razão, é sugerido um tratamento das observações discrepantes que não se reduz a uma simples regra de rejeição. O estudo da robustez em PCA é importante quando ocorrem, principalmente, dois tipos de observações influentes: os pontos de alavancagem e os pontos residuais. Ibazizen e Dauxois (2003) propõem a utilização de um critério baseado em uma função de perda convexa e suficientemente diferenciável. Dessa forma, as funções de influência de um vetor de médias e vetores principais foram desenvolvidas de modo a proporcionar a obtenção de um PCA robusto. As observações podem, também, ser ponderadas pela distância de Mahalanobis, fornecendo uma estimativa robusta de sua localização.

Esse tópico de PCA robusto vem recebendo maior atenção nas literaturas. Croux e Ruiz-Gazen (1996) introduziram um algoritmo, para obtenção de componentes principais robustos, baseado na busca de projeções. Bénasséni (1985) trata da influência dos pesos das observações sobre os autovalores. Os critérios utilizados neste trabalho serão baseados em Bénasséni

(2005), que trata do estudo das curvas de Lorenz e do associado índice de Gini como auxiliares à função de influência no estudo da sensibilidade dos autovalores em PCA.

2.4.1 *Outliers* na análise de componentes principais

De acordo com Hubert, Rousseeuw e Branden (2005), podem ser distinguidas quatro tipos de observações e entre elas estão os pontos de influência considerados bons e ruins, além das observações regulares e dos *outliers* ortogonais. A Figura 2, onde se observa um conjunto de dados tridimensional, projetado em um espaço bidimensional, deixa clara a diferença entre os quatro tipos de observações:

- a) Observações regulares: são observações que formam um grupo homogêneo (nuvem de observações) no mesmo plano do biplot;
- b) Pontos de influência considerados bons: são pontos que se encontram no mesmo plano do biplot, porém afastados da nuvem de observações, como exemplo, observam-se os pontos 1 e 4 na Figura 2;
- c) *Outliers* ortogonais: são pontos que têm uma grande distância ortogonal do plano do biplot, porém, sua projeção coincide com a nuvem de observações, um exemplo é a observação 5 na Figura 2;
- d) Pontos de influência considerados ruins: são pontos que têm uma grande distância ortogonal do plano do biplot e, além disso, sua projeção não coincide com a nuvem de pontos, os pontos 2 e 3 na Figura 2 são exemplos.

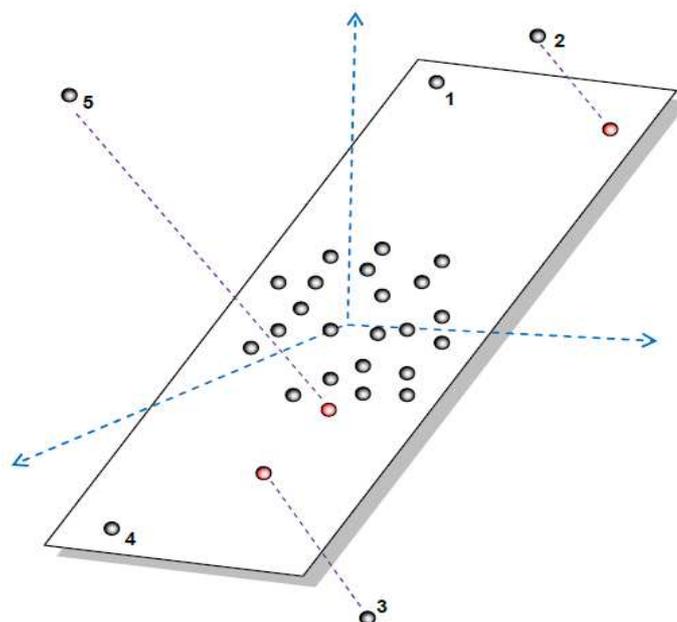


Figura 2 Diferentes tipos de *outliers* considerando um conjunto de dados tridimensional projetado sobre um espaço bidimensional

Os pontos de influência considerados bons, ruins e os *outliers* ortogonais, podem ser observações influentes em uma análise. Um dos métodos utilizados para identificar tais observações no cenário multivariado são as funções de influência, além das curvas de Lorenz e o associado índice de Gini, que são utilizados como auxiliares para tal detecção produzindo, assim, um diagnóstico estatístico e gráfico, como pode ser observado em Bénasséni (2005).

2.5 Curva de Lorenz e índice de concentração Gini

A curva de Lorenz e o associado índice Gini são medidas sensíveis que fornecem informação adicional à função de influência no estudo da contribuição que cada observação fornece para a variabilidade dos autovalores. A seguir serão detalhados os passos para a construção das curvas, desde o cálculo das

contribuições $c_i(j)$ de cada observação, até o gráfico. Além disso, será descrito como são obtidos os valores para o índice de Gini.

Voltando a atenção para o j -ésimo componente principal, define-se a contribuição $c_i(j)$ da i -ésima observação para a variância desse componente como:

$$c_i(j) = \frac{1}{n} \frac{z_{ij}^2}{\lambda_j} \times 100. \quad (7)$$

A contribuição $c_i(j)$ é a porcentagem da variância do componente explicada pela i -ésima observação. Se a variância estiver concentrada em poucas observações com valores grandes de $c_i(j)$, diz-se que o autovalor λ_j é muito dependente dessas observações, ou seja, ele é altamente sensível a essas observações. Caso contrário, se nenhuma das observações tem um valor significativamente alto de $c_i(j)$, então, λ_j é insensível a qualquer observação, ou em alguns casos, moderadamente sensível a elas (BÉNASSÉNI, 2005).

Agora, para $i = 1, 2, \dots, n$, consideremos $c_{(i)}(j)$ denotando as contribuições $c_i(j)$ classificadas em ordem crescente, definimos dessa forma a porcentagem da variância de λ_j explicada pelas k observações correspondentes aos k menores valores de $c_i(j)$ como:

$$C_k(j) = \sum_{i=1}^k c_{(i)}(j), \quad \text{para } k = 1, 2, \dots, n. \quad (8)$$

Construindo-se um gráfico de $C_k(j)$ versus a porcentagem acumulada para as observações, dada por $p_k = 100 \times k/n$, obtém-se a curva de Lorenz para λ_j , cujo exemplo está apresentado na Figura 3.

O clássico índice de concentração Gini, dado por G_j , mede a adequação da curva de Lorenz com base na linha de igualdade. O índice G_j é dado pelo dobro da área entre a curva de Lorenz e a linha de igualdade e é verificado como um indicador sensetivo variando entre 0 e 1. Na prática, um valor moderado indica que o correspondente autovalor é, relativamente, insensitivo a qualquer observação e, portanto, tem uma boa estabilidade. Um valor mais próximo de 1 indica que uma ou algumas observações apresentam grandes contribuições para o autovalor e, assim, ele é dependente dessas observações (BÉNASSÉNI, 2005). Vale salientar que é difícil estabelecer um ponto a partir do qual o índice é considerado elevado ou moderado, dessa forma, o índice de Gini é, geralmente, utilizado como critério de comparação.

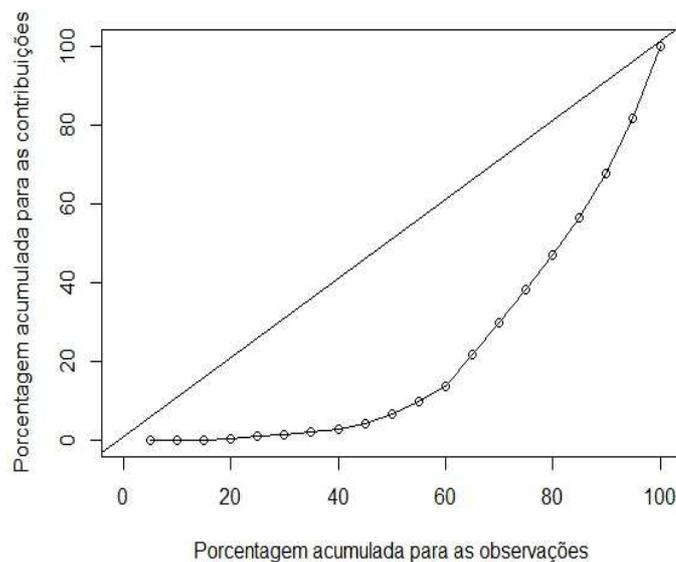


Figura 3 Curva de Lorenz para o j -ésimo autovalor λ_j

Como pode ser observado em (6) e (7), tanto $I_i(t)$ quanto $c_i(t)$, envolvem z_{ij}^2 como termo principal. Pode-se destacar uma relação interessante entre essas duas medidas. Geralmente, grandes valores de $c_i(t)$ correspondem a grandes valores de $I_i(t)$ e, portanto, há uma conexão entre o valor do índice Gini (G_j) e o valor da função de influência. Um valor moderado de G_j indica que a curva de Lorenz permanece relativamente perto da linha de igualdade. Portanto, não há, provavelmente, nenhum valor dominante de $c_i(t)$ e tem-se a mesma conclusão para $I_i(t)$. Por outro lado, altos valores de G_j tendem a indicar que há poucas observações que são potencialmente influentes, com valores bastante elevados de $c_i(t)$ e $I_i(t)$ (BÉNASSÉNI, 2005).

2.6 Café e agricultura familiar

A cafeicultura desempenha um papel fundamental na economia do Brasil, tanto em regiões que têm o café como principal fonte de sustentação econômica, quanto pela sua grande contribuição para os indicadores de exportação e ao superávit da balança comercial do país. Além disso, é o segundo mercado consumidor, ficando atrás dos Estados Unidos. No Brasil, o Estado de Minas Gerais merece um destaque especial, sendo responsável por, aproximadamente, 50% da produção nacional (BARBOSA, 2013).

Segundo a Organização Internacional do Café – OIC, o consumo mundial cresce em média 1,5 % ao ano. No Brasil, o consumo interno cresce, aproximadamente, 3,0 % ao ano, sendo o dobro da média internacional (ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE CAFÉ - ABIC, 2007).

Em razão desse aumento no consumo do café, é possível verificar em nível mundial, um alto investimento em novas tecnologias, objetivando-se o

aumento da produtividade e destaque mundial pela qualidade. Além disso, o sucesso da cafeicultura pode ser observado com o surgimento de novos países produtores e exportadores do produto (WARKEN et al., 2007).

A agricultura familiar, em que a família, além de ser proprietária dos meios de produção, também, assume o trabalho no estabelecimento produtivo, é a principal geradora de postos de trabalho no meio rural brasileiro. A cafeicultura desempenha um papel importante nesse segmento, sendo uma alternativa para a viabilização econômica e social. São levantadas várias questões relativas não só ao aspecto tecnológico da produção, como também à recomposição da identidade cultural da agricultura familiar e suas condições de trabalho (BARBOSA, 2013).

2.7 Agricultura familiar e ergonomia

No que diz respeito às condições de trabalho, torna-se importante o conceito de ergonomia, que é uma ciência que procura configurar, planejar e adaptar o trabalho ao homem, respondendo às questões que podem ser levantadas em condições insatisfatórias de trabalho (DUL; WEERDMEESTER, 2004). O que ocorre nos dias de hoje é uma busca constante por uma alta produtividade, que nada mais é do que uma alta produção mantendo-se uma boa qualidade do produto.

Um importante método, ligado diretamente à ergonomia, utilizado para melhorar as condições de trabalho é a Avaliação Ergonômica do Trabalho (AET), que tem como principais objetivos humanizar o trabalho, evitar doenças causadas pelo trabalho, assim como acidentes, melhorar a utilização dos recursos humanos e oferecer mais conforto, saúde e satisfação ao trabalhador (PINHEIRO; FRANÇA, 2006).

As atividades agrícolas, muitas vezes, demandam cargas de trabalho que podem ser prejudiciais ao organismo do trabalhador, levando-o ao desgaste e, assim, podendo causar problemas graves e crônicos de saúde. Quando o trabalhador é submetido a condições de trabalho desfavoráveis, podem ocorrer vários problemas, como fadiga, diminuição no rendimento, aumento dos erros, dores musculares e riscos de acidentes no trabalho (BARBOSA, 2013).

Além disso, a fadiga reduz os recursos (físicos e mentais) do trabalhador, aumentando, assim, o esforço exigido para realização de suas tarefas e induzindo, dessa forma, a perda de eficiência temporária, ou em casos mais graves, até uma perda irreversível, causando graves danos à saúde do trabalhador (FALZON, 2007).

Esse é um ponto crucial na agricultura familiar, pois como o produtor depende de recursos próprios, caso ele venha a adoecer ele fica desprovido de mão de obra. Por este motivo torna-se importante o estudo de alguns fatores que influenciam na saúde do trabalhador agrícola, como exemplo, podemos citar o esforço físico exigido para execução do trabalho e as posturas que o trabalhador executa as tarefas (BARBOSA, 2013).

2.8 Trabalho agrícola e ergonomia

O trabalho agrícola é uma das atividades mais antigas de que se tem registro, desde que os nossos ancestrais se fixaram na terra e dela passaram a tirar sua subsistência. As técnicas de plantio, na grande maioria das vezes, não proporcionam segurança e condições de trabalho adequadas aos que trabalham nessa atividade (BARBOSA, 2013).

Pela ótica da ergonomia, a atividade rural pode ser classificada como sendo não estruturada, ou seja, ao contrário do que pode ser observado na indústria, ela, geralmente, não possui posto de trabalho definido, além disso, o

conjunto de tarefas a serem executadas é muito variável. Em geral, são tarefas árduas e executadas em posturas inconvenientes, exigindo grande força muscular, além disso, o ambiente é desfavorável, com exposição direta ao sol, chuvas e ventos (IIDA, 2005).

A ergonomia apresenta grande importância no meio agrícola, realizando levantamentos antropométricos e propondo adequações. Podem ser observados vários benefícios alcançados em países que investiram em pesquisas ergonômicas do trabalho agrícola, como maior organização do trabalho, projetos de ferramentas e equipamentos adequados às tarefas agrícolas e planejamento dos postos de trabalho. Neste sentido, a ergonomia pode ser relevante em termos de produtividade, conforto e redução dos custos de operação (ABRAHÃO, 1993).

Um dos objetivos da ergonomia é desenvolver ambientes de trabalho que reduzam as exigências biomecânicas, colocando o trabalhador em uma boa postura de trabalho, os comandos dentro do alcance dos movimentos corporais e oferecendo facilidade de percepção das informações. O posto de trabalho deve envolver o operador como uma vestimenta bem adaptada, onde ele possa realizar o trabalho com conforto, eficiência e segurança (IIDA, 2005).

Além da aplicação de critérios ergonômicos na agricultura, que permitam estabelecer a correta adaptação dos componentes do sistema homem-trabalho, torna-se importante o investimento em pesquisas que possam melhorar as condições desse trabalho (WARKEN, 2007).

3 MATERIAIS E MÉTODOS

Nesta seção, são apresentados os dados utilizados para realização do estudo, bem como a descrição completa da metodologia utilizada na realização das análises.

3.1 Material

Os dados utilizados neste trabalho são oriundos de um experimento realizado por Barbosa (2013), considerando contatos com a Associação dos Agricultores Familiares de Santo Antônio do Amparo – Força Café, localizada no município de Santo Antônio do Amparo, Estado de Minas Gerais.

Como pode ser observado em Barbosa (2013), foram contatadas sete propriedades familiares, indicadas pela associação, que apresentavam em suas propriedades plantações de café em situações de plano e morro, com o objetivo de definir doze trabalhadores voluntários para a realização da pesquisa.

Os critérios de inclusão e exclusão da amostragem foram os seguintes: as propriedades de onde foram selecionados os trabalhadores deveriam participar ativamente da associação; as propriedades deveriam se dedicar ao cultivo do café, em tempo integral e apresentar lavouras em produção em áreas planas e com declividade superior a 10%.

Das sete propriedades indicadas, foram contatados e definidos os doze trabalhadores voluntários que participaram da pesquisa: um trabalhador em cada uma das propriedades 1, 2, 3 e 4; dois trabalhadores em cada uma das propriedades 5 e 6; e quatro trabalhadores na propriedade 7.

Posteriormente, foram coletados alguns dados pessoais dos voluntários (data de nascimento, gênero e senioridade) e, também, sobre a propriedade (endereço, área, produção e declividade). Os trabalhadores foram informados sobre os procedimentos relativos à sua participação na pesquisa e também, sobre os objetivos do estudo. Além disso, foi solicitado que cada voluntário assinasse um termo de livre consentimento para a participação no projeto.

Como exigência, para realização da pesquisa, o projeto passou pelo comitê de ética da UNICAMP, onde foi analisado e aprovado pelos membros.

Para o desenvolvimento deste estudo, foram selecionadas as seguintes tarefas, definidas anteriormente em Barbosa (2013), na lavoura em produção: a) Tratos culturais: adubação a lanço, desbrota, aplicação de herbicida e adubação foliar; b) Colheita: derriça manual; c) Pós-colheita: secagem no terreiro e armazenamento. As atividades de tratos culturais e colheita foram avaliadas tanto em áreas planas como em áreas com declividade, totalizando, assim, doze atividades.

3.1.1 Registro e processamento dos dados

Foram feitos registros de imagens de cada trabalhador em cada tarefa do sistema de trabalho na cafeicultura com a utilização de uma câmera digital. Cada um dos doze trabalhadores foi filmado por uma hora, realizando as tarefas de tratos culturais, colheita e pós-colheita de forma a relacionar a atividade de trabalho com as posturas adotadas. Portanto, foram doze horas de registros de imagens para cada um dos doze trabalhadores.

Para processamento das imagens, foi utilizado o CAPTIV (SOFTWARE..., 1998), que é um software elaborado para análise de tarefas e observações. O CAPTIV tem como principal característica a capacidade de ajustar o tempo real da tarefa, com base nos dados definidos pelo pesquisador,

ou seja, por meio do software se tem a possibilidade de agregar o protocolo de observação de comportamento (gestos, posturas, ações operacionais, eventos) com as atividades de trabalho. Dessa forma foram obtidas as porcentagens de tempo gastas por cada trabalhador, em cada atividade, em cada postura predefinida.

3.2 Métodos

Nesta seção, será apresentada a descrição completa da metodologia utilizada, desde a definição das combinações posturais até a descrição completa dos procedimentos estatísticos.

3.2.1 Posturas

Para que fossem definidas as combinações posturais, foi realizado um estudo piloto, onde três trabalhadores foram filmados, durante uma hora, realizando as tarefas de tratos culturais, colheita e pós-colheita e as tarefas de tratos culturais e colheita foram realizadas tanto no plano, como em terreno com declividade. O repertório postural típico de trabalho na cafeicultura foi definido com base em adaptações do método OWAS (IIDA, 2005) de classificação postural. A codificação das posturas foi realizada, em função da orientação dos três segmentos corporais, tronco, braços e pernas e a combinação das posturas básicas, que podem ser observadas na Figura 4, resultou em quarenta e cinco diferentes combinações.

Na Figura 5, observam-se as 16 combinações posturais, das 45 possíveis, que apareceram com maior frequência, durante a execução das doze atividades de trabalho na cafeicultura. Para cada atividade houve predominância de diferentes combinações posturais.

TRONCO	Neutra <20°	Flexão moderada 20° a 45°	Flexão severa >45°		
					
	1	2	3		
BRAÇOS	Dois braços acima da linha dos ombros	Um braço acima da linha dos ombros	Dois braços abaixo da linha dos ombros		
					
	1	2	3		
PERNAS	Pernas estendidas	Pernas flexionadas	Agachado	Ajoelhado	Sentado
					
	1	2	3	4	5

Figura 4 Registro das posturas básicas de tronco, pernas e braços

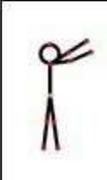
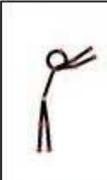
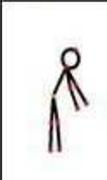
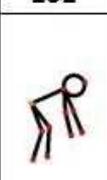
			
111	121	131	135
			
211	221	222	225
			
231	232	233	234
			
331	332	333	334

Figura 5 Combinações posturais

3.2.2 Procedimento das análises

Inicialmente, dentre as dezesseis combinações posturais consideradas (Figura 5), foram selecionadas as 14 que mais se destacaram e sua descrição é apresentada na Tabela 1.

Tabela 1 Descrição das 14 posturas selecionadas

Posturas	Descrição
111	tronco neutro, os dois braços acima da linha dos ombros e as pernas estendidas
121	tronco neutro, um braço acima da linha dos ombros e as pernas estendidas
131	tronco neutro, os dois braços abaixo da linha dos ombros e as pernas estendidas
211	flexão moderada de tronco, os dois braços acima da linha dos ombros e as pernas estendidas
222	flexão moderada de tronco, um braço acima da linha dos ombros e as pernas flexionadas
225	flexão moderada de tronco, um braço acima da linha dos ombros, sentado
231	flexão moderada de tronco, os dois braços abaixo da linha dos ombros e as pernas estendidas
232	flexão moderada de tronco, os dois braços abaixo da linha dos ombros e as pernas flexionadas
233	flexão moderada de tronco, os dois braços abaixo da linha dos ombros, agachado
234	flexão moderada de tronco, os dois braços abaixo da linha dos ombros, ajoelhado
331	flexão severa de tronco, os dois braços abaixo da linha dos ombros e as pernas estendidas
332	flexão severa de tronco, os dois braços abaixo da linha dos ombros e as pernas flexionadas
333	flexão severa de tronco, os dois braços abaixo da linha dos ombros, agachado
334	flexão severa de tronco, os dois braços abaixo da linha dos ombros, ajoelhado

As combinações posturais descartadas foram as 135 e 221, pois apresentaram porcentagem de tempo muito baixa para as doze atividades, sendo assim, consideradas desprezíveis para a realização das análises.

A estrutura do conjunto de dados foi definida da seguinte forma: os doze trabalhadores formaram as colunas, sendo considerados como variáveis, e as 14 posturas selecionadas formaram as linhas, sendo consideradas como observações. A estrutura foi assim definida em virtude da consequente aplicação da análise de PCA ao conjunto de dados. Cada valor do conjunto de dados se refere à porcentagem de tempo média das doze atividades (adubação a lanço, desbrota, aplicação de herbicida, adubação foliar e derriça manual, cada uma dessas atividades consideradas no plano e no morro, além das atividades de pós-colheita, que são a secagem no terreiro e armazenamento), correspondente ao j -ésimo trabalhador na i -ésima postura.

A execução das análises seguiu os seguintes passos:

- a) Foi aplicada a técnica multivariada de análise de componentes principais (PCA) aos dados, com base em qual foram obtidos os 12 autovalores e 12 autovetores da matriz de correlações das variáveis. Com isso, foram obtidos os componentes principais e assim, os escores, como pode ser observado em (2) e (3), respectivamente.
- b) Foi feito o cálculo das contribuições $c_i(t)$, conforme definido em (7), obtendo, assim, a porcentagem de contribuição que cada observação fornece para a variância de cada um dos 12 componentes.
- c) Posteriormente, essas contribuições foram ordenadas em ordem crescente e, portanto, calculadas as $C_k(t)$, como definido em (8), que correspondem às porcentagens da variância do componente em

questão explicada pelas k observações que correspondem aos k menores valores da contribuição c_i .

- d) Calculadas as porcentagens acumuladas, estas foram utilizadas na obtenção das curvas de Lorenz. Logo, foi também calculado o índice Gini, dado pelo dobro da área entre a curva de Lorenz e a linha de igualdade.
- e) O próximo passo foi o cálculo dos valores da função de influência I_i , como dado em (6). Assim, esses valores foram organizados juntamente com os valores de contribuição c_i , destacando-se a boa relação existente entre essas duas medidas.
- f) Finalmente foram feitas as interpretações da análise de componentes principais realizada, das curvas de Lorenz obtidas, dos valores do índice Gini (G_i), das contribuições e da função de influência, com o objetivo de avaliar a sensibilidade dos componentes principais selecionados, assim como a identificação das posturas que apresentaram maior influência na estimação e seleção desses componentes.

Na segunda etapa das análises, foram feitos os mesmos procedimentos aplicados anteriormente, porém, em vez de trabalhar com a média das doze atividades, foi selecionada a atividade aplicação de herbicida no morro, que foi aquela que apresentou maior variância dentre as doze. Isso foi feito com o objetivo de comparar a mudança ocorrida com relação à sensibilidade, comparando-se as curvas, os valores das contribuições, da função de influência e do índice de Gini, além de identificar as posturas que mais contribuíram para a variabilidade de cada autovalor.

Na terceira etapa, foram aplicados novamente, os mesmos procedimentos descritos (em a , b , c , d , e e f), a diferença nessa etapa, é que os

resultados foram obtidos por análises realizadas considerando a retirada das posturas, dentre as 14 selecionadas, que apresentaram maiores valores de contribuição na análise feita anteriormente para a atividade aplicação de herbicida no morro. Foram retiradas as posturas que apresentaram os maiores valores de contribuições para os componentes principais selecionados.

Novamente, foram observadas as mudanças ocorridas com relação à sensibilidade, observando as curvas de Lorenz, os valores das contribuições, da função de influência e do índice de Gini, além de identificar as posturas que mais contribuíram para a variabilidade de cada autovalor.

Para todas as análises foram feitas as discussões cabíveis e interpretação dos resultados de interesse obtidos. Todas as análises necessárias foram feitas com a utilização do software R (R DEVELOPMENT CORE TEAM, 2013). Maiores detalhes sobre os procedimentos realizados na obtenção e organização dos dados podem ser encontrados em Barbosa (2013).

4 RESULTADOS E DISCUSSÃO

Na apresentação dos resultados será feito, primeiramente, o estudo da sensibilidade para a média da porcentagem de tempo das 12 atividades. Posteriormente, apresentar-se-á o estudo para a atividade que apresentou maior variabilidade dentre as 12 e, finalmente, retiradas as posturas que apresentaram maior contribuição, será apresentada a análise feita para a atividade selecionada anteriormente sem a presença dessas posturas visando observar a mudança ocorrida no comportamento das curvas de Lorenz e índice de Gini, juntamente com a função de influência.

4.1 Estudo da sensibilidade para a porcentagem média de tempo das 12 atividades

Na Tabela 2 observam-se os valores das correlações entre os trabalhadores; cada um dos 12 trabalhadores foi considerado como uma variável para realização das análises.

Observa-se que os trabalhadores (variáveis) são positivamente correlacionados entre si e, em sua maioria, apresentam uma forte correlação. Esse resultado já era esperado, já que a porcentagem de tempo para os 12 trabalhadores nas 14 posturas predefinidas foram avaliadas em todas as doze atividades.

Tabela 2 Correlações entre as variáveis *

Var	Var											
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	
T1	1,00											
T2	0,80	1,00										
T3	0,66	0,78	1,00									
T4	0,61	0,82	0,95	1,00								
T5	0,74	0,89	0,92	0,96	1,00							
T6	0,76	0,81	0,96	0,95	0,97	1,00						
T7	0,64	0,82	0,48	0,56	0,66	0,55	1,00					
T8	0,53	0,84	0,51	0,59	0,66	0,52	0,96	1,00				
T9	0,58	0,86	0,57	0,60	0,67	0,56	0,85	0,91	1,00			
T10	0,62	0,89	0,55	0,62	0,71	0,57	0,85	0,90	0,95	1,00		
T11	0,71	0,92	0,53	0,59	0,73	0,60	0,86	0,87	0,89	0,93	1,00	
T12	0,65	0,90	0,66	0,67	0,72	0,64	0,92	0,95	0,93	0,92	0,87	1,00

*Cada variável corresponde a um dos 12 trabalhadores

Observa-se, na Tabela 3, os autovalores encontrados e, como foram observados 12 trabalhadores (variáveis), há 12 autovalores. O primeiro é aquele que representa uma maior parte da variabilidade encontrada nas variáveis, e os últimos são praticamente desprezíveis, apresentando valores muito pequenos.

Na Tabela 4 pode-se observar que os dois primeiros componentes principais acumulam cerca de 91,42% da variância total. Vale ressaltar que seria possível reduzir facilmente o sistema composto por 12 variáveis, para um sistema composto por 2 componentes principais. A perda na porcentagem da variabilidade explicada com essa redução é muito pequena, pois a proporção da variância total explicada pelos demais componentes é muito baixa, podendo,

assim, ser desprezada. Esse resultado é válido, já que um número de componentes que explique uma porcentagem da variabilidade acima de 70% já nos garante uma boa representação do conjunto de variáveis.

Tabela 3 Autovalores associados a cada componente principal para o estudo da porcentagem de tempo para a média das 12 atividades

Autovalores	
λ_1	8,64201600000
λ_2	1,54481100000
λ_3	0,45192660000
λ_4	0,21811800000
λ_5	0,14840520000
λ_6	0,05840668000
λ_7	0,04483496000
λ_8	0,02500983000
λ_9	0,00725241200
λ_{10}	0,00177649000
λ_{11}	0,00021551720
λ_{12}	0,00008436768

Tabela 4 Desvio Padrão, Proporção da variância e proporção da variância acumulada explicada por cada componente principal para o estudo da porcentagem de tempo para a média das 12 atividades

Componentes Principais	Desvio Padrão	Proporção da Variância Explicada	Proporção da Variância Acumulada
CP1	3,05070	0,77560	0,77560
CP2	1,28980	0,13860	0,91420
CP3	0,69760	0,04060	0,95476
CP4	0,48470	0,01960	0,97433
CP5	0,39980	0,01330	0,98765
CP6	0,25080	0,00520	0,99289
CP7	0,21970	0,00400	0,99692
CP8	0,16410	0,00220	0,99916
CP9	0,08840	0,00065	0,99981
CP10	0,04370	0,00016	0,99997
CP11	0,01520	0,00002	0,99999
CP12	0,00950	0,00001	1,00000

Nas Figuras 6 e 7 são apresentadas as curvas de Lorenz obtidas para os doze autovalores. A forma das curvas fornecem importantes informações sobre a sensibilidade que cada autovalor apresenta diante de cada observação (posturas). Assim, com uma inspeção visual podem-se detectar observações potencialmente influentes, ou seja, que contribuíram de forma altamente significativa para o autovalor em questão.

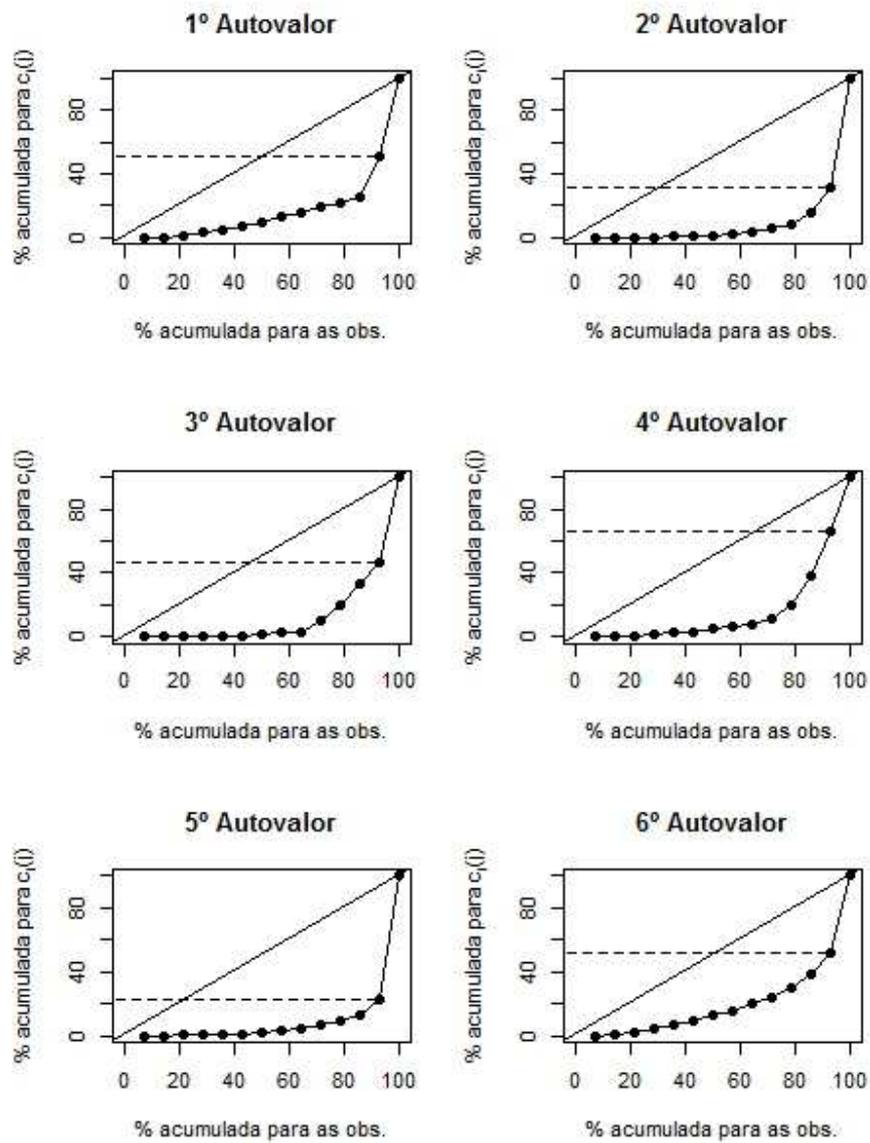


Figura 6 Curvas de Lorenz para os componentes de 1 a 6 (análise da percentagem de tempo média para as 12 atividades)

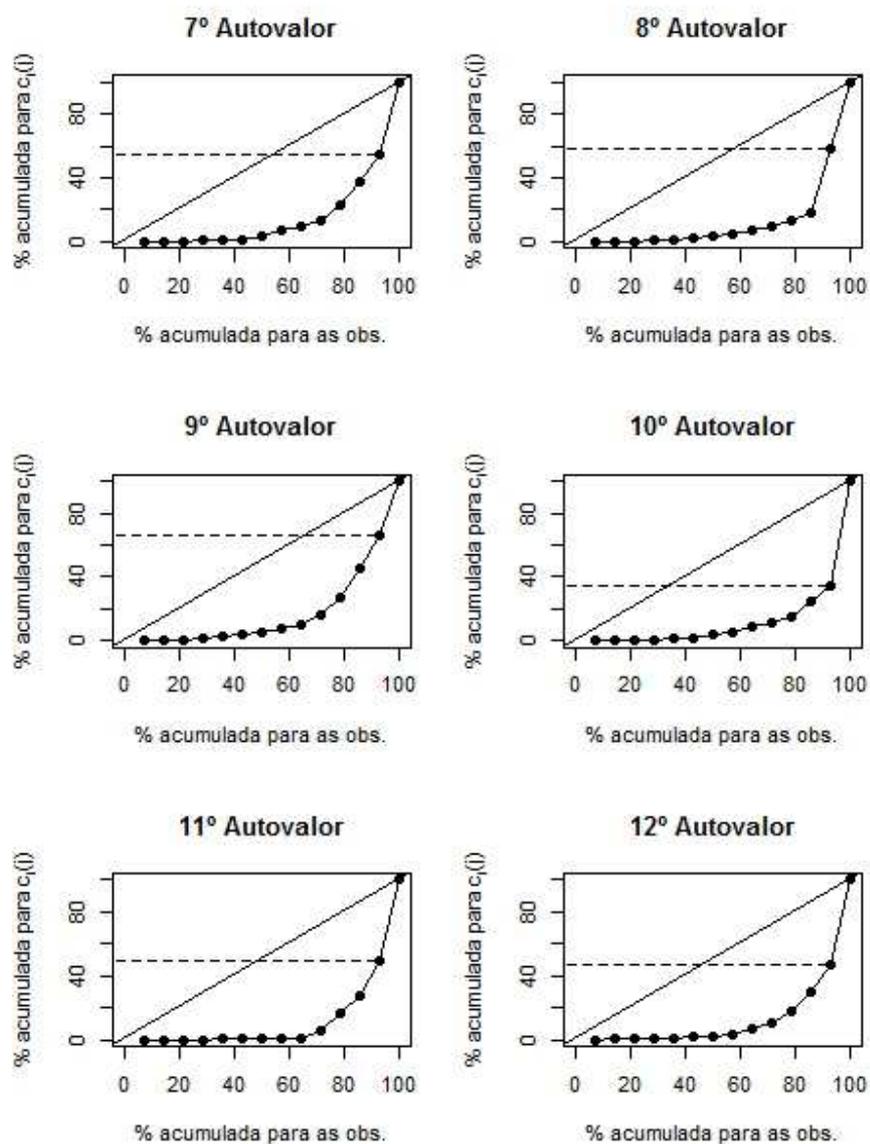


Figura 7 Curvas de Lorenz para os componentes de 7 a 12 (análise da porcentagem de tempo média para as 12 atividades)

Observa-se na Figura 6, pela curva de Lorenz, associada ao primeiro componente, que a observação mais influente responde por, aproximadamente, 50% da variabilidade associada a λ_1 , enquanto que para o segundo componente, a observação mais influente responde por, aproximadamente, 70% da variabilidade associada a λ_2 .

Observa-se, ainda, pela curva de Lorenz, associada ao terceiro componente, que a observação mais influente responde por, aproximadamente, 55% da variabilidade associada a λ_3 , enquanto que a observação mais influente para o quarto componente responde por, aproximadamente, 35% da variabilidade associada a λ_4 .

Por meio da curva de Lorenz, associada ao quinto componente, pode ser verificado que a observação mais influente representa, aproximadamente, 78% da variabilidade associada a λ_5 , por outro lado, para o sexto componente, a observação mais influente representa, aproximadamente, 50% da variabilidade associada a λ_6 .

Na Figura 7, observa-se pelas curvas de Lorenz, associadas ao sétimo e oitavo componentes, que a observação mais influente para cada um desses componentes, responde por, aproximadamente, 42% da variabilidade associada a λ_7 e λ_8 .

Observa-se, ainda, pela curva de Lorenz, associada ao nono componente, que a observação mais influente representa, aproximadamente, 35% da variabilidade associada a λ_9 , enquanto que para o décimo componente, a observação mais influente corresponde a, aproximadamente, 65% da variabilidade associada a λ_{10} .

Por meio das curvas de Lorenz, associadas ao décimo primeiro componente e décimo segundo, pode ser verificado que as observações que apresentaram maior contribuição para cada um deles, correspondem a, aproximadamente, 50% da variabilidade associada a λ_{11} e λ_{12} .

Tabela 5 Valores do índice de Gini para o estudo da porcentagem de tempo para a média das 12 atividades

G_j	
G_1	0,6797
G_2	0,8276
G_3	0,7596
G_4	0,6992
G_5	0,8393
G_6	0,6211
G_7	0,7073
G_8	0,7589
G_9	0,6637
G_{10}	0,7757
G_{11}	0,7856
G_{12}	0,7569

Na tabela 5 observam-se os valores do índice de Gini para as 12 curvas de Lorenz obtidas, como pode ser observado, em sua maioria, são valores elevados, aproximando-se de 1, principalmente, para os autovalores λ_2 e λ_5 , que apresentam valores do índice iguais a 0,8276 e 0,8393, respectivamente. Esses valores do índice de Gini serão comparados com os obtidos nas análises posteriores. Esse resultado indica, juntamente com a inspeção visual das curvas, que uma ou algumas observações apresentam grandes contribuições para esses autovalores e, dessa forma, sua variabilidade se torna dependente delas. Essas observações (posturas) podem ser identificadas subsequentemente pelas tabelas 6, 7 e 8, onde serão apresentadas as contribuições $c_i(j)$ em conjunto com os valores $I_i(j)$ da função de influência empírica para $i = 1, 2, \dots, 14$ e $j = 1, 2, \dots, 12$.

Como foi dito anteriormente, essas duas medidas apresentam uma boa relação, que pode ser confirmada por resultados obtidos. Verifica-se que observações que apresentam valores de contribuição mais elevados, no geral, apresentam, também, os maiores valores para a função de influência. Além disso, observações com pequenas contribuições apresentam valores pequenos para a função de influência, geralmente negativos. Assim, torna-se possível a confirmação dos resultados observados anteriormente pelas curvas de Lorenz e associado índice de Gini.

Tabela 6 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_1 , λ_2 , λ_3 e λ_4 (análise da porcentagem de tempo média para as 12 atividades)

Obs.	λ_1		λ_2		λ_3		λ_4	
	c_i (1)	I_i (1)	c_i (2)	I_i (2)	c_i (3)	I_i (3)	c_i (4)	I_i (4)
111	1,504	- 6,822	2,078	- 1,095	0,716	- 0,407	17,915	0,329
121	0,261	- 8,327	14,748	1,645	6,461	- 0,043	34,453	0,834
131	25,037	21,650	0,813	- 1,369	53,458	2,930	1,101	- 0,184
211	2,350	- 5,798	0,298	- 1,480	10,362	0,204	0,351	- 0,207
222	3,321	- 4,624	0,212	- 1,499	0,048	- 0,449	0,039	- 0,217
225	3,342	- 4,598	0,075	- 1,528	0,065	- 0,448	1,600	- 0,169
231	49,002	50,645	8,273	0,244	13,279	0,388	7,852	0,022
232	2,575	- 5,526	68,944	13,366	13,227	0,385	0,791	- 0,194
233	3,564	- 4,330	0,117	- 1,519	0,170	- 0,441	0,020	- 0,218
234	2,826	- 5,223	0,054	- 1,533	0,438	- 0,424	3,569	- 0,109
331	3,026	- 4,981	0,264	- 1,488	0,156	- 0,442	0,934	- 0,190
332	0,014	- 8,625	0,956	- 1,338	0,004	- 0,452	1,546	- 0,171
333	1,442	- 6,897	2,639	- 0,974	0,644	- 0,411	28,098	0,640
334	1,735	- 6,543	0,529	- 1,430	0,973	- 0,390	1,732	- 0,165

Tabela 7 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_5 , λ_6 , λ_7 e λ_8 (análise da porcentagem de tempo média para as 12 atividades)

Obs.	λ_5		λ_6		λ_7		λ_8	
	σ_i (5)	I_i (5)	σ_i (6)	I_i (6)	σ_i (7)	I_i (7)	σ_i (8)	I_i (8)
111	1,470	-0,118	1,538	-0,046	0,118	-0,044	0,0002	-0,025
121	0,655	-0,135	5,683	-0,012	0,743	-0,040	0,021	-0,025
131	1,207	-0,123	3,907	-0,026	2,924	-0,026	0,802	-0,022
211	0,431	-0,139	1,974	-0,042	15,047	0,050	41,712	0,121
222	0,247	-0,143	2,343	-0,039	0,059	-0,044	2,083	-0,018
225	3,061	-0,085	3,455	-0,030	8,570	0,009	0,652	-0,023
231	0,076	-0,147	8,597	0,012	2,464	-0,029	1,729	-0,019
232	2,203	-0,103	3,359	-0,031	0,046	-0,045	0,015	-0,025
233	0,039	-0,148	2,500	-0,038	0,339	-0,043	1,153	-0,021
234	9,611	0,051	13,953	0,056	17,520	0,065	4,751	-0,008
331	0,031	-0,148	0,00008	-0,058	0,297	-0,043	0,600	-0,023
332	77,862	1,469	4,430	-0,022	3,131	-0,025	4,280	-0,010
333	2,801	-0,090	47,611	0,331	3,768	-0,021	2,220	-0,017
334	0,303	-0,142	0,649	-0,053	44,973	0,237	39,982	0,115

Tabela 8 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_9 , λ_{10} , λ_{11} e λ_{12} (análise da porcentagem de tempo média para as 12 atividades)

Obs.	λ_9		λ_{10}		λ_{11}		λ_{12}	
	c_i (9)	I_i (9)	c_i (10)	I_i (10)	c_i (11)	I_i (11)	c_i (12)	I_i (12)
111	33,829	0,027	10,241	0,0008	0,043	- 0,0002	0,320	-0,0001
121	20,453	0,014	1,631	-0,001	0,024	- 0,0002	0,202	-0,0001
131	0,966	-0,006	0,640	-0,002	0,002	- 0,0002	0,931	-0,0001
211	10,914	0,004	2,186	-0,001	0,051	- 0,0002	3,409	-0,0001
222	19,023	0,012	3,491	-0,001	11,443	0,0001	17,304	0,0001
225	5,255	-0,002	0,336	-0,002	50,915	0,0013	7,953	0,00001
231	0,540	-0,007	0,310	-0,002	0,074	- 0,0002	0,004	-0,0001
232	1,287	-0,006	0,0008	-0,002	0,028	- 0,0002	0,173	-0,0001
233	1,320	-0,006	9,228	0,001	5,134	- 0,0001	52,714	0,0005
234	3,470	-0,004	3,175	-0,001	21,425	0,0004	12,048	0,0001
331	2,014	-0,005	65,407	0,014	10,548	0,0001	3,613	-0,0001
332	0,060	-0,007	0,027	-0,002	0,013	- 0,0002	0,273	-0,0001
333	0,036	-0,007	2,885	-0,001	0,049	- 0,0002	0,600	-0,0001
334	0,834	-0,006	0,443	-0,002	0,249	- 0,0002	0,456	-0,0001

Na Tabela 6, observa-se que para o primeiro autovalor λ_1 , as posturas 231 e 131, nessa ordem, foram as que apresentaram maiores contribuições e, conseqüentemente, maiores valores para a função de influência.

Considerando agora o autovalor λ_2 , as posturas 232 e 121, nessa ordem, destacaram-se apresentando os maiores valores para a contribuição e função de influência.

Para o terceiro autovalor λ_3 , destacaram-se as posturas 131, 231 e 232, respectivamente. Enquanto que para λ_4 , as posturas que apresentaram maiores valores de contribuição e da função de influência foram as posturas 121, 333 e 111, respectivamente.

Destaca-se a postura 332 para o autovalor λ_5 , enquanto que para λ_6 , destacaram-se as posturas 333 e 234, nessa ordem.

Para o sétimo autovalor λ_7 , destacaram-se as posturas 334, 234 e 211, nessa ordem, enquanto que para o autovalor λ_8 destacaram-se com maiores contribuições e maiores valores para a função de influência, as posturas 211 e 334, respectivamente.

Podem ser destacadas para o autovalor λ_9 as posturas 111, 121 e 222, respectivamente, enquanto que para o décimo autovalor λ_{10} , a postura que se destaca, apresentando o maior valor de contribuição e maior valor para a função de influência, é a postura 331.

Com relação ao autovalor λ_{11} , as posturas que apresentaram os maiores valores para a contribuição e função de influência foram as posturas 225 e 234, nessa ordem. Já para o autovalor λ_{12} se destacaram as posturas 233, 222 e 234, respectivamente.

É importante ressaltar que os resultados obtidos mostram que posturas que apresentam valores elevados para contribuição e função de influência para determinados autovalores não são as mesmas que apresentam para outros, ou seja, determinada postura pode ser muito influente para um autovalor e por outro lado, ser praticamente desprezível para outro.

4.2 Estudo da sensibilidade para a atividade aplicação de herbicida no

morro

O principal foco dessa seção é repetir todas as análises feitas anteriormente, porém, em vez de trabalhar com a média das 12 atividades, neste estudo serão considerados apenas os valores de porcentagem de tempo para uma atividade isolada, atividade essa que apresentou maior variância dentre as 12. Essas análises serão conduzidas visando comparar os resultados anteriores das curvas de Lorenz, índice de Gini e função de influência, com os resultados aqui obtidos.

Dessa forma, como pode ser observado na Tabela 9, a atividade que apresentou maior variabilidade, dentre as 12, foi a aplicação de herbicida no morro, seguida pela secagem no terreiro, adubação foliar no plano e aplicação de herbicida no plano, que, também, destacaram-se.

Tabela 9 Estimativa de Variância para cada atividade

Atividades	Variância
Adubação a lanço – Morro	321,4609
Adubação a lanço – Plano	377,3117
Adubação foliar – Morro	377,4660
Adubação foliar – Plano	457,5505
Aplicação de herbicida – Morro	519,0551
Aplicação de herbicida – Plano	443,1540
Armazenamento	241,0149
Colheita – Morro	156,8417
Colheita – Plano	175,6207
Desbrota– Morro	271,0989
Desbrota– Plano	213,2739

Secagem no terreiro

490,3707

Na Tabela 10, podem ser observados os autovalores associados a cada componente principal. Assim como na análise anterior, o primeiro autovalor corresponde a uma maior parte da variabilidade encontrada nas variáveis e os últimos são praticamente desprezíveis, pois apresentam valores muito próximos de zero. Os autovalores encontrados são fundamentais para o cálculo das contribuições e dos valores da função de influência, sendo assim, essenciais, também, para a construção das curvas de Lorenz.

Tabela 10 Autovalores associados a cada componente principal para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro

Autovalores	
λ_1	9,20478700000
λ_2	1,92591600000
λ_3	0,00802522000
λ_4	0,00302335800
λ_5	0,00087426120
λ_6	0,00022902740
λ_7	$1,594228 \times 10^{-6}$
λ_8	$8,624260 \times 10^{-7}$
λ_9	$2,296209 \times 10^{-8}$
λ_{10}	$1,936385 \times 10^{-31}$
λ_{11}	$8,372851 \times 10^{-32}$
λ_{12}	$2,300111 \times 10^{-32}$

Na Tabela 11, podem ser observados o desvio padrão e a proporção da variância explicada por cada componente, além da proporção da variância acumulada explicada pelos k primeiros componentes principais. Merece destaque o fato de que os dois primeiros componentes acumulam, praticamente, toda a variabilidade presente nas variáveis (trabalhadores), sendo responsáveis por cerca de 99,89% da variância total. Em consequência disso, a proporção da variância total explicada pelos demais componentes é muito baixa, sendo praticamente desprezível.

Tabela 11 Desvio padrão, proporção da variância e proporção da variância acumulada explicada por cada componente principal para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro

Componentes Principais	Desvio padrão	Proporção da variância explicada	Proporção da variância acumulada
CP1	3,14850	0,82610	0,82610
CP2	1,44020	0,17280	0,99890
CP3	0,09297	0,00072	0,99963
CP4	0,05706	0,00027	0,99990
CP5	0,03068	0,00008	0,99998
CP6	0,01570	0,00002	1,00000
CP7	0,00131	0	1,00000
CP8	0,00096	0	1,00000
CP9	0,00016	0	1,00000
CP10	$4,96 \times 10^{-17}$	0	1,00000
CP11	$2,35 \times 10^{-18}$	0	1,00000
CP12	$4,16 \times 10^{-19}$	0	1,00000

A seguir, nas figuras 8 e 9, são apresentadas as curvas de Lorenz obtidas para cada um dos doze autovalores.

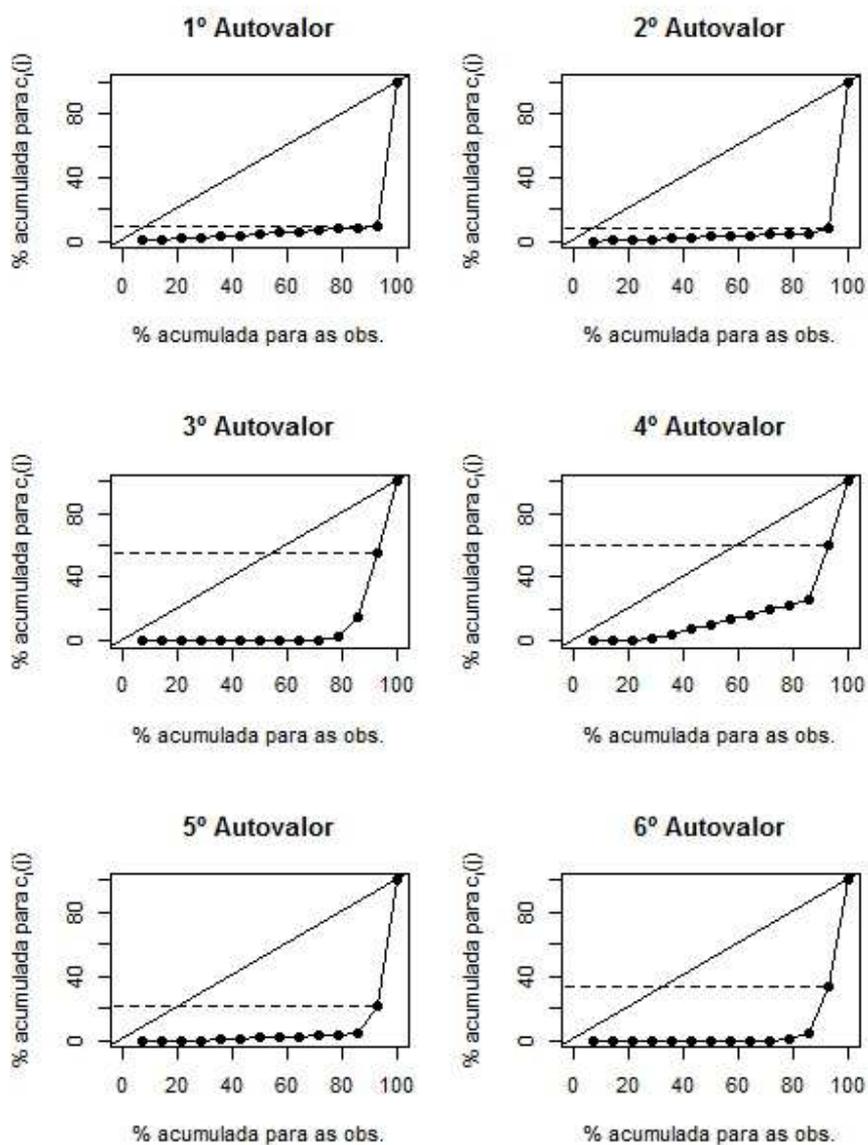


Figura 8 Curvas de Lorenz para os componentes de 1 a 6 (análise da porcentagem de tempo para atividade aplicação de herbicida no morro)

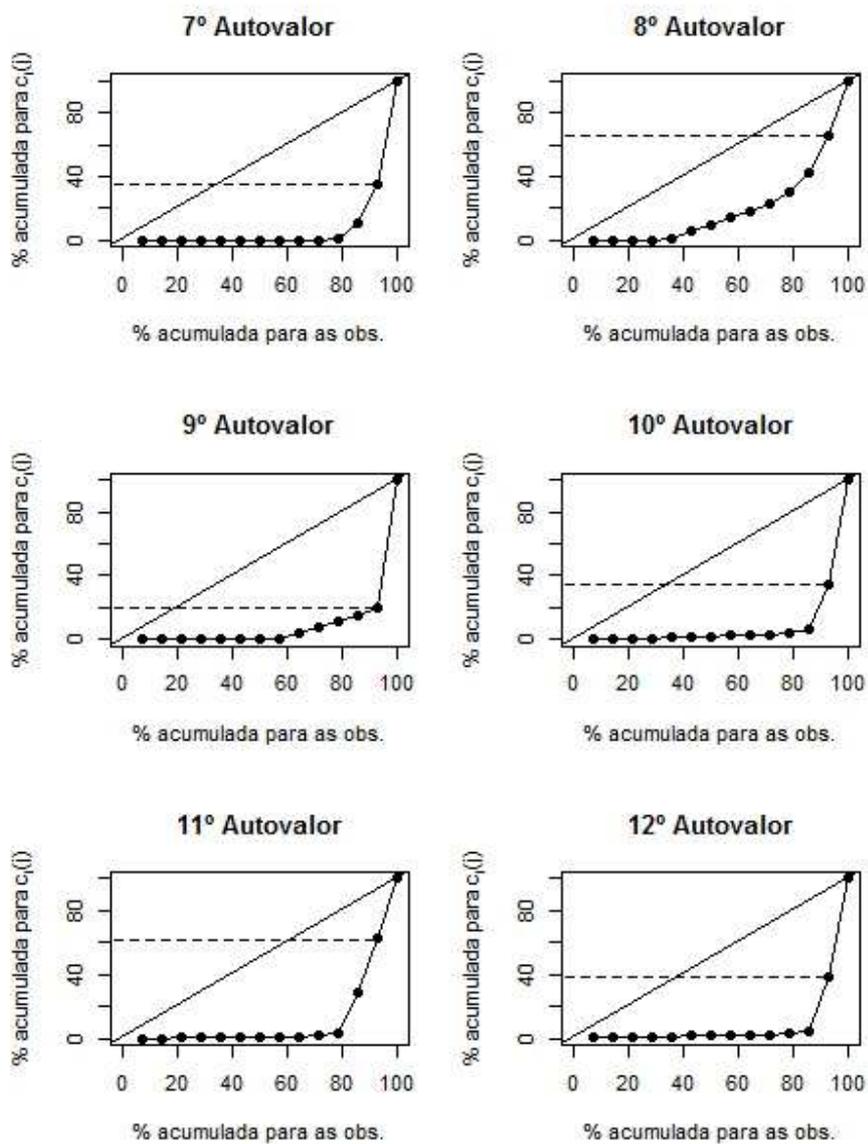


Figura 9 Curvas de Lorenz para os componentes de 7 a 12 (análise da porcentagem de tempo para atividade aplicação de herbicida no morro)

Na Figura 8, observa-se pelas curvas de Lorenz, associadas aos dois primeiros componentes, que a observação mais influente para cada um deles, responde por, aproximadamente, 90% da variabilidade associada a λ_1 e λ_2 .

Observa-se, ainda, por meio da curva de Lorenz, associada ao terceiro componente, que a observação mais influente responde por, aproximadamente, 47% da variabilidade associada a λ_3 , enquanto que a observação mais influente para o quarto componente responde por, aproximadamente, 40% da variabilidade associada a λ_4 .

Por meio da curva de Lorenz, associada ao quinto componente, pode ser notado que a observação mais influente representa, aproximadamente, 80% da variabilidade associada a λ_5 , por outro lado, para o sexto componente, a observação mais influente representa, aproximadamente, 68% da variabilidade associada a λ_6 .

Na Figura 9, observa-se pela curva de Lorenz, associada ao sétimo componente, que a observação mais influente responde por, aproximadamente, 65% da variabilidade associada a λ_7 , enquanto que para o oitavo componente, a observação mais influente representa, aproximadamente, 35% da variabilidade associada a λ_8 .

Por meio da curva de Lorenz, associada ao nono componente, pode ser verificado que a observação mais influente representa, aproximadamente, 80% da variabilidade associada a λ_9 , enquanto que para o décimo componente, a observação mais influente corresponde a, aproximadamente, 65% da variabilidade associada a λ_{10} .

Observa-se, ainda, pela curva de Lorenz, associada ao décimo primeiro componente, que a observação mais influente representa, aproximadamente,

38% da variabilidade associada a λ_{11} , enquanto que para o décimo segundo componente, a observação mais influente responde por, aproximadamente, 60% da variabilidade associada a λ_{12} .

Tabela 12 Valores do índice de Gini para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro

G_j	
G_1	0,8363
G_2	0,8684
G_3	0,8242
G_4	0,6691
G_5	0,8712
G_6	0,8739
G_7	0,8586
G_8	0,6271
G_9	0,8447
G_{10}	0,8456
G_{11}	0,7865
G_{12}	0,8435

Na Tabela 12 observam-se os valores do índice de Gini para as 12 curvas de Lorenz obtidas, como pode ser observado, em sua maioria, são valores elevados, aproximando-se de 1. Esse resultado indica, juntamente com a inspeção visual das curvas, que uma ou algumas observações (posturas) apresentam grandes contribuições para esses autovalores e, dessa forma, sua variabilidade se torna dependente delas. Essas observações (posturas) podem ser identificadas, subsequentemente, pelas das Tabelas 13, 14 e 15.

Comparando os valores aqui encontrados com os valores encontrados na análise realizada, utilizando-se a porcentagem de tempo das 12 atividades, o que

pode ser observado é que, na sua maioria, os valores do índice de Gini são maiores do que aqueles encontrados anteriormente, excetuando-se os valores do índice correspondentes ao quarto e oitavo componentes principais, que são menores. Esse resultado já era esperado, em virtude do fato de que a segunda análise foi realizada, utilizando os dados de porcentagem de tempo provenientes apenas da atividade aplicação de herbicida no morro, que foi aquela que apresentou maior variância, comparada às demais.

As Tabelas 13, 14 e 15 apresentam as contribuições $c_i(j)$ em conjunto com os valores $I_i(j)$ da função de influência empírica para $i = 1, 2, \dots, 14$ e $j = 1, 2, \dots, 12$, tornando possível a confirmação dos resultados observados anteriormente pelas curvas de Lorenz e do associado índice de Gini.

Tabela 13 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_1 , λ_2 , λ_3 e λ_4 (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro)

Obs.	λ_1		λ_2		λ_3		λ_4	
	$c_i(1)$	$I_i(1)$	$c_i(2)$	$I_i(2)$	$c_i(3)$	$I_i(3)$	$c_i(4)$	$I_i(4)$
111	0,770	-8,212	0,458	-1,802	0,307	-0,008	2,448	-0,002
121	0,610	-8,419	0,393	-1,820	40,172	0,037	34,169	0,011
131	0,727	-8,268	92,127	22,914	$2,5 \times 10^{-4}$	-0,008	0,003	-0,003
211	0,801	-8,173	0,454	-1,804	0,015	-0,008	3,003	-0,002
222	0,797	-8,178	0,455	-1,803	0,009	-0,008	3,341	-0,002
225	0,801	-8,173	0,454	-1,804	0,015	-0,008	3,003	-0,002
231	90,306	107,170	2,525	-1,245	$1,7 \times 10^{-4}$	-0,008	0,024	-0,003
232	0,801	-8,173	0,454	-1,804	0,015	-0,008	3,003	-0,002
233	0,801	-8,173	0,454	-1,804	0,015	-0,008	3,003	-0,002
234	0,801	-8,173	0,454	-1,803	0,015	-0,008	3,038	-0,002
331	0,551	-8,495	0,430	-1,810	45,285	0,043	40,077	0,014
332	0,676	-8,333	0,449	-1,805	11,511	0,005	1,538	-0,002
333	0,758	-8,229	0,439	-1,807	2,627	-0,005	0,350	-0,003
334	0,801	-8,173	0,454	-1,804	0,015	-0,008	3,003	-0,002

Tabela 14 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_5 , λ_6 , λ_7 e λ_8 (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro)

Obs.	λ_5		λ_6		λ_7		λ_8	
	c_i (5)	I_i (5)	c_i (6)	I_i (6)	c_i (7)	I_i (7)	c_i (8)	I_i (8)
111	0,764	-0,001	30,234	$7,4 \times 10^{-4}$	23,872	$3,7 \times 10^{-6}$	33,929	$3,2 \times 10^{-6}$
121	17,427	0,001	0,029	$-2,3 \times 10^{-4}$	0,039	$-1,6 \times 10^{-6}$	0,018	$-8,6 \times 10^{-7}$
131	$1,4 \times 10^{-4}$	-0,001	$1,5 \times 10^{-5}$	$-2,3 \times 10^{-4}$	$5,5 \times 10^{-8}$	$-1,6 \times 10^{-6}$	$3,9 \times 10^{-7}$	$-8,6 \times 10^{-7}$
211	0,502	-0,001	0,001	$-2,3 \times 10^{-4}$	0,015	$-1,6 \times 10^{-6}$	4,260	$-3,5 \times 10^{-7}$
222	0,034	-0,001	0,005	$-2,3 \times 10^{-4}$	64,734	$1,3 \times 10^{-5}$	23,332	$2,0 \times 10^{-6}$
225	0,502	-0,001	0,001	$-2,3 \times 10^{-4}$	0,015	$-1,6 \times 10^{-6}$	4,260	$-3,5 \times 10^{-7}$
231	$7,0 \times 10^{-4}$	-0,001	$5,6 \times 10^{-4}$	$-2,3 \times 10^{-4}$	$1,5 \times 10^{-6}$	$-1,6 \times 10^{-6}$	$4,8 \times 10^{-6}$	$-8,6 \times 10^{-7}$
232	0,502	-0,001	0,001	$-2,3 \times 10^{-4}$	0,015	$-1,6 \times 10^{-6}$	4,260	$-3,5 \times 10^{-7}$
233	0,502	-0,001	0,001	$-2,3 \times 10^{-4}$	0,015	$-1,6 \times 10^{-6}$	4,260	$-3,5 \times 10^{-7}$
234	0,512	-0,001	$1,6 \times 10^{-4}$	$-2,3 \times 10^{-4}$	0,075	$-1,6 \times 10^{-6}$	7,245	$1,2 \times 10^{-8}$
331	0,471	-0,001	3,594	$-1,1 \times 10^{-4}$	1,123	$-1,3 \times 10^{-6}$	1,324	$-7,0 \times 10^{-7}$
332	78,278	0,009	0,217	$-2,2 \times 10^{-4}$	0,146	$-1,6 \times 10^{-6}$	0,042	$-8,6 \times 10^{-7}$
333	$6,6 \times 10^{-4}$	-0,001	65,914	$1,9 \times 10^{-3}$	9,937	$6,2 \times 10^{-7}$	12,810	$6,8 \times 10^{-7}$
334	0,502	-0,001	0,001	$-2,3 \times 10^{-4}$	0,015	$-1,6 \times 10^{-6}$	4,260	$-3,5 \times 10^{-7}$

Tabela 15 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_9 , λ_{10} , λ_{11} e λ_{12} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro)

Obs.	λ_9		λ_{10}		λ_{11}		λ_{12}	
	c_i (9)	I_i (9)	c_i (10)	I_i (10)	c_i (11)	I_i (11)	c_i (12)	I_i (12)
111	0,076	$-2,3 \times 10^{-8}$	28,197	$5,7 \times 10^{-31}$	24,830	$2,1 \times 10^{-31}$	34,450	$8,8 \times 10^{-32}$
121	$1,9 \times 10^{-4}$	$-2,3 \times 10^{-8}$	0,064	$-1,9 \times 10^{-31}$	2,366	$-5,6 \times 10^{-32}$	0,957	$-2,0 \times 10^{-32}$
131	$2,9 \times 10^{-9}$	$-2,3 \times 10^{-8}$	2,842	$-1,2 \times 10^{-31}$	36,681	$3,5 \times 10^{-31}$	0,239	$-2,2 \times 10^{-32}$
211	3,806	$-1,1 \times 10^{-8}$	0,348	$-1,8 \times 10^{-31}$	0,066	$-8,3 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
222	0,150	$-2,2 \times 10^{-8}$	0,348	$-1,8 \times 10^{-31}$	0,263	$-8,1 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
225	3,806	$-1,1 \times 10^{-8}$	0,348	$-1,8 \times 10^{-31}$	0,066	$-8,3 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
231	$3,0 \times 10^{-7}$	$-2,3 \times 10^{-8}$	65,473	$1,6 \times 10^{-30}$	34,766	$3,2 \times 10^{-31}$	61,244	$1,7 \times 10^{-31}$
232	3,806	$-1,1 \times 10^{-8}$	0,348	$-1,8 \times 10^{-31}$	0,066	$-8,3 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
233	3,806	$-1,1 \times 10^{-8}$	0,348	$-1,8 \times 10^{-31}$	0,066	$-8,3 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
234	80,719	$2,4 \times 10^{-7}$	0,348	$-1,8 \times 10^{-31}$	0,066	$-8,3 \times 10^{-32}$	0,957	$-2,0 \times 10^{-32}$
331	$1,7 \times 10^{-3}$	$-2,3 \times 10^{-8}$	0,575	$-1,8 \times 10^{-31}$	0,642	$-7,6 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
332	$9,9 \times 10^{-5}$	$-2,3 \times 10^{-8}$	0,064	$-1,9 \times 10^{-31}$	0,000	$-8,4 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
333	0,022	$-2,3 \times 10^{-8}$	0,348	$-1,8 \times 10^{-31}$	0,058	$-8,3 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$
334	3,806	$-1,1 \times 10^{-8}$	0,348	$-1,8 \times 10^{-31}$	0,066	$-8,3 \times 10^{-32}$	0,239	$-2,2 \times 10^{-32}$

Na Tabela 13, observa-se que, para o primeiro autovalor λ_1 , a postura 231 foi a que apresentou maior contribuição e, conseqüentemente, maior valor para a função de influência.

Considerando o autovalor λ_2 , observa-se que a postura 131 se destacou, apresentando os maiores valores para a contribuição e função de influência.

Para o terceiro e quarto autovalores (λ_3 e λ_4), destacaram-se as posturas 331 e 121, respectivamente. Enquanto que para λ_5 , as posturas que apresentaram maiores valores de contribuição e da função de influência foram as posturas 332 e 121, respectivamente.

Destacam-se as posturas 333 e 111, respectivamente, para o autovalor λ_6 , enquanto que para λ_7 , destacaram-se as posturas 222 e 111, nessa ordem.

Para o oitavo autovalor λ_8 , destacaram-se as posturas 111 e 222, nessa ordem, enquanto que para o autovalor λ_9 se destacou, com maior contribuição e maior valor, para a função de influência, a postura 234.

Podem ser destacadas para o autovalor λ_{10} as posturas 231 e 111, respectivamente, enquanto que para o décimo primeiro autovalor λ_{11} , as posturas que se destacam foram as posturas 131, 231 e 111, respectivamente.

Com relação ao autovalor λ_{12} , as posturas que apresentaram os maiores valores para a contribuição e função de influência foram as posturas 231 e 111, nessa ordem.

Comparando as posturas que apresentaram maiores valores de contribuição e maiores valores para a função de influência, encontradas na análise em que se utilizou a porcentagem de tempo média das 12 atividades, com as posturas aqui encontradas, observa-se que na sua maioria são diferentes, ou seja, a maioria das posturas que apresentaram grande contribuição e maiores valores para a função de influência nessa análise, não são as mesmas que apresentaram anteriormente. Como é o caso da postura 131 para o segundo autovalor, na primeira análise apresentou um valor moderado de contribuição, enquanto que na segunda análise foi a que mais se destacou para esse componente. Além disso, observa-se que a maioria das posturas que apresentaram elevados valores de $c_i(t)$ e $I_i(t)$ anteriormente, apresentam aqui valores muito pequenos.

4.3 Estudo da sensibilidade para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231

Serão apresentados nessa seção, os resultados obtidos, por meio das análises realizadas, com base na retirada das duas posturas, dentre as 14

utilizadas, que apresentaram maiores valores de contribuição na análise feita anteriormente para a atividade aplicação de herbicida no morro.

As posturas retiradas foram a 131 e 231, que foram aquelas que apresentaram maiores valores de contribuição para os dois primeiros componentes principais que foram selecionados.

Tabela 16 Autovalores associados a cada componente principal para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231

Autovalores	
λ_1	3,736492000
λ_2	2,413964000
λ_3	1,730619000
λ_4	1,149401000
λ_5	1,000000000
λ_6	0,646298600
λ_7	0,178480000
λ_8	0,136792500
λ_9	0,007304689
λ_{10}	0,000648252
λ_{11}	$1,141996 \times 10^{-31}$
λ_{12}	$2,824697 \times 10^{-32}$

Na Tabela 16, podem ser observados os autovalores associados a cada componente principal. Assim, como nas análises anteriores e por definição, o primeiro autovalor é aquele que corresponde à maior parte da variabilidade encontrada nas variáveis. O que se observa na Tabela 16, é que a variabilidade que antes era praticamente toda concentrada nos dois primeiros autovalores,

passa a ser distribuída entre um número maior de autovalores com a retirada das duas posturas mais influentes. Observa-se, ainda, que para concentrar o equivalente à porcentagem de variância que antes era descrita pelos dois primeiros componentes principais, torna-se necessária a utilização dos 5 primeiros componentes. Isso pode ser confirmado na Tabela 17, os cinco primeiros componentes principais são responsáveis por, aproximadamente, 91,19% da variância total.

Tabela 17 Desvio padrão, proporção da variância e proporção da variância acumulada explicada por cada componente principal (atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)

Componentes Principais	Desvio Padrão	Proporção da Variância Explicada	Proporção da Variância Acumulada
CP1	2,01900	0,33970	0,33970
CP2	1,62280	0,21950	0,55910
CP3	1,37400	0,15730	0,71650
CP4	1,11980	0,10450	0,82090
CP5	1,04447	0,09091	0,91186
CP6	0,83967	0,05875	0,97062
CP7	0,44125	0,01623	0,98684
CP8	0,38630	0,01244	0,99928
CP9	0,08927	0,00066	0,99994
CP10	0,02659	0,00006	1,00000
CP11	$7,850 \times 10^{-17}$	0	1,00000
CP12	$2,434 \times 10^{-17}$	0	1,00000

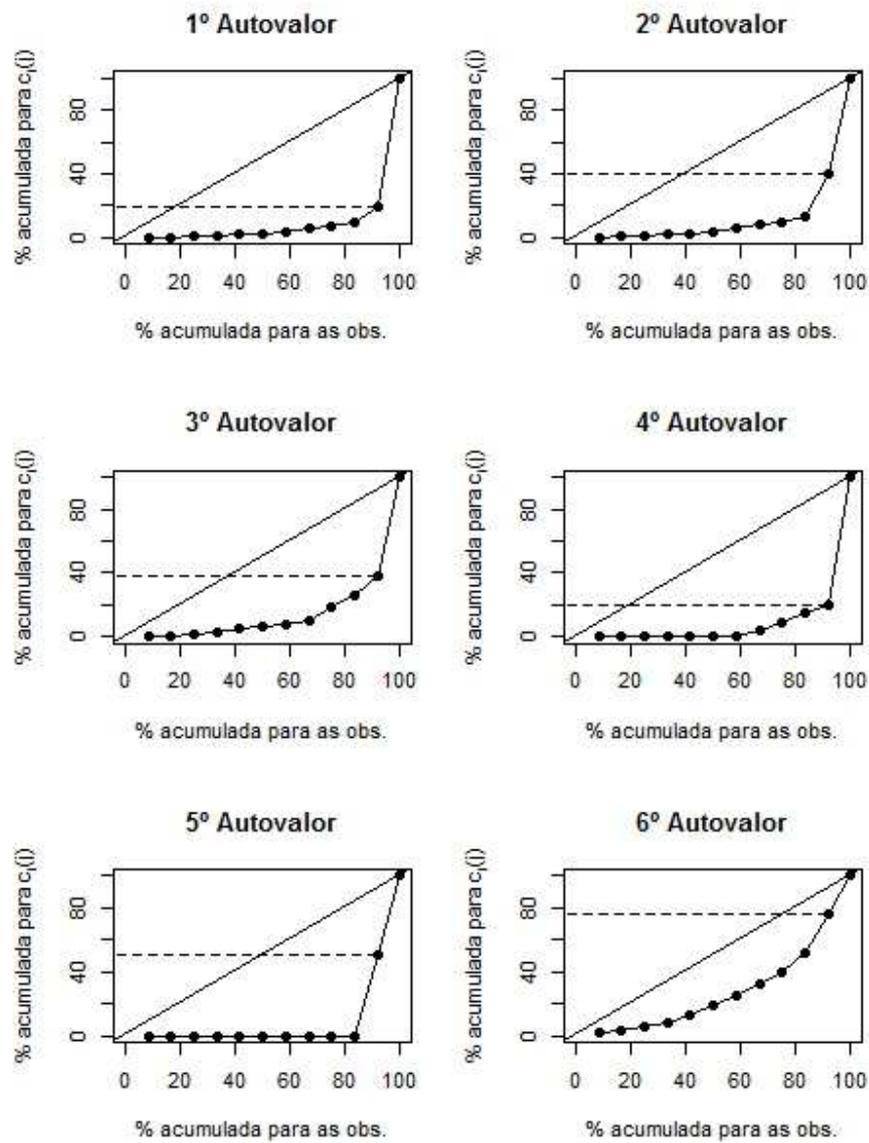


Figura 10 Curvas de Lorenz para os componentes de 1 a 6 (atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)

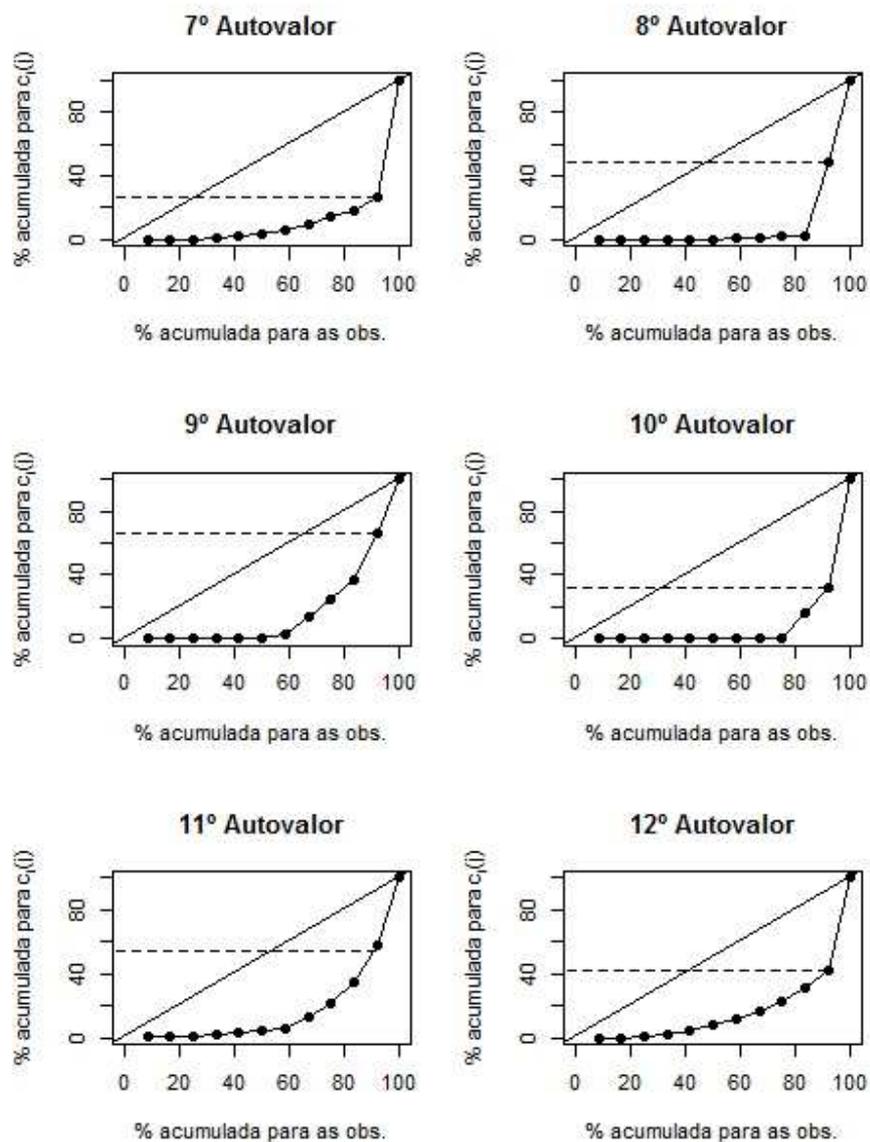


Figura 11 Curvas de Lorenz para os componentes de 7 a 12 (atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)

As Figuras 10 e 11 apresentadas anteriormente mostram as curvas de Lorenz obtidas para cada componente principal.

Na Figura 10, verifica-se pelas curvas de Lorenz, associadas ao primeiro componente, que a observação mais influente responde por, aproximadamente, 80% da variabilidade associada a λ_1 , por outro lado, a observação mais influente para o segundo componente corresponde a, aproximadamente, 60% da variabilidade associada a λ_2 .

Observa-se, ainda, pela curva de Lorenz, associada ao terceiro componente, que a observação mais influente responde por, aproximadamente, 63% da variabilidade associada a λ_3 , enquanto que a observação mais influente para o quarto componente responde por, aproximadamente, 80% da variabilidade associada a λ_4 .

Por meio da curva de Lorenz, associada ao quinto componente, pode ser visto que a observação mais influente representa, aproximadamente, 50% da variabilidade associada a λ_5 , por outro lado, para o sexto componente, a observação mais influente representa, aproximadamente, 25% da variabilidade associada a λ_6 .

Na Figura 11, observa-se pela curva de Lorenz, associada ao sétimo componente, que a observação mais influente responde por, aproximadamente, 75% da variabilidade associada a λ_7 , enquanto que para o oitavo componente, a observação mais influente representa, aproximadamente, 50% da variabilidade associada a λ_8 .

Por meio da curva de Lorenz, associada ao nono componente, pode ser visto que a observação mais influente representa, aproximadamente, 35% da variabilidade associada a λ_9 , enquanto que para o décimo componente, a observação mais influente corresponde a, aproximadamente, 65% da variabilidade associada a λ_{10} .

Observa-se, ainda, pela curva de Lorenz, associada ao décimo primeiro componente, que a observação mais influente representa, aproximadamente, 40% da variabilidade associada a λ_{11} , enquanto que para o décimo segundo componente, a observação mais influente responde por, aproximadamente, 60% da variabilidade associada a λ_{12} .

Na Tabela 18, observam-se os valores do índice de Gini para as 12 curvas de Lorenz obtidas. Comparando os valores aqui encontrados com os valores do índice de Gini encontrados na análise, realizada para a atividade aplicação de herbicida no morro antes da retirada das posturas 131 e 231, observa-se que os valores do índice tiveram um decréscimo, exceto para aqueles correspondentes ao quarto e oitavo componentes principais, que aumentaram em relação aos anteriores.

Esse resultado já era esperado, em virtude do fato de que foram retiradas as duas posturas que apresentaram maiores valores de contribuição para os dois componentes principais selecionados. Com a retirada, a variabilidade tende a diminuir, tornando as contribuições distribuídas de forma um pouco mais uniforme entre as posturas remanescentes. Por outro lado, como pode ser observado nas Tabelas 19, 20 e 21, novas posturas, que antes não haviam apresentado valores elevados de contribuição, passaram a se destacar, como é o caso da postura 331 para o primeiro autovalor, a postura 121 para o segundo, a 233 para o terceiro, a 332 para o quarto e as posturas 225 e 232 para o quinto autovalor.

Tabela 18 Valores do índice de Gini para o estudo da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231

G_j	
G_1	0,8299
G_2	0,7716
G_3	0,7227
G_4	0,8344
G_5	0,8333
G_6	0,4625
G_7	0,7809
G_8	0,8245
G_9	0,6761
G_{10}	0,8344
G_{11}	0,6812
G_{12}	0,6883

Tabela 19 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_1 , λ_2 , λ_3 e λ_4 (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)

Obs.	λ_1		λ_2		λ_3		λ_4	
	c_i (1)	I_i (1)	c_i (2)	I_i (2)	c_i (3)	I_i (3)	c_i (4)	I_i (4)
111	1,830	-2,916	0,113	-2,381	0,468	-1,633	0,041	-1,144
121	9,097	0,342	60,058	14,984	12,140	0,791	4,817	-0,485
211	0,370	-3,571	0,783	-2,187	1,489	-1,421	0,097	-1,136
222	0,414	-3,551	0,705	-2,210	1,648	-1,388	0,106	-1,135
225	0,689	-3,427	2,281	-1,753	8,355	0,004	5,768	-0,354
232	0,689	-3,427	2,281	-1,753	8,355	0,004	5,768	-0,354
233	1,712	-2,969	26,610	5,294	61,459	11,033	0,178	-1,125
234	1,243	-3,179	0,725	-2,204	0,065	-1,717	3,241	-0,702
331	81,132	32,641	3,234	-1,477	1,934	-1,329	0,049	-1,143
332	2,428	-2,648	1,968	-1,844	1,539	-1,411	79,761	9,852
333	0,025	-3,725	0,461	-2,280	1,058	-1,511	0,076	-1,139
334	0,370	-3,571	0,783	-2,187	1,489	-1,421	0,097	-1,136

Tabela 20 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_5 , λ_6 , λ_7 e λ_8 (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)

Obs.	λ_5		λ_6		λ_7		λ_8	
	c_i (5)	I_i (5)	c_i (6)	I_i (6)	c_i (7)	I_i (7)	c_i (8)	I_i (8)
111	$2,4 \times 10^{-30}$	-1,000	3,482	-0,376	0,970	-0,158	51,258	0,705
121	$5,6 \times 10^{-30}$	-1,000	1,775	-0,509	3,716	-0,099	0,034	-0,136
211	$7,2 \times 10^{-30}$	-1,000	7,294	-0,081	4,510	-0,082	0,337	-0,131
222	$2,4 \times 10^{-30}$	-1,000	6,131	-0,171	2,545	-0,124	0,247	-0,133
225	50,000	5,000	24,352	1,242	0,213	-0,174	0,008	-0,137
232	50,000	5,000	24,352	1,242	0,213	-0,174	0,008	-0,137
233	$5,9 \times 10^{-31}$	-1,000	1,662	-0,517	0,043	-0,178	0,002	-0,137
234	$3,1 \times 10^{-29}$	-1,000	11,975	0,282	73,583	1,397	0,745	-0,125
331	$3,7 \times 10^{-31}$	-1,000	1,535	-0,527	0,302	-0,172	0,888	-0,122
332	$2,1 \times 10^{-28}$	-1,000	3,897	-0,344	1,642	-0,143	0,032	-0,136
333	$9,5 \times 10^{-30}$	-1,000	6,251	-0,161	7,754	-0,012	46,104	0,620
334	$7,2 \times 10^{-30}$	-1,000	7,294	-0,081	4,510	-0,082	0,337	-0,131

Tabela 21 Comparação entre as contribuições e os valores da função de influência para os autovalores λ_9 , λ_{10} , λ_{11} e λ_{12} (análise da porcentagem de tempo para a atividade aplicação de herbicida no morro com a retirada das posturas 131 e 231)

Obs.	λ_9		λ_{10}		λ_{11}		λ_{12}	
	$c_i(9)$	$I_i(9)$	$c_i(10)$	$I_i(10)$	$c_i(11)$	$I_i(11)$	$c_i(12)$	$I_i(12)$
111	33,459	0,022	$4,5 \times 10^{-2}$	$-6,4 \times 10^{-4}$	0,172	$-1,1 \times 10^{-31}$	0,909	$-2,5 \times 10^{-32}$
121	$6,5 \times 10^{-3}$	-0,007	$2,4 \times 10^{-2}$	$-6,4 \times 10^{-4}$	7,761	$-7,8 \times 10^{-33}$	2,045	$-2,1 \times 10^{-32}$
211	10,954	0,002	15,834	$5,8 \times 10^{-4}$	1,268	$-9,7 \times 10^{-32}$	3,636	$-1,6 \times 10^{-32}$
222	12,031	0,003	67,840	$4,6 \times 10^{-3}$	0,285	$-1,1 \times 10^{-31}$	0	$-2,8 \times 10^{-32}$
225	$5,9 \times 10^{-4}$	-0,007	$6,7 \times 10^{-6}$	$-6,5 \times 10^{-4}$	23,052	$2,0 \times 10^{-31}$	5,682	$-9,0 \times 10^{-33}$
232	$5,9 \times 10^{-4}$	-0,007	$6,7 \times 10^{-6}$	$-6,5 \times 10^{-4}$	0,791	$-1,0 \times 10^{-31}$	8,182	$-5,1 \times 10^{-34}$
233	$1,5 \times 10^{-4}$	-0,007	$1,7 \times 10^{-6}$	$-6,5 \times 10^{-4}$	8,095	$-3,3 \times 10^{-33}$	58,182	$1,7 \times 10^{-31}$
234	$3,9 \times 10^{-2}$	-0,007	$5,0 \times 10^{-2}$	$-6,4 \times 10^{-4}$	1,549	$-9,3 \times 10^{-32}$	0	$-2,8 \times 10^{-32}$
331	2,589	-0,005	$3,5 \times 10^{-3}$	$-6,5 \times 10^{-4}$	42,513	$4,7 \times 10^{-31}$	5,682	$-9,0 \times 10^{-33}$
332	$7,2 \times 10^{-2}$	-0,007	$3,3 \times 10^{-1}$	$-6,2 \times 10^{-4}$	13,074	$6,5 \times 10^{-32}$	0,909	$-2,5 \times 10^{-32}$
333	29,896	0,019	$4,1 \times 10^{-2}$	$-6,5 \times 10^{-4}$	0,172	$-1,1 \times 10^{-31}$	11,136	$9,5 \times 10^{-33}$
334	10,954	0,002	15,834	$5,8 \times 10^{-4}$	1,268	$-9,7 \times 10^{-32}$	3,636	$-1,6 \times 10^{-32}$

As tabelas 19, 20 e 21 apresentam as contribuições $c_i(j)$ em conjunto com os valores $I_i(j)$ da função de influência empírica para $i = 1, 2, \dots, 14$ e $j = 1, 2, \dots, 12$, tornando possível a confirmação dos resultados observados anteriormente pelas curvas de Lorenz e do associado índice de Gini, assim como a identificação das posturas que se destacaram com valores elevados de contribuição.

Na tabela 19, observa-se que, para o primeiro autovalor λ_1 , a postura 331 foi a que apresentou maior contribuição e, conseqüentemente, maior valor para a função de influência.

Considerando o autovalor λ_2 , observa-se que a postura 121 se destacou, apresentando os maiores valores para a contribuição e função de influência.

Para o terceiro autovalor λ_3 , destacou-se a postura 233, enquanto que para λ_4 , a postura que apresentou maior valor de contribuição e da função de influência foi a postura 332.

Destacam-se as posturas 225 e 232 para os autovalores λ_5 e λ_6 .

Para o sétimo autovalor λ_7 , destaca-se a postura 234, enquanto que para os autovalores λ_8 e λ_9 destacaram-se com maiores valores de contribuição e maiores valores para a função de influência, as posturas 111 e 333.

Podem ser destacadas para o autovalor λ_{10} a postura 222, enquanto que para o décimo primeiro autovalor λ_{11} , a postura que se destacou apresentando o maior valor de contribuição e maior valor para a função de influência foi a postura 331. Já para o décimo segundo autovalor λ_{12} , destacou-se a postura 233.

É importante ressaltar que, dentre os resultados obtidos, observou-se que as combinações posturais que se destacaram, com valores elevados de contribuição, não foram as mesmas para todos os doze autovalores obtidos. Além disso, observou-se que, mesmo com a retirada daquelas combinações posturais consideradas mais influentes, destacaram-se outras, que antes não apresentavam contribuição significativa para os componentes em estudo.

Por outro lado, com a retirada das duas observações que apresentaram maiores valores de contribuição na análise realizada para a atividade aplicação de herbicida no morro, observou-se que os valores do índice de Gini sofreram um decréscimo para a maioria dos componentes principais, o que indica que as contribuições se tornaram um pouco mais homogêneas em relação às posturas remanescentes.

5 CONCLUSÃO

A metodologia proposta neste trabalho, aplicada ao estudo comportamental de trabalhadores em lavouras de café do Sul de Minas Gerais, permitiu identificar as observações discrepantes, cuja interpretação revela as posturas que exerceram maior influência nas atividades realizadas pelo trabalhador, uma vez que os trabalhadores foram considerados como variáveis e as posturas como unidades amostrais na organização dos dados para análise de componentes principais e sua sensibilidade.

REFERÊNCIAS

- ABRAHÃO, J. I. Ergonomia: modelo, métodos e técnicas. In: CONGRESSO LATINO AMERICANO DE ERGONOMIA, 2.; SEMINÁRIO BRASILEIRO DE ERGONOMIA, 6., 1993, Florianópolis. **Anais...** Florianópolis: UFSC, 1993. 1 CD-ROM.
- ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE CAFÉ. **A história do café: origem e trajetória.** São Paulo, 2007. Disponível em: <http://www.abic.com.br/scafe_historia.html>. Acesso em: 20 ago. 2013.
- BARBOSA, M. A. G. **Caracterização da carga física de trabalho na cafeicultura do sul de Minas Gerais.** 2013. 192 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual de Campinas, Campinas, 2013.
- BÉNASSÉNI, J. A concentration study of principal components. **Journal of Applied Statistics**, Abingdon, v. 32, n. 9, p. 947-957, Sept. 2005.
- CRITCHLEY, F. Influence in principal components analysis. **Biometrika**, London, v. 72, p. 627-636, 1985.
- CROUX, C.; HAESBROECK, G. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. **Biometrika**, London, v. 87, n. 3, p. 603-618, 2000.
- CROUX, C.; RUIZ-GAZEN, A. **A fast algorithm for principal components based on projection pursuit.** Berlin: Springer-Verlag, 1996. 216 p.
- DEVLIN, S.J.; GNANDESIKAN, R.; KETTENRING, J.R. Robust estimation of dispersion matrices and principal components. **Journal of the American Statistical Association**, New York, v. 76, p. 354-362, 1981.
- DUL, J.; WEERDMEESTER, B. **Ergonomia prática I.** São Paulo: E. Blucher, 2004. 137 p.
- FALZON, P. (Ed.). **Ergonomia.** São Paulo: Blucher, 2007. 640 p.
- FERREIRA, D. F. **Estatística multivariada.** 2. ed. Lavras: UFLA, 2011. 676 p.

GIRI, N. C. **Multivariate statistical analysis**. 2nded. New York: M. Dekker, 2004. 550 p.

HAMPEL, F. R. The influence curve and its role in robust estimation. **Journal of the American Statistical Association**, New York, v. 69, p. 383-393, 1974.

HUBER, P. J. **Robust statistics**. New York: Wiley, 1981. 320 p.

HUBERT, M.; ROUSSEEUW, P.J.; BRANDEN, K. van den. ROBPCA: a new approach to robust principal component analysis. **Technometrics**, Washington, v. 47, n. 1, p. 64-79, 2005.

IBAZIZEN, M.; DAUXOIS, J. A robust principal component analysis. **Statistics**, Abingdon, v. 37, n. 1, p. 73-83, 2003.

IIDA, I. **Ergonomia: projeto e produção**. 2.ed. São Paulo: E. Blucher, 2005. 614 p.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6thed. Upper Saddle River: Pearson Prentice Hall, 2007. 773 p.

JOLLIFFE, I. T. **Principal component analysis**. 2nded. New York: Springer-Verlag, 2002. 487 p. (Springer Series in Statistics).

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005. 297 p.

PINHEIRO, A.K. da S.; FRANÇA, M. B. A. **Ergonomia aplicada à anatomia e à fisiologia do trabalhador**. Goiânia: AB, 2006. 192 p.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013. Software.

RADHAKRISHNAN, R.; KSHIRSAGAR, A.M. Influence functions for certain parameters in multivariate analysis. **Communications in Statistics A**, New York, v. 10, p. 515-529, 1981.

SOFTWARE e hardware CAPTIV L 300: aquisição de dados e análise de sistemas. Ergobuyer: TEAERGO, 1998. Software.

WARKEN,C.L. et al. Análise ergonômica da colheita em um centro experimental de beneficiamento coletivo do café cerejeira. In: SIMPÓSIO ACADÊMICO DE ENGENHARIA DE PRODUÇÃO, 3., 2007, Viçosa, MG.**Anais...**Viçosa, MG: UFV, 2007.1 CD-ROM.

WEDLAKE, R. S. **Robust principal component analysis biplots**. 2008. 349 f. Dissertation (Master of Science in Mathematical Statistics) - University of Stellenbosch, Stellenbosch, 2008.